ADMINISTRATOR TRAINING: CDP PRIVATE CLOUD BASE

Take your knowledge to the next level

"Administrator training gave me an excellent jumpstart on acquiring the knowledge I needed to address my customers' big data and cloud challenges."

Canonical

Cloudera's four-day administrator training course for CDP Private Cloud Base provides participants with a comprehensive understanding of all the steps necessary to operate and maintain on-premises clusters using Cloudera Manager. From installation and configuration through load balancing and tuning, this Cloudera training course is the best preparation for the real-world challenges faced by administrators who run CDP Private Cloud Base.

What You'll Learn

Through instructor-led discussion and interactive, hands-on exercises, you will learn how to:

- Install Cloudera Manager
- Use Cloudera Manager to install a CDP Private Cloud Base cluster
- Configure and monitor the cluster using Cloudera Manager
- Understand, evaluate, and select the most appropriate data storage option
- Optimize cluster performance
- Perform routine cluster maintenance tasks
- Detect, troubleshoot, and repair problems with the cluster

What to Expect

This course is best suited to systems administrators who have at least basic Linux experience. Prior knowledge of CDP, nor earlier platforms such as Cloudera's CDH or Hortonworks HDP, is not required.

Course Details (page 1 of 2)

Cloudera Data Platform

- Industry Trends for Big Data
- The Challenge to Become Data-Driven
- The Enterprise Data Cloud
- CDP Overview
- CDP Form Factors

CDP Private Cloud Base Installation

- Installation Overview
- Cloudera Manager Installation
- CDP Runtime Overview
- Cloudera Manager Introduction

Cluster Configuration

- Overview
- Configuration Settings
- Modifying Service Configurations
- Configuration Files
- Managing Role Instances
- Adding New Services
- Adding and Removing Hosts

Data Storage

- Overview
- HDFS Topology and Roles
- HDFS Performance and Fault Tolerance
- HDFS and Hadoop Security Overview
- Working with HDFS
- HBase Overview
- Kudu Overview
- Cloud Storage Overview

Data Ingest

- Data Ingest Overview
- File Formats
- Ingesting Data using File Transfer or REST Interfaces
- Importing Data from Relational Databases with Apache Sqoop
- Ingesting Data Using NiFi
- Best Practices for Importing Data

Data Flow

- Overview of Cloudera Flow Management and NiFi
- NiFi Architecture
- Cloudera Edge Flow Management and MiNiFi
- Controller Services
- Apache Kafka Overview
- Apache Kafka Cluster Architecture
- Apache Kafka Command Line Tools

Data Access and Discovery

- Apache Hive
- Apache Impala
- Apache Impala Tuning
- Search Overview
- Hue Overview
- Managing and Configuring Hue
- Hue Authentication and Authorization
- CDSW Overview

Data Compute

- YARN Overview
- Running Applications on YARN

- Viewing YARN Applications
- YARN Application Logs
- MapReduce Applications
- YARN Memory and CPU Settings
- Tez Overview
- Hive on Tez
- ACID for Hive
- Spark Overview
- How Spark Applications Run on YARN
- Monitoring Spark Applications
- Phoenix Overview

Managing Resources

- Configuring cgroups with CPU Scheduling
- The Capacity Scheduler
- Managing Queues
- Impala Query Scheduling

Planning Your Cluster

- General Planning Considerations
- Choosing the Right Hardware
- Network Considerations
- CDP Private Cloud Considerations
- Configuring Nodes

Advanced Cluster Configuration

- Configuring Service Ports
- Tuning HDFS and MapReduce
- Managing Cluster Growth
- Erasure Coding
- Enabling HDFS High Availability

Course Details (page 2 of 2)

Cluster Maintenance
- Checking HDFS Status
- Copying Data Between Clusters
- Rebalancing Data in HDFS
- HDFS Directory Snapshots
- Host Maintenance
- Upgrading a Cluster

Cluster Monitoring
- Cloudera Manager Monitoring Features
- Health Tests
- Events and Alerts
- Charts and Reports
- Monitoring Recommendations

Cluster Troubleshooting
- Overview
- Troubleshooting Tools
- Misconfiguration Examples

Security
- Data Governance with SDX
- Hadoop Security Concepts
- Hadoop Authentication Using Kerberos
- Hadoop Authorization
- Hadoop Encryption
- Securing a Hadoop Cluster
- Apache Ranger
- Apache Atlas
- Backup and Recovery

Private Cloud / Public Cloud
- CDP Overview
- Private Cloud Capabilities
- Public Cloud Capabilities
- What is Kubernetes?
- WXM Overview
- Auto-scaling