HADOOP LAB

(Skill Based Lab Elective-I)

Course Code: 15CS11S2 **1 Credit**

Pre-requisites:

OOP through Java, Hadoop concepts.

Course Outcomes :

At the end of the Course, the Student will be able to:

CO 1 Install Hadoop on a Standalone system.

CO 2 Develop application for word count in any text document.

CO 3 Develop an application that calculates temperature variations on weather dataset.

CO 4 Install Hive on Hadoop system.

CO 5 Implement Hive queries on top of HDFS.

LIST OF EXPERIMENTS:

1) Create SSH connection on system installed with Ubuntu 14.04 LTS.

2) Download and Install Hadoop stable version on the above system.

3) Write necessary HDFS commands to place the data from Local file system to HDFS.

4) Develop a Mapper class for the word count on text document.

5) Develop a Reducer class for the word count on text document.

6) Develop an application to find the word count on text document.

7) Develop a Mapper class for the Maximum temperature calculation in the weather dataset.

8) Develop a Reducer class for the Maximum temperature calculation in the weather dataset.

9) Develop an application to find the Maximum temperature in the weather dataset.

10) Develop an application to find the Maximum temperature, using combiner function.

11) Download and Install Hive on Hadoop system.

12) Write a Hive query to find maximum temperature using MAX() function.

TEXT BOOK:

Tom White,"Hadoop: the Definitive Guide", 3rd edition, O'Reilly Media / Yahoo Press, 2012.

WEB REFERENCES:

1. https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/SingleNodeSetup.html
2. https://cwiki.apache.org/confluence/display/Hive/GettingStarted
3. https://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html