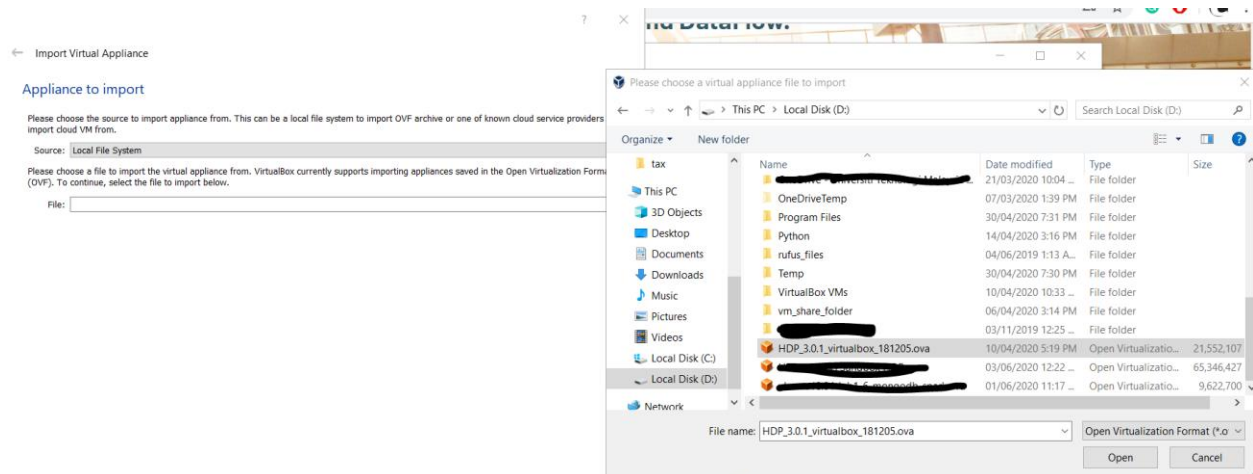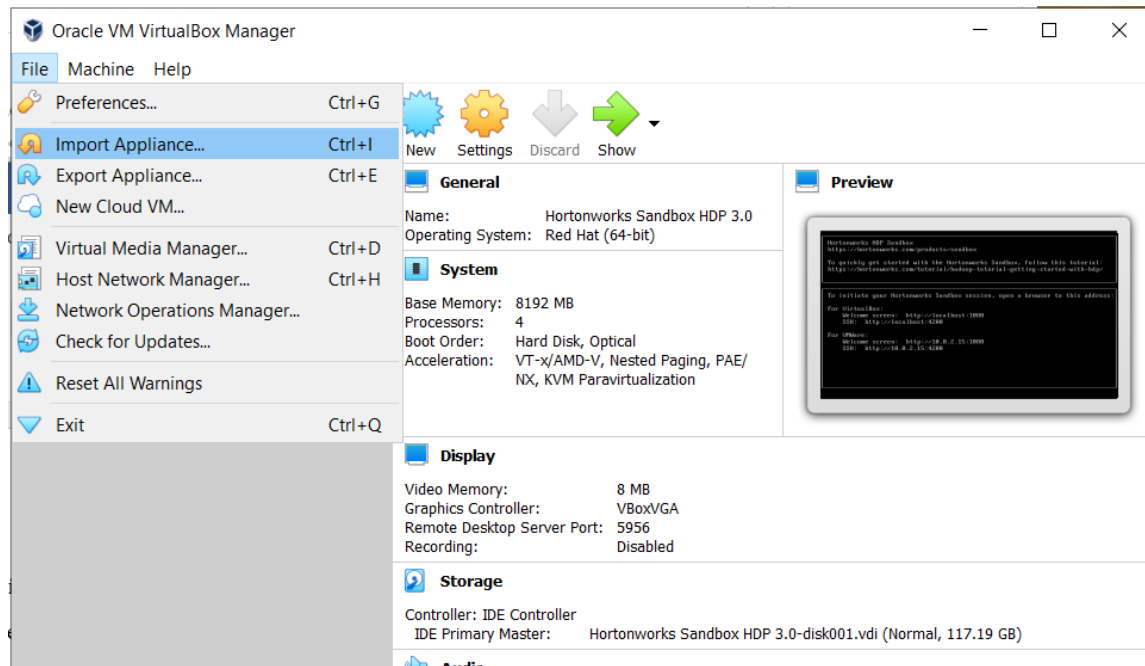## Milestone 2 – Store data into hive data warehouse

PART 1 Install Hortonworks HDP

Hortonworks sandbox will be used for the milestone 2.

1. Download Hortonworks HDP from https://www.cloudera.com/downloads/hortonworks-sandbox.html (20GB).

2. Import .ova into VM Virtualbox.

*Optimize the number of CPU cores and RAM resources before import.

3. Run the Hortonworks Sandbox HDP 3.0 for extraction and installation. Notice that first time installation will take around 15 minutes.
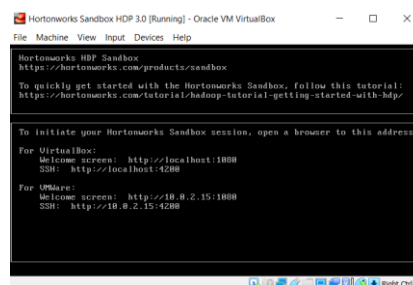
*Installation is completed

4. In the web browser, go to web shell client by accessing http://localhost:4200/. The default root user login credentials will be:

User: root

Password: hadoop

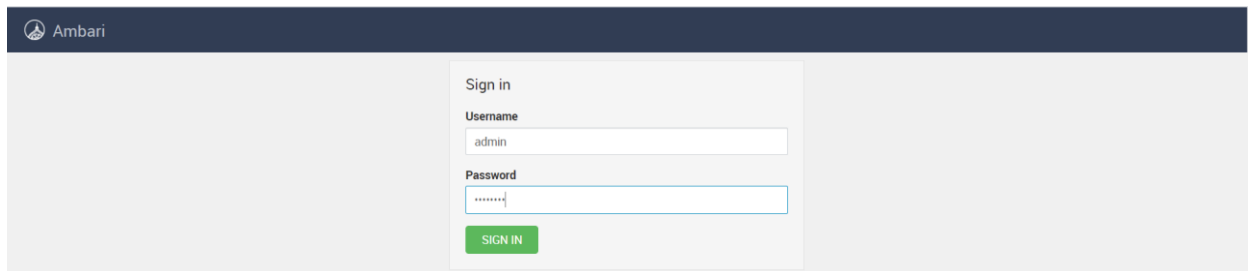After logging in by default password, you will be requested to change password. Please change the password for root user.

```
sandbox-hdp login: root
root@sandbox-hdp.hortonworks.com's password:
Last login: Wed Jun  3 23:18:14 2020
[root@sandbox-hdp ~]# 
```

5. Ambari enables system administrators to provision, manage and monitor a Hadoop cluster. Now, type 'ambari-admin-password-reset' to reset the Ambari' administrator password.

```
sandbox-hdp login: root
root@sandbox-hdp.hortonworks.com's password:
Last login: Wed Jun  3 23:18:14 2020
[root@sandbox-hdp ~]# ambari-admin-password-reset
Please set the password for admin: 
```

6. Access to http://localhost:8080/ for Ambari login.

7. After logging in, you will see the sandbox is taking some time to starting all the required services (around 15 mins).

8. The hive and HDFS are completely set up and ready to be used.

PART 2 Store data into hive database

1. Download and install WinSCP

https://winscp.net/eng/download.php

2. Send the local Windows' data into sandbox (VM) by using WinSCP.
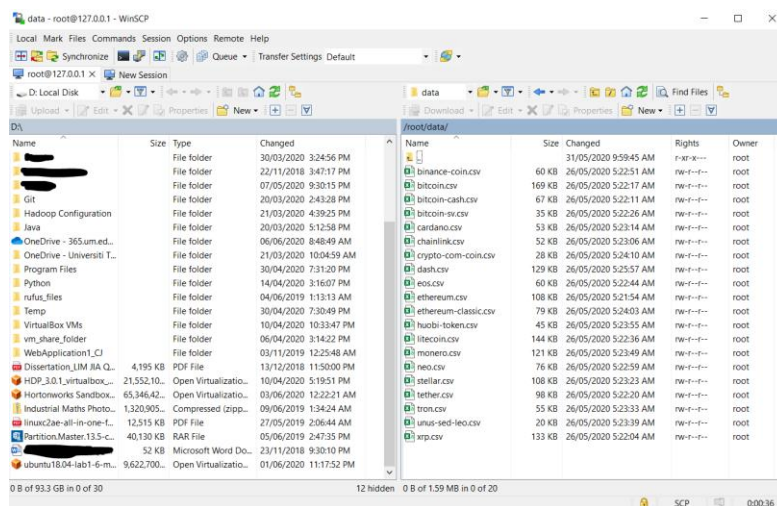
3. Go to web shell client,

hdfs fs -put /root/data/*.csv /user/root/datamining/data/

This is to copy the file from root directory into another directory of HDFS.

4. Open hive in web shell client, use the database preferred, then create a table under the database.

Eg. Table 'bitcoin'

```
CREATE TABLE bitcoin (MarketDate DATE, Open DOUBLE, High DOUBLE, Low DOUBLE, Close
DOUBLE, Volume DOUBLE, MarketCapacity DOUBLE)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

TBLPROPERTIES("skip.header.line.count"="1");
```

Then load the csv data from HDFS into bitcoin table.

```
LOAD DATA INPATH '/user/root/datamining/data/coin.csv' INTO TABLE bitcoin;
```