Big Data Hadoop Administration

Course Duration: 2 Months

1. **Big Data Overview**

   - **What is big data**

   - **Discussion over Databases**

   - **Databases v/s Hadoop**

   - **Problems with Large scale data**

2. **Overview of Apache Hadoop**

   - **Why Hadoop**

   - **Apache Hadoop Architecture**

   - **Apache Hadoop workflow**

   - **Apache hadoop Component**

3. **Installation of Apache Hadoop**

   - **Basic Prerequisites for Hadoop**

   - **Apache Hadoop Standalone Installation**

   - **Apache Hadoop Management**

4. **Hadoop Distributed File System (HDFS)**

   - **Understanding HDFS Architecture**

   - **Understanding HDFS Management and Core Component**

   - **HDFS Snapshot and Management**

   - **Understanding FSImage and edit logs management**

5. **Hadoop Component Management and Understanding**

    - **Understanding Apache Hadoop Namenode**

    - **Understanding Apache Hadoop Datanode**

    - **Understanding Apache Secondary NameNode**

    - **Apache Hadoop Backup & Management**

6. **Hadoop Cluster Installation and Management**

    - **Hadoop multi Node Cluster setup**

    - **Hadoop Cluster Management**

    - **Include and Exclude Data node in cluster**

7. **Hadoop Yarn Architecture & Management**

    - **Yarn service Introduction & Architecture**

    - **Yarn Resource manager & Node Manager**

    - **Yarn Scheduling and Management**

8. **Apache Hadoop Administration**

    - **Managing Hadoop Cli**

    - **Managing Hadoop Web**

    - **Managing Failover & Nodes**

9. **Apache Hive Management**

    - **Hive Architecture and Hadoop**

    - **Hive Installation and management**

    - **Data manipulation with Hive**

10. **Apache Pig Management**

    - **Pig Architecture**

    - **Pig Installation and management**

- **Pig Latin script for Data Manipulation**

11. Eco system process over Hadoop

- **Configure and Management of Sqoop**
- **Application Management using Flume**
- **Configure Management of HBASE**

12. High Availability over Hadoop

- **Deployment Namenode on HA using Zookeeper**
- **Namenode as a standby and active**

13. Overview of other tools and Related System

- **Cloudera manager**
- **Apache Strom**
- **Ambari**