Big Data Hadoop Course Contents

Total Duration: 60 Hours

Big data Concepts
- Distributed network and computation
- Challenges in data management and control
- Introduction to Big Data
- Types of data in detail
- Sources of Big Data
- Concept of Streaming data
- Batch and Streaming data processing
- Big data Hadoop future opportunities

Hadoop Overview
- Need of Hadoop technology
- Overview of Data centers and Cluster
- Hadoop Cluster and Racks in detail
- Learning Ubuntu for Hadoop
- Overview of Hadoop tools
- Overview of Map Reduce
- Understanding the Hadoop Installation and Configuration
- 5 daemons of Hadoop
- Name Node and its functionality
- Data Node and its functionality
- Secondary Name Node and its functionality
- Job Tracker and its functionality
- Task Tracker and its functionality

HDFS
- HDFS Daemons
- Introduction about Blocks
- Data replication
- Hadoop DFS and Processing
- Fault Tolerance in Hadoop
- Files operations in Hadoop
- FS shell commands in use

YARN (Yet Another Resource Negotiator)
- Introduction to YARN
- YARN Daemons
- Job assignment and Execution flow in Yarn
- Map Reduce Programming Model
- Word count program demonstration

Apache Pig

- Introduction to Apache Pig
- Apache pig Architecture
- Advantage of Pig over MapReduce programming
- Pig Latin and Grunt Shell
- Pig Latin basics
- Operators in Pig
- Group, Join
- Split and Combine
- Built in functions
- Load and Store function
- Date, Math and String functions
- Schema and Schema-less data in Pig
- Data processing in Pig
- Pig UDFs writing
- HCatalog
- Pig and Hive comparison

Apache Hive

- Data warehouse basics
- OLTP and OLAP Concepts
- Introduction to Hive
- Hive Architecture in detail
- Metastore DB and Metastore Service
- Hive Query Language (HQL)
- Types of table in hive
- Partitioning and Bucketing
- Built in Functions
- Case writing
- Views and Indexes
- Joins and Group by
- Sorting, Distribute by
- Query Optimization methods
- JDBC , ODBC connection to Hive
- Hive Transactions
- Hive UDFs and UDAFs
- Working with file formats

Sqoop

- Sqoop basic commands
- Sqoop practical implementation
- Sqoop Architecture
- Database operations

- Importing RDBMS data to HDFS
- Importing RDBMS data to Hive
- Exporting data to RDBMS
- Sqoop connectors

Flume
- Flume Architecture
- Flume Environment
- Data transfer and data flow
- Flume Configuration
- Configuration of Source, Channel and Sink
- Loading from web server or other storage data
- Loading from raw/flow data in HDFS using flume

Oozie
- Introduction to Oozie
- Designing workflow jobs
- Job scheduling using Oozie
- Time based job scheduling
- Oozie Conf file
- Crontab use and implementation

Hands on HUE (Hadoop User Interface) and Impala
- HUE usage
- User management
- Using Pig, Hive, Impala

Impala
- Impala overview
- Impala Architecture
- Impala Vs Hive
- Impala Built in functions
- Connecting to impala
- Java for impala JDBC
- Output creation

HBASE Basics
- Architecture and schema design
- HBase vs. RDBMS
- HMaster and Region Servers
- Column Families and Regions
- Write pipeline
- Read pipeline
- HBase commands

Domain Based Project With Real Time Data of E-Commerce, Banking domain
POC's
Assignment's

==========

Big Data Hadoop Workshop

Total Duration: 16 Hours (2 Days)

DAY-1

Big data Concepts
- Introduction to Big Data
- Types of data in detail
- Sources of Big Data
- Big data Hadoop future opportunities

Hadoop Overview
- Need of Hadoop technology
- Overview of Hadoop tools
- Overview of Map Reduce
- Understanding the Hadoop Installation and Configuration
- 5 daemons of Hadoop

HDFS
- HDFS Daemons
- Data replication
- FS shell commands in use

YARN (Yet Another Resource Negotiator)
- Introduction to YARN
- YARN Daemons
- Map Reduce Programming Model

Apache Pig
- Introduction to Apache Pig
- Apache pig Architecture
- Pig Latin basics
- Operators in Pig
- Group, Join
- Split and Combine
- Schema and Schema-less data in Pig

Data processing in Pig

DAY-2

Apache Hive

- Data warehouse basics
- OLTP and OLAP Concepts
- Hive Architecture in detail
- Hive Query Language (HQL)
- Types of table in hive
- Partitioning and Bucketing
- Joins and Group by
- Sorting, Distribute by

Sqoop

- Sqoop basic commands
- Sqoop Architecture
- Importing RDBMS data to HDFS
- Importing RDBMS data to Hive

Flume

- Flume Architecture
- Configuration of Source, Channel and Sink
- Loading from web server or other storage data
- Loading from raw/flow data in HDFS using flume

Crontab use and job scheduling Linux/Ubuntu

Impala

- Impala Architecture
- Impala Vs Hive
- Impala Built in functions

HBASE Basics

- Architecture and schema design
- HBase vs. RDBMS
- HMaster and Region Servers
- Column Families and Regions

POC's

Assignment's

==========