



**Continues Assessment**  
**WQD 7005 - DATA MINING**  
**Milestone 2 - Management of Data**  
**Store Data into HIVE Data Warehouse**

Instructor

Professor Dr. Teh Ying Wah

Group members

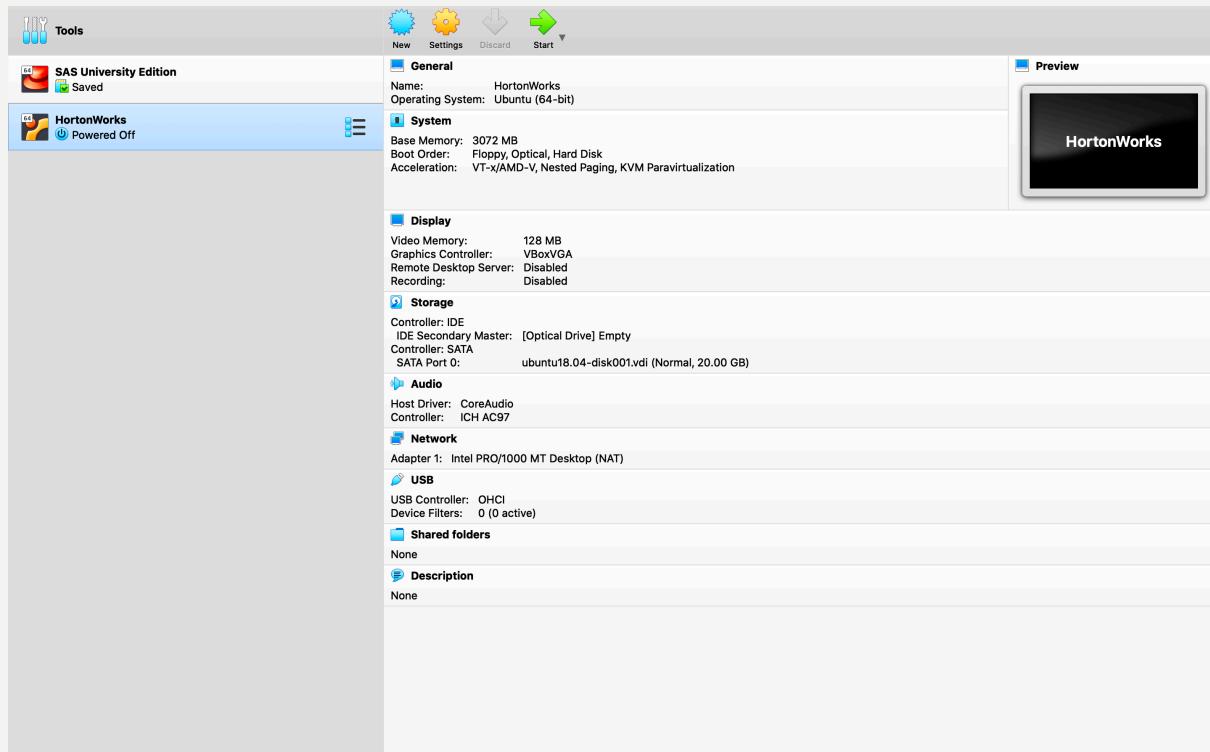
Gunasegarran Magadevan	17043640
Liu Hong Yang	17201091

## 1. Downloading Hortonworks Sandbox:

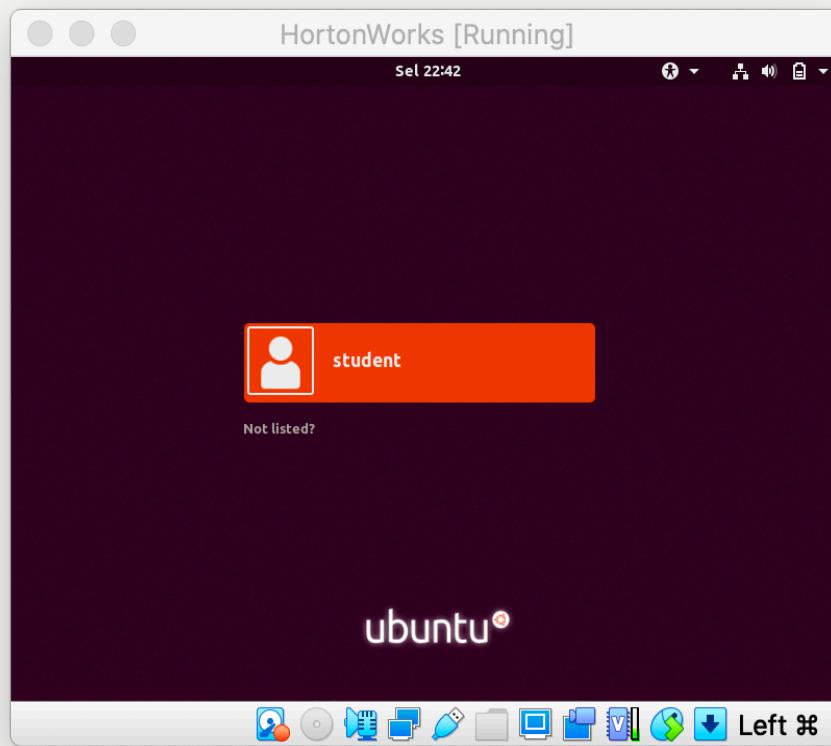
<https://www.cloudera.com/downloads/hortonworks-sandbox/hdp.html>

The screenshot shows the Cloudera website with the URL 'cloudera.com' in the address bar. The main header features the 'CLOUDERA' logo. Below it, there are navigation links for 'Why Cloudera', 'Products', 'Solutions', and 'Services & Support'. On the right side of the header are icons for phone, user profile, and search. The main content area has a large orange and blue background image of feathers. In the center, there's a section titled 'Get Started Now' with a dropdown menu set to 'Virtualbox' and a teal button labeled 'LET'S GO! →'. To the left, there's a section titled 'Hortonworks Data Platform (HDP®) on Hortonworks Sandbox' with a sub-section about HDP Sandbox making it easy to get started with Apache Hadoop, Apache Spark, Apache Hive, Apache HBase, Druid and Data Analytics Studio (DAS). Below this are links for 'HDP on Sandbox' (Getting Started with HDP, Learning the Ropes of the HDP Sandbox, Release Notes) and 'HDF on Sandbox' (Download HDF, Release Notes).

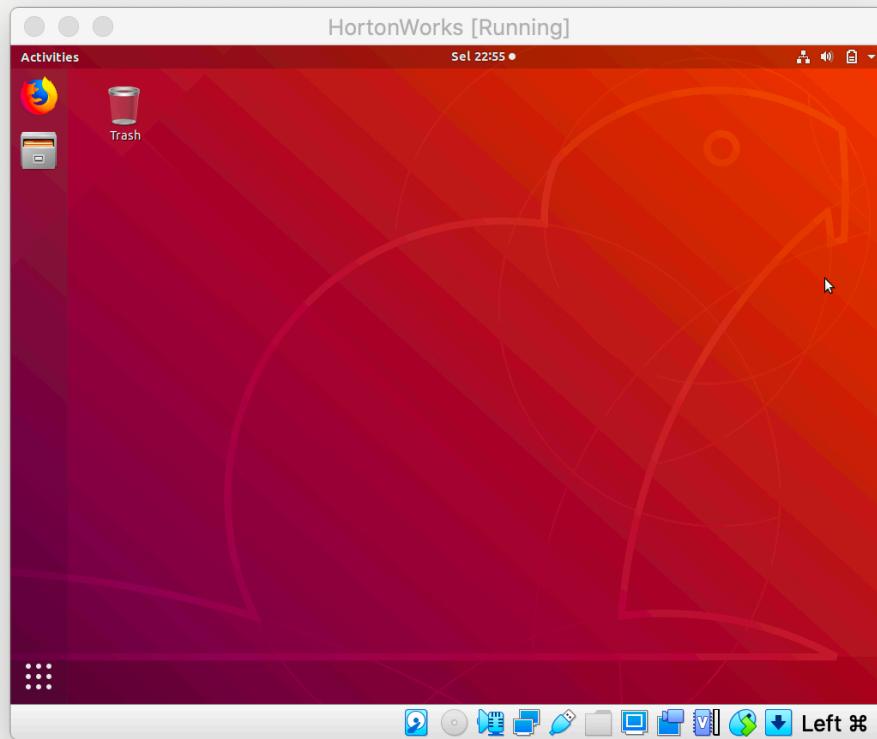
## 2. Importing the downloaded **HortonWorks.ova** in VirtualBox



### 3. Launch HortonWorks in VirtualBox



### 4. Successfully login to **HortonWorks** using **Student** credential

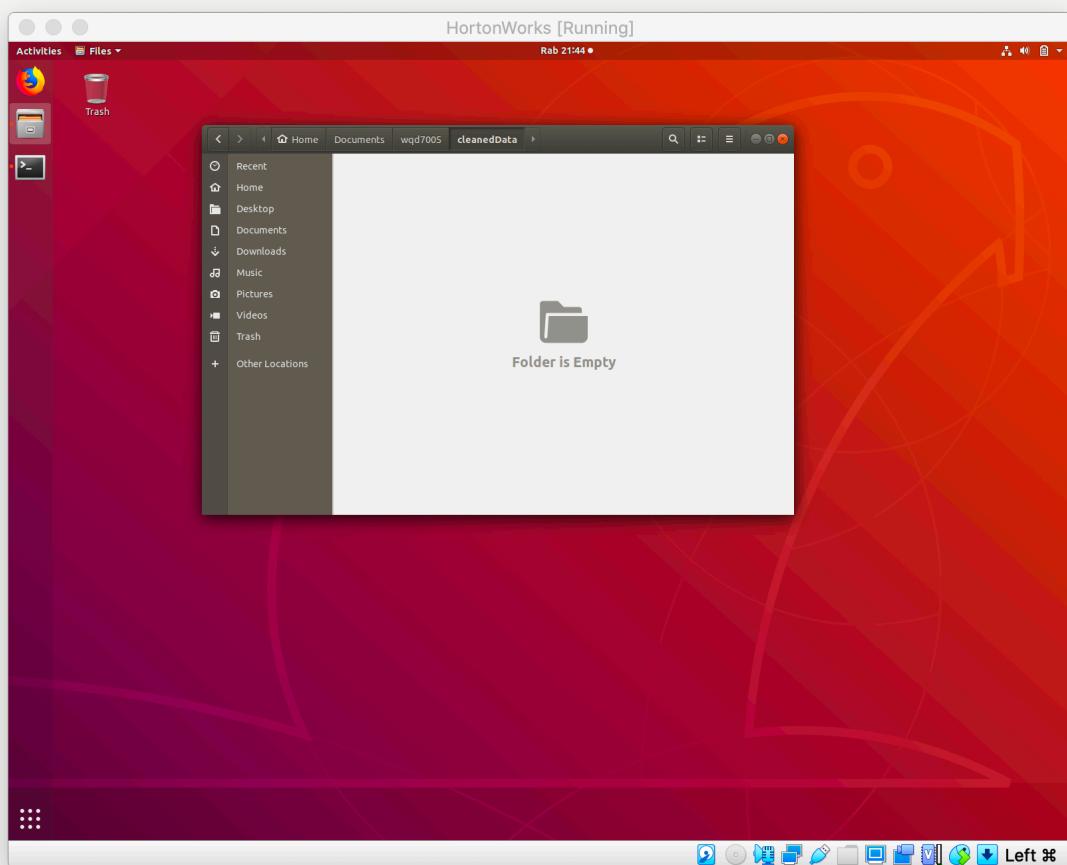


## 5. Created a directory into local system.

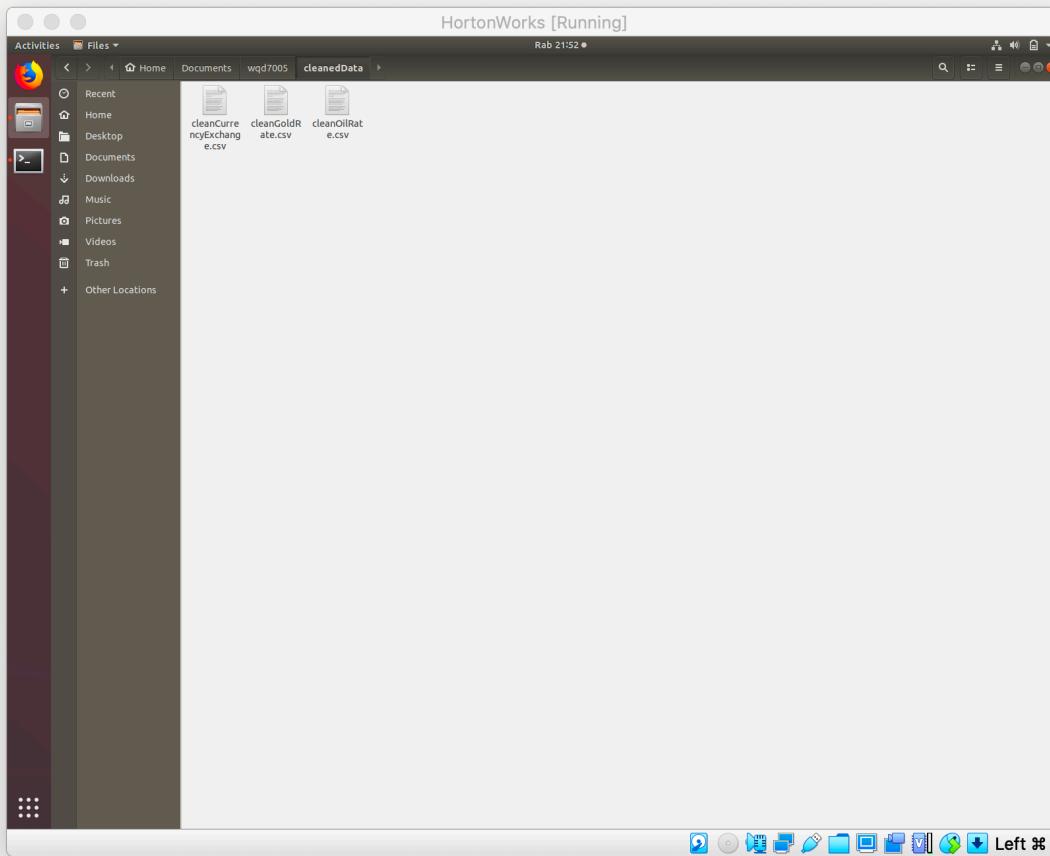
A screenshot of a Linux desktop environment. In the center is a terminal window titled "HortonWorks [Running]". The terminal shows the following command-line session:

```
student@student-VirtualBox: ~$ ls
apache-hive-1.2.2-bin.tar.gz  Downloads  hbase-1.2.12-bin.tar.gz  Pictures  Public  zookeeper-3.4.14.tar.gz
derby.log                    examples.desktop  input  pgp-0.16.v.tar.gz  snap  zookeeper.out
descriptions                 grep           metastore_db  pgp_15655452458.log  Templates
hadoop-2.7.7.tar.gz          Music  pig_15655509771.log  Videos
student@student-VirtualBox: ~$ cd Documents
student@student-VirtualBox: ~/Documents$ mkdir wqd7005
student@student-VirtualBox: ~/Documents$ cd wqd7005
student@student-VirtualBox: ~/Documents/wqd7005$ mkdir cleanedData
student@student-VirtualBox: ~/Documents/wqd7005$ ls
cleanedData
student@student-VirtualBox: ~/Documents/wqd7005$
```

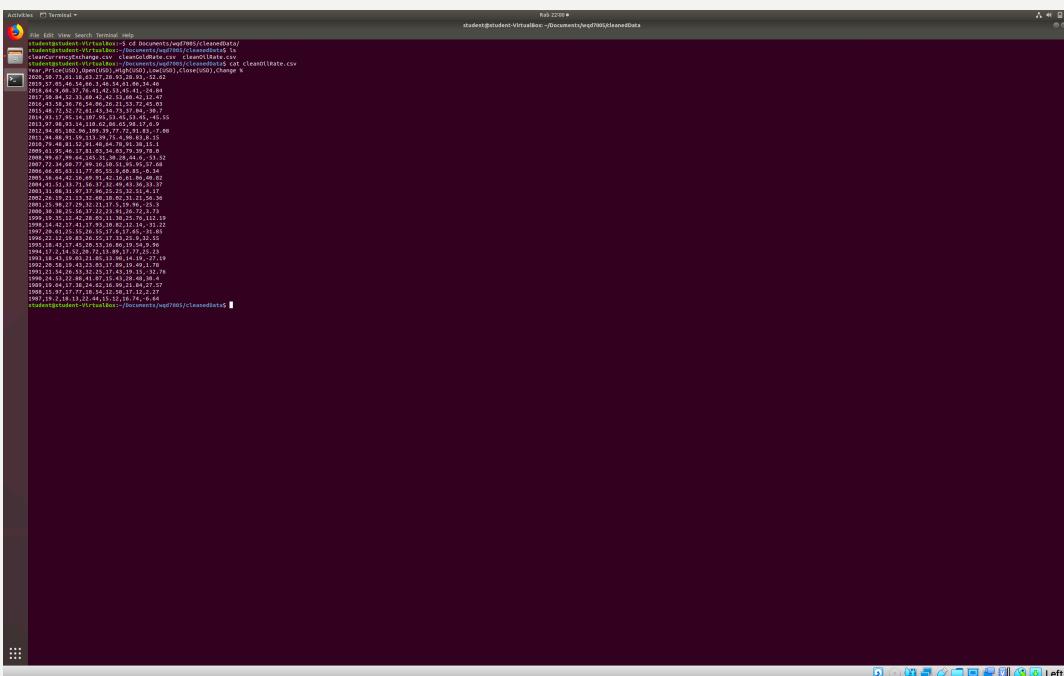
The desktop background is a dark red/orange gradient. The taskbar at the bottom has several icons, including a file manager icon.



6. Copy the cleaned data into newly created directory.

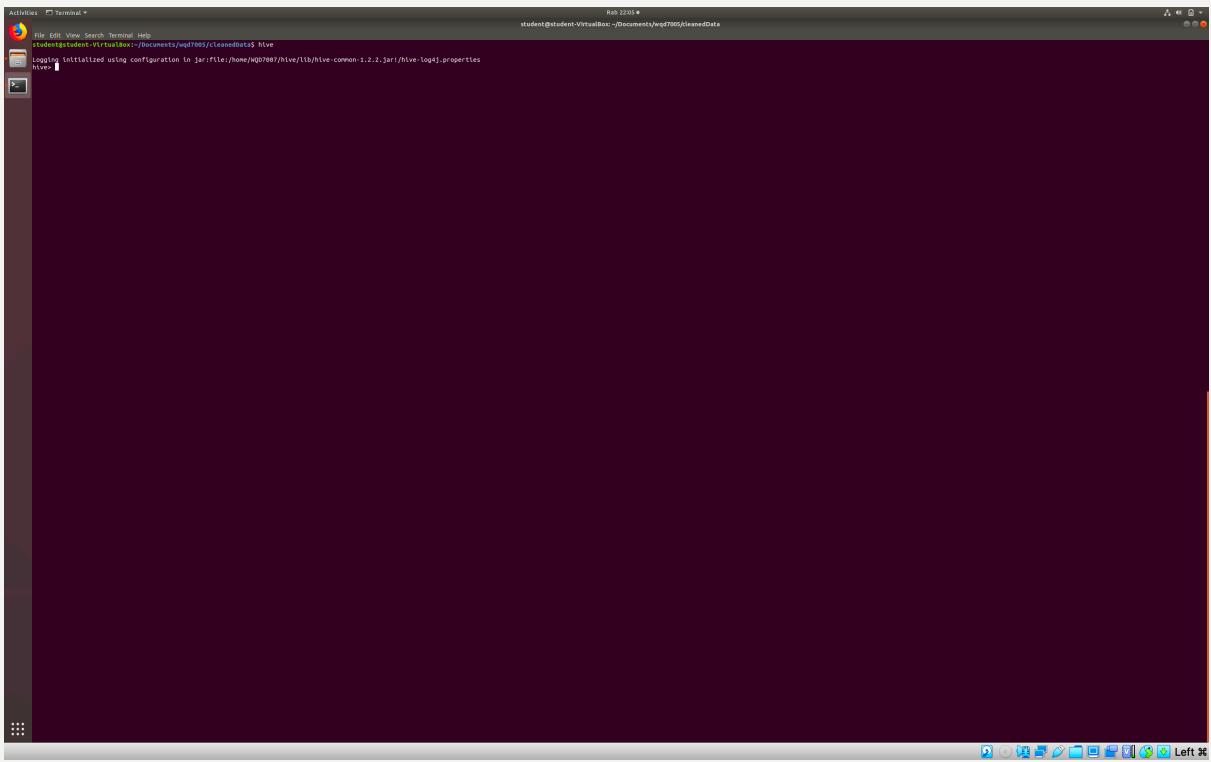


7. Locate the copied data and view the data files.



```
Activities Terminal
File Edit Help Search Terminal Help
Apr 87,454,10,437,09,478,59,415,09,11520,0.8,64
student@student-VirtualBox:~/.Documents/wqd7005/cleanedData$ cat cleanCurrencyExchange.csv
Mar 26,4,355,4,1235,4,3085,4,1595,-3,25
Mar 26,4,355,4,1235,4,3085,4,1595,-3,25
Jan 16,4,096,4,089,4,112,4,041,0,18
Jan 16,4,096,4,089,4,112,4,041,0,18
Feb 19,4,135,4,1765,4,3865,4,1215,-9,05
Feb 19,4,135,4,1765,4,3865,4,1215,-9,05
Jan 19,4,137,4,212,4,2205,4,162,-9,44
Jan 19,4,137,4,212,4,2205,4,162,-9,44
Jul 19,5,1205,4,2115,4,2455,4,1865,-8,15
Jul 19,5,1205,4,2115,4,2455,4,1865,-8,15
Jun 19,4,1325,4,1935,4,3935,-4,1115,-1,38
Jun 19,4,1325,4,1935,4,3935,-4,1115,-1,38
Apr 19,4,1345,4,0893,4,1475,4,4025,-2,27
Apr 19,4,1345,4,0893,4,1475,4,4025,-2,27
Mar 19,4,1345,4,0893,4,1475,4,4025,-2,27
Feb 19,4,0665,4,0855,4,1065,4,0665,9,72
Feb 19,4,0665,4,0855,4,1065,4,0665,9,72
Jan 18,4,0665,4,0855,4,1065,4,0665,9,72
Dec 18,4,1325,4,1865,4,1985,-1,1525,-1,24
Nov 18,4,1325,4,1865,4,1985,-1,1525,-1,24
Oct 18,4,1845,4,1365,4,1845,-4,1885,-1,11
Oct 18,4,1845,4,1365,4,1845,-4,1885,-1,11
Sep 18,4,1895,4,0015,4,1115,-4,0615,-1,09
Aug 18,4,1895,4,0015,4,1115,-4,0615,-1,09
Jul 18,4,0095,4,041,4,072,4,040,4,0615,-1,49
Jun 18,4,0095,4,041,4,072,4,040,4,0615,-1,49
May 18,3,98,3,924,3,995,-4,92,1,49
May 18,3,98,3,924,3,995,-4,92,1,49
Apr 18,3,98,3,924,3,995,-4,92,1,49
Mar 18,3,985,3,9395,3,9915,4,8585,-1,37
Feb 18,3,985,3,9395,3,9915,4,8585,-1,37
Jan 18,3,981,4,0455,4,0461,3,1,864,-1,61
Jan 18,3,981,4,0455,4,0461,3,1,864,-1,61
Oct 18,4,1355,4,138,4,228,-4,1825,-1,44
Oct 18,4,1355,4,138,4,228,-4,1825,-1,44
Sep 18,4,0895,4,2,154,-244,4,4082,-3,37
Sep 18,4,0895,4,2,154,-244,4,4082,-3,37
Jul 17,4,2385,4,2085,4,2085,-4,1815,-1,17
Jul 17,4,2385,4,2085,4,2085,-4,1815,-1,17
Jun 17,4,2385,4,2085,4,2085,-4,1815,-1,17
Jun 17,4,2385,4,2085,4,2085,-4,1815,-1,17
Jul 17,4,2385,4,2085,4,2085,-4,1815,-1,17
Jul 17,4,2385,4,2085,4,2085,-4,1815,-1,17
Jun 17,4,2385,4,2085,4,2085,-4,1815,-1,17
Apr 17,4,241,4,42,4,441,4,137,-1,93
Apr 17,4,241,4,42,4,441,4,137,-1,93
Mar 17,4,241,4,42,4,441,4,137,-1,93
Feb 17,4,44,4,432,4,46,4,425,-24
Jan 17,4,44,4,432,4,46,4,425,-24
Oct 16,4,465,4,4615,4,466,4,4205,-8,45
Oct 16,4,465,4,4615,4,466,4,4205,-8,45
Nov 16,4,1355,4,138,4,228,-4,1825,-1,44
Nov 16,4,1355,4,138,4,228,-4,1825,-1,44
Sep 16,4,0895,4,2,154,-244,4,4082,-3,37
Sep 16,4,0895,4,2,154,-244,4,4082,-3,37
Jul 16,4,2385,4,2085,4,2085,-4,1815,-1,17
Jul 16,4,2385,4,2085,4,2085,-4,1815,-1,17
Jun 16,4,2385,4,2085,4,2085,-4,1815,-1,17
Jun 16,4,2385,4,2085,4,2085,-4,1815,-1,17
Jul 16,4,2385,4,2085,4,2085,-4,1815,-1,17
Jul 16,4,2385,4,2085,4,2085,-4,1815,-1,17
Jun 16,4,2385,4,2085,4,2085,-4,1815,-1,17
Apr 16,4,241,4,42,4,441,4,137,-1,93
Apr 16,4,241,4,42,4,441,4,137,-1,93
Mar 16,4,241,4,42,4,441,4,137,-1,93
Feb 16,4,241,4,42,4,441,4,137,-1,93
Jan 16,4,241,4,42,4,441,4,137,-1,93
Oct 15,4,555,4,29,4,4,4,121,-3,23
Oct 15,4,555,4,29,4,4,4,121,-3,23
Sep 15,4,555,4,29,4,4,4,121,-3,23
Sep 15,4,555,4,29,4,4,4,121,-3,23
Jun 15,4,1355,4,138,4,228,-4,1825,-1,44
Jun 15,4,1355,4,138,4,228,-4,1825,-1,44
Oct 15,4,975,4,1955,4,4040,4,1415,-1,83
Oct 15,4,975,4,1955,4,4040,4,1415,-1,83
Jul 15,3,819,3,9734,3,8784,3,7808,-1,115
Jul 15,3,819,3,9734,3,8784,3,7808,-1,115
Jun 15,3,6655,3,6555,3,6555,3,6555,-1,89
Jun 15,3,6655,3,6555,3,6555,3,6555,-1,89
Apr 15,3,9625,3,9795,3,7253,-4,154,-3,981
Apr 15,3,9625,3,9795,3,7253,-4,154,-3,981
Mar 15,3,9625,3,9795,3,7253,-4,154,-3,981
Feb 15,3,911,3,629,3,645,3,539,-0,54
Feb 15,3,911,3,629,3,645,3,539,-0,54
Jan 15,3,911,3,629,3,645,3,539,-0,54
Dec 14,3,985,3,985,3,9855,4,3825,-3,37
Dec 14,3,985,3,985,3,9855,4,3825,-3,37
Nov 14,3,985,3,985,3,9855,4,3825,-3,37
Oct 14,3,985,3,985,3,9855,4,3825,-3,37
Sep 14,3,1515,3,1504,22,3,1345,-1,42
Sep 14,3,1515,3,1504,22,3,1345,-1,42
Aug 14,3,1515,3,1504,22,3,1345,-1,42
Aug 14,3,1515,3,1504,22,3,1345,-1,42
Jul 14,3,211,3,267,3,265,3,265,-0,44
Jul 14,3,211,3,267,3,265,3,265,-0,44
Jun 14,3,216,3,267,3,269,3,262,-1,43
Jun 14,3,216,3,267,3,269,3,262,-1,43
Mar 14,3,9445,3,7755,3,3112,-2,2695,-0,4
Mar 14,3,9445,3,7755,3,3112,-2,2695,-0,4
Feb 14,3,9445,3,7755,3,3112,-2,2695,-0,4
Feb 14,3,9445,3,7755,3,3112,-2,2695,-0,4
Jan 14,3,9465,3,785,3,348,-4,156,2,07
Jan 14,3,9465,3,785,3,348,-4,156,2,07
Dec 13,3,2425,3,3155,3,2395,-4,1652,-1,15
Dec 13,3,2425,3,3155,3,2395,-4,1652,-1,15
Nov 13,3,2425,3,3155,3,2395,-4,1652,-1,15
Nov 13,3,2425,3,3155,3,2395,-4,1652,-1,15
Oct 13,3,2425,3,3155,3,2395,-4,1652,-1,15
Oct 13,3,2425,3,3155,3,2395,-4,1652,-1,15
Sep 13,3,558,3,289,3,332,3,3485,-9,07
Sep 13,3,558,3,289,3,332,3,3485,-9,07
Aug 13,3,558,3,289,3,332,3,3485,-9,07
Aug 13,3,558,3,289,3,332,3,3485,-9,07
Jul 13,3,4665,3,1835,3,253,3,151,2,75
Jul 13,3,4665,3,1835,3,253,3,151,2,75
Jun 13,3,4665,3,1835,3,253,3,151,2,75
Jun 13,3,4665,3,1835,3,253,3,151,2,75
May 13,3,5995,3,681,3,899,2,169,-1,68
May 13,3,5995,3,681,3,899,2,169,-1,68
Apr 13,3,5945,3,6944,3,899,3,619,-1,65
Apr 13,3,5945,3,6944,3,899,3,619,-1,65
Feb 13,3,992,3,1061,3,112,3,0831,-0,48
Feb 13,3,992,3,1061,3,112,3,0831,-0,48
Jan 13,3,992,3,1061,3,112,3,0831,-0,48
Dec 12,3,958,3,04,3,072,-3,037,-0,5
Dec 12,3,958,3,04,3,072,-3,037,-0,5
Oct 12,3,9840,3,0555,3,084,3,0275,-0,34
```

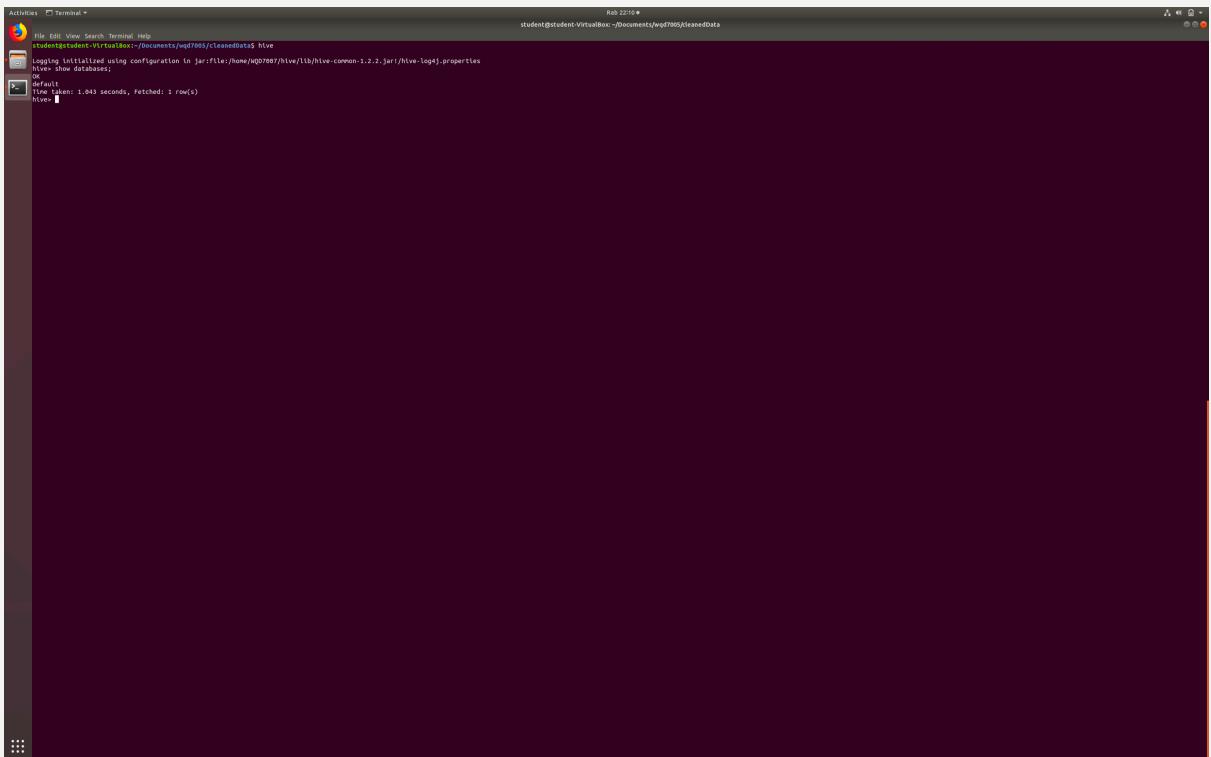
## 8. Launching “HIVE”.



A screenshot of a Linux desktop environment showing a terminal window titled "Terminal". The terminal window has a dark background and contains the following text:

```
Activities Terminal * student@student-VirtualBox: ~/Documents/wqd7005/cleanedData
student@student-VirtualBox:~/Documents/wqd7005/cleanedData$ hive
Logging initialized using configuration in jarfile:/home/wqd7005/hive/lib/hive-common-1.2.1.jar!/hive-log4j.properties
hive>
```

## 9. Checking any existing databases.



A screenshot of a Linux desktop environment showing a terminal window titled "Terminal". The terminal window has a dark background and contains the following text:

```
Activities Terminal * student@student-VirtualBox: ~/Documents/wqd7005/cleanedData
student@student-VirtualBox:~/Documents/wqd7005/cleanedData$ hive
Logging initialized using configuration in jarfile:/home/wqd7005/hive/lib/hive-common-1.2.2.jar!/hive-log4j.properties
hive> show databases;
OK
default
Time taken: 1.043 seconds, Fetched: 1 row(s)
hive>
```

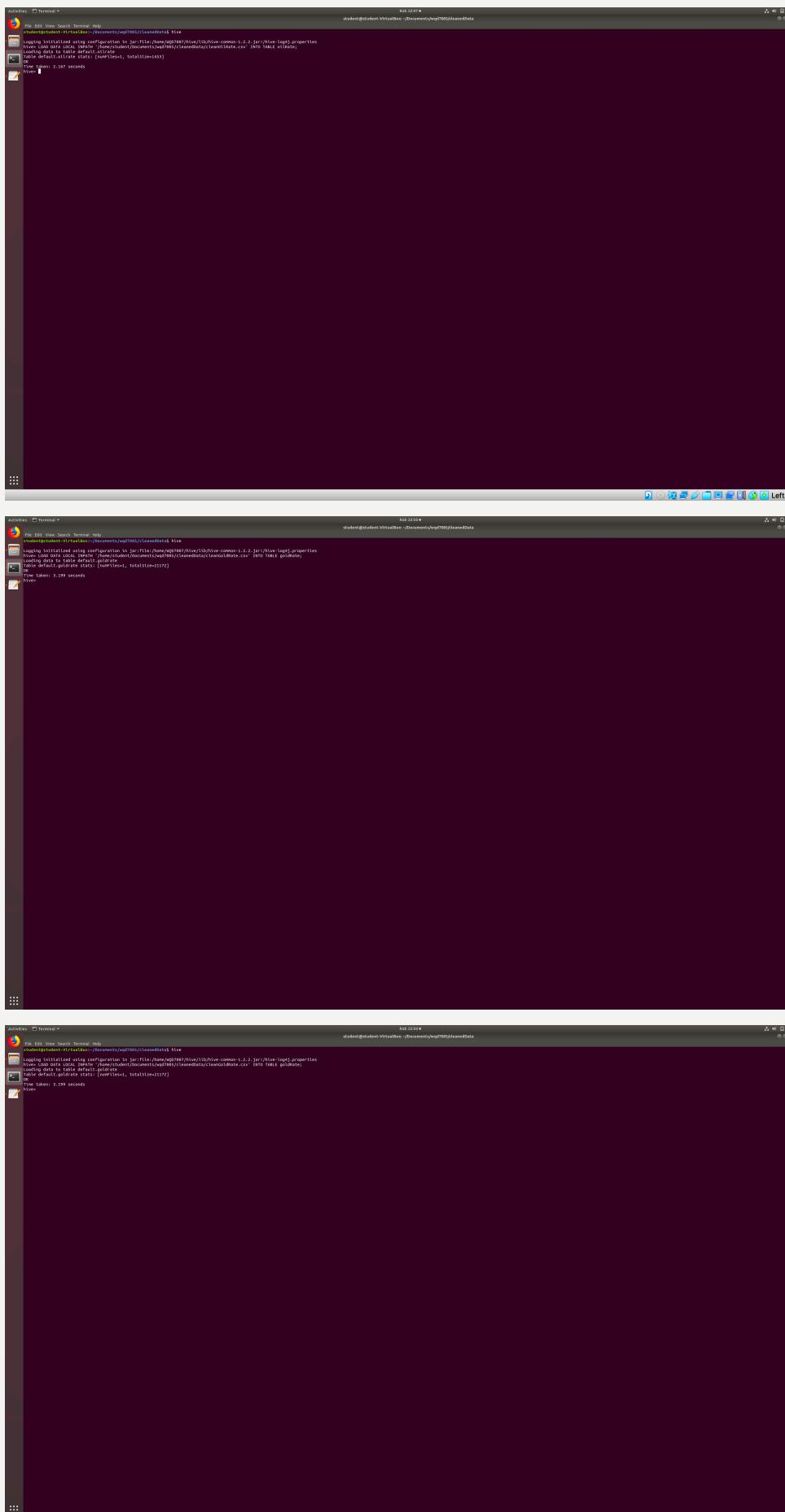
## 10. Create new tables for the cleaned data with defining all the required columns.

```
Analytics 3 Terminal 1
File Edit View Search Terminal Help
student@student-VirtualBox:~/Documents/uog705/cleanedData$ hive
Logging initialized using configuration in /etc/hive/conf.dist/hive-log4j.properties
hive> create table alldata (monthyear string,
>   avg double,
>   high double,
>   low double,
>   close double,
>   open double);
OK
Time taken: 1.45 seconds
hive>

Analytics 3 Terminal 2
File Edit View Search Terminal Help
student@student-VirtualBox:~/Documents/uog705/cleanedData$ hive
Logging initialized using configuration in /etc/hive/conf.dist/hive-log4j.properties
hive> create table currencyexchange (monthyear string,
>   avg double,
>   open double,
>   high double,
>   low double,
>   close double);
OK
Time taken: 0.23 seconds
hive>

Analytics 3 Terminal 3
File Edit View Search Terminal Help
student@student-VirtualBox:~/Documents/uog705/cleanedData$ hive
Logging initialized using configuration in /etc/hive/conf.dist/hive-log4j.properties
hive> create table goldrate (monthyear string,
>   avg double,
>   open double,
>   high double,
>   low double,
>   close double);
OK
Time taken: 0.17 seconds
hive>
```

## 11. Load data from local file system into hive table.



The image consists of three vertically stacked screenshots of a Linux terminal window. Each screenshot shows the command being run and its output. The terminal window has a dark background with light-colored text. The command is:

```
student@student-VirtualBox:~/Documents$ hdfs dfs -put /home/student/Downloads/cleandata.csv /student/hive/warehouse/cleandata
```

The output of each command is identical:

```
Logging initialized using configuration in /usr/lib/hadoop-mapreduce/hadoop-mapreduce-log4j.properties
HDFS Data Local Dir: /home/student/Downloads/cleandata
Input paths: /home/student/Downloads/cleandata
Output path: /student/hive/warehouse/cleandata
Table default: cleandata
Status: {numFiles=1, totalSize=2157}
Time taken: 3.099 seconds
hdfs:
```

12. Check and verified the data are successfully loaded into HIVE table (data warehouse).

```
student@student-VirtualBox:~/Documents$ ./hql/hive -e "load data local into table student_overview select * from student;"
```

**THANK  
YOU**