Semester-I

19CSP112 HADOOP – PRACTICAL 4H – 2C

Instruction Hours / week: L: 0 T: 0 P: 4 **Marks:** Internal:**40** External:**60** Total: **100**

End Semester Exam : 3 Hours

Course Objectives

- It provides grounding in basic and advanced methods to big data technology and tools like MapReduce and Hadoop and its ecosystem.
- Understand the Big Data Platform and web analytics
- Provide an overview of Apache Hadoop and Design of HDFS
- Understand Map Reduce features
- Provide hands on Hbase, Pig and HiveQL queries
- Introduction to NoSQL and Data Model
- Exposure to Graph databases Neo4J, Connecting your data and Dashboard

Course Outcomes(COs)

On successful completion of the course the student should be able to:

1. Apply Hadoop ecosystem components.
2. Access and Process Data on Hbase, Pig and HiveQL queries
3. Manage Job Execution in Hadoop Environment
4. Analyze Map Reduce Types
5. Apply Data Model and Connect your data and Dashboard
6. Participate data science and big data analytics projects

1. Perform setting up and Installing Hadoop in its three operating modes: Standalone, Pseudo distributed, Fully distributed.
2. Implement the following file management tasks in Hadoop: Adding files and directories• Retrieving files• Deleting files•
   Hint: A typical Hadoop workflow creates data files (such as log files) elsewhere and copies them into HDFS using one of the above command line utilities.
3. Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.
4. Write a Map Reduce program that mines weather data. Weather sensors collecting data every hour at many locations across the globe gather a large volume of log data, which is a good candidate for analysis with MapReduce, since it is semi structured and record-oriented.
5. Implement Matrix Multiplication with Hadoop Map Reduce
6. Write a Map Reduce program to implement Join operations on RDBMS.
7. Write a Map Reduce program to determine statistical measures a) Variance b) Max c) Min d) Range of a large data collection.
8. K-means clustering using map reduce
9. Page Rank Computation

SUGGESTED READINGS

1 Tom White. (2012). Hadoop: The Definitive Guide. 2nd Edition. OReilly.

2 Tom White. (2014). The Definitive Guide to Mongodb. 4th Edition. OReilly.

3 Rik Van Bruggen. (2014). Learning Neo4j. 1st Edition. Packt Publishing Ltd. UK.

4 Daniel G.Murray. (2016). Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software. 2nd Edition. Wiley.
5 Dirk deRoos, Paul Zikopoulos, Bruce Brown, Roman B. Melnyk,RafaelCoss.(2012). Hadoop For Dummies.
6 GauravVaish. (2013). Getting Started with NoSQL. Packt Publishing Ltd. UK.
7 Pramod J. Sadalage, Martin Fowler. (2013). NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Pearson Education.
8 Joshua N. Milligan. (2016). Learning Tableau. Packt Publishing Ltd. UK.

WEB SITES

https://www.tutorialspoint.com/big_data_analytics/
hadoop.apache.org/
https://www.mongodb.com/nosql-explained
https://neo4j.com/