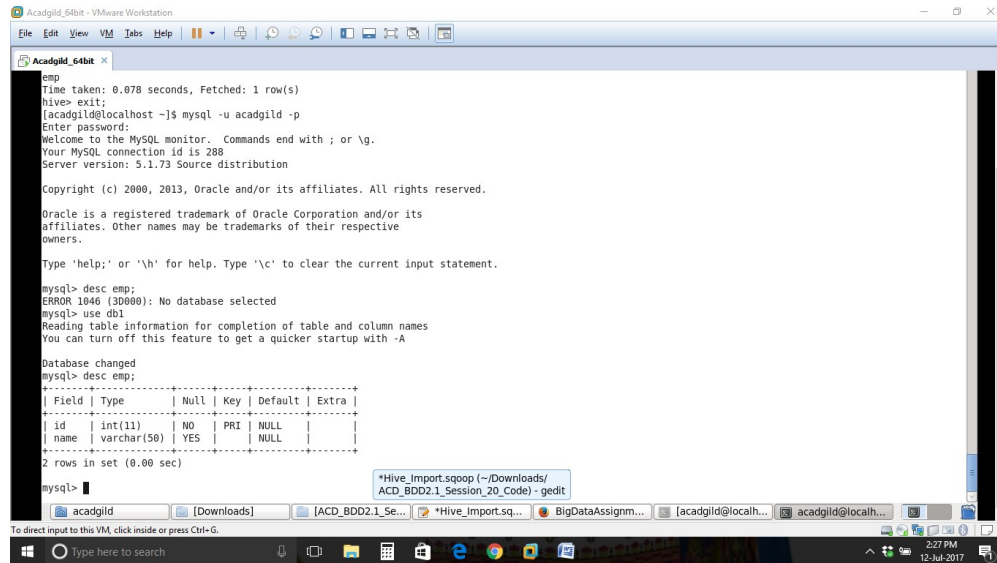


Assignment 20.2

Problem Statement

Perform incremental load in Hive

Table in MySQL



```
emp
Time taken: 0.078 seconds, Fetched: 1 row(s)
hive> exit;
[acadgild@localhost ~]$ mysql -u acadgild -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 288
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, 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.

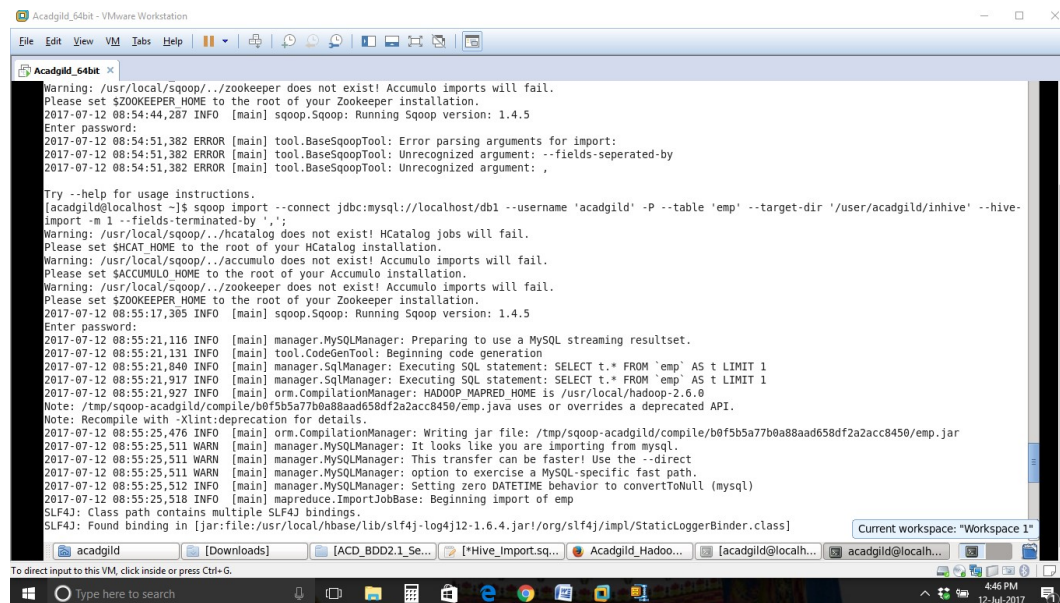
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> desc emp;
ERROR 1046 (3D000): No database selected
mysql> use db1
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> desc emp;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| id    | int(11) | NO   | PRI | NULL    |       |
| name  | varchar(50) | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
2 rows in set (0.00 sec)

mysql>
```

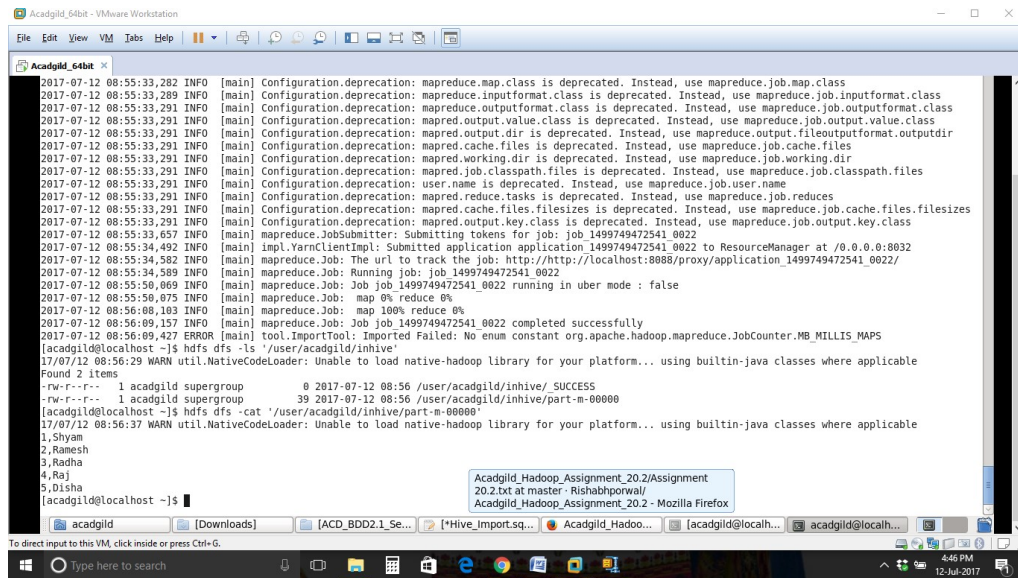
```
sqoop import --connect jdbc:mysql://localhost/sqoop --username 'root' -P --table 'emp' --target-dir
'/user/acadgild/inhive' --hive-import -m 1 --fields-terminated-by ',';
```



```
Warning: /usr/local/sqoop/./zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-07-12 08:54:44,287 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.5
Enter password:
2017-07-12 08:54:51,382 ERROR [main] tool.BaseSqoopTool: Error parsing arguments for import:
2017-07-12 08:54:51,382 ERROR [main] tool.BaseSqoopTool: Unrecognized argument: --fields-separated-by
2017-07-12 08:54:51,382 ERROR [main] tool.BaseSqoopTool: Unrecognized argument: ,

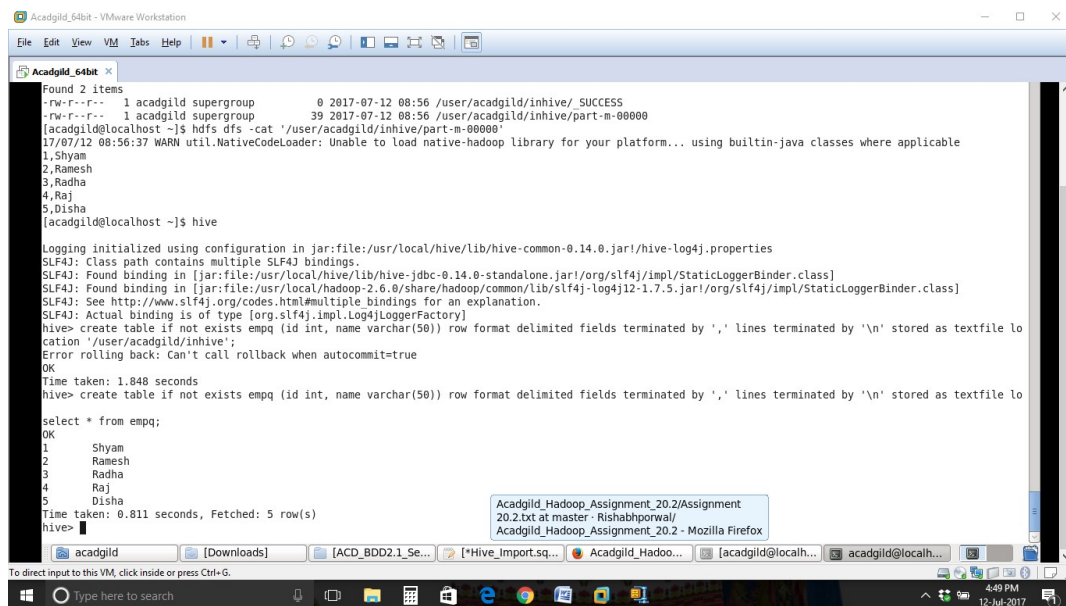
Try --help for usage instructions.
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/db1 --username 'acadgild' -P --table 'emp' --target-dir '/user/acadgild/inhive' --hive-
import -m 1 --fields-terminated-by ',';
Warning: /usr/local/sqoop/./hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/local/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /usr/local/sqoop/./zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-07-12 08:55:17,365 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.5
Enter password:
2017-07-12 08:55:21,116 INFO [main] manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2017-07-12 08:55:21,131 INFO [main] tool.CodeGenTool: Beginning code generation
2017-07-12 08:55:21,840 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'emp' AS t LIMIT 1
2017-07-12 08:55:21,917 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'emp' AS t LIMIT 1
2017-07-12 08:55:21,927 INFO [main] orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/hadoop-2.6.0
Note: /tmp/sqoop-acadgild/compile/b0f5b5a77b0a88aad658df2a2acc8450/emp.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
2017-07-12 08:55:25,476 INFO [main] orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/b0f5b5a77b0a88aad658df2a2acc8450/emp.jar
2017-07-12 08:55:25,511 WARN [main] manager.MySQLManager: It looks like you are importing from mysql.
2017-07-12 08:55:25,511 WARN [main] manager.MySQLManager: This transfer can be faster! Use the --direct
2017-07-12 08:55:25,511 INFO [main] manager.MySQLManager: option to exercise a MySQL-specific fast path.
2017-07-12 08:55:25,512 INFO [main] manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
2017-07-12 08:55:25,518 INFO [main] mapreduce.ImportJobBase: Beginning import of emp
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]
Current workspace: "Workspace 1"
```

Data imported to hdfs



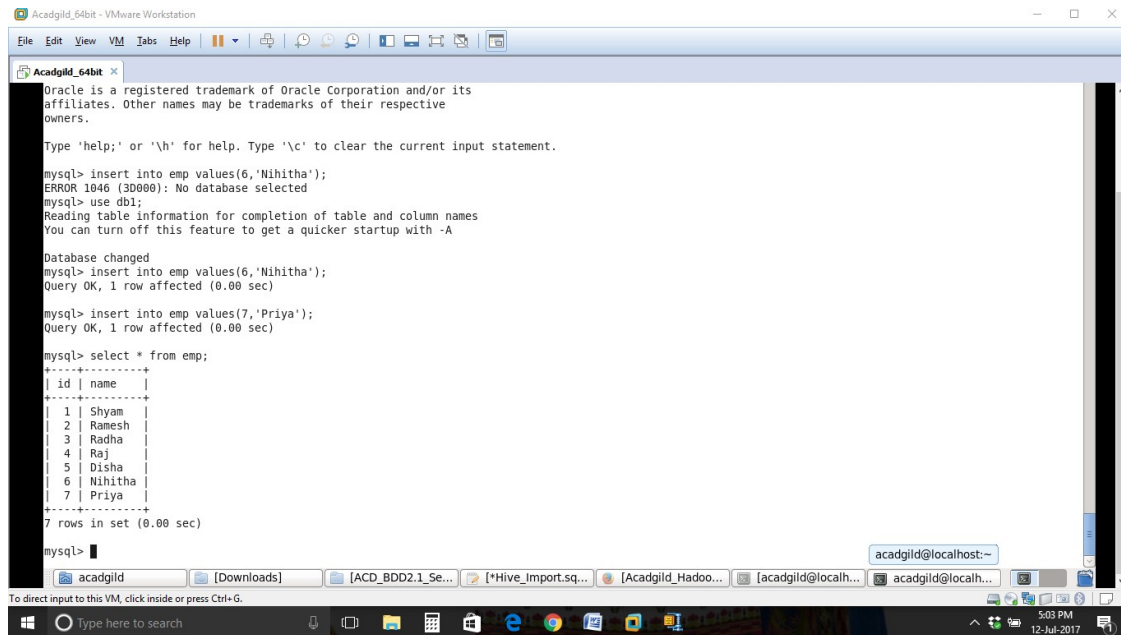
```
2017-07-12 08:55:33,282 INFO [main] Configuration.deprecation: mapreduce.map.class is deprecated. Instead, use mapreduce.job.map.class
2017-07-12 08:55:33,289 INFO [main] Configuration.deprecation: mapreduce.inputformat.class is deprecated. Instead, use mapreduce.job.inputformat.class
2017-07-12 08:55:33,291 INFO [main] Configuration.deprecation: mapreduce.outputformat.class is deprecated. Instead, use mapreduce.job.outputformat.class
2017-07-12 08:55:33,291 INFO [main] Configuration.deprecation: mapred.output.value.class is deprecated. Instead, use mapreduce.job.output.value.class
2017-07-12 08:55:33,291 INFO [main] Configuration.deprecation: mapred.output.dir is deprecated. Instead, use mapreduce.output.fileoutputformat.outputdir
2017-07-12 08:55:33,291 INFO [main] Configuration.deprecation: mapred.cache.files is deprecated. Instead, use mapreduce.job.cache.files
2017-07-12 08:55:33,291 INFO [main] Configuration.deprecation: mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir
2017-07-12 08:55:33,291 INFO [main] Configuration.deprecation: mapred.job.classpath.files is deprecated. Instead, use mapreduce.job.classpath.files
2017-07-12 08:55:33,291 INFO [main] Configuration.deprecation: user.name is deprecated. Instead, use mapreduce.job.user.name
2017-07-12 08:55:33,291 INFO [main] Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2017-07-12 08:55:33,291 INFO [main] Configuration.deprecation: mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.files.filesizes
2017-07-12 08:55:33,291 INFO [main] Configuration.deprecation: mapred.output.key.class is deprecated. Instead, use mapreduce.job.output.key.class
2017-07-12 08:55:33,657 INFO [main] mapreduce.JobSubmitter: Submitting tokens for job: job 1499749472541 0022
2017-07-12 08:55:34,492 INFO [main] impl.YarnClientImpl: Submitted application application 1499749472541 0022 to ResourceManager at /0.0.0.0:8032
2017-07-12 08:55:34,582 INFO [main] mapreduce.Job: The url to track the job: http://http://localhost:8088/proxy/application_1499749472541_0022/
2017-07-12 08:55:34,589 INFO [main] mapreduce.Job: Running job: job 1499749472541 0022
2017-07-12 08:55:50,069 INFO [main] mapreduce.Job: Job job 1499749472541 0022 running in uber mode : false
2017-07-12 08:55:50,075 INFO [main] mapreduce.Job: map 0% reduce 0%
2017-07-12 08:56:08,103 INFO [main] mapreduce.Job: map 100% reduce 0%
2017-07-12 08:56:09,157 INFO [main] mapreduce.Job: Job job 1499749472541 0022 completed successfully
2017-07-12 08:56:09,427 ERROR [main] tool.ImportTool: Imported Failed: No enum constant org.apache.hadoop.mapreduce.JobCounter.MB_MILLIS_MAPS
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/inhive
17/07/12 08:56:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2017-07-12 08:56 /user/acadgild/inhive/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 39 2017-07-12 08:56 /user/acadgild/inhive/part-m-00000
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/inhive/part-m-00000
1,Shyam
2,Ramesh
3,Radha
4,Raj
5,Disha
[acadgild@localhost ~]$
```

create table if not exists empq (id int, name varchar(50)) row format delimited fields terminated by ','
lines terminated by '\n' stored as textfile location '/user/acadgild/inhive';



```
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2017-07-12 08:56 /user/acadgild/inhive/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 39 2017-07-12 08:56 /user/acadgild/inhive/part-m-00000
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/inhive/part-m-00000
1,Shyam
2,Ramesh
3,Radha
4,Raj
5,Disha
[acadgild@localhost ~]$ hive
Logging initialized using configuration in jar:file:/usr/local/hive/lib/hive-common-0.14.0.jar!/hive-log4j.properties
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/hive-jdbc-0.14.0-standalone.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
hive> create table if not exists empq (id int, name varchar(50)) row format delimited fields terminated by ',' lines terminated by '\n' stored as textfile to
location '/user/acadgild/inhive';
Error rolling back: Can't call rollback when autocommit=true
OK
Time taken: 1.848 seconds
hive> create table if not exists empq (id int, name varchar(50)) row format delimited fields terminated by ',' lines terminated by '\n' stored as textfile to
select * from empq;
OK
1 Shyam
2 Ramesh
3 Radha
4 Raj
5 Disha
Time taken: 0.811 seconds, Fetched: 5 row(s)
hive>
```

Two rows inserted in mysql table



The screenshot shows a MySQL terminal window with the following commands and output:

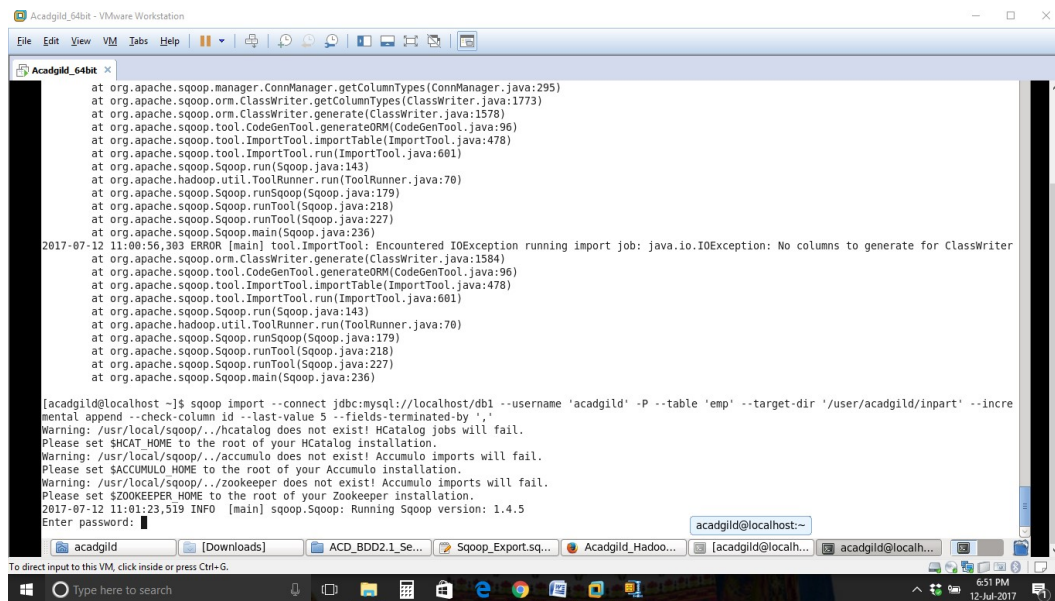
```
mysql> insert into emp values(6,'Nihitha');
ERROR 1046 (3D000): No database selected
mysql> use db1;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> insert into emp values(6,'Nihitha');
Query OK, 1 row affected (0.00 sec)

mysql> insert into emp values(7,'Priya');
Query OK, 1 row affected (0.00 sec)

mysql> select * from emp;
+----+-----+
| id | name  |
+----+-----+
| 1  | Shyam |
| 2  | Ramesh|
| 3  | Radha |
| 4  | Raj   |
| 5  | Disha |
| 6  | Nihitha|
| 7  | Priya |
+----+-----+
7 rows in set (0.00 sec)
```

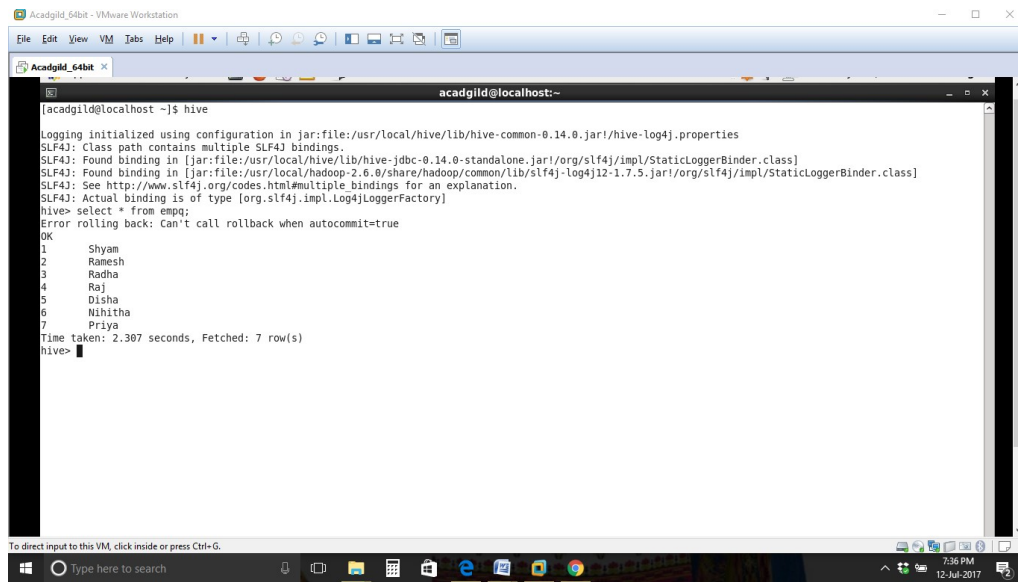
sqoop import --connect jdbc:mysql://localhost/db1 --username 'root' -P --table 'locations' --target-dir '/user/acadgild/inpart' --incremental append --check-column id --last-value 5 --fields-terminated-by ',';



The screenshot shows a terminal window with the following commands and output:

```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/db1 --username 'acadgild' -P --table 'emp' --target-dir '/user/acadgild/inpart' --incremental append --check-column id --last-value 5 --fields-terminated-by ','
Warning: /usr/local/sqoop/.hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/local/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /usr/local/sqoop/./zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-07-12 11:01:23,519 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.5
Enter password: 
```

Data appended to hive



The screenshot shows a terminal window titled 'Acadgild_64bit' running Hive. The user enters the command 'hive' at the prompt 'acadgild@localhost:~'. The terminal displays several SLF4J logging messages, followed by the execution of the query 'select * from empq;'. The output shows 7 rows of employee data. The terminal also displays the time taken for the query and the number of rows fetched.

```
acadgild@localhost:~$ hive
Logging initialized using configuration in jar:file:/usr/local/hive/lib/hive-common-0.14.0.jar!/hive-log4j.properties
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/hive-jdbc-0.14.0-standalone.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
hive> select * from empq;
Error rolling back: Can't call rollback when autocommit=true
OK
1      Shyam
2      Ramesh
3      Radha
4      Raj
5      Disha
6      Nihitha
7      Priya
Time taken: 2.307 seconds, Fetched: 7 row(s)
hive>
```