



Advanced Curriculum Greenplum Training 4.0+

Education Services
zData Inc.

40 E. Main St. Suite 610, Newark, DE 19711

T: 302-566-5351 • F: 419-715-6572

w w w . z d a . c o m

Greenplum Advanced System, Database and SQL Course Description

Overview

zData's 4 day Advanced Greenplum training class offers a unique opportunity for students to master the Greenplum database and DCA appliance. During this interactive course, students will be walked through the detailed steps involved in creating an instance, databases and schemas using a variety of distribution, partitioning, and storage methods. We will explain the required and optimal configuration settings at both the database and operating system levels. Students will receive hands-on training in implementing advanced techniques used in the field for increased performance and stability. The training will cover a wide area of topics including the required configuration settings, enabling compression and append only as columnar storage. Students will get to learn and try the different data loading methods like gpload, gpfdist, COPY and INSERT using multiple data sources. The training will provide hands-on experience with the key Greenplum features like resource queues and advanced workload management.

DURATION: 3 Days

Audience

This course is designed for individuals that have previously attended General Greenplum training or are actively using Greenplum.

Prerequisite Knowledge/Skills

To maximize what you learn from this course it is best to know

- Greenplum Fundamentals
- Basic Unix Administration
- SQL Syntax

Course Objectives

When you successfully complete this course, participants should be able to:

- Initialize and validate a Greenplum install
- Practical Distribution and Partitioning
- Column oriented storage techniques
- Tuning Techniques for GP SQL
- Utilize the explain plan to tune SQL
- Implement a complicated multi-source resource management matrix

- Loading techniques
- Backup techniques
- Greenplum scripting via Python or Perl

Course Outline

When you successfully complete this course, participants should be able to:

- **Module 1: DCA Hardware Maintenance**
 - dca_setup
 - connectEMC - Phone Home
 - Virtual IP takeover and HA improvements
 - XML functions
- **Module 2: Loading**
 - COPY vs INSERT vs parallel load
 - External tables
 - Writeable external table
 - External Web Tables
 - EWT + psql example
 - Connecting to Hadoop
 - Options when pulling data from Oracle
 - Gploading over SSH/pipes
 - Reusable external tables & gpload
 - Remote site loading considerations
 - Micro batching techniques
- **Module 3: Advanced Partitioning**
 - Subpartitioning
 - Combining row and columnar storage using partitions
 - Understanding Append Only tables
 - How much faster is append only?
 - When should you use columnar data storage?
 - It's append only, so no UPDATES/DELETES.
 - Working with append only partitions
 - Replacing and exchanging partitioning
 - Rolling data out of the warehouse
 - Gzip compression levels.
 - Should you ever go beyond level 1?
- **Module 4:Table Maintenance**
 - Free space map
 - Maintaining the system catalog
 - VACUUM
 - Bloat (Understanding and Fixing)
 - ANALYZE
 - CLUSTER
 - Vacuum Full
 - ALTER TABLE ... (reorganize=TRUE);
- **Module 5: Advanced SQL**

- Understanding JOIN types
- Advanced Partitioning
- Windowing functions
- Data types
- **Module 6: Encoding**
 - LATIN1, LATIN9
 - What is that character?
 - Iconv
 - Client_encoding vs databases encoding
 - How encoding affects sorting
- **Module 7: Query Tuning**
 - Query Tuning
 - EXPLAIN
 - EXPLAIN ANALYZE
 - Nested Loops
 - Merge Joins
- **Module 8: Workload Management**
 - Priority Queues
 - 3.X vs 4.X
 - Command Center
- **Module 9: Tools**
 - GP Mapreduce
 - Madlib, PGCrypto, R, PostGIS
- **Module 10: Tips and Tricks**
 - 3.X vs 4.X
 - Memory Management
 - Failure Scenarios
 - Multi-Rack Setup, VIP
 - Monitoring SNMP MIBs
 - GUC – Global User Config
 - Defaults and TCP Keepalive