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:

A)RDBMS Vs Hadoop:

1)difference in between Mysql and Hadoop

2)Why Hadoop is better that Mysql??

B)Introduction to Java:

1)Basics of Java required for Hadoop i.e.Core Java

C)Introduction to HDFS(Storage) & Understanding cluster environment:

1)NameNode and DataNodes

2)HDFS has a master/slave architecture

3)Overview of Hadoop Daemons

4)Hadoop FS and Processing Environment's UIs

5) Block Replication

6)How to read and write files

7)Hadoop FS shell commands

D)Understanding Map-Reduce Basics, Types & Formats:

1)The introduction of MapReduce.

2)MapReduce Architecture

3)Data flow in MapReduce

4)Role of RecordReader

5)Basic Configuration of MapReduce

6)How MapReduce Works

7)Writing and Executing the Basic MapReduce Program using Java

8)File Input/Output Formats in MapReduce Jobs

Text Input Format

Key Value Input Format

Sequence File Input Format

NLine Input Format

9)Word Count Example

10)Will cover five to Ten Map Reduce Examples with real time data.

E)TOOLS:

1)SQOOP:

-sqoop commands

-Sqoop practical implementation

-Importing data to HDFS

-Importing data to Hive

-Exporting data to RDBMS

-Sqoop connectors

2)HIVE:

-Hive Architecture

-Hive Query Language (HQL)

-Managed and External Tables

-Partitioning & Bucketing

-Query Optimization

-JDBC , ODBC connection to Hive

-Hands on Multiple Real Time datasets.

3)PIG:

-Pig Latin (Scripting language for Pig)

-Schema and Schema-less data in Pig

-Structured , Semi-Structure data processing in Pig

-piglatin builtin functions

4)HBASE:

-Introduction to HBASE

-Basic Configurations of HBASE

-Fundamentals of HBase

-What is NoSQL?

-HBase Data Model

-Table and Row.

-Column Family and Column Qualifier.

-Cell and its Versioning

-Get

-Scan -Put

5)ZOOKEEPER:

-configuration information

-naming

- providing distributed synchronization

- providing group services.

6)Basics of Spark:

-overview of spark modules(Spark core,Spark SQL,Spark MlLib,Spark Graphx)