**Hadoop Administration Online Training**

**Course Duration: 45 Hrs**

**1.Hadoop Admin**

* How the Hadoop Distributed File System and Map Reduce work
* What hardware configurations are optimal for Hadoop clusters
* How to configure Hadoop's options for best cluster performance
* How to configure NameNode High Availability
* How to configure NameNode Federation
* How to configure the FairScheduler to provide service-level agreements for multiple users of a cluster
* How to install and implement Kerberos-based security for your cluster
* What system administration issues exist with other Hadoop projects such as Hive, Pig, and HBase.

**2. Introduction to Big Data**

* Characteristics of Big Data
* Why is parallel computing important
* Discuss various products developed by vendor

**3. Introducing Hadoop**

* Components of Hadoop
* Starting Hadoop
* Identify various processes

**4. Working with HDFS**

* Basic file commands
* Web Based User Interface
* Reading & Writing to files
* Run a word count program
* View jobs in the Web UI

**5. Installation & Configuration of Hadoop**

* Types of installation (RPM’s & Tar files)
* Set up ‘ssh’ for the Hadoop cluster
* Tree structure
* XML, masters and slaves files
* Checking system health
* Discuss block size and replication factor
* Benchmarking the cluster

**6. Advanced administration activities**

* Adding and de-commissioning nodes
* Purpose of secondary name node
* Recovery from a failed name node
* Managing quotas
* Enabling trash

**7. Monitoring the Hadoop Cluster**

* Hadoop infrastructure monitoring
* Hadoop specific monitoring
* Install and configure Nagios / Ganglia
* Capture metrics

**8. Other Components of the Hadoop ecosystem**

* Discuss Hive, Sqoop, Pig, HBase, Flume
* Use cases of each
* Use Hadoop streaming to write code in Perl / Python
* Hands on Real time Tasks Explanation.

What is Hadoop in Big Data?

Hadoop is an open-source framework for distributed storage and processing of large data sets. It uses the Hadoop Distributed File System (HDFS) and MapReduce programming model to enable scalable and parallel data processing.

Introduction to Hadoop.

This introductory level training on Big Data Hadoop gives an overview of the MapReduce programming model with the use of a simple word counting mechanism with existing tools which highlights the challenges around processing data at a big scale. Thus, this will help you learn the use of Hadoop in different real-life situations to gain a deeper appreciation of its simplicity. Following are some other important details for you to know before applying for the program.