

## **Hadoop Development:- 25 days – 25 hrs**

### 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 analysing 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
- Understanding Hadoop Architecture
- Fundamental of HDFS (Blocks, Name Node, Data Node, Secondary Name Node)
- Rack Awareness
- Read/Write from HDFS
- HDFS Federation and High Availability

### Module 3- Starting Hadoop

- Setting up single node Hadoop cluster(Pseudo mode)
- Understanding Hadoop configuration files
- Hadoop Components- HDFS, MapReduce
- Overview Of Hadoop Processes
- Overview Of Hadoop Distributed File System
- The building blocks of Hadoop
- Hands-On Exercise: Using HDFS commands

### Module 4- MapReduce-1(MR V1)

- Understanding Map Reduce
- Job Tracker and Task Tracker
- Architecture of Map Reduce
- Map Function

- Reduce Function
- Data Flow of Map Reduce
- How Map Reduce Works
- Anatomy of Map Reduce Job (MR-1)
- Submission & Initialization of Map Reduce Job
- Assigning & Execution of Tasks
- Monitoring & Progress of Map Reduce Job
- Hadoop Writable and Comparable

#### Module 5- Hive

- Introduction to Apache Hive
- Architecture of Hive
- Installing Hive
- Hive data types
- Hive-HQL
- Types of Tables in Hive
- Partitions
- Buckets & Sampling
- Indexes
- Views
- Executing hive queries from Linux terminal
- Executing hive queries from a file
- Creating UDFs in HIVE
- Hands-On Exercise

#### Module 6- Pig

- Introduction to Apache Pig
- Install Pig
- Architecture
- Data types
- Working with various PIG Commands covering all the functions in PIG
- Working with un-structured data
- Working with Semi-structured data

- Creating UDFs
- Hands-On Exercise

#### Module 7-Sqoop

- Introduction to SQOOP& Architecture
- Installation of SQOOP
- Import data from RDBMS to HDFS
- Importing Data from RDBMS to HIVE
- Exporting data from HIVE to RDBMS
- Hands on exercise

#### Module 8-HBASE

- Introduction to HBASE
- Installation of HBase
- Exploring HBASE Master & Region server
- Exploring Zookeeper
- CRUD Operation of HBase with Examples
- HIVE integration with HBASE
- Hands on exercise on HBASE