

Assignment 11.2

1. Created a dataset employee.txt

```
[cloudera@quickstart ~]$ cat >employee.txt
101,Amitab,256,Finance,1
102,Shahrukh,78,IT_Dept,2
103,Akshay,110,HR,3
104,Anubhav,50,Network_Team,4
105,Pawan,250,Admin,5
106,Aamir,25,Finance,1
107,Salman,175,IT_Dept,2
108,Ranbir,142,HR,3
109,Katrina,100,Network team,4
110,Priyanka,222,Admin,5
```

2. Started MySQL commands and enter the MySQL shell. Created the database **sumona**

```
[cloudera@quickstart ~]$ mysql -uroot -pcloudera
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 21
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> show databases;
+-----+
| Database |
+-----+
| information_schema |
| cm |
| firehose |
| hue |
| metastore |
| mysql |
| nav |
| navms |
| oozie |
| retail_db |
| rman |
| sentry |
+-----+
12 rows in set (0.02 sec)

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

mysql> use sumona;
Database changed
```

3. Created the table **employee** and loaded the values into MySQL

```
mysql> create table employee(
-> emp_id int,
-> emp_name varchar (20),
-> emp_salary int,
-> emp_department varchar (20),
-> unit int
-> );
Query OK, 0 rows affected (0.02 sec)

mysql> LOAD DATA INFILE '/home/cloudera/employee.txt' INTO TABLE employee COLUMN
S TERMINATED BY ',';
Query OK, 10 rows affected, 2 warnings (0.01 sec)
Records: 10 Deleted: 0 Skipped: 0 Warnings: 1

mysql> select * from Employee;
ERROR 1146 (42S02): Table 'sumona.Employee' doesn't exist
mysql> select * from employee;
+-----+-----+-----+-----+-----+
| emp_id | emp_name | emp_salary | emp_department | unit |
+-----+-----+-----+-----+-----+
| 101 | Amitab.256 | 0 | 1 | NULL |
| 102 | Shahrukh | 78 | IT_Dept | 2 |
| 103 | Akshay | 110 | HR | 3 |
| 104 | Anubhav | 50 | Network_Team | 4 |
| 105 | Pawan | 250 | Admin | 5 |
| 106 | Aamir | 25 | Finance | 1 |
| 107 | Salman | 175 | IT_Dept | 2 |
| 108 | Ranbir | 142 | HR | 3 |
| 109 | Katrina | 100 | Network_team | 4 |
| 110 | Priyanka | 222 | Admin | 5 |
+-----+-----+-----+-----+-----+
10 rows in set (0.00 sec)
```

4. Since the data is present in table of MySQL and Sqoop is up and running, import the data into HIVE using the command

```
sqoop import --connect jdbc:mysql://localhost/sumona --username root --password cloudera --
table employee -m1 --target-dir /sqoopouthive --hive-import;
```

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/sumona --
username root --password cloudera --table employee -m1 --target-dir /sqoopouthiv
e --hive-import;
```

```
17/12/04 01:33:27 INFO hive.HiveImport: Loading uploaded data into Hive
Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-commo
n-1.1.0-cdh5.12.0.jar!/hive-log4j.properties
OK
Time taken: 2.742 seconds
Loading data to table default.employee
Table default.employee stats: [numFiles=5, numRows=0, totalSize=1313, rawDataSiz
e=0]
OK
Time taken: 0.886 seconds
[cloudera@quickstart ~]$
```

Let us now check the table in HIVE.

```
hive> select * from employee;
OK
101    Amitab.256      0      1      NULL
102    Shahrukh       78     IT_Dept 2
103    Akshay   110     HR      3
104    Anubhav   50     Network_Team 4
105    Pawan    250     Admin   5
106    Aamir    25      Finance 1
107    Salman   175     IT_Dept 2
108    Ranbir   142     HR      3
109    Katrina  100     Network_team 4
110    Priyanka  222     Admin   5
Time taken: 0.071 seconds, Fetched: 10 row(s)
hive>
```

5. Inserted new values to the table employee.

```
mysql> insert into employee values (111,'Amit',256,'Network_team',4);
Query OK, 1 row affected (0.01 sec)

mysql> insert into employee values (112,'Sumona',118,'Admin',5);
Query OK, 1 row affected (0.00 sec)

mysql> insert into employee values (113,'Hardik',200,'HR',3);
Query OK, 1 row affected (0.00 sec)

mysql> select * from employee;
+-----+-----+-----+-----+-----+
| emp_id | emp_name | emp_salary | emp_department | unit |
+-----+-----+-----+-----+-----+
| 101 | Amitab.256 | 0 | 1 | NULL |
| 102 | Shahrukh | 78 | IT_Dept | 2 |
| 103 | Akshay | 110 | HR | 3 |
| 104 | Anubhav | 50 | Network_Team | 4 |
| 105 | Pawan | 250 | Admin | 5 |
| 106 | Aamir | 25 | Finance | 1 |
| 107 | Salman | 175 | IT_Dept | 2 |
| 108 | Ranbir | 142 | HR | 3 |
| 109 | Katrina | 100 | Network_team | 4 |
| 110 | Priyanka | 222 | Admin | 5 |
| 111 | Amit | 256 | Network_team | 4 |
| 112 | Sumona | 118 | Admin | 5 |
| 113 | Hardik | 200 | HR | 3 |
+-----+-----+-----+-----+-----+
13 rows in set (0.00 sec)

mysql>
```

6. After inserting the new values in the table in MySQL shell, we shall import the updated values into HIVE by using the INCREMENTAL command

```
sqoop import --connect jdbc:mysql://localhost/sumona --username root -password cloudera --table employee --check-column emp_id --incremental append --last-value 110 -m1 --target-dir /sqoopouthive --hive-import;
```

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/sumona --username root --password cloudera --table employee --check-column emp_id --incremental append --last-value 110 -m1 --target-dir /sqoopouthive --hive-import;
```

In this command we are updating the new values by comparing the existing values of the table with the column emp_id.

```
17/12/04 01:49:38 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: 2.612 seconds
Loading data to table default.employee
Table default.employee stats: [numFiles=2, totalSize=317]
OK
Time taken: 0.872 seconds
[cloudera@quickstart ~]$
```

Now as you can see the new values have been updated into the HIVE table.

```
hive> select * from employee;
OK
101    Amitab.256      0      1      NULL
102    Shahrukh      78      IT_Dept 2
103    Akshay 110      HR      3
104    Anubhav 50      Network_Team 4
105    Pawan 250      Admin 5
106    Aamir 25      Finance 1
107    Salman 175      IT_Dept 2
108    Ranbir 142      HR      3
109    Katrina 100      Network_team 4
110    Priyanka 222      Admin 5
111    Amit 256      Network_team 4
112    Sumona 118      Admin 5
113    Hardik 200      HR      3
Time taken: 0.109 seconds, Fetched: 13 row(s)
hive>
```