

Subject Code	Subject Name	Credits
26CS110	BIG DATA TECHNOLOGY	4

Course Objectives:

1. To Introduce the field of Big Data and get a basic understanding of it
2. To know and learn the Hadoop Ecosystem and its Architecture

Learning Objectives:

On successful completion of the course, students will be able to:

1. To learn the advanced concepts of Big Data Technology

Unit 1 - Introduction to Big Data (12 hrs.):

Introduction – distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce.

Unit 2 - Introduction to Hadoop (12 hrs.):

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

Unit 3 - Hadoop Architecture (12 hrs.):

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, Task trackers - Cluster Setup – SSH & Hadoop Configuration – HDFS Administering –Monitoring & Maintenance.

Unit 4 - Hadoop Ecosystem and Yarn (12 hrs.):

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

Unit 5 - Hive And HiveQL, HBase (12 hrs.):

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: 12788126551071, 2015.
2. Chris Eaton, Dirk derooset al. , “Understanding Big data ”, McGraw Hill, 2012.
3. Tom White, “HADOOP: The definitive Guide” , O Reilly 2012. 6 IT2015 SRM(E&T)
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.

