Hewlett Packard Enterprise

HDP Operations: Install and Manage with Apache Ambari H7G70S

This 4-day course covers administration tasks for Hadoop 2.0 clusters. It presents content related to the deployment lifecycle for a multi-node Hadoop cluster including: installation, configuration, monitoring, scaling and how Hadoop works with Big Data. It is 50% Instructor-led lecture/ discussion and 50% hands-on labs.

HDP Operations: Install and Manage with Apache Ambari	
Price	USD $2,795
Links to local schedules, pricing and registration	**US/Canada** **Mexico/Latin America** **Brazil**
HP course #	H7G70S
Category	Big Data
Duration	4 days

Audience

- This course is designed for IT administrators and operators responsible for installing, configuring and supporting an Apache Hadoop 2.0 deployment in a Linux environment

Prerequisites

- Students should have a basic understanding of Hadoop and Linux environments

Course objectives

After completing this course, students should be able to:

- Describe various tools and frameworks in the Hadoop 2.0 ecosystem
- Describe the Hadoop Distributed File System (HDFS) architecture
- Install and configure an HDP 2.0 cluster
- Use Ambari to monitor and manage a cluster
- Describe how files are written to and stored in HDFS
- Perform a file system check using command line and browser-based tools
- Configure the replication factor of a file
- Mount HDFS to a local filesystem using the NFS Gateway
- Deploy and configure YARN on a cluster
- Configure and troubleshoot MapReduce jobs
- Describe how YARN jobs are scheduled
- Configure the capacity and fairschedulers of the ResourceManager
- Use WebHDFS to access a cluster over HTTP
- Configure a Hiveserver
- Describe how Hive tables are created and populated

- Use Sqoop to transfer data between Hadoop and a relational database
- Use Flume to ingest streaming data into HDFS
- Deploy and run an Oozie workflow
- Commission and decommission worker nodes
- Configure a cluster to be rack-aware
- Implement and configure NameNode HA
- Secure a Hadoop cluster

Benefits to you

- This course will provide in depth explanation on how to install, configure and support an Apache Hadoop 2.0 deployment in a Linux environment

Course outline

Day 1
- Unit 1: Introduction to HDP and Hadoop 2.0
- Unit 2: HDFS Architecture
- Unit 3: Installation Prerequisites and Planning
- Unit 4: Configuring Hadoop
- Unit 5: Ensuring Data Integrity

Day 2
- Unit 6: HDFS NFS Gateway
- Unit 7: YARN Architecture and MapReduce
- Unit 8: Job Schedulers
- Unit 9: Enterprise Data Movement
- Unit 10: HDFS Web Services

Day 3
- Unit 11: Hive Administration
- Unit 12: Transferring Data with Sqoop
- Unit 13: Flume
- Unit 14: Oozie
- Unit 15: Monitoring HDP2 Services

Day 4
- Unit 16: Commissioning and Decommissioning Nodes
- Unit 17: Backup and Recovery
- Unit 18: Rack Awareness and Topology
- Unit 19: NameNode HA
- Unit 20: Securing HDP

Lab Content

Students will work through the following lab exercises using the Hortonworks Data Platform 2.0:
- Install HDP 2.0 using Ambari
- Add a new node to the cluster
- Stop and start HDP services
- Use HDFS commands
- Verify data with block scanner and fsck
- Mount HDFS to a local file system

- Troubleshoot a MapReduce job
- Configure the capacity scheduler
- Use distcp to copy data from a remote cluster
- Use WebHDFS
- Understanding Hive tables
- Use Sqoop to transfer data
- Install and test Flume
- Run an Oozie workflow
- Commission and decommission a worker node
- Use HDFS snapshots
- Configure rack awareness
- Implement NameNode HA
- Secure an HDP cluster

Learn more at

hpe.com/us/training/bigdata

H7G70S Ver A.00

c04680810 April 2016

Hewlett Packard Enterprise