19162121031

SMIT R PATEL

PRACTICAL 6

BIG DATA AND ANALYTICS

CODES OR COMMANDS :-

$hadoop fs -mkdir smitrpatel19 ## make folder which name smitrpatel19

$hadoop fs -put SalesJan2009.csv smitrpatel19 ## move csv file into folder

$hadoop fs -ls smitrpatel19 ## for go inside to folder

new terminal and type --> pig

grunt> salesTable = LOAD 'smitrpatel19/SalesJan2009.csv' USING PigStorage(',') AS (Transaction\_date:chararray,Product:chararray,Price:chararray,Payment\_Type:chararray,Name:chararray,City:chararray,State:chararray,Country:chararray,Account\_Created:chararray,Last\_Login:chararray,Latitude:chararray,Longitude:chararray);

grunt> GroupbyCountry = GROUP salesTable by Country; ##load the file

grunt> CountbyCountry = FOREACH GroupbyCountry GENERATE CONCAT((chararray)$0,CONCAT(':',(chararray)COUNT($1)));

grunt> STORE CountbyCountry INTO 'pig\_output\_sales1' USING PigStorage('\t');

new terminal and type -->

$hdfs dfs -cat pig\_output\_sales1/part-r-00000

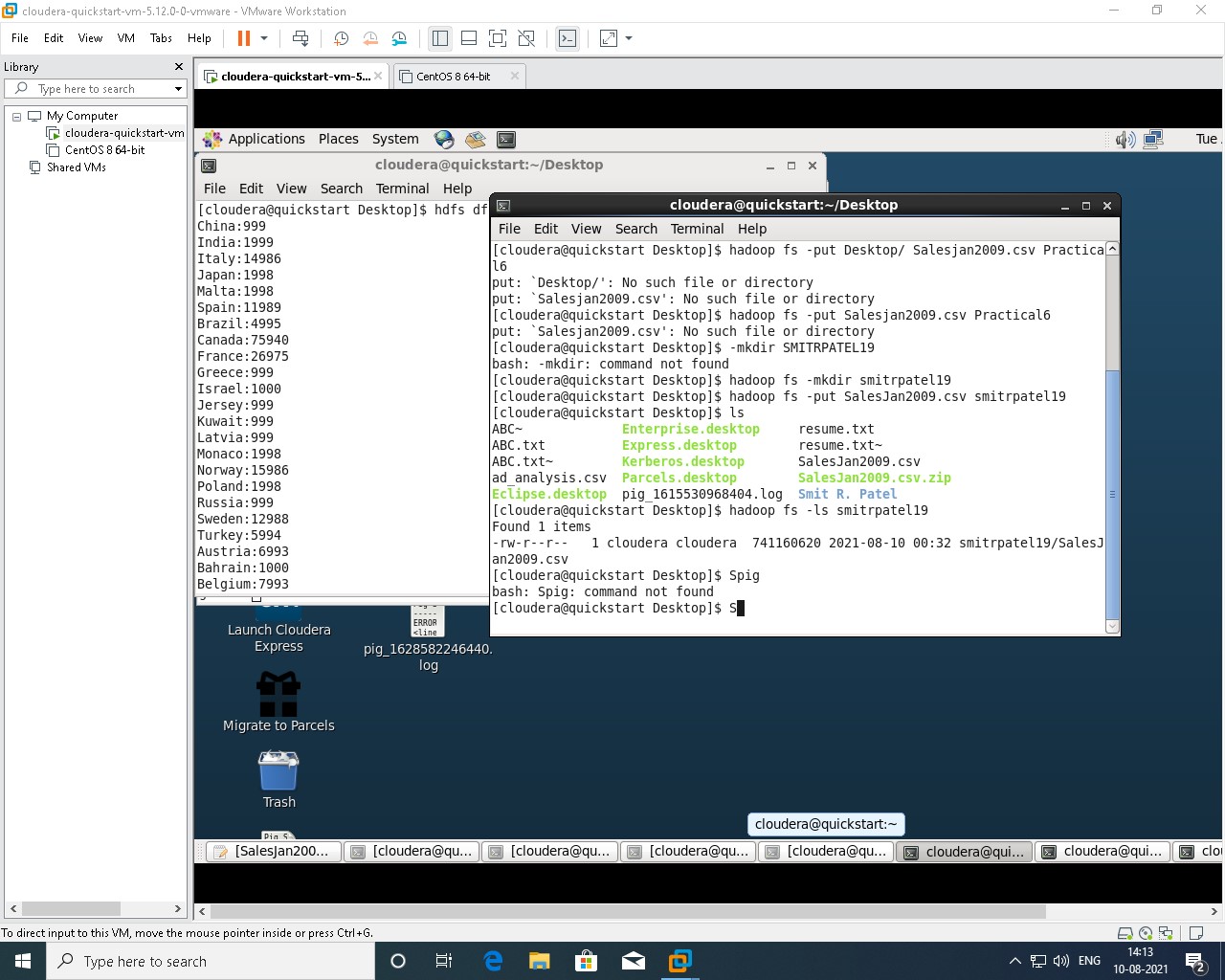
Code with output :-

Step 1 :-

$hadoop fs -mkdir smitrpatel19 ## make folder which name smitrpatel19

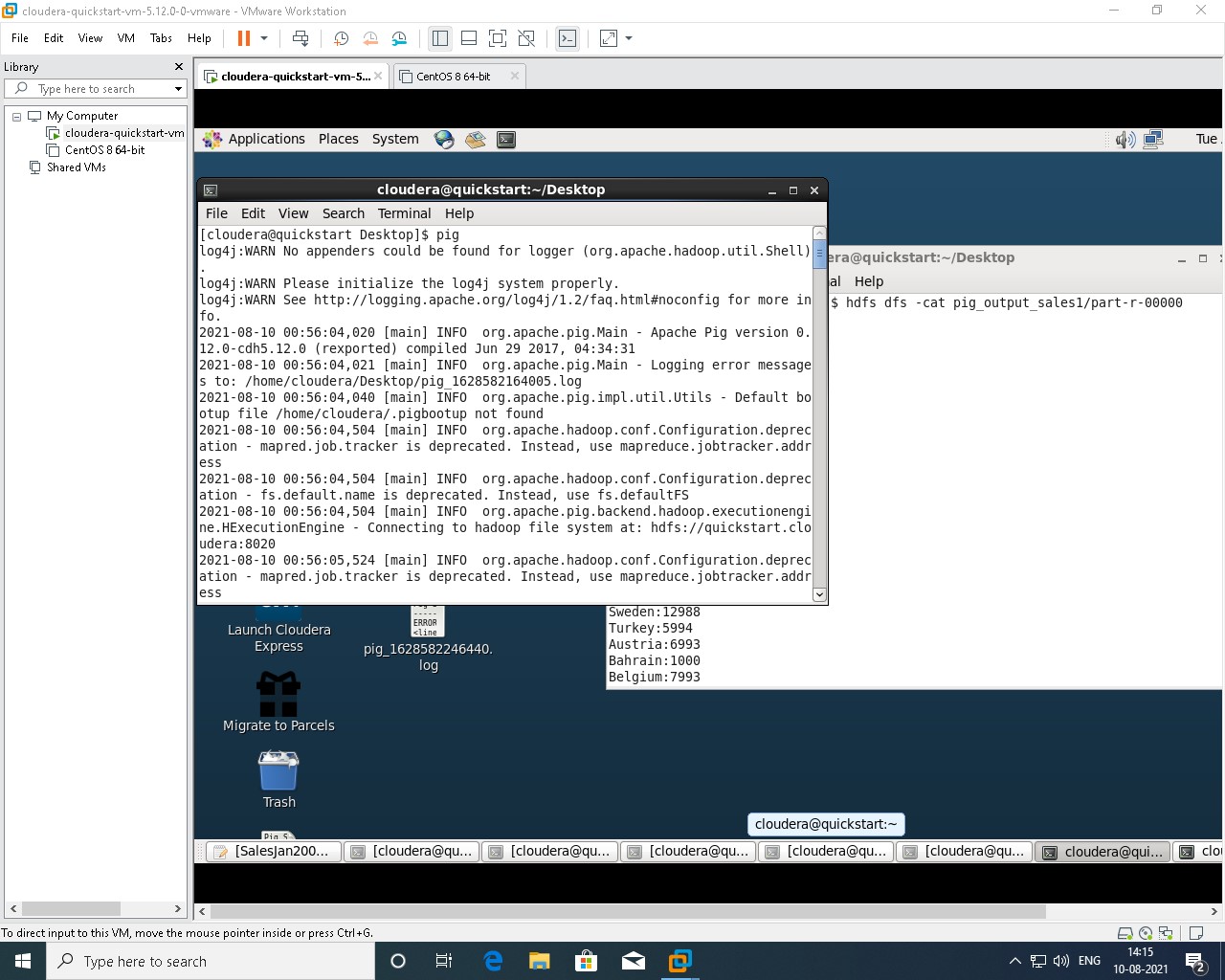
$hadoop fs -put SalesJan2009.csv smitrpatel19 ## move csv file into folder

$hadoop fs -ls smitrpatel19 ## for go inside to folder



Step 2 :-

new terminal and type --> pig

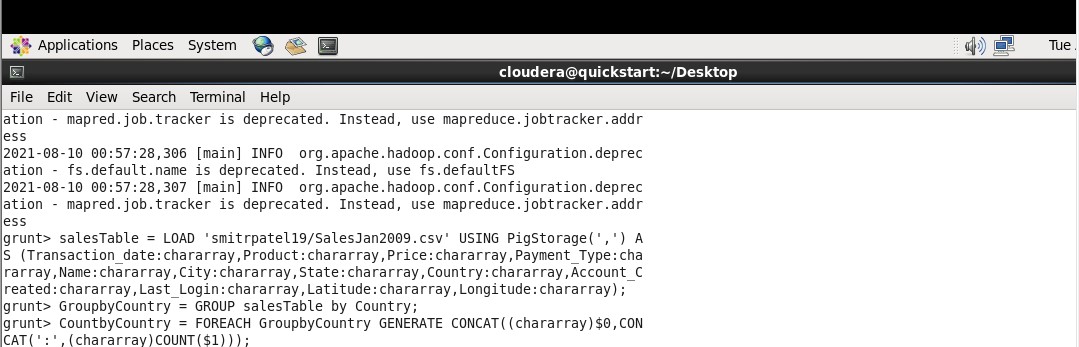


Step 3 :-

grunt> salesTable = LOAD 'smitrpatel19/SalesJan2009.csv' USING PigStorage(',') AS (Transaction\_date:chararray,Product:chararray,Price:chararray,Payment\_Type:chararray,Name:chararray,City:chararray,State:chararray,Country:chararray,Account\_Created:chararray,Last\_Login:chararray,Latitude:chararray,Longitude:chararray);

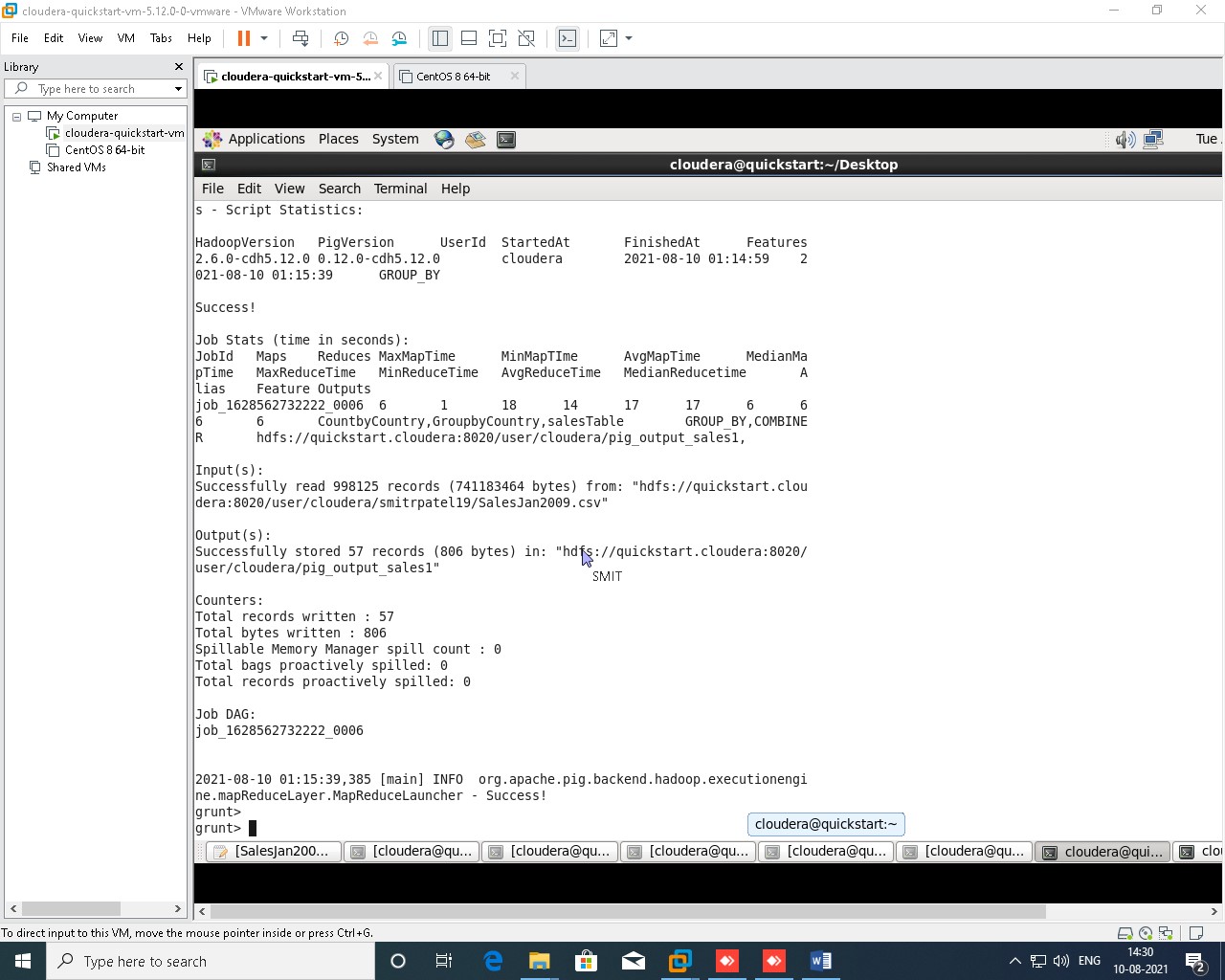
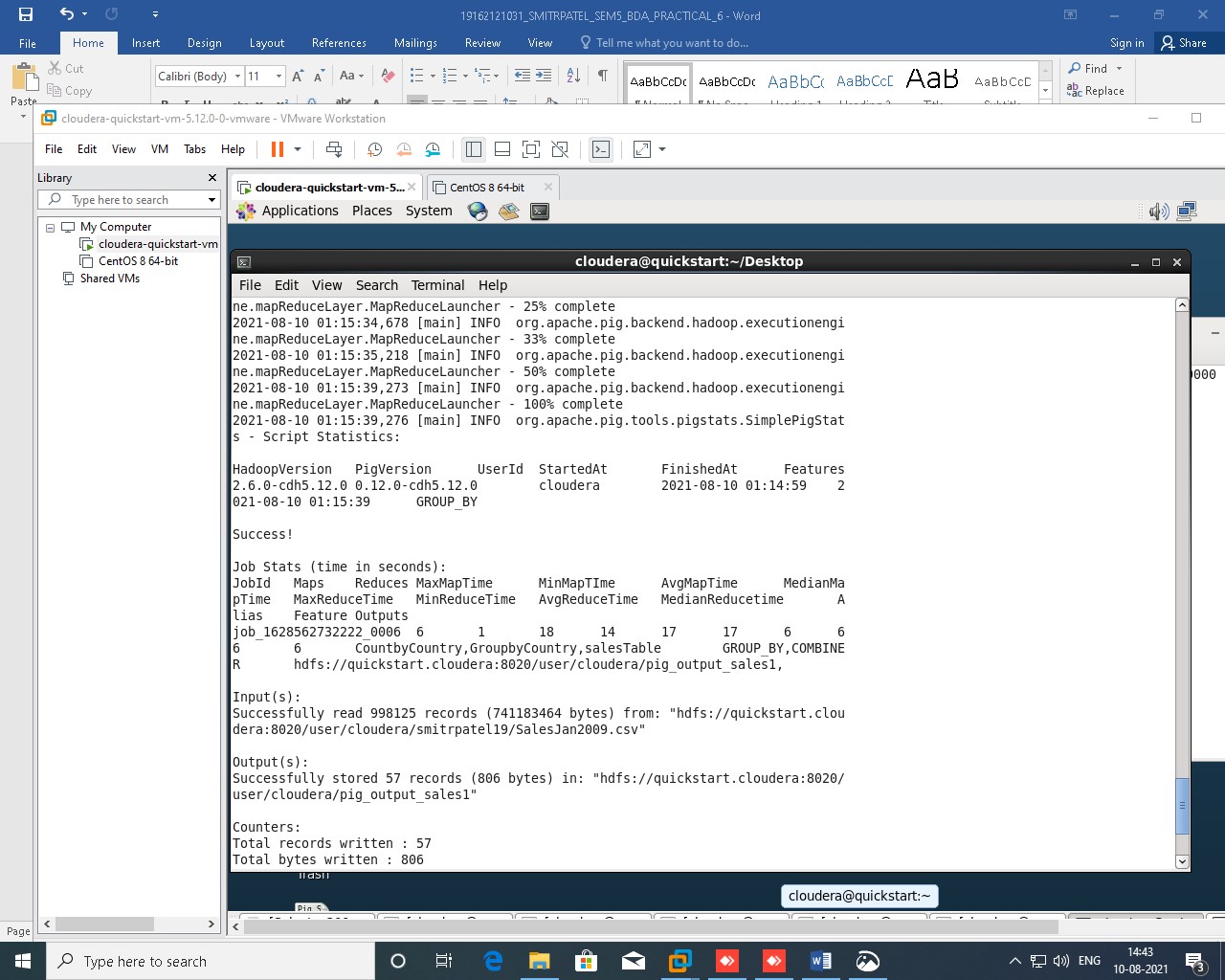
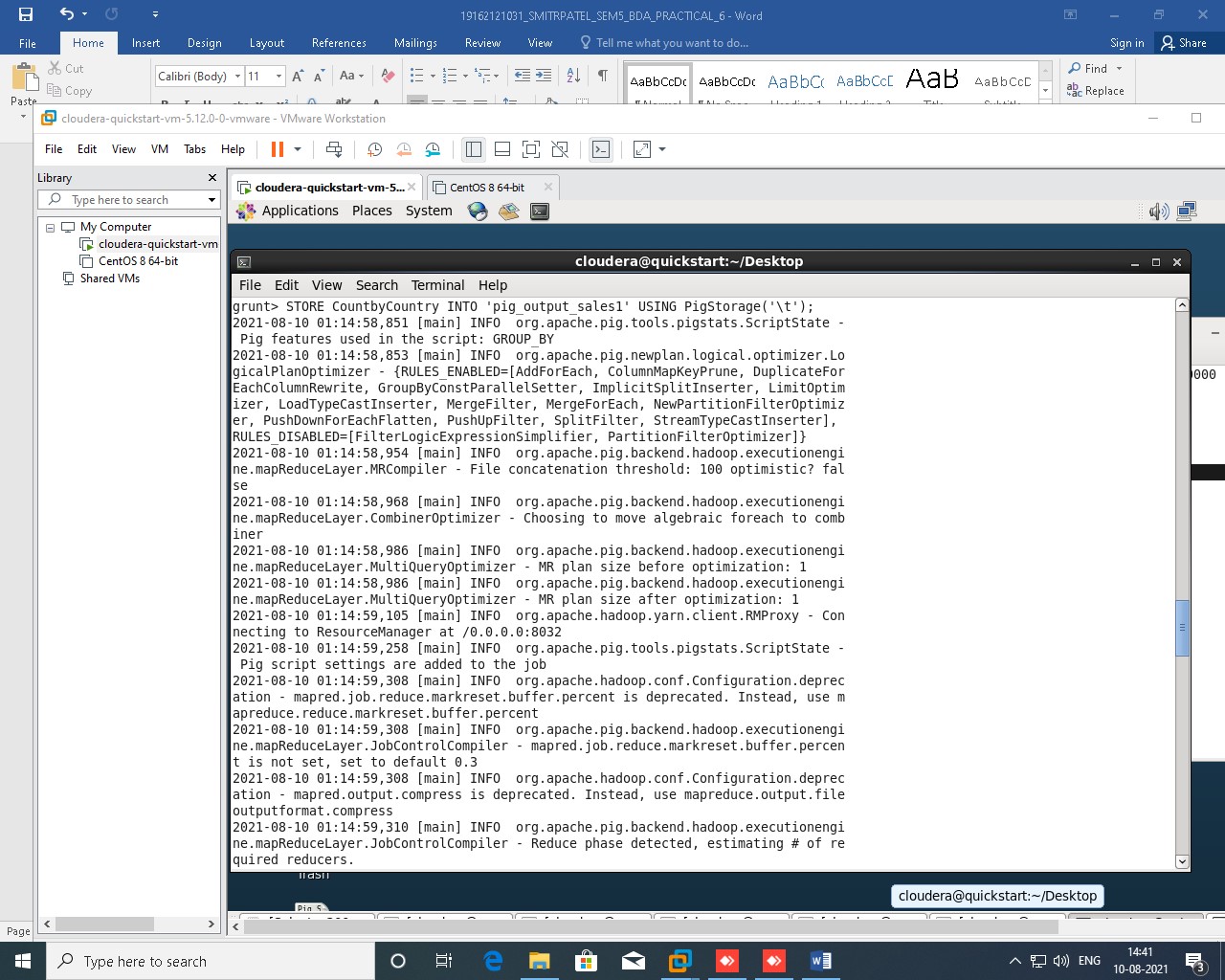
grunt> GroupbyCountry = GROUP salesTable by Country; ##load the file

grunt> CountbyCountry = FOREACH GroupbyCountry GENERATE CONCAT((chararray)$0,CONCAT(':',(chararray)COUNT($1)));



Step 4 :-

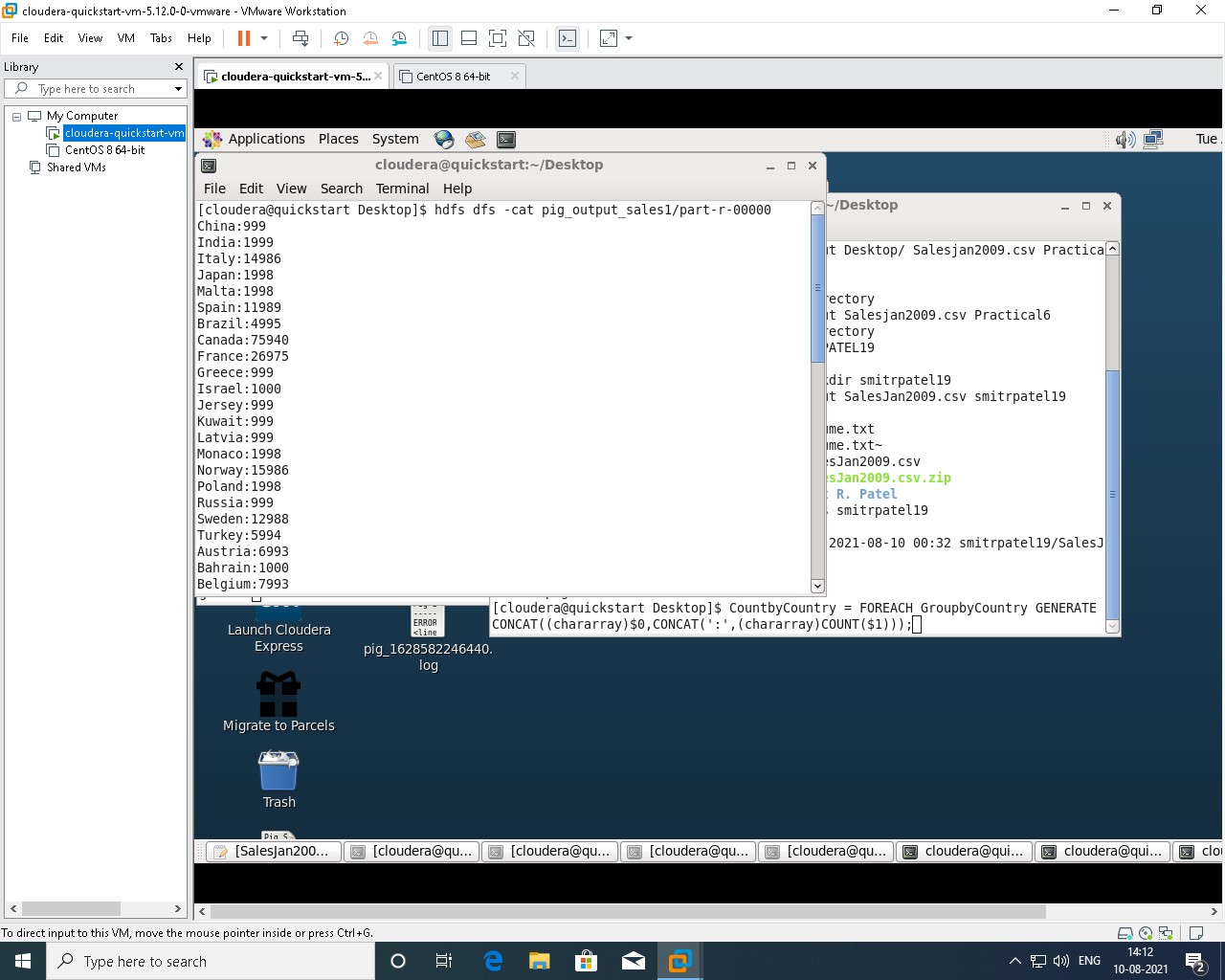
grunt> STORE CountbyCountry INTO 'pig\_output\_sales1' USING PigStorage('\t');



Final Step :-

new terminal and type -->

$hdfs dfs -cat pig\_output\_sales1/part-r-00000



Conclusion :-

Here, In this practical we learn load the file with all content – data and sort group data by field country after we generate results we show data flow in the directory ‘pig\_output  
\_sales’ and on hdfs, we get success all map reduce work flow and get final output.