***Steps***

1. **Prepare your computer network:**
   1. - Decide the number of nodes to set up in the cluster

* This example uses a cluster with 4 Datanodes + 1 Namenode
  1. - Proceed with a fresh new install of Ubuntu Desktop on each Machine and give name to each Machine
* This example uses Ubuntu 16.04 ‘Xenia Xerus’ distribution
* Our machine are 5 ‘OptiPlex 760’ PC Machines named HadoopMaster, HadoopSlave1, HadoopSlave2, HadoopSlave3, HadoopSlave4 and a common username and password for each machine. In this example username is ‘swl’.
  1. - For each Machine follow the next steps:
     1. **Installing Oracle’s Java 8**

Hadoop is a java framework and need the program to run. Hadoop supports all java version from 1.5 on but 1.8 is the latest stable version to date.

On HadoopMaster:

swl@HadoopMaster:~$ sudo add-apt-repository ppa:webupd8team/java

swl@HadoopMaster:~$ sudo apt-get update

swl@HadoopMaster:~$ sudo apt-get install oracle-java8-installer

On HadoopSlave :

swl@HadoopSlave1:~$ sudo add-apt-repository ppa:webupd8team/java

swl@HadoopSlave1:~$ sudo apt-get update

swl@HadoopSlave1:~$ sudo apt-get install oracle-java8-installer

On HadoopSlave2:

swl@HadoopSlave2:~$ sudo add-apt-repository ppa:webupd8team/java

swl@HadoopSlave2:~$ sudo apt-get update

swl@HadoopSlave2:~$ sudo apt-get install oracle-java8-installer

On HadoopSlave3:

swl@HadoopSlave3:~$ sudo add-apt-repository ppa:webupd8team/java

swl@HadoopSlave3:~$ sudo apt-get update

swl@HadoopSlave3:~$ sudo apt-get install oracle-java8-installer

On HadoopSlave4:

swl@HadoopSlave4:~$ sudo add-apt-repository ppa:webupd8team/java

swl@HadoopSlave4:~$ sudo apt-get update

swl@HadoopSlave4:~$ sudo apt-get install oracle-java8-installer

These will install java source in your machine at /usr/lib/jvm/java-8-oracle. Then verify your java installation using the following command:

swl@HadoopMaster:~$ java -version

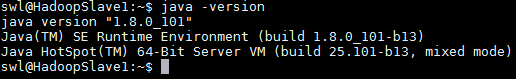
swl@HadoopSlave1:~$ java -version

swl@HadoopSlave2:~$ java -version

swl@HadoopSlave3:~$ java –version

swl@HadoopSlave4:~$ java –version

You should get something like the following output on each machine:



* + 1. **Creating a Hadoop user**

We will create hadoop as system group and hduser as system user on each machine, the same user, and add this user to sudo list. Then reboot machines.

In this example group is called ‘hadoop’ and user is ‘hduser’.

On HadoopMaster:

swl@HadoopMaster:~$ sudo addgroup hadoop

swl@HadoopMaster:~$ sudo adduser --ingroup hadoop hduser

swl@HadoopMaster:~$ sudo adduser hduser sudo

swl@HadoopMaster:~$ sudo reboot

On HadoopSlave :

swl@HadoopSlave1:~$ sudo addgroup hadoop

swl@HadoopSlave1:~$ sudo adduser --ingroup hadoop hduser

swl@HadoopSlave1:~$ sudo adduser hduser sudo

swl@ HadoopSlave1r:~$ sudo reboot

On HadoopSlave2:

swl@HadoopSlave2:~$ sudo addgroup hadoop

swl@HadoopSlave2:~$ sudo adduser --ingroup hadoop hduser

swl@HadoopSlave2:~$ sudo adduser hduser sudo

swl@ HadoopSlave2:~$ sudo reboot

On HadoopSlave3:

swl@HadoopSlave3:~$ sudo addgroup hadoop

swl@HadoopSlave3:~$ sudo adduser --ingroup hadoop hduser

swl@HadoopSlave3:~$ sudo adduser hduser sudo

swl@ HadoopSlave3:~$ sudo reboot

On HadoopSlave4:

swl@HadoopSlave4:~$ sudo addgroup hadoop

swl@HadoopSlave4:~$ sudo adduser --ingroup hadoop hduser

swl@HadoopSlave4:~$ sudo adduser hduser sudo

swl@ HadoopSlave4:~$ sudo reboot

* + 1. **Installing and Configuring SSH**

SSH (“Secure Shell”) is a protocol for securely accessing one machine from another using a protocol called SSH. Hadoop uses SSH for accessing another slaves nodes to start and manage all HDFS and MapReduce daemons.

On HadoopMaster:

swl@HadoopMaster:~$ sudo su hduser

hduser@HadoopMaster/:/home/swl $ cd ~

hduser@HadoopMaster:~$ sudo apt-get install openssh-server

hduser@HadoopMaster:~$ ssh-keygen -t rsa -P ""

hduser@HadoopMaster:~$ cat $HOME/.ssh/id\_rsa.pub >> $HOME/.ssh/authorized\_keys

On HadoopSlave :

swl@HadoopSlave1:~$ sudo su hduser

hduser@HadoopMaster/:/home/swl $ cd ~

hduser @HadoopSlave1:~$ sudo apt-get install openssh-server

hduser @HadoopSlave1:~$ ssh-keygen -t rsa -P ""

hduser @ HadoopSlave1r:~$ cat $HOME/.ssh/id\_rsa.pub >> $HOME/.ssh/authorized\_keys

On HadoopSlave2:

swl@HadoopSlave2:~$ sudo su hduser

hduser@HadoopMaster/:/home/swl $ cd ~

hduser @HadoopSlave2:~$ sudo apt-get install openssh-server

hduser @HadoopSlave2:~$ ssh-keygen -t rsa -P ""

hduser @ HadoopSlave2:~$ cat $HOME/.ssh/id\_rsa.pub >> $HOME/.ssh/authorized\_keys

On HadoopSlave3:

swl@HadoopSlave3:~$ sudo su hduser

hduser@HadoopMaster/:/home/swl $ cd ~

hduser @HadoopSlave3:~$ sudo apt-get install openssh-server

hduser @HadoopSlave3:~$ ssh-keygen -t rsa -P ""

hduser @ HadoopSlave3:~$ cat $HOME/.ssh/id\_rsa.pub >> $HOME/.ssh/authorized\_keys

On HadoopSlave4:

swl@HadoopSlave4:~$ sudo su hduser

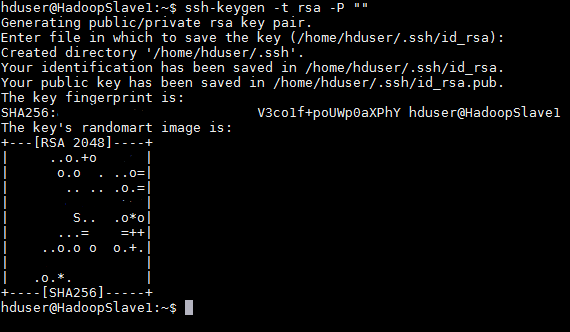
hduser@HadoopMaster/:/home/swl $ cd ~

hduser @HadoopSlave4:~$ sudo apt-get install openssh-server

hduser @HadoopSlave4:~$ ssh-keygen -t rsa -P ""

hduser @ HadoopSlave4:~$ cat $HOME/.ssh/id\_rsa.pub >> $HOME/.ssh/authorized\_keys

The first command logs you as hduser. You might be prompted to give password. Enter the password given at the creation of the hduser user. The second changes current directory to home of hduser. The third command install ssh and the fourth generates a ssh key for hduser account. In this step you might be prompted for a password. Leave it blank. An id\_rsa.pub is created with following type of output:



The fifth command copies the id\_rsa.pub generated to the authorized keys away of hduser.

* + 1. **Installing rsync**

Installing rsync permits the machines to share Hadoop source between all machines.

hduser@HadoopMaster:~$ sudo apt-get install rsync

hduser@HadoopSlave1:~$ sudo apt-get install rsync

hduser@HadoopSlave2:~$ sudo apt-get install rsync

hduser@HadoopSlave3:~$ sudo apt-get install rsync

hduser@HadoopSlave4:~$ sudo apt-get install rsync

* + 1. **Disabling IPv6**

For each machine, first run command:

hduser@HadoopSomething:~$ sudo nano /etc/sysctl.conf

An editable file opens within the CLI (command Line Interface) go to the end of the document then add these lines (you can copy paste):

net.ipv6.conf.all.disable\_ipv6 = 1

net.ipv6.conf.default.disable\_ipv6 = 1

net.ipv6.conf.lo.disable\_ipv6 = 1

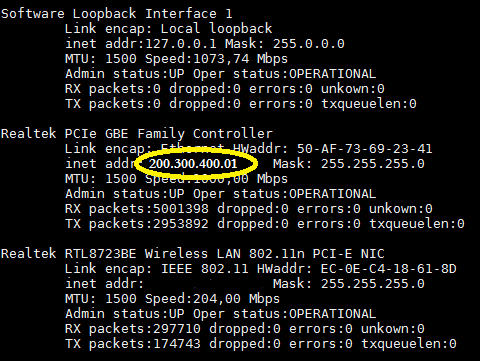
Then press **crtl-x**, then **y**, then **enter**. You have re-written the file.

* + 1. **The Network**

Prepare now the computer network, and make sure all your machines are plugged in to the internet. Get a pencil and a paper. Once you are done type the following command on each machine:

hduser@HadoopMaster:~$ ifconfig

Your output should look something like this:



Make sure you have your ip address, from the right network feed. Here look I have cable and wifi and I use cable. Once you do this for every machine, make a list as follows:

200.300.400.01 HadoopMaster

200.300.400.02 HadoopSlave1

200.300.400.03 HadoopSlave2

200.300.400.04 HadoopSlave3

200.300.400.05 HadoopSlave4

Once this list generated, for each machine, first run command:

hduser@HadoopSomething:~$ sudo nano /etc/hosts

An editable file opens within the CLI (command Line Interface) go to the end of the document then add the list above (you can copy paste).

Then press **crtl-x**, then **y**, then **enter**. You have re-written the file.

Now reboot all machines using:

hduser@HadoopSomething:~$ sudo reboot

1. **Hadoop Configuration Preparations**

**2.1 - Just on the MasterNode**

swl@HadoopMaster:~$ sudo su hduser

hduser@HadoopMaster: $ cd /usr/local

These lines permit us to change the user to hduser and go to/usr/local directory. Then:

hduser@HadoopMaster /usr/local: $ sudo wget ‘http://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-2.6.4/hadoop-2.6.4.tar.gz’

This command gets the tar ball of the lastest (2.6.4 version) Apache Hadoop Binary, and downloads it to the current folder.

hduser@HadoopMaster: /usr/local$ sudo tar -xzvf hadoop-2.6.0.tar.gz

hduser@HadoopMaster: /usr/local$ sudo mkdir hadoop

hduser@HadoopMaster: /usr/local$ sudo mv hadoop-2.6.4/\* /usr/local/hadoop

hduser@HadoopMaster: /usr/local$ sudo chown hduser:hadoop -R /usr/local/hadoop

The first command untars the tar ball into a file called ‘hadoop-2.6.4’. The second creates a directory called ‘hadoop’ to which the third command moves all the files in the tar from hadoop-2.6.4 directory. The fourth command assigns the ownership of the ‘hadoop’ directory to user ‘hduser’.

Then create Hadoop temp directories for Namenode and Datanode with the two following commands:

hduser@HadoopMaster: /usr/local$ sudo mkdir -p /hadoop\_tmp/hdfs/namenode

hduser@HadoopMaster: /usr/local$ sudo mkdir -p /hadoop\_tmp/hdfs/datanode

hduser@HadoopMaster: /usr/local$ sudo chown hduser:hadoop -R /hadoop\_tmp/

The last command assigns the ownership of these files to ‘hduser’ user. Then one needs to update some configuration files.

hduser@HadoopMaster: /usr/local$ sudo nano .bashrc

An editable file opens within the CLI (command Line Interface) go to the end of the document then add the following lines (you can copy paste):

export JAVA\_HOME=/usr/lib/jvm/java-8-oracle

export HADOOP\_HOME=/usr/local/hadoop

export PATH=$PATH:$HADOOP\_HOME/bin

export PATH=$PATH:$HADOOP\_HOME/sbin

export HADOOP\_MAPRED\_HOME=$HADOOP\_HOME

export HADOOP\_COMMON\_HOME=$HADOOP\_HOME

export HADOOP\_HDFS\_HOME=$HADOOP\_HOME

export YARN\_HOME=$HADOOP\_HOME

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native

export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_HOME/lib"

Then press **crtl-x**, then **y**, then **enter**. You have re-written the file.

Now reboot all machines using:

hduser@HadoopMaster: /usr/local$ cd /usr/local/hadoop/etc/hadoop

hduser@HadoopMaster:/usr/local/hadoop/etc/hadoop$ sudo nano hadoop-env.sh

An editable file opens within the CLI (command Line Interface) look in the document for this line and you can update it to:

JAVA\_HOME=/usr/lib/jvm/java-8-oracle

Then press **crtl-x**, then **y**, then **enter**. You have re-written the file.

Now reboot this machine with:

hduser@HadoopMaster:/usr/local/hadoop/etc/hadoop$ sudo reboot

1. **Hadoop Configuration**

cd /usr/local

sudo mkdir hadoop

hduser@HadoopMaster: /usr/local$ sudo chown hduser:hadoop -R /usr/local/hadoop