

## CASE STUDY - II HIVE

Created 2 Tables CUSTOMER , TRANSACTIONS as part of TODO assignment & loaded data for the same.

```
Acadgild Thu May 3, 12:04:49 PM Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hive
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/sha
re/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]

Logging initialized using configuration in jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hive-common-2.3.2.jar!/hive-log4j2.properties Async
: true
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 Hi
ve 1.X releases.
hive> use acadgilddb;
OK
Time taken: 6.799 seconds
hive> select * from customer;
OK
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: 2.511 seconds, Fetched: 8 row(s)
hive> select * from transactions;
OK
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
```

```
Acadgild Thu May 3, 12:04:49 PM Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hive
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/sha
re/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]

Logging initialized using configuration in jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hive-common-2.3.2.jar!/hive-log4j2.properties Async
: true
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 Hi
ve 1.X releases.
hive> use acadgilddb;
OK
Time taken: 6.799 seconds
hive> select * from customer;
OK
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: 2.511 seconds, Fetched: 8 row(s)
hive> select * from transactions;
OK
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
```

## Q1: Find out the number of transaction done by each customer

I have achieved the same using below query, results shown in screenshot

```
Acadgild Thu May 3, 12:12:45 PM Applications Places System acadgild@localhost:~
File Edit View Search Terminal Help
hive> select a.fname,a.lname,count(b.txnno) from CUSTOMER a join TRANSACTIONS b on a.custid = b.custno group by a.fname,a.lname;
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_20180503121122_1dbc6b1b-d931-4129-8eb9-34dff19e2477
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/StaticLoggerBind
er.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]
2018-05-03 12:11:32 Starting to launch local task to process map join; maximum memory = 518979584
2018-05-03 12:11:35 Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/c9421d20-4085-4fe5-a6ce-b0c384857cfc/hive_2018-05-03
_12-11-22_012_6951188983384884166-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile00-.hashtable
2018-05-03 12:11:35 Uploaded 1 File to: file:/tmp/acadgild/c9421d20-4085-4fe5-a6ce-b0c384857cfc/hive_2018-05-03_12-11-22_012_6951188983384884166-1/-local
-10005/HashTable-Stage-2/MapJoin-mapfile00-.hashtable (522 bytes)
2018-05-03 12:11:35 End of local task; Time Taken: 3.232 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_1525322803121_0005, Tracking URL = http://localhost:8080/proxy/application_1525322803121_0005/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1525322803121_0005
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-05-03 12:11:48,554 Stage-2 map = 0%, reduce = 0%
2018-05-03 12:12:00,238 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 4.35 sec
2018-05-03 12:12:09,117 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 6.25 sec
MapReduce Total cumulative CPU time: 6 seconds 250 msec
Ended Job = job_1525322803121_0005
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 6.25 sec HDFS Read: 13516 HDFS Write: 279 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 250 msec
```

```
Acadgild Thu May 3, 12:12:58 PM Applications Places System acadgild@localhost:~
File Edit View Search Terminal Help
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2018-05-03 12:11:32 Starting to launch local task to process map join; maximum memory = 518979584
2018-05-03 12:11:35 Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/c9421d20-4085-4fe5-a6ce-b0c384857cfc/hive_2018-05-03
_12-11-22_012_6951188983384884166-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile00-.hashtable
2018-05-03 12:11:35 Uploaded 1 File to: file:/tmp/acadgild/c9421d20-4085-4fe5-a6ce-b0c384857cfc/hive_2018-05-03_12-11-22_012_6951188983384884166-1/-local
-10005/HashTable-Stage-2/MapJoin-mapfile00-.hashtable (522 bytes)
2018-05-03 12:11:35 End of local task; Time Taken: 3.232 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_1525322803121_0005, Tracking URL = http://localhost:8080/proxy/application_1525322803121_0005/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1525322803121_0005
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-05-03 12:11:48,554 Stage-2 map = 0%, reduce = 0%
2018-05-03 12:12:00,238 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 4.35 sec
2018-05-03 12:12:09,117 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 6.25 sec
MapReduce Total cumulative CPU time: 6 seconds 250 msec
Ended Job = job_1525322803121_0005
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 6.25 sec HDFS Read: 13516 HDFS Write: 279 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 250 msec
OK
Aamir Null 1
Amitabh Bacchan 2
Anubhav Kumar 1
Pawan Trivedi 1
Ranbir Kapoor 1
Salman Khan 1
Sharukh Khan 1
Time taken: 48.314 seconds, Fetched: 7 row(s)
hive>
```

## Q2: Create a new table called TRANSACTIONS\_COUNT. This table should have two fields - fname and count

I have created table as described & loaded data from TRASACNTIONS table into TRASACTION\_COUNT table -- for the two coloumns FNAME,COUNT, result of table show below.

```
Acadgild Thu May 3, 12:16:38 PM Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
-10005/HashTable-Stage-2/MapJoin-mapfile00--.hashtable (522 bytes)
2018-05-03 12:11:35 End of local task; Time Taken: 3.232 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_1525322803121_0005, Tracking URL = http://localhost:8088/proxy/application_1525322803121_0005/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1525322803121_0005
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-05-03 12:11:48,554 Stage-2 map = 0%, reduce = 0%
2018-05-03 12:12:00,238 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 4.35 sec
2018-05-03 12:12:09,117 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 6.25 sec
MapReduce Total cumulative CPU time: 6 seconds 250 msec
Ended Job = job_1525322803121_0005
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 6.25 sec HDFS Read: 13516 HDFS Write: 279 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 250 msec
OK
Aamir Null 1
Amitabh Bacchan 2
Anubhav kumar 1
Pawan Trivedi 1
Ranbir Kapoor 1
Salman Khan 1
Sharukh Khan 1
Time taken: 48.314 seconds, Fetched: 7 row(s)
hive> CREATE TABLE IF NOT EXISTS TRANSACTION_COUNT
> ( fname string,
> count int)
> row format delimited fields terminated by ',';
OK
Time taken: 0.897 seconds
```

```
Acadgild Thu May 3, 12:23:02 PM Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
hive> INSERT OVERWRITE TABLE TRANSACTION_COUNT
> select a.fname,count(b.txnno) from CUSTOMER a join TRANSACTIONS b on a.custid = b.custno group by 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 Hive 1.X releases.
Query ID = acadgild_20180503121800_cd305369-d50f-4fe5-8a7f-a29d47a05c6b
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]
2018-05-03 12:18:10 Starting to launch local task to process map join; maximum memory = 518979584
2018-05-03 12:18:12 Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/c9421d20-4085-4fe5-a6ce-b0c384857cfc/hive_2018-05-03_12-18-00_979_3529013323770279398-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile10--.hashtable
2018-05-03 12:18:12 Uploaded 1 File to: file:/tmp/acadgild/c9421d20-4085-4fe5-a6ce-b0c384857cfc/hive_2018-05-03_12-18-00_979_3529013323770279398-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile10--.hashtable (469 bytes)
2018-05-03 12:18:12 End of local task; Time Taken: 2.368 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_1525322803121_0006, Tracking URL = http://localhost:8088/proxy/application_1525322803121_0006/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1525322803121_0006
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-05-03 12:18:22,698 Stage-2 map = 0%, reduce = 0%
2018-05-03 12:18:30,460 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.92 sec
2018-05-03 12:18:38,300 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.43 sec
MapReduce Total cumulative CPU time: 4 seconds 430 msec
```

```
Acadgild Thu May 3, 12:23:15 PM Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
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]
2018-05-03 12:18:10 Starting to launch local task to process map join; maximum memory = 518979584
2018-05-03 12:18:12 Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/c9421d20-4085-4fe5-a6ce-b0c384857cfc/hive_2018-05-03_12-18-00_979_3529013323770279398-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile10--.hashtable
2018-05-03 12:18:12 Uploaded 1 File to: file:/tmp/acadgild/c9421d20-4085-4fe5-a6ce-b0c384857cfc/hive_2018-05-03_12-18-00_979_3529013323770279398-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile10--.hashtable (469 bytes)
2018-05-03 12:18:12 End of local task; Time Taken: 2.368 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_1525322803121_0006, Tracking URL = http://localhost:8088/proxy/application_1525322803121_0006/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1525322803121_0006
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-05-03 12:18:22,698 Stage-2 map = 0%, reduce = 0%
2018-05-03 12:18:30,460 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.92 sec
2018-05-03 12:18:38,300 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.43 sec
MapReduce Total cumulative CPU time: 4 seconds 430 msec
Ended Job = job_1525322803121_0006
Loading data to table acadgild.transaction_count
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.43 sec HDFS Read: 13638 HDFS Write: 148 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 430 msec
OK
Time taken: 40.442 seconds
hive>
```

```
Acadgild Thu May 3, 12:25:23 PM Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
2018-05-03 12:18:12 Uploaded 1 File to: file:/tmp/acadgild/c9421d20-4085-4fe5-a6ce-b0c384857cfc/hive_2018-05-03_12-18-00_979_3529013323770279398-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile10--.hashtable (469 bytes)
2018-05-03 12:18:12 End of local task; Time Taken: 2.368 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_1525322803121_0006, Tracking URL = http://localhost:8088/proxy/application_1525322803121_0006/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1525322803121_0006
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-05-03 12:18:22,698 Stage-2 map = 0%, reduce = 0%
2018-05-03 12:18:30,460 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.92 sec
2018-05-03 12:18:38,300 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.43 sec
MapReduce Total cumulative CPU time: 4 seconds 430 msec
Ended Job = job_1525322803121_0006
Loading data to table acadgild.transaction_count
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.43 sec HDFS Read: 13638 HDFS Write: 148 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 430 msec
OK
Time taken: 40.442 seconds
hive> select * from transaction_count;
OK
Aamir 1
Amitabh 2
Anubhav 1
Pawan 1
Ranbir 1
Salman 1
Sharukh 1
Time taken: 0.298 seconds, Fetched: 7 row(s)
hive>
```