Problem Statement - Transfer data between Mysql and Hive (Import and Export only selected columns) using Sqoop.

1. **IMPORT-** We will import first from MySQL to HIVE.
   In order to do that we will create a database db1 in MySQL first as shown below-

```
mysql> create database db1;
Query OK, 1 row affected (0.04 sec)

mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| db1                |
| hive               |
| mysql              |
| test               |
+--------------------+
5 rows in set (0.00 sec)

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

Now inside database db1 we will create a table emp with columns emp_id, emp_name, emp_sal and emp_rating as shown below-

```
mysql> CREATE TABLE emp
    -> (
    -> emp_id int,
    -> emp_name varchar(20),
    -> emp_sal int,
    -> emp_rating int
    -> );
Query OK, 0 rows affected (0.13 sec)

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

mysql>
```

Now after creating tables we will insert some values in it as shown below-

```
mysql> insert into emp values(101, 'Amitabh' ,20000,1);
Query OK, 1 row affected (0.07 sec)

mysql> insert into emp values(102, 'Shahrukh' ,10000,2);
Query OK, 1 row affected (0.00 sec)

mysql> insert into emp values(103, 'Akshay' ,11000,3);
Query OK, 1 row affected (0.00 sec)

mysql> insert into emp values(104, 'Anubhav' ,5000,4);
Query OK, 1 row affected (0.00 sec)

mysql> insert into emp values(105, 'Pawan' ,2500,5);
Query OK, 1 row affected (0.00 sec)

mysql> insert into emp values(106, 'Aamir' ,25000,1);
Query OK, 1 row affected (0.00 sec)

mysql> insert into emp values(107, 'Salman' ,17500,2);
Query OK, 1 row affected (0.00 sec)

mysql> insert into emp values(108, 'Ranbir' ,14000,3);
Query OK, 1 row affected (0.00 sec)

mysql> insert into emp values(109, 'Katrina' ,1000,4);
Query OK, 1 row affected (0.00 sec)

mysql> insert into emp values(110, 'Priyanka' ,2000,5);
Query OK, 1 row affected (0.01 sec)
```

Below screenshot shows the values which we have inserted above in the emp table-

```
mysql> select * from emp;
+--------+----------+---------+------------+
| emp_id | emp_name | emp_sal | emp_rating |
+--------+----------+---------+------------+
|    101 | Amitabh  |   20000 |          1 |
|    102 | Shahrukh |   10000 |          2 |
|    103 | Akshay   |   11000 |          3 |
|    104 | Anubhav  |    5000 |          4 |
|    105 | Pawan    |    2500 |          5 |
|    106 | Aamir    |   25000 |          1 |
|    107 | Salman   |   17500 |          2 |
|    108 | Ranbir   |   14000 |          3 |
|    109 | Katrina  |    1000 |          4 |
|    110 | Priyanka |    2000 |          5 |
+--------+----------+---------+------------+
10 rows in set (0.05 sec)
```

Now we will go inside hive shell and use the 'default' database. So at present below tables are available inside default database. There is no table named as emp present in Hive inside default database-

```
hive (default)> show tables;
OK
tab_name
sample_07
sample_08
Time taken: 1.634 seconds, Fetched: 2 row(s)
hive (default)>
```

Now we will use below sqoop script to import data from MySQL to HIVE-

- ➢ **sqoop import --connect jdbc:mysql://localhost/db1 **
- ➢ **--username 'root' -P --columns emp_id,emp_name,emp_sal --table 'emp' **
- ➢ **--target-dir '/myhive' --fields-terminated-by ',' **
- ➢ **--hive-import --create-hive-table --hive-table 'default.emp' **
- ➢ **-m 1**

Here in above script we are making a JDBC connection with MySQL with database 'db1'. We have specified username as root and –P determines that it will ask for password. The table which we are importing is 'emp' and we are extracting only 3 columns out of 4 columns from the emp table which are emp_id, emp_name and emp_sal. We have also specified to create a hive table on the go inside default database named as emp-

```
[root@sandbox ~]# sqoop import --connect jdbc:mysql://localhost/db1 \
> --username 'root' -P --columns emp_id,emp_name,emp_sal --table 'emp'  \
> --target-dir '/myhive' --fields-terminated-by ',' \
> --hive-import --create-hive-table --hive-table 'default.emp' \
> -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.
```

Now we can see after running above command that inside default database emp table has been created-

```
hive (default)> show tables;
OK
tab_name
emp
sample_07
sample_08
Time taken: 1.993 seconds, Fetched: 3 row(s)
```

If we do a select we can see that only 3 columns have been imported to HIVE from MySQL table-

```
hive (default)> select * from emp;
OK
emp.emp_id      emp.emp_name      emp.emp_sal
101      Amitabh 20000
102      Shahrukh        10000
103      Akshay  11000
104      Anubhav 5000
105      Pawan   2500
106      Aamir   25000
107      Salman  17500
108      Ranbir  14000
109      Katrina 1000
110      Priyanka        2000
Time taken: 2.037 seconds, Fetched: 10 row(s)
```

2. **Export-** Now we will try to export data from HIVE to MySQL.

In order to do that the table should be present inside MySQL , so we will create a table first named emp2-

```
mysql> CREATE TABLE emp2
    -> (
    -> emp_id int,
    -> emp_name varchar(20),
    -> emp_sal int
    -> );
Query OK, 0 rows affected (0.08 sec)

mysql> select * from emp2;
Empty set (0.00 sec)
```

We are using below export command to export from HIVE to MySQL-

- ➢ **sqoop export --connect jdbc:mysql://localhost/db1 **
- ➢ **--username 'root' -P --columns emp_id,emp_name,emp_sal **
- ➢ **--table 'emp2' **
- ➢ **--export-dir /apps/hive/warehouse/emp **
- ➢ **--input-fields-terminated-by ',' **
- ➢ **-m 1**

Here are making a JDBC connection with MySQL database. We have specified username as root.
We are choosing only emp_id, emp_name and emp_sal from the table emp to load into MySQL table emp2. The directory- /apps/hive/warehouse/emp is the HDFS directory where HIVE table emp is stored as we now that HIVE stores data inside HIVE internally.
The delimiter has been specified as ','.
-m 1 specifies that there is only 1 mapper for this operation-

```
[root@sandbox ~]# sqoop export --connect jdbc:mysql://localhost/db1 \
> --username 'root' -P --columns emp_id,emp_name,emp_sal \
> --table 'emp2' \
> --export-dir /apps/hive/warehouse/emp \
> --input-fields-terminated-by ',' \
> -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.
```

Below screenshot shows that the operation is successful and 10 records were exported-

```
17/11/26 21:49:58 INFO mapreduce.ExportJobBase: Transferred 327 bytes in 101.048 seconds (3.2361 bytes/sec)
17/11/26 21:49:58 INFO mapreduce.ExportJobBase: Exported 10 records.
[root@sandbox ~]#
```

Now if we do a select from emp2 table in MySQL we can very well see all 10 records loaded inside the table emp2-

```
mysql> select * from emp2;
+--------+----------+---------+
| emp_id | emp_name | emp_sal |
+--------+----------+---------+
|    101 | Amitabh  |   20000 |
|    102 | Shahrukh |   10000 |
|    103 | Akshay   |   11000 |
|    104 | Anubhav  |    5000 |
|    105 | Pawan    |    2500 |
|    106 | Aamir    |   25000 |
|    107 | Salman   |   17500 |
|    108 | Ranbir   |   14000 |
|    109 | Katrina  |    1000 |
|    110 | Priyanka |    2000 |
+--------+----------+---------+
10 rows in set (0.01 sec)
```