CSCD467/567 Lab8 HDFS

When you answer the questions in this lab, please capture the command you used on the master node of hadoop cluster, as well as the output returned by your command. A screenshot is fine to use.

Submission: Wrap up all your answers into a single pdf file. Name your file as *FirstInitialYourLastName*CSCD467Lab8.pdf. For example, if your legal name is Will Smith, you should name your file as wSmithCSCD467Lab8.pdf.

Before you leave the laboratory, please show the TA or the instructor how your program works, they will give you a score for this Lab assignment.
For archive purpose, please also submit your single file on EWU Canvas by following CSCD467 → Assignments → Lab8 → Submit Assignment to upload your single pdf file.

Problem Description:
Please practice all commands listed below, which help you get familiar with Hadoop Distributed File System. Please show screeshots to prove you have sucessfully completed each step below.

1) In your .bashrc file ocated in your **home directory** on the master node of the cluster at 146.187.135.37
First add the following commands into your .bashrc file.

```
-----------------------------------------------
export PATH="$PATH":"$HADOOP_PREFIX"/bin
export PATH=$JAVA_HOME/bin:$PATH
export HADOOP_CLASSPATH=$JAVA_HOME/lib/tools.jar
-----------------------------------------------
```

Then run: source ~/.bashrc

2) Upload input files or source files into the cluster from your local computer,
using sftp command or WinSCP to remotely login the master node 146.187.135.37
Then you can upload or download files.
The files for practice are provided in this package, as well as in the demo code section on canvas.
Note that: after the files have been uploaded into the cluster, by default they are hosted on the NATIVE operating system of the master node, as opposed to being hosted by the HDFS. You have to use hadoop command to further transfer the input files into HDFS.

2.1 #Please upload your textfile: file1 and file2 and WordCount.java into the cluster, using sftp or WinSCP.

2.2 #This is a command on the native OS of the cluster to verify your files are successfully uploaded.

```
ytian@cscd-doop01:~/467hadoop$ ls
file1  file2  WordCount.java
```

3) Use HDFS
3.1 #command to list items in your home directory inside HDFS, **in the following**, change 'ytian' to your **own account name** PLEASE.
hadoop fs -ls /user/ytian

3.2 #create a folder named as **wc** in HDFS under your home there.
hadoop fs -mkdir /user/ytian/wc

3.3 #create a folder named as 'input' inside /user/ytian/wc on HDFS
hadoop fs -mkdir /user/ytian/wc/input

3.4 #list all items under folder /user/ytian/wc on HDFS. **Your output should be similar**.
hadoop fs -ls /user/ytian/wc
Found 1 items
drwxr-xr-x - ytian supergroup 0 2016-01-22 10:33 /user/ytian/wc/input

3.5 #Transfer file1 on the local NATIVE OS into a folder /user/ytian/wc/input on HDFS, #The folder /user/ytian/wc/input will contains all input files for WordCount mapreduce job
hadoop fs -copyFromLocal ./file1 /user/ytian/wc/input

3.6 #Transfer file2 on the local NATIVE OS into a folder /user/ytian/wc/input on HDFS,
hadoop fs -copyFromLocal ./file2 /user/ytian/wc/input

3.7 #Explore the content of file1 in HDFS directly
hadoop fs -cat /user/ytian/wc/input/file1

3.8 #You can also use the following command to copy from HDFS to local native system. Please copy file2 on the HDFS under /user/ytian/wc/input back to Local native OS file system /tmp/file2 Please show a screenshot to prove you successfully achieved this.
hadoop fs -copyToLocal sourceFileOnHDFS localFolder