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:char
array,City:chararray,State:chararray,Country:chararray,Account_Created:chararray,Last_Login:chara
rray,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
```

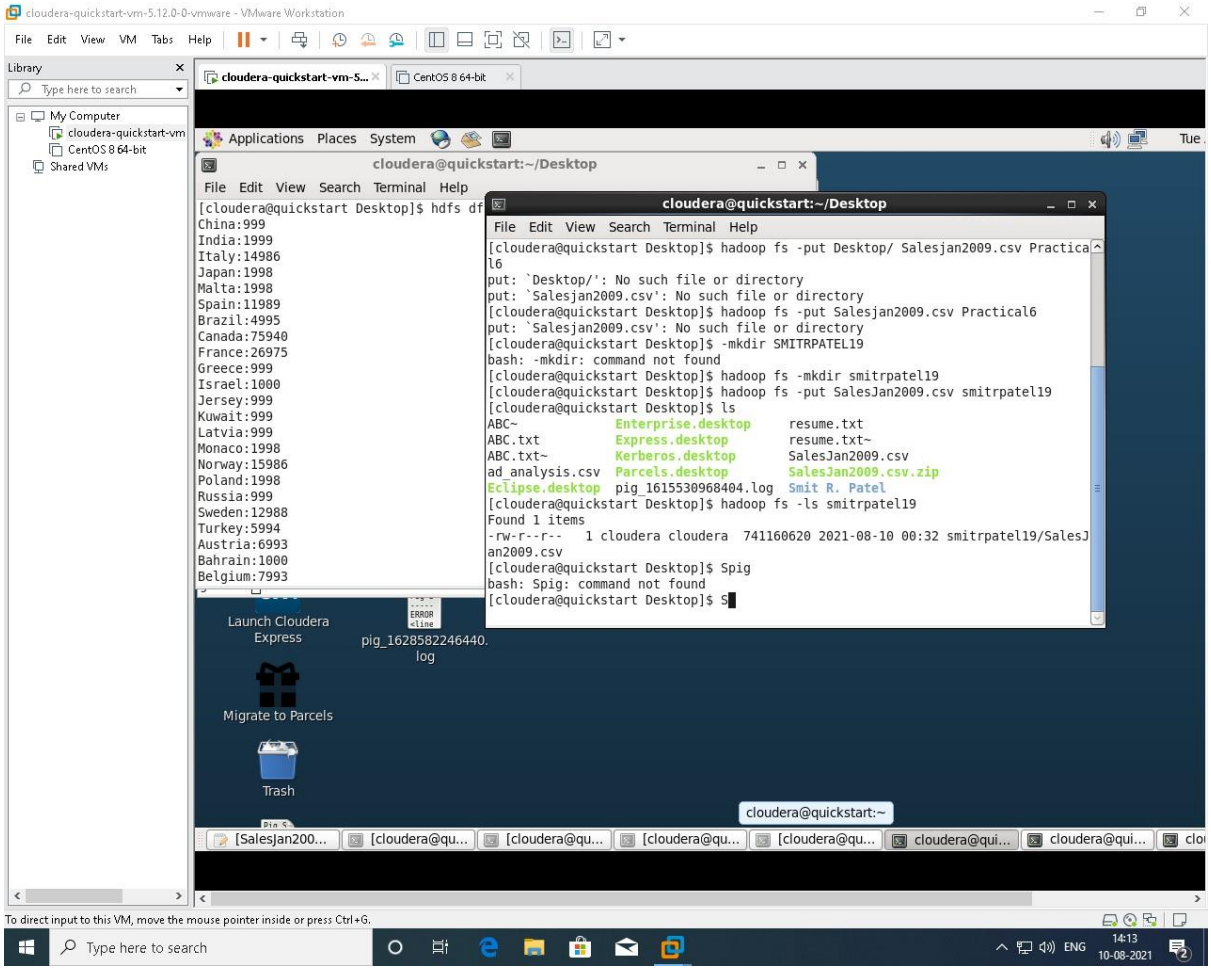
Code with output :-

Step 1 :-

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

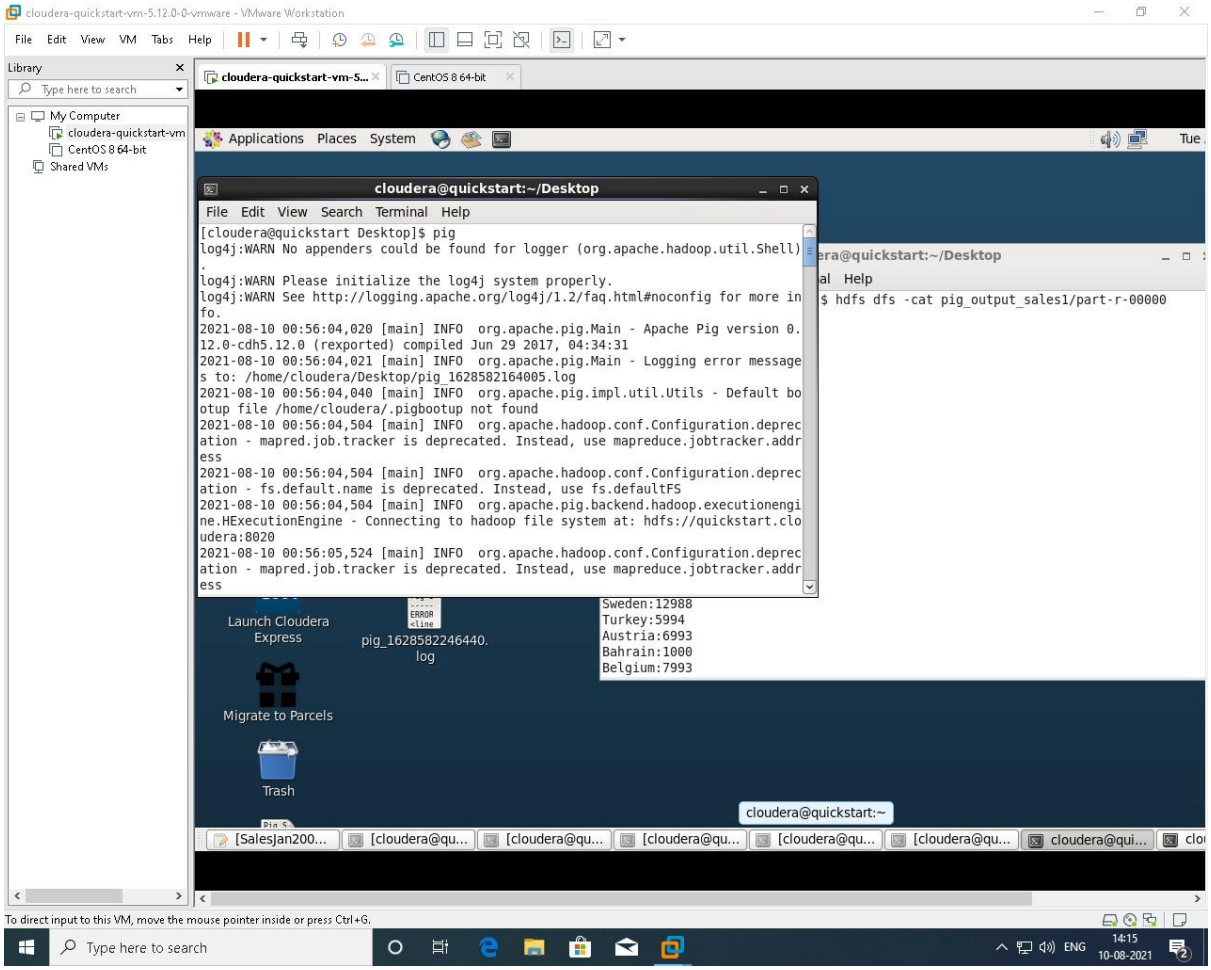
$hadoop fs -put SalesJan2009.csv smitrpatel19 ## move csv file into folder

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



Step 2 :-

new terminal and type --> pig

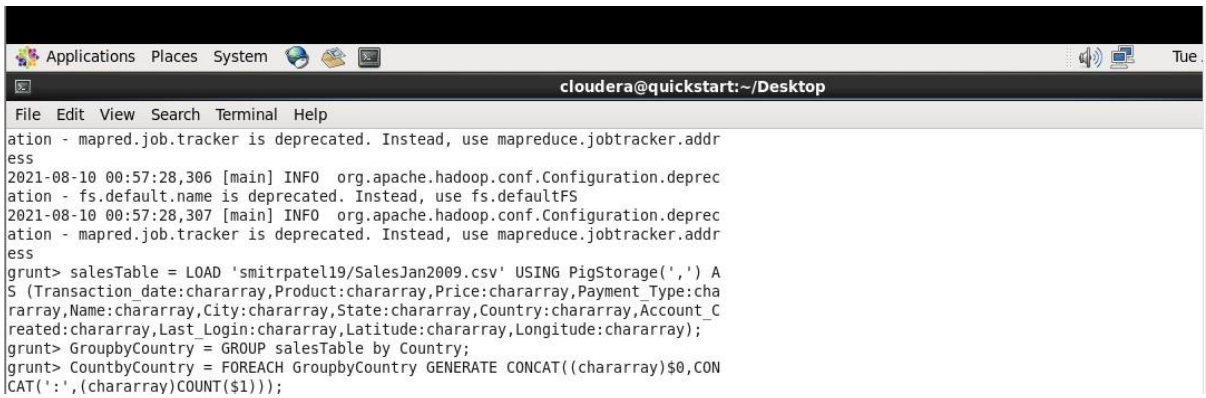


Step 3 :-

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);

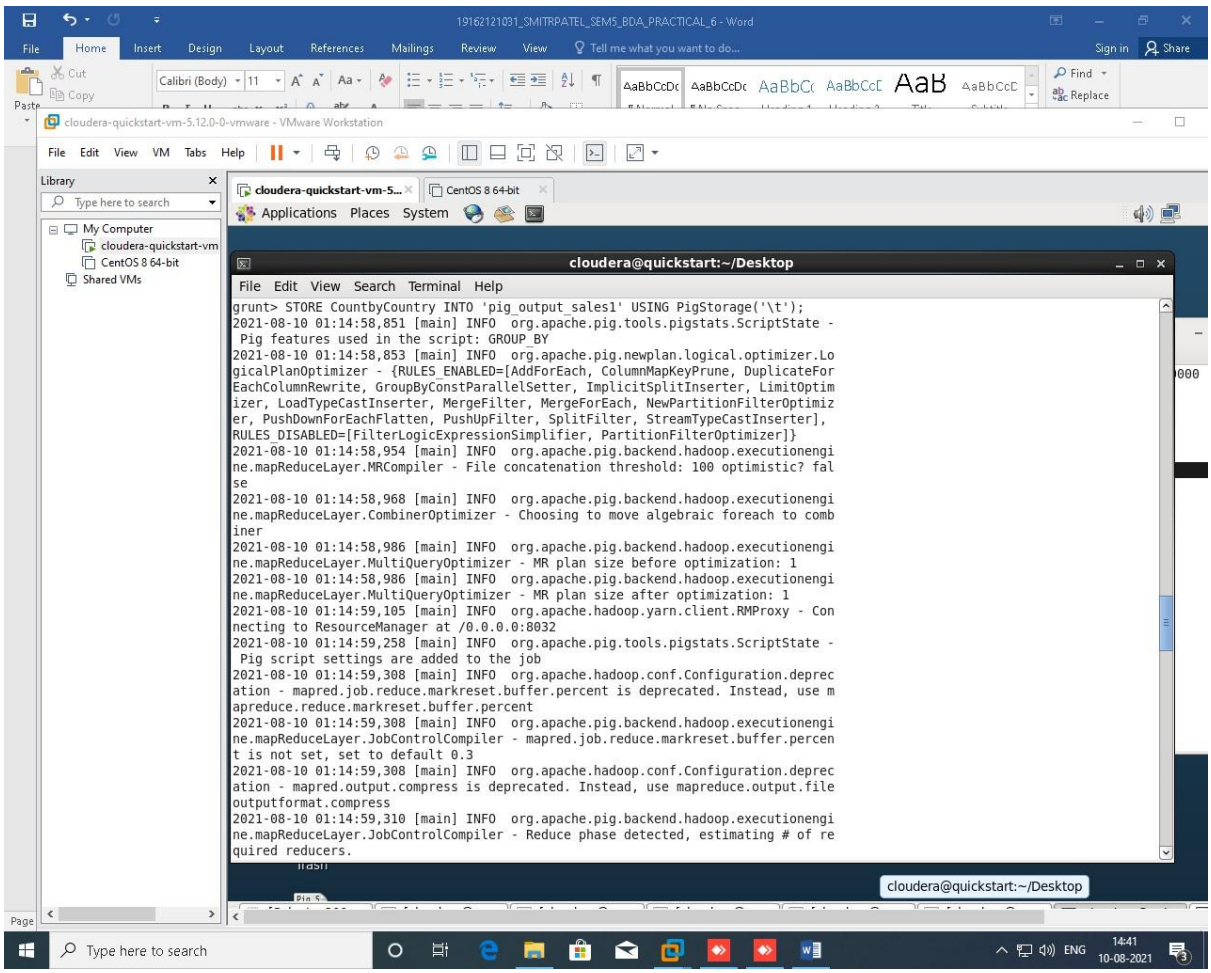
grunt> GroupbyCountry = GROUP salesTable by Country; ##load the file

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



Step 4 :-

grunt> STORE CountbyCountry INTO 'pig_output_sales1' USING PigStorage('\t');



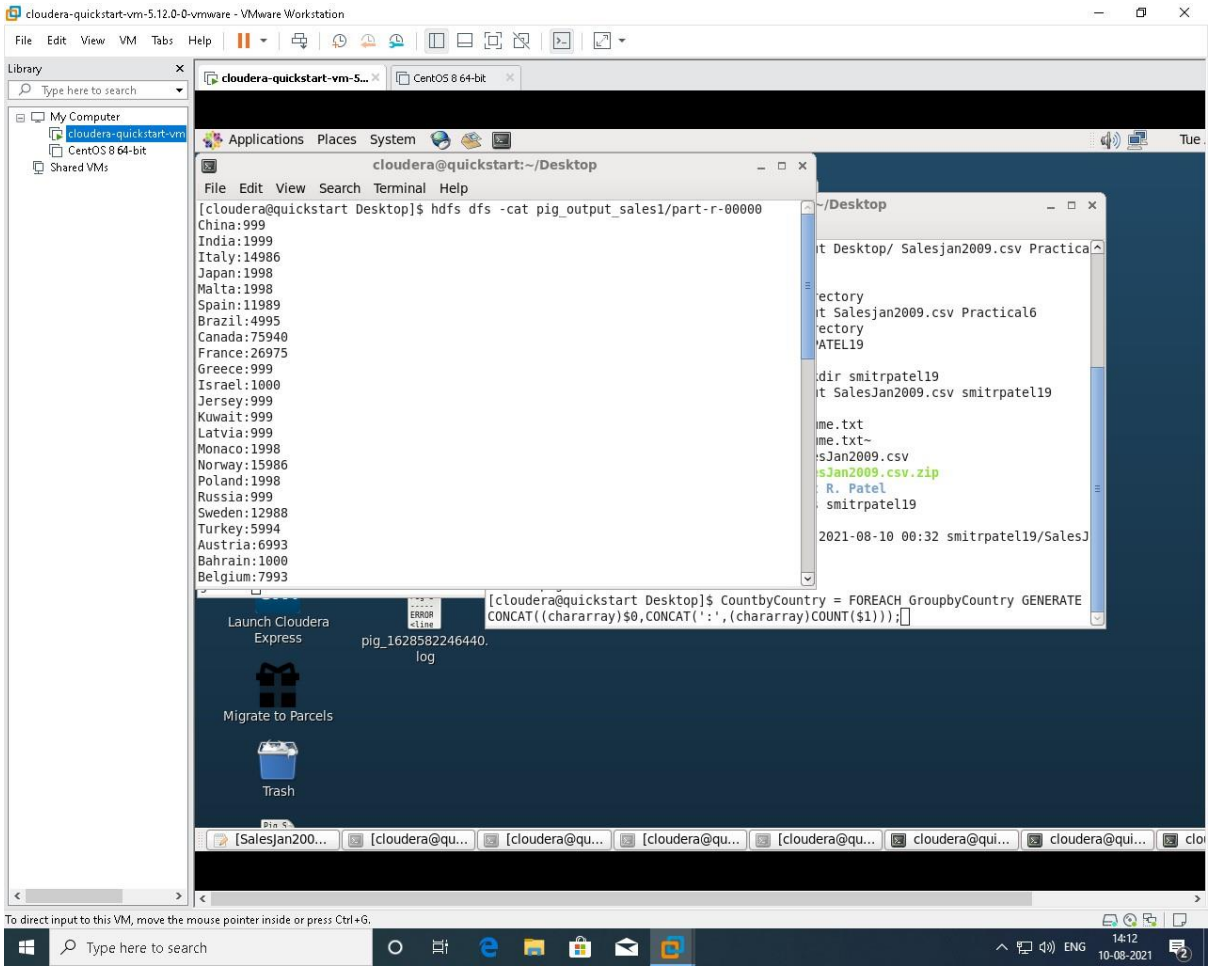
```
ne.mapReduceLayer.MapReduceLauncher - 25% complete
2021-08-10 01:15:34,678 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - 33% complete
2021-08-10 01:15:35,218 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - 50% complete
2021-08-10 01:15:39,273 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - 100% complete
2021-08-10 01:15:39,276 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStat
s - Script Statistics:

HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0        cloudera        2021-08-10 01:14:59    2
021-08-10 01:15:39      GROUP_BY

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTIme      AvgMapTime      MedianMa
pTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        A
lias    Feature Outputs
job_1628562732222_0006  6       1       18      14      17      17      6       6
6       6       CountbyCountry,GroupbyCountry,salesTable        GROUP_BY,COMBINE
R       hdfs://quickstart.cloudera:8020/user/cloudera/pig_output_sales1,

Input(s):
Successfully read 998125 records (741183464 bytes) from: "hdfs://quickstart.clou
dera:8020/user/cloudera/smitrpatel19/SalesJan2009.csv"

Output(s):
Successfully stored 57 records (806 bytes) in: "hdfs://quickstart.cloudera:8020/
user/cloudera/pig_output_sales1"

Counters:
Total records written : 57
Total bytes written : 806
```

```
s - Script Statistics:

HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0        cloudera        2021-08-10 01:14:59    2
021-08-10 01:15:39      GROUP_BY

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTIme      AvgMapTime      MedianMa
pTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime       A
lias    Feature Outputs
job_1628562732222_0006  6       1       18      14      17      17      6       6
6       6       CountbyCountry,GroupbyCountry,salesTable        GROUP_BY,COMBINE
R       hdfs://quickstart.cloudera:8020/user/cloudera/pig_output_sales1,

Input(s):
Successfully read 998125 records (741183464 bytes) from: "hdfs://quickstart.clou
dera:8020/user/cloudera/smitrpatel19/SalesJan2009.csv"

Output(s):
Successfully stored 57 records (806 bytes) in: "hdfs://quickstart.cloudera:8020/
user/cloudera/pig_output_sales1"

Counters:
Total records written : 57
Total bytes written : 806
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1628562732222_0006


2021-08-10 01:15:39,385 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
grunt>
grunt>
```

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.