TechMaster

TechMaster

Bigdata & Hadoop Course Content

Module 1 - INTRODUCTION TO BIG DATA

- ➢ What is Big Data?
- ➢ Examples of Big Data
- ➢ Reasons of Big Data Generation
- ➢ Why Big Data deserves your attention
- ➢ Use cases of Big Data
- ➢ Different options of analyzing Big Data

Module 2 - INTRODUCTION TO HADOOP

- ➢ What is Hadoop,
- ➢ History of Hadoop
- ➢ How Hadoop name was given
- ➢ Problems with Traditional Large-Scale Systems and Need for Hadoop
- ➢ Where Hadoop is being used
- ➢ Understanding distributed systems and Hadoop
- ➢ RDBMS and Hadoop

MODULE 3- STARTING HADOOP

- ➢ Hadoop Architecture
- ➢ Features of Hadoop
- ➢ Hadoop Components- HDFS, Map Reduce
- ➢ Anatomy of File write / read
- ➢ Introduce other components of Hadoop ecosystem

TechMaster

MODULE 4- HDFS

- ➢ HDFS Commands.
- ➢ Single node hadoop cluster
- ➢ Understanding hadoop configuration files
- ➢ Overview Of Hadoop Distributed File System
  - o Name nodes
  - o Data nodes
  - o The Command-Line Interface
- ➢ The building blocks of Hadoop.
- ➢ Running HDFS Commands
- ➢ Web-based cluster UI-Name Node UI, Map Reduce UI
- ➢ Hands-On Exercise: Using HDFS commands

MODULE 5- UNDERSTANDING MAP REDUCE

- ➢ How Map Reduce Works
  - o Data flow in MapReduce
  - o Map operation
  - o Reduce operation
- ➢ Input and Output Formats
- ➢ Partitions
- ➢ Combiners
- ➢ Schedulers
- ➢ MapReduce Program In JAVA using Eclipse
- ➢ Counting words with Hadoop—Running program
- ➢ Writing MapReduce Drivers, Mappers and Reducers in Java
- ➢ Real-world "MapReduce" problems
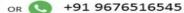 - o Hands-On Exercise: Writing a MapReduce Program and Running a MapReduce Job
- ➢ Java WordCount Code Walkthrough

MODULE 6- HADOOP ECOSYSTEM

- ➢ Hive
- ➢ Sqoop
- ➢ Pig

TechMaster

MODULE 7- EXTENDED SUBJECTS ON HIVE

- Installing Hive
- Introduction to Apache Hive
- Getting data into Hive
- Hive's architecture
- Hive-HQL
- Query execution
- Programming Practices and projects in Hive
- Troubleshooting
- Hands-On Exercise: Hive Programming

MODULE 8- EXTENDED SUBJECTS ON SQOOP

- Installing Sqoop
- Configure Sqoop
- Import RDBMS data to Hive using Sqoop
- Export from to Hive to RDBMS using Sqoop
- Hands-On Exercise: Import data from RDBMS to HDFS and Hive
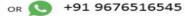- Hands-On Exercise: Export data from HDFS/Hive to RDBM

MODULE 9- EXTENDED SUBJECTS ON PIG

- Introduction to Apache Pig
- Install Pig
- Pig architecture
- Pig Latin - Reading and writing data using Pig
- Hands-On Exercise: Programming with pig, Load data, execute data processing statements

MODULE 10- ADVANCED TOPICS YARN

- Hadoop 2.0, namely, YARN, Name Node High Availability, HDFS Federation, support for Windows etc.

TechMaster

MODULE 11- REAL TIME PROJECT

> ➢ Clear explanation of real time Project by taking real time data
> ➢ Take the data from different source systems like text files, csv files, RDBMS
> ➢ Loading the data in to Hadoop & do some analytics using Map Reduce, HIVE & PIG