Join Query– Using Sqoop

Use Case

edureka!

Version 2.0

Join Query using Sqoop

Table of Contents

Join Query in Sqoop

Problem Statement:

We know how to perform the joins in MySQL, similarly we can perform the joins in Sqoop as well. In this we will see how to perform the joins in Sqoop. Sqoop has an argument --query that will help you in performing the joins.

Important Links:

Edureka VM Installation Guide:

Please refer to Installation guide section present in the LMS for accessing the Edureka VM Installation Guide.

Codes:

https://edureka.wistia.com/medias/1dihki22lq/download?media_file_id=67127797

Dataset:

Let us consider two tables:

emp - which consists of the employee details like employee id , employee name and the department to which the employee belongs to.

dept- which consists of the department details like department id, department name.

```
+------+----------+-------+
| id   | name     | depid |
+------+----------+-------+
|    1 | Ram      |     1 |
|    2 | sita     |     1 |
|    1 | priyanka |     2 |
+------+----------+-------+
```

```
+--------+----------+
| deptid | deptname |
+--------+----------+
|      1 | sales    |
|      2 | support  |
+--------+----------+
```

Implementation:

Let us first create two tables in MySQL and insert data into it.

Command: mysql -u root


```
[edureka@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>
```


Command: create database edureka;

Command: use edureka;

Command: create table emp(id int, name varchar(30), depid int);

Command: create table dept(deptid int, deptname varchar(30));

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

mysql> use edureka;
Database changed
mysql> create table emp( id int, name varchar(30), depid int);
Query OK, 0 rows affected (0.03 sec)

mysql> create table dept(deptid int, deptname varchar(30));
Query OK, 0 rows affected (0.04 sec)
```

Creating the database

Create the two tables

Command: insert into emp values(1, 'Ram',1);

Command: insert into emp values(2, 'sita',1);

Command: insert into emp values(1, 'priyanka',2);


```
mysql> insert into emp values(1, 'Ram',1);
Query OK, 1 row affected (0.01 sec)

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

mysql> insert into emp values(1, 'priyanka',2);
Query OK, 1 row affected (0.00 sec)
```

Command: insert into dept values(1, 'sales');

Command: insert into dept values(2,'support');


```
mysql> insert into dept values(1, 'sales');
Query OK, 1 row affected (0.02 sec)

mysql> insert into dept values(2,'support');
Query OK, 1 row affected (0.01 sec)
```

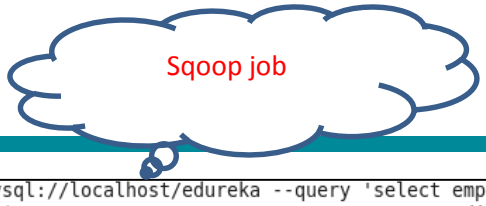
Command: grant all privileges on *.* to root@'localhost' identified by 'edureka' with grant option;

```
mysql> grant all privileges on *.* to root@'localhost' identified by 'edureka'
with grant option;
Query OK, 0 rows affected (0.0
```

Grant the privileges to the root user

Let us perform the join in sqoop by using --query parameter and its value would be the mysql query that we use to join the two tables.

Command: sqoop import --connect jdbc:mysql://localhost/edureka --query 'select emp.* , dept.* from emp join dept on (emp.depid=dept.deptid) where $CONDITIONS' --username root -P --target-dir /sqoopoutput -m 1




```
[edureka@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/edureka --query 'select emp.* , dept.*
 from emp join dept on (emp.depid=dept.deptid) where $CONDITIONS' --username root -P --target-dir /sqoopout
put -m 1
Warning: /usr/lib/hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Enter password:
15/01/28 13:40:21 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
15/01/28 13:40:21 INFO tool.CodeGenTool: Beginning code generation
15/01/28 13:40:23 INFO manager.SqlManager: Executing SQL statement: select emp.* , dept.* from emp join dep
t on (emp.depid=dept.deptid) where  (1 = 0)
15/01/28 13:40:23 INFO manager.SqlManager: Executing SQL statement: select emp.* , dept.* from emp join dep
t on (emp.depid=dept.deptid) where  (1 = 0)
15/01/28 13:40:23 INFO manager.SqlManager: Executing SQL statement: select emp.* , dept.* from emp join dep
t on (emp.depid=dept.deptid) where  (1 = 0)
```

```
        Map-Reduce Framework
                Map input records=3
                Map output records=3
                Input split bytes=87
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=465
                CPU time spent (ms)=3470
                Physical memory (bytes) snapshot=57851904
                Virtual memory (bytes) snapshot=357326848
                Total committed heap usage (bytes)=16252928
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=56
15/01/28 13:43:05 INFO mapreduce.ImportJobBase: Transferred 56 bytes in 141.2617 seconds (0.3964 bytes/sec)
15/01/28 13:43:05 INFO mapreduce.ImportJobBase: Retrieved 3 records.
[edureka@localhost ~]$
```

Let us check the output of Sqoop job.

```
[edureka@localhost ~]$ hadoop dfs -ls /sqoopoutput
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

15/01/28 13:57:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... usi
ng builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 edureka supergroup          0 2015-01-28 13:43 /sqoopoutput/_SUCCESS
-rw-r--r--   1 edureka supergroup         56 2015-01-28 13:42 /sqoopoutput/part-m-00000
[edureka@localhost ~]$ hadoop dfs -cat /sqoopoutput/part-m-00000
DEPRECATED: Use of this script to e                    deprecated.
Instead use the hdfs comm

15/01/28 13:58:02 WARN                                 ative-hadoop library for your platform... usi
ng builtin-java classe
1,Ram,1,1,sales
2,sita,1,1,sales
1,priyanka,2,2,support
```

Sqoop job output

We have successfully performed join operation using Sqoop!!!