Two Days Certified Workshop on Big Data: HADOOP

Pre-requisites:

- Internet Connectivity
- Systems with Linux Operating system or Virtual machine

Day 1

Session 1-Understanding Big Data and Hadoop

- How Big is this Big Data ?
- Definition with Real Time Examples
- How BigData is generated with Real Time Generation
- Use of BigData-How Industry is utilizing BigData
- Traditional Data Processing Technologies
- Future of BigData!!!
- Why Hadoop?
- What is Hadoop?
- Hadoop vs RDBMS, Hadoop vs BigData
- Brief history of Hadoop

Session 2- Hadoop Cluster Setup and Commands

- Anatomy of a Hadoop cluster
- Hadoop Setup and Installation
- Implementation of Distributed Storage : HDFS
- Hadoop Commands

- Run MapReduce Program

- Hadoop 1 vs Hadoop 2

- YARN Concepts

Day 2

Session 1- MapReduce Programming

- Understanding MapReduce Programming

- Develop Basic MapReduce Program for processing Real World data sets from our OWN develop Cluster

- How to Get Real data sets and work our cluster-WordCount Example

Session 2- Hadoop Ecosystem Tools

- Overview of Hbase,

- Pig and Pig Latin

- Hive and HiveQL

- Sqoop, Oozie

- HDFS Federation,

- NameNode HA