



Group Project of

DATA MINING (WQD 7005)

Milestone 2 - Management of data

Title

Store Data into HIVE Data Warehouse

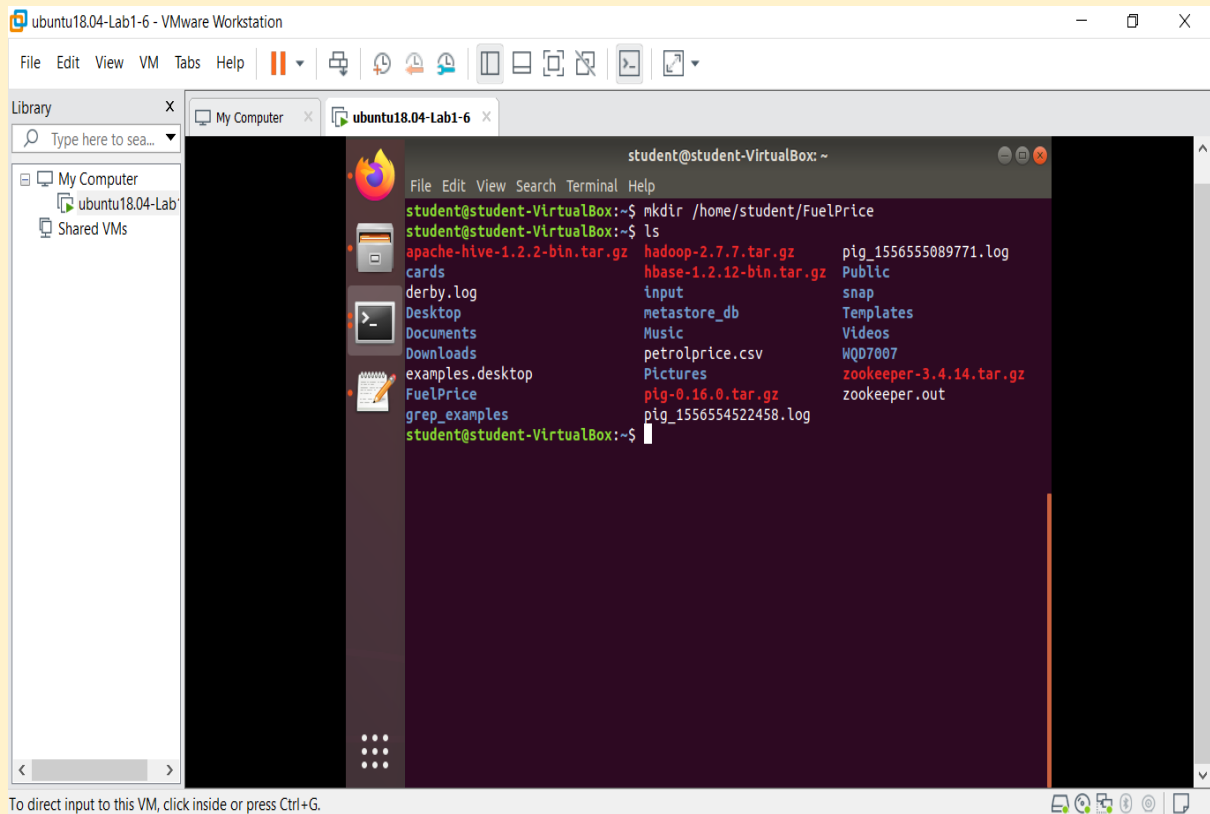
Instructor

Professor Dr. Teh Ying Wah

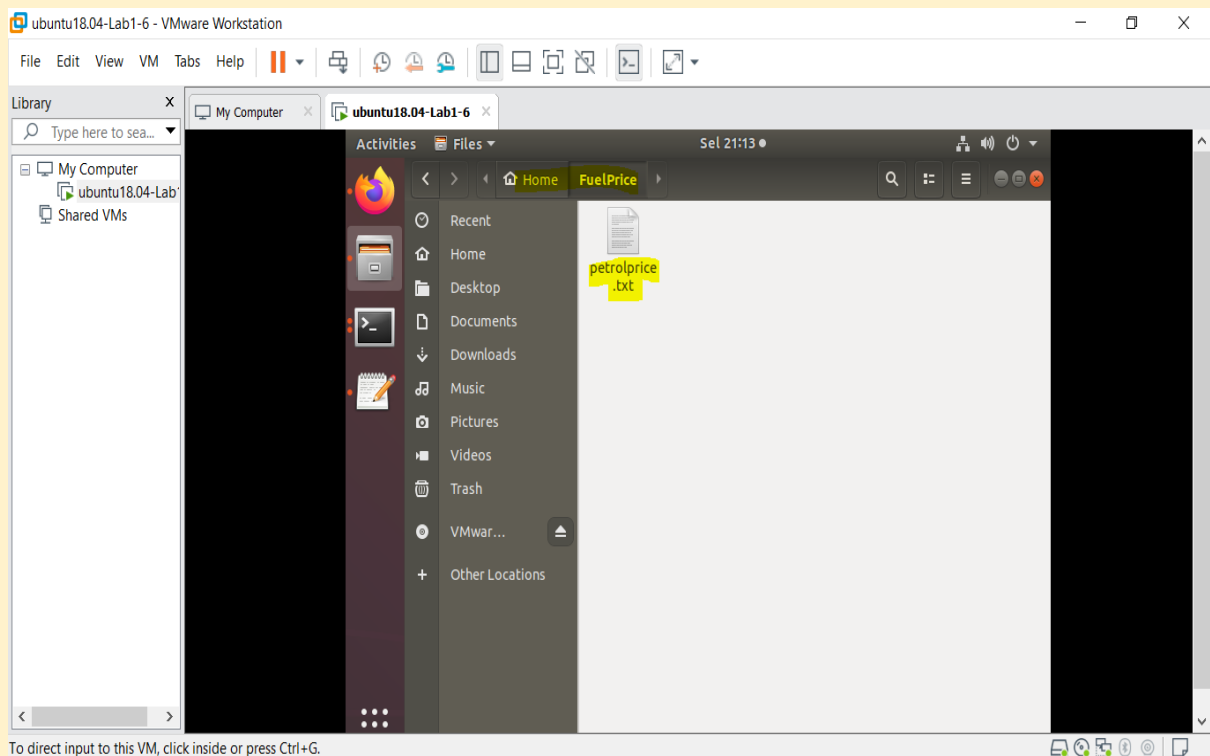
Group members

Sivanesan Pillai	WQD170074
Mathavan Chandrasegaram	WQD170075

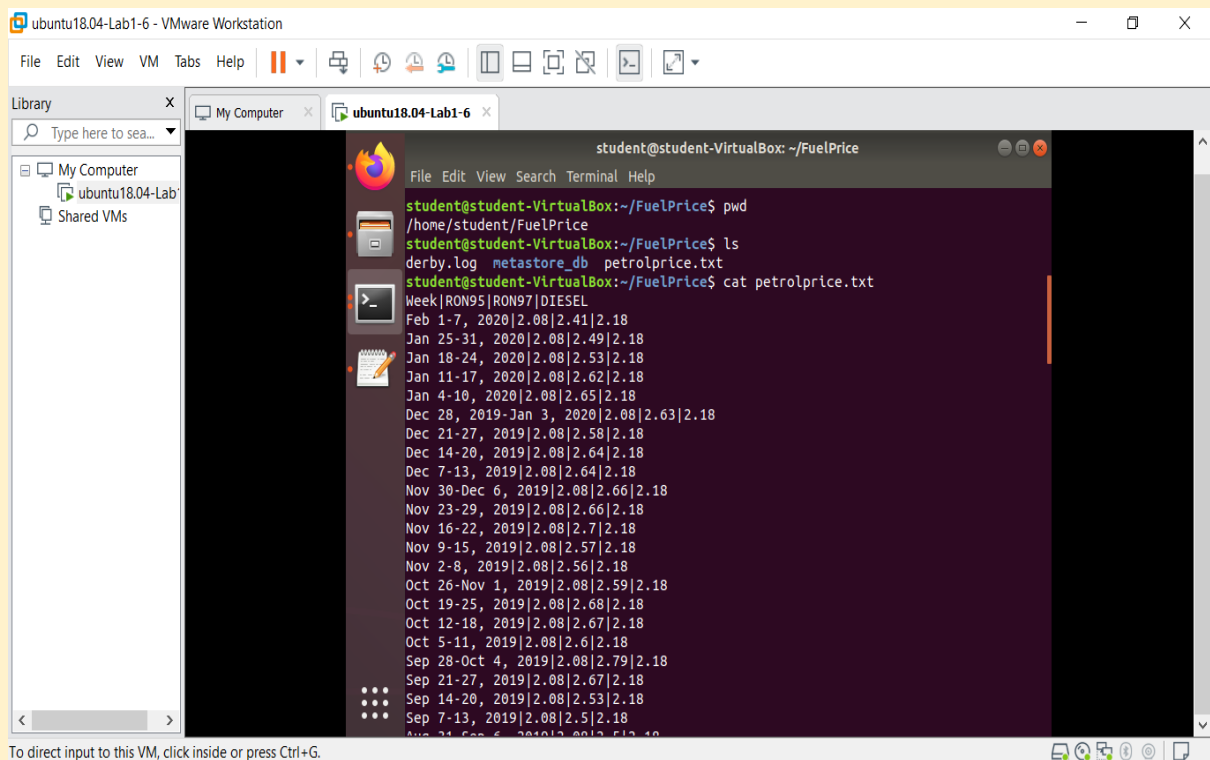
1. Created a directory into local system.



2. Save a data/file as in txt format into newly created directory.

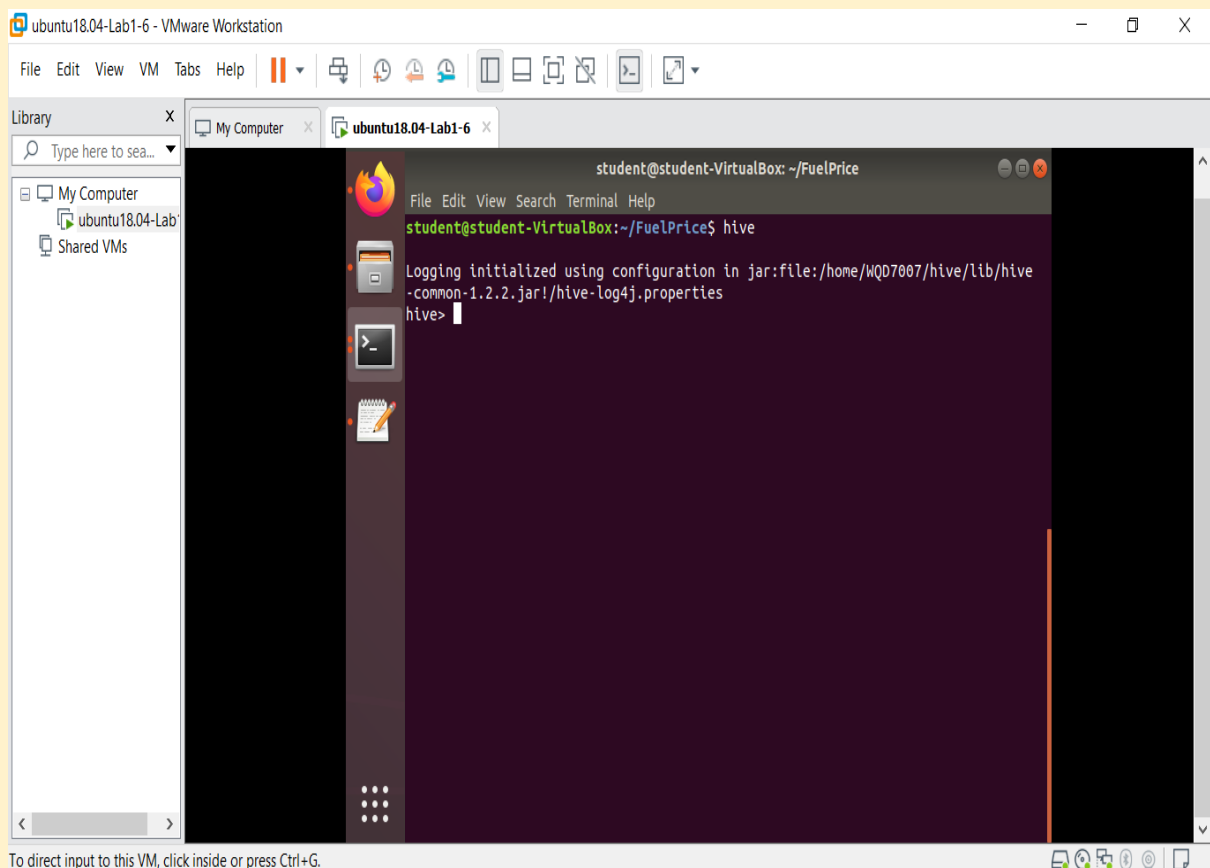


3. Locate and view the data file.



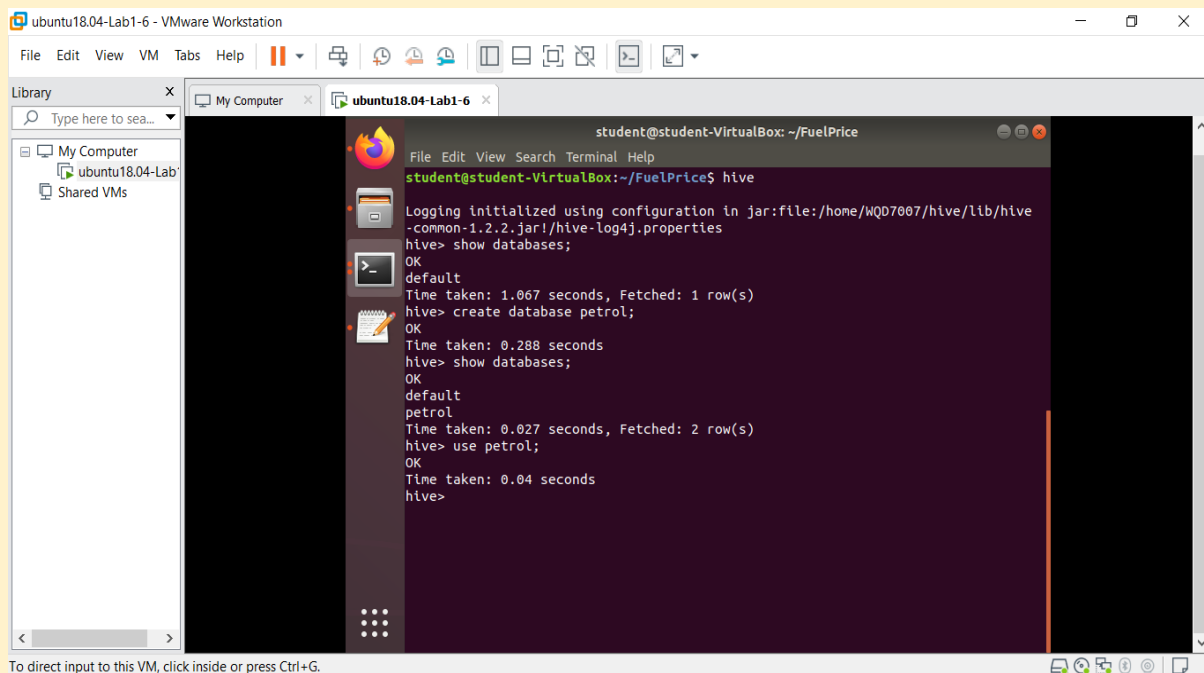
```
student@student-VirtualBox: ~/FuelPrice
File Edit View Search Terminal Help
student@student-VirtualBox:~/FuelPrice$ pwd
/home/student/FuelPrice
student@student-VirtualBox:~/FuelPrice$ ls
derby.log  metastore_db  petrolprice.txt
student@student-VirtualBox:~/FuelPrice$ cat petrolprice.txt
Week|RON95|RON97|DIESEL
Feb 1-7, 2020|2.08|2.41|2.18
Jan 25-31, 2020|2.08|2.49|2.18
Jan 18-24, 2020|2.08|2.53|2.18
Jan 11-17, 2020|2.08|2.62|2.18
Jan 4-10, 2020|2.08|2.65|2.18
Dec 28, 2019-Jan 3, 2020|2.08|2.63|2.18
Dec 21-27, 2019|2.08|2.58|2.18
Dec 14-20, 2019|2.08|2.64|2.18
Dec 7-13, 2019|2.08|2.64|2.18
Nov 30-Dec 6, 2019|2.08|2.66|2.18
Nov 23-29, 2019|2.08|2.66|2.18
Nov 16-22, 2019|2.08|2.7|2.18
Nov 9-15, 2019|2.08|2.57|2.18
Nov 2-8, 2019|2.08|2.56|2.18
Oct 26-Nov 1, 2019|2.08|2.59|2.18
Oct 19-25, 2019|2.08|2.68|2.18
Oct 12-18, 2019|2.08|2.67|2.18
Oct 5-11, 2019|2.08|2.6|2.18
Sep 28-Oct 4, 2019|2.08|2.79|2.18
Sep 21-27, 2019|2.08|2.67|2.18
Sep 14-20, 2019|2.08|2.53|2.18
Sep 7-13, 2019|2.08|2.5|2.18
Aug 31-Sep 6, 2019|2.08|2.5|2.18
```

4. Launch “HIVE” – data warehouse



```
student@student-VirtualBox: ~/FuelPrice
File Edit View Search Terminal Help
student@student-VirtualBox:~/FuelPrice$ hive
Logging initialized using configuration in jar:file:/home/MQD7007/hive/lib/hive
-common-1.2.2.jar!/hive-log4j.properties
hive>
```

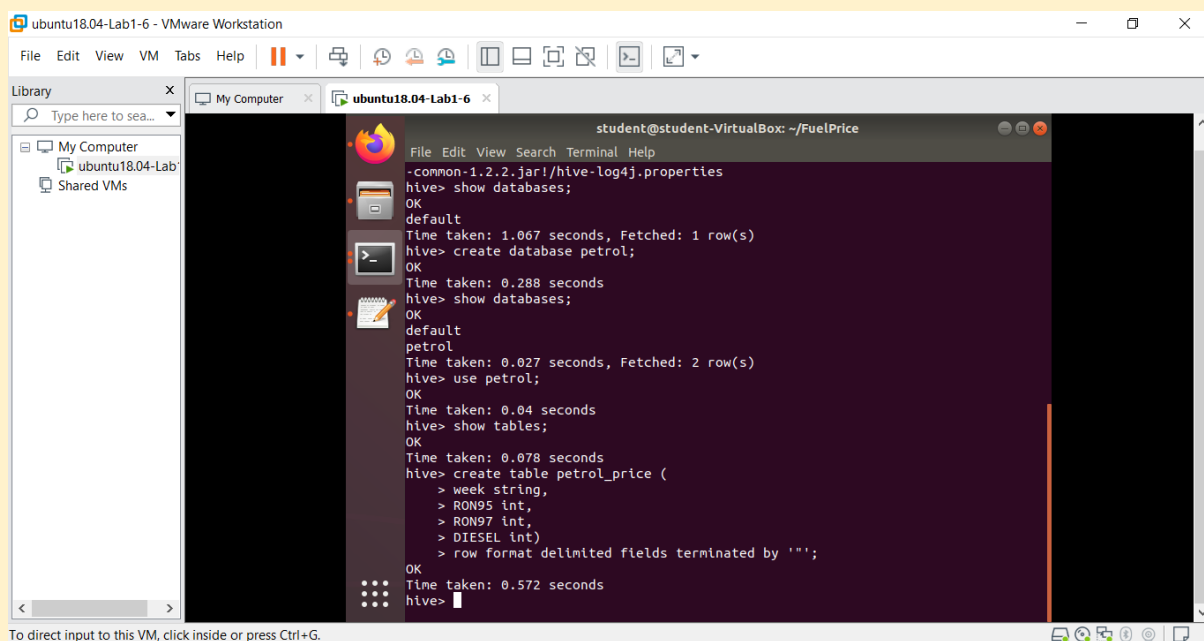
5. Check databases, create new database as petrol.



```
student@student-VirtualBox: ~/FuelPrice
student@student-VirtualBox:~/FuelPrice$ hive
Logging initialized using configuration in jar:file:/home/WQD7007/hive/lib/hive
-common-1.2.2.jar!/hive-log4j.properties
hive> show databases;
OK
default
Time taken: 1.067 seconds, Fetched: 1 row(s)
hive> create database petrol;
OK
Time taken: 0.288 seconds
hive> show databases;
OK
default
petrol
Time taken: 0.027 seconds, Fetched: 2 row(s)
hive> use petrol;
OK
Time taken: 0.04 seconds
hive>
```

6. Use the newly created database (petrol), check for existing table, and create new table as petrol_price_new with defining all the required columns.

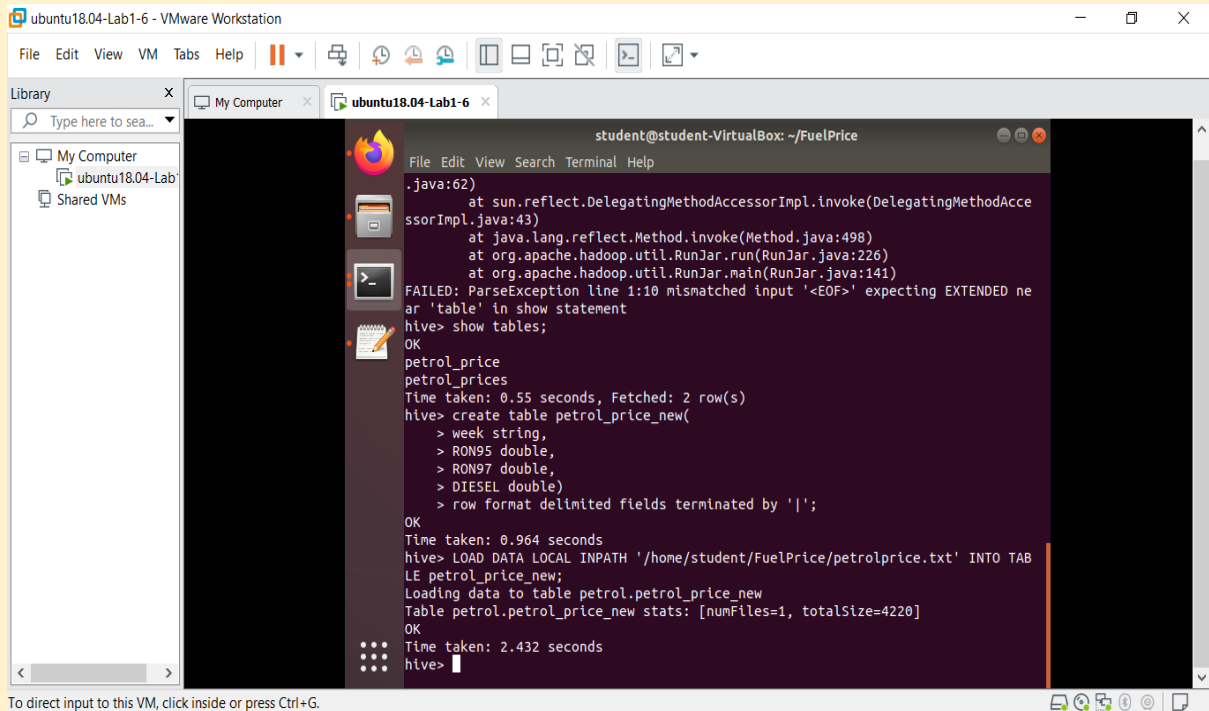
```
CREATE TABLE petrol_price_new (  
week string,  
RON95 double,  
RON97 double,  
DIESEL double)  
ROW FORMAT DELIMITED FIELDS TERMINATED BY '|';
```



```
student@student-VirtualBox: ~/FuelPrice
hive> show tables;
OK
Time taken: 0.078 seconds
hive> create table petrol_price (  
> week string,  
> RON95 int,  
> RON97 int,  
> DIESEL int)  
> row format delimited fields terminated by '|';  
OK  
Time taken: 0.572 seconds  
hive>
```

7. Load data from local file system into hive table.

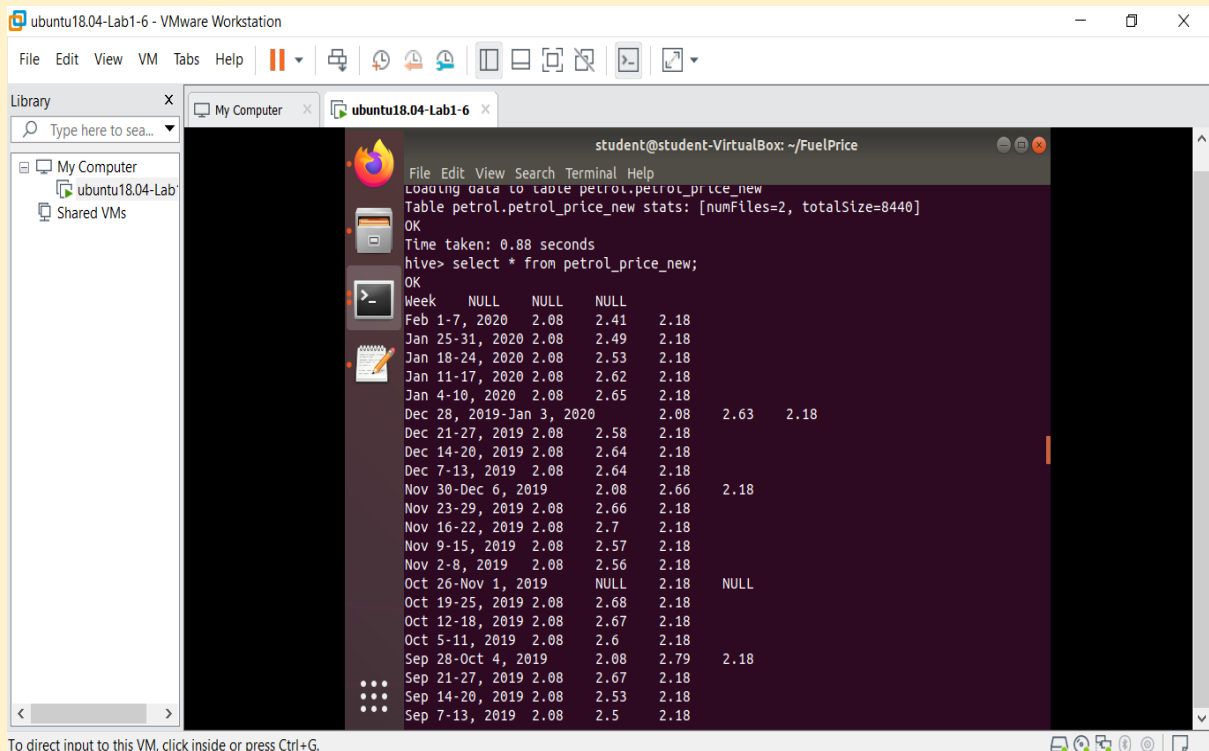
```
LOAD DATA LOCAL INPATH '/home/student/FuelPrice/petrolprice.txt' INTO TABLE petrol_price_new;
```



The screenshot shows a terminal window titled 'student@student-VirtualBox: ~/FuelPrice'. The terminal output is as follows:

```
student@student-VirtualBox: ~/FuelPrice
File Edit View Search Terminal Help
.java:62)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAcce
ssorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:226)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:141)
FAILED: ParseException line 1:10 mismatched input '<EOF>' expecting EXTENDED ne
ar 'table' in show statement
hive> show tables;
OK
petrol_price
petrol_prices
Time taken: 0.55 seconds, Fetched: 2 row(s)
hive> create table petrol_price_new(
> week string,
> RON95 double,
> RON97 double,
> DIESEL double)
> row format delimited fields terminated by '|';
OK
Time taken: 0.964 seconds
hive> LOAD DATA LOCAL INPATH '/home/student/FuelPrice/petrolprice.txt' INTO TAB
LE petrol_price_new;
Loading data to table petrol.ppetrol_price_new
Table petrol.ppetrol_price_new stats: [numFiles=1, totalSize=4220]
OK
Time taken: 2.432 seconds
hive>
```

8. Check and verified the data (petrolprice.txt) has successfully loaded into HIVE table (data warehouse)



The screenshot shows a terminal window titled 'student@student-VirtualBox: ~/FuelPrice'. The terminal output is as follows:

```
student@student-VirtualBox: ~/FuelPrice
File Edit View Search Terminal Help
Loading data to table petrol.ppetrol_price_new
Table petrol.ppetrol_price_new stats: [numFiles=2, totalSize=8440]
OK
Time taken: 0.88 seconds
hive> select * from petrol_price_new;
OK
Week      NULL      NULL      NULL
Feb 1-7, 2020  2.08      2.41      2.18
Jan 25-31, 2020 2.08      2.49      2.18
Jan 18-24, 2020 2.08      2.53      2.18
Jan 11-17, 2020 2.08      2.62      2.18
Jan 4-10, 2020  2.08      2.65      2.18
Dec 28, 2019-Jan 3, 2020 2.08      2.63      2.18
Dec 21-27, 2019 2.08      2.58      2.18
Dec 14-20, 2019 2.08      2.64      2.18
Dec 7-13, 2019  2.08      2.64      2.18
Nov 30-Dec 6, 2019 2.08      2.66      2.18
Nov 23-29, 2019 2.08      2.66      2.18
Nov 16-22, 2019 2.08      2.7       2.18
Nov 9-15, 2019  2.08      2.57      2.18
Nov 2-8, 2019   2.08      2.56      2.18
Oct 26-Nov 1, 2019 NULL      2.18      NULL
Oct 19-25, 2019 2.08      2.68      2.18
Oct 12-18, 2019 2.08      2.67      2.18
Oct 5-11, 2019  2.08      2.6       2.18
Sep 28-Oct 4, 2019 2.08      2.79      2.18
Sep 21-27, 2019 2.08      2.67      2.18
Sep 14-20, 2019 2.08      2.53      2.18
Sep 7-13, 2019  2.08      2.5       2.18
```