

Sqoop



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

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

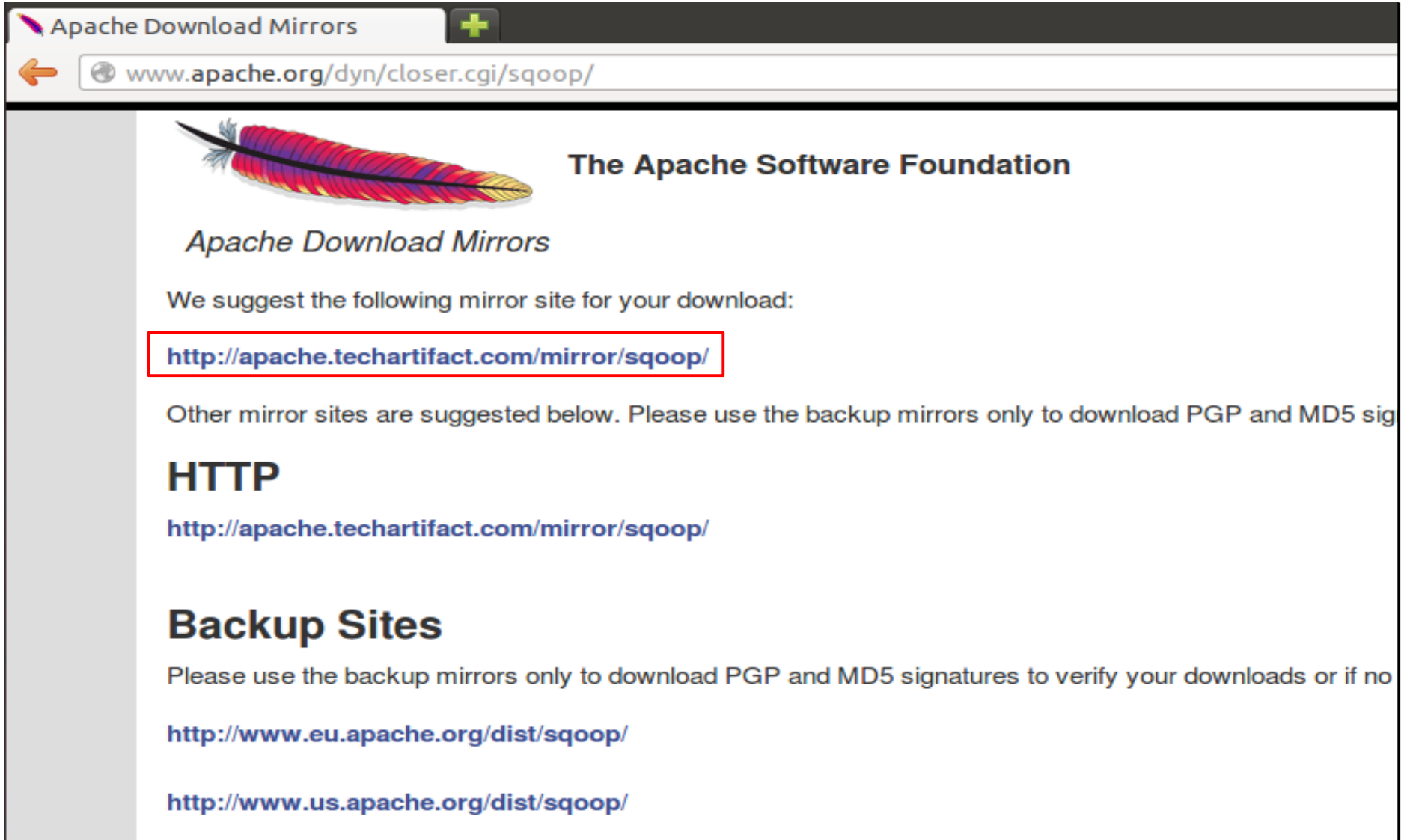
Export data from HBase to RDBMS

Import data from RDBMS to Hive

Export data from Hive to RDBMS

Visit Apache Sqoop website

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



The screenshot shows a web browser window with the title "Apache Download Mirrors" and a plus sign icon. The address bar displays the URL www.apache.org/dyn/closer.cgi/sqoop/. The page content features the Apache Software Foundation logo (a feather) and the text "The Apache Software Foundation". Below this, it says "Apache Download Mirrors" and "We suggest the following mirror site for your download:". A red box highlights the URL <http://apache.techartifact.com/mirror/sqoop/>. Further down, it states "Other mirror sites are suggested below. Please use the backup mirrors only to download PGP and MD5 sig" and "HTTP". Below this, it lists "Backup Sites" and provides two more URLs: <http://www.eu.apache.org/dist/sqoop/> and <http://www.us.apache.org/dist/sqoop/>.

Apache Download Mirrors

The Apache Software Foundation

Apache Download Mirrors

We suggest the following mirror site for your download:

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

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

HTTP

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

Backup Sites

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

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



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

Download sqoop-1.4.1

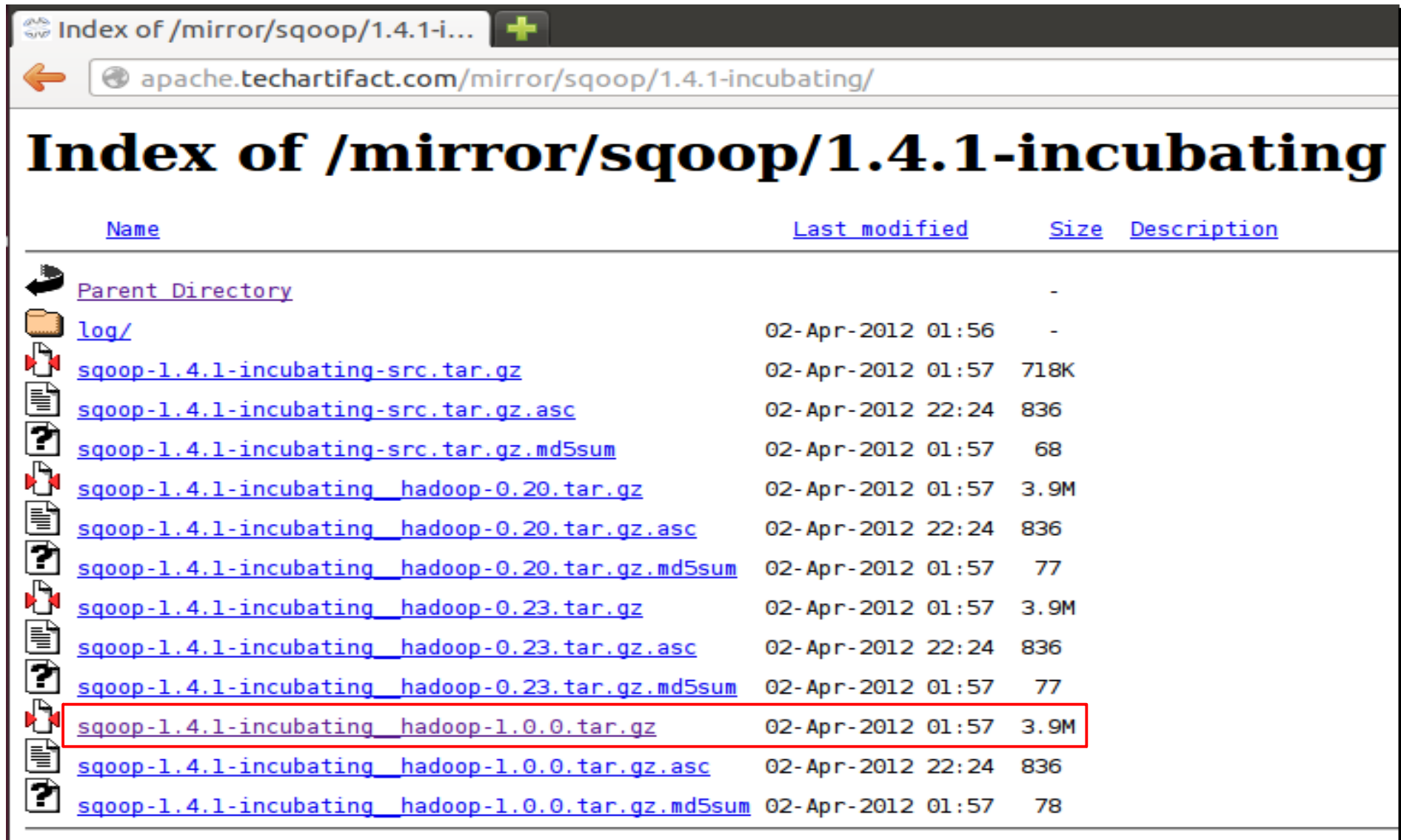
Index of /mirror/sqoop


← apache.techartifact.com/mirror/sqoop/


Index of /mirror/sqoop

	Name	Last modified	Size	Description
	Parent Directory		-	
	1.4.0-incubating/	02-Apr-2012 23:39	-	
	1.4.1-incubating/	02-Apr-2012 22:24	-	
	1.4.2/	23-Aug-2012 01:19	-	
	1.4.3/	05-Mar-2013 08:49	-	
	1.99.1/	24-Dec-2012 20:46	-	
	.revision	05-Mar-2013 08:49	5	
	KEYS	26-Feb-2013 08:14	25K	















Download sqoop tar.gz file



Index of /mirror/sqoop/1.4.1-i... 

←  apache.techartifact.com/mirror/sqoop/1.4.1-incubating/

Index of /mirror/sqoop/1.4.1-incubating

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
 Parent Directory		-	
 log/	02-Apr-2012 01:56	-	
 sqoop-1.4.1-incubating-src.tar.gz	02-Apr-2012 01:57	718K	
 sqoop-1.4.1-incubating-src.tar.gz.asc	02-Apr-2012 22:24	836	
 sqoop-1.4.1-incubating-src.tar.gz.md5sum	02-Apr-2012 01:57	68	
 sqoop-1.4.1-incubating_hadoop-0.20.tar.gz	02-Apr-2012 01:57	3.9M	
 sqoop-1.4.1-incubating_hadoop-0.20.tar.gz.asc	02-Apr-2012 22:24	836	
 sqoop-1.4.1-incubating_hadoop-0.20.tar.gz.md5sum	02-Apr-2012 01:57	77	
 sqoop-1.4.1-incubating_hadoop-0.23.tar.gz	02-Apr-2012 01:57	3.9M	
 sqoop-1.4.1-incubating_hadoop-0.23.tar.gz.asc	02-Apr-2012 22:24	836	
 sqoop-1.4.1-incubating_hadoop-0.23.tar.gz.md5sum	02-Apr-2012 01:57	77	
 sqoop-1.4.1-incubating_hadoop-1.0.0.tar.gz	02-Apr-2012 01:57	3.9M	
 sqoop-1.4.1-incubating_hadoop-1.0.0.tar.gz.asc	02-Apr-2012 22:24	836	
 sqoop-1.4.1-incubating_hadoop-1.0.0.tar.gz.md5sum	02-Apr-2012 01:57	78	

Save tar.gz file on your system

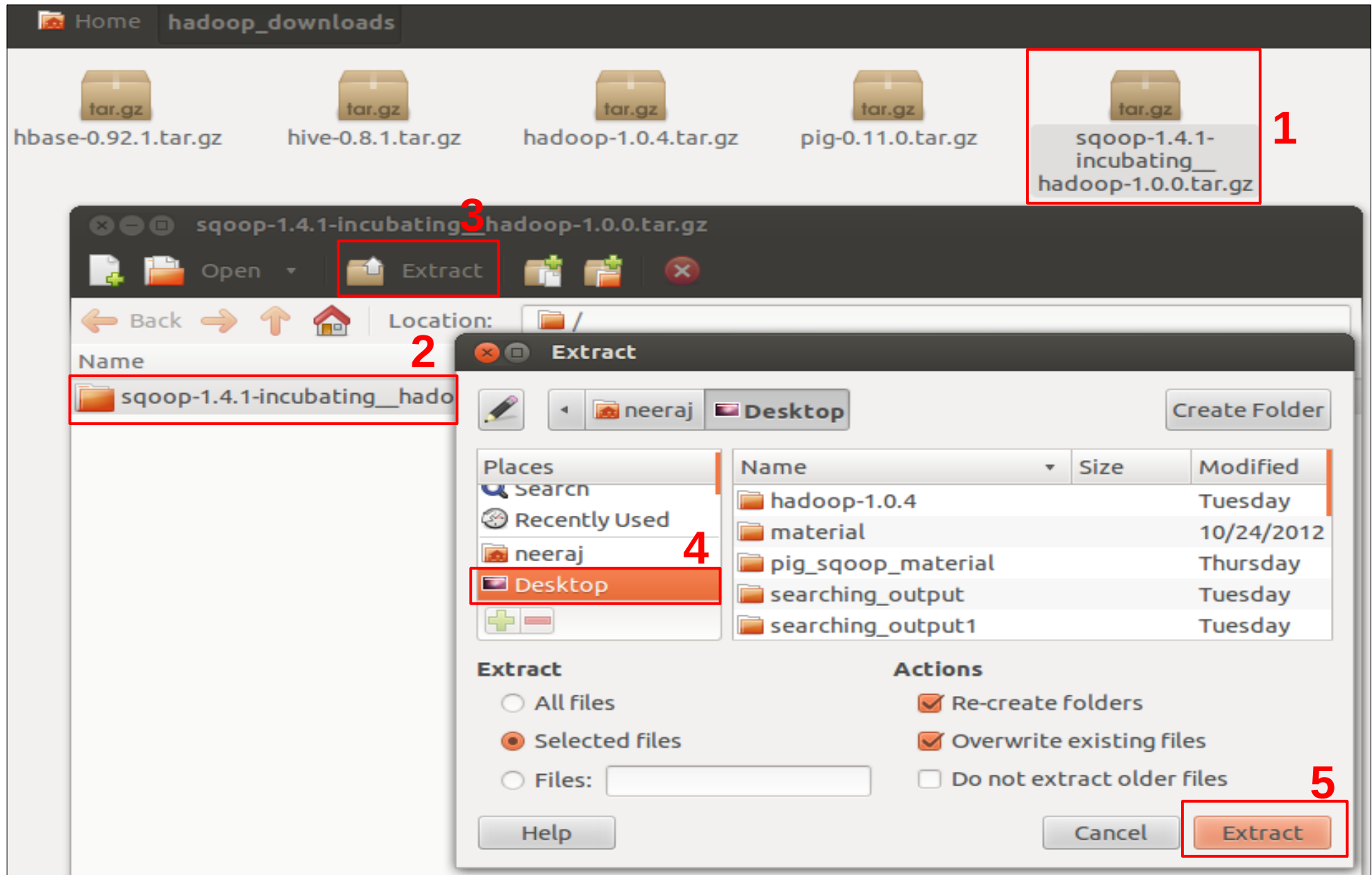
The screenshot shows a web browser window displaying a directory index for `/mirror/sqoop/1.4.1-incubating` on the website `apache.techartifact.com`. The browser's address bar shows the URL `apache.techartifact.com/mirror/sqoop/1.4.1-incubating/`. The page title is **Index of /mirror/sqoop/1.4.1-incubating**. Below the title, there is a table with columns: Name, Last modified, Size, and Description.

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
Parent Directory	-	-	
log/	02-Apr-2012 01:56	-	
sqoop-1.4.1-incubating-src.tar.gz			
sqoop-1.4.1-incubating-src.tar.gz.asc			
sqoop-1.4.1-incubating-src.tar.gz.md5sum			
sqoop-1.4.1-incubating_hadoop-0.20.tar.gz			
sqoop-1.4.1-incubating_hadoop-0.20.tar.gz.asc			
sqoop-1.4.1-incubating_hadoop-0.20.tar.gz.md5sum			
sqoop-1.4.1-incubating_hadoop-0.23.tar.gz			
sqoop-1.4.1-incubating_hadoop-0.23.tar.gz.asc			
sqoop-1.4.1-incubating_hadoop-0.23.tar.gz.md5sum			
sqoop-1.4.1-incubating_hadoop-1.0.0.tar.gz			
sqoop-1.4.1-incubating_hadoop-1.0.0.tar.gz.asc			
sqoop-1.4.1-incubating_hadoop-1.0.0.tar.gz.md5sum			

At the bottom of the browser window, the following text is visible: `Apache/2.2.23 (Unix) mod_ssl/2.2.23 OpenSSL/1.0.1 mod_qos/10.10 mod_perl/2.0.6 Perl/v5.10.1 Server at apache.techartifact.com Port 80`.

Overlaid on the right side of the browser window is a Firefox dialog box titled **Opening sqoop-1.4.1-incubating_hadoop-1.0.0.tar.gz**. The dialog box contains the following text: "You have chosen to open: **sqoop-1.4.1-incubating_hadoop-1.0.0.tar.gz** which is a: Gzip archive (3.9 MB) from: `http://apache.techartifact.com`". Below this, it asks "What should Firefox do with this file?". There are three options: ☐ Open with Archive Manager (default), ☒ Save File, and ☐ Do this automatically for files like this from now on. At the bottom right of the dialog box, there are two buttons: "Cancel" and "OK". The "OK" button is highlighted with a red rectangle, and a mouse cursor is pointing at it.

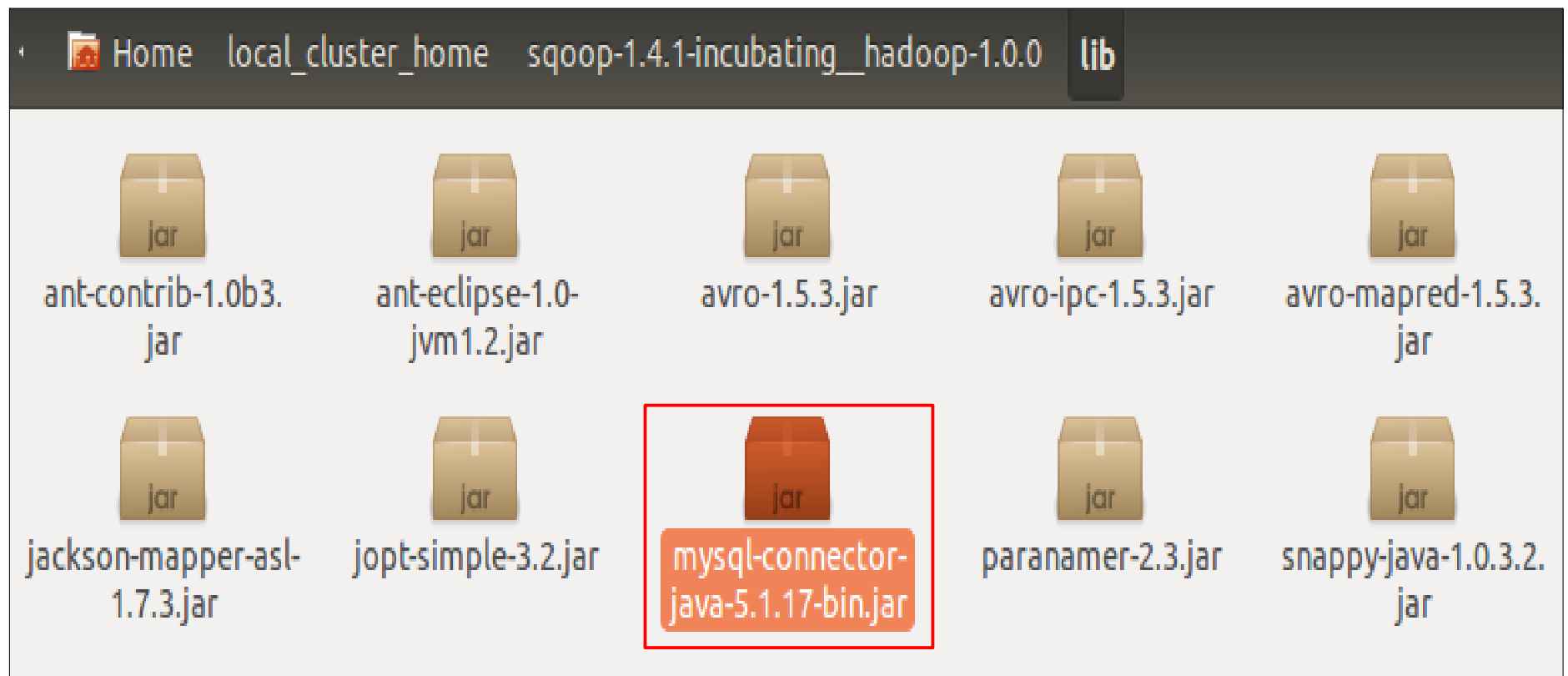
Untar the sqoop tar.gz file



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

```
Terminal
bash_profile is working...
neeraj@ubuntu:~$ pwd
/home/neeraj
neeraj@ubuntu:~$ vi .bash_profile
```

```
Terminal
export HADOOP_HOME=/home/neeraj/local_cluster_home/hadoop-1.0.3
export HBASE_HOME=/home/neeraj/local_cluster_home/hbase-0.92.1
export HIVE_HOME=/home/neeraj/local_cluster_home/hive-0.8.1
```

Install MYSQL on your machine

```
Terminal
bash_profile is working...
neeraj@ubuntu:~$ sudo apt-get install mysql-server
[sudo] password for neeraj:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  linux-headers-3.2.0-29 linux-headers-3.2.0-29-generic
Use 'apt-get autoremove' to remove them.
The following extra packages will be installed:
  libdbd-mysql-perl libdbi-perl libhtml-template-perl libmysqlclient18
  libmysqlclient18:i386 libnet-daemon-perl libplrpc-perl mysql-client-5.5
  mysql-client-core-5.5 mysql-common mysql-server-5.5 mysql-server-core-5.5
Suggested packages:
```

Run MYSQL on your machine

```
Terminal
bash_profile is working...
neeraj@ubuntu:~$ mysql
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 42
Server version: 5.5.29-0ubuntu0.12.04.2 (Ubuntu)

Copyright (c) 2000, 2012, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> show databases;
+-----+
| Database                |
+-----+
| information_schema       |
| test                    |
+-----+
2 rows in set (0.00 sec)
```

Create table in MYSQL

```
Terminal

mysql> use test;
Database changed
mysql> create table employee (
    -> empid int PRIMARY KEY,
    -> fname varchar(20),
    -> lname varchar(20),
    -> department varchar(20));
Query OK, 0 rows affected (0.39 sec)

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

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)

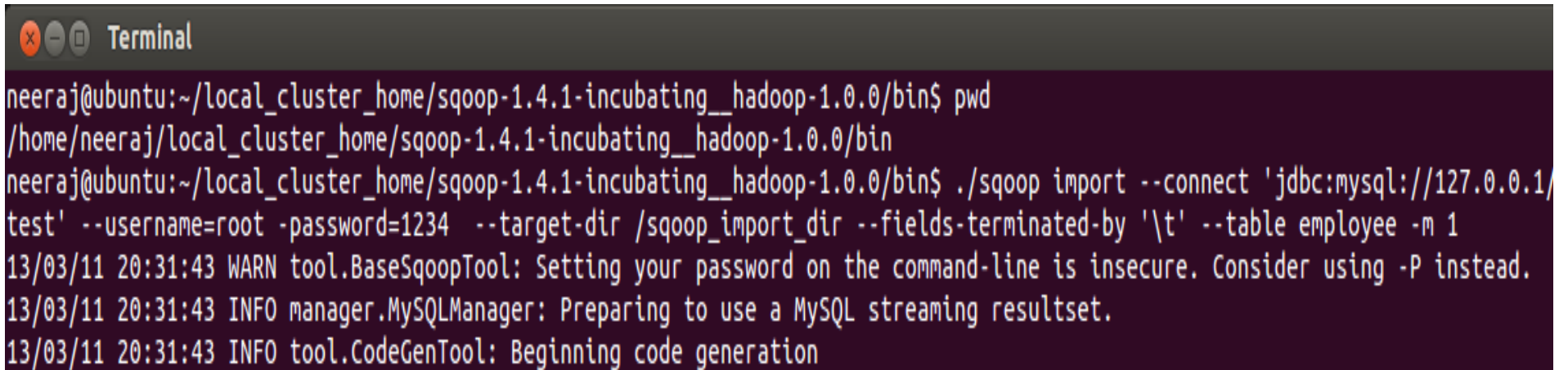
mysql> INSERT INTO employee VALUES (104, 'Sai', 'Kalyan', 'HBase');
Query OK, 1 row affected (0.21 sec)

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
```



A terminal window titled "Terminal" with a dark background and light-colored text. The window shows the execution of the sqoop import command. The user is in the directory ~/local_cluster_home/sqoop-1.4.1-incubating_hadoop-1.0.0/bin. The command is ./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. The output shows a warning about password security, information about the MySQL streaming resultset, and information about code generation.

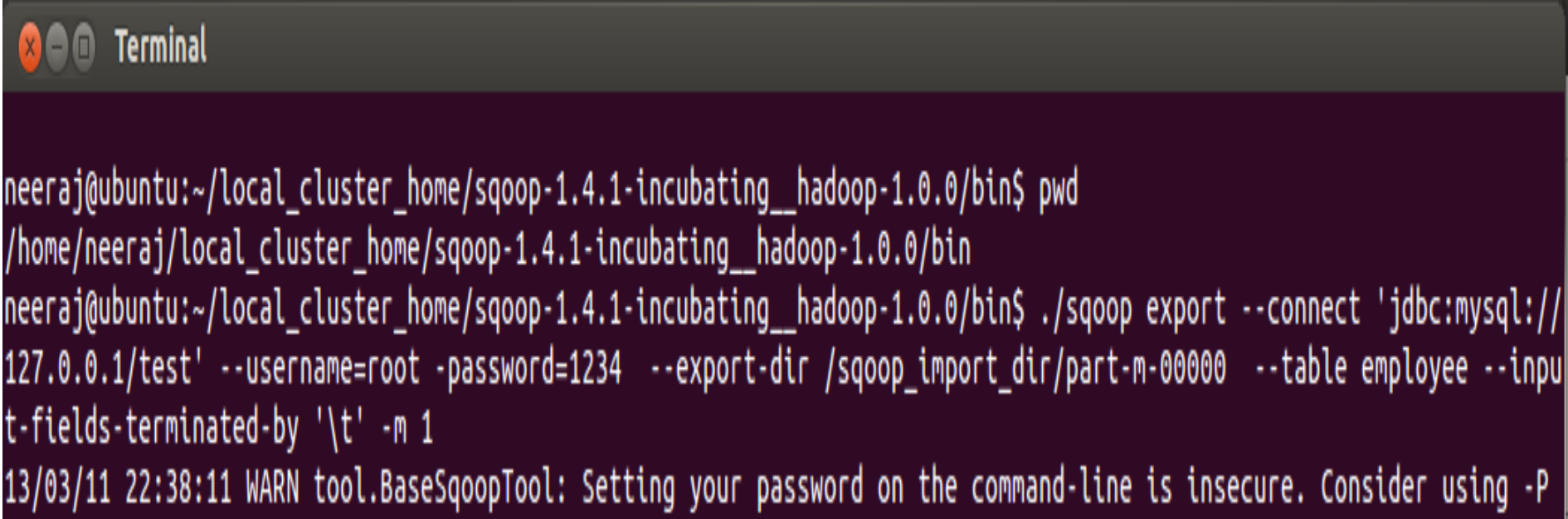
```
neeraj@ubuntu:~/local_cluster_home/sqoop-1.4.1-incubating_hadoop-1.0.0/bin$ pwd  
/home/neeraj/local_cluster_home/sqoop-1.4.1-incubating_hadoop-1.0.0/bin  
neeraj@ubuntu:~/local_cluster_home/sqoop-1.4.1-incubating_hadoop-1.0.0/bin$ ./sqoop import --connect 'jdbc:mysql://127.0.0.1/  
test' --username=root -password=1234 --target-dir /sqoop_import_dir --fields-terminated-by '\t' --table employee -m 1  
13/03/11 20:31:43 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.  
13/03/11 20:31:43 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.  
13/03/11 20:31:43 INFO tool.CodeGenTool: Beginning code generation
```

Verify imported data on HDFS

```
Terminal
neeraj@ubuntu:~/local_cluster_home/hadoop-1.0.3/bin$ ./hadoop fs -ls /
Found 8 items
drwxr-xr-x  - neeraj supergroup          0 2013-01-15 09:34 /dir1
drwxr-xr-x  - neeraj supergroup          0 2013-03-08 07:10 /dir2
drwxr-xr-x  - neeraj supergroup          0 2013-03-06 08:40 /hbase
drwxr-xr-x  - neeraj supergroup          0 2013-03-10 15:50 /home
drwxr-xr-x  - neeraj supergroup          0 2013-01-20 10:32 /input_files
drwxr-xr-x  - neeraj supergroup          0 2013-03-11 20:32 /sqoop_import_dir
drwxr-xr-x  - neeraj supergroup          0 2013-03-11 08:33 /tmp
drwxr-xr-x  - neeraj supergroup          0 2013-03-11 07:58 /user
neeraj@ubuntu:~/local_cluster_home/hadoop-1.0.3/bin$ ./hadoop fs -ls /sqoop_import_dir
Found 3 items
-rw-r--r--   1 neeraj supergroup          0 2013-03-11 20:32 /sqoop_import_dir/_SUCCESS
drwxr-xr-x  - neeraj supergroup          0 2013-03-11 20:31 /sqoop_import_dir/_logs
-rw-r--r--   1 neeraj supergroup        89 2013-03-11 20:32 /sqoop_import_dir/part-m-00000
neeraj@ubuntu:~/local_cluster_home/hadoop-1.0.3/bin$ ./hadoop fs -cat /sqoop_import_dir/part-m-00000
101      Neeraj   Kumar   Big Data
102      Sunil   Verma   Hadoop
103      Max     Basha   Hive
104      Sai     Kalyan  HBase
neeraj@ubuntu:~/local_cluster_home/hadoop-1.0.3/bin$
```

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
```



```
Terminal  
neeraj@ubuntu:~/local_cluster_home/sqoop-1.4.1-incubating__hadoop-1.0.0/bin$ pwd  
/home/neeraj/local_cluster_home/sqoop-1.4.1-incubating__hadoop-1.0.0/bin  
neeraj@ubuntu:~/local_cluster_home/sqoop-1.4.1-incubating__hadoop-1.0.0/bin$ ./sqoop export --connect 'jdbc:mysql://  
127.0.0.1/test' --username=root -password=1234 --export-dir /sqoop_import_dir/part-m-00000 --table employee --inpu  
t-fields-terminated-by '\t' -m 1  
13/03/11 22:38:11 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P
```


Verify exported data on RDBMS

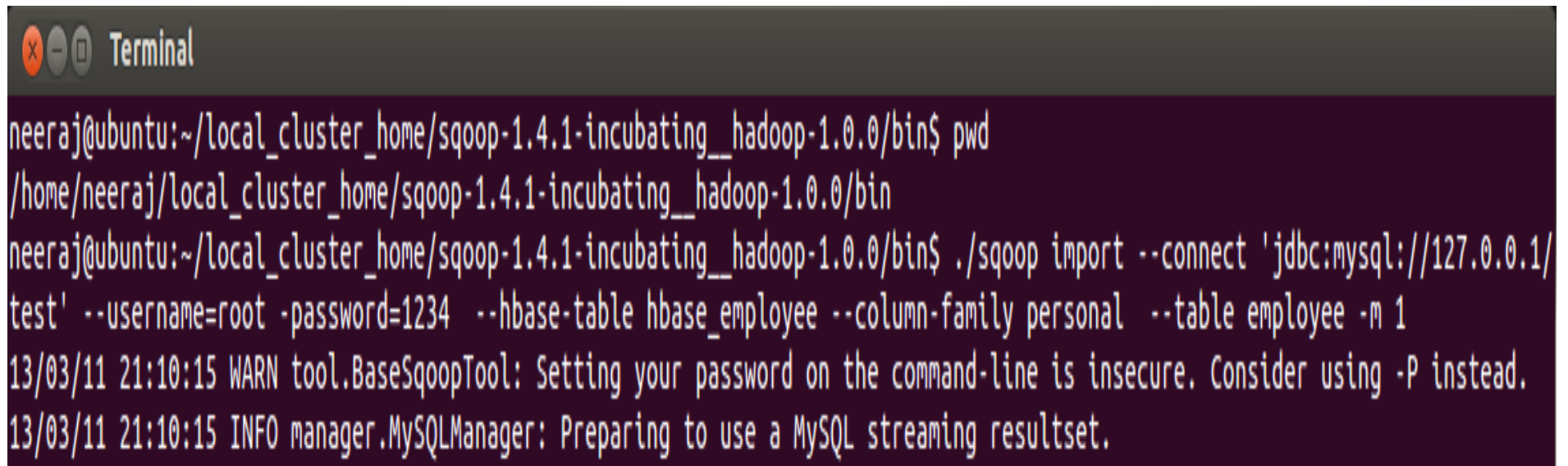
```
Terminal
mysql> SHOW TABLES;
+-----+
| Tables_in_test |
+-----+
| employee       |
+-----+
1 row in set (0.00 sec)

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)
```

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
```



```
Terminal  
neeraj@ubuntu:~/local_cluster_home/sqoop-1.4.1-incubating__hadoop-1.0.0/bin$ pwd  
/home/neeraj/local_cluster_home/sqoop-1.4.1-incubating__hadoop-1.0.0/bin  
neeraj@ubuntu:~/local_cluster_home/sqoop-1.4.1-incubating__hadoop-1.0.0/bin$ ./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  
13/03/11 21:10:15 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.  
13/03/11 21:10:15 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
```

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> 
```

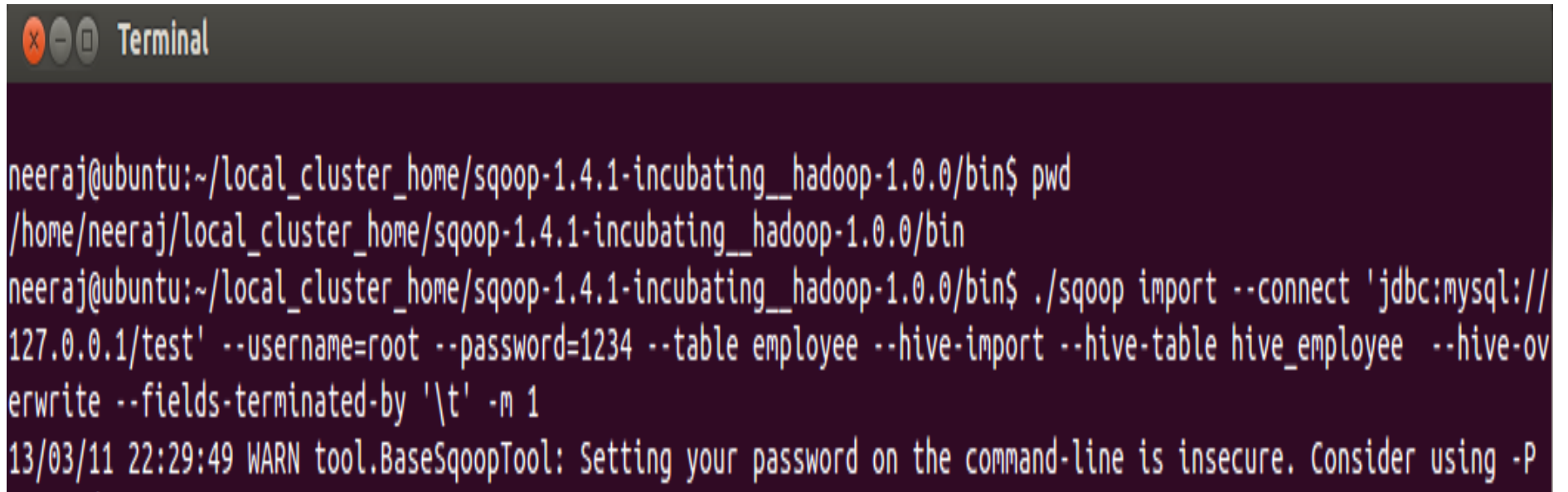
Export data from HBase to RDBMS



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

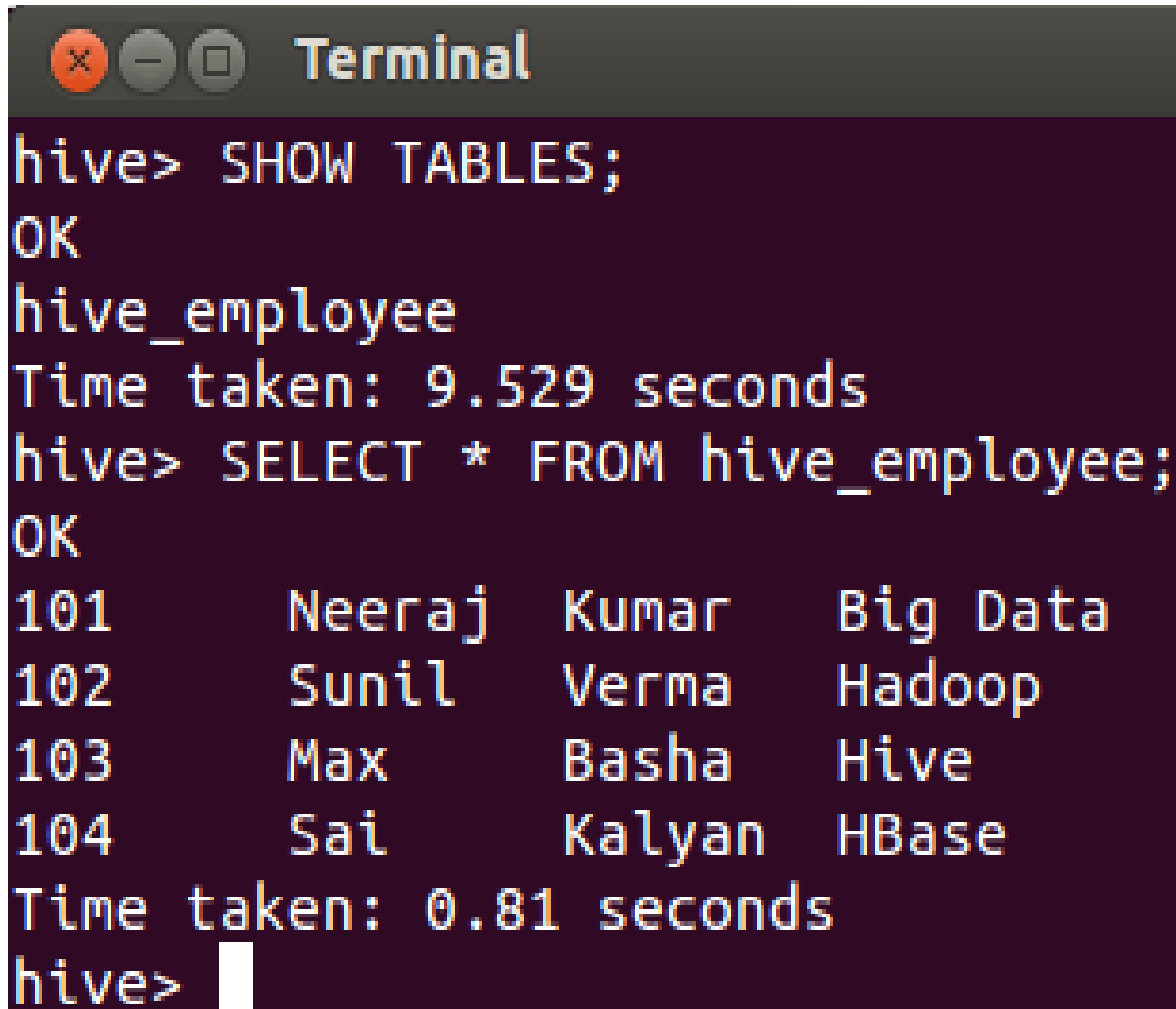
Import data from RDBMS to Hive

```
./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
```



```
Terminal  
neeraj@ubuntu:~/local_cluster_home/sqoop-1.4.1-incubating__hadoop-1.0.0/bin$ pwd  
/home/neeraj/local_cluster_home/sqoop-1.4.1-incubating__hadoop-1.0.0/bin  
neeraj@ubuntu:~/local_cluster_home/sqoop-1.4.1-incubating__hadoop-1.0.0/bin$ ./sqoop import --connect 'jdbc:mysql://  
127.0.0.1/test' --username=root --password=1234 --table employee --hive-import --hive-table hive_employee --hive-ov  
erwrite --fields-terminated-by '\t' -m 1  
13/03/11 22:29:49 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P
```

Verify imported data on Hive

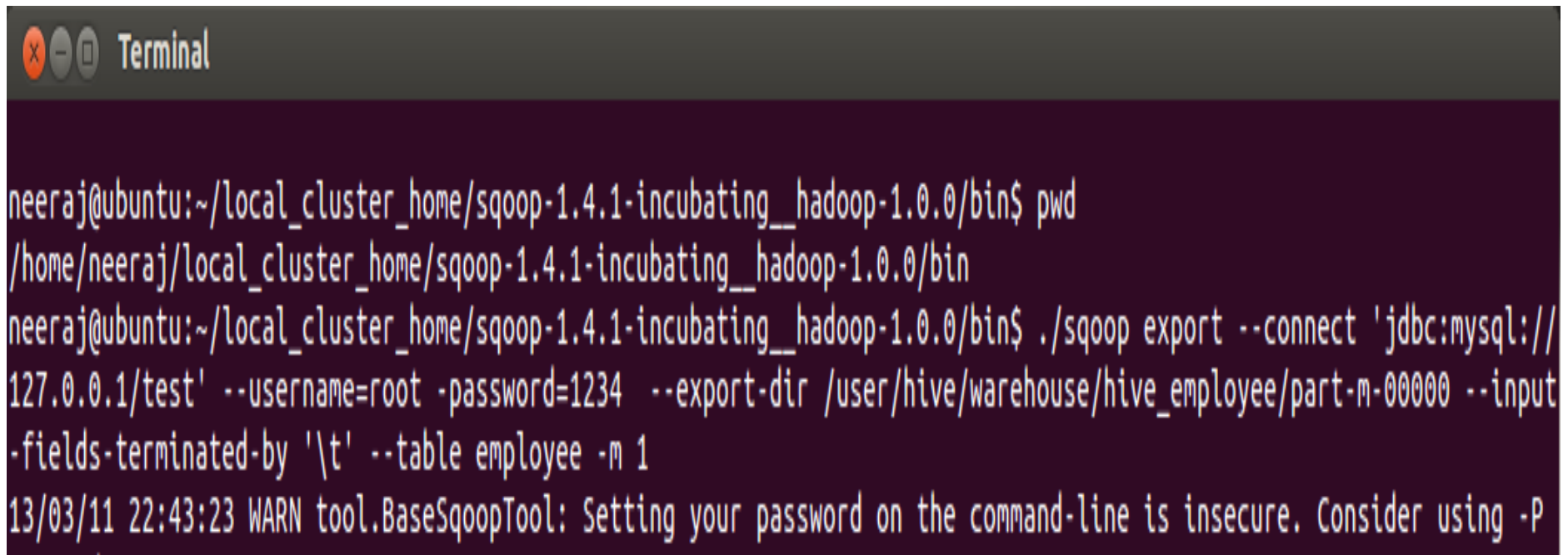


A terminal window titled "Terminal" with standard macOS window controls (close, minimize, maximize). The terminal shows the execution of Hive commands to verify imported data. The first command is `SHOW TABLES;`, which returns `OK` and lists `hive_employee`. The second command is `SELECT * FROM hive_employee;`, which also returns `OK` and displays a table with four columns: an ID, a name, a last name, and a technology. The table contains four rows of data. The terminal shows the time taken for each command: 9.529 seconds for the first and 0.81 seconds for the second.

```
hive> SHOW TABLES;
OK
hive_employee
Time taken: 9.529 seconds
hive> SELECT * FROM hive_employee;
OK
101      Neeraj   Kumar    Big Data
102      Sunil    Verma    Hadoop
103      Max      Basha    Hive
104      Sai      Kalyan   HBase
Time taken: 0.81 seconds
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
```

A terminal window titled "Terminal" with a dark background and light-colored text. The window shows the execution of the sqoop export command. The user's prompt is neeraj@ubuntu:~/local_cluster_home/sqoop-1.4.1-incubating__hadoop-1.0.0/bin\$. The first command is pwd, which returns /home/neeraj/local_cluster_home/sqoop-1.4.1-incubating__hadoop-1.0.0/bin. The second command is ./sqoop export --connect 'jdbc:mysql://127.0.0.1/test' --username=root -password=1234 --export-dir /user/hive/warehouse/hive_employee/part-m-000000 --input-fields-terminated-by '\t' --table employee -m 1. The output shows a warning message: 13/03/11 22:43:23 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P.

```
neeraj@ubuntu:~/local_cluster_home/sqoop-1.4.1-incubating__hadoop-1.0.0/bin$ pwd  
/home/neeraj/local_cluster_home/sqoop-1.4.1-incubating__hadoop-1.0.0/bin  
neeraj@ubuntu:~/local_cluster_home/sqoop-1.4.1-incubating__hadoop-1.0.0/bin$ ./sqoop export --connect 'jdbc:mysql://  
127.0.0.1/test' --username=root -password=1234 --export-dir /user/hive/warehouse/hive_employee/part-m-000000 --input  
-fields-terminated-by '\t' --table employee -m 1  
13/03/11 22:43:23 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P
```

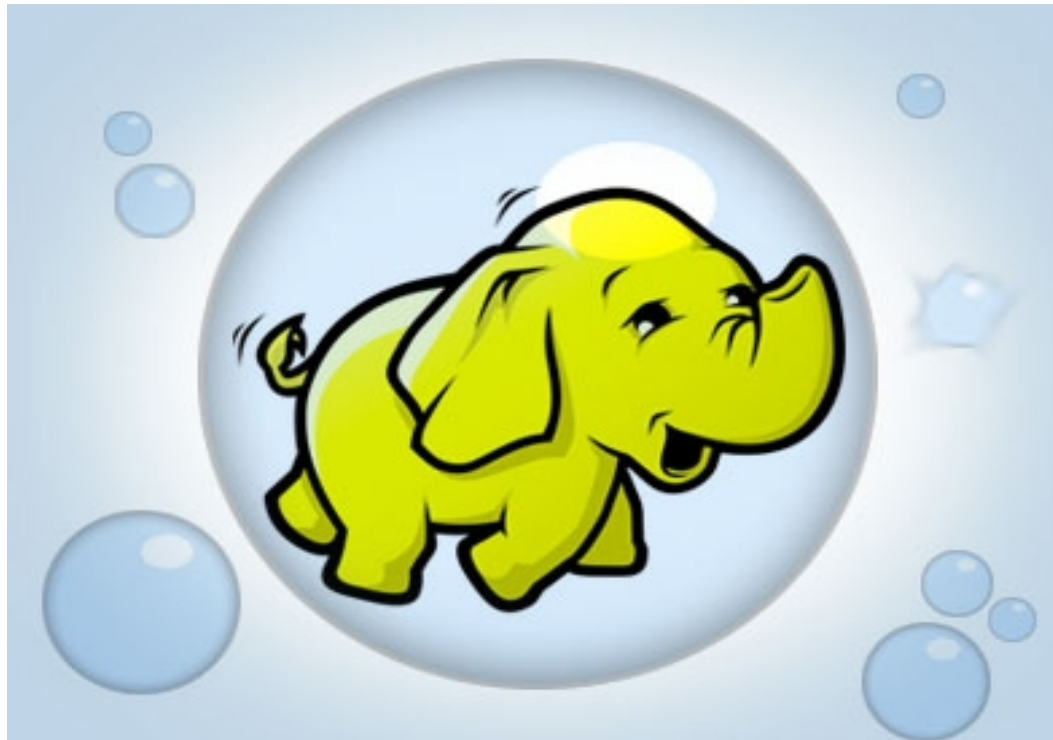
Verify exported data on RDBMS

```
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