



BIG DATA HADOOP

COURSE AGENDA

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.

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. AdvancedMapReduce

Custom Data Types
Input formats

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. Hadoop Administration and Troubleshooting

Different configurations of Hadoop cluster
Performance monitoring
Performance tuning
Troubleshooting and Log observation

13. YARN

Overview
YARN versus MapReduce1

14. Reporting Tool

Overview of various reporting available
Implementing Power Builder

15. Project Work

Hands on end to end Hadoop based live project



16. Technical Discussion

Extensive Interview Preparation & guidance

17. Certification

Cloudera Hadoop Developer certification overview & guidance

18. Hadoop Market

Overview

Cloudera vs Hortonworks

Palasia Centre: 20/1, Morden Mansion, Palasia square, Indore Ph: 0731-

4268656/4066541, 4091658

Bhanwarkuan Centre: 18- Professor Colony, Wadhvani Chambers, Bhanwarkuan square, Indore. Ph 0731-

4289083/84

Email: info@ssiedu.in