

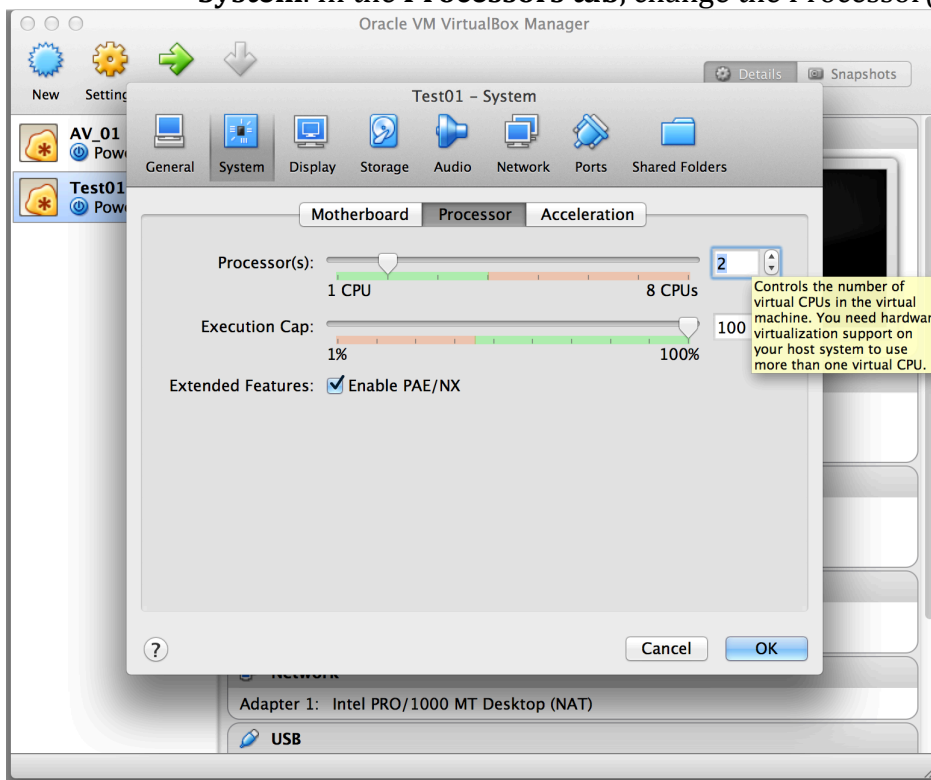
## TO BUILD YOUR OWN MAPR SANDBOX

### 1. Download and install virtualbox

<https://www.virtualbox.org/wiki/Downloads>

### 2. Build vm

1. Launch VirtualBox application
2. Click the **New** icon
3. Provide a **Name**, **Operating system type** (linux) and **Operating system version** (Redhat (64 bit))
4. Set the size to 6g ram- System Base Memory = 6 GB ~6144MB
5. Hard drive – **Create a virtual hard drive now >> Create >> VDI >> Dynamically allocated >> 16.00 GB >> Create**
6. Need to create 2 vcpus - From the VM VirtualBox Manager, click **System**. In the **Processors** tab, change the Processor(s) to 2.

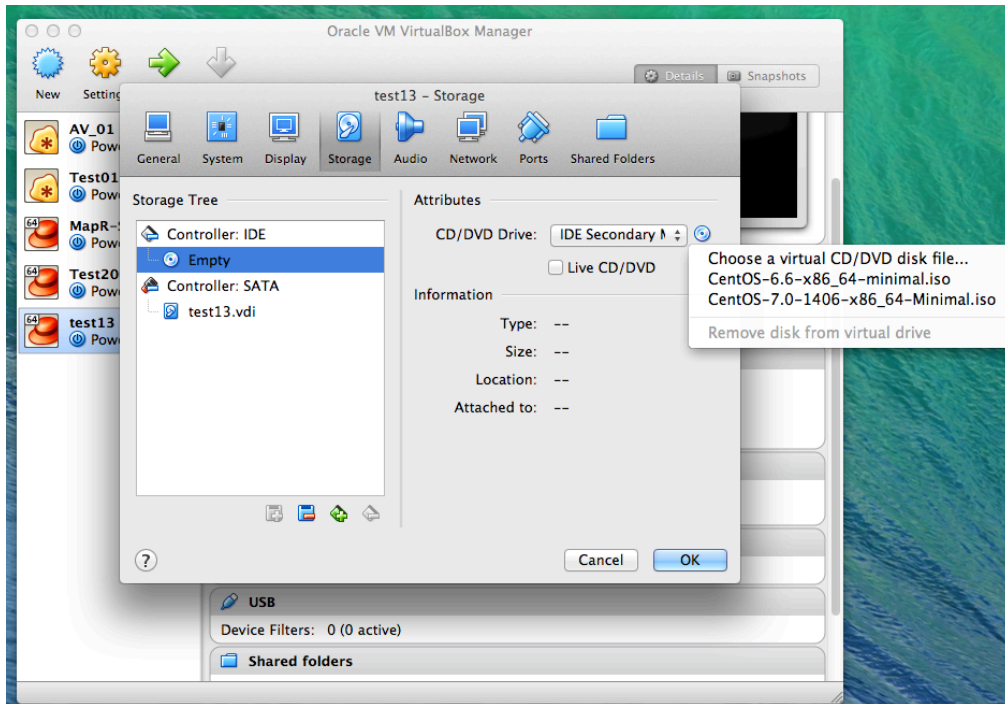


7. Click Network. Make sure NAT is selected.

### 3. Install centos 7 x86\_64 (minimal ISO)

1. Download the ISO from  
[http://mirrors.lug.mtu.edu/centos/7/isos/x86\\_64/CentOS-7-x86\\_64-Minimal-1708.iso](http://mirrors.lug.mtu.edu/centos/7/isos/x86_64/CentOS-7-x86_64-Minimal-1708.iso)

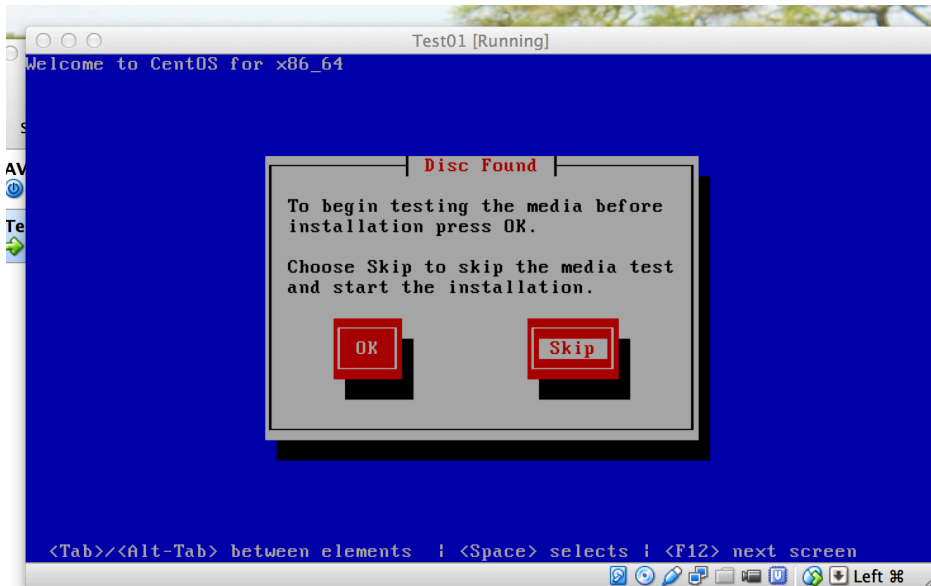
2. Attach CentOS ISO to the virtual machine -
  - a. Under Storage >>Controller:IDE, select the cdrom (empty).
  - b. Under the Attributes (to the right) click the CD icon and navigate to the location of the CentOS ISO file you downloaded.
  - c. Click OK



3. Then click Start (green arrow) button
4. Accept the default for install (install or upgrade an existing system)

*NOTE: you need to press the "command" key in MacOS or the right "control" key in Windows to get your mouse cursor out of the console window.*

5. Click Skip (not Test)



6. Choose languages >> Next

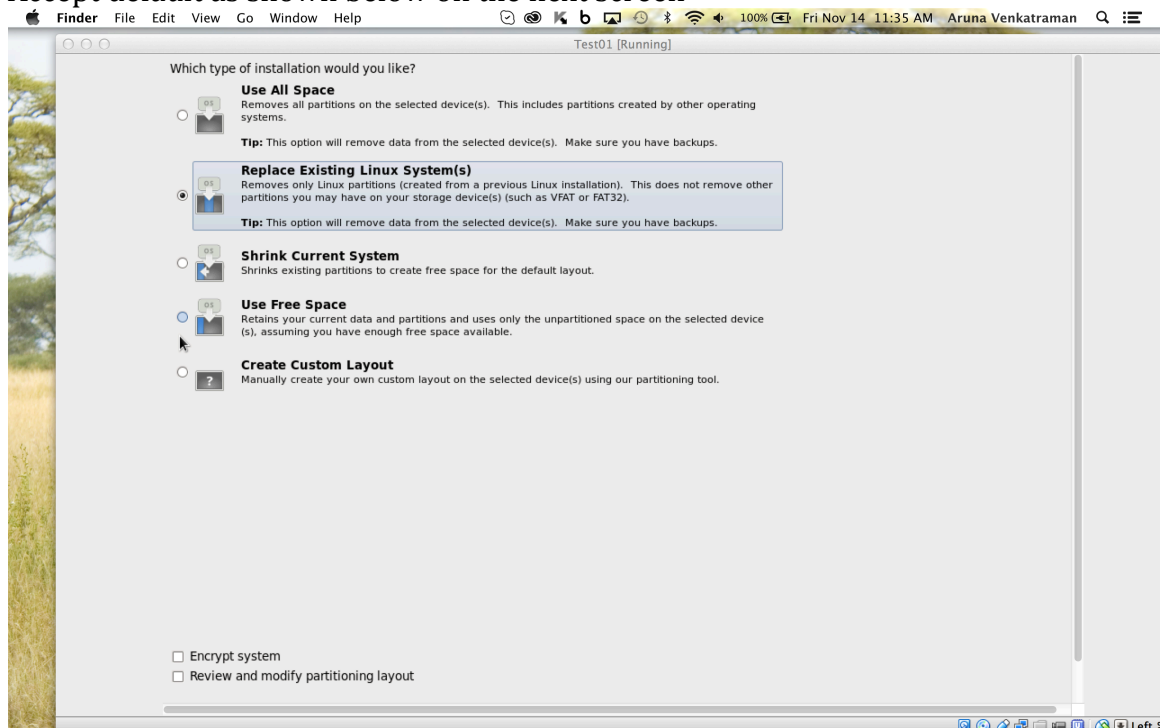
*NOTE: you may need to resize the console window in order to see all the buttons, drop-down lists, text fields, etc. The "Next" button, for example, is located at the bottom right-hand side of the installation wizard windows. Alternatively, you can change the view to "scaled mode" so you can see the entire console.*

7. **Basic Storage Devices** – Next >> Yes discard any data



- **Hostname** = cs185
  - Timezone
  - Root password (set to mapr)
- .....

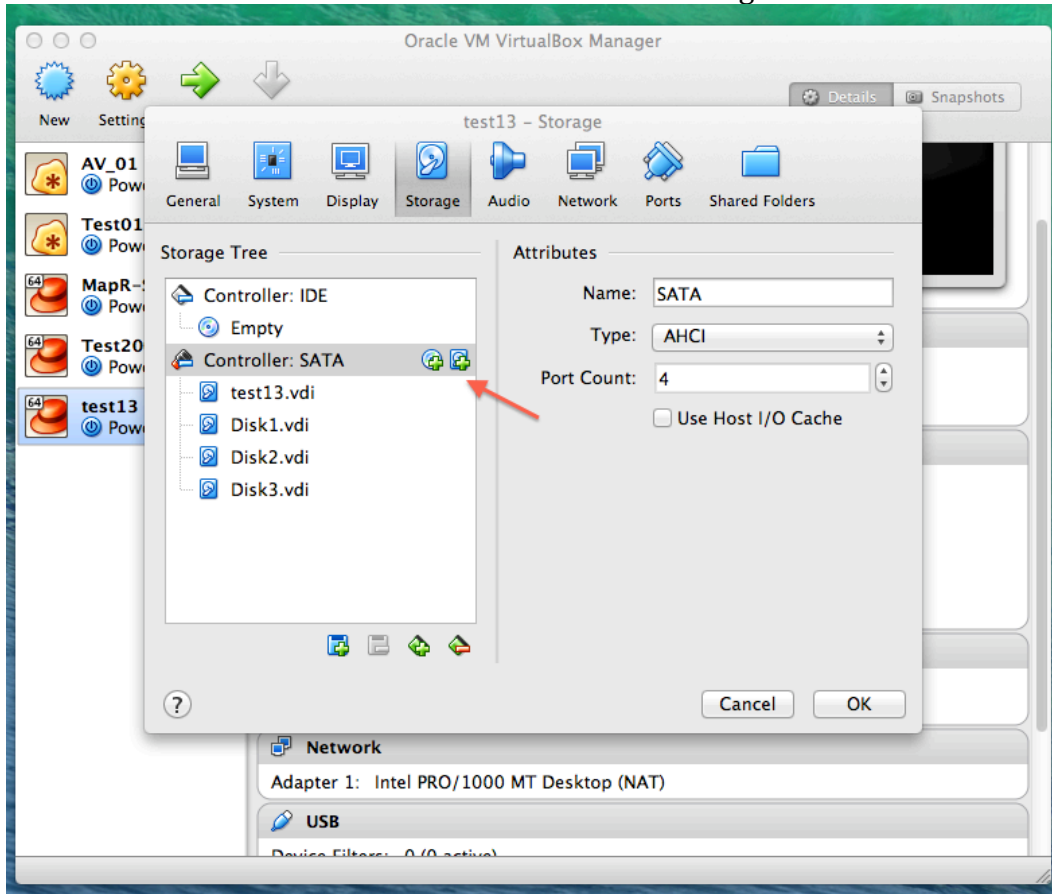
- Accept default as shown below on the next screen



From the next window- pick Write changes to disk  
This will take a little time.  
You will be asked to reboot. Go ahead and reboot.

**4. To Add additional disks**

1. Stop the VM (close the console window and select "power off the virtual machine")
2. Add 3 disks under the SATA controller in Storage

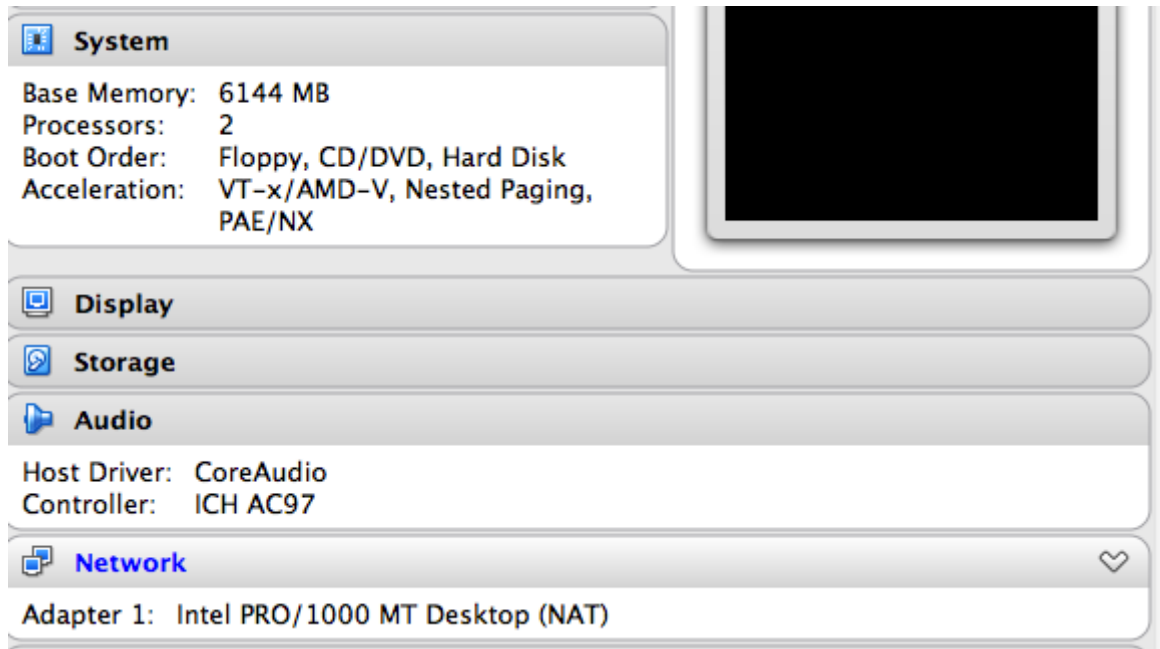


For each disk:

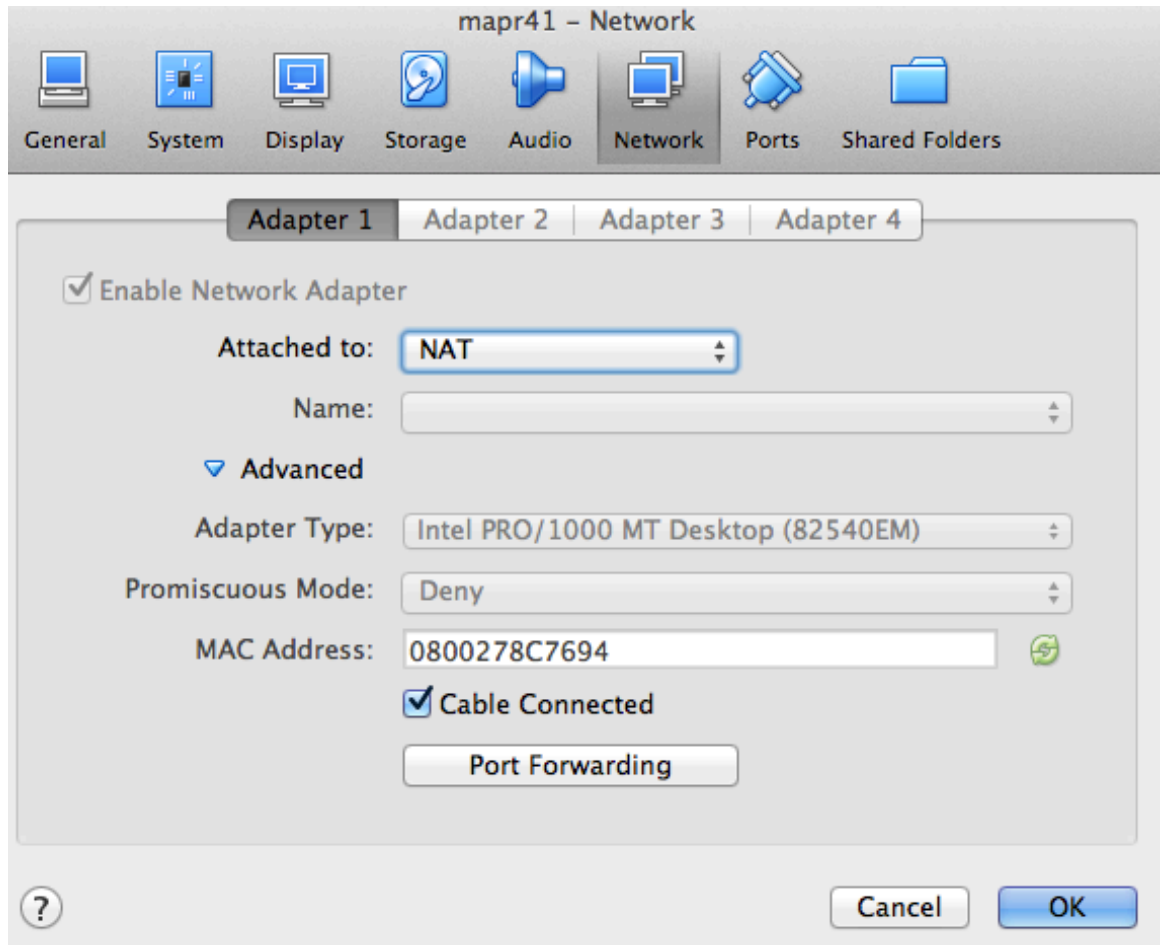
- Add Hard disk>>Create new>>VDI >>Dynamically allocated >> Name it (e.g. disk1, disk2, and disk3) and size = 8GB – Create

**5. Create port forwarding rules for NAT.**

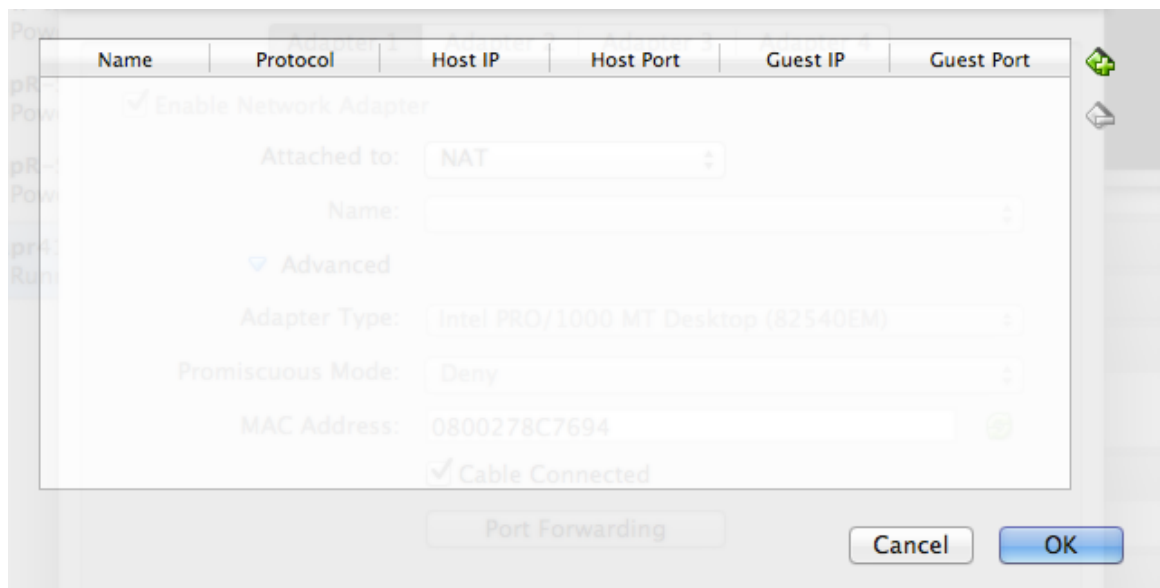
- a. In the VirtualBox Manager GUI, select your virtual machine and click the "Network" link.



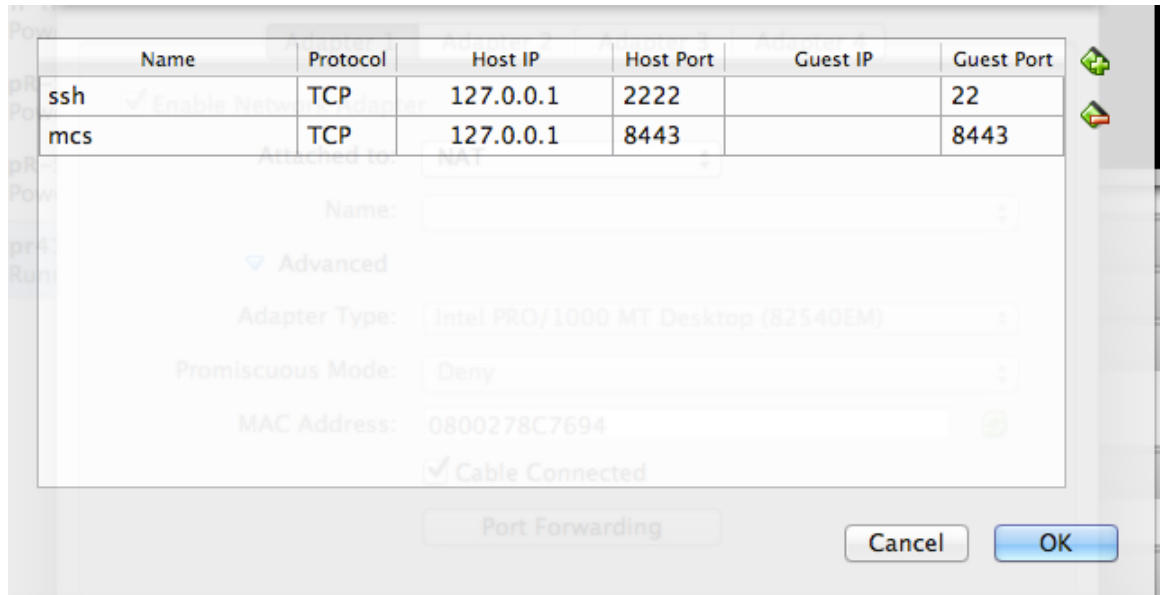
- b. In the Network wizard window, select "NAT" in the "Attached to" drop-down list and click the "Port Forwarding" button under "Advanced".



- c. In the port-forwarding pop-up window, click the icon with the "+" symbol at the top right-hand side.



- d. Create 2 port forwarding rules (one for SSH and one for the MCS)



- e. Click the "OK" buttons to save this network configuration.

6. Create a host-only network interface.
  - a. In the VirtualBox Manager GUI, select your virtual machine and click the "Network" link.
  - b. In the "Network" wizard window, select the "Adapter 2" tab.
  - c. In the "Adapter 2" tab, click the "Enable network adapter" checkbox
  - d. In the "Attached to" drop-down list, select "Host-only Adapter"
  - e. Click the "OK" button

7. Start the virtual machine and login to the console window as the `root` user.

8. Create `mapr` user and group

```
groupadd mapr
```

```
useradd -g mapr -d /home/mapr -s /bin/bash -m mapr
```

```
passwd mapr (set to mapr)
```

9. Create `/user` and `/mapr` mount points.

```
mkdir /user
```

```
mkdir /mapr
```



10. Update `/etc/sysconfig/network-scripts/ifcfg-enp0s3` as follows and bounce network service (service network restart).

(delete all the other properties from the file)

**BOOTPROTO="dhcp"**

**ONBOOT="yes"**

Restart the service –

**service network restart**

11. Disable iptables and selinux

**systemctl disable firewalld**

**systemctl stop firewalld**

**vi /etc/selinux/config** (set SELINUX=disabled)

12. Configure Linux EPEL repo

**yum install wget -y**

**cd**

**wget [http://dl.fedoraproject.org/pub/epel/7/x86\\_64/Packages/e/epel-release-7-11.noarch.rpm](http://dl.fedoraproject.org/pub/epel/7/x86_64/Packages/e/epel-release-7-11.noarch.rpm)**

**rpm -Uvh epel-release-7\*.rpm**

13. Update centos packages (this may take a while)

**yum update -y**

14. Install extra packages

**yum remove java-1.\* -y**

**yum install java-1.8.0-openjdk -y**

**yum install java-1.8.0-openjdk-devel -y**

**yum install nfs-utils -y**

**yum install zip unzip -y**

```
yum install git -y  
yum install acpid -y  
yum install net-tools -y
```

15. Create mapr repo for 6.0.0 in /etc/yum.repos.d

```
cd /etc/yum.repos.d  
  
vi mapr.repo ( add the following to the file )  
[maprtech]  
name=MapR Technologies  
baseurl=http://package.mapr.com/releases/v6.0.0/redhat  
enabled=1  
gpgcheck=0  
protect=1
```

16. Create mapr ecosystem repo

```
vi mapr-eco.repo (add the following to the file)  
  
[maprecosystem]  
name=MapR Technologies  
baseurl=http://package.mapr.com/releases/MEP/MEP-4.0.0/redhat  
enabled=1  
gpgcheck=0  
protect=1
```

17. Install MapR core software packages. This step will take a while, so make sure your laptop is plugged in and on a reliable network connection.

```
yum install mapr-core -y  
yum install mapr-fileserver -y  
yum install mapr-webserver -y  
yum install mapr-zookeeper -y  
yum install mapr-cldb -y  
yum install mapr-resourcemanager -y  
yum install mapr-nodemanager -y  
yum install mapr-historyserver -y  
yum install mapr-nfs -y
```

To validate that the above were installed, run the following command:

```
rpm -qa|grep mapr
```

18. Install the MapR ecosystem packages we'll be using during the class.

```
yum install mapr-spark -y
```

```
yum install mapr-spark-historyserver -y
```

```
yum install mapr-kafka -y
```

19. Update the `/etc/hosts` and `/opt/mapr/hostname` files.

- a. Determine the hostname (should be "cs185" if you've followed this document exactly)

```
hostname
```

- b. Edit the `/etc/hosts` file

```
vi /etc/hosts
```

```
127.0.0.1 localhost
```

```
10.0.2.15 cs185-nat
```

```
192.168.56.101 cs185
```

- c. Populate the `/opt/mapr/hostname` file

```
hostname > /opt/mapr/hostname
```

20. Run `configure.sh` to configure the MapR cluster. Replace `<hostname>` with the hostname you assigned (i.e. cs185). Replace `<clustername>` with `my.cluster.com`.

```
/opt/mapr/server/configure.sh -C <hostname>:7222 \  
-Z <hostname>:5181 -RM <hostname> -HS <hostname> \  
-N <clustername>
```

example:

```
/opt/mapr/server/configure.sh -C cs185:7222 \  
-Z cs185:5181 -RM cs185 -HS cs185 \  
-N my.cluster.com
```

21. Before you go to the next step – i.e. adding the disks, run the following to check which disks are mounted. Only those that are not mounted should be added to the file that you will create in the following step.

**lsblk --fs** ( this will give the list of all the disks – you should see sda, sdb, sdc, and sdd in your output.

**22.** Add 3 disks to MapR-FS

```
vi /tmp/disks.txt  
/dev/sdb  
/dev/sdc  
/dev/sdd
```

```
/opt/mapr/server/disksetup -F /tmp/disks.txt
```

Verify the disks were added by running the following command:

```
cat /opt/mapr/conf/disktab
```

**23.** Start mapr services

```
service mapr-zookeeper start
```

```
service mapr-warden start
```

Get a running tail of the warden log file:

```
tail -f /opt/mapr/logs/warden.log
```

Wait a few minutes for the services to start. Validate the MapR services (particularly the CLDB service) are running with the following command – look for "healthy" state. Note that the command will fail until the CLDB service is running.

```
maprcli node list
```

**24.** Add mapr user to mcs acl

```
maprcli acl edit -type cluster -user mapr:fc
```

**25.** Create /opt/mapr/conf/mapr\_fstab file for auto-mounting NFS exports at boot. Replace <clustername> appropriately. You can get cluster name from /opt/mapr/conf/mapr-clusters.conf.

```
vi /opt/mapr/conf/mapr_fstab (add the following 2 lines)  
localhost:/mapr /mapr soft,intr,nolock  
localhost:/mapr/<clustername>/user /user soft,intr,nolock
```

26. In a terminal window on your virtual machine, run the following commands as the root user:

- a. Determine the cluster id

```
cat /opt/mapr/conf/clusterid
```

- b. Determine the cluster name

```
maprcli dashboard info -json | grep name
```

27. Create a user account at <http://www.mapr.com>.

- a. Point your Web browser at <http://www.mapr.com>.

- b. Click the "Login" link at the top right-hand side of the Web page.

- c. Create a new account by following the instructions on the right side of the Web page.

**LOGIN**

Welcome to our new website. If you're an existing user, your account has been moved over, but you will have to reset your password. Please [click here](#) to request a new password.

E-mail \*

Enter your e-mail address.

Password \*

Enter the password that accompanies your e-mail.

**LOG IN**

**If you are not currently a member, please register here.**

E-mail \*

A valid e-mail address. All e-mails from the system will be sent to this address. The e-mail address is not made public and will only be used if you wish to receive a new password or wish to receive certain news or notifications by e-mail.

or

Password \*

Password strength:

Confirm password \*

Provide a password for the new account in both fields.

First name \*

Last name \*

Company \*

Leave this window open because you will need it in the next step.

- c. Back in your <http://mapr.com> browser window, add a new cluster by clicking "Register a cluster".
- d. Paste the cluster ID from your buffer into the "Cluster ID" textbox, and type the name of your cluster into the "Cluster Name" textbox. Select "M3 Community License" and click the "Register" button.

- e. Click the "view key" link for your newly registered cluster. Copy the entire contents of the key string into your "clipboard".

28. In the terminal window, create a file called `/tmp/license.txt` and paste the contents of your clipboard (i.e. the key string) into the file.

```
vi /tmp/license.txt
```

29. Run the `maprccli` command to add the license key to your cluster.

```
maprccli license add -cluster my.cluster.com \  
-license /tmp/license.txt -is_file true
```

30. Reboot vm

```
reboot
```

31. Login to your VM, give time for the cluster to start up, and validate cluster services are running (mapr password is 'mapr').

```
ssh -p 2222 mapr@localhost (use putty on windows laptops)
```

```
maprccli node list (look for "Healthy")
```

```
showmount -e (look for /mapr and /mapr/<clustername>)
```

```
df (look for /mapr and /user to be mounted)
```

```
jps -lm (look for NodeManager, JobHistoryServer, ResourceManager,  
WardenMain, and CLDB processes among others to be running)
```

32. Switch user to root

```
$ su - root
```

33. Create a temporary directory in HDFS for Spark.

```
hadoop fs -mkdir -p /apps/spark
```

```
hadoop fs -chmod 1777 /apps/spark
```

34. Enable log aggregation for YARN history server. Replace *<hostname>* appropriately.

- a. Edit the yarn-site.xml file

```
vi /opt/mapr/hadoop/hadoop-2.7.0/etc/hadoop/yarn-
```

**site.xml** (add the following lines inside the configuration)

```
<property>
  <name>yarn.log-aggregation-enable</name>
  <value>true</value>
</property>
```

b. Restart the node manager service

```
maprcli node services -action restart -name \
nodemanager -nodes <hostname>
```

- 35.** Install and configure tmpwatch to keep your /tmp directory uncluttered and prevent the boot disk root file system from filling up.

```
yum install tmpwatch -y
```

```
/usr/sbin/tmpwatch -am 1d /tmp
```

```
crontab -e (command will open vi session – put the following line inside)
0 0 * * * /usr/sbin/tmpwatch -am 1d /tmp
```

```
crontab -l (to verify you edited the crontab properly)
```

- 36.** Create an end user

```
groupadd group01
```

```
useradd -g group01 -d /home/user01 -s /bin/bash \
-m user01
```

```
passwd user01 (set to mapr)
```

```
mkdir /user/user01
```

```
chown user01:group01 /user/user01
```

37. Login as (or switch user to) end user user01, configure shell environment, and test the wordcount mapreduce application that's bundled with Hadoop.

```
su - user01
```

```
yarn jar /opt/mapr/hadoop/hadoop-  
2.7.0/share/hadoop/mapreduce/hadoop-mapreduce-  
examples-*.jar wordcount \  
file:///etc/passwd /tmp/out-$USER
```

```
hadoop fs -cat /tmp/out-$USER/part-r-00000
```

38. Test Spark by running the Spark Pi application.

```
cd /opt/mapr/spark/spark-2.1.0
```

```
./bin/spark-submit \  
--class org.apache.spark.examples.SparkPi \  
--master yarn --deploy-mode client --num-executors 1 \  
--driver-memory 512m --executor-memory 512m \  
--executor-cores 1 examples/jars/spark-examples*.jar 10
```

39. Change minimum replication factor of volumes to 1 (since there's only 1 node in this cluster).

```
su - root
```

```
for volume in $(maprcli volume list -columns \  
volumename | sed -n '2,$ p')  
> do  
> maprcli volume modify -name $volume -replication 1 \  
-minreplication 1  
> done
```

40. Update `warden.conf` to minimize memory utilization by warden services.  
d. Turn the warden service off.

```
service mapr-warden stop
```

- e. Make a backup copy of the `warden.conf` file.

```
cd /opt/mapr/conf
```

```
cp warden.conf warden.conf-orig
```



- f. Update the following lines in the `warden.conf` file (and leave all other lines as is).

```
service.command.cldb.heapsize.percent=8
service.command.cldb.heapsize.max=256
service.command.cldb.heapsize.min=256
service.command.mfs.heapsize.percent=15
service.command.mfs.heapsize.maxpercent=30
service.command.mfs.heapsize.min=512
service.command.webserver.heapsize.percent=3
service.command.webserver.heapsize.max=128
service.command.webserver.heapsize.min=128
service.command.nfs.heapsize.percent=3
service.command.nfs.heapsize.min=64
service.command.nfs.heapsize.max=64
service.command.os.heapsize.percent=10
service.command.os.heapsize.max=3000
service.command.os.heapsize.min=256
service.command.warden.heapsize.percent=1
service.command.warden.heapsize.max=256
service.command.warden.heapsize.min=64
service.command.zk.heapsize.percent=1
service.command.zk.heapsize.max=256
service.command.zk.heapsize.min=128
```

- g. Turn the warden service on.

```
service mapr-warden start
```