

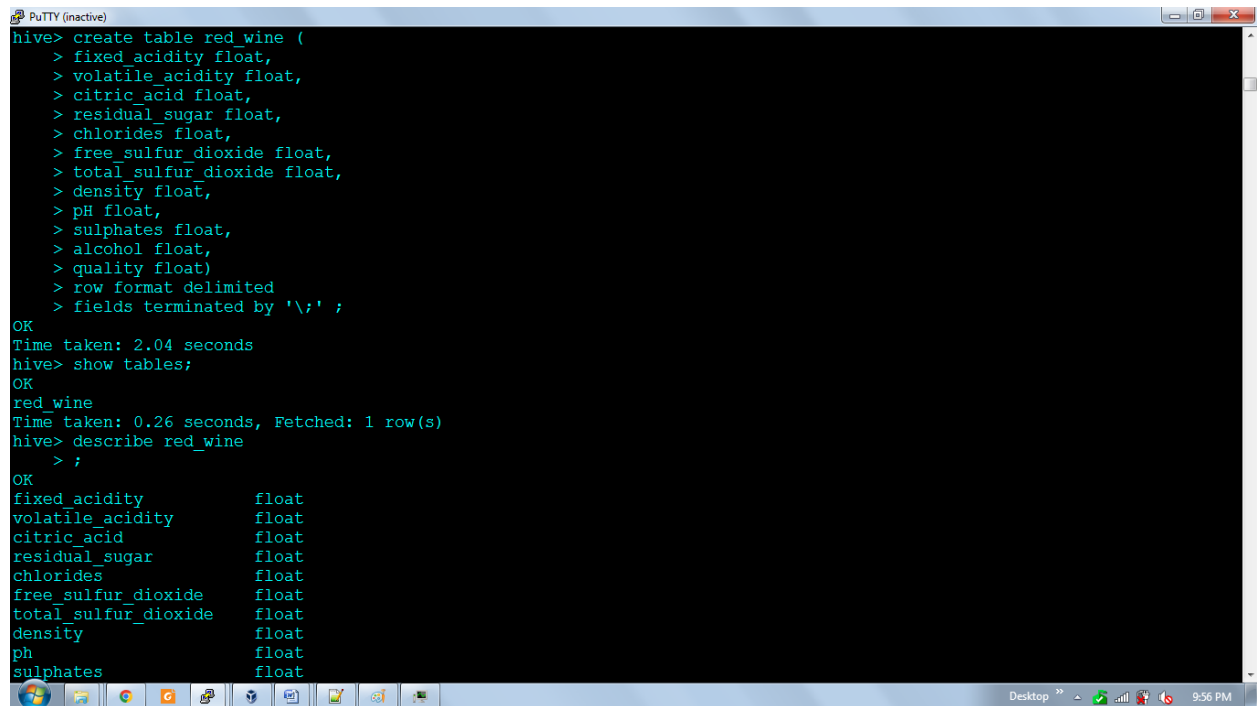
Big Data Assignment - 2

Copied two .csv files from Local to HDFS:

```
gpr@cloudera:~$ ssh cloudera@quickstart
[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 '\';
```

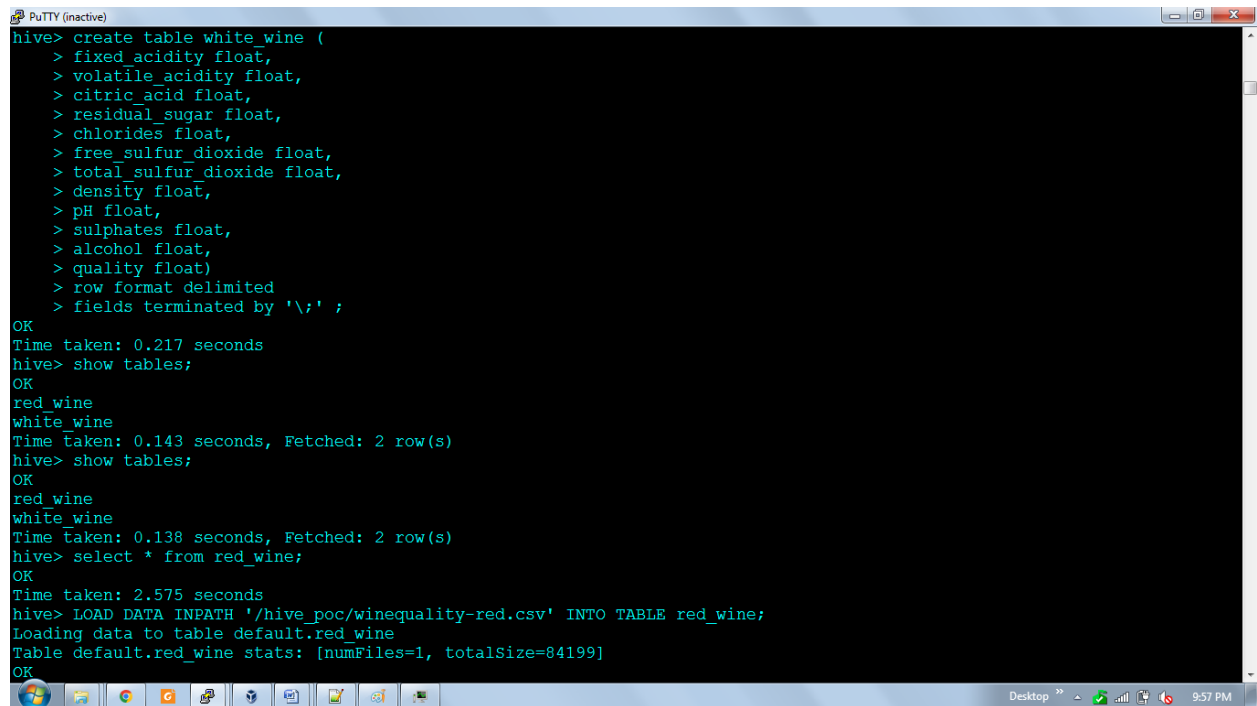


The screenshot shows a PuTTY terminal window titled "PuTTY (inactive)". The terminal displays the following Hive SQL commands and their outputs:

```
hive> 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 '\';  
OK  
Time taken: 2.04 seconds  
hive> show tables;  
OK  
red_wine  
Time taken: 0.26 seconds, Fetched: 1 row(s)  
hive> describe red_wine  
> ;  
OK  
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
```

The terminal window has a Windows taskbar at the bottom showing the time as 9:56 PM.

```
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 '\';
```



The screenshot shows a PuTTY terminal window titled "PuTTY (inactive)". The terminal displays the following Hive commands and their outputs:

```
hive> 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 '\';  
OK  
Time taken: 0.217 seconds  
hive> show tables;  
OK  
red_wine  
white_wine  
Time taken: 0.143 seconds, Fetched: 2 row(s)  
hive> show tables;  
OK  
red_wine  
white_wine  
Time taken: 0.138 seconds, Fetched: 2 row(s)  
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
```

The terminal window has a standard Windows taskbar at the bottom with various application icons and a system tray showing the time as 9:57 PM.

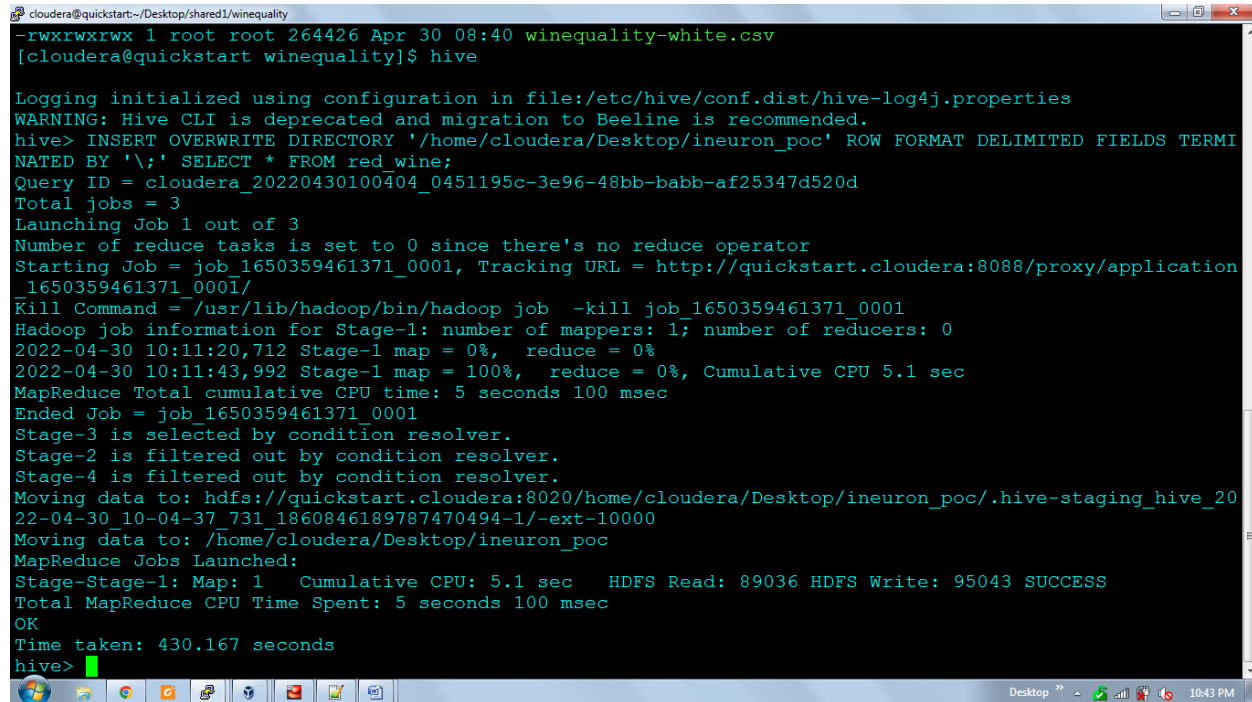
hive> LOAD DATA INPATH '/hive_poc/winequality-red.csv' INTO TABLE red_wine;

```
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
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;

```
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
```

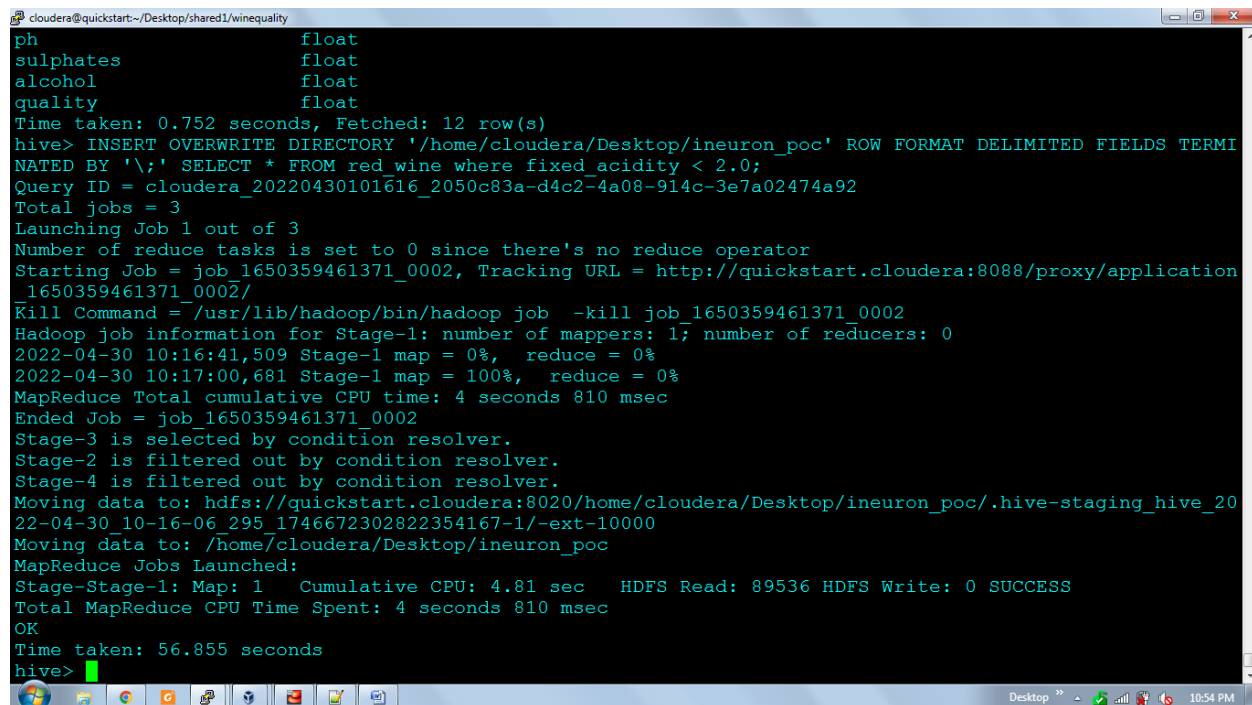
INSERT OVERWRITE DIRECTORY '/home/cloudera/Desktop/ineuron_poc' ROW FORMAT DELIMITED
FIELDS TERMINATED BY '\;' SELECT * FROM red_wine;



```
cloudera@quickstart:~/Desktop/shared1/winequality
-rwxrwxrwx 1 root root 264426 Apr 30 08:40 winequality-white.csv
[cloudera@quickstart winequality]$ 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> INSERT OVERWRITE DIRECTORY '/home/cloudera/Desktop/ineuron_poc' ROW FORMAT DELIMITED FIELDS TERMI
NATED BY '\;' SELECT * FROM red_wine;
Query ID = cloudera_20220430100404_0451195c-3e96-48bb-babb-af25347d520d
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1650359461371_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application
_1650359461371_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1650359461371_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-04-30 10:11:20,712 Stage-1 map = 0%, reduce = 0%
2022-04-30 10:11:43,992 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.1 sec
MapReduce Total cumulative CPU time: 5 seconds 100 msec
Ended Job = job_1650359461371_0001
Stage-3 is selected by condition resolver.
Stage-2 is filtered out by condition resolver.
Stage-4 is filtered out by condition resolver.
Moving data to: hdfs://quickstart.cloudera:8020/home/cloudera/Desktop/ineuron_poc/.hive-staging_hive_20
22-04-30_10-04-37_731_1860846189787470494-1/-ext-10000
Moving data to: /home/cloudera/Desktop/ineuron_poc
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 5.1 sec HDFS Read: 89036 HDFS Write: 95043 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 100 msec
OK
Time taken: 430.167 seconds
hive>
```

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/shared1/winequality
ph float
sulphates float
alcohol float
quality float
Time taken: 0.752 seconds, Fetched: 12 row(s)
hive> INSERT OVERWRITE DIRECTORY '/home/cloudera/Desktop/ineuron_poc' ROW FORMAT DELIMITED FIELDS TERMI
NATED BY '\;' SELECT * FROM red_wine where fixed_acidity < 2.0;
Query ID = cloudera_20220430101616_2050c83a-d4c2-4a08-914c-3e7a02474a92
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1650359461371_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application
_1650359461371_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1650359461371_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-04-30 10:16:41,509 Stage-1 map = 0%, reduce = 0%
2022-04-30 10:17:00,681 Stage-1 map = 100%, reduce = 0%
MapReduce Total cumulative CPU time: 4 seconds 810 msec
Ended Job = job_1650359461371_0002
Stage-3 is selected by condition resolver.
Stage-2 is filtered out by condition resolver.
Stage-4 is filtered out by condition resolver.
Moving data to: hdfs://quickstart.cloudera:8020/home/cloudera/Desktop/ineuron_poc/.hive-staging_hive_20
22-04-30_10-16-06_295_1746672302822354167-1/-ext-10000
Moving data to: /home/cloudera/Desktop/ineuron_poc
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 4.81 sec HDFS Read: 89536 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 810 msec
OK
Time taken: 56.855 seconds
hive>
```


hive> select quality from red_wine GROUP BY quality;

```
cloudera@quickstart:~/Desktop/shared1/winequality
hive> select quality from red_wine GROUP BY quality;
Query ID = cloudera_20220430111919_95cee23f-270b-4dcd-8a5e-102a23dc7f0b
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_1650359461371_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1650359461371_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1650359461371_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-04-30 11:20:24,534 Stage-1 map = 0%, reduce = 0%
2022-04-30 11:20:52,545 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.99 sec
2022-04-30 11:21:16,046 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.17 sec
MapReduce Total cumulative CPU time: 7 seconds 170 msec
Ended Job = job_1650359461371_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.17 sec HDFS Read: 92855 HDFS Write: 27 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 170 msec
OK
NULL
3.0
4.0
5.0
6.0
7.0
8.0
Time taken: 98.83 seconds, Fetched: 7 row(s)
hive>
```

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;

```
cloudera@quickstart:~/Desktop/shared1/winequality
hive>
>
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;
Query ID = cloudera_20220430113232_bba259f0-0046-4cc8-abce-162b1d175efb
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_1650359461371_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1650359461371_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1650359461371_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-04-30 11:32:32,436 Stage-1 map = 0%, reduce = 0%
2022-04-30 11:32:51,100 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.23 sec
2022-04-30 11:33:26,177 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.82 sec
MapReduce Total cumulative CPU time: 7 seconds 820 msec
Ended Job = job_1650359461371_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.82 sec HDFS Read: 94924 HDFS Write: 68 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 820 msec
OK
5.0 5.0
5.0 6.0
6.0 5.0
6.0 6.0
10.0 5.0
12.0 5.0
15.0 5.0
15.0 6.0
Time taken: 84.793 seconds, Fetched: 8 row(s)
hive>
```



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

```
coudera@quickstart~/Desktop/shared1/winequality
31.0
31.0
31.0
31.0
31.0
Time taken: 0.243 seconds, Fetched: 36 row(s)
hive> select free_sulfur_dioxide AS from red_wine WHERE free_sulfur_dioxide BETWEEN '35.0' AND '37.0';
OK
35.0
37.0
36.0
35.0
37.0
35.0
35.0
35.0
36.0
35.0
35.0
36.0
35.0
35.0
36.0
36.0
36.0
36.0
36.0
35.0
35.0
35.0
36.0
36.0
36.0
36.0
35.0
37.0
35.0
Time taken: 0.208 seconds, Fetched: 29 row(s)
hive>
```

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

```
coudera@quickstart-~/Desktop/shared1/winequality
```

```
hive> select free_sulfur_dioxide AS from red_wine WHERE free_sulfur_dioxide = '35.0';  
OK  
35.0  
35.0  
35.0  
35.0  
35.0  
35.0  
35.0  
35.0  
35.0  
35.0  
Time taken: 0.272 seconds, Fetched: 15 row(s)  
hive>  
hive>  
hive>  
hive>  
hive>  
hive>  
hive>;  
hive>  
hive>  
hive>  
hive>  
hive>;  
hive>  
hive>  
hive>  
hive>  
hive>  
hive>;  
hive>  
hive>  
hive>  
hive>  
hive>  
hive>
```

The screenshot shows a terminal window titled "coudera@quickstart-~/Desktop/shared1/winequality". The user has executed a Hive SQL query: `select free_sulfur_dioxide AS from red_wine WHERE free_sulfur_dioxide = '35.0';`. The output consists of 15 rows, each containing the value "35.0". Below the results, it states "Time taken: 0.272 seconds, Fetched: 15 row(s)". Following this, there are several blank prompts (`hive>`) and semicolon-separated commands (`hive>;`), indicating further interaction or execution attempts.

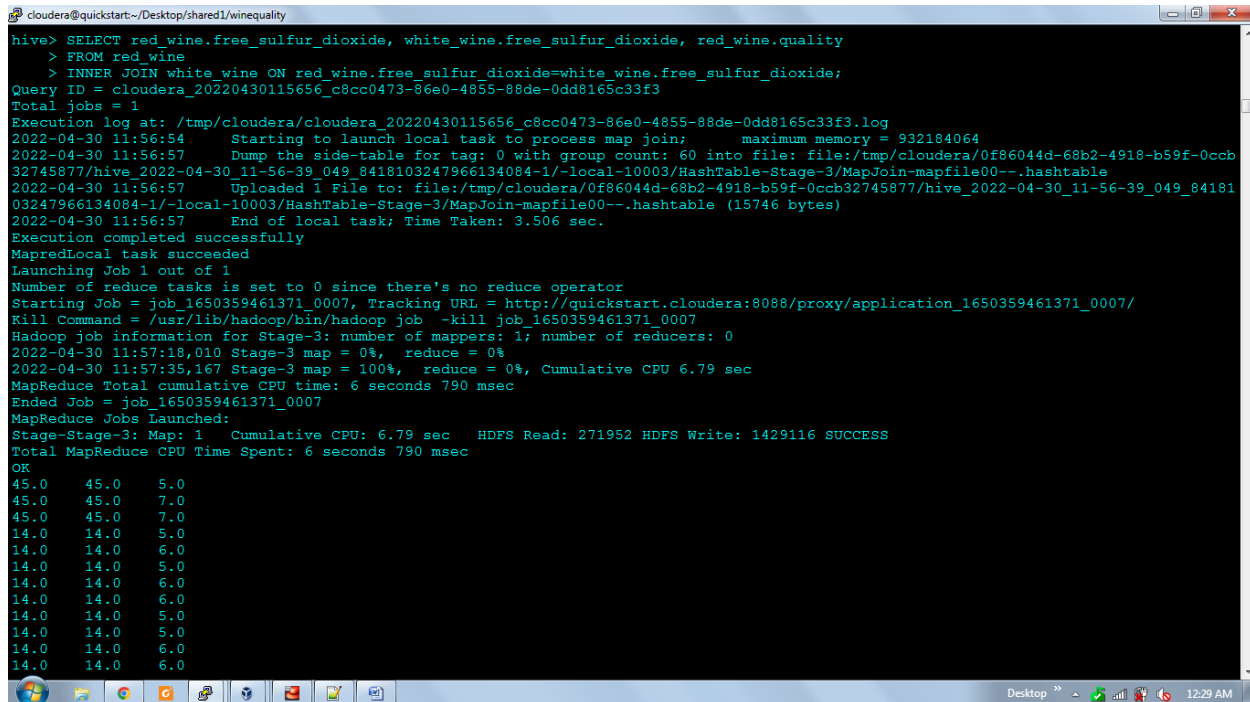
hive> select free_sulfur_dioxide AS from red_wine WHERE free_sulfur_dioxide != '35.0';

```
cloudera@quickstart:~/Desktop/shared1/winequality
hive> select free_sulfur_dioxide AS from red_wine WHERE free_sulfur_dioxide != '35.0';
OK
11.0
25.0
15.0
17.0
11.0
13.0
15.0
15.0
9.0
17.0
15.0
17.0
16.0
9.0
52.0
51.0
16.0
6.0
17.0
29.0
23.0
10.0
9.0
21.0
11.0
4.0
10.0
14.0
8.0
17.0
22.0
15.0
40.0
13.0
5.0
3.0
13.0
```

hive> SELECT DISTINCT total_sulfur_dioxide from white_wine;

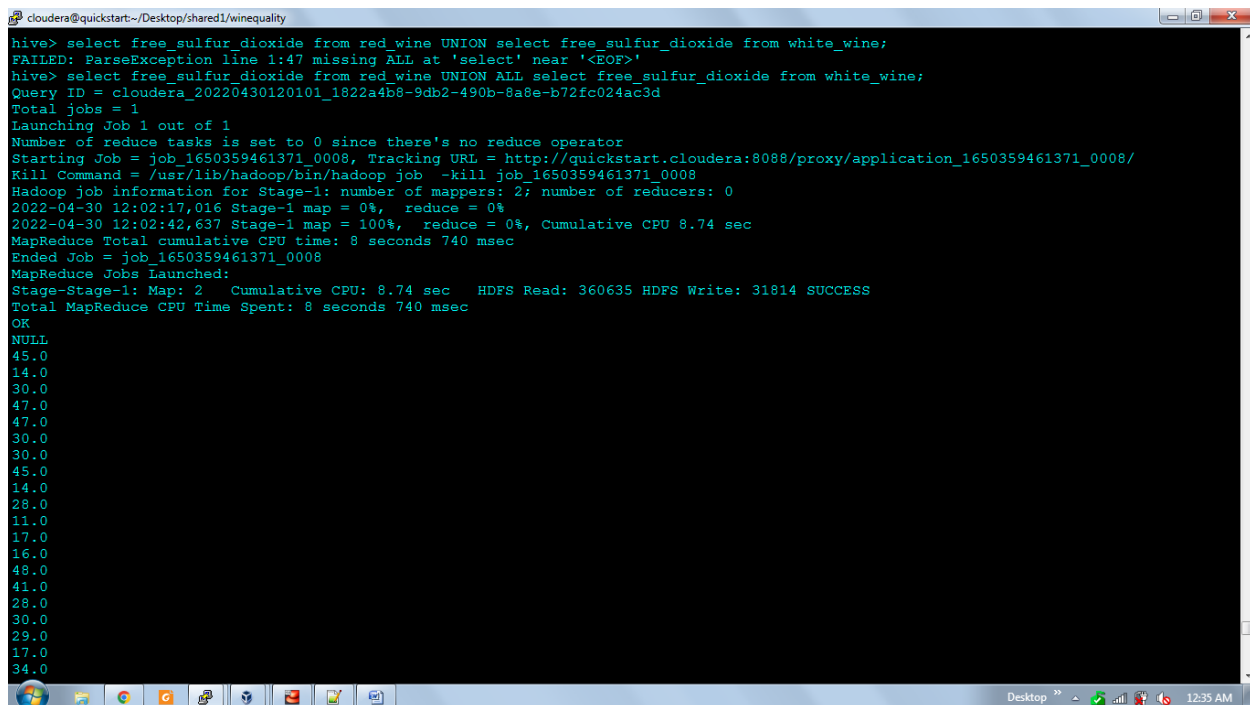
```
cloudera@quickstart:~/Desktop/shared1/winequality
hive> SELECT DISTINCT total_sulfur_dioxide from white_wine;
Query ID = cloudera_20220430114747_aa263205-6976-4df2-96a2-82e119c19066
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In 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_1650359461371_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1650359461371_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1650359461371_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-04-30 11:47:38,438 Stage-1 map = 0%, reduce = 0%
2022-04-30 11:47:54,797 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.74 sec
2022-04-30 11:48:14,883 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.34 sec
MapReduce Total cumulative CPU time: 7 seconds 340 msec
Ended Job = job_1650359461371_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.34 sec HDFS Read: 273061 HDFS Write: 1436 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 340 msec
OK
NULL
9.0
10.0
18.0
19.0
21.0
24.0
25.0
26.0
28.0
29.0
30.0
31.0
33.0
34.0
37.0
```

```
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;
```



```
cloudera@quickstart:~/Desktop/shared1/winequality
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;
Query ID = cloudera_20220430115656_c8cc0473-86e0-4855-88de-0dd8165c33f3
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20220430115656_c8cc0473-86e0-4855-88de-0dd8165c33f3.log
2022-04-30 11:56:54 Starting to launch local task to process map join; maximum memory = 932184064
2022-04-30 11:56:57 Dump the side-table for tag: 0 with group count: 60 into file: file:/tmp/cloudera/0f86044d-68b2-4918-b59f-0ccb32745877/hive_2022-04-30_11-56-39_049_8418103247966134084-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile00--.hashtable
2022-04-30 11:56:57 Uploaded 1 File to: file:/tmp/cloudera/0f86044d-68b2-4918-b59f-0ccb32745877/hive_2022-04-30_11-56-39_049_8418103247966134084-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile00--.hashtable (15746 bytes)
2022-04-30 11:56:57 End of local task; Time Taken: 3.506 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1650359461371_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1650359461371_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1650359461371_0007
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-04-30 11:57:18,010 Stage-3 map = 0%, reduce = 0%
2022-04-30 11:57:35,167 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 6.79 sec
MapReduce Total cumulative CPU time: 6 seconds 790 msec
Ended Job = job_1650359461371_0007
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 6.79 sec HDFS Read: 271952 HDFS Write: 1429116 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 790 msec
OK
45.0 45.0 5.0
45.0 45.0 7.0
45.0 45.0 7.0
14.0 14.0 5.0
14.0 14.0 6.0
14.0 14.0 5.0
14.0 14.0 6.0
14.0 14.0 6.0
14.0 14.0 5.0
14.0 14.0 6.0
14.0 14.0 5.0
14.0 14.0 6.0
14.0 14.0 6.0
14.0 14.0 6.0
```

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



```
cloudera@quickstart:~/Desktop/shared1/winequality
hive> select free_sulfur_dioxide from red_wine UNION select free_sulfur_dioxide from white_wine;
FAILED: ParseException line 1:47 missing ALL at 'select' near '<EOF>'
hive> select free_sulfur_dioxide from red_wine UNION ALL select free_sulfur_dioxide from white_wine;
Query ID = cloudera_20220430120101_1822a4b8-9db2-490b-8a8e-b72fc024ac3d
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1650359461371_0008, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1650359461371_0008/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1650359461371_0008
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 0
2022-04-30 12:02:17,016 Stage-1 map = 0%, reduce = 0%
2022-04-30 12:02:42,637 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.74 sec
MapReduce Total cumulative CPU time: 8 seconds 740 msec
Ended Job = job_1650359461371_0008
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Cumulative CPU: 8.74 sec HDFS Read: 360635 HDFS Write: 31814 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 740 msec
OK
NULL
45.0
14.0
30.0
47.0
47.0
30.0
30.0
45.0
14.0
28.0
11.0
17.0
16.0
48.0
41.0
28.0
30.0
29.0
17.0
34.0
```

hive> ALTER TABLE red_wine RENAME TO red_new_wine;

```
> ;
hive> show tables;
OK
red_wine
white_wine
Time taken: 0.137 seconds, Fetched: 2 row(s)
hive> ALTER TABLE red_wine RENAME TO red_new_wine;
OK
Time taken: 3.609 seconds
hive> show tables;
OK
red_new_wine
white_wine
Time taken: 0.083 seconds, Fetched: 2 row(s)
hive>
```

hive> select DISTINCT free_sulfur_dioxide from white_wine sort by free_sulfur_dioxide;

```
cloudera@quickstart:~/Desktop/shared1/winequality
hive> select DISTINCT free_sulfur_dioxide from white_wine sort by free_sulfur_dioxide;
Query ID = cloudera_20220430121313_462db291-3c91-4d33-a2dd-d34c34ab6241
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_1650359461371_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1650359461371_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1650359461371_0010
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-04-30 12:13:57,879 Stage-1 map = 0%, reduce = 0%
2022-04-30 12:14:11,822 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.99 sec
2022-04-30 12:14:32,098 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.96 sec
MapReduce Total cumulative CPU time: 6 seconds 960 msec
Ended Job = job_1650359461371_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.96 sec HDFS Read: 273122 HDFS Write: 668 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 960 msec
OK
NULL
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
11.5
12.0
13.0
14.0
15.0
```

hive> drop table if exists white_wine;

```
hive> drop table if exists white_wine;
OK
Time taken: 7.538 seconds
hive> show tables;
OK
red_new_wine
Time taken: 0.156 seconds, Fetched: 1 row(s)
hive> drop table red_new_wine;
OK
Time taken: 0.506 seconds
hive> show tables;
OK
Time taken: 0.095 seconds
hive>
```