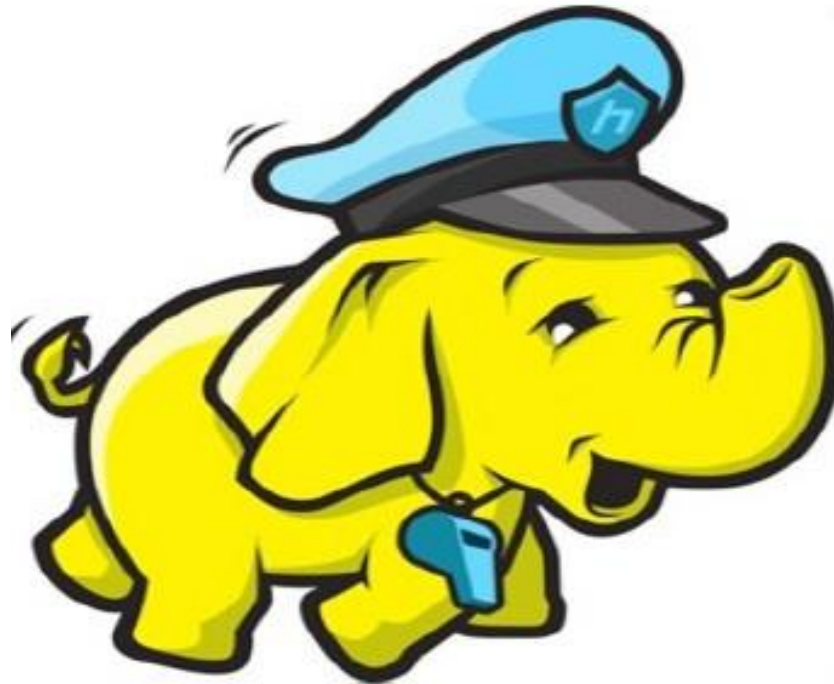


YARN (Hadoop 2.x.x)



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

Agenda

Hadoop 1.0 Vs Hadoop 2.0

Install YARN (Hadoop 2.x.x) on your system

Format your Namenode

Start HDFS

Start YARN

Start Job history server

Copy data from local to HDFS

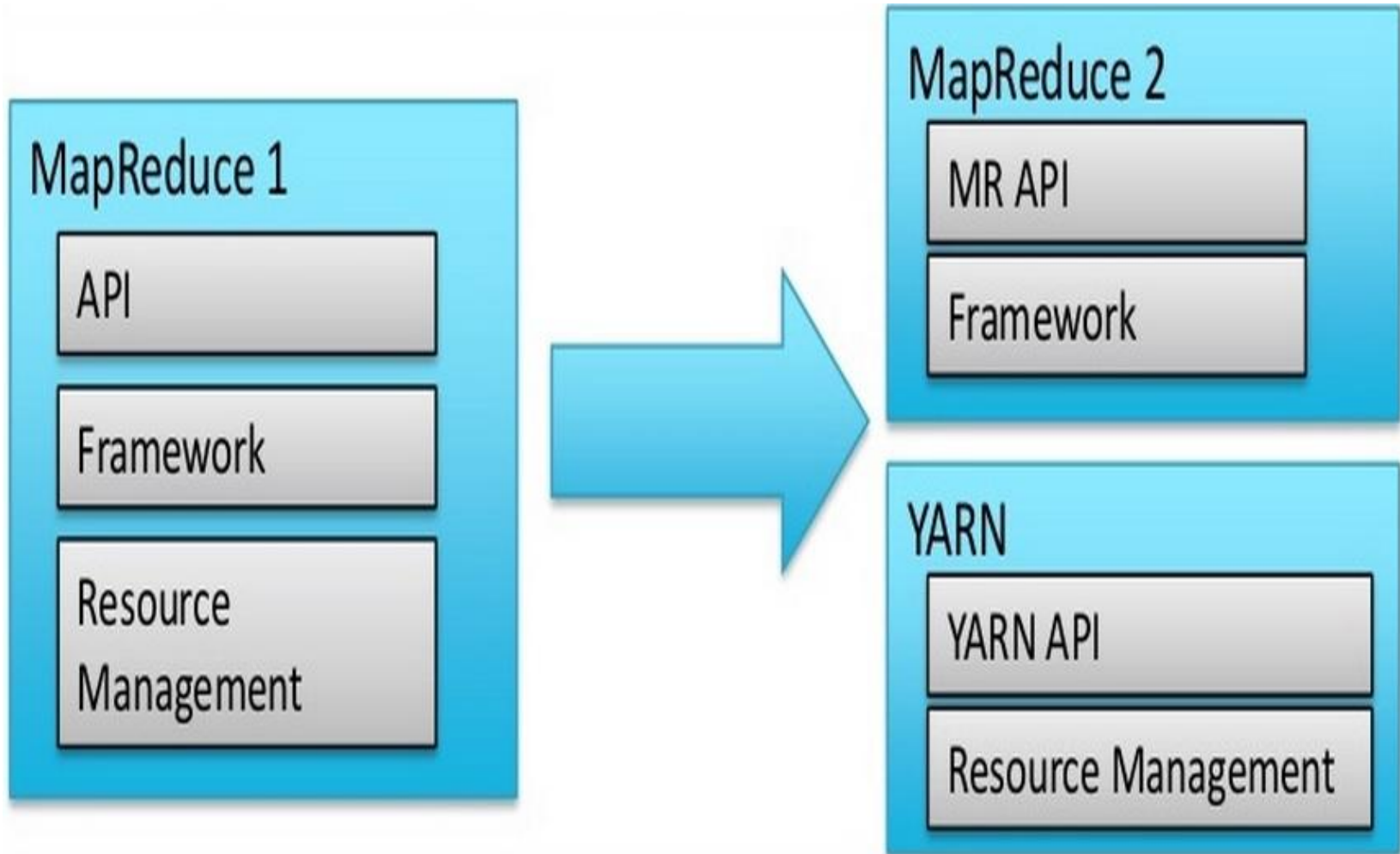
Execute Hadoop job on YARN

HDFS UI

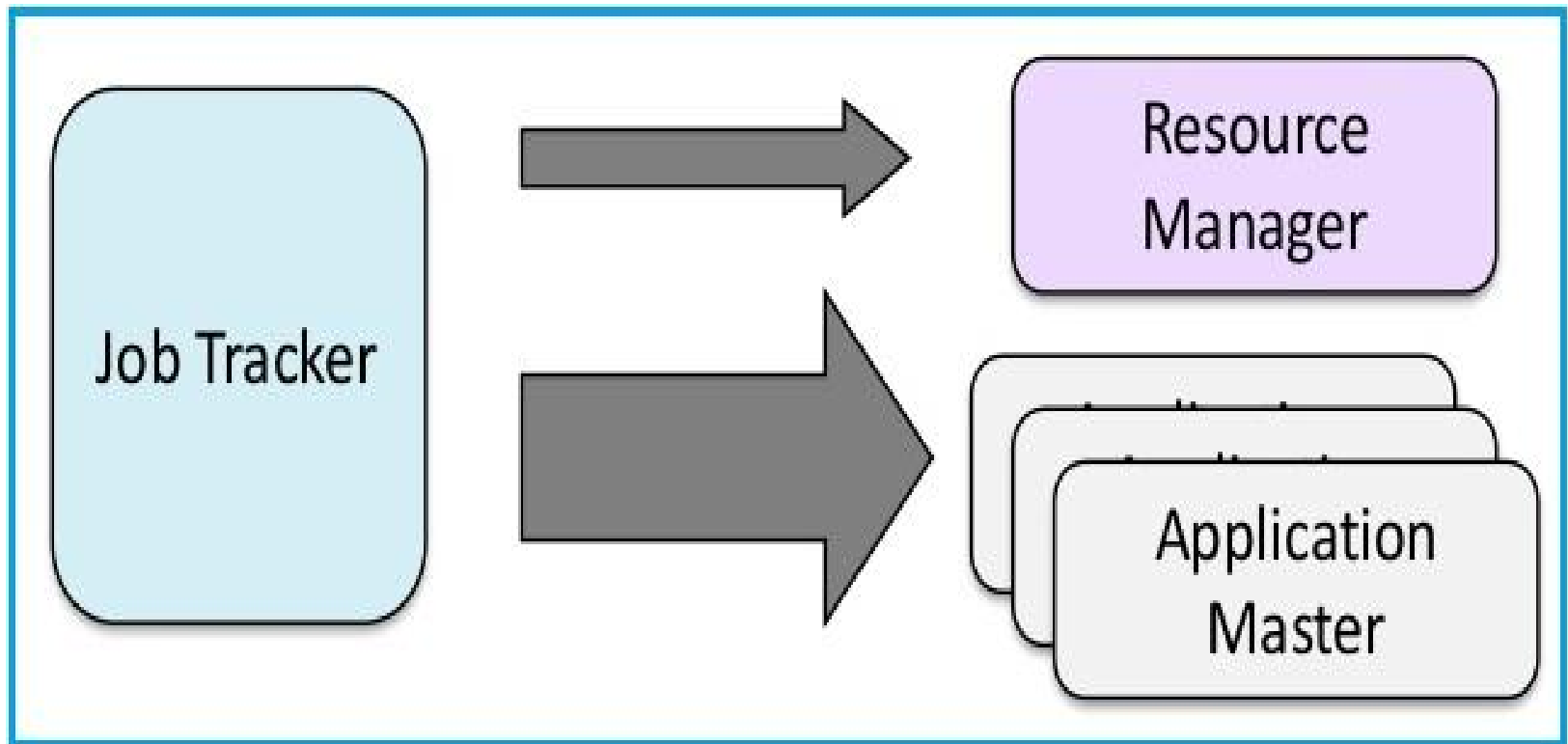
MR UI

Job history server UI

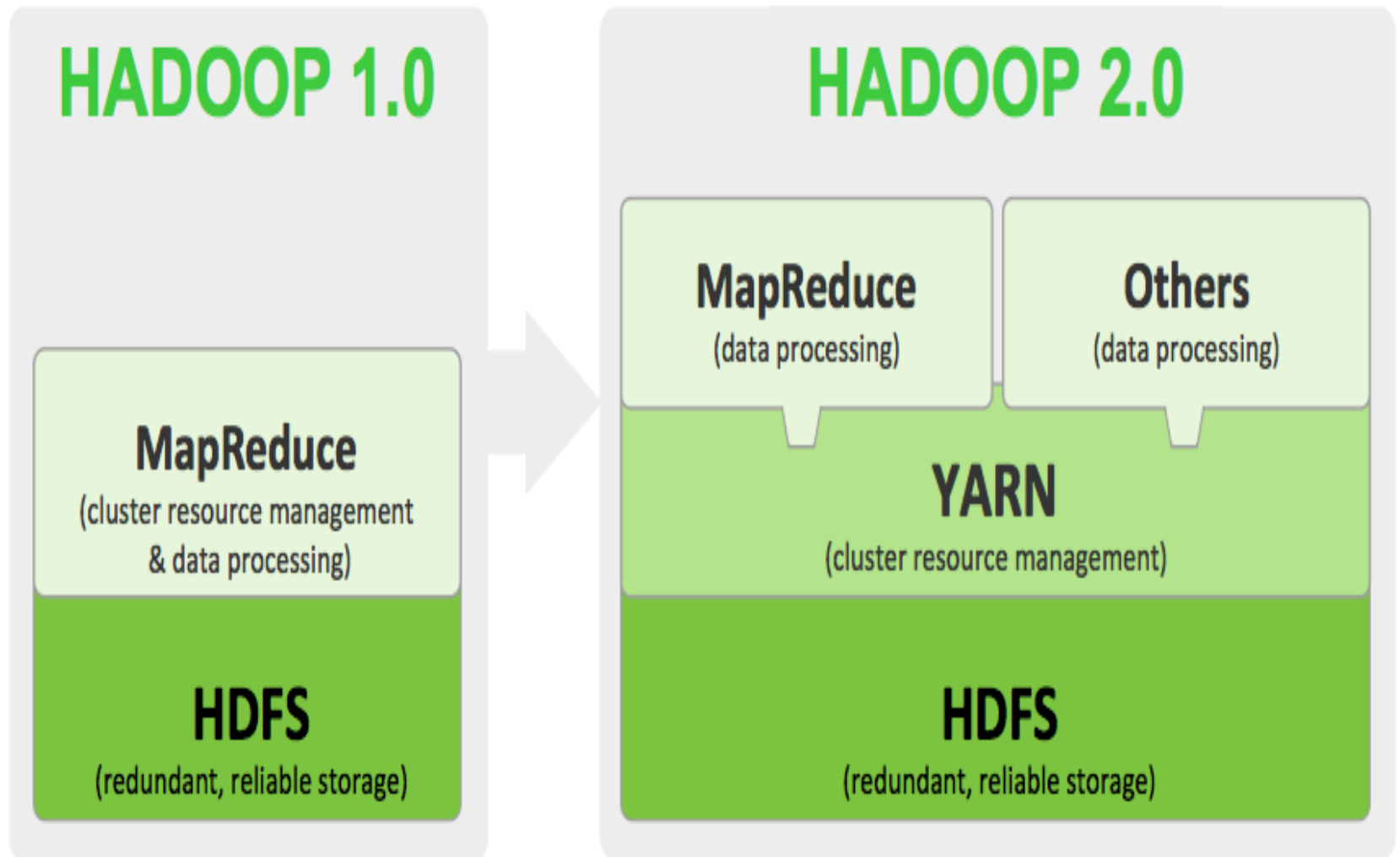
MR1 Vs MR2



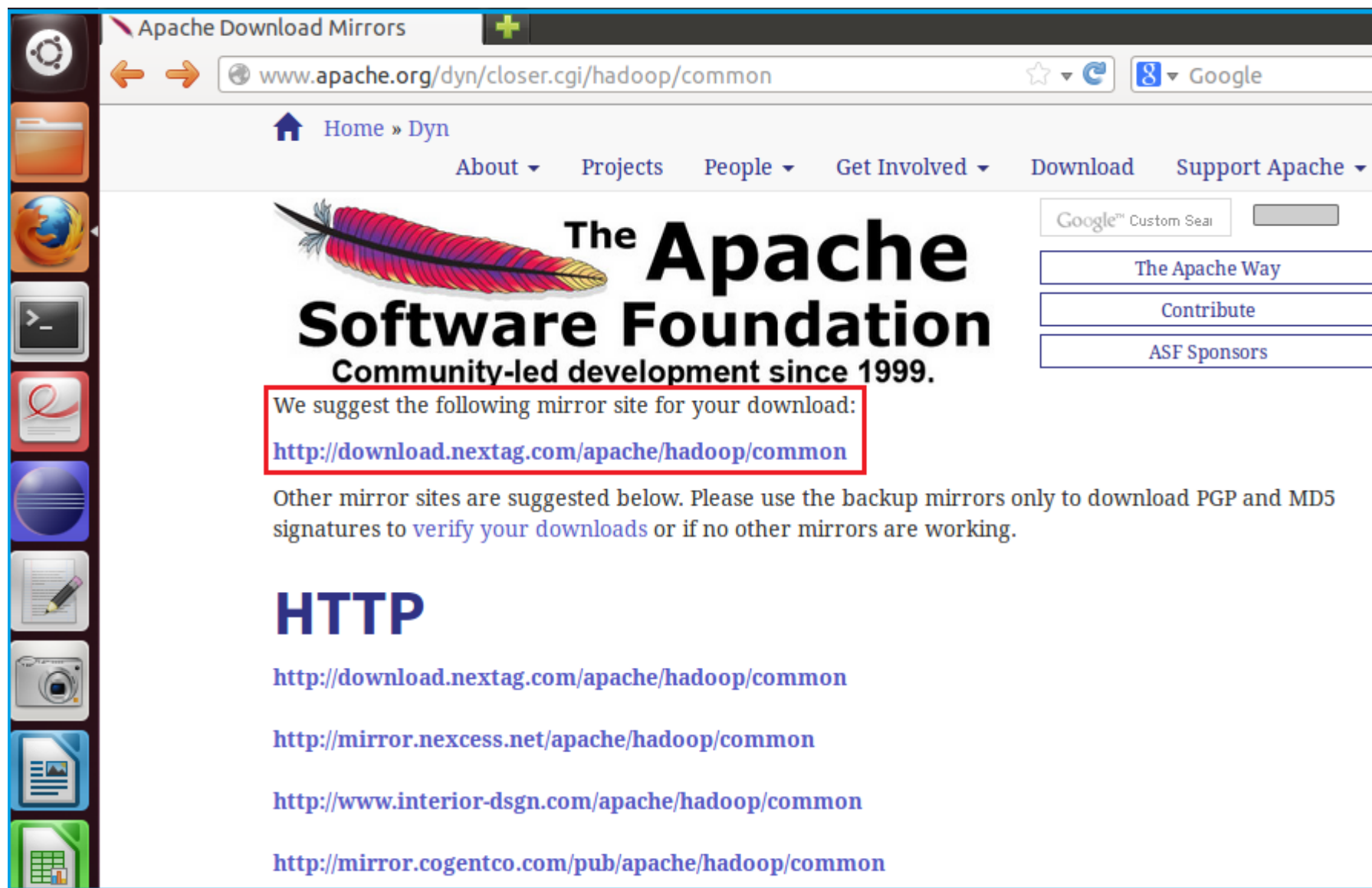
Job Tracker Vs Resource Manager/Application Master



Hadoop 1.0 Vs Hadoop 2.0



Download Hadoop from Apache website




The screenshot shows a web browser window with the address bar displaying `www.apache.org/dyn/closer.cgi/hadoop/common`. The page title is "Apache Download Mirrors". The main content area features the Apache Software Foundation logo (a colorful feather) and the text "The Apache Software Foundation Community-led development since 1999." Below this, a red-bordered box highlights the text "We suggest the following mirror site for your download:" followed by the URL <http://download.nextag.com/apache/hadoop/common>. To the right of the main content, there is a Google Custom Search bar and three buttons: "The Apache Way", "Contribute", and "ASF Sponsors". Below the highlighted URL, a paragraph states: "Other mirror sites are suggested below. Please use the backup mirrors only to download PGP and MD5 signatures to [verify your downloads](#) or if no other mirrors are working." This is followed by the heading "HTTP" and a list of four mirror URLs: <http://download.nextag.com/apache/hadoop/common>, <http://mirror.nexcess.net/apache/hadoop/common>, <http://www.interior-dsgn.com/apache/hadoop/common>, and <http://mirror.cogentco.com/pub/apache/hadoop/common>. The browser's left sidebar shows various application icons, and the top navigation bar includes links for "Home", "Dyn", "About", "Projects", "People", "Get Involved", "Download", and "Support Apache".

Apache Download Mirrors

Home » Dyn

About Projects People Get Involved Download Support Apache

 **The Apache Software Foundation**
Community-led development since 1999.

We suggest the following mirror site for your download:

<http://download.nextag.com/apache/hadoop/common>

Other mirror sites are suggested below. Please use the backup mirrors only to download PGP and MD5 signatures to [verify your downloads](#) or if no other mirrors are working.

HTTP

<http://download.nextag.com/apache/hadoop/common>

<http://mirror.nexcess.net/apache/hadoop/common>

<http://www.interior-dsgn.com/apache/hadoop/common>

<http://mirror.cogentco.com/pub/apache/hadoop/common>

Google™ Custom Search

The Apache Way

Contribute

ASF Sponsors





Click on stable directory



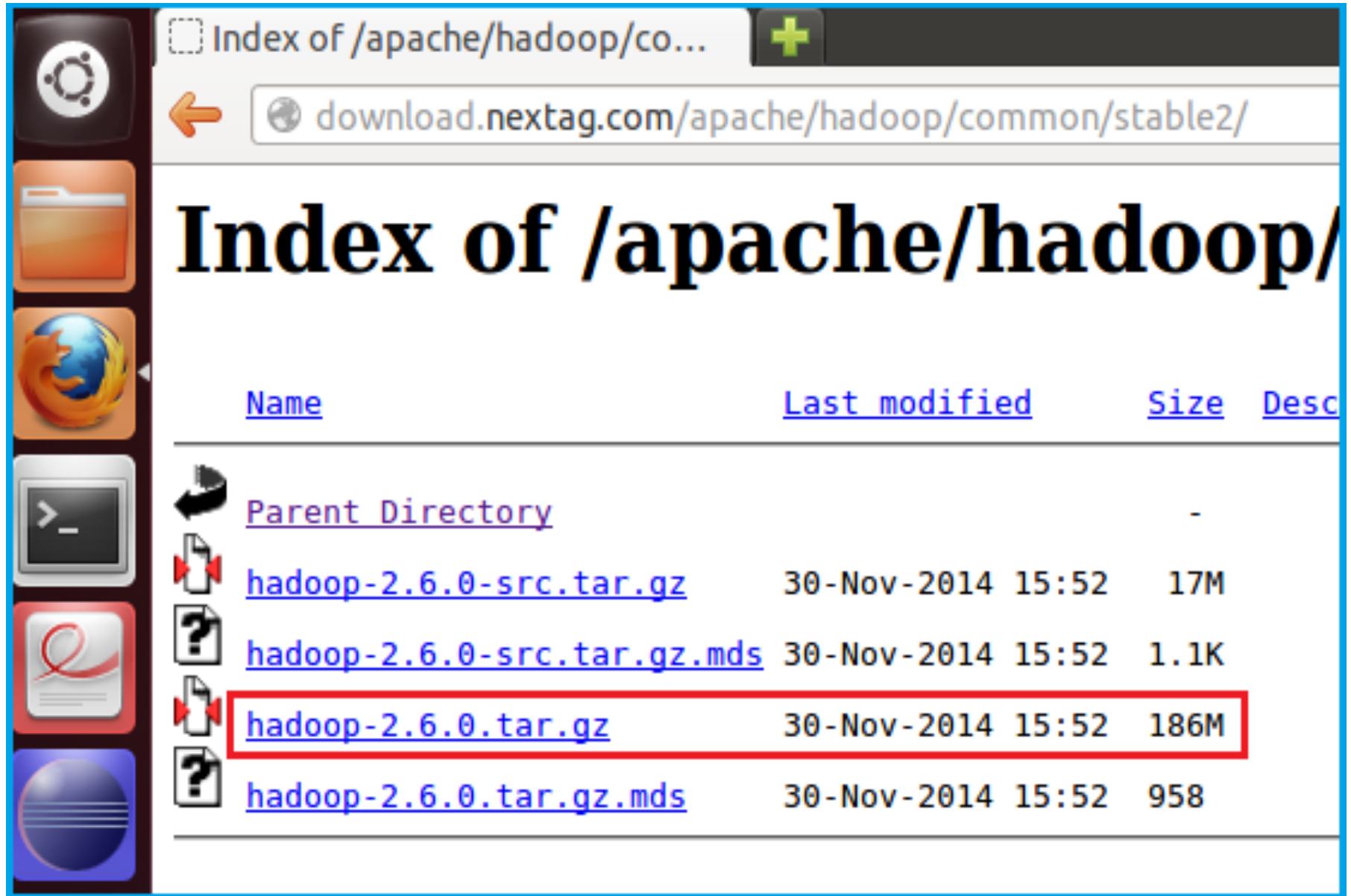
Index of /apache/hadoop/co...

download.nexttag.com/apache/hadoop/common/






Index of /apache/hadoop/common

Name	Last modified	Size	Description
 Parent Directory		-	
 current/	21-Apr-2015 09:47	-	
 hadoop-1.2.1/	29-Dec-2014 22:44	-	
 hadoop-2.5.2/	29-Dec-2014 22:45	-	
 hadoop-2.6.0/	29-Dec-2014 22:45	-	
 hadoop-2.7.0/	21-Apr-2015 09:47	-	
 readme.txt	20-Apr-2015 18:32	184	
 stable/	29-Dec-2014 22:45	-	
 stable1/	29-Dec-2014 22:44	-	
 stable2/	29-Dec-2014 22:45	-	

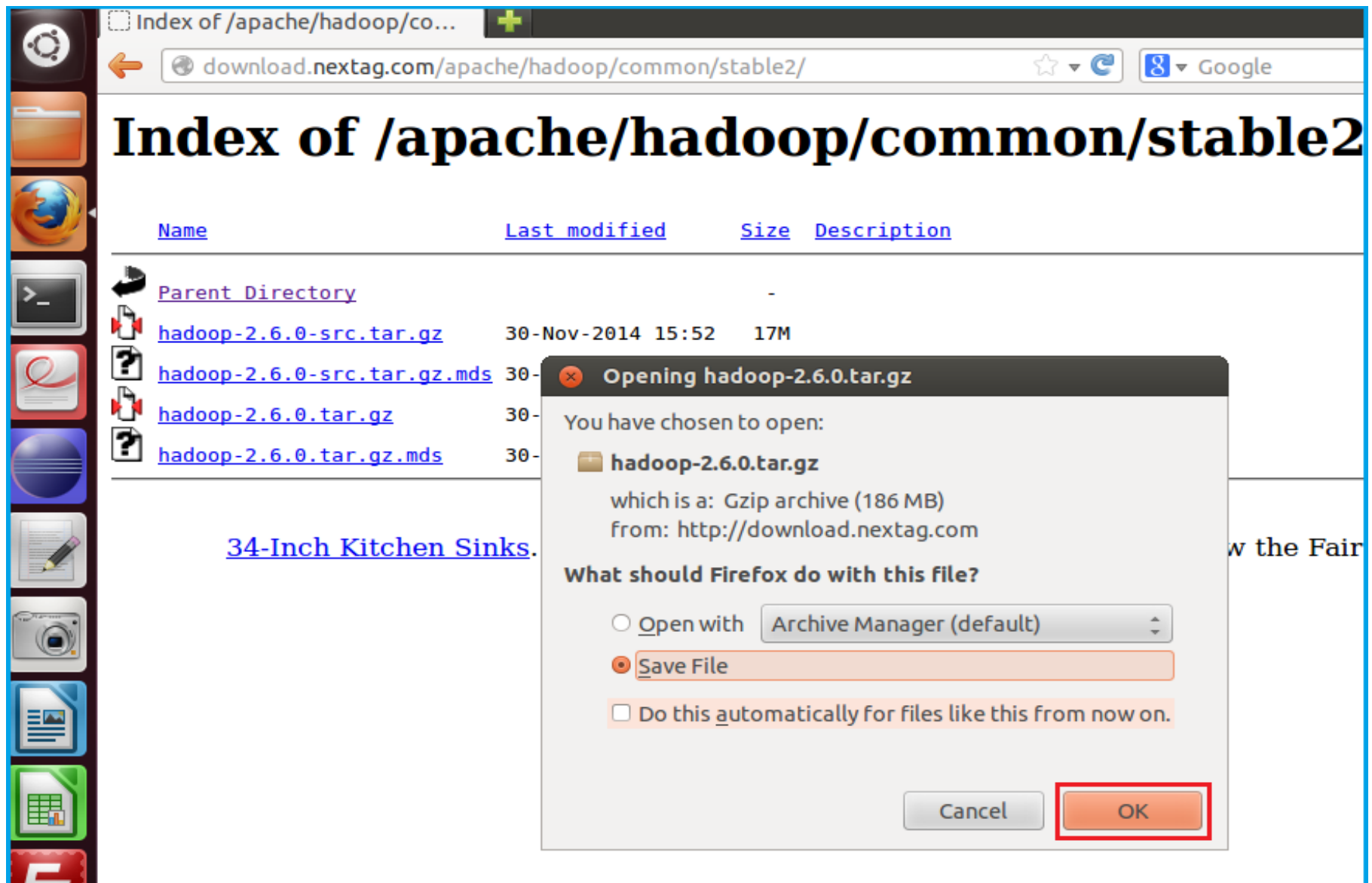
Click on tar.gz file



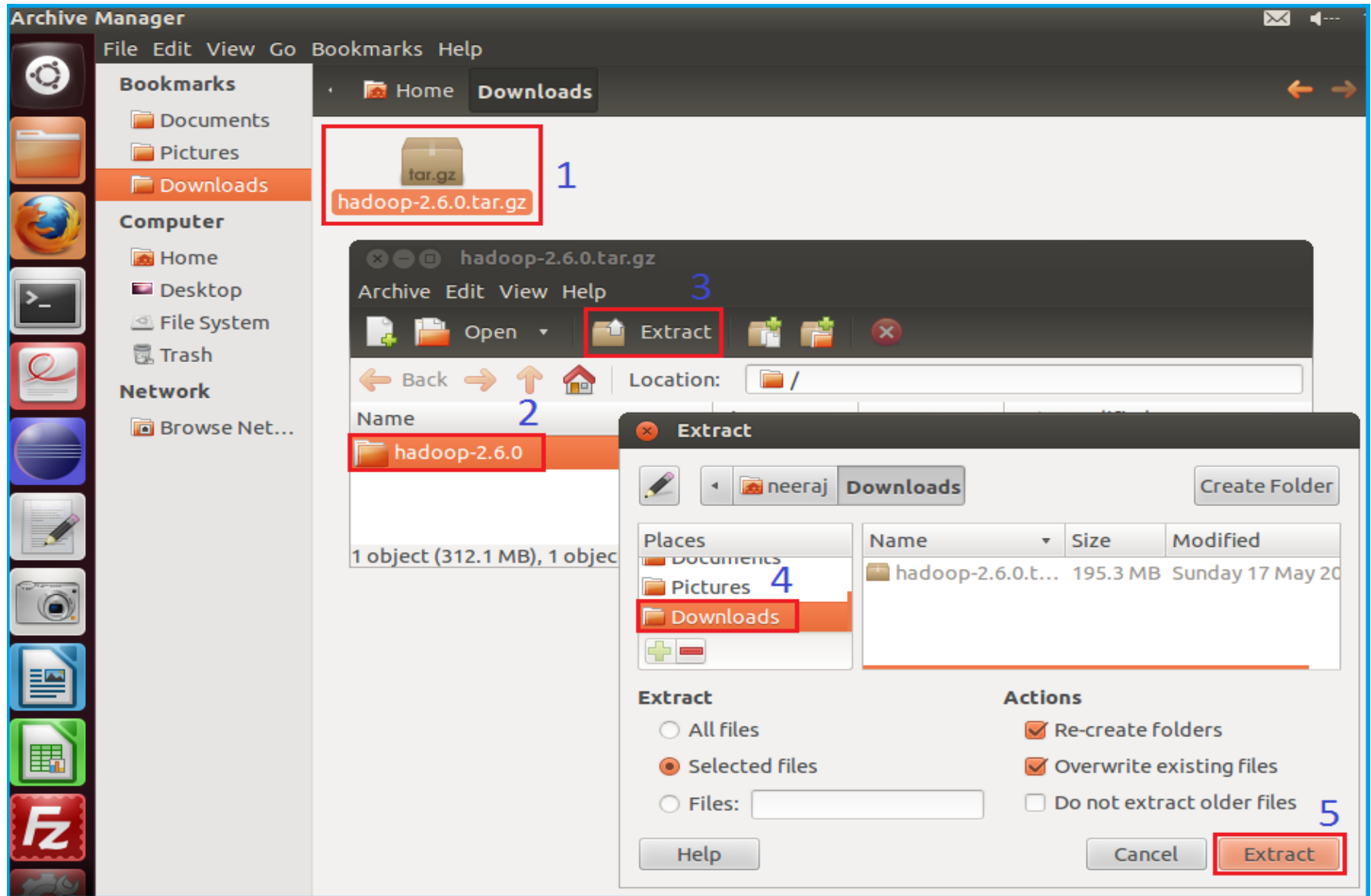
The screenshot shows a web browser window with the address bar displaying `download.nextag.com/apache/hadoop/common/stable2/`. The page title is "Index of /apache/hadoop/". The main content is a table listing files in the directory. The file `hadoop-2.6.0.tar.gz` is highlighted with a red box.

Name	Last modified	Size	Desc
 Parent Directory		-	
 hadoop-2.6.0-src.tar.gz	30-Nov-2014 15:52	17M	
 hadoop-2.6.0-src.tar.gz.mds	30-Nov-2014 15:52	1.1K	
 hadoop-2.6.0.tar.gz	30-Nov-2014 15:52	186M	
 hadoop-2.6.0.tar.gz.mds	30-Nov-2014 15:52	958	

Save tar.gz file on your system



Extract the content of tar.gz file



Directory structure of Hadoop

HADOOP_HOME

sbin [contains scripts to start/stop Hadoop]

bin [contains HDFS & Map-Reduce commands]

etc/hadoop [contains configuration files of Hadoop]

lib [contains libraries/JARs to run Hadoop]

logs [contains log files]

Modify etc/hadoop/hadoop-env.sh

export **JAVA_HOME**=/usr

export **HADOOP_HOME**=/home/neeraj/Desktop/hadoop-2.6.0

export **HADOOP_MAPRED_HOME**=/home/neeraj/Desktop/hadoop-2.6.0

export **HADOOP_COMMON_HOME**=/home/neeraj/Desktop/hadoop-2.6.0

export **HADOOP_HDFS_HOME**=/home/neeraj/Desktop/hadoop-2.6.0

export **HADOOP_YARN_HOME**=/home/neeraj/Desktop/hadoop-2.6.0

export **HADOOP_CONF_DIR**=/home/neeraj/Desktop/hadoop-2.6.0
/etc/hadoop

Modify etc/hadoop/core-site.xml

```
<configuration>
```

```
  <property>
```

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

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

```
  </property>
```

```
  <property>
```

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

```
    <value>/home/neeraj/Desktop/HDFS2/hdfs_temp</value>
```

```
  </property>
```

```
</configuration>
```

Modify etc/hadoop/hdfs-site.xml

```
<configuration>
```

```
  <property>
```

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

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

```
  </property>
```

```
  <property>
```

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

```
    <value>/home/neeraj/Desktop/HDFS2/hdfs_metadata</value>
```

```
  </property>
```

```
  <property>
```

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

```
    <value>/home/neeraj/Desktop/HDFS2/hdfs_data</value>
```

```
  </property>
```

```
</configuration>
```

Modify etc/hadoop/mapred-site.xml

```
<configuration>
```

```
  <property>
```

```
    <name>mapreduce.framework.name</name>
```

```
    <value>yarn</value>
```

```
  </property>
```

```
</configuration>
```

Modify etc/hadoop/yarn-site.xml

```
<configuration>
```

```
  <property>
```

```
    <name>yarn.nodemanager.aux-services</name>
```

```
    <value>mapreduce_shuffle</value>
```

```
  </property>
```

```
  <property>
```

```
    <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
```

```
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
```

```
  </property>
```

```
</configuration>
```


Format your Namenode

`./hdfs namenode -format`

```
Terminal
File Edit View Search Terminal Help
/home/neeraj/local_cluster_home/hadoop-2.6.0/bin
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/bin$ ./hdfs namenode -format
15/06/05 13:26:25 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = myubuntu/127.0.0.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 2.6.0
15/06/05 13:29:14 INFO namenode.FSImage: Allocated new BlockPoolId: BP-54795721-127.0.0.1-1433491154041
15/06/05 13:29:15 INFO common.Storage: Storage directory /home/neeraj/local_cluster_home/HDFS2/hdfs_metadata has been successfully formatted.
15/06/05 13:29:15 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
15/06/05 13:29:15 INFO util.ExitUtil: Exiting with status 0
15/06/05 13:29:15 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at myubuntu/127.0.0.1
*****/
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/bin$
```

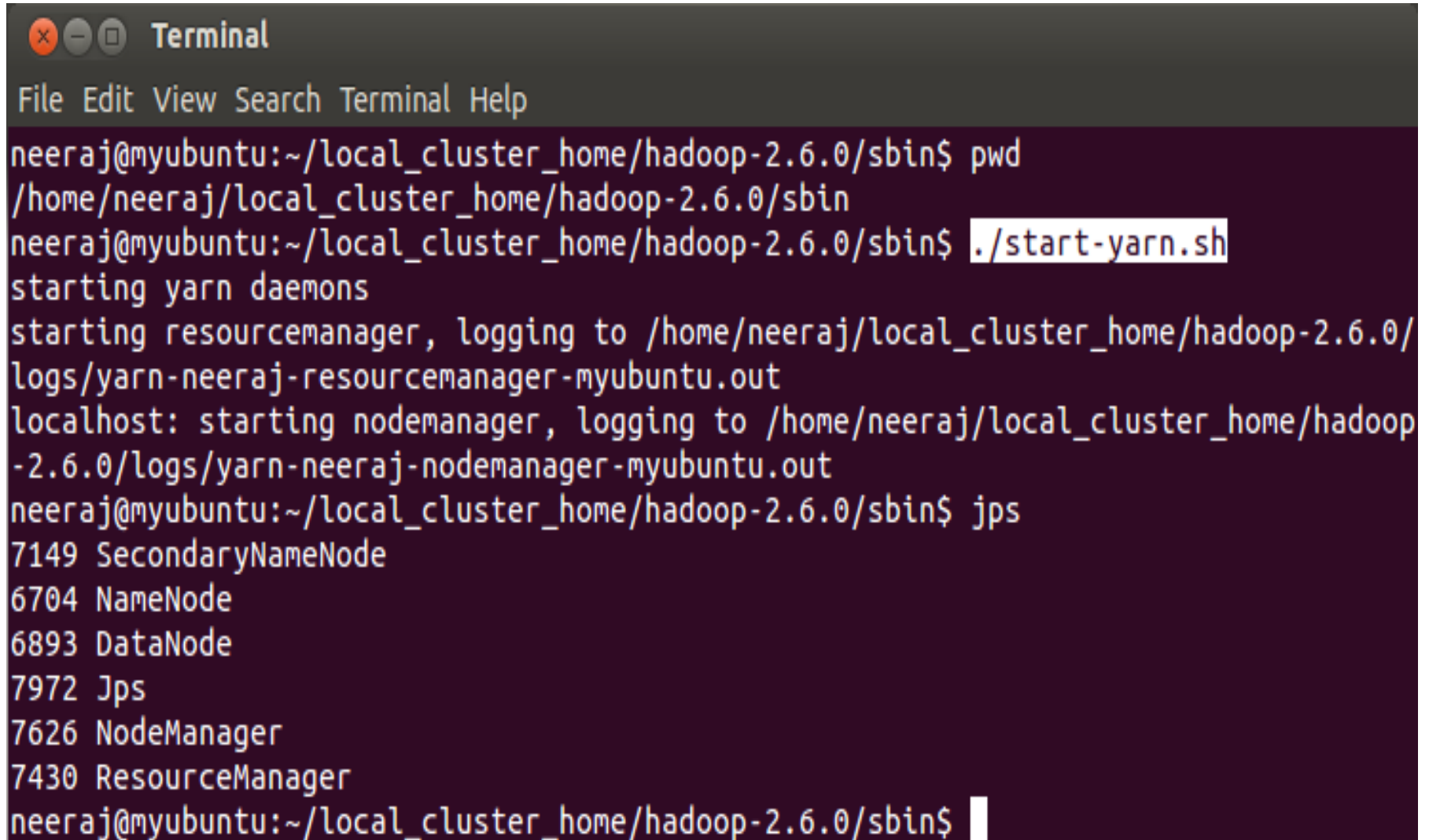
Start HDFS services

`./start-dfs.sh`

```
Terminal
File Edit View Search Terminal Help
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ pwd
/home/neeraj/local_cluster_home/hadoop-2.6.0/sbin
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ ./start-dfs.sh
15/06/05 13:35:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library
  for your platform... using builtin-java classes where applicable
Starting namenodes on [myubuntu]
myubuntu: starting namenode, logging to /home/neeraj/local_cluster_home/hadoop-2.6
.0/logs/hadoop-neeraj-namenode-myubuntu.out
localhost: starting datanode, logging to /home/neeraj/local_cluster_home/hadoop-2.
6.0/logs/hadoop-neeraj-datanode-myubuntu.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/neeraj/local_cluster_home/ha
doo-2.6.0/logs/hadoop-neeraj-secondarynamenode-myubuntu.out
15/06/05 13:36:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library
  for your platform... using builtin-java classes where applicable
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ jps
7149 SecondaryNameNode
6704 NameNode
6893 DataNode
7280 Jps
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$
```

Start YARN services

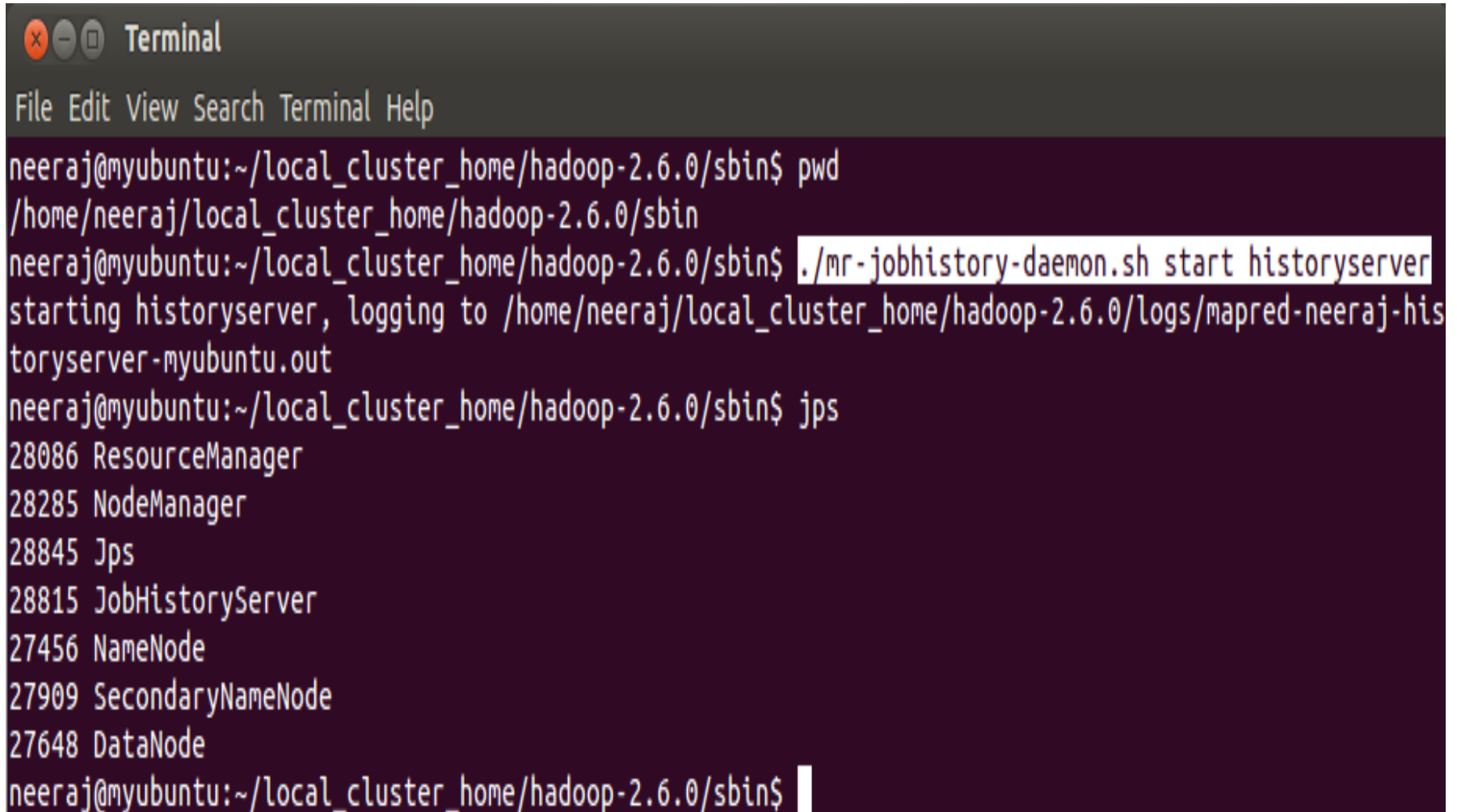
`./start-yarn.sh`



```
Terminal
File Edit View Search Terminal Help
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ pwd
/home/neeraj/local_cluster_home/hadoop-2.6.0/sbin
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ ./start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /home/neeraj/local_cluster_home/hadoop-2.6.0/
logs/yarn-neeraj-resourcemanager-myubuntu.out
localhost: starting nodemanager, logging to /home/neeraj/local_cluster_home/hadoop
-2.6.0/logs/yarn-neeraj-nodemanager-myubuntu.out
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ jps
7149 SecondaryNameNode
6704 NameNode
6893 DataNode
7972 Jps
7626 NodeManager
7430 ResourceManager
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$
```

Start job history server

`./mr-jobhistory-daemon.sh start historyserver`



```
Terminal
File Edit View Search Terminal Help
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ pwd
/home/neeraj/local_cluster_home/hadoop-2.6.0/sbin
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ ./mr-jobhistory-daemon.sh start historyserver
starting historyserver, logging to /home/neeraj/local_cluster_home/hadoop-2.6.0/logs/mapred-neeraj-his
toryserver-myubuntu.out
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ jps
28086 ResourceManager
28285 NodeManager
28845 Jps
28815 JobHistoryServer
27456 NameNode
27909 SecondaryNameNode
27648 DataNode
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$
```

Copy data from local file system to HDFS

`./hdfs dfs -copyFromLocal < local path > < HDFS path >`

```
Terminal
File Edit View Search Terminal Help
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/bin$ pwd
/home/neeraj/local_cluster_home/hadoop-2.6.0/bin
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/bin$ ./hdfs dfs -copyFromLocal /home/neeraj/sample_hadoop.txt /my_1st_HDFS_dir/
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/bin$
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/bin$ ./hdfs dfs -ls /my_1st_HDFS_dir
Found 1 items
-rw-r--r--  1 neeraj supergroup      124 2015-06-05 23:23 /my_1st_HDFS_dir/sample_hadoop.txt
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/bin$
```


Run Map-Reduce job on Hadoop cluster

`./hadoop jar < jar path > < job class name > < input path >
< non existing output path >`

```
Terminal
File Edit View Search Terminal Help
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/bin$ pwd
/home/neeraj/local_cluster_home/hadoop-2.6.0/bin
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/bin$ ./hadoop jar /home/neeraj/local_cluster_home/job_jars/WordCountJar.jar examples.WordCount /my_1st_HDFS_dir/ /wordcount_output
15/06/05 23:38:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
15/06/05 23:38:34 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
15/06/05 23:38:35 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
15/06/05 23:38:38 INFO input.FileInputFormat: Total input paths to process : 1
15/06/05 23:38:38 INFO mapreduce.JobSubmitter: number of splits:1
15/06/05 23:38:39 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1433527644845_0001
15/06/05 23:38:40 INFO impl.YarnClientImpl: Submitted application application_1433527644845_0001
15/06/05 23:38:40 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1433527644845_0001/
15/06/05 23:38:40 INFO mapreduce.Job: Running job: job_1433527644845_0001
15/06/05 23:39:17 INFO mapreduce.Job: Job job_1433527644845_0001 running in uber mode : false
15/06/05 23:39:17 INFO mapreduce.Job:  map 0% reduce 0%
15/06/05 23:39:33 INFO mapreduce.Job:  map 100% reduce 0%
15/06/05 23:39:50 INFO mapreduce.Job:  map 100% reduce 100%
15/06/05 23:39:51 INFO mapreduce.Job: Job job_1433527644845_0001 completed successfully
```

Browse HDFS data on browser

<http://localhost:50070>

Namenode information

localhost:50070/dfshealth.html#tab-overview

Google

Hadoop

Overview

Datanodes

Snapshot

Startup Progress

Utilities

Browse the file system

Logs

Overview 'myubuntu:9000' (active)

Started:	Sat Jun 27 10:44:54 IST 2015
Version:	2.6.0, re3496499ecb8d220fba99dc5ed4c99c8f9e33bb1
Compiled:	2014-11-13T21:10Z by jenkins from (detached from e349649)
Cluster ID:	CID-e3b47471-75fb-4c42-998a-f84b91786f28
Block Pool ID:	BP-902430900-127.0.0.1-1435028128080

Browse HDFS data on browser

<http://localhost:50070>

Browsing HDFS

localhost:50070/explorer.html#/

Google

HadoopOverviewDatanodesSnapshotStartup ProgressUtilities

Browse Directory

Go!

Permission	Owner	Group	Size	Replication	Block Size	Name
drwxr-xr-x	neeraj	supergroup	0 B	0	0 B	my_1st_HDFS_dir
-rw-r--r--	neeraj	supergroup	124 B	1	128 MB	sample_hadoop.txt
drwx-----	neeraj	supergroup	0 B	0	0 B	tmp
drwxr-xr-x	neeraj	supergroup	0 B	0	0 B	wordcount_output

Browse applications on browser


<http://localhost:8088>

All Applications

+

localhost:8088/cluster

Google



All Applications

Cluster

[About](#)
[Nodes](#)
[Applications](#)
[NEW](#)
[NEW SAVING](#)
[SUBMITTED](#)
[ACCEPTED](#)
[RUNNING](#)
[FINISHED](#)
[FAILED](#)
[KILLED](#)
[Scheduler](#)

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total
1	0	1	0	2	3 GB	8 GB	0 B	2	8

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	Finish
application_1433541519583_0001	neeraj	Word Count	MAPREDUCE	default	Fri, 05 Jun 2015 22:02:43 GMT	N/A

Showing 1 to 1 of 1 entries


Browse job history on browser

<http://localhost:19888>

JobHistory

localhost:19888/jobhistory

Google



JobHistory

▼ Application

[About](#)
[Jobs](#)

► Tools

Retired Jobs

Show 20 entries

Search:

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed
2015.06.06 03:32:43 IST	2015.06.06 03:34:07 IST	2015.06.06 03:35:37 IST	job_1433541519583_0001	Word Count	neeraj	default	SUCCEEDED	1	1
2015.06.05 23:38:39 IST	2015.06.05 23:39:15 IST	2015.06.05 23:39:49 IST	job_1433527644845_0001	Word Count	neeraj	default	SUCCEEDED	1	1
2015.06.05 23:33:53 IST	2015.06.05 23:34:11 IST	2015.06.05 23:34:31 IST	job_1433527289584_0001	Word Count	neeraj	default	SUCCEEDED	1	1
2015.06.05 23:26:07 IST	2015.06.05 23:26:25 IST	2015.06.05 23:26:53 IST	job_1433525702805_0001	Word Count	neeraj	default	SUCCEEDED	1	1

Showing 1 to 4 of 4 entries

First Pre

Stop job history server

`./mr-jobhistory-daemon.sh stop historyserver`

```
Terminal
File Edit View Search Terminal Help
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ pwd
/home/neeraj/local_cluster_home/hadoop-2.6.0/sbin
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ ./mr-jobhistory-daemon.sh stop historyserver
stopping historyserver
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ jps
5243 SecondaryNameNode
5657 NodeManager
6151 Jps
4792 NameNode
4984 DataNode
5459 ResourceManager
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$
```

Stop YARN services

`./stop-yarn.sh`

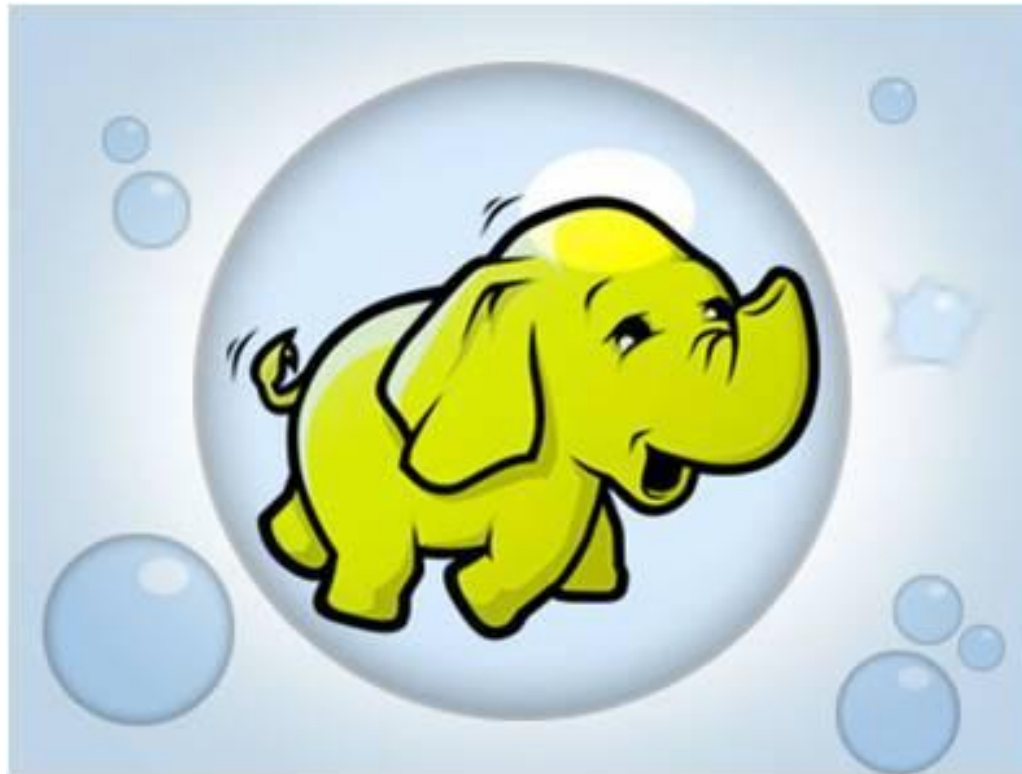
```
Terminal
File Edit View Search Terminal Help
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ pwd
/home/neeraj/local_cluster_home/hadoop-2.6.0/sbin
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ ./stop-yarn.sh
stopping yarn daemons
stopping resourcemanager
localhost: stopping nodemanager
no proxyserver to stop
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ jps
6467 Jps
5243 SecondaryNameNode
4792 NameNode
4984 DataNode
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$
```

Stop HDFS services

`./stop-dfs.sh`

```
Terminal
File Edit View Search Terminal Help
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ pwd
/home/neeraj/local_cluster_home/hadoop-2.6.0/sbin
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ ./stop-dfs.sh
15/06/27 11:42:40 WARN util.NativeCodeLoader: Unable to load native-hadoop
m... using builtin-java classes where applicable
Stopping namenodes on [myubuntu]
myubuntu: stopping namenode
localhost: stopping datanode
Stopping secondary namenodes [0.0.0.0]
0.0.0.0: stopping secondarynamenode
15/06/27 11:44:23 WARN util.NativeCodeLoader: Unable to load native-hadoop
m... using builtin-java classes where applicable
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$ jps
5503 Jps
neeraj@myubuntu:~/local_cluster_home/hadoop-2.6.0/sbin$
```

...Thanks...



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