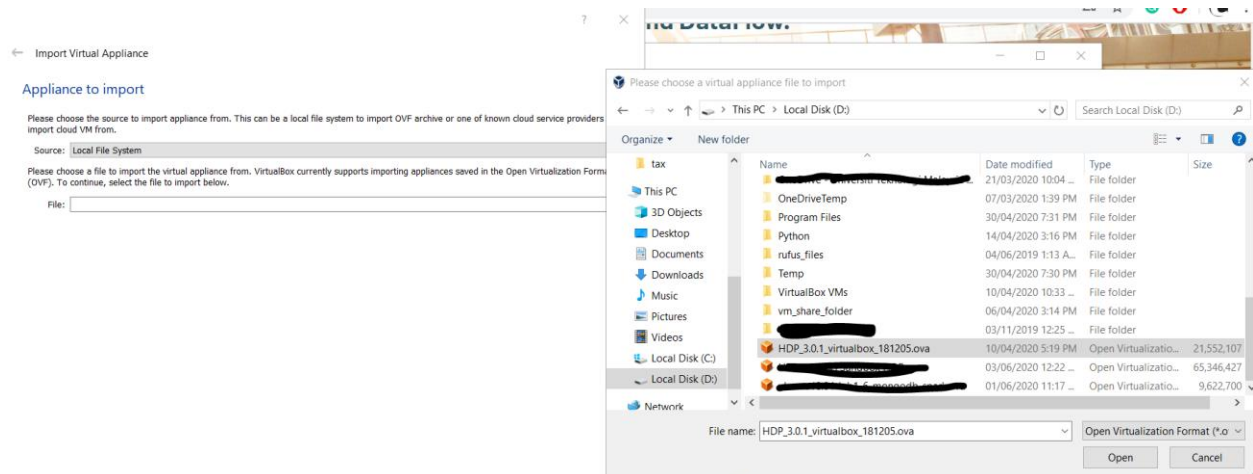
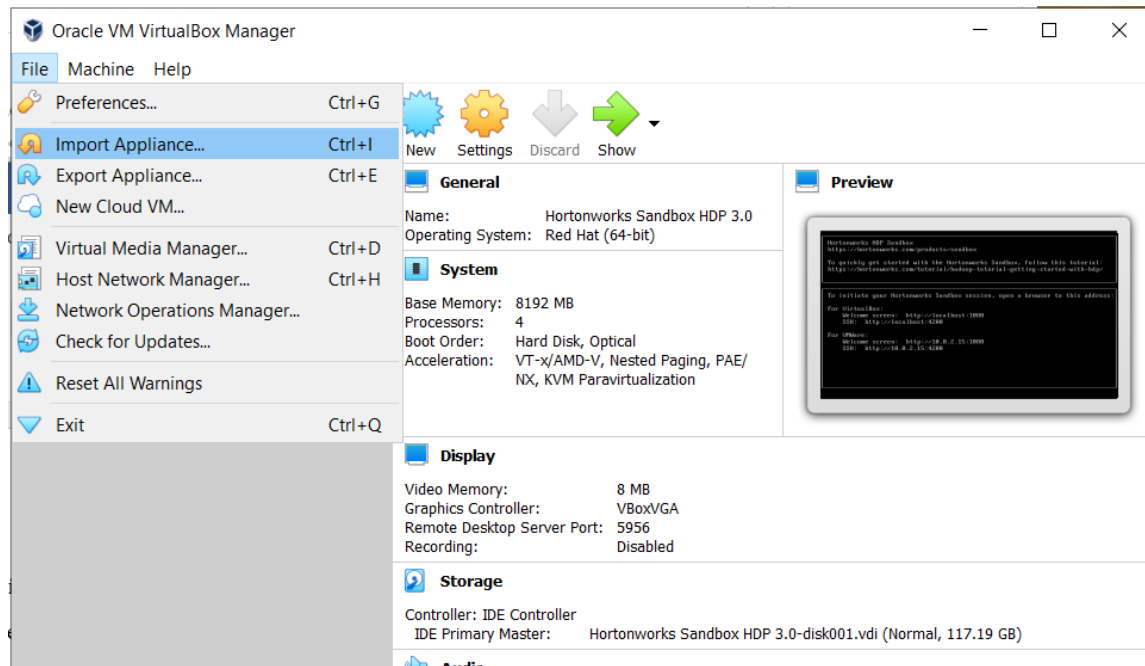


## Milestone 2 – Store data into hive data warehouse

### 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.



## ← Import Virtual Appliance

## Appliance settings

These are the virtual machines contained in the appliance and the suggested settings of the imported VirtualBox machines. You can change many of the properties shown by double-clicking on the items and disable others using the check boxes below.

Virtual System 1	
Name	Hortonworks Sandbox HDP 3.0 1
Guest OS Type	Red Hat (64-bit)
CPU	4
RAM	8192 MB
USB Controller	<input checked="" type="checkbox"/>
Sound Card	<input checked="" type="checkbox"/> ICH AC97
Network Adapter	<input checked="" type="checkbox"/> Intel PRO/1000 MT Desktop (82540EM)
Storage Controller (IDE)	PIIX4
Storage Controller (IDE)	PIIX4
Virtual Disk Image	Hortonworks Sandbox HDP 3.0-disk001.vmdk
Base Folder	C:\Users\Tan Chang Jung\VirtualBox VMs
Primary Group	/

Machine Base Folder:

MAC Address Policy:

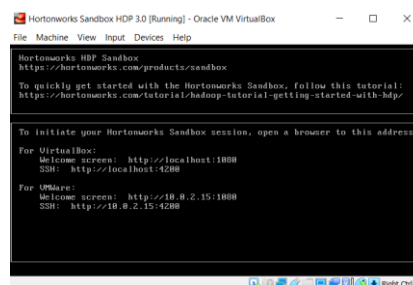
Additional Options: ☒ Import hard drives as VDI

Appliance is not signed

Restore Defaults **Import** Cancel

\*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

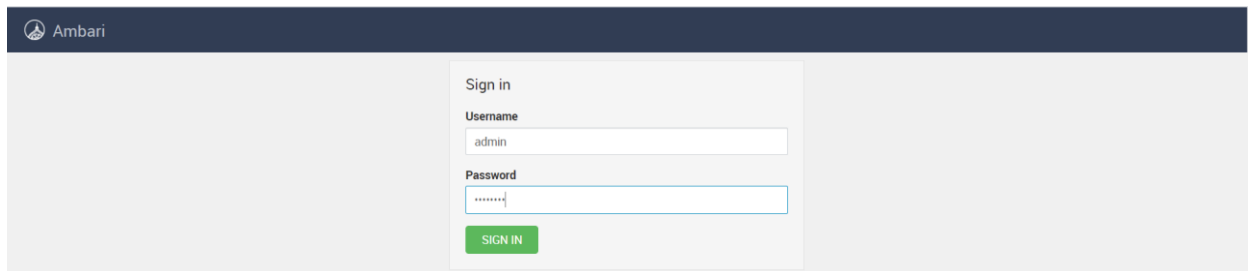
After logging in by default password, you will be requested to change password. Please change the password for root user.

```
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.



Ambari

Sign in

Username  
admin

Password  
\*\*\*\*\*

SIGN IN

## TAN CHANG JUNG

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

Background Operations

0 Background Operations Running

ALL (10)

✓ Start All Services	100%	raj_ops	Sun May 31 2020 16:01	11m 52s
✓ Start All Services	100%	raj_ops	Wed May 27 2020 07:52	14m 29s
✓ Restart all components with Stale Configs for Hive	100%	admin	Tue May 26 2020 18:05	10s
✓ Start Hive	100%	admin	Tue May 26 2020 18:02	1m 45s
➔ Start All Services	100%	raj_ops	Tue May 26 2020 17:52	7m 6s
✓ Stop required services	100%	admin	Fri Nov 30 2018 03:17	38s

☐ Do not show this dialog again when starting a background operation

OK

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

### Summary

#### Components

Started

HIVE METASTORE

Started

HIVESERVER2

Started

MYSQL SERVER

1 Installed

HIVE CLIENT

HIVESERVER2 JDBC URL

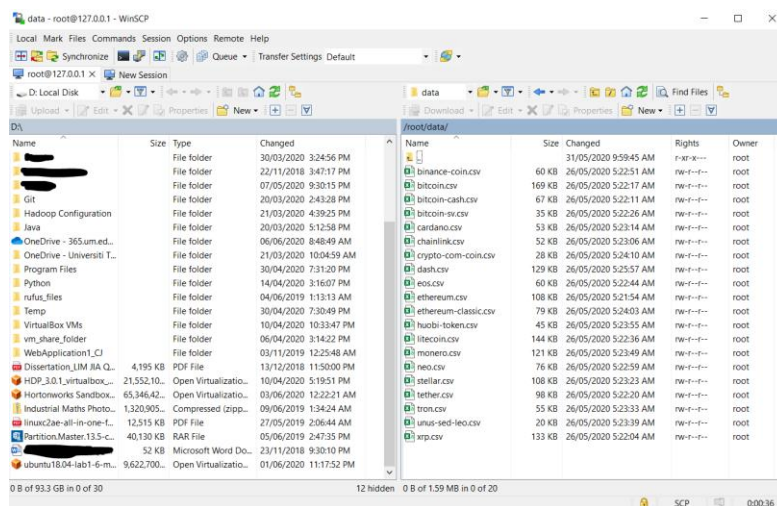
jdbc:hive2://sandbox-hdp.hortonworks.com:2181/?serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2

## 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.



## TAN CHANG JUNG

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;
```