ACADGILD

Installation Manual for Single Node Hadoop Cluster

Satyam Kumar
[Pick the date]

1. Table of Contents

1. Introduction 2
2. Procedure 2

1. Introduction

This is a manual to help you install VMware and Ubuntu 12.0.4 over it.

2. Procedure

Step 1: Download VMware (VM).

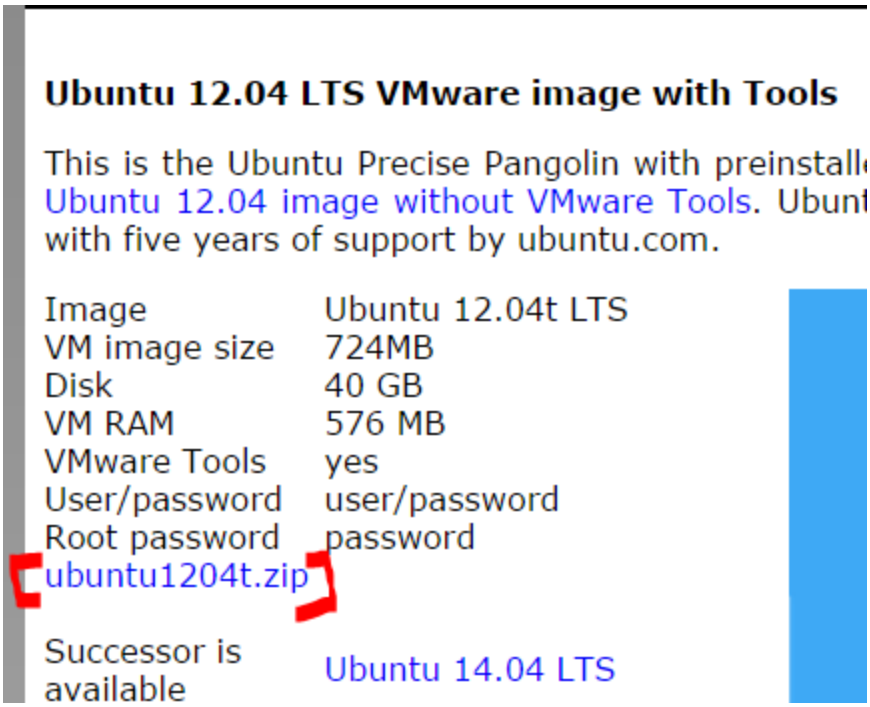
Refer the below link for the same:

Link: https://www.VMware.com/tryVMware/?p=player

Select the **VMware Player 5.0.4 for Windows 32-bit and 64-bit** for windows from the prompt as shown below:

Step 2: Download the image of Ubuntu from the below link, refer the below screenshot for more details.

Link: http://www.traffictool.net/VMware/ubuntu1204t.html

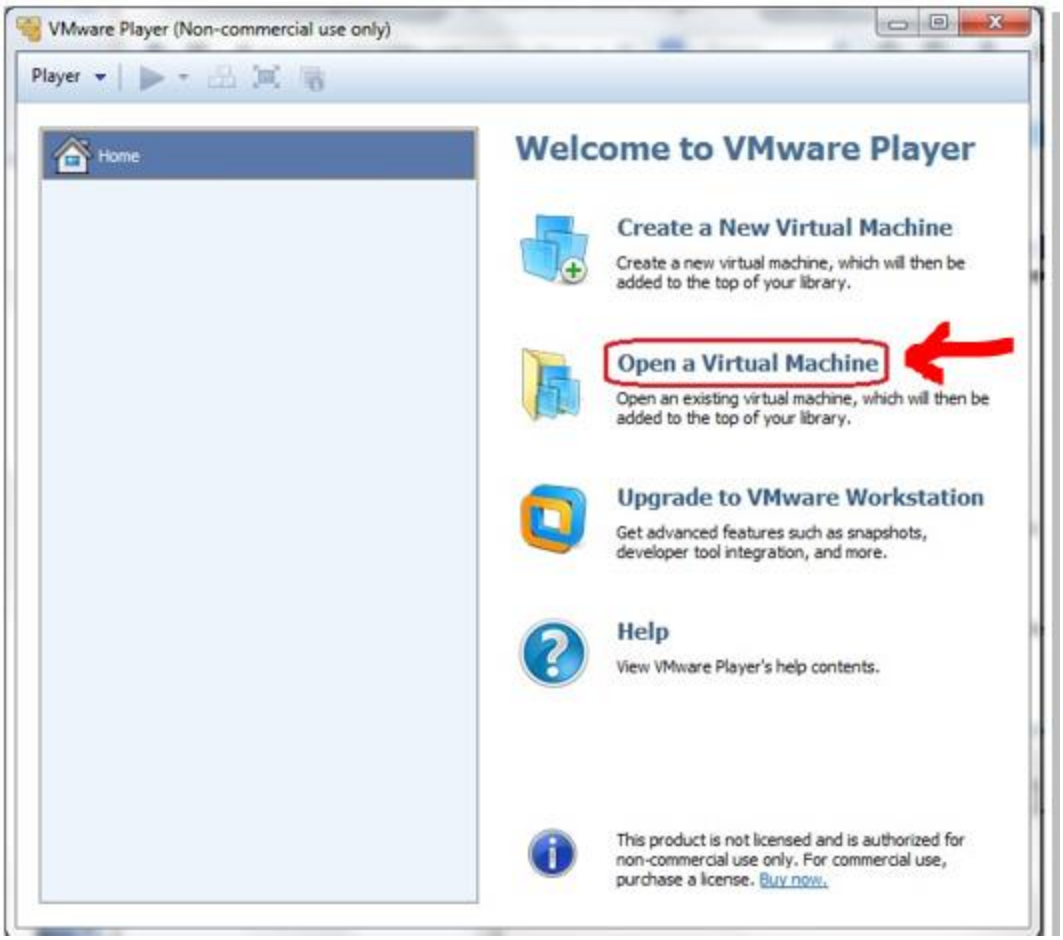


The above downloaded file is a zip file. Refer the below links to download the WinRAR to unzip the Ubuntu zipped image.

Link for windows 32 bit: http://download.cnet.com/WinRAR-32-bit/30002250_4-10007677.html

Link for windows 64 bit: http://download.cnet.com/WinRAR-64-bit/3000-2250_4-10965579.html

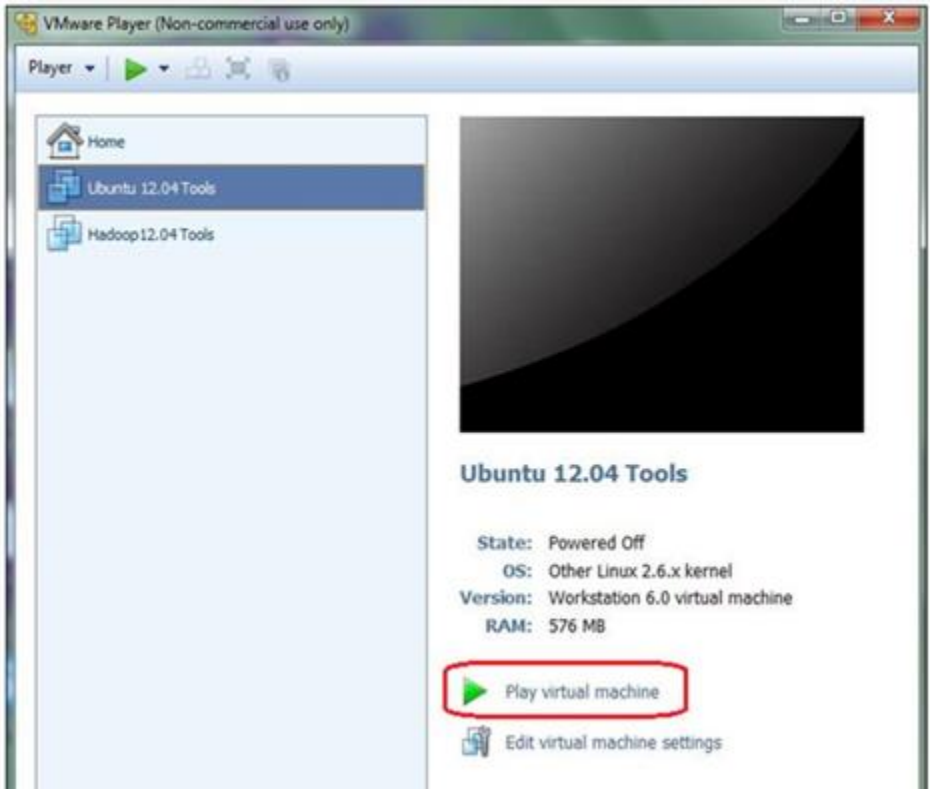
Step 3: Save the unzipped Ubuntu image in hard disk of your system and then open the VMware and then click on the option, **Open a Virtual Machine.**



Step 4: Browse to the location of your hard disk of the system and select the unzipped Ubuntu file.

On browsing the location of the unzipped Ubuntu you will get the file as shown in the below screenshot.

Step 5: Select the option **Open** shown in the above screenshot and then click on the option **Play Virtual Machine** to start the VM**.**



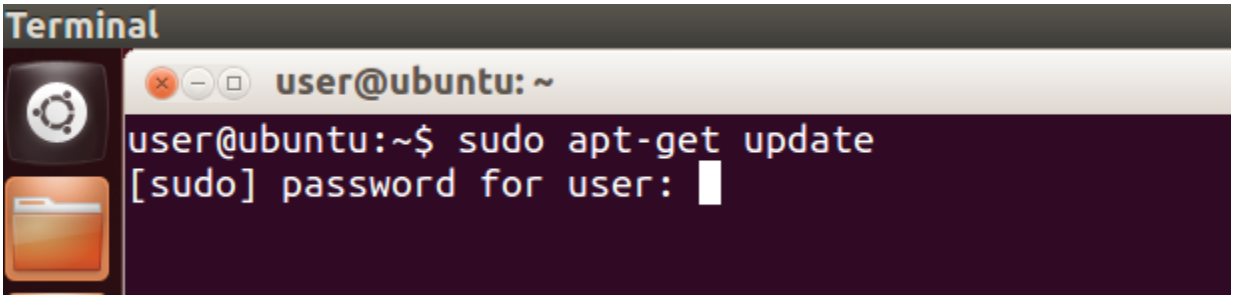
Note: Login into the Ubuntu with below credentials:

USER NAME: user

PASSWORD: password

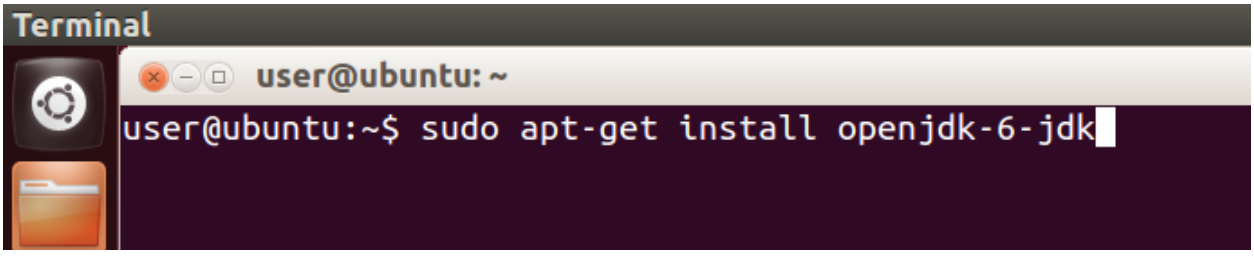
Step 6: The Ubuntu screen will prompt up as shown in the below screenshot and open the terminal as shown below.

Step 7: Update the repository of Ubuntu for the first time to proceed, as other dependencies are missing from the instance.



Step 8: Install java in Linux i.e. Ubuntu via the command line.

Command: sudo apt-get install openjdk-6-jdk



Step 9: Check at your end whether Java is installed or not, refer the below screenshot for the same.

```
user@ubuntu:~$ java -version
java version "1.6.0_36"
OpenJDK Runtime Environment (IcedTea6 1.13.8) (6b36-1.13.8-0ubuntu1~12.04)
OpenJDK Client VM (build 23.25-b01, mixed mode, sharing)
user@ubuntu:~$
```

Step 10: Download the instance of Ubuntu from command line using the below command.

In the case the below command does not work you can download Hadoop manually using the browser present in Ubuntu.

wget http://archive.apache.org/dist/hadoop/core/hadoop-1.2.0/hadoop-1.2.0.tar.gz

```
user@ubuntu: ~
user@ubuntu:~$ wget http://archive.apache.org/dist/hadoop/core/hadoop-1.2.0/hado
op-1.2.0.tar.gz
--2015-08-08 05:15:28--  http://archive.apache.org/dist/hadoop/core/hadoop-1.2.0
/hadoop-1.2.0.tar.gz
Resolving archive.apache.org (archive.apache.org)... 140.211.11.131, 192.87.106.
229, 2001:610:1:80bc:192:87:106:229
Connecting to archive.apache.org (archive.apache.org)|140.211.11.131|:80... conn
ected.
HTTP request sent, awaiting response... 200 OK
Length: 62984953 (60M) [application/x-gzip]
Saving to: `hadoop-1.2.0.tar.gz'

 1% [                                      ] 1,212,953   13.0K/s  eta 23m 53s
```

Step 11: Once Hadoop is downloaded, extract it as it is a tar file using the below command.

You see the extracted Hadoop file as shown below.

tar -xvf hadoop-1.2.0.tar.gz

```
user@ubuntu:~$ ls
Desktop     examples.desktop
Documents   hadoop-1.2.0
Downloads   hadoop-1.2.0.tar.gz
user@ubuntu:~$
```

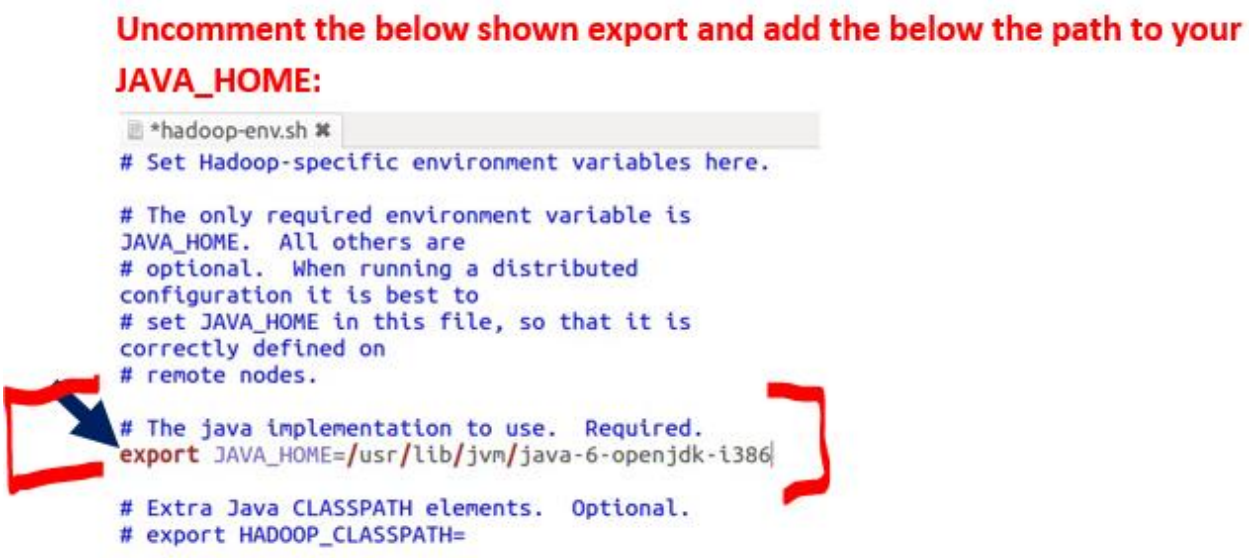
Step 12: Update the various configuration files to complete the set up.

Start with updating **hadoop-env.sh.**

Open the hadoop-env.sh using the below command:

Command: sudo gedit hadoop-1.2.0/conf/hadoop-env.sh

Set the java path in the hadoop-env.sh as shown below:



Step 13: Edit **core-site.xml** with below scripts.

Command: sudo gedit hadoop-1.2.0/conf/core-site.xml

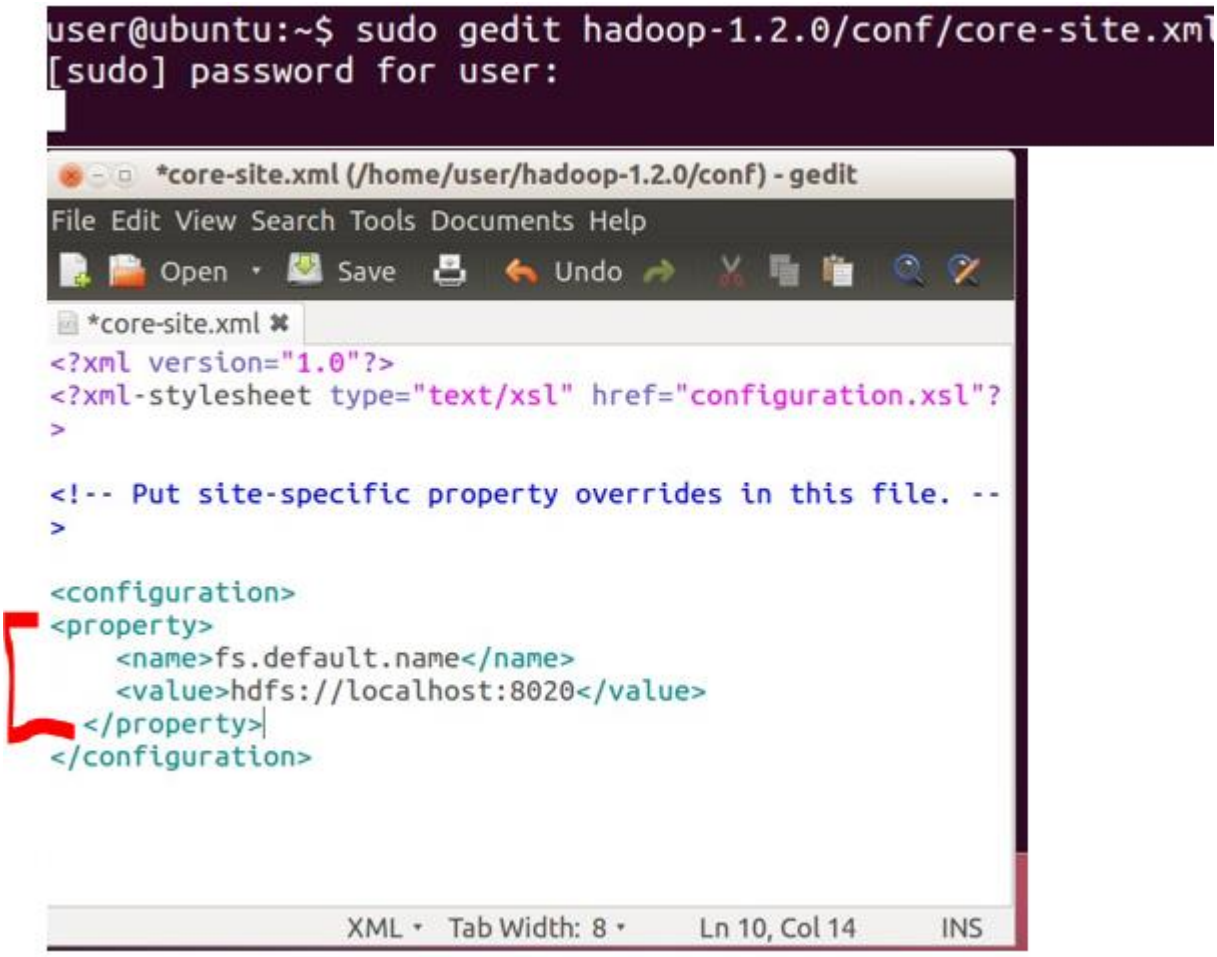
We need to add the following lines in the configuration tags.

<property>

<name>fs.default.name</name>

<value>hdfs://localhost:8020</value>

</property>



Step 14: Open the hdfs-site.xml with the below command:

Command: sudo gedit hadoop-1.2.0/conf/hdfs-site.xml

Add the below scripts in between configuration tags:

<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<property>
<name>dfs.permissions</name>
<value>false</value>
</property>
<name>dfs.namenode.name.dir</name>
<value>/home/acadgild/hadoop/namenode</value>
</property>

<property>
<name>dfs.datanode.data.dir</name>
<value>/home/acadgild/hadoop/datanode</value>

</property>
</configuration>

```
<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<property>
<name>dfs.permissions</name>
<value>false</value>
</property>
<name>dfs.namenode.name.dir</name>
<value>/home/acadgild/hadoop/namenode</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>/home/acadgild/hadoop/datanode</value>
</property>
</configuration>
```

Step 16: Open and edit the mapred-site.xml file.

Command to open the mampred-site.xml:

Command: sudo gedit hadoop-1.2.0/conf/mapred -site.xml

```
user@ubuntu:~$ sudo gedit hadoop-1.2.0/conf/mapred-site.xml
user@ubuntu:~$
```

Open the below scripts in between configuration tag of the mapred-site.xml:

<property>

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

<value>localhost:8021</value>

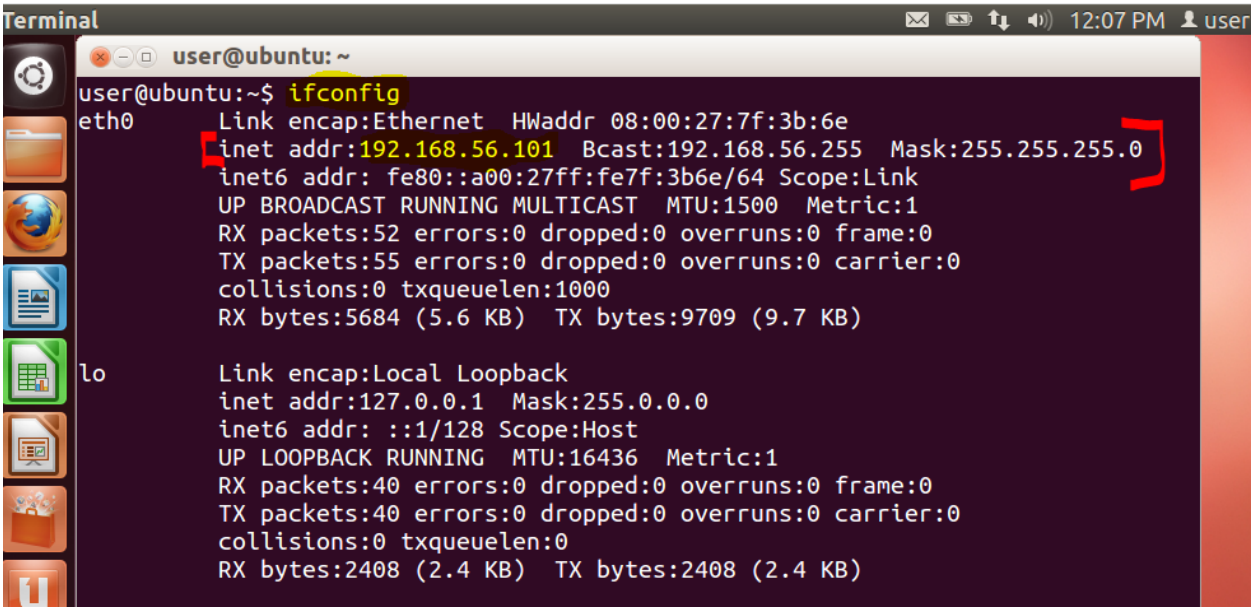
</property>

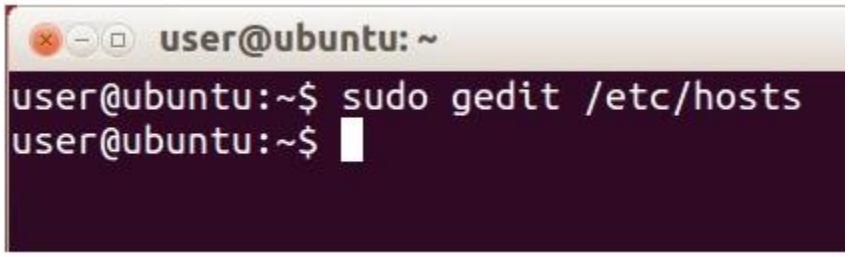
*mapred-site.xml

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
    <name>mapred.job.tracker</name>
    <value>localhost:8021</value>
  </property>
</configuration>
```

Step 17: We need to edit the hosts file with the ip address of our terminal.

Check the ip address of your terminal and the edit it as shown below:



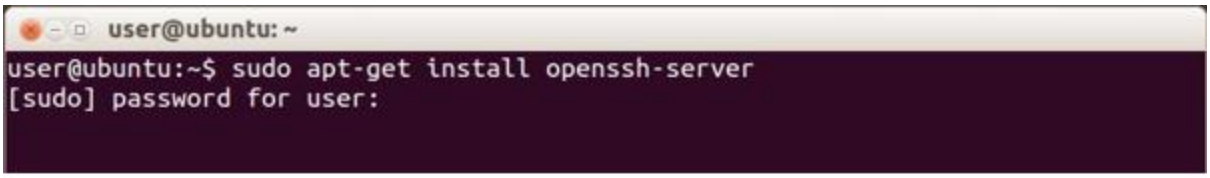


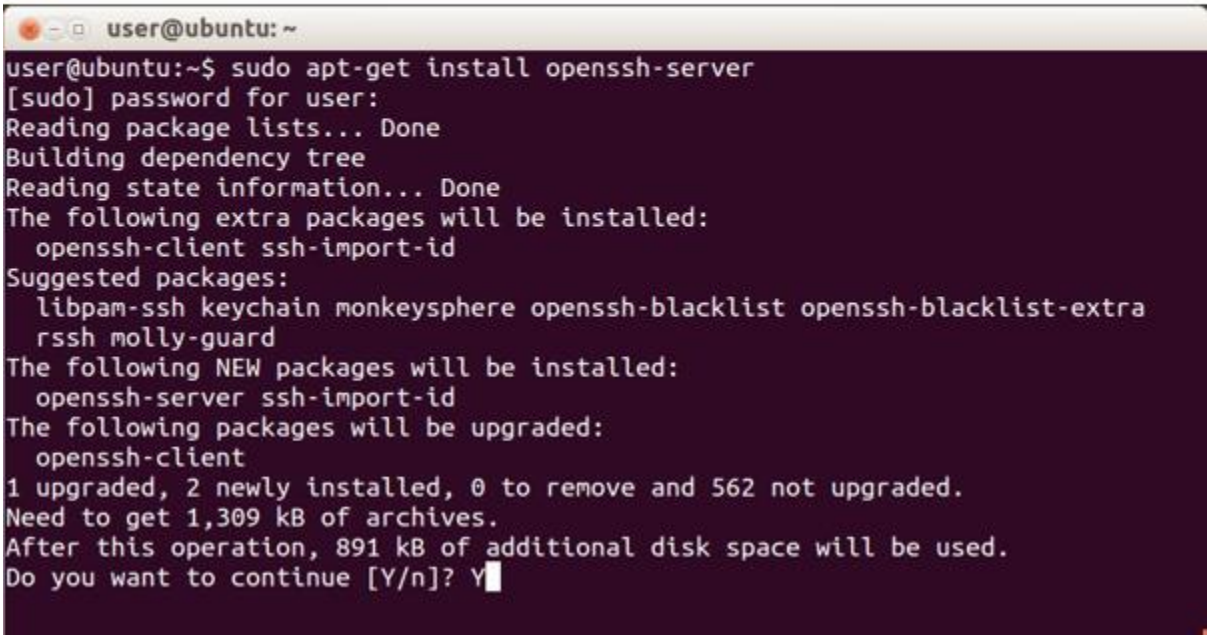
*host

```
192.168.56.101 localhost
```

Step 18: Install the openssh server in Ubuntu.

Command: sudo apt-get install openssh-server





Step 19: Create a password-less login (login without password).

ssh-keygen -t rsa -P

Key points to remember while typing the above command:

1. **Please do not copy paste this command**
2. **Give a space after keygen, after t, after rsa, and after P**
3. **P should in uppercase**
4. **Give single hyphen before t and P**
5. **Do not give any space between the double quotes.**

On entering the above command, a screen prompt appears asking for the location in which to save the key**.**

```
user@ubuntu:~$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/user/.ssh/id_rsa):
```

Note: We need to hit **Enter** once again**.**

Step 20: Move the key into authorized key folder as shown below using the below command.

Command: cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys

```
user@ubuntu:~$ cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
user@ubuntu:~$
```

Step 21: Procedure to start Hadoop Daemons:

Step 21.1 Change the directory of the Ubuntu to the location of the Hadoop where you have downloaded**.**

```
user@ubuntu:~$ cd hadoop-1.2.0/
user@ubuntu:~/hadoop-1.2.0$
```

Step 21.2 Format the NameNode before starting the daemon:

```
user@ubuntu:~$ cd hadoop-1.2.0/
user@ubuntu:~/hadoop-1.2.0$ bin/hadoop namenode -format
```

Step 21.3. Start the DFS daemons:

Command: bin/start-dfs.sh

Type the command to see the HDFS daemons including Namenode, SecondaryNamenode, and DataNode.

```
user@ubuntu:~/hadoop-1.2.0$ bin/start-dfs.sh
starting namenode, logging to /home/user/hadoop-1.2.0/libexec/../logs/hadoop-use
r-namenode-ubuntu.out
localhost: starting datanode, logging to /home/user/hadoop-1.2.0/libexec/../logs
/hadoop-user-datanode-ubuntu.out
localhost: starting secondarynamenode, logging to /home/user/hadoop-1.2.0/libexe
c/../logs/hadoop-user-secondarynamenode-ubuntu.out
user@ubuntu:~/hadoop-1.2.0$ jps
10819 NameNode
11030 DataNode
11244 SecondaryNameNode
11282 Jps
user@ubuntu:~/hadoop-1.2.0$
```

Step 21.4. Start the MapReduce daemons.

Command: bin/start-mapred.sh

Type the command **jps** to check the remaining two MapReduce daemons including TaskTracker and JobTracker.

```
user@ubuntu:~/hadoop-1.2.0$ bin/start-mapred.sh
starting jobtracker, logging to /home/user/hadoop-1.2.0/libexec/../logs/hadoop
-user-jobtracker-ubuntu.out
localhost: starting tasktracker, logging to /home/user/hadoop-1.2.0/libexec/..
/logs/hadoop-user-tasktracker-ubuntu.out
user@ubuntu:~/hadoop-1.2.0$ jps
3645 DataNode
3969 JobTracker
3863 SecondaryNameNode
4223 Jps
3430 NameNode
4184 TaskTracker
user@ubuntu:~/hadoop-1.2.0$
```

This completes the Hadoop Installation steps.