

“Administrator training gave me an excellent jumpstart on acquiring the Hadoop knowledge I needed to address my customers’ big data and cloud challenges. Cloudera saved me oodles of time!”

Canonical

## Cloudera Administrator Training for Apache Hadoop

### Take your knowledge to the next level with Cloudera’s Apache Hadoop Training and Certification

Cloudera University’s four-day administrator training course for Apache Hadoop provides participants with a comprehensive understanding of all the steps necessary to operate and maintain a Hadoop cluster using Cloudera Manager. From installation and configuration through load balancing and tuning, Cloudera’s training course is the best preparation for the real-world challenges faced by Hadoop administrators.

### Hands-On Hadoop

Through instructor-led discussion and interactive, hands-on exercises, participants will navigate the Hadoop ecosystem, learning topics such as:

- Cloudera Manager features that make managing your clusters easier, such as aggregated logging, configuration management, resource management, reports, alerts, and service management.
- The internals of YARN, MapReduce, Spark, and HDFS
- Determining the correct hardware and infrastructure for your cluster
- Proper cluster configuration and deployment to integrate with the data center
- How to load data into the cluster from dynamically-generated files using Flume and from RDBMS using Sqoop
- Configuring the FairScheduler to provide service-level agreements for multiple users of a cluster
- Best practices for preparing and maintaining Apache Hadoop in production
- Troubleshooting, diagnosing, tuning, and solving Hadoop issues

### Audience & Prerequisites

This course is best suited to systems administrators and IT managers who have basic Linux experience. Prior knowledge of Apache Hadoop is not required.

### Administrator Certification

Upon completion of the course, attendees are encouraged to continue their study and register for the Cloudera Certified Administrator for Apache Hadoop (CCA-H) exam. Certification is a great differentiator. It helps establish you as a leader in the field, providing employers and customers with tangible evidence of your skills and expertise.

## Course Outline: Cloudera Administrator Training for Apache Hadoop

### Introduction

#### The Case for Apache Hadoop

- Why Hadoop?
- Fundamental Concepts
- Core Hadoop Components

#### Hadoop Cluster Installation

- Rationale for a Cluster Management Solution
- Cloudera Manager Features
- Cloudera Manager Installation
- Hadoop (CDH) Installation

#### The Hadoop Distributed File System (HDFS)

- HDFS Features
- Writing and Reading Files
- NameNode Memory Considerations
- Overview of HDFS Security
- Web UIs for HDFS
- Using the Hadoop File Shell

#### MapReduce and Spark on YARN

- The Role of Computational Frameworks
- YARN: The Cluster Resource Manager
- MapReduce Concepts
- Apache Spark Concepts
- Running Computational Frameworks on YARN
- Exploring YARN Applications Through the Web UIs, and the Shell
- YARN Application Logs

#### Hadoop Configuration and Daemon Logs

- Cloudera Manager Constructs for Managing Configurations
- Locating Configurations and Applying Configuration Changes
- Managing Role Instances and Adding Services
- Configuring the HDFS Service
- Configuring Hadoop Daemon Logs
- Configuring the YARN Service

#### Getting Data Into HDFS

- Ingesting Data From External Sources With Flume
- Ingesting Data From Relational Databases With Sqoop
- REST Interfaces
- Best Practices for Importing Data

#### Planning Your Hadoop Cluster

- General Planning Considerations
- Choosing the Right Hardware
- Virtualization Options
- Network Considerations
- Configuring Nodes

#### Installing and Configuring Hive, Impala, and Pig

- Hive
- Impala
- Pig

#### Hadoop Clients Including Hue

- What Are Hadoop Clients?
- Installing and Configuring Hadoop Clients
- Installing and Configuring Hue
- Hue Authentication and Authorization

#### Advanced Cluster Configuration

- Advanced Configuration Parameters
- Configuring Hadoop Ports
- Configuring HDFS for Rack Awareness
- Configuring HDFS High Availability

#### Hadoop Security

- Why Hadoop Security Is Important
- Hadoop's Security System Concepts
- What Kerberos Is and how it Works
- Securing a Hadoop Cluster With Kerberos
- Other Security Concepts

#### Managing Resources

- Configuring cgroups with Static Service Pools
- The Fair Scheduler
- Configuring Dynamic Resource Pools
- YARN Memory and CPU Settings
- Impala Query Scheduling

#### Cluster Maintenance

- Checking HDFS Status
- Copying Data Between Clusters
- Adding and Removing Cluster Nodes
- Rebalancing the Cluster
- Directory Snapshots
- Cluster Upgrading

#### Cluster Monitoring and Troubleshooting

- Cloudera Manager Monitoring Features
- Monitoring Hadoop Clusters
- Troubleshooting Hadoop Clusters
- Common Misconfigurations

#### Conclusion