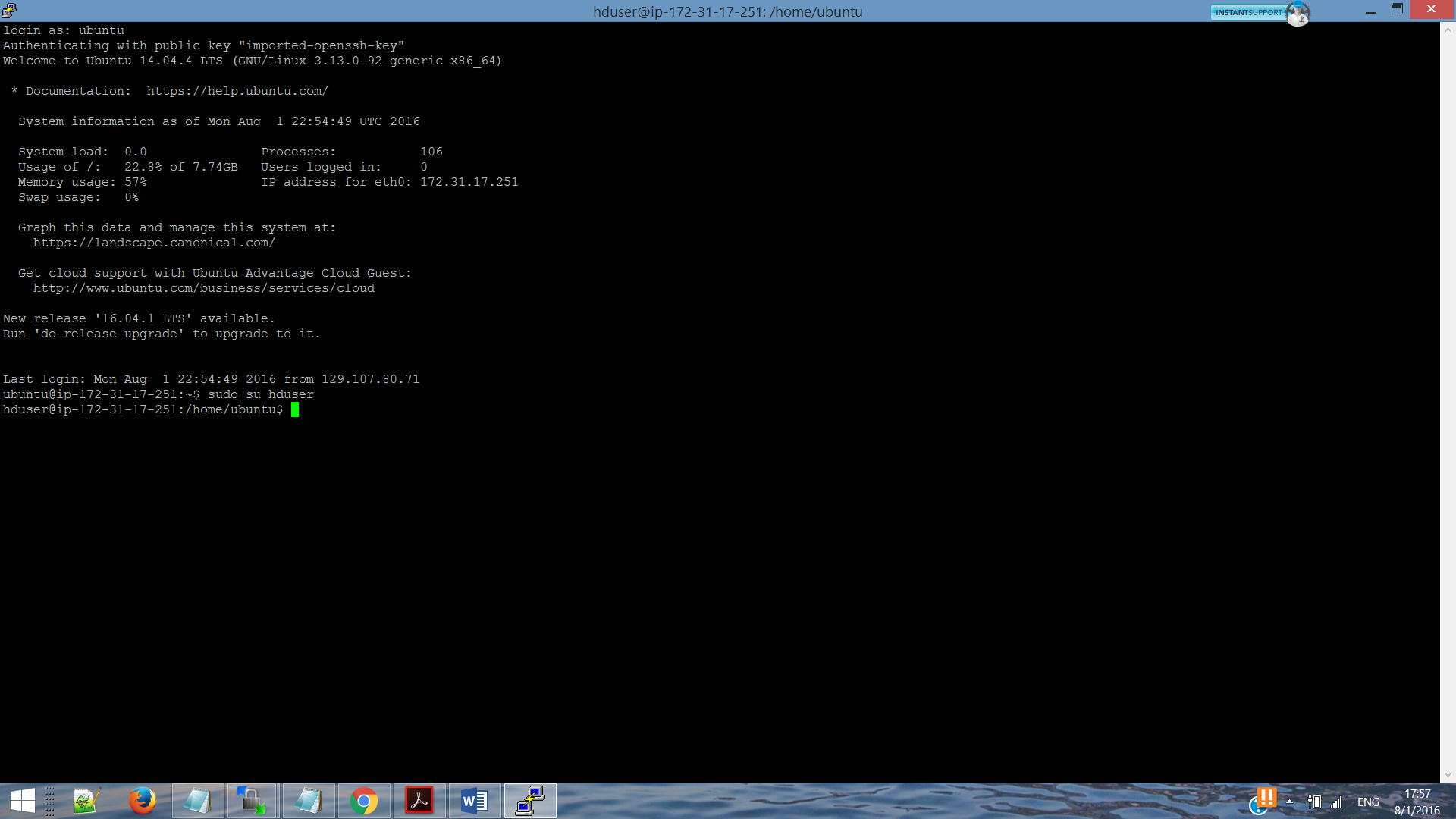
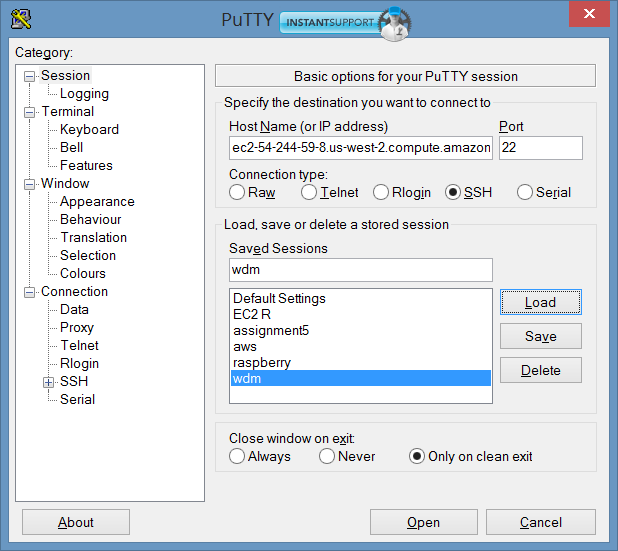
Amazon EC2 Commands for assignment 5

Started the instance in putty by giving the ec2 endpoint

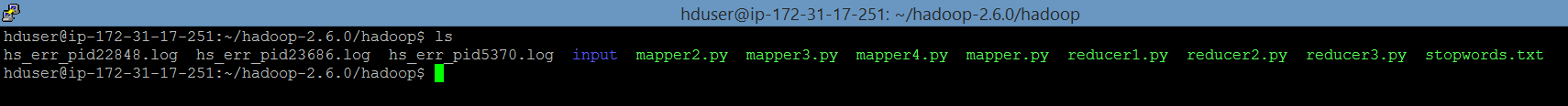


**List of commands hereafter**

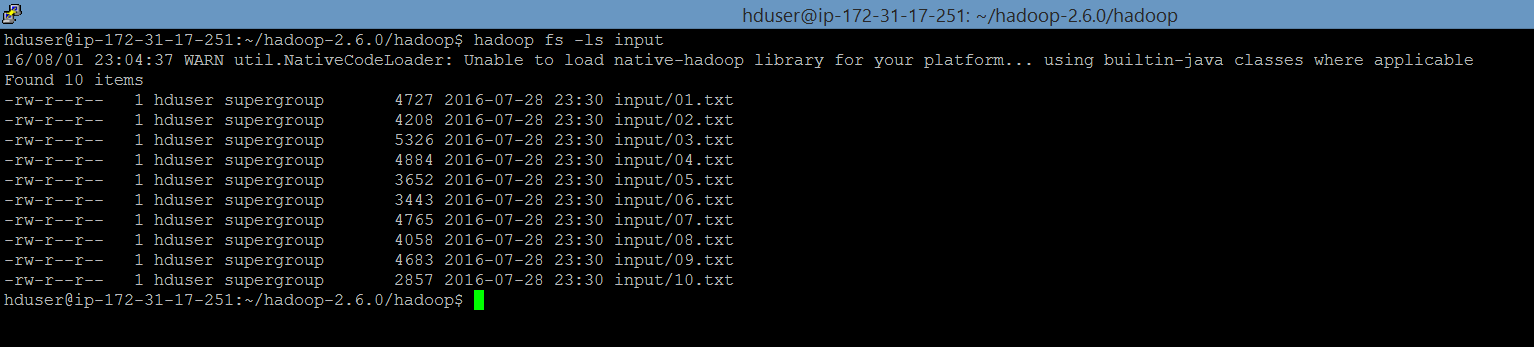
**Sudo su hduser**

**Ssh localhost**

I have mappers and reducers stored in the user hduser as follows. Also the stopwords.txt is used in the first mapper is kept here.



The 10 input files are loaded in the hdfs as follows



Note: the output of the first reducer will be put in the hdfs and commands are given in this document for the same.

**Running 1st mapper reducer file command**

hadoop jar /usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-2.6.0.jar -D mapred.map.tasks=1 -mapper mapper.py -reducer reducer1.py -input /user/hduser/input/ -output op

**Downloading the output in /home/hduser/**

hadoop fs -cat /user/hduser/op/part-00000 > input2.txt

**Putting input2.txt in hdfs for next mapper reducer**

hadoop fs -put input2.txt

**Running 2nd mapper reducer file command**

hadoop jar /usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-2.6.0.jar -D mapred.map.tasks=1 -mapper mapper2.py -reducer reducer2.py -input /user/hduser/input2.txt -output op2

**Downloading the output in /home/hduser/**

hadoop fs -cat /user/hduser/op2/part-00000 > input3.txt

**Putting input3.txt in hdfs for next mapper reducer**

hadoop fs -put input3.txt

**Running 3rd mapper reducer file command**

hadoop jar /usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-2.6.0.jar -D mapred.map.tasks=1 -mapper mapper3.py -reducer reducer3.py -input /user/hduser/input3.txt -output op3

**Downloading the output in /home/hduser/**

hadoop fs -cat /user/hduser/op3/part-00000 > input4.txt

**Putting input4.txt in hdfs for next and final mapper reducer**

hadoop fs -put input4.txt

**Running 4th and final mapper file command ( no reducer)**

hadoop jar /usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-2.6.0.jar -D mapred.map.tasks=1 -D mapred.reduce.tasks=0 -mapper mapper4.py -input /user/hduser/input4.txt -output op4

**Downloading the final output in /home/hduser/**

hadoop fs -cat /user/hduser/op4/part-00000 > finaloutput.txt

**The mapper and reducer files are attached in the submission folder**

Note:

to remove folder by force in hdfs

hadoop fs -rm -r directory\_name