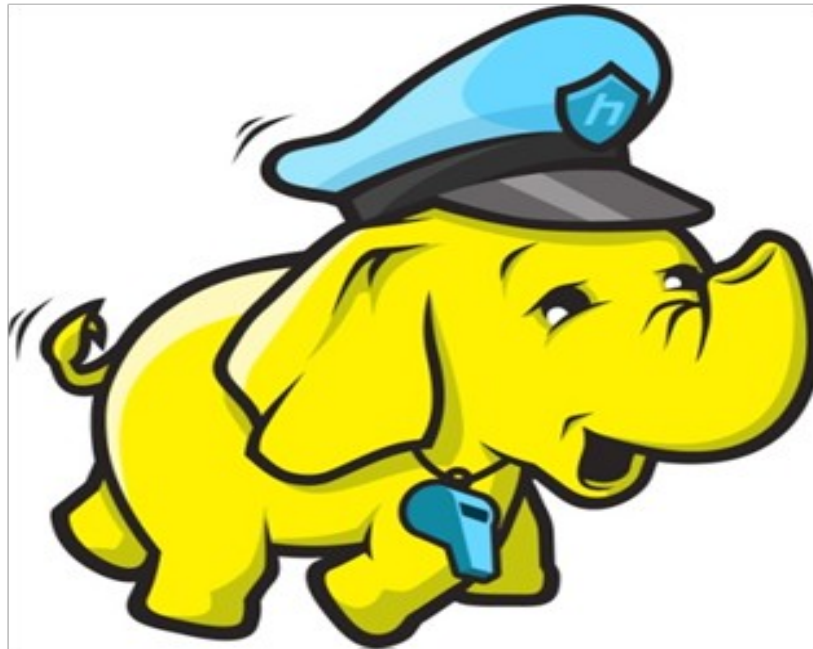


# Upgrade your Hadoop cluster



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

# Agenda

- What is Hadoop upgradation
- Why upgradation is required
- Save old Hadoop directory structure
- Stop old Hadoop
- Install new version of Hadoop
- Verify Namenode & data directory
- Start Namenode with upgrade option
- Start remaining Hadoop processes
- Check upgradation status
- Rollback the upgrade
- Finalize the upgrade

# What is Hadoop upgradation

Apache releases new versions of Hadoop.

We can use newer version of Hadoop without losing data.

Moving from old version of Hadoop to newer version of Hadoop is known as upgradation.

Moving from new version of Hadoop to older version of Hadoop is known as downgradation.

# Why upgradation is required

Newer version of Hadoop are better than older versions.

New versions has bugs fixed which were present in older version.

New version has new features.

New versions comes with updated functionality.

New version may be stable or unstable.

# Save old directory structure

Save the complete directory structure in a file for future use.

```
Terminal
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.0/bin$ pwd
/home/neeraj/local_cluster_home/test_upgrade/hadoop-1.2.0/bin
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.0/bin$ ./hadoop fs -lsr /
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:36 /dir1
-rw-r--r--  1 neeraj supergroup        123 2013-09-29 12:36 /dir1/sample_hadoop.txt
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:31 /home
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:31 /home/neeraj
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:31 /home/neeraj/local_cluster_home
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:31 /home/neeraj/local_cluster_home/test_upgrade
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:31 /home/neeraj/local_cluster_home/test_upgrade/HDFS
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:31 /home/neeraj/local_cluster_home/test_upgrade/HDFS/hdfs_te
mp
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:31 /home/neeraj/local_cluster_home/test_upgrade/HDFS/hdfs_te
mp/mapred
drwx----- - neeraj supergroup          0 2013-09-29 12:32 /home/neeraj/local_cluster_home/test_upgrade/HDFS/hdfs_te
mp/mapred/system
-rw-----  1 neeraj supergroup          4 2013-09-29 12:32 /home/neeraj/local_cluster_home/test_upgrade/HDFS/hdfs_te
mp/mapred/system/jobtracker.info
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.0/bin$ ./hadoop fs -lsr / > /home/neeraj/local_cluster_home
/test_upgrade/old-lsr.log
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.0/bin$
```

# Stop old Hadoop

```
Terminal
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.0/bin$ pwd
/home/neeraj/local_cluster_home/test_upgrade/hadoop-1.2.0/bin
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.0/bin$ jps
3338 TaskTracker
2581 NameNode
3034 SecondaryNameNode
3128 JobTracker
2815 DataNode
4040 Jps
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.0/bin$ ./stop-all.sh
stopping jobtracker
myubuntu: stopping tasktracker
stopping namenode
myubuntu: stopping datanode
myubuntu: stopping secondarynamenode
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.0/bin$ jps
4634 Jps
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.0/bin$
```

# Install new version of Hadoop

Download new version of Hadoop from Apache website.

Make the necessary changes in configuration files.

Repeat these steps on all machines on cluster.

Follow Hadoop Installation PDF for more details.

New version of Hadoop uses the old fsimage to create new metadata.

# Verify Namenode & data directory

Make sure that the Namenode directory & Data directory points to old version of Hadoop

```
<property>
```

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

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

```
  <description>This directory will be used to store the Namenode metadata </description>
```

```
</property>
```

```
<property>
```

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

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

```
  <description>This directory will be used to store the data </description>
```

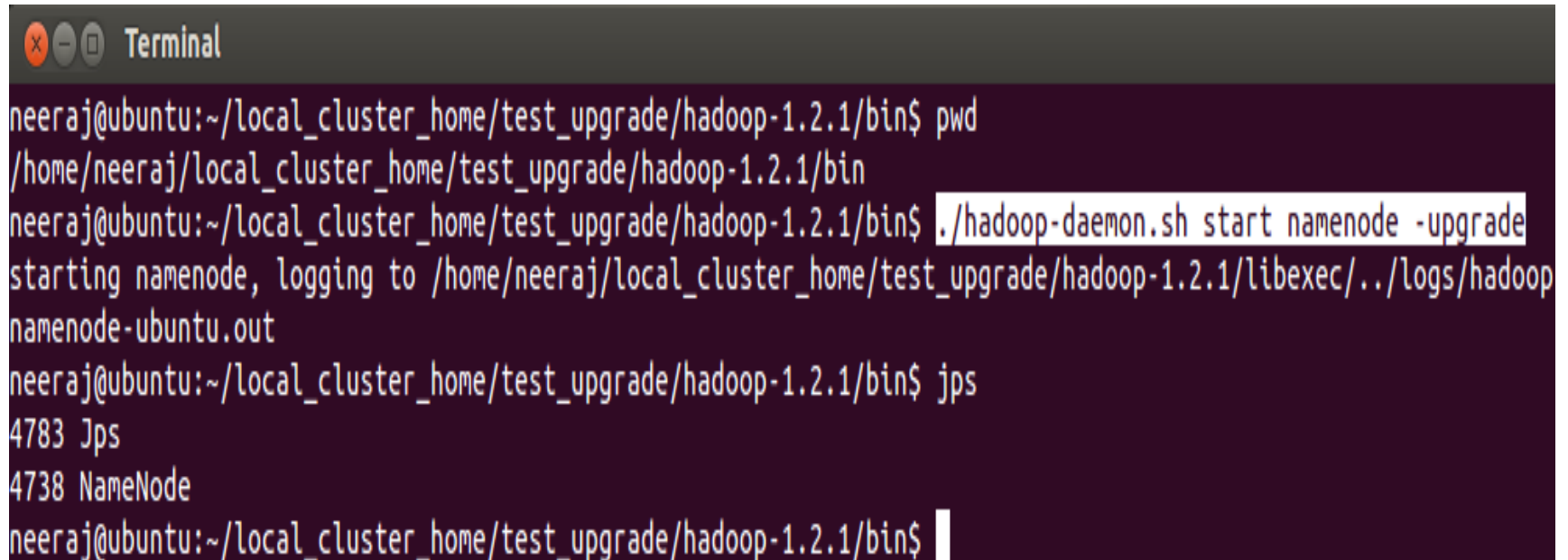
```
</property>
```



# Start Namenode with upgrade option

Fsimage contains all the metadata.

New Hadoop takes old fsimage & create a new fsimage for new version



```
Terminal
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$ pwd
/home/neeraj/local_cluster_home/test_upgrade/hadoop-1.2.1/bin
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$ ./hadoop-daemon.sh start namenode -upgrade
starting namenode, logging to /home/neeraj/local_cluster_home/test_upgrade/hadoop-1.2.1/libexec/./logs/hadoop
namenode-ubuntu.out
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$ jps
4783 Jps
4738 NameNode
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$
```

# Start new Hadoop

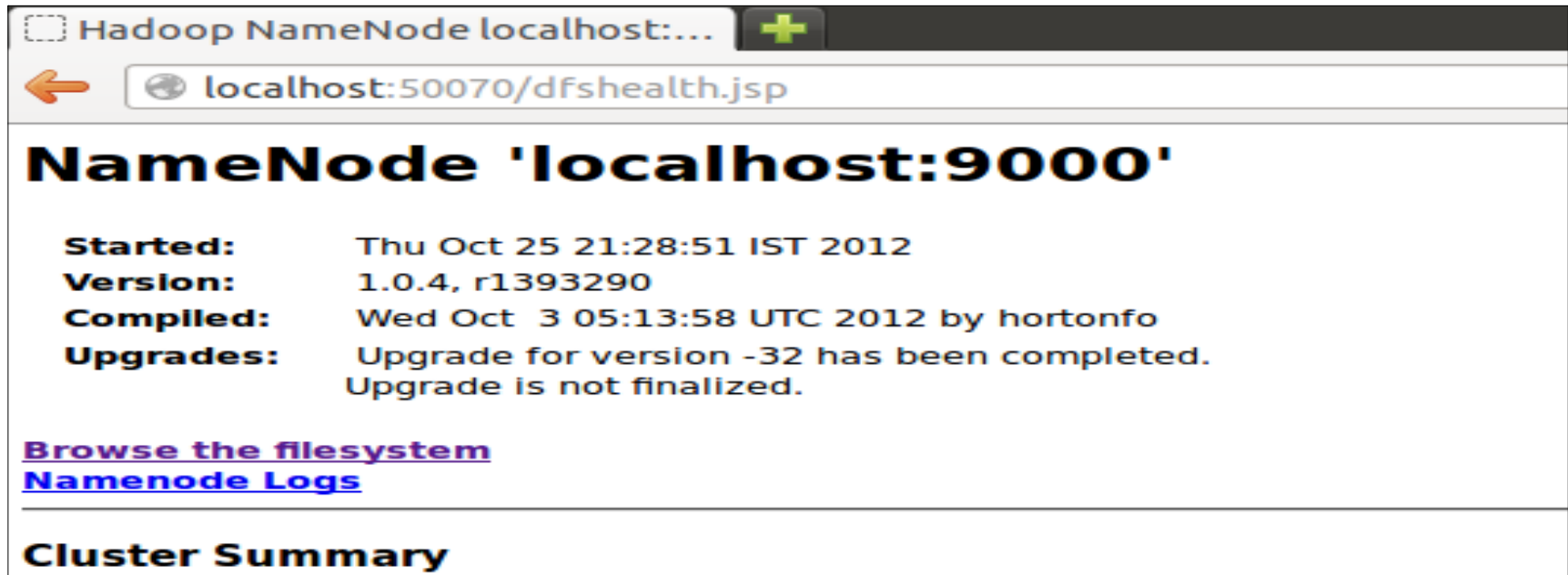
Start new Hadoop using the below command.  
`./start-all.sh`

```
Terminal
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$ pwd
/home/neeraj/local_cluster_home/test_upgrade/hadoop-1.2.1/bin
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$ jps
4874 Jps
4738 NameNode
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$ ./start-all.sh
namenode running as process 4738. Stop it first.
myubuntu: starting datanode, logging to /home/neeraj/local_cluster_home/test_upgrade/hadoop-neeraj-datanode-ubuntu.out
myubuntu: starting secondarynamenode, logging to /home/neeraj/local_cluster_home/test_upgrade/hadoop-neeraj-secondarynamenode-ubuntu.out
starting jobtracker, logging to /home/neeraj/local_cluster_home/test_upgrade/hadoop-neeraj-jobtracker-ubuntu.out
myubuntu: starting tasktracker, logging to /home/neeraj/local_cluster_home/test_upgrade/hadoop-neeraj-tasktracker-ubuntu.out
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$
```

# Verify HDFS data in new Hadoop

```
Terminal
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$ pwd
/home/neeraj/local_cluster_home/test_upgrade/hadoop-1.2.1/bin
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$ ./hadoop fs -lsr /
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:36 /dir1
-rw-r--r--  1 neeraj supergroup        123 2013-09-29 12:36 /dir1/sample_hadoop.txt
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:31 /home
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:31 /home/neeraj
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:31 /home/neeraj/local_cluster_home
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:31 /home/neeraj/local_cluster_home
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:31 /home/neeraj/local_cluster_home
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:31 /home/neeraj/local_cluster_home
mp
drwxr-xr-x  - neeraj supergroup          0 2013-09-29 12:59 /home/neeraj/local_cluster_home
mp/mapred
drwx-----  - neeraj supergroup          0 2013-09-29 12:59 /home/neeraj/local_cluster_home
mp/mapred/system
-rw-----  1 neeraj supergroup          4 2013-09-29 12:59 /home/neeraj/local_cluster_home
mp/mapred/system/jobtracker.info
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$
```

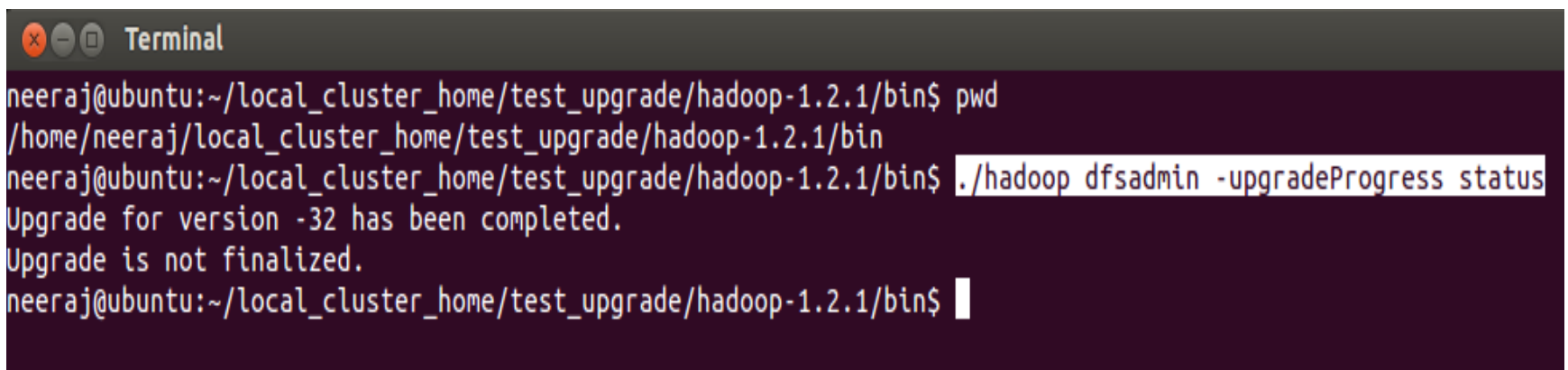
# Check upgradation status



A screenshot of a web browser window showing the Hadoop NameNode local status page. The browser tab is titled 'Hadoop NameNode localhost:...' and the address bar shows 'localhost:50070/dfshealth.jsp'. The main heading is 'NameNode 'localhost:9000''. Below this, there is a table of status information:

<b>Started:</b>	Thu Oct 25 21:28:51 IST 2012
<b>Version:</b>	1.0.4, r1393290
<b>Compiled:</b>	Wed Oct 3 05:13:58 UTC 2012 by hortonfo
<b>Upgrades:</b>	Upgrade for version -32 has been completed. Upgrade is not finalized.

Below the table, there are two links: [Browse the filesystem](#) and [Namenode Logs](#). At the bottom of the page, there is a section titled 'Cluster Summary'.



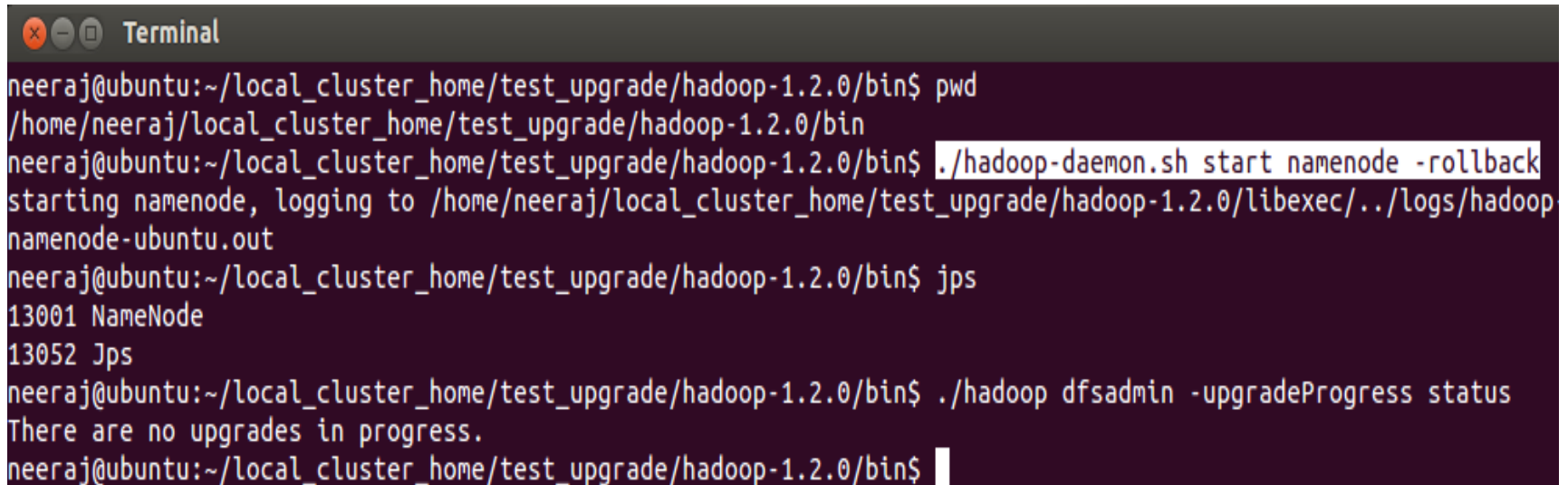
A screenshot of a terminal window titled 'Terminal'. The user 'neeraj' is at the prompt 'neeraj@ubuntu:~/local\_cluster\_home/test\_upgrade/hadoop-1.2.1/bin\$'. They run the command 'pwd' and get the output '/home/neeraj/local\_cluster\_home/test\_upgrade/hadoop-1.2.1/bin'. Then they run the command './hadoop dfsadmin -upgradeProgress status' and get the output 'Upgrade for version -32 has been completed. Upgrade is not finalized.'.

```
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$ pwd
/home/neeraj/local_cluster_home/test_upgrade/hadoop-1.2.1/bin
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$ ./hadoop dfsadmin -upgradeProgress status
Upgrade for version -32 has been completed.
Upgrade is not finalized.
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$
```

# Rollback the upgrade

Rollback the upgrade by running below command from old Hadoop bin directory

`./hadoop-daemon.sh start namenode -rollback`

A terminal window titled "Terminal" with a dark background. It shows a user named "neeraj" at an "ubuntu" machine. The user is in the directory "/local\_cluster\_home/test\_upgrade/hadoop-1.2.0/bin". They run "pwd" and get "/home/neeraj/local\_cluster\_home/test\_upgrade/hadoop-1.2.0/bin". Then they run the command ". /hadoop-daemon.sh start namenode -rollback", which outputs "starting namenode, logging to /home/neeraj/local\_cluster\_home/test\_upgrade/hadoop-1.2.0/libexec/../logs/hadoop-namenode-ubuntu.out". Next, they run "jps" and see "13001 NameNode" and "13052 Jps". Finally, they run ". /hadoop dfsadmin -upgradeProgress status" and get the output "There are no upgrades in progress.".

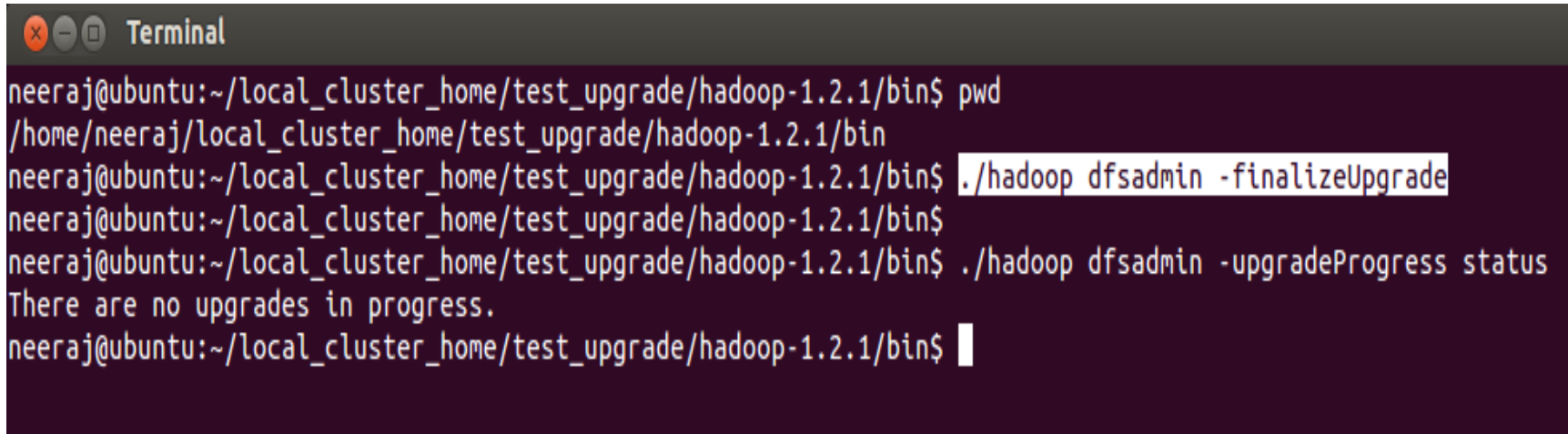
```
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.0/bin$ pwd
/home/neeraj/local_cluster_home/test_upgrade/hadoop-1.2.0/bin
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.0/bin$ ./hadoop-daemon.sh start namenode -rollback
starting namenode, logging to /home/neeraj/local_cluster_home/test_upgrade/hadoop-1.2.0/libexec/../logs/hadoop-namenode-ubuntu.out
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.0/bin$ jps
13001 NameNode
13052 Jps
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.0/bin$ ./hadoop dfsadmin -upgradeProgress status
There are no upgrades in progress.
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.0/bin$
```

You have successfully rolled back your Hadoop

# Finalize the upgrade

Finalize the upgrade by running below command from new Hadoop bin directory

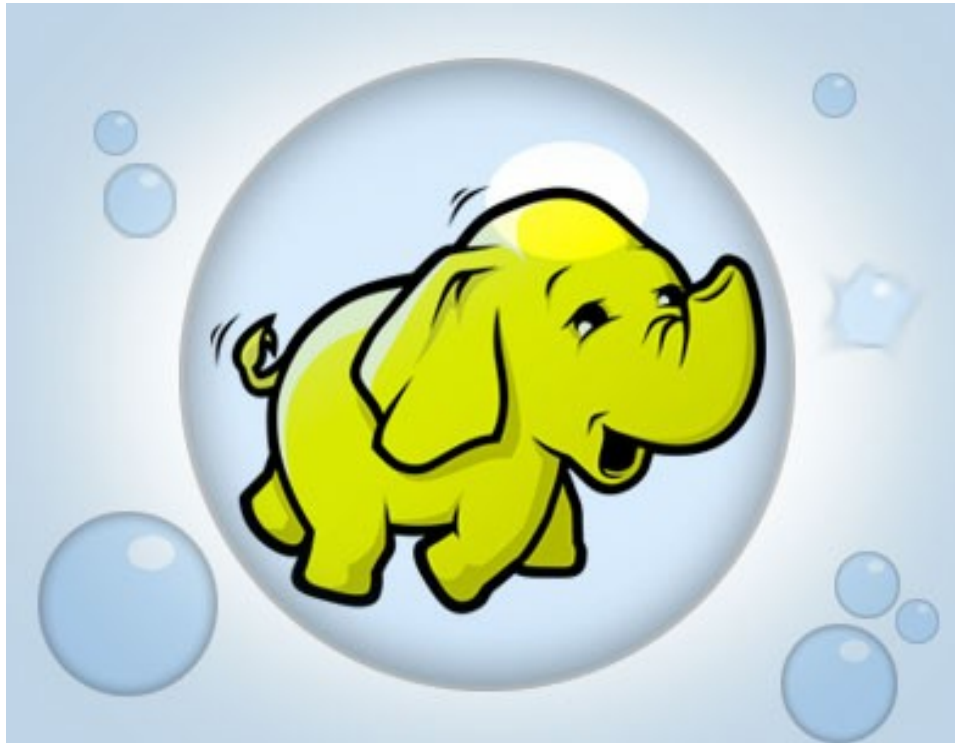
`./hadoop dfsadmin -finalizeUpgrade`

A terminal window titled "Terminal" with a dark background. The prompt is "neeraj@ubuntu:~/local\_cluster\_home/test\_upgrade/hadoop-1.2.1/bin\$". The user enters "pwd" and the output is "/home/neeraj/local\_cluster\_home/test\_upgrade/hadoop-1.2.1/bin". Then the user enters the command `./hadoop dfsadmin -finalizeUpgrade`, which is highlighted with a light blue background. The prompt returns. Finally, the user enters `./hadoop dfsadmin -upgradeProgress status` and the output is "There are no upgrades in progress." followed by a new line.

```
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$ pwd
/home/neeraj/local_cluster_home/test_upgrade/hadoop-1.2.1/bin
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$ ./hadoop dfsadmin -finalizeUpgrade
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$ ./hadoop dfsadmin -upgradeProgress status
There are no upgrades in progress.
neeraj@ubuntu:~/local_cluster_home/test_upgrade/hadoop-1.2.1/bin$
```

You have successfully upgraded your Hadoop

# ...Thanks...



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