

Hadoop Developer - Course Curriculum

Course Duration: 50 Hours

Module 1: Introduction to Big Data & Hadoop

Objective: In this module, you will be able to get big data context, its definition and its integration with Hadoop in terms of storage and processing.

Topics:

Big Data context with case studies, Big Data definition, structured vs unstructured, Business analytics lifecycle, Hadoop basics, Hadoop characteristics, Hadoop ecosystem, Hadoop core components, Secondary name node.

Module 2: HDFS Internals and YARN

Objective: In this module, you will learn the concept of Blocks, Rack awareness, HDFS architecture and a few Hadoop commands. This module also covers HDFS High Availability, HDFS Federation and YARN.

Topics:

Blocks, Data Replication and Rack Awareness, HDFS File Write Anatomy, HDFS File Read Anatomy, HDFS Architecture, Common Hadoop Shell Commands, High Availability, HDFS Federation, YARN, Firing our First Map Reduce Job, Checking the output of M/R Job and Understanding the dump of Map Reduce Job.

Module 3: Introduction to Map Reduce

Objective: In this module, you will learn about different Hadoop modes, configuration files, Split vs Blocks, traditional and Hadoop based distributed computing techniques. You will also learn advanced concepts like Combiners & Partitioners and Shuffle & Sort

Topics:

Hadoop Cluster Modes, Configuration Files, Web URLs, Split vs Blocks, Map Reduce use-cases, solving a problem in Traditional way, Understanding Map Reduce way, Map Reduce Anatomy, Advantages of Map Reduce, and Map Reduce Flow

Module 4: Advanced Map Reduce Concepts

Objective: In this module, you will learn the advanced concepts like MR Unit, Counters, Distributed Cache and Joins

Topics:

MR Unit, Counters, Distributed Cache, Joins, Secondary Sort and Total Order Sort.

Module 5: Pig and Advance Pig

Objective: In this module, you will be get a complete understanding of PIG & its association with Map Reduce.

Topics:

Pig Background, Need for Pig, Pig Vs M/R, Pig Definition, Pig Latin, Pig users, Pig usage at Yahoo, Pig Interaction Modes, Pig program execution, Pig data model, Pig data types, Pig operators and specialized Joins

Module 6: Hive and Advance Hive

Objective: In this module you will learn about Hive Background, Hive comparison with RDBMS and Hive design and architecture.

Topics:

Hive Background, Hive Definition, Pig vs Hive, RDBMS vs Hive, Hive components, Hive Architecture, Hive Meta Store, Hive Design, Hive Data Model, Partition and Buckets

Module 7: Hbase and Advanced HBase

Objective: In this module you will learn about NoSQL Background, Hive comparison with RDBMS and Hive design and architecture.

Topics:

NoSQL background and description, Real time scenarios, NoSQL landscapes, HBase definition, HBase characteristics, HBase history, HBase vs. RDBMS, HBase Data Model, HBase Data Model – Graphical representation, HBase Data Model – Logical Vs. Physical representation, Version concepts, Region and Region Servers and Zookeeper

Module 8: Oozie and Scoop

Objective: In this module, you will learn the concepts of Oozie as a Hadoop Workflow Framework and how it orchestrates the execution of Hadoop Components.

Topics:

Oozie workflow, Oozie server , Oozie co-ordinator, Oozie Bundles, Configuration XML and Properties file, Creating Oozie application, Oozie web console, Oozie scheduling, Scoop Setup between Hadoop and RDBMS, Exporting Data from Hadoop into RDBMS, Importing Data from RDBMS into Hadoop.

Module 9: Zookeeper and Flume

Objective: In this Module, you will learn about Zookeeper as a distributed cluster co-ordination system and how it can be used to keep clusters synchronized and avoid race conditions. You will also learn about Flume as a Hadoop sub-component to pull data from unstructured data sources.

Topics:

Zookeeper Master, Zookeeper Slave, concept of Ephemeral Nodes, Persistent and Optional Sequential Numbering, Configuration Management. Flume agent, Source, Sync, Defining the Flume flow, configuring individual components, configuring entire Flume set-up

Module 10: Project Set-up Discussion

Objective: In this session, you will get familiar with the project you would be working for your certification along with several other topics.

Topics:

Project Discussions, Evaluating Individual Approaches, Finalizing Optimal Approach