Case Study Description
Let us take up the CUSTOMER and TRANSACTIONS table we have created in the Let's Do Together section. Let us solve the following use cases using these tables :-

```
hive (default)> create database acadgilddb;
OK
Time taken: 3.09 seconds
hive (default)> use acadgilddb;
OK
Time taken: 0.112 seconds
hive (acadgilddb)> CREATE TABLE CUSTOMER(
                 > custid INT,
                 > fname STRING,
                 > lname STRING,
                 > age INT,
                 > profession STRING)
                 > row format delimited fields terminated by ',';
OK
Time taken: 1.821 seconds
```

```
hive (acadgilddb)> LOAD DATA LOCAL INPATH '/home/acadgild/Case_Study-2/custs.txt' into table CUSTOMER;
Loading data to table acadgilddb.customer
Table acadgilddb.customer stats: [numFiles=1, totalSize=238]
OK
Time taken: 3.723 seconds
hive (acadgilddb)> select * from CUSTOMER;
OK
customer.custid customer.fname  customer.lname  customer.age    customer.profession
101     Amitabh Bacchan 65      Actor
102     Sharukh Khan    45      Doctor
103     Akshay  Kumar   38      Dentist
104     Anubahv kumar   58      Business
105     Pawan   Trivedi 34      service
106     Aamir   Null    42      scientest
107     Salman  Khan    43      Surgen
108     Ranbir  Kapoor  26      Industrialist
Time taken: 3.324 seconds, Fetched: 8 row(s)
```

```
hive (acadgilddb)> CREATE TABLE TRANSACTIONS (
                 > txnno INT,
                 > txndate STRING,
                 > custno INT,
                 > amount DOUBLE,
                 > category STRING,
                 > product STRING,
                 > city STRING,
                 > state STRING,
                 > spendby STRING)
                 >
                 > row format delimited fields terminated by ',';
OK
Time taken: 0.202 seconds
```

```
hive (acadgilddb)> LOAD DATA LOCAL INPATH '/home/acadgild/Case_Study-2/txn.txt' into table TRANSACTIONS;
Loading data to table acadgilddb.transactions
Table acadgilddb.transactions stats: [numFiles=1, totalSize=493]
OK
Time taken: 1.313 seconds
hive (acadgilddb)> select * from TRANSACTIONS;
OK
transactions.txnno      transactions.txndate    transactions.custno     transactions.amount     transactions.category   transactions.product    transactions.city
ransactions.state       transactions.spendby
97834   05/02/2018      101     965.0   Entertainment   Movie   Pune    Maharashtra     Daughter
98396   12/01/2018      102     239.0   Food    Grocery Patna   Bihar   Self
34908   06/01/2018      101     875.0   Travel  Air     Bangalore       Karnataka       Spouse
70958   17/02/2018      104     439.0   Food    Restaurant      Delhi   Delhi   Wife
9874    21/01/2018      105     509.0   Entertainment   Park    Kolkata West Bengal     NULL
94585   19/01/2018      106     629.0   Rent    House   Hyderabad       Telangana       Self
45509   20/01/2018      107     953.0   Travel  Rail    Chennai Tamil Nadu      Brother
7864    01/02/2018      108     569.0   Rent    Parking Goa     Goa     Wife
Time taken: 0.232 seconds, Fetched: 8 row(s)
```

1. Find out the number of transaction done by each customer (These should be take up in module 8 itself)

```
hive (acadgilddb)> select custno,count(*) from TRANSACTIONS group by custno;
Query ID = acadgild_20180821043535_213c14c3-d268-4316-9866-9c0d8c17b4e9
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 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_1534799237082_0002, Tracking URL = http://localhost:8088/proxy/application_1534799237082_0002/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_1534799237082_0002
```

```
OK
custno  _c1
101     2
102     1
104     1
105     1
106     1
107     1
108     1
```

2. **Create a new table called TRANSACTIONS_COUNT. This table should have 3 fields - custid, fname and count. (Again to be done in module 8)**

```
hive (acadgilddb)> create table TRANSACTIONS_COUNT (
                 > custid int,
                 > fname string
                 > )
                 > row format delimited fields terminated by ',';
OK
Time taken: 0.154 seconds
```

3. **Now write a hive query in such a way that the query populates the data obtained in Step 1 above and populate the table in step 2 above. (This has to be done in module 9).**

```
hive (acadgilddb)> insert into table TRANSACTIONS_COUNT
                 > select custid,fname from CUSTOMER;
Query ID = acadgild_20180821045454_8c27e02a-181a-46d7-9af0-83b5afadfab3
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1534799237082_0003, Tracking URL = http://localhost:8088/proxy/application_1
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_1534799237082_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2018-08-21 04:55:36,226 Stage-1 map = 0%,  reduce = 0%
2018-08-21 04:56:16,229 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 8.49 sec
MapReduce Total cumulative CPU time: 8 seconds 490 msec
Ended Job = job_1534799237082_0003
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://localhost:9000/tmp/hive/acadgild/f5b95e53-8098-4816-9882-768e730dde61/hi
Loading data to table acadgilddb.transactions_count
Table acadgilddb.transactions_count stats: [numFiles=1, numRows=8, totalSize=89, rawDataSize=81
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 8.49 sec   HDFS Read: 466 HDFS Write: 174 SUCCESS
```

4.