cloudera®

TRAINING SHEET

> " 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. 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:

- The internals of YARN, MapReduce, 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 (CCAH) 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.

cloudera®

Course Outline: Cloudera Administrator Training for Apache Hadoop

Introduction
- The Case for Apache Hadoop
- Why Hadoop?
- Core Hadoop Components
- Fundamental Concepts

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

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

YARN and MapReduce
- What Is MapReduce?
- Basic MapReduce Concepts
- YARN Cluster Architecture
- Resource Allocation
- Failure Recovery
- Using the YARN Web UI
- MapReduce Version 1

Planning Your Hadoop Cluster
- General Planning Considerations
- Choosing the Right Hardware
- Network Considerations
- Configuring Nodes
- Planning for Cluster Management

Hadoop Installation and Initial Configuration
- Deployment Types
- Installing Hadoop
- Specifying the Hadoop Configuration
- Performing Initial HDFS Configuration
- Performing Initial YARN and MapReduce Configuration
- Hadoop Logging

Installing and Configuring Hive, Impala, and Pig
- Hive
- Impala
- Pig

Hadoop Clients
- What is a Hadoop Client?
- Installing and Configuring Hadoop Clients
- Installing and Configuring Hue
- Hue Authentication and Authorization

Cloudera Manager
- The Motivation for Cloudera Manager
- Cloudera Manager Features
- Express and Enterprise Versions
- Cloudera Manager Topology
- Installing Cloudera Manager
- Installing Hadoop Using Cloudera Manager
- Performing Basic Administration Tasks Using Cloudera Manager

Advanced Cluster Configuration
- Advanced Configuration Parameters
- Configuring Hadoop Ports
- Explicitly Including and Excluding Hosts
- 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

Managing and Scheduling Jobs
- Managing Running Jobs
- Scheduling Hadoop Jobs
- Configuring the FairScheduler
- Impala Query Scheduling

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

Cluster Monitoring and Troubleshooting
- General System Monitoring
- Monitoring Hadoop Clusters
- Common Troubleshooting Hadoop Clusters
- Common Misconfigurations

Conclusion

Cloudera Certified Administrator for Apache Hadoop (CCAH)

Establish yourself as a trusted and valuable resource by completing the certification exam for Apache Hadoop Administrators. CCAH certifies the core systems administrator skills sought by companies and organizations deploying Apache Hadoop. The exam can be demanding and will test your fluency with concepts and terminology in the following areas:

Hadoop Distributed File System (HDFS)

Recognize and identify daemons and understand the normal operation of an Apache Hadoop cluster, both in data storage and in data processing. Describe the current features of computing systems that motivate a system like Apache Hadoop:

HDFS Design

HDFS Daemons

HDFS Federation

HDFS HA

Securing HDFS (Kerberos)

File Read and Write Paths

MapReduce

Understand MapReduce core concepts and MapReduce v2 (MRv2/YARN).

Apache Hadoop Cluster Planning

Discuss the principal points to consider in choosing the hardware and operating systems to host an Apache Hadoop cluster

Apache Hadoop Cluster Installation and Administration

Analyze cluster handling of disk and machine failures. Recognize and identify tools for monitoring and managing HDFS.

Resource Management

Describe how the default FIFO scheduler and the FairScheduler handle the tasks in a mix of jobs running on a cluster

Monitoring and Logging

Discuss the functions and features of Apache Hadoop's logging and monitoring systems

Ecosystem

Understand ecosystem projects and what you need to do to deploy them on a cluster.

cloudera.com

1-888-789-1488 or 1-650-362-0488
Cloudera, Inc., 1001 Page Mill Road, Palo Alto, CA 94304, USA