



Fundamentals Curriculum

Pivotal Hadoop 2.1

HAWQ

Education Services
zData Inc.

Pivotal Hadoop & HAWQ Fundamentals

Course Description

Overview

zData's 5 day HAWQ and Pivotal Hadoop [PHD] Fundamentals training course offers a unique opportunity for students to be introduced to the fundamental concepts of the Pivotal Hadoop architecture and HAWQ data query appliance.

During this interactive course, students will review data warehouse essentials, use various Hadoop technologies, and experience how data warehouse functionality is augmented with Pivotal Hadoop's architecture with HAWQ. Students will be walked through the basic steps involved in creating a Pivotal Hadoop cluster instance and will review basic data warehouse table structure, storage & distribution strategies, access methods, and performance tuning, as they relate to HAWQ and PivotalHD.

Our training will also cover Pivotal Extension and Greenplum Extension Framework essentials, along with HAWQ SQL structures, user maintenance, roles & privileges, data types, and user defined functions in a Hadoop ecosystem.

We will teach Hadoop administrator functions such as namenode and master backup failover and recovery, datanode failure and recovery, and HAWQ segment expansion.

Training will provide hands-on experience with the key HAWQ and Pivotal Hadoop features throughout the course.

DURATION: 5 Days

Audience

This course is designed for individuals who are familiar with data warehouse and Greenplum database architecture.

Prerequisite Knowledge/Skills

Trademarks used herein are the property of their respective owners.

© Copyright 2014 zData Corporation. All rights reserved. Published in the USA. 1/12

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

- Experience with Unix or Linux
- Experience with SQL or Postgresql
- Experience with Greenplum database parallel architecture

Course Objectives

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

- Understand HAWQ and the Pivotal Hadoop architecture
- Initialize a Pivotal Hadoop cluster and install HAWQ
- Set up users, roles, and resource queues
- Understand basic data, cluster, storage, and external table design
- Understand basic HAWQ queries and user defined functions
- Implement basic Pivotal Hadoop cluster administration
- Load and query data
- Backup and restore database data and cluster nodes

Course Outline

- **Module 1: Hadoop Fundamentals**
 - Understanding Big Data and Hadoop
 - The Internals of YARN, Mapreduce, and HDFS
 - Fundamentals of Hadoop, ETL, and processing
 - Architecture and commands
- **Module 2: Pivotal Hadoop**
 - Introduction to Pivotal Hadoop
 - Pivotal Hadoop architecture
 - Pivotal Hadoop components
- **Module 3: HAWQ**
 - Introduction to HAWQ
 - HAWQ Architecture and internals
 - Features and benefits

- **Module 4: Installing HAWQ**
 - HAWQ install readiness
 - Install and configure HAWQ on Pivotal Hadoop

- **Module 5: Data Processing**
 - Options for loading data; Sqoop, Flume, SpringXD
 - Joining, organizing, and analyzing data sets
 - Transforming data
 - Choosing the right transformation tool
 - Developing and scheduling workflows
 - Writing custom user defined functions

- **Module 6: HAWQ Data Storage**
 - Storage & I/O
 - Storage options & comparison
 - Data distribution
 - Data partitioning
 - HAWQ segments

- **Module 7: Querying HAWQ**
 - Databases
 - Schemas
 - Tables
 - Data types
 - Constraints
 - Other Database Objects

- **Module 8: Unloading Data**
 - COPY
 - External tables
 - PXF and GPXF

- **Module 9: PXF and GPXF**
 - How to use PXF
 - How to use GPXF

- **Module 10: Debugging and Maintenance**
 - Pivotal Command Center
 - Ambari
 - Hardware infrastructure considerations
 - Best practices for preparing HAWQ for production
 - Best practices for maintaining HAWQ in production
 - Operations and performance tuning
 - HAWQ segment expansion
 - Backup and recovery

