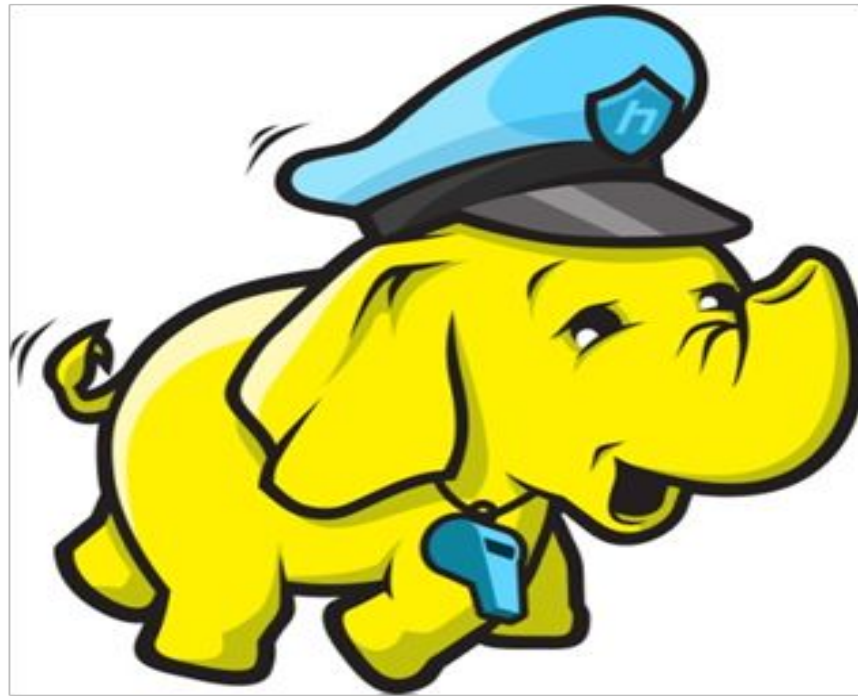


# Install Hadoop on Linux based OS

## Mac OS/Redhat/Ubuntu



For online Hadoop training, send mail to [neeraj.ymca.2k6@gmail.com](mailto:neeraj.ymca.2k6@gmail.com)

# Agenda

- Install Java & ssh on your Ubuntu
- Visit Apache Hadoop website
- Download Hadoop tar.gz file
- Untar the Hadoop tar.gz file
- Directory structure of Hadoop
- What is ssh (Secure Shell)
- How to configure password less ssh
- Configuration files in Hadoop
- Modify configuration files
- Format Namenode
- Start the services of Hadoop
- Verify Hadoop services
- Stop the services of Hadoop

# Install Java & ssh on your Ubuntu

**Run the below commands in terminal to install Java**

```
sudo add-apt-repository ppa:webupd8team/java  
sudo apt-get update  
sudo apt-get install oracle-java7-installer
```

**Run the below command in terminal to install ssh**


```
sudo apt-get install openssh-server openssh-client
```



**Make sure that your Ubuntu is connected to internet during execution of these commands**















# Download Hadoop tar.gz file

<http://www.motorlogy.com/apache/hadoop/common/hadoop-1.2.1/>

Index of /apache/hadoop/co... 


← [www.motorlogy.com/apache/hadoop/common/hadoop-1.2.1/](http://www.motorlogy.com/apache/hadoop/common/hadoop-1.2.1/)


## Index of /apache/hadoop/common/hadoop-1.2.1

Name	Last modified	Size	Description
 <a href="#">Parent Directory</a>	-	-	
 <a href="#">hadoop-1.2.1-1.i386.rpm</a>	22-Jul-2013 15:20	39M	
 <a href="#">hadoop-1.2.1-1.i386.rpm.mds</a>	22-Jul-2013 15:45	1.1K	
 <a href="#">hadoop-1.2.1-1.x86_64.rpm</a>	22-Jul-2013 15:30	35M	
 <a href="#">hadoop-1.2.1-1.x86_64.rpm.mds</a>	22-Jul-2013 15:45	1.1K	
 <a href="#">hadoop-1.2.1-bin.tar.gz</a>	22-Jul-2013 15:27	36M	
 <a href="#">hadoop-1.2.1-bin.tar.gz.mds</a>	22-Jul-2013 15:45	1.1K	
 <a href="#">hadoop-1.2.1.tar.gz</a>	22-Jul-2013 15:27	61M	
 <a href="#">hadoop-1.2.1.tar.gz.mds</a>	22-Jul-2013 15:45	958	
 <a href="#">hadoop_1.2.1-1_i386.deb</a>	22-Jul-2013 15:23	39M	
 <a href="#">hadoop_1.2.1-1_i386.deb.mds</a>	22-Jul-2013 15:45	1.1K	
 <a href="#">hadoop_1.2.1-1_x86_64.deb</a>	22-Jul-2013 15:33	35M	
 <a href="#">hadoop_1.2.1-1_x86_64.deb.mds</a>	22-Jul-2013 15:45	1.1K	
 <a href="#">releasenotes-1.2.1.html</a>	22-Jul-2013 15:15	519K	















Apache/2.2.22 Server at www.motorlogy.com Port 80

# Click on Save File

Index of /apache/hadoop/co... 

←  www.motorlogy.com/apache/hadoop/common/hadoop-1.2.1/


## Index of /apache/hadoop/common/hadoop-1.2.1

Name	Last modified	Size	Description
 <a href="#">Parent Directory</a>	-	-	-
 <a href="#">hadoop-1.2.1-1.i386.rpm</a>	22-Jul-2013 15:20	39M	
 <a href="#">hadoop-1.2.1-1.i386.rpm.mds</a>	22-Jul-2013 15:45	1.1K	
 <a href="#">hadoop-1.2.1-1.x86_64.rpm</a>	22-Jul-2013 15:45	39M	
 <a href="#">hadoop-1.2.1-1.x86_64.rpm.mds</a>	22-Jul-2013 15:45	1.1K	
 <a href="#">hadoop-1.2.1-bin.tar.gz</a>	22-Jul-2013 15:45	60.9 MB	
 <a href="#">hadoop-1.2.1-bin.tar.gz.mds</a>	22-Jul-2013 15:45	1.1K	
 <a href="#">hadoop-1.2.1.tar.gz</a>	22-Jul-2013 15:45	60.9 MB	
 <a href="#">hadoop-1.2.1.tar.gz.mds</a>	22-Jul-2013 15:45	1.1K	
 <a href="#">hadoop_1.2.1-1_i386.deb</a>	22-Jul-2013 15:45	39M	
 <a href="#">hadoop_1.2.1-1_i386.deb.mds</a>	22-Jul-2013 15:45	1.1K	
 <a href="#">hadoop_1.2.1-1_x86_64.deb</a>	22-Jul-2013 15:45	39M	
 <a href="#">hadoop_1.2.1-1_x86_64.deb.mds</a>	22-Jul-2013 15:45	1.1K	
 <a href="#">releasenotes-1.2.1.html</a>	22-Jul-2013 15:45	1.1K	

Apache/2.2.22 Server at www.motorlogy.com

**Opening hadoop-1.2.1.tar.gz**

You have chosen to open:

 **hadoop-1.2.1.tar.gz**  
which is a: Gzip archive (60.9 MB)  
from: http://www.motorlogy.com

**What should Firefox do with this file?**

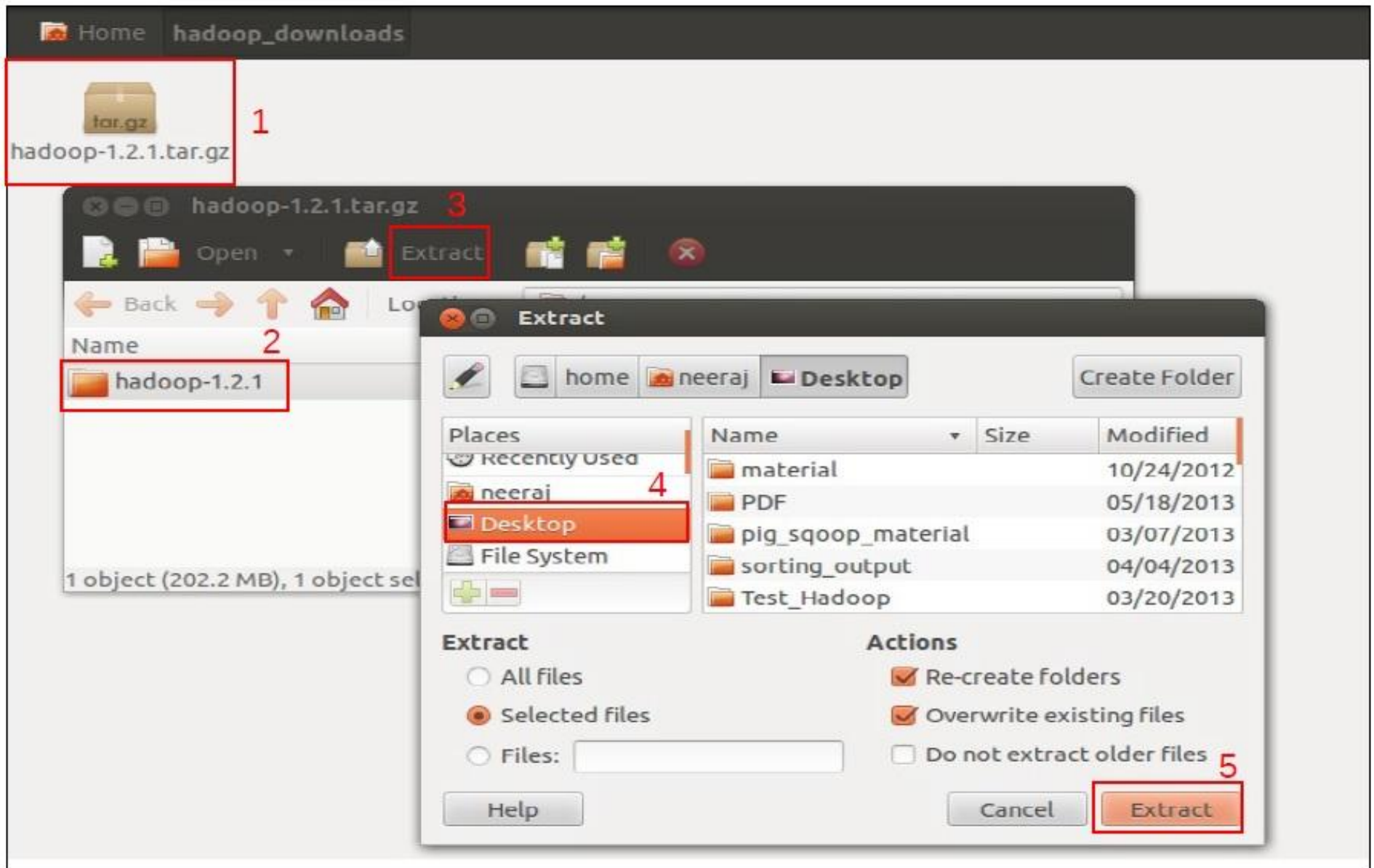
☐ Open with Archive Manager (default)

☒ **Save File**

☐ Do this automatically for files like this from now on.

Cancel OK

# Untar the Hadoop tar.gz file





# Directory structure of Hadoop

## **HADOOP\_HOME**

- bin [ contains scripts to start/stop Hadoop ]
- conf [ contains configuration files of Hadoop ]
- lib [ contains libraries to run Hadoop ]
- src [ contains the source code of Hadoop ]
- logs [ contains log files ]

# Configuration files in Hadoop

There are two types of important configuration files in Hadoop...

## **Default configuration files ( Read-only ) :**

src/core/core-default.xml  
src/hdfs/hdfs-default.xml  
src/mapred/mapred-default.xml

## **Site-specific configuration files ( can be modified ) :**

conf/core-site.xml  
conf/hdfs-site.xml  
conf/mapred-site.xml



# What is SSH ( Secure Shell )

Tool, used to remotely connect to other machine.

Namenode uses SSH for communicating with Datanodes.

Jobtracker uses SSH for communicating with Tasktrackers.

SSH must be installed, even if you are installing Hadoop on a single machine.

`ssh < IP of another machine >`

For example...

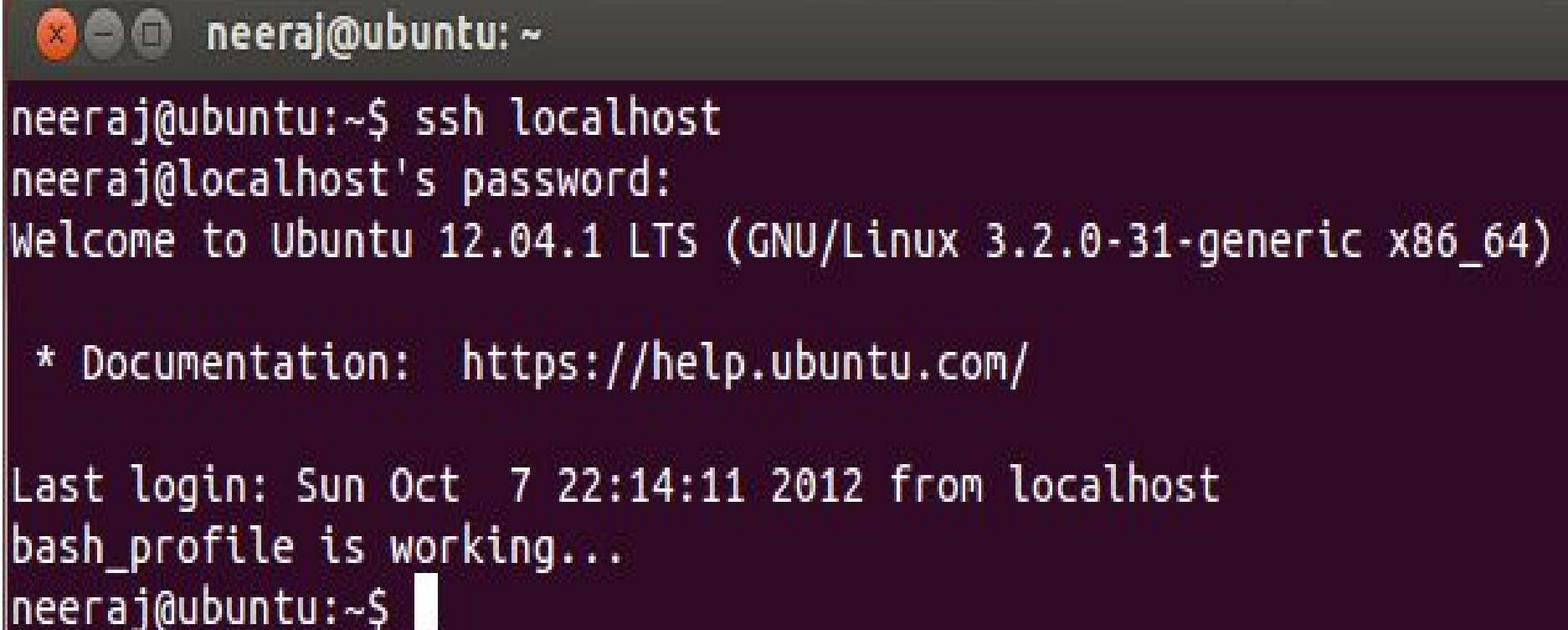
`ssh 192.168.342.187`

`ssh 127.0.0.1`

`ssh localhost`

Run the following command on your terminal

`ssh localhost`



A terminal window with a dark background and light text. The title bar shows three window control icons (close, minimize, maximize) and the text 'neeraj@ubuntu: ~'. The terminal content shows the user 'neeraj' at 'ubuntu' running the command 'ssh localhost'. It prompts for a password, then displays a welcome message for Ubuntu 12.04.1 LTS, including the kernel version and architecture. It also provides a link to the documentation. Finally, it shows the last login time and that the bash profile is working, returning to the shell prompt.

```
neeraj@ubuntu: ~  
neeraj@ubuntu:~$ ssh localhost  
neeraj@localhost's password:  
Welcome to Ubuntu 12.04.1 LTS (GNU/Linux 3.2.0-31-generic x86_64)  
  
* Documentation:  https://help.ubuntu.com/  
  
Last login: Sun Oct  7 22:14:11 2012 from localhost  
bash_profile is working...  
neeraj@ubuntu:~$
```

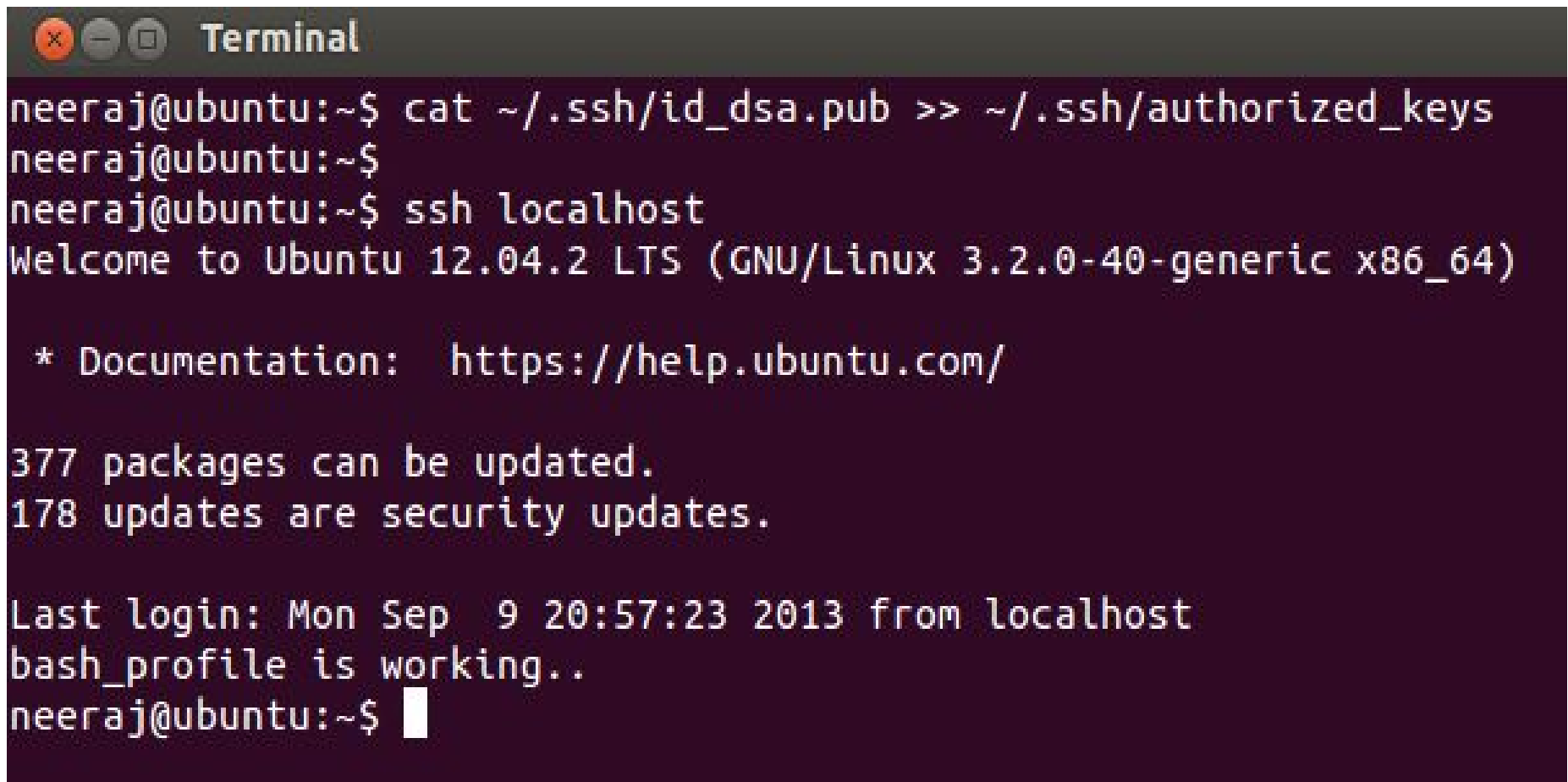
## Run this command from your home directory

```
ssh-keygen -t dsa -P "" -f ~/.ssh/id_dsa
```

```
neeraj@ubuntu: ~  
neeraj@ubuntu:~$ ssh-keygen -t dsa -P '' -f ~/.ssh/id_dsa  
Generating public/private dsa key pair.  
Your identification has been saved in /home/neeraj/.ssh/id_dsa.  
Your public key has been saved in /home/neeraj/.ssh/id_dsa.pub.  
The key fingerprint is:  
e8:71:21:1e:5f:47:6f:29:1f:bb:6b:8f:af:1f:9d:95 neeraj@ubuntu  
The key's randomart image is:  
+--[ DSA 1024]-----+  
|           .           |  
|        . . .         |  
|       o . . o =      |  
|      . = o . + o.    |  
|     + S       oE.    |  
|    . o         .+    |  
|     .          .o.   |  
|                 .o.  |  
|                 .+=+  |  
+-----+  
neeraj@ubuntu:~$
```

Run this command from your home directory

```
cat ~/.ssh/id_dsa.pub >> ~/.ssh/authorized_keys
```



```
Terminal
neeraj@ubuntu:~$ cat ~/.ssh/id_dsa.pub >> ~/.ssh/authorized_keys
neeraj@ubuntu:~$
neeraj@ubuntu:~$ ssh localhost
Welcome to Ubuntu 12.04.2 LTS (GNU/Linux 3.2.0-40-generic x86_64)

 * Documentation:  https://help.ubuntu.com/

377 packages can be updated.
178 updates are security updates.

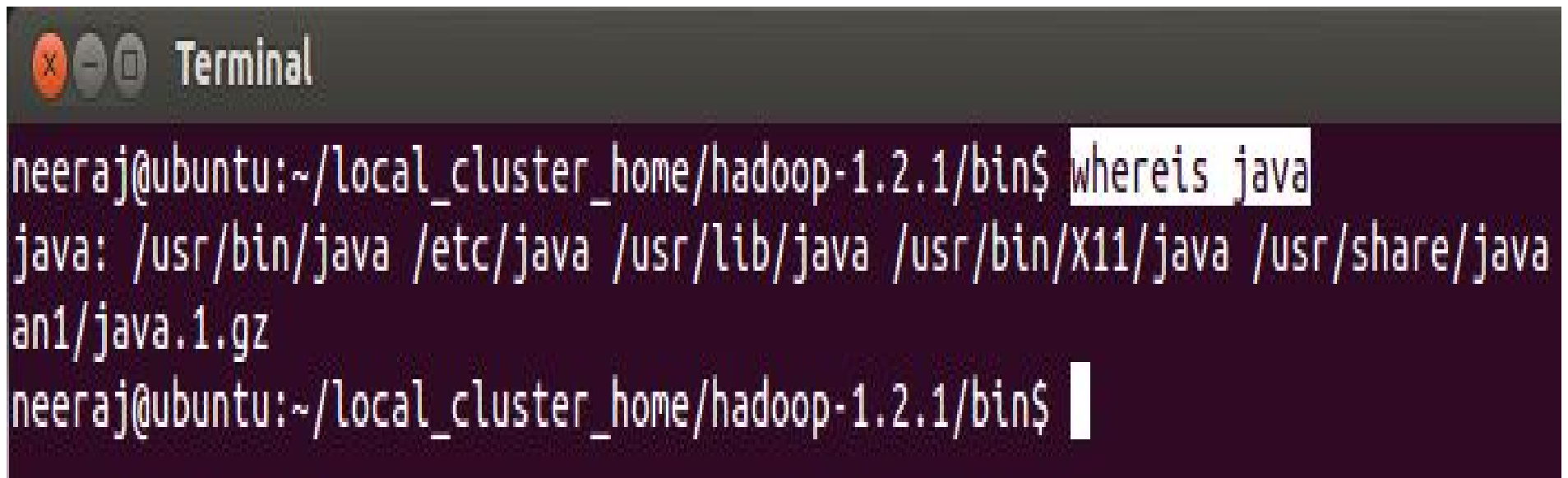
Last login: Mon Sep  9 20:57:23 2013 from localhost
bash_profile is working..
neeraj@ubuntu:~$
```

# Modify hadoop-env.sh file

**`hadoop-env.sh` file contains Hadoop environment settings.**

Make the following entry in `HADOOP_HOME/conf/hadoop-env.sh`

`export JAVA_HOME=/usr`



```
Terminal
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$ whereis java
java: /usr/bin/java /etc/java /usr/lib/java /usr/bin/X11/java /usr/share/java
an1/java.1.gz
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$
```

# Modify core-site.xml file

Make the following entries in HADOOP\_HOME/conf/**core-site.xml**...

```
<configuration>
```

```
  <property>
```

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

```
    <value>hdfs://localhost:9000</value>
```

```
  </property>
```

```
  <property>
```

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

```
    <value>/home/neeraj/local_cluster_home/HDFS/hdfs_temp</value>
```

```
  </property>
```

```
</configuration>
```

# Modify hdfs-site.xml file

Make the following entries in HADOOP\_HOME/conf/hdfs-site.xml...

```
<configuration>
```

```
  <property>
```

```
    <name>dfs.replication</name>
```

```
    <value>1</value>
```

```
  </property>
```

```
  <property>
```

```
    <name>dfs.name.dir</name>
```

```
    <value>/home/neeraj/local_cluster_home/HDFS/hdfs_metadata</value>
```

```
  </property>
```

```
  <property>
```

```
    <name>dfs.data.dir</name>
```

```
    <value>/home/neeraj/local_cluster_home/HDFS/hdfs_data</value>
```

```
  </property>
```

```
</configuration>
```



# Modify mapred-site.xml file

Make the following entries in `HADOOP_HOME/conf/mapred-site.xml`...

```
<configuration>
```

```
  <property>
```

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

```
    <value>localhost:9001</value>
```

```
    <description>The host and port on which Job tracker runs.
```

```
    </description>
```

```
  </property>
```

```
</configuration>
```

# Format your Namenode

Run the following command from bin directory of Hadoop

`./hadoop namenode -format`

```
neeraj@myubuntu:~/local_cluster_home/hadoop-1.2.1/bin$ pwd
/home/neeraj/local_cluster_home/hadoop-1.2.1/bin
neeraj@myubuntu:~/local_cluster_home/hadoop-1.2.1/bin$ ./hadoop namenode -format
15/11/26 08:53:23 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = myubuntu/127.0.0.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 1.2.1
STARTUP_MSG:   build = https://svn.apache.org/repos/asf/hadoop/common/branches/branch-1.2 -r 1503152; comp
iled by 'mattf' on Mon Jul 22 15:23:09 PDT 2013
STARTUP_MSG:   java = 1.6.0_27
*****/
15/11/26 08:53:23 INFO util.GSet: Computing capacity for map BlocksMap
15/11/26 08:53:23 INFO util.GSet: VM type           = 32-bit
15/11/26 08:53:23 INFO util.GSet: 2.0% max memory = 1013645312
15/11/26 08:53:23 INFO util.GSet: capacity          = 2^22 = 4194304 entries
15/11/26 08:53:23 INFO util.GSet: recommended=4194304, actual=4194304
15/11/26 08:53:23 INFO namenode.FSNamesystem: fsOwner=neeraj
15/11/26 08:53:23 INFO namenode.FSNamesystem: supergroup=supergroup
15/11/26 08:53:23 INFO namenode.FSNamesystem: isPermissionEnabled=true
15/11/26 08:53:23 INFO namenode.FSNamesystem: dfs.block.invalidate.limit=100
15/11/26 08:53:23 INFO namenode.FSNamesystem: isAccessTokenEnabled=false accessKeyUpdateInterval=0 min(s),
  accessTokenLifetime=0 min(s)
15/11/26 08:53:23 INFO namenode.FSEditLog: dfs.namenode.edits.toleration.length = 0
15/11/26 08:53:23 INFO namenode.NameNode: Caching file names occurring more than 10 times
15/11/26 08:53:23 INFO common.Storage: Image file /home/neeraj/local_cluster_home/HDFS/hdfs_metadata/curre
nt/fsimage of size 112 bytes saved in 0 seconds.
15/11/26 08:53:24 INFO namenode.FSEditLog: closing edit log: position=4, editlog=/home/neeraj/local_cluste
r_home/HDFS/hdfs_metadata/current/edits
15/11/26 08:53:24 INFO namenode.FSEditLog: close success: truncate to 4, editlog=/home/neeraj/local_cluste
r_home/HDFS/hdfs_metadata/current/edits
15/11/26 08:53:25 INFO common.Storage: Storage directory /home/neeraj/local_cluster_home/HDFS/hdfs_metadat
a has been successfully formatted.
15/11/26 08:53:25 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at myubuntu/127.0.0.1
*****/
neeraj@myubuntu:~/local_cluster_home/hadoop-1.2.1/bin$
```



# Start the services of Hadoop

Run the following command from bin directory of Hadoop

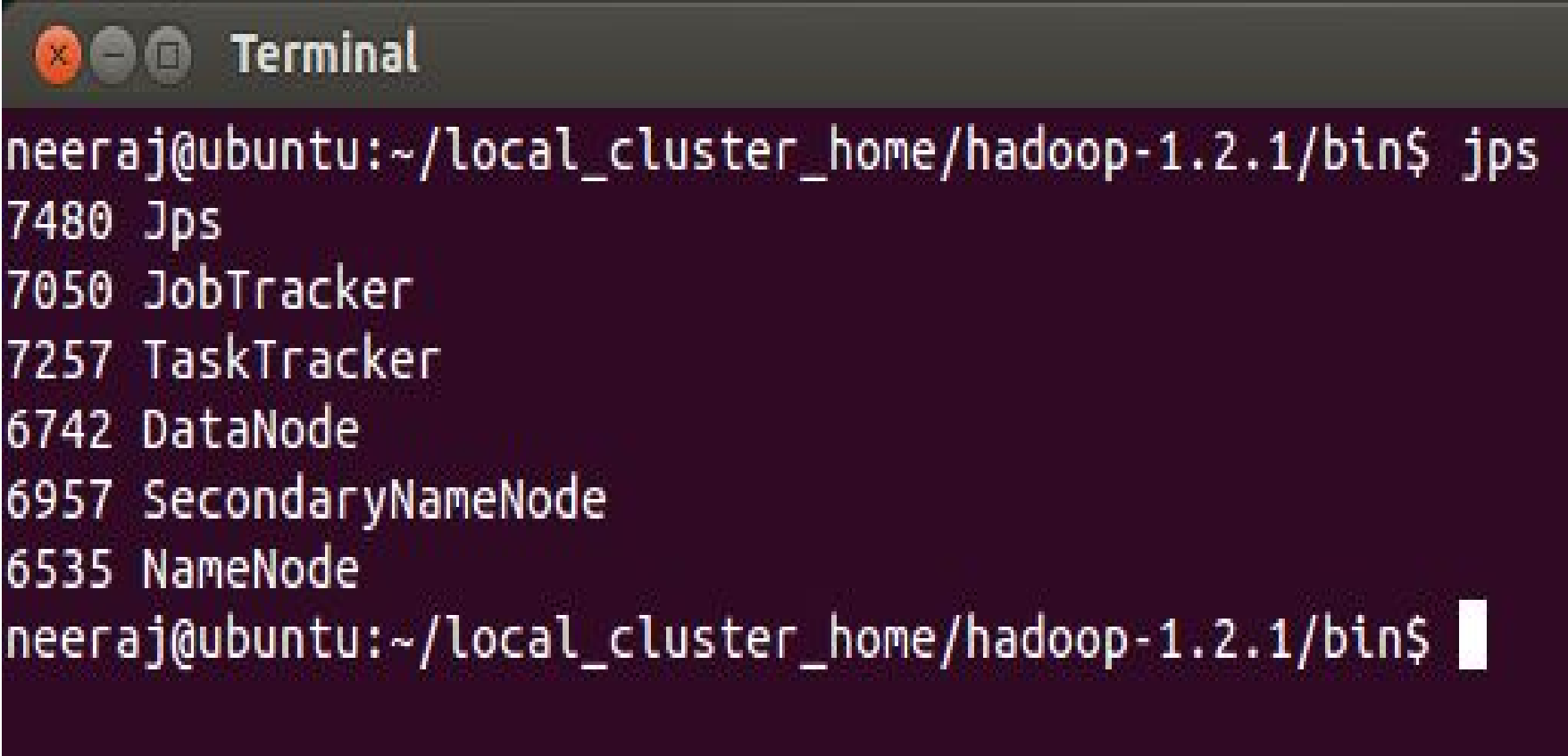
`./start-all.sh`

```
Terminal
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$ pwd
/home/neeraj/local_cluster_home/hadoop-1.2.1/bin
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$ ./start-all.sh
starting namenode, logging to /home/neeraj/local_cluster_home/hadoop-1.2.1/libexec/../logs/hadoop-neeraj-namenode-ubuntu.out
myubuntu: starting datanode, logging to /home/neeraj/local_cluster_home/hadoop-1.2.1/libexec/../logs/hadoop-neeraj-datanode-ubuntu.out
myubuntu: starting secondarynamenode, logging to /home/neeraj/local_cluster_home/hadoop-1.2.1/libexec/../logs/hadoop-neeraj-secondarynamenode-ubuntu.out
starting jobtracker, logging to /home/neeraj/local_cluster_home/hadoop-1.2.1/libexec/../logs/hadoop-neeraj-jobtracker-ubuntu.out
myubuntu: starting tasktracker, logging to /home/neeraj/local_cluster_home/hadoop-1.2.1/libexec/../logs/hadoop-neeraj-tasktracker-ubuntu.out
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$
```

# Verify Hadoop services

Run the following command from bin directory of Hadoop

jps

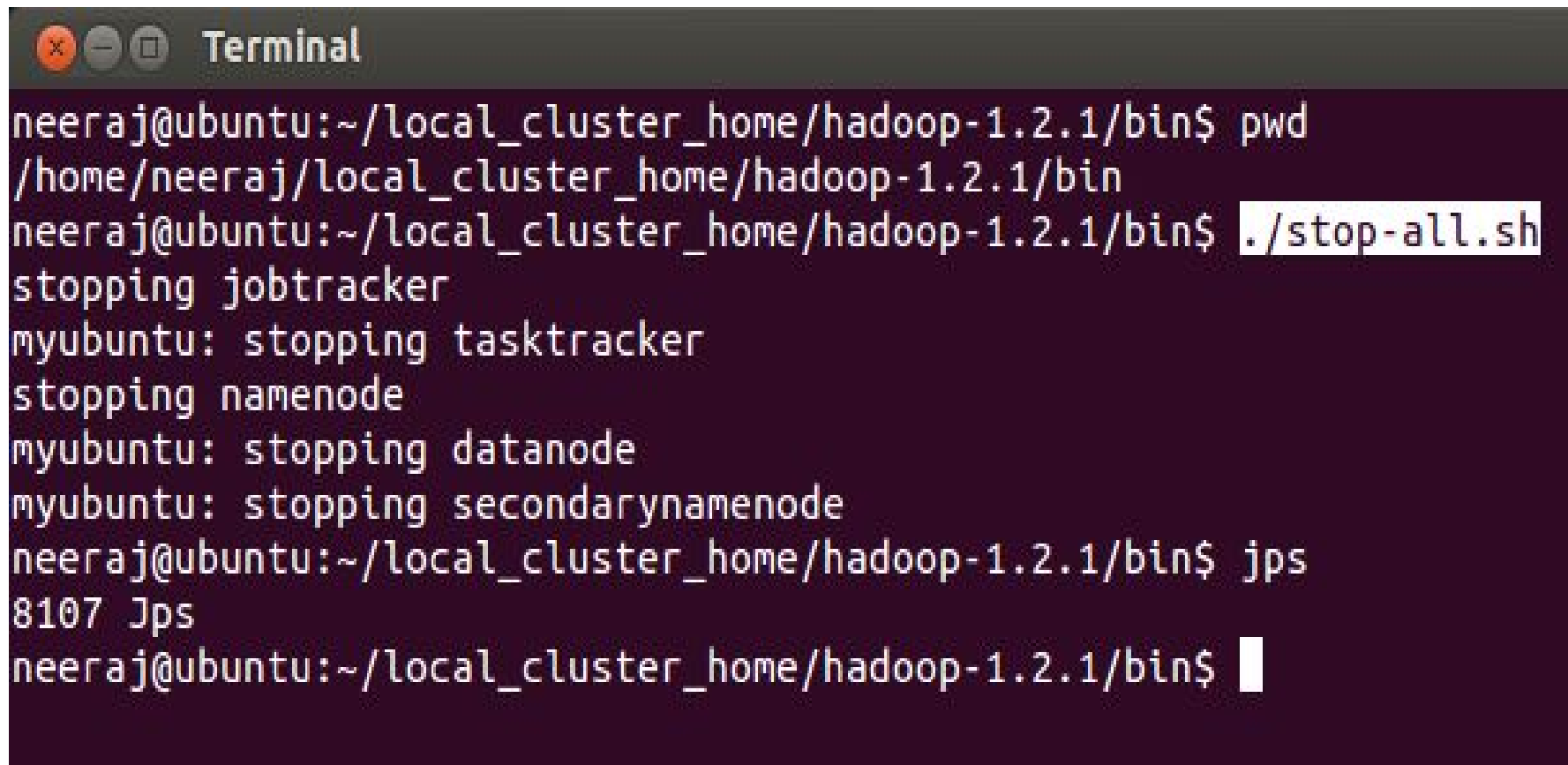
A terminal window titled "Terminal" with standard Ubuntu window controls (close, minimize, maximize). The prompt is "neeraj@ubuntu:~/local\_cluster\_home/hadoop-1.2.1/bin\$". The command "jps" has been executed, and the output lists six Hadoop processes with their respective PIDs: Jps (7480), JobTracker (7050), TaskTracker (7257), DataNode (6742), SecondaryNameNode (6957), and NameNode (6535). The prompt is now "neeraj@ubuntu:~/local\_cluster\_home/hadoop-1.2.1/bin\$" with a cursor at the end.

```
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$ jps
7480 Jps
7050 JobTracker
7257 TaskTracker
6742 DataNode
6957 SecondaryNameNode
6535 NameNode
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$
```

# Stop the services of Hadoop

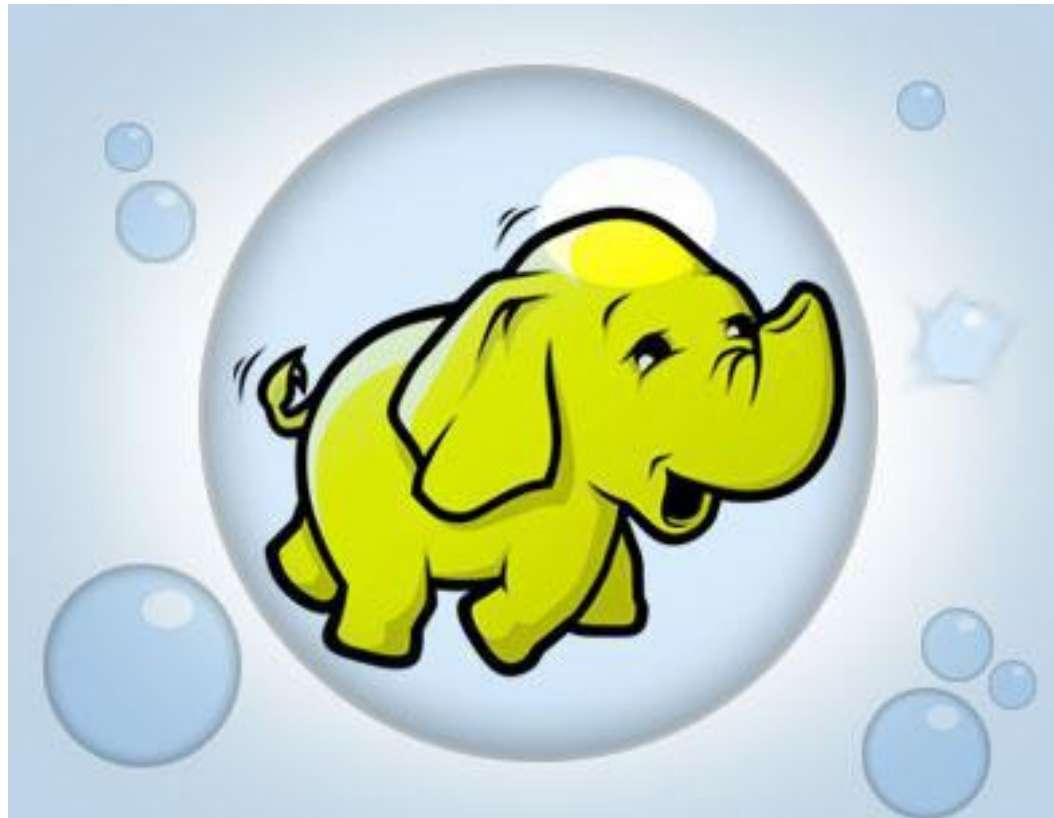
Run the following command from bin directory of Hadoop

`./stop-all.sh`



```
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$ pwd
/home/neeraj/local_cluster_home/hadoop-1.2.1/bin
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$ ./stop-all.sh
stopping jobtracker
myubuntu: stopping tasktracker
stopping namenode
myubuntu: stopping datanode
myubuntu: stopping secondarynamenode
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$ jps
8107 Jps
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$
```

...Thanks...



For online Hadoop training, send mail to [neeraj.ymca.2k6@gmail.com](mailto:neeraj.ymca.2k6@gmail.com)