ABOUT BIG DATA & HADOOP PROGRAM

Apache Hadoop is an open-source software framework used for distributed storage and processing of dataset of big data using the MapReduce programming model.

It consists of computer clusters built from commodity hardware. All the modules in Hadoop are designed with a fundamental assumption that hardware failures are common occurrences and should be automatically handled by the framework.The core part of Hadoop consist of a Storage part, known as Hadoop Distributed File system(HDFS),and a processing part which is a mapreduce programming model.

Apache Hadoop clusters which includes support for Hadoop HDFS, Hadoop MapReduce, Hive,

HBase, ZooKeeper, Oozie, Pig, and Sqoop

Career Opportunities:

Companies like Google, EMC Corporation, Yahoo, Apple, HortonWorks, Oracle, Amazon,

Cloudera, IBM, Cisco, Microsoft and many more have opened their doors for

hadoop professionals. Various positions like product managers, hadoop developers, testers,

database administrators, senior hadoop developers and alike are open.

Companies are searching for experienced candidates as well as freshers.

Course Structure and Pre-Requisites

1.Software Developers/Programmers/Engineers,who are into Database/Programming and

exploring for great job opportunities in Hadoop

2.Managers, who are looking for the latest technologies to be implemented in their organization,to meet the current and upcoming challenges of data management

3.Any Graduate/Post-Graduate, who is aspiring a great career towards the cutting edge

technologies.

Topics and Structure PRO package:

Architecture of Hadoop and relates how HDFS carry data storage and YARN,the other core constituent of Hadoop, gives the data processing capacity.

Hadoop administration comprises how to form and configure a single node, multinode homogenous clusters.

Data replication is the important part of HDFS, and we will practice about how HDFS arrange its data and how data replication works.

As an Hadoop administrator,important job is to arrange HDFS storage

1)Hadoop Administration:

-Introduction to Big Data and Hadoop

-Types Of Data

-Characteristics Of Big Data

-Hadoop And Traditional Rdbms

-Hadoop Core Services

2)Hadoop single node Cluster Setup for 1X series :

-Hadoop single node cluster(HADOOP-1.2.1)

-Tools installation for hadoop1x.

-Sqoop,Hive,Pig,Hbase,Zookeeper.

-Analyze the cluster using

a)NameNode UI

b)JobTracker UI

-SettingUp Replication Factor

3)Hadoop Distributed File System:

-Introduction to Hadoop Distributed File System

-Goals of HDFS

-HDFS Architecture

-Design of HDFS

-Hadoop Storage Mechanism

-Measures of Capacity Execution

-HDFS Commands

4)The MapReduce Framework:

-Understanding MapReduce

-The Map and Reduce Phase

-WordCount in MapReduce

-Running MapReduce Job

5)Hadoop single node Cluster Setup :

-Hadoop single node cluster(HADOOP-2.7.3)

-Tools installation for hadoop2x

-Sqoop,Hive,Pig,Hbase,Zookeeper.

6)YARN

-Introduction to YARN

-Need for YARN

-YARN Architecture

-YARN Installation and Configuration

7)Hadoop Multinode cluster setup:

-hadoop multinode cluster

-Checking HDFS Status

-Breaking the cluster

-Copying Data Between Clusters

-Adding and Removing Cluster Nodes

-Name Node Metadata Backup

-Cluster Upgrading

7) Hadoop ecosystem:

-Sqoop

-Hive

-Pig

-HBase

-zookeeper