

Downloading and Installing the Cloudera VM Instructions (Windows & Ubuntu)

Learning Goals

In this activity, you will:

- Download and Install VirtualBox.
- Download and Install Cloudera Virtual Machine (VM) Image.
- Launch the Cloudera VM.

Hardware Requirements: (A) Quad Core Processor (VT-x or AMD-V support recommended), 64-bit; (B) 8 GB RAM; (C) 20 GB disk free. How to find your hardware information: Open System by clicking the Start button, right-clicking Computer, and then clicking Properties. Most computers with 8 GB RAM purchased in the last 3 years will meet the minimum requirements. You will need a high speed internet connection because you will be downloading files up to 4 Gb in size.

Instructions

Please use the following instructions to download and install the Cloudera Quickstart VM with VirtualBox before proceeding to the Getting Started with the Cloudera VM Environment video. The screenshots are from a Mac but the instructions should be the same for Windows. Please see the discussion boards if you have any issues.

1. **Install VirtualBox.** Go to <https://www.virtualbox.org/wiki/Downloads> to download and install VirtualBox for your computer. The course uses Virtualbox 5.1.X, so we recommend clicking [VirtualBox 5.1 builds](#) on that page and downloading the older package for ease of following instructions and screenshots. However, it shouldn't be too different if you choose to use or upgrade to VirtualBox 5.2.X.

For Windows, select the link "**VirtualBox 5.1.X for Windows hosts x86/amd64**" where 'X' is the latest version.

For Ubuntu Select [VirtualBox 5.2.44](#) (released July 14 2020) *Virtual box for Linux Hosts:*

Ubuntu 32bit|64bit

Open Terminal -

```
$ sudo apt-get update
```

```
$ sudo apt-get install VirtualBox 5.2
```

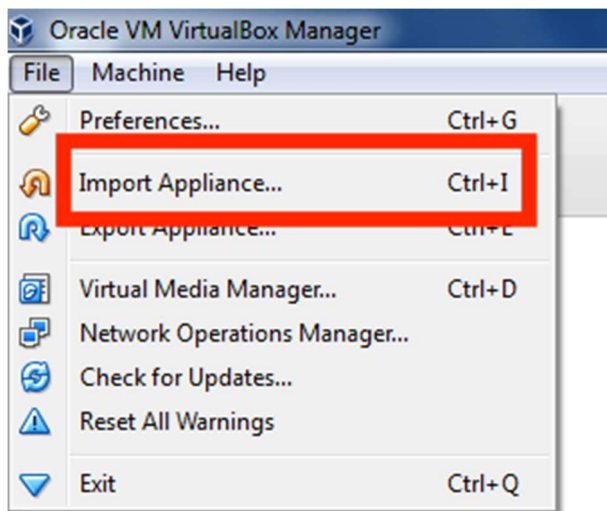
```
$ sudo apt-get -f install
```

2. **Download the Cloudera VM.** Download the Cloudera VM from https://downloads.cloudera.com/demo_vm/virtualbox/cloudera-quickstart-vm-5.4.2-0-virtualbox.zip. The VM is over 4GB, so will take some time to download.

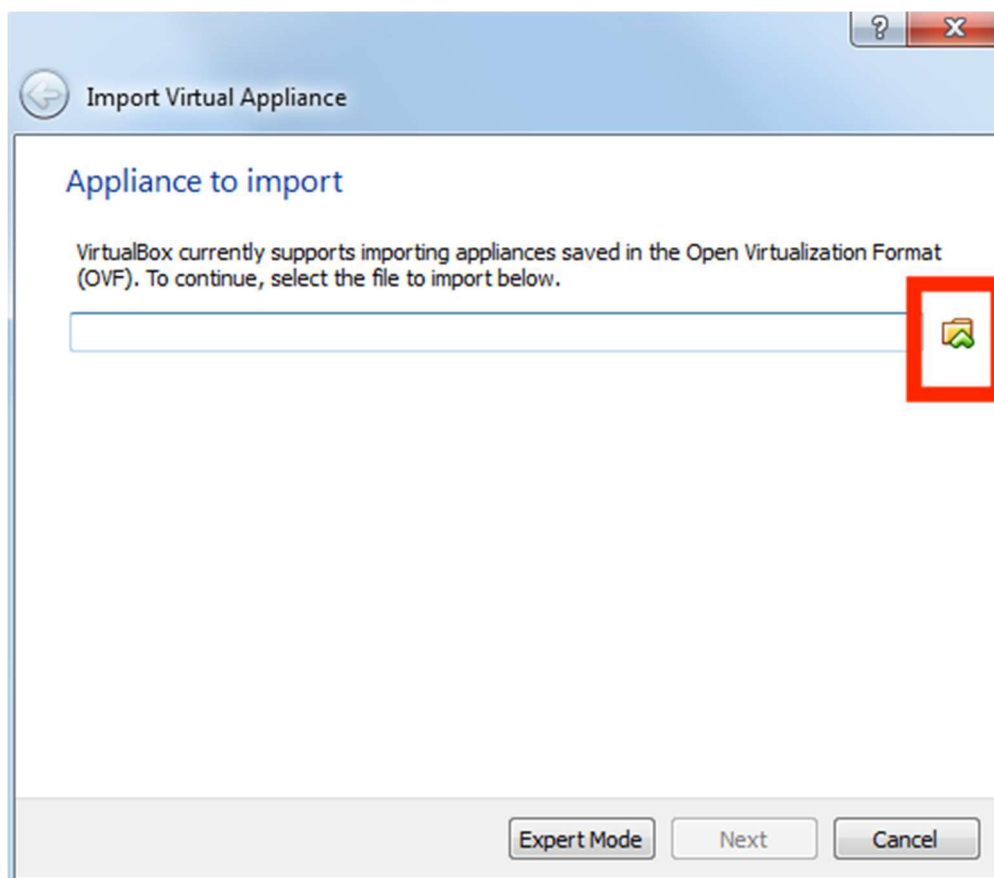
3. **Unzip the Cloudera VM:**

Right-click cloudera-quickstart-vm-5.4.2-0-virtualbox.zip and select "Extract All..."

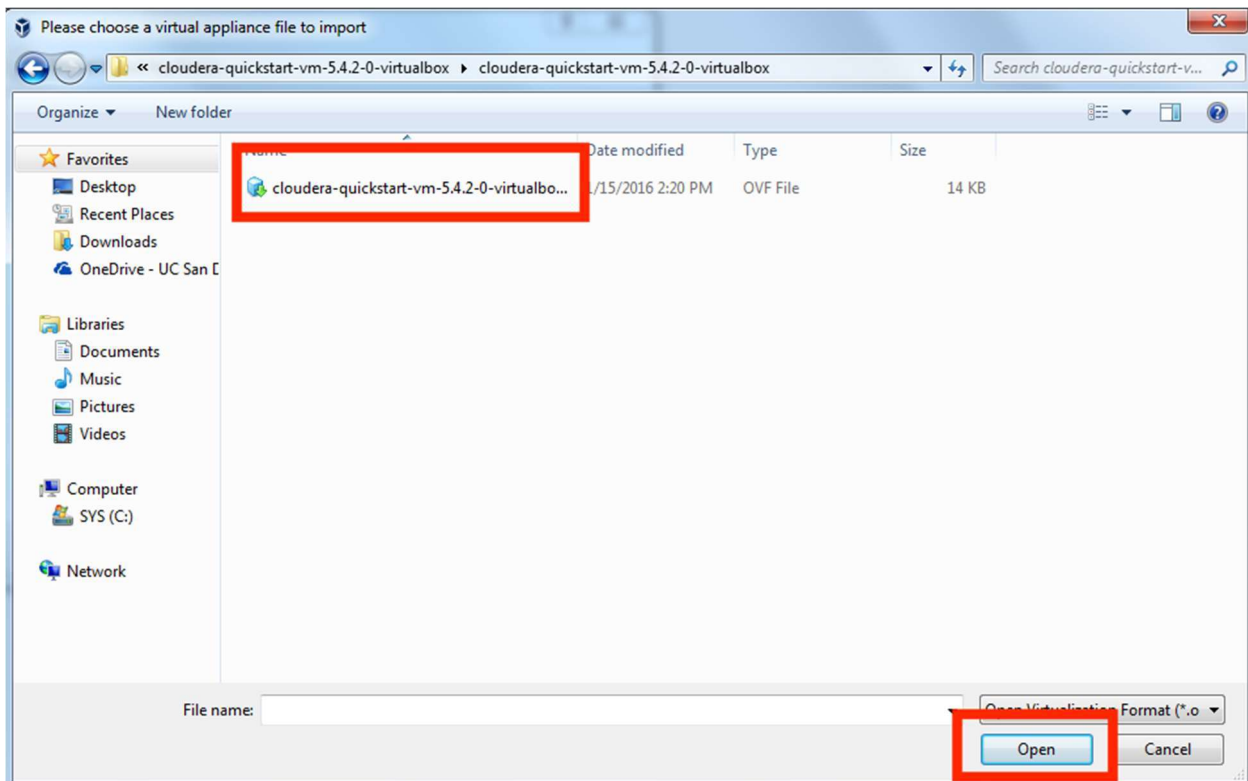
5. **Begin importing.** Import the VM by going to File -> Import Appliance



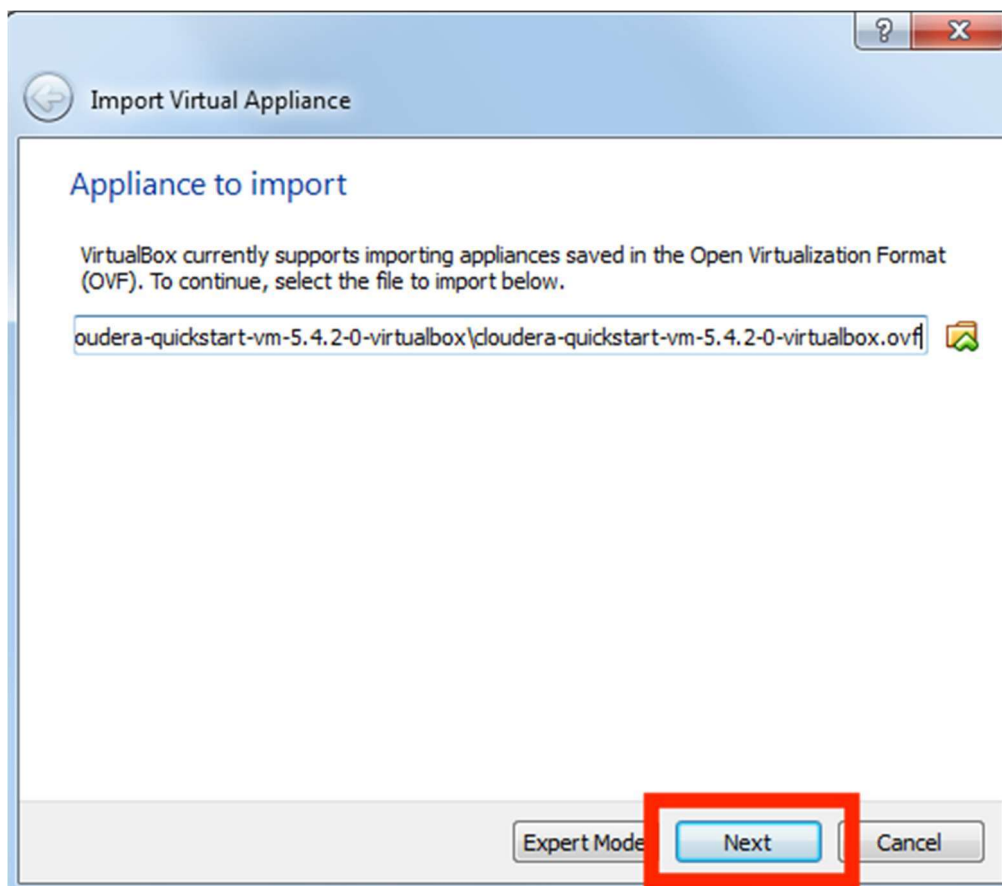
6. **Click the Folder icon.**



7. **Select the cloudera-quickstart-vm-5.4.2-0-virtualbox.ovf** from the Folder where you unzipped the VirtualBox VM and click Open.



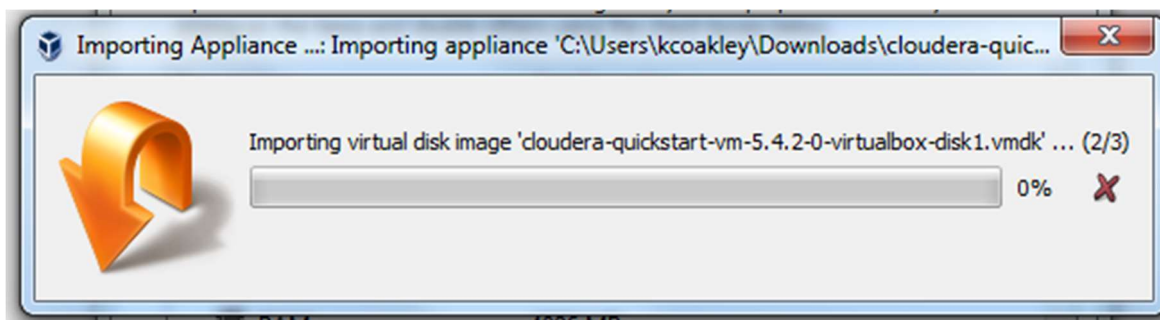
8. Click **Next** to proceed.



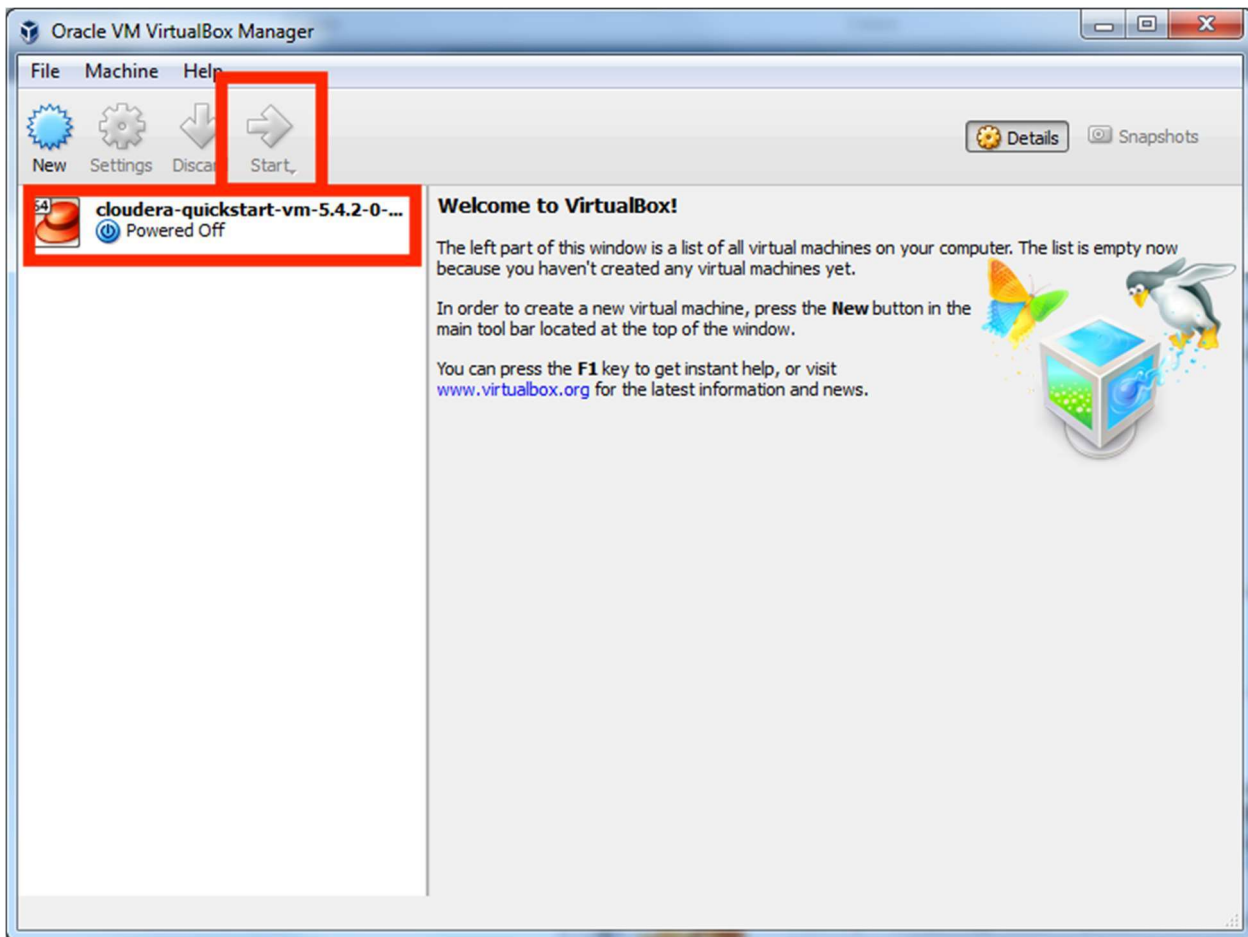
9. Click **Import**.



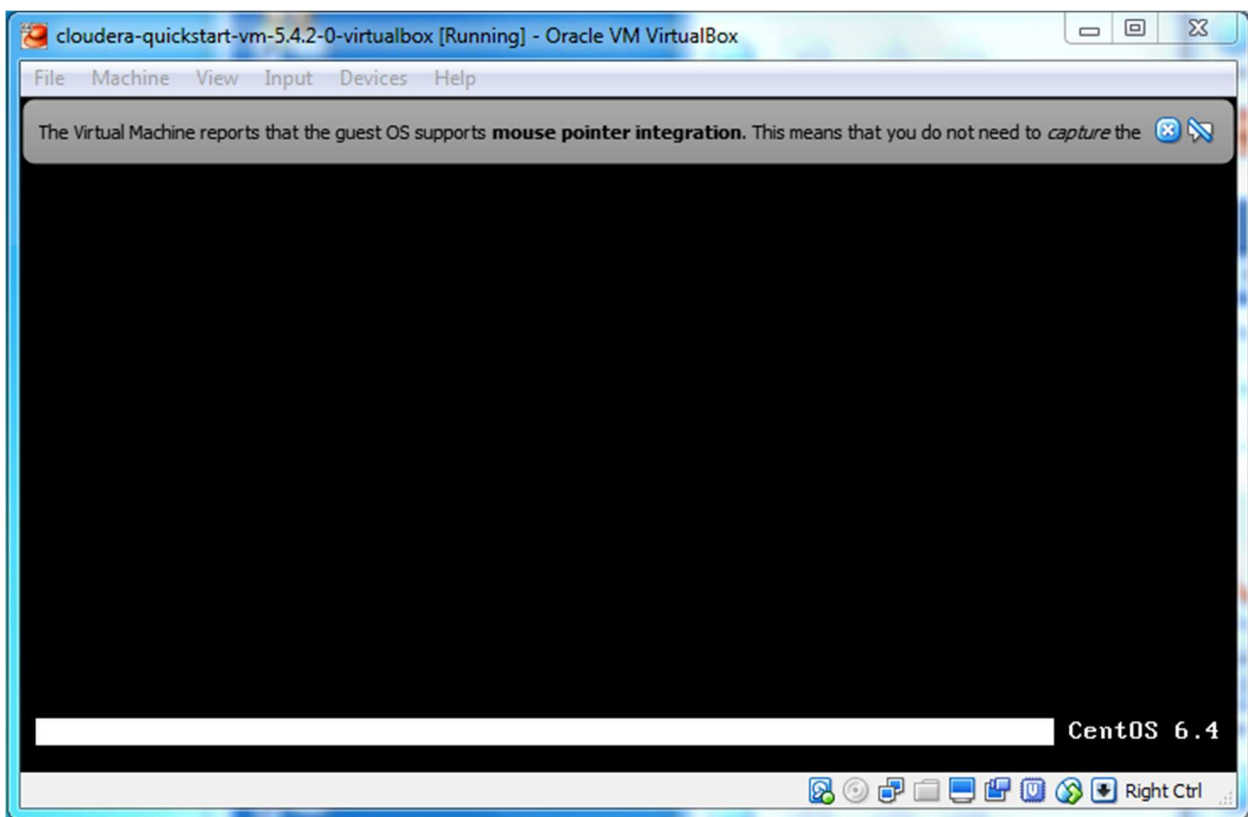
10. The virtual machine image will be imported. This can take several minutes.



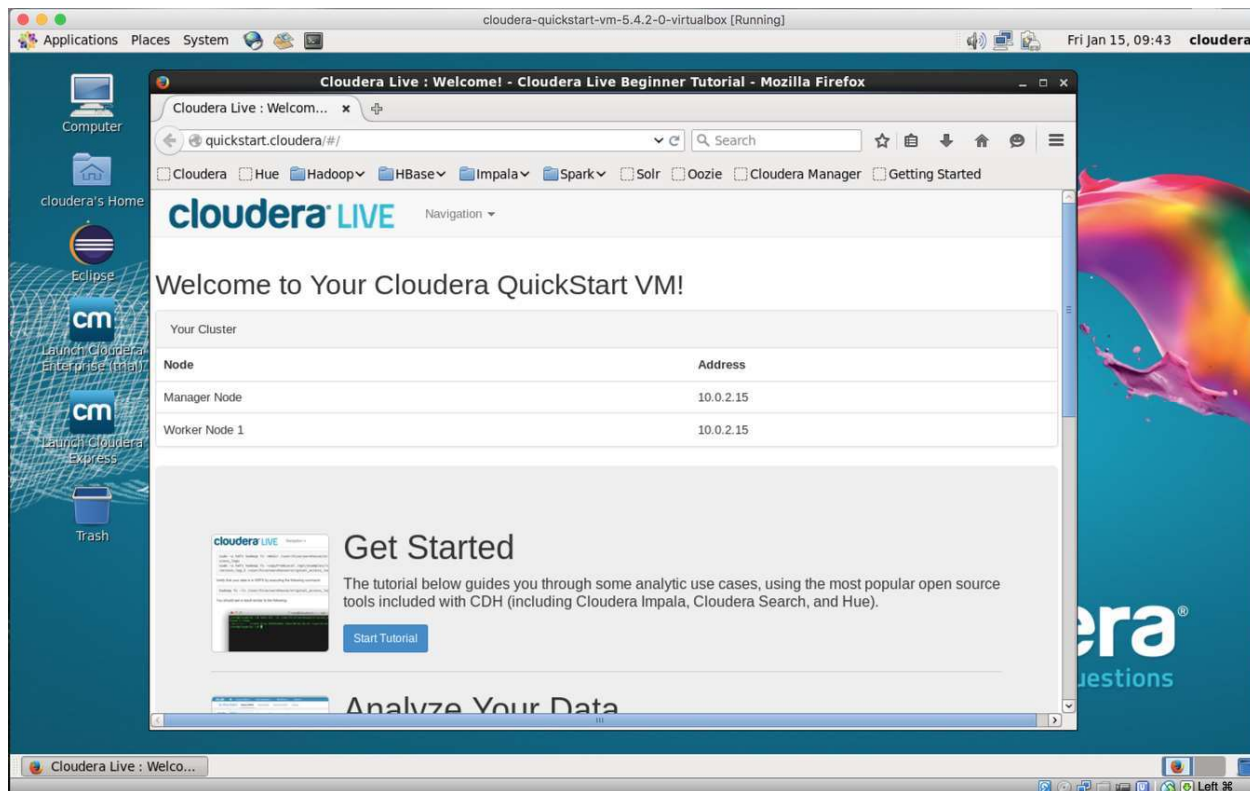
11. **Launch Cloudera VM.** When the importing is finished, the quickstart-vm-5.4.2-0 VM will appear on the left in the VirtualBox window. Select it and click the Start button to launch the VM.



12. **Cloudera VM booting.** It will take several minutes for the Virtual Machine to start. The booting process takes a long time since many Hadoop tools are started.



13. **The Cloudera VM desktop.** Once the booting process is complete, the desktop will appear with a browser.



b) MultipleNodes:

Install a Multi Node Hadoop Cluster on Ubuntu 16.04

This article is about multi-node installation of Hadoop cluster. You would need minimum of 2 ubuntu machines or virtual images to complete a multi-node installation. If you want to just try out a single node cluster, follow this article on [Installing Hadoop on Ubuntu 16.04](#).

I used Hadoop Stable version 2.6.0 for this article. I did this setup on a 3 node cluster. For simplicity, i will designate one node as **master**, and 2 nodes as **slaves (slave-1, and slave-2)**. Make sure all slave nodes are reachable from master node. To avoid any unreachable hosts error, make sure you add the slave hostnames and ip addresses in `/etc/hosts` file. Similarly, slave nodes should be able to resolve **master** hostname.

Installing Java on Master and Slaves

```
$sudo add-apt-repository ppa:webupd8team/java
$sudo apt-get update
$sudo apt-get install oracle-java7-installer
# Update Java runtime
$sudo update-java-alternatives -s java-7-oracle
```

Disable IPv6

As of now Hadoop does not support IPv6, and is tested to work only on IPv4 networks. If you are using IPv6, you need to switch Hadoop host machines to use IPv4. [The Hadoop Wiki](#) link provides a one liner command to disable the IPv6. If you are not using IPv6, skip this step:

```
sudo sed -i 's/net.ipv6.bindv6only\=\ 1/net.ipv6.bindv6only\=\ 0/'
/etc/sysctl.d/bindv6only.conf&& sudo invoke-rc.d procs restart
```

Setting up a Hadoop User