**Prerequisites:**

* Two Linux Machines (at least)
* Hadoop Binary Distribution software
* Passwordless ssh
* Java Platform
* Little bit knowledge on Java language

The following course outline for Hadoop Core Project only.

**Course Outline:**

**The Motivation for Hadoop**

* Big data
* Problems with traditional large-scale systems
* Requirements for a new approach

**Introduction:**

* An Overview of Hadoop
* The Hadoop Distributed File System
* MapReduce Programming model
* Hadoop common utilities
* Hadoop Ecosystem Components

**Hadoop Cluster Setup**

* Configuration details
* Local mode
* Pseudo distributed mode
* Distributed mode

**Components of Hadoop**

* Hadoop Distributed File System
* MapReduce Programming model
* Hadoop common utilities

**Hadoop Architecture**

* Master/Slave
* Name node(NN)
* Data node(DN)
* Job Tracker(JT)
* Task Tracker(TT)
* SecondaryNamenode(SNN)

**MapReduce Program**

* The MapReduce Flow
* Sample MapReduce Program
* Basic MapReduce API Concepts
* The Mapper
* The Reducer
* The Shuffle
* Hadoop Data types
* Hadoop Streaming API
* Some MapReduce Program Examples

**Common MapReduce Algorithms**

* Sorting and Searching
* Indexing
* Machine Learning With Mahout
* Term Frequency – Inverse Document Frequency
* Word Co-Occurrence

**Advanced MapReduce Programming**

* Custom Writables and WritableComparables
* Saving Binary Data using SequenceFiles and Avro Files
* Creating InputFormats and OutputFormats

**Hadoop Ecosystem**

* Pig
* Hive
* HBase
* Sqoop
* Zookeeper
* Cassandra
* Mahout
* Chukwa
* Avro
* Flume
* Others