ASSIGNMENT 12.1

Problem Statement:

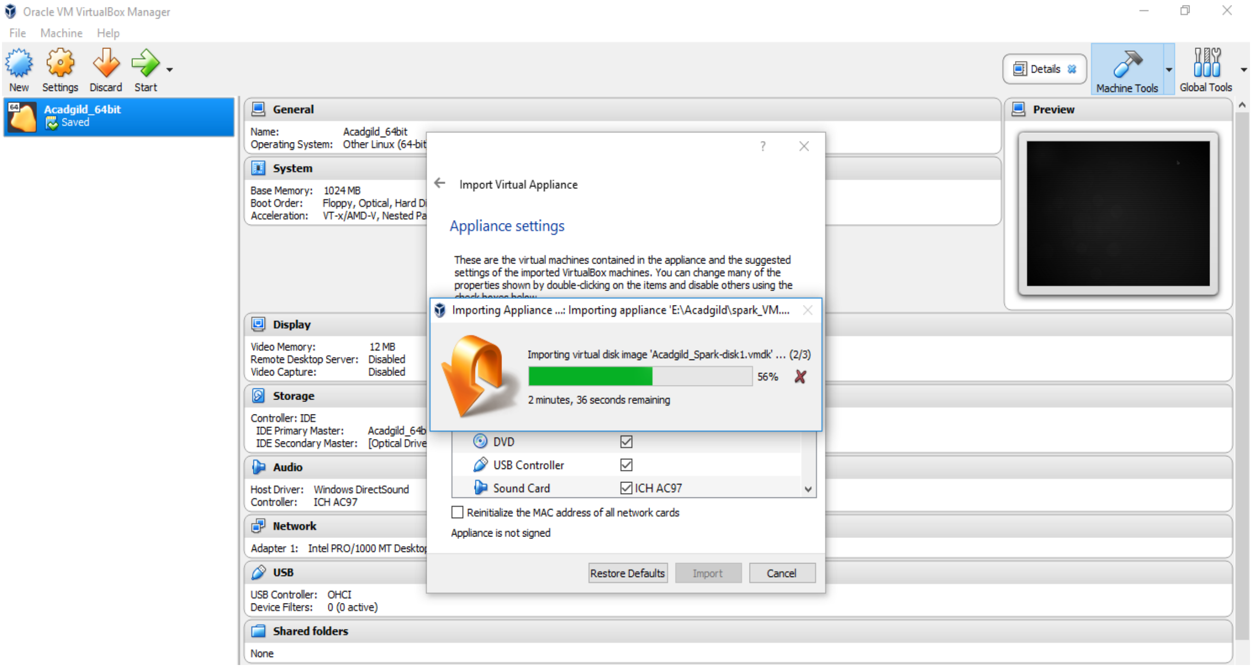
Follow the below link document steps to download and import Acadgild Spark VM in the Oracle Virtual Box.

https://drive.google.com/file/d/0ByJLBTmJojjzQ2hLc1RLX3pST0U/view

Solution:

Step 1: Download Acadgile_spark.ova file from given link and Import Acagild Spark VM into Oracle Virtual Box

We can import .ova template file by choosing 'Import Appliance' option in File menu of Oracle Virtual Box and follow the installation procedure by clicking 'Next' button. It will take some time to import virtual machine from given template and setup the virtual machine environment.



Name: Acadgild_64bit
Operating System: Other Linux (64-bit

Base Memory: 1024 MB
Boot Order: Floppy, Optical, Hard Di
Acceleration: VT-x/AMD-V, Nested Pa

Video Memory: 12 MB
Remote Desktop Server: Disabled
Video Capture: Disabled

Controller: IDE
IDE Primary Master: Acadgild_64b
IDE Secondary Master: [Optical Drive

Host Driver: Windows DirectSound
Controller: ICH AC97

Adapter 1: Intel PRO/1000 MT Deskto

USB Controller: OHCI
Device Filters: 0 (0 active)

None

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.

Importing Appliance ...: Importing appliance 'E:\Acadgild\spark_VM....

Importing virtual disk image 'Acadgild_Spark-disk1.vmdk' ... (2/3)

56%

2 minutes, 36 seconds remaining

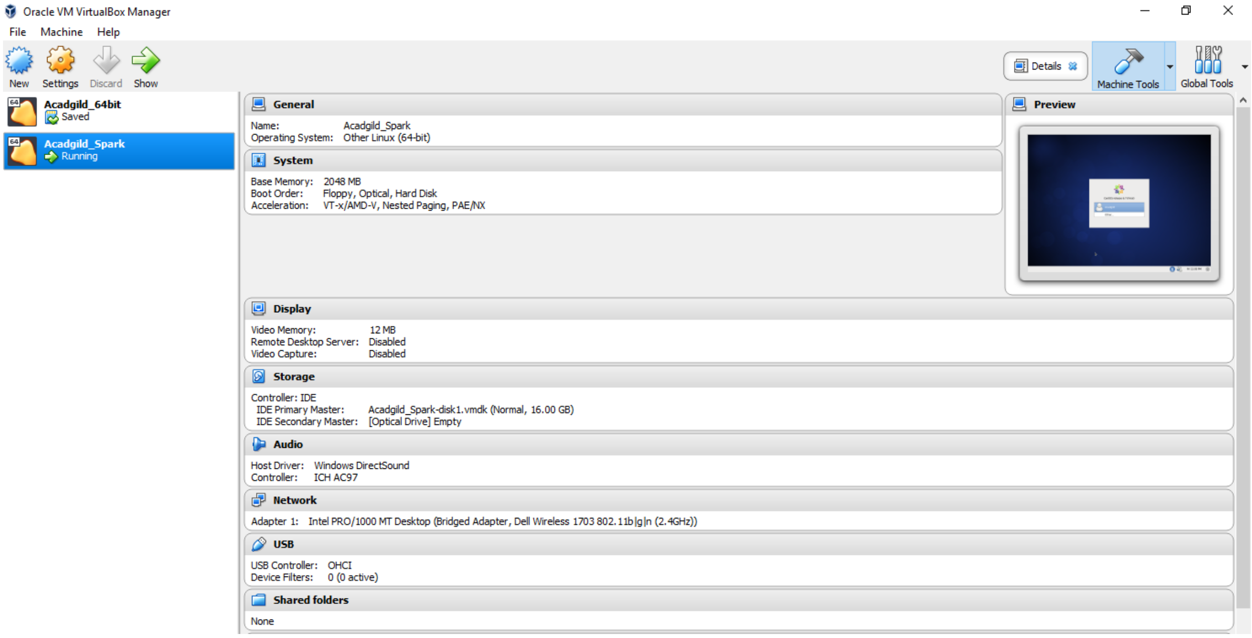
DVD ☑
USB Controller ☑
Sound Card ☑ ICH AC97

□ Reinitialize the MAC address of all network cards

Appliance is not signed

Once the installation is complete we will be able to see the virtual machine on the left hand side of Oracle Virtual Box main window.

Let's start this virtual machine and try to login to the system as user 'acadgild'.



General

Name: Acadgild_Spark
Operating System: Other Linux (64-bit)

System

Base Memory: 2048 MB
Boot Order: Floppy, Optical, Hard Disk
Acceleration: VT-x/AMD-V, Nested Paging, PAE/NX

Display

Video Memory: 12 MB
Remote Desktop Server: Disabled
Video Capture: Disabled

Storage

Controller: IDE
IDE Primary Master: Acadgild_Spark-disk1.vmdk (Normal, 16.00 GB)
IDE Secondary Master: [Optical Drive] Empty

Audio

Host Driver: Windows DirectSound
Controller: ICH AC97

Network

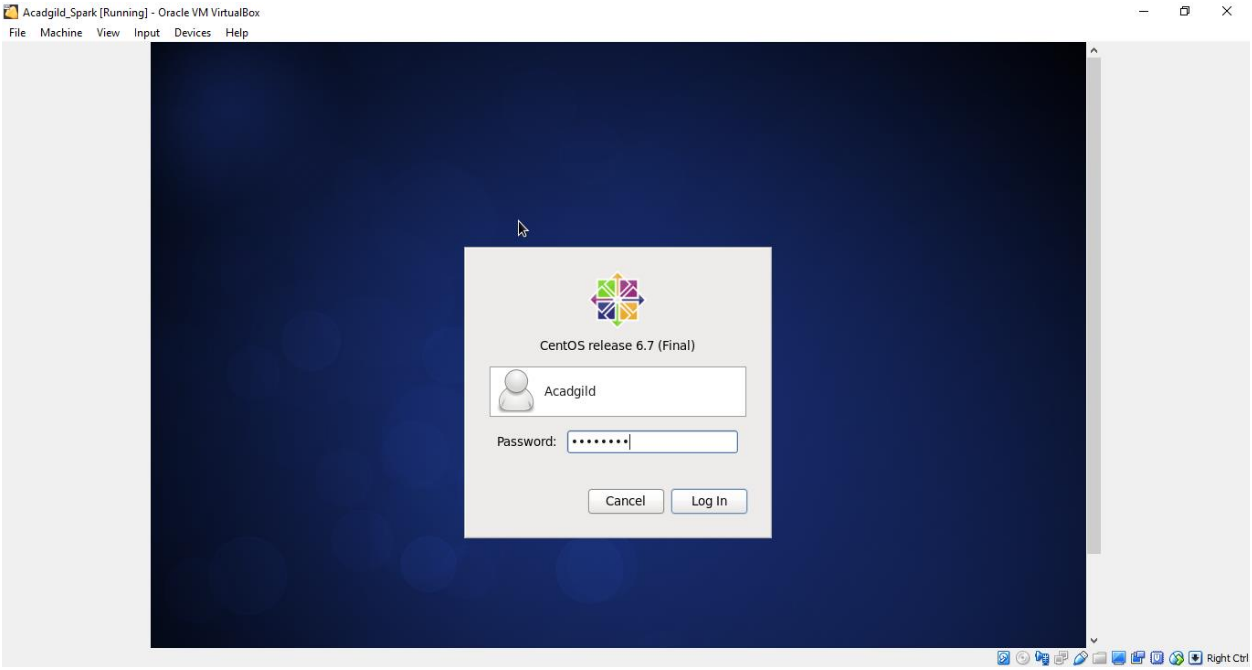
Adapter 1: Intel PRO/1000 MT Desktop (Bridged Adapter, Dell Wireless 1703 802.11b|g|n (2.4GHz))

USB

USB Controller: OHCI
Device Filters: 0 (0 active)

Shared folders

None



CentOS release 6.7 (Final)

Acadgild

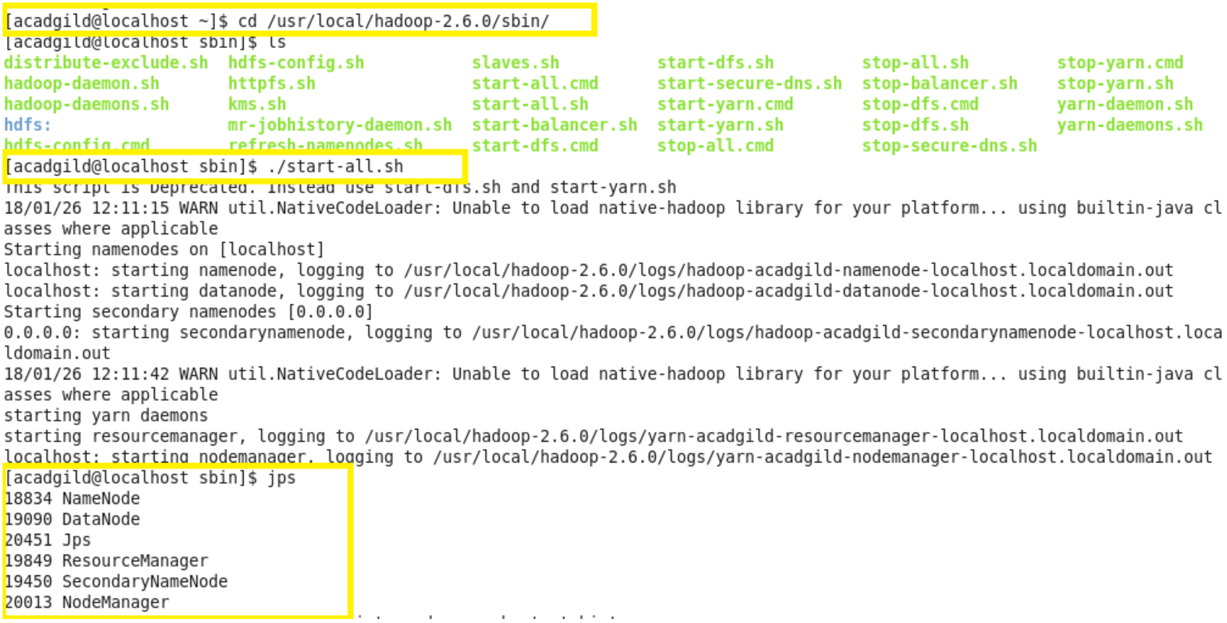
Password: ••••••••

Step 2: Start all Hadoop daemons/processes.

We need to open a terminal and move to the Hadoop home directory in the path: /usr/local/hadoop-2.6.0/sbin and then execute the shell script start-all.sh which will start all the Hadoop daemons.

```
$ cd /usr/local/hadoop-2.6.0/sbin
```

```
$ ./start-all.sh
```

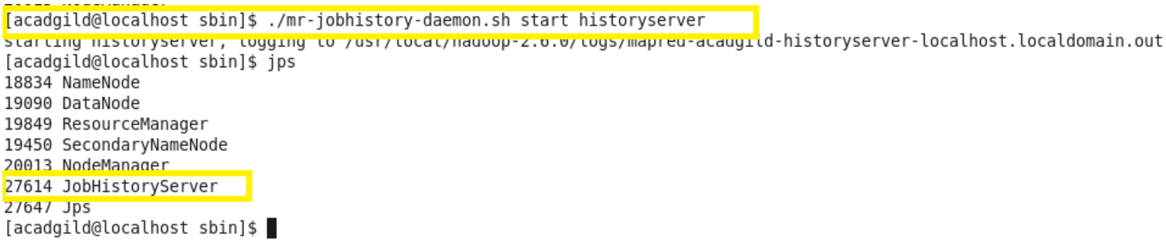


```
[acadgild@localhost ~]$ cd /usr/local/hadoop-2.6.0/sbin/
[acadgild@localhost sbin]$ ls
distribute-exclude.sh  hdfs-config.sh           slaves.sh           start-dfs.sh         stop-all.sh          stop-yarn.cmd
hadoop-daemon.sh       httpfs.sh                start-all.cmd       start-secure-dns.sh  stop-balancer.sh     stop-yarn.sh
hadoop-daemons.sh      kms.sh                   start-all.sh        start-yarn.cmd       stop-dfs.cmd         yarn-daemon.sh
hdfs:                  mr-jobhistory-daemon.sh  start-balancer.sh   start-yarn.sh        stop-dfs.sh          yarn-daemons.sh
hdfs-config.cmd        refresh-namenodes.sh     start-dfs.cmd       stop-all.cmd         stop-secure-dns.sh
[acadgild@localhost sbin]$ ./start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/01/26 12:11:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop-2.6.0/logs/hadoop-acadgild-namenode-localhost.localdomain.out
localhost: starting datanode, logging to /usr/local/hadoop-2.6.0/logs/hadoop-acadgild-datanode-localhost.localdomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop-2.6.0/logs/hadoop-acadgild-secondarynamenode-localhost.loca
ldomain.out
18/01/26 12:11:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop-2.6.0/logs/yarn-acadgild-resourcemanager-localhost.localdomain.out
localhost: starting nodemanager, logging to /usr/local/hadoop-2.6.0/logs/yarn-acadgild-nodemanager-localhost.localdomain.out
[acadgild@localhost sbin]$ jps
18834 NameNode
19090 DataNode
20451 Jps
19849 ResourceManager
19450 SecondaryNameNode
20013 NodeManager
```

We can verify this using ‘jps’ command as shown in the above screenshot.

We can also start the job history server by entering the following command:

$ ./mr-jobhistory-daemon.sh start historyserver



```
[acadgild@localhost sbin]$ ./mr-jobhistory-daemon.sh start historyserver
starting historyserver, logging to /usr/local/hadoop-2.6.0/logs/mapred-acadgild-historyserver-localhost.localdomain.out
[acadgild@localhost sbin]$ jps
18834 NameNode
19090 DataNode
19849 ResourceManager
19450 SecondaryNameNode
20013 NodeManager
27614 JobHistoryServer
27647 Jps
[acadgild@localhost sbin]$ 
```

Step 3: Start Spark related daemons

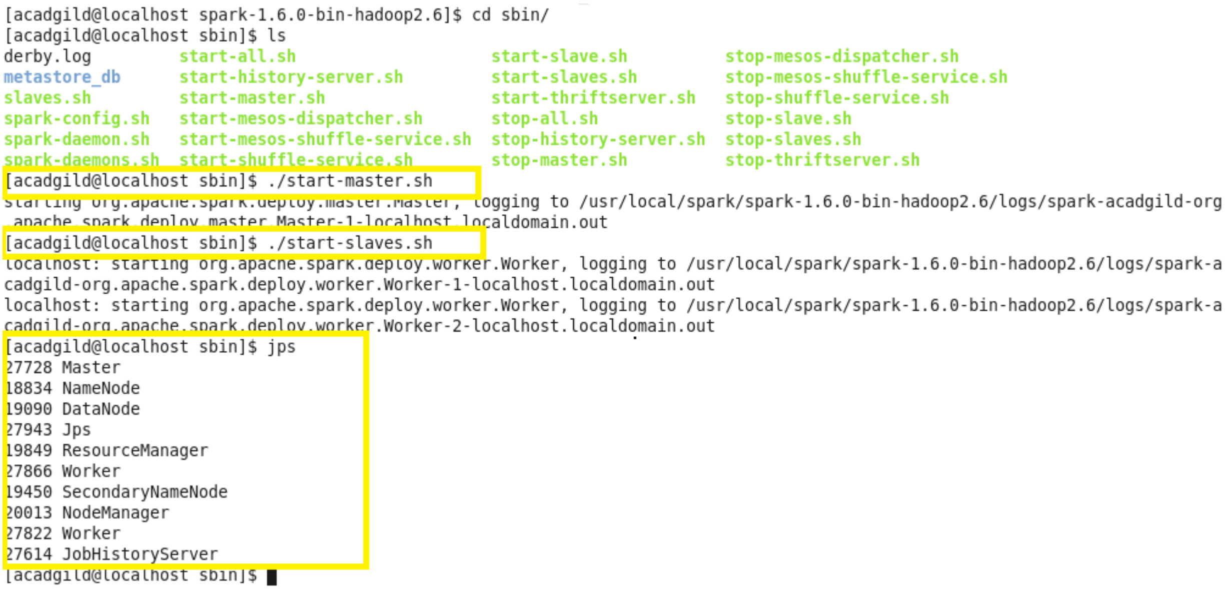
Let’s navigate to Spark home directory in the path: /usr/local/spark/spark-1.6.0-bin-hadoop-2.6/ sbin directory and execute a couple of shell scripts to start Spark Master and Slave nodes:

$ cd /usr/local/spark/spark-1.6.0-bin-hadoop-2.6/sbin

$ ./start-master.sh

$ ./start-slaves.sh

We can verify the start of Spark and Hadoop daemons by giving the command, jps as show below:

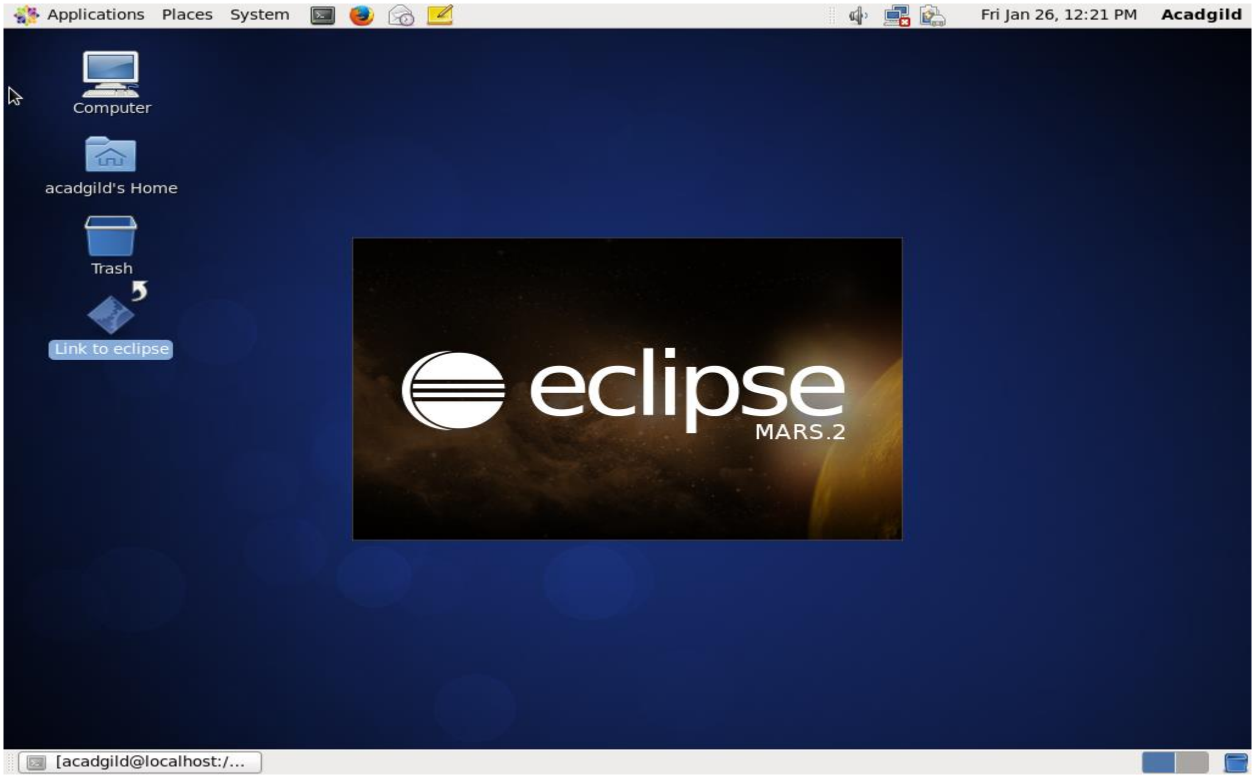


```
[acadgild@localhost spark-1.6.0-bin-hadoop2.6]$ cd sbin/
[acadgild@localhost sbin]$ ls
derby.log          start-all.sh                 start-slave.sh          stop-mesos-dispatcher.sh
metastore_db       start-history-server.sh      start-slaves.sh         stop-mesos-shuffle-service.sh
slaves.sh          start-master.sh              start-thriftserver.sh   stop-shuffle-service.sh
spark-config.sh    start-mesos-dispatcher.sh    stop-all.sh             stop-slave.sh
spark-daemon.sh    start-mesos-shuffle-service.sh stop-history-server.sh stop-slaves.sh
spark-daemons.sh   start-shuffle-service.sh     stop-master.sh          stop-thriftserver.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
[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
[acadgild@localhost sbin]$ jps
27728 Master
18834 NameNode
19090 DataNode
27943 Jps
19849 ResourceManager
27866 Worker
19450 SecondaryNameNode
20013 NodeManager
27822 Worker
27614 JobHistoryServer
[acadgild@localhost sbin]$
```

We can notice that a Spark Master and two Worker nodes are up and running at the moment.

Step 4: Start Eclipse to develop Spark applications

There is a shortcut to open 'eclipse' IDE on the desktop of the system. We can open it by double clicking on the shortcut.


Applications
Places
System
Fri Jan 26, 12:21 PM
Acadgild
Computer
acadgild's Home
Trash
Link to eclipse
eclipse
MARS.2
[acadgild@localhost:/...