



## COURSE AGENDA

### BIG DATA HADOOP - 4 DAYS CLASSROOM TRAINING

MORNING SESSION	CLASSROOM
<b>1. Introduction to Big Data and Hadoop</b> <ul style="list-style-type: none"> <li>• What is Big Data?</li> <li>• Why Big Data?</li> <li>• Limitations of Big Data</li> <li>• Hadoop Background</li> <li>• The Hadoop Way</li> </ul>	Day 1
<b>2. Getting Started With Hadoop</b> <ul style="list-style-type: none"> <li>• Setting up VM Hadoop Environment</li> <li>• Installing VMware Player</li> <li>• Setting up the Virtual Environment (Virtual Machine User Accounts; Running a Hadoop Job; Accessing the VM via ssh; Shutting Down the VM)</li> </ul>	Day 1
<b>3. Hadoop Architecture</b> <ul style="list-style-type: none"> <li>• Hadoop Cluster in commodity hardware</li> <li>• Hadoop core services and components</li> <li>• Regular file system vs. Hadoop</li> <li>• HDFS layer</li> <li>• HDFS operation principle</li> <li>• HDFS 1.0 &amp; HDFS 2.0</li> </ul>	Day 1
EVENING SESSION	
<b>4. Hadoop Deployment</b> <ul style="list-style-type: none"> <li>• Hadoop installation</li> <li>• Single node and multi node configuration</li> <li>• Hadoop Configuration in cluster environment</li> </ul>	Day 1

## MORNING SESSION

### 1. MapReduce

- MapReduce concepts
- Hadoop MapReduce example
- Hadoop MapReduce requirements
- Steps of Hadoop MapReduce
- MapReduce user supply
- MapReduce framework
- Basics of MapReduce programming

Day 2

## EVENING SESSION

### 2. Advanced MapReduce

- Custom Data Types
- Input formats
- Output formats
- Combiners and Partitioners
- Error handling and Unit Testing

Day 2

### 3. PIG

- Introduction to PIG
- Why PIG
- Comparison between PIG and SQL
- Installing and configuring PIG
- Running PIG
- PIG Latin

Day 2

## MORNING SESSION

### 1. 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

Day 3

## EVENING SESSION

### 2. 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

**Day 3**

### 3. 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

**Day 3**

## MORNING SESSION

### 1. 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)

**Day 4**

## 2. 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

Day 4

## EVENING SESSION

### 4. Hadoop Administration and Troubleshooting

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

Day 4

#### Disclaimers:

- 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