Masters in Big Data Hadoop Training Certification

IBig Data Hadoop training program helps you master the concepts of Big Data Hadoop and Spark framework to get ready for the Cloudera CCA Spark and master Hadoop Administration with real-time industry-oriented case-study projects. In this Big Data course, you Learn how various components of the Hadoop ecosystem fit into the Big Data processing lifecycle.

Big Data Hadoop Course Overview

Our Big Data Hadoop training is designed to give you an in-depth knowledge of the Big Data framework using Hadoop and Spark.It is a comprehensive Hadoop Big Data training course designed by industry experts considering current industry job requirements to help you learn Big Data Hadoop and Spark modules. In this hands-on Hadoop course, you will execute real-life, industry-based projects using Integrated Lab.This is an industry-recognized Big Data Hadoop certification training course that is a combination of the training courses in Hadoop developer, Hadoop administrator, Hadoop testing and analytics with Apache Spark.

Why should you go for Big Data Hadoop online training?

Upskilling in the Big Data and Analytics field is a smart career decision. According to the Market Research, the global Hadoop market will reach $84.6 Billion by 2021 and there is a shortage of 1.4-1.9 million Hadoop data analysts in the U.S. alone.Big Data is the fastest growing and the most promising technology for handling large volumes of data for doing data analytics. This Big Data Hadoop training will help you be up and running in the most demanding professional skills. Almost all top MNCs are trying to get into Big Data Hadoop; hence, there is a huge demand for certified Big Data professionals. Our Big Data online training will help you learn Big Data and upgrade your career in the Big Data domain. Getting the Big Data certification from us, can put you in a different league when it comes to applying for the best jobs.

What are the prerequisites for taking up this Big Data Hadoop certification training?

There are no prerequisites to take up this course and to master Hadoop. But professionals entering into Big Data Hadoop training should have a basic understanding of Core Java and SQL. If you wish to brush up your Core Java skills.

Eligibility:- Big Data Hadoop training is best suitable for IT, data management, and analytics professionals looking to acquire expertise in Big Data Hadoop, including Software Developers and Architects, Analytics Professionals, Senior IT professionals, Testing and Mainframe Professionals, Data Management Professionals, Business Intelligence Professionals, Project Managers, Aspiring Data Scientists, Graduates looking to begin a career in Big Data Analytics ,

Big Data Hadoop Course Curriculum:-

01 - Installation and Setup Hadoop

The architecture of Hadoop cluster
What is High Availability and Federation?
How to setup a production cluster?
Various shell commands in Hadoop
Understanding configuration files in Hadoop
Installing a single node cluster with Cloudera Manager
Understanding Spark, Scala, Sqoop, Pig, and Flume

02 - Introduction to Big Data Hadoop and Understanding HDFS and MapReduce

Introducing Big Data and Hadoop

What is Big Data and where does Hadoop fit in?

Two important Hadoop ecosystem components, namely, MapReduce and HDFS

In-depth Hadoop Distributed File System – Replications, Block Size, Secondary Name node, High Availability and in-depth YARN – resource manager and node manager

Hands-on Exercise:

1. HDFS working mechanism
2. Data replication process
3. How to determine the size of the block?
4. Understanding a data node and name node

03 - Deep Dive in MapReduce

Learning the working mechanism of MapReduce

Understanding the mapping and reducing stages in MR

Various terminologies in MR like Input Format, Output Format, Partitioners, Combiners, Shuffle, and Sort

Hands-on Exercise:

1. How to write a WordCount program in MapReduce?
2. How to write a Custom Partitioner?
3. What is a MapReduce Combiner?
4. How to run a job in a local job runner
5. Deploying a unit test
6. What is a map side join and reduce side join?
7. What is a tool runner?
8. How to use counters, dataset joining with map side, and reduce side joins?

04 - Introduction to Hive

Introducing Hadoop Hive

Detailed architecture of Hive

Comparing Hive with Pig and RDBMS

Working with Hive Query Language

Creation of a database, table, group by and other clauses

Various types of Hive tables, HCatalog

Storing the Hive Results, Hive partitioning, and Buckets

Hands-on Exercise:

1. Database creation in Hive
2. Dropping a database
3. Hive table creation
4. How to change the database?
5. Data loading
6. Dropping and altering table
7. Pulling data by writing Hive queries with filter conditions
8. Table partitioning in Hive
9. What is a group by clause?

05 - Advanced Hive and Impala

Indexing in Hive

The ap Side Join in Hive

Working with complex data types

The Hive user-defined functions

Introduction to Impala

Comparing Hive with Impala

The detailed architecture of Impala

Hands-on Exercise:

1. How to work with Hive queries?
2. The process of joining the table and writing indexes
3. External table and sequence table deployment
4. Data storage in a different table

06 - Introduction to Pig

Apache Pig introduction and its various features

Various data types and schema in Hive

The available functions in Pig, Hive Bags, Tuples, and Fields

Hands-on Exercise:

1. Working with Pig in MapReduce and local mode
2. Loading of data
3. Limiting data to 4 rows
4. Storing the data into files and working with Group By, Filter By, Distinct, Cross, Split in Hive

07 - Flume, Sqoop and HBase

Apache Sqoop introduction

Importing and exporting data

Performance improvement with Sqoop

Sqoop limitations

Introduction to Flume and understanding the architecture of Flume

What is HBase and the CAP theorem?

Hands-on Exercise:

1. Working with Flume to generate Sequence Number and consume it
2. Using the Flume Agent to consume the Twitter data
3. Using AVRO to create Hive Table
4. AVRO with Pig
5. Creating Table in HBase
6. Deploying Disable, Scan, and Enable Table

08 - Writing Spark Applications Using Scala

Using Scala for writing Apache Spark applications

Detailed study of Scala

The need for Scala

The concept of object-oriented programming

Executing the Scala code

Various classes in Scala like getters, setters, constructors, abstract, extending objects, overriding methods

The Java and Scala interoperability

The concept of functional programming and anonymous functions

Bobsrockets package and comparing the mutable and immutable collections

Scala REPL, Lazy Values, Control Structures in Scala, Directed Acyclic Graph (DAG), first Spark application using SBT/Eclipse, Spark Web UI, Spark in Hadoop ecosystem.

Hands-on Exercise:

1. Writing Spark application using Scala

2. Understanding the robustness of Scala for Spark real-time analytics operation

09 - Spark framework

Detailed Apache Spark and its various features

Comparing with Hadoop

Various Spark components

Combining HDFS with Spark and Scalding

Introduction to Scala

Importance of Scala and RDD

Hands-on Exercise:

1. The Resilient Distributed Dataset (RDD) in Spark

2. How does it help to speed up Big Data processing?

10 - RDD in Spark

Understanding the Spark RDD operations

Comparison of Spark with MapReduce

What is a Spark transformation?

Loading data in Spark

Types of RDD operations viz. transformation and action

What is a Key/Value pair?

Hands-on Exercise:

1. How to deploy RDD with HDFS?

2. Using the in-memory dataset

3. Using file for RDD

4. How to define the base RDD from an external file?

5. Deploying RDD via transformation

6. Using the Map and Reduce functions

7. Working on word count and count log severity

11 - Data Frames and Spark SQL

The detailed Spark SQL

The significance of SQL in Spark for working with structured data processing

Spark SQL JSON support

Working with XML data and parquet files

Creating Hive Context

Writing Dataframe to Hive

How to read a JDBC file?

Significance of a Spark data frame

How to create a data frame?

What is schema manual inferring?

Work with CSV files, JDBC table reading, data conversion from Data Frame to JDBC, Spark SQL user-defined functions, shared variable, and accumulators

How to query and transform data in Data Frames?

How data frames provide the benefits of both Spark RDD and Spark SQL?

Deploying Hive on Spark as the execution engine

Hands-on Exercise:

1. Data querying and transformation using Data Frames

2. Finding out the benefits of Data Frames over Spark SQL and Spark RDD

12 - Machine Learning Using Spark (MLlib)

Introduction to Spark MLlib

Understanding various algorithms

What is Spark iterative algorithm?

Spark graph processing analysis

Introducing Machine Learning

K-Means clustering

Spark variables like shared and broadcast variables

What are accumulators?

Various ML algorithms supported by MLlib

Linear regression, logistic regression, decision tree, random forest, and K-means clustering techniques

Hands-on Exercise:

1. Building a recommendation engine

13 - Integrating Apache Flume and Apache Kafka

Why Kafka?

What is Kafka?

Kafka architecture

Kafka workflow

Configuring Kafka cluster

Basic operations

Kafka monitoring tools

Integrating Apache Flume and Apache Kafka

Hands-on Exercise:

1. Configuring Single Node Single Broker Cluster

2. Configuring Single Node Multi Broker Cluster

3. Producing and consuming messages

4. Integrating Apache Flume and Apache Kafka.

14 - Spark Streaming

Introduction to Spark streaming

The architecture of Spark streaming

Working with the Spark streaming program

Processing data using Spark streaming

Requesting count and DStream

Multi-batch and sliding window operations

Working with advanced data sources

Features of Spark streaming

Spark Streaming workflow

Initializing StreamingContext

Discretized Streams (DStreams)

Input DStreams and Receivers

Transformations on DStreams

Output Operations on DStreams

Windowed operators and its uses

Important Windowed operators and Stateful operators

Hands-on Exercise:

1. Twitter Sentiment analysis
2. Streaming using Netcat server
3. Kafka-Spark streaming
4. Spark-Flume streaming

15 - Hadoop Administration – Multi-node Cluster Setup Using Amazon EC2

Create a 4-node Hadoop cluster setup

Running the MapReduce Jobs on the Hadoop cluster

Successfully running the MapReduce code

Working with the Cloudera Manager setup

Hands-on Exercise:

1. The method to build a multi-node Hadoop cluster using an Amazon EC2 instance
2. Working with the Cloudera Manager

16 - Hadoop Administration – Cluster Configuration

Overview of Hadoop configuration

The importance of Hadoop configuration file

The various parameters and values of configuration

The HDFS parameters and MapReduce parameters

Setting up the Hadoop environment

The Include and Exclude configuration files

The administration and maintenance of name node, data node directory structures, and files

What is a File system image?

Understanding Edit log

Hands-on Exercise:

1. The process of performance tuning in MapReduce

17 - Hadoop Administration – Maintenance, Monitoring and Troubleshooting

Introduction to the checkpoint procedure, name node failure

How to ensure the recovery procedure, Safe Mode, Metadata and Data backup, various potential problems and solutions, what to look for and how to add and remove nodes

Hands-on Exercise:

1. How to go about ensuring the MapReduce File System Recovery for different scenarios

2. JMX monitoring of the Hadoop cluster

3. How to use the logs and stack traces for monitoring and troubleshooting

4. Using the Job Scheduler for scheduling jobs in the same cluster

5. Getting the MapReduce job submission flow

6. FIFO schedule

7. Getting to know the Fair Scheduler and its configuration

18 - ETL Connectivity with Hadoop Ecosystem (Self-Paced)

How ETL tools work in the Big Data industry?

Introduction to ETL and data warehousing

Working with prominent use cases of Big Data in ETL industry

End-to-end ETL PoC showing Big Data integration with ETL tool

Hands-on Exercise:

1. Connecting to HDFS from ETL tool

2. Moving data from Local system to HDFS

3. Moving data from DBMS to HDFS,

4. Working with Hive with ETL Tool

5. Creating MapReduce job in ETL tool

19 - Project Solution Discussion and Cloudera Certification Tips and Tricks

Working towards the solution of the Hadoop project solution

Its problem statements and the possible solution outcomes

Preparing for the Cloudera certifications

Points to focus on scoring the highest marks

Tips for cracking Hadoop interview questions

Hands-on Exercise:

1. The project of a real-world high value Big Data Hadoop application

2. Getting the right solution based on the criteria set by the Intellipaat team

20 - Hadoop Application Testing

Importance of testing

Unit testing, Integration testing, Performance testing, Diagnostics, Nightly QA test, Benchmark and end-to-end tests, Functional testing, Release certification testing, Security testing, Scalability testing, Commissioning and Decommissioning of data nodes testing, Reliability testing, and Release testing

21 - Roles and Responsibilities of Hadoop Testing Professional

Understanding the Requirement

Preparation of the Testing Estimation

Test Cases, Test Data, Test Bed Creation, Test Execution, Defect Reporting, Defect Retest, Daily Status report delivery, Test completion, ETL testing at every stage (HDFS, Hive and HBase) while loading the input (logs, files, records, etc.) using Sqoop/Flume which includes but not limited to data verification, Reconciliation, User Authorization and Authentication testing (Groups, Users, Privileges, etc.), reporting defects to the development team or manager and driving them to closure

Consolidating all the defects and create defect reports

Validating new feature and issues in Core Hadoop

22 - Framework Called MRUnit for Testing of MapReduce Programs

Report defects to the development team or manager and driving them to closure

Consolidate all the defects and create defect reports

Responsible for creating a testing framework called MRUnit for testing of MapReduce programs

23 - Unit Testing

Automation testing using the OOZIE

Data validation using the query surge tool

24 - Test Execution

Test plan for HDFS upgrade

Test automation and result

25 - Test Plan Strategy and Writing Test Cases for Testing Hadoop Application

Test, install and configure

Big Data Hadoop Training & Certification

The entire training course content is in line with these certification programs and helps you clear these certifth ease and get the best jobs in the top MNCs.As part of this ication exams witraining, you will be working on real-time projects and assignments that have immense implications in the real-world industry scenarios, thus helping you fast-track your career effortlessly.Upon successful completion of the Big Data Hadoop training, you will be awarded the course completion certificate from our side.

FAQs

Why Should I Learn Hadoop from this Training Program:- The Big Data Hadoop training in Hyderabad is designed to help the candidates achieve Hadoop certification exam. It also gives a complete overview of the Big Data Framework using Hadoop and Spark. Learn to enhance your skills in using Spark for real-time data processing, including parallel processing in Spark, implementing Spark applications It is a known fact that the demand for Hadoop professionals far outstrips the supply. So, if you want to learn and make a career in Hadoop, then you need to enroll for our Hadoop course which is the most recognized name in Hadoop training and certification.Our entire Hadoop training has been created by industry professionals. You will get 24/7 lifetime support, high-quality course material and videos and free upgrade to the latest version of course material. Thus, it is clearly a one-time investment for a lifetime of benefits.

What is Hadoop :- Hadoop is an open-source framework which allows organizations to store and process big data in a parallel and distributed environment. It is used to store and combine data, and it scales up from one server to thousands of machines, each offering low-cost storage and as well as local computation

What is Hadoop :- Spark is considered by many to be a more advanced product than Hadoop. It's an open-source framework that provides several interconnected platforms, systems, and standards for big data projects.

We can Add other FAQs which relate to payments and refund and if we give job assistance then that can be also added.