CP5261 DATA ANALYTICS LABORATORY

DETAILED SYLLABUS

OBJECTIVES

- To implement Map Reduce programs for processing big data
- To realize storage of big data using H base, Mongo DB
- To analyse big data using linear models
- To analyse big data using machine learning techniques such as SVM / Decision tree classification and clustering

LIST OF EXPERIMENTS

Hadoop

1. Install, configure and run Hadoop and HDFS
2. Implement word count / frequency programs using MapReduce
3. Implement an MR program that processes a weather dataset R
4. Implement Linear and logistic Regression
5. Implement SVM / Decision tree classification techniques
6. Implement clustering techniques
7. Visualize data using any plotting framework
8. Implement an application that stores big data in H base/ MongoDB/ Pig using Hadoop/R.

REFERENCES

1. Alan Gates and Daniel Dai, "Programming Pig – Dataflow scripting with Hadoop", O'Reilley, 2nd Edition, 2016.
2. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, —An Introduction to Statistical Learning with Applications in R‖, Springer Publications, 2015(Corrected 6th Printing)
3. Hadley Wickham, ‖ggplot2 – Elegant Graphics for Data Analysis‖, Springer Publications,2nd Edition, 2016
4. Kristina Chodorow, "MongoDB: The Definitive Guide – Powerful and Scalable Data Storage", O'Reilley, 2nd Edition, 2013.
5. Lars George, "HBase: The Definitive Guide", O'Reilley, 2015.
6. Tom White, —Hadoop: The Definitive Guide – Storage and Analysis at Internet Scale‖, O'Reilley, 4th Edition, 2015.