BONAM VENKATA CHALAMAYYA INSTITUTE OF TECHNOLOGY & SCIENCE

(AUTONOMOUS)

DEPARTMENT OF CSE - ARTIFICIAL INTELLIGENCE & DATA SCIENCE

III Year II Semester

L	T	P	C
0	0	3	1.5

BIG DATA ANALYTICS LAB

Software Requirements:
1. **Hadoop** : https://hadoop.apache.org/release/2.7.6.html
2. **Java** : https://www.oracle.com/java/technologies/javase/javase8u211-later-archive-downloads.html
3. **Eclipse** : https://www.eclipse.org/downloads/

List of Experiments:

Experiment 1: Week 1, 2:
1. Implement the following Data structures in Java
   a) Linked Lists b) Stacks c) Queues d) Set e) Map

Experiment 2: Week 3:
2. (i)Perform setting up and Installing Hadoop in its three operating modes:
Standalone,   Pseudo distributed, Fully distributed
(ii)Use web based tools to monitor your Hadoop setup.

Experiment 3: Week 4:
3. 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.

Experiment 4: Week 5:
4. Run a basic Word Count MapReduce program to understand MapReduce Paradigm.

Dr. A S N CHAKRAVARTHY, Professor, Department of CSE, UCEK JNTUK, Kakinada	Dr. JIMSON MATHEW Professor Dept of Computer Science and Engg. Indian Institute of Technology Patna	Prof. CHAPRAM SUDHAKAR Professor, Department of CSE, National Institute of Technology, Warangal - 506 004 Telangana, INDIA	Mr. RAJESH BOBBURI COO & Director, HighQ Labs Pvt Ltd, Rajahmahendravaram	Mrs. G VIJAYA KUMARI Assoc Professor &HoD Department of CSE-AI & DS, AIML B.V.C.I.T.S, Batlapalem

Experiment 5: Week 6:

   5. 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 Map Reduce, since it is semi structured and record-oriented.

Experiment 6: Week 7:

   6. Use MapReduce to find the shortest path between two people in a social graph.

   Hint: Use an adjacency list to model a graph, and for each node store the distance from the original node, as well as a back pointer to the original node. Use the mappers to propagate the distance to the original node, and the reducer to restore the state of the graph. Iterate until the target node has been reached.

Experiment 7: Week 8:

   7. Implement Friends-of-friends algorithm in Map Reduce.

   Hint: Two Map Reduce jobs are required to calculate the FoFs for each user in a social network .The first job calculates the common friends for each user, and the second job sorts the common friends by the number of connections to your friends.

Experiment 8: Week 9:

   8. Implement an iterative PageRank graph algorithm in Map Reduce.

   Hint: PageRank can be implemented by iterating a Map Reduce job until the graph has converged. The mappers are responsible for propagating node PageRank values to their adjacent nodes, and the reducers are responsible for calculating new PageRank values for each node, and for re-creating the original graph with the updated PageRank values.

Experiment 9: Week 10:

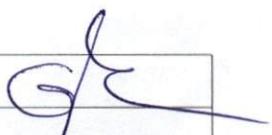   9. Perform an efficient semi-join in MapReduce.

   Hint: Perform a semi-join by having the mappers load a Bloom filter from the Distributed Cache, and then filter results from the actual MapReduce data source by performing membership queries against the Bloom filter to determine which data source records should be emitted to the reducers.

Experiment 10: Week 11:

   10. Install and Run Pig then write Pig Latin scripts to sort, group, join, project, and filter your data.

Experiment 12: Week 12:

   12. Install and Run Hive then use Hive to create, alter, and drop databases, tables, views, functions, and indexes

Dr. A S N CHAKRAVARTHY. Professor, Department of CSE. UCEK JNTUK, Kakinada	Dr. JIMSON MATHEW Professor Dept of Computer Science and Engg. Indian Institute of Technology Patna	Prof. CHAPRAM SUDHAKAR Professor. Department of CSE. National Institute of Technology. Warangal - 506 004 Telangana. INDIA	Mr. RAJESH BOBBURI COO & Director, HighQ Labs Pvt Ltd, Rajahmahendravaram	Mrs. G VIJAYA KUMARI Assoc Professor &HoD Department of CSE-AI & DS. AIML B.V.C.I.T.S, Batlapalem