



# Hadoop for Systems Administrators H6C60S

This course covers the essentials of deploying and managing an Apache™ Hadoop® cluster. The course is lab intensive with each participant creating their own Hadoop cluster using either the CDH (Cloudera's Distribution, including Apache Hadoop) or Hortonworks Data Platform stacks. Core Hadoop services are explored in depth with emphasis on troubleshooting and recovering from common cluster failures. The fundamentals of related services such as Ambari, Zookeeper, Pig, Hive, HBase, Sqoop, Flume, and Oozie are also covered. The course is approximately 60% lecture and 40% labs.

## Hadoop for Systems Administrators

---

**Price** USD \$2,400

**Links to local schedules, pricing and registration** [US/Canada](#)  
[Mexico/Latin America](#)  
[Brazil](#)

---

**HP course #** H6C60S

---

**Category** Big Data

---

**Duration** 3 days

---

## Audience

- Systems Administrators who will be responsible for managing and administering Hadoop clusters

## Prerequisites

- Qualified participants should be comfortable with the Linux commands and have some systems administration experience, but do not need previous Hadoop experience

## Supported Distributions

- Red Hat Enterprise Linux 6

## Benefits to you

- Hands-on coverage of Hadoop gives systems administrators the skills they need to properly deploy, manage, and maintain Hadoop clusters

## Next steps

- View our Big Data curriculum path
- HP also recommends the HP Roadmap Services for Hadoop from our HP Consulting team

## Course outline

### Chapter 1 - Hadoop: The Big Picture

- Data Analysis
- Big Data
- Hadoop Core Architecture
- Hadoop Ecosystem
- Hadoop Ecosystem continued
- Running Commands on Multiple Systems

#### Lab Tasks

- Running Commands on Multiple Hosts [R6]
- Preparing to Install Hadoop [CDH5 R6]

### Chapter 2 - HDFS

- Design Goals
- Design
- Blocks
- Block Replication
- Namenode Daemon
- Secondary Namenode Daemon
- Datanode Daemon
- Accessing HDFS
- Permissions and Users
- Adding and Removing Datanodes
- Balancing

#### Lab Tasks

- Single Node HDFS [CDH5 R6]
- Multi-node HDFS [CDH5 R6]
- Files and HDFS [CDH5 R6]
- Managing and Maintaining HDFS [CDH5 R6]

### Chapter 3 - MapReduce

- MapReduce
- Terminology and Data Flow
- MapReduce Daemons
- MapReduce Essential Configuration
- Failure and Recovery
- YARN

#### Lab Tasks

- MapReduce [CDH5 R6]

### Chapter 4 - MapReduce Schedulers

- Working with Jobs
- Scheduling Concepts
- FIFO Scheduler
- Fair Scheduler
- Fair Scheduler - Configuration

**Lab Tasks**

- MapReduce Schedulers [R6]

**Appendix A - Installing Hadoop with Ambari**

**Lab Tasks**

- Installing Hadoop with Ambari [R6]

Learn more at

**[hpe.com/us/training/bigdata](http://hpe.com/us/training/bigdata)**