

**Prasad V. Potluri Siddhartha Institute of Technology:: Vijayawada.
Department of Computer Science and Engineering**

I/II M.Tech. (CSE) (Second Semester)

17CSCS2L2

BIG DATA ANALYTICS LAB

Credits: 2

Lecture: 3 Periods/week

**Internal Assessment: 25 Marks
Semester end examination: 50 Marks**

Course Description:

This course prepares students to explore the capabilities and challenges of data-driven business decision making. The course will include hands-on work with Hadoop Framework and MapReduce Algorithms for data analytics.

Course Outcomes:

At the end of the course, students should be able to:

- CO1:** Install and run Hadoop in standalone mode, pseudo mode and fully distributed cluster environment.
- CO2:** Develop Hadoop MapReduce algorithms
- CO3:** Calculate basic analytics using Hadoop and MapReduce

Getting Hadoop Up and Running in a cluster:

1. Setting up Hadoop on standalone machine.
2. WordCount Map Reduce program using standalone Hadoop.
3. Adding the combiner step to the WordCount Map Reduce program.
4. Setting up HDFS.
5. Using HDFS monitoring UI
6. HDFS basic command-line file operations.
7. Setting Hadoop in a distributed cluster environment.
8. Running the WordCount program in a distributed cluster environment.
9. Using Map Reduce monitoring UI

Hadoop Map-Reduce Applications:

10. Choosing appropriate Hadoop data types.
11. Implementing a custom Hadoop Writable data type.

Analytics

12. Simple analytics using Map Reduce.
13. Performing Group-By using Map Reduce.
14. Calculating frequency distributions and sorting using Map Reduce.

Text Book:

1. Srinath Perera & Thilina Gunarathne, "Hadoop Map Reduce Cookbook", 2013, PACKT PUBLISHING.