

Big Data Hadoop Developer

1. Understanding Big Data and Hadoop

- Introduction to Big Data
- Limitations and Solutions of existing Data Analytics Architecture
- Hadoop Features
- Hadoop Ecosystem
- Hadoop 2.x core components
- Hadoop Storage: HDFS
- Hadoop Processing: MapReduce Framework
- Anatomy of File Read and File Write
- Rack Awareness.

2. Hadoop Architecture and HDFS

- Hadoop 2.x Cluster Architecture - Federation and High Availability
- Hadoop Cluster Modes
- Common Hadoop Shell Commands
- Hadoop 2.x Configuration Files
- Password-Less SSH
- MapReduce Job Execution

3. Introduction To MapReduce

- MapReduce introduction
- Understanding MapReduce Part1
- Understanding MapReduce Part2
- Running first MapReduce program
- Combiner and Tool runner

4. MapReduce Types and Formats

- MapReduce types and formats
- Experiments with Defaults
- IO Format classes
- Experiments with output formats advanced concepts

5. Classic MapReduce and YARN

- Anatomy of MapReduce job run
- Job run classic MapReduce
- Failure scenarios - classic MapReduce

- Job run YARN
- Failure scenarios -YARN
- Job scheduling in MapReduce
- Shuffle and Sort
- Performance tuning features

6. Advanced MapReduce concepts

- Looking at Counters
- Sorting ideas with partitioner-Part 1
- Sorting ideas with partitioner-Part 2
- Map side join operation
- Reduce side join operation
- Side distribution of data
- Hadoop streaming and Hadoop pipes

7. Sqoop

- Introduction
- List Databases, List Tables
- Eval command
- Import all the tables to HDFS
- Import data from mysql to HDFS
- Import data from mysql to Hive
- Import data from mysql to Hbase
- Export data from hdfs to mysql
- File formats
- Sqoop job
- Sqoop delimiters

8. Flume

- Introduction
- Flume Agent
- Spool to Logger
- Spool to HDFS
- HTTP to HDFS
- Spool to HBASE

9. Pig

- Introduction
- MapReduce Vs Pig
- Pig Latin scripting
- Modes of script execution

- Load and Store operator
- Diagnostic operators
- Grouping and Joining
- Combining and Splitting
- Filter
- Sorting
- Built in function
- Work with piggybank jar

10. Hive

- Introduction
- Hive Vs Pig
- Metastore in Hive
- Comparison with Traditional Database
- Hive Tables (Managed Tables and External Tables)
- Hive Data Types and Data Models, Partitions and Buckets
- Importing Data From local file system and HDFS
- Querying Data
- Run hive query using Tez
- Join in Hive
- Hive Script
- Hive UDF

11. HBase

- Introduction
- HBase v/s RDBMS
- HBase Architecture
- HBase Data Model
- HBase Shell
- HBase Client API
- Data Loading Techniques
- Filters in HBase.

12. Oozie

- Introduction
- Oozie Components
- Oozie Workflow
- Oozie property file
- Oozie coordinator
- Scheduling with Oozie
- Oozie Commands
- Oozie Web Console

