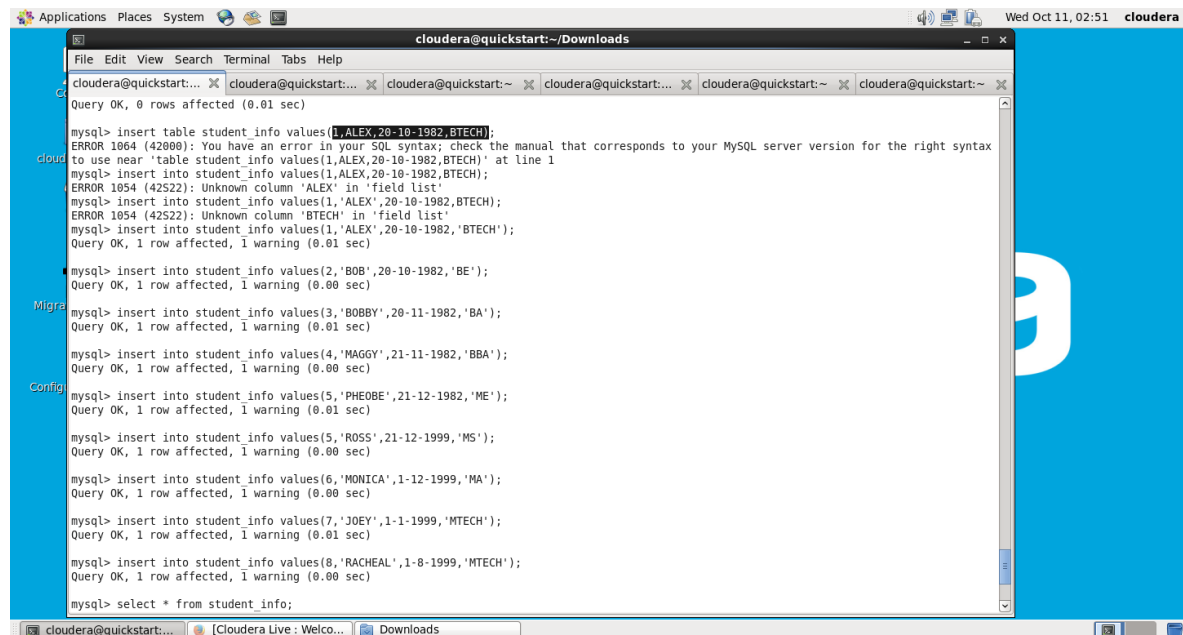


Incremental load in hive:

1. Create table in mysql and insert some data.



```
cloudera@quickstart:~/Downloads
mysql> create table student_info values(1,ALEX,20-10-1982,BTECH);
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 'table student_info values(1,ALEX,20-10-1982,BTECH)' at line 1
mysql> insert into student_info values(1,ALEX,20-10-1982,BTECH);
ERROR 1054 (42S22): Unknown column 'ALEX' in 'field list'
mysql> insert into student_info values(1,'ALEX',20-10-1982,BTECH);
ERROR 1054 (42S22): Unknown column 'BTECH' in 'field list'
mysql> insert into student_info values(1,'ALEX',20-10-1982,'BTECH');
Query OK, 1 row affected, 1 warning (0.01 sec)

mysql> insert into student_info values(2,'BOB',20-10-1982,'BE');
Query OK, 1 row affected, 1 warning (0.00 sec)

mysql> insert into student_info values(3,'BOBBY',20-11-1982,'BA');
Query OK, 1 row affected, 1 warning (0.01 sec)

mysql> insert into student_info values(4,'MAGGY',21-11-1982,'BBA');
Query OK, 1 row affected, 1 warning (0.00 sec)

mysql> insert into student_info values(5,'PHEOBE',21-12-1982,'ME');
Query OK, 1 row affected, 1 warning (0.01 sec)

mysql> insert into student_info values(5,'ROSS',21-12-1999,'MS');
Query OK, 1 row affected, 1 warning (0.00 sec)

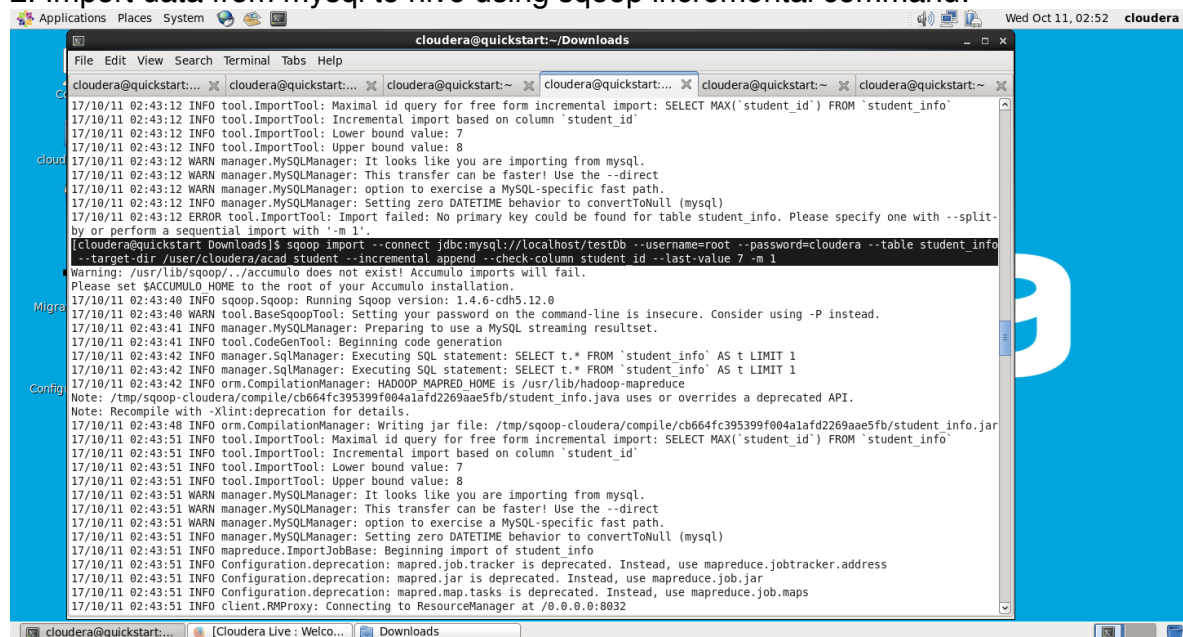
mysql> insert into student_info values(6,'MONICA',1-12-1999,'MA');
Query OK, 1 row affected, 1 warning (0.00 sec)

mysql> insert into student_info values(7,'JOEY',1-1-1999,'MTECH');
Query OK, 1 row affected, 1 warning (0.01 sec)

mysql> insert into student_info values(8,'RACHEAL',1-8-1999,'MTECH');
Query OK, 1 row affected, 1 warning (0.00 sec)

mysql> select * from student_info;
```

2. Import data from mysql to hive using sqoop incremental command:



```
cloudera@quickstart:~/Downloads
17/10/11 02:43:12 INFO tool.ImportTool: Maximal id query for free form incremental import: SELECT MAX('student_id') FROM 'student_info'
17/10/11 02:43:12 INFO tool.ImportTool: Incremental import based on column 'student_id'
17/10/11 02:43:12 INFO tool.ImportTool: Lower bound value: 7
17/10/11 02:43:12 INFO tool.ImportTool: Upper bound value: 8
17/10/11 02:43:12 WARN manager.MySQLManager: It looks like you are importing from mysql.
17/10/11 02:43:12 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
17/10/11 02:43:12 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
17/10/11 02:43:12 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
17/10/11 02:43:12 ERROR tool.ImportTool: Import failed: No primary key could be found for table student_info. Please specify one with --split-
by or perform a sequential import with --m 1.
cloudera@quickstart:~/Downloads$ sqoop import --connect jdbc:mysql://localhost/testDb --username=root --password=cloudera --table student_info
--target-dir /user/cloudera/acad student --incremental append --check-column student_id --last-value 7 -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.
17/10/11 02:43:40 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.12.0
17/10/11 02:43:40 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
17/10/11 02:43:41 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
17/10/11 02:43:41 INFO tool.CodeGenTool: Beginning code generation
17/10/11 02:43:42 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'student_info' AS t LIMIT 1
17/10/11 02:43:42 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'student_info' AS t LIMIT 1
17/10/11 02:43:42 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/cb664fc395399f004a1afd2269aae5fb/student_info.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
17/10/11 02:43:48 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/cb664fc395399f004a1afd2269aae5fb/student_info.jar
17/10/11 02:43:51 INFO tool.ImportTool: Maximal id query for free form incremental import: SELECT MAX('student_id') FROM 'student_info'
17/10/11 02:43:51 INFO tool.ImportTool: Incremental import based on column 'student_id'
17/10/11 02:43:51 INFO tool.ImportTool: Lower bound value: 7
17/10/11 02:43:51 INFO tool.ImportTool: Upper bound value: 8
17/10/11 02:43:51 WARN manager.MySQLManager: It looks like you are importing from mysql.
17/10/11 02:43:51 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
17/10/11 02:43:51 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
17/10/11 02:43:51 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
17/10/11 02:43:51 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
17/10/11 02:43:51 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
17/10/11 02:43:51 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
17/10/11 02:43:51 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
```

3. Once its complete, it should print incremental load complete!

```
Applications Places System cloudera@quickstart:~/Downloads Wed Oct 11, 02:52 cloudera

File Edit View Search Terminal Tabs Help
cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~
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)=17958
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=17958
  Total vcore-milliseconds taken by all map tasks=17958
  Total megabyte-milliseconds taken by all map tasks=18388992
Map-Reduce Framework
  Map input records=1
  Map output records=1
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=201
  CPU time spent (ms)=1500
  Physical memory (bytes) snapshot=116011008
  Virtual memory (bytes) snapshot=1510146048
  Total committed heap usage (bytes)=60751872
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=21
17/10/11 02:44:45 INFO mapreduce.ImportJobBase: Transferred 21 bytes in 54.0488 seconds (0.3885 bytes/sec)
17/10/11 02:44:45 INFO mapreduce.ImportJobBase: Retrieved 1 records.
17/10/11 02:44:45 INFO util.AppendUtils: Creating missing output directory - acad student
17/10/11 02:44:45 INFO tool.ImportTool: Incremental import complete! To run another incremental import of all data following this import, supply the following arguments:
17/10/11 02:44:45 INFO tool.ImportTool: --incremental append
17/10/11 02:44:45 INFO tool.ImportTool: --check-column student_id
17/10/11 02:44:45 INFO tool.ImportTool: --last-value 8
17/10/11 02:44:45 INFO tool.ImportTool: (Consider saving this with 'sqoop job --create')
[cloudera@quickstart Downloads]$
```