



Data Sheet

Administering Apache Hadoop

This three-day course provides training for administrations with the fundamentals required to successfully implement and maintain Hadoop clusters. The course consists of an effective mix of interactive lecture and extensive use of hands on lab exercises.

After successfully completing this each student will receive one free voucher for the Hadoop Certified Administration exam.

Course Objectives

By taking this course, administrators are enabled to perform the following:

- Install, configure and deploy clusters
- Deploy new Nodes
- Determine hardware needs
- Monitor Hadoop clusters using Nagios, Ganglia and Ambari
- Recover from NameNode failure
- Manage hardware upgrades including taking a node off-line, node installation and rebalancing clusters
- Manage log files
- Configure and maintain HBase
- Provide administration of Flume and Oozie

Extensive Hands-on Lab Experience

- Installing Hadoop
- Commissioning a New Node
- Benchmarking MapReduce Jobs
- Creating Users and using Quotas and Trash
- Work with Configuration Files
- NameNode and JobTracker Web UI, Monitor Job
- Logs
- Create Queues in Capacity Scheduler
- Verify Data with Block Scanner
- Using fsck
- Use Rack Topology Script for Rack Awareness
- Commissioning a New Node and Running Balancer
- Decommissioning a Node
- Use distcp to Copy Data from Remote Cluster
- Using Hadoop Archive
- Monitoring with Nagios, Ganglia, JMX and Ambari
- NameNode Failure and Recovery
- Upgrading a Hadoop Cluster
- HBase Configuration Files and Shell Commands
- HBase Administration Tools
- Oozie Administration
- Flume Administration

Course Outline

Day 1

- What is Big Data?
- Overview of Hadoop
- Hadoop Install Types and Daemons
- Installing Hadoop
- Installing a New Node
- HDFS Commands
- Benchmarking
- Multi-user Environment, Quotas and Trash
- Intermediate Sort
- Configuring Hadoop
- Hadoop Ports and Web UI
- Logs

Day 2

- Cluster Hardware and Topology
- Software Environment
- Schedulers
- Data Integrity
- Using Rack Topology
- Commissioning and Decommissioning Nodes
- Hadoop Security
- Hadoop Web Services
- Copying Cluster Data
- Hadoop Archive
- Monitoring Hadoop

Day 3

- NameNode Backup and Recovery
- Hadoop Ecosystem
- Upgrading Hadoop
- HBase Administration
- Oozie
- Flume
- Hadoop 2.0

Target Audience

Administrators who are interested in learning how to deploy and manage a Hadoop cluster. We recommend students have previous experience with UNIX.

Additional Information

- For availability for individual seats in our open enrollment classes please visit us at www.hortonworks.com/training
- Customized, onsite training is also available to be hosted at your offices. For more information around pricing and private training opportunities contact us at hwuniversity@hortonworks.com



Hortonworks certification identifies you as an expert in the Apache Hadoop ecosystem. Hortonworks offers a comprehensive certification program for students that attend a Hortonworks public or private on-site training course. Please visit hortonworks.com/training for more information.



Hortonworks University is your expert source for Apache Hadoop training and certification. Public and private on-site courses are available for developers, administrators, data analysts and other IT professionals involved in implementing big data solutions. Classes combine presentation material with industry-leading hands-on labs that fully prepare students for real-world Hadoop scenarios.



About Hortonworks

Hortonworks is accelerating the adoption of Apache Hadoop by bridging the technology and knowledge gaps that exist today. We are driving much of the design and development of both the current and future generations of Apache Hadoop and leveraging our development expertise to provide unmatched expert technical support, training and certification programs for enterprises, systems integrators and ISVs.

US: 1.855.846.7866
International: 1.408.916.4121
www.hortonworks.com

3460 West Bayshore Road
Palo Alto, CA 94303