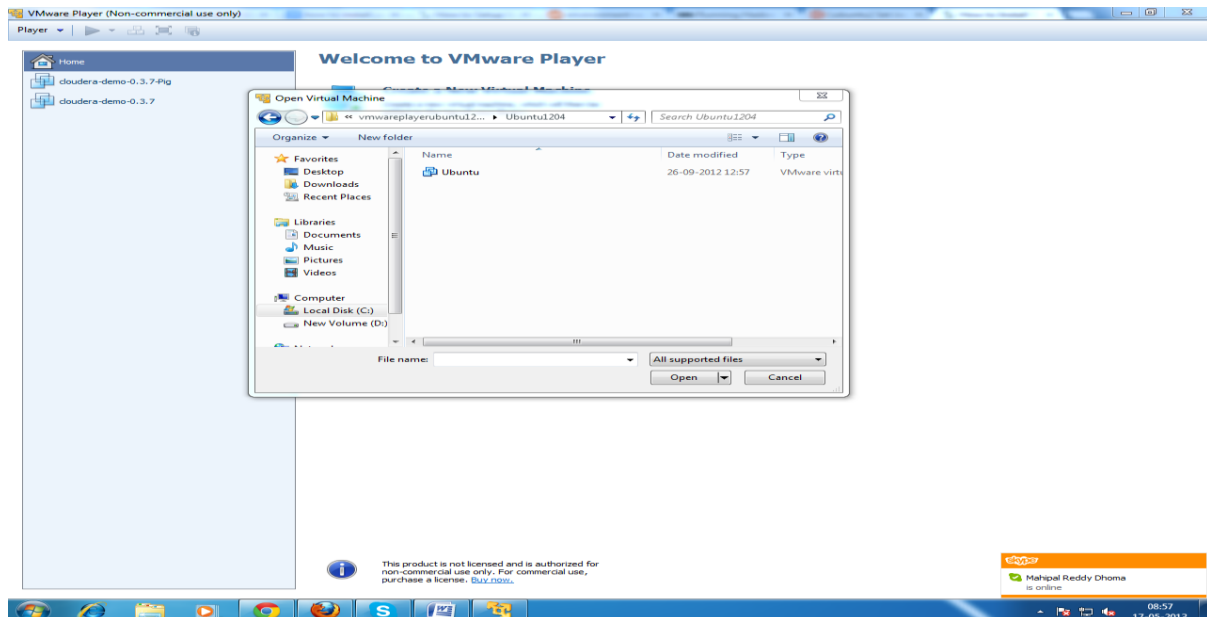
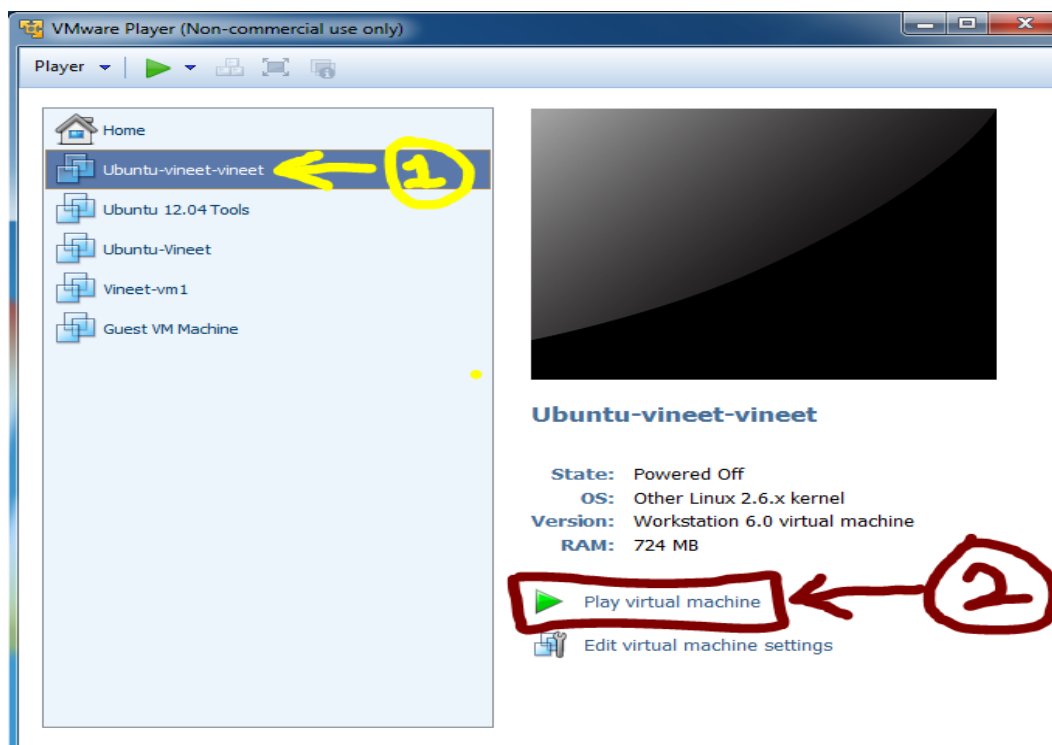


# Hadoop Installation with Single DataNode

- Go to the below link and download the image of ubuntu 12.04
  - <http://www.traffictool.net/vmware/ubuntu1204t.html>
- Open VMware Player and click open virtual machine and select path where you have extracted image of Ubuntu. After that select the .vmx file and click ok.



- Now you can see the below screen in VMware Player.

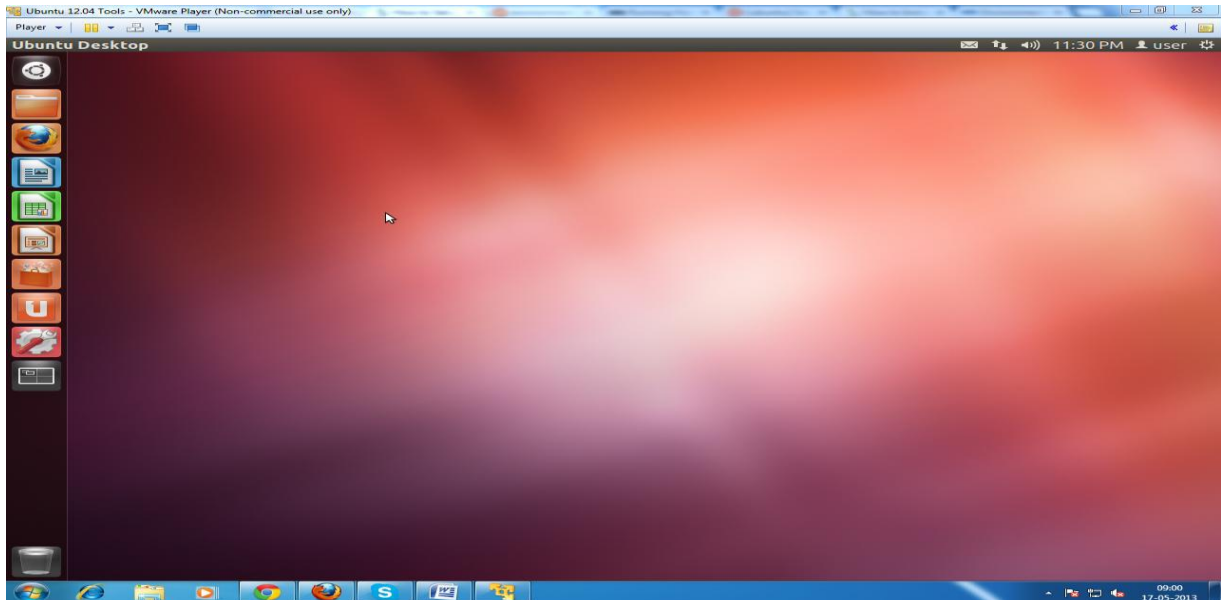


# edureka!

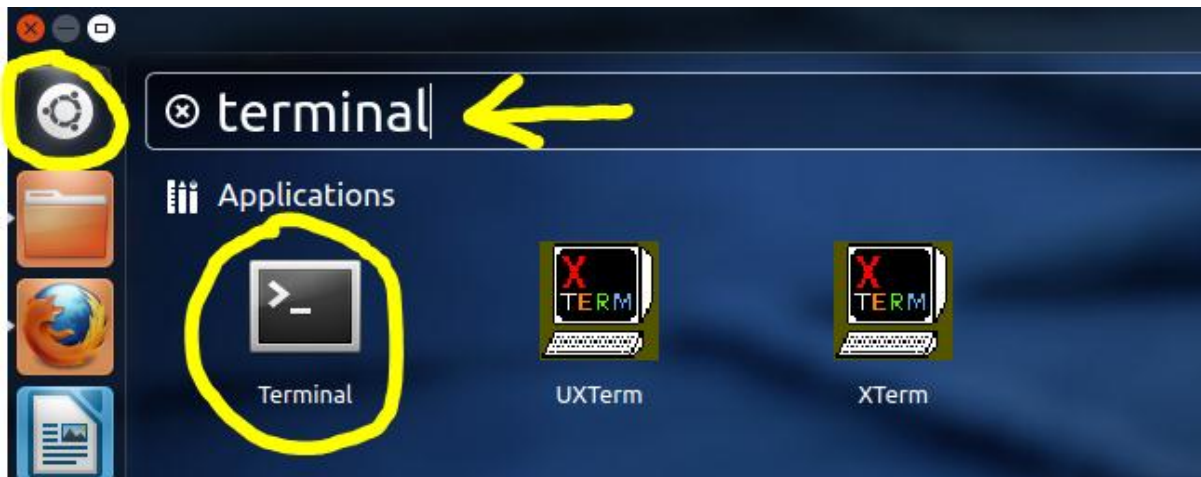
- Double click on ubuntu present in VMware Player. You will get a screen of the below image.

Username : **user**

Password : **password**



- Open a Terminal



- Update the repository:
  - Command: `sudo apt-get update`

```
user@ubuntu:~$ sudo apt-get update
```

# edureka!

➤ **Once the Update is complete :**

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

```
user@ubuntu:~$ sudo apt-get install openjdk-6-jdk
```

➤ **After Java has been Installed, To check whether Java is installed on your system or not give the below command :**

- **Command:** `java -version`

```
user@ubuntu:~$ java -version
java version "1.6.0_27"
OpenJDK Runtime Environment (IcedTea6 1.12.5) (6b27-1.12.5-0ubuntu0.12.04.1)
OpenJDK Client VM (build 20.0-b12, mixed mode, sharing)
user@ubuntu:~$
```

➤ **Install openssh-server:**

- **Command:** `sudo apt-get install openssh-server`

```
user@ubuntu:~$ sudo apt-get install openssh-server
[sudo] password for user:
```

```
user@ubuntu:~$ sudo apt-get install openssh-server
[sudo] password for user:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following extra packages will be installed:
  openssh-client ssh-import-id
Suggested packages:
  libpam-ssh keychain monkeysphere openssh-blacklist openssh-blacklist-extra
  rssh molly-guard
The following NEW packages will be installed:
  openssh-server ssh-import-id
The following packages will be upgraded:
  openssh-client
1 upgraded, 2 newly installed, 0 to remove and 562 not upgraded.
Need to get 1,309 kB of archives.
After this operation, 891 kB of additional disk space will be used.
Do you want to continue [Y/n]? Y
```

➤ Download and extract Hadoop:

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

```
user@ubuntu:~$ ls
Desktop      examples.desktop  Music      Templates
Documents    hadoop-1.2.0      Pictures    Videos
Downloads    hadoop-1.2.0.tar.gz  Public
```

➤ Edit core-site.xml:

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

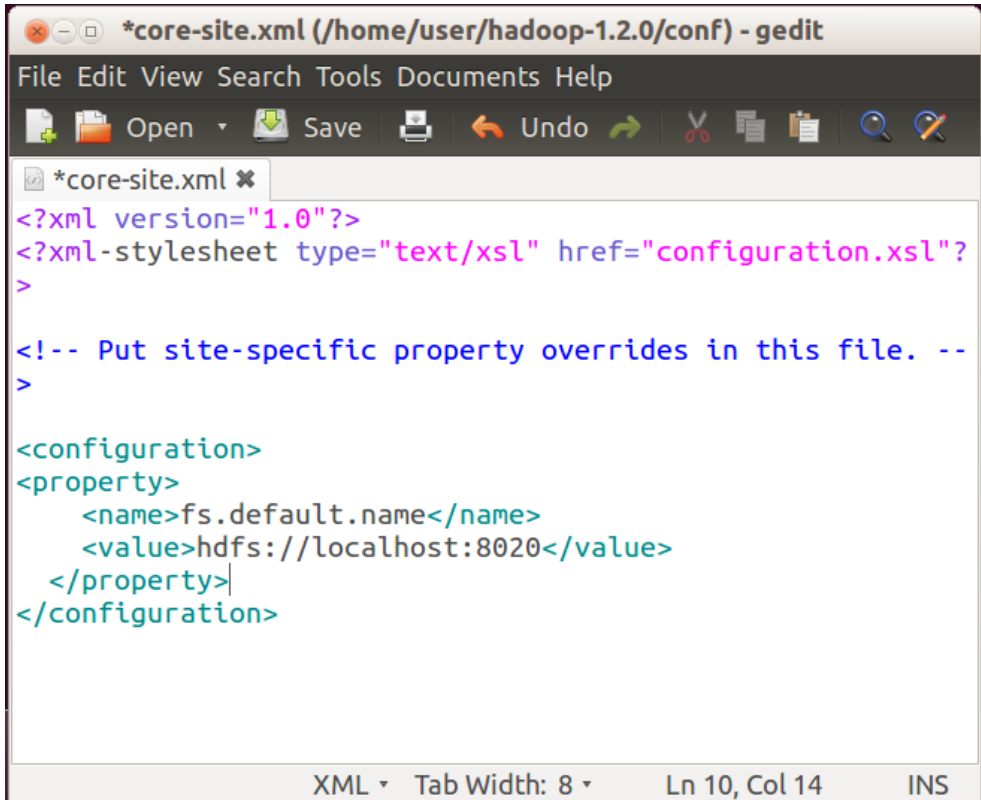
`<property>`

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

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

`</property>`

```
user@ubuntu:~$ sudo gedit hadoop-1.2.0/conf/core-site.xml
[sudo] password for user:
```



```
*core-site.xml (/home/user/hadoop-1.2.0/conf) - gedit
File Edit View Search Tools Documents Help
Open Save Print Undo
*core-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>fs.default.name</name>
  <value>hdfs://localhost:8020</value>
</property>
</configuration>
XML Tab Width: 8 Ln 10, Col 14 INS
```

➤ **Edit hdfs-site.xml:**

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

**<property>**

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

**<value>1</value>**

**</property>**

**<property>**

**<name>dfs.permissions</name>**

**<value>>false</value>**

**</property>**

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

# edureka!

```
hdfs-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>dfs.replication</name>
  <value>1</value>
</property>
<property>
  <name>dfs.permissions</name>
  <value>>false</value>
</property>
</configuration>
```

➤ **Edit mapred-site.xml:**

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

```
<property>

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

<value>localhost:8021</value>

</property>
```

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

# edureka!

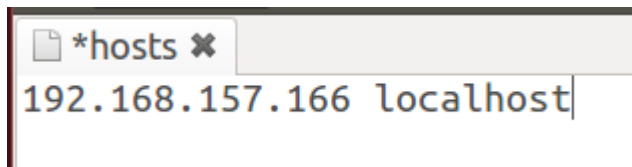
```
*mapred-site.xml ✕  
<?xml version="1.0"?>  
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>  
  
<!-- Put site-specific property overrides in this file. -->  
  
<configuration>  
<property>  
    <name>mapred.job.tracker</name>  
    <value>localhost:8021</value>  
</property>  
</configuration>
```

➤ **Get your ip address:**

- **Command:** ifconfig
- **Command:** sudo gedit /etc/hosts

```
user@ubuntu: ~  
user@ubuntu:~$ ifconfig  
eth0      Link encap:Ethernet  HWaddr 00:0c:29:3c:bf:09  
          inet addr:192.168.157.166  Bcast:192.168.157.255  Mask:255.255.255.0  
          inet6 addr: fe80::20c:29ff:fe3c:bf09/64 Scope:Link  
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1  
          RX packets:106707 errors:0 dropped:0 overruns:0 frame:0  
          TX packets:65046 errors:0 dropped:0 overruns:0 carrier:0  
          collisions:0 txqueuelen:1000  
          RX bytes:134094191 (134.0 MB)  TX bytes:4525014 (4.5 MB)  
          Interrupt:17 Base address:0x1080  
  
lo        Link encap:Local Loopback  
          inet addr:127.0.0.1  Mask:255.0.0.0  
          inet6 addr: ::1/128 Scope:Host  
          UP LOOPBACK RUNNING  MTU:16436  Metric:1  
          RX packets:24323 errors:0 dropped:0 overruns:0 frame:0  
          TX packets:24323 errors:0 dropped:0 overruns:0 carrier:0  
          collisions:0 txqueuelen:0  
          RX bytes:4315399 (4.3 MB)  TX bytes:4315399 (4.3 MB)  
  
user@ubuntu:~$
```

```
user@ubuntu: ~  
user@ubuntu:~$ sudo gedit /etc/hosts  
user@ubuntu:~$
```



```
*hosts *
192.168.157.166 localhost
```

➤ **Create a ssh key:**

- **Command:** `ssh-keygen -t rsa -P ""`

➤ **Moving the key to authorized key:**

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

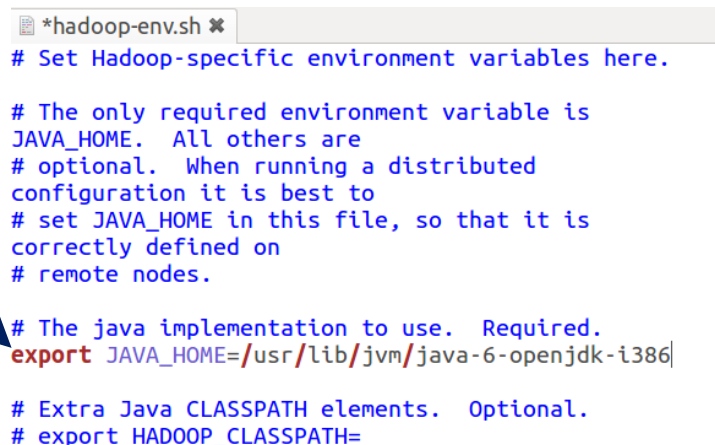
-----Reboot the system-----

## Configuration

➤ **Add JAVA\_HOME in hadoop-env.sh file:**

- **Command:** `sudo gedit hadoop-1.2.0/conf/hadoop-env.sh`
- **Type :** `export JAVA_HOME=/usr/lib/jvm/java-6-openjdk-i386`

**Uncomment the below shown export and add the below the path to your JAVA\_HOME:**



```
*hadoop-env.sh *
# Set Hadoop-specific environment variables here.

# The only required environment variable is
# JAVA_HOME. All others are
# optional. When running a distributed
# configuration it is best to
# set JAVA_HOME in this file, so that it is
# correctly defined on
# remote nodes.

# The java implementation to use. Required.
export JAVA_HOME=/usr/lib/jvm/java-6-openjdk-i386

# Extra Java CLASSPATH elements. Optional.
# export HADOOP_CLASSPATH=
```

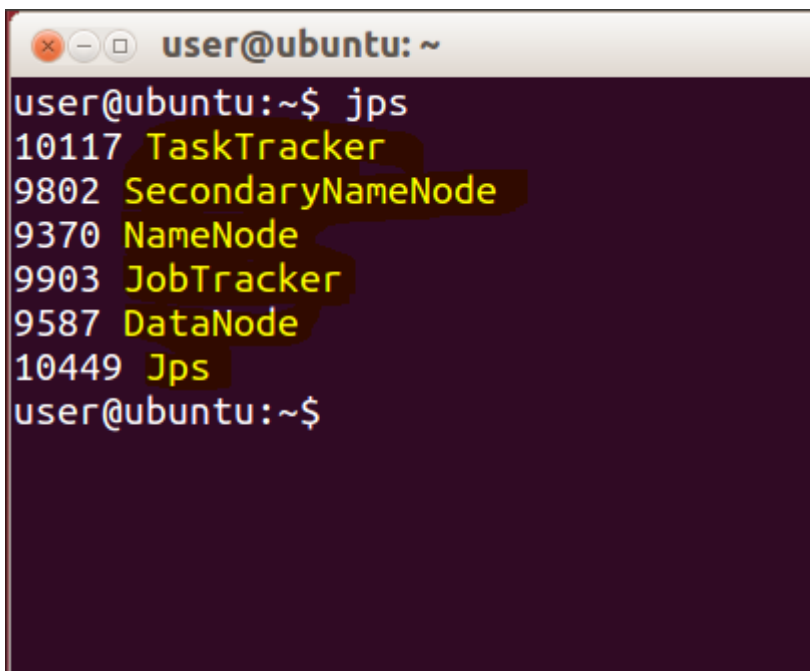
➤ **Change the directory where hadoop is installed.**

- **Command:** `cd hadoop-1.2.0`



# edureka!

- **Format the name node**
  - **Command:** bin/hadoop namenode -format
  
- **Start the namenode, datanode**
  - **Command:** bin/start-dfs.sh
  
- **Start the task tracker and job tracker**
  - **Command:** bin/start-mapred.sh
  
- **To check if Hadoop started correctly**
  - **Command:** jps

A terminal window titled 'user@ubuntu: ~' with standard window controls. The terminal shows the command 'jps' being executed, which returns a list of running Java processes. The output is: 10117 TaskTracker, 9802 SecondaryNameNode, 9370 NameNode, 9903 JobTracker, 9587 DataNode, and 10449 Jps. The prompt 'user@ubuntu:~\$' is shown at the bottom.

```
user@ubuntu:~$ jps
10117 TaskTracker
9802 SecondaryNameNode
9370 NameNode
9903 JobTracker
9587 DataNode
10449 Jps
user@ubuntu:~$
```