

CETPA INFOTECH PVT. LTD.
CURRICULUM OF BIG DATA HADOOP

Duration: 3 MONTHS

- **Introduction to Big Data**

- What is RDBMS?
- What is Big Data?
- Problems with the RDBMS and other existing systems
- Requirement for the new approach
- Solution to the problem with huge
- Difference between relational databases and NoSQL type databases
- Need of NoSQL type databases
- Problems in processing of Big Data with the traditional systems
- How to process and store Big Data?
- Where to use Hadoop?

- **Hadoop Basic Concepts**

- What is Hadoop?
- Why to use Hadoop?
- Architecture of Hadoop
- Difference between Hadoop 1.x and Hadoop 2.x
- What is YARN?
- Advantage of Hadoop 2.x over Hadoop 1.x
- Use cases for using Hadoop
- Components of Hadoop
- Hadoop Distributed File System (HDFS)
- Map Reduce

- **Hadoop Distributed File System**

- Components of HDFS
- What was the need of HDFS?
- Data Node, Name Node, Secondary name Node
- High Availability and Fault Tolerance
- Command Line interface
- Data Ingestion
- Hadoop Commands

- **Hadoop Cluster**

- Installation of Hadoop
- Understanding the Configuration of Hadoop
- Starting the Hadoop related Processes
- Visualization of Hadoop in UI
- Writing the files to the HDFS
- Reading the files from the Hadoop Cluster
- Work flow of the Job

- **HBASE**

- What is HBASE?
- Why HBASE is needed?
- HBASE Architecture and Schema Design
- Column Oriented and Row Oriented Databases
- HBASE Vs RDBMS

- **Map Reduce Programming**

- Overview of the Map Reduce
- History of Map Reduce
- Flow of Map Reduce
- Working of Map Reduce with simple example
- Difference Between Map phase and Reduce phase
- Concept of Partition and Combiner phase in Map Reduce
- Submission of a Map Reduce job in Hadoop cluster and it's completion
- File support in Hadoop
- Achieving different goals using Map Reduce programs

- **Sqoop**

- What is Sqoop ?
- Use Case for Sqoop?
- Configuring Sqoop
- Importing and Exporting Data using Sqoop
- Importing data into Hive using Sqoop
- Code Generation using sqoop
- Using Map Reduce with the Sqoop


- **PIG**

- Introduction to Apache Pig
- Architecture of Apache Pig
- Why Pig?
- RDBMS Vs Apache PIG
- Loading data using PIG
- Different Modes of execution of PIG Commands
- PIG Vs Map Reduce coding
- Diagnostic operations in Pig
- Combining and Filtering Operations in Pig

- **Flume**

- What is Flume?
- Architecture of Flume
- Why we need Flume?
- Problem with traditional export method
- Configuring Flume
- Different Channels in Flume
- Importing data using Flume
- Using Map Reduce with the Flume

| | |
|--|---|
| <ul style="list-style-type: none"> • HIVE <ul style="list-style-type: none"> ➤ Introduction to HIVE ➤ Architecture of HIVE ➤ Why HIVE? ➤ RDBMS Vs HIVE ➤ Introduction to HiveQL ➤ Loading data using HIVE ➤ HIVE Vs Map Reduce Coding ➤ Different functions supported in HIVE ➤ Partitioning, Bucketing in HIVE ➤ Hive Built-In Operators and Functions ➤ Why do we need Partitioning and Bucketing in HIVE? • MongoDB <ul style="list-style-type: none"> ➤ What is MongoDB? ➤ Difference between MongoDB and RDBMS ➤ Advantages of MongoDB over RDBMS ➤ Installing MongoDB ➤ What are Collections and Documents? ➤ Creating Databases and Collections. ➤ Working with Databases and Collections | <ul style="list-style-type: none"> • Analysis Using R Language <ul style="list-style-type: none"> ➤ Introduction to R Language ➤ Introduction to R Studio ➤ Why to use R? ➤ R Vs Other Languages ➤ Using R to analyze the data extracted using Map Reduce ➤ Introduction to ggplot package ➤ Plotting the graphs of the extracted data from Map Reduce using R • Mini Project to Use Hadoop and Related Technologies on a Dataset |
|--|---|

| | | |
|---|---|---|
| <p>HEAD OFFICE: Ganeshpur,</p> <p>CORPORATE OFFICE:</p> <p>BRANCH OFFICE:</p> <p>BRANCH OFFICE:</p> | <p>200 Purwavali , 2nd Floor, (Opp. Railway Ticket Agency), Railway Road , Roorkee – 247667, Ph.No.: 09219602769, 01332-270218 Fax - 1332 – 274960</p> <p>D-58, Sector-2, Near Red FM. Noida -201301, Uttar Pradesh Contact Us: +91-9212172602 , 0120-4535353</p> <p>401 A, 4th Floor, Lekhraj Khazana, Faizabad Road, Indira Nagar, Lucknow-220616 (U.P.) Ph. No: +91-522-6590802, +91-9258017974</p> <p>105, Mohit Vihar, Near Kamla Palace, GMS Road, Dehradun-248001, UK Contact: +91-9219602771, 0135-6006070 Toll Free- 1800-8333-999 (from any network)</p> |  <p>CETPA[®]</p> <hr/> <p><i>Because Knowledge Matters</i></p> <hr/> <p>ISO 9001:2015 Certified</p> |
|---|---|---|