Sqoop Assignment 2

For this assignment, I've used Hortonworks VM and used MySQl, Hive.

1) Starting the mySql using below syntax

sudo service mysqld start

2) Logging into mysql command line as user 'root'

mysql -u root

use db1;

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..

- For this problem, First i created a Employee table in MySql like this.

```
CREATE TABLE sqoop_employee (
   id int,
   name varchar(20),
   designation varchar(25),
   city varchar(15)
);
```

- Inserted few records in the employee table, by using the below syntax.

insert into employee values(1, 'kasthuri','Software Engg','Bangalore');

insert into employee values(2, 'malini','IT Analyst','Walnut');

insert into employee values(3, 'deepa','Charted Accountant','Bangalore');

Screenshot for creating sqoop_employee table in mySql

```
mysql> CREATE TABLE employee (
    ->     id int,
    ->     name varchar(20),
    -> designation varchar(25),
    ->     city varchar(15)
    -> );
Query OK, 0 rows affected (0.13 sec)
```

Screenshot for inserting records in sqoop_employee table in MySql

```
mysql> insert into sqoop_employee values(1, 'kasthuri','Software Engg','Bangalore');
Query OK, 1 row affected (0.00 sec)

mysql>
mysql> insert into sqoop_employee values(2, 'malini','IT Analyst','Walnut');
Query OK, 1 row affected (0.00 sec)

mysql>
mysql> insert into sqoop_employee values(3, 'deepa','Charted Accountant','Bangalore');
Query OK, 1 row affected (0.00 sec)

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

mysql> select * from sqoop_employee;
+------+----------+--------------------+-----------+
| id   | name     | designation        | city      |
+------+----------+--------------------+-----------+
|    1 | kasthuri | Software Engg      | Bangalore |
|    2 | malini   | IT Analyst         | Walnut    |
|    3 | deepa    | Charted Accountant | Bangalore |
+------+----------+--------------------+-----------+
3 rows in set (0.00 sec)
```

Importing the data from MySql to Hive:

Then on the command line, executed following command to create a sqoop job names 'incremental_job1' with incremental append.

- Sqoop job creates and saves the import and export commands. It specifies parameters to identify and recall the saved job.
- This re-calling or re-executing is used in the incremental import, which can import the updated rows from RDBMS table to HDFS.

```
sqoop job --create /incremental_job1 \
-- \
import --connect jdbc:mysql://localhost/db1 \
--username 'root' -P --table 'sqoop_employee' --target-dir '/incremental_job1' \
--hive-import \
--incremental append \
--check-column id \
-m 1;
```

Screenshot of Mobaxterm for creating a sqoop import job for MySql to hive import:

```
[root@sandbox ~]# sqoop job --create incremental_hive_job \
> -- \
> import --connect jdbc:mysql://localhost/db1 \
> --username 'root' -P --table 'sqoop_employee' --target-dir '/incremental_hive_job' \
> --hive-import \
> --incremental append \
> --check-column id \
> -m 1;
```

So once the job is created, we are executing the job to check if the first 3 records are imported into the hive table.

Screenshot of Mobaxterm for executing the sqoop job:

```
[root@sandbox ~]# sqoop job --exec /incremental_job_ex1
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
17/11/18 12:40:12 INFO sqoop.Sqoop: Running Sqoop version: 1.4.4.2.1.1.0-385
Enter password:
17/11/18 12:40:15 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
```

```
17/11/18 12:40:25 INFO mapreduce.Job: Running job: job_1510993803614_0029
17/11/18 12:40:35 INFO mapreduce.Job: Job job_1510993803614_0029 running in uber mode : false
17/11/18 12:40:35 INFO mapreduce.Job:  map 0% reduce 0%
17/11/18 12:40:43 INFO mapreduce.Job:  map 100% reduce 0%
17/11/18 12:40:43 INFO mapreduce.Job: Job job_1510993803614_0029 completed successfully
```

The sqoop job completed successfully. Now in the hive shell, we verify, if a new table sqoop_employee is automatically created and 3 rows are inserted into the sqoop_employee table.

In the below screenshot we can see, initally in show tables, sqoop_employee table doesn't exist, after executing the sqoop job, the employee_sqoop table is automatically created.

Screenshot of Mobaxterm for viewing the hive sqoop_employee table:

```
hive> show tables;
OK
college
employee
sample_07
sample_08
temperature_dataset
Time taken: 0.356 seconds, Fetched: 5 row(s)
hive> select * from sqoop_employee;
OK
1       kasthuri        Software Engg   Bangalore
2       malini  IT Analyst      Walnut
3       deepa   Charted Accountant      Bangalore
Time taken: 0.35 seconds, Fetched: 3 row(s)
```

Now to check incremental load, updated the sqoop_employee table in MySql with 3 more records

Screenshot for inserting more records in sqoop_employee table in MySql

```
mysql> insert into sqoop_employee values(4, 'Jyothi','IT Analyst','Mysore');
Query OK, 1 row affected (0.04 sec)

mysql>
mysql> insert into sqoop_employee values(5, 'divya','Charted Accountant','Chennai');
Query OK, 1 row affected (0.01 sec)

mysql>
mysql> insert into sqoop_employee values(6, 'Radha','Charted Accountant','Trichy');
Query OK, 1 row affected (0.00 sec)

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

mysql> select * from sqoop_employee;
+------+----------+--------------------+-----------+
| id   | name     | designation        | city      |
+------+----------+--------------------+-----------+
|    1 | kasthuri | Software Engg      | Bangalore |
|    2 | malini   | IT Analyst         | Walnut    |
|    3 | deepa    | Charted Accountant | Bangalore |
|    4 | Jyothi   | IT Analyst         | Mysore    |
|    5 | divya    | Charted Accountant | Chennai   |
|    6 | Radha    | Charted Accountant | Trichy    |
+------+----------+--------------------+-----------+
6 rows in set (0.09 sec)
```

Screenshot of Mobaxterm for re-executing the sqoop job:

```
[root@sandbox ~]# sqoop job --exec /incremental_job_ex1
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
17/11/18 12:41:34 INFO sqoop.Sqoop: Running Sqoop version: 1.4.4.2.1.1.0-385
Enter password:

17/11/18 12:41:44 INFO mapreduce.Job: Running job: job_1510993803614_0030
17/11/18 12:41:53 INFO mapreduce.Job: Job job_1510993803614_0030 running in uber mode : false
17/11/18 12:41:53 INFO mapreduce.Job:  map 0% reduce 0%
17/11/18 12:42:02 INFO mapreduce.Job:  map 100% reduce 0%
17/11/18 12:42:02 INFO mapreduce.Job: Job job_1510993803614_0030 completed successfully
```

The sqoop job completed successfully. Now in the hive shell, we verify, if the sqoop_employee is updated with 3 more.

In the below screenshot we can see, initally in the sqoop_employee table 3 records are there, after executing the sqoop job now 6 records are there implies that the incremental load happened successfully..

Screenshot of Mobaxterm for viewing the hive sqoop employee table:

```
hive> select * from sqoop_employee;
OK
1       kasthuri        Software Engg   Bangalore
2       malini  IT Analyst      Walnut
3       deepa   Charted Accountant      Bangalore
4       Jyothi  IT Analyst      Mysore
5       divya   Charted Accountant      Chennai
6       Radha   Charted Accountant      Trichy
Time taken: 0.699 seconds, Fetched: 6 row(s)
```