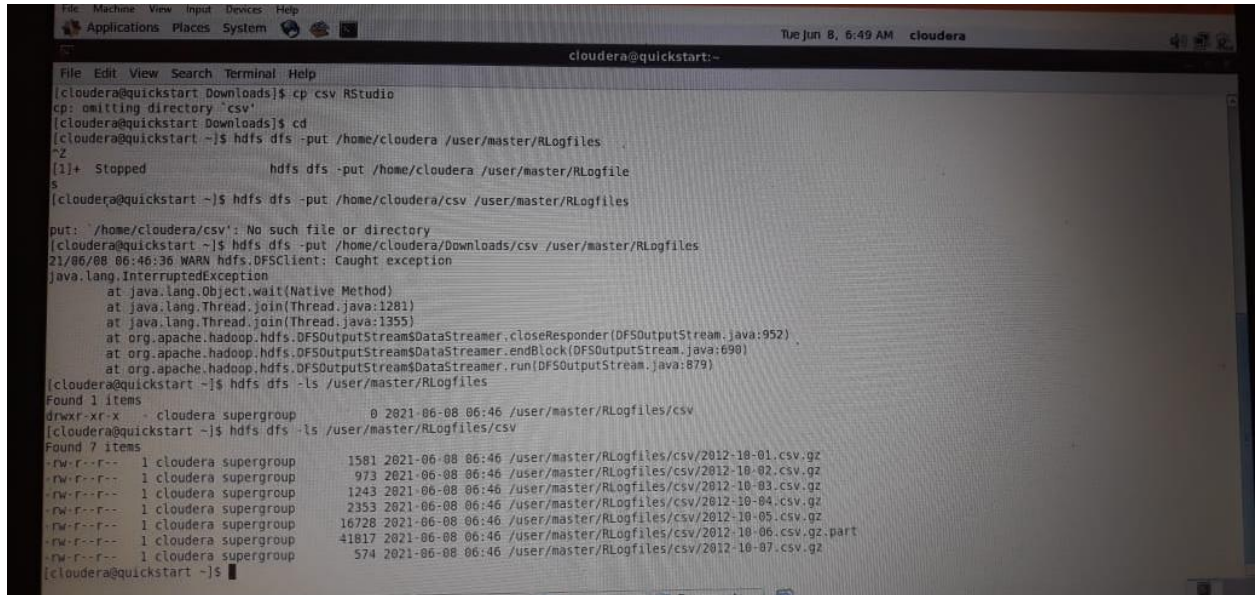
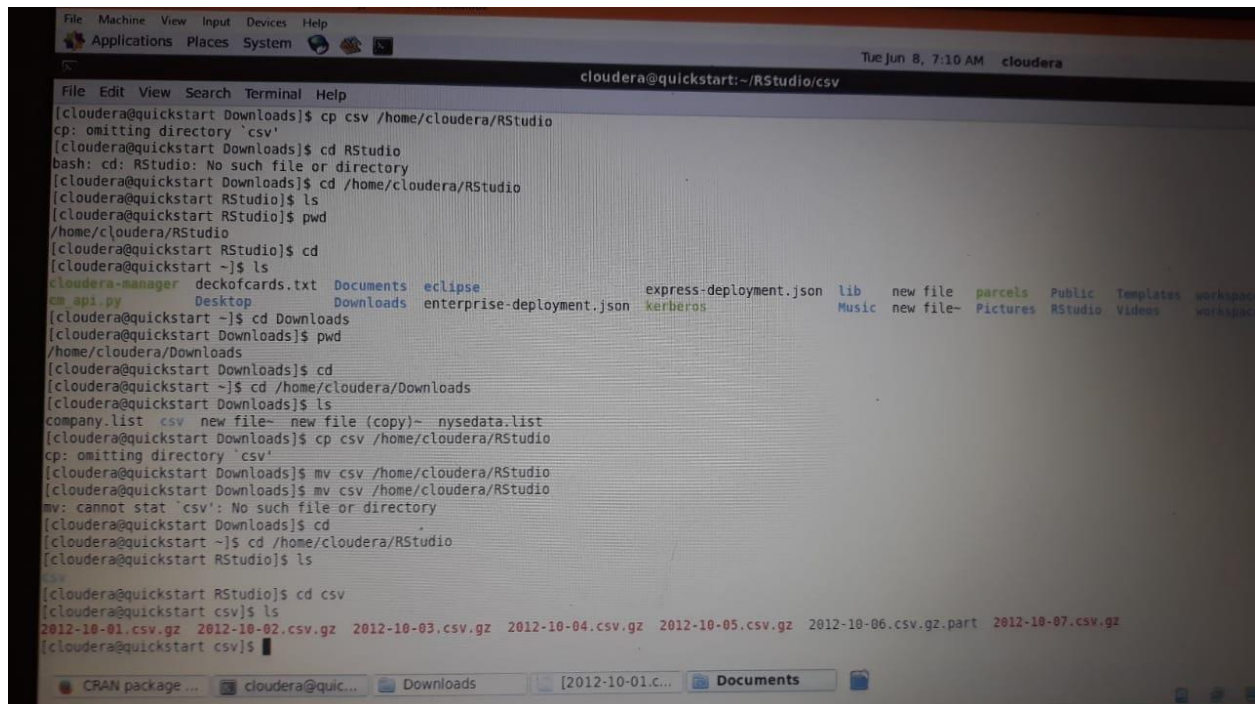


Hive Task 2:

1. Import RStudio Log Files from into HDFS.



```
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart Downloads]$ cp csv RStudio
cp: omitting directory 'csv'
[cloudera@quickstart Downloads]$ cd
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/user/master/RLogfiles
1+ Stopped hdfs dfs -put /home/cloudera/user/master/RLogfile
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/csv/user/master/RLogfiles
put: /home/cloudera/csv: No such file or directory
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Downloads/csv/user/master/RLogfiles
21/06/08 06:46:36 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
[cloudera@quickstart ~]$ hdfs dfs -ls /user/master/RLogfiles
Found 1 items
drwxr-xr-x 1 cloudera supergroup 0 2021-06-08 06:46 /user/master/RLogfiles/csv
[cloudera@quickstart ~]$ hdfs dfs -ls /user/master/RLogfiles/csv
Found 7 items
-rw-r--r-- 1 cloudera supergroup 1581 2021-06-08 06:46 /user/master/RLogfiles/csv/2012-10-01.csv.gz
-rw-r--r-- 1 cloudera supergroup 973 2021-06-08 06:46 /user/master/RLogfiles/csv/2012-10-02.csv.gz
-rw-r--r-- 1 cloudera supergroup 1243 2021-06-08 06:46 /user/master/RLogfiles/csv/2012-10-03.csv.gz
-rw-r--r-- 1 cloudera supergroup 2353 2021-06-08 06:46 /user/master/RLogfiles/csv/2012-10-04.csv.gz
-rw-r--r-- 1 cloudera supergroup 16728 2021-06-08 06:46 /user/master/RLogfiles/csv/2012-10-05.csv.gz
-rw-r--r-- 1 cloudera supergroup 41817 2021-06-08 06:46 /user/master/RLogfiles/csv/2012-10-06.csv.gz.part
-rw-r--r-- 1 cloudera supergroup 574 2021-06-08 06:46 /user/master/RLogfiles/csv/2012-10-07.csv.gz
[cloudera@quickstart ~]$
```



```
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:~/RStudio/csv
File Edit View Search Terminal Help
[cloudera@quickstart Downloads]$ cp csv /home/cloudera/RStudio
cp: omitting directory 'csv'
[cloudera@quickstart Downloads]$ cd RStudio
bash: cd: RStudio: No such file or directory
[cloudera@quickstart Downloads]$ cd /home/cloudera/RStudio
[cloudera@quickstart RStudio]$ ls
[cloudera@quickstart RStudio]$ pwd
/home/cloudera/RStudio
[cloudera@quickstart RStudio]$ cd
[cloudera@quickstart ~]$ ls
cloudera-manager  deckofcards.txt  Documents  eclipse  express-deployment.json  lib  new file  parcels  Public  Templates  workshop
api.py  Desktop  Downloads  enterprise-deployment.json  kerberos  Music  new file-  Pictures  RStudio  Videos  workshop
[cloudera@quickstart ~]$ cd Downloads
[cloudera@quickstart Downloads]$ pwd
/home/cloudera/Downloads
[cloudera@quickstart Downloads]$ cd
[cloudera@quickstart ~]$ cd /home/cloudera/Downloads
[cloudera@quickstart Downloads]$ ls
company.list  csv  new file-  new file (copy)-  nysedata.list
[cloudera@quickstart Downloads]$ cp csv /home/cloudera/RStudio
cp: omitting directory 'csv'
[cloudera@quickstart Downloads]$ mv csv /home/cloudera/RStudio
mv: cannot stat 'csv': No such file or directory
[cloudera@quickstart Downloads]$ cd
[cloudera@quickstart ~]$ cd /home/cloudera/RStudio
[cloudera@quickstart RStudio]$ ls
csv
[cloudera@quickstart RStudio]$ cd csv
[cloudera@quickstart csv]$ ls
2012-10-01.csv.gz  2012-10-02.csv.gz  2012-10-03.csv.gz  2012-10-04.csv.gz  2012-10-05.csv.gz  2012-10-06.csv.gz.part  2012-10-07.csv.gz
[cloudera@quickstart csv]$
```

Dump the first 10 entries on screen (attach a screen shot into your report) to check if it works or not

```
cloudera@quickstart:~
File Edit View Search Terminal Help
OK
Time taken: 0.3 seconds
hive> load data local inpath '/home/cloudera/Downloads/nysedata.list' into table nyse;
Loading data to table default.nyse
Table default.nyse stats: [numFiles=1, totalSize=25586560]
OK
Time taken: 1.887 seconds
hive> select * from nyse limit 5;
OK
A,01-Jan-2009,15.63,15.63,15.63,15.63,0 NULL NULL NULL NULL NULL NULL
AA,01-Jan-2009,11.26,11.26,11.26,11.26,0 NULL NULL NULL NULL NULL NULL
AAP,01-Jan-2009,33.65,33.65,33.65,33.65,0 NULL NULL NULL NULL NULL NULL
AAV,01-Jan-2009,4.21,4.21,4.21,4.21,0 NULL NULL NULL NULL NULL NULL
AB,01-Jan-2009,20.79,20.79,20.79,20.79,0 NULL NULL NULL NULL NULL NULL
Time taken: 0.141 seconds, Fetched: 5 row(s)
hive> create external table company(symbol string,name string,lastsale decimal,marketcap decimal,col_5 string,col_6 string,sector string,industry string,url string)
> row format delimited fields terminated by '|'
> stored as textfile location '/home/cloudera/Downloads/company.list';
OK
Time taken: 0.317 seconds
hive> load data local inpath '/home/cloudera/Downloads/company.list' into table company;
Loading data to table default.company
Table default.company stats: [numFiles=1, totalSize=391696]
OK
Time taken: 0.714 seconds
hive> select * from company limit 5;
OK
DDD 3M Systems Corporation 51 5340931605 n/a n/a Technology Computer Software: Prepackaged Software http://www.nasdaq.com/symbol/ddd
MMM 3M Company 142 NULL n/a n/a Health Care Medical/Dental Instruments http://www.nasdaq.com/symbol/mmm
WBAI 500.com Limited 34 1108170573 n/a 2013 Consumer Services Services-Misc, Amusement & Recreation http://www.nasdaq.com/symbol/wbai
WUBA 58.com Inc. 42 3302252043 n/a 2013 Technology Computer Software: Programming, Data Processing http://www.nasdaq.com/symbol/wuba
AHC A.H. Belo Corporation 12 256920010 n/a n/a Consumer Services Newspapers/Magazines http://www.nasdaq.com/symbol/ahc
```

```
cloudera@quickstart:~
File Edit View Search Terminal Help
2021-06-06 06:41:29,684 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.64
sec
MapReduce Total cumulative CPU time: 8 seconds 640 msec
Ended Job = job_1622984960986_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.64 sec HDFS Read: 923B HD
FS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 640 msec
OK
BLACK 2
Time taken: 105.112 seconds, Fetched: 1 row(s)
hive> select * from cards limit 5;
OK
BLACK SPADE 2
BLACK SPADE 3
BLACK SPADE 4
BLACK SPADE 5
BLACK SPADE 6
Time taken: 0.226 seconds, Fetched: 5 row(s)
hive> create external table cards1(color string,suit string,type string)
> row format delimited fields terminated by '|'
> stored as textfile location /user/firstdir/deckofcardtxt;
MismatchedTokenException(15!=307)
at org.apache.hadoop.hive.ql.parse.HiveParser.createTableStatement(HiveParser.java:5250)
at org.apache.hadoop.hive.ql.parse.HiveParser.ddlStatement(HiveParser.java:1589)
at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.java:1065)
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:201)
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:166)
```

```
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
Loading data to table default.cards
Table default.cards stats: [numFiles=1, totalSize=694]
OK
Time taken: 2.266 seconds
hive> show tables;
OK
cards
Time taken: 0.135 seconds, Fetched: 1 row(s)
hive> select * from cards;
OK
BLACK SPADE 2
BLACK SPADE 3
BLACK SPADE 4
BLACK SPADE 5
BLACK SPADE 6
BLACK SPADE 7
BLACK SPADE 8
BLACK SPADE 9
BLACK SPADE 10
BLACK SPADE J
BLACK SPADE Q
BLACK SPADE K
BLACK SPADE A
BLACK CLUB 2
BLACK CLUB 3
BLACK CLUB 4
BLACK CLUB 5
BLACK CLUB 6
BLACK CLUB 7
BLACK CLUB 8
BLACK CLUB 9
BLACK CLUB 10
```

```
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
BLACK 26
RED 26
Time taken: 82.003 seconds, Fetched: 3 row(s)
hive> select color,count(*) from cards where type='Q' and color='BLACK' group by
color,type;
FAILED: ParseException line 1:72 missing BY at 'bycolor' near '<EOF>'
hive> select color,count(*) from cards where type='Q' and color="BLACK" group by
color,type;
FAILED: ParseException line 1:72 missing BY at 'bycolor' near '<EOF>'
hive> select color,count(*) from cards where type='Q' and color="BLACK" group by
color,type;
Query ID = cloudera_20210606063939_9d18c6fe-92e6-44f7-a6d9-5305e26a7290
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1622984960986_0002, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1622984960986_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1622984960986_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-06 06:40:27,049 Stage-1 map = 0%, reduce = 0%
2021-06-06 06:41:06,079 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.59 se
c
2021-06-06 06:41:29,684 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.64
sec
MapReduce Total cumulative CPU time: 8 seconds 640 msec
Ended Job = job_1622984960986_0002
```



```
Applications Places System cloudera@quickstart:~
Sun Jun 6, 8:39 AM cloudera
File Edit View Search Terminal Help
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 3:28 mismatched input '/' expecting StringLiteral near 'location' in table location specification
hive> create external table cards1(color string,suit string,type string)
> row format delimited fields terminated by '|'
>
>
>
>
>
>
> stored as textfile location '/user/firstdir/deckofcards.txt';
OK
Time taken: 0.233 seconds
hive>
>
>
>
>
>
> select * from cards1 limit 5;
OK
Time taken: 0.191 seconds
hive> create external table cards1(color string,suit string,type string)
> row format delimited fields terminated by '|'
> stored as textfile location '/home/cloudera/Downloads/company.list';
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive ql.exec.DDLTask. AlreadyExistsException(message:Table cards1 already exists)
hive> create external table nyse(stockticker string,tradingdate string,open_price float,high_price float,low_price float,close_price float,volume bigint)
> row format delimited fields terminated by '|'
> stored as textfile location '/home/cloudera/Downloads/nysedata.list';
Mozilla Firefox cloudera@quickstart:~ Downloads nysedata.list (~/Dow...
```