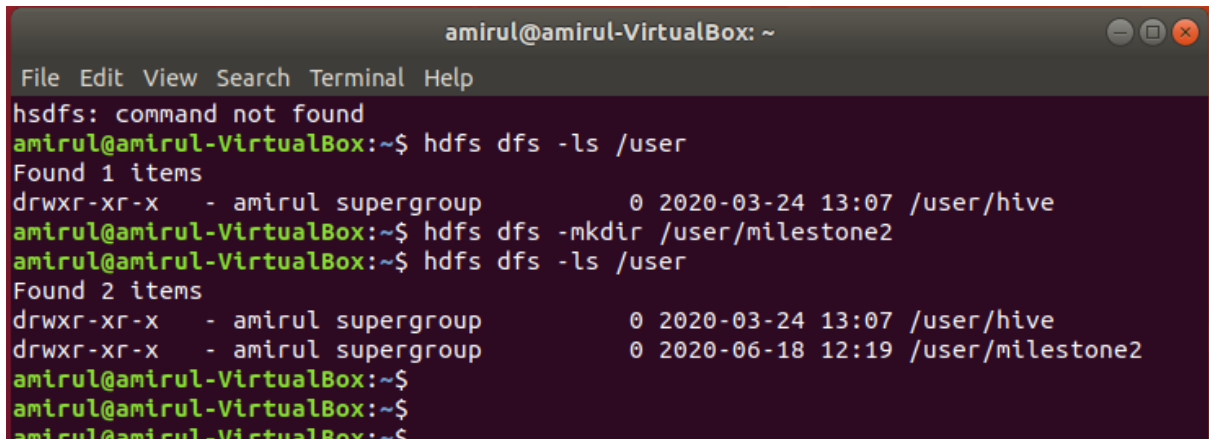Name: Mohd Amirul Shafiq Bin Shafiee
Matric No: WQD180114
Milestone 2: **Management of Data Store Data Into Hive Data Warehouse**

Step 1 – Create a directory
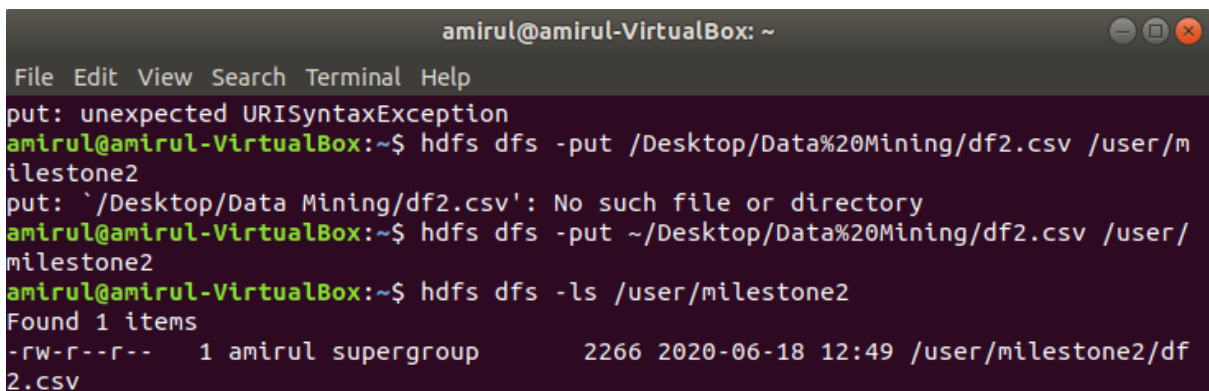
- Create directory in HDF, which we will call 'milestone2'
- Check if the directory exists

Step 2 – Import Data

- The data (df2) is on desktop. Import it to HDFS directory created earlier (milestone2)
- Check if data is successfully imported

Step 3 – Check the Data
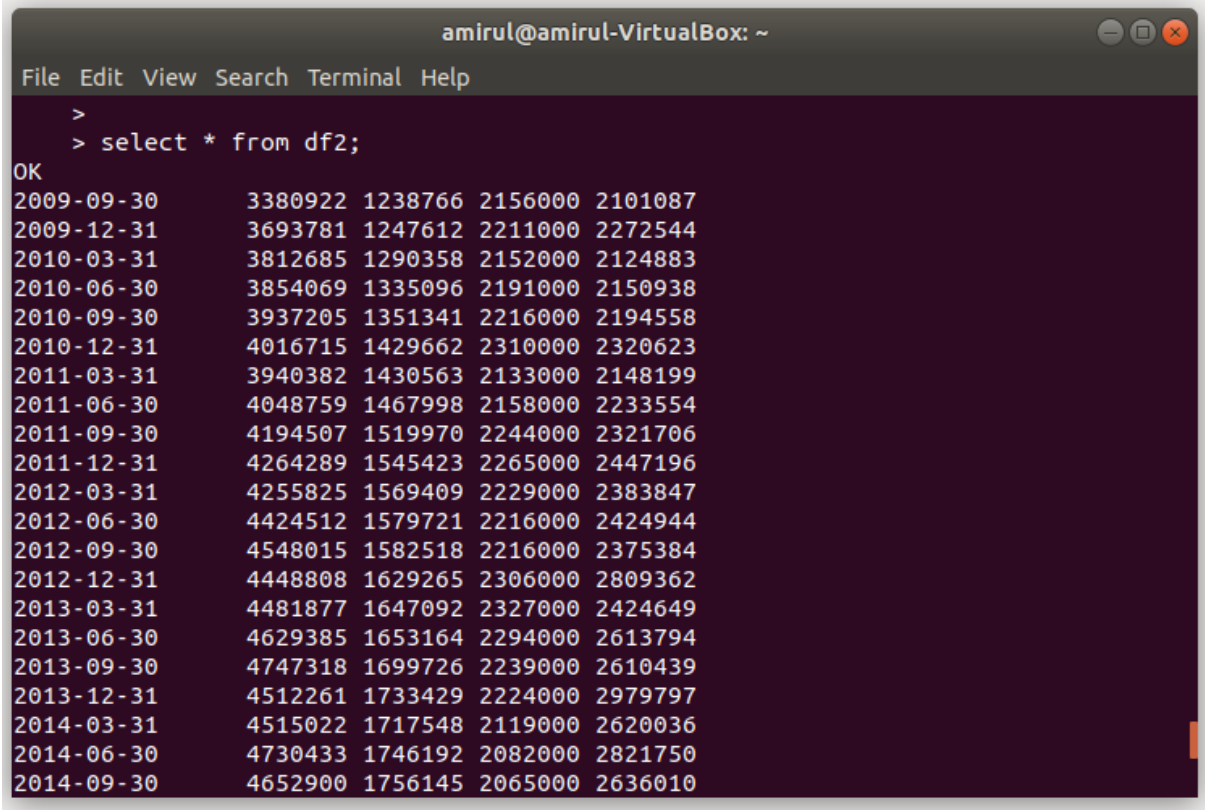
- Read the head of the data

```
amirul@amirul-VirtualBox: ~

File  Edit  View  Search  Terminal  Help
amirul@amirul-VirtualBox:~$ hdfs dfs -cat /user/milestone2/df2.csv |head -5
Quarter,Axiata,Digi,Maxis,TM
2009-09-30,3380922.0,1238766.0,2156000.0,2101087.0
2009-12-31,3693781.0,1247612.0,2211000.0,2272544.0
2010-03-31,3812685.0,1290358.0,2152000.0,2124883.0
2010-06-30,3854069.0,1335096.0,2191000.0,2150938.0
```

Step 4 – Create table in Hive

- Use query to create table with 5 columns and define the data type
- Remove first row as it is the column names

```
amirul@amirul-VirtualBox: ~

File  Edit  View  Search  Terminal  Help
hive> CREATE EXTERNAL TABLE IF NOT EXISTS df2 (QUARTER string, Axiata int, Digi
int, Maxis int, TM int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS
TEXTFILE LOCATION '/user/milestone2' TBLPROPERTIES ("skip.header.line.count"="1"
);
OK
Time taken: 1.535 seconds
hive> DESCRIBE DF2;
OK
quarter                 string
axiata                  int
digi                    int
maxis                   int
tm                      int
Time taken: 0.391 seconds, Fetched: 5 row(s)
hive>
```

Step 5 – Select all data in DF2 in HIVE

```
                              amirul@amirul-VirtualBox: ~

 File  Edit  View  Search  Terminal  Help
     >
     > select * from df2;
OK
2009-09-30      3380922 1238766 2156000 2101087
2009-12-31      3693781 1247612 2211000 2272544
2010-03-31      3812685 1290358 2152000 2124883
2010-06-30      3854069 1335096 2191000 2150938
2010-09-30      3937205 1351341 2216000 2194558
2010-12-31      4016715 1429662 2310000 2320623
2011-03-31      3940382 1430563 2133000 2148199
2011-06-30      4048759 1467998 2158000 2233554
2011-09-30      4194507 1519970 2244000 2321706
2011-12-31      4264289 1545423 2265000 2447196
2012-03-31      4255825 1569409 2229000 2383847
2012-06-30      4424512 1579721 2216000 2424944
2012-09-30      4548015 1582518 2216000 2375384
2012-12-31      4448808 1629265 2306000 2809362
2013-03-31      4481877 1647092 2327000 2424649
2013-06-30      4629385 1653164 2294000 2613794
2013-09-30      4747318 1699726 2239000 2610439
2013-12-31      4512261 1733429 2224000 2979797
2014-03-31      4515022 1717548 2119000 2620036
2014-06-30      4730433 1746192 2082000 2821750
2014-09-30      4652900 1756145 2065000 2636010
```