

# Setting up Hadoop made Easy

## Version details

---

Following are the details of components used, all license free:

1. Hadoop 1.2.1
2. Ubuntu LTS 12.04 (running on virtual Machine) 64 Bit
3. Windows 8. (The same thing can be done on mac, i.e., install a virtual machine on mac and follow the below procedure).

## Step 1. Installing Virtual Machine

---

### Step 1.1 Download

Free version of Oracle VirtualBox can be downloaded from:

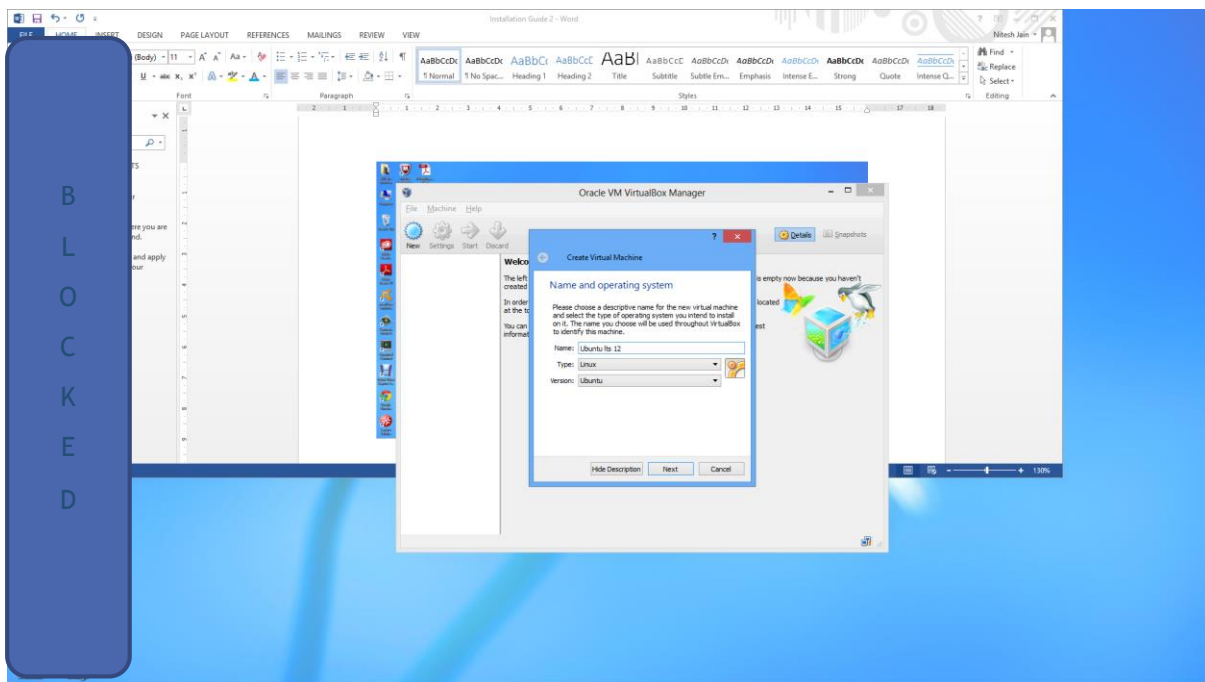
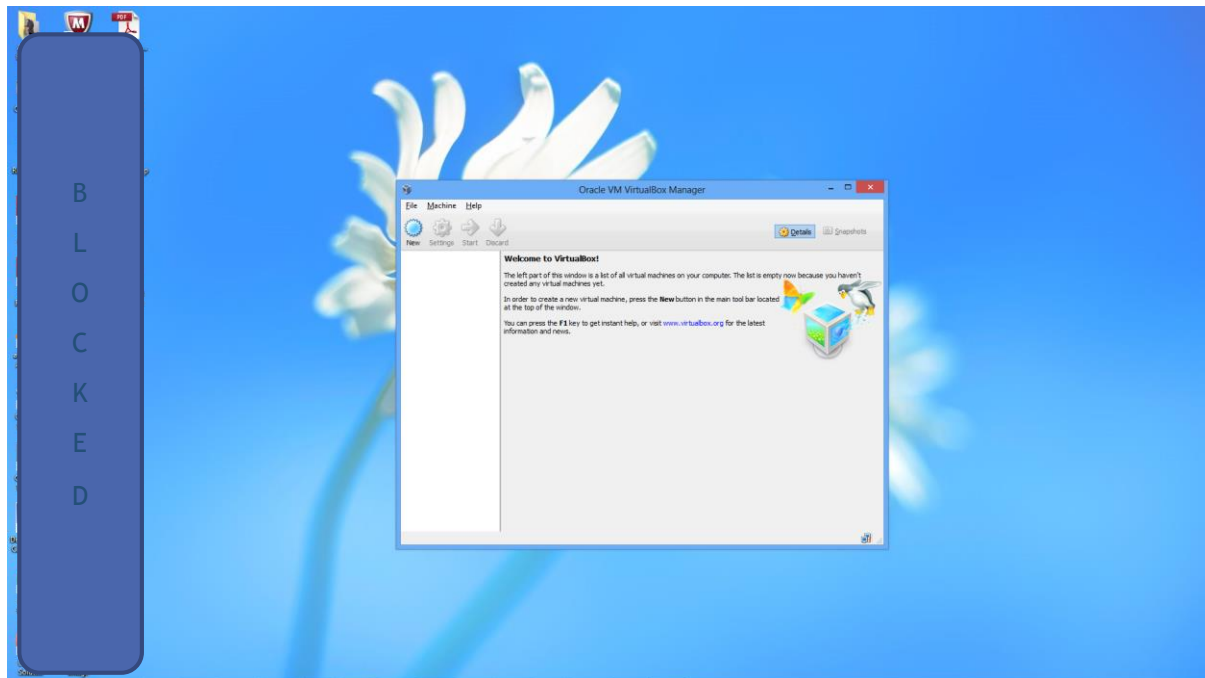
<https://www.virtualbox.org/wiki/Downloads>

Download UBUNTU LTS 64 bit from the following link (Make sure its ISO format and for 64 bit):

<http://www.ubuntu.com/download/desktop>

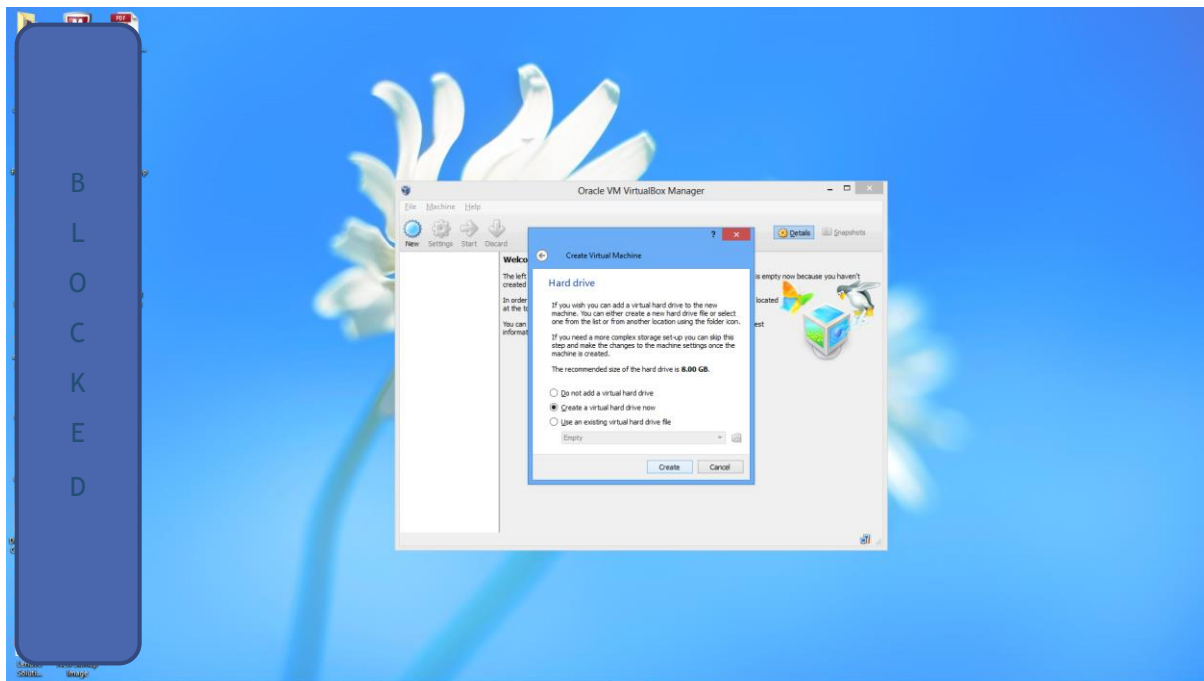
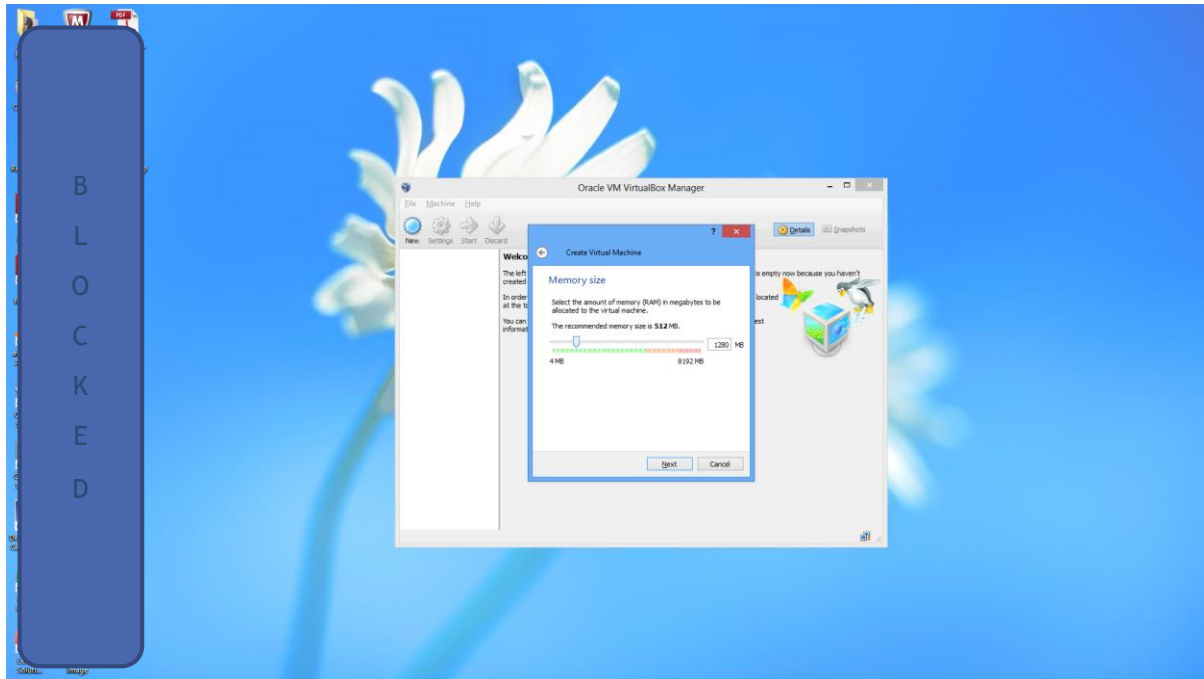
### Step 1.2 Installation

“Become a Certified Hadoop Developer” on udeMy by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)



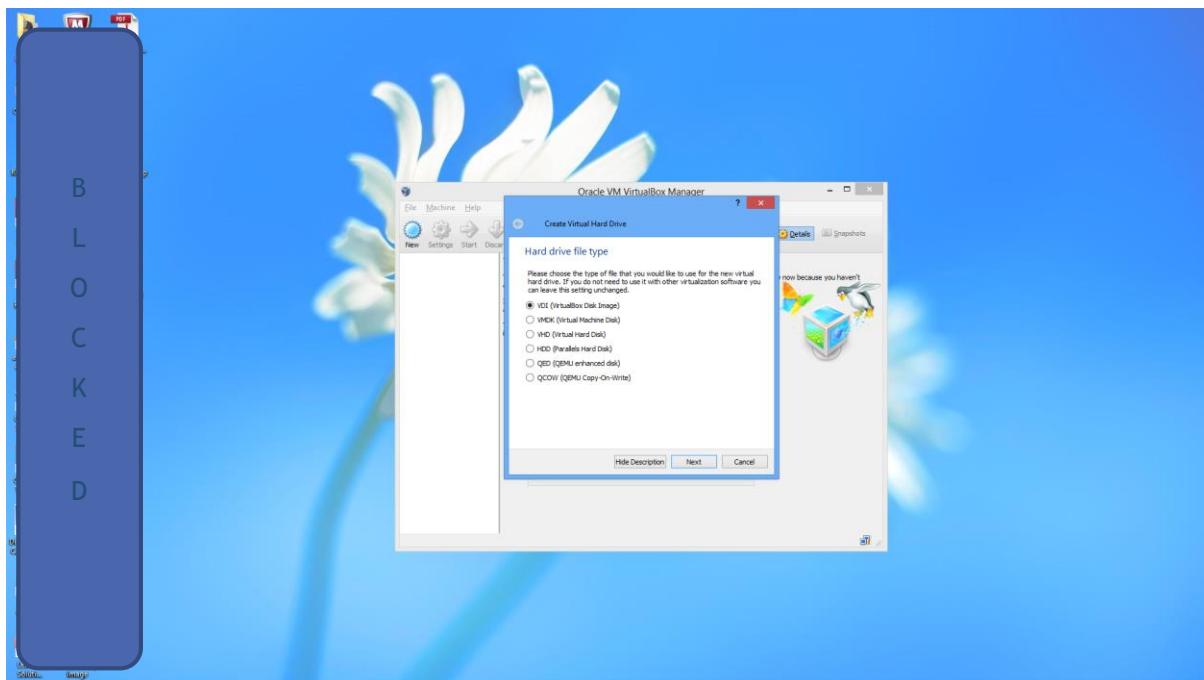
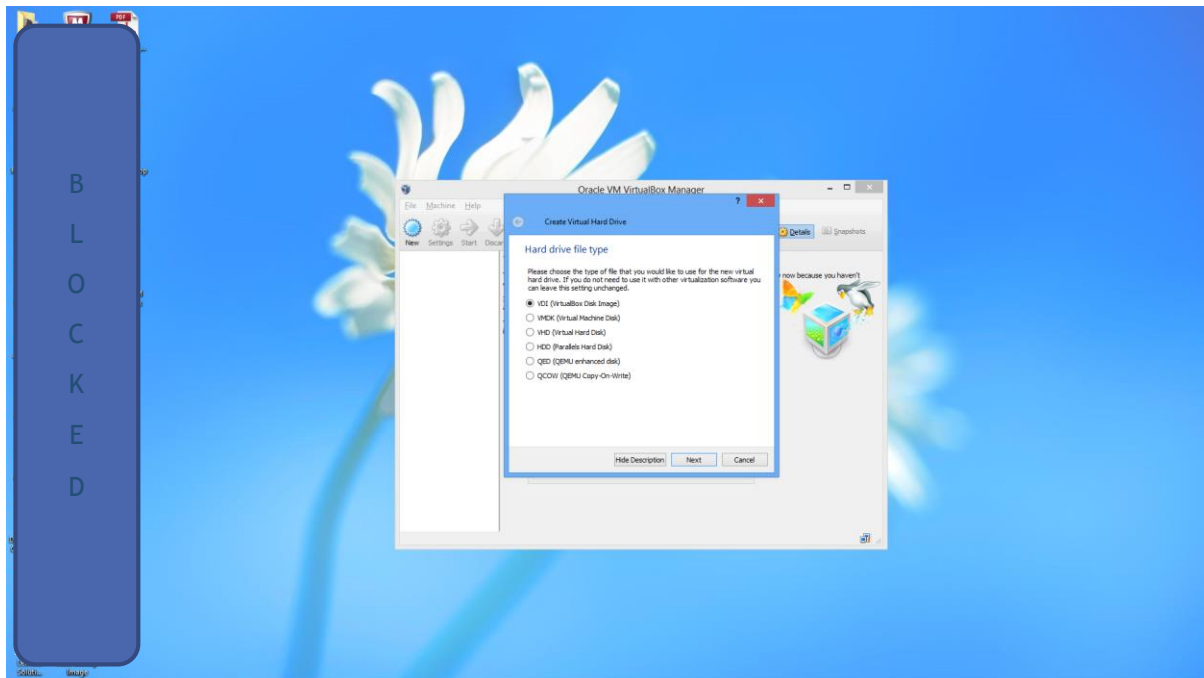
“Become a Certified Hadoop Developer” on udeMy by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)

“Become a Certified Hadoop Developer” on udeby by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)



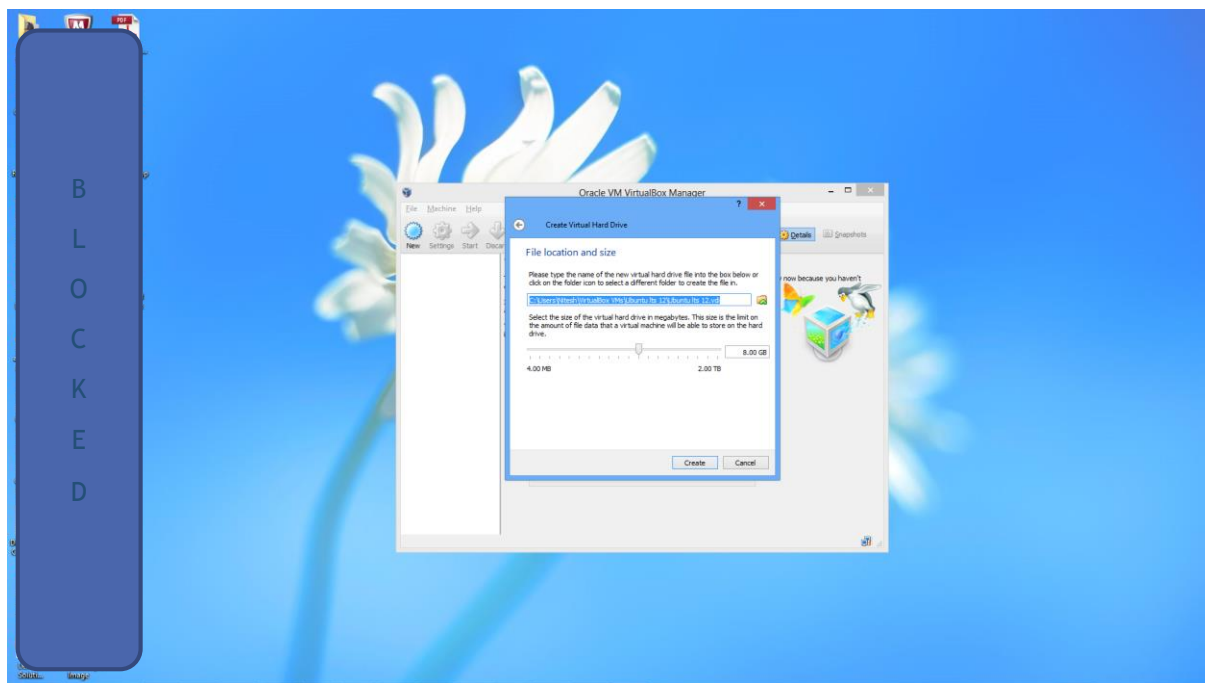
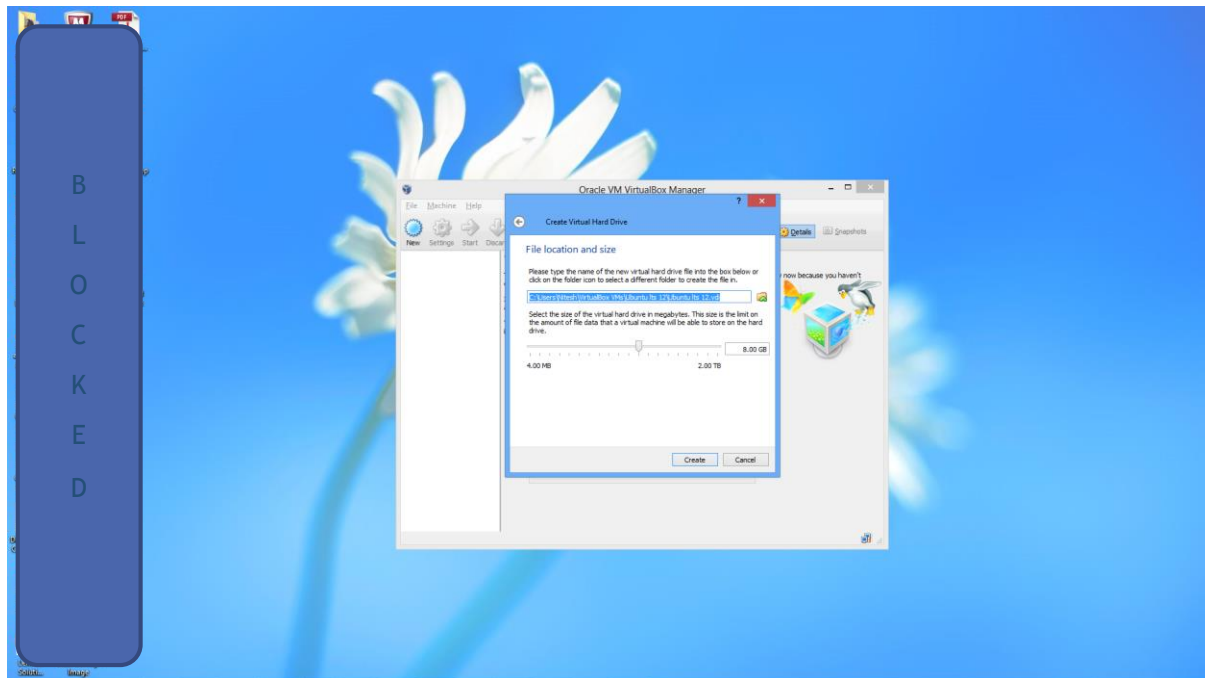
“Become a Certified Hadoop Developer” on udeby by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)

“Become a Certified Hadoop Developer” on udeby by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)



“Become a Certified Hadoop Developer” on udeby by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)

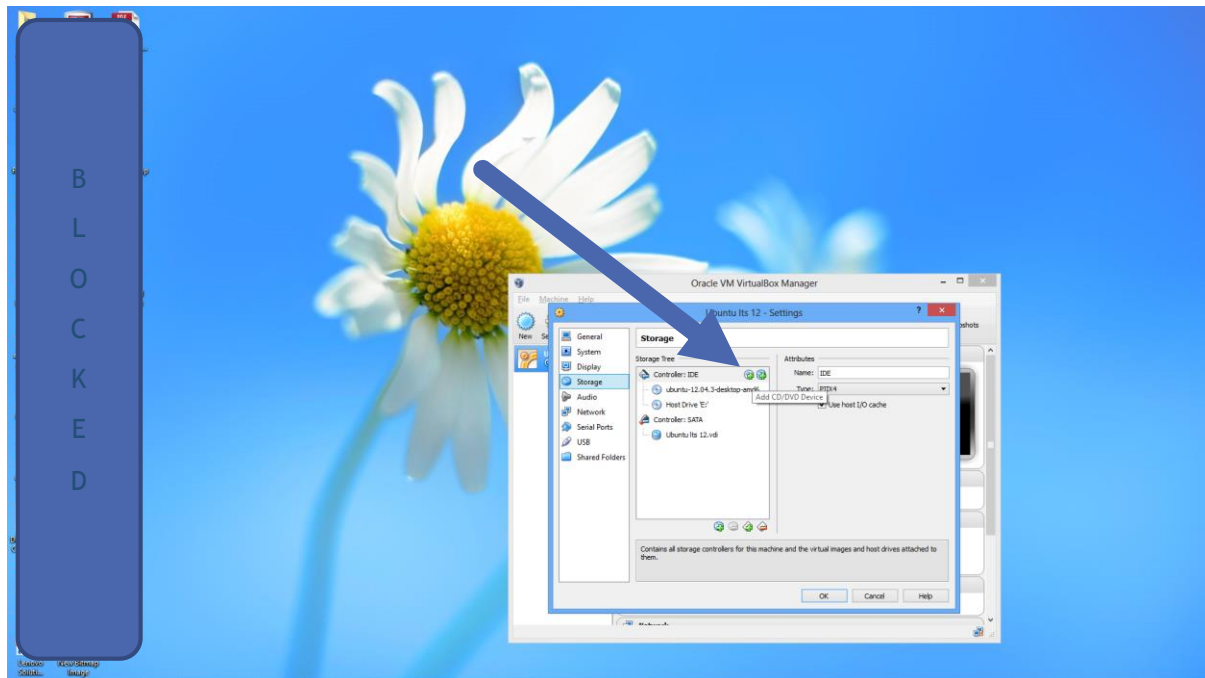
“Become a Certified Hadoop Developer” on udemy by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)



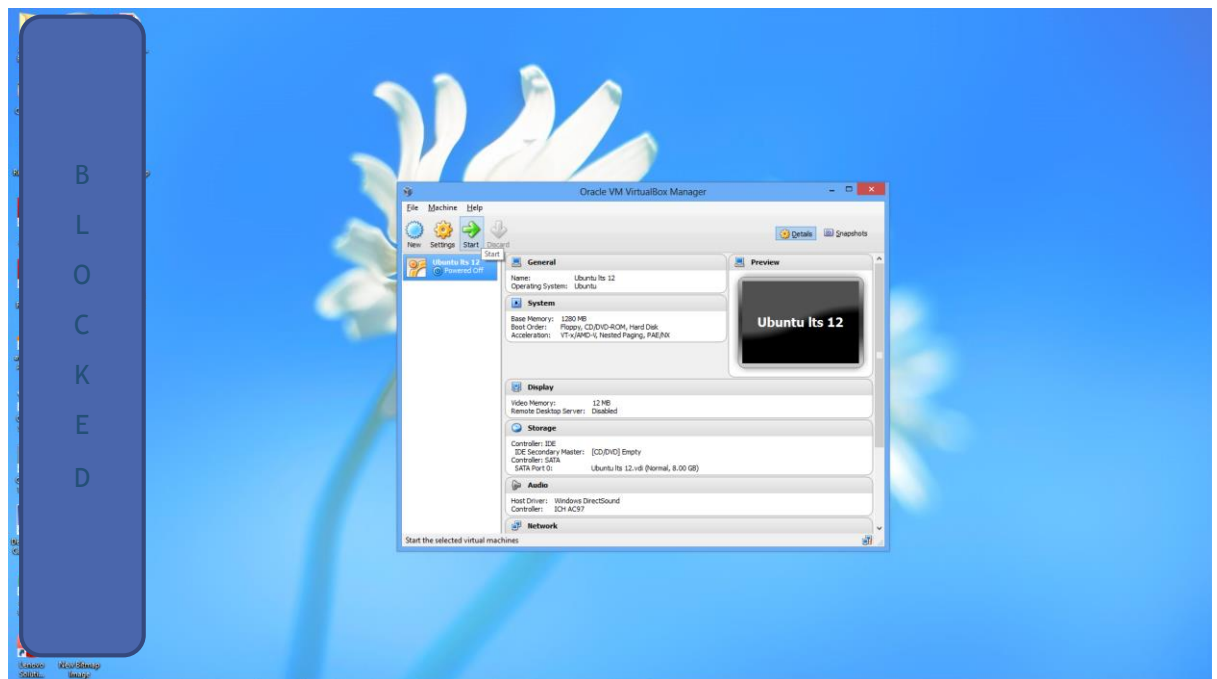
In the below screen shot click on the ‘+’ sign to add ISO which you have already downloaded to be loaded as CD drive.

“Become a Certified Hadoop Developer” on udemy by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)

“Become a Certified Hadoop Developer” on udeby by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)



Press Start.

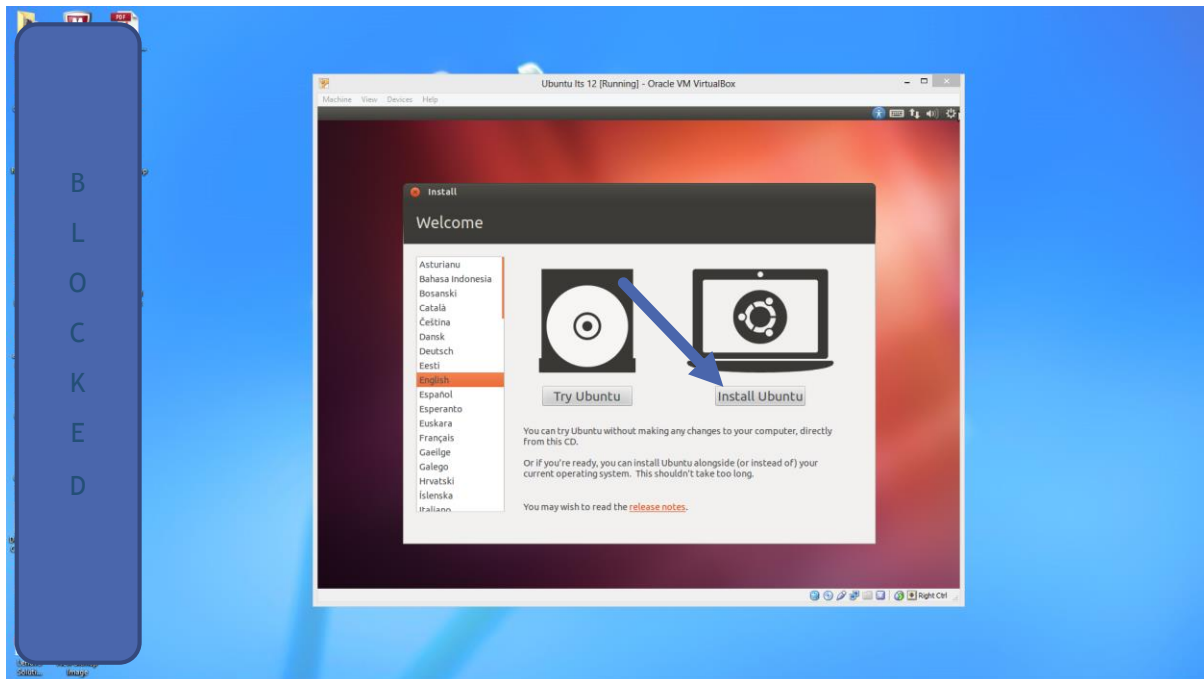


“Become a Certified Hadoop Developer” on udeby by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)

“Become a Certified Hadoop Developer” on udemy by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)

If throws an error, saying something about that 64 bit support and about VT-x/AMD-V,

- It means that your BIOS doesn't support virtualization.
- Perform the following steps. This is for my configuration yours may be a little different:
  - Restart you computer and go to BIOS setup
  - Goto UEFI Firmware>>Advanced>>CPU Setup >> Intel ® Virtualization Techonlogy. Enable this.
  - Save and exit.
- Now try to start the Ubuntu boot with the ISO image and it should work.

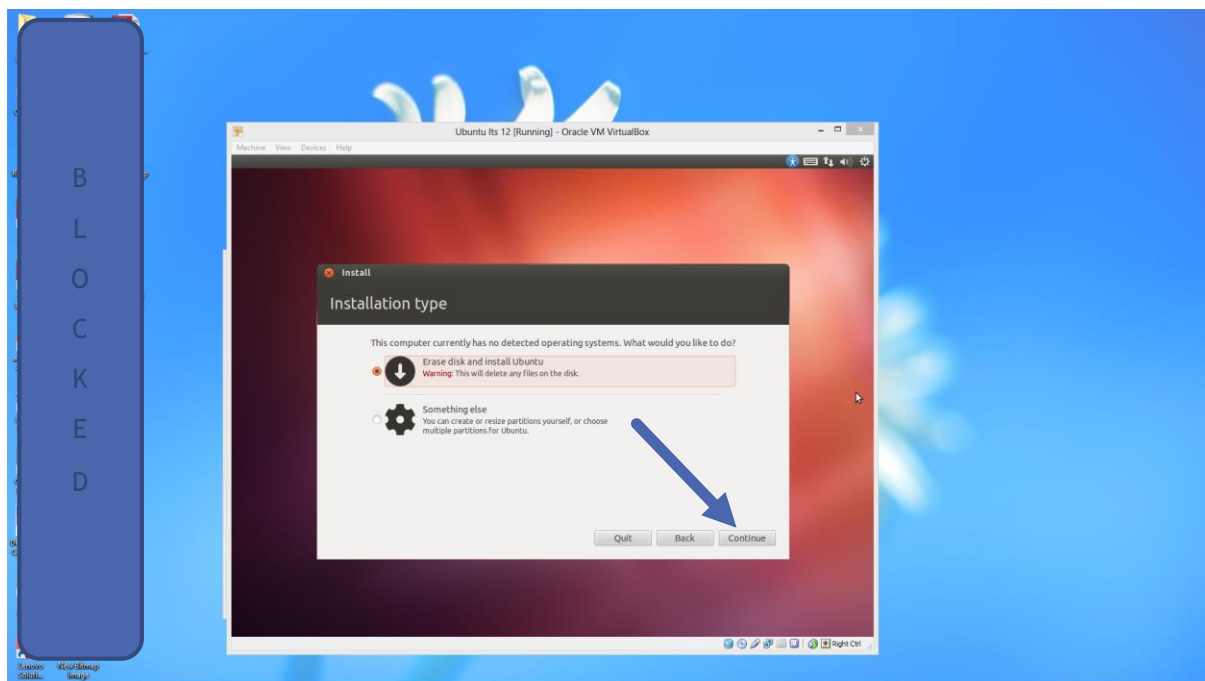
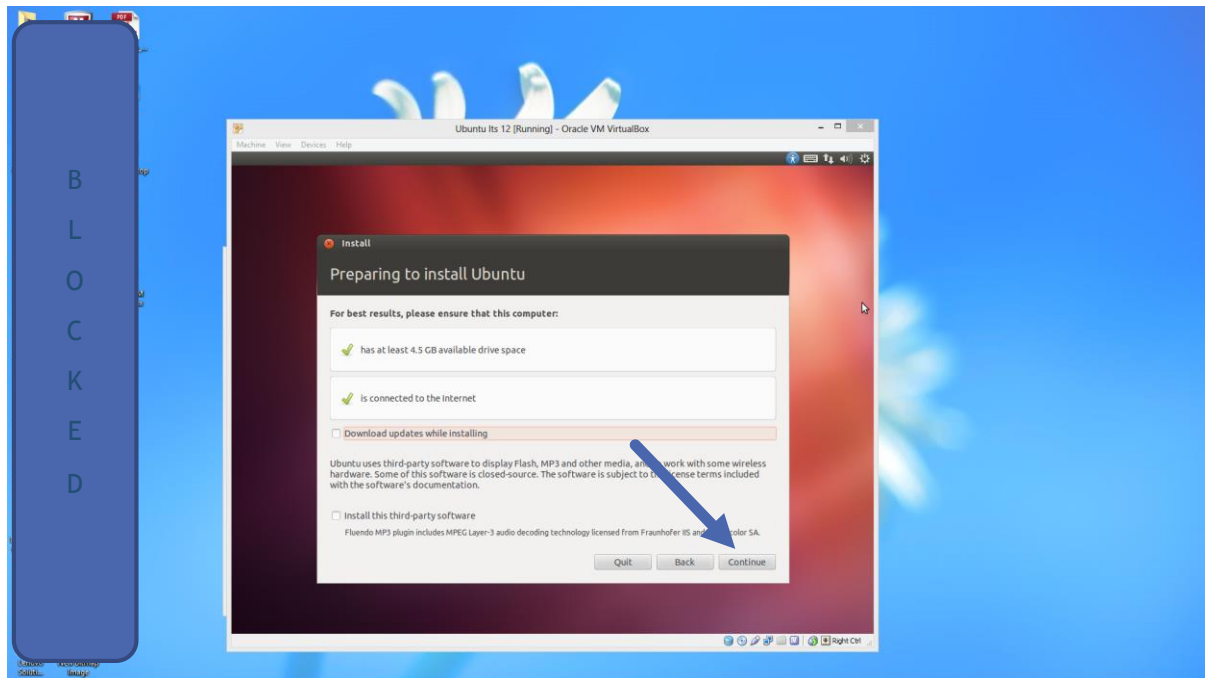


Click on install Ubuntu.

“Become a Certified Hadoop Developer” on udemy by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)



“Become a Certified Hadoop Developer” on udemy by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)

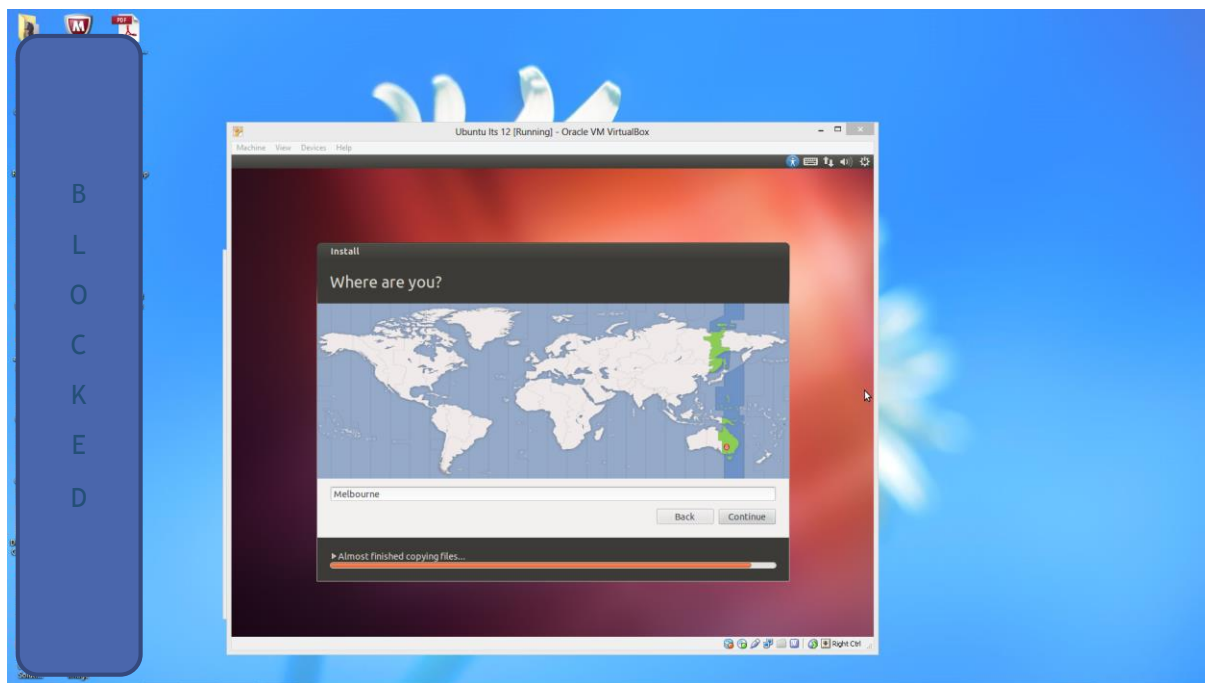
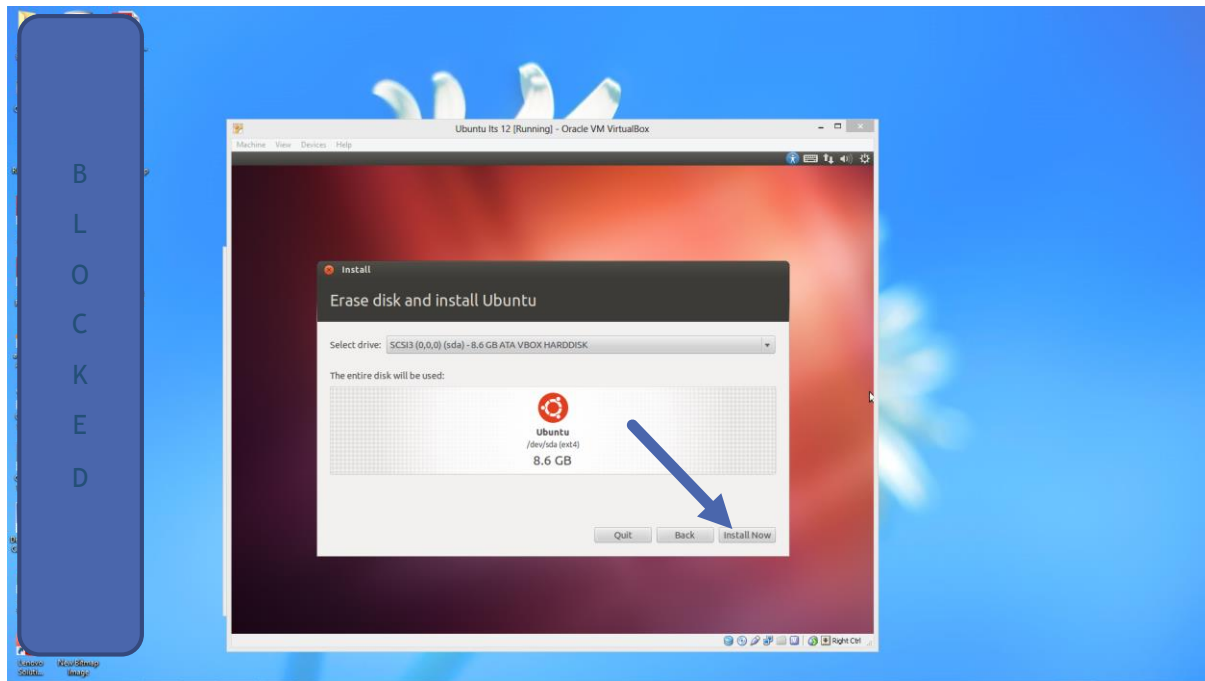


And after you have pressed continue the whole disk would be formatted. Nope just joking! (: Only the dynamic Disk allocated would be formatted.

“Become a Certified Hadoop Developer” on udemy by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)



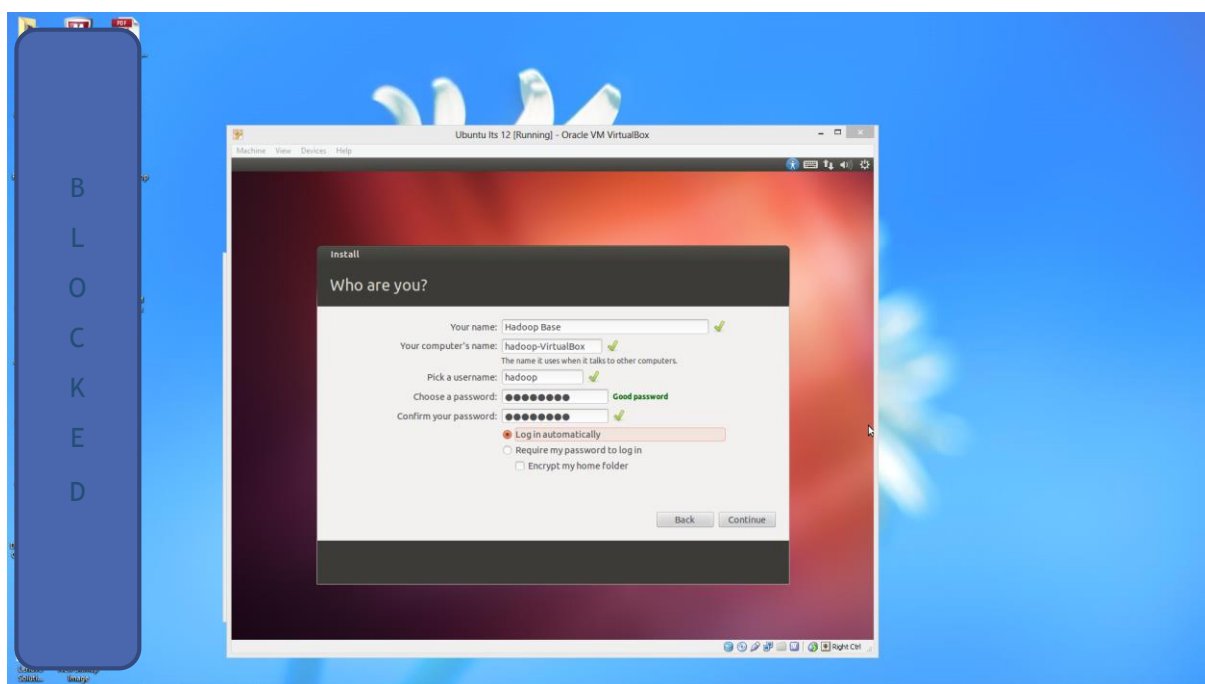
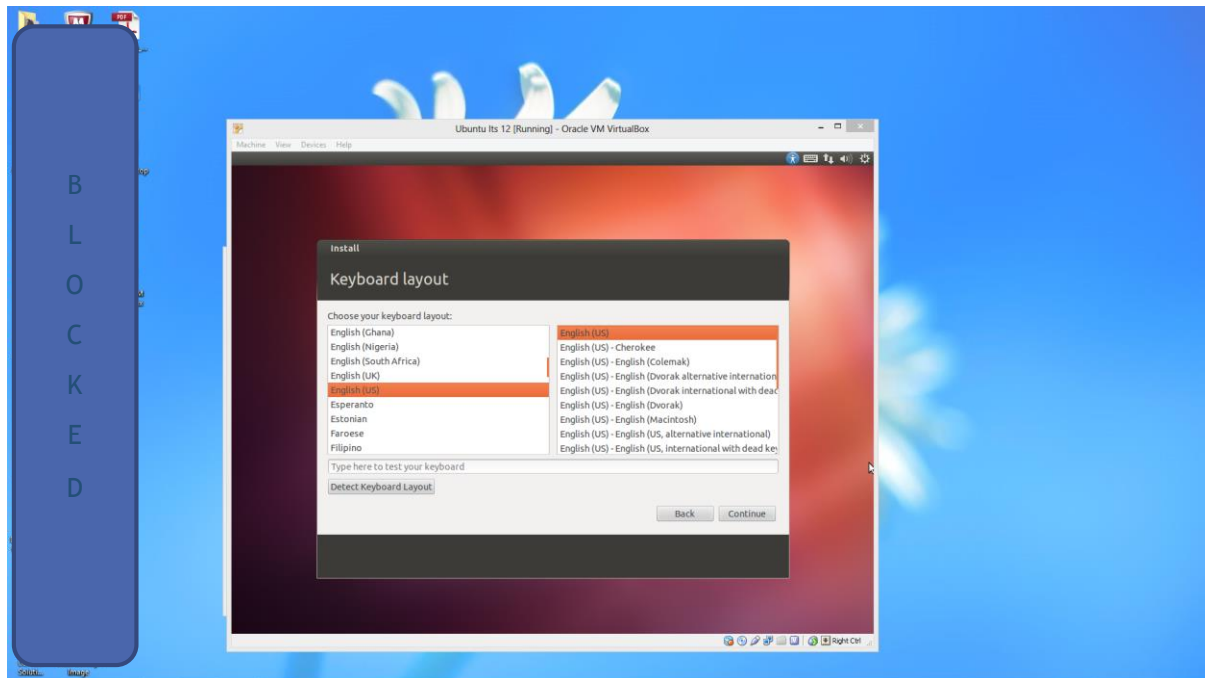
“Become a Certified Hadoop Developer” on udeby by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)



(I live in Melbourne. One of the loveliest cities in the world.)

“Become a Certified Hadoop Developer” on udeby by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)

“Become a Certified Hadoop Developer” on udemy by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)



## Step 2 Download Hadoop tar.gz

At this point you would like to reopen this document on Ubuntu. Transfer it by internet.

Most of the following steps are referred from:

“Become a Certified Hadoop Developer” on udemy by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)

[http://hadoop.apache.org/docs/stable/single\\_node\\_setup.html](http://hadoop.apache.org/docs/stable/single_node_setup.html)

1. Downloading a stable release copy ending with tar.gz
2. Create a new folder /home/hadoop
3. Move the file hadoop.x.y.z.tar.gz to the folder /home/Hadoop
4. *Type/Copy/Paste:* cd /home/hadoop
5. *Type/Copy/Paste:* tar xzf hadoop\*.tar.gz

## Step 3 Downloading and setting up Java

---

For more refer: <http://www.wikihow.com/Install-Oracle-Java-on-Ubuntu-Linux>

1. Check if Java is already present, by  
*Type/Copy/Paste :* java -version
2. If it is 1.7.\* then you can setup the JAVA\_HOME Variable according to where it is setup.
3. If you are confident to setup the JAVA\_HOME variable please go ahead to step X. If not don't worry and follow the following steps:
4. First we will purge the Java installed.

*Type/Copy/Paste :* sudo apt-get purge openjdk-\*

5. Make the directory where java would installed, by:

sudo mkdir -p /usr/local/java

6. Download Java JDK and JRE from the link, look for linux, 64 bit and tar.gz ending file:

<http://www.oracle.com/technetwork/java/javase/downloads/index.html>

7. Goto downloads folder and then copy to the folder we created for java:

*Type/Copy/Paste:* sudo cp -r jdk-\*.tar.gz /usr/local/java

*Type/Copy/Paste:* sudo cp -r jre-\*.tar.gz /usr/local/java

8. Extract and install Java:

*Type/Copy/Paste:* cd /usr/local/java

*Type/Copy/Paste:* sudo tar xvfz jdk\*.tar.gz

*Type/Copy/Paste:* sudo tar xvfz jre\*.tar.gz

9. Now put all the variables in the profile.

*Type/Copy/Paste:* sudo gedit /etc/profile

At the end copy paste the following. (Note: change the highlighted paths according to your installations. Version number would have changed from making this guide to your installation. So just make sure that the path you mention actually exists)

JAVA\_HOME=/usr/local/java/jdk1.7.0\_40

PATH=\$PATH:\$JAVA\_HOME/bin

JRE\_HOME=/usr/local/java/jre1.7.0\_40

PATH=\$PATH:\$JRE\_HOME/bin

HADOOP\_INSTALL=/home/hadoop/Hadoop/hadoop-1.2.1

“Become a Certified Hadoop Developer” on udemy by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)

```
PATH=$PATH:$HADOOP_INSTALL/bin
export JAVA_HOME
export JRE_HOME
export PATH
```

10. Do the following so that Linux knows where Java is, (Note that the highlighted following paths may be needed to be changed in accordance to your installation):

```
sudo update-alternatives --install "/usr/bin/java" "java" "/usr/local/java/jre1.7.0_40/bin/java" 1
sudo update-alternatives --install "/usr/bin/javac" "javac" "/usr/local/java/jdk1.7.0_40/bin/javac" 1
sudo update-alternatives --install "/usr/bin/javaws" "javaws" "/usr/local/java/jre1.7.0_40/bin/javaws" 1
sudo update-alternatives --set java /usr/local/java/jre1.7.0_40/bin/java
sudo update-alternatives --set javac /usr/local/java/jdk1.7.0_40/bin/javac
sudo update-alternatives --set javaws /usr/local/java/jre1.7.0_40/bin/javaws
```

11. Refresh the profile by:

Type/Copy/Paste: `. /etc/profile`

12. Test by typing `Java -version`.

## Step 4 Stand Alone mode installed! Congratulations!

---

At this point you should have had got to the point that you can run Hadoop in Stand Alone mode. You can practice almost anything for practicing developments in Map Reduce. Test if you are successful:

```
Type/Copy/Paste: cd $HADOOP_INSTALL      (going to the Hadoop directory)
Type/copy/Paste: mkdir input
Type/copy/Paste: bin/hadoop jar hadoop-examples-*.jar grep input output 'dfs[a-z.]+'
Type/copy/Paste: ls output/*
```

## Step 5 Pseudo Distribution Mode

---

1. Type/Copy/Paste: `sudo apt-get install ssh` (to install ssh)
2. Type/Copy/Paste: `sudo apt-get install rsync`
3. Change `conf/core-site.xml` to:

```
<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://localhost:9000</value>
  </property>
```

“Become a Certified Hadoop Developer” on udemy by Nitesh Jain.  
Look for Become a Certified Hadoop Developer on [www.udemy.com](http://www.udemy.com)

```
</configuration>
```

4. Change conf/hdfs-site.xml to:

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
</configuration>
```

5. Change conf/mapred-site.xml to:

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

6. Edit conf/hadoop-env.sh look for JAVA\_HOME and set it up  
export JAVA\_HOME=/usr/local/java/**jdk1.7.0\_40**

7. Setup passwordless ssh by:

Type/copy/paste: ssh-keygen -t dsa -P "" -f ~/.ssh/id\_dsa  
Type/copy/paste: cat ~/.ssh/id\_dsa.pub >> ~/.ssh/authorized\_keys

8. To confirm that passwordless ssh has been setup type the following and you should not be prompted for a password.

Type/copy/paste: ssh localhost

9. Format the name node:

Type/copy/paste: bin/hadoop namenode -format

10. Start all the demons:

Type/copy/paste: bin/start-all.sh

11. UI for Name node <http://localhost:50070/> and Jobtracker <http://localhost:50030/>

12. Stop all the demons:

Type/copy/paste: bin/stop-all.sh

I hope you made it this far. My heartiest Congratulations to you on Hadoop installation!

You are on the road to learn one of the most complex and promising new technology in the current times!

If this document has helped you please do share and like the video.

Thanks you!