

MARIS STELLA COLLEGE, VIJAYAWADA-8

(An autonomous college affiliated to Krishna University)

Department of Computer Science

Academic Year 2019 - 2020

Title : Big Data Technology

Semester : VI

Paper Code: COMPC099

Course Objectives:

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 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:

8 hours

Introduction to big data: What is Big Data, Structuring Big Data -Types of Big Data, Elements of big data- Volume, Velocity, Variety, Veracity, Big Data Analytics- Advantages of Big Data Analytics, Big Data Applications.

UNIT-II:

8 hours

Introduction to Hadoop: What is Hadoop, Understanding distributed systems &Hadoop, Comparing SQL databases and Hadoop, Understanding Map Reduce-scaling word count program manually, scaling word count program in Map reduce.

UNIT-III**8 hours**

Hadoop Eco System, HDFS-HDFS Architecture, concept of blocks in HDFS-namenode, datanode, secondary namenode, job tracker, task tracker). Introducing HBase-HBase architecture, Regions, storing Big Data with HBase, Why hive, pig, scoop, zookeeper, flume, oozie.

UNIT-IV**8 hours**

Working with files in HDFS-Basic file commands, reading & writing to HDFS programmatically, Anatomy of Map Reduce program-Hadoop data types, Mapper, Reducer, Partitioner, Combiner, word counting with pre-defined mapper and reducer, Reading & Writing-input format, output format.

UNIT-V:**8 hours**

Background of YARN, limitations of map reduce, advantages of YARN, YARN architecture, working of YARN. Introducing Hive, Hive Services, Hive Variables, Hive Queries, Data types, Hive Built in functions, Hive - DDL, DML, and Data Retrieval Queries.

Prescribed Books:

1. BIG DATA (covers hadoop2, map reduce, Hive, Yarn, Pig, R and Data Visualization) Black Book ,DreamTech Press.
2. Hadoop in Action by Chuck Lam, DreamTech Press.

MARIS STELLA COLLEGE, VIJAYAWADA-8

(An autonomous college affiliated to Krishna University)

Department of Computer Science

Practical : VII

Semester : VI

Title : Big Data Technology Lab

Paper Code: COMPC100

BIG DATA TECHNOLOGY LAB CYCLE

- 1. Introduction to Hadoop.**
- 2. Installation of Hadoop in 3 modes.**
- 3. Hadoop echo System.**
- 4. HDFS Architecture.**
- 5. Introduction to Hbase.**
- 6. Hbase Architecture.**
- 7. Create employee table in Hbase.**
- 8. HDFS Commands.**
- 9. Introduction to Hive.**
- 10. Installation of Hive.**
- 11. Building Functions in HIVE.**
- 12. Aggregate functions in HIVE.**
- 13. HIVE DDL Commands.**
- 14. HIVE DML Commands.**
- 15. Hive Data Retrieval Commands.**
- 16. Yarn Architecture.**