## Preface

- About This Guide ................................................................. 1
- Document Conventions .......................................................... 2
  - Text Conventions ............................................................... 2
  - Command Syntax Conventions ............................................ 3
- Getting Support ........................................................................ 3
  - Product information ............................................................ 3
  - Technical support .............................................................. 4

## Section I: Introduction to Greenplum

### Chapter 1: About the Greenplum Architecture .................. 6
- About the Greenplum Master ................................................. 7
- About the Greenplum Segments ............................................ 7
- About the Greenplum Interconnect ...................................... 7
- About Redundancy and Failover in Greenplum Database ......... 8
  - About Segment Mirroring ................................................ 8
  - About Master Mirroring .................................................. 9
  - About Interconnect Redundancy ...................................... 9
- About Parallel Data Loading .............................................. 10
- About Management and Monitoring .................................. 10

### Chapter 2: About Distributed Databases ......................... 12
- Understanding How Data is Stored ..................................... 12
- Understanding Greenplum Distribution Policies .................. 13

### Chapter 3: Summary of Greenplum Features .................... 14
- Greenplum SQL Standard Conformance ............................. 14
  - Core SQL Conformance ................................................ 14
  - SQL 1992 Conformance ................................................. 15
  - SQL 1999 Conformance ................................................. 16
  - SQL 2003 Conformance ................................................. 17
  - SQL 2008 Conformance ................................................. 17
- Greenplum and PostgreSQL Compatibility ....................... 18

### Chapter 4: About Greenplum Query Processing ............... 25
- Understanding Query Planning and Dispatch ....................... 25
- Understanding Greenplum Query Plans .............................. 26
- Understanding Parallel Query Execution ............................ 27

## Section II: Access Control and Security

### Chapter 5: Managing Roles and Privileges ....................... 30
- Creating New Roles (Users) ............................................... 31
  - Altering Role Attributes ............................................... 31
- Creating Groups (Role Membership) ................................ 32
- Managing Object Privileges ............................................ 33
  - Simulating Row and Column Level Access Control ............ 34
- Encrypting Data .............................................................. 34
Chapter 6: Configuring Client Authentication
Allowing Connections to Greenplum Database ........................................... 36
   Editing the pg_hba.conf File .................................................................. 37
Limiting Concurrent Connections ............................................................... 38
Encrypting Client/Server Connections ...................................................... 39

Chapter 7: Accessing the Database
Establishing a Database Session ............................................................... 40
Supported Client Applications .................................................................. 41
   Greenplum Database Client Applications ........................................... 41
   pgAdmin III for Greenplum Database .................................................. 42
   Database Application Interfaces ......................................................... 45
   Third-Party Client Tools ................................................................. 46
Troubleshooting Connection Problems ..................................................... 47

Chapter 8: Managing Workload and Resources
Overview of Greenplum Workload Management .................................... 48
   How Resource Queues Work in Greenplum Database ...................... 48
Steps to Enable Workload Management .................................................. 51
Configuring Workload Management ....................................................... 52
Creating Resource Queues ..................................................................... 53
   Creating Queues with an Active Query Limit ................................... 53
   Creating Queues with a Query Planner Cost Limits ......................... 53
Setting Priority Levels .......................................................................... 54
Assigning Roles (Users) to a Resource Queue ........................................ 55
   Removing a Role from a Resource Queue ......................................... 55
Modifying Resource Queues .................................................................. 55
   Altering a Resource Queue ............................................................. 56
   Dropping a Resource Queue .......................................................... 56
Checking Resource Queue Status .......................................................... 56
   Viewing Queued Statements and Resource Queue Status .............. 56
   Viewing Resource Queue Statistics ................................................. 57
   Viewing the Roles Assigned to a Resource Queue ......................... 57
   Viewing the Waiting Queries for a Resource Queue ...................... 57
   Clearing a Waiting Statement From a Resource Queue ................. 58
   Viewing the Priority of Active Statements .................................... 58
   Resetting the Priority of an Active Statement ............................... 59

Section III: Database Administration

Chapter 9: Defining Database Objects
Creating and Managing Databases .......................................................... 61
   About Template Databases ............................................................... 61
   Creating a Database ........................................................................ 61
   Viewing the List of Databases .......................................................... 62
   Altering a Database ........................................................................ 62
   Dropping a Database ...................................................................... 62
Creating and Managing Tablespaces ........................................................ 63
   Creating a Filespace ....................................................................... 63
   Creating a Tablespace ..................................................................... 64
   Using a Tablespace to Store Database Objects .............................. 64
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Viewing Existing Tablespaces and Filespaces</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>Dropping Tablespaces and Filespaces</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>Creating and Managing Schemas</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>The Default ‘Public’ Schema</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Creating a Schema</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Schema Search Paths</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Dropping a Schema</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>System Schemas</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Creating and Managing Tables</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Creating a Table</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Altering a Table</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Dropping a Table</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Partitioning Large Tables</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Understanding Table Partitioning in Greenplum Database</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Deciding on a Table Partitioning Strategy</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Creating Partitioned Tables</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Loading Partitioned Tables</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>Verifying Your Partition Strategy</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>Viewing Your Partition Design</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>Maintaining Partitioned Tables</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>Creating and Using Sequences</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Creating a Sequence</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Using a Sequence</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Altering a Sequence</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Dropping a Sequence</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Using Indexes in Greenplum Database</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Index Types</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Creating an Index</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Examining Index Usage</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Managing Indexes</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>Dropping an Index</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>Creating and Managing Views</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>Creating Views</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>Dropping Views</td>
<td>93</td>
</tr>
<tr>
<td>Chapter 10: Managing Data</td>
<td></td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>About Concurrency Control in Greenplum Database</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Inserting New Rows</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Updating Existing Rows</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>Deleting Rows</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>Truncating a Table</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>Working With Transactions</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>Transaction Isolation Levels</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>Vacuuming the Database</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Configuring the Free Space Map</td>
<td>100</td>
</tr>
<tr>
<td>Chapter 11: Querying Data</td>
<td></td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>Defining Queries</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>SQL Lexicon</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>SQL Value Expressions</td>
<td>101</td>
</tr>
</tbody>
</table>
# Greenplum Database Administrator Guide 4.0 - Contents

Using Functions and Operators......................................................110  
Using Functions in Greenplum Database.................................110  
User-Defined Functions............................................................111  
Built-in Functions and Operators..............................................111  
Query Profiling ..............................................................................119  
Reading EXPLAIN Output .........................................................119  
Reading EXPLAIN ANALYZE Output ..........................................120  
What to Look for in a Query Plan .............................................122

## Chapter 12: Loading and Unloading Data

About the Greenplum Database Loading Tools..............................124  
About External Tables..............................................................124  
About gpload ...........................................................................125  
About COPY .............................................................................125  
Formatting Your Data Files............................................................125  
Formatting Rows......................................................................125  
Formatting Columns ................................................................126  
Representing NULL Values .......................................................126  
Escaping ..................................................................................126  
Character Encoding..................................................................128  
Loading Data into Greenplum Database ........................................128  
Creating and Using File-Based External Tables.........................128  
Creating and Using Web External Tables..................................132  
Loading Data Using an External Table......................................133  
Handling Load Errors .............................................................133  
Loading Data with gpload ........................................................136  
Loading Data with COPY ..........................................................137  
Data Loading Performance Tips ................................................138  
Unloading Data from Greenplum Database.................................138  
Defining a File-Based Writable External Table..........................138  
Defining a Command-Based Writable External Web Table........139  
Unloading Data Using a Writable External Table ......................139  
Unloading Data Using COPY .....................................................140

## Section IV: System Administration

### Chapter 13: Starting and Stopping Greenplum

Overview.......................................................................................142  
Starting Greenplum Database .......................................................142  
Restarting Greenplum Database ....................................................142  
Uploading Configuration File Changes Only ..............................143  
Starting the Master in Maintenance Mode ................................143  
Stopping Greenplum Database.......................................................143

### Chapter 14: Configuring Your Greenplum System

About Greenplum Master and Local Parameters ..........................145  
Setting Configuration Parameters..............................................145  
Setting a Local Configuration Parameter ................................146  
Setting a Master Configuration Parameter ...............................146  
Viewing Settings of Server Configuration Parameters ...............147  
Configuration Parameter Categories ........................................147
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection and Authentication Parameters</td>
<td>148</td>
</tr>
<tr>
<td>System Resource Consumption Parameters</td>
<td>149</td>
</tr>
<tr>
<td>Query Tuning Parameters</td>
<td>150</td>
</tr>
<tr>
<td>Error Reporting and Logging Parameters</td>
<td>152</td>
</tr>
<tr>
<td>System Monitoring Parameters</td>
<td>152</td>
</tr>
<tr>
<td>Runtime Statistics Collection Parameters</td>
<td>153</td>
</tr>
<tr>
<td>Automatic Statistics Collection Parameters</td>
<td>153</td>
</tr>
<tr>
<td>Client Connection Default Parameters</td>
<td>154</td>
</tr>
<tr>
<td>Lock Management Parameters</td>
<td>154</td>
</tr>
<tr>
<td>Workload Management Parameters</td>
<td>154</td>
</tr>
<tr>
<td>External Table Parameters</td>
<td>155</td>
</tr>
<tr>
<td>Append-Only Table Parameters</td>
<td>155</td>
</tr>
<tr>
<td>Database and Tablespace/Filespace Parameters</td>
<td>155</td>
</tr>
<tr>
<td>Past PostgreSQL Version Compatibility Parameters</td>
<td>155</td>
</tr>
<tr>
<td>Greenplum Array Configuration Parameters</td>
<td>155</td>
</tr>
<tr>
<td><strong>Chapter 15: Enabling High Availability Features</strong></td>
<td>157</td>
</tr>
<tr>
<td>Overview of High Availability in Greenplum Database</td>
<td>157</td>
</tr>
<tr>
<td>Overview of Segment Mirroring</td>
<td>157</td>
</tr>
<tr>
<td>Overview of Master Mirroring</td>
<td>158</td>
</tr>
<tr>
<td>Overview of Fault Detection and Recovery</td>
<td>159</td>
</tr>
<tr>
<td>Enabling Mirroring in Greenplum Database</td>
<td>160</td>
</tr>
<tr>
<td>Enabling Segment Mirroring</td>
<td>160</td>
</tr>
<tr>
<td>Enabling Master Mirroring</td>
<td>161</td>
</tr>
<tr>
<td>Knowing When a Segment is Down</td>
<td>162</td>
</tr>
<tr>
<td>Enabling Alerts and Notifications</td>
<td>162</td>
</tr>
<tr>
<td>Checking for Failed Segments</td>
<td>162</td>
</tr>
<tr>
<td>Checking the Log Files</td>
<td>163</td>
</tr>
<tr>
<td>Recovering a Failed Segment</td>
<td>163</td>
</tr>
<tr>
<td>Recovering From Segment Failures</td>
<td>164</td>
</tr>
<tr>
<td>Recovering a Failed Master</td>
<td>167</td>
</tr>
<tr>
<td>Restoring Master Mirroring After a Recovery</td>
<td>168</td>
</tr>
<tr>
<td><strong>Chapter 16: Backing Up and Restoring Databases</strong></td>
<td>170</td>
</tr>
<tr>
<td>Overview of Backup and Restore Operations</td>
<td>170</td>
</tr>
<tr>
<td>About Parallel Backups</td>
<td>170</td>
</tr>
<tr>
<td>About Non-Parallel Backups</td>
<td>171</td>
</tr>
<tr>
<td>About Parallel Restores</td>
<td>171</td>
</tr>
<tr>
<td>About Non-Parallel Restores</td>
<td>172</td>
</tr>
<tr>
<td>Backing Up a Database</td>
<td>172</td>
</tr>
<tr>
<td>Backing Up a Database with gp_dump</td>
<td>173</td>
</tr>
<tr>
<td>Automating Parallel Backups with gpcrondump</td>
<td>174</td>
</tr>
<tr>
<td>Restoring From Parallel Backup Files</td>
<td>175</td>
</tr>
<tr>
<td>Restoring a Database with gp_restore</td>
<td>175</td>
</tr>
<tr>
<td>Restoring a Database Using gpdbrestore</td>
<td>176</td>
</tr>
<tr>
<td>Restoring to a Different Greenplum System Configuration</td>
<td>177</td>
</tr>
<tr>
<td><strong>Chapter 17: Expanding a Greenplum System</strong></td>
<td>179</td>
</tr>
<tr>
<td>Planning Greenplum System Expansion</td>
<td>179</td>
</tr>
<tr>
<td>System Expansion Overview</td>
<td>179</td>
</tr>
<tr>
<td>System Expansion Checklist</td>
<td>181</td>
</tr>
</tbody>
</table>
Planning New Hardware Platforms ........................................... 182
Planning Initialization of New Segments ............................... 182
Planning Table Redistribution................................................... 183
Preparing and Adding Nodes ......................................................... 186
Adding New Nodes to the Trusted Host Environment .......... 186
Verifying OS Settings............................................................... 188
Validating Disk I/O and Memory Bandwidth ...................... 188
Integrating New Hardware into the System ..................... 189
Initializing New Segments............................................................. 189
Creating an Input File for System Expansion ...................... 189
Running gpexpand to Initialize New Segments ............ 192
Rolling Back a Failed Expansion Setup ........................ 193
Redistributing Tables..................................................................... 193
Ranking Tables for Redistribution ............................................ 193
Redistributing Tables Using gpexpand...................................... 194
Monitoring Table Redistribution................................................ 194
Removing the Expansion Schema.................................................. 195

Chapter 18: Monitoring a Greenplum System ....................... 196
Monitoring Database Activity and Performance ..................... 196
Monitoring System State............................................................... 196
Enabling System Alerts and Notifications ....................... 197
Checking System State................................................................. 204
Checking Disk Space Usage ..................................................... 204
Checking for Data Distribution Skew ......................... 205
Viewing Metadata Information about Database Objects ....... 206
Viewing the Database Server Log Files ................................. 207
Log File Format........................................................................ 207
Searching the Greenplum Database Server Log Files ........... 209
Using gp_toolkit ............................................................................ 209

Chapter 19: Routine System Maintenance Tasks ................. 210
Routine Vacuum and Analyze .................................................... 210
Transaction ID Management ................................................... 210
System Catalog Maintenance ................................................... 210
Vacuum and Analyze for Query Optimization ............... 211
Routine Reindexing ................................................................. 211
Managing Greenplum Database Log Files ................. 212
Database Server Log Files ....................................................... 212
Management Utility Log Files ................................................... 212

Section V: Performance Tuning

Chapter 20: Defining Database Performance ....................... 214
Understanding the Performance Factors ............................ 214
System Resources ................................................................. 214
Workload .............................................................................. 214
Throughput............................................................................. 214
Contention........................................................................... 215
Optimization........................................................................ 215
Determining Acceptable Performance............................. 215
<table>
<thead>
<tr>
<th>Command</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALTER EXTERNAL TABLE</td>
<td>280</td>
</tr>
<tr>
<td>ALTER FILESPACE</td>
<td>283</td>
</tr>
<tr>
<td>ALTER FUNCTION</td>
<td>284</td>
</tr>
<tr>
<td>ALTER GROUP</td>
<td>287</td>
</tr>
<tr>
<td>ALTER INDEX</td>
<td>288</td>
</tr>
<tr>
<td>ALTER LANGUAGE</td>
<td>290</td>
</tr>
<tr>
<td>ALTER OPERATOR</td>
<td>291</td>
</tr>
<tr>
<td>ALTER OPERATOR CLASS</td>
<td>292</td>
</tr>
<tr>
<td>ALTER RESOURCE QUEUE</td>
<td>293</td>
</tr>
<tr>
<td>ALTER ROLE</td>
<td>295</td>
</tr>
<tr>
<td>ALTER SCHEMA</td>
<td>298</td>
</tr>
<tr>
<td>ALTER SEQUENCE</td>
<td>299</td>
</tr>
<tr>
<td>ALTER TABLE</td>
<td>302</td>
</tr>
<tr>
<td>ALTER TABLESPACE</td>
<td>314</td>
</tr>
<tr>
<td>ALTER TRIGGER</td>
<td>315</td>
</tr>
<tr>
<td>ALTER TYPE</td>
<td>316</td>
</tr>
<tr>
<td>ALTER USER</td>
<td>317</td>
</tr>
<tr>
<td>ANALYZE</td>
<td>318</td>
</tr>
<tr>
<td>BEGIN</td>
<td>320</td>
</tr>
<tr>
<td>CHECKPOINT</td>
<td>322</td>
</tr>
<tr>
<td>CLOSE</td>
<td>323</td>
</tr>
<tr>
<td>CLUSTER</td>
<td>324</td>
</tr>
<tr>
<td>COMMENT</td>
<td>327</td>
</tr>
<tr>
<td>COMMIT</td>
<td>330</td>
</tr>
<tr>
<td>COPY</td>
<td>331</td>
</tr>
<tr>
<td>CREATE AGGREGATE</td>
<td>339</td>
</tr>
<tr>
<td>CREATE CAST</td>
<td>343</td>
</tr>
<tr>
<td>CREATE CONVERSION</td>
<td>346</td>
</tr>
<tr>
<td>CREATE DATABASE</td>
<td>348</td>
</tr>
<tr>
<td>CREATE DOMAIN</td>
<td>350</td>
</tr>
<tr>
<td>CREATE EXTERNAL TABLE</td>
<td>352</td>
</tr>
<tr>
<td>CREATE FUNCTION</td>
<td>365</td>
</tr>
<tr>
<td>CREATE GROUP</td>
<td>371</td>
</tr>
<tr>
<td>CREATE INDEX</td>
<td>372</td>
</tr>
<tr>
<td>CREATE LANGUAGE</td>
<td>376</td>
</tr>
<tr>
<td>CREATE OPERATOR</td>
<td>380</td>
</tr>
<tr>
<td>CREATE OPERATOR CLASS</td>
<td>385</td>
</tr>
<tr>
<td>CREATE RESOURCE QUEUE</td>
<td>390</td>
</tr>
<tr>
<td>CREATE ROLE</td>
<td>393</td>
</tr>
<tr>
<td>CREATE RULE</td>
<td>398</td>
</tr>
<tr>
<td>CREATE SCHEMA</td>
<td>401</td>
</tr>
<tr>
<td>CREATE SEQUENCE</td>
<td>403</td>
</tr>
<tr>
<td>CREATE TABLE</td>
<td>407</td>
</tr>
<tr>
<td>CREATE TABLE AS</td>
<td>418</td>
</tr>
<tr>
<td>CREATE TABLESPACE</td>
<td>422</td>
</tr>
<tr>
<td>CREATE TRIGGER</td>
<td>424</td>
</tr>
<tr>
<td>CREATE TYPE</td>
<td>427</td>
</tr>
<tr>
<td>CREATE USER</td>
<td>434</td>
</tr>
<tr>
<td>CREATE VIEW</td>
<td>435</td>
</tr>
<tr>
<td>Command</td>
<td>Page</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------</td>
</tr>
<tr>
<td>DEALLOCATE</td>
<td>438</td>
</tr>
<tr>
<td>DECLARE</td>
<td>439</td>
</tr>
<tr>
<td>DELETE</td>
<td>442</td>
</tr>
<tr>
<td>DROP AGGREGATE</td>
<td>445</td>
</tr>
<tr>
<td>DROP CAST</td>
<td>446</td>
</tr>
<tr>
<td>DROP CONVERSION</td>
<td>447</td>
</tr>
<tr>
<td>DROP DATABASE</td>
<td>448</td>
</tr>
<tr>
<td>DROP DOMAIN</td>
<td>449</td>
</tr>
<tr>
<td>DROP EXTERNAL TABLE</td>
<td>450</td>
</tr>
<tr>
<td>DROP FILESPACE</td>
<td>452</td>
</tr>
<tr>
<td>DROP FUNCTION</td>
<td>453</td>
</tr>
<tr>
<td>DROP GROUP</td>
<td>455</td>
</tr>
<tr>
<td>DROP INDEX</td>
<td>456</td>
</tr>
<tr>
<td>DROP LANGUAGE</td>
<td>457</td>
</tr>
<tr>
<td>DROP OPERATOR</td>
<td>458</td>
</tr>
<tr>
<td>DROP OPERATOR CLASS</td>
<td>460</td>
</tr>
<tr>
<td>DROP OWNED</td>
<td>462</td>
</tr>
<tr>
<td>DROP RESOURCE QUEUE</td>
<td>464</td>
</tr>
<tr>
<td>DROP ROLE</td>
<td>466</td>
</tr>
<tr>
<td>DROP RULE</td>
<td>467</td>
</tr>
<tr>
<td>DROP SCHEMA</td>
<td>468</td>
</tr>
<tr>
<td>DROP SEQUENCE</td>
<td>469</td>
</tr>
<tr>
<td>DROP TABLE</td>
<td>470</td>
</tr>
<tr>
<td>DROP TABLESPACE</td>
<td>471</td>
</tr>
<tr>
<td>DROP TRIGGER</td>
<td>472</td>
</tr>
<tr>
<td>DROP TYPE</td>
<td>473</td>
</tr>
<tr>
<td>DROP USER</td>
<td>474</td>
</tr>
<tr>
<td>DROP VIEW</td>
<td>475</td>
</tr>
<tr>
<td>END</td>
<td>476</td>
</tr>
<tr>
<td>EXECUTE</td>
<td>477</td>
</tr>
<tr>
<td>EXPLAIN</td>
<td>478</td>
</tr>
<tr>
<td>FETCH</td>
<td>481</td>
</tr>
<tr>
<td>GRANT</td>
<td>485</td>
</tr>
<tr>
<td>INSERT</td>
<td>490</td>
</tr>
<tr>
<td>LOAD</td>
<td>492</td>
</tr>
<tr>
<td>LOCK</td>
<td>493</td>
</tr>
<tr>
<td>MOVE</td>
<td>497</td>
</tr>
<tr>
<td>PREPARE</td>
<td>499</td>
</tr>
<tr>
<td>REASSIGN OWNED</td>
<td>502</td>
</tr>
<tr>
<td>REINDEX</td>
<td>503</td>
</tr>
<tr>
<td>RELEASE SAVEPOINT</td>
<td>505</td>
</tr>
<tr>
<td>RESET</td>
<td>506</td>
</tr>
<tr>
<td>REVOKE</td>
<td>507</td>
</tr>
<tr>
<td>ROLLBACK</td>
<td>510</td>
</tr>
<tr>
<td>ROLLBACK TO SAVEPOINT</td>
<td>511</td>
</tr>
<tr>
<td>SAVEPOINT</td>
<td>513</td>
</tr>
<tr>
<td>SELECT</td>
<td>515</td>
</tr>
<tr>
<td>SELECT INTO</td>
<td>531</td>
</tr>
<tr>
<td>SET</td>
<td>533</td>
</tr>
</tbody>
</table>
SET ROLE .......................................................................................... 535
SET SESSION AUTHORIZATION .................................................... 537
SET TRANSACTION ........................................................................... 539
SHOW .............................................................................................. 542
START TRANSACTION .................................................................... 543
TRUNCATE ....................................................................................... 545
UPDATE .......................................................................................... 546
VACUUM ......................................................................................... 550
VALUES .......................................................................................... 553

Appendix B: Management Utility Reference .................................. 556
  Backend Server Programs .......................................................... 557
  Management Utility Summary ................................................... 558
  gp_dump ..................................................................................... 579
  gp_restore .................................................................................. 584
  gpaddmirrors ............................................................................ 588
  gpactivatestandby ........................................................................ 593
  gpbitmapreindex ........................................................................ 596
  gpcheck ..................................................................................... 598
  gpcheckperf .............................................................................. 600
  gpconfig ................................................................................... 604
  gpcrondump .............................................................................. 608
  gpdbrestore ................................................................................ 613
  gpdeletelesystem ....................................................................... 616
  gpdetective ............................................................................... 618
  gpexpand ................................................................................... 621
  gpfdist ..................................................................................... 625
  gpfilespase ............................................................................... 628
  gpinitstandby ............................................................................ 631
  gpinitssystem ............................................................................. 634
  gpload ....................................................................................... 641
  gplogfilter ............................................................................... 652
  gpmapreduce ............................................................................. 655
  gpmigrator ................................................................................ 658
  gpperfmon_install .................................................................... 663
  gprecoversseg .......................................................................... 665
  gpscp ......................................................................................... 669
  gseginstall ............................................................................... 671
  gpsetupsanfailover .................................................................... 674
  gpssnmpd .................................................................................. 677
  gpssh ......................................................................................... 680
  gpssh-exkeys .............................................................................. 682
  gpstart ....................................................................................... 685
  gpstate ...................................................................................... 688
  gpstop ....................................................................................... 692
  gpssys1 .................................................................................... 695

Appendix C: Client Utility Reference ............................................. 696
  Client Utility Summary .............................................................. 698
  clusterdb ................................................................................... 707
createdb .......................................................... 709
createlang .......................................................... 711
createuser ............................................................ 713
dropdb ................................................................. 716
gp_db_interfaces .................................................... 718
droplang ............................................................... 719
dropuser ............................................................... 721
ecpg ................................................................. 723
pg_config ............................................................. 725
pg_dump .............................................................. 728
pg_dumpall .......................................................... 735
pg_restore ........................................................... 739
psql ................................................................. 744
reindexdb ............................................................ 765
vacuumdb ............................................................ 767

Appendix D: Server Configuration Parameters .......................... 770
add_missing_from ................................................. 772
array_nulls ......................................................... 772
authentication_timeout ........................................ 772
backslash_quote ................................................... 772
block_size .......................................................... 772
bonjour_name ..................................................... 772
check_function_bodies ........................................ 772
client_encoding .................................................. 772
client_min_messages ........................................... 772
cpu_index_tuple_cost .......................................... 773
cpu_operator_cost ............................................... 773
cpu_tuple_cost ................................................... 773
cursor_tuple_fraction ......................................... 773
custom_variable_classes ..................................... 773
DateStyle ........................................................... 773
db_user_namespace .............................................. 773
deadlock_timeout ................................................ 774
debug_assertions ............................................... 774
debug_pretty_print ............................................. 774
debug_print_parse ............................................... 774
debug_print_plan ............................................... 774
debug_print_prelim_plan ..................................... 774
debug_print_rewritten ....................................... 774
debug_print_slice_table ..................................... 774
default_statistics_target ..................................... 774
default_tablespace ............................................. 774
default_transaction_isolation .............................. 775
default_transaction_read_only ............................. 775
dynamic_library_path ........................................ 775
effective_cache_size .......................................... 775
enable_bitmapscan ............................................. 775
enable_groupagg ............................................... 775
<table>
<thead>
<tr>
<th>Configuration Parameter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>enable_hashagg</td>
<td>775</td>
</tr>
<tr>
<td>enable_hashjoin</td>
<td>776</td>
</tr>
<tr>
<td>enable_indexscan</td>
<td>776</td>
</tr>
<tr>
<td>enable_mergejoin</td>
<td>776</td>
</tr>
<tr>
<td>enable_nestloop</td>
<td>776</td>
</tr>
<tr>
<td>enable_seqscan</td>
<td>776</td>
</tr>
<tr>
<td>enable_sort</td>
<td>776</td>
</tr>
<tr>
<td>enable_tidscan</td>
<td>776</td>
</tr>
<tr>
<td>escape_string_warning</td>
<td>776</td>
</tr>
<tr>
<td>explain_pretty_print</td>
<td>776</td>
</tr>
<tr>
<td>extra_float_digits</td>
<td>777</td>
</tr>
<tr>
<td>fromCollapse_limit</td>
<td>777</td>
</tr>
<tr>
<td>gp_adjust_selectivity_for_outerjoins</td>
<td>777</td>
</tr>
<tr>
<td>gp_analyze_relative_error</td>
<td>777</td>
</tr>
<tr>
<td>gp_autostats_mode</td>
<td>777</td>
</tr>
<tr>
<td>gp_autostats_on_change_threshold</td>
<td>778</td>
</tr>
<tr>
<td>gp_cached_segworkers_threshold</td>
<td>778</td>
</tr>
<tr>
<td>gp_command_count</td>
<td>778</td>
</tr>
<tr>
<td>gp_connections_per_thread</td>
<td>778</td>
</tr>
<tr>
<td>gp_content</td>
<td>778</td>
</tr>
<tr>
<td>gp_dbid</td>
<td>778</td>
</tr>
<tr>
<td>gp_debug_linger</td>
<td>778</td>
</tr>
<tr>
<td>gp_email_from</td>
<td>778</td>
</tr>
<tr>
<td>gp_email_password</td>
<td>779</td>
</tr>
<tr>
<td>gp_email_smtp_server</td>
<td>779</td>
</tr>
<tr>
<td>gp_email_smtp_userid</td>
<td>779</td>
</tr>
<tr>
<td>gp_email_to</td>
<td>779</td>
</tr>
<tr>
<td>gp_enable_adaptive_nestloop</td>
<td>779</td>
</tr>
<tr>
<td>gp_enable_agg_distinct</td>
<td>779</td>
</tr>
<tr>
<td>gp_enable_agg_distinct_pruning</td>
<td>779</td>
</tr>
<tr>
<td>gp_enable_direct_dispatch</td>
<td>779</td>
</tr>
<tr>
<td>gp_enable_fast_sri</td>
<td>780</td>
</tr>
<tr>
<td>gp_enable_gpperfmon</td>
<td>780</td>
</tr>
<tr>
<td>gp_enable_grouppext_distinct_gather</td>
<td>780</td>
</tr>
<tr>
<td>gp_enable_grouppext_distinct_pruning</td>
<td>780</td>
</tr>
<tr>
<td>gp_enable_multiphase_agg</td>
<td>780</td>
</tr>
<tr>
<td>gp_enable_predicate_propagation</td>
<td>780</td>
</tr>
<tr>
<td>gp_enable_preunique</td>
<td>780</td>
</tr>
<tr>
<td>gp_enable_sequential_window_plans</td>
<td>781</td>
</tr>
<tr>
<td>gp_enable_sort_distinct</td>
<td>781</td>
</tr>
<tr>
<td>gp_enable_sort_limit</td>
<td>781</td>
</tr>
<tr>
<td>gp_external_enable_exec</td>
<td>781</td>
</tr>
<tr>
<td>gp_external_grant_privileges</td>
<td>781</td>
</tr>
<tr>
<td>gp_external_max_segs</td>
<td>781</td>
</tr>
<tr>
<td>gp_fts_probe_interval</td>
<td>781</td>
</tr>
<tr>
<td>gp_fts_probe_threadcount</td>
<td>781</td>
</tr>
<tr>
<td>gp_fts_probe_timeout</td>
<td>781</td>
</tr>
<tr>
<td>gp_gpperfmon_send_interval</td>
<td>782</td>
</tr>
<tr>
<td>Topic</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>gp_hashjoin_tuples_per_bucket</td>
<td>782</td>
</tr>
<tr>
<td>gp_interconnect_hash_multiplier</td>
<td>782</td>
</tr>
<tr>
<td>gp_interconnect_queue_depth</td>
<td>782</td>
</tr>
<tr>
<td>gp_interconnect_setup_timeout</td>
<td>782</td>
</tr>
<tr>
<td>gp_interconnect_type</td>
<td>782</td>
</tr>
<tr>
<td>gp_log_format</td>
<td>782</td>
</tr>
<tr>
<td>gp_max_csv_line_length</td>
<td>782</td>
</tr>
<tr>
<td>gp_max_databases</td>
<td>782</td>
</tr>
<tr>
<td>gp_max_filespaces</td>
<td>782</td>
</tr>
<tr>
<td>gp_max_local_distributed_cache</td>
<td>782</td>
</tr>
<tr>
<td>gp_max_packet_size</td>
<td>782</td>
</tr>
<tr>
<td>gp_max_tablespaces</td>
<td>782</td>
</tr>
<tr>
<td>gp_motion_cost_per_row</td>
<td>782</td>
</tr>
<tr>
<td>gp_num_contents_in_cluster</td>
<td>782</td>
</tr>
<tr>
<td>gp_reject_percent_threshold</td>
<td>782</td>
</tr>
<tr>
<td>gp_reraise_signal</td>
<td>782</td>
</tr>
<tr>
<td>gp_resqueue_priority</td>
<td>782</td>
</tr>
<tr>
<td>gp_resqueue_priority_cpucores_per_segment</td>
<td>782</td>
</tr>
<tr>
<td>gp_resqueue_priority_sweeper_interval</td>
<td>782</td>
</tr>
<tr>
<td>gp_role</td>
<td>782</td>
</tr>
<tr>
<td>gp_safefswritesize</td>
<td>782</td>
</tr>
<tr>
<td>gp_segment_connect_timeout</td>
<td>782</td>
</tr>
<tr>
<td>gp_segments_for_planner</td>
<td>782</td>
</tr>
<tr>
<td>gp_session_id</td>
<td>782</td>
</tr>
<tr>
<td>gp_set_proc_affinity</td>
<td>782</td>
</tr>
<tr>
<td>gp_set_read_only</td>
<td>782</td>
</tr>
<tr>
<td>gp_snmp_community</td>
<td>782</td>
</tr>
<tr>
<td>gp_snmp_monitor_address</td>
<td>782</td>
</tr>
<tr>
<td>gp_snmp_use_inform_or_trap</td>
<td>782</td>
</tr>
<tr>
<td>gp_statistics_pullup_from_child_partition</td>
<td>782</td>
</tr>
<tr>
<td>gp_statistics_use_fkeys</td>
<td>782</td>
</tr>
<tr>
<td>gp_vmem_idle_resource_timeout</td>
<td>782</td>
</tr>
<tr>
<td>gp_vmem_protect_limit</td>
<td>782</td>
</tr>
<tr>
<td>gp_vmem_protect_segworker_cache_limit</td>
<td>782</td>
</tr>
<tr>
<td>gp_workfile_compress_algorithm</td>
<td>782</td>
</tr>
<tr>
<td>gpperfmon_port</td>
<td>782</td>
</tr>
<tr>
<td>integer_datetimes</td>
<td>782</td>
</tr>
<tr>
<td>IntervalStyle</td>
<td>782</td>
</tr>
<tr>
<td>joinCollapse_limit</td>
<td>782</td>
</tr>
<tr>
<td>krb_caseins_users</td>
<td>782</td>
</tr>
<tr>
<td>krb_server_keyfile</td>
<td>782</td>
</tr>
<tr>
<td>krb_srvname</td>
<td>782</td>
</tr>
<tr>
<td>lc_collate</td>
<td>782</td>
</tr>
<tr>
<td>lc_ctype</td>
<td>782</td>
</tr>
<tr>
<td>lc_messages</td>
<td>782</td>
</tr>
<tr>
<td>lc_monetary</td>
<td>782</td>
</tr>
<tr>
<td>lc_numeric</td>
<td>782</td>
</tr>
<tr>
<td>lc_time</td>
<td>782</td>
</tr>
<tr>
<td>listen_addresses</td>
<td>782</td>
</tr>
<tr>
<td>Configuration Setting</td>
<td>Page</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>local_preload_libraries</td>
<td>788</td>
</tr>
<tr>
<td>log_autostats</td>
<td>788</td>
</tr>
<tr>
<td>log_connections</td>
<td>788</td>
</tr>
<tr>
<td>log_disconnections</td>
<td>789</td>
</tr>
<tr>
<td>log_dispatch_stats</td>
<td>789</td>
</tr>
<tr>
<td>log_duration</td>
<td>789</td>
</tr>
<tr>
<td>log_error_verbosity</td>
<td>789</td>
</tr>
<tr>
<td>log_executor_stats</td>
<td>789</td>
</tr>
<tr>
<td>log_hostname</td>
<td>789</td>
</tr>
<tr>
<td>log_min_duration_statement</td>
<td>789</td>
</tr>
<tr>
<td>log_min_error_statement</td>
<td>789</td>
</tr>
<tr>
<td>log_min_messages</td>
<td>790</td>
</tr>
<tr>
<td>log_parser_stats</td>
<td>790</td>
</tr>
<tr>
<td>log_planner_stats</td>
<td>790</td>
</tr>
<tr>
<td>log_rotation_age</td>
<td>790</td>
</tr>
<tr>
<td>log_rotation_size</td>
<td>790</td>
</tr>
<tr>
<td>log_statement</td>
<td>790</td>
</tr>
<tr>
<td>log_statement_stats</td>
<td>790</td>
</tr>
<tr>
<td>log_timezone</td>
<td>790</td>
</tr>
<tr>
<td>log_truncate_on_rotation</td>
<td>791</td>
</tr>
<tr>
<td>maintenance_work_mem</td>
<td>791</td>
</tr>
<tr>
<td>max_appendonly_tables</td>
<td>791</td>
</tr>
<tr>
<td>max_connections</td>
<td>791</td>
</tr>
<tr>
<td>max_files_per_process</td>
<td>791</td>
</tr>
<tr>
<td>max_fsm_pages</td>
<td>792</td>
</tr>
<tr>
<td>max_fsm_relations</td>
<td>792</td>
</tr>
<tr>
<td>max_function_args</td>
<td>792</td>
</tr>
<tr>
<td>max_identifier_length</td>
<td>792</td>
</tr>
<tr>
<td>max_index_keys</td>
<td>792</td>
</tr>
<tr>
<td>max_locks_per_transaction</td>
<td>792</td>
</tr>
<tr>
<td>max_prepared_transactions</td>
<td>792</td>
</tr>
<tr>
<td>max_resource_portals_per_transaction</td>
<td>792</td>
</tr>
<tr>
<td>max_resource_queues</td>
<td>792</td>
</tr>
<tr>
<td>max_stack_depth</td>
<td>793</td>
</tr>
<tr>
<td>max_work_mem</td>
<td>793</td>
</tr>
<tr>
<td>password_encryption</td>
<td>793</td>
</tr>
<tr>
<td>pljava_classpath</td>
<td>793</td>
</tr>
<tr>
<td>pljava_statement_cache_size</td>
<td>793</td>
</tr>
<tr>
<td>pljava_release_lingering_savepoints</td>
<td>793</td>
</tr>
<tr>
<td>pljava_vmoptions</td>
<td>793</td>
</tr>
<tr>
<td>port</td>
<td>794</td>
</tr>
<tr>
<td>random_page_cost</td>
<td>794</td>
</tr>
<tr>
<td>regex_flavor</td>
<td>794</td>
</tr>
<tr>
<td>resource_cleanup_gangs_on_wait</td>
<td>794</td>
</tr>
<tr>
<td>resource_select_only</td>
<td>794</td>
</tr>
<tr>
<td>search_path</td>
<td>794</td>
</tr>
<tr>
<td>seq_page_cost</td>
<td>794</td>
</tr>
<tr>
<td>server_encoding</td>
<td>795</td>
</tr>
<tr>
<td>server_version</td>
<td>795</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>server_version_num</td>
<td>795</td>
</tr>
<tr>
<td>shared_buffers</td>
<td>795</td>
</tr>
<tr>
<td>shared_preload_libraries</td>
<td>795</td>
</tr>
<tr>
<td>ssl</td>
<td>795</td>
</tr>
<tr>
<td>ssl_ciphers</td>
<td>795</td>
</tr>
<tr>
<td>standard_conforming_strings</td>
<td>795</td>
</tr>
<tr>
<td>statement_timeout</td>
<td>795</td>
</tr>
<tr>
<td>stats_queue_level</td>
<td>796</td>
</tr>
<tr>
<td>superuser_reserved_connections</td>
<td>796</td>
</tr>
<tr>
<td>tcp_keepalives_count</td>
<td>796</td>
</tr>
<tr>
<td>tcp_keepalives_idle</td>
<td>796</td>
</tr>
<tr>
<td>tcp_keepalives_interval</td>
<td>796</td>
</tr>
<tr>
<td>temp_buffers</td>
<td>796</td>
</tr>
<tr>
<td>TimeZone</td>
<td>796</td>
</tr>
<tr>
<td>timezone_abbreviations</td>
<td>796</td>
</tr>
<tr>
<td>track_activities</td>
<td>797</td>
</tr>
<tr>
<td>track_counts</td>
<td>797</td>
</tr>
<tr>
<td>transaction_isolation</td>
<td>797</td>
</tr>
<tr>
<td>transaction_read_only</td>
<td>797</td>
</tr>
<tr>
<td>transform_null_equals</td>
<td>797</td>
</tr>
<tr>
<td>unix_socket_directory</td>
<td>797</td>
</tr>
<tr>
<td>unix_socket_group</td>
<td>797</td>
</tr>
<tr>
<td>unix_socket_permissions</td>
<td>797</td>
</tr>
<tr>
<td>update_process_title</td>
<td>797</td>
</tr>
<tr>
<td>vacuum_cost_delay</td>
<td>797</td>
</tr>
<tr>
<td>vacuum_cost_limit</td>
<td>798</td>
</tr>
<tr>
<td>vacuum_cost_page_dirty</td>
<td>798</td>
</tr>
<tr>
<td>vacuum_cost_page_hit</td>
<td>798</td>
</tr>
<tr>
<td>vacuum_cost_page_miss</td>
<td>798</td>
</tr>
<tr>
<td>vacuum_freeze_min_age</td>
<td>798</td>
</tr>
<tr>
<td>work_mem</td>
<td>798</td>
</tr>
</tbody>
</table>

**Appendix E: Greenplum MapReduce Specification** .......................... 799
  Greenplum MapReduce Document Format ........................................ 799
  Greenplum MapReduce Document Schema ....................................... 801
  Example Greenplum MapReduce Document....................................... 811
  MapReduce Flow Diagram ................................................................ 818

**Appendix F: Greenplum Environment Variables** .............................. 819
  Required Environment Variables ................................................ 819
  Optional Environment Variables ............................................... 820

**Appendix G: Greenplum Database Data Types** ................................ 822

**Appendix H: System Catalog Reference** ....................................... 825
  gp_configuration_history ....................................................... 828
  gp_distributed_log .................................................................. 829
  gp_distributed_xacts ............................................................. 830
  gp_distribution_policy ......................................................... 831
  gp_fastsequence ..................................................................... 832
  gp_fault_strategy .................................................................... 833
Checking Database Object Sizes and Disk Space.................933
  gp_size_of_all_table_indexes ..............................................934
  gp_size_of_database ........................................................934
  gp_size_of_index .............................................................934
  gp_size_of_partition_and_indexes_disk ..................................935
  gp_size_of_schema_disk ..................................................935
  gp_size_of_table_and_indexes_disk ....................................935
  gp_size_of_table_and_indexes_licensing ..................................936
  gp_size_of_table_disk ....................................................936
  gp_size_of_table_and_indexes_uncompressed .............................936
  gp_disk_free ......................................................................937
Checking for Uneven Data Distribution ..................................937
  gp_skew_coefficients .........................................................937
  gp_skew_idle_fractions .......................................................938

Appendix J: Character Set Support ..............................................939
  Setting the Character Set .......................................................940
  Character Set Conversion Between Server and Client ..................941

Appendix K: SQL 2008 Optional Feature Compliance ..................944

Glossary ................................................................................965

Index ..................................................................................974
Preface

This guide provides information for system administrators and database superusers responsible for administering a Greenplum Database system.

- About This Guide
- Document Conventions
- Getting Support

About This Guide

This guide provides information and instructions for configuring, maintaining and using a Greenplum Database system. This guide is intended for system and database administrators responsible for managing a Greenplum Database system.

This guide assumes knowledge of Linux/UNIX system administration, database management systems, database administration, and structured query language (SQL).

Because Greenplum Database is based on PostgreSQL 8.2.14, this guide assumes some familiarity with PostgreSQL. Links and cross-references to PostgreSQL documentation are provided throughout this guide for features that are similar to those in Greenplum Database.

This guide contains the following main sections:

- **Section I, “Introduction to Greenplum”** explains the distributed architecture and parallel processing concepts of Greenplum Database.
- **Section II, “Access Control and Security”** explains how clients connect to a Greenplum Database system, and how to configure access control and workload management.
- **Section III, “Database Administration”** explains how to do basic database administration tasks such as defining database objects, loading data, writing queries and managing data.
- **Section IV, “System Administration”** explains the various system administration tasks of Greenplum Database such as configuring the server, monitoring system activity, enabling high-availability, backing up and restoring databases, and other routine system administration tasks.
- **Section V, “Performance Tuning”** provides guidance on identifying and troubleshooting the most common causes of performance issues in Greenplum Database.
- **Section VI, “Extending Greenplum Database”** describes how to extend the functionality of Greenplum Database by developing your own functions and programs.
- **Section VII, “References”** contains reference documentation for SQL commands, command-line utilities, client programs, system catalogs, and configuration parameters.
**Document Conventions**

The following conventions are used throughout the Greenplum Database documentation to help you identify certain types of information.

- **Text Conventions**
- **Command Syntax Conventions**

### Text Conventions

<table>
<thead>
<tr>
<th>Text Convention</th>
<th>Usage</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>bold</strong></td>
<td>Button, menu, tab, page, and field names in GUI applications</td>
<td>Click <strong>Cancel</strong> to exit the page without saving your changes.</td>
</tr>
<tr>
<td><strong>italics</strong></td>
<td>New terms where they are defined Database objects, such as schema, table, or columns names</td>
<td>The <strong>master instance</strong> is the <strong>postgres</strong> process that accepts client connections. Catalog information for Greenplum Database resides in the <strong>pg_catalog</strong> schema.</td>
</tr>
<tr>
<td><strong>monospace</strong></td>
<td>File names and path names Programs and executables Command names and syntax Parameter names</td>
<td>Edit the <strong>postgresql.conf</strong> file. Use <strong>gpstart</strong> to start Greenplum Database.</td>
</tr>
<tr>
<td><strong>monospace italics</strong></td>
<td>Variable information within file paths and file names Variable information within command syntax</td>
<td><code>/home/gpadmin/config_file</code> <code>COPY tablename FROM 'filename'</code></td>
</tr>
<tr>
<td><strong>monospace bold</strong></td>
<td>Used to call attention to a particular part of a command, parameter, or code snippet.</td>
<td>Change the host name, port, and database name in the JDBC connection URL: <code>jdbc:postgresql://host:5432/mydb</code></td>
</tr>
<tr>
<td><strong>UPPERCASE</strong></td>
<td>Environment variables SQL commands Keyboard keys</td>
<td>Make sure that the Java <code>/bin</code> directory is in your <code>PATH</code>. <code>SELECT * FROM my_table;</code> Press <code>CTRL+C</code> to escape.</td>
</tr>
</tbody>
</table>
Command Syntax Conventions

Table 0.2 Command Syntax Conventions

<table>
<thead>
<tr>
<th>Text Convention</th>
<th>Usage</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>{ }</td>
<td>Within command syntax, curly braces group related command options. Do not type the curly braces.</td>
<td>FROM { 'filename'</td>
</tr>
<tr>
<td>[ ]</td>
<td>Within command syntax, square brackets denote optional arguments. Do not type the brackets.</td>
<td>TRUNCATE [ TABLE ] name</td>
</tr>
<tr>
<td>...</td>
<td>Within command syntax, an ellipsis denotes repetition of a command, variable, or option. Do not type the ellipsis.</td>
<td>DROP TABLE name [, ...]</td>
</tr>
<tr>
<td></td>
<td>Within command syntax, the pipe symbol denotes an “OR” relationship. Do not type the pipe symbol.</td>
<td>VACUUM [ FULL</td>
</tr>
<tr>
<td>$ system_command</td>
<td>Denotes a command prompt - do not type the prompt symbol. $ and # denote terminal command prompts. =&gt; and =&gt; denote Greenplum Database interactive program command prompts (psql or gpssh, for example).</td>
<td>$ createdb mydatabase # chown gpadmin -R /datadir =&gt; SELECT * FROM mytable; #=&gt; SELECT * FROM pg_database;</td>
</tr>
<tr>
<td># root_system_command =&gt; gpdb_command =&gt;# su_gpdb_command</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Getting Support

EMC support, product, and licensing information can be obtained as follows.

Product information

For documentation, release notes, software updates, or for information about EMC products, licensing, and service, go to the EMC Powerlink website (registration required) at:

http://Powerlink.EMC.com
Technical support

For technical support, go to Powerlink and choose Support. On the Support page, you will see several options, including one for making a service request. Note that to open a service request, you must have a valid support agreement. Please contact your EMC sales representative for details about obtaining a valid support agreement or with questions about your account.
Greenplum Database is a massively parallel processing (MPP) database server based on PostgreSQL open-source technology. MPP (also known as a \textit{shared nothing} architecture) refers to systems with two or more processors which cooperate to carry out an operation - each processor with its own memory, operating system and disks. Greenplum leverages this high-performance system architecture to distribute the load of multi-terabyte data warehouses, and is able to use all of a system’s resources in parallel to process a query.

Greenplum Database is essentially several PostgreSQL database instances acting together as one cohesive database management system. It is based on PostgreSQL 8.2.14, and in most cases is very similar to PostgreSQL with regards to SQL support, features, configuration options, and end-user functionality. Database users interact with Greenplum Database as they would a regular PostgreSQL DBMS.

The internals of PostgreSQL have been modified or supplemented to support the parallel structure of Greenplum Database. For example the system catalog, query planner, optimizer, query executor, and transaction manager components have been modified and enhanced to be able to execute queries in parallel across all of the PostgreSQL database instances at once. The Greenplum \textit{interconnect} (the networking layer) enables communication between the distinct PostgreSQL instances and allows the system to behave as one logical database.

Greenplum Database also includes features designed to optimize PostgreSQL for business intelligence (BI) workloads. For example, Greenplum has added parallel data loading (external tables), resource management, query optimizations and storage enhancements which are not found in regular PostgreSQL. Many features and optimizations developed by Greenplum do make their way back into the PostgreSQL community. For example, table partitioning is a feature developed by Greenplum which is now in standard PostgreSQL.

To learn more about Greenplum Database, refer to the following topics:

- About the Greenplum Architecture
- About Distributed Databases
- About Greenplum Query Processing
- Summary of Greenplum Features
1. About the Greenplum Architecture

Greenplum Database is able to handle the storage and processing of large amounts of data by distributing the load across several servers or hosts. A database in Greenplum is actually an array of individual PostgreSQL databases, all working together to present a single database image. The master is the entry point to the Greenplum Database system. It is the database instance where clients connect and submit SQL statements. The master coordinates the work with the other database instances in the system, the segments, which handle data processing and storage.

This section describes all of the components that comprise a Greenplum Database system, and how they work together:

- About the Greenplum Master
- About the Greenplum Segments
- About the Greenplum Interconnect
- About Redundancy and Failover in Greenplum Database
- About Parallel Data Loading
- About Management and Monitoring
About the Greenplum Master

The master is the entry point to the Greenplum Database system. It is the database process that accepts client connections and processes the SQL commands issued by the users of the system.

Since Greenplum Database is based on PostgreSQL, end-users interact with Greenplum Database (through the master) as they would a typical PostgreSQL database. They can connect to the database using client programs such as psql or application programming interfaces (APIs) such as JDBC or ODBC.

The master is where the global system catalog resides (the set of system tables that contain metadata about the Greenplum Database system itself), however the master does not contain any user data. Data resides only on the segments. The master does the work of authenticating client connections, processing the incoming SQL commands, distributing the work load between the segments, coordinating the results returned by each of the segments, and presenting the final results to the client program.

About the Greenplum Segments

In Greenplum Database, the segments are where the data is stored and where the majority of query processing takes place. User-defined tables and their indexes are distributed across the available number of segments in the Greenplum Database system, each segment containing a distinct portion of the data. Segment instances are the database server processes that serve segments. Users do not interact directly with the segments in a Greenplum Database system, but do so through the master.

In the recommended Greenplum Database hardware configuration, there is one active segment per effective CPU or CPU core. For example, if your segment hosts have two dual-core processors, you would have four primary segments per host.

About the Greenplum Interconnect

The interconnect is the networking layer of Greenplum Database. When a user connects to a database and issues a query, processes are created on each of the segments to handle the work of that query (see “Understanding Parallel Query Execution” on page 27). The interconnect refers to the inter-process communication between the segments, as well as the network infrastructure on which this communication relies. The interconnect uses a standard Gigabit Ethernet switching fabric.

By default, the interconnect uses UDP (User Datagram Protocol) to send messages over the network. The Greenplum software does the additional packet verification and checking not performed by UDP, so the reliability is equivalent to TCP (Transmission Control Protocol), and the performance and scalability exceeds TCP. With TCP, Greenplum has a scalability limit of 1000 segment instances. To remove this limit, UDP is now the default protocol for the interconnect.
About Redundancy and Failover in Greenplum Database

Greenplum Database has deployment options to provide for a system without a single point of failure. This section explains the redundancy components of Greenplum Database.

- About Segment Mirroring
- About Master Mirroring
- About Interconnect Redundancy

About Segment Mirroring

When you deploy your Greenplum Database system, you have the option to configure mirror segments. Mirror segments allow database queries to fail over to a backup segment if the primary segment becomes unavailable. To configure mirroring, you must have enough hosts in your Greenplum Database system so that the secondary segment always resides on a different host than its primary. Figure 1.2 shows how table data is distributed across the segments when mirroring is configured. The mirror segment always resides on a different host than its primary segment.

![Figure 1.2 Data Mirroring in Greenplum Database](image)

Segment Failover and Recovery

When mirroring is enabled in a Greenplum Database system, the system will automatically fail over to the mirror copy whenever a primary copy becomes unavailable. A Greenplum Database system can remain operational if a segment instance or host goes down as long as all portions of data are available on the remaining active segments.

Whenever the master cannot connect to a segment instance, it marks that segment instance as down in the Greenplum Database system catalog and brings up the mirror segment in its place. A failed segment instance will remain out of operation until steps are taken to bring that segment back online. A failed segment can be recovered while the system is up and running. The recovery process only copies over the changes that were missed while the segment was out of operation.
About Master Mirroring

You can also optionally deploy a backup or mirror of the master instance on a separate host from the master node. A backup master host serves as a warm standby in the event of the primary master host becoming unoperational. The standby master is kept up to date by a transaction log replication process, which runs on the standby master host and keeps the data between the primary and standby master hosts synchronized.

If the primary master fails, the log replication process is shutdown, and the standby master can be activated in its place. Upon activation of the standby master, the replicated logs are used to reconstruct the state of the master host at the time of the last successfully committed transaction. The activated standby master effectively becomes the Greenplum Database master, accepting client connections on the master port (which must be set to the same port number on the master host and the backup master host).

Since the master does not contain any user data, only the system catalog tables need to be synchronized between the primary and backup copies. These tables are not updated frequently, but when they are, changes are automatically copied over to the standby master so that it is always kept current with the primary.

![Diagram of Master Mirroring in Greenplum Database](image)

**Figure 1.3** Master Mirroring in Greenplum Database

About Interconnect Redundancy

The interconnect refers to the inter-process communication between the segments, as well as the network infrastructure on which this communication relies. A highly available interconnect can be achieved by deploying dual Gigabit Ethernet switches on your network, and redundant Gigabit connections to the Greenplum Database host servers.
About Parallel Data Loading

One challenge of large scale, multi-terabyte data warehouses is getting large amounts of data loaded within a given maintenance window. Greenplum supports fast, parallel data loading with its external tables feature. External tables can also be accessed in ‘single row error isolation’ mode, allowing administrators to filter out bad rows during a load operation into a separate error table, while still loading properly formatted rows. Administrators can control the acceptable error threshold for a load operation, giving them control over the quality and flow of data into the database.

By using external tables in conjunction with Greenplum Database’s parallel file server (gpfdist), administrators can achieve maximum parallelism and load bandwidth from their Greenplum Database system. Greenplum has demonstrated load rates in excess of 2 TB an hour.

![External Tables Using Greenplum Parallel File Server (gpfdist)](image)

Figure 1.4  External Tables Using Greenplum Parallel File Server (gpfdist)

About Management and Monitoring

Management of a Greenplum Database system is performed using a series of command-line utilities, which are located in $GPHOME/bin. Greenplum provides utilities for the following Greenplum Database administration tasks:

- Installing Greenplum Database on an Array
- Initializing a Greenplum Database System
- Starting and Stopping Greenplum Database
- Adding or Removing a Host
- Expanding the Array and Redistributing Tables among New Segments
- Managing Recovery for Failed Segment Instances
- Managing Failover and Recovery for a Failed Master Instance
• Backing Up and Restoring a Database (in Parallel)
• Loading Data in Parallel
• System State Reporting

Greenplum also provides an optional performance monitoring feature that administrators can install and enable with Greenplum Database. The Greenplum Performance Monitor has agents on each segment host that collect data on query execution and system utilization and store it in a database. Segment agents send their data to the Greenplum master at regular intervals (typically every 15 seconds). Users can query the Greenplum Performance Monitor database to see query and system performance data for both active queries and historical queries. Greenplum Performance Monitor also has a graphical web-based user interface for viewing these performance metrics, which can be installed separately from Greenplum Database.

![Greenplum Performance Monitor Architecture](image.png)

**Figure 1.5** Greenplum Performance Monitor Architecture
2. About Distributed Databases

Greenplum is a distributed database system. This means that the data stored in the database system is physically located on more than one database server (referred to as segments in Greenplum). These individual database servers are connected by a communications network (referred to as the interconnect in Greenplum). An essential feature of a true distributed database is that users and client programs work as if they were accessing one single database on a local machine (in Greenplum, this entry-point database is referred to as the master). The fact that the database is distributed across several machines is seamless to the users of the system.

Understanding How Data is Stored

To understand how Greenplum Database stores data across the various hosts and segment instances, consider the following simple logical database. In Figure 2.1, primary keys are shown in bold font and foreign key relationships are indicated by a line from the foreign key in the referring relation to the primary key of the referenced relation. In data warehouse terminology, this is referred to as a star schema. In this type of database schema, the sale table is usually called a fact table and the other tables (customer, vendor, product) are usually called the dimension tables.

![Figure 2.1 Sample Database Star Schema](image)

In Greenplum Database all tables are distributed, which means a table is divided into non-overlapping sets of rows or parts. Each part resides on a single database known as a segment within the Greenplum Database system. The parts are distributed across all of the available segments using a sophisticated hashing algorithm. Database administrators choose the hash key (one or more table columns) when defining the table.
The Greenplum Database physical database implements the logical database on an array of individual database instances — a master instance and two or more segment instances. The master instance does not contain any user data, only the global catalog tables. The segment instances contain disjoint parts (collections of rows) for each distributed table.

![Figure 2.2 Table Distribution in a Greenplum Database Physical Database](image)

**Understanding Greenplum Distribution Policies**

When you create or alter a table in Greenplum Database, there is an additional DISTRIBUTED clause to define the distribution policy of the table. The distribution policy determines how to divide the rows of a table across the Greenplum segments. Greenplum Database provides two types of distribution policy:

**Hash Distribution** - With hash distribution, one or more table columns is used as the distribution key for the table. The distribution key is used by a hashing algorithm to assign each row to a particular segment. Keys of the same value will always hash to the same segment. Choosing a unique distribution key, such as a primary key, will ensure the most even data distribution. Hash distribution is the default distribution policy for a table. If a DISTRIBUTED clause is not supplied, then either the PRIMARY KEY (if the table has one) or the first column of the table will be used as the table distribution key.

**Random Distribution** - With random distribution, rows are sent to the segments as they come in, cycling across the segments in a round-robin fashion. Rows with columns having the same values will not necessarily be located on the same segment. Although a random distribution ensures even data distribution, there are performance advantages to choosing a hash distribution policy whenever possible.
Greenplum SQL Standard Conformance

The SQL language was first formally standardized in 1986 by the American National Standards Institute (ANSI) as SQL 1986. Subsequent versions of the SQL standard have been released by ANSI and as International Organization for Standardization (ISO) standards: SQL 1989, SQL 1992, SQL 1999, SQL 2003, SQL 2006, and finally SQL 2008, which is the current SQL standard. The official name of the standard is ISO/IEC 9075-14:2008. In general, each new version adds more features, although occasionally features are deprecated or removed.

It is important to note that there are no commercial database systems that are fully compliant with the SQL standard. Greenplum Database is almost fully compliant with the SQL 1992 standard, with most of the features from SQL 1999. Several features from SQL 2003 have also been implemented (most notably the SQL OLAP features).

This section addresses the important conformance issues of Greenplum Database as they relate to the SQL standards. For a feature-by-feature list of Greenplum’s support of the latest SQL standard, see Appendix K, “SQL 2008 Optional Feature Compliance”.

Core SQL Conformance

In the process of building a parallel, shared-nothing database system and query optimizer, certain common SQL constructs are not currently implemented in Greenplum Database. The following SQL constructs are not supported:

1. **UPDATE** statements that update the distribution key columns of a hash-distributed Greenplum table. There is currently no way for the system to redistribute a row to a different segment when its hash value changes.

2. **UPDATE** and **DELETE** statements that require data to move from one segment to another. This restricts the use of joins in update and delete operations to hash-distributed tables that have the same distribution key column(s), and the join condition must specify equality on the distribution key column(s).

3. Correlated subqueries that Greenplum’s parallel optimizer cannot internally rewrite into non-correlated joins. Most simple uses of correlated subqueries do work. Those that do not can be manually rewritten using outer joins.

4. Certain rare cases of multi-row subqueries that Greenplum’s parallel optimizer cannot internally rewrite into equijoins.
5. Some set returning subqueries in \texttt{EXISTS} or \texttt{NOT EXISTS} clauses that Greenplum’s parallel optimizer cannot rewrite into joins.

6. \texttt{UNION ALL} of joined tables with subqueries.

7. Set-returning functions in the \texttt{FROM} clause of a subquery.

8. Backwards scrolling cursors, including the use of \texttt{FETCH PRIOR}, \texttt{FETCH FIRST}, \texttt{FETCH ABOLUTE}, and \texttt{FETCH RELATIVE}.

9. In \texttt{CREATE TABLE} statements (on hash-distributed tables): a \texttt{UNIQUE} or \texttt{PRIMARY KEY} clause must include all of (or a superset of) the distribution key columns. Because of this restriction, only one \texttt{UNIQUE} clause or \texttt{PRIMARY KEY} clause is allowed in a \texttt{CREATE TABLE} statement. \texttt{UNIQUE} or \texttt{PRIMARY KEY} clauses are not allowed on randomly-distributed tables.

10. \texttt{CREATE UNIQUE INDEX} statements that do not contain all of (or a superset of) the distribution key columns. \texttt{CREATE UNIQUE INDEX} is not allowed on randomly-distributed tables.

11. \texttt{VOLATILE} or \texttt{STABLE} functions cannot execute on the segments, and so are generally limited to being passed literal values as the arguments to their parameters.

12. Triggers are not supported since they typically rely on the use of \texttt{VOLATILE} functions.

13. Referential integrity constraints (foreign keys) are not enforced in Greenplum Database. Users can declare foreign keys and this information is kept in the system catalog, however.

14. Sequence manipulation functions \texttt{CURRVAL} and \texttt{LASTVAL}.

15. \texttt{DELETE WHERE CURRENT OF} and \texttt{UPDATE WHERE CURRENT OF} (positioned delete and positioned update operations).

SQL 1992 Conformance

The following features of SQL 1992 are not supported in Greenplum Database:

1. \texttt{NATIONAL CHARACTER (NCHAR) and NATIONAL CHARACTER VARYING (NVARCHAR)}. Users can declare the \texttt{NCHAR} and \texttt{NVARCHAR} types, however they are just synonyms for \texttt{CHAR} and \texttt{VARCHAR} in Greenplum Database.

2. \texttt{CREATE ASSERTION} statement.

3. \texttt{INTERVAL} literals are supported in Greenplum Database, but do not conform to the standard.

4. \texttt{GET DIAGNOSTICS} statement.

5. \texttt{GRANT INSERT} or \texttt{UPDATE} privileges on columns. Privileges can only be granted on tables in Greenplum Database.
6. **GLOBAL TEMPORARY TABLES** and **LOCAL TEMPORARY TABLES**. Greenplum **TEMPORARY TABLES** do not conform to the SQL standard, but many commercial database systems have implemented temporary tables in the same way. Greenplum temporary tables are the same as **VOLATILE TABLES** in Teradata.

7. **UNIQUE** predicate.

8. **MATCH PARTIAL** for referential integrity checks (most likely will not be implemented in Greenplum Database).

---

### SQL 1999 Conformance

The following features of SQL 1999 are not supported in Greenplum Database:

1. Large Object data types: **BLOB**, **CLOB**, **NCLOB**. However, the **BYTEA** and **TEXT** columns can store very large amounts of data in Greenplum Database (hundreds of megabytes).

2. Recursive **WITH** clause or the **WITH RECURSIVE** clause (recursive queries). Non-recursive **WITH** clauses can easily be rewritten by moving the common table expression into the **FROM** clause as a derived table.

3. **MODULE** (SQL client modules).

4. **CREATE PROCEDURE** (**SQL/PSM**). This can be worked around in Greenplum Database by creating a **FUNCTION** that returns **void**, and invoking the function as follows:
   
   ```sql
   SELECT myfunc(args);
   ```

5. The PostgreSQL/Greenplum function definition language (**PL/PGSQL**) is a subset of Oracle’s **PL/SQL**, rather than being compatible with the **SQL/PSM** function definition language. Greenplum Database also supports function definitions written in Python, Perl, and R.

6. **BIT** and **BIT VARYING** data types (intentionally omitted). These were deprecated in SQL 2003, and replaced in SQL 2008.

7. Greenplum supports identifiers up to 63 characters long. The SQL standard requires support for identifiers up to 128 characters long.

8. Prepared transactions (**PREPARE TRANSACTION**, **COMMIT PREPARED**, **ROLLBACK PREPARED**). This also means Greenplum does not support XA Transactions (2 phase commit coordination of database transactions with external transactions).

9. **CHARACTER SET** option on the definition of **CHAR()** or **VARCHAR()** columns.

10. Specification of **CHARACTERS** or **OCTETS (BYTES)** on the length of a **CHAR()** or **VARCHAR()** column. For example, **VARCHAR(15 CHARACTERS)** or **VARCHAR(15 OCTETS)** or **VARCHAR(15 BYTES)**.

11. **CURRENT_SCHEMA** function.
12. `CREATE DISTINCT TYPE` statement. `CREATE DOMAIN` can be used as a work-around in Greenplum.

13. The `explicit table` construct.

---

**SQL 2003 Conformance**

The following features of SQL 2003 are not supported in Greenplum Database:

1. **XML** data type (PostgreSQL does support this).

2. **MERGE** statements.

3. **IDENTITY** columns and the associated `GENERATED ALWAYS/GENERATED BY DEFAULT` clause. The `SERIAL` or `BIGSERIAL` data types are very similar to `INT` or `BIGINT` `GENERATED BY DEFAULT AS IDENTITY`.

4. **MULTISET** modifiers on data types.

5. **ROW** data type.

6. Greenplum Database syntax for using sequences is non-standard. For example, `nextval('seq')` is used in Greenplum instead of the standard `NEXT VALUE FOR seq`.

7. **GENERATED ALWAYS AS** columns. Views can be used as a work-around.

8. The sample clause (`TABLESAMPLE`) on `SELECT` statements. The `random()` function can be used as a work-around to get random samples from tables.

9. **NULLS FIRST/NULLS LAST** clause on `SELECT` statements and subqueries (nulls are always last in Greenplum Database).

10. The `partitioned join tables` construct (`PARTITION BY` in a join).

11. **GRANT** `SELECT` privileges on columns. Privileges can only be granted on tables in Greenplum Database. Views can be used as a work-around.

12. For `CREATE TABLE x (LIKE(y))` statements, Greenplum does not support the `[INCLUDING|EXCLUDING] [DEFAULTS|CONSTRAINTS|INDEXES]` clauses.

13. Greenplum array data types are almost SQL standard compliant with some exceptions. Generally customers should not encounter any problems using them.

---

**SQL 2008 Conformance**

The following features of SQL 2008 are not supported in Greenplum Database:

1. **BINARY** and **VARBINARY** data types. `BYTEA` can be used in place of `VARBINARY` in Greenplum Database.

2. **FETCH FIRST** or **FETCH NEXT** clause for `SELECT`, for example:

   ```sql
   SELECT id, name FROM tab1 ORDER BY id OFFSET 20 ROWS FETCH NEXT 10 ROWS ONLY;
   ```
Greenplum has \texttt{LIMIT} and \texttt{LIMIT OFFSET} clauses instead.

3. The \texttt{ORDER BY} clause is ignored in views and subqueries unless a \texttt{LIMIT} clause is also used. This is intentional, as the Greenplum optimizer cannot determine when it is safe to avoid the sort, causing an unexpected performance impact for such \texttt{ORDER BY} clauses. To work around, you can specify a really large \texttt{LIMIT}. For example: \texttt{SELECT * FROM mytable ORDER BY 1 LIMIT 9999999999}

4. The \texttt{row subquery} construct is not supported.

5. \texttt{TRUNCATE TABLE} does not accept the \texttt{CONTINUE IDENTITY} and \texttt{RESTART IDENTITY} clauses.

\section*{Greenplum and PostgreSQL Compatibility}

Greenplum Database is based on PostgreSQL 8.2 with a few features added in from the 8.3 release. To support the distributed nature and typical workload of a Greenplum Database system, some SQL commands have been added or modified, and there are a few PostgreSQL features that are not supported. Greenplum has also added features not found in PostgreSQL, such as physical data distribution, parallel query optimization, external tables, resource queues for workload management and enhanced table partitioning. For full SQL syntax and references, see “SQL Command Reference” on page 243.

\begin{table}[h]
\centering
\caption{SQL Support in Greenplum Database}
\begin{tabular}{|l|l|l|}
\hline
\textbf{SQL Command} & \textbf{Supported in Greenplum} & \textbf{Modifications, Limitations, Exceptions} \\
\hline
\texttt{ALTER AGGREGATE} & \texttt{YES} & \texttt{Greenplum Database parallel tablespace feature - not in PostgreSQL 8.2.14.} \\
\texttt{ALTER CONVERSION} & \texttt{YES} & \texttt{An alias for ALTER ROLE} \\
\texttt{ALTER DATABASE} & \texttt{YES} & \texttt{Greenplum Database workload management feature - not in PostgreSQL.} \\
\texttt{ALTER DOMAIN} & \texttt{YES} & \texttt{Resource Queue: RESOURCE QUEUE queue_name | none} \\
\texttt{ALTER FILESPACE} & \texttt{YES} & \\
\texttt{ALTER FUNCTION} & \texttt{YES} & \\
\texttt{ALTER GROUP} & \texttt{YES} & \\
\texttt{ALTER INDEX} & \texttt{YES} & \\
\texttt{ALTER LANGUAGE} & \texttt{YES} & \\
\texttt{ALTER OPERATOR} & \texttt{YES} & \\
\texttt{ALTER OPERATOR CLASS} & \texttt{NO} & \\
\texttt{ALTER RESOURCE QUEUE} & \texttt{YES} & \\
\texttt{ALTER ROLE} & \texttt{YES} & \\
\texttt{ALTER SCHEMA} & \texttt{YES} & \\
\hline
\end{tabular}
\end{table}
<table>
<thead>
<tr>
<th>SQL Command</th>
<th>Supported in Greenplum</th>
<th>Modifications, Limitations, Exceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALTER SEQUENCE</td>
<td>YES</td>
<td>Unsupported Clauses / Options:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CLUSTER ON</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ENABLE/DISABLE TRIGGER</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Greenplum Database Clauses:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ADD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PARTITION</td>
</tr>
<tr>
<td>ALTER TABLE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>ALTER TABLESPACE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>ALTER TRIGGER</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>ALTER TYPE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>ALTER USER</td>
<td>YES</td>
<td>An alias for ALTER ROLE</td>
</tr>
<tr>
<td>ANALYZE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>BEGIN</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>CHECKPOINT</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>CLOSE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>CLUSTER</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>COMMENT</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>COMMIT</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>COMMIT PREPARED</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>COPY</td>
<td>YES</td>
<td>Modified Clauses:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ESCAPE [ AS ] 'escape'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Greenplum Database Clauses:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[LOG ERRORS INTO error_table]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SEGMENT REJECT LIMIT count [ROWS</td>
</tr>
<tr>
<td>CREATE AGGREGATE</td>
<td>YES</td>
<td>Unsupported Clauses / Options:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ , SORTOP = sort_operator ]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Greenplum Database Clauses:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ , PREFUNC = prefunc ]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Limitations:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The functions used to implement the aggregate must be IMMUTABLE functions.</td>
</tr>
<tr>
<td>CREATE CAST</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>CREATE CONSTRAINT TRIGGER</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>CREATE CONVERSION</td>
<td>YES</td>
<td></td>
</tr>
</tbody>
</table>
### Table 3.1 SQL Support in Greenplum Database

<table>
<thead>
<tr>
<th>SQL Command</th>
<th>Supported in Greenplum</th>
<th>Modifications, Limitations, Exceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CREATE DATABASE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>CREATE DOMAIN</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>CREATE EXTERNAL TABLE</td>
<td>YES</td>
<td>Greenplum Database parallel ETL feature - not in PostgreSQL 8.2.14.</td>
</tr>
<tr>
<td>CREATE FUNCTION</td>
<td>YES</td>
<td><strong>Limitations:</strong> Functions defined as <strong>STABLE</strong> or <strong>VOLATILE</strong> can be executed in Greenplum Database provided that they are executed on the master only. <strong>STABLE</strong> and <strong>VOLATILE</strong> functions cannot be used in statements that execute at the segment level.</td>
</tr>
<tr>
<td>CREATE GROUP</td>
<td>YES</td>
<td>An alias for <strong>CREATE ROLE</strong></td>
</tr>
<tr>
<td>CREATE INDEX</td>
<td>YES</td>
<td><strong>Greenplum Database Clauses:</strong> USING bitmap (bitmap indexes) <strong>Limitations:</strong> UNIQUE indexes are allowed only if they contain all of (or a superset of) the Greenplum distribution key columns. On partitioned tables, a unique index is only supported within an individual partition - not across all partitions. <strong>CONCURRENTLY</strong> keyword not supported in Greenplum.</td>
</tr>
<tr>
<td>CREATE LANGUAGE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>CREATE OPERATOR</td>
<td>YES</td>
<td><strong>Limitations:</strong> The function used to implement the operator must be an <strong>IMMUTABLE</strong> function.</td>
</tr>
<tr>
<td>CREATE OPERATOR CLASS</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>CREATE OPERATOR FAMILY</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>CREATE RESOURCE QUEUE</td>
<td>YES</td>
<td>Greenplum Database workload management feature - not in PostgreSQL 8.2.14.</td>
</tr>
<tr>
<td>CREATE ROLE</td>
<td>YES</td>
<td><strong>Greenplum Database Clauses:</strong> RESOURCE QUEUE queue_name</td>
</tr>
<tr>
<td>CREATE RULE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>CREATE SCHEMA</td>
<td>YES</td>
<td></td>
</tr>
</tbody>
</table>
| CREATE SEQUENCE          | YES                    | **Limitations:**  
- The lastval and currval functions are not supported.  
- The setval function is only allowed in queries that do not operate on distributed data. |
### Table 3.1  SQL Support in Greenplum Database

<table>
<thead>
<tr>
<th>SQL Command</th>
<th>Supported in Greenplum</th>
<th>Modifications, Limitations, Exceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CREATE TABLE</td>
<td>YES</td>
<td>Unsupported Clauses / Options:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[GLOBAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>REFERENCES</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FOREIGN KEY</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[DEFERRABLE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Limited Clauses:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• UNIQUE or PRIMARY KEY constraints are only allowed on hash-distributed tables (DISTRIBUTED BY), and the constraint columns must be the same as or a superset of the table’s distribution key columns.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Greenplum Database Clauses:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DISTRIBUTED BY (column, [ ... ] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DISTRIBUTED RANDOMLY</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PARTITION BY type (column [, ...])</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( partition_specification, [ ... ] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WITH (appendonly=true</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ ,compresslevel=value,blocksize=value]</td>
</tr>
<tr>
<td>CREATE TABLE AS</td>
<td>YES</td>
<td>See CREATE TABLE</td>
</tr>
<tr>
<td>CREATE TABLESPACE</td>
<td>NO</td>
<td>Greenplum Database Clauses:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FILESPACE filespace_name</td>
</tr>
<tr>
<td>CREATE TRIGGER</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>CREATE TYPE</td>
<td>YES</td>
<td>Limitations:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The functions used to implement a new base type must be IMMUTABLE functions.</td>
</tr>
<tr>
<td>CREATE USER</td>
<td>YES</td>
<td>An alias for CREATE ROLE</td>
</tr>
<tr>
<td>CREATE VIEW</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DEALLOCATE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DECLARE</td>
<td>YES</td>
<td>Unsupported Clauses / Options:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCROLL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FOR UPDATE [ OF column [, ...] ]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Limitations:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cursors are non-updatable, and cannot be backward-scrolled. Forward scrolling is supported.</td>
</tr>
<tr>
<td>DELETE</td>
<td>YES</td>
<td>Unsupported Clauses / Options:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RETURNING</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Limitations:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Joins must be on a common Greenplum distribution key (equijoins)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Cannot use STABLE or VOLATILE functions in a DELETE statement if mirrors are enabled</td>
</tr>
<tr>
<td>SQL Command</td>
<td>Supported in Greenplum</td>
<td>Modifications, Limitations, Exceptions</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>DROP AGGREGATE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP CAST</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP CONVERSION</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP DATABASE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP DOMAIN</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP EXTERNAL TABLE</td>
<td>YES</td>
<td>Greenplum Database parallel ETL feature - not in PostgreSQL 8.2.14.</td>
</tr>
<tr>
<td>DROP FUNCTION</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP GROUP</td>
<td>YES</td>
<td>An alias for DROP ROLE</td>
</tr>
<tr>
<td>DROP INDEX</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP LANGUAGE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP OPERATOR</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP OPERATOR CLASS</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>DROP OWNED</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>DROP RESOURCE QUEUE</td>
<td>YES</td>
<td>Greenplum Database workload management feature - not in PostgreSQL 8.2.14.</td>
</tr>
<tr>
<td>DROP ROLE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP RULE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP SCHEMA</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP SEQUENCE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP TABLE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP TABLESPACE</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>DROP TRIGGER</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>DROP TYPE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>DROP USER</td>
<td>YES</td>
<td>An alias for DROP ROLE</td>
</tr>
<tr>
<td>DROP VIEW</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>END</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>EXECUTE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>EXPLAIN</td>
<td>YES</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.1  SQL Support in Greenplum Database

<table>
<thead>
<tr>
<th>SQL Command</th>
<th>Supported in Greenplum</th>
<th>Modifications, Limitations, Exceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>FETCH</td>
<td>YES</td>
<td>Unsupported Clauses / Options:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LAST</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PRIOR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BACKWARD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BACKWARD ALL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Limitations:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cannot fetch rows in a nonsequential</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fashion; backward scan is not</td>
</tr>
<tr>
<td></td>
<td></td>
<td>supported.</td>
</tr>
<tr>
<td>GRANT</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>INSERT</td>
<td>YES</td>
<td>Unsupported Clauses / Options:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RETURNING</td>
</tr>
<tr>
<td>LISTEN</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>LOAD</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>LOCK</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>MOVE</td>
<td>YES</td>
<td>See FETCH</td>
</tr>
<tr>
<td>NOTIFY</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>PREPARE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>PREPARE TRANSACTION</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>REASSIGN OWNED</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>REINDEX</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>RELEASE SAVEPOINT</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>RESET</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>REVOKE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>ROLLBACK</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>ROLLBACK PREPARED</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>ROLLBACK TO SAVEPOINT</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>SAVEPOINT</td>
<td>YES</td>
<td></td>
</tr>
</tbody>
</table>
### Table 3.1  SQL Support in Greenplum Database

<table>
<thead>
<tr>
<th>SQL Command</th>
<th>Supported in Greenplum</th>
<th>Modifications, Limitations, Exceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELECT</td>
<td>YES</td>
<td>Limitations:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Limited use of <code>VOLATILE</code> and <code>STABLE</code> functions in <code>FROM</code> or <code>WHERE</code> clauses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Limited use of correlated subquery expressions (See <code>SELECT</code>)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Text search (Tsearch2) is not supported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• FETCH FIRST or FETCH NEXT clauses not supported</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Greenplum Database Clauses (OLAP):</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[GROUP BY grouping_element [, ...]]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[WINDOW window_name AS (window_specification)]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[FILTER (WHERE condition)] applied to an aggregate function in the <code>SELECT</code> list</td>
</tr>
<tr>
<td>SELECT INTO</td>
<td>YES</td>
<td>See <code>SELECT</code></td>
</tr>
<tr>
<td>SET</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>SET CONSTRAINTS</td>
<td>NO</td>
<td>In PostgreSQL, this only applies to foreign key constraints, which are currently not enforced in Greenplum Database.</td>
</tr>
<tr>
<td>SET ROLE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>SET SESSION AUTHORIZATION</td>
<td>YES</td>
<td>Deprecated as of PostgreSQL 8.1 - see <code>SET ROLE</code></td>
</tr>
<tr>
<td>SET TRANSACTION</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>SHOW</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>START TRANSACTION</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>TRUNCATE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>UNLISTEN</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>UPDATE</td>
<td>YES</td>
<td><strong>Unsupported Clauses:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RETURNING</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Limitations:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <code>SET</code> not allowed for Greenplum distribution key columns.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Joins must be on a common Greenplum distribution key (equijoins).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Cannot use <code>STABLE</code> or <code>VOLATILE</code> functions in an <code>UPDATE</code> statement if mirrors are enabled.</td>
</tr>
<tr>
<td>VACUUM</td>
<td>YES</td>
<td><strong>Limitations:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><code>VACUUM FULL</code> is not recommended in Greenplum Database.</td>
</tr>
<tr>
<td>VALUES</td>
<td>YES</td>
<td></td>
</tr>
</tbody>
</table>
4. About Greenplum Query Processing

Users issue queries to Greenplum Database just as they would to any database management system (DBMS). They connect to the database instance on the Greenplum master host using a client application (such as `psql`) and submit an SQL statement.

Understanding Query Planning and Dispatch

The query is received through the master, which parses the query, optimizes the query, and creates either a parallel or targeted query plan (depending on the query). The master then dispatches the plan to the segments for execution. Each segment is then responsible for executing local database operations on its own particular set of data.

Most database operations—such as table scans, joins, aggregations, and sorts—execute in parallel across the segments simultaneously. Each operation is performed on a segment database independent of the data associated with the other segment databases.

![Figure 4.1 Dispatching the Parallel Query Plan](image)

Figure 4.1 Dispatching the Parallel Query Plan
Certain queries may only access data on a single segment, such as single-row INSERT, UPDATE, DELETE or SELECT operations, or queries that return a small set of rows and filter on the table distribution key column(s). In queries such as these, the query plan is not dispatched to all segments, but is targeted to the segment that contains the affected row(s).

**Figure 4.2  Dispatching a Targeted Query Plan**

**Understanding Greenplum Query Plans**

A query plan is the set of operations that Greenplum Database will perform to produce the answer to a given query. Each node or step in the plan represents a database operation such as a table scan, join, aggregation or sort. Plans are read and executed from bottom to top.

In addition to the typical database operations (tables scans, joins, etc.), Greenplum Database has an additional operation type called a *motion*. A motion operation involves moving tuples between the segments during query processing. Note that not every query requires a motion. For example, a query of the system catalog tables on the master does not require data to move across the interconnect.

In order to achieve maximum parallelism during query execution, Greenplum divides the work of the query plan into *slices*. A slice is a portion of the plan that can be worked on independently at the segment-level. A query plan is sliced wherever a *motion* operation occurs in the plan, one slice on each side of the motion.

For example, consider the following simple query involving a join between two tables:

```sql
SELECT customer, amount
```
FROM sales JOIN customer USING (cust_id)  
WHERE dateCol = '04-30-2008';

Figure 4.3 shows the query plan. Note that each segment gets a copy of the query plan and works on it in parallel. For this particular plan, there is a redistribute motion that moves tuples between the segments in order to complete the join. The plan is sliced on either side of the redistribute motion (slice 1 and slice 2). This query plan also has another type of motion operation called a gather motion. A gather motion is when the segments send results back up to the master for presentation to the client. Since a query plan is always sliced wherever a motion occurs, this plan also has an implicit slice at the very top of the plan (slice 3). Not all query plans involve a gather motion. For example, a CREATE TABLE x AS SELECT... statement would not have a gather motion (tuples are sent to the newly created table, not to the master).

Figure 4.3 Query Slice Plan

**Understanding Parallel Query Execution**

Greenplum creates a number of database processes to handle the work of a query. On the master, the query worker process is called the query dispatcher (QD). The QD is responsible for creating and dispatching the query plan, and for accumulating and presenting the final results. On the segments, a query worker process is called a query executor (QE). A QE is responsible for completing its portion of work and communicating its intermediate results to the other worker processes.
For each slice of the query plan there is at least one worker process assigned. A worker process works on its assigned portion of the query plan independently. During query execution, each segment will have a number of processes working on the query in parallel.

Related processes that are working on the same portion of the query plan are referred to as gangs. As a portion of work is completed, tuples flow up the query plan from one gang of processes to the next. This inter-process communication between the segments is what is referred to as the interconnect component of Greenplum Database.

Figure 4.4 shows the query worker processes on the master and two segment instances for the query plan illustrated in Figure 4.3.

Figure 4.4  Query Worker Processes
Section II: Access Control and Security

This section describes how to manage access to your Greenplum Database system. It contains the following chapters:

- **Managing Roles and Privileges** - Provides information on creating database roles (users and groups) and managing access privileges to database objects.
- **Configuring Client Authentication** - Provides information about the `pg_hba.conf` file, a configuration file used to control client access and authentication to Greenplum Database.
- **Accessing the Database** - Explains the various client tools you can use to connect to Greenplum Database, and how to establish a database session.
- **Managing Workload and Resources** - Describes the workload management feature of Greenplum Database, and explains the tasks involved in creating and managing resource queues.
5. Managing Roles and Privileges

Greenplum Database manages database access permissions using the concept of roles. The concept of roles subsumes the concepts of users and groups. A role can be a database user, a group, or both. Roles can own database objects (for example, tables) and can assign privileges on those objects to other roles to control access to the objects. Roles can be members of other roles, thus a member role can inherit the object privileges of its parent role.

Every Greenplum Database system contains a set of database roles (users and groups). Those roles are separate from the users and groups managed by the operating system on which the server runs. However, for convenience you may want to maintain a relationship between operating system user names and Greenplum Database role names, since many of the client applications use the current operating system user name as the default.

In Greenplum Database, users log in and connect through the master instance, which then verifies their role and access privileges. The master then issues out commands to the segment instances behind the scenes as the currently logged in role.

Roles are defined at the system level, meaning they are valid for all databases in the system.

In order to bootstrap the Greenplum Database system, a freshly initialized system always contains one predefined superuser role. This role will have the same name as the operating system user that initialized the Greenplum Database system. Customarily, this role is named gpadmin. In order to create more roles you first have to connect as this initial role.

Security Best Practices for Roles and Privileges

- Secure the gpadmin system user. Greenplum requires a UNIX user id to install and initialize the Greenplum Database system. This system user is referred to as gpadmin in the Greenplum documentation. This gpadmin user is the default database superuser in Greenplum Database, as well as the file system owner of the Greenplum installation and its underlying data files. This default administrator account is fundamental to the design of Greenplum Database. The system cannot run without it, and there is no way to limit the access of this gpadmin user id. This gpadmin user can bypass all security features of Greenplum Database. Anyone who logs on to a Greenplum host as this user id can read, alter or delete any data; including system catalog data and database access rights. Therefore, it is very important to secure the gpadmin user id and only provide access to essential system administrators. Administrators should only log in to Greenplum as gpadmin when performing certain system maintenance tasks (such as upgrade or expansion). Database users should never log on as gpadmin, and ETL or production workloads should never run as gpadmin.
• **Assign a distinct role to each user that logs in.** For logging and auditing purposes, each user that is allowed to log in to Greenplum Database should be given their own database role. For applications or web services, consider creating a distinct role for each application or service. See “Creating New Roles (Users)” on page 31.

• **Use groups to manage access privileges.** See “Creating Groups (Role Membership)” on page 32.

• **Limit users who have the SUPERUSER role attribute.** Roles that are superusers bypass all access privilege checks in Greenplum Database, as well as resource queuing. Only system administrators should be given superuser rights. See “Altering Role Attributes” on page 31.

### Creating New Roles (Users)

A user-level role is considered to be a database role that can log in to the database and initiate a database session. Therefore, when you create a new user-level role using the `CREATE ROLE` command, you must specify the `LOGIN` privilege. For example:

```sql
=# CREATE ROLE jsmith WITH LOGIN;
```

A database role may have a number of attributes that define what sort of tasks that role can perform in the database. You can set these attributes when you create the role, or later using the `ALTER ROLE` command. See Table 5.1, “Role Attributes” on page 31 for a description of the role attributes you can set.

### Altering Role Attributes

A database role may have a number of attributes that define what sort of tasks that role can perform in the database.

#### Table 5.1 Role Attributes

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUPERUSER</td>
<td>NOSUPERUSER</td>
</tr>
<tr>
<td>CREATEDB</td>
<td>NOCREATEDB</td>
</tr>
<tr>
<td>CREATEROLE</td>
<td>NOCREATEROLE</td>
</tr>
<tr>
<td>INHERIT</td>
<td>NOINHERIT</td>
</tr>
<tr>
<td>LOGIN</td>
<td>NOLOGIN</td>
</tr>
</tbody>
</table>
You can set these attributes when you create the role, or later using the `ALTER ROLE` command. For example:

```
=# ALTER ROLE jsmith WITH PASSWORD 'passwd123';
=# ALTER ROLE admin VALID UNTIL 'infinity';
=# ALTER ROLE jsmith LOGIN;
=# ALTER ROLE jsmith RESOURCE QUEUE adhoc;
```

A role can also have role-specific defaults for many of the server configuration settings. For example, to set the default schema search path for a role:

```
=# ALTER ROLE admin SET search_path TO myschema, public;
```

### Creating Groups (Role Membership)

It is frequently convenient to group users together to ease management of object privileges: that way, privileges can be granted to, or revoked from, a group as a whole. In Greenplum Database this is done by creating a role that represents the group, and then granting membership in the group role to individual user roles.

Use the `CREATE ROLE` SQL command to create a new group role. For example:

```
=# CREATE ROLE admin CREATEROLE CREATEDB;
```

Once the group role exists, you can add and remove members (user roles) using the `GRANT` and `REVOKE` commands. For example:

```
=# GRANT admin TO john, sally;
=# REVOKE admin FROM bob;
```
For managing object privileges, you would then grant the appropriate permissions to the group-level role only (see Table 5.2, “Object Privileges” on page 33). The member user roles then inherit the object privileges of the group role. For example:

```
=# GRANT ALL ON TABLE mytable TO admin;
=# GRANT ALL ON SCHEMA myschema TO admin;
=# GRANT ALL ON DATABASE mydb TO admin;
```

The role attributes LOGIN, SUPERUSER, CREATEDB, and CREATEROLE are never inherited as ordinary privileges on database objects are. User members must actually SET ROLE to a specific role having one of these attributes in order to make use of the attribute. In the above example, we gave CREATEDB and CREATEROLE to the admin role. If sally is a member of admin, she could issue the following command to assume the role attributes of the parent role:

```
=> SET ROLE admin;
```

### Managing Object Privileges

When an object (table, view, sequence, database, function, language, schema, or tablespace) is created, it is assigned an owner. The owner is normally the role that executed the creation statement. For most kinds of objects, the initial state is that only the owner (or a superuser) can do anything with the object. To allow other roles to use it, privileges must be granted. Greenplum Database supports the following privileges for each object type:

**Table 5.2 Object Privileges**

<table>
<thead>
<tr>
<th>Object Type</th>
<th>Privileges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tables, Views, Sequences</td>
<td>SELECT, INSERT, UPDATE, DELETE, RULE, ALL</td>
</tr>
<tr>
<td>External Tables</td>
<td>SELECT, RULE, ALL</td>
</tr>
<tr>
<td>Databases</td>
<td>CONNECT, CREATE, TEMPORARY</td>
</tr>
<tr>
<td>Functions</td>
<td>EXECUTE</td>
</tr>
<tr>
<td>Procedural Languages</td>
<td>USAGE</td>
</tr>
<tr>
<td>Schemas</td>
<td>CREATE, USAGE, ALL</td>
</tr>
</tbody>
</table>

Managing Object Privileges

When an object (table, view, sequence, database, function, language, schema, or tablespace) is created, it is assigned an owner. The owner is normally the role that executed the creation statement. For most kinds of objects, the initial state is that only the owner (or a superuser) can do anything with the object. To allow other roles to use it, privileges must be granted. Greenplum Database supports the following privileges for each object type:

**Table 5.2 Object Privileges**

<table>
<thead>
<tr>
<th>Object Type</th>
<th>Privileges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tables, Views, Sequences</td>
<td>SELECT, INSERT, UPDATE, DELETE, RULE, ALL</td>
</tr>
<tr>
<td>External Tables</td>
<td>SELECT, RULE, ALL</td>
</tr>
<tr>
<td>Databases</td>
<td>CONNECT, CREATE, TEMPORARY</td>
</tr>
<tr>
<td>Functions</td>
<td>EXECUTE</td>
</tr>
<tr>
<td>Procedural Languages</td>
<td>USAGE</td>
</tr>
<tr>
<td>Schemas</td>
<td>CREATE, USAGE, ALL</td>
</tr>
</tbody>
</table>
Privileges must be granted for each object individually. For example, granting `ALL` on a database does not grant full access to the objects within that database. It only grants all of the database-level privileges (`CONNECT`, `CREATE`, `TEMPORARY`) to the database itself.

Use the `GRANT` SQL command to give a specified role privileges on an object. For example:

```sql
=# GRANT INSERT ON mytable TO jsmith;
```

To revoke privileges, use the `REVOKE` command. For example:

```sql
=# REVOKE ALL PRIVILEGES ON mytable FROM jsmith;
```

You can also use the `DROP OWNED` and `REASSIGN OWNED` commands for managing objects owned by deprecated roles (Note: only an object’s owner or a superuser can drop an object or reassign ownership). For example:

```sql
=# REASSIGN OWNED BY sally TO bob;
=# DROP OWNED BY visitor;
```

Simulating Row and Column Level Access Control

Greenplum Database access control corresponds roughly to the Orange Book ‘C2’ level of security, not the ‘B1’ level. Greenplum Database currently supports access privileges at the object level. Row-level or column-level access is not supported, nor is labeled security.

Row-level and column-level access can be simulated using views to restrict the columns and/or rows that are selected. Row-level labels can be simulated by adding an extra column to the table to store sensitivity information, and then using views to control row-level access based on this column. Roles can then be granted access to the views rather than the base table. While these workarounds do not provide the same as "B1" level security, they may still be a viable alternative for many organizations that require more granular access control.

Encrypting Data

PostgreSQL provides an optional package of encryption/decryption functions called `pgcrypto`, which can also be installed and used in Greenplum Database. The `pgcrypto` package is not installed by default with Greenplum Database, however Greenplum can provide a platform-specific build of `pgcrypto` upon request. Contact Greenplum Customer Support to obtain a build of `pgcrypto` and its supporting documentation.

The `pgcrypto` functions allow database administrators to store certain columns of data in encrypted form. This adds an extra layer of protection for sensitive data, as data stored in Greenplum Database in encrypted form cannot be read by users who do not have the encryption key, nor be read directly from the disks.
It is important to note that the `pgcrypto` functions run inside database server. That means that all the data and passwords move between `pgcrypto` and the client application in clear-text. For optimal security, consider also using SSL connections between the client and the Greenplum master server.
6. Configuring Client Authentication

When a Greenplum Database system is first initialized, the system contains one predefined superuser role. This role will have the same name as the operating system user who initialized the Greenplum Database system. This role is referred to as gpadmin. By default, the system is configured to only allow local connections to the database from the gpadmin role. If you want to allow any other roles to connect, or if you want to allow connections from remote hosts, you have to configure Greenplum Database to allow such connections. This chapter explains how to configure client connections and authentication to Greenplum Database.

- Allowing Connections to Greenplum Database
- Limiting Concurrent Connections

Allowing Connections to Greenplum Database

Client access and authentication is controlled by a configuration file named pg_hba.conf (the standard PostgreSQL host-based authentication file). For detailed information about this file, see The pg_hba.conf File in the PostgreSQL documentation.

In Greenplum Database, the pg_hba.conf file of the master instance controls client access and authentication to your Greenplum system. The segments also have pg_hba.conf files, but these are already correctly configured to only allow client connections from the master host. The segments never accept outside client connections, so there is no need to alter the pg_hba.conf file on your segments.

The general format of the pg_hba.conf file is a set of records, one per line. Blank lines are ignored, as is any text after the # comment character. A record is made up of a number of fields which are separated by spaces and/or tabs. Fields can contain white space if the field value is quoted. Records cannot be continued across lines. Each remote client access record is in the format of:

```
host  database  role  CIDR-address  authentication-method
```

Each UNIX-domain socket access record is in the format of:

```
local  database  role  authentication-method
```

The meaning of the fields is as follows:

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>local</td>
<td>Matches connection attempts using UNIX-domain sockets. Without a record of this type, UNIX-domain socket connections are disallowed.</td>
</tr>
<tr>
<td>host</td>
<td>Matches connection attempts made using TCP/IP. Remote TCP/IP connections will not be possible unless the server is started with an appropriate value for the listen_addresses server configuration parameter.</td>
</tr>
</tbody>
</table>
Editing the pg_hba.conf File

This example shows how to edit the pg_hba.conf file of the master to allow remote client access to all databases from all roles using md5-encrypted password authentication.

**Note:** For a more secure system, consider removing all connections that use `trust` authentication from your master pg_hba.conf. Trust authentication means the role is granted access without any authentication, therefore bypassing all security. Replace trust entries with `ident` authentication if your system has an ident service available.

**Editing pg_hba.conf**

1. Open the file `$MASTER_DATA_DIRECTORY/pg_hba.conf` in a text editor.

2. Add a line to the file for each type of connection you want to allow. Records are read sequentially, so the order of the records is significant. Typically, earlier records will have tight connection match parameters and weaker authentication methods, while later records will have looser match parameters and stronger authentication methods. For example:

   ```
   # allow the gpadmin user local access to all databases
   ```
# using ident authentication
local   all   gpadmin   ident         sameuser
host   all   gpadmin   127.0.0.1/32  ident
host   all   gpadmin   ::1/128       ident

# allow the 'dba' role access to any database from any
# host with IP address 192.168.x.x and use md5 encrypted
# passwords to authenticate the user
host   all   dba   192.168.0.0/32  md5

# allow all roles access to any database from any
# host and use ldap to authenticate the user. Greenplum role
# names must match the LDAP common name.
host   all   all   192.168.0.0/32  ldap
ldapserver=usldap1
ldapport=1389
ldapprefix="cn="
ldapsuffix=",ou=People,dc=company,dc=com"

3. Save and close the file.
4. Reload the pg_hba.conf configuration file for your changes to take effect:
   $ gpstop -u

**Note:** Note that you can also control database access by setting object privileges as described in “Managing Object Privileges” on page 33. The pg_hba.conf file just controls who can initiate a database session and how those connections are authenticated.

---

**Limiting Concurrent Connections**

To limit the number of active concurrent sessions to your Greenplum Database system, you can configure the `max_connections` server configuration parameter. This is a local parameter, meaning that you must set it in the `postgresql.conf` file of the master, the standby master, and each segment instance (primary and mirror). The value of `max_connections` on segments must be 5-10 times the value on the master.

When you set `max_connections`, you must also set the dependent parameter `max_prepared_transactions`. This value must be at least as large as the value of `max_connections` on the master, and segment instances should be set to the same value as the master.

For example:

In `$MASTER_DATA_DIRECTORY/postgresql.conf` (including standby master):

```
max_connections=100
max_prepared_transactions=100
```

In `$SEGMENT_DATA_DIRECTORY/postgresql.conf` for all segment instances:

```
max_connections=500
max_prepared_transactions=100
```
To change the number of allowed connections

1. Stop your Greenplum Database system:
   
   $ gpstop

2. On your master host, edit $MASTER_DATA_DIRECTORY/postgresql.conf and change the following two parameters:
   
   max_connections (the number of active user sessions you want to allow plus the number of superuser_reserved_connections)
   max_prepared_transactions (must be greater than or equal to max_connections)

3. On each segment instance, edit SEGMENT_DATA_DIRECTORY/postgresql.conf and change the following two parameters:
   
   max_connections (must be 5-10 times the value on the master)
   max_prepared_transactions (must be equal to the value on the master)

4. Restart your Greenplum Database system:
   
   $ gpstart

Note: Raising the values of these parameters may cause Greenplum Database to request more shared memory. To mitigate this effect, consider decreasing other memory-related parameters such as gp_cached_segworkers_threshold.

Encrypting Client/Server Connections

Greenplum Database has native support for SSL connections between the client and the master server. SSL connections prevent third parties from snooping on the packets, and also prevent man-in-the-middle attacks. SSL should be used whenever the client connection goes through an insecure link, and must be used whenever client certificate authentication is used.

To enable SSL requires that OpenSSL be installed on both the client and the master server systems. Greenplum can be started with SSL enabled by setting the server configuration parameter ssl=on in the master postgresql.conf. When starting in SSL mode, the server will look for the files server.key (server private key) and server.crt (server certificate) in the master data directory. These files must be set up correctly before an SSL-enabled Greenplum system can start. If the private key is protected with a passphrase, the server will prompt for the passphrase and will not start until it has been entered.

For details on how to create your server private key and certificate, refer to the OpenSSL documentation. A self-signed certificate can be used for testing, but a certificate signed by a certificate authority (CA) (either one of the global CAs or a local one) should be used in production so the client can verify the server’s identity.
7. Accessing the Database

This chapter explains the various client tools you can use to connect to Greenplum Database, and how to establish a database session. It contains the following topics:

- Establishing a Database Session
- Supported Client Applications
- Troubleshooting Connection Problems

Establishing a Database Session

Users can connect to Greenplum Database using a PostgreSQL-compatible client program, such as `psql`. Users and administrators always connect to Greenplum Database through the master - the segments cannot accept client connections.

In order to establish a connection to the Greenplum Database master, you will need to know the following connection information and configure your client program accordingly.

<table>
<thead>
<tr>
<th>Connection Parameter</th>
<th>Description</th>
<th>Environment Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database name</td>
<td>The name of the database to which you want to connect. For a newly initialized system, use the <code>template1</code> database to connect for the first time.</td>
<td>$PGDATABASE</td>
</tr>
<tr>
<td>Host name</td>
<td>The host name of the Greenplum Database master. The default host is the local host.</td>
<td>$PGHOST</td>
</tr>
<tr>
<td>Port</td>
<td>The port number that the Greenplum Database master instance is running on. The default is 5432.</td>
<td>$PGPORT</td>
</tr>
<tr>
<td>User name</td>
<td>The database user (role) name to connect as. This is not necessarily the same as your OS user name. Check with your Greenplum administrator if you are not sure what you database user name is. Note that every Greenplum Database system has one superuser account that is created automatically at initialization time. This account has the same name as the OS name of the user who initialized the Greenplum system (typically <code>gpadmin</code>).</td>
<td>$PGUSER</td>
</tr>
</tbody>
</table>
Supported Client Applications

Users can connect to Greenplum Database using various client applications:

- A number of Greenplum Database Client Applications are provided with your Greenplum installation. The psql client application provides an interactive command-line interface to Greenplum Database.

- pgAdmin III for Greenplum Database is an enhanced version of the popular management tool pgAdmin III. Since version 1.10.0, the pgAdmin III client available from PostgreSQL Tools includes support for Greenplum-specific features. Installation packages are available for download from Greenplum Network and from the pgAdmin download site.

- Using standard Database Application Interfaces, such as ODBC and JDBC, users can create their own client applications that interface to Greenplum Database. Because Greenplum Database is based on PostgreSQL, it uses the standard PostgreSQL database drivers.

- Most Third-Party Client Tools that use standard database interfaces, such as ODBC and JDBC, can be configured to connect to Greenplum Database.

Greenplum Database Client Applications

Greenplum Database comes installed with a number of client applications located in $GPHOME/bin of your Greenplum Database master host installation. The following are the most commonly used client applications:

Table 7.2  Commonly used client applications

<table>
<thead>
<tr>
<th>Name</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>createdb</td>
<td>create a new database</td>
</tr>
<tr>
<td>createlang</td>
<td>define a new procedural language</td>
</tr>
<tr>
<td>createuser</td>
<td>define a new database role</td>
</tr>
<tr>
<td>dropdb</td>
<td>remove a database</td>
</tr>
<tr>
<td>droplang</td>
<td>remove a procedural language</td>
</tr>
<tr>
<td>dropuser</td>
<td>remove a role</td>
</tr>
<tr>
<td>psql</td>
<td>PostgreSQL interactive terminal</td>
</tr>
<tr>
<td>reindexdb</td>
<td>reindex a database</td>
</tr>
<tr>
<td>vacuumdb</td>
<td>garbage-collect and analyze a database</td>
</tr>
</tbody>
</table>

When using these client applications, you must connect to a database through the Greenplum master instance. You will need to know the name of your target database, the host name and port number of the master, and what database user name to connect as. This information can be provided on the command-line using the options -d, -h, -p, and -U respectively. If an argument is found that does not belong to any option, it will be interpreted as the database name first.
All of these options have default values which will be used if the option is not specified. The default host is the local host. The default port number is 5432. The default user name is your OS system user name, as is the default database name. Note that OS user names and Greenplum Database user names are not necessarily the same.

If the default values are not correct, you can save yourself some typing by setting the environment variables `PGDATABASE`, `PGHOST`, `PGPORT`, and `PGUSER` to the appropriate values. See “Greenplum Environment Variables” on page 819 for more information. It is also convenient to have a `~/.pgpass` file to avoid regularly having to type in passwords. See “psql” on page 744 for more information.

**Connecting with psql**

Depending on the default values used or the environment variables you have set, the following examples show how to access a database via `psql`:

```
$ psql -d gpdatabase -h master_host -p 5432 -U gpadmin
$ psql gpdatabase
$ psql
```

If a user-defined database has not yet been created, you can access the system by connecting to the `template1` database. For example:

```
$ psql template1
```

After connecting to a database, `psql` provides a prompt with the name of the database to which `psql` is currently connected, followed by the string `=>` (or `=#` if you are the database superuser). For example:

```
gpdatabase=>
```

At the prompt, you may type in SQL commands. A SQL command must end with a `;` (semicolon) in order to be sent to the server and executed. For example:

```
=> SELECT * FROM mytable;
```

For more information on using the `psql` client application, see “psql” on page 744. For more information on SQL commands and syntax, see “SQL Command Reference” on page 243.

**pgAdmin III for Greenplum Database**

If you prefer a graphic interface, use pgAdmin III for Greenplum Database. This GUI client supports PostgreSQL databases with all standard pgAdmin III features, while adding support for Greenplum-specific features.

pgAdmin III for Greenplum Database supports the following Greenplum-specific features:

- External tables
- Append-only tables, including compressed append-only tables
- Table partitioning
- Resource queues
- Graphical EXPLAIN ANALYZE
- Greenplum server configuration parameters

**Figure 7.1** Greenplum Options in pgAdmin III

**Installing pgAdmin III for Greenplum Database**

The installation package for pgAdmin III for Greenplum Database is available for download from the official pgAdmin III download site (http://www.pgadmin.org). Installation instructions are included in the installation package.

**Documentation for pgAdmin III for Greenplum Database**

For general help on the features of the graphical interface, select Help contents from the Help menu.

For help with Greenplum-specific SQL support, select Greenplum Database Help from the Help menu. If you have an active internet connection, you will be directed to online Greenplum SQL reference documentation. Alternately, you can install the Greenplum Client Tools package. This package contains SQL reference documentation that is accessible to the help links in pgAdmin III.

**Performing Administrative Tasks with pgAdmin III**

This section highlights two of the many Greenplum Database administrative tasks you can perform with pgAdmin III: editing the server configuration, and viewing a graphical representation of a query plan.
Editing Server Configuration

The pgAdmin III interface provides two ways to update the server configuration in `postgresql.conf`: locally, through the File menu, and remotely on the server through the Tools menu. Editing the server configuration remotely may be more convenient in many cases, because it does not require you to upload or copy `postgresql.conf`.

To edit server configuration remotely

1. Connect to the server whose configuration you want to edit. If you are connected to multiple servers, make sure that the correct server is highlighted in the object browser in the left pane.

2. Select Tools > Server Configuration > `postgresql.conf`. The Backend Configuration Editor opens, displaying the list of available and enabled server configuration parameters.

3. Locate the parameter you want to edit, and double click on the entry to open the Configuration settings dialog.

4. Enter the new value for the parameter, or select/deselect Enabled as desired and click OK.

5. If the parameter can be enabled by reloading server configuration, click the green reload icon, or select File > Reload server. Many parameters require a full restart of the server.

Viewing a Graphical Query Plan

Using the pgAdmin III query tool, you can run a query with EXPLAIN to view the details of the query plan. The output includes details about operations unique to Greenplum distributed query processing such as plan slices and motions between segments. You can view a graphical depiction of the plan as well as the text-based data output.

To view a graphical query plan

1. With the correct database highlighted in the object browser in the left pane, select Tools > Query tool.

2. Enter the query by typing in the SQL Editor, dragging objects into the Graphical Query Builder, or opening a file.

3. Select Query > Explain options and verify the following options:
   • Verbose — this must be deselected if you want to view a graphical depiction of the query plan
   • Analyze — select this option if you want to run the query in addition to viewing the plan

4. Trigger the operation by clicking the Explain query option at the top of the pane, or by selecting Query > Explain.

   The query plan displays in the Output pane at the bottom of the screen. Select the Explain tab to view the graphical output. For example:
Database Application Interfaces

You may want to develop your own client applications that interface to Greenplum Database. PostgreSQL provides a number of database drivers for the most commonly used database application programming interfaces (APIs), which can also be used with Greenplum Database. These drivers are not packaged with the Greenplum Database base distribution. Each driver is an independent PostgreSQL development project and must be downloaded, installed and configured to connect to Greenplum Database. The following drivers are available:

Table 7.3  Greenplum Database Interfaces

<table>
<thead>
<tr>
<th>API</th>
<th>PostgreSQL Driver</th>
<th>Download Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODBC</td>
<td>pgodbc</td>
<td>Available in the Greenplum Database Connectivity package, which can be downloaded from Greenplum Network.</td>
</tr>
<tr>
<td>JDBC</td>
<td>pgjdbc</td>
<td>Available in the Greenplum Database Connectivity package, which can be downloaded from Greenplum Network.</td>
</tr>
<tr>
<td>Perl DBI</td>
<td>pgperl</td>
<td><a href="http://gborg.postgresql.org/project/pgperl">http://gborg.postgresql.org/project/pgperl</a></td>
</tr>
<tr>
<td>Python DBI</td>
<td>pygresql</td>
<td><a href="http://www.pygresql.org">http://www.pygresql.org</a></td>
</tr>
</tbody>
</table>

General instructions for accessing a Greenplum Database with an API are:
1. Download your programming language platform and respective API from the appropriate source. For example, you can get the Java development kit (JDK) and JDBC API from Sun.

2. Write your client application according to the API specifications. When programming your application, be aware of the SQL support in Greenplum Database so you do not include any unsupported SQL syntax. See “SQL Command Reference” on page 243.

3. Download the appropriate PostgreSQL driver and configure connectivity to your Greenplum Database master instance. Greenplum provides a client tools package that contains the supported database drivers for Greenplum Database. The client tools package (and associated documentation) is available for download from Greenplum Network.

**Third-Party Client Tools**

Most third-party extract-transform-load (ETL) and business intelligence (BI) tools use standard database interfaces, such as ODBC and JDBC, and can be configured to connect to Greenplum Database. Greenplum has worked with the following tools on previous customer engagements and is in the process of becoming officially certified:

- Business Objects
- Microstrategy
- Informatica Power Center
- Microsoft SQL Server Integration Services (SSIS) and Reporting Services (SSRS)
- Ascential Datastage
- SAS
- Cognos

Greenplum Professional Services can assist users in configuring their chosen third-party tool for use with Greenplum Database.
# Troubleshooting Connection Problems

A number of things can prevent a client application from successfully connecting to Greenplum Database. This section explains some of the common causes of connection problems and how to correct them.

## Table 7.4 Common connection problems

<table>
<thead>
<tr>
<th>Problem</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>No <code>pg_hba.conf</code> entry for host or user</td>
<td>In order for Greenplum Database to be able to accept remote client connections, you must configure your Greenplum Database master instance so that connections are allowed from the client hosts and database users that will be connecting to Greenplum Database. This is done by adding the appropriate entries to the <code>pg_hba.conf</code> configuration file (located in the master instance’s data directory). For more detailed information, see “Allowing Connections to Greenplum Database” on page 36.</td>
</tr>
<tr>
<td>Greenplum Database is not running</td>
<td>If the Greenplum Database master instance is down, users will not be able to connect. You can verify that the Greenplum Database system is up by running the <code>gpstate</code> utility on the Greenplum master host.</td>
</tr>
<tr>
<td>Network problems Interconnect timeouts</td>
<td>If users are connecting to the Greenplum master host from a remote client, network problems may be preventing a connection (for example, DNS host name resolution problems, the host system is down, etc.). To ensure that network problems are not the cause, try connecting to the Greenplum master host from the remote client host. For example: <code>ping hostname</code>. If the system cannot resolve the host names and IP addresses of the hosts involved in Greenplum Database, queries and connections will fail. Keep in mind that for some operations, connections to the Greenplum Database master use <code>localhost</code> while others use the actual host name, so you must be able to resolve both. If you encounter this error, first make sure you can connect to each host in your Greenplum Database array from the master host over the network. In the <code>/etc/hosts</code> file of the master and all segments, make sure you have the correct host names and IP addresses for all hosts involved in the Greenplum Database array. The <code>127.0.0.1</code> IP should only resolve to <code>localhost</code>.</td>
</tr>
<tr>
<td>Too many clients already</td>
<td>By default, Greenplum Database is configured to allow a maximum of 25 concurrent user connections. A connection attempt that causes that limit to be exceeded will be refused. This limit is controlled by the <code>max_connections</code> parameter in the <code>postgresql.conf</code> configuration file of the Greenplum Database master. If you change this setting for the master, you must also make appropriate changes at the segments.</td>
</tr>
</tbody>
</table>
8. Managing Workload and Resources

This chapter describes the workload management feature of Greenplum Database, and explains the tasks involved in creating and managing resource queues. The following topics are covered in this chapter:

- Overview of Greenplum Workload Management
- Configuring Workload Management
- Creating Resource Queues
- Assigning Roles (Users) to a Resource Queue
- Modifying Resource Queues
- Checking Resource Queue Status

Overview of Greenplum Workload Management

The purpose of workload management is to limit the number of active queries in the system at any given time in order to avoid exhausting system resources such as memory, CPU, and disk I/O. This is accomplished in Greenplum Database with role-based resource queues. A resource queue has attributes that limit the size and/or total number of queries that can be executed by the users (or roles) in that queue. Also, you can assign a priority level that controls the relative share of available CPU used by queries associated with the resource queue. By assigning all database roles to the appropriate resource queue, administrators can control concurrent user queries and prevent the system from being overloaded.

How Resource Queues Work in Greenplum Database

Resource scheduling is enabled by default when you install Greenplum Database. All database roles must be assigned to a resource queue. If an administrator creates a role without explicitly assigning it to a resource queue, the role is assigned to the default resource queue, pg_default.

Greenplum recommends that administrators create resource queues for the various types of workloads in their organization. For example, you may have resource queues for power users, web users, and management reports. The administrator would then set limits on the resource queue based on his estimate of how resource-intensive the queries associated with that workload are likely to be. Currently, the configurable limits on a queue include:

- Active statement cost. This value is compared with the cost estimated by the query planner, measured in units of disk page fetches.
- Active statement count. The maximum number of statements that can run concurrently.
- Active statement priority. This value defines a queue’s priority relative to other queues in the usage of available CPU resources.
After resource queues are created, database roles (users) are then assigned to the appropriate resource queue. A resource queue can have multiple roles, but a role can have only one assigned resource queue.

**How Resource Queue Limits Work**

At runtime, when the user submits a query for execution, that query is evaluated against the resource queue’s limits. If the query does not cause the queue to exceed its resource limits, then that query will run immediately. If the query causes the queue to exceed its limits (for example, if the maximum number of active statement slots are currently in use), then the query must wait until queue resources are free before it can run. Queries are evaluated on a first in, first out basis. If query prioritization is enabled, the active workload on the system is periodically assessed and processing resources are reallocated according to query priority (see “How Resource Queue Priorities Work” on page 49).

**Figure 8.1  Resource Queue Example**

Roles with the `SUPERUSER` attribute are always exempt from resource queue limits. Superuser queries are always allowed to run immediately regardless of the limits of their assigned resource queue.

**How Resource Queue Priorities Work**

Resource limits on active statement count and query cost are *admission* limits, which determine whether a query is admitted into the group of actively running statements, or whether it is queued with other waiting statements. After a query becomes active, it must share available CPU resources as determined by the priority settings for its resource queue. When a statement from a high-priority queue enters the group of actively running statements, it may claim a significant share of the available CPU, reducing the share allotted to already-running statements.

The comparative size or complexity of the queries does not affect the allotment of CPU. If a simple, low-cost query is running simultaneously with a large, complex query, and their priority settings are the same, they will be allotted the same share of available CPU resources. When a new query becomes active, the exact percentage shares of CPU will be recalculated, but queries of equal priority will still have equal amounts of CPU allotted.
For example, an administrator creates three resource queues: *adhoc* for ongoing queries submitted by business analysts, *reporting* for scheduled reporting jobs, and *executive* for queries submitted by executive user roles. The administrator wants to ensure that scheduled reporting jobs are not heavily affected by unpredictable resource demands from ad-hoc analyst queries. Also, the administrator wants to make sure that queries submitted by executive roles are allotted a significant share of CPU. Accordingly, the resource queue priorities are set as shown:

- *adhoc* — Low priority
- *reporting* — High priority
- *executive* — Maximum priority

For more information about commands to set priorities, see “Setting Priority Levels” on page 54.

At runtime, the CPU share of active statements is determined by these priority settings. If queries 1 and 2 from the reporting queue are running simultaneously, they have equal shares of CPU. When an ad-hoc query becomes active, it claims a smaller share of CPU. The exact share used by the reporting queries is adjusted, but remains equal due to their equal priority setting:

![Diagram showing CPU share readjusted according to priority](image.png)

**Figure 8.2** CPU share readjusted according to priority

**Note:** The percentages shown in these graphics are approximate. CPU usage between high, low and maximum priority queues is not always calculated in precisely these proportions.
When an executive query enters the group of running statements, CPU usage is adjusted to account for its maximum priority setting. It may be a simple query compared to the analyst and reporting queries, but until it is completed, it will claim the largest share of CPU.

Figure 8.3  CPU share readjusted for maximum priority query

Types of Queries Evaluated for Resource Queues
Not all SQL statements submitted through a resource queue are evaluated against the queue limits. By default only SELECT, SELECT INTO, CREATE TABLE AS SELECT, and DECLARE CURSOR statements are evaluated. If the server configuration parameter resource_select_only is set to off, then INSERT, UPDATE, and DELETE statements will be evaluated as well.

Steps to Enable Workload Management
Enabling and using workload management in Greenplum Database involves the following high-level tasks:

1. Creating the resource queues and setting limits on them. See “Creating Resource Queues” on page 53.

2. Assigning a queue to one or more user roles. See “Assigning Roles (Users) to a Resource Queue” on page 55.

3. Using the workload management system views to monitor and manage the resource queues. See “Checking Resource Queue Status” on page 56.
Configuring Workload Management

Resource scheduling is enabled by default when you install Greenplum Database, and is required for all roles. The default resource queue, `pg_default`, has an active statement limit of 20, no cost limit, and a medium priority setting. Greenplum recommends that you create resource queues for the various types of workloads in your organization.

To configure workload management

1. You can configure the following workload management parameters:
   - `max_resource_queues` - Sets the maximum number of resource queues.
   - `max_resource_portals_per_transaction` - Sets the maximum number of simultaneously open cursors allowed per transaction. Note that an open cursor will hold an active query slot in a resource queue.
   - `resource_select_only` - If set to `on`, then `SELECT`, `SELECT INTO`, `CREATE TABLE AS SELECT`, and `DECLARE CURSOR` commands are evaluated. If set to `off` `INSERT`, `UPDATE`, and `DELETE` commands will be evaluated as well.
   - `resource_cleanup_gangs_on_wait` - Cleans up idle segment worker processes before taking a slot in the resource queue.
   - `stats_queue_level` - Enables statistics collection on resource queue usage, which can then be viewed by querying the `pg_stat_resqueues` system view.

2. The following parameters are related to query prioritization. Note that the following parameters are all `local` parameters, meaning they must be set in the `postgresql.conf` files of the master and all segments:
   - `gp_resqueue_priority` - The query prioritization feature is enabled by default.
   - `gp_resqueue_priority_sweeper_interval` - Sets the interval at which CPU usage is recalculated for all active statements. The default value for this parameter should be sufficient for typical database operations.
   - `gp_resqueue_priority_cpucores_per_segment` - Specifies the number of CPU cores per segment. The default is 4 for segments and 24 for the master, the correct values for the EMC Greenplum Data Computing Appliance. Each host checks its own `postgresql.conf` file for the value of this parameter. This parameter also affects the master node, where it should be set to a value reflecting the higher ratio of CPU cores. For example, on a cluster that has 8 CPU cores per host and 4 segments per host, you would use the following settings:
     - Master and standby master
       `gp_resqueue_priority_cpucores_per_segment = 8`
     - Segment hosts
       `gp_resqueue_priority_cpucores_per_segment = 2`
Important: If you have fewer than one segment per CPU core on your segment hosts, make sure you adjust this value accordingly. An improperly low value for this parameter can result in under-utilization of CPU resources.

3. If you wish to change any of the default parameter values, you can use the `gpconfig` utility. For example, to set one value on all segments and a different value on the master:

   ```
   $ gpconfig -c gp_resqueue_priority_cpus_per_segment -v 2 -m 8
   ```

4. Restart Greenplum Database to make the configuration changes effective:

   ```
   $ gpstop -r
   ```

Creating Resource Queues

Creating a resource queue involves giving it a name and setting either a query cost limit or an active query limit (or both), and optionally a query priority on the resource queue. Use the `CREATE RESOURCE QUEUE` command to create new resource queues.

Creating Queues with an Active Query Limit

Resource queues with an `ACTIVE_STATEMENTS` setting limit the number of queries that can be executed by roles assigned to that queue. For example, to create a resource queue named `adhoc` with an active query limit of three:

```
=# CREATE RESOURCE QUEUE adhoc WITH (ACTIVE_STATEMENTS=3);
```

This means that for all roles assigned to the `adhoc` resource queue, only three active queries can be running on the system at any given time. If this queue has three queries running, and a fourth query is submitted by a role in that queue, that query must wait until a slot is free before it can run.

Creating Queues with a Query Planner Cost Limits

Resource queues with a `MAX_COST` setting limit the total cost of queries that can be executed by roles assigned to that queue. Cost is specified as a floating point number (for example 100.0) or can also be specified as an exponent (for example 1e+2).

Cost is measured in the estimated total cost for the query as determined by the Greenplum query planner (as shown in the `EXPLAIN` output for a query). Therefore, an administrator must be familiar with the queries typically executed on the system in order to set an appropriate cost threshold for a queue. Cost is measured in units of disk page fetches; 1.0 equals one sequential disk page read.

For example, to create a resource queue named `webuser` with a query cost limit of 100000.0 (1e+5):

```
=# CREATE RESOURCE QUEUE webuser WITH (MAX_COST=100000.0);
```

or
Creating Resource Queues

```sql
=# CREATE RESOURCE QUEUE webuser WITH (MAX _COST=1e+5);
```

This means that for all roles assigned to the `webuser` resource queue, it will only allow queries into the system until the cost limit of 100000.0 is reached. So for example, if this queue has 200 queries with a 500.0 cost all running at the same time, and query 201 with a 1000.0 cost is submitted by a role in that queue, that query must wait until space is free before it can run.

**Allowing Queries to Run on Idle Systems**

If a resource queue is limited based on a cost threshold, then the administrator can allow `COST_OVERCOMMIT` (the default). Resource queues with a cost threshold and overcommit enabled will allow a query that exceeds the cost threshold to run, provided that there are no other queries in the system at the time the query is submitted. The cost threshold will still be enforced if there are concurrent workloads on the system.

If `COST_OVERCOMMIT` is false, then queries that exceed the cost limit will always be rejected and never allowed to run.

**Allowing Small Queries to Bypass Queue Limits**

Workloads may have certain small queries that administrators want to allow to run without taking up an active statement slot in the resource queue. For example, simple queries to look up metadata information in the system catalogs do not typically require significant resources or interfere with query processing on the segments. An administrator can set `MIN_COST` to denote a query planner cost associated with a small query. Any query that falls below the `MIN_COST` limit will be allowed to run immediately. `MIN_COST` can be used on resource queues with either an active statement or a maximum query cost limit. For example:

```sql
=# CREATE RESOURCE QUEUE adhoc WITH (ACTIVE_STATEMENTS=10, MIN_COST=100.0);
```

**Setting Priority Levels**

To control a resource queue’s consumption of available CPU resources, an administrator can assign an appropriate priority level. When high concurrency causes contention for CPU resources, queries and statements associated with a high-priority resource queue will claim a larger share of available CPU than lower priority queries and statements.

Priority settings are created or altered using the `WITH` parameter of the commands `CREATE RESOURCE QUEUE` and `ALTER RESOURCE QUEUE`. For example, to specify priority settings for the `adhoc` and `reporting` queues, an administrator would use the following commands:

```sql
=# ALTER RESOURCE QUEUE adhoc WITH (PRIORITY=LOW);
=# ALTER RESOURCE QUEUE reporting WITH (PRIORITY=HIGH);
```

To create the `executive` queue with maximum priority, an administrator would use the following command:

```sql
=# CREATE RESOURCE QUEUE executive WITH (ACTIVE_STATEMENTS=3, PRIORITY=MAX);
```
When the query prioritization feature is enabled, resource queues are given a MEDIUM priority by default if not explicitly assigned. For more information on how priority settings are evaluated at runtime, see “How Resource Queue Priorities Work” on page 49.

**Note:** In order for resource queue priority levels to be enforced on the active query workload, you must enable the query prioritization feature by setting the associated server configuration parameters. See “Configuring Workload Management” on page 52.

### Assigning Roles (Users) to a Resource Queue

Once a resource queue is created, you must assign roles (users) to their appropriate resource queue. If roles are not explicitly assigned to a resource queue, they will go to the default resource queue, `pg_default`. The default resource queue has an active statement limit of 20, no cost limit, and a medium priority setting.

Use the `ALTER ROLE` or `CREATE ROLE` commands to assign a role to a resource queue. For example:

```
=# ALTER ROLE name RESOURCE QUEUE queue_name;
=# CREATE ROLE name WITH LOGIN RESOURCE QUEUE queue_name;
```

A role can only be assigned to one resource queue at any given time, so you can use the `ALTER ROLE` command to initially assign or change a role’s resource queue.

Resource queues must be assigned on a user-by-user basis. If you have a role hierarchy (for example, a group-level role) then assigning a resource queue to the group does not propagate down to the users in that group.

Superusers are always exempt from resource queue limits. Superuser queries will always run regardless of the limits set on their assigned queue.

### Removing a Role from a Resource Queue

All users *must* be assigned to a resource queue. If not explicitly assigned to a particular queue, users will go into the default resource queue, `pg_default`.

If you wish to remove a role from a resource queue and put them in the default queue, change the role’s queue assignment to `none`. For example:

```
=# ALTER ROLE role_name RESOURCE QUEUE none;
```

### Modifying Resource Queues

After a resource queue has been created, you can change or reset the queue limits using the `ALTER RESOURCE QUEUE` command. You can remove a resource queue using the `DROP RESOURCE QUEUE` command. To change the roles (users) assigned to a resource queue, see “Assigning Roles (Users) to a Resource Queue” on page 55.
Altering a Resource Queue

The `ALTER RESOURCE QUEUE` command changes the limits of a resource queue. A resource queue must have either an `ACTIVE_STATEMENTS` or a `MAX_COST` value (or it can have both). To change the limits of a resource queue, specify the new values you want for the queue. For example:

```
=# ALTER RESOURCE QUEUE adhoc WITH (ACTIVE_STATEMENTS=5);
=# ALTER RESOURCE QUEUE exec WITH (MAX_COST=100000.0);
```

To reset active statements to no limit, enter a value of -1. To reset the maximum query cost to no limit, enter a value of -1.0. For example:

```
=# ALTER RESOURCE QUEUE adhoc WITH (MAX_COST=-1.0);
```

You can use the `ALTER RESOURCE QUEUE` command to change the priority of queries associated with a resource queue. For example, to set a queue to the minimum priority level:

```
ALTER RESOURCE QUEUE webuser WITH (PRIORITY=MIN);
```

Dropping a Resource Queue

The `DROP RESOURCE QUEUE` command drops a resource queue. To drop a resource queue, the queue cannot have any roles assigned to it, nor can it have any statements waiting in the queue. See “Removing a Role from a Resource Queue” on page 55 and “Clearing a Waiting Statement From a Resource Queue” on page 58 for instructions on emptying a resource queue. To drop a resource queue:

```
=# DROP RESOURCE QUEUE name;
```

Checking Resource Queue Status

Checking resource queue status involves the following tasks:

- Viewing Queued Statements and Resource Queue Status
- Viewing Resource Queue Statistics
- Viewing the Roles Assigned to a Resource Queue
- Viewing the Waiting Queries for a Resource Queue
- Clearing a Waiting Statement From a Resource Queue
- Viewing the Priority of Active Statements
- Resetting the Priority of an Active Statement

Viewing Queued Statements and Resource Queue Status

The `pg_resqueue_status` system view allows administrators to see status and activity for a workload management resource queue. It shows how many queries are waiting to run and how many queries are currently active in the system from a particular resource queue. To see the resource queues created in the system, their limit attributes, and their current status:
Checking Resource Queue Status

Viewing Resource Queue Statistics
If you want to track statistics and performance of resource queues over time, you can enable statistics collecting for resource queues. This is done by setting the following server configuration parameter in your master `postgresql.conf` file:

```
stats_queue_level = on
```

Once this is enabled, you can use the `pg_stat_resqueues` system view to see the statistics collected on resource queue usage. Note that enabling this feature does incur slight performance overhead, as each query submitted through a resource queue must be tracked. It may be useful to enable statistics collecting on resource queues for initial diagnostics and administrative planning, and then disable the feature for continued use.

See the section on the Statistics Collector in the PostgreSQL documentation for more information about collecting statistics in Greenplum Database.

Viewing the Roles Assigned to a Resource Queue
To see the roles assigned to a resource queue, perform the following query of the `pg_roles` and `pg_resqueue` system catalog tables:

```
=# SELECT rolname, rsqname FROM pg_roles, pg_resqueue
    WHERE pg_roles.rolresqueue=pg_resqueue.oid;
```

You may want to create a view of this query to simplify future inquiries. For example:

```
=# CREATE VIEW role2queue AS
    SELECT rolname, rsqname FROM pg_roles, pg_resqueue
    WHERE pg_roles.rolresqueue=pg_resqueue.oid;
```

Then you can just query the view:

```
=# SELECT * FROM role2queue;
```

Viewing the Waiting Queries for a Resource Queue
When a slot is in use for a resource queue, it is recorded in the `pg_locks` system catalog table. This is where you can see all of the currently active and waiting queries for all resource queues. To check that statements are being queued (even statements that are not waiting), you can also use the `pg_resqueue_status` system view.

To see all the currently active queries for all resource queues, perform the following query of the `pg_locks` table joined with the `pg_stat_activity` and `pg_resqueue` tables:

```
=# SELECT usename, rsqname, locktype, objid, transaction, pid, mode, granted, waiting
    FROM pg_stat_activity, pg_resqueue, pg_locks
    WHERE pg_stat_activity.procpid=pg_locks.pid
    AND pg_locks.objid=pg_resqueue.oid;
```
You may want to create a view of this query to simplify future inquiries. For example:

```
=# CREATE VIEW queued_statements AS
SELECT usename, rsqname, locktype, objid, transaction, pid, mode, granted, waiting
FROM pg_stat_activity, pg_resqueue, pg_locks
WHERE pg_stat_activity.procpid=pg_locks.pid
AND pg_locks.objid=pg_resqueue.oid;
```

Then you can just query the view:

```
=# SELECT * FROM queued_statements;
```

If this query returns no results, then that means there are currently no statements in a resource queue. A sample of a resource queue with two statements in it looks something like this:

<table>
<thead>
<tr>
<th>usename</th>
<th>rsqname</th>
<th>locktype</th>
<th>objid</th>
<th>transaction</th>
<th>pid</th>
<th>mode</th>
<th>granted</th>
<th>waiting</th>
</tr>
</thead>
<tbody>
<tr>
<td>sammy</td>
<td>webuser</td>
<td>resource queue</td>
<td>16510</td>
<td>1054</td>
<td>31905</td>
<td>ExclusiveLock</td>
<td>f</td>
<td>t</td>
</tr>
<tr>
<td>daria</td>
<td>webuser</td>
<td>resource queue</td>
<td>16510</td>
<td>1051</td>
<td>31861</td>
<td>ExclusiveLock</td>
<td>t</td>
<td>f</td>
</tr>
</tbody>
</table>

The `objid` is the object ID of the resource queue. Statements with a `granted` status of `t` and a `waiting` status of `f` are currently active. Statements with a `granted` status of `f` and a `waiting` status of `t` are waiting in the queue.

### Clearing a Waiting Statement From a Resource Queue

In some cases, you may want to clear a waiting statement from a resource queue. For example, you may want to remove a query that is waiting in the queue but has not been executed yet. You may also want to stop a query that has been started if it is taking too long to execute, or if it is sitting idle in a transaction and taking up resource queue slots that are needed by other users. To do this, you must first identify the statement you want to clear, determine its process id (pid), and then kill that process id.

For example, to see process information about all statements currently active or waiting in all resource queues, run the following query:

```
=# SELECT rolname, rsqname, pid, granted, current_query, datname
FROM pg_roles, pg_resqueue, pg_locks, pg_stat_activity
WHERE pg_roles.rolresqueue=pg_locks.objid
AND pg_locks.objid=pg_resqueue.oid
AND pg_stat_activity.procpid=pg_locks.pid;
```

You may want to create a view of this query to simplify future inquiries. For example:

```
=# CREATE VIEW resqueue_procs AS
SELECT rolname, rsqname, pid, granted, current_query, datname
FROM pg_roles, pg_resqueue, pg_locks, pg_stat_activity
WHERE pg_roles.rolresqueue=pg_locks.objid
```

---

**Checking Resource Queue Status**

58
AND pg_locks.objid=pg_resqueue.oid
AND pg_stat_activity.procpid=pg_locks.pid;

Then you can just query the view:

```sql
=# SELECT * FROM resqueue_procs;
```

If this query returns no results, then that means there are currently no statements in a resource queue. A sample of a resource queue with two statements in it looks something like this:

<table>
<thead>
<tr>
<th>rolname</th>
<th>rsqname</th>
<th>pid</th>
<th>granted</th>
<th>current_query</th>
<th>datname</th>
</tr>
</thead>
<tbody>
<tr>
<td>sammy</td>
<td>webuser</td>
<td>31861</td>
<td>t</td>
<td>&lt;IDLE&gt; in transaction</td>
<td>namesdb</td>
</tr>
<tr>
<td>daria</td>
<td>webuser</td>
<td>31905</td>
<td>f</td>
<td>SELECT * FROM topten;</td>
<td>namesdb</td>
</tr>
</tbody>
</table>

Use this output to identify the process id (pid) of the statement you want to clear from the resource queue. To clear the statement, you would then open a terminal window (as the `gpadmin` database superuser or as root) on the master host and kill the corresponding process. For example:

```sql
=# kill 31905
```

### Viewing the Priority of Active Statements

The `gp_toolkit` administrative schema has a view called `gp_resq_priority_statement`, which lists all statements currently being executed and provides the priority, session ID, and other information.

<table>
<thead>
<tr>
<th>datname</th>
<th>usename</th>
<th>session_id</th>
<th>command_count</th>
<th>priority</th>
<th>weight</th>
<th>current_query</th>
</tr>
</thead>
</table>
| namesdb | daria   | 752        | 24905         | LOW      | 200    | SELECT * FROM ...
| namesdb | sammy   | 699        | 28999         | HIGH     | 1000   | SELECT * FROM ...

This view is only available through the `gp_toolkit` administrative schema. See Appendix I, “The gp_toolkit Administrative Schema” for more information.

### Resetting the Priority of an Active Statement

Superusers can adjust the priority of a statement currently being executed using the built-in function `gp_adjust_priority(session_id, statement_count, priority)`. Using this function, superusers can raise or lower the priority of any query. For example:

```sql
=# SELECT gp_adjust_priority(752, 24905, 'HIGH')
```

To obtain the session ID and statement count parameters required by this function, Superusers can use the `gp_toolkit` administrative schema view, `gp_resq_priority_statement`. This function affects only the specified statement. Subsequent statements in the same resource queue are executed using the queue’s normally assigned priority.
Section III: Database Administration

This section describes how create, manage and access databases and database objects using SQL (structured query language).

• **Defining Database Objects** - This chapter covers data definition language (DDL) in Greenplum Database and how to create and manage database objects.

• **Managing Data** - This chapter covers data manipulation language (DML) in Greenplum Database and how transaction concurrency is handled.

• **Querying Data** - This chapter describes the use of the SQL language in Greenplum Database.

• **Loading and Unloading Data** - This chapter describes the various ways to load data into Greenplum Database.
9. Defining Database Objects

This chapter covers data definition language (DDL) in Greenplum Database and how to create and manage database objects.

- Creating and Managing Databases
- Creating and Managing Tablespaces
- Creating and Managing Schemas
- Creating and Managing Tables
- Partitioning Large Tables
- Creating and Using Sequences
- Using Indexes in Greenplum Database
- Creating and Managing Views

Creating and Managing Databases

A Greenplum Database system can have one or more databases. This is different from some other database management systems (such as Oracle) where the database instance is the database. Although you can create many databases in a Greenplum system, client programs can connect and access one database at a time — you cannot cross-query between databases.

About Template Databases

Every new database you create is based on a template. A default database called `template1` exists in every newly initialized Greenplum Database system. You can use this database to connect to Greenplum Database for the first time. This is the template used to create other databases by default if you do not explicitly declare a template when creating a new database. You do not want to create any objects in this database unless you want those objects to also be in every other database you create afterwards.

In addition to `template1`, every newly created Greenplum system has two other database templates, `template0` and `postgres`, which are used internally by the system and should not be dropped or modified. The `template0` database template can be used to create a completely clean database containing only the standard objects predefined by Greenplum Database at initialization. This is useful if you wish to avoid copying any objects that may have been added to `template1`.

Creating a Database

The `CREATE DATABASE` command creates a new database. For example:

```sql
=> CREATE DATABASE new_dbname;
```
In order to create a database, you must have privileges to create a database or be a
Greenplum superuser. If you do not have the correct privileges, then you will not be
able to create a database. Contact your Greenplum administrator to either give you the
necessary privilege or to create a database for you.

There is also a client program called createdb that you can use to create a database.
For example, running the following command in a command line terminal will make a
connection to Greenplum Database using the provided host name and port and create
the database named mydatabase:

$ createdb -h masterhost -p 5432 mydatabase

Cloning a Database

By default, a new database is created by cloning the standard system database
template, template1. Any database can be used as a template when creating a new
database, thereby providing the capability to ‘clone’ or copy an existing database and
all of the objects and data within that database. For example:

=> CREATE DATABASE new_dbname TEMPLATE old_dbname;

Viewing the List of Databases

If you are working in the psql client program, you can use the \l meta-command to
show the list of databases and templates in your Greenplum Database system. If using
another client program, you can query the list of databases from the pg_database
system catalog table (you must be a superuser). For example:

=> SELECT datname from pg_database;

Altering a Database

The ALTER DATABASE command is used to change certain attributes about a database,
such its owner, name or default configuration attributes. You must be either the owner
of the database or a superuser to alter it. Here is an example of altering a database to
set its default schema search path (the search_path configuration parameter):

=> ALTER DATABASE mydatabase SET search_path TO myschema,
   public, pg_catalog;

Dropping a Database

The DROP DATABASE command can be used to drop (or delete) a database. It removes
the system catalog entries for the database and also deletes the database directory on
disk containing the data. You must be the database owner or a superuser to drop a
database, and you cannot drop a database while you or anyone else is connected to it.
Connect into template1 (or another database) before dropping a database. For
example:

=> \c template1
=> DROP DATABASE mydatabase;
There is also a client program called `dropdb` that you can use to drop a database. For example, running the following command in a command line terminal will make a connection to Greenplum Database using the provided host name and port and drop the database named `mydatabase`:

```bash
$ dropdb -h masterhost -p 5432 mydatabase
```

**Warning:** Dropping a database cannot be undone, so use this command with care!

---

**Creating and Managing Tablespaces**

Tablespaces allow database administrators to have multiple file systems per machine and decide how to best use their physical storage to store database objects. Tablespaces are useful for a number of reasons, such as allowing different storage types for frequently versus infrequently used database objects, or controlling the I/O performance on certain database objects. For example, highly utilized tables can be placed on file systems that use high performance solid-state drives (SSD), while the remaining tables utilize standard hard drives.

A tablespace requires a file system location to store its database files. In Greenplum Database, the master and each segment (primary and mirror) needs its own distinct storage location. This collection of file system locations for all components in a Greenplum system is referred to as a *filespace*. Once a filespace is defined, it can be used by one or more tablespaces.

### Creating a Filespace

To create a filespace, you must first prepare the logical file systems on all of your Greenplum hosts. A file system location is needed for the master, each primary segment and each mirror segment on a host. Once you have your file systems prepared, use the `gpfilespace` utility to define the filespace. You must be a database superuser in order to create a filespace.

**Note:** Greenplum is not directly aware of the file system boundaries on your underlying systems. It will store files in the directories you tell it to use. Therefore, there is not much point in defining multiple filespaces that use the same logical file system since you cannot control the location on disk of individual files within a logical file system.

#### To create a filespace using `gpfilespace`

1. Log in to the Greenplum Database master as the `gpadmin` user.
   ```bash
   $ su - gpadmin
   ```

2. Create a filespace configuration file:
   ```bash
   $ gpfilespace -o gpfilespace_config
   ```
3. You will be prompted to enter a name for the filespace, the primary segment file system locations, the mirror segment file system locations, and a master file system location. For example, if your configuration has 2 primary and 2 mirror segments per host, you will be prompted for a total of 5 locations (including the master). For example:

   Enter a name for this filespace> fastdisk
   primary location 1> /gpfs1/seg1
   primary location 2> /gpfs1/seg2
   mirror location 1> /gpfs2/mir1
   mirror location 2> /gpfs2/mir2
   master location> /gpfs1/master

4. The utility will output a configuration file. Examine the file and make sure it reflects the filespace configuration you want.

5. Run the utility again to create the filespace based on the configuration file:

   $ gpfilespace -c gpfilespace_config

---

**Creating a Tablespace**

Once you have created a filespace, you can define a tablespace that uses that filespace. To define a tablespace, use the `CREATE TABLESPACE` command, for example:

   =# CREATE TABLESPACE fastspace FILESPACE fastdisk;

Creation of the tablespace itself must be done as a database superuser, but after that you can allow ordinary database users to make use of it. To do that, `GRANT` them the `CREATE` privilege on it. For example:

   =# GRANT CREATE ON TABLESPACE fastspace TO admin;

---

**Using a Tablespace to Store Database Objects**

Tables, indexes, and entire databases can be assigned to particular tablespaces. To do so, a user with the `CREATE` privilege on a given tablespace must pass the tablespace name as a parameter to the relevant command. For example, the following creates a table in the tablespace `space1`:

   CREATE TABLE foo(i int) TABLESPACE space1;

Alternatively, use the `default_tablespace` parameter:

   SET default_tablespace = space1;
   CREATE TABLE foo(i int);

When `default_tablespace` is set to anything but an empty string, it supplies an implicit `TABLESPACE` clause for `CREATE TABLE` and `CREATE INDEX` commands that do not have an explicit one.

The tablespace associated with a database is used to store the system catalogs of that database, as well as any temporary files created by server processes using that database. Furthermore, it is the default tablespace selected for tables and indexes created within the database, if no `TABLESPACE` clause is given (either explicitly or via
default_tablespace) when the objects are created. If a database is created without specifying a tablespace for it, it uses the same tablespace as the template database it is copied from.

Once created, a tablespace can be used from any database, provided the requesting user has sufficient privilege.

### Viewing Existing Tablespaces and Filespaces

Every Greenplum Database system has two default tablespaces: `pg_global` (used for shared system catalogs) and `pg_default` (the default tablespace used by the `template1` and `template0` databases). These tablespaces use the system default filesystem, `pg_system` (the data directory locations created at system initialization time).

To see filesystem information, look in the `pg_filespace` and `pg_filespace_entry` catalog tables. You can join these tables with `pg_tablespace` to see the full definition of a tablespace. For example:

```sql
=# SELECT spcname as tblspc, fsname as filespc, 
    fsedbid as seg_dbid, fselocation as datadir 
FROM pg_tablespace pgts, pg_filespace pgfs, 
   pg_filespace_entry pgfse 
WHERE pgts.spcfsoid=pgfse.fsefsoid 
AND pgfse.fsefsoid=pgfs.oid 
ORDER BY tblspc, seg_dbid;
```

### Dropping Tablespaces and Filespaces

A tablespace cannot be dropped until all objects in all databases using the tablespace have been removed. A filesystem cannot be dropped until all tablespaces using that filesystem have also been removed.

To remove an empty tablespace, use the `DROP TABLESPACE` command. A tablespace can only be dropped by its owner or a superuser.

To remove an empty filesystem, use the `DROP FILESPACE` command. Only a superuser can drop a filesystem.

### Creating and Managing Schemas

Schemas are a way to logically organize objects and data in a database. Schemas allow you to have more than one object (such as tables) with the same name in the database without conflict, as long as they are in different schemas.
The Default ‘Public’ Schema

Every newly created database has a default schema named public. If you do not create any schemas of your own, objects will be created in the public schema by default. All database roles (users) have CREATE and USAGE privileges in the public schema by default. Any other schemas you create, you will have to grant the appropriate privileges so that users can access the schema.

Creating a Schema

Use the CREATE SCHEMA command to create a new schema. For example:

```sql
=> CREATE SCHEMA myschema;
```

To create or access objects in a schema, write a qualified name consisting of the schema name and table name separated by a dot. For example:

```sql
schema.table
```

See “Schema Search Paths” on page 66 for more information about accessing a schema.

You may want to create a schema owned by someone else (since this is one of the ways to restrict the activities of your users to well-defined namespaces). The syntax for that is:

```sql
=> CREATE SCHEMA schemaname AUTHORIZATION username;
```

Schema Search Paths

For the database to know in which schema it should look for an object, you can always use the schema-qualified name. For example:

```sql
=> SELECT * FROM myschema.mytable;
```

If you do not want to type the schema-qualified name all the time, you can set the search_path configuration parameter. This tells the database in which order it should search the available schemas for objects. The schema listed first in the search path becomes the default schema. The default schema is where new objects will be created if a schema name is not explicitly declared.

Setting the Schema Search Path

The search_path configuration parameter is used to set the order in which schemas are searched. You can set search_path for a database using the ALTER DATABASE command. For example:

```sql
=> ALTER DATABASE mydatabase SET search_path TO myschema, public, pg_catalog;
```

You can also set search_path for a particular role (user) using the ALTER ROLE command. For example:

```sql
=> ALTER ROLE sally SET search_path TO myschema, public, pg_catalog;
```
**Viewing the Current Schema**

There may be times when you are not sure what schema you are currently in or what your `search_path` setting is. To find out this information, you can use the `current_schema()` function or the `SHOW` command. For example:

```sql
=> SELECT current_schema();
=> SHOW search_path;
```

**Dropping a Schema**

Use the `DROP SCHEMA` command to drop (delete) a schema. For example:

```sql
=> DROP SCHEMA myschema;
```

By default, the schema must be empty before you can drop it. If you want to drop a schema and all of the objects in that schema (tables, data, functions, etc.) use:

```sql
=> DROP SCHEMA myschema CASCADE;
```

**System Schemas**

The following system-level schemas also exist in every database:

- **pg_catalog** is the schema that has the system catalog tables, built-in data types, functions, and operators. It is always part of the schema search path, even if it is not explicitly named in the search path. For information on the system catalog tables in the `pg_catalog` schema, see “System Catalog Reference” on page 825.

- **information_schema** consists of a standardized set of views that contain information about the objects in the database. These views are used to get system information from the system catalog tables in a standardized way.

- **pg_toast** is a system schema where large objects are stored (records that exceed the page size). This schema is used internally by the Greenplum Database system and is not typically accessed by database administrators or users.

- **pg_bitmapindex** is the system schema where bitmap index objects are stored (list of values, etc.). This schema is used internally by the Greenplum Database system and is not typically accessed by database administrators or users.

- **pg_aoseg** is the system schema where append-only table objects are stored. This schema is used internally by the Greenplum Database system and is not typically accessed by database administrators or users.

- **gp_toolkit** is an administrative schema that you can use to view and query the system log files and other system metrics. The `gp_toolkit` schema contains a number of external tables, views and functions that you can access by using SQL commands. `gp_toolkit` is accessible by all database users. See Appendix I, “The gp_toolkit Administrative Schema” for more information.
**Creating and Managing Tables**

A table in Greenplum Database is much like a table in any other relational database, except that the records in the table are distributed across the different segments in the system. When you create a table, there is additional SQL syntax to declare the table’s distribution policy.

**Creating a Table**

The `CREATE TABLE` command is used to create a new table and define its structure. When creating a table, you will typically define the following aspects of the table:

- The columns of the table and their associated data types. See “Choosing Column Data Types” on page 68.
- Any table or column constraints to limit the data that a column or table can contain. See “Setting Table and Column Constraints” on page 69.
- The distribution policy of the table, which determines how the data is divided across the Greenplum Database segments. See “Choosing the Table Distribution Policy” on page 70.
- The way the table is stored on disk. See “Choosing the Table Storage Model” on page 71.
- The table partitioning strategy for large tables. See “Partitioning Large Tables” on page 76.

**Choosing Column Data Types**

The data type of a column determines the types of data values that column can contain. As a general rule, you want to choose the data type that uses the least possible space, yet can still accommodate your data. You should also select the data type that best constrains the data in that column. For example, use character data types for strings, date or timestamp data types for dates, and numeric data types for numbers.

For character column data types, there are no performance differences between the use of `CHAR`, `VARCHAR`, and `TEXT` data types, apart from the increased storage size when using the blank-padded type. While `CHAR` has performance advantages in some other database systems, it has no such advantages in Greenplum Database. In most situations, `TEXT` or `VARCHAR` should be used instead.

For numeric column data types, use the smallest data type in which the data will fit. A lot of space is wasted if, for example, the `BIGINT` data type is used when the data would always fit in `INT` or `SMALLINT`.

Also consider using identical data types for the columns you plan to use in cross-table joins. Joins work much more efficiently if the data types of the columns used in the join predicate (usually the primary key in one table and a foreign key in the other table) have identical data types. When the data types are different, the database has to convert one of them so that the data values can be compared correctly, and such conversion amounts to unnecessary overhead.

Greenplum Database has a rich set of native data types available to users. See “Greenplum Database Data Types” on page 822 for a list of the built-in data types.
Setting Table and Column Constraints

Data types are a way to limit the kind of data that can be stored in a table. For many applications, however, the constraint they provide is too coarse. SQL allows you to define constraints on columns and tables. Constraints give you more control over the data in your tables. If a user attempts to store data in a column that would violate a constraint, an error is raised.

There are some limitations and conditions when using constraints in Greenplum Database, most notably with regards to foreign key, primary key, and unique constraints. Otherwise constraints are supported as in PostgreSQL.

Check Constraints

A check constraint is the most generic constraint type. It allows you to specify that the value in a certain column must satisfy a Boolean (truth-value) expression. For instance, to require positive product prices, you could use:

```sql
=> CREATE TABLE products
    ( product_no integer,
      name text,
      price numeric CHECK (price > 0) );
```

Not-Null Constraints

A not-null constraint simply specifies that a column must not assume the null value. A not-null constraint is always written as a column constraint. For example:

```sql
=> CREATE TABLE products
    ( product_no integer NOT NULL,
      name text NOT NULL,
      price numeric );
```

Unique Constraints

Unique constraints ensure that the data contained in a column or a group of columns is unique with respect to all the rows in the table. For a unique constraint to be enforced in Greenplum Database, the table must be hash-distributed (not DISTRIBUTED RANDOMLY), and the constraint columns must be the same as (or a superset of) the table’s distribution key columns.

```sql
=> CREATE TABLE products
    ( product_no integer UNIQUE,
      name text, price numeric )
    DISTRIBUTED BY (product_no);
```

Primary Keys

A primary key constraint is simply a combination of a UNIQUE constraint and a NOT NULL constraint. For a primary key constraint to be enforced in Greenplum Database, the table must be hash-distributed (not DISTRIBUTED RANDOMLY), and the primary key columns must be the same as (or a superset of) the table’s distribution key columns. If a table has a primary key, this column (or group of columns) is chosen as the distribution key for the table by default. For example:

```sql
=> CREATE TABLE products
    ( product_no integer PRIMARY KEY,
      name text, price numeric )
    DISTRIBUTED BY (product_no);
```
name text,
    price numeric)
DISTRIBUTED BY (product_no);

Foreign Keys
Foreign keys are not supported in this release of Greenplum Database. It is possible to declare them, but referential integrity will not be enforced.

A foreign key constraint specifies that the values in a column (or a group of columns) must match the values appearing in some row of another table. This maintains the referential integrity between two related tables. In this release of Greenplum Database, referential integrity checks cannot be enforced between the distributed table segments of a Greenplum database.

Choosing the Table Distribution Policy
All Greenplum Database tables are distributed. When you create or alter a table, there is an optional DISTRIBUTED BY (hash distribution) or DISTRIBUTED RANDOMLY (round-robin distribution) clause to declare how the rows of the table should be distributed. See “Understanding Greenplum Distribution Policies” on page 13.

The following considerations should be taken into account when declaring a distribution policy for a table (in order of importance):

- **Even Data Distribution** — For the best possible performance, all of the segments should contain equal portions of data. If the data is unbalanced or skewed, then the segments with more data will have to work harder to perform their portion of the query processing. To ensure an even distribution of data, you want to choose a distribution key that is unique for each record, such as the primary key.

- **Local and Distributed Operations** — During query processing, it is faster if the work associated with join, sort or aggregation operations can be done locally at the segment-level rather than at the system-level (distributing tuples across the segments). When tables share a common distribution key in Greenplum Database, joining or sorting on their shared distribution key columns will result in the most efficient query processing, as the majority of the work is done locally at the segment-level. Local operations are approximately 5 times faster than distributed operations. With a random distribution policy, local operations are not an option.

- **Even Query Processing** — When a query is being processed, you want all of the segments to handle an equal amount of the query workload to get the best possible performance. In some cases, query processing workload can be skewed if the table’s data distribution policy and the query predicates are not well matched. For example, suppose you have a table of sales transactions. The table is distributed based on a column that contains corporate names as values. The hashing algorithm distributes the data based on the values of the distribution key, so if a predicate in a query references a single value from the distribution key, the work in the query will run on only one segment. This may be a viable distribution policy if your query predicates tend to select data on a criteria other than corporation name. However, for queries that do use corporation name in their predicates, you can potentially have just one segment instance handling all of the query workload.
Declaring Distribution Keys

When creating a table, there is an additional clause to declare the Greenplum Database distribution policy. If a `DISTRIBUTED BY` or `DISTRIBUTED RANDOMLY` clause is not supplied, then Greenplum assigns a hash distribution policy to the table using either the PRIMARY KEY (if the table has one) or the first column of the table as the distribution key. Columns of geometric or user-defined data types are not eligible as Greenplum distribution key columns. If a table does not have a column of an eligible data type, the rows are distributed based on a round-robin or random distribution.

To ensure an even distribution of data, you want to choose a distribution key that is unique for each record, or if that is not possible, then choose `DISTRIBUTED RANDOMLY`. For example:

```sql
=> CREATE TABLE products
    (name varchar(40),
     prod_id integer,
     supplier_id integer)
    DISTRIBUTED BY (prod_id);

=> CREATE TABLE random_stuff
    (things text,
     doodads text,
     etc text)
    DISTRIBUTED RANDOMLY;
```

Choosing the Table Storage Model

Greenplum Database provides an agile and flexible processing engine capable of supporting several storage models (or a hybrid of storage models). When you create a new table, you have several options as to how its data is stored on disk. This section explains the various options for table storage and how to decide on the best storage model for your workload.

- Choosing Heap or Append-Only Storage
- Choosing Row or Column-Oriented Storage
- Using Compression (Append-Only Tables Only)
- Checking the Compression and Distribution of an Append-Only Table

Choosing Heap or Append-Only Storage

By default, Greenplum Database uses the same heap storage model as PostgreSQL. Heap table storage favors OLTP-type workloads where the data is often modified after it is initially loaded. `UPDATE` and `DELETE` operations require row-level versioning information to be stored in order to ensure that database transactions are processed reliably. Heap tables are best suited for smaller tables, such as dimension tables, that are often updated after they are initially loaded.

Greenplum Database also offers an append-only table storage model. Append-only table storage favors denormalized fact tables in a data warehouse environment, which are typically the largest tables in the system. Fact tables are typically loaded in batches, and then accessed by read-only queries. Data is not updated after it is loaded. Moving large fact tables to an append-only storage model eliminates the storage overhead of the per-row update visibility information (about 20 bytes per row is saved). This allows for a leaner and easier-to-optimize page structure. Append-only
tables do not allow UPDATE and DELETE operations. The storage model of append-only tables is optimized for bulk data loading. Single row INSERT statements are not recommended.

To create a heap table
Row-oriented heap tables are the default storage type, so no extra CREATE TABLE command syntax is required to create a heap table. For example:

```sql
=> CREATE TABLE foo (a int, b text) DISTRIBUTED BY (a);
```

To create an append-only table
The WITH clause of the CREATE TABLE command is used to declare the storage options of the table. If not declared, the table will be created as a regular row-oriented heap-storage table. For example, to create an append-only table with no compression:

```sql
=> CREATE TABLE bar (a int, b text)
   WITH (appendonly=true)
   DISTRIBUTED BY (a);
```

Choosing Row or Column-Oriented Storage
Greenplum provides a choice of storage orientation models: row or column (or a hybrid of both). This section provides some general guidelines for choosing the correct storage orientation for a table; however you are encouraged to evaluate performance using your own data and query workloads.

For most general purpose or mixed workloads, row-oriented storage offers the best combination of flexibility and performance. However, there are certain specific use cases where a column-oriented storage model provides more efficient I/O and storage. Consider the following requirements when deciding on the storage orientation model of a table:

- **Updates of table data.** If table data must be updated after it is loaded, choose a row-oriented heap table. Column-oriented table storage is only available on append-only tables. See “Choosing Heap or Append-Only Storage” on page 71 for more information.

- **Frequent INSERTs.** If new rows are frequently inserted into the table, consider a row-oriented model. Column-oriented tables are not optimized for write operations, as column values for a row must be written to different places on disk.

- **Number of columns requested in queries.** If you typically request all or the majority of columns in the SELECT list or WHERE clause of your queries, consider a row-oriented model. Column-oriented tables are best suited to queries that aggregate many values of a single column where the WHERE or HAVING predicate is also on the aggregate column,

  ```sql
  SELECT SUM(salary)...
  SELECT AVG(salary)... WHERE salary > 10000
  or where the WHERE predicate is on a single column and is highly selective (returns a relatively small number of rows).
  SELECT salary, dept ... WHERE state='CA'
  ```
• **Number of columns in the table.** Row-oriented storage is more efficient when many columns are required at the same time, or when the row-size of a table is relatively small. Column-oriented tables can offer better query performance on wide tables (lots of columns) where you typically only access a small subset of columns in your queries.

• **Compression.** Since column data is of the same data type, there are some storage size optimizations available in column-oriented data that are not available in row-oriented data. For example, many compression schemes make use of the similarity of adjacent data to compress. However, the greater adjacent compression achieved, the more difficult random access may become, as data might need to be uncompressed to be read.

**To create a column-oriented table**

The `WITH` clause of the `CREATE TABLE` command is used to declare the storage options of the table. If not declared, the table will be created as a row-oriented heap table. Tables that use column-oriented storage must also be append-only tables. For example, to create a column-oriented table:

```sql
=> CREATE TABLE bar (a int, b text)
    WITH (appendonly=true, orientation=column)
    DISTRIBUTED BY (a);
```

**Using Compression (Append-Only Tables Only)**

Tables that utilize the append-only storage model also have the option of using in-database compression (with zlib or QuickLZ) to save disk space. Using in-database compression requires that your segment systems have the available CPU power to compress and uncompress the data. Compressed append-only tables should not be used on file systems that also are using compression. If the file system where your segment data directory resides is a compressed file system, your append-only table should *not* use compression.

When choosing a compression type and level for append-only tables, consider these factors:

• CPU usage
• Compression ratio/disk size
• Speed of compression
• Speed of decompression/scan rate

Though minimizing disk size may be the main goal in compressing tables, the time and CPU capacity required to compress and scan data is also important to consider. Every system has an optimal range of settings where compression most efficiently reduces data size without causing excessively long compression times or slow scan rates.

QuickLZ compression generally uses less CPU capacity and compresses data faster at a lower compression ratio than zlib. Conversely, zlib provides higher compression ratios at lower speeds. At compression level 1 (`compresslevel=1`), QuickLZ and zlib may yield comparable compression ratios (though at different speeds). However,
using zlib compression at a higher level of 6 might dramatically increase its advantage over QuickLZ in compression ratio (though consequently lowering the speed of compression).

Performance with compressed append-only tables depends on hardware, query tuning settings, and other factors. Greenplum recommends performing comparison testing to determine the actual performance in your environment.

To create a compressed table

The WITH clause of the CREATE TABLE command is used to declare the storage options of the table. Tables that use compression must also be append-only tables. For example, to create an append-only table with zlib compression at a compression level of 5:

```sql
=> CREATE TABLE foo (a int, b text)
    WITH (appendonly=true, comprresstype=zlib,
          compresslevel=5);
```

**Note:** QuickLZ compression level can only be set to level 1; no other options are available. Compression level with zlib can be set at any value from 1 - 9.

Checking the Compression and Distribution of an Append-Only Table

Greenplum provides built-in functions to check the compression ratio and the distribution of an append-only table. Both functions take either the object ID or name of a table. The table name may be qualified with a schema name.

<table>
<thead>
<tr>
<th>Function</th>
<th>Return Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>get_ao_distribution(oid, name)</td>
<td>Set of (dbid, tuplecount)</td>
<td>Shows the distribution of rows of an append-only table across the array.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Returns a set of rows, each of which includes a segment dbid and the number of tuples stored on the segment.</td>
</tr>
<tr>
<td>get_ao_compression_ratio(oid, name)</td>
<td>float8</td>
<td>Calculates the compression ratio for a compressed append-only table.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>If information is not available, this function returns a value of -1.</td>
</tr>
</tbody>
</table>

The compression ratio is returned as a common ratio. For example, a returned value of 3.19, or 3.19:1, means that the uncompressed table is slightly larger than three times the size of the compressed table.

The distribution of the table is returned as a set of rows indicating how many tuples are stored on each segment. For example, in a system with four primary segments with dbid values ranging from 0 - 3, the function returns four rows similar to the following:

```sql
=# SELECT get_ao_distribution('lineitem_comp');
get_ao_distribution
```
Altering a Table

The `ALTER TABLE` command is used to change the definition of an existing table. With `ALTER TABLE`, you can change various table attributes such as column definitions, distribution policy, storage model, and partition structure (see also “Maintaining Partitioned Tables” on page 83). For example, to add a not-null constraint to a table column:

```sql
=> ALTER TABLE address ALTER COLUMN street SET NOT NULL;
```

**Altering Table Distribution**

`ALTER TABLE` provides options to change the distribution policy of a table. When the distribution options of a table change, the table data is redistributed on disk, which can be resource intensive. There is also an option to redistribute table data using the existing distribution policy.

**Changing the Distribution Policy**

You can use the `ALTER TABLE` command to change the distribution policy for a table. For partitioned tables, changes to the distribution policy recursively apply to the child partitions. This operation preserves the ownership and all other attributes of the table. For example, the following command redistributes the table `sales` across all segments using the `customer_id` column as the distribution key:

```sql
ALTER TABLE sales SET DISTRIBUTED BY (customer_id);
```

When you change the hash distribution of a table, table data is automatically redistributed. However, changing the distribution policy to a random distribution will not cause the data to be redistributed. For example:

```sql
ALTER TABLE sales SET DISTRIBUTED RANDOMLY;
```

**Redistributing Table Data**

To redistribute table data for tables with a random distribution policy (or when the hash distribution policy has not changed) use `REORGANIZE=TRUE`. This sometimes may be necessary to correct a data skew problem, or when new segment resources have been added to the system. For example:

```sql
ALTER TABLE sales SET WITH (REORGANIZE=TRUE);
```

This command rebalances a table evenly across all segments using the current distribution policy (including random distribution).
Altering the Table Storage Model

It is not possible to alter the storage model of a table that has already been created. Storage, compression, and orientation can only be declared at `CREATE TABLE` time. If you have an existing table for which you want to change the storage model, you must recreate the table with the correct storage options, reload the data into the newly created table, drop the original table and rename the new table to the original name. You must also re-grant any table permissions. For example:

```sql
CREATE TABLE sales2 (LIKE sales)
WITH (appendonly=true, compressstype=quicklz,
compresslevel=1, orientation=column);

INSERT INTO sales2 SELECT * FROM sales;
DROP TABLE sales;
ALTER TABLE sales2 RENAME TO sales;
GRANT ALL PRIVILEGES ON sales TO admin;
GRANT SELECT ON sales TO guest;
```

See also, “Exchanging a Partition” on page 86 for instructions on changing the storage model of a partitioned table.

Dropping a Table

The `DROP TABLE` command removes tables from the database. For example:

```sql
DROP TABLE mytable;
```

To empty a table of rows without removing the table definition, use `DELETE` or `TRUNCATE`. For example:

```sql
DELETE FROM mytable;
TRUNCATE mytable;
```

`DROP TABLE` always removes any indexes, rules, triggers, and constraints that exist for the target table. However, to drop a table that is referenced by a view, `CASCADE` must be specified. `CASCADE` will remove a dependent view entirely.

Partitioning Large Tables

Table partitioning addresses the problem of supporting very large tables, such as fact tables, by allowing you to divide them into smaller and more manageable pieces. Partitioned tables can improve query performance by allowing the Greenplum Database query planner to scan only the relevant data needed to satisfy a given query rather than scanning the entire contents of a large table. Partitioned tables can also be used to facilitate database maintenance tasks, such as rolling old data out of the data warehouse.
Understanding Table Partitioning in Greenplum Database

Tables are partitioned at `CREATE TABLE` time using the `PARTITION BY` (and optionally the `SUBPARTITION BY`) clause. When you partition a table in Greenplum Database, you are actually creating a top-level (or parent) table with one or more levels of sub-tables (or child tables). Internally, Greenplum Database creates an inheritance relationship between the top-level table and its underlying partitions (similar to the functionality of the `INHERITS` clause of PostgreSQL).

Using the partition criteria defined during table creation, each partition is created with a distinct `CHECK` constraint, which limits the data that table can contain. The `CHECK` constraints are also used by the query planner to determine which table partitions to scan in order to satisfy a given query predicate.

Partition hierarchy information is stored in the Greenplum system catalog so that rows inserted into the top-level parent table appropriately propagate to the child table partitions. Any changes to the partition design or table structure must be done through the parent table using the `PARTITION` clauses of the `ALTER TABLE` command.

Greenplum Database supports both range partitioning (division of data based on a numerical range, such as date or price) or list partitioning (division of data based on a list of values, such as sales territory or product line), or a combination of both types.

![Example Multi-level Partition Design](image)

Partitioned tables are also distributed across Greenplum Database segments as is any non-partitioned table. Table distribution in Greenplum Database physically divides a table across the Greenplum segments to enable parallel query processing. Table partitioning is a tool to logically divide big tables to improve query performance and facilitate data warehouse maintenance tasks. Partitioning does not change the physical distribution of the table data across the segments.
Deciding on a Table Partitioning Strategy

Not all tables are good candidates for partitioning. If the answer is yes to all or most of the following questions, then table partitioning is a viable database design strategy for improving query performance. If the answer is no to most of the following questions, then table partitioning is not the right solution for that table:

- **Is the table large enough?** Large fact tables are good candidates for table partitioning. If you have millions or billions of records in a table, you will see performance benefits from logically breaking that data up into smaller chunks. For smaller tables with only a few thousand rows or less, the administrative overhead of maintaining the partitions will outweigh any performance benefits you might see.

- **Are you experiencing unsatisfactory performance?** As with any performance tuning initiative, a table should be partitioned only if queries against that table are producing slower response times than desired.

- **Do your query predicates have identifiable access patterns?** Examine the WHERE clauses of your query workload and look for table columns that are consistently used to access data. For example, if most of your queries tend to look up records by date, then a monthly or weekly date-partitioning design might be beneficial. Or if you tend to access records by region, consider a list-partitioning design to divide the table by region.

- **Does your data warehouse maintain a window of historical data?** Another consideration for partition design is your organization’s business requirements for maintaining historical data. For example, your data warehouse may only require you to keep the past twelve months worth of data. If the data is partitioned by month, you can easily drop the oldest monthly partition from the warehouse, and load current data into the most recent monthly partition.

- **Can the data be divided into somewhat equal parts based on some defining criteria?** You should choose partitioning criteria that will divide your data as evenly as possible. If the partitions contain a relatively equal number of records, query performance improves based on the number of partitions created. For example, by dividing a large table into 10 partitions, a query will execute 10 times faster than it would against the unpartitioned table (provided that the partitions are designed to support the query’s criteria).

Creating Partitioned Tables

A table can only be partitioned at creation time using the `CREATE TABLE` command.

The first step in partitioning a table is to decide on the partition design (date range, numeric range, or list of values) and choose the column(s) on which to partition the table. Decide how many levels of partitions you want. For example, you may want to date range partition a table by month and then further subpartition the monthly partitions by sales region. This section shows examples of SQL syntax for creating a table with various partition designs.

- **Defining Date Range Table Partitions**
- **Defining Numeric Range Table Partitions**
- **Defining List Table Partitions**
• Defining Multi-level Partitions
• Partitioning an Existing Table

Defining Date Range Table Partitions
A date range partitioned table uses a single date or timestamp column as the partition key column. You can use the same partition key column to further subpartition a table if necessary (for example, to partition by month and then subpartition by day). When date partitioning a table, consider partitioning by the most granular level you are interested in. For example, partition by day and have 365 daily partitions, rather than partition by year then subpartition by month then subpartition by day. A multi-level design can reduce query planning time, but a flat partition design will execute faster at query run time.

You can have Greenplum Database automatically generate partitions by giving a START value, an END value, and an EVERY clause that defines the partition increment value. By default, START values are always inclusive and END values are always exclusive. For example:

```sql
CREATE TABLE sales (id int, date date, amt decimal(10,2))
DISTRIBUTED BY (id)
PARTITION BY RANGE (date)
( START (date '2008-01-01') INCLUSIVE
  END (date '2009-01-01') EXCLUSIVE
  EVERY (INTERVAL '1 day') );
```

You can also declare and name each partition individually. For example:

```sql
CREATE TABLE sales (id int, date date, amt decimal(10,2))
DISTRIBUTED BY (id)
PARTITION BY RANGE (date)
( PARTITION Jan08 START (date '2008-01-01') INCLUSIVE,
  PARTITION Feb08 START (date '2008-02-01') INCLUSIVE,
  PARTITION Mar08 START (date '2008-03-01') INCLUSIVE,
  PARTITION Apr08 START (date '2008-04-01') INCLUSIVE,
  PARTITION May08 START (date '2008-05-01') INCLUSIVE,
  PARTITION Jun08 START (date '2008-06-01') INCLUSIVE,
  PARTITION Jul08 START (date '2008-07-01') INCLUSIVE,
  PARTITION Aug08 START (date '2008-08-01') INCLUSIVE,
  PARTITION Sep08 START (date '2008-09-01') INCLUSIVE,
  PARTITION Oct08 START (date '2008-10-01') INCLUSIVE,
  PARTITION Nov08 START (date '2008-11-01') INCLUSIVE,
  PARTITION Dec08 START (date '2008-12-01') INCLUSIVE
  END (date '2009-01-01') EXCLUSIVE );
```

Note that you do not need to declare an END value for each partition, only the last one. In this example, Jan08 would end where Feb08 starts.
Defining Numeric Range Table Partitions

A numeric range partitioned table uses a single numeric data type column as the partition key column. For example:

```sql
CREATE TABLE rank (id int, rank int, year int, gender char(1), count int)
DISTRIBUTED BY (id)
PARTITION BY RANGE (year)
  DEFAULT PARTITION extra );
```

For more information on default partitions, see “Adding a Default Partition” on page 85.

Defining List Table Partitions

A list partitioned table can use any data type column that allows equality comparisons as its partition key column. A list partition can also have a multi-column (composite) partition key, whereas a range partition only allows a single column as the partition key. For list partitions, you must declare a partition specification for every partition (list value) you want to create. For example:

```sql
CREATE TABLE rank (id int, rank int, year int, gender char(1), count int )
DISTRIBUTED BY (id)
PARTITION BY LIST (gender)
( PARTITION girls VALUES ('F'),
  PARTITION boys VALUES ('M'),
  DEFAULT PARTITION other );
```

For more information on default partitions, see “Adding a Default Partition” on page 85.

Defining Multi-level Partitions

It is possible to create a multi-level partition design where you have subpartitions of partitions. Using a subpartition template ensures that every partition has the same subpartition design, even partitions that are added later. For example, to create the two-level partition design illustrated in Figure 9.1, “Example Multi-level Partition Design” on page 77:

```sql
CREATE TABLE sales (trans_id int, date date, amount decimal(9,2), region text)
DISTRIBUTED BY (trans_id)
PARTITION BY RANGE (date)
SUBPARTITION BY LIST (region)
SUBPARTITION TEMPLATE
( SUBPARTITION usa VALUES ('usa'),
  SUBPARTITION asia VALUES ('asia'),
  SUBPARTITION europe VALUES ('europe'),
  DEFAULT SUBPARTITION other_regions)
( START (date '2008-01-01') INCLUSIVE
```
Below is a similar example illustrating a three-level partition design where the sales table is partitioned by year, then month, then region. The SUBPARTITION TEMPLATE clauses ensure that each yearly partition has the same subpartition structure. Also note that a DEFAULT partition is declared at each level of the hierarchy:

```sql
CREATE TABLE sales (id int, year int, month int, day int, region text)
DISTRIBUTED BY (id)
PARTITION BY RANGE (year)
    SUBPARTITION BY RANGE (month)
        SUBPARTITION TEMPLATE (
            START (1) END (13) EVERY (1),
            DEFAULT SUBPARTITION other_months
        )
    SUBPARTITION BY LIST (region)
        SUBPARTITION usa VALUES ('usa'),
        SUBPARTITION europe VALUES ('europe'),
        SUBPARTITION asia VALUES ('asia'),
        DEFAULT SUBPARTITION other_regions
    DEFAULT PARTITION outlying_years );
```

**Partitioning an Existing Table**

It is not possible to partition a table that has already been created. Tables can only be partitioned at CREATE TABLE time. If you have an existing table that you want to partition, you must recreate the table as a partitioned table, reload the data into the newly partitioned table, drop the original table and rename the partitioned table to the original name. You must also regrant any table permissions. For example:

```sql
CREATE TABLE sales2 (LIKE sales)
PARTITION BY RANGE (date)
( START (date '2008-01-01') INCLUSIVE
    END (date '2009-01-01') EXCLUSIVE
    EVERY (INTERVAL '1 month') );

INSERT INTO sales2 SELECT * FROM sales;
DROP TABLE sales;
ALTER TABLE sales2 RENAME TO sales;
GRANT ALL PRIVILEGES ON sales TO admin;
GRANT SELECT ON sales TO guest;
```
Limitations of Partitioned Tables

- To be able to enforce a primary key or unique constraint, the primary or unique key columns must start with the partitioning key column.

Loading Partitioned Tables

Once you have created your partitioned table structure, top-level parent tables are always empty. Data is routed to the bottom-level child table partitions only. In a multi-level partition design, only the subpartitions at the bottom of the hierarchy can contain data.

If a row cannot be mapped to a child table partition, it will be rejected and the load will fail. If you do not want unmapped rows to be rejected at load time, you can define your partition hierarchy with a DEFAULT partition. Any rows that do not match to an existing partition’s CHECK constraints will then load into the DEFAULT partition. See “Adding a Default Partition” on page 85.

At runtime, the query planner scans the entire table inheritance hierarchy and uses the CHECK table constraints to determine which of the child table partitions to scan in order to satisfy the query’s conditions. The DEFAULT partition (if your hierarchy has one) is always scanned. If the DEFAULT partition contains data, this will slow down the overall scan time.

When you use COPY or INSERT to load data into a parent table, it automatically gets rerouted to the correct partition by default. Therefore, you can load a partitioned table as you would a regular table.

You can also load data into the child table partitions directly if needed. You can also create an intermediate staging table, load it, and then exchange it into your partition design. See “Exchanging a Partition” on page 86.

Verifying Your Partition Strategy

The purpose for partitioning a table is to reduce the number of rows that must be scanned in order to satisfy a given query. If a table is partitioned based on the query predicate, you can verify that the query planner is selectively scanning the relevant data by using EXPLAIN to look at the query plan.

For example, suppose we have a sales table that is date-range partitioned by month and subpartitioned by region as illustrated in Figure 9.1, “Example Multi-level Partition Design” on page 77. For the following query:

```
EXPLAIN SELECT * FROM sales WHERE date='01-07-08' AND region='usa';
```

The query plan for this query should show a table scan of the following tables only:

- the default partition returning 0-1 rows (if your partition design has one)
- the January 2008 partition (sales_1_prt_1) returning 0-1 rows
- the USA region subpartition (sales_1_2_prt_usa) returning some number of rows.

Below is an example of the relevant portion of the query plan:

```
-> Seq Scan on sales_1_prt_1 sales (cost=0.00..0.00 rows=0
```
Make sure that the query planner is not scanning unnecessary partitions or subpartitions (for example, scans of other months or regions not specified in the query predicate), and that scans of the top-level tables are returning 0-1 rows.

**Some Limitations of Selective Partition Scanning**

If the query plan shows that your partition hierarchy is not being selectively scanned, it may be due to one of the following limitations:

- The query planner is only able to selectively scan partitioned tables when the query contains a direct and simple restriction of the table using immutable operators such as: 
  
  `=` \ `\lt` \ `\le` \ `\gt` \ `\ge` \ `\neq`  

- Selective scanning does not currently recognize VOLATILE functions within a query (STABLE and IMMUTABLE functions are recognized). For example, WHERE clauses such as `date > CURRENT_DATE` will cause the query planner to selectively scan partitioned tables, but `time > TIMEMOFDAY` will not.

- Selective scanning does not currently support the runtime evaluation of table data as a basis for determining which partitions to eliminate. For example, a query with a WHERE predicate such as the following would not cause selective scanning of partitions because the evaluation criteria is based on the runtime evaluation of the value in `some_column`:

  ```sql
  SELECT * from partition_table PT, other_table OT WHERE PT.id=OT.id and OT.some_column = 'value';
  ```

**Viewing Your Partition Design**

You can look up information about your partition design using the `pg_partitions` view. For example to see the partition design of the `sales` table:

```sql
SELECT partitionboundary, partitiontablename, partitionname, partitionlevel, partitionrank FROM pg_partitions WHERE tablename='sales';
```

There are also the following views that show information about partitioned tables:

- `pg_partition_templates` - Shows subpartitions that were created using a subpartition template.
- `pg_partition_columns` - Shows the partition key columns used in a partition design.

**Maintaining Partitioned Tables**

You must maintain a partitioned table using the `ALTER TABLE` command against the top-level parent table. The most common scenario is dropping old partitions and adding new ones in order to maintain a rolling window of data in a range partition design. You may also want to convert (exchange) older partitions to the append-only
compressed storage format in order to save space. If you have a default partition in your partition design, the procedure for adding a new partition is to *split* the default partition.

- Adding a New Partition
- Renaming a Partition
- Adding a Default Partition
- Dropping a Partition
- Truncating a Partition
- Exchanging a Partition
- Splitting a Partition
- Modifying a Subpartition Template

**Note:** When defining and altering partition designs, use the given *partition name*, not the table object name. Although you can query and load any table (including partitioned tables) directly using SQL commands, you can only modify the structure of a partitioned table using the `ALTER TABLE...PARTITION` clauses.

### Adding a New Partition

You can add a new partition to an existing partition design using the `ALTER TABLE` command. If the original partition design included subpartitions defined by a *subpartition template*, then the newly added partition will also be subpartitioned according to that template. For example:

```sql
ALTER TABLE sales ADD PARTITION
    START (date '2009-02-01') INCLUSIVE
    END (date '2009-03-01') EXCLUSIVE;
```

If a subpartition template was not used when you created the table, you would then define subpartitions when adding a new partition:

```sql
ALTER TABLE sales ADD PARTITION
    START (date '2009-02-01') INCLUSIVE
    END (date '2009-03-01') EXCLUSIVE
    ( SUBPARTITION usa VALUES ('usa'),
      SUBPARTITION asia VALUES ('asia'),
      SUBPARTITION europe VALUES ('europe') );
```

If you want to add a new subpartition to an existing partition, you can specify a particular partition to alter. For example:

```sql
ALTER TABLE sales ALTER PARTITION FOR (RANK(12))
ADD PARTITION africa VALUES ('africa');
```

**Note:** You cannot add a new partition to a partition design that has a default partition. You must split the default partition in order to add a new partition. See “Splitting a Partition” on page 86.
Renaming a Partition

Partitioned tables are created using the naming convention:

<parentname>_<level>_prt_<partition_name>

For example:

sales_1_prt_jan08

Or for auto-generated range partitions (a number is assigned when no name is given):

sales_1_prt_1

It is not possible to rename a partitioned child table directly by altering the table name. However, you can rename the top-level parent table, and the associated <parentname> will change in the table names of all associated child table partitions. For example:

ALTER TABLE sales RENAME TO globalsales;

Would change the associated table names accordingly:

globalsales_1_prt_1

You can also change the partition name of a partition to make it easier to identify. For example:

ALTER TABLE sales RENAME PARTITION FOR ('2008-01-01') TO jan08;

Would change the associated table name accordingly:

sales_1_prt_jan08

When altering partitioned tables with the ALTER TABLE command, they are always referred to by their partition name (jan08) and not their full table name (sales_1_prt_jan08).

Adding a Default Partition

You can add a default partition to an existing partition design using the ALTER TABLE command.

ALTER TABLE sales ADD DEFAULT PARTITION other;

If your partition design is multi-leveled, then each level in the hierarchy needs a default partition. For example:

ALTER TABLE sales ALTER PARTITION FOR (RANK(1)) ADD DEFAULT PARTITION other;
ALTER TABLE sales ALTER PARTITION FOR (RANK(2)) ADD DEFAULT PARTITION other;
ALTER TABLE sales ALTER PARTITION FOR (RANK(3)) ADD DEFAULT PARTITION other;

Partition designs that do not have a default partition will reject incoming rows that do not match to an existing partition's CHECK constraint. If a partitioned table has a default partition, incoming data that does not match to an existing partition is instead inserted into the default partition.
Dropping a Partition
You can drop a partition from your partition design using the `ALTER TABLE` command. When you drop a partition that has subpartitions, the subpartitions (and all data in them) are automatically dropped as well. For range partitions, it is common to drop the older partitions from the range as old data is rolled out of the data warehouse. For example:

```
ALTER TABLE sales DROP PARTITION FOR (RANK(1));
```

Truncating a Partition
You can truncate a partition using the `ALTER TABLE` command. When you truncate a partition that has subpartitions, the subpartitions are automatically truncated as well.

```
ALTER TABLE sales TRUNCATE PARTITION FOR (RANK(1));
```

Exchanging a Partition
Exchanging a partition involves swapping in another table in place of an existing partition. You can exchange a partition using the `ALTER TABLE` command. You can only exchange partitions at the lowest level of your partition hierarchy (only partitions that contain data can be exchanged).

This can be useful for data loading. For example, you could load a staging table and then swap the loaded table into your partition design. You can also use exchange to change the storage type of older partitions to append-only tables. For example:

```
CREATE TABLE jan08 (LIKE sales) WITH (appendonly=true);
INSERT INTO jan08 SELECT * FROM sales_1_prt_1;
ALTER TABLE sales EXCHANGE PARTITION FOR ('2008-01-01') WITH TABLE jan08;
```

Splitting a Partition
Splitting a partition involves dividing an existing partition into two. You can split a partition using the `ALTER TABLE` command. You can only split partitions at the lowest level of your partition hierarchy (only partitions that contain data can be split). The split value you specify will go into the `latter` partition.

For example, to split a monthly partition into two with the first partition containing dates January 1-15 and the second partition containing dates January 16-31:

```
ALTER TABLE sales SPLIT PARTITION FOR ('2008-01-01')
AT ('2008-01-16')
INTO (PARTITION jan081to15, PARTITION jan0816to31);
```

If your partition design has a default partition, you must split the default partition in order to add a new partition. You can only split default partitions at the lowest level of your partition hierarchy (only default partitions that contain data can be split).

When using the `INTO` clause, the second partition name specified should always be that of the existing default partition. For example, to split a default range partition to add a new monthly partition for January 2009:

```
ALTER TABLE sales SPLIT DEFAULT PARTITION
START ('2009-01-01') INCLUSIVE
```
Creating and Using Sequences

Sequences are often used to auto-increment unique ID columns of a table whenever a new record is added.

Creating a Sequence

The `CREATE SEQUENCE` command creates and initializes a new special single-row sequence generator table with the given sequence name. The sequence name must be distinct from the name of any other sequence, table, index, or view in the same schema. For example:

```
CREATE SEQUENCE myserial START 101;
```

Using a Sequence

Once you have created a sequence generator table using `CREATE SEQUENCE`, you can use the `nextval` function to operate on the sequence. For example, to insert a row into a table that gets the next value of a sequence:
INSERT INTO vendors VALUES (nextval('myserial'), 'acme');

You can also use the \texttt{setval} function to operate on a sequence to reset a sequence’s counter value. For example:

\begin{verbatim}
SELECT setval('myserial', 201);
\end{verbatim}

A \texttt{nextval} operation is never rolled back. Once a value has been fetched it is considered used, even if the transaction that did the \texttt{nextval} later aborts. This means that aborted transactions may leave unused holes in the sequence of assigned values. \texttt{setval} operations are never rolled back, either.

Note that the \texttt{nextval} function is currently not allowed in \texttt{UPDATE} or \texttt{DELETE} statements if mirroring is enabled, and the \texttt{currval} and \texttt{lastval} functions are currently not supported in Greenplum Database.

To then examine the current settings of a sequence you can simply query the sequence table directly:

\begin{verbatim}
SELECT * FROM myserial;
\end{verbatim}

### Altering a Sequence

The \texttt{ALTER SEQUENCE} command changes the parameters of an existing sequence generator. For example:

\begin{verbatim}
ALTER SEQUENCE myserial RESTART WITH 105;
\end{verbatim}

Any parameters not specifically set in the \texttt{ALTER SEQUENCE} command retain their prior settings.

### Dropping a Sequence

The \texttt{DROP SEQUENCE} command removes a sequence generator table. For example:

\begin{verbatim}
DROP SEQUENCE myserial;
\end{verbatim}

### Using Indexes in Greenplum Database

In most traditional databases, indexes can greatly improve data access times. However, in a distributed database such as Greenplum, indexes should be used more sparingly. Greenplum Database is very fast at sequential scanning (indexes use a random seek pattern to locate records on disk). Also, unlike a traditional database, the data is distributed across the segments. This means each segment scans a smaller portion of the overall data in order to get the result. If using table partitioning, the total data to scan may be even a fraction of that. Typically, a business intelligence (BI) query workload returns very large data sets, and thus does not make efficient use of indexes.

Greenplum recommends that you first try your query workload without adding any additional indexes. Indexes are more likely to improve performance for OLTP type workloads, where the query is returning a single record or a small subset of data. Indexes can also improve performance on compressed append-only tables for queries.
that return a targeted set of rows, as the optimizer has the option to use an index access method rather than a full table scan when appropriate. For compressed data, an index access method means only the necessary rows are uncompressed.

Note that Greenplum Database will automatically create **PRIMARY KEY** indexes for tables with primary keys. If you want to create an index on a partitioned table, you must index each partitioned child table directly. Indexes on the parent table are not passed down to child table partitions.

Indexes do add some database overhead — they take up storage space and have to be maintained whenever the table is updated. Make sure that the indexes you create are actually being used by your query workload. Also, check to see that the indexes you add are indeed improving query performance (as compared to a sequential scan of the table). You can look at the **EXPLAIN** plans for a query to determine if indexes are being used. See “Query Profiling” on page 119.

Some other general considerations when creating indexes are:

- **Your Query Workload.** Indexes are more likely to improve performance for OLTP type workloads, where the query is returning a single record or a very small data set. Typically, a business intelligence (BI) query workload returns very large data sets, and thus does not make efficient use of indexes. For this type of workload, it is better to use sequential scans to locate large chunks of data on disk rather than to randomly seek the disk using index scans.

- **Compressed Tables.** Indexes can improve performance on compressed append-only tables for queries that return a targeted set of rows, as the optimizer has the option to use an index access method rather than a full table scan when appropriate. For compressed data, an index access method means only the necessary rows are uncompressed.

- **Avoid indexes on frequently updated columns.** Creating an index on a column that is frequently updated increases the amount of writes required when the column is updated.

- **Create selective B-tree indexes.** Index selectivity is a ratio of the number of distinct values a column has divided by the number of rows in a table. For example, if a table has 1000 rows and a column has 800 distinct values, the selectivity of the index is 0.8, which is considered good. Unique indexes always have a selectivity ratio of 1.0, which is the best possible. Note that unique indexes are only allowed on distribution key columns in Greenplum Database.

- **Use Bitmap indexes for low selectivity columns.** Greenplum Database has an additional index type called a Bitmap index, which is not available in regular PostgreSQL. See “About Bitmap Indexes” on page 90 for details.

- **Index columns used in joins.** An index on a column used for frequent joins (such as a foreign key column) may improve join performance, as it can enable additional join methods for the query planner to use.

- **Index columns frequently used in predicates.** For queries on large tables, examine the WHERE predicates for the columns that are referenced most often. These may be good candidates for indexes.

- **Avoid overlapping indexes.** Overlapping indexes (those that have the same leading column) are redundant and unnecessary.
• **Drop indexes for bulk loads.** For mass loads of data into a table, consider dropping the indexes and re-creating them after the load is complete. This is often faster than updating the indexes.

• **Consider a clustered index.** Clustering an index means that the records are physically ordered on disk according to the index. If the records you need are distributed randomly on disk, then the database has to seek across the disk to get the records requested. If those records are stored more closely together, then the fetching from disk is more sequential. A good example for a clustered index is on a date column where the data is ordered sequentially by date. A query against a specific date range will result in an ordered fetch from the disk, which leverages fast sequential access.

**To cluster an index in Greenplum Database**

For very large tables, using the `CLUSTER` command to physically reorder a table based on an index can take an extremely long time. To achieve the same results much faster, you can manually reorder the data on disk by creating an intermediate table and loading the data in the desired order. For example:

```
CREATE TABLE new_table (LIKE old_table)
    AS SELECT * FROM old_table ORDER BY myixcolumn;
DROP old_table;
ALTER TABLE new_table RENAME TO old_table;
CREATE INDEX myixcolumn_ix ON old_table;
VACUUM ANALYZE old_table;
```

**Index Types**

Greenplum Database provides the following index types as does PostgreSQL: B-tree and GiST (Hash and GIN indexes are disabled in Greenplum Database). Each index type uses a different algorithm that is best suited to different types of queries. By default, the `CREATE INDEX` command will create a B-tree index, which fits the most common situations. See Index Types in the PostgreSQL documentation for a description of these types.

**Note:** Greenplum Database has some special considerations concerning unique indexes on a table. Unique indexes are allowed only if the columns of the index key are the same as (or a superset of) the Greenplum distribution key. Unique indexes are not supported on append-only tables. On partitioned tables, a unique index can not be enforced across all child table partitions of a partitioned table - a unique index is only supported within an individual partition.

**About Bitmap Indexes**

In addition to the index types provided by PostgreSQL, Greenplum Database has an additional Bitmap index type. Bitmap indexes are one of the most promising strategies for indexing high dimensional data in data warehousing applications and decision support systems. These types of applications typically have large amounts of data and ad hoc queries, but a low number of DML transactions.

An index provides pointers to the rows in a table that contain a given key value. A regular index stores a list of tuple ids for each key corresponding to the rows with that key value. In a bitmap index, a bitmap for each key value replaces a list of tuple ids.
Fully indexing a large table with a traditional B-tree index can be expensive in terms of space because the indexes can be several times larger than the data in the table. Bitmap indexes are typically only a fraction of the size of the indexed data in the table.

Each bit in the bitmap corresponds to a possible tuple id, and if the bit is set, it means that the row with the corresponding tuple id contains the key value. A mapping function converts the bit position to an actual tuple id, so that the bitmap index provides the same functionality as a regular index. Bitmap indexes store the bitmaps in a compressed way. If the number of distinct key values is small, bitmap indexes compress better and the space saving benefit compared to a B-tree index becomes even better.

Bitmap indexes are most effective for queries that contain multiple conditions in the **WHERE** clause. Rows that satisfy some, but not all, conditions are filtered out before the table itself is accessed. This improves response time, often dramatically.

### When to Use Bitmap Indexes

Bitmap indexes perform best for columns that have between 100 and 100,000 distinct values. Columns with fewer than 100 distinct values usually do not benefit much from any type of index. For example, a gender column with only two distinct values for male and female would not be a good candidate. On a column with more than 100,000 distinct values, the performance and space efficiency of a bitmap index decline. The size of a bitmap index is proportional to the number of rows in the table times the number of distinct values in the indexed column.

When used on columns that have the right characteristics, a bitmap index can outperform a B-tree index, particularly when the indexed column is often queried in conjunction with other indexed columns.

Bitmap indexes can dramatically improve query performance for ad hoc queries. **AND** and **OR** conditions in the **WHERE** clause of a query can be resolved quickly by performing the corresponding Boolean operations directly on the bitmaps before converting the resulting bitmap to tuple ids. If the resulting number of rows is small, the query can be answered quickly without resorting to a full table scan.

### When Not to Use Bitmap Indexes

Bitmap indexes should not be used for unique columns or columns with high cardinality data, such as customer names or phone numbers. The performance gains and disk space advantages of bitmap indexes start to diminish on columns with 100,000 or more unique values, regardless of the number of rows in the table.

Bitmap indexes are primarily intended for data warehousing applications where users query the data rather than update it. They are not suitable for OLTP applications with large numbers of concurrent transactions modifying the data.

As with B-tree indexes, bitmap indexes should be used sparingly. Testing should demonstrate improved query performance after adding the bitmap index (as opposed to doing a sequential scan of the table).
Creating an Index

The `CREATE INDEX` command defines a new index on a table. By default, a B-tree index is created if you do not declare an index type. For example, to create a B-tree index on the column `title` in the table `films`:

```
CREATE INDEX title_idx ON films (title);
```

To create a bitmap index on the column `gender` in the table `employee`:

```
CREATE INDEX gender_bmp_idx ON employee USING bitmap (gender);
```

Examining Index Usage

Although indexes in Greenplum Database do not need maintenance and tuning, it is still important to check which indexes are actually used by the real-life query workload. Examining index usage for an individual query is done with the `EXPLAIN` command.

The query plan shows the different steps or `plan nodes` that the database will take to answer a particular query, along with time estimates for each plan node. To examine the use of indexes, look for the following query plan node types in your `EXPLAIN` output:

- **Index Scan** - A scan of an index.
- **Bitmap Heap Scan** - Retrieves all rows from the bitmap generated by `BitmapAnd`, `BitmapOr`, or `BitmapIndexScan` and accesses the heap to retrieve the relevant rows.
- **Bitmap Index Scan** - Compute a bitmap by OR-ing all bitmaps that satisfy the query predicates from the underlying index.
- **BitmapAnd** or **BitmapOr** - Takes the bitmaps generated from multiple `BitmapIndexScan` nodes, ANDs or ORs them together, and generates a new bitmap as its output.

It is difficult to formulate a general procedure for determining which indexes to set up. A good deal of experimentation will be necessary in most cases.

- **Always run `ANALYZE` after creating or updating an index.** This command collects table statistics that are used by the query planner. This information is required to guess the number of rows returned by a query, which is needed by the planner to assign realistic costs to each possible query plan.
- **Use real data for experimentation.** Using test data for setting up indexes will tell you what indexes you need for the test data, but that is all.
- **It is especially fatal to use very small test data sets.** While selecting 1000 out of 1,000,000 rows could be a candidate for an index, selecting 1 out of 100 rows will hardly be, because the 100 rows will probably fit within a single disk page, and there is no plan that can beat sequentially fetching one disk page.
- **Also be careful when making up test data, which is often unavoidable when the application is not in production use yet.** Values that are very similar, completely random, or inserted in sorted order will skew the statistics away from the distribution that real data would have.
• When indexes are not used, it can be useful for testing to force their use. There are run-time parameters that can turn off various plan types. For instance, turning off sequential scans (`enable_seqscan`) and nested-loop joins (`enable_nestloop`), which are the most basic plans, will force the system to use a different plan. Time your query with and without indexes. The `EXPLAIN ANALYZE` command can be useful here.

Managing Indexes
In certain conditions, a poorly performing index may need to be rebuilt using the `REINDEX` command. This rebuilds an index using the data stored in the index’s table, replacing the old copy of the index.

Updates or delete operations do not update bitmap indexes. Therefore, if you have deleted rows or updated columns in a table that has bitmap indexes, you will need to rebuild the indexes using the `REINDEX` command.

To rebuild all indexes on a table

```
REINDEX my_table;
```

To rebuild a particular index

```
REINDEX my_index;
```

Dropping an Index
The `DROP INDEX` command removes an index. For example:

```
DROP INDEX title_idx;
```

When loading data, it is often faster to drop all indexes, load, then recreate the indexes afterwards.

Creating and Managing Views
Views are a way to save frequently used or complex queries and then access them in a `SELECT` statement as if they were a table. A view is not physically materialized on disk as is a table. The query is instead ran as a subquery whenever the view is accessed.

Creating Views
The `CREATE VIEW` command defines a view of a query. For example:

```
CREATE VIEW comedies AS SELECT * FROM films WHERE kind = 'comedy';
```

Note that currently views ignore `ORDER BY` or `SORT` operations stored in the view.

Dropping Views
The `DROP VIEW` command removes a view. For example:

```
DROP VIEW topten;
```
Managing Data

This chapter provides information about managing data and concurrent access in Greenplum Database. It contains the following topics:

- About Concurrency Control in Greenplum Database
- Inserting New Rows
- Updating Existing Rows
- Deleting Rows
- Working With Transactions
- Vacuuming the Database

About Concurrency Control in Greenplum Database

Unlike traditional database systems which use locks for concurrency control, Greenplum Database (as does PostgreSQL) maintains data consistency by using a multiversion model (Multiversion Concurrency Control, MVCC). This means that while querying a database, each transaction sees a snapshot of data which protects the transaction from viewing inconsistent data that could be caused by (other) concurrent updates on the same data rows. This provides transaction isolation for each database session.

MVCC, by eschewing explicit locking methodologies of traditional database systems, minimizes lock contention in order to allow for reasonable performance in multiuser environments. The main advantage to using the MVCC model of concurrency control rather than locking is that in MVCC locks acquired for querying (reading) data do not conflict with locks acquired for writing data, and so reading never blocks writing and writing never blocks reading.

Greenplum Database provides various lock modes to control concurrent access to data in tables. Most Greenplum Database SQL commands automatically acquire locks of appropriate modes to ensure that referenced tables are not dropped or modified in incompatible ways while the command executes. For applications that cannot adapt easily to MVCC behavior, the `LOCK` command can be used to acquire explicit locks. However, proper use of MVCC will generally provide better performance than locks.

<table>
<thead>
<tr>
<th>Lock Mode</th>
<th>Associated SQL Commands</th>
<th>Conflicts With</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCESS SHARE</td>
<td>SELECT</td>
<td>ACCESS EXCLUSIVE</td>
</tr>
<tr>
<td>ROW SHARE</td>
<td>SELECT FOR UPDATE, SELECT FOR SHARE</td>
<td>EXCLUSIVE, ACCESS EXCLUSIVE</td>
</tr>
<tr>
<td>ROW EXCLUSIVE</td>
<td>INSERT, COPY</td>
<td>SHARE, SHARE ROW EXCLUSIVE, EXCLUSIVE, ACCESS EXCLUSIVE</td>
</tr>
</tbody>
</table>
Inserting New Rows

When a table is created, it contains no data. The first thing to do before a database can be of much use is to insert data. To create a new row, use the `INSERT` command. This command requires the table name and a value for each of the columns of the table. The data values are listed in the order in which the columns appear in the table, separated by commas. For example:

```
INSERT INTO products VALUES (1, 'Cheese', 9.99);
```

If you do not know the order of the columns in the table, you can also list the columns explicitly. Many users consider it good practice to always list the column names. For example:

```
INSERT INTO products (name, price, product_no) VALUES ('Cheese', 9.99, 1);
```

Usually, the data values will be literals (constants), but scalar expressions are also allowed. For example:

```
INSERT INTO films SELECT * FROM tmp_films WHERE date_prod < '2004-05-07';
```

You can also insert multiple rows in a single command. For example:

```
INSERT INTO products (product_no, name, price) VALUES
(1, 'Cheese', 9.99),
(2, 'Bread', 1.99),
(3, 'Milk', 2.99);
```

Table 10.1 Lock Modes in Greenplum Database

<table>
<thead>
<tr>
<th>Lock Mode</th>
<th>Associated SQL Commands</th>
<th>Conflicts With</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHARE UPDATE EXCLUSIVE</td>
<td>VACUUM (without FULL), ANALYZE</td>
<td>SHARE UPDATE EXCLUSIVE, SHARE, SHARE ROW EXCLUSIVE, EXCLUSIVE, ACCESS EXCLUSIVE</td>
</tr>
<tr>
<td>SHARE</td>
<td>CREATE INDEX</td>
<td>ROW EXCLUSIVE, SHARE UPDATE EXCLUSIVE, SHARE, SHARE ROW EXCLUSIVE, EXCLUSIVE, ACCESS EXCLUSIVE</td>
</tr>
<tr>
<td>SHARE ROW EXCLUSIVE</td>
<td></td>
<td>ROW EXCLUSIVE, SHARE UPDATE EXCLUSIVE, SHARE, SHARE ROW EXCLUSIVE, EXCLUSIVE, ACCESS EXCLUSIVE</td>
</tr>
<tr>
<td>EXCLUSIVE</td>
<td>DELETE, UPDATE¹</td>
<td>ROW SHARE, ROW EXCLUSIVE, SHARE UPDATE EXCLUSIVE, SHARE, SHARE ROW EXCLUSIVE, EXCLUSIVE, ACCESS EXCLUSIVE</td>
</tr>
<tr>
<td>ACCESS EXCLUSIVE</td>
<td>ALTER TABLE, DROP TABLE, TRUNCATE, REINDEX, CLUSTER, VACUUM FULL</td>
<td>ACCESS SHARE, ROW SHARE, ROW EXCLUSIVE, SHARE UPDATE EXCLUSIVE, SHARE, SHARE ROW EXCLUSIVE, EXCLUSIVE, ACCESS EXCLUSIVE</td>
</tr>
</tbody>
</table>

¹. UPDATE and DELETE in Greenplum Database acquire a more restrictive lock - EXCLUSIVE rather than ROW EXCLUSIVE.
When inserting a lot of data at the same time, considering using external tables (CREATE EXTERNAL TABLE) or the COPY command. These load mechanisms are more efficient than INSERT when loading a large number of rows. See “Loading and Unloading Data” on page 124 for more information about bulk data loading.

The storage model of append-only tables is optimized for bulk data loading. Single row INSERT statements are not recommended for append-only tables.

### Updating Existing Rows

The term update refers to the modification of data that is already in the database. You can update individual rows, all the rows in a table, or a subset of all rows. Each column can be updated separately without affecting the other columns.

To perform an update, you need three pieces of information:

1. The name of the table and columns to update,
2. The new values of the columns,
3. The row(s) to update by specifying the conditions a row must meet in order to be updated.

The **UPDATE** command updates rows in a table. For example, this command updates all products that have a price of 5 to have a price of 10:

```sql
UPDATE products SET price = 10 WHERE price = 5;
```

Using **UPDATE** in Greenplum Database has the following restrictions:

- The Greenplum distribution key columns may not be updated.
- If the WHERE clause joins two or more tables to perform the update, the tables must have the same Greenplum distribution key and be joined on the distribution key column(s). This is referred to as an equijoin.
- Cannot use STABLE or VOLATILE functions in an UPDATE statement if mirrors are enabled.
- The RETURNING clause is not supported in Greenplum Database.

### Deleting Rows

The **DELETE** command deletes rows that satisfy the WHERE clause from the specified table. If the WHERE clause is absent, the effect is to delete all rows in the table. The result is a valid, but empty table. For example, to remove all rows from the products table that have a price of 10:

```sql
DELETE FROM products WHERE price = 10;
```

Or to delete all rows from a table:

```sql
DELETE FROM products;
```

Using **DELETE** in Greenplum Database has the following restrictions:
• If the \texttt{WHERE} clause joins two or more tables to perform the update, the tables must have the same Greenplum distribution key and be joined on the distribution key column(s). This is referred to as an \textit{equijoin}.
• Cannot use \texttt{STABLE} or \texttt{VOLATILE} functions in a \texttt{DELETE} statement if mirrors are enabled.
• The \texttt{RETURNING} clause is not supported in Greenplum Database.

\textbf{Truncating a Table}

If you want to quickly remove all rows in a table, consider using the \texttt{TRUNCATE} command. For example:

\begin{verbatim}
TRUNCATE mytable;
\end{verbatim}

This command empties the table of all rows in one operation. Note that \texttt{TRUNCATE} does not scan the table, therefore it does not process inherited child tables or \texttt{ON DELETE} rewrite rules. Only rows in the named table will be truncated.

\textbf{Working With Transactions}

Transactions allow you to bundle together multiple SQL statements in one all-or-nothing operation.

The SQL commands used to perform transactions in Greenplum Database are:
• \texttt{BEGIN} or \texttt{START TRANSACTION} to start a transaction block.
• \texttt{END} or \texttt{COMMIT} to commit the results of the transaction.
• \texttt{ROLLBACK} to abandon the transaction without making any changes.
• \texttt{SAVEPOINT} to allow you to selectively discard parts of the transaction, while committing the rest. After defining a savepoint with \texttt{SAVEPOINT}, you can if needed roll back to the savepoint with \texttt{ROLLBACK TO SAVEPOINT}. To destroy a savepoint within a transaction, use \texttt{RELEASE SAVEPOINT}.

\textbf{Transaction Isolation Levels}

The SQL standard defines four transaction isolation levels. In Greenplum Database, you can request any of the four standard transaction isolation levels. But internally, there are only two distinct isolation levels — \textit{read committed} and \textit{serializable}. 
• **read committed** — When a transaction runs on this isolation level, a `SELECT` query sees only data committed before the query began. It never sees either uncommitted data or changes committed during query execution by concurrent transactions. However, the `SELECT` does see the effects of previous updates executed within its own transaction, even though they are not yet committed. In effect, a `SELECT` query sees a snapshot of the database as of the instant that query begins to run. Notice that two successive `SELECT` commands can see different data, even though they are within a single transaction, if other transactions commit changes during execution of the first `SELECT`. `UPDATE` and `DELETE` commands behave the same as `SELECT` in terms of searching for target rows. They will only find target rows that were committed as of the command start time. However, such a target row may have already been updated (or deleted or locked) by another concurrent transaction by the time it is found. The partial transaction isolation provided by read committed mode is adequate for many applications, and this mode is fast and simple to use. However, for applications that do complex queries and updates, it may be necessary to guarantee a more rigorously consistent view of the database than the read committed mode provides.

• **serializable** — This is the strictest transaction isolation. This level emulates serial transaction execution, as if transactions had been executed one after another, serially, rather than concurrently. Applications using this level must be prepared to retry transactions due to serialization failures. When a transaction is on the serializable level, a `SELECT` query sees only data committed before the transaction began. It never sees either uncommitted data or changes committed during transaction execution by concurrent transactions. However, the `SELECT` does see the effects of previous updates executed within its own transaction, even though they are not yet committed. Successive `SELECT` commands within a single transaction always see the same data. `UPDATE` and `DELETE` commands behave the same as `SELECT` in terms of searching for target rows. They will only find target rows that were committed as of the transaction start time. However, such a target row may have already been updated (or deleted or locked) by another concurrent transaction by the time it is found. In this case, the serializable transaction will wait for the first updating transaction to commit or roll back (if it is still in progress). If the first updater rolls back, then its effects are negated and the serializable transaction can proceed with updating the originally found row. But if the first updater commits (and actually updated or deleted the row, not just locked it) then the serializable transaction will be rolled back.

• **read uncommitted** — Treated the same as `read committed` in Greenplum Database.

• **repeatable read** — Treated the same as `serializable` in Greenplum Database.

The default transaction isolation level in Greenplum Database is `read committed`. To change the isolation level for a transaction, you can declare the isolation level when you `BEGIN` the transaction, or else use the `SET TRANSACTION` command after the transaction is started.
**Vacuuming the Database**

Because of the MVCC transaction concurrency model, data rows that are deleted or updated still occupy physical space on disk even though they are not visible to any new transactions. If you have a database with lots of updates and deletes, you will generate a lot of expired rows. Periodically running the `VACUUM` command will remove these expired rows. For example:

```
VACUUM mytable;
```

The `VACUUM` command also collects table-level statistics such as number of rows and pages, so it is necessary to vacuum all tables after loading data, including append-only tables. Recommended routine vacuum operations are described in “Routine Vacuum and Analyze” on page 210.

**Configuring the Free Space Map**

Expired rows are held in what is called the *free space map*. The free space map must be sized large enough to cover the expired rows of all tables in your database. If not sized large enough, space occupied by expired rows that overflow the free space map cannot be reclaimed by a regular `VACUUM` command.

A `VACUUM FULL` will reclaim all expired row space, but is a very expensive operation and may take an unacceptably long time to finish on large, distributed Greenplum Database tables. If you do get into a situation where the free space map has overflowed, it may be more timely to recreate the table with a `CREATE TABLE AS` statement and drop the old table. A `VACUUM FULL` is not recommended in Greenplum Database.

It is best to size the free space map appropriately. The free space map is configured with the following server configuration parameters:

```
max_fsm_pages
max_fsm_relations
```

11. Querying Data

This chapter describes the use of the SQL language in Greenplum Database. SQL commands are typically entered using the standard PostgreSQL interactive terminal \texttt{psql}, but other programs that have similar functionality can be used as well.

- Defining Queries
- Using Functions and Operators
- Query Profiling

Defining Queries

A query is a SQL command that views, changes or analyzes the data in a database. This section describes how to construct SQL queries in Greenplum Database.

- SQL Lexicon
- SQL Value Expressions

SQL Lexicon

SQL (structured query language) is the language used to access the database. The SQL language has a specific lexicon (words, special characters, etc.) used to construct queries or commands that the database engine can understand.

SQL input consists of a sequence of commands. A command is composed of a sequence of tokens, terminated by a semicolon (\texttt{;}). Which tokens are valid depends on the syntax of the particular command. The syntax rules for each command are described in “SQL Command Reference” on page 243.

Greenplum Database is based on PostgreSQL and adheres to the same SQL structure and syntax (with some minor exceptions). In most cases, the syntax is identical to PostgreSQL, however some commands may have additional or restricted syntax in Greenplum Database. For a complete explanation of the SQL rules and concepts as implemented in PostgreSQL, see SQL Syntax in the PostgreSQL documentation.

SQL Value Expressions

Value expressions are used in a variety of contexts, such as in the target list of the \texttt{SELECT} command, as new column values in \texttt{INSERT} or \texttt{UPDATE}, or in search conditions in a number of commands. The result of a value expression is sometimes called a \textit{scalar}, to distinguish it from the result of a table expression (which is a table). Value expressions are therefore also called scalar expressions (or even simply expressions). The expression syntax allows the calculation of values from primitive parts using arithmetic, logical, set, and other operations.

A value expression is one of the following:

- A constant or literal value.
• A column reference.
• A positional parameter reference, in the body of a function definition or prepared statement.
• A subscripted expression.
• A field selection expression.
• An operator invocation.
• A function call.
• An aggregate expression.
• A window expression.
• A type cast.
• A scalar subquery.
• An array constructor.
• A row constructor.
• Another value expression in parentheses, useful to group subexpressions and override precedence.

In addition to this list, there are a number of constructs that can be classified as an expression but do not follow any general syntax rules. These generally have the semantics of a function or operator and are explained in “Using Functions and Operators” on page 110.

Column References
A column can be referenced in the form:

\[ \text{correlation.columnname} \]

Where \text{correlation} is the name of a table (possibly qualified with a schema name), or an alias for a table defined by means of a \text{FROM} clause, or one of the key words \text{NEW} or \text{OLD}. (\text{NEW} and \text{OLD} can only appear in rewrite rules, while other correlation names can be used in any SQL statement.) The correlation name and separating dot may be omitted if the column name is unique across all the tables being used in the current query.

Positional Parameters
A positional parameter reference is used to indicate a value that is supplied externally to an SQL statement. Parameters are used in SQL function definitions and in prepared queries. Some client libraries also support specifying data values separately from the SQL command string, in which case parameters are used to refer to the out-of-line data values. The form of a parameter reference is:

\[ \$\text{number} \]

For example, consider the definition of a function, \text{dept}, as:

```
CREATE FUNCTION dept(text) RETURNS dept
AS $$ SELECT * FROM dept WHERE name = $1 $$
LANGUAGE SQL;
```
Here the $1$ references the value of the first function argument whenever the function is invoked.

**Subscripts**

If an expression yields a value of an array type, then a specific element of the array value can be extracted by writing:

```
expression[subscript]
```

Or multiple adjacent elements (an ‘array slice’) can be extracted by writing:

```
expression[lower_subscript:upper_subscript]
```

Here, the brackets [ ] are meant to appear literally. Each subscript is itself an expression, which must yield an integer value.

In general the array expression must be parenthesized, but the parentheses may be omitted when the expression to be subscripted is just a column reference or positional parameter. Also, multiple subscripts can be concatenated when the original array is multidimensional. For example:

```plaintext
mytable.arraycolumn[4]
mytable.two_d_column[17][34]
$1[10:42]
(arrayfunction(a,b))[42]
```

The parentheses in the last example are required.

**Field Selection**

If an expression yields a value of a composite type (row type), then a specific field of the row can be extracted by writing:

```
expression.fieldname
```

In general the row expression must be parenthesized, but the parentheses may be omitted when the expression to be selected from is just a table reference or positional parameter. For example:

```plaintext
mytable.mycolumn
$1.somecolumn
(rowfunction(a,b)).col3
```

Thus, a qualified column reference is actually just a special case of the field selection syntax.

**Operator Invocations**

There are three possible syntaxes for an operator invocation:

```
expression operator expression (binary infix operator)
operator expression (unary prefix operator)
expression operator (unary postfix operator)
```

Where `operator` is an operator token, one of the key words `AND`, `OR`, or `NOT`, or is a qualified operator name in the form:

```
OPERATOR(schema.operatorname)
```
Which particular operators exist and whether they are unary or binary depends on what operators have been defined by the system or the user. “Built-in Functions and Operators” on page 111 describes the built-in operators.

**Function Calls**

The syntax for a function call is the name of a function (possibly qualified with a schema name), followed by its argument list enclosed in parentheses:

```
function ([expression [, expression ... ]])
```

For example, the following function call computes the square root of 2:

```
sqrt(2)
```

The list of built-in functions is listed in “Built-in Functions and Operators” on page 111. Other functions may be added by the user.

**Aggregate Expressions**

An aggregate expression represents the application of an aggregate function across the rows selected by a query. An aggregate function reduces multiple inputs to a single output value, such as the sum or average of the inputs. The syntax of an aggregate expression is one of the following:

```
aggregate_name (expression [, ... ] ) [FILTER (WHERE condition)]
```

Where `aggregate_name` is a previously defined aggregate (possibly qualified with a schema name), and `expression` is any value expression that does not itself contain an aggregate expression.

The first form of aggregate expression invokes the aggregate across all input rows for which the given expression(s) yield non-null values. The second form is the same as the first, since `ALL` is the default. The third form invokes the aggregate for all distinct non-null values of the expressions found in the input rows. The last form invokes the aggregate once for each input row regardless of null or non-null values; since no particular input value is specified, it is generally only useful for the `count(*)` aggregate function.

For example, `count(*)` yields the total number of input rows; `count(f1)` yields the number of input rows in which `f1` is non-null; `count(distinct f1)` yields the number of distinct non-null values of `f1`.

The `FILTER` clause allows you to specify a condition to limit the input rows to the aggregate function. For example:

```
SELECT count(*) FILTER (WHERE gender='F') FROM employee;
```
The \texttt{WHERE} condition of the \texttt{FILTER} clause cannot contain a set returning function, subquery, a window function, or an outer reference. If using a user-defined aggregate function, the state transition function must be declared as \texttt{STRICT} (see \texttt{CREATE AGGREGATE}).

The predefined aggregate functions are described in \textit{“Aggregate Functions” on page 112}. Other aggregate functions may be added by the user.

An aggregate expression may only appear in the result list or \texttt{HAVING} clause of a \texttt{SELECT} command. It is forbidden in other clauses, such as \texttt{WHERE}, because those clauses are logically evaluated before the results of aggregates are formed.

When an aggregate expression appears in a subquery (see \textit{“Scalar Subqueries” on page 107} and \textit{“Subquery Expressions” on page 112}), the aggregate is normally evaluated over the rows of the subquery. But an exception occurs if the aggregate’s arguments contain only outer-level variables: the aggregate then belongs to the nearest such outer level, and is evaluated over the rows of that query. The aggregate expression as a whole is then an outer reference for the subquery it appears in, and acts as a constant over any one evaluation of that subquery. The restriction about appearing only in the result list or \texttt{HAVING} clause applies with respect to the query level that the aggregate belongs to.

Greenplum Database currently does not support \texttt{DISTINCT} with more than one input expression.

\section*{Window Expressions}

Window expressions allow application developers to more easily compose complex online analytical processing (OLAP) queries using standard SQL commands. For example, moving averages or sums can be calculated over various intervals; aggregations and ranks can be reset as selected column values change; and complex ratios can be expressed in simple terms.

A window expression represents the application of a \textit{window function} applied to a \textit{window frame}, which is defined in a special \texttt{OVER()} clause. A window partition is a set of rows that are grouped together for the purpose of applying an window function. Unlike aggregate functions, which return a result value for each group of rows, window functions return a result value for every row, but that value is calculated with respect to the rows in a particular window partition. If no partition is specified, the window function is computed over the complete intermediate result set.

The syntax of a window expression is:

\begin{verbatim}
    window_function ( [expression [, ...]] ) OVER ( \\
    window_specification )
\end{verbatim}

Where \textit{window_function} is one of the functions listed in \textit{“Window Functions” on page 113}, \textit{expression} is any value expression that does not itself contain a window expression, and \textit{window_specification} is:

\begin{verbatim}
    [window_name] \\
    [PARTITION BY expression [, ...]] \\
    [[ORDER BY expression [ASC | DESC | USING operator] [, ...] \\
    {{RANGE | ROWS} \\
    } UNBOUNDED PRECEDING
\end{verbatim}
Defining Queries

| expression PRECEDING |
| CURRENT ROW |
| BETWEEN window_frame_bound AND window_frame_bound ]]]

and where window_frame_bound can be one of:

- UNBOUNDED PRECEDING
- expression PRECEDING
- CURRENT ROW
- expression FOLLOWING
- UNBOUNDED FOLLOWING

A window expression may only appear in the select list of a SELECT command. For example:

```sql
SELECT count(*) OVER(PARTITION BY customer_id), * FROM sales;
```

The OVER clause is what differentiates window functions from other aggregate or reporting functions. The OVER clause defines the window_specification to which the window function is applied. A window specification has the following characteristics:

- The PARTITION BY clause, which defines the window partitions to which the window function is applied. If omitted, the entire result set is treated as one partition.

- The ORDER BY clause defines the expression(s) for sorting rows within a window partition. Note that the ORDER BY of a window specification is separate and distinct from the ORDER BY clause of a regular query expression (see “The ORDER BY Clause” on page 525). The ORDER BY clause is required for the window functions that calculate rankings, as it identifies the measure(s) for the ranking values. For OLAP aggregations, the ORDER BY clause is required in order to use window frames (the ROWS | RANGE clause).

**NOTE:** Columns of data types that lack a coherent ordering, such as time, are not good candidates for use in the ORDER BY clause of a window specification. Time, with or without time zone, lacks a coherent ordering because addition and subtraction do not have the expected effects. For example, the following is not generally true: `x::time < x::time + '2 hour'::interval`

- The ROWS/RANGE clause is used to define a window frame for aggregate (non-ranking) window functions. A window frame defines a set of rows within a window partition. When a window frame is defined, the window function is computed with respect to the contents of this moving frame rather than the fixed contents of the entire window partition. Window frames can be row-based (ROWS) or value-based (RANGE).

**Type Casts**

A type cast specifies a conversion from one data type to another. Greenplum Database (as does PostgreSQL) accepts two equivalent syntaxes for type casts:

```sql
CAST ( expression AS type )

expression::type
```

The CAST syntax conforms to SQL; the syntax with :: is historical PostgreSQL usage.
When a cast is applied to a value expression of a known type, it represents a run-time type conversion. The cast will succeed only if a suitable type conversion function has been defined. Notice that this is subtly different from the use of casts with constants. A cast applied to an unadorned string literal represents the initial assignment of a type to a literal constant value, and so it will succeed for any type (if the contents of the string literal are acceptable input syntax for the data type).

An explicit type cast may usually be omitted if there is no ambiguity as to the type that a value expression must produce (for example, when it is assigned to a table column); the system will automatically apply a type cast in such cases. However, automatic casting is only done for casts that are marked “OK to apply implicitly” in the system catalogs. Other casts must be invoked with explicit casting syntax. This restriction is intended to prevent surprising conversions from being applied silently.

**Scalar Subqueries**

A scalar subquery is an ordinary `SELECT` query in parentheses that returns exactly one row with one column. The `SELECT` query is executed and the single returned value is used in the surrounding value expression. It is an error to use a query that returns more than one row or more than one column as a scalar subquery. The subquery can refer to variables from the surrounding query, which will act as constants during any one evaluation of the subquery. For example, the following finds the largest city population in each state:

```
SELECT name, (SELECT max(pop) FROM cities WHERE cities.state = states.name) FROM states;
```

**Array Constructors**

An array constructor is an expression that builds an array value from values for its member elements. A simple array constructor consists of the key word `ARRAY`, a left square bracket `[`, one or more expressions (separated by commas) for the array element values, and finally a right square bracket `]`. For example,

```
SELECT ARRAY[1,2,3+4];
array
-------
{1,2,7}
```

The array element type is the common type of the member expressions, determined using the same rules as for `UNION` or `CASE` constructs.

Multidimensional array values can be built by nesting array constructors. In the inner constructors, the key word `ARRAY` may be omitted. For example, these two `SELECT` statements produce the same result:

```
SELECT ARRAY[ARRAY[1,2], ARRAY[3,4]];
SELECT ARRAY[[1,2],[3,4]];
array
-------------
{{1,2},{3,4}}
```

Since multidimensional arrays must be rectangular, inner constructors at the same level must produce sub-arrays of identical dimensions.
Multidimensional array constructor elements can be anything yielding an array of the proper kind, not only a sub-ARRAY construct. For example:

```
CREATE TABLE arr(f1 int[], f2 int[]);
INSERT INTO arr VALUES (ARRAY[[1,2],[3,4]],
ARRAY[[5,6],[7,8]]);
SELECT ARRAY[f1, f2, '((9,10),(11,12))':int[]] FROM arr;
```

```
array
------------------------------------------------
{{{1,2},{3,4}},{{5,6},{7,8}},{{9,10},{11,12}}}
```

It is also possible to construct an array from the results of a subquery. In this form, the array constructor is written with the key word ARRAY followed by a parenthesized (not bracketed) subquery. For example:

```
SELECT ARRAY(SELECT oid FROM pg_proc WHERE proname LIKE 'bytea%');
```

```
?column?
-----------------------------------------------------------
```

(The subquery must return a single column. The resulting one-dimensional array will have an element for each row in the subquery result, with an element type matching that of the subquery’s output column. The subscripts of an array value built with ARRAY always begin with one.)

### Row Constructors

A row constructor is an expression that builds a row value (also called a composite value) from values for its member fields. A row constructor consists of the key word ROW, a left parenthesis, zero or more expressions (separated by commas) for the row field values, and finally a right parenthesis. For example,

```
SELECT ROW(1,2.5,'this is a test');
```

A row constructor can include the syntax rowvalue.*, which will be expanded to a list of the elements of the row value, just as occurs when the .* syntax is used at the top level of a SELECT list. For example, if table t has columns f1 and f2, these are the same:

```
SELECT ROW(t.*, 42) FROM t;
SELECT ROW(t.f1, t.f2, 42) FROM t;
```

By default, the value created by a ROW expression is of an anonymous record type. If necessary, it can be cast to a named composite type — either the row type of a table, or a composite type created with CREATE TYPE AS. An explicit cast may be needed to avoid ambiguity. For example:

```
CREATE TABLE mytable(f1 int, f2 float, f3 text);
CREATE FUNCTION getf1(mytable) RETURNS int AS 'SELECT $1.f1'
LANGUAGE SQL;

-- No cast needed since only one getf1() exists
SELECT getf1(ROW(1,2.5,'this is a test'));
```
CREATE TYPE myrowtype AS (f1 int, f2 text, f3 numeric);
CREATE FUNCTION getf1(myrowtype) RETURNS int AS 'SELECT $1.f1' LANGUAGE SQL;
-- Now we need a cast to indicate which function to call:
SELECT getf1(ROW(1,2.5,'this is a test'));
ERROR: function getf1(record) is not unique
SELECT getf1(ROW(1,2.5,'this is a test')::mytable);
getf1
-----
  1
SELECT getf1(CAST(ROW(11,'this is a test',2.5) AS myrowtype));
getf1
-----
  11
Row constructors can be used to build composite values to be stored in a composite-type table column, or to be passed to a function that accepts a composite parameter.

**Expression Evaluation Rules**

The order of evaluation of subexpressions is not defined. In particular, the inputs of an operator or function are not necessarily evaluated left-to-right or in any other fixed order.

Furthermore, if the result of an expression can be determined by evaluating only some parts of it, then other subexpressions might not be evaluated at all. For instance, if one wrote:

```sql
SELECT true OR somefunc();
```

then `somefunc()` would (probably) not be called at all. The same would be the case if one wrote:

```sql
SELECT somefunc() OR true;
```

Note that this is not the same as the left-to-right ‘forced evaluation order’ of Boolean operators that is found in some programming languages.

As a consequence, it is unwise to use functions with side effects as part of complex expressions. It is particularly dangerous to rely on side effects or evaluation order in **WHERE** and **HAVING** clauses, since those clauses are extensively reprocessed as part of developing an execution plan. Boolean expressions (**AND**/**OR**/**NOT** combinations) in those clauses may be reorganized in any manner allowed by the laws of Boolean algebra.
When it is essential to force evaluation order, a `CASE` construct may be used. For example, this is an untrustworthy way of trying to avoid division by zero in a `WHERE` clause:

```sql
SELECT ... WHERE x <> 0 AND y/x > 1.5;
```

But this is safe:

```sql
SELECT ... WHERE CASE WHEN x <> 0 THEN y/x > 1.5 ELSE false END;
```

A `CASE` construct used in this fashion will defeat optimization attempts, so it should only be done when necessary.

---

**Using Functions and Operators**

- Using Functions in Greenplum Database
- User-Defined Functions
- Built-in Functions and Operators
- Window Functions
- Advanced Analytic Functions

---

**Using Functions in Greenplum Database**

A function can be one of three types: **IMMUTABLE**, **STABLE**, or **VOLATILE**. Greenplum Database offers full support of all **IMMUTABLE** functions. An immutable function is a function that relies only on information directly present in its argument list and will always return the same result when given the same argument values.

The use of **STABLE** functions is supported in most cases in Greenplum Database. **STABLE** indicates that within a single table scan the function will consistently return the same result for the same argument values, but that its result could change across SQL statements. Functions whose results depend on database lookups or parameter variables are classified as **STABLE**. Also note that the `current_timestamp` family of functions qualify as stable, since their values do not change within a transaction.

The use of **VOLATILE** functions is restricted in Greenplum Database. **VOLATILE** indicates that the function value can change even within a single table scan. Relatively few database functions are volatile in this sense; some examples are `random()`, `currval()`, `timeofday()`. But note that any function that has side-effects must be classified volatile, even if its result is quite predictable (for example, `setval()`).

In Greenplum Database, the data is divided up across the segments — each segment is, in a sense, its own distinct PostgreSQL database. To prevent data from becoming out-of-sync across the segments, any function classified **VOLATILE** cannot be executed at the segment level if it contains SQL or modifies the database in any way. For example, functions such as `random()` or `timeofday()` are not allowed to execute on distributed data in Greenplum Database because they could potentially cause inconsistent data between the segment instances.
To ensure data consistency, VOLATILE functions can safely be used in statements that are evaluated on and execute from the master. For example, the following statements are always executed on the master (statements without a FROM clause):

```sql
SELECT setval('myseq', 201);
SELECT foo();
```

In cases where a statement has a FROM clause containing a distributed table and the function used in the FROM clause simply returns a set of rows, execution may be allowed on the segments:

```sql
SELECT * from foo();
```

One exception to this rule are functions that return a table reference (rangeFuncs) or functions that use the refCursor datatype. These types of functions cannot be used at all in Greenplum Database.

**User-Defined Functions**

Greenplum Database supports user-defined functions as does PostgreSQL. See the section on Extending SQL in the PostgreSQL documentation for more information. You can use the CREATE FUNCTION command to register user-defined functions as long as they are used as described in “Using Functions in Greenplum Database” on page 110. By default, functions are declared as VOLATILE, so it is important to specify the correct volatility level if you are registering a user-defined function that is IMMUTABLE or STABLE.

When creating user-defined functions, avoid using fatal errors or any kind of destructive call. The Greenplum Database server may respond to such errors with a sudden shutdown or restart.

Note that in Greenplum Database, the shared library files for user-created functions must reside in the same library path location on every host in the Greenplum Database array (masters, segments, and mirrors).

**Built-in Functions and Operators**

The following table lists the categories of built-in functions and operators supported by PostgreSQL. All functions and operators are supported in Greenplum Database as in PostgreSQL with the exception of STABLE and VOLATILE functions, which are subject to the restrictions noted in “Using Functions in Greenplum Database” on page 110. See the Functions and Operators section of the PostgreSQL documentation for more information about these built-in functions and operators.

<table>
<thead>
<tr>
<th>Operator/Function Category</th>
<th>VOLATILE Functions</th>
<th>STABLE Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical Operators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison Operators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematical Functions and Operators</td>
<td>random setseed</td>
<td></td>
</tr>
</tbody>
</table>
### Table 11.1  Built-in functions and operators

<table>
<thead>
<tr>
<th>Operator/Function Category</th>
<th>VOLATILE Functions</th>
<th>STABLE Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>String Functions and Operators</td>
<td>All built-in conversion functions</td>
<td>convert pg_client_encoding</td>
</tr>
<tr>
<td>Binary String Functions and Operators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bit String Functions and Operators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pattern Matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Type Formatting Functions</td>
<td>to_char to_timestamp</td>
<td></td>
</tr>
<tr>
<td>Date/Time Functions and Operators</td>
<td>timeofday</td>
<td>age current_date current_time current_timestamp localtime localtimestamp now</td>
</tr>
<tr>
<td>Geometric Functions and Operators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network Address Functions and Operators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequence Manipulation Functions</td>
<td>curval lastval nextval setval</td>
<td></td>
</tr>
<tr>
<td>Conditional Expressions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Array Functions and Operators</td>
<td>All array functions</td>
<td></td>
</tr>
<tr>
<td>Aggregate Functions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subquery Expressions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row and Array Comparisons</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set Returning Functions</td>
<td>generate_series</td>
<td></td>
</tr>
</tbody>
</table>
Window Functions

Greenplum Database has the following built-in window functions that are not available in PostgreSQL. Window functions are often used to compose complex OLAP (online analytical processing) queries. Window functions are applied to partitioned result sets within the scope of a single query expression. A window partition is a subset of rows returned by a query, as defined in a special `OVER()` clause. See “Window Expressions” on page 105. All window functions are immutable functions.

Note that any aggregate function (as described in Aggregate Functions of the PostgreSQL documentation) may also be used with an `OVER` clause, thereby making it a window aggregate function. A window aggregate function returns the aggregated value of the expression argument(s) for the specified window frame corresponding to a particular row. If the `OVER` clause does not specify ordering (`ORDER BY`) or window framing (`ROWS | RANGE`), the value is aggregated over the defined window partition. In the case where ordering and framing are not specified, if the aggregate function allows

### Table 11.1  Built-in functions and operators

<table>
<thead>
<tr>
<th>Operator/Function Category</th>
<th>VOLATILE Functions</th>
<th>STABLE Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Information Functions</td>
<td></td>
<td>All session information functions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All access privilege inquiry functions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All schema visibility inquiry functions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All system catalog information functions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All comment information functions</td>
</tr>
<tr>
<td>System Administration Functions</td>
<td>set_config</td>
<td>current_setting</td>
</tr>
<tr>
<td></td>
<td>pg_cancel_backend</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pg_reload_conf</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pg_rotate_logfile</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pg_start_backup</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pg_stop_backup</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pg_size_pretty</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pg_ls_dir</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pg_read_file</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pg_stat_file</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>All database object size functions</td>
</tr>
</tbody>
</table>
the **DISTINCT** qualifier, then so does its corresponding window aggregate function. **DISTINCT** is not allowed for window aggregate functions that specify ordering or framing.

<table>
<thead>
<tr>
<th>Function</th>
<th>Return Type</th>
<th>Full Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>cume_dist()</code></td>
<td>double precision</td>
<td><code>CUME_DIST()</code> OVER ( [PARTITION BY expr] ORDER BY expr )</td>
<td>Calculates the cumulative distribution of a value in a group of values. Rows with equal values always evaluate to the same cumulative distribution value.</td>
</tr>
<tr>
<td><code>dense_rank()</code></td>
<td>bigint</td>
<td><code>DENSE_RANK () OVER ( [PARTITION BY expr] ORDER BY expr)</code></td>
<td>Computes the rank of a row in an ordered group of rows without skipping rank values. Rows with equal values are given the same rank value.</td>
</tr>
<tr>
<td><code>first_value(expr)</code></td>
<td>same as input expr type</td>
<td>`FIRST_VALUE(expr) OVER ( [PARTITION BY expr] ORDER BY expr [ROWS</td>
<td>RANGE frame_expr] )`</td>
</tr>
<tr>
<td><code>lag(expr [,offset] [,default])</code></td>
<td>same as input expr type</td>
<td><code>LAG(expr [,offset] [,default]) OVER ( [PARTITION BY expr] ORDER BY expr )</code></td>
<td>Provides access to more than one row of the same table without doing a self join. Given a series of rows returned from a query and a position of the cursor, <strong>LAG</strong> provides access to a row at a given physical offset prior to that position. If <code>offset</code> is not specified, the default offset is 1. <code>default</code> sets the value that is returned if the offset goes beyond the scope of the window. If <code>default</code> is not specified, the default value is null.</td>
</tr>
<tr>
<td><code>last_value(expr)</code></td>
<td>same as input expr type</td>
<td>`LAST_VALUE(expr) OVER ( [PARTITION BY expr] ORDER BY expr [ROWS</td>
<td>RANGE frame_expr] )`</td>
</tr>
<tr>
<td><code>lead(expr [,offset] [,default])</code></td>
<td>same as input expr type</td>
<td><code>LEAD(expr [,offset] [,default]) OVER ( [PARTITION BY expr] ORDER BY expr )</code></td>
<td>Provides access to more than one row of the same table without doing a self join. Given a series of rows returned from a query and a position of the cursor, <strong>LEAD</strong> provides access to a row at a given physical offset after that position. If <code>offset</code> is not specified, the default offset is 1. <code>default</code> sets the value that is returned if the offset goes beyond the scope of the window. If <code>default</code> is not specified, the default value is null.</td>
</tr>
</tbody>
</table>
### Table 11.2  Window functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Return Type</th>
<th>Full Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ntile(expr)</td>
<td>bigint</td>
<td>NTILE(expr) OVER ( [PARTITION BY expr] ORDER BY expr )</td>
<td>Divides an ordered dataset into a number of buckets (as defined by expr) and assigns a bucket number to each row.</td>
</tr>
<tr>
<td>percent_rank()</td>
<td>double precision</td>
<td>PERCENT_RANK () OVER ( [PARTITION BY expr] ORDER BY expr )</td>
<td>Calculates the rank of a hypothetical row ( R ) minus 1, divided by 1 less than the number of rows being evaluated (within a window partition).</td>
</tr>
<tr>
<td>rank()</td>
<td>bigint</td>
<td>RANK () OVER ( [PARTITION BY expr] ORDER BY expr )</td>
<td>Calculates the rank of a row in an ordered group of values. Rows with equal values for the ranking criteria receive the same rank. The number of tied rows are added to the rank number to calculate the next rank value. Ranks may not be consecutive numbers in this case.</td>
</tr>
<tr>
<td>row_number()</td>
<td>bigint</td>
<td>ROW_NUMBER () OVER ( [PARTITION BY expr] ORDER BY expr )</td>
<td>Assigns a unique number to each row to which it is applied (either each row in a window partition or each row of the query).</td>
</tr>
</tbody>
</table>

### Advanced Analytic Functions

Greenplum Database has the following built-in advanced analytic functions that are not available in PostgreSQL. All of these analytic functions are immutable functions.

### Table 11.3  Advanced Analytic Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Return Type</th>
<th>Full Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>matrix_add(array[], array[])</td>
<td>smallint[], int[], bigint[], float[]</td>
<td>matrix_add(array[[1,1],[2,2]], array[[3,4],[5,6]])</td>
<td>Adds two two-dimensional matrices. The matrices must be conformable.</td>
</tr>
<tr>
<td>matrix_multiply(array [], array [])</td>
<td>smallint[], int[], bigint[], float[]</td>
<td>matrix_multiply(array[[2,0,0],[0,2,0],[0,0,2]], [[3,0,3],[0,3,0]])</td>
<td>Multiplies two two-dimensional arrays. The matrices must be conformable.</td>
</tr>
<tr>
<td>matrix_multiply(array [], expr)</td>
<td>int[], float[]</td>
<td>matrix_multiply(array[[1,1,1],[2,2,2],[3,3,3]], 2)</td>
<td>Multiplies a two-dimensional array and a scalar numeric value.</td>
</tr>
<tr>
<td>matrix_transpose(array [])</td>
<td>Same as input array type.</td>
<td>matrix_transpose(array [[1,1,1],[2,2,2]])</td>
<td>Transposes a two-dimensional array.</td>
</tr>
</tbody>
</table>
### Table 11.3 Advanced Analytic Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Return Type</th>
<th>Full Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pinv(array [])</td>
<td>smallint[], int[], bigint[], float[]</td>
<td>pinv(array[[2.5, 0, 0], [0, 1, 0], [0, 0, .5]])</td>
<td>Calculates the Moore-Penrose pseudoinverse of a matrix.</td>
</tr>
<tr>
<td>unnest (int [])</td>
<td>set of anyelement</td>
<td>unnest(array['one', 'row', 'per', 'item'])</td>
<td>Transforms a one dimensional array into rows. Returns a set of anyelement, a polymorphic pseudotype in PostgreSQL.</td>
</tr>
</tbody>
</table>

### Table 11.4 Advanced Aggregate Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Return Type</th>
<th>Full Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sum(array [])</td>
<td>smallint[], int[], bigint[], float[]</td>
<td>sum(array[[1, 2], [3, 4]])</td>
<td>Performs matrix summation. Can take as input a two-dimensional array that is treated as a matrix.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Example:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CREATE TABLE mymatrix (myvalue int[]);</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>INSERT INTO mymatrix VALUES (array[1,2],[3,4]);</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>INSERT INTO mymatrix VALUES (array[0,1],[1,0]);</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SELECT sum(myvalue) FROM mymatrix;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>sum</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>{{1,3},{4,4}}</td>
<td></td>
</tr>
<tr>
<td>pivot_sum(label[], label, expr)</td>
<td>int[], bigint[], float[]</td>
<td>pivot_sum(array['A1','A2'], attr, value)</td>
<td>A pivot aggregation using sum to resolve duplicate entries.</td>
</tr>
<tr>
<td>mregr_coef(expr, array[])</td>
<td>float[]</td>
<td>mregr_coef(y, array[1, x1, x2])</td>
<td>The functions mregr_coef and mregr_r2 are both multiple linear regression aggregates. mregr_coef calculates the regression coefficients. The size of the return array for mregr_coef is the same as the size of the input array for mregr_r2. The array contains the coefficient for each independent variable.</td>
</tr>
</tbody>
</table>
Greenplum Database Administrator Guide 4.0 – Chapter 11: Querying Data

Advanced Analytic Function Examples

These examples illustrate selected advanced analytic functions in queries on simplified example data. Examples are given for the multiple linear regression aggregate functions, and for Naive Bayes Classification with \texttt{nb\_classify}.

Linear Regression Functions Example

This example uses the linear regression functions \texttt{mregr\_coef} and \texttt{mregr\_r2} in a query on the example table \texttt{regr\_example}. In this example query, both functions take the dependent variable as the first parameter and an array of independent variables as the second parameter.

```sql
SELECT mregr_coef(y, array[1, x1, x2]), mregr_r2(y, array[1, x1, x2]) from regr_example;
```

Table \texttt{regr\_example}:

<table>
<thead>
<tr>
<th>id</th>
<th>y</th>
<th>x1</th>
<th>x2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

Running the example query against this simple table yields one row of data displaying these values:

\texttt{mregr\_coef}:

\{-3.90798504668055e-14, 1.00000000000003, 2.99999999999994\}

\texttt{mregr\_r2}:

0.99999999999953

If an intercept term is desired, it must be manually entered.

<table>
<thead>
<tr>
<th>Function</th>
<th>Return Type</th>
<th>Full Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\texttt{mregr_r2} ( (\text{expr}, \text{array}[]) )</td>
<td>\texttt{float()}</td>
<td>\texttt{mregr_r2(\text{y, array}[1, \text{x1}, \text{x2}])}</td>
<td>The functions \texttt{mregr_coef} and \texttt{mregr_r2} are both multiple linear regression aggregates. \texttt{mregr_r2} calculates the r-squared error value for the regression.</td>
</tr>
<tr>
<td>\texttt{nb_classify} ( (\text{text}[], \text{bigint}, \text{bigint}[], \text{bigint}[]) )</td>
<td>\texttt{text[]}</td>
<td>\texttt{nb_classify(\text{classes, attr_count, class_count, class_total})}</td>
<td>Classify rows using a naive Bayes classifier. This function is used with a baseline of training data to predict the classification of new rows.</td>
</tr>
</tbody>
</table>

Table 11.4 Advanced Aggregate Functions
Naive Bayes Classification Example

The function `nb_classify` is used within a larger classification process that involves the creation of tables and views for training data. This example begins with the normalized data in the example table `class_example` and proceeds through four discrete steps, as described below.

**Table `class_example`:**

<table>
<thead>
<tr>
<th>id</th>
<th>class</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>C1</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>C2</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>C1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>C2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>C2</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

1. **Unpivot the data**

   For use as training data, the data in `class_example` must be unpivoted. In cases where the data is already in denormalized form, this step is not required.

   ```sql
   CREATE view class_example_unpivot AS
   SELECT id, class, unnest(array['A1', 'A2', 'A3']) as attr,
   unnest(array[a1,a2,a3]) as value FROM class_example;
   ```

2. **Create a training table from the unpivoted data**

   ```sql
   CREATE table class_example_nb_training AS
   SELECT attr, value, pivot_sum(array['C1', 'C2'], class, 1)
   as class_count
   FROM   class_example_unpivot
   GROUP BY attr, value
   DISTRIBUTED by (attr);
   ```

3. **Create a summary view over the training data**

   ```sql
   CREATE VIEW class_example_nb_classify AS
   SELECT attr, value, class_count, array['C1', 'C2'] as classes,
   sum(class_count) over (wa)::integer[] as class_total,
   count(distinct value) over (wa) as attr_count
   FROM class_example_nb_training
   WINDOW wa as (partition by attr);
   ```

4. **Classify rows with `nb_classify`**

   Once the view is prepared, the training data is ready for use as a baseline for determining the class of incoming rows. The following query predicts whether rows are of class C1 or C2 by using the `nb_classify` aggregate:

   ```sql
   SELECT nb_classify(classes, attr_count, class_count, class_total) as class
   FROM class_example_nb_classify
   where (attr = 'A1' and value = 0) or (attr = 'A2' and value =
This query yields the expected single-row result of C1.

Actual data in production scenarios will be more extensive than this simple example data, and will therefore yield better results. Accuracy of classification with `nb_classify` improves significantly with larger sets of training data.

---

### Query Profiling

Greenplum Database devises a query plan for each query it is given. Choosing the right query plan to match the query and data structure is absolutely critical for good performance. A query plan defines how the query will be executed in Greenplum Database’s parallel execution environment. By examining the query plans of poorly performing queries, you can identify possible performance tuning opportunities.

The query planner uses the database statistics it has to choose a query plan with the lowest possible cost. Cost is measured in disk I/O and CPU effort (shown as units of disk page fetches). The goal is to minimize the total execution cost for the plan.

You can view the plan for a given query using the `EXPLAIN` command. This will show the query planner’s estimated plan for the query. For example:

```sql
EXPLAIN SELECT * FROM names WHERE id=22;
```

`EXPLAIN ANALYZE` causes the statement to be actually executed, not only planned. This is useful for seeing whether the planner’s estimates are close to reality. For example:

```sql
EXPLAIN ANALYZE SELECT * FROM names WHERE id=22;
```

### Reading EXPLAIN Output

Query plans are a tree plan of nodes. Each node in the plan represents a single operation, such as table scan, join, aggregation or a sort.

Plans should be read from the bottom up as each node feeds rows into the node directly above it. The bottom nodes of a plan are usually table scan operations (sequential, index or bitmap index scans). If the query requires joins, aggregations, or sorts (or other operations on the raw rows) then there will be additional nodes above the scan nodes to perform these operations. The topmost plan nodes are usually the Greenplum Database motion nodes (redistribute, broadcast, or gather motions). These are the operations responsible for moving rows between the segment instances during query processing.

The output of `EXPLAIN` has one line for each node in the plan tree, showing the basic node type plus the following cost estimates that the planner made for the execution of that plan node:
• **cost** - measured in units of disk page fetches; that is, 1.0 equals one sequential disk page read. The first estimate is the start-up cost (cost of getting to the first row) and the second is the total cost (cost of getting all rows). Note that the total cost assumes that all rows will be retrieved, which may not always be the case (if using `LIMIT` for example).

• **rows** - the total number of rows output by this plan node. This is usually less than the actual number of rows processed or scanned by the plan node, reflecting the estimated selectivity of any `WHERE` clause conditions. Ideally the top-level nodes estimate will approximate the number of rows actually returned, updated, or deleted by the query.

• **width** - total bytes of all the rows output by this plan node.

It is important to note that the cost of an upper-level node includes the cost of all its child nodes. The topmost node of the plan has the estimated total execution cost for the plan. This is this number that the planner seeks to minimize. It is also important to realize that the cost only reflects things that the query planner cares about. In particular, the cost does not consider the time spent transmitting result rows to the client.

### EXPLAIN Example

To illustrate how to read an `EXPLAIN` query plan, consider the following example for a very simple query:

```
EXPLAIN SELECT * FROM names WHERE name = 'Joelle';
```

```
QUERY PLAN
------------------------------------------------------------
Gather Motion 2:1 (slice1) (cost=0.00..20.88 rows=1 width=13)
  -> Seq Scan on 'names' (cost=0.00..20.88 rows=1 width=13)
        Filter: name::text ~~ 'Joelle'::text
```

If we read the plan from the bottom up, the query planner starts by doing a sequential scan of the `names` table. Notice that the `WHERE` clause is being applied as a `filter` condition. This means that the scan operation checks the condition for each row it scans, and outputs only the ones that pass the condition.

The results of the scan operation are passed up to a *gather motion* operation. In Greenplum Database, a gather motion is when segments send rows up to the master. In this case we have 2 segment instances sending to 1 master instance (2:1). This operation is working on `slice1` of the parallel query execution plan. In Greenplum Database a query plan is divided into `slices` so that portions of the query plan can be worked on in parallel by the segments.

The estimated startup cost for this plan is **00.00** (no cost) and a total cost of **20.88** disk page fetches. The planner is estimating that this query will return one row.

### Reading EXPLAIN ANALYZE Output

`EXPLAIN ANALYZE` causes the statement to be actually executed, not only planned. The `EXPLAIN ANALYZE` plan shows the actual results along with the planner’s estimates. This is useful for seeing whether the planner’s estimates are close to reality. In addition to the information shown in the `EXPLAIN` plan, `EXPLAIN ANALYZE` will show the following additional information:
- The total elapsed time (in milliseconds) that it took to run the query.
- The number of *workers* (segments) involved in a plan node operation. Only segments that return rows are counted.
- The maximum number of rows returned by the segment that produced the most rows for an operation. If multiple segments produce an equal number of rows, the one with the longest *time to end* is the one chosen.
- The segment id number of the segment that produced the most rows for an operation.
- For relevant operations, the *work_mem* used by the operation. If *work_mem* was not sufficient to perform the operation in memory, the plan will show how much data was spilled to disk and how many passes over the data were required for the lowest performing segment. For example:

  Work_mem used: 64K bytes avg, 64K bytes max (seg0).
  Work_mem wanted: 90K bytes avg, 90K bytes max (seg0) to abate workfile I/O affecting 2 workers.
  [seg0] pass 0: 488 groups made from 488 rows; 263 rows written to workfile
  [seg0] pass 1: 263 groups made from 263 rows
- The time (in milliseconds) it took to retrieve the first row from the segment that produced the most rows, and the total time taken to retrieve all rows from that segment. The `<time> to first row` may be omitted if it is the same as the `<time> to end`.

**EXPLAIN ANALYZE Example**

To illustrate how to read an EXPLAIN ANALYZE query plan, we will use the same simple query we used in the “EXPLAIN Example” on page 120. Notice that there is some additional information in this plan that is not in a regular EXPLAIN plan. The parts of the plan in **bold** show the actual timing and rows returned for each plan node:

```
EXPLAIN ANALYZE SELECT * FROM names WHERE name = 'Joelle';
```

```
QUERY PLAN
------------------------------------------------------------
Gather Motion 2:1 (slice1) (cost=0.00..20.88 rows=1 width=13)
  recv: Total 1 rows with 0.305 ms to first row, 0.537 ms to end.
  -> Seq Scan on 'names' (cost=0.00..20.88 rows=1 width=13)
    Total 1 rows (seg0) with 0.255 ms to first row, 0.486 ms to end.
    Filter: name::text ~~ 'Joelle'::text

22.548 ms elapsed
```

If we read the plan from the bottom up, you will see some additional information for each plan node operation. The total elapsed time it took to run this query was 22.548 milliseconds.

The *sequential scan* operation had only one segment (*seg0*) that returned rows, and it returned just 1 row. It took 0.255 milliseconds to find the first row and 0.486 to scan all rows. Notice that this is pretty close to the planner’s estimate — the query planner estimated that it would return one row for this query, which it did. The *gather motion* operation then received 1 row (segments sending up to the master). The total elapsed time for this operation was 0.537 milliseconds.
What to Look for in a Query Plan

If a query is performing poorly, looking at its query plan can help identify problem areas. Here are some things to look for:

- When looking through the query plan, is there one operation that is consuming the majority of the query processing time? For example, if an index scan is taking longer than expected, perhaps the index is out-of-date and needs to be reindexed. You could also temporarily experiment with the enable parameters to see if you can force the planner to choose a different (and potentially better) plan by disabling a particular query plan operator for that query.

- **Are the planner’s estimates close to reality?** Run an `EXPLAIN ANALYZE` and see if the number of rows estimated by the planner is close to the number of rows actually returned by the query operation. If there is a huge discrepancy, you may need to collect more statistics on the relevant columns. See “Maintaining Database Statistics” on page 218.

- **Are selective predicates applied early in the plan?** The most selective filters should be applied early in the plan so that less rows move up the plan tree. If the query plan is not doing a good job at estimating the selectivity of a query predicate, you may need to collect more statistics on the relevant columns. See “Maintaining Database Statistics” on page 218. You can also try reordering the `WHERE` clause of your SQL statement.

- **Is the planner choosing the best join order?** When you have a query that joins multiple tables, make sure that the planner is choosing the most selective join order. Joins that eliminate the largest number of rows should be done earlier in the plan so that less rows move up the plan tree. If the plan is not choosing the optimal join order, you can set `join_collapse_limit=1` and use explicit `JOIN` syntax in your SQL statement to force the planner to the specified join order. You can also collect more statistics on the relevant join columns. See “Maintaining Database Statistics” on page 218.

- **Is the planner selectively scanning partitioned tables?** If you are using table partitioning, is the planner selectively scanning only the child tables required to satisfy the query predicates? Do scans of the parent tables return 0 rows (they should, since the parent tables should not contain any data). For an example of a query plan that shows a selective partition scan, see “Verifying Your Partition Strategy” on page 82.

- **Is the planner choosing hash aggregate and hash join operations where applicable?** Hash operations are typically much faster than other types of joins or aggregations. Row comparison and sorting is done in memory rather than reading/writing from disk. In order for hash operations to be chosen, there has to be sufficient work memory available to hold the number of estimated rows. Try increasing work memory to see if you can get better performance for a given query. If possible run an `EXPLAIN ANALYZE` for the query, which will show you which plan operations spilled to disk, how much work memory they used, and how much was required to not spill to disk. For example:

```
Work_mem used: 23430K bytes avg, 23430K bytes max (seg0).
Work_mem wanted: 33649K bytes avg, 33649K bytes max (seg0) to lessen workfile I/O affecting 2 workers.
```
Note that the bytes wanted message from `EXPLAIN ANALYZE` is only a hint, based on the amount of data written to work files and is not exact. The minimum `work_mem` needed could be more or less than the suggested value.
12. Loading and Unloading Data

This chapter describes the various ways to load data into and get data out of Greenplum Database. Greenplum supports fast, parallel data loading and unloading for large amounts of data, in addition to single file (non-parallel) import and export for small amounts of data. This chapter contains the following topics:

- About the Greenplum Database Loading Tools
- Formatting Your Data Files
- Loading Data into Greenplum Database
- Unloading Data from Greenplum Database

About the Greenplum Database Loading Tools

This section provides a brief overview of the tools Greenplum provides for loading and unloading data.

- About External Tables
- About gpload
- About COPY

About External Tables

Greenplum Database provides two types of external tables — readable external tables for data loading, and writable external tables for data unloading. Used in conjunction with Greenplum’s parallel file distribution program (gpfdist), external tables provide full parallelism by utilizing the resources of all Greenplum segments when loading or unloading data.

Readable external tables provide an easy way to perform basic extraction, transformation, and loading (ETL) tasks that are common in data warehousing. External table data is read in parallel by all Greenplum Database segment instances, so large load operations are processed as quickly as possible. Once the external table is defined, its data can be queried directly (and in parallel) using SQL commands. You can, for example, select, join, or sort external table data. You can also create views for external tables. However, readable external tables cannot be modified. DML operations (UPDATE, INSERT, DELETE, or TRUNCATE) are not allowed.

There are two types of readable external tables — regular and web. The main difference between regular external tables and web external tables are their data sources. Regular external tables access static flat files, whereas web external tables access dynamic data sources — either on a web server or by executing OS commands or scripts. When a query is planned using a regular external table, the external table is considered rescannable since the data is thought to be static for the course of the query. For web external tables, the data is not rescannable because there is the possibility that the data could change during the course of the query’s execution.
A writable external table allows you to select rows from other database tables and output the rows to files, named pipes, or other executable programs. For example, you could unload data from Greenplum Database and send it to an executable that connects to another database or ETL tool to load the data elsewhere. Writable external tables can also be used as output targets for Greenplum parallel MapReduce calculations.

**About gpload**

gpload is a data loading utility that acts as an interface to Greenplum’s external table parallel loading feature. Using a load specification defined in a YAML formatted control file, gpload executes a load by invoking the Greenplum parallel file server program (gpfdist), creating an external table definition based on the source data defined, and executing an `INSERT`, `UPDATE` or `MERGE` operation to load the source data into the target table in the database.

**About COPY**

Greenplum Database also offers the standard PostgreSQL `COPY` command for loading and unloading data. `COPY` is a non-parallel load/unload mechanism, meaning that data is loaded/unloaded in a single process via the Greenplum master instance. In cases where a small amount of data is involved, `COPY` offers a simple way to move data into or out of the database in a single transaction, without the administrative overhead of setting up an external table.

**Formatting Your Data Files**

When you use the various Greenplum tools for loading and unloading data, you need to specify how your data is formatted. `COPY`, `CREATE EXTERNAL TABLE` and `gpload` have clauses that allow you to specify how your data is formatted. Data can be in either delimited text (TEXT) or comma separated values (CSV) format. External data must be formatted correctly in order to be read by Greenplum Database. This section explains the format of data files expected by Greenplum Database.

- Formatting Rows
- Formatting Columns
- Representing NULL Values
- Escaping
- Character Encoding

**Formatting Rows**

Greenplum Database expects rows of data to be separated by the LF character (Line feed, 0x0A), CR (Carriage return, 0x0D), or CR followed by LF (CR+LF, 0x0D 0x0A). LF is the standard newline representation on UNIX or UNIX-like operating systems. Other operating systems (such as Windows or Mac OS 9) may use CR individually, or CR+LF. All of these representations of a newline are supported by Greenplum Database as a row delimiter.
Formatting Columns

The default column or field delimiter is the horizontal TAB character (0x09) for text files and the comma character (0x2C) for CSV files. However, it is possible to declare another single character delimiter using the DELIMITER clause of COPY, CREATE EXTERNAL TABLE or gpload when defining your data format. The delimiter character must only appear between any two data value fields. Do not place a delimiter at the beginning or end of a row. For example, if using the pipe character ( | ) as your delimiter:

```
data value 1|data value 2|data value 3
```

Representing NULL Values

NULL is the value used to represent an unknown piece of data in a column or field. Within your data files you can designate a string to represent null values. The default string is \N (backslash-N) in TEXT mode, or an empty value with no quotations in CSV mode. You can also declare a different string using the NULL clause of COPY, CREATE EXTERNAL TABLE or gpload when defining your data format. For example, you might prefer an empty string for cases where you do not want to distinguish nulls from empty strings. When using the Greenplum Database loading tools, any data item that matches the designated null string will be considered a null value.

Escaping

The data file has two reserved characters that have special meaning to Greenplum Database:

- The designated delimiter character, which is used to separate columns or fields in the data file.
- The newline character used to designate a new row in the data file.

If your data contains either of these characters, you must escape the character so Greenplum treats it as data and not as a field separator or new row. By default, the escape character is a \ (backslash) for text-formatted files and a double quote (") for csv-formatted files.

Escaping in Text Formatted Files

By default, the escape character is a \ (backslash) for text-formatted files. If you want to use a different escape character, use the ESCAPE clause of COPY, CREATE EXTERNAL TABLE or gpload to declare a different escape character. In cases where your selected escape character is present in your data, you can use it to escape itself.

For example, suppose you have a table with three columns and you want to load the following three fields:

- \ = backslash
- | = vertical bar
- ! = exclamation point

Your designated delimiter character is | (pipe character), and your designated escape character is \ (backslash). The formatted row in your data file would look like this:
backslash = \ | vertical bar = | | exclamation point = !

Notice how the backslash character that is part of the data is escaped with another backslash character, and the pipe character that is part of the data is escaped with a backslash character.

The escape character can also be used to escape octal and hexadecimal sequences. When used in this way, the escaped value will get converted to the equivalent character when loaded into Greenplum Database. For example, to load the ampersand character (&), you could use the escape character to escape its equivalent hexadecimal (\0x26) or octal (\046) representation.

If there is no need to escape the data in text-formatted files, you can disable escaping using the ESCAPE clause of COPY, CREATE EXTERNAL TABLE or gpload as follows:

```
ESCAPE 'OFF'
```

This is useful for input data that contains a lot of backslash characters within the data itself (such as web log data).

### Escaping in CSV Formatted Files

By default, the escape character is a " (double quote) for CSV-formatted files. If you want to use a different escape character, use the ESCAPE clause of COPY, CREATE EXTERNAL TABLE or gpload to declare a different escape character. In cases where your selected escape character is present in your data, you can use it to escape itself.

For example, suppose you have a table with three columns and you want to load the following three fields:

- Free trip to A,B
- 5.89
- Special rate "1.79"

Your designated delimiter character is , (comma), and your designated escape character is " (double quote). The formatted row in your data file would look like this:

```
"Free trip to A,B","5.89","Special rate "1.79"
```

Notice how that for the comma character that is part of the data, the entire data value is enclosed in double quotes. Also notice how the double quotes that are part of the data are also escaped with a double quote even though the field value is enclosed in double quotes.

Embedding the entire field inside a set of double quotes also guarantees preservation of leading and trailing whitespace characters:

```
"Free trip to A,B ","5.89 ","Special rate ""1.79""
```

**Note:** In CSV mode, all characters are significant. A quoted value surrounded by white space, or any characters other than DELIMITER, will include those characters. This can cause errors if you import data from a system that pads CSV lines with white space out to some fixed width. If such a situation arises you might need to preprocess the CSV file to remove the trailing white space, before importing the data into Greenplum Database.
Character Encoding

A character encoding system consists of a code that pairs each character from a given repertoire with something else, such as a sequence of numbers or octets, in order to facilitate the transmission and storage of data. The character set support in Greenplum Database allows you to store text in a variety of character sets, including single-byte character sets such as the ISO 8859 series and multiple-byte character sets such as EUC (Extended UNIX Code), UTF-8, and Mule internal code. All supported character sets can be used transparently by clients, but a few are not supported for use within the server (that is, as a server-side encoding).

Data files must be in a character encoding recognized by Greenplum Database. See “Character Set Support” on page 939 for the supported character sets. Data files that contain invalid or unsupported encoding sequences will encounter errors when loading into Greenplum Database.

**Note:** On data files generated on a Microsoft Windows operating system, try running the dos2unix system command to remove any Windows-only characters prior to loading into Greenplum Database.

Loading Data into Greenplum Database

This section describes how to load data into Greenplum Database using both parallel load operations (file-based external tables, web external tables and gpload) and non-parallel load operations (COPY).

- “Creating and Using File-Based External Tables” on page 128
- “Creating and Using Web External Tables” on page 132
- “Loading Data Using an External Table” on page 133
- “Handling Load Errors” on page 133
- “Loading Data with gpload” on page 136
- “Loading Data with COPY” on page 137
- “Data Loading Performance Tips” on page 138

Creating and Using File-Based External Tables

The CREATE EXTERNAL TABLE command creates a regular readable external table definition in Greenplum Database. When you create an external table definition, you must specify the format of your input files (see “Formatting Your Data Files” on page 125), and the location of your external data sources. There are two protocols that you can use to access external table data sources; gpfdist (Greenplum’s parallel file server program) or the file protocol (file locations on your Greenplum segment machines). See CREATE EXTERNAL TABLE for details on these two protocols.
For the best performance and ease of administrative setup, the `gpfdist` protocol is preferred. One advantage of using the `gpfdist` file server program (as opposed to the file protocol) is that it ensures that all of the segments in your Greenplum Database system are fully utilized when reading from external table data files. See “Using the Greenplum Parallel File Server (gpfdist)” on page 129 for more information.

By default, if external table data contains an error, the entire command fails and no data is loaded into the target database table. To isolate data errors in external table data while still loading correctly formatted rows, define the external table with single row error handling. See “Handling Load Errors” on page 133.

**Defining the External Table**

Use the `CREATE EXTERNAL TABLE` command to define the external table and specify the location and format of the external table data files.

For example, if using a single Greenplum file server (`gpfdist`) instance on a machine with multiple NICs (see Figure 12.1):

```
=# CREATE EXTERNAL TABLE ext_expenses ( name text, 
    date date,  amount float4, category text, desc text ) 
  LOCATION ('gpfdist://etlhos:1:8081/*', 
            'gpfdist://etlhos:2:8081/*') 
  FORMAT 'TEXT' (DELIMITER ',');
```

For example, if using multiple Greenplum file server (`gpfdist`) instances (see Figure 12.2):

```
=# CREATE EXTERNAL TABLE ext_expenses ( name text, 
    date date,  amount float4, category text, desc text ) 
  LOCATION ('gpfdist://etlhos:1:8081/*', 
            'gpfdist://etlhos:2:8082/*') 
  FORMAT 'TEXT' (DELIMITER ',');
```

**Using the Greenplum Parallel File Server (gpfdist)**

This section describes the tasks involved in setting up and managing the Greenplum parallel file server program (`gpfdist`) for use with external tables.

- About gpfdist Setup and Performance
- Controlling Segment Parallelism
- Installing gpfdist
- Starting and Stopping gpfdist
- Troubleshooting gpfdist

**About gpfdist Setup and Performance**

The `gpfdist` program can serve data to the segment instances at an average rate of about 350 MB/s for delimited text formatted files and 200 MB/s for CSV formatted files. Therefore, you should consider the following options when running `gpfdist` in order to maximize the network bandwidth of your ETL systems:
• If your ETL machine is configured with multiple network interface cards (NICs), run one instance of `gpfdist` on your ETL host and then define your external table definition so that the host name of each NIC is declared in the `LOCATION` clause (see “Defining the External Table” on page 129 for an example). This allows network traffic between your Greenplum segment hosts and your ETL host to use all NICs simultaneously.

![Figure 12.1](image1.png)

**Figure 12.1**  External Table Using Single `gpfdist` Instance with Multiple NICs

• Run multiple `gpfdist` instances on your ETL host and divide your external data files equally between each instance. For example, if you have an ETL system with two network interface cards (NICs), then you could run two `gpfdist` instances on that machine to maximize your load performance. You would then divide the external table data files evenly between the two `gpfdist` programs.

![Figure 12.2](image2.png)

**Figure 12.2**  External Tables Using Multiple `gpfdist` Instances with Multiple NICs
**Controlling Segment Parallelism**

You can use the following server configuration parameter to control how many segment instances access a single gpfdist program at a time. 64 is the default. This allows you to control the number of segments processing external table files, while reserving some segments for other database processing. This parameter can be set in the postgresql.conf file of your master instance:

```
gp_external_max_segs
```

**Installing gpfdist**

The gpfdist program is installed in $GPHOME/bin of your Greenplum Database master host installation. Most likely you will want to run gpfdist from a machine other than your Greenplum Database master, such as on a machine devoted to ETL processing. To install gpfdist on another machine:

1. Copy gpfdist from $GPHOME/bin in your Greenplum installation to the remote machine.
2. Add gpfdist to your $PATH.

**Starting and Stopping gpfdist**

To start gpfdist, you must tell it from which directory it will be serving files and optionally the port to run on (defaults to HTTP port 8080).

To start gpfdist in the background (and log output messages and errors to a log file):

```
$ gpfdist -d /var/load_files -p 8081 -l /home/gpadmin/log &
```

If starting multiple gpfdist instances on the same ETL host (see Figure 12.2), then each should use a different port and base directory. For example:

```
$ gpfdist -d /var/load_files1 -p 8081 -l /home/gpadmin/log1 &
$ gpfdist -d /var/load_files2 -p 8082 -l /home/gpadmin/log2 &
```

To stop gpfdist when its running in the background:

```
--First find its process id:
$ ps ax | grep gpfdist (Linux)
$ ps -ef | grep gpfdist (Solaris)
--Then kill the process, for example:
$ kill 3456
```

**Troubleshooting gpfdist**

Keep in mind that gpfdist is accessed at runtime by the segment instances. Therefore, you must ensure that the Greenplum segment hosts have network access to gpfdist. The gpfdist program is a simple web server, so to test connectivity you can run the following command from each host in your Greenplum array (segments and master):

```
$ wget http://gpfdist_hostname:port,filename
```
Also, make sure that your `CREATE EXTERNAL TABLE` definition has the correct host name, port, and file names for `gpfdist`. The file names and paths specified should be relative to the directory from which `gpfdist` is serving files (the directory path used when you started the `gpfdist` program). See “Defining the External Table” on page 129.

Creating and Using Web External Tables

`CREATE EXTERNAL WEB TABLE` creates a web table definition in Greenplum Database. A web table is a type of external table that allows you to access dynamic data sources as though they were a regular database table. Web table data is considered *dynamic* (meaning the data could possibly change midstream during the execution of a query). Therefore, the query planner must choose plans that do not allow for rescanning of the web table data.

- **Web URLs.** Specify the `LOCATION` of files on a web server using the `http://` protocol. The web data file(s) must reside on a web server that is accessible by the Greenplum segment hosts. The number of URLs specified corresponds to the number of segment instances that will work in parallel to access the web table. So for example, if you have a Greenplum Database system with 8 primary segments and you specify 2 external files, only 2 of the 8 segments will access the web table in parallel at query runtime.

- **OS Command.** Specify a shell command or script to `EXECUTE` on one or more segments, the output of which will comprise the web table’s data at the time of access. A web table defined with an `EXECUTE` clause will execute the given OS shell command or script on the specified segment host or hosts. By default, the command is executed by all active segment instances on all segment hosts. For example, if each segment host has four primary segment instances running, the command will be executed four times per segment host. You can optionally limit the number of segment instances that execute the web table command.

Web table data is comprised of the output of the command at the time the web table statement is executed. All segment instances included in the web table definition (as specified by the `ON` clause) execute the command in parallel.

Defining Command-Based Web External Tables

The command or program specified in your external table definition must reside on all of the Greenplum segment hosts. If you use environment variables in external web table commands (such as `$PATH`), keep in mind that the command is executed from within the database and not from a login shell. Therefore the `bashrc` or `profile` of the current user will not be sourced. However, you can set desired environment variables from within the `EXECUTE` clause of your external web table definition, for example:

```sql
=# CREATE EXTERNAL WEB TABLE output (output text)
   EXECUTE 'PATH=/home/gpadmin/programs; export PATH;
   myprogram.sh'
   FORMAT 'TEXT';
```

Also, any scripts that you want to execute must be present in the same location on the segment hosts and be executable by the `gpadmin` user.
For example, here is a command that defines a web table that executes a script once per segment host:

```
#$ CREATE EXTERNAL WEB TABLE log_output
    (linenum int, message text)
    EXECUTE '/var/load_scripts/get_log_data.sh' ON HOST
    FORMAT 'TEXT' (DELIMITER '|');
```

### Defining URL-Based Web External Tables

A URL-based web table is similar to a file-based external table except that the data is accessed from a web server via the HTTP protocol. Also, web table data is assumed to be dynamic, so unlike file-based external tables, the data is not rescannable.

For example, here is a command that defines a web table that gets data from several different URLs:

```
#$ CREATE EXTERNAL WEB TABLE ext_expenses (name text,
    date date, amount float4, category text, description text)
    LOCATION ( 'http://intranet.company.com/expenses/sales/file.csv',
                'http://intranet.company.com/expenses/exec/file.csv',
                'http://intranet.company.com/expenses/finance/file.csv',
                'http://intranet.company.com/expenses/ops/file.csv',
                'http://intranet.company.com/expenses/marketing/file.csv',
                'http://intranet.company.com/expenses/eng/file.csv',
            )
    FORMAT 'CSV' ( HEADER );
```

### Loading Data Using an External Table

Once you have defined the external table, and have placed the necessary data files in the correct location (and have started the Greenplum files server(s) if using the `gpfdist` protocol), you can then select from the external table as you would an ordinary database table. For example, if you wanted to load a portion of the external table data into a database table, you could do something like:

```
#$ INSERT INTO expenses_travel
    SELECT * from ext_expenses where category='travel';
```

Or if you wanted to quickly load all data into a new database table:

```
#$ CREATE TABLE expenses AS SELECT * from ext_expenses;
```

### Handling Load Errors

The most common use of external tables is selecting data from them to load into regular database tables. This is typically done by issuing a `CREATE TABLE AS SELECT` or `INSERT INTO SELECT` command, where the `SELECT` statement queries external table data. By default, if the external table data contains an error, the entire command fails and no data is loaded into the target database table. To isolate data errors in external table data while still loading correctly formatted rows, you can
define an external table with a \texttt{SEGMENT REJECT LIMIT} clause at \texttt{CREATE EXTERNAL TABLE} time. If you would like to keep error rows for further examination, you can optionally declare an error table using the \texttt{LOG ERRORS INTO} clause.

### Defining an External Table with Single Row Error Isolation

You can specify the number of error rows acceptable (on a per-segment basis), after which the entire external table operation will be aborted and no rows will be processed or loaded. Note that the count of error rows is per-segment, not per entire operation. If the per-segment reject limit is not reached, then all rows not containing an error will be processed. If the limit is not reached, all good rows will be processed and any error rows discarded. If you would like to keep error rows for further examination, you can optionally declare an error table using the \texttt{LOG ERRORS INTO} clause. Any rows containing a format error would then be logged to the specified error table.

For example:

```sql
=# CREATE EXTERNAL TABLE ext_expenses ( name text, date date, amount float4, category text, desc text ) LOCATION ('gpfdist://etlhost-1:8081/*', 'gpfdist://etlhost-2:8082/*') FORMAT 'TEXT' (DELIMITER '|') LOG ERRORS INTO err_expenses SEGMENT REJECT LIMIT 10 ROWS;
```

When \texttt{SEGMENT REJECT LIMIT} is used, then the external data will be scanned in single row error isolation mode. This can be helpful in isolating errors when loading data from an external table using \texttt{CREATE TABLE AS SELECT} or \texttt{INSERT INTO SELECT}. Single row error isolation mode only applies to external data rows with format errors — for example, extra or missing attributes, attributes of a wrong data type, or invalid client encoding sequences. Constraint errors are not checked, however it is easy to filter out constraint errors when using external tables by limiting the \texttt{SELECT} from an external table at runtime. For example, to eliminate duplicate key errors:

```sql
=# INSERT INTO table_with_pkeys
 SELECT DISTINCT * FROM external_table;
```

### Viewing Bad Rows in the Error Table

If you are using the single row error isolation feature (see “Defining an External Table with Single Row Error Isolation” on page 134 or “Running COPY in Single Row Error Isolation Mode” on page 137), any rows with formatting errors are logged into an error table. The error table has the following columns:

**Table 12.1  Error Table Format**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>cmdtime</td>
<td>timestampz</td>
<td>Timestamp when the error occurred.</td>
</tr>
<tr>
<td>relname</td>
<td>text</td>
<td>The name of the external table that was scanned or the target table of a COPY command.</td>
</tr>
<tr>
<td>filename</td>
<td>text</td>
<td>The name of the load file that contained the error.</td>
</tr>
</tbody>
</table>
You can use SQL commands to query the error table and view the rows that were not loaded. For example:

```sql
=# SELECT * from err_expenses;
```

### Identifying Invalid CSV Files in Error Table Data

If a comma separated values (CSV) file has invalid formatting, the `rawdata` field in the error table may contain several combined rows. For example, if a closing quote for a specific field is missing, all the following newlines are treated as embedded newlines. When this happens, Greenplum stops parsing a row when it reaches 64K, puts that 64K of data into the error table as a single row, resets the quote flag, and continues. If this happens three times during load processing, the load file is determined to be invalid and the entire load fails with the message “rejected N or more rows”. See “Escaping in CSV Formatted Files” on page 127 for more information on the correct use of quotes in CSV files.
Loading Data with gpload

gpload is a Greenplum utility that can be used to load data using readable external tables and the Greenplum parallel file server (gpfdist). It facilitates some of the maintenance tasks of setting up parallel file-based external tables by allowing users to configure their data format, external table definition and gpfdist setup into one easy-to-manage configuration file.

To use gpload

1. First, make sure your environment is set up to run gpload. You’ll need some dependent files from your Greenplum Database installation such as gpfdist and Python, as well as network access to the Greenplum segment hosts. See “Prerequisites” on page 641 for details.

2. Create your load control file. This is a YAML-formatted file that specifies the Greenplum Database connection information, gpfdist configuration information, external table options, and data format. See “Control File Format” on page 643 for details. For example:

```yaml
---
VERSION: 1.0.0.1
DATABASE: ops
USER: gpadmin
HOST: mdw-1
PORT: 5432
GPLOAD:
  INPUT:
    - SOURCE:
      LOCAL_HOSTNAME:
        - etl1-1
        - etl1-2
        - etl1-3
        - etl1-4
      PORT: 8081
    FILE:
      - /var/load/data/*
    COLUMNS:
      - name: text
      - amount: float4
      - category: text
      - desc: text
      - date: date
    FORMAT: text
    DELIMITER: '|' 
    ERROR_LIMIT: 25
    ERROR_TABLE: payables.err_expenses
  OUTPUT:
    - TABLE: payables.expenses
    - MODE: INSERT
```
SQL:
- BEFORE: "INSERT INTO audit VALUES('start', current_timestamp)"
- AFTER: "INSERT INTO audit VALUES('end', current_timestamp)"

3. Run `gpload` passing in the load control file. For example:

```bash
gpload -f my_load.yml
```

---

### Loading Data with COPY

`COPY FROM` copies data from a file (or standard input) into a table (appending the data to whatever is in the table already). If copying data from a file, the file must be accessible to the master host and the name must be specified from the viewpoint of the master host. When `STDIN` or `STDOUT` is specified, data is transmitted via the connection between the client and the master server. `COPY` is non-parallel, meaning that data is loaded in a single process via the Greenplum master instance.

To maximize the performance and throughput of `COPY`, consider running multiple `COPY` commands concurrently in separate sessions and dividing the data to be loaded evenly across all concurrent processes. To maximize throughput, run one concurrent `COPY` operation per CPU.

### Running COPY in Single Row Error Isolation Mode

By default, `COPY` stops an operation at the first error, meaning if the data being loaded contains an error, the entire operation fails and no data is loaded. Optionally, a `COPY FROM` command can be run in single row error isolation mode. In this mode, rows containing format errors — for example, extra or missing attributes, attributes of a wrong data type, or invalid client encoding sequences — can be skipped over while still loading all properly formatted rows. In this release, single row error isolation mode only applies to rows in the input file with format errors. Constraint errors such as violation of a `NOT NULL`, `CHECK`, or `UNIQUE` constraint will still be handled in ‘all-or-nothing’ input mode.

The `SEGMENT REJECT LIMIT` clause when added to a `COPY FROM` operation will run the command in single row error isolation mode. The user can specify the number of error rows acceptable (on a per-segment basis), after which the entire `COPY FROM` operation will be aborted and no rows will be loaded. Note that the count of error rows is per Greenplum segment, not per entire load operation. If the per-segment reject limit is not reached, then all rows not containing an error will be loaded. If the limit is not reached, all good rows will be loaded and any error rows discarded. If you would like to keep error rows for further examination, you can optionally declare an error table using the `LOG ERRORS INTO` clause. Any rows containing a format error would then be logged to the specified error table. For example:

```sql
=> COPY country FROM '/data/gpdb/country_data'
    WITH DELIMITER '|' LOG ERRORS INTO err_country
    SEGMENT REJECT LIMIT 10 ROWS;
```

See also “Viewing Bad Rows in the Error Table” on page 134 for information on investigating rows that were not loaded.
Data Loading Performance Tips

Drop Indexes Before Loading — If you are loading a freshly created table, the fastest way is to create the table, load the data, then create any indexes needed for the table. Creating an index on pre-existing data is quicker than updating it incrementally as each row is loaded. If you are adding large amounts of data to an existing table, it may be a faster to drop the index, load the table, and then recreate the index. Temporarily increasing the `maintenance_work_mem` server configuration parameter will help to speed up `CREATE INDEX` commands, although it will not help performance of the load itself. Dropping and recreating indexes should be done when there are no users on the system.

Run ANALYZE After Loads — Whenever you have significantly altered the data in a table, running `ANALYZE` is strongly recommended. Running `ANALYZE` (or `VACUUM ANALYZE`) ensures that the query planner has up-to-date statistics about the table. With no statistics or obsolete statistics, the planner may make poor decisions during query planning, leading to poor performance on any tables with inaccurate or nonexistent statistics.

Run VACUUM After Load Errors — If not running in single row error isolation mode, a load operation stops at the first encountered error. The target table will already have received earlier rows before the error occurred. These rows will not be visible or accessible, but they still occupy disk space. This may amount to a considerable amount of wasted disk space if the failure happened well into a large load operation. Invoking the `VACUUM` command will recover the wasted space.

Unloading Data from Greenplum Database

This section describes how unload data from Greenplum Database using both parallel unload (writable external tables) and non-parallel unload (`COPY`).

- “Defining a File-Based Writable External Table” on page 138
- “Defining a Command-Based Writable External Web Table” on page 139
- “Unloading Data Using a Writable External Table” on page 139
- “Unloading Data Using COPY” on page 140

Defining a File-Based Writable External Table

Writable external tables that output data to files use the Greenplum parallel file server program (`gpfdist`), just as file-based readable external tables do. With a writable external table, the Greenplum segments send their output rows to the `gpfdist` process, which then writes out the data to the named file.

Use the `CREATE WRITABLE EXTERNAL TABLE` command to define the external table and specify the location and format of the output files. File-based writable external tables must use the Greenplum parallel file server (`gpfdist`) to output files. See “Using the Greenplum Parallel File Server (gpfdist)” on page 129 for instructions on setting up `gpfdist` for use with an external table.
Unlike readable external tables, writable external tables also have the option to declare a distribution policy. By default, writable external tables use a random distribution policy. If the source table you are exporting data from has a hash distribution policy, defining the same distribution key column(s) for the writable external table will improve unload performance by eliminating the need to move rows over the interconnect. If you are unloading data from a particular table, you can use the LIKE clause to copy the column definitions and distribution policy from the source table. For example:

```sql
=# CREATE WRITABLE EXTERNAL TABLE unload_expenses
   ( LIKE expenses )
   LOCATION ('gpfdist://etlhost-1:8081/expenses1.out',
             'gpfdist://etlhost-2:8081/expenses2.out')
   FORMAT 'TEXT' (DELIMITER ',')
   DISTRIBUTED BY (exp_id);
```

**Defining a Command-Based Writable External Web Table**

Similar to readable web external tables that execute a command or program, writable external web tables can also be defined to send output rows to an executable program or script. The executable program used in the writable external table definition must be prepared to accept an input stream, reside in the same location on all of the Greenplum segment hosts, and be executable by the `gpadmin` user. The command specified in your writable external table definition is executed by all segments in the Greenplum system, whether or not a segment has output rows to process or not.

Use the `CREATE WRITABLE EXTERNAL WEB TABLE` command to define the external table and specify the executable command or program to run on the segment hosts. If you use environment variables in the executable command (such as `$PATH`), keep in mind that the command is executed from within the database and not from a login shell. Therefore the `.bashrc` or `.profile` of the current user will not be sourced. However, you can set desired environment variables from within the `EXECUTE` clause of your writable external table definition, for example:

```sql
=# CREATE WRITABLE EXTERNAL WEB TABLE output (output text)
   EXECUTE 'export PATH=$PATH:/home/gpadmin/programs;
             myprogram.sh'
   FORMAT 'TEXT'
   DISTRIBUTED RANDOMLY;
```

**Unloading Data Using a Writable External Table**

Because writable external tables only allow `INSERT` operations, users who are not the table owner or a superuser must be granted `INSERT` permissions on the table in order to use it. For example:

```sql
GRANT INSERT ON writable_ext_table TO admin;
```

To unload data using a writable external table, select the data from the source table(s) and insert it into the writable external table. The rows produced by the `SELECT` statement will be output to the writable external table. For example:
INSERT INTO writable_ext_table SELECT * FROM regular_table;

Unloading Data Using COPY

COPY TO copies data from a table to a file (or standard input) on the Greenplum master host. COPY is non-parallel, meaning that data is unloaded in a single process via the Greenplum master instance. You can use COPY to output a table’s entire contents, or filter the output using a SELECT statement. For example:

COPY (SELECT * FROM country WHERE country_name LIKE 'A%') TO '/home/gpadmin/a_list_countries.out';
Section IV: System Administration

This section describes how to monitor and administer a Greenplum Database system.

- Starting and Stopping Greenplum
- Configuring Your Greenplum System
- Enabling High Availability Features
- Backing Up and Restoring Databases
- Expanding a Greenplum System
- Monitoring a Greenplum System
- Routine System Maintenance Tasks
Starting and Stopping Greenplum

This chapter describes how to start, stop, or restart a Greenplum Database system. This chapter contains the following topics:

- Overview
- Starting Greenplum Database
- Stopping Greenplum Database

Overview

Because a Greenplum Database system is distributed across many machines, the process for starting and stopping a Greenplum Database DBMS is different than a regular PostgreSQL DBMS. With a regular PostgreSQL DBMS, you run a utility called pg_ctl, which starts, stops, or restarts the database server process (postgres). pg_ctl also takes care of tasks such as redirecting log output and properly detaching from the terminal and process group.

In a Greenplum Database DBMS, each database server instance (the master and all segments) must be started or stopped across all of the hosts in the system in such a way that they can all work together as a unified DBMS.

Greenplum provides similar functionality of the pg_ctl utility with the gpstart and gpstop utilities, which are located in $GPHOME/bin of your Greenplum Database master host installation.

Starting Greenplum Database

The database initialization process (gpinitsystem) will start your Greenplum Database system for you once it has completed successfully. However, there may be times when you will need to restart the system, for example to reset operational configuration parameters or to troubleshoot a failed segment.

Use the gpstart utility to start your Greenplum Database system. This utility starts all of the postgres database listener processes in the system (the master and all of the segment instances). gpstart is always run on the master host.

Each instance is started in parallel (up to 60 parallel processes by default).

To start Greenplum Database

$ gpstart

Restarting Greenplum Database

The gpstop utility has a restart option that will restart the Greenplum Database system once a successful shutdown has completed.
To restart Greenplum Database
$ gpstop -r

Uploading Configuration File Changes Only
The gpstop utility has an option that will upload changes made to the pg_hba.conf configuration file and to runtime parameters in the master postgresql.conf file without interruption of service. Note that any active sessions will not pickup the changes until they reconnect to the database, and that many server configuration parameters require a full system restart (gpstop -r) to be activated. For more information, see “Server Configuration Parameters” on page 770.

To upload runtime configuration file changes without restarting
$ gpstop -u

Starting the Master in Maintenance Mode
There may be cases where you want to start only the master. This is called maintenance mode. In this mode, you can do things such as connect to a database on the master instance only in utility mode and edit settings in the system catalog, without affecting user data on the segment instances. See “System Catalog Reference” on page 825 for more information about the system catalog tables.

To start the master in utility mode
1. Run gpstart using the -m option:
   $ gpstart -m
2. Connect to the master in utility mode to do catalog maintenance. For example:
   $ PGOPTIONS='"c gp_session_role=utility"' psql template1
3. After completing your administrative tasks, you must stop the master in utility mode before you can restart it again in production mode.
   $ gpstop -m

Stopping Greenplum Database
Use the gpstop utility to stop or restart your Greenplum Database system. This utility stops all of the postgres processes in the system (the master and all of the segment instances). gpstop is always run on the master host.

Each instance is stopped in parallel (up to 60 parallel processes by default). By default the system will wait for any active transactions to finish before shutting down, and will not shutdown if the database has any active client connections.

To stop Greenplum Database
$ gpstop
To stop Greenplum Database in fast mode
(all active transactions are interrupted and rolled back, all active client sessions are cancelled)

$ gpstop -M fast
14. Configuring Your Greenplum System

There are many server configuration parameters that affect the behavior of a Greenplum Database system. Most of these parameters are the same as in regular PostgreSQL, but there are also several Greenplum-specific configuration parameters.

- About Greenplum Master and Local Parameters
- Setting Configuration Parameters
- Configuration Parameter Categories

About Greenplum Master and Local Parameters

In most database management systems, you have a server configuration file that configures various aspects of the server. In Greenplum Database (as in PostgreSQL) this file is called `postgresql.conf`. This configuration file is located in the data directory of the database instance.

In Greenplum Database, the master and each segment instance has its own `postgresql.conf` file (located in their respective data directories). Some parameters are considered local parameters, meaning that each segment instance looks to its own `postgresql.conf` file to get the value of that parameter. You must set local parameters on every instance in the system (master and segments).

Others parameters are considered master parameters. Master parameters need only be set at the master instance, and the value is passed down to (or in some cases ignored by) the segment instances at query run time.

See “Server Configuration Parameters” on page 770 to see if a particular parameter is classified as local or master.

Setting Configuration Parameters

Many of the configuration parameters have limitations on who can change them and where or when they can be set. For example, to change certain parameters, you must be a Greenplum Database superuser. Other parameters can only be set at the system-level in the `postgresql.conf` file or require a restart of the system for the changes to take effect.

Many configuration parameters are considered session parameters. A session parameter can be set at the system-level, the database-level, the role-level or the session-level. Most session parameters can be changed by any database user within their session, but a few may require superuser permissions. See “Server Configuration Parameters” on page 770 to see the set classifications of particular parameters.
Setting a Local Configuration Parameter

If you want to change a configuration parameter that is classified as local, you must change it in every postgresql.conf file where you want that change to take effect. In most cases, this means making the change at the master and at every segment (primary and mirror). To set a parameter in all postgresql.conf files of a Greenplum system, you can use the gpconfig utility. For example:

$ gpconfig -c gp_vmem_protect_limit -v 4096MB

Then restart Greenplum Database to make the configuration changes effective:

$ gpstop -r

Setting a Master Configuration Parameter

If a parameter is classified as master, you only need to set it at the Greenplum master instance. If it is also a session parameter, you have the additional flexibility of setting the parameter for a particular database, role or session. If a parameter is set at multiple levels, then the more granular level takes precedence. For example, session overrides role, role overrides database, and database overrides system.

Setting Parameters at the System Level

Setting master parameters in the master postgresql.conf file makes that setting the new system-wide default.

1. Edit the $MASTER_DATA_DIRECTORY/postgresql.conf file.

2. Find the parameter you want to change, uncomment it (remove the preceding # character), and set it to the desired value.

3. Save and close the file.

4. For session parameters that do not require a server restart, you can upload the postgresql.conf changes as follows:

   $ gpstop -u

5. For parameter changes that require a server restart, restart Greenplum Database as follows:

   $ gpstop -r

Note: See “Server Configuration Parameters” on page 770 for detailed descriptions of the server configuration parameters.

Setting Parameters at the Database Level

When a session parameter is set at the database level, every session that connects to that database will pick up that parameter setting. Settings at the database level override those at the system level. Use the ALTER DATABASE command to set a parameter at the database level. For example:

=# ALTER DATABASE mydatabase SET search_path TO myschema;
Setting Parameters at the Role Level
When a session parameter is set at the role level, every session initiated by that role will pick up that parameter setting. Settings at the role level override those at the database level. Use the ALTER ROLE command to set a parameter at the role level. For example:

```sql
=# ALTER ROLE bob SET search_path TO bobschema;
```

Setting Parameters in a Session
Any session parameter can also be set in an active database session using the SET command. That parameter setting is then valid for the rest of that session (or until a RESET command is issued). Settings at the session level override those at the role level. For example:

```sql
=# SET work_mem TO '200MB';
=# RESET work_mem;
```

Viewing Settings of Server Configuration Parameters
The SHOW SQL command allows you to see the settings of the server configuration parameters used by the Greenplum Database system. For example, to see the settings for all parameters:

```
$ psql -c 'SHOW ALL;'
```

Running SHOW will show the settings for the master instance only. If you want to see the value of a particular parameter across the entire system (master and all segments), you can use the gpconfig utility. For example:

```
$ gpconfig --show max_connections
```

Configuration Parameter Categories
There are many configuration parameters that affect the behavior of a Greenplum Database system. This section explains the general categories of configuration parameters. See “Server Configuration Parameters” on page 770 to see the details of particular parameters.

- Connection and Authentication Parameters
- System Resource Consumption Parameters
- Query Tuning Parameters
- Error Reporting and Logging Parameters
- System Monitoring Parameters
- Runtime Statistics Collection Parameters
- Automatic Statistics Collection Parameters
- Client Connection Default Parameters
- Client Connection Default Parameters
Connection and Authentication Parameters
These parameters control how clients connect and authenticate to Greenplum Database. See also, “Configuring Client Authentication” on page 36.

### Connection Parameters
- `gp_vmem_idle_resource_timeout`
- `listen_addresses`
- `max_connections`
- `max_prepared_transactions`
- `port` (see “Changing the Master Port”)
- `superuser_reserved_connections`
- `tcp_keepalives_count`
- `tcp_keepalives_idle`
- `tcp_keepalives_interval`
- `unix_socket_directory`
- `unix_socket_group`
- `unix_socket_permissions`

### Changing the Master Port
Greenplum Database comes configured with the default master port of 5432. Administrators should follow this process to change the master port number. Perform these steps on your Greenplum master host.

1. Shut down your Greenplum system in fast mode, interrupting and rolling back any active transactions. There cannot be active queries in the system when the system configuration is being altered.
   ```bash
   $ gpstop -M fast
   ```

2. Edit the `postgresql.conf` file of the master. For example:
   ```bash
   $ vi $MASTER_DATA_DIRECTORY/postgresql.conf
   ```

3. Change the following parameter to the new port number. For example if the new port is 54321:
   ```bash
   port=54321
   ```
   **NOTE:** If you have a standby master host, you must also edit the `$MASTER_DATA_DIRECTORY/postgresql.conf` file on your standby master host. After you are finished, return to your primary master host to complete the process.

4. Restart your Greenplum Database system.
   ```bash
   $ gpstart
   ```

5. Start a `psql` client session using the new port and edit the `gp_segment_configuration` system catalog to change the master port. For example, if the new port is 54321 and the old port is 5432:
psql -p 54321 -c "UPDATE gp_segment_configuration SET port=54321 WHERE port=5432"

6. Make sure that any client programs you are using are updated to connect to the new master port. For example, if using psql update the $PGPORT environment variable in the profile of your Greenplum superuser (gpadmin).

Security and Authentication Parameters

- authentication_timeout
- db_user_namespace
- krb_caseins_users
- krb_server_keyfile
- krb_srvname
- password_encryption
- ssl
- ssl_ciphers

System Resource Consumption Parameters

Memory Consumption Parameters

These parameters control system memory usage. The most commonly used tuning parameter is work_mem, but you may also want to adjust gp_vmem_protect_limit and max_work_mem to avoid running out of memory at the segment hosts during query processing.

- gp_vmem_idle_resource_timeout
- gp_vmem_protect_limit
- gp_vmem_protect_segworker_cache_limit
- maintenance_work_mem
- max_appendonly_tables
- max_prepared_transactions
- max_stack_depth
- max_work_mem
- shared_buffers
- temp_buffers
- work_mem

Free Space Map Parameters

These parameters control the sizing of the free space map where expired rows are held. Disk space in the free space map is reclaimed by a VACUUM command. See also, “Vacuuming the Database” on page 100.

- max_fsm_pages
- max_fsm_relations

OS Resource Parameters

- max_files_per_process
- shared_preload_libraries

Cost-Based Vacuum Delay Parameters

Standard PostgreSQL has a number of parameters for configuring the execution cost of VACUUM and ANALYZE commands. The intent of this feature is to allow administrators to reduce the I/O impact of these commands on concurrent database
activity. When the accumulated cost of the various I/O operations reaches the limit, the process performing the operation will sleep for a while. Then it will reset the counter and continue execution.

**Warning:** Cost-based vacuum delay is not recommended in Greenplum Database because it runs asynchronously between the segment instances. The vacuum cost limit and delay is invoked locally at the segment level without taking into account the state of the entire Greenplum array.

- vacuum_cost_delay
- vacuum_cost_limit
- vacuum_cost_page_dirty
- vacuum_cost_page_hit
- vacuum_cost_page_miss

### Query Tuning Parameters

### Query Plan Operator Control Parameters

The following parameters control the types of plan operations the query planner has to choose from. Enabling or disabling certain plan operations is a way to force the planner to choose a different plan. These can be useful in testing a query using different plan types to see which plan offers the best performance.

- enable_bitmapscan
- enable_groupagg
- enable_hashagg
- enable_hashjoin
- enable_indexscan
- enable_mergejoin
- enable_nestloop
- enable_seqscan
- enable_sort
- enable_tidscan
- gp_enable_adaptive_nestloop
- gp_enable_agg_distinct

### Query Planner Costing Parameters

**Note:** Greenplum recommends that you do not adjust these query costing parameters. They have already been tuned to reflect Greenplum Database hardware configurations and typical workloads. All of these parameters are related. Changing one without changing the others can have adverse affects on performance.
Database Statistics Sampling Parameters
These parameters adjust the amount of data sampled by an *ANALYZE* operation. Adjusting these parameters will affect statistics collection system-wide. You can also configure statistics collection on particular tables and columns by using the *ALTER TABLE SET STATISTICS* clause.

- `default_statistics_target`
- `gp_analyze_relative_error`

Sort Operator Configuration Parameters
- `gp_enable_sort_distinct`
- `gp_enable_sort_limit`

Aggregate Operator Configuration Parameters
- `gp_enable_agg_distinct`
- `gp_enable_agg_distinct_pruning`
- `gp_enable_multiphase_agg`
- `gp_enable_preunique`
- `gp_enable_grouplex_distinct_gather`
- `gp_enable_grouplex_distinct_pruning`
- `gp_workfile_compress_algorithm`

Join Operator Configuration Parameters
- `join_collapse_limit`
- `gp_adjust_selectivity_for_outerjoins`
- `gp_hashjoin_tuples_per_bucket`
- `gp_statistics_use_fkeys`
- `gp_workfile_compress_algorithm`

Other Query Planner Configuration Parameters
- `fromCollapseLimit`
- `gp_enable_predicate_propagation`
- `gp_statistics_pullup_from_child_partition`
Error Reporting and Logging Parameters

Log Rotation

- `log_rotation_age`
- `log_rotation_size`
- `log_truncate_on_rotation`

When to Log

- `client_min_messages`
- `log_error_verbosity`
- `log_min_duration_statement`
- `log_min_error_statement`
- `log_min_messages`

What to Log

- `debug_pretty_print`
- `debug_print_parse`
- `debug_print_plan`
- `debug_print_prelim_plan`
- `debug_print_rewritten`
- `debug_print_slice_table`
- `log_autostats`
- `log_connections`
- `log_disconnections`
- `log_dispatch_stats`
- `log_duration`
- `log_executor_stats`
- `log_hostname`
- `log_parser_stats`
- `log_planner_stats`
- `log_statement`
- `log_statement_stats`
- `log_timezone`
- `gp_debug_linger`
- `gp_log_format`
- `gp_max_csv_line_length`
- `gp_reraise_signal`

System Monitoring Parameters

SNMP Alerts

The following parameters are used to send SNMP notifications when events occur on a Greenplum Database system.

- `gp_snmp_community`
- `gp_snmp_monitor_address`
- `gp_snmp_use_inform_or_trap`
Email Alerts
The following parameters are used to configure the system to send email alerts for fatal error events, such as a segment going down or a server crash and reset.

- gp_email_from
- gp_email_smtp_password
- gp_email_smtp_server
- gp_email_smtp_userid
- gp_email_to

Greenplum Performance Monitor Agent
The following parameters are used to configure the data collection agents for the Greenplum Performance Monitor.

- gp_enable_gpperfmon
- gp_gpperfmon_send_interval
- gpperfmon_port

Runtime Statistics Collection Parameters
These parameters control the PostgreSQL server statistics collection feature. When statistics collection is enabled, the data that is produced can be accessed via the pg_stat and pg_statio family of system catalog views.

- stats_queue_level
- track_activities
- track_counts
- update_process_title

Automatic Statistics Collection Parameters
When automatic statistics collection is enabled, ANALYZE can be run automatically in the same transaction as an INSERT, UPDATE, DELETE, COPY or CREATE TABLE...AS SELECT statement when a certain threshold of rows is affected (on_change), or when a newly generated table has no statistics (on_no_stats). To enable this feature, set the following server configuration parameters in your Greenplum master postgresql.conf file and restart Greenplum Database:

- gp_autostats_mode
- gp_autostats_on_change_threshold
- log_autostats

Warning: Depending on the specific nature of your database operations, automatic statistics collection may have a negative performance impact. Carefully evaluate whether the default setting of on_no_stats is appropriate for your system.
Client Connection Default Parameters

Statement Behavior Parameters

- check_function_bodies
- default_tablespace
- default_transaction_isolation
- default_transaction_read_only
- search_path
- statement_timeout
- vacuum_freeze_min_age

Locale and Formatting Parameters

- client_encoding
- DateStyle
- extra_float_digits
- IntervalStyle
- lc_collate
- lc_ctype
- lc_messages
- lc_monetary
- lc_numeric
- lc_time
- TimeZone

Other Client Default Parameters

- dynamic_library_path
- explain_pretty_print
- local_preload_libraries

Lock Management Parameters

- deadlock_timeout
- max_locks_per_transaction

Workload Management Parameters

The following configuration parameters are used to configure the Greenplum Database workload management feature (resource queues), query prioritization, memory utilization and concurrency control.

- gp_resqueue_priority
- gp_resqueue_priority_cpucores_per_segment
- gp_resqueue_priority_sweeper_interval
- gp_vmem_idle_resource_timeout
- gp_vmem_protect_limit
- gp_vmem_protect_segworker_cache_limit
- max_resource_queues
- max_resource_portals_per_transaction
- resource_cleanup_gangs_on_wait
- resource_select_only
- stats_queue_level
External Table Parameters
The following parameters are used to configure the external tables feature of Greenplum Database. See “About External Tables” on page 124.

- `gp_external_enable_exec`
- `gp_external_grant_privileges`
- `gp_external_max_segs`
- `gp_reject_percent_threshold`

Append-Only Table Parameters
The following parameters are used to configure the append-only tables feature of Greenplum Database. See “Choosing the Table Storage Model” on page 71 for more information about append-only tables.

- `max_appendonly_tables`

Database and Tablespace/Filespace Parameters
The following parameters are used to configure the maximum number of databases, tablespaces, and filespaces allowed in a system.

- `gp_max_tablespaces`
- `gp_max_filespaces`
- `gp_max_databases`

Past PostgreSQL Version Compatibility Parameters
The following parameters are for compatibility with older PostgreSQL versions. Most likely, you do not need to change these in Greenplum Database.

- `add_missing_from`
- `array_nulls`
- `backslash_quote`
- `escape_string_warning`
- `regex_flavor`
- `standard_conforming_strings`
- `transform_null_equals`

Greenplum Array Configuration Parameters
The parameters in this section control the configuration of the Greenplum Database array and its various components (segments, master, distributed transaction manager, and interconnect).
Interconnect Configuration Parameters

- gp_interconnect_hash_multiplier
- gp_interconnect_queue_depth
- gp_interconnect_setup_timeout
- gp_interconnect_type
- gp_max_packet_size

Dispatch Configuration Parameters

- gp_cached_segworkers_threshold
- gp_connections_per_thread
- gp_enable_direct_dispatch
- gp_segment_connect_timeout
- gp_set_proc_affinity

Fault Operation Parameters

- gp_set_read_only
- gp_fts_probe_interval
- gp_fts_probe_threadcount

Distributed Transaction Management Parameters

- gp_max_local_distributed_cache

Read-Only Parameters

- gp_command_count
- gp_content
- gp_dbid
- gp_num_contents_in_cluster
- gp_role
- gp_session_id
15. Enabling High Availability Features

This chapter describes the high availability features of Greenplum Database, and explains the tasks involved for segment or master instance recovery. The following topics are covered in this chapter:

- Overview of High Availability in Greenplum Database
- Enabling Mirroring in Greenplum Database
- Knowing When a Segment is Down
- Recovering a Failed Segment
- Recovering a Failed Master

Overview of High Availability in Greenplum Database

Greenplum Database provides several optional features to ensure maximum uptime and high-availability of your system. This section provides an overview of these features:

- Overview of Segment Mirroring
- Overview of Master Mirroring
- Overview of Fault Detection and Recovery

Overview of Segment Mirroring

Mirror segments allow database queries to fail over to a backup segment if the primary segment is unavailable. To configure mirroring, you must have enough nodes in your Greenplum Database system so that the mirror segment always resides on a different host than its primary.
The mirror segment for a distributed table resides on a different host than its primary segment. During database operations, only the primary segment is active. Changes to a primary segment are copied over to its mirror using a file block replication process. Until a failure occurs on the primary segment, there is no live segment instance running on the mirror host -- only the replication process.

![Data Mirroring in Greenplum Database](image)

**Figure 15.1  Data Mirroring in Greenplum Database**

In the event of a segment failure, the file replication process is stopped and the mirror segment is automatically brought up as the active segment instance. All database operations then continue using the mirror. While the mirror is active, it is also logging all transactional changes made to the database. This system state is known as *Change Tracking* mode. When the failed segment is ready to be brought back online, administrators initiate a recovery process to bring it back into operation. The recovery process synchronizes with the mirror and only copies over the changes that were missed while the segment was down. This system state is known as *Resynchronizing* mode. Once all mirrors and their primaries are synchronized again, the system state becomes *Synchronized*.

**Overview of Master Mirroring**

You can also optionally deploy a backup or ‘mirror’ of the master instance on a separate host machine. A backup master (or *standby master*) host serves as a ‘warm standby’ in the event of the primary master host becoming unoperational.

The standby master is kept up to date by a replication process (*gpsyncagent*), which runs on the standby master host and keeps the data between the primary and standby masters synchronized. Until a failure occurs in the primary master, no actual Greenplum Database master server is running on the standby master host -- only the replication process.

If the primary master fails, the replication process is shut down, and the standby master can be activated in its place. Upon activation of the standby master, the replicated logs are used to reconstruct the state of the master host at the time of the last successfully committed transaction. The activated standby then functions as the Greenplum Database master, accepting connections on the same port used for the failed primary master.
Since the master does not contain any user data, only the system catalog tables need to be synchronized between the primary and standby copies. These tables are not updated frequently, but when they are, changes are automatically copied over to the standby master so that it is always kept current with the primary.

![Figure 15.2 Master Mirroring in Greenplum Database](image)

**Overview of Fault Detection and Recovery**

Fault detection is handled by a Greenplum Database server (postgres) subprocess named `ftsprobe`. This fault detection process monitors the Greenplum array, scanning all segments and database processes at configurable intervals.

Whenever the fault detection process cannot connect to a segment, it marks that segment instance as down in the Greenplum Database system catalog. Once a segment is down, it will remain out of operation until an administrator initiates the recovery process to bring that segment back online.

When mirroring is enabled in a Greenplum Database system, the system will automatically failover to the mirror copy whenever a primary copy becomes unavailable. A Greenplum Database system can remain operational if a segment instance or host goes down as long as all portions of data are available on the remaining active segments.

To recover failed segments in the system, the Greenplum administrator runs a recovery utility (`gprecoverseg`). This utility locates the down segments, checks if they are valid, and compares the transactional state with the currently active segment to find out what changes were missed when the segment was offline. It then synchronizes only the changed database files with the active segment and brings the segment back online. This recovery process is performed while the Greenplum Database system is up and running.

If you do not have mirroring enabled, the system will automatically shut down if a segment instance becomes invalid. You must manually recover all failed segments before operations can continue.
Enabling Mirroring in Greenplum Database

You can either configure your Greenplum Database system with mirroring at setup time (using \texttt{gpinitsystem}), or enable mirroring in an existing system that was initially configured without mirroring (using \texttt{gpaddmirrors} and \texttt{gpinitstandby}). This section assumes you are adding mirrors to an existing system that was originally initialized without mirrors.

There are two types of mirroring you can configure:

- Enabling Segment Mirroring
- Enabling Master Mirroring

### Enabling Segment Mirroring

Mirror segments allow database queries to fail over to a backup segment if the primary segment is unavailable. To configure mirroring, you must have enough nodes in your Greenplum Database system so that the mirror segment always resides on a different host than its primary. By default, mirrors are configured on the same array of hosts as your primary segments. You can also optionally choose a completely different set of hosts for your mirror segments so they do not share machines with your primary segments.

#### To add segment mirrors to an existing system (same hosts as primaries)

1. Allocate the data storage area for mirror data on all segment hosts. This must be a different file system location than your primary segments are using.

2. The segment hosts must be able to SSH and SCP to each other without a password prompt. See “\texttt{gpssh-exkeys}” on page 682.

3. Run the \texttt{gpaddmirrors} utility to enable mirroring in your Greenplum Database system. For example (where \texttt{-p} specifies the number to add to your primary segment port numbers to calculate the mirror segment port numbers):

   $$\texttt{gpaddmirrors -p 10000}$$

#### To add segment mirrors to an existing system (different hosts from primaries)

1. Make sure that the Greenplum Database software has been installed on all hosts. See the \textit{Greenplum Database Installation Guide} for detailed instructions on running the Greenplum Database installer.

2. Allocate the data storage area for mirror data on all segment hosts.

3. The segment hosts must be able to SSH and SCP to each other without a password prompt. See “\texttt{gpssh-exkeys}” on page 682.

4. Create a configuration file that lists the host names, ports, and data directories where you want your mirrors to be created. You can create a sample configuration file to use as a starting point by running:

   $$\texttt{gpaddmirrors -o filename}$$
The format of the mirror configuration file is:

```
filespaceOrder=[filespace1_fsname[ :filespace2_fsname: ... ]
mirror[content]=content:address:mir_replication_port:pri_replication_port:fseolocation[:fseolocation: ...]
```

For example (2 segment hosts, 2 segments per host, and you do not have additional filesystems configured besides the default `pg_system` filesystem):

```
filespaceOrder=
mirror0=0:sdw1:sdw1-1:52001:53001:54001:/gpdata/mir1/gp0
mirror1=1:sdw1:sdw1-2:52002:53002:54002:/gpdata/mir1/gp1
mirror2=2:sdw2:sdw2-1:52001:53001:54001:/gpdata/mir1/gp2
```

5. Run the `gpaddmirrors` utility to enable mirroring in your Greenplum Database system (where `-i` names the mirror configuration file you just created):

```
$ gpaddmirrors -i mirror_config_file
```

---

## Enabling Master Mirroring

You can either configure your Greenplum Database system with a standby master at setup time (using `gpinitsystem`), or enable it in an existing system that was initially configured without a standby master (using `gpinitstandby`). This section assumes you are adding a standby master to an existing system that was originally initialized without one.

### To add a standby master to an existing system

1. Make sure the standby master host is installed and configured (`gpadmin` system user created, Greenplum Database binaries installed, environment variables set, ssh keys exchanged, and data directory created). See the **Greenplum Database Installation Guide** for detailed instructions.

2. Run the `gpinitstandby` utility on the currently active primary master host. This will add a standby master host to your Greenplum Database system. For example (where `-s` specifies the standby master host name):

```
$ gpinitstandby -s smdw
```

3. To switch operations to a standby master, see “Recovering a Failed Master” on page 167.

### To check the status of the log synchronization process

If the synchronization process (`gpsyncagent`) fails on the standby master, it may not always be noticeable to users of the system. The `gp_master_mirroring` catalog is a place where Greenplum Database administrators can check to see if the standby master is currently synchronized. For example:

```
$ psql dbname -c 'SELECT * FROM gp_master_mirroring;
```

If the result indicates that the standby master is “Not Synchronized,” check the `detail_state` and `error_message` columns to attempt to determine the cause of the errors.
Knowing When a Segment is Down

If mirroring is enabled, Greenplum Database will automatically failover to a mirror segment when a primary segment goes down. As long as one segment instance is alive per portion of data, it will not typically be apparent to users of the system that a segment instance is down. If a transaction is in progress when a fault occurs, the in progress transaction will roll back, and then automatically restart on the reconfigured set of segments.

If the entire Greenplum Database system becomes unoperational due to a segment failure (for example if mirroring is not enabled or there are not enough segments alive to serve all the portions of user data), users will see errors when trying to connect to a database. The errors returned to the client program may give some indication of the failure. For example:

ERROR: All segment databases are unavailable

Enabling Alerts and Notifications

Greenplum Database administrators can enable email and/or SNMP alerts to be notified on system events such as segment failures. See “Enabling System Alerts and Notifications” on page 197 for more information.

Checking for Failed Segments

With mirroring enabled, you may have failed segments in the system without interruption of service or any indication that a failure has occurred. One way to verify the status of your system is to use the gpstate utility. This utility provides the status of each individual component of a Greenplum Database system (primary segments, mirror segments, master, and standby master).

To check for failed segments

1. On the master, run the gpstate utility with the -e option. This will show any segments with error conditions:

   $ gpstate -e

2. Segments that are in Change Tracking mode indicate that the corresponding mirror segment is down.

   A segment that is not in its preferred role means that the segment is not operating in the role it was assigned at system initialization time. This means that the system is in a potentially unbalanced state, as some segment hosts may have more active segments that is optimal for top system performance. See “To return all segments to their preferred role” on page 165 for instructions on how to fix this situation.

3. To get detailed information about a failed segment, look it up in the gp_segment_configuration catalog table. For example:
$ psql -c "SELECT * FROM gp_segment_configuration WHERE status='d';"

4. For failed segment instances, note the host, port, preferred role, and data directory. This will help you determine the host and segment instances to troubleshoot.

5. To see the primary to mirror segment instance mapping, run:
   $ gpstate -m

Checking the Log Files
The log files may provide more information to help you determine the cause of an error. The master and each segment instance has its own log file, which is located in `pg_log` of its data directory. The master log file contains the most information and should always be checked first.

You can use the `gplogfilter` utility to check the Greenplum Database log files for any additional information. To check the segment log files, you can run `gplogfilter` on the segment hosts using `gpssh`.

To check the log files
1. Check the master log file for WARNING, ERROR, FATAL or PANIC log level messages:
   $ gplogfilter -t

2. Using `gpssh`, check for WARNING, ERROR, FATAL or PANIC log level messages on each segment instance. For example:
   $ gpssh -f seg_hosts_file -e 'source /usr/local/greenplum-db/greenplum_path.sh ; gplogfilter -t /data1/primary/*/pg_log/gpdb*.log' > seglog.out

Recovering a Failed Segment
Whenever the master cannot connect to a segment instance, it marks that segment down in the Greenplum Database system catalog. The segment instance will remain out of operation until steps are taken to bring that segment back online.

The process for recovering a failed segment instance or host depends on the cause of the failure and whether or not you have mirroring enabled. A segment instance can become unavailable for several reasons, such as:

- A segment host is not available (network errors, hardware failures).
- A segment instance is not running (no `postgres` database listener process).
- The data directory of the segment instance is corrupted or missing (data not accessible, corrupted file system, disk failure).

Figure Figure 15.3 shows the high-level steps to take for each of these failure scenarios.
Recovering From Segment Failures

When an entire segment host goes down, this typically causes several segment failures, as all primary and mirror segments on that host will be marked down and taken out of operation. If your Greenplum Database system was deployed without mirroring enabled, the system will automatically become unoperational whenever a segment host goes down.

To recover with mirroring enabled

1. First, make sure that you can connect to the segment host from the master host. For example:

   $ ping failed_seg_host_address
2. Troubleshoot the problem that is preventing the master host from connecting to the segment host. For example, the host machine may need to be restarted or replaced.

3. Once the host is up again, and you have verified that you can connect to it, run the `gprecoverseg` utility to reactivate the failed segment instances on that host. For example:
   
   ```
   $ gprecoverseg
   ```

4. The recovery process will bring up the failed segments and identify the changed files that need to be synchronized. Do not cancel `gprecoverseg` during this time; wait for it to complete. During this transition, database write activity is momentarily suspended.

5. After `gprecoverseg` completes, the system goes into `Resynchronizing` mode and begins copying over the changed files. This process is performed in the background while the system is online and still accepting database requests.

6. When the resynchronization process is complete, the system state will become `Synchronized`. Run the `gpstate` utility to verify the status of the resynchronization process:
   
   ```
   $ gpstate -m
   ```

**To return all segments to their preferred role**

When a primary segment goes down, the mirror is activated and becomes the primary segment. After running `gprecoverseg`, the currently active segment remains the primary and the failed segment is then brought up as the mirror. The segment instances are not returned to the preferred role that they were given at system initialization time. This can leave the system in a potentially unbalanced state, as some segment hosts may still have more active segments that is optimal for top system performance.

1. Run `gpstate -m` to make sure that all mirrors are `Synchronized`.
   
   ```
   $ gpstate -m
   ```

2. If any mirrors are in `Resynchronizing` mode, wait for them to complete.

3. Restart Greenplum Database. A restart will bring up all segments in their preferred role. The Greenplum Database system will be unavailable for a brief period while it restarts.
   
   ```
   $ gpstop -r
   ```

**To recover from a double fault**

A double fault is when a primary and its corresponding mirror are both down at the same time. This can occur in rare situations, such as multiple hardware failures on different segment hosts at the same time. When you have a double fault, Greenplum Database is down and not available for query processing. To recover from a double fault:

1. Restart Greenplum Database:
To recover without mirroring enabled

1. First, make sure that you can connect to the segment host from the master host. For example:
   
   ```bash
   $ ping failed_seg_host_address
   ```

2. Troubleshoot the problem that is preventing the master host from connecting to the segment host. For example, the host machine may need to be restarted.

3. Once the host is up again, and you have verified that you can connect to it, do a restart of the Greenplum Database system. For example:
   
   ```bash
   $ gpstop -r
   ```

4. Run the `gpstate` utility to verify that all segment instances are up:
   
   ```bash
   $ gpstate
   ```

If a segment host is not recoverable and you have lost one or more segments, you will have to recreate your Greenplum Database system from backup files. See “Backing Up and Restoring Databases” on page 170.

When a segment host is not recoverable

If a host is no longer operational (due to hardware failures, for example) you will need to recover the segments on to a spare set of hardware resources. If you have mirroring enabled, you can recover a segment from its mirror copy on to an alternate host using `gprecoverseg`. For example:

```bash
$ gprecoverseg -i recover_config_file
```

Where `recover_config_file` is in the format of:

```
filespaceOrder=[filespace1_name[:filespace2_name:...]]
failed_host_address:port:fselocation[:fselocation:...]
[recovery_host_address:port:replication_port:fselocation[:fselocation:...]]
```

For example (if recovering to a different host than the failed host, and you do not have additional filespaces configured besides the default `pg_system` filesystem):

```
filespaceOrder=
sdw5-2:50002:/gpdata/gpseg2 sdw9-2:50002:53002:/gpdata/gpseg2
```

The `gp_segment_configuration` and `pg_filespace_entry` system catalog tables can help you determine your current segment configuration so that you can plan your mirror recovery configuration. For example, run the following query:
The new recovery segment host must be pre-installed with the Greenplum Database software and configured exactly the same as the existing segment hosts.

### Recovering a Failed Master

If the primary master fails, the log replication process is shut down, and the standby master can be activated in its place by using the `gpactivatestandby` utility. Upon activation of the standby master, the replicated logs are used to reconstruct the state of the master host at the time of the last successfully committed transaction. You can also specify a new standby master host when you activate your currently configured standby master.

#### To activate the standby master

1. First, you must have a standby master host configured for the system. See “Enabling Master Mirroring” on page 161.

2. Run the `gpactivatestandby` utility from the standby master host you are activating. For example, where `-d` specifies the data directory of the master host you are activating:

   ```
   $ gpactivatestandby -d /data/master/gpseg-1
   ```

   Note that once you activate your standby, it then becomes the active or primary master for your Greenplum Database array. If you want to configure another host to be your new standby at this time, you can optionally use the `-c` option when running the `gpactivatestandby` utility. For example:

   ```
   $ gpactivatestandby -d /data/master/gpseg-1 -c new_standby_hostname
   ```

3. After the utility has finished, run `gpstate` to check the status:

   ```
   $ gpstate -f
   ```

   The newly activated master should have `Active` status, and if you configured a new standby host, it should have `Passive` status (if not configured, the status is displayed as `Not Configured`).

4. After switching over, run `ANALYZE` on the newly active master host. For example:

   ```
   $ psql dbname -c 'ANALYZE;'
   ```

5. (optional) If you did not specify a new standby host when activating, you can use `gpinitstandby` to configure a new standby master at a later time. Run this utility on your currently active master host. For example:

   ```
   $ gpinitstandby -s new_standby_master_hostname
   ```
**Restoring Master Mirroring After a Recovery**

After activating a standby master in a recovery scenario and making it your current primary master, you can continue running that instance as your primary master. This assumes that the capabilities and dependability of that host machine are equivalent to the original master host.

Unless you have already done so while activating the prior standby master, you must initialize a new standby master in order to continue to provide master mirroring. Run `gpinitstandby` on the currently active master host to configure a new standby master.

Optionally, you may prefer to restore the primary and standby master instances on the original hosts. This process essentially swaps back the roles of the primary and standby master hosts, and should be performed only if you strongly prefer to run the master instances on the same hosts they occupied prior to the recovery scenario.

**To restore the master and standby instances on original hosts (optional)**

1. Make sure the original master host is in dependable running condition. The conditions that caused the original failure should be fully remedied.

2. Initialize a standby master on the original master host. For example:
   
   ```bash
   $ gpinitstandby -s original_master_hostname
   ```

3. Stop the master process on the current master host (which was the original standby master host). For example:
   
   ```bash
   $ gpstop -m
   ```

4. Run the `gpactivatestandby` utility from the original master host (currently a standby master). For example, where `-d` specifies the data directory of the host you are activating:
   
   ```bash
   $ gpactivatestandby -d $MASTER_DATA_DIRECTORY
   ```

5. After the utility has finished, run `gpstate` to check the status:
   
   ```bash
   $ gpstate -f
   ```

   The original primary master should now have an `Active` status, and the status for the standby master should be `Not Configured`.

6. Once the original master host is again running the primary Greenplum Database master, you can initialize a standby master on the original standby master host. For example:
   
   ```bash
   $ gpinitstandby -s original_standby_master_hostname
   ```

**To resynchronize a standby master**

There may be times when the log synchronization process between the primary and standby master has stopped or has fallen behind, and your standby master is then out of date. Check the `gp_master_mirroring` system catalog table to see when the standby master was last updated. For example:

```bash
$ psql dbname -c 'SELECT * FROM gp_master_mirroring;'
```
To recover a standby master and bring it up to date again, run the following `gpinitstandby` command (using the `-n` option):

$$ gpinitstandby -s \text{standby_master_hostname} -n $$
16. **Backing Up and Restoring Databases**

This chapter provides information on backing up and restoring databases and user data in a Greenplum Database system. It contains the following topics:

- Overview of Backup and Restore Operations
- Backing Up a Database
- Restoring From Parallel Backup Files

**Overview of Backup and Restore Operations**

Greenplum recommends that you take regular backups of your databases. These backups can be used to restore your data or to rebuild a Greenplum Database system in the event of a system failure or data corruption. You can also use the backups to migrate data from one Greenplum Database system to another.

**About Parallel Backups**

Greenplum provides a parallel dump utility called `gp_dump`. This utility backs up the Greenplum master instance and each active segment instance at the same time. See “Backing Up a Database with `gp_dump`” on page 173.

Because the segments are dumped in parallel, the time it takes to do a backup should scale regardless of the number of segments in your system. The dump files on the master host consist of DDL statements and the Greenplum-specific system catalog tables (such as `gp_segment_configuration`). On the segment hosts, there is one dump file created for each segment instance. The segment dump files contain the data for an individual segment instance. All of the dump files that comprise a total backup set are identified by a unique 14-digit timestamp key.

![Parallel Backups in Greenplum Database](image)
In order to automate routine backups, Greenplum also provides the `gpcrondump` utility, which is a wrapper for `gp_dump` that can be called directly or from a scheduled CRON job. `gpcrondump` also allows you to backup additional objects besides your databases and data, such as database roles and server configuration files. See “Automating Parallel Backups with gpcrondump” on page 174.

**About Non-Parallel Backups**

Greenplum also supports the regular PostgreSQL dump utilities: `pg_dump` and `pg_dumpall`. The PostgreSQL dump utilities (when used on Greenplum Database) will create one big dump file on the master host containing the data from all active segments. In most cases, this is probably not practical, as there is most likely not enough disk space on the master host to create a single backup file of an entire distributed database. These utilities are mostly supported for users who are migrating from regular PostgreSQL to Greenplum Database.

Another useful command for getting data out of a database is the `COPY TO SQL` command. This allows you to copy all or a portion of a table out of the database to a text-delimited file on the master host.

If you are migrating your data to another Greenplum Database system with a different segment configuration (for example, if the system you are migrating to has greater or fewer segment instances), Greenplum recommends using your parallel dump files created by `gp_dump` or `gpcrondump` and following the restore process described in “Restoring to a Different Greenplum System Configuration” on page 177.

**About Parallel Restores**

To do a parallel restore, you must have a complete backup set created by `gp_dump` or `gpcrondump`. Greenplum provides a parallel restore utility called `gp_restore`. This utility takes the timestamp key generated by `gp_dump`, validates the backup set, and restores the database objects and data into a distributed database. As with a parallel dump, each segment’s data is restored in parallel.

![Figure 16.2 Parallel Restores in Greenplum Database](image)
Greenplum also provides the `gpdbrestore` utility, which is a wrapper for `gp_restore`. `gpdbrestore` provides additional flexibility and verification options, which are useful if you are using automated backup files produced by `gpcrondump`, or have moved your backup files off of the Greenplum array to an alternate location. See “Restoring a Database Using gpdbrestore” on page 176.

### About Non-Parallel Restores

Greenplum also supports the regular PostgreSQL restore utility: `pg_restore`. This utility is mostly supported for users who are migrating to Greenplum Database from regular PostgreSQL, and have compressed dump files created by `pg_dump` or `pg_dumpall`. Before restoring PostgreSQL dump files into Greenplum Database, make sure to modify the `CREATE TABLE` statements in the dump files to include the Greenplum `DISTRIBUTED` clause.

It may also sometimes be necessary to do a non-parallel restore from a parallel backup set. For example, suppose you are migrating from a Greenplum system that has four segments to one that has five segments. You cannot do a parallel restore in this case, because your backup set only has four backup files and would not be evenly distributed across the new expanded system. A non-parallel restore using parallel backup files involves collecting each backup file from the segment hosts, copying them to the master host, and loading them through the master. See “Restoring to a Different Greenplum System Configuration” on page 177.

![Non-parallel Restore Using Parallel Backup Files](image)

**Figure 16.3** Non-parallel Restore Using Parallel Backup Files

### Backing Up a Database

There are three ways to backup a database:
• **Create a dump file for each segment with gp_dump.** Use this option if you are dumping a database for backup purposes, or if you want to migrate your data to another system with the same segment configuration (for example, same number of segment instances, but on different hosts). To restore, you must use the corresponding gp_restore utility. See “Backing Up a Database with gp_dump” on page 173. You can also use dump files created by gp_dump to restore to a different Greenplum system configuration if needed.

• **Schedule routine dumps with gpcrondump.** This is a wrapper utility for gp_dump. It allows you to schedule routine backups of a Greenplum Database database using cron (a scheduling utility for UNIX operating systems). Cron jobs that call gpcrondump should be scheduled on the Greenplum master host. gpcrondump also allows you to backup additional objects besides your databases and data, such as database roles and server configuration files.

• **Create a single dump file with pg_dump or pg_dumpall.** Use this option if you are migrating your data to another database vendor’s system. If restoring to a PostgreSQL or Greenplum database, you can use the corresponding pg_restore utility (if the dump file is in archive format), or you can use a client program such as psql (if the dump file is in plain text format). If you plan to restore to another Greenplum Database system, Greenplum recommends that you do a parallel dump using gp_dump or gpcrondump and then do a non-parallel restore.

### Backing Up a Database with gp_dump

The gp_dump utility dumps the contents of a Greenplum Database system into a series of SQL utility files, which can then be used to restore a Greenplum Database system configuration, database, and data. During a dump operation users will still have access to the database.

The gp_dump utility performs the following actions and produces the following dump files:

**On the master host**

- Dumps the Greenplum configuration system catalog tables into a SQL file in the master data directory. The default naming convention of this file is `gp_catalog_1_<dbid>_<timestamp>`.
- Dumps a `CREATE DATABASE` SQL statement into a file in the master data directory. The default naming convention of this file is `gp_cdatabase_1_<dbid>_<timestamp>`. This statement can be run on the master instance to recreate the database.
- Dumps the database schema(s) into a SQL file in the master data directory. The default naming convention of this file is `gp_dump_1_<dbid>_<timestamp>`. This file is used to recreate the database objects.
- Creates a log file in the master data directory named `gp_dump_status_1_<dbid>_<timestamp>`.
- gp_dump launches a gp_dump_agent for each segment instance to be backed up. gp_dump_agent processes run on the segment hosts and report status back to the gp_dump process running on the master host.

**On the segment hosts**
• Dumps the data for each segment instance into a SQL file in the segment instance’s data directory. By default, only primary (or active) segment instances are backed up. The default naming convention of this file is gp_dump_0_<dbid>_<timestamp>. This file is used to recreate that particular segment of data.

• Creates a log file in each segment instance’s data directory named gp_dump_status_0_<dbid>_<timestamp>.

Note that the 14 digit timestamp is the number that uniquely identifies the backup job, and is part of the filename for each dump file created by a gp_dump operation. This timestamp must be passed to the gp_restore utility when restoring a Greenplum database.

To backup a Greenplum database using gp_dump

1. From the master, run the gp_dump utility. For example (where mydatabase is the name of the database you are backing up):

   $ gp_dump mydatabase

Automating Parallel Backups with gpcrondump

gpcrondump is a wrapper utility for gp_dump, which can be called directly or from a crontab entry. It also allows you to backup additional objects besides your databases and data, such as database roles and server configuration files.

gpcrondump creates the dump files in the master and each segment instance’s data directory in <data_directory>/db_dumps/YYYYMMDD. The segment data dump files are compressed using gzip.

To schedule a dump operation using CRON

1. On the master, log in as the Greenplum superuser (gpadmin).

2. Define a crontab entry that calls gpcrondump. For example, to schedule a nightly dump of the sales database at one minute past midnight (note that the SHELL is set to /bin/bash and the PATH includes the location of the Greenplum Database management utilities):

   **Linux Example:**
   
   SHELL=/bin/bash
   GPHOME=/usr/local/greenplum-db-4.0.x.x
   MASTER_DATA_DIRECTORY=/data/gpdb_p1/gp-1
   PATH=$PATH:$GPHOME/bin
   01 0 * * * gpadmin gpcrondump -x sales -c -g -G -a -q >> gp_salesdump.log

   **Solaris Example** (no line breaks):
   
   01 0 * * * SHELL=/bin/bash
   GPHOME=/usr/local/greenplum-db-4.0.x.x PATH=$PATH:$GPHOME/bin
   HOME=/export/home/gpadmin
   MASTER_DATA_DIRECTORY=/data/gpdb_p1/gp-1
/usr/local/greenplum-db/bin/gpcrondump -x sales -c -g -G -a -q >> gp_salesdump.log

3. Create a file named mail_contacts in either the Greenplum superuser’s home directory or in $GPHOME/bin. For example:
   $ vi /home/gpadmin/mail_contacts
   $ vi /export/home/gpadmin/mail_contacts

4. In this file, type one email address per line. For example:
   dba@mycompany.com
   jjones@mycompany.com

5. Save and close the mail_contacts file. gpcrondump will send email notifications to the email addresses listed in this file.

### Restoring From Parallel Backup Files

The procedure for restoring a database from parallel backup files depends on a few factors. To determine the restore procedure to use, determine your answers to the following questions:

1. **Where are your backup files located?** If your backup files reside in their original location on the segment hosts where they were created by gp_dump, you can do a simple restore using gp_restore. If you have moved your backup files off of your Greenplum array, for example to an archive server, use gpdbrestore.

2. **Do you need to restore your entire system, or just your data?** If you have your Greenplum Database up and running and just need to restore your data, you can do a restore using gp_restore or gpdbrestore. If you have lost your entire array and need to rebuild the entire system from backup, use gpinitsystem.

3. **Are you restoring to a system with the same number of segment instances as your backup set?** If you are restoring to an array with the same number of segment hosts and segment instances per host, use gp_restore or gpdbrestore. If you are migrating to a different array configuration, you must do a non-parallel restore. See “Restoring to a Different Greenplum System Configuration” on page 177.

### Restoring a Database with gp_restore

The gp_restore utility recreates the data definitions (schema) and user data in a database using the backup files created by a gp_dump operation. To do a restore, you must have:

1. Backup files created by a gp_dump operation.

2. The backup files reside on the segment hosts in the location where gp_dump created them.

3. The Greenplum Database system up and running.
4. A Greenplum Database system with the exact same number of primary segment instances as the system that was backed up using gp_dump.

5. The database you are restoring to is created in the system.

6. If you used the options -s (schema only), -a (data only), --gp-c (compressed), --gp-d (alternate dump file location) when performing the gp_dump operation, you must specify these options when doing the gp_restore as well.

The gp_restore utility performs the following actions:

On the master host
- Runs the SQL DDL commands in the gp_dump_1_<dbid>_<timestamp> file created by gp_dump to recreate the database schema and objects.
- Creates a log file in the master data directory named gp_restore_status_1_<dbid>_<timestamp>.
- gp_restore launches a gp_restore_agent for each segment instance to be restored. gp_restore_agent processes run on the segment hosts and report status back to the gp_restore process running on the master host.

On the segment hosts
- Restores the user data for each segment instance using the gp_dump_1_<dbid>_<timestamp> files created by gp_dump. Each primary and mirror segment instance on a host is restored.
- Creates a log file for each segment instance named gp_restore_status_1_<dbid>_<timestamp>.

Note that the 14 digit timestamp is the number that uniquely identifies the backup job to be restored, and is part of the filename for each dump file created by a gp_dump operation. This timestamp must be passed to the gp_restore utility when restoring a database.

To restore from a backup created by gp_dump

1. Make sure the backup files created by gp_dump reside on the master host and segment hosts for the Greenplum Database system you are restoring.

2. Make sure the database you are restoring to has been created in the system. For example:
   ```
   createdb database_name
   ```

3. From the master, run the gp_restore utility. For example (where --gp-k specifies the timestamp key of the backup job and -d specifies the database to connect to):
   ```
   gp_restore -gp-k=2007103112453 -d database_name
   ```

Restoring a Database Using gpdbrestore

The gpdbrestore utility is a wrapper around gp_restore, which provides some convenience and flexibility in restoring from a set of backup files created by gpcrondump. To do a restore using gpdbrestore, you must have:
Restoring From Parallel Backup Files

1. Backup files created by a `gpcrondump` operation.
2. The Greenplum Database system up and running.
3. A Greenplum Database system with the *exact* same number of primary segment instances as the system that was backed up.
4. The database you are restoring to is created in the system.

To restore from an archive host using `gpdbrestore`

(This procedure assumes your backup set has been moved off of your Greenplum array to another host in the network)

1. First, make sure that the archive host is reachable from the Greenplum master host:
   
   ```sh
   $ ping archive_host
   ```

2. Make sure the database you are restoring to has been created in the system. For example:
   
   ```sh
   $ createdb database_name
   ```

3. From the master, run the `gpdbrestore` utility. For example (where `-R` specifies the host name and path to a complete backup set):
   
   ```sh
   $ gpdbrestore -R archive_host:/gpdb/backups/archive/20080714
   ```

Restoring to a Different Greenplum System Configuration

In order to do a parallel restore operation using `gp_restore` or `gpdbrestore`, the system you are restoring to must be the same configuration as the system that was backed up. If you want to restore your database objects and data into a different system configuration (for example, if you are expanding to a system with more segments), you can still use your parallel backup files and restore them by loading them through the Greenplum master. To do a non-parallel restore, you must have:

1. A complete backup set created by a `gp_dump` or `gpcrondump` operation. The backup file of the master contains the DDL to recreate your database objects. The backup files of the segments contain the data.
2. A Greenplum Database system up and running.
3. The database you are restoring to is created in the system.

If you look at the contents of a segment dump file, it simply contains a `COPY` command for each table followed by the data in delimited text format. If you collect all of the dump files for all of the segment instances and run them through the master, you will have restored all of your data and redistributed it across the new system configuration.

To restore a database to a different system configuration

1. First make sure you have a complete backup set. This includes a dump file of the master (`gp_dump_1_* <timestamp>`) and one for each segment instance (`gp_dump_0_2 <timestamp>, gp_dump_0_3 <timestamp>`),
gp_dump_0_4_<timestamp>, and so on. The individual dump files should all have the same timestamp key. By default, `gp_dump` creates the dump files in each segment instance’s data directory, so you will need to collect all of the dump files and move them to a place on the master host. If you do not have a lot of disk space on the master, you can copy each segment dump file to the master, load it, and then delete it once it has loaded successfully.

2. Make sure the database you are restoring to has been created in the system. For example:
   
   $ createdb database_name

3. Load the master dump file to restore the database objects. For example:

   $ psql database_name -f /gpdb/backups/gp_dump_1_1_20080714

4. Load each segment dump file to restore the data. For example:

   $ psql database_name -f /gpdb/backups/gp_dump_0_2_20080714
   $ psql database_name -f /gpdb/backups/gp_dump_0_3_20080714
   $ psql database_name -f /gpdb/backups/gp_dump_0_4_20080714
   $ psql database_name -f /gpdb/backups/gp_dump_0_5_20080714
   ...

...
17. Expanding a Greenplum System

This chapter provides information on adding additional resources to an existing Greenplum Database system in order to scale performance and storage capacity. It contains the following topics:

- Planning Greenplum System Expansion
- Preparing and Adding Nodes
- Initializing New Segments
- Redistributing Tables
- Removing the Expansion Schema

Though this chapter provides some general information on preparing hardware platforms, it focuses chiefly on software aspects of expansion. Greenplum recommends that you work with our platform engineers when configuring hardware resources for expanding Greenplum Database.

Planning Greenplum System Expansion

Careful planning is critical to the success of a system expansion operation. By thoroughly preparing all new hardware and carefully planning all the steps of the expansion procedure, you can minimize risk and downtime for Greenplum Database.

Administrators of large-scale systems may want to give special attention to the performance-related considerations covered in “Planning Table Redistribution” on page 183.

This section provides an overview and a checklist for the system expansion process.

System Expansion Overview

System expansion consists of three phases:

- Adding and testing new hardware platforms
- Initializing new segments
- Redistributing tables

**Adding and testing new hardware** — General considerations for deploying new hardware are described in “Planning New Hardware Platforms” on page 182. For more information on hardware platforms, consult Greenplum platform engineers. After the new hardware platforms are provisioned and networked, you must run performance tests using Greenplum utilities.

**Initializing new segments** — Once Greenplum Database is installed on new hardware, you must initialize the new segments using gpexpand (not gpinitsystem). In this process, the utility creates a data directory and copies all user tables from all
existing databases on the new segments, capturing metadata for each table in an expansion schema for status tracking. After this process has completed successfully, the expansion operation is committed, and cannot be reversed.

These operations are performed with the system offline. The `gpexpand` utility will shut down the database during initialization if you have not already done so.

**Redistributing tables** — As part of the initialization process, `gpexpand` nullifies hash distribution policies (except for the parent tables of a partitioned table) and sets the distribution policy for all tables to random distribution. This action is performed on all tables in all existing databases in the Greenplum Database instance. Users can continue to access Greenplum Database after initialization is complete and the system is back online, though they may experience some performance degradation on systems that rely heavily on hash distribution of tables. During this process, normal operations such as ETL jobs, user queries, and reporting can continue, although users might experience slower response times.

**Important:** When a table has a random distribution policy, Greenplum Database cannot enforce unique constraints (such as `PRIMARY KEY`). This might affect your ETL and loading processes until table redistribution is completed, as duplicate rows will not issue a constraint violation error as expected.

To complete system expansion, you must run `gpexpand` to redistribute data tables across the newly added segments. Depending on the size and scale of your system, this might be accomplished in a single session during low-use hours, or it might require you to divide the process into batches over an extended period. Each table or partition will be unavailable for read or write operations during the period in which it is being redistributed. As each table is successfully redistributed across the new segments according to its distribution key (if any), the performance of the database should incrementally improve until it equals and then exceeds pre-expansion performance levels.

In a typical operation, you will run the `gpexpand` utility four times with different options during the complete expansion process.

- To interactively create an expansion input file:
  
  `gpexpand -f hosts_file`

- To initialize segments and create expansion schema:
  
  `gpexpand -i input_file -D database_name`

- To redistribute tables:
  
  `gpexpand -d duration`

- To remove the expansion schema:
  
  `gpexpand -c`

In systems whose large scale requires multiple redistribution sessions, `gpexpand` may need to be run several more times to complete the expansion, and may benefit from explicit table redistribution ranking. For more information, see “Planning Table Redistribution” on page 183.”
System Expansion Checklist
This checklist provides a quick overview of the steps required for a system expansion.

## Online Pre-Expansion Preparation

* System is up and available

- [ ] Devise and execute a plan for ordering, building, and networking new hardware platforms.
- [ ] Devise a database expansion plan. Map the number of segments per host, schedule the offline period for testing performance and creating the expansion schema, and schedule the intervals for table redistribution.
- [ ] Install Greenplum Database binaries on new hosts.
- [ ] Copy SSH keys to the new hosts (`gpssh-exkeys`).
- [ ] Validate the operating system environment of the new hardware (`gpcheck`).
- [ ] Validate disk I/O and memory bandwidth of the new hardware (`gpcheckperf`).
- [ ] Prepare an expansion input file (`gpexpand`).

## Offline Expansion Tasks

* The system will be locked down to all user activity during this process

- [ ] Validate the operating system environment of the combined existing and new hardware (`gpcheck`).
- [ ] Validate disk I/O and memory bandwidth of the combined existing and new hardware (`gpcheckperf`).
- [ ] Initialize new segments into the array and create an expansion schema (`gpexpand -i input_file`).

## Online Expansion and Table Redistribution

* System is up and available

- [ ] Before beginning table redistribution, stop any automated snapshot processes or other processes that consume disk space.
- [ ] Redistribute tables through the expanded system (`gpexpand`).
- [ ] Remove expansion schema (`gpexpand -c`).
Planning New Hardware Platforms

Careful preparation of new hardware for system expansion is extremely important. Deliberate and thorough deployment of compatible hardware can greatly minimize the risk of issues developing later in the system expansion process.

All new segment hosts for the expanded Greenplum Database array should have hardware resources and configurations matching those of the existing hosts. Greenplum recommends that you work with Greenplum Platform Engineering prior to making a hardware purchase decision for expanding Greenplum Database.

The steps to plan and set up new hardware platforms will vary greatly for each unique deployment. Some of the possible considerations include:

- Preparing the physical space for the new hardware. Consider cooling, power supply, and other physical factors.
- Determining the physical networking and cabling required to connect the new and existing hardware.
- Mapping the existing IP address spaces and developing a networking plan for the expanded system.
- Capturing the system configuration (users, profiles, NICs, etc.) from existing hardware to list it in detail for ordering the new hardware.
- Creating a custom build plan for deploying hardware with the desired configuration in the particular site and environment.

After selecting and adding new hardware to your network environment, make sure you perform the burn-in tasks described in “Verifying OS Settings” on page 188.

Planning Initialization of New Segments

Expanding Greenplum Database requires a limited period of system down time. During this period, you must run `gpexpand` to initialize new segments into the array and create an expansion schema.

The time required will depend on the number of schema objects in the Greenplum system, and other factors related to hardware performance. In most environments, the initialization of new segments will require less than thirty minutes offline.

**Note:** After you begin initializing new segments, you can no longer restore the system using `gp_dump` files created for the pre-expansion system. When initialization is successfully completed, the expansion is committed and cannot be rolled back.

Planning Mirror Segments

If your existing array has mirror segments, the new segments are required to have mirroring configured. Conversely, if there are no mirrors configured for existing segments, you cannot add mirrors to new hosts with the `gpexpand` utility.

For Greenplum Database arrays with mirror segments, you must make sure that you have added enough new host machines to accommodate the new mirror segments. The number of new hosts required depends on your mirroring strategy:
Spread Mirroring — add at least one more host to the array than the number of segments per host. The number of separate hosts must be greater than the number of segment instances per host to ensure even spreading.

Grouped Mirroring — add no fewer than two new hosts. In a minimum case with two hosts, this ensures that the mirrors for the first host can reside on the second host, and vice versa.

For more information, see “About Segment Mirroring” on page 8.

Increasing Segments Per Host

By default, new hosts are initialized with the same number of primary segments as existing hosts. Optionally, you can increase the number of segments per host, or add new segments only to existing hosts.

For example, if existing hosts currently have two segments per host, you can use `gpexpand` to initialize two additional segments on existing hosts (for a total of four), and four new segments on new hosts.

The interactive process for creating an expansion input file prompts for this option, and the input file format allows you to specify new segment directories manually as well. For more information, see “Creating an Input File for System Expansion” on page 189.

About the Expansion Schema

At initialization time, `gpexpand` creates an expansion schema. If you do not specify a particular database at initialization time (`gpexpand -D`), the schema is created in the database indicated by the `PGDATABASE` environment variable.

The expansion schema stores metadata for each table in the system so that its status can be tracked throughout the expansion process. It consists of two tables and a view for tracking the progress of an expansion operation:

- `gpexpand.status`
- `gpexpand.status_detail`
- `gpexpand.expansion_progress`

You can control aspects of the expansion process by modifying `gpexpand.status_detail`. For example, removing a record from this table prevents the table from being expanded across new segments. By updating the `rank` value for a record, you can control the order in which tables are processes for redistribution. For more information, see “Ranking Tables for Redistribution” on page 193.

Planning Table Redistribution

The redistribution of tables is performed with the system online. For many Greenplum systems, table redistribution can be completed in a single `gpexpand` session scheduled during a low-use period. Larger systems may require you to plan multiple sessions and to set the order of table redistribution in order to minimize the performance impact. Greenplum recommends completing the table redistribution in one session if your database size and design permit it.
Important: To perform table redistribution, you must have enough disk space on your segment hosts to temporarily hold a copy of your largest table. Each table is unavailable for read and write operations while `gpexpand` is redistributing it among the segments.

The performance impact of table redistribution depends on the size, storage type, and partitioning design of a table. Redistributing a table with `gpexpand` takes approximately as much time per table as a `CREATE TABLE AS SELECT` operation would take. When redistributing a terabyte-scale fact table, the expansion utility may use a significant portion of available system resources, with resulting impact on the performance of queries or other database workload.

**Managing Redistribution in Large-Scale Greenplum Systems**

You can manage the order in which tables are redistributed by adjusting their ranking as described in “Ranking Tables for Redistribution” on page 193. Manipulating the redistribution order can help you adjust for limited disk space and restore optimal query performance more quickly.

When planning the redistribution phase, you should consider the impact of the exclusive lock taken on each table while it is actively being redistributed. User activity on a table may delay the beginning of its scheduled redistribution. Likewise, a table will be unavailable for all other operations while the `gpexpand` utility is redistributing it among the new segments.

**Systems with Abundant Free Disk Space**

In systems with abundant free disk space (required to store a copy of the largest table), you can focus on restoring optimum query performance as soon as possible by first redistributing important tables that are heavily used by common queries. Accordingly, assign high ranking to these tables, and schedule the redistribution operations for times of low system usage. Run only one redistribution process at a time until large or critical tables have been successfully redistributed.

**Systems with Limited Free Disk Space**

If your existing hosts have limited disk space, you may prefer to first redistribute smaller tables (such as dimension tables) in order to clear the space needed to store a copy of the largest table. Disk space on the original segments will increase as each table is redistributed across the expanded array. Once the amount of free space on all segments is adequate to store a copy of the largest table, you can redistribute large and critical tables. Again, because of the exclusive lock required, you should schedule large tables for off-hours.

Also consider the following:

- Run multiple parallel redistribution processes in off-hours to maximize the available system resources.
- When running multiple process, make sure you operate within the connection limits for your Greenplum system. For more information, see “Limiting Concurrent Connections” on page 38.
Redistributing Append-Only and Compressed Tables
Append-only and compressed append-only tables are redistributed by `gpexpand` at different rates from heap tables. The CPU capacity required to compress and decompress data tends to increase the impact on system performance. For similar-sized tables with similar data, you may find overall performances differences like the following:

- Uncompressed append-only tables expand 10% faster than heap tables
- Zlib-compressed append-only tables expand at a significantly slower rate than uncompressed append-only tables, potentially up to 80% slower.
- Systems with data compression such as ZFS/LZJB experience a significantly slower rate of redistribution.

**Important:** If file system compression is used on the existing nodes in the system, identical compression must be used on new nodes to avoid disk space shortage.

Redistributing Tables with Primary Key Constraints
Between the initialization of new segments and the successful redistribution of tables, there is a window during which primary key constraints cannot be enforced. Any duplicate data inserted into tables during this window will prevent the expansion utility from redistributing the affected tables. Once a table is successfully redistributed, the primary key constraint is again properly enforced.

If constraints are violated during the expansion process, the expansion utility logs errors and prints warnings to the screen after attempting to redistribute all tables. You have the following options to remedy constraint violations:

- Clean up duplicate data in the primary key columns, and re-run `gpexpand`.
- Drop the primary key constraints, and re-run `gpexpand`.

Redistributing Tables with User-Defined Data Types
With tables that have dropped columns of user-defined data types, you cannot perform redistribution with the expansion utility. To redistribute tables with dropped columns of user-defined types, first re-create the table using `CREATE TABLE AS SELECT`. Once the dropped columns are removed by this process, you can proceed to redistribute the table with the expansion utility.

Redistributing Partitioned Tables
Because the expansion utility can process a large table partition by partition, an efficient partition design reduces the performance impact of table redistribution. Only the child tables of a partitioned table are set to a random distribution policy, and the read/write lock for redistribution applies to only one child table at a time.

Redistributing Indexed Tables
Because the `gpexpand` utility must re-index each indexed table after redistribution, a high level of indexing has a high performance impact. Systems with intensive indexing will have significantly slower rates of table redistribution.
Preparing and Adding Nodes

To prepare new system nodes for expansion, install the Greenplum Database software binaries, exchange the required SSH keys and run performance tests. Greenplum recommends running performance tests at least twice: first on the new nodes only, and then on both the new and existing nodes together. The second set of tests must be run with the system offline in order to prevent user activity from distorting test results.

Beyond these general guidelines, Greenplum recommends running performance tests any time that the networking of nodes is modified, or for any special conditions in the system environment. For example, if you plan to run the expanded system on two network clusters, run the performance tests on each cluster.

This rest of this section describes how to run Greenplum administrative utilities to verify that your new nodes are ready for integration into the existing Greenplum system.

Adding New Nodes to the Trusted Host Environment

New nodes must exchange SSH keys with the existing nodes in order to enable Greenplum administrative utilities to connect to all segments without a password prompt.

Greenplum recommends performing the key exchange process twice: once as root (for administration convenience) and once as the gpadmin user (required for the Greenplum management utilities). Perform the following tasks in this order:

- “To exchange SSH keys as root” on page 186
- “To create the gpadmin user” on page 187
- “To exchange SSH keys as the gpadmin user” on page 188

To exchange SSH keys as root

1. Create two separate host list files: one that has all of the existing host names in your Greenplum Database array, and one that has all of the new expansion hosts. For existing hosts, you can use the same host file that you used for the initial setup of SSH keys in the system.
The files should include all hosts (master, backup master and segment hosts) and list one host name per line. If using a multi-NIC configuration, make sure to exchange SSH keys using all of the configured host names for a given host. Make sure there are no blank lines or extra spaces. For example:

```
mdw  OR  masterhost
sdw1-1
sdw1-2
sdw1-3
sdw1-4
sdw2-1
sdw2-2
sdw2-3
sdw2-4
sdw3-1
sdw3-2
sdw3-3
sdw3-4
```

2. Log in as root on the master host, and source the greenplum_path.sh file from your Greenplum installation.

```
$ su -
# source /usr/local/greenplum-db/greenplum_path.sh
```

3. Run the gpssh-exkeys utility referencing the host list files. For example:

```
# gpssh-exkeys -f /home/gpadmin/existing_hosts_file -x /home/gpadmin/new_hosts_file
```

4. gpssh-exkeys will check the remote hosts and perform the key exchange between all hosts. Enter the root user password when prompted. For example:

```
***Enter password for root@hostname: <root_password>
```

**To create the gpadmin user**

1. Use gpssh to create the gpadmin user on all of the new segment hosts (if it does not exist already). Use the list of new hosts that you created for the key exchange. For example:

```
# gpssh -f new_hosts_file '/usr/sbin/useradd gpadmin -d /home/gpadmin -s /bin/bash'
```

2. Set the new gpadmin user’s password. On Linux, you can do this on all segment hosts at once using gpssh. For example:

```
# gpssh -f new_hosts_file 'echo gpadmin_password | passwd gpadmin --stdin'
```

On Solaris, you must log in to each segment host and set the gpadmin user’s password on each host. For example:

```
# ssh segment_hostname
# passwd gpadmin
# New password: <gpadmin_password>
# Retype new password: <gpadmin_password>

3. Verify that the gpadmin user has been created by looking for its home directory:
   # gpssh -f new_hosts_file ls -l /home

To exchange SSH keys as the gpadmin user

1. Log in as gpadmin, and run the gpssh-exkeys utility referencing the host list files. For example:
   # gpssh-exkeys -f /home/gpadmin/existing_hosts_file -x /home/gpadmin/new_hosts_file

2. gpssh-exkeys will check the remote hosts and perform the key exchange between all hosts. Enter the gpadmin user password when prompted. For example:
   ***Enter password for gpadmin@hostname: <gpadmin_password>

Verifying OS Settings

Use the gpcheck utility to verify that all the new hosts in your array have the correct OS settings for running the Greenplum Database software.

To run gpcheck

1. Log in on the master host as the user who will be running your Greenplum Database system (for example, gpadmin).
   $ su - gpadmin

2. Run the gpcheck utility using your host file for new hosts. For example:
   $ gpcheck -f new_hosts_file

Validating Disk I/O and Memory Bandwidth

Use the gpcheckperf utility to test disk I/O and memory bandwidth.

To run gpcheckperf

1. Run the gpcheckperf utility using the host file for new hosts. Use the -d option to specify the file systems you want to test on each host (you must have write access to these directories). For example:
   $ gpcheckperf -f new_hosts_file -d /data1 -d /data2 -v

2. The utility may take a while to perform the tests as it is copying very large files between the hosts. When it is finished, you will see the summary results for the Disk Write, Disk Read, and Stream tests.

If your network is divided into subnets, repeat this procedure with a separate host file for each subnet.
Integrating New Hardware into the System

Before initializing the system with all new segments, repeat the performance tests on all nodes in the system, new and existing. Shut down the system and run these same tests using host files that include all nodes, existing and new:

- Verifying OS Settings
- Validating Disk I/O and Memory Bandwidth

Because user activity may skew the results of these test, you must shut down Greenplum Database (gpstop) before running them.

Initializing New Segments

Use the gpexpand utility to initialize the new segments, create the expansion schema, and set a system-wide random distribution policy for the database. The utility performs these tasks by default the first time you run it with a valid input file on a Greenplum Database master. Subsequently, it will detect that an expansion schema has been created, and will perform table redistribution.

Creating an Input File for System Expansion

To begin expansion, the gpexpand utility requires an input file containing information about the new segments and hosts. If you run gpexpand without specifying an input file, the utility displays an interactive interview that collects the required information and automatically creates an input file for you.

If you choose to create the input file using the interactive interview, you can optionally specify a file containing a list of expansion hosts. If your platform or command shell limits the length of the list of hostnames that you can type when prompted in the interview, specifying the hosts with -f may be mandatory.

Creating an input file in Interactive Mode

Before running gpexpand to create an input file in interactive mode, make sure you have the required information:

- Number of new hosts (or a hosts file)
- New hostnames (or a hosts file)
- The mirroring strategy used in existing hosts, if any
- Number of segments to add per host, if any

The utility automatically generates an input file based on this information and dbid, content ID, and data directory values stored in gp_segment_configuration, and saves it in the current directory.

To create an input file in interactive mode

1. Log in on the master host as the user who will be running your Greenplum Database system (for example, gpadmin).
2. Run `gpexpand`. The utility displays messages about preparing for an expansion operation and prompts you to quit or continue.
   Optionally, specify a hosts file using `-f`. For example:
   ```bash
   $ gpexpand -f /home/gpadmin/new_hosts_file
   ```

3. At the prompt, select `Y` to continue.

4. Unless you specified a hosts file using `-f`, you will be prompted to enter hostnames. Enter a comma separated list of the hostnames of the new expansion hosts. Do not include interface hostnames. For example:
   ```
   > sdw4-1, sdw4-2, sdw4-3, sdw4-4
   ```
   To add segments to existing hosts only, enter a blank line at this prompt. Do not specify `localhost` or any existing host name.

5. Enter the mirroring strategy used in your system, if any. Options are `spread|grouped|none`, default is `grouped`.
   Make sure you have enough hosts for your selected grouping strategy. For more information, see “Planning Mirror Segments” on page 182.

6. Enter the number of new primary segments to add, if any. By default, new hosts are initialized with the same number of primary segments as existing hosts. Optionally, you can increase the number of segments per host.
   If you want to increase the number of segments per host, enter a number greater than zero. This number of additional segments will be initialized on all hosts. For example, if existing hosts currently have two segments per host, entering a value of 2 will initialize two additional segments on existing hosts, and four new segments on new hosts.

7. If you are adding new primary segments, enter the new primary data directory root for the new segments. Do not specify the actual data directory name, which is created automatically by `gpxpand` based on the existing data directory names.
   For example, if your existing data directories are as follows:
   ```
   /gpdata/primary/gp0
   /gpdata/primary/gp1
   ```
   you should enter the following (one at each prompt) to specify the data directories for two new primary segments:
   ```
   /gpdata/primary
   /gpdata/primary
   ```
   When the initialization is run, the utility will create the new directories `gp2` and `gp3` under `/gpdata/primary`

8. If you are adding new mirror segments, enter the new mirror data directory root for the new segments. Do not specify the actual data directory name, which is created automatically by `gpxpand` based on the existing data directory names.
   For example, if your existing data directories are as follows:
   ```
   /gpdata/mirror/gp0
   /gpdata/mirror/gp1
   ```
you should enter the following (one at each prompt) to specify the data directories for two new mirror segments:

/gpdata/mirror
/gpdata/mirror

When the initialization is run, the utility will create the new directories `gp2` and `gp3` under `/gpdata/mirror`

**Important:** These primary and mirror root directories for new segments must exist on the hosts, and the user running `gpexpand` must have permissions to create directories in them.

After you have entered all required information, the utility generates an input file and saves it in the current directory. For example:

`gpexpand_inputfile_yyyymmdd_145134`

**Expansion Input File Format**

You can create your own input file in the required format. Unless you have special needs for your expansion scenario, Greenplum recommends creating the input file using the interactive interview process.

The format for expansion input files is:

`hostname:address:port:fselocation:dbid:content:preferred_role:replication_port`

For example:

`sdw5:sdw5-1:50011:/gpdata/primary/gp9:11:9:p:53011`
`sdw5:sdw5-2:50012:/gpdata/primary/gp10:12:10:p:53011`
`sdw5:sdw5-1:60012:/gpdata/mirror/gp10:14:10:m:63011`

An expansion input file in this format requires the following information for each new segment:

<table>
<thead>
<tr>
<th>Table 17.1 Data for the expansion configuration file</th>
<th>Parameter</th>
<th>Valid Values</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>hostname</td>
<td>Hostname</td>
<td>Hostname for the segment host.</td>
<td></td>
</tr>
<tr>
<td>port</td>
<td>An available port number</td>
<td>Database listener port for the segment, incremented on the existing segment port base number.</td>
<td></td>
</tr>
<tr>
<td>fselocation</td>
<td>Directory name</td>
<td>The data directory (filesystem) location for a segment as per the <code>pg_filespace_entry</code> system catalog.</td>
<td></td>
</tr>
</tbody>
</table>
Running gpexpand to Initialize New Segments

After you have created an input file, run gpexpand to initialize new segments. The utility will automatically stop Greenplum Database for the time required to initialize the segments, and then restart the system when finished.

To run gpexpand with an input file

1. Log in on the master host as the user who will be running your Greenplum Database system (for example, gpadmin).

2. Run the gpexpand utility, specifying the input file with -i. Optionally, use -D to specify the database in which to create the expansion schema. For example:

   $ gpexpand -i input_file -D database1

   The utility detects if there is an existing expansion schema for the Greenplum system. If there is an existing schema, you must remove it with gpexpand -c before beginning a new expansion operation. See “Removing the Expansion Schema” on page 195.

   When the new segments are initialized and the expansion schema is successfully created, the utility prints a success message and exits. When the initialization process is complete, you can connect to Greenplum Database and view the expansion schema. The schema resides in the database you specified with -D, or in the database specified by the PGDATABASE environment variable. For more information, see “About the Expansion Schema” on page 183.

---

### Table 17.1  Data for the expansion configuration file

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Valid Values</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbid</td>
<td>Integer. Must not conflict with existing dbid values.</td>
<td>Database ID for the segment. The values you enter should be incremented sequentially from existing dbid values shown in the system catalog gp_segment_configuration. For example, to add four nodes to an existing ten-segment array with dbid values of 1-10, list new dbid values of 11, 12, 13 and 14.</td>
</tr>
<tr>
<td>content</td>
<td>Integer. Must not conflict with existing content values.</td>
<td>The content ID of the segment. A primary segment and its mirror should have the same content ID, incremented sequentially from existing values. For more information, see content in the reference for gp_segment_configuration.</td>
</tr>
<tr>
<td>preferred_role</td>
<td>p</td>
<td>m</td>
</tr>
<tr>
<td>replication_port</td>
<td>An available port number</td>
<td>File replication port for the segment, incremented on the existing segment replication_port base number.</td>
</tr>
</tbody>
</table>


**Rolling Back an Failed Expansion Setup**

You can roll back a failed expansion setup operation using the command `gpexpand -r | --rollback`. However, this command is only allowed in a failure scenario. Once a setup operation has completed successfully, the expansion is committed, and you cannot roll back.

To roll back a failed expansion setup, use the following command, specifying the database that contains the expansion schema:

```
gpexpand --rollback -D database_name
```

---

**Redistributing Tables**

After successfully creating an expansion schema, you can bring Greenplum Database back online and redistribute tables across the entire array. You can redistribute tables with `gpexpand` at specified intervals, targeting low-use hours when the utility’s CPU usage and table locks will have the least impact on database operations. Also, you can rank tables to ensure that the largest or most critical tables are redistributed in your preferred order.

While the redistribution of tables is underway:

- Any new tables or partitions created will be distributed across all segments exactly as they would be under normal operating conditions.
- Queries will use all segments, even though the relevant data may not have yet been redistributed to the tables on the new segments.
- The table or partition currently being redistributed will be locked and unavailable for read or write operations. When its redistribution is completed, normal operations resume.

**Ranking Tables for Redistribution**

For large systems, it is recommended to control the order in which tables are redistributed by adjusting their `rank` values in the expansion schema. This allows you to redistribute heavily-used tables first and minimize the performance hit on the system. The amount of free disk space available can affect table ranking; see “Managing Redistribution in Large-Scale Greenplum Systems” on page 184 for more information.

To rank tables for redistribution by updating `rank` values in `gpexpand.status_detail`, connect to Greenplum Database using `psql` or another supported client. Update `gpexpand.status_detail` with commands like the following:

```
=> UPDATE gpexpand.status_detail SET rank= 10;
UPDATE gpexpand.status_detail SET rank=1 WHERE fq_name = 'public.lineitem';
UPDATE gpexpand.status_detail SET rank=2 WHERE fq_name = 'public.orders';
```
These commands first lower the priority of all tables to 10, and then assign a rank of 1 to `lineitem` and then a rank of 2 to `orders`. When table redistribution begins, `lineitem` will be redistributed first, followed by `orders` and then all other tables in `gpexpand.status_detail`.

**Note:** For any table that you do not want to redistribute, you must remove the corresponding entry from `gpexpand.status_detail`.

---

### Redistributing Tables Using `gpexpand`

**To redistribute tables with `gpexpand`**

1. Log in on the master host as the user who will be running your Greenplum Database system (for example, `gpadmin`).

2. Run the `gpexpand` utility. Optionally, you can use either the `-d` or `-e` option to define the time period for the expansion session. For example, to run the utility for a maximum of 60 consecutive hours:
   ```
   $ gpexpand -d 60:00:00
   ```
   The utility redistributes tables until the last table in the schema is successfully marked completed, or until the specified duration or end time is reached. Each time a session is started or finished, the utility updates the status and updated time in `gpexpand.status`.

---

### Monitoring Table Redistribution

At any time during the process of redistributing tables, you can query the expansion schema. The view `gpexpand.expansion_progress` provides a summary of the current progress, including calculations of the estimated rate of table redistribution and estimated time to completion. The table `gpexpand.status_detail` can be queried for per-table status information.

**Viewing Expansion Status**

Because the estimates in `gpexpand.expansion_progress` are based on the rates achieved for each table, the view cannot calculate an accurate estimate until the first table has completed. Calculations are restarted each time you re-run `gpexpand` to start a new table redistribution session.

To monitor progress by querying `gpexpand.expansion_progress`, connect to Greenplum Database using `psql` or another supported client. Query `gpexpand.expansion_progress` with a command like the following:

```
=\$ select * from gpexpand.expansion_progress;

<table>
<thead>
<tr>
<th>name</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bytes Left</td>
<td>5534842880</td>
</tr>
<tr>
<td>Bytes Done</td>
<td>142475264</td>
</tr>
<tr>
<td>Estimated Expansion Rate</td>
<td>680.75667095996092 MB/s</td>
</tr>
</tbody>
</table>
```
### Viewing Table Status

The table `gpexpand.status_detail` stores status, last updated time, and other useful information about each table in the schema. To monitor the status of a particular table by querying `gpexpand.status_detail`, connect to Greenplum Database using `psql` or another supported client. Query `gpexpand.status_detail` with a command like the following:

```
=> SELECT status, expansion_started, source_bytes FROM gpexpand.status_detail WHERE fq_name = 'public.sales';
```

<table>
<thead>
<tr>
<th>status</th>
<th>expansion_started</th>
<th>source_bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPLETED</td>
<td>2009-02-20 10:54:10.043869</td>
<td>4929748992</td>
</tr>
</tbody>
</table>

(1 row)

### Removing the Expansion Schema

The expansion schema can safely be removed after the expansion operation is completed and verified. To run another expansion operation on a Greenplum system, you must first remove the existing expansion schema.

**To remove the expansion schema**

1. Log in on the master host as the user who will be running your Greenplum Database system (for example, `gpadmin`).

2. Run the `gpexpand` utility with the `-c` option. For example:

   ```
   $ gpexpand -c
   ```
18. Monitoring a Greenplum System

As the Greenplum administrator, it is often necessary to know what is going on in the Greenplum Database system in order to plan workflow and troubleshoot problems. This chapter discusses the various tools available for monitoring database performance and activity.

- Monitoring Database Activity and Performance
- Monitoring System State
- Viewing the Database Server Log Files
- Using gp_toolkit

Also, once one has identified a poorly-performing query, further investigation may be needed using the `EXPLAIN` command. See “Query Profiling” on page 119 for more information.

Monitoring Database Activity and Performance

Greenplum provides an optional performance monitoring feature that administrators can enable within Greenplum Database 4.0.

Enabling Greenplum Performance Monitor is a two part process. First you must enable the Greenplum Database server to collect and store query performance data and system metrics. After data collection is enabled on the server side, the next step is to install and configure the Greenplum Performance Monitor Console (a Web application used to view the performance monitor data stored in Greenplum Database).

The Greenplum Performance Monitor console is shipped separately from your Greenplum Database 4.0 installation. You can download the Greenplum Performance Monitor console package and documentation from Greenplum Network. See the Greenplum Database Performance Monitor Administrator Guide for more information on installing and using the Greenplum Performance Monitor Console.

Monitoring System State

As a Greenplum Database administrator, you will need to monitor the system for certain problem events (such as a segment going down or running out of disk space on a segment host). This section describes how to monitor the health of a Greenplum Database system and how to examine certain state information for a Greenplum Database system.

- Enabling System Alerts and Notifications
- Checking System State
- Checking Disk Space Usage
- Checking for Data Distribution Skew
Enabling System Alerts and Notifications
You can configure a Greenplum Database system to trigger SNMP alerts or send email notifications to system administrators whenever certain database events occur. These events include:

- All PANIC-level error conditions
- All FATAL-level error conditions
- ERROR-level conditions that are "internal errors" (for example, SIGSEGV errors)
- Database system shutdown and restart
- Segment failure and recovery
- Standby master out-of-sync conditions
- Master host manual shutdown or other software problem (in certain failure scenarios, Greenplum Database will not be able to send an alert or notification)

This section includes the following topics:

- Using SNMP with a Greenplum Database System
- Enabling Email Notifications

Note that SNMP alerts and email notifications report the same event information. There is no difference in the event information that either tool reports.

Using SNMP with a Greenplum Database System
Greenplum’s gpssnmpd agent is an SNMP (Simple Network Management Protocol) daemon that support SNMP requests on the health and state of a Greenplum Database system by using a set of MIBs (Management Information Bases).

MIBs are a collection of objects that describe an SNMP-manageable entity — in this case, a Greenplum Database system. An Agent is any SNMP software running on a managed device that responds to queries or set requests. The gpssnmpd daemon currently supports the generic RDBMS MIB and typically operates on the master host.

gpssnmpd works in conjunction with the SNMP support that already exists on the Greenplum Database system. You can install and run gpssnmpd as a AgentX (Agent Extensibility Protocol) sub-agent to the operating system’s SNMP agent (usually called snmpd). This allows a Network Management System to get hardware and operating system information, as well as Greenplum Database information from the same port (161) and IP address. It also enables the auto-discovery of Greenplum Database instances.

However, if required, you can run the Greenplum SNMP agent as a stand-alone agent and use the gpssnmpd agent or SNMP notification features independently of each other. As a stand-alone SNMP agent, gpssnmpd listens (on a network socket) for SNMP queries, and requires the same extensive configuration as the system SNMP agent.

Greenplum recommends that you run gpssnmpd as a sub-agent to the system agent. When it starts, the gpssnmpd sub-agent registers itself with the system-level SNMP agent, and communicates to the system agent the parts of the MIB of which it is
The system agent communicates with the SNMP client/network monitoring application and forwards requests for particular sections of the MIB to the \texttt{gpsnmpd} sub-agent.

To get information about a Greenplum Database instance, \texttt{gpsnmpd} logs into Greenplum Database through a normal \texttt{libpq} client session. SNMP GetRequests trigger \texttt{SELECT} statements against Greenplum Database to get the requested information. It is important to note that \texttt{gpsnmpd} does not send SNMP trap notifications; Greenplum Database itself sends this information to the network monitor application.

The \texttt{gpsnmpd} sub-agent communicates with the system agent through UNIX sockets; it does not listen on network sockets when used as a sub-agent.

\textbf{Prerequisites}

Before setting up SNMP support on Greenplum Database, make sure SNMP is installed on the master host. If the \texttt{snmpd} file is not present in the \texttt{/usr/sbin} directory (\texttt{/usr/sfw/sbin/} on Solaris), then SNMP is not installed on the system. Depending on the platform on which you are running Greenplum Database, install the following:

\begin{table}[h!]
\centering
\begin{tabular}{|l|l|}
\hline
\textbf{Operating System} & \textbf{Packages} \textsuperscript{1} \\
\hline
Red Hat Enterprise & net-snmp  \\
& net-snmp-libs  \\
& net-snmp-utils  \\
\hline
CentOS & net-snmp  \\
\hline
SUSE, Solaris, OSX & N/A  \\
\hline
\end{tabular}
\end{table}

\textsuperscript{1.} SNMP is installed by default on SUSE, Solaris, and OSX platforms.

The \texttt{snmp.conf} configuration file is located in \texttt{/etc/snmp/}. On Solaris platforms, the \texttt{snmp.conf} file is in \texttt{/etc/sma/snmp/}.

\textbf{Pre-installation Tasks}

After installing SNMP (or determining that SNMP is already installed), log in as \texttt{root}, open a text editor, and edit the \texttt{path_to/snmp/snmpd.conf} file. To use SNMP with Greenplum Database, the minimum configuration change required to the \texttt{snmpd.conf} file it to specify a community name. For example:

\begin{verbatim}
rocommunity public
\end{verbatim}

\textbf{Note:} Replace \texttt{public} with the name of your SNMP community. Greenplum also recommends configuring \texttt{syslocation} and \texttt{syscontact}. Configure other SNMP settings as required for your environment and save the file.

For more information about the \texttt{snmpd.conf} file, enter:

\begin{verbatim}
man snmpd.conf
\end{verbatim}
After you finish configuring the `snmpd.conf` file, start the system `snmpd` daemon:

```bash
# /sbin/chkconfig snmpd on
```

Then, verify that the system `snmpd` daemon is running. Enter:

```bash
# snmpwalk -v 1 -c community_name localhost .1.3.6.1.2.1.1.1.0
```

For example:

```bash
# snmpwalk -v 1 -c public localhost .1.3.6.1.2.1.1.1.0
```

If this command returns “Timeout: No Response from localhost”, then the system `snmpd` daemon is not running. If the daemon is running, output similar to the following displays:

```plaintext
SNMPv2-MIB::sysDescr.0 = STRING: Linux hostname
2.6.18-92.el5 #1 SMP Tue Jun 10 18:51:06 EDT 2010 x86_64
```

## Installing and Configuring Greenplum’s SNMP Agent

The following information describes how to configure the Greenplum SNMP agent (`gpsnmpd`) to collect and return database information to a network monitor.

If required, `gpsnmpd` can run as a stand-alone agent on a port other than port 161.

1. Log in as `root` and source the `greenplum_path.sh` file from your Greenplum installation.
   ```bash
   $ su -
   # source $GPHOME/greenplum_path.sh
   ```

2. Determine the default location for MIBs on the system:
   ```bash
   # net-snmp-config --default-mibdirs
   ```
   Typically, the default location is `path_to/snmp/mibs`.

3. Copy `NETWORK-SERVICES-MIB.txt`, `GPDB-MIB.txt`, and `RDBMS-MIB.txt` from the `$GPHOME/share/postgres` directory to the default location for MIBs on the system. For example:
   ```bash
   # cp NETWORK-SERVICES-MIB.txt GPDB-MIB.txt RDBMS-MIB.txt \ path_to/snmp/mibs
   ```

4. Determine how you want to use `gpsnmpd`. If you want run `gpsnmpd` as a sub-agent to the existing SNMP agent, open `path_to/snmp/snmpd.conf` in a text editor and enter the following line:
   ```bash
   master agentx
   ```
   On SUSE Linux, enter:
   ```bash
   agentXSocket /var/run/agentx/master
   ```

---

**Note:** On SUSE Linux platforms, make sure to review and configure security settings in the `snmp.conf` file so that `snmpd` accepts connections from sub-agents and returns all available Object IDs (OIDs).
Greenplum recommends running `gpsnmpd` as a sub-agent. This allows the monitoring application to use standard SNMP ports to communicate with both the system agent and `gpsnmpd`. This enables you to monitor both system and Greenplum Database status and events.

Alternatively, you can run `gpsnmpd` as stand-alone agent. To do this, skip this step and go to Step 7b.

5. Perform platform-specific tasks:
   a. On SUSE Linux platforms, create the following link:
      ```bash
      /var $ ln -s /var/run/agentx /var/agent
      ```
   b. On Solaris platforms, start the System Management Agent:
      ```bash
      enable svc:/application/management/sma
      ```

6. Restart the `snmpd` daemon.
   ```bash
   # /etc/init.d/snmpd restart
   ```

7. Start `gpsnmpd`.
   a. To start `gpsnmpd` as an agentx subagent, enter the following command (as root):
      ```bash
      # gpsnmpd -s -b -C "dbname=postgres user=username \ password=password"
      ```
      For example:
      ```bash
      # gpsnmpd -s -b -C "dbname=postgres user=gpadmin \ password=secret"
      ```
   b. To start `gpsnmpd` as a stand-alone agent, enter the following command (as root):
      ```bash
      # gpsnmpd -b -c path_to/snmp/snmpd.conf -x \
      nic_ip_address:port -C "dbname=postgres user=username \ password=password"
      ```
      For example:
      ```bash
      # gpsnmpd -b -c /etc/snmp/snmpd.conf -x \
      192.168.100.24:10161 -C "dbname=postgres user=gpadmin \ password=secret"
      ```

Greenplum recommends using the `postgres` database in the connection string (dbname=postgres).

You do not need to specify the `-C` option if you create a database role (user id) called `root`, and add the following line in the `pg_hba.conf` file:

```bash
local postgres root ident
```

This allows the UNIX user `root` to connect to the `postgres` database over the local connection. The `root` user does not require special permissions. The `user` and `password` parameters are only required when starting `gpsnmpd` as a user other than `root`.

Note that you can specify any configuration file to run `gpsnmpd` as a stand-alone agent; you do not have to use the `path_to/snmp/snmpd.conf` file. The configuration file you use must include a value for `rocommunity`. 
The `-x` option allows you to specify an IP address for a network interface card on the host, and specify a port other than the default SNMP port of 161. This enables you to run `gpsnmpd` without root permissions (you must have root permissions to use ports 1024 and lower). You do not need to specify this option if you are running `gpsnmpd` as an AgentX sub-agent (`-s`).

By default, `gpsnmpd` loads a default set of MIBs. However, you can use `-m` option to load specific MIBs when starting `gpsnmpd`. You can also use `-M` option to specify a list of directories from which you want to load MIBs.

**Note:** Greenplum recommends not entering a password in the `gpsnmpd` command as shown in Step 7, or using the `PGPASSWORD` environment variable. Some operating systems allow non-root users to see process environment variables and passwords through the `ps` command.

Instead, use the `.pgpass` file. If you do not specify a password in the `gpsnmpd` command, the agent reads from the `.pgpass` file in the home directory of the user you specify in the command. See the The Password File section in the PostgreSQL documentation for more information.

8. To verify that the `gpsnmpd` agent is enabled and responding, perform the following tests:

   a. Test server access to the Greenplum Database:

```
# snmpwalk -c communityname hostname:161 -v2c \
RDBMS-MIB::rdbmsRelState
```

You should see the following output:

```
Resetting connectionapplIndex for 0 is 0
applIndex for 1 is 1
RDBMS-MIB::rdbmsRelState.1.1 = INTEGER: active(2)
RDBMS-MIB::rdbmsRelState.2.0 = INTEGER: active(2)
RDBMS-MIB::rdbmsRelState.2.1 = INTEGER: active(2)
```

   b. Verify that Greenplum Database appears in the `rdbmsDbTable`. This table describes each database monitored by `gpsnmpd` agent.

```
# snmpwalk -c communityname hostname:161 -v2c \
RDBMS-MIB::rdbmsDbTable
```

You should see output similar to the following:

```
... 
RDBMS-MIB::rdbmsDbPrivateMibOID.10888 = OID: SNMPv2-SMI::enterprises.31327.10888
RDBMS-MIB::rdbmsDbPrivateMibOID.10889 = OID: SNMPv2-SMI::enterprises.31327.10889
RDBMS-MIB::rdbmsDbVendorName.1 = STRING: Greenplum Corporation
RDBMS-MIB::rdbmsDbVendorName.10888 = STRING: Greenplum Corporation
RDBMS-MIB::rdbmsDbVendorName.10889 = STRING: Greenplum Corporation
RDBMS-MIB::rdbmsDbName.1 = STRING: postgres
```
RDBMS-MIB::rdbmsDbName.10888 = STRING: template0
RDBMS-MIB::rdbmsDbName.10889 = STRING: postgres

9. In your network monitor application (such as Nagios, Cacti, OpenView), import RDBMS-MIB.txt, GPDB-MIB.txt, and NETWORK-SERVICES-MIB.txt. Make sure to specify the host name of the Greenplum master in your monitoring application.

Setting up SNMP Notifications

1. To configure a Greenplum Database system to send SNMP notifications when alerts occur, open the postgresql.conf file on the master host, and uncomment and set the following parameters:
   - **gp_snmp_community**: Set this parameter to the community name you specified for your environment.
   - **gp_snmp_monitor_address**: Enter the hostname:port of your network monitor application. Typically, the port number is 162. If there are multiple monitor addresses, separate them with a comma.
   - **gp_snmp_use_inform_or_trap**: Enter either `trap` or `inform`. Trap notifications are SNMP messages sent from one application to another (for example, between Greenplum Database and a network monitoring application). These messages are unacknowledged by the monitoring application, but generate less network overhead. Inform notifications are the same as trap messages, except that the application sends an acknowledgement to the application that generated the alert. In this case, the monitoring application sends acknowledgement messages to Greenplum Database-generated traps. While informs create more overhead, they inform Greenplum Database that the traps are being received by the monitoring application.

2. To test SNMP notifications, you can use the `snmptrapd` trap receiver. As root, enter:
   ```bash
   # /usr/sbin/snmptrapd -m ALL -Lf ~filename.log
   ```
   On Solaris, enter:
   ```bash
   # /usr/sfw/sbin/snmptrapd -m ALL -Lf ~filename.log
   ```
   `-Lf` indicates that traps are logged to a file. `-Le` indicates that traps are logged to `stderr` instead. `-m ALL` loads all available MIBs (you can also specify specific MIBs if required).

Enabling Email Notifications

To enable Greenplum Database to send email notifications to system administrators whenever certain database events occur, do the following steps.

1. Open `$MASTER_DATA_DIRECTORY/postgresql.conf` in a text editor.

2. In the `EMAIL ALERTS` section, uncomment the following parameters and enter the appropriate values for your email server and domain. For example:
   ```bash
   gp_email_smtp_server='smtp.company.com:25'
gp_email_smtp_userid='gpadmin@company.com'
   ```
gp_email_smtp_password='mypass'
gp_email_from='Greenplum Database <gpadmin@company.com>'
gp_email_to='dba@company.com;John Smith <jsmith@company.com>'

You might want to create specific email accounts/groups in your email system that send and receive email alerts from the Greenplum Database system. For example:

   gp_email_from='GPDB Production Instance <gpdb@company.com>'
gp_email_to='gpdb_dba_group@company.com'

You can also specify multiple email addresses for both gp_email parameters. Use a semi-colon ( ; ) to separate each email address. For example:

   gp_email_to='gpdb_dba_group@company.com';'admin@company.com'

3. Save and close the postgresql.conf file.

4. Restart Greenplum Database:
   $ gpstop -r

Testing Email Notifications

The Greenplum Database master host must be able to connect to the SMTP email server you specify for the gp_email_smtp_server parameter. To test connectivity, use the ping command:

   $ ping my_email_server

If the master host can contact the SMTP server, log in to psql and test email notifications with the following command:

   $ psql template1
  =# SELECT gp_elog('Test GPDB Email',true); gp_elog

The address you specified for the gp_email_to parameter should receive an email with Test GPDB Email in the subject line.

You can also test email notifications by using a public SMTP server, such as Google’s Gmail SMTP server, and an external email address. For example:

   gp_email_smtp_server='smtp.gmail.com:25'
   #gp_email_smtp_userid=''
   #gp_email_smtp_password=''
   gp_email_from='gpadmin@company.com'
gp_email_to='test_account@gmail.com'

Note: If you are having difficulty sending and receiving email notifications, verify the security settings for your organization’s email server and firewall.
Checking System State

A Greenplum Database system is comprised of multiple PostgreSQL instances (the master and segments) spanning multiple machines. To monitor a Greenplum Database system, you need to know information about the system as a whole, as well as status information of the individual instances. The `gpstate` utility provides status information about a Greenplum Database system.

Viewing Master and Segment Status and Configuration

The default behavior of `gpstate` is to check the segment instances and show a brief status of the valid and failed segments. For example, to see a quick status of the state of your Greenplum Database system:

```
$ gpstate
```

To see more detailed information about your Greenplum array configuration, use `gpstate` with the `-s` option:

```
$ gpstate -s
```

Viewing Your Mirroring Configuration and Status

If you are using mirroring for data redundancy, you may want to see the list of mirror segment instances in the system, their current synchronization status, and the mirror to primary mapping. For example, to see the mirror segments in the system and their status:

```
$ gpstate -m
```

To see the primary to mirror segment mappings:

```
$ gpstate -c
```

To see the status of the standby master mirror:

```
$ gpstate -f
```

Checking Disk Space Usage

The most important monitoring task of a database administrator is to make sure that the file systems where the master and segment data directories reside do not grow to more than 70 percent full. A filled data disk will not result in data corruption, but it may prevent normal database activity from occurring. If the disk grows too full, it can cause the database server to panic and consequent shutdown may occur.

You can use the `gp_disk_free` external table in the `gp_toolkit` administrative schema to check for remaining free space (in bytes) on the segment host file systems. For example:

```
=# SELECT * FROM gp_toolkit.gp_disk_free
    ORDER BY dfsegment;
```
Checking Sizing of Distributed Databases and Tables

The `gp_toolkit` administrative schema contains a number of views that can be used to determine the disk space usage for a distributed Greenplum database, schema, table, or index. See “Checking Database Object Sizes and Disk Space” on page 933 for the list of all available sizing views.

Viewing Disk Space Usage for a Database

To see the total size of a database (in bytes), use the `gp_size_of_database` view in the `gp_toolkit` administrative schema. For example:

```
=> SELECT * FROM gp_toolkit.gp_size_of_database
    ORDER BY soddatname;
```

Viewing Disk Space Usage for a Table

The `gp_toolkit` administrative schema contains a number of views for checking the size of a table. The table sizing views list the table by object ID (not by name). To check the size of a table by name, you must look up the relation name (`relname`) in the `pg_class` table. For example:

```
=> SELECT relname as name, sotdsize as size, sotdtoastsize as toast, sotdadditionalsize as other
    FROM gp_size_of_table_DISK as sotd, pg_class
    WHERE sotd.sotdoid=pg_class.oid ORDER BY relname;
```

See “Checking Database Object Sizes and Disk Space” on page 933 for the list of all available table sizing views.

Viewing Disk Space Usage for Indexes

The `gp_toolkit` administrative schema contains a number of views for checking index sizes. To see the total size of all index(es) on a table, use the `gp_size_of_all_table_indexes` view. To see the size of a particular index, use the `gp_size_of_index` view. The index sizing views list tables and indexes by object ID (not by name). To check the size of an index by name, you must look up the relation name (`relname`) in the `pg_class` table. For example:

```
=> SELECT soisize, relname as indexname
    FROM pg_class, gp_size_of_index
    WHERE pg_class.oid=gp_size_of_index.sooid
    AND pg_class.relkind='i';
```

Checking for Data Distribution Skew

All tables in Greenplum Database are distributed, meaning their data is divided across all of the segments in the system. If the data is not distributed evenly, then query processing performance may suffer. A table’s distribution policy is determined at table creation time. See “Choosing the Table Distribution Policy” on page 70 for more information.

- Viewing a Table’s Distribution Key
- Viewing Data Distribution
- Checking for Query Processing Skew
The `gp_toolkit` administrative schema also contains a number of views for checking data distribution skew on a table. See “Checking for Uneven Data Distribution” on page 937.

### Viewing a Table’s Distribution Key
To see the columns used as the data distribution key for a table, you can use the `\d+` meta-command in `psql` to examine the definition of a table. For example:

```sql
=# \d+ sales
```

<table>
<thead>
<tr>
<th>Column</th>
<th>Type</th>
<th>Modifiers</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sale_id</td>
<td>integer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>amt</td>
<td>float</td>
<td></td>
<td></td>
</tr>
<tr>
<td>date</td>
<td>date</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Has OIDs: no

Distributed by: (sale_id)

### Viewing Data Distribution
To see the data distribution of a table’s rows (the number of rows on each segment), you can run a query such as:

```sql
=# SELECT gp_segment_id, count(*)
    FROM table_name GROUP BY gp_segment_id;
```

A table is considered to have a balanced distribution if all of the segments have roughly the same number of rows.

### Checking for Query Processing Skew
When a query is being processed, you want all of the segments to handle an equal amount of the query workload to get the best possible performance. In some cases, query processing workload can be skewed if the table’s data distribution policy and the query predicates are not well matched. To check for processing skew, you can run a query such as:

```sql
=# SELECT gp_segment_id, count(*)
    FROM table_name
    WHERE column='value'
    GROUP BY gp_segment_id;
```

This will show the number of rows returned by segment for the given `WHERE` predicate.

### Viewing Metadata Information about Database Objects
Greenplum Database tracks various metadata information in its system catalogs about the objects stored in a database, such as tables, views, indexes and so on, as well as global objects such as roles and tablespaces. For example, administrators may want to examine information about a table, such as its definition, when it was created or the last operation performed.
Viewing the Last Operation Performed

The system views `pg_stat_operations` and `pg_stat_partition_operations` can be used to look up actions performed on an object, such as a table. For example, to see the actions performed on a table, such as when it was created and when it was last vacuumed and analyzed:

```sql
=> SELECT schemaname as schema, objname as table, username as role, actionname as action, subtype as type, statime as time
    FROM pg_stat_operations
    WHERE objname='cust';
```

<table>
<thead>
<tr>
<th>schema</th>
<th>table</th>
<th>role</th>
<th>action</th>
<th>type</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>sales</td>
<td>cust</td>
<td>main</td>
<td>CREATE</td>
<td>TABLE</td>
<td>2010-02-09 18:10:07.867977-08</td>
</tr>
<tr>
<td>sales</td>
<td>cust</td>
<td>main</td>
<td>VACUUM</td>
<td></td>
<td>2010-02-10 13:32:39.068219-08</td>
</tr>
<tr>
<td>sales</td>
<td>cust</td>
<td>main</td>
<td>ANALYZE</td>
<td></td>
<td>2010-02-25 16:07:01.157168-08</td>
</tr>
</tbody>
</table>

(3 rows)

Viewing the Definition of an Object

To see the definition of an object, such as a table or view, you can use the `\d+` meta command when working in `psql`. For example to see the definition of a table:

```sql
=> \d+ mytable
```

Viewing the Database Server Log Files

Every database instance in Greenplum Database (master and segments) is a running a PostgreSQL database server with its own server log file. Daily log files are created in the `pg_log` directory of the master and each segment data directory.

Log File Format

The server log files are written in comma-separated values (CSV) format. Not all log entries will have values for all of the log fields. For example, only log entries associated with a query worker process will have the `slice_id` populated. Related log entries of a particular query can be identified by its session identifier (`gp_session_id`) and command identifier (`gp_command_count`).

The following fields are written to the log:

<table>
<thead>
<tr>
<th>#</th>
<th>Field Name</th>
<th>Data Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><code>event_time</code></td>
<td>timestamp with time zone</td>
<td>Time that the log entry was written to the log</td>
</tr>
<tr>
<td>2</td>
<td><code>user_name</code></td>
<td>varchar(100)</td>
<td>The database user name</td>
</tr>
<tr>
<td>3</td>
<td><code>database_name</code></td>
<td>varchar(100)</td>
<td>The database name</td>
</tr>
<tr>
<td>4</td>
<td><code>process_id</code></td>
<td>varchar(10)</td>
<td>The system process id (prefixed with &quot;p&quot;)</td>
</tr>
<tr>
<td>#</td>
<td>Field Name</td>
<td>Data Type</td>
<td>Description</td>
</tr>
<tr>
<td>----</td>
<td>---------------------</td>
<td>------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>5</td>
<td>thread_id</td>
<td>varchar(50)</td>
<td>The thread count (prefixed with &quot;th&quot;)</td>
</tr>
<tr>
<td>6</td>
<td>remote_host</td>
<td>varchar(100)</td>
<td>The segment or master host name</td>
</tr>
<tr>
<td>7</td>
<td>remote_port</td>
<td>varchar(10)</td>
<td>The segment or master port number</td>
</tr>
<tr>
<td>8</td>
<td>session_start_time</td>
<td>timestamp with time zone</td>
<td>Time session connection was opened</td>
</tr>
<tr>
<td>9</td>
<td>transaction_id</td>
<td>int</td>
<td>Global transaction ID</td>
</tr>
<tr>
<td>10</td>
<td>gp_session_id</td>
<td>text</td>
<td>Session identifier number (prefixed with &quot;con&quot;)</td>
</tr>
<tr>
<td>11</td>
<td>gp_command_count</td>
<td>text</td>
<td>The command number within a session (prefixed with &quot;cmd&quot;)</td>
</tr>
<tr>
<td>12</td>
<td>gp_segment</td>
<td>text</td>
<td>The segment content identifier (prefixed with &quot;seg&quot; for primaries or &quot;mir&quot; for mirrors). The master always has a content id of -1.</td>
</tr>
<tr>
<td>13</td>
<td>slice_id</td>
<td>text</td>
<td>The slice id (portion of the query plan being executed)</td>
</tr>
<tr>
<td>14</td>
<td>distr_tranx_id</td>
<td>text</td>
<td>Distributed transaction ID</td>
</tr>
<tr>
<td>15</td>
<td>local_tranx_id</td>
<td>text</td>
<td>Local transaction ID</td>
</tr>
<tr>
<td>16</td>
<td>sub_tranx_id</td>
<td>text</td>
<td>Subtransaction ID</td>
</tr>
<tr>
<td>17</td>
<td>event_severity</td>
<td>varchar(10)</td>
<td>Values include: LOG, ERROR, FATAL, PANIC, DEBUG1, DEBUG2</td>
</tr>
<tr>
<td>18</td>
<td>sql_state_code</td>
<td>varchar(10)</td>
<td>SQL state code associated with the log message</td>
</tr>
<tr>
<td>19</td>
<td>event_message</td>
<td>text</td>
<td>Log or error message text</td>
</tr>
<tr>
<td>20</td>
<td>event_detail</td>
<td>text</td>
<td>Detail message text associated with an error or warning message</td>
</tr>
<tr>
<td>21</td>
<td>event_hint</td>
<td>text</td>
<td>Hint message text associated with an error or warning message</td>
</tr>
<tr>
<td>22</td>
<td>internal_query</td>
<td>text</td>
<td>The internally-generated query text</td>
</tr>
<tr>
<td>23</td>
<td>internal_query_pos</td>
<td>int</td>
<td>The cursor index into the internally-generated query text</td>
</tr>
<tr>
<td>24</td>
<td>event_context</td>
<td>text</td>
<td>The context in which this message gets generated</td>
</tr>
<tr>
<td>25</td>
<td>debug_query_string</td>
<td>text</td>
<td>Query string with full detail for debugging</td>
</tr>
<tr>
<td>26</td>
<td>error_cursor_pos</td>
<td>int</td>
<td>The cursor index into the query string</td>
</tr>
<tr>
<td>27</td>
<td>func_name</td>
<td>text</td>
<td>The function in which this message is generated</td>
</tr>
<tr>
<td>28</td>
<td>file_name</td>
<td>text</td>
<td>The internal code file where the message originated</td>
</tr>
<tr>
<td>29</td>
<td>file_line</td>
<td>int</td>
<td>The line of the code file where the message originated</td>
</tr>
<tr>
<td>30</td>
<td>stack_trace</td>
<td>text</td>
<td>Stack trace text associated with this message</td>
</tr>
</tbody>
</table>
Searching the Greenplum Database Server Log Files

Greenplum provides a utility called `gplogfilter` that can be used to search through a Greenplum Database log file for entries matching the specified criteria. By default, this utility searches through the Greenplum master log file in the default logging location. For example, to display the last three lines of the master log file:

```
$ gplogfilter -n 3
```

You can also use `gplogfilter` to search through all segment log files at once by running it through the `gpssh` utility. For example, to display the last three lines of each segment log file:

```
$ gpssh -f seg_host_file
  => source /usr/local/greenplum-db/greenplum_path.sh
  => gplogfilter -n 3 /gpdata/gp*/pg_log/gpdb*.log
```

Using `gp_toolkit`

Greenplum provides an administrative schema called `gp_toolkit` that you can use to query the system catalogs, log files, and operating environment for system status information. The `gp_toolkit` schema contains a number of views that you can access using SQL commands. The `gp_toolkit` schema is accessible to all database users, although some objects may require superuser permissions. For convenience, you may want to add the `gp_toolkit` schema to your schema search path. For example:

```
=> ALTER ROLE myrole SET search_path TO myschema,gp_toolkit;
```

See Appendix I, “The `gp_toolkit` Administrative Schema” for a description of the views and their usages.
19. Routine System Maintenance Tasks

Greenplum Database, like any database software, requires that certain tasks be performed regularly to achieve optimum performance. The tasks discussed here are required, but they are repetitive in nature and can easily be automated using standard UNIX tools such as cron scripts. But it is the database administrator’s responsibility to set up appropriate scripts, and to check that they execute successfully.

- Routine Vacuum and Analyze
- Routine Reindexing
- Managing Greenplum Database Log Files

Routine Vacuum and Analyze

Because of the MVCC transaction concurrency model used in Greenplum Database, data rows that are deleted or updated still occupy physical space on disk even though they are not visible to any new transactions. If you have a database with lots of updates and deletes, you will generate a lot of expired rows. The VACUUM command also collects table-level statistics such as number of rows and pages, so it is also necessary to vacuum append-only tables. Vacuuming append-only tables should be instantaneous since there will be no space to reclaim. See “Vacuuming the Database” on page 100.

Transaction ID Management

Greenplum’s MVCC transaction semantics depend on being able to compare transaction ID (XID) numbers to determine visibility to other transactions. But since transaction IDs have limited size, a Greenplum system that runs for a long time (more than 4 billion transactions) would suffer transaction ID wraparound: the XID counter wraps around to zero, and all of a sudden transactions that were in the past appear to be in the future — which means their outputs become invisible. To avoid this, it is necessary to VACUUM every table in every database at least once every two billion transactions.

System Catalog Maintenance

Numerous database updates with CREATE and DROP commands can cause growth in the size of the system catalog that affects system performance. For example, after a large number of DROP TABLE statements, the overall performance of the system begins to degrade due to excessive data scanning during metadata operations on the catalog tables. Depending on your system, the performance loss may occur between thousands to tens of thousands of DROP TABLE statements.

Greenplum recommends that you regularly run a system catalog maintenance procedure to reclaim the space occupied by deleted objects. If a regular procedure has not been run for a long time, you may need to run a more intensive procedure to clear the system catalog. Both types of procedures are described in this section.
Regular System Catalog Maintenance

Greenplum recommends that you periodically run `VACUUM` on the system catalog to clear the space occupied by deleted objects. If numerous `DROP` statements are a part of regular database operations, it is safe and appropriate to run a system catalog maintenance procedure with `VACUUM` daily at off-peak hours. This can be done while the system is running and available.

The following example script performs a `VACUUM` of the Greenplum Database system catalog:

```bash
#!/bin/bash
DBNAME="<database_name>"
VCOMMAND="VACUUM ANALYZE"
psql -tc "select '$VCOMMAND' || ' pg_catalog.' || relname || ';' from pg_class a,pg_namespace b where a.relnamespace=b.oid and b.nspname= 'pg_catalog' and a.relkind='r'" $DBNAME | psql -a $DBNAME
```

Intensive System Catalog Maintenance

If a system catalog maintenance procedure has not been performed in a long time, the catalog may become bloated with dead space, causing excessively long wait times for simple metadata operations. A wait of more than one or two seconds to list user tables, such as with the \d metacommand from within `psql`, is an indication of catalog bloat.

If you see indications of system catalog bloat, an intensive system catalog maintenance procedure with `VACUUM FULL` must be performed during a scheduled downtime period. During this period you must stop all catalog activity on the system due to the exclusive locks taken against the system catalog by the `FULL` system catalog maintenance procedure.

Running regular system catalog maintenance procedures can prevent the need for the more costly intensive procedure.

Vacuum and Analyze for Query Optimization

Greenplum Database uses a cost-based query planner that relies on database statistics. Accurate statistics allow the query planner to better estimate selectivity and the number of rows retrieved by a query operation in order to choose the most efficient query plan. The `ANALYZE` command collects column-level statistics needed by the query planner.

Both `VACUUM` and `ANALYZE` operations can be run in the same command. For example:

```
=# VACUUM ANALYZE mytable;
```

Routine Reindexing

For B-tree indexes, a freshly-constructed index is somewhat faster to access than one that has been updated many times, because logically adjacent pages are usually also physically adjacent in a newly built index. It might be worthwhile to reindex periodically to improve access speed. Also, if all but a few index keys on a page have
been deleted, there will be wasted space on the index page. A reindex will reclaim that wasted space. In Greenplum Database it is often faster to drop an index (DROP INDEX) and then recreate it (CREATE INDEX) than it is to use the REINDEX command.

Bitmap indexes are not updated when changes are made to the indexed column(s). If you have updated a table that has a bitmap index, you must drop and recreate the index for it to remain current.

Managing Greenplum Database Log Files

- Database Server Log Files
- Management Utility Log Files

Database Server Log Files

Greenplum Database log output tends to be voluminous (especially at higher debug levels) and you do not need to save it indefinitely. Administrators need to rotate the log files periodically so that new log files are started and old ones are removed after a reasonable period of time.

Greenplum Database has log file rotation enabled on the master and all segment instances. Daily log files are created in pg_log of the master and each segment data directory using the naming convention of: gpdb-YYYY-MM-DD.log. Although log files are rolled over daily, they are not automatically truncated or deleted. Administrators will need to implement some script or program to periodically clean up old log files in the pg_log directory of the master and each segment instance.

See also, “Viewing the Database Server Log Files” on page 207.

Management Utility Log Files

Log files for the Greenplum Database management utilities are written to ~/gpAdminLogs by default. The naming convention for management log files is:

<script_name>_<date>.log

The log entry format is:

<timestamp>:<utility>:<host>:<user>:[INFO|WARN|FATAL]:<message>

The log file for a particular utility execution is appended to its daily log file each time that utility is run.
Section V: Performance Tuning

This section describes the different performance tuning opportunities of Greenplum Database and how to troubleshoot certain performance problems. It contains the following topics:

- Defining Database Performance
- Common Causes of Performance Issues
- Investigating a Performance Problem
20. **Defining Database Performance**

For a data warehouse database system, such as Greenplum Database, database performance can be defined as the rate at which the database management system (DBMS) supplies information to those requesting it.

- **Understanding the Performance Factors**
- **Determining Acceptable Performance**

---

**Understanding the Performance Factors**

There are several factors that influence database performance. Understanding the key performance factors can help avoid performance problems or identify performance opportunities:

- **System Resources**
- **Workload**
- **Throughput**
- **Contention**
- **Optimization**

---

**System Resources**

Database performance relies heavily on disk I/O and memory usage. Knowing the baseline performance of the hardware on which your DBMS is deployed is essential is setting performance expectations. Performance of hardware components such as CPUs, hard disks, disk controllers, RAM, and network interfaces (and the interaction of these resources) will have a profound effect on how fast your database performs.

---

**Workload**

Your workload equals the total demand from the DBMS. The total workload is a combination of ad-hoc user queries, applications, batch jobs, transactions, and system commands directed through the DBMS at any given time. Workload, or demand, can change over time. For example, it may increase when month-end reports need to be run, or decrease on weekends when most users are out of the office. Workload is a major influence on database performance. Knowing your workload and peak demand times will help you plan for the most efficient use of your system resources, and enable the largest possible workload to be processed.

---

**Throughput**

A system’s throughput defines its overall capability to process data. DBMS throughput can be measured in queries per second, transactions per second, or average response times. DBMS throughput is closely related to the processing capacity of the
underlying systems (disk I/O, CPU speed, memory bandwidth, and so on), so it is important to know the throughput capacity of your hardware when setting DBMS throughput goals.

**Contention**

Contention is the condition in which two or more components of the workload are attempting to use the system in a conflicting way — for example, trying to update the same piece of data at the same time, or running multiple large workloads at once that compete with each other for system resources. As contention increases, throughput decreases.

**Optimization**

Optimizations that you make to your DBMS can affect the overall performance of your system. SQL formulation, database configuration parameters, table design, data distribution, and so on can enable the database query planner and optimizer to create the most efficient access plans.

**Determining Acceptable Performance**

When approaching a performance tuning initiative, it is important to know your system’s expected level of performance and to define measurable performance requirements. Without setting an acceptable threshold for database performance, you will end up chasing a carrot always out of reach. Consider the following when setting performance goals:

- **Baseline Hardware Performance**
- **Performance Benchmarks**

**Baseline Hardware Performance**

The majority of database performance problems are caused not by the database itself, but by the underlying systems on which the database is running. I/O bottlenecks, memory problems, and network problems can significantly degrade database performance. Therefore, it is important to know the baseline capabilities of your hardware and operating system (OS). This will help you to identify and troubleshoot hardware-related problems before undertaking database-level or query-level tuning initiatives. See “gpcheckperf” on page 600 for more information on running this utility to validate hardware and network performance.

**Performance Benchmarks**

In order to maintain good performance or improve upon performance issues, you need to know the capabilities of your DBMS on a defined workload. A benchmark is a predefined workload that produces a known result set, which can then be used for comparison purposes. Periodically running the same benchmark tests can help identify
system-related performance degradation over time. Benchmarks can also be used as a comparison to other workloads in an effort to identify queries or applications in need of optimization.

There are many third-party organizations which provide benchmark tools for the database industry, one of those being the Transaction Processing Performance Council (TPC). TPC-H is the ad-hoc, decision support benchmark. This benchmark illustrates decision support systems that examine large volumes of data, execute queries with a high degree of complexity, and give answers to critical business questions. For more information about TPC-H, go to:

http://www.tpc.org/tpch

This chapter describes the common causes of performance issues and potential solutions to these issues. The following issues are the most common cause of performance problems in Greenplum Database:

- Hardware failures and invalid segments
- Multiple workloads competing for system resources
- Contention between concurrent transactions
- Inaccurate database statistics
- Unbalanced data distribution across the segments
- Unoptimized database design

Identifying Hardware and Segment Failures

As with any database system, the performance of Greenplum Database is dependant upon the hardware and IT infrastructure on which it is running. Greenplum Database is comprised of several servers (or hosts) acting together as one cohesive system. Greenplum Database’s performance will be as fast as the slowest host in the array. Problems with CPU utilization, memory management, I/O processing, or network load will affect performance. Common hardware-related issues are:

- **Disk Failure** – Although a single disk failure should not dramatically effect database performance if you are using RAID, there is some impact caused by disk resynching consuming resources on the host with failed disks. The `gpcheckperf` utility can help identify segment hosts that have disk I/O issues.

- **Host Failure** – When a host is offline the segments on that host are out of operation. This means that other hosts in the array are doing double duty as they are running both the primary segments and a number of mirrors. If mirrors are not enabled – service is interrupted. There is also a temporary interruption of service to recover failed segments. The `gpstate` utility can help identify failed segments.

- **Network Failure** – Failure of a network interface card, a switch, or DNS server can bring down segments. If host names or IP addresses cannot be resolved within your Greenplum array, these manifest themselves as interconnect errors in Greenplum Database. The `gpcheckperf` utility can help identify segment hosts that have network issues.

- **Disk Capacity** – Disk capacity on your segment hosts should never exceed 70 percent full. Greenplum needs some free space for runtime processing. You can reclaim disk space occupied by deleted rows by running `VACUUM` after loads or updates. The `gp_toolkit` administrative schema has a number of views for checking the size of distributed database objects. See “Checking Database Object Sizes and Disk Space” on page 933.
Managing Workload

A database system has a limited capacity of CPU, memory, and disk I/O resources. When multiple workloads compete for access to these resources, database performance suffers. Workload management can be used to maximize system throughput while still meeting varied business requirements. Greenplum Database workload management limits the number of active queries in the system at any given time in order to avoid exhausting system resources. This is accomplished by creating role-based resource queues. A resource queue has attributes that limit the size and/or total number of queries that can be executed by the users (or roles) in that queue. By assigning all of your database roles to the appropriate resource queue, administrators can control concurrent user queries and prevent the system from being overloaded. See Chapter 8, “Managing Workload and Resources” for more information about setting up resource queues.

As the Greenplum Database administrator, run maintenance workloads such as data loads and VACUUM ANALYZE operations after regular business hours. Do not compete with database users for system resources by performing administrative tasks at peak usage times.

Avoiding Contention

Contention arises when two or more users or workloads try to use the system in a conflicting way. For example, if two transactions are trying to update the same table at once. A transaction seeking either a table-level or row-level lock will wait indefinitely for conflicting locks to be released. This means it is a bad idea for applications to hold transactions open for long periods of time (e.g., while waiting for user input).

Maintaining Database Statistics

Greenplum Database uses a cost-based query planner that relies on database statistics. Accurate statistics allow the query planner to better estimate the number of rows retrieved by a query in order to choose the most efficient query plan. Without database statistics, the query planner can not estimate how many records might be returned, and therefore cannot assume it has sufficient memory to perform certain operations such as aggregations. In this case, the planner always takes the safe route and does aggregations by reading/writing from disk, which is significantly slower than doing them in memory. The ANALYZE command collects statistics about the database needed by the query planner.

Identifying Statistics Problems in Query Plans

When looking at the query plan for a query using EXPLAIN or EXPLAIN ANALYZE, it helps to know your data in order to identify possible statistics problems. Check the plan for the following indicators of inaccurate statistics:

- Are the planner’s estimates close to reality? Run an EXPLAIN ANALYZE and see if the number of rows estimated by the planner is close to the number of rows actually returned by the query operation.
• **Are selective predicates applied early in the plan?** The most selective filters should be applied early in the plan so that less rows move up the plan tree.

• **Is the planner choosing the best join order?** When you have a query that joins multiple tables, make sure that the planner is choosing the most selective join order. Joins that eliminate the largest number of rows should be done earlier in the plan so that less rows move up the plan tree.

See also, “Query Profiling” on page 119 for more information on reading query plans.

---

**Tuning Statistics Collection**

The following configuration parameters control the amount of data sampled for statistics collection:

- `default_statistics_target`
- `gp_analyze_relative_error`

These parameters control statistics sampling at the system level. It is probably better to only sample increased statistics for the columns used most frequently in query predicates. You can adjust statistics for a particular column using the `ALTER TABLE ... SET STATISTICS` command. For example:

```
ALTER TABLE sales ALTER COLUMN region SET STATISTICS 50;
```

This is equivalent to increasing `default_statistics_target` for a particular column. Subsequent `ANALYZE` operations will then gather more statistics data for that column, and hopefully produce better query plans as a result.

---

**Optimizing Data Distribution**

When you create a table in Greenplum Database, it is important to declare a distribution key that allows for even data distribution across all segments in the system. Because the segments work on a query in parallel, Greenplum Database will always be as fast as the slowest segment. If the data is unbalanced, the segments that have more data will return their results slower.

---

**Optimizing Your Database Design**

Many performance issues can be improved by database design. Examine your database design and ask yourself the following:

- Does the schema reflect the way the data is accessed?
- Can larger tables be broken down into partitions?
- Are you using the smallest data type possible to store column values?
- Are columns used to join tables of the same datatype?
- Are your indexes being used?
Greenplum Database Maximum Limits

Knowing these maximum limits supported by Greenplum Database can help you optimize database design:

**Table 21.1** Maximum Limits of Greenplum Database

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database Size</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Table Size</td>
<td>128 TB per partition per segment</td>
</tr>
<tr>
<td>Row Size</td>
<td>1 GB</td>
</tr>
<tr>
<td>Field Size</td>
<td>1 GB</td>
</tr>
<tr>
<td>Rows per Table</td>
<td>$2^{48}$ (2 to the power of 48)</td>
</tr>
<tr>
<td>Columns per Table/View</td>
<td>1600</td>
</tr>
<tr>
<td>Indexes per Table</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Columns per Index</td>
<td>32</td>
</tr>
<tr>
<td>Table-level Constraints per Table</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Table Name Length</td>
<td>64 Bytes (Limited by name data type)</td>
</tr>
</tbody>
</table>

Dimensions listed as unlimited are not intrinsically limited by Greenplum Database. However, they are limited in practice to available disk space and memory/swap space. Performance may suffer when these values get unusually large.

**Note:** Limits for all database objects are limited by the number of available OIDs. There are 4 GB of OIDs in total, and all objects consume OIDs from a single namespace per database.
22. Investigating a Performance Problem

Many of the performance management steps taken by database administrators are reactive — a business user calls with a response time problem, a batch job fails, the system suddenly becomes unavailable. This section lists some steps an administrator can take to help identify the cause of a performance problem. If the problem is related to a particular workload or query, then you can focus your efforts on tuning that particular workload. If the performance problem is system-wide, then hardware problems, system failures, or resource contention may be the cause.

Checking System State

The `gpstate` utility can be used to identify failed segments. A Greenplum Database system will incur performance degradation when it has segment instances down because it requires some hosts to pick up the processing responsibilities of the downed segments.

Failed segments can be an indicator of some type of hardware failure, such as a failed disk drive or network card. Greenplum Database provides the hardware verification tool `gpcheckperf` to help identify segment hosts that have hardware issues.

Checking Database Activity

- Checking for Active Sessions (Workload)
- Checking for Locks (Contention)
- Checking Query Status and System Utilization

Checking for Active Sessions (Workload)

The `pg_stat_activity` system catalog view shows one row per server process, showing database OID, database name, process ID, user OID, user name, current query, time at which the current query began execution, time at which the process was started, and client address and port number. Querying this view can provide more information about the current workload on the system. For example:

```
SELECT * FROM pg_stat_activity;
```

This view should be queried as the database superuser to obtain the most information possible. Also note that the information does not update instantaneously.

Checking for Locks (Contention)

If a transaction is holding a lock on an object, there may be other queries that are waiting for that lock to be released before they can continue. This may appear to the user as if their query is hanging. The `pg_locks` system catalog view allows you to view information about outstanding locks. Examining `pg_locks` for ungranted locks can help identify contention between database client sessions. `pg_locks` provides a global
view of all locks in the database system, not only those relevant to the current database. Although its relation column can be joined against `pg_class.oid` to identify locked relations (such as tables), this will only work correctly for relations in the current database. The `pid` column can be joined to the `pg_stat_activity.procpid` to get more information on the session holding or waiting to hold a lock. For example:

```sql
SELECT locktype, database, c.relname, l.relation, l.transactionid, l.transaction, l.pid, l.mode, l.granted, a.current_query
FROM pg_locks l, pg_class c, pg_stat_activity a
WHERE l.relation=c.oid AND l.pid=a.procpid
ORDER BY c.relname;
```

If you are using resource queues for workload management, queries that are waiting in a queue will also show in `pg_locks`. To see how many queries are waiting to run from a particular resource queue, use the `pg_resqueue_status` system catalog view. For example:

```sql
SELECT * FROM pg_resqueue_status;
```

### Checking Query Status and System Utilization

System monitoring utilities such as `ps`, `top`, `iostat`, `vmstat`, `netstat` and so on can be used to monitor database activity on the hosts in your Greenplum Database array. These tools can be used to help identify Greenplum Database processes (`postgres` processes) currently running on the system and the most resource intensive tasks with regards to CPU, memory, disk I/O, or network activity. Looking at these system statistics can help identify queries that are overloading the system by consuming excessive resources and thereby degrading database performance. Greenplum Database comes with an management tool called `gpssh`, which allows you to run these system monitoring commands on several hosts at once.

The Greenplum Performance Monitor also collects query and system utilization metrics. See the [Greenplum Database Performance Monitor Guide](#) for information on enabling Greenplum Performance Monitor.

### Troubleshooting Problem Queries

If a query is performing poorly, looking at its query plan can help identify problem areas. You can use the `EXPLAIN` command to see the query plan for a given query. See “Query Profiling” on page 119 for more information on reading query plans and identifying problems with a plan.

### Investigating Error Messages

Greenplum Database log messages are written to files in the `pg_log` directory within the master’s or segment’s data directory. Log files are rolled over daily and are named using the naming convention: `gpdb-YYYY-MM-DD`. For example, to locate the log files on the master host:
$ cd $MASTER_DATA_DIRECTORY/pg_log

The master log file contains the most information and should always be checked first. Log lines have the format of:

```
timestamp | user | database | statement_id | con# | cmd# |:-- LOG_LEVEL: log_message
```

You may want to focus your search for `WARNING`, `ERROR`, `FATAL` or `PANIC` log level messages. The Greenplum utility, `gplogfilter`, can be used to search through Greenplum Database log files. For example, the following command (when run on the master host) will check for problem log messages in the standard logging locations:

```
$ gplogfilter -t
```

To search for any related log entries in the segment log files, you can run `gplogfilter` on the segment hosts using `gpssh`. One way to identify corresponding log entries is by the `statement_id` or `con#` (session identifier). For example, to search for log messages in the segment log files containing the string `con6` and save output to a file:

```
gpssh -f seg_hosts_file -e 'source /usr/local/greenplum-db/greenplum_path.sh ; gplogfilter -f con6 /gpdata/*/pg_log/gpdb*.log' > seglog.out
```

**Gathering Information for Greenplum Support**

The `gpdetective` utility collects information from a running Greenplum Database system and creates a bzip2-compressed tar output file. This output file can then be sent to Greenplum Customer Support to help with the diagnosis of Greenplum Database errors or system failures. Run `gpdetective` on your master host, for example:

```
$ gpdetective -f /var/data/my043008gp.tar
```
Section VI: Extending Greenplum Database

This section describes how to extend the functionality of Greenplum Database by developing your own functions and programs.

- Using Greenplum MapReduce
23 Using Greenplum MapReduce

- About Greenplum MapReduce
- Programming Greenplum MapReduce
- Submitting MapReduce Jobs for Execution
- Troubleshooting Problems with MapReduce Jobs

About Greenplum MapReduce

MapReduce is a programming model developed by Google for processing and generating large data sets on an array of commodity servers. Greenplum MapReduce allows programmers who are familiar with the MapReduce paradigm to write map and reduce functions and submit them to the Greenplum Database parallel data flow engine for processing. The Greenplum Database system takes care of the details of distributing the input data, executing the program across a set of machines, handling machine failures, and managing the required inter-machine communication.

The Basics of MapReduce

In general, MapReduce is a simple dataflow programming model that utilizes user-defined functions to pass data items from one stage of processing to the next. MapReduce programs typically start with a large data file that is broken down into smaller, contiguous pieces called splits, which are akin to database rows. Typically each split is parsed into (key, value) pairs that are then sent to a Map module. The Map module invokes a user-defined Map function on each pair and produces new (key, output_list) pairs. Each new (key, output_list) pair is then passed to a Reduce module. The Reduce module gathers these pairs together, grouping them by key, and then calls a user-defined Reduce function to produce one reduced output list for each distinct input key. Both the Map and Reduce modules utilize parallelism to enable many Map tasks and Reduce tasks to be worked on at the same time on a number of machines.
The MapReduce engine acts as an abstraction allowing programmers to focus on their desired data computations while hiding the details of parallelism, distribution, load balancing and fault tolerance.

**Figure 23.1** Basic MapReduce Data Flow

### How Greenplum MapReduce Works

In order for Greenplum to be able to process the Map and Reduce functions written by a user, the functions need to be defined in a specially formatted Greenplum MapReduce document, which is then passed to the Greenplum MapReduce program, `gpmapreduce`, for execution by the Greenplum Database parallel engine.

The Greenplum MapReduce document defines the parts that comprise a complete MapReduce job:

- **Input Data** - Input data can come from a number of sources either inside or outside the database. Greenplum Database supports a number of file formats for external data as well as SQL for data already stored in the database.

- **Map Function** - Users provide their own map function(s) written in Python, PERL, or C.

- **Reduce Function** - Users provide their own reduce function(s) written in Python, PERL, or C or use one of the built-in reduce functions.

- **Output Data** - Output can be persistently stored in the database, sent to a writable external table, or directed to standard output or an external file.
Once you have defined your MapReduce job in the specification document, you then use the `gpmapreduce` client program to submit your job to the Greenplum Database parallel data flow engine for execution.

## Programming Greenplum MapReduce

In order to submit a MapReduce job to Greenplum Database for execution, you must define your MapReduce job in a special Greenplum MapReduce specification document. Greenplum MapReduce uses the YAML document framework and implements its own YAML schema. See “Greenplum MapReduce Specification” on page 799.

This section walks through the stages of the Greenplum MapReduce specification and provides examples for each stage:

- Defining Inputs
- Defining Map Functions
- Defining Reduce Functions
- Defining Outputs
- Putting Together a Complete MapReduce Specification

These stages are specified in the `DEFINE` section of a Greenplum MapReduce specification document.

### Defining Inputs

Every MapReduce job requires at least one `INPUT` data source. A data source can be a single file, files served by the Greenplum parallel file distribution program (`gpfdist`), a table in the database, an SQL `SELECT` statement, or an operating system command that outputs data.

#### External File Inputs

A `FILE` input describes a single file located on a machine that is a Greenplum segment host. The file must be in either text-delimited or comma-separated values (CSV) format. See “Formatting Your Data Files” on page 125 for more information about the
format requirements of input data files. If COLUMNS (delimited fields in the file) are not specified, the entire file is treated as one big text column named value by default. You must be a Greenplum Database superuser to run MapReduce jobs with a FILE input.

```
- INPUT:
  NAME: my_file_input
  FILE: seghostname:/var/data/gpfiles/employees.txt
  COLUMNS
    - first_name text
    - last_name text
    - dept text
    - hire_date text
  FORMAT: TEXT
  DELIMITER: '|
```

A GPFDIST input is similar to FILE, except that the file is served by gpfdist, rather than the file system on a single segment host. See “Using the Greenplum Parallel File Server (gpfdist)” on page 129 for instructions on setting up a gpfdist file server. One advantage of using the gpfdist file server (as opposed to the FILE input) is that it ensures that all of the segments in your Greenplum Database system are fully utilized when reading the external data file(s). You must be a Greenplum Database superuser to run MapReduce jobs with GPFDIST input unless the server configuration parameter gp_external_grant_privileges is set to on.

```
- INPUT:
  NAME: my_distributed_input
  # specifies the host, port and the desired files served # by gpfdist. /* denotes all files on the gpfdist server
  GPFDIST:
    - gpfdisthost:8080/*
  COLUMNS
    - first_name text
    - last_name text
    - dept text
    - hire_date text
  FORMAT: TEXT
  DELIMITER: '|
```
Database Inputs

You can use data already stored in Greenplum Database as your MapReduce data source. A `TABLE` input simply takes all data from the table name you specify. The columns and data types are already defined by the table definition.

```
- INPUT:
  NAME: my_table_input
  TABLE: sales
```

Similarly, a `QUERY` input specifies a `SELECT` statement that returns selected data from one or more tables. The columns and data types are already defined by the sourced table columns.

```
- INPUT:
  NAME: my_query_input
  QUERY: SELECT vendor, amt FROM sales WHERE region='usa';
```

OS Executable Inputs

An `EXEC` input allows you to specify a shell command or script that will be executed by all Greenplum segments. The combined output of all the segment processes comprise your data source. The command is executed by all active segment instances on all segment hosts. For example, if each segment host has four primary segment instances running, the command will be executed four times per segment host. Data is comprised of the output of the command at the time the MapReduce job is executed on each segment instance. All segment instances execute the command in parallel. If your command calls a script or program, that executable must reside on all Greenplum segment hosts.

```
- INPUT:
  NAME: my_query_input
  EXEC: /var/load_scripts/get_log_data.sh
  COLUMNS
    - url text
    - date timestamp
  FORMAT: TEXT
  DELIMITER: '|' 
```

If you use environment variables in external web table commands (such as `$PATH`), keep in mind that the command is executed from within the database and not from a login shell. Therefore the `.bashrc` or `.profile` of the current user will not be sourced. However, you can set desired environment variables from within the `EXEC` definition itself, for example:

```
EXEC 'export PATH=/var/myscripts:$PATH; get_log_data.sh;'
```
In order to use this input type, you must be a Greenplum Database superuser and have the server configuration parameter `gp_external_enable_exec` set to `on` in your master `postgresql.conf` file.

---

**Defining Map Functions**

To borrow from database terminology, a Map function takes as input a single row (a set of values assigned to parameters), and produces zero or more rows of output. By default, the input and output are both defined to have two parameters of type `text`, called `key` and `value`. However, Greenplum MapReduce allows arbitrary parameter lists to be defined for both the input and the output, in the style of SQL table definitions. The input format is specified in the `PARAMETERS` definition in the `MAP` specification; the output format is specified in the `RETURNS` definition. The `RETURNS` definition requires each output parameter to be assigned a SQL data type for use in subsequent steps involving table outputs or SQL query inputs. When in doubt, SQL’s `text` data type will usually work fine, since both PERL and Python will interpret text strings appropriately.

The `MAP` specification also includes a `FUNCTION` definition that provides the code for the function, in a scripting language specified via `LANGUAGE`. Supported languages currently include `PERL` and `PYTHON`.

A typical Map function definition uses the values in the `PARAMETERS` in some way to generate appropriate output values matching the format of the `RETURNS` declaration. So the main issue for defining a Map function is to know how to access the `PARAMETERS` from within the scripting language, and how to prepare the output needed for `RETURNS`.

**Map Functions in PERL**

Following PERL style, the `PARAMETERS` to a Map function are available in the usual `@_` parameters list. A typical first step in a PERL function is to extract the parameters into local variables via an assignment statement.

The output of a Map function must be a PERL hash, with a hash-key for each parameter in the `RETURNS` definition. Output is typically returned to the MapReduce runtime engine via a special PERL function called `return_next`. This function behaves like a normal return, except that when the map code is re-invoked to generate another output row, it will pick up processing on the line after the last `return_next` that was executed (analogous to Python’s `yield` statement). This programming style makes it possible to take a single row as input, and return multiple outputs (each being passed back via `return_next` from within a PERL loop). When there are no more results to pass back, a standard PERL `return undef` call will tell the MapReduce harness to proceed with the next row of input, starting at the top of the Map function.

If you know that your Map function will only return one row of output for every input, you can specify the definition `MODE: SINGLE` in your `MAP` specification, and pass back a hash using a standard PERL `return` call, rather than `return_next`.
The following simple Map example converts a row containing a comma-separated value into multiple rows, one per value. Note the vertical bar (the YAML ‘literal’ marker) after the FUNCTION: declaration, indicating that the subsequent indented lines are to be considered a single literal string.

```yaml
- MAP:
  NAME: perl_splitter
  LANGUAGE: PERL
  PARAMETERS: [key, value]
  RETURNS: [key text, value text]
  FUNCTION: |
    my ($key, $value) = @_;  
    my @list = split(/,/, $value);  
    for my $item(@list) {
        return_next({"key" => $key, "value" => $item});
    }

    return undef;
```

**Map Functions in Python**

In Python, the PARAMETERS specified for a Map function are available as local Python variables. No PERL-style parameter interpretation is necessary.

The output of a Map function must be a (Python) dictionary, with a key for each parameter in the RETURNS definition. Output is typically returned to the MapReduce harness via the Python yield construct - but when the Map code is re-invoked to generate another output row, it will pick up processing on the line after the last yield that was executed. This programming style makes it possible to take a single row as input, and return multiple outputs (each being passed back via yield from within a Python loop). When there are no more results to pass back, the Python code should simply ‘drop through’ to the end of the script. This tells the MapReduce harness to proceed with the next row of input, starting at the top of the Map function.

If you know that your Map function will only return one row of output for every input, you can specify the definition MODE: SINGLE in your MAP specification, and pass back a hash using a standard Python return call, rather than yield.
The following simple Map example converts a row containing a comma-separated value into multiple rows, one per value. Note the vertical bar - a YAML ‘literal’ marker - after the `FUNCTION:` declaration, indicating that the subsequent indented lines are to be considered a single literal string.

```
- MAP:
  
  NAME: py_splitter
  LANGUAGE: PYTHON
  PARAMETERS: [key, value]
  RETURNS: [key text, value text]
  FUNCTION: |
          list = value.split(',')
          for item in list:
              yield {'key': key, 'value': item}
```

**Defining Reduce Functions**

Reduce functions handle a set of input rows that have matching values in a particular attribute (or set of attributes), and produce a single ‘reduced’ row.

Greenplum Database provides several predefined REDUCE functions you can execute, which all operate over a column named `value`:

- **IDENTITY** - returns (key, value) pairs unchanged
- **SUM** - calculates the sum of numeric data
- **AVG** - calculates the average of numeric data
- **COUNT** - calculates the count of input data
- **MIN** - calculates minimum value of numeric data
- **MAX** - calculates maximum value of numeric data

To use one of the predefined REDUCE jobs, you can simply declare it by name in the EXECUTE portion of your MapReduce specification document. For example:

```
EXECUTE
  - RUN

  SOURCE: input_or_task_name
  MAP: map_function_name

  REDUCE: IDENTITY
```

Writing custom Reduce functions is a bit more involved than writing Map functions, because the Reduce has to be defined to work through a set of input rows, not just a single row. To achieve this, you must define a TRANSITION function associated with the REDUCE, which is called once for each input row. In order to allow you to remember information between calls of the transition function, it takes as its first input parameter a variable called `state`. Before a set of tuples is to be Reduced, the `state` variable is initialized to the value specified in your INITIALIZE definition. This value
must be an SQL data type. For example, a (single-quoted) SQL text string. During the processing of a set, the \textit{state} variable records the most recent return value of the \texttt{TRANSITION} function. After the last row in the set is processed by the \texttt{TRANSITION} function, the \textit{state} variable is passed to the \texttt{FINALIZE} function, which can return multiple rows (via \texttt{PERL}'s \texttt{return\_next} or \texttt{Python}'s \texttt{yield}). Each row returned must be a hash representing the reduced output row.

By default, the parameters to a Reduce are \texttt{(key, value)} pairs. However, for custom Reduce functions, an arbitrary list of columns can be passed in. The \texttt{KEYS} definition defines the column or columns used to partition the input into subsets to be Reduced; the default value of the \texttt{KEYS} definition is the column called \texttt{key}. In the absence of a \texttt{KEYS} definition, the \texttt{key} is defined to be the set of parameters not mentioned in the \texttt{TRANSITION} function's \texttt{PARAMETERS} list.

As a performance optimization, you can optionally define a \texttt{CONSOLIDATE} function, which consolidates multiple \textit{state} variables into a single \textit{state} variable. This allows Greenplum Database to send a \textit{state} variable between machines in lieu of a set of input tuples, substantially lowering the amount of network traffic over the Greenplum interconnect. \texttt{CONSOLIDATE} is similar to \texttt{TRANSITION} in its structure, taking two \textit{state} variables at each invocation and returning a single \textit{state}.

Below is a complete \texttt{PERL} Reduce function definition for computing the average of all positive values:

\begin{quote}
\texttt{- REDUCE:}
\begin{verbatim}
NAME: perl_pos_avg
TRANSITION: perl_pos_avg_trans
CONSOLIDATE: perl_pos_avg_cons
FINALIZE: perl_pos_avg_final
INITIALIZE: '0,0'
KEYS: [key]
\end{verbatim}
\end{quote}

\begin{quote}
\texttt{- TRANSITION:}
\begin{verbatim}
NAME: perl_pos_avg_trans
PARAMETERS: [state, value]
RETURNS: [state text]
LANGUAGE: perl
FUNCTION: |
\begin{verbatim}
my ($state, $value) = @_;  
my ($count, $sum) = split('/,/,$state);  
if ($value > 0) {  
    $sum += $value;  
    $count++;  
    $state = $count . "," . $sum;  
}
return $state;
\end{verbatim}
\end{verbatim}
\end{quote}
- CONSOLIDATE:

   NAME: perl_pos_avg_cons
   PARAMETERS: [state, value]
   RETURNS: [state text]
   LANGUAGE: perl
   FUNCTION: |
   my ($state, $value) = @_; 
   my ($scount, $ssum) = split(/,,/, $state); 
   my ($vcount, $vsun) = split(/,,/, $value); 
   my $count = $scount + $vcount; 
   my $sum = $ssum + $vsun; 
   return ($count . "","" . $sum);

- FINALIZE:

   NAME: perl_pos_avg_final
   PARAMETERS: [state]
   RETURNS: [value float]
   LANGUAGE: perl
   FUNCTION: |
   my ($state) = @_; 
   my ($count, $sum) = split(/,,/, $state); 
   return_next ($count*1.0/$sum); 
   return undef;

### Defining Outputs

Defining an OUTPUT specification is optional. If no output is defined, the default is to send the final results to standard output of the Greenplum MapReduce client. You can also direct output to a file on the Greenplum MapReduce client or to a table in the database by defining an OUTPUT specification.

#### Table Outputs

A TABLE output defines a table in the database where the final output of your MapReduce job will be stored. By default, a table of the given TABLE name will be created in the database if it does not already exist. If the named table does exist in the database, you must declare a MODE to specify if you want to add the output to the table
(APPEND) or drop and recreate the table (REPLACE). By default, the table will be distributed by the REDUCE keys or you can optionally declare a distribution column using the KEYS specification.

- OUTPUT:
  - NAME: gpmr_output
  - TABLE: wordcount_out
  - KEYS:
    - value
  - MODE: REPLACE

An output table may also be a writable external table that pipes data out of the database and into a file or another program for further processing. If using writable external tables, the table must already be created prior to running your MapReduce job (see CREATE EXTERNAL TABLE). MODE must be set to APPEND for output tables that are writable external tables.

**File Outputs**

A FILE output defines a file location on the Greenplum MapReduce client where the output data will be written to. The named file will be created when the MapReduce job is run. The file cannot already exist or else the job will return an error.

- OUTPUT:
  - NAME: gpmr_output
  - FILE: /var/data/mapreduce/wordcount.out

**Defining Tasks**

A TASK specification is optional, but can be useful in multi-stage MapReduce jobs. A task defines a complete end-to-end INPUT/MAP/REDUCE stage within a complete Greenplum MapReduce job pipeline. Once defined, a TASK object can be called as input to further processing stages.
For example, suppose you have defined a table INPUT called documents and another called keywords. Each respective table input is processed by its own MAP function document_map and keyword_map. If you wanted to use the results of these processing stages as input to further stages in your MapReduce job, you could define two tasks as follows:

- TASK:
  NAME: document_prep
  SOURCE: documents
  MAP: document_map

- TASK:
  NAME: keyword_prep
  SOURCE: keywords
  MAP: keyword_map

These named tasks can then be called as input in a later processing stage. In this example, we are defining an SQL QUERY input that joins the results of the two tasks we defined earlier (document_prep and keyword_prep).

- INPUT:
  NAME: term_join
  QUERY: |
  SELECT doc.doc_id, kw.keyword_id, kw.term, kw.nterms,
          doc.positions as doc_positions,
          kw.positions as kw_positions
  FROM document_prep doc INNER JOIN keyword_prep kw
  ON (doc.term = kw.term)

**Putting Together a Complete MapReduce Specification**

Once you have defined all of the stages of your MapReduce job in the DEFINE section of your Greenplum MapReduce Specification document, you must define an EXECUTE section to specify the final INPUT/MAP/REDUCE stage. All of the objects named in the EXECUTE section are defined earlier in the Greenplum MapReduce specification DEFINE section.

EXECUTE:
- RUN:
  SOURCE: input_or_task_name
  TARGET: output_name
  MAP: map_function_name
  REDUCE: reduce_function_name
Submitting MapReduce Jobs for Execution

Once you have defined your MapReduce program in a Greenplum MapReduce Specification document, you can submit the document to Greenplum Database for execution using the `gpmapreduce` client program. This is a database client program similar to `psql` in that it requires you to supply connection information such as the database you are connecting to, the database user you are connecting as, the host and port of the Greenplum master, and so on. You can either specify the connection options on the command-line or use the environment variable settings `$PGDATABASE`, `$PGUSER`, `$PGHOST` and `$PGPORT` (if set). For example:

```
gpmapreduce -h gpmasterhost -p 54321 -f my_gpmr_spec.yml mydatabase
```

Creating the Languages in the Database

During execution of your map and reduce functions, Greenplum MapReduce makes use of the procedural languages built in to the database. Greenplum requires these languages to be created in the database prior to executing your MapReduce jobs. Use the `CREATE LANGUAGE` command to create the language in the database you will be using to execute your MapReduce jobs. You must be a database superuser to create a language. For example to create the PERL procedural language:

```
$ psql -c 'CREATE LANGUAGE plperl;
```

And to create the Python procedural language:

```
$ psql -c 'CREATE LANGUAGE plpythonu;
```

Installing Library Files (for Custom C Functions)

If you are using custom C map or reduce functions, the library files containing your function definitions must be installed in the same location on all Greenplum hosts (master and segments). For example, if you have a C function definition in your MapReduce specification YAML file such as:

```
DEFINE:
  - TRANSITION:
    NAME:       int4_accum
    PARAMETERS: [state int8, value int4]
    RETURNS:    [state int8]
    LANGUAGE:   C
    LIBRARY:    $libdir/gpmr.so
    FUNCTION:   int4_accum
```

The `$libdir/gpmr.so` file that is declared in the LIBRARY attribute must exist in `$GPHOME/lib` (the default library directory) on all Greenplum hosts. The FUNCTION attribute specifies the name of the function to call in the specified library file.
Troubleshooting Problems with MapReduce Jobs

This section describes some common errors encountered when executing Greenplum MapReduce jobs and how to resolve them.

Language Does Not Exist

**Symptom:**

```plaintext
ERROR:  language "pl*" does not exist
HINT:  Use CREATE LANGUAGE to load the language into the database.
```

**Explanation:**

During execution of your map and reduce functions, Greenplum MapReduce makes use of the procedural languages built in to the database. Greenplum requires these languages to be created in the database prior to executing your MapReduce jobs.

**Solution:**

If the language you are trying to use has not been created in the database you are connecting to, you (or your Greenplum administrator) will need to create the language for you before you can proceed. For example:

```plaintext
psql database_name -c 'CREATE LANGUAGE plperl;'
```

See [CREATE LANGUAGE](#) for more information. Some languages may require database superuser privileges to use.

Generic Python Iterator Error

**Symptom:**

```plaintext
ERROR: plpython: function "function_name" error fetching next item from iterator
DETAIL:  <type 'exceptions.IOError'>: [Errno 2] No such file or directory: '/tmp/file/doesnt/exist' (, line 39)
```

**Explanation:**

This is an error returned from Python that occurs whenever there was an error in an iterator function (such as a function which makes use of `yield`). This usually indicates a bug in your Python code.

**Solution:**

The easiest way to debug the problem is to wrap your function with additional exception handling. For example:

```plaintext
FUNCTION: |
    try: |
        ... |
        user code |
        ... |
    except Exception, e:
```
plpy.warning('my function name:' + str(e))

This will generate a warning message that is more useful for debugging.

### Function Defined Using Wrong MODE

**Symptom:**

- **ERROR:** set-returning PERL function must return reference to array or use return_next

```
ERR...------------------------------------------------------------------------
ERROR: composite-returning PERL function must return reference to hash
```

```
ERR...------------------------------------------------------------------------
ERROR: returned sequence's length must be same as tuple's length
```

```
ERR...------------------------------------------------------------------------
ERROR: no attribute named "key"
HINT: to return null in specific column, let returned object to have attribute named after column with value None
```

**Explanation:**

There are two primary modes that a Greenplum MapReduce function can be in:

- **MODE:** **SINGLE** returns precisely one row for every row received.
- **MODE:** **MULTI** may return any number of rows from 0-N for every row received.

TRANSITION and CONSOLIDATE functions only support **SINGLE** mode since they are effectively finite state functions and must return the next state.

MAP and FINALIZE functions support both modes defaulting to **MULTI** mode since it is more general.

**Solution:**

In a **SINGLE** mode function you must use the language’s return method to return a single row of data.

In **MULTI** mode it is best to write the function as a generator function. In Python this is done by using `yield`. In PERL this is done using `return_next`. For example:

```
- MAP:

  NAME: perl_single
  MODE: SINGLE
  PARAMETERS: [key text, value text]
  RETURNS: [key text, value text]
  LANGUAGE: perl
  FUNCTION: |
      my ($key, $value) = @_;
      return {'key' => $key, 'value' => $value}
```
The most common occurrence of the ‘returned sequence’s length ...’ error occurs when you try to use return from a MULTI mode function. In this context, the single row is interpreted not as a single row with columns, but as a set of rows each one column wide. Each column is too small to be a full row, so the returned sequence’s length does not match and an error is given.
A less common scenario that can generate the same error is when the number of columns returned does not match the declared number. For example:

```
- MAP:
  NAME:       python_error
  MODE:       SINGLE
  PARAMETERS: [key text, value text]
  RETURNS:    [key text, value text]
  LANGUAGE:   python
  FUNCTION:   |
               return {'key': key}
```

Because the function declared that it returns a key and a value but actually only returns a key, this can also cause the ‘returned sequence’s length . . . ’ error.
This section contains the following references:

- SQL Command Reference
- Management Utility Reference
- Client Utility Reference
- Server Configuration Parameters
- Greenplum MapReduce Specification
- Greenplum Environment Variables
- Greenplum Database Data Types
- System Catalog Reference
- Character Set Support
- SQL 2008 Optional Feature Compliance
- Glossary
A. SQL Command Reference

This appendix provides references for the SQL commands available in Greenplum Database:

- ABORT
- ALTER AGGREGATE
- ALTER CONVERSION
- ALTER DATABASE
- ALTER DOMAIN
- ALTER EXTERNAL TABLE
- ALTER FILESPACE
- ALTER FOREIGN DATA WRAPPER
- ALTER FOREIGN TABLE
- ALTER FUNCTION
- ALTER GROUP
- ALTER INDEX
- ALTER LANGUAGE
- ALTER OPERATOR
- ALTER OPERATOR CLASS
- ALTER RESOURCE QUEUE
- ALTER ROLE
- ALTER SCHEMA
- ALTER SEQUENCE
- ALTER SERVER
- ALTER TABLE
- ALTER TABLESPACE
- ALTER TRIGGER
- ALTER TYPE
- ALTER USER
- ALTER USER MAPPING
- ANALYZE
- BEGIN
- CHECKPOINT
- CLOSE
- CLUSTER
- COMMENT
- COMMIT
- COPY
- CREATE AGGREGATE
- CREATE CAST
- CREATE CONVERSION
- CREATE DATABASE
- CREATE DOMAIN
- CREATE EXTERNAL TABLE
- CREATE FOREIGN DATA WRAPPER
- CREATE FOREIGN TABLE
- CREATE FUNCTION
- CREATE GROUP
- CREATE INDEX
- CREATE LANGUAGE
- CREATE OPERATOR
- CREATE OPERATOR CLASS
- CREATE RESOURCE QUEUE
- CREATE ROLE
- CREATE RULE
- CREATE SCHEMA
- CREATE SEQUENCE
- CREATE SERVER
- CREATE TABLE
- CREATE TABLE AS
- CREATE TABLESPACE
- CREATE TRIGGER
- CREATE TYPE
- CREATE USER
- CREATE USER MAPPING*
- CREATE VIEW
- DEALLOCATE
- DECLARE
- DELETE
- DROP AGGREGATE
- DROP CAST
- DROP CONVERSION
- DROP DATABASE
- DROP DOMAIN
- DROP EXTERNAL TABLE
- DROP FILESPACE
- DROP FOREIGN DATA WRAPPER*
- DROP FOREIGN TABLE*
- DROP FUNCTION
- DROP GROUP
- DROP INDEX
- DROP LANGUAGE
- DROP OPERATOR
- DROP OPERATOR CLASS
- DROP OWNED
- DROP RESOURCE QUEUE
- DROP ROLE
- DROP RULE
- DROP SCHEMA
- DROP SEQUENCE
- DROP SERVER*
- DROP TABLE
- DROP TABLESPACE
- DROP TRIGGER
- DROP TYPE
- DROP USER

- DROP USER MAPPING*
- DROP VIEW
- END
- EXECUTE
- EXPLAIN
- FETCH
- GRANT
- INSERT
- LOAD
- LOCK
- MOVE
- PREPARE
- REASSIGN OWNED
- REINDEX
- RELEASE SAVEPOINT
- RESET
- REVOKE
- ROLLBACK
- ROLLBACK TO SAVEPOINT
- SAVEPOINT
- SELECT
- SELECT INTO
- SET
- SET ROLE
- SET SESSION AUTHORIZATION
- SET TRANSACTION
- SHOW
- START TRANSACTION
- TRUNCATE
- UPDATE
- VACUUM
- VALUES

* Not implemented in 4.0
**SQL Syntax Summary**

**ABORT**
Aborts the current transaction.

ABORT [WORK | TRANSACTION]

**ALTER AGGREGATE**
Changes the definition of an aggregate function

ALTER AGGREGATE name ( type [ , ... ] ) RENAME TO new_name
ALTER AGGREGATE name ( type [ , ... ] ) OWNER TO new_owner
ALTER AGGREGATE name ( type [ , ... ] ) SET SCHEMA new_schema

**ALTER CONVERSION**
Changes the definition of a conversion.

ALTER CONVERSION name RENAME TO newname
ALTER CONVERSION name OWNER TO newowner

**ALTER DATABASE**
Changes the attributes of a database.

ALTER DATABASE name [ WITH CONNECTION LIMIT conllimit ]
ALTER DATABASE name SET parameter { TO | = } { value | DEFAULT }
ALTER DATABASE name RESET parameter
ALTER DATABASE name RENAME TO newname
ALTER DATABASE name OWNER TO new_owner

**ALTER DOMAIN**
Changes the definition of a domain.

ALTER DOMAIN name ( SET DEFAULT expression | DROP DEFAULT )
ALTER DOMAIN name ( SET | DROP ) NOT NULL
ALTER DOMAIN name ADD domain_constraint
ALTER DOMAIN name DROP CONSTRAINT constraint_name [RESTRICT | CASCADE]
ALTER DOMAIN name OWNER TO new_owner
ALTER DOMAIN name SET SCHEMA new_schema

**ALTER EXTERNAL TABLE**
Changes the definition of an external table.

ALTER EXTERNAL TABLE name RENAME [COLUMN] column TO new_column
ALTER EXTERNAL TABLE name RENAME TO new_name
ALTER EXTERNAL TABLE name SET SCHEMA new_schema
ALTER EXTERNAL TABLE name action [, ... ]

where action is one of:
- ADD [COLUMN] column_name type
- DROP [COLUMN] column
- ALTER [COLUMN] column TYPE type [USING expression]
- OWNER TO new_owner
**ALTER FILESPACE**
Changes the definition of a filespace.

ALTER FILESPACE name RENAME TO newname
ALTER FILESPACE name OWNER TO newowner

**ALTER FUNCTION**
Changes the definition of a function.

ALTER FUNCTION name ( [ argmode ] [ argname ] argtype [, ... ] ) action [, ... ]
[RESTRICT]
ALTER FUNCTION name ( [ argmode ] [ argname ] argtype [, ... ] ) RENAME TO new_name
ALTER FUNCTION name ( [ argmode ] [ argname ] argtype [, ... ] ) OWNER TO new_owner
ALTER FUNCTION name ( [ argmode ] [ argname ] argtype [, ... ] ) SET SCHEMA new_schema

where action is one of:
{CALLED ON NULL INPUT | RETURNS NULL ON NULL INPUT | STRICT}
{IMMUTABLE | STABLE | VOLATILE}
{[EXTERNAL] SECURITY INVOKER | [EXTERNAL] SECURITY DEFINER}

**ALTER GROUP**
Changes a role name or membership.

ALTER GROUP groupname ADD USER username [, ... ]
ALTER GROUP groupname DROP USER username [, ... ]
ALTER GROUP groupname RENAME TO newname

**ALTER INDEX**
Changes the definition of an index.

ALTER INDEX name RENAME TO new_name
ALTER INDEX name SET TABLESPACE tablespace_name
ALTER INDEX name SET ( FILLFACTOR = value )
ALTER INDEX name RESET ( FILLFACTOR )

**ALTER LANGUAGE**
Changes the name of a procedural language.

ALTER LANGUAGE name RENAME TO newname

**ALTER OPERATOR**
Changes the definition of an operator.

ALTER OPERATOR name ( [ {lefttype | NONE} , [righttype | NONE} ) OWNER TO newowner

**ALTER OPERATOR CLASS**
Changes the definition of an operator class.

ALTER OPERATOR CLASS name USING index_method RENAME TO newname
ALTER OPERATOR CLASS name USING index_method OWNER TO newowner
**ALTER RESOURCE QUEUE**
Changes the limits of a resource queue.

```
ALTER RESOURCE QUEUE name WITH ( queue_attribute=value [, ... ] )
```

where `queue_attribute` is:
- `ACTIVE_STATEMENTS=integer`
- `MAX_COST=float`
- `COST_OVERCOMMIT={TRUE|FALSE}`
- `MIN_COST=float`
- `PRIORITY={MIN|LOW|MEDIUM|HIGH|MAX}`

```
ALTER RESOURCE QUEUE name WITHOUT ( queue_attribute=value [, ... ] )
```

where `queue_attribute` is:
- `ACTIVE_STATEMENTS=integer`
- `MAX_COST=float`
- `COST_OVERCOMMIT={TRUE|FALSE}`
- `MIN_COST=float`

**ALTER ROLE**
Changes a database role (user or group).

```
ALTER ROLE name RENAME TO newname
```

```
ALTER ROLE name SET config_parameter {TO | =} {value | DEFAULT}
```

```
ALTER ROLE name RESET config_parameter
```

```
ALTER ROLE name RESOURCE QUEUE {queue_name | NONE}
```

```
ALTER ROLE name [ [WITH] option [ ... ] ]
```

where `option` can be:
- `SUPERUSER | NOSUPERUSER`
- `CREATEDB | NOCREATEDB`
- `CREATEROLE | NOCREATEROLE`
- `CREATEEXTTABLE | NOCREATEEXTTABLE`
- `[ ( attribute='value'[, ... ] ) ]`

where attributes and values are:
- `type='readable'|'writable'
- `protocol='gpfdist'|'http'
- `INHERIT | NOINHERIT`
- `LOGIN | NOLOGIN`
- `CONNECTION LIMIT connlimit`
- `[ENCRYPTED | UNENCRYPTED] PASSWORD 'password'
- `VALID UNTIL 'timestamp'

**ALTER SCHEMA**
Changes the definition of a schema.

```
ALTER SCHEMA name RENAME TO newname
```

```
ALTER SCHEMA name OWNER TO newowner
```
**ALTER SEQUENCE**
Changes the definition of a sequence generator.

```
ALTER SEQUENCE name [INCREMENT [ BY ] increment]
    [MINVALUE minvalue | NO MINVALUE]
    [MAXVALUE maxvalue | NO MAXVALUE]
    [RESTART [ WITH ] start]
    [CACHE cache] [[ NO ] CYCLE]
    [OWNED BY {table.column | NONE}]
```

ALTER SEQUENCE name SET SCHEMA new_schema

**ALTER TABLE**
Changes the definition of a table.

```
ALTER TABLE [ONLY] name RENAME [COLUMN] column TO new_column
ALTER TABLE name RENAME TO new_name
ALTER TABLE name SET SCHEMA new_schema
ALTER TABLE [ONLY] name SET
    DISTRIBUTED BY (column, [ ... ])
    | DISTRIBUTED RANDOMLY
    | WITH (REORGANIZE=true|false)

ALTER TABLE [ONLY] name action [, ...]
ALTER TABLE name
    [ ALTER PARTITION { partition_name | FOR (RANK(number)) | FOR (value) }
```
partition_action [...] ]
    partition_action

where action is one of:
ADD [COLUMN] column_name type [column_constraint [ ... ]]
DROP [COLUMN] column [RESTRICT | CASCADE]
ALTER [COLUMN] column TYPE type [USING expression]
ALTER [COLUMN] column SET DEFAULT expression
ALTER [COLUMN] column DROP DEFAULT
ALTER [COLUMN] column { SET | DROP } NOT NULL
ALTER [COLUMN] column SET STATISTICS integer
ADD table_constraint
DROP CONSTRAINT constraint_name [RESTRICT | CASCADE]
DISABLE TRIGGER [trigger_name | ALL | USER]
ENABLE TRIGGER [trigger_name | ALL | USER]
CLUSTER ON index_name
SET WITHOUT CLUSTER
SET WITHOUT OIDS
SET (FILLFACTOR = value)
RESET (FILLFACTOR)
INHERIT parent_table
NO INHERIT parent_table
OWNER TO new_owner
SET TABLESPACE new_tablespace
ALTER DEFAULT PARTITION
DROP DEFAULT PARTITION [IF EXISTS]
DROP PARTITION [IF EXISTS] { partition_name | FOR (RANK(number)) | FOR (value) } [CASCADE]
TRUNCATE DEFAULT PARTITION
TRUNCATE PARTITION { partition_name | FOR (RANK(number)) | FOR (value) }
RENAME DEFAULT PARTITION TO new_partition_name
RENAME PARTITION { partition_name | FOR (RANK(number)) | FOR (value) } TO new_partition_name
ADD DEFAULT PARTITION [ ( subpartition_spec ) ]
ADD PARTITION [name] partition_element [ ( subpartition_spec ) ]
EXCHANGE PARTITION { partition_name | FOR (RANK(number)) | FOR (value) } WITH TABLE table_name [ WITH | WITHOUT VALIDATION ]
EXCHANGE DEFAULT PARTITION WITH TABLE table_name [ WITH | WITHOUT VALIDATION ]
SET SUBPARTITION TEMPLATE (subpartition_spec)
SPLIT DEFAULT PARTITION
{ AT (list_value)
  | START([datatype] range_value) [INCLUSIVE | EXCLUSIVE]
  | END([datatype] range_value) [INCLUSIVE | EXCLUSIVE] }
[ INTO ( PARTITION new_partition_name,
  PARTITION default_partition_name ) ]
SPLIT PARTITION { partition_name | FOR (RANK(number)) | FOR (value) } AT (value)
[ INTO (PARTITION partition_name, PARTITION partition_name)]

where partition_element is:
VALUES (list_value [ ,... ])
  | START ([datatype] 'start_value') [INCLUSIVE | EXCLUSIVE]
  | END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE] [ ,... ]
[ WITH ( partition_storage_parameter=value [, ... ] ) ]
[ TABLESPACE tablespace ]

where subpartition_spec is:
subpartition_element [, ...]

and subpartition_element is:
  DEFAULT SUBPARTITION name
  | [SUBPARTITION name] VALUES (list_value [, ...] )
  | [SUBPARTITION name]
    START ( [datatype] 'start_value') [INCLUSIVE | EXCLUSIVE]
    [ END ( [datatype] 'end_value') [INCLUSIVE | EXCLUSIVE] ]
    [ EVERY ( [number | datatype] 'interval_value') ]
  | [SUBPARTITION name]
    END ( [datatype] 'end_value') [INCLUSIVE | EXCLUSIVE]
    [ EVERY ( [number | datatype] 'interval_value') ]
[ WITH ( partition_storage_parameter=value [, ... ] ) ]
[ TABLESPACE tablespace ]

where storage_parameter is:
  APPENDONLY={TRUE|FALSE}
  COMPRESSTYPE={ZLIB|QUICKLZ}
  COMPRESSLEVEL={0-9|1}
  ORIENTATION={COLUMN|ROW}
  FILLFACTOR={10-100}

**ALTER TABLESPACE**
Changes the definition of a tablespace.

ALTER TABLESPACE name RENAME TO newname
ALTER TABLESPACE name OWNER TO newowner

**ALTER TRIGGER**
Changes the definition of a trigger.

ALTER TRIGGER name ON table RENAME TO newname

**ALTER TYPE**
Changes the definition of a data type.

ALTER TYPE name OWNER TO new_owner
ALTER TYPE name SET SCHEMA new_schema
**ALTER USER**
Changes the definition of a database role (user).

```
ALTER USER name RENAME TO newname
ALTER USER name SET config_parameter {TO | =} {value | DEFAULT}
ALTER USER name RESET config_parameter
ALTER USER name [ [WITH] option [ ... ] ]
```

where **option** can be:
- SUPERUSER | NOSUPERUSER
- CREATEDB | NOCREATEDB
- CREATEROLE | NOCREATEROLE
- CREATEUSER | NOCREATEUSER
- INHERIT | NOINHERIT
- LOGIN | NOLOGIN
- [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
- VALID UNTIL 'timestamp'

**ANALYZE**
Collects statistics about a database.

```
ANALYZE [VERBOSE] [table [ (column [, ...] ) ]]
```

**BEGIN**
Starts a transaction block.

```
BEGIN [WORK | TRANSACTION] [SERIALIZABLE | REPEATABLE READ | READ COMMITTED | READ UNCOMMITTED] [READ WRITE | READ ONLY]
```

**CHECKPOINT**
Forces a transaction log checkpoint.

```
CHECKPOINT
```

**CLOSE**
Closes a cursor.

```
CLOSE cursor_name
```

**CLUSTER**
Physically reorders a heap storage table on disk according to an index. Not a recommended operation in Greenplum Database.

```
CLUSTER indexname ON tablename
CLUSTER tablename
```
COMMENT
Defines or change the comment of an object.

COMMENT ON
{ TABLE object_name | COLUMN table_name.column_name | AGGREGATE agg_name (agg_type [, ...]) | CAST (sourcetype AS targettype) | CONSTRAINT constraint_name ON table_name | CONVERSION object_name | DATABASE object_name | DOMAIN object_name | FILESPACE object_name | FUNCTION func_name ([[argmode] [argname] argtype [, ...]]) | INDEX object_name | LARGE OBJECT large_object_oid | OPERATOR op (leftoperand_type, rightoperand_type) | OPERATOR CLASS object_name USING index_method | [PROCEDURAL] LANGUAGE object_name | RESOURCE QUEUE object_name | ROLE object_name | RULE rule_name ON table_name | SCHEMA object_name | SEQUENCE object_name | TABLESPACE object_name | TRIGGER trigger_name ON table_name | TYPE object_name | VIEW object_name }
IS 'text'

COMMIT
Commits the current transaction.

COMMIT [WORK | TRANSACTION]
COPY
Copies data between a file and a table.

COPY table [(column [, ...])] FROM '{file} | STDIN'
[ [WITH]
  [OIDS]
  [HEADER]
  [DELIMITER [ AS ] 'delimiter']
  [NULL [ AS ] 'null string']
  [ESCAPE [ AS ] 'escape' | 'OFF']
  [NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
  [CSV [QUOTE [ AS ] 'quote']
    [FORCE NOT NULL column [, ...]]
  [FILL MISSING FIELDS]
  [ LOG ERRORS INTO error_table] [KEEP]
  SEGMENT REJECT LIMIT count [ROWS | PERCENT] ]
COPY {table [(column [, ...])] | (query)} TO '{file} | STDOUT'
[ [WITH]
  [OIDS]
  [HEADER]
  [DELIMITER [ AS ] 'delimiter']
  [NULL [ AS ] 'null string']
  [ESCAPE [ AS ] 'escape' | 'OFF']
  [CSV [QUOTE [ AS ] 'quote']
    [FORCE QUOTE column [, ...]]
 CREATE AGGREGATE
Defines a new aggregate function.

CREATE AGGREGATE name (input_data_type [, ...])
  ( SFUNC = sfunc,
    STYPE = state_data_type
  , PREFUNC = prefunc] 
  [, FINALFUNC = ffunc] 
  [, INITCOND = initial_condition] 
  [, SORTOP = sort_operator] )

CREATE CAST
Defines a new cast.

CREATE CAST (sourcetype AS targettype)
  WITH FUNCTION funcname (argtypes)
  [AS ASSIGNMENT | AS IMPLICIT]
CREATE CAST (sourcetype AS targettype) WITHOUT FUNCTION
  [AS ASSIGNMENT | AS IMPLICIT]

CREATE CONVERSION
Defines a new encoding conversion.

CREATE [DEFAULT] CONVERSION name FOR source_encoding TO dest_encoding FROM funcname
CREATE DATABASE
Creates a new database.

```
CREATE DATABASE name [ [WITH] [OWNER [=] dbowner]
    [TEMPLATE [=] template]
    [ENCODING [=] encoding]
    [TABLESPACE [=] tablespace]
    [CONNECTION LIMIT [=] connlimit ] ]
```

CREATE DOMAIN
Defines a new domain.

```
CREATE DOMAIN name [AS] data_type [DEFAULT expression]
    [CONSTRAINT constraint_name
    | NOT NULL | NULL
    | CHECK (expression) [...] ]
```
CREATE EXTERNAL TABLE
Defines a new external table.

CREATE [READABLE] EXTERNAL TABLE table_name
( column_name data_type [, ...] | LIKE other_table )
LOCATION ('file://seghost[:port]/path/file' [, ...])
| ('gpfdist://filehost[:port]/file_pattern' [, ...])
FORMAT 'TEXT'

| ([ [HEADER]
|   [DELIMITER [AS] 'delimiter' | 'OFF']
|   [NULL [AS] 'null string']
|   [ESCAPE [AS] 'escape' | 'OFF']
|   [NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
|   [FILL MISSING FIELDS] )
| | 'CSV'
| ([ [HEADER]
|   [QUOTE [AS] 'quote']
|   [DELIMITER [AS] 'delimiter']
|   [NULL [AS] 'null string']
|   [FORCE NOT NULL column [, ...]]
|   [ESCAPE [AS] 'escape']
|   [NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
|   [FILL MISSING FIELDS] ]
| [ ENCODING 'encoding' ]
| [ [LOG ERRORS INTO error_table] SEGMENT REJECT LIMIT count
| [ ROWS | PERCENT ] ]

CREATE [READABLE] EXTERNAL WEB TABLE table_name
( column_name data_type [, ...] | LIKE other_table )
LOCATION ('http://webhost[:port]/path/file' [, ...])
| EXECUTE 'command' [ON ALL
| MASTER
| number_of_segments
| HOST ['segment_hostname']
| SEGMENT segment_id ]
FORMAT 'TEXT'

| ([ [HEADER]
|   [DELIMITER [AS] 'delimiter' | 'OFF']
|   [NULL [AS] 'null string']
|   [ESCAPE [AS] 'escape' | 'OFF']
|   [NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
|   [FILL MISSING FIELDS] ]
| | 'CSV'
| ([ [HEADER]
|   [QUOTE [AS] 'quote']
|   [DELIMITER [AS] 'delimiter']
|   [NULL [AS] 'null string']
|   [FORCE NOT NULL column [, ...]]
|   [ESCAPE [AS] 'escape']
|   [NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
|   [FILL MISSING FIELDS] ]
| [ ENCODING 'encoding' ]
| [ [LOG ERRORS INTO error_table] SEGMENT REJECT LIMIT count
| [ ROWS | PERCENT ] ]

CREATE WRITABLE EXTERNAL TABLE table_name
( column_name data_type [, ...] | LIKE other_table )
LOCATION ('gpfdist://outputhost[:port]/filename' [, ...])
FORMAT 'TEXT'
CREATE FUNCTION

Defines a new function.

CREATE [OR REPLACE] FUNCTION name
    ( [ [argmode] [argname] argtype [, ...] ] )
    [RETURNS [SETOF] rettype]
    { LANGUAGE langname
      | IMMUTABLE | STABLE | VOLATILE
      | CALLED ON NULL INPUT | RETURNS NULL ON NULL INPUT | STRICT
      | [EXTERNAL] SECURITY INVOKER | [EXTERNAL] SECURITY DEFINER
      | AS 'definition'
      | AS 'obj_file', 'link_symbol' } ...
CREATE GROUP
Defines a new database role.

CREATE GROUP name [ [WITH] option [ ... ] ]

where option can be:

  SUPERUSER | NOSUPERUSER
  | CREATEDB | NOCREATEDB
  | CREATEROLE | NOCREATEROLE
  | CREATEUSER | NOCREATEUSER
  | INHERIT | NOINHERIT
  | LOGIN | NOLOGIN
  | [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
  | VALID UNTIL 'timestamp'
  | IN ROLE rolename [, ...]
  | IN GROUP rolename [, ...]
  | ROLE rolename [, ...]
  | ADMIN rolename [, ...]
  | USER rolename [, ...]
  | SYSID uid

CREATE INDEX
Defines a new index.

CREATE [UNIQUE] INDEX name ON table
  [USING btree|bitmap|gist]
  ( {column | (expression) | [opclass] [, ...] }
  | WITH ( FILLFACTOR = value ) ]
  [TABLSPACE tablespace]
  [WHERE predicate]

CREATE LANGUAGE
Defines a new procedural language.

CREATE [PROCEDURAL] LANGUAGE name
CREATE [TRUSTED] [PROCEDURAL] LANGUAGE name
  HANDLER call_handler [VALIDATOR valfunction]

CREATE OPERATOR
Defines a new operator.

CREATE OPERATOR name ( 
  [PROCEDURE = funcname
  [, LEFTARG = lefttype] [, RIGHTARG = righttype]
  [, COMMUTATOR = com_op] [, NEGATOR = neg_op]
  [, RESTRICT = res_proc] [, JOIN = join_proc]
  [, HASHES] [, MERGES]
  [, SORT1 = left_sort_op] [, SORT2 = right_sort_op]
  [, LTCMP = less_than_op] [, GTCMP = greater_than_op] )
CREATE OPERATOR CLASS
Defines a new operator class.

CREATE OPERATOR CLASS name [DEFAULT] FOR TYPE data_type
  USING index_method AS
  {
    OPERATOR strategy_number op_name [(op_type, op_type)] [RECHECK]
    | FUNCTION support_number funcname (argument_type [, ...] )
    | STORAGE storage_type
  } [, ... ]

CREATE RESOURCE QUEUE
Defines a new resource queue.

CREATE RESOURCE QUEUE name WITH (queue_attribute=value [, ... ])
where queue_attribute is:
  ACTIVE_STATEMENTS=integer
    [ MAX_COST=Float [COST_OVERCOMMIT={TRUE|FALSE}] ]
    [ MIN_COST=Float ]
    [ PRIORITY={MIN|LOW|MEDIUM|HIGH|MAX} ]
  MAX_COST=Float [ COST_OVERCOMMIT={TRUE|FALSE} ]
    [ ACTIVE_STATEMENTS=integer ]
    [ MIN_COST=Float ]
    [ PRIORITY={MIN|LOW|MEDIUM|HIGH|MAX} ]

CREATE ROLE
Defines a new database role (user or group).

CREATE ROLE name [[WITH] option [ ... ]]
where option can be:
  SUPERUSER | NOSUPERUSER
  | CREATEDB | NOCREATEDB
  | CREATEROLE | NOCREATEROLE
  | CREATEEXTTABLE | NOCREATEEXTTABLE
  | ( attribute=value [, ... ] )
  where attributes and values are:
    type='readable'|'writable'
    protocol='gpfdist'|'http'
    INHERIT | NOINHERIT
    LOGIN | NOLOGIN
    CONNECTION LIMIT connlimit
  | [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
  | VALID UNTIL 'timestamp'
  | IN ROLE rolename [, ...]
  | ROLE rolename [, ...]
  | ADMIN rolename [, ...]
  | RESOURCE QUEUE queue_name

CREATE RULE
Defines a new rewrite rule.

CREATE [OR REPLACE] RULE name AS ON event
  TO table [WHERE condition]
  DO [ALSO | INSTEAD] { NOTHING | command | (command; command ... ) }
CREATE SCHEMA
Defines a new schema.

CREATE SCHEMA schema_name [AUTHORIZATION username] [schema_element [ ... ]]
CREATE SCHEMA AUTHORIZATION rolename [schema_element [ ... ]]

CREATE SEQUENCE
Defines a new sequence generator.

CREATE [TEMPORARY | TEMP] SEQUENCE name
  [INCREMENT [BY] value]
  [MINVALUE minvalue | NO MINVALUE]
  [MAXVALUE maxvalue | NO MAXVALUE]
  [START [ WITH ] start]
  [CACHE cache]
  [[NO] CYCLE]
  [OWNED BY { table.column | NONE }]}
CREATE TABLE
Defines a new table.

CREATE [GLOBAL | LOCAL] [TEMPORARY | TEMP] TABLE table_name (  
   [ { column_name data_type [DEFAULT default_expr]  
      [column_constraint [ ... ]]  
   | table_constraint  
   | LIKE other_table [(INCLUDING | EXCLUDING)  
      {DEFAULTS | CONSTRAINTS}] ...}  
   [, ... ] ]  
   [ INHERITS ( parent_table [, ... ] ) ]  
   [ WITH ( storage_parameter=value [, ... ] ) ]  
   [ ON COMMIT (PRESERVE ROWS | DELETE ROWS | DROP) ]  
   [ TABLESPACE tablespace ]  
   [ DISTRIBUTED BY (column, [ ... ] ) | DISTRIBUTED RANDOMLY ]  
   [ PARTITION BY partition_type (column) ]  
      [ SUBPARTITION BY partition_type (column) ]  
      [ SUBPARTITION TEMPLATE ( template_spec ) ]  
      [ ... ]  
      ( partition_spec )  
      [ SUBPARTITION BY partition_type (column) ]  
      [ ... ]  
      ( partition_spec )  
      [ ( subpartition_spec  
      [ (...)]  
      ) ]  
   )  

where storage_parameter is:
   APPENDONLY={TRUE|FALSE}  
   ORIENTATION={COLUMN|ROW}  
   COMPRESSTYPE={ZLIB|QUICKLZ}  
   COMPRESSIONLEVEL={0-9 | 1}  
   FILLFACTOR={10-100}  
   OIDS=[=TRUE|FALSE]

where column_constraint is:
   [CONSTRAINT constraint_name]  
   NOT NULL | NULL  
   | UNIQUE [USING INDEX TABLESPACE tablespace]  
   [ WITH ( FILLFACTOR = value ) ]  
   | PRIMARY KEY [USING INDEX TABLESPACE tablespace]  
   [ WITH ( FILLFACTOR = value ) ]  
   | CHECK ( expression )

and table_constraint is:
   [CONSTRAINT constraint_name]  
   UNIQUE ( column_name [, ... ] )  
   [USING INDEX TABLESPACE tablespace]  
   [ WITH ( FILLFACTOR=value ) ]  
   | PRIMARY KEY ( column_name [, ... ] )  
   [USING INDEX TABLESPACE tablespace]  
   [ WITH ( FILLFACTOR=value ) ]  
   | CHECK ( expression )

where partition_type is:
   LIST  
   | RANGE

where partition_specification is:
   partition_element [, ...]
and partition_element is:

```sql
| DEFAULT PARTITION name |
| [PARTITION name] VALUES (list_value [, ... ] ) |
| [PARTITION name] START ([datatype] 'start_value') [INCLUSIVE | EXCLUSIVE] |
| END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE] |
| EVERY ([datatype] [number | INTERVAL] 'interval_value') ] |
| [WITH (partition_storage_parameter=value [, ... ] ) ] |
| TABLESPACE tablespace ]
```

where subpartition_spec or template_spec is:

```sql
subpartition_element [, ... ]
```

and subpartition_element is:

```sql
| DEFAULT SUBPARTITION name |
| [SUBPARTITION name] VALUES (list_value [, ... ] ) |
| [SUBPARTITION name] START ([datatype] 'start_value') [INCLUSIVE | EXCLUSIVE] |
| END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE] |
| EVERY ([datatype] [number | INTERVAL] 'interval_value') ] |
| [WITH (partition_storage_parameter=value [, ... ] ) ] |
| TABLESPACE tablespace ]
```

**CREATE TABLE AS**

Defines a new table from the results of a query.

```sql
CREATE [ [GLOBAL | LOCAL] {TEMPORARY | TEMP} ] TABLE table_name |
| (column_name [, ... ] ) |
| [WITH (storage_parameter=value [, ... ] ) ] |
| [ON COMMIT {PRESERVE ROWS | DELETE ROWS | DROP}] |
| [TABLESPACE tablespace] |
| AS query |
| [DISTRIBUTED BY (column, [ ... ] ) | DISTRIBUTED RANDOMLY] |
```

where storage_parameter is:

- APPENDONLY={TRUE|FALSE}
- ORIENTATION={COLUMN|ROW}
- COMPRESSTYPE={ZLIB|QUICKLZ}
- COMPRESSLEVEL={0-9 | 1}
- FILLFACTOR={10-100}
- OIDS [=TRUE|FALSE]

**CREATE TABLESPACE**

Defines a new tablespace.

```sql
CREATE TABLESPACE tablespace_name [OWNER username] |
| FILESPACE filespace_name
```
**CREATE TRIGGER**
Defines a new trigger. User-defined triggers are not supported in Greenplum Database.

```
CREATE TRIGGER name {BEFORE | AFTER} {event [OR ...]} ON table [ FOR [EACH] [ROW | STATEMENT] ]
   EXECUTE PROCEDURE funcname ( arguments )
```

**CREATE TYPE**
Defines a new data type.

```
CREATE TYPE name AS ( attribute_name data_type [, ... ] )
CREATE TYPE name ( 
   INPUT = input_function,
   OUTPUT = output_function
 [, RECEIVE = receive_function]
 [, SEND = send_function]
 [, ANALYZE = analyze_function]
 [, INTERNALLENGTH = {internallength | VARIABLE}]
 [, PASSEDBYVALUE]
 [, ALIGNMENT = alignment]
 [, STORAGE = storage]
 [, DEFAULT = default]
 [, ELEMENT = element]
 [, DELIMITER = delimiter]
 )
CREATE TYPE name
```

**CREATE USER**
Defines a new database role with the `LOGIN` privilege by default.

```
CREATE USER name [ [WITH] option [ ... ] ]
where option can be:
   SUPERUSER | NOSUPERUSER
   CREATEDB | NOCREATEDB
   CREATEROLE | NOCREATEROLE
   CREATEUSER | NOCREATEUSER
   INHERIT | NOINHERIT
   LOGIN | NOLOGIN
   [ [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
   VALID UNTIL 'timestamp'
   IN ROLE rolename [, ...]
   IN GROUP rolename [, ...]
   ROLE rolename [, ...]
   ADMIN rolename [, ...]
   USER rolename [, ...]
   SYSID uid
   RESOURCE QUEUE queue_name
```

**CREATE VIEW**
Defines a new view.

```
CREATE [OR REPLACE] [TEMP | TEMPORARY] VIEW name 
   [ ( column_name [, ...] ) ]
   AS query
```
**DEALLOCATE**
Deallocates a prepared statement.

DEALLOCATE [PREPARE] name

**DECLARE**
Defines a cursor.

DECLARE name [BINARY] [INSENSITIVE] [NO SCROLL] CURSOR
   [{WITH | WITHOUT} HOLD]
   FOR query [FOR READ ONLY]

**DELETE**
Deletes rows from a table.

DELETE FROM [ONLY] table [[AS] alias]
   [USING usinglist]
   [WHERE condition]

**DROP AGGREGATE**
Removes an aggregate function.

DROP AGGREGATE [IF EXISTS] name ( type [, ...] ) [CASCADE | RESTRICT]

**DROP CAST**
Removes a cast.

DROP CAST [IF EXISTS] (sourcetype AS targettype) [CASCADE | RESTRICT]

**DROP CONVERSION**
Removes a conversion.

DROP CONVERSION [IF EXISTS] name [CASCADE | RESTRICT]

**DROP DATABASE**
Removes a database.

DROP DATABASE [IF EXISTS] name

**DROP DOMAIN**
Removes a domain.

DROP DOMAIN [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

**DROP EXTERNAL TABLE**
Removes an external table definition.

DROP EXTERNAL [WEB] TABLE [IF EXISTS] name [CASCADE | RESTRICT]

**DROP FILESPACE**
Removes a filesystem.

DROP FILESPACE [IF EXISTS] filespacename

**DROP FUNCTION**
Removes a function.

DROP FUNCTION [IF EXISTS] name ( [ [argmode] [argname] argtype [, ...] ] ) [CASCADE | RESTRICT]
DROP GROUP
Removes a database role.

DROP GROUP [IF EXISTS] name [, ...]

DROP INDEX
Removes an index.

DROP INDEX [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

DROP LANGUAGE
Removes a procedural language.

DROP [PROCEDURAL] LANGUAGE [IF EXISTS] name [CASCADE | RESTRICT]

DROP OPERATOR
Removes an operator.

DROP OPERATOR [IF EXISTS] name ( {lefttype | NONE} , {righttype | NONE} ) [CASCADE | RESTRICT]

DROP OPERATOR CLASS
Removes an operator class.

DROP OPERATOR CLASS [IF EXISTS] name USING index_method [CASCADE | RESTRICT]

DROP OWNED
Removes database objects owned by a database role.

DROP OWNED BY name [, ...] [CASCADE | RESTRICT]

DROP RESOURCE QUEUE
Removes a resource queue.

DROP RESOURCE QUEUE queue_name

DROP ROLE
Removes a database role.

DROP ROLE [IF EXISTS] name [, ...]

DROP RULE
Removes a rewrite rule.

DROP RULE [IF EXISTS] name ON relation [CASCADE | RESTRICT]

DROP SCHEMA
Removes a schema.

DROP SCHEMA [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

DROP SEQUENCE
Removes a sequence.

DROP SEQUENCE [IF EXISTS] name [, ...] [CASCADE | RESTRICT]
DROP TABLE
Removes a table.

DROP TABLE [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

DROP TABLESPACE
Removes a tablespace.

DROP TABLESPACE [IF EXISTS] tablespacename

DROP TRIGGER
Removes a trigger.

DROP TRIGGER [IF EXISTS] name ON table [CASCADE | RESTRICT]

DROP TYPE
Removes a data type.

DROP TYPE [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

DROP USER
Removes a database role.

DROP USER [IF EXISTS] name [, ...]

DROP VIEW
Removes a view.

DROP VIEW [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

END
Commits the current transaction.

END [WORK | TRANSACTION]

EXECUTE
Executes a prepared SQL statement.

EXECUTE name [ (parameter [, ...] ) ]

EXPLAIN
Shows the query plan of a statement.

EXPLAIN [ANALYZE] [VERBOSE] statement
**FETCH**
Retrieves rows from a query using a cursor.

```sql
FETCH [ forward_direction { FROM | IN } ] cursorname
```

where `forward_direction` can be empty or one of:
- NEXT
- FIRST
- LAST
- ABSOLUTE count
- RELATIVE count
- count
- ALL
- FORWARD
- FORWARD count
- FORWARD ALL

**GRANT**
Defines access privileges.

```sql
GRANT { {SELECT | INSERT | UPDATE | DELETE | REFERENCES | TRIGGER} [, ...] | ALL [PRIVILEGES] }
  ON [TABLE] tablename [, ...]
  TO { rolename | PUBLIC } [, ...] [WITH GRANT OPTION]
GRANT { {USAGE | SELECT | UPDATE} [, ...] | ALL [PRIVILEGES] }
  ON SEQUENCE sequencename [, ...]
  TO { rolename | PUBLIC } [, ...] [WITH GRANT OPTION]
GRANT { {CREATE | CONNECT | TEMPORARY | TEMP} [, ...] | ALL [PRIVILEGES] }
  ON DATABASE dbname [, ...]
  TO { rolename | PUBLIC } [, ...] [WITH GRANT OPTION]
GRANT { EXECUTE | ALL [PRIVILEGES] }
  ON FUNCTION funcname ( [ [argmode] argname argtype [, ...] ] ) [, ...]
  TO { rolename | PUBLIC } [, ...] [WITH GRANT OPTION]
GRANT { USAGE | ALL [PRIVILEGES] }
  ON LANGUAGE langname [, ...]
  TO { rolename | PUBLIC } [, ...] [WITH GRANT OPTION]
GRANT { {CREATE | USAGE} [, ...] | ALL [PRIVILEGES] }
  ON SCHEMA schemaname [, ...]
  TO { rolename | PUBLIC } [, ...] [WITH GRANT OPTION]
GRANT { CREATE | ALL [PRIVILEGES] }
  ON TABLESPACE tablespacename [, ...]
  TO { rolename | PUBLIC } [, ...] [WITH GRANT OPTION]
GRANT parent_role [, ...]
  TO member_role [, ...] [WITH ADMIN OPTION]
```

**INSERT**
Creates new rows in a table.

```sql
INSERT INTO table ( [ column [, ...] ] )
  {DEFAULT VALUES | VALUES ( {expression | DEFAULT} [, ...] ) [, ...] | query}
```

**LOAD**
Loads or reloads a shared library file.

```sql
LOAD 'filename'
```
**LOCK**
Locks a table.

LOCK [TABLE] name [, ...] [IN lockmode MODE] [NOWAIT]

where *lockmode* is one of:
ACCESS SHARE | ROW SHARE | ROW EXCLUSIVE | SHARE UPDATE EXCLUSIVE | SHARE | SHARE ROW EXCLUSIVE | EXCLUSIVE | ACCESS EXCLUSIVE

**MOVE**
Positions a cursor.

MOVE [ forward_direction {FROM | IN} ] cursorname

where *direction can be empty or one of*:
NEXT
FIRST
LAST
ABSOLUTE count
RELATIVE count
count
ALL
FORWARD
FORWARD count
FORWARD ALL

**PREPARE**
Prepare a statement for execution.

PREPARE name [ (datatype [, ...] ) ] AS statement

**REASSIGN OWNED**
Changes the ownership of database objects owned by a database role.

REASSIGN OWNED BY old_role [, ...] TO new_role

**REINDEX**
Rebuilds indexes.

REINDEX [INDEX | TABLE | DATABASE | SYSTEM] name

**RELEASE SAVEPOINT**
Destroys a previously defined savepoint.

RELEASE [SAVEPOINT] savepoint_name

**RESET**
Restores the value of a system configuration parameter to the default value.

RESET configuration_parameter
RESET ALL
REVOKE
Removes access privileges.

REVOKE [GRANT OPTION FOR] { [SELECT | INSERT | UPDATE | DELETE
| REFERENCES | TRIGGER] [, ... ] | ALL [PRIVILEGES] }
ON [TABLE] tablename [, ...]
FROM { rolename | PUBLIC } [, ...]
[CASCADE | RESTRICT]

REVOKE [GRANT OPTION FOR] { [USAGE | SELECT | UPDATE] [, ...]
| ALL [PRIVILEGES] }
ON SEQUENCE sequencename [, ...]
FROM { rolename | PUBLIC } [, ...]
[CASCADE | RESTRICT]

REVOKE [GRANT OPTION FOR] { [CREATE | CONNECT
| TEMPORARY | TEMP] [, ... ] | ALL [PRIVILEGES] }
ON DATABASE dbname [, ...]
FROM { rolename | PUBLIC } [, ...]
[CASCADE | RESTRICT]

REVOKE [GRANT OPTION FOR] { EXECUTE | ALL [PRIVILEGES]}
ON FUNCTION funcname ( [ [argmode] argname ] argtype
[, ... ] ) [, ...]
FROM { rolename | PUBLIC } [, ...]
[CASCADE | RESTRICT]

REVOKE [GRANT OPTION FOR] { [CREATE | USAGE] [, ...]
| ALL [PRIVILEGES] }
ON SCHEMA schemaname [, ...]
FROM { rolename | PUBLIC } [, ...]
[CASCADE | RESTRICT]

REVOKE [GRANT OPTION FOR] { CREATE | ALL [PRIVILEGES] }
ON TABLESPACE tablespacename [, ...]
FROM { rolename | PUBLIC } [, ...]
[CASCADE | RESTRICT]

REVOKE [ADMIN OPTION FOR] parent_role [, ...]
FROM member_role [, ...]
[CASCADE | RESTRICT]

ROLLBACK
Aborts the current transaction.

ROLLBACK [WORK | TRANSACTION]

ROLLBACK TO SAVEPOINT
Rolls back the current transaction to a savepoint.

ROLLBACK [WORK | TRANSACTION] TO [SAVEPOINT] savepoint_name

SAVEPOINT
Defines a new savepoint within the current transaction.

SAVEPOINT savepoint_name
SELECT

Retrieves rows from a table or view.

SELECT [ALL | DISTINCT [ON (expression [, ...])]]
    * | expression [[AS] output_name] [, ...]
    [FROM from_item [, ...]]
    [WHERE condition]
    [GROUP BY grouping_element [, ...]]
    [HAVING condition [, ...]]
    [WINDOW window_name AS (window_specification)]
    [[UNION | INTERSECT | EXCEPT] [ALL] select]
    [ORDER BY expression [ASC | DESC | USING operator] [, ...]]
    [LIMIT {count | ALL}]
    [OFFSET start]
    [FOR {UPDATE | SHARE} [OF table_name [, ...]] [NOWAIT] [...]]

where grouping_element can be one of:

()  
expression  
ROLLUP (expression [, ...])  
CUBE (expression [, ...])  
GROUPING SETS ((grouping_element [, ...]))

where window_specification can be:

[window_name]
    [PARTITION BY expression [, ...]]
    [ORDER BY expression [ASC | DESC | USING operator] [, ...]]
    [[RANGE | ROWS]  
        { UNBOUNDED PRECEDING
        | expression PRECEDING
        | CURRENT ROW
        | BETWEEN window_frame_bound AND window_frame_bound }]

where window_frame_bound can be one of:

UNBOUNDED PRECEDING  
expression PRECEDING  
CURRENT ROW  
expression FOLLOWING  
UNBOUNDED FOLLOWING

where from_item can be one of:

[ONLY] table_name [[AS] alias [(column_alias [, ...])]  
(select) [AS] alias [(column_alias [, ...])]  
function_name ( [argument [, ...]] ) [AS] alias  
    [(column_alias [, ...])  
        | column_definition [, ...])  
function_name ( [argument [, ...]] ) AS  
    (column_definition [, ...])  
from_item [NATURAL] join_type from_item
    [ON join_condition | USING (join_column [, ...])]

SELECT

Retrieves rows from a table or view.
SELECT INTO
Defines a new table from the results of a query.

SELECT [ALL | DISTINCT [ON ( expression [, ...] )]]
* | expression [AS output_name] [, ...]
INTO [TEMPORARY | TEMP] [TABLE] new_table
[FROM from_item [, ...]]
[WHERE condition]
[GROUP BY expression [, ...]]
[HAVING condition [, ...]]
[UNION | INTERSECT | EXCEPT] [ALL] select
[ORDER BY expression [ASC | DESC | USING operator] [, ...]]
[LIMIT {count | ALL}]
[OFFSET start]
[FOR {UPDATE | SHARE} [OF table_name [, ...]] [NOWAIT] [...]]

SET
Changes the value of a Greenplum Database configuration parameter.

SET [SESSION | LOCAL] configuration_parameter {TO | =} value | 'value' | DEFAULT
SET [SESSION | LOCAL] TIME_ZONE {timezone | LOCAL | DEFAULT}

SET ROLE
Sets the current role identifier of the current session.

SET [SESSION | LOCAL] ROLE rolename
SET [SESSION | LOCAL] ROLE NONE
RESET ROLE

SET SESSION AUTHORIZATION
Sets the session role identifier and the current role identifier of the current session.

SET [SESSION | LOCAL] SESSION AUTHORIZATION rolename
SET [SESSION | LOCAL] SESSION AUTHORIZATION DEFAULT
RESET SESSION AUTHORIZATION

SET TRANSACTION
Sets the characteristics of the current transaction.

SET TRANSACTION transaction_mode [, ...]
SET SESSION CHARACTERISTICS AS TRANSACTION transaction_mode [, ...]

where transaction_mode is one of:
ISOLATION LEVEL {SERIALIZABLE | REPEATABLE READ | READ COMMITTED | READ UNCOMMITTED}
READ WRITE | READ ONLY

SHOW
Shows the value of a system configuration parameter.

SHOW configuration_parameter
SHOW ALL

START TRANSACTION
Starts a transaction block.

START TRANSACTION [SERIALIZABLE | REPEATABLE READ | READ COMMITTED | READ UNCOMMITTED]
[READ WRITE | READ ONLY]
TRUNCATE
Empties a table of all rows.
TRUNCATE [TABLE] name [, ...] [CASCADE | RESTRICT]

UPDATE
Updates rows of a table.
UPDATE [ONLY] table [[AS] alias]
   SET {column = {expression | DEFAULT} |
    (column [, ...]) = {(expression | DEFAULT) [, ...]}) [, ...]
    [FROM fromlist]
    [WHERE condition]

VACUUM
Garbage-collects and optionally analyzes a database.
VACUUM [FULL] [FREEZE] [VERBOSE] [table]
VACUUM [FULL] [FREEZE] [VERBOSE] ANALYZE
   [table [(column [, ...])]]

VALUES
Computes a set of rows.
VALUES ( expression [, ...] ) [, ...]
   [ORDER BY sort_expression [ASC | DESC | USING operator] [, ...]]
   [LIMIT {count | ALL}] [OFFSET start]
ABORT

Aborts the current transaction.

Synopsis

ABORT [WORK | TRANSACTION]

Description

ABORT rolls back the current transaction and causes all the updates made by the transaction to be discarded. This command is identical in behavior to the standard SQL command ROLLBACK, and is present only for historical reasons.

Parameters

WORK
TRANSACTION

Optional key words. They have no effect.

Notes

Use COMMIT to successfully terminate a transaction. Issuing ABORT when not inside a transaction does no harm, but it will provoke a warning message.

Compatibility

This command is a Greenplum Database extension present for historical reasons. ROLLBACK is the equivalent standard SQL command.

See Also

BEGIN, COMMIT, ROLLBACK
ALTER AGGREGATE

Changes the definition of an aggregate function

Synopsis

ALTER AGGREGATE name ( type [ , ... ] ) RENAME TO new_name
ALTER AGGREGATE name ( type [ , ... ] ) OWNER TO new_owner
ALTER AGGREGATE name ( type [ , ... ] ) SET SCHEMA new_schema

Description

ALTER AGGREGATE changes the definition of an aggregate function.

You must own the aggregate function to use ALTER AGGREGATE. To change the schema of an aggregate function, you must also have CREATE privilege on the new schema. To alter the owner, you must also be a direct or indirect member of the new owning role, and that role must have CREATE privilege on the aggregate function’s schema. (These restrictions enforce that altering the owner does not do anything you could not do by dropping and recreating the aggregate function. However, a superuser can alter ownership of any aggregate function anyway.)

Parameters

name
The name (optionally schema-qualified) of an existing aggregate function.

type
An input data type on which the aggregate function operates. To reference a zero-argument aggregate function, write * in place of the list of input data types.

new_name
The new name of the aggregate function.

new_owner
The new owner of the aggregate function.

new_schema
The new schema for the aggregate function.

Examples

To rename the aggregate function myavg for type integer to my_average:

    ALTER AGGREGATE myavg(integer) RENAME TO my_average;

To change the owner of the aggregate function myavg for type integer to joe:

    ALTER AGGREGATE myavg(integer) OWNER TO joe;
To move the aggregate function *myavg* for type *integer* into schema *myschema*:

```
ALTER AGGREGATE myavg(integer) SET SCHEMA myschema;
```

**Compatibility**

There is no `ALTER AGGREGATE` statement in the SQL standard.

**See Also**

`CREATE AGGREGATE`, `DROP AGGREGATE`
ALTER CONVERSION

Changes the definition of a conversion.

Synopsis

ALTER CONVERSION name RENAME TO newname
ALTER CONVERSION name OWNER TO newowner

Description

ALTER CONVERSION changes the definition of a conversion.

You must own the conversion to use ALTER CONVERSION. To alter the owner, you must also be a direct or indirect member of the new owning role, and that role must have CREATE privilege on the conversion’s schema. (These restrictions enforce that altering the owner does not do anything you could not do by dropping and recreating the conversion. However, a superuser can alter ownership of any conversion anyway.)

Parameters

name
The name (optionally schema-qualified) of an existing conversion.

newname
The new name of the conversion.

newowner
The new owner of the conversion.

Examples

To rename the conversion iso_8859_1_to_utf8 to latin1_to_unicode:

    ALTER CONVERSION iso_8859_1_to_utf8 RENAME TO latin1_to_unicode;

To change the owner of the conversion iso_8859_1_to_utf8 to joe:

    ALTER CONVERSION iso_8859_1_to_utf8 OWNER TO joe;

Compatibility

There is no ALTER CONVERSION statement in the SQL standard.

See Also

CREATE CONVERSION, DROP CONVERSION
ALTER DATABASE

Changes the attributes of a database.

Synopsis

ALTER DATABASE name [ WITH CONNECTION LIMIT connlimit ]
ALTER DATABASE name SET parameter { TO | = } { value | DEFAULT }
ALTER DATABASE name RESET parameter
ALTER DATABASE name RENAME TO newname
ALTER DATABASE name OWNER TO new_owner

Description

ALTER DATABASE changes the attributes of a database.

The first form changes the allowed connection limit for a database. Only the database owner or a superuser can change this setting.

The second and third forms change the session default for a configuration parameter for a Greenplum database. Whenever a new session is subsequently started in that database, the specified value becomes the session default value. The database-specific default overrides whatever setting is present in the server configuration file (postgresql.conf). Only the database owner or a superuser can change the session defaults for a database. Certain parameters cannot be set this way, or can only be set by a superuser.

The fourth form changes the name of the database. Only the database owner or a superuser can rename a database; non-superuser owners must also have the CREATEDB privilege. You cannot rename the current database. Connect to a different database first.

The fifth form changes the owner of the database. To alter the owner, you must own the database and also be a direct or indirect member of the new owning role, and you must have the CREATEDB privilege. (Note that superusers have all these privileges automatically.)

Parameters

name

The name of the database whose attributes are to be altered.

connlimit

Warning: Setting a connection limit at the database level cannot be enforced in Greenplum Database and may cause queries to fail. Leave this set at the default of -1 (no limit). Connection limits may only be set at the system level in Greenplum Database. See “Limiting Concurrent Connections” on page 38 for more information.
**parameter value**

Set this database’s session default for the specified configuration parameter to the given value. If value is **DEFAULT** or, equivalently, **RESET** is used, the database-specific setting is removed, so the system-wide default setting will be inherited in new sessions. Use **RESET ALL** to clear all database-specific settings. See “Server Configuration Parameters” on page 770 for information about all user-settable configuration parameters.

**newname**

The new name of the database.

**new_owner**

The new owner of the database.

**Notes**

It is also possible to set a configuration parameter session default for a specific role (user) rather than to a database. Role-specific settings override database-specific ones if there is a conflict. See **ALTER ROLE**.

**Examples**

To set the default schema search path for the *mydatabase* database:

```
ALTER DATABASE mydatabase SET search_path TO myschema, public, pg_catalog;
```

**Compatibility**

The **ALTER DATABASE** statement is a Greenplum Database extension.

**See Also**

*CREATE DATABASE*, *DROP DATABASE*, *SET*
ALTER DOMAIN

Changes the definition of a domain.

Synopsis

ALTER DOMAIN name { SET DEFAULT expression | DROP DEFAULT }
ALTER DOMAIN name { SET | DROP } NOT NULL
ALTER DOMAIN name ADD domain_constraint
ALTER DOMAIN name DROP CONSTRAINT constraint_name [RESTRICT | CASCADE]
ALTER DOMAIN name OWNER TO new_owner
ALTER DOMAIN name SET SCHEMA new_schema

Description

ALTER DOMAIN changes the definition of an existing domain. There are several sub-forms:

- **SET/DROP DEFAULT** — These forms set or remove the default value for a domain. Note that defaults only apply to subsequent INSERT commands. They do not affect rows already in a table using the domain.

- **SET/DROP NOT NULL** — These forms change whether a domain is marked to allow NULL values or to reject NULL values. You may only SET NOT NULL when the columns using the domain contain no null values.

- **ADD domain_constraint** — This form adds a new constraint to a domain using the same syntax as CREATE DOMAIN. This will only succeed if all columns using the domain satisfy the new constraint.

- **DROP CONSTRAINT** — This form drops constraints on a domain.

- **OWNER** — This form changes the owner of the domain to the specified user.

- **SET SCHEMA** — This form changes the schema of the domain. Any constraints associated with the domain are moved into the new schema as well.

You must own the domain to use ALTER DOMAIN. To change the schema of a domain, you must also have CREATE privilege on the new schema. To alter the owner, you must also be a direct or indirect member of the new owning role, and that role must have CREATE privilege on the domain’s schema. (These restrictions enforce that altering the owner does not do anything you could not do by dropping and recreating the domain. However, a superuser can alter ownership of any domain anyway.)

Parameters

*name*

The name (optionally schema-qualified) of an existing domain to alter.
domain_constraint
New domain constraint for the domain.

constraint_name
Name of an existing constraint to drop.

CASCADE
Automatically drop objects that depend on the constraint.

RESTRICT
Refuse to drop the constraint if there are any dependent objects. This is the default behavior.

new_owner
The user name of the new owner of the domain.

new_schema
The new schema for the domain.

Examples
To add a NOT NULL constraint to a domain:
ALTER DOMAIN zipcode SET NOT NULL;

To remove a NOT NULL constraint from a domain:
ALTER DOMAIN zipcode DROP NOT NULL;

To add a check constraint to a domain:
ALTER DOMAIN zipcode ADD CONSTRAINT zipchk CHECK (char_length(VALUE) = 5);

To remove a check constraint from a domain:
ALTER DOMAIN zipcode DROP CONSTRAINT zipchk;

To move the domain into a different schema:
ALTER DOMAIN zipcode SET SCHEMA customers;

Compatibility
ALTER DOMAIN conforms to the SQL standard, except for the OWNER and SET SCHEMA variants, which are Greenplum Database extensions.

See Also
CREATE DOMAIN, DROP DOMAIN
**ALTER EXTERNAL TABLE**

Changes the definition of an external table.

**Synopsis**

ALTER EXTERNAL TABLE name RENAME [COLUMN] column TO new_column

ALTER EXTERNAL TABLE name RENAME TO new_name

ALTER EXTERNAL TABLE name SET SCHEMA new_schema

ALTER EXTERNAL TABLE name action [, ... ]

where *action* is one of:

- ADD [COLUMN] column_name type
- DROP [COLUMN] column
- ALTER [COLUMN] column TYPE type [USING expression]
- OWNER TO new_owner

**Description**

ALTER EXTERNAL TABLE changes the definition of an existing external table. There are several subforms:

- **ADD COLUMN** — Adds a new column to the external table definition.
- **DROP COLUMN** — Drops a column from the external table definition. Note that if you drop readable external table columns, it only changes the table definition in Greenplum Database. External data files are not changed.
- **ALTER COLUMN TYPE** — Changes the data type of a column of a table. The optional `USING` clause specifies how to compute the new column value from the old. If omitted, the default conversion is the same as an assignment cast from old data type to new. A `USING` clause must be provided if there is no implicit or assignment cast from the old to new type.
- **OWNER** — Changes the owner of the external table to the specified user.
- **RENAME** — Changes the name of an external table or the name of an individual column in the table. There is no effect on the external data.
- **SET SCHEMA** — Moves the external table into another schema.

You must own the external table to use **ALTER EXTERNAL TABLE**. To change the schema of an external table, you must also have CREATE privilege on the new schema. To alter the owner, you must also be a direct or indirect member of the new owning role, and that role must have CREATE privilege on the external table’s schema. A superuser has these privileges automatically.

In this release, **ALTER EXTERNAL TABLE** cannot modify the external table type, the data format, or the location of the external data. To modify this information, you must drop and recreate the external table definition.
Parameters

name
The name (possibly schema-qualified) of an existing external table definition to alter.

column
Name of a new or existing column.

new_column
New name for an existing column.

new_name
New name for the external table.

type
Data type of the new column, or new data type for an existing column.

new_owner
The role name of the new owner of the external table.

new_schema
The name of the schema to which the external table will be moved.

Examples

Add a new column to an external table definition:

```
ALTER EXTERNAL TABLE ext_expenses ADD COLUMN manager text;
```

Change the name of an external table:

```
ALTER EXTERNAL TABLE ext_data RENAME TO ext_sales_data;
```

Change the owner of an external table:

```
ALTER EXTERNAL TABLE ext_data OWNER TO jojo;
```

Change the schema of an external table:

```
ALTER EXTERNAL TABLE ext_leads SET SCHEMA marketing;
```

Compatibility

ALTER EXTERNAL TABLE is a Greenplum Database extension. There is no ALTER EXTERNAL TABLE statement in the SQL standard or regular PostgreSQL.

See Also

CREATE EXTERNAL TABLE, DROP EXTERNAL TABLE
**ALTER FILESPACE**

Changes the definition of a filespace.

**Synopsis**

ALTER FILESPACE name RENAME TO newname  
ALTER FILESPACE name OWNER TO newowner

**Description**

ALTER FILESPACE changes the definition of a filespace.  

You must own the filespace to use ALTER FILESPACE. To alter the owner, you must also be a direct or indirect member of the new owning role (note that superusers have these privileges automatically).

**Parameters**

*name*

The name of an existing filespace.

*newname*

The new name of the filespace. The new name cannot begin with `pg_` or `gp_` (reserved for system filespaces).

*newowner*

The new owner of the filespace.

**Examples**

Rename filesystem *myfs* to *fast_ssd*:  

```
ALTER FILESPACE myfs RENAME TO fast_ssd;
```

Change the owner of tablespace *mysf*:

```
ALTER FILESPACE myfs OWNER TO dba;
```

**Compatibility**

There is no ALTER FILESPACE statement in the SQL standard or in PostgreSQL.

**See Also**

gpfilespace, DROP FILESPACE
**ALTER FUNCTION**

Changes the definition of a function.

**Synopsis**

```
ALTER FUNCTION name ( [ [argmode] argname argtype [, ...] ] )
action [, ... ] [RESTRICT]

ALTER FUNCTION name ( [ [argmode] argname argtype [, ...] ] )
RENAME TO new_name

ALTER FUNCTION name ( [ [argmode] argname argtype [, ...] ] )
OWNER TO new_owner

ALTER FUNCTION name ( [ [argmode] argname argtype [, ...] ] )
SET SCHEMA new_schema
```

where `action` is one of:

```
(CALLED ON NULL INPUT | RETURNS NULL ON NULL INPUT | STRICT)
(IMMUTABLE | STABLE | VOLATILE)
([EXTERNAL] SECURITY INVOKER | [EXTERNAL] SECURITY DEFINER)
```

**Description**

`ALTER FUNCTION` changes the definition of a function.

You must own the function to use `ALTER FUNCTION`. To change a function’s schema, you must also have `CREATE` privilege on the new schema. To alter the owner, you must also be a direct or indirect member of the new owning role, and that role must have `CREATE` privilege on the function’s schema. (These restrictions enforce that altering the owner does not do anything you could not do by dropping and recreating the function. However, a superuser can alter ownership of any function anyway.)

**Parameters**

- `name`
  The name (optionally schema-qualified) of an existing function.

- `argmode`
  The mode of an argument: either `IN`, `OUT`, or `INOUT`. If omitted, the default is `IN`. Note that `ALTER FUNCTION` does not actually pay any attention to `OUT` arguments, since only the input arguments are needed to determine the function’s identity. So it is sufficient to list the `IN` and `INOUT` arguments.

- `argname`
  The name of an argument. Note that `ALTER FUNCTION` does not actually pay any attention to argument names, since only the argument data types are needed to determine the function’s identity.
**argtype**

The data type(s) of the function’s arguments (optionally schema-qualified), if any.

**new_name**

The new name of the function.

**new_owner**

The new owner of the function. Note that if the function is marked `SECURITY DEFINER`, it will subsequently execute as the new owner.

**new_schema**

The new schema for the function.

**CALLED ON NULL INPUT**

**RETURNS NULL ON NULL INPUT**

**STRICT**

`CALLED ON NULL INPUT` changes the function so that it will be invoked when some or all of its arguments are null. `RETURNS NULL ON NULL INPUT` or `STRICT` changes the function so that it is not invoked if any of its arguments are null; instead, a null result is assumed automatically. See `CREATE FUNCTION` for more information.

**IMMUTABLE**

**STABLE**

**VOLATILE**

Change the volatility of the function to the specified setting. See `CREATE FUNCTION` for details.

```
[ EXTERNAL ] SECURITY INVOKER
[ EXTERNAL ] SECURITY DEFINER
```

Change whether the function is a security definer or not. The key word `EXTERNAL` is ignored for SQL conformance. See `CREATE FUNCTION` for more information about this capability.

**RESTRICT**

Ignored for conformance with the SQL standard.

---

**Notes**

Greenplum Database has limitations on the use of functions defined as `STABLE` or `VOLATILE`. See `CREATE FUNCTION` for more information.

---

**Examples**

To rename the function `sqrt` for type `integer` to `square_root`:

```
ALTER FUNCTION sqrt(integer) RENAME TO square_root;
```

To change the owner of the function `sqrt` for type `integer` to `joe`:

```
ALTER FUNCTION sqrt(integer) OWNER TO joe;
```
To change the schema of the function sqrt for type integer to math:

    ALTER FUNCTION sqrt(integer) SET SCHEMA math;

**Compatibility**

This statement is partially compatible with the `ALTER FUNCTION` statement in the SQL standard. The standard allows more properties of a function to be modified, but does not provide the ability to rename a function, make a function a security definer, or change the owner, schema, or volatility of a function. The standard also requires the `RESTRICT` key word, which is optional in Greenplum Database.

**See Also**

`CREATE FUNCTION`, `DROP FUNCTION`
ALTER GROUP

Changes a role name or membership.

Synopsis

ALTER GROUP groupname ADD USER username [, ... ]
ALTER GROUP groupname DROP USER username [, ... ]
ALTER GROUP groupname RENAME TO newname

Description

ALTER GROUP is an obsolete command, though still accepted for backwards compatibility. Groups (and users) have been superseded by the more general concept of roles. See ALTER ROLE for more information.

Parameters

groupname

The name of the group (role) to modify.

username

Users (roles) that are to be added to or removed from the group. The users (roles) must already exist.

newname

The new name of the group (role).

Examples

To add users to a group:

    ALTER GROUP staff ADD USER karl, john;

To remove a user from a group:

    ALTER GROUP workers DROP USER beth;

Compatibility

There is no ALTER GROUP statement in the SQL standard.

See Also

ALTER ROLE, GRANT, REVOKE
ALTER INDEX

Changes the definition of an index.

Synopsis

```
ALTER INDEX name RENAME TO new_name
ALTER INDEX name SET TABLESPACE tablespace_name
ALTER INDEX name SET ( FILLFACTOR = value )
ALTER INDEX name RESET ( FILLFACTOR )
```

Description

ALTER INDEX changes the definition of an existing index. There are several subforms:

- **RENAME** — Changes the name of the index. There is no effect on the stored data.
- **SET TABLESPACE** — Changes the index’s tablespace to the specified tablespace and moves the data file(s) associated with the index to the new tablespace. See also CREATE TABLESPACE.
- **SET FILLFACTOR** — Changes the index-method-specific storage parameters for the index. The built-in index methods all accept a single parameter: FILLFACTOR. The fillfactor for an index is a percentage that determines how full the index method will try to pack index pages. Index contents will not be modified immediately by this command. Use REINDEX to rebuild the index to get the desired effects.
- **RESET FILLFACTOR** — Resets FILLFACTOR to the default. As with SET, a REINDEX may be needed to update the index entirely.

Parameters

**name**

The name (optionally schema-qualified) of an existing index to alter.

**new_name**

New name for the index.

**tablespace_name**

The tablespace to which the index will be moved.

**FILLFACTOR**

The fillfactor for an index is a percentage that determines how full the index method will try to pack index pages. For B-trees, leaf pages are filled to this percentage during initial index build, and also when extending the index at the right (largest key values). If pages subsequently become completely full, they will be split, leading to gradual degradation in the index’s efficiency.
B-trees use a default fillfactor of 90, but any value from 10 to 100 can be selected. If the table is static then fillfactor 100 is best to minimize the index’s physical size, but for heavily updated tables a smaller fillfactor is better to minimize the need for page splits. The other index methods use fillfactor in different but roughly analogous ways; the default fillfactor varies between methods.

Notes
These operations are also possible using ALTER TABLE.
Changing any part of a system catalog index is not permitted.

Examples
To rename an existing index:

```
ALTER INDEX distributors RENAME TO suppliers;
```

To move an index to a different tablespace:

```
ALTER INDEX distributors SET TABLESPACE fasttablespace;
```

To change an index’s fill factor (assuming that the index method supports it):

```
ALTER INDEX distributors SET (fillfactor = 75);
REINDEX INDEX distributors;
```

Compatibility
ALTER INDEX is a Greenplum Database extension.

See Also
CREATE INDEX, REINDEX, ALTER TABLE
ALTER LANGUAGE

Changes the name of a procedural language.

Synopsis
ALTER LANGUAGE name RENAME TO newname

Description
ALTER LANGUAGE changes the name of a procedural language. Only a superuser can rename languages.

Parameters

name
   Name of a language.

newname
   The new name of the language.

Compatibility
There is no ALTER LANGUAGE statement in the SQL standard.

See Also
CREATE LANGUAGE, DROP LANGUAGE
**ALTER OPERATOR**

Changes the definition of an operator.

**Synopsis**

```sql
ALTER OPERATOR name ( {lefttype | NONE} , {righttype | NONE} )
OWNER TO newowner
```

**Description**

`ALTER OPERATOR` changes the definition of an operator. The only currently available functionality is to change the owner of the operator.

You must own the operator to use `ALTER OPERATOR`. To alter the owner, you must also be a direct or indirect member of the new owning role, and that role must have `CREATE` privilege on the operator’s schema. (These restrictions enforce that altering the owner does not do anything you could not do by dropping and recreating the operator. However, a superuser can alter ownership of any operator anyway.)

**Parameters**

*name*

The name (optionally schema-qualified) of an existing operator.

*lefttype*

The data type of the operator’s left operand; write `NONE` if the operator has no left operand.

*righttype*

The data type of the operator’s right operand; write `NONE` if the operator has no right operand.

*newowner*

The new owner of the operator.

**Examples**

Change the owner of a custom operator `a @@ b` for type `text`:

```sql
ALTER OPERATOR @@ (text, text) OWNER TO joe;
```

**Compatibility**

There is no `ALTER OPERATOR` statement in the SQL standard.

**See Also**

`CREATE OPERATOR`, `DROP OPERATOR`
ALTER OPERATOR CLASS

Changes the definition of an operator class.

Synopsis

ALTER OPERATOR CLASS name USING index_method RENAME TO newname
ALTER OPERATOR CLASS name USING index_method OWNER TO newowner

Description

ALTER OPERATOR CLASS changes the definition of an operator class.

You must own the operator class to use ALTER OPERATOR CLASS. To alter the owner, you must also be a direct or indirect member of the new owning role, and that role must have CREATE privilege on the operator class’s schema. (These restrictions enforce that altering the owner does not do anything you could not do by dropping and recreating the operator class. However, a superuser can alter ownership of any operator class anyway.)

Parameters

name

The name (optionally schema-qualified) of an existing operator class.

index_method

The name of the index method this operator class is for.

newname

The new name of the operator class.

newowner

The new owner of the operator class

Compatibility

There is no ALTER OPERATOR CLASS statement in the SQL standard.

See Also

CREATE OPERATOR CLASS, DROP OPERATOR CLASS
ALTER RESOURCE QUEUE

Changes the limits of a resource queue.

Synopsis

ALTER RESOURCE QUEUE name WITH ( queue_attribute=value [, ... ] )

where queue_attribute is:

ACTIVE_STATEMENTS=integer
MAX_COST=FLOAT
COST_OVERCOMMIT={TRUE|FALSE}
MIN_COST=FLOAT
PRIORITY={MIN|LOW|MEDIUM|HIGH|MAX}

ALTER RESOURCE QUEUE name WITHOUT ( queue_attribute=value [, ... ] )

where queue_attribute is:

ACTIVE_STATEMENTS=integer
MAX_COST=FLOAT
COST_OVERCOMMIT={TRUE|FALSE}
MIN_COST=FLOAT

Description

ALTER RESOURCE QUEUE changes the limits of a resource queue. Only a superuser can alter a resource queue. A resource queue must have either an ACTIVE_STATEMENTS or a MAX_COST value (or it can have both). You can also set or reset priority for a resource queue to control the relative share of available CPU resources used by queries associated with the queue.

Parameters

name
The name of the resource queue whose limits are to be altered.

ACTIVE_STATEMENTS integer
The number of active statements submitted from users in this resource queue allowed on the system at any one time. The value for ACTIVE_STATEMENTS should be an integer greater than 0. To reset ACTIVE_STATEMENTS to have no limit, enter a value of -1.

MAX_COST float
The total query planner cost of statements submitted from users in this resource queue allowed on the system at any one time. The value for MAX_COST is specified as a floating point number (for example 100.0) or can also be specified as an exponent (for example 1e+2). To reset MAX_COST to have no limit, enter a value of -1.0.
COST_OVERCOMMIT boolean

If a resource queue is limited based on query cost, then the administrator can allow cost overcommit (COST_OVERCOMMIT=TRUE, the default). This means that a query that exceeds the allowed cost threshold will be allowed to run but only when the system is idle. If COST_OVERCOMMIT=FALSE is specified, queries that exceed the cost limit will always be rejected and never allowed to run.

MIN_COST float

Queries with a cost under this limit will not be queued and run immediately. Cost is measured in units of disk page fetches; 1.0 equals one sequential disk page read. The value for MIN_COST is specified as a floating point number (for example 100.0) or can also be specified as an exponent (for example 1e+2). To reset MIN_COST to have no limit, enter a value of -1.0.

PRIORITY={MIN|LOW|MEDIUM|HIGH|MAX}

Sets the priority of queries associated with a resource queue. Queries or statements in queues with higher priority levels will receive a larger share of available CPU resources in case of contention. Queries in low-priority queues may be delayed while higher priority queries are executed.

Notes

Use CREATE ROLE or ALTER ROLE to add a role (user) to a resource queue.

Examples

Change the active query limit for a resource queue:

```
ALTER RESOURCE QUEUE myqueue WITH (ACTIVE_STATEMENTS=20);
```

Reset the maximum and minimum query cost limit for a resource queue to no limit:

```
ALTER RESOURCE QUEUE myqueue WITH (MAX_COST=-1.0, MIN_COST=-1.0);
```

Reset the query cost limit for a resource queue to 310 (or 30000000000.0) and do not allow overcommit:

```
ALTER RESOURCE QUEUE myqueue WITH (MAX_COST=3e+10,
COST_OVERCOMMIT=FALSE);
```

Reset the priority of queries associated with a resource queue to the minimum level:

```
ALTER RESOURCE QUEUE myqueue WITH (PRIORITY=MIN);
```

Compatibility

The ALTER RESOURCE QUEUE statement is a Greenplum Database extension. This command does not exist in standard PostgreSQL.

See Also

CREATE RESOURCE QUEUE, DROP RESOURCE QUEUE, CREATE ROLE, ALTER ROLE
**ALTER ROLE**

Changes a database role (user or group).

**Synopsis**

ALTER ROLE name RENAME TO newname

ALTER ROLE name SET config_parameter {TO | =} {value | DEFAULT}

ALTER ROLE name RESET config_parameter

ALTER ROLE name RESOURCE QUEUE {queue_name | NONE}

ALTER ROLE name [ [WITH] option [ ... ] ]

where option can be:

SUPERUSER | NOSUPERUSER
CREATEDB | NOCREATEDB
CREATEROLE | NOCREATEROLE
CREATEEXTTABLE | NOCREATEEXTTABLE

[ ( attribute='value'[, ...] ) ]

where attributes and values are:

type='readable'|'writable'
protocol='gpfdist'|'http'
| INHERIT | NOINHERIT
| LOGIN | NOLOGIN
| CONNECTION LIMIT conlimit
| [ENCRYPTED | UNENCRYPTED] PASSWORD 'password'
| VALID UNTIL 'timestamp'

**Description**

**ALTER ROLE** changes the attributes of a Greenplum Database role. There are several variants of this command:

- **RENAME** — Changes the name of the role. Database superusers can rename any role. Roles having CREATEROLE privilege can rename non-superuser roles. The current session user cannot be renamed (connect as a different user to rename a role). Because MD5-encrypted passwords use the role name as cryptographic salt, renaming a role clears its password if the password is MD5-encrypted.

- **SET | RESET** — changes a role’s session default for a specified configuration parameter. Whenever the role subsequently starts a new session, the specified value becomes the session default, overriding whatever setting is present in server configuration file (`postgresql.conf`). For a role without LOGIN privilege, session defaults have no effect. Ordinary roles can change their own session defaults. Superusers can change anyone’s session defaults. Roles having CREATEROLE privilege can change defaults for non-superuser roles. See “Server Configuration Parameters” on page 770 for more information on all user-settable configuration parameters.
• **RESOURCE QUEUE** — Assigns the role to a workload management resource queue. The role would then be subject to the limits assigned to the resource queue when issuing queries. Specify `NONE` to assign the role to the default resource queue. A role can only belong to one resource queue. For a role without `LOGIN` privilege, resource queues have no effect. See `CREATE RESOURCE QUEUE` for more information.

• **WITH option** — Changes many of the role attributes that can be specified in `CREATE ROLE`. Attributes not mentioned in the command retain their previous settings. Database superusers can change any of these settings for any role. Roles having `CREATEROLE` privilege can change any of these settings, but only for non-superuser roles. Ordinary roles can only change their own password.

### Parameters

- **name**
  The name of the role whose attributes are to be altered.

- **newname**
  The new name of the role.

- **config_parameter=value**
  Set this role’s session default for the specified configuration parameter to the given value. If `value` is `DEFAULT` or if `RESET` is used, the role-specific variable setting is removed, so the role will inherit the system-wide default setting in new sessions. Use `RESET ALL` to clear all role-specific settings. See `SET` and “Server Configuration Parameters” on page 770 for more information on user-settable configuration parameters.

- **queue_name**
  The name of the resource queue to which the user-level role is to be assigned. Only roles with `LOGIN` privilege can be assigned to a resource queue. To unassign a role from a resource queue and put it in the default resource queue, specify `NONE`. A role can only belong to one resource queue.

**SUPERUSER | NOSUPERUSER**
**CREATEDB | NOCREATEDB**
**CREATEROLE | NOCREATEROLE**
**CREATEEXTTABLE | NOCREATEEXTTABLE [(attribute='value')]**
**INHERIT | NOINHERIT**
**LOGIN | NOLOGIN**
**CONNECTION LIMIT connlimit**
**PASSWORD password**
**ENCRYPTED | UNENCRYPTED**
**VALID UNTIL 'timestamp'**

These clauses alter role attributes originally set by `CREATE ROLE`.

**Warning:** Setting a connection limit at the role level cannot be enforced in Greenplum Database and may cause queries to fail. Leave this set at the default of `-1` (no limit). Connection limits may only be set at the system level in Greenplum Database. See “Limiting Concurrent Connections” on page 38 for more information.
Notes
Use `GRANT` and `REVOKE` for adding and removing role memberships.
Caution must be exercised when specifying an unencrypted password with this command. The password will be transmitted to the server in clear text, and it might also be logged in the client’s command history or the server log. The `psql` command-line client contains a meta-command `\password` that can be used to safely change a role’s password.
It is also possible to tie a session default to a specific database rather than to a role. Role-specific settings override database-specific ones if there is a conflict. See `ALTER DATABASE`.

Examples
Change a role’s password:
```
ALTER ROLE daria WITH PASSWORD 'passwd123';
```
Change a password expiration date:
```
ALTER ROLE scott VALID UNTIL 'May 4 12:00:00 2015 +1';
```
Make a password valid forever:
```
ALTER ROLE luke VALID UNTIL 'infinity';
```
Give a role the ability to create other roles and new databases:
```
ALTER ROLE joelle CREATEROLE CREATEDB;
```
Give a role a non-default setting of the `maintenance_work_mem` parameter:
```
ALTER ROLE admin SET maintenance_work_mem = 100000;
```
Assign a role to a resource queue:
```
ALTER ROLE sammy RESOURCE QUEUE poweruser;
```
Give a role permission to create writable external tables:
```
ALTER ROLE load CREATEEXTTABLE (type='writable');
```

Compatibility
The `ALTER ROLE` statement is a Greenplum Database extension.

See Also
CREATE ROLE, DROP ROLE, SET, CREATE RESOURCE QUEUE, GRANT, REVOKE
ALTER SCHEMA

Changes the definition of a schema.

Synopsis

ALTER SCHEMA name RENAME TO newname
ALTER SCHEMA name OWNER TO newowner

Description

ALTER SCHEMA changes the definition of a schema.

You must own the schema to use ALTER SCHEMA. To rename a schema you must also have the CREATE privilege for the database. To alter the owner, you must also be a direct or indirect member of the new owning role, and you must have the CREATE privilege for the database. Note that superusers have all these privileges automatically.

Parameters

name

The name of an existing schema.

newname

The new name of the schema. The new name cannot begin with pg_, as such names are reserved for system schemas.

newowner

The new owner of the schema.

Compatibility

There is no ALTER SCHEMA statement in the SQL standard.

See Also

CREATE SCHEMA, DROP SCHEMA
ALTER SEQUENCE

Changes the definition of a sequence generator.

Synopsis

ALTER SEQUENCE name [INCREMENT [ BY ] increment]
    [MINVALUE minvalue | NO MINVALUE]
    [MAXVALUE maxvalue | NO MAXVALUE]
    [RESTART [ WITH ] start]
    [CACHE cache] [[ NO ] CYCLE]
    [OWNED BY {table.column | NONE}]

ALTER SEQUENCE name SET SCHEMA new_schema

Description

ALTER SEQUENCE changes the parameters of an existing sequence generator. Any parameters not specifically set in the ALTER SEQUENCE command retain their prior settings.

You must own the sequence to use ALTER SEQUENCE. To change a sequence’s schema, you must also have CREATE privilege on the new schema. Note that superusers have all these privileges automatically.

Parameters

name

The name (optionally schema-qualified) of a sequence to be altered.

increment

The clause INCREMENT BY increment is optional. A positive value will make an ascending sequence, a negative one a descending sequence. If unspecified, the old increment value will be maintained.

minvalue

The optional clause MINVALUE minvalue determines the minimum value a sequence can generate. If NO MINVALUE is specified, the defaults of 1 and -263-1 for ascending and descending sequences, respectively, will be used. If neither option is specified, the current minimum value will be maintained.

maxvalue

The optional clause MAXVALUE maxvalue determines the maximum value for the sequence. If NO MAXVALUE is specified, the defaults are 263-1 and -1 for ascending and descending sequences, respectively, will be used. If neither option is specified, the current maximum value will be maintained.
### start

The optional clause `RESTART WITH start` changes the current value of the sequence.

### cache

The clause `CACHE cache` enables sequence numbers to be preallocated and stored in memory for faster access. The minimum value is 1 (only one value can be generated at a time, i.e., no cache). If unspecified, the old cache value will be maintained.

### CYCLE

The optional `CYCLE` key word may be used to enable the sequence to wrap around when the `maxvalue` or `minvalue` has been reached by an ascending or descending sequence. If the limit is reached, the next number generated will be the respective `minvalue` or `maxvalue`.

### NO CYCLE

If the optional `NO CYCLE` key word is specified, any calls to `nextval` after the sequence has reached its maximum value will return an error. If neither `CYCLE` or `NO CYCLE` are specified, the old cycle behavior will be maintained.

### OWNED BY table.column

**OWNED BY NONE**

The `OWNED BY` option causes the sequence to be associated with a specific table column, such that if that column (or its whole table) is dropped, the sequence will be automatically dropped as well. If specified, this association replaces any previously specified association for the sequence. The specified table must have the same owner and be in the same schema as the sequence. Specifying `OWNED BY NONE` removes any existing table column association.

### new_schema

The new schema for the sequence.

### Notes

To avoid blocking of concurrent transactions that obtain numbers from the same sequence, `ALTER SEQUENCE`’s effects on the sequence generation parameters are never rolled back; those changes take effect immediately and are not reversible. However, the `OWNED BY` and `SET SCHEMA` clauses are ordinary catalog updates and can be rolled back.

`ALTER SEQUENCE` will not immediately affect `nextval` results in sessions, other than the current one, that have preallocated (cached) sequence values. They will use up all cached values prior to noticing the changed sequence generation parameters. The current session will be affected immediately.

Some variants of `ALTER TABLE` can be used with sequences as well. For example, to rename a sequence use `ALTER TABLE RENAME`. 
Examples
Restart a sequence called *serial*, at 105:

    ALTER SEQUENCE serial RESTART WITH 105;

Compatibility
`ALTER SEQUENCE` conforms to the SQL standard, except for the `OWNED BY` and `SET SCHEMA` clauses, which are Greenplum Database extensions.

See Also
`CREATE SEQUENCE`, `DROP SEQUENCE`, `ALTER TABLE`
ALTER TABLE

Changes the definition of a table.

Synopsis

ALTER TABLE [ONLY] name RENAME [COLUMN] column TO new_column
ALTER TABLE name RENAME TO new_name
ALTER TABLE name SET SCHEMA new_schema
ALTER TABLE [ONLY] name SET
    DISTRIBUTED BY (column, [ ... ])
    | DISTRIBUTED RANDOMLY
    | WITH (REORGANIZE=true|false)

ALTER TABLE [ONLY] name action [, ... ]
ALTER TABLE name
    [ ALTER PARTITION { partition_name | FOR (RANK(number)) | FOR (value) } partition_action [...] ]

where action is one of:
ADD [COLUMN] column_name type [column_constraint [ ... ]]
DROP [COLUMN] column [RESTRICT | CASCADE]
ALTER [COLUMN] column TYPE type [USING expression]
ALTER [COLUMN] column SET DEFAULT expression
ALTER [COLUMN] column DROP DEFAULT
ALTER [COLUMN] column { SET | DROP } NOT NULL
ALTER [COLUMN] column SET STATISTICS integer
ADD table_constraint
DROP CONSTRAINT constraint_name [RESTRICT | CASCADE]
DISABLE TRIGGER [trigger_name | ALL | USER]
ENABLE TRIGGER [trigger_name | ALL | USER]
CLUSTER ON index_name
SET WITHOUT CLUSTER
SET WITHOUT OIDS
SET (FILLFACTOR = value)
RESET (FILLFACTOR)
INHERIT parent_table
NO INHERIT parent_table
OWNER TO new_owner
SET TABLESPACE new_tablespace

where partition_action is one of:
ALTER DEFAULT PARTITION
DROP DEFAULT PARTITION [IF EXISTS]
DROP PARTITION [IF EXISTS] { partition_name | FOR (RANK(number)) | FOR (value) } [CASCADE]
TRUNCATE DEFAULT PARTITION
TRUNCATE PARTITION { partition_name | FOR (RANK(number)) | FOR (value) }
RENAME DEFAULT PARTITION TO new_partition_name
RENAME PARTITION { partition_name | FOR (RANK(number)) | FOR (value) } TO new_partition_name
ADD DEFAULT PARTITION name [ ( subpartition_spec ) ]
ADD PARTITION [name] partition_element
[ ( subpartition_spec ) ]
EXCHANGE PARTITION { partition_name | FOR (RANK(number)) | FOR (value) } WITH TABLE table_name
[ WITH | WITHOUT VALIDATION ]
EXCHANGE DEFAULT PARTITION WITH TABLE table_name
[ WITH | WITHOUT VALIDATION ]
SET SUBPARTITION TEMPLATE (subpartition_spec)
SPLIT DEFAULT PARTITION
{ AT (list_value)
 | START(([datatype] range_value) [INCLUSIVE | EXCLUSIVE]
 | END(([datatype] range_value) [INCLUSIVE | EXCLUSIVE] )
 | INTO ( PARTITION new_partition_name,
 | PARTITION default_partition_name ) ]
SPLIT PARTITION { partition_name | FOR (RANK(number)) | FOR (value) } AT (value)
[ INTO (PARTITION partition_name, PARTITION partition_name) ]
where partition_element is:
VALUES (list_value [, ...] )
| START ([datatype] 'start_value') [INCLUSIVE | EXCLUSIVE]
 | END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE] ]
| END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE]
[ WITH ( partition_storage_parameter=value [, ... ] ) ]
[ TABLESPACE tablespace ]
where subpartition_spec is:
subpartition_element [, ...]
and subpartition_element is:
DEFAULT SUBPARTITION name
| [ [SUBPARTITION name] VALUES (list_value [, ...] )
| [SUBPARTITION name]
 | START ([datatype] 'start_value') [INCLUSIVE | EXCLUSIVE]
 | END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE] ]
 | EVERY ( [number | datatype] 'interval_value' ) ]
| [SUBPARTITION name]
 | END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE]
 | EVERY ( [number | datatype] 'interval_value' ) ]
| WITH ( partition_storage_parameter=value [, ... ] ) ]

ALTER TABLE
[ TABLESPACE tablespace ]

where storage_parameter is:
  APPENDONLY={TRUE|FALSE}
  COMPRESSTYPE={ZLIB|QUICKLZ}
  COMPRESSLEVEL={0-9|1}
  ORIENTATION={COLUMN|ROW}
  FILLFACTOR={10-100}

Description

ALTER TABLE changes the definition of an existing table. There are several subforms:

- **ADD COLUMN** — Adds a new column to the table, using the same syntax as CREATE TABLE.

- **DROP COLUMN** — Drops a column from a table. Note that if you drop table columns that are being used as the Greenplum Database distribution key, the distribution policy for the table will be changed to DISTRIBUTED RANDOMLY. Indexes and table constraints involving the column will be automatically dropped as well. You will need to say CASCADE if anything outside the table depends on the column (such as views).

- **ALTER COLUMN TYPE** — Changes the data type of a column of a table. Note that you cannot alter column data types that are being used as the Greenplum Database distribution key. Indexes and simple table constraints involving the column will be automatically converted to use the new column type by reparsing the originally supplied expression. The optional USING clause specifies how to compute the new column value from the old. If omitted, the default conversion is the same as an assignment cast from old data type to new. A USING clause must be provided if there is no implicit or assignment cast from old to new type.

- **SET/DROP DEFAULT** — Sets or removes the default value for a column. The default values only apply to subsequent INSERT commands. They do not cause rows already in the table to change. Defaults may also be created for views, in which case they are inserted into statements on the view before the view’s ON INSERT rule is applied.

- **SET/DROP NOT NULL** — Changes whether a column is marked to allow null values or to reject null values. You can only use SET NOT NULL when the column contains no null values.

- **SET STATISTICS** — Sets the per-column statistics-gathering target for subsequent ANALYZE operations. The target can be set in the range 0 to 1000, or set to -1 to revert to using the system default statistics target (default_statistics_target).

- **ADD table_constraint** — Adds a new constraint to a table using the same syntax as CREATE TABLE.

- **DROP CONSTRAINT** — Drops the specified constraint on a table.
ALTER TABLE

• **DISABLE/ENABLE TRIGGER** — Disables or enables trigger(s) belonging to the table. A disabled trigger is still known to the system, but is not executed when its triggering event occurs. For a deferred trigger, the enable status is checked when the event occurs, not when the trigger function is actually executed. One may disable or enable a single trigger specified by name, or all triggers on the table, or only user-created triggers. Disabling or enabling constraint triggers requires superuser privileges. Note that foreign key constraint triggers are not currently supported in Greenplum Database, and triggers in general have very limited functionality due to the parallelism of Greenplum Database. See `CREATE TRIGGER` for more information.

• **CLUSTER/SET WITHOUT CLUSTER** — Selects or removes the default index for future `CLUSTER` operations. It does not actually re-cluster the table. Note that `CLUSTER` is not the recommended way to physically reorder a table in Greenplum Database because it takes so long. It is better to recreate the table with `CREATE TABLE AS` and order it by the index column(s).

• **SET WITHOUT OIDS** — Removes the OID system column from the table. Note that there is no variant of `ALTER TABLE` that allows OIDs to be restored to a table once they have been removed.

• **SET (FILLFACTOR = value) / RESET (FILLFACTOR)** — Changes the fillfactor for the table. The fillfactor for a table is a percentage between 10 and 100. 100 (complete packing) is the default. When a smaller fillfactor is specified, `INSERT` operations pack table pages only to the indicated percentage; the remaining space on each page is reserved for updating rows on that page. This gives `UPDATE` a chance to place the updated copy of a row on the same page as the original, which is more efficient than placing it on a different page. For a table whose entries are never updated, complete packing is the best choice, but in heavily updated tables smaller fillfactors are appropriate. Note that the table contents will not be modified immediately by this command. You will need to rewrite the table to get the desired effects.

• **SET DISTRIBUTED** — Changes the distribution policy of a table. Changes to a hash distribution policy will cause the table data to be physically redistributed on disk, which can be resource intensive.

• **INHERIT parent_table / NO INHERIT parent_table** — Adds or removes the target table as a child of the specified parent table. Queries against the parent will include records of its child table. To be added as a child, the target table must already contain all the same columns as the parent (it could have additional columns, too). The columns must have matching data types, and if they have `NOT NULL` constraints in the parent then they must also have `NOT NULL` constraints in the child. There must also be matching child-table constraints for all `CHECK` constraints of the parent.

• **OWNER** — Changes the owner of the table, sequence, or view to the specified user.

• **SET TABLESPACE** — Changes the table’s tablespace to the specified tablespace and moves the data file(s) associated with the table to the new tablespace. Indexes on the table, if any, are not moved; but they can be moved separately with additional `SET TABLESPACE` commands. See also `CREATE TABLESPACE`. If changing the tablespace of a partitioned table, all child table partitions will also be moved to the new tablespace.
- **RENAME** — Changes the name of a table (or an index, sequence, or view) or the name of an individual column in a table. There is no effect on the stored data. Note that Greenplum Database distribution key columns cannot be renamed.

- **SET SCHEMA** — Moves the table into another schema. Associated indexes, constraints, and sequences owned by table columns are moved as well.

- **ALTER PARTITION | DROP PARTITION | RENAME PARTITION | TRUNCATE PARTITION | ADD PARTITION | SPLIT PARTITION | EXCHANGE PARTITION | SET SUBPARTITION TEMPLATE** — Changes the structure of a partitioned table. In most cases, you must go through the parent table to alter one of its child table partitions.

You must own the table to use `ALTER TABLE`. To change the schema of a table, you must also have `CREATE` privilege on the new schema. To add the table as a new child of a parent table, you must own the parent table as well. To alter the owner, you must also be a direct or indirect member of the new owning role, and that role must have `CREATE` privilege on the table’s schema. A superuser has these privileges automatically.

### Parameters

- **ONLY**
  Only perform the operation on the table name specified. If the `ONLY` keyword is not used, the operation will be performed on the named table and any child table partitions associated with that table.

- **name**
  The name (possibly schema-qualified) of an existing table to alter. If `ONLY` is specified, only that table is altered. If `ONLY` is not specified, the table and all its descendant tables (if any) are updated.

- **column**
  Name of a new or existing column. Note that Greenplum Database distribution key columns must be treated with special care. Altering or dropping these columns can change the distribution policy for the table.

- **new_column**
  New name for an existing column.

- **new_name**
  New name for the table.

- **type**
  Data type of the new column, or new data type for an existing column. If changing the data type of a Greenplum distribution key column, you are only allowed to change it to a compatible type (for example, `text` to `varchar` is OK, but `text` to `int` is not).
**table_constraint**
New table constraint for the table. Note that foreign key constraints are currently not supported in Greenplum Database. Also a table is only allowed one unique constraint and the uniqueness must be within the Greenplum Database distribution key.

**constraint_name**
Name of an existing constraint to drop.

**CASCADE**
Automatically drop objects that depend on the dropped column or constraint (for example, views referencing the column).

**RESTRICT**
Refuse to drop the column or constraint if there are any dependent objects. This is the default behavior.

**trigger_name**
Name of a single trigger to disable or enable. Note that Greenplum Database has limited support of triggers. See CREATE TRIGGER for more information.

**ALL**
Disable or enable all triggers belonging to the table including constraint related triggers. This requires superuser privilege.

**USER**
Disable or enable all user-created triggers belonging to the table.

**index_name**
The index name on which the table should be marked for clustering. Note that CLUSTER is not the recommended way to physically reorder a table in Greenplum Database because it takes so long. It is better to recreate the table with CREATE TABLE AS and order it by the index column(s).

**FILLFACTOR**
Set the fillfactor percentage for a table.

**value**
The new value for the FILLFACTOR parameter, which is a percentage between 10 and 100. 100 is the default.

**DISTRIBUTED BY (column) | DISTRIBUTED RANDOMLY**
Specifies the distribution policy for a table. Changing a hash distribution policy will cause the table data to be physically redistributed on disk, which can be resource intensive. If you declare the same hash distribution policy or change from hash to random distribution, data will not be redistributed unless you declare SET WITH (REORGANIZE=true).
REORGANIZE=true|false
Use REORGANIZE=true when the hash distribution policy has not changed or when you have changed from a hash to a random distribution, and you want to redistribute the data anyways.

parent_table
A parent table to associate or de-associate with this table.

default_owner
The role name of the new owner of the table.

default_tablespace
The name of the tablespace to which the table will be moved.

default_schema
The name of the schema to which the table will be moved.

parent_table_name
When altering a partitioned table, the name of the top-level parent table.

ALTER [DEFAULT] PARTITION
If altering a partition deeper than the first level of partitions, the ALTER PARTITION clause is used to specify which subpartition in the hierarchy you want to alter.

DROP [DEFAULT] PARTITION
Drops the specified partition. If the partition has subpartitions, the subpartitions are automatically dropped as well.

TRUNCATE [DEFAULT] PARTITION
Truncates the specified partition. If the partition has subpartitions, the subpartitions are automatically truncated as well.

RENAME [DEFAULT] PARTITION
Changes the partition name of a partition (not the relation name). Partitioned tables are created using the naming convention:
<parentname>_<level>_prt_<partition_name>.

ADD DEFAULT PARTITION
Adds a default partition to an existing partition design. When data does not match to an existing partition, it is inserted into the default partition. Partition designs that do not have a default partition will reject incoming rows that do not match to an existing partition. Default partitions must be given a name.

ADD PARTITION
partition_element - Using the existing partition type of the table (range or list), defines the boundaries of new partition you are adding.

default_name - A name for this new partition.
VALUES - For list partitions, defines the value(s) that the partition will contain.

START - For range partitions, defines the starting range value for the partition. By default, start values are INCLUSIVE. For example, if you declared a start date of ‘2008-01-01’, then the partition would contain all dates greater than or equal to ‘2008-01-01’. Typically the data type of the START expression is the same type as the partition key column. If that is not the case, then you must explicitly cast to the intended data type.

END - For range partitions, defines the ending range value for the partition. By default, end values are EXCLUSIVE. For example, if you declared an end date of ‘2008-02-01’, then the partition would contain all dates less than but not equal to ‘2008-02-01’. Typically the data type of the END expression is the same type as the partition key column. If that is not the case, then you must explicitly cast to the intended data type.

WITH - Sets the table storage options for a partition. For example, you may want older partitions to be append-only tables and newer partitions to be regular heap tables. See “CREATE TABLE” on page 407 for a description of the storage options.

TABLESPACE - The name of the tablespace in which the partition is to be created.

subpartition_spec - Only allowed on partition designs that were created without a subpartition template. Declares a subpartition specification for the new partition you are adding. If the partitioned table was originally defined using a subpartition template, then the template will be used to generate the subpartitions automatically.

EXCHANGE [DEFAULT] PARTITION
Exchanges another table into the partition hierarchy into the place of an existing partition. In a multi-level partition design, you can only exchange the lowest level partitions (those that contain data).

WITH TABLE table_name - The name of the table you are swapping in to the partition design.

WITH | WITHOUT VALIDATION - Validates that the data in the table matches the CHECK constraint of the partition you are exchanging. The default is to validate the data against the CHECK constraint.

SET SUBPARTITION TEMPLATE
Modifies the subpartition template for an existing partition. After a new subpartition template is set, all new partitions added will have the new subpartition design (existing partitions are not modified).

SPLIT DEFAULT PARTITION
Splits a default partition. In a multi-level partition design, you can only split the lowest level default partitions (those that contain data). Splitting a default partition creates a new partition containing the values specified and leaves the default partition containing any values that do not match to an existing partition.
AT - For list partitioned tables, specifies a single list value that should be used as the criteria for the split.

START - For range partitioned tables, specifies a starting value for the new partition.

END - For range partitioned tables, specifies an ending value for the new partition.

INTO - Allows you to specify a name for the new partition. When using the INTO clause to split a default partition, the second partition name specified should always be that of the existing default partition. If you do not know the name of the default partition, you can look it up using the pg_partitions view.

SPLIT PARTITION
Splits an existing partition into two partitions. In a multi-level partition design, you can only split the lowest level partitions (those that contain data).

AT - Specifies a single value that should be used as the criteria for the split. The partition will be divided into two new partitions with the split value specified being the starting range for the latter partition.

INTO - Allows you to specify names for the two new partitions created by the split.

partition_name
The given name of a partition.

FOR (RANK(number))
For range partitions, the rank of the partition in the range.

FOR ('value')
Specifies a partition by declaring a value that falls within the partition boundary specification. If the value declared with FOR matches to both a partition and one of its subpartitions (for example, if the value is a date and the table is partitioned by month and then by day), then FOR will operate on the first level where a match is found (for example, the monthly partition). If your intent is to operate on a subpartition, you must declare so as follows:

```sql
ALTER TABLE name ALTER PARTITION FOR ('2008-10-01') DROP PARTITION FOR ('2008-10-01');
```

Notes
Take special care when altering or dropping columns that are part of the Greenplum Database distribution key as this can change the distribution policy for the table.

Greenplum Database does not currently support foreign key constraints. For a unique constraint to be enforced in Greenplum Database, the table must be hash-distributed (not DISTRIBUTED RANDOMLY), and all of the distribution key columns must be the same as the initial columns of the unique constraint columns.

Adding a CHECK or NOT NULL constraint requires scanning the table to verify that existing rows meet the constraint.
When a column is added with `ADD COLUMN`, all existing rows in the table are initialized with the column's default value (`NULL` if no `DEFAULT` clause is specified). Adding a column with a non-null default or changing the type of an existing column will require the entire table to be rewritten. This may take a significant amount of time for a large table; and it will temporarily require double the disk space.

You can specify multiple changes in a single `ALTER TABLE` command, which will be done in a single pass over the table.

The `DROP COLUMN` form does not physically remove the column, but simply makes it invisible to SQL operations. Subsequent insert and update operations in the table will store a null value for the column. Thus, dropping a column is quick but it will not immediately reduce the on-disk size of your table, as the space occupied by the dropped column is not reclaimed. The space will be reclaimed over time as existing rows are updated.

The fact that `ALTER TYPE` requires rewriting the whole table is sometimes an advantage, because the rewriting process eliminates any dead space in the table. For example, to reclaim the space occupied by a dropped column immediately, the fastest way is:

```
ALTER TABLE table ALTER COLUMN anycol TYPE sametype;
```

Where `anycol` is any remaining table column and `sametype` is the same type that column already has. This results in no semantically-visible change in the table, but the command forces rewriting, which gets rid of no-longer-useful data.

If a table is partitioned or has any descendant tables, it is not permitted to add, rename, or change the type of a column in the parent table without doing the same to the descendants. This ensures that the descendants always have columns matching the parent.

To see the structure of a partitioned table, you can use the view `pg_partitions`. This view can help identify the particular partitions you may want to alter.

A recursive `DROP COLUMN` operation will remove a descendant table’s column only if the descendant does not inherit that column from any other parents and never had an independent definition of the column. A nonrecursive `DROP COLUMN` (`ALTER TABLE ONLY ... DROP COLUMN`) never removes any descendant columns, but instead marks them as independently defined rather than inherited.

The `TRIGGER`, `CLUSTER`, `OWNER`, and `TABLESPACE` actions never recurse to descendant tables; that is, they always act as though `ONLY` were specified. Adding a constraint can recurse only for `CHECK` constraints.

Changing any part of a system catalog table is not permitted.

**Examples**

Add a column to a table:

```
ALTER TABLE distributors ADD COLUMN address varchar(30);
```

Rename an existing column:

```
ALTER TABLE distributors RENAME COLUMN address TO city;
```

Rename an existing table:

```
ALTER TABLE distributors RENAME TO suppliers;
```
Add a not-null constraint to a column:

   ALTER TABLE distributors ALTER COLUMN street SET NOT NULL;

Add a check constraint to a table:

   ALTER TABLE distributors ADD CONSTRAINT zipchk CHECK (char_length(zipcode) = 5);

Move a table to a different schema:

   ALTER TABLE myschema.distributors SET SCHEMA yourschema;

Add a new partition to a partitioned table:

   ALTER TABLE sales ADD PARTITION
       START (date '2009-02-01') INCLUSIVE
       END (date '2009-03-01') EXCLUSIVE;

Add a default partition to an existing partition design:

   ALTER TABLE sales ADD DEFAULT PARTITION other;

Rename a partition:

   ALTER TABLE sales RENAME PARTITION FOR ('2008-01-01') TO jan08;

Drop the first (oldest) partition in a range sequence:

   ALTER TABLE sales DROP PARTITION FOR (RANK(1));

Exchange a table into your partition design:

   ALTER TABLE sales EXCHANGE PARTITION FOR ('2008-01-01') WITH TABLE jan08;

Split the default partition (where the existing default partition’s name is ‘other’) to add a new monthly partition for January 2009:

   ALTER TABLE sales SPLIT DEFAULT PARTITION
       START ('2009-01-01') INCLUSIVE
       END ('2009-02-01') EXCLUSIVE
       INTO (PARTITION jan09, PARTITION other);

Split a monthly partition into two with the first partition containing dates January 1-15 and the second partition containing dates January 16-31:

   ALTER TABLE sales SPLIT PARTITION FOR ('2008-01-01')
       AT ('2008-01-16')
       INTO (PARTITION jan081to15, PARTITION jan0816to31);

**Compatibility**

The **ADD, DROP, and SET DEFAULT** forms conform with the SQL standard. The other forms are Greenplum Database extensions of the SQL standard. Also, the ability to specify more than one manipulation in a single **ALTER TABLE** command is an extension.
ALTER TABLE DROP COLUMN can be used to drop the only column of a table, leaving a zero-column table. This is an extension of SQL, which disallows zero-column tables.

See Also

CREATE TABLE, DROP TABLE
ALTER TABLESPACE

Changes the definition of a tablespace.

**Synopsis**

ALTER TABLESPACE name RENAME TO newname
ALTER TABLESPACE name OWNER TO newowner

**Description**

ALTER TABLESPACE changes the definition of a tablespace.

You must own the tablespace to use ALTER TABLESPACE. To alter the owner, you must also be a direct or indirect member of the new owning role. (Note that superusers have these privileges automatically.)

**Parameters**

*name*

The name of an existing tablespace.

*newname*

The new name of the tablespace. The new name cannot begin with `pg_` or `gp_` (reserved for system tablespaces).

*newowner*

The new owner of the tablespace.

**Examples**

Rename tablespace `index_space` to `fast_raid`:

```sql
ALTER TABLESPACE index_space RENAME TO fast_raid;
```

Change the owner of tablespace `index_space`:

```sql
ALTER TABLESPACE index_space OWNER TO mary;
```

**Compatibility**

There is no ALTER TABLESPACE statement in the SQL standard.

**See Also**

CREATE TABLESPACE, DROP TABLESPACE
ALTER TRIGGER

Changes the definition of a trigger.

Synopsis

ALTER TRIGGER name ON table RENAME TO newname

Description

ALTER TRIGGER changes properties of an existing trigger. The RENAME clause changes the name of the given trigger without otherwise changing the trigger definition. You must own the table on which the trigger acts to be allowed to change its properties.

Parameters

name
The name of an existing trigger to alter.

table
The name of the table on which this trigger acts.

newname
The new name for the trigger.

Notes

The ability to temporarily enable or disable a trigger is provided by ALTER TABLE, not by ALTER TRIGGER, because ALTER TRIGGER has no convenient way to express the option of enabling or disabling all of a table’s triggers at once.

Note that Greenplum Database has limited support of triggers in this release. See CREATE TRIGGER for more information.

Examples

To rename an existing trigger:

ALTER TRIGGER emp_stamp ON emp RENAME TO emp_track_chgs;

Compatibility

ALTER TRIGGER is a Greenplum Database extension of the SQL standard.

See Also

ALTER TABLE, CREATE TRIGGER, DROP TRIGGER
ALTER TYPE

Changes the definition of a data type.

Synopsis

ALTER TYPE name OWNER TO new_owner
ALTER TYPE name SET SCHEMA new_schema

Description

ALTER TYPE changes the definition of an existing type. The only currently available capabilities are changing the owner and schema of a type.

You must own the type to use ALTER TYPE. To change the schema of a type, you must also have CREATE privilege on the new schema. To alter the owner, you must also be a direct or indirect member of the new owning role, and that role must have CREATE privilege on the type’s schema. (These restrictions enforce that altering the owner doesn't do anything you couldn't do by dropping and recreating the type. However, a superuser can alter ownership of any type anyway.)

Parameters

name
The name (optionally schema-qualified) of an existing type to alter.

new_owner
The user name of the new owner of the type.

new_schema
The new schema for the type.

Examples

To change the owner of the user-defined type email to joe:

    ALTER TYPE email OWNER TO joe;

To change the schema of the user-defined type email to customers:

    ALTER TYPE email SET SCHEMA customers;

Compatibility

There is no ALTER TYPE statement in the SQL standard.

See Also

CREATE TYPE, DROP TYPE
ALTER USER

Changes the definition of a database role (user).

Synopsis

ALTER USER name RENAME TO newname
ALTER USER name SET config_parameter {TO | =} {value | DEFAULT}
ALTER USER name RESET config_parameter
ALTER USER name [ [WITH] option [ ... ] ]

where option can be:

- SUPERUSER | NOSUPERUSER
- CREATEDB | NOCREATEDB
- CREATEROLE | NOCREATEROLE
- CREATEUSER | NOCREATEUSER
- INHERIT | NOINHERIT
- LOGIN | NOLOGIN
- [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
- VALID UNTIL 'timestamp'

Description

ALTER USER is a deprecated command but is still accepted for historical reasons. It is an alias for ALTER ROLE. See ALTER ROLE for more information.

Compatibility

The ALTER USER statement is a Greenplum Database extension. The SQL standard leaves the definition of users to the implementation.

See Also

ALTER ROLE
ANALYZE

Collects statistics about a database.

Synopsis

ANALYZE [VERBOSE] [table [ (column [, ...] ) ]]

Description

ANALYZE collects statistics about the contents of tables in the database, and stores the results in the system table `pg_statistic`. Subsequently, the query planner uses these statistics to help determine the most efficient execution plans for queries.

With no parameter, ANALYZE examines every table in the current database. With a parameter, ANALYZE examines only that table. It is further possible to give a list of column names, in which case only the statistics for those columns are collected.

Parameters

VERBOSE

Enables display of progress messages. When specified, ANALYZE emits progress messages to indicate which table is currently being processed. Various statistics about the tables are printed as well.

table

The name (possibly schema-qualified) of a specific table to analyze. Defaults to all tables in the current database.

column

The name of a specific column to analyze. Defaults to all columns.

Notes

It is a good idea to run ANALYZE periodically, or just after making major changes in the contents of a table. Accurate statistics will help the query planner to choose the most appropriate query plan, and thereby improve the speed of query processing. A common strategy is to run VACUUM and ANALYZE once a day during a low-usage time of day.

ANALYZE requires only a read lock on the target table, so it can run in parallel with other activity on the table.

The statistics collected by ANALYZE usually include a list of some of the most common values in each column and a histogram showing the approximate data distribution in each column. One or both of these may be omitted if ANALYZE deems them uninteresting (for example, in a unique-key column, there are no common values) or if the column data type does not support the appropriate operators.
For large tables, `ANALYZE` takes a random sample of the table contents, rather than examining every row. This allows even very large tables to be analyzed in a small amount of time. Note, however, that the statistics are only approximate, and will change slightly each time `ANALYZE` is run, even if the actual table contents did not change. This may result in small changes in the planner’s estimated costs shown by `EXPLAIN`. In rare situations, this non-determinism will cause the query optimizer to choose a different query plan between runs of `ANALYZE`. To avoid this, raise the amount of statistics collected by `ANALYZE` by adjusting the `default_statistics_target` configuration parameter, or on a column-by-column basis by setting the per-column statistics target with `ALTER TABLE ... ALTER COLUMN ... SET STATISTICS` (see `ALTER TABLE`). The target value sets the maximum number of entries in the most-common-value list and the maximum number of bins in the histogram. The default target value is 10, but this can be adjusted up or down to trade off accuracy of planner estimates against the time taken for `ANALYZE` and the amount of space occupied in `pg_statistic`. In particular, setting the statistics target to zero disables collection of statistics for that column. It may be useful to do that for columns that are never used as part of the `WHERE`, `GROUP BY`, or `ORDER BY` clauses of queries, since the planner will have no use for statistics on such columns.

The largest statistics target among the columns being analyzed determines the number of table rows sampled to prepare the statistics. Increasing the target causes a proportional increase in the time and space needed to do `ANALYZE`.

**Examples**

Collect statistics for the table `mytable`:

```sql
ANALYZE mytable;
```

**Compatibility**

There is no `ANALYZE` statement in the SQL standard.

**See Also**

`ALTER TABLE`, `EXPLAIN`, `VACUUM`
BEGIN

Starts a transaction block.

**Synopsis**

BEGIN [WORK | TRANSACTION] [SERIALIZABLE | REPEATABLE READ | READ COMMITTED | READ UNCOMMITTED] [READ WRITE | READ ONLY]

**Description**

BEGIN initiates a transaction block, that is, all statements after a BEGIN command will be executed in a single transaction until an explicit COMMIT or ROLLBACK is given. By default (without BEGIN), Greenplum Database executes transactions in autocommit mode, that is, each statement is executed in its own transaction and a commit is implicitly performed at the end of the statement (if execution was successful, otherwise a rollback is done).

Statements are executed more quickly in a transaction block, because transaction start/commit requires significant CPU and disk activity. Execution of multiple statements inside a transaction is also useful to ensure consistency when making several related changes: other sessions will be unable to see the intermediate states wherein not all the related updates have been done.

If the isolation level or read/write mode is specified, the new transaction has those characteristics, as if SET TRANSACTION was executed.

**Parameters**

**WORK**
**TRANSACTION**

Optional key words. They have no effect.

**SERIALIZABLE**
**REPEATABLE READ**
**READ COMMITTED**
**READ UNCOMMITTED**

The SQL standard defines four transaction isolation levels: READ COMMITTED, READ UNCOMMITTED, SERIALIZABLE, and REPEATABLE READ. The default behavior is that a statement can only see rows committed before it began (READ COMMITTED). In Greenplum Database READ UNCOMMITTED is treated the same as READ COMMITTED. SERIALIZABLE is supported the same as REPEATABLE READ wherein all statements of the current transaction can only see rows committed before the first statement was executed in the transaction. SERIALIZABLE is the strictest transaction isolation. This level emulates serial transaction execution, as if transactions had been executed one after another, serially, rather than concurrently. Applications using this level must be prepared to retry transactions due to serialization failures.
**READ WRITE**

**READ ONLY**

Determines whether the transaction is read/write or read-only. Read/write is the default. When a transaction is read-only, the following SQL commands are disallowed: INSERT, UPDATE, DELETE, and COPY FROM if the table they would write to is not a temporary table; all CREATE, ALTER, and DROP commands; GRANT, REVOKE, TRUNCATE; and EXPLAIN ANALYZE and EXECUTE if the command they would execute is among those listed.

---

**Notes**

START TRANSACTION has the same functionality as BEGIN.

Use COMMIT or ROLLBACK to terminate a transaction block.

Issuing BEGIN when already inside a transaction block will provoke a warning message. The state of the transaction is not affected. To nest transactions within a transaction block, use savepoints (see SAVEPOINT).

---

**Examples**

To begin a transaction block:

```
  BEGIN;
```

---

**Compatibility**

BEGIN is a Greenplum Database language extension. It is equivalent to the SQL-standard command START TRANSACTION.

Incidentally, the BEGIN key word is used for a different purpose in embedded SQL. You are advised to be careful about the transaction semantics when porting database applications.

---

**See Also**

COMMIT, ROLLBACK, START TRANSACTION, SAVEPOINT
CHECKPOINT

Forces a transaction log checkpoint.

Synopsis

CHECKPOINT

Description

Write-Ahead Logging (WAL) puts a checkpoint in the transaction log every so often. The automatic checkpoint interval is set per Greenplum Database segment instance by the server configuration parameters `checkpoint_segments` and `checkpoint_timeout`. The `CHECKPOINT` command forces an immediate checkpoint when the command is issued, without waiting for a scheduled checkpoint.

A checkpoint is a point in the transaction log sequence at which all data files have been updated to reflect the information in the log. All data files will be flushed to disk.

Only superusers may call `CHECKPOINT`. The command is not intended for use during normal operation.

Compatibility

The `CHECKPOINT` command is a Greenplum Database language extension.
CLOSE

Closes a cursor.

Synopsis

CLOSE cursor_name

Description

CLOSE frees the resources associated with an open cursor. After the cursor is closed, no subsequent operations are allowed on it. A cursor should be closed when it is no longer needed.

Every non-holdable open cursor is implicitly closed when a transaction is terminated by COMMIT or ROLLBACK. A holdable cursor is implicitly closed if the transaction that created it aborts via ROLLBACK. If the creating transaction successfully commits, the holdable cursor remains open until an explicit CLOSE is executed, or the client disconnects.

Parameters

cursor_name

The name of an open cursor to close.

Notes

Greenplum Database does not have an explicit OPEN cursor statement. A cursor is considered open when it is declared. Use the DECLARE statement to declare (and open) a cursor.

You can see all available cursors by querying the pg_cursors system view.

Examples

Close the cursor portala:

    CLOSE portala;

Compatibility

CLOSE is fully conforming with the SQL standard.

See Also

DECLARE, FETCH, MOVE
**CLUSTER**

Physically reorders a heap storage table on disk according to an index. Not a recommended operation in Greenplum Database.

**Synopsis**

```sql
CLUSTER indexname ON tablename
CLUSTER tablename
CLUSTER
```

**Description**

`CLUSTER` orders a heap storage table based on an index. `CLUSTER` is not supported on append-only storage tables. Clustering an index means that the records are physically ordered on disk according to the index information. If the records you need are distributed randomly on disk, then the database has to seek across the disk to get the records requested. If those records are stored more closely together, then the fetching from disk is more sequential. A good example for a clustered index is on a date column where the data is ordered sequentially by date. A query against a specific date range will result in an ordered fetch from the disk, which leverages faster sequential access.

Clustering is a one-time operation: when the table is subsequently updated, the changes are not clustered. That is, no attempt is made to store new or updated rows according to their index order. If one wishes, one can periodically recluster by issuing the command again.

When a table is clustered using this command, Greenplum Database remembers on which index it was clustered. The form `CLUSTER tablename` reclusters the table on the same index that it was clustered before. `CLUSTER` without any parameter reclusters all previously clustered tables in the current database that the calling user owns, or all tables if called by a superuser. This form of `CLUSTER` cannot be executed inside a transaction block.

When a table is being clustered, an `ACCESS EXCLUSIVE` lock is acquired on it. This prevents any other database operations (both reads and writes) from operating on the table until the `CLUSTER` is finished.

**Parameters**

- `indexname`
  
  The name of an index.

- `tablename`

  The name (optionally schema-qualified) of a table.
Notes

In cases where you are accessing single rows randomly within a table, the actual order of the data in the table is unimportant. However, if you tend to access some data more than others, and there is an index that groups them together, you will benefit from using \texttt{CLUSTER}. If you are requesting a range of indexed values from a table, or a single indexed value that has multiple rows that match, \texttt{CLUSTER} will help because once the index identifies the table page for the first row that matches, all other rows that match are probably already on the same table page, and so you save disk accesses and speed up the query.

During the cluster operation, a temporary copy of the table is created that contains the table data in the index order. Temporary copies of each index on the table are created as well. Therefore, you need free space on disk at least equal to the sum of the table size and the index sizes.

Because the query planner records statistics about the ordering of tables, it is advisable to run \texttt{ANALYZE} on the newly clustered table. Otherwise, the planner may make poor choices of query plans.

There is another way to cluster data. The \texttt{CLUSTER} command reorders the original table by scanning it using the index you specify. This can be slow on large tables because the rows are fetched from the table in index order, and if the table is disordered, the entries are on random pages, so there is one disk page retrieved for every row moved. (Greenplum Database has a cache, but the majority of a big table will not fit in the cache.) The other way to cluster a table is to use a statement such as:

\begin{verbatim}
CREATE TABLE newtable AS SELECT * FROM table ORDER BY column;
DROP table;
ALTER TABLE newtable RENAME TO table;
CREATE INDEX column_ix ON table (column);
VACUUM ANALYZE table;
\end{verbatim}

This uses the Greenplum Database sorting code to produce the desired order, which is usually much faster than an index scan for disordered data. Then you drop the old table, use \texttt{ALTER TABLE ... RENAME} to rename \texttt{newtable} to the old name, and recreate the table’s indexes. The big disadvantage of this approach is that it does not preserve OIDs, constraints, granted privileges, and other ancillary properties of the table — all such items must be manually recreated. Another disadvantage is that this way requires a sort temporary file about the same size as the table itself, so peak disk usage is about three times the table size instead of twice the table size.

Examples

Cluster the table \texttt{employees} on the basis of its index \texttt{emp_ind}:

\begin{verbatim}
CLUSTER emp_ind ON emp;
\end{verbatim}

Cluster a large table by recreating it and loading it in the correct index order:

\begin{verbatim}
CREATE TABLE newtable AS SELECT * FROM table ORDER BY column;
DROP table;
ALTER TABLE newtable RENAME TO table;
CREATE INDEX column_ix ON table (column);
VACUUM ANALYZE table;
\end{verbatim}
**Compatibility**

There is no `CLUSTER` statement in the SQL standard.

**See Also**

CREATE TABLE AS, CREATE INDEX
**COMMENT**

Defines or change the comment of an object.

**Synopsis**

```sql
COMMENT ON
{ TABLE object_name |
  COLUMN table_name.column_name | 
  AGGREGATE agg_name (agg_type [, ...]) | 
  CAST (sourcetype AS targettype) |
  CONSTRAINT constraint_name ON table_name |
  CONVERSION object_name |
  DATABASE object_name |
  DOMAIN object_name |
  FILESPACE object_name |
  FUNCTION func_name ([argmode] [argname] argtype [, ...]) |
  INDEX object_name |
  LARGE OBJECT large_object_oid |
  OPERATOR op (leftoperand_type, rightoperand_type) |
  OPERATOR CLASS object_name USING index_method |
  [PROCEDURAL] LANGUAGE object_name |
  RESOURCE QUEUE object_name |
  ROLE object_name |
  RULE rule_name ON table_name |
  SCHEMA object_name |
  SEQUENCE object_name |
  TABLESPACE object_name |
  TRIGGER trigger_name ON table_name |
  TYPE object_name |
  VIEW object_name }
IS 'text'
```

**Description**

`COMMENT` stores a comment about a database object. To modify a comment, issue a new `COMMENT` command for the same object. Only one comment string is stored for each object. To remove a comment, write `NULL` in place of the text string. Comments are automatically dropped when the object is dropped.

Comments can be easily retrieved with the `psql` meta-commands `\dd`, `\d+`, and `\l+`. Other user interfaces to retrieve comments can be built atop the same built-in functions that `psql` uses, namely `obj_description`, `col_description`, and `shobj_description`.

**Parameters**

- `object_name`  
- `table_name.column_name`
agg_name
constraint_name
func_name
op
rule_name
trigger_name

The name of the object to be commented. Names of tables, aggregates, domains, functions, indexes, operators, operator classes, sequences, types, and views may be schema-qualified.

agg_type

An input data type on which the aggregate function operates. To reference a zero-argument aggregate function, write * in place of the list of input data types.

sourcetype

The name of the source data type of the cast.

targettype

The name of the target data type of the cast.

argmode

The mode of a function argument: either IN, OUT, or INOUT. If omitted, the default is IN. Note that COMMENT ON FUNCTION does not actually pay any attention to OUT arguments, since only the input arguments are needed to determine the function’s identity. So it is sufficient to list the IN and INOUT arguments.

argname

The name of a function argument. Note that COMMENT ON FUNCTION does not actually pay any attention to argument names, since only the argument data types are needed to determine the function’s identity.

argtype

The data type(s) of the function’s arguments (optionally schema-qualified), if any.

large_object_oid

The OID of the large object.

PROCEDURAL

This is a noise word.

text

The new comment, written as a string literal; or NULL to drop the comment.

Notes

There is presently no security mechanism for comments: any user connected to a database can see all the comments for objects in that database (although only superusers can change comments for objects that they do not own). For shared objects
such as databases, roles, and table spaces comments are stored globally and any user connected to any database can see all the comments for shared objects. Therefore, do not put security-critical information in comments.

**Examples**

Attach a comment to the table *mytable*:

```sql
COMMENT ON TABLE mytable IS 'This is my table.';
```

Remove it again:

```sql
COMMENT ON TABLE mytable IS NULL;
```

**Compatibility**

There is no `COMMENT` statement in the SQL standard.
COMMIT

Commits the current transaction.

Synopsis

COMMIT [WORK | TRANSACTION]

Description

COMMIT commits the current transaction. All changes made by the transaction become visible to others and are guaranteed to be durable if a crash occurs.

Parameters

WORK
TRANSACTION

Optional key words. They have no effect.

Notes

Use ROLLBACK to abort a transaction.

Issuing COMMIT when not inside a transaction does no harm, but it will provoke a warning message.

Examples

To commit the current transaction and make all changes permanent:

    COMMIT;

Compatibility

The SQL standard only specifies the two forms COMMIT and COMMIT WORK. Otherwise, this command is fully conforming.

See Also

BEGIN, END, START TRANSACTION, ROLLBACK
COPY

Copies data between a file and a table.

Synopsis
COPY table [((column [, ...]))] FROM ('file' | STDIN)
[ [WITH]
  [OIDS]
  [HEADER]
  [DELIMITER [ AS ] 'delimiter']
  [NULL [ AS ] 'null string']
  [ESCAPE [ AS ] 'escape' | 'OFF']
  [NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
  [CSV [QUOTE [ AS ] 'quote']
    [FORCE NOT NULL column [, ...]]
  [FILL MISSING FIELDS]
 [ [LOG ERRORS INTO error_table] [KEEP]
  SEGMENT REJECT LIMIT count [ROWS | PERCENT] ]
COPY (table [((column [, ...]))] | (query)) TO ('file' | STDOUT)
[ [WITH]
  [OIDS]
  [HEADER]
  [DELIMITER [ AS ] 'delimiter']
  [NULL [ AS ] 'null string']
  [ESCAPE [ AS ] 'escape' | 'OFF']
  [CSV [QUOTE [ AS ] 'quote']
    [FORCE QUOTE column [, ...]] ]

Description
COPY moves data between Greenplum Database tables and standard file-system files. COPY TO copies the contents of a table to a file, while COPY FROM copies data from a file to a table (appending the data to whatever is in the table already). COPY TO can also copy the results of a SELECT query.

If a list of columns is specified, COPY will only copy the data in the specified columns to or from the file. If there are any columns in the table that are not in the column list, COPY FROM will insert the default values for those columns.

COPY with a file name instructs the Greenplum Database master host to directly read from or write to a file. The file must be accessible to the master host and the name must be specified from the viewpoint of the master host. When STDIN or STDOUT is specified, data is transmitted via the connection between the client and the master.

If SEGMENT REJECT LIMIT is used, then a COPY FROM operation will operate in single row error isolation mode. In this release, single row error isolation mode only applies to rows in the input file with format errors — for example, extra or missing attributes, attributes of a wrong data type, or invalid client encoding sequences. Constraint errors such as violation of a NOT NULL, CHECK, or UNIQUE constraint will still be handled in
‘all-or-nothing’ input mode. The user can specify the number of error rows acceptable (on a per-segment basis), after which the entire COPY FROM operation will be aborted and no rows will be loaded. Note that the count of error rows is per-segment, not per entire load operation. If the per-segment reject limit is not reached, then all rows not containing an error will be loaded. If the limit is not reached, all good rows will be loaded and any error rows discarded. If you would like to keep error rows for further examination, you can optionally declare an error table using the LOG ERRORS INTO clause. Any rows containing a format error would then be logged to the specified error table.

**Outputs**

On successful completion, a COPY command returns a command tag of the form, where `count` is the number of rows copied:

```
COPY count
```

If running a COPY FROM command in single row error isolation mode, the following notice message will be returned if any rows were not loaded due to format errors, where `count` is the number of rows rejected:

```
NOTICE: Rejected count badly formatted rows.
```

**Parameters**

**table**

The name (optionally schema-qualified) of an existing table.

**column**

An optional list of columns to be copied. If no column list is specified, all columns of the table will be copied.

**query**

A SELECT or VALUES command whose results are to be copied. Note that parentheses are required around the query.

**file**

The absolute path name of the input or output file.

**STDIN**

Specifies that input comes from the client application.

**STDOUT**

Specifies that output goes to the client application.

**OIDS**

Specifies copying the OID for each row. (An error is raised if OIDS is specified for a table that does not have OIDs, or in the case of copying a query.)
**delimiter**

The single ASCII character that separates columns within each row (line) of the file. The default is a tab character in text mode, a comma in CSV mode.

**null string**

The string that represents a null value. The default is \N (backslash-N) in text mode, and a empty value with no quotes in CSV mode. You might prefer an empty string even in text mode for cases where you don’t want to distinguish nulls from empty strings. When using COPY FROM, any data item that matches this string will be stored as a null value, so you should make sure that you use the same string as you used with COPY TO.

**escape**

Specifies the single character that is used for C escape sequences (such as \n, \t, \100, and so on) and for quoting data characters that might otherwise be taken as row or column delimiters. Make sure to choose an escape character that is not used anywhere in your actual column data. The default escape character is \ (backslash) for text files or " (double quote) for CSV files, however it is possible to specify any other character to represent an escape. It is also possible to disable escaping on text-formatted files by specifying the value 'OFF' as the escape value. This is very useful for data such as web log data that has many embedded backslashes that are not intended to be escapes.

**NEWLINE**

Specifies the newline used in your data files — LF (Line feed, 0x0A), CR (Carriage return, 0x0D), or CRLF (Carriage return plus line feed, 0x0D 0x0A). If not specified, a Greenplum Database segment will detect the newline type by looking at the first row of data it receives and using the first newline type encountered.

**CSV**

Selects Comma Separated Value (CSV) mode.

**HEADER**

Specifies that a file contains a header line with the names of each column in the file. On output, the first line contains the column names from the table, and on input, the first line is ignored.

**quote**

Specifies the quotation character in CSV mode. The default is double-quote.

**FORCE QUOTE**

In CSV COPY TO mode, forces quoting to be used for all non-NULL values in each specified column. NULL output is never quoted.
FORCE NOT NULL

In CSV COPY FROM mode, process each specified column as though it were quoted and hence not a NULL value. For the default null string in CSV mode (nothing between two delimiters), this causes missing values to be evaluated as zero-length strings.

FILL MISSING FIELDS

In COPY FROM more for both TEXT and CSV, specifying FILL MISSING FIELDS will set missing trailing field values to NULL (instead of reporting an error) when a row of data has missing data fields at the end of a line or row. Blank rows, fields with a NOT NULL constraint, and trailing delimiters on a line will still report an error.

LOG ERRORS INTO error_table [KEEP]

This is an optional clause that may precede a SEGMENT REJECT LIMIT clause. It specifies an error table where rows with formatting errors will be logged when running in single row error isolation mode. You can then examine this error table to see error rows that were not loaded (if any). If the error_table specified already exists, it will be used. If it does not exist, it will be automatically generated. If the command auto-generates the error table and no errors are produced, the default is to drop the error table after the operation completes unless KEEP is specified. If the table is auto-generated and the error limit is exceeded, the entire transaction is rolled back and no error data is saved. If you want the error table to persist in this case, create the error table prior to running the COPY. An error table is defined as follows:

```
CREATE TABLE error_table_name ( cmdtime timestamptz, relname text, filename text, linenum int, bytenum int, errmsg text, rawdata text, rawbytes bytea ) DISTRIBUTED RANDOMLY;
```

SEGMENT REJECT LIMIT count [ROWS | PERCENT]

Runs a COPY FROM operation in single row error isolation mode. If the input rows have format errors they will be discarded provided that the reject limit count is not reached on any Greenplum segment instance during the load operation. The reject limit count can be specified as number of rows (the default) or percentage of total rows (1-100). If PERCENT is used, the percentage of rows per segment is calculated based on the parameter gp_reject_percent_threshold (default is 300 rows). Constraint errors such as violation of a NOT NULL, CHECK, or UNIQUE constraint will still be handled in ‘all-or-nothing’ input mode. If the limit is not reached, all good rows will be loaded and any error rows discarded.

Notes

COPY can only be used with tables, not with views. However, you can write COPY (SELECT * FROM viewname) TO ....

The BINARY key word causes all data to be stored/read as binary format rather than as text. It is somewhat faster than the normal text mode, but a binary-format file is less portable across machine architectures and Greenplum Database versions. Also, you cannot run COPY FROM in single row error isolation mode if the data is in binary format.
You must have `SELECT` privilege on the table whose values are read by `COPY TO`, and insert privilege on the table into which values are inserted by `COPY FROM`.

Files named in a `COPY` command are read or written directly by the database server, not by the client application. Therefore, they must reside on or be accessible to the Greenplum Database master host machine, not the client. They must be accessible to and readable or writable by the Greenplum Database administrative user (the user ID the server runs as), not the client. `COPY` naming a file is only allowed to database superusers, since it allows reading or writing any file that the server has privileges to access.

`COPY FROM` will invoke any triggers and check constraints on the destination table. However, it will not invoke rewrite rules. Note that in this release, violations of constraints are not evaluated for single row error isolation mode.

`COPY` input and output is affected by `DateStyle`. To ensure portability to other Greenplum Database installations that might use non-default `DateStyle` settings, `DateStyle` should be set to ISO before using `COPY TO`.

By default, `COPY` stops operation at the first error. This should not lead to problems in the event of a `COPY TO`, but the target table will already have received earlier rows in a `COPY FROM`. These rows will not be visible or accessible, but they still occupy disk space. This may amount to a considerable amount of wasted disk space if the failure happened well into a large `COPY FROM` operation. You may wish to invoke `VACUUM` to recover the wasted space. Another option would be to use single row error isolation mode to filter out error rows while still loading good rows.

### File Formats

#### Text Format

When `COPY` is used without the `BINARY` or `CSV` options, the data read or written is a text file with one line per table row. Columns in a row are separated by the *delimiter* character (tab by default). The column values themselves are strings generated by the output function, or acceptable to the input function, of each attribute’s data type. The specified null string is used in place of columns that are null.

`COPY FROM` will raise an error if any line of the input file contains more or fewer columns than are expected. If `OIDS` is specified, the OID is read or written as the first column, preceding the user data columns.

The data file has two reserved characters that have special meaning to `COPY`:

- The designated delimiter character (tab by default), which is used to separate fields in the data file.
- A UNIX-style line feed (`\n` or `0x0a`), which is used to designate a new row in the data file. It is strongly recommended that applications generating `COPY` data convert data line feeds to UNIX-style line feeds rather than Microsoft Windows style carriage return line feeds (`\r\n` or `0x0a 0x0d`).

If your data contains either of these characters, you must escape the character so `COPY` treats it as data and not as a field separator or new row.
By default, the escape character is a \ (backslash) for text-formatted files and a " (double quote) for csv-formatted files. If you want to use a different escape character, you can do so using the `ESCAPE AS` clause. Make sure to choose an escape character that is not used anywhere in your data file as an actual data value. You can also disable escaping in text-formatted files by using `ESCAPE 'OFF'`.

For example, suppose you have a table with three columns and you want to load the following three fields using COPY.

- percentage sign = %
- vertical bar = |
- backslash = \ 

Your designated `DELIMITER` character is | (pipe character), and your designated `ESCAPE` character is * (asterisk). The formatted row in your data file would look like this:

```
percentage sign = % | vertical bar = *| | backslash = \
```

Notice how the pipe character that is part of the data has been escaped using the asterisk character (*). Also notice that we do not need to escape the backslash since we are using an alternative escape character.

The following characters must be preceded by the escape character if they appear as part of a column value: the escape character itself, newline, carriage return, and the current delimiter character. You can specify a different escape character using the `ESCAPE AS` clause.

**CSV Format**

This format is used for importing and exporting the Comma Separated Value (CSV) file format used by many other programs, such as spreadsheets. Instead of the escaping used by Greenplum Database standard text mode, it produces and recognizes the common CSV escaping mechanism.

The values in each record are separated by the `DELIMITER` character. If the value contains the delimiter character, the `QUOTE` character, the `ESCAPE` character (which is double quote by default), the `NULL` string, a carriage return, or line feed character, then the whole value is prefixed and suffixed by the `QUOTE` character. You can also use `FORCE QUOTE` to force quotes when outputting non-`NULL` values in specific columns.

The CSV format has no standard way to distinguish a `NULL` value from an empty string. Greenplum Database `COPY` handles this by quoting. A `NULL` is output as the `NULL` string and is not quoted, while a data value matching the `NULL` string is quoted. Therefore, using the default settings, a `NULL` is written as an unquoted empty string, while an empty string is written with double quotes ("""). Reading values follows similar rules. You can use `FORCE NOT NULL` to prevent `NULL` input comparisons for specific columns.

Because backslash is not a special character in the CSV format, \. \., the end-of-data marker, could also appear as a data value. To avoid any misinterpretation, a \ . data value appearing as a lone entry on a line is automatically quoted on output, and on input, if quoted, is not interpreted as the end-of-data marker. If you are loading a file created by another application that has a single unquoted column and might have a value of \., you might need to quote that value in the input file.
Note: In CSV mode, all characters are significant. A quoted value surrounded by white space, or any characters other than \texttt{DELIMITER}, will include those characters. This can cause errors if you import data from a system that pads CSV lines with white space out to some fixed width. If such a situation arises you might need to preprocess the CSV file to remove the trailing white space, before importing the data into Greenplum Database.

Note: CSV mode will both recognize and produce CSV files with quoted values containing embedded carriage returns and line feeds. Thus the files are not strictly one line per table row like text-mode files.

Note: Many programs produce strange and occasionally perverse CSV files, so the file format is more a convention than a standard. Thus you might encounter some files that cannot be imported using this mechanism, and \texttt{COPY} might produce files that other programs cannot process.

\textbf{Binary Format}

The \texttt{BINARY} format consists of a file header, zero or more tuples containing the row data, and a file trailer. Headers and data are in network byte order.

- \textbf{File Header} — The file header consists of 15 bytes of fixed fields, followed by a variable-length header extension area. The fixed fields are:
  - \textbf{Signature} — 11-byte sequence \texttt{PGCOPY\n\3 77\r\0} — note that the zero byte is a required part of the signature. (The signature is designed to allow easy identification of files that have been munged by a non-8-bit-clean transfer. This signature will be changed by end-of-line-translation filters, dropped zero bytes, dropped high bits, or parity changes.)
  - \textbf{Flags field} — 32-bit integer bit mask to denote important aspects of the file format. Bits are numbered from 0 (LSB) to 31 (MSB). Note that this field is stored in network byte order (most significant byte first), as are all the integer fields used in the file format. Bits 16-31 are reserved to denote critical file format issues; a reader should abort if it finds an unexpected bit set in this range. Bits 0-15 are reserved to signal backwards-compatible format issues; a reader should simply ignore any unexpected bits set in this range. Currently only one flag is defined, and the rest must be zero (Bit 16: 1 if data has OIDs, 0 if not).
  - \textbf{Header extension area length} — 32-bit integer, length in bytes of remainder of header, not including self. Currently, this is zero, and the first tuple follows immediately. Future changes to the format might allow additional data to be present in the header. A reader should silently skip over any header extension data it does not know what to do with. The header extension area is envisioned to contain a sequence of self-identifying chunks. The flags field is not intended to tell readers what is in the extension area. Specific design of header extension contents is left for a later release.
  - \textbf{Tuples} — Each tuple begins with a 16-bit integer count of the number of fields in the tuple. (Presently, all tuples in a table will have the same count, but that might not always be true.) Then, repeated for each field in the tuple, there is a 32-bit length word followed by that many bytes of field data. (The length word does not include itself, and can be zero.) As a special case, -1 indicates a NULL field value. No value bytes follow in the NULL case. There is no alignment padding or any other extra data between fields.
Presently, all data values in a COPY BINARY file are assumed to be in binary format (format code one). It is anticipated that a future extension may add a header field that allows per-column format codes to be specified.

If OIDs are included in the file, the OID field immediately follows the field-count word. It is a normal field except that it's not included in the field-count. In particular it has a length word — this will allow handling of 4-byte vs. 8-byte OIDs without too much pain, and will allow OIDs to be shown as null if that ever proves desirable.

- **File Trailer** — The file trailer consists of a 16-bit integer word containing $-1$. This is easily distinguished from a tuple’s field-count word. A reader should report an error if a field-count word is neither $-1$ nor the expected number of columns. This provides an extra check against somehow getting out of sync with the data.

### Examples

**Copy a table to the client using the vertical bar (|) as the field delimiter:**

```sql
COPY country TO STDOUT WITH DELIMITER '|';
```

**Copy data from a file into the country table:**

```sql
COPY country FROM '/home/usr1/sql/country_data';
```

**Copy into a file just the countries whose names start with 'A':**

```sql
COPY (SELECT * FROM country WHERE country_name LIKE 'A%') TO '/home/usr1/sql/a_list_countries.copy';
```

**Create an error table called err_sales to use with single row error isolation mode:**

```sql
CREATE TABLE err_sales ( cmdtime timestamptz, relname text, filename text, linenum int, bytenum int, errmsg text, rawdata text, rawbytes bytea ) DISTRIBUTED RANDOMLY;
```

**Copy data from a file into the sales table using single row error isolation mode:**

```sql
COPY sales FROM '/home/usr1/sql/sales_data' LOG ERRORS INTO err_sales SEGMENT REJECT LIMIT 10 ROWS;
```

### Compatibility

There is no **COPY** statement in the SQL standard.

### See Also

- **CREATE EXTERNAL TABLE**
CREATE AGGREGATE

Defines a new aggregate function.

Synopsis

```
CREATE AGGREGATE name (input_data_type [ , ... ]) 
    ( SFUNC = sfunc, 
      STYPE = state_data_type 
      [, PREFUNC = prefunc] 
      [, FINALFUNC = ffunc] 
      [, INITCOND = initial_condition] 
      [, SORTOP = sort_operator] )
```

Description

CREATE AGGREGATE defines a new aggregate function. Some basic and commonly-used aggregate functions such as `count`, `min`, `max`, `sum`, `avg` and so on are already provided in Greenplum Database. If one defines new types or needs an aggregate function not already provided, then `CREATE AGGREGATE` can be used to provide the desired features.

An aggregate function is identified by its name and input data type(s). Two aggregates in the same schema can have the same name if they operate on different input types. The name and input data type(s) of an aggregate must also be distinct from the name and input data type(s) of every ordinary function in the same schema.

An aggregate function is made from one, two or three ordinary functions (all of which must be `IMMUTABLE` functions): a state transition function `sfunc`, an optional preliminary segment-level calculation function `prefunc`, and an optional final calculation function `ffunc`. These are used as follows:

```
sfunc( internal-state, next-data-values ) ---> next-internal-state
prefunc( internal-state, internal-state ) ---> next-internal-state
ffunc( internal-state ) ---> aggregate-value
```

Greenplum Database creates a temporary variable of data type `stype` to hold the current internal state of the aggregate. At each input row, the aggregate argument value(s) are calculated and the state transition function is invoked with the current state value and the new argument value(s) to calculate a new internal state value. After all the rows have been processed, the final function is invoked once to calculate the aggregate’s return value. If there is no final function then the ending state value is returned as-is.

An aggregate function may provide an initial condition, that is, an initial value for the internal state value. This is specified and stored in the database as a value of type text, but it must be a valid external representation of a constant of the state value data type. If it is not supplied then the state value starts out null.

If the state transition function is declared strict, then it cannot be called with null inputs. With such a transition function, aggregate execution behaves as follows. Rows with any null input values are ignored (the function is not called and the previous state
value is retained). If the initial state value is null, then at the first row with all-nonnull input values, the first argument value replaces the state value, and the transition function is invoked at subsequent rows with all-nonnull input values. This is handy for implementing aggregates like \texttt{max}. Note that this behavior is only available when \texttt{state_data_type} is the same as the first \texttt{input_data_type}. When these types are different, you must supply a nonnull initial condition or use a nonstrict transition function.

If the state transition function is not strict, then it will be called unconditionally at each input row, and must deal with null inputs and null transition values for itself. This allows the aggregate author to have full control over the aggregate’s handling of null values.

If the final function is declared strict, then it will not be called when the ending state value is null; instead a null result will be returned automatically. (Of course this is just the normal behavior of strict functions.) In any case the final function has the option of returning a null value. For example, the final function for \texttt{avg} returns null when it sees there were zero input rows.

Single argument aggregate functions, such as \texttt{min} or \texttt{max}, can sometimes be optimized by looking into an index instead of scanning every input row. If this aggregate can be so optimized, indicate it by specifying a sort operator. The basic requirement is that the aggregate must yield the first element in the sort ordering induced by the operator; in other words

\begin{verbatim}
SELECT \texttt{agg(col)} FROM \texttt{tab};
\end{verbatim}

must be equivalent to:

\begin{verbatim}
SELECT \texttt{col} FROM \texttt{tab} ORDER BY \texttt{col} USING \texttt{sortop} LIMIT 1;
\end{verbatim}

Further assumptions are that the aggregate ignores null inputs, and that it delivers a null result if and only if there were no non-null inputs. Ordinarily, a data type’s \texttt{<} operator is the proper sort operator for \texttt{MIN}, and \texttt{>} is the proper sort operator for \texttt{MAX}. Note that the optimization will never actually take effect unless the specified operator is the “less than” or “greater than” strategy member of a B-tree index operator class.

\section*{Parameters}

\begin{description}
\item[\texttt{name}] The name (optionally schema-qualified) of the aggregate function to create.
\item[\texttt{input_data_type}] An input data type on which this aggregate function operates. To create a zero-argument aggregate function, write \texttt{*} in place of the list of input data types. An example of such an aggregate is \texttt{count(*)}.
\item[\texttt{sfunc}] The name of the state transition function to be called for each input row. For an \texttt{N}-argument aggregate function, the \texttt{sfunc} must take \texttt{N+1} arguments, the first being of type \texttt{state_data_type} and the rest matching the declared input data type(s) of
the aggregate. The function must return a value of type \texttt{state_data_type}. This function takes the current state value and the current input data value(s), and returns the next state value.

\textbf{state_data_type}

The data type for the aggregate’s state value.

\textbf{prefunc}

The name of a preliminary aggregation function. This is a function of two arguments, both of type \texttt{state_data_type}. It must return a value of \texttt{state_data_type}. A preliminary function takes two transition state values and returns a new transition state value representing the combined aggregation. In Greenplum Database, if the result of the aggregate function is computed in a segmented fashion, the preliminary aggregation function is invoked on the individual internal states in order to combine them into an ending internal state.

\textbf{ffunc}

The name of the final function called to compute the aggregate’s result after all input rows have been traversed. The function must take a single argument of type \texttt{state_data_type}. The return data type of the aggregate is defined as the return type of this function. If \texttt{ffunc} is not specified, then the ending state value is used as the aggregate's result, and the return type is \texttt{state_data_type}.

\textbf{initial_condition}

The initial setting for the state value. This must be a string constant in the form accepted for the data type \texttt{state_data_type}. If not specified, the state value starts out null.

\textbf{sort_operator}

The associated sort operator for a MIN- or MAX-like aggregate. This is just an operator name (possibly schema-qualified). The operator is assumed to have the same input data types as the aggregate (which must be a single-argument aggregate).

\section*{Notes}

The ordinary functions used to define a new aggregate function must be defined first. Note that in this release of Greenplum Database, it is required that the \texttt{sfunc}, \texttt{ffunc}, and \texttt{prefunc} functions used to create the aggregate are defined as \texttt{IMMUTABLE}.

Any compiled code (shared library files) for custom functions must be placed in the same location on every host in your Greenplum Database array (master and all segments). This location must also be in the \texttt{LD_LIBRARY_PATH} so that the server can locate the files.

\section*{Examples}

Create a sum of cubes aggregate:

\begin{verbatim}
CREATE FUNCTION scube_accum(numeric, numeric) RETURNS numeric
\end{verbatim}
CREATE AGGREGATE
AS 'select $1 + $2 * $2 * $2'
LANGUAGE SQL
IMMUTABLE
RETURNS NULL ON NULL INPUT;
CREATE AGGREGATE scube(numeric) (
SFUNC = scube_accum,
STYPE = numeric,
INITCOND = 0);

To test this aggregate:
CREATE TABLE x(a INT);
INSERT INTO x VALUES (1),(2),(3);
SELECT scube(a) FROM x;

Correct answer for reference:
SELECT sum(a*a*a) FROM x;

Compatibility
CREATE AGGREGATE is a Greenplum Database language extension. The SQL standard
does not provide for user-defined aggregate functions.

See Also
ALTER AGGREGATE, DROP AGGREGATE, CREATE FUNCTION
CREATE CAST

Defines a new cast.

Synopsis

CREATE CAST (sourcetype AS targettype)
WITH FUNCTION funcname (argtypes)
[AS ASSIGNMENT | AS IMPLICIT]
CREATE CAST (sourcetype AS targettype) WITHOUT FUNCTION
[AS ASSIGNMENT | AS IMPLICIT]

Description

CREATE CAST defines a new cast. A cast specifies how to perform a conversion between two data types. For example,

SELECT CAST(42 AS text);

converts the integer constant 42 to type text by invoking a previously specified function, in this case text(int4). If no suitable cast has been defined, the conversion fails.

Two types may be binary compatible, which means that they can be converted into one another without invoking any function. This requires that corresponding values use the same internal representation. For instance, the types text and varchar are binary compatible.

By default, a cast can be invoked only by an explicit cast request, that is an explicit CAST(x AS typename) or x::typename construct.

If the cast is marked AS ASSIGNMENT then it can be invoked implicitly when assigning a value to a column of the target data type. For example, supposing that foo.f1 is a column of type text, then

INSERT INTO foo (f1) VALUES (42);

will be allowed if the cast from type integer to type text is marked AS ASSIGNMENT, otherwise not. The term assignment cast is typically used to describe this kind of cast.

If the cast is marked AS IMPLICIT then it can be invoked implicitly in any context, whether assignment or internally in an expression. The term implicit cast is typically used to describe this kind of cast. For example, since || takes text operands,

SELECT 'The time is ' || now();

will be allowed only if the cast from type timestamp to text is marked AS IMPLICIT. Otherwise, it will be necessary to write the cast explicitly, for example

SELECT 'The time is ' || CAST(now() AS text);

It is wise to be conservative about marking casts as implicit. An overabundance of implicit casting paths can cause Greenplum Database to choose surprising interpretations of commands, or to be unable to resolve commands at all because there
are multiple possible interpretations. A good rule of thumb is to make a cast implicitly
invokable only for information-preserving transformations between types in the same
general type category. For example, the cast from \texttt{int2} to \texttt{int4} can reasonably be
implicit, but the cast from \texttt{float8} to \texttt{int4} should probably be assignment-only.
Cross-type-category casts, such as \texttt{text} to \texttt{int4}, are best made explicit-only.

To be able to create a cast, you must own the source or the target data type. To create a
binary-compatible cast, you must be superuser.

### Parameters

- **sourcetype**
  
  The name of the source data type of the cast.

- **targettype**
  
  The name of the target data type of the cast.

- **funcname([argtypes])**
  
  The function used to perform the cast. The function name may be schema-qualified.
  If it is not, the function will be looked up in the schema search path. The function’s
  result data type must match the target type of the cast.

  Cast implementation functions may have one to three arguments. The first argument
type must be identical to the cast’s source type. The second argument, if present,
must be type \texttt{integer}; it receives the type modifier associated with the destination

type, or \texttt{-1} if there is none. The third argument, if present, must be type \texttt{boolean};
it receives \texttt{true} if the cast is an explicit cast, \texttt{false} otherwise. The SQL specification
demands different behaviors for explicit and implicit casts in some cases. This
argument is supplied for functions that must implement such casts. It is not
recommended that you design your own data types this way.

Ordinarily a cast must have different source and target data types. However, it is
allowed to declare a cast with identical source and target types if it has a cast
implementation function with more than one argument. This is used to represent

type-specific length coercion functions in the system catalogs. The named function
is used to coerce a value of the type to the type modifier value given by its second
argument. (Since the grammar presently permits only certain built-in data types to
have type modifiers, this feature is of no use for user-defined target types.)

When a cast has different source and target types and a function that takes more than
one argument, it represents converting from one type to another and applying a
length coercion in a single step. When no such entry is available, coercion to a type
that uses a type modifier involves two steps, one to convert between data types and a
second to apply the modifier.

### WITHOUT FUNCTION

Indicates that the source type and the target type are binary compatible, so no
function is required to perform the cast.
**AS ASSIGNMENT**
Indicates that the cast may be invoked implicitly in assignment contexts.

**AS IMPLICIT**
Indicates that the cast may be invoked implicitly in any context.

**Notes**
Note that in this release of Greenplum Database, user-defined functions used in a user-defined cast must be defined as **IMMUTABLE**. Any compiled code (shared library files) for custom functions must be placed in the same location on every host in your Greenplum Database array (master and all segments). This location must also be in the **LD_LIBRARY_PATH** so that the server can locate the files.

Remember that if you want to be able to convert types both ways you need to declare casts both ways explicitly.

It is recommended that you follow the convention of naming cast implementation functions after the target data type, as the built-in cast implementation functions are named. Many users are used to being able to cast data types using a function-style notation, that is $\text{typename}(x)$.

**Examples**
To create a cast from type `text` to type `int4` using the function `int4(text)` (This cast is already predefined in the system.):

```
CREATE CAST (text AS int4) WITH FUNCTION int4(text);
```

**Compatibility**
The `CREATE CAST` command conforms to the SQL standard, except that SQL does not make provisions for binary-compatible types or extra arguments to implementation functions. **AS IMPLICIT** is a Greenplum Database extension, too.

**See Also**
`CREATE FUNCTION`, `CREATE TYPE`, `DROP CAST`
**CREATE CONVERSION**

Defines a new encoding conversion.

**Synopsis**

```
CREATE [DEFAULT] CONVERSION name FOR source_encoding TO dest_encoding FROM funcname
```

**Description**

`CREATE CONVERSION` defines a new conversion between character set encodings. Conversion names may be used in the `convert` function to specify a particular encoding conversion. Also, conversions that are marked `DEFAULT` can be used for automatic encoding conversion between client and server. For this purpose, two conversions, from encoding A to B and from encoding B to A, must be defined.

To create a conversion, you must have `EXECUTE` privilege on the function and `CREATE` privilege on the destination schema.

**Parameters**

**DEFAULT**

Indicates that this conversion is the default for this particular source to destination encoding. There should be only one default encoding in a schema for the encoding pair.

**name**

The name of the conversion. The conversion name may be schema-qualified. If it is not, the conversion is defined in the current schema. The conversion name must be unique within a schema.

**source_encoding**

The source encoding name.

**dest_encoding**

The destination encoding name.

**funcname**

The function used to perform the conversion. The function name may be schema-qualified. If it is not, the function will be looked up in the path. The function must have the following signature:

```c
conv_proc(
    integer,  -- source encoding ID
    integer,  -- destination encoding ID
    cstring,  -- source string (null terminated C string)
    internal, -- destination (fill with a null terminated C string)
    integer   -- source string length
);```
Notes
Note that in this release of Greenplum Database, user-defined functions used in a user-defined conversion must be defined as IMMUTABLE. Any compiled code (shared library files) for custom functions must be placed in the same location on every host in your Greenplum Database array (master and all segments). This location must also be in the LD_LIBRARY_PATH so that the server can locate the files.

Examples
To create a conversion from encoding UTF8 to LATIN1 using myfunc:

```sql
CREATE CONVERSION myconv FOR 'UTF8' TO 'LATIN1' FROM myfunc;
```

Compatibility
There is no CREATE CONVERSION statement in the SQL standard.

See Also
ALTER CONVERSION, CREATE FUNCTION, DROP CONVERSION
CREATE DATABASE

Creates a new database.

Synopsis

CREATE DATABASE name [ [WITH] [OWNER [=] dbowner] [TEMPLATE [=] template] [ENCODING [=] encoding] [TABLESPACE [=] tablespace] [CONNECTION LIMIT [=] connlimit ] ]

Description

CREATE DATABASE creates a new database. To create a database, you must be a superuser or have the special CREATEDB privilege.

The creator becomes the owner of the new database by default. Superusers can create databases owned by other users by using the OWNER clause. They can even create databases owned by users with no special privileges. Non-superusers with CREATEDB privilege can only create databases owned by themselves.

By default, the new database will be created by cloning the standard system database template1. A different template can be specified by writing TEMPLATE name. In particular, by writing TEMPLATE template0, you can create a clean database containing only the standard objects predefined by Greenplum Database. This is useful if you wish to avoid copying any installation-local objects that may have been added to template1.

Parameters

name

The name of a database to create.

dbowner

The name of the database user who will own the new database, or DEFAULT to use the default owner (the user executing the command).

template

The name of the template from which to create the new database, or DEFAULT to use the default template (template1).

encoding

Character set encoding to use in the new database. Specify a string constant (such as 'SQL_ASCII'), an integer encoding number, or DEFAULT to use the default encoding. See “Character Set Support” on page 939.
**tablespace**

The name of the tablespace that will be associated with the new database, or `DEFAULT` to use the template database’s tablespace. This tablespace will be the default tablespace used for objects created in this database.

**connlimit**

**Warning:** Setting a connection limit at the database level cannot be enforced in Greenplum Database and may cause queries to fail. Leave this set at the default of `-1` (no limit). Connection limits may only be set at the system level in Greenplum Database. See “Limiting Concurrent Connections” on page 38 for more information.

**Notes**

`CREATE DATABASE` cannot be executed inside a transaction block.

When you copy a database by specifying its name as the template, no other sessions can be connected to the template database while it is being copied. New connections to the template database are locked out until `CREATE DATABASE` completes.

The `CONNECTION LIMIT` is not enforced against superusers.

**Examples**

To create a new database:

```
CREATE DATABASE gpdb;
```

To create a database `sales` owned by user `salesapp` with a default tablespace of `salespace`:

```
CREATE DATABASE sales OWNER salesapp TABLESPACE salespace;
```

To create a database `music` which supports the ISO-8859-1 character set:

```
CREATE DATABASE music ENCODING 'LATIN1';
```

**Compatibility**

There is no `CREATE DATABASE` statement in the SQL standard. Databases are equivalent to catalogs, whose creation is implementation-defined.

**See Also**

`ALTER DATABASE`, `DROP DATABASE`
CREATE DOMAIN

Defines a new domain.

Synopsis

```
CREATE DOMAIN name [AS] data_type [DEFAULT expression]
  [CONSTRAINT constraint_name]
  | NOT NULL | NULL
  | CHECK (expression) [...]
```

Description

CREATE DOMAIN creates a new domain. A domain is essentially a data type with optional constraints (restrictions on the allowed set of values). The user who defines a domain becomes its owner. The domain name must be unique among the data types and domains existing in its schema.

Domains are useful for abstracting common constraints on fields into a single location for maintenance. For example, several tables might contain email address columns, all requiring the same CHECK constraint to verify the address syntax. It is easier to define a domain rather than setting up a column constraint for each table that has an email column.

Parameters

- **name**
  The name (optionally schema-qualified) of a domain to be created.

- **data_type**
  The underlying data type of the domain. This may include array specifiers.

- **DEFAULT expression**
  Specifies a default value for columns of the domain data type. The value is any variable-free expression (but subqueries are not allowed). The data type of the default expression must match the data type of the domain. If no default value is specified, then the default value is the null value. The default expression will be used in any insert operation that does not specify a value for the column. If a default value is defined for a particular column, it overrides any default associated with the domain. In turn, the domain default overrides any default value associated with the underlying data type.

- **CONSTRAINT constraint_name**
  An optional name for a constraint. If not specified, the system generates a name.

- **NOT NULL**
  Values of this domain are not allowed to be null.
**NULL**

Values of this domain are allowed to be null. This is the default. This clause is only intended for compatibility with nonstandard SQL databases. Its use is discouraged in new applications.

**CHECK (expression)**

CHECK clauses specify integrity constraints or tests which values of the domain must satisfy. Each constraint must be an expression producing a Boolean result. It should use the key word VALUE to refer to the value being tested. Currently, CHECK expressions cannot contain subqueries nor refer to variables other than VALUE.

**Examples**

Create the `us_zip_code` data type. A regular expression test is used to verify that the value looks like a valid US zip code.

```sql
CREATE DOMAIN us_zip_code AS TEXT CHECK
( VALUE ~ '^\d{5}$' OR VALUE ~ '^\d{5}-\d{4}$' );
```

**Compatibility**

CREATE DOMAIN conforms to the SQL standard.

**See Also**

ALTER DOMAIN, DROP DOMAIN
CREATE EXTERNAL TABLE

Defines a new external table.

Synopsis

CREATE [READABLE] EXTERNAL TABLE table_name
( column_name data_type [, ...] | LIKE other_table )
LOCATION ('file://seghost[:port]/path/file' [, ...])
| ('gpfdist://filehost[:port]/file_pattern' [, ...])
| EXECUTE 'command' [ON ALL | MASTER | number_of_segments | HOST ['segment_hostname'] | SEGMENT segment_id ]
FORMAT 'TEXT'

| 'CSV'

CREATE [READABLE] EXTERNAL WEB TABLE table_name
( column_name data_type [, ...] | LIKE other_table )
LOCATION ('http://webhost[:port]/path/file' [, ...])
| EXECUTE 'command' [ON ALL | MASTER | number_of_segments | HOST ['segment_hostname'] | SEGMENT segment_id ]
FORMAT 'TEXT'

| 'CSV'
CREATE EXTERNAL TABLE

\[
\text{CREATE EXTERNAL TABLE} \quad \text{table_name} \quad \left( \begin{array}{l}
\text{column_name} \quad \text{data_type} \\
, \ldots \\
| \text{LIKE} \quad \text{other_table}
\end{array} \right) \quad \text{LOCATION} \quad \left( \begin{array}{c}
gpfdist://outphost[:port]/filename \quad [, \ldots] \\
\end{array} \right) \quad \text{FORMAT} \quad \text{'TEXT'} \quad \left( \begin{array}{l}
[D] \quad \text{[DELIMITER [AS] 'delimiter']} \\
[N] \quad \text{[NULL [AS] 'null string']} \\
[E] \quad \text{[ESCAPE [AS] 'escape']} \\
[\text{NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF'}] \\
[\text{FILL MISSING FIELDS}] \\
[\text{ENCODING 'encoding'}] \\
[\text{[LOG ERRORS INTO error_table]} \quad \text{SEGMENT REJECT LIMIT count}] \\
[\text{ROWS | PERCENT}] \\
\end{array} \right) \\
\text{[ENCODE 'encoding']}
\]

CREATE WRITABLE EXTERNAL TABLE

\[
\text{CREATE WRITABLE EXTERNAL TABLE} \quad \text{table_name} \quad \left( \begin{array}{l}
\text{column_name} \quad \text{data_type} \\
, \ldots \\
| \text{LIKE} \quad \text{other_table}
\end{array} \right) \quad \text{LOCATION} \quad \left( \begin{array}{c}
gpfdist://outphost[:port]/filename \quad [, \ldots] \\
\end{array} \right) \quad \text{FORMAT} \quad \text{'TEXT'} \quad \left( \begin{array}{l}
[D] \quad \text{[DELIMITER [AS] 'delimiter']} \\
[N] \quad \text{[NULL [AS] 'null string']} \\
[E] \quad \text{[ESCAPE [AS] 'escape']} \\
[\text{NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF'}] \\
[\text{FILL MISSING FIELDS}] \\
[\text{ENCODING 'encoding'}] \\
[\text{[LOG ERRORS INTO error_table]} \quad \text{SEGMENT REJECT LIMIT count}] \\
[\text{ROWS | PERCENT}] \\
\end{array} \right) \\
\text{[ENCODE 'encoding']}
\]

CREATE WRITABLE EXTERNAL WEB TABLE

\[
\text{CREATE WRITABLE EXTERNAL WEB TABLE} \quad \text{table_name} \quad \left( \begin{array}{l}
\text{column_name} \quad \text{data_type} \\
, \ldots \\
| \text{LIKE} \quad \text{other_table}
\end{array} \right) \quad \text{EXECUTE} \quad \text{'command'} \quad \text{[ON ALL]} \\
\text{FORMAT} \quad \text{'TEXT'} \quad \left( \begin{array}{l}
[D] \quad \text{[DELIMITER [AS] 'delimiter']} \\
[N] \quad \text{[NULL [AS] 'null string']} \\
[E] \quad \text{[ESCAPE [AS] 'escape']} \\
[\text{NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF'}] \\
[\text{FILL MISSING FIELDS}] \\
[\text{ENCODING 'write_encoding'}] \\
[\text{[LOG ERRORS INTO error_table]} \quad \text{SEGMENT REJECT LIMIT count}] \\
[\text{ROWS | PERCENT}] \\
\end{array} \right) \\
\text{[ENCODE 'write_encoding']}
\]

\text{[DISTRIBUTED BY (column, [ ... ])} \quad \text{[DISTRIBUTED RANDOMLY]}
\]

\text{CREATE EXTERNAL TABLE} or \text{CREATE EXTERNAL WEB TABLE} creates a new readable external table definition in Greenplum Database. Readable external tables are typically used for fast, parallel data loading. Once an external table is defined, its data can be queried directly (and in parallel) using SQL commands. You can, for example, select, join, or sort external table data. You can also create views for external tables. DML operations (\text{UPDATE}, \text{INSERT}, \text{DELETE}, \text{or TRUNCATE}) are not allowed on readable external tables, and you cannot create indexes on readable external tables.
CREATE WRITABLE EXTERNAL TABLE or CREATE WRITABLE EXTERNAL WEB TABLE creates a new writable external table definition in Greenplum Database. Writable external tables are typically used for unloading data from the database into a set of files or named pipes. Writable external web tables can be used to output data to an executable program. Once a writable external table is defined, data can be selected from database tables and inserted into the writable external table. Writable external tables only allow INSERT operations — SELECT, UPDATE, DELETE or TRUNCATE are not allowed.

About Readable External Tables

A readable external table allows you to access flat files as though they were a regular database table. External table data is considered static (meaning the data does not change midstream during the execution of a query). This allows the query planner to choose plans that allow for rescanning of the external table data.

You may specify multiple external data sources or URIs (uniform resource identifiers) with the LOCATION clause — up to the number of primary segment instances in your Greenplum Database array. Each URI points to an external data file or data source. These URIs do not need to exist prior to defining an external table (CREATE EXTERNAL TABLE does not validate the URIs specified). However you will get an error if they cannot be found when querying the external table.

There are two protocols that you can use to access the external table data sources. You may use one of the following protocols per CREATE EXTERNAL TABLE statement (cannot mix protocols):

- **gpfdist** — If using the gpfdist:// protocol, you must have the Greenplum file distribution program (gpfdist) running on the host where the external data files reside. This program points to a given directory on the file host and serves external data files to all Greenplum Database segments in parallel. If files are compressed using gzip or bzip2 (have a .gz or .bz2 file extension), gpfdist will uncompress the files automatically (provided that gunzip or bunzip2 is in your $PATH).

  All primary segments access the external file(s) in parallel regardless of how many URIs you specify when defining the external table. You can use multiple gpfdist data sources in a CREATE EXTERNAL TABLE statement to scale the scan performance of the external table.

  When specifying which files to get using gpfdist, you can use the wildcard character (*) or other C-style pattern matching to denote multiple files. The files specified are assumed to be relative to the directory from which gpfdist is serving files (the directory specified when you started the gpfdist program).

  gpfdist is located in $GPHOME/bin on your Greenplum Database master host. See the gpfdist reference documentation for more information on using this file distribution program with external tables.
- **file** — If using the `file://` protocol the external data file(s) must reside on a segment host in a location accessible by the Greenplum superuser (`gpadmin`). The number of URLs specified corresponds to the number of segment instances that will work in parallel to access the external table. So for example, if you have a Greenplum Database system with 8 primary segments and you specify 2 external files, only 2 of the 8 segments will access the external table in parallel at query time. The number of external files per segment host cannot exceed the number of primary segment instances on that host. For example, if your array has 4 primary segment instances per segment host, you may place 4 external files on each segment host. Also, the host name used in the URI must match the segment host name as registered in the `gp_configuration` system catalog table.

The **FORMAT** clause is used to describe how the external table files are formatted. The files can be in delimited text (`TEXT`) or comma separated values (`CSV`) format, similar to the formatting options available with the PostgreSQL `COPY` command. If the data in the file does not use the default column delimiter, escape character, null string and so on, you must specify the additional formatting options so that the data in the external file is read correctly by Greenplum Database.

**About Readable Web External Tables**

The main difference between regular external tables and web external tables is their data sources. Regular readable external tables access static flat files, whereas web external tables access dynamic data sources — either on a web server or by executing OS commands or scripts. When a query is planned using a regular external table, the external table is considered rescannable since the data is thought to be static for the course of the query. For web external tables, the data is not rescannable because there is the possibility that the data could change during the course of the query’s execution. Therefore, the query planner must choose plans that do not allow for rescanning of data when the external table is defined as a web external table.

There are two forms of defining a web external table. You may use one of the following forms per `CREATE EXTERNAL WEB TABLE` statement (you cannot mix the two forms):

- **Web URLs.** Specify the `LOCATION` of files on a web server using the `http://` protocol. The web data file(s) must reside on a web server that is accessible by the Greenplum segment hosts. The number of URLs specified corresponds to the number of segment instances that will work in parallel to access the web external table. So for example, if you have a Greenplum Database system with 8 primary segments and you specify 2 external files, only 2 of the 8 segments will access the web external table in parallel at query runtime.

- **OS Command.** Specify a shell command or script to `EXECUTE` on one or more segments, the output of which will comprise the web external table’s data at the time of access. A web external table defined with an `EXECUTE` clause will execute the given OS shell command or script on the specified segment host or hosts. By default, the command is executed by all active segment instances on all segment hosts. For example, if each segment host has four primary segment instances running, the command will be executed four times per segment host. You can optionally limit the number of segment instances that execute the web external table command.
Web external table data is comprised of the output of the command at the time the statement that queries the external table is executed. All segment instances included in the web external table definition (as specified by the ON clause) execute the command in parallel.

**Disabling EXECUTE for Web or Writable External Tables**

External tables that execute OS commands or scripts have a certain security risk associated with them. Some database administrators may decide that they do not want their Greenplum Database systems exposed to this functionality. If this is the case, you can disable the use of EXECUTE in web and writable external table definitions by setting the following server configuration parameters in your master `postgresql.conf` file:

```bash
  gp_external_enable_exec = off
```

**Using Environment Variables in EXECUTE commands**

If you use environment variables in external table EXECUTE commands (such as `$PATH`), keep in mind that the command is executed from within the database and not from a login shell. Therefore the `.bashrc` or `.profile` of the current user will not be sourced. However, you can set desired environment variables from within the EXECUTE clause of your external table definition, for example:

```sql
CREATE EXTERNAL WEB TABLE test (blah text) EXECUTE
  'BLAH=blah-blah-blah; export BLAH; echo $BLAH' FORMAT 'TEXT';
SELECT * FROM test;
```

```
blah
-------------
blah-blah-blah
blah-blah-blah
(2 rows)
```

The following additional Greenplum Database variables are also available for use in OS commands executed by a web or writable external table. These variables are set as environment variables in the shell that executes the command(s). They can be used to identify a set of requests made by an external table statement across the Greenplum Database array of hosts and segment instances.

**Table A.1  External Table EXECUTE Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>$GP_CID</code></td>
<td>Command count of the session executing the external table statement.</td>
</tr>
<tr>
<td><code>$GP_DATABASE</code></td>
<td>The database that the external table definition resides in.</td>
</tr>
<tr>
<td><code>$GP_DATE</code></td>
<td>The date the external table command was executed.</td>
</tr>
<tr>
<td><code>$GP_MASTER_HOST</code></td>
<td>The host name of the Greenplum master host from which the external table statement was dispatched.</td>
</tr>
<tr>
<td><code>$GP_MASTER_PORT</code></td>
<td>The port number of the Greenplum master instance from which the external table statement was dispatched.</td>
</tr>
<tr>
<td><code>$GP_SEG_DATADIR</code></td>
<td>The location of the data directory of the segment instance executing the external table command.</td>
</tr>
</tbody>
</table>
Readable External Tables and Single Row Error Isolation

The most common use of readable external tables is selecting data from them to load into regular database tables. This is typically done by issuing a `CREATE TABLE AS SELECT` or `INSERT INTO SELECT` command, where the `SELECT` statement queries external table data. By default, if the external table data contains an error, the entire command fails and no data is loaded into the target database table. To isolate data errors in external table data while still loading correctly formatted rows, you can define a readable external table with a `SEGMENT REJECT LIMIT` clause. The reject limit `count` can be specified as number of `ROWS` (the default) or a `PERCENT` of total rows (1-100).

When `SEGMENT REJECT LIMIT` is used, then the external data will be scanned in single row error isolation mode. This can be helpful in isolating errors when loading data from and external table using `CREATE TABLE AS SELECT` or `INSERT INTO SELECT`. Single row error isolation mode only applies to external data rows with format errors — for example, extra or missing attributes, attributes of a wrong data type, or invalid client encoding sequences. Constraint errors such as violation of a `NOT NULL`, `CHECK`, or `UNIQUE` constraint will still be handled in ‘all-or-nothing’ input mode. The user can specify the number or percentage of error rows acceptable (on a per-segment basis), after which the entire external table operation will be aborted and no rows will be processed or loaded. Note that the count or percent of error rows is per-segment, not per entire operation. If the per-segment reject limit is not reached, then all rows not containing an error will be processed. If the limit is not reached, all good rows will be processed and any error rows discarded. If you would like to keep error rows for further examination, you can optionally declare an error table using the `LOG ERRORS INTO` clause. Any rows containing a format error would then be logged to the specified error table.
**Readable External Tables and Query Planner Statistics**

Because the data sources for readable external tables are outside of the database, statistics needed by the query planner are not collected when you run the `ANALYZE` command. You can set some rough statistics for a readable external table by manually editing the system catalog table `pg_class` and specifying the number of rows and database pages (calculated as `data_size / 32k`). By default, `pg_class.reltuples` (number of rows) is set to 1000000 and `pg_class.relpages` (number of pages) is set to 1000 for all external tables when they are first defined. To change these defaults, you can update these values for your external table in `pg_class` (as the database superuser).

For example:

```
UPDATE pg_class SET reltuples=500000, relpages=150 WHERE relname='my_ext_table';
```

**About Writable External Tables**

A writable external table allows you to select rows from other database tables and output the rows to files, named pipes, or other executable programs. Writable external tables can also be used as output targets for Greenplum parallel MapReduce calculations.

There are two forms of defining a writable external table:

- A regular writable external table outputs data to files using the `gpfdist` parallel file distribution program. You must have the Greenplum file distribution program (`gpfdist`) running on a host that is accessible over the network by the Greenplum segments. This program points to a given file location on the output host and writes output data received from the Greenplum segments to this file. You can list multiple `gpfdist` URIs in your writable external table definition if you want your output data to be divided across multiple files.

- A writable external web table outputs data to an executable program. For example, you could unload data from Greenplum Database and send it to an executable that connects to another database or ETL tool and loads the data elsewhere. Web tables use the `EXECUTE` clause to specify a shell command, script, or executable program to run on the segment hosts. For writable external web tables, this command must be prepared to accept an input stream of data. See also, “Disabling EXECUTE for Web or Writable External Tables” and “Using Environment Variables in EXECUTE commands” for more information on using `EXECUTE` commands in a writable external table definition.

Because writable external tables only allow INSERT operations, users who are not the table owner or a superuser must be granted INSERT permissions on the table in order to use it. For example:

```
GRANT INSERT ON writable_ext_table TO user;
```

**Parameters**

**READABLE | WRITABLE**

Specifies the type of external table, readable being the default. Readable external tables are used for loading data into Greenplum Database. Writable external tables are used for unloading data.
WEB

Creates a readable or writable web external table definition in Greenplum Database. There are two forms of readable web external tables — those that access files via the `http://` protocol or those that access data by executing OS commands. Writable web external tables output data to an executable program that can accept an input stream of data. Web external tables are not rescannable during query execution.

**table_name**

The name of the new external table.

**column_name**

The name of a column to create in the external table definition. Unlike regular tables, external tables do not have column constraints or default values, so do not specify those.

**LIKE other_table**

The `LIKE` clause specifies a table from which the new external table automatically copies all column names, data types and Greenplum distribution policy. If the original table specifies any column constraints or default column values, those will not be copied over to the new external table definition.

**data_type**

The data type of the column.

**LOCATION ("protocol://host[:port]/path/file" [, ...])**

For readable external tables, specifies the URI of the external data source(s) to be used to populate the external table or web table. Regular readable external tables allow the `gpfdist` or `file` protocols. Web external tables allow the `http` protocol. If `port` is omitted for `http` and `gpfdist` protocols, port 8080 is assumed. If using the `gpfdist` protocol, the `path` is relative to the directory from which `gpfdist` is serving files (the directory specified when you started the `gpfdist` program). Also, `gpfdist` can use wildcards (or other C-style pattern matching) to denote multiple files in a directory. For example:

- `gpfdist://filehost:8081/*`
- `gpfdist://masterhost/my_load_file`
- `file://seghost1/dbfast1/external/myfile.txt`
- `http://intranet.mycompany.com/finance/expenses.csv`

For writable external tables, specifies the URI location of the `gpfdist` process that will collect data output from the Greenplum segments and write it to the named file. The `path` is relative to the directory from which `gpfdist` is serving files (the directory specified when you started the `gpfdist` program). If multiple `gpfdist` locations are listed, the segments sending data will be evenly divided across the available output locations. For example:

- `gpfdist://outputhost:8081/data1.out`
- `gpfdist://outputhost:8081/data2.out`
With two `gpfdist` locations listed as in the above example, half of the segments would send their output data to the `data1.out` file and the other half to the `data2.out` file.

**EXECUTE 'command' [ON ...]**

Allowed for readable web external tables or writable external tables only. For readable web external tables, specifies the OS command to be executed by the segment instances. The `command` can be a single OS command or a script. The `ON` clause is used to specify which segment instances will execute the given command.

- **ON ALL** is the default. The command will be executed by every active (primary) segment instance on all segment hosts in the Greenplum Database system. If the command executes a script, that script must reside in the same location on all of the segment hosts and be executable by the Greenplum superuser (`gpadmin`).
- **ON MASTER** runs the command on the master host only.
- **ON number** means the command will be executed by the specified number of segments. The particular segments are chosen randomly at runtime by the Greenplum Database system. If the command executes a script, that script must reside in the same location on all of the segment hosts and be executable by the Greenplum superuser (`gpadmin`).
- **HOST** means the command will be executed by one segment on each segment host (once per segment host), regardless of the number of active segment instances per host.
- **HOST segment_hostname** means the command will be executed by all active (primary) segment instances on the specified segment host.
- **SEGMENT segment_id** means the command will be executed only once by the specified segment. You can determine a segment instance’s ID by looking at the content number in the system catalog table `gp_segment_configuration`. The content ID of the Greenplum Database master is always -1.

For writable external tables, the `command` specified in the `EXECUTE` clause must be prepared to have data piped into it. Since all segments that have data to send will write their output to the specified command or program, the only available option for the `ON` clause is `ON ALL`.

**FORMAT 'TEXT | CSV' (options)**

Specifies the format of the external or web table data - either plain text (TEXT) or comma separated values (CSV) format.

**DELIMITER**

Specifies a single ASCII character that separates columns within each row (line) of data. The default is a tab character in TEXT mode, a comma in CSV mode. In TEXT mode for readable external tables, the delimiter can be set to OFF for special use cases in which unstructured data is loaded into a single-column table.
**NULL**

Specifies the string that represents a null value. The default is `\N` (backslash-N) in TEXT mode, and an empty value with no quotations in CSV mode. You might prefer an empty string even in TEXT mode for cases where you do not want to distinguish nulls from empty strings. When using external and web tables, any data item that matches this string will be considered a null value.

**ESCAPE**

Specifies the single character that is used for C escape sequences (such as `\n`, `\t`, `\100`, and so on) and for escaping data characters that might otherwise be taken as row or column delimiters. Make sure to choose an escape character that is not used anywhere in your actual column data. The default escape character is a \ (backslash) for text-formatted files and a " (double quote) for csv-formatted files, however it is possible to specify another character to represent an escape. It is also possible to disable escaping in text-formatted files by specifying the value 'OFF' as the escape value. This is very useful for data such as text-formatted web log data that has many embedded backslashes that are not intended to be escapes.

**NEWLINE**

Specifies the newline used in your data files — LF (Line feed, 0x0A), CR (Carriage return, 0x0D), or CRLF (Carriage return plus line feed, 0x0D 0x0A). If not specified, a Greenplum Database segment will detect the newline type by looking at the first row of data it receives and using the first newline type encountered.

**HEADER**

For readable external tables, specifies that the first line in the data file(s) is a header row (contains the names of the table columns) and should not be included as data for the table. If using multiple data source files, all files must have a header row.

**QUOTE**

Specifies the quotation character for CSV mode. The default is double-quote (").

**FORCE NOT NULL**

In CSV mode, processes each specified column as though it were quoted and hence not a NULL value. For the default null string in CSV mode (nothing between two delimiters), this causes missing values to be evaluated as zero-length strings.

**FORCE QUOTE**

In CSV mode for writable external tables, forces quoting to be used for all non-NULL values in each specified column. NULL output is never quoted.

**FILL MISSING FIELDS**

In both TEXT and CSV mode for readable external tables, specifying FILL MISSING FIELDS will set missing trailing field values to NULL (instead of reporting an error) when a row of data has missing data fields at the end of a line or row. Blank rows, fields with a NOT NULL constraint, and trailing delimiters on a line will still report an error.
ENCODING 'encoding'

Character set encoding to use for the external table. Specify a string constant (such as 'SQL ASCII'), an integer encoding number, or DEFAULT to use the default client encoding. See “Character Set Support” on page 939.

LOG ERRORS INTO error_table

This is an optional clause that may precede a SEGMENT REJECT LIMIT clause. It specifies an error table where rows with formatting errors will be logged when running in single row error isolation mode. You can then examine this error table to see error rows that were not loaded (if any). If the error_table specified already exists, it will be used. If it does not exist, it will be automatically generated.

SEGMENT REJECT LIMIT count [ROWS | PERCENT]

Runs a SELECT from a readable external table operation in single row error isolation mode. If the input rows have format errors they will be discarded provided that the reject limit count is not reached on any Greenplum segment instance during the load operation. The reject limit count can be specified as number of rows (the default) or percentage of total rows (1-100). If PERCENT is used, the percentage of rows per segment is calculated based on the parameter gp_reject_percent_threshold (default is 300 rows). Constraint errors such as violation of a NOT NULL, CHECK, or UNIQUE constraint will still be handled in ‘all-or-nothing’ mode, meaning a single constraint error will cause the entire external table operation to fail. If the limit is not reached, all good rows will be processed and any error rows discarded.

DISTRIBUTED BY (column, [ ... ])
DISTRIBUTED RANDOMLY

Used to declare the Greenplum Database distribution policy for a writable external table. By default, writable external tables are distributed randomly. If the source table you are exporting data from has a hash distribution policy, defining the same distribution key column(s) for the writable external table will improve unload performance by eliminating the need to move rows over the interconnect. When you issue an unload command such as INSERT INTO wex_table SELECT * FROM source_table, the rows that are unloaded can be sent directly from the segments to the output location if the two tables have the same hash distribution policy.

Notes

There is a system view named pg_max_external_files that you can use to determine how many external table files are permitted per external table. This view lists the available file slots per segment host (if using the file:// protocol). For example:

    SELECT * FROM pg_max_external_files;

If using the gpfdist:// protocol, the following server configuration parameter can be used to set the maximum number of segment instances that will go to one gpfdist file distribution program to get external table data in parallel. The default is 64 segment instances.

    gp_external_max_segs = <integer>
The `gpfdist` program serves files via the HTTP protocol. Queries of external tables that use a `LIMIT` clause will break off the HTTP connection after retrieving the rows causing an HTTP socket error. If using `LIMIT` in queries of external tables that use the `gpfdist://` or `http://` protocols, it is safe to ignore these HTTP socket errors from the `gpfdist` or web server — data is still returned to the database client as expected.

During dump/restore operations, only external and web external table definitions will be backed up and restored. The data sources will not be included.

You can use `CREATE TABLE AS` or `INSERT...SELECT` to load external and web external table data into another (non-external) database table, and the data will be loaded in parallel according to the external or web external table definition.

If an external table file or web external table data source has an error, any operation that reads from that table will fail. Similar to `COPY`, loading from external and web external tables is an all or nothing operation.

### Examples

Start the `gpfdist` file server program in the background on port 8081 serving files from directory `/var/data/staging`:

```
gpfdist -p 8081 -d /var/data/staging -l /home/gpadmin/log &
```

Create a readable external table named `ext_customer` using the `gpfdist` protocol and any text formatted files (`*.txt`) found in the `gpfdist` directory. The files are formatted with a pipe (`|`) as the column delimiter and an empty space as null. Also access the external table in single row error isolation mode:

```
CREATE EXTERNAL TABLE ext_customer
(id int, name text, sponsor text)
LOCATION ( 'gpfdist://filehost:8081/*.txt' )
FORMAT 'TEXT' ( DELIMITER '|' NULL ' ')
LOG ERRORS INTO err_customer SEGMENT REJECT LIMIT 5;
```

Create the same readable external table definition as above, but with CSV formatted files:

```
CREATE EXTERNAL TABLE ext_customer
(id int, name text, sponsor text)
LOCATION ( 'gpfdist://filehost:8081/*.csv' )
FORMAT 'CSV' ( DELIMITER ',' );
```

Create a readable external table named `ext_expenses` using the `file` protocol and several CSV formatted files that have a header row:

```
CREATE EXTERNAL TABLE ext_expenses (name text, date date, amount float4, category text, description text)
LOCATION ( 'file://seghost1/dbfast/external/expenses1.csv',
            'file://seghost1/dbfast/external/expenses2.csv',
            'file://seghost2/dbfast/external/expenses3.csv',
            'file://seghost2/dbfast/external/expenses4.csv',
            'file://seghost3/dbfast/external/expenses5.csv',
            'file://seghost3/dbfast/external/expenses6.csv',
            )
```

```
FORMAT 'CSV' ( HEADER );

Create a readable web external table that executes a script once per segment host:

CREATE EXTERNAL WEB TABLE log_output (linenum int, message text) EXECUTE '/var/load_scripts/get_log_data.sh' ON HOST FORMAT 'TEXT' (DELIMITER '|');

Create a writable external table named sales_out that uses gpfdist to write output data to a file named sales.out. The files are formatted with a pipe (|) as the column delimiter and an empty space as null.

CREATE WRITABLE EXTERNAL TABLE sales_out (LIKE sales)
    LOCATION ('gpfdist://etl1:8081/sales.out')
    FORMAT 'TEXT' ( DELIMITER '|' NULL ' ')
    DISTRIBUTED BY (txn_id);

Create a writable external web table that pipes output data received by the segments to an executable script named to_adreport_etl.sh:

CREATE WRITABLE EXTERNAL WEB TABLE campaign_out
    (LIKE campaign)
    EXECUTE '/var/unload_scripts/to_adreport_etl.sh'
    FORMAT 'TEXT' (DELIMITER '|');

Use the writable external table defined above to unload selected data:

INSERT INTO campaign_out SELECT * FROM campaign WHERE customer_id=123;

Compatibility

CREATE EXTERNAL TABLE is a Greenplum Database extension. The SQL standard makes no provisions for external tables.

See Also

CREATE TABLE AS, CREATE TABLE, COPY, SELECT INTO, INSERT
CREATE FUNCTION

Defines a new function.

Synopsis

CREATE [OR REPLACE] FUNCTION name
( [ [argmode] [argname] argtype [, ...] ] )
[RETURNS [SETOF] rettype]
{ LANGUAGE langname
| IMMUTABLE | STABLE | VOLATILE
| CALLED ON NULL INPUT | RETURNS NULL ON NULL INPUT | STRICT
| [EXTERNAL] SECURITY INVOKER | [EXTERNAL] SECURITY DEFINER
| AS 'definition'
| AS 'obj_file', 'link_symbol' } ...

Description

CREATE FUNCTION defines a new function. CREATE OR REPLACE FUNCTION will either create a new function, or replace an existing definition.

The name of the new function must not match any existing function with the same argument types in the same schema. However, functions of different argument types may share a name (overloading).

To update the definition of an existing function, use CREATE OR REPLACE FUNCTION. It is not possible to change the name or argument types of a function this way (this would actually create a new, distinct function). Also, CREATE OR REPLACE FUNCTION will not let you change the return type of an existing function. To do that, you must drop and recreate the function. If you drop and then recreate a function, you will have to drop existing objects (rules, views, triggers, and so on) that refer to the old function. Use CREATE OR REPLACE FUNCTION to change a function definition without breaking objects that refer to the function.

For more information about creating functions, see the User Defined Functions section of the PostgreSQL documentation.

Limited Use of VOLATILE and STABLE Functions

To prevent data from becoming out-of-sync across the segments in Greenplum Database, any function classified as STABLE or VOLATILE cannot be executed at the segment level if it contains SQL or modifies the database in any way. For example, functions such as random() or timeofday() are not allowed to execute on distributed data in Greenplum Database because they could potentially cause inconsistent data between the segment instances.

To ensure data consistency, VOLATILE and STABLE functions can safely be used in statements that are evaluated on and execute from the master. For example, the following statements are always executed on the master (statements without a FROM clause):

```sql
SELECT setval('myseq', 201);
SELECT foo();
```
In cases where a statement has a `FROM` clause containing a distributed table and the function used in the `FROM` clause simply returns a set of rows, execution may be allowed on the segments:

```
SELECT * FROM foo();
```

One exception to this rule are functions that return a table reference (rangeFuncs) or functions that use the refCursor data type. These types of functions cannot be used at all in Greenplum Database. However, they are not very commonly used anyways.

### Parameters

**name**

The name (optionally schema-qualified) of the function to create.

**argmode**

The mode of an argument: either `IN`, `OUT`, or `INOUT`. If omitted, the default is `IN`.

**argname**

The name of an argument. Some languages (currently only PL/pgSQL) let you use the name in the function body. For other languages the name of an input argument is just extra documentation. But the name of an output argument is significant, since it defines the column name in the result row type. (If you omit the name for an output argument, the system will choose a default column name.)

**argtype**

The data type(s) of the function's arguments (optionally schema-qualified), if any. The argument types may be base, composite, or domain types, or may reference the type of a table column.

Depending on the implementation language it may also be allowed to specify pseudotypes such as `cstring`. Pseudotypes indicate that the actual argument type is either incompletely specified, or outside the set of ordinary SQL data types.

The type of a column is referenced by writing `tablename.columnname%TYPE`. Using this feature can sometimes help make a function independent of changes to the definition of a table.

**rettype**

The return data type (optionally schema-qualified). The return type may be a base, composite, or domain type, or may reference the type of a table column. Depending on the implementation language it may also be allowed to specify pseudotypes such as `cstring`. If the function is not supposed to return a value, specify `void` as the return type.

When there are `OUT` or `INOUT` parameters, the `RETURNS` clause may be omitted. If present, it must agree with the result type implied by the output parameters: `RECORD` if there are multiple output parameters, or the same type as the single output parameter.
The \texttt{SETOF} modifier indicates that the function will return a set of items, rather than a single item.

The type of a column is referenced by writing \texttt{tablename.columnname\%TYPE}.

\textit{langname}

The name of the language that the function is implemented in. May be \texttt{SQL}, \texttt{C}, \texttt{internal}, or the name of a user-defined procedural language. See \texttt{CREATE LANGUAGE} for the procedural languages supported in Greenplum Database. For backward compatibility, the name may be enclosed by single quotes.

\textbf{IMMUTABLE} \textbf{STABLE} \textbf{VOLATILE}

These attributes inform the query optimizer about the behavior of the function. At most one choice may be specified. If none of these appear, \texttt{VOLATILE} is the default assumption. Since Greenplum Database currently has limited use of \texttt{VOLATILE} functions, if a function is truly \texttt{IMMUTABLE}, you must declare it as so to be able to use it without restrictions.

\texttt{IMMUTABLE} indicates that the function cannot modify the database and always returns the same result when given the same argument values. It does not do database lookups or otherwise use information not directly present in its argument list. If this option is given, any call of the function with all-constant arguments can be immediately replaced with the function value.

\texttt{STABLE} indicates that the function cannot modify the database, and that within a single table scan it will consistently return the same result for the same argument values, but that its result could change across SQL statements. This is the appropriate selection for functions whose results depend on database lookups, parameter values (such as the current time zone), and so on. Also note that the \texttt{current_timestamp} family of functions qualify as stable, since their values do not change within a transaction.

\texttt{VOLATILE} indicates that the function value can change even within a single table scan, so no optimizations can be made. Relatively few database functions are volatile in this sense; some examples are \texttt{random()}, \texttt{currval()}, \texttt{timeofday()}. But note that any function that has side-effects must be classified volatile, even if its result is quite predictable, to prevent calls from being optimized away; an example is \texttt{setval()}.

\textbf{CALLED ON NULL INPUT} \textbf{RETURNS NULL ON NULL INPUT} \textbf{STRICT}

\texttt{CALLED ON NULL INPUT} (the default) indicates that the function will be called normally when some of its arguments are null. It is then the function author’s responsibility to check for null values if necessary and respond appropriately. \texttt{RETURNS NULL ON NULL INPUT} or \texttt{STRICT} indicates that the function always returns null whenever any of its arguments are null. If this parameter is specified, the function is not executed when there are null arguments; instead a null result is assumed automatically.
[EXTERNAL] SECURITY INVOKER

SECURITY INVOKER (the default) indicates that the function is to be executed with the privileges of the user that calls it. SECURITY DEFINER specifies that the function is to be executed with the privileges of the user that created it. The key word EXTERNAL is allowed for SQL conformance, but it is optional since, unlike in SQL, this feature applies to all functions not just external ones.

definition

A string constant defining the function; the meaning depends on the language. It may be an internal function name, the path to an object file, an SQL command, or text in a procedural language.

obj_file, link_symbol

This form of the AS clause is used for dynamically loadable C language functions when the function name in the C language source code is not the same as the name of the SQL function. The string obj_file is the name of the file containing the dynamically loadable object, and link_symbol is the name of the function in the C language source code. If the link symbol is omitted, it is assumed to be the same as the name of the SQL function being defined. It is recommended to locate shared libraries either relative to $libdir (which is located at $GPHOME/lib) or through the dynamic library path (set by the dynamic_library_path server configuration parameter). This simplifies version upgrades if the new installation is at a different location.

Notes

Any compiled code (shared library files) for custom functions must be placed in the same location on every host in your Greenplum Database array (master and all segments). This location must also be in the LD_LIBRARY_PATH so that the server can locate the files. It is recommended to locate shared libraries either relative to $libdir (which is located at $GPHOME/lib) or through the dynamic library path (set by the dynamic_library_path server configuration parameter) on all master segment instances in the Greenplum array.

The full SQL type syntax is allowed for input arguments and return value. However, some details of the type specification (such as the precision field for type numeric) are the responsibility of the underlying function implementation and are not recognized or enforced by the CREATE FUNCTION command.

Greenplum Database allows function overloading. The same name can be used for several different functions so long as they have distinct argument types. However, the C names of all functions must be different, so you must give overloaded C functions different C names (for example, use the argument types as part of the C names).

Two functions are considered the same if they have the same names and input argument types, ignoring any OUT parameters. Thus for example these declarations conflict:

```
CREATE FUNCTION foo(int) ...
CREATE FUNCTION foo(int, out text) ...
```
When repeated `CREATE FUNCTION` calls refer to the same object file, the file is only loaded once. To unload and reload the file, use the `LOAD` command.

To be able to define a function, the user must have the `USAGE` privilege on the language.

It is often helpful to use dollar quoting to write the function definition string, rather than the normal single quote syntax. Without dollar quoting, any single quotes or backslashes in the function definition must be escaped by doubling them. A dollar-quoted string constant consists of a dollar sign ($), an optional tag of zero or more characters, another dollar sign, an arbitrary sequence of characters that makes up the string content, a dollar sign, the same tag that began this dollar quote, and a dollar sign. Inside the dollar-quoted string, single quotes, backslashes, or any character can be used without escaping. The string content is always written literally. For example, here are two different ways to specify the string “Dianne’s horse” using dollar quoting:

```
$$Dianne's horse$$
$$SomeTag$Dianne's horse$SomeTag$$
```

### Examples

A very simple addition function:

```sql
CREATE FUNCTION add(integer, integer) RETURNS integer
AS 'select $1 + $2;'
LANGUAGE SQL
IMMUTABLE
RETURNS NULL ON NULL INPUT;
```

Increment an integer, making use of an argument name, in PL/pgSQL:

```sql
CREATE OR REPLACE FUNCTION increment(i integer) RETURNS integer AS $$
BEGIN
RETURN i + 1;
END;
$$ LANGUAGE plpgsql;
```

Return a record containing multiple output parameters:

```sql
CREATE FUNCTION dup(in int, out f1 int, out f2 text)
AS $$ SELECT $1, CAST($1 AS text) || ' is text' $$
LANGUAGE SQL;
SELECT * FROM dup(42);
```

You can do the same thing more verbosely with an explicitly named composite type:

```sql
CREATE TYPE dup_result AS (f1 int, f2 text);
CREATE FUNCTION dup(int) RETURNS dup_result
AS $$ SELECT $1, CAST($1 AS text) || ' is text' $$
LANGUAGE SQL;
```

SELECT * FROM dup(42);

**Compatibility**

CREATE FUNCTION is defined in SQL:1999 and later. The Greenplum Database version is similar but not fully compatible. The attributes are not portable, neither are the different available languages.

For compatibility with some other database systems, `argmode` can be written either before or after `argname`. But only the first way is standard-compliant.

**See Also**

ALTER FUNCTION, DROP FUNCTION, LOAD
CREATE GROUP

Defines a new database role.

**Synopsis**

CREATE GROUP *name* [ [WITH] *option* [ ... ] ]

where *option* can be:

- SUPERUSER | NOSUPERUSER
- CREATEDB | NOCREATEDB
- CREATEROLE | NOCREATEROLE
- CREATEUSER | NOCREATEUSER
- INHERIT | NOINHERIT
- LOGIN | NOLOGIN
- [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
- VALID UNTIL 'timestamp'
- IN ROLE *rolename* [, ...]
- IN GROUP *rolename* [, ...]
- ROLE *rolename* [, ...]
- ADMIN *rolename* [, ...]
- USER *rolename* [, ...]
- SYSID *uid"

**Description**

As of Greenplum Database release 2.2, `CREATE GROUP` has been replaced by `CREATE ROLE`, although it is still accepted for backwards compatibility.

**Compatibility**

There is no `CREATE GROUP` statement in the SQL standard.

**See Also**

`CREATE ROLE`
CREATE INDEX

Defines a new index.

Synopsis

CREATE [UNIQUE] INDEX name ON table
  [USING btree|bitmap|gist]
  ( { column | ( expression ) } [ opclass ] [, ... ] )
  [ WITH ( FILLFACTOR = value ) ]
  [ TABLESPACE tablespace ]
  [ WHERE predicate ]

Description

CREATE INDEX constructs an index on the specified table. Indexes are primarily used to enhance database performance (though inappropriate use can result in slower performance).

The key field(s) for the index are specified as column names, or alternatively as expressions written in parentheses. Multiple fields can be specified if the index method supports multicolumn indexes.

An index field can be an expression computed from the values of one or more columns of the table row. This feature can be used to obtain fast access to data based on some transformation of the basic data. For example, an index computed on upper(col) would allow the clause WHERE upper(col) = 'JIM' to use an index.

Greenplum Database provides the index methods B-tree, bitmap, and GiST. Users can also define their own index methods, but that is fairly complicated.

When the WHERE clause is present, a partial index is created. A partial index is an index that contains entries for only a portion of a table, usually a portion that is more useful for indexing than the rest of the table. For example, if you have a table that contains both billed and unbilled orders where the unbilled orders take up a small fraction of the total table and yet is most often selected, you can improve performance by creating an index on just that portion.

The expression used in the WHERE clause may refer only to columns of the underlying table, but it can use all columns, not just the ones being indexed. Subqueries and aggregate expressions are also forbidden in WHERE. The same restrictions apply to index fields that are expressions.

All functions and operators used in an index definition must be immutable. Their results must depend only on their arguments and never on any outside influence (such as the contents of another table or a parameter value). This restriction ensures that the behavior of the index is well-defined. To use a user-defined function in an index expression or WHERE clause, remember to mark the function IMMUTABLE when you create it.
Parameters

**UNIQUE**
Checks for duplicate values in the table when the index is created and each time data is added. Duplicate entries will generate an error. Unique indexes only apply to B-tree indexes. In Greenplum Database, unique indexes are allowed only if the columns of the index key are the same as (or a superset of) the Greenplum distribution key. On partitioned tables, a unique index is only supported within an individual partition - not across all partitions.

**name**
The name of the index to be created. The index is always created in the same schema as its parent table.

**table**
The name (optionally schema-qualified) of the table to be indexed.

**btree | bitmap | gist**
The name of the index method to be used. Choices are `btree`, `bitmap`, and `gist`. The default method is `btree`.

**column**
The name of a column of the table on which to create the index. Only the B-tree, bitmap, and GiST index methods support multicolumn indexes.

**expression**
An expression based on one or more columns of the table. The expression usually must be written with surrounding parentheses, as shown in the syntax. However, the parentheses may be omitted if the expression has the form of a function call.

**opclass**
The name of an operator class. The operator class identifies the operators to be used by the index for that column. For example, a B-tree index on four-byte integers would use the `int4_ops` class (this operator class includes comparison functions for four-byte integers). In practice the default operator class for the column’s data type is usually sufficient. The main point of having operator classes is that for some data types, there could be more than one meaningful ordering. For example, a complex-number data type could be sorted by either absolute value or by real part. We could do this by defining two operator classes for the data type and then selecting the proper class when making an index.

**FILLFACTOR**
The fillfactor for an index is a percentage that determines how full the index method will try to pack index pages. For B-trees, leaf pages are filled to this percentage during initial index build, and also when extending the index at the right (largest key values). If pages subsequently become completely full, they will be split, leading to gradual degradation in the index’s efficiency.
B-trees use a default fillfactor of 90, but any value from 10 to 100 can be selected. If the table is static then fillfactor 100 is best to minimize the index’s physical size, but for heavily updated tables a smaller fillfactor is better to minimize the need for page splits. The other index methods use fillfactor in different but roughly analogous ways; the default fillfactor varies between methods.

**tablespace**

The tablespace in which to create the index. If not specified, the default tablespace is used.

**predicate**

The constraint expression for a partial index.

**Notes**

**UNIQUE** indexes are allowed only if the index columns are the same as (or a superset of) the Greenplum distribution key columns. On partitioned tables, a unique index is only supported within an individual partition - not across all partitions.

**UNIQUE** indexes are not allowed on append-only tables.

Indexes are not used for **IS NULL** clauses by default. The best way to use indexes in such cases is to create a partial index using an **IS NULL** predicate.

Prior releases of Greenplum Database also had an R-tree index method. This method has been removed because it had no significant advantages over the GiST method. If **USING rtree** is specified, **CREATE INDEX** will interpret it as **USING gist**.

For more information on the GiST index type, refer to the PostgreSQL documentation. The use of hash and GIN indexes has been disabled in Greenplum Database.

**Examples**

To create a B-tree index on the column **title** in the table **films**:

```
CREATE UNIQUE INDEX title_idx ON films (title);
```

To create a bitmap index on the column **gender** in the table **employee**:

```
CREATE INDEX gender_bmp_idx ON employee USING bitmap (gender);
```

To create an index on the expression **lower(title)**, allowing efficient case-insensitive searches:

```
CREATE INDEX lower_title_idx ON films ((lower(title)));
```

To create an index with non-default fill factor:

```
CREATE UNIQUE INDEX title_idx ON films (title) WITH (fillfactor = 70);
```

To create an index on the column **code** in the table **films** and have the index reside in the tablespace **indexespace**:

```
CREATE INDEX code_idx ON films(code) TABLESPACE indexespace;
```
**Compatibility**

`CREATE INDEX` is a Greenplum Database language extension. There are no provisions for indexes in the SQL standard.

Greenplum Database does not support the concurrent creation of indexes (`CONCURRENTLY` keyword not supported).

**See Also**

`ALTER INDEX`, `DROP INDEX`, `CREATE TABLE`, `CREATE OPERATOR CLASS`
CREATE LANGUAGE

Defines a new procedural language.

**Synopsis**

CREATE [PROCEDURAL] LANGUAGE name
CREATE [TRUSTED] [PROCEDURAL] LANGUAGE name
  HANDLER call_handler [VALIDATOR valfunction]

**Description**

CREATE LANGUAGE registers a new procedural language with a Greenplum database. Subsequently, functions and trigger procedures can be defined in this new language. You must be a superuser to register a new language. The PL/pgSQL language is already registered in all databases by default.

CREATE LANGUAGE effectively associates the language name with a call handler that is responsible for executing functions written in that language. For a function written in a procedural language (a language other than C or SQL), the database server has no built-in knowledge about how to interpret the function’s source code. The task is passed to a special handler that knows the details of the language. The handler could either do all the work of parsing, syntax analysis, execution, and so on or it could serve as a bridge between Greenplum Database and an existing implementation of a programming language. The handler itself is a C language function compiled into a shared object and loaded on demand, just like any other C function. There are currently four procedural language packages included in the standard Greenplum Database distribution: PL/pgSQL, PL/Perl, PL/Python, and PL/Java. A language handler has also been added for PL/R, but the PL/R language package is not pre-installed with Greenplum Database. See the section on Procedural Languages in the PostgreSQL documentation for more information on developing functions using these procedural languages.

Users who wish to use the PL/Perl procedural language must make sure that the systems that run Greenplum Database (master and all segments) have a shared version of Perl installed. For 64-bit systems, you will need a 64-bit shared version of Perl. Solaris does not have a 64-bit shared version of Perl by default. Most Linux distributions typically have a shared Perl and Python preinstalled in `/usr/lib64` (or `/usr/lib` on 32-bit systems). Greenplum provides a 64-bit shared version of Perl for both Solaris and Linux platforms. If you need a 64-bit shared Perl install package, download it from the Greenplum Network Download page.

Users who wish to use the PL/Java procedural language must make sure that the systems that run Greenplum Database (master and all segments) have a JDK version 1.6 or higher installed. Any Java archive (jar) files you are using must be added to `$GPHOME/lib/postgresql/java` and also listed in the `pljava_classpath` server configuration parameter. See the PLJAVA_README file (located in `$GPHOME/share/postgresql/pljava`) for more information on using PL/Java in Greenplum Database.
Users who wish to use the PL/R procedural language must make sure that the systems that run Greenplum Database (master and all segments) have the R language installed and the PL/R package library (plr.so) added to their Greenplum installation on all hosts. Greenplum provides compiled packages for R and PL/R that you can install. Contact Greenplum Customer Support if you need the install packages for R or PL/R.

There are two forms of the CREATE LANGUAGE command. In the first form, the user supplies just the name of the desired language, and the Greenplum Database server consults the pg_pltemplate system catalog to determine the correct parameters. In the second form, the user supplies the language parameters along with the language name. The second form can be used to create a language that is not defined in pg_pltemplate.

When the server finds an entry in the pg_pltemplate catalog for the given language name, it will use the catalog data even if the command includes language parameters. This behavior simplifies loading of old dump files, which are likely to contain out-of-date information about language support functions.

### Parameters

**TRUSTED**

Ignored if the server has an entry for the specified language name in pg_pltemplate. Specifies that the call handler for the language is safe and does not offer an unprivileged user any functionality to bypass access restrictions. If this key word is omitted when registering the language, only users with the superuser privilege can use this language to create new functions.

**PROCEDURAL**

This is a noise word.

**name**

The name of the new procedural language. The language name is case insensitive. The name must be unique among the languages in the database. Built-in support is included for plpgsql, plperl, plpython, plpythonu, and plr.plpgsql is already installed by default in Greenplum Database.

**HANDLER call_handler**

Ignored if the server has an entry for the specified language name in pg_pltemplate. The name of a previously registered function that will be called to execute the procedural language functions. The call handler for a procedural language must be written in a compiled language such as C with version 1 call convention and registered with Greenplum Database as a function taking no arguments and returning the language_handler type, a placeholder type that is simply used to identify the function as a call handler.

**VALIDATOR valfunction**

Ignored if the server has an entry for the specified language name in pg_pltemplate. valfunction is the name of a previously registered function that will be called when a new function in the language is created, to validate the new function. If no
validator function is specified, then a new function will not be checked when it is created. The validator function must take one argument of type oid, which will be the OID of the to-be-created function, and will typically return void.

A validator function would typically inspect the function body for syntactical correctness, but it can also look at other properties of the function, for example if the language cannot handle certain argument types. To signal an error, the validator function should use the ereport() function. The return value of the function is ignored.

Notes
The PL/pgSQL language is installed by default in Greenplum Database.

The system catalog pg_language records information about the currently installed languages.

To create functions in a procedural language, a user must have the USAGE privilege for the language. By default, USAGE is granted to PUBLIC (everyone) for trusted languages. This may be revoked if desired.

Procedural languages are local to individual databases. However, a language can be installed into the template1 database, which will cause it to be available automatically in all subsequently-created databases.

The call handler function and the validator function (if any) must already exist if the server does not have an entry for the language in pg_pltemplate. But when there is an entry, the functions need not already exist; they will be automatically defined if not present in the database.

Any shared library that implements a language must be located in the same LD_LIBRARY_PATH location on all segment hosts in your Greenplum Database array.

Examples
The preferred way of creating any of the standard procedural languages:

```
CREATE LANGUAGE plpgsql;
CREATE LANGUAGE plr;
```

For a language not known in the pg_pltemplate catalog:

```
CREATE FUNCTION plsample_call_handler() RETURNS language_handler
language_handler
    AS '$libdir/plsample'
    LANGUAGE C;
CREATE LANGUAGE plsample
    HANDLER plsample_call_handler;
```

Compatibility
CREATE LANGUAGE is a Greenplum Database extension.
See Also

ALTER LANGUAGE, CREATE FUNCTION, DROP LANGUAGE
CREATE OPERATOR

Defines a new operator.

Synopsis

```
CREATE OPERATOR name (  
  PROCEDURE = funcname  
  [, LEFTARG = lefttype] [, RIGHTARG = righttype]  
  [, COMMUTATOR = com_op] [, NEGATOR = neg_op]  
  [, RESTRICT = res_proc] [, JOIN = join_proc]  
  [, HASHES] [, MERGES]  
  [, SORT1 = left_sort_op] [, SORT2 = right_sort_op]  
  [, LTCMP = less_than_op] [, GTCMP = greater_than_op]  
)
```

Description

CREATE OPERATOR defines a new operator. The user who defines an operator becomes its owner.

The operator name is a sequence of up to NAMELEN - 1 (63 by default) characters from the following list: + - * / < > = ~ ! @ # % ^ & | ` ?

There are a few restrictions on your choice of name:

- -- and /* cannot appear anywhere in an operator name, since they will be taken as the start of a comment.
- A multicharacter operator name cannot end in + or -, unless the name also contains at least one of these characters: ~ ! @ # % ^ & | ` ?

For example, @- is an allowed operator name, but *- is not. This restriction allows Greenplum Database to parse SQL-compliant commands without requiring spaces between tokens.

The operator != is mapped to <> on input, so these two names are always equivalent.

At least one of LEFTARG and RIGHTARG must be defined. For binary operators, both must be defined. For right unary operators, only LEFTARG should be defined, while for left unary operators only RIGHTARG should be defined.

The funcname procedure must have been previously defined using CREATE FUNCTION, must be IMMUTABLE, and must be defined to accept the correct number of arguments (either one or two) of the indicated types.

The other clauses specify optional operator optimization clauses. These clauses should be provided whenever appropriate to speed up queries that use the operator. But if you provide them, you must be sure that they are correct. Incorrect use of an optimization clause can result in server process crashes, subtly wrong output, or other unexpected results. You can always leave out an optimization clause if you are not sure about it.
Parameters

name
The (optionally schema-qualified) name of the operator to be defined. Two operators in the same schema can have the same name if they operate on different data types.

funcname
The function used to implement this operator (must be an IMMUTABLE function).

lefttype
The data type of the operator’s left operand, if any. This option would be omitted for a left-unary operator.

righttype
The data type of the operator’s right operand, if any. This option would be omitted for a right-unary operator.

com_op
The optional COMMUTATOR clause names an operator that is the commutator of the operator being defined. We say that operator A is the commutator of operator B if (x A y) equals (y B x) for all possible input values x, y. Notice that B is also the commutator of A. For example, operators < and > for a particular data type are usually each others commutators, and operator + is usually commutative with itself. But operator - is usually not commutative with anything. The left operand type of a commutable operator is the same as the right operand type of its commutator, and vice versa. So the name of the commutator operator is all that needs to be provided in the COMMUTATOR clause.

neg_op
The optional NEGATOR clause names an operator that is the negator of the operator being defined. We say that operator A is the negator of operator B if both return Boolean results and (x A y) equals NOT (x B y) for all possible inputs x, y. Notice that B is also the negator of A. For example, < and >= are a negator pair for most data types. An operator’s negator must have the same left and/or right operand types as the operator to be defined, so only the operator name need be given in the NEGATOR clause.

res_proc
The optional RESTRICT names a restriction selectivity estimation function for the operator. Note that this is a function name, not an operator name. RESTRICT clauses only make sense for binary operators that return boolean. The idea behind a restriction selectivity estimator is to guess what fraction of the rows in a table will satisfy a WHERE-clause condition of the form:

column OP constant
for the current operator and a particular constant value. This assists the optimizer by giving it some idea of how many rows will be eliminated by \texttt{WHERE} clauses that have this form.

You can usually just use one of the following system standard estimator functions for many of your own operators:

\begin{verbatim}
   eqsel for =
   neqsel for <>
   scalarltssel for < or <=
   scalargtssel for > or >=
\end{verbatim}

\textbf{join\_proc}

The optional \texttt{JOIN} clause names a join selectivity estimation function for the operator. Note that this is a function name, not an operator name. \texttt{JOIN} clauses only make sense for binary operators that return \texttt{boolean}. The idea behind a join selectivity estimator is to guess what fraction of the rows in a pair of tables will satisfy a \texttt{WHERE}-clause condition of the form

\begin{verbatim}
table1.column1 OP table2.column2
\end{verbatim}

for the current operator. This helps the optimizer by letting it figure out which of several possible join sequences is likely to take the least work.

You can usually just use one of the following system standard join selectivity estimator functions for many of your own operators:

\begin{verbatim}
   eqjoinsel for =
   neqjoinsel for <>
   scalarltjoinsel for < or <=
   scalargtjoinsel for > or >=
   areajoinsel for 2D area-based comparisons
   positionjoinsel for 2D position-based comparisons
   contjoinsel for 2D containment-based comparisons
\end{verbatim}

\textbf{HASHES}

The optional \texttt{HASHES} clause tells the system that it is permissible to use the hash join method for a join based on this operator. \texttt{HASHES} only makes sense for a binary operator that returns \texttt{boolean}. The hash join operator can only return true for pairs of left and right values that hash to the same hash code. If two values get put in different hash buckets, the join will never compare them at all, implicitly assuming that the result of the join operator must be false. So it never makes sense to specify \texttt{HASHES} for operators that do not represent equality.

To be marked \texttt{HASHES}, the join operator must appear in a hash index operator class. Attempts to use the operator in hash joins will fail at run time if no such operator class exists. The system needs the operator class to find the data-type-specific hash function for the operator’s input data type. You must also supply a suitable hash
function before you can create the operator class. Care should be exercised when preparing a hash function, because there are machine-dependent ways in which it might fail to do the right thing.

**MERGES**

The `MERGES` clause, if present, tells the system that it is permissible to use the merge-join method for a join based on this operator. `MERGES` only makes sense for a binary operator that returns `boolean`, and in practice the operator must represent equality for some data type or pair of data types.

Merge join is based on the idea of sorting the left- and right-hand tables into order and then scanning them in parallel. So, both data types must be capable of being fully ordered, and the join operator must be one that can only succeed for pairs of values that fall at the same place in the sort order. In practice this means that the join operator must behave like equality. It is possible to merge-join two distinct data types so long as they are logically compatible. For example, the smallint-versus-integer equality operator is merge-joinable. We only need sorting operators that will bring both data types into a logically compatible sequence.

Execution of a merge join requires that the system be able to identify four operators related to the merge-join equality operator: less-than comparison for the left operand data type, less-than comparison for the right operand data type, less-than comparison between the two data types, and greater-than comparison between the two data types. It is possible to specify these operators individually by name, as the `SORT1`, `SORT2`, `LTCMP`, and `GTCMP` options respectively. The system will fill in the default names if any of these are omitted when `MERGES` is specified.

**left_sort_op**

If this operator can support a merge join, the less-than operator that sorts the left-hand data type of this operator. `<` is the default if not specified.

**right_sort_op**

If this operator can support a merge join, the less-than operator that sorts the right-hand data type of this operator. `<` is the default if not specified.

**less_than_op**

If this operator can support a merge join, the less-than operator that compares the input data types of this operator. `<` is the default if not specified.

**greater_than_op**

If this operator can support a merge join, the greater-than operator that compares the input data types of this operator. `>` is the default if not specified.

To give a schema-qualified operator name in optional arguments, use the `OPERATOR()` syntax, for example:

```
COMMUTATOR = OPERATOR(myschema.===),
```

**Notes**

Any functions used to implement the operator must be defined as `IMMUTABLE`. 

---

CREATE OPERATOR 383
Examples
Here is an example of creating an operator for adding two complex numbers, assuming we have already created the definition of type complex. First define the function that does the work, then define the operator:

```
CREATE FUNCTION complex_add(complex, complex)
    RETURNS complex
    AS 'filename', 'complex_add'
    LANGUAGE C IMMUTABLE STRICT;

CREATE OPERATOR + (
    leftarg = complex,
    rightarg = complex,
    procedure = complex_add,
    commutator = +
);
```

To use this operator in a query:

```
SELECT (a + b) AS c FROM test_complex;
```

Compatibility
CREATE OPERATOR is a Greenplum Database language extension. The SQL standard does not provide for user-defined operators.

See Also
CREATE FUNCTION, CREATE TYPE, ALTER OPERATOR, DROP OPERATOR
CREATE OPERATOR CLASS

Defines a new operator class.

Synopsis

CREATE OPERATOR CLASS name [DEFAULT] FOR TYPE data_type
  USING index_method AS
  { OPERATOR strategy_number op_name [(op_type, op_type)] [RECHECK]
    | FUNCTION support_number funcname (argument_type [, ...] )
    | STORAGE storage_type
  } [, ... ]

Description

CREATE OPERATOR CLASS creates a new operator class. An operator class defines how a particular data type can be used with an index. The operator class specifies that certain operators will fill particular roles or strategies for this data type and this index method. The operator class also specifies the support procedures to be used by the index method when the operator class is selected for an index column. All the operators and functions used by an operator class must be defined before the operator class is created. Any functions used to implement the operator class must be defined as IMMUTABLE.

CREATE OPERATOR CLASS does not presently check whether the operator class definition includes all the operators and functions required by the index method, nor whether the operators and functions form a self-consistent set. It is the user’s responsibility to define a valid operator class.

You must be a superuser to create an operator class.

Parameters

name

The (optionally schema-qualified) name of the operator class to be defined. Two operator classes in the same schema can have the same name only if they are for different index methods.

DEFAULT

Makes the operator class the default operator class for its data type. At most one operator class can be the default for a specific data type and index method.

data_type

The column data type that this operator class is for.

index_method

The name of the index method this operator class is for. Choices are btree, bitmap, and gist.
**strategy_number**

The operators associated with an operator class are identified by *strategy numbers*, which serve to identify the semantics of each operator within the context of its operator class. For example, B-trees impose a strict ordering on keys, lesser to greater, and so operators like *less than* and *greater than or equal to* are interesting with respect to a B-tree. These strategies can be thought of as generalized operators. Each operator class specifies which actual operator corresponds to each strategy for a particular data type and interpretation of the index semantics. The corresponding strategy numbers for each index method are as follows:

**Table A.1** B-tree and Bitmap Strategies

<table>
<thead>
<tr>
<th>Operation</th>
<th>Strategy Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>less than</td>
<td>1</td>
</tr>
<tr>
<td>less than or equal</td>
<td>2</td>
</tr>
<tr>
<td>equal</td>
<td>3</td>
</tr>
<tr>
<td>greater than or equal</td>
<td>4</td>
</tr>
<tr>
<td>greater than</td>
<td>5</td>
</tr>
</tbody>
</table>

**Table A.2** GiST Two-Dimensional Strategies (R-Tree)

<table>
<thead>
<tr>
<th>Operation</th>
<th>Strategy Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>strictly left of</td>
<td>1</td>
</tr>
<tr>
<td>does not extend to right of</td>
<td>2</td>
</tr>
<tr>
<td>overlaps</td>
<td>3</td>
</tr>
<tr>
<td>does not extend to left of</td>
<td>4</td>
</tr>
<tr>
<td>strictly right of</td>
<td>5</td>
</tr>
<tr>
<td>same</td>
<td>6</td>
</tr>
<tr>
<td>contains</td>
<td>7</td>
</tr>
<tr>
<td>contained by</td>
<td>8</td>
</tr>
<tr>
<td>does not extend above</td>
<td>9</td>
</tr>
<tr>
<td>strictly below</td>
<td>10</td>
</tr>
<tr>
<td>strictly above</td>
<td>11</td>
</tr>
<tr>
<td>does not extend below</td>
<td>12</td>
</tr>
</tbody>
</table>

**operator_name**

The name (optionally schema-qualified) of an operator associated with the operator class.

**op_type**

The operand data type(s) of an operator, or *NONE* to signify a left-unary or right-unary operator. The operand data types may be omitted in the normal case where they are the same as the operator class data type.
RECHECK

If present, the index is “lossy” for this operator, and so the rows retrieved using the index must be rechecked to verify that they actually satisfy the qualification clause involving this operator.

**support_number**

Index methods require additional support routines in order to work. These operations are administrative routines used internally by the index methods. As with strategies, the operator class identifies which specific functions should play each of these roles for a given data type and semantic interpretation. The index method defines the set of functions it needs, and the operator class identifies the correct functions to use by assigning them to the *support function numbers* as follows:

### Table A.3 B-tree and Bitmap Support Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Support Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compare two keys and return an integer less than zero, zero, or greater than zero, indicating whether the first key is less than, equal to, or greater than the second.</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table A.4 GiST Support Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Support Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>consistent - determine whether key satisfies the query qualifier.</td>
<td>1</td>
</tr>
<tr>
<td>union - compute union of a set of keys.</td>
<td>2</td>
</tr>
<tr>
<td>compress - compute a compressed representation of a key or value to be indexed.</td>
<td>3</td>
</tr>
<tr>
<td>decompress - compute a decompressed representation of a compressed key.</td>
<td>4</td>
</tr>
<tr>
<td>penalty - compute penalty for inserting new key into subtree with given subtree’s key.</td>
<td>5</td>
</tr>
<tr>
<td>picksplit - determine which entries of a page are to be moved to the new page and compute the union keys for resulting pages.</td>
<td>6</td>
</tr>
<tr>
<td>equal - compare two keys and return true if they are equal.</td>
<td>7</td>
</tr>
</tbody>
</table>

**funcname**

The name (optionally schema-qualified) of a function that is an index method support procedure for the operator class.

**argument_types**

The parameter data type(s) of the function.

**storage_type**

The data type actually stored in the index. Normally this is the same as the column data type, but the GiST index method allows it to be different. The *STORAGE* clause must be omitted unless the index method allows a different type to be used.
Notes
Because the index machinery does not check access permissions on functions before using them, including a function or operator in an operator class is the same as granting public execute permission on it. This is usually not an issue for the sorts of functions that are useful in an operator class.

The operators should not be defined by SQL functions. A SQL function is likely to be inlined into the calling query, which will prevent the optimizer from recognizing that the query matches an index.

Any functions used to implement the operator class must be defined as IMMUTABLE.

Examples
The following example command defines a GiST index operator class for the data type _int4 (array of int4):

```
CREATE OPERATOR CLASS gist__int_ops
DEFAULT FOR TYPE _int4 USING gist AS
  OPERATOR 3 &&,
  OPERATOR 6 = RECHECK,
  OPERATOR 7 @>,
  OPERATOR 8 <@,
  OPERATOR 20 @@ (_int4, query_int),
  FUNCTION 1 g_int_consistent (internal, _int4, int4),
  FUNCTION 2 g_int_union (bytea, internal),
  FUNCTION 3 g_int_compress (internal),
  FUNCTION 4 g_int_decompress (internal),
  FUNCTION 5 g_int_penalty (internal, internal, internal),
  FUNCTION 6 g_int_picksplit (internal, internal),
  FUNCTION 7 g_int_same (_int4, _int4, internal);
```

Compatibility
CREATE OPERATOR CLASS is a Greenplum Database extension. There is no CREATE OPERATOR CLASS statement in the SQL standard.

See Also
ALTER OPERATOR CLASS, DROP OPERATOR CLASS, CREATE FUNCTION
CREATE RESOURCE QUEUE

Defines a new resource queue.

Synopsis

CREATE RESOURCE QUEUE name WITH (queue_attribute=value [, ... ])

where queue_attribute is:

  ACTIVE_STATEMENTS=integer 
    [ MAX_COST=float [COST_OVERCOMMIT={TRUE|FALSE}] ] 
    [ MIN_COST=float ] 
    [ PRIORITY={MIN|LOW|MEDIUM|HIGH|MAX} ] 
  | MAX_COST=float [COST_OVERCOMMIT={TRUE|FALSE}] 
    [ ACTIVE_STATEMENTS=integer ] 
    [ MIN_COST=float ] 
    [ PRIORITY={MIN|LOW|MEDIUM|HIGH|MAX} ]

Description

Creates a new resource queue for Greenplum Database workload management. A resource queue must have either an ACTIVE_STATEMENTS or a MAX_COST value (or it can have both). You can optionally assign a priority to a resource queue to control the relative share of available CPU resources used by queries associated with the queue. Only a superuser can create a resource queue.

Resource queues with an ACTIVE_STATEMENTS threshold set a maximum limit on the number of queries that can be executed by roles assigned to that queue. It controls the number of active queries that are allowed to run at the same time. The value for ACTIVE_STATEMENTS should be an integer greater than 0.

Resource queues with a MAX_COST threshold set a maximum limit on the total cost of queries that can be executed by roles assigned to that queue. Cost is measured in the estimated total cost for the query as determined by the Greenplum Database query planner (as shown in the EXPLAIN output for a query). Therefore, an administrator must be familiar with the queries typically executed on the system in order to set an appropriate cost threshold for a queue. Cost is measured in units of disk page fetches; 1.0 equals one sequential disk page read. The value for MAX_COST is specified as a floating point number (for example 100.0) or can also be specified as an exponent (for example 1e+2). If a resource queue is limited based on a cost threshold, then the administrator can allow COST_OVERCOMMIT=TRUE (the default). This means that a query that exceeds the allowed cost threshold will be allowed to run but only when the system is idle. If COST_OVERCOMMIT=FALSE is specified, queries that exceed the cost limit will always be rejected and never allowed to run. Specifying a value for MIN_COST allows the administrator to define a cost for small queries that will be exempt from resource queueing.

If a value is not defined for ACTIVE_STATEMENTS or MAX_COST, it is set to -1 by default (meaning no limit). After defining a resource queue, you must assign roles to the queue using the ALTER ROLE or CREATE ROLE command.
If a value is not defined for `PRIORITY`, queries associated with the queue have a default priority of `MEDIUM`.

**Parameters**

**name**

The name of the resource queue.

**ACTIVE_STATEMENTS integer**

Resource queues with an `ACTIVE_STATEMENTS` threshold limit the number of queries that can be executed by roles assigned to that queue. It controls the number of active queries that are allowed to run at the same time. The value for `ACTIVE_STATEMENTS` should be an integer greater than 0.

**MAX_COST float**

Resource queues with a `MAX_COST` threshold set a maximum limit on the total cost of queries that can be executed by roles assigned to that queue. Cost is measured in the estimated total cost for the query as determined by the Greenplum Database query planner (as shown in the `EXPLAIN` output for a query). Therefore, an administrator must be familiar with the queries typically executed on the system in order to set an appropriate cost threshold for a queue. Cost is measured in units of disk page fetches; 1.0 equals one sequential disk page read. The value for `MAX_COST` is specified as a floating point number (for example 100.0) or can also be specified as an exponent (for example 1e+2).

**COST_OVERCOMMIT boolean**

If a resource queue is limited based on `MAX_COST`, then the administrator can allow `COST_OVERCOMMIT` (the default). This means that a query that exceeds the allowed cost threshold will be allowed to run but only when the system is idle. If `COST_OVERCOMMIT=FALSE` is specified, queries that exceed the cost limit will always be rejected and never allowed to run.

**MIN_COST float**

The minimum query cost limit of what is considered a small query. Queries with a cost under this limit will not be queued and run immediately. Cost is measured in the estimated total cost for the query as determined by the Greenplum Database query planner (as shown in the `EXPLAIN` output for a query). Therefore, an administrator must be familiar with the queries typically executed on the system in order to set an appropriate cost for what is considered a small query. Cost is measured in units of disk page fetches; 1.0 equals one sequential disk page read. The value for `MIN_COST` is specified as a floating point number (for example 100.0) or can also be specified as an exponent (for example 1e+2).
PRIORITY={MIN|LOW|MEDIUM|HIGH|MAX}

Sets the priority of queries associated with a resource queue. Queries or statements in queues with higher priority levels will receive a larger share of available CPU resources in case of contention. Queries in low-priority queues may be delayed while higher priority queries are executed. If no priority is specified, queries associated with the queue have a priority of MEDIUM.

**Notes**

Use the `pg_resqueue_status` system view to see the limit settings and current status of a resource queue:

```
SELECT * FROM pg_resqueue_status WHERE rsqname='queue_name';
```

There is also another system view named `pg_stat_resqueues` which shows statistical metrics for a resource queue over time. To use this view, however, you must enable the `stats_queue_level` server configuration parameter. See “Managing Workload and Resources” on page 48 for more information about using resource queues.

`CREATE RESOURCE QUEUE` cannot be run within a transaction.

**Examples**

Create a resource queue with an active query limit of 20:

```
CREATE RESOURCE QUEUE myqueue WITH (ACTIVE_STATEMENTS=20);
```

Create a resource queue with a query cost limit of 3000.0:

```
CREATE RESOURCE QUEUE myqueue WITH (MAX_COST=3000.0);
```

Create a resource queue with a query cost limit of $3^{10}$ (or 30000000000.0) and do not allow overcommit. Allow small queries with a cost under 500 to run immediately:

```
CREATE RESOURCE QUEUE myqueue WITH (MAX_COST=3e+10, COST_OVERCOMMIT=FALSE, MIN_COST=500.0);
```

Create a resource queue with both an active query limit and a query cost limit:

```
CREATE RESOURCE QUEUE myqueue WITH (ACTIVE_STATEMENTS=30, MAX_COST=5000.00);
```

Create a resource queue with an active query limit of 5 and a maximum priority setting:

```
CREATE RESOURCE QUEUE myqueue WITH (ACTIVE_STATEMENTS=5, PRIORITY=MAX);
```

**Compatibility**

`CREATE RESOURCE QUEUE` is a Greenplum Database extension. There is no provision for resource queues or workload management in the SQL standard.

**See Also**

`ALTER ROLE, CREATE ROLE, ALTER RESOURCE QUEUE, DROP RESOURCE QUEUE`
CREATE ROLE

Defines a new database role (user or group).

Synopsis

CREATE ROLE name [[WITH] option [ ... ]]

where option can be:

- SUPERUSER | NOSUPERUSER
- CREATEDB | NOCREATEDB
- CREATEROLE | NOCREATEROLE
- CREATEEXTTABLE | NOCREATEEXTTABLE
- [ ( attribute='value'[, ...] ) ]
  where attributes and values are:
  - type='readable'|'writable'
  - protocol='gpfdist'|'http'
  - INHERIT | NOINHERIT
  - LOGIN | NOLOGIN
  - CONNECTION LIMIT conlimit
  - [ ENCRYPTED | UNENCrypted ] PASSWORD 'password'
  - VALID UNTIL 'timestamp'
  - IN ROLE rolename [, ...]
  - ROLE rolename [, ...]
  - ADMIN rolename [, ...]
  - RESOURCE QUEUE queue_name

Description

CREATE ROLE adds a new role to a Greenplum Database system. A role is an entity that can own database objects and have database privileges. A role can be considered a user, a group, or both depending on how it is used. You must have CREATEROLE privilege or be a database superuser to use this command.

Note that roles are defined at the system-level and are valid for all databases in your Greenplum Database system.

Parameters

name

The name of the new role.

SUPERUSER
NOSUPERUSER

If SUPERUSER is specified, the role being defined will be a superuser, who can override all access restrictions within the database. Superuser status is dangerous and should be used only when really needed. You must yourself be a superuser to create a new superuser. NOSUPERUSER is the default.
**CREATEDB**
**NOCREATEDB**
If `CREATEDB` is specified, the role being defined will be allowed to create new databases. `NOCREATEDB` (the default) will deny a role the ability to create databases.

**CREATEROLE**
**NOCREATEROLE**
If `CREATEDB` is specified, the role being defined will be allowed to create new roles, alter other roles, and drop other roles. `NOCREATEROLE` (the default) will deny a role the ability to create roles or modify roles other than their own.

**CREATEEXTTABLE**
**NOCREATEEXTTABLE**
If `CREATEEXTTABLE` is specified, the role being defined will be allowed to create external tables. The default type is `readable` and the default protocol is `gpfdist` if not specified. `NOCREATEEXTTABLE` (the default) will deny a role the ability to create external tables. Note that external tables that use the `file` or `execute` protocols can only be created by superusers.

**INHERIT**
**NOCREATEROLE**
If specified, `INHERIT` (the default) allows the role to use whatever database privileges have been granted to all roles it is directly or indirectly a member of. With `NOCREATEROLE`, membership in another role only grants the ability to set `ROLE` to that other role.

**LOGIN**
**NOLOGIN**
If specified, `LOGIN` allows a role to log in to a database. A role having the `LOGIN` attribute can be thought of as a user. Roles with `NOLOGIN` (the default) are useful for managing database privileges, and can be thought of as groups.

**CONNECTION LIMIT connlimit**
*Warning:* Setting a connection limit at the role level cannot be enforced in Greenplum Database and may cause queries to fail. Leave this set at the default of `-1` (no limit). Connection limits may only be set at the system level in Greenplum Database. See “Limiting Concurrent Connections” on page 38 for more information.

**PASSWORD password**
Sets the user password for roles with the `LOGIN` attribute. If you do not plan to use password authentication you can omit this option. If no password is specified, the password will be set to null and password authentication will always fail for that user. A null password can optionally be written explicitly as `PASSWORD NULL`.

**ENCRYPTED**
**UNENCRYPTED**
These key words control whether the password is stored encrypted in the system catalogs. (If neither is specified, the default behavior is determined by the configuration parameter `password_encryption`.) If the presented password string is
already in MD5-encrypted format, then it is stored encrypted as-is, regardless of whether ENCRYPTED or UNENCRYPTED is specified (since the system cannot decrypt the specified encrypted password string). This allows reloading of encrypted passwords during dump/restore.

Note that older clients may lack support for the MD5 authentication mechanism that is needed to work with passwords that are stored encrypted.

**VALID UNTIL 'timestamp'**

The VALID UNTIL clause sets a date and time after which the role’s password is no longer valid. If this clause is omitted the password will never expire.

**IN ROLE rolename**

Adds the new role as a member of the named roles. Note that there is no option to add the new role as an administrator; use a separate `GRANT` command to do that.

**ROLE rolename**

Adds the named roles as members of this role, making this new role a group.

**ADMIN rolename**

The `ADMIN` clause is like `ROLE`, but the named roles are added to the new role with `ADMIN OPTION`, giving them the right to grant membership in this role to others.

**RESOURCE QUEUE queue_name**

The name of the resource queue to which the new user-level role is to be assigned. Only roles with `LOGIN` privilege can be assigned to a resource queue. The special keyword `NONE` means that the role is assigned to the default resource queue. A role can only belong to one resource queue.

**Notes**

The preferred way to add and remove role members (manage groups) is to use `GRANT` and `REVOKE`.

The `VALID UNTIL` clause defines an expiration time for a password only, not for the role. The expiration time is not enforced when logging in using a non-password-based authentication method.

The `INHERIT` attribute governs inheritance of grantable privileges (access privileges for database objects and role memberships). It does not apply to the special role attributes set by `CREATE ROLE` and `ALTER ROLE`. For example, being a member of a role with `CREATEDB` privilege does not immediately grant the ability to create databases, even if `INHERIT` is set.

The `INHERIT` attribute is the default for reasons of backwards compatibility. In prior releases of Greenplum Database, users always had access to all privileges of groups they were members of. However, `NOINHERIT` provides a closer match to the semantics specified in the SQL standard.
Be careful with the `CREATEROLE` privilege. There is no concept of inheritance for the privileges of a `CREATEROLE`-role. That means that even if a role does not have a certain privilege but is allowed to create other roles, it can easily create another role with different privileges than its own (except for creating roles with superuser privileges). For example, if a role has the `CREATEROLE` privilege but not the `CREATEDB` privilege, it can create a new role with the `CREATEDB` privilege. Therefore, regard roles that have the `CREATEROLE` privilege as almost-superuser-roles.

The `CONNECTION LIMIT` option is never enforced for superusers.

Caution must be exercised when specifying an unencrypted password with this command. The password will be transmitted to the server in clear-text, and it might also be logged in the client’s command history or the server log. The client program `createuser`, however, transmits the password encrypted. Also, psql contains a command `\password` that can be used to safely change the password later.

### Examples

Create a role that can log in, but don't give it a password:

```sql
CREATE ROLE jonathan LOGIN;
```

Create a role that belongs to a resource queue:

```sql
CREATE ROLE jonathan LOGIN RESOURCE QUEUE poweruser;
```

Create a role with a password that is valid until the end of 2009 (`CREATE USER` is the same as `CREATE ROLE` except that it implies `LOGIN`):

```sql
CREATE USER joelle WITH PASSWORD 'jw8s0F4' VALID UNTIL '2010-01-01';
```

Create a role that can create databases and manage other roles:

```sql
CREATE ROLE admin WITH CREATEDB CREATEROLE;
```

### Compatibility

The SQL standard defines the concepts of users and roles, but it regards them as distinct concepts and leaves all commands defining users to be specified by the database implementation. In Greenplum Database users and roles are unified into a single type of object. Roles therefore have many more optional attributes than they do in the standard.

`CREATE ROLE` is in the SQL standard, but the standard only requires the syntax:

```sql
CREATE ROLE name [WITH ADMIN rolename]
```

Allowing multiple initial administrators, and all the other options of `CREATE ROLE`, are Greenplum Database extensions.

The behavior specified by the SQL standard is most closely approximated by giving users the `NOINHERIT` attribute, while roles are given the `INHERIT` attribute.

### See Also

`SET ROLE`, `ALTER ROLE`, `DROP ROLE`, `GRANT`, `REVOKE`, `CREATE RESOURCE QUEUE`
CREATE RULE

Defines a new rewrite rule.

Synopsis

CREATE [OR REPLACE] RULE name AS ON event
    TO table [WHERE condition]
    DO [ALSO | INSTEAD] { NOTHING | command | (command; command ... ) }

Description

CREATE RULE defines a new rule applying to a specified table or view. CREATE OR REPLACE RULE will either create a new rule, or replace an existing rule of the same name for the same table.

The Greenplum Database rule system allows one to define an alternate action to be performed on insertions, updates, or deletions in database tables. A rule causes additional or alternate commands to be executed when a given command on a given table is executed. Rules can be used on views as well. It is important to realize that a rule is really a command transformation mechanism, or command macro. The transformation happens before the execution of the commands starts. It does not operate independently for each physical row as does a trigger.

ON SELECT rules must be unconditional INSTEAD rules and must have actions that consist of a single SELECT command. Thus, an ON SELECT rule effectively turns the table into a view, whose visible contents are the rows returned by the rule’s SELECT command rather than whatever had been stored in the table (if anything). It is considered better style to write a CREATE VIEW command than to create a real table and define an ON SELECT rule for it.

You can create the illusion of an updatable view by defining ON INSERT, ON UPDATE, and ON DELETE rules to replace update actions on the view with appropriate updates on other tables. If you want to support INSERT RETURNING and so on, then be sure to put a suitable RETURNING clause into each of these rules.

Rules are also helpful for managing partitioned tables. You can define ON INSERT rules on the parent table to route inserted rows to the correct partitioned child table. Note that rules do not work for COPY commands.

There is a catch if you try to use conditional rules for view updates: there must be an unconditional INSTEAD rule for each action you wish to allow on the view. If the rule is conditional, or is not INSTEAD, then the system will still reject attempts to perform the update action, because it thinks it might end up trying to perform the action on the dummy table of the view in some cases. If you want to handle all the useful cases in conditional rules, add an unconditional DO INSTEAD NOTHING rule to ensure that the system understands it will never be called on to update the dummy table. Then make the conditional rules non-INSTEAD; in the cases where they are applied, they add to the default INSTEAD NOTHING action. (This method does not currently work to support RETURNING queries, however.)
**Parameters**

**name**

The name of a rule to create. This must be distinct from the name of any other rule for the same table. Multiple rules on the same table and same event type are applied in alphabetical name order.

**event**

The event is one of `SELECT`, `INSERT`, `UPDATE`, or `DELETE`.

**table**

The name (optionally schema-qualified) of the table or view the rule applies to.

**condition**

Any SQL conditional expression (returning boolean). The condition expression may not refer to any tables except `NEW` and `OLD`, and may not contain aggregate functions. `NEW` and `OLD` refer to values in the referenced table. `NEW` is valid in `ON INSERT` and `ON UPDATE` rules to refer to the new row being inserted or updated. `OLD` is valid in `ON UPDATE` and `ON DELETE` rules to refer to the existing row being updated or deleted.

**INSTEAD**

`INSTEAD` indicates that the commands should be executed instead of the original command.

**ALSO**

`ALSO` indicates that the commands should be executed in addition to the original command. If neither `ALSO` nor `INSTEAD` is specified, `ALSO` is the default.

**command**

The command or commands that make up the rule action. Valid commands are `SELECT`, `INSERT`, `UPDATE`, or `DELETE`. The special table names `NEW` and `OLD` may be used to refer to values in the referenced table. `NEW` is valid in `ON INSERT` and `ON UPDATE` rules to refer to the new row being inserted or updated. `OLD` is valid in `ON UPDATE` and `ON DELETE` rules to refer to the existing row being updated or deleted.

**Notes**

You must be the owner of a table to create or change rules for it.

In a rule for `INSERT`, `UPDATE`, or `DELETE` on a view, you can add a `RETURNING` clause that emits the view’s columns. This clause will be used to compute the outputs if the rule is triggered by an `INSERT RETURNING`, `UPDATE RETURNING`, or `DELETE RETURNING` command respectively. When the rule is triggered by a command without `RETURNING`, the rule’s `RETURNING` clause will be ignored. The current implementation allows only unconditional `INSTEAD` rules to contain `RETURNING`; furthermore there can be at most one `RETURNING` clause among all the rules for the same event. (This
ensures that there is only one candidate RETURNING clause to be used to compute the results.) RETURNING queries on the view will be rejected if there is no RETURNING clause in any available rule.

It is very important to take care to avoid circular rules. Recursive rules are not validated at rule create time, but will report an error at execution time.

Examples
Create a rule that inserts rows into the child table `b2001` when a user tries to insert into the partitioned parent table `rank`:

```
CREATE RULE b2001 AS ON INSERT TO rank WHERE gender='M' and year='2001' DO INSTEAD INSERT INTO b2001 VALUES (NEW.id, NEW.rank, NEW.year, NEW.gender, NEW.count);
```

Compatibility
CREATE RULE is a Greenplum Database language extension, as is the entire query rewrite system.

See Also
DROP RULE, CREATE TABLE, CREATE VIEW
CREATE SCHEMA

Defines a new schema.

Synopsis

CREATE SCHEMA schema_name [AUTHORIZATION username] [schema_element [ ... ]]
CREATE SCHEMA AUTHORIZATION rolename [schema_element [ ... ]]

Description

CREATE SCHEMA enters a new schema into the current database. The schema name must be distinct from the name of any existing schema in the current database.

A schema is essentially a namespace: it contains named objects (tables, data types, functions, and operators) whose names may duplicate those of other objects existing in other schemas. Named objects are accessed either by qualifying their names with the schema name as a prefix, or by setting a search path that includes the desired schema(s). A CREATE command specifying an unqualified object name creates the object in the current schema (the one at the front of the search path, which can be determined with the function current_schema).

Optionally, CREATE SCHEMA can include subcommands to create objects within the new schema. The subcommands are treated essentially the same as separate commands issued after creating the schema, except that if the AUTHORIZATION clause is used, all the created objects will be owned by that role.

Parameters

schema_name

The name of a schema to be created. If this is omitted, the user name is used as the schema name. The name cannot begin with pg_, as such names are reserved for system catalog schemas.

rolename

The name of the role who will own the schema. If omitted, defaults to the role executing the command. Only superusers may create schemas owned by roles other than themselves.

schema_element

An SQL statement defining an object to be created within the schema. Currently, only CREATE TABLE, CREATE VIEW, CREATE INDEX, CREATE SEQUENCE, CREATE TRIGGER and GRANT are accepted as clauses within CREATE SCHEMA. Other kinds of objects may be created in separate commands after the schema is created.
Notes
To create a schema, the invoking user must have the `CREATE` privilege for the current database or be a superuser.

Examples
Create a schema:

```sql
CREATE SCHEMA myschema;
```

Create a schema for role `joe` (the schema will also be named `joe`):

```sql
CREATE SCHEMA AUTHORIZATION joe;
```

Compatibility
The SQL standard allows a `DEFAULT CHARACTER SET` clause in `CREATE SCHEMA`, as well as more subcommand types than are presently accepted by Greenplum Database.

The SQL standard specifies that the subcommands in `CREATE SCHEMA` may appear in any order. The present Greenplum Database implementation does not handle all cases of forward references in subcommands; it may sometimes be necessary to reorder the subcommands in order to avoid forward references.

According to the SQL standard, the owner of a schema always owns all objects within it. Greenplum Database allows schemas to contain objects owned by users other than the schema owner. This can happen only if the schema owner grants the `CREATE` privilege on the schema to someone else.

See Also
`ALTER SCHEMA`, `DROP SCHEMA`
CREATE SEQUENCE

Defines a new sequence generator.

Synopsis
CREATE [TEMPORARY | TEMP] SEQUENCE name
    [INCREMENT [BY] value]
    [MINVALUE minvalue | NO MINVALUE]
    [MAXVALUE maxvalue | NO MAXVALUE]
    [START [ WITH ] start]
    [CACHE cache]
    [[NO] CYCLE]
    [OWNED BY { table.column | NONE }]

Description
CREATE SEQUENCE creates a new sequence number generator. This involves creating and initializing a new special single-row table. The generator will be owned by the user issuing the command.

If a schema name is given, then the sequence is created in the specified schema. Otherwise it is created in the current schema. Temporary sequences exist in a special schema, so a schema name may not be given when creating a temporary sequence. The sequence name must be distinct from the name of any other sequence, table, index, or view in the same schema.

After a sequence is created, you use the nextval function to operate on the sequence. For example, to insert a row into a table that gets the next value of a sequence:

    INSERT INTO distributors VALUES (nextval('myserial'),
                                'acme');

You can also use the function setval to operate on a sequence, but only for queries that do not operate on distributed data. For example, the following query is allowed because it resets the sequence counter value for the sequence generator process on the master:

    SELECT setval('myserial', 201);

But the following query will be rejected in Greenplum Database because it operates on distributed data:

    INSERT INTO product VALUES (setval('myserial', 201),
                                'gizmo');

In a regular (non-distributed) database, functions that operate on the sequence go to the local sequence table to get values as they are needed. In Greenplum Database, however, keep in mind that each segment is its own distinct database process. Therefore the segments need a single point of truth to go for sequence values so that all segments get incremented correctly and the sequence moves forward in the right order. A sequence server process runs on the master and is the point-of-truth for a sequence in a Greenplum distributed database. Segments get sequence values at runtime from the master.
Because of this distributed sequence design, there are some limitations on the functions that operate on a sequence in Greenplum Database:

- `lastval` and `currval` functions are not supported.
- `setval` can only be used to set the value of the sequence generator on the master, it cannot be used in subqueries to update records on distributed table data.
- `nextval` sometimes grabs a block of values from the master for a segment to use, depending on the query. So values may sometimes be skipped in the sequence if all of the block turns out not to be needed at the segment level. Note that a regular PostgreSQL database does this too, so this is not something unique to Greenplum Database.

Although you cannot update a sequence directly, you can use a query like:

```
SELECT * FROM sequence_name;
```

to examine the parameters and current state of a sequence. In particular, the `last_value` field of the sequence shows the last value allocated by any session.

### Parameters

**TEMPORARY | TEMP**

If specified, the sequence object is created only for this session, and is automatically dropped on session exit. Existing permanent sequences with the same name are not visible (in this session) while the temporary sequence exists, unless they are referenced with schema-qualified names.

**name**

The name (optionally schema-qualified) of the sequence to be created.

**increment**

Specifies which value is added to the current sequence value to create a new value. A positive value will make an ascending sequence, a negative one a descending sequence. The default value is 1.

**minvalue**

**NO MINVALUE**

Determines the minimum value a sequence can generate. If this clause is not supplied or `NO MINVALUE` is specified, then defaults will be used. The defaults are 1 and -263-1 for ascending and descending sequences, respectively.

**maxvalue**

**NO MAXVALUE**

Determines the maximum value for the sequence. If this clause is not supplied or `NO MAXVALUE` is specified, then default values will be used. The defaults are 263-1 and -1 for ascending and descending sequences, respectively.

**start**

Allows the sequence to begin anywhere. The default starting value is `minvalue` for ascending sequences and `maxvalue` for descending ones.
**cache**

Specifies how many sequence numbers are to be preallocated and stored in memory for faster access. The minimum (and default) value is 1 (no cache).

**CYCLE**

**NO CYCLE**

Allows the sequence to wrap around when the `maxvalue` (for ascending) or `minvalue` (for descending) has been reached. If the limit is reached, the next number generated will be the `minvalue` (for ascending) or `maxvalue` (for descending). If **NO CYCLE** is specified, any calls to `nextval` after the sequence has reached its maximum value will return an error. If not specified, **NO CYCLE** is the default.

**OWNED BY** `table.column`

**OWNED BY** `NONE`

Causes the sequence to be associated with a specific table column, such that if that column (or its whole table) is dropped, the sequence will be automatically dropped as well. The specified table must have the same owner and be in the same schema as the sequence. **OWNED BY** `NONE`, the default, specifies that there is no such association.

**Notes**

Sequences are based on bigint arithmetic, so the range cannot exceed the range of an eight-byte integer (-9223372036854775808 to 9223372036854775807).

Although multiple sessions are guaranteed to allocate distinct sequence values, the values may be generated out of sequence when all the sessions are considered. For example, session A might reserve values 1..10 and return `nextval=1`, then session B might reserve values 11..20 and return `nextval=11` before session A has generated `nextval=2`. Thus, you should only assume that the `nextval` values are all distinct, not that they are generated purely sequentially. Also, `last_value` will reflect the latest value reserved by any session, whether or not it has yet been returned by `nextval`.

**Examples**

Create a sequence named `myseq`:

```sql
CREATE SEQUENCE myseq START 101;
```

Insert a row into a table that gets the next value:

```sql
INSERT INTO distributors VALUES (nextval('myseq'), 'acme');
```

Reset the sequence counter value on the master:

```sql
SELECT setval('myseq', 201);
```

Illegal use of `setval` in Greenplum Database (setting sequence values on distributed data):

```sql
INSERT INTO product VALUES (setval('myseq', 201), 'gizmo');
```
Compatibility

`CREATE SEQUENCE` conforms to the SQL standard, with the following exceptions:

- The `AS data_type` expression specified in the SQL standard is not supported.
- Obtaining the next value is done using the `nextval()` function instead of the `NEXT VALUE FOR` expression specified in the SQL standard.
- The `OWNED BY` clause is a Greenplum Database extension.

See Also

`ALTER SEQUENCE`, `DROP SEQUENCE`
CREATE TABLE

Defines a new table.

Synopsis

CREATE [[GLOBAL | LOCAL] {TEMPORARY | TEMP}] TABLE table_name (  
  [ { column_name data_type [DEFAULT default_exp] }  
  [column_constraint [ ... ] ]  
  [table_constraint  
  | LIKE other_table [{INCLUDING | EXCLUDING}  
  (DEFAULTS | CONSTRAINTS)] ...}  
  [, ... ] ] )  
[ INHERITS ( parent_table [, ... ] ) ]  
[ WITH ( storage_parameter=value [, ... ] ) ]  
[ ON COMMIT {PRESERVE ROWS | DELETE ROWS | DROP} ]  
[ TABLESPACE tablespace ]  
[ DISTRIBUTED BY (column, [ ... ] ) | DISTRIBUTED RANDOMLY ]  
[ PARTITION BY partition_type (column)  
  [ SUBPARTITION BY partition_type (column) ]  
  [ SUBPARTITION TEMPLATE (template_spec) ] ]  
[...]
  (partition_spec)  
  | [ SUBPARTITION BY partition_type (column) ]  
  [...]
  (partition_spec
  [ (subpartition_spec  
  [([...])  
  ) ]  
)

where storage_parameter is:
APPENDONLY={TRUE|FALSE}
ORIENTATION={COLUMN|ROW}
COMPRESSTYPE={ZLIB|QUICKLZ}
COMPRESSLEVEL={0-9 | 1}
FILLFACTOR={10-100}
OIDS [=TRUE|FALSE]

where column_constraint is:
[CONSTRAINT constraint_name]
NOT NULL | NULL  
| UNIQUE [USING INDEX TABLESPACE tablespace]  
  [WITH ( FILLFACTOR = value )]
| PRIMARY KEY [USING INDEX TABLESPACE tablespace]  
  [WITH ( FILLFACTOR = value )]
| CHECK ( expression )

and table_constraint is:
[CONSTRAINT constraint_name]
CREATE TABLE

   UNIQUE ( column_name [, ... ] )
   [USING INDEX TABLESPACE tablespace]
   [WITH ( FILLFACTOR=value )]
| PRIMARY KEY ( column_name [, ... ] )
   [USING INDEX TABLESPACE tablespace]
   [WITH ( FILLFACTOR=value )]
| CHECK ( expression )

where partition_type is:
   LIST
| RANGE

where partition_specification is:
   partition_element [, ...]

and partition_element is:
   DEFAULT PARTITION name

| [PARTITION name] VALUES (list_value [, ...] )
| [PARTITION name]
   START ([datatype] 'start_value') [INCLUSIVE | EXCLUSIVE]
   [ END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE] ]
   [ EVERY ([datatype] [number | INTERVAL] 'interval_value') ]
| [PARTITION name]
   END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE]
   [ EVERY ([datatype] [number | INTERVAL] 'interval_value') ]

[ WITH ( partition_storage_parameter=value [, ... ] ) ]
[ TABLESPACE tablespace ]

where subpartition_spec or template_spec is:
   subpartition_element [, ...]

and subpartition_element is:
   DEFAULT SUBPARTITION name

| [SUBPARTITION name] VALUES (list_value [, ...] )
| [SUBPARTITION name]
   START ([datatype] 'start_value') [INCLUSIVE | EXCLUSIVE]
   [ END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE] ]
   [ EVERY ([datatype] [number | INTERVAL] 'interval_value') ]
| [SUBPARTITION name]
   END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE]
   [ EVERY ([datatype] [number | INTERVAL] 'interval_value') ]

[ WITH ( partition_storage_parameter=value [, ... ] ) ]
[ TABLESPACE tablespace ]

Description

CREATE TABLE will create a new, initially empty table in the current database. The table will be owned by the user issuing the command.
If a schema name is given then the table is created in the specified schema. Otherwise it is created in the current schema. Temporary tables exist in a special schema, so a schema name may not be given when creating a temporary table. The name of the table must be distinct from the name of any other table, external table, sequence, index, or view in the same schema.

The optional constraint clauses specify conditions that new or updated rows must satisfy for an insert or update operation to succeed. A constraint is an SQL object that helps define the set of valid values in the table in various ways.

There are two ways to define constraints: table constraints and column constraints. A column constraint is defined as part of a column definition. A table constraint definition is not tied to a particular column, and it can encompass more than one column. Every column constraint can also be written as a table constraint; a column constraint is only a notational convenience for use when the constraint only affects one column.

When creating a table, there is an additional clause to declare the Greenplum Database distribution policy. If a `DISTRIBUTED BY` or `DISTRIBUTED RANDOMLY` clause is not supplied, then Greenplum assigns a hash distribution policy to the table using either the `PRIMARY KEY` (if the table has one) or the first column of the table as the distribution key. Columns of geometric or user-defined data types are not eligible as Greenplum distribution key columns. If a table does not have a column of an eligible data type, the rows are distributed based on a round-robin or random distribution. To ensure an even distribution of data in your Greenplum Database system, you want to choose a distribution key that is unique for each record, or if that is not possible, then choose `DISTRIBUTED RANDOMLY`.

The `PARTITION BY` clause allows you to divide the table into multiple sub-tables (or child tables) that inherit from the parent table. Table partitioning creates a persistent relationship between the child table partitions and the parent table, so that all of the schema information from the parent table propagates to the child table partitions. Each child table partition is created with a distinct `CHECK` constraint, which limits the data the table can contain based on some defining criteria. The `CHECK` constraints are also used by the query planner to determine which table partitions to scan in order to satisfy a given query predicate.

### Parameters

**GLOBAL | LOCAL**

These keywords are present for SQL standard compatibility, but have no effect in Greenplum Database.

**TEMPORARY | TEMP**

If specified, the table is created as a temporary table. Temporary tables are automatically dropped at the end of a session, or optionally at the end of the current transaction (see `ON COMMIT`). Existing permanent tables with the same name are not visible to the current session while the temporary table exists, unless they are referenced with schema-qualified names. Any indexes created on a temporary table are automatically temporary as well.
**table_name**

The name (optionally schema-qualified) of the table to be created.

**column_name**

The name of a column to be created in the new table.

**data_type**

The data type of the column. This may include array specifiers.

**DEFAULT default_expr**

The DEFAULT clause assigns a default data value for the column whose column definition it appears within. The value is any variable-free expression (subqueries and cross-references to other columns in the current table are not allowed). The data type of the default expression must match the data type of the column. The default expression will be used in any insert operation that does not specify a value for the column. If there is no default for a column, then the default is null.

**INHERITS**

The optional INHERITS clause specifies a list of tables from which the new table automatically inherits all columns. Use of INHERITS creates a persistent relationship between the new child table and its parent table(s). Schema modifications to the parent(s) normally propagate to children as well, and by default the data of the child table is included in scans of the parent(s).

In Greenplum Database, the INHERITS clause is not used when creating partitioned tables. Although the concept of inheritance is used in partition hierarchies, the inheritance structure of a partitioned table is created using the PARTITION BY clause.

If the same column name exists in more than one parent table, an error is reported unless the data types of the columns match in each of the parent tables. If there is no conflict, then the duplicate columns are merged to form a single column in the new table. If the column name list of the new table contains a column name that is also inherited, the data type must likewise match the inherited column(s), and the column definitions are merged into one. However, inherited and new column declarations of the same name need not specify identical constraints: all constraints provided from any declaration are merged together and all are applied to the new table. If the new table explicitly specifies a default value for the column, this default overrides any defaults from inherited declarations of the column. Otherwise, any parents that specify default values for the column must all specify the same default, or an error will be reported.

**LIKE other_table [{INCLUDING | EXCLUDING} {DEFAULTS | CONSTRAINTS}]**

The LIKE clause specifies a table from which the new table automatically copies all column names, data types, not-null constraints, and distribution policy. Storage properties like append-only or partition structure are not copied. Unlike INHERITS, the new table and original table are completely decoupled after creation is complete.
Default expressions for the copied column definitions will only be copied if \texttt{INCLUDING DEFAULTS} is specified. The default behavior is to exclude default expressions, resulting in the copied columns in the new table having null defaults.

Not-null constraints are always copied to the new table. \texttt{CHECK} constraints will only be copied if \texttt{INCLUDING CONSTRAINTS} is specified; other types of constraints will \textit{never} be copied. Also, no distinction is made between column constraints and table constraints — when constraints are requested, all check constraints are copied.

Note also that unlike \texttt{INHERITS}, copied columns and constraints are not merged with similarly named columns and constraints. If the same name is specified explicitly or in another \texttt{LIKE} clause an error is signalled.

\textbf{CONSTRAINT} \texttt{constraint\_name}

An optional name for a column or table constraint. If the constraint is violated, the constraint name is present in error messages, so constraint names like \textit{column must be positive} can be used to communicate helpful constraint information to client applications. (Double-quotes are needed to specify constraint names that contain spaces.) If a constraint name is not specified, the system generates a name.

\textbf{NULL} | \textbf{NOT NULL}

Specifies if the column is or is not allowed to contain null values. \texttt{NULL} is the default.

\textbf{UNIQUE} ( \texttt{column\_constraint} )
\texttt{UNIQUE} ( \texttt{column\_name [, ... ]} ) ( \texttt{table\_constraint} )

The \texttt{UNIQUE} constraint specifies that a group of one or more columns of a table may contain only unique values. The behavior of the unique table constraint is the same as that for column constraints, with the additional capability to span multiple columns. For the purpose of a unique constraint, null values are not considered equal. The column(s) that are unique must also contain all of the Greenplum distribution key column(s).

\textbf{PRIMARY KEY} ( \texttt{column\_constraint} )
\texttt{PRIMARY KEY} ( \texttt{column\_name [, ... ]} ) ( \texttt{table\_constraint} )

The primary key constraint specifies that a column or columns of a table may contain only unique (non-duplicate), non-null values. Technically, \texttt{PRIMARY KEY} is merely a combination of \texttt{UNIQUE} and \texttt{NOT NULL}, but identifying a set of columns as primary key also provides metadata about the design of the schema, as a primary key implies that other tables may rely on this set of columns as a unique identifier for rows. For a table to have a primary key, it must be hash distributed (not randomly distributed), and the primary key must contain all (or a superset) of the Greenplum distribution key column(s).

\textbf{CHECK} ( \texttt{expression} )

The \texttt{CHECK} clause specifies an expression producing a Boolean result which new or updated rows must satisfy for an insert or update operation to succeed. Expressions evaluating to \texttt{TRUE} or \texttt{UNKNOWN} succeed. Should any row of an insert or update operation produce a \texttt{FALSE} result an error exception is raised and the insert or update does not alter the database. A check constraint specified as a column constraint
should reference that column’s value only, while an expression appearing in a table constraint may reference multiple columns. CHECK expressions cannot contain subqueries nor refer to variables other than columns of the current row.

**WITH ( storage_option=value )**

The **WITH** clause can be used to set storage options for the table or its indexes. Note that you can also set storage parameters on a particular partition or subpartition by declaring the **WITH** clause in the partition specification. The following storage options are available:

**APPENDONLY** - Set to **TRUE** to create the table as an append-only table. If **FALSE** or not declared, the table will be created as a regular heap-storage table.

**ORIENTATION** - Set to **column** for column-oriented storage, or **row** (the default) for row-oriented storage. This option is only valid if **APPENDONLY=TRUE**. Heap-storage tables can only be row-oriented.

**COMPRESSTYPE** - Set to **ZLIB** (the default) or **QUICKLZ** to specify the type of compression used. QuickLZ uses less CPU power and compresses data faster at a lower compression ratio than zlib. Conversely, zlib provides more compact compression ratios at lower speeds. This option is only valid if **APPENDONLY=TRUE**.

**COMPRESSLEVEL** - For zlib compression of append-only tables, set to a value between 1 (fastest compression) to 9 (highest compression ratio). QuickLZ compression level can only be set to 1. If not declared, the default is 0 (no compression). This option is only valid if **APPENDONLY=TRUE**.

**FILLFACTOR** - See **CREATE INDEX** for more information about this index storage parameter.

**OIDS** - Set to **OIDS=FALSE** (the default) so that rows do not have object identifiers assigned to them. Greenplum strongly recommends that you do not enable OIDS when creating a table. On large tables, such as those in a typical Greenplum Database system, using OIDs for table rows can cause wrap-around of the 32-bit OID counter. Once the counter wraps around, OIDs can no longer be assumed to be unique, which not only makes them useless to user applications, but can also cause problems in the Greenplum Database system catalog tables. In addition, excluding OIDs from a table reduces the space required to store the table on disk by 4 bytes per row, slightly improving performance. OIDS are not allowed on partitioned tables or append-only column-oriented tables.

**ON COMMIT**

The behavior of temporary tables at the end of a transaction block can be controlled using **ON COMMIT**. The three options are:

**PRESERVE ROWS**

No special action is taken at the ends of transactions for temporary tables. This is the default behavior.

**DELETE ROWS**
All rows in the temporary table will be deleted at the end of each transaction block. Essentially, an automatic TRUNCATE is done at each commit.

**DROP**

The temporary table will be dropped at the end of the current transaction block.

**TABLESPACE** *tablespace*

The name of the tablespace in which the new table is to be created. If not specified, the database’s default tablespace is used.

**USING INDEX TABLESPACE** *tablespace*

This clause allows selection of the tablespace in which the index associated with a UNIQUE or PRIMARY KEY constraint will be created. If not specified, the database’s default tablespace is used.

**DISTRIBUTED BY** (column, [ ... ])

**DISTRIBUTED RANDOMLY**

Used to declare the Greenplum Database distribution policy for the table. DISTRIBUTED BY uses hash distribution with one or more columns declared as the distribution key. For the most even data distribution, the distribution key should be the primary key of the table or a unique column (or set of columns). If that is not possible, then you may choose DISTRIBUTED RANDOMLY, which will send the data round-robin to the segment instances. If not supplied, then hash distribution is chosen using the PRIMARY KEY (if the table has one) or the first eligible column of the table as the distribution key.

**PARTITION BY**

Declares one or more columns by which to partition the table.

**partition_type**

Declares partition type: LIST (list of values) or RANGE (a numeric or date range).

**partition_specification**

Declares the individual partitions to create. Each partition can be defined individually or, for range partitions, you can use the EVERY clause (with a START and optional END clause) to define an increment pattern to use to create the individual partitions.

**DEFAULT PARTITION** *name* - Declares a default partition. When data does not match to an existing partition, it is inserted into the default partition. Partition designs that do not have a default partition will reject incoming rows that do not match to an existing partition.

**PARTITION** *name* - Declares a name to use for the partition. Partitions are created using the following naming convention:

`parentname_level#_prt_givenname`.

**VALUES** - For list partitions, defines the value(s) that the partition will contain.
**START** - For range partitions, defines the starting range value for the partition. By default, start values are **INCLUSIVE**. For example, if you declared a start date of ‘2008-01-01’, then the partition would contain all dates greater than or equal to ‘2008-01-01’. Typically the data type of the **START** expression is the same type as the partition key column. If that is not the case, then you must explicitly cast to the intended data type.

**END** - For range partitions, defines the ending range value for the partition. By default, end values are **EXCLUSIVE**. For example, if you declared an end date of ‘2008-02-01’, then the partition would contain all dates less than but not equal to ‘2008-02-01’. Typically the data type of the **END** expression is the same type as the partition key column. If that is not the case, then you must explicitly cast to the intended data type.

**EVERY** - For range partitions, defines how to increment the values from **START** to **END** to create individual partitions. Typically the data type of the **EVERY** expression is the same type as the partition key column. If that is not the case, then you must explicitly cast to the intended data type.

**WITH** - Sets the table storage options for a partition. For example, you may want older partitions to be append-only tables and newer partitions to be regular heap tables.

**TABLESPACE** - The name of the tablespace in which the partition is to be created.

**SUBPARTITION BY**
Declares one or more columns by which to subpartition the first-level partitions of the table. The format of the subpartition specification is similar to that of a partition specification described above.

**SUBPARTITION TEMPLATE**
Instead of declaring each subpartition definition individually for each partition, you can optionally declare a subpartition template to be used to create the subpartitions. This subpartition specification would then apply to all parent partitions.

**Notes**
Using OIDs in new applications is not recommended: where possible, using a **SERIAL** or other sequence generator as the table’s primary key is preferred. However, if your application does make use of OIDs to identify specific rows of a table, it is recommended to create a unique constraint on the OID column of that table, to ensure that OIDs in the table will indeed uniquely identify rows even after counter wrap-around. Avoid assuming that OIDs are unique across tables; if you need a database-wide unique identifier, use the combination of table OID and row OID for the purpose.

Greenplum Database has some special conditions for primary key and unique constraints with regards to columns that are the **distribution key** in a Greenplum table. For a unique constraint to be enforced in Greenplum Database, the table must be hash-distributed (not **DISTRIBUTED RANDOMLY**), and the constraint columns must be the same as (or a superset of) the table’s distribution key columns.
A primary key constraint is simply a combination of a unique constraint and a not-null constraint.

Greenplum Database automatically creates an index for each unique constraint or primary key constraint to enforce uniqueness. Thus, it is not necessary to create an index explicitly for primary key columns.

Foreign key constraints are not supported in Greenplum Database.

For inherited tables, unique constraints, primary key constraints, indexes and table privileges are not inherited in the current implementation.

Examples

Create a table named `rank` in the schema named `baby` and distribute the data using the columns `rank`, `gender`, and `year`:

```
CREATE TABLE baby.rank (id int, rank int, year smallint, gender char(1), count int ) DISTRIBUTED BY (rank, gender, year);
```

Create table `films` and table `distributors` (the primary key will be used as the Greenplum distribution key by default):

```
CREATE TABLE films (  
  code        char(5) CONSTRAINT firstkey PRIMARY KEY,  
  title       varchar(40) NOT NULL,  
  did         integer NOT NULL,  
  date_prod   date,  
  kind        varchar(10),  
  len         interval hour to minute  
);

CREATE TABLE distributors (  
  did    integer PRIMARY KEY DEFAULT nextval('serial'),  
  name   varchar(40) NOT NULL CHECK (name <> '')  
);
```

Create a gzip-compressed, append-only table:

```
CREATE TABLE sales (txn_id int, qty int, date date)  
WITH (appendonly=true, compresslevel=5)  
DISTRIBUTED BY (txn_id);
```

Create a three level partitioned table using subpartition templates and default partitions at each level:

```
CREATE TABLE sales (id int, year int, month int, day int, region text)  
DISTRIBUTED BY (id)  
PARTITION BY RANGE (year)  
SUBPARTITION BY RANGE (month)  
SUBPARTITION TEMPLATE (  
  START (1) END (13) EVERY (1),  
  DEFAULT SUBPARTITION other_months )
```
SUBPARTITION BY LIST (region)
SUBPARTITION TEMPLATE (  
    SUBPARTITION usa VALUES ('usa'),  
    SUBPARTITION europe VALUES ('europe'),  
    SUBPARTITION asia VALUES ('asia'),  
    DEFAULT SUBPARTITION other_regions) 
.DEFAULT PARTITION outlying_years);

Compatibility
CREATE TABLE command conforms to the SQL standard, with the following exceptions:

- **Temporary Tables** — In the SQL standard, temporary tables are defined just once and automatically exist (starting with empty contents) in every session that needs them. Greenplum Database instead requires each session to issue its own CREATE TEMPORARY TABLE command for each temporary table to be used. This allows different sessions to use the same temporary table name for different purposes, whereas the standard’s approach constrains all instances of a given temporary table name to have the same table structure.

  The standard’s distinction between global and local temporary tables is not in Greenplum Database. Greenplum Database will accept the GLOBAL and LOCAL keywords in a temporary table declaration, but they have no effect.

  If the ON COMMIT clause is omitted, the SQL standard specifies that the default behavior as ON COMMIT DELETE ROWS. However, the default behavior in Greenplum Database is ON COMMIT PRESERVE ROWS. The ON COMMIT DROP option does not exist in the SQL standard.

- **Column Check Constraints** — The SQL standard says that CHECK column constraints may only refer to the column they apply to; only CHECK table constraints may refer to multiple columns. Greenplum Database does not enforce this restriction; it treats column and table check constraints alike.

- **NULL Constraint** — The NULL constraint is a Greenplum Database extension to the SQL standard that is included for compatibility with some other database systems (and for symmetry with the NOT NULL constraint). Since it is the default for any column, its presence is not required.

- **Inheritance** — Multiple inheritance via the INHERITS clause is a Greenplum Database language extension. SQL:1999 and later define single inheritance using a different syntax and different semantics. SQL:1999-style inheritance is not yet supported by Greenplum Database.

- **Partitioning** — Table partitioning via the PARTITION BY clause is a Greenplum Database language extension.

- **Zero-column tables** — Greenplum Database allows a table of no columns to be created (for example, CREATE TABLE foo();). This is an extension from the SQL standard, which does not allow zero-column tables. Zero-column tables are not in themselves very useful, but disallowing them creates odd special cases for ALTER TABLE DROP COLUMN, so Greenplum decided to ignore this spec restriction.
- **WITH clause** — The WITH clause is a Greenplum Database extension; neither storage parameters nor OIDs are in the standard.

- **Tablespaces** — The Greenplum Database concept of tablespaces is not part of the SQL standard. The clauses TABLESPACE and USING INDEX TABLESPACE are extensions.

- **Data Distribution** — The Greenplum Database concept of a parallel or distributed database is not part of the SQL standard. The DISTRIBUTED clauses are extensions.

**See Also**

ALTER TABLE, DROP TABLE, CREATE EXTERNAL TABLE, CREATE TABLE AS
CREATE TABLE AS

Defines a new table from the results of a query.

Synopsis

CREATE [ [GLOBAL | LOCAL] {TEMPORARY | TEMP} ] TABLE table_name
   [(column_name [, ] ... )]
   [ WITH ( storage_parameter=value [, ... ] ) ]
   [ON COMMIT {PRESERVE ROWS | DELETE ROWS | DROP}]
   [TABLESPACE tablespace]
   AS query
   [DISTRIBUTED BY (column, [ ... ] ) | DISTRIBUTED RANDOMLY]

where storage_parameter is:
   APPENDONLY={TRUE|FALSE}
   ORIENTATION={COLUMN|ROW}
   COMPRESSION={ZLIB|QUICKLZ}
   COMPRESSLEVEL={0-9 | 1}
   FILLFACTOR={10-100}
   OIDS [=TRUE|FALSE]

Description

CREATE TABLE AS creates a table and fills it with data computed by a SELECT command. The table columns have the names and data types associated with the output columns of the SELECT, however you can override the column names by giving an explicit list of new column names.

CREATE TABLE AS creates a new table and evaluates the query just once to fill the new table initially. The new table will not track subsequent changes to the source tables of the query.

Parameters

GLOBAL | LOCAL

These keywords are present for SQL standard compatibility, but have no effect in Greenplum Database.

TEMPORARY | TEMP

If specified, the new table is created as a temporary table. Temporary tables are automatically dropped at the end of a session, or optionally at the end of the current transaction (see ON COMMIT). Existing permanent tables with the same name are not visible to the current session while the temporary table exists, unless they are referenced with schema-qualified names. Any indexes created on a temporary table are automatically temporary as well.

table_name

The name (optionally schema-qualified) of the new table to be created.
**column_name**

The name of a column in the new table. If column names are not provided, they are taken from the output column names of the query. If the table is created from an `EXECUTE` command, a column name list cannot be specified.

**WITH ( storage_parameter=value )**

The `WITH` clause can be used to set storage options for the table or its indexes. Note that you can also set different storage parameters on a particular partition or subpartition by declaring the `WITH` clause in the partition specification. The following storage options are available:

- **APPENDONLY** - Set to **TRUE** to create the table as an append-only table. If **FALSE** or not declared, the table will be created as a regular heap-storage table.
- **ORIENTATION** - Set to **column** for column-oriented storage, or **row** (the default) for row-oriented storage. This option is only valid if `APPENDONLY=TRUE`. Heap-storage tables can only be row-oriented.
- **COMPRESSTYPE** - Set to **ZLIB** (the default) or **QUICKLZ** to specify the type of compression used. QuickLZ uses less CPU power and compresses data faster at a lower compression ratio than zlib. Conversely, zlib provides more compact compression ratios at lower speeds. This option is only valid if `APPENDONLY=TRUE`.
- **COMPRESSLEVEL** - For zlib compression of append-only tables, set to a value between 1 (fastest compression) to 9 (highest compression ratio). QuickLZ compression level can only be set to 1. If not declared, the default is 0 (no compression). This option is only valid if `APPENDONLY=TRUE`.
- **FILLFACTOR** - See `CREATE INDEX` for more information about this index storage parameter.
- **OIDS** - Set to **OIDS=FALSE** (the default) so that rows do not have object identifiers assigned to them. Greenplum strongly recommends that you do not enable OIDS when creating a table. On large tables, such as those in a typical Greenplum Database system, using OIDs for table rows can cause wrap-around of the 32-bit OID counter. Once the counter wraps around, OIDs can no longer be assumed to be unique, which not only makes them useless to user applications, but can also cause problems in the Greenplum Database system catalog tables. In addition, excluding OIDs from a table reduces the space required to store the table on disk by 4 bytes per row, slightly improving performance. OIDS are not allowed on column-oriented tables.

**ON COMMIT**

The behavior of temporary tables at the end of a transaction block can be controlled using **ON COMMIT**. The three options are:

- **PRESERVE ROWS**
  No special action is taken at the ends of transactions for temporary tables. This is the default behavior.
- **DELETE ROWS**
All rows in the temporary table will be deleted at the end of each transaction block. Essentially, an automatic `TRUNCATE` is done at each commit.

**DROP**

The temporary table will be dropped at the end of the current transaction block.

**TABLESPACE** `tablespace`

The tablespace is the name of the tablespace in which the new table is to be created. If not specified, the database’s default tablespace is used.

**AS** `query`

A `SELECT` or `VALUES` command, or an `EXECUTE` command that runs a prepared `SELECT` or `VALUES` query.

**DISTRIBUTED BY** `(column, [ ... ] )`  
**DISTRIBUTED RANDOMLY**

Used to declare the Greenplum Database distribution policy for the table. One or more columns can be used as the distribution key, meaning those columns are used by the hashing algorithm to divide the data evenly across all of the segments. The distribution key should be the primary key of the table or a unique column (or set of columns). If that is not possible, then you may choose to distribute randomly, which will send the data round-robin to the segment instances. If not supplied, then either the `PRIMARY KEY` (if the table has one) or the first eligible column of the table will be used.

---

**Notes**

This command is functionally similar to `SELECT INTO`, but it is preferred since it is less likely to be confused with other uses of the `SELECT INTO` syntax. Furthermore, `CREATE TABLE AS` offers a superset of the functionality offered by `SELECT INTO`.

`CREATE TABLE AS` can be used for fast data loading from external table data sources. See `CREATE EXTERNAL TABLE`.

---

**Examples**

Create a new table `films_recent` consisting of only recent entries from the table `films`:

```sql
CREATE TABLE films_recent AS SELECT * FROM films WHERE date_prod >= '2007-01-01';
```

Create a new temporary table `films_recent`, consisting of only recent entries from the table `films`, using a prepared statement. The new table has OIDs and will be dropped at commit:

```sql
PREPARE recentfilms(date) AS SELECT * FROM films WHERE date_prod > $1;
CREATE TEMP TABLE films_recent WITH (OIDS) ON COMMIT DROP AS EXECUTE recentfilms('2007-01-01');
```
Compatibility

CREATE TABLE AS conforms to the SQL standard, with the following exceptions:

- The standard requires parentheses around the subquery clause; in Greenplum Database, these parentheses are optional.
- The standard defines a WITH [NO] DATA clause; this is not currently implemented by Greenplum Database. The behavior provided by Greenplum Database is equivalent to the standard’s WITH DATA case. WITH NO DATA can be simulated by appending LIMIT 0 to the query.
- Greenplum Database handles temporary tables differently from the standard; see CREATE TABLE for details.
- The WITH clause is a Greenplum Database extension; neither storage parameters nor OIDs are in the standard.
- The Greenplum Database concept of tablespaces is not part of the standard. The TABLESPACE clause is an extension.

See Also

CREATE EXTERNAL TABLE, CREATE EXTERNAL TABLE, EXECUTE, SELECT, SELECT INTO, VALUES
CREATE TABLESPACE

Defines a new tablespace.

Synopsis

CREATE TABLESPACE tablespace_name [OWNER username] FILESPACE filespace_name

Description

CREATE TABLESPACE registers a new tablespace for your Greenplum Database system. The tablespace name must be distinct from the name of any existing tablespace in the system.

A tablespace allows superusers to define an alternative location on the file system where the data files containing database objects (such as tables and indexes) may reside.

A user with appropriate privileges can pass a tablespace name to CREATE DATABASE, CREATE TABLE, or CREATE INDEX to have the data files for these objects stored within the specified tablespace.

In Greenplum Database, there must be a file system location defined for the master, each primary segment, and each mirror segment in order for the tablespace to have a location to store its objects across an entire Greenplum system. This collection of file system locations is defined in a filespace object. A filespace must be defined before you can create a tablespace. See gpfilespace for more information.

Parameters

tablespacename

The name of a tablespace to be created. The name cannot begin with pg_ or gp_, as such names are reserved for system tablespaces.

OWNER username

The name of the user who will own the tablespace. If omitted, defaults to the user executing the command. Only superusers may create tablespaces, but they can assign ownership of tablespaces to non-superusers.

FILESPACE

The name of a Greenplum Database filespace that was defined using the CREATE FILESPACE command or the gpfilespace management utility.

Notes

You must first create a filespace to be used by the tablespace. See gpfilespace. Tablespaces are only supported on systems that support symbolic links.

CREATE TABLESPACE cannot be executed inside a transaction block.
Examples
Create a new tablespace by specifying the corresponding filespace to use:

   CREATE TABLESPACE mytblspace FILESPACE myfilespace;

Compatibility
CREATE TABLESPACE is a Greenplum Database extension.

See Also
gpfilespace, CREATE DATABASE, CREATE TABLE, CREATE INDEX, DROP TABLESPACE, ALTER TABLESPACE
CREATE TRIGGER

Defines a new trigger. User-defined triggers are not supported in Greenplum Database.

Synopsis

```
CREATE TRIGGER name {BEFORE | AFTER} {event [OR ...]} 
  ON table [ FOR [EACH] {ROW | STATEMENT} ] 
  EXECUTE PROCEDURE funcname ( arguments )
```

Description

CREATE TRIGGER creates a new trigger. The trigger will be associated with the specified table and will execute the specified function when certain events occur.

Due to the distributed nature of a Greenplum Database system, the use of triggers is very limited in Greenplum Database. The function used in the trigger must be IMMUTABLE, meaning it cannot use information not directly present in its argument list. The function specified in the trigger also cannot execute any SQL or modify distributed database objects in any way. Given that triggers are most often used to alter tables (for example, update these other rows when this row is updated), these limitations offer very little practical use of triggers in Greenplum Database. For that reason, Greenplum does not support the use of user-defined triggers in Greenplum Database. Triggers cannot be used on append-only tables.

If multiple triggers of the same kind are defined for the same event, they will be fired in alphabetical order by name.

SELECT does not modify any rows so you can not create SELECT triggers. Rules and views are more appropriate in such cases.

Parameters

**name**

The name to give the new trigger. This must be distinct from the name of any other trigger for the same table.

**BEFORE**

**AFTER**

Determines whether the function is called before or after the event. If the trigger fires before the event, the trigger may skip the operation for the current row, or change the row being inserted (for INSERT and UPDATE operations only). If the trigger fires after the event, all changes, including the last insertion, update, or deletion, are visible to the trigger.

**event**

 Specifies the event that will fire the trigger (INSERT, UPDATE, or DELETE). Multiple events can be specified using OR.
table
The name (optionally schema-qualified) of the table the trigger is for.

FOR EACH ROW
FOR EACH STATEMENT
This specifies whether the trigger procedure should be fired once for every row
affected by the trigger event, or just once per SQL statement. If neither is specified,
FOR EACH STATEMENT is the default. A trigger that is marked FOR EACH ROW is
called once for every row that the operation modifies. In contrast, a trigger that is
marked FOR EACH STATEMENT only executes once for any given operation,
regardless of how many rows it modifies.

funcname
A user-supplied function that is declared as IMMUTABLE, taking no arguments, and
returning type trigger, which is executed when the trigger fires. This function
must not execute SQL or modify the database in any way.

arguments
An optional comma-separated list of arguments to be provided to the function when
the trigger is executed. The arguments are literal string constants. Simple names and
numeric constants may be written here, too, but they will all be converted to strings.
Please check the description of the implementation language of the trigger function
about how the trigger arguments are accessible within the function; it may be
different from normal function arguments.

Notes
To create a trigger on a table, the user must have the TRIGGER privilege on the table.

Examples
Declare the trigger function and then a trigger:

```
CREATE FUNCTION sendmail() RETURNS trigger AS
'$$GPHOME/lib/emailtrig.so' LANGUAGE C IMMUTABLE;

CREATE TRIGGER t_sendmail AFTER INSERT OR UPDATE OR DELETE
ON mytable FOR EACH STATEMENT EXECUTE PROCEDURE sendmail();
```

Compatibility
The CREATE TRIGGER statement in Greenplum Database implements a subset of the
SQL standard. The following functionality is currently missing:

- Greenplum Database has strict limitations on the function that is called by a
  trigger, which makes the use of triggers very limited in Greenplum Database. For
  this reason, triggers are not officially supported in Greenplum Database.
- SQL allows triggers to fire on updates to specific columns (e.g., AFTER UPDATE
  OF col1, col2).
SQL allows you to define aliases for the ‘old’ and ‘new’ rows or tables for use in the definition of the triggered action (e.g., CREATE TRIGGER ... ON tablename REFERENCING OLD ROW AS somename NEW ROW AS othername ...). Since Greenplum Database allows trigger procedures to be written in any number of user-defined languages, access to the data is handled in a language-specific way.

Greenplum Database only allows the execution of a user-defined function for the triggered action. The standard allows the execution of a number of other SQL commands, such as CREATE TABLE as the triggered action. This limitation is not hard to work around by creating a user-defined function that executes the desired commands.

SQL specifies that multiple triggers should be fired in time-of-creation order. Greenplum Database uses name order, which was judged to be more convenient.

SQL specifies that BEFORE DELETE triggers on cascaded deletes fire after the cascaded DELETE completes. The Greenplum Database behavior is for BEFORE DELETE to always fire before the delete action, even a cascading one. This is considered more consistent. There is also unpredictable behavior when BEFORE triggers modify rows that are later to be modified by referential actions. This can lead to constraint violations or stored data that does not honor the referential constraint.

The ability to specify multiple actions for a single trigger using OR is a Greenplum Database extension of the SQL standard.

See Also
CREATE FUNCTION, ALTER TRIGGER, DROP TRIGGER, CREATE RULE
CREATE TYPE

Defines a new data type.

Synopsis
CREATE TYPE name AS ( attribute_name data_type [, ... ] )
CREATE TYPE name ( INPUT = input_function,
OUTPUT = output_function
[, RECEIVE = receive_function]
[, SEND = send_function]
[, ANALYZE = analyze_function]
[, INTERNALLENGTH = {internallength | VARIABLE}]
[, PASSEDBYVALUE]
[, ALIGNMENT = alignment]
[, STORAGE = storage]
[, DEFAULT = default]
[, ELEMENT = element]
[, DELIMITER = delimiter]
)
CREATE TYPE name

Description
CREATE TYPE registers a new data type for use in the current database. The user who
defines a type becomes its owner.

If a schema name is given then the type is created in the specified schema. Otherwise
it is created in the current schema. The type name must be distinct from the name of
any existing type or domain in the same schema. The type name must also be distinct
from the name of any existing table in the same schema.

Composite Types
The first form of CREATE TYPE creates a composite type. The composite type is
specified by a list of attribute names and data types. This is essentially the same as the
row type of a table, but using CREATE TYPE avoids the need to create an actual table
when all that is wanted is to define a type. A stand-alone composite type is useful as
the argument or return type of a function.

Base Types
The second form of CREATE TYPE creates a new base type (scalar type). The
parameters may appear in any order, not only that shown in the syntax, and most are
optional. You must register two or more functions (using CREATE FUNCTION) before
defining the type. The support functions input_function and output_function
are required, while the functions receive_function, send_function and
analyze_function are optional. Generally these functions have to be coded in C or
another low-level language. In Greenplum Database, any function used to implement
a data type must be defined as IMMUTABLE.
The `input_function` converts the type’s external textual representation to the internal representation used by the operators and functions defined for the type. `output_function` performs the reverse transformation. The input function may be declared as taking one argument of type `cstring`, or as taking three arguments of types `cstring`, `oid`, and `integer`. The first argument is the input text as a C string, the second argument is the type’s own OID (except for array types, which instead receive their element type’s OID), and the third is the `typmod` of the destination column, if known (`-1` will be passed if not). The input function must return a value of the data type itself. Usually, an input function should be declared `STRICT`; if it is not, it will be called with a `NULL` first parameter when reading a `NULL` input value. The function must still return `NULL` in this case, unless it raises an error. (This case is mainly meant to support domain input functions, which may need to reject `NULL` inputs.) The output function must be declared as taking one argument of the new data type. The output function must return type `cstring`. Output functions are not invoked for `NULL` values.

The optional `receive_function` converts the type’s external binary representation to the internal representation. If this function is not supplied, the type cannot participate in binary input. The binary representation should be chosen to be cheap to convert to internal form, while being reasonably portable. (For example, the standard integer data types use network byte order as the external binary representation, while the internal representation is in the machine’s native byte order.) The receive function should perform adequate checking to ensure that the value is valid. The receive function may be declared as taking one argument of type `internal`, or as taking three arguments of types `internal`, `oid`, and `integer`. The first argument is a pointer to a `StringInfo` buffer holding the received byte string; the optional arguments are the same as for the text input function. The receive function must return a value of the data type itself. Usually, a receive function should be declared `STRICT`; if it is not, it will be called with a `NULL` first parameter when reading a `NULL` input value. The function must still return `NULL` in this case, unless it raises an error. (This case is mainly meant to support domain receive functions, which may need to reject `NULL` inputs.) Similarly, the optional `send_function` converts from the internal representation to the external binary representation. If this function is not supplied, the type cannot participate in binary output. The send function must be declared as taking one argument of the new data type. The send function must return type `bytea`. Send functions are not invoked for `NULL` values.

You should at this point be wondering how the input and output functions can be declared to have results or arguments of the new type, when they have to be created before the new type can be created. The answer is that the type should first be defined as a shell type, which is a placeholder type that has no properties except a name and an owner. This is done by issuing the command `CREATE TYPE name`, with no additional parameters. Then the I/O functions can be defined referencing the shell type. Finally, `CREATE TYPE` with a full definition replaces the shell entry with a complete, valid type definition, after which the new type can be used normally.

The optional `analyze_function` performs type-specific statistics collection for columns of the data type. By default, `ANALYZE` will attempt to gather statistics using the type’s ‘equals’ and ‘less-than’ operators, if there is a default b-tree operator class for the type. For non-scalar types this behavior is likely to be unsuitable, so it can be overridden by specifying a custom analysis function. The analysis function must be
declared to take a single argument of type `internal`, and return a `boolean` result. The detailed API for analysis functions appears in

\[ \text{\$GPHOME/include/commands/vacuum.h} \]

While the details of the new type's internal representation are only known to the I/O functions and other functions you create to work with the type, there are several properties of the internal representation that must be declared to Greenplum Database. Foremost of these is `internallength`. Base data types can be fixed-length, in which case `internallength` is a positive integer, or variable length, indicated by setting `internallength` to `VARIABLE`. (Internally, this is represented by setting `typlen` to `-1`.) The internal representation of all variable-length types must start with a 4-byte integer giving the total length of this value of the type.

The optional flag `PASSEDBYVALUE` indicates that values of this data type are passed by value, rather than by reference. You may not pass by value types whose internal representation is larger than the size of the `Datum` type (4 bytes on most machines, 8 bytes on a few).

The `alignment` parameter specifies the storage alignment required for the data type. The allowed values equate to alignment on 1, 2, 4, or 8 byte boundaries. Note that variable-length types must have an alignment of at least 4, since they necessarily contain an `int4` as their first component.

The `storage` parameter allows selection of storage strategies for variable-length data types. (Only `plain` is allowed for fixed-length types.) `plain` specifies that data of the type will always be stored in-line and not compressed. `extended` specifies that the system will first try to compress a long data value, and will move the value out of the main table row if it's still too long. `external` allows the value to be moved out of the main table, but the system will not try to compress it. `main` allows compression, but discourages moving the value out of the main table. (Data items with this storage strategy may still be moved out of the main table if there is no other way to make a row fit, but they will be kept in the main table preferentially over `extended` and `external` items.)

A default value may be specified, in case a user wants columns of the data type to default to something other than the null value. Specify the default with the `DEFAULT` keyword. (Such a default may be overridden by an explicit `DEFAULT` clause attached to a particular column.)

To indicate that a type is an array, specify the type of the array elements using the `ELEMENT` keyword. For example, to define an array of 4-byte integers (int4), specify `ELEMENT = int4`. More details about array types appear below.

To indicate the delimiter to be used between values in the external representation of arrays of this type, `delimiter` can be set to a specific character. The default delimiter is the comma (`,`). Note that the delimiter is associated with the array element type, not the array type itself.

**Array Types**

Whenever a user-defined base data type is created, Greenplum Database automatically creates an associated array type, whose name consists of the base type’s name prepended with an underscore. The parser understands this naming convention, and
translates requests for columns of type `foo[]` into requests for type `_foo`. The implicitly-created array type is variable length and uses the built-in input and output functions `array_in` and `array_out`.

You might reasonably ask why there is an `ELEMENT` option, if the system makes the correct array type automatically. The only case where it’s useful to use `ELEMENT` is when you are making a fixed-length type that happens to be internally an array of a number of identical things, and you want to allow these things to be accessed directly by subscripting, in addition to whatever operations you plan to provide for the type as a whole. For example, type `name` allows its constituent `char` elements to be accessed this way. A 2-D point type could allow its two component numbers to be accessed like `point[0]` and `point[1]`. Note that this facility only works for fixed-length types whose internal form is exactly a sequence of identical fixed-length fields. A subscriptable variable-length type must have the generalized internal representation used by `array_in` and `array_out`. For historical reasons, subscripting of fixed-length array types starts from zero, rather than from one as for variable-length arrays.

---

**Parameters**

**name**

The name (optionally schema-qualified) of a type to be created.

**attribute_name**

The name of an attribute (column) for the composite type.

**data_type**

The name of an existing data type to become a column of the composite type.

**input_function**

The name of a function that converts data from the type’s external textual form to its internal form.

**output_function**

The name of a function that converts data from the type’s internal form to its external textual form.

**receive_function**

The name of a function that converts data from the type’s external binary form to its internal form.

**send_function**

The name of a function that converts data from the type’s internal form to its external binary form.

**analyze_function**

The name of a function that performs statistical analysis for the data type.
**internallength**
A numeric constant that specifies the length in bytes of the new type’s internal representation. The default assumption is that it is variable-length.

**alignment**
The storage alignment requirement of the data type. Must be one of `char`, `int2`, `int4`, or `double`. The default is `int4`.

**storage**
The storage strategy for the data type. Must be one of `plain`, `external`, `extended`, or `main`. The default is `plain`.

**default**
The default value for the data type. If this is omitted, the default is null.

**element**
The type being created is an array; this specifies the type of the array elements.

**delimiter**
The delimiter character to be used between values in arrays made of this type.

---

**Notes**
User-defined type names cannot begin with the underscore character (`_`) and can only be 62 characters long (or in general `NAMEDATALEN - 2`, rather than the `NAMEDATALEN - 1` characters allowed for other names). Type names beginning with underscore are reserved for internally-created array type names.

Because there are no restrictions on use of a data type once it’s been created, creating a base type is tantamount to granting public execute permission on the functions mentioned in the type definition. (The creator of the type is therefore required to own these functions.) This is usually not an issue for the sorts of functions that are useful in a type definition. But you might want to think twice before designing a type in a way that would require ‘secret’ information to be used while converting it to or from external form.

Before Greenplum Database version 2.4, the syntax **CREATE TYPE** *name* did not exist. The way to create a new base type was to create its input function first. In this approach, Greenplum Database will first see the name of the new data type as the return type of the input function. The shell type is implicitly created in this situation, and then it can be referenced in the definitions of the remaining I/O functions. This approach still works, but is deprecated and may be disallowed in some future release. Also, to avoid accidentally cluttering the catalogs with shell types as a result of simple typos in function definitions, a shell type will only be made this way when the input function is written in C.

---

**Examples**
This example creates a composite type and uses it in a function definition:
CREATE TYPE compfoo AS (f1 int, f2 text);
CREATE FUNCTION getfoo() RETURNS SETOF compfoo AS $$
    SELECT fooid, fooname FROM foo
$$ LANGUAGE SQL;

This example creates the base data type box and then uses the type in a table definition:

CREATE TYPE box;
CREATE FUNCTION my_box_in_function(cstring) RETURNS box AS ...
CREATE FUNCTION my_box_out_function(box) RETURNS cstring AS ...

CREATE TYPE box (  
    INTERNALLENGTH = 16,  
    INPUT = my_box_in_function,  
    OUTPUT = my_box_out_function
);
CREATE TABLE myboxes (  
    id integer,  
    description box
);

If the internal structure of box were an array of four float4 elements, we might instead use:

CREATE TYPE box (  
    INTERNALLENGTH = 16,  
    INPUT = my_box_in_function,  
    OUTPUT = my_box_out_function,  
    ELEMENT = float4
);

which would allow a box value’s component numbers to be accessed by subscripting. Otherwise the type behaves the same as before.

This example creates a large object type and uses it in a table definition:

CREATE TYPE bigobj (  
    INPUT = lo_filein, OUTPUT = lo_fileout,  
    INTERNALLENGTH = VARIABLE
);
CREATE TABLE big_objs (  
    ...)
id integer,
obj bigobj
);

**Compatibility**

This `CREATE TYPE` command is a Greenplum Database extension. There is a `CREATE TYPE` statement in the SQL standard that is rather different in detail.

**See Also**

`CREATE FUNCTION`, `ALTER TYPE`, `DROP TYPE`, `CREATE DOMAIN`
CREATE USER

Defines a new database role with the LOGIN privilege by default.

**Synopsis**

CREATE USER name [ [WITH] option [ ... ] ]

where option can be:

- SUPERUSER | NOSUPERUSER
- CREATEDB | NOCREATEDB
- CREATEROLE | NOCREATEROLE
- CREATEUSER | NOCREATEUSER
- INHERIT | NOINHERIT
- LOGIN | NOLOGIN
- [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
- VALID UNTIL 'timestamp'
- IN ROLE rolename [, ...]
- IN GROUP rolename [, ...]
- ROLE rolename [, ...]
- ADMIN rolename [, ...]
- USER rolename [, ...]
- SYSID uid
- RESOURCE QUEUE queue_name

**Description**

As of Greenplum Database release 2.2, CREATE USER has been replaced by CREATE ROLE, although it is still accepted for backwards compatibility.

The only difference between CREATE ROLE and CREATE USER is that LOGIN is assumed by default with CREATE USER, whereas NOLOGIN is assumed by default with CREATE ROLE.

**Compatibility**

There is no CREATE USER statement in the SQL standard.

**See Also**

CREATE ROLE
CREATE VIEW

Defines a new view.

Synopsis

```
CREATE [OR REPLACE] [TEMP | TEMPORARY] VIEW name
   [ ( column_name [, ...] ) ]
   AS query
```

Description

CREATE VIEW defines a view of a query. The view is not physically materialized. Instead, the query is run every time the view is referenced in a query.

CREATE OR REPLACE VIEW is similar, but if a view of the same name already exists, it is replaced. You can only replace a view with a new query that generates the identical set of columns (same column names and data types).

If a schema name is given then the view is created in the specified schema. Otherwise it is created in the current schema. Temporary views exist in a special schema, so a schema name may not be given when creating a temporary view. The name of the view must be distinct from the name of any other view, table, sequence, or index in the same schema.

Parameters

TEMPORARY | TEMP

If specified, the view is created as a temporary view. Temporary views are automatically dropped at the end of the current session. Existing permanent relations with the same name are not visible to the current session while the temporary view exists, unless they are referenced with schema-qualified names. If any of the tables referenced by the view are temporary, the view is created as a temporary view (whether TEMPORARY is specified or not).

name

The name (optionally schema-qualified) of a view to be created.

column_name

An optional list of names to be used for columns of the view. If not given, the column names are deduced from the query.

query

A SELECT or VALUES command which will provide the columns and rows of the view.
Notes
Views in Greenplum Database are read only. The system will not allow an insert, update, or delete on a view. You can get the effect of an updatable view by creating rewrite rules on the view into appropriate actions on other tables. For more information see CREATE RULE.

Be careful that the names and data types of the view’s columns will be assigned the way you want. For example:

    CREATE VIEW vista AS SELECT 'Hello World';

is bad form in two ways: the column name defaults to `column?`, and the column data type defaults to `unknown`. If you want a string literal in a view’s result, use something like:

    CREATE VIEW vista AS SELECT text 'Hello World' AS hello;

Access to tables referenced in the view is determined by permissions of the view owner not the current user (even if the current user is a superuser). This can be confusing in the case of superusers, since superusers typically have access to all objects. In the case of a view, even superusers must be explicitly granted access to tables referenced in the view if they are not the owner of the view.

However, functions called in the view are treated the same as if they had been called directly from the query using the view. Therefore the user of a view must have permissions to call any functions used by the view.

If you create a view with an ORDER BY clause, the ORDER BY clause is ignored when you do a SELECT from the view.

Examples
Create a view consisting of all comedy films:

    CREATE VIEW comedies AS SELECT * FROM films WHERE kind = 'comedy';

Create a view that gets the top ten ranked baby names:

    CREATE VIEW topten AS SELECT name, rank, gender, year FROM names, rank WHERE rank < '11' AND names.id=rank.id;

Compatibility
The SQL standard specifies some additional capabilities for the CREATE VIEW statement that are not in Greenplum Database. The optional clauses for the full SQL command in the standard are:

- **CHECK OPTION** — This option has to do with updatable views. All INSERT and UPDATE commands on the view will be checked to ensure data satisfy the view-defining condition (that is, the new data would be visible through the view). If they do not, the update will be rejected.

- **LOCAL** — Check for integrity on this view.
CREATE VIEW

- **CASCADED** — Check for integrity on this view and on any dependent view. **CASCADED** is assumed if neither **CASCADED** nor **LOCAL** is specified.

**CREATE OR REPLACE VIEW** is a Greenplum Database language extension. So is the concept of a temporary view.

**See Also**

**SELECT, DROP VIEW**
DEALLOCATE

Deallocates a prepared statement.

Synopsis

DEALLOCATE [PREPARE] name

Description

DEALLOCATE is used to deallocate a previously prepared SQL statement. If you do not explicitly deallocate a prepared statement, it is deallocated when the session ends.

For more information on prepared statements, see PREPARE.

Parameters

PREPARE

Optional key word which is ignored.

name

The name of the prepared statement to deallocate.

Examples

Deallocated the previously prepared statement named insert_names:

    DEALLOCATE insert_names;

Compatibility

The SQL standard includes a DEALLOCATE statement, but it is only for use in embedded SQL.

See Also

EXECUTE, PREPARE
DECLARE

Defines a cursor.

Synopsis

DECLARE name [BINARY] [INSENSITIVE] [NO SCROLL] CURSOR
  [{WITH | WITHOUT} HOLD]
FOR query [FOR READ ONLY]

Description

DECLARE allows a user to create cursors, which can be used to retrieve a small number of rows at a time out of a larger query. Cursors can return data either in text or in binary format using FETCH.

Normal cursors return data in text format, the same as a SELECT would produce. Since data is stored natively in binary format, the system must do a conversion to produce the text format. Once the information comes back in text form, the client application may need to convert it to a binary format to manipulate it. In addition, data in the text format is often larger in size than in the binary format. Binary cursors return the data in a binary representation that may be more easily manipulated. Nevertheless, if you intend to display the data as text anyway, retrieving it in text form will save you some effort on the client side.

As an example, if a query returns a value of one from an integer column, you would get a string of 1 with a default cursor whereas with a binary cursor you would get a 4-byte field containing the internal representation of the value (in big-endian byte order).

Binary cursors should be used carefully. Many applications, including psql, are not prepared to handle binary cursors and expect data to come back in the text format.

**Note:** When the client application uses the ‘extended query’ protocol to issue a FETCH command, the Bind protocol message specifies whether data is to be retrieved in text or binary format. This choice overrides the way that the cursor is defined. The concept of a binary cursor as such is thus obsolete when using extended query protocol — any cursor can be treated as either text or binary.

Parameters

*name*

The name of the cursor to be created.

*binary*

Causes the cursor to return data in binary rather than in text format.
INSENSITIVE
Indicates that data retrieved from the cursor should be unaffected by updates to the tables underlying the cursor while the cursor exists. In Greenplum Database, all cursors are insensitive. This key word currently has no effect and is present for compatibility with the SQL standard.

NO SCROLL
A cursor cannot be used to retrieve rows in a nonsequential fashion. This is the default behavior in Greenplum Database, since scrollable cursors (SCROLL) are not supported.

WITH HOLD
WITHOUT HOLD
WITH HOLD specifies that the cursor may continue to be used after the transaction that created it successfully commits. WITHOUT HOLD specifies that the cursor cannot be used outside of the transaction that created it. WITHOUT HOLD is the default.

query
A SELECT or VALUES command which will provide the rows to be returned by the cursor.

FOR READ ONLY
Cursors can only be used in a read-only mode in Greenplum Database. Greenplum Database does not support updatable cursors (FOR UPDATE), so this is the default behavior.

Notes
Unless WITH HOLD is specified, the cursor created by this command can only be used within the current transaction. Thus, DECLARE without WITH HOLD is useless outside a transaction block: the cursor would survive only to the completion of the statement. Therefore Greenplum Database reports an error if this command is used outside a transaction block. Use BEGIN, COMMIT and ROLLBACK to define a transaction block.

If WITH HOLD is specified and the transaction that created the cursor successfully commits, the cursor can continue to be accessed by subsequent transactions in the same session. (But if the creating transaction is aborted, the cursor is removed.) A cursor created with WITH HOLD is closed when an explicit CLOSE command is issued on it, or the session ends. In the current implementation, the rows represented by a held cursor are copied into a temporary file or memory area so that they remain available for subsequent transactions.

Scrollable cursors are not currently supported in Greenplum Database. You can only use FETCH to move the cursor position forward, not backwards.

You can see all available cursors by querying the pg_cursors system view.

Examples
Declare a cursor:
DECLARE mycursor CURSOR FOR SELECT * FROM mytable;

**Compatibility**

SQL standard allows cursors only in embedded SQL and in modules. Greenplum Database permits cursors to be used interactively.

Greenplum Database does not implement an `OPEN` statement for cursors. A cursor is considered to be open when it is declared.

The SQL standard allows cursors to update table data. All Greenplum Database cursors are read only.

The SQL standard allows cursors to move both forward and backward. All Greenplum Database cursors are forward moving only (not scrollable).

Binary cursors are a Greenplum Database extension.

**See Also**

CLOSE, FETCH, MOVE, SELECT
**DELETE**

Deletes rows from a table.

---

**Synopsis**

```sql
DELETE FROM [ONLY] table [[AS] alias]
[USING usinglist]
[WHERE condition]
```

---

**Description**

`DELETE` deletes rows that satisfy the `WHERE` clause from the specified table. If the `WHERE` clause is absent, the effect is to delete all rows in the table. The result is a valid, but empty table.

By default, `DELETE` will delete rows in the specified table and all its child tables. If you wish to delete only from the specific table mentioned, you must use the `ONLY` clause.

There are two ways to delete rows in a table using information contained in other tables in the database: using sub-selects, or specifying additional tables in the `USING` clause. Which technique is more appropriate depends on the specific circumstances. Note that if you do join two or more tables in a `DELETE` command, those tables must have the same Greenplum distribution key column(s) and also be joined on the distribution key column(s).

You must have the `DELETE` privilege on the table to delete from it.

---

**Outputs**

On successful completion, a `DELETE` command returns a command tag of the form

```sql
DELETE count
```

The count is the number of rows deleted. If count is 0, no rows matched the condition (this is not considered an error).

---

**Parameters**

**ONLY**

If specified, delete rows from the named table only. When not specified, any tables inheriting from the named table are also processed.

**table**

The name (optionally schema-qualified) of an existing table.

**alias**

A substitute name for the target table. When an alias is provided, it completely hides the actual name of the table. For example, given `DELETE FROM foo AS f`, the remainder of the `DELETE` statement must refer to this table as `f` not `foo`. 
**usinglist**
A list of table expressions, allowing columns from other tables to appear in the
WHERE condition. This is similar to the list of tables that can be specified in the FROM
Clause of a SELECT statement; for example, an alias for the table name can be
specified. Do not repeat the target table in the usinglist, unless you wish to set up
a self-join.

**condition**
An expression returning a value of type boolean, which determines the rows that
are to be deleted.

---

**Notes**
You cannot use STABLE or VOLATILE functions in a DELETE statement if mirrors are
enabled. This can potentially cause the primary segment and its mirror to become
out-of-sync because the command is run first on the primary and then a second time
on the mirror in the current Greenplum Database implementation.

Greenplum Database lets you reference columns of other tables in the WHERE
condition by specifying the other tables in the USING clause. For example, to the name
Hannah from the rank table, one might do:

```
DELETE FROM rank USING names WHERE names.id = rank.id AND
name = 'Hannah';
```

What is essentially happening here is a join between rank and names, with all
successfully joined rows being marked for deletion. This syntax is not standard.
However, this join style is usually easier to write and faster to execute than a more
standard sub-select style, such as:

```
DELETE FROM rank WHERE id IN (SELECT id FROM names WHERE name
= 'Hannah');
```

Note also that in Greenplum Database, any tables that are joined in a delete operation
must have the same Greenplum distribution key and be joined on the column(s) that
make up the distribution key (equijoin).

When using DELETE to remove all the rows of a table (for example: DELETE * FROM
table;), Greenplum Database adds an implicit TRUNCATE command (when user
permissions allow). The added TRUNCATE command frees the disk space occupied by
the deleted rows without requiring a VACUUM of the table. This improves scan
performance of subsequent queries, and benefits ELT workloads that frequently insert
and delete from temporary tables.

---

**Examples**
Delete all films but musicals:

```
DELETE FROM films WHERE kind <> 'Musical';
```

Clear the table films:

```
DELETE FROM films;
```
Delete using an equijoin (assuming both Greenplum Database tables are distributed on the id column):

```
DELETE FROM rank USING names WHERE names.id = rank.id AND name = 'Hannah';
```

**Compatibility**

This command conforms to the SQL standard, except that the **USING** clause is a Greenplum Database extension.

**See Also**

TRUNCATE
DROP AGGREGATE

Removes an aggregate function.

Synopsis

DROP AGGREGATE [IF EXISTS] name ( type [, ...] ) [CASCADE | RESTRICT]

Description

DROP AGGREGATE will delete an existing aggregate function. To execute this command the current user must be the owner of the aggregate function.

Parameters

IF EXISTS

Do not throw an error if the aggregate does not exist. A notice is issued in this case.

name

The name (optionally schema-qualified) of an existing aggregate function.

type

An input data type on which the aggregate function operates. To reference a zero-argument aggregate function, write * in place of the list of input data types.

CASCADE

Automatically drop objects that depend on the aggregate function.

RESTRICT

Refuse to drop the aggregate function if any objects depend on it. This is the default.

Examples

To remove the aggregate function myavg for type integer:

```
DROP AGGREGATE myavg(integer);
```

Compatibility

There is no DROP AGGREGATE statement in the SQL standard.

See Also

ALTER AGGREGATE, CREATE AGGREGATE
**DROP CAST**

Removes a cast.

**Synopsis**

```
DROP CAST [IF EXISTS] (sourcetype AS targettype) [CASCADE | RESTRICT]
```

**Description**

DROP CAST will delete a previously defined cast. To be able to drop a cast, you must own the source or the target data type. These are the same privileges that are required to create a cast.

**Parameters**

**IF EXISTS**

Do not throw an error if the cast does not exist. A notice is issued in this case.

**sourcetype**

The name of the source data type of the cast.

**targettype**

The name of the target data type of the cast.

**CASCADE**

**RESTRICT**

These keywords have no effect since there are no dependencies on casts.

**Examples**

To drop the cast from type `text` to type `int`:

```
DROP CAST (text AS int);
```

**Compatibility**

There **DROP CAST** command conforms to the SQL standard.

**See Also**

CREATE CAST
DROP CONVERSION

Removes a conversion.

Synopsis

DROP CONVERSION [IF EXISTS] name [CASCADE | RESTRICT]

Description

DROP CONVERSION removes a previously defined conversion. To be able to drop a conversion, you must own the conversion.

Parameters

IF EXISTS

Do not throw an error if the conversion does not exist. A notice is issued in this case.

name

The name of the conversion. The conversion name may be schema-qualified.

CASCADE

RESTRICT

These keywords have no effect since there are no dependencies on conversions.

Examples

Drop the conversion named myname:

    DROP CONVERSION myname;

Compatibility

There is no DROP CONVERSION statement in the SQL standard.

See Also

ALTER CONVERSION, CREATE CONVERSION
**DROP DATABASE**

Removes a database.

**Synopsis**

DROP DATABASE [IF EXISTS] name

**Description**

DROP DATABASE drops a database. It removes the catalog entries for the database and deletes the directory containing the data. It can only be executed by the database owner. Also, it cannot be executed while you or anyone else are connected to the target database. (Connect to template1 or any other database to issue this command.)

DROP DATABASE cannot be undone. Use it with care!

**Parameters**

**IF EXISTS**

Do not throw an error if the database does not exist. A notice is issued in this case.

**name**

The name of the database to remove.

**Notes**

DROP DATABASE cannot be executed inside a transaction block.

This command cannot be executed while connected to the target database. Thus, it might be more convenient to use the program dropdb instead, which is a wrapper around this command.

**Examples**

Drop the database named testdb:

```sql
    DROP DATABASE testdb;
```

**Compatibility**

There is no DROP DATABASE statement in the SQL standard.

**See Also**

ALTER DATABASE, CREATE DATABASE
DROP DOMAIN

Removes a domain.

Synopsis

DROP DOMAIN [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

Description

DROP DOMAIN removes a previously defined domain. You must be the owner of a domain to drop it.

Parameters

IF EXISTS
Do not throw an error if the domain does not exist. A notice is issued in this case.

name
The name (optionally schema-qualified) of an existing domain.

CASCADE
Automatically drop objects that depend on the domain (such as table columns).

RESTRICT
Refuse to drop the domain if any objects depend on it. This is the default.

Examples

Drop the domain named zipcode:

    DROP DOMAIN zipcode;

Compatibility

This command conforms to the SQL standard, except for the IF EXISTS option, which is a Greenplum Database extension.

See Also

ALTER DOMAIN, CREATE DOMAIN
**DROP EXTERNAL TABLE**

Removes an external table definition.

**Synopsis**

```
DROP EXTERNAL [WEB] TABLE [IF EXISTS] name [CASCADE | RESTRICT]
```

**Description**

`DROP EXTERNAL TABLE` drops an existing external table definition from the database system. The external data sources or files are not deleted. To execute this command you must be the owner of the external table.

**Parameters**

- **WEB**
  - Optional keyword for dropping external web tables.

- **IF EXISTS**
  - Do not throw an error if the external table does not exist. A notice is issued in this case.

- **name**
  - The name (optionally schema-qualified) of an existing external table.

- **CASCADE**
  - Automatically drop objects that depend on the external table (such as views).

- **RESTRICT**
  - Refuse to drop the external table if any objects depend on it. This is the default.

**Examples**

Remove the external table named `staging` if it exists:

```
DROP EXTERNAL TABLE IF EXISTS staging;
```

**Compatibility**

There is no `DROP EXTERNAL TABLE` statement in the SQL standard.

**See Also**

- `CREATE EXTERNAL TABLE`
DROP FILESPACE

Removes a filespace.

Synopsis

DROP FILESPACE [IF EXISTS] filespacename

Description

DROP FILESPACE removes a filespace definition and its system-generated data directories from the system.

A filespace can only be dropped by its owner or a superuser. The filespace must be empty of all tablespace objects before it can be dropped. It is possible that tablespaces in other databases may still be using a filespace even if no tablespaces in the current database are using the filespace.

Parameters

IF EXISTS

Do not throw an error if the filespace does not exist. A notice is issued in this case.

tablespacename

The name of the filespace to remove.

Examples

Remove the tablespace myfs:

    DROP FILESPACE myfs;

Compatibility

There is no DROP FILESPACE statement in the SQL standard or in PostgreSQL.

See Also

ALTER FILESPACE, gpfilespace, DROP TABLESPACE
DROP FUNCTION

Removes a function.

Synopsis

DROP FUNCTION [IF EXISTS] name ( [ [argmode] [argname] argtype [, ...] ] ) [CASCADE | RESTRICT]

Description

DROP FUNCTION removes the definition of an existing function. To execute this command the user must be the owner of the function. The argument types to the function must be specified, since several different functions may exist with the same name and different argument lists.

Parameters

IF EXISTS

Do not throw an error if the function does not exist. A notice is issued in this case.

name

The name (optionally schema-qualified) of an existing function.

argmode

The mode of an argument: either IN, OUT, or INOUT. If omitted, the default is IN. Note that DROP FUNCTION does not actually pay any attention to OUT arguments, since only the input arguments are needed to determine the function’s identity. So it is sufficient to list the IN and INOUT arguments.

argname

The name of an argument. Note that DROP FUNCTION does not actually pay any attention to argument names, since only the argument data types are needed to determine the function’s identity.

argtype

The data type(s) of the function’s arguments (optionally schema-qualified), if any.

CASCADE

Automatically drop objects that depend on the function (such as operators or triggers).

RESTRICT

Refuse to drop the function if any objects depend on it. This is the default.
Examples
Drop the square root function:

    DROP FUNCTION sqrt(integer);

Compatibility
A DROP FUNCTION statement is defined in the SQL standard, but it is not compatible with this command.

See Also
CREATE FUNCTION, ALTER FUNCTION
DROP GROUP

Removes a database role.

**Synopsis**

DROP GROUP [IF EXISTS] name [, ...]

**Description**

DROP GROUP is an obsolete command, though still accepted for backwards compatibility. Groups (and users) have been superseded by the more general concept of roles. See DROP ROLE for more information.

**Parameters**

**IF EXISTS**

Do not throw an error if the role does not exist. A notice is issued in this case.

**name**

The name of an existing role.

**Compatibility**

There is no DROP GROUP statement in the SQL standard.

**See Also**

DROP ROLE
**DROP INDEX**

Removes an index.

**Synopsis**

```
DROP INDEX [IF EXISTS] name [, ...] [CASCADE | RESTRICT]
```

**Description**

`DROP INDEX` drops an existing index from the database system. To execute this command you must be the owner of the index.

**Parameters**

**IF EXISTS**

Do not throw an error if the index does not exist. A notice is issued in this case.

**name**

The name (optionally schema-qualified) of an existing index.

**CASCADE**

Automatically drop objects that depend on the index.

**RESTRICT**

Refuse to drop the index if any objects depend on it. This is the default.

**Examples**

Remove the index `title_idx`:

```
DROP INDEX title_idx;
```

**Compatibility**

`DROP INDEX` is a Greenplum Database language extension. There are no provisions for indexes in the SQL standard.

**See Also**

`ALTER INDEX`, `CREATE INDEX`, `REINDEX`
DROP LANGUAGE

Removes a procedural language.

Synopsis
DROP [PROCEDURAL] LANGUAGE [IF EXISTS] name [CASCADE | RESTRICT]

Description
DROP LANGUAGE will remove the definition of the previously registered procedural language. You must be a superuser to drop a language.

Parameters
PROCEDURAL
Optional keyword - has no effect.

IF EXISTS
Do not throw an error if the language does not exist. A notice is issued in this case.

name
The name of an existing procedural language. For backward compatibility, the name may be enclosed by single quotes.

CASCADE
Automatically drop objects that depend on the language (such as functions written in that language).

RESTRICT
Refuse to drop the language if any objects depend on it. This is the default.

Examples
Remove the procedural language pls sample:

    DROP LANGUAGE pls sample;

Compatibility
There is no DROP LANGUAGE statement in the SQL standard.

See Also
ALTER LANGUAGE, CREATE LANGUAGE
DROP OPERATOR

Removes an operator.

Synopsis

DROP OPERATOR [IF EXISTS] name ( {lefttype | NONE} , {righttype | NONE} ) [CASCADE | RESTRICT]

Description

DROP OPERATOR drops an existing operator from the database system. To execute this command you must be the owner of the operator.

Parameters

IF EXISTS
Do not throw an error if the operator does not exist. A notice is issued in this case.

name
The name (optionally schema-qualified) of an existing operator.

lefttype
The data type of the operator’s left operand; write NONE if the operator has no left operand.

righttype
The data type of the operator’s right operand; write NONE if the operator has no right operand.

CASCADE
Automatically drop objects that depend on the operator.

RESTRICT
Refuse to drop the operator if any objects depend on it. This is the default.

Examples

Remove the power operator \( a^b \) for type integer:

```sql
DROP OPERATOR ^ (integer, integer);
```

Remove the left unary bitwise complement operator \( \sim b \) for type bit:

```sql
DROP OPERATOR ~ (none, bit);
```

Remove the right unary factorial operator \( x! \) for type bigint:

```sql
DROP OPERATOR ! (bigint, none);
```
**Compatibility**

There is no `DROP OPERATOR` statement in the SQL standard.

**See Also**

`ALTER OPERATOR`, `CREATE OPERATOR`
DROP OPERATOR CLASS

Removes an operator class.

Synopsis
DROP OPERATOR CLASS [IF EXISTS] name USING index_method [CASCADE | RESTRICT]

Description
DROP OPERATOR drops an existing operator class. To execute this command you must be the owner of the operator class.

Parameters

IF EXISTS
Do not throw an error if the operator class does not exist. A notice is issued in this case.

name
The name (optionally schema-qualified) of an existing operator class.

index_method
The name of the index access method the operator class is for.

CASCADE
Automatically drop objects that depend on the operator class.

RESTRICT
Refuse to drop the operator class if any objects depend on it. This is the default.

Examples
Remove the B-tree operator class widget_ops:

    DROP OPERATOR CLASS widget_ops USING btree;

This command will not succeed if there are any existing indexes that use the operator class. Add CASCADE to drop such indexes along with the operator class.

Compatibility
There is no DROP OPERATOR CLASS statement in the SQL standard.

See Also
ALTER OPERATOR CLASS, CREATE OPERATOR CLASS
DROP OWNED

Removes database objects owned by a database role.

Synopsis

DROP OWNED BY name [, ...] [CASCADE | RESTRICT]

Description

DROP OWNED drops all the objects in the current database that are owned by one of the specified roles. Any privileges granted to the given roles on objects in the current database will also be revoked.

Parameters

name

The name of a role whose objects will be dropped, and whose privileges will be revoked.

CASCADE

Automatically drop objects that depend on the affected objects.

RESTRICT

Refuse to drop the objects owned by a role if any other database objects depend on one of the affected objects. This is the default.

Notes

DROP OWNED is often used to prepare for the removal of one or more roles. Because DROP OWNED only affects the objects in the current database, it is usually necessary to execute this command in each database that contains objects owned by a role that is to be removed.

Using the CASCADE option may make the command recurse to objects owned by other users.

The REASSIGN OWNED command is an alternative that reassigns the ownership of all the database objects owned by one or more roles.

Examples

Remove any database objects owned by the role named sally:

    DROP OWNED BY sally;

Compatibility

The DROP OWNED statement is a Greenplum Database extension.
See Also

REASSIGN OWNED, DROP ROLE
DROP RESOURCE QUEUE

Removes a resource queue.

Synopsis

DROP RESOURCE QUEUE queue_name

Description

This command removes a workload management resource queue from Greenplum Database. To drop a resource queue, the queue cannot have any roles assigned to it, nor can it have any statements waiting in the queue. Only a superuser can drop a resource queue.

Parameters

queue_name

The name of a resource queue to remove.

Notes

Use ALTER ROLE to remove a user from a resource queue.

To see all the currently active queries for all resource queues, perform the following query of the pg_locks table joined with the pg_roles and pg_resqueue tables:

```
SELECT rolname, rsqname, locktype, objid, transaction, pid, mode, granted FROM pg_roles, pg_resqueue, pg_locks WHERE pg_roles.rolresqueue=pg_locks.objid AND pg_locks.objid=pg_resqueue.oid;
```

To see the roles assigned to a resource queue, perform the following query of the pg_roles and pg_resqueue system catalog tables:

```
SELECT rolname, rsqname FROM pg_roles, pg_resqueue WHERE pg_roles.rolresqueue=pg_resqueue.oid;
```

Examples

Remove a role from a resource queue (and move the role to the default resource queue, pg_default):

```
ALTER ROLE bob RESOURCE QUEUE NONE;
```

Remove the resource queue named adhoc:

```
DROP RESOURCE QUEUE adhoc;
```

Compatibility

The DROP RESOURCE QUEUE statement is a Greenplum Database extension.
See Also

ALTER RESOURCE QUEUE, CREATE RESOURCE QUEUE, ALTER ROLE
DROP ROLE

Removes a database role.

Synopsis
DROP ROLE [IF EXISTS] name [, ...]

Description
DROP ROLE removes the specified role(s). To drop a superuser role, you must be a superuser yourself. To drop non-superuser roles, you must have CREATEROLE privilege.

A role cannot be removed if it is still referenced in any database; an error will be raised if so. Before dropping the role, you must drop all the objects it owns (or reassign their ownership) and revoke any privileges the role has been granted. The REASSIGN OWNED and DROP OWNED commands can be useful for this purpose.

However, it is not necessary to remove role memberships involving the role; DROP ROLE automatically revokes any memberships of the target role in other roles, and of other roles in the target role. The other roles are not dropped nor otherwise affected.

Parameters

IF EXISTS
Do not throw an error if the role does not exist. A notice is issued in this case.

name
The name of the role to remove.

Examples
Remove the roles named sally and bob:

    DROP ROLE sally, bob;

Compatibility
The SQL standard defines DROP ROLE, but it allows only one role to be dropped at a time, and it specifies different privilege requirements than Greenplum Database uses.

See Also
REASSIGN OWNED, DROP OWNED, CREATE ROLE, ALTER ROLE, SET ROLE
DROP RULE

Removes a rewrite rule.

Synopsis

DROP RULE [IF EXISTS] name ON relation [CASCADE | RESTRICT]

Description

DROP RULE drops a rewrite rule from a table or view.

Parameters

IF EXISTS
Do not throw an error if the rule does not exist. A notice is issued in this case.

name
The name of the rule to remove.

relation
The name (optionally schema-qualified) of the table or view that the rule applies to.

CASCADE
Automatically drop objects that depend on the rule.

RESTRICT
Refuse to drop the rule if any objects depend on it. This is the default.

Examples

Remove the rewrite rule sales_2006 on the table sales:

    DROP RULE sales_2006 ON sales;

Compatibility

There is no DROP RULE statement in the SQL standard.

See Also

CREATE RULE
DROP SCHEMA

Removes a schema.

Synopsis

DROP SCHEMA [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

Description

DROP SCHEMA removes schemas from the database. A schema can only be dropped by its owner or a superuser. Note that the owner can drop the schema (and thereby all contained objects) even if he does not own some of the objects within the schema.

Parameters

IF EXISTS

Do not throw an error if the schema does not exist. A notice is issued in this case.

name

The name of the schema to remove.

CASCADE

Automatically drops any objects contained in the schema (tables, functions, etc.).

RESTRICT

Refuse to drop the schema if it contains any objects. This is the default.

Examples

Remove the schema mystuff from the database, along with everything it contains:

    DROP SCHEMA mystuff CASCADE;

Compatibility

DROP SCHEMA is fully conforming with the SQL standard, except that the standard only allows one schema to be dropped per command. Also, the IF EXISTS option is a Greenplum Database extension.

See Also

CREATE SCHEMA, ALTER SCHEMA
DROP SEQUENCE

Removes a sequence.

Synopsis

DROP SEQUENCE [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

Description

DROP SEQUENCE removes a sequence generator table. You must own the sequence to drop it (or be a superuser).

Parameters

IF EXISTS
Do not throw an error if the sequence does not exist. A notice is issued in this case.

name
The name (optionally schema-qualified) of the sequence to remove.

CASCADE
Automatically drop objects that depend on the sequence.

RESTRICT
Refuse to drop the sequence if any objects depend on it. This is the default.

Examples

Remove the sequence myserial:

    DROP SEQUENCE myserial;

Compatibility

DROP SEQUENCE is fully conforming with the SQL standard, except that the standard only allows one sequence to be dropped per command. Also, the IF EXISTS option is a Greenplum Database extension.

See Also

ALTER SEQUENCE, CREATE SEQUENCE
DROP TABLE

Removes a table.

Synopsis
DROP TABLE [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

Description
DROP TABLE removes tables from the database. Only its owner may drop a table. To empty a table of rows without removing the table definition, use DELETE or TRUNCATE.

DROP TABLE always removes any indexes, rules, triggers, and constraints that exist for the target table. However, to drop a table that is referenced by a view, CASCADE must be specified. CASCADE will remove a dependent view entirely.

Parameters
IF EXISTS
Do not throw an error if the table does not exist. A notice is issued in this case.

name
The name (optionally schema-qualified) of the table to remove.

CASCADE
Automatically drop objects that depend on the table (such as views).

RESTRICT
Refuse to drop the table if any objects depend on it. This is the default.

Examples
Remove the table mytable:

    DROP TABLE mytable;

Compatibility
DROP TABLE is fully conforming with the SQL standard, except that the standard only allows one table to be dropped per command. Also, the IF EXISTS option is a Greenplum Database extension.

See Also
CREATE TABLE, ALTER TABLE, TRUNCATE
DROP TABLESPACE

Removes a tablespace.

**Synopsis**

DROP TABLESPACE [IF EXISTS] `tablespacename`

**Description**

DROP TABLESPACE removes a tablespace from the system.

A tablespace can only be dropped by its owner or a superuser. The tablespace must be empty of all database objects before it can be dropped. It is possible that objects in other databases may still reside in the tablespace even if no objects in the current database are using the tablespace.

**Parameters**

- **IF EXISTS**
  Do not throw an error if the tablespace does not exist. A notice is issued in this case.

- **tablespacename**
  The name of the tablespace to remove.

**Examples**

Remove the tablespace `mystuff`:

```sql
DROP TABLESPACE mystuff;
```

**Compatibility**

DROP TABLESPACE is a Greenplum Database extension.

**See Also**

CREATE TABLESPACE, ALTER TABLESPACE
DROP TRIGGER

Removes a trigger.

Synopsis

DROP TRIGGER [IF EXISTS] name ON table [CASCADE | RESTRICT]

Description

DROP TRIGGER will remove an existing trigger definition. To execute this command, the current user must be the owner of the table for which the trigger is defined.

Parameters

IF EXISTS

Do not throw an error if the trigger does not exist. A notice is issued in this case.

name

The name of the trigger to remove.

table

The name (optionally schema-qualified) of the table for which the trigger is defined.

CASCADE

Automatically drop objects that depend on the trigger.

RESTRICT

Refuse to drop the trigger if any objects depend on it. This is the default.

Examples

Remove the trigger sendmail on table expenses;

DROP TRIGGER sendmail ON expenses;

Compatibility

The DROP TRIGGER statement in Greenplum Database is not compatible with the SQL standard. In the SQL standard, trigger names are not local to tables, so the command is simply DROP TRIGGER name.

See Also

ALTER TRIGGER, CREATE TRIGGER
DROP TYPE

Removes a data type.

Synopsis

DROP TYPE [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

Description

DROP TYPE will remove a user-defined data type. Only the owner of a type can remove it.

Parameters

IF EXISTS
Do not throw an error if the type does not exist. A notice is issued in this case.

name
The name (optionally schema-qualified) of the data type to remove.

CASCADE
Automatically drop objects that depend on the type (such as table columns, functions, operators).

RESTRICT
Refuse to drop the type if any objects depend on it. This is the default.

Examples

Remove the data type box;

    DROP TYPE box;

Compatibility

This command is similar to the corresponding command in the SQL standard, apart from the IF EXISTS option, which is a Greenplum Database extension. But note that the CREATE TYPE command and the data type extension mechanisms in Greenplum Database differ from the SQL standard.

See Also

ALTER TYPE, CREATE TYPE
DROP USER

Removes a database role.

Synopsis

DROP USER [IF EXISTS] name [, ...]

Description

DROP USER is an obsolete command, though still accepted for backwards compatibility. Groups (and users) have been superseded by the more general concept of roles. See DROP ROLE for more information.

Parameters

IF EXISTS

Do not throw an error if the role does not exist. A notice is issued in this case.

name

The name of an existing role.

Compatibility

There is no DROP USER statement in the SQL standard. The SQL standard leaves the definition of users to the implementation.

See Also

DROP ROLE
DROP VIEW

Removes a view.

Synopsis

DROP VIEW [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

Description

DROP VIEW will remove an existing view. Only the owner of a view can remove it.

Parameters

IF EXISTS
Do not throw an error if the view does not exist. A notice is issued in this case.

name
The name (optionally schema-qualified) of the view to remove.

CASCADE
Automatically drop objects that depend on the view (such as other views).

RESTRICT
Refuse to drop the view if any objects depend on it. This is the default.

Examples

Remove the view topten;

DROP VIEW topten;

Compatibility

DROP VIEW is fully conforming with the SQL standard, except that the standard only allows one view to be dropped per command. Also, the IF EXISTS option is a Greenplum Database extension.

See Also

CREATE VIEW
END

Commits the current transaction.

Synopsis

END [WORK | TRANSACTION]

Description

END commits the current transaction. All changes made by the transaction become visible to others and are guaranteed to be durable if a crash occurs. This command is a Greenplum Database extension that is equivalent to COMMIT.

Parameters

WORK
TRANSACTION

Optional keywords. They have no effect.

Examples

Commit the current transaction:

    END;

Compatibility

END is a Greenplum Database extension that provides functionality equivalent to COMMIT, which is specified in the SQL standard.

See Also

BEGIN, ROLLBACK, COMMIT
EXECUTE

Executes a prepared SQL statement.

Synopsis

EXECUTE name [ (parameter [, ...] ) ]

Description

EXECUTE is used to execute a previously prepared statement. Since prepared statements only exist for the duration of a session, the prepared statement must have been created by a PREPARE statement executed earlier in the current session.

If the PREPARE statement that created the statement specified some parameters, a compatible set of parameters must be passed to the EXECUTE statement, or else an error is raised. Note that (unlike functions) prepared statements are not overloaded based on the type or number of their parameters; the name of a prepared statement must be unique within a database session.

For more information on the creation and usage of prepared statements, see PREPARE.

Parameters

name

The name of the prepared statement to execute.

parameter

The actual value of a parameter to the prepared statement. This must be an expression yielding a value that is compatible with the data type of this parameter, as was determined when the prepared statement was created.

Examples

Create a prepared statement for an INSERT statement, and then execute it:

```
PREPARE fooplan (int, text, bool, numeric) AS INSERT INTO foo VALUES($1, $2, $3, $4);
EXECUTE fooplan(1, 'Hunter Valley', 't', 200.00);
```

Compatibility

The SQL standard includes an EXECUTE statement, but it is only for use in embedded SQL. This version of the EXECUTE statement also uses a somewhat different syntax.

See Also

DEALLOCATE, PREPARE
**EXPLAIN**

Shows the query plan of a statement.

**Synopsis**

```
EXPLAIN [ANALYZE] [VERBOSE] statement
```

**Description**

`EXPLAIN` displays the query plan that the Greenplum planner generates for the supplied statement. Query plans are a tree plan of nodes. Each node in the plan represents a single operation, such as table scan, join, aggregation or a sort.

Plans should be read from the bottom up as each node feeds rows into the node directly above it. The bottom nodes of a plan are usually table scan operations (sequential, index or bitmap index scans). If the query requires joins, aggregations, or sorts (or other operations on the raw rows) then there will be additional nodes above the scan nodes to perform these operations. The topmost plan nodes are usually the Greenplum Database motion nodes (redistribute, broadcast, or gather motions). These are the operations responsible for moving rows between the segment instances during query processing.

The output of `EXPLAIN` has one line for each node in the plan tree, showing the basic node type plus the following cost estimates that the planner made for the execution of that plan node:

- **cost** - measured in units of disk page fetches; that is, 1.0 equals one sequential disk page read. The first estimate is the start-up cost (cost of getting to the first row) and the second is the total cost (cost of getting all rows). Note that the total cost assumes that all rows will be retrieved, which may not always be the case (if using `LIMIT` for example).

- **rows** - the total number of rows output by this plan node. This is usually less than the actual number of rows processed or scanned by the plan node, reflecting the estimated selectivity of any `WHERE` clause conditions. Ideally the top-level nodes estimate will approximate the number of rows actually returned, updated, or deleted by the query.

- **width** - total bytes of all the rows output by this plan node.

It is important to note that the cost of an upper-level node includes the cost of all its child nodes. The topmost node of the plan has the estimated total execution cost for the plan. This is this number that the planner seeks to minimize. It is also important to realize that the cost only reflects things that the query planner cares about. In particular, the cost does not consider the time spent transmitting result rows to the client.

`EXPLAIN ANALYZE` causes the statement to be actually executed, not only planned. The `EXPLAIN ANALYZE` plan shows the actual results along with the planner’s estimates. This is useful for seeing whether the planner’s estimates are close to reality. In addition to the information shown in the `EXPLAIN` plan, `EXPLAIN ANALYZE` will show the following additional information:
The total elapsed time (in milliseconds) that it took to run the query.

- The number of workers (segments) involved in a plan node operation. Only segments that return rows are counted.
- The maximum number of rows returned by the segment that produced the most rows for an operation. If multiple segments produce an equal number of rows, the one with the longest time to end is the one chosen.
- The segment id number of the segment that produced the most rows for an operation.
- For relevant operations, the work_mem used by the operation. If work_mem was not sufficient to perform the operation in memory, the plan will show how much data was spilled to disk and how many passes over the data were required for the lowest performing segment. For example:

```
Work_mem used: 64K bytes avg, 64K bytes max (seg0).
Work_mem wanted: 90K bytes avg, 90K bytes max (seg0) to abate workfile
I/O affecting 2 workers.
```

```
[seg0] pass 0: 488 groups made from 488 rows; 263 rows written to
workfile
[seg0] pass 1: 263 groups made from 263 rows
```
- The time (in milliseconds) it took to retrieve the first row from the segment that produced the most rows, and the total time taken to retrieve all rows from that segment. The <time> to first row may be omitted if it is the same as the <time> to end.

**Important:** Keep in mind that the statement is actually executed when EXPLAIN ANALYZE is used. Although EXPLAIN ANALYZE will discard any output that a SELECT would return, other side effects of the statement will happen as usual. If you wish to use EXPLAIN ANALYZE on a DML statement without letting the command affect your data, use this approach:

```
BEGIN;
EXPLAIN ANALYZE ...;
ROLLBACK;
```

### Parameters

**name**

The name of the prepared statement to execute.

**parameter**

The actual value of a parameter to the prepared statement. This must be an expression yielding a value that is compatible with the data type of this parameter, as was determined when the prepared statement was created.

### Notes

In order to allow the query planner to make reasonably informed decisions when optimizing queries, the `ANALYZE` statement should be run to record statistics about the distribution of data within the table. If you have not done this (or if the statistical
distribution of the data in the table has changed significantly since the last time
ANALYZE was run), the estimated costs are unlikely to conform to the real properties of
the query, and consequently an inferior query plan may be chosen.

See also, “Query Profiling” on page 119.

**Examples**

To illustrate how to read an EXPLAIN query plan, consider the following example for a
very simple query:

```
EXPLAIN SELECT * FROM names WHERE name = 'Joelle';
```

```
QUERY PLAN
-------------------------------------------------------------
Gather Motion 2:1 (slice1) (cost=0.00..20.88 rows=1 width=13)
  -> Seq Scan on 'names' (cost=0.00..20.88 rows=1 width=13)
       Filter: name::text ~~ 'Joelle'::text
```

If we read the plan from the bottom up, the query planner starts by doing a sequential
scan of the names table. Notice that the WHERE clause is being applied as a filter
condition. This means that the scan operation checks the condition for each row it
scans, and outputs only the ones that pass the condition.

The results of the scan operation are passed up to a gather motion operation. In
Greenplum Database, a gather motion is when segments send rows up to the master. In
this case we have 2 segment instances sending to 1 master instance (2:1). This
operation is working on slice1 of the parallel query execution plan. In Greenplum
Database a query plan is divided into slices so that portions of the query plan can be
worked on in parallel by the segments.

The estimated startup cost for this plan is 00.00 (no cost) and a total cost of 20.88
disk page fetches. The planner is estimating that this query will return one row.

**Compatibility**

There is no EXPLAIN statement defined in the SQL standard.

**See Also**

ANALYZE
**FETCH**

Retrieves rows from a query using a cursor.

**Synopsis**

```
FETCH [ forward_direction { FROM | IN } ] cursorname
```

where `forward_direction` can be empty or one of:

- NEXT
- FIRST
- LAST
- ABSOLUTE count
- RELATIVE count
- count
- ALL
- FORWARD
- FORWARD count
- FORWARD ALL

**Description**

`FETCH` retrieves rows using a previously-created cursor.

A cursor has an associated position, which is used by `FETCH`. The cursor position can be before the first row of the query result, on any particular row of the result, or after the last row of the result. When created, a cursor is positioned before the first row. After fetching some rows, the cursor is positioned on the row most recently retrieved. If `FETCH` runs off the end of the available rows then the cursor is left positioned after the last row. `FETCH ALL` will always leave the cursor positioned after the last row.

The forms `NEXT`, `FIRST`, `LAST`, `ABSOLUTE`, `RELATIVE` fetch a single row after moving the cursor appropriately. If there is no such row, an empty result is returned, and the cursor is left positioned before the first row or after the last row as appropriate.

The forms using `FORWARD` retrieve the indicated number of rows moving in the forward direction, leaving the cursor positioned on the last-returned row (or after all rows, if the count exceeds the number of rows available). Note that it is not possible to move a cursor position backwards in Greenplum Database, since scrollable cursors are not supported. You can only move a cursor forward in position using `FETCH`.

`RELATIVE 0` and `FORWARD 0` request fetching the current row without moving the cursor, that is, re-fetching the most recently fetched row. This will succeed unless the cursor is positioned before the first row or after the last row, in which case no row is returned.

**Outputs**

On successful completion, a `FETCH` command returns a command tag of the form

```
FETCH count
```
The count is the number of rows fetched (possibly zero). Note that in psql, the command tag will not actually be displayed, since psql displays the fetched rows instead.

**Parameters**

*forward_direction*

Defines the fetch direction and number of rows to fetch. Only forward fetches are allowed in Greenplum Database. It can be one of the following:

**NEXT**

Fetch the next row. This is the default if direction is omitted.

**FIRST**

Fetch the first row of the query (same as `ABSOLUTE 1`). Only allowed if it is the first `FETCH` operation using this cursor.

**LAST**

Fetch the last row of the query (same as `ABSOLUTE -1`).

**ABSOLUTE count**

Fetch the specified row of the query. Position after last row if count is out of range. Only allowed if the row specified by `count` moves the cursor position forward.

**RELATIVE count**

Fetch the specified row of the query `count` rows ahead of the current cursor position. RELATIVE 0 re-fetches the current row, if any. Only allowed if `count` moves the cursor position forward.

**count**

Fetch the next `count` number of rows (same as `FORWARD count`).

**ALL**

Fetch all remaining rows (same as `FORWARD ALL`).

**FORWARD**

Fetch the next row (same as `NEXT`).

**FORWARD count**

Fetch the next `count` number of rows. FORWARD 0 re-fetches the current row.

**FORWARD ALL**

Fetch all remaining rows.

**cursoname**

The name of an open cursor.
Notes
Greenplum Database does not support scrollable cursors, so you can only use `FETCH` to move the cursor position forward.

`ABSOLUTE` fetches are not any faster than navigating to the desired row with a relative move: the underlying implementation must traverse all the intermediate rows anyway.

Updating data via a cursor is currently not supported by Greenplum Database.

`DECLARE` is used to define a cursor. Use `MOVE` to change cursor position without retrieving data.

Examples
-- Start the transaction:

```
BEGIN;
```

-- Set up a cursor:

```
DECLARE mycursor CURSOR FOR SELECT * FROM films;
```

-- Fetch the first 5 rows in the cursor `mycursor`:

```
FETCH FORWARD 5 FROM mycursor;
```

```
code   |          title          | did | date_prod  |   kind   |  len  
-------+-------------------------+-----+------------+----------+------
BL101  | The Third Man           | 101 | 1949-12-23 | Drama    | 01:44
BL102  | The African Queen       | 101 | 1951-08-11 | Romantic | 01:43
JL201  | Une Femme est une Femme | 102 | 1961-03-12 | Romantic | 01:25
P_301  | Vertigo                 | 103 | 1958-11-14 | Action   | 02:08
P_302  | Becket                  | 103 | 1964-02-03 | Drama    | 02:28
```

-- Close the cursor and end the transaction:

```
CLOSE mycursor;
COMMIT;
```

Compatibility
SQL standard allows cursors only in embedded SQL and in modules. Greenplum Database permits cursors to be used interactively.

The variant of `FETCH` described here returns the data as if it were a `SELECT` result rather than placing it in host variables. Other than this point, `FETCH` is fully upward-compatible with the SQL standard.

The `FETCH` forms involving `FORWARD`, as well as the forms `FETCH count` and `FETCH ALL`, in which `FORWARD` is implicit, are Greenplum Database extensions. `BACKWARD` is not supported.

The SQL standard allows only `FROM` preceding the cursor name; the option to use `IN` is an extension.
See Also

DECLARE, CLOSE, MOVE
**GRANT**

Defines access privileges.

---

**Synopsis**

```sql
GRANT { {SELECT | INSERT | UPDATE | DELETE | REFERENCES | TRIGGER} [, ...] | ALL [PRIVILEGES] } 
    ON [TABLE] tablename [, ...] 
    TO { rolename | PUBLIC} [, ...] [WITH GRANT OPTION]
```

```sql
GRANT { {USAGE | SELECT | UPDATE} [, ...] | ALL [PRIVILEGES] } 
    ON SEQUENCE sequencename [, ...] 
    TO { rolename | PUBLIC } [, ...] [WITH GRANT OPTION]
```

```sql
GRANT { {CREATE | CONNECT | TEMPORARY | TEMP} [, ...] | ALL [PRIVILEGES] } 
    ON DATABASE dbname [, ...] 
    TO { rolename | PUBLIC} [, ...] [WITH GRANT OPTION]
```

```sql
GRANT { EXECUTE | ALL [PRIVILEGES] } 
    ON FUNCTION funcname ( [ [argmode] [argname] argtype [, ...] ] ) [, ...] 
    TO { rolename | PUBLIC} [, ...] [WITH GRANT OPTION]
```

```sql
GRANT { USAGE | ALL [PRIVILEGES] } 
    ON LANGUAGE langname [, ...] 
    TO { rolename | PUBLIC} [, ...] [WITH GRANT OPTION]
```

```sql
GRANT { {CREATE | USAGE} [, ...] | ALL [PRIVILEGES] } 
    ON SCHEMA schemaname [, ...] 
    TO { rolename | PUBLIC} [, ...] [WITH GRANT OPTION]
```

```sql
GRANT { CREATE | ALL [PRIVILEGES] } 
    ON TABLESPACE tablespacename [, ...] 
    TO { rolename | PUBLIC} [, ...] [WITH GRANT OPTION]
```

```sql
GRANT parent_role [, ...] 
    TO member_role [, ...] [WITH ADMIN OPTION]
```

---

**Description**

The **GRANT** command has two basic variants: one that grants privileges on a database object (table, view, sequence, database, function, procedural language, schema, or tablespace), and one that grants membership in a role.

**GRANT on Database Objects**

This variant of the **GRANT** command gives specific privileges on a database object to one or more roles. These privileges are added to those already granted, if any.
The key word PUBLIC indicates that the privileges are to be granted to all roles, including those that may be created later. PUBLIC may be thought of as an implicitly defined group-level role that always includes all roles. Any particular role will have the sum of privileges granted directly to it, privileges granted to any role it is presently a member of, and privileges granted to PUBLIC.

IF WITH GRANT OPTION is specified, the recipient of the privilege may in turn grant it to others. Without a grant option, the recipient cannot do that. Grant options cannot be granted to PUBLIC.

There is no need to grant privileges to the owner of an object (usually the role that created it), as the owner has all privileges by default. The right to drop an object, or to alter its definition in any way is not described by a grantable privilege; it is inherent in the owner, and cannot be granted or revoked. The owner implicitly has all grant options for the object, too.

Depending on the type of object, the initial default privileges may include granting some privileges to PUBLIC. The default is no public access for tables, schemas, and tablespaces; CONNECT privilege and TEMP table creation privilege for databases; EXECUTE privilege for functions; and USAGE privilege for languages. The object owner may of course revoke these privileges.

**Grant on Roles**

This variant of the GRANT command grants membership in a role to one or more other roles. Membership in a role is significant because it conveys the privileges granted to a role to each of its members.

IF WITH ADMIN OPTION is specified, the member may in turn grant membership in the role to others, and revoke membership in the role as well. Database superusers can grant or revoke membership in any role to anyone. Roles having CREATEROLE privilege can grant or revoke membership in any role that is not a superuser.

Unlike the case with privileges, membership in a role cannot be granted to PUBLIC.

**Parameters**

**SELECT**

Allows SELECT from any column of the specified table, view, or sequence. Also allows the use of COPY TO. For sequences, this privilege also allows the use of the currval function.

**INSERT**

Allows INSERT of a new row into the specified table. Also allows COPY FROM.

**UPDATE**

Allows UPDATE of any column of the specified table. SELECT ... FOR UPDATE and SELECT ... FOR SHARE also require this privilege (as well as the SELECT privilege). For sequences, this privilege allows the use of the nextval and setval functions.
DELETE
   Allows DELETE of a row from the specified table.

REFERENCES
   This keyword is accepted, although foreign key constraints are currently not supported in Greenplum Database. To create a foreign key constraint, it is necessary to have this privilege on both the referencing and referenced tables.

TRIGGER
   Allows the creation of a trigger on the specified table.

CREATE
   For databases, allows new schemas to be created within the database.
   For schemas, allows new objects to be created within the schema. To rename an existing object, you must own the object and have this privilege for the containing schema.
   For tablespaces, allows tables and indexes to be created within the tablespace, and allows databases to be created that have the tablespace as their default tablespace. (Note that revoking this privilege will not alter the placement of existing objects.)

CONNECT
   Allows the user to connect to the specified database. This privilege is checked at connection startup (in addition to checking any restrictions imposed by pg_hba.conf).

TEMPORARY TEMP
   Allows temporary tables to be created while using the database.

EXECUTE
   Allows the use of the specified function and the use of any operators that are implemented on top of the function. This is the only type of privilege that is applicable to functions. (This syntax works for aggregate functions, as well.)

USAGE
   For procedural languages, allows the use of the specified language for the creation of functions in that language. This is the only type of privilege that is applicable to procedural languages.
   For schemas, allows access to objects contained in the specified schema (assuming that the objects’ own privilege requirements are also met). Essentially this allows the grantee to look up objects within the schema.
   For sequences, this privilege allows the use of the currval and nextval functions.

ALL PRIVILEGES
   Grant all of the available privileges at once. The PRIVILEGES key word is optional in Greenplum Database, though it is required by strict SQL.
PUBLIC

A special group-level role that denotes that the privileges are to be granted to all roles, including those that may be created later.

WITH GRANT OPTION

The recipient of the privilege may in turn grant it to others.

WITH ADMIN OPTION

The member of a role may in turn grant membership in the role to others.

Notes

Database superusers can access all objects regardless of object privilege settings. One exception to this rule is view objects. Access to tables referenced in the view is determined by permissions of the view owner not the current user (even if the current user is a superuser).

If a superuser chooses to issue a GRANT or REVOKE command, the command is performed as though it were issued by the owner of the affected object. In particular, privileges granted via such a command will appear to have been granted by the object owner. For role membership, the membership appears to have been granted by the containing role itself.

GRANT and REVOKE can also be done by a role that is not the owner of the affected object, but is a member of the role that owns the object, or is a member of a role that holds privileges WITH GRANT OPTION on the object. In this case the privileges will be recorded as having been granted by the role that actually owns the object or holds the privileges WITH GRANT OPTION.

Granting permission on a table does not automatically extend permissions to any sequences used by the table, including sequences tied to SERIAL columns. Permissions on a sequence must be set separately.

Greenplum Database does not support granting or revoking privileges for individual columns of a table. One possible workaround is to create a view having just the desired columns and then grant privileges to that view.

Use psql’s \z meta-command to obtain information about existing privileges for an object.

Examples

Grant insert privilege to all roles on table mytable:

```
GRANT INSERT ON mytable TO PUBLIC;
```

Grant all available privileges to role sally on the view topten. Note that while the above will indeed grant all privileges if executed by a superuser or the owner of topten, when executed by someone else it will only grant those permissions for which the granting role has grant options.

```
GRANT ALL PRIVILEGES ON topten TO sally;
```

Grant membership in role admins to user joe:

```
GRANT TO PUBLIC;
GRANT admins TO joe;

**Compatibility**

The `PRIVILEGES` key word in is required in the SQL standard, but optional in Greenplum Database. The SQL standard does not support setting the privileges on more than one object per command.

Greenplum Database allows an object owner to revoke his own ordinary privileges: for example, a table owner can make the table read-only to himself by revoking his own `INSERT`, `UPDATE`, and `DELETE` privileges. This is not possible according to the SQL standard. Greenplum Database treats the owner’s privileges as having been granted by the owner to himself; therefore he can revoke them too. In the SQL standard, the owner’s privileges are granted by an assumed `system` entity.

The SQL standard allows setting privileges for individual columns within a table.

The SQL standard provides for a `USAGE` privilege on other kinds of objects: character sets, collations, translations, domains.

Privileges on databases, tablespaces, schemas, and languages are Greenplum Database extensions.

**See Also**

REVOKE
**INSERT**

Creates new rows in a table.

**Synopsis**

```sql
INSERT INTO table [( column [, ...] )]
 {DEFAULT VALUES | VALUES ( {expression | DEFAULT} [, ...] )
 [, ...] | query}
```

**Description**

**INSERT** inserts new rows into a table. One can insert one or more rows specified by value expressions, or zero or more rows resulting from a query.

The target column names may be listed in any order. If no list of column names is given at all, the default is the columns of the table in their declared order. The values supplied by the **VALUES** clause or query are associated with the explicit or implicit column list left-to-right.

Each column not present in the explicit or implicit column list will be filled with a default value, either its declared default value or null if there is no default.

If the expression for any column is not of the correct data type, automatic type conversion will be attempted.

You must have **INSERT** privilege on a table in order to insert into it.

**Outputs**

On successful completion, an **INSERT** command returns a command tag of the form:

```
INSERT oid count
```

The **count** is the number of rows inserted. If count is exactly one, and the target table has OIDs, then **oid** is the OID assigned to the inserted row. Otherwise **oid** is zero.

**Parameters**

**table**

The name (optionally schema-qualified) of an existing table.

**column**

The name of a column in table. The column name can be qualified with a subfield name or array subscript, if needed. (Inserting into only some fields of a composite column leaves the other fields null.)

**DEFAULT VALUES**

All columns will be filled with their default values.

**expression**

An expression or value to assign to the corresponding column.
**DEFAULT**

The corresponding column will be filled with its default value.

**query**

A query (SELECT statement) that supplies the rows to be inserted. Refer to the SELECT statement for a description of the syntax.

---

**Examples**

Insert a single row into table `films`:

```
INSERT INTO films VALUES ('UA502', 'Bananas', 105, '1971-07-13', 'Comedy', '82 minutes');
```

In this example, the `length` column is omitted and therefore it will have the default value:

```
INSERT INTO films (code, title, did, date_prod, kind) VALUES ('T_601', 'Yojimbo', 106, '1961-06-16', 'Drama');
```

This example uses the DEFAULT clause for the `date_prod` column rather than specifying a value:

```
INSERT INTO films VALUES ('UA502', 'Bananas', 105, DEFAULT, 'Comedy', '82 minutes');
```

To insert a row consisting entirely of default values:

```
INSERT INTO films DEFAULT VALUES;
```

To insert multiple rows using the multirow VALUES syntax:

```
INSERT INTO films (code, title, did, date_prod, kind) VALUES ('B6717', 'Tampopo', 110, '1985-02-10', 'Comedy'), ('HG120', 'The Dinner Game', 140, DEFAULT, 'Comedy');
```

This example inserts some rows into table `films` from a table `tmp_films` with the same column layout as `films`:

```
INSERT INTO films SELECT * FROM tmp_films WHERE date_prod < '2004-05-07';
```

---

**Compatibility**

`INSERT` conforms to the SQL standard. The case in which a column name list is omitted, but not all the columns are filled from the VALUES clause or query, is disallowed by the standard.

Possible limitations of the query clause are documented under SELECT.

---

**See Also**

COPY, SELECT, CREATE EXTERNAL TABLE
LOAD

Loads or reloads a shared library file.

**Synopsis**

`LOAD 'filename'`

**Description**

This command loads a shared library file into the Greenplum Database server address space. If the file had been loaded previously, it is first unloaded. This command is primarily useful to unload and reload a shared library file that has been changed since the server first loaded it. To make use of the shared library, function(s) in it need to be declared using the `CREATE FUNCTION` command.

The file name is specified in the same way as for shared library names in `CREATE FUNCTION`; in particular, one may rely on a search path and automatic addition of the system’s standard shared library file name extension.

Note that in Greenplum Database the shared library file (.so file) must reside in the same path location on every host in the Greenplum Database array (masters, segments, and mirrors).

Only database superusers can load shared library files.

**Parameters**

`filename`

The path and file name of a shared library file. This file must exist in the same location on all hosts in your Greenplum Database array.

**Examples**

Load a shared library file:

```
LOAD '/usr/local/greenplum-db/lib/myfuncs.so';
```

**Compatibility**

LOAD is a Greenplum Database extension.

**See Also**

`CREATE FUNCTION`
**LOCK**

Locks a table.

**Synopsis**

```
LOCK [TABLE] name [, ...] [IN lockmode MODE] [NOWAIT]
```

where `lockmode` is one of:

- ACCESS
- SHARE
- ROW SHARE
- ROW EXCLUSIVE
- SHARE UPDATE
- EXCLUSIVE
- SHARE ROW EXCLUSIVE
- ACCESS EXCLUSIVE

**Description**

`LOCK TABLE` obtains a table-level lock, waiting if necessary for any conflicting locks to be released. If `NOWAIT` is specified, `LOCK TABLE` does not wait to acquire the desired lock: if it cannot be acquired immediately, the command is aborted and an error is emitted. Once obtained, the lock is held for the remainder of the current transaction. There is no `UNLOCK TABLE` command; locks are always released at transaction end.

When acquiring locks automatically for commands that reference tables, Greenplum Database always uses the least restrictive lock mode possible. `LOCK TABLE` provides for cases when you might need more restrictive locking. For example, suppose an application runs a transaction at the Read Committed isolation level and needs to ensure that data in a table remains stable for the duration of the transaction. To achieve this you could obtain `SHARE` lock mode over the table before querying. This will prevent concurrent data changes and ensure subsequent reads of the table see a stable view of committed data, because `SHARE` lock mode conflicts with the `ROW EXCLUSIVE` lock acquired by writers, and your `LOCK TABLE name IN SHARE MODE` statement will wait until any concurrent holders of `ROW EXCLUSIVE` mode locks commit or roll back. Thus, once you obtain the lock, there are no uncommitted writes outstanding; furthermore none can begin until you release the lock.

To achieve a similar effect when running a transaction at the Serializable isolation level, you have to execute the `LOCK TABLE` statement before executing any `SELECT` or data modification statement. A serializable transaction’s view of data will be frozen when its first `SELECT` or data modification statement begins. A `LOCK TABLE` later in the transaction will still prevent concurrent writes — but it won’t ensure that what the transaction reads corresponds to the latest committed values.

If a transaction of this sort is going to change the data in the table, then it should use `SHARE ROW EXCLUSIVE` lock mode instead of `SHARE` mode. This ensures that only one transaction of this type runs at a time. Without this, a deadlock is possible: two transactions might both acquire `SHARE` mode, and then be unable to also acquire `ROW EXCLUSIVE` mode to actually perform their updates. Note that a transaction’s own locks never conflict, so a transaction can acquire `ROW EXCLUSIVE` mode when it holds `SHARE` mode — but not if anyone else holds `SHARE` mode. To avoid deadlocks, make
sure all transactions acquire locks on the same objects in the same order, and if multiple lock modes are involved for a single object, then transactions should always acquire the most restrictive mode first.

**Parameters**

**name**

The name (optionally schema-qualified) of an existing table to lock.

If multiple tables are given, tables are locked one-by-one in the order specified in the `LOCK TABLE` command.

**lockmode**

The lock mode specifies which locks this lock conflicts with. If no lock mode is specified, then `ACCESS EXCLUSIVE`, the most restrictive mode, is used. Lock modes are as follows:

- **ACCESS SHARE** — Conflicts with the `ACCESS EXCLUSIVE` lock mode only. The commands `SELECT` and `ANALYZE` automatically acquire a lock of this mode on referenced tables. In general, any query that only reads a table and does not modify it will acquire this lock mode.

- **ROW SHARE** — Conflicts with the `EXCLUSIVE` and `ACCESS EXCLUSIVE` lock modes. The commands `SELECT FOR UPDATE` and `SELECT FOR SHARE` commands automatically acquire a lock of this mode on the target table(s) (in addition to `ACCESS SHARE` locks on any other tables that are referenced but not selected `FOR UPDATE/FOR SHARE`).

- **ROW EXCLUSIVE** — Conflicts with the `SHARE`, `SHARE ROW EXCLUSIVE`, `EXCLUSIVE`, and `ACCESS EXCLUSIVE` lock modes. The commands `INSERT` and `COPY` automatically acquire this lock mode on the target table (in addition to `ACCESS SHARE` locks on any other referenced tables).

- **SHARE UPDATE EXCLUSIVE** — Conflicts with the `SHARE UPDATE EXCLUSIVE`, `SHARE`, `SHARE ROW EXCLUSIVE`, `EXCLUSIVE`, and `ACCESS EXCLUSIVE` lock modes. This mode protects a table against concurrent schema changes and `VACUUM` runs. Acquired automatically by `VACUUM` (without `FULL`).

- **SHARE** — Conflicts with the `ROW EXCLUSIVE`, `SHARE UPDATE EXCLUSIVE`, `SHARE ROW EXCLUSIVE`, `EXCLUSIVE`, and `ACCESS EXCLUSIVE` lock modes. This mode protects a table against concurrent data changes. Acquired automatically by `CREATE INDEX`.

- **SHARE ROW EXCLUSIVE** — Conflicts with the `ROW EXCLUSIVE`, `SHARE UPDATE EXCLUSIVE`, `SHARE ROW EXCLUSIVE`, `EXCLUSIVE`, and `ACCESS EXCLUSIVE` lock modes. This lock mode is not automatically acquired by any Greenplum Database command.
• **EXCLUSIVE** — Conflicts with the row share, row exclusive, share update exclusive, share, share row exclusive, exclusive, and access exclusive lock modes. This mode allows only concurrent access share locks, i.e., only reads from the table can proceed in parallel with a transaction holding this lock mode. This lock mode is automatically acquired for update and delete in Greenplum Database (which is more restrictive locking than in regular PostgreSQL).

• **ACCESS EXCLUSIVE** — Conflicts with locks of all modes (access share, row share, row exclusive, share update exclusive, share, share row exclusive, exclusive, and access exclusive). This mode guarantees that the holder is the only transaction accessing the table in any way. Acquired automatically by the alter table, drop table, reindex, cluster, and vacuum full commands. This is also the default lock mode for lock table statements that do not specify a mode explicitly.

**NOWAIT**

Specifies that lock table should not wait for any conflicting locks to be released: if the specified lock(s) cannot be acquired immediately without waiting, the transaction is aborted.

**Notes**

Lock table ... in access share mode requires select privileges on the target table. All other forms of lock require update and/or delete privileges.

Lock table is useful only inside a transaction block (begin/commit pair), since the lock is dropped as soon as the transaction ends. A lock table command appearing outside any transaction block forms a self-contained transaction, so the lock will be dropped as soon as it is obtained.

Lock table only deals with table-level locks, and so the mode names involving row are all misnomers. These mode names should generally be read as indicating the intention of the user to acquire row-level locks within the locked table. Also, row exclusive mode is a sharable table lock. Keep in mind that all the lock modes have identical semantics so far as lock table is concerned, differing only in the rules about which modes conflict with which. For information on how to acquire an actual row-level lock, see the for update/for share clause in the select reference documentation.

**Examples**

Obtain a share lock on the films table when going to perform inserts into the films_user_comments table:

```
BEGIN WORK;
LOCK TABLE films IN SHARE MODE;
SELECT id FROM films
WHERE name = 'Star Wars: Episode I - The Phantom Menace';
-- Do ROLLBACK if record was not returned
INSERT INTO films_user_comments VALUES
```
( _id_ , 'GREAT! I was waiting for it for so long!');
COMMIT WORK;

Take a SHARE ROW EXCLUSIVE lock on a table when performing a delete operation:
BEGIN WORK;
LOCK TABLE films IN SHARE ROW EXCLUSIVE MODE;
DELETE FROM films_user_comments WHERE id IN
   (SELECT id FROM films WHERE rating < 5);
DELETE FROM films WHERE rating < 5;
COMMIT WORK;

Compatibility
There is no LOCK TABLE in the SQL standard, which instead uses SET TRANSACTION
to specify concurrency levels on transactions. Greenplum Database supports that too.

Except for ACCESS SHARE, ACCESS EXCLUSIVE, and SHARE UPDATE EXCLUSIVE
lock modes, the Greenplum Database lock modes and the LOCK TABLE syntax are
compatible with those present in Oracle.

See Also
BEGIN, SET TRANSACTION, SELECT
MOVE

Positions a cursor.

Synopsis

MOVE [ forward_direction {FROM | IN} ] cursorname

where direction can be empty or one of:

NEXT
FIRST
LAST
ABSOLUTE count
RELATIVE count
count
ALL
FORWARD
FORWARD count
FORWARD ALL

Description

MOVE repositions a cursor without retrieving any data. MOVE works exactly like the FETCH command, except it only positions the cursor and does not return rows.

Note that it is not possible to move a cursor position backwards in Greenplum Database, since scrollable cursors are not supported. You can only move a cursor forward in position using MOVE.

Outputs

On successful completion, a MOVE command returns a command tag of the form

MOVE count

The count is the number of rows that a FETCH command with the same parameters would have returned (possibly zero).

Parameters

forward_direction

See FETCH for more information.

cursorname

The name of an open cursor.

Examples

-- Start the transaction:

BEGIN;
-- Set up a cursor:
    DECLARE mycursor CURSOR FOR SELECT * FROM films;

-- Move forward 5 rows in the cursor mycursor:
    MOVE FORWARD 5 IN mycursor;
    MOVE 5

-- Fetch the next row after that (row 6):
    FETCH 1 FROM mycursor;

<table>
<thead>
<tr>
<th>code</th>
<th>title</th>
<th>did</th>
<th>date_prod</th>
<th>kind</th>
<th>len</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_303</td>
<td>48 Hrs</td>
<td>103</td>
<td>1982-10-22</td>
<td>Action</td>
<td>01:37</td>
</tr>
</tbody>
</table>

(1 row)

-- Close the cursor and end the transaction:
    CLOSE mycursor;
    COMMIT;

Compatibility
There is no MOVE statement in the SQL standard.

See Also
DECLARE, FETCH, CLOSE
**PREPARE**

Prepare a statement for execution.

**Synopsis**

PREPARE name [ (datatype [, ...] ) ] AS statement

**Description**

`PREPARE` creates a prepared statement. A prepared statement is a server-side object that can be used to optimize performance. When the `PREPARE` statement is executed, the specified statement is parsed, rewritten, and planned. When an `EXECUTE` command is subsequently issued, the prepared statement need only be executed. Thus, the parsing, rewriting, and planning stages are only performed once, instead of every time the statement is executed.

Prepared statements can take parameters: values that are substituted into the statement when it is executed. When creating the prepared statement, refer to parameters by position, using $1, $2, etc. A corresponding list of parameter data types can optionally be specified. When a parameter’s data type is not specified or is declared as unknown, the type is inferred from the context in which the parameter is used (if possible). When executing the statement, specify the actual values for these parameters in the `EXECUTE` statement.

Prepared statements only last for the duration of the current database session. When the session ends, the prepared statement is forgotten, so it must be recreated before being used again. This also means that a single prepared statement cannot be used by multiple simultaneous database clients; however, each client can create their own prepared statement to use. The prepared statement can be manually cleaned up using the `DEALLOCATE` command.

Prepared statements have the largest performance advantage when a single session is being used to execute a large number of similar statements. The performance difference will be particularly significant if the statements are complex to plan or rewrite, for example, if the query involves a join of many tables or requires the application of several rules. If the statement is relatively simple to plan and rewrite but relatively expensive to execute, the performance advantage of prepared statements will be less noticeable.

**Parameters**

`name`

An arbitrary name given to this particular prepared statement. It must be unique within a single session and is subsequently used to execute or deallocate a previously prepared statement.
**datatype**

The data type of a parameter to the prepared statement. If the data type of a particular parameter is unspecified or is specified as unknown, it will be inferred from the context in which the parameter is used. To refer to the parameters in the prepared statement itself, use $1, $2, etc.

**statement**

Any SELECT, INSERT, UPDATE, DELETE, or VALUES statement.

---

**Notes**

In some situations, the query plan produced for a prepared statement will be inferior to the query plan that would have been chosen if the statement had been submitted and executed normally. This is because when the statement is planned and the planner attempts to determine the optimal query plan, the actual values of any parameters specified in the statement are unavailable. Greenplum Database collects statistics on the distribution of data in the table, and can use constant values in a statement to make guesses about the likely result of executing the statement. Since this data is unavailable when planning prepared statements with parameters, the chosen plan may be suboptimal. To examine the query plan Greenplum Database has chosen for a prepared statement, use EXPLAIN.

For more information on query planning and the statistics collected by Greenplum Database for that purpose, see the ANALYZE documentation.

You can see all available prepared statements of a session by querying the `pg_prepared_statements` system view.

---

**Examples**

Create a prepared statement for an INSERT statement, and then execute it:

```sql
PREPARE fooplan (int, text, bool, numeric) AS INSERT INTO foo VALUES($1, $2, $3, $4);
EXECUTE fooplan(1, 'Hunter Valley', 't', 200.00);
```

Create a prepared statement for a SELECT statement, and then execute it. Note that the data type of the second parameter is not specified, so it is inferred from the context in which $2 is used:

```sql
PREPARE usrrptplan (int) AS SELECT * FROM users u, logs l WHERE u.usrid=$1 AND u.usrid=l.usrid AND l.date = $2;
EXECUTE usrrptplan(1, current_date);
```

---

**Compatibility**

The SQL standard includes a PREPARE statement, but it is only for use in embedded SQL. This version of the PREPARE statement also uses a somewhat different syntax.
See Also

EXECUTE, DEALLOCATE
REASSIGN OWNED

Changes the ownership of database objects owned by a database role.

Synopsis

REASSIGN OWNED BY old_role [, ...] TO new_role

Description

REASSIGN OWNED reassigns all the objects in the current database that are owned by old_role to new_role. Note that it does not change the ownership of the database itself.

Parameters

old_role

The name of a role. The ownership of all the objects in the current database owned by this role will be reassigned to new_role.

new_role

The name of the role that will be made the new owner of the affected objects.

Notes

REASSIGN OWNED is often used to prepare for the removal of one or more roles. Because REASSIGN OWNED only affects the objects in the current database, it is usually necessary to execute this command in each database that contains objects owned by a role that is to be removed.

The DROP OWNED command is an alternative that drops all the database objects owned by one or more roles.

The REASSIGN OWNED command does not affect the privileges granted to the old roles in objects that are not owned by them. Use DROP OWNED to revoke those privileges.

Examples

Reassign any database objects owned by the role named sally and bob to admin;

    REASSIGN OWNED BY sally, bob TO admin;

Compatibility

The REASSIGN OWNED statement is a Greenplum Database extension.

See Also

DROP OWNED, DROP ROLE
**REINDEX**

Rebuilds indexes.

**Synopsis**

```
REINDEX {INDEX | TABLE | DATABASE | SYSTEM} name
```

**Description**

REINDEX rebuilds an index using the data stored in the index’s table, replacing the old copy of the index. There are several scenarios in which to use REINDEX:

- An index has become corrupted, and no longer contains valid data. Although in theory this should never happen, in practice indexes may become corrupted due to software bugs or hardware failures. REINDEX provides a recovery method.

- An index has become bloated, that it is contains many empty or nearly-empty pages. This can occur with B-tree indexes in Greenplum Database under certain uncommon access patterns. REINDEX provides a way to reduce the space consumption of the index by writing a new version of the index without the dead pages.

- You have altered the fillfactor storage parameter for an index, and wish to ensure that the change has taken full effect.

**Parameters**

**INDEX**

Recreate the specified index.

**TABLE**

Recreate all indexes of the specified table. If the table has a secondary TOAST table, that is reindexed as well.

**DATABASE**

Recreate all indexes within the current database. Indexes on shared system catalogs are skipped. This form of REINDEX cannot be executed inside a transaction block.

**SYSTEM**

Recreate all indexes on system catalogs within the current database. Indexes on user tables are not processed. Also, indexes on shared (global) system catalogs are skipped. This form of REINDEX cannot be executed inside a transaction block.

**name**

The name of the specific index, table, or database to be reindexed. Index and table names may be schema-qualified. Presently, REINDEX DATABASE and REINDEX SYSTEM can only reindex the current database, so their parameter must match the current database’s name.
Notes

`REINDEX` is similar to a drop and recreate of the index in that the index contents are rebuilt from scratch. However, the locking considerations are rather different. `REINDEX` locks out writes but not reads of the index’s parent table. It also takes an exclusive lock on the specific index being processed, which will block reads that attempt to use that index. In contrast, `DROP INDEX` momentarily takes exclusive lock on the parent table, blocking both writes and reads. The subsequent `CREATE INDEX` locks out writes but not reads; since the index is not there, no read will attempt to use it, meaning that there will be no blocking but reads may be forced into expensive sequential scans. Another important point is that the drop/create approach invalidates any cached query plans that use the index, while `REINDEX` does not.

Reindexing a single index or table requires being the owner of that index or table. Reindexing a database requires being the owner of the database (note that the owner can therefore rebuild indexes of tables owned by other users). Of course, superusers can always reindex anything.

If you suspect that shared global system catalog indexes are corrupted, they can only be reindexed in Greenplum utility mode. The typical symptom of a corrupt shared index is “index is not a btree” errors, or else the server crashes immediately at startup due to reliance on the corrupted indexes. Contact Greenplum Customer Support for assistance in this situation.

Examples

Rebuild a single index:

```
REINDEX INDEX my_index;
```

Rebuild all the indexes on the table `my_table`:

```
REINDEX TABLE my_table;
```

Compatibility

There is no `REINDEX` command in the SQL standard.

See Also

`CREATE INDEX`, `DROP INDEX`, `VACUUM`
RELEASE SAVEPOINT

Destroys a previously defined savepoint.

Synopsis

RELEASE [SAVEPOINT] savepoint_name

Description

RELEASE SAVEPOINT destroys a savepoint previously defined in the current transaction.

Destroying a savepoint makes it unavailable as a rollback point, but it has no other user visible behavior. It does not undo the effects of commands executed after the savepoint was established. (To do that, see ROLLBACK TO SAVEPOINT.) Destroying a savepoint when it is no longer needed may allow the system to reclaim some resources earlier than transaction end.

RELEASE SAVEPOINT also destroys all savepoints that were established after the named savepoint was established.

Parameters

savepoint_name

The name of the savepoint to destroy.

Examples

To establish and later destroy a savepoint:

BEGIN;
    INSERT INTO table1 VALUES (3);
    SAVEPOINT my_savepoint;
    INSERT INTO table1 VALUES (4);
    RELEASE SAVEPOINT my_savepoint;
    COMMIT;

The above transaction will insert both 3 and 4.

Compatibility

This command conforms to the SQL standard. The standard specifies that the key word SAVEPOINT is mandatory, but Greenplum Database allows it to be omitted.

See Also

BEGIN, SAVEPOINT, ROLLBACK TO SAVEPOINT, COMMIT
RESET

Restores the value of a system configuration parameter to the default value.

Synopsis

RESET configuration_parameter
RESET ALL

Description

RESET restores system configuration parameters to their default values. RESET is an alternative spelling for SET configuration_parameter TO DEFAULT.

The default value is defined as the value that the parameter would have had, had no SET ever been issued for it in the current session. The actual source of this value might be a compiled-in default, the master postgresql.conf configuration file, command-line options, or per-database or per-user default settings. See “Server Configuration Parameters” on page 770.

Parameters

configuration_parameter

The name of a system configuration parameter. See “Server Configuration Parameters” on page 770 for details.

ALL

Resets all settable configuration parameters to their default values.

Examples

Set the work_mem configuration parameter to its default value:

    RESET work_mem;

Compatibility

RESET is a Greenplum Database extension.

See Also

SET
REVOKE

Removes access privileges.

Synopsis

REVOKE [GRANT OPTION FOR] { {SELECT | INSERT | UPDATE | DELETE
   | REFERENCES | TRIGGER} [, . . .] | ALL [PRIVILEGES] }
ON [TABLE] tablename [, . . .]
FROM {rolename | PUBLIC} [, . . .]
[CASCADE | RESTRICT]

REVOKE [GRANT OPTION FOR] { {USAGE | SELECT | UPDATE} [, . . .]
   | ALL [PRIVILEGES] }
ON SEQUENCE sequencename [, . . .]
FROM {rolename | PUBLIC} [, . . .]
[CASCADE | RESTRICT]

REVOKE [GRANT OPTION FOR] { {CREATE | CONNECT
   | TEMPORARY | TEMP} [, . . .] | ALL [PRIVILEGES] }
ON DATABASE dbname [, . . .]
FROM {rolename | PUBLIC} [, . . .]
[CASCADE | RESTRICT]

REVOKE [GRANT OPTION FOR] {EXECUTE | ALL [PRIVILEGES]}
ON FUNCTION funcname ( [ [argmode] [argname] argtype
   [, . . .] ] ) [, . . .]
FROM {rolename | PUBLIC} [, . . .]
[CASCADE | RESTRICT]

REVOKE [GRANT OPTION FOR] { USAGE | ALL [PRIVILEGES]}
ON LANGUAGE langname [, . . .]
FROM {rolename | PUBLIC} [, . . .]
[CASCADE | RESTRICT]

REVOKE [GRANT OPTION FOR] { {CREATE | USAGE} [, . . .]
   | ALL [PRIVILEGES] }
ON SCHEMA schemaname [, . . .]
FROM {rolename | PUBLIC} [, . . .]
[CASCADE | RESTRICT]

REVOKE [GRANT OPTION FOR] { CREATE | ALL [PRIVILEGES] }
ON TABLESPACE tablespacename [, . . .]
FROM {rolename | PUBLIC} [, . . .]
[CASCADE | RESTRICT]

REVOKE [ADMIN OPTION FOR] parent_role [, . . .]
FROM member_role [, . . .]
[CASCADE | RESTRICT]
Description

The `REVOKE` command revokes previously granted privileges from one or more roles. The key word `PUBLIC` refers to the implicitly defined group of all roles.

See the description of the `GRANT` command for the meaning of the privilege types.

Note that any particular role will have the sum of privileges granted directly to it, privileges granted to any role it is presently a member of, and privileges granted to `PUBLIC`. Thus, for example, revoking `SELECT` privilege from `PUBLIC` does not necessarily mean that all roles have lost `SELECT` privilege on the object: those who have it granted directly or via another role will still have it.

If `GRANT OPTION FOR` is specified, only the grant option for the privilege is revoked, not the privilege itself. Otherwise, both the privilege and the grant option are revoked.

If a role holds a privilege with grant option and has granted it to other roles then the privileges held by those other roles are called dependent privileges. If the privilege or the grant option held by the first role is being revoked and dependent privileges exist, those dependent privileges are also revoked if `CASCADE` is specified, else the revoke action will fail. This recursive revocation only affects privileges that were granted through a chain of roles that is traceable to the role that is the subject of this `REVOKE` command. Thus, the affected roles may effectively keep the privilege if it was also granted through other roles.

When revoking membership in a role, `GRANT OPTION` is instead called `ADMIN OPTION`, but the behavior is similar.

Parameters

See `GRANT`.

Examples

- Revoke insert privilege for the public on table `films`:
  ```sql
  REVOKE INSERT ON films FROM PUBLIC;
  ```

- Revoke all privileges from role `sally` on view `topten`. Note that this actually means revoke all privileges that the current role granted (if not a superuser).
  ```sql
  REVOKE ALL PRIVILEGES ON topten FROM sally;
  ```

- Revoke membership in role `admins` from user `joe`:
  ```sql
  REVOKE admins FROM joe;
  ```

Compatibility

The compatibility notes of the `GRANT` command also apply to `REVOKE`.

One of `RESTRICT` or `CASCADE` is required according to the standard, but Greenplum Database assumes `RESTRICT` by default.
See Also

GRANT
ROLLBACK

Aborts the current transaction.

Synopsis

ROLLBACK [WORK | TRANSACTION]

Description

ROLLBACK rolls back the current transaction and causes all the updates made by the transaction to be discarded.

Parameters

WORK
TRANSACTION

Optional key words. They have no effect.

Notes

Use COMMIT to successfully end the current transaction.

Issuing ROLLBACK when not inside a transaction does no harm, but it will provoke a warning message.

Examples

To discard all changes made in the current transaction:

ROLLBACK;

Compatibility

The SQL standard only specifies the two forms ROLLBACK and ROLLBACK WORK. Otherwise, this command is fully conforming.

See Also

BEGIN, COMMIT, SAVEPOINT, ROLLBACK TO SAVEPOINT
ROLLBACK TO SAVEPOINT

Rolls back the current transaction to a savepoint.

Synopsis

ROLLBACK [WORK | TRANSACTION] TO [SAVEPOINT] savepoint_name

Description

This command will roll back all commands that were executed after the savepoint was established. The savepoint remains valid and can be rolled back to again later, if needed.

ROLLBACK TO SAVEPOINT implicitly destroys all savepoints that were established after the named savepoint.

Parameters

WORK

TRANSACTION

Optional key words. They have no effect.

savepoint_name

The name of a savepoint to roll back to.

Notes

Use RELEASE SAVEPOINT to destroy a savepoint without discarding the effects of commands executed after it was established.

Specifying a savepoint name that has not been established is an error.

Cursors have somewhat non-transactional behavior with respect to savepoints. Any cursor that is opened inside a savepoint will be closed when the savepoint is rolled back. If a previously opened cursor is affected by a FETCH command inside a savepoint that is later rolled back, the cursor position remains at the position that FETCH left it pointing to (that is, FETCH is not rolled back). Closing a cursor is not undone by rolling back, either. A cursor whose execution causes a transaction to abort is put in a can’t-execute state, so while the transaction can be restored using ROLLBACK TO SAVEPOINT, the cursor can no longer be used.

Examples

To undo the effects of the commands executed after my_savepoint was established:

ROLLBACK TO SAVEPOINT my_savepoint;

Cursor positions are not affected by a savepoint rollback:

BEGIN;
    DECLARE foo CURSOR FOR SELECT 1 UNION SELECT 2;
SAVEPOINT foo;
FETCH 1 FROM foo;
column
---------
 1
ROLLBACK TO SAVEPOINT foo;
FETCH 1 FROM foo;
column
---------
 2
COMMIT;

**Compatibility**

The SQL standard specifies that the key word `SAVEPOINT` is mandatory, but Greenplum Database (and Oracle) allow it to be omitted. SQL allows only `WORK`, not `TRANSACTION`, as a noise word after `ROLLBACK`. Also, SQL has an optional clause `AND [NO] CHAIN` which is not currently supported by Greenplum Database. Otherwise, this command conforms to the SQL standard.

**See Also**

`BEGIN`, `COMMIT`, `SAVEPOINT`, `RELEASE SAVEPOINT`, `ROLLBACK`
SAVEPOINT

Defines a new savepoint within the current transaction.

Synopsis

SAVEPOINT savepoint_name

Description

SAVEPOINT establishes a new savepoint within the current transaction.

A savepoint is a special mark inside a transaction that allows all commands that are executed after it was established to be rolled back, restoring the transaction state to what it was at the time of the savepoint.

Parameters

savepoint_name

The name of the new savepoint.

Notes

Use ROLLBACK TO SAVEPOINT to rollback to a savepoint. Use RELEASE SAVEPOINT to destroy a savepoint, keeping the effects of commands executed after it was established.

Savepoints can only be established when inside a transaction block. There can be multiple savepoints defined within a transaction.

Examples

To establish a savepoint and later undo the effects of all commands executed after it was established:

```
BEGIN;
    INSERT INTO table1 VALUES (1);
    SAVEPOINT my_savepoint;
    INSERT INTO table1 VALUES (2);
    ROLLBACK TO SAVEPOINT my_savepoint;
    INSERT INTO table1 VALUES (3);
COMMIT;
```

The above transaction will insert the values 1 and 3, but not 2.

To establish and later destroy a savepoint:

```
BEGIN;
    INSERT INTO table1 VALUES (3);
```

SAVEPOINT my_savepoint;
INSERT INTO table1 VALUES (4);
RELEASE SAVEPOINT my_savepoint;
COMMIT;

The above transaction will insert both 3 and 4.

**Compatibility**

SQL requires a savepoint to be destroyed automatically when another savepoint with the same name is established. In Greenplum Database, the old savepoint is kept, though only the more recent one will be used when rolling back or releasing. (Releasing the newer savepoint will cause the older one to again become accessible to **ROLLBACK TO SAVEPOINT** and **RELEASE SAVEPOINT**.) Otherwise, **SAVEPOINT** is fully SQL conforming.

**See Also**

**BEGIN**, **COMMIT**, **ROLLBACK**, **RELEASE SAVEPOINT**, **ROLLBACK TO SAVEPOINT**
SELECT

Retrieves rows from a table or view.

Synopsis

```
SELECT [ALL | DISTINCT [ON (expression [, ...])] ]
   * | expression [[AS] output_name] [, ...]
   [FROM from_item [, ...]]
   [WHERE condition]
   [GROUP BY grouping_element [, ...]]
   [HAVING condition [, ...]]
   [WINDOW window_name AS (window_specification)]
   [{UNION | INTERSECT | EXCEPT} [ALL] select]
   [ORDER BY expression [ASC | DESC | USING operator] [, ...]]
   [LIMIT {count | ALL}]
   [OFFSET start]
   [FOR {UPDATE | SHARE} {OF table_name [, ...]} [NOWAIT] [ ...]]
```

where grouping_element can be one of:

- ()
- expression
- ROLLUP (expression [, ...])
- CUBE (expression [, ...])
- GROUPING SETS ((grouping_element [, ...]))

where window_specification can be:

```
[window_name]
[PARTITION BY expression [, ...]]
[ORDER BY expression [ASC | DESC | USING operator] [, ...]
   [{RANGE | ROWS}
    { UNBOUNDED PRECEDING
    | expression PRECEDING
    | CURRENT ROW
    | BETWEEN window_frame_bound AND window_frame_bound }
    ]]
```

where window_frame_bound can be one of:

- UNBOUNDED PRECEDING
- expression PRECEDING
- CURRENT ROW
- expression FOLLOWING
- UNBOUNDED FOLLOWING

where from_item can be one of:

```
[ONLY] table_name [[AS] alias [( column_alias [, ...] )]]
(select) [AS] alias [( column_alias [, ...] )]
function_name ( [argument [, ...]] ) [AS] alias
   [( column_alias [, ...]
     | column_definition [, ...] )]
function_name ( [argument [, ...]] ) AS
   ( column_definition [, ...] )
```
**from_item [NATURAL] join_type from_item**

\[
\text{[ON join_condition | USING (join_column [, ...])]}
\]

**Description**

`SELECT` retrieves rows from zero or more tables. The general processing of `SELECT` is as follows:

1. All elements in the `FROM` list are computed. (Each element in the `FROM` list is a real or virtual table.) If more than one element is specified in the `FROM` list, they are cross-joined together.

2. If the `WHERE` clause is specified, all rows that do not satisfy the condition are eliminated from the output.

3. If the `GROUP BY` clause is specified, the output is divided into groups of rows that match on one or more of the defined grouping elements. If the `HAVING` clause is present, it eliminates groups that do not satisfy the given condition.

4. If a window expression is specified (and optional `WINDOW` clause), the output is organized according to the positional (row) or value-based (range) window frame.

5. `DISTINCT` eliminates duplicate rows from the result. `DISTINCT ON` eliminates rows that match on all the specified expressions. `ALL` (the default) will return all candidate rows, including duplicates.

6. The actual output rows are computed using the `SELECT` output expressions for each selected row.

7. Using the operators `UNION`, `INTERSECT`, and `EXCEPT`, the output of more than one `SELECT` statement can be combined to form a single result set. The `UNION` operator returns all rows that are in one or both of the result sets. The `INTERSECT` operator returns all rows that are strictly in both result sets. The `EXCEPT` operator returns the rows that are in the first result set but not in the second. In all three cases, duplicate rows are eliminated unless `ALL` is specified.

8. If the `ORDER BY` clause is specified, the returned rows are sorted in the specified order. If `ORDER BY` is not given, the rows are returned in whatever order the system finds fastest to produce.

9. If the `LIMIT` or `OFFSET` clause is specified, the `SELECT` statement only returns a subset of the result rows.

10. If `FOR UPDATE` or `FOR SHARE` is specified, the `SELECT` statement locks the entire table against concurrent updates.

You must have `SELECT` privilege on a table to read its values. The use of `FOR UPDATE` or `FOR SHARE` requires `UPDATE` privilege as well.
Parameters

The SELECT List

The SELECT list (between the key words SELECT and FROM) specifies expressions that form the output rows of the SELECT statement. The expressions can (and usually do) refer to columns computed in the FROM clause.

Using the clause [AS] output_name, another name can be specified for an output column. This name is primarily used to label the column for display. It can also be used to refer to the column’s value in ORDER BY and GROUP BY clauses, but not in the WHERE or HAVING clauses; there you must write out the expression instead. The AS keyword is optional in most cases (such as when declaring an alias for column names, constants, function calls, and simple unary operator expressions). In cases where the declared alias is a reserved SQL keyword, the output_name must be enclosed in double quotes to avoid ambiguity.

An expression in the SELECT list can be a constant value, a column reference, an operator invocation, a function call, an aggregate expression, a window expression, a scalar subquery, and so on. See “SQL Value Expressions” on page 101 for more information. There are a number of constructs that can be classified as an expression but do not follow any general syntax rules. These generally have the semantics of a function or operator and are explained in “Function Calls” on page 104.

Instead of an expression, * can be written in the output list as a shorthand for all the columns of the selected rows. Also, one can write table_name.* as a shorthand for the columns coming from just that table.

The FROM Clause

The FROM clause specifies one or more source tables for the SELECT. If multiple sources are specified, the result is the Cartesian product (cross join) of all the sources. But usually qualification conditions are added to restrict the returned rows to a small subset of the Cartesian product. The FROM clause can contain the following elements:

**table_name**

The name (optionally schema-qualified) of an existing table or view. If ONLY is specified, only that table is scanned. If ONLY is not specified, the table and all its descendant tables (if any) are scanned.

**alias**

A substitute name for the FROM item containing the alias. An alias is used for brevity or to eliminate ambiguity for self-joins (where the same table is scanned multiple times). When an alias is provided, it completely hides the actual name of the table or function; for example given FROM foo AS f, the remainder of the SELECT must refer to this FROM item as f not foo. If an alias is written, a column alias list can also be written to provide substitute names for one or more columns of the table.

**select**

A sub-SELECT can appear in the FROM clause. This acts as though its output were created as a temporary table for the duration of this single SELECT command. Note that the sub-SELECT must be surrounded by parentheses, and an alias must be
provided for it. A `VALUES` command can also be used here. See “Limited Correlated Subquery Syntax” on page 529 for limitations of using correlated sub-selects in Greenplum Database.

**function_name**

Function calls can appear in the `FROM` clause. (This is especially useful for functions that return result sets, but any function can be used.) This acts as though its output were created as a temporary table for the duration of this single `SELECT` command. An alias may also be used. If an alias is written, a column alias list can also be written to provide substitute names for one or more attributes of the function’s composite return type. If the function has been defined as returning the record data type, then an alias or the key word AS must be present, followed by a column definition list in the form `( column_name data_type [, ... ] )`. The column definition list must match the actual number and types of columns returned by the function.

**join_type**

One of:

- `[INNER] JOIN`
- `LEFT [OUTER] JOIN`
- `RIGHT [OUTER] JOIN`
- `FULL [OUTER] JOIN`
- `CROSS JOIN`

For the `INNER` and `OUTER` join types, a join condition must be specified, namely exactly one of `NATURAL`, `ON join_condition`, or `USING (join_column [, ...])`. See below for the meaning. For `CROSS JOIN`, none of these clauses may appear.

A `JOIN` clause combines two `FROM` items. Use parentheses if necessary to determine the order of nesting. In the absence of parentheses, `JOINs` nest left-to-right. In any case `JOIN` binds more tightly than the commas separating `FROM` items.

`CROSS JOIN` and `INNER JOIN` produce a simple Cartesian product, the same result as you get from listing the two items at the top level of `FROM`, but restricted by the join condition (if any). `CROSS JOIN` is equivalent to `INNER JOIN ON (TRUE)`, that is, no rows are removed by qualification. These join types are just a notational convenience, since they do nothing you could not do with plain `FROM` and `WHERE`.

`LEFT OUTER JOIN` returns all rows in the qualified Cartesian product (i.e., all combined rows that pass its join condition), plus one copy of each row in the left-hand table for which there was no right-hand row that passed the join condition. This left-hand row is extended to the full width of the joined table by inserting null values for the right-hand columns. Note that only the `JOIN` clause’s own condition is considered while deciding which rows have matches. Outer conditions are applied afterwards.

Conversely, `RIGHT OUTER JOIN` returns all the joined rows, plus one row for each unmatched right-hand row (extended with nulls on the left). This is just a notational convenience, since you could convert it to a `LEFT OUTER JOIN` by switching the left and right inputs.
FULL OUTER JOIN returns all the joined rows, plus one row for each unmatched
left-hand row (extended with nulls on the right), plus one row for each unmatched
right-hand row (extended with nulls on the left).

ON join_condition

join_condition is an expression resulting in a value of type boolean (similar to a
WHERE clause) that specifies which rows in a join are considered to match.

USING (join_column [, ...])

A clause of the form USING (a, b, ... ) is shorthand for ON left_table.a =
right_table.a AND left_table.b = right_table.b .... Also, USING
implies that only one of each pair of equivalent columns will be included in the join
output, not both.

NATURAL

NATURAL is shorthand for a USING list that mentions all columns in the two tables
that have the same names.

The WHERE Clause

The optional WHERE clause has the general form:

    WHERE condition

Where condition is any expression that evaluates to a result of type boolean. Any
row that does not satisfy this condition will be eliminated from the output. A row
satisfies the condition if it returns true when the actual row values are substituted for
any variable references.

The GROUP BY Clause

The optional GROUP BY clause has the general form:

    GROUP BY grouping_element [, ...]

where grouping_element can be one of:

    ()
    expression
    ROLLUP (expression [, ...])
    CUBE (expression [, ...])
    GROUPING SETS ((grouping_element [, ...]))

GROUP BY will condense into a single row all selected rows that share the same values
for the grouped expressions. expression can be an input column name, or the name
or ordinal number of an output column (SELECT list item), or an arbitrary expression
formed from input-column values. In case of ambiguity, a GROUP BY name will be
interpreted as an input-column name rather than an output column name.

Aggregate functions, if any are used, are computed across all rows making up each
group, producing a separate value for each group (whereas without GROUP BY, an
aggregate produces a single value computed across all the selected rows). When
GROUP BY is present, it is not valid for the SELECT list expressions to refer to ungrouped columns except within aggregate functions, since there would be more than one possible value to return for an ungrouped column.

Greenplum Database has the following additional OLAP grouping extensions (often referred to as supergroups):

**ROLLUP**

A ROLLUP grouping is an extension to the GROUP BY clause that creates aggregate subtotals that roll up from the most detailed level to a grand total, following a list of grouping columns (or expressions). ROLLUP takes an ordered list of grouping columns, calculates the standard aggregate values specified in the GROUP BY clause, then creates progressively higher-level subtotals, moving from right to left through the list. Finally, it creates a grand total. A ROLLUP grouping can be thought of as a series of grouping sets. For example:

```
GROUP BY ROLLUP (a,b,c)
```

is equivalent to:

```
GROUP BY GROUPING SETS( (a,b,c), (a,b), (a), () )
```

Notice that the $n$ elements of a ROLLUP translate to $n+1$ grouping sets. Also, the order in which the grouping expressions are specified is significant in a ROLLUP.

**CUBE**

A CUBE grouping is an extension to the GROUP BY clause that creates subtotals for all of the possible combinations of the given list of grouping columns (or expressions). In terms of multidimensional analysis, CUBE generates all the subtotals that could be calculated for a data cube with the specified dimensions. For example:

```
GROUP BY CUBE (a,b,c)
```

is equivalent to:

```
GROUP BY GROUPING SETS( (a,b,c), (a,b), (a,c), (b,c), (a), (b), (c), () )
```

Notice that $n$ elements of a CUBE translate to $2^n$ grouping sets. Consider using CUBE in any situation requiring cross-tabular reports. CUBE is typically most suitable in queries that use columns from multiple dimensions rather than columns representing different levels of a single dimension. For instance, a commonly requested cross-tabulation might need subtotals for all the combinations of month, state, and product.

**GROUPING SETS**

You can selectively specify the set of groups that you want to create using a GROUPING SETS expression within a GROUP BY clause. This allows precise specification across multiple dimensions without computing a whole ROLLUP or CUBE. For example:

```
GROUP BY GROUPING SETS( (a,c), (a,b) )
```
If using the grouping extension clauses `ROLLUP`, `CUBE`, or `GROUPING SETS`, two challenges arise. First, how do you determine which result rows are subtotals, and then the exact level of aggregation for a given subtotal. Or, how do you differentiate between result rows that contain both stored `NULL` values and “NULL” values created by the `ROLLUP` or `CUBE`. Secondly, when duplicate grouping sets are specified in the `GROUP BY` clause, how do you determine which result rows are duplicates? There are two additional grouping functions you can use in the `SELECT` list to help with this:

- **grouping(column [, ...])** The grouping function can be applied to one or more grouping attributes to distinguish super-aggregated rows from regular grouped rows. This can be helpful in distinguishing a “NULL” representing the set of all values in a super-aggregated row from a `NULL` value in a regular row. Each argument in this function produces a bit — either 1 or 0, where 1 means the result row is super-aggregated, and 0 means the result row is from a regular grouping. The grouping function returns an integer by treating these bits as a binary number and then converting it to a base-10 integer.

- **group_id()** For grouping extension queries that contain duplicate grouping sets, the group_id function is used to identify duplicate rows in the output. All unique grouping set output rows will have a group_id value of 0. For each duplicate grouping set detected, the group_id function assigns a group_id number greater than 0. All output rows in a particular duplicate grouping set are identified by the same group_id number.

### The WINDOW Clause

The `WINDOW` clause is used to define a window that can be used in the `OVER()` expression of a window function such as `rank` or `avg`. For example:

```sql
SELECT vendor, rank() OVER (mywindow) FROM sale
WINDOW mywindow AS (ORDER BY sum(prc*qty));
```

A `WINDOW` clause is has this general form:

```sql
WINDOW window_name AS (window_specification)
```

where `window_specification` can be:

- `[window_name]`
- `[PARTITION BY expression [, ...]]`
- `[ORDER BY expression [ASC | DESC | USING operator] [, ...]` ([RANGE | ROWS])
  - `{ UNBOUNDED PRECEDING
    | expression PRECEDING
    | CURRENT ROW
    | BETWEEN window_frame_bound AND window_frame_bound }]

where `window_frame_bound` can be one of:

UNBOUNDED PRECEDING
expression PRECEDING
CURRENT ROW
expression FOLLOWING
UNBOUNDED FOLLOWING
window_name

Gives a name to the window specification.

PARTITION BY

The PARTITION BY clause organizes the result set into logical groups based on the unique values of the specified expression. When used with window functions, the functions are applied to each partition independently. For example, if you follow PARTITION BY with a column name, the result set is partitioned by the distinct values of that column. If omitted, the entire result set is considered one partition.

ORDER BY

The ORDER BY clause defines how to sort the rows in each partition of the result set. If omitted, rows are returned in whatever order is most efficient and may vary.

NOTE: Columns of data types that lack a coherent ordering, such as time, are not good candidates for use in the ORDER BY clause of a window specification. Time, with or without time zone, lacks a coherent ordering because addition and subtraction do not have the expected effects. For example, the following is not generally true: x::time < x::time + '2 hour':interval

ROWS | RANGE

Use either a ROWS or RANGE clause to express the bounds of the window. The window bound can be one, many, or all rows of a partition. You can express the bound of the window either in terms of a range of data values offset from the value in the current row (RANGE), or in terms of the number of rows offset from the current row (ROWS). When using the RANGE clause, you must also use an ORDER BY clause. This is because the calculation performed to produce the window requires that the values be sorted. Additionally, the ORDER BY clause cannot contain more than one expression, and the expression must result in either a date or a numeric value. When using the ROWS or RANGE clauses, if you specify only a starting row, the current row is used as the last row in the window.

PRECEDING

The PRECEDING clause defines the first row of the window using the current row as a reference point. The starting row is expressed in terms of the number of rows preceding the current row. For example, in the case of ROWS framing, 5 PRECEDING sets the window to start with the fifth row preceding the current row. In the case of RANGE framing, it sets the window to start with the first row whose ordering column value precedes that of the current row by 5 in the given order. If the specified order is ascending by date, this will be the first row within 5 days before the current row. UNBOUNDED PRECEDING sets the first row in the window to be the first row in the partition.

BETWEEN

The BETWEEN clause defines the first and last row of the window, using the current row as a reference point. First and last rows are expressed in terms of the number of rows preceding and following the current row, respectively. For example, BETWEEN 3 PRECEDING AND 5 FOLLOWING sets the window to start with the third row preceding the current row, and end with the fifth row following the current row. Use BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED
FOLLOWING to set the first and last rows in the window to be the first and last row in the partition, respectively. This is equivalent to the default behavior if no ROW or RANGE clause is specified.

FOLLOWING

The FOLLOWING clause defines the last row of the window using the current row as a reference point. The last row is expressed in terms of the number of rows following the current row. For example, in the case of ROWS framing, 5 FOLLOWING sets the window to end with the fifth row following the current row. In the case of RANGE framing, it sets the window to end with the last row whose ordering column value follows that of the current row by 5 in the given order. If the specified order is ascending by date, this will be the last row within 5 days after the current row. Use UNBOUNDED FOLLOWING to set the last row in the window to be the last row in the partition.

If you do not specify a ROW or a RANGE clause, the window bound starts with the first row in the partition (UNBOUNDED PRECEDING) and ends with the current row (CURRENT ROW) if ORDER BY is used. If an ORDER BY is not specified, the window starts with the first row in the partition (UNBOUNDED PRECEDING) and ends with last row in the partition (UNBOUNDED FOLLOWING).

The HAVING Clause

The optional HAVING clause has the general form:

```
HAVING condition
```

Where condition is the same as specified for the WHERE clause. HAVING eliminates group rows that do not satisfy the condition. HAVING is different from WHERE: WHERE filters individual rows before the application of GROUP BY, while HAVING filters group rows created by GROUP BY. Each column referenced in condition must unambiguously reference a grouping column, unless the reference appears within an aggregate function.

The presence of HAVING turns a query into a grouped query even if there is no GROUP BY clause. This is the same as what happens when the query contains aggregate functions but no GROUP BY clause. All the selected rows are considered to form a single group, and the SELECT list and HAVING clause can only reference table columns from within aggregate functions. Such a query will emit a single row if the HAVING condition is true, zero rows if it is not true.

The UNION Clause

The UNION clause has this general form:

```
select_statement UNION [ALL] select_statement
```

Where select_statement is any SELECT statement without an ORDER BY, LIMIT, FOR UPDATE, or FOR SHARE clause. (ORDER BY and LIMIT can be attached to a subquery expression if it is enclosed in parentheses. Without parentheses, these clauses will be taken to apply to the result of the UNION, not to its right-hand input expression.)
The **UNION** operator computes the set union of the rows returned by the involved **SELECT** statements. A row is in the set union of two result sets if it appears in at least one of the result sets. The two **SELECT** statements that represent the direct operands of the **UNION** must produce the same number of columns, and corresponding columns must be of compatible data types.

The result of **UNION** does not contain any duplicate rows unless the **ALL** option is specified. **ALL** prevents elimination of duplicates. (Therefore, **UNION ALL** is usually significantly quicker than **UNION**; use **ALL** when you can.)

Multiple **UNION** operators in the same **SELECT** statement are evaluated left to right, unless otherwise indicated by parentheses.

Currently, **FOR UPDATE** and **FOR SHARE** may not be specified either for a **UNION** result or for any input of a **UNION**.

**The INTERSECT Clause**

The **INTERSECT** clause has this general form:

```
select_statement INTERSECT [ALL] select_statement
```

Where **select_statement** is any **SELECT** statement without an **ORDER BY**, **LIMIT**, **FOR UPDATE**, or **FOR SHARE** clause.

The **INTERSECT** operator computes the set intersection of the rows returned by the involved **SELECT** statements. A row is in the intersection of two result sets if it appears in both result sets.

The result of **INTERSECT** does not contain any duplicate rows unless the **ALL** option is specified. With **ALL**, a row that has \(m\) duplicates in the left table and \(n\) duplicates in the right table will appear \(\min(m, n)\) times in the result set.

Multiple **INTERSECT** operators in the same **SELECT** statement are evaluated left to right, unless parentheses dictate otherwise. **INTERSECT** binds more tightly than **UNION**. That is, \(A \cup B \cap C\) will be read as \(A \cup (B \cap C)\).

Currently, **FOR UPDATE** and **FOR SHARE** may not be specified either for an **INTERSECT** result or for any input of an **INTERSECT**.

**The EXCEPT Clause**

The **EXCEPT** clause has this general form:

```
select_statement EXCEPT [ALL] select_statement
```

Where **select_statement** is any **SELECT** statement without an **ORDER BY**, **LIMIT**, **FOR UPDATE**, or **FOR SHARE** clause.

The **EXCEPT** operator computes the set of rows that are in the result of the left **SELECT** statement but not in the result of the right one.

The result of **EXCEPT** does not contain any duplicate rows unless the **ALL** option is specified. With **ALL**, a row that has \(m\) duplicates in the left table and \(n\) duplicates in the right table will appear \(\max(m-n,0)\) times in the result set.

Multiple **EXCEPT** operators in the same **SELECT** statement are evaluated left to right, unless parentheses dictate otherwise. **EXCEPT** binds at the same level as **UNION**.
Currently, \texttt{FOR UPDATE} and \texttt{FOR SHARE} may not be specified either for an \texttt{EXCEPT} result or for any input of an \texttt{EXCEPT}.

**The ORDER BY Clause**

The optional \texttt{ORDER BY} clause has this general form:

\begin{verbatim}
ORDER BY expression [ASC | DESC | USING operator] [, \ldots ]
\end{verbatim}

Where \texttt{expression} can be the name or ordinal number of an output column (\texttt{SELECT} list item), or it can be an arbitrary expression formed from input-column values.

The \texttt{ORDER BY} clause causes the result rows to be sorted according to the specified expressions. If two rows are equal according to the left-most expression, they are compared according to the next expression and so on. If they are equal according to all specified expressions, they are returned in an implementation-dependent order.

The ordinal number refers to the ordinal (left-to-right) position of the result column. This feature makes it possible to define an ordering on the basis of a column that does not have a unique name. This is never absolutely necessary because it is always possible to assign a name to a result column using the \texttt{AS} clause.

It is also possible to use arbitrary expressions in the \texttt{ORDER BY} clause, including columns that do not appear in the \texttt{SELECT} result list. Thus the following statement is valid:

\begin{verbatim}
SELECT name FROM distributors ORDER BY code;
\end{verbatim}

A limitation of this feature is that an \texttt{ORDER BY} clause applying to the result of a \texttt{UNION}, \texttt{INTERSECT}, or \texttt{EXCEPT} clause may only specify an output column name or number, not an expression.

If an \texttt{ORDER BY} expression is a simple name that matches both a result column name and an input column name, \texttt{ORDER BY} will interpret it as the result column name. This is the opposite of the choice that \texttt{GROUP BY} will make in the same situation. This inconsistency is made to be compatible with the SQL standard.

Optionally one may add the key word \texttt{ASC} (ascending) or \texttt{DESC} (descending) after any expression in the \texttt{ORDER BY} clause. If not specified, \texttt{ASC} is assumed by default. Alternatively, a specific ordering operator name may be specified in the \texttt{USING} clause. \texttt{ASC} is usually equivalent to \texttt{USING <} and \texttt{DESC} is usually equivalent to \texttt{USING >}. (But the creator of a user-defined data type can define exactly what the default sort ordering is, and it might correspond to operators with other names.)

The null value sorts higher than any other value. In other words, with ascending sort order, null values sort at the end, and with descending sort order, null values sort at the beginning.

Character-string data is sorted according to the locale-specific collation order that was established when the Greenplum Database system was initialized.

**The DISTINCT Clause**

If \texttt{DISTINCT} is specified, all duplicate rows are removed from the result set (one row is kept from each group of duplicates). \texttt{ALL} specifies the opposite: all rows are kept. \texttt{ALL} is the default.
DISTINCT ON (expression [, ...]) keeps only the first row of each set of rows where the given expressions evaluate to equal. The DISTINCT ON expressions are interpreted using the same rules as for ORDER BY. Note that the ‘first row’ of each set is unpredictable unless ORDER BY is used to ensure that the desired row appears first. For example:

```
SELECT DISTINCT ON (location) location, time, report FROM weather_reports ORDER BY location, time DESC;
```

retrieves the most recent weather report for each location. But if we had not used ORDER BY to force descending order of time values for each location, we would have gotten a report from an unpredictable time for each location.

The DISTINCT ON expression(s) must match the left-most ORDER BY expression(s). The ORDER BY clause will normally contain additional expression(s) that determine the desired precedence of rows within each DISTINCT ON group.

### The LIMIT Clause

The LIMIT clause consists of two independent sub-clauses:

```
LIMIT {count | ALL}
OFFSET start
```

Where `count` specifies the maximum number of rows to return, while `start` specifies the number of rows to skip before starting to return rows. When both are specified, start rows are skipped before starting to count the count rows to be returned.

When using LIMIT, it is a good idea to use an ORDER BY clause that constrains the result rows into a unique order. Otherwise you will get an unpredictable subset of the query’s rows — you may be asking for the tenth through twentieth rows, but tenth through twentieth in what ordering? You don’t know what ordering unless you specify ORDER BY.

The query planner takes LIMIT into account when generating a query plan, so you are very likely to get different plans (yielding different row orders) depending on what you use for LIMIT and OFFSET. Thus, using different LIMIT/OFFSET values to select different subsets of a query result will give inconsistent results unless you enforce a predictable result ordering with ORDER BY. This is not a defect; it is an inherent consequence of the fact that SQL does not promise to deliver the results of a query in any particular order unless ORDER BY is used to constrain the order.

### FOR UPDATE/FOR SHARE Clause

The FOR UPDATE clause has this form:

```
FOR UPDATE [OF table_name [, ...]] [NOWAIT]
```

The closely related FOR SHARE clause has this form:

```
FOR SHARE [OF table_name [, ...]] [NOWAIT]
```

FOR UPDATE causes the tables accessed by the SELECT statement to be locked as though for update. This prevents the table from being modified or deleted by other transactions until the current transaction ends. That is, other transactions that attempt UPDATE, DELETE, or SELECT FOR UPDATE of this table will be blocked until the current transaction ends. Also, if an UPDATE, DELETE, or SELECT FOR UPDATE from
another transaction has already locked a selected table, SELECT FOR UPDATE will wait for the other transaction to complete, and will then lock and return the updated table.

To prevent the operation from waiting for other transactions to commit, use the NOWAIT option. SELECT FOR UPDATE NOWAIT reports an error, rather than waiting, if a selected row cannot be locked immediately. Note that NOWAIT applies only to the row-level lock(s) — the required ROW SHARE table-level lock is still taken in the ordinary way. You can use the NOWAIT option of LOCK if you need to acquire the table-level lock without waiting (see LOCK).

FOR SHARE behaves similarly, except that it acquires a shared rather than exclusive lock on the table. A shared lock blocks other transactions from performing UPDATE, DELETE, or SELECT FOR UPDATE on the table, but it does not prevent them from performing SELECT FOR SHARE.

If specific tables are named in FOR UPDATE or FOR SHARE, then only those tables are locked; any other tables used in the SELECT are simply read as usual. A FOR UPDATE or FOR SHARE clause without a table list affects all tables used in the command. If FOR UPDATE or FOR SHARE is applied to a view or subquery, it affects all tables used in the view or subquery.

Multiple FOR UPDATE and FOR SHARE clauses can be written if it is necessary to specify different locking behavior for different tables. If the same table is mentioned (or implicitly affected) by both FOR UPDATE and FOR SHARE clauses, then it is processed as FOR UPDATE. Similarly, a table is processed as NOWAIT if that is specified in any of the clauses affecting it.

Examples

To join the table films with the table distributors:

```sql
SELECT f.title, f.did, d.name, f.date_prod, f.kind FROM distributors d, films f WHERE f.did = d.did
```

To sum the column length of all films and group the results by kind:

```sql
SELECT kind, sum(length) AS total FROM films GROUP BY kind;
```

To sum the column length of all films, group the results by kind and show those group totals that are less than 5 hours:

```sql
SELECT kind, sum(length) AS total FROM films GROUP BY kind HAVING sum(length) < interval '5 hours';
```

Calculate the subtotals and grand totals of all sales for movie kind and distributor.

```sql
SELECT kind, distributor, sum(prc*qty) FROM sales GROUP BY ROLLUP(kind, distributor)
ORDER BY 1,2,3;
```

Calculate the rank of movie distributors based on total sales:

```sql
SELECT distributor, sum(prc*qty),
   rank() OVER (ORDER BY sum(prc*qty) DESC)
FROM sale
```
GROUP BY distributor ORDER BY 2 DESC;

The following two examples are identical ways of sorting the individual results according to the contents of the second column (name):

```sql
SELECT * FROM distributors ORDER BY name;
SELECT * FROM distributors ORDER BY 2;
```

The next example shows how to obtain the union of the tables distributors and actors, restricting the results to those that begin with the letter W in each table. Only distinct rows are wanted, so the key word ALL is omitted:

```sql
SELECT distributors.name FROM distributors WHERE distributors.name LIKE 'W%' UNION SELECT actors.name FROM actors WHERE actors.name LIKE 'W%';
```

This example shows how to use a function in the FROM clause, both with and without a column definition list:

```sql
CREATE FUNCTION distributors(int) RETURNS SETOF distributors AS $$ SELECT * FROM distributors WHERE did = $1; $$ LANGUAGE SQL;
SELECT * FROM distributors(111);

CREATE FUNCTION distributors_2(int) RETURNS SETOF record AS $$ SELECT * FROM distributors WHERE did = $1; $$ LANGUAGE SQL;
SELECT * FROM distributors_2(111) AS (dist_id int, dist_name text);
```

**Compatibility**

The SELECT statement is compatible with the SQL standard, but there are some extensions and some missing features.

**Omitted FROM Clauses**

Greenplum Database allows one to omit the FROM clause. It has a straightforward use to compute the results of simple expressions. For example:

```sql
SELECT 2+2;
```

Some other SQL databases cannot do this except by introducing a dummy one-row table from which to do the SELECT.

Note that if a FROM clause is not specified, the query cannot reference any database tables. For compatibility with applications that rely on this behavior the add_missing_from configuration variable can be enabled.

**The AS Key Word**

In the SQL standard, the optional key word AS is just noise and can be omitted without affecting the meaning. The Greenplum Database parser requires this key word when renaming output columns because the type extensibility features lead to parsing ambiguities without it. AS is optional in FROM items, however.
Namespace Available to GROUP BY and ORDER BY

In the SQL-92 standard, an ORDER BY clause may only use result column names or numbers, while a GROUP BY clause may only use expressions based on input column names. Greenplum Database extends each of these clauses to allow the other choice as well (but it uses the standard’s interpretation if there is ambiguity). Greenplum Database also allows both clauses to specify arbitrary expressions. Note that names appearing in an expression will always be taken as input-column names, not as result-column names.

SQL:1999 and later use a slightly different definition which is not entirely upward compatible with SQL-92. In most cases, however, Greenplum Database will interpret an ORDER BY or GROUP BY expression the same way SQL:1999 does.

Nonstandard Clauses

The clauses DISTINCT ON, LIMIT, and OFFSET are not defined in the SQL standard.

Limited Correlated Subquery Syntax

A correlated subquery is a nested SELECT statement that refers to a column from an outer SELECT statement. For example:

```
SELECT * FROM product WHERE exists (SELECT * FROM sale WHERE qty>0 AND pn = product.pn);
```

Subqueries that do not need any input from the parent query and produce just a single value or row are not subject to the limitations described in this section.

Greenplum Database (as does PostgreSQL) supports correlated subqueries by flattening (transforming the subquery into a join with the table(s) of the outer query). If a correlated subquery has none of the characteristics mentioned below that prevent flattening, then it will be transformed to a join and be executed as expected. Correlated subqueries with the following syntax cannot be flattened and will be rejected:

- `expression operator ANY/ALL (subquery)` where `expression` contains a volatile function
- `EXISTS with aggregate in HAVING`
- `(subquery) returning a single value or row`
- `ARRAY (subquery) returning an array of values or rows`
- `Correlated subquery used in WHERE clause`
- `Correlated subquery used in HAVING clause`
- `Correlated subquery used in OR expression`
- `Correlated subquery used in select-list`
- `Set-returning function in select-list of subquery`
- `Aggregate function in subquery (other than EXISTS)`
- `DISTINCT in value, row or ARRAY subquery`
- `GROUP BY or DISTINCT in subquery (other than EXISTS)`
- `LIMIT in subquery (other than EXISTS)`
- `OFFSET in subquery`
• Subquery without a FROM clause

**Limited Use of STABLE and VOLATILE Functions**

To prevent data from becoming out-of-sync across the segments in Greenplum Database, any function classified as **STABLE** or **VOLATILE** cannot be executed at the segment database level if it contains SQL or modifies the database in any way. See **CREATE FUNCTION** for more information.

---

**See Also**

**EXPLAIN**
### SELECT INTO

Defines a new table from the results of a query.

#### Synopsis

```
SELECT [ALL | DISTINCT [ON ( expression [, ...] )]]
   * | expression [AS output_name] [, ...]
INTO [TEMPORARY | TEMP] [TABLE] new_table
   [FROM from_item [, ...]]
   [WHERE condition]
   [GROUP BY expression [, ...]]
   [HAVING condition [, ...]]
   [{UNION | INTERSECT | EXCEPT} [ALL] select]
   [ORDER BY expression [ASC | DESC | USING operator] [, ...]]
   [LIMIT {count | ALL}]
   [OFFSET start]
   [FOR {UPDATE | SHARE} [OF table_name [, ...]] [NOWAIT]]
```

#### Description

SELECT INTO creates a new table and fills it with data computed by a query. The data is not returned to the client, as it is with a normal SELECT. The new table’s columns have the names and data types associated with the output columns of the SELECT.

#### Parameters

The majority of parameters for SELECT INTO are the same as SELECT.

**TEMPORARY**

**TEMP**

If specified, the table is created as a temporary table.

**new_table**

The name (optionally schema-qualified) of the table to be created.

#### Examples

Create a new table `films_recent` consisting of only recent entries from the table `films`:

```
SELECT * INTO films_recent FROM films WHERE date_prod >= '2006-01-01';
```
**Compatibility**

The SQL standard uses `SELECT INTO` to represent selecting values into scalar variables of a host program, rather than creating a new table. The Greenplum Database usage of `SELECT INTO` to represent table creation is historical. It is best to use `CREATE TABLE AS` for this purpose in new applications.

**See Also**

`SELECT`, `CREATE TABLE AS`
**SET**

Changes the value of a Greenplum Database configuration parameter.

**Synopsis**

```
SET [SESSION | LOCAL] configuration_parameter {TO | =} value | 
'value' | DEFAULT

SET [SESSION | LOCAL] TIME_ZONE {timezone | LOCAL | DEFAULT}
```

**Description**

The `SET` command changes server configuration parameters. Any configuration parameter classified as a *session* parameter can be changed on-the-fly with `SET` (see “Server Configuration Parameters” on page 770 for details). `SET` only affects the value used by the current session.

If `SET` or `SET SESSION` is issued within a transaction that is later aborted, the effects of the `SET` command disappear when the transaction is rolled back. Once the surrounding transaction is committed, the effects will persist until the end of the session, unless overridden by another `SET`.

The effects of `SET LOCAL` last only till the end of the current transaction, whether committed or not. A special case is `SET` followed by `SET LOCAL` within a single transaction: the `SET LOCAL` value will be seen until the end of the transaction, but afterwards (if the transaction is committed) the `SET` value will take effect.

**Parameters**

**SESSION**

Specifies that the command takes effect for the current session. This is the default.

**LOCAL**

Specifies that the command takes effect for only the current transaction. After `COMMIT` or `ROLLBACK`, the session-level setting takes effect again. Note that `SET LOCAL` will appear to have no effect if it is executed outside of a transaction.

**configuration_parameter**

The name of a Greenplum Database configuration parameter. Only parameters classified as *session* can be changed with `SET`. See “Server Configuration Parameters” on page 770 for details.

**value**

New value of parameter. Values can be specified as string constants, identifiers, numbers, or comma-separated lists of these. `DEFAULT` can be used to specify resetting the parameter to its default value. If specifying memory sizing or time units, enclose the value in single quotes.
**TIME ZONE**

SET TIME ZONE value is an alias for SET timezone TO value. The syntax SET TIME ZONE allows special syntax for the time zone specification. Here are examples of valid values:

- 'PST8PDT'
- 'Europe/Rome'
-7 (time zone 7 hours west from UTC)
INTERVAL '-08:00' HOUR TO MINUTE (time zone 8 hours west from UTC).

**LOCAL DEFAULT**

Set the time zone to your local time zone (the one that the server’s operating system defaults to). See the Time zone section of the PostgreSQL documentation for more information about time zones in Greenplum Database.

**Examples**

Set the schema search path:

```sql
SET search_path TO my_schema, public;
```

Increase work memory to 200 MB:

```sql
SET work_mem TO '200MB';
```

Set the style of date to traditional POSTGRES with “day before month” input convention:

```sql
SET datestyle TO postgres, dmy;
```

Set the time zone for San Mateo, California:

```sql
SET TIME ZONE 'PST8PDT';
```

Set the time zone for Italy:

```sql
SET TIME ZONE 'Europe/Rome';
```

**Compatibility**

SET TIME ZONE extends syntax defined in the SQL standard. The standard allows only numeric time zone offsets while Greenplum Database allows more flexible time-zone specifications. All other SET features are Greenplum Database extensions.

**See Also**

RESET, SHOW
## SET ROLE

Sets the current role identifier of the current session.

### Synopsis

```
SET [SESSION | LOCAL] ROLE rolename
SET [SESSION | LOCAL] ROLE NONE
RESET ROLE
```

### Description

This command sets the current role identifier of the current SQL-session context to be *rolename*. The role name may be written as either an identifier or a string literal. After `SET ROLE`, permissions checking for SQL commands is carried out as though the named role were the one that had logged in originally.

The specified *rolename* must be a role that the current session user is a member of. If the session user is a superuser, any role can be selected.

The `NONE` and `RESET` forms reset the current role identifier to be the current session role identifier. These forms may be executed by any user.

### Parameters

**SESSION**

Specifies that the command takes effect for the current session. This is the default.

**LOCAL**

Specifies that the command takes effect for only the current transaction. After `COMMIT` or `ROLLBACK`, the session-level setting takes effect again. Note that `SET LOCAL` will appear to have no effect if it is executed outside of a transaction.

**rolename**

The name of a role to use for permissions checking in this session.

**NONE**

**RESET**

Reset the current role identifier to be the current session role identifier (that of the role used to log in).

### Notes

Using this command, it is possible to either add privileges or restrict privileges. If the session user role has the `INHERITS` attribute, then it automatically has all the privileges of every role that it could `SET ROLE` to; in this case `SET ROLE` effectively drops all the privileges assigned directly to the session user and to the other roles it is a member of, leaving only the privileges available to the named role. On the other
hand, if the session user role has the \texttt{NOINHERITS} attribute, \texttt{SET ROLE} drops the privileges assigned directly to the session user and instead acquires the privileges available to the named role.

In particular, when a superuser chooses to \texttt{SET ROLE} to a non-superuser role, she loses her superuser privileges.

\texttt{SET ROLE} has effects comparable to \texttt{SET SESSION AUTHORIZATION}, but the privilege checks involved are quite different. Also, \texttt{SET SESSION AUTHORIZATION} determines which roles are allowable for later \texttt{SET ROLE} commands, whereas changing roles with \texttt{SET ROLE} does not change the set of roles allowed to a later \texttt{SET ROLE}.

\underline{Examples}

\begin{verbatim}
SELECT SESSION_USER, CURRENT_USER;
 session_user | current_user
--------------+--------------
 peter        | peter

SET ROLE 'paul';

SELECT SESSION_USER, CURRENT_USER;
 session_user | current_user
--------------+--------------
 peter        | paul
\end{verbatim}

\underline{Compatibility}

Greenplum Database allows identifier syntax (\texttt{rolename}), while the SQL standard requires the role name to be written as a string literal. SQL does not allow this command during a transaction; Greenplum Database does not make this restriction. The \texttt{SESSION} and \texttt{LOCAL} modifiers are a Greenplum Database extension, as is the \texttt{RESET} syntax.

\underline{See Also}

\texttt{SET SESSION AUTHORIZATION}
SET SESSION AUTHORIZATION

Sets the session role identifier and the current role identifier of the current session.

**Synopsis**

```sql
SET [SESSION | LOCAL] SESSION AUTHORIZATION rolename
SET [SESSION | LOCAL] SESSION AUTHORIZATION DEFAULT
RESET SESSION AUTHORIZATION
```

**Description**

This command sets the session role identifier and the current role identifier of the current SQL-session context to be `rolename`. The role name may be written as either an identifier or a string literal. Using this command, it is possible, for example, to temporarily become an unprivileged user and later switch back to being a superuser.

The session role identifier is initially set to be the (possibly authenticated) role name provided by the client. The current role identifier is normally equal to the session user identifier, but may change temporarily in the context of `setuid` functions and similar mechanisms; it can also be changed by `SET ROLE`. The current user identifier is relevant for permission checking.

The session user identifier may be changed only if the initial session user (the authenticated user) had the superuser privilege. Otherwise, the command is accepted only if it specifies the authenticated user name.

The `DEFAULT` and `RESET` forms reset the session and current user identifiers to be the originally authenticated user name. These forms may be executed by any user.

**Parameters**

- **SESSION**
  
  Specifies that the command takes effect for the current session. This is the default.

- **LOCAL**
  
  Specifies that the command takes effect for only the current transaction. After `COMMIT` or `ROLLBACK`, the session-level setting takes effect again. Note that `SET LOCAL` will appear to have no effect if it is executed outside of a transaction.

- **rolename**
  
  The name of the role to assume.

- **NONE**
  
  Reset the session and current role identifiers to be that of the role used to log in.
Examples

```
SELECT SESSION_USER, CURRENT_USER;
session_user | current_user
--------------+--------------
peter         | peter

SET SESSION AUTHORIZATION 'paul';

SELECT SESSION_USER, CURRENT_USER;
session_user | current_user
--------------+--------------
paul         | paul
```

Compatibility

The SQL standard allows some other expressions to appear in place of the literal `rolename`, but these options are not important in practice. Greenplum Database allows identifier syntax ("`rolename`"), which SQL does not. SQL does not allow this command during a transaction; Greenplum Database does not make this restriction. The `SESSION` and `LOCAL` modifiers are a Greenplum Database extension, as is the `RESET` syntax.

See Also

`SET ROLE`
SET TRANSACTION

Sets the characteristics of the current transaction.

Synopsis

SET TRANSACTION transaction_mode [, ...]
SET SESSION CHARACTERISTICS AS TRANSACTION transaction_mode [, ...]

where transaction_mode is one of:

- ISOLATION LEVEL {SERIALIZABLE | REPEATABLE READ | READ COMMITTED | READ UNCOMMITTED}
- READ WRITE | READ ONLY

Description

The SET TRANSACTION command sets the characteristics of the current transaction. It has no effect on any subsequent transactions.

The available transaction characteristics are the transaction isolation level and the transaction access mode (read/write or read-only).

The isolation level of a transaction determines what data the transaction can see when other transactions are running concurrently.

- **READ COMMITTED** — A statement can only see rows committed before it began. This is the default.
- **SERIALIZABLE** — All statements of the current transaction can only see rows committed before the first query or data-modification statement was executed in this transaction.

The SQL standard defines two additional levels, READ UNCOMMITTED and REPEATABLE READ. In Greenplum Database READ UNCOMMITTED is treated as READ COMMITTED, while REPEATABLE READ is treated as SERIALIZABLE.

The transaction isolation level cannot be changed after the first query or data-modification statement (SELECT, INSERT, DELETE, UPDATE, FETCH, or COPY) of a transaction has been executed.

The transaction access mode determines whether the transaction is read/write or read-only. Read/write is the default. When a transaction is read-only, the following SQL commands are disallowed: INSERT, UPDATE, DELETE, and COPY FROM if the table they would write to is not a temporary table; all CREATE, ALTER, and DROP commands; GRANT, REVOKE, TRUNCATE; and EXPLAIN ANALYZE and EXECUTE if the command they would execute is among those listed. This is a high-level notion of read-only that does not prevent all writes to disk.

Parameters

SESSION CHARACTERISTICS

Sets the default transaction characteristics for subsequent transactions of a session.
SERIALIZABLE
REPEATABLE READ
READ COMMITTED
READ UNCOMMITTED

The SQL standard defines four transaction isolation levels: READ COMMITTED, READ UNCOMMITTED, SERIALIZABLE, and REPEATABLE READ. The default behavior is that a statement can only see rows committed before it began (READ COMMITTED). In Greenplum Database READ UNCOMMITTED is treated the same as READ COMMITTED. SERIALIZABLE is supported the same as REPEATABLE READ wherein all statements of the current transaction can only see rows committed before the first statement was executed in the transaction. SERIALIZABLE is the strictest transaction isolation. This level emulates serial transaction execution, as if transactions had been executed one after another, serially, rather than concurrently. Applications using this level must be prepared to retry transactions due to serialization failures.

READ WRITE
READ ONLY

Determines whether the transaction is read/write or read-only. Read/write is the default. When a transaction is read-only, the following SQL commands are disallowed: INSERT, UPDATE, DELETE, and COPY FROM if the table they would write to is not a temporary table; all CREATE, ALTER, and DROP commands; GRANT, REVOKE, TRUNCATE; and EXPLAIN ANALYZE and EXECUTE if the command they would execute is among those listed.

Notes

If SET TRANSACTION is executed without a prior START TRANSACTION or BEGIN, it will appear to have no effect.

It is possible to dispense with SET TRANSACTION by instead specifying the desired transaction_modes in BEGIN or START TRANSACTION.

The session default transaction modes can also be set by setting the configuration parameters default_transaction_isolation and default_transaction_read_only.

Examples

Set the transaction isolation level for the current transaction:

```
BEGIN;
SET TRANSACTION SERIALizable;
```

Compatibility

Both commands are defined in the SQL standard. SERIALIZABLE is the default transaction isolation level in the standard. In Greenplum Database the default is READ COMMITTED. Because of lack of predicate locking, the SERIALIZABLE level is not truly serializable. Essentially, a predicate-locking system prevents phantom reads by restricting what is written, whereas a multi-version concurrency control model (MVCC) as used in Greenplum Database prevents them by restricting what is read.
In the SQL standard, there is one other transaction characteristic that can be set with these commands: the size of the diagnostics area. This concept is specific to embedded SQL, and therefore is not implemented in the Greenplum Database server.

The SQL standard requires commas between successive `transaction_modes`, but for historical reasons Greenplum Database allows the commas to be omitted.

**See Also**

`BEGIN`, `LOCK`
SHOW

Shows the value of a system configuration parameter.

Synopsis

SHOW configuration_parameter
SHOW ALL

Description

SHOW will display the current settings of Greenplum Database system configuration parameters. These parameters can be set using the SET statement, or by editing the postgresql.conf configuration file of the Greenplum Database master. Note that some parameters viewable by SHOW are read-only — their values can be viewed but not set. See “Server Configuration Parameters” on page 770 for details.

Parameters

configuration_parameter
  The name of a system configuration parameter. See “Server Configuration Parameters” on page 770.

ALL
  Shows the current value of all configuration parameters.

Examples

Show the current setting of the parameter search_path:

  SHOW search_path;

Show the current setting of all parameters:

  SHOW ALL;

Compatibility

SHOW is a Greenplum Database extension.

See Also

SET, RESET
**START TRANSACTION**

Starts a transaction block.

---

**Synopsis**

\[
\text{START TRANSACTION [SERIALIZABLE | REPEATABLE READ | READ COMMITTED | READ UNCOMMITTED] [READ WRITE | READ ONLY]}
\]

---

**Description**

`START TRANSACTION` begins a new transaction block. If the isolation level or read/write mode is specified, the new transaction has those characteristics, as if `SET TRANSACTION` was executed. This is the same as the `BEGIN` command.

**Parameters**

- `SERIALIZABLE`
- `REPEATABLE READ`
- `READ COMMITTED`
- `READ UNCOMMITTED`

The SQL standard defines four transaction isolation levels: `READ COMMITTED`, `READ UNCOMMITTED`, `SERIALIZABLE`, and `REPEATABLE READ`. The default behavior is that a statement can only see rows committed before it began (`READ COMMITTED`). In Greenplum Database, `READ UNCOMMITTED` is treated the same as `READ COMMITTED`. `SERIALIZABLE` is supported the same as `REPEATABLE READ` wherein all statements of the current transaction can only see rows committed before the first statement was executed in the transaction. `SERIALIZABLE` is the strictest transaction isolation. This level emulates serial transaction execution, as if transactions had been executed one after another, serially, rather than concurrently. Applications using this level must be prepared to retry transactions due to serialization failures.

- `READ WRITE`
- `READ ONLY`

Determines whether the transaction is read/write or read-only. Read/write is the default. When a transaction is read-only, the following SQL commands are disallowed: `INSERT`, `UPDATE`, `DELETE`, and `COPY FROM` if the table they would write to is not a temporary table; all `CREATE`, `ALTER`, and `DROP` commands; `GRANT`, `REVOKE`, `TRUNCATE`; and `EXPLAIN ANALYZE` and `EXECUTE` if the command they would execute is among those listed.

---

**Examples**

To begin a transaction block:

```
START TRANSACTION;
```
**Compatibility**

In the standard, it is not necessary to issue `START TRANSACTION` to start a transaction block: any SQL command implicitly begins a block. Greenplum Database behavior can be seen as implicitly issuing a `COMMIT` after each command that does not follow `START TRANSACTION` (or `BEGIN`), and it is therefore often called ‘autocommit’. Other relational database systems may offer an autocommit feature as a convenience.

The SQL standard requires commas between successive `transaction_modes`, but for historical reasons Greenplum Database allows the commas to be omitted.

See also the compatibility section of `SET TRANSACTION`.

---

**See Also**

`BEGIN, SET TRANSACTION`
TRUNCATE

Empties a table of all rows.

Synopsis
TRUNCATE [TABLE] name [, ...] [CASCADE | RESTRICT]

Description
TRUNCATE quickly removes all rows from a table or set of tables. It has the same effect as an unqualified DELETE on each table, but since it does not actually scan the tables it is faster. This is most useful on large tables.

Parameters
name
The name (optionally schema-qualified) of a table to be truncated.

CASCADE
Since this key word applies to foreign key references (which are not supported in Greenplum Database) it has no effect.

RESTRICT
Since this key word applies to foreign key references (which are not supported in Greenplum Database) it has no effect.

Notes
Only the owner of a table may TRUNCATE it.
TRUNCATE will not run any user-defined ON DELETE triggers that might exist for the tables.
TRUNCATE will not truncate any tables that inherit from the named table. Only the named table is truncated, not its child tables.

Examples
Empty the table films:

    TRUNCATE films;

Compatibility
There is no TRUNCATE command in the SQL standard.

See Also
DELETE, DROP TABLE
**UPDATE**

Updates rows of a table.

**Synopsis**

```
UPDATE [ONLY] table [[AS] alias]
    SET {column = {expression | DEFAULT} |
    (column [, ...]) = ({expression | DEFAULT} [, ...])} [, ...]
    [FROM fromlist]
    [WHERE condition]
```

**Description**

`UPDATE` changes the values of the specified columns in all rows that satisfy the condition. Only the columns to be modified need be mentioned in the `SET` clause; columns not explicitly modified retain their previous values.

By default, `UPDATE` will update rows in the specified table and all its subtables. If you wish to only update the specific table mentioned, you must use the `ONLY` clause.

There are two ways to modify a table using information contained in other tables in the database: using sub-selects, or specifying additional tables in the `FROM` clause. Which technique is more appropriate depends on the specific circumstances. Note that if you do join two or more tables in an `UPDATE` command, those tables must have the same Greenplum distribution key column(s) and also be joined on the distribution key column(s).

You must have the `UPDATE` privilege on the table to update it, as well as the `SELECT` privilege to any table whose values are read in the expressions or condition.

**Outputs**

On successful completion, an `UPDATE` command returns a command tag of the form:

```
UPDATE count
```

Where count is the number of rows updated. If count is 0, no rows matched the condition (this is not considered an error).

**Parameters**

**ONLY**

If specified, update rows from the named table only. When not specified, any tables inheriting from the named table are also processed.

**table**

The name (optionally schema-qualified) of an existing table.
**alias**
A substitute name for the target table. When an alias is provided, it completely hides the actual name of the table. For example, given `UPDATE foo AS f`, the remainder of the `UPDATE` statement must refer to this table as `f` not `foo`.

**column**
The name of a column in table. The column name can be qualified with a subfield name or array subscript, if needed. Do not include the table’s name in the specification of a target column.

**expression**
An expression to assign to the column. The expression may use the old values of this and other columns in the table.

**DEFAULT**
Set the column to its default value (which will be NULL if no specific default expression has been assigned to it).

**fromlist**
A list of table expressions, allowing columns from other tables to appear in the `WHERE` condition and the update expressions. This is similar to the list of tables that can be specified in the `FROM` clause of a `SELECT` statement. Note that the target table must not appear in the `fromlist`, unless you intend a self-join (in which case it must appear with an alias in the `fromlist`).

**condition**
An expression that returns a value of type boolean. Only rows for which this expression returns true will be updated.

**output_expression**
An expression to be computed and returned by the `UPDATE` command after each row is updated. The expression may use any column names of the table or table(s) listed in `FROM`. Write `*` to return all columns.

**output_name**
A name to use for a returned column.

---

**Notes**
*SET* is not allowed on the Greenplum distribution key columns of a table.

You cannot use `STABLE` or `VOLATILE` functions in an `UPDATE` statement if mirrors are enabled. This can potentially cause the primary segment and its mirror to become out-of-sync because the command is run first on the primary and then a second time on the mirror in the current Greenplum Database implementation.

When a `FROM` clause is present, what essentially happens is that the target table is joined to the tables mentioned in the from list, and each output row of the join represents an update operation for the target table. When using `FROM` you should
ensure that the join produces at most one output row for each row to be modified. In other words, a target row should not join to more than one row from the other table(s). If it does, then only one of the join rows will be used to update the target row, but which one will be used is not readily predictable.

Because of this indeterminacy, referencing other tables only within sub-selects is safer, though often harder to read and slower than using a join.

Note also that in Greenplum Database, any tables that are joined in an update operation must have the same Greenplum distribution key and be joined on the column(s) that make up the distribution key (equijoin).

### Examples

Change the word *Drama* to *Dramatic* in the column *kind* of the table *films*:

```
UPDATE films SET kind = 'Dramatic' WHERE kind = 'Drama';
```

Adjust temperature entries and reset precipitation to its default value in one row of the table *weather*:

```
UPDATE weather SET temp_lo = temp_lo+1, temp_hi = temp_lo+15, prcp = DEFAULT
WHERE city = 'San Francisco' AND date = '2006-07-03';
```

Use the alternative column-list syntax to do the same update:

```
UPDATE weather SET (temp_lo, temp_hi, prcp) = (temp_lo+1, temp_lo+15, DEFAULT)
WHERE city = 'San Francisco' AND date = '2006-07-03';
```

Increment the sales count of the salesperson who manages the account for Acme Corporation, using the FROM clause syntax (assuming both tables being joined are distributed in Greenplum Database on the *id* column):

```
UPDATE employees SET sales_count = sales_count + 1 FROM accounts
WHERE accounts.name = 'Acme Corporation'
AND employees.id = accounts.id;
```

Perform the same operation, using a sub-select in the WHERE clause:

```
UPDATE employees SET sales_count = sales_count + 1 WHERE id =
(SELECT id FROM accounts WHERE name = 'Acme Corporation');
```

Attempt to insert a new stock item along with the quantity of stock. If the item already exists, instead update the stock count of the existing item. To do this without failing the entire transaction, use savepoints.

```
BEGIN;
-- other operations
SAVEPOINT sp1;
INSERT INTO wines VALUES('Chateau Lafite 2003', '24');
-- Assume the above fails because of a unique key violation,
-- so now we issue these commands:
```
ROLLBACK TO sp1;
UPDATE wines SET stock = stock + 24 WHERE winename = 'Chateau Lafite 2003';
-- continue with other operations, and eventually
COMMIT;

Compatibility
This command conforms to the SQL standard, except that the FROM clause is a
Greenplum Database extension.

According to the standard, the column-list syntax should allow a list of columns to be
assigned from a single row-valued expression, such as a sub-select:

```
UPDATE accounts SET (contact_last_name, contact_first_name) =
(SELECT last_name, first_name FROM salesmen
WHERE salesmen.id = accounts.sales_id);
```

This is not currently implemented — the source must be a list of independent
expressions.

Some other database systems offer a FROM option in which the target table is supposed
to be listed again within FROM. That is not how Greenplum Database interprets FROM.
Be careful when porting applications that use this extension.

See Also
DELETE, SELECT, INSERT
**VACUUM**

Garbage-collects and optionally analyzes a database.

**Synopsis**

```
VACUUM [FULL] [FREEZE] [VERBOSE] [table]
VACUUM [FULL] [FREEZE] [VERBOSE] ANALYZE
    [table [(column [, ...] )]]
```

**Description**

VACUUM reclaims storage occupied by deleted tuples. In normal Greenplum Database operation, tuples that are deleted or obsoleted by an update are not physically removed from their table; they remain present on disk until a VACUUM is done. Therefore it is necessary to do VACUUM periodically, especially on frequently-updated tables.

With no parameter, VACUUM processes every table in the current database. With a parameter, VACUUM processes only that table.

`VACUUM ANALYZE` performs a VACUUM and then an ANALYZE for each selected table. This is a handy combination form for routine maintenance scripts. See `ANALYZE` for more details about its processing.

Plain VACUUM (without FULL) simply reclaims space and makes it available for re-use. This form of the command can operate in parallel with normal reading and writing of the table, as an exclusive lock is not obtained. VACUUM FULL does more extensive processing, including moving of tuples across blocks to try to compact the table to the minimum number of disk blocks. This form is much slower and requires an exclusive lock on each table while it is being processed.

**Outputs**

When VERBOSE is specified, VACUUM emits progress messages to indicate which table is currently being processed. Various statistics about the tables are printed as well.

**Parameters**

**FULL**

Selects a full vacuum, which may reclaim more space, but takes much longer and exclusively locks the table.

*Warning:* A VACUUM FULL is not recommended in Greenplum Database. See the “Notes” section.

**FREEZE**

Specifying FREEZE is equivalent to performing VACUUM with the `vacuum_freeze_min_age` server configuration parameter set to zero. The FREEZE option is deprecated and will be removed in a future release. Set the parameter in the master `postgresql.conf` file instead.
**VACUUM**

Prints a detailed vacuum activity report for each table.

**ANALYZE**

Updates statistics used by the planner to determine the most efficient way to execute a query.

**table**

The name (optionally schema-qualified) of a specific table to vacuum. Defaults to all tables in the current database.

**column**

The name of a specific column to analyze. Defaults to all columns.

---

**Notes**

**VACUUM** cannot be executed inside a transaction block.

Greenplum recommends that active production databases be vacuumed frequently (at least nightly), in order to remove expired rows. After adding or deleting a large number of rows, it may be a good idea to issue a **VACUUM ANALYZE** command for the affected table. This will update the system catalogs with the results of all recent changes, and allow the Greenplum query planner to make better choices in planning queries.

**VACUUM** causes a substantial increase in I/O traffic, which can cause poor performance for other active sessions. Therefore, it is advisable to vacuum the database at low usage times.

Regular PostgreSQL has a separate optional server process called the **autovacuum daemon**, whose purpose is to automate the execution of **VACUUM** and **ANALYZE** commands. This feature is currently disabled in Greenplum Database.

Expired rows are held in what is called the **free space map**. The free space map must be sized large enough to cover the dead rows of all tables in your database. If not sized large enough, space occupied by dead rows that overflow the free space map cannot be reclaimed by a regular **VACUUM** command.

A **VACUUM FULL** will reclaim all expired row space, but is a very expensive operation and may take an unacceptably long time to finish on large, distributed Greenplum Database tables. If you do get into a situation where the free space map has overflowed, it may be more timely to recreate the table with a **CREATE TABLE AS** statement and drop the old table. A **VACUUM FULL** is not recommended in Greenplum Database.

It is best to size the free space map appropriately. The free space map is configured with the following server configuration parameters:

```bash
max_fsm_pages
max_fsm_relations
```

See “About Concurrency Control in Greenplum Database” on page 95 for more information.
Examples
Vacuum all tables in the current database:
    VACUUM;
Vacuum a specific table only:
    VACUUM mytable;
Vacuum all tables in the current database and collect statistics for the query planner:
    VACUUM ANALYZE;

Compatibility
There is no VACUUM statement in the SQL standard.

See Also
ANALYZE
VALUES

Computes a set of rows.

**Synopsis**

VALUES ( expression [, ...] ) [, ...]
[ORDER BY sort_expression [ASC | DESC | USING operator] [, ...]]
[LIMIT {count | ALL}] [OFFSET start]

**Description**

VALUES computes a row value or set of row values specified by value expressions. It is most commonly used to generate a ‘constant table’ within a larger command, but it can be used on its own.

When more than one row is specified, all the rows must have the same number of elements. The data types of the resulting table’s columns are determined by combining the explicit or inferred types of the expressions appearing in that column, using the same rules as for UNION.

Within larger commands, VALUES is syntactically allowed anywhere that SELECT is. Because it is treated like a SELECT by the grammar, it is possible to use the ORDER BY, LIMIT, and OFFSET clauses with a VALUES command.

**Parameters**

*expression*

A constant or expression to compute and insert at the indicated place in the resulting table (set of rows). In a VALUES list appearing at the top level of an INSERT, an expression can be replaced by DEFAULT to indicate that the destination column’s default value should be inserted. DEFAULT cannot be used when VALUES appears in other contexts.

*sort_expression*

An expression or integer constant indicating how to sort the result rows. This expression may refer to the columns of the VALUES result as column1, column2, etc. For more details see “The ORDER BY Clause” on page 525.

*operator*

A sorting operator. For details see “The ORDER BY Clause” on page 525.

*LIMIT count*

OFFSET start

The maximum number of rows to return. For details see “The LIMIT Clause” on page 526.
Notes
VALUES lists with very large numbers of rows should be avoided, as you may encounter out-of-memory failures or poor performance. VALUES appearing within INSERT is a special case (because the desired column types are known from the INSERT's target table, and need not be inferred by scanning the VALUES list), so it can handle larger lists than are practical in other contexts.

Examples
A bare VALUES command:

VALUES (1, 'one'), (2, 'two'), (3, 'three');

This will return a table of two columns and three rows. It is effectively equivalent to:

SELECT 1 AS column1, 'one' AS column2
UNION ALL
SELECT 2, 'two'
UNION ALL
SELECT 3, 'three';

More usually, VALUES is used within a larger SQL command. The most common use is in INSERT:

INSERT INTO films (code, title, did, date_prod, kind)
VALUES ('T_601', 'Yojimbo', 106, '1961-06-16', 'Drama');

In the context of INSERT, entries of a VALUES list can be DEFAULT to indicate that the column default should be used here instead of specifying a value:

INSERT INTO films VALUES
('UA502', 'Bananas', 105, DEFAULT, 'Comedy', '82 minutes'),
('T_601', 'Yojimbo', 106, DEFAULT, 'Drama', DEFAULT);

VALUES can also be used where a sub-SELECT might be written, for example in a FROM clause:

SELECT f.* FROM films f, (VALUES('MGM', 'Horror'), ('UA', 'Sci-Fi')) AS t (studio, kind) WHERE f.studio = t.studio AND f.kind = t.kind;
UPDATE employees SET salary = salary * v.increase FROM (VALUES(1, 200000, 1.2), (2, 400000, 1.4)) AS v (depno, target, increase) WHERE employees.depno = v.depno AND employees.sales >= v.target;

Note that an AS clause is required when VALUES is used in a FROM clause, just as is true for SELECT. It is not required that the AS clause specify names for all the columns, but it is good practice to do so. The default column names for VALUES are column1, column2, etc. in Greenplum Database, but these names might be different in other database systems.
When `VALUES` is used in `INSERT`, the values are all automatically coerced to the data type of the corresponding destination column. When it is used in other contexts, it may be necessary to specify the correct data type. If the entries are all quoted literal constants, coercing the first is sufficient to determine the assumed type for all:

```
SELECT * FROM machines WHERE ip_address IN
  (VALUES('192.168.0.1'::inet), ('192.168.0.10'),
   ('192.168.1.43'));
```

**Note:** For simple `IN` tests, it is better to rely on the list-of-scalars form of `IN` than to write a `VALUES` query as shown above. The list of scalars method requires less writing and is often more efficient.

**Compatibility**

`VALUES` conforms to the SQL standard, except that `LIMIT` and `OFFSET` are Greenplum Database extensions.

**See Also**

`INSERT`, `SELECT`
B. Management Utility Reference

This appendix provides references for the command-line management utilities provided with Greenplum Database. Greenplum Database utilizes the standard PostgreSQL client and server programs, and also has additional management utilities to facilitate the administration of a distributed Greenplum Database DBMS.

The following Greenplum Database management utilities are located in $GPHOME/bin:

- gp_dump
- gp_restore
- gpactivatestandby
- gpaddmirrors
- gpcheck
- gpchecknet (deprecated)
- gpcheckos (deprecated)
- gpcheckperf
- gpcrondump
- gpconfig
- gpdbrestore
- gpdeleteesystem
- gpdetective
- gpexpand
- gpfdist
- gpfilesystem
- gpinitstandby
- gpinitssystem
- gpload
- gplogfilter
- gpmapreduce
- gpmigrator
- gprecoverseg
- gprebuildsystem (deprecated)
- gpsizecalc (deprecated)
- gpscp
- gpskew (deprecated)
- gsnmpd
- gpssh
- gpssh-exkeys
- gpstart
- gpstate
- gpstop
Backend Server Programs

The following server programs are also located in $GPHOME/bin of your Greenplum Database installation. These are the standard PostgreSQL server programs, which have been modified to handle the parallelism and distribution of a Greenplum Database system. Keep in mind that Greenplum Database is essentially several PostgreSQL database instances working together as a single DBMS, so Greenplum Database relies on PostgreSQL for its underlying functionality. Users and administrators do not access these programs directly, but do so through the Greenplum Database management tools and utilities.

<table>
<thead>
<tr>
<th>Program Name</th>
<th>Description</th>
<th>Use Instead</th>
</tr>
</thead>
<tbody>
<tr>
<td>initdb</td>
<td>This program is called by gpinitsystem when initializing a Greenplum Database array. It is used internally to create the individual segment instances and the master instance.</td>
<td>gpinitsystem</td>
</tr>
<tr>
<td>ipcclean</td>
<td>Not used in Greenplum Database</td>
<td>N/A</td>
</tr>
<tr>
<td>gpsyncmaster</td>
<td>This is the Greenplum program that starts the gpsyncagent process on the standby master host. Administrators do not call this program directly, but do so through the management scripts that initialize and/or activate a standby master for a Greenplum Database system. This process is responsible for keeping the standby master up to date with the primary master via a transaction log replication process.</td>
<td>gpinitstandby, gpactivatestandby</td>
</tr>
<tr>
<td>pg_controldata</td>
<td>Not used in Greenplum Database</td>
<td>gpstate</td>
</tr>
<tr>
<td>pg_ctl</td>
<td>This program is called by gpstart and gpstop when starting or stopping a Greenplum Database array. It is used internally to stop and start the individual segment instances and the master instance in parallel and with the correct options.</td>
<td>gpstart, gpstop</td>
</tr>
<tr>
<td>pg_resetxlog</td>
<td>Not used in Greenplum Database</td>
<td>N/A</td>
</tr>
<tr>
<td>postgres</td>
<td>The postgres executable is the actual PostgreSQL server process that processes queries.</td>
<td>The main postgres process (postmaster) creates other postgres subprocesses and postgres session as needed to handle client connections.</td>
</tr>
<tr>
<td>postmaster</td>
<td>postmaster starts the postgres database server listener process that accepts client connections. In Greenplum Database, a postgres database listener process runs on the Greenplum Master Instance and on each Segment Instance.</td>
<td>In Greenplum Database, you use gpstart and gpstop to start all postmasters (postgres processes) in the system at once in the correct order and with the correct options.</td>
</tr>
</tbody>
</table>
Management Utility Summary

**gp_dump**

Writes out a database to SQL script files, which can then be used to restore the database using `gp_restore`.

```
```

```
gp_dump -? | --help
```

```
gp_dump --version
```

```
-a | --data-only
-s | --schema-only
-c | --clean
-d | --inserts
-D | --column-inserts
-n schema | --schema=schema
-o | --oids
-O | --no-owner
-t table | --table=table
-x | --no-privileges | --no-acl
-h hostname | --host=hostname
-p port | --port=port
-U username | --username=user
-W (force password prompt)
-i | --ignore-version
-v | --verbose
--gp-c (use gzip)
--gp-d=directoryname
--gp-r=reportfile
--gp-s=dbid (backup certain segments)
```

**gp_restore**

Restores Greenplum databases that were backed up using `gp_dump`.

```
```

```
gp_dump
```

558
gpaddmirrors

Adds mirror segments to a Greenplum Database system that was initially configured without mirroring.


gpaddmirrors -i mirror_config_file [-s] [-a] [-d master_data_directory] [-B

[--gp-r=reportfile] [--gp-l=dbid [, ...]]

gp_restore -? | -h | --help

gp_restore --version

--gp-k=timestamp_key

-d database_name | --dbname=dbname

-i | --ignore-version

-v | --verbose

-a | --data-only

-c | --clean

-s | --schema-only

-h hostname | --host=hostname

-p port | --port=port

-U username | --username=username

-W (force password prompt)

--gp-c (use gunzip)

--gp-i (ignore errors)

--gp-d=directoryname

--gp-r=reportfile

--gp-l=dbid [, ...] (restore certain segments)

-? | -h | --help (help)

--version (show utility version)
parallel_processes] [-l logfile_directory] [-v]
gpaddmirrors -o output_sample_mirror_config [-m datadir_config_file]
gpaddmirrors -?
gpaddmirrors --version
-a (do not prompt)
-B parallel_processes
d-master_data_directory
-i mirror_config_file
-l logfile_directory
-m datadir_config_file
-o output_sample_mirror_config
-p port_offset
-s (spread mirrors)
-v (verbose)
--version (show utility version)
?- (help)

gpactivatestandby
Activates a standby master host and makes it the active master for the Greenplum Database system.
gpactivatestandby -d standby_master_datadir [-c new_standby_master] [-f] [-a]
[ -q] [-l logfile_directory]
gpactivatestandby -? | -h | --help
gpactivatestandby -v
-a (do not prompt)
c new_standby_master_hostname
d standby_master_datadir
-f (force activation)
l logfile_directory
-q (no screen output)
v (show utility version)
?- | -h | --help (help)
**gpbitmapreindex**

Rebuilds bitmap indexes after a 3.3.x to 4.0.x upgrade.

```
gpbitmapreindex [-m {r | d | l [-o output_sql_file]} ]
[-h master_host] [-p master_port] [-n number_of_processes] [-v]
gpmigrator --version
```

**gpmigrator**

```
gpmigrator --help | -?
-m {r|d|l} | --mode {reindex|drop|list}
-n number_of_processes | --parallel number_of_processes
-o output_sql_file | --outfile output_sql_file
-v | --verbose
--version
-? | --help
```

**gpcheck**

Verifies and validates Greenplum Database platform settings.

```
gpcheck -f hostfile_gpcheck [-m master_host] [-s standby_master_host] [-stdout | --zipout] [-v]
--config config_file
```

```
gpcheck --zipin gpcheck_zipfile

gpcheck -?
```

```
gpcheck --version
```

```
gpcheck --config config_file
-f hostfile_gpcheck
-m master_host
-s standby_master_host
-stdout
--zipout
--zipin gpcheck_zipfile
-? (help)
```

```
--version
```
gpcheckperf

Verifies the baseline hardware performance of the specified hosts.

gpcheckperf -d test_directory [-d test_directory ...]
  {-f hostfile_gpcheckperf | - h hostname [-h hostname ...]}
  [-r ds] [-B Block_size] [-S file_size] [-D] [-v|-V]

gpcheckperf -d temp_directory
  {-f hostfile_gpchecknet | - h hostname [-h hostname ...]}
  [-r n|M|--duration time] [--netperf] [-D] [-v|-V]

gpcheckperf -?

gpcheckperf --version

-B block_size
-d test_directory
-d temp_directory
-D (display per-host results)
--duration time
-f hostfile_gpcheckperf
-f hostfile_gpchecknet
-h hostname
--netperf
-r ds{n|M}
-S file_size
-v (verbose) | -V (very verbose)
--version
-? (help)
gpconfig

Sets server configuration parameters on all segments within a Greenplum Database system.

```
gpconfig -c param_name -v value [-m master_value | --masteronly]
    | -r param_name [--masteronly]
    | -l
    [--skipvalidation] [--verbose] [--debug]

gpconfig -s param_name [--verbose] [--debug]

gpconfig --help
```

-gpcrondump

A wrapper utility for gp_dump, which can be called directly or from a crontab entry.

```
gpcrondump -x database_name
            [-s schema | -t schema.table | -T schema.table]
            [-w dbid[, ...]] [-u backup_directory]
            [-B parallel_processes] [-a] [-q]
            [-y reportfile] [-l logfile_directory] [-D]
            [-E encoding] [-inserts | --column-inserts] [--oids] [--no-owner |
gpdbrestore
A wrapper utility around gp_restore. Restores a database from a set of dump files generated by
gpcreondump.

`gpdbrestore` { -t timestamp_key [ -L ] | -b YYYYMMDD | -R hostname:path_to_dumpset | -s database_name } [-T schema.table [...] ] [ -e ] [ -G ] [ -B parallel_processes ] [ -d master_data_directory ] [ -a ] [ -q ] [ -l logfile_directory ] [ -D ]

`gpdbrestore -?`

`gpdbrestore -v`

- `a` (do not prompt)
- `-b YYYYMMDD`
- `B parallel_processes`
- `-d master_data_directory`
- `-D` (debug)
- `-e` (drop target database before restore)
- `-G` (restore global objects)
- `-l logfile_directory`
- `-L` (list tablenames in backup set)
- `-q` (no screen output)
- `-R hostname:path_to_dumpset`
- `-s database_name`
- `-t timestamp_key`
- `-T schema.table_name`
- `-v` (show utility version)
- `-?` (help)

`gpdeletesystem`

Deletes a Greenplum Database system that was initialized using `gpinitsystem`.

`gpdeletesystem -d master_data_directory [ -B parallel_processes ] [ -f ] [ -l logfile_directory ] [ -D ]`

`gpdeletesystem -?`

`gpdeletesystem -v`

- `-d data_directory`
- `B parallel_processes`
- `-f` (force)
- `-l logfile_directory`
- `-D` (debug)
- `-?` (help)
- `-v` (show utility version)
**gpdetective**

Collects diagnostic information from a running Greenplum Database system.

```
gpdetective [-h hostname] [-p port] [-U username] [-P password]
            [---start_date number_of_days | YYYY-MM-DD]
            [---end_date YYYY-MM-DD]
            [---diagnostics a|n|s|o|c]
            [---logs a|n|dbid[,dbid,... | -dbid]]
            [---cores t|f]
            [---pg_dumpall t|f] [---pg_dump_options option[,...]]
            [---tempdir temp_dir]
            [---connect t|f]
```

```
gpdetective -?
gpdetective -v
--connect t|f
--cores t|f
--diagnostics a|n|s|o|c
--end_date YYYY-MM-DD
-h hostname
--logs a|n|dbid_list
--pg_dumpall t|f
--pg_dump_options option[,...]
-p password
--start_date number_of_days | YYYY-MM-DD
--tempdir temp_dir
-U gp_superuser
-v (show utility version)
-? (help)
```
**gpexpand**

Expands an existing Greenplum Database across new hosts in the array.

```
gpexpand
  [-f hosts_file]  # Use a file to specify the hosts
  [-i input_file]  # Use an input file
  [-B batch_size]  # specify the batch size
  [-V]            # verbose output
  [-d hh:mm:ss]   # specify the duration
  [-e 'YYYY-MM-DD hh:mm:ss']  # end date and time
  [-n parallel_processes]  # number of parallel processes
  [-r]            # rollback
  [-s]            # silent
  [-v]            # very verbose

  -D database_name  # database name
  --rollback
  --clean

  gpexpand -? | -h | --help
  gpexpand --version
```

**gpfdist**

Serves data files to or writes data files out from Greenplum Database segments.

```

gpfdist -?
  gpfdist --version
  -d directory
  -l log_file
  -p http_port
  -t timeout
  -m max_length
  -? (help)
  -v (verbose)
  -V (very verbose)
  --version
```
**gpfilespace**

Creates a filespace using a configuration file that defines per-segment file system locations. Filespaces describe the physical file system resources to be used by a tablespace.

```
gpfilespace [connection_option ...] [-l logfile_directory] [-o output_file_name]
gpfilespace [connection_option ...] [-l logfile_directory] -c fs_config_file
```

```
gpfilespace -v | -?  
-c | --config fs_config_file  
-l | --logdir logfile_directory  
-o | --output output_file_name  
-v | --version (show utility version)  
-? | --help (help)  
-h host | --host host  
-p port | --port port  
-U username | --username superuser_name  
-W | --password
```

**gpinitstandby**

Adds and/or initializes a standby master host for a Greenplum Database system.

```
gpinitstandby -? | -v
```

```
-a (do not prompt)  
-D (debug)  
-l logfile_directory  
-L (leave database stopped)  
-M fast (fast shutdown - rollback)  
-M smart (smart shutdown - warn)  
-n (resynchronize)  
-q (no screen output)  
-r (remove standby master)  
-s standby_hostname  
-v (show utility version)  
-? (help)
```

**gpinitsystem**

Initializes a Greenplum Database system using configuration parameters specified in the
gpinit system_config file.

```
gpinit system -c gpinit system_config
   [-h hostfile_gpinit system] [-o]
   [-B parallel_processes]
   [-p postgresql conf_param_file]
   [-s standby master host]
   [--max_connections number] [--shared_buffers size]
   [--locale locale] [--lc-collate locale]
   [--lc-ctype locale] [--lc-messages locale]
   [--lc-monetary locale] [--lc-numeric locale]
   [--lc-time locale] [--su_password password]
gpinit system -?
gpinit system -v
   -a (do not prompt)
   -B parallel_processes
   -c gpinit system_config
   -D (debug)
   -h hostfile_gpinit system
   --locale | -n locale
   --lc-collate locale
   --lc-ctype locale
   --lc-messages locale
   --lc-monetary locale
   --lc-numeric locale
   --lc-time locale
   -l logfile_directory
   --max_connections | -m number
   -o (create master only)
   -p postgresql conf_param_file
   -q (no screen output)
   --shared_buffers | -b size
   -s standby master host
   --su_password | -e superuser_password
   -S (spread mirror configuration)
   -v (show utility version)
   -? (help)
```

gpload

Runs a load job as defined in a YAML formatted control file.

```
```
gpload

-?
gpload --version
-f control_file
--gpfdist_timeout seconds
-l log_file
-v (verbose mode)
-V (very verbose mode)
-q (no screen output)
-D (debug mode)
-? (show help)
--version
-d database
-h hostname
-p port
-U username
-W (force password prompt)

gplogfilter

Searches through Greenplum Database log files for specified entries.

gplogfilter [timestamp_options] [pattern_options] [output_options]
gplogfilter --help
gplogfilter --version
-b datetime | --begin=datetime
-e datetime | --end=datetime
-d time | --duration=time
-c i[gnore]|r[espect] | --case=i[gnore]|r[espect]
-C '<string>' | --columns='<string>'
-f 'string' | --find='string'
-F 'string' | --nofind='string'
-m regex | --match=regex
-M regex | --nomatch=regex
-t | --trouble
-n integer | --tail=integer
-s offset [limit] | --slice=offset [limit]
-o output_file | --out=output_file
-z 0-9 | --zip=0-9
-a | --append
input_file
-u | --unzip
--help
--version

gpmapreduce
Runs Greenplum MapReduce jobs as defined in a YAML specification document.

gpmapreduce -f yaml_file [dbname [username]] [-k name=value | --key name=value]
gpmapreduce -V | --version
gpmapreduce -h | --help
-f yaml_file
-? | --help
-V | --version
-v | --verbose
-k | --key name=value
-h host | --host host
-p port | --port port
-U username | --username username
-W | --password
**gpmigrator**

Upgrades an existing Greenplum Database 3.3.x system to 4.0.x

```
gpmigrator old_GPHOME_path new_GPHOME_path
   --replication_port_base port
   --mirror_replication_port_base port
   --mirror-mode {redundant | single}
   --gpperfmonhome path
   [--nosnapshot] [-d master_data_directory]
   [-l logfile_directory] [-q] [--debug] [-R]
```

```
gpmigrator old_GPHOME_path new_GPHOME_path --resume
```

```
gpmigrator --version | -v
```

```
gpmigrator --help | -h
```

```
old_GPHOME_path
new_GPHOME_path
   --replication_port_base port
   --mirror_replication_port_base port
   --mirror-mode {redundant | single}
   --gpperfmonhome path
   --nosnapshot
   --resume
   -R (revert)
   -d master_data_directory
   -l logfile_directory
   -q (quiet mode)
   --help | -h
   --debug
   --version | -v
```

**gpperfmon_install**

Installs the gpperfmon database and optionally enables the data collection agents.

```
gpperfmon_install
   [--enable --password gpmon_password --port gpdb_port]
   [--pgpass path_to_file]
   [--verbose]
```

```
gpperfmon_install --help | -h | -?
```

```
--enable
--password gpmon_password
--port gpdb_port
--pgpass path_to_file
--verbose
--help | -h | -?
```
gprecoverseg
Recovers a primary or mirror segment instance that has been marked as down (if mirroring is enabled).

```
gprecoverseg -o output_recover_config_file
       | -S output_filespace_config_file [-p new_recover_host]
gprecoverseg -?
gprecoverseg --version
-a (do not prompt)
-B parallel_processes
-d master_data_directory
-F (full recovery)
-i recover_config_file
-l logfile_directory
-o output_recover_config_file
-p new_recover_host
-q (no screen output)
-s filespace_config_file
-S output_filespace_config_file
-v (verbose)
--version (version)
-? (help)
```

gpscp
Copies files between multiple hosts at once.

```
gpscp [-f hostfile_gpssh | - h hostname [-h hostname ...] ] [-J character] [-v] [[user@hostname:]file_to_copy [...] [[user@host-
name:]copy_to_path]
gpscp -?
gpscp --version
-f hostfile_gpssh
-h hostname
-J character
-v (verbose mode)
file_to_copy
copy_to_path
-? (help)
--version
```
gpseginstall
Installs the Greenplum Database software on multiple hosts.

```
gpseginstall -f hostfile_exkeys [-u gpdb_admin_user] [-p password] 
  [-c u|i|p|c|s|E|e|l|v]
gpseginstall --help
-c | --commands option_list
-f | --file hostfile_exkeys
-p | --password password
-u | --user user
--help (help)
```

gpsetupsanfailover
Specify failover host/mountpoints for Greenplum Database in a SAN environment.

```
gpsetupsanfailover <output_san_sql_file>
gpsetupsanfailover [-d master_data_directory]
  <output_san_sql_file> [-p port_offset] [-D] 
  [-l logfile_directory]
gpsetupsanfailover --version
 gpsetupsanfailover -?
-d master_data_directory
-l logfile_directory
-p port_offset
-D (debug)
--version (show utility version)
-? (help)
```

gpsnmpd
Reports on the health and state of a Greenplum Database system through SNMP.

```
gpsnmpd -s -C connect_string [-b] [-g] [-m MIB:...] 
  [-M directory:...]
gpsnmpd -c FILE -C connect_string [-x address:port] [-b] [-g]
```
[-m MIB:...] [-M directory:...]
gpsnmpd -?
gpsnmpd --version
-s (sub-agent)
-b (background)
-c (configuration file)
-g (use syslog)
-C (libpq connection string)
-x (address:port of a network interface)
-m (MIB:...)
-M (directory:...)
-? (help)
-V

**gpssh**

Provides ssh access to multiple hosts at once.

```
gpssh { -f hostfile_gpssh | - h hostname [-h hostname ...] } [-v] [-e] [bash_command]
gpssh -?
gpssh --version
bash_command
-e (echo)
-f hostfile_gpssh
-h hostname
-v (verbose mode)
--version
-? (help)
```
**gpssh-exkeys**

Exchanges SSH public keys between hosts.

```
gpssh-exkeys -f hostfile_exkeys | - h hostname [-h hostname ...]
gpssh-exkeys -e hostfile_exkeys -x hostfile_gpexpand
```

**gpstart**

Starts a Greenplum Database system.

```
gpstart -? | -h | --help
```

**gpstate**

Shows the status of a running Greenplum Database system.

```
gpstate [-d master_data_directory] [-B parallel_processes]
```
gpstate

gpstate -? | -h | --help

-b (brief status)
-B parallel_processes
-c (show primary to mirror mappings)
-d master_data_directory
-e (show segments with mirror status issues)
-f (show standby master details)
-i (show Greenplum Database version)
-l logfile_directory
-m (list mirrors)
-p (show ports)
-q (no screen output)
-Q (quick status)
-s (detailed status)
-v (verbose output)
-? | -h | --help (help)

gpstop

Stops or restarts a Greenplum Database system.

gpstop [-d master_data_directory] [-B parallel_processes]
[-M smart | fast | immediate] [-t timeout_seconds] [-r] [-y] [-a]
[-l logfile_directory] [-v | -q]
gpstop -m [-d master_data_directory] [-y] [-l logfile_directory] [-v | -q]
gpstop -u [-d master_data_directory] [-l logfile_directory] [-v | -q]
gpstop --version

gpstop -? | -h | --help
-a (do not prompt)
-B parallel_processes
-d master_data_directory
-l logfile_directory
-m (master only)
-M fast (fast shutdown - rollback)
-M immediate (immediate shutdown - abort)
-M smart (smart shutdown - warn)
-q (no screen output)
-r (restart)
-t timeout_seconds
-u (reload pg_hba.conf and postgresql.conf files only)
-v (verbose output)
--version (show utility version)
-y (do not stop standby master)
-? | -h | --help (help)

gpsys1
Displays information about your operating system.

gpsys1 [-a | -m | -p ]
gpsys1 -? | --version
-a (show all)
-m (show memory only)
-p (show platform only)
-? (help)
--version
gp_dump

Writes out a database to SQL script files, which can then be used to restore the database using gp_restore.

Synopsis


gp_dump -? | --help

gp_dump --version

Description

The gp_dump utility dumps the contents of a database into SQL script files, which can then be used to restore the database schema and user data at a later time using gp_restore. During a dump operation, users will still have full access to the database.

The functionality of gp_dump is analogous to PostgreSQL’s pg_dump utility, which writes out (or dumps) the content of a database into a script file. The script file contains SQL commands that can be used to restore the databases, data, and global objects such as users, groups, and access permissions.

The functionality of gp_dump is modified to accommodate the distributed nature of a Greenplum database. Keep in mind that a database in Greenplum Database is actually comprised of several PostgreSQL instances (the master and all segments), each of which must be dumped individually. The gp_dump utility takes care of dumping all of the individual instances across the system.

The gp_dump utility performs the following actions and produces the following dump files by default:

On the master host

- Dumps the Greenplum Database system configuration tables into a SQL file in the master data directory. The default naming convention of this file is gp_catalog_1_<dbid>_<timestamp>.

- Dumps CREATE DATABASE SQL statements into a file in the master data directory. The default naming convention of this file is gp_cdatabase_1_<dbid>_<timestamp>. This statement can be run on the master instance to recreate the user database(s).

- Dumps the user database schema(s) into a SQL file in the master data directory. The default naming convention of this file is gp_dump_1_<dbid>_<timestamp>. This file is used by gp_restore to recreate the database schema(s).

- Creates a log file in the master data directory named gp_dump_status_1_<dbid>_<timestamp>. 
- gp_dump launches a gp_dump_agent for each segment instance to be backed up.
gp_dump_agent processes run on the segment hosts and report status back to the
gp_dump process running on the master host.

On the segment hosts
- Dumps the user data for each segment instance into a SQL file in the segment
instance’s data directory. By default, only primary (or active) segment instances
are backed up. The default naming convention of this file is
gp_dump_0_<dbid>_<timestamp>. This file is used by gp_restore to recreate
that particular segment of user data.
- Creates a log file in each segment instance’s data directory named
gp_dump_status_0_<dbid>_<timestamp>.

Note that the 14 digit timestamp is the number that uniquely identifies the backup job,
and is part of the filename for each dump file created by a gp_dump operation. This
timestamp must be passed to the gp_restore utility when restoring a Greenplum
database.

Options

-a | --data-only
Dump only the data, not the schema (data definitions).

-s | --schema-only
Dump only the object definitions (schema), not data.

-c | --clean
Output commands to clean (drop) database objects prior to (the commands for)
creating them.

-d | --inserts
Dump data as INSERT commands (rather than COPY). This will make restoration
very slow; it is mainly useful for making dumps that can be loaded into
non-PostgreSQL based databases. Note that the restore may fail altogether if you
have rearranged column order. The -D option is safer, though even slower.

-D | --column-inserts
Dump data as INSERT commands with explicit column names (INSERT INTO
table (column, ...) VALUES ...). This will make restoration very slow; it is
mainly useful for making dumps that can be loaded into non-PostgreSQL based
databases.

-n schema | --schema=schema
Dumps the contents of the named schema only. If this option is not specified, all
non-system schemas in the target database will be dumped. You cannot backup
system catalog schemas (such as pg_catalog) with gp_dump.

-
Caution: In this mode, gp_dump makes no attempt to dump any other database objects that objects in the selected schema may depend upon. Therefore, there is no guarantee that the results of a single-schema dump can be successfully restored by themselves into a clean database.

-o | --oids
Dump object identifiers (OIDs) as part of the data for every table. Use of OIDs is not recommended in Greenplum Database, so this option should not be used if restoring data to another Greenplum Database installation.

-O | --no-owner
Do not output commands to set ownership of objects to match the original database. By default, gp_dump issues ALTER OWNER or SET SESSION AUTHORIZATION statements to set ownership of created database objects. These statements will fail when the script is run unless it is started by a superuser (or the same user that owns all of the objects in the script). To make a script that can be restored by any user, but will give that user ownership of all the objects, specify -O.

-t table | --table=table
Dump data for the named table only. It is possible for there to be multiple tables with the same name in different schemas; if that is the case, all matching tables will be dumped. Specify both --schema and --table to select just one table.

Caution: In this mode, gp_dump makes no attempt to dump any other database objects that the selected table may depend upon. Therefore, there is no guarantee that the results of a single-table dump can be successfully restored by themselves into a clean database.

-x | --no-privileges | --no-acl
Prevents the dumping of access privileges (GRANT/REVOKE commands).

-h hostname | --host=hostname
The host name of the Greenplum Database master host. If not provided, the value of $PGHOST or the local host is used.

-p port | --port=port
The Greenplum Database master port. If not provided, the value of $PGPORT or the port number provided at compile time is used.

-U username | --username=user
The database superuser account name, for example gpadmin. If not provided, the value of $PGUSER or the current OS user name is used.

-W (force password prompt)
Forces a password prompt. This will happen automatically if the server requires password authentication.
-i  |  --ignore-version
    Ignores a version mismatch between gp_dump and the database server.

-v  |  --verbose
    Specifies verbose mode. This will cause gp_dump to output detailed object
    comments and start/stop times to the dump file, and progress messages to standard
    error.

--gp-c (use gzip)
    Use gzip for inline compression.

--gp-d=directoryname
    Specifies the relative or absolute path where the backup files will be placed on each
    host. If this is a relative path, it is considered to be relative to the data directory. If
    the path does not exist, it will be created, if possible. If not specified, defaults to the
    data directory of each instance to be backed up. Using this option may be desirable
    if each segment host has multiple segment instances — it will create the dump files
    in a centralized location.

--gp-r=reportfile
    Specifies the full path name where the backup job log file will be placed on the
    master host. If not specified, defaults to the master data directory or if running
    remotely, the current working directory.

--gp-s=dbid (backup certain segments)
    Specifies the set of active segment instances to back up with a comma-separated list
    of the segments dbid. The default is to backup all active segment instances.

database_name
    Required. The name of the database you want to dump. If not specified, the value of
    $PGDATABASE will be used. The database name must be stated last after all other
    options have been specified.

-?  |  --help (help)
    Displays the online help.

--version (show utility version)
    Displays the version of this utility.

Examples
Back up a database:
    gp_dump gpdb
Back up a database, and create dump files in a centralized location on all hosts:
    gp_dump --gp-d=/home/gpadmin/backups gpdb
Back up a particular schema only:
gp_dump -n myschema mydatabase

Back up a single segment instance only (by noting the dbid of the segment instance):

gp_dump --gp-s=5 gpdb

See Also

gp_restore, gpdbrestore, gpcrondump, pg_dump
gp_restore

Restores Greenplum databases that were backed up using gp_dump.

Synopsis

gp_restore -? | -h | --help
gp_restore --version```

Description

The `gp_restore` utility recreates the data definitions (schema) and user data in a Greenplum database using the script files created by an `gp_dump` operation. The use of this utility assumes:

1. You have backup files created by an `gp_dump` operation.
2. Your Greenplum Database system up and running.
3. Your Greenplum Database system has the exact same number of segment instances (primary and mirror) as the system that was backed up using `gp_dump`.
4. (optional) The `gp_restore` utility uses the information in the Greenplum system catalog tables to determine the hosts, ports, and data directories for the segment instances it is restoring. If you want to change any of this information (for example, move the system to a different array of hosts) you must use the `gpbuildsystem` and `gpbuildseg` scripts to reconfigure your array before restoring.
5. The databases you are restoring have been created in the system.
6. If you used the options `-s` (schema only), `-a` (data only), `-gp-c` (compressed), `-gp-d` (alternate dump file location) when performing the `gp_dump` operation, you must specify these options when doing the `gp_restore` as well.

The functionality of `gp_restore` is analogous to PostgreSQL’s `pg_restore` utility, which restores a database from files created by the database backup process. It issues the commands necessary to reconstruct the database to the state it was in at the time it was saved.

The functionality of `gp_restore` is modified to accommodate the distributed nature of a Greenplum database, and to use files created by an `gp_dump` operation. Keep in mind that a database in Greenplum is actually comprised of several PostgreSQL database instances (the master and all segments), each of which must be restored individually. The `gp_restore` utility takes care of populating each segment in the system with its own distinct portion of data.

The `gp_restore` utility performs the following actions:
On the master host

- Creates the user database schema(s) using the
  `gp_dump_1_<dbid>_<timestamp>` SQL file created by `gp_dump`.
- Creates a log file in the master data directory named
  `gp_restore_status_1_<dbid>_<timestamp>`.
- `gp_restore` launches a `gp_restore_agent` for each segment instance to be
  restored. `gp_restore_agent` processes run on the segment hosts and report
  status back to the `gp_restore` process running on the master host.

On the segment hosts

- Restores the user data for each segment instance using the
  `gp_dump_0_<dbid>_<timestamp>` files created by `gp_dump`. Each segment
  instance on a host (primary and mirror instances) are restored.
- Creates a log file for each segment instance named
  `gp_restore_status_0_<dbid>_<timestamp>`.

Note that the 14 digit timestamp is the number that uniquely identifies the backup job
be restored, and is part of the filename for each dump file created by a `gp_dump`
operation. This timestamp must be passed to the `gp_restore` utility when restoring a
Greenplum database.

Options

`--gp-k=timestamp_key`
Required. The 14 digit timestamp key that uniquely identifies the backup set of data
to restore. This timestamp can be found in the `gp_dump` log file output, as well as at
the end of the dump files created by a `gp_dump` operation. It is of the form
`YYYYMMDDHHMMSS`.

`-d database_name | --dbname=dbname`
Required. The name of the database to connect to in order to restore the user data.
The database(s) you are restoring must exist, `gp_restore` does not create the
database.

`-i | --ignore-version`
Ignores a version mismatch between `gp_restore` and the database server.

`-v | --verbose`
Specifies verbose mode.

`-a | --data-only`
Restore only the data, not the schema (data definitions).

`-c | --clean`
Clean (drop) database objects before recreating them.
-s | --schema-only
Restores only the schema (data definitions), no user data is restored.

-h hostname | --host=hostname
The host name of the Greenplum master host. If not provided, the value of PGHOST or the local host is used.

-p port | --port=port
The Greenplum master port. If not provided, the value of PGPORT or the port number provided at compile time is used.

-U username | --username=hostname
The database superuser account name, for example gpadmin. If not provided, the value of PGUSER or the current OS user name is used.

-W (force password prompt)
Forces a password prompt. This will happen automatically if the server requires password authentication.

--gp-c (use gunzip)
Use gunzip for inline decompression.

--gp-i (ignore errors)
Specifies that processing should ignore any errors that occur. Use this option to continue restore processing on errors.

--gp-d=directoryname
Specifies the relative or absolute path to backup files on the hosts. If this is a relative path, it is considered to be relative to the data directory. If not specified, defaults to the data directory of each instance being restored. Use this option if you created your backup files in an alternate location when running gp_dump.

--gp-r=reportfile
Specifies the full path name where the restore job report file will be placed on the master host. If not specified, defaults to the master data directory.

--gp-l=dbname [, ...] (restore certain segments)
Specifies whether to check for backup files on only the specified active segment instances (followed by a comma-separated list of the segments dbname). The default is to check for backup files on all active segments, restore the active segments, and then synchronize the mirrors.

-? | -h | --help (help)
Displays the online help.

--version (show utility version)
Displays the version of this utility.
Examples

Restore an Greenplum database using backup files created by gp_dump:

```
gp_restore --gp-k=2005103112453 -d gpdb
```

Restore a single segment instance only (by noting the dbid of the segment instance):

```
gp_restore --gp-k=2005103112453 -d gpdb --gp-s=5
```

See Also

gp_dump, pg_restore, gpdbrestore
gpaddmirrors

Adds mirror segments to a Greenplum Database system that was initially configured without mirroring.

Synopsis

gpaddmirrors -o output_sample_mirror_config [-m datadir_config_file]
gpaddmirrors -?
gpaddmirrors --version

Description

The gpaddmirrors utility configures mirror segment instances for an existing Greenplum Database system that was initially configured with primary segment instances only. The utility will create the mirror instances and begin the online replication process between the primary and mirror segment instances. Once all mirrors are synchronized with their primaries, your Greenplum Database system is fully data redundant.

By default, the utility will prompt you for the file system location(s) where it will create the mirror segment data directories. If you do not want to be prompted, you can pass in a file containing the file system locations using the -m option.

The mirror locations and ports must be different than your primary segment data locations and ports. If you have created additional filespaces, you will also be prompted for mirror locations for each of your filespaces.

The utility will create a unique data directory for each mirror segment instance in the specified location using the predefined naming convention. There must be the same number of file system locations declared for mirror segment instances as for primary segment instances. It is OK to specify the same directory name multiple times if you want your mirror data directories created in the same location, or you can enter a different data location for each mirror. Enter the absolute path. For example:

Enter mirror segment data directory location 1 of 2 > /gpdb/mirror
Enter mirror segment data directory location 2 of 2 > /gpdb/mirror

OR

Enter mirror segment data directory location 1 of 2 > /gpdb/m1
Enter mirror segment data directory location 2 of 2 > /gpdb/m2
Alternatively, you can run the `gpaddmirrors` utility and supply a detailed configuration file using the `-i` option. This is useful if you want your mirror segments on a completely different set of hosts than your primary segments. The format of the mirror configuration file is:

```plaintext
filespaceOrder=[filespace1_fsname[:filespace2_fsname:...]
mirror[content]=content:address:port:mir_replication_port:pri_replication_port:fselocation[:fselocation:...]
```

For example (if you do not have additional filespaces configured besides the default `pg_system` filesystem):

```plaintext
filespaceOrder=
mirror0=0:sdw1-1:60000:61000:62000:/gpdata/mir1/gp0
mirror1=1:sdw1-1:60001:61001:62001:/gpdata/mir2/gp1
```

The `gp_segment_configuration`, `pg_filespace`, and `pg_filespace_entry` system catalog tables can help you determine your current primary segment configuration so that you can plan your mirror segment configuration. For example, run the following query:

```sql
=# SELECT dbid, content, address as host_address, port, replication_port, fselocation as datadir
   FROM gp_segment_configuration, pg_filespace_entry
   WHERE dbid=fsedbid
   ORDER BY dbid;
```

If creating your mirrors on alternate mirror hosts, the new mirror segment hosts must be pre-installed with the Greenplum Database software and configured exactly the same as the existing primary segment hosts.

You must make sure that the user who runs `gpaddmirrors` (the `gpadmin` user) has permissions to write to the data directory locations specified. You may want to create these directories on the segment hosts and `chown` them to the appropriate user before running `gpaddmirrors`.

**Options**

- **-a** *(do not prompt)*
  Run in quiet mode - do not prompt for information. Must supply a configuration file with either `-m` or `-i` if this option is used.

- **-B parallel_processes**
  The number of mirror setup processes to start in parallel. If not specified, the utility will start up to 10 parallel processes depending on how many mirror segment instances it needs to set up.

- **-d master_data_directory**
  The master data directory. If not specified, the value set for `$MASTER_DATA_DIRECTORY` will be used.
-i mirror_config_file
A configuration file containing one line for each mirror segment you want to create. You must have one mirror segment listed for each primary segment in the system. The format of this file is as follows (as per attributes in the gp_segment_configuration, pg_filespace, and pg_filespace_entry catalog tables):

```
filespaceOrder=[filespace1_fsname[:filespace2_fsname:...]
mirror[content]=content:address:port:mir_replication_port:pri_replication_port:fselocation[:fselocation:...]
```

Note that you only need to specify a name for filespaceOrder if your system has multiple filesystems configured. If your system does not have additional filesystems configured besides the default pg_system filesystem, this file will only have one location (for the default data directory filesystem, pg_system). pg_system does not need to be listed in the filespaceOrder line. It will always be the first fselocation listed after replication_port.

-1 logfile_directory
The directory to write the log file. Defaults to ~/gpAdminLogs.

-m datadir_config_file
A configuration file containing a list of file system locations where the mirror data directories will be created. If not supplied, the utility will prompt you for locations. Each line in the file specifies a mirror data directory location. For example:

```
/gpdata/m1
/gpdata/m2
/gpdata/m3
/gpdata/m4
```

If your system has additional filesystems configured in addition to the default pg_system filesystem, you must also list file system locations for each filesystem as follows:

```
filespace filesystem
/gpfs1/m1
/gpfs1/m2
/gpfs1/m3
/gpfs1/m4
```

-o output_sample_mirror_config
If you are not sure how to lay out the mirror configuration file used by the -i option, you can run gpaddmirrors with this option to generate a sample mirror configuration file based on your primary segment configuration. The utility will prompt you for your mirror segment data directory locations (unless you provide these in a file using -m). You can then edit this file to change the host names to alternate mirror hosts if necessary.
-p port_offset
Optional. This number is used to calculate the database ports and replication ports used for mirror segments. The default offset is 100. Mirror port assignments are calculated as follows:

primary port + offset = mirror database port
primary port + (2 * offset) = mirror replication port
primary port + (3 * offset) = primary replication port

For example, if a primary segment has port 50001, then its mirror will use a database port of 51001, a mirror replication port of 52001, and a primary replication port of 53001 by default.

-s (spread mirrors)
Spreads the mirror segments across the available hosts. The default is to group a set of mirror segments together on an alternate host from their primary segment set. Mirror spreading will place each mirror on a different host within the Greenplum Database array. Spreading is only allowed if there is a sufficient number of hosts in the array (number of hosts is greater than or equal to the number of segment instances per host).

-v (verbose)
Sets logging output to verbose.

--version (show utility version)
Displays the version of this utility.

-? (help)
Displays the online help.

Examples

Add mirroring to an existing Greenplum Database system using the same set of hosts as your primary data. Calculate the mirror database and replication ports by adding 100 to the current primary segment port numbers:

$ gpaddmirrors -p 100

Add mirroring to an existing Greenplum Database system using a different set of hosts from your primary data:

$ gpaddmirrors -i mirror_config_file

Where the mirror_config_file looks something like this (if you do not have additional filespaces configured besides the default pg_system filespace):

```
filespaceOrder=
mirror0=0:sdw1-1:52001:53001:54001:/gpdata/mir1/gp0
mirror1=1:sdw1-2:52002:53002:54002:/gpdata/mir2/gp1
mirror2=2:sdw2-1:52001:53001:54001:/gpdata/mir1/gp2
```
Output a sample mirror configuration file to use with `gpaddmirrors -i`:

```
$ gpaddmirrors -o /home/gpadmin/sample_mirror_config
```

---

**See Also**

gpinitsystem, gpinitstandby, gactivatestandby
gpactivatestandby

Activates a standby master host and makes it the active master for the Greenplum Database system.

Synopsis

gpactivatestandby -d standby_master_datadir
[ -c new_standby_master ] [ -f ] [ -a ] [ -q ] [ -l logfile_directory ]

gpactivatestandby -? | -h | --help

gpactivatestandby -v

Description

The gpactivatestandby utility activates a backup master host and brings it into operation as the active master instance for a Greenplum Database system. The activated standby master effectively becomes the Greenplum Database master, accepting client connections on the master port (which must be set to the same port number on the master host and the backup master host).

You must run this utility from the master host you are activating, not the failed master host you are disabling. Running this utility assumes you have a backup master host configured for the system (see gpinitstandby).

The utility will perform the following steps:

- Stop the synchronization process (gpsyncagent) on the backup master
- Update the system catalog tables of the backup master using the logs
- Activate the backup master to be the new active master for the system
- (optional) Make the host specified with the -c option the new standby master host
- Restart the Greenplum Database system with the new master host

A backup Greenplum master host serves as a ‘warm standby’ in the event of the primary Greenplum master host becoming unoperational. The backup master is kept up to date by a transaction log replication process (gpsyncagent), which runs on the backup master host and keeps the data between the primary and backup master hosts synchronized.

If the primary master fails, the log replication process is shutdown, and the backup master can be activated in its place by using the gpactivatestandby utility. Upon activation of the backup master, the replicated logs are used to reconstruct the state of the Greenplum master host at the time of the last successfully committed transaction. To specify a new standby master host after making your current standby master active, use the -c option.

In order to use gpactivatestandby to activate a new primary master host, the master host that was previously serving as the primary master cannot be running. The utility checks for a postmaster.pid file in the data directory of the disabled master host, and if it finds it there, it will assume the old master host is still active. In some
cases, you may need to remove the postmaster.pid file from the disabled master host data directory before running gpactivatestandby (for example, if the disabled master host process was terminated unexpectedly).

After activating a standby master, run ANALYZE to update the database query statistics. For example:

```
psql dbname -c 'ANALYZE;'
```

### Options

-a (do not prompt)

Do not prompt the user for confirmation.

-c new_standby_master_hostname

Optional. After you activate your standby master you may want to specify another host to be the new standby, otherwise your Greenplum Database system will no longer have a standby master configured. Use this option to specify the hostname of the new standby master host. You can also use gpinitstandby at a later time to configure a new standby master host.

-d standby_master_datadir

Required. The absolute path of the data directory for the master host you are activating.

-f (force activation)

Use this option to force activation of the backup master host when the synchronization process (gpsyncagent) is not running. Only use this option if you are sure that the backup and primary master hosts are consistent, and you know the gpsyncagent process is not running on the backup master host. This option may be useful if you have just initialized a new backup master using gpinitstandby, and want to activate it immediately.

-l logfile_directory

The directory to write the log file. Defaults to ~/gpAdminLogs.

-q (no screen output)

Run in quiet mode. Command output is not displayed on the screen, but is still written to the log file.

-v (show utility version)

Displays the version, status, last updated date, and check sum of this utility.

-? | -h | --help (help)

Displays the online help.
**Examples**

Activate the backup master host and make it the active master instance for a Greenplum Database system (run from backup master host you are activating):

```
gpactivatestandby -d /gpdata
```

Activate the backup master host and at the same time configure another host to be your new standby master:

```
gpactivatestandby -d /gpdata -c new_standby_hostname
```

**See Also**

gpinitsystem, gpinitstandby
gpbitmapreindex

Rebuilds bitmap indexes after a 3.3.x to 4.0.x upgrade.

Synopsis

```
gpbitmapreindex -m { r | d | l [ -o output_sql_file] } 
[-h master_host] [-p master_port] [-n number_of_processes] [-v]
gpmigrator --version

gpmigrator --help | -?
```

Description

The on-disk format of bitmap indexes has changed from release 3.3.x to 4.0.x. Users who upgrade must rebuild all bitmap indexes after upgrading to 4.0. The `gpbitmapreindex` utility facilitates the upgrade of bitmap indexes by either running the `REINDEX` command to reindex them, or running the `DROP INDEX` command to simply remove them. If you decide to drop your bitmap indexes rather than reindex, run `gpbitmapreindex` in `list` -- `outfile` mode first to output a SQL file that you can use to recreate the indexes later. You must be the Greenplum Database superuser (`gpadmin`) in order to run `gpbitmapreindex`.

Options

- `-h host | --host host`
  Specifies the host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable `PGHOST` or defaults to `localhost`.

- `-m {r|d|l} | --mode {reindex|drop|list}`
  Required. The bitmap index upgrade mode: either `reindex`, `drop`, or `list` all bitmap indexes in the system.

- `-n number_of_processes | --parallel number_of_processes`
  The number of bitmap indexes to reindex or drop in parallel. Valid values are 1-16. The default is 1.

- `-o output_sql_file | --outfile output_sql_file`
  When used with `list` mode, outputs a SQL file that can be used to recreate the bitmap indexes.

- `-p port | --port port`
  Specifies the TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable `PGPORT` or defaults to 5432.
-v  | --verbose
   Show verbose output.

--version
   Displays the version of this utility.

-?  | --help
   Displays the online help.

Examples
Reindex all bitmap indexes:
   gpbitmapreindex -m r

Output a file of SQL commands that can be used to recreate all bitmap indexes:
   gpbitmapreindex -m list --outfile /home/gpadmin/bmp_ix.sql

Drop all bitmap indexes and run in verbose mode:
   gpbitmapreindex -m d -v

See Also
REINDEX, DROP INDEX, CREATE INDEX
gpcheck

Verifies and validates Greenplum Database platform settings.

Synopsis

gpcheck -f hostfile_gpcheck [-m master_host] [-s standy_master_host] [--stdout | --zipout] [--config config_file]
gpcheck --zipin gpcheck_zipfile

Description

The gpcheck utility determines the platform on which you are running Greenplum Database and validates various platform-specific configuration settings. gpcheck can use a host file or a file previously created with the --zipout option to validate platform settings. At the end of a successful validation process, GPCHECK_NORMAL message displays. If GPCHECK_ERROR displays, one or more validation checks failed. You can use also gpcheck to gather and view platform settings on hosts without running validation checks.

Greenplum recommends that you run gpcheck as root. If you do not run gpcheck as root, the utility displays a warning message and will not be able to validate all configuration settings; Only some of these settings will be validated.

Options

--config config_file

The name of a configuration file to use instead of the default file $GPHOME/etc/gpcheck.cnf (or ~/gpconfigs/gpcheck_dca_config on the EMC Greenplum Data Computing Appliance). This file specifies the OS-specific checks to run.

-f hostfile_gpcheck

The name of a file that contains a list of hosts that gpcheck uses to validate platform-specific settings. This file should contain a single host name for all hosts in your Greenplum Database system (master, standby master, and segments).

-m master_host

Perform special master host-specific validation tasks on this host.

-s standy_master_host

Perform special standby master host-specific validation tasks on this host.
--stdout
Display collected host information from gpcheck. No checks or validations are performed.

--zipout
Save all collected data to a .zip file in the current working directory. gpcheck automatically creates the .zip file and names it gpcheck_timestamp.tar.gz. No checks or validations are performed.

--zipin gpcheck_zipfile
Use this option to decompress and check a .zip file created with the --zipout option. gpcheck performs validation tasks against the file you specify in this option.

-? (help)
Displays the online help.

--version
Displays the version of this utility.

Examples
Verify and validate the Greenplum Database platform settings by entering a host file and specifying the master host and the standby master host:

# gpcheck -f hostfile_gpcheck -m mdw -s smdw

Save Greenplum Database platform settings to a zip file:

# gpcheck -f hostfile_gpcheck -m mdw -s smdw --zipout

Verify and validate the Greenplum Database platform settings using a zip file created with the --zipout option:

# gpcheck --zipin gpcheck_timestamp.tar.gz

View collected Greenplum Database platform settings:

# gpcheck -f hostfile_gpcheck -m mdw -s smdw --stdout

See Also
gpcheckperf
gpcheckperf

Verifies the baseline hardware performance of the specified hosts.

Synopsis

gpcheckperf -d test_directory [-d test_directory ...]
   {-f hostfile_gpcheckperf | - h hostname [-h hostname ...]}

gpcheckperf -d temp_directory
   {-f hostfile_gpchecknet | - h hostname [-h hostname ...]}

gpcheckperf -?

gpcheckperf --version

Description

The gpcheckperf utility starts a session on the specified hosts and runs the following performance tests:

- **Disk I/O Test (dd test)** — To test the sequential throughput performance of a logical disk or file system, the utility uses the dd command, which is a standard UNIX utility. It times how long it takes to write and read a large file to and from disk and calculates your disk I/O performance in megabytes (MB) per second. By default, the file size that is used for the test is calculated at two times the total random access memory (RAM) on the host. This ensures that the test is truly testing disk I/O and not using the memory cache.

- **Memory Bandwidth Test (stream)** — To test memory bandwidth, the utility uses the STREAM benchmark program to measure sustainable memory bandwidth (in MB/s). This tests that your system is not limited in performance by the memory bandwidth of the system in relation to the computational performance of the CPU. In applications where the data set is large (as in Greenplum Database), low memory bandwidth is a major performance issue. If memory bandwidth is significantly lower than the theoretical bandwidth of the CPU, then it can cause the CPU to spend significant amounts of time waiting for data to arrive from system memory.

- **Network Performance Test (gpnetbench*)** — To test network performance (and thereby the performance of the Greenplum Database interconnect), the utility runs a network benchmark program that transfers a 5 second stream of data from the current host to each remote host included in the test. The data is transferred in parallel to each remote host and the minimum, maximum, average and median network transfer rates are reported in megabytes (MB) per second. If the summary transfer rate is slower than expected (less than 100 MB/s), you can run the network test serially using the -r n option to obtain per-host results. To run a full-matrix bandwidth test, you can specify -r M which will cause every host to send and receive data from every other host specified. This test is best used to validate if the switch fabric can tolerate a full-matrix workload.
To specify the hosts to test, use the `-f` option to specify a file containing a list of host names, or use the `-h` option to name single host names on the command-line. If running the network performance test, all entries in the host file must be for network interfaces within the same subnet. If your segment hosts have multiple network interfaces configured on different subnets, run the network test once for each subnet.

You must also specify at least one test directory (with `-d`). The user who runs `gpcheckperf` must have write access to the specified test directories on all remote hosts. For the disk I/O test, the test directories should correspond to your segment data directories (primary and/or mirrors). For the memory bandwidth and network tests, a temporary directory is required for the test program files.

Before using `gpcheckperf`, you must have a trusted host setup between the hosts involved in the performance test. You can use the utility `gpssh-exkeys` to update the known host files and exchange public keys between hosts if you have not done so already. Note that `gpcheckperf` calls to `gpssh` and `gpscp`, so these Greenplum utilities must also be in your `$PATH`.

### Options

- **-B block_size**
  Specifies the block size to be used for disk I/O test. The default is 32KB, which is the same as the Greenplum Database page size. You can specify sizing in KB, MB, or GB.

- **-d test_directory**
  For the disk I/O test, specifies the file system directory locations to test. You must have write access to the test directory on all hosts involved in the performance test. You can use the `-d` option multiple times to specify multiple test directories (for example, to test disk I/O of your primary and mirror data directories).

- **-d temp_directory**
  For the network and stream tests, specifies a single directory where the test program files will be copied for the duration of the test. You must have write access to this directory on all hosts involved in the test.

- **-D (display per-host results)**
  Reports performance results for each host for the disk I/O tests. The default is to report results for just the hosts with the minimum and maximum performance, as well as the total and average performance of all hosts.

- **--duration time**
  Specifies the duration of the network test in seconds (the default), minutes (m), hours (h), or days (d). The default is 5 seconds.
-f hostfile_gpcheckperf
For the disk I/O and stream tests, specifies the name of a file that contains one host
name per host that will participate in the performance test. The host name is
required, and you can optionally specify an alternate user name and/or SSH port
number per host. The syntax of the host file is one host per line as follows:

[username@]hostname[:ssh_port]

-f hostfile_gpchecknet
For the network performance test, all entries in the host file must be for host
addresses within the same subnet. If your segment hosts have multiple network
interfaces configured on different subnets, run the network test once for each subnet.
For example (a host file containing segment host address names for interconnect
subnet 1):
sdw1-1
sdw2-1
sdw3-1

-h hostname
Specifies a single host name (or host address) that will participate in the
performance test. You can use the -h option multiple times to specify multiple host
names.

--netperf
Specifies that the netperf binary should be used to perform the network test
instead of the Greenplum network test. To use this option, you must download
netperf from www.netperf.org and install it into $GPHOME/bin/lib on all
Greenplum hosts (master and segments).

-r ds{n|N|M}
Specifies which performance tests to run. The default is dsn:

- Disk I/O test (d)
- Stream test (s)
- Network performance test in sequential (n), parallel (N), or full-matrix (M)
  mode. The optional --duration option specifies how long (in seconds) to
  run the network test. To use the parallel (N) mode, you must run the test on an
even number of hosts.

If you would rather use netperf (www.netperf.org) instead of the Greenplum
network test, you can download it and install it into $GPHOME/bin/lib on all
Greenplum hosts (master and segments). You would then specify the optional
--netperf option to use the netperf binary instead of the default
gpnetbench* utilities.
-S file_size
Specifies the total file size to be used for the disk I/O test for all directories specified with -d. file_size should equal two times total RAM on the host. If not specified, the default is calculated at two times the total RAM on the host where gpcheckperf is executed. This ensures that the test is truly testing disk I/O and not using the memory cache. You can specify sizing in KB, MB, or GB.

-v (verbose) | -V (very verbose)
Verbose mode shows progress and status messages of the performance tests as they are run. Very verbose mode shows all output messages generated by this utility.

--version
Displays the version of this utility.

-? (help)
Displays the online help.

Examples
Run the disk I/O and memory bandwidth tests on all the hosts in the file host_file using the test directory of /data1 and /data2:

$ gpcheckperf -f hostfile_gpcheckperf -d /data1 -d /data2 -r ds

Run only the disk I/O test on the hosts named sdw1 and sdw2 using the test directory of /data1. Show individual host results and run in verbose mode:

$ gpcheckperf -h sdw1 -h sdw2 -d /data1 -r d -D -v

Run the parallel network test using the test directory of /tmp, where hostfile_gpcheck_ic* specifies all network interface host address names within the same interconnect subnet:

$ gpcheckperf -f hostfile_gpchecknet_ic1 -r N -d /tmp
$ gpcheckperf -f hostfile_gpchecknet_ic2 -r N -d /tmp

Run the same test as above, but use netperf instead of the Greenplum network test (note that netperf must be installed in $GPHOME/bin/lib on all Greenplum hosts):

$ gpcheckperf -f hostfile_gpchecknet_ic1 -r N --netperf -d /tmp
$ gpcheckperf -f hostfile_gpchecknet_ic2 -r N --netperf -d /tmp

See Also
gpssh, gpscp, gpcheck
gpconfig

Sets server configuration parameters on all segments within a Greenplum Database system.

Synopsis

\texttt{gpconfig -c \ param\_name \ -v \ value \ [-m \ master\_value \ | \ --masteronly]}
\texttt{| \ -r \ param\_name \ [--masteronly]}
\texttt{| \ -l}
\texttt{[--skipvalidation] \ [--verbose] \ [--debug]}

\texttt{gpconfig -s \ param\_name \ [--masteronly]}

\texttt{gpconfig --help}

Description

The \texttt{gpconfig} utility allows you to set, unset, or view configuration parameters from the \texttt{postgresql.conf} files of all instances (master, segments, and mirrors) in your Greenplum Database system. When setting a parameter, you can also specify a different value for the master if necessary. For example, parameters such as \texttt{max\_connections} require a different setting on the master than what is used for the segments. If you want to set or unset a global or master only parameter, use the \texttt{--masteronly} option.

\texttt{gpconfig} can only be used to manage certain parameters. For example, you cannot use it to set parameters such as \texttt{port}, which is required to be distinct for every segment instance. Use the \texttt{-l (list)} option to see a complete list of configuration parameters supported by \texttt{gpconfig}.

When \texttt{gpconfig} sets a configuration parameter in a segment \texttt{postgresql.conf} file, the new parameter setting always displays at the bottom of the file. When you use \texttt{gpconfig} to remove a configuration parameter setting, \texttt{gpconfig} comments out the parameter in all segment \texttt{postgresql.conf} files, thereby restoring the system default setting. For example, if you use \texttt{gpconfig} to remove (comment out) a parameter and later add it back (set a new value), there will be two instances of the parameter; one that is commented out, and one that is enabled and inserted at the bottom of the \texttt{postgresql.conf} file.

After setting a parameter, you must restart your Greenplum Database system or reload the \texttt{postgresql.conf} files in order for the change to take effect. Whether you require a restart or a reload depends on the parameter. See the Server Configuration Parameters reference for more information about the server configuration parameters.

To show the currently set values for a parameter across the system, use the \texttt{-s} option.

\texttt{gpconfig} uses the following environment variables to connect to the Greenplum Database master instance and obtain system configuration information:

- \texttt{PGHOST}
- \texttt{PGPORT}
- \texttt{PGUSER}
Options

-c | --change param_name
Changes a configuration parameter setting by adding the new setting to the bottom of the postgresql.conf files.

-v | --value value
The value to use for the configuration parameter you specified with the -c option. By default, this value is applied to all segments, their mirrors, the master, and the standby master.

-m | --mastervalue master_value
The master value to use for the configuration parameter you specified with the -c option. If specified, this value only applies to the master and standby master. This option can only be used with -v.

--masteronly
When specified, gpconfig will only edit the master postgresql.conf file.

-r | --remove param_name
Removes a configuration parameter setting by commenting out the entry in the postgresql.conf files.

-l | --list
Lists all configuration parameters supported by the gpconfig utility.

-s | --show param_name
Shows the value for a configuration parameter used on all instances (master and segments) in the Greenplum Database system. If there is a discrepancy in a parameter value between segment instances, the gpconfig utility displays an error message. Note that the gpconfig utility reads parameter values directly from the database, and not the postgresql.conf file. If you are using gpconfig to set configuration parameters across all segments, then running gpconfig -s to verify the changes, you might still see the previous (old) values. You must reload the configuration files (gpstop -u) or restart the system (gpstop -r) for changes to take effect.

--skipvalidation
Overrides the system validation checks of gpconfig and allows you to operate on any server configuration parameter, including hidden parameters and restricted parameters that cannot be changed by gpconfig. When used with the -l option (list), it shows the list of restricted parameters. This option should only be used to set parameters when directed by Greenplum Customer Support.
--verbose
Displays additional log information during gpconfig command execution.

--debug
Sets logging output to debug level.

-? | -h | --help
Displays the online help.

**Examples**
Set the work_mem parameter to 120MB in the master host file only:

```
gpconfig -c work_mem -v 120MB --masteronly
```

Set the max_connections setting to 100 on all segments and 10 on the master:

```
gpconfig -c max_connections -v 100 -m 10
```

Comment out all instances of the default_statistics_target configuration parameter, and restore the system default:

```
gpconfig -r default_statistics_target
```

List all configuration parameters supported by gpconfig:

```
gpconfig -l
```

Show the values of a particular configuration parameter across the system:

```
gpconfig -s max_connections
```

**See Also**
gpstop
**gpcrondump**

A wrapper utility for *gp_dump*, which can be called directly or from a *crontab* entry.

### Synopsis

```
gpcrondump -x database_name
[-s schema | -t schema.table | -T schema.table]
[-w dbid[,..]] [-u backup_directory]
[-R post_dump_script] [-c | -o] [-z] [-r] [-f free_space_percent]
[-B parallel_processes] [-a] [-q]
[-y reportfile] [-l logfile_directory] [-D]
{ [-E encoding] [--inserts | --column-inserts] [--oids]
  [--no-owner | --use-set-session-authorization]
  [--no-privileges] }
```

### Description

The *gpcrondump* utility is a wrapper utility for *gp_dump*. By default, dump files are created in their respective master and segment data directories in a directory named `db_dumps/YYYYMMDD`. The data dump files are compressed by default using gzip.

The *gpcrondump* utility allows you to schedule routine backups of a Greenplum database using cron (a scheduling utility for UNIX operating systems). Cron jobs that call *gpcrondump* should be scheduled on the master host.

### Email Notifications

To have *gpcrondump* send out status email notifications, you must place a file named `mail_contacts` in the home directory of the Greenplum superuser (`gpadmin`) or in the same directory as the *gpcrondump* utility (`$GPHOME/bin`). This file should contain one email address per line. *gpcrondump* will issue a warning if it cannot locate a `mail_contacts` file in either location. If both locations have a `mail_contacts` file, then the one in `$HOME` takes precedence.

### Options

- **-a (do not prompt)**
  Do not prompt the user for confirmation.

- **-b (bypass disk space check)**
  Bypass disk space check. The default is to check for available disk space.
-B parallel_processes
  The number of segments to check in parallel for pre/post-dump validation. If not specified, the utility will start up to 60 parallel processes depending on how many segment instances it needs to dump.

-c (clear old dump files first)
  Clear out old dump files before doing the dump. The default is not to clear out old dump files. This will remove all old dump directories in the db_dumps directory, except for the dump directory of the current date.

-C (clean old catalog dumps)
  Clean out old catalog schema dump files prior to create.

--column-inserts
  Dump data as INSERT commands with column names.

-d master_data_directory
  The master host data directory. If not specified, the value set for $MASTER_DATA_DIRECTORY will be used.

-D (debug)
  Sets logging level to debug.

-E encoding
  Character set encoding of dumped data. Defaults to the encoding of the database being dumped. See “Character Set Support” on page 939 for the list of supported character sets.

-f free_space_percent
  Overestimate the amount of free disk space by this percentage when performing the pre-dump disk space check. This overestimation is how the pre-dump check accounts for the disk space saved when compression is used. The default is 10.

-g (copy config files)
  Secure a copy of the master and segment configuration files postgresql.conf, pg_ident.conf, and pg_hba.conf. These configuration files are dumped in the master or segment data directory to db_dumps/YYYYMMDD/config_files_<timestamp>.tar

-G (dump global objects)
  Use pg_dumpall to dump global objects such as roles and tablespaces. Global objects are dumped in the master data directory to db_dumps/YYYYMMDD/gp_global_1_1_<timestamp>.

-i (ignore parameter check)
  Ignore the initial parameter check phase.
--inserts
Dump data as INSERT, rather than COPY commands.

-j (vacuum before dump)
Run VACUUM before the dump starts.

-k (vacuum after dump)
Run VACUUM after the dump has completed successfully.

-l logfile_directory
The directory to write the log file. Defaults to ~/gpAdminLogs.

--no-owner
Do not output commands to set object ownership.

--no-privileges
Do not output commands to set object privileges (GRANT/REVOKE commands).

-o (clear old dump files only)
Clear out old dump files only, but do not run a dump. This will remove the oldest dump directory except the current date’s dump directory. All dump sets within that directory will be removed.

--oids
Include object identifiers (oid) in dump data.

-p (primary segments only)
Dump all primary segments, which is the default behavior. Note: this option is deprecated.

-q (no screen output)
Run in quiet mode. Command output is not displayed on the screen, but is still written to the log file.

-r (no rollback on failure)
Do not rollback the dump files (delete a partial dump) if a failure is detected. The default is to rollback (delete partial dump files).

-R post_dump_script
The absolute path of a script to run after a successful dump operation. For example, you might want a script that moves completed dump files to a backup host. This script must reside in the same location on the master and all segment hosts.

-s schema_name
Dump only the named schema in the named database.
-t schema.table_name
  Dump only the named table in this database. The -t option can be specified multiple times.

-T schema.table_name
  A table name to exclude from the database dump. The -T option can be specified multiple times.

-u backup_directory
  Specifies the absolute path where the backup files will be placed on each host. If the path does not exist, it will be created, if possible. If not specified, defaults to the data directory of each instance to be backed up. Using this option may be desirable if each segment host has multiple segment instances as it will create the dump files in a centralized location rather than the segment data directories.

--use-set-session-authorization
  Use SET SESSION AUTHORIZATION commands instead of ALTER OWNER commands to set object ownership.

-v (show utility version)
  Displays the version, status, last updated date, and check sum of this utility.

-w dbid[,....] (backup certain segments only)
  Specifies a set of active segment instances to back up as a comma-separated list of the segments’ dbid. The master is added to the list automatically. The default is to backup all active segment instances.

-x database_name
  Required. The name of the Greenplum database to dump.

-y reportfile
  Specifies the full path name where the backup job log file will be placed on the master host. If not specified, defaults to the master data directory or if running remotely, the current working directory.

-z (no compression)
  Do not use compression. Default is to compress the dump files using gzip.

-? (help)
  Displays the online help.

Examples
Call gpcrondump directly and dump mydatabase (and global objects):

    gpcrondump -x mydatabase -c -g -G

A Linux crontab entry that runs a backup of the sales database (and global objects) nightly at one past midnight:
SHELL=/bin/bash
GPHOME=/usr/local/greenplum-db-4.0.x.x
MASTER_DATA_DIRECTORY=/data/gpdb_p1/gp-1
PATH=$PATH:$GPHOME/bin
01 0 * * * gpadmin gpcrondump -x sales -c -g -G -a -q >>
gp_salesdump.log

A Solaris crontab entry that runs a backup of the sales database (and global objects) nightly at one past midnight (no line breaks):

01 0 * * * SHELL=/bin/bash ;
GPHOME=/usr/local/greenplum-db-4.0.x.x ;
PATH=$PATH:$GPHOME/bin ; HOME=/export/home/gpadmin ;
MASTER_DATA_DIRECTORY=/data/gpdb_p1/gp-1 ;
/usr/local/greenplum-db/bin/gpcrondump -x sales -c -g -G -a
-q >> gp_salesdump.log

See Also

gp_dump, gpdbrestore
gpdbrestore

A wrapper utility around `gp_restore`. Restores a database from a set of dump files generated by `gpcrondump`.

**Synopsis**

```
gpdbrestore { -t timestamp_key [-L] | -b YYYYMMDD | 
-R hostname:path_to_dumpset | -s database_name } 
[-T schema.table [, ...]] [-e] [-G] [-B parallel_processes] 
gpdbrestore -?
gpdbrestore -v
```

**Description**

`gpdbrestore` is a wrapper around `gp_restore`, which provides some convenience and flexibility in restoring from a set of backup files created by `gpcrondump`. This utility provides the following additional functionality on top of `gp_restore`:

- Automatically reconfigures for compression.
- Validates the number of dump files are correct (For primary only, mirror only, primary and mirror, or a subset consisting of some mirror and primary segment dump files).
- If a failed segment is detected, restores to active segment instances.
- Do not need to know the complete timestamp key (`-t`) of the backup set to restore. Additional options are provided to instead give just a date (`-b`), backup set directory location (`-R`), or database name (`-s`) to restore.
- The `-R` option allows the ability to restore from a backup set located on a host outside of the Greenplum Database array (archive host). Ensures that the correct dump file goes to the correct segment instance.
- Identifies the database name automatically from the backup set.
- Allows you to restore particular tables only (`-T` option) instead of the entire database. Note that single tables are not automatically dropped or truncated prior to restore.
- Can restore global objects such as roles and tablespaces (`-G` option).
- Detects if the backup set is primary segments only or primary and mirror segments and passes the appropriate options to `gp_restore`.
- Allows you to drop the target database before a restore in a single operation.

**Options**

- `-a (do not prompt)`
  Do not prompt the user for confirmation.
-b YYYMMDD
Looks for dump files in the segment data directories on the Greenplum Database array of hosts in db_dumps/YYYYMMDD

-B parallel_processes
The number of segments to check in parallel for pre/post-restore validation. If not specified, the utility will start up to 60 parallel processes depending on how many segment instances it needs to restore.

-d master_data_directory
Optional. The master host data directory. If not specified, the value set for $MASTER_DATA_DIRECTORY will be used.

-D (debug)
Sets logging level to debug.

-e (drop target database before restore)
Drops the target database before doing the restore and then recreates it.

-G (restore global objects)
Restores global objects such as roles and tablespaces if the global object dump file db_dumps/<date>/gp_global_1_1_<timestamp> is found in the master data directory.

-l logfile_directory
The directory to write the log file. Defaults to ~/gpAdminLogs.

-L (list tablenames in backup set)
When used with the -t option, lists the table names that exist in the named backup set and exits. Does not do a restore.

-q (no screen output)
Run in quiet mode. Command output is not displayed on the screen, but is still written to the log file.

-R hostname:path_to_dumpset
Allows you to provide a hostname and full path to a set of dump files. The host does not have to be in the Greenplum Database array of hosts, but must be accessible from the Greenplum master.

-s database_name
Looks for latest set of dump files for the given database name in the segment data directories db_dumps directory on the Greenplum Database array of hosts.
-t timestamp_key
The 14 digit timestamp key that uniquely identifies a backup set of data to restore. It is of the form YYYYMMDDHHMMSS. Looks for dump files matching this timestamp key in the segment data directories db_dumps directory on the Greenplum Database array of hosts.

-T schema.table_name
A comma-separated list of specific table names to restore. The named table(s) must exist in the backup set of the database being restored. Existing tables are not automatically truncated before data is restored from backup. If your intention is to replace existing data in the table from backup, truncate the table prior to running gpdbrestore -T.

-v (show utility version)
Displays the version, status, last updated date, and check sum of this utility.

-? (help)
Displays the online help.

Examples
Restore the sales database from the latest backup files generated by gpcrondump (assumes backup files are in the segment data directories in db_dumps):

    gpdbrestore -s sales

Restore a database from backup files that reside on an archive host outside the Greenplum Database array (command issued on the Greenplum master host):

    gpdbrestore -R archivehostname:/data_p1/db_dumps/20080214

Restore global objects only (roles and tablespaces):

    gpdbrestore -G

See Also
gpcrondump, gp_restore
gpdeletesystem

Deletes a Greenplum Database system that was initialized using gpinitdb.

Synopsis

gpdeletesystem -?
gpdeletesystem -v

Description

The gpdeletesystem utility will perform the following actions:

- Stop all postgres processes (the segment instances and master instance).
- Deletes all data directories.

Before running gpdeletesystem:

- Move any backup files out of the master and segment data directories.
- Make sure that Greenplum Database is running.
- If you are currently in a segment data directory, change directory to another location. The utility fails with an error when run from within a segment data directory.

This utility will not uninstall the Greenplum Database software.

Options

- **-d data_directory**
  Required. The master host data directory.

- **-B parallel_processes**
  The number of segments to delete in parallel. If not specified, the utility will start up to 60 parallel processes depending on how many segment instances it needs to delete.

- **-f (force)**
  Force a delete even if backup files are found in the data directories. The default is to not delete Greenplum Database instances if backup files are present.

- **-l logfile_directory**
  The directory to write the log file. Defaults to ~/gpAdminLogs.

- **-D (debug)**
  Sets logging level to debug.
-? (help)
  Displays the online help.

-v (show utility version)
  Displays the version, status, last updated date, and check sum of this utility.

Examples
Delete a Greenplum Database system:
  `gpdeletesystem -d /gpdata/gp-1`
Delete a Greenplum Database system even if backup files are present:
  `gpdeletesystem -d /gpdata/gp-1 -f`
gpdetective

Collects diagnostic information from a running Greenplum Database system.

Synopsis

gpdetective [-h hostname] [-p port] [-U username] [-P password]
  [--start_date number_of_days | YYYY-MM-DD]
  [--end_date YYYY-MM-DD]
  [--diagnostics a|n|s|o|c]
  [--logs a|n|dbid[,dbid,... | -dbid]]
  [--cores t|f]
  [--pg_dumpall t|f] [--pg_dump_options option[, ...]]
  [--tempdir temp_dir]
  [--connect t|f]

gpdetective -?
gpdetective -v

Description

The gpdetective utility collects information from a running Greenplum Database
system and creates a bzip2-compressed tar output file. This output file can then be sent
to Greenplum Customer Support to help with the diagnosis of Greenplum Database
errors or system failures. The gpdetective utility runs the following diagnostic tests:

- \texttt{gpstate} to check the system status
- \texttt{gpcheck} to make sure the recommended OS settings are set on all hosts
- \texttt{gpcheckcat} and \texttt{gpcheckdb} to check the system catalog tables for
  inconsistencies

\texttt{gpdetective} captures the following files and Greenplum system information:

- \texttt{postgresql.conf} configuration files (master and segments)
- log files (master and segments)
- Greenplum Database system configuration information
- (optional) Core files
- (optional) Schema DDL dumps for all databases and global objects (using
  \texttt{pg_dumpall})

A bzip2-compressed tar output file containing this information is created in the current
directory with a file name of \texttt{gpdetective<timestamp>.tar.bz2}. 
Options

--connect t|f
Specifies if gpdetective should connect to the database to obtain system information. The default is true (t). If false (f), gpdetective only gathers information it can obtain without making a connection to the database. This information includes (from the master host):

- Log files
- The master_data_directory/postgresql.conf file
- The ~/.gpAdminLogs directory
- gpcheck output
- Core files

--cores t|f
Determines whether or not the utility retrieves core files. The default is true (t).

--diagnostics a|n|s|o|c
Specifies the diagnostic tests to run: all (a), none (n), operating system (o) diagnostics, or catalog (c) diagnostics. The default is all (a).

--end_date YYYY-MM-DD
Sets the end date for the diagnostic information collected. The collected information ends at 00:00:00 of the specified date.

-h hostname
The host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable PGHOST or defaults to localhost.

--logs a|n|dbid_list
Specifies which log file(s) to retrieve: all (a), none (n), a comma separated list of segment dbid numbers, or a range of dbid numbers divided by a dash (-) (for example, 3-6 retrieves logs from segments 3, 4, 5, and 6). The default is all (a).

-P password
If Greenplum Database is configured to use password authentication, you must also supply the database superuser password. If not specified, reads from ~/.pgpass if it exists.

--pg_dumpall t|f
Determines whether or not the utility runs pg_dumpall to collect schema DDL for all databases and global objects. The default is true (t).
**--pg_dump_options** option[,...]
If --pg_dumpall is true, specifies a comma separated list of dump options to use when the pg_dumpall utility is called. See pg_dumpall for a valid list of dump options.

**-p port**
The TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable PGPORT or defaults to 5432.

**--start_date** number_of_days | YYYY-MM-DD
Sets the start date for the diagnostic information collected. Specify either the number of days prior, or an explicit past date.

**--tempdir** temp_dir
Specifies the temporary directory used by gpdetective. The default value is determined by the $TEMP, $TMP and $TMPDIR environment variables.

**-U gp_superuser**
The Greenplum database superuser role name to connect as (typically gpadmin). If not specified, reads from the environment variable PGUSER or defaults to the current system user name.

**-v (show utility version)**
Displays the version of this utility.

**-? (help)**
Displays the utility usage and syntax.

**Examples**
Collect all diagnostic information for a Greenplum Database system and supply the required connection information for the master host:

```bash
gpdetective -h mdw -p 54320 -U gpadmin -P mypassword
```

Run diagnostics and collect all logs and system information for the past two days:

```bash
gpdetective --start_date 2
```

Do not run diagnostic tests or schema dumps, just collect the log files of the master and segment 3:

```bash
gpdetective --diagnostics n --logs -1,3 --pg_dumpall f
```

**See Also**
gpstate, gpcheck, pg_dumpall
gpexpand

Expands an existing Greenplum Database across new hosts in the array.

**Synopsis**

```
gpexpand
   [-f hosts_file]
   | -i input_file [-B batch_size] [-V]
   | { -d hh:mm:ss | -e 'YYYY-MM-DD hh:mm:ss' } [-analyze]
   | -n parallel_processes
   | --rollback
   | --clean
   [-D database_name] [--verbose] [--silent]
```  
gpexpand -? | -h | --help

**Prerequisites**

- You are logged in as the Greenplum Database superuser (gpadmin).
- The new segment hosts have been installed and configured as per the existing segment hosts. This involves:
  - Configuring the hardware and OS
  - Installing the Greenplum software
  - Creating the gpadmin user account
  - Exchanging SSH keys.
- Enough disk space on your segment hosts to temporarily hold a copy of your largest table.

**Description**

The gpexpand utility performs system expansion in two phases: segment initialization and then table redistribution.

In the initialization phase, gpexpand runs with an input file that specifies data directories, dbid values, and other characteristics of the new segments. You can create the input file manually, or by following the prompts in an interactive interview.

If you choose to create the input file using the interactive interview, you can optionally specify a file containing a list of expansion hosts. If your platform or command shell limits the length of the list of hostnames that you can type when prompted in the interview, specifying the hosts with `-f` may be mandatory.

In addition to initializing the segments, the initialization phase performs these actions:

- Creates an expansion schema to store the status of the expansion operation, including detailed status for tables.
- Changes the distribution policy for all tables to DISTRIBUTED RANDOMLY. The original distribution policies are later restored in the redistribution phase.

To begin the redistribution phase, you must run gpexpand with either the \-d (duration) or \-e (end time) options. Until the specified end time or duration is reached, the utility will redistribute tables in the expansion schema. Each table is reorganized using ALTER TABLE commands to rebalance the tables across new segments, and to set tables to their original distribution policy. If gpexpand completes the reorganization of all tables before the specified duration, it displays a success message and ends.

## Options

- **-a | --analyze**
  
  Run ANALYZE to update the table statistics after expansion. The default is to not run ANALYZE.

- **-B batch_size**
  
  Batch size of remote commands to send to a given host before making a one-second pause. Default is 8. Valid values are 1-128.

  The gpexpand utility issues a number of setup commands that may exceed the host’s maximum threshold for authenticated connections as defined by MaxStartups in the SSH daemon configuration. The one-second pause allows authentications to be completed before gpexpand issues any more commands.

  The default value does not normally need to be changed. However, it may be necessary to reduce the maximum number of commands if gpexpand fails with connection errors such as 'ssh_exchange_identification: Connection closed by remote host.'

- **-c | --clean**
  
  Remove the expansion schema.

- **-d | --duration \texttt{hh:mm:ss}**
  
  Duration of the expansion session from beginning to end.

- **-D database_name**
  
  Specifies the database in which to create the expansion schema and tables. If this option is not given, the setting for the environment variable PGDATABASE is used. The database templates template1 and template0 cannot be used.

- **-e | --end \texttt{YYYY-MM-DD \texttt{hh:mm:ss}}**
  
  Ending date and time for the expansion session.

- **-f | --hosts-file filename**
  
  Specifies the name of a file that contains a list of new hosts for system expansion. Each line of the file must contain a single host name.
This file can contain hostnames with or without network interfaces specified. The `gpexpand` utility handles either case, adding interface numbers to end of the hostname if the original nodes are configured with multiple network interfaces.

`-i | --input input_file`
Specifies the name of the expansion configuration file, which contains one line for each segment to be added in the format of:

```
hostname:address:port:fselocation:dbid:content:preferred_role:replication_port
```

If your system has filespaces, `gpexpand` will expect a filesystem configuration file (`input_file_name.fs`) to exist in the same directory as your expansion configuration file. The filesystem configuration file is in the format of:

```
filespaceOrder=filespace1_name:filespace2_name: ... 
dbid:/path/for/filespace1:/path/for/filespace2: ... 
```

`-n parallel_processes`
The number of tables to redistribute simultaneously. Valid values are 1 - 16.

Each table redistribution process requires two database connections: one to alter the table, and another to update the table’s status in the expansion schema. Before increasing `-n`, check the current value of the server configuration parameter `max_connections` and make sure the maximum connection limit is not exceeded.

`-r | --rollback`
Roll back a failed expansion setup operation. If the rollback command fails, attempt again using the `-D` option to specify the database that contains the expansion schema for the operation that you want to roll back.

`-s | --silent`
Runs in silent mode. Does not prompt for confirmation to proceed on warnings.

`-v | --verbose`
Verbose debugging output. With this option, the utility will output all DDL and DML used to expand the database.

`--version`
Display the utility’s version number and exit.

`-V | --novacuum`
Do not vacuum catalog tables before creating schema copy.

`-? | -h | --help`
Displays the online help.
Examples

Run `gpexpand` with an input file to initialize new segments and create the expansion schema in the default database:

```
$ gpexpand -i input_file
```

Run `gpexpand` for sixty hours maximum duration to redistribute tables to new segments:

```
$ gpexpand -d 60:00:00
```

See Also

`gpssh-exkeys`
gpfdist

Serves data files to or writes data files out from Greenplum Database segments.

Synopsis

gpfdist -?
gpfdist --version

Description

gpfdist is Greenplum’s parallel file distribution program. It is used by readable external tables and gpload to serve external table files to all Greenplum Database segments in parallel. It is used by writable external tables to accept output streams from Greenplum Database segments in parallel and write them out to a file.

In order for gpfdist to be used by an external table, the LOCATION clause of the external table definition must specify the correct file location using the gpfdist:// protocol (see CREATE EXTERNAL TABLE).

The benefit of using gpfdist is that you are guaranteed maximum parallelism while reading from or writing to external tables, thereby offering the best performance as well as easier administration of external tables.

For readable external tables, gpfdist parses and serves data files evenly to all the segment instances in the Greenplum Database system when users SELECT from the external table. For writable external tables, gpfdist accepts parallel output streams from the segments when users INSERT into the external table, and writes to an output file.

For readable external tables, if load files are compressed using gzip or bzip2 (have a .gz or .bz2 file extension), gpfdist will uncompress the files automatically before loading provided that gunzip or bunzip2 is in your path. Writable external tables do not currently support compression.

Most likely, you will want to run gpfdist on your ETL machines rather than the hosts where Greenplum Database is installed. To install gpfdist on another host, simply copy the utility over to that host and add gpfdist to your $PATH.

Options

- **d directory**
  The directory from which gpfdist will serve files for readable external tables or create output files for writable external tables. If not specified, defaults to the current directory.
-l **log_file**
   The fully qualified path and log file name where standard output messages are to be logged.

-**p** **http_port**
   The HTTP port on which gpfdist will serve files. Defaults to 8080.

-**t** **timeout**
   Sets the time allowed for Greenplum Database to establish a connection to a gpfdist process. Default is 5 seconds. Allowed values are 2 to 30 seconds. May need to be increased on systems with a lot of network traffic.

-**m** **max_length**
   Sets the maximum allowed data row length in bytes. Default is 32768. Should be used when user data includes very wide rows (or when line too long error message occurs). Should not be used otherwise as it increases resource allocation. Valid range is 32K to 1MB.

-**?** (**help**)  
   Displays the online help.

-**v** (**verbose**)  
   Verbose mode shows progress and status messages.

-**V** (**very verbose**)  
   Verbose mode shows all output messages generated by this utility.

--**version**
   Displays the version of this utility.

**Examples**

Serve files from a specified directory using port 8081 (and start gpfdist in the background):

```
gpfdist -d /var/load_files -p 8081 &
```

Start gpfdist in the background and redirect output and errors to a log file:

```
gpfdist -d /var/load_files -p 8081 -l /home/gpadmin/log &
```

To stop gpfdist when it is running in the background:

```
--First find its process id:
ps ax | grep gpfdist
```

OR on Solaris:

```
ps -ef | grep gpfdist
```

```
--Then kill the process, for example:
k 3456
```
See Also

CREATE EXTERNAL TABLE, gpload
gpfilespace

Creates a filespace using a configuration file that defines per-segment file system locations. Filespaces describe the physical file system resources to be used by a tablespace.

Synopsis

```
gpfilespace [connection_option ...] [-l logfile_directory] [-o [output_file_name]]
gpfilespace [connection_option ...] [-l logfile_directory] -c fs_config_file
gpfilespace -v | -?
```

Description

A tablespace requires a file system location to store its database files. In Greenplum Database, the master and each segment (primary and mirror) needs its own distinct storage location. This collection of file system locations for all components in a Greenplum system is referred to as a filespace. Once a filespace is defined, it can be used by one or more tablespaces.

When used with the -o option, the gpfilespace utility looks up your system configuration information in the Greenplum Database catalog tables and prompts you for the appropriate file system locations needed to create the filespace. It then outputs a configuration file that can be used to create a filespace. If a file name is not specified, a `gpfilespace_config_#` file will be created in the current directory by default.

Once you have a configuration file, you can run gpfilespace with the -c option to create the filespace in Greenplum Database.

Options

-c | --config fs_config_file

A configuration file containing:

- An initial line denoting the new filespace name. For example:
  ```
  filespace:myfs
  ```
- One line each for the master, the primary segments, and the mirror segments. A line describes a file system location that a particular segment database instance should use as its data directory location to store database files associated with a tablespace. Each line is in the format of:
  ```
  hostname:dbid:/filesystem_dir/seg_datadir_name
  ```

-1 | --logdir logfile_directory

The directory to write the log file. Defaults to `~/gpAdminLogs`.
-o | --output output_file_name
The directory location and file name to output the generated filespace configuration file. You will be prompted to enter a name for the filespace, a master file system location, the primary segment file system locations, and the mirror segment file system locations. For example, if your configuration has 2 primary and 2 mirror segments per host, you will be prompted for a total of 5 locations (including the master). The file system locations must exist on all hosts in your system prior to running the gpfilespace utility. The utility will designate segment-specific data directories within the location(s) you specify, so it is possible to use the same location for multiple segments. However, primaries and mirrors cannot use the same location. After the utility creates the configuration file, you can manually edit the file to make any required changes to the filespace layout before creating the filespace in Greenplum Database.

-v | --version (show utility version)
Displays the version of this utility.

-? | --help (help)
Displays the utility usage and syntax.

Connection Options

-h host | --host host
The host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable PGHOST or defaults to localhost.

-p port | --port port
The TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable PGPORT or defaults to 5432.

-U username | --username superuser_name
The database superuser role name to connect as. If not specified, reads from the environment variable PGUSER or defaults to the current system user name. Only database superusers are allowed to create filespaces.

-W | --password
Force a password prompt.

Examples
Create a filespace configuration file. You will be prompted to enter a name for the filespace, a master file system location, the primary segment file system locations, and the mirror segment file system locations. For example, if your configuration has 2 primary and 2 mirror segments per host, you will be prompted for a total of 5 locations (including the master). The file system locations must exist on all hosts in your system prior to running the gpfilespace utility:
$ gpfilespace -o .
Enter a name for this filespace
> fastdisk

Checking your configuration:
Your system has 2 hosts with 2 primary and 2 mirror segments per host.

Configuring hosts: [sdw1, sdw2]

Please specify 2 locations for the primary segments, one per line:
primary location 1> /gp_pri_filespc
primary location 2> /gp_pri_filespc

Please specify 2 locations for the mirror segments, one per line:
mirror location 1> /gp_mir_filespc
mirror location 2> /gp_mir_filespc

Enter a file system location for the master:
master location> /gp_master_filespc

Example filespace configuration file:
    filespace:fastdisk
    mdw:1:/gp_master_filespc/gp-1
    sdw1:2:/gp_pri_filespc/gp0
    sdw1:3:/gp_mir_filespc/gp1
    sdw2:4:/gp_mir_filespc/gp0
    sdw2:5:/gp_pri_filespc/gp1

Execute the configuration file to create the filespace in Greenplum Database:
$ gpfilespace -c gpfilespace_config_1

See Also
CREATE TABLESPACE
gpinitstandby

Adds and/or initializes a standby master host for a Greenplum Database system.

Synopsis

gpinitstandby { -s standby_hostname | -r }
[-D]
gpinitstandby -? | -v

Description

The gpinitstandby utility adds a backup master host to your Greenplum Database system. If your system has an existing backup master host configured, use the -r option to remove it before adding the new standby master host.

Before running this utility, make sure that the Greenplum Database software is installed on the backup master host and that you have exchanged SSH keys between hosts. Also make sure that the master port is set to the same port number on the master host and the backup master host.

See the Greenplum Database Installation Guide for instructions. This utility should be run on the currently active primary master host.

The utility will perform the following steps:

- Shutdown your Greenplum Database system
- Update the Greenplum Database system catalog to remove the existing backup master host information (if the -r option is supplied)
- Update the Greenplum Database system catalog to add the new backup master host information (use the -n option to skip this step)
- Edit the pg_hba.conf files of the segment instances to allow access from the newly added standby master.
- Setup the backup master instance on the alternate master host
- Start the synchronization process
- Restart your Greenplum Database system

A backup master host serves as a ‘warm standby’ in the event of the primary master host becoming unoperational. The backup master is kept up to date by a transaction log replication process (gpsyncagent), which runs on the backup master host and keeps the data between the primary and backup master hosts synchronized. If the primary master fails, the log replication process is shut down, and the backup master can be activated in its place by using the utility. Upon activation of the backup master, the replicated logs are used to reconstruct the state of the master host at the time of the last successfully committed transaction.

The activated standby master effectively becomes the Greenplum Database master, accepting client connections on the master port and performing normal master operations such as SQL command processing and workload management.
Options

-a (do not prompt)
Do not prompt the user for confirmation.

-D (debug)
Sets logging level to debug.

-l logfile_directory
The directory to write the log file. Defaults to ~/gpAdminLogs.

-L (leave database stopped)
Leave Greenplum Database in a stopped state after removing the warm standby master.

-M fast (fast shutdown - rollback)
Use fast shut down when stopping Greenplum Database at the beginning of the standby initialization process. Any transactions in progress are interrupted and rolled back.

-M smart (smart shutdown - warn)
Use smart shut down when stopping Greenplum Database at the beginning of the standby initialization process. If there are active connections, this command fails with a warning. This is the default shutdown mode.

-n (resynchronize)
Use this option if you already have a standby master configured, and just want to resynchronize the data between the primary and backup master host. The Greenplum system catalog tables will not be updated.

-q (no screen output)
Run in quiet mode. Command output is not displayed on the screen, but is still written to the log file.

-r (remove standby master)
Removes the currently configured standby master host from your Greenplum Database system.

-s standby_hostname
Required. The host name of the standby master host.

-v (show utility version)
Displays the version, status, last updated date, and check sum of this utility.

-? (help)
Displays the online help.
Examples

Add a backup master host to your Greenplum Database system and start the synchronization process:

```bash
gpinitstandby -s host09
```

Remove the existing backup master from your Greenplum system configuration:

```bash
gpinitstandby -r
```

Start an existing backup master host and synchronize the data with the primary master host - do not add a new Greenplum backup master host to the system catalog:

```bash
gpinitstandby -s host09 -n
```
gpinit/system

Initializes a Greenplum Database system using configuration parameters specified in the gpinit/system_config file.

Synopsis

```

`gpinit/system -?`

`gpinit/system -v`

Description

The gpinit/system utility will create a Greenplum Database instance using the values defined in a configuration file. See “Initialization Configuration File Format” on page 637 for more information about this configuration file. Before running this utility, make sure that you have installed the Greenplum Database software on all the hosts in the array.

In a Greenplum Database DBMS, each database instance (the master and all segments) must be initialized across all of the hosts in the system in such a way that they can all work together as a unified DBMS. The gpinit/system utility takes care of initializing the Greenplum master and each segment instance, and configuring the system as a whole.

Before running gpinit/system, you must set the $GPHOME environment variable to point to the location of your Greenplum Database installation on the master host and exchange SSH keys between all host addresses in the array using gpssh-exkeys.

This utility performs the following tasks:

- Verifies that the parameters in the configuration file are correct.
- Ensures that a connection can be established to each host address. If a host address cannot be reached, the utility will exit.
- Verifies that the locale settings.
- Displays the configuration that will be used and prompts the user for confirmation.
- Initializes the master instance.
• Initializes the standby master instance (if specified).
• Initializes the primary segment instances.
• Initializes the mirror segment instances (if mirroring is configured).
• Configures the Greenplum Database system and checks for errors.
• Starts the Greenplum Database system.

## Options

-a (do not prompt)
Do not prompt the user for confirmation.

-B parallel_processes
The number of segments to create in parallel. If not specified, the utility will start up to 4 parallel processes at a time.

-c gpinitconfig
Required. The full path and filename of the configuration file, which contains all of the defined parameters to configure and initialize a new Greenplum system. See “Initialization Configuration File Format” on page 637 for a description of this file.

-D (debug)
Sets log output level to debug.

-h hostfile_gpinitsystem
Optional. The full path and filename of a file that contains the host addresses of your segment hosts. If not specified on the command line, you can specify the host file using the MACHINE_LIST_FILE parameter in the gpinitconfig file.

--locale | -n locale
Sets the default locale used by Greenplum Database. If not specified, the LC_ALL, LC_COLLATE, or LANG environment variable of the master host determines the locale. If these are not set, the default locale is C (POSIX). A locale identifier consists of a language identifier and a region identifier, and optionally a character set encoding. For example, sv_SE is Swedish as spoken in Sweden, en_US is U.S. English, and fr_CA is French Canadian. If more than one character set can be useful for a locale, then the specifications look like this: en_US.UTF-8 (locale specification and character set encoding). On most systems, the command locale will show the locale environment settings and locale -a will show a list of all available locales.

--lc-collate locale
Similar to --locale, but sets the locale used for collation (sorting data). The sort order cannot be changed after Greenplum Database is initialized, so it is important to choose a collation locale that is compatible with the character set encodings that
you plan to use for your data. There is a special collation name of \texttt{C} or \texttt{POSIX} (byte-order sorting as opposed to dictionary-order sorting). The \texttt{C} collation can be used with any character encoding.

\texttt{--lc-ctype locale}
Similar to \texttt{--locale}, but sets the locale used for character classification (what character sequences are valid and how they are interpreted). This cannot be changed after Greenplum Database is initialized, so it is important to choose a character classification locale that is compatible with the data you plan to store in Greenplum Database.

\texttt{--lc-messages locale}
Similar to \texttt{--locale}, but sets the locale used for messages output by Greenplum Database. The current version of Greenplum Database does not support multiple locales for output messages (all messages are in English), so changing this setting will not have any effect.

\texttt{--lc-monetary locale}
Similar to \texttt{--locale}, but sets the locale used for formatting currency amounts.

\texttt{--lc-numeric locale}
Similar to \texttt{--locale}, but sets the locale used for formatting numbers.

\texttt{--lc-time locale}
Similar to \texttt{--locale}, but sets the locale used for formatting dates and times.

\texttt{-l logfile_directory}
The directory to write the log file. Defaults to \texttt{~/gpAdminLogs}.

\texttt{--max_connections | -m number}
Sets the maximum number of client connections allowed to the master. The default is 25.

\texttt{-o (create master only)}
Creates a master instance only. Does not configure any segment hosts.

\texttt{-p postgresql_conf_param_file}
Optional. The name of a file that contains \texttt{postgresql.conf} parameter settings that you want to set for Greenplum Database. These settings will be used when the individual master and segment instances are initialized. You can also set parameters after initialization using the \texttt{gpconfig} utility.

\texttt{-q (no screen output)}
Run in quiet mode. Command output is not displayed on the screen, but is still written to the log file.
--shared_buffers | -b size
Sets the amount of memory a Greenplum server instance uses for shared memory buffers. You can specify sizing in kilobytes (kB), megabytes (MB) or gigabytes (GB). The default is 125MB.

-s standby_master_host
Optional. If you wish to configure a backup master host, specify the host name using this option. The Greenplum Database software must already be installed and configured on this host.

--su_password | -e superuser_password
The password to set for the Greenplum Database superuser. Defaults to gparray. You can always change the superuser password at a later time using the ALTER ROLE command. Client connections over the network require a password login for the database superuser account (for example, the gpadmin user).

-S (spread mirror configuration)
If mirroring parameters are specified, spreads the mirror segments across the available hosts. The default is to group the set of mirror segments together on an alternate host from their primary segment set. Mirror spreading will place each mirror on a different host within the Greenplum Database array. Spreading is only allowed if there is a sufficient number of hosts in the array (number of hosts is greater than the number of segment instances).

-v (show utility version)
Displays the version of this utility.

-? (help)
Displays the online help.

Initialization Configuration File Format
gpinit system requires a configuration file with the following parameters defined. An example initialization configuration file can be found in $GPHOME/docs/cli_help/gpconfigs/gpinit system_config.

ARRAY_NAME
Required. A name for the array you are configuring. You can use any name you like. Enclose the name in quotes if the name contains spaces.

MACHINE_LIST_FILE
Optional. Can be used in place of the -h option. This specifies the file that contains the list of segment host address names that comprise the Greenplum system. The master host is assumed to be the host from which you are running the utility and should not be included in this file. If your segment hosts have multiple network interfaces, then this file would include all addresses for the host. Give the absolute path to the file.
SEG_PREFIX
Required. This specifies a prefix that will be used to name the data directories on the master and segment instances. The naming convention for data directories in a Greenplum Database system is \texttt{SEG\_PREFIXnumber} where \texttt{number} starts with 0 for segment instances (the master is always -1). So for example, if you choose the prefix \texttt{gpseg}, your master instance data directory would be named \texttt{gpseg-1}, and the segment instances would be named \texttt{gpseg0, gpseg1, gpseg2, gpseg3}, and so on.

PORT_BASE
Required. This specifies the base number by which primary segment port numbers are calculated. The first primary segment port on a host is set as \texttt{PORT\_BASE}, and then incremented by one for each additional primary segment on that host. Valid values range from 1 through 65535.

DATA_DIRECTORY
Required. This specifies the data storage location(s) where the utility will create the primary segment data directories. The number of locations in the list dictate the number of primary segments that will get created per physical host (if multiple addresses for a host are listed in the host file, the number of segments will be spread evenly across the specified interface addresses). It is OK to list the same data storage area multiple times if you want your data directories created in the same location. The user who runs \texttt{gpinitsystem} (for example, the \texttt{gpadmin} user) must have permission to write to these directories. For example, this will create six primary segments per host:

\begin{verbatim}
declare -a DATA_DIRECTORY=(/data1/primary /data1/primary /data1/primary /data2/primary /data2/primary /data2/primary)
\end{verbatim}

MASTER_HOSTNAME
Required. The host name of the master instance. This host name must exactly match the configured host name of the machine (run the \texttt{hostname} command to determine the correct hostname).

MASTER_DIRECTORY
Required. This specifies the location where the data directory will be created on the master host. You must make sure that the user who runs \texttt{gpinitsystem} (for example, the \texttt{gpadmin} user) has permissions to write to this directory.

MASTER_PORT
Required. The port number for the master instance. This is the port number that users and client connections will use when accessing the Greenplum Database system.

TRUSTED_SHELL
Required. The shell the \texttt{gpinitsystem} utility uses to execute commands on remote hosts. Allowed values are \texttt{ssh}. You must set up your trusted host environment before running the \texttt{gpinitsystem} utility (you can use \texttt{gpssh-exkeys} to do this).
CHECK_POINT_SEGMENTS
Required. Maximum distance between automatic write ahead log (WAL) checkpoints, in log file segments (each segment is normally 16 megabytes). This will set the `checkpoint_segments` parameter in the `postgresql.conf` file for each segment instance in the Greenplum Database system.

ENCODING
Required. The character set encoding to use. This character set must be compatible with the `--locale` settings used, especially `--lc-collate` and `--lc-ctype`. Greenplum Database supports the same character sets as PostgreSQL.

DATABASE_NAME
Optional. The name of a Greenplum Database database to create after the system is initialized. You can always create a database later using the `CREATE DATABASE` command or the `createdb` utility.

MIRROR_PORT_BASE
Optional. This specifies the base number by which mirror segment port numbers are calculated. The first mirror segment port on a host is set as `MIRROR_PORT_BASE`, and then incremented by one for each additional mirror segment on that host. Valid values range from 1 through 65535 and cannot conflict with the ports calculated by `PORT_BASE`.

REPLICATION_PORT_BASE
Optional. This specifies the base number by which the port numbers for the primary file replication process are calculated. The first replication port on a host is set as `REPLICATION_PORT_BASE`, and then incremented by one for each additional primary segment on that host. Valid values range from 1 through 65535 and cannot conflict with the ports calculated by `PORT_BASE` or `MIRROR_PORT_BASE`.

MIRROR_REPLICATION_PORT_BASE
Optional. This specifies the base number by which the port numbers for the mirror file replication process are calculated. The first mirror replication port on a host is set as `MIRROR_REPLICATION_PORT_BASE`, and then incremented by one for each additional mirror segment on that host. Valid values range from 1 through 65535 and cannot conflict with the ports calculated by `PORT_BASE`, `MIRROR_PORT_BASE`, or `REPLICATION_PORT_BASE`.

MIRROR_DATA_DIRECTORY
Optional. This specifies the data storage location(s) where the utility will create the mirror segment data directories. There must be the same number of data directories declared for mirror segment instances as for primary segment instances (see the `DATA_DIRECTORY` parameter). The user who runs `gpinitssystem` (for example, the `gpadmin` user) must have permission to write to these directories. For example:

```
declare -a MIRROR_DATA_DIRECTORY=(/data1/mirror /data1/mirror /data1/mirror /data2/mirror /data2/mirror)
```
Examples

Initialize a Greenplum Database array by supplying a configuration file and a segment host address file, and set up a spread mirroring (-S) configuration:

$ gpinitsystem -c gpinitsystem_config -h hostfile_gpinitsystem -S

Initialize a Greenplum Database array and set the superuser remote password:

$ gpinitsystem -c gpinitsystem_config -h hostfile_gpinitsystem --su-password mypassword

Initialize a Greenplum Database array with an optional standby master host:

$ gpinitsystem -c gpinitsystem_config -h hostfile_gpinitsystem -s host09

See Also

gpssh-exkeys, gpdeletesystem
gpload

Runs a load job as defined in a YAML formatted control file.

Synopsis

```
gpload --version
```

Prerequisites

The client machine where `gpload` is executed must have the following:

- Python 2.6.2 or later, `psycopg2` (the Python interface to PostgreSQL), and `pyyaml`. Note that Python and the required Python libraries are included with the Greenplum Database server installation, so if you have Greenplum Database installed on the machine where `gpload` is running, you do not need a separate Python installation.

- The `gpfdist` parallel file distribution program installed and in your `$PATH`. This program is located in `$GPHOME/bin` of your Greenplum Database server installation.

- Network access to and from all hosts in your Greenplum Database array (master and segments).

- Network access to and from the hosts where the data to be loaded resides (ETL servers).

Description

`gpload` is a data loading utility that acts as an interface to Greenplum Database’s external table parallel loading feature. Using a load specification defined in a YAML formatted control file, `gpload` executes a load by invoking the Greenplum parallel file server (`gpfdist`), creating an external table definition based on the source data defined, and executing an `INSERT`, `UPDATE` or `MERGE` operation to load the source data into the target table in the database.

Options

- `-f control_file`

  Required. A YAML file that contains the load specification details. See “Control File Format” on page 643.
---gpdist_timeout seconds
Sets the timeout for the gpfdist parallel file distribution program to send a response. Enter a value from 0 to 30 seconds (entering “0” to disables timeouts). Note that you might need to increase this value when operating on high-traffic networks.

-l log_file
Specifies where to write the log file. Defaults to ~/gpAdminLogs/gpload_YYYYMMDD. See also, “Log File Format” on page 650.

-v (verbose mode)
Show verbose output of the load steps as they are executed.

-V (very verbose mode)
Shows very verbose output.

-q (no screen output)
Run in quiet mode. Command output is not displayed on the screen, but is still written to the log file.

-D (debug mode)
Check for error conditions, but do not execute the load.

-? (show help)
Show help, then exit.

--version
Show the version of this utility, then exit.

Connection Options

-d database
The database to load into. If not specified, reads from the load control file, the environment variable $PGDATABASE or defaults to the current system user name.

-h hostname
Specifies the host name of the machine on which the Greenplum master database server is running. If not specified, reads from the load control file, the environment variable $PGHOST or defaults to localhost.

-p port
Specifies the TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the load control file, the environment variable $PGPORT or defaults to 5432.

-U username
The database role name to connect as. If not specified, reads from the load control file, the environment variable $PGUSER or defaults to the current system user name.
-W (force password prompt)

Force a password prompt. If not specified, reads the password from the environment variable \$PGPASSWORD or from a password file specified by \$PGPASSFILE or in ~/.pgpass. If these are not set, then gpload will prompt for a password even if -W is not supplied.

Control File Format

The gpload control file uses the YAML 1.1 document format and then implements its own schema for defining the various steps of a Greenplum Database load operation. The control file must be a valid YAML document.

The gpload program processes the control file document in order and uses indentation (spaces) to determine the document hierarchy and the relationships of the sections to one another. The use of white space is significant. White space should not be used simply for formatting purposes, and tabs should not be used at all.

The basic structure of a load control file is:

```yaml
---
VERSION: 1.0.0.1
DATABASE: db_name
USER: db_username
HOST: master_hostname
PORT: master_port
GPLOAD:
  INPUT:
    - SOURCE:
      LOCAL_HOSTNAME:
        - hostname_or_ip
      PORT: http_port
      | PORT_RANGE: [start_port_range, end_port_range]
    FILE:
      - /path/to/input_file
    - COLUMNS:
      - field_name: data_type
    - FORMAT: text | csv
    - DELIMITER: 'delimiter_character'
    - ESCAPE: 'escape_character' | 'OFF'
    - NULL_AS: 'null_string'
    - FORCE_NOT_NULL: true | false
    - QUOTE: 'csv_quote_character'
    - HEADER: true | false
    - ENCODING: database_encoding
```
- **ERROR_LIMIT**: integer
- **ERROR_TABLE**: schema.table_name

**OUTPUT:**
- **TABLE**: schema.table_name
- **MODE**: insert | update | merge
- **MATCH_COLUMNS**:
  - target_column_name
- **UPDATE_COLUMNS**:
  - target_column_name
- **UPDATE_CONDITION**: 'boolean_condition'
- **MAPPING**:
  - target_column_name: source_column_name | 'expression'

**PRELOAD**:
- **TRUNCATE**: true | false

**SQL**:
- **BEFORE**: "sql_command"
- **AFTER**: "sql_command"

**VERSION**
Optional. The version of the gpload control file schema. The current version is 1.0.0.1.

**DATABASE**
Optional. Specifies which database in Greenplum to connect to. If not specified, defaults to $PGDATABASE if set or the current system user name. You can also specify the database on the command line using the -d option.

**USER**
Optional. Specifies which database role to use to connect. If not specified, defaults to the current user or $PGUSER if set. You can also specify the database role on the command line using the -U option.

If the user running gpload is not a Greenplum superuser, then the server configuration parameter gp_external_grant_privileges must be set to on in order for the load to be processed.

**HOST**
Optional. Specifies Greenplum master host name. If not specified, defaults to localhost or $PGHOST if set. You can also specify the master host name on the command line using the -h option.
PORT
Optional. Specifies Greenplum master port. If not specified, defaults to 5432 or $PGPORT if set. You can also specify the master port on the command line using the -p option.

GPLOAD
Required. Begins the load specification section. A gpload specification must have an INPUT and an OUTPUT section defined.

INPUT
Required. Defines the location and the format of the input data to be loaded. gpload will start one or more instances of the gpfdist file distribution program on the current host and create the required external table definition(s) in Greenplum Database that point to the source data. Note that the host from which you run gpload must be accessible over the network by all Greenplum hosts (master and segments).

SOURCE
Required. The SOURCE block of an INPUT specification defines the location of a source file. An INPUT section can have more than one SOURCE block defined. Each SOURCE block defined corresponds to one instance of the gpfdist file distribution program that will be started on the local machine. Each SOURCE block defined must have a FILE specification.

See “Using the Greenplum Parallel File Server (gpfdist)” on page 129 for more information about single and multiple gpfdist instances.

LOCAL_HOSTNAME
Optional. Specifies the host name or IP address of the local machine on which gpload is running. If this machine is configured with multiple network interface cards (NICs), you can specify the host name or IP of each individual NIC to allow network traffic to use all NICs simultaneously. The default is to use the local machine’s primary host name or IP only.

PORT
Optional. Specifies the specific port number that the gpfdist file distribution program should use. You can also supply a PORT_RANGE to select an available port from the specified range. If both PORT and PORT_RANGE are defined, then PORT takes precedence. If neither PORT or PORT_RANGE are defined, the default is to select an available port between 8000 and 9000.

If multiple host names are declared in LOCAL_HOSTNAME, this port number is used for all hosts. This configuration is desired if you want to use all NICs to load the same file or set of files in a given directory location.
**PORT_RANGE**
Optional. Can be used instead of **PORT** to supply a range of port numbers from which *gpload* can choose an available port for this instance of the *gpfdist* file distribution program.

**FILE**
Required. Specifies the location of a file, named pipe, or directory location on the local file system that contains data to be loaded. You can declare more than one file so long as the data is of the same format in all files specified.

If the files are compressed using *gzip* or *bzip2* (have a .gz or .bz2 file extension), the files will be uncompressed automatically (provided that *gunzip* or *bunzip2* is in your path).

When specifying which source files to load, you can use the wildcard character (*) or other C-style pattern matching to denote multiple files. The files specified are assumed to be relative to the current directory from which *gpload* is executed (or you can declare an absolute path).

**COLUMNS**
Optional. Specifies the schema of the source data file(s) in the format of *field_name: data_type*. The **DELIMITER** character in the source file is what separates two data value fields (columns). A row is determined by a line feed character (0x0a).

If the input **COLUMNS** are not specified, then the schema of the output **TABLE** is implied, meaning that the source data must have the same column order, number of columns, and data format as the target table.

The default source-to-target mapping is based on a match of column names as defined in this section and the column names in the target **TABLE**. This default mapping can be overridden using the **MAPPING** section.

**FORMAT**
Optional. Specifies the format of the source data file(s) - either plain text (**TEXT**) or comma separated values (**CSV**) format. Defaults to **TEXT** if not specified. For more information about the format of the source data, see “Formatting Your Data Files” on page 125.

**DELIMITER**
Optional. Specifies a single ASCII character that separates columns within each row (line) of data. The default is a tab character in **TEXT** mode, a comma in **CSV** mode. You can also specify a non-printable ASCII character via an escape sequence using the decimal representation of the ASCII character. For example, \014 represents the shift out character.
ESCAPE
Specifies the single character that is used for C escape sequences (such as \n, \t, \l00, and so on) and for escaping data characters that might otherwise be taken as row or column delimiters. Make sure to choose an escape character that is not used anywhere in your actual column data. The default escape character is a \ (backslash) for text-formatted files and a " (double quote) for csv-formatted files, however it is possible to specify another character to represent an escape. It is also possible to disable escaping in text-formatted files by specifying the value 'OFF' as the escape value. This is very useful for data such as text-formatted web log data that has many embedded backslashes that are not intended to be escapes.

NULL_AS
Optional. Specifies the string that represents a null value. The default is \N (backslash-N) in TEXT mode, and an empty value with no quotations in CSV mode. You might prefer an empty string even in TEXT mode for cases where you do not want to distinguish nulls from empty strings. Any source data item that matches this string will be considered a null value.

FORCE_NOT_NULL
Optional. In CSV mode, processes each specified column as though it were quoted and hence not a NULL value. For the default null string in CSV mode (nothing between two delimiters), this causes missing values to be evaluated as zero-length strings.

QUOTE
Required when FORMAT is CSV. Specifies the quotation character for CSV mode. The default is double-quote (").

HEADER
Optional. Specifies that the first line in the data file(s) is a header row (contains the names of the columns) and should not be included as data to be loaded. If using multiple data source files, all files must have a header row. The default is to assume that the input files do not have a header row.

ENCODING
Optional. Character set encoding of the source data. Specify a string constant (such as 'SQL_ASCII'), an integer encoding number, or 'DEFAULT' to use the default client encoding. See “Character Set Support” on page 939. If not specified, the default client encoding is used.

ERROR_LIMIT
Optional. Enables single row error isolation mode for this load operation. When enabled, input rows that have format errors will be discarded provided that the error limit count is not reached on any Greenplum segment instance during input processing. If the error limit is not reached, all good rows will be loaded and any error rows will either be discarded or logged to the table.
specified in **ERROR_TABLE**. The default is to abort the load operation on the first error encountered. Note that single row error isolation only applies to data rows with format errors; for example, extra or missing attributes, attributes of a wrong data type, or invalid client encoding sequences. Constraint errors, such as primary key violations, will still cause the load operation to abort if encountered. See “Handling Load Errors” on page 133 for more information.

**ERROR_TABLE**

Optional when **ERROR_LIMIT** is declared. Specifies an error table where rows with formatting errors will be logged when running in single row error isolation mode. You can then examine this error table to see error rows that were not loaded (if any). If the **error_table** specified already exists, it will be used. If it does not exist, it will be automatically generated. See “Handling Load Errors” on page 133 for more information about error tables.

**OUTPUT**

Required. Defines the target table and final data column values that are to be loaded into the database.

**TABLE**

Required. The name of the target table to load into.

**MODE**

Optional. Defaults to INSERT if not specified. There are three available load modes:

**INSERT** - Loads data into the target table using the following method:

```
INSERT INTO target_table SELECT * FROM input_data;
```

**UPDATE** - Updates the **UPDATE_COLUMNS** of the target table where the rows have **MATCH_COLUMNS** attribute values equal to those of the input data, and the optional **UPDATE_CONDITION** is true. **UPDATE** is not supported on tables with a random distribution policy.

**MERGE** - Inserts new rows and updates the **UPDATE_COLUMNS** of existing rows where **MATCH_COLUMNS** attribute values are equal to those of the input data, and the optional **UPDATE_CONDITION** is true. New rows are identified when the **MATCH_COLUMNS** value in the source data does not have a corresponding value in the existing data of the target table. If there are multiple new **MATCH_COLUMNS** values that are the same, only one new row for that value will be inserted (use **UPDATE_CONDITION** to filter out the rows you want to discard). **MERGE** is not supported on tables with a random distribution policy.

**MATCH_COLUMNS**

Required if **MODE** is **UPDATE** or **MERGE**. Specifies the column(s) to use as the join condition for the update. The attribute value in the specified target column(s) must be equal to that of the corresponding source data column(s) in
order for the row to be updated in the target table. The \texttt{MATCH\_COLUMNS}
declared must be the same as the Greenplum distribution key columns for the
table.

\textbf{UPDATE\_COLUMNS}
Required if \texttt{MODE} is \texttt{UPDATE} or \texttt{MERGE}. Specifies the column(s) to update for
the rows that meet the \texttt{MATCH\_COLUMNS} criteria and the optional
\texttt{UPDATE\_CONDITION}. Update columns cannot be columns that are used for the
Greenplum distribution key for the table.

\textbf{UPDATE\_CONDITION}
Optional. Specifies a Boolean condition (similar to what you would declare in
a \texttt{WHERE} clause) that must be met in order for a row in the target table to be
updated (or inserted in the case of a \texttt{MERGE}).

\textbf{MAPPING}
Optional. If a mapping is specified, it overrides the default source-to-target
column mapping. The default source-to-target mapping is based on a match
of column names as defined in the source \texttt{COLUMNS} section and the column
names of the target \texttt{TABLE}. A mapping is specified as either:
\begin{itemize}
  \item \texttt{target\_column\_name: source\_column\_name}
  \item \texttt{target\_column\_name: 'expression'}
\end{itemize}
Where \texttt{expression} is any expression that you would specify in the \texttt{SELECT}
list of a query, such as a constant value, a column reference, an operator
invocation, a function call, and so on.

\textbf{PRELOAD}
Optional. Specifies operations to run prior to the load operation. Right now the
only preload operation is \texttt{TRUNCATE}.

\textbf{TRUNCATE}
Optional. If set to true, \texttt{gpload} will remove all rows in the target table prior
to loading it.

\textbf{SQL}
Optional. Defines SQL commands to run before and/or after the load operation.
You can specify multiple \texttt{BEFORE} and/or \texttt{AFTER} commands. List commands in the
order of desired execution.

\textbf{BEFORE}
Optional. An SQL command to run before the load operation starts. Enclose
commands in quotes.

\textbf{AFTER}
Optional. An SQL command to run after the load operation completes.
Enclose commands in quotes.
Notes
If your database object names were created using a double-quoted identifier (delimited identifier), you must specify the delimited name within single quotes in the `gpload` control file. For example, if you create a table as follows:

```sql
CREATE TABLE "MyTable" ("MyColumn" text);
```

Your YAML-formatted `gpload` control file would refer to the above table and column names as follows:

```
- COLUMNS:
  - "MyColumn": text
OUTPUT:
  - TABLE: public."MyTable"
```

Log File Format
Log files output by `gpload` have the following format:

```
timestamp|level|message
```

Where `timestamp` takes the form: `YYYY-MM-DD HH:MM:SS`, `level` is one of `DEBUG`, `LOG`, `INFO`, `ERROR`, and `message` is a normal text message.

Some `INFO` messages that may be of interest in the log files are (where `#` corresponds to the actual number of seconds, units of data, or failed rows):

```
INFO|running time: #.## seconds
INFO|transferred #.# kB of #.# kB.
INFO|gpload succeeded
INFO|gpload succeeded with warnings
INFO|gpload failed
INFO|1 bad row
INFO|# bad rows
```

Examples
Run a load job as defined in `my_load.yml`:

```
gpload -f my_load.yml
```

Example load control file:
```
---
VERSION: 1.0.0.1
DATABASE: ops
USER: gpadmin
HOST: mdw-1
PORT: 5432
GPLOAD:
  INPUT:
    - SOURCE:
      LOCAL_HOSTNAME:
        - et11-1
```
- etl1-2
- etl1-3
- etl1-4
PORT: 8081
FILE:
- /var/load/data/*
- COLUMNS:
  - name: text
  - amount: float4
  - category: text
  - desc: text
  - date: date
- FORMAT: text
- DELIMITER: '|' 
- ERROR_LIMIT: 25
- ERROR_TABLE: payables.err_expenses
OUTPUT:
- TABLE: payables.expenses
- MODE: INSERT
SQL:
- BEFORE: "INSERT INTO audit VALUES('start', current_timestamp)"
- AFTER: "INSERT INTO audit VALUES('end', current_timestamp)"

See Also

gpfdist, CREATE EXTERNAL TABLE
gplogfilter

Searches through Greenplum Database log files for specified entries.

Synopsis

```
gplogfilter [timestamp_options] [pattern_options] [output_options] [input_options] [input_file]
gplogfilter --help
 gplogfilter --version
```

Description

The `gplogfilter` utility can be used to search through a Greenplum Database log file for entries matching the specified criteria. If an input file is not supplied, then `gplogfilter` will use the `$MASTER_DATA_DIRECTORY` environment variable to locate the Greenplum master log file in the standard logging location. To read from standard input, use a dash (`-`) as the input file name. Input files may be compressed using `gzip`. In an input file, a log entry is identified by its timestamp in `YYYY-MM-DD [hh:mm[:ss]]` format.

You can also use `gplogfilter` to search through all segment log files at once by running it through the `gpssh` utility. For example, to display the last three lines of each segment log file:

```
gpssh -f seg_host_file
 => source /usr/local/greenplum-db/greenplum_path.sh
 => gplogfilter -n 3 /gpdata/*/pg_log/gpdb*.log
```

By default, the output of `gplogfilter` is sent to standard output. Use the `-o` option to send the output to a file or a directory. If you supply an output file name ending in `.gz`, the output file will be compressed by default using maximum compression. If the output destination is a directory, the output file is given the same name as the input file.

Options

### Timestamp Options

**-b datetime | --begin=datatime**

Specifies a starting date and time to begin searching for log entries in the format of `YYYY-MM-DD [hh:mm[:ss]]`.

**-e datetime | --end=datatime**

Specifies an ending date and time to stop searching for log entries in the format of `YYYY-MM-DD [hh:mm[:ss]]`. 
gplogfilter

-d time | --duration=time
Specifies a time duration to search for log entries in the format of [hh][:mm[:ss]]. If used without either the -b or -e option, will use the current time as a basis.

Pattern Matching Options

-c i[gnore]|r[espect] | --case=i[gnore]|r[espect]
Matching of alphabetic characters is case sensitive by default unless proceeded by the --case=ignore option.

-C '<string>' | --columns='<string>'
Selects specific columns from the log file. Specify the desired columns as a comma-delimited string of column numbers beginning with 1, where the second column from left is 2, the third is 3, and so on. See “Log File Format” on page 207 for a list of the available columns and their associated number.

-f 'string' | --find='string'
Finds the log entries containing the specified string.

-F 'string' | --nofind='string'
Rejects the log entries containing the specified string.

-m regex | --match=regex
Finds log entries that match the specified Python regular expression. See http://docs.python.org/library/re.html for Python regular expression syntax.

-M regex | --nomatch=regex
Rejects log entries that match the specified Python regular expression. See http://docs.python.org/library/re.html for Python regular expression syntax.

-t | --trouble
Finds only the log entries that have ERROR:, FATAL:, or PANIC: in the first line.

Output Options

-n integer | --tail=integer
Limits the output to the last integer of qualifying log entries found.

-s offset [limit] | --slice=offset [limit]
From the list of qualifying log entries, returns the limit number of entries starting at the offset entry number, where an offset of zero (0) denotes the first entry in the result set and an offset of any number greater than zero counts back from the end of the result set.

-o output_file | --out=output_file
Writes the output to the specified file or directory location instead of STDOUT.
-z 0-9 | --zip=0-9
Compresses the output file to the specified compression level using gzip, where 0 is no compression and 9 is maximum compression. If you supply an output file name ending in .gz, the output file will be compressed by default using maximum compression.

-a | --append
If the output file already exists, appends to the file instead of overwriting it.

Input Options

input_file
The name of the input log file(s) to search through. If an input file is not supplied, gplogfilter will use the $MASTER_DATA_DIRECTORY environment variable to locate the Greenplum master log file. To read from standard input, use a dash (-) as the input file name.

-u | --unzip
Uncompress the input file using gunzip. If the input file name ends in .gz, it will be uncompressed by default.

--help
Displays the online help.

--version
Displays the version of this utility.

Examples
Display the last three error messages in the master log file:

gplogfilter -t -n 3

Display all log messages in the master log file timestamped in the last 10 minutes:

gplogfilter -d :10

Display log messages in the master log file containing the string |con6 cmd11|:

gplogfilter -f '|con6 cmd11|

Using gpssh, run gplogfilter on the segment hosts and search for log messages in the segment log files containing the string con6 and save output to a file.

gpssh -f seg_hosts_file -e 'source /usr/local/greenplum-db/greenplum_path.sh ; gplogfilter -f con6 /gpdata/*/pg_log/gpdb.log' > seglog.out

See Also

gpssh, gscp
gpmapreduce

Runs Greenplum MapReduce jobs as defined in a YAML specification document.

Synopsis

gpmapreduce -f yaml_file [dbname [username]] [-k name=value | --key name=value] [-h hostname | --host hostname] [-p port | --port port] [-U username | --username username] [-W] [-v]
gpmapreduce -V | --version

Prerequisites

The following are required prior to running this program:

- You must have your MapReduce job defined in a YAML file. See “Greenplum MapReduce Specification” on page 799.
- You must be a Greenplum Database superuser to run MapReduce jobs written in untrusted Perl or Python.
- You must be a Greenplum Database superuser to run MapReduce jobs with EXEC and FILE inputs.
- You must be a Greenplum Database superuser to run MapReduce jobs with GPFDIST input unless the server configuration parameter `gp_external_grant_privileges` is set to `on`.

Description

MapReduce is a programming model developed by Google for processing and generating large data sets on an array of commodity servers. Greenplum MapReduce allows programmers who are familiar with the MapReduce paradigm to write map and reduce functions and submit them to the Greenplum Database parallel engine for processing.

In order for Greenplum to be able to process MapReduce functions, the functions need to be defined in a YAML document, which is then passed to the Greenplum MapReduce program, `gpmapreduce`, for execution by the Greenplum Database parallel engine. The Greenplum system takes care of the details of distributing the input data, executing the program across a set of machines, handling machine failures, and managing the required inter-machine communication.

Options

-f yaml_file

Required. The YAML file that contains the Greenplum MapReduce job definitions. See “Greenplum MapReduce Specification” on page 799.
-? | --help
    Show help, then exit.

-V | --version
    Show version information, then exit.

-v | --verbose
    Show verbose output.

-k | --key name=value
    Sets a YAML variable. A value is required. Defaults to “key” if no variable name is specified.

### Connection Options

-h host | --host host
    Specifies the host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable PGHOST or defaults to localhost.

-p port | --port port
    Specifies the TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable PGPORT or defaults to 5432.

-U username | --username username
    The database role name to connect as. If not specified, reads from the environment variable PGUSER or defaults to the current system user name.

-W | --password
    Force a password prompt.

### Examples

Run a MapReduce job as defined in *my_yaml.txt* and connect to the database *mydatabase*:

    gpmapreduce -f my_yaml.txt mydatabase

### See Also

Greenplum MapReduce Specification
gpmigrator

Upgrades an existing Greenplum Database 3.3.x system to 4.0.x

Synopsis

gpmigrator old_GPHOME_path new_GPHOME_path
   --replication_port_base port
   --mirror_replication_port_base port
   --mirror-mode {redundant | single}
   --gpperfmonhome path
   [ --nosnapshot] [-d master_data_directory]
   [-l logfile_directory] [-q] [- debug] [-R]

Prerequisites

The following tasks should be performed prior to executing an upgrade:

- Make sure you are logged in to the master host as the Greenplum Database superuser (gpadmin).
- Install the Greenplum Database 4.0 binaries on all Greenplum hosts.
- Copy any custom modules from your existing installation to your 4.0 installation on all Greenplum hosts. For example, shared library files for user-defined functions in $GPHOME/lib or PostgreSQL add-on modules (such as plr.so or pgcrypto.so) in $GPHOME/lib/postgresql.
- Copy or preserve any additional folders or files (such as backup folders) that you have added in the Greenplum data directories or $GPHOME directory. Only files or folders strictly related to Greenplum Database operations are preserved by the migration utility.
- (Optional) Run VACUUM on all databases, and remove old server log files from pg_log in your master and segment data directories. This is not required, but will reduce the size of Greenplum Database files to be backed up and migrated.
- Remove the gp_jetpack schema from all databases (if it exists). Also note that gp_ is a reserved prefix for schema names in Greenplum Database 4.0 (as is pg_). If you have named any schemas with a gp_ prefix, Greenplum recommends that you rename them prior to upgrading.
- The use of OIDs (per-row system-generated object identifiers) has been disabled for partitioned and column-oriented tables in 4.0. If you have these types of tables that were created with OIDS=TRUE, you must alter them to be with OIDS=FALSE.
- Check for and recover any failed segments in your current Greenplum Database system (gpstat, gprecoverset).
- (Optional) Verify the integrity of your current system catalogs using the gpcheckcat utility.
- (Optional, but highly recommended) Backup your current databases (gpcrondump or ZFS snapshots). If you find any issues when testing your upgraded system, you can restore this backup.

- Remove the standby master from your system configuration (gpinitstandby -r).

- Do a clean shutdown of your current system (gpstop).

- Update your environment to source the 4.0 installation.

- Inform all database users of the upgrade and lockout time frame. Once the upgrade is in process, users will not be allowed on the system until the upgrade is complete.

Note: Failing to meet any of these prerequisites can result in an aborted upgrade, placing your system in an unusable or even unrecoverable state. Make sure you meet all the prerequisites before running gpmigrator.

Description
The gpmigrator utility upgrades an existing Greenplum Database 3.3.x.x system to 4.0. This utility updates the system catalog and internal version number, but not the actual software binaries.

During the migration process, all client connections to the master will be locked out. The migration utility locks all client access to the master, however it does not block direct utility mode access to the segments. To ensure a safe upgrade, make sure users are aware that any connections to a Greenplum segment are not safe during the upgrade time frame.

Options

old_GPHOME_path
Required. The absolute path to the current version of Greenplum Database software you want to migrate away from.

new_GPHOME_path
Required. The absolute path to the new version of Greenplum Database software you want to migrate to.

--replication_port_base port
Required if upgrading a system with mirrors. Specifies the base number by which the port numbers for the primary file replication process are calculated. The first replication port on a host is set as REPLICATION_PORT_BASE, and then incremented by one for each additional primary segment on that host. Valid values range from 1 through 65535 and cannot conflict with other port assignments on a segment host.
--mirror_replication_port_base port
Required if upgrading a system with mirrors. Specifies the base number by which
the port numbers for the mirror file replication process are calculated. The first
mirror replication port on a host is set as MIRROR_REPLICATION_PORT_BASE, and
then incremented by one for each additional mirror segment on that host. Valid
values range from 1 through 65535 and cannot conflict with other port assignments
on a segment host.

--mirror-mode {redundant | single}
Required if upgrading a system with mirrors. Specifies how to upgrade mirror
segments to the new 4.0 format. Redundant mode will keep a copy of the old 3.3
mirror until its new mirror is fully synchronized. Redundant mode always ensures
that the system always has a valid mirror copy for every segment, and is the
preferred mode provided that your systems have enough available disk space. Single
mode removes the old mirror before creating the new mirror on the target node, and
can be used on systems where disk space is limited. All segments are upgraded in
parallel using the maximum parallelism allowed by the available network, I/O, and
CPU bandwidth.

--gpperfmonhome path
Required only if your current system has the Greenplum Performance Monitor
installed and enabled. Provide the full path to your 3.3.x Greenplum Performance
Monitor installation directory ($GPPERFMONHOME). The upgrade process will stop
the monitor data collection agents and web application. You must then upgrade the
Greenplum Performance Monitor Web Application to 4.0 after upgrade is complete.

--nosnapshot
(Solaris only) On systems that support zfs snapshots the default behavior is to pause
the upgrade after the mirrors have been synchronized, but before the upgraded
catalog has been installed. This allows users to take a zfs snapshot before upgrade
progresses beyond the point of no return. The --nosnapshot option runs the
complete upgrade as a single step, just as it would do on systems that do not support
zfs.

--resume
(Solaris only) On systems that support zfs snapshots, allows the upgrade process to
resume where it left off after it has been paused to allow for a zfs snapshot to be
taken.

-R (revert)
Restores the backup directory in the event of a failed upgrade. This is only available
if running in redundant mode (mirror-mode=redundant).

If the migration procedure fails and you cannot start the system using the binary
files of the prior version, it is probable that gpmigrator was not able to revert
automatically. In these cases, you can run gpmigrator -R to attempt to restore the
backup directory.

Cases in which the migration is not automatically reverted include:
- The `gpmigrator` process is killed with a command such as `kill -9`
- The system crashes during the migration procedure
- You are running `gpmigrator` in `mirror-mode=single` mode

If a ZFS snapshot is available, backing up from those resources is preferred.

- **-d master_data_directory**
  Optional. The current master host data directory. If not specified, the value set for `$MASTER_DATA_DIRECTORY` will be used.

- **-l logfile_directory**
  The directory to write the log file. Defaults to `~/gpAdminLogs`.

- **-q (quiet mode)**
  Run in quiet mode. Command output is not displayed on the screen, but is still written to the log file.

- **--help | -h**
  Displays the online help.

- **--debug**
  Sets logging level to debug.

- **--version | -v**
  Displays the version of this utility.

### Examples

Upgrade to version 4.0 from version 3.3.6.1 (make sure you are using the 4.0 version of `gpmigrator`):

```
/usr/local/greenplum-db-4.0.x.x/bin/gpmigrator \
   --replication_port_base 53000 \
   --mirror_replication_port_base 54000 \
   --mirror-mode redundant \
   --gpperfmonhome /usr/local/greenplum-perfmon-2.1.0.0 \
   /usr/local/greenplum-db-3.3.6.0 \
   /usr/local/greenplum-db-4.0.0.0
```

Resume an upgrade that has been paused for a snapshot (make sure you are using the 4.0 version of `gpmigrator`):

```
/usr/local/greenplum-db-4.0.x.x/bin/gpmigrator \
/usr/local/greenplum-db-3.3.6.1 \
/usr/local/greenplum-db-4.0.x.x --resume
```

### See Also

`gpstop, gpstate, gprecoverseg, gpcrondump, gpbitmapreindex`
**gpperfmon_install**

Installs the `gpperfmon` database and optionally enables the data collection agents.

**Synopsis**

```
gpperfmon_install
   [--enable --password gpmon_password --port gpdb_port]
   [--pgpass path_to_file]
   [--verbose]
```

**Description**

The `gpperfmon_install` utility automates the steps required to enable the performance monitor data collection agents. You must be the Greenplum Database administrative user (`gpadmin`) in order to run this utility. If using the `--enable` option, Greenplum Database must be restarted after the utility completes.

When run without any options, the utility will just create the `gpperfmon` database (the database used to store performance monitor data). When run with the `--enable` option, the utility will also run the following additional tasks necessary to enable the performance monitor data collection agents:

1. Creates the `gpmon` superuser role in Greenplum Database. The performance monitor data collection agents require this role to connect to the database and write their data. The `gpmon` superuser role uses MD5-encrypted password authentication by default. Use the `--password` option to set the `gpmon` superuser’s password. Use the `--port` option to supply the port of the Greenplum Database master instance.

2. Updates the `$MASTER_DATA_DIRECTORY/pg_hba.conf` file. The utility will add the following line to the host-based authentication file (`pg_hba.conf`). This allows the `gpmon` user to locally connect to any database using MD5-encrypted password authentication:
   
   ```
   local all gpmon md5
   ```

3. Updates the password file (`.pgpass`). In order to allow the data collection agents to connect as the `gpmon` role without a password prompt, you must have a password file that has an entry for the `gpmon` user. The utility add the following entry to your password file (if the file does not exist, the utility will create it):
   
   ```
   *:5432:gpperfmon:gpmon:gpmon_password
   ```
   
   If your password file is not located in the default location (`~/.pgpass`), use the `--pgpass` option to specify the file location.

4. Sets the server configuration parameters for performance monitor. The following parameters must be enabled in order for the data collection agents to begin collecting data. The utility will set the following parameters in the Greenplum Database `postgresql.conf` configuration files:
   
   ```
   gp_enable_gpperfmon=on
   ```
   
   (in all `postgresql.conf` files)
Options

--enable
In addition to creating the gpperfmon database, performs the additional steps required to enable the performance monitor data collection agents. When --enable is specified the utility will also create and configure the gpmon superuser account and set the performance monitor server configuration parameters in the postgresql.conf files.

--password gpmon_password
Required if --enable is specified. Sets the password of the gpmon superuser.

--port gpdb_port
Required if --enable is specified. Specifies the connection port of the Greenplum Database master.

--pgpass path_to_file
Optional if --enable is specified. If the password file is not in the default location of ~/.pgpass, specifies the location of the password file.

--verbose
Sets the logging level to verbose.

--help | -h | -?
Displays the online help.

Examples
Create the gpperfmon database only:
$ su - gpadmin
$ gpperfmon_install

Create the gpperfmon database, create the gpmon superuser, and enable the performance monitor agents:
$ su - gpadmin
$ gpperfmon_install --enable --password p@$word --port 5432
$ gpstop -r

See Also
gpstop


**gprecoverseg**

Recovers a primary or mirror segment instance that has been marked as down (if mirroring is enabled).

**Synopsis**

```
```

**Description**

In a system with mirrors enabled, the gprecoverseg utility reactivates a failed segment instance and identifies the changed database files that require resynchronization. Once gprecoverseg completes this process, the system goes into resynchronization mode until the recovered segment is brought up to date. The system is online and fully operational during resynchronization.

A segment instance can fail for several reasons, such as a host failure, network failure, or disk failure. When a segment instance fails, its status is marked as down in the Greenplum Database system catalog, and its mirror is activated in change tracking mode. In order to bring the failed segment instance back into operation again, you must first correct the problem that made it fail in the first place, and then recover the segment instance in Greenplum Database using gprecoverseg.

Segment recovery using gprecoverseg requires that you have an active mirror to recover from. For systems that do not have mirroring enabled, or in the event of a double fault (a primary and mirror pair both down at the same time) — do a system restart to bring the segments back online (gpstop -r).

By default, a failed segment is recovered in place, meaning that the system brings the segment back online on the same host and data directory location on which it was originally configured. In some cases, this may not be possible (for example, if a host was physically damaged and cannot be recovered). In this situation, gprecoverseg allows you to recover failed segments to a completely new host (using -p), on an alternative data directory location on your remaining live segment hosts (using -s), or by supplying a recovery configuration file (using -i) in the format of:

```
filespaceOrder=[filespace1_fsnamed:filespace2_fsnamed:...]
failed_host_address:port:fselocation:[fselocation:...]
[recovery_host_address:port:replication_port:fselocation:[fselocation:...]]
```

For example (if recovering to a different host than the failed host, and you do not have additional filespaces configured besides the default pg_system filespace):

```
filespaceOrder=
```
The `gp_segment_configuration`, `pg_filespace`, and `pg_filespace_entry` system catalog tables can help you determine your current segment configuration so that you can plan your mirror recovery configuration. For example, run the following query:

```
=# SELECT dbid, content, address, port, 
    replication_port, fselocation as datadir 
FROM gp_segment_configuration, pg_filespace_entry 
WHERE dbid=fseidbid 
ORDER BY dbid;
```

The new recovery segment host must be pre-installed with the Greenplum Database software and configured exactly the same as the existing segment hosts. A spare data directory location must exist on all currently configured segment hosts and have enough disk space to accommodate the failed segments.

The recovery process marks the segment as up again in the Greenplum Database system catalog, and then initiates the resynchronization process to bring the transactional state of the segment up-to-date with the latest changes. The system is online and available during resynchronization. To check the status of the resynchronization process run:

```
gpstate -m
```

If you do not have mirroring enabled or if you have both a primary and its mirror down, you must take manual steps to recover the failed segment instances and then restart the system, for example:

```
gpstop -r
```

### Options

- **-a (do not prompt)**
  Do not prompt the user for confirmation.

- **-B parallel_processes**
  The number of segments to recover in parallel. If not specified, the utility will start up to four parallel processes depending on how many segment instances it needs to recover.

- **-d master_data_directory**
  Optional. The master host data directory. If not specified, the value set for $MASTER_DATA_DIRECTORY will be used.

- **-F (full recovery)**
  Optional. Perform a full copy of the active segment instance in order to recover the failed segment. The default is to only copy over the incremental changes that occurred while the segment was down.
-i recover_config_file
Specifications the name of a configuration file that has information about which failed segments to recover, and where to recover them. Each line in the file is in the following format. Note that you only need to specify a name for filespaceOrder if your system has multiple filesystems configured. If your system does not have additional filesystems configured besides the default pg_system filesystem, this file will only have one location per segment (for the default data directory filesystem, pg_system). pg_system does not need to be listed in the filespaceOrder line. It will always be the first fileselocation listed after replication_port.

filespaceOrder=[filespace1_fileselocation[:filespace2_fileselocation:...]]
failed_host_address:port:fileselocation
[recovery_host_address:port:replication_port:fileselocation[:fileselocation: ...]]

You can use the -o option to output a sample recovery configuration file to use as a starting point.

-l logfile_directory
The directory to write the log file. Defaults to ~/.gpAdminLogs.

-o output_recover_config_file
Specifies a file name and location to output a sample recovery configuration file. The output file lists the currently invalid segments and their default recovery location in the format that is required by the -i option. Use together with the -p option to output a sample file for recovering on a different host. This file can be edited to supply alternate recovery locations if needed.

-p new_recover_host
Specifies a spare hostname outside of the currently configured Greenplum Database array on which to recover invalid segments. This spare host must have the Greenplum Database software installed and configured, and have the same hardware and OS configuration as the current segment hosts (same OS version, locales, gpadmin user account, data directory locations created, ssh keys exchanged, etc.).

-q (no screen output)
Run in quiet mode. Command output is not displayed on the screen, but is still written to the log file.

-s filespace_config_file
Specifies the name of a configuration file that contains file system locations on the currently configured segment hosts where you can recover failed segment instances. The filesystem configuration file is in the format of:

pg_system=default_fselocation
filespace1_name=filespace1_fselocation
filespace2_name=filespace2_fselocation
...
If your system does not have additional filespaces configured, this file will only have one location (for the default filespace, `pg_system`). These file system locations must exist on all segment hosts in the array and have sufficient disk space to accommodate recovered segments.

`-S output_filespace_config_file`  
Specifies a file name and location to output a sample filespace configuration file in the format that is required by the `-s` option. This file should be edited to supply the correct alternate filespace locations.

`-v (verbose)`  
Sets logging output to verbose.

`--version (version)`  
Displays the version of this utility.

`-? (help)`  
Displays the online help.

**Examples**

Recover any failed segment instances in place:

```
gprecoverseg
```

Recover any failed segment instances to a newly configured spare segment host:

```
gprecoverseg -i recover_config_file
```

Output the default recovery configuration file:

```
gprecoverseg -o /home/gpadmin/recover_config_file
```

**See Also**

`gpstart`, `gpstop`
gpscp

Copies files between multiple hosts at once.

**Synopsis**

```
gpscp { -f hostfile_gpssh | -h hostname [-h hostname ...] } [-J character] [-v] [[user@]hostname:]file_to_copy [...] [[user@]hostname:]copy_to_path
```

```
gpscp -?
gpscp --version
```

**Description**

The `gpscp` utility allows you to copy one or more files from the specified hosts to other specified hosts in one command using SCP (secure copy). For example, you can copy a file from the Greenplum Database master host to all of the segment hosts at the same time.

To specify the hosts involved in the SCP session, use the `-f` option to specify a file containing a list of host names, or use the `-h` option to name single host names on the command-line. At least one host name (`-h`) or a host file (`-f`) is required. The `-J` option allows you to specify a single character to substitute for the `hostname` in the copy from and to destination strings. If `-J` is not specified, the default substitution character is an equal sign (`=`). For example, the following command will copy `.bashrc` from the local host to `/home/gpadmin` on all hosts named in `hostfile_gpssh`:

```
gpscp -f hostfile_gpssh .bashrc =:/home/gpadmin
```

If a user name is not specified in the host list or with `user@` in the file path, `gpscp` will copy files as the currently logged in user. To determine the currently logged in user, do a `whoami` command. By default, `gpscp` goes to `$HOME` of the session user on the remote hosts after login. To ensure the file is copied to the correct location on the remote hosts, it is recommended that you use absolute paths.

Before using `gpscp`, you must have a trusted host setup between the hosts involved in the SCP session. You can use the utility `gpssh-exkeys` to update the known host files and exchange public keys between hosts if you have not done so already.

**Options**

- `-f hostfile_gpssh`
  
  Specifies the name of a file that contains a list of hosts that will participate in this SCP session. The host name is required, and you can optionally specify an alternate user name and/or ssh port number per host. The syntax of the host file is one host per line as follows:

  ```
  [username@]hostname[:ssh_port]
  ```
-h hostname
  Specifies a single host name that will participate in this SCP session. You can use
  the -h option multiple times to specify multiple host names.

-J character
  The -J option allows you to specify a single character to substitute for the hostname
  in the copy from and to destination strings. If -J is not specified, the default
  substitution character is an equal sign (=).

-v (verbose mode)
  Optional. Reports additional messages in addition to the SCP command output.

file_to_copy
  Required. The file name (or absolute path) of a file that you want to copy to other
  hosts (or file locations). This can be either a file on the local host or on another
  named host.

copy_to_path
  Required. The path where you want the file(s) to be copied on the named hosts. If an
  absolute path is not used, the file will be copied relative to $HOME of the session user.
  You can also use the equal sign ‘=’ (or another character that you specify with the -J
  option) in place of a hostname. This will then substitute in each host name as
  specified in the supplied host file (-f) or with the -h option.

-? (help)
  Displays the online help.

--version
  Displays the version of this utility.

Examples
Copy the file named installer.tar to / on all the hosts in the file hostfile_gpssh.
  gscp -f hostfile_gpssh installer.tar =:/

Copy the file named myfuncs.so to the specified location on the hosts named sdwl and
  gscp -h sdwl -h sdw2 myfuncs.so \
  =:/usr/local/greenplum-db/lib
gpseginstall

Installs the Greenplum Database software on multiple hosts.

Synopsis

gpseginstall -f hostfile_exkeys [-u gpdb_admin_user] [-p password] [-c u|p|c|s|E|e|l|v]
gpseginstall --help

Description

The gpseginstall utility allows system administrators to install the Greenplum
Database software on multiple hosts at once. When run as root, it also automates
other system configuration tasks such as creating the Greenplum system user
(gpadmin), setting the system user’s password, setting the ownership of the
Greenplum Database installation directory, and exchanging ssh keys between all
specified host address names.

When run as a non-root user, the gpseginstall utility will only install the
Greenplum Database software (provided the user has the appropriate permissions to
the install location), exchange ssh keys as the current user, and verify that the active
version of the Greenplum Database software is the same on all hosts.

To run this utility, the master host must have the Greenplum Database software
installed and the $GPHOME environment variable must be set.

Options

-c | --commands option_list

Optional. Controls which gpseginstall actions are executed. If run as root, all
actions will be executed by default. If run as a regular system user, only s e v
actions will be executed (provided the user has write permissions to the installation
location on all hosts).

- u: Add the Greenplum system user, gpadmin by default.
- p: Set the password for the specified system user, changeme by default.
- s: Compress, copy, and decompress (install) Greenplum Database on all
  specified hosts.
- c: Change the ownership of the Greenplum Database installation directory to
  the specified system user.
- E: Exchange ssh keys as root between all specified hosts.
- e: Exchange ssh keys as the specified system user (gpadmin) between all
  specified hosts.
- l: (Linux only) Check and modify the user limits configuration file
  (/etc/security/limits.conf) when adding a new system user.
• **v**: Verify the active version of the Greenplum Database software on all hosts by checking the `$GPHOME` environment variable and `greenplum-db` symbolic link to make sure that all hosts are referencing the same installation directory.

**-f | --file hostfile_exkeys**
Required. The name of file containing the machine configured host names and host addresses (interface names) for each host in your Greenplum system (master, standby master and segments). Make sure there are no blank lines or extra spaces. For example, if you have a master, standby master and three segments with two network interfaces per host, your file would look something like this:

```
mdw
mdw-1
mdw-2
smdw
smdw-1
smdw-2
sdw1
sdw1-1
sdw1-2
sdw2
sdw2-1
sdw2-2
sdw3
sdw3-1
sdw3-2
```

**-p | --password password**
Optional. Sets the password for the system user specified with the `-u` option. The default password is `changeme`.

**-u | --user user**
Optional. Creates the specified system user on all hosts. This is the user that owns your Greenplum Database installation and is used for all database administration tasks. The default is `gpadmin`.

**--help (help)**
Displays the online help.

**Examples**
As root, install a Greenplum Database on all segments, leave the system user as the default (`gpadmin`) and set the `gpadmin` password to `secret123`:

```
# gpseginstall -f hostfile_exkeys -p secret123
```
As a non-root user, install the Greenplum Database binaries on all hosts and exchange ssh keys as the current user.
$ gpseginstall -f hostfile_exkeys

See Also

gpinit system, gpssh-exkeys
**gpsetupSANfailover**

Specify failover host/mountpoints for Greenplum Database in a SAN environment.

---

**Synopsis**

gpsetupSANfailover <output_san_sql_file>


gpsetupSANfailover --version

gpsetupSANfailover -?

---

**Description**

This command creates a file that contains SQL code that is used for setting up SAN/failover-specific catalog entries in the master database. It prompts you to specify the failover mirror host/mountpoints for each primary host/mountpoint in the system. `gpsetupSANfailover` then verifies that the mountpoint was properly exported to the mirror failover host and that the failover mountpoint exists.

It then identifies the host’s identifiers on the SAN and the storage group name for each primary mountpoint. If there are no errors, `gpsetupSANfailover` generates the SQL output file. This file is used for setting up SAN/failover-specific catalog entries in the master database.

When the master host is running, enter:

```
$ PGOPTIONS='-c gp_session_role=utility' psql template1 \\
output_san_sql_file
```

This adds SAN/failover-specific catalog entries in the following master database tables:

- `gp_segment_configuration`
- `gp_fault_strategy`
- `gp_san_configuration`
- `gp_filespace_entry`

---

**Options**

- `-d master_data_directory`
  The master data directory. If not specified, the value set for `$MASTER_DATA_DIRECTORY` is used.

- `-l logfile_directory`
  The directory to write the log file. Defaults to `~/gpAdminLogs`.
-p port_offset
Optional. This number is used to calculate the database ports and failover ports used for mirror segments. The default offset is 1000. Mirror port assignments are calculated as follows:

- primary port + offset = mirror database port
- primary port + (2 * offset) = mirror failover port
- primary port + (3 * offset) = primary failover port

For example, if a primary segment has port 50001, then its mirror will use (listen on) database port of 51001, a mirror failover port of 52001, and a primary failover port of 53001 by default.

-D (debug)
Sets logging output to debug level.

--version (show utility version)
Displays the version of this utility.

-? (help)
Displays the online help.

Examples
Generate a SQL output file that you can use to add SAN/failover-specific catalog entries in the master database tables:

    gpsetupsanfailover output_san_sql_file

Generate a SQL output file that includes port offset of 2000.

    gpsetupsanfailover output_san_sql_file -p 2000

See Also
**gpsnmpd**

Reports on the health and state of a Greenplum Database system through SNMP.

**Synopsis**

```
gpsnmpd -s -C connect_string [-b] [-g] [-m MIB:...]
    [-M directory:...]
gpsnmpd -c FILE -C connect_string [-x address:port] [-b] [-g]
    [-m MIB:...] [-M directory:...]
gpsnmpd -?
gpsnmpd --version
```

**Description**

Greenplum’s `gpsnmpd` agent is an SNMP (Simple Network Management Protocol) daemon that support SNMP requests on the health and state of a Greenplum Database system by using a set of MIBs (Management Information Bases).

MIBs are a collection of objects that describe an SNMP-manageable entity — in this case, a Greenplum Database system. An *Agent* is any SNMP software running on a managed device that responds to queries or set requests. The `gpsnmpd` daemon currently supports the generic RDBMS MIB and typically operates on the master host.

`gpsnmpd` works in conjunction with the SNMP support that already exists on the Greenplum Database system. You can install and run `gpsnmpd` as a AgentX (Agent Extensibility Protocol) sub-agent to the operating system’s SNMP agent (usually called `snmpd`). This allows a Network Management System to get hardware and operating system information, as well as Greenplum Database information from the same port (161) and IP address. It also enables the auto-discovery of Greenplum Database instances.

However, if required, you can run the Greenplum SNMP agent as a stand-alone agent and use the `gpsnmpd` agent or SNMP notification features independently of each other. As a standalone SNMP agent, `gpsnmpd` listens (on a network socket) for SNMP queries, and requires the same extensive configuration as the system SNMP agent.

Greenplum recommends that you run `gpsnmpd` as a sub-agent to the system agent. When it starts, the `gpsnmpd` sub-agent registers itself with the system-level SNMP agent, and communicates to the system agent the parts of the MIB of which it is aware. The system agent communicates with the SNMP client/network monitoring application and forwards requests for particular sections of the MIB to the `gpsnmpd` sub-agent.

To get information about a Greenplum Database instance, `gpsnmpd` logs into Greenplum Database through a normal libpq client session. SNMP GetRequests trigger `SELECT` statements against Greenplum Database to get the requested information. It is important to note that `gpsnmpd` does not send SNMP trap notifications; Greenplum Database itself sends this information to the network monitor application.
The gpsnmpd sub-agent communicates with the system agent through UNIX sockets; it does not listen on network sockets when used as a sub-agent.

**Options**

*-s (sub-agent)*

Run gpsnmpd as an AgentX sub-agent to the system snmpd process. You do not need to use the -x option when using this option.

*-b (background)*

Run gpsnmpd as a background process.

*-c (configuration file)*

Specify the SNMP configuration file to use when starting gpsnmpd as a stand-alone agent. Note that you can specify any configuration file to run gpsnmpd as a stand-alone agent; you do not have to use the /etc/snmp/snmpd.conf file (/etc/sma/snmp/ on Solaris platforms). The configuration file you use must include a value for rocommunity.

*-g (use syslog)*

Logs gpsnmpd error messages to syslog.

*-C (libpq connection string)*

The libpq connection string to connect to Greenplum Database. Note that you can run gpsnmpd from a remote system. Depending on your network configuration, the gpsnmpd agent can establish a connection and monitor a remote Greenplum Database database instance. The configuration string can contain the database name, the port number, the username, the password, and other information if required. Greenplum recommends using the postgres database in the connection string (dbname=postgres).

You do not need to specify the -C option if you create a database role (user id) called root, and add the following line in the pg_hba.conf file:

```
local postgres root ident
```

This allows the UNIX user root to connect to the postgres database over the local connection. The root user does not require special permissions. The user and password parameters are only required when starting gpsnmpd as a user other than root.

*-x (address:port of a network interface)*

Specify an IP address for a network interface card on the host system, and specify a port other than the default SNMP port of 161. This enables you to run gpsnmpd without root permissions (you must have root permissions to use ports 1024 and lower).

You do not need to specify this option if you are running gpsnmpd as an AgentX sub-agent (-s).
-m (MIB:...)  
Loads one or more MIBs when starting gpsnmpd. Use a colon (:) to separate the MIBs. Enter ALL to load all MIBs. If you do not enter -m in the gpsnmpd command, a default set of MIBs are loaded by default.

-M (directory:...)  
Loads all MIBs from one or more directories when starting gpsnmpd. Use a colon (:) to separate the each directory. Enter the full path to each directory you specify for this option. If you do not enter -M in the gpsnmpd command, a default set of MIBs are loaded by default.

-? (help)  
Displays the online help.

-V  
Displays the version of this utility.

---

**Examples**

Start gpsnmpd as an agentx subagent:

```
# gpsnmpd -s -b -m ALL -C "dbname=postgres user=gpadmin \ password=secret"
```

Start gpsnmpd as a stand-alone agent:

```
# gpsnmpd -b -c /etc/snmp/snmpd.conf -x \ 192.168.100.12:10161 -M /usr/mibs/mymibs -C \ "dbname=postgres user=gpadmin password=secret"
```
gpssh

Provides ssh access to multiple hosts at once.

Synopsis

```bash
gpssh { -f hostfile_gpssh | - h hostname [-h hostname ...] } [-v]
[-e] [bash_command]
gpssh -?
gpssh --version
```

Description

The `gpssh` utility allows you to run bash shell commands on multiple hosts at once using SSH (secure shell). You can execute a single command by specifying it on the command-line, or omit the command to enter into an interactive command-line session.

To specify the hosts involved in the SSH session, use the `-f` option to specify a file containing a list of host names, or use the `-h` option to name single host names on the command-line. At least one host name (`-h`) or a host file (`-f`) is required. Note that the current host is not included in the session by default — to include the local host, you must explicitly declare it in the list of hosts involved in the session.

Before using `gpssh`, you must have a trusted host setup between the hosts involved in the SSH session. You can use the utility `gpssh-exkeys` to update the known host files and exchange public keys between hosts if you have not done so already.

If you do not specify a command on the command-line, `gpssh` will go into interactive mode. At the `gpssh` command prompt (`=>`), you can enter a command as you would in a regular bash terminal command-line, and the command will be executed on all hosts involved in the session. To end an interactive session, press `CTRL+D` on the keyboard or type `exit` or `quit`.

If a user name is not specified in the host file, `gpssh` will execute commands as the currently logged in user. To determine the currently logged in user, do a `whoami` command. By default, `gpssh` goes to `$HOME` of the session user on the remote hosts after login. To ensure commands are executed correctly on all remote hosts, you should always enter absolute paths.

Options

`bash_command`

A bash shell command to execute on all hosts involved in this session (optionally enclosed in quotes). If not specified, `gpssh` will start an interactive session.

`-e (echo)`

Optional. Echoes the commands passed to each host and their resulting output while running in non-interactive mode.
-f hostfile_gpssh
Specifies the name of a file that contains a list of hosts that will participate in this SSH session. The host name is required, and you can optionally specify an alternate user name and/or SSH port number per host. The syntax of the host file is one host per line as follows:

[username@]hostname[:ssh_port]

-h hostname
Specifies a single host name that will participate in this SSH session. You can use the -h option multiple times to specify multiple host names.

-v (verbose mode)
Optional. Reports additional messages in addition to the command output when running in non-interactive mode.

--version
Displays the version of this utility.

-? (help)
Displays the online help.

Examples
Start an interactive group SSH session with all hosts listed in the file hostfile_gpssh:

$ gpssh -f hostfile_gpssh

At the gpssh interactive command prompt, run a shell command on all the hosts involved in this session.

=> ls -a /data/primary/*

Exit an interactive session:

=> exit
=> quit

Start a non-interactive group SSH session with the hosts named dw1 and dw2 and pass a file containing several commands named command_file to gpssh:

$ gpssh -h sdw1 -h sdw2 -v -e < command_file

Execute single commands in non-interactive mode on hosts sdw2 and localhost:

$ gpssh -h sdw2 -h localhost -v -e 'ls -a /data/primary/*'
$ gpssh -h sdw2 -h localhost -v -e 'echo $GPHOME'
$ gpssh -h sdw2 -h localhost -v -e 'ls -l | wc -l'
gpssh-exkeys

Exchanges SSH public keys between hosts.

Synopsis

gpssh-exkeys -f hostfile_exkeys | - h hostname [-h hostname ...]
gpssh-exkeys -e hostfile_exkeys -x hostfile_gpexpand

gpssh-exkeys -?
gpssh-exkeys --version

Description

The gpssh-exkeys utility exchanges SSH keys between the specified host names (or host addresses). This allows SSH connections between Greenplum hosts and network interfaces without a password prompt. The utility is used to initially prepare a Greenplum Database system for password-free SSH access, and also to add additional ssh keys when expanding a Greenplum Database system.

To specify the hosts involved in an initial SSH key exchange, use the -f option to specify a file containing a list of host names (recommended), or use the -h option to name single host names on the command-line. At least one host name (-h) or a host file is required. Note that the local host is included in the key exchange by default.

To specify new expansion hosts to be added to an existing Greenplum Database system, use the -e and -x options. The -e option specifies a file containing a list of existing hosts in the system that already have SSH keys. The -x option specifies a file containing a list of new hosts that need to participate in the SSH key exchange.

Keys are exchanged as the currently logged in user. Greenplum recommends performing the key exchange process twice: once as root and once as the gpadmin user (the user designated to own your Greenplum Database installation). The Greenplum Database management utilities require that the same non-root user be created on all hosts in the Greenplum Database system, and the utilities must be able to connect as that user to all hosts without a password prompt.

The gpssh-exkeys utility performs key exchange using the following steps:

- Creates an RSA identification key pair for the current user if one does not already exist. The public key of this pair is added to the authorized_keys file of the current user.
- Updates the known_hosts file of the current user with the host key of each host specified using the -h, -f, -e, and -x options.
- Connects to each host using ssh and obtains the authorized_keys, known_hosts, and id_rsa.pub files to set up password-free access.
- Adds keys from the id_rsa.pub files obtained from each host to the authorized_keys file of the current user.
- Updates the authorized_keys, known_hosts, and id_rsa.pub files on all hosts with new host information (if any).
Options

\texttt{-e hostfile\_exkeys}
When doing a system expansion, this is the name and location of a file containing all configured host names and host addresses (interface names) for each host in your current Greenplum system (master, standby master and segments), one name per line without blank lines or extra spaces. Hosts specified in this file cannot be specified in the host file used with \texttt{-x}.

\texttt{-f hostfile\_exkeys}
Specifies the name and location of a file containing all configured host names and host addresses (interface names) for each host in your Greenplum system (master, standby master and segments), one name per line without blank lines or extra spaces.

\texttt{-h hostname}
Specifies a single host name (or host address) that will participate in the SSH key exchange. You can use the \texttt{-h} option multiple times to specify multiple host names and host addresses.

\texttt{--version}
Displays the version of this utility.

\texttt{-x hostfile\_gpexpand}
When doing a system expansion, this is the name and location of a file containing all configured host names and host addresses (interface names) for each new segment host you are adding to your Greenplum system, one name per line without blank lines or extra spaces. Hosts specified in this file cannot be specified in the host file used with \texttt{-e}.

\texttt{-? (help)}
Displays the online help.

Examples
Exchange SSH keys between all host names and addresses listed in the file \textit{hostfile\_exkeys}:

\begin{verbatim}
$ gpssh-exkeys -f hostfile_exkeys
\end{verbatim}

Exchange SSH keys between the hosts \texttt{sdw1}, \texttt{sdw2}, and \texttt{sdw3}:

\begin{verbatim}
$ gpssh-exkeys -h sdw1 -h sdw2 -h sdw3
\end{verbatim}

Exchange SSH keys between existing hosts \texttt{sdw1}, \texttt{sdw2} and \texttt{sdw3}, and new hosts \texttt{sdw4} and \texttt{sdw5} as part of a system expansion operation:

\begin{verbatim}
$ cat hostfile_exkeys
mdw
mdw-1
\end{verbatim}
mdw-2
smdw
smdw-1
smdw-2
sdw1
sdw1-1
sdw1-2
sdw2
sdw2-1
sdw2-2
sdw3
sdw3-1
sdw3-2

$ cat hostfile_gpexpand
  sdw4
  sdw4-1
  sdw4-2
  sdw5
  sdw5-1
  sdw5-2

$ gpssh-exkeys -e hostfile_exkeys -x hostfile_gpexpand

See Also

gpssh, gpscp
gpstart

Starts a Greenplum Database system.

Synopsis

gpstart [-d master_data_directory] [-B parallel_processes] [-R]
[-m] [-y] [-a] [-t timeout_seconds] [-l logfile_directory] [-v | -q]
gpstart -? | -h | --help

gpstart --version

Description

The gpstart utility is used to start the Greenplum Database server processes. When you start a Greenplum Database system, you are actually starting several postgres database server listener processes at once (the master and all of the segment instances). The gpstart utility handles the startup of the individual instances. Each instance is started in parallel.

The first time an administrator runs gpstart, the utility creates a hosts cache file named .gphostcache in the user’s home directory. Subsequently, the utility uses this list of hosts to start the system more efficiently. If new hosts are added to the system, you must manually remove this file from the gpadmin user’s home directory. The utility will create a new hosts cache file at the next startup.

Before you can start a Greenplum Database system, you must have initialized the system using gpinit系统 first.

Options

-a (do not prompt)

Do not prompt the user for confirmation.

-B parallel_processes

The number of segments to start in parallel. If not specified, the utility will start up to 60 parallel processes depending on how many segment instances it needs to start.

-d master_data_directory

Optional. The master host data directory. If not specified, the value set for $MASTER_DATA_DIRECTORY will be used.

-l logfile_directory

The directory to write the log file. Defaults to ~/gpAdminLogs.
-m (master only)
  Optional. Starts the master instance only, which may be useful for maintenance
tasks. This mode only allows connections to the master in utility mode. For
example:

    PGOPTIONS='-c gp_session_role=utility' psql

-q (no screen output)
  Run in quiet mode. Command output is not displayed on the screen, but is still
  written to the log file.

-R (restricted mode)
  Starts Greenplum Database in restricted mode (only database superusers are allowed
to connect).

-t timeout_seconds
  Specifies a timeout in seconds to wait for a segment instance to start up. If a segment
  instance was shutdown abnormally (due to power failure or killing its postgres
database listener process, for example), it may take longer to start up due to the
database recovery and validation process. If not specified, the default timeout is 60
seconds.

-v (verbose output)
  Displays detailed status, progress and error messages output by the utility.

-y (do not start standby master)
  Optional. Do not start the standby master host. The default is to start the standby
  master host and synchronization process.

-? | -h | --help (help)
  Displays the online help.

--version (show utility version)
  Displays the version of this utility.
**Examples**

Start a Greenplum Database system:

```
   gpstart
```

Start a Greenplum Database system in restricted mode (only allow superuser connections):

```
   gpstart -R
```

Start the Greenplum master instance only and connect in utility mode:

```
   gpstart -m
   POPTIONS='"-c gp_session_role=utility"' psql
```

Display the online help for the `gpstart` utility:

```
   gpstart -?
```

**See Also**

`gpstop`
gpstate

Shows the status of a running Greenplum Database system.

Synopsis

gpstate [-d master_data_directory] [-B parallel_processes]
[-s | -b | -Q | -e] [-m | -c] [-p | -i] [-f] [-v | -q]
[-l log_directory]
gpstate -? | -h | --help

Description

The gpstate utility displays information about a running Greenplum Database instance. There is additional information you may want to know about a Greenplum Database system, since it is comprised of multiple PostgreSQL database instances (segments) spanning multiple machines. The gpstate utility provides additional status information for a Greenplum Database system, such as:

- Which segments are down.
- Master and segment configuration information (hosts, data directories, etc.).
- The ports used by the system.
- A mapping of primary segments to their corresponding mirror segments.

Options

-b (brief status)
Optional. Display a brief summary of the state of the Greenplum Database system. This is the default option.

-B parallel_processes
The number of segments to check in parallel. If not specified, the utility will start up to 60 parallel processes depending on how many segment instances it needs to check.

-c (show primary to mirror mappings)
Optional. Display mapping of primary segments to their corresponding mirror segments.

-d master_data_directory
Optional. The master data directory. If not specified, the value set for $MASTER_DATA_DIRECTORY will be used.

-e (show segments with mirror status issues)
Show details on primary/mirror segment pairs that have potential issues such as 1) the active segment is running in change tracking mode, meaning a segment is down 2) the active segment is in resynchronization mode, meaning it is catching up
changes to the mirror 3) a segment is not in its preferred role, for example a segment that was a primary at system initialization time is now acting as a mirror, meaning you may have one or more segment hosts with unbalanced processing load.

- **f** *(show standby master details)*
  Display details of the standby master host if configured.

- **i** *(show Greenplum Database version)*
  Display the Greenplum Database software version information for each instance.

- **l logfile_directory**
  The directory to write the log file. Defaults to `~/gpAdminLogs`.

- **m** *(list mirrors)*
  Optional. List the mirror segment instances in the system, their current role, and synchronization status.

- **p** *(show ports)*
  List the port numbers used throughout the Greenplum Database system.

- **q** *(no screen output)*
  Optional. Run in quiet mode. Except for warning messages, command output is not displayed on the screen. However, this information is still written to the log file.

- **Q** *(quick status)*
  Optional. Checks segment status in the system catalog on the master host. Does not poll the segments for status.

- **s** *(detailed status)*
  Optional. Displays detailed status information for the Greenplum Database system.

- **v** *(verbose output)*
  Optional. Displays error messages and outputs detailed status and progress information.

- **? | -h | --help** *(help)*
  Displays the online help.

**Output Field Definitions**

The following output fields are reported by `gpstate -s` for the master:

<table>
<thead>
<tr>
<th>Output Data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master host</td>
<td>host name of the master</td>
</tr>
<tr>
<td>Master postgres process ID</td>
<td>PID of the master database listener process</td>
</tr>
</tbody>
</table>
The following output fields are reported by `gpstate -s` for each segment:

### Table B.2  gpstate output data for segments

<table>
<thead>
<tr>
<th>Output Data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hostname</td>
<td>system-configured host name</td>
</tr>
<tr>
<td>Address</td>
<td>network address host name (NIC name)</td>
</tr>
<tr>
<td>Datadir</td>
<td>file system location of segment data directory</td>
</tr>
<tr>
<td>Port</td>
<td>port number of segment postgres database listener process</td>
</tr>
<tr>
<td>Current Role</td>
<td>current role of a segment: <strong>Mirror or Primary</strong></td>
</tr>
<tr>
<td>Preferred Role</td>
<td>role at system initialization time: <strong>Mirror or Primary</strong></td>
</tr>
</tbody>
</table>
| Mirror Status                            | status of a primary/mirror segment pair:  
|                                          | *Synchronized* = data is up to date on both                                                           |
|                                          | *Resynchronization* = data is currently being copied from one to the other                             |
|                                          | *Change Tracking* = segment down and active segment is logging changes                                  |
| Change tracking data size                | when in *Change Tracking* mode, the size of the change log file (may grow and shrink as compression is applied) |
| Estimated total data to synchronize      | when in *Resynchronization* mode, the estimated size of data left to synchronize                       |
| Data synchronized                        | when in *Resynchronization* mode, the estimated size of data that has already been synchronized       |
Table B.2  gpstate output data for segments

<table>
<thead>
<tr>
<th>Output Data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated resync progress with mirror</td>
<td>When in Resynchronization mode, the estimated percentage of completion</td>
</tr>
<tr>
<td>Estimated resync end time</td>
<td>when in Resynchronization mode, the estimated time to complete</td>
</tr>
<tr>
<td>File postmaster.pid</td>
<td>status of postmaster.pid lock file: Found or Missing</td>
</tr>
<tr>
<td>PID from postmaster.pid file</td>
<td>PID found in the postmaster.pid file</td>
</tr>
<tr>
<td>Lock files in /tmp</td>
<td>a segment port lock file for its postgres process is created in /tmp (file is removed when a segment shuts down)</td>
</tr>
<tr>
<td>Active PID</td>
<td>active process ID of a segment</td>
</tr>
<tr>
<td>Master reports status as</td>
<td>segment status as reported in the system catalog: Up or Down</td>
</tr>
<tr>
<td>Database status</td>
<td>status of Greenplum Database to incoming requests: Up, Down, or Suspended. A Suspended state means database activity is temporarily paused while a segment transitions from one state to another.</td>
</tr>
</tbody>
</table>

Examples

Show detailed status information of a Greenplum Database system:

```
gpstate -s
```

Do a quick check for down segments in the master host system catalog:

```
gpstate -Q
```

Show information about mirror segment instances:

```
gpstate -m
```

Show information about the standby master configuration:

```
gpstate -f
```

Display the Greenplum software version information:

```
gpstate -i
```

See Also

gpstart, gplogfilter
gpstop

Stops or restarts a Greenplum Database system.

Synopsis

```
gpstop [-d master_data_directory] [-B parallel_processes] 
[-M smart | fast | immediate] [-t timeout_seconds] [-r] [-y] [-a] 
[-l logfile_directory] [-v | -q]
gpstop -m [-d master_data_directory] [-y] [-l logfile_directory]
[-v | -q]
gpstop -u [-d master_data_directory] [-l logfile_directory] [-v | 
-q]
gpstop --version

gpstop -? | -h | --help
```

Description

The gpstop utility is used to stop the database servers that comprise a Greenplum Database system. When you stop a Greenplum Database system, you are actually stopping several postgres database server processes at once (the master and all of the segment instances). The gpstop utility handles the shutdown of the individual instances. Each instance is shutdown in parallel.

By default, you are not allowed to shut down Greenplum Database if there are any client connections to the database. Use the -M fast option to roll back all in progress transactions and terminate any connections before shutting down. If there are any transactions in progress, the default behavior is to wait for them to commit before shutting down.

With the -u option, the utility uploads changes made to the master pg_hba.conf file or to runtime configuration parameters in the master postgresql.conf file without interruption of service. Note that any active sessions will not pickup the changes until they reconnect to the database.

Options

- **-a (do not prompt)**
  Do not prompt the user for confirmation.

- **-B parallel_processes**
  The number of segments to stop in parallel. If not specified, the utility will start up to 60 parallel processes depending on how many segment instances it needs to stop.

- **-d master_data_directory**
  Optional. The master host data directory. If not specified, the value set for $MASTER_DATA_DIRECTORY will be used.
-l logfile_directory
   The directory to write the log file. Defaults to ~/gpAdminLogs.

-m (master only)
   Optional. Shuts down a Greenplum master instance that was started in maintenance mode.

-M fast (fast shutdown - rollback)
   Fast shut down. Any transactions in progress are interrupted and rolled back.

-M immediate (immediate shutdown - abort)
   Immediate shut down. Any transactions in progress are aborted. This shutdown mode is not recommended. This mode kills all postgres processes without allowing the database server to complete transaction processing or clean up any temporary or in-process work files.

-M smart (smart shutdown - warn)
   Smart shut down. If there are active connections, this command fails with a warning. This is the default shutdown mode.

-q (no screen output)
   Run in quiet mode. Command output is not displayed on the screen, but is still written to the log file.

-r (restart)
   Restart after shutdown is complete.

-t timeout_seconds
   Specifies a timeout threshold (in seconds) to wait for a segment instance to shutdown. If a segment instance does not shutdown in the specified number of seconds, gpstop displays a message indicating that one or more segments are still in the process of shutting down and that you cannot restart Greenplum Database until the segment instance(s) are stopped. This option is useful in situations where gpstop is executed and there are very large transactions that need to rollback. These large transactions can take over a minute to rollback and surpass the default timeout period of 60 seconds.

-u (reload pg_hba.conf and postgresql.conf files only)
   This option reloads the pg_hba.conf files of the master and segments and the runtime parameters of the postgresql.conf files but does not shutdown the Greenplum Database array. Use this option to make new configuration settings active after editing postgresql.conf or pg_hba.conf. Note that this only applies to configuration parameters that are designated as runtime parameters.

-v (verbose output)
   Displays detailed status, progress and error messages output by the utility.
gpstop

```
--version (show utility version)
   Displays the version of this utility.

-y (do not stop standby master)
   Do not stop the standby master process. The default is to stop the standby master.

-? | -h | --help (help)
   Displays the online help.
```

**Examples**

Stop a Greenplum Database system in smart mode:
```
gpstop
```

Stop a Greenplum Database system in fast mode:
```
gpstop -M fast
```

Stop all segment instances and then restart the system:
```
gpstop -r
```

Stop a master instance that was started in maintenance mode:
```
gpstop -m
```

Reload the `postgresql.conf` and `pg_hba.conf` files after making configuration changes but do not shutdown the Greenplum Database array:
```
gpstop -u
```

**See Also**

gpstart
gpsys1

Displays information about your operating system.

Synopsis

    gpsys1 [ -a | -m | -p ]
    gpsys1 -? | --version

Description

gpsys1 displays the platform and installed memory (in bytes) of the current host. For example:

    linux 1073741824

Options

- **-a (show all)**
  
  Shows both platform and memory information for the current host. This is the default.

- **-m (show memory only)**
  
  Shows system memory installed in bytes.

- **-p (show platform only)**
  
  Shows the OS platform. Platform can be `linux`, `darwin` or `sunos5`.

- **-? (help)**
  
  Displays the online help.

- **--version**
  
  Displays the version of this utility.

Examples

Show information about the current host operating system:

    gpsys1

See Also

gpcheckperf
This appendix provides references for the command-line client utilities provided with Greenplum Database. Greenplum Database utilizes the standard PostgreSQL client programs, and also has additional management utilities to facilitate the administration of a distributed Greenplum Database DBMS.

The following Greenplum Database client programs are located in $GPHOME/bin:

- clusterdb
- createdb
- createlang
- createuser
- dropdb
- droplang
- dropuser
- ecpg
- pg_config
- pg_dump
- pg_dumpall
- pg_restore
- psql
- reindexdb
- vacuumdb
**Client Utility Summary**

- `host` | `--host host`
- `port` | `--port port`

### clusterdb

Reclusters tables that were previously clustered with `CLUSTER`.

```bash
clusterdb [connection-option...] [-v] [-t table ] [{-d} dbname]
```

```bash
clusterdb [connection-option...] [-a] [-v]
```

```bash
clusterdb --help | --version
```

- `-a` | `--all`
- `[-d] dbname` | `[--dbname] dbname`
- `-e` | `--echo`
- `-q` | `--quiet`
- `-t table` | `--table table`
- `-v` | `--verbose`
- `-h host` | `--host host`
- `-p port` | `--port port`
- `-U username` | `--username username`
- `-w` | `--no-password`
- `-W` | `--password`

### createdb

Creates a new database.

```bash
createdb [connection_option ...] [-D tablespace] [-E encoding] [-O owner] [-T template] [-e] [dbname ['description']]
```

```bash
createdb --help | --version
```

```bash
dbname
description
```

- `-D tablespace` | `--tablespace tablespace`
- `-e echo`
- `-E encoding` | `--encoding encoding`
- `-O owner` | `--owner owner`
- `-T template` | `--template template`
- `-h host` | `--host host`
- `-p port` | `--port port`
- `-U username` | `--username username`
- `-w` | `--no-password`
- `-W` | `--password`
createlang
Defines a new procedural language for a database.
createlang [connection_option ...] [-e] langname [(-d) dbname]
createlang [connection-option ...] -l dbname
createlang --help | --version

langname
[-d] dbname | [--dbname] dbname
-e | --echo
-l dbname | --list dbname
-h host | --host host
-p port | --port port
-U username | --username username
-w | --no-password
-W | --password

createuser
Creates a new database role.
createuser [connection_option ...] [role_attribute ...] [-e] role_name
createuser --help | --version

role_name
-c number | --connection-limit number
-D | --no-createdb
-d | --createdb
-e | --echo
-E | --encrypted
-i | --inherit
-I | --no-inherit
-l | --login
-L | --no-login
-N | --unencrypted
-P | --pwprompt
-r | --createrole
-R | --no-createrole
-s | --superuser
-S | --no-superuser
-h host | --host host
-p port | --port port
-U username | --username username
-w | --no-password
-W | --password
dropdb
Removes a database.

```
dropdb [connection_option ...] [-e] [-i] dbname
dropdb --help | --version
dbname
-e | --echo
-i | --interactive
-h host | --host host
-p port | --port port
-U username | --username username
-w | --no-password
-W | --password
```

droplang
Removes a procedural language.

```
droplang [connection-option ...] [-e] langname [[-d] dbname]
droplang [connection-option ...] [-e] -l dbname
droplang --help | --version
langname
[-d] dbname | [[-dbname] dbname
-e | --echo
-l | --list
-h host | --host host
-p port | --port port
-U username | --username username
-w | --no-password
-W | --password
```

dropuser
Removes a database role.

```
dropuser [connection_option ...] [-e] [-i] role_name
dropuser --help | --version
role_name
-e | --echo
-i | --interactive
-h host | --host host
-p port | --port port
-U username | --username username
-w | --no-password
-W | --password
```
**ecpg**

Embedded SQL C preprocessor.

```
ecpg [option ...] file ...
```

```
file
-c
-C mode
-D symbol
-i
-I directory
-o filename
-r option
-t
-v
--help
--version
```

**pg_config**

Retrieves information about the installed version of Greenplum Database.

```
pg_config [option ...]
```

```
--bindir
--docdir
--includedir
--pkgincludedir
--includedir-server
--libdir
--pkglibdir
--localedir
--mandir
--sharedir
--sysconfdir
--pgxs
--configure
--cc
--cppflags
--cflags
--cflags_sl
--ldflags
--ldflags_sl
--libs
--version
```
pg_dump

Extracts a database into a single script file or other archive file.

```
pg_dump [connection_option ...] [dump_option ...] dbname
dbname
  -a | --data-only
  -b | --blobs
  -c | --clean
  -C | --create
  -d | --inserts
  -D | --column-inserts | --attribute-inserts
  -E encoding | --encoding=encoding
  -f file | --file=file
  -F p|c|t | --format=plain|custom|tar
  -i | --ignore-version
  -n schema | --schema=schema
  -N schema | --exclude-schema=schema
  -o | --oids
  -O | --no-owner
  -s | --schema-only
  -S username | --superuser=username
  -t table | --table=table
  -T table | --exclude-table=table
  -v | --verbose
  -x | --no-privileges | --no-acl
  --disable-dollar-quoting
  --disable-triggers
  --use-set-session-authorization
  --gp-syntax | --no-gp-syntax
  -Z 0..9 | --compress=0..9
  -h host | --host host
  -p port | --port port
  -U username | --username username
  -W | --password
```
**pg_dumpall**

Extracts all databases in a Greenplum Database system to a single script file or other archive file.

```markdown
pg_dumpall [connection_option ...] [dump_option ...]
- a | --data-only
- c | --clean
- d | --inserts
- D | --column-inserts | --attribute-inserts
- f | --filesaces
- g | --globals-only
- i | --ignore-version
- o | --oids
- O | --no-owner
- r | --resource-queues
- s | --schema-only
- S username | --superuser=username
- v | --verbose
- x | --no-privileges | --no-acl
--disable-dollar-quoting
--disable-triggers
--use-set-session-authorization
--gp-syntax
- h host | --host host
- p port | --port port
- U username | --username username
- W | --password
```

pkg_restore

Restores a database from an archive file created by pkg_dump.

pkg_restore [connection_option ...] [restore_option ...] filename

filename
-a | --data-only
-c | --clean
-C | --create
-d dbname | --dbname=dbname
-e | --exit-on-error
-f outfile | --file=outfile
-F t|c | --format=tar|custom
-i | --ignore-version
-I index | --index=index
-1 | --list
-L list-file | --use-list=list-file
-n schema | --schema=schema
-O | --no-owner
-P 'function-name(argtype [, ...])' | --function='function-name(argtype [, ...])'
-s | --schema-only
-S username | --superuser=username
-t table | --table=table
-T trigger | --trigger=trigger
-v | --verbose
-x | --no-privileges | --no-acl
--disable-triggers
--no-data-for-failed-tables
-h host | --host host
-p port | --port port
-U username | --username username
-W | --password
-1 | --single-transaction
psql
Interactive command-line interface for Greenplum Database

psql [option...] [dbname [username]]
-a | --echo-all
-A | --no-align
-c 'command' | --command 'command'
-d dbname | --dbname dbname
-e | --echo-queries
-E | --echo-hidden
-f filename | --file filename
-F separator | --field-separator separator
-H | --html
-l | --list
-L filename | --log-file filename
-o filename | --output filename
-P assignment | --pset assignment
-q | --quiet
-R separator | --record-separator separator
-s | --single-step
-S | --single-line
-t | --tuples-only
-T table_options | --table-attr table_options
-v assignment | --set assignment | --variable assignment
-V | --version
-x | --expanded
-X | --no-psqlrc
-1 | --single-transaction
-? | --help
-h host | --host host
-p port | --port port
-U username | --username username
-W | --password

reindexdb
Rebuilds indexes in a database.

reindexdb [connection-option...] [-table | -t table] [-index | -i index] [db-
name]

reindexdb [connection-option...] [--all | -a]
reindexdb [connection-option...] [--system | -s] [dbname]
reindexdb --help | --version
-a | --all
[-d] dbname | [--dbname] dbname
-e | --echo
-i index | --index index
-q | --quiet
-s | --system
-t table | --table table
-h host | --host host
-p port | --port port
-U username | --username username
-w | --no-password
-W | --password

vacuumdb
Garbage-collects and analyzes a database.
vacuumdb --help | --version
-a | --all
[-d] dbname | [--dbname] dbname
-e | --echo
-f | --full
-F | --freeze
-q | --quiet
-t table [{(column)}] | --table table [{(column)}]
-v | --verbose
-z | --analyze
-h host | --host host
-p port | --port port
-U username | --username username
-w | --no-password
-W | --password


**clusterdb**

Reclusters tables that were previously clustered with CLUSTER.

---

**Synopsis**

```
clusterdb [connection-option...] [-v] [-t table] [ [-d] dbname]
clusterdb [connection-option...] [-a] [-v]
```

**Description**

To cluster a table means to physically reorder a table on disk according to an index so that index scan operations can access data on disk in a somewhat sequential order, thereby improving index seek performance for queries that use that index.

The `clusterdb` utility will find any tables in a database that have previously been clustered with the `CLUSTER` SQL command, and clusters them again on the same index that was last used. Tables that have never been clustered are not affected.

`clusterdb` is a wrapper around the SQL command `CLUSTER`. Although clustering a table in this way is supported in Greenplum Database, it is not recommended because the `CLUSTER` operation itself is extremely slow.

If you do need to order a table in this way to improve your query performance, Greenplum recommends using a `CREATE TABLE AS` statement to reorder the table on disk rather than using `CLUSTER`. If you do ‘cluster’ a table in this way, then `clusterdb` would not be relevant.

---

**Options**

- **-a | --all**
  
  Cluster all databases.

- **[-d] dbname | [--dbname] dbname**
  
  Specifies the name of the database to be clustered. If this is not specified, the database name is read from the environment variable `PGDATABASE`. If that is not set, the user name specified for the connection is used.

- **-e | --echo**
  
  Echo the commands that `clusterdb` generates and sends to the server.

- **-q | --quiet**
  
  Do not display a response.

- **-t table | --table table**
  
  Cluster the named table only.
-v | --verbose
Print detailed information during processing.

Connection Options

-h host | --host host
The host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable PGHOST or defaults to localhost.

-p port | --port port
The TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable PGPORT or defaults to 5432.

-U username | --username username
The database role name to connect as. If not specified, reads from the environment variable PGUSER or defaults to the current system role name.

-w | --no-password
Never issue a password prompt. If the server requires password authentication and a password is not available by other means such as a .pgpass file, the connection attempt will fail. This option can be useful in batch jobs and scripts where no user is present to enter a password.

-W | --password
Force a password prompt.

Examples
To cluster the database test:
    clusterdb test

To cluster a single table foo in a database named xyzzy:
    clusterdb --table foo xyzzyb

See Also
CLUSTER
**createdb**

Creates a new database.

### Synopsis

```
createdb [connection_option ...] [-D tablespace] [-E encoding]
[-O owner] [-T template] [-e] [dbname ['description']]  
createdb --help | --version
```

### Description

`createdb` creates a new database in a Greenplum Database system.

Normally, the database user who executes this command becomes the owner of the new database. However a different owner can be specified via the `-O` option, if the executing user has appropriate privileges.

`createdb` is a wrapper around the SQL command `CREATE DATABASE`.

### Options

**dbname**

The name of the database to be created. The name must be unique among all other databases in the Greenplum system. If not specified, reads from the environment variable `PGDATABASE`, then `PGUSER` or defaults to the current system user.

**description**

A comment to be associated with the newly created database. Descriptions containing white space must be enclosed in quotes.

**-D tablespace | --tablespace tablespace**

The default tablespace for the database.

**-e echo**

Echo the commands that `createdb` generates and sends to the server.

**-E encoding | --encoding encoding**

Character set encoding to use in the new database. Specify a string constant (such as `'UTF8'`), an integer encoding number, or `DEFAULT` to use the default encoding. See “Character Set Support” on page 939 for the list of supported character sets.

**-O owner | --owner owner**

The name of the database user who will own the new database. Defaults to the user executing this command.
-T template | --template template
The name of the template from which to create the new database. Defaults to template1.

Connection Options

-h host | --host host
The host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable PGHOST or defaults to localhost.

-p port | --port port
The TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable PGPORT or defaults to 5432.

-U username | --username username
The database role name to connect as. If not specified, reads from the environment variable PGUSER or defaults to the current system role name.

-w | --no-password
Never issue a password prompt. If the server requires password authentication and a password is not available by other means such as a .pgpass file, the connection attempt will fail. This option can be useful in batch jobs and scripts where no user is present to enter a password.

-W | --password
Force a password prompt.

Examples
To create the database test using the default options:
```
createdb test
```

To create the database demo using the Greenplum master on host gpmaster, port 54321, using the LATIN1 encoding scheme:
```
createdb -p 54321 -h gpmaster -E LATIN1 demo
```

See Also
CREATE DATABASE
**createlang**

Defines a new procedural language for a database.

**Synopsis**

```
createlang [connection_option ...] [-e] langname [[-d]dbname]
createlang [connection-option ...] -l dbname
createlang --help | --version
```

**Description**

`createlang` is a utility for adding a new programming language to a database. `createlang` is a wrapper around the `CREATE LANGUAGE` SQL command.

There are currently three procedural language packages included in the standard Greenplum Database distribution: PL/pgSQL, PL/Perl and PL/Python. The PL/pgSQL language is already registered in all databases by default.

A language handler has also been added for PL/R, but the PL/R language package is not pre-installed with Greenplum Database. There is also a package available for PL/Tcl, which can be enabled if needed. See the section on Procedural Languages in the PostgreSQL documentation for more information.

**Options**

- **langname**
  
  Specifies the name of the procedural programming language to be defined.

- **[-d] dbname | [--dbname] dbname**
  
  Specifies to which database the language should be added. The default is to use the `PGDATABASE` environment variable setting, or the same name as the current system user.

- **-e | --echo**
  
  Echo the commands that `createlang` generates and sends to the server.

- **-l dbname | --list dbname**
  
  Show a list of already installed languages in the target database.

**Connection Options**

- **-h host | --host host**
  
  The host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable `PGHOST` or defaults to localhost.
-p port | --port port

The TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable PGPORT or defaults to 5432.

-U username | --username username

The database role name to connect as. If not specified, reads from the environment variable PGUSER or defaults to the current system role name.

-w | --no-password

Never issue a password prompt. If the server requires password authentication and a password is not available by other means such as a .pgpass file, the connection attempt will fail. This option can be useful in batch jobs and scripts where no user is present to enter a password.

-W | --password

Force a password prompt.

Examples
To install the language plperl into the database template1:

    createlang plperl template1

See Also
CREATE LANGUAGE, droplang
**createuser**

Creates a new database role.

**Synopsis**

```
createuser [connection_option ...] [role_attribute ...] [-e]
role_name

createuser --help | --version
```

**Description**

`createuser` creates a new Greenplum Database role. Only superusers and users with `CREATEROLE` privilege can create new roles, so `createuser` must be invoked by someone who can connect as a superuser or a role with `CREATEROLE` privilege.

If you wish to create a new superuser, you must connect as a superuser, not merely with `CREATEROLE` privilege. Being a superuser implies the ability to bypass all access permission checks within the database, so superuser privileges should not be granted lightly.

`createuser` is a wrapper around the SQL command `CREATE ROLE`.

**Options**

**role_name**

The name of the role to be created. This name must be different from all existing roles in this Greenplum Database installation.

- `-c number | --connection-limit number`
  
  Set a maximum number of connections for the new role. The default is to set no limit.

- `-D | --no-createdb`
  
  The new role will not be allowed to create databases. This is the default.

- `-d | --createdb`
  
  The new role will be allowed to create databases.

- `-e | --echo`
  
  Echo the commands that `createuser` generates and sends to the server.

- `-E | --encrypted`
  
  Encrypts the role’s password stored in the database. If not specified, the default password behavior is used.
-i | --inherit
   The new role will automatically inherit privileges of roles it is a member of. This is the default.

-I | --no-inherit
   The new role will not automatically inherit privileges of roles it is a member of.

-l | --login
   The new role will be allowed to log in to Greenplum Database. This is the default.

-L | --no-login
   The new role will not be allowed to log in (a group-level role).

-N | --unencrypted
   Does not encrypt the role’s password stored in the database. If not specified, the default password behavior is used.

-P | --pwprompt
   If given, createuser will issue a prompt for the password of the new role. This is not necessary if you do not plan on using password authentication.

-r | --createrole
   The new role will be allowed to create new roles (CREATEROLE privilege).

-R | --no-createrole
   The new role will not be allowed to create new roles. This is the default.

-s | --superuser
   The new role will be a superuser.

-S | --no-superuser
   The new role will not be a superuser. This is the default.

Connection Options

-h host | --host host
   The host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable PGHOST or defaults to localhost.

-p port | --port port
   The TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable PGPORT or defaults to 5432.
-U username | --username username
The database role name to connect as. If not specified, reads from the environment
variable PGUSER or defaults to the current system role name.

-w | --no-password
Never issue a password prompt. If the server requires password authentication and a
password is not available by other means such as a .pgpass file, the connection
attempt will fail. This option can be useful in batch jobs and scripts where no user is
present to enter a password.

-W | --password
Force a password prompt.

Examples
Create a role named joe using the default options:

```
createuser joe
```

Shall the new role be a superuser? (y/n) n
Shall the new role be allowed to create databases? (y/n) n
Shall the new role be allowed to create more new roles? (y/n) n

CREATE ROLE

To create the same role joe using connection options and avoiding the prompts and
taking a look at the underlying command:

```
createuser -h masterhost -p 54321 -S -D -R -e joe
```

CREATE ROLE joe NOSUPERUSER NOCREATEDB NOCREATEROLE INHERIT
LOGIN;
CREATE ROLE

To create the role joe as a superuser, and assign a password immediately:

```
createuser -P -s -e joe
```

Enter password for new role: admin123
Enter it again: admin123

CREATE ROLE joe PASSWORD 'admin123' SUPERUSER CREATEDB
CREATEROLE INHERIT LOGIN;
CREATE ROLE

In the above example, the new password is not actually echoed when typed, but we
show what was typed for clarity. However the password will appear in the echoed
command, as illustrated if the -e option is used.

See Also
CREATE ROLE
**Synopsis**

```bash
dropdb [connection_option ...] [-e] [-i] dbname
dropdb --help | --version
```

**Description**

`dropdb` destroys an existing database. The user who executes this command must be a superuser or the owner of the database being dropped.

`dropdb` is a wrapper around the SQL command `DROP DATABASE`.

**Options**

- **dbname**
  The name of the database to be removed.

- **-e** | **--echo**
  Echo the commands that `dropdb` generates and sends to the server.

- **-i** | **--interactive**
  Issues a verification prompt before doing anything destructive.

**Connection Options**

- **-h host** | **--host host**
  The host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable `PGHOST` or defaults to localhost.

- **-p port** | **--port port**
  The TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable `PGPORT` or defaults to 5432.

- **-U username** | **--username username**
  The database role name to connect as. If not specified, reads from the environment variable `PGUSER` or defaults to the current system role name.

- **-w** | **--no-password**
  Never issue a password prompt. If the server requires password authentication and a password is not available by other means such as a `.pgpass` file, the connection attempt will fail. This option can be useful in batch jobs and scripts where no user is present to enter a password.
-W | --password
Force a password prompt.

Examples
To destroy the database named demo using default connection parameters:

```
dropdb demo
```

To destroy the database named demo using connection options, with verification, and a peek at the underlying command:

```
dropdb -p 54321 -h masterhost -i -e demo
```

Database "demo" will be permanently deleted.

Are you sure? (y/n) y

DROP DATABASE "demo"
DROP DATABASE

See Also

DROP DATABASE
gp_db_interfaces

The gp_db_interfaces table contains information about the relationship of segments to network interfaces. This information, joined with data from gp_interfaces, is used by the system to optimize the usage of available network interfaces for various purposes, including fault detection.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbid</td>
<td>smallint</td>
<td>gp_segment_configuration.dbid</td>
<td>System-assigned ID. The unique identifier of a segment (or master) instance.</td>
</tr>
<tr>
<td>interfaceid</td>
<td>smallint</td>
<td>gp_interfaces.interfaceid</td>
<td>System-assigned ID for a network interface.</td>
</tr>
<tr>
<td>priority</td>
<td>smallint</td>
<td></td>
<td>Priority of the network interface for this segment.</td>
</tr>
</tbody>
</table>
**droplang**

Removes a procedural language.

**Synopsis**

droplang [connection-option ...] [-e] langname [(-d) dbname]
droplang [connection-option ...] [-e] -l dbname
droplang --help | --version

**Description**

droplang is a utility for removing an existing programming language from a database. droplang can drop any procedural language, even those not supplied by the Greenplum Database distribution.

Although programming languages can be removed directly using several SQL commands, it is recommended to use droplang because it performs a number of checks and is much easier to use.

droplang is a wrapper for the SQL command DROP LANGUAGE.

**Options**

*langname*

  Specifies the name of the programming language to be removed.

[-d] *dbname* | [--dbname] *dbname*

  Specifies from which database the language should be removed. The default is to use the PGDATABASE environment variable setting, or the same name as the current system user.

-e | --echo

  Echo the commands that droplang generates and sends to the server.

-l | --list

  Show a list of already installed languages in the target database.

**Connection Options**

-h *host* | --host *host*

  The host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable PGHOST or defaults to localhost.
-p port | --port port
The TCP port on which the Greenplum master database server is listening for
connections. If not specified, reads from the environment variable PGPORT or
defaults to 5432.

-U username | --username username
The database role name to connect as. If not specified, reads from the environment
variable PGUSER or defaults to the current system role name.

-w | --no-password
Never issue a password prompt. If the server requires password authentication and a
password is not available by other means such as a .pgpass file, the connection
attempt will fail. This option can be useful in batch jobs and scripts where no user is
present to enter a password.

-W | --password
Force a password prompt.

Examples
To remove the language pltcl:

droplang pltcl mydatabase

See Also
DROP LANGUAGE
**dropuser**

Removes a database role.

**Synopsis**

```
dropuser [connection_option ...] [-e] [-i] role_name
```

```
dropuser --help | --version
```

**Description**

dropuser removes an existing role from Greenplum Database. Only superusers and users with the `CREATEROLE` privilege can remove roles. To remove a superuser role, you must yourself be a superuser.

dropuser is a wrapper around the SQL command `DROP ROLE`.

**Options**

*role_name*

The name of the role to be removed. You will be prompted for a name if not specified on the command line.

*e | --echo*

Echo the commands that `dropuser` generates and sends to the server.

*i | --interactive*

Prompt for confirmation before actually removing the role.

**Connection Options**

*host | --host host*

The host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable `PGHOST` or defaults to localhost.

*port | --port port*

The TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable `PGPORT` or defaults to 5432.

*username | --username username*

The database role name to connect as. If not specified, reads from the environment variable `PGUSER` or defaults to the current system role name.
**-w | --no-password**

Never issue a password prompt. If the server requires password authentication and a password is not available by other means such as a .pgpass file, the connection attempt will fail. This option can be useful in batch jobs and scripts where no user is present to enter a password.

**-W | --password**

Force a password prompt.

---

**Examples**

To remove the role *joe* using default connection options:

```
dropuser joe
DROP ROLE
```

To remove the role *joe* using connection options, with verification, and a peek at the underlying command:

```
dropuser -p 54321 -h masterhost -i -e joe
Role "joe" will be permanently removed.
Are you sure? (y/n) y
DROP ROLE "joe"
DROP ROLE
```

---

**See Also**

DROP ROLE
**Synopsis**

epg [option ...] file ...

**Description**

epg is the embedded SQL preprocessor for C programs. It converts C programs with embedded SQL statements to normal C code by replacing the SQL invocations with special function calls. The output files can then be processed with any C compiler tool chain.

epg will convert each input file given on the command line to the corresponding C output file. Input files preferably have the extension .pgc, in which case the extension will be replaced by .c to determine the output file name. If the extension of the input file is not .pgc, then the output file name is computed by appending .c to the full file name. The output file name can also be overridden using the `-o` option.

This reference page does not describe the embedded SQL language. See the ECPG - Embedded SQL in C chapter of the PostgreSQL documentation for more information.

**Options**

`file`

The file to convert.

`-c`

Automatically generate certain C code from SQL code. Currently, this works for EXEC SQL TYPE.

`-C mode`

Set a compatibility mode. mode may be INFORMIX or INFORMIX_SE.

`-D symbol`

Define a C preprocessor symbol.

`-i`

Parse system include files as well.

`-I directory`

Specify an additional include path, used to find files included via EXEC SQL INCLUDE. Defaults are . (current directory), /usr/local/include, the Greenplum Database include directory (/usr/local/greenplum-db-4.0.x.x/include), and /usr/include, in that order.
-o filename
Specifies that `ecpg` should write all its output to the given filename.

-r option
Selects a run-time behavior. Currently, option can only be `no_indicator`.

-t
Turn on autocommit of transactions. In this mode, each SQL command is automatically committed unless it is inside an explicit transaction block. In the default mode, commands are committed only when `EXEC SQL COMMIT` is issued.

-v
Print additional information including the version and the include path.

--help
Show a brief summary of the command usage, then exit.

--version
Output version information, then exit.

**Examples**

If you have an embedded SQL C source file named `progl.pgc`, you can create an executable program using the following sequence of commands:

```
ecpg progl.pgc
cc -I/usr/local/pgsql/include -c progl.c
cc -o progl progl.o -L/usr/local/greenplum-db-4.0.x.x/lib
-lecpg
```
**pg_config**

Retrieves information about the installed version of Greenplum Database.

---

**Synopsis**

```
pg_config [option ...]
```

---

**Description**

The `pg_config` utility prints configuration parameters of the currently installed version of Greenplum Database. It is intended, for example, to be used by software packages that want to interface to Greenplum Database to facilitate finding the required header files and libraries. Note that information printed out by `pg_config` is for the Greenplum Database master only.

If more than one option is given, the information is printed in that order, one item per line. If no options are given, all available information is printed, with labels.

---

**Options**

- **--bindir**
  
  Print the location of user executables. Use this, for example, to find the `psql` program. This is normally also the location where the `pg_config` program resides.

- **--docdir**
  
  Print the location of documentation files.

- **--includedir**
  
  Print the location of C header files of the client interfaces.

- **--pkgincludedir**
  
  Print the location of other C header files.

- **--includedir-server**
  
  Print the location of C header files for server programming.

- **--libdir**
  
  Print the location of object code libraries.

- **--pkglibdir**
  
  Print the location of dynamically loadable modules, or where the server would search for them. (Other architecture-dependent data files may also be installed in this directory.)

- **--localedir**
  
  Print the location of locale support files.
--mandir
Print the location of manual pages.

--sharedir
Print the location of architecture-independent support files.

--sysconfdir
Print the location of system-wide configuration files.

--pgxs
Print the location of extension makefiles.

--configure
Print the options that were given to the configure script when Greenplum Database was configured for building.

--cc
Print the value of the CC variable that was used for building Greenplum Database. This shows the C compiler used.

--cppflags
Print the value of the CPPFLAGS variable that was used for building Greenplum Database. This shows C compiler switches needed at preprocessing time.

--cflags
Print the value of the CFLAGS variable that was used for building Greenplum Database. This shows C compiler switches.

--cflags_sl
Print the value of the CFLAGS_SL variable that was used for building Greenplum Database. This shows extra C compiler switches used for building shared libraries.

--ldflags
Print the value of the LDFLAGS variable that was used for building Greenplum Database. This shows linker switches.

--ldflags_sl
Print the value of the LDFLAGS_SL variable that was used for building Greenplum Database. This shows linker switches used for building shared libraries.

--libs
Print the value of the LIBS variable that was used for building Greenplum Database. This normally contains -l switches for external libraries linked into Greenplum Database.

--version
Print the version of Greenplum Database.
Examples

To reproduce the build configuration of the current Greenplum Database installation, run the following command:

```bash
eval ./configure 'pg_config --configure'
```

The output of `pg_config --configure` contains shell quotation marks so arguments with spaces are represented correctly. Therefore, using `eval` is required for proper results.
**pg_dump**

Extracts a database into a single script file or other archive file.

### Synopsis

```
pg_dump [connection_option ...] [dump_option ...] dbname
```

### Description

**pg_dump** is a standard PostgreSQL utility for backing up a database, and is also supported in Greenplum Database. It creates a single (non-parallel) dump file. For routine backups of Greenplum Database it is better to use Greenplum’s parallel dump utility, **gp_dump**, for the best performance.

Use **pg_dump** if you are migrating your data to another database vendor’s system, or to another Greenplum Database system with a different segment configuration (for example, if the system you are migrating to has greater or fewer segment instances). To restore, you must use the corresponding **pg_restore** utility (if the dump file is in archive format), or you can use a client program such as **psql** (if the dump file is in plain text format).

Since **pg_dump** is compatible with regular PostgreSQL, it can be used to migrate data into Greenplum Database. The **pg_dump** utility in Greenplum Database is very similar to the PostgreSQL **pg_dump** utility, with the following exceptions and limitations:

- If using **pg_dump** to backup a Greenplum database, keep in mind that the dump operation can take a long time (several hours) for very large databases. Also, you must make sure you have sufficient disk space to create the dump file.

- If you are migrating data from one Greenplum Database system to another, use the **--gp-syntax** command-line option to include the DISTRIBUTED BY clause in CREATE TABLE statements. This ensures that Greenplum Database table data is distributed with the correct distribution key columns upon restore.

**pg_dump** makes consistent backups even if the database is being used concurrently. **pg_dump** does not block other users accessing the database (readers or writers).

When used with one of the archive file formats and combined with **pg_restore**, **pg_dump** provides a flexible archival and transfer mechanism. **pg_dump** can be used to backup an entire database, then **pg_restore** can be used to examine the archive and/or select which parts of the database are to be restored. The most flexible output file format is the **custom** format (`-Fc`). It allows for selection and reordering of all archived items, and is compressed by default. The tar format (`-Ft`) is not compressed and it is not possible to reorder data when loading, but it is otherwise quite flexible. It can be manipulated with standard UNIX tools such as **tar**.
Options

dbname

Specifies the name of the database to be dumped. If this is not specified, the environment variable PGDATABASE is used. If that is not set, the user name specified for the connection is used.

Dump Options

-a | --data-only

Dump only the data, not the schema (data definitions). This option is only meaningful for the plain-text format. For the archive formats, you may specify the option when you call pg_restore.

-b | --blobs

Include large objects in the dump. This is the default behavior except when --schema, --table, or --schema-only is specified, so the -b switch is only useful to add large objects to selective dumps.

-c | --clean

Adds commands to the text output file to clean (drop) database objects prior to (the commands for) creating them. Note that objects are not dropped before the dump operation begins, but DROP commands are added to the DDL dump output files so that when you use those files to do a restore, the DROP commands are run prior to the CREATE commands. This option is only meaningful for the plain-text format. For the archive formats, you may specify the option when you call pg_restore.

-C | --create

Begin the output with a command to create the database itself and reconnect to the created database. (With a script of this form, it doesn’t matter which database you connect to before running the script.) This option is only meaningful for the plain-text format. For the archive formats, you may specify the option when you call pg_restore.

-d | --inserts

Dump data as INSERT commands (rather than COPY). This will make restoration very slow; it is mainly useful for making dumps that can be loaded into non-PostgreSQL-based databases. Also, since this option generates a separate command for each row, an error in reloading a row causes only that row to be lost rather than the entire table contents. Note that the restore may fail altogether if you have rearranged column order. The -D option is safe against column order changes, though even slower.

-D | --column-inserts | --attribute-inserts

Dump data as INSERT commands with explicit column names (INSERT INTO table (column, ...) VALUES ...). This will make restoration very slow; it is mainly useful for making dumps that can be loaded into non-PostgreSQL-based
databases. Also, since this option generates a separate command for each row, an error in reloading a row causes only that row to be lost rather than the entire table contents.

-E encoding | --encoding=encoding
Create the dump in the specified character set encoding. By default, the dump is created in the database encoding. (Another way to get the same result is to set the PGCLIENTENCODING environment variable to the desired dump encoding.)

-f file | --file=file
Send output to the specified file. If this is omitted, the standard output is used.

-F p|c|t | --format=plain|custom|tar
Selects the format of the output. format can be one of the following:

p | plain — Output a plain-text SQL script file (the default).
c | custom — Output a custom archive suitable for input into pg_restore. This is the most flexible format in that it allows reordering of loading data as well as object definitions. This format is also compressed by default.
t | tar — Output a tar archive suitable for input into pg_restore. Using this archive format allows reordering and/or exclusion of database objects at the time the database is restored. It is also possible to limit which data is reloaded at restore time.

-i | --ignore-version
Ignore version mismatch between pg_dump and the database server. pg_dump can dump from servers running previous releases of Greenplum Database (or PostgreSQL), but very old versions may not be supported anymore. Use this option if you need to override the version check.

-n schema | --schema=schema
Dump only schemas matching the schema pattern; this selects both the schema itself, and all its contained objects. When this option is not specified, all non-system schemas in the target database will be dumped. Multiple schemas can be selected by writing multiple -n switches. Also, the schema parameter is interpreted as a pattern according to the same rules used by psql’s \d commands, so multiple schemas can also be selected by writing wildcard characters in the pattern. When using wildcards, be careful to quote the pattern if needed to prevent the shell from expanding the wildcards.

Note: When -n is specified, pg_dump makes no attempt to dump any other database objects that the selected schema(s) may depend upon. Therefore, there is no guarantee that the results of a specific-schema dump can be successfully restored by themselves into a clean database.

Note: Non-schema objects such as blobs are not dumped when -n is specified. You can add blobs back to the dump with the --blobs switch.
\texttt{-N schema \mid --exclude-schema=schema}

Do not dump any schemas matching the schema pattern. The pattern is interpreted according to the same rules as for \texttt{-n}. \texttt{-N} can be given more than once to exclude schemas matching any of several patterns. When both \texttt{-n} and \texttt{-N} are given, the behavior is to dump just the schemas that match at least one \texttt{-n} switch but no \texttt{-N} switches. If \texttt{-N} appears without \texttt{-n}, then schemas matching \texttt{-N} are excluded from what is otherwise a normal dump.

\texttt{-o \mid --oids}

Dump object identifiers (OIDs) as part of the data for every table. Use of this option is not recommended for files that are intended to be restored into Greenplum Database.

\texttt{-O \mid --no-owner}

Do not output commands to set ownership of objects to match the original database. By default, \texttt{pg_dump} issues ALTER OWNER or SET SESSION AUTHORIZATION statements to set ownership of created database objects. These statements will fail when the script is run unless it is started by a superuser (or the same user that owns all of the objects in the script). To make a script that can be restored by any user, but will give that user ownership of all the objects, specify \texttt{-O}. This option is only meaningful for the plain-text format. For the archive formats, you may specify the option when you call \texttt{pg_restore}.

\texttt{-s \mid --schema-only}

Dump only the object definitions (schema), not data.

\texttt{-S username \mid --superuser=typename}

Specify the superuser user name to use when disabling triggers. This is only relevant if \texttt{--disable-triggers} is used. It is better to leave this out, and instead start the resulting script as a superuser.

\texttt{-t table \mid --table=table}

Dump only tables (or views or sequences) matching the table pattern. Multiple tables can be selected by writing multiple \texttt{-t} switches. Also, the table parameter is interpreted as a pattern according to the same rules used by \texttt{psql}'s \texttt{\!d} commands, so multiple tables can also be selected by writing wildcard characters in the pattern. When using wildcards, be careful to quote the pattern if needed to prevent the shell from expanding the wildcards. The \texttt{-n} and \texttt{-N} switches have no effect when \texttt{-t} is used, because tables selected by \texttt{-t} will be dumped regardless of those switches, and non-table objects will not be dumped.

\textbf{Note:} When \texttt{-t} is specified, \texttt{pg_dump} makes no attempt to dump any other database objects that the selected table(s) may depend upon. Therefore, there is no guarantee that the results of a specific-table dump can be successfully restored by themselves into a clean database.

\textbf{Note:} \texttt{-t} cannot be used to specify a child table partition. To dump a partitioned table, you must specify the parent table name.
-T table | --exclude-table=table
Do not dump any tables matching the table pattern. The pattern is interpreted according to the same rules as for -t. -T can be given more than once to exclude tables matching any of several patterns. When both -t and -T are given, the behavior is to dump just the tables that match at least one -t switch but no -T switches. If -T appears without -t, then tables matching -T are excluded from what is otherwise a normal dump.

-v | --verbose
Specifies verbose mode. This will cause pg_dump to output detailed object comments and start/stop times to the dump file, and progress messages to standard error.

-x | --no-privileges | --no-acl
Prevent dumping of access privileges (GRANT/REVOKE commands).

--disable-dollar-quoting
This option disables the use of dollar quoting for function bodies, and forces them to be quoted using SQL standard string syntax.

--disable-triggers
This option is only relevant when creating a data-only dump. It instructs pg_dump to include commands to temporarily disable triggers on the target tables while the data is reloaded. Use this if you have triggers on the tables that you do not want to invoke during data reload. The commands emitted for --disable-triggers must be done as superuser. So, you should also specify a superuser name with -S, or preferably be careful to start the resulting script as a superuser. This option is only meaningful for the plain-text format. For the archive formats, you may specify the option when you call pg_restore.

--use-set-session-authorization
Output SQL-standard SET SESSION AUTHORIZATION commands instead of ALTER OWNER commands to determine object ownership. This makes the dump more standards compatible, but depending on the history of the objects in the dump, may not restore properly. A dump using SET SESSION AUTHORIZATION will require superuser privileges to restore correctly, whereas ALTER OWNER requires lesser privileges.

--gp-syntax | --no-gp-syntax
Use --gp-syntax to dump Greenplum Database syntax in the CREATE TABLE statements. This allows the distribution policy (DISTRIBUTED BY or DISTRIBUTED RANDOMLY clauses) of a Greenplum Database table to be dumped, which is useful for restoring into other Greenplum Database systems. The default is to include Greenplum Database syntax when connected to a Greenplum system, and to exclude it when connected to a regular PostgreSQL system.
Specify the compression level to use in archive formats that support compression. Currently only the *custom* archive format supports compression.

**Connection Options**

`-h host | --host host`

The host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable `PGHOST` or defaults to localhost.

`-p port | --port port`

The TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable `PGPORT` or defaults to 5432.

`-U username | --username username`

The database role name to connect as. If not specified, reads from the environment variable `PGUSER` or defaults to the current system role name.

`-W | --password`

Force a password prompt.

**Notes**

When a data-only dump is chosen and the option `--disable-triggers` is used, `pg_dump` emits commands to disable triggers on user tables before inserting the data and commands to re-enable them after the data has been inserted. If the restore is stopped in the middle, the system catalogs may be left in the wrong state.

Members of `tar` archives are limited to a size less than 8 GB. (This is an inherent limitation of the `tar` file format.) Therefore this format cannot be used if the textual representation of any one table exceeds that size. The total size of a tar archive and any of the other output formats is not limited, except possibly by the operating system.

The dump file produced by `pg_dump` does not contain the statistics used by the optimizer to make query planning decisions. Therefore, it is wise to run `ANALYZE` after restoring from a dump file to ensure good performance.

**Examples**

Dump a database called `mydb` into a SQL-script file:

```
pg_dump mydb > db.sql
```

To reload such a script into a (freshly created) database named `newdb`:

```
psql -d newdb -f db.sql
```

Dump a Greenplum database in tar file format and include distribution policy information:

```
pg_dump -Ft --gp-syntax mydb > db.tar
```
To dump a database into a custom-format archive file:

```
pg_dump -Fc mydb > db.dump
```

To reload an archive file into a (freshly created) database named `newdb`:

```
pg_restore -d newdb db.dump
```

To dump a single table named `mytab`:

```
pg_dump -t mytab mydb > db.sql
```

To specify an upper-case or mixed-case name in `-t` and related switches, you need to double-quote the name; else it will be folded to lower case. But double quotes are special to the shell, so in turn they must be quoted. Thus, to dump a single table with a mixed-case name, you need something like:

```
pg_dump -t "MixedCaseName" mydb > mytab.sql
```

**See Also**

`gp_dump`, `pg_dumpall`, `pg_restore`, `gp_restore`, `psql`
**pg_dumpall**

Extracts all databases in a Greenplum Database system to a single script file or other archive file.

---

**Synopsis**

```
pg_dumpall [connection_option ...] [dump_option ...]
```

---

**Description**

`pg_dumpall` is a standard PostgreSQL utility for backing up all databases in a Greenplum Database (or PostgreSQL) instance, and is also supported in Greenplum Database. It creates a single (non-parallel) dump file. For routine backups of Greenplum Database it is better to use Greenplum’s parallel dump utility, `gp_dump`, for the best performance.

`pg_dumpall` creates a single script file that contains SQL commands that can be used as input to `psql` to restore the databases. It does this by calling `pg_dump` for each database. `pg_dumpall` also dumps global objects that are common to all databases. (`pg_dump` does not save these objects.) This currently includes information about database users and groups, and access permissions that apply to databases as a whole.

Since `pg_dumpall` reads tables from all databases you will most likely have to connect as a database superuser in order to produce a complete dump. Also you will need superuser privileges to execute the saved script in order to be allowed to add users and groups, and to create databases.

The SQL script will be written to the standard output. Shell operators should be used to redirect it into a file.

`pg_dumpall` needs to connect several times to the Greenplum Database master server (once per database). If you use password authentication it is likely to ask for a password each time. It is convenient to have a `-/.pgpass` file in such cases.

---

**Options**

**Dump Options**

- `-a` | `--data-only`
  
  Dump only the data, not the schema (data definitions). This option is only meaningful for the plain-text format. For the archive formats, you may specify the option when you call `pg_restore`.

- `-c` | `--clean`
  
  Output commands to clean (drop) database objects prior to (the commands for) creating them. This option is only meaningful for the plain-text format. For the archive formats, you may specify the option when you call `pg_restore`. 

---

```
-d | --inserts
Dump data as INSERT commands (rather than COPY). This will make restoration very slow; it is mainly useful for making dumps that can be loaded into non-PostgreSQL-based databases. Also, since this option generates a separate command for each row, an error in reloading a row causes only that row to be lost rather than the entire table contents. Note that the restore may fail altogether if you have rearranged column order. The -D option is safe against column order changes, though even slower.

-D | --column-inserts | --attribute-inserts
Dump data as INSERT commands with explicit column names (INSERT INTO table (column, ...) VALUES ...). This will make restoration very slow; it is mainly useful for making dumps that can be loaded into non-PostgreSQL-based databases. Also, since this option generates a separate command for each row, an error in reloading a row causes only that row to be lost rather than the entire table contents.

-f | --filespaces
Dump filesystem definitions.

-g | --globals-only
Dump only global objects (roles and tablespaces), no databases.

-i | --ignore-version
Ignore version mismatch between pg_dump and the database server. pg_dump can dump from servers running previous releases of Greenplum Database (or PostgreSQL), but very old versions may not be supported anymore. Use this option if you need to override the version check.

-o | --oids
Dump object identifiers (OIDs) as part of the data for every table. Use of this option is not recommended for files that are intended to be restored into Greenplum Database.

-O | --no-owner
Do not output commands to set ownership of objects to match the original database. By default, pg_dump issues ALTER OWNER or SET SESSION AUTHORIZATION statements to set ownership of created database objects. These statements will fail when the script is run unless it is started by a superuser (or the same user that owns all of the objects in the script). To make a script that can be restored by any user, but will give that user ownership of all the objects, specify -O. This option is only meaningful for the plain-text format. For the archive formats, you may specify the option when you call pg_restore.

-r | --resource-queues
Dump resource queue definitions.
-s | --schema-only
Dump only the object definitions (schema), not data.

-S username | --superuser=username
Specify the superuser user name to use when disabling triggers. This is only relevant if --disable-triggers is used. It is better to leave this out, and instead start the resulting script as a superuser.

-v | --verbose
Specifies verbose mode. This will cause `pg_dump` to output detailed object comments and start/stop times to the dump file, and progress messages to standard error.

-x | --no-privileges | --no-acl
Prevent dumping of access privileges (GRANT/REVOKE commands).

--disable-dollar-quoting
This option disables the use of dollar quoting for function bodies, and forces them to be quoted using SQL standard string syntax.

--disable-triggers
This option is only relevant when creating a data-only dump. It instructs `pg_dumpall` to include commands to temporarily disable triggers on the target tables while the data is reloaded. Use this if you have triggers on the tables that you do not want to invoke during data reload. The commands emitted for --disable-triggers must be done as superuser. So, you should also specify a superuser name with -S, or preferably be careful to start the resulting script as a superuser.

--use-set-session-authorization
Output SQL-standard SET SESSION AUTHORIZATION commands instead of ALTER OWNER commands to determine object ownership. This makes the dump more standards compatible, but depending on the history of the objects in the dump, may not restore properly. A dump using SET SESSION AUTHORIZATION will require superuser privileges to restore correctly, whereas ALTER OWNER requires lesser privileges.

--gp-syntax
Output Greenplum Database syntax in the CREATE TABLE statements. This allows the distribution policy (DISTRIBUTED BY or DISTRIBUTED RANDOMLY clauses) of a Greenplum Database table to be dumped, which is useful for restoring into other Greenplum Database systems.
Connection Options

-h host | --host host
The host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable PGHOST or defaults to localhost.

-p port | --port port
The TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable PGPORT or defaults to 5432.

-U username | --username username
The database role name to connect as. If not specified, reads from the environment variable PGUSER or defaults to the current system role name.

-W | --password
Force a password prompt.

Notes
Since pg_dumpall calls pg_dump internally, some diagnostic messages will refer to pg_dump.

Once restored, it is wise to run ANALYZE on each database so the query planner has useful statistics. You can also run vacuumdb -a -z to analyze all databases.

pg_dumpall requires all needed tablespace (filesystem) directories to exist before the restore or database creation will fail for databases in non-default locations.

Examples
To dump all databases:
   pg_dumpall > db.out
To reload this file:
   psql template1 -f db.out
To dump only global objects (including filesystems and resource queues):
   pg_dumpall -g -f -r

See Also
gp_dump, pg_dump
**pg_restore**

Restores a database from an archive file created by `pg_dump`.

**Synopsis**

```
pg_restore [connection_option ...] [restore_option ...] filename
```

**Description**

`pg_restore` is a utility for restoring a database from an archive created by `pg_dump` in one of the non-plain-text formats. It will issue the commands necessary to reconstruct the database to the state it was in at the time it was saved. The archive files also allow `pg_restore` to be selective about what is restored, or even to reorder the items prior to being restored.

`pg_restore` can operate in two modes. If a database name is specified, the archive is restored directly into the database. Otherwise, a script containing the SQL commands necessary to rebuild the database is created and written to a file or standard output. The script output is equivalent to the plain text output format of `pg_dump`. Some of the options controlling the output are therefore analogous to `pg_dump` options.

`pg_restore` cannot restore information that is not present in the archive file. For instance, if the archive was made using the “dump data as INSERT commands” option, `pg_restore` will not be able to load the data using `COPY` statements.

**Options**

- **filename**
  - Specifies the location of the archive file to be restored. If not specified, the standard input is used.

- **-a | --data-only**
  - Restore only the data, not the schema (data definitions).

- **-c | --clean**
  - Clean (drop) database objects before recreating them.

- **-C | --create**
  - Create the database before restoring into it. (When this option is used, the database named with `-d` is used only to issue the initial `CREATE DATABASE` command. All data is restored into the database name that appears in the archive.)

- **-d dbname | --dbname=dbname**
  - Connect to this database and restore directly into this database. The default is to use the `PGDATABASE` environment variable setting, or the same name as the current system user.
-e | --exit-on-error
   Exit if an error is encountered while sending SQL commands to the database. The default is to continue and to display a count of errors at the end of the restoration.

-f outfilename | --file=outfilename
   Specify output file for generated script, or for the listing when used with -l. Default is the standard output.

-F t|c | --format=tar|custom
   The format of the archive produced by pg_dump. It is not necessary to specify the format, since pg_restore will determine the format automatically. Format can be either tar or custom.

-i | --ignore-version
   Ignore database version checks.

-I index | --index=index
   Restore definition of named index only.

-l | --list
   List the contents of the archive. The output of this operation can be used with the -L option to restrict and reorder the items that are restored.

-L list-file | --use-list=list-file
   Restore elements in the list-file only, and in the order they appear in the file. Lines can be moved and may also be commented out by placing a ; at the start of the line.

-n schema | --schema=schema
   Restore only objects that are in the named schema. This can be combined with the -t option to restore just a specific table.

-O | --no-owner
   Do not output commands to set ownership of objects to match the original database. By default, pg_restore issues ALTER OWNER or SET SESSION AUTHORIZATION statements to set ownership of created schema elements. These statements will fail unless the initial connection to the database is made by a superuser (or the same user that owns all of the objects in the script). With -O, any user name can be used for the initial connection, and this user will own all the created objects.

-P 'function-name(argtype [, ...])' | --function='function-name(argtype [, ...])'
   Restore the named function only. The function name must be enclosed in quotes. Be careful to spell the function name and arguments exactly as they appear in the dump file’s table of contents (as shown by the --list option).
-s | --schema-only
    Restore only the schema (data definitions), not the data (table contents). Sequence current values will not be restored, either. (Do not confuse this with the --schema option, which uses the word schema in a different meaning.)

-S username | --superuser=username
    Specify the superuser user name to use when disabling triggers. This is only relevant if --disable-triggers is used.

-t table | --table=table
    Restore definition and/or data of named table only.

-T trigger | --trigger=trigger
    Restore named trigger only.

-v | --verbose
    Specifies verbose mode.

-x | --no-privileges | --no-acl
    Prevent restoration of access privileges (GRANT/REVOKE commands).

--disable-triggers
    This option is only relevant when performing a data-only restore. It instructs pg_restore to execute commands to temporarily disable triggers on the target tables while the data is reloaded. Use this if you have triggers on the tables that you do not want to invoke during data reload. The commands emitted for --disable-triggers must be done as superuser. So, you should also specify a superuser name with -S, or preferably run pg_restore as a superuser.

--no-data-for-failed-tables
    By default, table data is restored even if the creation command for the table failed (e.g., because it already exists). With this option, data for such a table is skipped. This behavior is useful when the target database may already contain the desired table contents. Specifying this option prevents duplicate or obsolete data from being loaded. This option is effective only when restoring directly into a database, not when producing SQL script output.

Connection Options

-h host | --host host
    The host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable PGHOST or defaults to localhost.

-p port | --port port
    The TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable PGPORT or defaults to 5432.
-U username | --username username
  The database role name to connect as. If not specified, reads from the environment
  variable PGUSER or defaults to the current system role name.

-W | --password
  Force a password prompt.

-1 | --single-transaction
  Execute the restore as a single transaction. This ensures that either all the commands
  complete successfully, or no changes are applied.

Notes
If your installation has any local additions to the template1 database, be careful to
load the output of pg_restore into a truly empty database; otherwise you are likely
to get errors due to duplicate definitions of the added objects. To make an empty
database without any local additions, copy from template0 not template1, for
database:

    CREATE DATABASE foo WITH TEMPLATE template0;

When restoring data to a pre-existing table and the option --disable-triggers is
used, pg_restore emits commands to disable triggers on user tables before inserting
the data then emits commands to re-enable them after the data has been inserted. If the
restore is stopped in the middle, the system catalogs may be left in the wrong state.

pg_restore will not restore large objects for a single table. If an archive contains
large objects, then all large objects will be restored.

See also the pg_dump documentation for details on limitations of pg_dump.

Once restored, it is wise to run ANALYZE on each restored table so the query planner
has useful statistics.

Examples
Assume we have dumped a database called mydb into a custom-format dump file:

    pg_dump -Fc mydb > db.dump

To drop the database and recreate it from the dump:

    dropdb mydb
    pg_restore -C -d template1 db.dump

To reload the dump into a new database called newdb. Notice there is no -C, we
instead connect directly to the database to be restored into. Also note that we clone the
new database from template0 not template1, to ensure it is initially empty:

    createdb -T template0 newdb
    pg_restore -d newdb db.dump

To reorder database items, it is first necessary to dump the table of contents of the
archive:
pg_restore -l db.dump > db.list

The listing file consists of a header and one line for each item, for example,

```
; Archive created at Fri Jul 28 22:28:36 2006
;    dbname: mydb
;    TOC Entries: 74
;    Compression: 0
;    Dump Version: 1.4-0
;    Format: CUSTOM
;
; Selected TOC Entries:
;
2; 145344 TABLE species postgres
3; 145344 ACL species
4; 145359 TABLE nt_header postgres
5; 145359 ACL nt_header
6; 145402 TABLE species_records postgres
7; 145402 ACL species_records
8; 145416 TABLE ss_old postgres
9; 145416 ACL ss_old
10; 145433 TABLE map_resolutions postgres
11; 145433 ACL map_resolutions
12; 145443 TABLE hs_old postgres
13; 145443 ACL hs_old
```

Semicolons start a comment, and the numbers at the start of lines refer to the internal archive ID assigned to each item. Lines in the file can be commented out, deleted, and reordered. For example,

```
10; 145433 TABLE map_resolutions postgres
;2; 145344 TABLE species postgres
;4; 145359 TABLE nt_header postgres
6; 145402 TABLE species_records postgres
;8; 145416 TABLE ss_old postgres
```

Could be used as input to `pg_restore` and would only restore items 10 and 6, in that order:

```
pg_restore -L db.list db.dump
```

**See Also**

`pg_dump`, `gp_restore`, `gp_dump`
**psql**

Interactive command-line interface for Greenplum Database

**Synopsis**

```
psql [option...] [dbname [username]]
```

**Description**

`psql` is a terminal-based front-end to Greenplum Database. It enables you to type in queries interactively, issue them to Greenplum Database, and see the query results. Alternatively, input can be from a file. In addition, it provides a number of meta-commands and various shell-like features to facilitate writing scripts and automating a wide variety of tasks.

**Options**

- `a` | `--echo-all`
  
  Print all input lines to standard output as they are read. This is more useful for script processing rather than interactive mode.

- `A` | `--no-align`
  
  Switches to unaligned output mode. (The default output mode is aligned.)

- `c 'command'` | `--command 'command'`
  
  Specifies that `psql` is to execute the specified command string, and then exit. This is useful in shell scripts. `command` must be either a command string that is completely parseable by the server, or a single backslash command. Thus you cannot mix SQL and `psql` meta-commands with this option. To achieve that, you could pipe the string into `psql`, like this: `echo '\x \ SELECT * FROM foo;' | psql`. (\x is the separator meta-command.)

  If the command string contains multiple SQL commands, they are processed in a single transaction, unless there are explicit `BEGIN/COMMIT` commands included in the string to divide it into multiple transactions. This is different from the behavior when the same string is fed to `psql`’s standard input.

- `d dbname` | `--dbname dbname`
  
  Specifies the name of the database to connect to. This is equivalent to specifying `dbname` as the first non-option argument on the command line.

- `e` | `--echo-queries`
  
  Copy all SQL commands sent to the server to standard output as well.

- `E` | `--echo-hidden`
  
  Echo the actual queries generated by `\d` and other backslash commands. You can use this to study `psql`’s internal operations.
-f filename | --file filename

Use a file as the source of commands instead of reading commands interactively. After the file is processed, psql terminates. If filename is - (hyphen), then standard input is read. Using this option is subtly different from writing psql < filename. In general, both will do what you expect, but using -f enables some nice features such as error messages with line numbers.

-F separator | --field-separator separator

Use the specified separator as the field separator for unaligned output.

-H | --html

Turn on HTML tabular output.

-l | --list

List all available databases, then exit. Other non-connection options are ignored.

-L filename | --log-file filename

Write all query output into the specified log file, in addition to the normal output destination.

-o filename | --output filename

Put all query output into the specified file.

-P assignment | --pset assignment

Allows you to specify printing options in the style of \pset on the command line. Note that here you have to separate name and value with an equal sign instead of a space. Thus to set the output format to LaTeX, you could write -P format=latex.

-q | --quiet

Specifies that psql should do its work quietly. By default, it prints welcome messages and various informational output. If this option is used, none of this happens. This is useful with the -c option.

-R separator | --record-separator separator

Use separator as the record separator for unaligned output.

-s | --single-step

Run in single-step mode. That means the user is prompted before each command is sent to the server, with the option to cancel execution as well. Use this to debug scripts.

-S | --single-line

Runs in single-line mode where a new line terminates an SQL command, as a semicolon does.

-t | --tuples-only

Turn off printing of column names and result row count footers, etc.
-T table_options | --table-attr table_options
   Allows you to specify options to be placed within the HTML table tag. See \pset for details.

-v assignment | --set assignment | --variable assignment
   Perform a variable assignment, like the \set internal command. Note that you must separate name and value, if any, by an equal sign on the command line. To unset a variable, leave off the equal sign. To just set a variable without a value, use the equal sign but leave off the value. These assignments are done during a very early stage of start-up, so variables reserved for internal purposes might get overwritten later.

-V | --version
   Print the psql version and exit.

-x | --expanded
   Turn on the expanded table formatting mode.

-X | --no-psqlrc
   Do not read the start-up file (neither the system-wide psqlrc file nor the user’s ~/.psqlrc file).

-1 | --single-transaction
   When psql executes a script with the -f option, adding this option wraps BEGIN/COMMIT around the script to execute it as a single transaction. This ensures that either all the commands complete successfully, or no changes are applied.

   If the script itself uses BEGIN, COMMIT, or ROLLBACK, this option will not have the desired effects. Also, if the script contains any command that cannot be executed inside a transaction block, specifying this option will cause that command (and hence the whole transaction) to fail.

-? | --help
   Show help about psql command line arguments, and exit.

Connection Options

-h host | --host host
   The host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable PGHOST or defaults to localhost.

-p port | --port port
   The TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable PGPORT or defaults to 5432.
-U *username* | --username *username*

The database role name to connect as. If not specified, reads from the environment variable `PGUSER` or defaults to the current system role name.

-W | --password

Force a password prompt. `psql` should automatically prompt for a password whenever the server requests password authentication. However, currently password request detection is not totally reliable, hence this option to force a prompt. If no password prompt is issued and the server requires password authentication, the connection attempt will fail.

**Exit Status**

`psql` returns 0 to the shell if it finished normally, 1 if a fatal error of its own (out of memory, file not found) occurs, 2 if the connection to the server went bad and the session was not interactive, and 3 if an error occurred in a script and the variable `ON_ERROR_STOP` was set.

**Usage**

**Connecting To A Database**

`psql` is a client application for Greenplum Database. In order to connect to a database you need to know the name of your target database, the host name and port number of the Greenplum master server and what database user name you want to connect as. `psql` can be told about those parameters via command line options, namely `-d`, `-h`, `-p`, and `-U` respectively. If an argument is found that does not belong to any option it will be interpreted as the database name (or the user name, if the database name is already given). Not all these options are required; there are useful defaults. If you omit the host name, `psql` will connect via a UNIX-domain socket to a master server on the local host, or via TCP/IP to `localhost` on machines that do not have UNIX-domain sockets. The default master port number is 5432. If you use a different port for the master, you must specify the port. The default database user name is your UNIX user name, as is the default database name. Note that you cannot just connect to any database under any user name. Your database administrator should have informed you about your access rights.

When the defaults are not right, you can save yourself some typing by setting the environment variables `PGDATABASE`, `PGHOST`, `PGPORT` and/or `PGUSER` to appropriate values.

It is also convenient to have a `~/.pgpass` file to avoid regularly having to type in passwords. This file should reside in your home directory and contain lines of the following format:

```
hostname:port:database:username:password
```

The permissions on `.pgpass` must disallow any access to world or group (for example: `chmod 0600 ~/.pgpass`). If the permissions are less strict than this, the file will be ignored. (The file permissions are not currently checked on Microsoft Windows clients, however.)
If the connection could not be made for any reason (insufficient privileges, server is not running, etc.), `psql` will return an error and terminate.

**Entering SQL Commands**

In normal operation, `psql` provides a prompt with the name of the database to which `psql` is currently connected, followed by the string `=>` for a regular user or `=#` for a superuser. For example:

```
testdb=>
testdb=#
```

At the prompt, the user may type in SQL commands. Ordinarily, input lines are sent to the server when a command-terminating semicolon is reached. An end of line does not terminate a command. Thus commands can be spread over several lines for clarity. If the command was sent and executed without error, the results of the command are displayed on the screen.

**Meta-Commands**

Anything you enter in `psql` that begins with an unquoted backslash is a `psql` meta-command that is processed by `psql` itself. These commands help make `psql` more useful for administration or scripting. Meta-commands are more commonly called slash or backslash commands.

The format of a `psql` command is the backslash, followed immediately by a command verb, then any arguments. The arguments are separated from the command verb and each other by any number of whitespace characters.

To include whitespace into an argument you may quote it with a single quote. To include a single quote into such an argument, use two single quotes. Anything contained in single quotes is furthermore subject to C-like substitutions for `\n` (new line), `\t` (tab), `\digits` (octal), and `\xdigits` (hexadecimal).

If an unquoted argument begins with a colon (`:`), it is taken as a `psql` variable and the value of the variable is used as the argument instead.

Arguments that are enclosed in backquotes (``) are taken as a command line that is passed to the shell. The output of the command (with any trailing newline removed) is taken as the argument value. The above escape sequences also apply in backquotes.

Some commands take an SQL identifier (such as a table name) as argument. These arguments follow the syntax rules of SQL: Unquoted letters are forced to lowercase, while double quotes ("`) protect letters from case conversion and allow incorporation of whitespace into the identifier. Within double quotes, paired double quotes reduce to a single double quote in the resulting name. For example, `FOO"BAR"BAZ` is interpreted as `fooBARbaz`, and "A weird" name becomes A weird name.

Parsing for arguments stops when another unquoted backslash occurs. This is taken as the beginning of a new meta-command. The special sequence `\ \` (two backslashes) marks the end of arguments and continues parsing SQL commands, if any. That way SQL and `psql` commands can be freely mixed on a line. But in any case, the arguments of a meta-command cannot continue beyond the end of the line.

The following meta-commands are defined:
\a

If the current table output format is unaligned, it is switched to aligned. If it is not unaligned, it is set to unaligned. This command is kept for backwards compatibility. See \pset for a more general solution.

\cd [directory]

Changes the current working directory. Without argument, changes to the current user’s home directory. To print your current working directory, use \!pwd.

\C [title]

Sets the title of any tables being printed as the result of a query or unset any such title. This command is equivalent to \pset title.

\c | \connect [dbname [username] [host] [port]]

Establishes a new connection. If the new connection is successfully made, the previous connection is closed. If any of dbname, username, host or port are omitted, the value of that parameter from the previous connection is used. If the connection attempt failed, the previous connection will only be kept if psql is in interactive mode. When executing a non-interactive script, processing will immediately stop with an error. This distinction was chosen as a user convenience against typos, and a safety mechanism that scripts are not accidentally acting on the wrong database.

\copy {table [(column_list)] | (query)} {from | to} {filename | stdin | stdout | pstdin | pstdout} [with] [binary] [oids] [delimiter [as] 'character'] [null [as] 'string'] [csv [header] [quote [as] 'character'] [force quote column_list] [force not null column_list]]

Performs a frontend (client) copy. This is an operation that runs an SQL COPY command, but instead of the server reading or writing the specified file, psql reads or writes the file and routes the data between the server and the local file system. This means that file accessibility and privileges are those of the local user, not the server, and no SQL superuser privileges are required.

The syntax of the command is similar to that of the SQL COPY command. Note that, because of this, special parsing rules apply to the \copy command. In particular, the variable substitution rules and backslash escapes do not apply.

\copy ... from stdin | to stdout reads/writes based on the command input and output respectively. All rows are read from the same source that issued the command, continuing until \. is read or the stream reaches EOF. Output is sent to the same place as command output. To read/write from psql’s standard input or output, use pstdin or pstdout. This option is useful for populating tables in-line within a SQL script file.

This operation is not as efficient as the SQL COPY command because all data must pass through the client/server connection.

\copyright

Shows the copyright and distribution terms of PostgreSQL on which Greenplum Database is based.
\d [relation_pattern] | \d+ [relation_pattern]
For each relation (table, external table, view, index, or sequence) matching the
relation pattern, show all columns, their types, the tablespace (if not the default) and
any special attributes such as NOT NULL or defaults, if any. Associated indexes,
constraints, rules, and triggers are also shown, as is the view definition if the relation
is a view.

The command form \d+ is identical, except that more information is displayed: any
comments associated with the columns of the table are shown, as is the presence of
OIDs in the table.

If \d is used without a pattern argument, it is equivalent to \dt vs which will show a
list of all tables, views, and sequences.

\da [aggregate_pattern]
Lists all available aggregate functions, together with the data types they operate on.
If a pattern is specified, only aggregates whose names match the pattern are shown.

\db [tablespace_pattern] | \db+ [tablespace_pattern]
Lists all available tablespaces and their corresponding filespace locations. If pattern
is specified, only tablespaces whose names match the pattern are shown. If + is
appended to the command name, each object is listed with its associated
permissions.

\dc [conversion_pattern]
Lists all available conversions between character-set encodings. If pattern is
specified, only conversions whose names match the pattern are listed.

\dC
Lists all available type casts.

\dd [object_pattern]
Lists all available objects. If pattern is specified, only matching objects are shown.

\dD [domain_pattern]
Lists all available domains. If pattern is specified, only matching domains are
shown.

\df [function_pattern] | \df+ [function_pattern ]
Lists available functions, together with their argument and return types. If pattern is
specified, only functions whose names match the pattern are shown. If the form
\df+ is used, additional information about each function, including language and
description, is shown. To reduce clutter, \df does not show data type I/O functions.
This is implemented by ignoring functions that accept or return type cstring.

\dg [role_pattern]
Lists all database roles. If pattern is specified, only those roles whose names match
the pattern are listed.
\distPvxS [index | sequence | table | parent table | view
| external_table | system_object]

This is not the actual command name: the letters i, s, t, p, v, x, S stand for index, sequence, table, parent table, view, external table, and system table, respectively. You can specify any or all of these letters, in any order, to obtain a listing of all the matching objects. The letter S restricts the listing to system objects; without S, only non-system objects are shown. If + is appended to the command name, each object is listed with its associated description, if any. If a pattern is specified, only objects whose names match the pattern are listed.

\dl
This is an alias for \lo_list, which shows a list of large objects.

\dn [schema_pattern] | \dn+ [schema_pattern]
Lists all available schemas (namespaces). If pattern is specified, only schemas whose names match the pattern are listed. Non-local temporary schemas are suppressed. If + is appended to the command name, each object is listed with its associated permissions and description, if any.

\do [operator_pattern]
Lists available operators with their operand and return types. If pattern is specified, only operators whose names match the pattern are listed.

\dp [relation_pattern_to_show_privileges]
Produces a list of all available tables, views and sequences with their associated access privileges. If pattern is specified, only tables, views and sequences whose names match the pattern are listed. The GRANT and REVOKE commands are used to set access privileges.

\dT [datatype_pattern] | \dT+ [datatype_pattern]
Lists all data types or only those that match pattern. The command form \dT+ shows extra information.

\du [role_pattern]
Lists all database roles, or only those that match pattern.

\e | \edit [filename]
If a file name is specified, the file is edited; after the editor exits, its content is copied back to the query buffer. If no argument is given, the current query buffer is copied to a temporary file which is then edited in the same fashion. The new query buffer is then re-parsed according to the normal rules of psql, where the whole buffer is treated as a single line. (Thus you cannot make scripts this way. Use \i for that.) This means also that if the query ends with (or rather contains) a semicolon, it is immediately executed. In other cases it will merely wait in the query buffer.

psql searches the environment variables PSQL_EDITOR, EDITOR, and VISUAL (in that order) for an editor to use. If all of them are unset, vi is used on UNIX systems, notepad.exe on Windows systems.
\echo text [ ... ]
Prints the arguments to the standard output, separated by one space and followed by
a newline. This can be useful to intersperse information in the output of scripts.

If you use the \o command to redirect your query output you may wish to use
\qecho instead of this command.

\encoding [encoding]
Sets the client character set encoding. Without an argument, this command shows
the current encoding.

\f [field_separator_string]
Sets the field separator for unaligned query output. The default is the vertical bar
(|). See also \pset for a generic way of setting output options.

\g [{filename | |command }]
Sends the current query input buffer to the server and optionally stores the query’s
output in a file or pipes the output into a separate UNIX shell executing command.
A bare \g is virtually equivalent to a semicolon. A \g with argument is a one-shot
alternative to the \o command.

\h | \help [sql_command]
Gives syntax help on the specified SQL command. If a command is not specified,
then psql will list all the commands for which syntax help is available. Use an
asterisk (*) to show syntax help on all SQL commands. To simplify typing,
commands that consists of several words do not have to be quoted.

\H
Turns on HTML query output format. If the HTML format is already on, it is
switched back to the default aligned text format. This command is for compatibility
and convenience, but see \pset about setting other output options.

\i input_filename
Reads input from a file and executes it as though it had been typed on the keyboard.
If you want to see the lines on the screen as they are read you must set the variable
ECHO to all.

\l | \list | \l+ | \list+
List the names, owners, and character set encodings of all the databases in the
server. If + is appended to the command name, database descriptions are also
displayed.

\lo_export loid filename
Reads the large object with OID loid from the database and writes it to filename.
Note that this is subtly different from the server function lo_export, which acts
with the permissions of the user that the database server runs as and on the server’s
file system. Use \lo_list to find out the large object’s OID.
\lo_import large_object_filename [comment]
Stores the file into a large object. Optionally, it associates the given comment with
the object. Example:
mydb=> \lo_import '/home/gpadmin/pictures/photo.xcf' 'a picture of me'
lo_import 152801
The response indicates that the large object received object ID 152801 which one
ought to remember if one wants to access the object ever again. For that reason it is
recommended to always associate a human-readable comment with every object.
Those can then be seen with the \lo_list command. Note that this command is
subtly different from the server-side \lo_import because it acts as the local user on
the local file system, rather than the server’s user and file system.

\lo_list
Shows a list of all large objects currently stored in the database, along with any
comments provided for them.

\lo_unlink largeobject_oid
Deletes the large object of the specified OID from the database. Use \lo_list to
find out the large object’s OID.

\o [ {query_result_filename | |command} ]
Saves future query results to a file or pipes them into a UNIX shell command. If no
arguments are specified, the query output will be reset to the standard output. Query
results include all tables, command responses, and notices obtained from the
database server, as well as output of various backslash commands that query the
database (such as \d), but not error messages. To intersperse text output in between
query results, use \qecho.

\p
Print the current query buffer to the standard output.

\password [username]
Changes the password of the specified user (by default, the current user). This
command prompts for the new password, encrypts it, and sends it to the server as an
ALTER ROLE command. This makes sure that the new password does not appear in
cleartext in the command history, the server log, or elsewhere.

\pset print_option [value]
This command sets options affecting the output of query result tables.
print_option describes which option is to be set. Adjustable printing options are:

- **format** — Sets the output format to one of unaligned, aligned, html,
  latex, or troff-ms. First letter abbreviations are allowed. Unaligned writes
  all columns of a row on a line, separated by the currently active field
  separator. This is intended to create output that might be intended to be read in
  by other programs. Aligned mode is the standard, human-readable, nicely
  formatted text output that is default. The HTML and LaTeX modes put out
tables that are intended to be included in documents using the respective mark-up language. They are not complete documents! (This might not be so dramatic in HTML, but in LaTeX you must have a complete document wrapper.)

- **border** — The second argument must be a number. In general, the higher the number the more borders and lines the tables will have, but this depends on the particular format. In HTML mode, this will translate directly into the `border=...` attribute, in the others only values 0 (no border), 1 (internal dividing lines), and 2 (table frame) make sense.

- **expanded | x** — Toggles between regular and expanded format. When expanded format is enabled, query results are displayed in two columns, with the column name on the left and the data on the right. This mode is useful if the data would not fit on the screen in the normal horizontal mode. Expanded mode is supported by all four output formats.

- **null** — The second argument is a string that should be printed whenever a column is null. The default is not to print anything, which can easily be mistaken for, say, an empty string. Thus, one might choose to write `\pset null ' (null) '`.  

- **fieldsep** — Specifies the field separator to be used in unaligned output mode. That way one can create, for example, tab- or comma-separated output, which other programs might prefer. To set a tab as field separator, type `\pset fieldsep ' \t '.` The default field separator is ' | ' (a vertical bar).

- **footer** — Toggles the display of the default footer (x rows).

- **numericlocale** — Toggles the display of a locale-aware character to separate groups of digits to the left of the decimal marker. It also enables a locale-aware decimal marker.

- **recordsep** — Specifies the record (line) separator to use in unaligned output mode. The default is a newline character.

- **tuples_only | t** — Toggles between tuples only and full display. Full display may show extra information such as column headers, titles, and various footers. In tuples only mode, only actual table data is shown.

- **title [text]** — Sets the table title for any subsequently printed tables. This can be used to give your output descriptive tags. If no argument is given, the title is unset.

- **tableattr | T [text]** — Allows you to specify any attributes to be placed inside the HTML table tag. This could for example be `cellpadding` or `bgcolor`. Note that you probably don’t want to specify `border` here, as that is already taken care of by `\pset border`.

- **pager** — Controls the use of a pager for query and `psql` help output. When `on`, if the environment variable `PAGER` is set, the output is piped to the specified program. Otherwise a platform-dependent default (such as `more`) is used. When `off`, the pager is not used. When `on`, the pager is used only when appropriate. Pager can also be set to `always`, which causes the pager to be always used.

`\q`  
Quits the `psql` program.
\qecho text [ ... ]
This command is identical to \echo except that the output will be written to the query output channel, as set by \o.

\r
Resets (clears) the query buffer.

\s [history_filename]
Print or save the command line history to filename. If filename is omitted, the history is written to the standard output.

\set [name [value [ ... ]]]
Sets the internal variable name to value or, if more than one value is given, to the concatenation of all of them. If no second argument is given, the variable is just set with no value. To unset a variable, use the \unset command.

Valid variable names can contain characters, digits, and underscores. See “Variables” on page 757. Variable names are case-sensitive.

Although you are welcome to set any variable to anything you want, psql treats several variables as special. They are documented in the section about variables.

This command is totally separate from the SQL command SET.

\t
Toggles the display of output column name headings and row count footer.

\T table_options
Allows you to specify attributes to be placed within the table tag in HTML tabular output mode.

\timing
Toggles a display of how long each SQL statement takes, in milliseconds.

\w {filename | |command}
Outputs the current query buffer to a file or pipes it to a UNIX command.

\x
Toggles expanded table formatting mode.

\z [relation_to_show_privileges]
Produces a list of all available tables, views and sequences with their associated access privileges. If a pattern is specified, only tables, views and sequences whose names match the pattern are listed. This is an alias for \dp.

\! [command]
Escapes to a separate UNIX shell or executes the UNIX command. The arguments are not further interpreted, the shell will see them as is.
Shows help information about the `psql` backslash commands.

**Patterns**

The various `\d` commands accept a pattern parameter to specify the object name(s) to be displayed. In the simplest case, a pattern is just the exact name of the object. The characters within a pattern are normally folded to lower case, just as in SQL names; for example, `\dt FOO` will display the table named `foo`. As in SQL names, placing double quotes around a pattern stops folding to lower case. Should you need to include an actual double quote character in a pattern, write it as a pair of double quotes within a double-quote sequence; again this is in accord with the rules for SQL quoted identifiers. For example, `\dt "FOO""BAR"` will display the table named `FOO"BAR` (not `foo"bar`). Unlike the normal rules for SQL names, you can put double quotes around just part of a pattern, for instance `\dt FOO"FOO"BAR` will display the table named `fooFOO"BAR`.

Within a pattern, `*` matches any sequence of characters (including no characters) and `?` matches any single character. (This notation is comparable to UNIX shell file name patterns.) For example, `\dt int*` displays all tables whose names begin with `int`. But within double quotes, `*` and `?` lose these special meanings and are just matched literally.

A pattern that contains a dot (`.`) is interpreted as a schema name pattern followed by an object name pattern. For example, `\dt foo*.bar*` displays all tables whose table name starts with `bar` that are in schemas whose schema name starts with `foo`. When no dot appears, then the pattern matches only objects that are visible in the current schema search path. Again, a dot within double quotes loses its special meaning and is matched literally.

Advanced users can use regular-expression notations. All regular expression special characters work as specified in the PostgreSQL documentation on regular expressions, except for `.`, which is taken as a separator as mentioned above, `*` which is translated to the regular-expression notation `.*`, and `?` which is translated to `..`. You can emulate these pattern characters at need by writing `?` for `.`, `(R+|)` for `R*`, or `(R|)` for `R?`. Remember that the pattern must match the whole name, unlike the usual interpretation of regular expressions; write `*` at the beginning and/or end if you don’t wish the pattern to be anchored. Note that within double quotes, all regular expression special characters lose their special meanings and are matched literally. Also, the regular expression special characters are matched literally in operator name patterns (such as the argument of `\do`).

Whenever the pattern parameter is omitted completely, the `\d` commands display all objects that are visible in the current schema search path — this is equivalent to using the pattern `*`. To see all objects in the database, use the pattern `*.*`. 
Advanced Features

Variables

psql provides variable substitution features similar to common UNIX command shells. Variables are simply name/value pairs, where the value can be any string of any length. To set variables, use the psql meta-command \set:

```sql
testdb=> \set foo bar
```

sets the variable `foo` to the value `bar`. To retrieve the content of the variable, precede the name with a colon and use it as the argument of any slash command:

```sql
testdb=> \echo :foo
bar
```

**Note:** The arguments of `\set` are subject to the same substitution rules as with other commands. Thus you can construct interesting references such as `\set :foo 'something'` and get ‘soft links’ or ‘variable variables’ of Perl or PHP fame, respectively. Unfortunately, there is no way to do anything useful with these constructs. On the other hand, `\set bar :foo` is a perfectly valid way to copy a variable.

If you call `\set` without a second argument, the variable is set, with an empty string as value. To unset (or delete) a variable, use the command `\unset`.

psql’s internal variable names can consist of letters, numbers, and underscores in any order and any number of them. A number of these variables are treated specially by psql. They indicate certain option settings that can be changed at run time by altering the value of the variable or represent some state of the application. Although you can use these variables for any other purpose, this is not recommended, as the program behavior might behave unexpectedly. By convention, all specially treated variables consist of all upper-case letters (and possibly numbers and underscores). To ensure maximum compatibility in the future, avoid using such variable names for your own purposes. A list of all specially treated variables are as follows:

**AUTOCOMMIT**

When on (the default), each SQL command is automatically committed upon successful completion. To postpone commit in this mode, you must enter a `BEGIN` or `START TRANSACTION` SQL command. When off or unset, SQL commands are not committed until you explicitly issue `COMMIT` or `END`. The autocommit-on mode works by issuing an implicit `BEGIN` for you, just before any command that is not already in a transaction block and is not itself a `BEGIN` or other transaction-control command, nor a command that cannot be executed inside a transaction block (such as `VACUUM`).

In autocommit-off mode, you must explicitly abandon any failed transaction by entering `ABORT` or `ROLLBACK`. Also keep in mind that if you exit the session without committing, your work will be lost.

The autocommit-on mode is PostgreSQL’s traditional behavior, but autocommit-off is closer to the SQL spec. If you prefer autocommit-off, you may wish to set it in your `~/.psqlrc` file.
**DBNAME**

The name of the database you are currently connected to. This is set every time you connect to a database (including program start-up), but can be unset.

**ECHO**

If set to all, all lines entered from the keyboard or from a script are written to the standard output before they are parsed or executed. To select this behavior on program start-up, use the switch `-a`. If set to queries, `psql` merely prints all queries as they are sent to the server. The switch for this is `-e`.

**ECHO_HIDDEN**

When this variable is set and a backslash command queries the database, the query is first shown. This way you can study the Greenplum Database internals and provide similar functionality in your own programs. (To select this behavior on program start-up, use the switch `-E`.) If you set the variable to the value `noexec`, the queries are just shown but are not actually sent to the server and executed.

**ENCODING**

The current client character set encoding.

**FETCH_COUNT**

If this variable is set to an integer value > 0, the results of `SELECT` queries are fetched and displayed in groups of that many rows, rather than the default behavior of collecting the entire result set before display. Therefore only a limited amount of memory is used, regardless of the size of the result set. Settings of 100 to 1000 are commonly used when enabling this feature. Keep in mind that when using this feature, a query may fail after having already displayed some rows.

Although you can use any output format with this feature, the default aligned format tends to look bad because each group of `FETCH_COUNT` rows will be formatted separately, leading to varying column widths across the row groups. The other output formats work better.

**HISTCONTROL**

If this variable is set to `ignorespace`, lines which begin with a space are not entered into the history list. If set to a value of `ignoredups`, lines matching the previous history line are not entered. A value of `ignoreboth` combines the two options. If unset, or if set to any other value than those above, all lines read in interactive mode are saved on the history list.

**HISTFILE**

The file name that will be used to store the history list. The default value is `~/.psql_history`. For example, putting

```
\set HISTFILE ~/.psql_history- :DBNAME
```

in `~/.psqlrc` will cause `psql` to maintain a separate history for each database.

**HISTSIZE**

The number of commands to store in the command history. The default value is 500.
HOST
The database server host you are currently connected to. This is set every time you connect to a database (including program start-up), but can be unset.

IGNOREEOF
If unset, sending an EOF character (usually CTRL+D) to an interactive session of psql will terminate the application. If set to a numeric value, that many EOF characters are ignored before the application terminates. If the variable is set but has no numeric value, the default is 10.

LASTOID
The value of the last affected OID, as returned from an INSERT or lo_insert command. This variable is only guaranteed to be valid until after the result of the next SQL command has been displayed.

ON_ERROR_ROLLBACK
When on, if a statement in a transaction block generates an error, the error is ignored and the transaction continues. When interactive, such errors are only ignored in interactive sessions, and not when reading script files. When off (the default), a statement in a transaction block that generates an error aborts the entire transaction. The on_error_rollback-on mode works by issuing an implicit SAVEPOINT for you, just before each command that is in a transaction block, and rolls back to the savepoint on error.

ON_ERROR_STOP
By default, if non-interactive scripts encounter an error, such as a malformed SQL command or internal meta-command, processing continues. This has been the traditional behavior of psql but it is sometimes not desirable. If this variable is set, script processing will immediately terminate. If the script was called from another script it will terminate in the same fashion. If the outermost script was not called from an interactive psql session but rather using the -f option, psql will return error code 3, to distinguish this case from fatal error conditions (error code 1).

PORT
The database server port to which you are currently connected. This is set every time you connect to a database (including program start-up), but can be unset.

PROMPT1
PROMPT2
PROMPT3
These specify what the prompts psql issues should look like. See “Prompting” on page 761.

QUIET
This variable is equivalent to the command line option -q. It is not very useful in interactive mode.
This variable is equivalent to the command line option `-S`.

**SINGLESTEP**
This variable is equivalent to the command line option `-s`.

**USER**
The database user you are currently connected as. This is set every time you connect to a database (including program start-up), but can be unset.

**VERBOSITY**
This variable can be set to the values `default`, `verbose`, or `terse` to control the verbosity of error reports.

**SQL Interpolation**
An additional useful feature of `psql` variables is that you can substitute (interpolate) them into regular SQL statements. The syntax for this is again to prepend the variable name with a colon (`:`).

```
testdb=> \set foo 'my_table'
```
```
testdb=> SELECT * FROM :foo;
```

would then query the table `my_table`. The value of the variable is copied literally, so it can even contain unbalanced quotes or backslash commands. You must make sure that it makes sense where you put it. Variable interpolation will not be performed into quoted SQL entities.

A popular application of this facility is to refer to the last inserted OID in subsequent statements to build a foreign key scenario. Another possible use of this mechanism is to copy the contents of a file into a table column. First load the file into a variable and then proceed as above.

```
testdb=> \set content '''' `cat my_file.txt` ''''
```
```
testdb=> INSERT INTO my_table VALUES (:content);
```

One problem with this approach is that `my_file.txt` might contain single quotes. These need to be escaped so that they don’t cause a syntax error when the second line is processed. This could be done with the program `sed`:

```
testdb=> \set content '''' `sed -e "s/'/''/g" < my_file.txt` ''''
```

If you are using non-standard-conforming strings then you’ll also need to double backslashes. This is a bit tricky:

```
testdb=> \set content '''' `sed -e "s/'/''/g" -e 's/\//\\/\//g' < my_file.txt` ''''
```

Note the use of different shell quoting conventions so that neither the single quote marks nor the backslashes are special to the shell. Backslashes are still special to `sed`, however, so we need to double them.
Since colons may legally appear in SQL commands, the following rule applies: the character sequence "name" is not changed unless "name" is the name of a variable that is currently set. In any case you can escape a colon with a backslash to protect it from substitution. (The colon syntax for variables is standard SQL for embedded query languages, such as ECPG. The colon syntax for array slices and type casts are Greenplum Database extensions, hence the conflict.)

**Prompting**

The prompts `psql` issues can be customized to your preference. The three variables `PROMPT1`, `PROMPT2`, and `PROMPT3` contain strings and special escape sequences that describe the appearance of the prompt. Prompt 1 is the normal prompt that is issued when `psql` requests a new command. Prompt 2 is issued when more input is expected during command input because the command was not terminated with a semicolon or a quote was not closed. Prompt 3 is issued when you run an SQL `COPY` command and you are expected to type in the row values on the terminal.

The value of the selected prompt variable is printed literally, except where a percent sign (%) is encountered. Depending on the next character, certain other text is substituted instead. Defined substitutions are:

- **%M**
  The full host name (with domain name) of the database server, or `[local]` if the connection is over a UNIX domain socket, or `[local:/dir/name]`, if the UNIX domain socket is not at the compiled in default location.

- **%m**
  The host name of the database server, truncated at the first dot, or `[local]` if the connection is over a UNIX domain socket.

- **%>**
  The port number at which the database server is listening.

- **%n**
  The database session user name. (The expansion of this value might change during a database session as the result of the command `SET SESSION AUTHORIZATION`.)

- **%/**
  The name of the current database.

- **%~**
  Like %/, but the output is ~ (tilde) if the database is your default database.

- **%#**
  If the session user is a database superuser, then a #, otherwise a >. (The expansion of this value might change during a database session as the result of the command `SET SESSION AUTHORIZATION`.)
%R
In prompt 1 normally =, but ^ if in single-line mode, and ! if the session is disconnected from the database (which can happen if \connect fails). In prompt 2 the sequence is replaced by -, *, a single quote, a double quote, or a dollar sign, depending on whether psql expects more input because the command wasn’t terminated yet, because you are inside a /* ... */ comment, or because you are inside a quoted or dollar-escaped string. In prompt 3 the sequence doesn’t produce anything.

%x
Transaction status: an empty string when not in a transaction block, or * when in a transaction block, or ! when in a failed transaction block, or ? when the transaction state is indeterminate (for example, because there is no connection).

%digits
The character with the indicated octal code is substituted.

%:name:
The value of the psql variable name. See “Variables” on page 757 for details.

% `command`
The output of command, similar to ordinary back-tick substitution.

% [ ... %]
Prompts may contain terminal control characters which, for example, change the color, background, or style of the prompt text, or change the title of the terminal window. In order for line editing to work properly, these non-printing control characters must be designated as invisible by surrounding them with %[ and %]. Multiple pairs of these may occur within the prompt. For example,

```
testdb=> \set PROMPT1 '%[033[1;33;40m]n@/%R%[033[0m]%# '
```
results in a boldfaced (1;) yellow-on-black (33;40) prompt on VT100-compatible, color-capable terminals. To insert a percent sign into your prompt, write %%%. The default prompts are '%/%R%# ' for prompts 1 and 2, and '>> ' for prompt 3.

Command-Line Editing
psql supports the NetBSD libedit library for convenient line editing and retrieval. The command history is automatically saved when psql exits and is reloaded when psql starts up. Tab-completion is also supported, although the completion logic makes no claim to be an SQL parser. If for some reason you do not like the tab completion, you can turn it off by putting this in a file named .inputrc in your home directory:

```
$if psql
  set disable-completion on
$endif
```
**Environment**

**PAGER**

If the query results do not fit on the screen, they are piped through this command. Typical values are `more` or `less`. The default is platform-dependent. The use of the pager can be disabled by using the \pset command.

**PGDATABASE**
**PGHOST**
**PGPORT**
**PGUSER**

Default connection parameters.

**PSQL_EDITOR**
**EDITOR**
**VISUAL**

Editor used by the \e command. The variables are examined in the order listed; the first that is set is used.

**SHELL**

Command executed by the \! command.

**TMPDIR**

Directory for storing temporary files. The default is /tmp.

**Files**

Before starting up, `psql` attempts to read and execute commands from the user’s ~/.psqlrc file.

The command-line history is stored in the file ~/.psql_history.

**Notes**

`psql` only works smoothly with servers of the same version. That does not mean other combinations will fail outright, but subtle and not-so-subtle problems might come up. Backslash commands are particularly likely to fail if the server is of a different version.

**Notes for Windows users**

`psql` is built as a console application. Since the Windows console windows use a different encoding than the rest of the system, you must take special care when using 8-bit characters within `psql`. If `psql` detects a problematic console code page, it will warn you at startup. To change the console code page, two things are necessary:

Set the code page by entering `cmd.exe /c chcp 1252` (1252 is a character encoding of the Latin alphabet, used by Microsoft Windows for English and some other Western languages.) If you are using Cygwin, you can put this command in /etc/profile.
Set the console font to Lucida Console, because the raster font does not work with the ANSI code page.

**Examples**

Start `psql` in interactive mode:

```
psql -p 54321 -U sally mydatabase
```

In `psql` interactive mode, spread a command over several lines of input. Notice the changing prompt:

```
testdb=> CREATE TABLE my_table (first integer not null default 0, second text);
CREATE TABLE
```

Look at the table definition:

```
testdb=> \d my_table
    Table "my_table"
    Attribute | Type   | Modifier
              |---------|-------------------
         first | integer | not null default 0
         second | text    |
```

Run `psql` in non-interactive mode by passing in a file containing SQL commands:

```
psql -f /home/gpadmin/test/myscript.sql
```
reindexdb

Rebuilds indexes in a database.

Synopsis

reindexdb [connection-option...] [--table | -t table ] [--index | -i index ] [dbname]
reindexdb [connection-option...] [--all | -a]
reindexdb [connection-option...] [--system | -s] [dbname]
reindexdb --help | --version

Description

reindexdb is a utility for rebuilding indexes in Greenplum Database, and is a wrapper around the SQL command REINDEX.

Options

-a | --all
Reindex all databases.

[-d] dbname | [--dbname] dbname
Specifies the name of the database to be reindexed. If this is not specified and --all is not used, the database name is read from the environment variable PGDATABASE. If that is not set, the user name specified for the connection is used.

-e | --echo
Echo the commands that reindexdb generates and sends to the server.

-i index | --index index
Recreate index only.

-q | --quiet
Do not display a response.

-s | --system
Reindex system catalogs.

-t table | --table table
Reindex table only.
Connection Options

- **h host | --host host**
  Specifies the host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable `PGHOST` or defaults to localhost.

- **p port | --port port**
  Specifies the TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable `PGPORT` or defaults to 5432.

- **U username | --username username**
  The database role name to connect as. If not specified, reads from the environment variable `PGUSER` or defaults to the current system user name.

- **w | --no-password**
  Never issue a password prompt. If the server requires password authentication and a password is not available by other means such as a `.pgpass` file, the connection attempt will fail. This option can be useful in batch jobs and scripts where no user is present to enter a password.

- **W | --password**
  Force a password prompt.

Notes

reindexdb might need to connect several times to the master server, asking for a password each time. It is convenient to have a ~/.pgpass file in such cases.

Examples

To reindex the database *mydb*:

    reindexdb mydb

To reindex the table *foo* and the index *bar* in a database named *abcd*:

    reindexdb --table foo --index bar abcd

See Also

REINDEX
vacuumdb

Garbage-collects and analyzes a database.

Synopsis

```
vacuumdb [connection-option...] [--full | -f] [-F] [--verbose | -v] [--analyze | -z] [--table | -t table [{ column [, ...] }] ] [dbname]
```

```
```

Description

vacuumdb is a utility for cleaning a PostgreSQL database. vacuumdb will also generate internal statistics used by the PostgreSQL query optimizer.

vacuumdb is a wrapper around the SQL command VACUUM. There is no effective difference between vacuuming databases via this utility and via other methods for accessing the server.

Options

```
-a | --all
  Vacuums all databases.
```

```
[-d] dbname | [--dbname] dbname
  The name of the database to vacuum. If this is not specified and --all is not used, the database name is read from the environment variable PGDATABASE. If that is not set, the user name specified for the connection is used.
```

```
-e | --echo
  Echo the commands that reindexdb generates and sends to the server.
```

```
-f | --full
  Selects a full vacuum, which may reclaim more space, but takes much longer and exclusively locks the table.
  Warning: A VACUUM FULL is not recommended in Greenplum Database.
```

```
-F | --freeze
  Freeze row transaction information.
```

```
-q | --quiet
  Do not display a response.
```
-t table [(column)] | --table table [(column)]
Clean or analyze this table only. Column names may be specified only in conjunction with the --analyze option. If you specify columns, you probably have to escape the parentheses from the shell.

-v | --verbose
Print detailed information during processing.

-z | --analyze
Collect statistics for use by the query planner.

Connection Options

-h host | --host host
Specifies the host name of the machine on which the Greenplum master database server is running. If not specified, reads from the environment variable PGHOST or defaults to localhost.

-p port | --port port
Specifies the TCP port on which the Greenplum master database server is listening for connections. If not specified, reads from the environment variable PGPORT or defaults to 5432.

-U username | --username username
The database role name to connect as. If not specified, reads from the environment variable PGUSER or defaults to the current system user name.

-w | --no-password
Never issue a password prompt. If the server requires password authentication and a password is not available by other means such as a .pgpass file, the connection attempt will fail. This option can be useful in batch jobs and scripts where no user is present to enter a password.

-W | --password
Force a password prompt.

Notes
vacuumdb might need to connect several times to the master server, asking for a password each time. It is convenient to have a ~/.pgpass file in such cases.

Examples
To clean the database test:

    vacuumdb test

To clean and analyze a database named bigdb:

    vacuumdb --analyze bigdb
To clean a single table foo in a database named mydb, and analyze a single column bar of the table. Note the quotes around the table and column names to escape the parentheses from the shell:

    vacuumdb --analyze --verbose --table 'foo(bar)' mydb

---

**See Also**

VACUUM, ANALYZE
D. Server Configuration Parameters

There are many configuration parameters that affect the behavior of the Greenplum Database system. Many of these configuration parameters have the same names, settings, and behaviors as in a regular PostgreSQL database system.

Parameter Types and Values

All parameter names are case-insensitive. Every parameter takes a value of one of four types: boolean, integer, floating point, or string. Boolean values may be written as ON, OFF, TRUE, FALSE, YES, NO, 1, 0 (all case-insensitive).

Some settings specify a memory size or time value. Each of these has an implicit unit, which is either kilobytes, blocks (typically eight kilobytes), milliseconds, seconds, or minutes. Valid memory size units are \texttt{kB} (kilobytes), \texttt{MB} (megabytes), and \texttt{GB} (gigabytes). Valid time units are \texttt{ms} (milliseconds), \texttt{s} (seconds), \texttt{min} (minutes), \texttt{h} (hours), and \texttt{d} (days). Note that the multiplier for memory units is 1024, not 1000. When specifying a memory or time unit using the \texttt{SET} command, enclose the value in quotes. For example:

\begin{verbatim}
    SET work_mem TO '200MB';
\end{verbatim}

Setting Parameters

Many of the configuration parameters have limitations on who can change them and where or when they can be set. For example, to change certain parameters, you must be a Greenplum Database superuser. Other parameters require a restart of the system for the changes to take effect. A parameter that is classified as \textit{session} can be set at the system level (in the \texttt{postgresql.conf} file), at the database-level (using \texttt{ALTER DATABASE}), at the role-level (using \texttt{ALTER ROLE}), or at the session-level (using \texttt{SET}). System parameters can only be set in the \texttt{postgresql.conf} file.

In Greenplum Database, the master and each segment instance has its own \texttt{postgresql.conf} file (located in their respective data directories). Some parameters are considered \textit{local} parameters, meaning that each segment instance looks to its own \texttt{postgresql.conf} file to get the value of that parameter. You must set local parameters on every instance in the system (master and segments). Others parameters are considered \textit{master} parameters. Master parameters need only be set at the master instance.

<table>
<thead>
<tr>
<th>Set Classification</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>master or local</td>
<td>A \textit{master} parameter only needs to be set in the \texttt{postgresql.conf} file of the Greenplum master instance. The value for this parameter is then either passed to (or ignored by) the segments at run time. A local parameter must be set in the \texttt{postgresql.conf} file of the master AND each segment instance. Each segment instance looks to its own configuration to get the value for the parameter. Local parameters always requires a system restart for changes to take effect.</td>
</tr>
<tr>
<td>session or system</td>
<td>Session parameters can be changed on the fly within a database session, and can have a hierarchy of settings: at the system level (\texttt{postgresql.conf}), at the database level (\texttt{ALTER DATABASE...SET}), at the role level (\texttt{ALTER ROLE...SET}), or at the session level (\texttt{SET}). If the parameter is set at multiple levels, then the most granular setting takes precedence (for example, session overrides role, role overrides database, and database overrides system). A \textit{system} parameter can only be changed via the \texttt{postgresql.conf} file(s).</td>
</tr>
<tr>
<td>restart or reload</td>
<td>When changing parameter values in the \texttt{postgresql.conf} file(s), some require a \textit{restart} of Greenplum Database for the change to take effect. Other parameter values can be refreshed by just reloading the server configuration file (using \texttt{gpstop -u}), and do not require stopping the system.</td>
</tr>
</tbody>
</table>
These session parameters can only be set by a database superuser. Regular database users cannot set this parameter.

These parameters are not settable by database users or superusers. The current value of the parameter can be shown but not altered.
### Table D.2 Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>add_missing_from</td>
<td>boolean</td>
<td>off</td>
<td>Automatically adds missing table references to FROM clauses. Present for compatibility with releases of PostgreSQL prior to 8.1, where this behavior was allowed by default.</td>
<td>master session reload</td>
</tr>
<tr>
<td>array_nulls</td>
<td>boolean</td>
<td>on</td>
<td>This controls whether the array input parser recognizes unquoted NULL as specifying a null array element. By default, this is on, allowing array values containing null values to be entered. Greenplum Database versions before 3.0 did not support null values in arrays, and therefore would treat NULL as specifying a normal array element with the string value 'NULL'.</td>
<td>master session reload</td>
</tr>
<tr>
<td>authentication_timeout</td>
<td>number of seconds</td>
<td>1min</td>
<td>Maximum time to complete client authentication. This prevents hung clients from occupying a connection indefinitely.</td>
<td>local system restart</td>
</tr>
<tr>
<td>backslash_quote</td>
<td>on (allow \ always)</td>
<td>safe_encoding</td>
<td>This controls whether a quote mark can be represented by \ in a string literal. The preferred, SQL-standard way to represent a quote mark is by doubling it (&quot;&quot;&quot;) but PostgreSQL has historically also accepted .. However, use of \ creates security risks because in some client character set encodings, there are multibyte characters in which the last byte is numerically equivalent to ASCII ..</td>
<td>master session reload</td>
</tr>
<tr>
<td>block_size</td>
<td>number of bytes</td>
<td>32768</td>
<td>Reports the size of a disk block.</td>
<td>read only</td>
</tr>
<tr>
<td>bonjour_name</td>
<td>string</td>
<td>unset</td>
<td>Specifies the Bonjour broadcast name. By default, the computer name is used, specified as an empty string. This option is ignored if the server was not compiled with Bonjour support.</td>
<td>master system restart</td>
</tr>
<tr>
<td>check_function_bodies</td>
<td>boolean</td>
<td>on</td>
<td>When set to off, disables validation of the function body string during CREATE FUNCTION. Disabling validation is occasionally useful to avoid problems such as forward references when restoring function definitions from a dump.</td>
<td>master session reload</td>
</tr>
<tr>
<td>client_encoding</td>
<td>character set</td>
<td>UTF8</td>
<td>Sets the client-side encoding (character set). The default is to use the same as the database encoding. See Supported Character Sets in the PostgreSQL documentation.</td>
<td>master session reload</td>
</tr>
<tr>
<td>Parameter</td>
<td>Value Range</td>
<td>Default</td>
<td>Description</td>
<td>Set Classifications</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------------------</td>
<td>---------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>client_min_messages</td>
<td>DEBUG5, DEBUG4, DEBUG3, DEBUG2, DEBUG1, LOG, NOTICE, WARNING, ERROR, FATAL, PANIC</td>
<td>NOTICE</td>
<td>Controls which message levels are sent to the client. Each level includes all the levels that follow it. The later the level, the fewer messages are sent.</td>
<td>master session reload</td>
</tr>
<tr>
<td>cpu_index_tuple_cost</td>
<td>floating point</td>
<td>0.005</td>
<td>Sets the planner’s estimate of the cost of processing each index row during an index scan. This is measured as a fraction of the cost of a sequential page fetch.</td>
<td>master session reload</td>
</tr>
<tr>
<td>cpu_operator_cost</td>
<td>floating point</td>
<td>0.0025</td>
<td>Sets the planner’s estimate of the cost of processing each operator in a WHERE clause. This is measured as a fraction of the cost of a sequential page fetch.</td>
<td>master session reload</td>
</tr>
<tr>
<td>cpu_tuple_cost</td>
<td>floating point</td>
<td>0.01</td>
<td>Sets the planner’s estimate of the cost of processing each row during a query. This is measured as a fraction of the cost of a sequential page fetch.</td>
<td>master session reload</td>
</tr>
<tr>
<td>cursor_tuple_fraction</td>
<td>integer</td>
<td>1</td>
<td>Tells the query planner how many rows are expected to be fetched in a cursor query, thereby allowing the planner to use this information to optimize the query plan. The default of 1 means all rows will be fetched.</td>
<td>master session reload</td>
</tr>
<tr>
<td>custom_variable_classes</td>
<td>comma-separated list of class names</td>
<td>unset</td>
<td>Specifies one or several class names to be used for custom variables. A custom variable is a variable not normally known to the server but used by some add-on module. Such variables must have names consisting of a class name, a dot, and a variable name.</td>
<td>local system restart</td>
</tr>
<tr>
<td>DateStyle</td>
<td>&lt;format&gt;, &lt;date style&gt;</td>
<td>ISO, MDY</td>
<td>Sets the display format for date and time values, as well as the rules for interpreting ambiguous date input values. This variable contains two independent components: the output format specification and the input/output specification for year/month/day ordering.</td>
<td>master session reload</td>
</tr>
<tr>
<td>db_user_namespace</td>
<td>boolean</td>
<td>off</td>
<td>This enables per-database user names. If on, you should create users as username@dbname. To create ordinary global users, simply append @ when specifying the user name in the client.</td>
<td>local system restart</td>
</tr>
</tbody>
</table>
### Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>deadlock_timeout</td>
<td>number of milliseconds</td>
<td>1s</td>
<td>The number of milliseconds to wait on a lock before checking to see if there is a deadlock condition. On a heavily loaded server you might want to raise this value. Ideally the setting should exceed your typical transaction time, so as to improve the odds that a lock will be released before the waiter decides to check for deadlock.</td>
<td>local system restart</td>
</tr>
<tr>
<td>debug_assertions</td>
<td>boolean</td>
<td>off</td>
<td>Turns on various assertion checks.</td>
<td>local system restart</td>
</tr>
<tr>
<td>debug_pretty_print</td>
<td>boolean</td>
<td>off</td>
<td>Indents debug output to produce a more readable but much longer output format. client_min_messages or log_min_messages must be DEBUG1 or lower.</td>
<td>master session reload</td>
</tr>
<tr>
<td>debug_print_parse</td>
<td>boolean</td>
<td>off</td>
<td>For each executed query, prints the resulting parse tree. client_min_messages or log_min_messages must be DEBUG1 or lower.</td>
<td>master session reload</td>
</tr>
<tr>
<td>debug_print_plan</td>
<td>boolean</td>
<td>off</td>
<td>For each executed query, prints the Greenplum parallel query execution plan. client_min_messages or log_min_messages must be DEBUG1 or lower.</td>
<td>master session reload</td>
</tr>
<tr>
<td>debug_print_prelim_plan</td>
<td>boolean</td>
<td>off</td>
<td>For each executed query, prints the preliminary query plan. client_min_messages or log_min_messages must be DEBUG1 or lower.</td>
<td>master session reload</td>
</tr>
<tr>
<td>debug_print_rewritten</td>
<td>boolean</td>
<td>off</td>
<td>For each executed query, prints the query rewriter output. client_min_messages or log_min_messages must be DEBUG1 or lower.</td>
<td>master session reload</td>
</tr>
<tr>
<td>debug_print_slice_table</td>
<td>boolean</td>
<td>off</td>
<td>For each executed query, prints the Greenplum query slice plan. client_min_messages or log_min_messages must be DEBUG1 or lower.</td>
<td>master session reload</td>
</tr>
<tr>
<td>default_statistics_target</td>
<td>integer &gt; 0</td>
<td>25</td>
<td>Sets the default statistics target for table columns that have not had a column-specific target set via ALTER TABLE SET STATISTICS. Larger values increase the time needed to do ANALYZE, but may improve the quality of the planner’s estimates.</td>
<td>master session reload</td>
</tr>
<tr>
<td>default_tablespace</td>
<td>name of a tablespace</td>
<td>unset</td>
<td>The default tablespace in which to create objects (tables and indexes) when a CREATE command does not explicitly specify a tablespace.</td>
<td>master session reload</td>
</tr>
<tr>
<td>Parameter</td>
<td>Value Range</td>
<td>Default</td>
<td>Description</td>
<td>Set Classifications</td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------------------------------------------------</td>
<td>---------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>default_transaction_isolation</td>
<td>read committed, read uncommitted, repeatable, read serializable</td>
<td>read committed</td>
<td>Controls the default isolation level of each new transaction.</td>
<td>master session reload</td>
</tr>
<tr>
<td>default_transaction_read_only</td>
<td>boolean</td>
<td>off</td>
<td>Controls the default read-only status of each new transaction. A read-only SQL transaction cannot alter non-temporary tables.</td>
<td>master session reload</td>
</tr>
<tr>
<td>dynamic_library_path</td>
<td>a list of absolute directory paths separated by colons</td>
<td>$libdir</td>
<td>If a dynamically loadable module needs to be opened and the file name specified in the CREATE FUNCTION or LOAD command does not have a directory component (i.e. the name does not contain a slash), the system will search this path for the required file. The compiled-in PostgreSQL package library directory is substituted for $libdir. This is where the modules provided by the standard PostgreSQL distribution are installed.</td>
<td>local system restart</td>
</tr>
<tr>
<td>effective_cache_size</td>
<td>floating point</td>
<td>512MB</td>
<td>Sets the planner’s assumption about the effective size of the disk cache that is available to a single query. This is factored into estimates of the cost of using an index; a higher value makes it more likely index scans will be used, a lower value makes it more likely sequential scans will be used. This parameter has no effect on the size of shared memory allocated by a Greenplum server instance, nor does it reserve kernel disk cache; it is used only for estimation purposes.</td>
<td>master session reload</td>
</tr>
<tr>
<td>enable_bitmapscan</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables the query planner’s use of bitmap-scan plan types. Note that this is different than a Bitmap Index Scan. A Bitmap Scan means that indexes will be dynamically converted to bitmaps in memory when appropriate, giving faster index performance on complex queries against very large tables. It is used when there are multiple predicates on different indexed columns. Each bitmap per column can be compared to create a final list of selected tuples.</td>
<td>master session reload</td>
</tr>
<tr>
<td>enable_groupagg</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables the query planner’s use of group aggregation plan types.</td>
<td>master session reload</td>
</tr>
<tr>
<td>enable_hashagg</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables the query planner’s use of hash aggregation plan types.</td>
<td>master session reload</td>
</tr>
<tr>
<td>enable_hashjoin</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables the query planner’s use of hash-join plan types.</td>
<td>master session reload</td>
</tr>
<tr>
<td>Parameter</td>
<td>Value Range</td>
<td>Default</td>
<td>Description</td>
<td>Set Classifications</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------</td>
<td>---------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>enable_indexscan</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables the query planner’s use of index-scan plan types.</td>
<td>master session reload</td>
</tr>
<tr>
<td>enable_mergejoin</td>
<td>boolean</td>
<td>off</td>
<td>Enables or disables the query planner’s use of merge-join plan types.</td>
<td>master session reload</td>
</tr>
<tr>
<td>enable_nestloop</td>
<td>boolean</td>
<td>off</td>
<td>Enables or disables the query planner’s use of nested-loop join plans.</td>
<td>master session reload</td>
</tr>
<tr>
<td>enable_seqscan</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables the query planner’s use of sequential scan plan types.</td>
<td>master session reload</td>
</tr>
<tr>
<td>enable_sort</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables the query planner’s use of explicit sort steps.</td>
<td>master session reload</td>
</tr>
<tr>
<td>enable_tidscan</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables the query planner’s use of tuple identifier (TID) scan plan types.</td>
<td>master session reload</td>
</tr>
<tr>
<td>escape_string_warning</td>
<td>boolean</td>
<td>on</td>
<td>When on, a warning is issued if a backslash () appears in an ordinary string literal (&quot;...&quot;) syntax. Escape string syntax (E&quot;...&quot;') should be used for escapes, because in future versions, ordinary strings will have the SQL standard-conforming behavior of treating backslashes literally.</td>
<td>master session reload</td>
</tr>
<tr>
<td>explain.pretty_print</td>
<td>boolean</td>
<td>on</td>
<td>Determines whether EXPLAIN VERBOSE uses the indented or non-indented format for displaying detailed query-tree dumps.</td>
<td>master session reload</td>
</tr>
</tbody>
</table>
### Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>extra_float_digits</td>
<td>integer</td>
<td>0</td>
<td>Adjusts the number of digits displayed for floating-point values, including float4, float8, and geometric data types. The parameter value is added to the standard number of digits. The value can be set as high as 2, to include partially-significant digits; this is especially useful for dumping float data that needs to be restored exactly. Or it can be set negative to suppress unwanted digits.</td>
<td>master session reload</td>
</tr>
<tr>
<td>from_collapse_limit</td>
<td>1-n</td>
<td>16</td>
<td>The planner will merge sub-queries into upper queries if the resulting FROM list would have no more than this many items. Smaller values reduce planning time but may yield inferior query plans.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_adjust_selectivity_for_outerjoins</td>
<td>boolean</td>
<td>on</td>
<td>Enables the selectivity of NULL tests over outer joins.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_analyze_relative_error</td>
<td>floating point &lt; 1.0</td>
<td>0.25</td>
<td>Sets the estimated acceptable error in the cardinality of the table — a value of 0.5 is supposed to be equivalent to an acceptable error of 50% (this is the default value used in PostgreSQL). If the statistics collected during ANALYZE are not producing good estimates of cardinality for a particular table attribute, decreasing the relative error fraction (accepting less error) tells the system to sample more rows.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_autostats_mode</td>
<td>none, on_change, on_no_stats</td>
<td>on_no_stats</td>
<td>Specifies the mode for triggering automatic statistics collection with ANALYZE. The on_no_stats option triggers statistics collection for CREATE TABLE AS SELECT, INSERT, or COPY operations on any table that has no existing statistics. The on_change option triggers statistics collection only when the number of rows affected meets or exceeds the threshold defined by gp_autostats_on_change_threshold. Operations that can trigger automatic statistics collection with on_change are: CREATE TABLE AS SELECT UPDATE DELETE INSERT COPY. Default is on_no_stats.</td>
<td>master session reload</td>
</tr>
</tbody>
</table>
### Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp_autostats_on_change_threshold</td>
<td>integer</td>
<td>2147483647</td>
<td>Specifies the threshold for automatic statistics collection when gp_autostats_mode is set to on_change. When a triggering table operation affects a number of rows exceeding this threshold, <strong>ANALYZE</strong> is added and statistics are collected for the table.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_cached_segworkers_threshold</td>
<td>integer &gt; 0</td>
<td>5</td>
<td>When a user starts a session with Greenplum Database and issues a query, the system creates groups or ‘gangs’ of worker processes on each segment to do the work. After the work is done, the segment worker processes are destroyed except for a cached number which is set by this parameter. A lower setting conserves system resources on the segment hosts, but a higher setting may improve performance for power-users that want to issue many complex queries in a row.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_command_count</td>
<td>integer &gt; 0</td>
<td>1</td>
<td>Shows how many commands the master has received from the client. Note that a single SQL command might actually involve more than one command internally, so the counter may increment by more than one for a single query. This counter also is shared by all of the segment processes working on the command.</td>
<td>read only</td>
</tr>
<tr>
<td>gp_connections_per_thread</td>
<td>integer</td>
<td>512</td>
<td>A value larger than or equal to the number of primary segments means that each slice in a query plan will get its own thread when dispatching to the segments. A value of 0 indicates that the dispatcher should use a single thread when dispatching all query plan slices to a segment. Lower values will use more threads, which utilizes more resources on the master. Typically, the default does not need to be changed unless there is a known throughput performance problem.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_content</td>
<td>integer</td>
<td></td>
<td>The local content id if a segment.</td>
<td>read only</td>
</tr>
<tr>
<td>gp_dbid</td>
<td>integer</td>
<td></td>
<td>The local content dbid if a segment.</td>
<td>read only</td>
</tr>
<tr>
<td>gp_debug_linger</td>
<td>number of seconds</td>
<td>0</td>
<td>Number of seconds for a Greenplum process to linger after a fatal internal error.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_email_from</td>
<td>string</td>
<td></td>
<td>The email address used to send email alerts, in the format of: '<a href="mailto:username@domain.com">username@domain.com</a>' or 'Name <a href="mailto:username@domain.com">username@domain.com</a>'</td>
<td>master system restart</td>
</tr>
</tbody>
</table>
### Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp_email_smtp_password</td>
<td>string</td>
<td></td>
<td>The password/passphrase used to authenticate with the SMTP server.</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_email_smtp_server</td>
<td>string</td>
<td></td>
<td>The fully qualified domain name or IP address and port of the SMTP server to use to send the email alerts. Must be in the format of: smtp_servername.domain.com:port</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_email_smtp_userid</td>
<td>string</td>
<td></td>
<td>The user id used to authenticate with the SMTP server.</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_email_to</td>
<td>string</td>
<td></td>
<td>A semi-colon (;) separated list of email addresses to receive email alert messages to in the format of: '<a href="mailto:username@domain.com">username@domain.com</a>' or 'Name <a href="mailto:username@domain.com">username@domain.com</a>' If this parameter is not set, then email alerts are disabled.</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_enable_adaptive_nestloop</td>
<td>boolean</td>
<td>on</td>
<td>Enables the query planner to use a new type of join node called &quot;Adaptive Nestloop&quot; at query execution time. This causes the planner to favor a hash-join over a nested-loop join if the number of rows on the outer side of the join exceeds a precalculated threshold. This parameter improves performance of index operations, which previously favored slower nested-loop joins.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_enable_agg_distinct</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables two-phase aggregation to compute a single distinct-qualified aggregate. This applies only to subqueries that include a single distinct-qualified aggregate function.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_enable_agg_distinct_pruning</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables three-phase aggregation and join to compute distinct-qualified aggregates. This applies only to subqueries that include one or more distinct-qualified aggregate functions.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_enable_direct_dispatch</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables the dispatching of targeted query plans for queries that access data on a single segment. When on, queries that target rows on a single segment will only have their query plan dispatched to that segment (rather than to all segments). This significantly reduces the response time of qualifying queries as there is no interconnect setup involved. Direct dispatch does require more CPU utilization on the master.</td>
<td>master system restart</td>
</tr>
</tbody>
</table>
### Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp_enable_fallback_plan</td>
<td>boolean</td>
<td>on</td>
<td>Allows use of disabled plan types when a query would not be feasible without them.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_enable_fast_sri</td>
<td>boolean</td>
<td>on</td>
<td>When set to on, the query planner plans single row inserts so that they are sent directly to the correct segment instance (no motion operation required). This significantly improves performance of single-row-insert statements.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_enable_gpperfmon</td>
<td>boolean</td>
<td>off</td>
<td>Enables or disables the data collection agents of Greenplum Performance Monitor.</td>
<td>local system restart</td>
</tr>
<tr>
<td>gp_enable_grouptext_distinct_gather</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables gathering data to a single node to compute distinct-qualified aggregates on grouping extension queries. When this parameter and gp_enable_grouptext_distinct_pruning are both enabled, the planner uses the cheaper plan.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_enable_grouptext_distinct_pruning</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables three-phase aggregation and join to compute distinct-qualified aggregates on grouping extension queries. Usually, enabling this parameter generates a cheaper query plan that the planner will use in preference to existing plan.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_enable_multiphase_agg</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables the query planner’s use of two or three-stage parallel aggregation plans. This approach applies to any subquery with aggregation. If gp_enable_multiphase_agg is off, then gp_enable_agg_distinct and gp_enable_agg_distinct_pruning are disabled.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_enable_predicate_propagation</td>
<td>boolean</td>
<td>on</td>
<td>When enabled, the query planner applies query predicates to both table expressions in cases where the tables are joined on their distribution key column(s). Filtering both tables prior to doing the join (when possible) is more efficient.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_enable_preunique</td>
<td>boolean</td>
<td>on</td>
<td>Enables two-phase duplicate removal for SELECT DISTINCT queries (not SELECT COUNT(DISTINCT)). When enabled, it adds an extra SORT DISTINCT set of plan nodes before motioning. In cases where the distinct operation greatly reduces the number of rows, this extra SORT DISTINCT is much cheaper than the cost of sending the rows across the Interconnect.</td>
<td>master session reload</td>
</tr>
</tbody>
</table>
Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp_enable_sequential_window_plans</td>
<td>boolean</td>
<td>on</td>
<td>If on, enables non-parallel (sequential) query plans for queries containing window function calls. If off, evaluates compatible window functions in parallel and rejoins the results. This is an experimental parameter.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_enable_sort_distinct</td>
<td>boolean</td>
<td>on</td>
<td>Enable duplicates to be removed while sorting.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_enable_sort_limit</td>
<td>boolean</td>
<td>on</td>
<td>Enable LIMIT operation to be performed while sorting. Sorts more efficiently when the plan requires the first <code>limit_number</code> of rows at most.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_external_enable_exec</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables the use of external tables that execute OS commands or scripts on the segment hosts (CREATE EXTERNAL TABLE EXECUTE syntax). Must be enabled if using the Performance Monitor or MapReduce features.</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_external_grant_privileges</td>
<td>boolean</td>
<td>off</td>
<td>In releases prior to 4.0, enables or disables non-superusers to issue a CREATE EXTERNAL [WEB] TABLE command in cases where the LOCATION clause specifies http or gpfdist. In releases after 4.0, the ability to create an external table can be granted to a role using CREATE ROLE or ALTER ROLE.</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_external_max_segs</td>
<td>integer</td>
<td>64</td>
<td>Sets the number of segments that will scan external table data during an external table operation, the purpose being not to overload the system with scanning data and take away resources from other concurrent operations. This only applies to external tables that use the gpfdist:// protocol to access external table data.</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_fts_probe_interval</td>
<td>10 seconds or greater</td>
<td>1min</td>
<td>Specifies the polling interval for the fault detection process (ftsprobe). The ftsprobe process will take approximately this amount of time to detect a segment failure.</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_fts_probe_threadcount</td>
<td>1 - 128</td>
<td>5</td>
<td>Specifies the number of ftsprobe threads to create. This parameter should be set to a value equal to or greater than the number of segments per host.</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_fts_probe_timeout</td>
<td>10 seconds or greater</td>
<td>3min</td>
<td>Specifies the allowed timeout for the fault detection process (ftsprobe) to establish a connection to a segment before declaring it down.</td>
<td>master system restart</td>
</tr>
</tbody>
</table>
### Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp_gpperfmon_send_interval</td>
<td>number of seconds</td>
<td>1</td>
<td>Sets the frequency that the Greenplum Database server processes send query execution updates to the Performance Monitor agent processes. Query operations (iterators) executed during this interval are sent through UDP to the segment monitor agents. If you find that an excessive number of UDP packets are dropped during long-running, complex queries, you may consider increasing this value.</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_hashjoin_tuples_per_bucket</td>
<td>integer</td>
<td>5</td>
<td>Sets the target density of the hash table used by HashJoin operations. A smaller value will tend to produce larger hash tables, which can increase join performance.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_interconnect_hash_multiplier</td>
<td>2-25</td>
<td>2</td>
<td>Sets the size of the hash table used by the UDP interconnect to track connections. This number is multiplied by the number of segments to determine the number of buckets in the hash table. Increasing the value may increase interconnect performance for complex multi-slice queries (while consuming slightly more memory on the segment hosts).</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_interconnect_queue_depth</td>
<td>1-2048</td>
<td>4</td>
<td>Sets the amount of data per-peer to be queued by the UDP interconnect on receivers (when data is received but no space is available to receive it the data will be dropped, and the transmitter will need to resend it). Increasing the depth from its default value will cause the system to use more memory; but may increase performance. It is reasonable for this to be set between 1 and 10. Queries with data skew potentially perform better when this is increased. Increasing this may radically increase the amount of memory used by the system.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_interconnect_setup_timeout</td>
<td>0-n seconds</td>
<td>20s</td>
<td>Number of seconds to wait for the Interconnect to complete setup before it times out.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_interconnect_type</td>
<td>TCP, UDP</td>
<td>UDP</td>
<td>Sets the networking protocol used for Interconnect traffic. With the TCP protocol, Greenplum Database has an upper limit of 1000 segment instances - less than that if the query workload involves complex, multi-slice queries. UDP allows for greater interconnect scalability. Note that the Greenplum software does the additional packet verification and checking not performed by UDP, so reliability and performance is equivalent to TCP.</td>
<td>master session reload</td>
</tr>
</tbody>
</table>
Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp_log_format</td>
<td>csv, text</td>
<td>csv</td>
<td>Specifies the format of the server log files. If using gp_toolkit administrative schema, the log files must be in csv format.</td>
<td>local system restart</td>
</tr>
<tr>
<td>gp_max_csv_line_length</td>
<td>number of bytes</td>
<td>1048576</td>
<td>The maximum length of a line in a CSV formatted file that will be imported into the system. The default is 1MB (1048576 bytes). Maximum allowed is 4MB (4194304 bytes). The default may need to be increased if using the gp_toolkit administrative schema to read Greenplum Database log files.</td>
<td>local system restart</td>
</tr>
<tr>
<td>gp_max_databases</td>
<td>integer</td>
<td>16</td>
<td>The maximum number of databases allowed in a Greenplum Database system.</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_max_filespaces</td>
<td>integer</td>
<td>8</td>
<td>The maximum number of filespaces allowed in a Greenplum Database system.</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_max_local_distributed_cache</td>
<td>integer</td>
<td>1024</td>
<td>Sets the number of local to distributed transactions to cache. Higher settings may improve performance.</td>
<td>local system restart</td>
</tr>
<tr>
<td>gp_max_packet_size</td>
<td>512-65536</td>
<td>8192</td>
<td>Sets the size (in bytes) of messages sent by the UDP interconnect, and sets the tuple-serialization chunk size for both the UDP and TCP interconnect.</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_max_tablespaces</td>
<td>integer</td>
<td>16</td>
<td>The maximum number of tablespaces allowed in a Greenplum Database system.</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_motion_cost_per_row</td>
<td>floating point</td>
<td>0</td>
<td>Sets the query planner cost estimate for a Motion operator to transfer a row from one segment to another, measured as a fraction of the cost of a sequential page fetch. If 0, then the value used is two times the value of cpu_tuple_cost.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_num_contents_in_cluster</td>
<td>-</td>
<td>-</td>
<td>The number of primary segments in the Greenplum Database system.</td>
<td>read only</td>
</tr>
<tr>
<td>gp_reject_percent_threshold</td>
<td>1-n</td>
<td>300</td>
<td>For single row error handling on COPY and external table SELECTs, sets the number of rows processed before SEGMENT REJECT LIMIT n PERCENT starts calculating.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_reraise_signal</td>
<td>boolean</td>
<td>on</td>
<td>If enabled, will attempt to dump core if a fatal server error occurs.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_resqueue_priority</td>
<td>boolean</td>
<td>on</td>
<td>Enables or disables query prioritization. When this parameter is disabled, existing priority settings are not evaluated at query run time.</td>
<td>local system restart</td>
</tr>
</tbody>
</table>
### Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
</table>
| gp_resqueue_priority_cpucores_per_segment     | 0.1 - 25.0           | segments = 4  
master = 24 | Specifies the number of CPU units per segment. In a configuration where one segment is configured per CPU core on a host, this unit is 1.0 (default). If an 8-core host is configured with four segments, the value would be 2.0. A master host typically only has one segment running on it (the master instance), so the value for the master should reflect the usage of all available CPU cores. Incorrect settings can result in CPU under-utilization. The default values are appropriate for the Greenplum Data Computing Appliance. | local system restart     |
| gp_resqueue_priority_sweeper_interval         | 500 - 15000 ms       | 1000    | Specifies the interval at which the sweeper process evaluates current CPU usage. When a new statement becomes active, its priority is evaluated and its CPU share determined when the next interval is reached.                                                                 | local system restart     |
| gp_role                                        | dispatch  
execute  
utility |                     | The role of this server process — set to dispatch for the master and execute for a segment.                                                                                                                                     | read only                |
| gp_safefwritesize                              | integer              | 0       | Specifies a minimum size for safe write operations to append-only tables in a non-mature file system. When a number of bytes greater than zero is specified, the append-only writer adds padding data up to that number in order to prevent data corruption due to file system errors. Each non-mature file system has a known safe write size that must be specified here when using Greenplum Database with that type of file system. This is commonly set to a multiple of the extent size of the file system; for example, Linux ext3 is 4096 bytes, so a value of 32768 is commonly used. | local system restart     |
| gp_segment_connect_timeout                     | 0-n seconds          | 1min    | Time that the Greenplum interconnect will try to connect to a segment instance over the network before timing out. Controls the network connection timeout between master and primary segments, and primary to mirror segment replication processes. | local system reload      |
| gp_segments_for_planner                        | 0-n                  | 0       | Sets the number of primary segment instances for the planner to assume in its cost and size estimates. If 0, then the value used is the actual number of primary segments. This variable affects the planner’s estimates of the number of rows handled by each sending and receiving process in Motion operators. | master session reload    |
### Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp_session_id</td>
<td>1-n</td>
<td></td>
<td>A system assigned ID number for a client session. Starts counting from 1 when the master instance is first started.</td>
<td>read only</td>
</tr>
<tr>
<td>gp_set_proc_affinity</td>
<td>boolean</td>
<td>off</td>
<td>If enabled, when a Greenplum server process (postmaster) is started it will bind to a CPU.</td>
<td>master system restart</td>
</tr>
<tr>
<td>gp_set_read_only</td>
<td>boolean</td>
<td>off</td>
<td>Set to on to disable writes to the database. Any in progress transactions must finish before read-only mode takes affect.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_snmp_community</td>
<td>SNMP community name</td>
<td>public</td>
<td>Set to the community name you specified for your environment.</td>
<td>master system reload</td>
</tr>
<tr>
<td>gp_snmp_monitor_address</td>
<td>hostname:port</td>
<td></td>
<td>The hostname:port of your network monitor application. Typically, the port number is 162. If there are multiple monitor addresses, separate them with a comma.</td>
<td>master system reload</td>
</tr>
<tr>
<td>gp_snmp_use_inform_or_trap</td>
<td>inform, trap</td>
<td>trap</td>
<td>Trap notifications are SNMP messages sent from one application to another (for example, between Greenplum Database and a network monitoring application). These messages are unacknowledged by the monitoring application, but generate less network overhead. Inform notifications are the same as trap messages, except that the application sends an acknowledgement to the application that generated the alert.</td>
<td>master system reload</td>
</tr>
<tr>
<td>gp_statistics_pullup_from_child_partition</td>
<td>boolean</td>
<td>on</td>
<td>Enables the query planner to utilize statistics from child tables when planning queries on the parent table.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_statistics_use_fkeys</td>
<td>boolean</td>
<td>off</td>
<td>When enabled, allows the optimizer to use foreign key information stored in the system catalog to optimize joins between foreign keys and primary keys.</td>
<td>master session reload</td>
</tr>
<tr>
<td>gp_vmem_idle_resource_timeout</td>
<td>time in seconds</td>
<td>18s</td>
<td>If a database session is idle for longer than the time specified, the session will free system resources (such as shared memory), but remain connected to the database. This allows more concurrent connections to the database at one time.</td>
<td>master system restart</td>
</tr>
</tbody>
</table>
### gp_vmem_protect_limit

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
</table>
| gp_vmem_protect_limit | integer | 8192 | Sets the amount of memory (in number of MBs) that all postgres processes of an active segment instance can consume. To prevent over allocation of memory, set to: <br>\[
\frac{(X \times \text{physical_memory})}{\text{primary_segments}}
\] Where \( X \) is a value between 1.0 and 1.5. \( X=1 \) offers the best system performance. \( X=1.5 \) may cause more swapping on the system, but less queries will be cancelled. For example, on a segment host with 16GB physical memory and 4 primary segment instances the calculation would be: <br>\[
\frac{(1 \times 16)}{4} = 4GB
\] <br>\[
4 \times 1024 = 4096MB
\] If a query causes this limit to be exceeded, memory will not be allocated and the query will fail. Note that this is a local parameter and must be set for every segment in the system (primary and mirrors). | local system restart |

### gp_vmem_protect_segworker_cache_limit

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp_vmem_protect_segworker_cache_limit</td>
<td>number of megabytes</td>
<td>500</td>
<td>If a query executor process consumes more than this configured amount, then the process will not be cached for use in subsequent queries after the process completes. Systems with lots of connections or idle processes may want to reduce this number to free more memory on the segments. Note that this is a local parameter and must be set for every segment.</td>
<td>local system restart</td>
</tr>
</tbody>
</table>

### gp_workfile_compress_algorithm

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp_workfile_compress_algorithm</td>
<td>none zlib</td>
<td>none</td>
<td>When a hash aggregation or hash join operation spills to disk during query processing, specifies the compression algorithm to use on the spill files. If using zlib, it must be in your $PATH on all segments.</td>
<td>master session reload</td>
</tr>
</tbody>
</table>

### gpperfmon_port

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>gpperfmon_port</td>
<td>integer</td>
<td>8888</td>
<td>Sets the port on which all performance monitor agents communicate with the master.</td>
<td>master system restart</td>
</tr>
</tbody>
</table>

### integer_datetimes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>integer_datetimes</td>
<td>boolean</td>
<td>on</td>
<td>Reports whether PostgreSQL was built with support for 64-bit-integer dates and times.</td>
<td>read only</td>
</tr>
</tbody>
</table>
### Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>IntervalStyle</td>
<td>postgres</td>
<td>postgres</td>
<td>Sets the display format for interval values. The value <code>sql_standard</code> produces output matching SQL standard interval literals. The value <code>postgres</code> produces output matching PostgreSQL releases prior to 8.4 when the <code>DateStyle</code> parameter was set to ISO. The value <code>postgres_verbose</code> produces output matching Greenplum releases prior to 3.3 when the <code>DateStyle</code> parameter was set to non-ISO output. The value <code>iso_8601</code> will produce output matching the time interval format with designators defined in section 4.4.3.2 of ISO 8601. See the PostgreSQL 8.4 documentation for more information.</td>
<td>master session reload</td>
</tr>
<tr>
<td>joinCollapseLimit</td>
<td>1-(n)</td>
<td>16</td>
<td>The planner will rewrite explicit inner <code>JOIN</code> constructs into lists of <code>FROM</code> items whenever a list of no more than this many items in total would result. By default, this variable is set the same as <code>fromCollapseLimit</code>, which is appropriate for most uses. Setting it to 1 prevents any reordering of inner JOINs. Setting this variable to a value between 1 and <code>fromCollapseLimit</code> might be useful to trade off planning time against the quality of the chosen plan (higher values produce better plans).</td>
<td>master session reload</td>
</tr>
<tr>
<td>krbCaseinsUsers</td>
<td>boolean</td>
<td>off</td>
<td>Sets whether Kerberos user names should be treated case-insensitively. The default is case sensitive (off).</td>
<td>master system restart</td>
</tr>
<tr>
<td>krbServerKeyfile</td>
<td>path and file name</td>
<td>unset</td>
<td>Sets the location of the Kerberos server key file.</td>
<td>master system restart</td>
</tr>
<tr>
<td>krbSrvname</td>
<td>service name</td>
<td>postgres</td>
<td>Sets the Kerberos service name.</td>
<td>master system restart</td>
</tr>
<tr>
<td>lcCollate</td>
<td><code>&lt;system dependent&gt;</code></td>
<td></td>
<td>Reports the locale in which sorting of textual data is done. The value is determined when the Greenplum Database array is initialized.</td>
<td>read only</td>
</tr>
<tr>
<td>lcCtype</td>
<td><code>&lt;system dependent&gt;</code></td>
<td></td>
<td>Reports the locale that determines character classifications. The value is determined when the Greenplum Database array is initialized.</td>
<td>read only</td>
</tr>
</tbody>
</table>
## Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>lc_messages</td>
<td>&lt;system dependent&gt;</td>
<td></td>
<td>Sets the language in which messages are displayed. The locales available depends on what was installed with your operating system - use <code>locale -a</code> to list available locales. The default value is inherited from the execution environment of the server. On some systems, this locale category does not exist. Setting this variable will still work, but there will be no effect. Also, there is a chance that no translated messages for the desired language exist. In that case you will continue to see the English messages.</td>
<td>local system restart</td>
</tr>
<tr>
<td>lc_monetary</td>
<td>&lt;system dependent&gt;</td>
<td></td>
<td>Sets the locale to use for formatting monetary amounts, for example with the <code>to_char</code> family of functions. The locales available depends on what was installed with your operating system - use <code>locale -a</code> to list available locales. The default value is inherited from the execution environment of the server.</td>
<td>local system restart</td>
</tr>
<tr>
<td>lc_numeric</td>
<td>&lt;system dependent&gt;</td>
<td></td>
<td>Sets the locale to use for formatting numbers, for example with the <code>to_char</code> family of functions. The locales available depends on what was installed with your operating system - use <code>locale -a</code> to list available locales. The default value is inherited from the execution environment of the server.</td>
<td>local system restart</td>
</tr>
<tr>
<td>lc_time</td>
<td>&lt;system dependent&gt;</td>
<td></td>
<td>This parameter currently does nothing, but may in the future.</td>
<td>local system restart</td>
</tr>
<tr>
<td>listen_addresses</td>
<td>localhost, host names, IP addresses, *</td>
<td></td>
<td>Specifies the TCP/IP address(es) on which the server is to listen for connections from client applications - a comma-separated list of host names and/or numeric IP addresses. The special entry * corresponds to all available IP interfaces. If the list is empty, only UNIX-domain sockets can connect.</td>
<td>master system restart</td>
</tr>
<tr>
<td>local_preload_libraries</td>
<td></td>
<td></td>
<td>Comma separated list of shared library files to preload at the start of a client session.</td>
<td>local system restart</td>
</tr>
<tr>
<td>log_autostats</td>
<td>boolean</td>
<td>on</td>
<td>Logs information about automatic <code>ANALYZE</code> operations related to <code>gp_autostats_mode</code> and <code>gp_autostats_on_change_threshold</code>.</td>
<td>master session reload superuser</td>
</tr>
<tr>
<td>log_connections</td>
<td>boolean</td>
<td>off</td>
<td>This outputs a line to the server log detailing each successful connection. Some client programs, like <code>psql</code>, attempt to connect twice while determining if a password is required, so duplicate &quot;connection received&quot; messages do not always indicate a problem.</td>
<td>local system restart</td>
</tr>
</tbody>
</table>
### Table D.2 Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_disconnections</td>
<td>boolean</td>
<td>off</td>
<td>This outputs a line in the server log at termination of a client session, and includes the duration of the session.</td>
<td>local system restart</td>
</tr>
<tr>
<td>log_dispatch_stats</td>
<td>boolean</td>
<td>off</td>
<td>When set to &quot;on,&quot; this parameter adds a log message with verbose information about the dispatch of the statement.</td>
<td>local system restart</td>
</tr>
<tr>
<td>log_duration</td>
<td>boolean</td>
<td>off</td>
<td>Causes the duration of every completed statement which satisfies log_statement to be logged.</td>
<td>master session reload superuser</td>
</tr>
<tr>
<td>log_error_verbosity</td>
<td>TERSE DEFAULT VERBOSE</td>
<td>DEFAULT</td>
<td>Controls the amount of detail written in the server log for each message that is logged.</td>
<td>master session reload superuser</td>
</tr>
<tr>
<td>log_executor_stats</td>
<td>boolean</td>
<td>off</td>
<td>For each query, write performance statistics of the query executor to the server log. This is a crude profiling instrument. Cannot be enabled together with log_statement_stats.</td>
<td>local system restart</td>
</tr>
<tr>
<td>log_hostname</td>
<td>boolean</td>
<td>off</td>
<td>By default, connection log messages only show the IP address of the connecting host. Turning on this option causes logging of the host name as well. Note that depending on your host name resolution setup this might impose a non-negligible performance penalty.</td>
<td>local system restart</td>
</tr>
<tr>
<td>log_min_duration_statement</td>
<td>number of milliseconds, 0, -1</td>
<td>-1</td>
<td>Logs the statement and its duration on a single log line if its duration is greater than or equal to the specified number of milliseconds. Setting this to 0 will print all statements and their durations. -1 disables the feature. For example, if you set it to 250 then all SQL statements that run 250ms or longer will be logged. Enabling this option can be useful in tracking down unoptimized queries in your applications.</td>
<td>master session reload superuser</td>
</tr>
<tr>
<td>log_min_error_statement</td>
<td>DEBUG5 DEBUG4 DEBUG3 DEBUG2, DEBUG1 INFO NOTICE WARNING ERROR FATAL PANIC</td>
<td>ERROR</td>
<td>Controls whether or not the SQL statement that causes an error condition will also be recorded in the server log. All SQL statements that cause an error of the specified level or higher are logged. The default is PANIC (effectively turning this feature off for normal use). Enabling this option can be helpful in tracking down the source of any errors that appear in the server log.</td>
<td>master session reload superuser</td>
</tr>
</tbody>
</table>
### Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_min_messages</td>
<td>DEBUG5, DEBUG4, DEBUG3, DEBUG2, DEBUG1, INFO, NOTICE, WARNING, ERROR, LOG, FATAL, PANIC</td>
<td>NOTICE</td>
<td>Controls which message levels are written to the server log. Each level includes all the levels that follow it. The later the level, the fewer messages are sent to the log.</td>
<td>master session reload superuser</td>
</tr>
<tr>
<td>log_parser_stats</td>
<td>boolean</td>
<td>off</td>
<td>For each query, write performance statistics of the query parser to the server log. This is a crude profiling instrument. Cannot be enabled together with log_statement_stats.</td>
<td>master session reload superuser</td>
</tr>
<tr>
<td>log_planner_stats</td>
<td>boolean</td>
<td>off</td>
<td>For each query, write performance statistics of the query planner to the server log. This is a crude profiling instrument. Cannot be enabled together with log_statement_stats.</td>
<td>master session reload superuser</td>
</tr>
<tr>
<td>log_rotation_age</td>
<td>number of minutes</td>
<td>1d</td>
<td>Determines the maximum lifetime of an individual log file. After this time has elapsed, a new log file will be created. Set to zero to disable time-based creation of new log files.</td>
<td>local system restart</td>
</tr>
<tr>
<td>log_rotation_size</td>
<td>number of kilobytes</td>
<td>0</td>
<td>Determines the maximum size of an individual log file. After this many kilobytes have been emitted into a log file, a new log file will be created. Set to zero to disable size-based creation of new log files.</td>
<td>local system restart</td>
</tr>
<tr>
<td>log_statement</td>
<td>NONE, DDL, MOD, ALL</td>
<td>ALL</td>
<td>Controls which SQL statements are logged. DDL logs all data definition commands like CREATE, ALTER, and DROP commands. MOD logs all DDL statements, plus INSERT, UPDATE, DELETE, TRUNCATE, and COPY FROM. PREPARE and EXPLAIN ANALYZE statements are also logged if their contained command is of an appropriate type.</td>
<td>master session reload superuser</td>
</tr>
<tr>
<td>log_statement_stats</td>
<td>boolean</td>
<td>off</td>
<td>For each query, write total performance statistics of the query parser, planner, and executor to the server log. This is a crude profiling instrument.</td>
<td>master session reload superuser</td>
</tr>
<tr>
<td>log_timezone</td>
<td>string</td>
<td>unknown</td>
<td>Sets the time zone used for timestamps written in the log. Unlike TimeZone, this value is system-wide, so that all sessions will report timestamps consistently. The default is unknown, which means to use whatever the system environment specifies as the time zone.</td>
<td>local system restart</td>
</tr>
</tbody>
</table>
### Table D.2 Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_truncate_on_rotation</td>
<td>boolean</td>
<td>off</td>
<td>Truncates (overwrites), rather than appends to, any existing log file of the same name. Truncation will occur only when a new file is being opened due to time-based rotation. For example, using this setting in combination with a log_filename such as <code>gpseg#-%H.log</code> would result in generating twenty-four hourly log files and then cyclically overwriting them. When off, pre-existing files will be appended to in all cases.</td>
<td>local system restart</td>
</tr>
<tr>
<td>maintenance_work_mem</td>
<td>number of kilobytes</td>
<td>64MB</td>
<td>Specifies the maximum amount of memory (in kilobytes) to be used in maintenance operations, such as VACUUM or CREATE INDEX. Since only one of these operations can be executed at a time by a database session, and an installation normally doesn’t have very many of them happening concurrently, it’s safe to set this value significantly larger than <code>work_mem</code>. Larger settings may improve performance for vacuuming and for restoring database dumps.</td>
<td>master session reload</td>
</tr>
<tr>
<td>max_appendonly_tables</td>
<td>2048</td>
<td></td>
<td>Sets the maximum number of append-only relations that can be written to or loaded concurrently. Append-only table partitions and subpartitions are considered as unique tables against this limit. Increasing the limit will allocate more shared memory at server start.</td>
<td>master system restart</td>
</tr>
<tr>
<td>max_connections</td>
<td>10-n</td>
<td>25 on master 125 on segments</td>
<td>The maximum number of concurrent connections to the database server. In a Greenplum Database system, user client connections go through the Greenplum master instance only. Segment instances should allow 5-10 times the amount as the master. When you increase this parameter, <code>max_prepared_transactions</code> must be increased as well. For more information, see Limiting Concurrent Connections. Increasing this parameter may cause Greenplum Database to request more shared memory.</td>
<td>local system restart</td>
</tr>
<tr>
<td>max_files_per_process</td>
<td>integer</td>
<td>1000</td>
<td>Sets the maximum number of simultaneously open files allowed to each server subprocess. If the kernel is enforcing a safe per-process limit, you don’t need to worry about this setting. Some platforms such as BSD, the kernel will allow individual processes to open many more files than the system can really support.</td>
<td>local system restart</td>
</tr>
</tbody>
</table>
### Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>max_fsm_pages</td>
<td>integer &gt; 16 * max_fsm_relations</td>
<td>200000</td>
<td>Sets the maximum number of disk pages for which free space will be tracked in the shared free-space map. Six bytes of shared memory are consumed for each page slot.</td>
<td>local system restart</td>
</tr>
<tr>
<td>max_fsm_relations</td>
<td>integer</td>
<td>1000</td>
<td>Sets the maximum number of relations for which free space will be tracked in the shared memory free-space map. Should be set to a value larger than the total number of: tables + indexes + system tables. It costs about 60 bytes of memory for each relation per segment instance. It is better to allow some room for overhead and set too high rather than too low.</td>
<td>local system restart</td>
</tr>
<tr>
<td>max_function_args</td>
<td>integer</td>
<td>100</td>
<td>Reports the maximum number of function arguments.</td>
<td>read only</td>
</tr>
<tr>
<td>max_identifier_length</td>
<td>integer</td>
<td>63</td>
<td>Reports the maximum identifier length.</td>
<td>read only</td>
</tr>
<tr>
<td>max_index_keys</td>
<td>integer</td>
<td>32</td>
<td>Reports the maximum number of index keys.</td>
<td>read only</td>
</tr>
<tr>
<td>max_locks_per_transaction</td>
<td>integer</td>
<td>64</td>
<td>The shared lock table is created with room to describe locks on max_locks_per_transaction * (max_connections + max_prepared_transactions) objects, so no more than this many distinct objects can be locked at any one time. This is not a hard limit on the number of locks taken by any one transaction, but rather a maximum average value. You might need to raise this value if you have clients that touch many different tables in a single transaction.</td>
<td>local system restart</td>
</tr>
<tr>
<td>max_prepared_transactions</td>
<td>integer</td>
<td>50 on master 50 on segments</td>
<td>Sets the maximum number of transactions that can be in the prepared state simultaneously. Greenplum uses prepared transactions internally to ensure data integrity across the segments. This value must be at least as large as the value of max_connections on the master. Segment instances should be set to the same value as the master. For more information, see Limiting Concurrent Connections.</td>
<td>local system restart</td>
</tr>
<tr>
<td>max_resource_portals_per_transaction</td>
<td>integer</td>
<td>64</td>
<td>Sets the maximum number of simultaneously open user-declared cursors allowed per transaction. Note that an open cursor will hold an active query slot in a resource queue. Used for workload management.</td>
<td>master system restart</td>
</tr>
<tr>
<td>max_resource_queues</td>
<td>integer</td>
<td>8</td>
<td>Sets the maximum number of resource queues that can be created in a Greenplum Database system. Note that resource queues are system-wide (as are roles) so they apply to all databases in the system.</td>
<td>master system restart</td>
</tr>
<tr>
<td>Parameter</td>
<td>Value Range</td>
<td>Default</td>
<td>Description</td>
<td>Set Classifications</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-------------------</td>
<td>---------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>max_stack_depth</td>
<td>number of kilobytes</td>
<td>2MB</td>
<td>Specifies the maximum safe depth of the server’s execution stack. The ideal setting for this parameter is the actual stack size limit enforced by the kernel (as set by <code>ulimit -s</code> or local equivalent), less a safety margin of a megabyte or so. Setting the parameter higher than the actual kernel limit will mean that a runaway recursive function can crash an individual backend process.</td>
<td>local system restart</td>
</tr>
</tbody>
</table>
| max_work_mem                      | number of kilobytes | 1000MB  | Sets the maximum amount of memory that can be used by a single segment worker process during query processing. Helps avoid out-of-memory errors on a segment host during query processing as a result of setting `work_mem` too high. Taking into account the configuration of a single segment host, calculate this setting as follows: 

\[
\text{((physical_memory/segments) / (average_query_slice_count))}
\]

The average number of query slices can be determined from looking at your `EXPLAIN` query plans. | master session reload superuser |
| password_encryption               | boolean           | on      | When a password is specified in CREATE USER or ALTER USER without writing either ENCRYPTED or UNENCRYPTED, this option determines whether the password is to be encrypted. | master session reload       |
| pljava_classpath                  | string            |         | A colon (:) separated list of the jar files containing the Java classes used in any PL/Java functions. The jar files listed here must also be installed on all Greenplum hosts in the following location: 

\[
\$GPHOME/lib/postgresql/java/
\]

| pljava_statement_cache_size       | number of kilobytes | 10      | Sets the size in KB of the JRE MRU (Most Recently Used) cache for prepared statements. | master system restart superuser |
| pljava_release_lingering_savepoints| boolean            | true    | If true, lingering savepoints used in PL/Java functions will be released on function exit. If false, savepoints will be rolled back. | master system restart superuser |
| pljava_vmoptions                  | string             | -Xmx64M | Defines the startup options for the Java VM. | master system restart superuser |
### Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>port</td>
<td>any valid port number</td>
<td>5432</td>
<td>The database listener port for a Greenplum instance. The master and each segment has its own port. Port numbers for the Greenplum system must also be changed in the gp_segment_configuration catalog. You must shut down your Greenplum Database system before changing port numbers.</td>
<td>local system restart</td>
</tr>
<tr>
<td>random_page_cost</td>
<td>floating point</td>
<td>100</td>
<td>Sets the planner’s estimate of the cost of a nonsequentially fetched disk page. This is measured as a multiple of the cost of a sequential page fetch. A higher value makes it more likely a sequential scan will be used, a lower value makes it more likely an index scan will be used.</td>
<td>master session reload</td>
</tr>
<tr>
<td>regex_flavor</td>
<td>advanced, extended, basic</td>
<td>advanced</td>
<td>The ‘extended’ setting may be useful for exact backwards compatibility with pre-7.4 releases of PostgreSQL.</td>
<td>master session reload</td>
</tr>
<tr>
<td>resource_cleanup_gangs_on_wait</td>
<td>boolean</td>
<td>on</td>
<td>If a statement is submitted through a resource queue, clean up any idle query executor worker processes before taking a lock on the resource queue.</td>
<td>master system restart</td>
</tr>
<tr>
<td>resource_select_only</td>
<td>boolean</td>
<td>off</td>
<td>Sets the types of queries managed by resource queues. If set to on, then SELECT, SELECT INTO, CREATE TABLE AS SELECT, and DECLARE CURSOR commands are evaluated. If set to off INSERT, UPDATE, and DELETE commands will be evaluated as well.</td>
<td>master system restart</td>
</tr>
<tr>
<td>search_path</td>
<td>a comma-separated list of schema names</td>
<td>$user,public</td>
<td>Specifies the order in which schemas are searched when an object is referenced by a simple name with no schema component. When there are objects of identical names in different schemas, the one found first in the search path is used. The system catalog schema, pg_catalog, is always searched, whether it is mentioned in the path or not. When objects are created without specifying a particular target schema, they will be placed in the first schema listed in the search path. The current effective value of the search path can be examined via the SQL function current_schemas(). current_schemas() shows how the requests appearing in search_path were resolved.</td>
<td>master session reload</td>
</tr>
<tr>
<td>seq_page_cost</td>
<td>floating point</td>
<td>1</td>
<td>Sets the planner’s estimate of the cost of a disk page fetch that is part of a series of sequential fetches.</td>
<td>master session reload</td>
</tr>
</tbody>
</table>
### Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>server_encoding</td>
<td>&lt;system dependent&gt;</td>
<td>UTF8</td>
<td>Reports the database encoding (character set). It is determined when the Greenplum Database array is initialized. Ordinarily, clients need only be concerned with the value of <code>client_encoding</code>.</td>
<td>read only</td>
</tr>
<tr>
<td>server_version</td>
<td>string</td>
<td>8.2.14</td>
<td>Reports the version of PostgreSQL that this release of Greenplum Database is based on.</td>
<td>read only</td>
</tr>
<tr>
<td>server_version_num</td>
<td>integer</td>
<td>80214</td>
<td>Reports the version of PostgreSQL that this release of Greenplum Database is based on as an integer.</td>
<td>read only</td>
</tr>
<tr>
<td>shared_buffers</td>
<td>integer &gt; 16K * $max_connections</td>
<td>125MB</td>
<td>Sets the amount of memory a Greenplum server instance uses for shared memory buffers. This setting must be at least 128 kilobytes and at least 16 kilobytes times $max_connections$.</td>
<td>local system restart</td>
</tr>
<tr>
<td>shared_preload_libraries</td>
<td></td>
<td></td>
<td>A comma-separated list of shared libraries that are to be preloaded at server start. PostgreSQL procedural language libraries can be preloaded in this way, typically by using the syntax '$libdir/plXXX' where XXX is pgsql, perl, tcl, or python. By preloading a shared library, the library startup time is avoided when the library is first used. If a specified library is not found, the server will fail to start.</td>
<td>local system restart</td>
</tr>
<tr>
<td>ssl</td>
<td>boolean</td>
<td>off</td>
<td>Enables SSL connections.</td>
<td>master system restart</td>
</tr>
<tr>
<td>ssl_ciphers</td>
<td>string</td>
<td>ALL</td>
<td>Specifies a list of SSL ciphers that are allowed to be used on secure connections. See the openssl manual page for a list of supported ciphers.</td>
<td>master system restart</td>
</tr>
<tr>
<td>standard_conforming_strings</td>
<td>boolean</td>
<td>off</td>
<td>Reports whether ordinary string literals ('...') treat backslashes literally, as specified in the SQL standard. The value is currently always off, indicating that backslashes are treated as escapes. It is planned that this will change to on in a future release when string literal syntax changes to meet the standard. Applications may check this parameter to determine how string literals will be processed. The presence of this parameter can also be taken as an indication that the escape string syntax (E'...') is supported.</td>
<td>read only</td>
</tr>
<tr>
<td>statement_timeout</td>
<td>number of milliseconds</td>
<td>0</td>
<td>Abort any statement that takes over the specified number of milliseconds. 0 turns off the limitation.</td>
<td>master session reload</td>
</tr>
</tbody>
</table>
Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>stats_queue_level</td>
<td>boolean</td>
<td>off</td>
<td>Collects resource queue statistics on database activity.</td>
<td>master session reload</td>
</tr>
<tr>
<td>superuser_reserved_connections</td>
<td>integer &lt; max_connections</td>
<td>3</td>
<td>Determines the number of connection slots that are reserved for Greenplum Database superusers.</td>
<td>local system restart</td>
</tr>
<tr>
<td>tcp_keepalives_count</td>
<td>number of lost keepalives</td>
<td>0</td>
<td>How many keepalives may be lost before the connection is considered dead. A value of 0 uses the system default. If TCP_KEEPCNT is not supported, this parameter must be 0.</td>
<td>local system restart</td>
</tr>
<tr>
<td>tcp_keepalives_idle</td>
<td>number of seconds</td>
<td>0</td>
<td>Number of seconds between sending keepalives on an otherwise idle connection. A value of 0 uses the system default. If TCP_KEEPIDLE is not supported, this parameter must be 0.</td>
<td>local system restart</td>
</tr>
<tr>
<td>tcp_keepalives_interval</td>
<td>number of seconds</td>
<td>0</td>
<td>How many seconds to wait for a response to a keepalive before retransmitting. A value of 0 uses the system default. If TCP_KEEPINTVL is not supported, this parameter must be 0.</td>
<td>local system restart</td>
</tr>
<tr>
<td>temp_buffers</td>
<td>integer</td>
<td>1024</td>
<td>Sets the maximum number of temporary buffers used by each database session. These are session-local buffers used only for access to temporary tables. The setting can be changed within individual sessions, but only up until the first use of temporary tables within a session. The cost of setting a large value in sessions that do not actually need a lot of temporary buffers is only a buffer descriptor, or about 64 bytes, per increment. However if a buffer is actually used, an additional 8192 bytes will be consumed.</td>
<td>master session reload</td>
</tr>
<tr>
<td>TimeZone</td>
<td>time zone abbreviation</td>
<td></td>
<td>Sets the time zone for displaying and interpreting time stamps. The default is to use whatever the system environment specifies as the time zone. See Date/Time Keywords in the PostgreSQL documentation.</td>
<td>master session reload</td>
</tr>
<tr>
<td>timezoneabbreviations</td>
<td>string</td>
<td>Default</td>
<td>Sets the collection of time zone abbreviations that will be accepted by the server for date time input. The default is Default, which is a collection that works in most of the world. Australia and India, and other collections can be defined for a particular installation. Possible values are names of configuration files stored in /share/timezonesets/ in the installation directory.</td>
<td>master session reload</td>
</tr>
</tbody>
</table>
Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>track_activities</td>
<td>boolean</td>
<td>on</td>
<td>Enables the collection of statistics on the currently executing command of each session, along with the time at which that command began execution. When enabled, this information is not visible to all users, only to superusers and the user owning the session. This data can be accessed via the <code>pg_stat_activity</code> system view.</td>
<td>master session reload</td>
</tr>
<tr>
<td>track_counts</td>
<td>boolean</td>
<td>off</td>
<td>Enables the collection of row and block level statistics on database activity. If enabled, the data that is produced can be accessed via the <code>pg_stat</code> and <code>pg_statio</code> family of system views.</td>
<td>local system restart</td>
</tr>
<tr>
<td>transaction_isolation</td>
<td>read committed</td>
<td>read committed</td>
<td>Sets the current transaction's isolation level.</td>
<td>master session reload</td>
</tr>
<tr>
<td>transaction_read_only</td>
<td>boolean</td>
<td>off</td>
<td>Sets the current transaction's read-only status.</td>
<td>master session reload</td>
</tr>
<tr>
<td>transform_null_equals</td>
<td>boolean</td>
<td>off</td>
<td>When on, expressions of the form <code>expr = NULL</code> or <code>NULL = expr</code> are treated as <code>expr IS NULL</code>, that is, they return true if <code>expr</code> evaluates to the null value, and false otherwise. The correct SQL-spec-compliant behavior of <code>expr = NULL</code> is to always return null (unknown).</td>
<td>master session reload</td>
</tr>
<tr>
<td>unix_socket_directory</td>
<td>directory path</td>
<td>unset</td>
<td>Specifies the directory of the UNIX-domain socket on which the server is to listen for connections from client applications.</td>
<td>local system restart</td>
</tr>
<tr>
<td>unix_socket_group</td>
<td>UNIX group name</td>
<td>unset</td>
<td>Sets the owning group of the UNIX-domain socket. By default this is an empty string, which uses the default group for the current user.</td>
<td>local system restart</td>
</tr>
<tr>
<td>unix_socket_permissions</td>
<td>numeric UNIX file permission mode</td>
<td>511</td>
<td>Sets the access permissions of the UNIX-domain socket. UNIX-domain sockets use the usual UNIX file system permission set. Note that for a UNIX-domain socket, only write permission matters.</td>
<td>local system restart</td>
</tr>
<tr>
<td>update_process_title</td>
<td>boolean</td>
<td>on</td>
<td>Enables updating of the process title every time a new SQL command is received by the server. The process title is typically viewed by the <code>ps</code> command.</td>
<td>local system restart</td>
</tr>
<tr>
<td>vacuum_cost_delay</td>
<td>milliseconds &lt; 0 (in multiples of 10)</td>
<td>0</td>
<td>The length of time that the process will sleep when the cost limit has been exceeded. 0 disables the cost-based vacuum delay feature.</td>
<td>local system restart</td>
</tr>
</tbody>
</table>
### Table D.2  Server Configuration Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Default</th>
<th>Description</th>
<th>Set Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>vacuum_cost_limit</td>
<td>integer &gt; 0</td>
<td>200</td>
<td>The accumulated cost that will cause the vacuuming process to sleep.</td>
<td>local, system restart</td>
</tr>
<tr>
<td>vacuum_cost_page_dirty</td>
<td>integer &gt; 0</td>
<td>20</td>
<td>The estimated cost charged when vacuum modifies a block that was previously clean. It represents the extra I/O required to flush the dirty block out to disk again.</td>
<td>local, system restart</td>
</tr>
<tr>
<td>vacuum_cost_page_hit</td>
<td>integer &gt; 0</td>
<td>1</td>
<td>The estimated cost for vacuuming a buffer found in the shared buffer cache. It represents the cost to lock the buffer pool, lookup the shared hash table and scan the content of the page.</td>
<td>local, system restart</td>
</tr>
<tr>
<td>vacuum_cost_page_miss</td>
<td>integer &gt; 0</td>
<td>10</td>
<td>The estimated cost for vacuuming a buffer that has to be read from disk. This represents the effort to lock the buffer pool, lookup the shared hash table, read the desired block in from the disk and scan its content.</td>
<td>local, system restart</td>
</tr>
<tr>
<td>vacuum_freeze_min_age</td>
<td>integer 0-1000000000000</td>
<td>1000000000000</td>
<td>Specifies the cutoff age (in transactions) that VACUUM should use to decide whether to replace transaction IDs with FrozenXID while scanning a table. VACUUM will limit the effective value to half the value of autovacuum_freeze_max_age, so that there is not an unreasonably short time between forced autovacuums.</td>
<td>local, system restart</td>
</tr>
<tr>
<td>work_mem</td>
<td>number of kilobytes</td>
<td>32MB</td>
<td>Specifies the amount of memory to be used by internal sort operations and hash tables before switching to temporary disk files. Note that for a complex query, several sort or hash operations might be running in parallel; each one will be allowed to use as much memory as this value specifies before it starts to put data into temporary files. Also, several running sessions could be doing such operations concurrently. So the total memory used could be many times the value of work_mem; it is necessary to keep this fact in mind when choosing the value. Sort operations are used for ORDER BY, DISTINCT, and merge joins. Hash tables are used in hash joins, hash-based aggregation, and hash-based processing of IN subqueries. See also, max_work_mem.</td>
<td>master, session, reload</td>
</tr>
</tbody>
</table>
E. Greenplum MapReduce Specification

MapReduce is a programming model developed by Google for processing and generating large data sets on an array of commodity servers. Greenplum MapReduce allows programmers who are familiar with the MapReduce paradigm to write map and reduce functions and submit them to the Greenplum Database parallel engine for processing.

In order for Greenplum to be able to process MapReduce functions, the functions need to be defined in a document, which is then passed to the Greenplum MapReduce program, `gpmapreduce`, for execution by the Greenplum Database parallel engine. The Greenplum Database system takes care of the details of distributing the input data, executing the program across a set of machines, handling machine failures, and managing the required inter-machine communication.

This specification describes the document format and schema for defining Greenplum MapReduce jobs.

Greenplum MapReduce Document Format

This section explains some basics of the Greenplum MapReduce document format to help you get started creating your own Greenplum MapReduce documents. Greenplum uses the YAML 1.1 document format and then implements its own schema for defining the various steps of a MapReduce job.

All Greenplum MapReduce files must first declare the version of the YAML specification they are using. After that, three dashes (---) denote the start of a document, and three dots (....) indicate the end of a document without starting a new one. Comment lines are prefixed with a pound symbol (#). It is possible to declare multiple Greenplum MapReduce documents in the same file:

```
%YAML 1.1
---
# Begin Document 1
# ...
---
# Begin Document 2
# ...
```

Within a Greenplum MapReduce document, there are three basic types of data structures or nodes: scalars, sequences and mappings.

A scalar is a basic string of text indented by a space. If you have a scalar input that spans multiple lines, a preceding pipe ( | ) denotes a literal style, where all line breaks are significant. Alternatively, a preceding angle bracket ( > ) folds a single line break
to a space for subsequent lines that have the same indentation level. If a string contains characters that have reserved meaning, the string must be quoted or the special character must be escaped with a backslash (\ ).

```plaintext
# Read each new line literally
somekey: |
  this value contains two lines
  and each line is read literally
# Treat each new line as a space
anotherkey: >
  this value contains two lines
  but is treated as one continuous line
# This quoted string contains a special character
ThirdKey: "This is a string: not a mapping"
```

A *sequence* is a list with each entry in the list on its own line denoted by a dash and a space (- ). Alternatively, you can specify an inline sequence as a comma-separated list within square brackets. A sequence provides a set of data and gives it an order. When you load a list into the Greenplum MapReduce program, the order is kept.

```plaintext
# list sequence
- this
- is
- a list
- with
- five scalar values
# inline sequence
[this, is, a list, with, five scalar values]
```

A *mapping* is used to pair up data values with identifiers called *keys*. Mappings use a colon and space ( : ) for each key: value pair, or can also be specified inline as a comma-separated list within curly braces. The *key* is used as an index for retrieving data from a mapping.

```plaintext
# a mapping of items
title: War and Peace
author: Leo Tolstoy
date: 1865
# same mapping written inline
{title: War and Peace, author: Leo Tolstoy, date: 1865}
```

Keys are used to associate meta information with each node and specify the expected node type (*scalar*, *sequence* or *mapping*). See “Greenplum MapReduce Document Schema” on page 801 for the keys expected by the Greenplum MapReduce program.
The Greenplum MapReduce program processes the nodes of a document in order and uses indentation (spaces) to determine the document hierarchy and the relationships of the nodes to one another. The use of white space is significant. White space should not be used simply for formatting purposes, and tabs should not be used at all.

**Greenplum MapReduce Document Schema**

Greenplum MapReduce uses the YAML document framework and implements its own YAML schema. The basic structure of a Greenplum MapReduce document is:

```yaml
%YAML 1.1
---
VERSION: 1.0.0.2
DATABASE: dbname
USER: db_username
HOST: master_hostname
PORT: master_port

DEFINE:

- INPUT:
  NAME: input_name
  FILE:
  - hostname:/path/to/file
  GPFDIST:
  - hostname:port:/file_pattern
  TABLE: table_name
  QUERY: SELECT_statement
  EXEC: command_string
  COLUMNS:
  - field_name data_type
  FORMAT: TEXT | CSV
  DELIMITER: delimiter_character
  ESCAPE: escape_character
  NULL: null_string
  QUOTE: csv_quote_character
  ERROR_LIMIT: integer
  ENCODING: database_encoding
```
- **OUTPUT**:
  
  `NAME`: `output_name`
  
  `FILE`: `file_path_on_client`
  
  `TABLE`: `table_name`
  
  `KEYS`:
  
  - `column_name`
  
  `MODE`: `REPLACE` | `APPEND`

- **MAP**:
  
  `NAME`: `function_name`
  
  `FUNCTION`: `function_definition`
  
  `LANGUAGE`: `perl` | `python` | `c`
  
  `LIBRARY`: `/path/filename.so`
  
  `PARAMETERS`:
  
  - `name` `type`
  
  `RETURNS`:
  
  - `name` `type`
  
  `OPTIMIZE`: `STRICT IMMUTABLE`
  
  `MODE`: `SINGLE` | `MULTI`

- **TRANSITION | CONSOLIDATE | FINALIZE**:
  
  `NAME`: `function_name`
  
  `FUNCTION`: `function_definition`
  
  `LANGUAGE`: `perl` | `python` | `c`
  
  `LIBRARY`: `/path/filename.so`
  
  `PARAMETERS`:
  
  - `name` `type`
  
  `RETURNS`:
  
  - `name` `type`
  
  `OPTIMIZE`: `STRICT IMMUTABLE`
  
  `MODE`: `SINGLE` | `MULTI`
- **REDUCE**:
  
  **NAME**: reduce_job_name  
  **TRANSITION**: transition_function_name  
  **CONSOLIDATE**: consolidate_function_name  
  **FINALIZE**: finalize_function_name  
  **INITIALIZE**: value  
  **KEYS**:
  - key_name

- **TASK**:
  
  **NAME**: task_name  
  **SOURCE**: input_name  
  **MAP**: map_function_name  
  **REDUCE**: reduce_function_name

**EXECUTE**:

- **RUN**:
  
  **SOURCE**: input_or_task_name  
  **TARGET**: output_name  
  **MAP**: map_function_name  
  **REDUCE**: reduce_function_name

**VERSION**

Required. The version of the Greenplum MapReduce YAML specification. Current versions are 1.0.0.1.

**DATABASE**

Optional. Specifies which database in Greenplum to connect to. If not specified, defaults to the default database or $PGDATABASE if set.

**USER**

Optional. Specifies which database role to use to connect. If not specified, defaults to the current user or $PGUSER if set. You must be a Greenplum superuser to run functions written in untrusted Python and Perl. Regular database users can run functions written in trusted Perl. You also must be a database superuser to run MapReduce jobs that contain FILE, GPFDIST or EXEC input types.

**HOST**

Optional. Specifies Greenplum master host name. If not specified, defaults to localhost or $PGHOST if set.
**PORT**
Optional. Specifies Greenplum master port. If not specified, defaults to 5432 or $PGPORT if set.

**DEFINE**
Required. A sequence of definitions for this MapReduce document. The DEFINE section must have at least one INPUT definition.

**INPUT**
Required. Defines the input data. Every MapReduce document must have at least one input defined. Multiple input definitions are allowed in a document, but each input definition can specify only one of these access types: a file, a gpfdist file distribution program, a table in the database, an SQL command, or an operating system command.

**NAME**
A name for this input. Names must be unique with regards to the names of other objects in this MapReduce job (such as map function, task, reduce function and output names). Also, names cannot conflict with existing objects in the database (such as tables, functions or views).

**FILE**
A sequence of one or more input files in the format: `seghostname:/path/to/filename`. You must be a Greenplum Database superuser to run MapReduce jobs with FILE input. The file must reside on a Greenplum segment host.

**GPFDIST**
A sequence of one or more running gpfdist file distribution programs in the format: `hostname[:port]/file_pattern`. You must be a Greenplum Database superuser to run MapReduce jobs with GPFDIST input, unless the server configuration parameter `gp_external_grant_privileges` is set to on.

**TABLE**
The name of an existing table in the database.

**QUERY**
An SQL SELECT command to run within the database.

**EXEC**
An operating system command to run on the Greenplum segment hosts. The command is run by all segment instances in the system by default. For example, if you have four segment instances per segment host, the command will be run four times on each host. You must be a Greenplum Database superuser to run MapReduce jobs with EXEC input and the server configuration parameter `gp_external_enable_exec` is set to on.
**COLUMNS**
Optional. Columns are specified as: `column_name [data_type]`. If not specified, the default is `value text`. The **DELIMITER** character is what separates two data value fields (columns). A row is determined by a line feed character (0x0a).

**FORMAT**
Optional. Specifies the format of the data - either delimited text (TEXT) or comma separated values (CSV) format. If the data format is not specified, defaults to TEXT.

**DELIMITER**
Optional for **FILE**, **GPFDIST** and **EXEC** inputs. Specifies a single character that separates data values. The default is a tab character in TEXT mode, a comma in CSV mode. The delimiter character must only appear between any two data value fields. Do not place a delimiter at the beginning or end of a row.

**ESCAPE**
Optional for **FILE**, **GPFDIST** and **EXEC** inputs. Specifies the single character that is used for C escape sequences (such as \n, \t, \100, and so on) and for escaping data characters that might otherwise be taken as row or column delimiters. Make sure to choose an escape character that is not used anywhere in your actual column data. The default escape character is a \ (backslash) for text-formatted files and a ” (double quote) for csv-formatted files, however it is possible to specify another character to represent an escape. It is also possible to disable escaping by specifying the value 'OFF' as the escape value. This is very useful for data such as text-formatted web log data that has many embedded backslashes that are not intended to be escapes.

**NULL**
Optional for **FILE**, **GPFDIST** and **EXEC** inputs. Specifies the string that represents a null value. The default is \N in TEXT format, and an empty value with no quotations in CSV format. You might prefer an empty string even in TEXT mode for cases where you do not want to distinguish nulls from empty strings. Any input data item that matches this string will be considered a null value.

**QUOTE**
Optional for **FILE**, **GPFDIST** and **EXEC** inputs. Specifies the quotation character for CSV formatted files. The default is a double quote (”). In CSV formatted files, data value fields must be enclosed in double quotes if they contain any commas or embedded new lines. Fields that contain double quote characters must be surrounded by double quotes, and the embedded double quotes must each be represented by a pair of consecutive double quotes. It is important to always open and close quotes correctly in order for data rows to be parsed correctly.
ERROR_LIMIT
If the input rows have format errors they will be discarded provided that the
error limit count is not reached on any Greenplum segment instance during
input processing. If the error limit is not reached, all good rows will be
processed and any error rows discarded.

ENCODING
Character set encoding to use for the data. Specify a string constant (such as
'SQL_ASCII'), an integer encoding number, or DEFAULT to use the default
client encoding. See “Character Set Support” on page 939.

OUTPUT
Optional. Defines where to output the formatted data of this MapReduce job. If
output is not defined, the default is STDOUT (standard output of the client). You
can send output to a file on the client host or to an existing table in the database.

NAME
A name for this output. The default output name is STDOUT. Names must be
unique with regards to the names of other objects in this MapReduce job
(such as map function, task, reduce function and input names). Also, names
cannot conflict with existing objects in the database (such as tables, functions
or views).

FILE
Specifies a file location on the MapReduce client machine to output data in
the format: /path/to/filename

TABLE
Specifies the name of a table in the database to output data. If this table does
not exist prior to running the MapReduce job, it will be created using the
distribution policy specified with KEYS.

KEYS
Optional for TABLE output. Specifies the column(s) to use as the Greenplum
Database distribution key. If the EXECUTE task contains a REDUCE definition,
then the REDUCE keys will be used as the table distribution key by default.
Otherwise, the first column of the table will be used as the distribution key.

MODE
Optional for TABLE output. If not specified, the default is to create the table if
it does not already exist, but error out if it does exist. Declaring APPEND adds
output data to an existing table (provided the table schema matches the output
format) without removing any existing data. Declaring REPLACE will drop the
table if it exists and then recreate it. Both APPEND and REPLACE will create a
new table if one does not exist.
**MAP**

Required. Each MAP function takes data structured in \((key, value)\) pairs, processes each pair, and generates zero or more output \((key, value)\) pairs. The Greenplum MapReduce framework then collects all pairs with the same key from all output lists and groups them together. This output is then passed to the REDUCE task, which is comprised of TRANSITION | CONSOLIDATE | FINALIZE functions. There is one predefined MAP function named IDENTITY that returns \((key, value)\) pairs unchanged. Although \((key, value)\) are the default parameters, you can specify other prototypes as needed.

**TRANSITION | CONSOLIDATE | FINALIZE**

TRANSITION, CONSOLIDATE and FINALIZE are all component pieces of REDUCE. A TRANSITION function is required. CONSOLIDATE and FINALIZE functions are optional. By default, all take state as the first of their input PARAMETERS, but other prototypes can be defined as well.

A TRANSITION function iterates through each value of a given key and accumulates values in a state variable. When the transition function is called on the first value of a key, the state is set to the value specified by INITIALIZE of a REDUCE job (or the default state value for the data type). A transition takes two arguments as input; the current state of the key reduction, and the next value, which then produces a new state.

If a CONSOLIDATE function is specified, TRANSITION processing is performed at the segment-level before redistributing the keys across the Greenplum interconnect for final aggregation (two-phase aggregation). Only the resulting state value for a given key is redistributed, resulting in lower interconnect traffic and greater parallelism. CONSOLIDATE is handled like a TRANSITION, except that instead of \((state + value) \Rightarrow state\), it is \((state + state) \Rightarrow state\).

If a FINALIZE function is specified, it takes the final state produced by CONSOLIDATE (if present) or TRANSITION and does any final processing before emitting the final result. TRANSITION and CONSOLIDATE functions cannot return a set of values. If you need a REDUCE job to return a set, then a FINALIZE is necessary to transform the final state into a set of output values.

**NAME**

Required. A name for the function. Names must be unique with regards to the names of other objects in this MapReduce job (such as function, task, input and output names). You can also specify the name of a function built-in to Greenplum Database. If using a built-in function, do not supply LANGUAGE or a FUNCTION body.

**FUNCTION**

Optional. Specifies the full body of the function using the specified LANGUAGE. If FUNCTION is not specified, then a built-in database function corresponding to NAME is used.
**LANGUAGE**

Required when FUNCTION is used. Specifies the implementation language used to interpret the function. This release has language support for perl, python and C. If calling a built-in database function, LANGUAGE should not be specified.

**LIBRARY**

Required when LANGUAGE is C (not allowed for other language functions). To use this attribute, VERSION must be 1.0.0.2. The specified library file must be installed prior to running the MapReduce job, and it must exist in the same file system location on all Greenplum hosts (master and segments).

**PARAMETERS**

Optional. Function input parameters. The default type is text.

- MAP default - key text, value text
- TRANSITION default - state text, value text
- CONSOLIDATE default - state1 text, state2 text (must have exactly two input parameters of the same data type)
- FINALIZE default - state text (single parameter only)

**RETURNS**

Optional. The default return type is text.

- MAP default - key text, value text
- TRANSITION default - state text (single return value only)
- CONSOLIDATE default - state text (single return value only)
- FINALIZE default - value text

**OPTIMIZE**

Optional optimization parameters for the function:

- STRICT - function is not affected by NULL values
- IMMUTABLE - function will always return the same value for a given input

**MODE**

Optional. Specifies the number of rows returned by the function.

- MULTI - returns 0 or more rows per input record. The return value of the function must be an array of rows to return, or the function must be written as an iterator using `yield` in Python or `return_next` in Perl. MULTI is the default mode for MAP and FINALIZE functions.
- SINGLE - returns exactly one row per input record. SINGLE is the only mode supported for TRANSITION and CONSOLIDATE functions. When used with MAP and FINALIZE functions, SINGLE mode can provide modest performance improvement.
REDUCE
Required. A REDUCE definition names the TRANSITION | CONSOLIDATE | FINALIZE functions that comprise the reduction of (key, value) pairs to the final result set. There are also several predefined REDUCE jobs you can execute, which all operate over a column named value:

IDENTITY - returns (key, value) pairs unchanged
SUM - calculates the sum of numeric data
AVG - calculates the average of numeric data
COUNT - calculates the count of input data
MIN - calculates minimum value of numeric data
MAX - calculates maximum value of numeric data

NAME
Required. The name of this REDUCE job. Names must be unique with regards to the names of other objects in this MapReduce job (function, task, input and output names). Also, names cannot conflict with existing objects in the database (such as tables, functions or views).

TRANSITION
Required. The name of the TRANSITION function.

CONSOLIDATE
Optional. The name of the CONSOLIDATE function.

FINALIZE
Optional. The name of the FINALIZE function.

INITIALIZE
Optional for text and float data types. Required for all other data types. The default value for text is '' . The default value for float is 0.0 . Sets the initial state value of the TRANSITION function.

KEYS
Optional. Defaults to [key, *]. When using a multi-column reduce it may be necessary to specify which columns are key columns and which columns are value columns. By default, any input columns that are not passed to the TRANSITION function are key columns, and a column named key is always a key column even if it is passed to the TRANSITION function. The special indicator * indicates all columns not passed to the TRANSITION function. If this indicator is not present in the list of keys then any unmatched columns are discarded.
**TASK**
Optional. A **TASK** defines a complete end-to-end **INPUT/MAP/REDUCE** stage within a Greenplum MapReduce job pipeline. It is similar to **EXECUTE** except it is not immediately executed. A task object can be called as **INPUT** to further processing stages.

**NAME**
Required. The name of this task. Names must be unique with regards to the names of other objects in this MapReduce job (such as map function, reduce function, input and output names). Also, names cannot conflict with existing objects in the database (such as tables, functions or views).

**SOURCE**
The name of an **INPUT** or another **TASK**.

**MAP**
Optional. The name of a **MAP** function. If not specified, defaults to **IDENTITY**.

**REDUCE**
Optional. The name of a **REDUCE** function. If not specified, defaults to **IDENTITY**.

**EXECUTE**
Required. **EXECUTE** defines the final **INPUT/MAP/REDUCE** stage within a Greenplum MapReduce job pipeline.

**RUN**

**SOURCE**
Required. The name of an **INPUT** or **TASK**.

**TARGET**
Optional. The name of an **OUTPUT**. The default output is **STDOUT**.

**MAP**
Optional. The name of a **MAP** function. If not specified, defaults to **IDENTITY**.

**REDUCE**
Optional. The name of a **REDUCE** function. Defaults to **IDENTITY**.
Example Greenplum MapReduce Document

# This example MapReduce job processes documents and looks for keywords in them.
# It takes two database tables as input:
# - documents (doc_id integer, url text, data text)
# - keywords (keyword_id integer, keyword text)
# The documents data is searched for occurrences of keywords and returns results of
# url, data and keyword (a keyword can be multiple words, such as "high performance
# computing")

%YAML 1.1
---

VERSION: 1.0.0.1
# Connect to Greenplum Database using this database and role
DATABASE: webdata
USER: jsmith
# Begin definition section
DEFINE:
  # Declare the input, which selects all columns and rows from the
  # 'documents' and 'keywords' tables.
  - INPUT:
    NAME: doc
    TABLE: documents

  - INPUT:
    NAME: kw
    TABLE: keywords

# Define the map functions to extract terms from documents and keyword
# This example simply splits on white space, but it would be possible
# to make use of a python library like nltk (the natural language toolkit)
# to perform more complex tokenization and word stemming.
- MAP:

  NAME: doc_map
  LANGUAGE: python
  FUNCTION: |
  
  # the index of a word within the document
  i = 0
  # a hash of terms and their indexes within the document
  terms = {}
  
  # Lower-case and split the text string on space
  for term in data.lower().split():
    i = i + 1
    # increment i (the index)
    # Check for the term in the terms list:
    # if stem word already exists, append the i value to the array entry
    # corresponding to the term. This counts multiple occurences of the word.
    # If stem word does not exist, add it to the dictionary with position i.
    # For example:
    # data: "a computer is a machine that manipulates data"
    # "a" [1, 4]
    # "computer" [2]
    # "machine" [3]
    # ...
    if term in terms:
      terms[term] += ','+str(i)
    else:
      terms[term] = str(i)
  
  # Return multiple lines for each document. Each line consists of
  # the doc_id, a term and the positions in the data where the term appeared.
  # For example:
  # (doc_id => 100, term => "a", [1,4]
  # (doc_id => 100, term => "computer", [2]
  # ...
  for term in terms:
    yield([doc_id, term, terms[term]])

  OPTIMIZE: STRICT IMMUTABLE

  PARAMETERS:
  - doc_id integer
  - data text

  RETURNS:
  - doc_id integer
  - term text
  - positions text
# The map function for keywords is almost identical to the one for documents
# but it also counts of the number of terms in the keyword.

- MAP:
  
  NAME: kw_map
  LANGUAGE: python
  
  FUNCTION:
  ```python
  i = 0
  terms = {}

  for term in keyword.lower().split():
      i = i + 1
      if term in terms:
          terms[term] += ','+str(i)
      else:
          terms[term] = str(i)
  # output 4 values including i (the total count for term in terms):
  yield([keyword_id, i, term, terms[term]])
  ```

  OPTIMIZE: STRICT IMMUTABLE
  
  PARAMETERS:
  - keyword_id integer
  - keyword text
  
  RETURNS:
  - keyword_id integer
  - nterms integer
  - term text
  - positions text

# A TASK is an object that defines an entire INPUT/MAP/REDUCE stage
# within a Greenplum MapReduce pipeline. It is like EXECUTION, but it is
# executed only when called as input to other processing stages.

# Identify a task called 'doc_prep' which takes in the 'doc' INPUT defined earlier
# and runs the 'doc_map' MAP function which returns doc_id, term, [term_position]

- TASK:
  
  NAME: doc_prep
  SOURCE: doc
  MAP: doc_map
# Identify a task called 'kw_prep' which takes in the 'kw' INPUT defined earlier
# and runs the kw_map MAP function which returns kw_id, term, [term_position]
- TASK:
  NAME: kw_prep
  SOURCE: kw
  MAP: kw_map

# One advantage of Greenplum MapReduce is that MapReduce tasks can be
# used as input to SQL operations and SQL can be used to process a MapReduce task.
# This INPUT defines a SQL query that joins the output of the 'doc_prep'
# TASK to that of the 'kw_prep' TASK. Matching terms are output to the 'candidate'
# list (any keyword that shares at least one term with the document).
- INPUT:
  NAME: term_join
  QUERY: |
  
  SELECT doc.doc_id, kw.keyword_id, kw.term, kw.nterms,
          doc.positions as doc_positions,
          kw.positions as kw_positions
  FROM doc_prep doc INNER JOIN kw_prep kw ON (doc.term = kw.term)

# In Greenplum MapReduce, a REDUCE function is comprised of one or more functions.
# A REDUCE has an initial 'state' variable defined for each grouping key. that is
# A TRANSITION function adjusts the state for every value in a key grouping.
# If present, an optional CONSOLIDATE function combines multiple
# 'state' variables. This allows the TRANSITION function to be executed locally at
# the segment-level and only redistribute the accumulated 'state' over
# the network. If present, an optional FINALIZE function can be used to perform
# final computation on a state and emit one or more rows of output from the state.
#
# This REDUCE function is called 'term_reducer' with a TRANSITION function
# called 'term_transition' and a FINALIZE function called 'term_finalizer'

- REDUCE:
  NAME: term_reducer
  TRANSITION: term_transition
  FINALIZE: term_finalizer
- TRANSITION:
  
  NAME: term_transition
  LANGUAGE: python
  PARAMETERS:
  - state text
  - term text
  - nterms integer
  - doc_positions text
  - kw_positions text
  
  FUNCTION: |
  # 'state' has an initial value of '' and is a colon delimited set
  # of keyword positions. keyword positions are comma delimited sets of
  # integers. For example, '1,3,2:4:'
  # If there is an existing state, split it into the set of keyword positions
  # otherwise construct a set of 'nterms' keyword positions - all empty
  if state:
    kw_split = state.split('::')
  else:
    kw_split = []
    for i in range(0,nterms):
      kw_split.append('')
  # 'kw_positions' is a comma delimited field of integers indicating what
  # position a single term occurs within a given keyword.
  # Splitting based on ',' converts the string into a python list.
  # add doc_positions for the current term
  for kw_p in kw_positions.split(', '):
    kw_split[int(kw_p) - 1] = doc_positions
  # This section takes each element in the 'kw_split' array and strings
  # them together placing a ':' in between each element from the array.
  # For example: for the keyword "computer software computer hardware",
  # the 'kw_split' array matched up to the document data of
  # "in the business of computer software software engineers"
  # would look like: ['5', '6,7', '5', '']
  # and the outstate would look like: 5:6,7:5:
  outstate = kw_split[0]
  for s in kw_split[1:]:
    outstate = outstate + ':' + s
  return outstate
- FINALIZE:
  NAME: term_finalizer
  LANGUAGE: python
  RETURNS:
    - count integer
  MODE: MULTI
  FUNCTION: |
    if not state:
      return 0
  kw_split = state.split(':
  # This function does the following:
  # 1) Splits 'kw_split' on ':'
  # for example, 1,5,7:2,8 creates '1,5,7' and '2,8'
  # 2) For each group of positions in 'kw_split', splits the set on ',',
  # to create ['1','5','7'] from Set 0: 1,5,7 and
  # eventually ['2', '8'] from Set 1: 2,8
  # 3) Checks for empty strings
  # 4) Adjusts the split sets by subtracting the position of the set
  # in the 'kw_split' array
  # ['1','5','7'] - 0 from each element = ['1','5','7']
  # ['2', '8'] - 1 from each element = ['1', '7']
  # 5) Resulting arrays after subtracting the offset in step 4 are
  # intersected and their overlapping values kept:
  # ['1','5','7'].intersect['1', '7'] = [1,7]
  # 6) Determines the length of the intersection, which is the number of
  # times that an entire keyword (with all its pieces) matches in the
  # document data.
  previous = None
  for i in range(0,len(kw_split)):
    isplit = kw_split[i].split(',
    if any(map(lambda(x): x == ',', isplit)):
      return 0
    adjusted = set(map(lambda(x): int(x)-i, isplit))
    if (previous):
      previous = adjusted.intersection(previous)
    else:
      previous = adjusted
  # return the final count
  if previous:
    return len(previous)
  return 0
# Define the 'term_match' task which is then executed as part
# of the 'final_output' query. It takes the INPUT 'term_join' defined
# earlier and uses the REDUCE function 'term_reducer' defined earlier

- TASK:
  NAME: term_match
  SOURCE: term_join
  REDUCE: term_reducer

- INPUT:
  NAME: final_output
  QUERY:
  | SELECT doc.*, kw.*, tm.count
  | FROM documents doc, keywords kw, term_match tm
  | WHERE doc.doc_id = tm.doc_id
  | AND kw.keyword_id = tm.keyword_id
  | AND tm.count > 0

# Execute this MapReduce job and send output to STDOUT

EXECUTE:

- RUN:
  SOURCE: final_output
  TARGET: STDOUT
MapReduce Flow Diagram

The following diagram shows the job flow of the MapReduce job defined in the example:

**Task: doc_prep**

**Input:** doc

**Table:** documents

**Example row:** `<doc_id=100, data="A mapreduce tutorial for keyword matching in Greenplum mapreduce">`

**Map function:** doc_map

**Returns:** doc_id int, term text, positions [text]

**Example:**
- `<doc_id=100, term="A", positions=["1"]`
- `<doc_id=100, term="mapreduce", positions=[2,3]>`

**Task: kw_prep**

**Input:** kw

**Table:** keywords

**Example row:** `<keyword_id=1, keyword="mapreduce">`

**Map function:** kw_map

**Returns:** keyword_id int, terms int, term text, positions [text]

**Example:**
- `<keyword_id=1, term="mapreduce", positions=["1"]`
- `<keyword_id=1, term="tutorial", positions=["2"]`

**Task: term_match**

**Input:** term_join

**Query:** `SELECT doc.doc_id, kw.keyword_id, kw.terms, kw.terms, doc.positions as doc_positions, kw.positions as kw_positions FROM doc_prep.doc INNER JOIN kw_prep kw ON doc.term = kw.term`

**Example:**
- `<doc_id=100, keyword_id=1, term="mapreduce", terms=2, doc_positions=[2,3], kw_positions=["1"]`
- `<doc_id=100, keyword_id=1, term="tutorial", terms=2, doc_positions=[3], kw_positions=["2"]`

**Reduce function:** term_reducer

**Keys:** [doc_id, keyword_id]

**Transition function:** term_transition

**Returns:** state text

**Example:** `<state="2,9,3">`

**Example:** `<doc_id=100, keyword_id=1, count=1>`

**Finalize function:** term_finalizer

**Returns:** count integer

**Example:** `<count=1>`

**Execute:**

**Input:** final_output

**Query:** `SELECT * FROM documents.doc, keywords kw, term_match tm WHERE doc.doc_id = tm.doc_id AND kw.keyword_id = tm.keyword_id AND tm.count > 0`

**Example:** `<doc_id=100, data="A mapreduce tutorial for keyword matching in Greenplum mapreduce", keyword_id=1, keyword="mapreduce tutorial", count=1>`

**Stdout**
F. Greenplum Environment Variables

This is a reference of the environment variables to set for Greenplum Database. Set these in your user’s startup shell profile (such as ~/.bashrc or ~/.bash_profile), or in /etc/profile if you want to set them for all users.

Required Environment Variables

<table>
<thead>
<tr>
<th>GPHOME</th>
</tr>
</thead>
<tbody>
<tr>
<td>This is the installed location of your Greenplum Database software. For example:</td>
</tr>
<tr>
<td>GPHOME=/usr/local/greenplum-db-4.0.x.x</td>
</tr>
<tr>
<td>export GPHOME</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PATH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your PATH environment variable should point to the location of the Greenplum Database bin directory. Solaris users must also add /usr/sfw/bin and /opt/sfw/bin to their PATH. For example:</td>
</tr>
<tr>
<td>PATH=$GPHOME/bin:$PATH</td>
</tr>
<tr>
<td>PATH=$GPHOME/bin:/usr/local/bin:/usr/sbin:/usr/sfw/bin:/opt/sfw/bin:$PATH</td>
</tr>
<tr>
<td>export PATH</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LD_LIBRARY_PATH</th>
</tr>
</thead>
<tbody>
<tr>
<td>The LD_LIBRARY_PATH environment variable should point to the location of the Greenplum Database/PostgreSQL library files. For Solaris, this also points to the GNU compiler and readline library files as well (readline libraries may be required for Python support on Solaris). For example:</td>
</tr>
<tr>
<td>LD_LIBRARY_PATH=$GPHOME/lib</td>
</tr>
<tr>
<td>LD_LIBRARY_PATH=$GPHOME/lib:/usr/sfw/lib</td>
</tr>
<tr>
<td>export LD_LIBRARY_PATH</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MASTER_DATA_DIRECTORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>This should point to the directory created by the gpinit system utility in the master data directory location. For example:</td>
</tr>
<tr>
<td>MASTER_DATA_DIRECTORY=/data/master/gpseg-1</td>
</tr>
<tr>
<td>export MASTER_DATA_DIRECTORY</td>
</tr>
</tbody>
</table>

Note: GPHOME, PATH and LD_LIBRARY_PATH can be set by sourcing the greenplum_path.sh file from your Greenplum Database installation directory.
Optional Environment Variables

The following are standard PostgreSQL environment variables, which are also recognized in Greenplum Database. You may want to add the connection-related environment variables to your profile for convenience, so you do not have to type so many options on the command line for client connections. Note that these environment variables should be set on the Greenplum Database master host only.

**PGDATABASE**
The name of the default database to use when connecting.

**PGHOST**
The Greenplum Database master host name.

**PGHOSTADDR**
The numeric IP address of the master host. This can be set instead of or in addition to PGHOST to avoid DNS lookup overhead.

**PGPASSWORD**
The password used if the server demands password authentication. Use of this environment variable is not recommended for security reasons (some operating systems allow non-root users to see process environment variables via `ps`). Instead consider using the `~/.pgpass` file.

**PGPASSFILE**
The name of the password file to use for lookups. If not set, it defaults to `~/.pgpass`. See the section about The Password File in the PostgreSQL documentation for more information.

**PGOPTIONS**
Sets additional configuration parameters for the Greenplum Database master server.

**PGPORT**
The port number of the Greenplum Database server on the master host. The default port is 5432.

**PGUSER**
The Greenplum Database user name used to connect.

**PGDATESTYLE**
Sets the default style of date/time representation for a session. (Equivalent to `SET datestyle TO ....`)  

**PGTZ**
Sets the default time zone for a session. (Equivalent to `SET timezone TO ....`) 

**PGCLIENTENCODING**
Sets the default client character set encoding for a session. (Equivalent to `SET client_encoding TO ....`)
Greenplum Database Data Types

Greenplum Database has a rich set of native data types available to users. Users may also define new data types using the `CREATE TYPE` command. This reference shows all of the built-in data types. In addition to the types listed here, there are also some internally used data types, such as `oid` (object identifier), but those are not documented in this guide.

The following data types are specified by SQL: `bit`, `bit varying`, `boolean`, `character varying`, `varchar`, `character`, `char`, `date`, `double precision`, `integer`, `interval`, `numeric`, `decimal`, `real`, `smallint`, `time` (with or without time zone), and `timestamp` (with or without time zone).

Each data type has an external representation determined by its input and output functions. Many of the built-in types have obvious external formats. However, several types are either unique to PostgreSQL (and Greenplum Database), such as geometric paths, or have several possibilities for formats, such as the date and time types. Some of the input and output functions are not invertible. That is, the result of an output function may lose accuracy when compared to the original input.

<table>
<thead>
<tr>
<th>Name</th>
<th>Alias</th>
<th>Size</th>
<th>Range</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bigint</td>
<td>int8</td>
<td>8 bytes</td>
<td>-922372036854775808 to 922372036854775807</td>
<td>large range integer</td>
</tr>
<tr>
<td>bigserial</td>
<td>serial8</td>
<td>8 bytes</td>
<td>1 to 922372036854775807</td>
<td>large autoincrementing integer</td>
</tr>
<tr>
<td>bit [ (n) ]</td>
<td>n bits</td>
<td></td>
<td>bit string constant</td>
<td>fixed-length bit string</td>
</tr>
<tr>
<td>bit varying [ (n) ]</td>
<td>varbit</td>
<td>actual number of bits</td>
<td>bit string constant</td>
<td>variable-length bit string</td>
</tr>
<tr>
<td>boolean</td>
<td>bool</td>
<td>1 byte</td>
<td>true/false, t/f, yes/no, y/n, 1/0</td>
<td>logical boolean (true/false)</td>
</tr>
<tr>
<td>box</td>
<td></td>
<td>32 bytes</td>
<td>((x1,y1),(x2,y2))</td>
<td>rectangular box in the plane - not allowed in distribution key columns.</td>
</tr>
<tr>
<td>bytea</td>
<td></td>
<td>1 byte + binary string</td>
<td>sequence of octets</td>
<td>variable-length binary string</td>
</tr>
<tr>
<td>character [ (n) ]</td>
<td>char [ (n) ]</td>
<td>1 byte + n</td>
<td>strings up to n characters in length</td>
<td>fixed-length, blank padded</td>
</tr>
<tr>
<td>character varying [ (n) ]</td>
<td>varchar [ (n) ]</td>
<td>1 byte + string size</td>
<td>strings up to n characters in length</td>
<td>variable-length with limit</td>
</tr>
<tr>
<td>cidr</td>
<td></td>
<td>12 or 24 bytes</td>
<td>IPv4 and IPv6 networks</td>
<td>IPv4 and IPv6 networks</td>
</tr>
</tbody>
</table>

Table G.1  Greenplum Database Built-in Data Types
<table>
<thead>
<tr>
<th>Name</th>
<th>Alias</th>
<th>Size</th>
<th>Range</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>circle</td>
<td></td>
<td>24 bytes</td>
<td>&lt;(x,y),r&gt; (center and radius)</td>
<td>circle in the plane - not allowed in distribution key columns.</td>
</tr>
<tr>
<td>date</td>
<td></td>
<td>4 bytes</td>
<td>4713 BC - 5874897 AD</td>
<td>calendar date (year, month, day)</td>
</tr>
<tr>
<td>decimal [ (p, s) ]</td>
<td>numeric [ (p, s) ]</td>
<td>variable</td>
<td>no limit</td>
<td>user-specified precision, exact</td>
</tr>
<tr>
<td>double precision</td>
<td>float8</td>
<td>8 bytes</td>
<td>15 decimal digits precision</td>
<td>variable-precision, inexact</td>
</tr>
<tr>
<td>inet</td>
<td>float</td>
<td>12 or 24 bytes</td>
<td></td>
<td>IPV4 and IPv6 hosts and networks</td>
</tr>
<tr>
<td>integer</td>
<td>int, int4</td>
<td>4 bytes</td>
<td>-2147483648 to +2147483647</td>
<td>usual choice for integer</td>
</tr>
<tr>
<td>interval [ (p) ]</td>
<td></td>
<td>12 bytes</td>
<td>-178000000 years - 178000000 years</td>
<td>time span</td>
</tr>
<tr>
<td>lseg</td>
<td></td>
<td>32 bytes</td>
<td>((x1,y1),(x2,y2))</td>
<td>line segment in the plane - not allowed in distribution key columns.</td>
</tr>
<tr>
<td>macaddr</td>
<td></td>
<td>6 bytes</td>
<td></td>
<td>MAC addresses</td>
</tr>
<tr>
<td>money</td>
<td></td>
<td>4 bytes</td>
<td>-21474836.48 to +21474836.47</td>
<td>currency amount</td>
</tr>
<tr>
<td>path</td>
<td></td>
<td>16+16n bytes</td>
<td>[(x1,y1),…]</td>
<td>geometric path in the plane - not allowed in distribution key columns.</td>
</tr>
<tr>
<td>point</td>
<td></td>
<td>16 bytes</td>
<td>(x,y)</td>
<td>geometric point in the plane - not allowed in distribution key columns.</td>
</tr>
<tr>
<td>polygon</td>
<td></td>
<td>40+16n bytes</td>
<td>((x1,y1),…),</td>
<td>closed geometric path in the plane - not allowed in distribution key columns.</td>
</tr>
<tr>
<td>real</td>
<td>float4</td>
<td>4 bytes</td>
<td>6 decimal digits precision</td>
<td>variable-precision, inexact</td>
</tr>
<tr>
<td>serial</td>
<td>serial4</td>
<td>4 bytes</td>
<td>1 to 2147483647</td>
<td>autoincrementing integer</td>
</tr>
<tr>
<td>smallint</td>
<td>int2</td>
<td>2 bytes</td>
<td>-32768 to +32767</td>
<td>small range integer</td>
</tr>
<tr>
<td>text</td>
<td></td>
<td>1 byte + string size</td>
<td>strings of any length</td>
<td>variable unlimited length</td>
</tr>
<tr>
<td>time [ (p) ] [ without time zone ]</td>
<td></td>
<td>8 bytes</td>
<td>00:00:00[.000000] - 24:00:00[.000000]</td>
<td>time of day only</td>
</tr>
<tr>
<td>time [ (p) ] with time zone</td>
<td>timetz</td>
<td>12 bytes</td>
<td>00:00:00+1359 - 24:00:00-1359</td>
<td>time of day only, with time zone</td>
</tr>
<tr>
<td>timestamp [ (p) ] [ without time zone ]</td>
<td></td>
<td>8 bytes</td>
<td>4713 BC - 5874897 AD</td>
<td>both date and time</td>
</tr>
<tr>
<td>timestamp [ (p) ] with time zone</td>
<td>timestampzt</td>
<td>8 bytes</td>
<td>4713 BC - 5874897 AD</td>
<td>both date and time, with time zone</td>
</tr>
</tbody>
</table>

1. For variable length data types (such as char, varchar, text, etc.) if the data is greater than or equal to 127 bytes, the storage overhead is 4 bytes instead of 1.
H. System Catalog Reference

This is a reference of the system catalog tables and views of Greenplum Database. All system tables related to the parallel features of Greenplum Database are prefixed with `gp_`. Tables prefixed with `pg_` are either standard PostgreSQL system catalog tables (which are also used in Greenplum Database), or are related to features Greenplum has added to enhance PostgreSQL for data warehousing workloads. Note that the global system catalog for Greenplum Database resides on the master instance.

**System Tables**

- `gp_configuration` (no longer used, see `gp_segment_configuration`)
- `gp_configuration_history`
- `gp_db_interfaces`
- `gp_distribution_policy`
- `gp_fastsequence`
- `gp_fault_strategy`
- `gp_global_sequence`
- `gp_id`
- `gp_interfaces`
- `gp_master_mirroring`
- `gp_persistent_database_node`
- `gp_persistent_filespace_node`
- `gp_persistent_relation_node`
- `gp_persistent_tablespace_node`
- `gp_relation_node`
- `gp_san_configuration`
- `gp_segment_configuration`
- `gp_version_at_initdb`
- `gpexpand.status`
- `gpexpand.status_detail`
- `pg_aggregate`
- `pg_am`
- `pg_amop`
- `pg_amproc`
- `pg_appendonly`
- `pg_appendonly_alter_column` (not implemented in 4.0 - for a future release)
- pg_attrdef
- pg_attribute
- pg_auth_members
- pg_authid
- pg_autovacuum
- pg_cast
- pg_class
- pg_constraint
- pg_conversion
- pg_database
- pg_depend
- pg_description
- pg_exttable
- pg_filespace
- pg_filespace_entry
- pg_foreign_data_wrapper (not implemented in 4.0 - for a future release)
- pg_foreign_server (not implemented in 4.0 - for a future release)
- pg_foreign_table (not implemented in 4.0 - for a future release)
- pg_index
- pg Inherits
- pg_language
- pg_largeobject
- pg_listener
- pg_namespace
- pg_opclass
- pg_operator
- pg_partition
- pg_partition_rule
- pg_pltemplate
- pg_proc
- pg_resourcetype
- pg_resqueue
- pg_resqueuencapability
- pg_rewrite
- pg_shdepend
- `pg_shdescription`
- `pg_stat_last_operation`
- `pg_stat_last_shoperation`
- `pg_statistic`
- `pg_tablespace`
- `pg_trigger`
- `pg_type`
- `pg_user_mapping` (not implemented in 4.0 - for a future release)
- `pg_window`

**System Views**

Greenplum Database also contains the following system views currently not available in PostgreSQL:

- `gp_distributed_log`
- `gp_distributed_xacts`
- `gp_pgdatabase`
- `gp_transaction_log`
- `gpexpand.expansion_progress`
- `pg_max_external_files` (shows number of external table files allowed per segment host when using the file protocol)
- `pg_partition_columns`
- `pg_partition_templates`
- `pg_partitions`
- `pg_resqueue_attributes`
- `pg_resqueue_status`
- `pg_stat_resqueues`
- `pg_user_mappings` (not implemented in 4.0 - for a future release)

For more information on the standard system views in PostgreSQL (which are also used in Greenplum Database), see the following sections of the PostgreSQL documentation:

- **System Views**
- **Statistics Collector Views**
- **The Information Schema**
The gp_configuration_history table contains information about system changes related to fault detection and recovery operations. The fts_probe process logs data to this table, as do certain related management utilities such as gpcheck, gprecoverseg, and gpinitsystem. For example, when you add a new segment and mirror segment to the system, records for these events are logged to gp_configuration_history.

The event descriptions stored in this table may be helpful for troubleshooting serious system issues in collaboration with Greenplum support technicians.

This table is populated only on the master. This table is defined in the pg_global tablespace, meaning it is globally shared across all databases in the system.

Table H.1  pg_catalog.gp_configuration_history

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>timestamp with time zone</td>
<td>gp_segment_configuration.dbid</td>
<td>Timestamp for the event recorded.</td>
</tr>
<tr>
<td>dbid</td>
<td>smallint</td>
<td></td>
<td>System-assigned ID. The unique identifier of a segment (or master) instance.</td>
</tr>
<tr>
<td>desc</td>
<td>text</td>
<td></td>
<td>Text description of the event.</td>
</tr>
</tbody>
</table>
The `gp_distributed_log` view contains status information about distributed transactions and their associated local transactions. A distributed transaction is a transaction that involves modifying data on the segment instances. Greenplum’s distributed transaction manager ensures that the segments stay in synch. This view allows you to see the status of distributed transactions.

Table H.1  pg_catalog.gp_distributed_log

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>segment_id</td>
<td>smallint</td>
<td><code>gp_segment_configuration.content</code></td>
<td>The content id if the segment. The master is always -1 (no content).</td>
</tr>
<tr>
<td>dbid</td>
<td>small_int</td>
<td><code>gp_segment_configuration.dbid</code></td>
<td>The unique id of the segment instance.</td>
</tr>
<tr>
<td>distributed_xid</td>
<td>xid</td>
<td></td>
<td>The global transaction id.</td>
</tr>
<tr>
<td>distributed_id</td>
<td>text</td>
<td></td>
<td>A system assigned ID for a distributed transaction.</td>
</tr>
<tr>
<td>status</td>
<td>text</td>
<td></td>
<td>The status of the distributed transaction (Committed or Aborted).</td>
</tr>
<tr>
<td>local_transaction</td>
<td>xid</td>
<td></td>
<td>The local transaction ID.</td>
</tr>
</tbody>
</table>
gp_distributed_xacts

The gp_distributed_xacts view contains information about Greenplum Database distributed transactions. A distributed transaction is a transaction that involves modifying data on the segment instances. Greenplum’s distributed transaction manager ensures that the segments stay in sync. This view allows you to see the currently active sessions and their associated distributed transactions.

Table H.1  pg_catalog.gp_distributed_xacts

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>distributed_xid</td>
<td>xid</td>
<td></td>
<td>The transaction ID used by the distributed transaction across the Greenplum Database array.</td>
</tr>
<tr>
<td>distributed_id</td>
<td>text</td>
<td></td>
<td>The distributed transaction identifier. It has 2 parts — a unique timestamp and the distributed transaction number.</td>
</tr>
<tr>
<td>state</td>
<td>text</td>
<td></td>
<td>The current state of this session with regards to distributed transactions.</td>
</tr>
<tr>
<td>gp_session_id</td>
<td>int</td>
<td></td>
<td>The ID number of the Greenplum Database session associated with this transaction.</td>
</tr>
<tr>
<td>xmin_distributed_snapshot</td>
<td>xid</td>
<td></td>
<td>The minimum distributed transaction number found among all open transactions when this transaction was started. It is used for MVCC distributed snapshot purposes.</td>
</tr>
</tbody>
</table>
The `gp_distribution_policy` table contains information about Greenplum Database tables and their policy for distributing table data across the segments. This table is populated only on the master. This table is not globally shared, meaning each database has its own copy of this table.

### Table H.1 pg_catalog.gp_distribution_policy

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>localoid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The table object identifier (OID).</td>
</tr>
<tr>
<td>attnums</td>
<td>smallint[]</td>
<td><code>pg_attribute.attnum</code></td>
<td>The column number(s) of the distribution column(s).</td>
</tr>
</tbody>
</table>
gp_fastsequence

The `gp_fastsequence` table contains information about indexes on append-only column-oriented tables. It is used to track the maximum row number used by a file segment of an append-only column-oriented table.

**Table H.1  pg_catalog.gp_fastsequence**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>objid</td>
<td>oid</td>
<td>pg_class.oid</td>
<td>Object id of the pg_aoseg.pg_aocsseg_* table used to track append-only file segments.</td>
</tr>
<tr>
<td>objmod</td>
<td>bigint</td>
<td></td>
<td>Object modifier.</td>
</tr>
<tr>
<td>last_sequence</td>
<td>bigint</td>
<td></td>
<td>The last sequence number used by the object.</td>
</tr>
</tbody>
</table>
### gp_fault_strategy

The `gp_fault_strategy` table specifies the fault action.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>fault_strategy</td>
<td>char</td>
<td></td>
<td>The mirror failover action to take when a segment failure occurs: n = nothing. f = file-based failover. s = SAN-based failover.</td>
</tr>
</tbody>
</table>

Table H.1  pg_catalog.gp_fault_strategy
gp_global_sequence

The `gp_global_sequence` table contains the log sequence number position in the transaction log, which is used by the file replication process to determine the file blocks to replicate from a primary to a mirror segment.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sequence_num</td>
<td>bigint</td>
<td></td>
<td>log sequence number position in the transaction log</td>
</tr>
</tbody>
</table>
gpexpand.status

The \textit{gpexpand.status} table contains information about the status of a system expansion operation. Status for specific tables involved in the expansion is stored in \textit{gpexpand.status\_detail}.

In a normal expansion operation it is not necessary to modify the data stored in this table.

\textbf{Table H.1} gpexpand.status

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>status</td>
<td>text</td>
<td></td>
<td>Tracks the status of an expansion operation. Valid values are:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SETUP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SETUP DONE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>EXPANSION STARTED</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>EXPANSION STOPPED</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>COMPLETED</td>
</tr>
<tr>
<td>updated</td>
<td>timestamp with time zone</td>
<td></td>
<td>Timestamp of the last change in status.</td>
</tr>
</tbody>
</table>
gpexpand.status_detail

The gpexpand.status_detail table contains information about the status of tables involved in a system expansion operation. You can query this table to determine the status of tables being expanded, or to view the start and end time for completed tables.

This table also stores related information about the table such as the oid, disk size, and normal distribution policy and key. Overall status information for the expansion is stored in gpexpand.status.

In a normal expansion operation it is not necessary to modify the data stored in this table.

Table H.1 gpexpand.status_detail

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbname</td>
<td>text</td>
<td></td>
<td>Name of the database to which the table belongs.</td>
</tr>
<tr>
<td>fq_name</td>
<td>text</td>
<td></td>
<td>Fully qualified name of the table.</td>
</tr>
<tr>
<td>schema_oid</td>
<td>oid</td>
<td></td>
<td>OID for the schema of the database to which the table belongs.</td>
</tr>
<tr>
<td>table_oid</td>
<td>oid</td>
<td></td>
<td>OID of the table.</td>
</tr>
<tr>
<td>distribution_policy</td>
<td>smallint()</td>
<td></td>
<td>Array of column IDs for the distribution key of the table.</td>
</tr>
<tr>
<td>distribution_policy_names</td>
<td>text</td>
<td></td>
<td>Column names for the hash distribution key.</td>
</tr>
<tr>
<td>distribution_policy_coloids</td>
<td>text</td>
<td></td>
<td>Column IDs for the distribution keys of the table.</td>
</tr>
<tr>
<td>storage_options</td>
<td>text</td>
<td></td>
<td>Not enabled in this release. Do not update this field.</td>
</tr>
<tr>
<td>rank</td>
<td>int</td>
<td></td>
<td>Rank determines the order in which tables are expanded. The expansion utility will sort on rank and expand the lowest-ranking tables first.</td>
</tr>
<tr>
<td>status</td>
<td>text</td>
<td></td>
<td>Status of expansion for this table. Valid values are: NOT STARTED IN PROGRESS FINISHED</td>
</tr>
<tr>
<td>last_updated</td>
<td>timestamp with time zone</td>
<td></td>
<td>Timestamp of the last change in status for this table.</td>
</tr>
<tr>
<td>expansion started</td>
<td>timestamp with time zone</td>
<td></td>
<td>Timestamp for the start of the expansion of this table. This field is only populated after a table is successfully expanded.</td>
</tr>
</tbody>
</table>
## Table H.1  gpexpand.status_detail

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>expansion finished</td>
<td>timestamp with time zone</td>
<td></td>
<td>Timestamp for the completion of expansion of this table.</td>
</tr>
<tr>
<td>source bytes</td>
<td></td>
<td></td>
<td>The size of disk space associated with the source table. Due to table bloat in heap tables and differing numbers of segments after expansion, it is not expected that the final number of bytes will equal the source number. This information is tracked to help provide progress measurement to aid in duration estimation for the end-to-end expansion operation.</td>
</tr>
</tbody>
</table>
gp_id

The gp_id system catalog table identifies the Greenplum Database system name and number of segments for the system. It also has local values for the particular database instance (segment or master) on which the table resides. This table is defined in the pg_global tablespace, meaning it is globally shared across all databases in the system.

Table H.1  pg_catalog.gp_id

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gpname</td>
<td>name</td>
<td></td>
<td>The name of this Greenplum Database system.</td>
</tr>
<tr>
<td>numsegments</td>
<td>integer</td>
<td></td>
<td>The number of segments in the Greenplum Database system.</td>
</tr>
<tr>
<td>dbid</td>
<td>integer</td>
<td></td>
<td>The unique identifier of this segment (or master) instance.</td>
</tr>
<tr>
<td>content</td>
<td>integer</td>
<td></td>
<td>The ID for the portion of data on this segment instance. A primary and its mirror will have the same content ID. For a segment the value is from 0-N, where N is the number of segments in Greenplum Database. For the master, the value is -1.</td>
</tr>
</tbody>
</table>
The `gp_interfaces` table contains information about network interfaces on segment hosts. This information, joined with data from `gp_db_interfaces`, is used by the system to optimize the usage of available network interfaces for various purposes, including fault detection.

### Table H.1  gp_interfaces

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>interfaceid</td>
<td>smallint</td>
<td></td>
<td>System-assigned ID. The unique identifier of a network interface.</td>
</tr>
<tr>
<td>address</td>
<td>name</td>
<td></td>
<td>Hostname address for the segment host containing the network interface. Can be a numeric IP address or a hostname.</td>
</tr>
<tr>
<td>status</td>
<td>smallint</td>
<td></td>
<td>Status for the network interface. A value of 0 indicates that the interface is unavailable.</td>
</tr>
</tbody>
</table>
The `gp_master_mirroring` table contains state information about the standby master host and its associated write-ahead log (WAL) replication process. If this synchronization process (`gpsyncagent`) fails on the standby master, it may not always be noticeable to users of the system. This catalog is a place where Greenplum Database administrators can check to see if the standby master is current and fully synchronized.

**Table H.1  pg_catalog.gp_master_mirroring**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>summary_state</td>
<td>text</td>
<td></td>
<td>The current state of the log replication process between the master and standby master - logs are either 'Synchronized' or 'Not Synchronized'</td>
</tr>
<tr>
<td>detail_state</td>
<td>text</td>
<td></td>
<td>If not synchronized, this column will have information about the cause of the error.</td>
</tr>
<tr>
<td>log_time</td>
<td>timestampz</td>
<td></td>
<td>This contains the timestamp of the last time the master sent its logs to the standby master.</td>
</tr>
<tr>
<td>error_message</td>
<td>text</td>
<td></td>
<td>If not synchronized, this column will have the error message from the failed synchronization attempt.</td>
</tr>
</tbody>
</table>
gp_persistent_database_node

The `gp_persistent_database_node` table keeps track of the status of file system objects in relation to the transaction status of database objects. This information is used to make sure the state of the system catalogs and the file system files persisted to disk are synchronized. This information is used by the primary to mirror file replication process.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tablespace_oid</td>
<td>oid</td>
<td>pg_tablespace.oid</td>
<td>Table space object id.</td>
</tr>
<tr>
<td>database_oid</td>
<td>oid</td>
<td>pg_database.oid</td>
<td>Database object id.</td>
</tr>
</tbody>
</table>
| persistent_state| smallint    |                  | 1 - create pending  
2 - created  
3 - drop pending  
4 - dropped |
| mirror_state    | smallint    |                  | 0 - rsync pending or no mirrors  
1 - rsync complete |
| parent_xid      | integer     |                  | Global transaction id.                                           |
| persistent_serial_num | bigint  |                  | Log sequence number position in the transaction log for a file block. |
| previous_free_tid | tid        |                  | Used by Greenplum Database to internally manage persistent representations of file system objects. |
gp_persistent_filespace_node

The `gp_persistent_filespace_node` table keeps track of the status of file system objects in relation to the transaction status of filespace objects. This information is used to make sure the state of the system catalogs and the file system files persisted to disk are synchronized. This information is used by the primary to mirror file replication process.

Table H.1  pg_catalog.gp_persistent_filespace_node

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>filespace_oid</td>
<td>oid</td>
<td>pg_filespace.oid</td>
<td>object id of the filespace</td>
</tr>
<tr>
<td>db_id_1</td>
<td>smallint</td>
<td></td>
<td>primary segment id</td>
</tr>
<tr>
<td>location_1</td>
<td>text</td>
<td></td>
<td>primary filesystem location</td>
</tr>
<tr>
<td>db_id_2</td>
<td>smallint</td>
<td></td>
<td>mirror segment id</td>
</tr>
<tr>
<td>location_2</td>
<td>text</td>
<td></td>
<td>mirror filesystem location</td>
</tr>
<tr>
<td>persistent_state</td>
<td>smallint</td>
<td></td>
<td>1 - create pending</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - created</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - drop pending</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 - dropped</td>
</tr>
<tr>
<td>mirror_state</td>
<td>smallint</td>
<td></td>
<td>0 - rsync pending or no mirrors</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 - rsync complete</td>
</tr>
<tr>
<td>parent_xid</td>
<td>integer</td>
<td></td>
<td>Global transaction id.</td>
</tr>
<tr>
<td>persistent_serial_num</td>
<td>bigint</td>
<td></td>
<td>Log sequence number position in the transaction log for a file block.</td>
</tr>
<tr>
<td>previous_free_tid</td>
<td>tid</td>
<td></td>
<td>Used by Greenplum Database to internally manage persistent representations of file system objects.</td>
</tr>
</tbody>
</table>
gp_persistent_relation_node

The `gp_persistent_relation_node` table keeps track of the status of file system objects in relation to the transaction status of relation objects (tables, view, indexes, and so on). This information is used to make sure the state of the system catalogs and the file system files persisted to disk are synchronized. This information is used by the primary to mirror file replication process.

**Table H.1** pg_catalog.gp_persistent_relation_node

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>tablespace_oid</code></td>
<td>oid</td>
<td><code>pg_tablespace.oid</code></td>
<td>Tablespace object id</td>
</tr>
<tr>
<td><code>database_oid</code></td>
<td>oid</td>
<td><code>pg_database.oid</code></td>
<td>Database object id</td>
</tr>
<tr>
<td><code>relfilenode_oid</code></td>
<td>oid</td>
<td><code>pg_class.relfilenode</code></td>
<td>The object id of the relation file node.</td>
</tr>
<tr>
<td><code>segment_file_num</code></td>
<td>integer</td>
<td></td>
<td>For append-only tables, the append-only segment file number.</td>
</tr>
<tr>
<td><code>relation_storage_manager</code></td>
<td>smallint</td>
<td></td>
<td>Whether the relation is heap storage or append-only storage.</td>
</tr>
</tbody>
</table>
| `persistent_state` | smallint |                     | 1 - create pending  
|                   |           |                     | 2 - created  
|                   |           |                     | 3 - drop pending  
|                   |           |                     | 4 - dropped      |
| `mirror_state`   | smallint  |                     | 0 - rsync pending or no mirrors  
|                   |           |                     | 1 - rsync complete           |
| `mirror_append_only_eof` | bigint |                     | For append-only tables, the last end of file position on the mirror.       |
| `parent_xid`     | integer   |                     | Global transaction id.                                                     |
| `persistent_serial_num` | bigint |                     | Log sequence number position in the transaction log for a file block.    |
| `previous_free_tid` | tid        |                     | Used by Greenplum Database to internally manage persistent representations of file system objects. |
gp_persistent_tablespace_node

The gp_persistent_tablespace_node table keeps track of the status of file system objects in relation to the transaction status of tablespace objects. This information is used to make sure the state of the system catalogs and the file system files persisted to disk are synchronized. This information is used by the primary to mirror file replication process.

Table H.1  pg_catalog.gp_persistent_tablespace_node

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>filespace_oid</td>
<td>oid</td>
<td>pg_filespace.oid</td>
<td>Filespace object id</td>
</tr>
<tr>
<td>tablespace_oid</td>
<td>oid</td>
<td>pg_tablespace.oid</td>
<td>Tablespace object id</td>
</tr>
</tbody>
</table>
| persistent_state| smallint|                  | 1 - create pending  
2 - created  
3 - drop pending  
4 - dropped    |
| mirror_state    | smallint|                  | 0 - rsync pending or no mirrors  
1 - rsync complete |
| parent_xid      | integer |                  | Global transaction id.                         |
| persistent_serial_num | bigint |               | Log sequence number position in the transaction log for a file block. |
| previous_free_tid| tid     |                  | Used by Greenplum Database to internally manage persistent representations of file system objects. |
gp_relation_node

The `gp_relation_node` table contains information about the file system objects for a relation (table, view, index, and so on).

**Table H.1**  pg_catalog.gp_relation_node

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tablespace_oid</td>
<td>oid</td>
<td>pg_tablespace.oid</td>
<td>Tablespace object id.</td>
</tr>
<tr>
<td>relfilenode_oid</td>
<td>oid</td>
<td>pg_class.relfilenode</td>
<td>The object id of the relation file node.</td>
</tr>
<tr>
<td>segment_file_num</td>
<td>integer</td>
<td></td>
<td>For append-only tables, the append-only segment file number.</td>
</tr>
<tr>
<td>persistent_tid</td>
<td>tid</td>
<td></td>
<td>Used by Greenplum Database to internally manage persistent representations of file system objects.</td>
</tr>
<tr>
<td>persistent_serial_num</td>
<td>bigint</td>
<td></td>
<td>Log sequence number position in the transaction log for a file block.</td>
</tr>
</tbody>
</table>
**gp_san_configuration**

The `gp_san_configuration` table contains mount-point information for SAN failover.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>mountid</td>
<td>smallint</td>
<td></td>
<td>A value that identifies the mountpoint for the primary and mirror hosts. This is the primary key which is referred to by the value that appears in the <code>san_mounts</code> structure in <code>gp_segment_configuration</code>.</td>
</tr>
<tr>
<td>active_host</td>
<td>char</td>
<td></td>
<td>The current active host. <code>p</code> indicates primary, and <code>m</code> indicates mirror.</td>
</tr>
<tr>
<td>san_type</td>
<td>char</td>
<td></td>
<td>The type of shared storage in use. <code>n</code> indicates NFS, and <code>e</code> indicates EMC SAN.</td>
</tr>
<tr>
<td>primary_host</td>
<td>text</td>
<td></td>
<td>The name of the primary host system.</td>
</tr>
<tr>
<td>primary_mountpoint</td>
<td>text</td>
<td></td>
<td>The mount point for the primary host.</td>
</tr>
<tr>
<td>primary_device</td>
<td>text</td>
<td></td>
<td>A string specifying the device to mount on the primary mountpoint. For NFS, this string is similar to: <code>nfs-server:/exported/fs</code>. For EMC this is a larger string that includes the WWN for the storage processor, the storage-processor IP, and the storage-group name.</td>
</tr>
<tr>
<td>mirror_host</td>
<td>text</td>
<td></td>
<td>The name or the mirror/backup host system. The <code>primary_device</code> field is identical to the <code>mirror_device</code> field.</td>
</tr>
</tbody>
</table>
### Table H.1  pg_catalog.gp_san_configuration

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>mirror_mountpoint</td>
<td>text</td>
<td></td>
<td>The mount point for the mirror/backup host.</td>
</tr>
<tr>
<td>mirror_device</td>
<td>text</td>
<td></td>
<td>A string specifying the device to mount on the mirror mountpoint. For NFS, this string is similar to: nfs-server:/exported/fs. For EMC this is a larger string that includes the WWN for the storage processor, the storage-processor IP, and the storage-group name. The mirror_device field is identical to the primary_device field.</td>
</tr>
</tbody>
</table>

The mirror_device field is identical to the primary_device field.
The `gp_segment_configuration` table contains information about mirroring and segment configuration.

### Table H.1  pg_catalog.gp_segment_configuration

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbid</td>
<td>smallint</td>
<td></td>
<td>The unique identifier of a segment (or master) instance.</td>
</tr>
<tr>
<td>content</td>
<td>smallint</td>
<td></td>
<td>The content identifier for a segment instance. A primary segment instance and its corresponding mirror will always have the same content identifier. For a segment the value is from 0-N, where N is the number of primary segments in the system. For the master, the value is always -1.</td>
</tr>
<tr>
<td>role</td>
<td>char</td>
<td></td>
<td>The role that a segment is currently running as. Values are p (primary) or m (mirror).</td>
</tr>
<tr>
<td>preferred_role</td>
<td>char</td>
<td></td>
<td>The role that a segment was originally assigned at initialization time. Values are p (primary) or m (mirror).</td>
</tr>
<tr>
<td>mode</td>
<td>char</td>
<td></td>
<td>The synchronization status of a segment with its mirror copy. Values are s (synchronized), c (change logging), or r (resyncing).</td>
</tr>
<tr>
<td>status</td>
<td>char</td>
<td></td>
<td>The fault status of a segment. Values are u (up) or d (down).</td>
</tr>
<tr>
<td>port</td>
<td>integer</td>
<td></td>
<td>The TCP port the database server listener process is using.</td>
</tr>
<tr>
<td>hostname</td>
<td>text</td>
<td></td>
<td>The hostname of a segment host.</td>
</tr>
<tr>
<td>address</td>
<td>text</td>
<td></td>
<td>The hostname used to access a particular segment on a segment host. This value may be the same as <code>hostname</code> in systems upgraded from 3.x or on systems that do not have per-interface hostnames configured.</td>
</tr>
<tr>
<td>replication_port</td>
<td>integer</td>
<td></td>
<td>The TCP port the file block replication process is using to keep primary and mirror segments synchronized.</td>
</tr>
<tr>
<td>san_mounts</td>
<td>int2vector</td>
<td></td>
<td>An array of references to the <code>gp_san_configuration</code> table. Only used on systems that were initialized using shared storage.</td>
</tr>
</tbody>
</table>
The `gp.pgdatabase` view shows status information about the Greenplum segment instances and whether they are acting as the mirror or the primary. This view is used internally by the Greenplum fault detection and recovery utilities to determine failed segments.

### Table H.1  `pg_catalog.gp_pgdatabase`

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>dbid</code></td>
<td><code>smallint</code></td>
<td><code>gp_segment_configuration.dbid</code></td>
<td>System-assigned ID. The unique identifier of a segment (or master) instance.</td>
</tr>
<tr>
<td><code>isprimary</code></td>
<td><code>boolean</code></td>
<td><code>gp_segment_configuration.role</code></td>
<td>Whether or not this instance is active. Is it currently acting as the primary segment (as opposed to the mirror).</td>
</tr>
<tr>
<td><code>content</code></td>
<td><code>smallint</code></td>
<td><code>gp_segment_configuration.content</code></td>
<td>The ID for the portion of data on an instance. A primary segment instance and its mirror will have the same content ID. For a segment the value is from 0-N, where N is the number of segments in Greenplum Database. For the master, the value is -1.</td>
</tr>
<tr>
<td><code>definedprimary</code></td>
<td><code>boolean</code></td>
<td><code>gp_segment_configuration.preferred_role</code></td>
<td>Whether or not this instance was defined as the primary (as opposed to the mirror) at the time the system was initialized.</td>
</tr>
</tbody>
</table>
The `gp_transaction_log` view contains status information about transactions local to a particular segment. This view allows you to see the status of local transactions.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>segment_id</td>
<td>smallint</td>
<td>gp_segment_configuration.content</td>
<td>The content id if the segment. The master is always -1 (no content).</td>
</tr>
<tr>
<td>dbid</td>
<td>smallint</td>
<td>gp_segment_configuration.dbid</td>
<td>The unique id of the segment instance.</td>
</tr>
<tr>
<td>transaction</td>
<td>xid</td>
<td></td>
<td>The local transaction ID.</td>
</tr>
<tr>
<td>status</td>
<td>text</td>
<td></td>
<td>The status of the local transaction (Committed or Aborted).</td>
</tr>
</tbody>
</table>
The `gp_version_at_initdb` table is populated on the master and each segment in the Greenplum Database system. It identifies the version of Greenplum Database used when the system was first initialized. This table is defined in the `pg_global` tablespace, meaning it is globally shared across all databases in the system.

**Table H.1  pg_catalog.gp_version**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>schemaversion</td>
<td>integer</td>
<td></td>
<td>Schema version number.</td>
</tr>
<tr>
<td>productversion</td>
<td>text</td>
<td></td>
<td>Product version number.</td>
</tr>
</tbody>
</table>
**gpexpand.expansion_progress**

The `gpexpand.expansion_progress` view contains information about the status of a system expansion operation. The view provides calculations of the estimated rate of table redistribution and estimated time to completion.

Status for specific tables involved in the expansion is stored in `gpexpand.status_detail`.

**Table H.1  gpexpand.expansion_progress**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
<td>text</td>
<td></td>
<td>Name for the data field provided Includes:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bytes Left</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bytes Done</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Estimated Expansion Rate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Estimated Time to Completion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tables Expanded</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tables Left</td>
</tr>
<tr>
<td>value</td>
<td>text</td>
<td></td>
<td>The value for the progress data. For example:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Estimated Expansion Rate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- 9.75667095996092 MB/s</td>
</tr>
</tbody>
</table>
**pg_aggregate**

The `pg_aggregate` table stores information about aggregate functions. An aggregate function is a function that operates on a set of values (typically one column from each row that matches a query condition) and returns a single value computed from all these values. Typical aggregate functions are `sum`, `count`, and `max`. Each entry in `pg_aggregate` is an extension of an entry in `pg_proc`. The `pg_proc` entry carries the aggregate’s name, input and output data types, and other information that is similar to ordinary functions.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aggfoid</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>Aggregate function OID</td>
</tr>
<tr>
<td>aggtransfn</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>Transition function OID</td>
</tr>
<tr>
<td>aggpelimfn</td>
<td>regproc</td>
<td></td>
<td>Preliminary function OID (zero if none)</td>
</tr>
<tr>
<td>aggfinalfn</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>Final function OID (zero if none)</td>
</tr>
<tr>
<td>agginvtransfn</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>The OID in <code>pg_proc</code> of the inverse function of <code>aggtransfn</code></td>
</tr>
<tr>
<td>agginvprelimfn</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>The OID in <code>pg_proc</code> of the inverse function of <code>aggprelimfn</code></td>
</tr>
<tr>
<td>aggsortop</td>
<td>oid</td>
<td><code>pg_operator.oid</code></td>
<td>Associated sort operator OID (zero if none)</td>
</tr>
<tr>
<td>aggtranstype</td>
<td>oid</td>
<td><code>pg_type.oid</code></td>
<td>Data type of the aggregate function’s internal transition (state) data</td>
</tr>
<tr>
<td>agginitval</td>
<td>text</td>
<td></td>
<td>The initial value of the transition state. This is a text field containing the initial value in its external string representation. If this field is NULL, the transition state value starts out NULL</td>
</tr>
</tbody>
</table>
The `pg_am` table stores information about index access methods. There is one row for each index access method supported by the system.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>amname</td>
<td>name</td>
<td></td>
<td>Name of the access method</td>
</tr>
<tr>
<td>amstrategies</td>
<td>int2</td>
<td></td>
<td>Number of operator strategies for this access method</td>
</tr>
<tr>
<td>amsupport</td>
<td>int2</td>
<td></td>
<td>Number of support routines for this access method</td>
</tr>
<tr>
<td>amorderstrategy</td>
<td>int2</td>
<td></td>
<td>Zero if the index offers no sort order, otherwise the strategy number of the</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>strategy operator that describes the sort order</td>
</tr>
<tr>
<td>amcanunique</td>
<td>boolean</td>
<td></td>
<td>Does the access method support unique indexes?</td>
</tr>
<tr>
<td>amcanmulticol</td>
<td>boolean</td>
<td></td>
<td>Does the access method support multicol indexes?</td>
</tr>
<tr>
<td>amoptionalkey</td>
<td>boolean</td>
<td></td>
<td>Does the access method support a scan without any constraint for the first</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>index column?</td>
</tr>
<tr>
<td>amindexnulls</td>
<td>boolean</td>
<td></td>
<td>Does the access method support null index entries?</td>
</tr>
<tr>
<td>amstorage</td>
<td>boolean</td>
<td></td>
<td>Can index storage data type differ from column data type?</td>
</tr>
<tr>
<td>amclusterable</td>
<td>boolean</td>
<td></td>
<td>Can an index of this type be clustered on?</td>
</tr>
<tr>
<td>aminsert</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>“Insert this tuple” function</td>
</tr>
<tr>
<td>ambeginscan</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>“Start new scan” function</td>
</tr>
<tr>
<td>amgettuple</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>“Next valid tuple” function</td>
</tr>
<tr>
<td>amgetmulti</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>“Fetch multiple tuples” function</td>
</tr>
<tr>
<td>amrescan</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>“Restart this scan” function</td>
</tr>
<tr>
<td>amendscan</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>“End this scan” function</td>
</tr>
<tr>
<td>ammarkpos</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>“Mark current scan position” function</td>
</tr>
<tr>
<td>amrestrpos</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>“Restore marked scan position” function</td>
</tr>
<tr>
<td>ambuild</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>“Build new index” function</td>
</tr>
<tr>
<td>ambulkdelete</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>Bulk-delete function</td>
</tr>
<tr>
<td>amvacuumcleanup</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>Post-VACUUM cleanup function</td>
</tr>
<tr>
<td>column</td>
<td>type</td>
<td>references</td>
<td>description</td>
</tr>
<tr>
<td>--------------</td>
<td>----------</td>
<td>----------------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>amcostestimate</td>
<td>regproc</td>
<td>pg_proc.oid</td>
<td>Function to estimate cost of an index scan</td>
</tr>
<tr>
<td>amoptions</td>
<td>regproc</td>
<td>pg_proc.oid</td>
<td>Function to parse and validate reloptions for an index</td>
</tr>
</tbody>
</table>
The `pg_amop` table stores information about operators associated with index access method operator classes. There is one row for each operator that is a member of an operator class.

**Table H.1  pg_catalog.pg_amop**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>amopclaid</td>
<td>oid</td>
<td><code>pg_opclass.oid</code></td>
<td>The index operator class this entry is for</td>
</tr>
<tr>
<td>amopsubtype</td>
<td>oid</td>
<td><code>pg_type.oid</code></td>
<td>Subtype to distinguish multiple entries for one strategy; zero for default</td>
</tr>
<tr>
<td>amopstrategy</td>
<td>int2</td>
<td></td>
<td>Operator strategy number</td>
</tr>
<tr>
<td>amopreqcheck</td>
<td>boolean</td>
<td></td>
<td>Index hit must be rechecked</td>
</tr>
<tr>
<td>amopopr</td>
<td>oid</td>
<td><code>pg_operator.oid</code></td>
<td>OID of the operator</td>
</tr>
</tbody>
</table>
pg_amproc

The `pg_amproc` table stores information about support procedures associated with index access method operator classes. There is one row for each support procedure belonging to an operator class.

Table H.1  pg_catalog.pg_amproc

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>amopclaid</td>
<td>oid</td>
<td><code>pg_opclass.oid</code></td>
<td>The index operator class this entry is for</td>
</tr>
<tr>
<td>amprocssubtype</td>
<td>oid</td>
<td><code>pg_type.oid</code></td>
<td>Subtype, if cross-type routine, else zero</td>
</tr>
<tr>
<td>amprocnum</td>
<td>int2</td>
<td></td>
<td>Support procedure number</td>
</tr>
<tr>
<td>amproc</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>OID of the procedure</td>
</tr>
</tbody>
</table>
The **pg_appendonly** table contains information about the storage options and other characteristics of append-only tables. This table is populated only on the master.

**Table H.1  pg_catalog.pgappendonly**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>relid</td>
<td>oid</td>
<td></td>
<td>The table object identifier (OID) of the compressed table.</td>
</tr>
<tr>
<td>blocksize</td>
<td>integer</td>
<td></td>
<td>Block size used for compression of append-only tables. Valid values are 8K - 2M. Default is 32K.</td>
</tr>
<tr>
<td>safefswritesize</td>
<td>integer</td>
<td></td>
<td>Minimum size for safe write operations to append-only tables in a non-mature file system. Commonly set to a multiple of the extent size of the file system; for example, Linux ext3 is 4096 bytes, so a value of 32768 is commonly used.</td>
</tr>
<tr>
<td>majorversion</td>
<td>smallint</td>
<td></td>
<td>The major version number of the pgappendonly table.</td>
</tr>
<tr>
<td>minorversion</td>
<td>smallint</td>
<td></td>
<td>The minor version number of the pgappendonly table.</td>
</tr>
<tr>
<td>checksum</td>
<td>boolean</td>
<td></td>
<td>A checksum value that is stored to compare the state of a block of data at compression time and at scan time to ensure data integrity. This data is stored only if gp_appendonly_verify_block_checksums is enabled (this parameter is disabled by default to optimize performance).</td>
</tr>
<tr>
<td>compressstype</td>
<td>text</td>
<td></td>
<td>Type of compression used to compress append-only tables. Valid values are zlib (gzip compression) and quicklz.</td>
</tr>
<tr>
<td>compresslevel</td>
<td>smallint</td>
<td></td>
<td>The compression level, with compression ratio increasing from 1 to 9. When quicklz is specified for compressstype, valid values are 1 or 3. With zlib specified, valid values are 1-9.</td>
</tr>
<tr>
<td>columnstore</td>
<td>boolean</td>
<td></td>
<td>1 for column-oriented storage, 0 for row-oriented storage.</td>
</tr>
<tr>
<td>segrelid</td>
<td>oid</td>
<td></td>
<td>Table on-disk segment file id.</td>
</tr>
<tr>
<td>segidxid</td>
<td>oid</td>
<td></td>
<td>Index on-disk segment file id.</td>
</tr>
<tr>
<td>column</td>
<td>type</td>
<td>references</td>
<td>description</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>------------</td>
<td>------------------------------------------------------</td>
</tr>
<tr>
<td>blkdirrelid</td>
<td>oid</td>
<td></td>
<td>Block used for on-disk column-oriented table file.</td>
</tr>
<tr>
<td>blkdiridxid</td>
<td>oid</td>
<td></td>
<td>Block used for on-disk column-oriented index file.</td>
</tr>
</tbody>
</table>
The `pg_attrdef` table stores column default values. The main information about columns is stored in `pg_attribute`. Only columns that explicitly specify a default value (when the table is created or the column is added) will have an entry here.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>adrelid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The table this column belongs to</td>
</tr>
<tr>
<td>adnum</td>
<td>int2</td>
<td><code>pg_attribute.attnum</code></td>
<td>The number of the column</td>
</tr>
<tr>
<td>adbin</td>
<td>text</td>
<td></td>
<td>The internal representation of the column default value</td>
</tr>
<tr>
<td>adsrc</td>
<td>text</td>
<td></td>
<td>A human-readable representation of the default value. This field is historical, and is best not used.</td>
</tr>
</tbody>
</table>
The `pg_attribute` table stores information about table columns. There will be exactly one `pg_attribute` row for every column in every table in the database. (There will also be attribute entries for indexes, and all objects that have `pg_class` entries.) The term attribute is equivalent to column.

**Table H.1  pg_catalog.pg_attribute**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>attrelid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The table this column belongs to</td>
</tr>
<tr>
<td>attname</td>
<td>name</td>
<td></td>
<td>The column name</td>
</tr>
<tr>
<td>atttypid</td>
<td>oid</td>
<td><code>pg_type.oid</code></td>
<td>The data type of this column</td>
</tr>
<tr>
<td>attstattarget</td>
<td>int4</td>
<td></td>
<td>Controls the level of detail of statistics accumulated for this column by <code>ANALYZE</code>. A zero value indicates that no statistics should be collected. A negative value says to use the system default statistics target. The exact meaning of positive values is data type-dependent. For scalar data types, it is both the target number of &quot;most common values&quot; to collect, and the target number of histogram bins to create.</td>
</tr>
<tr>
<td>attlen</td>
<td>int2</td>
<td></td>
<td>A copy of <code>pg_type.typlen</code> of this column's type</td>
</tr>
<tr>
<td>attnum</td>
<td>int2</td>
<td></td>
<td>The number of the column. Ordinary columns are numbered from 1 up. System columns, such as oid, have (arbitrary) negative numbers.</td>
</tr>
<tr>
<td>attndims</td>
<td>int4</td>
<td></td>
<td>Number of dimensions, if the column is an array type; otherwise 0. (Presently, the number of dimensions of an array is not enforced, so any nonzero value effectively means it is an array)</td>
</tr>
<tr>
<td>attcacheoff</td>
<td>int4</td>
<td></td>
<td>Always −1 in storage, but when loaded into a row descriptor in memory this may be updated to cache the offset of the attribute within the row</td>
</tr>
<tr>
<td>atttypmod</td>
<td>int4</td>
<td></td>
<td>Records type-specific data supplied at table creation time (for example, the maximum length of a varchar column). It is passed to type-specific input functions and length coercion functions. The value will generally be −1 for types that do not need it.</td>
</tr>
<tr>
<td>attbyval</td>
<td>boolean</td>
<td></td>
<td>A copy of <code>pg_type.typbyval</code> of this column's type</td>
</tr>
<tr>
<td>column</td>
<td>type</td>
<td>references</td>
<td>description</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>attstorage</td>
<td>char</td>
<td></td>
<td>Normally a copy of pg_type.typstorage of this column's type. For TOAST-able data types, this can be altered after column creation to control storage policy.</td>
</tr>
<tr>
<td>attalign</td>
<td>char</td>
<td></td>
<td>A copy of pg_type.typalign of this column's type</td>
</tr>
<tr>
<td>attnotnull</td>
<td>boolean</td>
<td></td>
<td>This represents a not-null constraint. It is possible to change this column to enable or disable the constraint.</td>
</tr>
<tr>
<td>atthasdef</td>
<td>boolean</td>
<td></td>
<td>This column has a default value, in which case there will be a corresponding entry in the pg_attrdef catalog that actually defines the value</td>
</tr>
<tr>
<td>attisdropped</td>
<td>boolean</td>
<td></td>
<td>This column has been dropped and is no longer valid. A dropped column is still physically present in the table, but is ignored by the parser and so cannot be accessed via SQL</td>
</tr>
<tr>
<td>attislocal</td>
<td>boolean</td>
<td></td>
<td>This column is defined locally in the relation. Note that a column may be locally defined and inherited simultaneously</td>
</tr>
<tr>
<td>attinhcount</td>
<td>int4</td>
<td></td>
<td>The number of direct ancestors this column has. A column with a nonzero number of ancestors cannot be dropped nor renamed</td>
</tr>
</tbody>
</table>
The `pg_auth_members` system catalog table shows the membership relations between roles. Any non-circular set of relationships is allowed. Because roles are system-wide, `pg_auth_members` is shared across all databases of a Greenplum Database system.

**Table H.1  pg_catalog.pg_auth_members**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>roleid</td>
<td>oid</td>
<td><code>pg_authid.oid</code></td>
<td>ID of the parent-level (group) role</td>
</tr>
<tr>
<td>member</td>
<td>oid</td>
<td><code>pg_authid.oid</code></td>
<td>ID of a member role</td>
</tr>
<tr>
<td>grantor</td>
<td>oid</td>
<td><code>pg_authid.oid</code></td>
<td>ID of the role that granted this membership</td>
</tr>
<tr>
<td>admin_option</td>
<td>boolean</td>
<td></td>
<td>True if role member may grant membership to others</td>
</tr>
</tbody>
</table>
**pg_authid**

The *pg_authid* table contains information about database authorization identifiers (roles). A role subsumes the concepts of users and groups. A user is a role with the *rolcanlogin* flag set. Any role (with or without *rolcanlogin*) may have other roles as members. See *pg_auth_members*.

Since this catalog contains passwords, it must not be publicly readable. *pg_roles* is a publicly readable view on *pg_authid* that blanks out the password field.

Because user identities are system-wide, *pg_authid* is shared across all databases in a Greenplum Database system: there is only one copy of *pg_authid* per system, not one per database.

**Table H.1  pg_catalog.pg_authid**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rolname</td>
<td>name</td>
<td></td>
<td>Role name</td>
</tr>
<tr>
<td>rolsuper</td>
<td>boolean</td>
<td></td>
<td>Role has superuser privileges</td>
</tr>
<tr>
<td>rolinherit</td>
<td>boolean</td>
<td></td>
<td>Role automatically inherits privileges of roles it is a member of</td>
</tr>
<tr>
<td>rolcreaterole</td>
<td>boolean</td>
<td></td>
<td>Role may create more roles</td>
</tr>
<tr>
<td>rolccreatedb</td>
<td>boolean</td>
<td></td>
<td>Role may create databases</td>
</tr>
<tr>
<td>rolcatupdate</td>
<td>boolean</td>
<td></td>
<td>Role may update system catalogs directly. (Even a superuser may not do this unless this column is true)</td>
</tr>
<tr>
<td>rolcantable</td>
<td>boolean</td>
<td></td>
<td>Role may log in. That is, this role can be given as the initial session authorization identifier</td>
</tr>
<tr>
<td>rolc◗nlimit</td>
<td>int4</td>
<td></td>
<td>For roles that can log in, this sets maximum number of concurrent connections this role can make. -1 means no limit</td>
</tr>
<tr>
<td>rolpASSWORD</td>
<td>text</td>
<td></td>
<td>Password (possibly encrypted); NULL if none</td>
</tr>
<tr>
<td>rolvaliduntil</td>
<td>timestampTZ</td>
<td></td>
<td>Password expiry time (only used for password authentication); NULL if no expiration</td>
</tr>
<tr>
<td>roleconfig</td>
<td>text[]</td>
<td></td>
<td>Session defaults for server configuration parameters</td>
</tr>
</tbody>
</table>
**pg_autovacuum**

The `pg_autovacuum` system catalog table stores optional per-relation configuration parameters for the autovacuum daemon. If there is an entry here for a particular relation, the given parameters will be used for autovacuuming that table. If no entry is present, the system-wide defaults will be used.

The autovacuum daemon will initiate a `VACUUM` operation on a particular table when the number of updated or deleted tuples exceeds `vac_base_thresh` plus `vac_scale_factor` times the number of live tuples currently estimated to be in the relation. Similarly, it will initiate an `ANALYZE` operation when the number of inserted, updated or deleted tuples exceeds `anl_base_thresh` plus `anl_scale_factor` times the number of live tuples currently estimated to be in the relation.

Also, the autovacuum daemon will perform a `VACUUM` operation to prevent transaction ID wraparound if the table’s `pg_class.relfrozenxid` field attains an age of more than `freeze_max_age` transactions, whether the table has been changed or not. The system will launch autovacuum to perform such `VACUUM`s even if autovacuum is otherwise disabled.

Any of the numerical fields can contain -1 to indicate that the system-wide default should be used for this particular value. Observe that the `vac_cost_delay` variable inherits its default value from the `autovacuum_vacuum_cost_delay` configuration parameter, or from `vacuum_cost_delay` if the former is set to a negative value. The same applies to `vac_cost_limit`. Also, autovacuum will ignore attempts to set a per-table `freeze_max_age` larger than the system-wide setting (it can only be set smaller), and the `freeze_min_age` value will be limited to half the system-wide `autovacuum_freeze_max_age` setting.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>vacrelid</code></td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The table this entry is for</td>
</tr>
<tr>
<td><code>enabled</code></td>
<td>boolean</td>
<td></td>
<td>If false, this table is never autovacuumed</td>
</tr>
<tr>
<td><code>vac_base_thresh</code></td>
<td>integer</td>
<td></td>
<td>Minimum number of modified tuples before vacuum</td>
</tr>
<tr>
<td><code>vac_scale_factor</code></td>
<td>float4</td>
<td></td>
<td>Multiplier for <code>reftuples</code> to add to <code>vac_base_thresh</code></td>
</tr>
<tr>
<td><code>anl_base_thresh</code></td>
<td>integer</td>
<td></td>
<td>Minimum number of modified tuples before analyze</td>
</tr>
<tr>
<td><code>anl_scale_factor</code></td>
<td>float4</td>
<td></td>
<td>Multiplier for <code>reftuples</code> to add to <code>anl_base_thresh</code></td>
</tr>
<tr>
<td><code>vac_cost_delay</code></td>
<td>integer</td>
<td></td>
<td>Custom <code>vacuum_cost_delay</code> parameter</td>
</tr>
<tr>
<td><code>vac_cost_limit</code></td>
<td>integer</td>
<td></td>
<td>Custom <code>vacuum_cost_limit</code> parameter</td>
</tr>
<tr>
<td><code>freeze_min_age</code></td>
<td>integer</td>
<td></td>
<td>Custom <code>vacuum_freeze_min_age</code> parameter</td>
</tr>
<tr>
<td><code>freeze_max_age</code></td>
<td>integer</td>
<td></td>
<td>Custom <code>autovacuum_freeze_max_age</code> parameter</td>
</tr>
</tbody>
</table>
The catalog *pg_cast* stores data type conversion paths, both built-in paths and those defined with `CREATE CAST`. The cast functions listed in *pg_cast* must always take the cast source type as their first argument type, and return the cast destination type as their result type. A cast function can have up to three arguments. The second argument, if present, must be type `integer`; it receives the type modifier associated with the destination type, or −1 if there is none. The third argument, if present, must be type `boolean`; it receives `true` if the cast is an explicit cast, `false` otherwise.

It is legitimate to create a *pg_cast* entry in which the source and target types are the same, if the associated function takes more than one argument. Such entries represent ‘length coercion functions’ that coerce values of the type to be legal for a particular type modifier value. Note however that at present there is no support for associating non-default type modifiers with user-created data types, and so this facility is only of use for the small number of built-in types that have type modifier syntax built into the grammar.

When a *pg_cast* entry has different source and target types and a function that takes more than one argument, it represents converting from one type to another and applying a length coercion in a single step. When no such entry is available, coercion to a type that uses a type modifier involves two steps, one to convert between data types and a second to apply the modifier.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>castsoure</td>
<td>oid</td>
<td><code>pg_type.oid</code></td>
<td>OID of the source data type.</td>
</tr>
<tr>
<td>casttarget</td>
<td>oid</td>
<td><code>pg_type.oid</code></td>
<td>OID of the target data type.</td>
</tr>
<tr>
<td>castfunc</td>
<td>oid</td>
<td><code>pg_proc.oid</code></td>
<td>The OID of the function to use to perform this cast. Zero is stored if the data types are binary compatible (that is, no run-time operation is needed to perform the cast).</td>
</tr>
<tr>
<td>castcontext</td>
<td>char</td>
<td></td>
<td>Indicates what contexts the cast may be invoked in. e means only as an explicit cast (using <code>CAST</code> or <code>::</code> syntax). a means implicitly in assignment to a target column, as well as explicitly. i means implicitly in expressions, as well as the other cases.</td>
</tr>
</tbody>
</table>
**pg_class**

The system catalog table *pg_class* catalogs tables and most everything else that has columns or is otherwise similar to a table (also known as *relations*). This includes indexes (see also *pg_index*), sequences, views, composite types, and TOAST tables. Not all columns are meaningful for all relation types.

**Table H.1  pg_catalog.pg_class**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>relname</td>
<td>name</td>
<td></td>
<td>Name of the table, index, view, etc.</td>
</tr>
<tr>
<td>relnamespace</td>
<td>oid</td>
<td><em>pg_namespace.oid</em></td>
<td>The OID of the namespace (schema) that contains this relation</td>
</tr>
<tr>
<td>reltype</td>
<td>oid</td>
<td><em>pg_type.oid</em></td>
<td>The OID of the data type that corresponds to this table’s row type, if any (zero for indexes, which have no <em>pg_type</em> entry)</td>
</tr>
<tr>
<td>relowner</td>
<td>oid</td>
<td><em>pg_authid.oid</em></td>
<td>Owner of the relation</td>
</tr>
<tr>
<td>relam</td>
<td>oid</td>
<td><em>pg_am.oid</em></td>
<td>If this is an index, the access method used (B-tree, Bitmap, hash, etc.)</td>
</tr>
<tr>
<td>relfilenode</td>
<td>oid</td>
<td></td>
<td>Name of the on-disk file of this relation; 0 if none.</td>
</tr>
<tr>
<td>reltablespace</td>
<td>oid</td>
<td><em>pg_tablespace.oid</em></td>
<td>The tablespace in which this relation is stored. If zero, the database’s default tablespace is implied. (Not meaningful if the relation has no on-disk file.)</td>
</tr>
<tr>
<td>repages</td>
<td>int4</td>
<td></td>
<td>Size of the on-disk representation of this table in pages (of 32K each). This is only an estimate used by the planner. It is updated by VACUUM, ANALYZE, and a few DDL commands.</td>
</tr>
<tr>
<td>reltuples</td>
<td>float4</td>
<td></td>
<td>Number of rows in the table. This is only an estimate used by the planner. It is updated by VACUUM, ANALYZE, and a few DDL commands.</td>
</tr>
<tr>
<td>reltoastrelid</td>
<td>oid</td>
<td><em>pg_class.oid</em></td>
<td>OID of the TOAST table associated with this table, 0 if none. The TOAST table stores large attributes “out of line” in a secondary table.</td>
</tr>
<tr>
<td>reltoastidxid</td>
<td>oid</td>
<td><em>pg_class.oid</em></td>
<td>For a TOAST table, the OID of its index. 0 if not a TOAST table.</td>
</tr>
<tr>
<td>relaosegidxid</td>
<td>oid</td>
<td></td>
<td>Deprecated in Greenplum Database 3.4.</td>
</tr>
<tr>
<td>relaosegrelid</td>
<td>oid</td>
<td></td>
<td>Deprecated in Greenplum Database 3.4.</td>
</tr>
</tbody>
</table>
### Table H.1  pg_catalog.pg_class

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>relhasindex</td>
<td>boolean</td>
<td></td>
<td>True if this is a table and it has (or recently had) any indexes. This is set</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>by <code>CREATE INDEX</code>, but not cleared immediately by <code>DROP INDEX</code>. <code>VACUUM</code> will</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>clear if it finds the table has no indexes.</td>
</tr>
<tr>
<td>relisshared</td>
<td>boolean</td>
<td></td>
<td>True if this table is shared across all databases in the system. Only certain</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>system catalog tables are shared.</td>
</tr>
<tr>
<td>relkind</td>
<td>char</td>
<td></td>
<td>The type of object <code>r</code> = heap or append-only table, <code>i</code> = index, <code>S</code> = sequence, <code>v</code> = view, <code>c</code> = composite type, <code>t</code> = TOAST value, <code>o</code> = internal append-only segment files and EOFs, <code>c</code> = composite type, <code>u</code> = uncataloged temporary heap table</td>
</tr>
<tr>
<td>relstorage</td>
<td>char</td>
<td></td>
<td>The storage mode of a table <code>a</code> = append-only, <code>h</code> = heap, <code>v</code> = virtual, <code>x</code> = external table.</td>
</tr>
<tr>
<td>relnatts</td>
<td>int2</td>
<td></td>
<td>Number of user columns in the relation (system columns not counted). There must</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>be this many corresponding entries in <code>pg_attribute</code>.</td>
</tr>
<tr>
<td>relchecks</td>
<td>int2</td>
<td></td>
<td>Number of check constraints on the table.</td>
</tr>
<tr>
<td>reltriggers</td>
<td>int2</td>
<td></td>
<td>Number of triggers on the table.</td>
</tr>
<tr>
<td>relukeys</td>
<td>int2</td>
<td></td>
<td>Unused</td>
</tr>
<tr>
<td>relfkeys</td>
<td>int2</td>
<td></td>
<td>Unused</td>
</tr>
<tr>
<td>relrefs</td>
<td>int2</td>
<td></td>
<td>Unused</td>
</tr>
<tr>
<td>relhasoids</td>
<td>boolean</td>
<td></td>
<td>True if an OID is generated for each row of the relation.</td>
</tr>
<tr>
<td>relhaspkey</td>
<td>boolean</td>
<td></td>
<td>True if the table has (or once had) a primary key.</td>
</tr>
<tr>
<td>relhasrules</td>
<td>boolean</td>
<td></td>
<td>True if table has rules.</td>
</tr>
<tr>
<td>relhassubclass</td>
<td>boolean</td>
<td></td>
<td>True if table has (or once had) any inheritance children.</td>
</tr>
<tr>
<td>column</td>
<td>type</td>
<td>references</td>
<td>description</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------</td>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>relfrozenxid</td>
<td>xid</td>
<td></td>
<td>All transaction IDs before this one have been replaced with a permanent (frozen) transaction ID in this table. This is used to track whether the table needs to be vacuumed in order to prevent transaction ID wraparound or to allow <code>pg_clog</code> to be shrunk. Zero (<code>InvalidTransactionId</code>) if the relation is not a table.</td>
</tr>
<tr>
<td>relacl</td>
<td>aclitem[]</td>
<td></td>
<td>Access privileges assigned by <code>GRANT</code> and <code>REVOKE</code>.</td>
</tr>
<tr>
<td>reloptions</td>
<td>text[]</td>
<td></td>
<td>Access-method-specific options, as “keyword=value” strings.</td>
</tr>
</tbody>
</table>
The `pg_constraint` system catalog table stores check, primary key, unique, and foreign key constraints on tables. Column constraints are not treated specially. Every column constraint is equivalent to some table constraint. Not-null constraints are represented in the `pg_attribute` catalog. Check constraints on domains are stored here, too.

**Table H.1  pg_catalog.pg_constraint**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>conname</td>
<td>name</td>
<td></td>
<td>Constraint name (not necessarily unique!)</td>
</tr>
<tr>
<td>connamespace</td>
<td>oid</td>
<td><code>pg_namespace.oid</code></td>
<td>The OID of the namespace (schema) that contains this constraint.</td>
</tr>
<tr>
<td>contype</td>
<td>char</td>
<td></td>
<td><code>c</code> = check constraint, <code>f</code> = foreign key constraint, <code>p</code> = primary key constraint, <code>u</code> = unique constraint.</td>
</tr>
<tr>
<td>condeferrable</td>
<td>boolean</td>
<td></td>
<td>Is the constraint deferrable?</td>
</tr>
<tr>
<td>condeferred</td>
<td>boolean</td>
<td></td>
<td>Is the constraint deferred by default?</td>
</tr>
<tr>
<td>conrelid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The table this constraint is on; 0 if not a table constraint.</td>
</tr>
<tr>
<td>contypid</td>
<td>oid</td>
<td><code>pg_type.oid</code></td>
<td>The domain this constraint is on; 0 if not a domain constraint.</td>
</tr>
<tr>
<td>confrelid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>If a foreign key, the referenced table; else 0.</td>
</tr>
<tr>
<td>confupdtype</td>
<td>char</td>
<td></td>
<td>Foreign key update action code.</td>
</tr>
<tr>
<td>confdeltype</td>
<td>char</td>
<td></td>
<td>Foreign key deletion action code.</td>
</tr>
<tr>
<td>confmatchtype</td>
<td>char</td>
<td></td>
<td>Foreign key match type.</td>
</tr>
<tr>
<td>conkey</td>
<td>int2[]</td>
<td><code>pg_attribute.attnum</code></td>
<td>If a table constraint, list of columns which the constraint constrains.</td>
</tr>
<tr>
<td>confkey</td>
<td>int2[]</td>
<td><code>pg_attribute.attnum</code></td>
<td>If a foreign key, list of the referenced columns.</td>
</tr>
<tr>
<td>conbin</td>
<td>text</td>
<td></td>
<td>If a check constraint, an internal representation of the expression.</td>
</tr>
<tr>
<td>consrc</td>
<td>text</td>
<td></td>
<td>If a check constraint, a human-readable representation of the expression. This is not updated when referenced objects change; for example, it won’t track renaming of columns. Rather than relying on this field, it is best to use <code>pg_get_constraintdef()</code> to extract the definition of a check constraint.</td>
</tr>
</tbody>
</table>
The `pg_conversion` system catalog table describes the available encoding conversion procedures as defined by `CREATE CONVERSION`.

**Table H.1  pg_catalog.pg_conversion**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>conname</td>
<td>name</td>
<td></td>
<td>Conversion name (unique within a namespace).</td>
</tr>
<tr>
<td>connamespace</td>
<td>oid</td>
<td><code>pg_namespace.oid</code></td>
<td>The OID of the namespace (schema) that contains this conversion.</td>
</tr>
<tr>
<td>conowner</td>
<td>oid</td>
<td><code>pg_authid.oid</code></td>
<td>Owner of the conversion.</td>
</tr>
<tr>
<td>conforencoding</td>
<td>int4</td>
<td></td>
<td>Source encoding ID.</td>
</tr>
<tr>
<td>contoencoding</td>
<td>int4</td>
<td></td>
<td>Destination encoding ID.</td>
</tr>
<tr>
<td>conproc</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>Conversion procedure.</td>
</tr>
<tr>
<td>condefault</td>
<td>boolean</td>
<td></td>
<td>True if this is the default conversion.</td>
</tr>
</tbody>
</table>
The *pg_database* system catalog table stores information about the available databases. Databases are created with the `CREATE DATABASE` SQL command. Unlike most system catalogs, *pg_database* is shared across all databases in the system. There is only one copy of *pg_database* per system, not one per database.

**Table H.1  pg_catalog.pg_database**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>datname</td>
<td>name</td>
<td></td>
<td>Database name.</td>
</tr>
<tr>
<td>datdba</td>
<td>oid</td>
<td><em>pg_authid.oid</em></td>
<td>Owner of the database, usually the user who created it.</td>
</tr>
<tr>
<td>encoding</td>
<td>int4</td>
<td></td>
<td>Character encoding for this database. <em>pg_encoding_to_char()</em> can translate this number to the encoding name.</td>
</tr>
<tr>
<td>datistemplate</td>
<td>boolean</td>
<td></td>
<td>If true then this database can be used in the <code>TEMPLATE</code> clause of <code>CREATE DATABASE</code> to create a new database as a clone of this one.</td>
</tr>
<tr>
<td>datallowconn</td>
<td>boolean</td>
<td></td>
<td>If false then no one can connect to this database. This is used to protect the <code>template0</code> database from being altered.</td>
</tr>
<tr>
<td>datconnlimit</td>
<td>int4</td>
<td></td>
<td>Sets the maximum number of concurrent connections that can be made to this database. <code>-1</code> means no limit.</td>
</tr>
<tr>
<td>datlastsysoid</td>
<td>oid</td>
<td></td>
<td>Last system OID in the database; useful particularly to <code>pg_dump/gp_dump</code>.</td>
</tr>
<tr>
<td>datfrozenxid</td>
<td>xid</td>
<td></td>
<td>All transaction IDs before this one have been replaced with a permanent (frozen) transaction ID in this database. This is used to track whether the database needs to be vacuumed in order to prevent transaction ID wraparound or to allow <code>pg_clog</code> to be shrunk. It is the minimum of the per-table <code>pg_class.relfrozenxid</code> values.</td>
</tr>
<tr>
<td>dattablespace</td>
<td>oid</td>
<td><em>pg_tablespace.oid</em></td>
<td>The default tablespace for the database. Within this database, all tables for which <code>pg_class.reltablespace</code> is zero will be stored in this tablespace. All non-shared system catalogs will also be there.</td>
</tr>
</tbody>
</table>
### Table H.1  pg_catalog.pg_database

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>datconfig</td>
<td>text[]</td>
<td></td>
<td>Session defaults for user-settable server configuration parameters.</td>
</tr>
<tr>
<td>datacl</td>
<td>aclitem[]</td>
<td></td>
<td>Database access privileges as given by <code>GRANT</code> and <code>REVOKE</code>.</td>
</tr>
</tbody>
</table>
The `pg_depend` system catalog table records the dependency relationships between database objects. This information allows `DROP` commands to find which other objects must be dropped by `DROP CASCADE` or prevent dropping in the `DROP RESTRICT` case. See also `pg_shdepend`, which performs a similar function for dependencies involving objects that are shared across a Greenplum system.

In all cases, a `pg_depend` entry indicates that the referenced object may not be dropped without also dropping the dependent object. However, there are several subflavors identified by `deptype`:

- **DEPENDENCY_NORMAL (n)** — A normal relationship between separately-created objects. The dependent object may be dropped without affecting the referenced object. The referenced object may only be dropped by specifying `CASCADE`, in which case the dependent object is dropped, too. Example: a table column has a normal dependency on its data type.

- **DEPENDENCY_AUTO (a)** — The dependent object can be dropped separately from the referenced object, and should be automatically dropped (regardless of `RESTRICT` or `CASCADE` mode) if the referenced object is dropped. Example: a named constraint on a table is made autodependent on the table, so that it will go away if the table is dropped.

- **DEPENDENCY_INTERNAL (i)** — The dependent object was created as part of creation of the referenced object, and is really just a part of its internal implementation. A `DROP` of the dependent object will be disallowed outright (we’ll tell the user to issue a `DROP` against the referenced object, instead). A `DROP` of the referenced object will be propagated through to drop the dependent object whether `CASCADE` is specified or not. Example: a trigger that’s created to enforce a foreign-key constraint is made internally dependent on the constraint’s `pg_constraint` entry.

- **DEPENDENCY_PIN (p)** — There is no dependent object; this type of entry is a signal that the system itself depends on the referenced object, and so that object must never be deleted. Entries of this type are created only by system initialization. The columns for the dependent object contain zeroes.

### Table H.1  pg_catalog.pg_depend

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>classid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The OID of the system catalog the dependent object is in.</td>
</tr>
<tr>
<td>objid</td>
<td>oid</td>
<td>any OID column</td>
<td>The OID of the specific dependent object.</td>
</tr>
<tr>
<td>objsubid</td>
<td>int4</td>
<td></td>
<td>For a table column, this is the column number. For all other object types, this column is zero.</td>
</tr>
<tr>
<td>refclassid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The OID of the system catalog the referenced object is in.</td>
</tr>
<tr>
<td>refobjid</td>
<td>oid</td>
<td>any OID column</td>
<td>The OID of the specific referenced object.</td>
</tr>
<tr>
<td>refobjsubid</td>
<td>int4</td>
<td></td>
<td>For a table column, this is the referenced column number. For all other object types, this column is zero.</td>
</tr>
<tr>
<td>deptype</td>
<td>char</td>
<td></td>
<td>A code defining the specific semantics of this dependency relationship.</td>
</tr>
</tbody>
</table>
**pg_description**

The `pg_description` system catalog table stores optional descriptions (comments) for each database object. Descriptions can be manipulated with the `COMMENT` command and viewed with `psql`'s `\d` meta-commands. Descriptions of many built-in system objects are provided in the initial contents of `pg_description`. See also `pg_shdescription`, which performs a similar function for descriptions involving objects that are shared across a Greenplum system.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>objoid</td>
<td>oid</td>
<td>any OID column</td>
<td>The OID of the object this description pertains to.</td>
</tr>
<tr>
<td>classoid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The OID of the system catalog this object appears in</td>
</tr>
<tr>
<td>objsubid</td>
<td>int4</td>
<td></td>
<td>For a comment on a table column, this is the column number. For all other object types, this column is zero.</td>
</tr>
<tr>
<td>description</td>
<td>text</td>
<td></td>
<td>Arbitrary text that serves as the description of this object.</td>
</tr>
</tbody>
</table>
**pg_exttable**

The `pg_exttable` system catalog table is used to track external tables and web tables created by the `CREATE EXTERNAL TABLE` command.

**Table H.1  pg_catalog.pg_exttable**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>reloid</td>
<td>oid</td>
<td>pg_class.oid</td>
<td>The OID of this external table.</td>
</tr>
<tr>
<td>location</td>
<td>text[]</td>
<td></td>
<td>The URI location(s) of the external table files.</td>
</tr>
<tr>
<td>fmttype</td>
<td>char</td>
<td></td>
<td>Format of the external table files: ( t ) for text, or ( c ) for csv.</td>
</tr>
</tbody>
</table>
| fmtopts   | text       |            | Formatting options of the external table files, such as the field delimiter,
|           |            |            | null string, escape character, etc.                                        |
| command   | text       |            | The OS command to execute when the external table is accessed.             |
| rejectlimit | integer   |            | The per segment reject limit for rows with errors, after which the load will fail. |
| rejectlimittype | char   |            | Type of reject limit threshold: \( r \) for number of rows.               |
| fmterrtbl | oid        | pg_class.oid | The object id of the error table where format errors will be logged.       |
| encoding  | text       |            | The client encoding.                                                       |
| writable  | boolean    |            | 0 for readable external tables, 1 for writable external tables.             |
The `pg_filespace` table contains information about the filespaces created in a Greenplum Database system. Every system contains a default filespace, `pg_system`, which is a collection of all the data directory locations created at system initialization time.

A tablespace requires a file system location to store its database files. In Greenplum Database, the master and each segment (primary and mirror) needs its own distinct storage location. This collection of file system locations for all components in a Greenplum system is referred to as a filespace.

**Table H.1  pg_catalog.pg_filespace**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>fsname</td>
<td>name</td>
<td></td>
<td>The name of the filespace.</td>
</tr>
<tr>
<td>fsowner</td>
<td>oid</td>
<td>pg_roles.oid</td>
<td>The object id of the role that created the filespace.</td>
</tr>
</tbody>
</table>
A tablespace requires a file system location to store its database files. In Greenplum Database, the master and each segment (primary and mirror) needs its own distinct storage location. This collection of file system locations for all components in a Greenplum system is referred to as a *filespace*. The *pg_filespace_entry* table contains information about the collection of file system locations across a Greenplum Database system that comprise a Greenplum Database filespace.

**Table H.1  pg_catalog.pg_filespace_entry**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>fsefsoid</td>
<td>OID</td>
<td>pg_filespace.oid</td>
<td>Object id of the filespace.</td>
</tr>
<tr>
<td>fsedbid</td>
<td>integer</td>
<td>gp_segment_configuration.dbid</td>
<td>Segment id.</td>
</tr>
<tr>
<td>fselocation</td>
<td>text</td>
<td></td>
<td>File system location for this segment id.</td>
</tr>
</tbody>
</table>
The `pg_index` system catalog table contains part of the information about indexes. The rest is mostly in `pg_class`.

### Table H.1  pg_catalog.pg_index

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>indexrelid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The OID of the <code>pg_class</code> entry for this index.</td>
</tr>
<tr>
<td>indrelid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The OID of the <code>pg_class</code> entry for the table this index is for.</td>
</tr>
<tr>
<td>indnatts</td>
<td>int2</td>
<td></td>
<td>The number of columns in the index (duplicates <code>pg_class.relnatts</code>).</td>
</tr>
<tr>
<td>indisunique</td>
<td>boolean</td>
<td></td>
<td>If true, this is a unique index.</td>
</tr>
</tbody>
</table>
| indisprimary | boolean   |             | If true, this index represents the primary key of the table.  
(`indisunique` should always be true when this is true.) |
| indisclustered | boolean |             | If true, the table was last clustered on this index via the `CLUSTER` command. |
| indisvalid | boolean    |             | If true, the index is currently valid for queries. False means the index is possibly incomplete: it must still be modified by `INSERT/UPDATE` operations, but it cannot safely be used for queries. |
| indkey   | int2vector | `pg_attribute.attnum` | This is an array of `indnatts` values that indicate which table columns this index indexes. For example a value of 1 3 would mean that the first and the third table columns make up the index key. A zero in this array indicates that the corresponding index attribute is an expression over the table columns, rather than a simple column reference. |
| indclass | oidvector  | `pg_opclass.oid` | For each column in the index key this contains the OID of the operator class to use. |
### Table H.1  pg_catalog.pg_index

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>indexprs</td>
<td>text</td>
<td></td>
<td>Expression trees (in nodeToString() representation) for index attributes that are not simple column references. This is a list with one element for each zero entry in <code>indkey</code>. NULL if all index attributes are simple references.</td>
</tr>
<tr>
<td>indpred</td>
<td>text</td>
<td></td>
<td>Expression tree (in nodeToString() representation) for partial index predicate. NULL if not a partial index.</td>
</tr>
</tbody>
</table>
**pg_inherits**

The `pg_inherits` system catalog table records information about table inheritance hierarchies. There is one entry for each direct child table in the database. (Indirect inheritance can be determined by following chains of entries.) In Greenplum Database, inheritance relationships are created by both the `INHERITS` clause (standalone inheritance) and the `PARTITION BY` clause (partitioned child table inheritance) of `CREATE TABLE`.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>inhrelid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The OID of the child table.</td>
</tr>
<tr>
<td>inhparent</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The OID of the parent table.</td>
</tr>
<tr>
<td>inhseqno</td>
<td>int4</td>
<td></td>
<td>If there is more than one direct parent for a child table (multiple inheritance), this number tells the order in which the inherited columns are to be arranged. The count starts at 1.</td>
</tr>
</tbody>
</table>
The `pg_language` system catalog table registers languages in which you can write functions or stored procedures. It is populated by `CREATE LANGUAGE`.

Table H.1  `pg_catalog.pg_language`

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lanname</td>
<td>name</td>
<td></td>
<td>Name of the language.</td>
</tr>
<tr>
<td>lanispl</td>
<td>boolean</td>
<td></td>
<td>This is false for internal languages (such as SQL) and true for user-defined languages. Currently, <code>pg_dump</code> still uses this to determine which languages need to be dumped, but this may be replaced by a different mechanism in the future.</td>
</tr>
<tr>
<td>lanpltrusted</td>
<td>boolean</td>
<td></td>
<td>True if this is a trusted language, which means that it is believed not to grant access to anything outside the normal SQL execution environment. Only superusers may create functions in untrusted languages.</td>
</tr>
<tr>
<td>lanplcallfoid</td>
<td>oid</td>
<td><code>pg_proc.oid</code></td>
<td>For noninternal languages this references the language handler, which is a special function that is responsible for executing all functions that are written in the particular language.</td>
</tr>
<tr>
<td>lanvalidator</td>
<td>oid</td>
<td><code>pg_proc.oid</code></td>
<td>This references a language validator function that is responsible for checking the syntax and validity of new functions when they are created. Zero if no validator is provided.</td>
</tr>
<tr>
<td>lanacl</td>
<td>aclitem[]</td>
<td></td>
<td>Access privileges for the language.</td>
</tr>
</tbody>
</table>
The `pg_largeobject` system catalog table holds the data making up ‘large objects’. A large object is identified by an OID assigned when it is created. Each large object is broken into segments or ‘pages’ small enough to be conveniently stored as rows in `pg_largeobject`. The amount of data per page is defined to be `LOBLKSIZE` (which is currently `BLCKSZ/4`, or typically 8K).

Each row of `pg_largeobject` holds data for one page of a large object, beginning at byte offset (`pageno * LOBLKSIZE`) within the object. The implementation allows sparse storage: pages may be missing, and may be shorter than `LOBLKSIZE` bytes even if they are not the last page of the object. Missing regions within a large object read as zeroes.

**Table H.1  pg_catalog.pg_largeobject**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>loid</td>
<td>oid</td>
<td></td>
<td>Identifier of the large object that includes this page.</td>
</tr>
<tr>
<td>pageno</td>
<td>int4</td>
<td></td>
<td>Page number of this page within its large object (counting from zero).</td>
</tr>
<tr>
<td>data</td>
<td>bytea</td>
<td></td>
<td>Actual data stored in the large object. This will never be more than <code>LOBLKSIZE</code> bytes and may be less.</td>
</tr>
</tbody>
</table>
The `pg_listener` system catalog table supports the `LISTEN` and `NOTIFY` commands. A listener creates an entry in `pg_listener` for each notification name it is listening for. A notifier scans and updates each matching entry to show that a notification has occurred. The notifier also sends a signal (using the PID recorded in the table) to awaken the listener from sleep.

This table is not currently used in Greenplum Database.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>relname</td>
<td>name</td>
<td></td>
<td>Notify condition name. (The name need not match any actual relation in the database.</td>
</tr>
</tbody>
</table>
The view `pg_locks` provides access to information about the locks held by open transactions within Greenplum Database.

`pg_locks` contains one row per active lockable object, requested lock mode, and relevant transaction. Thus, the same lockable object may appear many times, if multiple transactions are holding or waiting for locks on it. However, an object that currently has no locks on it will not appear at all.

There are several distinct types of lockable objects: whole relations (such as tables), individual pages of relations, individual tuples of relations, transaction IDs, and general database objects. Also, the right to extend a relation is represented as a separate lockable object.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>locktype</td>
<td>text</td>
<td></td>
<td>Type of the lockable object: relation, extend, page, tuple, transactionid, object, userlock, resource queue, or advisory</td>
</tr>
<tr>
<td>database</td>
<td>oid</td>
<td><code>pg_database.oid</code></td>
<td>OID of the database in which the object exists, zero if the object is a shared object, or NULL if the object is a transaction ID</td>
</tr>
<tr>
<td>relation</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>OID of the relation, or NULL if the object is not a relation or part of a relation</td>
</tr>
<tr>
<td>page</td>
<td>integer</td>
<td></td>
<td>Page number within the relation, or NULL if the object is not a tuple or relation page</td>
</tr>
<tr>
<td>tuple</td>
<td>smallint</td>
<td></td>
<td>Tuple number within the page, or NULL if the object is not a tuple</td>
</tr>
<tr>
<td>transactionid</td>
<td>xid</td>
<td></td>
<td>ID of a transaction, or NULL if the object is not a transaction ID</td>
</tr>
<tr>
<td>classid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>OID of the system catalog containing the object, or NULL if the object is not a general database object</td>
</tr>
<tr>
<td>objid</td>
<td>oid</td>
<td>any OID column</td>
<td>OID of the object within its system catalog, or NULL if the object is not a general database object</td>
</tr>
<tr>
<td>objsubid</td>
<td>smallint</td>
<td></td>
<td>For a table column, this is the column number (the <code>classid</code> and <code>objid</code> refer to the table itself). For all other object types, this column is zero. NULL if the object is not a general database object</td>
</tr>
<tr>
<td>transaction</td>
<td>xid</td>
<td></td>
<td>ID of the transaction that is holding or awaiting this lock</td>
</tr>
</tbody>
</table>
### Table H.1 pg_catalog.pg_locks

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pid</td>
<td>integer</td>
<td></td>
<td>Process ID of the server process holding or awaiting this lock. NULL if the lock is held by a prepared transaction</td>
</tr>
<tr>
<td>mode</td>
<td>text</td>
<td></td>
<td>Name of the lock mode held or desired by this process</td>
</tr>
<tr>
<td>granted</td>
<td>boolean</td>
<td></td>
<td>True if lock is held, false if lock is awaited</td>
</tr>
</tbody>
</table>
**pg_namespace**

The `pg_namespace` system catalog table stores namespaces. A namespace is the structure underlying SQL schemas: each namespace can have a separate collection of relations, types, etc. without name conflicts.

### Table H.1  pg_catalog.pg_namespace

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>nspname</td>
<td>name</td>
<td></td>
<td>Name of the namespace</td>
</tr>
<tr>
<td>nspowner</td>
<td>oid</td>
<td>pg_authid.oid</td>
<td>Owner of the namespace</td>
</tr>
<tr>
<td>nspacl</td>
<td>aclitem[]</td>
<td></td>
<td>Access privileges as given by GRANT and REVOKE.</td>
</tr>
</tbody>
</table>
**pg_opclass**

The `pg_opclass` system catalog table defines index access method operator classes. Each operator class defines semantics for index columns of a particular data type and a particular index access method. Note that there can be multiple operator classes for a given data type/access method combination, thus supporting multiple behaviors. The majority of the information defining an operator class is actually not in its `pg_opclass` row, but in the associated rows in `pg_amop` and `pg_amproc`. Those rows are considered to be part of the operator class definition — this is not unlike the way that a relation is defined by a single `pg_class` row plus associated rows in `pg_attribute` and other tables.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>opcamid</td>
<td>oid</td>
<td><code>pg_am.oid</code></td>
<td>Index access method operator class is for.</td>
</tr>
<tr>
<td>opcname</td>
<td>name</td>
<td></td>
<td>Name of this operator class</td>
</tr>
<tr>
<td>opcnamespace</td>
<td>oid</td>
<td><code>pg_namespace.oid</code></td>
<td>Namespace of this operator class</td>
</tr>
<tr>
<td>opcowner</td>
<td>oid</td>
<td><code>pg_authid.oid</code></td>
<td>Owner of the operator class</td>
</tr>
<tr>
<td>opcintype</td>
<td>oid</td>
<td><code>pg_type.oid</code></td>
<td>Data type that the operator class indexes.</td>
</tr>
<tr>
<td>opcvdefault</td>
<td>boolean</td>
<td></td>
<td>True if this operator class is the default for the data type <code>opcintype</code>.</td>
</tr>
<tr>
<td>opckeytype</td>
<td>oid</td>
<td><code>pg_type.oid</code></td>
<td>Type of data stored in index, or zero if same as <code>opcintype</code>.</td>
</tr>
</tbody>
</table>
The *pg_operator* system catalog table stores information about operators, both built-in and those defined by `CREATE OPERATOR`. Unused column contain zeroes. For example, `oprleft` is zero for a prefix operator.

**Table H.1  pg_catalog.pg_operator**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>oprname</td>
<td>name</td>
<td></td>
<td>Name of the operator.</td>
</tr>
<tr>
<td>opnamespace</td>
<td>oid</td>
<td>pg_namespace.oid</td>
<td>The OID of the namespace that contains this operator.</td>
</tr>
<tr>
<td>oprowner</td>
<td>oid</td>
<td>pg_authid.oid</td>
<td>Owner of the operator.</td>
</tr>
<tr>
<td>oprkind</td>
<td>char</td>
<td></td>
<td><code>b</code> = infix (both), <code>l</code> = prefix (left), <code>r</code> = postfix (right)</td>
</tr>
<tr>
<td>oprcanhash</td>
<td>boolean</td>
<td></td>
<td>This operator supports hash joins.</td>
</tr>
<tr>
<td>oprleft</td>
<td>oid</td>
<td>pg_type.oid</td>
<td>Type of the left operand.</td>
</tr>
<tr>
<td>oprright</td>
<td>oid</td>
<td>pg_type.oid</td>
<td>Type of the right operand.</td>
</tr>
<tr>
<td>oprresult</td>
<td>oid</td>
<td>pg_type.oid</td>
<td>Type of the result.</td>
</tr>
<tr>
<td>oprcom</td>
<td>oid</td>
<td>pg_operator.oid</td>
<td>Commutator of this operator, if any.</td>
</tr>
<tr>
<td>opmlegate</td>
<td>oid</td>
<td>pg_operator.oid</td>
<td>Negator of this operator, if any.</td>
</tr>
<tr>
<td>oprlsortop</td>
<td>oid</td>
<td>pg_operator.oid</td>
<td>If this operator supports merge joins, the operator that sorts the type of the left-hand operand (L&lt;L).</td>
</tr>
<tr>
<td>oprrsortop l</td>
<td>oid</td>
<td>pg_operator.oid</td>
<td>If this operator supports merge joins, the operator that sorts the type of the right-hand operand (R&lt;R).</td>
</tr>
<tr>
<td>oprltcmpop</td>
<td>oid</td>
<td>pg_operator.oid</td>
<td>If this operator supports merge joins, the less-than operator that compares the left and right operand types (L&lt;R).</td>
</tr>
<tr>
<td>oprgtcmpop</td>
<td>oid</td>
<td>pg_operator.oid</td>
<td>If this operator supports merge joins, the greater-than operator that compares the left and right operand types (L&gt;R).</td>
</tr>
<tr>
<td>oprcode</td>
<td>regproc</td>
<td>pg_proc.oid</td>
<td>Function that implements this operator.</td>
</tr>
<tr>
<td>oprrest</td>
<td>regproc</td>
<td>pg_proc.oid</td>
<td>Restriction selectivity estimation function for this operator.</td>
</tr>
<tr>
<td>oprjoin</td>
<td>regproc</td>
<td>pg_proc.oid</td>
<td>Join selectivity estimation function for this operator.</td>
</tr>
</tbody>
</table>
**pg_partition**

The `pg_partition` system catalog table is used to track partitioned tables and their inheritance level relationships. Each row of `pg_partition` represents either the level of a partitioned table in the partition hierarchy, or a subpartition template description. The value of the attribute `paristemplate` determines what a particular row represents.

**Table H.1  pg_catalog.pg_partition**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>parrelid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The object identifier of the table.</td>
</tr>
<tr>
<td>parkind</td>
<td>char</td>
<td></td>
<td>The partition type - R for range or L for list.</td>
</tr>
<tr>
<td>parlevel</td>
<td>smallint</td>
<td></td>
<td>The partition level of this row: 0 for the top-level parent table, 1 for the first level under the parent table, 2 for the second level, and so on.</td>
</tr>
<tr>
<td>paristemplate</td>
<td>boolean</td>
<td></td>
<td>Whether or not this row represents a subpartition template definition (true) or an actual partitioning level (false).</td>
</tr>
<tr>
<td>parnatts</td>
<td>smallint</td>
<td><code>pg_attribute.oid</code></td>
<td>The number of attributes that define this level.</td>
</tr>
<tr>
<td>paratts</td>
<td>smallint()</td>
<td></td>
<td>An array of the attribute numbers (as in <code>pg_attribute.attnum</code>) of the attributes that participate in defining this level.</td>
</tr>
<tr>
<td>parclass</td>
<td>oidvector</td>
<td><code>pg_opclass.oid</code></td>
<td>The operator class identifier(s) of the partition columns.</td>
</tr>
</tbody>
</table>
**pg_partition_columns**

The *pg_partition_columns* system view is used to show the partition key columns of a partitioned table.

**Table H.1  pg_catalog.pg_partition_columns**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>schemaname</td>
<td>name</td>
<td></td>
<td>The name of the schema the partitioned table is in.</td>
</tr>
<tr>
<td>tablename</td>
<td>name</td>
<td></td>
<td>The table name of the top-level parent table.</td>
</tr>
<tr>
<td>columnname</td>
<td>name</td>
<td></td>
<td>The name of the partition key column.</td>
</tr>
<tr>
<td>partitionlevel</td>
<td>smallint</td>
<td></td>
<td>The level of this subpartition in the hierarchy.</td>
</tr>
<tr>
<td>position_in_partition_key</td>
<td>integer</td>
<td></td>
<td>For list partitions you can have a composite (multi-column) partition key. This shows the position of the column in a composite key.</td>
</tr>
</tbody>
</table>
pg_partition_rule

The `pg_partition_rule` system catalog table is used to track partitioned tables, their check constraints, and data containment rules. Each row of `pg_partition_rule` represents either a leaf partition (the bottom level partitions that contain data), or a branch partition (a top or mid-level partition that is used to define the partition hierarchy, but does not contain any data).

**Table H.1  pg_catalog.pg_partition_rule**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>paroid</td>
<td>oid</td>
<td><code>pg_partition.oid</code></td>
<td>Row identifier of the partitioning level (from <code>pg_partition</code>) to which this partition belongs. In the case of a branch partition, the corresponding table (identified by <code>parchildrelid</code>) is an empty container table. In case of a leaf partition, the table contains the rows for that partition containment rule.</td>
</tr>
<tr>
<td>parchildrelid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The table identifier of the partition (child table).</td>
</tr>
<tr>
<td>parparentrule</td>
<td>oid</td>
<td><code>pg_partition_rule.paroid</code></td>
<td>The row identifier of the rule associated with the parent table of this partition.</td>
</tr>
<tr>
<td>parname</td>
<td>name</td>
<td></td>
<td>The given name of this partition.</td>
</tr>
<tr>
<td>parisdefault</td>
<td>boolean</td>
<td></td>
<td>Whether or not this partition is a default partition.</td>
</tr>
<tr>
<td>parruleord</td>
<td>smallint</td>
<td></td>
<td>For range partitioned tables, the rank of this partition on this level of the partition hierarchy.</td>
</tr>
<tr>
<td>parrangestartincl</td>
<td>boolean</td>
<td></td>
<td>For range partitioned tables, whether or not the starting value is inclusive.</td>
</tr>
<tr>
<td>parrangeendincl</td>
<td>boolean</td>
<td></td>
<td>For range partitioned tables, whether or not the ending value is inclusive.</td>
</tr>
<tr>
<td>parrangestart</td>
<td>text</td>
<td></td>
<td>For range partitioned tables, the starting value of the range.</td>
</tr>
<tr>
<td>parrangeend</td>
<td>text</td>
<td></td>
<td>For range partitioned tables, the ending value of the range.</td>
</tr>
<tr>
<td>partlistvalues</td>
<td>text</td>
<td></td>
<td>For list partitioned tables, the list of values assigned to this partition.</td>
</tr>
<tr>
<td>parreloptions</td>
<td>text</td>
<td></td>
<td>An array describing the storage characteristics of the particular partition.</td>
</tr>
</tbody>
</table>
### pg_partition_templates

The `pg_partition_templates` system view is used to show the subpartitions that were created using a subpartition template.

**Table H.1  pg_catalog.pg_partition_templates**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>schemaname</td>
<td>name</td>
<td></td>
<td>The name of the schema the partitioned table is in.</td>
</tr>
<tr>
<td>tablename</td>
<td>name</td>
<td></td>
<td>The table name of the top-level parent table.</td>
</tr>
<tr>
<td>partitionname</td>
<td>name</td>
<td></td>
<td>The name of the subpartition (this is the name to use if referring to the partition in an <code>ALTER TABLE</code> command). <code>NULL</code> if the partition was not given a name at create time or generated by an <code>EVERY</code> clause.</td>
</tr>
<tr>
<td>partitiontype</td>
<td>text</td>
<td></td>
<td>The type of subpartition (range or list).</td>
</tr>
<tr>
<td>partitionlevel</td>
<td>smallint</td>
<td></td>
<td>The level of this subpartition in the hierarchy.</td>
</tr>
<tr>
<td>partitionrank</td>
<td>bigint</td>
<td></td>
<td>For range partitions, the rank of the partition compared to other partitions of the same level.</td>
</tr>
<tr>
<td>partitionposition</td>
<td>smallint</td>
<td></td>
<td>The rule order position of this subpartition.</td>
</tr>
<tr>
<td>partitionlistvalues</td>
<td>text</td>
<td></td>
<td>For list partitions, the list value(s) associated with this subpartition.</td>
</tr>
<tr>
<td>partitionrangestart</td>
<td>text</td>
<td></td>
<td>For range partitions, the start value of this subpartition.</td>
</tr>
<tr>
<td>partitionstartinclusive</td>
<td>boolean</td>
<td></td>
<td>T if the start value is included in this subpartition. F if it is excluded.</td>
</tr>
<tr>
<td>partitionrangeend</td>
<td>text</td>
<td></td>
<td>For range partitions, the end value of this subpartition.</td>
</tr>
<tr>
<td>partitionendinclusive</td>
<td>boolean</td>
<td></td>
<td>T if the end value is included in this subpartition. F if it is excluded.</td>
</tr>
<tr>
<td>partitioneveryclause</td>
<td>text</td>
<td></td>
<td>The <code>EVERY</code> clause (interval) of this subpartition.</td>
</tr>
<tr>
<td>partitionisdefault</td>
<td>boolean</td>
<td></td>
<td>T if this is a default subpartition, otherwise F.</td>
</tr>
<tr>
<td>partitionboundary</td>
<td>text</td>
<td></td>
<td>The entire partition specification for this subpartition.</td>
</tr>
</tbody>
</table>
**pg_partitions**

The `pg_partitions` system view is used to show the structure of a partitioned table.

**Table H.1  pg_catalog.pg_partitions**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>schemaname</td>
<td>name</td>
<td></td>
<td>The name of the schema the partitioned table is in.</td>
</tr>
<tr>
<td>tablename</td>
<td>name</td>
<td></td>
<td>The name of the top-level parent table.</td>
</tr>
<tr>
<td>partitiontablename</td>
<td>name</td>
<td></td>
<td>The relation name of the partitioned table (this is the table name to use if accessing the partition directly).</td>
</tr>
<tr>
<td>partitionname</td>
<td>name</td>
<td></td>
<td>The name of the partition (this is the name to use if referring to the partition in an <code>ALTER TABLE</code> command). <code>NULL</code> if the partition was not given a name at create time or generated by an <code>EVERY</code> clause.</td>
</tr>
<tr>
<td>parentpartitiontablename</td>
<td>name</td>
<td></td>
<td>The relation name of the parent table one level up from this partition.</td>
</tr>
<tr>
<td>parentpartitionname</td>
<td>name</td>
<td></td>
<td>The given name of the parent table one level up from this partition.</td>
</tr>
<tr>
<td>partitiontype</td>
<td>text</td>
<td></td>
<td>The type of partition (range or list).</td>
</tr>
<tr>
<td>partitionlevel</td>
<td>smallint</td>
<td></td>
<td>The level of this partition in the hierarchy.</td>
</tr>
<tr>
<td>partitionrank</td>
<td>bigint</td>
<td></td>
<td>For range partitions, the rank of the partition compared to other partitions of the same level.</td>
</tr>
<tr>
<td>partitionposition</td>
<td>smallint</td>
<td></td>
<td>The rule order position of this partition.</td>
</tr>
<tr>
<td>partitionlistvalues</td>
<td>text</td>
<td></td>
<td>For list partitions, the list value(s) associated with this partition.</td>
</tr>
<tr>
<td>partitionrangestart</td>
<td>text</td>
<td></td>
<td>For range partitions, the start value of this partition.</td>
</tr>
<tr>
<td>partitionstartinclusive</td>
<td>boolean</td>
<td></td>
<td><code>T</code> if the start value is included in this partition. <code>F</code> if it is excluded.</td>
</tr>
<tr>
<td>partitionrangeend</td>
<td>text</td>
<td></td>
<td>For range partitions, the end value of this partition.</td>
</tr>
<tr>
<td>partitionendinclusive</td>
<td>boolean</td>
<td></td>
<td><code>T</code> if the end value is included in this partition. <code>F</code> if it is excluded.</td>
</tr>
<tr>
<td>partitioneveryclause</td>
<td>text</td>
<td></td>
<td>The <code>EVERY</code> clause (interval) of this partition.</td>
</tr>
</tbody>
</table>
### Table H.1  pg_catalog.pg_partitions

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>partitionisdefault</td>
<td>boolean</td>
<td></td>
<td>T if this is a default partition, otherwise F.</td>
</tr>
<tr>
<td>partitionboundary</td>
<td>text</td>
<td></td>
<td>The entire partition specification for this partition.</td>
</tr>
</tbody>
</table>
**pg_pltemplate**

The `pg_pltemplate` system catalog table stores template information for procedural languages. A template for a language allows the language to be created in a particular database by a simple `CREATE LANGUAGE` command, with no need to specify implementation details. Unlike most system catalogs, `pg_pltemplate` is shared across all databases of Greenplum system: there is only one copy of `pg_pltemplate` per system, not one per database. This allows the information to be accessible in each database as it is needed.

There are not currently any commands that manipulate procedural language templates; to change the built-in information, a superuser must modify the table using ordinary `INSERT`, `DELETE`, or `UPDATE` commands.

**Table H.1  pg_catalog.pg_pltemplate**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tmplname</td>
<td>name</td>
<td></td>
<td>Name of the language this template is for</td>
</tr>
<tr>
<td>tmpltrusted</td>
<td>boolean</td>
<td></td>
<td>True if language is considered trusted</td>
</tr>
<tr>
<td>tmplhandler</td>
<td>text</td>
<td></td>
<td>Name of call handler function</td>
</tr>
<tr>
<td>tmplvalidator</td>
<td>text</td>
<td></td>
<td>Name of validator function, or NULL if none</td>
</tr>
<tr>
<td>tmpllibrary</td>
<td>text</td>
<td></td>
<td>Path of shared library that implements language</td>
</tr>
<tr>
<td>tmplacl</td>
<td>aclitem[]</td>
<td></td>
<td>Access privileges for template (not yet implemented).</td>
</tr>
</tbody>
</table>
The *pg_proc* system catalog table stores information about functions (or procedures), both built-in functions and those defined by `CREATE FUNCTION`. The table contains data for aggregate and window functions as well as plain functions. If `proisagg` is true, there should be a matching row in *pg_aggregate*. If `proiswin` is true, there should be a matching row in *pg_window*.

For compiled functions, both built-in and dynamically loaded, `prosrc` contains the function’s C-language name (link symbol). For all other currently-known language types, `prosrc` contains the function’s source text. `probin` is unused except for dynamically-loaded C functions, for which it gives the name of the shared library file containing the function.

**Table H.1  pg_catalog.pg_proc**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>praname</td>
<td>name</td>
<td></td>
<td>Name of the function.</td>
</tr>
<tr>
<td>pronamespace</td>
<td>oid</td>
<td><code>pg_namespace.oid</code></td>
<td>The OID of the namespace that contains this function.</td>
</tr>
<tr>
<td>proowner</td>
<td>oid</td>
<td><code>pg_authid.oid</code></td>
<td>Owner of the function.</td>
</tr>
<tr>
<td>prolang</td>
<td>oid</td>
<td><code>pg_language.oid</code></td>
<td>Implementation language or call interface of this function.</td>
</tr>
<tr>
<td>proisagg</td>
<td>boolean</td>
<td></td>
<td>Function is an aggregate function.</td>
</tr>
<tr>
<td>prosecdef</td>
<td>boolean</td>
<td></td>
<td>Function is a security definer (for example, a ‘setuid’ function).</td>
</tr>
<tr>
<td>proisstrict</td>
<td>boolean</td>
<td></td>
<td>Function returns NULL if any call argument is NULL. In that case the function will not actually be called at all. Functions that are not strict must be prepared to handle NULL inputs.</td>
</tr>
<tr>
<td>proretset</td>
<td>boolean</td>
<td></td>
<td>Function returns a set (multiple values of the specified data type).</td>
</tr>
<tr>
<td>provolatile</td>
<td>char</td>
<td></td>
<td>Tells whether the function’s result depends only on its input arguments, or is affected by outside factors. (i) = immutable (always delivers the same result for the same inputs), (s) = stable (results (for fixed inputs) do not change within a scan), (v) = volatile (results may change at any time or functions with side-effects).</td>
</tr>
<tr>
<td>pronargs</td>
<td>int2</td>
<td></td>
<td>Number of arguments.</td>
</tr>
<tr>
<td>prorettype</td>
<td>oid</td>
<td><code>pg_type.oid</code></td>
<td>Data type of the return value.</td>
</tr>
<tr>
<td>proiswin</td>
<td>boolean</td>
<td></td>
<td>Function is neither an aggregate nor a scalar function, but a pure window function.</td>
</tr>
<tr>
<td>column</td>
<td>type</td>
<td>references</td>
<td>description</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------</td>
<td>------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>proargtypes</td>
<td>oidvector</td>
<td>pg_type.oid</td>
<td>An array with the data types of the function arguments. This includes only input arguments (including INOUT arguments), and thus represents the call signature of the function.</td>
</tr>
<tr>
<td>proallargtypes</td>
<td>oid[]</td>
<td>pg_type.oid</td>
<td>An array with the data types of the function arguments. This includes all arguments (including OUT and INOUT arguments); however, if all the arguments are IN arguments, this field will be null. Note that subscripting is 1-based, whereas for historical reasons proargtypes is subscripted from 0.</td>
</tr>
<tr>
<td>proargmodes</td>
<td>char[]</td>
<td></td>
<td>An array with the modes of the function arguments: i = IN, o = OUT, b = INOUT. If all the arguments are IN arguments, this field will be null. Note that subscripts correspond to positions of proallargtypes not proargtypes.</td>
</tr>
<tr>
<td>proargnames</td>
<td>text[]</td>
<td></td>
<td>An array with the names of the function arguments. Arguments without a name are set to empty strings in the array. If none of the arguments have a name, this field will be null. Note that subscripts correspond to positions of proallargtypes not proargtypes.</td>
</tr>
<tr>
<td>prosrc</td>
<td>text</td>
<td></td>
<td>This tells the function handler how to invoke the function. It might be the actual source code of the function for interpreted languages, a link symbol, a file name, or just about anything else, depending on the implementation language/call convention.</td>
</tr>
<tr>
<td>probin</td>
<td>bytea</td>
<td></td>
<td>Additional information about how to invoke the function. Again, the interpretation is language-specific.</td>
</tr>
<tr>
<td>proacl</td>
<td>aclitem[]</td>
<td></td>
<td>Access privileges for the function as given by GRANT/REVOKE.</td>
</tr>
</tbody>
</table>
The `pg_resqueue` system catalog table contains information about Greenplum Database resource queues, which are used for the workload management feature. This table is populated only on the master. This table is defined in the `pg_global` tablespace, meaning it is globally shared across all databases in the system.

**Table H.1  pg_catalog.pg_resqueue**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rsqname</td>
<td>name</td>
<td></td>
<td>The name of the resource queue.</td>
</tr>
<tr>
<td>rsqcountlimit</td>
<td>real</td>
<td></td>
<td>The active query threshold of the resource queue.</td>
</tr>
<tr>
<td>rsqcostlimit</td>
<td>real</td>
<td></td>
<td>The query cost threshold of the resource queue.</td>
</tr>
<tr>
<td>rsqovercommit</td>
<td>boolean</td>
<td></td>
<td>Allows queries that exceed the cost threshold to run when the system is idle.</td>
</tr>
<tr>
<td>rsqignorecostlimit</td>
<td>real</td>
<td></td>
<td>The query cost limit of what is considered a ‘small query’. Queries with a cost under this limit will not be queued and run immediately.</td>
</tr>
</tbody>
</table>
The `pg_resourcetype` system catalog table contains information about the extended attributes that can be assigned to Greenplum Database resource queues. Each row details an attribute and inherent qualities such as its default setting, whether it is required, and the value to disable it (when allowed).

This table is populated only on the master. This table is defined in the `pg_global` tablespace, meaning it is globally shared across all databases in the system.

**Table H.1  pg_catalog.pg_resourcetype**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>restypid</td>
<td>smallint</td>
<td></td>
<td>The resource type ID.</td>
</tr>
<tr>
<td>resname</td>
<td>name</td>
<td></td>
<td>The name of the resource type.</td>
</tr>
<tr>
<td>resrequired</td>
<td>boolean</td>
<td></td>
<td>Whether the resource type is required for a valid resource queue.</td>
</tr>
<tr>
<td>reshasdefault</td>
<td>boolean</td>
<td></td>
<td>Whether the resource type has a default value. When true, the default value is specified in <code>reshasdefaultsetting</code>.</td>
</tr>
<tr>
<td>rescandisable</td>
<td>boolean</td>
<td></td>
<td>Whether the type can be removed or disabled. When true, the default value is specified in <code>resdisabledsetting</code>.</td>
</tr>
<tr>
<td>resdefaultsetting</td>
<td>text</td>
<td></td>
<td>Default setting for the resource type, when applicable.</td>
</tr>
<tr>
<td>resdisabledsetting</td>
<td>text</td>
<td></td>
<td>The value that disables this resource type (when allowed).</td>
</tr>
</tbody>
</table>
The `pg_resqueue_attributes` view allows administrators to see the attributes set for a resource queue, such as its active statement limit, query cost limits, and priority.

**Table H.1**  pg_catalog.pg_resqueue_attributes

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rsqname</td>
<td>name</td>
<td>pg_resqueue.rsqname</td>
<td>The name of the resource queue.</td>
</tr>
<tr>
<td>resname</td>
<td>text</td>
<td></td>
<td>The name of the resource queue attribute.</td>
</tr>
<tr>
<td>ressetting</td>
<td>text</td>
<td></td>
<td>The current value of a resource queue attribute.</td>
</tr>
<tr>
<td>restypid</td>
<td>integer</td>
<td></td>
<td>System assigned resource type id.</td>
</tr>
</tbody>
</table>
pg_resqueue_status

The *pg_resqueue_status* view allows administrators to see status and activity for a workload management resource queue. It shows how many queries are waiting to run and how many queries are currently active in the system from a particular resource queue.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rsqname</td>
<td>name</td>
<td>pg_resqueue.rsqname</td>
<td>The name of the resource queue.</td>
</tr>
<tr>
<td>rsqcountlimit</td>
<td>real</td>
<td>pg_resqueue.countlimit</td>
<td>The active query threshold of the resource queue. A value of -1 means no limit.</td>
</tr>
<tr>
<td>rsqcountvalue</td>
<td>real</td>
<td></td>
<td>The number of active query slots currently being used in the resource queue.</td>
</tr>
<tr>
<td>rsqcostlimit</td>
<td>real</td>
<td>pg_resqueue.costlimit</td>
<td>The query cost threshold of the resource queue. A value of -1 means no limit.</td>
</tr>
<tr>
<td>rsqcostvalue</td>
<td>real</td>
<td></td>
<td>The total cost of all statements currently in the resource queue.</td>
</tr>
<tr>
<td>rsqwaiters</td>
<td>integer</td>
<td></td>
<td>The number of statements currently waiting in the resource queue.</td>
</tr>
<tr>
<td>rsqholders</td>
<td>integer</td>
<td></td>
<td>The number of statements currently running on the system from this resource queue.</td>
</tr>
</tbody>
</table>
**pg_resqueuecapability**

The `pg_resqueuecapability` system catalog table contains information about the extended attributes, or capabilities, of existing Greenplum Database resource queues. Only resource queues that have been assigned an extended capability, such as a priority setting, are recorded in this table. This table is joined to the `pg_resqueue` table by resource queue object ID, and to the `pg_resourcetype` table by resource type ID (`restypid`).

This table is populated only on the master. This table is defined in the `pg_global` tablespace, meaning it is globally shared across all databases in the system.

<table>
<thead>
<tr>
<th><strong>Table H.1</strong> pg_catalog.pg_resqueuecapability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>column</strong></td>
</tr>
<tr>
<td>rsqueueid</td>
</tr>
<tr>
<td>restypid</td>
</tr>
<tr>
<td>ressetting</td>
</tr>
</tbody>
</table>
**pg_rewrite**

The *pg_rewrite* system catalog table stores rewrite rules for tables and views. *pg_class.relhasrules* must be true if a table has any rules in this catalog.

**Table H.1  pg_catalog.pg_rewrite**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rulename</td>
<td>name</td>
<td></td>
<td>Rule name.</td>
</tr>
<tr>
<td>ev_class</td>
<td>oid</td>
<td><em>pg_class.oid</em></td>
<td>The table this rule is for.</td>
</tr>
<tr>
<td>ev_attr</td>
<td>int2</td>
<td></td>
<td>The column this rule is for (currently, always zero to indicate the whole table).</td>
</tr>
<tr>
<td>ev_type</td>
<td>char</td>
<td></td>
<td>Event type that the rule is for: 1 = SELECT, 2 = UPDATE, 3 = INSERT, 4 = DELETE.</td>
</tr>
<tr>
<td>is_instead</td>
<td>boolean</td>
<td></td>
<td>True if the rule is an INSTEAD rule.</td>
</tr>
<tr>
<td>ev_qual</td>
<td>text</td>
<td></td>
<td>Expression tree (in the form of a nodeToString() representation) for the rule's qualifying condition.</td>
</tr>
<tr>
<td>ev_action</td>
<td>text</td>
<td></td>
<td>Query tree (in the form of a nodeToString() representation) for the rule's action.</td>
</tr>
</tbody>
</table>
The view *pg_roles* provides access to information about database roles. This is simply a publicly readable view of *pg_authid* that blanks out the password field. This view explicitly exposes the OID column of the underlying table, since that is needed to do joins to other catalogs.

### Table H.1  *pg_catalog.pg_roles*

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rolname</td>
<td>name</td>
<td></td>
<td>Role name</td>
</tr>
<tr>
<td>rolsuper</td>
<td>bool</td>
<td></td>
<td>Role has superuser privileges</td>
</tr>
<tr>
<td>rolinherit</td>
<td>bool</td>
<td></td>
<td>Role automatically inherits privileges of roles it is a member of</td>
</tr>
<tr>
<td>rolcreaterole</td>
<td>bool</td>
<td></td>
<td>Role may create more roles</td>
</tr>
<tr>
<td>rolcreatedb</td>
<td>bool</td>
<td></td>
<td>Role may create databases</td>
</tr>
<tr>
<td>rolcatupdate</td>
<td>bool</td>
<td></td>
<td>Role may update system catalogs directly. (Even a superuser may not do this unless this column is true.)</td>
</tr>
<tr>
<td>rolcanlogin</td>
<td>bool</td>
<td></td>
<td>Role may log in. That is, this role can be given as the initial session authorization identifier</td>
</tr>
<tr>
<td>rolconnlimit</td>
<td>int4</td>
<td></td>
<td>For roles that can log in, this sets maximum number of concurrent connections this role can make. -1 means no limit</td>
</tr>
<tr>
<td>rolpassword</td>
<td>text</td>
<td></td>
<td>Not the password (always reads as ********)</td>
</tr>
<tr>
<td>rolvaliduntil</td>
<td>timestamptz</td>
<td></td>
<td>Password expiry time (only used for password authentication); NULL if no expiration</td>
</tr>
<tr>
<td>rolconfig</td>
<td>text[]</td>
<td></td>
<td>Session defaults for run-time configuration variables</td>
</tr>
<tr>
<td>rolresqueue</td>
<td>oid</td>
<td>pg_resqueue.oid</td>
<td>Object ID of the resource queue this role is assigned to.</td>
</tr>
<tr>
<td>oid</td>
<td>oid</td>
<td>pg_authid.oid</td>
<td>Object ID of role</td>
</tr>
<tr>
<td>rolcreaterextgpfd</td>
<td>bool</td>
<td></td>
<td>Role may create readable external tables that use the gpfdist protocol.</td>
</tr>
<tr>
<td>rolcreaterexthttp</td>
<td>bool</td>
<td></td>
<td>Role may create readable external tables that use the gpfdist protocol.</td>
</tr>
<tr>
<td>rolcreatewextgpfd</td>
<td>bool</td>
<td></td>
<td>Role may create writable external tables that use the gpfdist protocol.</td>
</tr>
</tbody>
</table>
The `pg_shdepend` system catalog table records the dependency relationships between database objects and shared objects, such as roles. This information allows Greenplum Database to ensure that those objects are unreferenced before attempting to delete them. See also `pg_depend`, which performs a similar function for dependencies involving objects within a single database. Unlike most system catalogs, `pg_shdepend` is shared across all databases of Greenplum system: there is only one copy of `pg_shdepend` per system, not one per database.

In all cases, a `pg_shdepend` entry indicates that the referenced object may not be dropped without also dropping the dependent object. However, there are several subflavors identified by `deptype`:

- **SHAREDEDPENDENCY_OWNER (o)** — The referenced object (which must be a role) is the owner of the dependent object.
- **SHAREDEDPENDENCY_ACL (a)** — The referenced object (which must be a role) is mentioned in the ACL (access control list) of the dependent object.
- **SHAREDEDPENDENCY_PIN (p)** — There is no dependent object; this type of entry is a signal that the system itself depends on the referenced object, and so that object must never be deleted. Entries of this type are created only by system initialization. The columns for the dependent object contain zeroes.

### Table H.1  pg_catalog.pg_shdepend

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbid</td>
<td>oid</td>
<td><code>pg_database.oid</code></td>
<td>The OID of the database the dependent object is in, or zero for a shared object.</td>
</tr>
<tr>
<td>classid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The OID of the system catalog the dependent object is in.</td>
</tr>
<tr>
<td>objid</td>
<td>oid</td>
<td>any OID column</td>
<td>The OID of the specific dependent object.</td>
</tr>
<tr>
<td>objsubid</td>
<td>int4</td>
<td></td>
<td>For a table column, this is the column number. For all other object types, this column is zero.</td>
</tr>
<tr>
<td>refclassid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The OID of the system catalog the referenced object is in (must be a shared catalog).</td>
</tr>
<tr>
<td>refobjid</td>
<td>oid</td>
<td>any OID column</td>
<td>The OID of the specific referenced object.</td>
</tr>
<tr>
<td>refobjsubid</td>
<td>int4</td>
<td></td>
<td>For a table column, this is the referenced column number. For all other object types, this column is zero.</td>
</tr>
<tr>
<td>deptype</td>
<td>char</td>
<td></td>
<td>A code defining the specific semantics of this dependency relationship.</td>
</tr>
</tbody>
</table>
The `pg_shdescription` system catalog table stores optional descriptions (comments) for shared database objects. Descriptions can be manipulated with the `COMMENT` command and viewed with `psql`’s `\d` meta-commands. See also `pg_description`, which performs a similar function for descriptions involving objects within a single database. Unlike most system catalogs, `pg_shdescription` is shared across all databases of a Greenplum system: there is only one copy of `pg_shdescription` per system, not one per database.

**Table H.1  pg_catalog.pg_shdescription**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>objoid</td>
<td>oid</td>
<td>any OID column</td>
<td>The OID of the object this description pertains to.</td>
</tr>
<tr>
<td>classoid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The OID of the system catalog this object appears in.</td>
</tr>
<tr>
<td>description</td>
<td>text</td>
<td></td>
<td>Arbitrary text that serves as the description of this object.</td>
</tr>
</tbody>
</table>
**pg_stat_activity**

The view *pg_stat_activity* shows one row per server process and details about it associated user session and query. The columns that report data on the current query are available unless the parameter *stats_command_string* has been turned off. Furthermore, these columns are only visible if the user examining the view is a superuser or the same as the user owning the process being reported on.

**Table H.1  pg_catalog.pg_stat_activity**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>datid</td>
<td>oid</td>
<td><em>pg_database.oid</em></td>
<td>Database OID</td>
</tr>
<tr>
<td>datname</td>
<td>name</td>
<td></td>
<td>Database name</td>
</tr>
<tr>
<td>procpid</td>
<td>integer</td>
<td></td>
<td>Process ID of the server process</td>
</tr>
<tr>
<td>sess_id</td>
<td>integer</td>
<td></td>
<td>Session ID</td>
</tr>
<tr>
<td>usesysid</td>
<td>oid</td>
<td><em>pg_authid.oid</em></td>
<td>Role OID</td>
</tr>
<tr>
<td>username</td>
<td>name</td>
<td></td>
<td>Role name</td>
</tr>
<tr>
<td>current_query</td>
<td>text</td>
<td></td>
<td>Current query that process is running</td>
</tr>
<tr>
<td>waiting</td>
<td>boolean</td>
<td></td>
<td>True if waiting on a lock, false if not waiting</td>
</tr>
<tr>
<td>query_start</td>
<td>timestampz</td>
<td></td>
<td>Time query began execution</td>
</tr>
<tr>
<td>backend_start</td>
<td>timestampz</td>
<td></td>
<td>Time backend process was started</td>
</tr>
<tr>
<td>client_addr</td>
<td>inet</td>
<td></td>
<td>Client address</td>
</tr>
<tr>
<td>client_port</td>
<td>integer</td>
<td></td>
<td>Client port</td>
</tr>
</tbody>
</table>
**pg_stat_operations**

The view *pg_stat_operations* shows details about the last operation performed on a database object (such as a table, index, view or database) or a global object (such as a role).

**Table H.1 pg_catalog.pg_stat_operations**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>classname</td>
<td>text</td>
<td></td>
<td>The name of the system table in the pg_catalog schema where the record about this object is stored (pg_class=relations, pg_database=databases, pg_namespace=schemas, pg_authid=roles)</td>
</tr>
<tr>
<td>objname</td>
<td>name</td>
<td></td>
<td>The name of the object.</td>
</tr>
<tr>
<td>objid</td>
<td>oid</td>
<td></td>
<td>The OID of the object.</td>
</tr>
<tr>
<td>schemaname</td>
<td>name</td>
<td></td>
<td>The name of the schema where the object resides.</td>
</tr>
<tr>
<td>usestatus</td>
<td>text</td>
<td></td>
<td>The status of the role who performed the last operation on the object (CURRENT=a currently active role in the system, DROPPED=a role that no longer exists in the system, CHANGED=a role name that exists in the system, but has changed since the last operation was performed).</td>
</tr>
<tr>
<td>username</td>
<td>name</td>
<td></td>
<td>The name of the role that performed the operation on this object.</td>
</tr>
<tr>
<td>actionname</td>
<td>name</td>
<td></td>
<td>The action that was taken on the object.</td>
</tr>
<tr>
<td>subtype</td>
<td>text</td>
<td></td>
<td>The type of object operated on or the subclass of operation performed.</td>
</tr>
<tr>
<td>statime</td>
<td>timestampz</td>
<td></td>
<td>The timestamp of the operation. This is the same timestamp that is written to the Greenplum Database server log files in case you need to look up more detailed information about the operation in the logs.</td>
</tr>
</tbody>
</table>
The view `pg_stat_partition_operations` shows details about the last operation performed on a partitioned table.

### Table H.1  `pg_catalog.pg_stat_partition_operations`

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>classname</code></td>
<td>text</td>
<td></td>
<td>The name of the system table in the <code>pg_catalog</code> schema where the record about this object is stored (always <code>pg_class</code> for tables and partitions).</td>
</tr>
<tr>
<td><code>objname</code></td>
<td>name</td>
<td></td>
<td>The name of the object.</td>
</tr>
<tr>
<td><code>objid</code></td>
<td>oid</td>
<td></td>
<td>The OID of the object.</td>
</tr>
<tr>
<td><code>schemaname</code></td>
<td>name</td>
<td></td>
<td>The name of the schema where the object resides.</td>
</tr>
<tr>
<td><code>usestatus</code></td>
<td>text</td>
<td></td>
<td>The status of the role who performed the last operation on the object (CURRENT=a currently active role in the system, DROPPED=a role that no longer exists in the system, CHANGED=a role name that exists in the system, but its definition has changed since the last operation was performed).</td>
</tr>
<tr>
<td><code>usename</code></td>
<td>name</td>
<td></td>
<td>The name of the role that performed the operation on this object.</td>
</tr>
<tr>
<td><code>actionname</code></td>
<td>name</td>
<td></td>
<td>The action that was taken on the object.</td>
</tr>
<tr>
<td><code>subtype</code></td>
<td>text</td>
<td></td>
<td>The type of object operated on or the subclass of operation performed.</td>
</tr>
<tr>
<td><code>statime</code></td>
<td>timestampz</td>
<td></td>
<td>The timestamp of the operation. This is the same timestamp that is written to the Greenplum Database server log files in case you need to look up more detailed information about the operation in the logs.</td>
</tr>
<tr>
<td><code>partitionlevel</code></td>
<td>smallint</td>
<td></td>
<td>The level of this partition in the hierarchy.</td>
</tr>
<tr>
<td><code>parenttablename</code></td>
<td>name</td>
<td></td>
<td>The relation name of the parent table one level up from this partition.</td>
</tr>
<tr>
<td><code>parentschemaname</code></td>
<td>name</td>
<td></td>
<td>The name of the schema where the parent table resides.</td>
</tr>
<tr>
<td><code>parent_relid</code></td>
<td>oid</td>
<td></td>
<td>The OID of the parent table one level up from this partition.</td>
</tr>
</tbody>
</table>
The `pg_stat_resqueues` view allows administrators to view metrics about a resource queue’s workload over time. To allow statistics to be collected for this view, you must enable the `stats_queue_level` server configuration parameter on the Greenplum Database master instance. Enabling the collection of these metrics does incur a small performance penalty, as each statement submitted through a resource queue must be logged in the system catalog tables.

### Table H.1  `pg_catalog.pg_stat_resqueues`

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>queueoid</td>
<td>oid</td>
<td></td>
<td>The OID of the resource queue.</td>
</tr>
<tr>
<td>queuename</td>
<td>name</td>
<td></td>
<td>The name of the resource queue.</td>
</tr>
<tr>
<td>n_queries_exec</td>
<td>bigint</td>
<td></td>
<td>Number of queries submitted for execution from this resource queue.</td>
</tr>
<tr>
<td>n_queries_wait</td>
<td>bigint</td>
<td></td>
<td>Number of queries submitted to this resource queue that had to wait before they could execute.</td>
</tr>
<tr>
<td>elapsed_exec</td>
<td>bigint</td>
<td></td>
<td>Total elapsed execution time for statements submitted through this resource queue.</td>
</tr>
<tr>
<td>elapsed_wait</td>
<td>bigint</td>
<td></td>
<td>Total elapsed time that statements submitted through this resource queue had to wait before they were executed.</td>
</tr>
</tbody>
</table>
The `pg_stat_last_operation` table contains metadata tracking information about database objects (tables, views, etc.).

**Table H.1  pg_catalog.pg_stat_last_operation**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>classid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>OID of the system catalog containing the object.</td>
</tr>
<tr>
<td>objid</td>
<td>oid</td>
<td>any OID column</td>
<td>OID of the object within its system catalog.</td>
</tr>
<tr>
<td>staactionname</td>
<td>name</td>
<td></td>
<td>The action that was taken on the object.</td>
</tr>
<tr>
<td>stasysid</td>
<td>oid</td>
<td><code>pg_authid.oid</code></td>
<td>A foreign key to <code>pg_authid.oid</code>.</td>
</tr>
<tr>
<td>stausename</td>
<td>name</td>
<td></td>
<td>The name of the role that performed the operation on this object.</td>
</tr>
<tr>
<td>stasubtype</td>
<td>text</td>
<td></td>
<td>The type of object operated on or the subclass of operation performed.</td>
</tr>
<tr>
<td>statime</td>
<td>timestamp with timezone</td>
<td></td>
<td>The timestamp of the operation. This is the same timestamp that is written to the Greenplum Database server log files in case you need to look up more detailed information about the operation in the logs.</td>
</tr>
</tbody>
</table>
The `pg_stat_last_shoperation` table contains metadata tracking information about global objects (roles, tablespaces, etc.).

**Table H.1  pg_catalog.pg_stat_last_shoperation**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>classid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>OID of the system catalog containing the object.</td>
</tr>
<tr>
<td>objid</td>
<td>oid</td>
<td>any OID column</td>
<td>OID of the object within its system catalog.</td>
</tr>
<tr>
<td>staactionname</td>
<td>name</td>
<td></td>
<td>The action that was taken on the object.</td>
</tr>
<tr>
<td>stasysid</td>
<td>oid</td>
<td></td>
<td>The name of the role that performed the operation on this object.</td>
</tr>
<tr>
<td>stausename</td>
<td>name</td>
<td></td>
<td>The type of object operated on or the subclass of operation performed.</td>
</tr>
<tr>
<td>stasubtype</td>
<td>text</td>
<td></td>
<td>The timestamp of the operation. This is the same timestamp that is written to the Greenplum Database server log files in case you need to look up more detailed information about the operation in the logs.</td>
</tr>
<tr>
<td>statime</td>
<td>timestamp with timezone</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
pg_statistic

The `pg_statistic` system catalog table stores statistical data about the contents of the database. Entries are created by `ANALYZE` and subsequently used by the query planner. There is one entry for each table column that has been analyzed. Note that all the statistical data is inherently approximate, even assuming that it is up-to-date.

`pg_statistic` also stores statistical data about the values of index expressions. These are described as if they were actual data columns; in particular, `starelid` references the index. No entry is made for an ordinary non-expression index column, however, since it would be redundant with the entry for the underlying table column.

Since different kinds of statistics may be appropriate for different kinds of data, `pg_statistic` is designed not to assume very much about what sort of statistics it stores. Only extremely general statistics (such as nullness) are given dedicated columns in `pg_statistic`. Everything else is stored in slots, which are groups of associated columns whose content is identified by a code number in one of the slot’s columns.

`pg_statistic` should not be readable by the public, since even statistical information about a table’s contents may be considered sensitive (for example: minimum and maximum values of a salary column). `pg_stats` is a publicly readable view on `pg_statistic` that only exposes information about those tables that are readable by the current user.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>starelid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The table or index that the described column belongs to.</td>
</tr>
<tr>
<td>staattnum</td>
<td>int2</td>
<td><code>pg_attribute.attnum</code></td>
<td>The number of the described column.</td>
</tr>
<tr>
<td>stanullfrac</td>
<td>float4</td>
<td></td>
<td>The fraction of the column’s entries that are null.</td>
</tr>
<tr>
<td>stawidth</td>
<td>int4</td>
<td></td>
<td>The average stored width, in bytes, of nonnull entries.</td>
</tr>
<tr>
<td>stadistinct</td>
<td>float4</td>
<td></td>
<td>The number of distinct nonnull data values in the column. A value greater than zero is the actual number of distinct values. A value less than zero is the negative of a fraction of the number of rows in the table (for example, a column in which values appear about twice on the average could be represented by stadistinct = -0.5). A zero value means the number of distinct values is unknown.</td>
</tr>
<tr>
<td>stakindN</td>
<td>int2</td>
<td></td>
<td>A code number indicating the kind of statistics stored in the Nth slot of the <code>pg_statistic</code> row.</td>
</tr>
</tbody>
</table>
An operator used to derive the statistics stored in the Nth slot. For example, a histogram slot would show the \(<\) operator that defines the sort order of the data.

Numerical statistics of the appropriate kind for the Nth slot, or NULL if the slot kind does not involve numerical values.

Column data values of the appropriate kind for the Nth slot, or NULL if the slot kind does not store any data values. Each array's element values are actually of the specific column's data type, so there is no way to define these columns' type more specifically than anyarray.
The `pg_tablespace` system catalog table stores information about the available tablespaces. Tables can be placed in particular tablespaces to aid administration of disk layout. Unlike most system catalogs, `pg_tablespace` is shared across all databases of a Greenplum system: there is only one copy of `pg_tablespace` per system, not one per database.

**Table H.1  pg_catalog.pg_tablespace**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>spcname</td>
<td>name</td>
<td></td>
<td>Tablespace name.</td>
</tr>
<tr>
<td>spcowner</td>
<td>oid</td>
<td><code>pg_authid.oid</code></td>
<td>Owner of the tablespace, usually the user who created it.</td>
</tr>
<tr>
<td>spclocation</td>
<td>text[]</td>
<td></td>
<td>Deprecated.</td>
</tr>
<tr>
<td>spcacl</td>
<td>aclitem[]</td>
<td></td>
<td>Tablespace access privileges.</td>
</tr>
<tr>
<td>spcprilocations</td>
<td>text[]</td>
<td></td>
<td>Deprecated.</td>
</tr>
<tr>
<td>spcmrilocations</td>
<td>text[]</td>
<td></td>
<td>Deprecated.</td>
</tr>
<tr>
<td>spcsfsoid</td>
<td>oid</td>
<td><code>pg_filespace.oid</code></td>
<td>The object id of the filespace used by this tablespace. A filespace defines directory locations on the primary, mirror and master segments.</td>
</tr>
</tbody>
</table>
**pg_trigger**

The `pg_trigger` system catalog table stores triggers on tables.

**Table H.1  pg_catalog.pg_trigger**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tgrelid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The table this trigger is on.</td>
</tr>
<tr>
<td>tgnname</td>
<td>name</td>
<td></td>
<td>Trigger name (must be unique among triggers of same table).</td>
</tr>
<tr>
<td>tgfoid</td>
<td>oid</td>
<td><code>pg_proc.oid</code></td>
<td>The function to be called.</td>
</tr>
<tr>
<td>tgtype</td>
<td>int2</td>
<td></td>
<td>Bit mask identifying trigger conditions.</td>
</tr>
<tr>
<td>tgenabled</td>
<td>boolean</td>
<td></td>
<td>True if trigger is enabled.</td>
</tr>
<tr>
<td>tgisconstraint</td>
<td>boolean</td>
<td></td>
<td>True if trigger implements a referential integrity constraint.</td>
</tr>
<tr>
<td>tgconstname</td>
<td>name</td>
<td></td>
<td>Referential integrity constraint name.</td>
</tr>
<tr>
<td>tgconstrelid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>The table referenced by a referential integrity constraint.</td>
</tr>
<tr>
<td>tgdeferrable</td>
<td>boolean</td>
<td></td>
<td>True if deferrable.</td>
</tr>
<tr>
<td>tginitdeferred</td>
<td>boolean</td>
<td></td>
<td>True if initially deferred.</td>
</tr>
<tr>
<td>tgnargs</td>
<td>int2</td>
<td></td>
<td>Number of argument strings passed to trigger function.</td>
</tr>
<tr>
<td>tgattr</td>
<td>int2vector</td>
<td></td>
<td>Currently unused.</td>
</tr>
<tr>
<td>tgargs</td>
<td>bytea</td>
<td></td>
<td>Argument strings to pass to trigger, each NULL-terminated.</td>
</tr>
</tbody>
</table>
**pg_type**

The `pg_type` system catalog table stores information about data types. Base types (scalar types) are created with `CREATE TYPE`, and domains with `CREATE DOMAIN`. A composite type is automatically created for each table in the database, to represent the row structure of the table. It is also possible to create composite types with `CREATE TYPE AS`.

**Table H.1  pg_catalog.pg_type**

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>typname</td>
<td>name</td>
<td></td>
<td>Data type name.</td>
</tr>
<tr>
<td>typnamespace</td>
<td>oid</td>
<td><code>pg_namespace.oid</code></td>
<td>The OID of the namespace that contains this type.</td>
</tr>
<tr>
<td>typowner</td>
<td>oid</td>
<td><code>pg_authid.oid</code></td>
<td>Owner of the type.</td>
</tr>
<tr>
<td>typlen</td>
<td>int2</td>
<td></td>
<td>For a fixed-size type, <code>typlen</code> is the number of bytes in the internal representation of the type. But for a variable-length type, <code>typlen</code> is negative. -1 indicates a 'varlena' type (one that has a length word), -2 indicates a null-terminated C string.</td>
</tr>
<tr>
<td>typbyval</td>
<td>boolean</td>
<td></td>
<td>Determines whether internal routines pass a value of this type by value or by reference. <code>typbyval</code> had better be false if <code>typlen</code> is not 1, 2, or 4 (or 8 on machines where Datum is 8 bytes). Variable-length types are always passed by reference. Note that <code>typbyval</code> can be false even if the length would allow pass-by-value; this is currently true for type <code>float4</code>, for example.</td>
</tr>
<tr>
<td>typtype</td>
<td>char</td>
<td></td>
<td>b for a base type, c for a composite type, d for a domain, or p for a pseudo-type.</td>
</tr>
<tr>
<td>typisdefined</td>
<td>boolean</td>
<td></td>
<td>True if the type is defined; false if this is a placeholder entry for a not-yet-defined type. When false, nothing except the type name, namespace, and OID can be relied on.</td>
</tr>
<tr>
<td>typdelim</td>
<td>char</td>
<td></td>
<td>Character that separates two values of this type when parsing array input. Note that the delimiter is associated with the array element data type, not the array data type.</td>
</tr>
<tr>
<td>typrelid</td>
<td>oid</td>
<td><code>pg_class.oid</code></td>
<td>If this is a composite type, then this column points to the <code>pg_class</code> entry that defines the corresponding table. (For a free-standing composite type, the <code>pg_class</code> entry does not really represent a table, but it is needed anyway for the type’s <code>pg_attribute</code> entries to link to.) Zero for non-composite types.</td>
</tr>
</tbody>
</table>
The current type can then be subscripted like an array yielding values of type `typelem`. A true array type is variable length (`typlen = -1`), but some fixed-length (`typlen > 0`) types also have nonzero `typelem`, for example `name` and `point`. If a fixed-length type has a `typelem` then its internal representation must be some number of values of the `typelem` data type with no other data. Variable-length array types have a header defined by the array subroutines.

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>typelem</td>
<td>oid</td>
<td><code>pg_type.oid</code></td>
<td>If not 0 then it identifies another row in <code>pg_type</code>. The current type can then be subscripted like an array yielding values of type <code>typelem</code>. A true array type is variable length (<code>typlen = -1</code>), but some fixed-length (<code>typlen &gt; 0</code>) types also have nonzero <code>typelem</code>, for example <code>name</code> and <code>point</code>. If a fixed-length type has a <code>typelem</code> then its internal representation must be some number of values of the <code>typelem</code> data type with no other data. Variable-length array types have a header defined by the array subroutines.</td>
</tr>
<tr>
<td>typinput</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>Input conversion function (text format).</td>
</tr>
<tr>
<td>typoutput</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>Output conversion function (text format).</td>
</tr>
<tr>
<td>typreceive</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>Input conversion function (binary format), or 0 if none.</td>
</tr>
<tr>
<td>tsend</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>Output conversion function (binary format), or 0 if none.</td>
</tr>
<tr>
<td>tpanalyze</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>Custom ANALYZE function, or 0 to use the standard function.</td>
</tr>
<tr>
<td>typalign</td>
<td>char</td>
<td></td>
<td>The alignment required when storing a value of this type. It applies to storage on disk as well as most representations of the value inside Greenplum Database. When multiple values are stored consecutively, such as in the representation of a complete row on disk, padding is inserted before a datum of this type so that it begins on the specified boundary. The alignment reference is the beginning of the first datum in the sequence. Possible values are: <code>c</code> = char alignment (no alignment needed), <code>s</code> = short alignment (2 bytes on most machines), <code>i</code> = int alignment (4 bytes on most machines), <code>d</code> = double alignment (8 bytes on many machines, but not all).</td>
</tr>
<tr>
<td>typstorage</td>
<td>char</td>
<td></td>
<td>For varlena types (those with <code>typlen = -1</code>) tells if the type is prepared for toasting and what the default strategy for attributes of this type should be. Possible values are: <code>p</code> = Value must always be stored plain, <code>e</code> = Value can be stored in a secondary relation (if relation has one, see <code>pg_class.reltoastrelid</code>), <code>m</code> = Value can be stored compressed inline, <code>x</code> = Value can be stored compressed inline or stored in secondary storage. Note that <code>m</code> columns can also be moved out to secondary storage, but only as a last resort (<code>e</code> and <code>x</code> columns are moved first).</td>
</tr>
<tr>
<td>typnotnull</td>
<td>boolean</td>
<td></td>
<td>Represents a not-null constraint on a type. Used for domains only.</td>
</tr>
</tbody>
</table>
### Table H.1  pg_catalog.pg_type

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>typbasetype</td>
<td>oid</td>
<td>pg_type.oid</td>
<td>Identifies the type that a domain is based on. Zero if this type is not a domain.</td>
</tr>
<tr>
<td>typtypmod</td>
<td>int4</td>
<td></td>
<td>Domains use typtypmod to record the typmod to be applied to their base type (-1 if base type does not use a typmod). -1 if this type is not a domain.</td>
</tr>
<tr>
<td>typndims</td>
<td>int4</td>
<td></td>
<td>The number of array dimensions for a domain that is an array (if typbasetype is an array type; the domain’s typelem will match the base type’s typelem). Zero for types other than array domains.</td>
</tr>
<tr>
<td>typdefaultbin</td>
<td>text</td>
<td></td>
<td>If not null, it is the nodeToString() representation of a default expression for the type. This is only used for domains.</td>
</tr>
<tr>
<td>typdefault</td>
<td>text</td>
<td></td>
<td>Null if the type has no associated default value. If not null, typdefault must contain a human-readable version of the default expression represented by typdefaultbin. If typdefaultbin is null and typdefault is not, then typdefault is the external representation of the type’s default value, which may be fed to the type’s input converter to produce a constant.</td>
</tr>
</tbody>
</table>
The `pg_window` table stores information about window functions. Window functions are often used to compose complex OLAP (online analytical processing) queries. Window functions are applied to partitioned result sets within the scope of a single query expression. A window partition is a subset of rows returned by a query, as defined in a special `OVER()` clause. Typical window functions are `rank`, `dense_rank`, and `row_number`. Each entry in `pg_window` is an extension of an entry in `pg_proc`. The `pg_proc` entry carries the window function’s name, input and output data types, and other information that is similar to ordinary functions.

### Table H.1  pg_catalog.pg_window

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>winfnoid</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>The OID in <code>pg_proc</code> of the window function.</td>
</tr>
<tr>
<td>winrequireorder</td>
<td>boolean</td>
<td></td>
<td>The window function requires its window specification to have an <code>ORDER BY</code> clause.</td>
</tr>
<tr>
<td>winallowframe</td>
<td>boolean</td>
<td></td>
<td>The window function permits its window specification to have a <code>ROWS</code> or <code>RANGE</code> framing clause.</td>
</tr>
<tr>
<td>winpeercount</td>
<td>boolean</td>
<td></td>
<td>The peer group row count is required to compute this window function, so the Window node implementation must ‘look ahead’ as necessary to make this available in its internal state.</td>
</tr>
<tr>
<td>wincount</td>
<td>boolean</td>
<td></td>
<td>The partition row count is required to compute this window function.</td>
</tr>
<tr>
<td>winfunc</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>The OID in <code>pg_proc</code> of a function to compute the value of an immediate-type window function.</td>
</tr>
<tr>
<td>winprefunc</td>
<td>regproc</td>
<td><code>pg_proc.oid</code></td>
<td>The OID in <code>pg_proc</code> of a preliminary window function to compute the partial value of a deferred-type window function.</td>
</tr>
<tr>
<td>winpretype</td>
<td>oid</td>
<td><code>pg_type.oid</code></td>
<td>The OID in <code>pg_type</code> of the preliminary window function’s result type.</td>
</tr>
</tbody>
</table>
## Table H.1  pg_catalog.pg_window

<table>
<thead>
<tr>
<th>column</th>
<th>type</th>
<th>references</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>winfinfunc</td>
<td>regproc</td>
<td>pg_proc.oid</td>
<td>The OID in pg_proc of a function to compute the final value of a deferred-type window function from the partition row count and the result of winprefunc.</td>
</tr>
</tbody>
</table>
| winkind | char   |            | A character indicating membership of the window function in a class of related functions:  
|         |        |            |   w - ordinary window functions  
|         |        |            |   n - NTILE functions  
|         |        |            |   f - FIRST_VALUE functions  
|         |        |            |   l - LAST_VALUE functions  
|         |        |            |   g - LAG functions  
|         |        |            |   d - LEAD functions  |
I. The gp_toolkit Administrative Schema

Greenplum provides an administrative schema called *gp_toolkit* that you can use to query the system catalogs, log files, and operating environment for system status information. The *gp_toolkit* schema contains a number of views that you can access using SQL commands. The *gp_toolkit* schema is accessible to all database users, although some objects may require superuser permissions. For convenience, you may want to add the *gp_toolkit* schema to your schema search path. For example:

```sql
=> ALTER ROLE myrole SET search_path TO myschema,gp_toolkit;
```

This documentation describes the most useful views in *gp_toolkit*. You may notice other objects (views, functions, and external tables) within the *gp_toolkit* schema that are not described in this documentation (these are supporting objects to the views described in this section).

Checking for Tables that Need Routine Maintenance

The following views can help identify tables that need routine table maintenance (VACUUM and/or ANALYZE).

- *gp_bloat_diag*
- *gp_stats_missing*

The VACUUM or VACUUM FULL command reclaims disk space occupied by deleted or obsolete rows. Because of the MVCC transaction concurrency model used in Greenplum Database, data rows that are deleted or updated still occupy physical space on disk even though they are not visible to any new transactions. Expired rows increase table size on disk and eventually slow down scans of the table.

The ANALYZE command collects column-level statistics needed by the query planner. Greenplum Database uses a cost-based query planner that relies on database statistics. Accurate statistics allow the query planner to better estimate selectivity and the number of rows retrieved by a query operation in order to choose the most efficient query plan.
gp_bloat_diag
This view shows tables that have bloat (the actual number of pages on disk exceeds the expected number of pages given the table statistics). Tables that are bloated require a VACUUM or a VACUUM FULL in order to reclaim disk space occupied by deleted or obsolete rows. This view is accessible to all users, however non-superusers will only be able to see the tables that they have permission to access.

Table I.1  gp_bloat_diag view

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bdirelid</td>
<td>Table object id.</td>
</tr>
<tr>
<td>bdinspname</td>
<td>Schema name.</td>
</tr>
<tr>
<td>bdirelname</td>
<td>Table name.</td>
</tr>
<tr>
<td>bdirelpages</td>
<td>Actual number of pages on disk.</td>
</tr>
<tr>
<td>bdixpppages</td>
<td>Expected number of pages given the table data.</td>
</tr>
<tr>
<td>bdidiag</td>
<td>Bloat diagnostic message.</td>
</tr>
</tbody>
</table>

gp_stats_missing
This view shows tables that do not have statistics and therefore may require an ANALYZE be run on the table.

Table I.2  gp_stats_missing view

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>smischema</td>
<td>Schema name.</td>
</tr>
<tr>
<td>smitable</td>
<td>Table name.</td>
</tr>
<tr>
<td>smisize</td>
<td>Does this table have statistics? False if the table does not have row count and row sizing statistics recorded in the system catalog, which may indicate that the table needs to be analyzed. This will also be false if the table does not contain any rows. For example, the parent tables of partitioned tables are always empty and will always return a false result.</td>
</tr>
<tr>
<td>smicols</td>
<td>Number of columns in the table.</td>
</tr>
<tr>
<td>smirecs</td>
<td>Number of rows in the table.</td>
</tr>
</tbody>
</table>

Checking for Locks
When a transaction accesses a relation (such as a table), it acquires a lock. Depending on the type of lock acquired, subsequent transactions may have to wait before they can access the same relation. For more information on the types of locks, see “Lock Modes in Greenplum Database” on page 95. Greenplum Database resource queues (used for workload management) also use locks to control the admission of queries into the system.
The \textit{gp\_locks\_*} family of views can help diagnose queries and sessions that are waiting to access an object due to a lock.

- \textit{gp\_locks\_on\_relation}
- \textit{gp\_locks\_on\_resqueue}

\textbf{gp\_locks\_on\_relation}

This view shows any locks currently being held on a relation, and the associated session information about the query associated with the lock. For more information on the types of locks, see “Lock Modes in Greenplum Database” on page 95. This view is accessible to all users, however non-superusers will only be able to see the locks for relations that they have permission to access.

\textbf{Table I.3} \textit{gp\_locks\_on\_relation} view

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lorlocktype</td>
<td>Type of the lockable object: relation, extend, page, tuple, transactionid, object, userlock, resource queue, or advisory</td>
</tr>
<tr>
<td>lordatabase</td>
<td>Object ID of the database in which the object exists, zero if the object is a shared object.</td>
</tr>
<tr>
<td>lorrelname</td>
<td>The name of the relation.</td>
</tr>
<tr>
<td>lorrelation</td>
<td>The object ID of the relation.</td>
</tr>
<tr>
<td>lortransaction</td>
<td>The transaction ID that is affected by the lock.</td>
</tr>
<tr>
<td>lorpid</td>
<td>Process ID of the server process holding or awaiting this lock. NULL if the lock is held by a prepared transaction.</td>
</tr>
<tr>
<td>lormode</td>
<td>Name of the lock mode held or desired by this process.</td>
</tr>
<tr>
<td>lorgranted</td>
<td>Displays whether the lock is granted (true) or not granted (false).</td>
</tr>
<tr>
<td>lorcurrentquery</td>
<td>The current query in the session.</td>
</tr>
</tbody>
</table>

\textbf{gp\_locks\_on\_resqueue}

This view shows any locks currently being held on a resource queue, and the associated session information about the query associated with the lock. This view is accessible to all users, however non-superusers will only be able to see the locks associated with their own sessions.

\textbf{Table I.4} \textit{gp\_locks\_on\_resqueue} view

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lorusename</td>
<td>Name of the user executing the session.</td>
</tr>
<tr>
<td>lorrsqname</td>
<td>The resource queue name.</td>
</tr>
<tr>
<td>lorlocktype</td>
<td>Type of the lockable object: resource queue</td>
</tr>
<tr>
<td>lorobjid</td>
<td>The ID of the locked transaction.</td>
</tr>
<tr>
<td>lortransaction</td>
<td>The ID of the transaction that is affected by the lock.</td>
</tr>
</tbody>
</table>
Viewing Greenplum Database Server Log Files

Each component of a Greenplum Database system (master, standby master, primary segments, and mirror segments) keeps its own server log files. The `gp_log_*` family of views allows you to issue SQL queries against the server log files to find particular entries of interest. The use of these views require superuser permissions.

- `gp_log_command_timings`
- `gp_log_database`
- `gp_log_master_concise`
- `gp_log_system`

**gp_log_command_timings**

This view uses an external table to read the log files on the master and report the execution time of SQL commands executed in a database session. The use of this view requires superuser permissions.

**Table I.5  gp_log_command_timings view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>logsession</td>
<td>The session identifier (prefixed with &quot;con&quot;).</td>
</tr>
<tr>
<td>logcmdcount</td>
<td>The command number within a session (prefixed with &quot;cmd&quot;).</td>
</tr>
<tr>
<td>logdatabase</td>
<td>The name of the database.</td>
</tr>
<tr>
<td>loguser</td>
<td>The name of the database user.</td>
</tr>
<tr>
<td>logpid</td>
<td>The process id (prefixed with &quot;p&quot;).</td>
</tr>
<tr>
<td>logtimemin</td>
<td>The time of the first log message for this command.</td>
</tr>
<tr>
<td>logtimemax</td>
<td>The time of the last log message for this command.</td>
</tr>
<tr>
<td>logduration</td>
<td>Statement duration from start to end time.</td>
</tr>
</tbody>
</table>
gp_log_database

This view uses an external table to read the server log files of the entire Greenplum system (master, segments, and mirrors) and lists log entries associated with the current database. Associated log entries can be identified by the session id (logsession) and command id (logcmdcount). The use of this view requires superuser permissions.

Table I.6  gp_log_database view

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>logtime</td>
<td>The timestamp of the log message.</td>
</tr>
<tr>
<td>loguser</td>
<td>The name of the database user.</td>
</tr>
<tr>
<td>logdatabase</td>
<td>The name of the database.</td>
</tr>
<tr>
<td>logpid</td>
<td>The associated process id (prefixed with &quot;p&quot;).</td>
</tr>
<tr>
<td>logthread</td>
<td>The associated thread count (prefixed with &quot;th&quot;).</td>
</tr>
<tr>
<td>loghost</td>
<td>The segment or master host name.</td>
</tr>
<tr>
<td>logport</td>
<td>The segment or master port.</td>
</tr>
<tr>
<td>logsessiontime</td>
<td>Time session connection was opened.</td>
</tr>
<tr>
<td>logtransaction</td>
<td>Global transaction id.</td>
</tr>
<tr>
<td>logsession</td>
<td>The session identifier (prefixed with &quot;con&quot;).</td>
</tr>
<tr>
<td>logcmdcount</td>
<td>The command number within a session (prefixed with &quot;cmd&quot;).</td>
</tr>
<tr>
<td>logsegment</td>
<td>The segment content identifier (prefixed with &quot;seg&quot; for primary or &quot;mir&quot; for mirror. The master always has a content id of -1).</td>
</tr>
<tr>
<td>logslice</td>
<td>The slice id (portion of the query plan being executed).</td>
</tr>
<tr>
<td>logdistxact</td>
<td>Distributed transaction id.</td>
</tr>
<tr>
<td>loglocalxact</td>
<td>Local transaction id.</td>
</tr>
<tr>
<td>logsubxact</td>
<td>Subtransaction id.</td>
</tr>
<tr>
<td>logseverity</td>
<td>LOG, ERROR, FATAL, PANIC, DEBUG1 or DEBUG2.</td>
</tr>
<tr>
<td>logstate</td>
<td>SQL state code associated with the log message.</td>
</tr>
<tr>
<td>logmessage</td>
<td>Log or error message text.</td>
</tr>
<tr>
<td>logdetail</td>
<td>Detail message text associated with an error message.</td>
</tr>
<tr>
<td>loghint</td>
<td>Hint message text associated with an error message.</td>
</tr>
<tr>
<td>logquery</td>
<td>The internally-generated query text.</td>
</tr>
<tr>
<td>logquerypos</td>
<td>The cursor index into the internally-generated query text.</td>
</tr>
<tr>
<td>logcontext</td>
<td>The context in which this message gets generated.</td>
</tr>
<tr>
<td>logdebug</td>
<td>Query string with full detail for debugging.</td>
</tr>
<tr>
<td>logcursorpos</td>
<td>The cursor index into the query string.</td>
</tr>
<tr>
<td>logfunction</td>
<td>The function in which this message is generated.</td>
</tr>
<tr>
<td>logfile</td>
<td>The log file in which this message is generated.</td>
</tr>
</tbody>
</table>
Table I.6  *gp_log_database* view

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>logline</td>
<td>The line in the log file in which this message is generated.</td>
</tr>
<tr>
<td>logstack</td>
<td>Full text of the stack trace associated with this message.</td>
</tr>
</tbody>
</table>

**gp_log_master_concise**

This view uses an external table to read a subset of the log fields from the master log file. The use of this view requires superuser permissions.

Table I.7  *gp_log_master_concise* view

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>logtime</td>
<td>The timestamp of the log message.</td>
</tr>
<tr>
<td>logdatabase</td>
<td>The name of the database.</td>
</tr>
<tr>
<td>logsession</td>
<td>The session identifier (prefixed with &quot;con&quot;).</td>
</tr>
<tr>
<td>logcmdcount</td>
<td>The command number within a session (prefixed with &quot;cmd&quot;).</td>
</tr>
<tr>
<td>logmessage</td>
<td>Log or error message text.</td>
</tr>
</tbody>
</table>

**gp_log_system**

This view uses an external table to read the server log files of the entire Greenplum system (master, segments, and mirrors) and lists all log entries. Associated log entries can be identified by the session id (*logsession*) and command id (*logcmdcount*). The use of this view requires superuser permissions.

Table I.8  *gp_log_system* view

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>logtime</td>
<td>The timestamp of the log message.</td>
</tr>
<tr>
<td>loguser</td>
<td>The name of the database user.</td>
</tr>
<tr>
<td>logdatabase</td>
<td>The name of the database.</td>
</tr>
<tr>
<td>logpid</td>
<td>The associated process id (prefixed with &quot;p&quot;).</td>
</tr>
<tr>
<td>logthread</td>
<td>The associated thread count (prefixed with &quot;th&quot;).</td>
</tr>
<tr>
<td>loghost</td>
<td>The segment or master host name.</td>
</tr>
<tr>
<td>logport</td>
<td>The segment or master port.</td>
</tr>
<tr>
<td>logsessiontime</td>
<td>Time session connection was opened.</td>
</tr>
<tr>
<td>logtransaction</td>
<td>Global transaction id.</td>
</tr>
<tr>
<td>logsession</td>
<td>The session identifier (prefixed with &quot;con&quot;).</td>
</tr>
<tr>
<td>logcmdcount</td>
<td>The command number within a session (prefixed with &quot;cmd&quot;).</td>
</tr>
<tr>
<td>logsegment</td>
<td>The segment content identifier (prefixed with &quot;seg&quot; for primary or &quot;mir&quot; for mirror. The master always has a content id of -1).</td>
</tr>
</tbody>
</table>
Checking Server Configuration Files

Each component of a Greenplum Database system (master, standby master, primary segments, and mirror segments) has its own server configuration file (`postgresql.conf`). The following `gp_toolkit` objects can be used to check parameter settings across all primary `postgresql.conf` files in the system:

- `gp_param_setting('parameter_name')`
- `gp_param_settings_seg_value_diffs`

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>logslice</td>
<td>The slice id (portion of the query plan being executed).</td>
</tr>
<tr>
<td>logdistxact</td>
<td>Distributed transaction id.</td>
</tr>
<tr>
<td>loglocalxact</td>
<td>Local transaction id.</td>
</tr>
<tr>
<td>logsubxact</td>
<td>Subtransaction id.</td>
</tr>
<tr>
<td>logseverity</td>
<td>LOG, ERROR, FATAL, PANIC, DEBUG1 or DEBUG2.</td>
</tr>
<tr>
<td>logstate</td>
<td>SQL state code associated with the log message.</td>
</tr>
<tr>
<td>logmessage</td>
<td>Log or error message text.</td>
</tr>
<tr>
<td>logdetail</td>
<td>Detail message text associated with an error message.</td>
</tr>
<tr>
<td>loghint</td>
<td>Hint message text associated with an error message.</td>
</tr>
<tr>
<td>logquery</td>
<td>The internally-generated query text.</td>
</tr>
<tr>
<td>logquerypos</td>
<td>The cursor index into the internally-generated query text.</td>
</tr>
<tr>
<td>logcontext</td>
<td>The context in which this message gets generated.</td>
</tr>
<tr>
<td>logdebug</td>
<td>Query string with full detail for debugging.</td>
</tr>
<tr>
<td>logcursorpos</td>
<td>The cursor index into the query string.</td>
</tr>
<tr>
<td>logfunction</td>
<td>The function in which this message is generated.</td>
</tr>
<tr>
<td>logfile</td>
<td>The log file in which this message is generated.</td>
</tr>
<tr>
<td>logline</td>
<td>The line in the log file in which this message is generated.</td>
</tr>
<tr>
<td>logstack</td>
<td>Full text of the stack trace associated with this message.</td>
</tr>
</tbody>
</table>
gp_param_setting('parameter_name')

This function takes the name of a server configuration parameter and returns the `postgresql.conf` value for the master and each active segment. This function is accessible to all users.

**Table I.9  gp_param_setting('parameter_name') function**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>paramsegment</td>
<td>The segment content id (only active segments are shown). The master content id is always -1.</td>
</tr>
<tr>
<td>paramname</td>
<td>The name of the parameter.</td>
</tr>
<tr>
<td>paramvalue</td>
<td>The value of the parameter.</td>
</tr>
</tbody>
</table>

**Example:**

```
SELECT * FROM gp_param_setting('max_connections');
```

gp_param_settings_seg_value_diffs

Server configuration parameters that are classified as *local* parameters (meaning each segment gets the parameter value from its own `postgresql.conf` file), should be set identically on all segments. This view shows local parameter settings that are inconsistent. Parameters that are supposed to have different values (such as `port`) are not included. This view is accessible to all users.

**Table I.10  gp_param_settings_seg_value_diffs view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>psdname</td>
<td>The name of the parameter.</td>
</tr>
<tr>
<td>psdvalue</td>
<td>The value of the parameter.</td>
</tr>
<tr>
<td>psdcount</td>
<td>The number of segments that have this value.</td>
</tr>
</tbody>
</table>

Checking for Failed Segments

The `gp_pgdatabase_invalid` view can be used to check for down segments.

**gp_pgdatabase_invalid**

This view shows information about segments that are marked as down in the system catalog. This view is accessible to all users.

**Table I.11  gp_pgdatabase_invalid view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pgdbidbid</td>
<td>The segment dbid. Every segment has a unique dbid.</td>
</tr>
<tr>
<td>pgdbisprimary</td>
<td>Is the segment currently acting as the primary (active) segment? (t or f)</td>
</tr>
</tbody>
</table>
Checking Resource Queue Activity and Status

The purpose of resource queues is to limit the number of active queries in the system at any given time in order to avoid exhausting system resources such as memory, CPU, and disk I/O. All database users are assigned to a resource queue, and every statement submitted by a user is first evaluated against the resource queue limits before it can run. The `gp_resq_*` family of views can be used to check the status of statements currently submitted to the system through their respective resource queue. Note that statements issued by superusers are exempt from resource queuing.

- `gp_resq_activity`
- `gp_resq_activity_by_queue`
- `gp_resq_priority_statement`
- `gp_resq_role`

**gp_resq_activity**

For the resource queues that have active workload, this view shows one row for each active statement submitted through a resource queue. This view is accessible to all users.

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>resqprocid</td>
<td>Process ID assigned to this statement (on the master).</td>
</tr>
<tr>
<td>resqrole</td>
<td>User name.</td>
</tr>
<tr>
<td>resqoid</td>
<td>Resource queue object id.</td>
</tr>
<tr>
<td>resqname</td>
<td>Resource queue name.</td>
</tr>
<tr>
<td>resqstart</td>
<td>Time statement was issued to the system.</td>
</tr>
<tr>
<td>resqstatus</td>
<td>Status of statement: running, waiting or cancelled.</td>
</tr>
</tbody>
</table>

**Table I.11  gp_pgdatabase_invalid view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pgdbicontent</td>
<td>The content id of this segment. A primary and mirror will have the same content id.</td>
</tr>
<tr>
<td>pgdbivalid</td>
<td>Is this segment up and valid? (t or f)</td>
</tr>
<tr>
<td>pgdbidefinedprimary</td>
<td>Was this segment assigned the role of primary at system initialization time? (t or f)</td>
</tr>
</tbody>
</table>
**gp_resq_activity_by_queue**

For the resource queues that have active workload, this view shows a summary of queue activity. This view is accessible to all users.

**Table I.13  gp_resq_activity_by_queue Column**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>resqoid</td>
<td>Resource queue object id.</td>
</tr>
<tr>
<td>resqname</td>
<td>Resource queue name.</td>
</tr>
<tr>
<td>resqlast</td>
<td>Time of the last statement issued to the queue.</td>
</tr>
<tr>
<td>reqstatus</td>
<td>Status of last statement: running, waiting or cancelled.</td>
</tr>
<tr>
<td>resqtotal</td>
<td>Total statements in this queue.</td>
</tr>
</tbody>
</table>

**gp_resq_priority_statement**

This view shows the resource queue priority, session ID, and other information for all statements currently running in the Greenplum Database system. This view is accessible to all users.

**Table I.14  gp_resq_priority_statement view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rqpdatname</td>
<td>The database name that the session is connected to.</td>
</tr>
<tr>
<td>rqpusename</td>
<td>The user who issued the statement.</td>
</tr>
<tr>
<td>rqpsession</td>
<td>The session ID.</td>
</tr>
<tr>
<td>rqpcommand</td>
<td>The number of the statement within this session (the command id and session id uniquely identify a statement).</td>
</tr>
<tr>
<td>rqppriority</td>
<td>The resource queue priority for this statement (MAX, HIGH, MEDIUM, LOW).</td>
</tr>
<tr>
<td>rqpweight</td>
<td>An integer value associated with the priority of this statement.</td>
</tr>
<tr>
<td>rqpquery</td>
<td>The query text of the statement.</td>
</tr>
</tbody>
</table>

**gp_resq_role**

This view shows the resource queue associated with a role. This view is accessible to all users.

**Table I.15  gp_resq_role view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rrrolname</td>
<td>Role (user) name.</td>
</tr>
<tr>
<td>rrsqname</td>
<td>The resource queue name assigned to this role. If a role has not been explicitly assigned to a resource queue, it will be in the default resource queue (pg_default).</td>
</tr>
</tbody>
</table>
Viewing Users and Groups (Roles)

It is frequently convenient to group users (roles) together to ease management of object privileges: that way, privileges can be granted to, or revoked from, a group as a whole. In Greenplum Database this is done by creating a role that represents the group, and then granting membership in the group role to individual user roles.

The `gp_roles_assigned` view can be used to see all of the roles in the system, and their assigned members (if the role is also a group role).

**gp_roles_assigned**

This view shows all of the roles in the system, and their assigned members (if the role is also a group role). This view is accessible to all users.

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>raroleid</td>
<td>The role object ID. If this role has members (users), it is considered a group role.</td>
</tr>
<tr>
<td>rrolename</td>
<td>The role (user or group) name.</td>
</tr>
<tr>
<td>ramemberid</td>
<td>The role object ID of the role that is a member of this role.</td>
</tr>
<tr>
<td>ramembername</td>
<td>Name of the role that is a member of this role.</td>
</tr>
</tbody>
</table>

Checking Database Object Sizes and Disk Space

The `gp_size_*` family of views can be used to determine the disk space usage for a distributed Greenplum database, schema, table, or index. The following views calculate the total size of an object across all primary segments (mirrors are not included in the size calculations).

- `gp_size_of_all_table_indexes`
- `gp_size_of_database`
- `gp_size_of_index`
- `gp_size_of_partition_and_indexes_disk`
- `gp_size_of_schema_disk`
- `gp_size_of_table_and_indexes_disk`
- `gp_size_of_table_and_indexes_licensing`
- `gp_size_of_table_disk`
- `gp_size_of_table_uncompressed`
- `gp_disk_free`

The table and index sizing views list the relation by object ID (not by name). To check the size of a table or index by name, you must look up the relation name (`relname`) in the `pg_class` table. For example:
SELECT relname as name, sotdsize as size, sotdtoastsize as toast, sotdadditionalsize as other
FROM gp_size_of_table_disk as sotd, pg_class
WHERE sotd.sotdoid=pg_class.oid ORDER BY relname;

**gp_size_of_all_table_indexes**
This view shows the total size of all indexes for a table. This view is accessible to all users, however non-superusers will only be able to see relations that they have permission to access.

**Table I.17 gp_size_of_all_table_indexes view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>soatioid</td>
<td>Object ID of the table.</td>
</tr>
<tr>
<td>soatitisize</td>
<td>Total size of all table indexes in bytes.</td>
</tr>
</tbody>
</table>

**gp_size_of_database**
This view shows the total size of a database. This view is accessible to all users, however non-superusers will only be able to see databases that they have permission to access.

**Table I.18 gp_size_of_database view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sodddatname</td>
<td>The name of the database.</td>
</tr>
<tr>
<td>sodddatsize</td>
<td>The size of the database in bytes.</td>
</tr>
</tbody>
</table>

**gp_size_of_index**
This view shows the total size of an index. This view is accessible to all users, however non-superusers will only be able to see relations that they have permission to access.

**Table I.19 gp_size_of_index view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>soioid</td>
<td>The object ID of the index.</td>
</tr>
<tr>
<td>soitableoid</td>
<td>Object ID of the table to which the index belongs.</td>
</tr>
<tr>
<td>soisize</td>
<td>The size of the index in bytes.</td>
</tr>
</tbody>
</table>
**gp_size_of_partition_and_indexes_disk**

This view shows the size on disk of partitioned child tables and their indexes. This view is accessible to all users, however non-superusers will only be able to see relations that they have permission to access.

**Table I.20 gp_size_of_partition_and_indexes_disk view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sopaidparentoid</td>
<td>Object ID of the parent table.</td>
</tr>
<tr>
<td>sopaidpartitionoid</td>
<td>Object ID of the partition table.</td>
</tr>
<tr>
<td>sopaidpartitiontablesizer</td>
<td>Partition table size in bytes.</td>
</tr>
<tr>
<td>sopaidpartitionindexesizer</td>
<td>Total size of all indexes on this partition.</td>
</tr>
</tbody>
</table>

**gp_size_of_schema_disk**

This view shows schema sizes for the schemas in the current database. This view is accessible to all users, however non-superusers will only be able to see schemas that they have permission to access.

**Table I.21 gp_size_of_schema_disk view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sosdnnsp</td>
<td>The name of the schema.</td>
</tr>
<tr>
<td>sosdschematablesizer</td>
<td>The total size of tables in the schema in bytes.</td>
</tr>
<tr>
<td>sosdschemaidxsize</td>
<td>The total size of indexes in the schema in bytes.</td>
</tr>
</tbody>
</table>

**gp_size_of_table_and_indexes_disk**

This view shows the size on disk of tables and their indexes. This view is accessible to all users, however non-superusers will only be able to see relations that they have permission to access.

**Table I.22 gp_size_of_table_and_indexes_disk view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sotaidoid</td>
<td>The object ID of the parent table.</td>
</tr>
<tr>
<td>sotaidatablesize</td>
<td>The disk size of the table.</td>
</tr>
<tr>
<td>sotaididxsize</td>
<td>The total size of all indexes on the table.</td>
</tr>
</tbody>
</table>
gp_size_of_table_and_indexes_licensing

This view shows the total size of tables and their indexes for licensing purposes. The use of this view requires superuser permissions.

**Table I.23 gp_size_of_table_and_indexes_licensing view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sotailoid</td>
<td>The object ID of the table.</td>
</tr>
<tr>
<td>sotailtablesizedish</td>
<td>The total disk size of the table.</td>
</tr>
<tr>
<td>sotailtablesizeuncompressed</td>
<td>If the table is a compressed append-only table, shows the uncompressed table size in bytes.</td>
</tr>
<tr>
<td>sotailindexessize</td>
<td>Total size of all indexes in the table.</td>
</tr>
</tbody>
</table>

gp_size_of_table_disk

This view shows the size of a table on disk. This view is accessible to all users, however non-superusers will only be able to see tables that they have permission to access.

**Table I.24 gp_size_of_table_disk view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sotdoid</td>
<td>Object ID of the table.</td>
</tr>
<tr>
<td>sotdsize</td>
<td>Total size of the table in bytes (main relation, plus oversized (toast) attributes, plus additional storage objects for AO tables).</td>
</tr>
<tr>
<td>sotdtoastsize</td>
<td>The size of the TOAST table (oversized attribute storage), if there is one.</td>
</tr>
<tr>
<td>sotdadditionalsize</td>
<td>Reflects the segment and block directory table sizes for append-only (AO) tables.</td>
</tr>
</tbody>
</table>

gp_size_of_table_uncompressed

This view shows the uncompressed table size for append-only (AO) tables. Otherwise, the table size on disk is shown. The use of this view requires superuser permissions.

**Table I.25 gp_size_of_table_uncompressed view**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sotuoid</td>
<td>The object ID of the table.</td>
</tr>
<tr>
<td>sotusize</td>
<td>The uncompressed size of the table in bytes if it is a compressed AO table. Otherwise, the table size on disk.</td>
</tr>
</tbody>
</table>
gp_disk_free

This external table runs the `df` (disk free) command on the active segment hosts and reports back the results. Inactive mirrors are not included in the calculation. The use of this external table requires superuser permissions.

Table I.26  gp_disk_free external table

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dfsegment</td>
<td>The content id of the segment (only active segments are shown).</td>
</tr>
<tr>
<td>dfhostname</td>
<td>Hostname of the segment host.</td>
</tr>
<tr>
<td>dfdevice</td>
<td>Device name.</td>
</tr>
<tr>
<td>dfspace</td>
<td>Free disk space in the segment file system in bytes.</td>
</tr>
</tbody>
</table>

Checking for Uneven Data Distribution

All tables in Greenplum Database are distributed, meaning their data is divided across all of the segments in the system. If the data is not distributed evenly, then query processing performance may suffer. The following views can help diagnose if a table has uneven data distribution:

- `gp_skew_coefficients`
- `gp_skew_idle_fractions`

gp_skew_coefficients

This view shows data distribution skew by calculating the coefficient of variation (CV) for the data stored on each segment. This view is accessible to all users, however non-superusers will only be able to see tables that they have permission to access.

Table I.27  gp_skew_coefficients view

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>skcoid</td>
<td>The object id of the table.</td>
</tr>
<tr>
<td>skcnamespace</td>
<td>The namespace where the table is defined.</td>
</tr>
<tr>
<td>skcrelname</td>
<td>The table name.</td>
</tr>
<tr>
<td>skccoef</td>
<td>The coefficient of variation (CV) is calculated as the standard deviation divided by the average. It takes into account both the average and variability around the average of a data series. The lower the value, the better. Higher values indicate greater data skew.</td>
</tr>
</tbody>
</table>
gp_skew_idle_fractions

This view shows data distribution skew by calculating the percentage of the system that is idle during a table scan, which is an indicator of processing data skew. This view is accessible to all users, however non-superusers will only be able to see tables that they have permission to access.

Table I.28  gp_skew_idle_fractions view

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sifoid</td>
<td>The object id of the table.</td>
</tr>
<tr>
<td>sifnamespace</td>
<td>The namespace where the table is defined.</td>
</tr>
<tr>
<td>sifrelname</td>
<td>The table name.</td>
</tr>
<tr>
<td>siffraction</td>
<td>The percentage of the system that is idle during a table scan, which is an indicator of uneven data distribution or query processing skew. For example, a value of 0.1 indicates 10% skew, a value of 0.5 indicates 50% skew, and so on. Tables that have more than 10% skew should have their distribution policies evaluated.</td>
</tr>
</tbody>
</table>
J. Character Set Support

The character set support in Greenplum Database allows you to store text in a variety of character sets, including single-byte character sets such as the ISO 8859 series and multiple-byte character sets such as EUC (Extended Unix Code), UTF-8, and Mule internal code. All supported character sets can be used transparently by clients, but a few are not supported for use within the server (that is, as a server-side encoding). The default character set is selected while initializing your Greenplum Database array using `gpinitsystem`. It can be overridden when you create a database, so you can have multiple databases each with a different character set.

### Table J.1 Greenplum Database Character Sets

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Language</th>
<th>Server?</th>
<th>Bytes/Char</th>
<th>Aliases</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIG5</td>
<td>Big Five</td>
<td>Traditional Chinese</td>
<td>No</td>
<td>1-2</td>
<td>WIN950, Windows950</td>
</tr>
<tr>
<td>EUC-CN</td>
<td>Extended UNIX Code-CN</td>
<td>Simplified Chinese</td>
<td>Yes</td>
<td>1-3</td>
<td></td>
</tr>
<tr>
<td>EUC_TW</td>
<td>Extended UNIX Code-TW</td>
<td>Traditional Chinese, Taiwanese</td>
<td>Yes</td>
<td>1-3</td>
<td></td>
</tr>
<tr>
<td>GB18030</td>
<td>National Standard</td>
<td>Chinese</td>
<td>No</td>
<td>1-2</td>
<td></td>
</tr>
<tr>
<td>GBK</td>
<td>Extended National Standard</td>
<td>Simplified Chinese</td>
<td>No</td>
<td>1-2</td>
<td>WIN936, Windows936</td>
</tr>
<tr>
<td>ISO_8859_5</td>
<td>ISO 8859-5, ECMA 113</td>
<td>Latin/Cyrillic</td>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ISO_8859_6</td>
<td>ISO 8859-6, ECMA 114</td>
<td>Latin/Arabic</td>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ISO_8859_7</td>
<td>ISO 8859-7, ECMA 118</td>
<td>Latin/Greek</td>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ISO_8859_8</td>
<td>ISO 8859-8, ECMA 121</td>
<td>Latin/Hebrew</td>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>JOHAB</td>
<td>JOHA</td>
<td>Korean (Hangul)</td>
<td>Yes</td>
<td>1-3</td>
<td></td>
</tr>
<tr>
<td>KOI8</td>
<td>KOI8-R(U)</td>
<td>Cyrillic</td>
<td>Yes</td>
<td>1</td>
<td>KOI8R</td>
</tr>
<tr>
<td>LATIN1</td>
<td>ISO 8859-1, ECMA 94</td>
<td>Western European</td>
<td>Yes</td>
<td>1</td>
<td>ISO88591</td>
</tr>
<tr>
<td>LATIN2</td>
<td>ISO 8859-2, ECMA 94</td>
<td>Central European</td>
<td>Yes</td>
<td>1</td>
<td>ISO88592</td>
</tr>
<tr>
<td>LATIN3</td>
<td>ISO 8859-3, ECMA 94</td>
<td>South European</td>
<td>Yes</td>
<td>1</td>
<td>ISO88593</td>
</tr>
<tr>
<td>LATIN4</td>
<td>ISO 8859-4, ECMA 94</td>
<td>North European</td>
<td>Yes</td>
<td>1</td>
<td>ISO88594</td>
</tr>
<tr>
<td>LATIN5</td>
<td>ISO 8859-9, ECMA 128</td>
<td>Turkish</td>
<td>Yes</td>
<td>1</td>
<td>ISO88599</td>
</tr>
<tr>
<td>LATIN6</td>
<td>ISO 8859-10, ECMA 144</td>
<td>Nordic</td>
<td>Yes</td>
<td>1</td>
<td>ISO885910</td>
</tr>
<tr>
<td>LATIN7</td>
<td>ISO 8859-13</td>
<td>Baltic</td>
<td>Yes</td>
<td>1</td>
<td>ISO885913</td>
</tr>
<tr>
<td>LATIN8</td>
<td>ISO 8859-14</td>
<td>Celtic</td>
<td>Yes</td>
<td>1</td>
<td>ISO885914</td>
</tr>
<tr>
<td>LATIN9</td>
<td>ISO 8859-15</td>
<td>LATIN1 with Euro and accents</td>
<td>Yes</td>
<td>1</td>
<td>ISO885915</td>
</tr>
</tbody>
</table>
Setting the Character Set

gpinit��统 defines the default character set for a Greenplum Database system by reading the setting of the ENCODING parameter in the gp_init_config file at initialization time. The default character set is UNICODE or UTF8.

You can create a database with a different character set besides what is used as the system-wide default. For example:

```sql
=> CREATE DATABASE korean WITH ENCODING 'EUC_KR';
```

Important: Although you can specify any encoding you want for a database, it is unwise to choose an encoding that is not what is expected by the locale you have selected. The LC_COLLATE and LC_CTYPE settings imply a particular encoding, and locale-dependent operations (such as sorting) are likely to misinterpret data that is in an incompatible encoding.

---

**Table J.1** Greenplum Database Character Sets

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Language</th>
<th>Server?</th>
<th>Bytes/Char</th>
<th>Aliases</th>
</tr>
</thead>
<tbody>
<tr>
<td>LATIN10</td>
<td>ISO 8859-16, ASRO SR 14111</td>
<td>Romanian</td>
<td>Yes</td>
<td>1</td>
<td>ISO885916</td>
</tr>
<tr>
<td>MULE_INTERNAL</td>
<td>Mule internal code</td>
<td>Multilingual Emacs</td>
<td>Yes</td>
<td>1-4</td>
<td>ISO885916, MULE_INTERNAL</td>
</tr>
<tr>
<td>SJIS</td>
<td>Shift JIS</td>
<td>Japanese</td>
<td>No</td>
<td>1-2</td>
<td>ShiftJIS, WIN932, Windows932</td>
</tr>
<tr>
<td>SQL_ASCII</td>
<td>unspecified²</td>
<td>any</td>
<td>No</td>
<td>1</td>
<td>ISO885916</td>
</tr>
<tr>
<td>UHC</td>
<td>Unified Hangul Code</td>
<td>Korean</td>
<td>No</td>
<td>1-2</td>
<td>WIN949, Windows949</td>
</tr>
<tr>
<td>UTF8</td>
<td>Unicode, 8-bit</td>
<td>all</td>
<td>Yes</td>
<td>1-4</td>
<td>Unicode</td>
</tr>
<tr>
<td>WIN866</td>
<td>Windows CP866</td>
<td>Cyrillic</td>
<td>Yes</td>
<td>1</td>
<td>ALTVIN949</td>
</tr>
<tr>
<td>WIN874</td>
<td>Windows CP874</td>
<td>Thai</td>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>WIN1250</td>
<td>Windows CP1250</td>
<td>Central European</td>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>WIN1251</td>
<td>Windows CP1251</td>
<td>Cyrillic</td>
<td>Yes</td>
<td>1</td>
<td>WIN</td>
</tr>
<tr>
<td>WIN1252</td>
<td>Windows CP1252</td>
<td>Western European</td>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>WIN1253</td>
<td>Windows CP1253</td>
<td>Greek</td>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>WIN1254</td>
<td>Windows CP1254</td>
<td>Turkish</td>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>WIN1255</td>
<td>Windows CP1255</td>
<td>Hebrew</td>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>WIN1256</td>
<td>Windows CP1256</td>
<td>Arabic</td>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>WIN1257</td>
<td>Windows CP1257</td>
<td>Baltic</td>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>WIN1258</td>
<td>Windows CP1258</td>
<td>Vietnamese</td>
<td>Yes</td>
<td>1</td>
<td>ABC, TCVN, TCVN5712, VSCII</td>
</tr>
</tbody>
</table>

1. Not all APIs support all the listed character sets. For example, the JDBC driver does not support MULE_INTERNAL, LATIN6, LATIN8, and LATIN10.
2. The SQL ASCII setting behaves considerably differently from the other settings. The ASCII standard, while byte values 128-255 are taken as uninterpreted characters. If you are working with any non-ASCII data, it is unwise to use the SQL_ASCII setting as a client encoding. SQL_ASCII is not supported as a server encoding.
Since these locale settings are frozen by `gpinitparam`, the apparent flexibility to use different encodings in different databases is more theoretical than real.

One way to use multiple encodings safely is to set the locale to `C` or `POSIX` during initialization time, thus disabling any real locale awareness.

### Character Set Conversion Between Server and Client

Greenplum Database supports automatic character set conversion between server and client for certain character set combinations. The conversion information is stored in the master `pg_conversion` system catalog table. Greenplum Database comes with some predefined conversions or you can create a new conversion using the SQL command `CREATE CONVERSION`.

**Table J.2 Client/Server Character Set Conversions**

<table>
<thead>
<tr>
<th>Server Character Set</th>
<th>Available Client Character Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIG5</td>
<td>not supported as a server encoding</td>
</tr>
<tr>
<td>EUC_CN</td>
<td>EUC_CN, MULE_INTERNAL, UTF8</td>
</tr>
<tr>
<td>EUC_JP</td>
<td>EUC_JP, MULE_INTERNAL, SJIS, UTF8</td>
</tr>
<tr>
<td>EUC_KR</td>
<td>EUC_KR, MULE_INTERNAL, UTF8</td>
</tr>
<tr>
<td>EUC_TW</td>
<td>EUC_TW, BIG5, MULE_INTERNAL, UTF8</td>
</tr>
<tr>
<td>GB18030</td>
<td>not supported as a server encoding</td>
</tr>
<tr>
<td>GBK</td>
<td>not supported as a server encoding</td>
</tr>
<tr>
<td>ISO_8859_5</td>
<td>ISO_8859_5, KOI8, MULE_INTERNAL, UTF8, WIN866, WIN1251</td>
</tr>
<tr>
<td>ISO_8859_6</td>
<td>ISO_8859_6, UTF8</td>
</tr>
<tr>
<td>ISO_8859_7</td>
<td>ISO_8859_7, UTF8</td>
</tr>
<tr>
<td>ISO_8859_8</td>
<td>ISO_8859_8, UTF8</td>
</tr>
<tr>
<td>JOHAB</td>
<td>JOHAB, UTF8</td>
</tr>
<tr>
<td>KOI8</td>
<td>KOI8, ISO_8859_5, MULE_INTERNAL, UTF8, WIN866, WIN1251</td>
</tr>
<tr>
<td>LATIN1</td>
<td>LATIN1, MULE_INTERNAL, UTF8</td>
</tr>
<tr>
<td>LATIN2</td>
<td>LATIN2, MULE_INTERNAL, UTF8, WIN1250</td>
</tr>
<tr>
<td>LATIN3</td>
<td>LATIN3, MULE_INTERNAL, UTF8</td>
</tr>
<tr>
<td>LATIN4</td>
<td>LATIN4, MULE_INTERNAL, UTF8</td>
</tr>
<tr>
<td>LATIN5</td>
<td>LATIN5, UTF8</td>
</tr>
<tr>
<td>LATIN6</td>
<td>LATIN6, UTF8</td>
</tr>
<tr>
<td>LATIN7</td>
<td>LATIN7, UTF8</td>
</tr>
<tr>
<td>LATIN8</td>
<td>LATIN8, UTF8</td>
</tr>
<tr>
<td>LATIN9</td>
<td>LATIN9, UTF8</td>
</tr>
<tr>
<td>LATIN10</td>
<td>LATIN10, UTF8</td>
</tr>
</tbody>
</table>
To enable automatic character set conversion, you have to tell Greenplum Database the character set (encoding) you would like to use in the client. There are several ways to accomplish this:

- **Using the \encoding command in psql**, which allows you to change client encoding on the fly.

- **Using SET client_encoding TO**. Setting the client encoding can be done with this SQL command:
  
  => SET CLIENT_ENCODING TO 'value';

  To query the current client encoding:
  
  => SHOW client_encoding;

  To return to the default encoding:
  
  => RESET client_encoding;

- **Using the PGCLIENTENCODING environment variable**. When PGCLIENTENCODING is defined in the client’s environment, that client encoding is automatically selected when a connection to the server is made. (This can subsequently be overridden using any of the other methods mentioned above.)

---

**Table J.2** Client/Server Character Set Conversions

<table>
<thead>
<tr>
<th>Server Character Set</th>
<th>Available Client Character Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>MULE_INTERNAL</td>
<td>MULE_INTERNAL, BIG5, EUC_CN, EUC_JP, EUC_KR, EUC_TW, ISO_8859_5, KOI8, LATIN1 to LATIN4, SJIS, WIN866, WIN1250, WIN1251</td>
</tr>
<tr>
<td>SJIS</td>
<td>not supported as a server encoding</td>
</tr>
<tr>
<td>SQL_ASCII</td>
<td>not supported as a server encoding</td>
</tr>
<tr>
<td>UHC</td>
<td>not supported as a server encoding</td>
</tr>
<tr>
<td>UTF8</td>
<td>all supported encodings</td>
</tr>
<tr>
<td>WIN866</td>
<td>WIN866</td>
</tr>
<tr>
<td>ISO_8859_5</td>
<td>KOI8, MULE_INTERNAL, UTF8, WIN1251</td>
</tr>
<tr>
<td>WIN874</td>
<td>WIN874, UTF8</td>
</tr>
<tr>
<td>WIN1250</td>
<td>WIN1250, LATIN2, MULE_INTERNAL, UTF8</td>
</tr>
<tr>
<td>WIN1251</td>
<td>WIN1251, ISO_8859_5, KOI8, MULE_INTERNAL, UTF8, WIN866</td>
</tr>
<tr>
<td>WIN1252</td>
<td>WIN1252, UTF8</td>
</tr>
<tr>
<td>WIN1253</td>
<td>WIN1253, UTF8</td>
</tr>
<tr>
<td>WIN1254</td>
<td>WIN1254, UTF8</td>
</tr>
<tr>
<td>WIN1255</td>
<td>WIN1255, UTF8</td>
</tr>
<tr>
<td>WIN1256</td>
<td>WIN1256, UTF8</td>
</tr>
<tr>
<td>WIN1257</td>
<td>WIN1257, UTF8</td>
</tr>
<tr>
<td>WIN1258</td>
<td>WIN1258, UTF8</td>
</tr>
</tbody>
</table>
• Setting the configuration parameter `client_encoding`. If `client_encoding` is set in the master `postgresql.conf` file, that client encoding is automatically selected when a connection to Greenplum Database is made. (This can subsequently be overridden using any of the other methods mentioned above.)

If the conversion of a particular character is not possible — suppose you chose `EUC_JP` for the server and `LATIN1` for the client, then some Japanese characters do not have a representation in `LATIN1` — then an error is reported.

If the client character set is defined as `SQL_ASCII`, encoding conversion is disabled, regardless of the server’s character set. The use of `SQL_ASCII` is unwise unless you are working with all-ASCII data. `SQL_ASCII` is not supported as a server encoding.
SQL 2008 Optional Feature Compliance

The following table is list of features described in the 2008 SQL standard. Features that are supported in Greenplum Database are marked as YES in the ‘Supported’ column, features that are not implemented are marked as NO.

For more information on Greenplum features and SQL compliance, see “Summary of Greenplum Features” on page 14.

Table K.1  SQL 2008 Optional Feature Compliance Details

<table>
<thead>
<tr>
<th>ID</th>
<th>Feature</th>
<th>Supported</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>B011</td>
<td>Embedded Ada</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B012</td>
<td>Embedded C</td>
<td>NO</td>
<td>Due to issues with PostgreSQL ecpg</td>
</tr>
<tr>
<td>B013</td>
<td>Embedded COBOL</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B014</td>
<td>Embedded Fortran</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B015</td>
<td>Embedded MUMPS</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B016</td>
<td>Embedded Pascal</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B017</td>
<td>Embedded PL/I</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B021</td>
<td>Direct SQL</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>B031</td>
<td>Basic dynamic SQL</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B032</td>
<td>Extended dynamic SQL</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B033</td>
<td>Untyped SQL-invoked function arguments</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B034</td>
<td>Dynamic specification of cursor attributes</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B035</td>
<td>Non-extended descriptor names</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B041</td>
<td>Extensions to embedded SQL exception declarations</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B051</td>
<td>Enhanced execution rights</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B111</td>
<td>Module language Ada</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B112</td>
<td>Module language C</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B113</td>
<td>Module language COBOL</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B114</td>
<td>Module language Fortran</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B115</td>
<td>Module language MUMPS</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B116</td>
<td>Module language Pascal</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B117</td>
<td>Module language PL/I</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B121</td>
<td>Routine language Ada</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B122</td>
<td>Routine language C</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B123</td>
<td>Routine language COBOL</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B124</td>
<td>Routine language Fortran</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Feature</td>
<td>Supported</td>
<td>Comments</td>
</tr>
<tr>
<td>------</td>
<td>-------------------------------------------------------------------------</td>
<td>-----------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>B125</td>
<td>Routine language MUMPS</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B126</td>
<td>Routine language Pascal</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B127</td>
<td>Routine language PL/I</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>B128</td>
<td>Routine language SQL</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>E011</td>
<td>Numeric data types</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E011-01</td>
<td>INTEGER and SMALLINT data types</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E011-02</td>
<td>DOUBLE PRECISION and FLOAT data types</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E011-03</td>
<td>DECIMAL and NUMERIC data types</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E011-04</td>
<td>Arithmetic operators</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E011-05</td>
<td>Numeric comparison</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E011-06</td>
<td>Implicit casting among the numeric data types</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E021</td>
<td>Character data types</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E021-01</td>
<td>CHARACTER data type</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E021-02</td>
<td>CHARACTER VARYING data type</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E021-03</td>
<td>Character literals</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E021-04</td>
<td>CHARACTER_LENGTH function</td>
<td>YES</td>
<td>Trims trailing spaces from CHARACTER values before counting</td>
</tr>
<tr>
<td>E021-05</td>
<td>OCTET_LENGTH function</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E021-06</td>
<td>SUBSTRING function</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E021-07</td>
<td>Character concatenation</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E021-08</td>
<td>UPPER and LOWER functions</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E021-09</td>
<td>TRIM function</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E021-10</td>
<td>Implicit casting among the character string types</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E021-11</td>
<td>POSITION function</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E021-12</td>
<td>Character comparison</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E031</td>
<td>Identifiers</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E031-01</td>
<td>Delimited identifiers</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E031-02</td>
<td>Lower case identifiers</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E031-03</td>
<td>Trailing underscore</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E051</td>
<td>Basic query specification</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E051-01</td>
<td>SELECT DISTINCT</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E051-02</td>
<td>GROUP BY clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E051-03</td>
<td>GROUP BY can contain columns not in SELECT list</td>
<td>YES</td>
<td></td>
</tr>
</tbody>
</table>
### Table K.1  SQL 2008 Optional Feature Compliance Details

<table>
<thead>
<tr>
<th>ID</th>
<th>Feature</th>
<th>Supported</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>E051-04</td>
<td>SELECT list items can be renamed</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E051-05</td>
<td>HAVING clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E051-06</td>
<td>Qualified * in SELECT list</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E051-07</td>
<td>Correlation names in the FROM clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E051-08</td>
<td>Rename columns in the FROM clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E061</td>
<td>Basic predicates and search conditions</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E061-01</td>
<td>Comparison predicate</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E061-02</td>
<td>BETWEEN predicate</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E061-03</td>
<td>IN predicate with list of values</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E061-04</td>
<td>LIKE predicate</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E061-05</td>
<td>LIKE predicate_ESCAPE clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E061-06</td>
<td>NULL predicate</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E061-07</td>
<td>Quantified comparison predicate</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E061-08</td>
<td>EXISTS predicate</td>
<td>YES</td>
<td>Not all uses work in Greenplum</td>
</tr>
<tr>
<td>E061-09</td>
<td>Subqueries in comparison predicate</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E061-11</td>
<td>Subqueries in IN predicate</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E061-12</td>
<td>Subqueries in quantified comparison predicate</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E061-13</td>
<td>Correlated subqueries</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>E061-14</td>
<td>Search condition</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E071</td>
<td>Basic query expressions</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E071-01</td>
<td>UNION DISTINCT table operator</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E071-02</td>
<td>UNION ALL table operator</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E071-03</td>
<td>EXCEPT DISTINCT table operator</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E071-05</td>
<td>Columns combined via table operators need not have exactly the same data type</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E071-06</td>
<td>Table operators in subqueries</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E081</td>
<td>Basic Privileges</td>
<td>NO</td>
<td>Partial sub-feature support</td>
</tr>
<tr>
<td>E081-01</td>
<td>SELECT privilege</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E081-02</td>
<td>DELETE privilege</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E081-03</td>
<td>INSERT privilege at the table level</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E081-04</td>
<td>UPDATE privilege at the table level</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E081-05</td>
<td>UPDATE privilege at the column level</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>E081-06</td>
<td>REFERENCES privilege at the table level</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>E081-07</td>
<td>REFERENCES privilege at the column level</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Feature</td>
<td>Supported</td>
<td>Comments</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------</td>
<td>-----------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>E081-08</td>
<td>WITH GRANT OPTION</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E081-09</td>
<td>USAGE privilege</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E081-10</td>
<td>EXECUTE privilege</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E091</td>
<td>Set Functions</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E091-01</td>
<td>AVG</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E091-02</td>
<td>COUNT</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E091-03</td>
<td>MAX</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E091-04</td>
<td>MIN</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E091-05</td>
<td>SUM</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E091-06</td>
<td>ALL quantifier</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E091-07</td>
<td>DISTINCT quantifier</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E101</td>
<td>Basic data manipulation</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E101-01</td>
<td>INSERT statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E101-03</td>
<td>Searched UPDATE statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E101-04</td>
<td>Searched DELETE statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E111</td>
<td>Single row SELECT statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E121</td>
<td>Basic cursor support</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E121-01</td>
<td>DECLARE CURSOR</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E121-02</td>
<td>ORDER BY columns need not be in select list</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E121-03</td>
<td>Value expressions in ORDER BY clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E121-04</td>
<td>OPEN statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E121-06</td>
<td>Positioned UPDATE statement</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>E121-07</td>
<td>Positioned DELETE statement</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>E121-08</td>
<td>CLOSE statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E121-10</td>
<td>FETCH statement implicit NEXT</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E121-17</td>
<td>WITH HOLD cursors</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E131</td>
<td>Null value support</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E141</td>
<td>Basic integrity constraints</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E141-01</td>
<td>NOT NULL constraints</td>
<td>YES</td>
<td>Must be the same as or a superset of the Greenplum distribution key</td>
</tr>
<tr>
<td>E141-02</td>
<td>UNIQUE constraints of NOT NULL columns</td>
<td>YES</td>
<td>Must be the same as or a superset of the Greenplum distribution key</td>
</tr>
<tr>
<td>E141-03</td>
<td>PRIMARY KEY constraints</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Feature</td>
<td>Supported</td>
<td>Comments</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------------------------------</td>
<td>-----------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>E141-04</td>
<td>Basic FOREIGN KEY constraint with the NO ACTION default for both referential delete action and referential update action</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>E141-06</td>
<td>CHECK constraints</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E141-07</td>
<td>Column defaults</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E141-08</td>
<td>NOT NULL inferred on PRIMARY KEY</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E141-10</td>
<td>Names in a foreign key can be specified in any order</td>
<td>YES</td>
<td>Foreign keys can be declared but are not enforced in Greenplum</td>
</tr>
<tr>
<td>E151</td>
<td>Transaction support</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E151-01</td>
<td>COMMIT statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E151-02</td>
<td>ROLLBACK statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E152</td>
<td>Basic SET TRANSACTION statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E152-01</td>
<td>ISOLATION LEVEL SERIALIZABLE clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E152-02</td>
<td>READ ONLY and READ WRITE clauses</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E153</td>
<td>Updatable queries with subqueries</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>E161</td>
<td>SQL comments using leading double minus</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E171</td>
<td>SQLSTATE support</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>E182</td>
<td>Module language</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F021</td>
<td>Basic information schema</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F021-01</td>
<td>COLUMNS view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F021-02</td>
<td>TABLES view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F021-03</td>
<td>VIEWS view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F021-04</td>
<td>TABLE_CONSTRAINTS view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F021-05</td>
<td>REFERENTIAL_CONSTRAINTS view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F021-06</td>
<td>CHECK_CONSTRAINTS view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F031</td>
<td>Basic schema manipulation</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F031-01</td>
<td>CREATE TABLE statement to create persistent base tables</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F031-02</td>
<td>CREATE VIEW statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F031-03</td>
<td>GRANT statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F031-04</td>
<td>ALTER TABLE statement: ADD COLUMN clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F031-13</td>
<td>DROP TABLE statement: RESTRICT clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F031-16</td>
<td>DROP VIEW statement: RESTRICT clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F031-19</td>
<td>REVOKE statement: RESTRICT clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F032</td>
<td>CASCADE drop behavior</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Feature</td>
<td>Supported</td>
<td>Comments</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------------------------------------------------</td>
<td>-----------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>F033</td>
<td>ALTER TABLE statement: DROP COLUMN clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F034</td>
<td>Extended REVOKE statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F034-01 REVOKE statement performed by other than the owner of a schema object</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F034-02 REVOKE statement: GRANT OPTION FOR clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F034-03 REVOKE statement to revoke a privilege that the grantee has WITH GRANT OPTION</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F041</td>
<td>Basic joined table</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F041-01 Inner join (but not necessarily the INNER keyword)</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F041-02 INNER keyword</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F041-03 LEFT OUTER JOIN</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F041-04 RIGHT OUTER JOIN</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F041-05 Outer joins can be nested</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F041-07 The inner table in a left or right outer join can also be used in an inner join</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F041-08 All comparison operators are supported (rather than just =)</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F051</td>
<td>Basic date and time</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F051-01 DATE data type (including support of DATE literal)</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F051-02 TIME data type (including support of TIME literal) with fractional seconds precision of at least 0</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F051-03 TIMESTAMP data type (including support of TIMESTAMP literal) with fractional seconds precision of at least 0 and 6</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F051-04 Comparison predicate on DATE, TIME, and TIMESTAMP data types</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F051-05 Explicit CAST between datetime types and character string types</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F051-06 CURRENT_DATE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F051-07 LOCALTIME</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F051-08 LOCALLTIME</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F052</td>
<td>Intervals and datetime arithmetic</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F053</td>
<td>OVERLAPS predicate</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F081</td>
<td>UNION and EXCEPT in views</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F111</td>
<td>Isolation levels other than SERIALIZABLE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F111-01 READ UNCOMMITTED isolation level</td>
<td>NO</td>
<td>Can be declared but is treated as a synonym for READ COMMITTED</td>
</tr>
</tbody>
</table>
### Table K.1 SQL 2008 Optional Feature Compliance Details

<table>
<thead>
<tr>
<th>ID</th>
<th>Feature</th>
<th>Supported</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>F111-02</td>
<td>READ COMMITTED isolation level</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F111-03</td>
<td>REPEATABLE READ isolation level</td>
<td>NO</td>
<td>Can be declared but is treated as a synonym for SERIALIZABLE</td>
</tr>
<tr>
<td>F121</td>
<td>Basic diagnostics management</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F122</td>
<td>Enhanced diagnostics management</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F123</td>
<td>All diagnostics</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F131-01</td>
<td>Grouped operations</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F131-02</td>
<td>WHERE, GROUP BY, and HAVING clauses supported in queries with grouped views</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F131-03</td>
<td>Multiple tables supported in queries with grouped views</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F131-04</td>
<td>Set functions supported in queries with grouped views</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F131-05</td>
<td>Subqueries with GROUP BY and HAVING clauses and grouped views</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F131-06</td>
<td>Single row SELECT with GROUP BY and HAVING clauses and grouped views</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F171</td>
<td>Multiple schemas per user</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F181</td>
<td>Multiple module support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F191</td>
<td>Referential delete actions</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F200</td>
<td>TRUNCATE TABLE statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F201</td>
<td>CAST function</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F202</td>
<td>TRUNCATE TABLE: identity column restart option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F221</td>
<td>Explicit defaults</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F222</td>
<td>INSERT statement: DEFAULT VALUES clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F231</td>
<td>Privilege tables</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F231-01</td>
<td>TABLE_PRIVILEGES view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F231-02</td>
<td>COLUMN_PRIVILEGES view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F231-03</td>
<td>USAGE_PRIVILEGES view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F251</td>
<td>Domain support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F261</td>
<td>CASE expression</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F261-01</td>
<td>Simple CASE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F261-02</td>
<td>Searched CASE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F261-03</td>
<td>NULLIF</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F261-04</td>
<td>COALESCE</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F262</td>
<td>Extended CASE expression</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Feature</td>
<td>Supported</td>
<td>Comments</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------------------------------------------------------</td>
<td>-----------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>F263</td>
<td>Comma-separated predicates in simple <code>CASE</code> expression</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F271</td>
<td>Compound character literals</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F281</td>
<td><code>LIKE</code> enhancements</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F291</td>
<td><code>UNIQUE</code> predicate</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F301</td>
<td><code>CORRESPONDING</code> in query expressions</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F302</td>
<td><code>INTERSECT</code> table operator</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F302-01</td>
<td><code>INTERSECT DISTINCT</code> table operator</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F302-02</td>
<td><code>INTERSECT ALL</code> table operator</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F304</td>
<td><code>EXCEPT ALL</code> table operator</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F311</td>
<td><code>Schema definition statement</code></td>
<td>YES</td>
<td>Partial sub-feature support</td>
</tr>
<tr>
<td>F311-01</td>
<td><code>CREATE SCHEMA</code></td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F311-02</td>
<td><code>CREATE TABLE</code> for persistent base tables</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F311-03</td>
<td><code>CREATE VIEW</code></td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F311-04</td>
<td><code>CREATE VIEW: WITH CHECK OPTION</code></td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F311-05</td>
<td><code>GRANT</code> statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F312</td>
<td><code>MERGE</code> statement</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F313</td>
<td>Enhanced <code>MERGE</code> statement</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F321</td>
<td>User authorization</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F341</td>
<td>Usage Tables</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F361</td>
<td>Subprogram support</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F381</td>
<td><code>Extended schema manipulation</code></td>
<td>YES</td>
<td>Some limitations on altering distribution key columns</td>
</tr>
<tr>
<td>F381-01</td>
<td><code>ALTER TABLE</code> statement: <code>ALTER COLUMN</code> clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F381-02</td>
<td><code>ALTER TABLE</code> statement: <code>ADD CONSTRAINT</code> clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F381-03</td>
<td><code>ALTER TABLE</code> statement: <code>DROP CONSTRAINT</code> clause</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F382</td>
<td>Alter column data type</td>
<td>YES</td>
<td>Some limitations on altering distribution key columns</td>
</tr>
<tr>
<td>F391</td>
<td>Long identifiers</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F392</td>
<td>Unicode escapes in identifiers</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F393</td>
<td>Unicode escapes in literals</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F394</td>
<td>Optional normal form specification</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F401</td>
<td><code>Extended joined table</code></td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F401-01</td>
<td><code>NATURAL JOIN</code></td>
<td>YES</td>
<td></td>
</tr>
</tbody>
</table>
## Table K.1  SQL 2008 Optional Feature Compliance Details

<table>
<thead>
<tr>
<th>ID</th>
<th>Feature</th>
<th>Supported</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>F401-02</td>
<td>FULL OUTER JOIN</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F401-04</td>
<td>CROSS JOIN</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F402</td>
<td>Named column joins for LOBs, arrays, and multisets</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F403</td>
<td>Partitioned joined tables</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F411</td>
<td>Time zone specification</td>
<td>YES</td>
<td>Differences regarding literal interpretation</td>
</tr>
<tr>
<td>F421</td>
<td>National character</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F431</td>
<td>Read-only scrollable cursors</td>
<td>YES</td>
<td>Forward scrolling only</td>
</tr>
<tr>
<td>01</td>
<td>FETCH with explicit NEXT</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>02</td>
<td>FETCH FIRST</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>03</td>
<td>FETCH LAST</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>04</td>
<td>FETCH PRIOR</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>05</td>
<td>FETCH ABSOLUTE</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>06</td>
<td>FETCH RELATIVE</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F441</td>
<td>Extended set function support</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F442</td>
<td>Mixed column references in set functions</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F451</td>
<td>Character set definition</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F461</td>
<td>Named character sets</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F471</td>
<td>Scalar subquery values</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F481</td>
<td>Expanded NULL predicate</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F491</td>
<td>Constraint management</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F501</td>
<td>Features and conformance views</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F501-01</td>
<td>SQL_FEATURES view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F501-02</td>
<td>SQL_SIZING view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F501-03</td>
<td>SQL_LANGUAGES view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F502</td>
<td>Enhanced documentation tables</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F502-01</td>
<td>SQL_SIZING_PROFILES view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F502-02</td>
<td>SQL_IMPLEMENTATION_INFO view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F502-03</td>
<td>SQL_PACKAGES view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F521</td>
<td>Assertions</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F531</td>
<td>Temporary tables</td>
<td>YES</td>
<td>Non-standard form</td>
</tr>
<tr>
<td>F555</td>
<td>Enhanced seconds precision</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F561</td>
<td>Full value expressions</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F571</td>
<td>Truth value tests</td>
<td>YES</td>
<td></td>
</tr>
</tbody>
</table>
## Table K.1 SQL 2008 Optional Feature Compliance Details

<table>
<thead>
<tr>
<th>ID</th>
<th>Feature</th>
<th>Supported</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>F591</td>
<td>Derived tables</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F611</td>
<td>Indicator data types</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F641</td>
<td>Row and table constructors</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F651</td>
<td>Catalog name qualifiers</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F661</td>
<td>Simple tables</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F671</td>
<td>Subqueries in CHECK</td>
<td>NO</td>
<td>Intentionally ommitted</td>
</tr>
<tr>
<td>F672</td>
<td>Retrospective check constraints</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F690</td>
<td>Collation support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F692</td>
<td>Enhanced collation support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F693</td>
<td>SQL-session and client module collations</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F695</td>
<td>Translation support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F696</td>
<td>Additional translation documentation</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F701</td>
<td>Referential update actions</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F711</td>
<td>ALTER domain</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F721</td>
<td>Deferrable constraints</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F731</td>
<td>INSERT column privileges</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F741</td>
<td>Referential MATCH types</td>
<td>NO</td>
<td>No partial match</td>
</tr>
<tr>
<td>F751</td>
<td>View CHECK enhancements</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F761</td>
<td>Session management</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F762</td>
<td>CURRENT_CATALOG</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F763</td>
<td>CURRENT_SCHEMA</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F771</td>
<td>Connection management</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F781</td>
<td>Self-referencing operations</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F791</td>
<td>Inensitive cursors</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F801</td>
<td>Full set function</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F812</td>
<td>Basic flagging</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F813</td>
<td>Extended flagging</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F831</td>
<td>Full cursor update</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F841</td>
<td>LIKE_REGEX predicate</td>
<td>NO</td>
<td>Non-standard syntax for regex</td>
</tr>
<tr>
<td>F842</td>
<td>OCCURENCES_REGEX function</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F843</td>
<td>POSITION_REGEX function</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F844</td>
<td>SUBSTRING_REGEX function</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F845</td>
<td>TRANSLATE_REGEX function</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F846</td>
<td>Octet support in regular expression operators</td>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>
## Table K.1 SQL 2008 Optional Feature Compliance Details

<table>
<thead>
<tr>
<th>ID</th>
<th>Feature</th>
<th>Supported</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>F847</td>
<td>Nonconstant regular expressions</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F850</td>
<td>Top-level ORDER BY clause in query expression</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>F851</td>
<td>Top-level ORDER BY clause in subqueries</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F852</td>
<td>Top-level ORDER BY clause in views</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F855</td>
<td>Nested ORDER BY clause in query expression</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F856</td>
<td>Nested FETCH FIRST clause in query expression</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F857</td>
<td>Top-level FETCH FIRST clause in query expression</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F858</td>
<td>FETCH FIRST clause in subqueries</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F859</td>
<td>Top-level FETCH FIRST clause in views</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F860</td>
<td>FETCH FIRST ROW count in FETCH FIRST clause</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F861</td>
<td>Top-level RESULT OFFSET clause in query expression</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F862</td>
<td>RESULT OFFSET clause in subqueries</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F863</td>
<td>Nested RESULT OFFSET clause in query expression</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F864</td>
<td>Top-level RESULT OFFSET clause in views</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>F865</td>
<td>OFFSET ROW count in RESULT OFFSET clause</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S011</td>
<td>Distinct data types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S023</td>
<td>Basic structured types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S024</td>
<td>Enhanced structured types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S025</td>
<td>Final structured types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S026</td>
<td>Self-referencing structured types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S027</td>
<td>Create method by specific method name</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S028</td>
<td>Permutable UDT options list</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S041</td>
<td>Basic reference types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S043</td>
<td>Enhanced reference types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S051</td>
<td>Create table of type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S071</td>
<td>SQL paths in function and type name resolution</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>S091</td>
<td>Basic array support</td>
<td>NO</td>
<td>Greenplum has arrays, but is not fully standards compliant</td>
</tr>
<tr>
<td>S091-01</td>
<td>Arrays of built-in data types</td>
<td>NO</td>
<td>Partially compliant</td>
</tr>
<tr>
<td>S091-02</td>
<td>Arrays of distinct types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S091-03</td>
<td>Array expressions</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S092</td>
<td>Arrays of user-defined types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S094</td>
<td>Arrays of reference types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Feature</td>
<td>Supported</td>
<td>Comments</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------------------------</td>
<td>-----------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>S095</td>
<td>Array constructors by query</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S096</td>
<td>Optional array bounds</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S097</td>
<td>Array element assignment</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S098</td>
<td>array_agg</td>
<td>NO</td>
<td>ORDER BY clause not supported</td>
</tr>
<tr>
<td>S111</td>
<td>only in query expressions</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>S151</td>
<td>Type predicate</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S161</td>
<td>Subtype treatment</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S162</td>
<td>Subtype treatment for references</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S201</td>
<td>SQL-invoked routines on arrays</td>
<td>NO</td>
<td>Functions can be passed</td>
</tr>
<tr>
<td>S202</td>
<td>SQL-invoked routines on multisets</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S211</td>
<td>User-defined cast functions</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>S231</td>
<td>Structured type locators</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S232</td>
<td>Array locators</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S233</td>
<td>Multiset locators</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S241</td>
<td>Transform functions</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S242</td>
<td>Alter transform statement</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S251</td>
<td>User-defined orderings</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S261</td>
<td>Specific type method</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S271</td>
<td>Basic multiset support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S272</td>
<td>Multisets of user-defined types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S274</td>
<td>Multisets of reference types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S275</td>
<td>Advanced multiset support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S281</td>
<td>Nested collection types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S291</td>
<td>Unique constraint on entire row</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S301</td>
<td>enhanced unnest</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S401</td>
<td>Distinct types based on array types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S402</td>
<td>Distinct types based on distinct types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S403</td>
<td>max_cardinality</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>S404</td>
<td>trim_array</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T011</td>
<td>Timestamp in Information Schema</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T021</td>
<td>binary and varbinary data types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T022</td>
<td>Advanced support for binary and varbinary data types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T023</td>
<td>Compound binary literal</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Feature</td>
<td>Supported</td>
<td>Comments</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td>T024</td>
<td>Spaces in binary literals</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T031</td>
<td><strong>BOOLEAN</strong> data type</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T041</td>
<td>Basic <strong>LOB</strong> data type support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T042</td>
<td>Extended <strong>LOB</strong> data type support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T043</td>
<td>Multiplier T</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T044</td>
<td>Multiplier P</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T051</td>
<td>Row types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T052</td>
<td><strong>MAX</strong> and <strong>MIN</strong> for row types</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T053</td>
<td>Explicit aliases for all-fields reference</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T061</td>
<td>UCS support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T071</td>
<td><strong>BIGINT</strong> data type</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T101</td>
<td>Enhanced nullability determination</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T111</td>
<td>Updatable joins, unions, and columns</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T121</td>
<td>WITH (excluding <strong>RECURSIVE</strong>) in query expression</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T122</td>
<td>WITH (excluding <strong>RECURSIVE</strong>) in subquery</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T131</td>
<td>Recursive query</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T132</td>
<td>Recursive query in subquery</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T141</td>
<td><strong>SIMILAR</strong> predicate</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T151</td>
<td><strong>DISTINCT</strong> predicate</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T152</td>
<td><strong>DISTINCT</strong> predicate with negation</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T171</td>
<td><strong>LIKE</strong> clause in table definition</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T172</td>
<td><strong>AS</strong> subquery clause in table definition</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T173</td>
<td>Extended <strong>LIKE</strong> clause in table definition</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T174</td>
<td>Identity columns</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T175</td>
<td>Generated columns</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T176</td>
<td>Sequence generator support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T177</td>
<td>Sequence generator support: simple restart option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T178</td>
<td>Identity columns: simple restart option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T191</td>
<td>Referential action <strong>RESTRICT</strong></td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T201</td>
<td>Comparable data types for referential constraints</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T211</td>
<td><strong>Basic trigger capability</strong></td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T211-01</td>
<td>Triggers activated on <strong>UPDATE</strong>, <strong>INSERT</strong>, or <strong>DELETE</strong> of one base table</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T211-02</td>
<td><strong>BEFORE</strong> triggers</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Feature</td>
<td>Supported</td>
<td>Comments</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------------------------------</td>
<td>-----------</td>
<td>------------------------</td>
</tr>
<tr>
<td>T211-03</td>
<td>AFTER triggers</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T211-04</td>
<td>FOR EACH ROW triggers</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T211-05</td>
<td>Ability to specify a search condition that must be true before the trigger is invoked</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T211-06</td>
<td>Support for run-time rules for the interaction of triggers and constraints</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T211-07</td>
<td>TRIGGER privilege</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T211-08</td>
<td>Multiple triggers for the same event are executed in the order in which they were created in the catalog</td>
<td>NO</td>
<td>Intentionally omitted</td>
</tr>
<tr>
<td>T212</td>
<td>Enhanced trigger capability</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T213</td>
<td>INSTEAD OF triggers</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T231</td>
<td>Sensitive cursors</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T241</td>
<td>START TRANSACTION statement</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T251</td>
<td>SET TRANSACTION statement: LOCAL option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T261</td>
<td>Chained transactions</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T271</td>
<td>Savepoints</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T272</td>
<td>Enhanced savepoint management</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T281</td>
<td>SELECT privilege with column granularity</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T285</td>
<td>Enhanced derived column names</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T301</td>
<td>Functional dependencies</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T312</td>
<td>OVERLAY function</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T321</td>
<td>Basic SQL-invoked routines</td>
<td>NO</td>
<td>Partial support</td>
</tr>
<tr>
<td>T321-01</td>
<td>User-defined functions with no overloading</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T321-02</td>
<td>User-defined stored procedures with no overloading</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T321-03</td>
<td>Function invocation</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T321-04</td>
<td>CALL statement</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T321-05</td>
<td>RETURN statement</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T321-06</td>
<td>ROUTINES view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T321-07</td>
<td>PARAMETERS view</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T322</td>
<td>Overloading of SQL-invoked functions and procedures</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T323</td>
<td>Explicit security for external routines</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T324</td>
<td>Explicit security for SQL routines</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T325</td>
<td>Qualified SQL parameter references</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T326</td>
<td>Table functions</td>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>
Table K.1  SQL 2008 Optional Feature Compliance Details

<table>
<thead>
<tr>
<th>ID</th>
<th>Feature</th>
<th>Supported</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>T331</td>
<td>Basic roles</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T332</td>
<td>Extended roles</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T351</td>
<td>Bracketed SQL comments (/<em>...</em>/ comments)</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T431</td>
<td>Extended grouping capabilities</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T432</td>
<td>Nested and concatenated GROUPING SETS</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T433</td>
<td>Multiargument GROUPING function</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T434</td>
<td>GROUP BY DISTINCT</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T441</td>
<td>ABS and MOD functions</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T461</td>
<td>Symmetric BETWEEN predicate</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T471</td>
<td>Result sets return value</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T491</td>
<td>LATERAL derived table</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T501</td>
<td>Enhanced EXISTS predicate</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T511</td>
<td>Transaction counts</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T541</td>
<td>Updatable table references</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T561</td>
<td>Holdable locators</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T571</td>
<td>Array-returning external SQL-invoked functions</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T572</td>
<td>Multiset-returning external SQL-invoked functions</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T581</td>
<td>Regular expression substring function</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T591</td>
<td>UNIQUE constraints of possibly null columns</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T601</td>
<td>Local cursor references</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T611</td>
<td>Elementary OLAP operations</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T612</td>
<td>Advanced OLAP operations</td>
<td>NO</td>
<td>Partially supported</td>
</tr>
<tr>
<td>T613</td>
<td>Sampling</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T614</td>
<td>NTILE function</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T615</td>
<td>LEAD and LAG functions</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T616</td>
<td>Null treatment option for LEAD and LAG functions</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T617</td>
<td>FIRST_VALUE and LAST_VALUE function</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T618</td>
<td>NTH_VALUE</td>
<td>NO</td>
<td>Function exists in Greenplum but not all options are supported</td>
</tr>
<tr>
<td>T621</td>
<td>Enhanced numeric functions</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>T631</td>
<td>N predicate with one list element</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T641</td>
<td>Multiple column assignment</td>
<td>NO</td>
<td>Some syntax variants supported</td>
</tr>
<tr>
<td>T651</td>
<td>SQL-schema statements in SQL routines</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T652</td>
<td>SQL-dynamic statements in SQL routines</td>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>
Table K.1  SQL 2008 Optional Feature Compliance Details

<table>
<thead>
<tr>
<th>ID</th>
<th>Feature</th>
<th>Supported</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>T653</td>
<td>SQL-schema statements in external routines</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T654</td>
<td>SQL-dynamic statements in external routines</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>T655</td>
<td>Cyclically dependent routines</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M001</td>
<td>Datalinks</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M002</td>
<td>Datalinks via SQL/CLI</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M003</td>
<td>Datalinks via Embedded SQL</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M004</td>
<td>Foreign data support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M005</td>
<td>Foreign schema support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M006</td>
<td>GetSQLString routine</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M007</td>
<td>TransmitRequest</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M009</td>
<td>GetOpts and GetStatistics routines</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M010</td>
<td>Foreign data wrapper support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M011</td>
<td>Datalinks via Ada</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M012</td>
<td>Datalinks via C</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M013</td>
<td>Datalinks via COBOL</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M014</td>
<td>Datalinks via Fortran</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M015</td>
<td>Datalinks via M</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M016</td>
<td>Datalinks via Pascal</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M017</td>
<td>Datalinks via PL/I</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M018</td>
<td>Foreign data wrapper interface routines in Ada</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M019</td>
<td>Foreign data wrapper interface routines in C</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M020</td>
<td>Foreign data wrapper interface routines in COBOL</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M021</td>
<td>Foreign data wrapper interface routines in Fortran</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M022</td>
<td>Foreign data wrapper interface routines in MUMPS</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M023</td>
<td>Foreign data wrapper interface routines in Pascal</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M024</td>
<td>Foreign data wrapper interface routines in PL/I</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M030</td>
<td>SQL-server foreign data support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>M031</td>
<td>Foreign data wrapper general routines</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X010</td>
<td>XML type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X011</td>
<td>Arrays of XML type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X012</td>
<td>Multisets of XML type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X013</td>
<td>Distinct types of XML type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X014</td>
<td>Attributes of XML type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X015</td>
<td>Fields of XML type</td>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>
#### Table K.1  SQL 2008 Optional Feature Compliance Details

<table>
<thead>
<tr>
<th>ID</th>
<th>Feature</th>
<th>Supported</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>X016</td>
<td>Persistent XML values</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X020</td>
<td>XMLConcat</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X025</td>
<td>XMLCast</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X030</td>
<td>XMLDocument</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X031</td>
<td>XMLElement</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X032</td>
<td>XMLForest</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X034</td>
<td>XMLAgg</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X035</td>
<td>XMLAgg: ORDER BY option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X036</td>
<td>XMLComment</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X037</td>
<td>XMLPI</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X038</td>
<td>XMLText</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X040</td>
<td>Basic table mapping</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X041</td>
<td>Basic table mapping: nulls absent</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X042</td>
<td>Basic table mapping: null as nil</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X043</td>
<td>Basic table mapping: table as forest</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X044</td>
<td>Basic table mapping: table as element</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X045</td>
<td>Basic table mapping: with target namespace</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X046</td>
<td>Basic table mapping: data mapping</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X047</td>
<td>Basic table mapping: metadata mapping</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X048</td>
<td>Basic table mapping: base64 encoding of binary strings</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X049</td>
<td>Basic table mapping: hex encoding of binary strings</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X051</td>
<td>Advanced table mapping: nulls absent</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X052</td>
<td>Advanced table mapping: null as nil</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X053</td>
<td>Advanced table mapping: table as forest</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X054</td>
<td>Advanced table mapping: table as element</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X055</td>
<td>Advanced table mapping: target namespace</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X056</td>
<td>Advanced table mapping: data mapping</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X057</td>
<td>Advanced table mapping: metadata mapping</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X058</td>
<td>Advanced table mapping: base64 encoding of binary strings</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X059</td>
<td>Advanced table mapping: hex encoding of binary strings</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X060</td>
<td>XMLParse: Character string input and CONTENT option</td>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>
### Table K.1 SQL 2008 Optional Feature Compliance Details

<table>
<thead>
<tr>
<th>ID</th>
<th>Feature</th>
<th>Supported</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>X061</td>
<td>XMLParse: Character string input and DOCUMENT option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X065</td>
<td>XMLParse: BLOB input and CONTENT option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X066</td>
<td>XMLParse: BLOB input and DOCUMENT option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X068</td>
<td>XMLSerialize: BOM</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X069</td>
<td>XMLSerialize: INDENT</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X070</td>
<td>XMLSerialize: Character string serialization and CONTENT option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X071</td>
<td>XMLSerialize: Character string serialization and DOCUMENT option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X072</td>
<td>XMLSerialize: Character string serialization</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X073</td>
<td>XMLSerialize: BLOB serialization and CONTENT option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X074</td>
<td>XMLSerialize: BLOB serialization and DOCUMENT option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X075</td>
<td>XMLSerialize: BLOB serialization</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X076</td>
<td>XMLSerialize: VERSION</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X077</td>
<td>XMLSerialize: explicit ENCODING option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X078</td>
<td>XMLSerialize: explicit XML declaration</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X080</td>
<td>Namespaces in XML publishing</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X081</td>
<td>Query-level XML namespace declarations</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X082</td>
<td>XML namespace declarations in DML</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X083</td>
<td>XML namespace declarations in DDL</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X084</td>
<td>XML namespace declarations in compound statements</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X085</td>
<td>Predefined namespace prefixes</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X086</td>
<td>XML namespace declarations in XMLTable</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X090</td>
<td>XML document predicate</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X091</td>
<td>XML content predicate</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X096</td>
<td>XMLExists</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X100</td>
<td>Host language support for XML: CONTENT option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X101</td>
<td>Host language support for XML: DOCUMENT option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X110</td>
<td>Host language support for XML: VARCHAR mapping</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X111</td>
<td>Host language support for XML: CLOB mapping</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X112</td>
<td>Host language support for XML: BLOB mapping</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X113</td>
<td>Host language support for XML: STRIP WHITESPACE option</td>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>
### Table K.1 SQL 2008 Optional Feature Compliance Details

<table>
<thead>
<tr>
<th>ID</th>
<th>Feature</th>
<th>Supported</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>X114</td>
<td>Host language support for XML: PRESERVE WHITESPACE option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X120</td>
<td>XML parameters in SQL routines</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X121</td>
<td>XML parameters in external routines</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X131</td>
<td>Query-level XMLBINARY clause</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X132</td>
<td>XMLBINARY clause in DML</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X133</td>
<td>XMLBINARY clause in DDL</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X134</td>
<td>XMLBINARY clause in compound statements</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X135</td>
<td>XMLBINARY clause in subqueries</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X141</td>
<td>IS VALID predicate: data-driven case</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X142</td>
<td>IS VALID predicate: ACCORDING TO clause</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X143</td>
<td>IS VALID predicate: ELEMENT clause</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X144</td>
<td>IS VALID predicate: schema location</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X145</td>
<td>IS VALID predicate outside check constraints</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X151</td>
<td>IS VALID predicate with DOCUMENT option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X152</td>
<td>IS VALID predicate with CONTENT option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X153</td>
<td>IS VALID predicate with SEQUENCE option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X155</td>
<td>IS VALID predicate: NAMESPACE without ELEMENT clause</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X157</td>
<td>IS VALID predicate: NO NAMESPACE with ELEMENT clause</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X160</td>
<td>Basic Information Schema for registered XML Schemas</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X161</td>
<td>Advanced Information Schema for registered XML Schemas</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X170</td>
<td>XML null handling options</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X171</td>
<td>NIL ON NO CONTENT option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X181</td>
<td>XML(DOCUMENT(UNTYPED)) type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X182</td>
<td>XML(DOCUMENT(ANY)) type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X190</td>
<td>XML(SEQUENCE) type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X191</td>
<td>XML(DOCUMENT(XMLSCHEMA)) type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X192</td>
<td>XML(CONTENT(XMLSCHEMA)) type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X200</td>
<td>XMLQuery</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X201</td>
<td>XMLQuery: RETURNING CONTENT</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X202</td>
<td>XMLQuery: RETURNING SEQUENCE</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X203</td>
<td>XMLQuery: passing a context item</td>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>
### Table K.1  SQL 2008 Optional Feature Compliance Details

<table>
<thead>
<tr>
<th>ID</th>
<th>Feature</th>
<th>Supported</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>X204</td>
<td>XMLQuery: initializing an XQuery variable</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X205</td>
<td>XMLQuery: EMPTY ON EMPTY option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X206</td>
<td>XMLQuery: NULL ON EMPTY option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X211</td>
<td>XML 1.1 support</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X221</td>
<td>XML passing mechanism BY VALUE</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X222</td>
<td>XML passing mechanism BY REF</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X231</td>
<td>XML(CONTENT(UNTYPED)) type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X232</td>
<td>XML(CONTENT(ANY)) type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X241</td>
<td>RETURNING CONTENT in XML publishing</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X242</td>
<td>RETURNING SEQUENCE in XML publishing</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X251</td>
<td>Persistent XML values of XML(DOCUMENT(UNTYPED)) type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X252</td>
<td>Persistent XML values of XML(DOCUMENT(ANY)) type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X253</td>
<td>Persistent XML values of XML(CONTENT(UNTYPED)) type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X254</td>
<td>Persistent XML values of XML(CONTENT(ANY)) type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X255</td>
<td>Persistent XML values of XML(SEQUENCE) type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X256</td>
<td>Persistent XML values of XML(DOCUMENT(XMLSCHEMA)) type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X257</td>
<td>Persistent XML values of XML(CONTENT(XMLSCHEMA)) type</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X260</td>
<td>XML type: ELEMENT clause</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X261</td>
<td>XML type: NAMESPACE without ELEMENT clause</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X263</td>
<td>XML type: NO NAMESPACE with ELEMENT clause</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X264</td>
<td>XML type: schema location</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X271</td>
<td>XMLValidate: data-driven case</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X272</td>
<td>XMLValidate: ACCORDING TO clause</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X273</td>
<td>XMLValidate: ELEMENT clause</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X274</td>
<td>XMLValidate: schema location</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X281</td>
<td>XMLValidate: with DOCUMENT option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X282</td>
<td>XMLValidate with CONTENT option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X283</td>
<td>XMLValidate with SEQUENCE option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X284</td>
<td>XMLValidate NAMESPACE without ELEMENT clause</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X286</td>
<td>XMLValidate: NO NAMESPACE with ELEMENT clause</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Feature</td>
<td>Supported</td>
<td>Comments</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td>X300</td>
<td>XMLTable</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X301</td>
<td>XMLTable: derived column list option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X302</td>
<td>XMLTable: ordinality column option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X303</td>
<td>XMLTable: column default option</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X304</td>
<td>XMLTable: passing a context item</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X305</td>
<td>XMLTable: initializing an XQuery variable</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>X400</td>
<td>Name and identifier mapping</td>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>
Glossary

A

append-only tables
An append-only (AO) table is a storage representation that allows only appending new rows to a table, but does not allow updating or deleting existing rows. This allows for more compact storage on disk because each row does not need to store the MVCC transaction visibility info. This saves 20 bytes per row. AO tables can also be compressed.

array
The set of physical devices (hosts, servers, network switches, etc.) used to house a Greenplum Database system.

B

bandwidth
Bandwidth is the maximum amount of information that can be transmitted along a channel, such as a network or I/O channel. This data transfer rate is usually measured in megabytes per second (MB/s).

C

catalog
See system catalog.

column-oriented table
Greenplum provides a choice of storage orientation models for a table: row or column. A column-oriented table stores its content on disk by column rather than by row. This storage model has performance advantages for certain types of queries. Only append-only tables can be column-oriented; heap tables are always row-oriented.

correlated subquery
A correlated subquery is a nested SELECT statement that refers to a column from an outer SELECT statement. For example:

```
SELECT * FROM product WHERE exists (SELECT * FROM sale WHERE qty>0 AND pn = product.pn);
```
D

data directory
The data directory is the file system location on disk where database data is stored. The master data directory contains the global system catalog only — no user data is stored on the master. The data directory on the segment instances has user data for that segment plus a local copy of the system catalog. The data directory contains several subdirectories, control files, and configuration files as well.

distributed
Certain database objects in Greenplum Database, such as tables and indexes, are distributed. They are divided into equal parts and spread out among the segment instances based on a hashing algorithm. To the end-user and client software, however, a distributed object appears as a conventional database object.

distribution key
In a Greenplum table that uses hash distribution, one or more columns are used as the distribution key, meaning those columns are used to divide the data among all of the segments. The distribution key should be the primary key of the table or a unique column or set of columns.

distribution policy
The distribution policy determines how to divide the rows of a table among the Greenplum segments. Greenplum Database provides two types of distribution policy: hash distribution and random distribution.

DDL
Data Definition Language. A subset of SQL commands used for defining the structure of a database.

DML
Database Manipulation Language. SQL commands that store, manipulate, and retrieve data from tables. INSERT, UPDATE, DELETE, and SELECT are DML commands.

G

gang
For each slice of the query plan there is at least one query executor worker process assigned. During query execution, each segment will have a number of processes working on the query in parallel. Related processes that are working on the same portion of the query plan on different segments are referred to as gangs.
Greenplum Database

Greenplum Database is the industry’s first massively parallel processing (MPP) database server based on open-source technology. It is explicitly designed to support business intelligence (BI) applications and large, multi-terabyte data warehouses. Greenplum Database is based on PostgreSQL.

Greenplum Database system

An associated set of segment instances and a master instance running on an array, which can be composed of one or more hosts.

Greenplum instance

The process that serves a database. An instance of Greenplum Database is comprised of a master instance and two or more segment instances, however users and administrators always connect to the database via the master instance.

GUC

GUC is a PostgreSQL acronym that stands for global user configuration, and is an obsolete term in Greenplum Database. However, the term GUC may still occasionally appear in Greenplum Database error messages and other materials. A GUC is equivalent to a global server configuration parameter. See also, postgresql.conf.

H

hash distribution

With hash distribution, one or more table columns is used as the distribution key for the table. The distribution key is used by a hashing algorithm to assign each row to a particular segment. Keys of the same value will always hash to the same segment.

heap tables

Whenever you create a table without specifying a storage structure, the default is a heap storage structure. In a heap structure, the table is an unordered collection of data that allows multiple copies or versions of a row. Heap tables have row-level versioning information and allow updates and deletes. See also append-only tables and multiversion concurrency control.

host

A host represents a physical machine or compute node in a Greenplum Database system. In Greenplum Database, one host is designated as the master. The other hosts in the system have one or more segments on them.
**I**

**interconnect**

The interconnect is the networking layer of Greenplum Database. When a user connects to a database and issues a query, processes are created on each of the **segments** to handle the work of that query. The interconnect refers to the inter-process communication between the segments and **master**, as well as the network infrastructure on which this communication relies. The interconnect typically uses a standard Gigabit Ethernet switching fabric.

**I/O**

Input/Output (I/O) refers to the transfer of data to and from a system or device using a communication channel.

**J**

**JDBC**

Java Database Connectivity is an application program interface (API) specification for connecting programs written in Java to data in a database management system (DBMS). The application program interface lets you encode access request statements in SQL that are then passed to the program that manages the database.

**M**

**master**

The master is the entry point to a Greenplum Database system. It is the database listener process (**postmaster**) that accepts client connections and dispatches the SQL commands issued by the users of the system.

The master is where the global **system catalog** resides. However, the master does not contain any user data. User data resides only on the **segments**. The master does the work of authenticating user connections, parsing and planning the incoming SQL commands, distributing the query plan to the segments for execution, coordinating the results returned by each of the segments, and presenting the final results to the user.

**master instance**

The database process that serves the Greenplum master. See **master**.

**mirror**

A mirror is a backup copy of a **segment** (or master) that is stored on a different host than the primary copy. Mirrors are useful for maintaining operations if a host in your Greenplum Database system fails. Mirroring is an optional feature of Greenplum Database. Mirror segments are evenly distributed among other hosts in the array. If a host that holds a primary segment fails, Greenplum Database will switch to the mirror or secondary host.
motion node

A motion node is a portion of a query execution plan that indicates data movement between the various database instances of Greenplum Database (segments and the master). Some operations, such as joins, require segments to send and receive tuples to one another in order to satisfy the operation. A motion node can also indicate data movement from the segments back up to the master.

MPP

Massive Parallel Processing.

multiversion concurrency control

Unlike traditional database systems which use locks for concurrency control, Greenplum Database (as does PostgreSQL) maintains data consistency by using a multiversion model (multiversion concurrency control or MVCC). This means that while querying a database, each transaction sees a snapshot of data which protects the transaction from viewing inconsistent data that could be caused by (other) concurrent updates on the same data rows. This provides transaction isolation for each database session.

MVCC, by eschewing explicit locking methodologies of traditional database systems, minimizes lock contention in order to allow for reasonable performance in multiuser environments. The main advantage to using the MVCC model of concurrency control rather than locking is that in MVCC locks acquired for querying (reading) data do not conflict with locks acquired for writing data, and so reading never blocks writing and writing never blocks reading.

MVCC

See multiversion concurrency control.

ODBC

Open Database Connectivity, a standard database access method that makes it possible to access any data from any client application, regardless of which database management system (DBMS) is handling the data. ODBC manages this by inserting a middle layer, called a database driver, between a client application and the DBMS. The purpose of this layer is to translate the application’s data queries into commands that the DBMS understands.
OLAP

Online Analytical Processing (OLAP) is a category of technologies for collecting, managing, processing and presenting multidimensional data for analysis and management. OLAP leverages existing data from a relational schema or data warehouse (data source) by placing key performance indicators (measures) into context (dimensions). As of release 3.1, OLAP functions are supported in Greenplum Database. In practice, OLAP functions allow application developers to compose analytic business queries more easily and more efficiently. For example, moving averages and moving sums can be calculated over various intervals; aggregations and ranks can be reset as selected column values change; and complex ratios can be expressed in simple terms.

P

partitioned

Partitioning is a way to logically divide the data in a table for better performance and easier maintenance. In Greenplum Database, partitioning is a procedure that creates multiple sub-tables (or child tables) from a single large table (or parent table). The primary purpose is to improve performance by scanning only the relevant data needed to satisfy a query. Note that partitioned tables are also distributed.

Perl DBI

Perl Database Interface (DBI) is an API for connecting programs written in Perl to database management systems (DBMS). Perl DBI (DataBase Interface) is the most common database interface for the Perl programming language.

PostgreSQL

PostgreSQL is a SQL compliant, open source relational database management system (RDBMS). Greenplum Database uses a modified version of PostgreSQL as its underlying database server. For more information on PostgreSQL go to http://www.postgresql.org.

postgresql.conf

The server configuration file that configures various aspects of the database server. This configuration file is located in the data directory of the database instance. In Greenplum Database, the master and each segment instance has its own postgresql.conf file.

postres process

The postres executable is the actual PostgreSQL server process that processes queries. The database listener postres process (also known as the postmaster) creates other postres subprocesses as needed to handle client connections.

postmaster

The postmaster server program starts the postres database server listener process that accepts client connections. In Greenplum Database, a postres database listener process runs on the Greenplum master instance and on each segment instance.
psql

This is the interactive terminal to PostgreSQL and Greenplum Database. You can use psql to access a database and issue SQL commands. For more information on psql, see “psql” on page 744.

Q

QD

See query dispatcher.

QE

See query executor.

query dispatcher

The query dispatcher (QD) is a process that is initiated when users connect to the master and issue SQL commands. This process represents a user session and is responsible for sending the query plan to the segments and coordinating the results it gets back. The query dispatcher process spawns one or more query executor processes to assist in the execution of SQL commands.

query executor

A query executor process (QE) is associated with a query dispatcher (QD) process and operates on its behalf. Query executor processes run on the segment instances and execute their slice of the query plan on a segment.

query plan

A query plan is the set of operations that Greenplum Database will perform to produce the answer to a given query. Each node or step in the plan represents a database operation such as a table scan, join, aggregation or sort. Plans are read and executed from bottom to top. Greenplum Database supports an additional plan node type called a motion node. See also slice.

R

rack

A type of shelving to which computer components can be attached vertically, one on top of the other. Components are normally screwed into front-mounted, tapped metal strips with holes which are spaced so as to accommodate the height of devices of various U-sizes. Racks usually have their height denominated in U-units.

RAID

Redundant Array of Independent (or Inexpensive) Disks. RAID is a system of using multiple hard drives for sharing or replicating data among the drives. The benefit of RAID is increased data integrity, fault-tolerance and/or performance. Multiple hard drives are grouped and seen by the OS as one logical hard drive.
**RAM**

Random Access Memory. The main memory of a computer system used for storing programs and data. RAM provides temporary read/write storage while hard disks offer semi-permanent storage.

**random distribution**

With random distribution, table rows are sent to the segments as they come in, cycling across the segments in a round-robin fashion. Rows with columns having the same values will not necessarily be located on the same segment. Although a random distribution ensures even data distribution, there are performance advantages to choosing a hash distribution policy whenever possible.

**S**

**segment**

A segment represents a portion of data in a Greenplum database. User-defined tables and their indexes are distributed across the available number of segment instances in the Greenplum Database system. Each segment instance contains a distinct portion of the user data. A primary segment instance and its mirror both store the same segment of data.

**segment instance**

The segment instance is the database server process (postmaster) that serves segments. Users do not connect to segment instances directly, but through the master.

**slice**

In order to achieve maximum parallelism during query execution, Greenplum divides the work of the query plan into slices. A slice is a portion of the plan that can be worked on independently at the segment level. A query plan is sliced wherever a motion node occurs in the plan, one slice on each side of the motion. Plans that do not require data movement (such as catalog lookups on the master) are known as single-slice plans.

**star schema**

A relational database schema often used in data warehousing. The star schema is organized around a central table (fact table) joined to a few smaller tables (dimension tables) using foreign key references. The fact table contains raw numeric items that represent relevant business facts (price, number of units sold, etc.).

**system catalog**

The system catalogs are the place where a relational database management system stores schema metadata, such as information about tables and columns, and internal bookkeeping information. The system catalog in Greenplum Database is the same as the PostgreSQL catalog with some additional tables to support the distributed nature of the Greenplum system and databases. In Greenplum Database, the master contains the global system catalog tables. The segments also maintain their own local copy of the system catalog.
TPC-H

The Transaction Processing Performance Council (TPC) is a third-party organization that provides database benchmark tools for the industry. TPC-H is their ad-hoc, decision support benchmark. This benchmark illustrates decision support systems that examine large volumes of data, execute queries with a high degree of complexity, and give answers to critical business questions. The TPC-H toolkit is used for Greenplum Database functional and performance testing.

tuple

A tuple is another name for a row or record in a relational database table.

W

WAL

Write-Ahead Logging (WAL) is a standard approach to transaction logging. WAL’s central concept is that changes to data files (where tables and indexes reside) are logged before they are written to permanent storage. Data pages do not need to be flushed to disk on every transaction commit. In the event of a crash, data changes not yet applied to the database can be recovered from the log. A major benefit of using WAL is a significantly reduced number of disk writes.
## Index

### Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>__gp_masterid</td>
<td>923</td>
</tr>
<tr>
<td>.pgpass</td>
<td>42</td>
</tr>
</tbody>
</table>

### A

- **ABORT**: 272
- **access and permissions**: 30
- **ACTIVE_STATEMENTS**: 53
- **add_missing_from**: 772
- **administration utilities**: 10
- **aggregate functions**: 112
  - user-defined: 339
- **ALTER AGGREGATE**: 273
- **ALTER CONVERSION**: 275
- **ALTER DATABASE**: 276
- **ALTER DOMAIN**: 278
- **ALTER EXTERNAL TABLE**: 280
- **ALTER FILESPACE**: 283
- **ALTER FUNCTION**: 284
- **ALTER GROUP**: 287
- **ALTER INDEX**: 288
- **ALTER LANGUAGE**: 290
- **ALTER OPERATOR**: 291
- **ALTER OPERATOR CLASS**: 292
- **ALTER RESOURCE QUEUE**: 293
- **ALTER ROLE**: 295
- **ALTER SCHEMA**: 298
- **ALTER SEQUENCE**: 299
- **ALTER TABLE**: 302
- **ALTER TABLESPACE**: 314
- **ALTER TRIGGER**: 315
- **ALTER TYPE**: 316
- **ALTER USER**: 317
- **ANALYZE**: 92, 210, 318
- **APIs**: 45
- **array**: 71
  - configuration parameters: 155
  - functions: 112
  - in Greenplum Database: 965
  - operators: 112
- **ARRAY_NAME**: 637
- **array-nulls**: 772
- **Ascential Datastage**: 46
- **attributes**: 31
  - role attributes: 31
- **attrnums**: 831
- **authentication parameters**: 148
- **authentication_timeout**: 772
- **automatic database backup**: 174

### B

- **backslash_quote**: 772
- **backup**: 170, 172
  - database: 170, 172
  - non-parallel: 171
  - parallel: 170
- **backup master**: 158
  - configuring: 161
- **BEGIN**: 320
- **bigint**: 822
- **bigserial**: 822
- **binary functions**: 112
- **binary_operators**: 112
- **bit**: 822
- **bit_string functions**: 112
- **bit_string_operators**: 112
- **bit_varying**: 822
- **bitmap**: 90
- **bitmap index**: 67
- **bitmap_indexes**: 90
- **block_size**: 772
- **bonjour_name**: 772
- **boolean**: 822
- **box**: 822
- **Business Objects**: 46
- **bytea**: 822

### C

- **cast, user-defined**: 343
- **catalog**: 965
- **char**: 822
- **character**: 822
- **character_varying**: 822
- **check constraints**: 69, 76
- **check_function_bodies**: 772
- **CHECK_POINT_SEGMENTS**: 639
- **checking for locks**: 221
- **CHECKPOINT**: 322
- **cidr**: 822
- **circle**: 823
- **classid**: 912, 913
- **client**: 40
  - connection parameters: 154
  - connections: 40
  - interfaces: 45
  - third-party: 46
  - troubleshooting: 47
- **client applications**: 40
- **client tools**: 707
  - clusterdb: 707
  - createdb: 709
  - createlang: 711
  - createuser: 713
  - dropdb: 716
  - droplang: 719
  - dropuser: 721
contention: 218, 221
conversions: 346
COPY: 331
  non-parallel data loading: 137
cpu_index_tuple_cost: 773
cpu_operator_cost: 773
cpu_tuple_cost: 773
CREATE AGGREGATE: 339
CREATE CAST: 343
CREATE CONVERSION: 346
CREATE DATABASE: 348
CREATE DOMAIN: 350
CREATE EXTERNAL TABLE: 129, 352
CREATE EXTERNAL WEB TABLE: 352
CREATE FUNCTION: 365
CREATE GROUP: 371
CREATE LANGUAGE: 376
CREATE OPERATOR: 380
CREATE OPERATOR CLASS: 385
CREATE RESOURCE QUEUE: 390
CREATE ROLE: 393
CREATE RULE: 398
CREATE SCHEMA: 401
CREATE SEQUENCE: 403
CREATE TABLE: 407
CREATE TABLE AS: 418
CREATE TABLESPACE: 422
CREATE TRIGGER: 424
CREATE TYPE: 427
CREATE USER: 434
CREATE VIEW: 435
createdb: 709
createlang: 711
createuser: 713
creating
  databases: 348
  filespaces: 63
  functions: 365
  indexes: 372
  resource queues: 53
  roles: 393
  rules: 398
  schemas: 401
  sequences: 403
  tables: 407
cron: 608, 613
cursor_tuple_fraction: 773
cursors
  closing: 323
  fetching data from: 481
  moving position: 497
  opening: 439
custom_variable_classes: 773
customer support: 223

D
data
distribution: 219
redundancy: 157
data directory: 966
data loading: 331, 352, 641
  error isolation mode: 133
  from external tables: 133
  parallel: 10
  performance tips: 138
  with COPY: 137
  with gpfdist: 10
data types
  bigint: 822
  bigserial: 822
  bit: 822
  bit varying: 822
  boolean: 822
  box: 822
  bytea: 822
  char: 822
  character: 822
  character varying: 822
  choosing: 68
  cidr: 822
  circle: 823
  date: 823
  decimal: 823
  double precision: 823
  float4: 823
  float8: 823
  formatting: 112
  inet: 823
  int, int4: 823
  int2: 823
  int8: 822
  integer: 823
  interval: 823
  lseg: 823
  macaddr: 823
  money: 823
  numeric: 823
  path: 823
  point: 823
  polygon: 823
  real: 823
  serial: 823
  serial4: 823
  serial8: 822
  smallint: 823
  text: 823
  time: 823
  timestamp: 823
  timestamptz: 823
  timetz: 823
  user-defined: 427
  varbit: 822
varchar: 822
DATA_DIRECTORY: 638
database
  access and permissions: 30
  administration: 61
  analyze: 210
  backup: 170, 172
  backup automatically: 171, 174
  client
    applications: 40
    connections: 36, 40
    creating: 348
    design, optimizing: 219
    log files: 210, 212
    maintenance: 76
    management: 61
    migration: 170
    reindex: 210
    restore: 170, 171, 175
    restore from archive host: 177
    schema: 65
    statistics: 218
    templates: 61
    vacuum: 210
Datastage: 46
date: 823
date functions: 112
date operators: 112
date range partitioning: 79
DateStyle: 773
db_user_namespace: 773
dbid: 828, 848
DDL: 966
deadlock_timeout: 774
DEALLOCATE: 438
debug_assertions: 774
debug_pretty_print: 774
debug_print_parse: 774
debug_print_plan: 774
debug_print_prelim_plan: 774
debug_print_rewritten: 774
debug_print_slice_table: 774
debugging: 223
decimal: 823
DECLARE: 439
default_statistics_target: 774
default_tablespace: 64, 774
default_transaction_isolation: 775
default_transaction_read_only: 775
DELETE: 442
direct dispatch: 779
disk
  failure: 217
  usage: 204, 217
dispatcher: 971
distributed: 966
  database system architecture: 12
  table skew: 205
  tables: 12, 68
  tables, altering: 75
tables, distribution policy: 70
tables, optimizing: 219
distribution key: 75, 205
distribution policy: 13
  hash: 13
  random: 13
DML: 966
  INSERT command: 96
  TRUNCATE command: 98
  UPDATE command: 97
document
  MapReduce: 799
domains: 350
double precision: 823
DROP AGGREGATE: 445
DROP CAST: 446
DROP CONVERSION: 447
DROP DATABASE: 448
DROP DOMAIN: 449
DROP EXTERNAL TABLE: 450
DROP FILESPACE: 452
DROP FUNCTION: 453
DROP GROUP: 455
DROP INDEX: 456
DROP LANGUAGE: 457
DROP OPERATOR: 458
DROP OPERATOR CLASS: 460
DROP OWNED: 462
DROP RESOURCE QUEUE: 464
DROP ROLE: 466
DROP RULE: 467
DROP SCHEMA: 468
DROP SEQUENCE: 469
DROP TABLE: 470
DROP TABLESPACE: 471
DROP TRIGGER: 472
DROP TYPE: 473
DROP USER: 474
DROP VIEW: 475
dropdb: 716
droplang: 719
dropuser: 721
dynamic_library_path: 775

E
ecpg: 723
effective_cache_size: 775
Email Notifications: 202
  Testing: 203
enable_bitmapscan: 775
enable_groupagg: 775
enable_hashagg: 775
enable_hashjoin: 776
enable_indexscan: 776
enable_mergejoin: 776
enable_nestloop: 776
enable_seqscan: 776
enable_sort: 776
enable_tidscan: 776
ENCODING: 639
encoding conversions: 346
END: 476
environment variables: 40, 819, 822
error isolation mode: 129, 133
  with COPY: 137
error reporting and logging parameters: 152
errors: 162
error messages: 222
escape_string_warning: 776
etc/hosts: 47
ETL
  with gpfdist: 129
EXECUTE: 477
executor: 971
EXPLAIN: 478
EXPLAIN ANALYZE: 218
explain_pretty_print: 776
external tables: 352
defining: 129, 133
  error isolation mode: 133
parameters: 155
extra_float_digits: 777

F
failover: 8
failures: 159
diagnosing: 162
  of master: 167
  of segments: 163
recovering from: 163
faults: 159
FETCH: 481
file server, gpfdist: 129
filespace
catalog tables: 65
creating: 63
dropping: 65
pg_system: 65
float4: 823
float8: 823
foreign keys: 70
fromCollapseLimit: 777
functions
aggregate: 112
array: 112
binary: 112
bit string: 112
data type formatting: 112
date: 112
geometric: 112
MapReduce: 799
mathematical: 111
network address: 112
sequence: 112
set returning: 112
string: 112
system administration: 113
system information: 113
time: 112
user defined: 111
user-defined: 365

G
gangs: 27
gather motion: 26
geometric functions: 112
geometric operators: 112
GiST indexes: 90, 374
global parameters: 145
  setting: 146
global system catalog: 7, 972
gp_adjust_selectivity_for_outerjoins: 777
gp_analyze_relative_error: 777
gp_autostats_mode: 777
gp_autostats_on_change_threshold: 778
gp_bloat_diag: 924
gpcachedsegworkers_threshold: 778
gp_command_count: 778
gp_configuration: 825, 828
gp_configuration_history: 828
gp_connections_per_thread: 778
gp_crondump: 171
gp_db_interfaces: 718
gp_debug_linger: 778
gp_disk_free: 937
gp_distributed_log: 829
gp_distributed_xacts: 830
gp_distribution_policy: 831
gp_dump: 170, 172, 176, 579
gp_email_from: 778
gp_email_smtp_password: 779
gp_email_smtp_server: 779
gp_email_smtp_userid: 779
gp_email_to: 779
gp_enable_adaptive_nestloop: 779
gp_enable_agg_distinct: 779
  gp_enable_agg_distinct_pruning: 779
  gp_enable_direct_dispatch: 779
  gp_enable_fast_sri: 780
  gp_enable_gpperfmon: 780
  gp_enable_multiphase_agg: 780
  gp_enable_predicate_propagation: 780
  gp_enable_preunique: 780
  gp_enable_sequential_window_plans: 781
  gp_enable_sort_distinct: 781
  gp_enable_sort_limit: 781
  gp_external_enable_exec: 781
  gp_external_enable_exec: 781
  gp_external_grant_privileges: 781
  gp_external_max_segs: 781
  gp_fastsequence: 832
  gp_fault_strategy: 833
  gp_fts_probe_threadcount: 781
  gp_fts_probe_timeout: 781
  gp_global_sequence: 834
  gp_gpperfmon_send_interval: 782
  gp_hashjoin_tuples_per_bucket: 782
  gp_id: 838
  gp_interconnect_hash_multiplier: 782
  gp_interconnect_queue_depth: 782
gp_interconnect_setup_timeout: 782
 gp_interconnect_type: 782
 gp_interfaces: 839
 gp_locks_on_relation: 925
 gp_locks_on_resqueue: 925
 gp_log_command_timings: 926
 gp_log_database: 927
 gp_log_format: 783
 gp_log_master_concise: 928
 gp_log_system: 928
 gp_max_csv_line_length: 783
 gp_max_databases: 783
 gp_max_filespaces: 783
 gp_max_local_distributed_cache: 783
 gp_max_packet_size: 783
 gp_max_tablespaces: 783
 gp_motion_cost_per_row: 783
 gp_num_contents_in_cluster: 783
 gp_param_setting("param"): 930
 gp_param_settings_seg_value_diffs: 930
 gp_persistent_database_node: 841
 gp_persistent_filespace_node: 842
 gp_persistent_relation_node: 843
 gp_persistent_tablespace_node: 844
 gp_pgdatabase_invalid: 930
 gp_reject_percent_threshold: 783
 gp_relation_node: 845
 gp_reraise_signal: 783
 gp_resq_activity: 931
 gp_resq_activity_by_queue: 932
 gp_resq_priority_statement: 932
 gp_resq_role: 932
 gp_resqueue_priority: 783
 gp_resqueue_priority_cpuscores_per_segment: 784
 gp_resqueue_priority_sweeper_interval: 784
 gpRestore: 175, 584
 gp_role: 784
 gp_roles_assigned: 933
 gp_safefswritesize: 784
 gp_san_configuration: 846
 gp_segment_configuration: 848
 gp_segment_connect_timeout: 784
 gp_segments_for_planner: 784
 gp_session_id: 785
 gp_set_proc_affinity: 785
 gp_set_read_only: 785
 gp_size_of_all_table_indexes: 934
 gp_size_of_database: 934
 gp_size_of_index: 934
 gp_size_of_partition_and_indexes_disk: 935
 gp_size_of_schema_disk: 935
 gp_size_of_table_and_indexes_disk: 935
 gp_size_of_table_and_indexes_encoding: 936
 gp_size_of_table_disk: 936
 gp_size_of_table_uncompressed: 936
 gp_skew_coefficients: 937
 gp_skew_idle_fractions: 938
 gp_statistics_pullup_from_child_partition: 785
 gp_statistics_use_fkeys: 785
 gp_stats_missing: 924
 gp_toolkit: 209
   external tables
   functions
     gp_param_setting("param"): 930
 supporting external tables
     __gp_local_id: 923
     __gp_log_master_ext: 923
     __gp_log_segment_ext: 923
 supporting functions
     __gp_param_local_setting(varchar): 923
     __gp_skew_coefficients(): 923
     __gp_skew_idle_fractions(): 923
 gp_bloat_diag(int, int, boolean): 923
 gp_bloat_expected_pages: 923
 gp_bloat_diag: 924
 gp_locks_on_relation: 925
 gp_locks_on_resqueue: 925
 gp_log_command_timings: 926
 gp_log_database: 927
 gp_log_master_concise: 928
 gp_log_system: 928
 gp_param_settings_seg_value_diffs: 930
 gp_pgdatabase_invalid: 930
 gp_resq_activity: 931
 gp_resq_activity_by_queue: 932
 gp_resq_priority_statement: 932
 gp_resq_role: 932
 gp_roles_assigned: 933
 gp_size_of_all_table_indexes: 934
 gp_size_of_database: 934
 gp_size_of_index: 934
 gp_size_of_partition_and_indexes_disk: 935
 gp_size_of_schema_disk: 935
 gp_size_of_table_and_indexes_disk: 935
 gp_size_of_table_and_indexes_encoding: 936
 gp_size_of_table_disk: 936
 gp_size_of_table_uncompressed: 936
 gp_skew_coefficients: 937
 gp_skew_idle_fractions: 938
 gp_stats_missing: 924
 gp_transaction_log: 830
 gp_version_at_initdb: 851
 gp_vmem_idle_resource_timeout: 785
gp_vmem_protect_limit: 786
gp_vmem_protect_segworker_cache_limit: 786
gp_workfile_compress_algorithm: 786
gpactivatestandby: 593
gpaddmirrors: 588
gpadmin user: 30, 36
gpcheck: 598
gpcheckperf: 221, 600
gpconfig: 604
gpcordova: 172, 608
gpdbrestore: 172, 176, 613
gpdelete: 616
gpexpand: 852, 835, 836
gpexpand: 621
gpexpand.expansion_progress: 852
gpexpand: 621
gpcrondump: 172, 608
gpdist: 625
gpfs: 628
GPHOME: 819
gpinitstandby: 631
gpins: 634
gpload: 641
gplogfilter: 652
gpmapreduce: 655
gpperfmon_port: 786
gprecover: 665
gpscp: 669
gpseginstall: 671
gpsetupsanfailover: 674
gpsizecalc: 204
gpskew: 205
gpsnmpd: 677
gpssh: 680
gpssh-exkeys: 682
gpstart: 212, 685
gpstate: 196, 221, 688
gpstop: 692
gpsyncmaster: 557
gpsys1: 695
GRANT: 485
Greenplum Database
array: 965
connecting to: 40
starting: 142
stopping: 143
groups
creating: 32
gzip compression: 71
hash distribution: 13
hash key: 12
heap tables: 71
high availability: 9
historical data: 78
host: 967
host failure: 217
I
idle processes: 778
idle sessions: 785
indexes: 90, 372
btree: 90
examining usage: 92
GIST: 90
managing: 93
operator classes: 385
types: 90
using with Greenplum: 88
inet: 823
Informatica: 46
inheritance: 76
initdb: 557
initialization parameters
ARRAY_NAME: 637
CHECK_POINT_SEGMENTS: 639
DATA_DIRECTORY: 638
ENCODING: 639
MACHINE_LIST_FILE: 637
MASTER_DIRECTORY: 638
MASTER_HOSTNAME: 638
MASTER_PORT: 638
MIRROR_DATA_DIRECTORY: 639
MIRROR_PORT_BASE: 639
MIRROR_REPLICATION_PORT_BASE: 639
PORT_BASE: 638
REPLICATION_PORT_BASE: 639
SEG_PREFIX: 638
TRUSTED_SHELL: 638
INSERT command: 96, 490
int: 823
int2: 823
int4: 823
int8: 822
integer: 823
integer_datetimes: 786
interconnect: 12
about: 7
redundancy: 9
interface
database: 40
interval: 823
IntervalStyle: 787
ipccheck: 557
J
JDBC: 40, 45
join order: 219

H
hardware issues: 217
join_collapse_limit: 787

K
key: 79, 80
krb_caseins_users: 787
krb_server_keyfile: 787
krb_srvname: 787

L
lc_collate: 787
lc_ctype: 787
lc_messages: 788
lc_monetary: 788
lc_numeric: 788
lc_time: 788
LD_LIBRARY_PATH: 819
list partitioning: 77
listen_addresses: 788
loader: 641
loading data: 129, 331, 352
error isolation mode: 133
from external tables: 133
parallel: 10
performance tips: 138
with COPY: 137
with gpfdist: 10, 129
loading partitioned tables: 82
local parameters: 145
setting: 146
local_preload_libraries: 788
localoid: 831
LOCK: 493
lock: 221
management parameters: 154
modes: 95
row-level: 218
table-level: 218
log files
error messages: 222
format of: 207
managing: 212
viewing: 207
log rotation: 212
log_autostats: 788
log_connections: 788
log_disconnections: 789
log_dispatch_stats: 789
log_duration: 789
log_errorverbosity: 789
log_executor_stats: 789
log_hostname: 789
log_min_duration_statement: 789
log_min_error_statement: 789
log_min_messages: 790
log_parser_stats: 790
log_planner_stats: 790
log_rotation_age: 790
log_rotation_size: 790
log_statement: 790
log_statement_stats: 790
log_timezone: 790
log_truncate_on_rotation: 791
logical operators: 111
logs
on segments: 163
parameters: 152
searching: 652
lseg: 823

M
macaddr: 823
MACHINE_LIST_FILE: 637
maintenance: 76, 78, 83, 93
maintenance_work_mem: 791
management scripts
gp_dump: 579
gp_reactivatestandby: 593
gpaddmirrors: 588
gpcheck: 598
gpcheckperf: 600
gpconfig: 604
gpcrondump: 608
gpdbrestore: 613
gpdeletelesystem: 616
gpdetective: 618
gpexpandsystem: 621
gpfldist: 625
gpfilespace: 628
gpinitstandby: 631
gpinitsystem: 634
gpload: 641
gpmapreduce: 655
gprecoverseg: 665
gpscp: 669
gpseinstall: 671
gpsetupsanfailover: 674
gpsnmpd: 677
gpssh: 680
gpssh-exkeys: 682
gpsstop: 685
gpstate: 688
gpsvstop: 692
gpsys1: 695
MapReduce
document format: 799
document schema: 801
interface: 655
specification: 799
master: 12
mirroring: 158
recovery: 167
master instance: 968
master parameters: 145
setting: 146
master port
changing the default: 148
<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>MASTER_DATA_DIRECTORY</td>
<td>819</td>
</tr>
<tr>
<td>MASTER_DIRECTORY</td>
<td>638</td>
</tr>
<tr>
<td>MASTER_HOSTNAME</td>
<td>638</td>
</tr>
<tr>
<td>MASTER_PORT</td>
<td>638</td>
</tr>
<tr>
<td>mathematical functions</td>
<td>111</td>
</tr>
<tr>
<td>mathematical operators</td>
<td>111</td>
</tr>
<tr>
<td>max_appendonly_tables</td>
<td>791</td>
</tr>
<tr>
<td>max_connections</td>
<td>47, 791</td>
</tr>
<tr>
<td>max_connections parameter</td>
<td>38</td>
</tr>
<tr>
<td>MAX_COST</td>
<td>53</td>
</tr>
<tr>
<td>max_files_per_process</td>
<td>791</td>
</tr>
<tr>
<td>max_fsm_pages</td>
<td>792</td>
</tr>
<tr>
<td>max_fsm_relations</td>
<td>792</td>
</tr>
<tr>
<td>max_function_args</td>
<td>792</td>
</tr>
<tr>
<td>max_identifier_length</td>
<td>792</td>
</tr>
<tr>
<td>max_index_keys</td>
<td>792</td>
</tr>
<tr>
<td>max_locks_per_transaction</td>
<td>792</td>
</tr>
<tr>
<td>max_prepared_transactions</td>
<td>792</td>
</tr>
<tr>
<td>max_resource_portals_per_transaction</td>
<td>792</td>
</tr>
<tr>
<td>max_resource_queues</td>
<td>792</td>
</tr>
<tr>
<td>max_stack_depth</td>
<td>793</td>
</tr>
<tr>
<td>max_work_mem</td>
<td>793</td>
</tr>
<tr>
<td>Microsoft SQL Server</td>
<td>46</td>
</tr>
<tr>
<td>Microstrategy</td>
<td>46</td>
</tr>
<tr>
<td>MIN_COST</td>
<td>54</td>
</tr>
<tr>
<td>mirror</td>
<td>968</td>
</tr>
<tr>
<td>MIRROR_DATA_DIRECTORY</td>
<td>639</td>
</tr>
<tr>
<td>MIRROR_PORT_BASE</td>
<td>639</td>
</tr>
<tr>
<td>MIRROR_REPLICATION_PORT_BASE</td>
<td>639</td>
</tr>
<tr>
<td>mirroring</td>
<td>8</td>
</tr>
<tr>
<td>of data</td>
<td>157</td>
</tr>
<tr>
<td>of master</td>
<td>158</td>
</tr>
<tr>
<td>of segments</td>
<td>157</td>
</tr>
<tr>
<td>setting up</td>
<td>160</td>
</tr>
<tr>
<td>viewing mirror configuration</td>
<td>204</td>
</tr>
<tr>
<td>money</td>
<td>823</td>
</tr>
<tr>
<td>monitor performance</td>
<td>11</td>
</tr>
<tr>
<td>monitoring</td>
<td>10, 196</td>
</tr>
<tr>
<td>disk space usage</td>
<td>204</td>
</tr>
<tr>
<td>skew</td>
<td>205</td>
</tr>
<tr>
<td>system state</td>
<td>196</td>
</tr>
<tr>
<td>motion</td>
<td>26</td>
</tr>
<tr>
<td>gather</td>
<td>26</td>
</tr>
<tr>
<td>redistribute</td>
<td>26</td>
</tr>
<tr>
<td>motion node</td>
<td>969</td>
</tr>
<tr>
<td>motion, defined</td>
<td>26</td>
</tr>
<tr>
<td>MOVE</td>
<td>497</td>
</tr>
<tr>
<td>MPP</td>
<td>969</td>
</tr>
<tr>
<td>Multiversion Concurrency Control</td>
<td>95</td>
</tr>
<tr>
<td>MVCC</td>
<td>95, 210</td>
</tr>
<tr>
<td>network</td>
<td></td>
</tr>
<tr>
<td>address functions</td>
<td>112</td>
</tr>
<tr>
<td>address operators</td>
<td>112</td>
</tr>
<tr>
<td>failure</td>
<td>217</td>
</tr>
<tr>
<td>infrastructure</td>
<td>7</td>
</tr>
<tr>
<td>protocols</td>
<td>7</td>
</tr>
<tr>
<td>network protocol</td>
<td></td>
</tr>
<tr>
<td>TCP</td>
<td>7</td>
</tr>
<tr>
<td>UDP</td>
<td>7</td>
</tr>
<tr>
<td>node, query plan</td>
<td>26</td>
</tr>
<tr>
<td>Notifications</td>
<td></td>
</tr>
<tr>
<td>Enabling Email</td>
<td>202</td>
</tr>
<tr>
<td>Testing Email</td>
<td>203</td>
</tr>
<tr>
<td>not-null constraints</td>
<td>69</td>
</tr>
<tr>
<td>numeric</td>
<td>823</td>
</tr>
<tr>
<td>numeric range partitioning</td>
<td>80</td>
</tr>
<tr>
<td>object privileges</td>
<td>33</td>
</tr>
<tr>
<td>objid</td>
<td>912, 913</td>
</tr>
<tr>
<td>ODBC</td>
<td>40, 45</td>
</tr>
<tr>
<td>operator classes</td>
<td>385</td>
</tr>
<tr>
<td>operators</td>
<td></td>
</tr>
<tr>
<td>array</td>
<td>112</td>
</tr>
<tr>
<td>binary</td>
<td>112</td>
</tr>
<tr>
<td>bit string</td>
<td>112</td>
</tr>
<tr>
<td>comparison</td>
<td>111</td>
</tr>
<tr>
<td>date</td>
<td>112</td>
</tr>
<tr>
<td>geometric</td>
<td>112</td>
</tr>
<tr>
<td>logical</td>
<td>111</td>
</tr>
<tr>
<td>mathematical</td>
<td>111</td>
</tr>
<tr>
<td>network address</td>
<td>112</td>
</tr>
<tr>
<td>sequence</td>
<td>112</td>
</tr>
<tr>
<td>string</td>
<td>112</td>
</tr>
<tr>
<td>time</td>
<td>112</td>
</tr>
<tr>
<td>operators, user-defined</td>
<td>380</td>
</tr>
<tr>
<td>parameters</td>
<td></td>
</tr>
<tr>
<td>append-only table</td>
<td>155</td>
</tr>
<tr>
<td>append-only tables</td>
<td>155</td>
</tr>
<tr>
<td>array configuration</td>
<td>155</td>
</tr>
<tr>
<td>categories</td>
<td>147</td>
</tr>
<tr>
<td>client connection</td>
<td>154</td>
</tr>
<tr>
<td>connection and authentication</td>
<td>148</td>
</tr>
<tr>
<td>connection parameters</td>
<td>40</td>
</tr>
<tr>
<td>error reporting and logging</td>
<td>152</td>
</tr>
<tr>
<td>external table</td>
<td>155</td>
</tr>
<tr>
<td>lock management</td>
<td>154</td>
</tr>
<tr>
<td>master, global and local</td>
<td>145</td>
</tr>
<tr>
<td>query tuning</td>
<td>150</td>
</tr>
<tr>
<td>runtime statistics collection</td>
<td>153</td>
</tr>
<tr>
<td>setting level</td>
<td>146</td>
</tr>
<tr>
<td>system resource and consumption</td>
<td>149</td>
</tr>
<tr>
<td>version compatibility</td>
<td>155</td>
</tr>
<tr>
<td>workload management</td>
<td>154</td>
</tr>
<tr>
<td>partition</td>
<td>79, 80</td>
</tr>
<tr>
<td>by date range</td>
<td>79</td>
</tr>
<tr>
<td>by numeric range</td>
<td>80</td>
</tr>
<tr>
<td>existing table</td>
<td>81</td>
</tr>
<tr>
<td>key</td>
<td>79, 80</td>
</tr>
<tr>
<td>list</td>
<td>77</td>
</tr>
<tr>
<td>range</td>
<td>77</td>
</tr>
<tr>
<td>strategy</td>
<td>78</td>
</tr>
</tbody>
</table>
strategy, verifying: 82
tables: 76
viewing: 83
viewing design: 82
partitioned: 970
password file: 42
password_encryption: 793
password, default superuser: 637
PATH: 819
path: 823
pattern matching: 112
performance
benchmarks: 215
defining: 214
factors: 214
hardware baseline: 215
issues and causes: 78, 217
troubleshooting: 221
performance metrics: 196
performance monitor: 11
architecture: 11
Perl DBI: 45
permissions: 485
database: 30
pg_appendonly: 858
pg_bitmapindex: 67
pg_catalog: 67
pg_conf: 725
pg_controldata: 557
pg_ctl: 557
pg_default: 55, 65
pg_dump: 728
pg_dumpall: 735
pg_filespace: 877
pg_filespace_entry: 878
pg_foreign_data_wrapper: 826
pg_foreign_server: 826
pg_foreign_table: 826
pg_global: 65
pg_hba.conf: 47
editing: 37
fields: 36
pg_max_external_files: 827
pg_partition_columns: 82, 83
pg_partition_templates: 82, 83
pg_partitions: 82, 83
pg_resetxlog: 557
pg_resqueue_attributes: 901
pg_resqueue_status: 902
pg_resqueuecapability: 903
pg_restore: 739
pg_stat_last_operation: 912
pg_stat_last_shoperation: 913
pg_system: 65
pg_user_mapping: 827
PGCLIENTENCODING: 820
PGDATABASE: 820
PGDATESTYLE: 820
PGHOST: 820
PGHOSTADDR: 820
pgjdbc: 45
pgodbc: 45
PGOPTIONS: 820
PGPASSFILE: 820
PGPASSWORD: 820
pg_perl: 45
PGPORT: 820
PGTZ: 820
PGUSER: 820
PL/Java: 376
configuration parameters: 793
PL/Per: 376
PL/Python: 376
PL/R: 377
pljava_classpath: 793
point: 823
polygon: 823
port: 794
master, changing the default: 148
PORT_BASE: 638
postgres: 557
PostgreSQL: 970
postmaster: 557, 970
PREPARE: 499
prepared statements: 499
primary key: 13, 69
restrictions on using: 15
privileges: 485
managing object privileges: 33
procedural languages: 376
pljava: 376
plperl: 376
plpy: 376
plr: 377
psql: 40, 744, 971
public key exchange: 682
pygresql: 45

Q
QD: 971
QE: 971
query
cost: 218
dispatcher: 27, 971
execution plan: 971
executor: 27, 971
monitoring active: 11
parsing: 25
planner: 218
status check: 222
troubleshooting: 222
tuning parameters: 150
worker processes example: 27
query plans: 26, 478
about: 26
example: 26
parallel: 25
query prioritization
about: 49
configuring: 52
setting on a resource queue: 54
queued statements: 56

R
rack: 971
RAID: 971
random distribution: 13
random_page_cost: 794
range partitioning: 77
read uncommitted: 99
read-committed: 98
real: 823
REASSIGN OWNED: 502
recovery: 159
  of master: 167
  of segments: 163
redistribute motion: 26
redundancy: 8, 157
referential integrity: 70
regex_flavor: 794
REINDEX: 503
reindex: 210
reindexdb: 765
RELEASE SAVEPOINT: 505
repeatable read: 99
REPLICATION_PORT_BASE: 639
RESET: 506
resource contention: 221
resource queues: 48, 390
  active statement limit: 53
  altering: 55
  assigning roles: 55
  configuring: 51
  creating: 53
  default resource queue: 55
  evaluated statements: 51
  ignoring small queries: 54
  limits: 49
  monitoring: 56
  priority: 49
  query cost limit: 53
  setting priority: 54
  superusers: 55
resource_cleanup_gangs_on_wait: 794
resource_select_only: 51, 794
restarting Greenplum Database: 143
restore database: 170, 171, 175
  non-parallel: 172
    with different configuration: 177
REVOKE: 507
rewrite rules: 398
roles: 30
  adding to a resource queue: 55
  creating: 31, 393
  creating group membership: 32
  list of attributes: 31
  resource queues: 48
ROLLBACK: 510
ROLLBACK TO SAVEPOINT: 511
rotating log files: 212
row comparisons: 112
row-level lock: 218
Rows: 97
runtime statistics collection parameters: 153

S
SAS: 46
SAVEPOINT: 513
schema: 401
  inheritance: 76
  managing: 65
  MapReduce YAML document schema: 801
  pg_catalog: 67
  star: 12, 972
  system-level: 67
scp, group session: 669
search_path: 794
SEG_PREFIX: 638
segment: 12, 972
  about: 7
  failures: 162
  mirroring: 157
  primary per host: 7
  recovery: 163
segment instance: 972
SELECT: 515
SELECT INTO: 531
seq_page_cost: 794
sequence functions: 112
sequence operators: 112
sequences: 403
  creating and using: 87
serial: 823
serial4: 823
serial8: 822
serializable: 98
server configuration: 770
server logs: 207
server programs: 557
  gpsyncmaster: 557
  initdb: 557
  ipcclean: 557
  pg_controldata: 557
  pg_ctl: 557
  pg_resetxlog: 557
  postgres: 557
  postmaster: 557
server_encoding: 795
server_version: 795
server_version_num: 795
sessions: 40
SET: 533
set returning functions: 112
SET ROLE: 535
SET SESSION AUTHORIZATION: 535
SET TRANSACTION: 539
shared_buffers: 795
shared_preload_libraries: 795
SHOW: 542
skew: 205
correcting skew: 75
slice: 26
slice, defined: 26
smallint: 823
SNMP
  Installing: 199
  Notifications: 202
  Using with Greenplum Database: 197
ssh
  group session: 680
  key exchange: 682
ssl: 795
  ssl_ciphers: 795
staactionname: 912, 913
standard_conforming_strings: 795
standby master
  configuring: 161
star schema: 972
START TRANSACTION: 543
starting Greenplum Database: 142
stasubtype: 912, 913
stasysid: 912, 913
statement_timeout: 795
statime: 912, 913
statistics: 318
database: 218
stats_queue_level: 796
stausename: 912, 913
stopping Greenplum Database: 143
stored procedures: 365
string functions: 112
string operators: 112
subquery expressions: 112
superuser: 30, 36
superuser password: 637
superuser_reserved_connections: 796
support, customer: 223
system catalog: 972
system catalogs
global: 7
gp_configuration: 825, 828
gp_configuration_history: 828
gp_db_interfaces: 718
gp_distributed_log: 829
gp_distributed_xacts: 830
gp_distribution_policy: 831
gp_fastsequence: 832
gp_fault_strategy: 833
gp_global_sequence: 834
gp_id: 838
gp_interfaces: 839
gp_persistent_database_node: 841
gp_persistent_filespace_node: 842
gp_persistent_relation_node: 843
gp_persistent_tablespace_node: 844
gp_relation_node: 845
gp_san_configuration: 846
gp_segment_configuration: 848
gp_transaction_log: 850
gp_version_at_initdb: 851
gpexpand.expansion_progress: 852
gpexpand.status: 835
gpexpand.status_detail: 836
pg_appendonly: 858
pg_filesapce: 877
pg_filespace_entry: 878
pg_foreign_data_wrapper: 826
pg_foreign_server: 826
pg_foreign_table: 826
pg_resqueue_attributes: 901
pg_resqueue_status: 902
pg_resqueuecapability: 903
pg_stat_last_operation: 912
pg_stat_last_operation: 912
pg_stat_last_shoperation: 913
pg_user_mapping: 827
system resource and consumption parameters: 149
system resources and performance: 214
system state: 196
system utilization metrics: 11, 196
system-levels schemas: 67
T
  tables
    append-only: 71
column-oriented: 71
  compressed append-only: 71
  creating: 407
  distributed: 68
  distributed: 75
distribution policy: 70
  heap type: 71
  loading partitioned: 82
  locking: 218, 493
  maintaining partitioned: 83
  partition existing: 81
  partition strategy: 78
  partitioning: 76
  row-oriented: 71
  storage model: 71
tablespace
catalog tables: 65
default: 64
dropping: 65
file system locations: 63
pg_default: 65
pg_global: 65
using: 64
  tablespaces: 422
TCP: 7
tcp_keepalives_count: 796
tcp_keepalives_idle: 796
tcp_keepalives_interval: 796
temp_buffers: 796
templates, database: 61
text: 823
time: 823
time functions: 112
time operators: 112
time with time zone: 823
timestamp: 823
timestamp with time zone: 823
timetz: 823
Time Zone: 796
timezone_abbreviations: 796
TPC-H: 973
track_activities: 797
track_counts: 797
transaction_isolation: 797
transaction_read_only: 797
transactions
committing: 330
isolation level: 98, 320
management: 218
rolling back: 510
savepoints: 513
starting: 320
working with: 98
transform_null_equals: 797
triggers: 424
troubleshooting: 618
client connections: 47
data loading errors: 133
gpfdist: 131
performance problems: 221
TRUNCATE: 98, 545
trusted hosts
setup: 682
TRUSTED_SHELL: 638
tuple: 973

U
UDP: 7
unique constraints: 69
unix_socket_directory: 797
unix_socket_group: 797
unix_socket_permissions: 797
UPDATE: 546
UPDATE command: 97
update_process_title: 797
utilities
administration: 10
clusterdb: 707
createdb: 709
createlang: 711
createuser: 713
dropdb: 716
droplang: 719
dropuser: 721
ecpg: 723
gp_dump: 579
gp_restore: 584
gpactivatestandby: 593
gpaddmirrors: 588
gpcheck: 598
gpcheckperf: 600
gpconfig: 604
gpcrondump: 608
gpdbrestore: 613
gpdeleteesystem: 616
gpdevelop: 618
gpexpandsystem: 621
gpfdist: 625
gpf filespace: 628
gpinitstandby: 631
gpinit system: 634
gpload: 641
gplogfilter: 652
gpmapreduce: 655
gprecoverseg: 665
gpscp: 669
gpseinstall: 671
gpsetupsanfailover: 674
gpsnmapped: 677
gpssh: 680
gpssh-exkeys: 682
gpstart: 685
gpstate: 688
gpstop: 692
gpsys1: 695
pg_config: 725
pg_dump: 728
pg_dumpall: 735
pg_dumpall: 735
pg_restore: 739
psql: 744
reindexdb: 765
vacuumdb: 767
utility mode: 143

V
VACUUM: 550
vacuum: 210
vacuum_cost_delay: 797
vacuum_cost_limit: 798
vacuum_cost_page_dir: 798
vacuum_cost_page_hit: 798
vacuum_cost_page_miss: 798
vacuum_freeze_min_age: 798
vacuumdb: 767
VALUES: 553
varbit: 822
varchar: 822
version compatibility parameters: 155
viewing
partition design: 82
partitions: 83
query plans: 478
views: 435
creating and managing: 93
W

waiting queries: 56
web tables: 352
  creating and using: 132, 133
  defining command-based: 132, 133
  defining URL-based: 133
work_mem: 798
worker processes (query): 27
workload and system performance: 214
workload management: 48, 154, 218, 390
  configuring: 51
  parameters: 154