**Hadoop: Setting up Multiple Node Cluster**

In the previous tutorial, we know how to set up a Hadoop single node cluster. Now, I will show how to set up a **Hadoop Multi Node Cluster.**

Prerequisites:

Master Node

Slave Node 2

Slave Node 1

Configuring single-node clusters first, here we have used two single node clusters.

If you don’t have single-node cluster VM, you can access the following link to download.

***https://drive.google.com/open?id=1WYIOJfupaq2X411KWEROJF1anAD5mfYQ***

(Paste the link on web browser to download **single\_node\_VM.zip**. User/Password: root/123456 or hadoopuser/123456)

* CentOS 7 Minimal 64bit
* Java SE Development Kit 8u161
* hadoop-2.8.1

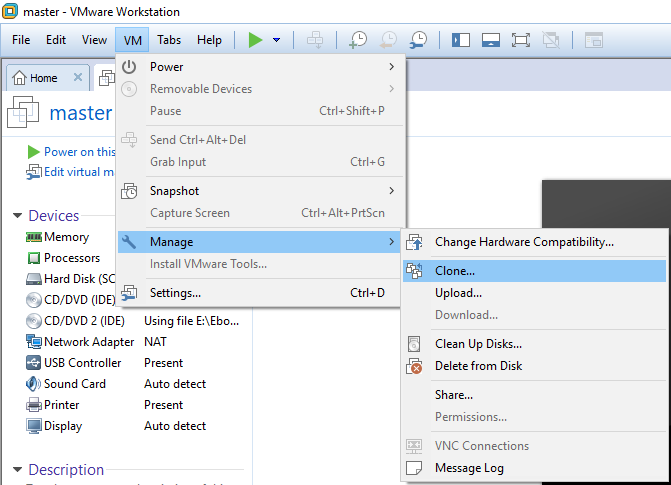
**Setup Detail**

We have two machines (master and slave) with IP:

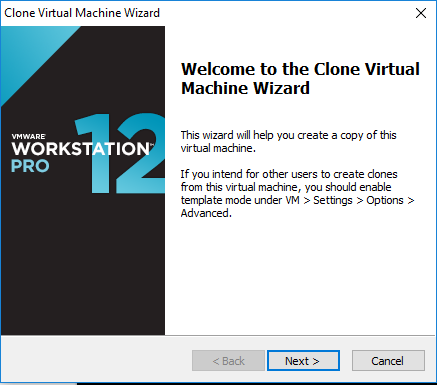
* Master IP: 192.168.201.50
* Node1: 192.168.201.51
* Node2: 192.168.201.52

1. **Clone a virtual machine**

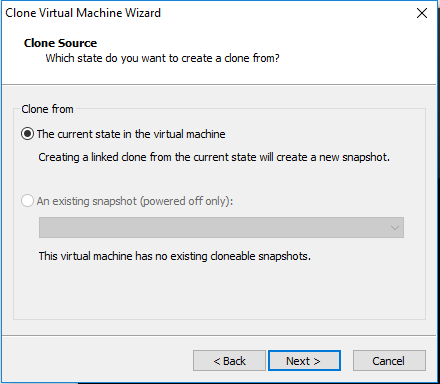
* Shutdown your virtual machine
* Select **VM -> Manage** -> **Clone…**



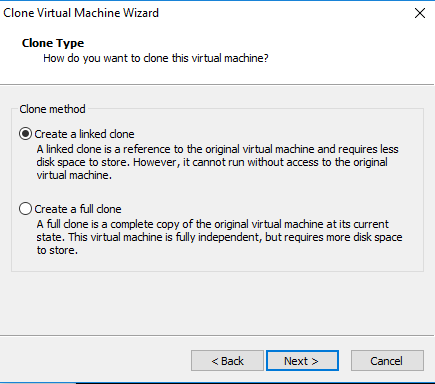
* Click **Next**



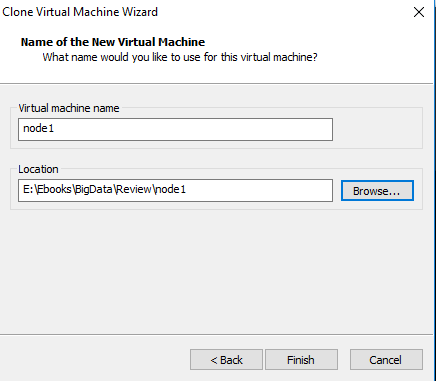
* Select “**The current state in the virtual machine**” and click **Next**



* Select “**Created a linked clone**” and “**Next**”



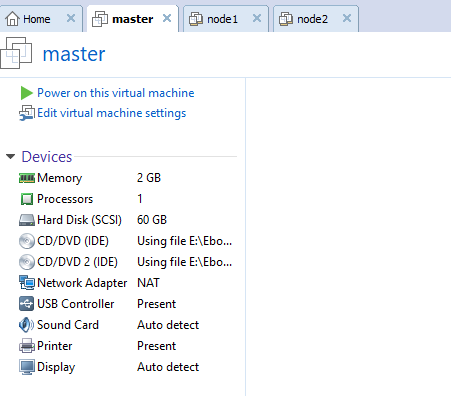
* Enter “**Virtual machine name**” and “**Location**”



* We do the same steps for node2 VM

1. **Configure IP and hosts, disable firewall**

You can make your network adapter as Bridge, NAT or Host-Only depending upon your requirement. Here I select NAT for my network type.

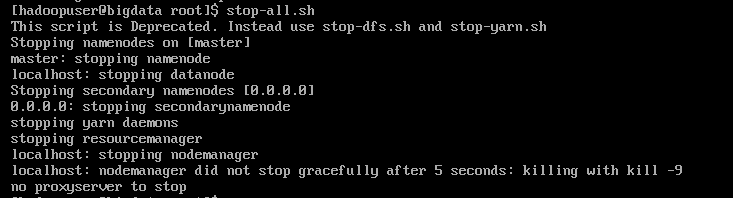


**On master,** I already set the IP to **192.168.201.50** and hostname to **master** (in single node cluster setup tutorial).

**On node1,** I will set the IP to **192.168.201.51** and hostname to **node1**.

**On node2,** I will set the IP to **192.168.201.52** and hostname to **node2**.

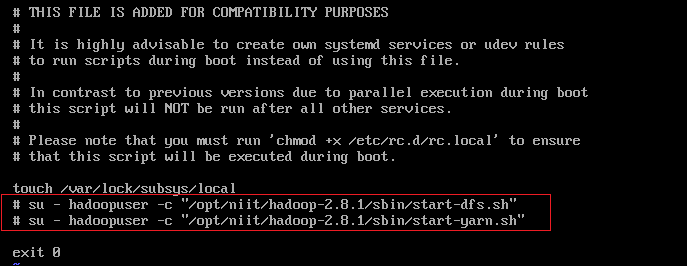
* First, start **node1 VM**, then stop all Hadoop daemons using command **stop-all.sh**



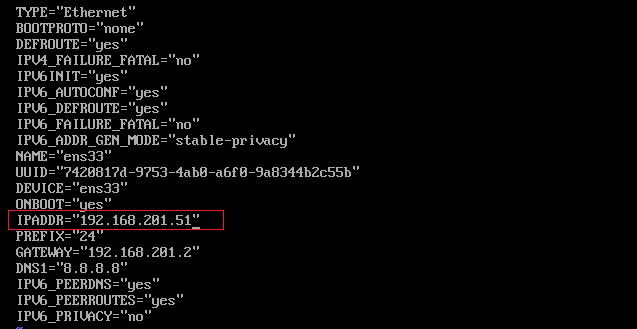
* Check Hadoop daemons again



* Comment two lines below in **/etc/rc.local**

****

* Configure IP by editing **/etc/sysconfig/network-scripts/ifcfg-ens33**

****

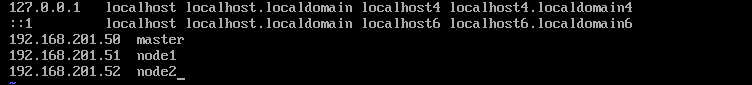
* Set hostname



* Add the following lines to **/etc/hosts**

***192.168.201.51 node1***

***192.168.201.52 node2***



* Restart network



* Disable firewall



* Check service



**And we do the same above steps on node2 VM**

**On master,** start machine, then stop all Hadoop daemons using command **stop-all.sh**



* Edit **/etc/hosts** on master. It is the same as node1 and node2

1. **Configure SSH on all nodes**

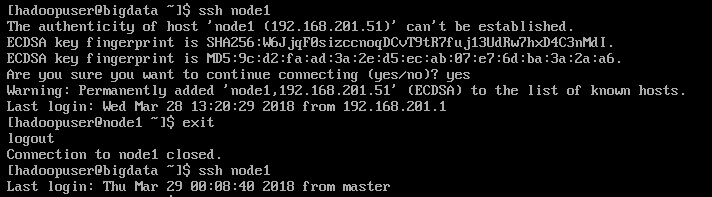
Check ssh in every node such that they can communicate with one another without any prompt for password.

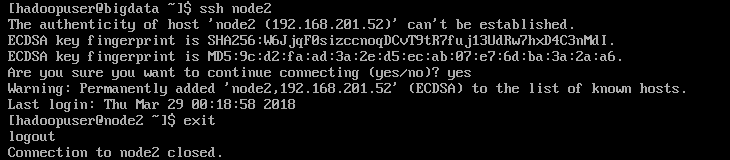
**ssh master**

**ssh node1**

**ssh node2**









1. **Edit configuration files:**

We will edit the common configuration files on master, then copy them to the slave nodes.

* 1. **mapred-site.xml**

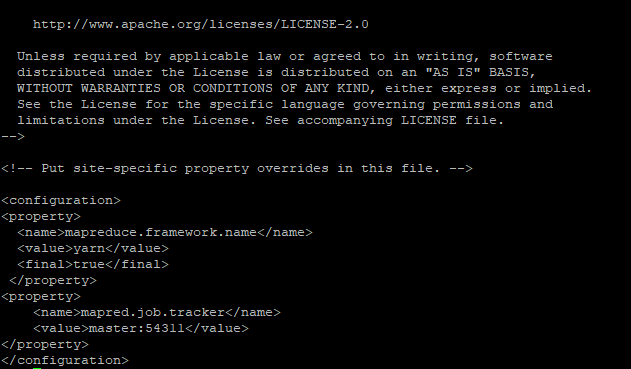
Edit configuration file mapred-site.xml (located in HADOOP\_HOME/etc/hadoop) and add following line.

*<property>*

*<name>mapred.job.tracker</name>*

*<value>master:54311</value>*

*</property>*



* 1. **hdfs-site.xml**

Update **dfs.replication** value to **3**

*<property>*

*<name>dfs.replication</name>*

*<value>****3****</value>*

*</property>*

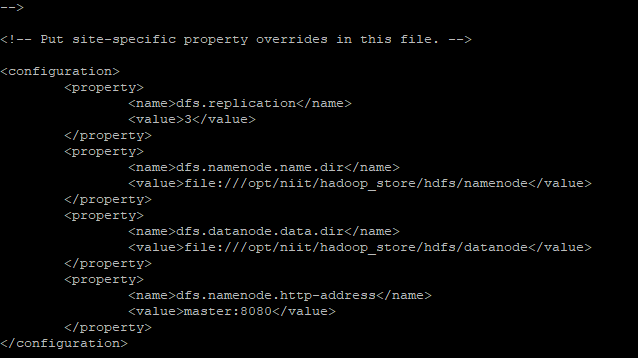
And add following line.

*<property>*

*<name>dfs.namenode.http-address</name>*

*<value>master:8080</value>*

*</property>*



Copy configuration files from master to slave nodes using the following command.

*cd $HADOOP\_HOME/etc/hadoop*

*scp \** [*hadoopuser@node1:/opt/niit/hadoop-2.8.1/etc/hadoop/*](mailto:hadoopuser@node1:/opt/niit/hadoop-2.8.1/etc/hadoop/)

*scp \* hadoopuser@node2:/opt/niit/hadoop-2.8.1/etc/hadoop/*

**On master,** edit **/opt/niit/hadoop-2.8.1/etc/hadoop/masters** as below (it will be created if it does not exist).



**On master,** edit **/opt/niit/hadoop-2.8.1/etc/hadoop/slaves** as below. Because we want master to be datanode as well, so we add master into this file.



1. **Remove existing Hadoop data folder on both master and slaves (which was created while single node hadoop setup.)**

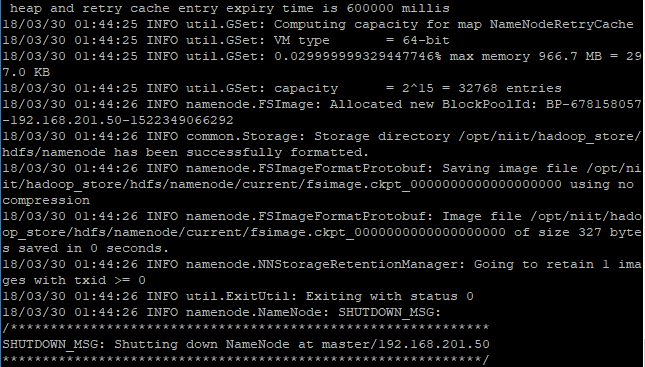
*rm -rf /opt/niit/hadoop\_store/hdfs/namenode/\**

*rm -rf /opt/niit/hadoop\_store/hdfs/datanode/\**

1. **Format namenode on master**

*hdfs namenode –format*

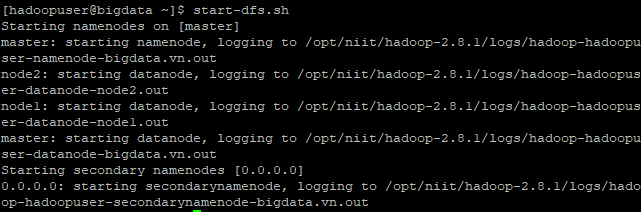




1. **Starting up Hadoop cluster daemons : (Run on MasterNode)**

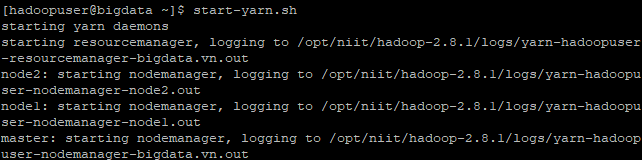
* Start HDFS daemons:

*start-dfs.sh*

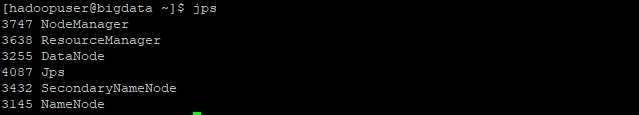


* Start Yarn daemons:

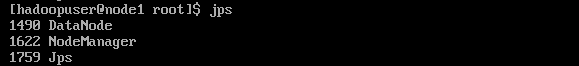
*start-yarn.sh*



* Verify Hadoop daemons on Master: *jps*

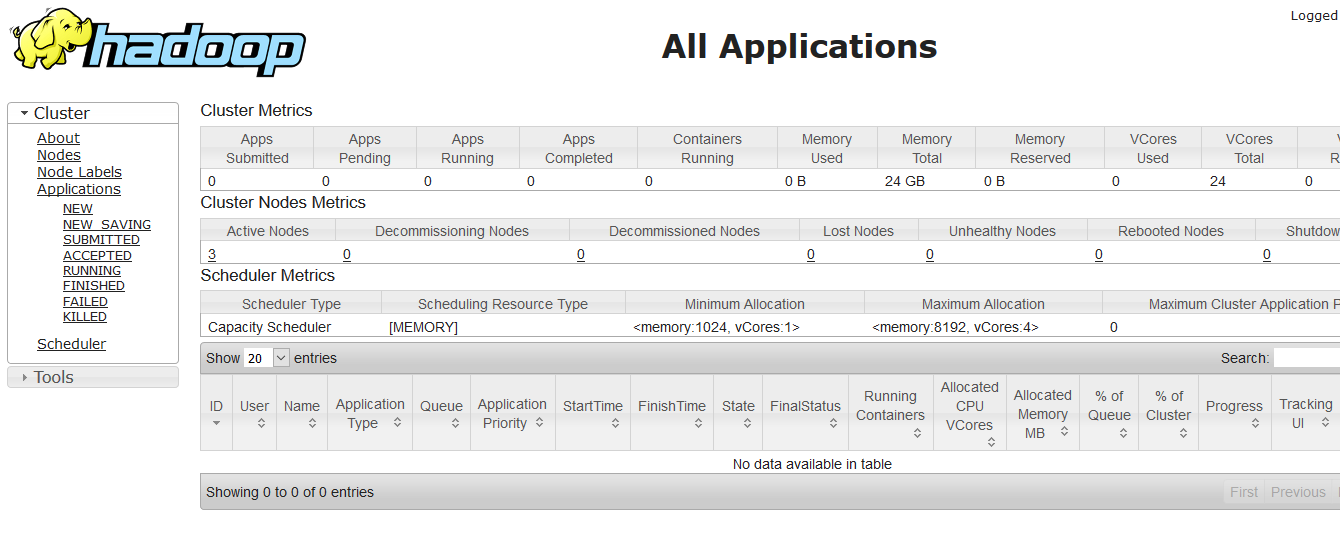


* Verify Hadoop daemons on all slave nodes: *jps*



1. **Monitor HDFS Cluster**

*For ResourceManager –* [*http://192.168.201.50:8088*](http://192.168.201.50:8088)



*For NameNode –*[*http://192.168.201.50:8080*](http://192.168.201.50:8080/)

