Case study

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

```
File  Edit  View  Search  Terminal  Help
hive> select a.custid, a.fname, count(b.txnno) from CUSTOMER a join TRANSACTIONS b on a.custid=b.custno group by a.custid,a.fname
    > ;
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 Hiv
e 1.X releases.
Query ID = acadgild_20180804100641_c0bb197e-2662-4e65-8373-67fa8bd34b3d
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]
2018-08-04 10:07:03     Starting to launch local task to process map join;      maximum memory = 518979584
2018-08-04 10:07:06     Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/4fca165d-a5c5-494b-874b-cc8a88cb7855/hive_2018-08-04_10-06-41_
719_3103819070842978357-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile10--.hashtable
2018-08-04 10:07:07     Uploaded 1 File to: file:/tmp/acadgild/4fca165d-a5c5-494b-874b-cc8a88cb7855/hive_2018-08-04_10-06-41_719_3103819070842978357-1/-local-10005/Has
hTable-Stage-2/MapJoin-mapfile10--.hashtable (469 bytes)
2018-08-04 10:07:07     End of local task; Time Taken: 4.074 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_1533349119711_0003, Tracking URL = http://localhost:8088/proxy/application_1533349119711_0003/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1533349119711_0003
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-08-04 10:07:28,853 Stage-2 map = 0%,  reduce = 0%
2018-08-04 10:07:45,939 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 4.59 sec
2018-08-04 10:08:00,060 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 7.68 sec
MapReduce Total cumulative CPU time: 7 seconds 680 msec
Ended Job = job_1533349119711_0003
```

```
2018-08-04 10:07:06     Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/4fca165d-a5c5-
719_3103819070842978357-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile10--.hashtable
2018-08-04 10:07:07     Uploaded 1 File to: file:/tmp/acadgild/4fca165d-a5c5-494b-874b-cc8a88cb7855/hive_2018-08-04_10-
hTable-Stage-2/MapJoin-mapfile10--.hashtable (469 bytes)
2018-08-04 10:07:07     End of local task; Time Taken: 4.074 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_1533349119711_0003, Tracking URL = http://localhost:8088/proxy/application_1533349119711_0003/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1533349119711_0003
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-08-04 10:07:28,853 Stage-2 map = 0%,  reduce = 0%
2018-08-04 10:07:45,939 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 4.59 sec
2018-08-04 10:08:00,060 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 7.68 sec
MapReduce Total cumulative CPU time: 7 seconds 680 msec
Ended Job = job_1533349119711_0003
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 7.68 sec   HDFS Read: 13372 HDFS Write: 263 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 680 msec
OK
101     Amitabh 2
102     Sharukh 1
104     Anubahv 1
105     Pawan   1
106     Aamir   1
107     Salman  1
108     Ranbir  1
Time taken: 80.878 seconds, Fetched: 7 row(s)
hive>
```

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)

```
Time taken: 80.878 seconds, Fetched: 7 row(s)
hive> CREATE TABLE TRANSACTIONS_COUNT(
    > custid INT,
    > fname STRING,
    > count INT)
    > row format delimited fields terminated by ',';
OK
Time taken: 1.063 seconds
hive> show tables;
OK
customer
transactions
transactions_count
Time taken: 0.487 seconds, Fetched: 3 row(s)
hive> describe transactions_count;
OK
custid                  int
fname                   string
count                   int
Time taken: 0.153 seconds, Fetched: 3 row(s)
hive>
```

acadgild@localhost:~ | acadgild@localhost:~ | Student Dashboard

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).

File Edit View Search Terminal Help

```
hive> describe transactions_count;
OK
custid                  int
fname                   string
count                   int
Time taken: 0.153 seconds, Fetched: 3 row(s)
hive> insert overwrite table transactions_count select a.custid, a.fname, count(b.txnno) from CUSTOMER a join TRANSACTIONS b on a.custid=b.custno group by
name;
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)
e 1.X releases.
Query ID = acadgild_20180804102907_0667cd84-107f-41ab-bc5f-90856200cd84
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/StaticLoggerBi
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2018-08-04 10:29:38     Starting to launch local task to process map join;      maximum memory = 518979584
2018-08-04 10:29:42     Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/4fca165d-a5c5-494b-874b-cc8a88cb7855/hive_2018-08-
384_7549071872107172955-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile20--.hashtable
2018-08-04 10:29:42     Uploaded 1 File to: file:/tmp/acadgild/4fca165d-a5c5-494b-874b-cc8a88cb7855/hive_2018-08-04_10-29-07_384_7549071872107172955-1/-loc
hTable-Stage-2/MapJoin-mapfile20--.hashtable (469 bytes)
2018-08-04 10:29:42     End of local task; Time Taken: 4.605 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_1533349119711_0004, Tracking URL = http://localhost:8088/proxy/application_1533349119711_0004/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job 1533349119711 0004
```

```
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_1533349119711_0004, Tracking URL = http://localhost:8088/proxy/application_1533349119711_0004/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1533349119711_0004
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-08-04 10:30:03,516 Stage-2 map = 0%,  reduce = 0%
2018-08-04 10:30:42,725 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 14.32 sec
2018-08-04 10:31:00,609 Stage-2 map = 100%,  reduce = 67%, Cumulative CPU 18.69 sec
2018-08-04 10:31:04,106 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 21.09 sec
MapReduce Total cumulative CPU time: 21 seconds 90 msec
Ended Job = job_1533349119711_0004
Loading data to table acadgilddb.transactions_count
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 21.09 sec   HDFS Read: 14106 HDFS Write: 177 SUCCESS
Total MapReduce CPU Time Spent: 21 seconds 90 msec
OK
Time taken: 120.13 seconds
hive> select * from transactions_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: 1.082 seconds, Fetched: 7 row(s)
hive>
```

acadgild@localhost:~ | acadgild@localhost:~ | Student Dashboard - ... | Session8 Todo Hive B... | New Note 3

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)

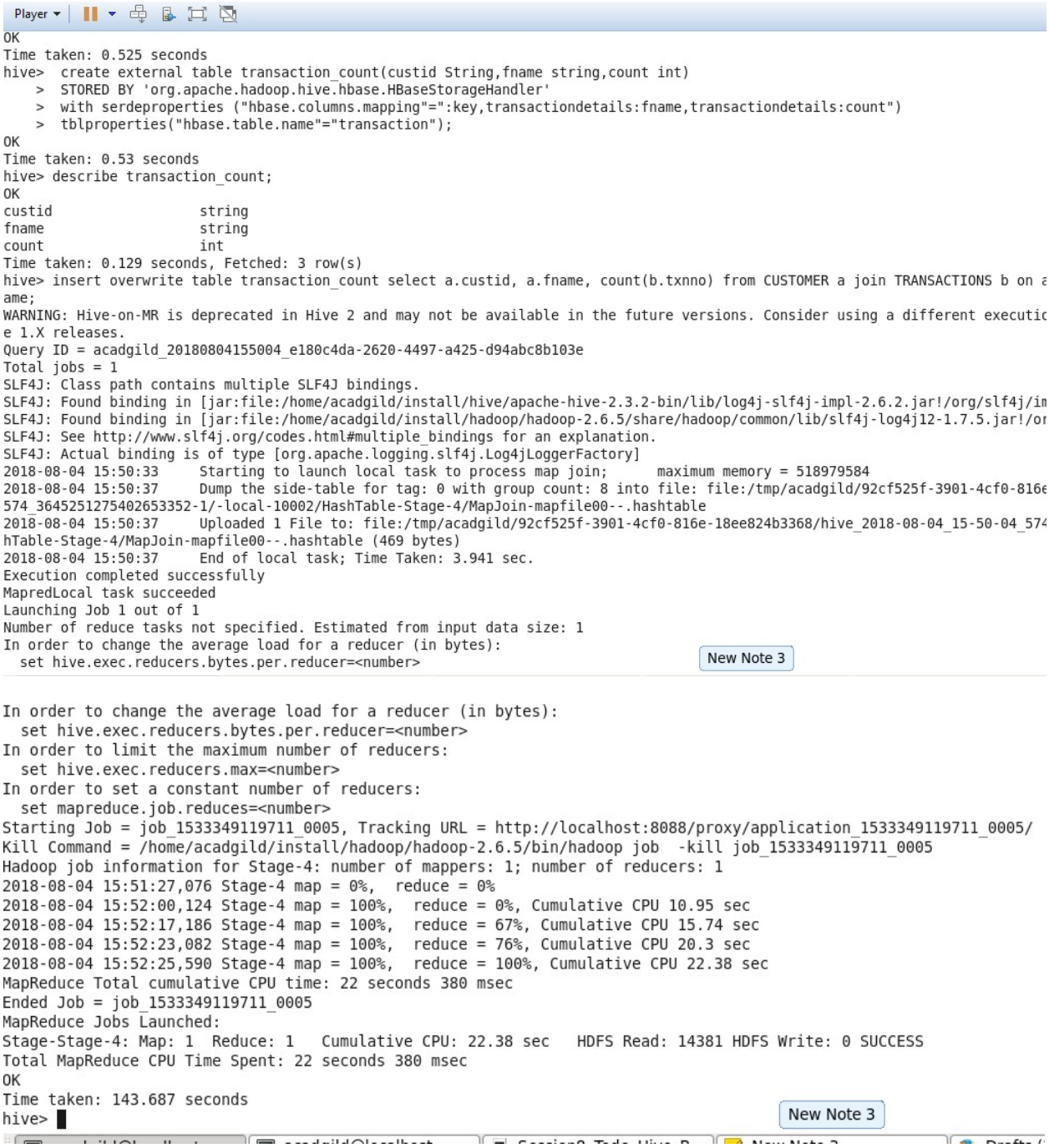
5. Now insert the data in TRANSACTIONS_COUNT table using the query in step 3 again, this should populate the Hbase TRANSACTIONS table automatically (This has to be done in module 10)

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

```
hbase(main):006:0> create 'transaction','transactiondetails'
0 row(s) in 1.2730 seconds

=> Hbase::Table - transaction
hbase(main):007:0> list
TABLE
transaction
1 row(s) in 0.0120 seconds

=> ["transaction"]
```


Player
OK
Time taken: 0.525 seconds
hive> create external table transaction_count(custid String,fname string,count int)
> STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
> with serdeproperties ("hbase.columns.mapping"=":key,transactiondetails:fname,transactiondetails:count")
> tblproperties("hbase.table.name"="transaction");
OK
Time taken: 0.53 seconds
hive> describe transaction_count;
OK
custid string
fname string
count int
Time taken: 0.129 seconds, Fetched: 3 row(s)
hive> insert overwrite table transaction_count select a.custid, a.fname, count(b.txnno) from CUSTOMER a join TRANSACTIONS b on a
ame;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different executic
e 1.X releases.
Query ID = acadgild_20180804155004_e180c4da-2620-4497-a425-d94abc8b103e
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/im
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/or
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2018-08-04 15:50:33 Starting to launch local task to process map join; maximum memory = 518979584
2018-08-04 15:50:37 Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/92cf525f-3901-4cf0-816e
574_3645251275402653352-1/-local-10002/HashTable-Stage-4/MapJoin-mapfile00--.hashtable
2018-08-04 15:50:37 Uploaded 1 File to: file:/tmp/acadgild/92cf525f-3901-4cf0-816e-18ee824b3368/hive_2018-08-04_15-50-04_574
hTable-Stage-4/MapJoin-mapfile00--.hashtable (469 bytes)
2018-08-04 15:50:37 End of local task; Time Taken: 3.941 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>
New Note 3
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_1533349119711_0005, Tracking URL = http://localhost:8088/proxy/application_1533349119711_0005/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1533349119711_0005
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 1
2018-08-04 15:51:27,076 Stage-4 map = 0%, reduce = 0%
2018-08-04 15:52:00,124 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 10.95 sec
2018-08-04 15:52:17,186 Stage-4 map = 100%, reduce = 67%, Cumulative CPU 15.74 sec
2018-08-04 15:52:23,082 Stage-4 map = 100%, reduce = 76%, Cumulative CPU 20.3 sec
2018-08-04 15:52:25,590 Stage-4 map = 100%, reduce = 100%, Cumulative CPU 22.38 sec
MapReduce Total cumulative CPU time: 22 seconds 380 msec
Ended Job = job_1533349119711_0005
MapReduce Jobs Launched:
Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 22.38 sec HDFS Read: 14381 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 22 seconds 380 msec
OK
Time taken: 143.687 seconds
hive>
New Note 3

```
hbase(main):009:0> scan 'transaction'
ROW                                   COLUMN+CELL
0 row(s) in 0.0570 seconds

hbase(main):010:0> scan 'transaction'
ROW                                   COLUMN+CELL
 101                                  column=transactiondetails:count, timestamp=1533378144966, value=2
 101                                  column=transactiondetails:fname, timestamp=1533378144966, value=Amitabh
 102                                  column=transactiondetails:count, timestamp=1533378144966, value=1
 102                                  column=transactiondetails:fname, timestamp=1533378144966, value=Sharukh
 104                                  column=transactiondetails:count, timestamp=1533378144966, value=1
 104                                  column=transactiondetails:fname, timestamp=1533378144966, value=Anubahv
 105                                  column=transactiondetails:count, timestamp=1533378144966, value=1
 105                                  column=transactiondetails:fname, timestamp=1533378144966, value=Pawan
 106                                  column=transactiondetails:count, timestamp=1533378144966, value=1
 106                                  column=transactiondetails:fname, timestamp=1533378144966, value=Aamir
 107                                  column=transactiondetails:count, timestamp=1533378144966, value=1
 107                                  column=transactiondetails:fname, timestamp=1533378144966, value=Salman
 108                                  column=transactiondetails:count, timestamp=1533378144966, value=1
 108                                  column=transactiondetails:fname, timestamp=1533378144966, value=Ranbir
7 row(s) in 0.2410 seconds

hbase(main):011:0>
```

New Note 3