ICP -1

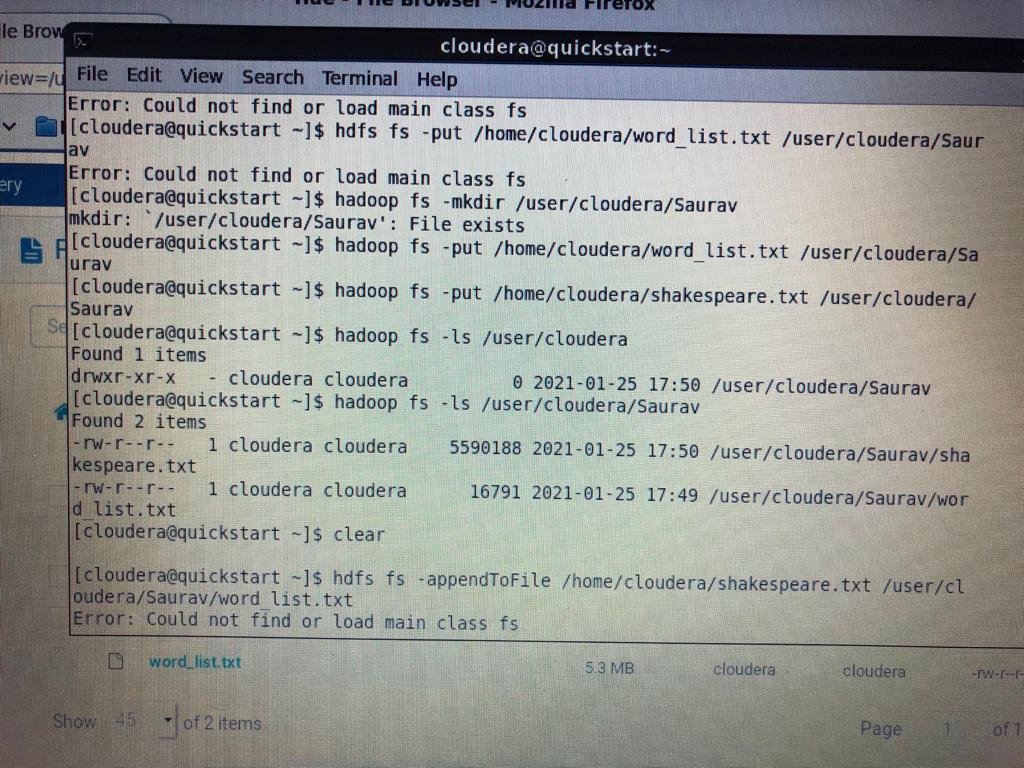
Name: Saurav Pawar

Email: [spp3dh@mail.umkc.edu](mailto:spp3dh@mail.umkc.edu)

Answer:

* Create directory in Hadoop hdfs:

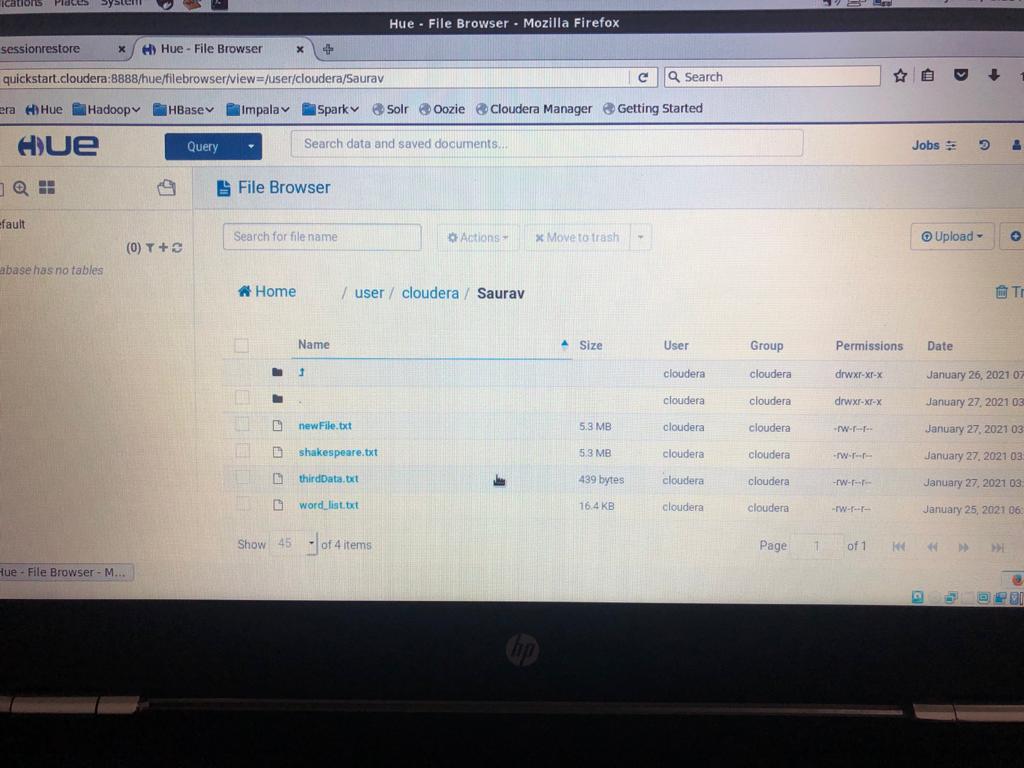
hadoop fs -mkdir /user/cloudera/Saurav



* Load Datasets in Hadoop hdfs:

hadoop fs -put /home/cloudera/shakespeare.txt /user/cloudera/Saurav

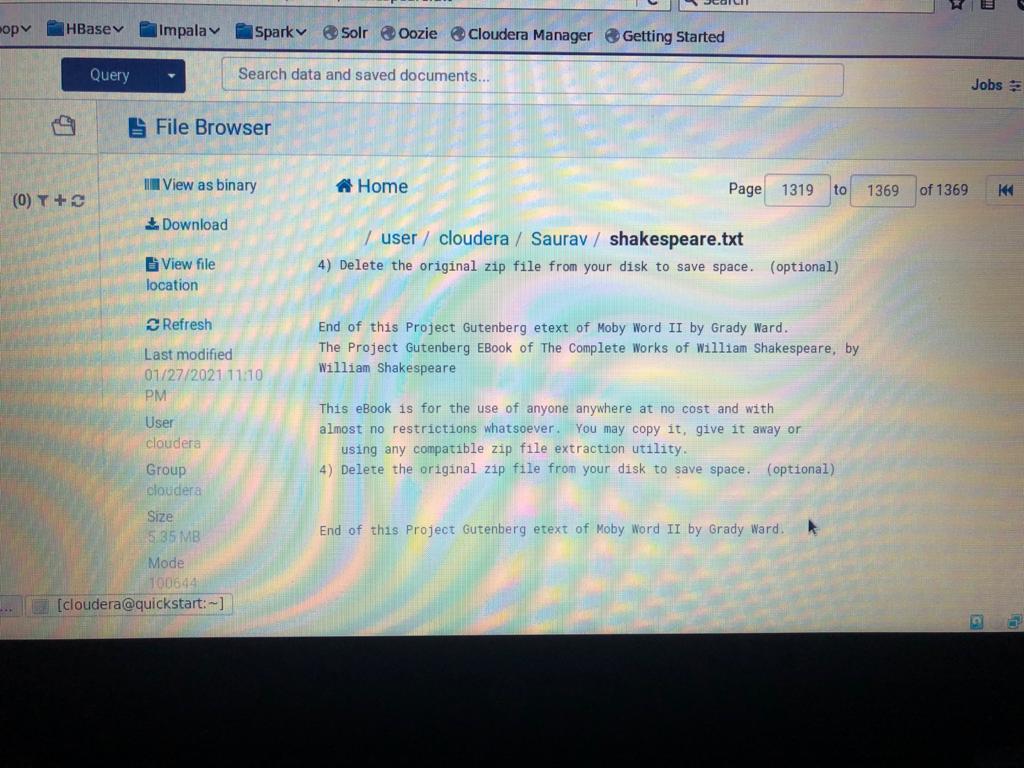
hadoop fs -put /home/cloudera/word\_list.txt /user/cloudera/Saurav



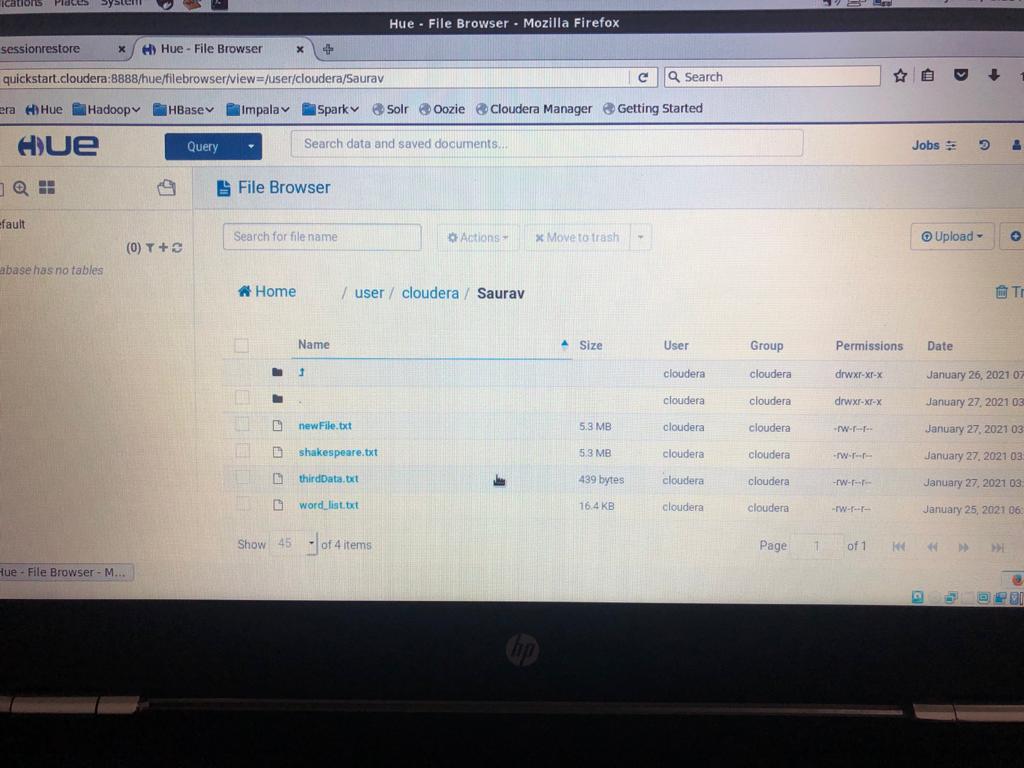
* Append the files

hadoop fs -appendToFile /home/cloudera/word\_list.txt /user/cloudera/Saurav/Shakespeare.txt

This command appends the word\_list.txt to shakespeare.txt.



* Visualize with HUE:



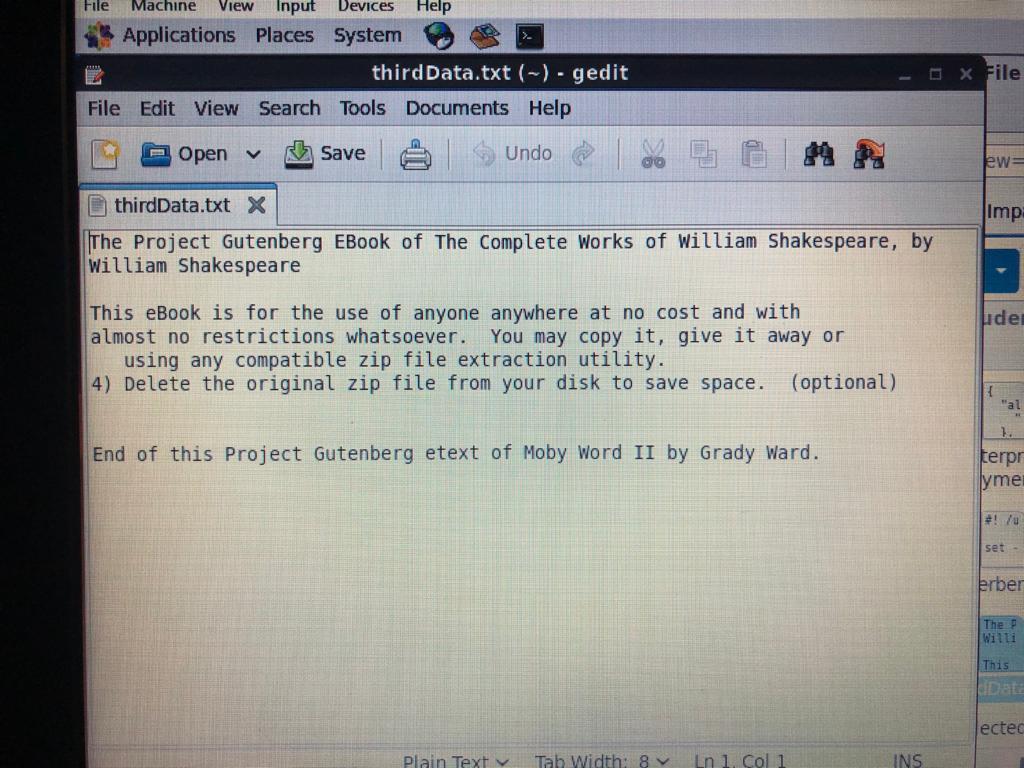
* Merge first and last 5 lines using hdfs commands:

To do this I pipe out the first and last 5 lines in new file (third\_dat.txt) in home directory using,

hadoop fs -cat /user/cloudera/Saurav/shakespeare.txt | head -5 >> /home/cloudera/third\_data.txt

hadoop fs -cat /user/cloudera/Saurav/shakespeare.txt | tail -5 >> /home/cloudera/third\_data.txt

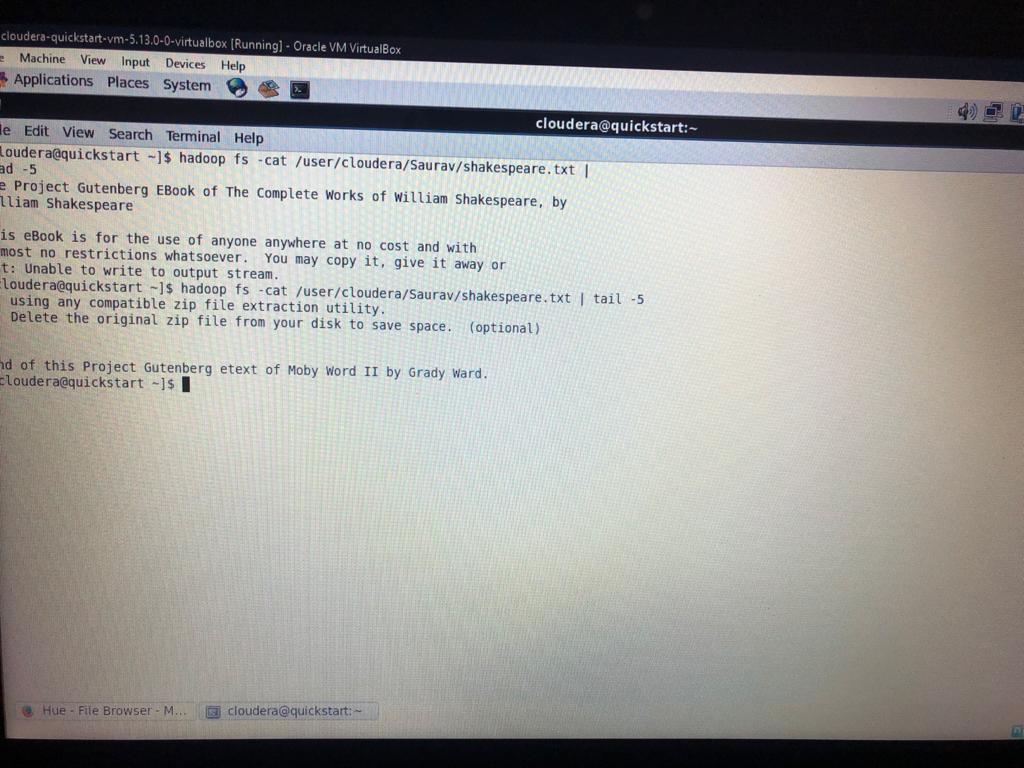
So now I have first and last 5 lines of Shakespeare.txt (which Is combination of 2 datasets) in third\_data.txt and I can put or append on hadoop hdfs.



* View first and last 5 lines with hdfs commands:

hadoop fs -cat /user/cloudera/Saurav/Shakespeare.txt | head -5

hadoop fs -cat /user/cloudera/Saurav/Shakespeare.txt | tail -5



* Create new file and append 3 datasets:

Shakespeare.txt has all 3 data sets appended to it.

I created new file,

Hadoop fs -touchz /user/cloudera/Saurav/newFile.txt

I put the Shakespeare.txt in new file,

hadoop fs -cat /user/cloudera/Saurav/shakespeare.txt | hadoop fs -put -f - /user/cloudera/Saurav/newFile.txt

