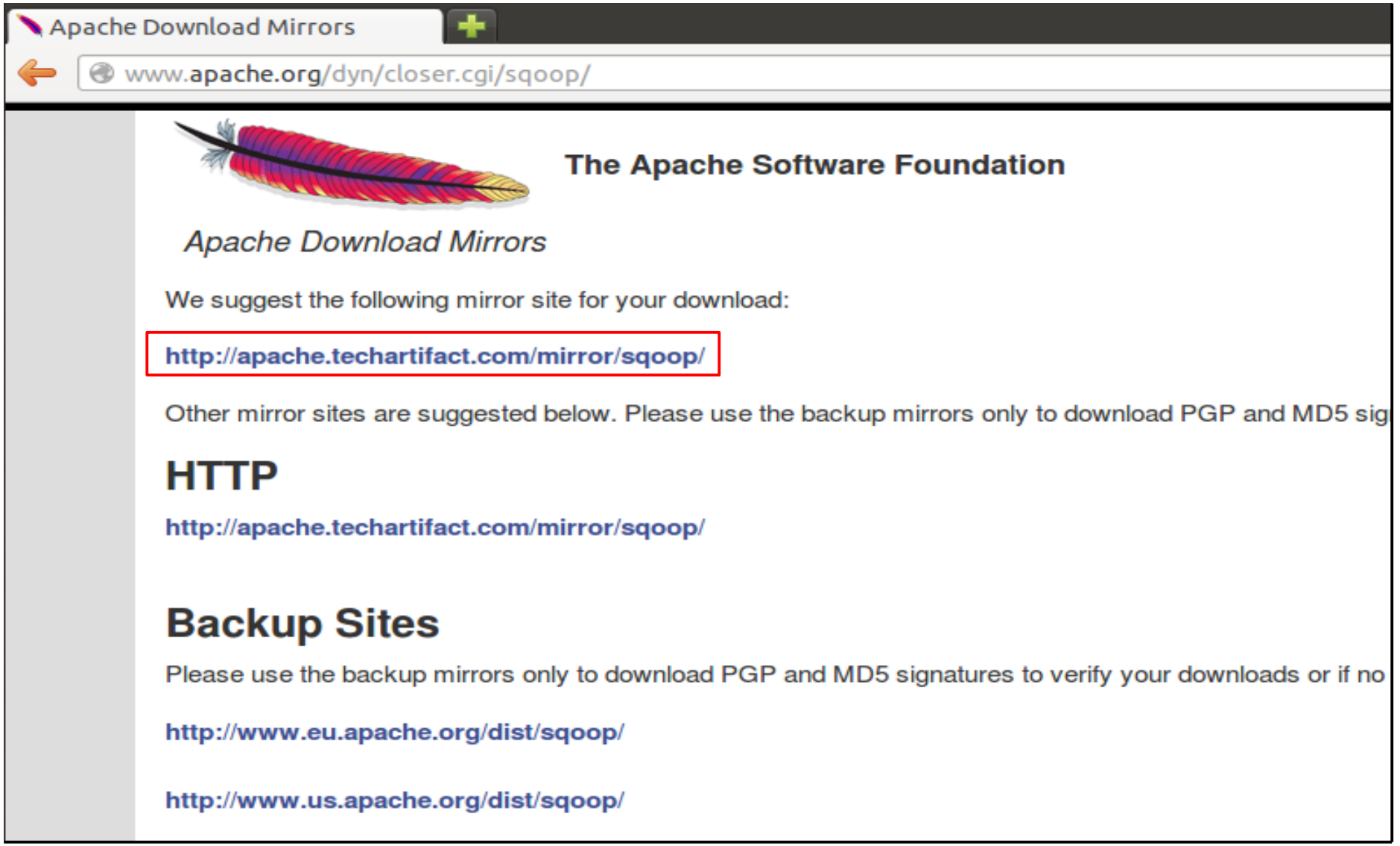
Sqoop

Agenda

Download Sqoop tar.gz file
Extract content of Sqoop tar.gz file
Configure Sqoop
Install MYSQL on your system
Create table in MYSQL.
Insert sample data into MYSQL table
Import data from RDBMS to HDFS
Export data from HDFS to RDBMS
Import data from RDBMS to HBase
Export data from HBase to RDBMS
Import data from RDBMS to Hive
Export data from Hive to RDBMS

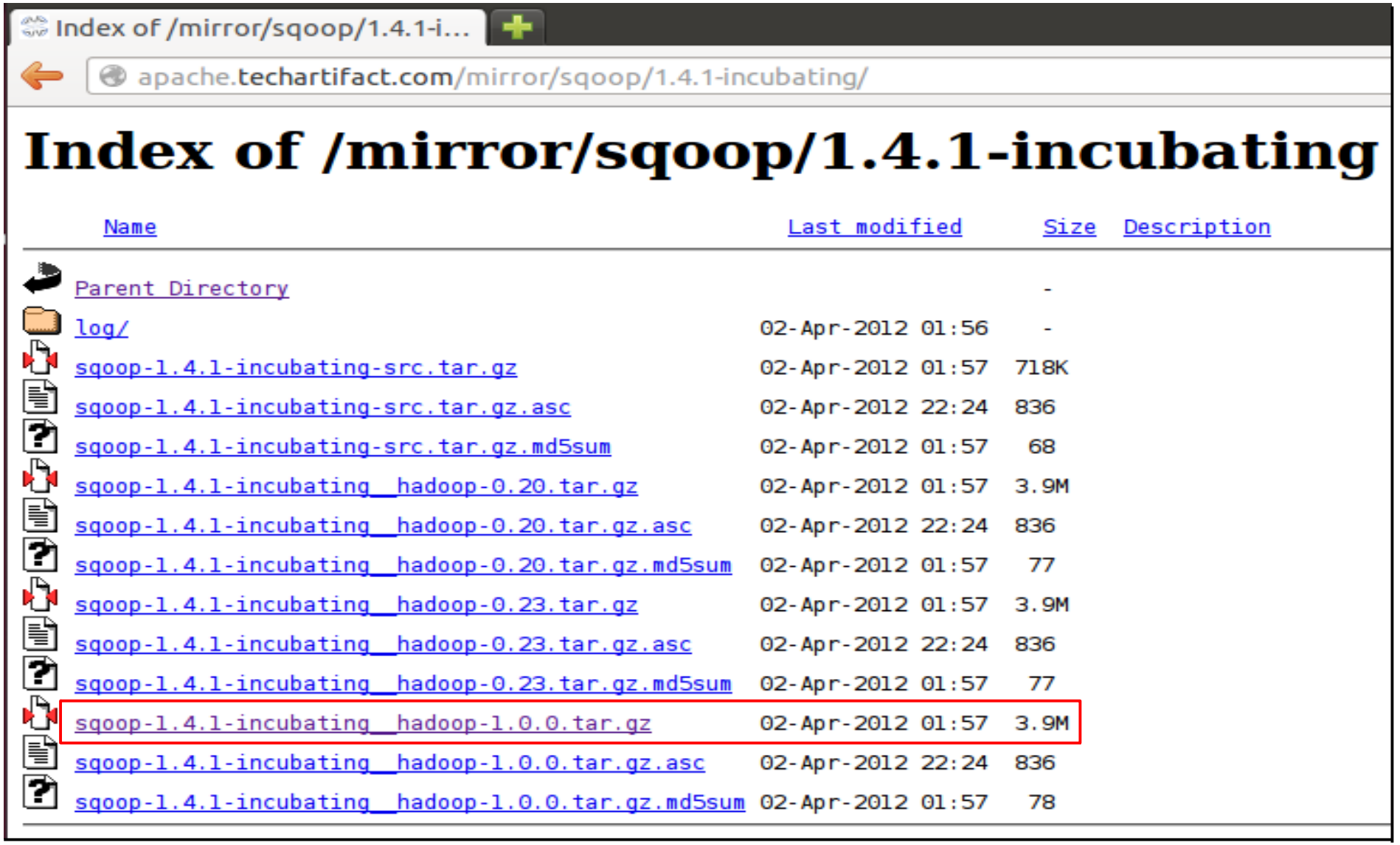
Visit Apache Sqoop website

www.apache.org/dyn/closer.cgi/sqoop/



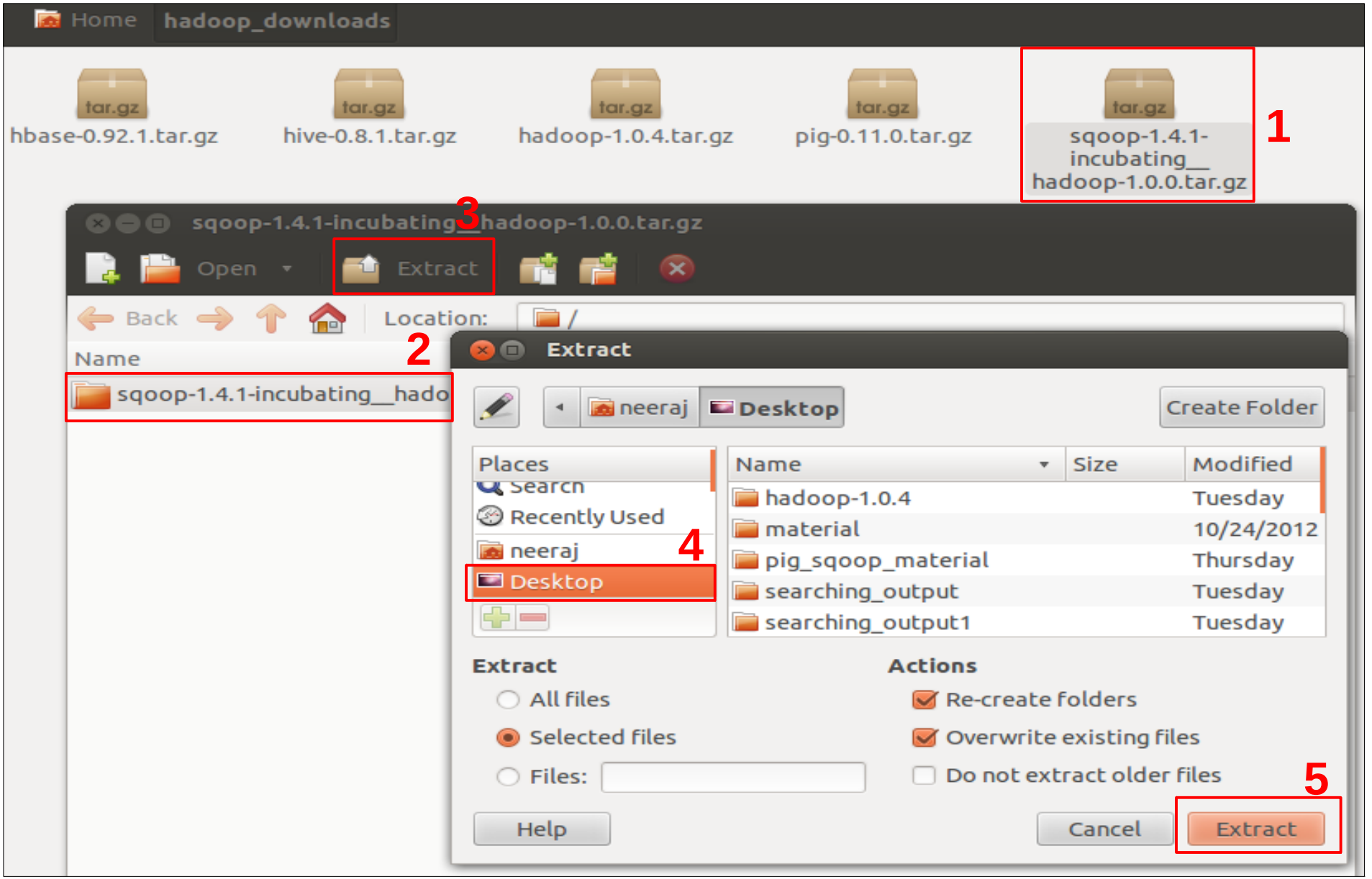
Download sqoop-1.4.1

Download sqoop tar.gz file



Save tar.gz file on your system

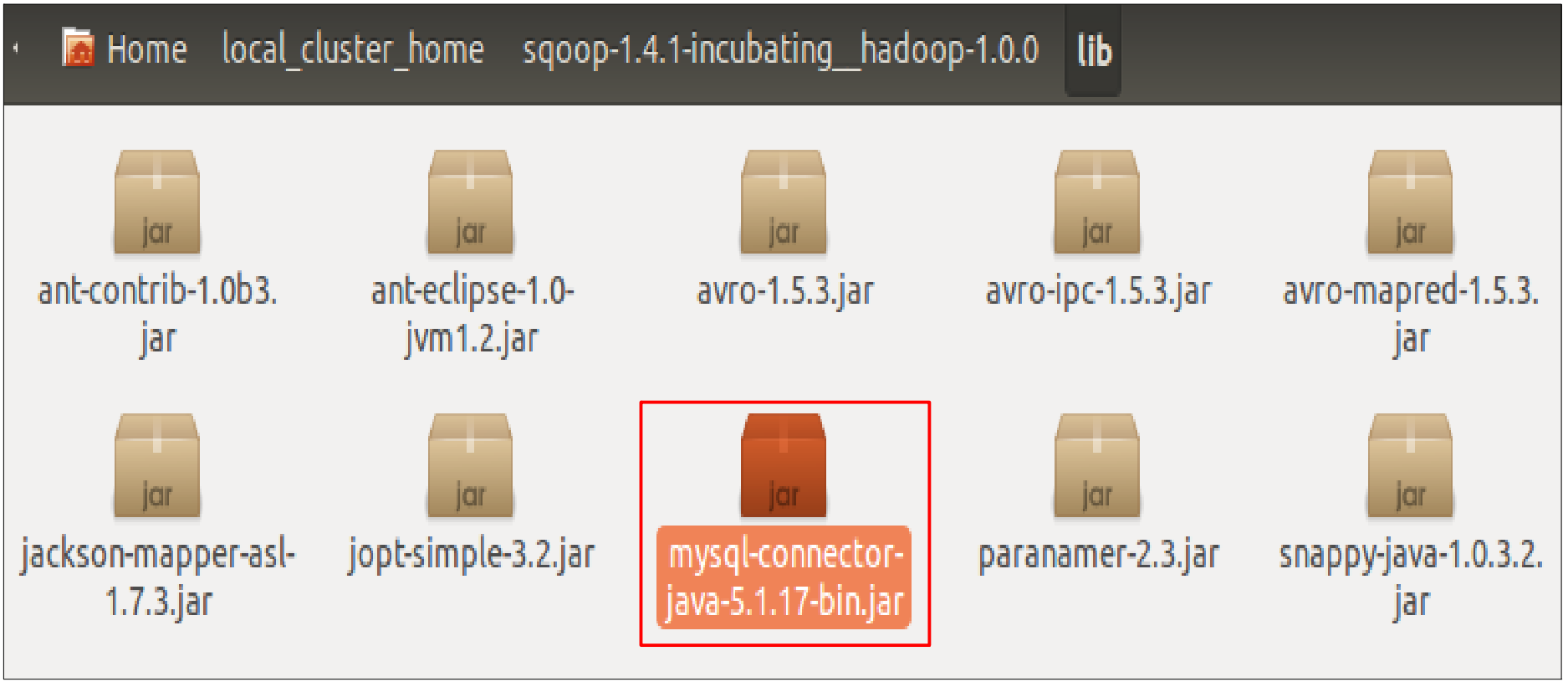
Untar the sqoop tar.gz file



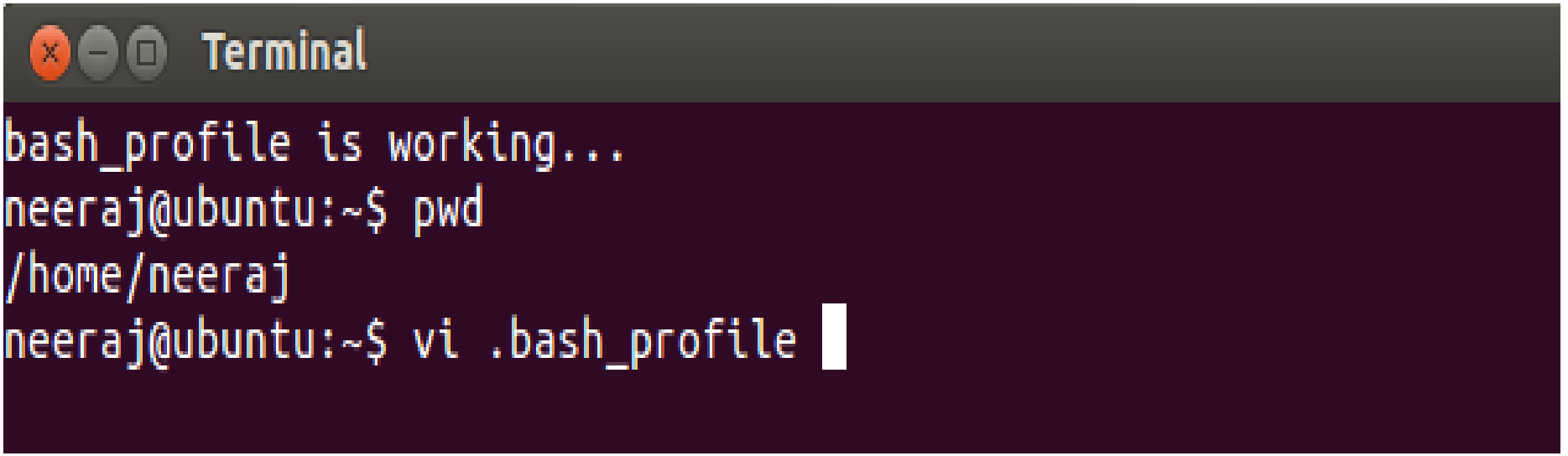
Download MySQL Connector jar

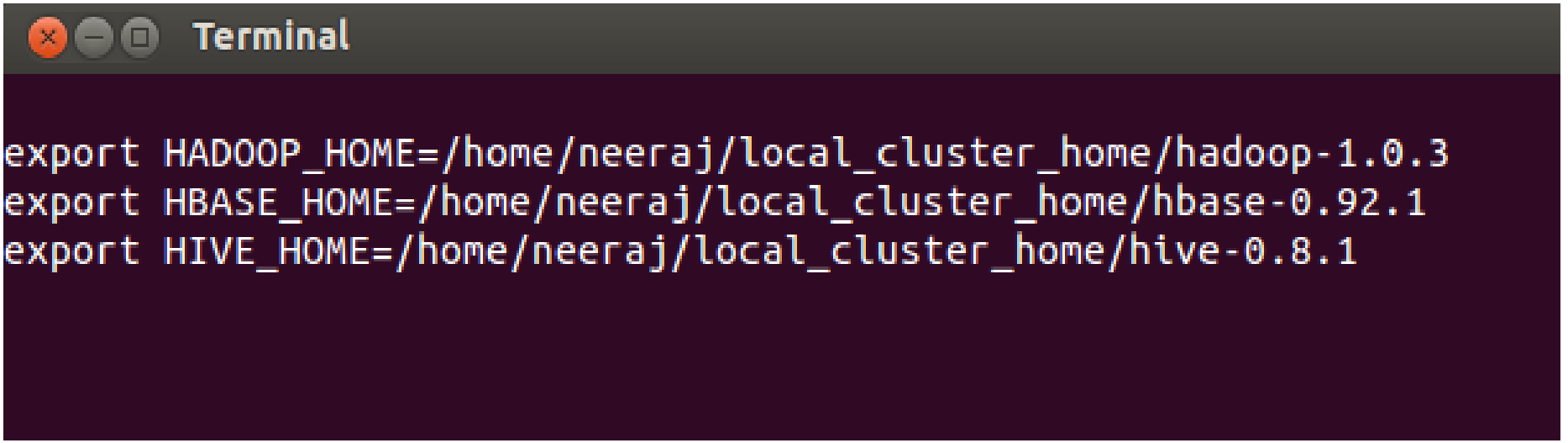
Download mysql-connector-java-5.1.17-bin.jar from internet

Copy the downloaded jar into SQOOP_HOME/lib directory

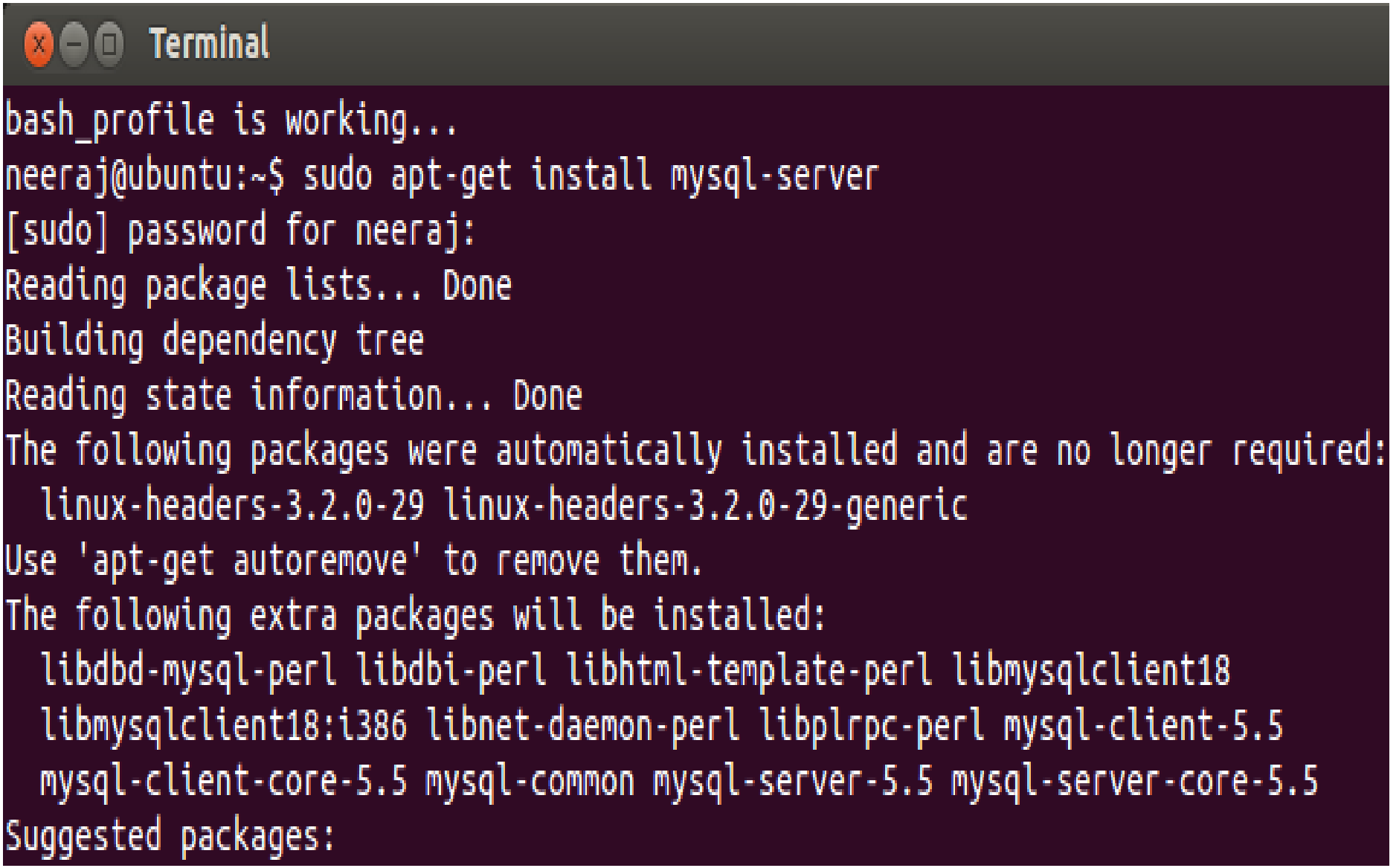


Configure .bash_profile for Sqoop

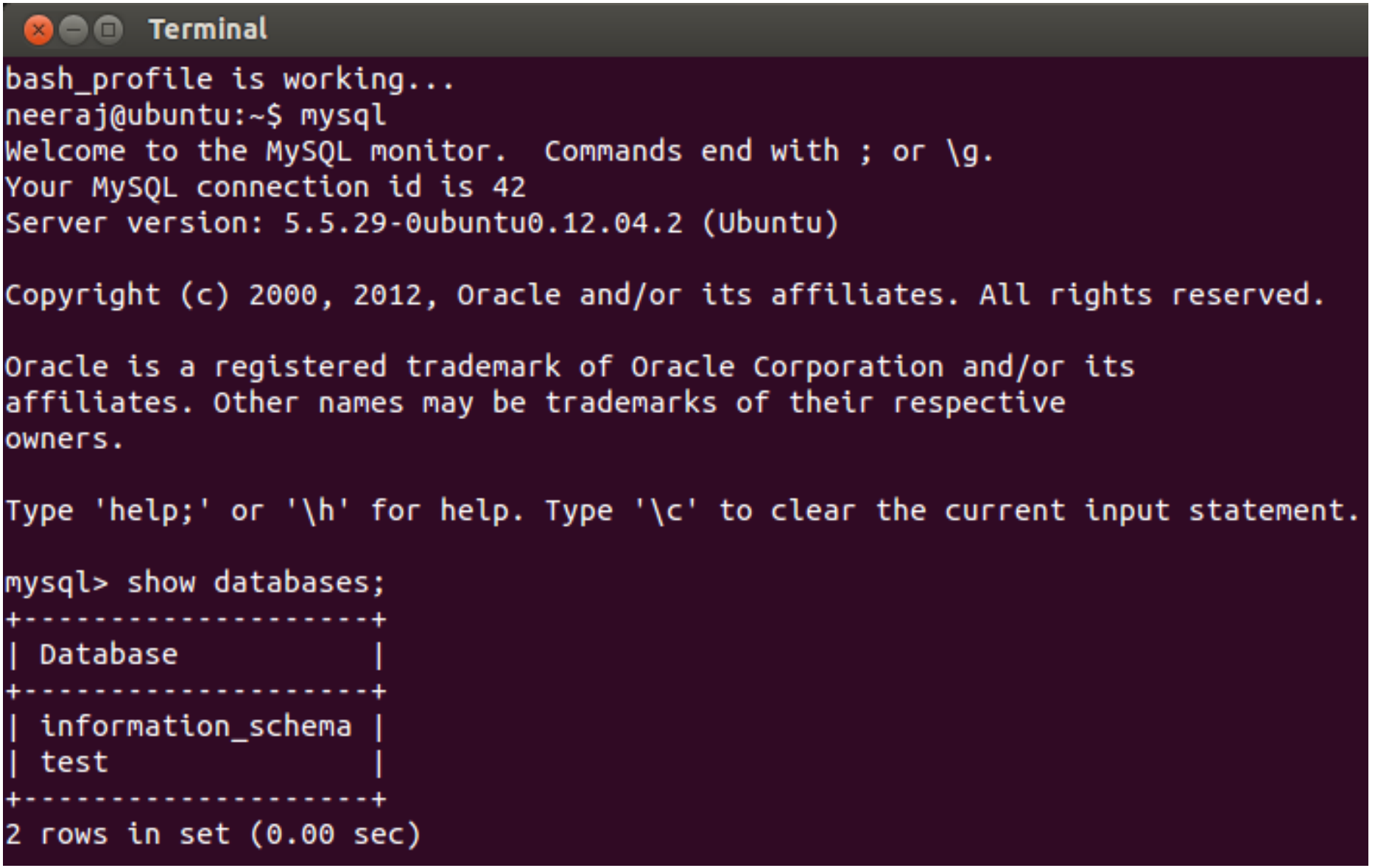




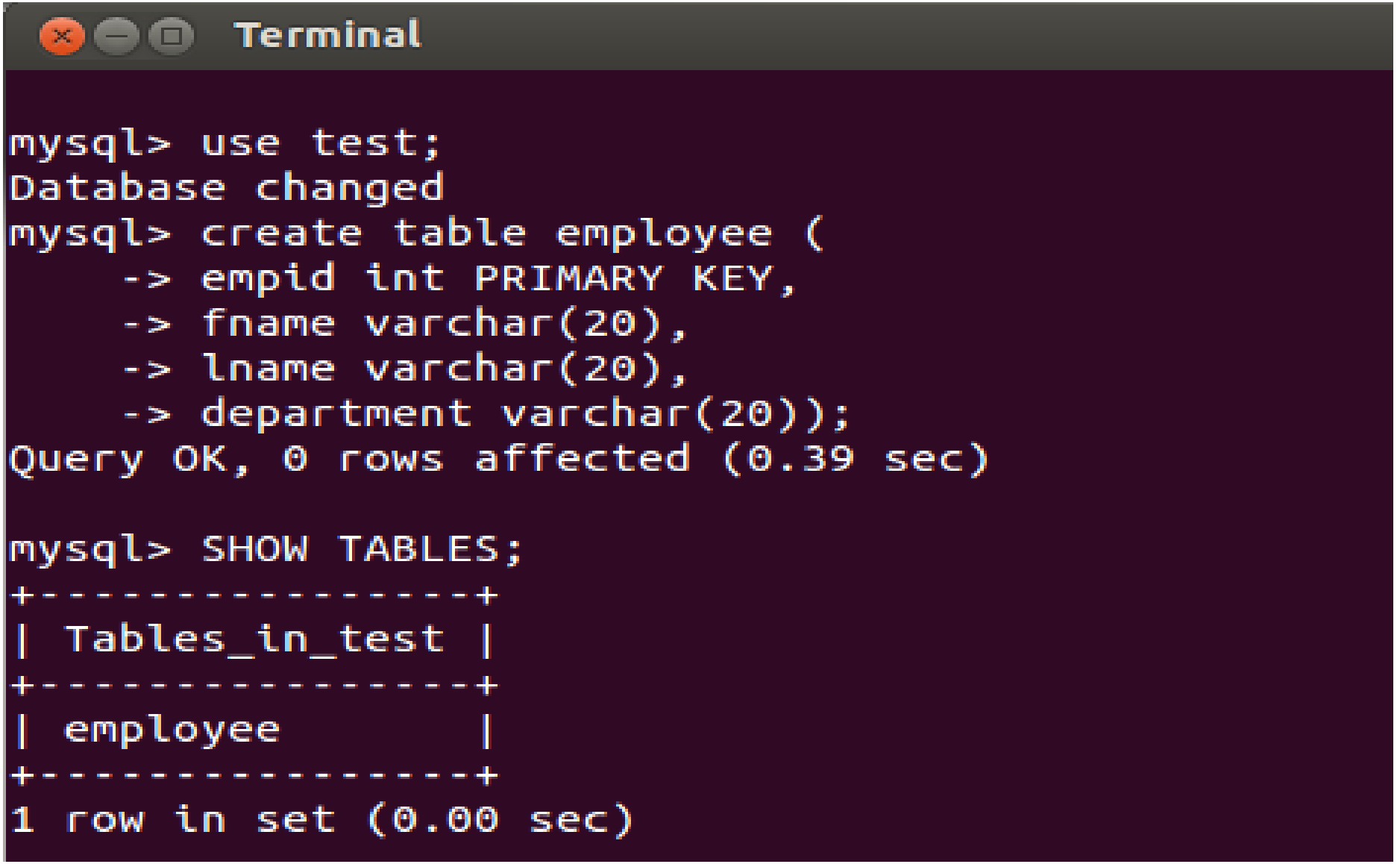
Install MYSQL on your machine



Run MYSQL on your machine



Create table in MYSQL



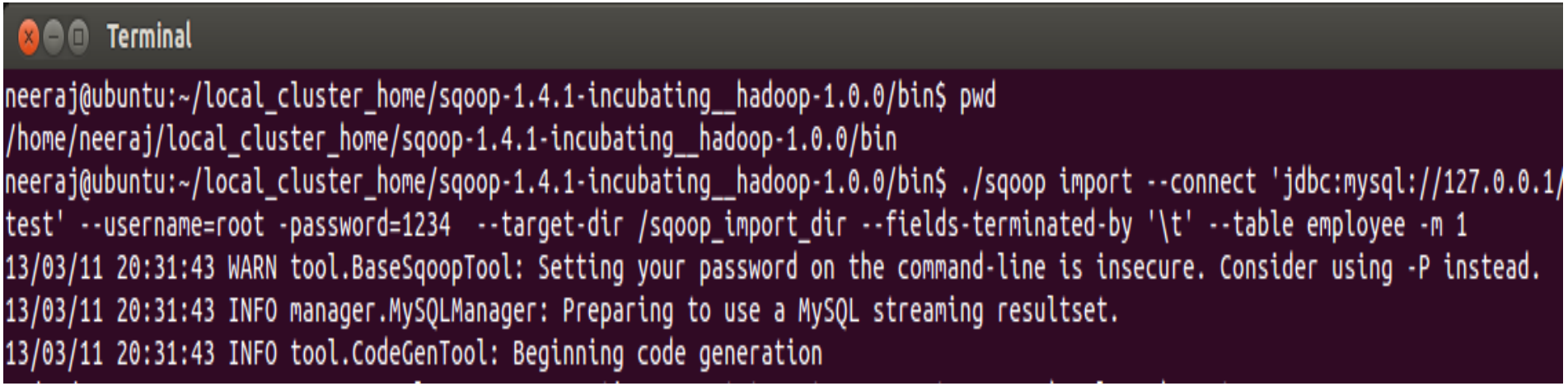
Insert data into MYSQL table

Terminal

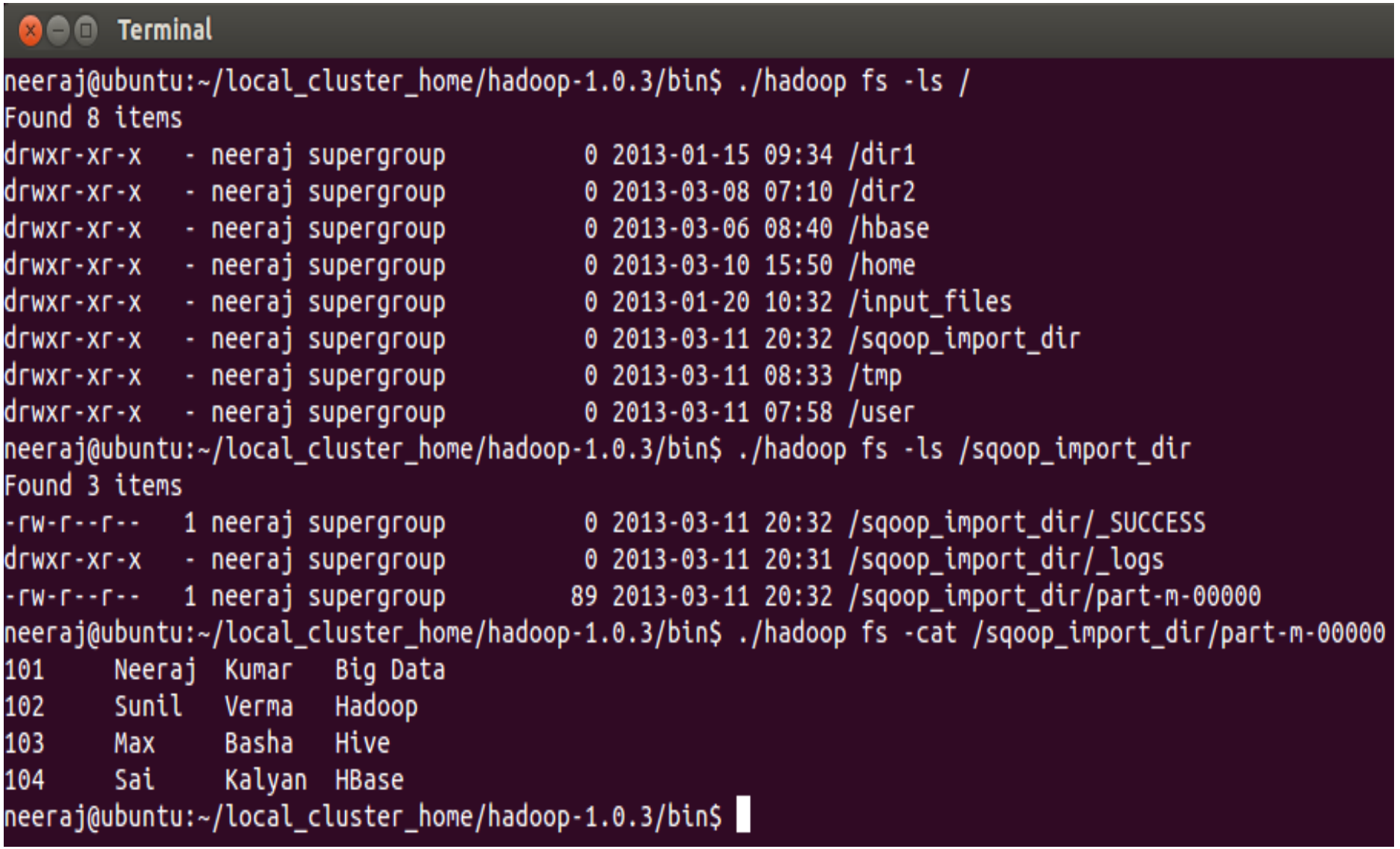
```
mysql> INSERT INTO employee VALUES (101, 'Neeraj', 'Kumar', 'Big Data');
Query OK, 1 row affected (0.20 sec)

mysql> INSERT INTO employee VALUES (102, 'Sunil', 'Verma', 'Hadoop');
^[[AQuery OK, 1 row affected (0.82 sec)

mysql> INSERT INTO employee VALUES (103, 'Max', 'Basha', 'Hive');
Query OK, 1 row affected (0.57 sec)

mysql> INSERT INTO employee VALUES (104, 'Sai', 'Kalyan', 'HBase');
Query OK, 1 row affected (0.21 sec)

mysql> SELECT * FROM employee;
+-------+--------+--------+------------+
| empid | fname  | lname  | department |
+-------+--------+--------+------------+
|   101 | Neeraj | Kumar  | Big Data   |
|   102 | Sunil  | Verma  | Hadoop     |
|   103 | Max    | Basha  | Hive       |
|   104 | Sai    | Kalyan | HBase      |
+-------+--------+--------+------------+
4 rows in set (0.00 sec)

mysql>
```

Import data from RDBMS to HDFS

./sqoop import --connect 'jdbc:mysql://127.0.0.1/test' --username=root
--password=1234 --target-dir /sqoop_import_dir
--fields-terminated-by '\t' --table employee -m 1

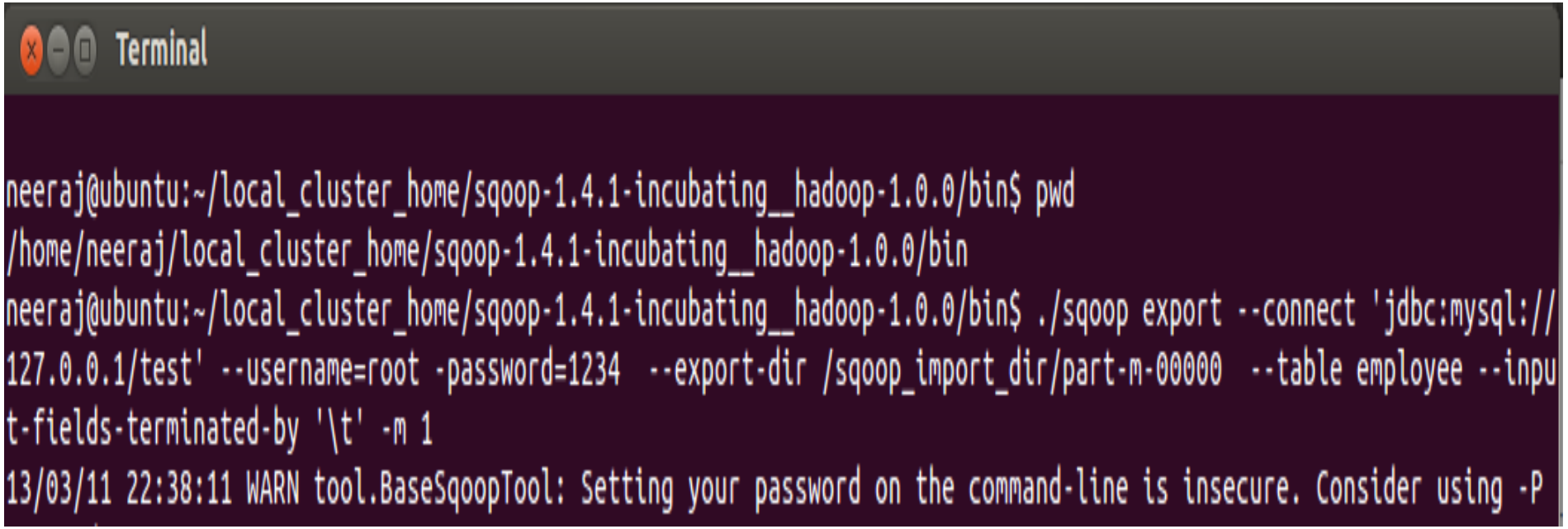


Verify imported data on HDFS

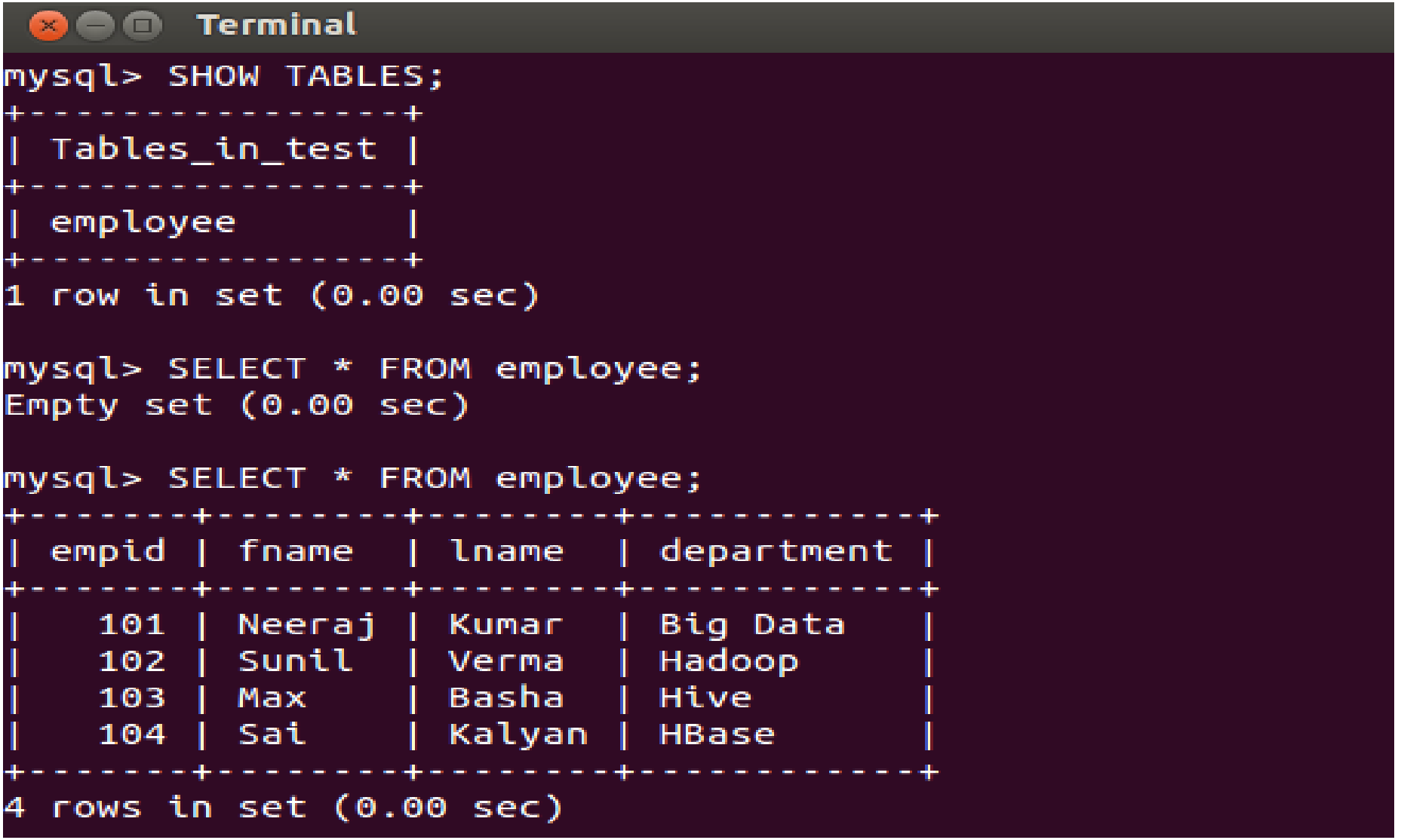


Export data from HDFS to RDBMS

./sqoop export --connect 'jdbc:mysql://127.0.0.1/test' --username=root --password=1234 --export-dir /sqoop_import_dir/part-m-00000 --table employee --input-fields-terminated-by '\t' -m 1

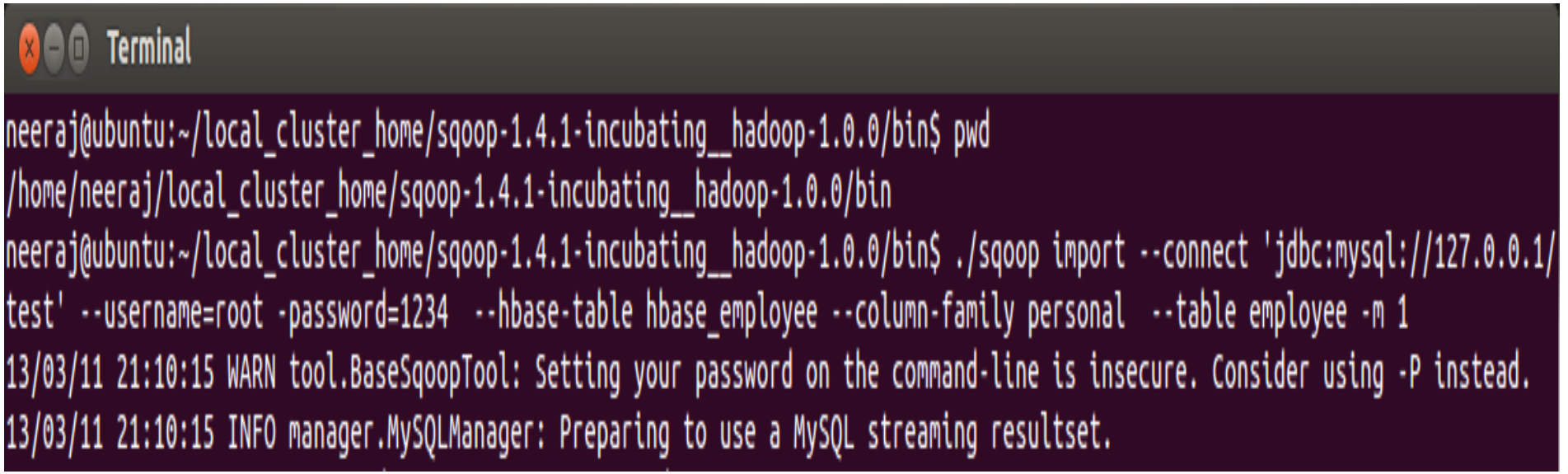


Verify exported data on RDBMS



Import data from RDBMS to HBase

./sqoop import --connect 'jdbc:mysql://127.0.0.1/test' --username=root
--password=1234 --hbase-table hbase_employee
--column-family personal --table employee -m 1



Verify imported data on HBase

Terminal

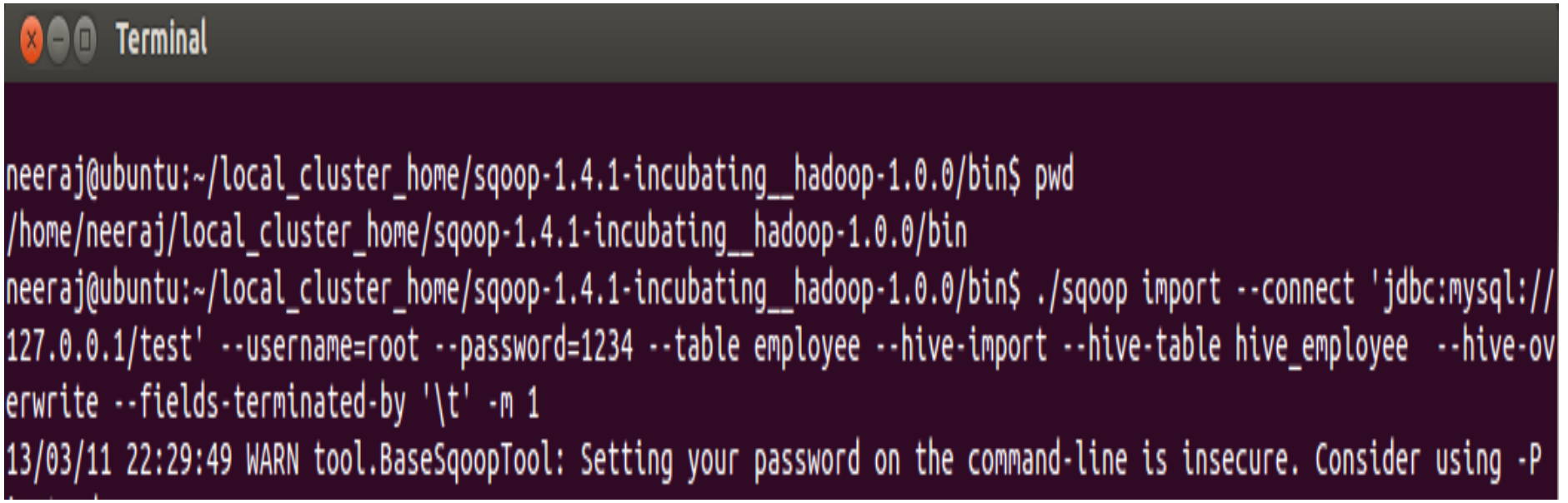
```
hbase(main):009:0> list
TABLE
hbase_employee
sample_table
2 row(s) in 0.0540 seconds

hbase(main):010:0> scan 'hbase_employee'
ROW                           COLUMN+CELL
0 row(s) in 0.2130 seconds

hbase(main):011:0> scan 'hbase_employee'
ROW                           COLUMN+CELL
 101                          column=personal:department, timestamp=1363016538719, value=Big Data
 101                          column=personal:fname, timestamp=1363016538719, value=Neeraj
 101                          column=personal:lname, timestamp=1363016538719, value=Kumar
 102                          column=personal:department, timestamp=1363016538719, value=Hadoop
 102                          column=personal:fname, timestamp=1363016538719, value=Sunil
 102                          column=personal:lname, timestamp=1363016538719, value=Verma
 103                          column=personal:department, timestamp=1363016538719, value=Hive
 103                          column=personal:fname, timestamp=1363016538719, value=Max
 103                          column=personal:lname, timestamp=1363016538719, value=Basha
 104                          column=personal:department, timestamp=1363016538719, value=HBase
 104                          column=personal:fname, timestamp=1363016538719, value=Sai
 104                          column=personal:lname, timestamp=1363016538719, value=Kalyan
4 row(s) in 0.2440 seconds

hbase(main):012:0>
```

Export data from HBase to RDBMS

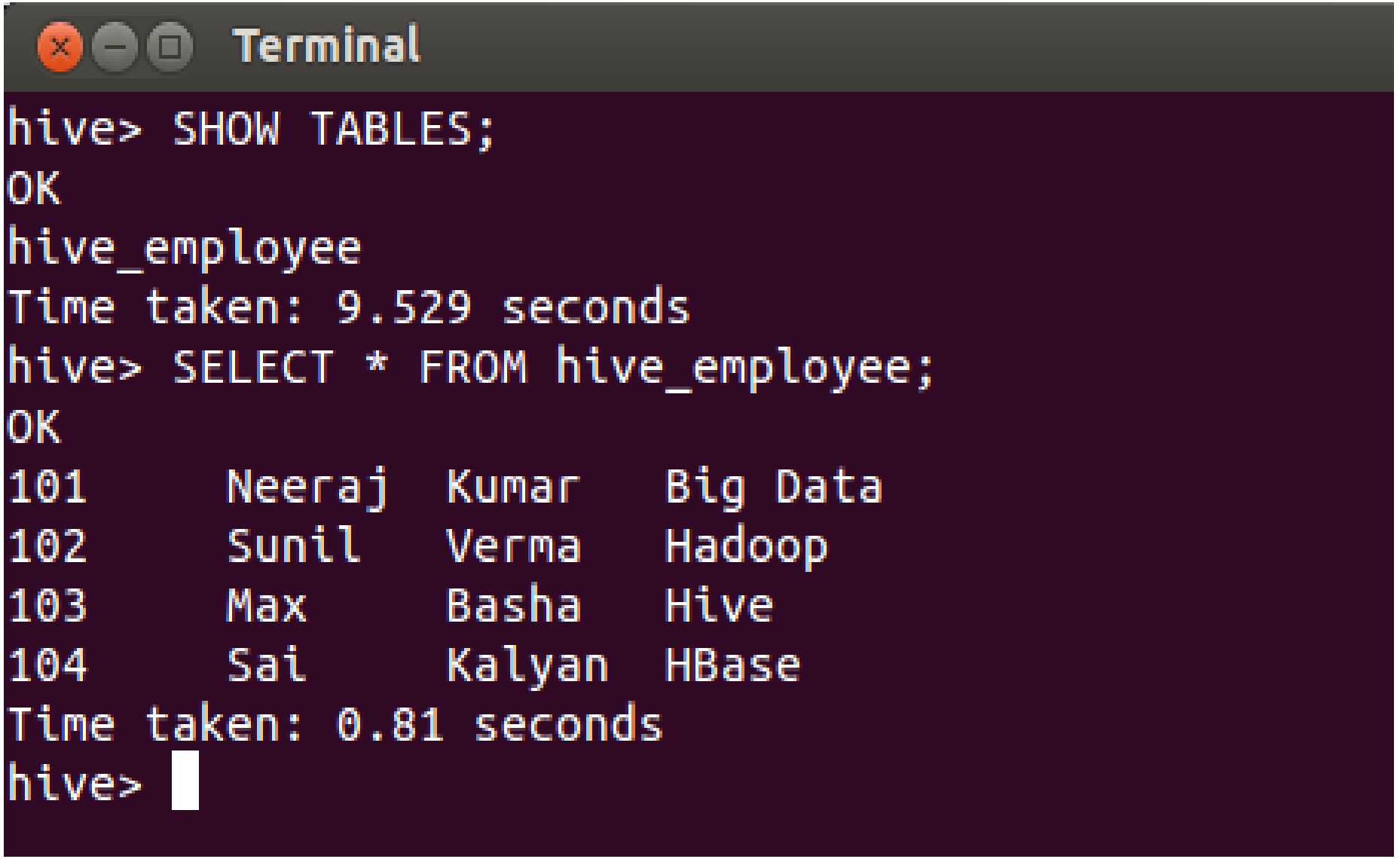
Sqoop doesn't support exporting of data from HBase to RDBMS.

Import data from RDBMS to Hive

./sqoop import --connect 'jdbc:mysql://127.0.0.1/test' --username=root
--password=1234 --table employee --hive-import
--hive-table hive_employee --hive-overwrite --fields-terminated-by '\t' -m 1

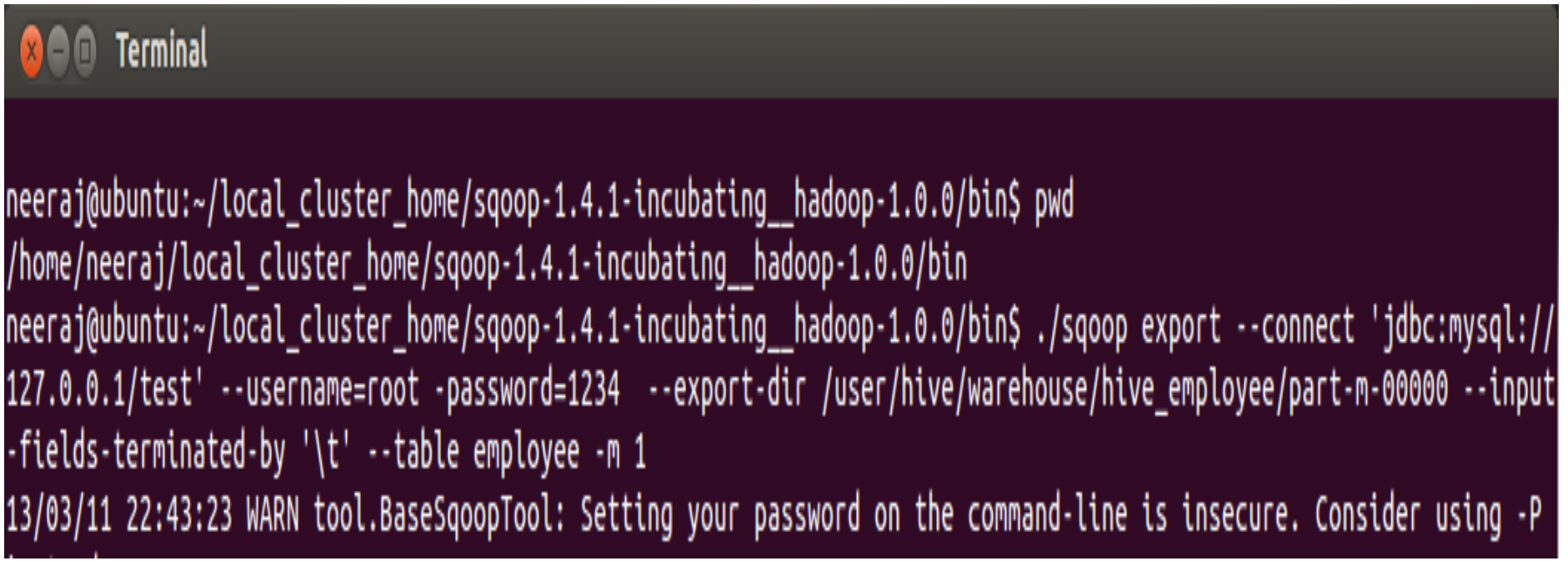


Verify imported data on Hive



Export data from Hive to RDBMS

./sqoop export --connect 'jdbc:mysql://127.0.0.1/test' --username=root --password=1234 --export-dir /user/hive/warehouse/hive_employee/input_data.txt --input-fields-terminated-by '\t' --table employee -m 1

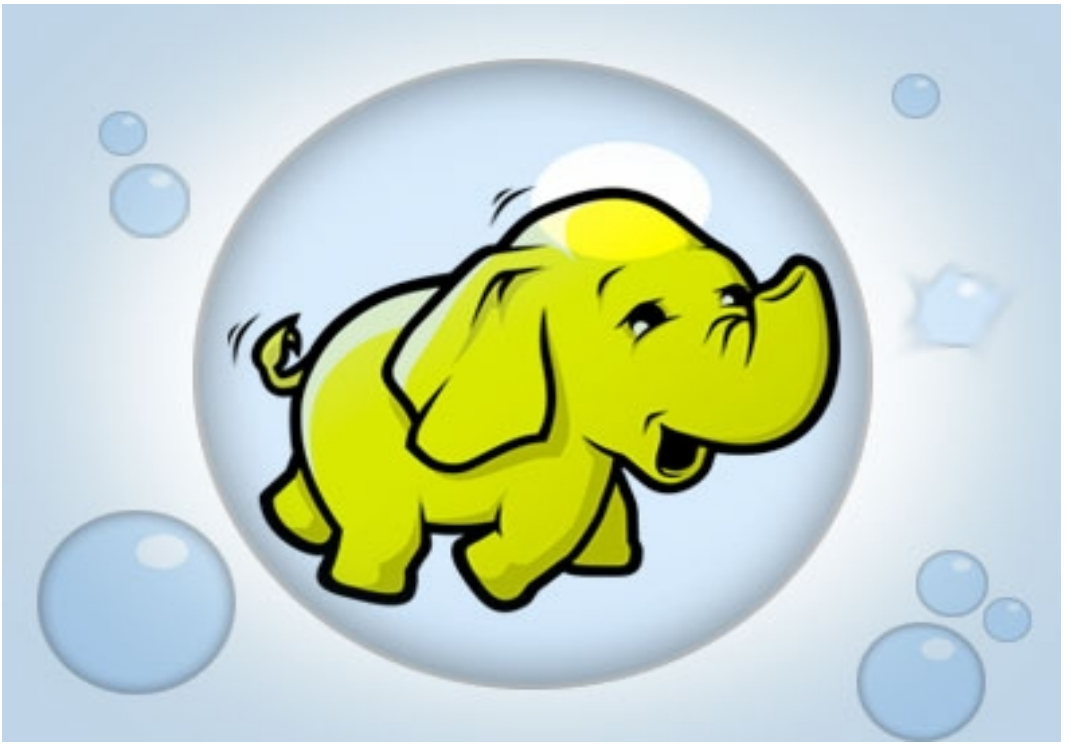


Verify exported data on RDBMS

Terminal

```
mysql> SHOW TABLES;
+----------------+
| Tables_in_test |
+----------------+
| employee       |
+----------------+
1 row in set (0.00 sec)

mysql> SELECT * FROM employee;
Empty set (0.00 sec)

mysql> SELECT * FROM employee;
+-------+--------+--------+------------+
| empid | fname  | lname  | department |
+-------+--------+--------+------------+
|   101 | Neeraj | Kumar  | Big Data   |
|   102 | Sunil  | Verma  | Hadoop     |
|   103 | Max    | Basha  | Hive       |
|   104 | Sai    | Kalyan | HBase      |
+-------+--------+--------+------------+
4 rows in set (0.00 sec)
```

...Thanks...



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