



COURSE AGENDA

BIG DATA AND HADOOP - ONLINE TRAINING

LESSONS

1. Introduction to Big Data and Hadoop

- a. What is Big Data?
- b. Why Big Data?
- c. Limitations of Big Data
- d. Hadoop Background
- e. The Hadoop Way

2. Getting Started With Hadoop

- a. Setting up VM Hadoop Environment
- b. Installing VMware Player
- c. Setting up the Virtual Environment (Virtual Machine User Accounts; Running a Hadoop Job; Accessing the VM via ssh; Shutting Down the VM)

3. Hadoop Architecture

- a. Hadoop Cluster in commodity hardware
- b. Hadoop core services and components
- c. Regular file system vs. Hadoop
- d. HDFS layer
- e. HDFS operation principle
- f. HDFS 1.0 & HDFS 2.0

4. Hadoop Deployment

- a. Hadoop installation
- b. Single node and multi node configuration
- c. Hadoop Configuration in cluster environment

5. MapReduce

- a. MapReduce concepts
- b. Hadoop MapReduce example
- c. Hadoop MapReduce requirements
- d. Steps of Hadoop MapReduce
- e. MapReduce user supply
- f. MapReduce framework
- g. Basics of MapReduce programming

6. Advanced MapReduce

- a. Custom Data Types
- b. Input formats
- c. Output formats
- d. Combiners and Partitioners
- e. Error handling and Unit Testing

7. PIG

- a. Introduction to PIG
- b. Why PIG
- c. Comparison between PIG and SQL
- d. Installing and configuring PIG
- e. Running PIG
- f. PIG Latin

8. HIVE

- a. Why another data warehousing system
- b. What is HIVE
- c. Type System
- d. Data Model - Tables, Partitions, Buckets, External Tables
- e. Serialization/De-serialization
- f. Hive file formats
- g. System Architecture and components
- h. Hive Query Language
- i. HIVE: Installing, running, and programming
- j. Difference between Hive and PIG

9. HBase

- a. HBase introduction
- b. HBase history
- c. Who uses Hbase
- d. When to use HBase
- e. HBase Data Model
- f. HBase Families
- g. HBase Components
- h. Row Distribution between region servers
- i. Data Storage
- j. HBase Master
- k. HBase and Zookeeper

10. Cloudera

- a. What is Cloudera
- b. Cloudera enterprise pictorial view
- c. Downloading Cloudera Quickstart VM
- d. Starting the Cloudera VM
- e. Exploring the Welcome Page
- f. Understanding Hue
- g. Understanding Cloudera Manager

11. ZooKeeper and Sqoop

- a. Introduction to ZooKeeper
- b. What is ZooKeeper
- c. Challenges faced in distributed applications
- d. Coordination
- e. ZooKeeper: Goals and Uses
- f. ZooKeeper: Entities, Data Model, Services
- g. Client APIs
- h. Introduction to Sqoop (Why, what, processing, under the hood)
- i. Importing data into Hive
- j. Importing data into HBase
- k. Exporting data from Hadoop using Sqoop
- l. Sqoop Connectors
- m. Connecting MongoDB (NoSQL database)

12. Ecosystem and its Components

a. Why Flume and Chukwa

b. What is Apache Flume

c. Flume Model

d. Scalability in Flume

e. Chukwa

f. Chukwa Architecture

g. Chukwa Capability

h. Chukwa Agent

i. Introduction to Apache Oozie

j. Apache Oozie Workflow

k. Introduction to Mahout

l. Introduction to YARN

m. YARN Architecture

n. Apache Cassandra

o. Why Apache Cassandra

13. Hadoop Administration and Troubleshooting

a. Different configurations of Hadoop cluster

b. Performance monitoring

c. Performance tuning

d. Troubleshooting and Log observation

**Need help? Ask a question or contact our Support team on
+1 281 816 3008 (US) OR +91 80 6435 0979 (India)**

