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 :-

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

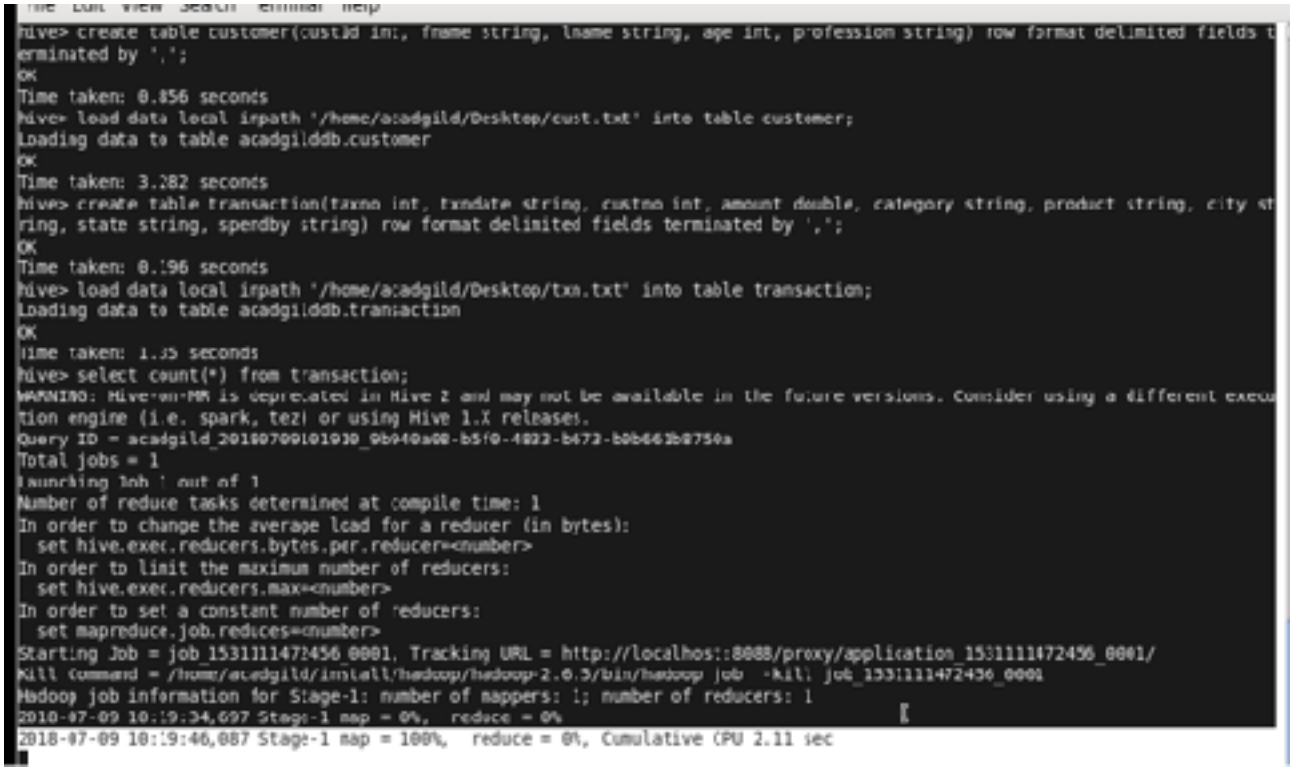
2. Create a new table called TRANSACTIONS_COUNT. This table should have 3 fields - custid, fname and count. (Again to be done in module 8)

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).

4. Now lets make the TRANSACTIONS_COUNT table Hbase complaint. In the sence, use Ser Des And Storate handler features of hive to change the TRANSACTIONS_COUNT table to be able to create a TRANSACTIONS table in Hbase. (This has to be done in module 10)

5. Now insert the data in TRANSACTIONS_COUNT table using the query in step 3 again, this should populate the Hbase TRANSACTIONS table automatically (This has to be done in module 10)

6. Now from the Hbase level, write the Hbase java API code to access and scan the TRANSACTIONS table data from java level.


File Edit View Search Terminal Help
hive> create table customer(custId int, fname string, lname string, age int, profession string) row format delimited fields t
erminated by ',';
OK
Time taken: 0.856 seconds
hive> load data local inpath '/home/acadgild/Desktop/cust.txt' into table customer;
Loading data to table acadgilddb.customer
OK
Time taken: 3.282 seconds
hive> create table transaction(txnno int, txndate string, custno int, amount double, category string, product string, city st
ring, state string, spendby string) row format delimited fields terminated by ',';
OK
Time taken: 0.196 seconds
hive> load data local inpath '/home/acadgild/Desktop/txn.txt' into table transaction;
Loading data to table acadgilddb.transaction
OK
Time taken: 1.35 seconds
hive> select count(*) from transaction;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180709101930_9b940a08-b5f0-4833-b673-b0b661b8758a
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_1531111472456_0001, Tracking URL = http://localhost:8088/proxy/application_1531111472456_0001/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1531111472456_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-07-09 10:19:34,697 Stage-1 map = 0%, reduce = 0%
2018-07-09 10:19:46,087 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.11 sec


hive> select count(*) from transaction;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180709101930_9b940a08-b5f0-4833-b673-b0b661b8758a
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_1531111472456_0001, Tracking URL = http://localhost:8088/proxy/application_1531111472456_0001/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1531111472456_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-07-09 10:19:34,697 Stage-1 map = 0%, reduce = 0%
2018-07-09 10:19:46,087 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.11 sec
2018-07-09 10:20:00,310 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.94 sec
MapReduce Total cumulative CPU time: 4 seconds 940 msec
Ended Job = job_1531111472456_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1   Cumulative CPU: 4.94 sec   HDFS Read: 9315 HDFS Write: 101 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 940 msec
OK
8
Time taken: 51.751 seconds, Fetched: 1 row(s)

```
hive>
hive> select * from customer a join transaction b on a.custid = b.custno;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180709110946_58f153c5-b6bb-447f-9b54-93a05ad717f3
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j
/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!
/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2018-07-09 11:09:54     Starting to launch local task to process map join;      maximum memory = 518979584
2018-07-09 11:09:57     Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/ff1ca6b8-7e24-4de4-8
427-2758c4eac766/hive_2018-07-09_11-09-46_736_7905087743141238463-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile29--.hashta
ble
2018-07-09 11:09:57     Uploaded 1 File to: file:/tmp/acadgild/ff1ca6b8-7e24-4de4-8427-2758c4eac766/hive_2018-07-09_11-09-46_
736_7905087743141238463-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile20--.hashtable (599 bytes)
2018-07-09 11:09:57     End of local task; Time Taken: 2.187 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1531111472456_0004, Tracking URL = http://localhost:8088/proxy/application_1531111472456_0004/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1531111472456_0004
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2018-07-09 11:10:07,176 Stage-3 map = 0%,  reduce = 0%
2018-07-09 11:10:17,663 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 3.44 sec
MapReduce Total cumulative CPU time: 3 seconds 440 msec
Ended Job = job_1531111472456_0004
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 3.44 sec   HDFS Read: 8887 HDFS Write: 920 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 440 msec
OK
101     Amitabh Bacchan 65      Actor   97834   06/02/2018      101     965.0   Entertainment   Movie   Pune    Maharashtra D
aughter
102     Sharukh Khan    45      Doctor  98396   12/01/2018      102     239.0   Food    Grocery Patna   Bihar   Self
101     Amitabh Bacchan 65      Actor   34988   06/01/2018      101     875.0   Travel  Air     Bangalore       Karnataka   S
```

```
hive> create table transaction_count as select custId, fname, count(custId) from customer a join transaction b on a.custid =
b.custno group by custid,fname;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180709113022_c8e12f1f-de19-412c-87f3-b049e83d4e32
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j
/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!
/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2018-07-09 11:30:32     Starting to launch local task to process map join;      maximum memory = 518979584
2018-07-09 11:30:34     Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/ff1ca6b8-7e24-4de4-8
427-2758c4eac766/hive_2018-07-09_11-30-22_600_2817484850305340574-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile50--.hashta
ble
2018-07-09 11:30:34     Uploaded 1 File to: file:/tmp/acadgild/ff1ca6b8-7e24-4de4-8427-2758c4eac766/hive_2018-07-09_11-30-22_
600_2817484850305340574-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile50--.hashtable (469 bytes)
2018-07-09 11:30:34     End of local task; Time Taken: 2.514 sec.
Execution completed successfully
MapredLocal task succeeded
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_1531111472456_0007, Tracking URL = http://localhost:8088/proxy/application_1531111472456_0007/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1531111472456_0007
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-07-09 11:30:46,116 Stage-2 map = 0%,  reduce = 0%
2018-07-09 11:30:58,734 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 4.2 sec
2018-07-09 11:31:14,053 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 7.1 sec
MapReduce Total cumulative CPU time: 7 seconds 300 msec
Ended Job = job_1531111472456_0007
Moving data to directory hdfs://localhost:8020/user/hive/warehouse/acadgilddb.db/transaction_count
```

```
MapReduce Total cumulative CPU time: 7 seconds 300 msec
Ended Job = job_1531111472456_0007
Moving data to directory hdfs://localhost:9020/user/hive/warehouse/acadgilddb.db/transaction_count
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 7.3 sec   HDFS Read: 12993 HDFS Write: 176 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 300 msec
OK
Time taken: 53.727 seconds
hive> select * from transaction_count;
OK
101     Amitabh 2
102     Sharukh 1
104     Anubahv 1
105     Pawan   1
106     Aamir   1
107     Salman  1
108     Ranbir  1
Time taken: 0.234 seconds, Fetched: 7 row(s)
hive>
```

```
hive> create external table transaction_count_hbase(custId int, fname string, count int)
    > STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
    >  with serdeproperties ("hbase.columns.mapping"=":key,cf1:custId,cf1:fname")
    > tblproperties('hbase.table.name'="transaction_count");
OK
Time taken: 0.681 seconds
hive> select * from transaction_count_hbase;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'transaction_count_hbase'
hive> select * from transaction_count_hbase;
OK
1       111     NULL
Time taken: 0.657 seconds, Fetched: 1 row(s)
hive> set hbase.mapred.output.outputtable=transaction_count;
hive> insert into table transaction_count_hbase values(2,'aaa',33);
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180709132550_d4004cc9-7f39-4b16-94cd-b795517c59fb
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1531111472456_0008, Tracking URL = http://localhost:8088/proxy/application_1531111472456_0008/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1531111472456_0008
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2018-07-09 13:26:19,415 Stage-3 map = 0%,  reduce = 0%
2018-07-09 13:26:38,201 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 3.87 sec
MapReduce Total cumulative CPU time: 3 seconds 870 msec
Ended Job = job_1531111472456_0008
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 3.87 sec   HDFS Read: 4961 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 870 msec
OK
Time taken: 49.311 seconds
hive> select * from transaction_count_hbase;
OK
1       111     NULL
2       aaa     33
Time taken: 0.663 seconds, Fetched: 2 row(s)
hive>
```

```
hbase(main):012:0> create 'transaction_count','cf1'
0 row(s) in 1.2310 seconds

=> Hbase::Table - transaction_count
hbase(main):013:0> put 'transaction_count','1','cf1:custId','111'
0 row(s) in 0.0140 seconds

hbase(main):014:0> put 'transaction_count','1','cf1:fname','shishir'
0 row(s) in 0.0150 seconds

hbase(main):015:0> scan 'transaction_count'
ROW                           COLUMN+CELL
 1                            column=cf1:custId, timestamp=1531122173675, value=111
 1                            column=cf1:fname, timestamp=1531122190628, value=shishir
1 row(s) in 0.0170 seconds

hbase(main):016:0> scan 'transaction_count'
ROW                           COLUMN+CELL
 1                            column=cf1:custId, timestamp=1531122173675, value=111
 1                            column=cf1:fname, timestamp=1531122190628, value=shishir
 2                            column=cf1:custId, timestamp=1531122996651, value=aaa
 2                            column=cf1:fname, timestamp=1531122996651, value=33
2 row(s) in 0.2140 seconds

hbase(main):017:0>
```