

40 hrs



1. System Requirements:-

Hardware Requirements (Minimum)

Core i5, 1.8 GHz, 4 gig RAM, 500 meg disk space.

Hardware Requirements (Recommended)

Core i5, 2.8 GHz, 8 gig RAM, 500 meg disk space.

Operating System Tested on Windows 7/8, Mac OS 10.8.5.

Network and Security Limited privileges required -- please see our standard security requirements.

Software Requirements All free downloadable tools.

2. Course Content:-

Introduction to Hadoop and Big Data:

- What is Big Data?
- What are the challenges for processing big data?
- What technologies support big data?
- What is Hadoop?
- Why Hadoop?
- History of Hadoop
- Use cases of Hadoop
- RDBMS vs Hadoop
- When to use and when not to use Hadoop
- Ecosystem tour
- Vendor comparison
- Hardware Recommendations & Statistics

HDFS: Hadoop Distributed File System:

– Significance of HDFS in Hadoop

• Features of HDFS

• 5 daemons of Hadoop

1. Name Node and its functionality
2. Data Node and its functionality
3. Secondary Name Node and its functionality
4. Job Tracker and its functionality
5. Task Tracker and its functionality

• Data Storage in HDFS

1. Introduction about Blocks
2. Data replication

• Accessing HDFS

1. CLI (Command Line Interface) and admin commands
2. Java Based Approach

• Fault tolerance

• Download Hadoop

• Installation and set-up of Hadoop

1.Start-up & Shut down process

• HDFS Federation

Map Reduce:

• Map Reduce Story

• Map Reduce Architecture

• How Map Reduce works

• Developing Map Reduce

• Map Reduce Programming Model

1. Different phases of Map Reduce Algorithm.
2. Different Data types in Map Reduce.
3. how Write a basic Map Reduce Program.

- Driver Code
- 3Mapper
- Reducer
- Creating Input and Output Formats in Map Reduce Jobs
- 1. Text Input Format
- 2. Key Value Input Format
- 3. Sequence File Input Format
- Data localization in Map Reduce
- Combiner (Mini Reducer) and Partitioner
- Hadoop I/O
- Distributed cache

PIG:

- Introduction to Apache Pig
- Map Reduce Vs. Apache Pig
- SQL vs. Apache Pig
- Different data types in Pig
- Modes of Execution in Pig
- Grunt shell
- Loading data
- Exploring Pig
- Latin commands

HIVE:

- Hive introduction
- Hive architecture
- Hive vs RDBMS
- HiveQL and the shell
- Managing tables (external vs managed)

- Data types and schemas
- Partitions and buckets

Flume

SQOOP