VUEMA Technologies

Contact Us : +91 910 0036 789
Email : info@vuematech.com
Visit Us : www.vuematech.com

HADOOP Course Content   Duration: 40 Days

HADOOP Training Course Prerequisites

- Knowledge of Core Java and SQL will be beneficial, but certainly not a mandate.

Section: 1 - Introduction to Big Data & Hadoop

- Big Data
- What is Big Data?
- Importance of Data
- Big data use cases
- Challenges of Big Data.
- Hadoop
- What is Hadoop?
- History of Hadoop.
- Hadoop Architecture
- Databases and Hadoop
- Hadoop Echo system
- Hadoop in real time
- Setting up Hortonworks Hadoop sandbox (single node setup in laptop)

Section: 2 – HDFS (Hadoop Distributed File System)

- Features of HDFS
- HDFS architecture
- About HDFS block
- HDFS components
- Read and write operation
- How HDFS is Fault Tolerance
- HDFS commands
- Hadoop HDFS JAVA API

Section: 3 - MapReduce

- What is MapReduce?
- MapReduce algorithm
- How MapReduce works with Example
- Understanding the MapReduce Job Life Cycle

- Partitioner and Combiners in MapReduce
- Input Output formats in Mapreduce
- MapReduce API
- MapReduce Implementation
- Compression techniques
- Types of Joins

Section: 4 - Apache PIG

- Introduction to Pig
- MR vs Pig
- SQL vs Pig
- Pig Vs Hive
- Data Types in PIG
- Pig Execution Modes
- Pig Latin
- Pig Latin Macro & UDF statements
- Programming with pig, Load data, execute data processing statements

Section: 5 - HIVE

- Introduction to HIVE
- Architecture of Hive
- Data Types
- Hive Query Language
- HQL Vs SQL
- Tables
- UDF in HIVE
- HIVE SerDe
- Querying Data
- Reading and writing in HIVE table using PIG
- HIVE table Export and Import

Section: 6 - SQOOP

- Introduction to SQOOP
- How Sqoop Works?
- SQOOP connectors
- SQOOP commands
- Import
- Export
- Hands on exercise

Section: 7 - HBASE

- About NoSql and Big Table
- Introduction to Hbase
- Storage Mechanism in HBase
- HBase vs RDBMS
- Hbase architecture
- HBase Shell
- Bulk load in Hbase table
- Hbase Java API

Section: 8 - OOZIE

- Introduction to OOZIE
- OOZIE architecture
- Workflow
- Property file
- Coordinator
- Action types
- Oozie CLI
- OOZIE workflow example
- Running OOZIE workflow

Section: 9 - Apache Flume

- Introduction to Flume
- Flume architecture
- Flume Data Flow
- Flume Configuration
- Fetching Twitter Data

Section: 10 -Apache KAFKA

- Introduction to KAFKA
- Kafka Architecture: Topics, Producers and Consumers
- Kafka and Avro
- Kafka and Schema Registry
- Kafka Tools
- Simple producer/consumer application

Section: 11 - YARN

- Introduction to YARN
- MapReduce vs Yarn
- YARN architecture
- Benefits of YARN

Section: 12 - ZOOKEEPER

- Introduction to Zookeeper
- Architecture of Zookeeper
- Zookeeper workflow
- Zookeeper CLI
- Zookeeper Applications

Section: 13 - Apache SPARK with SCALA

- SCALA
- Introduction to SCALA
- Scala Basics
- Data types
- Variables
- Writing Scala scripts
- Defining functions
- Functional Programming in SCALA
- SCALA environment setup
- SPARK
- Introduction to SPARK
- Spark core
- Spark SQL
- Real time use case

Section:14 - Multinode Cluster setup

Section: 15 - Working session on real time projects.

- Work session on at least two real time projects