2/13/16 4:48 AM

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.
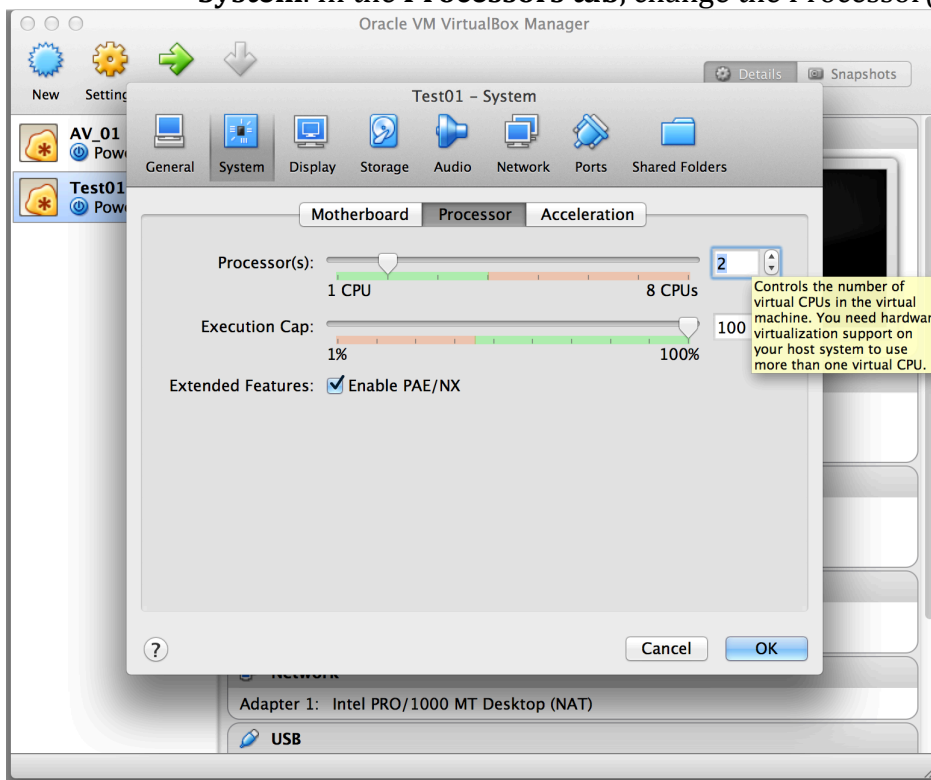

   7. Click Network. Make sure NAT is selected.

3. **Install centos 6.7 x64 (minimal ISO)**
   1. Download the ISO from
      http://isoredirect.centos.org/centos/6/isos/x86_64/

   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 6.7 ISO file you downloaded.
      c. Click OK
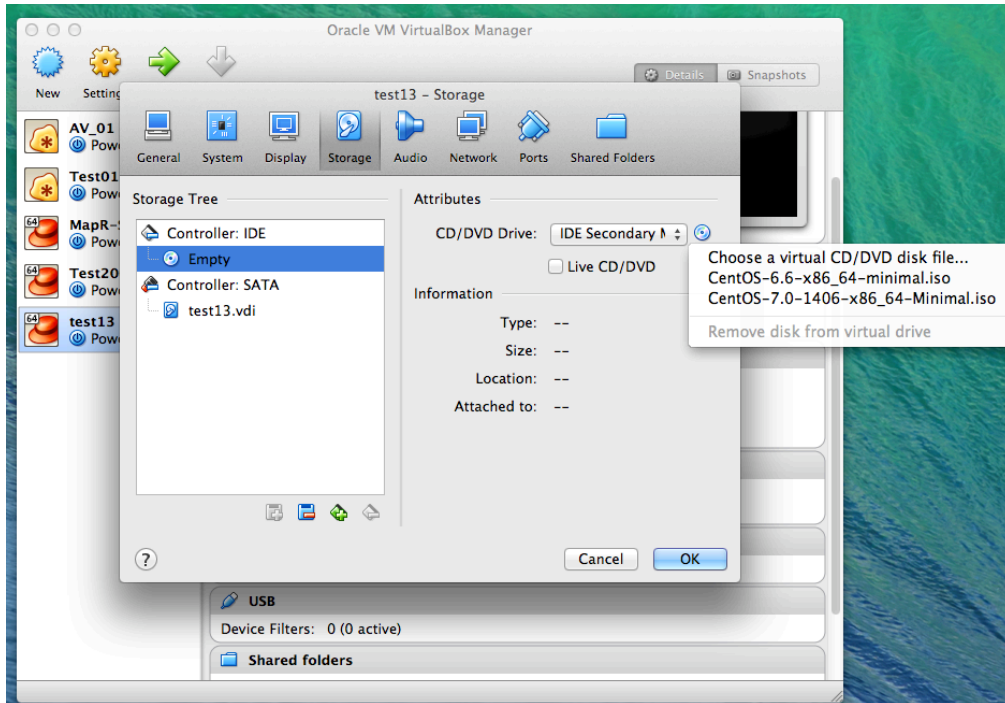
3. Then click Start (green arrow) button
4. Accept the default for install (install or upgrade an existing system)

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)
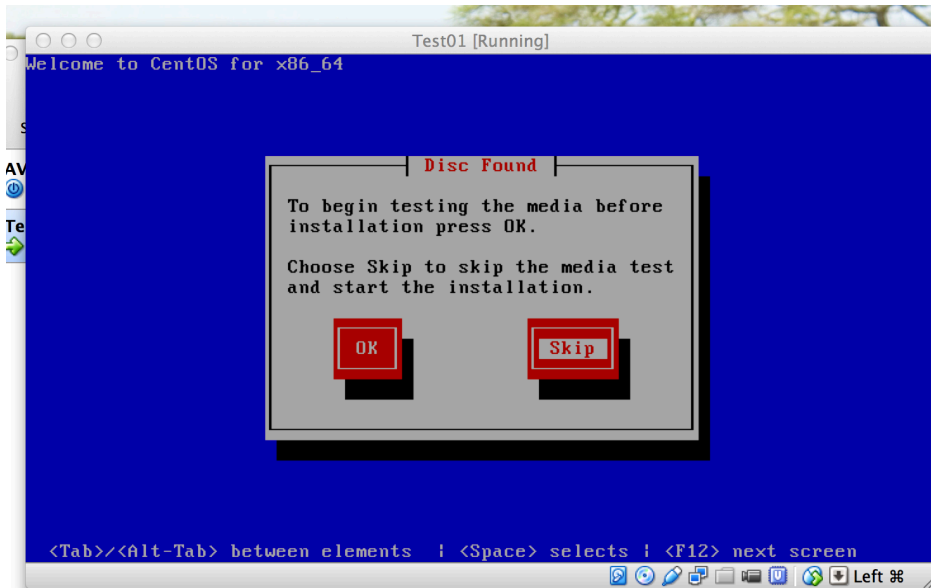
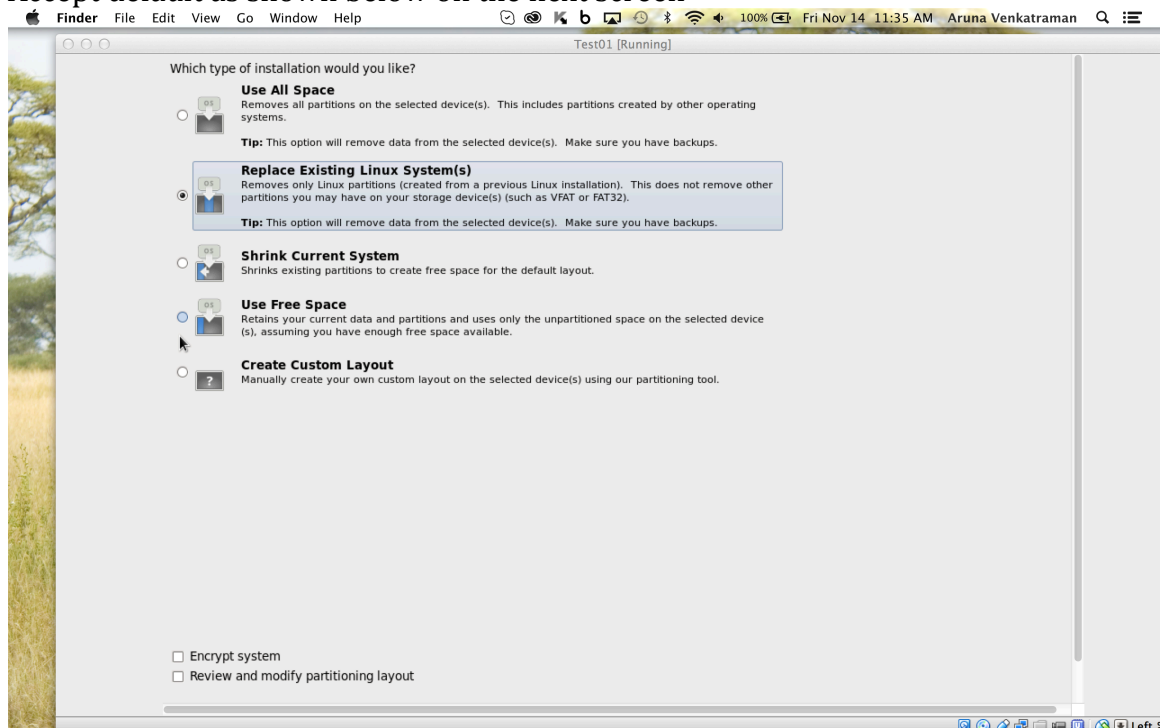6. Choose languages >> Next

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

Storage Device Warning

⚠️ **The storage device below may contain data.**

🖴 **ATA VBOX HARDDISK**
16384.0 MB   pci-0000:00:01.1-scsi-0:0:0:0

We could not detect partitions or filesystems on this device.

This could be because the device is **blank**, **unpartitioned**, or **virtual**. If
not, there may be data on the device that can not be recovered if you use it
in this installation. We can remove the device from this installation to
protect the data.

Are you sure this device does not contain valuable data?

☑ Apply my choice to all devices with undetected partitions or filesystems

[Yes, discard any data]   [No, keep any data]

- **Hostname** = mapr1node (or whatever you wish – this is used later)
- Timezone
- Root password (set to !mapr123!)

- Accept default as shown below on the next screen

Which type of installation would you like?

Use All Space
Removes all partitions on the selected device(s). This includes partitions created by other operating systems.
Tip: This option will remove data from the selected device(s). Make sure you have backups.

Replace Existing Linux System(s)
Removes only Linux partitions (created from a previous Linux installation). This does not remove other partitions you may have on your storage device(s) (such as VFAT or FAT32).
Tip: This option will remove data from the selected device(s). Make sure you have backups.

Shrink Current System
Shrinks existing partitions to create free space for the default layout.

Use Free Space
Retains your current data and partitions and uses only the unpartitioned space on the selected device(s), assuming you have enough free space available.

Create Custom Layout
Manually create your own custom layout on the selected device(s) using our partitioning tool.

☐ Encrypt system
☐ Review and modify partitioning layout

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")
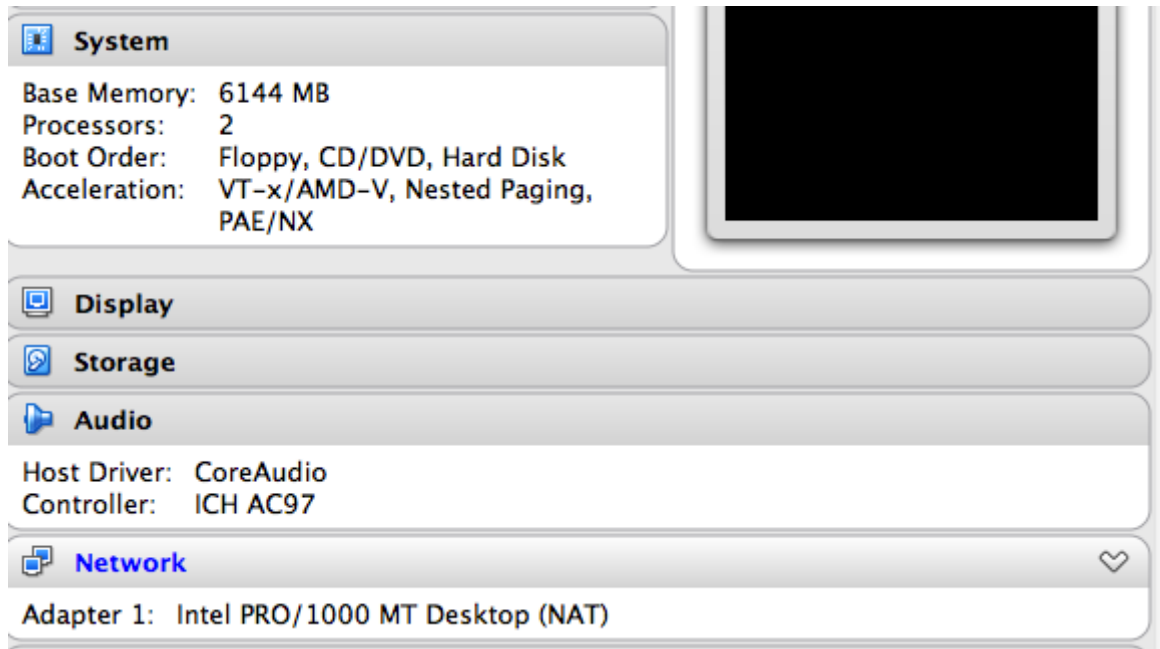   2. Add 3 disks under the SATA controller in Storage
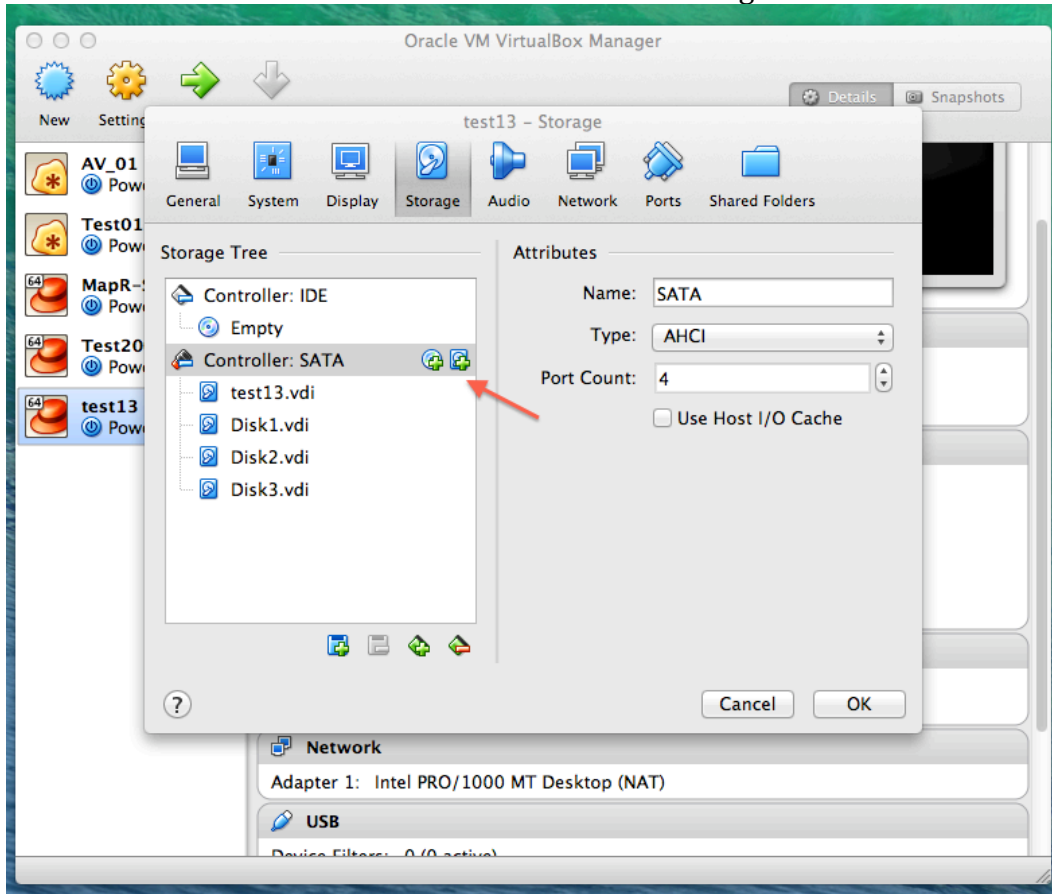
For each disk:
-   Add Hard disk>>Create new>>VDI >>Dynamically allocated >> Name it (e.g. disk1, disk2, and disk3) and size = 8GB – Create
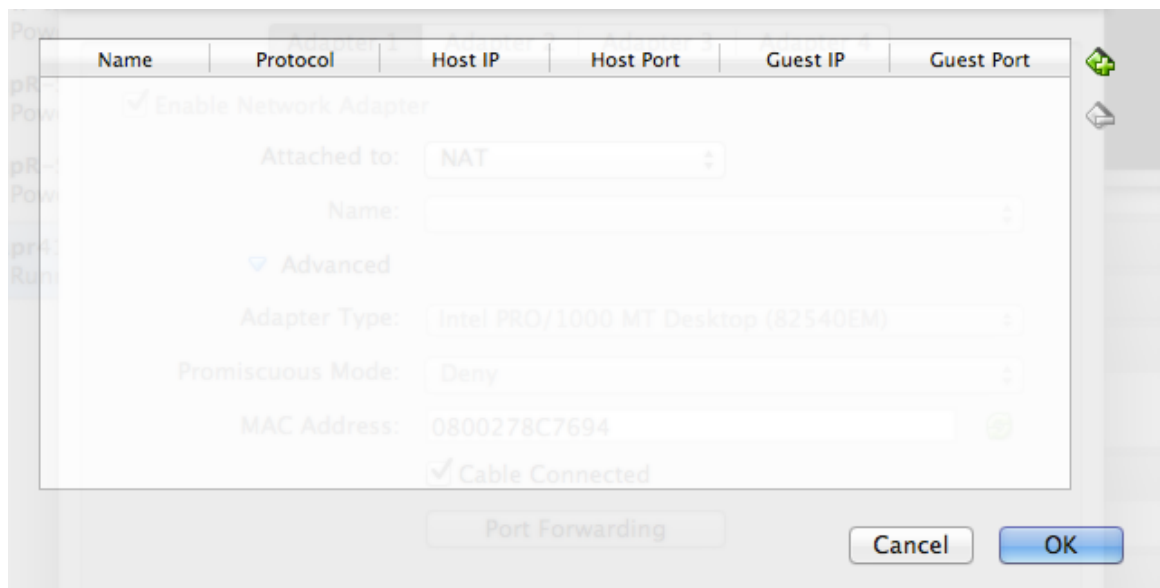
5. Create port forwarding rules for NAT.
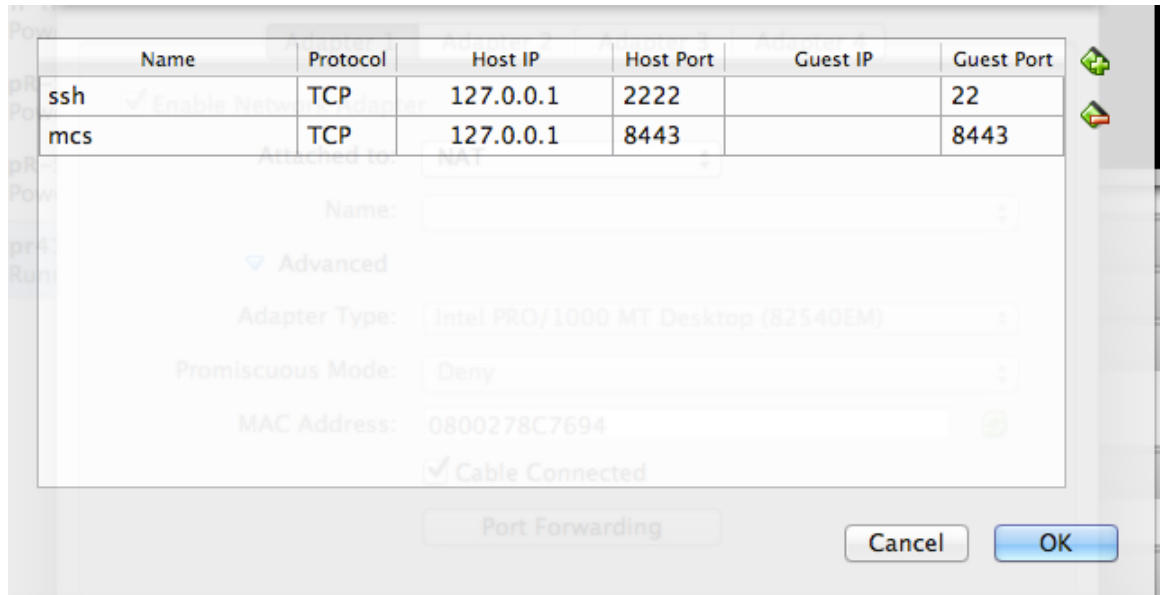   a. In the VirtualBox Manager GUI, select your virtual machine and click the "Network" link.

b. In the Network wizard window, select "NAT" in the "Attached to" drop-down list and click the "Port Forwarding" button under "Advanced".
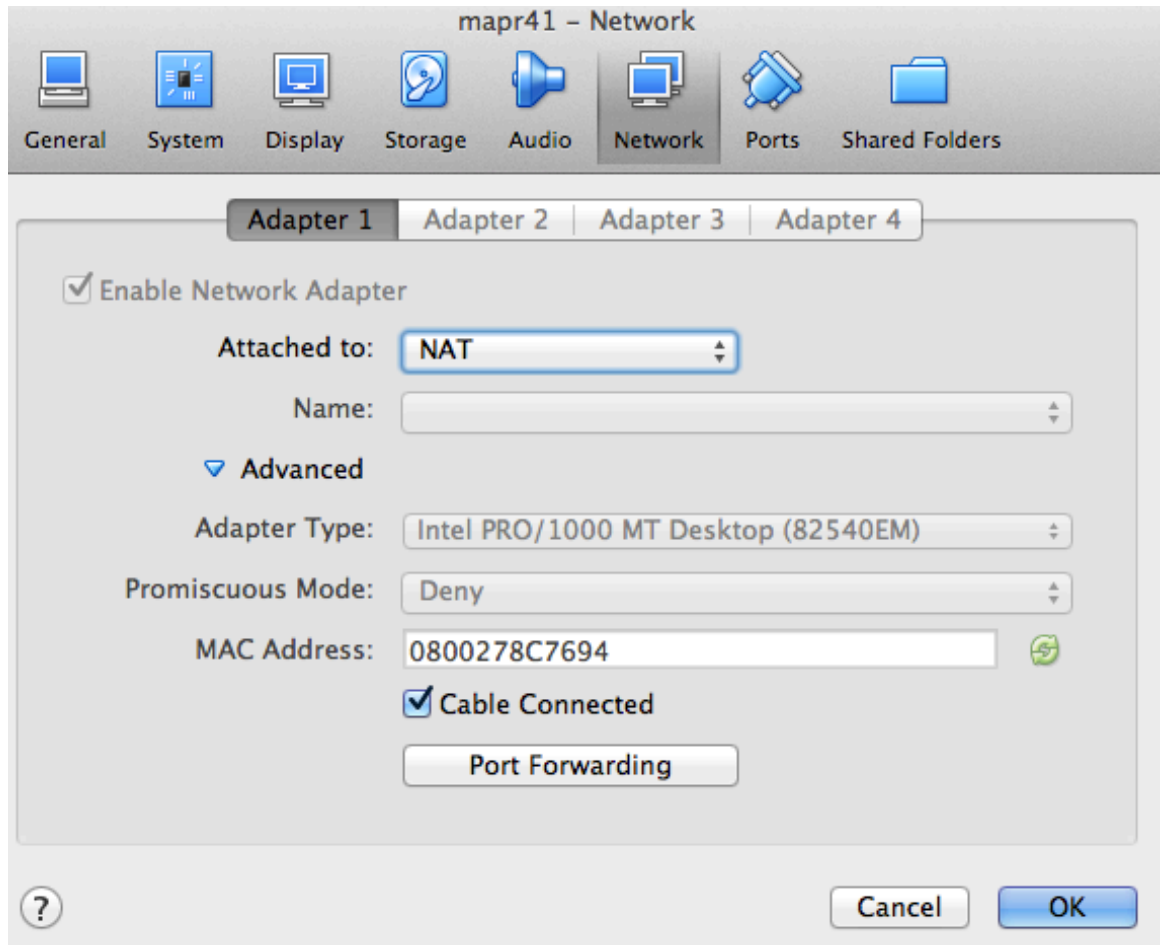
c. In the port-forwarding pop-up window, click the icon with the "+" symbol at the top right-hand side.

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

Name	Protocol	Host IP	Host Port	Guest IP	Guest Port
ssh	TCP	127.0.0.1	2222		22
mcs	TCP	127.0.0.1	8443		8443

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

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

7. Create `mapr` user and group

```
groupadd mapr
useradd –g mapr –d /home/mapr –s /bin/bash –m mapr
passwd mapr (set to mapr)
```

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

```
mkdir /user
mkdir /mapr
```

9. Update `/etc/sysconfig/network-scripts/ifcfg-eth0` as follows and bounce network service (service network restart).

(delete all the other properties from the file)
```
DEVICE="eth0"
BOOTPROTO="dhcp"
ONBOOT="yes"
```

Restart the service –

```
service network restart
```

10. Disable iptables and selinux

```
service iptables stop
chkconfig iptables off
vi /etc/selinux/config (set SELINUX=disabled)
```

11. Configure Linux EPEL repo

```
yum install wget -y

cd

wget http://dl.fedoraproject.org/pub/epel/6/x86_64/\
epel-release-6-8.noarch.rpm

rpm -Uvh epel-release-6*.rpm
```

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

```
yum update -y
```

13. Install extra packages

```
yum remove java-1.* -y
yum install java-1.7.0-openjdk -y
yum install java-1.7.0-openjdk-devel -y
yum install nfs-utils -y
yum install zip unzip -y
yum install git -y
yum install acpid -y
```

14. Create mapr repo for 5.0.0 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/v5.0.0/redhat
enabled=1
gpgcheck=0
protect=1
```

15. Create mapr ecosystem repo

vi mapr-eco.repo (add the following to the file)

```
[maprecosystem]
name=MapR Technologies
baseurl=http://package.mapr.com/releases/ecosystem-5.x/redhat
enabled=1
gpgcheck=0
protect=1
```

16. Install MapR 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
yum install mapr-gateway -y
```

to see that the above were installed
rpm —qa|grep mapr

17. Update the /etc/hosts and /opt/mapr/hostname files.
   a. Determine the hostname

   hostname

   b. Determine the IP address.

   ifconfig eth0

   c. Edit the /etc/hosts file

vi etc/hosts
keep the existing line -
127.0.0.1   localhost (remove the rest)

Add line
10.0.2.15 mapr1node
(whatever you named the host during the installation – use the hostname command to find the name if you forgot)

      d. Populate the `/opt/mapr/hostname` file

`hostname > /opt/mapr/hostname`

18. Run `configure.sh` to configure the MapR cluster. Replace the hostname `mapr1node` with whatever you named the host during the installation. Run the `hostname` command first if you want to verify.

```
/opt/mapr/server/configure.sh –C mapr1node:7222 \
–Z mapr1node:5181  –RM mapr1node –HS mapr1node  \
–N mycluster
```

19. Before you go to the next step – i.e. adding the disks, run the following to check which disk is the mount disk. Only those that are not mounted should be added to the file that you will create in the following step.
      a. **`fdisk –l|more`** (this will give the list of all the disks – you should see /dev/sdb, /dev/sdc, and /dev/sdd in your output. If you don't, you cannot proceed in the installation.)
      b. **`df`**   (will give the mounted disks)

20. 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`

~~**21.** adjust warden.conf memory settings (if needed)~~

22. Start mapr services

```
service mapr-zookeeper start
service mapr-warden start
```

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`**

23. Add `mapr` user to mcs acl

`/opt/mapr/bin/maprcli acl edit –type cluster \`
`–user mapr:fc`

24. Create `/opt/mapr/conf/mapr_fstab` file for auto-mounting NFS exports at boot

`vi /opt/mapr/conf/mapr_fstab` (add the following 2 lines)
`  localhost:/mapr /mapr    hard,intr,nolock`
`  localhost:/mapr/mycluster/user /user hard,intr,nolock`

25. Login to Hadoop Web UI / MCS
   a. Open a Web browser on your laptop.

   b. Type the following URL in your browser
   https://localhost:8443

   c. Depending on which Web browser you're using, you'll need to "accept" the security risks associated with this site. This is because the MapR Web server uses a self-signed SSL certificate.

   d. Login as the mapr user (password is 'mapr' or whatever you set it to previously in this installation document when you created the mapr user).
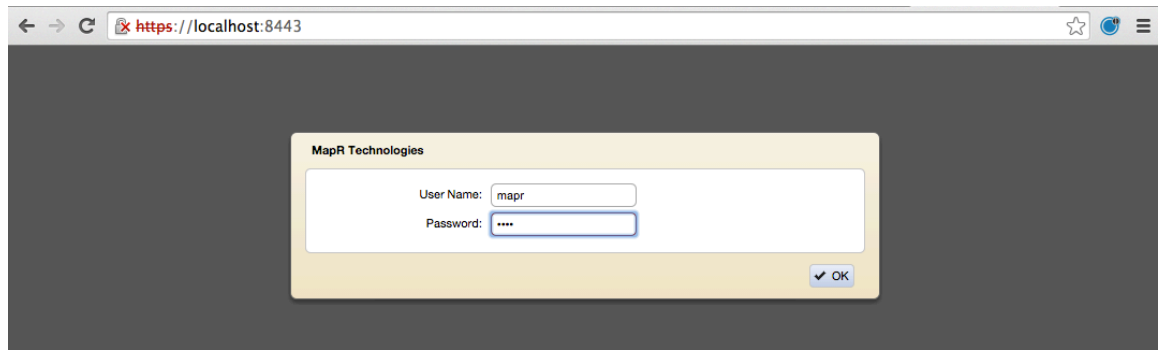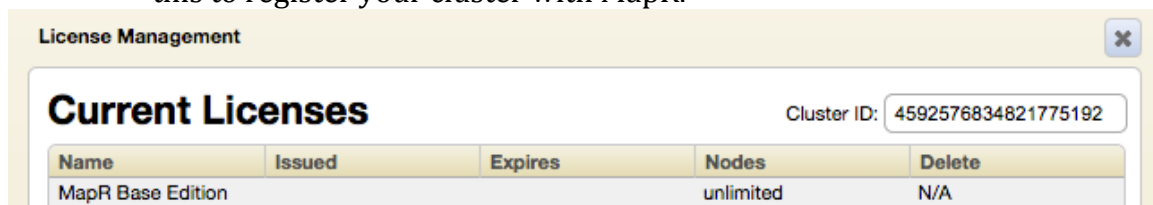
e. The first time you log in to the MCS, you'll need to accept or decline using dial-home metrics in a pop-up dialog.

f. Click the "Manage Licenses" link at the top right-hand side of the GUI.
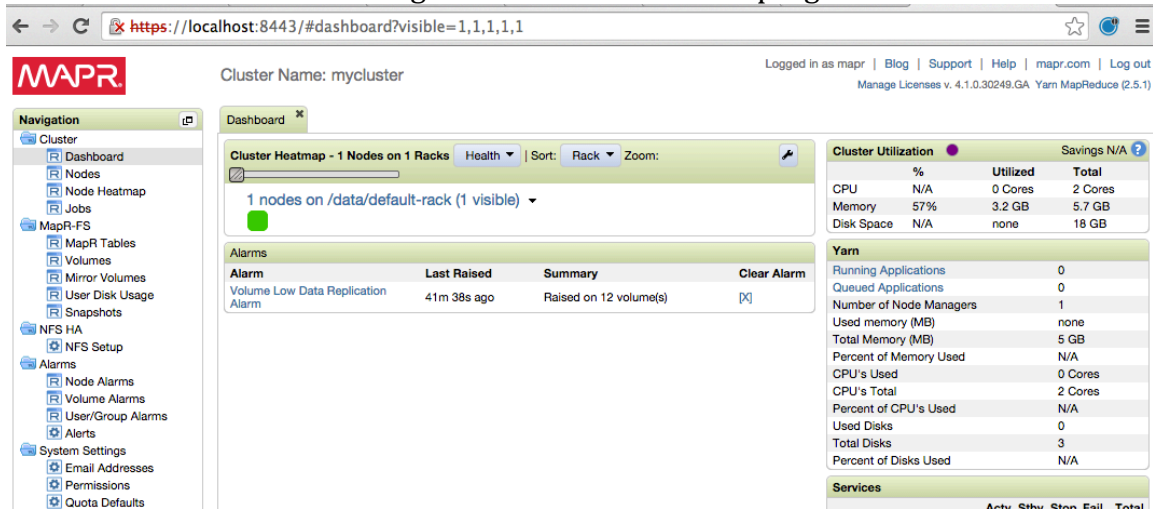
g. Copy the "cluster ID" (or write it down somewhere). You will need this to register your cluster with MapR.

26. 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

or

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.

Password *

Password strength:

Confirm password *

Provide a password for the new account in both fields.

First name *

Last name *

Company *

27. Register your cluster in the MCS
   a. a. Click the "Manage Licenses" link.

   b. Click the "Add licenses via Web" link in the popup window.

      This will open a page in http://mapr.com in which your list of registered clusters is displayed.

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

   d. Back in the MCS window, click the "add licenses via copy/paste" link in the "Manage Licenses" popup window. Paste the key string from your clipboard into the text area and click the "apply licenses" button.

28. Reboot vm

    `reboot`

29. Login to your VM, give time for the cluster to start up, and validate cluster services are running

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

14

> **`maprcli node list`** (look for "Healthy")
>
> **`showmount -e`** (look for `/mapr` and `/mapr/mycluster` to be exported)
>
> **`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)

30. Install and configure tmpwatch to keep your /tmp directory uncluttered.

> **`su - root`**
>
> **`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 3d /tmp`**
>
> **`crontab -l`** (to verify you edited the crontab properly)

31. Create an end user

> **`groupadd user01`**
>
> **`useradd –g user01 –d /home/user01 –s /bin/bash \`**
> **`–m user01`**
>
> **`passwd user01`** (set to mapr)
>
> **`mkdir /user/user01`**
>
> **`chown user01:user01 /user/user01`**

32. 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`**
>
> **`export JAVA_HOME=/usr`**
>
> **`hadoop jar /opt/mapr/hadoop/hadoop-\`**
> **`2.7.0/share/hadoop/mapreduce/hadoop-mapreduce-\`**

```
examples-2.7.0-mapr-1506.jar wordcount \
file:///etc/passwd /tmp/out-$USER

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

33. Add startup script to resolve IP addresses automatically for when you change between networking modes in your VM (i.e. NAT, bridged, host-only). Perform the following steps as the root user.
   a. Create the script
`vi /etc/init.d/mapr-ipaddr`
```
#! /bin/bash
#
# mapr-ipaddr      set local IP in /etc/hosts
#
# chkconfig: 2345 10 90
# description:
#
### BEGIN INIT INFO
# Provides: $mapr-ipaddr
# Should-Start: network
# Short-Description: set ip
# Description: set ip
### END INIT INFO

# Source function library.
. /etc/init.d/functions

if [ ! -f /etc/sysconfig/network ]; then
    exit 6
fi

. /etc/sysconfig/network

if [ -f /etc/sysconfig/pcmcia ]; then
 . /etc/sysconfig/pcmcia
fi


# Check that networking is up.
[ "${NETWORKING}" = "no" ] && exit 6

# if the ip configuration isn't around we can't function
[ -x /sbin/ip ] || exit 1

CWD=$(pwd)
cd /etc/sysconfig/network-scripts

. ./network-functions


# See how we were called.
case "$1" in
  start)
```

```
    myHostname=$(/bin/hostname)
        myNewIpAddress=$(/sbin/ifconfig eth0 | grep inet| grep -v \
inet6 | awk '{print $2}' | cut -d: -f2)
        myOldIpAddress=$(grep $myHostname /etc/hosts | \
awk '{print $1}')
        if [ "$myNewIpAddress" != "$myOldIpAddress" ]
        then
            sed -i "s/$myOldIpAddress/$myNewIpAddress/" /etc/hosts
        fi


        ;;
  stop)
        ;;
  status)
        cat /etc/hosts
 ;;
 restart|reload|force-reload)
 $0 stop
 $0 start
 rc=$?
 ;;
 *)
 echo $"Usage: $0 {start|stop|status|restart|reload|force-reload}"
        exit 2
 esac
```

       b.   make the script executable

chmod a+x /etc/init.d/mapr-ipaddr

       c.   add the service

chkconfig --add /etc/init.d/mapr-ipaddr

       d.   reboot the VM

reboot

34. If you want to increase the amount of memory available to your YARN applications (or otherwise wish to manage the memory allocations for your warden-controlled services), perform the following steps as the root user:
    a. Turn off the warden service.

    service mapr-warden stop

      b. Make a backup copy of the `warden.conf` file.

```
cd /opt/mapr/conf
cp warden.conf warden.conf.orig
```

      c. Edit the `warden.conf` file with the following changes.

        Delete the following lines:

```
service.command.jt.start=/opt/mapr/hadoop/hadoop-0.20.2/bin/hadoop-
daemon.sh start jobtracker
service.command.tt.start=/opt/mapr/hadoop/hadoop-0.20.2/bin/hadoop-
daemon.sh start tasktracker
service.command.hbmaster.start=/opt/mapr/hbase/hbase-0.90.4/bin/hbase-
daemon.sh start master
service.command.hbregion.start=/opt/mapr/hbase/hbase-0.90.4/bin/hbase-
daemon.sh start regionserver

service.command.jt.stop=/opt/mapr/hadoop/hadoop-0.20.2/bin/hadoop-
daemon.sh stop jobtracker
service.command.tt.stop=/opt/mapr/hadoop/hadoop-0.20.2/bin/hadoop-
daemon.sh stop tasktracker
service.command.hbmaster.stop=/opt/mapr/hbase/hbase-0.90.4/bin/hbase-
daemon.sh stop master
service.command.hbregion.stop=/opt/mapr/hbase/hbase-0.90.4/bin/hbase-
daemon.sh stop regionserver

service.command.jt.type=BACKGROUND
service.command.tt.type=BACKGROUND
service.command.hbmaster.type=BACKGROUND

service.command.jt.monitor=org.apache.hadoop.mapred.JobTracker
service.command.tt.monitor=org.apache.hadoop.mapred.TaskTracker
service.command.hbmaster.monitor=org.apache.hadoop.hbase.master.HMaster
start
service.command.hbregion.monitor=org.apache.hadoop.hbase.regionserver.H
RegionServer start

service.command.jt.monitorcommand=/opt/mapr/hadoop/hadoop-
0.20.2/bin/hadoop-daemon.sh status jobtracker
service.command.tt.monitorcommand=/opt/mapr/hadoop/hadoop-
0.20.2/bin/hadoop-daemon.sh status tasktracker
service.command.hbmaster.monitorcommand=/opt/mapr/hbase/hbase-
0.90.4/bin/hbase-daemon.sh status master
service.command.hbregion.monitorcommand=/opt/mapr/hbase/hbase-
0.90.4/bin/hbase-daemon.sh status regionserver

service.command.tt.heapsize.percent=2
service.command.tt.heapsize.max=325
service.command.tt.heapsize.min=64
service.command.hbmaster.heapsize.percent=4
service.command.hbmaster.heapsize.max=512
service.command.hbmaster.heapsize.min=64
```

```
service.command.hbregion.heapsize.percent=25
service.command.hbregion.heapsize.max=4000
service.command.hbregion.heapsize.min=1000

service.command.jt.heapsize.percent=10
service.command.jt.heapsize.max=5000
service.command.jt.heapsize.min=256

service.command.tt.heapsize.percent=2
service.command.tt.heapsize.max=325
service.command.tt.heapsize.min=64
service.command.hbmaster.heapsize.percent=4
service.command.hbmaster.heapsize.max=512
service.command.hbmaster.heapsize.min=64

hbmaster.port=60000


jt.port=9001
jt.http.port=50030
```

Edit the properties as follows:

```
service.command.cldb.heapsize.percent=8
service.command.cldb.heapsize.max=2000
service.command.cldb.heapsize.min=256
service.command.mfs.heapsize.percent=20
service.command.mfs.heapsize.maxpercent=40
service.command.mfs.heapsize.min=256
service.command.webserver.heapsize.percent=2
service.command.webserver.heapsize.max=512
service.command.webserver.heapsize.min=256
service.command.nfs.heapsize.percent=2
service.command.nfs.heapsize.min=32
service.command.nfs.heapsize.max=512
service.command.os.heapsize.percent=10
service.command.os.heapsize.max=2000
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=512
service.command.zk.heapsize.min=256
```

d. Run the `configure.sh` command.

/opt/mapr/server/configure.sh -R

e. Start the warden service.

service mapr-warden start

19