

Hortonworks HDP Administration Fast Track

Course Summary

Description

This 5-day training course is designed for primarily for systems administrators and platform architects who need to understand HDP cluster capabilities, and manage HDP clusters. Topics include: Understanding HDF capabilities, Apache Hadoop, Apache YARN, HDFS, and other Hadoop ecosystem components. Students will understand how to administer, manage, and monitor HDP clusters.

Topics

- An Introduction to Apache Hadoop and HDFS
- Working with HDFS
- Working with Apache YARN
- High Availability, Backups and Configuring Centralized Cache
- Performing a Rolling Upgrade

Audience

This course is for students who range from having an understanding of server software concepts, to system administrators and platform architects who plan on administering HDP clusters.

Prerequisites

Before taking this course, students should be familiar with server or platform software concepts and have a basic understanding of system administration.

Duration

Five days

Hortonworks HDP Administration Fast Track

Course Outline

I. *An Introduction to Apache Hadoop and HDFS*

- A. Big Data, Hadoop and the Hortonworks Data Platform
- B. Installing the Hortonworks Data Platform
- C. Using HDFS Storage
- D. Managing Apache Ambari Users and Groups
- E. Managing Hadoop Services

Labs:

- Setting Up the Lab Environment
- Installing HDP
- Managing Apache Ambari Users and Groups
- Managing Hadoop Services

II. *Working with HDFS*

- A. Using HDFS Storage
- B. Managing HDFS Storage
- C. Adding, Deleting, and Replacing Worker Nodes
- D. Configuring Rack Awareness

Labs:

- Using Hadoop Storage
- Using WebHDFS
- Using HDFS Access Control Lists
- Managing Hadoop Storage
- Managing HDFS Quotas
- Adding, Decommissioning, and Recommissioning Worker Nodes
- Configuring Rack Awareness

III. *Working with Apache YARN*

- A. YARN Resource Management
- B. YARN Applications
- C. YARN Capacity Scheduler

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
- YARN Node Labels

IV. *High Availability, Backups and Configuring Centralized Cache*

- A. HDFS and YARN High Availability
- B. Monitoring a Cluster
- C. Protecting a Cluster with Backups
- D. Configuring Heterogenous HDFS Storage
- E. Managing the HDFS NFS Gateway
- F. Configuring HDFS Centralized Cache

Labs:

- Configuring NameNode High Availability
- Configuring ResourceManager High Availability
- Managing Apache Ambari Alerts
- Managing HDFS Snapshots
- Using DistCP
- Configuring HDFS Storage Policies
- Configuring an NFS Gateway
- Configuring HDFS Centralized Cache

V. *Performing a Rolling Upgrade*

- A. Apache Hive Tuning
- B. Managing Workflows Using Apache Oozie
- C. Integrating Ambari with LDAP
- D. Automating Cluster Provisioning Using Ambari Blueprints
- E. Performing an HDP Rolling Upgrade

Labs:

- Configuring Apache Hive High Availability
- Managing Workflows Using Apache Oozie
- Integrating Apache Ambari with AD/LDAP
- Automating Cluster Provisioning using Apache Ambari
- Performing an HDP Upgrade