CPRE/SE 419

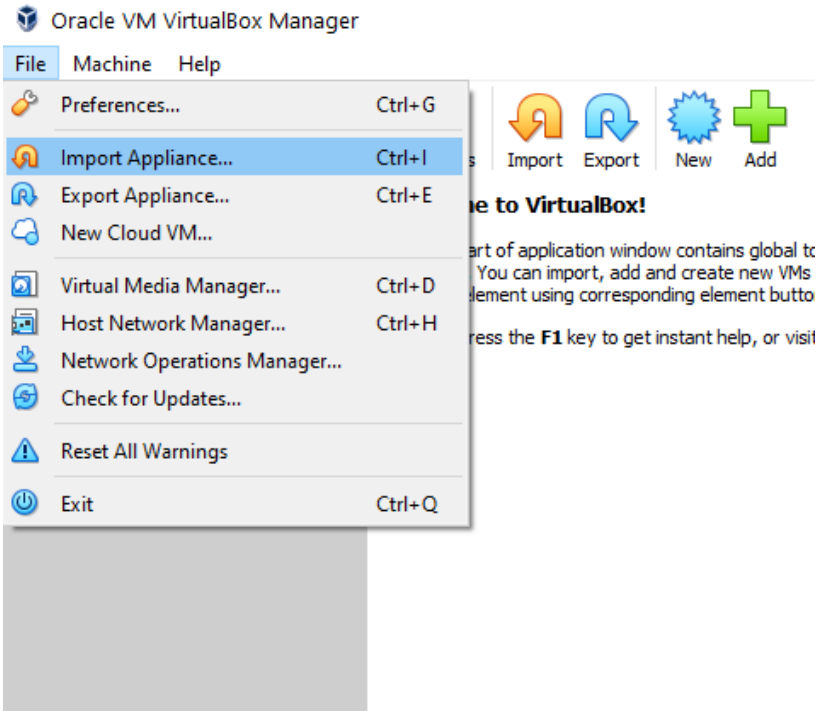
Install Virtual Machine Instance on local machine

1. Download and import .ova VM instance

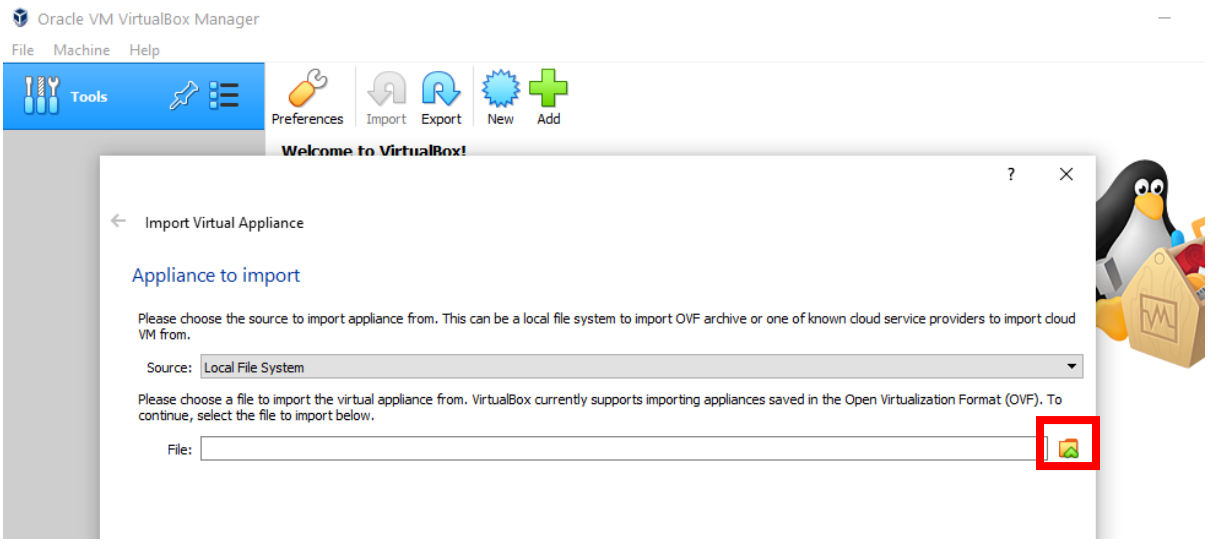
1) Please make sure Oracle VirtualBox has been installed on your machine. If not, here is the link for download, https://www.virtualbox.org/wiki/Downloads

2) Find .ova virtual machine instance from Cybox, https://iastate.box.com/s/3llu4dwpioh6l2ylrm94ybtfcba0uok3 . You should be able to download zip file "cpre419_2021_ova.zip".

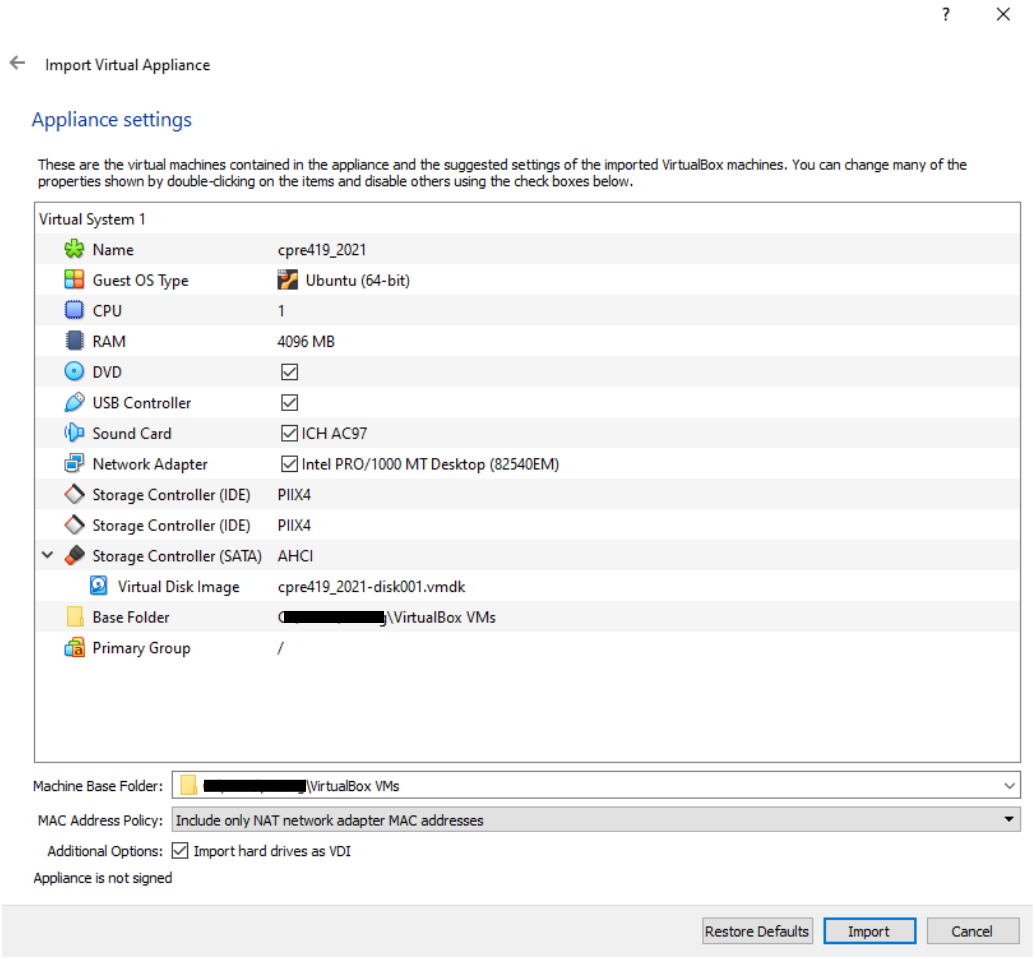
3) Unzip it when download has been completed. Open Oracle VirtualBox and choose "File" -> "Import Virtual Appliance".



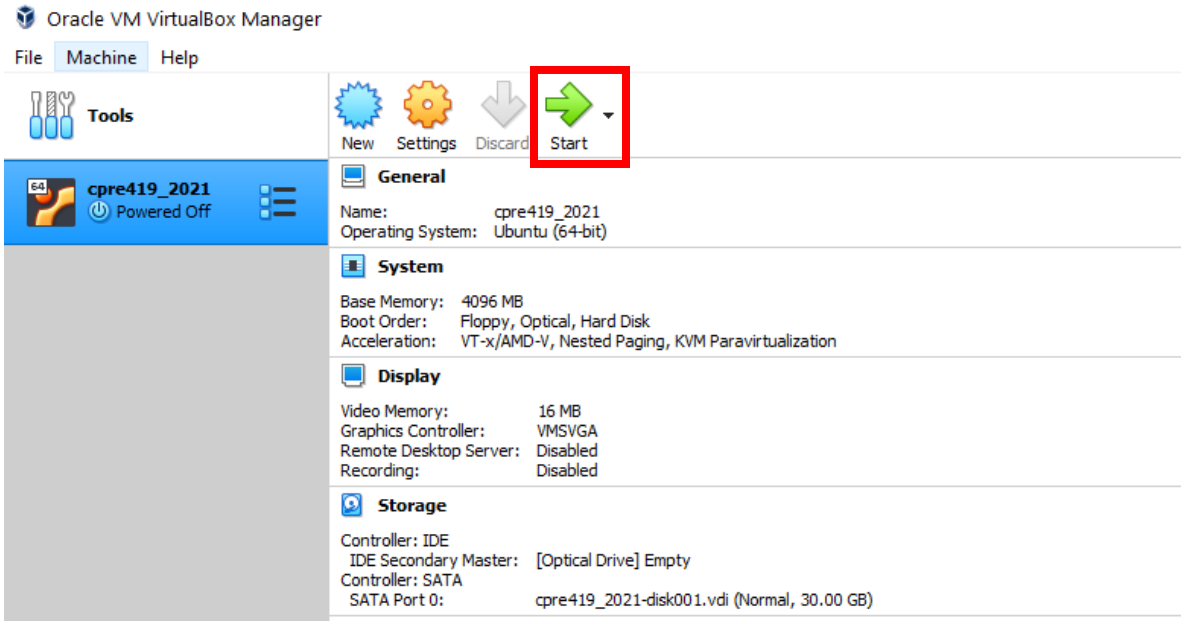
4) Keep "Source" as "Local File System".
"File": Find/Browse where your .ova is and select it.



Then next page will be some information about “cpre419_2021”, and “Import”.

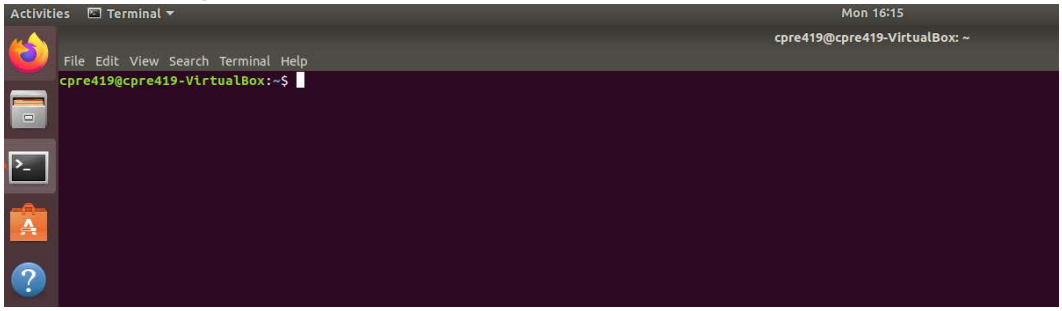


2. **After Import process has been completed, there will be a "cpre419_2021" on the left in Oracle VirtualBox. Start the VM instance**

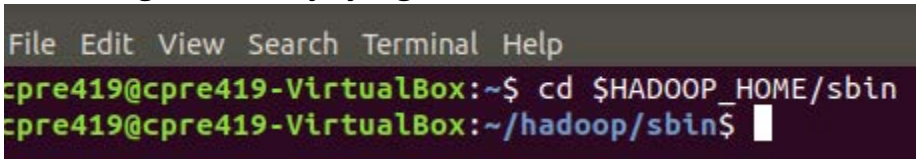


3. **Wait couple minutes until VM instance has powered on. At login page, the password should be "hadoop"**
4. **Starting HDFS and Yarn**
   Before you run any Hadoop, Pig, Spark jar file, please <u>make sure HDFS and Yarn have already started!</u>

   1) After entering Ubuntu VM instance, find "Terminal" on the left side



   2) Go to target folder by tying "cd $HADOOP_HOME/sbin"



3) Start HDFS and Yarn by typing "./start-all.sh"

```
File  Edit  View  Search  Terminal  Help
cpre419@cpre419-VirtualBox:~$ cd $HADOOP_HOME/sbin
cpre419@cpre419-VirtualBox:~/hadoop/sbin$ ./start-all.sh
```

Waiting…… Until the starting process is done

```
cpre419@cpre419-VirtualBox:~/hadoop/sbin$ ./start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as cpre419 in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [cpre419-VirtualBox]
Starting resourcemanager
Starting nodemanagers
cpre419@cpre419-VirtualBox:~/hadoop/sbin$
```

Type "jps" in terminal and check whether all that is need has started

```
cpre419@cpre419-VirtualBox:~/hadoop/sbin$ jps
3345 Jps
2850 ResourceManager
2355 DataNode
2599 SecondaryNameNode
2184 NameNode
3017 NodeManager
cpre419@cpre419-VirtualBox:~/hadoop/sbin$
```

Try a few Hadoop command lines, like

```
cpre419@cpre419-VirtualBox:~/hadoop/sbin$ hadoop fs -ls /
Found 3 items
drwxr-xr-x   - cpre419 supergroup          0 2020-03-01 22:00 /data
drwxr-xr-x   - cpre419 supergroup          0 2020-03-02 11:15 /spark-logs
drwxr-xr-x   - cpre419 supergroup          0 2020-02-29 23:23 /user
cpre419@cpre419-VirtualBox:~/hadoop/sbin$
```

Stop HDFS and Yarn

Before you shut down the machine, <u>make sure to stop HDFS and Yarn first</u>

"cd $HADOOP_HOME/sbin" and stop all by "./stop-all.sh"

```
File  Edit  View  Search  Terminal  Help
cpre419@cpre419-VirtualBox:~/hadoop/sbin$ cd $HADOOP_HOME/sbin
cpre419@cpre419-VirtualBox:~/hadoop/sbin$ ./stop-all.sh
```

```
cpre419@cpre419-VirtualBox:~/hadoop/sbin$ cd $HADOOP_HOME/sbin
cpre419@cpre419-VirtualBox:~/hadoop/sbin$ ./stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as cpre419 in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [cpre419-VirtualBox]
Stopping nodemanagers
Stopping resourcemanager
cpre419@cpre419-VirtualBox:~/hadoop/sbin$
```

```
cpre419@cpre419-VirtualBox:~/hadoop/sbin$ jps
5349 Jps
cpre419@cpre419-VirtualBox:~/hadoop/sbin$
```

<u>NOTE: we pre-installed Eclipse Java in Virtual Machine, which you can find under /home/eclipse/java-2019-12/eclipse/. Thus, you can write your code directly in the Virtual Machine</u>