

Big Data Hadoop Administrator Certification Training | Course Agenda

Lesson 1: Big Data & Hadoop Introduction

In this lesson you will learn about Big Data characteristics, need of a framework such as Hadoop & its ecosystem. You will also be introduced to important daemons that support functioning of a Hadoop cluster. Topics covered are:

- Data & Existing Solutions
- Welcome to the world of Big Data—What, Why & Where
- Case studies
- Hadoop & its Ecosystem
- Hadoop Core components
- Hadoop & its capabilities

Lesson 2: HDFS - Hadoop Distributed File System & Hadoop's Distributions

In this lesson you will learn about Hadoop Distributed file System, its architecture, working & internals, Hadoop different distributions and about their similarities & differences. Topics covered are:

- Gain knowledge on HDFS its internals, working & features
- Learn about possibilities without HDFS
- Differentiate or find similarities in different distributions of Hadoop.
- Identify the requirements to setup a Hadoop cluster

Lesson 3: Hadoop Cluster Setup & Working with Hadoop Cluster

In this lesson you will learn about steps to setup Apache Hadoop (core distribution) & Cloudera Distribution of Hadoop (vendor specific), cluster management solutions and their benefits and nut & bolts of Cloudera Distribution of Hadoop. You will also learn how to verify your cluster. Topics covered are:

- Need of Cluster Management Solution
- Choice of Installation methods—Automated/ Manual
- Linux machines setup—Virtualization & Cloud
- **Hadoop Cluster Setup**—Apache Hadoop V3 & Cloudera Distribution of Hadoop (CDH)
- Cloudera manager features and capabilities
- Working with Hadoop cluster, HDFS & data
- Working with management console/ UI (user interfaces) & Linux terminals
- Understand administration scenarios

Lesson 4: Hadoop Configurations & Daemon Logs

In this lesson you will learn about configuration files, ports & properties that relate to functioning of Hadoop cluster. You will also learn about Hadoop daemons logs and how they help in problem scenarios for diagnosing & gathering information. Topics covered are:

- List and describe the files that control Hadoop configuration
- Explain how to manage Hadoop configuration with Cloudera Manager
- Locate configuration files and make changes
- Explain how to deal with stale configurations
- Explain the properties of addresses and ports of RPC and HTTP servers run by Hadoop Daemons

- Locate log files generated on hosts
- Filter information in log files
- Explain how to get diagnostic information from log files

Lesson 5: Hadoop Cluster Maintenance & Administration

In this lesson you will learn Hadoop cluster maintenance and administration activities. You will also learn the shortcomings of Hadoop v1 and how they are fulfilled by Hadoop V3 features. Topics covered are:

- Explain how to add and remove nodes in an adhoc way
- Explain how to add and remove nodes in a systematic way, otherwise known as commissioning and decommissioning of nodes
- Explain how to balance a cluster
- List the steps for managing services including adding, deleting, starting, stopping and checking status of services
- Explain the procedure to enable rack awareness
- List the steps to add, remove and move role instances and hosts
- Cite the challenges faced with the first version of Hadoop
- Explain the features in the second version that help overcome the challenges faced with the first version

Lesson 6: Hadoop Computational Frameworks

In this lesson you will learn about different types of computational frameworks, MapReduce & YARN concepts & configurations and how YARN manages applications. Topics covered are:

- Describe the role of computational frameworks
- Explain MapReduce concepts

- Explain YARN framework and concepts
- Describe MRv2 on YARN
- Explain configuring and understanding of YARN
- Describe YARN applications
- Describe YARN memory and CPU settings

Lesson 7: Scheduling—Managing resources via Schedulers

In this lesson you will learn cluster scheduling concepts, managing resources in your YARN cluster by usage of schedulers & queue management to manage jobs/applications. Topics covered are:

- Describe the scheduling concepts
- Identify the Schedulers
- Explain the ways to manage resources using Schedulers
- Describe FIFO, Fair Scheduler, and Capacity Scheduler
- Explain how to configure Schedulers
- Explain queue management

Lesson 8: Hadoop Cluster Planning

In this lesson you will learn about how to plan your Hadoop cluster, considerations for cluster sizing & workload patterns in Hadoop cluster, making choices pertaining to variables such as hardware, software & different cluster deployment options. Topics covered are:

- Planning Hadoop Cluster
- General Planning considerations
- Workload and cluster sizing

- Hadoop Cluster Setup Options—Physical, Virtualization, Cloud or Hybrid
- Making Choices—Hardware, Software & Network
- Making Choices—Master/Slave considerations
- News from the world—Existing Setups

Lesson 9: Hadoop Clients & HUE interface

In this lesson you will learn about Hadoop clients, nodes that support Hadoop clients and web interface such as HUE which can be used to work with Hadoop cluster and its components. Topics covered are:

- Explain the concepts of Hadoop client, edge nodes, and gateway nodes
- Install and configure Hadoop clients
- Explain how Hue works
- Install and configure Hue
- Describe how authentication and authorization is managed in Hue

Lesson 10: Data Ingestion in Hadoop Cluster

In this lesson you will learn about data ingestion types & tools. You will learn more about tools such as Flume, Sqoop that can be used for data import/export. Topics covered are:

- Understand Data Ingestion & its types
- Knowing about various data ingestion tools & their capabilities
- Understanding how Flume works
- Understanding how sqoop works

Lesson 11: Hadoop Ecosystem components/services

In this lesson you will learn about open-source components (also known as services in CDH) that work within Hadoop ecosystem such as Hive, Hbase, kafka & Spark. Topics covered are:

- List some of the services and open-source components that work within the Hadoop ecosystem
- List the advantages and key features of Hive
- Describe briefly about the components of Hive
- Explain how to configure Hive in different modes
- Explain the architecture of HBase and cite the advantages of using HBase
- Explain the working of Apache Kafka
- Describe the architecture of Apache Spark

Lesson 12: Hadoop Security—Securing Hadoop Cluster

In this lesson you will learn about security aspects and security implementation in a Hadoop cluster to secure data & cluster. Topics covered are:

- Describe the different ways to avoid risks and secure data
- Identify the different threat categories
- Describe the security aspects for different nodes
- Describe operating system security
- Describe Kerberos and how it works
- Describe Service Level Authorization

Lesson 13: Cluster Monitoring—Monitoring Hadoop Cluster

In this lesson you will learn about basics of cluster monitoring, choosing right monitoring solutions, Hadoop metrics categories & types and Cloudera manager's features and capabilities that can be used for monitoring your Hadoop cluster. Topics covered are:

- Describe cluster monitoring
- Describe the ways to choose the right monitoring solutions
- List the features and considerations of Cloudera manager for monitoring
- Describe the different categories of Hadoop Metrics
- List the different types of Hadoop Metrics
- List the steps to monitor a cluster by using Cloudera Manager

For information on the course, visit: <https://www.simplilearn.com/big-data-and-analytics/big-data-and-hadoop-administrator-certification-training>



LIVE SUPPORT 24x7



REQUEST A CALL BACK



EMAIL US

LIVE SUPPORT 24x7

REQUEST A CALL BACK

EMAIL US