HDP® HADOOP Administration Foundations

Course number: ADM-221, 4 days

What you will learn

This course is intended for systems administrators who will be responsible for the design, installation, configuration, and management of the Hortonworks Data Platform (HDP). The course provides in-depth knowledge and experience in using Apache Ambari as the operational management platform for HDP. This course presumes no prior knowledge or experience with Hadoop.

Prerequisites

Students must have experience working in a Linux environment with standard Linux system commands. Students should be able to read and execute basic Linux shell scripts. Basic knowledge of SQL statements is recommended, but not a requirement. In addition, it is recommended for students to have some operational experience in data center practices, such as change management, release management, incident management, and problem management. It is also strongly recommended that you complete HDP-123 Hortonworks Hadoop Essentials before taking this course.

Target Audience

Linux administrators and system operators responsible for installing, configuring and managing an HDP cluster.

Format

50% Lecture
50% Hands-on Labs

Agenda Summary

Day 1:
Introduction to Big Data, Hadoop and the Hortonworks Data Platform

Day 2:
Managing HDFS Storage, Rack Awareness, HDFS Snapshots and HDFS Centralized Cache

Day 3:
Introduction to YARN

Day 4:
High Availability with HDP, Deploying HDP with Blueprints, and the HDP Upgrade Process

REGISTRATION AND INFORMATION
education@ecs.com.sg
www.ecs.com.sg/training

Page 1 of 3

HDP® HADOOP Administration Foundations
Course number: ADM-221, 4 days

DAY 1 OBJECTIVES

- Describe Apache Hadoop
- Summarize the Purpose of the Hortonworks Data Platform Software Frameworks
- List Hadoop Cluster Management Choices
- Describe Apache Ambari
- Identify Hadoop Cluster Deployment Options
- Plan for a Hadoop Cluster Deployment
- Perform an Interactive HDP Installation using Apache Ambari
- Install Apache Ambari
- Describe the Differences Between Hadoop Users, Hadoop Service Owners, and Apache Ambari Users
- Manage Users, Groups and Permissions
- Identify Hadoop Configuration Files
- Summarize Operations of the Web UI Tool
- Manage Hadoop Service Configuration Properties Using the Apache Ambari Web UI
- Describe the Hadoop Distributed File System (HDFS)
- Perform HDFS Shell Operations
- Use WebHDFS
- Protect Data Using HDFS Access Control Lists (ACLs)

DAY 1 LABS

- Setting Up the Environment
- Installing HDP
- Managing Ambari Users and Groups
- Managing Hadoop Services
- Using HDFS Storage
- Using WebHDFS
- Using HDFS Access Control Lists

DAY 2 OBJECTIVES

- Describe HDFS Architecture and Operation
- Manage HDFS using Ambari Web, NameNode and DataNode UIs
- Manage HDFS using Command-line Tools
- Summarize the Purpose and Benefits of Rack Awareness
- Configure Rack Awareness
- Summarize Hadoop Backup Considerations
- Enable and Manage HDFS Snapshots
- Copy Data Using DistCP
- Use Snapshots and DistCP Together
- Identify the Purpose and Operation of Heterogeneous HDFS Storage
- Summarize the Purpose and Operation of HDFS Centralized Caching
- Configure HDFS Centralized Cache
- Define and Manage Cache Pools and Cache Directives
- Identify HDFS NFS Gateway Use Cases
- Recall HDFS NFS Gateway Architecture and Operation
- Install and Configure an HDFS NFS Gateway
- Configure an HDFS NFS Gateway Client

DAY 2 LABS

- Managing HDFS Storage
- Managing HDFS Quotas
- Configuring Rack Awareness
- Managing HDFS Snapshots
- Using DistCP
- Configuring HDFS Storage Policies
- Configuring HDFS Centralized Cache
- Configuring an NFS Gateway

HDP® HADOOP Administration Foundations
Course number: ADM-221, 4 days

DAY 3 OBJECTIVES

- Describe YARN Resource Management
- Summarize YARN Architecture and Operation
- Identify and Use YARN Management Options
- Summarize YARN Response to Component Failure
- Understand the Basics of Running Simple YARN Applications
- Summarize the Purpose and Operation of the YARN Capacity Scheduler
- Configure and Manage YARN Queues
- Control Access to YARN Queues
- Summarize the Purpose and Operation of YARN Node Labels
- Describe the Process used to Create Node Labels
- Describe the Process Used to Add, Modify and Remove Node Labels
- Configure Queues to Access Node Label Resources
- Run Test Jobs to Confirm Node Label Behavior

DAY 3 LABS

- Managing YARN Using Ambari
- Managing YARN Using CLI
- Running Sample YARN Applications
- Setting Up for Capacity Scheduler
- Managing YARN Containers and Queues
- Managing YARN ACLs and User Limits
- Working with YARN Node Labels

DAY 4 OBJECTIVES

- Summarize the Purpose of NameNode HA
- Configure NameNode HA Using Ambari
- Summarize the Purpose of ResourceManager HA
- Configure ResourceManager HA using Apache Ambari
- Identify Reasons to Add, Replace and Delete Worker Nodes
- Demonstrate How to Add a Worker Node
- Configure and Run the HDFS Balancer
- Decommission and Re-commission a Worker Node
- Describe the Process of Moving a Master Component
- Summarize the Purpose and Operation of Apache Ambari Metrics
- Describe the Features and Benefits of the Apache Ambari Dashboard
- Summarize the Purpose and Benefits of Apache Ambari Blueprints
- Recall the Process Used to Deploy a Cluster Using Ambari Blueprints
- Recall the Definition of an HDP Stack and Interpret its Version Number
- View the Current Stack and Identify Compatible Apache Ambari Software Versions
- Recall the Types of Methods and Upgrades Available in HDP
- Describe the Upgrade Process, Restrictions and Pre-upgrade Checklist
- Perform an Upgrade Using the Apache Ambari Web UI

DAY 4 LABS

- Configuring NameNode HA
- Configuring Resource Manager HA
- Adding, Decommissioning and Re-commissioning a Worker Node
- Configuring Ambari Alerts
- Deploying an HDP Cluster Using Ambari Blueprints
- Performing an HDP Upgrade – Express