Assignment 20.1

1. Perform UPSERT in Sqoop export.

Read a file from HDFS and based on the id field, perform UPSERT in MySQL table.

In UPSERT, if the field exists, then it is updated else it is inserted.

Step 1: use 'customers.dat' file as an input to the sqoop.

Data is ',' separated and file contains field as id,name,location,age.

```
[acadgild@localhost dataset]$ cat customers.dat
1,Amit,IND,18
2,Sumit,PAK,20
3,Rohit,AUS,26
4,Namit,UK,24[acadgild@localhost dataset]$
```

Step 2: move 'customers.dat' file to HDFS

$ start-all.sh

$ hadoop fs -mkdir /sqoop

$ hadoop fs -put customers.dat /sqoop

$ hadoop fs -cat /sqoop/customers.dat

```
[acadgild@localhost dataset]$ hadoop fs -mkdir /sqoop
17/08/24 00:21:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost dataset]$ hadoop fs -put customers.dat /sqoop
17/08/24 00:21:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost dataset]$ hadoop fs -ls /sqoop
17/08/24 00:22:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r--   1 acadgild supergroup         60 2017-08-24 00:21 /sqoop/customers.dat
[acadgild@localhost dataset]$ hadoop fs -cat /sqoop/customers.dat
17/08/24 00:22:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1,Amit,IND,18
2,Sumit,PAK,20
3,Rohit,AUS,26
4,Namit,UK,24[acadgild@localhost dataset]$
```

Step 3: start mysql.

Change user to root

$ sudo su

Start mysqld service

 # service mysqld start

Start mysql as user root

 # mysql -u root

```
[acadgild@localhost ~]$ sudo su
[sudo] password for acadgild:
[root@localhost acadgild]# mysql -u root
ERROR 2002 (HY000): Can't connect to local MySQL server through socket '/var/lib/mysql/mysql.sock' (2)
[root@localhost acadgild]# service mysqld start
Starting mysqld:                                           [  OK  ]
[root@localhost acadgild]# 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.
```

Step 4: create table with the above mentioned fields (columns)

use db1;

show tables;

create table customer

(

id int(5),

name varchar(20),

location varchar(20),

age int(3),

PRIMARY KEY (id)

);

insert into customer values(1,'Yogesh','IND',25);

*select * from customer;*

commit;

```
mysql> use db1;
Database changed
mysql> show tables;
+---------------------+
| Tables_in_db1       |
+---------------------+
| statewiseBPL80      |
| statewiseBPLacheived |
+---------------------+
2 rows in set (0.00 sec)

mysql> create table customer
    -> (
    -> id int(5),
    -> name varchar(20),
    -> location varchar(20),
    -> age int(3),
    -> PRIMARY KEY (id)
    -> );
Query OK, 0 rows affected (0.00 sec)

mysql> insert into customer values(1,'Yogesh','IND',25);
Query OK, 1 row affected (0.00 sec)

mysql> select * from customer;
+----+--------+----------+------+
| id | name   | location | age  |
+----+--------+----------+------+
|  1 | Yogesh | IND      |   25 |
+----+--------+----------+------+
1 row in set (0.00 sec)

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

mysql>
```

Step 5: Perform sqoop export to allow UPSERT in the SQL table from HDFS file.

*sqoop export --connect jdbc:mysql://localhost/db1 *

*--username 'root' -P --table 'customer' --export-dir '/sqoop' *

*--input-fields-terminated-by ',' -m 1 --columns id,name,location,age *

*--update-key id *

--update-mode allowinsert

```
[acadgild@localhost dataset]$ sqoop export --connect jdbc:mysql://localhost/db1 \
> --username 'root' -P --table 'customer' --export-dir '/sqoop' \
> --input-fields-terminated-by ',' -m 1 --columns id,name,location,age \
> --update-key id \
> --update-mode allowinsert
Warning: /usr/local/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/local/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /usr/local/sqoop/../zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-08-24 00:33:06,753 INFO  [main] sqoop.Sqoop: Running Sqoop version: 1.4.5
Enter password:
2017-08-24 00:33:16,592 INFO  [main] manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2017-08-24 00:33:16,592 INFO  [main] tool.CodeGenTool: Beginning code generation
2017-08-24 00:33:21,887 INFO  [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM `customer` AS t LIMIT 1
2017-08-24 00:33:21,932 INFO  [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM `customer` AS t LIMIT 1
2017-08-24 00:33:22,843 INFO  [main] orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/hadoop-2.6.0
Note: /tmp/sqoop-acadgild/compile/e37dd1c57365c3a123599108f2312766/customer.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
```

```
2017-08-24 00:33:40,201 INFO  [main] Configuration.deprecation: mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.files.filesizes
2017-08-24 00:33:40,756 INFO  [main] mapreduce.JobSubmitter: Submitting tokens for job: job_1503514160743_0001
2017-08-24 00:33:41,679 INFO  [main] impl.YarnClientImpl: Submitted application application_1503514160743_0001 to ResourceManager at /0.0.0.0:8032
2017-08-24 00:33:41,731 INFO  [main] mapreduce.Job: The url to track the job: http://http://localhost:8088/proxy/application_1503514160743_0001/
2017-08-24 00:33:41,737 INFO  [main] mapreduce.Job: Running job: job_1503514160743_0001
2017-08-24 00:33:52,723 INFO  [main] mapreduce.Job: Job job_1503514160743_0001 running in uber mode : false
2017-08-24 00:33:52,723 INFO  [main] mapreduce.Job:  map 0% reduce 0%
2017-08-24 00:33:58,789 INFO  [main] mapreduce.Job:  map 100% reduce 0%
2017-08-24 00:33:58,804 INFO  [main] mapreduce.Job: Job job_1503514160743_0001 completed successfully
```

Step 6: Check table content in SQL.

*select * from customer;*

```
mysql> select * from customer;
+----+-------+----------+------+
| id | name  | location | age  |
+----+-------+----------+------+
|  1 | Amit  | IND      |   18 |
|  2 | Sumit | PAK      |   20 |
|  3 | Rohit | AUS      |   26 |
|  4 | Namit | UK       |   24 |
+----+-------+----------+------+
4 rows in set (0.00 sec)

mysql>
```