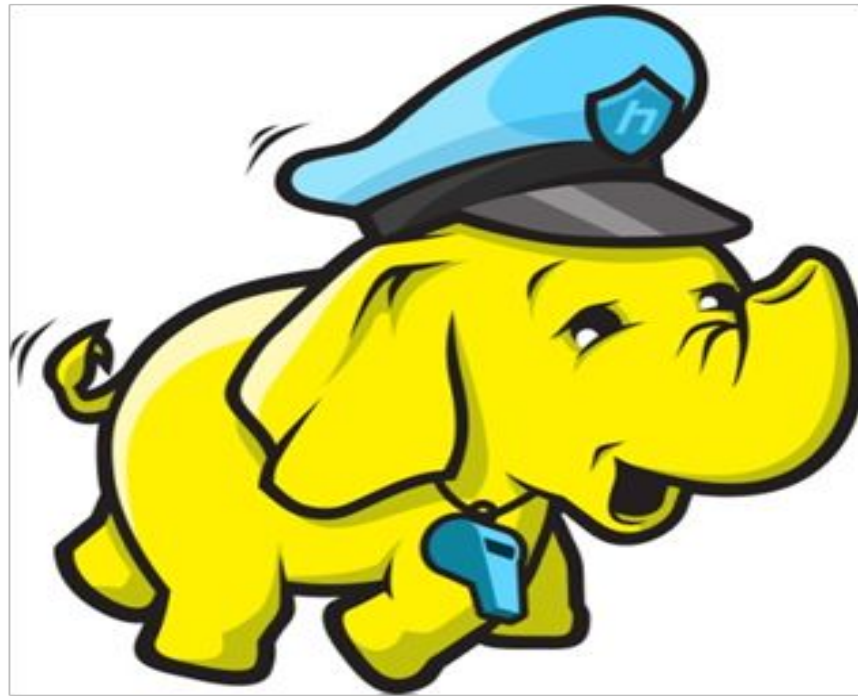# Install Hadoop on Linux based OS

Mac OS/Redhat/Ubuntu

For online Hadoop training, send mail to 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/

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

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

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

Click on Save File

Untar the Hadoop tar.gz file
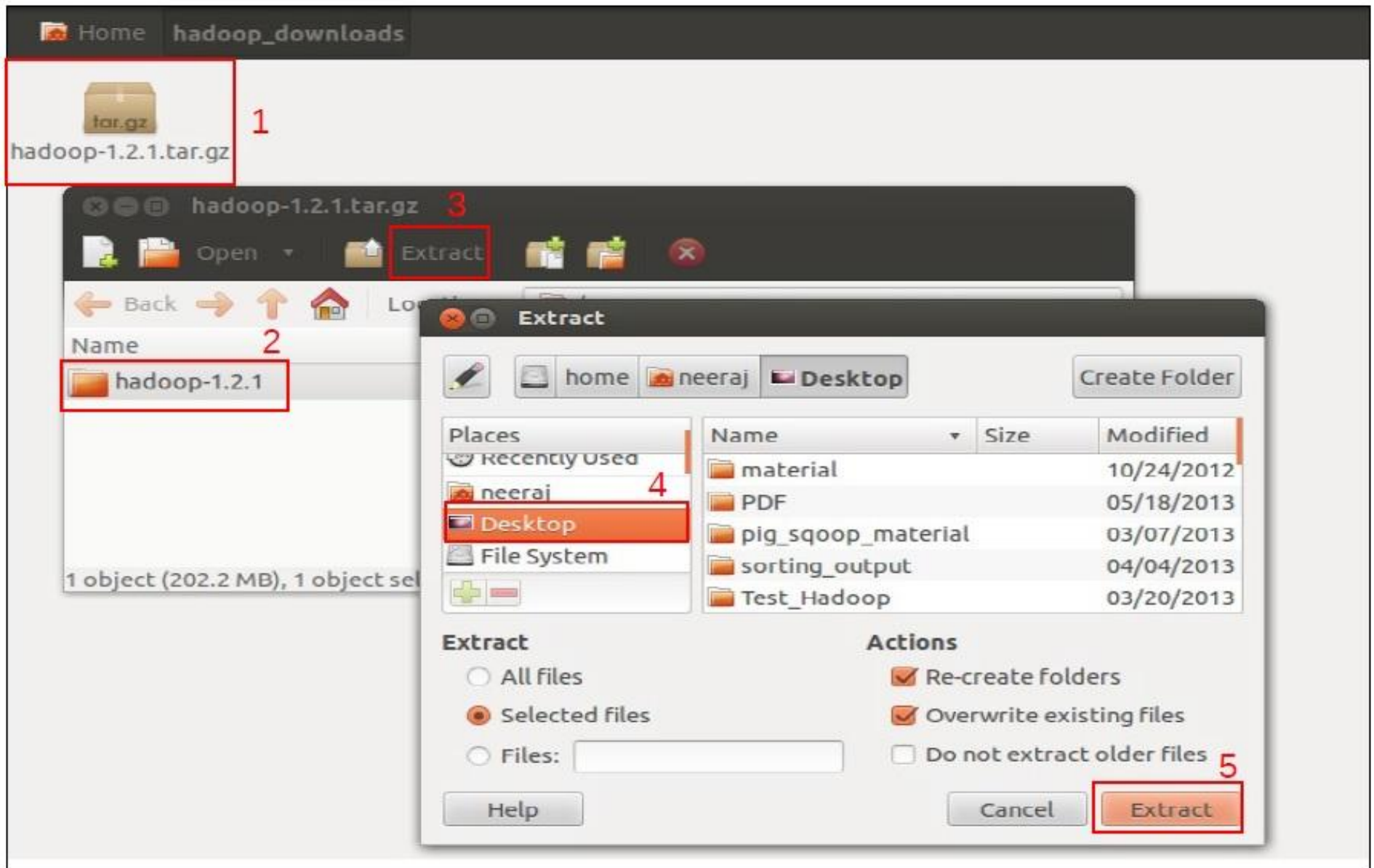
Directory structure of Hadoop

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 >
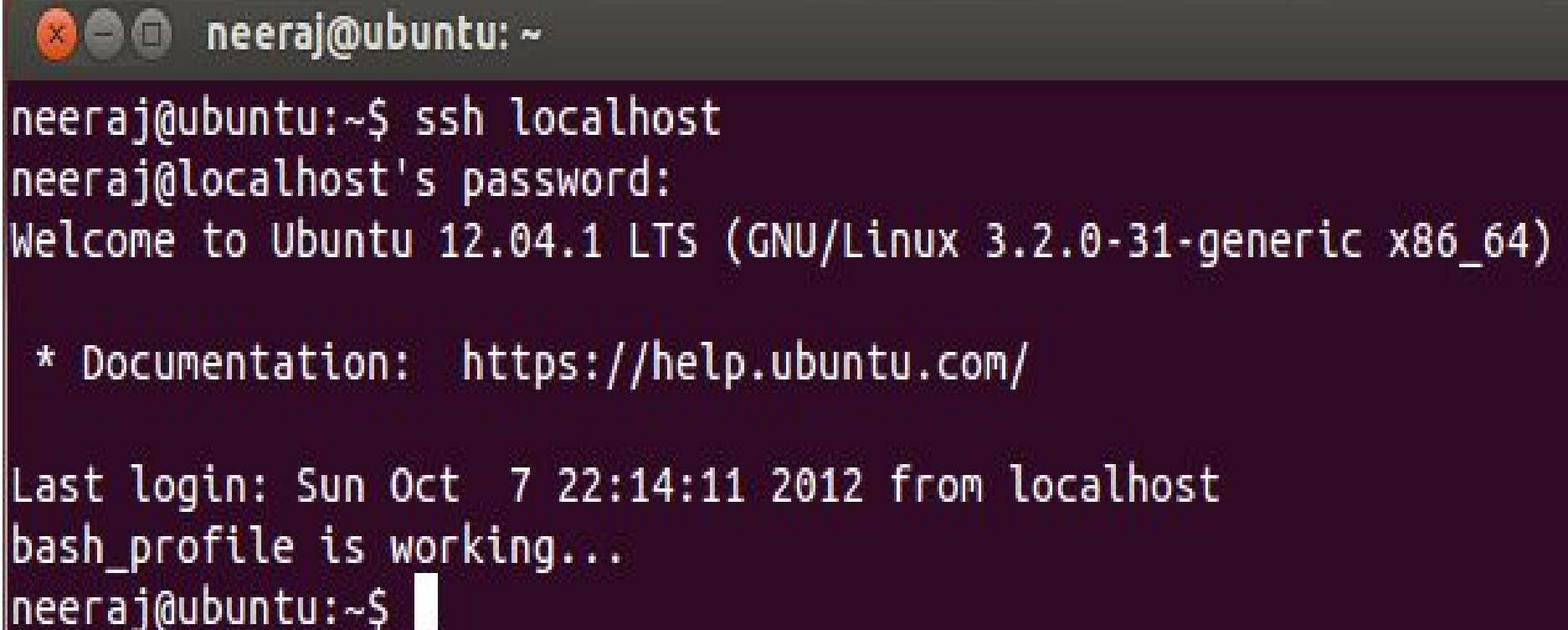
For example...   ssh 192.168.342.187
                        ssh 127.0.0.1
                        ssh localhost

Run the following command on your terminal

ssh localhost

Run this command from your home directory

ssh-keygen -t dsa -P '' -f ~/.ssh/id_dsa

Run this command from your home directory

cat ~/.ssh/id_dsa.pub >> ~/.ssh/authorized_keys
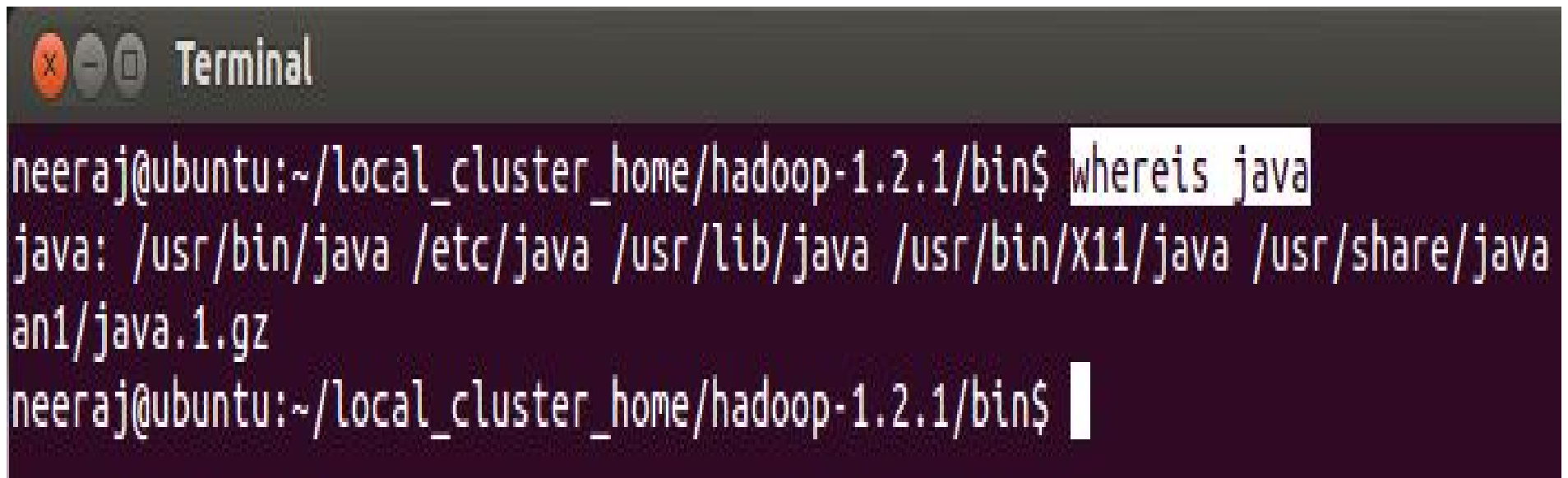
Modify hadoop-env.sh file

hadoop-env.sh file contains Hadoop environment settings.

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

export JAVA_HOME=/usr

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...**

```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 occuring 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

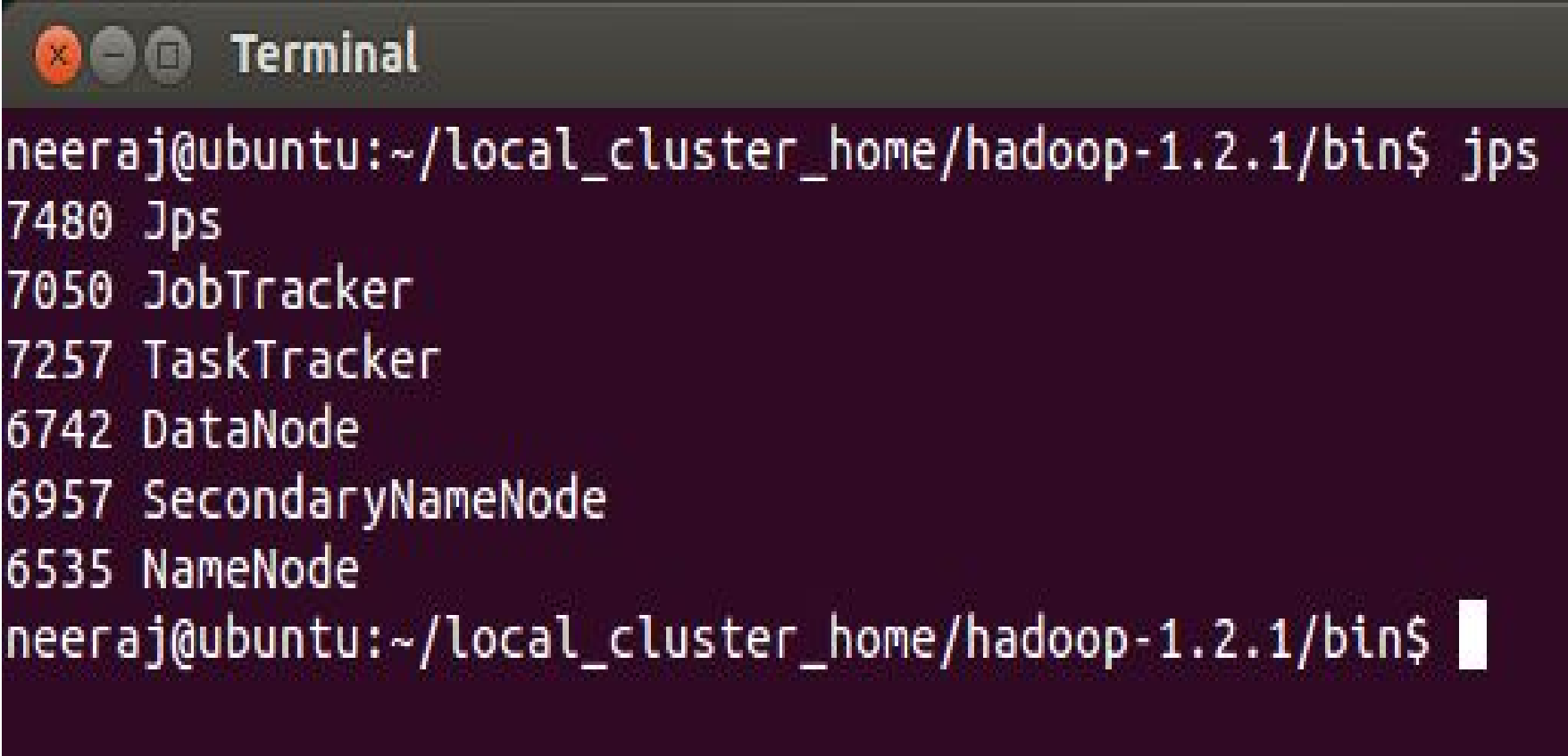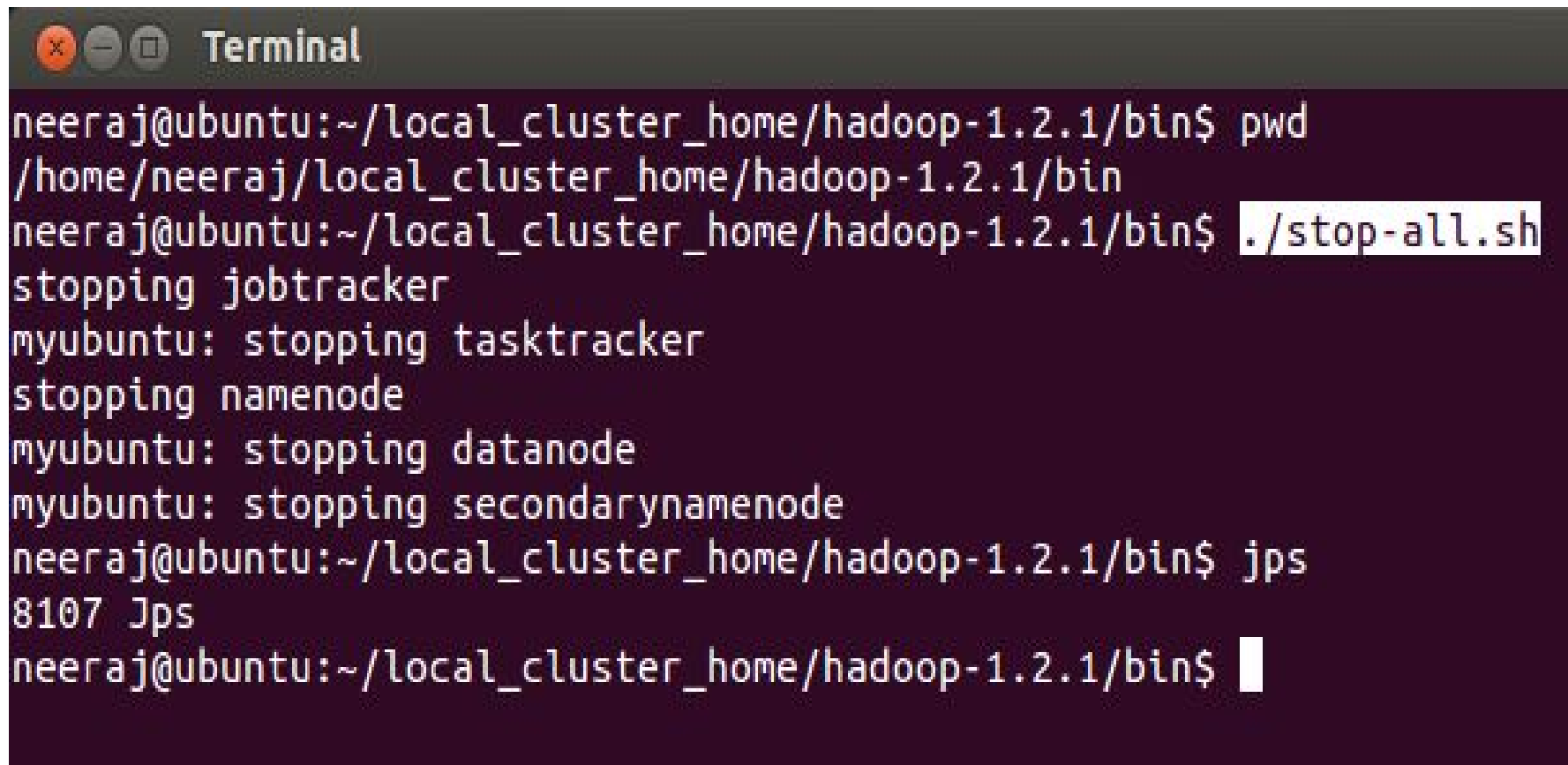Run the following command from bin directory of Hadoop

./start-all.sh

Verify Hadoop services

Run the following command from bin directory of Hadoop

jps

Stop the services of Hadoop

Run the following command from bin directory of Hadoop

./stop-all.sh

…Thanks…

For online Hadoop training, send mail to neeraj.ymca.2k6@gmail.com