

HIVE TASK

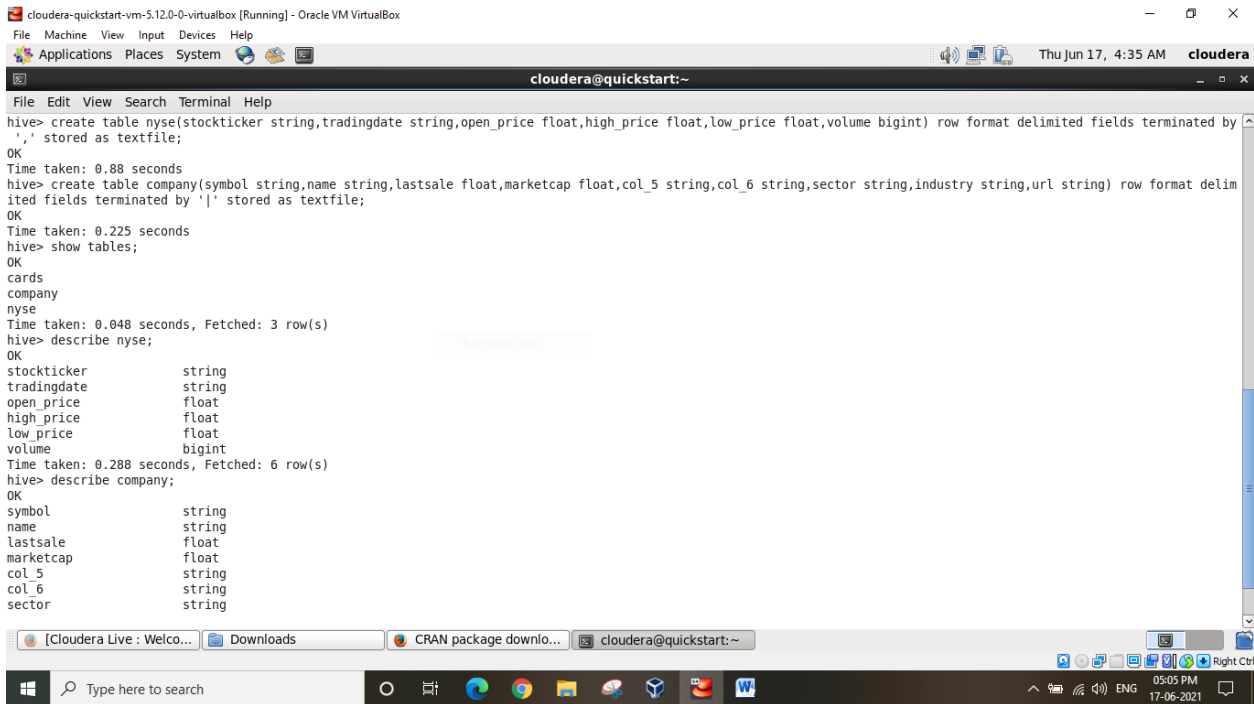
HIVE Task-1

Q: Write a query to calculate maximum no of shares traded in 2009 industry wise, sector wise and company wise.

Sample Data used: nyse_2009 and companylist_noheader

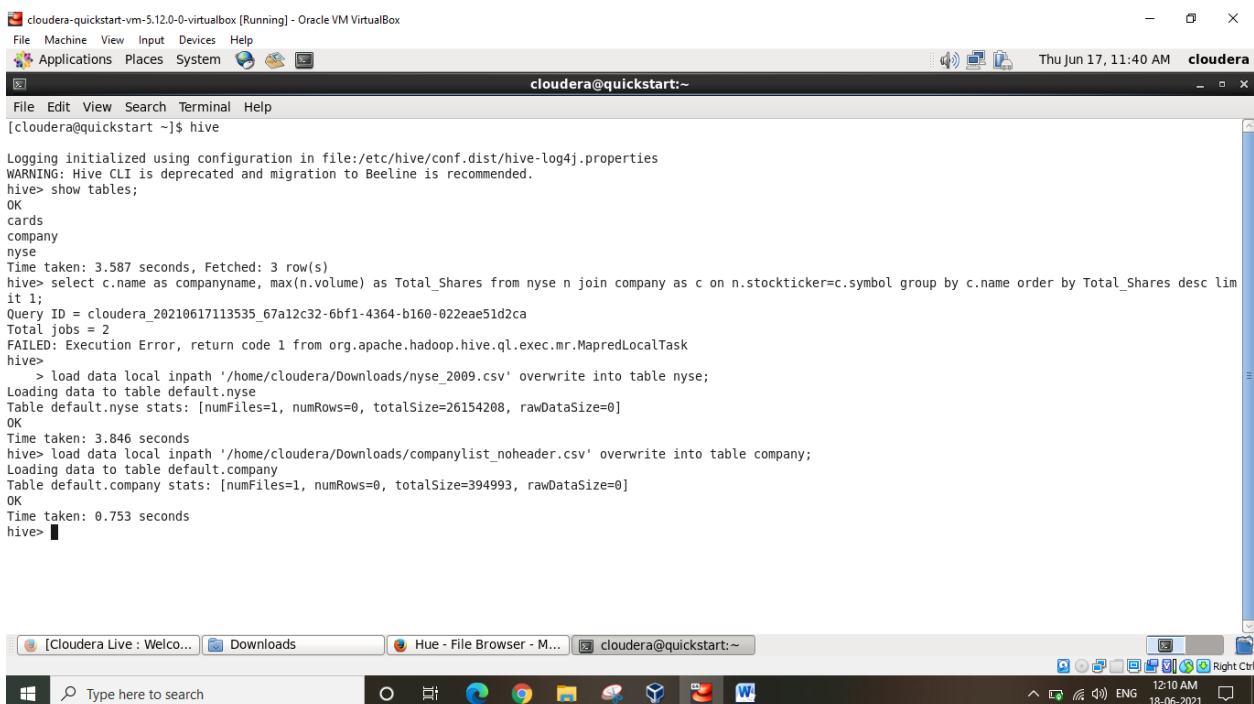
1. Create table nyse

2. Create table Company



```
cloudera-quickstart-vm-5.12.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~
File Edit View Search Terminal Help
hive> create table nyse(stockticker string,tradingdate string,open_price float,high_price float,low_price float,volume bigint) row format delimited fields terminated by ',' stored as textfile;
OK
Time taken: 0.88 seconds
hive> create table company(symbol string,name string,lastsale float,marketcap float,col_5 string,col_6 string,sector string,industry string,url string) row format delimited fields terminated by '|' stored as textfile;
OK
Time taken: 0.225 seconds
hive> show tables;
OK
cards
company
nyse
Time taken: 0.048 seconds, Fetched: 3 row(s)
hive> describe nyse;
OK
stockticker      string
tradingdate      string
open_price       float
high_price       float
low_price        float
volume           bigint
Time taken: 0.288 seconds, Fetched: 6 row(s)
hive> describe company;
OK
symbol           string
name             string
lastsale         float
marketcap        float
col_5            string
col_6            string
sector           string
```

3. load data in nyse and company table.



```
cloudera-quickstart-vm-5.12.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> show tables;
OK
cards
company
nyse
Time taken: 3.587 seconds, Fetched: 3 row(s)
hive> select c.name as companyname, max(n.volume) as Total_Shares from nyse n join company as c on n.stockticker=c.symbol group by c.name order by Total_Shares desc limit 1;
Query ID = cloudera_20210617113535_67a12c32-6bf1-4364-b160-022eae51d2ca
Total jobs = 2
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.mr.MapredLocalTask
hive>
> load data local inpath '/home/cloudera/Downloads/nyse_2009.csv' overwrite into table nyse;
Loading data to table default.nyse
Table default.nyse stats: [numFiles=1, numRows=0, totalSize=26154208, rawDataSize=0]
OK
Time taken: 3.846 seconds
hive> load data local inpath '/home/cloudera/Downloads/companylist_noheader.csv' overwrite into table company;
Loading data to table default.company
Table default.company stats: [numFiles=1, numRows=0, totalSize=394993, rawDataSize=0]
OK
Time taken: 0.753 seconds
hive>
```

4. Write a query to calculate maximum no of shares traded in 2009 Industry wise.

```
cloudera-quickstart-vm-5.12.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> show tables;
OK
cards
company
nyse
Time taken: 3.587 seconds, Fetched: 3 row(s)
hive> select c.name as companyname, max(n.volume) as Total_Shares from nyse n join company as c on n.stockticker=c.symbol group by c.name order by Total_Shares desc limit 1;
Query ID = cloudera_20210617113535_67a12c32-6bf1-4364-b160-022eae51d2ca
Total jobs = 2
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.mr.MapredLocalTask
hive>
```

5. Write a query to calculate maximum no of shares traded in 2009 Sector wise.

6. Write a query to calculate maximum no of shares traded in 2009 Company wise