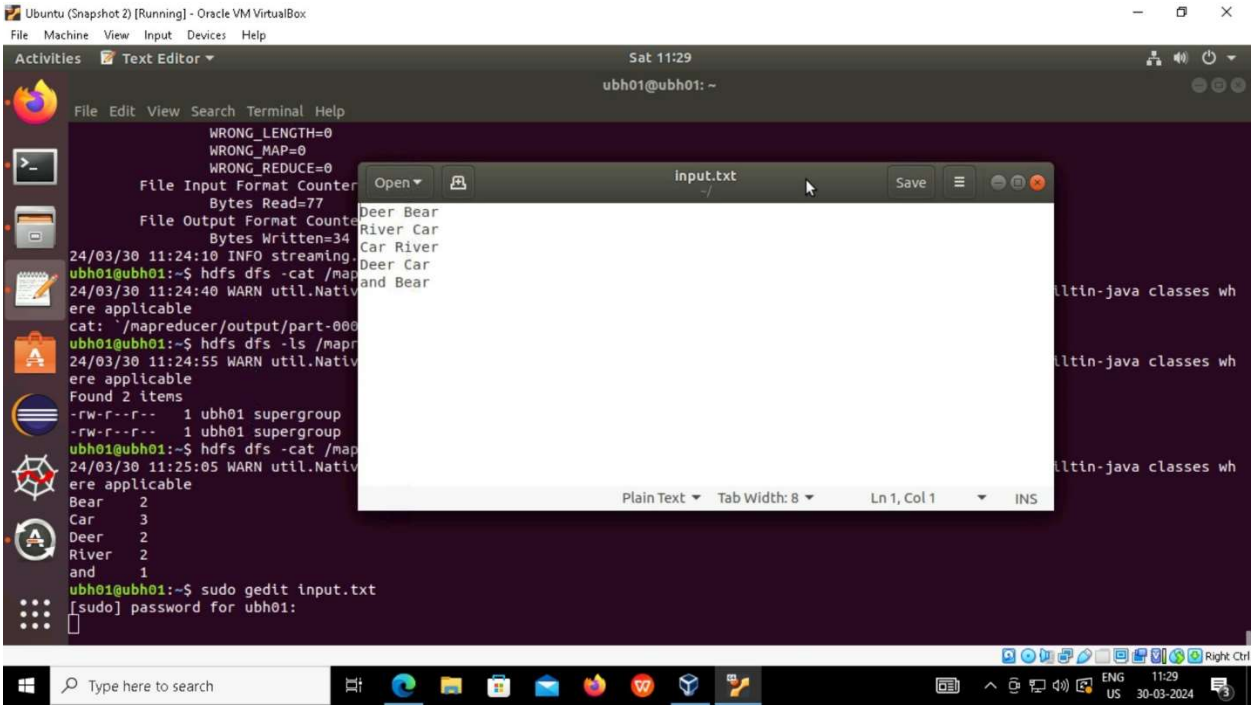
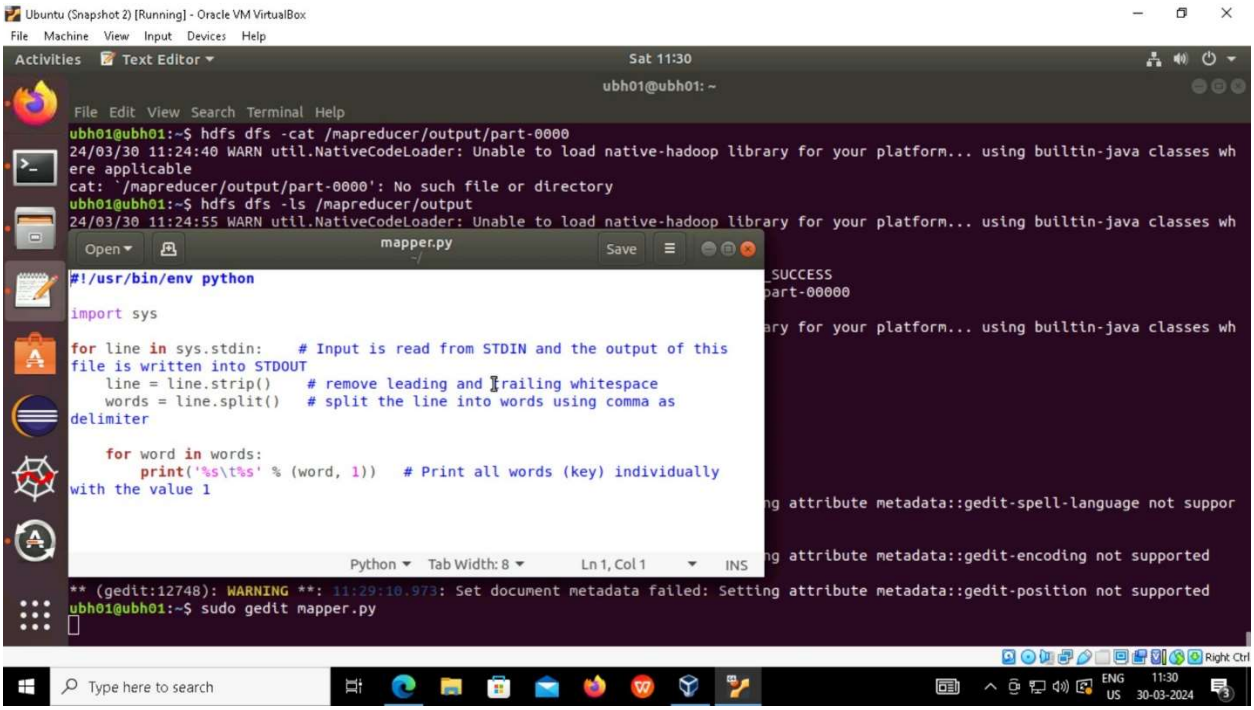
MAP REDUCER ASSIGNEMENT - 4

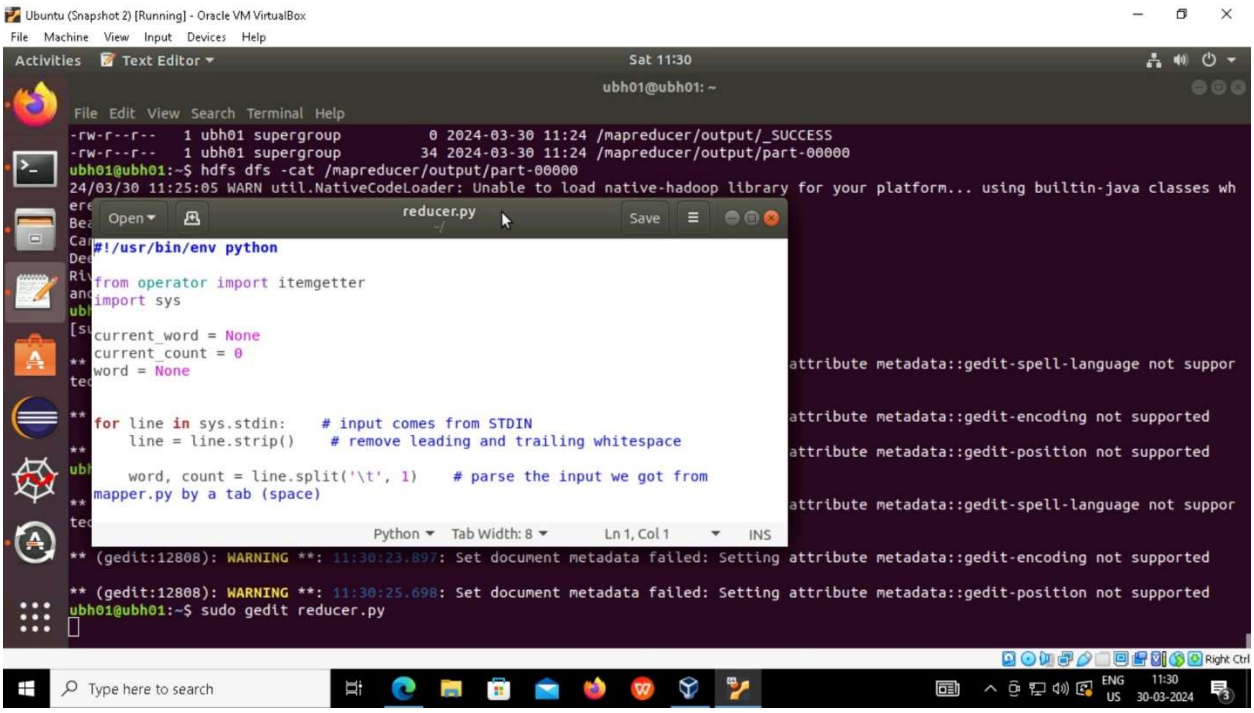
A. JOSHI SATYA VARDAN
CSDAIA24AZ003
EMP ID:2320097



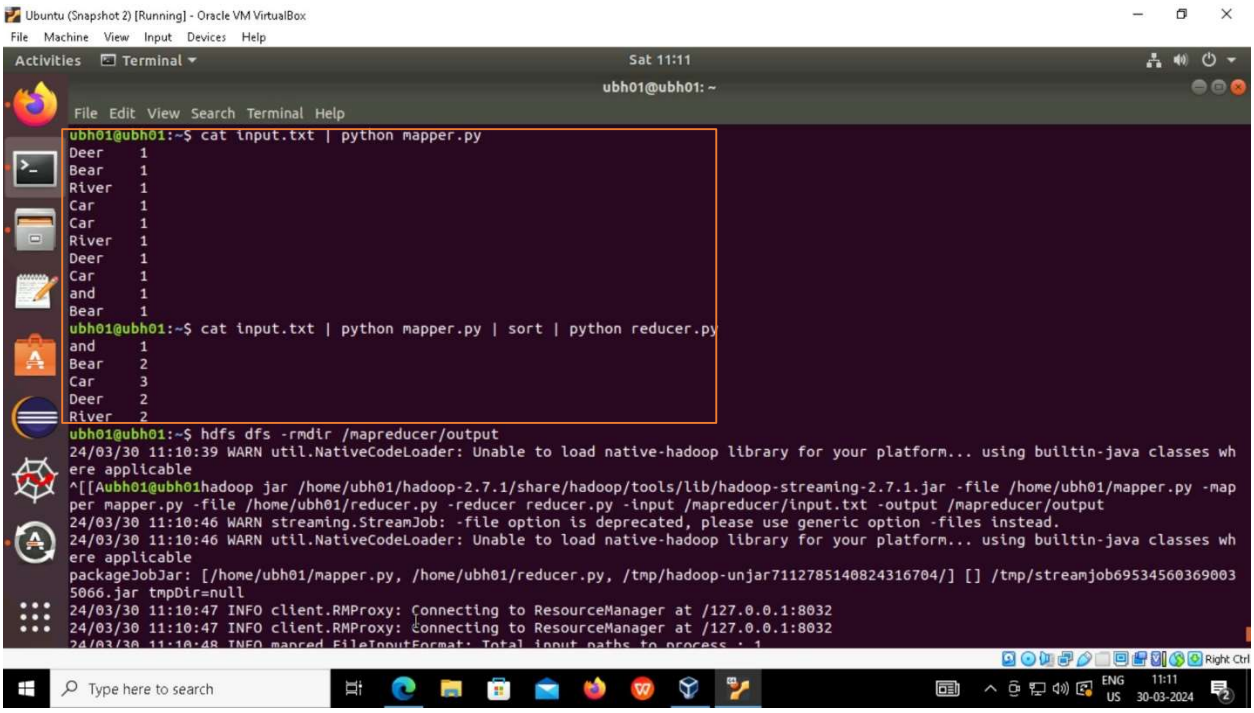
- First create the file by using command **sudo gedit input.txt** file by adding some text as shown like above picture.



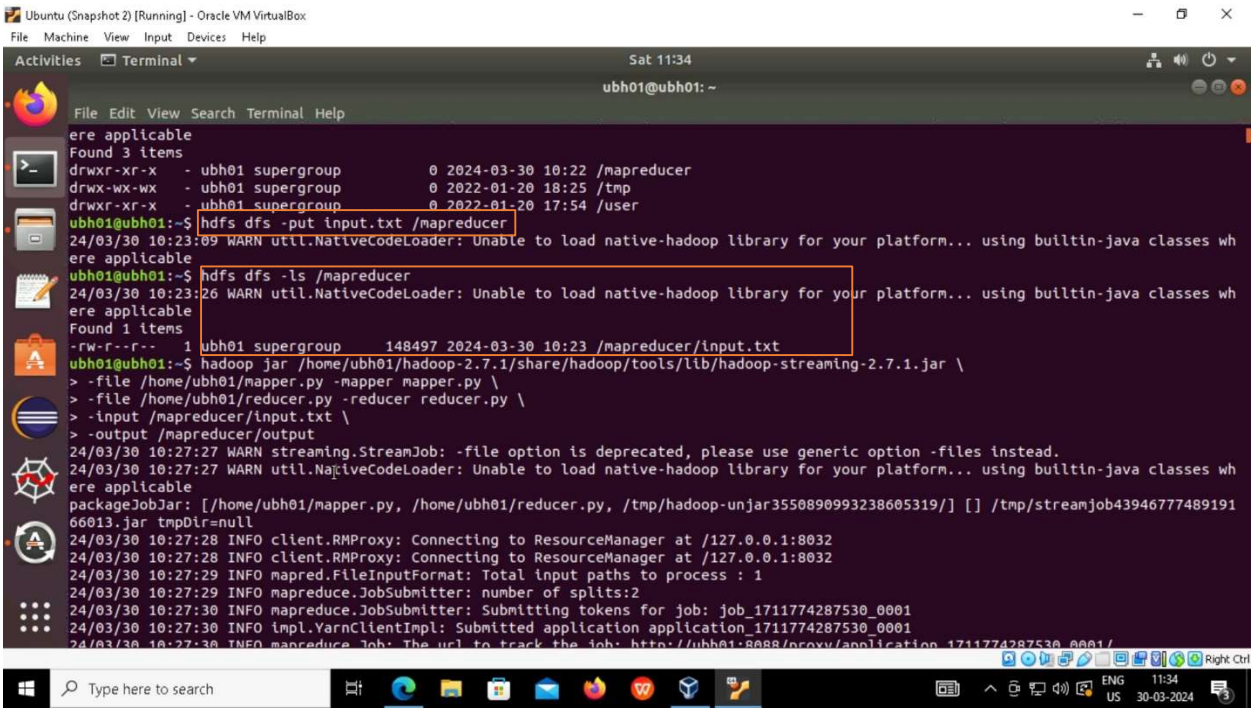
- Create another file by using command “**sudo gedit mapper.py** “and written some mapper function code inside that file.



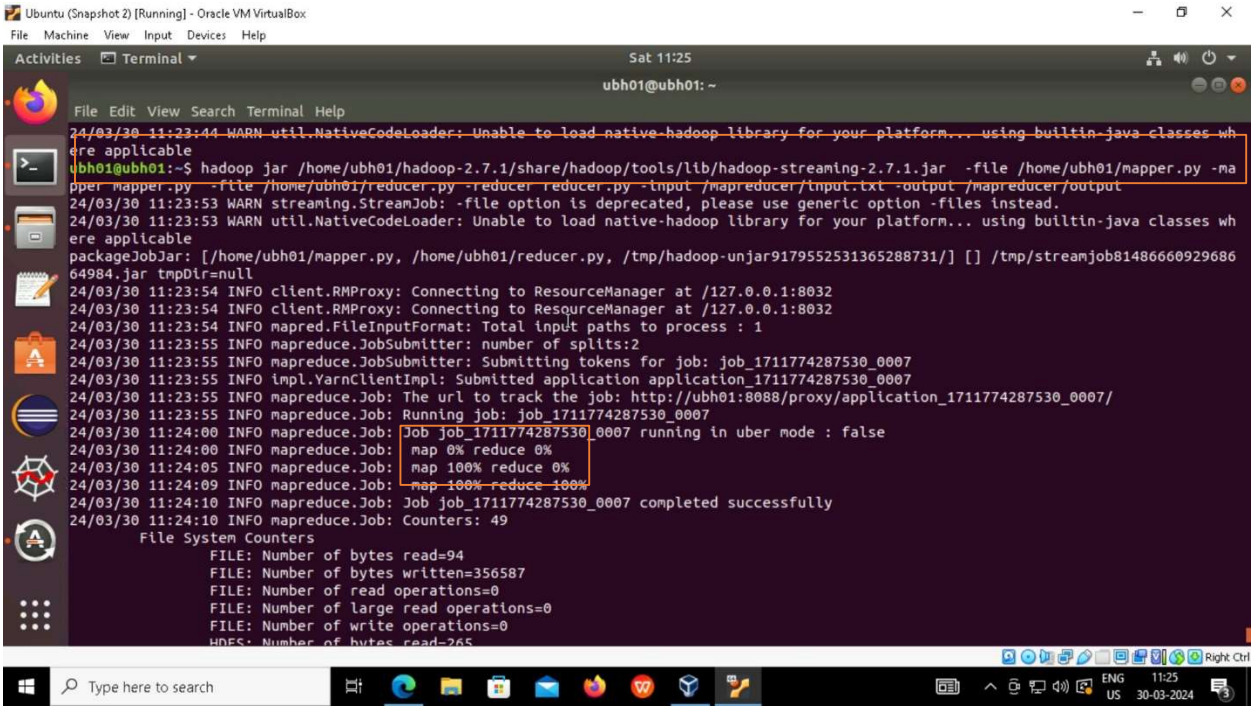
- Create another file by using command “**sudo gedit reducer.py”** and written some reducer function code inside that file.



- By usind command "**cat input.txt | python mapper.py**" and check whether mapper function is working on the given text file in local
- By using command "**cat input.txt | python mapper.py | sort | python reducer.py**" and checking whether the mapper and reducer functions are working on the given text file in local.



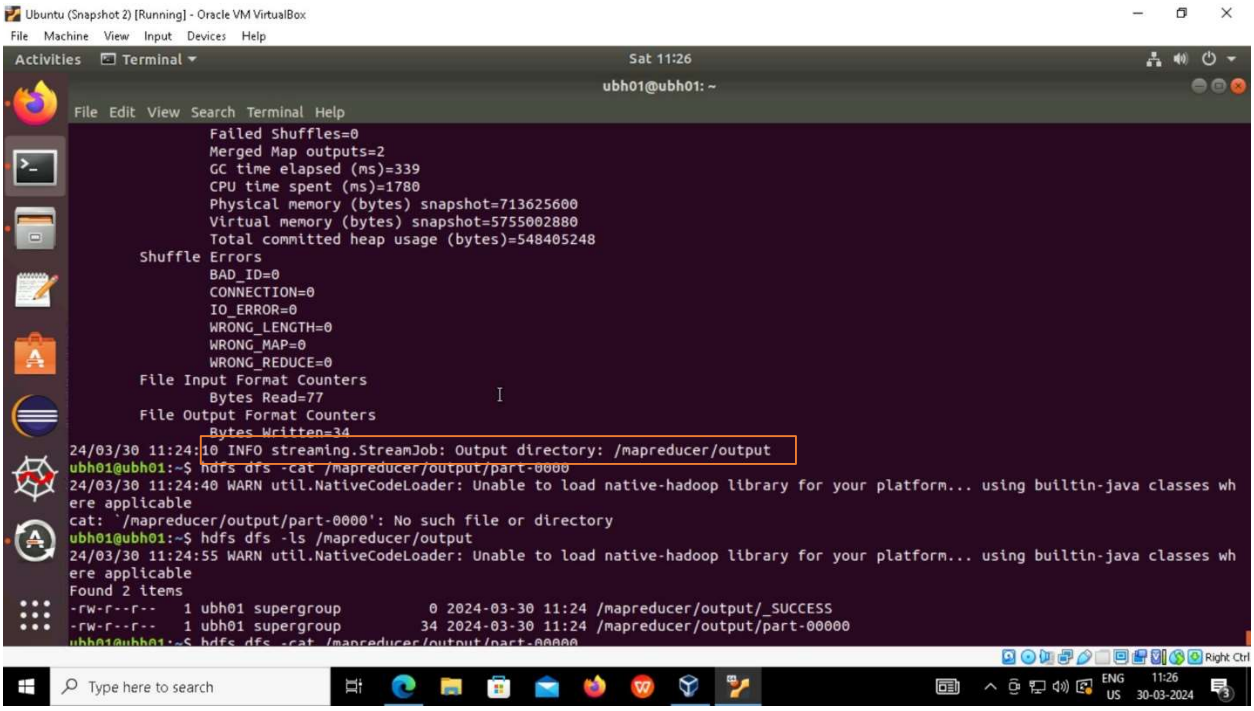
- By using this command "**hdfs dfs -mkdir /mapreducer**" create a directory in hdfs
- By using this command "**hdfs dsf -put input.txt /mapreducer**" moving the file input.txt from local to mapreducer directory in Hadoop.



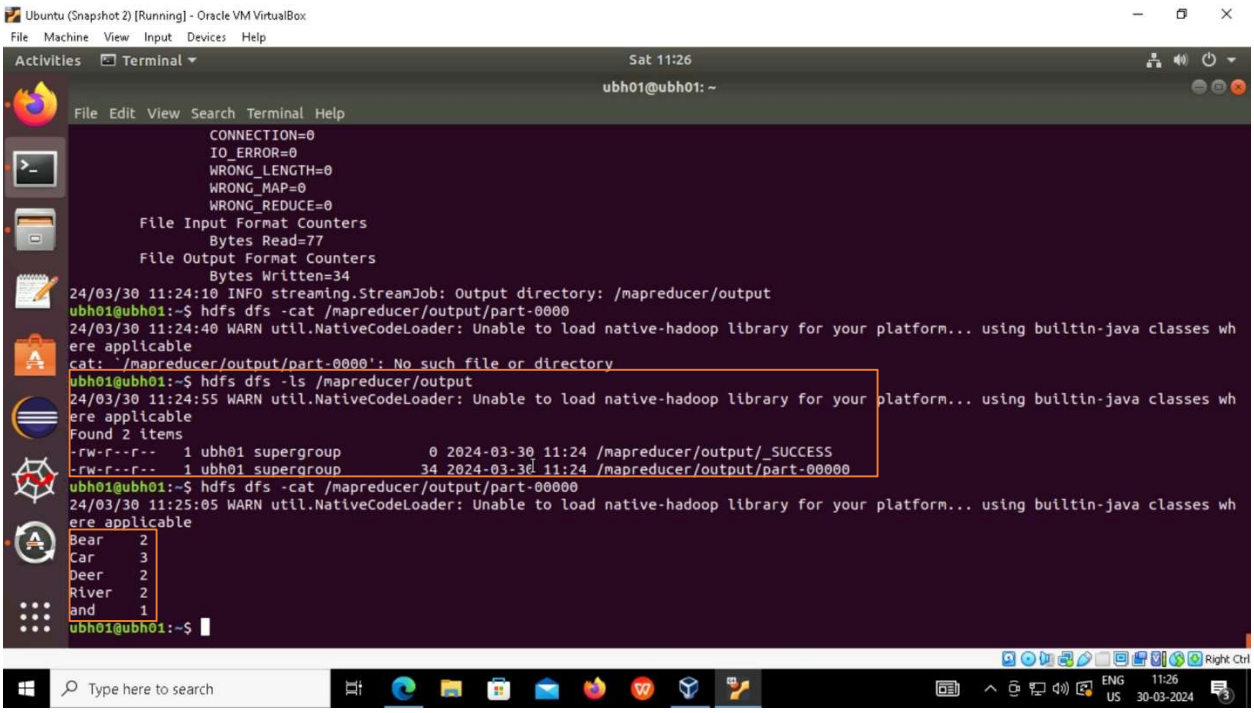
- By using this command

  “ hadoop jar /home/ubh01/hadoop-2.7.1/share/hadoop/tools/lib/hadoop/streaming-2.7.1.jar -file /home/ubh01/mapper.py -mapper mapper.py -file /home/ubh01/reducer.py -reducer reducer.py -input /mapreducer/input.txt -output /mapreducer/output “

  we are running mapreducer job by using hadoop streaming-2.7.1.jar and checking whether mapper and reducer are working 100% on the given input.txt file.



- After running map reducer job output get succeed and output file given by map reducer is stored in hadoop mapreducer/output directory.



- By using this command "**hdfs dfs -ls /mapreducer/output"** we are checking whether the output file s given by mapreducer is stored in output directory or not in Hadoop.
- By using this command "**hdfs -cat /mapreducer/output/part-0000** "we are checking the output which is given by the mapreducer is correct comparing with the local output.