Big Data - Hadoop Developer Course

Course Modules

Basic Primitives of UNIX and SQL:

- Origin of Unix, Overview of System Administration
- Unix basic commands
- Working with editors – vi and sed
- Linking files with symbolic link and hard link
- Understanding job processes.
- Killing processes in different ways.

Big Data

- Understanding Data & Hadoop: Basic Concepts
- What is BigData
- Characteristics of BigData
- Challenges with Traditional Systems
- Problems with BigData
- Handling BigData

HADOOP Core Concepts

- Problems with Existing Distributed Systems to deal Big Data
- Why Hadoop and An Overview and History of Hadoop
- Requirements of New Approach
- The Hadoop Project and Hadoop Components

Hadoop Distributed File System – HDFS

- What is HDFS, Why it is required for running Map-Reduce.
- How it differs from other distributed file systems.
- Design of HDFS & Concepts
- Command Line Interface, Hadoop File Systems, Java Interface
- Anatomy of a File Read, Anatomy of a File Write
- Hadoop Archives.
- Hands-on Exercise on HDFS

Understanding - Map-Reduce Basics and Map-Reduce Types and Formats

- Describe Map-Reduce framework works & Flow
- Why Map-Reduce is tightly coupled with HDFS.
- What are the different types of Input and Output formats and why they are required?
- Architecture of map reduce framework
- Hadoop Data Types
- Concept of Mappers & Reducers
- Concept of Partitioners & Combiners
- MapReduce Join types
- Secondary sort
- Input Formats (Input Splits and Records, Text Input, SequentialFile Input, Multiple Inputs, KeyvalueTextInput, Database Input and Output)
- Output Formats (TextOutput, BinaryOutPut, Multiple Outputs, Databaseoutput).
- Hands-on Exercise

Developing Map Reduce Programs

- Setting up Eclipse Development Environment.
- Eclipse integration with HADOOP for Rapid Application Development
- Understanding HADOOP API
- Creating Map Reduce Projects,
- Writing MapReduce Drivers, Mappers and Reducers in Java
- Driver Code
- Mapper Code
- Reducer Code
- Map Reduce Code
- Differences Between the Old and New MapReduce APIs
- Hands-on Exercise

Understanding ToolRunner

- More about ToolRunner
- Combiner
- Reducer
- Configure and close methods
- Hands-on Exercise

Hadoop Eco-Components

Flume

- Extract records from the third party resources using Hadoop API
- Process the records using Sentiment analysis
- Integrating with netcat server and pulling out the data

Sqoop

- Importing data to and from RDBMS to Hadoop.
- Exporting data to RDBMS
- Creating sqoop job and implementing updates.

Hive

- Getting Data into Hive
- Manipulating Data with Hive
- Partitioning and Bucketing Data

Assessments will happen for every week to analyse the understanding of the students on the ongoing topics.

Once all the workarounds and assessments are done, the students will be given five real time projects based on what they have learned in the course.