Module I. Basic Understanding

➢ Understanding Big Data Basics
➢ Big Data Use Cases
➢ Introduction to Hadoop
➢ Understanding Hadoop Ecosystem
➢ Introduction to HDFS Introduction to Namenode Introduction to Datanode Introduction to Secondary Namenode
➢ Introduction to MapReduce Introduction to JobTracker Introduction to TaskTracker
➢ Summarizing Hadoop Architecture h. Roles and Responsibilities of a Hadoop Administrator
➢ Roles and Responsibilities of a Hadoop Administrator

Module II. Introduction To Hadoop

➢ Linux internals Commands that are required Linux basics
➢ Hadoop Cluster Installation Pre-requisites Pre-requisites of Hadoop Installation
  o Software's Download
  o Preparing your VM
  o Enabling VM with VMware
  o Understanding mandatory changes in the operating system

➢ Installation and Configuration Understanding Hadoop cluster installation modes Understanding Hadoop version 1 installation and configuration Passwordless SSH setup

Module III.

➢ Hands-On Practice for creating a Hadoop cluster Helping individually in practicing Hadoop cluster installation Job Tracker, Task Tracker and its functionality

Module IV. Hadoop 14

➢ Hadoop Cluster Planning Recommended Hadoop cluster configuration
➢ Hardware/Software/Network
➢ Recommended configuration for Master and Slave Nodes
➢ Sample Base configuration
➢ Hadoop Different Distributions in the market b. Hadoop performance tuning Important Hadoop tuning parameters to understand Hadoop Cluster Benchmarking Jobs – How to run the jobs

Module V.

➢ Job Schedulers FIFO Scheduler Fair Scheduler
➢ Backup and Recovery Data backup Meta-data backup Hadoop Quotas Safemode Hadoop Ports
➢ DistCP
➢ Security How to secure your cluster using Kerberos
➢ Upgrades Upgrading Hadoop cluster from Hadoop 1 to Hadoop 2

Module VI. Understanding Hadoop Internals

- ➢ Hadoop 2.0 new features
- ➢ YARN Understanding Resource Manager Understanding Application Master Understanding Node Manager Understanding Hadoop
- ➢ Job Execution Framework
- ➢ Hadoop 2 Multi-node cluster creation
  - o Pre-requisites of Hadoop Installation
  - o Software's Download
  - o Preparing your VM
  - o Enabling VM with VMware
  - o Understanding mandatory changes in the operating system
  - o Installation and Configuration
  - o Understanding Hadoop version 2 installation and configuration
  - o Password less SSH setup

Module VII. Hadoop 2

- ➢ Practice Hadoop 2 multi-node Cluster Creation Helping individuals in practicing Hadoop
- ➢ cluster installation
- ➢ Sample Yarn Job execution
- ➢ Understanding Issues of Hadoop 1
- ➢ Understanding improvements in Hadoop 2
- ➢ Namenode Federation Enable segregation of HDFS using multiple namenodes d. Namenode – High Availability Achieving Namenode High-Availability using Quorum Journal Manager Achieving Namenode High-Availability using Network File System
- ➢ Implementation of NN High Availability Helping individuals achieving Namenode High Availability
- ➢ Hadoop Ecosystem Introduction Understanding the integration of Hadoop ecosystem
- ➢ Touchbase with Hive What is Hive Architecture of Hive Understanding Hive metastore concepts
- ➢ HBase Understading HBase Basics Understanding HBase storage Model Understanding HBase Architecture Cluster Installation and Configuration
- ➢ Pig What is Pig? How Pig integrates with Hadoop cluster? Demo of Pig Jobs using MapReduce
- ➢ Sqoop What is Sqoop? How to import and export the data from Sqoop to RDBMS? Example of Sqoop jobs using MySQL
- ➢ Flume what is Flume? Sample Flume jobs

Module VIII. Hadoop Ecosystem

- ➢ Understanding the internals of Cloudera Manager

- Understanding the automation of Hadoop installation using Cloudera Manager
- Understanding Cloudera Hadoop Distribution and Cloudera Manager
- Understanding the underlying directory structure of Cloudera Hadoop
- Cloudera Hadoop Cluster Installation – CDH