**Update the Source list:**

$sudo apt-get update

**Check if Java is installed**

$java -version

**Add a dedicated Hadoop user**

$ sudo addgroup hadoop

**Add user to group**

$sudo adduser --ingroup hadoop hduser

Adding user 'hduser' ...

Adding new user `hduser' (1001) with group `hadoop' ...

Creating home directory `/home/hduser' ...

Copying files from '/etc/skel' ...

Enter new UNIX password:

Retype new UNIX password:

passwd: password updated successfully

Changing the user information for hduser

Enter the new value, or press ENTER for the default

Full Name []:

Room Number []:

Work Phone []:

Home Phone []:

Other []:

Is the information correct? [Y/n] Y

**Install ssh**

$ sudo apt-get install ssh

Let’s check if it is installed properly:

$ which ssh

/usr/bin/ssh

$ which sshd

/usr/sbin/sshd

**Create and setup ssh certificates**

Hadoop requires ssh access to manage its nodes, i.e. remote machines plus our local machine. For our single-node setup of Hadoop, we, therefore need to configure SSH access to localhost.

So, we need to have ssh up and running on our machine and configured it to allow ssh public key authentication.

Hadoop uses ssh (to access its nodes) which would normally require the user to enter a password. However, this requirement can be eliminated by creating and setting up ssh certificates using the following commands. If asked for a filename just leave it blank and press the enter key to continue.

**switch user**

$ su hduser

Password:

hduser@ranjankumar:~$ ssh-keygen -t rsa -P ""

Generating public/private rsa key pair.

Enter file in which to save the key (/home/hduser/.ssh/id\_rsa):

Created directory '/home/hduser/.ssh'.

Your identification has been saved in /home/hduser/.ssh/id\_rsa.

Your public key has been saved in /home/hduser/.ssh/id\_rsa.pub.

The key fingerprint is:

SHA256:YHgzw/A+4JBwq1/iEoKrIzbD7UsP4Z1hotf4LrG6dmI hduser@ranjankumar

The key's randomart image is:

+---[RSA 2048]----+

|. . . |

| o o = |

| + o X |

|.. o = = |

|= .oooo S |

|.=o+B o. |

|+.+\*o+ |

|+Eo=+ |

|\*+Oo++ |

+----[SHA256]-----+

hduser@ranjankumar:~$ cat $HOME/.ssh/id\_rsa.pub >> $HOME/.ssh/authorized\_keys

The second command adds the newly created key to the list of authorized keys so that Hadoop can use ssh without prompting for a password.

We can check if ssh works:

hduser@ranjankumar:~$ ssh localhost

The authenticity of host 'localhost (127.0.0.1)' can't be established.

ECDSA key fingerprint is e1:8b:a0:a5:75:ef:f4:b4:5e:a9:ed:be:64:be:5c:2f.

Are you sure you want to continue connecting (yes/no)? yes

Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.

Welcome to Ubuntu 14.04.1 LTS (GNU/Linux 3.13.0-40-generic x86\_64)

...

**Install Hadoop**

hduser@ranjankumar:~$ wget http://mirrors.sonic.net/apache/hadoop/common/hadoop-2.6.0/hadoop-2.6.0.tar.gz

hduser@ranjankumar:~$ tar xvzf hadoop-2.6.0.tar.gz

**Move the Hadoop installation to the /usr/local/hadoop directory using the following command:**

hduser@ranjankumar:~/hadoop-2.6.0$ sudo mv \* /usr/local/hadoop

[sudo] password for hduser:

hduser is not in the sudoers file. This incident will be reported.

To resolve this error, login in as a root user, and then add hduser to sudo:

hduser@ranjankumar:~/hadoop-2.6.0$ su ranjank

Password:

ranjank@ranjankumar:/home/hduser$ sudo adduser hduser sudo

[sudo] password for k:

Adding user `hduser' to group `sudo' ...

Adding user hduser to group sudo

Done.

Let’s now move the Hadoop installation to the /usr/local/hadoop directory:

**ranjank@ranjankumar:~$ sudo su hduser**

**hduser@ranjankumar:~/hadoop-2.6.0$ sudo mv \* /usr/local/hadoop**

**hduser@ranjankumar:~/hadoop-2.6.0$ sudo chown -R hduser:hadoop /usr/local/hadoop**

**Setup Configuration Files**

We need to modify following files to complete the Hadoop setup:

1. ~/.bashrc

2. /usr/local/hadoop/etc/hadoop/hadoop-env.sh

3. /usr/local/hadoop/etc/hadoop/core-site.xml

4. /usr/local/hadoop/etc/hadoop/mapred-site.xml.template

5. /usr/local/hadoop/etc/hadoop/hdfs-site.xml

**~/.bashrc:**

**hduser@ranjankumar:~$ nano ~/.bashrc**

#HADOOP VARIABLES START

export JAVA\_HOME=/usr/lib/jvm/jdk1.7.0\_79

export HADOOP\_INSTALL=/usr/local/hadoop

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

export PATH=$PATH:$HADOOP\_INSTALL/sbin

export HADOOP\_MAPRED\_HOME=$HADOOP\_INSTALL

export HADOOP\_COMMON\_HOME=$HADOOP\_INSTALL

export HADOOP\_HDFS\_HOME=$HADOOP\_INSTALL

export YARN\_HOME=$HADOOP\_INSTALL

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_INSTALL/lib/native

export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_INSTALL/lib"

#HADOOP VARIABLES END

hduser@ranjankumar:~$ source ~/.bashrc

7.2. /usr/local/hadoop/etc/hadoop/hadoop-env.sh

We need to set JAVA\_HOME by modifying hadoop-env.sh file.

**hduser@ranjankumar:~$ vi /usr/local/hadoop/etc/hadoop/hadoop-env.sh**

...

export JAVA\_HOME=/usr/lib/jvm/jdk1.7.0\_79

...

**/usr/local/hadoop/etc/hadoop/core-site.xml:**

The /usr/local/hadoop/etc/hadoop/core-site.xml file contains configuration properties that Hadoop uses when starting up. This file can be used to override the default settings that Hadoop starts with.

**hduser@ranjankumar:~$ sudo mkdir -p /app/hadoop/tmp**

**hduser@ranjankumar:~$ sudo chown hduser:hadoop /app/hadoop/tmp**

Open the file and enter the following in between the <configuration> </configuration> tag:

**hduser@ranjankumar:~$ vi /usr/local/hadoop/etc/hadoop/core-site.xml**

<configuration>

<property>

<name>hadoop.tmp.dir</name>

<value>/app/hadoop/tmp</value>

<description>A base for other temporary directories.</description>

</property>

<property>

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

<value>hdfs://localhost:54310</value>

<description>The name of the default file system. A URI whose

scheme and authority determine the FileSystem implementation. The

uri's scheme determines the config property (fs.SCHEME.impl) naming

the FileSystem implementation class. The uri's authority is used to

determine the host, port, etc. for a filesystem.</description>

</property>

</configuration>

**/usr/local/hadoop/etc/hadoop/mapred-site.xml**

By default, the /usr/local/hadoop/etc/hadoop/ folder contains /usr/local/hadoop/etc/hadoop/mapred-site.xml.template file which has to be renamed/copied with the name mapred-site.xml:

**hduser@ranjankumar:~$ cp /usr/local/hadoop/etc/hadoop/mapred-site.xml.template /usr/local/hadoop/etc/hadoop/mapred-site.xml**

The mapred-site.xml file is used to specify which framework is being used for MapReduce. We need to enter the following content in between the <configuration> </configuration> tag: