

Hive Installation & Basic Commands



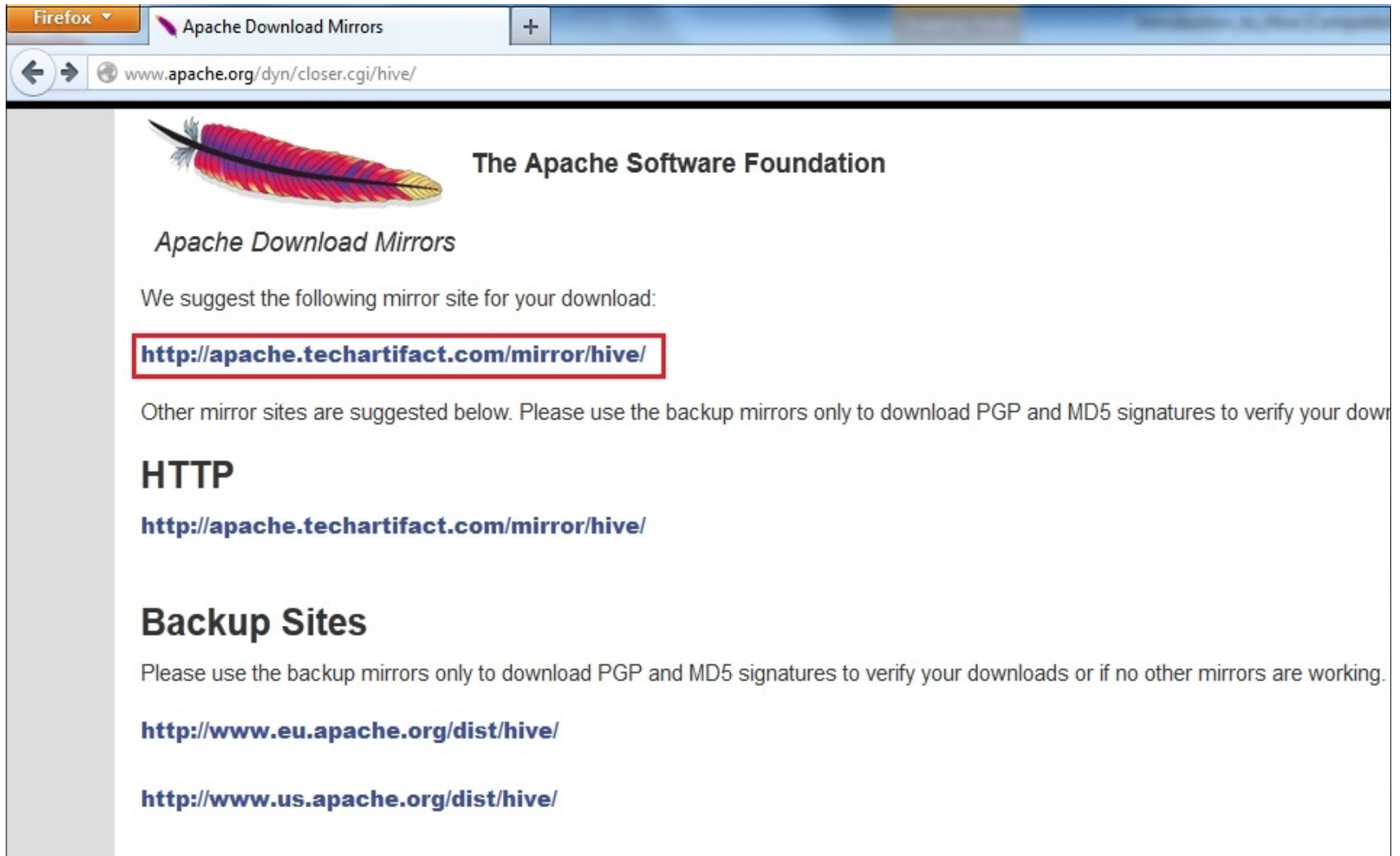
For online Hadoop training, send mail to neeraj.ymca.2k6@gmail.com

Agenda

- Download Hive from Apache website
- Extract the Hive tar.gz file
- Configure your Hive
- Start your Hadoop
- Check safe mode status
- Start Hive shell
- Input data for Hive table
- Create table in Hive
- Load data into Hive table
- Where clause in Hive
- Count number of rows in a table
- Delete table in Hive
- Exit from Hive shell

Download Hive from Apache website


<http://www.apache.org/dyn/closer.cgi/hive/>



The screenshot shows a Firefox browser window with the address bar displaying www.apache.org/dyn/closer.cgi/hive/. The page content includes the Apache Software Foundation logo (a feather) and the text "The Apache Software Foundation". Below this, it says "Apache Download Mirrors". A message states: "We suggest the following mirror site for your download:". The suggested mirror URL, <http://apache.techartifact.com/mirror/hive/>, is highlighted with a red rectangular border. Below this, it says "Other mirror sites are suggested below. Please use the backup mirrors only to download PGP and MD5 signatures to verify your download". The page then lists "HTTP" mirrors with the URL <http://apache.techartifact.com/mirror/hive/>. Finally, it lists "Backup Sites" with two URLs: <http://www.eu.apache.org/dist/hive/> and <http://www.us.apache.org/dist/hive/>.

Firefox ▾ Apache Download Mirrors +

← → www.apache.org/dyn/closer.cgi/hive/



The Apache Software Foundation

Apache Download Mirrors

We suggest the following mirror site for your download:

<http://apache.techartifact.com/mirror/hive/>

Other mirror sites are suggested below. Please use the backup mirrors only to download PGP and MD5 signatures to verify your download.

HTTP

<http://apache.techartifact.com/mirror/hive/>

Backup Sites

Please use the backup mirrors only to download PGP and MD5 signatures to verify your downloads or if no other mirrors are working.

<http://www.eu.apache.org/dist/hive/>

<http://www.us.apache.org/dist/hive/>

Select the stable version

 Index of /hive 

   apache.mesi.com.ar/hive/

Index of /hive

	<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
	Parent Directory	10-Aug-2013 07:00	-	
	hive-0.10.0/	18-Dec-2012 20:30	-	
	hive-0.11.0/	11-May-2013 14:02	-	
	stable/	18-Dec-2012 20:30	-	

Apache/1.3.42 Server at apache.mesi.com.ar Port 80

Download the tar.gz file

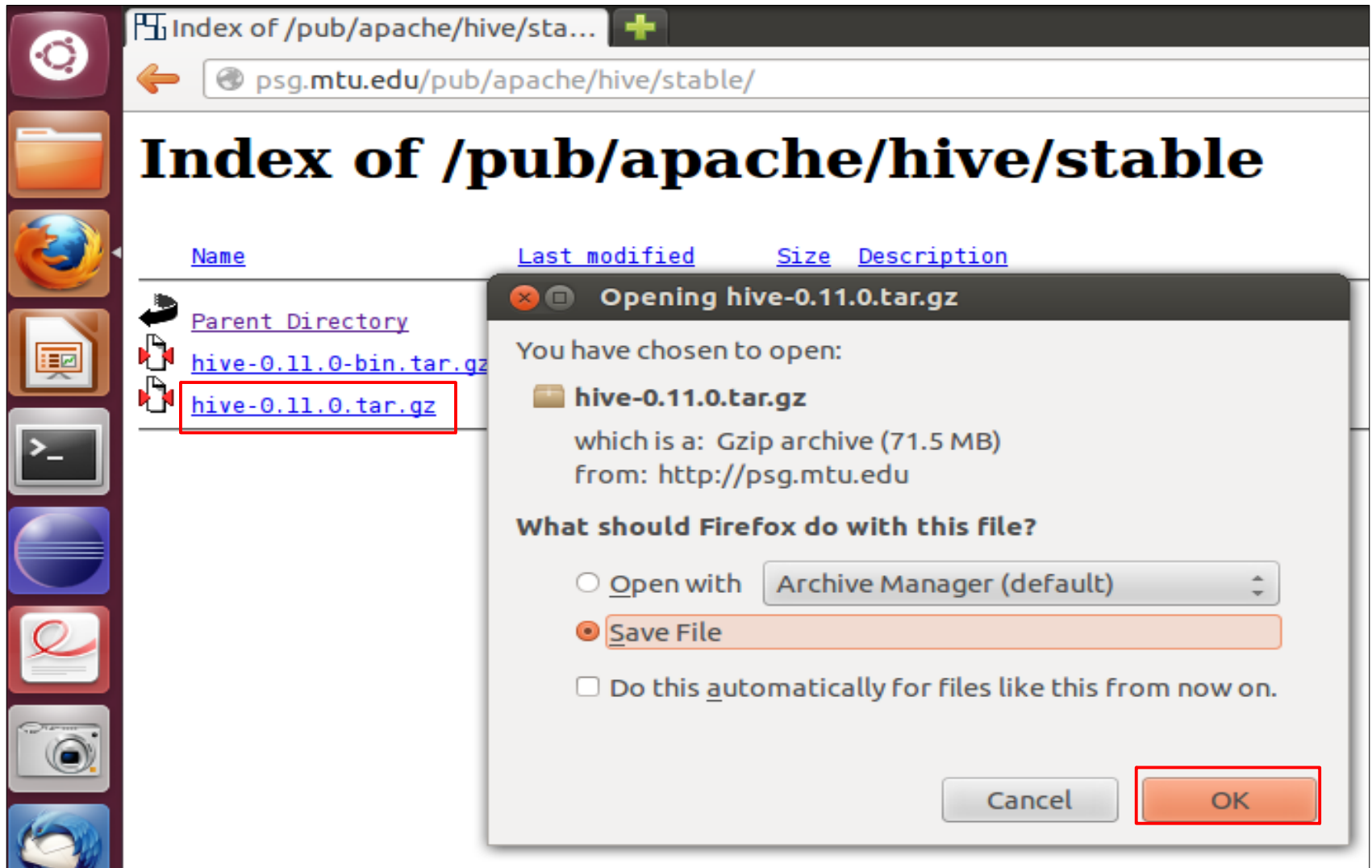


The screenshot shows a web browser window with the address bar displaying `psg.mtu.edu/pub/apache/hive/stable/`. The page title is "Index of /pub/apache/hive/stable". Below the title is a table with columns: Name, Last modified, Size, and Description. The table lists the following items:

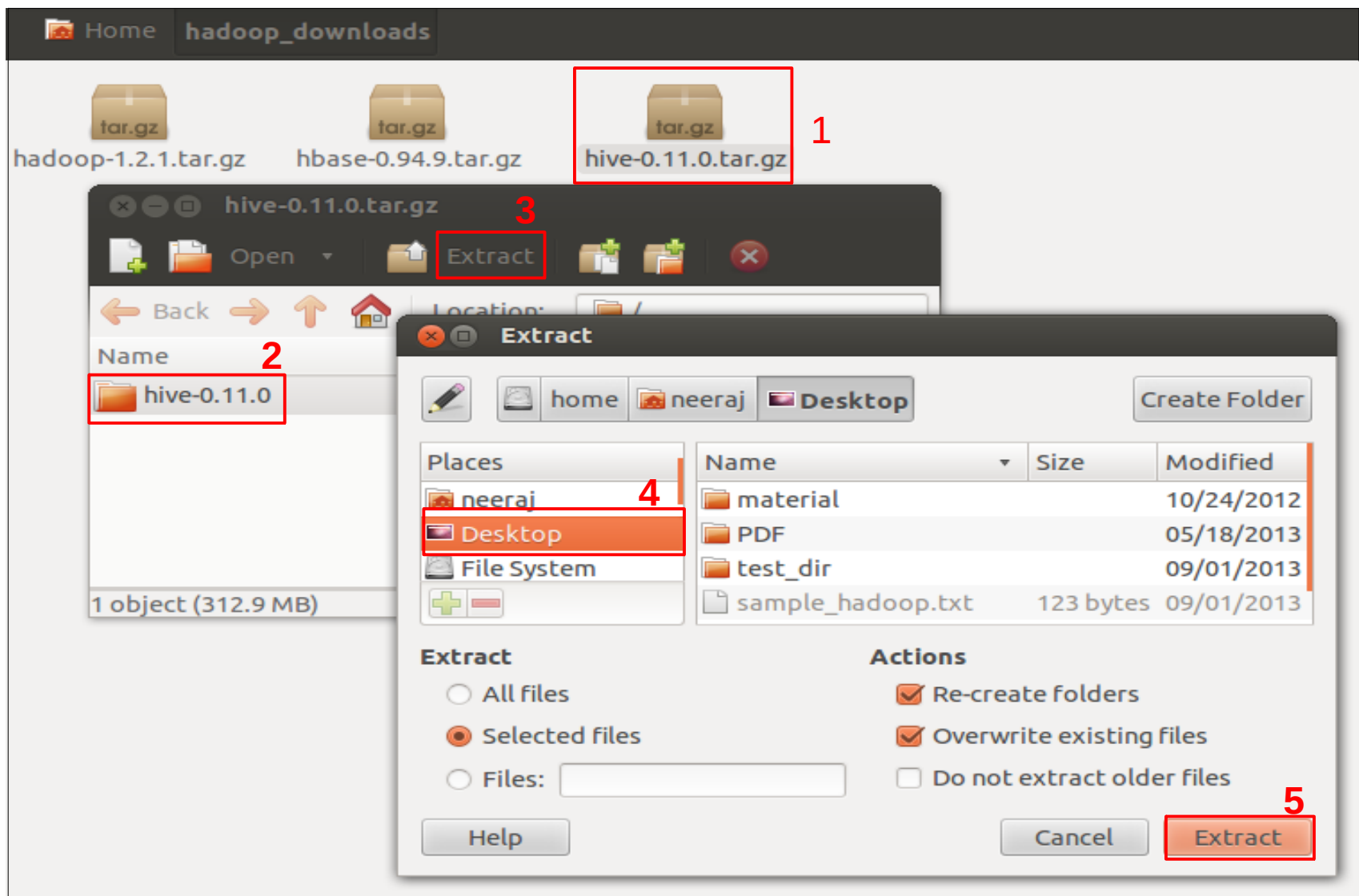
Name	Last modified	Size	Description
Parent Directory	-	-	-
hive-0.11.0-bin.tar.gz	11-May-2013 13:00	57M	
hive-0.11.0.tar.gz	11-May-2013 13:02	71M	

The file `hive-0.11.0.tar.gz` is highlighted with a red dashed border. The browser's left sidebar shows various application icons, including a terminal, a file manager, and a web browser.

Download the tar.gz file



Untar the Hive tar.gz file



Configure your Hive

Rename HIVE_HOME/conf/hive-default.xml.template to hive-default.xml

Rename HIVE_HOME/conf/hive-env.sh.template to hive-env.sh

Make the following entry in HIVE_HOME/conf/hive-env.sh file

```
export HADOOP_HOME=/home/neeraj/local_cluster_home/hadoop-1.2.1
```


Start your Hadoop

```
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$
```

Check Hadoop processes & Safemode

Make sure that safe mode is off before you start Hive

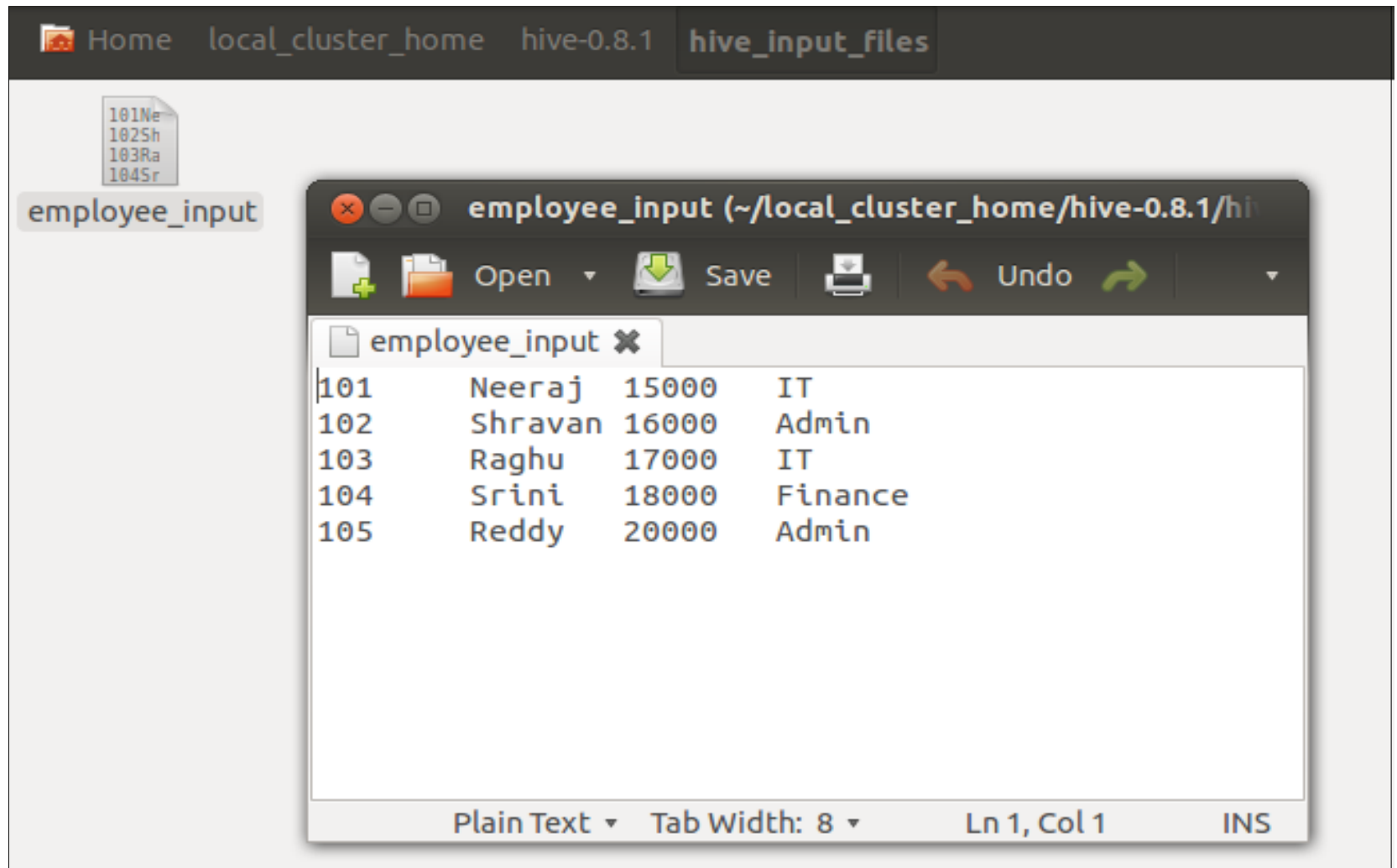
```
Terminal
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$ pwd
/home/neeraj/local_cluster_home/hadoop-1.2.1/bin
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$ jps
3915 TaskTracker
4178 Jps
3610 SecondaryNameNode
3399 DataNode
3159 NameNode
3701 JobTracker
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$ ./hadoop dfsadmin -safemode get
Safe mode is OFF
neeraj@ubuntu:~/local_cluster_home/hadoop-1.2.1/bin$
```

Start Hive Shell

```
Terminal
neeraj@ubuntu:~/local_cluster_home/hive-0.11.0/bin$ pwd
/home/neeraj/local_cluster_home/hive-0.11.0/bin
neeraj@ubuntu:~/local_cluster_home/hive-0.11.0/bin$ ./hive

Logging initialized using configuration in jar:file:/home/neeraj/l
on-0.11.0.jar!/hive-log4j.properties
Hive history file=/tmp/neeraj/hive_job_log_neeraj_4418@ubuntu_2013
hive>
>
> SHOW TABLES;
OK
employee
Time taken: 7.727 seconds, Fetched: 1 row(s)
hive> 
```

Input data for Hive table



The screenshot shows a file editor window titled "employee_input (~/.local_cluster_home/hive-0.8.1/hive_input_files)". The editor contains a text file named "employee_input" with the following content:

101	Neeraj	15000	IT
102	Shravan	16000	Admin
103	Raghu	17000	IT
104	Srini	18000	Finance
105	Reddy	20000	Admin

The editor interface includes a menu bar with "Open", "Save", and "Undo" options. The status bar at the bottom indicates "Plain Text", "Tab Width: 8", "Ln 1, Col 1", and "INS".

Create table in Hive

```
Terminal
hive> show tables;
OK
Time taken: 0.05 seconds
hive> CREATE TABLE employee(
  > id int,
  > name String,
  > salary int,
  > department String
  > )
  > ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE;
OK
Time taken: 0.822 seconds
hive> show tables;
OK
employee
Time taken: 0.083 seconds
hive> 
```

Load data into Hive table

```
Terminal
hive> SHOW TABLES;
OK
employee
Time taken: 6.5 seconds
hive> LOAD DATA LOCAL INPATH
    > "/home/neeraj/local_cluster_home/hive-0.8.1/hive_input_files/employee_input"
    > INTO TABLE employee;
Copying data from file:/home/neeraj/local_cluster_home/hive-0.8.1/hive_input_files/employee_input
Copying file: file:/home/neeraj/local_cluster_home/hive-0.8.1/hive_input_files/employee_input
Loading data to table default.employee
OK
Time taken: 1.186 seconds
hive> SELECT * FROM employee;
OK
101      Neeraj  15000   IT
102      Shravan 16000   Admin
103      Raghu   17000   IT
104      Srini   18000   Finance
105      Reddy   20000   Admin
Time taken: 0.755 seconds
hive> 
```

Where clause in Hive

```
Terminal
hive> SHOW TABLES;
OK
employee
Time taken: 0.063 seconds
hive> SELECT * FROM employee WHERE id=103;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_201212300844_0003, Tracking URL = http://myubuntu:50030/jobdetails
44_0003
Kill Command = /home/neeraj/local_cluster_home/hadoop-1.0.3/libexec/./bin/hadoop job
ubuntu:9001 -kill job_201212300844_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2012-12-30 09:57:06,019 Stage-1 map = 0%, reduce = 0%
2012-12-30 09:57:12,063 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.32 sec
2012-12-30 09:57:13,073 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.32 sec
2012-12-30 09:57:14,084 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.32 sec
2012-12-30 09:57:15,094 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.32 sec
2012-12-30 09:57:16,104 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.32 sec
2012-12-30 09:57:17,114 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.32 sec
2012-12-30 09:57:18,124 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 1.32 sec
MapReduce Total cumulative CPU time: 1 seconds 320 msec
Ended Job = job_201212300844_0003
MapReduce Jobs Launched:
Job 0: Map: 1 Accumulative CPU: 1.32 sec HDFS Read: 327 HDFS Write: 19 SUCESS
Total MapReduce CPU Time Spent: 1 seconds 320 msec
OK
103 Raghu 17000 IT
Time taken: 27.768 seconds
```

Count no. of rows in a table

```
Terminal
hive> SHOW TABLES;
OK
employee
Time taken: 5.523 seconds
hive> SELECT COUNT(*) FROM employee;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_201212300844_0002, Tracking URL = http://myubuntu:50030/jobdetails.jsp?jobid=job_201212300844_0002
Kill Command = /home/neeraj/local_cluster_home/hadoop-1.0.3/libexec/./bin/hadoop job -Dmapred.job.tracker=myubuntu:9001 -kill job_201212300844_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2012-12-30 09:32:51,718 Stage-1 map = 0%, reduce = 0%
2012-12-30 09:32:57,761 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:32:58,770 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:32:59,791 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:33:00,802 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:33:01,811 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:33:02,826 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:33:03,836 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:33:04,846 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:33:05,855 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:33:06,865 Stage-1 map = 100%, reduce = 33%, Cumulative CPU 1.11 sec
2012-12-30 09:33:07,875 Stage-1 map = 100%, reduce = 33%, Cumulative CPU 1.11 sec
2012-12-30 09:33:08,886 Stage-1 map = 100%, reduce = 33%, Cumulative CPU 1.11 sec
2012-12-30 09:33:09,896 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.0 sec
```

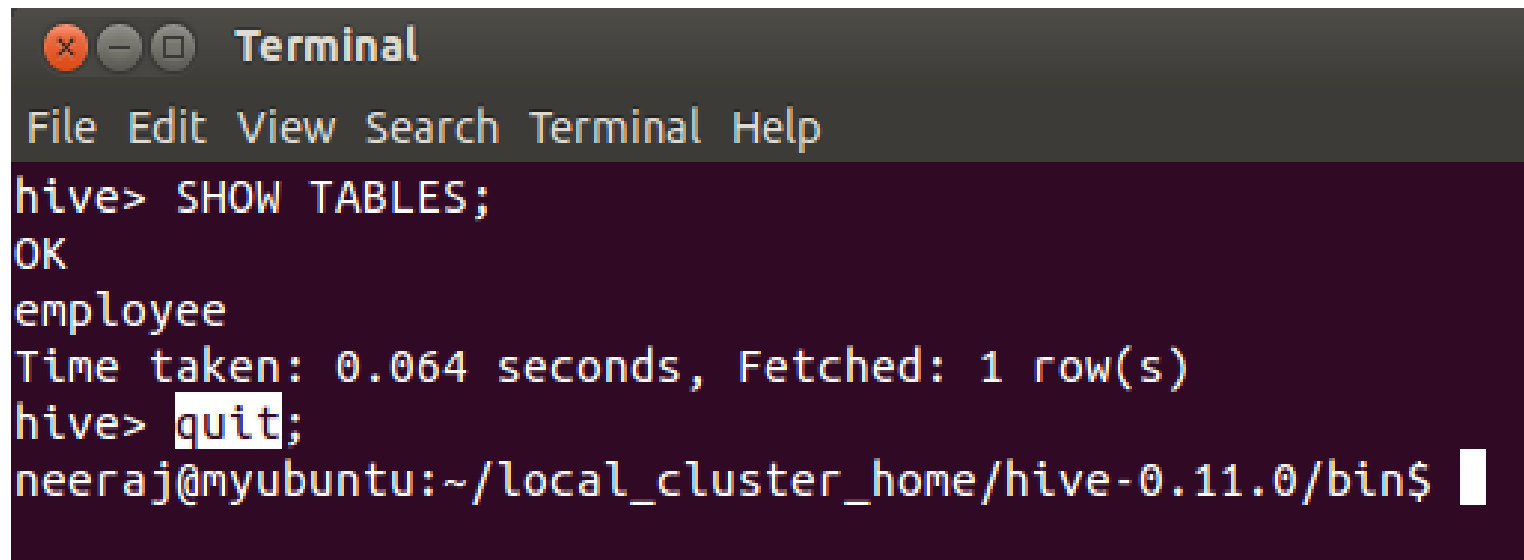

Count no. of rows in a table

```
2012-12-30 09:32:51,718 Stage-1 map = 0%, reduce = 0%
2012-12-30 09:32:57,761 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:32:58,770 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:32:59,791 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:33:00,802 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:33:01,811 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:33:02,826 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:33:03,836 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:33:04,846 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:33:05,855 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.11 sec
2012-12-30 09:33:06,865 Stage-1 map = 100%, reduce = 33%, Cumulative CPU 1.11 sec
2012-12-30 09:33:07,875 Stage-1 map = 100%, reduce = 33%, Cumulative CPU 1.11 sec
2012-12-30 09:33:08,886 Stage-1 map = 100%, reduce = 33%, Cumulative CPU 1.11 sec
2012-12-30 09:33:09,896 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.0 sec
2012-12-30 09:33:10,904 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.0 sec
2012-12-30 09:33:12,007 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.0 sec
2012-12-30 09:33:13,019 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.0 sec
2012-12-30 09:33:14,030 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.0 sec
2012-12-30 09:33:15,043 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.0 sec
2012-12-30 09:33:16,052 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.0 sec
MapReduce Total cumulative CPU time: 3 seconds 0 msec
Ended Job = job_201212300844_0002
MapReduce Jobs Launched:
Job 0: Map: 1 Reduce: 1 Accumulative CPU: 3.0 sec HDFS Read: 327 HDFS Write: 2 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 0 msec
OK
5
Time taken: 47.153 seconds
```

Delete table in Hive

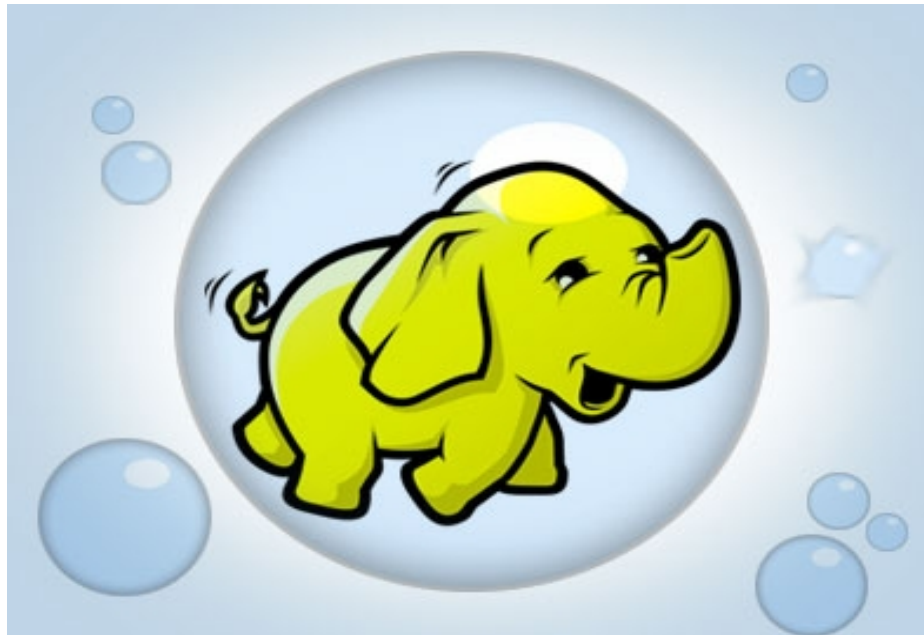
```
Terminal
hive> SHOW TABLES;
OK
employee
Time taken: 0.062 seconds
hive> DROP TABLE employee;
OK
Time taken: 0.246 seconds
hive> SHOW TABLES;
OK
Time taken: 0.061 seconds
hive> 
```

Exit from Hive shell



```
Terminal
File Edit View Search Terminal Help
hive> SHOW TABLES;
OK
employee
Time taken: 0.064 seconds, Fetched: 1 row(s)
hive> quit;
neeraj@myubuntu:~/local_cluster_home/hive-0.11.0/bin$
```

...Thanks...



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