

# Cloudera Administrator Training for Apache Hadoop

---

<b>Code:</b>	HADOOP-ADMIN
<b>Length:</b>	4 days
<b>URL:</b>	<a href="#">View Online</a>

---

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.

## Skills Gained

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
- Configuring and deploying production-scale clusters that provide key Hadoop-related services, including YARN, HDFS, Impala, Hive, Spark, Kudu, and Kafka
- Determining the correct hardware and infrastructure for your cluster
- Proper cluster configuration and deployment to integrate with the data center
- Ingesting, storing, and accessing data in HDFS, Kudu, and cloud object stores such as Amazon S3
- How to load file-based and streaming data into the cluster using Kafka and Flume
- Configuring automatic resource management to ensure service-level agreements are met for multiple users of a cluster
- Best practices for preparing, tuning, and maintaining a production cluster
- Troubleshooting, diagnosing, and solving cluster issues

## Who Can Benefit

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

## Course Details

### The Cloudera Enterprise Data Hub

- Cloudera Enterprise Data Hub
- CDH Overview
- Cloudera Manager Overview

- Hadoop Administrator Responsibilities

## **Installing Cloudera Manager and CDH**

- Cluster Installation Overview
- Cloudera Manager Installation
- CDH Installation
- CDH Cluster Services

## **Configuring a Cloudera Cluster**

- Overview
- Configuration Settings
- Modifying Service Configurations
- Configuration Files
- Managing Role Instances
- Adding New Services
- Adding and Removing Hosts

## **Hadoop Distributed File System**

- Overview
- HDFS Topology and Roles
- Edit Logs and Checkpointing
- HDFS Performance and Fault Tolerance
- HDFS and Hadoop Security Overview
- Web User Interfaces for HDFS
- Using the HDFS Command Line Interface
- Other Command Line Utilities

## **HDFS Data Ingest**

- Data Ingest Overview
- File Formats
- Ingesting Data using File Transfer or REST Interfaces
- Importing Data from Relational Databases with Apache Sqoop
- Ingesting Data From External Sources with Apache Flume
- Best Practices for Importing Data

## **Hive and Impala**

- Apache Hive
- Apache Impala

## **YARN and MapReduce**

- YARN Overview
- Running Applications on YARN

- Viewing YARN Applications
- YARN Application Logs
- MapReduce Applications
- YARN Memory and CPU Settings

## **Apache Spark**

- Spark Overview
- Spark Applications
- How Spark Applications Run on YARN
- Monitoring Spark Applications

## **Planning Your Cluster**

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

## **Advanced Cluster Configuration**

- Configuring Service Ports
- Tuning HDFS and MapReduce
- Enabling HDFS High Availability

## **Managing Resources**

- Configuring cgroups with Static Service Pools
- The Fair Scheduler
- Configuring Dynamic Resource Pools
- Impala Query Scheduling

## **Cluster Maintenance**

- Checking HDFS Status
- Copying Data Between Clusters
- Rebalancing Data in HDFS
- HDFS Directory Snapshots
- Upgrading a Cluster

## **Monitoring Clusters**

- Cloudera Manager Monitoring Features
- Health Tests
- Events and Alerts
- Charts and Reports

- Monitoring Recommendations

## **Cluster Troubleshooting**

- Overview
- Troubleshooting Tools
- Misconfiguration Examples
- Essential Points

## **Installing and Managing Hue**

- Overview
- Managing and Configuring Hue
- Hue Authentication and Authorization

## **Security**

- Hadoop Security Concepts
- Hadoop Authentication Using Kerberos
- Hadoop Authorization
- Hadoop Encryption
- Securing a Hadoop Cluster

## **Apache Kudu**

- Kudu Overview
- Architecture
- Installation and Configuration
- Monitoring and Management Tools

## **Apache Kafka**

- What Is Apache Kafka?
- Apache Kafka Overview
- Apache Kafka Cluster Architecture
- Apache Kafka Command Line Tools
- Using Kafka with Flume

## **Object Storage in the Cloud**

- Object Storage
- Connecting Hadoop to Object Storage

---

Download Whitepaper: Accelerate Your Modernization Efforts with a Cloud-Native Strategy

Get Your Free Copy Now

