



Fundamentals 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 Fundamentals

Course Description

Overview

zData's 5 day Greenplum Fundamentals training course offers a unique opportunity for students to be introduced to the fundamental concepts of the Greenplum database and EMC DCA appliance. During this interactive course, students will be will review data warehouse essentials and how Greenplum's "Shared Nothing " architecture effects data warehouse functionality. Students will be walked through the basic steps involved in creating an instance. We will review basic data warehouse table structures and review access methods. Our training will cover postgresql essentials including basic SQL structures, user maintenance, Roles and privileges, data types and functions. We will teach basic Greenplum administrator functions such as tuning and backup/recovery. Training will provide hands-on experience with the key Greenplum features throughout the course.

DURATION: 5 Days

Audience

This course is designed for individuals who are just starting out with the Greenplum Database.

Prerequisite Knowledge/Skills

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

- Experience with Unix or Linux
- Experience with SQL or Postgresql

Course Objectives

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

- Understand Greenplum and DCA architectures
- Initialize a Greenplum database
- Set up users, roles and resource queues
- Understand basic data design and table design
- Understand basic postgresql and functions
- Basic database administration
- Load data
- Backup and restore data

Course Outline

- **Module 1: Greenplum Overview**
 - Greenplum Database Architecture and Components
 - EMC DCA Architecture and Components
 - Data Redundancy components
 - Monitoring components
- **Module 2: Distributed Data and query processing**
 - Distributed Table architecture
 - Parallel Query plans and execution
 - *Lab – Review Module 1 and 2*
- **Module 3: Hardware Setup**
 - Software only setup considerations
 - DCA considerations
 - Storage Considerations
 - OS parameters
 - Hardware verification and Tuning
 - *Lab – System Verification utilities*
- **Module 4: Installation and Initialization**
 - Installation overview
 - Environment settings
 - Database initialization
 - Software Upgrade concepts
 - *Lab – Database initialization*
- **Module 5: Postgresql PSQL**
 - Connecting to a database
 - Running SQL statements
 - PSQL Meta Commands
 - *Lab – Using PSQL*
- **Module 6: Database Parameters**
 - Master-only, Global and Local Parameters
 - Postgresql.conf configuration file
 - How to set parameters
 - How to view parameters
 - Pg_hba.conf configuration file
 - *Lab – Setting Configuration parameters on Greenplum*
- **Module 7: Data Definition Language**
 - Databases
 - Schemas
 - Tables
 - Data types
 - Constraints
 - Other Database Objects
 - *Lab – Databases and Schemas*
 - *Lab – Tables and Distribution Keys*

- *Lab – Views, Indexes and Sequences*
- **Module 8: Roles, Privileges and Resource Queues**
 - Overview
 - Database Users
 - Database Groups
 - Database privileges
 - Resource queues and workload management
 - *Lab – Users and Privileges*
 - *Lab – Resource queues*
- **Module 9: Working with Tables**
 - Table partitioning
 - How to partition a table
 - Append only (AO Tables)
 - *Lab – Setting up a partitioned table*
 - *Lab – Setting up an AO table*
- **Module 10: Data Loading**
 - External Tables
 - GPFDIST and GPLOAD
 - COPY
 - Data Loading Performance
 - *Lab – Data Loads*
- **Module 11: Postgresql PSQL basics**
 - Inserts, Updates, Deletes
 - Select
 - Grouping Data
 - Transactions
 - Data Locking overview
 - *Lab – PSQL data manipulation*
- **Module 12: Performance tuning**
 - Performance tuning considerations
 - Common Causes
 - Hardware issues
 - Resource capacity issues
 - Database statistics
 - Data Distribution
 - Database design
 - Data Locks
 - PSQL Tuning
 - Explain Plans
 - *Lab – Query Tuning*
- **Module 13: Database administration**
 - Stopping and Starting an Database
 - Monitoring system state
 - Checking for data skew
 - Checking for Disk space usage

- Log Files
- Vacuum
- Analyze
- Reindex
- *Lab – Database administration examples*
- **Module 14: Fault Tolerance, Redundancy and High Availability**
 - Primary segments
 - Mirror segments
 - Fault Detection and recovery
 - Standby Master
 - *Lab – Standby master recovery*
- **Module 15: Backup and Recovery**
 - Backing up data
 - Restoring data
 - Automating backups
 - Other utilities
 - *Lab – backup and restore of data*
- **Module 16: Database internals**
 - System Catalog Tables
 - Physical storage of Database
 - Database processes
- **Module 17: Wrap up and Review**
 - Question and Answer