

# Syllabus for Hadoop Administration

## Course Duration For Hadoop Administration Course :

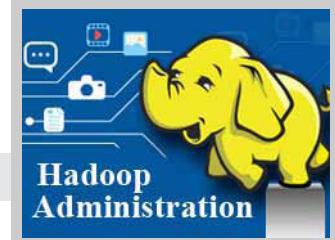
- 8 Weekend (Weekend batches)

## Objective For Hadoop Administration Course :

- Understand the fundamental concepts of Hadoop
- Understand HDFS features
- Provide Insights About the Roles of a Data Scientist
- Ability to Analyze Big Data
- Make predictions using machine learning
- Learn to apply hypotheses and data into actionable predictions

## Eligibility For Hadoop Administration Course :

- BSc, BCS, BCA, BE, B.Tech, MSc, MCS, MCA, M.Tech
- Fundamental knowledge of Java and Linux environment shall be preferred



## Hadoop Administration

### Hadoop Administration

- Introduction to Big Data and Hadoop
- Types Of Data
- Characteristics Of Big Data
- Business Benefits Of Big Data Technology
- Hadoop And Traditional Rdbms
- Hadoop Core Services

### Hadoop Installation and Configuration

- Ubuntu Server-Introduction
- Hadoop and Multi-Node Installation
- Create a Clone of Hadoop Virtual Machine
- Perform Clustering of the Hadoop Environment

### Hadoop Distributed File System

- Introduction to Hadoop Distributed File System
- Goals of HDFS
- HDFS Architecture
- Design of HDFS
- Hadoop Storage Mechanism
- Measures of Capacity Execution
- HDFS Storage Architecture Heterogeneous Storage
- HDFS Commands

### The MapReduce Framework

- Understanding MapReduce
- The Map and Reduce Phase
- WordCount in MapReduce
- Running MapReduce Job

# Syllabus for

## Hadoop Administration

### Planning Hadoop Cluster

- Architecture of Hadoop Cluster
- Workflow of Hadoop Cluster
- HDFS Writes
- Preparing for HDFS Writes
- Pipelined HDFS Write
- NameNode Functionality
- Replicating Missing Replicas
- HDFS Reads
- Factors for Planning Hadoop Cluster
- Single-Node and Multi-Node Cluster Configuration
- HDFS Block replication and rack awareness
- Topology and Components of Hadoop Cluster

### Cluster Maintenance

- Checking HDFS Status
- Breaking the cluster
- Copying Data Between Clusters
- Adding and Removing Cluster Nodes
- Rebalancing the cluster
- Name Node Metadata Backup
- Cluster Upgrading

### Advanced Cluster Configuration Features

- Hadoop Configuration Overview
- Types of Configuration Files
- Hadoop Cluster and Map Reduce Configuration Parameters with Values
- Hadoop Environment Setup
- Include and Exclude Configuration Files

### Managing and Scheduling Jobs

- Managing Jobs
- The FIFO and Fair Schedule
- How to stop and start jobs running on the cluster

### Cluster Monitoring, Troubleshooting and Optimizing

- General System conditions to Monitor
- Name Node and Job Tracker Web Uis
- View and Manage Hadoop's Log files
- Ganglia Monitoring Tool
- Common cluster issues and their resolutions

### YARN

- Introduction to YARN
- Need for YARN
- YARN Architecture
- YARN Installation and Configuration

### Extending Hadoop

- Simplifying information access
- Enabling SQL-like querying with Hive
- Installing Pig to create MapReduce jobs
- Imposing a tabular view on HDFS with HBase
- Configuring Oozie to schedule workflows

### Installing and Managing Hadoop Ecosystem

- Sqoop
- Flume
- Hive
- Pig
- HBase
- Oozie