Downloading and importing Acadgild spark VM

1) First I've downloaded the acadgild spark VM from the link and imported in oracle virtual box.

2) Once imported, logged into the spark vm using acadgild.

3) Opened the terminal and type the command jps to check the list of daemons, we need to start all the daemons for the first time.

```
[acadgild@localhost ~]$ jps
2615 Jps
```

Starting Hadoop Demons:

In the VM hadoop-2.6.0 is present in /user/local directory. So navigating inside hadoop using below command.

Command: $ cd /usr/local/hadoop-2.6.0

$ cd sbin

```
[acadgild@localhost ~]$ cd /usr/local/hadoop-2.6.0
[acadgild@localhost hadoop-2.6.0]$ cd sbin
```

Once inside hadoop sbin need to start namenode, datanode, resource manager and node manager

1) Starting namenode:

Command: ./hadoop-daemon.sh start namenode

2) Starting datanode as below

Command: ./hadoop-daemon.sh start namenode

```
[acadgild@localhost sbin]$ ./hadoop-daemon.sh start namenode
starting namenode, logging to /usr/local/hadoop-2.6.0/logs/hadoop-acadgild-namen
ode-localhost.localdomain.out
[acadgild@localhost sbin]$ ./hadoop-daemon.sh start datanode
starting datanode, logging to /usr/local/hadoop-2.6.0/logs/hadoop-acadgild-datan
ode-localhost.localdomain.out
```

3) Starting resourcemanager:

Command: ./yarn-daemon.sh start resourcemanager

4) Starting nodemanager:

Command: ./yarn-daemon.sh start resourcemanager

```
[acadgild@localhost sbin]$ ./yarn-daemon.sh start resourcemanager
starting resourcemanager, logging to /usr/local/hadoop-2.6.0/logs/yarn-acadgild-
resourcemanager-localhost.localdomain.out
[acadgild@localhost sbin]$ ./yarn-daemon.sh start nodemanager
starting nodemanager, logging to /usr/local/hadoop-2.6.0/logs/yarn-acadgild-node
manager-localhost.localdomain.out
```

5) Once done, checked the health of the daemons, if all the daemons will be running then cluster is ready.

```
[acadgild@localhost ~]$ jps
3252 Jps
2661 NameNode
3097 NodeManager
2747 DataNode
2846 ResourceManager
```

Starting Spark Demons:

First located the spark location using the below command.

```
[acadgild@localhost ~]$ which spark
/usr/bin/which: no spark in (/usr/local/spark/spark-1.6.0-bin-hadoop2.6/bin:/usr/local/scala/scala-2.10.4/bin:/usr/local/spar
k/spark-1.6.0-bin-hadoop2.6/bin:/usr/local/scala/scala-2.10.4/bin:/usr/lib64/qt-3.3/bin:/usr/local/bin:/usr/bin:/bin:/usr/loc
al/sbin:/usr/sbin:/sbin:/usr/local/java/bin:/usr/local/hadoop-2.6.0/bin:/usr/local/hadoop-2.6.0/sbin:/home/acadgild/bin:/usr/
local/java/bin:/usr/local/hadoop-2.6.0/bin:/usr/local/hadoop-2.6.0/sbin)
```

In the VM spark-1.6.0 is present in /user/local directory. So navigating inside spark using below command.

Command: $ cd /usr/local/spark/spark-1.6.0-bin-hadoop2.6

$ cd sbin

```
[acadgild@localhost ~]$ cd /usr
[acadgild@localhost usr]$ cd local
[acadgild@localhost local]$ cd spark
[acadgild@localhost spark]$ ls
spark-1.6.0-bin-hadoop2.6

[acadgild@localhost spark]$ cd spark-1.6.0-bin-hadoop2.6
[acadgild@localhost spark-1.6.0-bin-hadoop2.6]$ cd sbin
```

Once inside spark sbin need to start master and slave using below commands

1) Starting Master:

Command: ./ start-master.sh

```
[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
```

2) Starting Slaves:

Command: ./start-slaves.sh

```
[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-a
cadgild-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-a
cadgild-org.apache.spark.deploy.worker.Worker-2-localhost.localdomain.out
```

3) Once done, checked the health of the daemons, if the master and slave are running

```
[acadgild@localhost sbin]$ jps
2661 NameNode
3288 Master
3097 NodeManager
2747 DataNode
3563 Worker
3517 Worker
2846 ResourceManager
3614 Jps
[acadgild@localhost sbin]$ █
```

All the Hadoop and Spark daemons are Running Successfully.