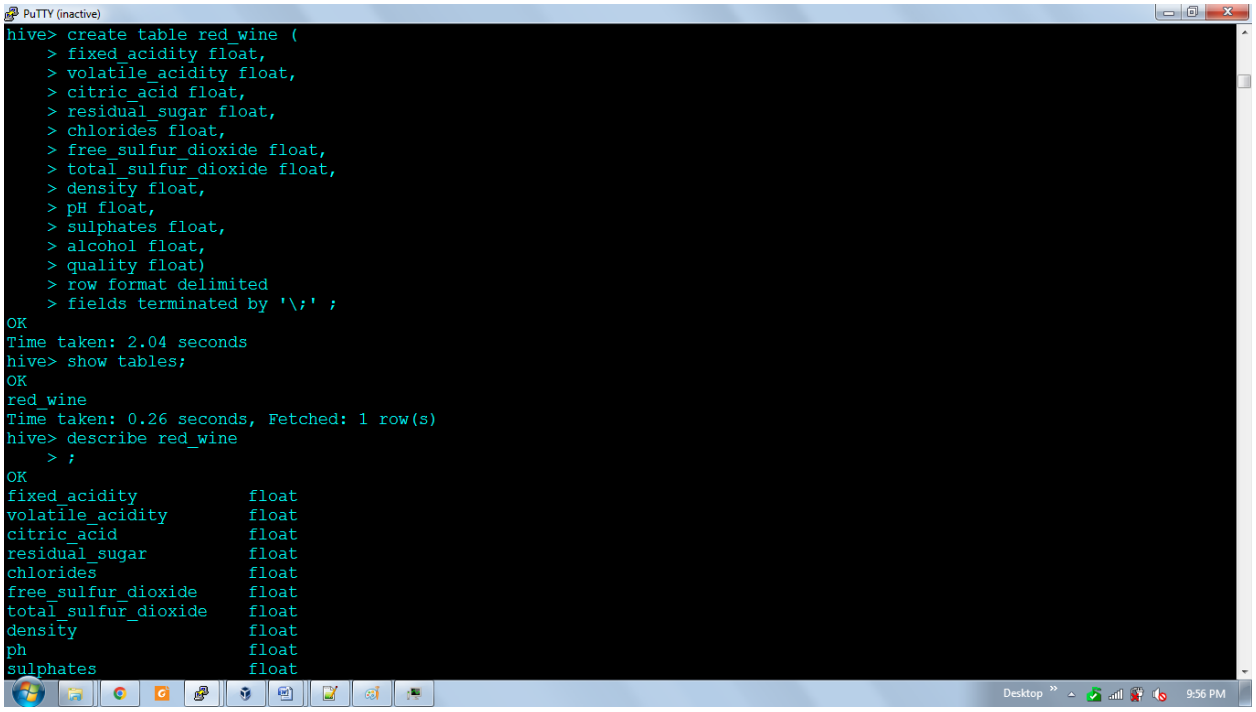
Big Data Assignment - 2

Copied two .csv files from Local to HDFS:

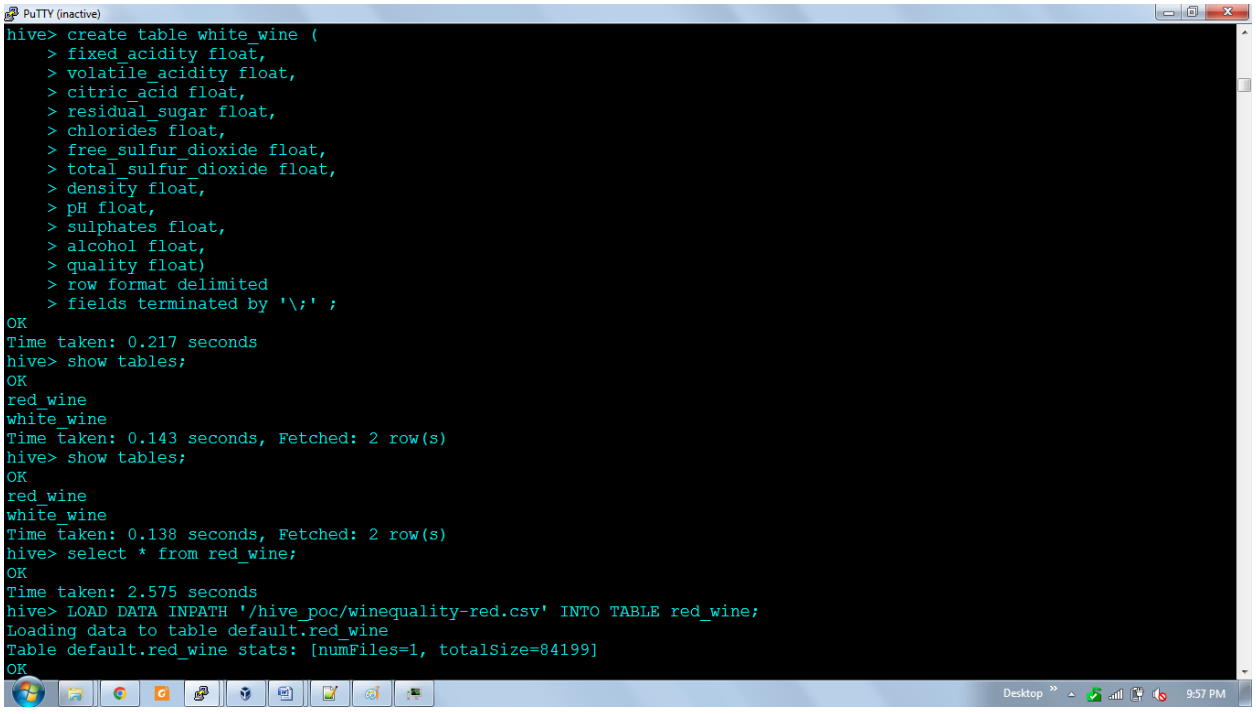
```
[cloudera@quickstart winequality]$ hadoop fs -mkdir /hive_poc
[cloudera@quickstart winequality]$ ls -l
total 352
-rwxrwxrwx 1 root root   2838 Apr 30 08:40 winequality-names.txt
-rwxrwxrwx 1 root root   2838 Apr 30 08:40 winequality-names.txt.bak
-rwxrwxrwx 1 root root  84199 Apr 30 08:40 winequality-red.csv
-rwxrwxrwx 1 root root 264426 Apr 30 08:40 winequality-white.csv
[cloudera@quickstart winequality]$ hadoop -put *.csv /hive_poc
Error: No command named `-put' was found. Perhaps you meant `hadoop put'
[cloudera@quickstart winequality]$ hadoop fs -put *.csv /hive_poc
[cloudera@quickstart winequality]$ hadoop fs -ls /hive_poc
Found 2 items
-rw-r--r--   1 cloudera supergroup      84199 2022-04-30 08:43 /hive_poc/winequality-red.csv
-rw-r--r--   1 cloudera supergroup     264426 2022-04-30 08:43 /hive_poc/winequality-white.csv
[cloudera@quickstart winequality]$ hive
```

```
hive> create database ronak;
OK
Time taken: 56.487 seconds
hive> show databases;
OK
default
ronak
Time taken: 0.162 seconds, Fetched: 2 row(s)
```

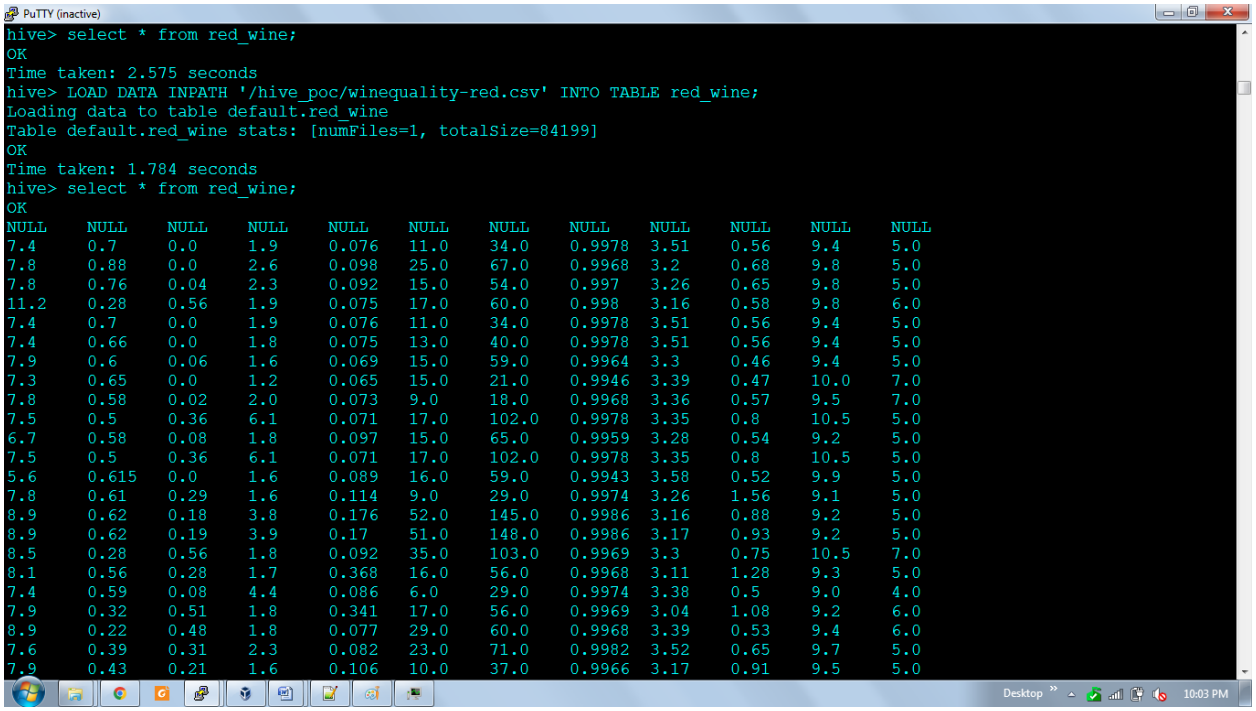
```
create table red_wine (
fixed_acidity float,
volatile_acidity float,
citric_acid float,
residual_sugar float,
chlorides float,
free_sulfur_dioxide float,
total_sulfur_dioxide float,
density float,
pH float,
sulphates float,
alcohol float,
quality float)
row format delimited
fields terminated by '\;' ;
```



```
create table white_wine (
fixed_acidity float,
volatile_acidity float,
citric_acid float,
residual_sugar float,
chlorides float,
free_sulfur_dioxide float,
total_sulfur_dioxide float,
density float,
pH float,
sulphates float,
alcohol float,
quality float)
row format delimited
fields terminated by '\;' ;
```

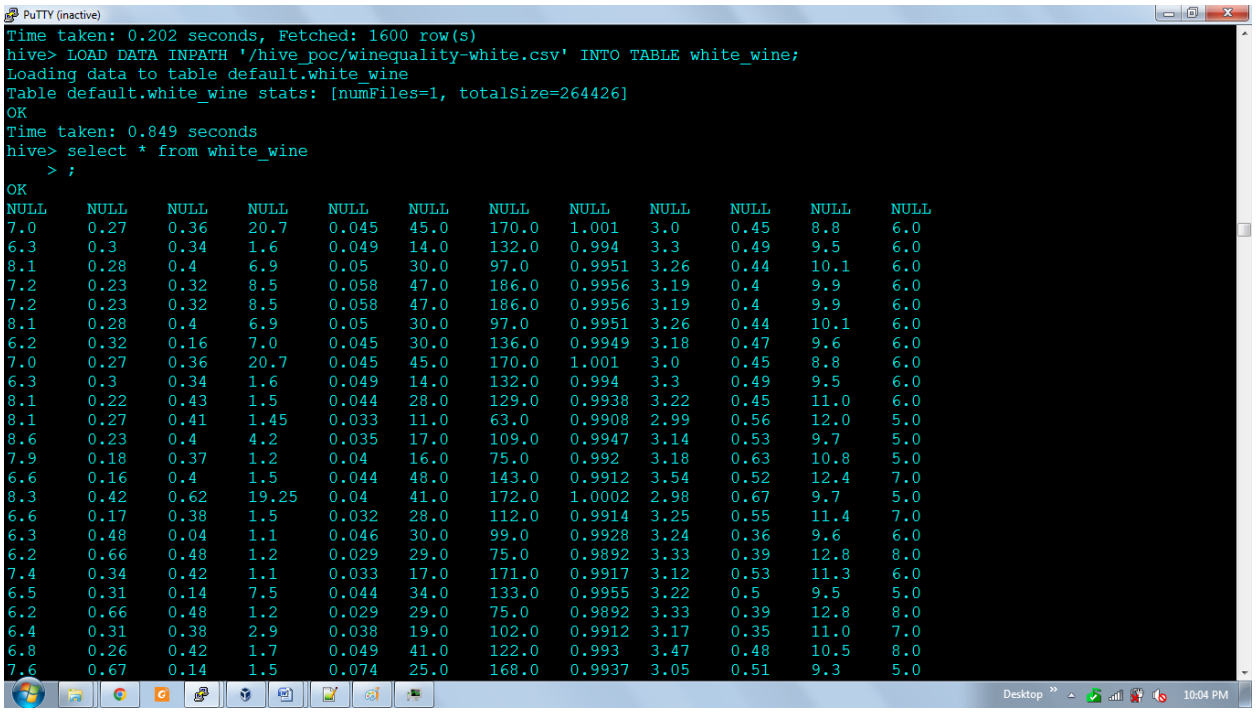


hive> LOAD DATA INPATH '/hive_poc/winequality-red.csv' INTO TABLE red_wine;



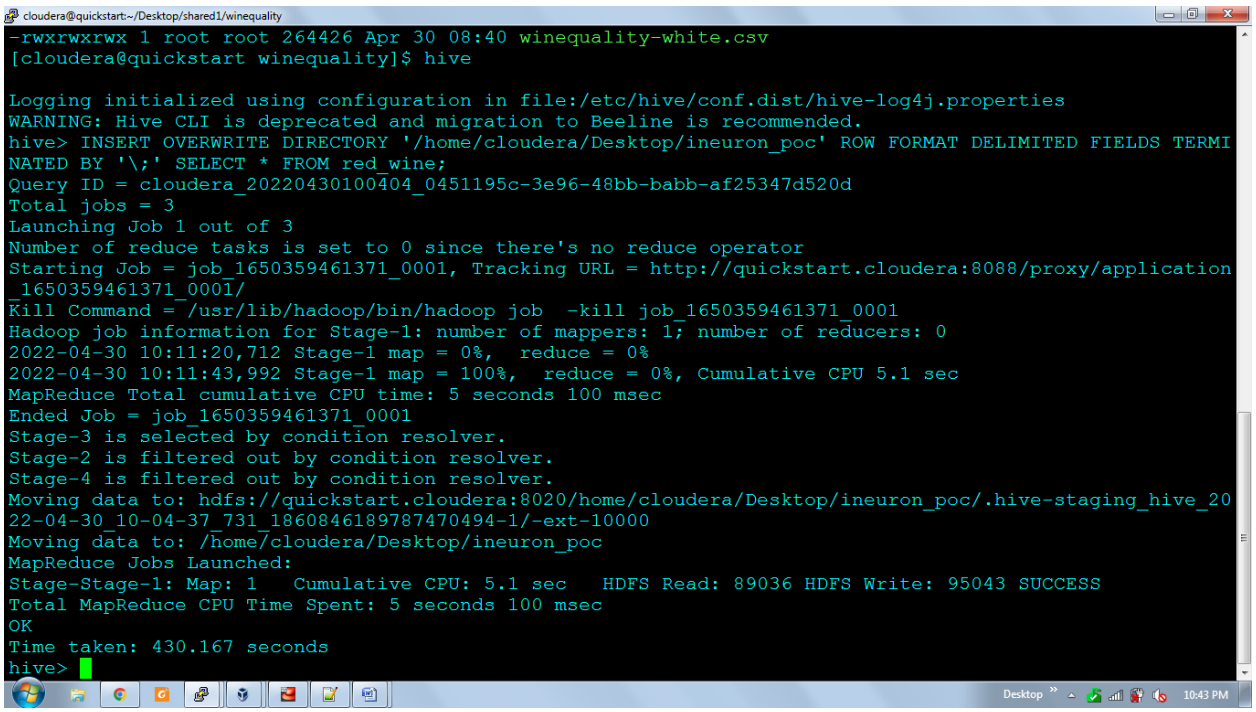
```
PuTTY (inactive)
hive> select * from red_wine;
OK
Time taken: 2.575 seconds
hive> LOAD DATA INPATH '/hive_poc/winequality-red.csv' INTO TABLE red_wine;
Loading data to table default.red_wine
Table default.red_wine stats: [numFiles=1, totalSize=84199]
OK
Time taken: 1.784 seconds
hive> select * from red_wine;
OK
NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL
7.4     0.7     0.0     1.9     0.076   11.0    34.0    0.9978  3.51    0.56    9.4     5.0
7.8     0.88    0.0     2.6     0.098   25.0    67.0    0.9968  3.2     0.68    9.8     5.0
7.8     0.76    0.04    2.3     0.092   15.0    54.0    0.997   3.26    0.65    9.8     5.0
11.2    0.28    0.56    1.9     0.075   17.0    60.0    0.998   3.16    0.58    9.8     6.0
7.4     0.7     0.0     1.9     0.076   11.0    34.0    0.9978  3.51    0.56    9.4     5.0
7.4     0.66    0.0     1.8     0.075   13.0    40.0    0.9978  3.51    0.56    9.4     5.0
7.9     0.6     0.06    1.6     0.069   15.0    59.0    0.9964  3.3     0.46    9.4     5.0
7.3     0.65    0.0     1.2     0.065   15.0    21.0    0.9946  3.39    0.47    10.0    7.0
7.8     0.58    0.02    2.0     0.073   9.0     18.0    0.9968  3.36    0.57    9.5     7.0
7.5     0.5     0.36    6.1     0.071   17.0    102.0   0.9978  3.35    0.8     10.5    5.0
6.7     0.58    0.08    1.8     0.097   15.0    65.0    0.9959  3.28    0.54    9.2     5.0
7.5     0.5     0.36    6.1     0.071   17.0    102.0   0.9978  3.35    0.8     10.5    5.0
5.6     0.615   0.0     1.6     0.089   16.0    59.0    0.9943  3.58    0.52    9.9     5.0
7.8     0.61    0.29    1.6     0.114   9.0     29.0    0.9974  3.26    1.56    9.1     5.0
8.9     0.62    0.18    3.8     0.176   52.0    145.0   0.9986  3.16    0.88    9.2     5.0
8.9     0.62    0.19    3.9     0.17    51.0    148.0   0.9986  3.17    0.93    9.2     5.0
8.5     0.28    0.56    1.8     0.092   35.0    103.0   0.9969  3.3     0.75    10.5    7.0
8.1     0.56    0.28    1.7     0.368   16.0    56.0    0.9968  3.11    1.28    9.3     5.0
7.4     0.59    0.08    4.4     0.086   6.0     29.0    0.9974  3.38    0.5     9.0     4.0
7.9     0.32    0.51    1.8     0.341   17.0    56.0    0.9969  3.04    1.08    9.2     6.0
8.9     0.22    0.48    1.8     0.077   29.0    60.0    0.9968  3.39    0.53    9.4     6.0
7.6     0.39    0.31    2.3     0.082   23.0    71.0    0.9982  3.52    0.65    9.7     5.0
7.9     0.43    0.21    1.6     0.106   10.0    37.0    0.9966  3.17    0.91    9.5     5.0
```

hive> LOAD DATA INPATH '/hive_poc/winequality-white.csv' INTO TABLE white_wine;

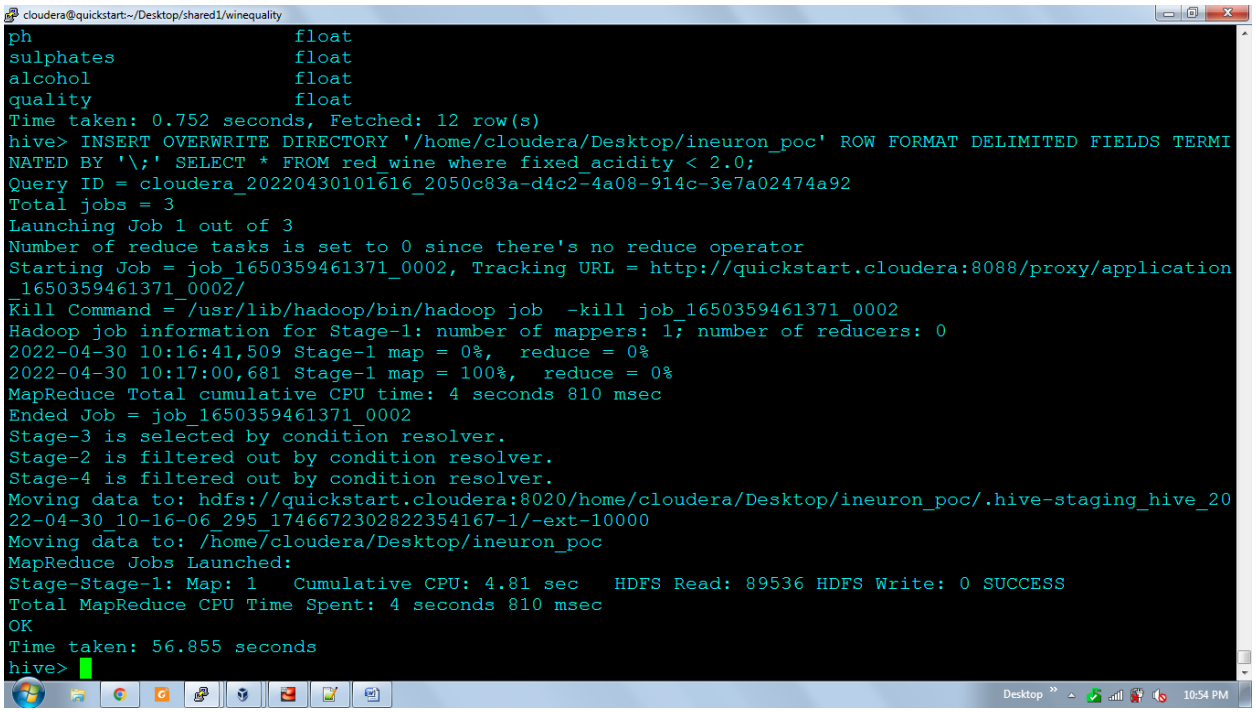


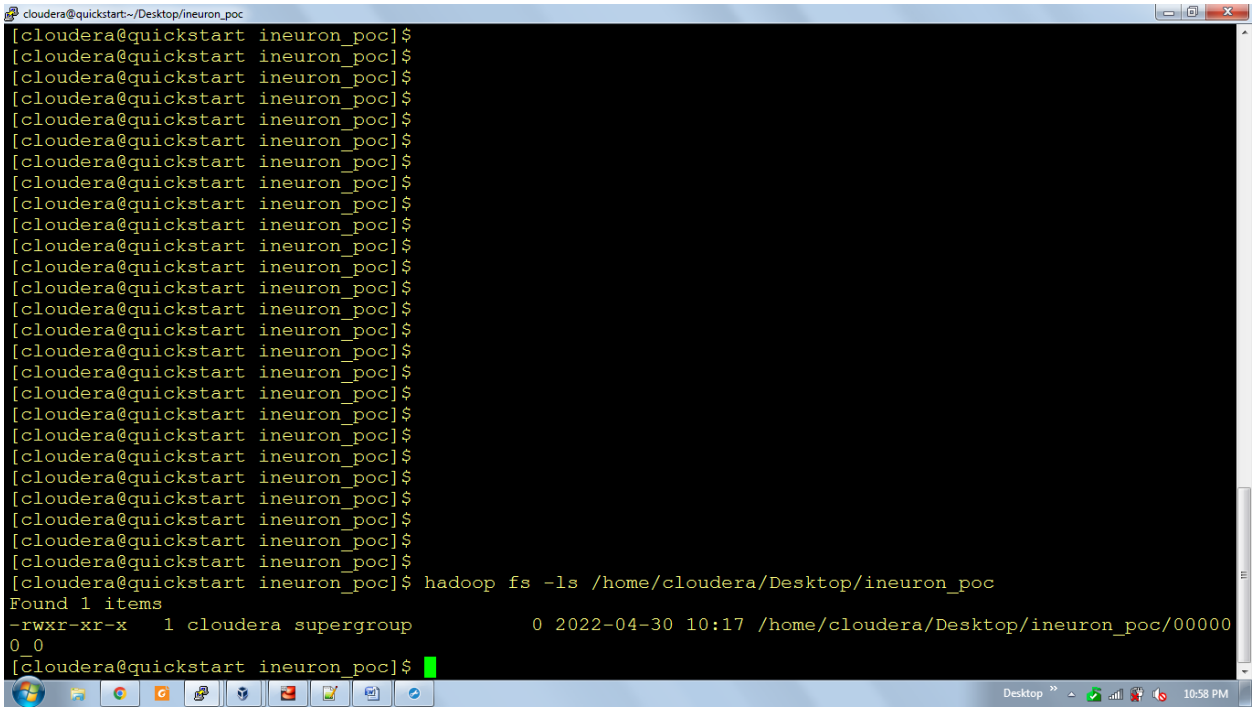
```
PuTTY (inactive)
Time taken: 0.202 seconds, Fetched: 1600 row(s)
hive> LOAD DATA INPATH '/hive_poc/winequality-white.csv' INTO TABLE white_wine;
Loading data to table default.white_wine
Table default.white_wine stats: [numFiles=1, totalSize=264426]
OK
Time taken: 0.849 seconds
hive> select * from white_wine
    > ;
OK
NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL
7.0     0.27    0.36    20.7    0.045   45.0    170.0   1.001   3.0     0.45    8.8     6.0
6.3     0.3     0.34    1.6     0.049   14.0    132.0   0.994   3.3     0.49    9.5     6.0
8.1     0.28    0.4     6.9     0.05    30.0    97.0    0.9951  3.26    0.44    10.1    6.0
7.2     0.23    0.32    8.5     0.058   47.0    186.0   0.9956  3.19    0.4     9.9     6.0
7.2     0.23    0.32    8.5     0.058   47.0    186.0   0.9956  3.19    0.4     9.9     6.0
8.1     0.28    0.4     6.9     0.05    30.0    97.0    0.9951  3.26    0.44    10.1    6.0
6.2     0.32    0.16    7.0     0.045   30.0    136.0   0.9949  3.18    0.47    9.6     6.0
7.0     0.27    0.36    20.7    0.045   45.0    170.0   1.001   3.0     0.45    8.8     6.0
6.3     0.3     0.34    1.6     0.049   14.0    132.0   0.994   3.3     0.49    9.5     6.0
8.1     0.22    0.43    1.5     0.044   28.0    129.0   0.9938  3.22    0.45    11.0    6.0
8.1     0.27    0.41    1.45    0.033   11.0    63.0    0.9908  2.99    0.56    12.0    5.0
8.6     0.23    0.4     4.2     0.035   17.0    109.0   0.9947  3.14    0.53    9.7     5.0
7.9     0.18    0.37    1.2     0.04    16.0    75.0    0.992   3.18    0.63    10.8    5.0
6.6     0.16    0.4     1.5     0.044   48.0    143.0   0.9912  3.54    0.52    12.4    7.0
8.3     0.42    0.62    19.25   0.04    41.0    172.0   1.0002  2.98    0.67    9.7     5.0
6.6     0.17    0.38    1.5     0.032   28.0    112.0   0.9914  3.25    0.55    11.4    7.0
6.3     0.48    0.04    1.1     0.046   30.0    99.0    0.9928  3.24    0.36    9.6     6.0
6.2     0.66    0.48    1.2     0.029   29.0    75.0    0.9892  3.33    0.39    12.8    8.0
7.4     0.34    0.42    1.1     0.033   17.0    171.0   0.9917  3.12    0.53    11.3    6.0
6.5     0.31    0.14    7.5     0.044   34.0    133.0   0.9955  3.22    0.5     9.5     5.0
6.2     0.66    0.48    1.2     0.029   29.0    75.0    0.9892  3.33    0.39    12.8    8.0
6.4     0.31    0.38    2.9     0.038   19.0    102.0   0.9912  3.17    0.35    11.0    7.0
6.8     0.26    0.42    1.7     0.049   41.0    122.0   0.993   3.47    0.48    10.5    8.0
7.6     0.67    0.14    1.5     0.074   25.0    168.0   0.9937  3.05    0.51    9.3     5.0
```

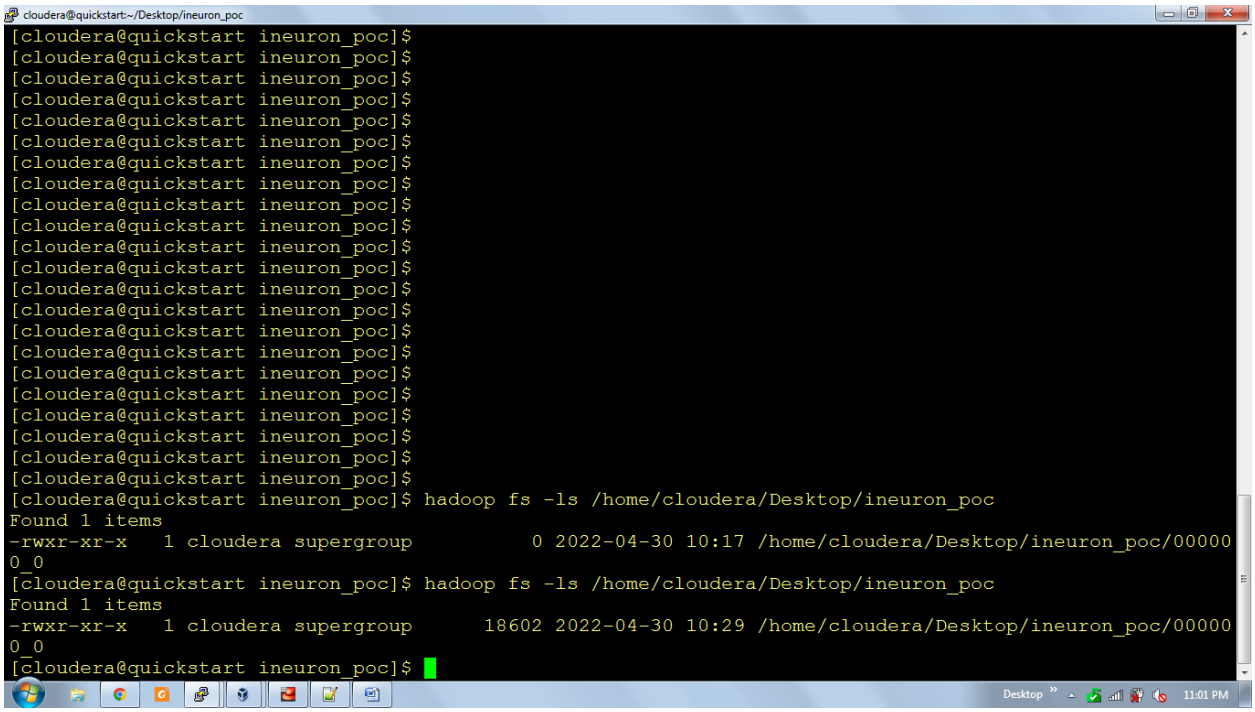
INSERT OVERWRITE DIRECTORY '/home/cloudera/Desktop/ineuron_poc' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\;' SELECT * FROM red_wine;

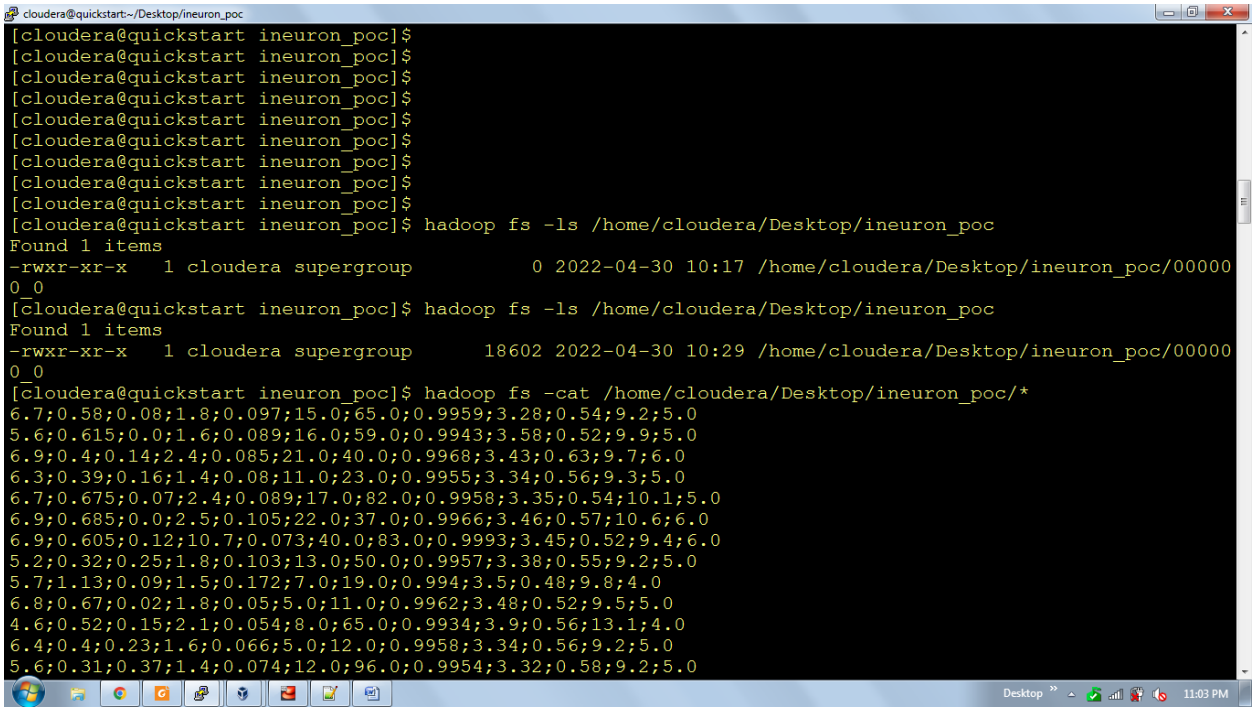


hive> INSERT OVERWRITE DIRECTORY '/home/cloudera/Desktop/ineuron_poc' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\;' SELECT * FROM red_wine where fixed_acidity < 2.0;

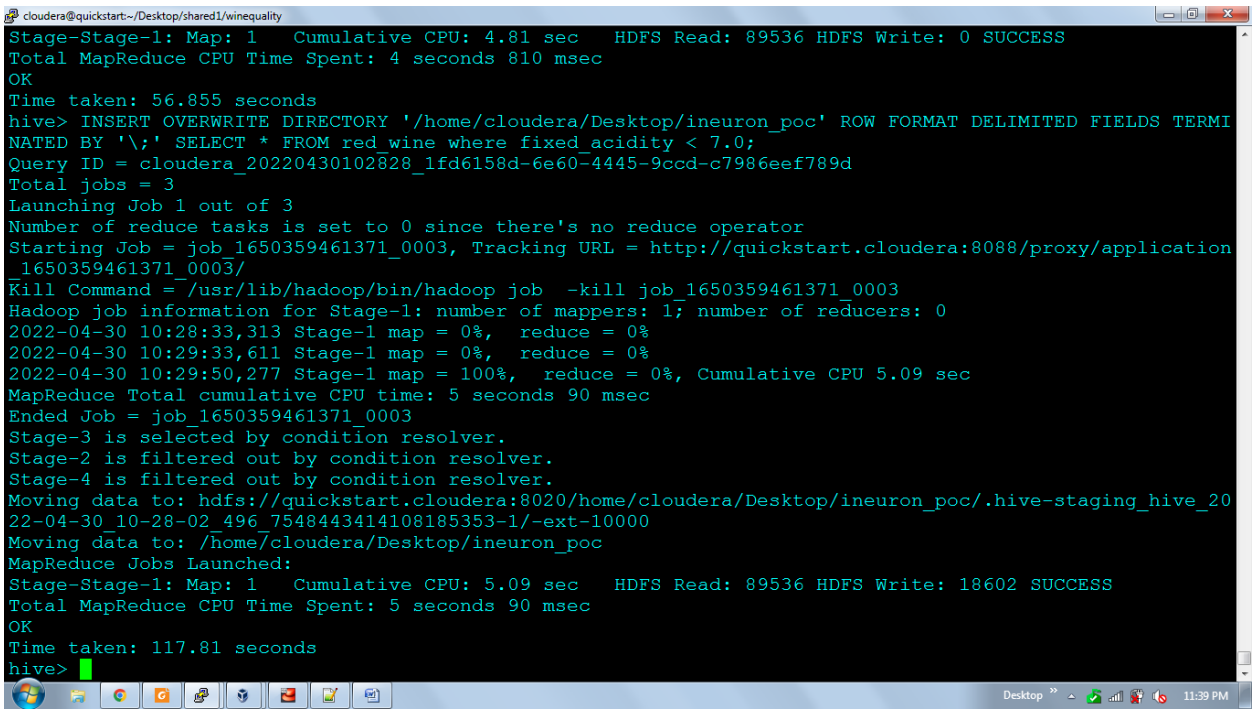



cloudera@quickstart:~/Desktop/ineuron_poc
[cloudera@quickstart ineuron_poc]$ hadoop fs -ls /home/cloudera/Desktop/ineuron_poc
Found 1 items
-rwxr-xr-x 1 cloudera supergroup 0 2022-04-30 10:17 /home/cloudera/Desktop/ineuron_poc/000000_0
[cloudera@quickstart ineuron_poc]$
Desktop
10:58 PM

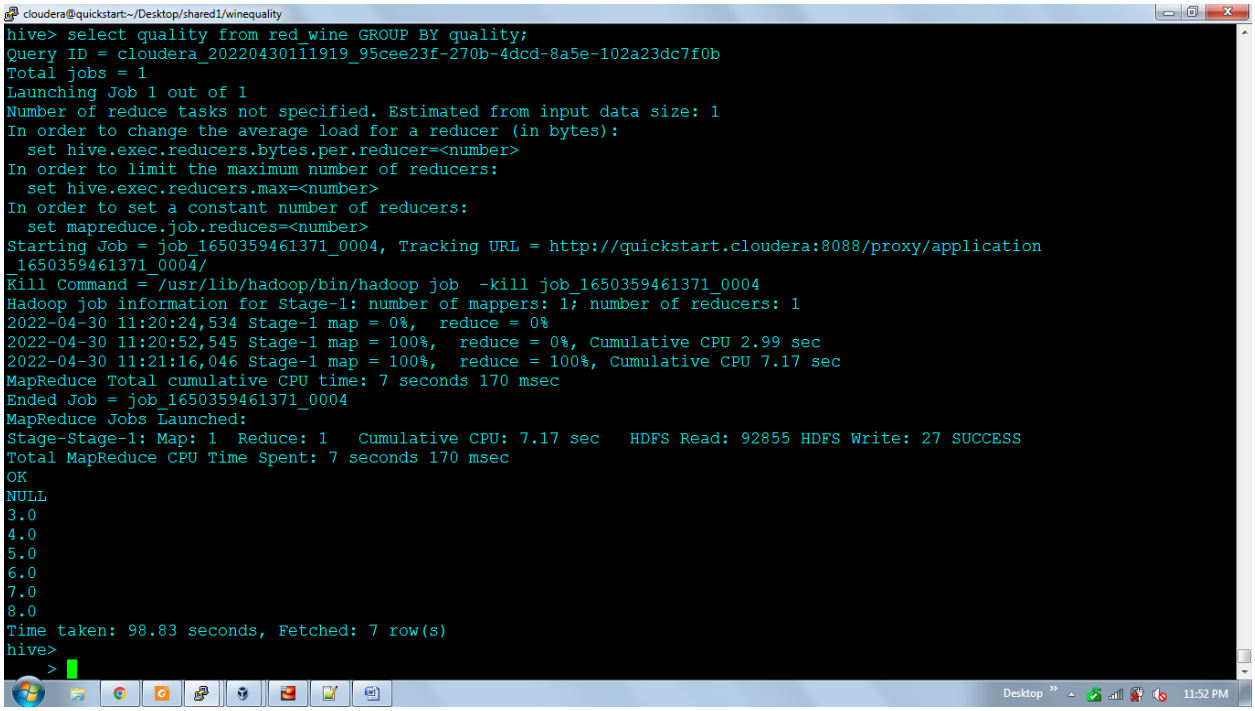

cloudera@quickstart:~/Desktop/ineuron_poc
[cloudera@quickstart ineuron_poc]$ hadoop fs -ls /home/cloudera/Desktop/ineuron_poc
Found 1 items
-rwxr-xr-x 1 cloudera supergroup 0 2022-04-30 10:17 /home/cloudera/Desktop/ineuron_poc/000000_0
[cloudera@quickstart ineuron_poc]$ hadoop fs -ls /home/cloudera/Desktop/ineuron_poc
Found 1 items
-rwxr-xr-x 1 cloudera supergroup 18602 2022-04-30 10:29 /home/cloudera/Desktop/ineuron_poc/000000_0
[cloudera@quickstart ineuron_poc]$
Desktop
11:01 PM



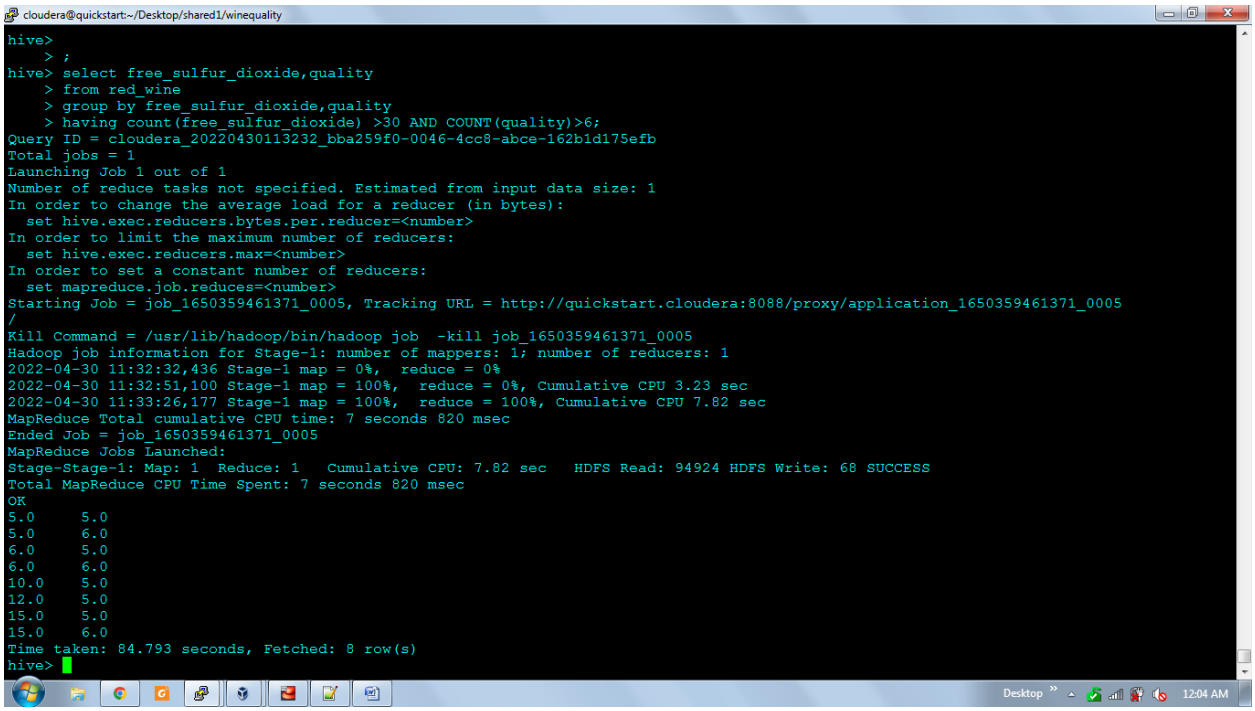
hive> INSERT OVERWRITE DIRECTORY '/home/cloudera/Desktop/ineuron_poc' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\;' SELECT * FROM red_wine where fixed_acidity < 7.0;



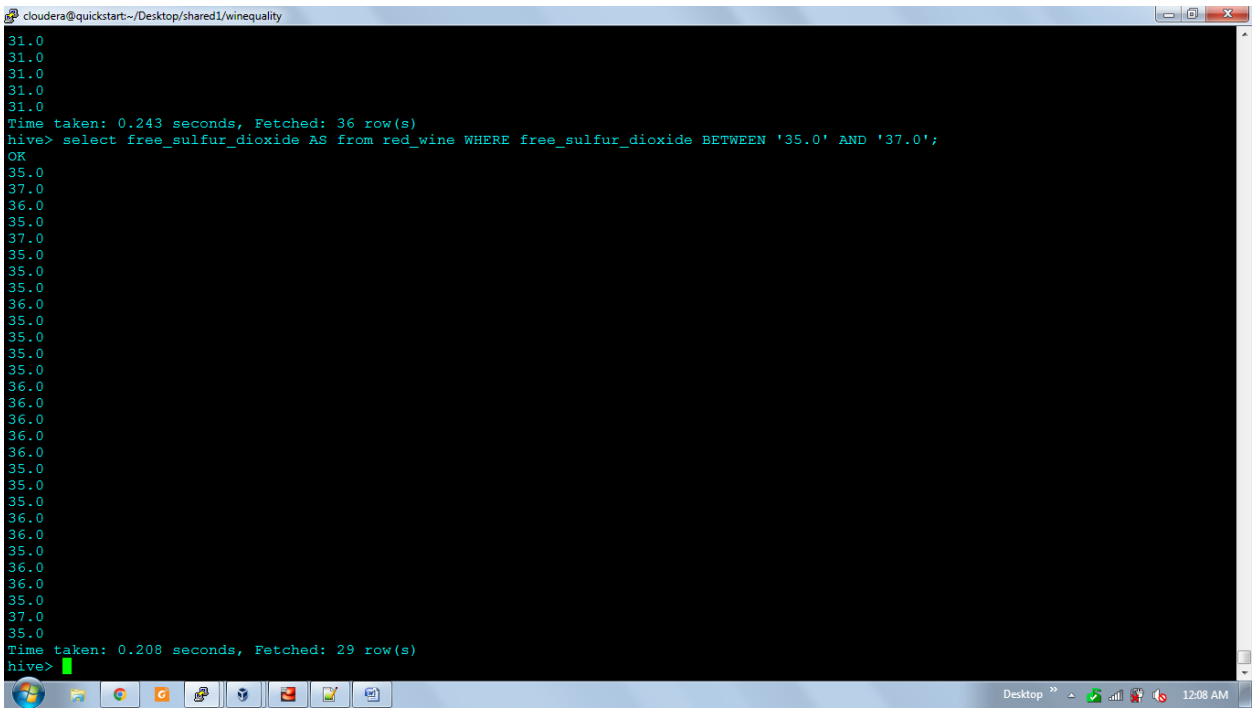
hive> select quality from red_wine GROUP BY quality;



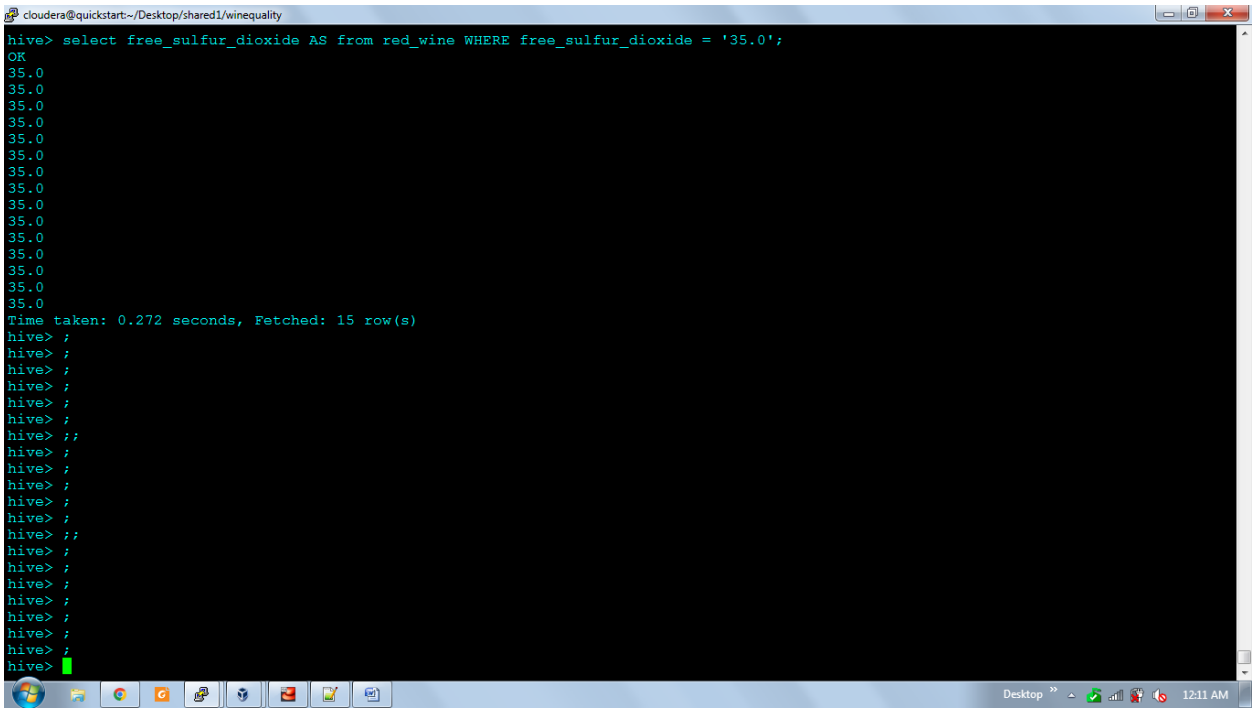
hive> select free_sulfur_dioxide,quality
from red_wine
group by free_sulfur_dioxide,quality
having count(free_sulfur_dioxide) >30 AND COUNT(quality)>6;



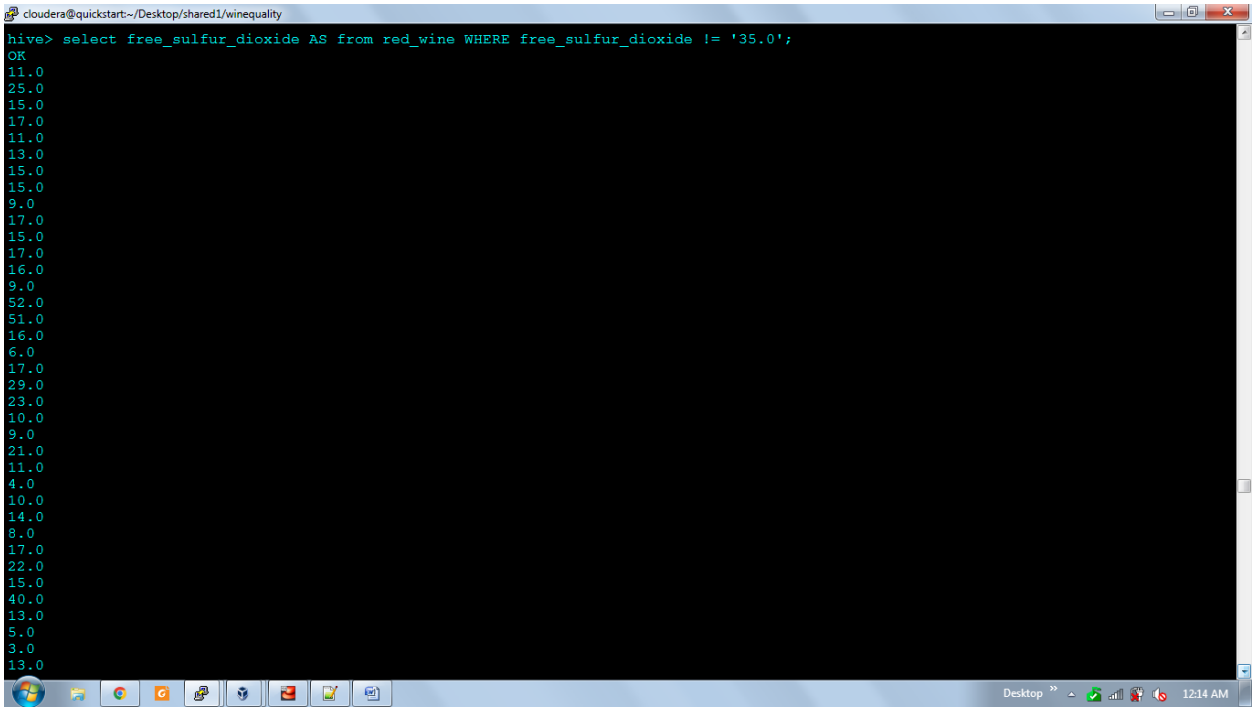
hive> select free_sulfur_dioxide AS from red_wine WHERE free_sulfur_dioxide BETWEEN '35.0' AND '37.0';



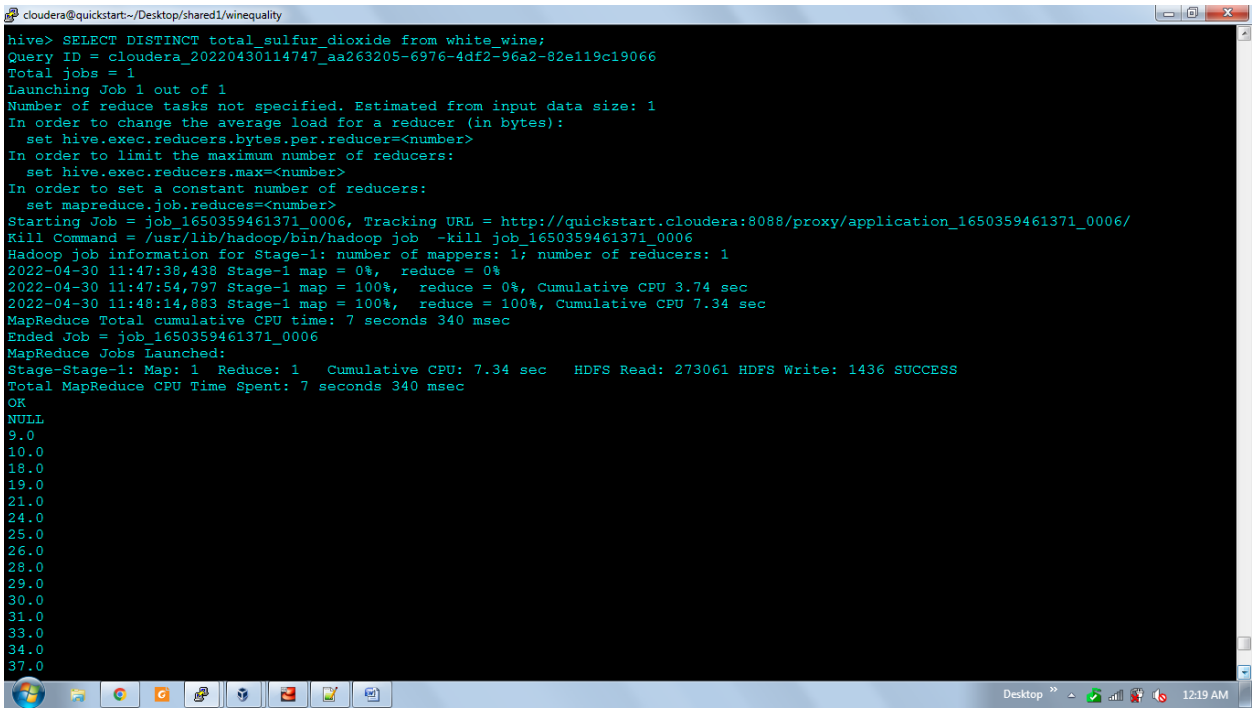
hive> select free_sulfur_dioxide AS from red_wine WHERE free_sulfur_dioxide = '35.0';



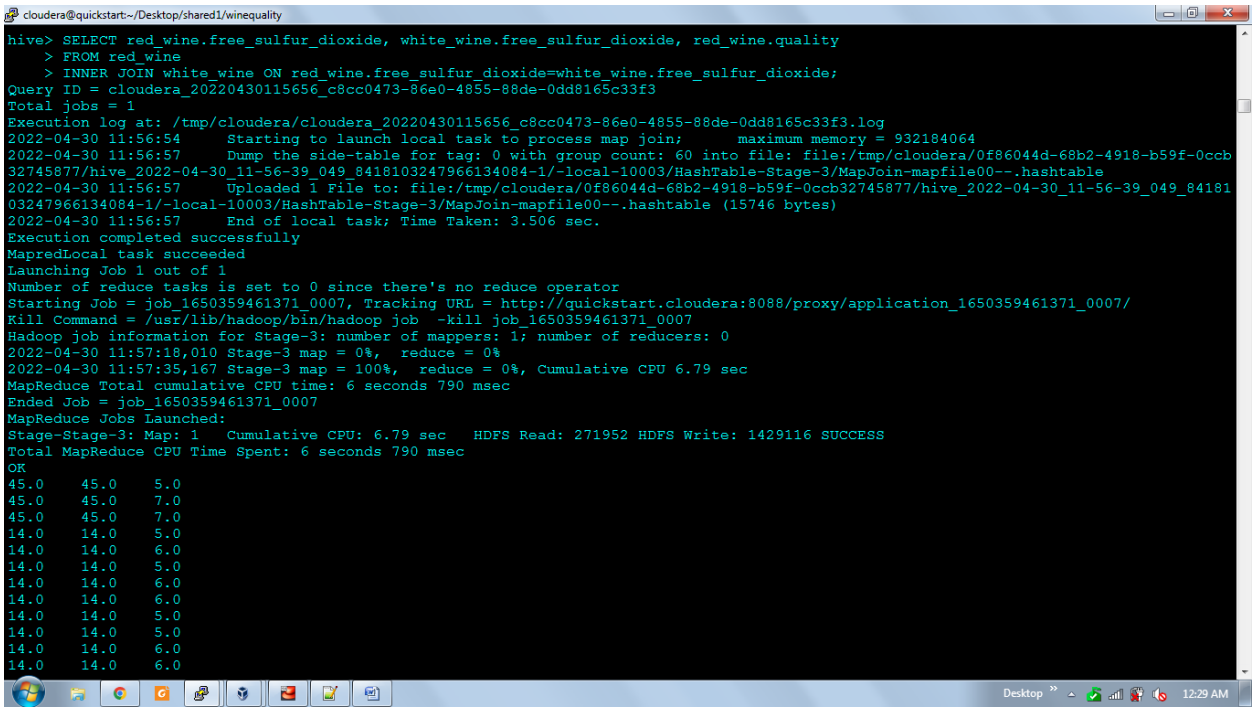
hive> select free_sulfur_dioxide AS from red_wine WHERE free_sulfur_dioxide != '35.0';



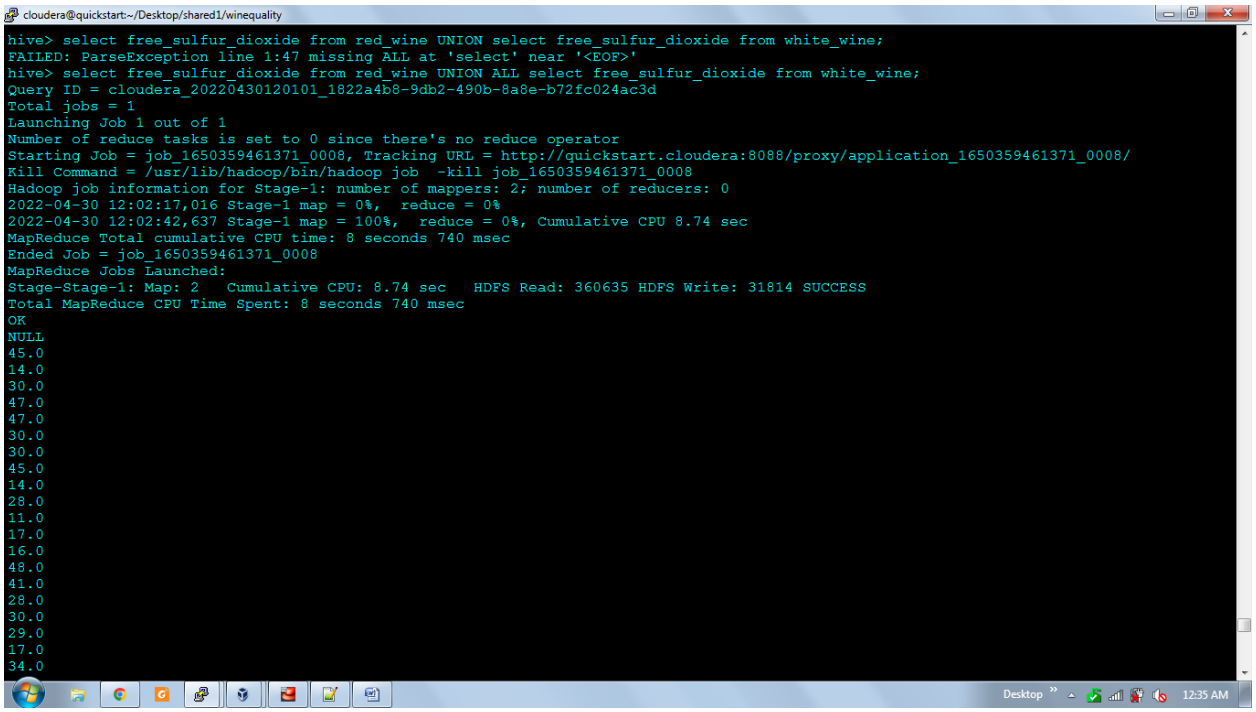
hive> SELECT DISTINCT total_sulfur_dioxide from white_wine;



```
hive> SELECT red_wine.free_sulfur_dioxide, white_wine.free_sulfur_dioxide, red_wine.quality
    FROM red_wine
    INNER JOIN white_wine ON red_wine.free_sulfur_dioxide=white_wine.free_sulfur_dioxide;
```

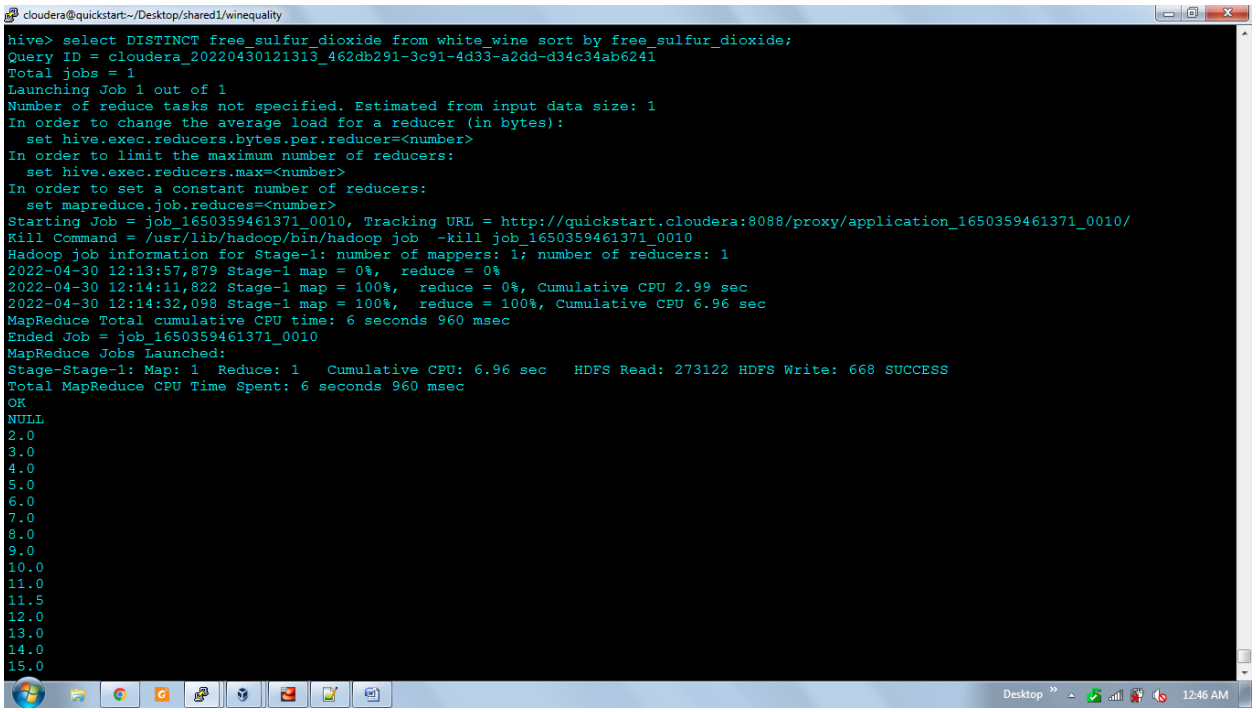



```
hive> select free_sulfur_dioxide from red_wine UNION ALL select free_sulfur_dioxide from white_wine;
```



hive> ALTER TABLE red_wine RENAME TO red_new_wine;

hive> select DISTINCT free_sulfur_dioxide from white_wine sort by free_sulfur_dioxide;



hive> drop table if exists white_wine;