



**Hadoop**

## Table of Contents

<i>Overview</i> .....	2
<i>Students</i> .....	3
<i>Curriculum</i> .....	4



# Overview

Hadoop Course mainly focuses on students with interest towards data and design of proper data sources to reduce loss of data and duplication of data. This course primarily focuses on Hadoop Cluster and its Architecture, Cluster Setup and Working, Cluster Administration and Maintenance, computational frameworks, Managing Resources, Scheduling, Planning, Security and cluster monitoring.

By the end of the course, students will be able to:

- Create and Administer the HDFS cluster
- Work and master the components of Hadoop
- Basics of Linux



## Students

- Fresh Graduates** Students who want to improve the skill set and meet the current trends in the job market. They will be able to make their own Career in the world of Information Technology.
- Career Shift** Someone looking for a career shift or start the career after having some gap in their career. This program helps to bridge the gap and make their career in Information Technology.





# Curriculum

## Understanding Big Data and Hadoop

- Introduction to Big data
- Common big data domain scenarios
- Limitations of traditional solutions
- Hadoop 1.0 & 2.0 ecosystem and its core components
- Application submission in YARN

## Hadoop Cluster and its Architecture

- Distributed File System
- Hadoop Cluster Architecture
- Replication Rules
- Hadoop Cluster Modes
- Rack awareness theory
- Hadoop Cluster Administrator responsibilities
- Understand working of HDFS
- NTP Server
- Configurations for installing Hadoop
- Deploying Hadoop in a pseudo-distributed mode

## Hadoop Cluster Setup and Working

- OS Tuning for Hadoop Performance
- Pre-requisite for installing Hadoop
- Hadoop Configuration files
- Stale Configuration
- RPC and HTTP Server Properties
- Properties of Namenode, Datanode and Secondary Namenode
- Log Files in Hadoop
- Deploying a multi-node Hadoop cluster

## Hadoop Cluster Administration and Maintenance

- Commissioning and Decommissioning of Node
- HDFS Balancer
- Namenode Federation in Hadoop
- High Availability in Hadoop
- Trash Functionality
- Checkpointing in Hadoop
- Distcp
- Disk Balancer
- 

## Computational Frameworks, Managing Resources and Scheduling

- Different processing frameworks
- Different phases in MapReduce
- Spark and its Features
- Application Workflow in YARN
- YARN Metrics
- YARN Capacity Scheduler



## COURSE CURRICULUM HADOOP

- Service Level Authorization

### **Hadoop 2.x Cluster Planning and Management**

- Planning Hadoop 2.x Cluster
- Cluster Sizing
- Hardware, Network and Software considerations
- Popular Hadoop distributions
- Workload and usage patterns

### **Hadoop Security and Cluster Monitoring**

- Monitoring Hadoop Clusters
- Hadoop Security System Concepts
- Securing a Hadoop Cluster Kerberos
- Common misconfigurations
- Troubleshooting

### **Cloudera Hadoop 2.x and its Features**

- Visualize Cloudera Manager
- Features of Cloudera Manager
- Build Cloudera Hadoop Cluster using CDH
- Installation choices in Cloudera
- HUE and its Architecture
- HUE interface
- HUE Features



### **Pig, Hive installation and working**

- Pig setup and working
- Hive setup and configuration
- Working with Hive
- Setting Hive in local and remote metastore mode

### **HBase, Zookeeper Installation and Working**

- HBase data model, Architecture and Hfile
- HBase setup
- MemStore, WAL, BlockCache
- Installing Zookeeper

#### **Oozie**

- Overview and Features
- Workflow, Coordinator and bundle
- Start, End and Error Node
- Action Node
- Join and Fork
- Decision Node
- Oozie CLI
- Install Oozie



## **Data Ingestion using Sqoop and Flume**

- Types of Data Ingestion
- HDFS data loading commands
- Purpose and features of Sqoop
- Perform operations like, Sqoop Import, Export and Hive Import
- Sqoop 2
- Install Sqoop
- Import data from RDBMS into HDFS
- Flume features and architecture
- Types of flow
- Install Flume
- Ingest Data from External Sources with Flume
- Best Practices for Importing Data

