



COURSE AGENDA

BIG DATA HADOOP – ONLINE TRAINING

LESSONS

1. Introduction to Big Data and Hadoop

What is Big Data?

Why Big Data?

Limitations of Big Data

Hadoop Background

The Hadoop Way

2. Getting Started With Hadoop

Setting up VM Hadoop Environment

Installing VMware Player

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

Hadoop Cluster in commodity hardware

Hadoop core services and components

Regular file system vs. Hadoop

HDFS layer

HDFS operation principle

HDFS 1.0 & HDFS 2.0

4. Hadoop Deployment

Hadoop installation

Single node and multi node configuration

Hadoop Configuration in cluster environment

5. MapReduce

MapReduce concepts

Hadoop MapReduce example

Hadoop MapReduce requirements

Steps of Hadoop MapReduce

MapReduce user supply

MapReduce framework

Basics of MapReduce programming

6. Advanced MapReduce

Custom Data Types

Input formats

Output formats

Combiners and Partitioners

Error handling and Unit Testing

7. PIG

Introduction to PIG

Why PIG

Comparison between PIG and SQL

Installing and configuring PIG

Running PIG

PIG Latin

8. HIVE

Why another data warehousing system

What is HIVE

Type System

Data Model - Tables, Partitions, Buckets, External Tables

Serialization/De-serialization

Hive file formats

System Architecture and components

Hive Query Language

HIVE: Installing, running, and programming

Difference between Hive and PIG

9. HBase

HBase introduction

HBase history

Who uses Hbase

When to use HBase

HBase Data Model

HBase Families

HBase Components

Row Distribution between region servers

Data Storage

HBase Master

HBase and Zookeeper

HBase Deployment

Installation of HBase

Configuration of HBase

10. Cloudera

What is Cloudera

Cloudera enterprise pictorial view

Downloading Cloudera Quickstart VM

Starting the Cloudera VM

Exploring the Welcome Page

Understanding Hue

Understanding Cloudera Manager

11. ZooKeeper and Sqoop

Introduction to ZooKeeper

What is ZooKeeper

Challenges faced in distributed applications

Coordination

ZooKeeper: Goals and Uses

ZooKeeper: Entities, Data Model, Services

Client APIs

Introduction to Sqoop (Why, what, processing, under the hood)

Importing data into Hive

Importing data into HBase

Exporting data from Hadoop using Sqoop

Sqoop Connectors

Connecting MongoDB (NoSQL database)

12. Ecosystem and its Components

Why Flume and Chukwa

What is Apache Flume

Flume Model

Scalability in Flume

Chukwa

Chukwa Architecture

Chukwa Capability

Chukwa Agent

Introduction to Apache Oozie

Apache Oozie Workflow

Introduction to Mahout

Introduction to YARN

YARN Architecture

Apache Cassandra

Why Apache Cassandra

13. Hadoop Administration and Troubleshooting

Different configurations of Hadoop cluster

Performance monitoring

Performance tuning

For information on the course, visit:<http://www.simplilearn.com/big-data-and-cloud-computing/big-data-and-hadoop-training>

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



LIVE SUPPORT **24/7**



REQUEST A CALL BACK



EMAIL US