

Assignment 12.1

Problem Statement :-

Follow the below link document steps to download and import Acadgild Spark VM in the Oracle Virtual Box.

<https://drive.google.com/file/d/0ByJLBTmJojzQ2hLc1RLX3pST0U/view>

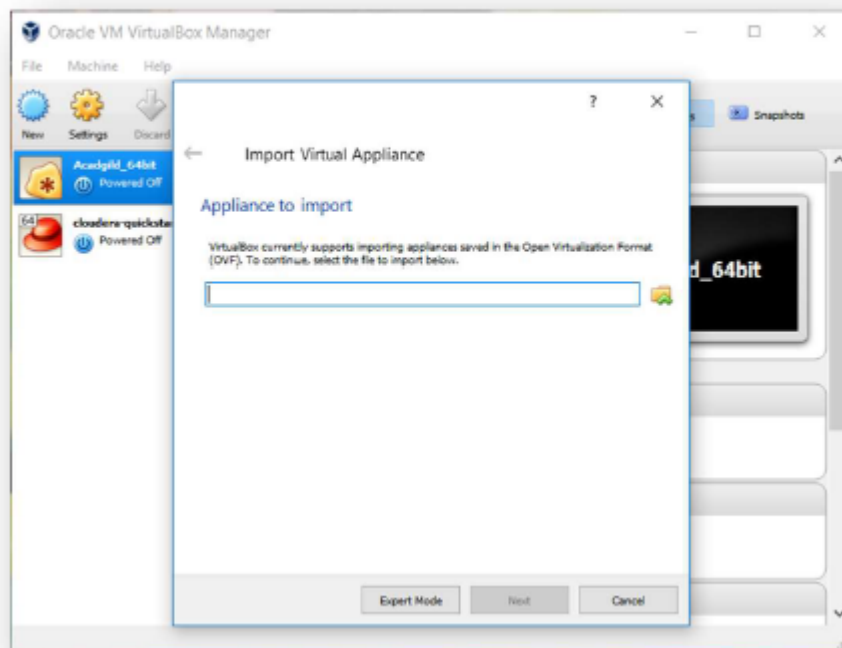
Solution:-

1. Install Acadgild-Spark VM from the link

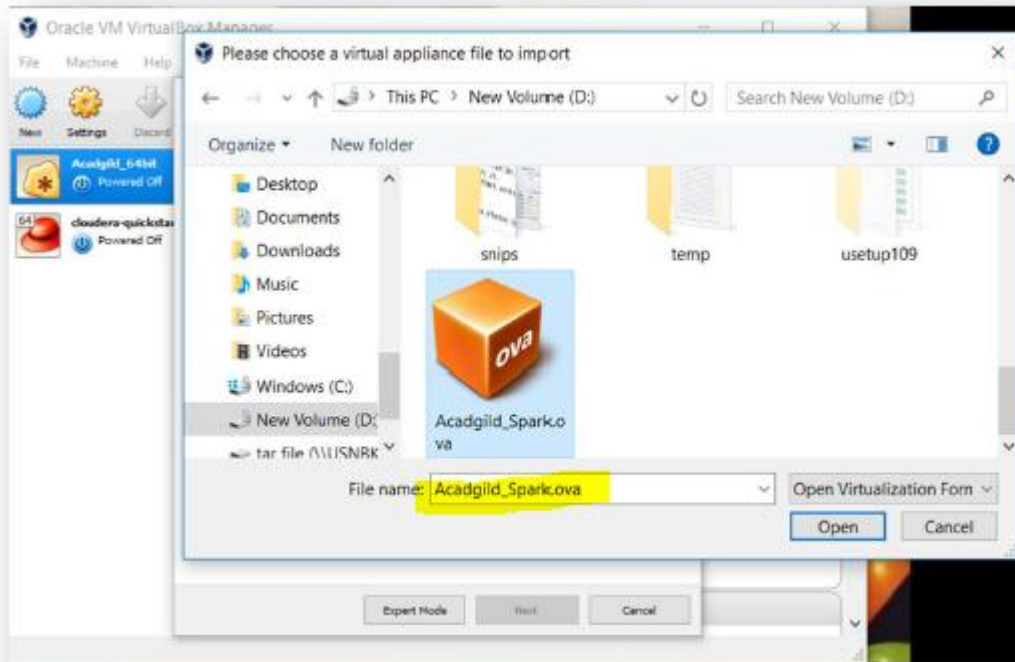
<https://drive.google.com/open?id=0ByJLBTmJojzY0pvVW9YUIFwLUk>

2. Now we shall import the Acadgild-Spark VM into Oracle VM.

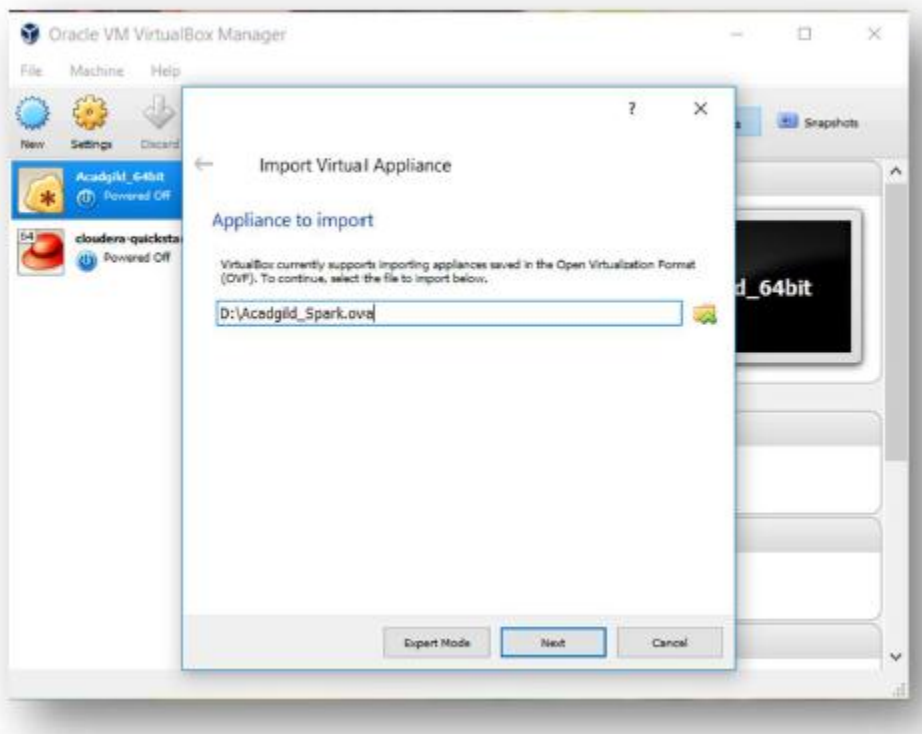
Click on File -> Import Appliance



3. Browse to the location where the Acadgild-Spark VM is installed

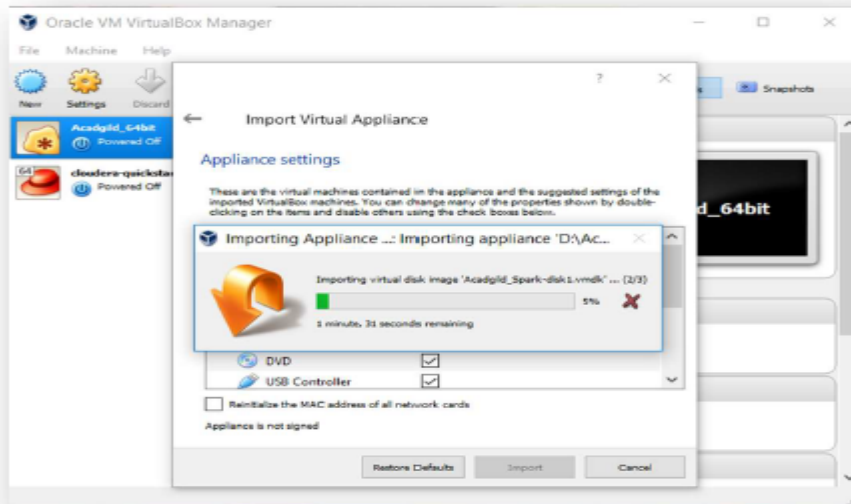


4. Then click on Next

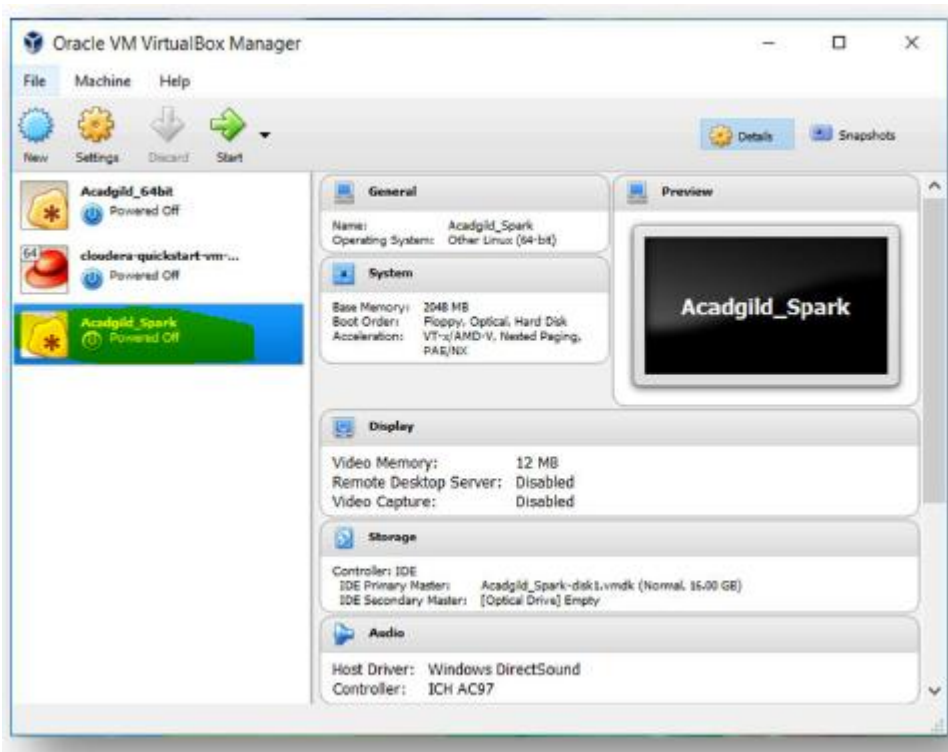


5. Then, the Acadgild-Spark VM will start installing.

Then, the Acadgild-Spark VM will start installing.



After the installation, the VM will show on the left pane of Oracle VM. Then Click on start to run the Acadgild-Spark VM



Login Username: Acadgild

Login Password: Acadgild



Open a command prompt and start the Hadoop daemons by using the command start-all.sh. Following Hadoop daemons get started

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ start-all.sh  
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh  
17/12/05 23:04:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
Starting namenodes on [localhost]  
localhost: namenode running as process 2703. Stop it first.  
localhost: datanode running as process 2805. Stop it first.  
Starting secondary namenodes [0.0.0.0]  
0.0.0.0: secondarynamenode running as process 2996. Stop it first.  
17/12/05 23:04:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
starting yarn daemons  
resourcemanager running as process 3140. Stop it first.  
localhost: nodemanager running as process 3243. Stop it first.  
[acadgild@localhost ~]$ jps  
2996 SecondaryNameNode  
3140 ResourceManager  
2805 DataNode  
4058 Jps  
3243 NodeManager  
2703 NameNode  
[acadgild@localhost ~]$
```

6. Now to start the spark daemons MASTER and WORKERS

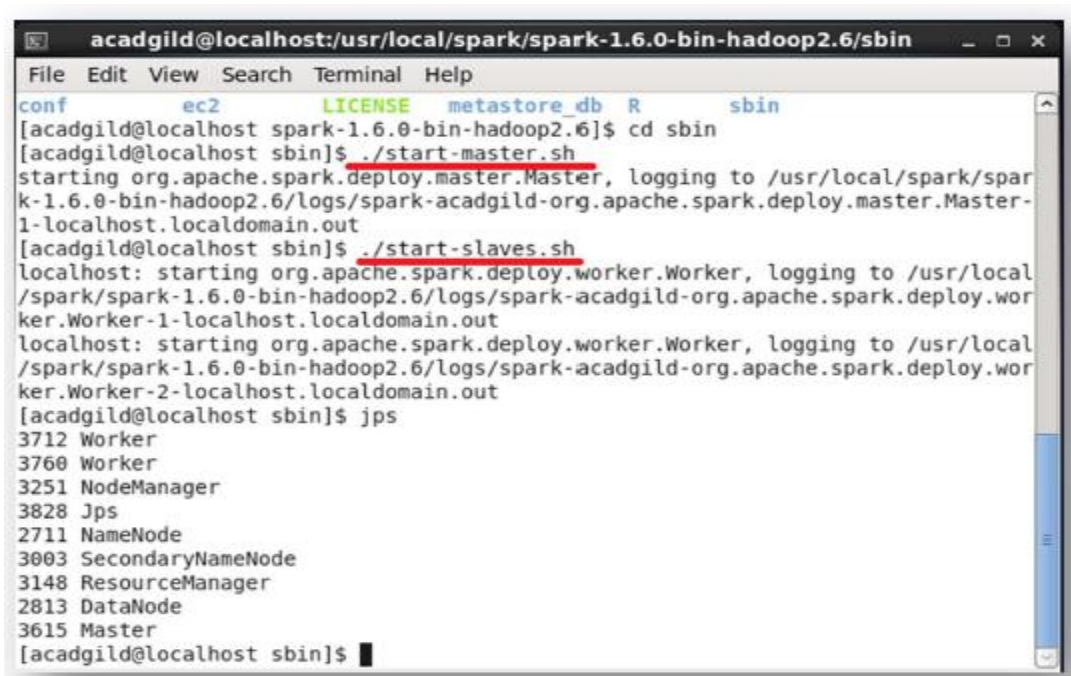
First navigate into the folder /usr/local/spark/spark-1.6.0-bin-hadoop2.6/sbin

And run the following commands

To start master: ./start-master.sh

To start worker: `./start-slaves.sh`

As you can see in the below screenshot, the worker and the master daemons have started running.



```
acadgild@localhost: /usr/local/spark/spark-1.6.0-bin-hadoop2.6/sbin
File Edit View Search Terminal Help
conf ec2 LICENSE metastore_db R sbin
[acadgild@localhost spark-1.6.0-bin-hadoop2.6]$ cd sbin
[acadgild@localhost sbin]$ ./start-master.sh
starting org.apache.spark.deploy.master.Master, logging to /usr/local/spark/spark-1.6.0-bin-hadoop2.6/logs/spark-acadgild-org.apache.spark.deploy.master.Master-1-localhost.localdomain.out
[acadgild@localhost sbin]$ ./start-slaves.sh
localhost: starting org.apache.spark.deploy.worker.Worker, logging to /usr/local/spark/spark-1.6.0-bin-hadoop2.6/logs/spark-acadgild-org.apache.spark.deploy.worker.Worker-1-localhost.localdomain.out
localhost: starting org.apache.spark.deploy.worker.Worker, logging to /usr/local/spark/spark-1.6.0-bin-hadoop2.6/logs/spark-acadgild-org.apache.spark.deploy.worker.Worker-2-localhost.localdomain.out
[acadgild@localhost sbin]$ jps
3712 Worker
3760 Worker
3251 NodeManager
3828 Jps
2711 NameNode
3003 SecondaryNameNode
3148 ResourceManager
2813 DataNode
3615 Master
[acadgild@localhost sbin]$
```

7. To work with eclipse on Acadgild-Spark VM, click on the icon “Link to eclipse”



Submitted By:-

Hardik Kaushik