# HADOOP 3.2.0 MULTI NODE INSTALLATION

- sudo apt-get update && sudo apt-get -y dist-upgrade
    **OR**
- sudo -- sh -c 'apt-get update; apt-get upgrade -y; apt-get dist-upgrade -y; apt-get autoremove -y; apt-get autoclean -y'


- sudo apt-get install openjdk-8-jdk
- java -version


- sudo addgroup hadoop
- sudo adduser --ingroup hadoop hduser
- sudo adduser hduser sudo
- sudo apt-get install ssh

 **After cloning:**

**for all nodes:**
**1-** Change host name:
- sudo gedit /etc/hostname
Type ---> master, etc.

**2-** Change hosts:
- sudo gedit /etc/hosts
Type --->
192.168.56.10 master
192.168.56.11 slave1
192.168.56.12 slave2
192.168.56.6 dataStore

**3-** Make the Ip stable:
- sudo gedit /etc/network/interfaces


        ------------
# interfaces(5) file used by ifup(8) and ifdown(8)
auto lo
iface lo inet loopback

auto enp0s8
iface enp0s8 inet static
address 192.168.56.10
netmask 255.255.255.0
------------


**4-** sudo reboot
**5-** check them ---> ifconfig

**Then for SSH:**
- sudo su - hduser
-  sudo ufw disable
- ssh-keygen -t rsa -P "" -f ~/.ssh/id_rsa
- cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
- chmod 700 ~/.ssh/id_rsa
- ssh-copy-id -i /home/hduser/.ssh/id_rsa.pub hduser@192.168.56.10
- ssh-copy-id -i /home/hduser/.ssh/id_rsa.pub hduser@192.168.56.11
- ssh-copy-id -i /home/hduser/.ssh/id_rsa.pub hduser@192.168.56.12
- ssh-copy-id -i /home/hduser/.ssh/id_rsa.pub hduser@192.168.56.13
-  ssh-copy-id -i /home/hduser/.ssh/id_rsa.pub umit@192.168.56.6

Then try to ping each other  --->   ping 192.168.56.6
And try SSH                              --->   ssh hduser@192.168.56.6
                                                         OR
                                              ssh hduser@dataStore


**Download Hadoop**
- wget https://archive.apache.org/dist/hadoop/core/hadoop-3.2.0/hadoop-3.2.0.tar.gz -P ~/Downloads
- wget https://archive.apache.org/dist/hadoop/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz -P ~/Downloads
- sudo tar zxvf ~/Downloads/hadoop-* -C /usr/local
- sudo mv /usr/local/hadoop-* /usr/local/hadoop
- sudo chown -R hduser:hadoop /usr/local/hadoop

- sudo gedit ~/.bashrc
--------------
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PATH=$PATH:$JAVA_HOME/bin
export HADOOP_HOME=/usr/local/hadoop
export PATH=$PATH:$HADOOP_HOME/bin
export HADOOP_CONF_DIR=/usr/local/hadoop/etc/hadoop
export HADOOP_MAPRED_HOME=${HADOOP_HOME}
export HADOOP_COMMON_HOME=${HADOOP_HOME}
export HADOOP_HDFS_HOME=${HADOOP_HOME}
export YARN_HOME=${HADOOP_HOME}

--------------
- source ~/.bashrc

        **on ALL HADOOP NODES**

- cd $HADOOP_CONF_DIR
- sudo gedit hadoop-env.sh

**Replace this line:**

export JAVA_HOME=${JAVA_HOME}
**with the following line:**
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

## on ALL HADOOP NODES
- **sudo gedit core-site.xml**

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://master:9000</value>
  </property>
</configuration>
```

## ONLY namenode: Create namenode dir
- sudo rm -rf /usr/local/hadoop/hdfs/namenodedata
- sudo mkdir -p /usr/local/hadoop/hdfs/namenodedata
- sudo chown -R hduser:hadoop /usr/local/hadoop/hdfs/namenodedata
- sudo chmod 700 -R '/usr/local/hadoop/hdfs/namenodedata'

## All Nodes: Create Data dir (do this on namenode if it is also a datanode. Otherwise do this for only datanodes)
- sudo rm -rf /usr/local/hadoop/hdfs/datanodedata
- sudo mkdir -p /usr/local/hadoop/hdfs/datanodedata
- sudo chown -R hduser:hadoop /usr/local/hadoop/hdfs/datanodedata
- sudo chmod 700 -R '/usr/local/hadoop/hdfs/datanodedata'

## Namenode: Setup HDFS Properties
- **sudo gedit hdfs-site.xml**

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>3</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///usr/local/hadoop/hdfs/namenodedata</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
```

```
            <value>file:///usr/local/hadoop/hdfs/datanodedata</value>
        </property>
</configuration>
```

## Namenode: Setup MapReduce Properties ( with JobHistoryServer )
- **sudo gedit mapred-site.xml**

```
<configuration>
  <property>
        <name>mapreduce.jobtracker.address</name>
        <value>master:54311</value>
  </property>
  <property>
        <name>mapreduce.framework.name</name>
        <value>yarn</value>
  </property>
  <property>
        <name>mapreduce.jobhistory.address</name>
        <value>master:10020</value>
  </property>
  <property>
        <name>mapreduce.jobhistory.webapp.address</name>
        <value>master:19888</value>
  </property>
  <property>
        <name>yarn.app.mapreduce.am.env</name>
        <value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
  </property>
  <property>
        <name>mapreduce.map.env</name>
        <value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
  </property>
  <property>
        <name>mapreduce.map.speculative</name>
        <value>true</value>
  </property>
  <property>
        <name>mapreduce.reduce.speculative</name>
        <value>false</value>
  </property>
  <property>
        <name>mapreduce.reduce.env</name>
        <value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
  </property>
  <property>
        <name>mapreduce.job.reduce.slowstart.completedmaps</name>
        <value>0.99</value>
```

```
    </property>
</configuration>
```

- **sudo gedit yarn-site.xml**

```
<configuration>
    <property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
    </property>
    <property>
        <name>yarn.nodemanager.aux-services.mapreduce_shuffle.class</name>
        <value>org.apache.hadoop.mapred.ShuffleHandler</value>
    </property>
    <property>
        <name>yarn.resourcemanager.hostname</name>
        <value>master</value>
    </property>
    <property>
        <name>yarn.resourcemanager.resource-tracker.address</name>
        <value>master:8025</value>
    </property>
    <property>
        <name>yarn.resourcemanager.scheduler.address</name>
        <value>master:8030</value>
    </property>
    <property>
        <name>yarn.resourcemanager.address</name>
        <value>master:8050</value>
    </property>
    <property>
        <name>yarn.log-aggregation-enable</name>
        <value>true</value>
    </property>
    <property>
        <name>yarn.resourcemanager.scheduler.class</name>
        <value>org.apache.hadoop.yarn.server.resourcemanager.scheduler.fair.FairScheduler</value>
    </property>
</configuration>
```

**Old one:**
```
<configuration>
<!— Site specific YARN configuration properties -->
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
```

```
</property>
<property>
<name>yarn.resourcemanager.hostname</name>
<value>namenode</value>
</property>
</configuration>
```

## Namenode: Setup Master and Slaves

- cd $HADOOP_CONF_DIR
- **sudo cp workers masters**
- **sudo gedit masters**

**Delete localhost and add only the line below:**
master

- **sudo gedit workers**

**Delete localhost and add only the line below: (add only datanode's Private IPs)**
master
slave1
slave2

**Configuring Data Nodes (SLAVES) only** (don't include namenode even if it is also a datanode, as we did this above**)**

## datanode: Setup HDFS Properties

- **sudo gedit hdfs-site.xml**

```
<configuration>
  <property>
   <name>dfs.replication</name>
   <value>3</value>
  </property>
  <property>
   <name>dfs.datanode.data.dir</name>
   <value>file:///usr/local/hadoop/hdfs/datanodedata</value>
  </property>
</configuration>
```

**Starting the Cluster**
on namenode:

- $HADOOP_HOME/bin/hdfs namenode -format
- $HADOOP_HOME/sbin/start-dfs.sh
- $HADOOP_HOME/sbin/start-yarn.sh
- $HADOOP_HOME/sbin/mr-jobhistory-daemon.sh start historyserver
- hdfs dfsadmin -safemode leave
- hdfs dfsadmin -report

**STARTING:**
sudo su - hduser  ----> qwe

sudo ufw disable

$HADOOP_HOME/sbin/start-dfs.sh && $HADOOP_HOME/sbin/start-yarn.sh && $HADOOP_HOME/sbin/mr-jobhistory-daemon.sh start historyserver && hdfs dfsadmin -safemode leave && hdfs dfsadmin -report

**STOPPING:**
$HADOOP_HOME/sbin/stop-dfs.sh && $HADOOP_HOME/sbin/stop-yarn.sh && $HADOOP_HOME/sbin/mr-jobhistory-daemon.sh stop historyserver

**STOP:**
- $HADOOP_HOME/sbin/stop-dfs.sh
- $HADOOP_HOME/sbin/stop-yarn.sh

**ls problem--------------**

hadoop fs -mkdir /user
hadoop fs -mkdir /user/hduser
hadoop fs -ls