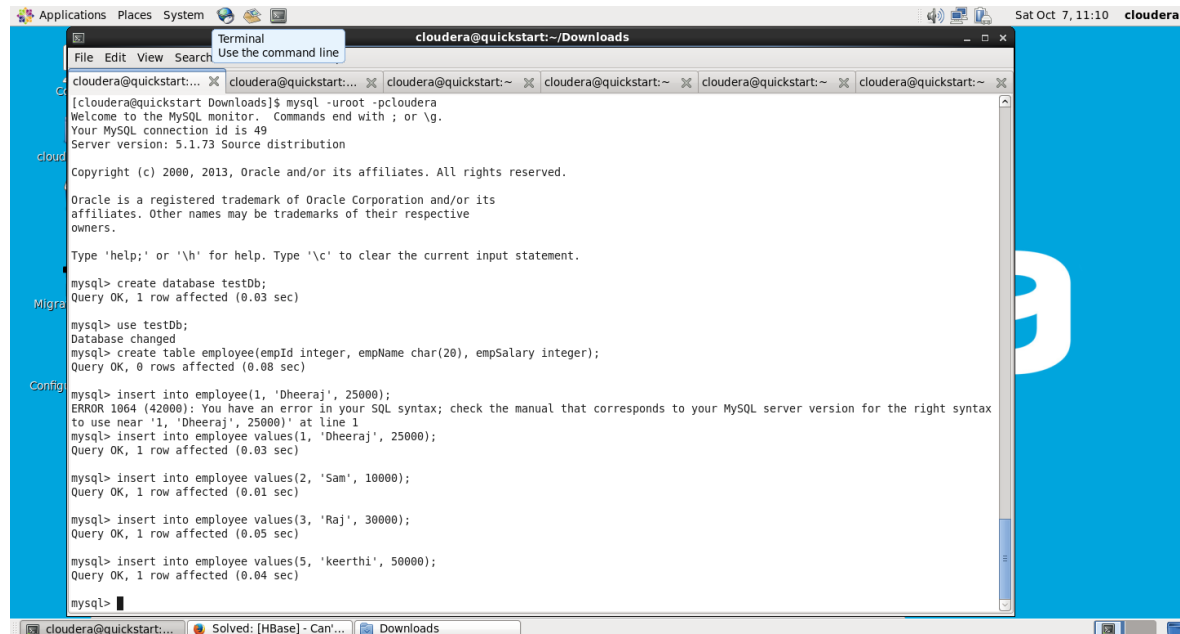


# Transfer data between Mysql and HDFS using sqoop:

## Importing data from mysql to HDFS:

1. Create database 'testDb' and create table student and add records to the table:



```
cloudera@quickstart:~/Downloads
mysql> create database testDb;
Query OK, 1 row affected (0.03 sec)

mysql> use testDb;
Database changed
mysql> create table employee(empId integer, empName char(20), empSalary integer);
Query OK, 0 rows affected (0.08 sec)

mysql> insert into employee values(1, 'Dheeraj', 25000);
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near '1, 'Dheeraj', 25000)' at line 1
mysql> insert into employee values(1, 'Dheeraj', 25000);
Query OK, 1 row affected (0.03 sec)

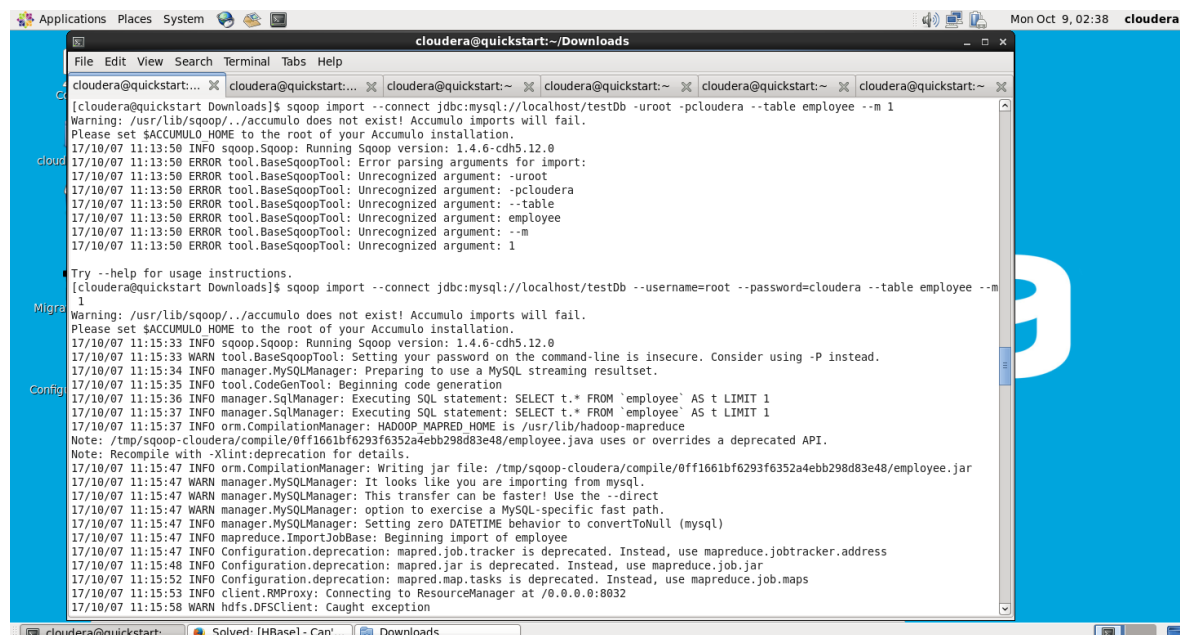
mysql> insert into employee values(2, 'Sam', 10000);
Query OK, 1 row affected (0.01 sec)

mysql> insert into employee values(3, 'Raj', 30000);
Query OK, 1 row affected (0.05 sec)

mysql> insert into employee values(5, 'keerthi', 50000);
Query OK, 1 row affected (0.04 sec)

mysql>
```

2. Import a table into HDFS using sqoop import command:



```
cloudera@quickstart:~/Downloads
[cloudera@quickstart Downloads]$ sqoop import --connect jdbc:mysql://localhost/testDb -uroot -pcloudera --table employee --m 1
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
17/10/07 11:13:50 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.12.0
17/10/07 11:13:50 ERROR tool.BaseSqoopTool: Error parsing arguments for import:
17/10/07 11:13:50 ERROR tool.BaseSqoopTool: Unrecognized argument: -uroot
17/10/07 11:13:50 ERROR tool.BaseSqoopTool: Unrecognized argument: -pcloudera
17/10/07 11:13:50 ERROR tool.BaseSqoopTool: Unrecognized argument: --table
17/10/07 11:13:50 ERROR tool.BaseSqoopTool: Unrecognized argument: employee
17/10/07 11:13:50 ERROR tool.BaseSqoopTool: Unrecognized argument: --m
17/10/07 11:13:50 ERROR tool.BaseSqoopTool: Unrecognized argument: 1

Try --help for usage instructions.
[cloudera@quickstart Downloads]$ sqoop import --connect jdbc:mysql://localhost/testDb --username=root --password=cloudera --table employee --m 1
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
17/10/07 11:15:33 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.12.0
17/10/07 11:15:33 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
17/10/07 11:15:34 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
17/10/07 11:15:35 INFO tool.CodeGenTool: Beginning code generation
17/10/07 11:15:36 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee` AS t LIMIT 1
17/10/07 11:15:37 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee` AS t LIMIT 1
17/10/07 11:15:37 INFO orm.CompilationManager: HADOOP MAPRED HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/0ff1661bf6293f6352a4ebb298d83e48/employee.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
17/10/07 11:15:47 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/0ff1661bf6293f6352a4ebb298d83e48/employee.jar
17/10/07 11:15:47 WARN manager.MySQLManager: It looks like you are importing from mysql.
17/10/07 11:15:47 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
17/10/07 11:15:47 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
17/10/07 11:15:47 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
17/10/07 11:15:47 INFO mapreduce.ImportJobBase: Beginning import of employee
17/10/07 11:15:47 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
17/10/07 11:15:48 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
17/10/07 11:15:52 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
17/10/07 11:15:53 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
17/10/07 11:15:58 WARN hdfs.DFSClient: Caught exception
```

```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Tabs Help
cloudera@quickstart:~
cloudera@quickstart:~
cloudera@quickstart:~
cloudera@quickstart:~
cloudera@quickstart:~
cloudera@quickstart:~
cloudera@quickstart:~

FILE: Number of bytes read=0
FILE: Number of bytes written=151260
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=87
HDFS: Number of bytes written=56
HDFS: Number of read operations=4
HDFS: Number of large read operations=0
HDFS: Number of write operations=2

Job Counters
  Launched map tasks=1
  Other local map tasks=1
  Total time spent by all maps in occupied slots (ms)=19833
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=19833
  Total vcore-milliseconds taken by all map tasks=19833
  Total megabyte-milliseconds taken by all map tasks=20308992

Map-Reduce Framework
  Map input records=4
  Map output records=4
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=236
  CPU time spent (ms)=1480
  Physical memory (bytes) snapshot=112713728
  Virtual memory (bytes) snapshot=1518117376
  Total committed heap usage (bytes)=60751872

File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=56
17/10/07 11:17:00 INFO mapreduce.ImportJobBase: Transferred 56 bytes in 67.6688 seconds (0.8276 bytes/sec)
17/10/07 11:17:00 INFO mapreduce.ImportJobBase: Retrieved 4 records.
[cloudera@quickstart Downloads]$
```

3. Once import is done, data from the student table should be present in the HDFS at /user/cloudera/employee/part-m-00000 location:

```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Tabs Help
cloudera@quickstart:~
cloudera@quickstart:~
cloudera@quickstart:~
cloudera@quickstart:~
cloudera@quickstart:~
cloudera@quickstart:~
cloudera@quickstart:~

File Output Format Counters
  Bytes Written=56
17/10/07 11:17:00 INFO mapreduce.ImportJobBase: Transferred 56 bytes in 67.6688 seconds (0.8276 bytes/sec)
17/10/07 11:17:00 INFO mapreduce.ImportJobBase: Retrieved 4 records.
[cloudera@quickstart Downloads]$ hadoop fs -ls /
Found 7 items
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:34 /benchmarks
drwxr-xr-x - hbase supergroup 0 2017-10-07 10:39 /hbase
drwxr-xr-x - solr solr 0 2017-07-19 05:37 /solr
-rw-r--r-- 1 cloudera supergroup 26204 2017-10-07 10:31 /student.txt
drwxr-xrwt - hdfs supergroup 0 2017-10-07 10:52 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /user
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /var
[cloudera@quickstart Downloads]$ hadoop fs -ls /user
Found 8 items
drwxr-xr-x - cloudera cloudera 0 2017-10-07 11:16 /user/cloudera
drwxr-xr-x - mapred hadoop 0 2017-07-19 05:34 /user/history
drwxr-xrwx - hive supergroup 0 2017-07-19 05:36 /user/hive
drwxr-xrwx - hue supergroup 0 2017-07-19 05:35 /user/hue
drwxr-xrwx - jenkins supergroup 0 2017-07-19 05:35 /user/jenkins
drwxr-xrwx - oozie supergroup 0 2017-10-06 05:16 /user/oozie
drwxr-xrwx - root supergroup 0 2017-07-19 05:35 /user/root
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /user/spark
[cloudera@quickstart Downloads]$ hadoop fs -ls /user/cloudera/
Found 2 items
drwxr-xr-x - cloudera cloudera 0 2017-10-07 11:16 /user/cloudera/employee
drwxr-xr-x - cloudera cloudera 0 2017-10-06 21:59 /user/cloudera/oozie-oozi
[cloudera@quickstart Downloads]$ hadoop fs -ls /user/cloudera/employee/
Found 2 items
-rw-r--r-- 1 cloudera cloudera 0 2017-10-07 11:16 /user/cloudera/employee/_SUCCESS
-rw-r--r-- 1 cloudera cloudera 56 2017-10-07 11:16 /user/cloudera/employee/part-m-00000
[cloudera@quickstart Downloads]$ hadoop fs -cat /user/cloudera/employee/part-m-00000
1,Dheeraj,25000
2,Sam,10000
3,Raj,30000
5,keerthi,50000
[cloudera@quickstart Downloads]$
```

## Task: Export Data from HDFS to mysql using swoop:

1. Create a new tables in the mysql in existing database testDb:

```

cloudera@quickstart:~/Downloads
File Edit View Search Terminal Tabs Help
cloudera@quickstart:~
mysql> use testDb;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> CREATE TABLE acad_student(student_name varchar(20), sector varchar(20), DOB date, qualification varchar(10), score float(8,4), state va
rchar(10),random_name varchar(20));
Query OK, 0 rows affected (0.05 sec)

mysql> select * from acad_student;
Empty set (0.00 sec)

mysql> describe acad_student;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| student_name | varchar(20) | YES | | NULL | |
| sector | varchar(20) | YES | | NULL | |
| DOB | date | YES | | NULL | |
| qualification | varchar(10) | YES | | NULL | |
| score | float(8,4) | YES | | NULL | |
| state | varchar(10) | YES | | NULL | |
| random_name | varchar(20) | YES | | NULL | |
+-----+-----+-----+-----+-----+-----+
7 rows in set (0.01 sec)

mysql> select * from acad_student;
Empty set (0.00 sec)

mysql> DROP TABLE acad_student;
Query OK, 0 rows affected (0.05 sec)

mysql> describe acad student;
ERROR 1146 (42S02): Table 'testDb.acad_student' doesn't exist
mysql> CREATE TABLE acad_student(student_name varchar(20), sector varchar(20),qualification varchar(10), state varchar(10));
Query OK, 0 rows affected (0.01 sec)

```

2. Copy the file that needs to be exported from HDFS using command

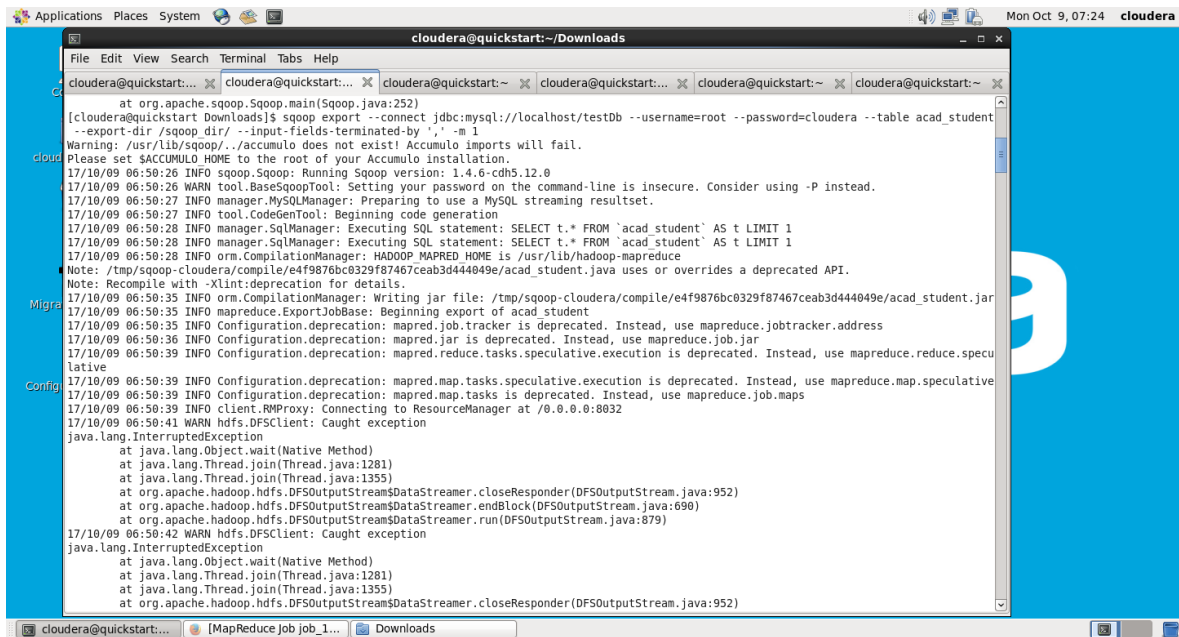
1. Hadoop fs -put <file\_name> <path\_in\_hdfs\_file\_system>

```

cloudera@quickstart:~/Downloads
File Edit View Search Terminal Tabs Help
cloudera@quickstart:~
ABROSER, government, 18-11-2002, MBBS, 3.5, Pennsylvania, prattville*
ALEXANDER, government, 20-10-2000, BSC, 2.5, Vermont, gadsden+
ALEXANDER, private, 20-10-2000, BE, 8.5, arizona, decatur!
ALEXANDER, government, 01-01-2003, BTECH, 4.5, oregon, huntsville/
AGNEW, government, 20-10-2000, BCOM, 7.5, california, dothan@
ATNEST, government, 20-10-2000, MTECH, 8.5, arizona, decatur!
BELL, government, 10-07-2004, BBA, 9.5, alaska, auburn~
BURR, government, 12-12-2001, BE, 100, alabama, madison'
BURD, government, 20-10-2000, ME, 6.5, louisiana, hoover#
BACHTEL, government, 28-04-2005, BE, 100, alabama, madison'
MULVETS, government, 20-10-2000, MS, 8.5, arizona, decatur!
BURTNER, government, 20-10-2000, BE, 100, alabama, madison'
BLOOM, government, 20-10-2000, BE, 4.5, oregon, huntsville/
BECHTEL, government, 20-10-2000, BE, 5.5, Maryland, tuscaloosa$
[cloudera@quickstart Downloads]$ ls
acad student.java employee.java student info.txt student.txt
[cloudera@quickstart Downloads]$ vi student info.txt
[cloudera@quickstart Downloads]$ hadoop fs -rm /sqaop_dir/student_info.txt
Deleted /sqaop_dir/student_info.txt
[cloudera@quickstart Downloads]$ hadoop fs -put student_info.txt /sqaop_dir/
[cloudera@quickstart Downloads]$ hadoop fs -rm /sqaop_dir/student_info.txt
Deleted /sqaop_dir/student_info.txt
[cloudera@quickstart Downloads]$ vi student info.txt
[cloudera@quickstart Downloads]$ hadoop fs -put student_info.txt /sqaop_dir/
[cloudera@quickstart Downloads]$ cat student_info.txt
ALEXANDER, government, BTECH, oregon
AGNEW, government, BCOM, california
ATNEST, government, MTECH, arizona
BELL, government, BBA, alaska
BURR, government, BE, alabama
BURD, government, ME, louisiana
BACHTEL, government, BE, alabama
MULVETS, government, MS, arizona
BURTNER, government, BE, alabama
BLOOM, government, BE, oregon
BECHTEL, government, BE, Maryland
[cloudera@quickstart Downloads]$

```

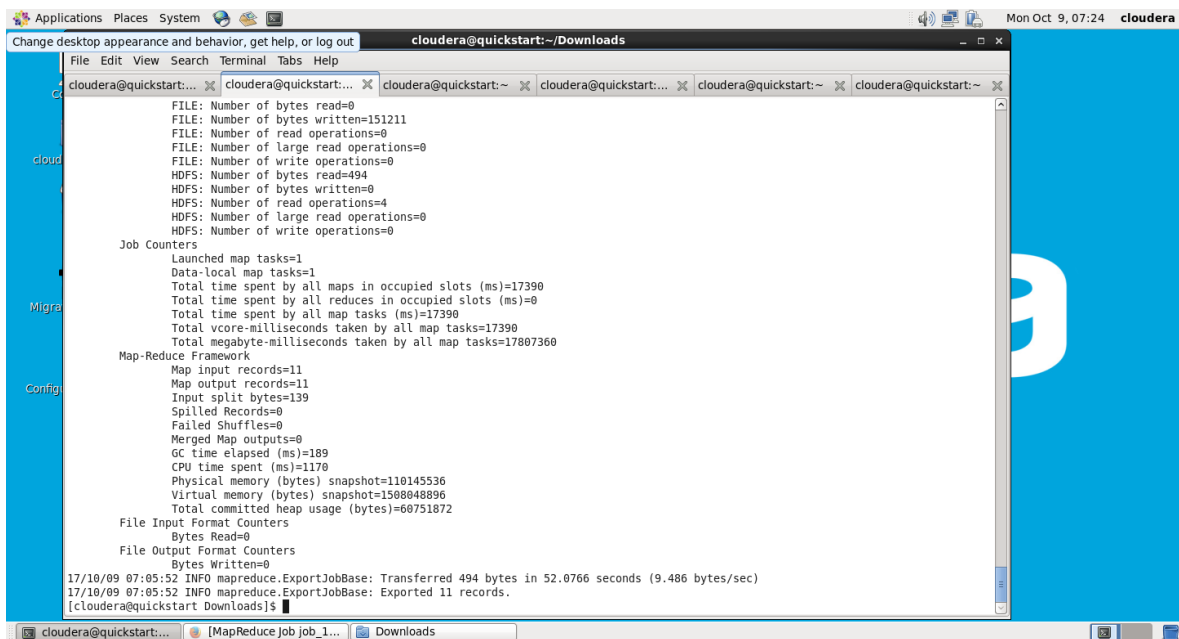
3. Export file from HDFS to mysql using sqoop export command:



```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Tabs Help

cloudera@quickstart:~/Downloads$ sqoop export --connect jdbc:mysql://localhost/testDb --username=root --password=cldudera --table acad_student
--export-dir /sqoop dir/ --input-fields-terminated-by ',' -m 1
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
17/10/09 06:50:26 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.12.0
17/10/09 06:50:26 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
17/10/09 06:50:27 INFO tool.CodeGenTool: Beginning code generation
17/10/09 06:50:28 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'acad student' AS t LIMIT 1
17/10/09 06:50:28 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'acad student' AS t LIMIT 1
17/10/09 06:50:28 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/e4f9876bc0329f87467ceab3d444049e/acad_student.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
17/10/09 06:50:35 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/e4f9876bc0329f87467ceab3d444049e/acad_student.jar
17/10/09 06:50:35 INFO mapreduce.ExportJobBase: Beginning export of acad student
17/10/09 06:50:35 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
17/10/09 06:50:36 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
17/10/09 06:50:39 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
17/10/09 06:50:39 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
17/10/09 06:50:39 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
17/10/09 06:50:41 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
17/10/09 06:50:42 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
```

4. Job completion statement will be displayed as



```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Tabs Help

cloudera@quickstart:~/Downloads$ sqoop export --connect jdbc:mysql://localhost/testDb --username=root --password=cldudera --table acad_student
--export-dir /sqoop dir/ --input-fields-terminated-by ',' -m 1
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
17/10/09 06:50:26 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.12.0
17/10/09 06:50:26 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
17/10/09 06:50:27 INFO tool.CodeGenTool: Beginning code generation
17/10/09 06:50:28 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'acad student' AS t LIMIT 1
17/10/09 06:50:28 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'acad student' AS t LIMIT 1
17/10/09 06:50:28 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/e4f9876bc0329f87467ceab3d444049e/acad_student.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
17/10/09 06:50:35 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/e4f9876bc0329f87467ceab3d444049e/acad_student.jar
17/10/09 06:50:35 INFO mapreduce.ExportJobBase: Beginning export of acad student
17/10/09 06:50:35 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
17/10/09 06:50:36 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
17/10/09 06:50:39 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
17/10/09 06:50:39 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
17/10/09 06:50:39 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
17/10/09 06:50:41 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
17/10/09 06:50:42 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)

17/10/09 07:05:52 INFO mapreduce.ExportJobBase: Transferred 494 bytes in 52.0766 seconds (9.486 bytes/sec)
17/10/09 07:05:52 INFO mapreduce.ExportJobBase: Exported 11 records.
cloudera@quickstart:~/Downloads$
```