Big Data And Hadoop

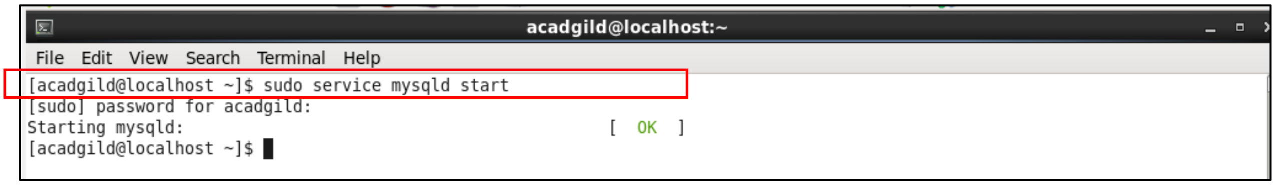
Session 20 – Assignment 2

Problem Statement:

Perform incremental load in Hive
Read from MySQL Table and load it in Hive table.
Create hive table if it does not exist.
If it exists, perform the incremental load.

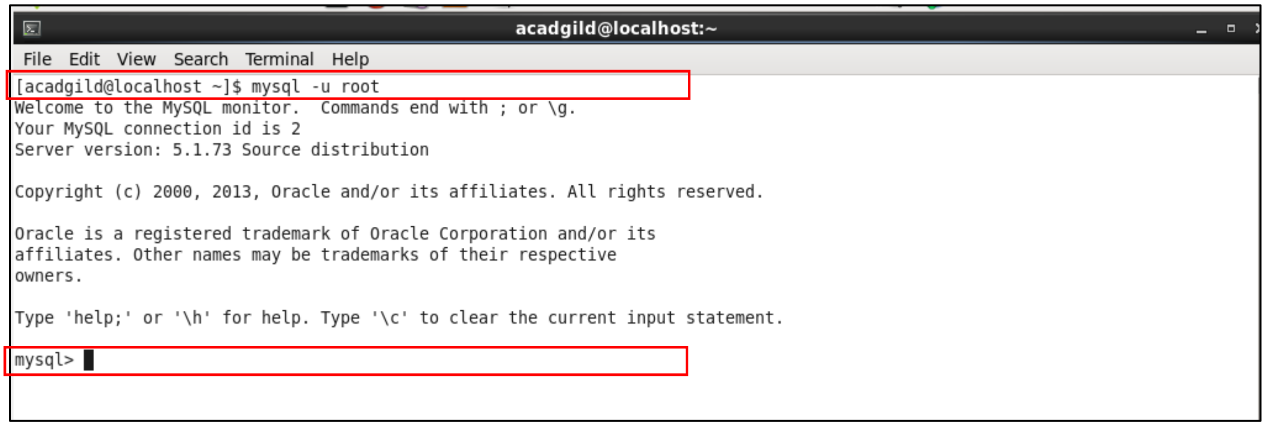
Solution:

1. We first start mysql service using the command **sudo service mysqld start** as follows:



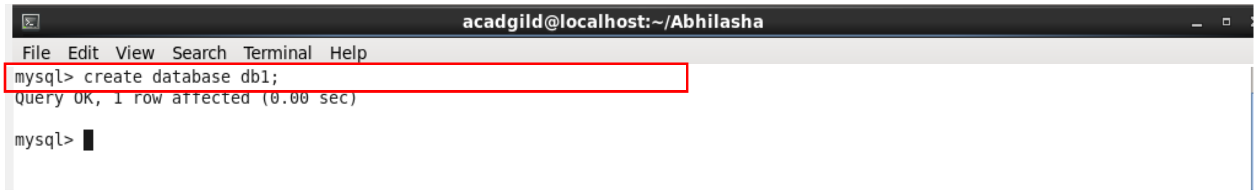
```
[acadgild@localhost ~]$ sudo service mysqld start
[sudo] password for acadgild:
Starting mysqld:                                           [  OK  ]
[acadgild@localhost ~]$ 
```

2. Next, we start the command line interface for mysql as follows:



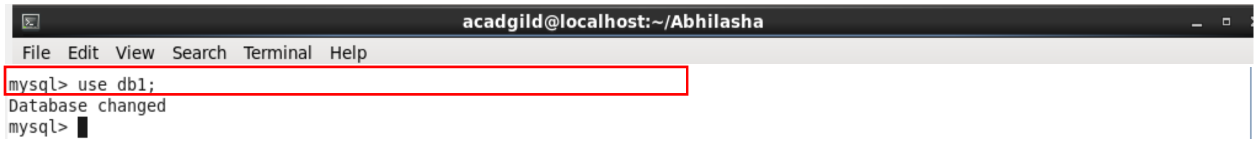
```
[acadgild@localhost ~]$ mysql -u root
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 2
Server version: 5.1.73 Source distribution

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> 
```

3. We will create the table in a database named **db1.** So we will first create the database using the command **create database db1;** as follows:


acadgild@localhost:~/Abhilasha
File Edit View Search Terminal Help

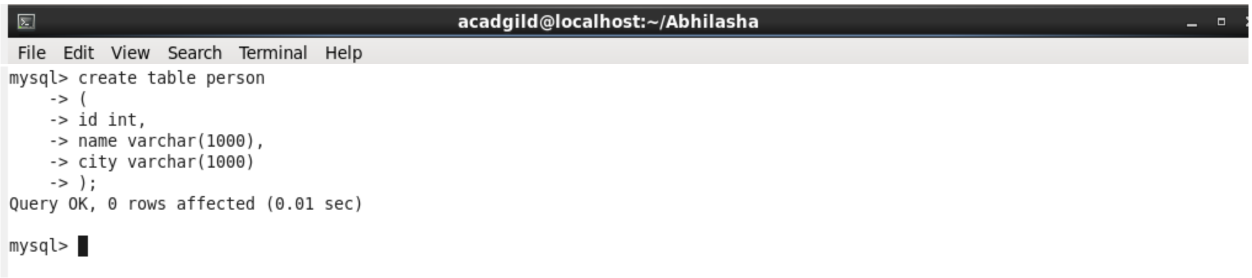
```
mysql> create database db1;
Query OK, 1 row affected (0.00 sec)

mysql> 
```

4. We specify which database to work in using the command **use db1;**


acadgild@localhost:~/Abhilasha
File Edit View Search Terminal Help

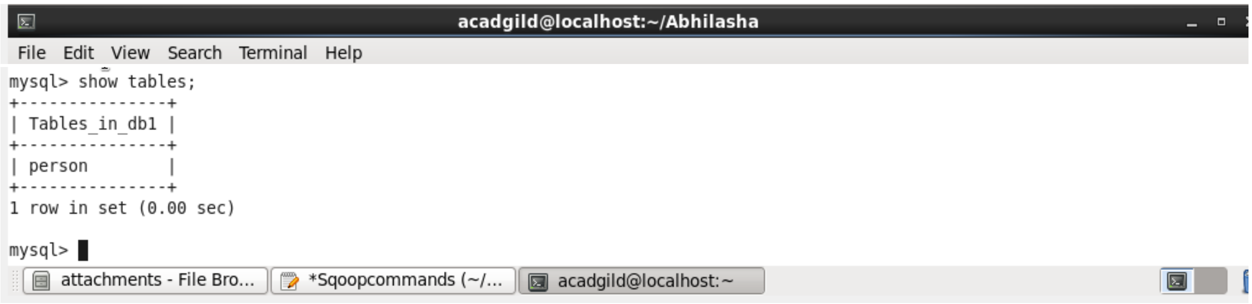
```
mysql> use db1;
Database changed
mysql> 
```

5. We now create the table named **person** in MySQL as follows:


acadgild@localhost:~/Abhilasha
File Edit View Search Terminal Help

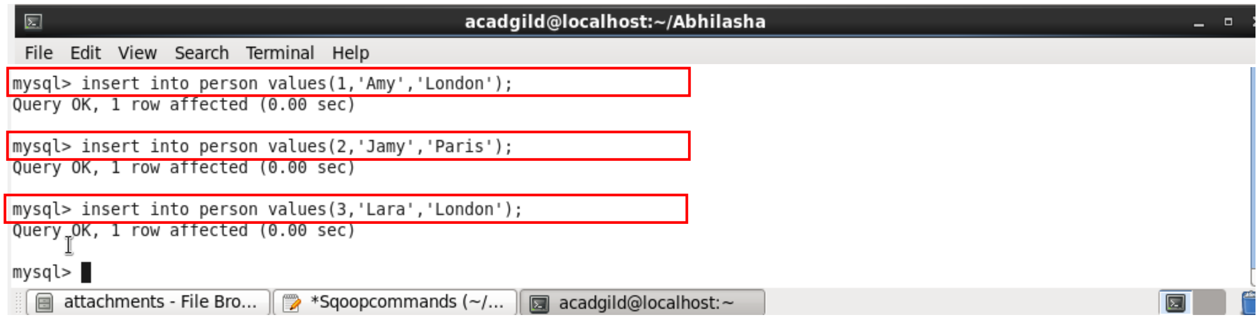
```
mysql> create table person
    -> (
    -> id int,
    -> name varchar(1000),
    -> city varchar(1000)
    -> );
Query OK, 0 rows affected (0.01 sec)

mysql> 
```

6. Using **show tables** command, we see the table that we created is listed as follows:


acadgild@localhost:~/Abhilasha
File Edit View Search Terminal Help
attachments - File Bro...
*Sqoopcommands (~/...
acadgild@localhost:~

```
mysql> show tables;
+---------------+
| Tables_in_db1 |
+---------------+
| person        |
+---------------+
1 row in set (0.00 sec)

mysql> 
```

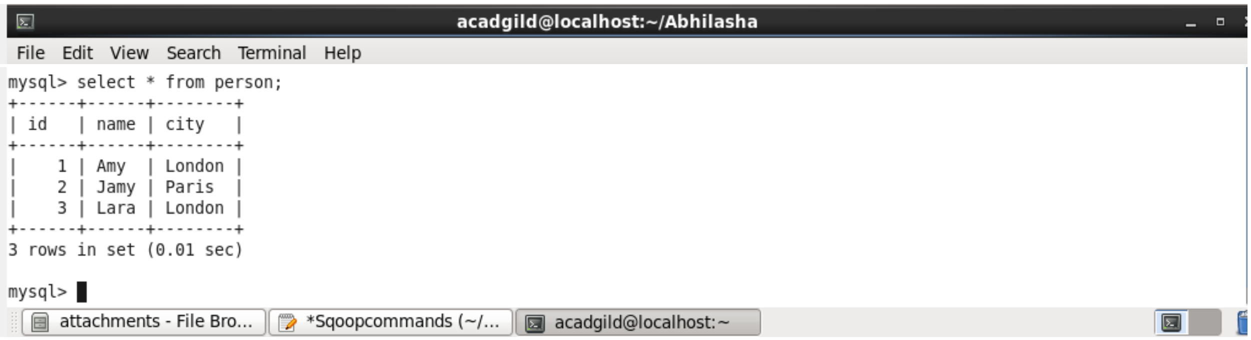
7. We now insert records in the table in MySQL as follows:


acadgild@localhost:~/Abhilasha
File Edit View Search Terminal Help
attachments - File Bro...
*Sqoopcommands (~/...
acadgild@localhost:~

```
mysql> insert into person values(1,'Amy','London');
Query OK, 1 row affected (0.00 sec)

mysql> insert into person values(2,'Jamy','Paris');
Query OK, 1 row affected (0.00 sec)

mysql> insert into person values(3,'Lara','London');
Query OK, 1 row affected (0.00 sec)

mysql> 
```

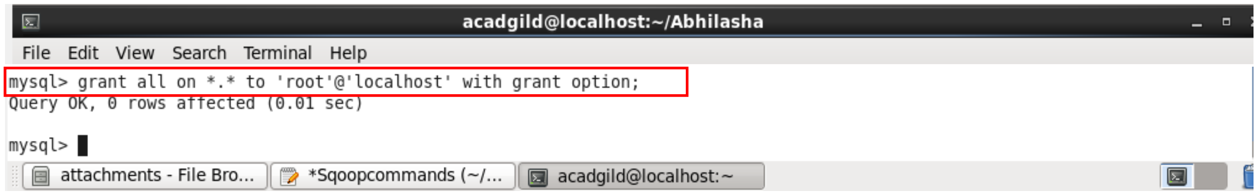
We have inserted three records here.

8. The inserted records can be seen using **select * from person;** query as follows:


acadgild@localhost:~/Abhilasha
File Edit View Search Terminal Help
attachments - File Bro...
*Sqoopcommands (~/...
acadgild@localhost:~

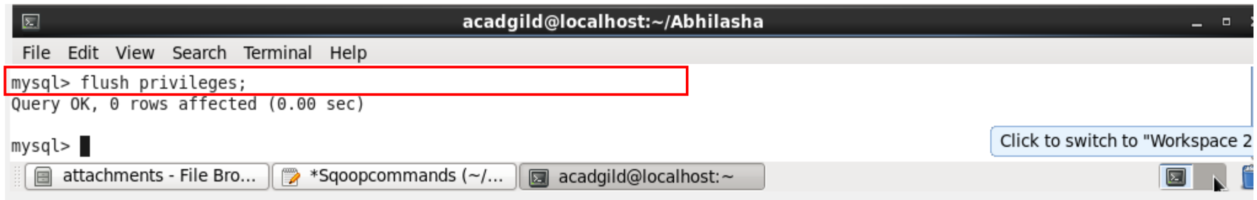
```
mysql> select * from person;
+------+------+--------+
| id   | name | city   |
+------+------+--------+
|    1 | Amy  | London |
|    2 | Jamy | Paris  |
|    3 | Lara | London |
+------+------+--------+
3 rows in set (0.01 sec)

mysql> 
```

9. We also grant access to root user to that import/export is error free. The command used is as follows:


acadgild@localhost:~/Abhilasha
File Edit View Search Terminal Help
attachments - File Bro...
*Sqoopcommands (~/...
acadgild@localhost:~

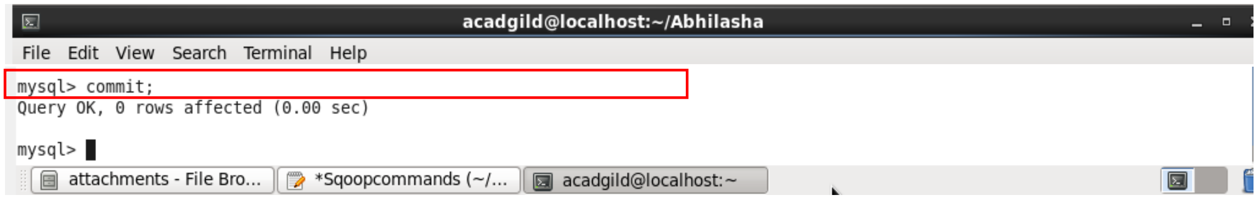
```
mysql> grant all on *.* to 'root'@'localhost' with grant option;
Query OK, 0 rows affected (0.01 sec)

mysql> 
```

10. We then flush privileges as follows:


acadgild@localhost:~/Abhilasha
File Edit View Search Terminal Help
Click to switch to "Workspace 2
attachments - File Bro...
*Sqoopcommands (~/...
acadgild@localhost:~

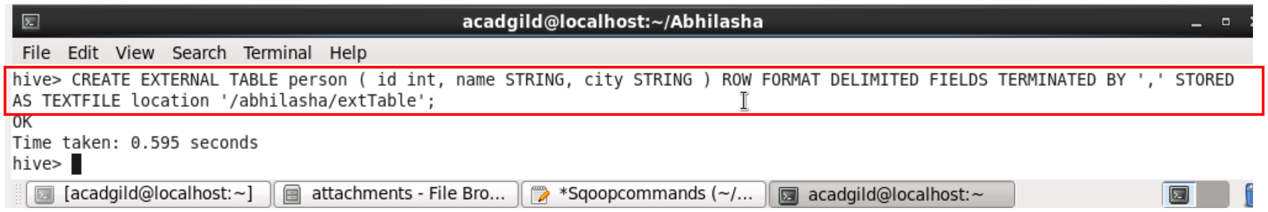
```
mysql> flush privileges;
Query OK, 0 rows affected (0.00 sec)

mysql> 
```

11. We now commit the data inserted as follows:


acadgild@localhost:~/Abhilasha
File Edit View Search Terminal Help
attachments - File Bro...
*Sqoopcommands (~/...
acadgild@localhost:~

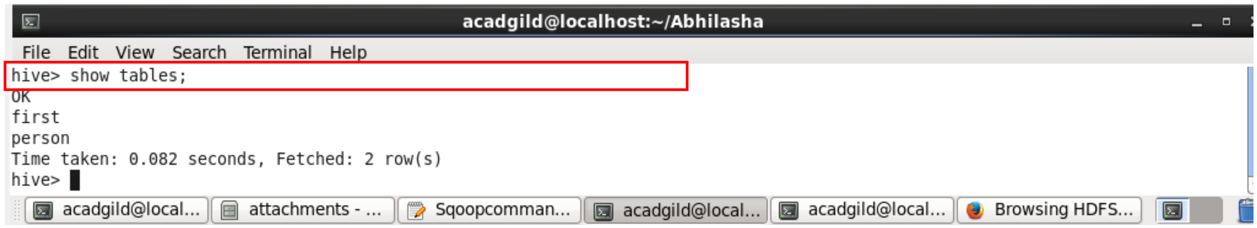
```
mysql> commit;
Query OK, 0 rows affected (0.00 sec)

mysql> 
```

12. Now, inorder to import data into hive, incrementally, we need to use incremental hdfs import. We are not directly using hive incremental import as it is not supported. Hence, this work around. We first need to create external table in hive. We are using external table as it will read data from a location in HDFS. This is the shared location where data from MySQL will be stored.
Query to create external table is as follows:


acadgild@localhost:~/Abhilasha
File Edit View Search Terminal Help
[acadgild@localhost:~]
attachments - File Bro...
*Sqoopcommands (~/...
acadgild@localhost:~

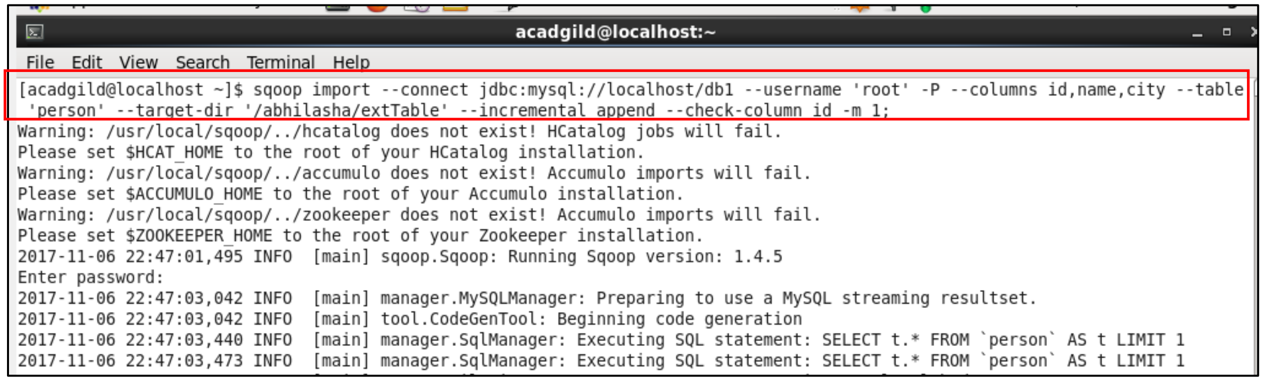
```
hive> CREATE EXTERNAL TABLE person ( id int, name STRING, city STRING ) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED
AS TEXTFILE location '/abhilasha/extTable';
OK
Time taken: 0.595 seconds
hive>
```

13. The created table can be listed using the command **show tables;** as follows:



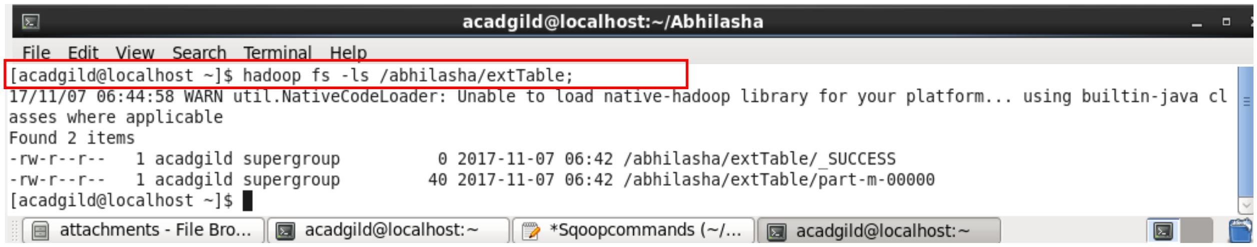
```
hive> show tables;
OK
first
person
Time taken: 0.082 seconds, Fetched: 2 row(s)
hive>
```

14. Now we perform an import to place the data from MySQL to hive using incremental HDFS import as follows. Here, we mention the path on HDFS to be **/abhilasha/extTable.** We also have mentioned **incremental** to be **append.** Also, the column on which check will be performed is **id**.



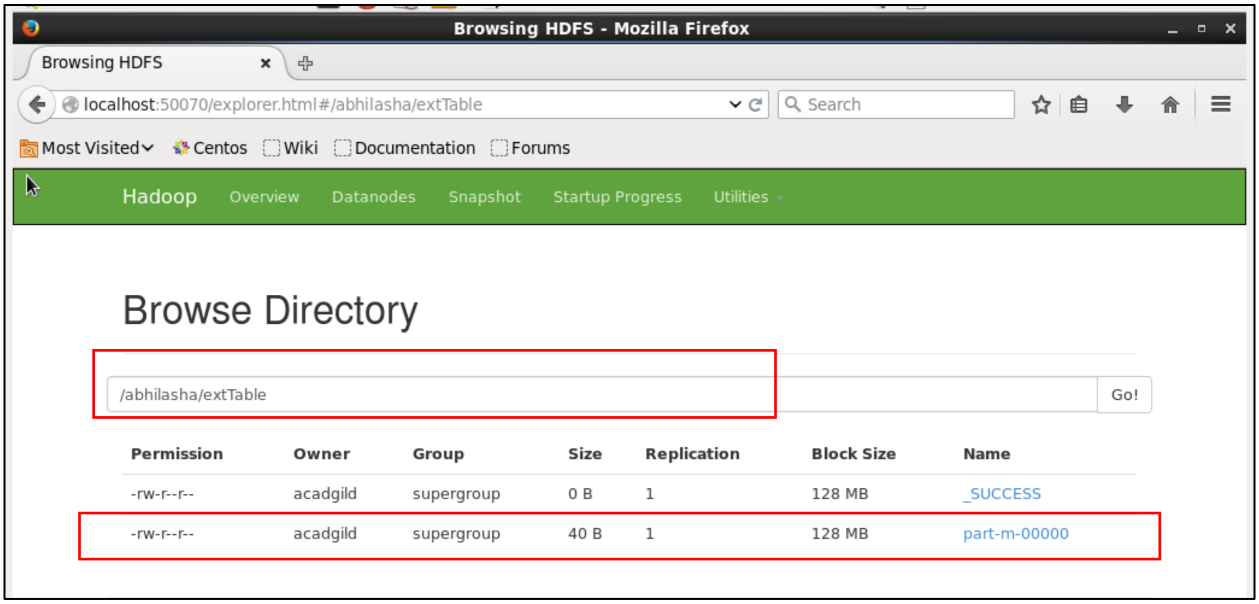
```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/db1 --username 'root' -P --columns id,name,city --table
 'person' --target-dir '/abhilasha/extTable' --incremental append --check-column id -m 1;
Warning: /usr/local/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please 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.
2017-11-06 22:47:01,495 INFO  [main] sqoop.Sqoop: Running Sqoop version: 1.4.5
Enter password:
2017-11-06 22:47:03,042 INFO  [main] manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2017-11-06 22:47:03,042 INFO  [main] tool.CodeGenTool: Beginning code generation
2017-11-06 22:47:03,440 INFO  [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM `person` AS t LIMIT 1
2017-11-06 22:47:03,473 INFO  [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM `person` AS t LIMIT 1
```

15. We will see the file created at the location mentioned above using **ls** command as follows:


acadgild@localhost:~/Abhilasha
File Edit View Search Terminal Help
attachments - File Bro...
acadgild@localhost:~
*Sqoopcommands (~/...
acadgild@localhost:~

```
[acadgild@localhost ~]$ hadoop fs -ls /abhilasha/extTable;
17/11/07 06:44:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2017-11-07 06:42 /abhilasha/extTable/_SUCCESS
-rw-r--r--   1 acadgild supergroup         40 2017-11-07 06:42 /abhilasha/extTable/part-m-00000
[acadgild@localhost ~]$
```

16. The file can also be seen through HDFS UI as follows:

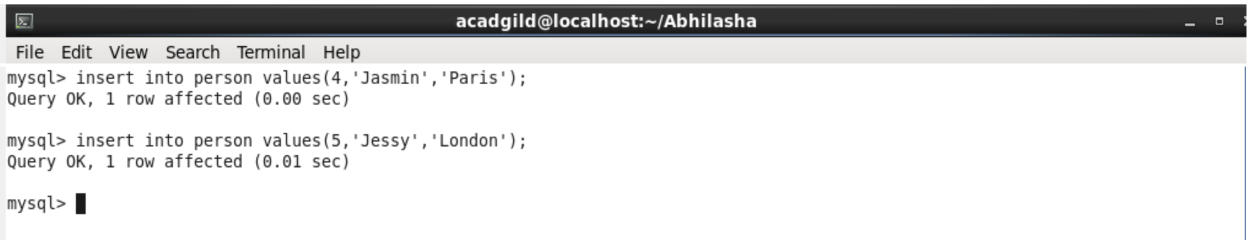


Browse Directory

/abhilasha/extTable

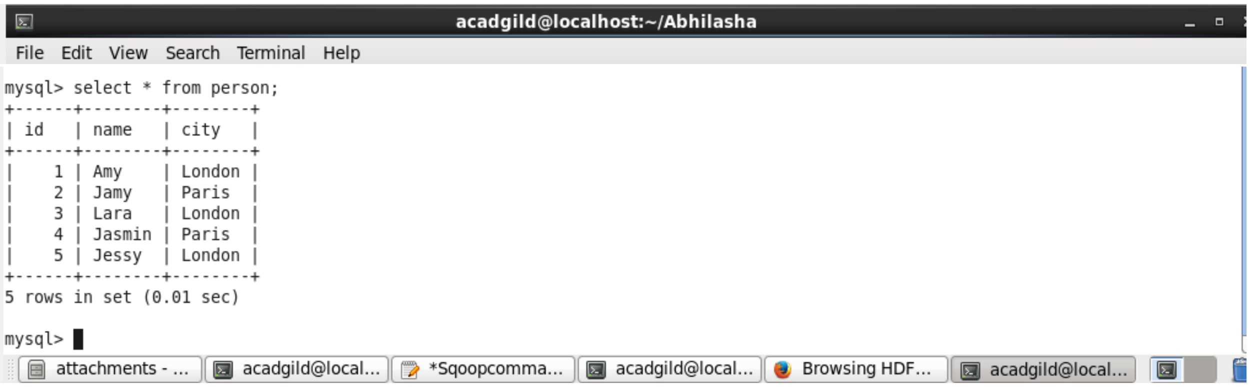
Permission	Owner	Group	Size	Replication	Block Size	Name
-rw-r--r--	acadgild	supergroup	0 B	1	128 MB	_SUCCESS
-rw-r--r--	acadgild	supergroup	40 B	1	128 MB	part-m-00000

17. To test incremental import, we add some more records in MySQL as follows:


acadgild@localhost:~/Abhilasha
File Edit View Search Terminal Help

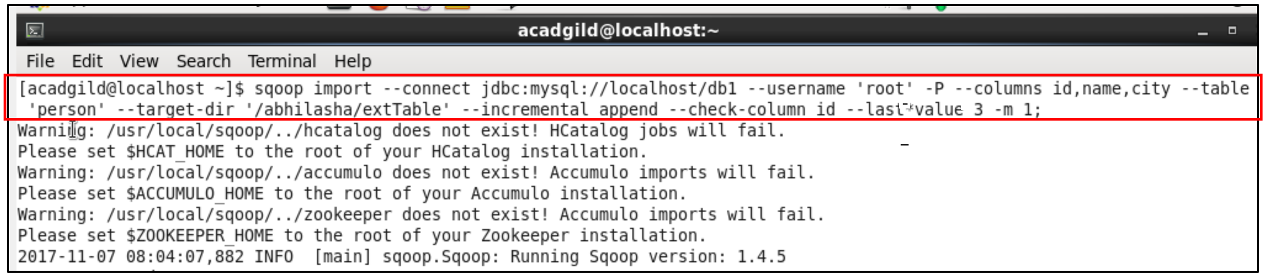
```
mysql> insert into person values(4,'Jasmin','Paris');
Query OK, 1 row affected (0.00 sec)

mysql> insert into person values(5,'Jessy','London');
Query OK, 1 row affected (0.01 sec)

mysql> 
```

18. The table in MySQL now has 5 records in all as follows:



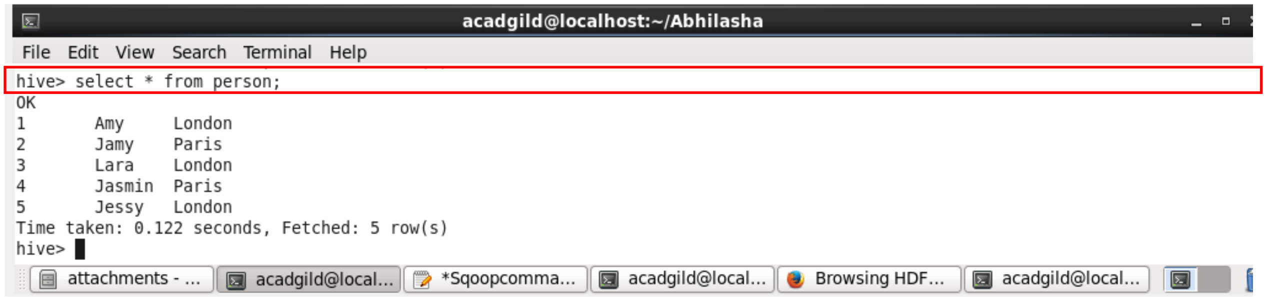
```
mysql> select * from person;
+------+--------+--------+
| id   | name   | city   |
+------+--------+--------+
|    1 | Amy    | London |
|    2 | Jamy   | Paris  |
|    3 | Lara   | London |
|    4 | Jasmin | Paris  |
|    5 | Jessy  | London |
+------+--------+--------+
5 rows in set (0.01 sec)

mysql> 
```

19. To perform incremental import, we give a predicate on the check-column (i.e., id) that value of last id is 3 and run the import as follows:



```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/db1 --username 'root' -P --columns id,name,city --table
 'person' --target-dir '/abhilasha/extTable' --incremental append --check-column id --last-value 3 -m 1;
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.
2017-11-07 08:04:07,882 INFO  [main] sqoop.Sqoop: Running Sqoop version: 1.4.5
```

20. The output can be seen in hive using **select * from person;** command as follows:



```
hive> select * from person;
OK
1       Amy     London
2       Jamy    Paris
3       Lara    London
4       Jasmin  Paris
5       Jessy   London
Time taken: 0.122 seconds, Fetched: 5 row(s)
hive> 
```

Here, we see that all five records are in place. None of the records are repeated.