G.PULLA REDDY ENGINEERING COLLEGE (AUTONOMOUS): KURNOOL
COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
B.TECH. – VI SEMISTER (S-2020)
BIG DATA TECHNOLOGIES LABORATORY

List of Experiments

1. Perform Hadoop setup in Local and Pseudo mode and monitor through web based UI.
2. Implementation of Hadoop shell commands on files.
3. Implementation of word count example using Hadoop MapReduce.
4. Write a MapReduce program that works on Gutenberg data.
5. Write a MapReduce program that mines weather data.
6. Write pig latin scripts on Describe, for each and order by operator.
7. Write pig latin scripts to perform set and sort operation.
8. Perform DDL operations on Hive.
9. Implementation of data management using NOSQL databases.

Video Tutorials	
https://www.youtube.com/channel/UC_6mhzMATOtsC1UXO0sHpwA	
Topic	**Youtube link**
Install Ubuntu in Virtualbox	https://www.youtube.com/watch?v=2QVz7715n5g
run Wordcount MapReduce	https://www.youtube.com/watch?v=G0xyw1ODi5A
MapReduce on Gutenberg	https://www.youtube.com/watch?v=q8INOCrU9HE
Pig Latin Operators	https://www.youtube.com/watch?v=2N9gP1l9_F4

G.PULLA REDDY ENGINEERING COLLEGE (AUTONOMOUS): KURNOOL
COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
B.TECH. – VI SEMISTER (S-2020)
BIG DATA TECHNOLOGIES LABORATORY

01.	**Perform Hadoop setup in Local and Pseudo mode and monitor through web based UI.**
Expected Output	a) Successful installation of Hadoop in local, pseudo mode **hadoop version** b) Monitor Namenode,secondarynamenode,datanode,YARN RM, YARN NM information

Local (Standalone) mode:

Step	Details
1.	Prerequisites: a) VMWare b) Ubuntu 18.04 c) Jdk 8 d) Hadoop 2.10.0
2.	*Open Terminal and type in the following command* sudo apt-get install openjdk-8-jdk
3.	*Check whether java is installed or not using the command* java -version
4.	*Download Hadoop 2.10.0*
5.	cd /Downloads
6.	sudo tar xvf hadoop-2.10.0.tar.gz
7.	sudo mv hadoop-2.10.0 /opt
8.	cd /
9.	cd opt
10.	sudo chmod 777 hadoop-2.10.0
11.	cd /home/Sreedhar
12.	sudo gedit .bashrc *At the end of the file (after fi) add the following (export JAVA_HOME…)*

```
# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi


export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64/
alias jps='/usr/lib/jvm/java-8-openjdk-amd64/bin/jps'
export HADOOP_HOME=/opt/hadoop-2.10.0/
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export HADOOP_CLASSPATH=${JAVA_HOME}/lib/tools.jar
```

13.	source .bashrc
14.	hadoop version

G.PULLA REDDY ENGINEERING COLLEGE (AUTONOMOUS): KURNOOL
COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
B.TECH. – VI SEMISTER (S-2020)
BIG DATA TECHNOLOGIES LABORATORY

Pseudo mode

Step	Details
1.	Prerequisites: a) VMWare b) Ubuntu 18.04 c) Jdk 8 d) Hadoop 2.10.0
2.	*Open Terminal and type in the following command* sudo apt-get install openjdk-8-jdk
3.	*Check whether java is installed or not using the command* java -version
4.	sudo su
5.	adduser hduser *(Give password)*
6.	usermod -aG sudo hduser
7.	sudo su hduser
8.	sudo apt-get purge openssh-server
9.	sudo apt-get install openssh-server
10.	ssh-keygen -t rsa
11.	cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
12.	ssh localhost
13.	cd /home/hduser
14.	*Download Hadoop 2.10.0*
15.	sudo tar xvf hadoop-2.10.0.tar.gz
16.	sudo mv /home/hduser/hadoop-2.10.0 /opt
17.	cd /
18.	cd opt
19.	sudo chmod 777 hadoop-2.10.0
20.	cd /home/hduser
21.	sudo gedit .bashrc *At the end of the file add export JAVA_HOME...(Same as local mode)*
22.	source .bashrc
23.	cd /
24.	cd opt
25.	cd hadoop-2.10.0
26.	cd etc
27.	cd hadoop
28.	sudo gedit hadoop-env.sh *replace the following export JAVA_HOME=${JAVA_HOME}*

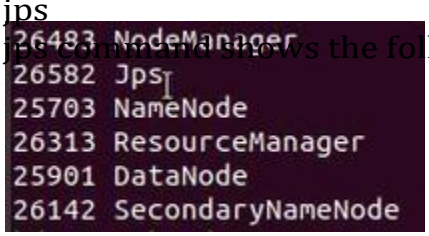
```
# The java implementation to use.
#export JAVA_HOME=${JAVA_HOME}
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

29.	sudo gedit core-site.xml

	add the following between *<configuration> </configuration>*
30.	sudo gedit hdfs-site.xml *add the following between <configuration> </configuration>*
31.	sudo gedit yarn-site.xml *add the following between <configuration> </configuration>*
32.	sudo cp mapred-site.xml.template mapred-site.xml
33.	sudo gedit yarn-site.xml *add the following between <configuration> </configuration>*
34.	cd /home/hduser
35.	sudo mkdir -p hadoop_tmp/hdfs/namenode
36.	sudo mkdir -p hadoop_tmp/hdfs/datanode
37.	sudo chmod 777 -R hadoop_tmp/hdfs/namenode
38.	sudo chmod 777 -R hadoop_tmp/hdfs/datanode
39.	sudo chown -R hduser hadoop_tmp/hdfs/datanode
40.	hdfs namenode -format
41.	start-dfs.sh

42. start-yarn.sh
43. jps

jps command shows the following output



44. To stop all hadoop daemon services, use the following command
stop-dfs.sh
stop-yarn.sh

Monitor through Web based UI	
Namenode information	localhost:50070
Secondarynamenode information	localhost:50090
Datanode information	localhost:50075
YARN Resource Manager	localhost:8088
YARN Node Manager	localhost:8042

02. Implementation of Hadoop shell commands on files

Syntax and Description	Example (Usage)
hadoop **version** displays the version of hadoop installed in the system	hadoop version
hadoop fs -**ls** / *Displays List of Files and Directories in HDFS file Path*	hadoop fs -ls /
hadoop fs -**mkdir** *create a directory on an HDFS environment.*	hadoop fs -mkdir /user/hadoop/
hadoop fs -**put** *used to copy files from the local file system to the HDFS filesystem*	hadoop fs -put sample.txt /user/data/
hadoop fs -**get** *used to copy files from HDFS file system to the local file system, just the opposite to put command.*	hadoop fs -get /user/data/sample.txt workspace/
hadoop fs -**cat** URI [URI ...] *used for displaying the contents of a file on the console.*	hadoop fs -cat /user/data/sampletext.txt
hadoop fs -**cp** URI [URI ...] <dest> *Copy files from source to destination. This command allows multiple sources as well in which case the destination must be a directory.*	hadoop fs -cp /user/hadoop/file1 /user/hadoop/file2

hadoop fs -**appendToFile** <localsrc> ... <dst> *Append single src, or multiple srcs from local file system to the destination file system. Also reads input from stdin and appends to destination file system.*	hadoop fs -appendToFile localfile /user/hadoop/hadoopfile
hadoop fs -df URI [URI ...] Displays free space	hadoop dfs -df /user/hadoop/dir1
hadoop fs –help	hadoop fs –help
hadoop fs -touchz URI [URI ...] Create a file of zero length. An error is returned if the file exists with non-zero length	hadoop -touchz pathname
hadoop fs -rmdir URI [URI ...] Delete a directory	hadoop fs -rmdir /user/hadoop/emptydir

hadoop fs -mv URI [URI ...] <dest> Moves files from source to destination. This command allows multiple sources as well in which case the destination needs to be a directory.	hadoop fs -mv /user/hadoop/file1 /user/hadoop/file2