CDH – CREATE 4 NODES FOR CLUSTER

By www.HadoopExam.com

Note: These instructions should be used with the HadoopExam Apache OOzie: Professional Trainings. Where it is executed and you can do hands on with trainer.

Step 1: Clone the VMs (First shutdown, if vm is running)

init 0

Step 2: Now start cloning the VM from workstation
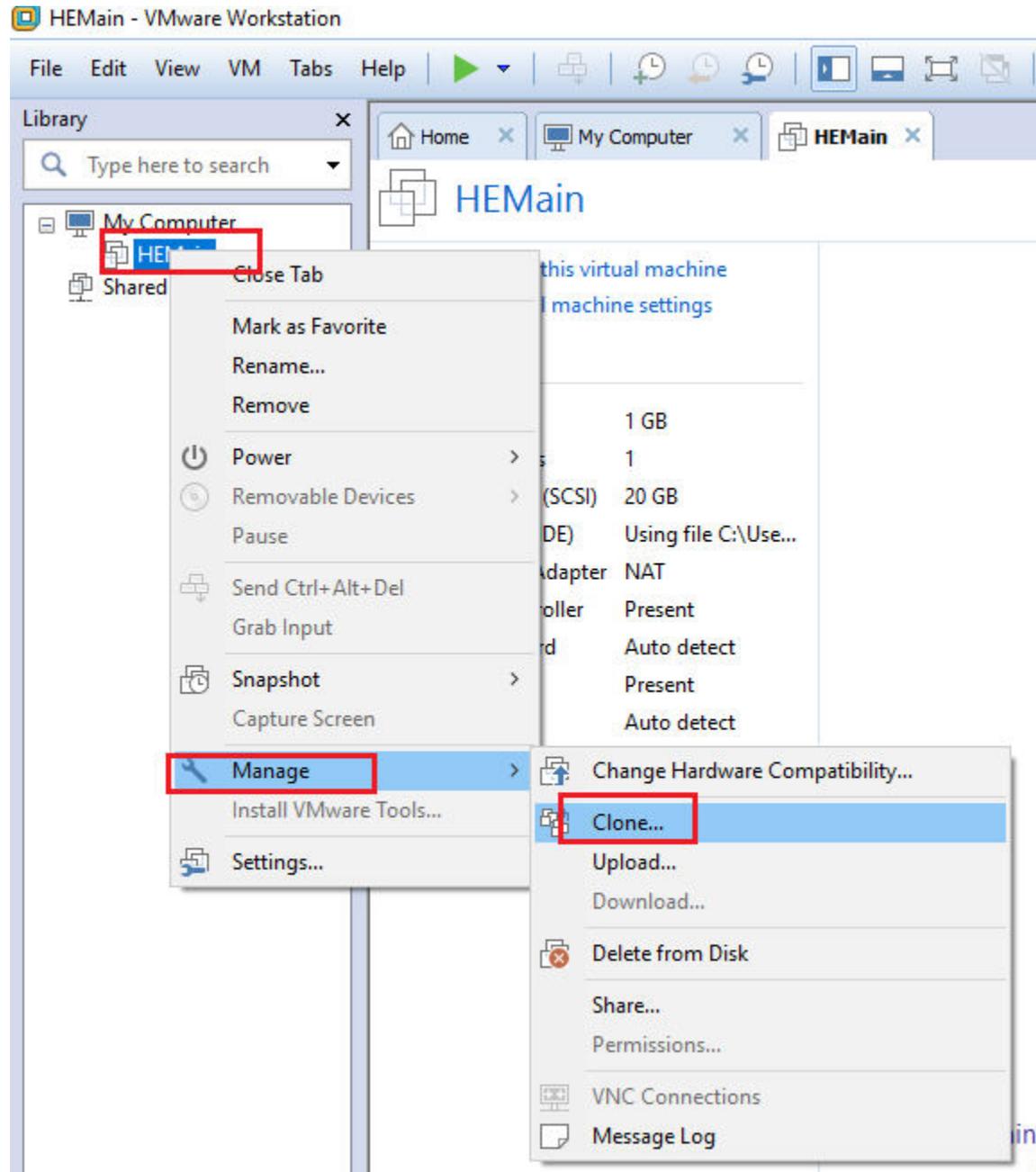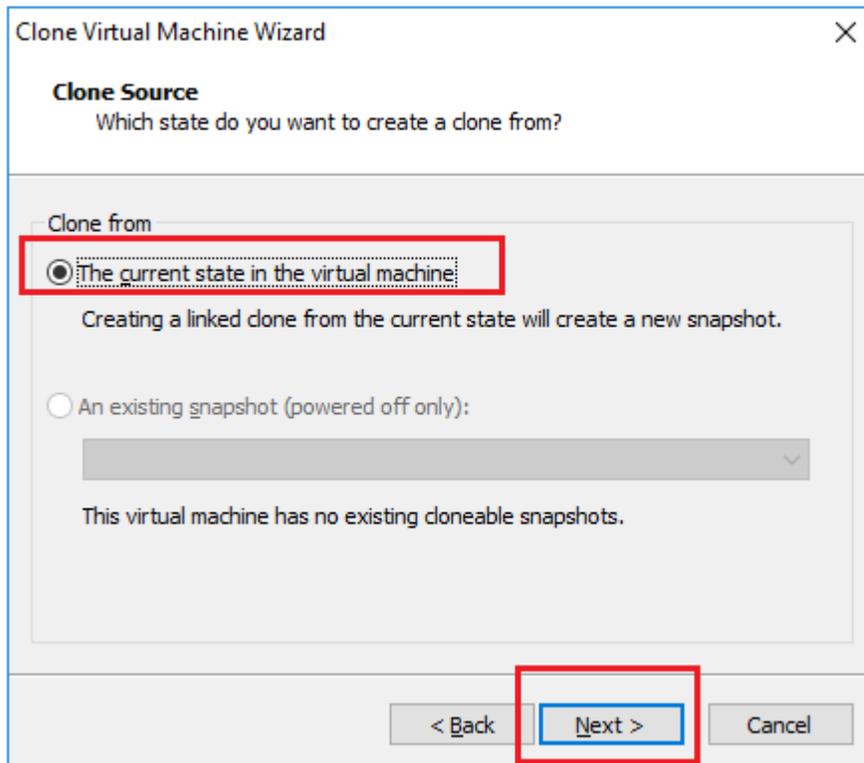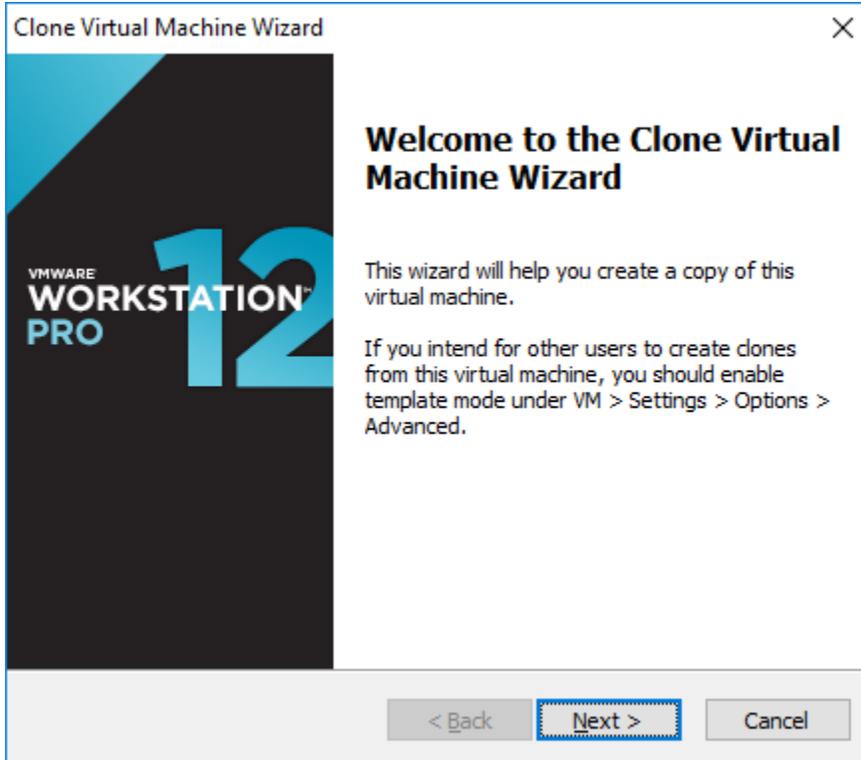
It is important you create full clone.

Clone Virtual Machine Wizard   ✕

Clone Type
How do you want to clone this virtual machine?

Clone method

○ Create a linked clone
A linked clone is a reference to the original virtual machine and requires less
disk space to store. However, it cannot run without access to the original
virtual machine.

◉ Create a full clone
A full clone is a complete copy of the original virtual machine at its current
state. This virtual machine is fully independent, but requires more disk space
to store.

< Back   Next >   Cancel

Give node name like master101.hadoopexam.com

Clone Virtual Machine Wizard   ✕

Name of the New Virtual Machine
What name would you like to use for this virtual machine?

Virtual machine name

master101.hadoopexam.com

Location

C:\Users\hadoo\Documents\Virtual Machines\master101.hadoop   Browse...

< Back   Finish   Cancel

Cloning will be finished in few minutes

Clone Virtual Machine Wizard ✕

Cloning Virtual Machine

✔ Preparing clone operation

VMware Workstation

Cloning...

[████████████████████████████]

Cancel

Close

Click close

Clone Virtual Machine Wizard ✕

Cloning Virtual Machine

✔ Preparing clone operation

✔ Creating full clone

✔ Done
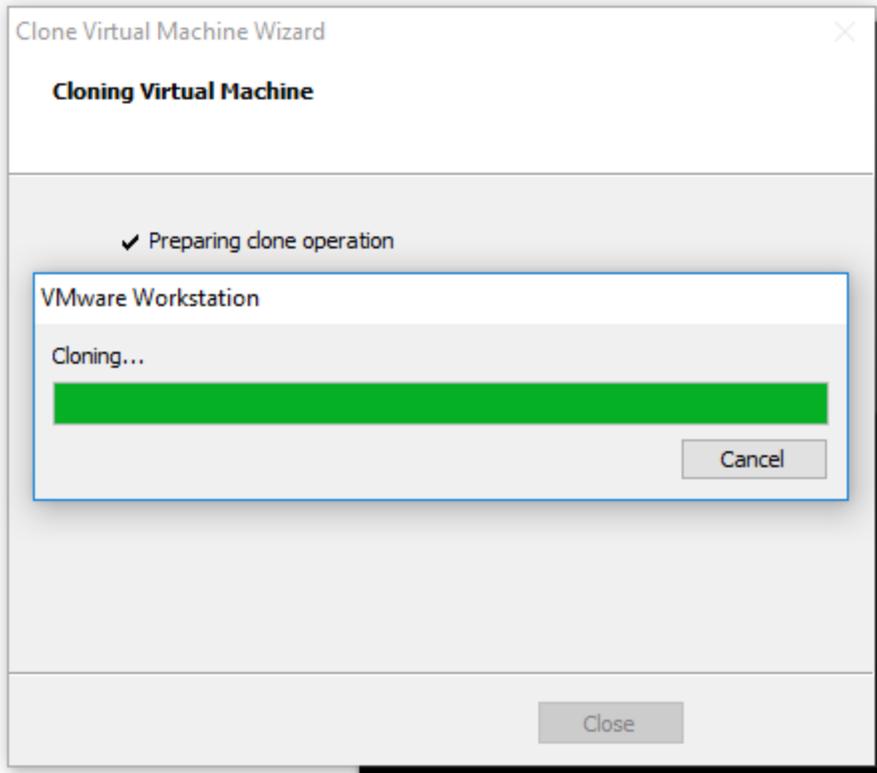
Close
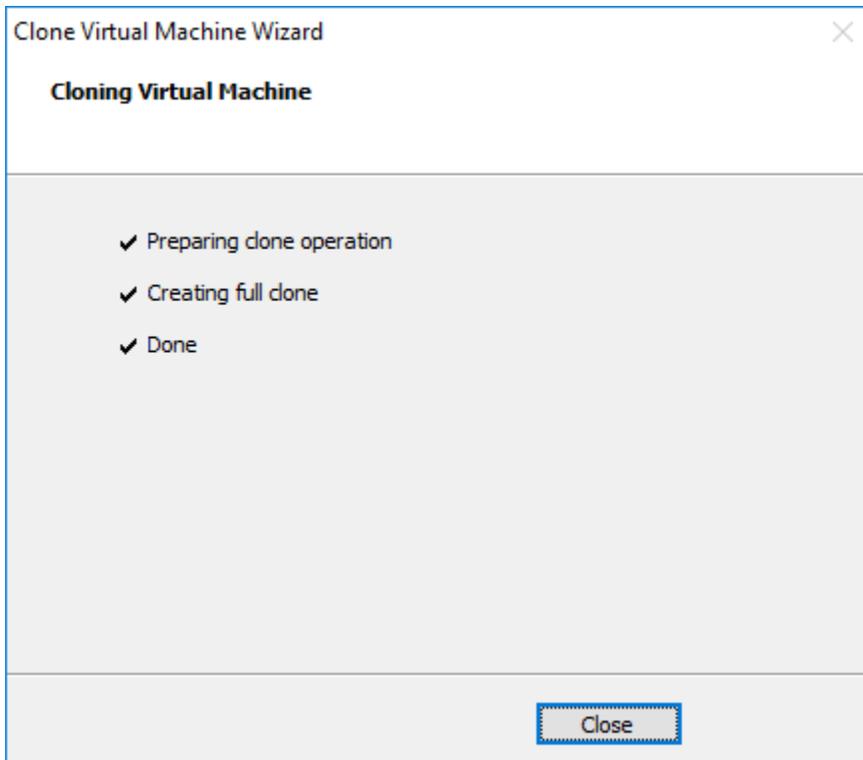
Now we have one copy of our vm created. As we going to have 4 node cluster, we will first do some configuratios, which is required on all 4 nodes.

Step 3: Now start the cloned vm (master101.hadoopexam.com)
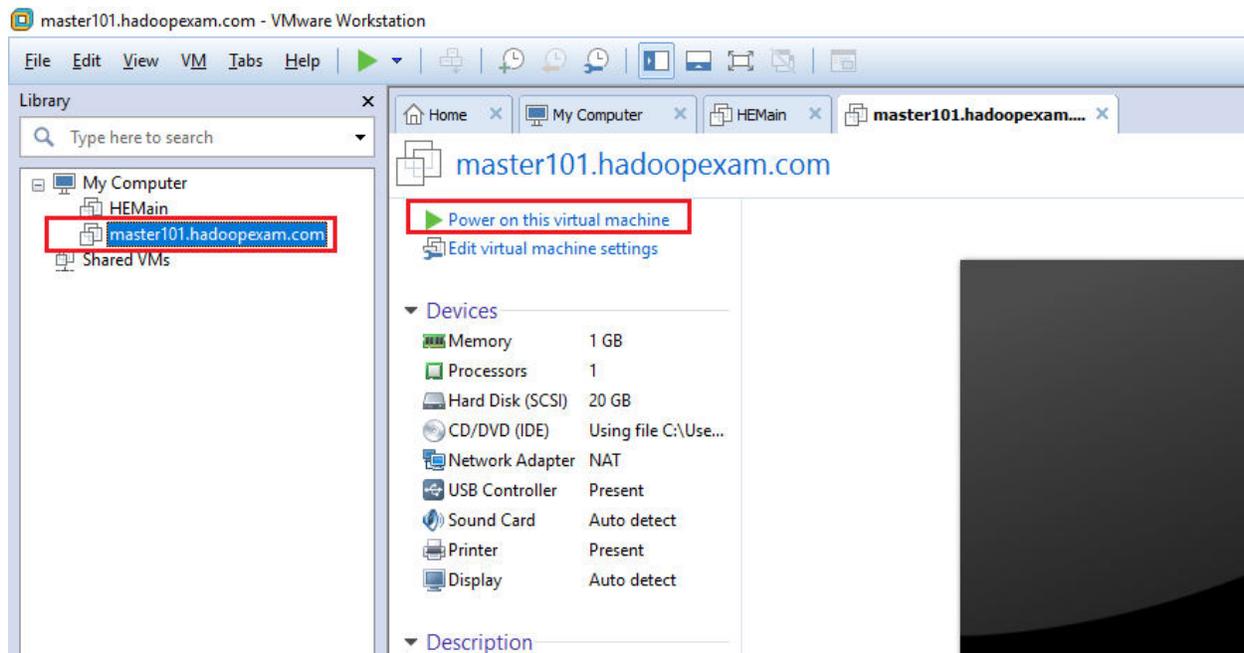
Step 4: Change the hostname

```
vi /etc/sysconfig/network
```

Step 5: Change the ip address of this hosts. If you type command ifconfig –a, it will show previous ip. We need to change that ip

Change the following file

```
vi /etc/sysconfig/network-scripts/ifcfg-ens33
```

Step 6: Now change the hosts file

```
vi /etc/hosts
```

```
127.0.0.1    localhost localhost.localdomain localhost4 localhost4.localdomain4
::1          localhost localhost.localdomain localhost6 localhost6.localdomain6
192.168.239.11          master101.hadoopexam.com
~
~
```

Step 7: Now restart the network

```
service network restart
```

Check the ip address of this machine

```
[root@hemain ~]# service network restart
Restarting network (via systemctl):                    [  OK  ]
[root@hemain ~]#
[root@hemain ~]# ifconfig -a
ens33: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.239.11  netmask 255.255.255.0  broadcast 192.168.239.255
        inet6 fe80::20c:29ff:fe78:8b1d  prefixlen 64  scopeid 0x20<link>
        ether 00:0c:29:78:8b:1d  txqueuelen 1000  (Ethernet)
        RX packets 104  bytes 7887 (7.7 KiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 72  bytes 5686 (5.5 KiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

Step 8: Now try to connect from windows Cygwin

```
ssh root@192.168.239.11
```

```
ssh root@192.168.239.11
e authenticity of host '192.168.239.11 (192.168.239.11)' can't be established.
DSA key fingerprint is SHA256:BSUBvtOq4oSz6+hyBVoHPxmYGR4SFCm44vx6iuj2e5k.
e you sure you want to continue connecting (yes/no)? yes
rning: Permanently added '192.168.239.11' (ECDSA) to the list of known hosts.
st login: Mon Jul 31 00:08:34 2017
oot@master101 ~]#
```

Step 9: Now change the/etc/hosts file for other nodes (Which we will be creating soon)

```
echo '192.168.239.12     node102.hadoopexam.com' >> /etc/hosts
echo '192.168.239.13     node103.hadoopexam.com' >> /etc/hosts
echo '192.168.239.14     node104.hadoopexam.com' >> /etc/hosts
```

Step 10: Add the same entry in our windows hosts file.

```
echo '192.168.239.11 master101' >> /etc/hosts
echo '192.168. 239.11 master101.hadoopexam.com' >> /etc/hosts

echo '192.168.239.12 node102' >> /etc/hosts
echo '192.168.239.12 node102.hadoopexam.com' >> /etc/hosts

echo '192.168.239.13 node103' >> /etc/hosts
echo '192.168.239.13 node103.hadoopexam.com' >> /etc/hosts

echo '192.168.239.14 node104' >> /etc/hosts
echo '192.168.239.14 node104.hadoopexam.com' >> /etc/hosts
```

```
#         ::1              localhost
127.0.0.1        sandbox.hortonworks.com

192.168.239.101                   hemain
192.168.239.11 master101
192.168. 239.11 master101.hadoopexam.com
192.168.239.12 node102
192.168.239.12 node102.hadoopexam.com
192.168.239.13 node103
192.168.239.13 node103.hadoopexam.com
192.168.239.14 node104
192.168.239.14 node104.hadoopexam.com
```

Step 11: Now try to connect master101 from windows, using hostname

```
ssh root@master101
```

```
adoo@HadoopExam ~/.ssh
 ssh root@master101
he authenticity of host 'master101 (192.168.239.11)' can't be established.
CDSA key fingerprint is SHA256:BSUBvtOq4oSz6+hyBVoHPxmYGR4SFCm44vx6iuj2e5k.
re you sure you want to continue connecting (yes/no)? yes
arning: Permanently added 'master101' (ECDSA) to the list of known hosts.
ast login: Mon Jul 31 00:13:40 2017 from 192.168.239.1
```

Stop the VM and create 3 more VMs , using master101 VM

Step 12: Now create remaining 3 VMs by cloning master101 vms

Once you are done, you should have following 4 equals VM created
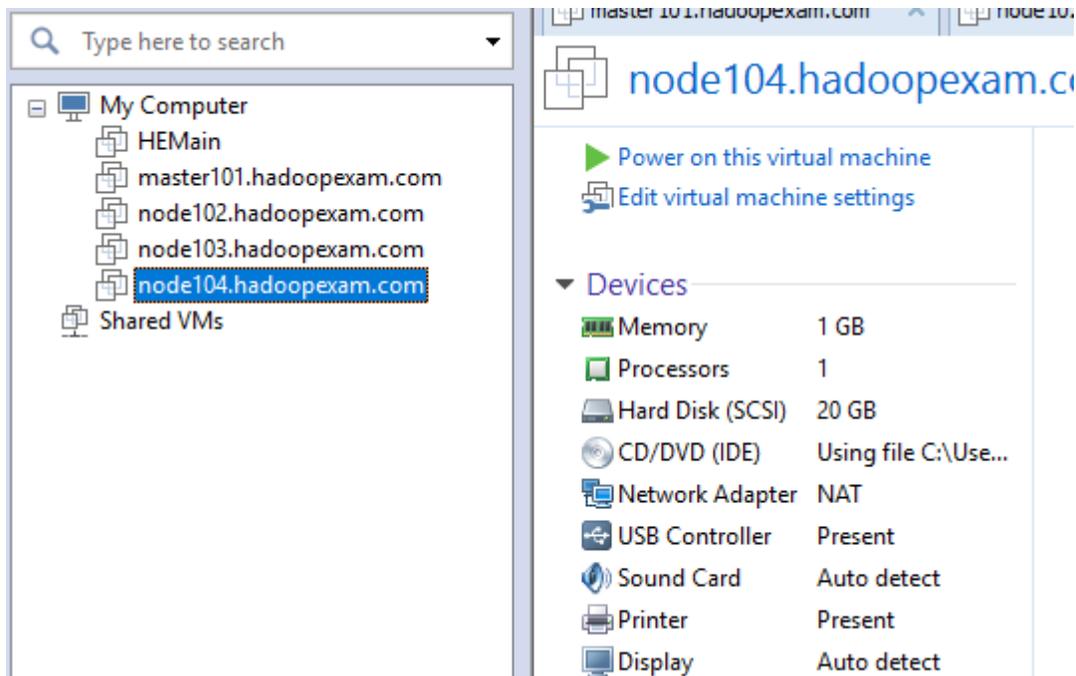
Step 13: Now start VM one by one and change their ip addresses.

Change the following file

```
vi /etc/sysconfig/network-scripts/ifcfg-ens33
```

Change the IP addresses as below

master101.hadoopexam.com	192.168.239.11
node102.hadoopexam.com	192.168.239.12
node103.hadoopexam.com	192.168.239.13
node104.hadoopexam.com	192.168.239.14

Similarly change the hostname on all the 4 servers in following file. One entry on each respective nodes

```
vi /etc/sysconfig/network
```

| HOSTNAME=master101.hadoopexam.com |
| HOSTNAME=node102.hadoopexam.com |
| HOSTNAME=node103.hadoopexam.com |
| HOSTNAME=node104.hadoopexam.com |

Example , you can see on node102

```
GATEWAY=192.168.239.2
HOSTNAME=node102.hadoopexam.com
```

Now restart network service on each VM

```
service network restart
```

Step 14: Check IP config of each VM

```
ifconfig -a
```

Step 15: Now check ip address configuration on respective vm

```
Ifconfig -a
```

Step 16: Now connect from cycgwin to each node.

```
ssh root@master101
exit
ssh root@node102
exit
ssh root@node103
exit
ssh root@node104
```

Step 17: Setup nameserver, so nodes can be connected to internet. (On each node)

Now define the nameserver on virtual machine. We will use the same Gateway ip (**192.168.239.2**) as a nameserver. By running below command, it will add nameserver

```
echo 'nameserver 192.168.239.2' >> /etc/resolv.conf
```

the +i takes care that the file won't be reset on a boot even by root.

```
sudo chattr +i /etc/resolv.conf
```

Step 18: Now we need to setup ssh keys on all the vms, so that master101 can connect to other node without password.

- Login to master101
- Generate ssh keys
  ```
  ssh-keygen
  ```

- Append public key on authorized keys
  ```
  cat .ssh/id_rsa.pub >> .ssh/authorized_keys
  ```

- Now check we should have keys appended

```
cat .ssh/authorized_keys
```

- Now copy this authorized keys on all remaining 3 nodes, using scp command

```
Change permissions of .ssh folder on each node
ssh root@node102.hadoopexam.com chmod 700 ~/.ssh
ssh root@node103.hadoopexam.com chmod 700 ~/.ssh
ssh root@node104.hadoopexam.com chmod 700 ~/.ssh

Now copy the content of .ssh folder on each host (One by one)
scp /root/.ssh/* root@node102.hadoopexam.com:~/.ssh/

scp /root/.ssh/* root@node103.hadoopexam.com:~/.ssh/

scp /root/.ssh/* root@node104.hadoopexam.com:~/.ssh/
```

- Now try to login each hosts from master node

```
ssh root@node102.hadoopexam.com
ssh root@node103.hadoopexam.com
ssh root@node104.hadoopexam.com
```

Now we have 4 nodes each running with 50GB Disks and 1GB RAM

TRAINING'S (AVAILABLE)

- Hadoop BigData Professional Training
- HBase (NoSQL) Professional Training
- Apache Spark Professional Training
- Apache OOZie (Hadoop workflow) Professional Training
- Beginner AWS Training Course- **(HETRNAWS101**)
- Core Java 1z0-808 Exam training
- JAX-WS (Java WebService HandsOn Training)
- Scala Programming Training
- Python Programming Training
- Hortonworks Administration Professional Trainings

MAPR HADOOP AND NOSQL CERTIFICATION (AVAILABLE)

- **MapR Hadoop Developer Certification**
- **MapR HBase NoSQL Certification**
- **MapR Spark Developer Certification (In Progress)**

HORTONWORKS HADOOP AND NOSQL CERTIFICATION (AVAILABLE)

- HDPCD : NO Java (Hortonworks Developer Certification)
- HDPCD : Spark (Spark Developer Certifications)
- HDPCA : Hortonworks Administration Certification
- Hortonworks Administration Professional Trainings

CLOUDERA HADOOP AND SPARK CERTIFICATION (AVAILABLE)

- CCA131 : Hadoop Administrator
- CCA-175 Cloudera® (Hadoop and Spark Developer)
- CCP:DE575 : Cloudera® Data Engineer Certification
- CCA159 : Cloudera Data Analyst Certifications

DATABRICKSA OREILLY SPARK CERTIFICATION (AVAILABLE)

- Apache Spark Developer

AWS: AMAZON WEBSERVICE CERTIFICATION (AVAILABLE)

- AWS Solution Architect : Associate
- AWS Solution Architect: Professional
- AWS Developer : Associate
- AWS Sysops Admin : Associate

MICROSOFT AZURE CERTIFICATION (AVAILABLE)

- Azure 70-532
- Azure 70-533

DATA SCIENCE CERTIFICATION (AVAILABLE)

- EMC E20-007

EMC CERTIFICATIONS (AVAILABLE)

- EMC E20-007

SAS ANALYTICS CERTIFICATION (AVAILABLE)

- SAS Base A00-211
- SAS Advanced A00-212
- SAS Analytics : A00-240
- SAS Administrator : A00-250

ORACLE JAVA CERTIFICATION (AVAILABLE)

- Java 1z0-808
- Java 1z0-809
- Java 1z0-897 (Java WebService Certification)

ORACLE DATABASE CLOUD CERTIFICATION (AVAILABLE)

- 1z0-060 (Oracle 12c)
- 1z0-061 (Oracle 12c)

Subscribe Here for Regular Updates: Like New Training Module launched

---

Become Author and Trainer: We are looking for Author (Writing Technical Books) and Trainer (Creating Training Material): **No Compromise on Quality**.

Benefit: You will get very good revenue sharing. Please drop us an email to hadoopexam@gmail.com (For the skills, you feel you are master)

We are sure, you are good at least one technology. Don't limit your potential, contact us immediately with your skill. Our expert team will contact you with more detail. You training and Books will reach to all our existing network and with our expert marketing team we will help you to reach as much as technical professional, with our Smart Advertising network. Contact us with sending an email hadoopexam@gmail.com

Opportunity to share your knowledge with all learners who are in need. We are helping 1000's of learners since last 4 years and established ourselves with Quality low cost material.