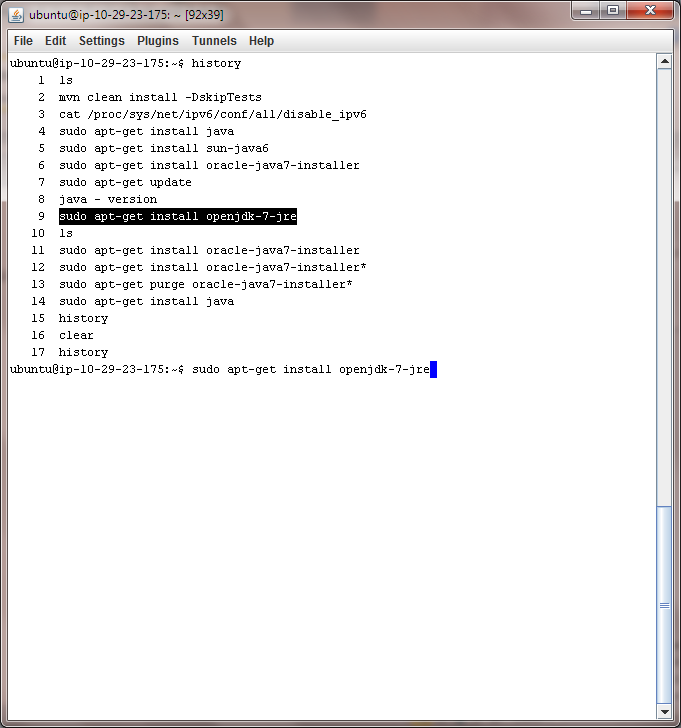
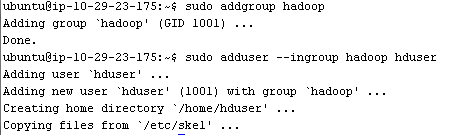
**Pre-requisite:**

**Installing Java in Ubuntu:**

Enter below command to install java. If you have any problems try this link <https://help.ubuntu.com/community/Java>.

**Dedicated User for Hadoop system user (Recommended):**

A dedicated user for Hadoop system will help in separating Hadoop installation from other applications on same machine. It also helps in improved security.

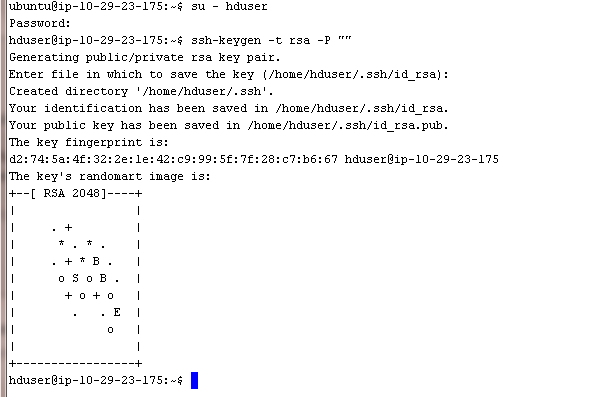


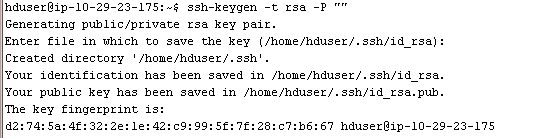
This will add the user **hduser** and the group **hadoop** to your local machine. Create a password and profile for the user.

**Configuring SSH:**

Hadoop requires SSH access to manage its nodes. For single node setup, we need to configure SSH access to **localhost** for **hduser** we created in before section.

Enter the below commands to create RSA Key pair with empty password. Generally, using an empty password is not recommended, but in this case it is needed to unlock the key without your interaction (you don’t want to enter the passphrase every time Hadoop interacts with its nodes).

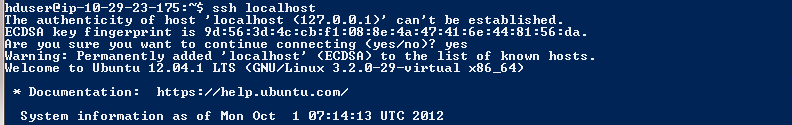




Now enable access ssh access to local machine with newly created RSA key pairs.



Test SSH connection to local host.

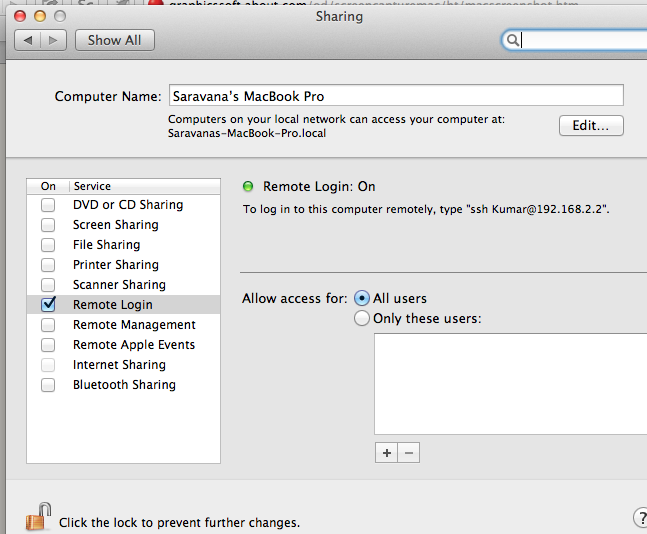


If the SSH connect fails, these general tips might help:

* Enable debugging with ssh -vvv localhost and investigate the error in detail.
* Check the SSH server configuration in /etc/ssh/sshd\_config, in particular the options Pubkey Authentication (which should be set to yes) and AllowUsers (if this option is active, add the hduser user to it). If you made any changes to the SSH server configuration file, you can force a configuration reload with sudo /etc/init.d/ssh reload.

**In Mac:**

You have to enable the ssh by enabling remote login in system preference > Sharing

****

**Hadoop and IPv6:**

Apache Hadoop does not currently support IPv6 networks; it uses IPv4 addresses for communicating between nodes. This is because Hadoop is designed to work in private datacenters, which usually have private IP addresses in the 10.x.x.x address space (Apache Hadoop, 2010).

**Disabling IPv6**

***[This section is referred from*** <http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-single-node-cluster/>]

To disable IPv6 on Ubuntu 12.04 LTS, open /etc/sysctl.conf in the editor of your choice and add the following lines to the end of the file:

#disable ipv6

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

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

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

You have to reboot your machine in order to make the changes take effect.

You can check whether IPv6 is enabled on your machine with the following command:

$ cat /proc/sys/net/ipv6/conf/all/disable\_ipv6

A return value of 0 means IPv6 is enabled, a value of 1 means disabled (that’s what we want).

**In Mac:**

Refer this document for <http://support.apple.com/kb/HT4667> disabling IPV6 on mac.

# Hadoop

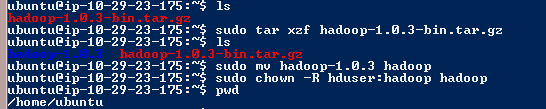
**Installation:**

Download Hadoop from its mirrors <http://www.apache.org/dyn/closer.cgi/hadoop/core>.

To download from latest Hadoop use below command



Extract the downloaded file and then move it a folder Hadoop.



Place your Hadoop folder in a folder you wish, here I am keeping it in /usr/local.



**Update $HOME/.bashrc**

Add the following lines to the end of the $HOME/.bashrc file of user hduser. If you use a shell other than bash, you should of course update its appropriate configuration files instead of .bashrc.

***[This section is referred from*** <http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-single-node-cluster/>]

**Below lines are useful shortcuts for Hadoop that you can use later, it not needed to run Hadoop.**

# Set Hadoop-related environment variables

export HADOOP\_HOME=/usr/local/hadoop

# Set JAVA\_HOME (we will also configure JAVA\_HOME directly for Hadoop later on)

export JAVA\_HOME=/usr/lib/jvm/java-6-sun

# Some convenient aliases and functions for running Hadoop-related commands

unalias fs &> /dev/null

alias fs="hadoop fs"

unalias hls &> /dev/null

alias hls="fs -ls"

# If you have LZO compression enabled in your Hadoop cluster and

# compress job outputs with LZOP (not covered in this tutorial):

# Conveniently inspect an LZOP compressed file from the command

# line; run via:

# $ lzohead /hdfs/path/to/lzop/compressed/file.lzo

# Requires installed 'lzop' command.

lzohead () {

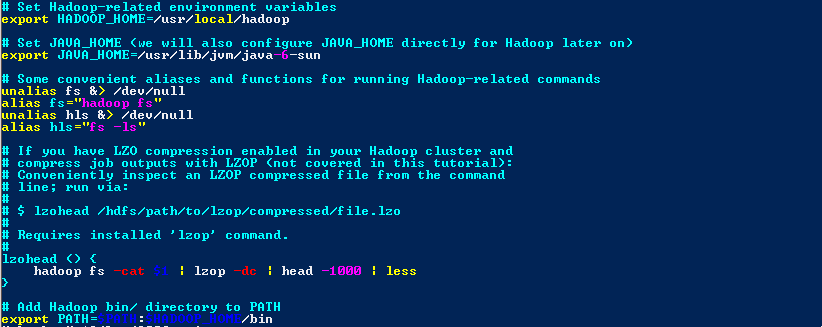
hadoop fs -cat $1 | lzop -dc | head -1000 | less

}

# Add Hadoop bin/ directory to PATH

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

You can repeat this exercise also for other users who want to use Hadoop.



Hadoop in Virtual Machine Environment:

Use this [Hadoop Instance](http://ydn.zenfs.com/site/hadoop/hadoop-vm-appliance-0-18-0_v1.zip) to install a preconfigured Linux virtual machine instead of manual setup.

**References:**

1. <http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-single-node-cluster/>
2. <http://wiki.apache.org/hadoop/HadoopIPv6>
3. <http://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/SingleCluster.html>