

Multi Node Hadoop Set Up

K Som Shekhar Sharma

Table of Contents

1.	How many Nodes you want?	3
2.	Change the Network	3
3.	Change the virtual machine name	4
4.	Start your VM machine	5
5.	Change host name of your machines	6
6.	Stop the firewall	6
7.	Change hosts files	6
8.	Modify your Hadoop's configuration files	7
9.	Delete hadoop-temp folder	7
10.	Format and start your cluster	7
11.	Adding new machines to the existing cluster	7

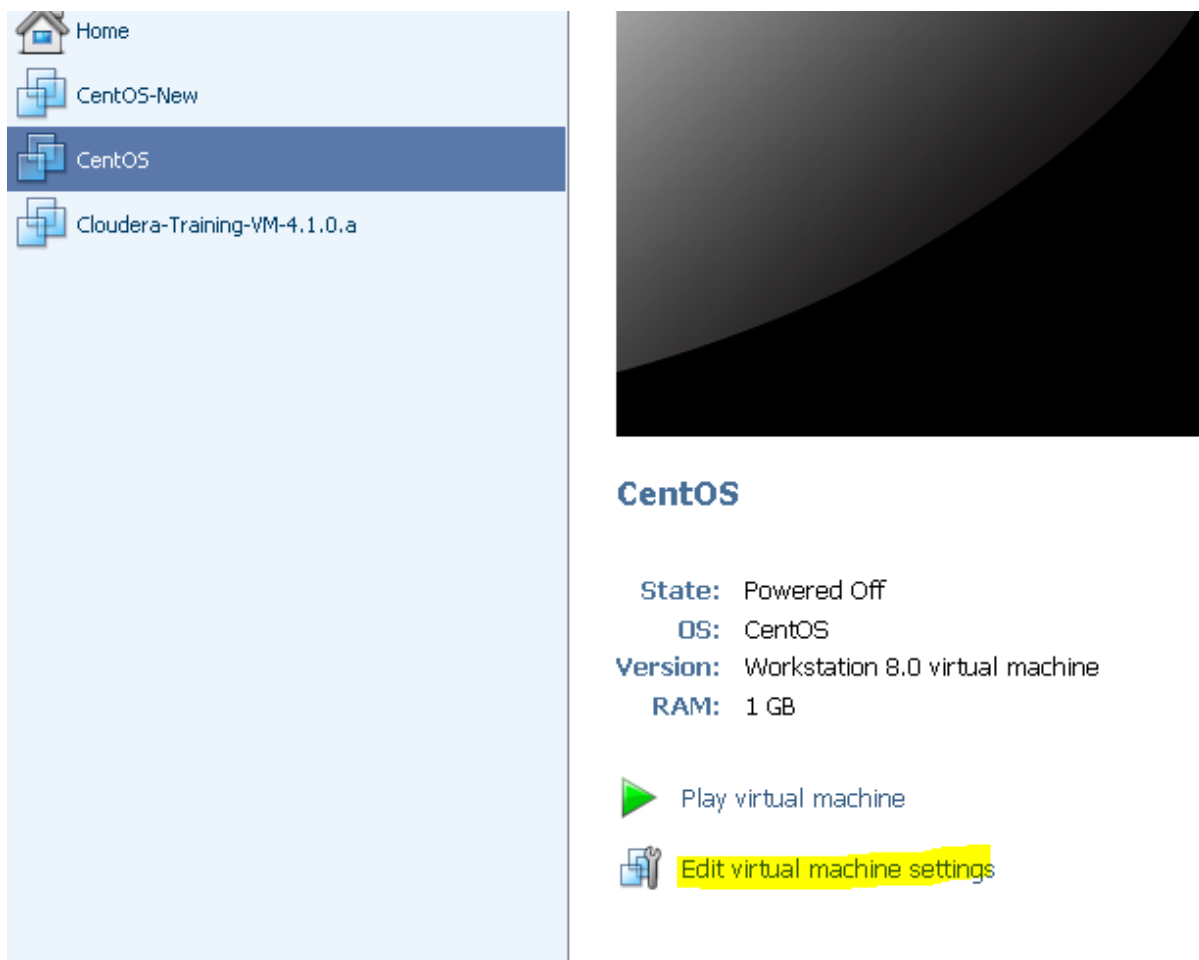
1. How many Nodes you want?

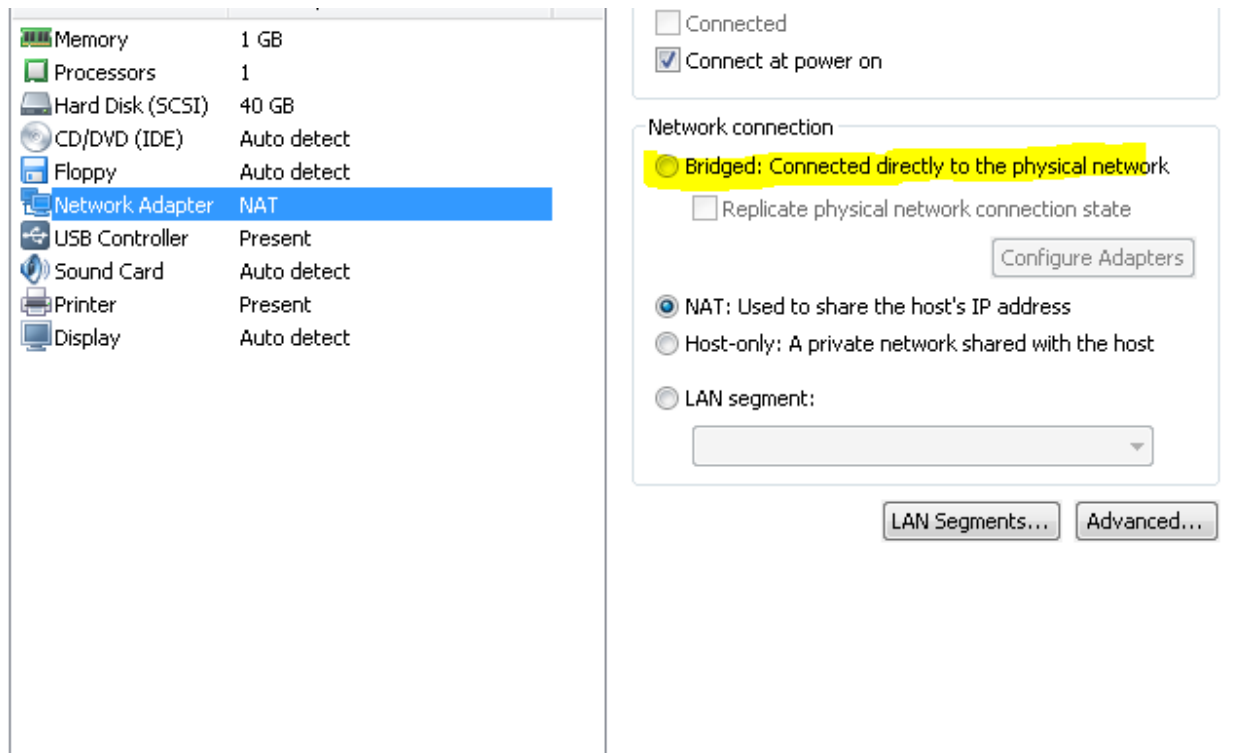
Create the copies of the virtual machine file which I have given to you. Create as many copies as many nodes you want. And open them in VM ware player.

NOTE: Please don't click on "Play Virtual machine" first change the networking and VM name and then start your virtual machine

2. Change the Network

Since you need the communication between your nodes, you need to change the networking from NAT to bridge. See the screen shot below

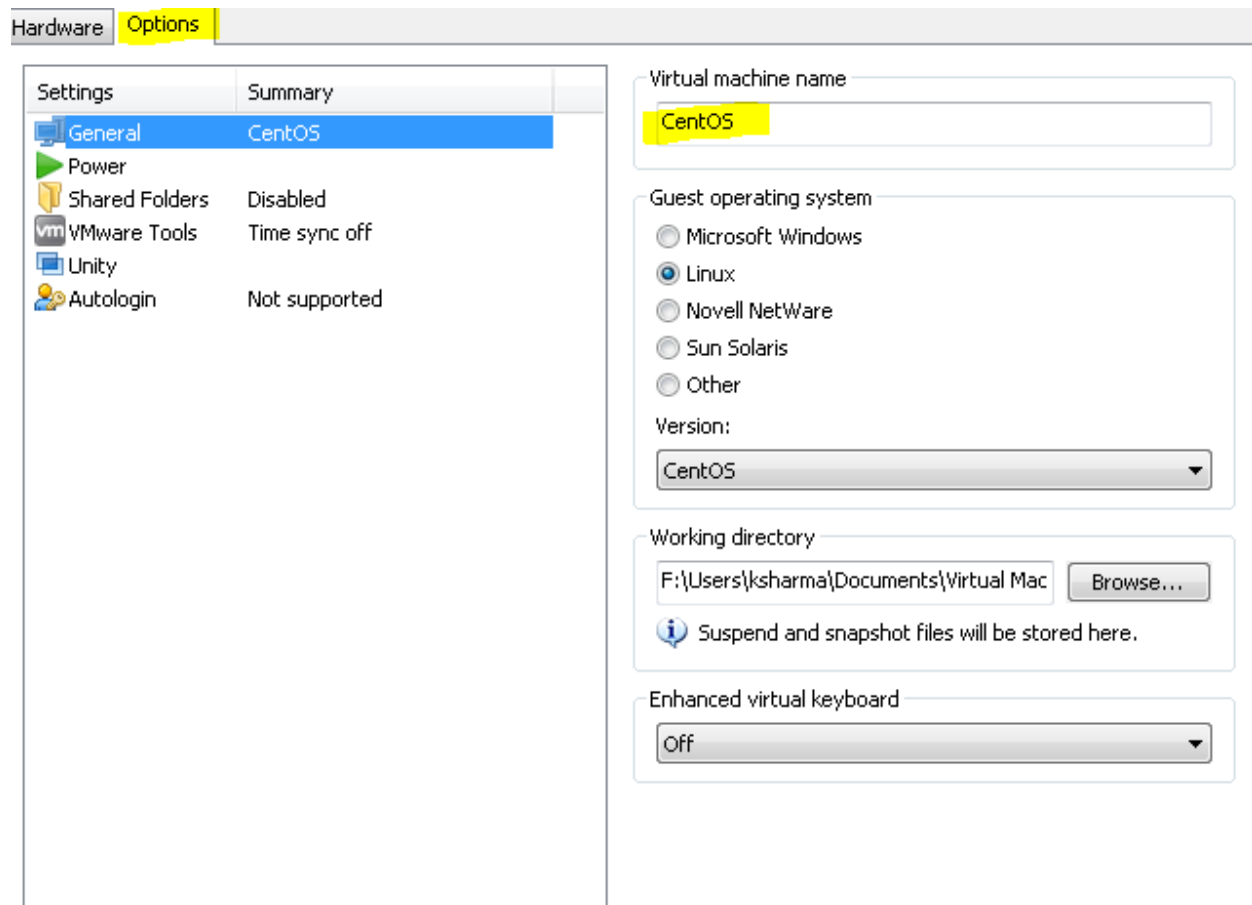




3. Change the virtual machine name

Consider changing your name of the node. Give appropriate names like NameNode, DataNode etc.

You can change the name of your machine as shown in below screen shot



Let's assume that you have two machines with the following IP address (after changing the networking from NAT to bridge)

Run the "*ifconfig*" command from the terminal of the machine and get the IP address of that machine

10.99.64.182 (This machine I would like to make it as my NameNode) 10.99.64.82 (This machine I would like to make it as my DataNode)
--

4. Start your VM machine

Now you are ready to play your virtual machine. Click on the option "Play Virtual machine".

When it starts, it will ask you for some options, choose "I copied" option

5. Change host name of your machines

Once the virtual machine is up and you have entered into your virtual change the hostname of your machine. By default your host name of your machine would be "localhost.localdomain". This will not work with multi node

Go to each of the machine and you need to change the host name of the machine

```
$su root
$vi /etc/sysconfig/network

#Edit the hostname to NameNode or DataNode depending on which machine you are
#close this file and reboot your machine to take the hostname into effect

$reboot

# once you rebooted your machine and when you run the below command, it should show the
proper hostname which you have given

$hostname
```

6. Stop the firewall

Do the following operation in each of the machines

```
$su root

##Enter the password for the root and run the following commands to stop the firewall in each of
the machines

$service iptables save
$service iptables stop
$chkconfig iptables off
```

7. Change hosts files

After doing this, go in each of the machines and change the "/etc/hosts" file

```
$su root

$vi /etc/hosts
# edit the file by providing the mapping the IP address and hostname
#for example you can append the following entries like as follows depending on ur IP address
10.99.64.182 NameNode
```

10.99.64.82 DataNode

8. Modify your Hadoop's configuration files

Now change the Configuration files of the hadoop, replace localhost with "NameNode"

In the masters file give the entry as NameNode

And in the slaves file give the entries of you slaves (Give the hostname)

Make sure that slave machines atleast know the address of the NameNode and job tracker and masters file must have the entry for your NameNode

9. Delete hadoop-temp folder

Delete the hadoop-temp folder from each of the machines

10. Format and start your cluster

From the namenode , format your hadoop cluster

\$hadoop namenode -format

Now go into the NameNode machine and run the command start-all.sh

Check your NameNode UI(10.99.64.182:50070), and verify whether you can see two node cluster

Note: You will 2 nodes only when your Master machine is also acting as slaves, else it will show you only 1 node

11. Adding new machines to the existing cluster

For adding new machines to the existing cluster, you don't have to stop the existing running cluster and don't have to change.

For the new machine, do the following

- Change its hostname (don't forget to reboot your machine after changing the hostname)
- Stop the firewall
- In the "/etc/hosts" file give map IP address of the host name..better map the IP address of all the machines with their hostname.
- Delete the hadoop-temp folder
- Manually run the following command to start datanode and tasktracker on this machine

```
$hadoop datanode
```

```
#the above command will not daemonize so open a new terminal and run the following command
```

```
$hadoop tasktracker
```

Again open your NameNode UI and check whether you can see the new data node or not