HADOOP:

MULTI-NODE CLUSTER

TABLE OF CONTENTS

SETUP DETAILS:

Create 4 separate machines i.e., 1master and 3slaves with defined IP addresses
master 192.168.10.10
slave1 192.168.10.11
slave2 192.168.10.12
slave3 192.168.10.13

STEP 1: INSTALL JDK7

Before installing hadoop make sure you have java installed on all nodes of hadoop cluster systems.

Download JDK7 for Linux-x64 from official Oracle site.
```
[root@master]# cd ~/Download
[root@master]# yum localinstall jdk-7u80-linux-x64.rpm
[root@master]# alternatives --install /usr/bin/java java /usr/java/jdk1.7.0_80/bin/java 210000
```

To check java version and also alternatives
```
[root@master]# java -version
[root@master]# alternatives --display java
```

This is need to done all the 4 machines.

STEP 2: CREATE USER ACCOUNT

Create a system user account on both master and slave systems to use for hadoop installation

```
[root@master]# useradd huser
[root@master]# passwd huser
```

STEP 3: ADD FQDN MAPPING

Edit /etc/hosts file on master and slave machines and add following entries.
```
[root@master]# gedit /etc/hosts
```

Append the following lines at the end of the file:
```
192.168.10.10 master
192.168.10.11 slave1
192.168.10.12 slave2
192.168.10.13 slave3
```

STEP 4: CONFIGURING KEY BASED LOGIN

It's required to set up hadoop user to ssh itself without password. Use following commands to configure auto login between all hadoop cluster servers..

```
[root@master]# su - huser
[root@huser]$ ssh-keygen
[root@huser]$ ssh-copy-id -i ~/.ssh/id_rsa.pub huser@192.168.10.10
[root@huser]$ ssh-copy-id -i ~/.ssh/id_rsa.pub huser@192.168.10.11
[root@huser]$ ssh-copy-id -i ~/.ssh/id_rsa.pub huser@192.168.10.12
[root@huser]$ ssh-copy-id -i ~/.ssh/id_rsa.pub huser@192.168.10.13
[root@huser]$ chmod 0600 ~/.ssh/authorized_keys
[root@huser]$ exit
```

To avoid typing password for each time we login:
```
[root@master]# gedit /etc/ssh/ssh_config
```
And search for "StrickHostKeyChecking"
Remove "#" and make it like this "StrickHostKeyChecking no" without double quote and save it.

STEP 5: DOWNLOAD AND EXTRACT HADOOP SOURCE

Download hadoop 1.2.1 version from its official site at huser-master server only.

```
[root@master]# cd ~/Downloads
[root@master]# weget http://apache.javapipe.com/hadoop/common/hadoop-1.2.1/hadoop-1.2.1.tar.gz
[root@master]# mkdir /opt/hadoop
[root@master]# cp ~/Downloads/hadoop-1.2.1.tar.gz /opt/hadoop
[root@master]# cd /opt/hadoop/
[root@master]# tar -xzf hadoop-1.2.1.tar.gz
[root@master]# chown -R huser /opt/hadoop
[root@master]# cd /opt/hadoop/hadoop-1.2.1/
```

STEP 6: CONFIGURE HADOOP

First edit hadoop configuration files and make following changes.

```
[root@master]# ls –ls
```

6.1 Edit core-site.xml

```
[root@master]# gedit conf/core-site.xml
```
Add the following inside the <configuration> tag

```
<property>
   <name>fs.default.name</name>
   <value>hdfs:// master:9000/</value>
</property>
<property>
   <name>dfs.permissions</name>
   <value>false</value>
</property>
```

6.2 Edit hdfs-site.xml

```
[root@master]# vim conf/hdfs-site.xml
```
Add the following inside the <configuration> tag

```
<property>
      <name>dfs.data.dir</name>
      <value>/opt/hadoop/hadoop/dfs/name/data</value>
      <final>true</final>
</property>
<property>
      <name>dfs.name.dir</name>
      <value>/opt/hadoop/hadoop/dfs/name</value>
      <final>true</final>
</property>
<property>
      <name>dfs.replication</name>
      <value>3</value>
</property>
```

6.3 Edit mapred-site.xml

```
[root@master]# gedit conf/mapred-site.xml
```

Add the following inside the <configuration> tag

```
<property>
    <name>mapred.job.tracker</name>
        <value>master:9001</value>
</property>
```

6.4 Edit hadoop-env.sh

```
[root@master]# gedit conf/hadoop-env.sh
```

Append the following lines at the end of the file:

```
export JAVA_HOME=/usr/java/jdk1.7.0_80
export HADOOP_OPTS=-Djava.net.preferIPv4Stack=true
export HADOOP_CONF_DIR=/opt/hadoop/hadoop-1.2.1/conf
```

STEP 7: COPY HADOOP SOURCE TO SLAVE SERVERS

After updating above configuration, we need to copy the source files to all slave servers.

```
[root@master]# scp -rp /opt/hadoop slave1:/opt/
[root@master]# scp -rp /opt/hadoop slave2:/opt/
[root@master]# scp -rp /opt/hadoop slave3:/opt/
```

STEP 8: CONFIGURE HADOOP ON MASTER SERVER ONLY

Go to hadoop source folder on huser-master and do following settings.

```
[root@master]# su - huser
[root@huser]$ cd /opt/hadoop/hadoop-1.2.1/
[root@huser]$ gedit conf/masters
```

And this line:

```
master
```

```
[root@huser]$ gedit conf/slaves
```

Add this lines:

```
slave1
slave2
slave3
```

STEP 9: SETTING UP THE ENVIRONMENT FOR JAVA AND HADOOP

We need to source the environment files

```
[root@master]# su - huser
[root@huser]$ gedit ~/.bash_profile
```
Append the following lines at the end of the file:

```
## JAVA env variables
export JAVA_HOME=/usr/java/jdk1.7.0_80
export PATH=$PATH:$JAVA_HOME/bin
export CLASSPATH=.:$JAVA_HOME/jre/lib:$JAVA_HOME/lib:$JAVA_HOME/lib/tools.jar

## HADOOP env variables
export HADOOP_HOME=/opt/hadoop
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_YARN_HOME=$HADOOP_HOME
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
```

```
[root@huser]$ source ~/.bash_profile
[root@huser]$ echo $HADOOP_HOME
[root@huser]$ echo $JAVA_HOME
[root@huser]$ exit
```

SCP the ~/.bash_profile to other slave machines
```
[root@master]# scp -rp /root/huser/.bash_profile slave1:~/
[root@master]# ssh slave1
[root@slave1]$ source ~/.bash_profile
[root@slave1]$ exit
```

```
[root@master]# scp -rp /root/huser/.bash_profile slave2:~/
[root@master]# ssh slave1
[root@slave2]$ source ~/.bash_profile
[root@slave2]$ exit
```

```
[root@master]# scp -rp /root/huser/.bash_profile slave3:~/
[root@master]# ssh slave1
[root@slave3]$ source ~/.bash_profile
[root@slave3]$ exit
```

STEP 10: FORMAT THE NODE

Format Name Node on Hadoop Master only

```
[root@master]# su - huser
[root@huser]$ cd /opt/hadoop/hadoop-1.2.1/
[root@huser]$ bin/hadoop namenode –format
```

STEP 11: START HADOOP SERVICES

Use the following command to start all hadoop services on huser-master

```
[root@huser]$ bin/start-all.sh
```

STEP 12: CHECK RUNNING SERVICES

```
[root@huser]$ jps
```

Open browser and type on address bar "master:50070" without double quote
And u can see 3 live nodes