IV/IV B. TECH. FIRST SEMESTER
DATA ANALYTICS LAB(Required)

Course Code: CS7L1 **Credits: 2**
Lecture:-- **Internal assessment: 25 Marks**
Lab: 3period/week **Semester end examination: 50 Marks**

Prerequisites: Big Data Concepts

At the end of this course student will:

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.

Syllabus:

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:

1. Choosing appropriate Hadoop data types.
2. Implementing a custom Hadoop Writable data type.
3. Implementing a custom Hadoop key type.
4. Emitting data of different value types from a mapper.
5. Choosing a suitable Hadoop Input Format for your input data format.
6. Formatting the results of Map Reduce Computation – using Hadoop Output Formats.

Analytics

1. Simple analytics using Map Reduce.

2. Performing Group-By using Map Reduce.
3. Calculating frequency distributions and sorting using Map Reduce.
4. Plotting the Hadoop results using GNU plot.
5. Calculating histograms using Map Reduce.
6. Calculating scatter plots using Map Reduce.
7. Parsing a Complex dataset with Hadoop.
8. Joining two datasets using Map Reduce

Learning Resource

Text Books

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