VTEAM IT SOLUTION

Duration: 45 days Faculty:Veera Exp: 3+Years

HADOOP Course Content

==

NTRODUCTION 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

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

Software Installation

- Pre requisites
- Understanding decencies software and installing
- Understanding hadoop configuration files
- Setup single node hadoop cluster
- Configuring Hadoop
- The Command-Line Interface
- Understanding hadoop Shell and shell commands

Understanding Hadoop

- Understanding Hadoop Architecture
- Hadoop Components- HDFS, Map Reduce
- Overview of Hadoop Processes
- Overview of Hadoop Distributed File System
- Name Nodes and Data Nodes
- The building blocks of Hadoop

Hive

- Installing Hive

- Running Hadoop
- Web-based cluster UI-Name Node UI, Map Reduce UI
- **Hands-On Exercise:** Using HDFS commands

Understanding MapReduce

- How MapReduce Works
- Data flow in MapReduce
- Map operation
- Reduce operation
- Split
- Distributed Cache
- Record Reader
- Sorter
- Shuffler
- Partitioner
- MapReduce Program in JAVA using Eclipse
- Running your first program

Combiner Operation

- Writing MapReduce Drivers, Mappers and Reducers in Java

Hands-On Exercise:

- Writing a MapReduce Program and Running a MapReduce Job
- Code Walkthrough

Errorhandling in MapReduce.

- Real-world "MapReduce" problems
- MapReduce Job execution In-depth
- Resource allocation
- Yarn-MapReduce

Hadoop Ecosystem

- Hive-Introduction
- Sqoop-Introduction
- Pig-Introduction
- HBase-Introduction
- Flume
- Spark-Introduction
  - Introduction to Apache Hive
  - Getting data into Hive

- Hive architecture

- Hive-HQL
- Query execution
- UDFS in hive
- Partitions in hive
- Buckets in hive

- Compression in hive (ORC)
- SEARDE in hive
- Executing hive queries in real time
- Programming Practices and projects in Hive
- Troubleshooting
- **Hands-On Exercise:** Hive Queries and UDF

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 RDBMS

Pig

- Introduction to Apache Pig
- Install Pig
- Pig architecture
- Pig Latin – Reading and writing data using Pig
- Parameter passing with pig
- UDFS in PIG
- Managing multiple pig scripts in real-time Load data, execute data processing statements.

case
- Executing pig scripts in real-time projects

Hands-On Exercise: Programming with pig, Load data,
execute data processing statements.

HBASE

- What is HBase?
- Install HBase
- HBase Architecture
- Command line interface Exercise
- Programming In HBase
- MapReducePrograms in HBase
- Internals of HBase
- Filters in HBase
- Load and Managing large volumes data sets with HBase
- HBase-Hive integration
- Real time projects in HBase(use cases MapReduce)

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

Note *: **Also we will give you some real time raw data, ask you to load in to Hadoop & cater some requirements which you have to submit as assignment Mandatory to submit.**

Note: We will conduct one TEST at the End of the course bas