Assignment 11.1

1. Transfer data between MySQL and HDFS (Import and export) using Sqoop.

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
```

Created a table in 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 (0.03 sec)
```

Loaded the data from local file to MySQL

```
mysql> LOAD DATA INFILE '/hoe/cloudera/employee.txt' INTO TABLE employee COLUMNS
 TERMINATED BY ',';
ERROR 13 (HY000): Can't get stat of '/hoe/cloudera/employee.txt' (Errcode: 2)
mysql> LOAD DATA INFILE '/home/cloudera/employee.txt' INTO TABLE employee COLUMN
S TERMINATED BY ',';
Query OK, 10 rows affected, 2 warnings (0.00 sec)
Records: 10  Deleted: 0  Skipped: 0  Warnings: 1

mysql> select * from employee;
+--------+------------+------------+----------------+------+
| emp_id | emp_name   | emp_salary | emp_department | unit |
+--------+------------+------------+----------------+------+
|    101 | Amitabh    |        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 |
|    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 |
+--------+------------+------------+----------------+------+
20 rows in set (0.00 sec)

mysql>
```

Now we transfer data from MySQL to HDFS using sqoop command:

Sqoop import –connect jdbc:mysql:localhost/Sumona –username root –password cloudera –table employee –m1 –taget-dir /sqoopout1

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/sumona --
username root --password cloudera --table employee -m1 --target-dir /sqoopout1
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
```

```
17/12/02 03:54:44 INFO mapreduce.ImportJobBase: Transferred 248 bytes in 39.5855
 seconds (6.2649 bytes/sec)
17/12/02 03:54:44 INFO mapreduce.ImportJobBase: Retrieved 10 records.
[cloudera@quickstart ~]$
```

```
[cloudera@quickstart ~]$ hadoop fs -ls /sqoopout1
Found 2 items
-rw-r--r--   1 cloudera supergroup          0 2017-12-02 03:54 /sqoopout1/_SUCCE
SS
-rw-r--r--   1 cloudera supergroup        248 2017-12-02 03:54 /sqoopout1/part-m
-00000
[cloudera@quickstart ~]$ hadoop fs -cat /sqoopout1/part-m-00000
101,Amitabh,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
[cloudera@quickstart ~]$ 
```

Now we will drop the MySQL table and export the contents to the table using Sqoop

```
mysql> delete from employee;
Query OK, 10 rows affected (0.04 sec)

mysql> select * from employee;
Empty set (0.00 sec)

mysql> 
```

Sqoop export –connect jdbc:mysql://localhost/employee –username root –password cloudera –table employee –m1 –export-dir /sqooput –input-fields-terminated-by ',' –columns emp_id,emp_name,emp_salary,emp_department,unit

```
[cloudera@quickstart ~]$ sqoop export --connect jdbc:mysql://localhost/sumona --username root -password cloudera --ta
ble employee -m1 --export-dir /sqoopout --input-fields-terminated-by ',' --columns emp_id,emp_name,emp_salary,emp_dep
artment,unit
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
```

```
         bytes written=0
17/12/02 04:19:05 INFO mapreduce.ExportJobBase: Transferred 386 bytes in 23.2316 seconds (16.6153 bytes/sec)
17/12/02 04:19:05 INFO mapreduce.ExportJobBase: Exported 10 records.
[cloudera@quickstart ~]$
```

Now let us check the table in MySQL

```
mysql> select * from employee;
+--------+----------+------------+-----------------+------+
| emp_id | emp_name | emp_salary | emp_department  | unit |
+--------+----------+------------+-----------------+------+
|    101 | Amitabh  |        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 |
+--------+----------+------------+-----------------+------+
10 rows in set (0.00 sec)

mysql>
```

As we can see the Sqoop export has exported the values to MySQL from HDFS

2. Transfer data between MySQL and HIVE (import and export only selected columns) using sqoop

Taking the same dataset from the above.

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/sumona --username root --password cloudera --t
able employee -m1 --target-dir /sqoopout1 --hove-import;
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO HOME to the root of your Accumulo installation.
```

```
Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-1.1.0-cdh5.12.0.jar!/hive-log4j.pro
perties
OK
Time taken: 3.418 seconds
Loading data to table default.employee
Table default.employee stats: [numFiles=2, numRows=0, totalSize=274, rawDataSize=0]
OK
Time taken: 3.194 seconds
[cloudera@quickstart ~]$ █
```

Now let us check the table in HIVE

```
hive> show databases;
OK
default
Time taken: 1.013 seconds, Fetched: 1 row(s)
hive> show databases;
OK
default
Time taken: 0.052 seconds, Fetched: 1 row(s)
hive> use default;
OK
Time taken: 0.093 seconds
hive> show tables;
OK
employee
Time taken: 0.049 seconds, Fetched: 1 row(s)
hive> select * from employee;
OK
101     Amitabh 256     Finance 1
101     Amitabh 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
Time taken: 1.027 seconds, Fetched: 11 row(s)
hive> █
```