

HADOOP

Curriculum



Duration: - 45 Days

1. Introduction to Big Data and Hadoop

- * What is Big Data?
- * Challenges for processing big data?
- * Technologies support big data
- * Overview and History of Hadoop
- * Hadoop Advantages and Disadvantages

2. HDFS (Hadoop Distributed File System)

- * HDFS Design Goals
- * HDFS architecture
- * Data Storage in HDFS
- * HDFS Block Size and Commands
- * Accessing HDFS using CLI

3. Working with HDFS

- * Configuring Hadoop Cluster
- * Hadoop 2.x.x version features
- * Difference between 1.x.x and 2.x.x versions

4. Apache MapReduce

- * MapReduce Architecture
- * Data Types in Hadoop
- * Programing Model
- * Configuring and Writing MapReduce jobs in IDE

- * MapReduce Commands

5. Apache Hive

- * Introduction
- * Hive Architecture
- * Installation and Configuration
- * Hive Services and metastore
- * Data Types and File Formats
- * Views in Hive

6. YARN (Next Generation MapReduce)

- * What is YARN?
- * Difference between YARN and MapReduce
- * YARN Architecture
- * Installation of YARN

7. Apache PIG

- * Introduction to Pig
- * MapReduce vs Pig
- * Data Types in Pig
- * Modes of execution in Pig

8. Apache Hbase

- * Hbase Introduction
- * Hbase use cases
- * Hbase Architecture and Installation