BIG DATA TECHNOLOGY

Course Objective

The Objective of this course is to provide practical foundation level training that enables immediate and effective participation in big data projects. The course provides grounding in basic and advanced methods to big data technology and tools, including MapReduce and Hadoop and its ecosystem.

Course Outcome

1. Learn tips and tricks for Big Data use cases and solutions.

2. Learn to build and maintain reliable, scalable, distributed systems with Apache Hadoop.

3. Able to apply Hadoop ecosystem components.

UNIT I

INTRODUCTION TO BIG DATA: Introduction – distributed file system – Big Data and its importance, Four V's in bigdata, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce.

UNIT II

 INTRODUCTION HADOOP : Big Data – Apache Hadoop & Hadoop EcoSystem – Moving Data in and out of Hadoop – Understanding inputs and outputs of MapReduce - Data Serialization.

UNIT- III

HADOOP ARCHITECTURE: Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands , Anatomy of File Write and Read., NameNode, Secondary NameNode, and DataNode, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Tasktrackers - Cluster Setup – SSH & Hadoop Configuration – HDFS Administering –Monitoring & Maintenance.

UNIT-IV

HADOOP ECOSYSTEM AND YARN : Hadoop ecosystem components - Schedulers - Fair and Capacity, Hadoop 2.0 New Features- NameNode High Availability, HDFS Federation, MRv2, YARN, Running MRv1 in YARN.

UNIT-V

HIVE AND HIVEQL, HBASE:-Hive Architecture and Installation, Comparison with Traditional Database, HiveQL - Querying Data - Sorting And Aggregating, Map Reduce Scripts, Joins &

Subqueries, HBase concepts- Advanced Usage, Schema Design, Advance Indexing - PIG, Zookeeper - how it helps in monitoring a cluster, HBase uses Zookeeper and how to Build Applications with Zookeeper.

Reference Books

1. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, "Professional Hadoop Solutions", Wiley, ISBN: 9788126551071, 2015.
2. Chris Eaton, Dirk deroos et al. , "Understanding Big data ", McGraw Hill, 2012.
3. Tom White, "HADOOP: The definitive Guide" , O Reilly 2012.
4. Vignesh Prajapati, "Big Data Analytics with R and Haoop", Packet Publishing 2013.
5. Tom Plunkett, Brian Macdonald et al, "Oracle Big Data Handbook", Oracle Press, 2014.
6. Jy Liebowitz, "Big Data and Business analytics",CRC press, 2013.

Student Activity:

1. Collect real time data and justify how it has become Big Data

2. Reduce the dimensionality of a big data using your own map reducer