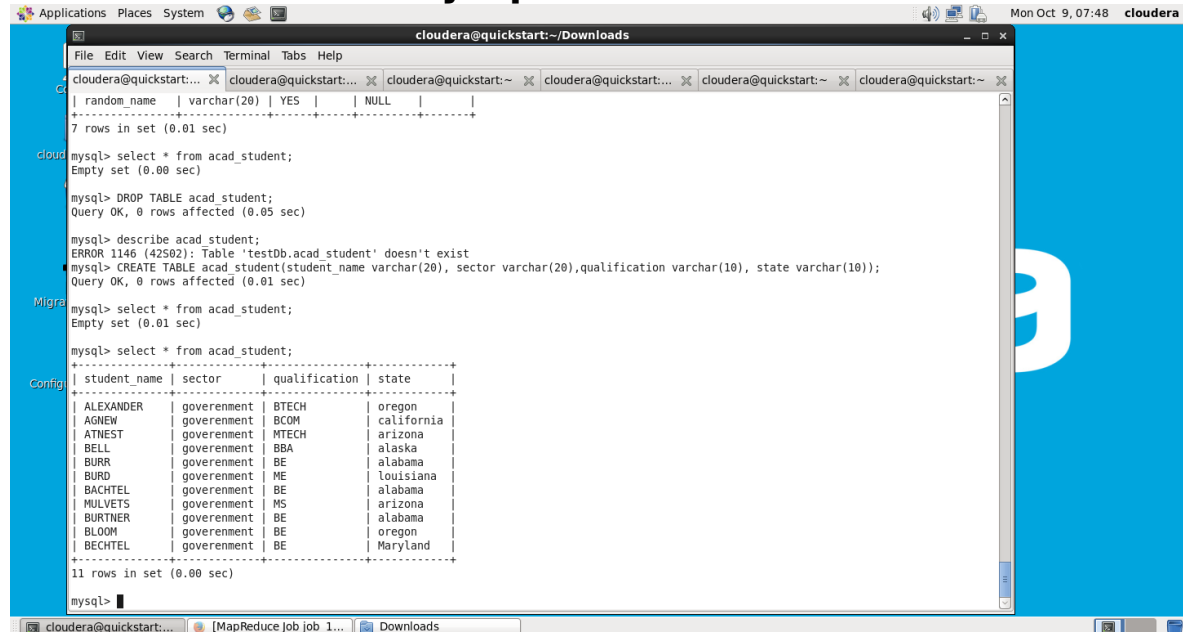


Transfer data between mysql and Hive using sqoop:

Import data from mysql to Hive:

1. Create table in mysql and load some data:



```
cloudera@quickstart:~/Downloads
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 (42502): 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)

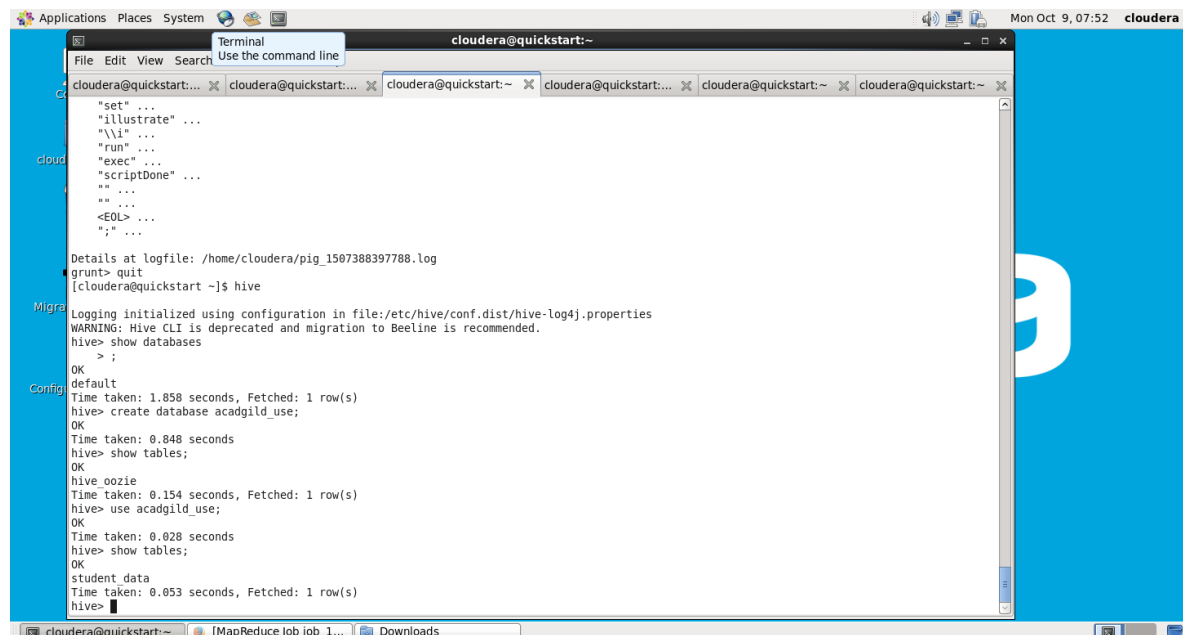
mysql> select * from acad_student;
Empty set (0.01 sec)

mysql> select * from acad_student;
+-----+-----+-----+-----+
| student_name | sector | qualification | state |
+-----+-----+-----+-----+
| ALEXANDER    | goverment | BTECH | oregon |
| AGNEW        | goverment | BCOM  | california |
| ATNEST       | goverment | MTECH | arizona |
| BELL         | goverment | BBA   | alaska |
| BURR         | goverment | BE    | alabama |
| BURD        | goverment | ME    | louisiana |
| BACHTEL      | goverment | BE    | alabama |
| MULVETS      | goverment | MS    | arizona |
| BURTNER      | goverment | BE    | alabama |
| BLOOM        | goverment | BE    | oregon |
| BECHTEL      | goverment | BE    | Maryland |
+-----+-----+-----+-----+
11 rows in set (0.00 sec)

mysql>
```

acad_student table have total 11 rows.

2. Create database in hive table and copy the acad_student table to hive:



```
cloudera@quickstart:~
"set" ...
"illustrate" ...
"\l" ...
"run" ...
"exec" ...
"scriptDone" ...
"..." ...
"<EOL>" ...
";" ...

Details at logfile: /home/cloudera/pig_1507388397788.log
grunt> quit
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> show databases;
+-----+
| > |
+-----+
OK
default
Time taken: 1.858 seconds, Fetched: 1 row(s)
hive> create database acadgild_use;
OK
Time taken: 0.848 seconds
hive> show tables;
+-----+
| > |
+-----+
OK
hive_oozie
Time taken: 0.154 seconds, Fetched: 1 row(s)
hive> use acadgild_use;
OK
Time taken: 0.028 seconds
hive> show tables;
+-----+
| > |
+-----+
OK
student_data
Time taken: 0.053 seconds, Fetched: 1 row(s)
hive>
```

3. Scoop import command to import data directly to hive table:

```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Tabs Help
cloudera@quickstart:~$ sqoop import --connect jdbc:mysql://localhost/testdb --username=root --password=cloudera --columns student_name,qualification --table acad_student --target-dir /user/cloudera/acad_student --fields-terminated-by "," --hive-import --create-hive-table --hive-table acadgild_use.student_data -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 07:43:38 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.12.0
17/10/09 07:43:38 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
17/10/09 07:43:39 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
17/10/09 07:43:39 INFO tool.CodeGenTool: Beginning code generation
17/10/09 07:43:40 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `acad_student` AS t LIMIT 1
17/10/09 07:43:40 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `acad_student` AS t LIMIT 1
17/10/09 07:43:40 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/223e9260a2f26057b6766f842d11f065/acad_student.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
17/10/09 07:43:47 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/223e9260a2f26057b6766f842d11f065/acad_student.jar
17/10/09 07:43:47 WARN manager.MySQLManager: It looks like you are importing from mysql.
17/10/09 07:43:47 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
17/10/09 07:43:47 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
17/10/09 07:43:47 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
17/10/09 07:43:47 INFO mapreduce.ImportJobBase: Beginning import of acad_student
17/10/09 07:43:47 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
17/10/09 07:43:48 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
17/10/09 07:43:52 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
17/10/09 07:43:53 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
17/10/09 07:43:56 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 07:43:57 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
```

4. Once the import is complete, data is copied to hive table using acadgild_use database:

```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Tabs Help
cloudera@quickstart:~$ sqoop export --connect jdbc:mysql://localhost/testdb --username=root --password=cloudera --table acad_student --export-dir /user/cloudera/acad_student --fields-terminated-by "," --hive-table acadgild_use.student_data -m 1
Job Counters
  Launched map tasks=1
  Other local map tasks=1
  Total time spent by all maps in occupied slots (ms)=18245
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=18245
  Total vcore-millisecons taken by all map tasks=18245
  Total megabyte-millisecons taken by all map tasks=18682880
Map-Reduce Framework
  Map input records=11
  Map output records=11
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=229
  CPU time spent (ms)=1840
  Physical memory (bytes) snapshot=132018176
  Virtual memory (bytes) snapshot=1510146048
  Total committed heap usage (bytes)=60751872
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=118
17/10/09 07:44:50 INFO mapreduce.ImportJobBase: Transferred 118 bytes in 58.4056 seconds (2.0204 bytes/sec)
17/10/09 07:44:50 INFO mapreduce.ImportJobBase: Retrieved 11 records.
17/10/09 07:44:50 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `acad_student` AS t LIMIT 1
17/10/09 07:44:50 INFO hive.HiveImport: Loading uploaded data into Hive
Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-1.1.0-cdh5.12.0.jar!/hive-log4j.properties
OK
Time taken: 6.979 seconds
Loading data to table acadgild_use.student_data
Table acadgild_use.student_data stats: [numFiles=1, totalSize=118]
OK
Time taken: 2.402 seconds
cloudera@quickstart:~/Downloads$
```

5. Check the data in hive table to make sure the data is copied properly:

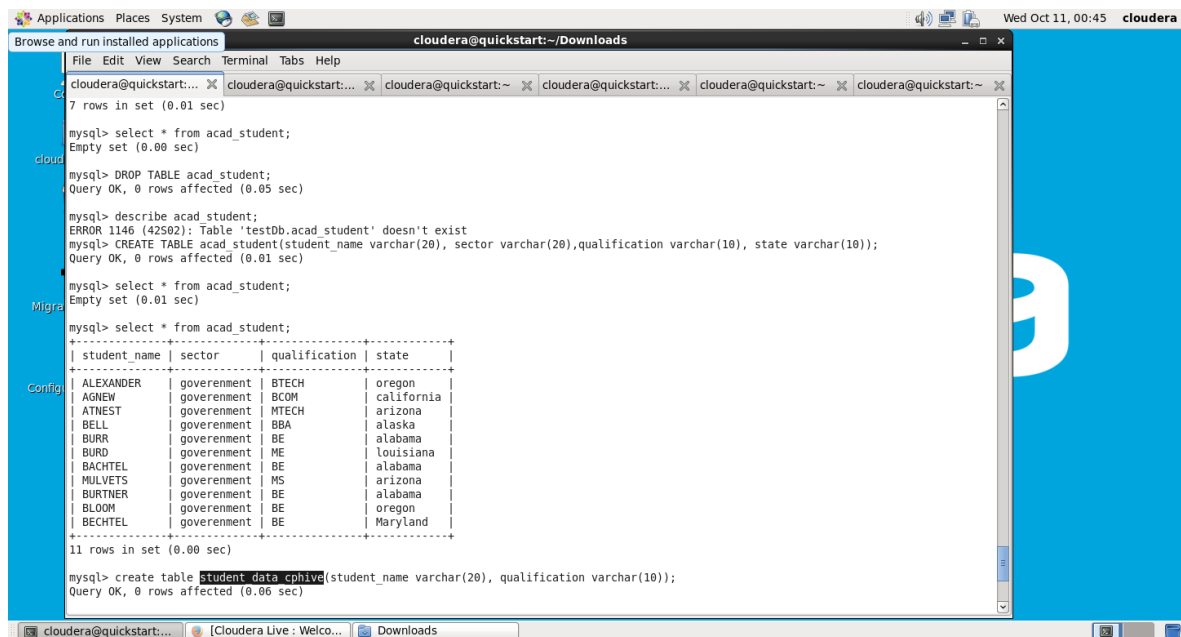
```
cloudera@quickstart:~  
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties  
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.  
hive> show databases  
OK  
> ;  
OK  
default  
Time taken: 1.858 seconds, Fetched: 1 row(s)  
hive> create database acadgild_use;  
OK  
Time taken: 0.848 seconds  
hive> show tables;  
OK  
hive_oozie  
Time taken: 0.154 seconds, Fetched: 1 row(s)  
hive> use acadgild_use;  
OK  
Time taken: 0.028 seconds  
hive> show tables;  
OK  
student_data  
Time taken: 0.053 seconds, Fetched: 1 row(s)  
hive> select * from student_data;  
OK  
ALEXANDER      BTECH  
AGNEW          BCOM  
ATNEST         MTECH  
BELL           BBA  
BURR           BE  
BURD           ME  
BACHTEL        BE  
MULVETS        MS  
BURTNER        BE  
BLOOM          BE  
BECHTEL        BE  
Time taken: 1.295 seconds, Fetched: 11 row(s)  
hive>
```

Export data from hive to mysql using sqoop:

1.create table in hive with some data in table or use the existing table with data for export. Here I am using the existing table student_data with columns student_name and qualification:

```
cloudera@quickstart:~  
Time taken: 0.043 seconds, Fetched: 1 row(s)  
hive> select * from student_data;  
OK  
ALEXANDER      BTECH  
AGNEW          BCOM  
ATNEST         MTECH  
BELL           BBA  
BURR           BE  
BURD           ME  
BACHTEL        BE  
MULVETS        MS  
BURTNER        BE  
BLOOM          BE  
BECHTEL        BE  
Time taken: 0.129 seconds, Fetched: 11 row(s)  
hive> show databases;  
OK  
acadgild_use  
default  
Time taken: 0.018 seconds, Fetched: 2 row(s)  
hive> use default  
OK  
> ;  
OK  
Time taken: 0.063 seconds  
hive> show tables;  
OK  
hive_oozie  
Time taken: 0.026 seconds, Fetched: 1 row(s)  
hive> select * from hive_oozie;  
OK  
Time taken: 0.109 seconds  
hive> use acadgild_use;  
OK  
Time taken: 0.037 seconds  
hive> show tables;  
OK  
student_data
```

2. Create table in mysql with the same number of columns as hive:



```
cloudera@quickstart:~/Downloads
mysql> select * from acad_student;
Empty set (0.00 sec)

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

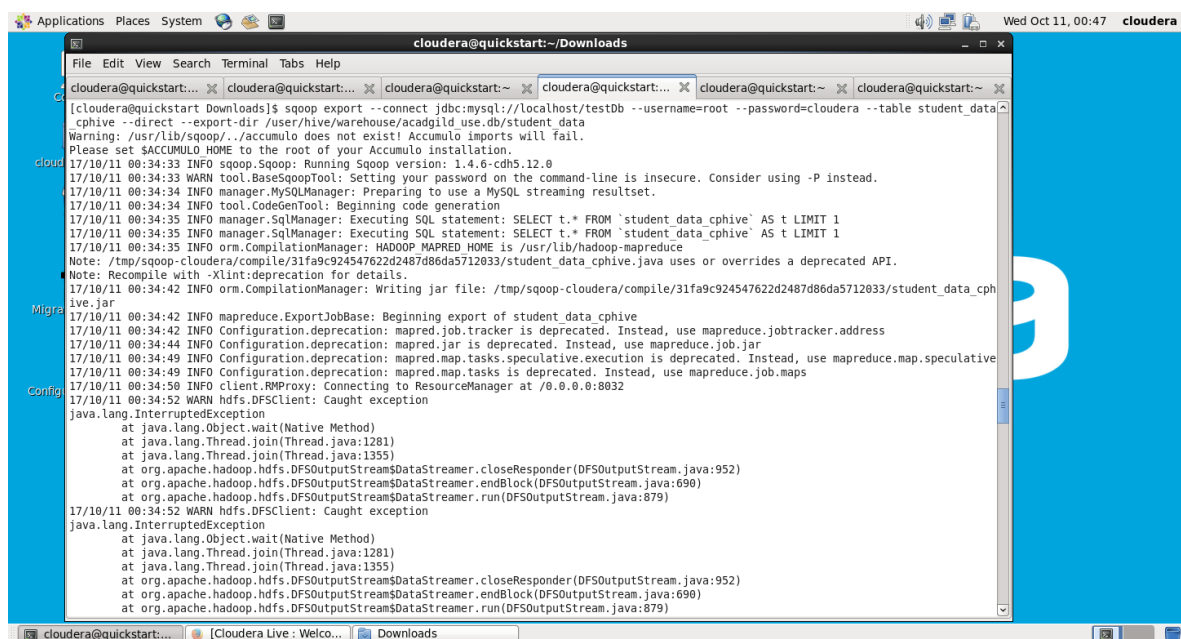
mysql> describe acad_student;
ERROR 1146 (42502): Table 'testDb.acad_student' doesn't exist
mysql> CREATE TABLE acad_student(student_name varchar(20), sector varchar(20), qualification varchar(10), state varchar(10));
Query OK, 0 rows affected (0.01 sec)

mysql> select * from acad_student;
Empty set (0.01 sec)

mysql> select * from acad_student;
+-----+-----+-----+-----+
| student_name | sector | qualification | state |
+-----+-----+-----+-----+
| 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 |
+-----+-----+-----+-----+
11 rows in set (0.00 sec)

mysql> create table student_data cphive(student_name varchar(20), qualification varchar(10));
Query OK, 0 rows affected (0.06 sec)
```

3. Transfer data from hive to mysql using sqoop export command:



```
cloudera@quickstart:~/Downloads
[cloudera@quickstart Downloads]$ sqoop export --connect jdbc:mysql://localhost/testDb --username=root --password=cloudera --table student_data
cphive --direct --export-dir /user/hive/warehouse/acadgild use.db/student_data
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
17/10/11 00:34:33 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.12.0
17/10/11 00:34:33 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
17/10/11 00:34:34 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
17/10/11 00:34:34 INFO tool.CodeGenTool: Beginning code generation
17/10/11 00:34:35 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'student_data cphive' AS t LIMIT 1
17/10/11 00:34:35 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'student_data cphive' AS t LIMIT 1
17/10/11 00:34:35 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/31fa9c924547622d2487d86da5712033/student_data_cphive.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
17/10/11 00:34:42 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/31fa9c924547622d2487d86da5712033/student_data_cphive.jar
17/10/11 00:34:42 INFO mapreduce.ExportJobBase: Beginning export of student_data cphive
17/10/11 00:34:42 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
17/10/11 00:34:44 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
17/10/11 00:34:49 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
17/10/11 00:34:50 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
17/10/11 00:34:52 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedExce...
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$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)
17/10/11 00:34:52 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedExce...
    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)
```

4. Once it is completed, check for the exported success command:

```
cloudera@quickstart:~/Downloads
FILE: Number of bytes read=0
FILE: Number of bytes written=607776
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=1135
HDFS: Number of bytes written=0
HDFS: Number of read operations=19
HDFS: Number of large read operations=0
HDFS: Number of write operations=0
Job Counters
  Launched map tasks=4
  Data-local map tasks=4
  Total time spent by all maps in occupied slots (ms)=178504
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=178504
  Total vcore-milliseconds taken by all map tasks=178504
  Total megabyte-milliseconds taken by all map tasks=182788096
Map-Reduce Framework
  Map input records=11
  Map output records=0
  Input split bytes=806
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=891
  CPU time spent (ms)=2030
  Physical memory (bytes) snapshot=410935936
  Virtual memory (bytes) snapshot=6027870208
  Total committed heap usage (bytes)=243007408
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=0
17/10/11 00:36:18 INFO mapreduce.ExportJobBase: Transferred 1.1084 KB in 88.6072 seconds (12.8093 bytes/sec)
17/10/11 00:36:18 INFO mapreduce.ExportJobBase: Exported 11 records.
[cloudera@quickstart Downloads]$
```

5. Check the data in mysql table for the successful transfer of data:

```
cloudera@quickstart:~/Downloads
mysql> create table student_data_cphive(student_name varchar(20), qualification varchar(10));
Query OK, 0 rows affected (0.06 sec)

mysql> select * from student_data_cphive
->
+-----+-----+
| student_name | qualification |
+-----+-----+
| BELL         | BBA          |
| BURR        | BE           |
| BURD        | ME           |
| ALEXANDER    | BTECH        |
| AGNEW       | BCOM         |
| ATNEST      | MTECH        |
| BLOOM       | BE           |
| BECHTEL     | BE           |
| BACHTEL     | BE           |
| MULVETS     | MS           |
| BURTNER    | BE           |
+-----+-----+
11 rows in set (0.02 sec)

mysql>
```