**Hadoop** **Basics**

**Duration:** 3 days

**Pre-Requisites:** OOPS Concept

**Course Contents:**

* **Hadoop Introduction**
  + Why Hadoop?
  + What is hadoop?
  + Brief History
  + Architecture of Hadoop and eco system
* **Hadoop Installation**
* **HDFS commands**
  + HDFS commands
  + Blocks, input split
* **Mapreduce Introduction**
  + Input Formats
  + MapReduce Flow and program
* **MapReduce programs**
  + Mapreduce program
  + More scenarios/programs on Mapreduce
* **Mapreduce phases**
  + Mapper
  + partitioner
  + sorting
  + shuffling
  + Reducer
  + combiner
* **MapReduce unstructure programs**
* **MapReduce Image programs**
* **Pig**
  + Pig Commands
  + Pig Programs
* **Hive**
  + Hive queries
  + joib vs mapjoin
  + Distributed cache
* **Impala**
  + Impala introduction
* **Sqoop**
  + sqoop data transfer b/n traditional and Hadoop
* **Flume**
  + Flume flow and transfer tweets to HDFS
* **Spark**
  + Few programs in Spark (RDD)
* **Entire Hadoop Flow**

**Lab Setup:**

* System with 4GB RAM
* Linux Operation System
* VMware Workstation should be installed