

Assignment 24thDec

SCD1 Achived in the hive

Customer id 73

```
hive> SELECT * FROM cus_transactions_external LIMIT 1;
OK
cus_transactions_external.custid      cus_transactions_external.username      cus_transactions_external.quote_count  cus_transactions_external.ip      cus_transactions_external.entry_time  cus_transactions_external.prp_1  cus_transactions_external.prp_2  cus_transactions_external.prp_3  cus_transactions_external.ms      cus_transactions_external.http_type  cus_transactions_external.purchase_category  cus_transactions_external.total_count  cus_transactions_external.purchase_sub_category  cus_transactions_external.http_info  cus_transactions_external.status_code  cus_transactions_external.year  cus_transactions_external.month
73      beautifulbear      9350      361.631.17.30      27/Oct/2011:08:31:27 -0500      8      31      27      -0500      GET      /demo      0      /demo      Jakarta Commons-HttpCl
ient/3.0-rc4      100      2011      10
Time taken: 0.171 seconds, Fetched: 1 row(s)
hive>
```

Only one record I updated

Updated CustomerID 73

```
hive (default)> SELECT * FROM cus_transactions_external WHERE cus_transactions_external.custid=73;
OK
cus_transactions_external.custid      cus_transactions_external.username      cus_transactions_external.quote_count  cus_transactions_external.ip      cus_transactions_external.entry_time  cus_transactions_external.prp_1  cus_transactions_external.prp_2  cus_transactions_external.prp_3  cus_transactions_external.ms      cus_transactions_external.http_type  cus_transactions_external.purchase_category  cus_transactions_external.total_count  cus_transactions_external.purchase_sub_category  cus_transactions_external.http_info  cus_transactions_external.status_code  cus_transactions_external.year  cus_transactions_external.month
73      blackmouse      9185      322.0.648.325      27/Oct/2011:08:31:27 -0500      17      53      39      -0500      GET      /download/download3.zip      0      /download/download3.zip      Mozilla/4.0 (compatible; MSIE 5.00; Windows 98) 201      2011      10
Time taken: 0.505 seconds, Fetched: 1 row(s)
hive (default)>
```

Verifying the before hadfs and after hadfs in the SQL table

```
mysql> select COUNT(*) from cus_transactions_check;
+-----+
| COUNT(*) |
+-----+
|      166349 |
+-----+
1 row in set (0.07 sec)

mysql> select COUNT(*) from cus_transactions_export;
+-----+
| COUNT(*) |
+-----+
|      166349 |
+-----+
1 row in set (0.06 sec)

mysql>
```

Total counts are same

Validating the rows in the external table HIVE as there was only one row which got updated

```
hive (default)> SELECT COUNT(*) FROM cus_transactions_external;
Query ID = saif_20211227072629_1bbae680-71e5-491e-bfa7-67794c400892
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1640565533810_0079, Tracking URL = http://localhost:8088/proxy/application_1640565533810_0079/
Kill Command = /usr/local/hadoop-env/hadoop-3.2.1/bin/mapred job -kill job_1640565533810_0079
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-12-27 07:26:33,986 Stage-1 map = 0%, reduce = 0%
2021-12-27 07:26:38,067 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.02 sec
2021-12-27 07:26:43,207 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.37 sec
MapReduce Total cumulative CPU time: 3 seconds 370 msec
Ended Job = job_1640565533810_0079
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.37 sec HDFS Read: 30827745 HDFS Write: 106 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 370 msec
OK
_c0
166348
Time taken: 14.706 seconds, Fetched: 1 row(s)
hive (default)>
```