Hive Installation
&
Basic Commands

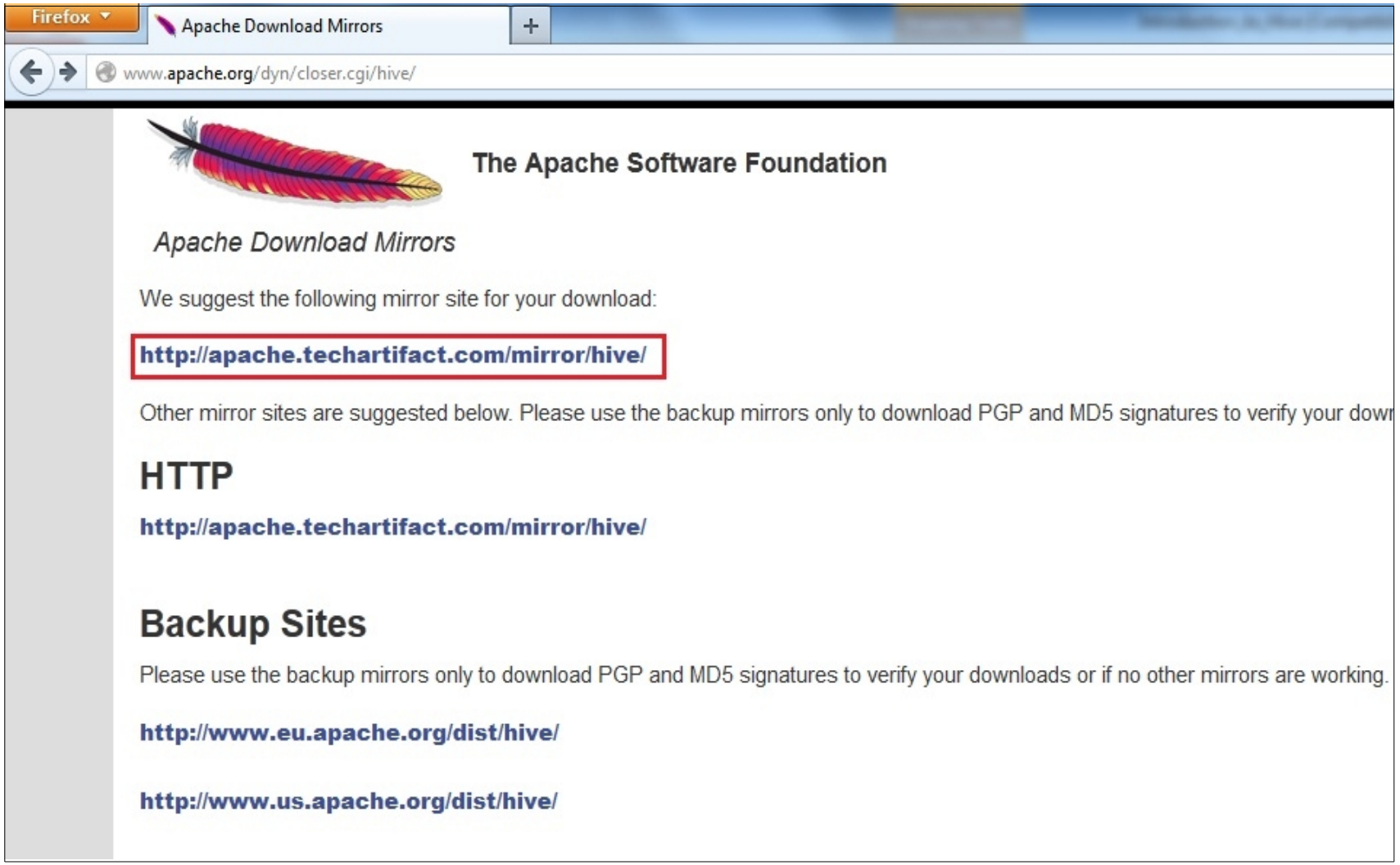
For online Hadoop training, send mail to neeraj.ymca.2k6@gmail.com

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

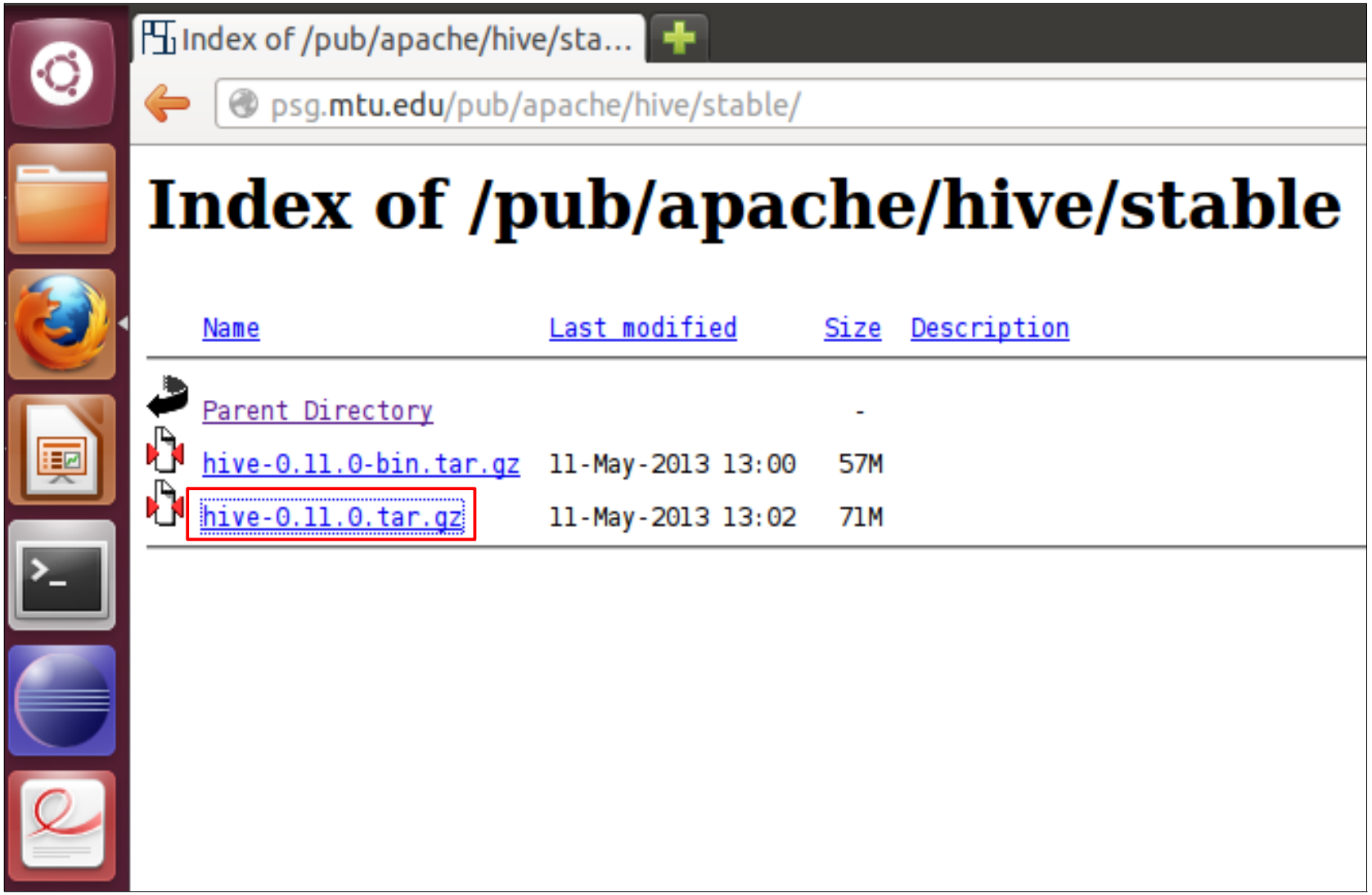
Download Hive from Apache website

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

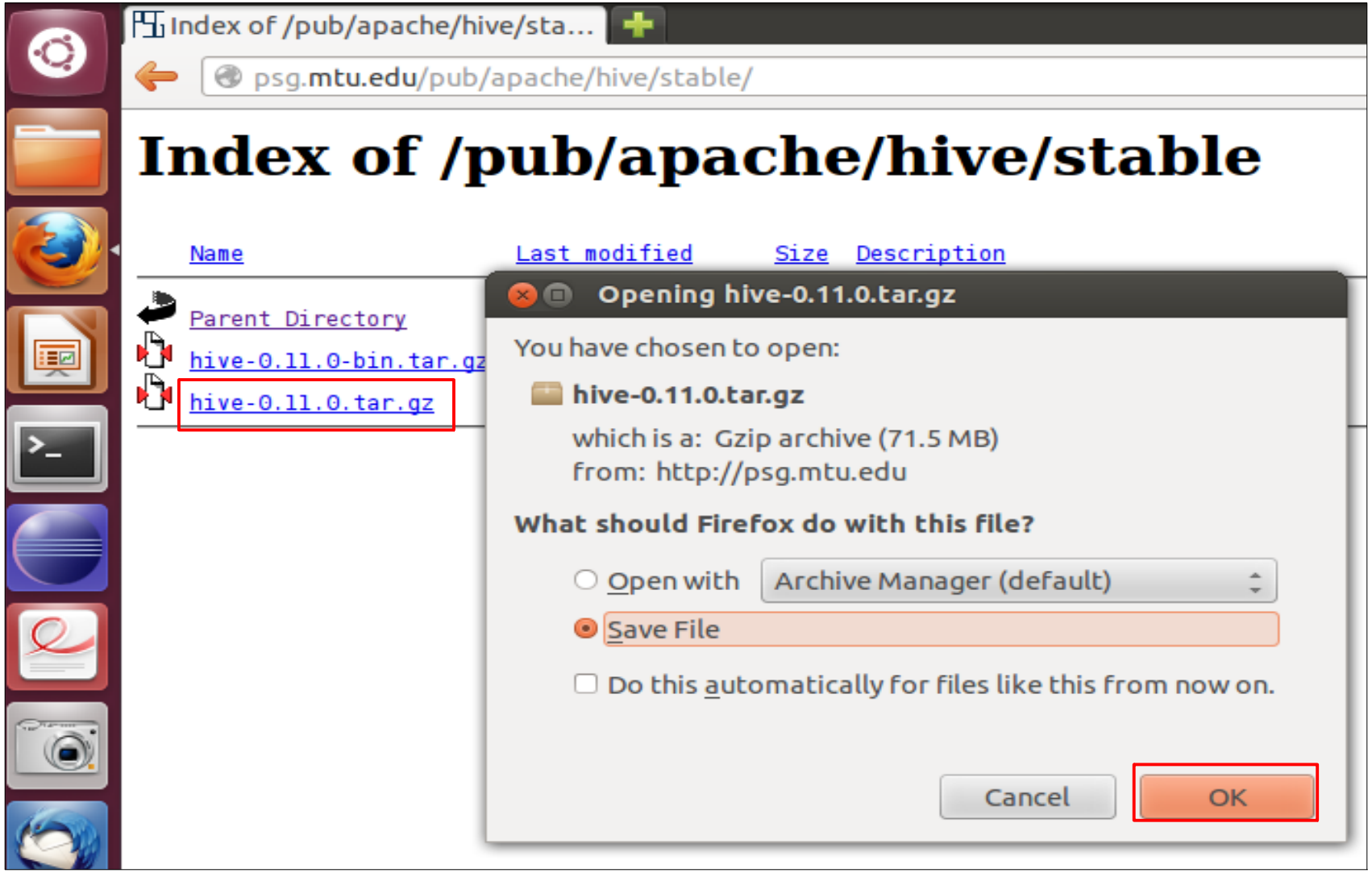


Select the stable version

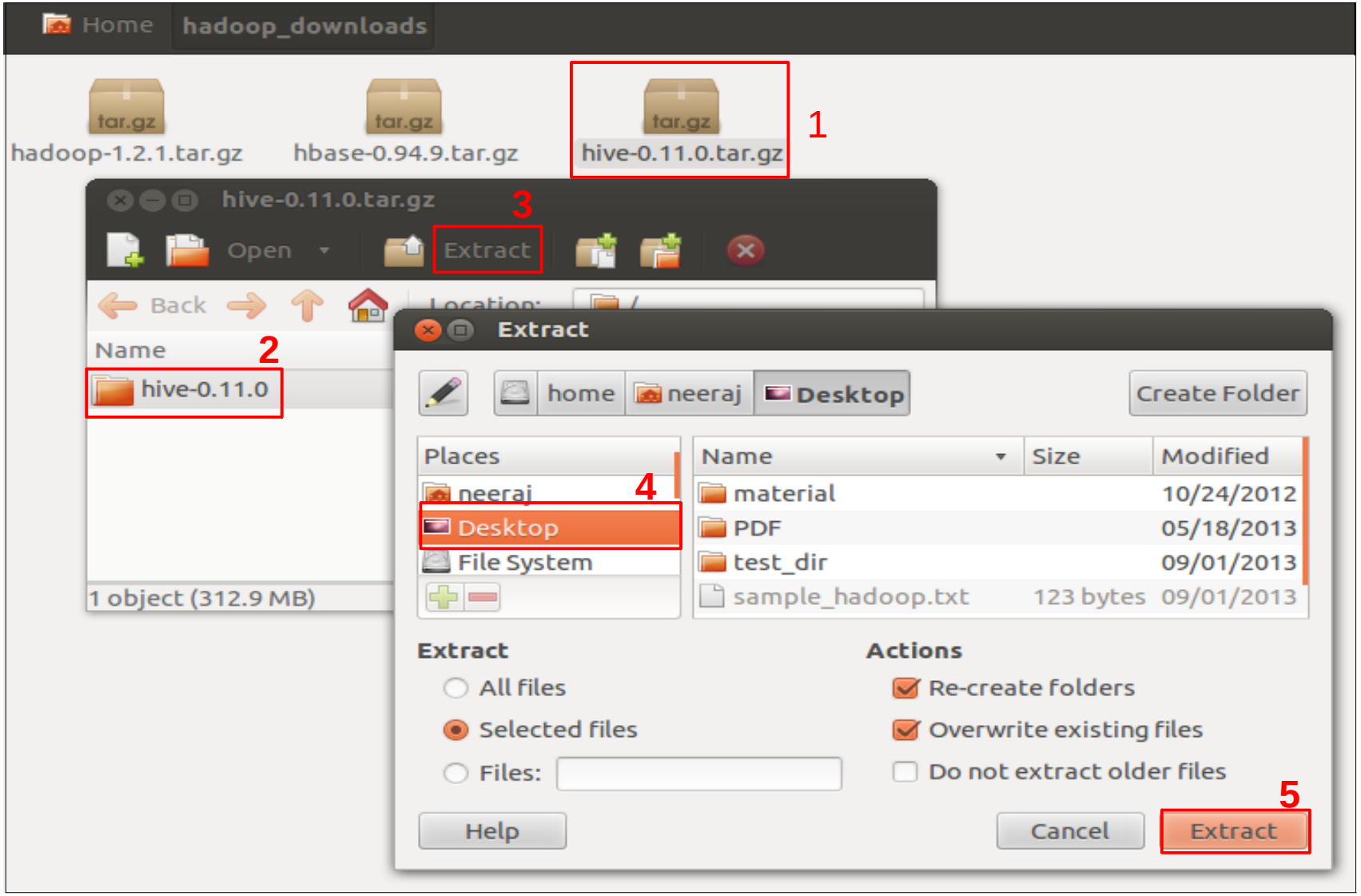
Download the tar.gz file



Download the tar.gz file



Untar the Hive tar.gz file



Configure your Hive

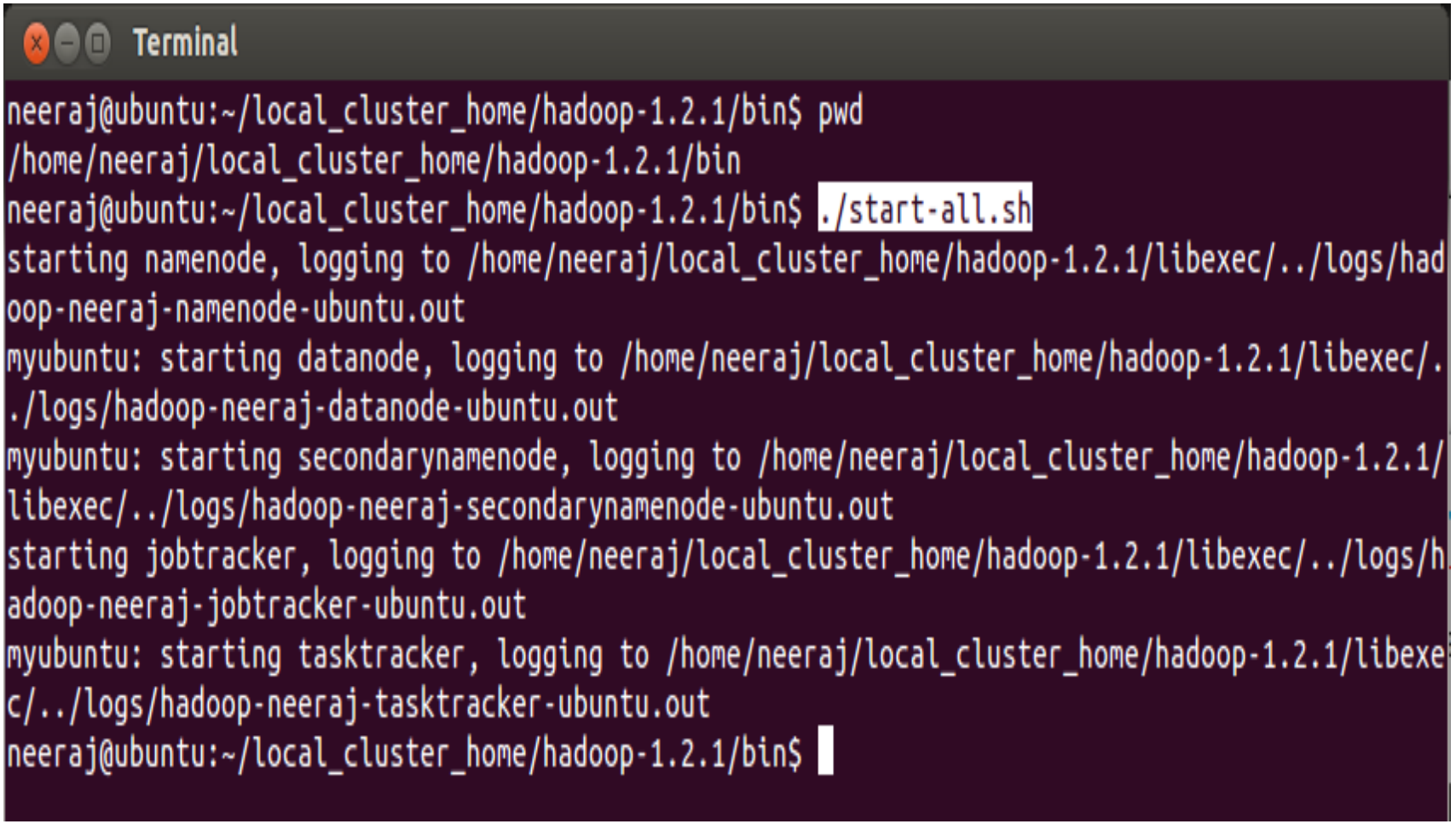
Rename HIVE_HOME/conf/hive-default.xml.template to hive-default.xml

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

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

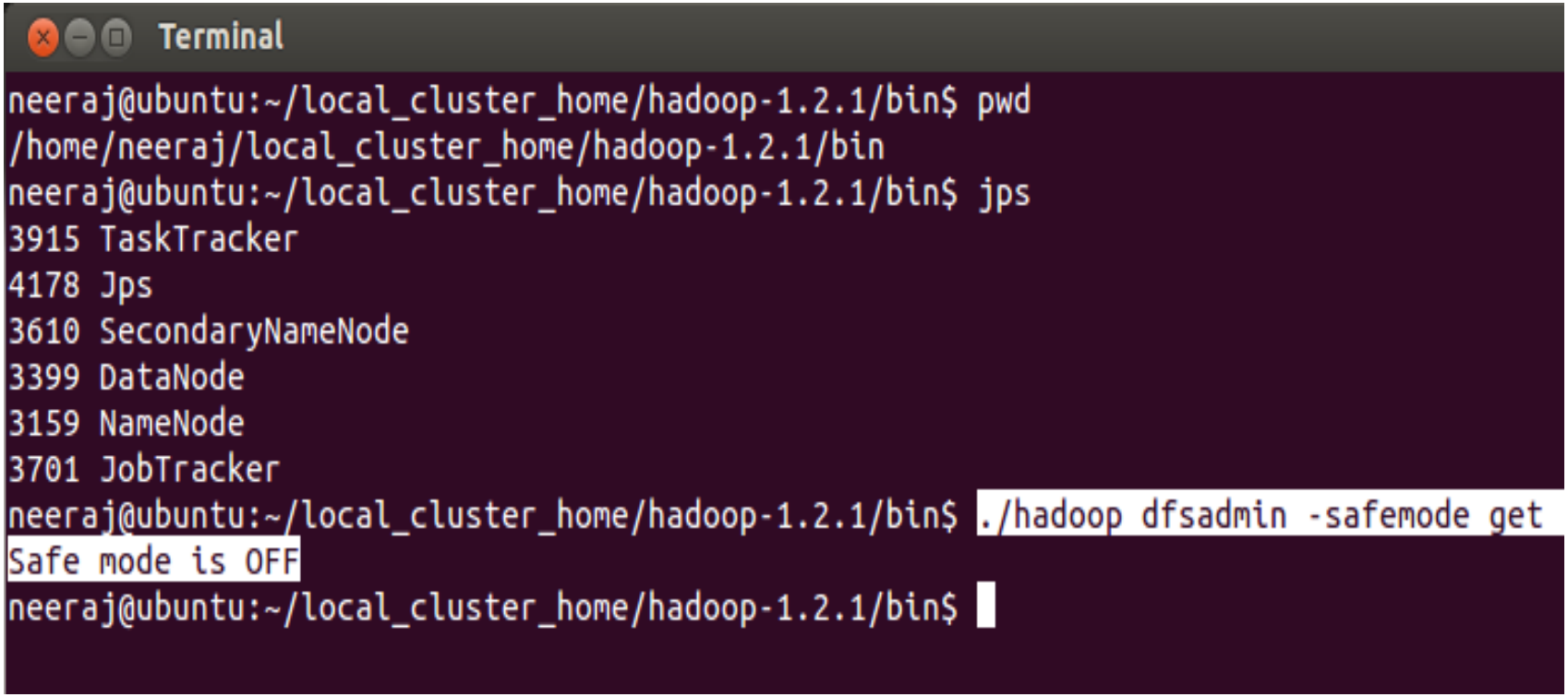
export HADOOP_HOME=/home/neeraj/local_cluster_home/hadoop-1.2.1

Start your Hadoop



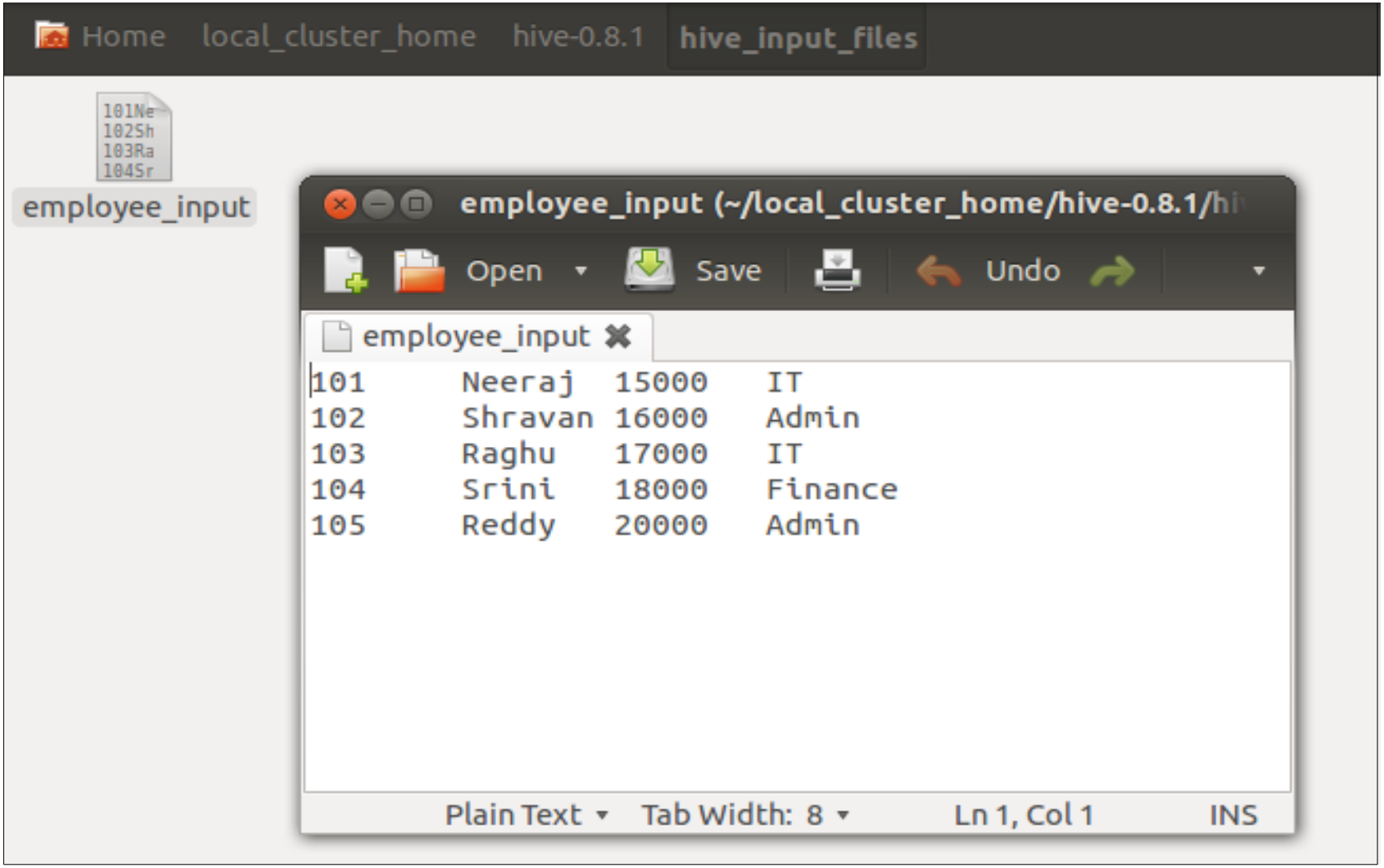
Check Hadoop processes
&
Safemode

Make sue that safe mode is off before you start Hive

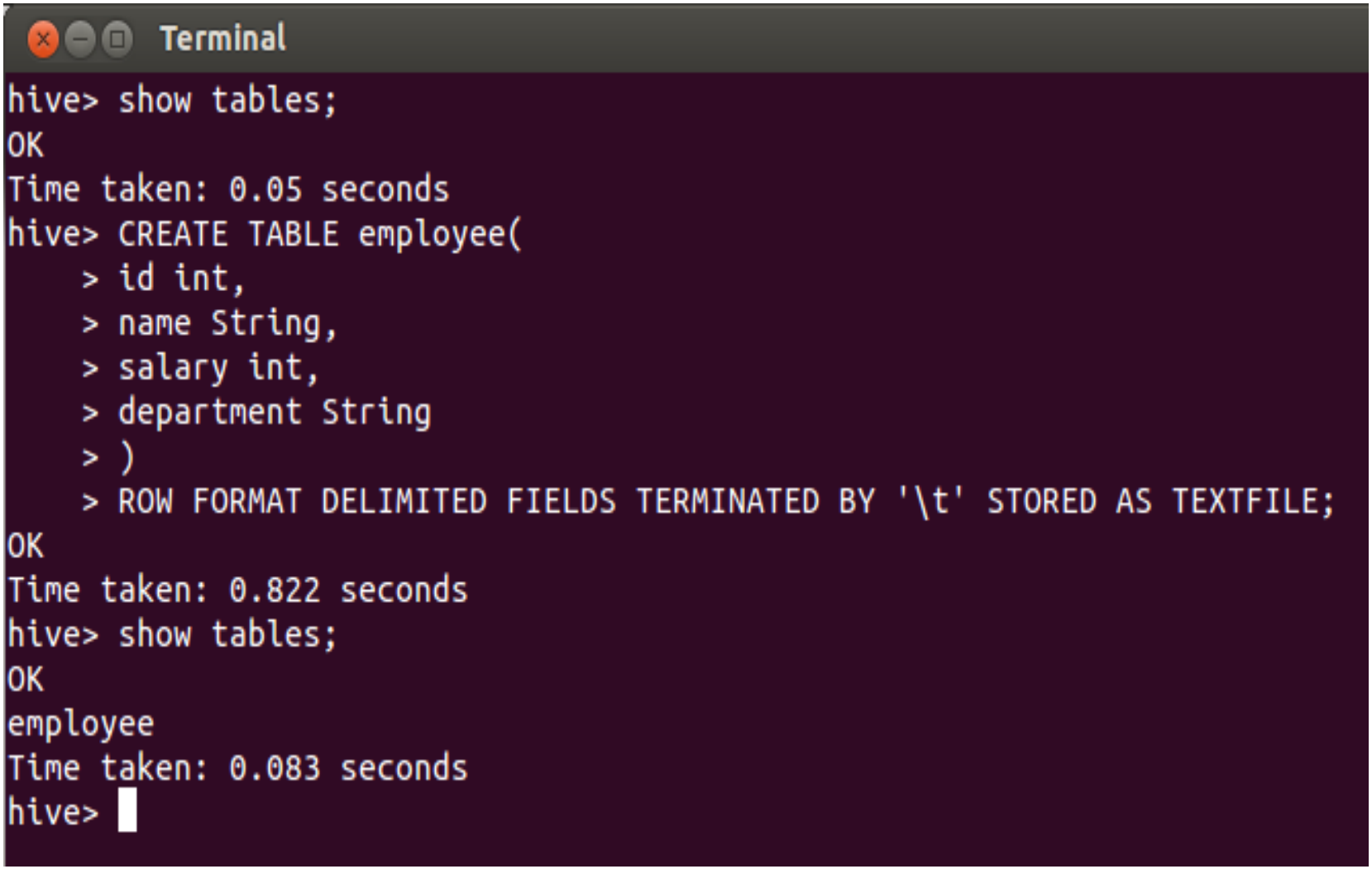


Start Hive Shell

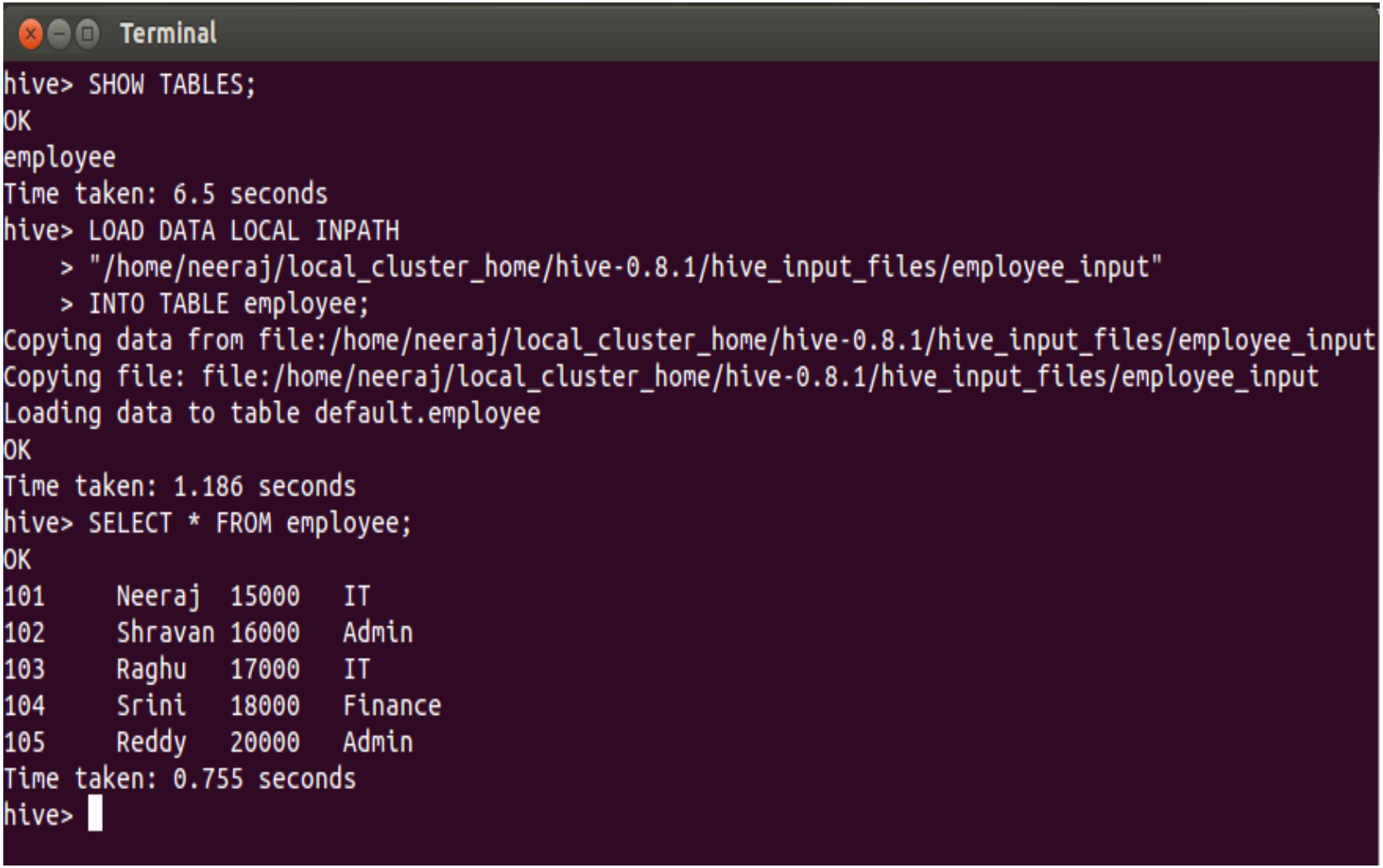
Input data for Hive table



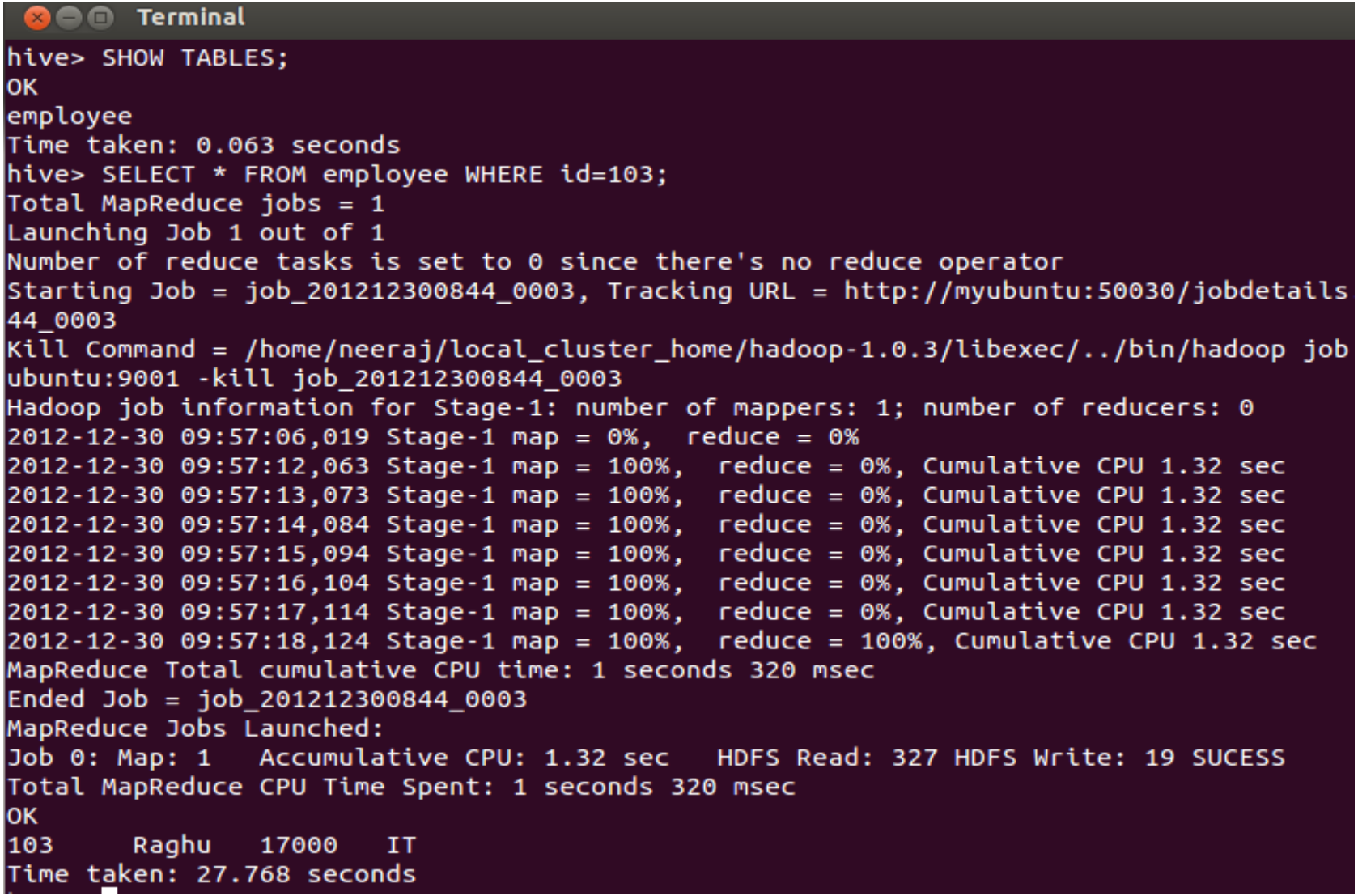
Create table in Hive



Load data into Hive table



Where clause in Hive



Count no. of rows in a table

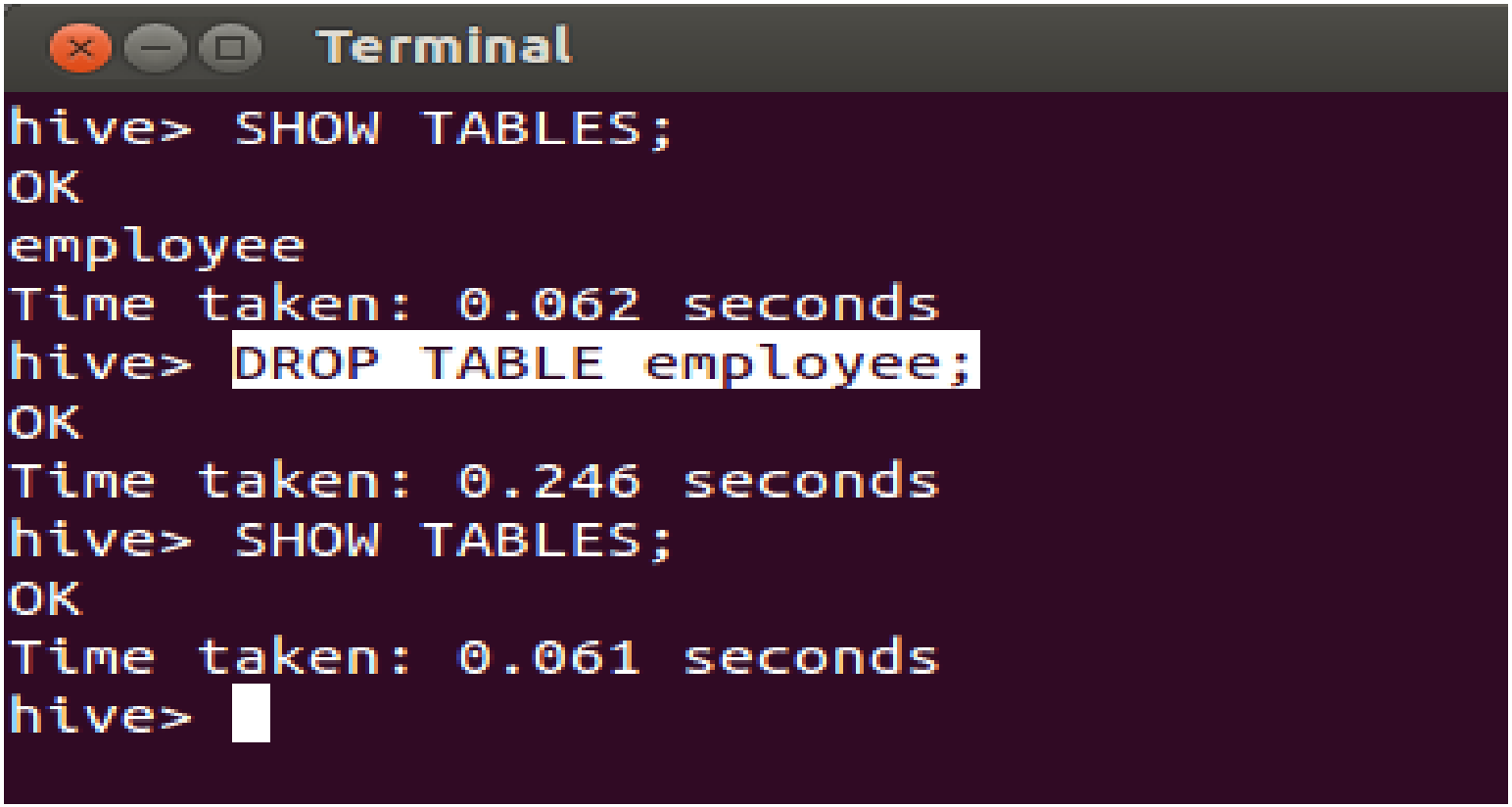
Terminal

```
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_2012123008
44_0002
Kill Command = /home/neeraj/local_cluster_home/hadoop-1.0.3/libexec/../bin/hadoop job  -Dmapred.job.tracker=my
ubuntu: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
```

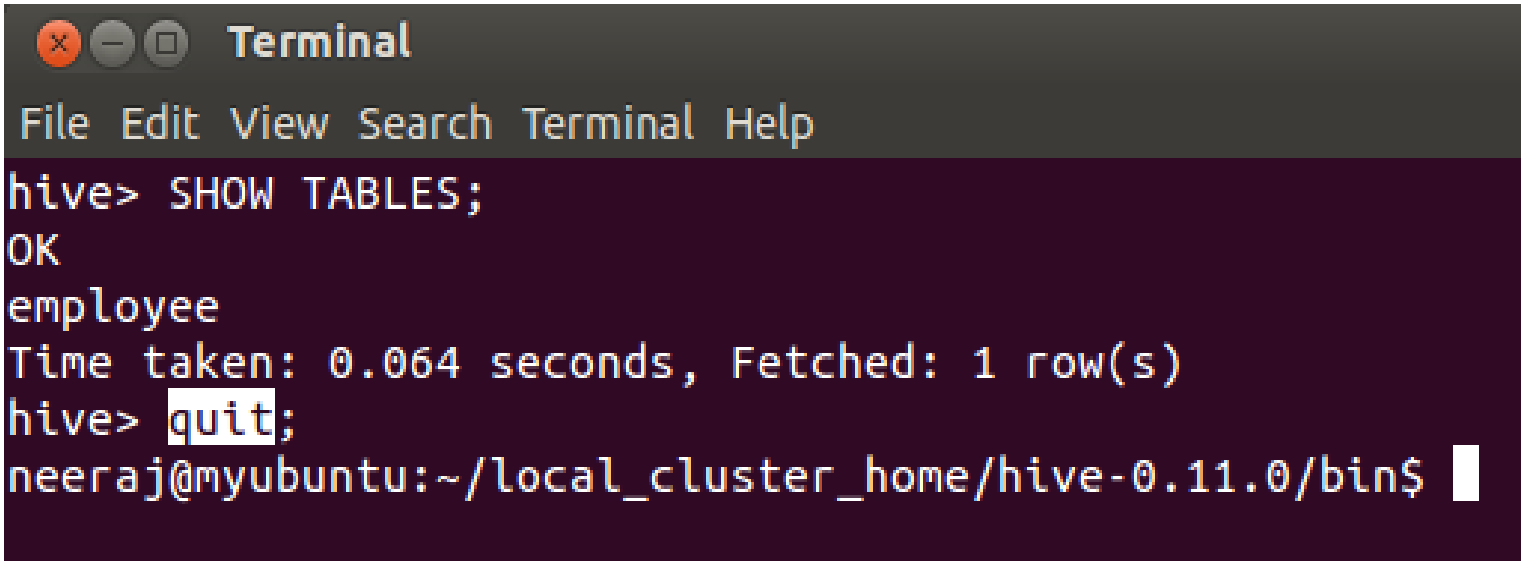
Count no. of rows in a table

```
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 SUCESS
Total MapReduce CPU Time Spent: 3 seconds 0 msec
OK
5
Time taken: 47.153 seconds
```

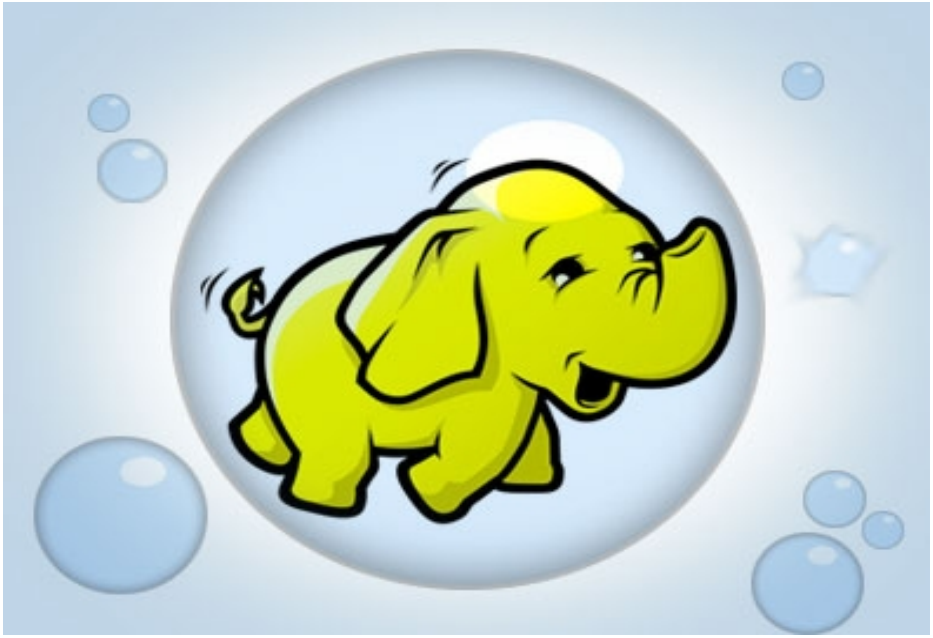
Delete table in Hive



Exit from Hive shell



...Thanks...



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