

CASE STUDY 2

Customer Transactions Analysis

Assignment 22.2: Case Study Customer Transaction Analysis

Problem Statement

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 :-

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.

Dataset :

- **customer.txt**
- **transaction.txt**

Dataset Descriptions :

1. customer.txt

Terminal Execution :

```
[acadgild@localhost ~]$ cat /home/acadgild/Desktop/customer.txt
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
```

2. transaction.txt

Terminal Execution :

```
[acadgild@localhost ~]$ cat /home/acadgild/Desktop/transaction.txt
97834,05/02/2018,101,965,Entertainment,Movie,Pune,Maharashtra,Daughter
98396,12/01/2018,102,239,Food,Grocery,Patna,Bihar,Self
34908,06/01/2018,101,875,Travel,Air,Bangalore,Karnataka,Spouse
70958,17/02/2018,104,439,Food,Restaurant,Delhi,Delhi,Wife
09874,21/01/2018,105,509,Entertainment,Park,Kolkata,West Bengal
94585,19/01/2018,106,629,Rent,House,Hyderabad,Telangana,Self
45509,20/01/2018,107,953,Travel,Rail,Chennai,Tamil Nadu,Brother
07864,01/02/2018,108,569,Rent,Parking,Goa,Goa,Wife
[acadgild@localhost ~]$
```

INITIAL EXECUTION :

```
[acadgild@localhost ~]$ sudo service sshd start
[sudo] password for acadgild:
[acadgild@localhost ~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/09/09 19:20:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-namenode-localhost.localdomain.
out
localhost: starting datanode, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-datanode-localhost.localdomain.o
ut
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-secondarynamenode-localhost.loc
aldomain.out
18/09/09 19:20:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-resourcemanager-localhost.localdom
ain.out
localhost: starting nodemanager, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-nodemanager-localhost.localdomain.
out
[acadgild@localhost ~]$ jps
13617 Jps
13041 SecondaryNameNode
12754 NameNode
13284 NodeManager
12855 DataNode
13183 ResourceManager
[acadgild@localhost ~]$

[acadgild@localhost ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in
[jar:file:/home/acadgild/install/hive/apache-hive-2.3.3-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/imp
l/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.ja
r!/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-2.3.3-bin/lib/hive-common-2.3.3.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 Hive 1.X releases.

```
hive> create database acadgilddb;  
OK  
Time taken: 0.422 seconds  
hive> show databases;  
OK  
acadgilddb  
custom  
default  
Time taken: 0.054 seconds, Fetched: 3 row(s)  
hive>
```

```
hive> use acadgilddb;  
OK  
Time taken: 0.05 seconds
```

```
hive> CREATE TABLE CUSTOMER(  
  > custid INT,  
  > fname STRING,  
  > lname STRING,  
  > age INT,  
  > profession STRING)  
  > row format delimited fields terminated by ',';  
OK  
Time taken: 1.281 seconds
```

```
  > LOAD DATA LOCAL INPATH '/home/acadgild/Desktop/customer.txt' into table CUSTOMER;  
Loading data to table acadgilddb.customer  
OK  
Time taken: 3.933 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: 0.315 seconds, Fetched: 8 row(s)

hive> 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: 0.265 seconds

hive>

```
> LOAD DATA LOCAL INPATH '/home/acadgild/Desktop/transaction.txt' into tabLoading data  
to table acadgilddb.transactions
```

OK

Time taken: 1.196 seconds

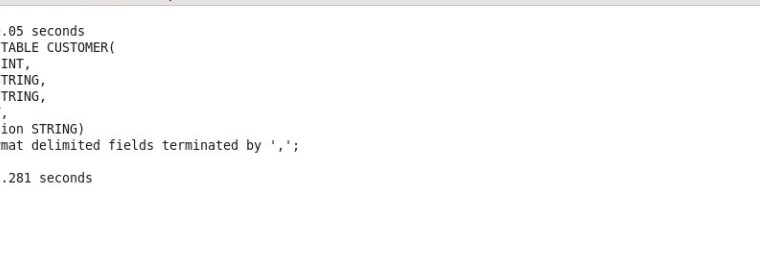
hive> SELECT * FROM TRANSACTIONS;

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

Time taken: 0.366 seconds, Fetched: 8 row(s)

OUTPUT :



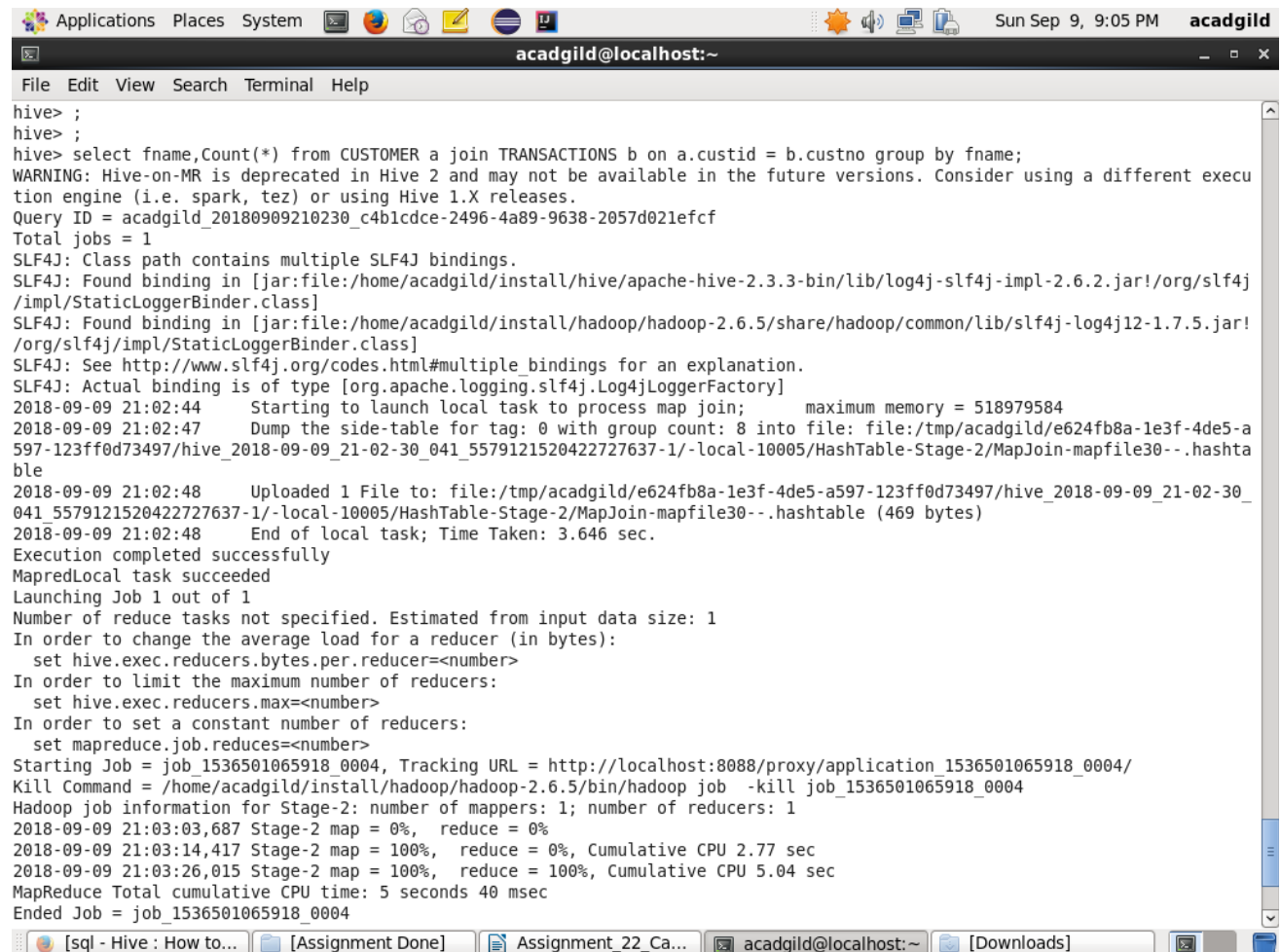
```
acadgild@localhost:~  
File Edit View Search Terminal Help  
OK  
Time taken: 0.05 seconds  
hive> CREATE TABLE CUSTOMER(  
  > custid INT,  
  > fname STRING,  
  > lname STRING,  
  > age INT,  
  > profession STRING)  
  > row format delimited fields terminated by ',';  
OK  
Time taken: 1.281 seconds  
hive>  
  >  
  >  
  >  
  >  
  >  
  >  
  >  
  >  
  > LOAD DATA LOCAL INPATH '/home/acadgild/Desktop/customer.txt' into table CUSTOMER;  
Loading data to table acadgild.db.customer  
OK  
Time taken: 3.933 seconds  
hive> SELECT * FROM CUSTOMER;  
OK  
101      Amitabh Bacchan 65      Actor  
102      Sharukh Khan  45      Doctor  
103      Akshay Kumar  38      Dentist  
104      Anubhav 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: 0.315 seconds, Fetched: 8 row(s)  
hive> CREATE TABLE TRANSACTIONS (  
  > txnno INT,
```

```
Applications  Places  System  acadgild@localhost:~
File Edit View Search Terminal Help
102 Sharukh Khan 45 Doctor
103 Akshay Kumar 38 Dentist
104 Anubhav 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: 0.315 seconds, Fetched: 8 row(s)
hive> 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: 0.265 seconds
hive>
> LOAD DATA LOCAL INPATH '/home/acadgild/Desktop/transaction.txt' into tab
Loading data to table acadgilddb.transactions
OK
Time taken: 1.196 seconds
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
45599 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: 0.366 seconds, Fetched: 8 row(s)
hive>
```

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

Terminal Execution :

```
hive> select fname,Count(*) from CUSTOMER a join TRANSACTIONS b on a.custid = b.custno
group by 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_20180909210230_c4b1cdce-2496-4a89-9638-2057d021efcf
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in
[jar:file:/home/acadgild/install/hive/apache-hive-2.3.3-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/imp
l/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.ja
r!/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]
2018-09-09 21:02:44 Starting to launch local task to process map join;    maximum memory =
518979584
2018-09-09 21:02:47 Dump the side-table for tag: 0 with group count: 8 into file:
file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-02-30_041_557912
1520422727637-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile30--.hashtable
2018-09-09 21:02:48 Uploaded 1 File to:
file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-02-30_041_557912
1520422727637-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile30--.hashtable (469 bytes)
2018-09-09 21:02:48 End of local task; Time Taken: 3.646 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_1536501065918_0004, Tracking URL =
http://localhost:8088/proxy/application_1536501065918_0004/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill
job_1536501065918_0004
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-09-09 21:03:03,687 Stage-2 map = 0%, reduce = 0%
2018-09-09 21:03:14,417 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.77 sec
2018-09-09 21:03:26,015 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 5.04 sec
```

```
Applications Places System [Icons] [System Tray] Sun Sep 9, 9:06 PM acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2018-09-09 21:02:44 Starting to launch local task to process map join; maximum memory = 518979584
2018-09-09 21:02:47 Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-02-30_041_5579121520422727637-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile30--.hashtable
2018-09-09 21:02:48 Uploaded 1 File to: file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-02-30_041_5579121520422727637-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile30--.hashtable (469 bytes)
2018-09-09 21:02:48 End of local task; Time Taken: 3.646 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_1536501065918_0004, Tracking URL = http://localhost:8088/proxy/application_1536501065918_0004/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1536501065918_0004
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-09-09 21:03:03,687 Stage-2 map = 0%, reduce = 0%
2018-09-09 21:03:14,417 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.77 sec
2018-09-09 21:03:26,015 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 5.04 sec
MapReduce Total cumulative CPU time: 5 seconds 40 msec
Ended Job = job_1536501065918_0004
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 5.04 sec HDFS Read: 12690 HDFS Write: 235 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 40 msec
OK
Aamir 1
Amitabh 2
Anubhav 1
Pawan 1
Ranbir 1
Salman 1
Sharukh 1
Time taken: 58.529 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)

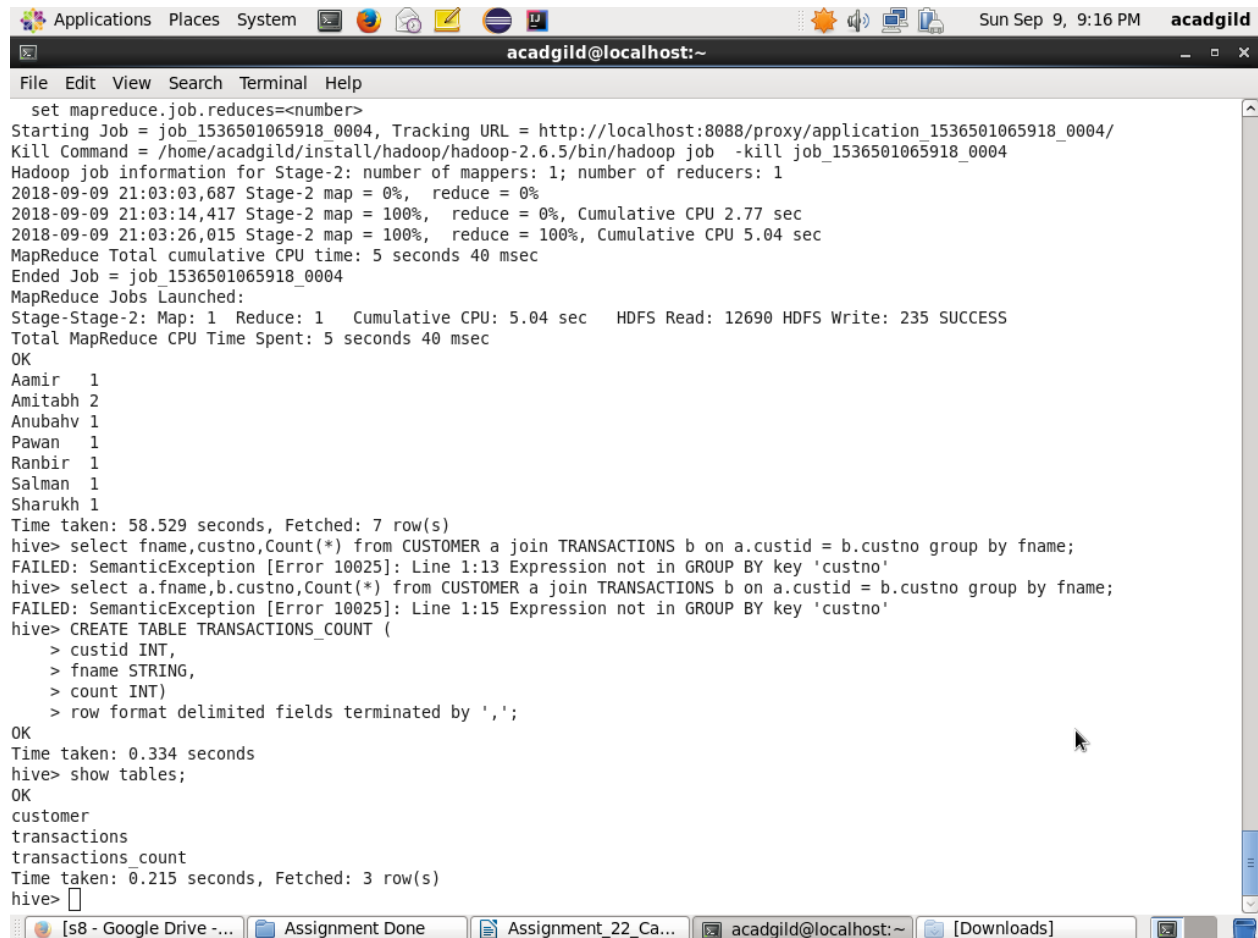
Terminal Execution :

```
hive> CREATE TABLE TRANSACTIONS_COUNT (  
  > custid INT,  
  > fname STRING,  
  > count INT)  
  > row format delimited fields terminated by ',';
```

OK

Time taken: 0.334 seconds

OUTPUT :



```
set mapreduce.job.reduces=<number>  
Starting Job = job_1536501065918_0004, Tracking URL = http://localhost:8088/proxy/application_1536501065918_0004/  
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1536501065918_0004  
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1  
2018-09-09 21:03:03,687 Stage-2 map = 0%, reduce = 0%  
2018-09-09 21:03:14,417 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.77 sec  
2018-09-09 21:03:26,015 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 5.04 sec  
MapReduce Total cumulative CPU time: 5 seconds 40 msec  
Ended Job = job_1536501065918_0004  
MapReduce Jobs Launched:  
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 5.04 sec HDFS Read: 12690 HDFS Write: 235 SUCCESS  
Total MapReduce CPU Time Spent: 5 seconds 40 msec  
OK  
Aamir 1  
Amitabh 2  
Anubahv 1  
Pawan 1  
Ranbir 1  
Salman 1  
Sharukh 1  
Time taken: 58.529 seconds, Fetched: 7 row(s)  
hive> select fname,custno,Count(*) from CUSTOMER a join TRANSACTIONS b on a.custid = b.custno group by fname;  
FAILED: SemanticException [Error 10025]: Line 1:13 Expression not in GROUP BY key 'custno'  
hive> select a.fname,b.custno,Count(*) from CUSTOMER a join TRANSACTIONS b on a.custid = b.custno group by fname;  
FAILED: SemanticException [Error 10025]: Line 1:15 Expression not in GROUP BY key 'custno'  
hive> CREATE TABLE TRANSACTIONS_COUNT (  
  > custid INT,  
  > fname STRING,  
  > count INT)  
  > row format delimited fields terminated by ',';  
OK  
Time taken: 0.334 seconds  
hive> show tables;  
OK  
customer  
transactions  
transactions_count  
Time taken: 0.215 seconds, Fetched: 3 row(s)  
hive>
```

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).

Terminal Execution :

```
hive> CREATE TABLE TRANSACTIONS_COUNT (
  > custid INT,
  > fname STRING,
  > count INT)
  > row format delimited fields terminated by ',';
OK
Time taken: 0.334 seconds
hive> show tables;
OK
customer
transactions
transactions_count
Time taken: 0.215 seconds, Fetched: 3 row(s)
hive> select a.custid,a.fname,Count(*) 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 Hive 1.X releases.
Query ID = acadgild_20180909211928_662c1ec2-e2a0-4b20-9ee2-dc0cfd5878fd
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in
[jar:file:/home/acadgild/install/hive/apache-hive-2.3.3-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/imp
l/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.ja
r!/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]
2018-09-09 21:19:44 Starting to launch local task to process map join;    maximum memory =
518979584
2018-09-09 21:19:47 Dump the side-table for tag: 0 with group count: 8 into file:
file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-19-28_935_830969
2770408811156-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile40--.hashtable
2018-09-09 21:19:47 Uploaded 1 File to:
file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-19-28_935_830969
2770408811156-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile40--.hashtable (469 bytes)
2018-09-09 21:19:47 End of local task; Time Taken: 3.284 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_1536501065918_0005, Tracking URL =

http://localhost:8088/proxy/application_1536501065918_0005/

Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill

job_1536501065918_0005

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

2018-09-09 21:20:01,310 Stage-2 map = 0%, reduce = 0%

2018-09-09 21:20:11,932 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 3.02 sec

2018-09-09 21:20:23,378 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 5.29 sec

MapReduce Total cumulative CPU time: 5 seconds 290 msec

Ended Job = job_1536501065918_0005

MapReduce Jobs Launched:

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 5.29 sec HDFS Read: 13091 HDFS Write: 263 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 290 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: 55.665 seconds, Fetched: 7 row(s)

hive> INSERT OVERWRITE TABLE TRANSACTIONS_COUNT

> select a.custid,a.fname,Count(*) 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 Hive 1.X releases.

Query ID = acadgild_20180909213126_6f098c97-1b14-45fa-8c06-06c813dd40bf

Total jobs = 1

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in

[jar:file:/home/acadgild/install/hive/apache-hive-2.3.3-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]

2018-09-09 21:31:47 Starting to launch local task to process map join; maximum memory = 518979584

2018-09-09 21:31:51 Dump the side-table for tag: 0 with group count: 8 into file:

```

file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-31-26_176_805639
9582483140555-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile50--.hashtable
2018-09-09 21:31:51 Uploaded 1 File to:
file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-31-26_176_805639
9582483140555-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile50--.hashtable (469 bytes)
2018-09-09 21:31:51 End of local task; Time Taken: 4.071 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_1536501065918_0006, Tracking URL =
http://localhost:8088/proxy/application_1536501065918_0006/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill
job_1536501065918_0006
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-09-09 21:32:08,926 Stage-2 map = 0%, reduce = 0%
2018-09-09 21:32:20,672 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.86 sec
2018-09-09 21:32:32,470 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 6.16 sec
MapReduce Total cumulative CPU time: 6 seconds 160 msec
Ended Job = job_1536501065918_0006
Loading data to table acadgilddb.transactions_count
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 6.16 sec HDFS Read: 13825 HDFS Write:
177 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 160 msec
OK
Time taken: 70.939 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: 0.416 seconds, Fetched: 7 row(s)

```

OUTPUT:

```
Applications Places System Sun Sep 9, 9:20 PM acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
hive> show tables;
OK
customer
transactions
transactions count
Time taken: 0.215 seconds, Fetched: 3 row(s)
hive> select a.custid,a.fname,Count(*) 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 Hive 1.X releases.
Query ID = acadgild_20180909211928_662c1ec2-e2a0-4b20-9ee2-dc0cfd5878fd
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.3-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]
2018-09-09 21:19:44 Starting to launch local task to process map join; maximum memory = 518979584
2018-09-09 21:19:47 Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-19-28_935_8309692770408811156-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile40--.hashtable (469 bytes)
2018-09-09 21:19:47 End of local task; Time Taken: 3.284 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_1536501065918_0005, Tracking URL = http://localhost:8088/proxy/application_1536501065918_0005/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1536501065918_0005
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-09-09 21:20:01,310 Stage-2 map = 0%, reduce = 0%
```

```
Applications Places System Sun Sep 9, 9:21 PM acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2018-09-09 21:19:44 Starting to launch local task to process map join; maximum memory = 518979584
2018-09-09 21:19:47 Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-19-28_935_8309692770408811156-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile40--.hashtable (469 bytes)
2018-09-09 21:19:47 End of local task; Time Taken: 3.284 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_1536501065918_0005, Tracking URL = http://localhost:8088/proxy/application_1536501065918_0005/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1536501065918_0005
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-09-09 21:20:01,310 Stage-2 map = 0%, reduce = 0%
2018-09-09 21:20:11,932 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 3.02 sec
2018-09-09 21:20:23,378 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 5.29 sec
MapReduce Total cumulative CPU time: 5 seconds 290 msec
Ended Job = job_1536501065918_0005
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 5.29 sec HDFS Read: 13091 HDFS Write: 263 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 290 msec
OK
101 Amitabh 2
102 Sharukh 1
104 Anubhav 1
105 Pawan 1
106 Aamir 1
107 Salman 1
108 Ranbir 1
Time taken: 55.665 seconds, Fetched: 7 row(s)
hive>
```



```
Applications Places System Sun Sep 9, 9:32 PM acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
107 Salman 1
108 Ranbir 1
Time taken: 55.665 seconds, Fetched: 7 row(s)
hive> INSERT OVERWRITE TABLE TRANSACTIONS_COUNT
> select a.custid,a.fname,Count(*) 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 Hive 1.X releases.
Query ID = acadgild_20180909213126_6f098c97-1b14-45fa-8c06-06c813dd40bf
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.3-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]
2018-09-09 21:31:47 Starting to launch local task to process map join; maximum memory = 518979584
2018-09-09 21:31:51 Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-31-26_176_8056399582483140555-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile50--.hashtable
2018-09-09 21:31:51 Uploaded 1 File to: file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-31-26_176_8056399582483140555-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile50--.hashtable (469 bytes)
2018-09-09 21:31:51 End of local task; Time Taken: 4.071 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_1536501065918_0006, Tracking URL = http://localhost:8088/proxy/application_1536501065918_0006/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1536501065918_0006
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-09-09 21:32:08,926 Stage-2 map = 0%, reduce = 0%
2018-09-09 21:32:20,672 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.86 sec
[Student Dashboard... Assignment Done Assignment_22_Ca... acadgild@localhost:~ [Downloads]
```

```
Applications Places System Sun Sep 9, 9:33 PM acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.3-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]
2018-09-09 21:31:47 Starting to launch local task to process map join; maximum memory = 518979584
2018-09-09 21:31:51 Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-31-26_176_8056399582483140555-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile50--.hashtable
2018-09-09 21:31:51 Uploaded 1 File to: file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-31-26_176_8056399582483140555-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile50--.hashtable (469 bytes)
2018-09-09 21:31:51 End of local task; Time Taken: 4.071 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_1536501065918_0006, Tracking URL = http://localhost:8088/proxy/application_1536501065918_0006/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1536501065918_0006
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-09-09 21:32:08,926 Stage-2 map = 0%, reduce = 0%
2018-09-09 21:32:20,672 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.86 sec
2018-09-09 21:32:32,470 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 6.16 sec
MapReduce Total cumulative CPU time: 6 seconds 160 msec
Ended Job = job_1536501065918_0006
Loading data to table acadgild.transactions_count
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 6.16 sec HDFS Read: 13825 HDFS Write: 177 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 160 msec
OK
Time taken: 70.939 seconds
hive>
```



```
Applications Places System [Icons] [System Tray] Sun Sep 9, 9:34 PM acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
ble
2018-09-09 21:31:51 Uploaded 1 File to: file:/tmp/acadgild/e624fb8a-1e3f-4de5-a597-123ff0d73497/hive_2018-09-09_21-31-26_
176_8056399582483140555-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile50--.hashtable (469 bytes)
2018-09-09 21:31:51 End of local task; Time Taken: 4.071 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_1536501065918_0006, Tracking URL = http://localhost:8088/proxy/application_1536501065918_0006/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1536501065918_0006
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-09-09 21:32:08,926 Stage-2 map = 0%, reduce = 0%
2018-09-09 21:32:20,672 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.86 sec
2018-09-09 21:32:32,470 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 6.16 sec
MapReduce Total cumulative CPU time: 6 seconds 160 msec
Ended Job = job_1536501065918_0006
Loading data to table acadgild.db.transactions_count
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 6.16 sec HDFS Read: 13825 HDFS Write: 177 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 160 msec
OK
Time taken: 70.939 seconds
hive> SELECT * FROM TRANSACTIONS_COUNT;
OK
101 Amitabh 2
102 Sharukh 1
104 Anubhav 1
105 Pawan 1
106 Aamir 1
107 Salman 1
108 Ranbir 1
Time taken: 0.416 seconds, Fetched: 7 row(s)
hive>
```

[Student Dashboard...] [Assignment Done] [Assignment_22_Ca...] acadgild@localhost:~ [Downloads]

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)

Explanation:

For integrating HBase with Hive, Storage Handlers in Hive is used.

Storage Handlers are a combination of InputFormat, OutputFormat, SerDe, and specific code that Hive uses to identify an external entity as a Hive table. This allows the user to issue SQL queries seamlessly, whether the table represents a text file stored in Hadoop or a column family stored in a NoSQL database such as Apache Hbase, Apache Cassandra, and Amazon DynamoDB. Storage Handlers are not only limited to NoSQL databases, a storage handler could be designed for several different kinds of data stores.

Here the example for connecting Hive with HBase using HiveStorageHandler.

Create the HBase Table:

```
create 'TRANSACTIONS','customertransdetails'
```

The above statement will create ‘TRANSACTIONS’ with one columns families that is

Now create the Hive table pointing to HBase table.

If there are multiple columns family in Hbase, then create one table for each column families. In this case, there is 1 column families and hence, 1 table is required to be created, for one column family.

```
create external table TRANSACTIONS_COUNT(custid INT, fname STRING, count INT)
STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
with serdeproperties
("hbase.columns.mapping"=":key,customertransdetails:custid,customertransdetails:fname,customertr
ansdetails: count")
tblproperties("hbase.table.name"="TRANSACTIONS");
STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
```

If the Non-native Hive table is created using Storage Handler then STORED BY clause should be specified.

hbase.columns.mapping : It is used to map the Hive columns with the HBase columns. The first column must be the key column which would also be same as the HBase’s row key column.

Terminal Execution :

```
[acadgild@localhost ~]$ jps
22448 Jps
21573 NameNode
22007 ResourceManager
22121 NodeManager
21833 SecondaryNameNode
21674 DataNode
```

```
[acadgild@localhost ~]$ start-hbase.sh
localhost: running zookeeper, logging to
/home/acadgild/install/hbase/hbase-1.4.4/logs/hbase-acadgild-zookeeper-localhost.localdomain.out
running master, logging to
/home/acadgild/install/hbase/hbase-1.4.4/logs/hbase-acadgild-master-localhost.localdomain.out
: running regionserver, logging to
/home/acadgild/install/hbase/hbase-1.4.4/logs/hbase-acadgild-regionserver-localhost.localdomain.out
```

```
[acadgild@localhost ~]$ jps
22757 HMaster
21573 NameNode
22007 ResourceManager
22856 HRegionServer
22664 HQuorumPeer
22121 NodeManager
21833 SecondaryNameNode
21674 DataNode
23098 Jps
```

```
[acadgild@localhost ~]$ hbase shell
2018-09-10 00:23:05,072 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in
[jar:file:/home/acadgild/install/hbase/hbase-1.4.4/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLo
ggerBinder.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.slf4j.impl.Log4jLoggerFactory]
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 1.4.4, rfe146eb48c24d56dbcd2f669bb5ff8197e6c918b, Sun Apr 22 20:42:02 PDT 2018
```

```
hbase(main):001:0>
```

```
=> Hbase::Table - TRANSACTIONS
hbase(main):002:0>
```

OUTPUT :

The screenshot shows a terminal window titled 'acadgild@localhost:~'. The user has installed HBase and is configuring it. The terminal output shows the following commands and their results:

```

21674 DataNode
[acadgild@localhost ~]$ start-hbase.sh
localhost: running zookeeper, logging to /home/acadgild/install/hbase/hbase-1.4.4/logs/hbase-acadgild-zookeeper-localhost.localdomain.out
running master, logging to /home/acadgild/install/hbase/hbase-1.4.4/logs/hbase-acadgild-master-localhost.localdomain.out
: running regionserver, logging to /home/acadgild/install/hbase/hbase-1.4.4/logs/hbase-acadgild-regionserver-localhost.localdomain.out
[acadgild@localhost ~]$ jps
22757 HMaster
21573 NameNode
22007 ResourceManager
22856 HRegionServer
22664 HQuorumPeer
22121 NodeManager
21833 SecondaryNameNode
21674 DataNode
23098 Jps
[acadgild@localhost ~]$ create 'TRANSACTIONS','customertransdetails'
bash: create: command not found
[acadgild@localhost ~]$ hbase-shell
bash: hbase-shell: command not found
[acadgild@localhost ~]$ hbase shell
2018-09-10 00:23:05,072 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.4.4/lib/slf4j-log4j12-1.7.10.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.slf4j.impl.Log4jLoggerFactory]
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 1.4.4, rfe146eb48c24d56dbcd2f669bb5ff8197e6c918b, Sun Apr 22 20:42:02 PDT 2018

hbase(main):001:0> create 'TRANSACTIONS','customertransdetails'
0 row(s) in 2.0610 seconds

```

The terminal window also shows the top of the file manager and the taskbar at the bottom with various open applications like 'Integrating...', 'Assignment...', 'Downloads', and 'PDF'.

```
Applications Places System Mon Sep 10, 1:10 AM acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
0 row(s) in 2.0610 seconds

=> Hbase::Table - TRANSACTIONS
hbase(main):002:0> exit
[acadgild@localhost ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.3-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]

Logging initialized using configuration in jar:file:/home/acadgild/install/hive/apache-hive-2.3.3-bin/lib/hive-common-2.3.3.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 Hive 1.X releases.
hive> create external table TRANSACTIONS COUNT(custid INT, fname STRING, count INT)
> STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
> with serdeproperties ("hbase.columns.mapping"=":key,customertransdetails:custid,customertransdetails:fname,customertransdetails: count")
> tblproperties("hbase.table.name"="TRANSACTIONS");
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. java.lang.RuntimeException: MetaException (message:org.apache.hadoop.hive.serde2.SerDeException org.apache.hadoop.hive.hbase.HBaseSerDe: columns has 3 elements while hbase.columns.mapping has 4 elements (counting the key if implicit))
hive> STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
>
NoViableAltException(275@[])
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:1300)
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:208)
at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:77)
at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:70)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:468)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1317)
at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1457)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1237)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1227)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:233)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:184)
```

```
Applications Places System Mon Sep 10, 1:11 AM acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:0 cannot recognize input near 'STORED' 'BY' 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
'
hive> create external table TRANSACTIONS COUNT HBASE(custid INT, fname STRING, count INT)
> STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
> with serdeproperties ("hbase.columns.mapping"=":key,customertransdetails:custid,customertransdetails:fname,customertransdetails: count")
> tblproperties("hbase.table.name"="TRANSACTIONS");
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. java.lang.RuntimeException: MetaException (message:org.apache.hadoop.hive.serde2.SerDeException org.apache.hadoop.hive.hbase.HBaseSerDe: columns has 3 elements while hbase.columns.mapping has 4 elements (counting the key if implicit))
hive> create external table TRANSACTIONS COUNT_HBASE(custid INT, fname STRING, count INT)
> with serdeproperties ("hbase.columns.mapping"=":key,customertransdetails:fname,customertransdetails: count")
> STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
> ;;;
FAILED: ParseException line 2:1 missing EOF at 'with' near ''
hive> create external table TRANSACTIONS COUNT_HBASE(custid INT, fname STRING, count INT)
> STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
> with serdeproperties ("hbase.columns.mapping"=":key,customertransdetails:fname,customertransdetails: count")
> tblproperties("hbase.table.name"="TRANSACTIONS");
OK
Time taken: 3.704 seconds
hive> STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
>
NoViableAltException(275@[])
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:1300)
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:208)
at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:77)
at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:70)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:468)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1317)
at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1457)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1237)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1227)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:233)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:184)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:403)
at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:821)
```

```
Applications Places System acadgild@localhost:~
File Edit View Search Terminal Help
hive> select a.custid,a.fname,Count(*) from acadgilddb.CUSTOMER a join acadgilddb.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 Hive 1.X releases.
Query ID = acadgild_20180910004840_0d948595-b712-4a0d-8fcc-ef0f844074e1
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.3-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]
2018-09-10 00:48:57 Starting to launch local task to process map join; maximum memory = 518979584
2018-09-10 00:49:00 Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/a4081cac-7e8b-46e8-9ecf-a0af83235fch/hive_2018-09-10_00-48-40_126_1532323502154986158-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile000--.hashtable
2018-09-10 00:49:00 Uploaded 1 File to: file:/tmp/acadgild/a4081cac-7e8b-46e8-9ecf-a0af83235fch/hive_2018-09-10_00-48-40_126_1532323502154986158-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile000--.hashtable (469 bytes)
2018-09-10 00:49:00 End of local task; Time Taken: 3.649 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_1536517870700_0001, Tracking URL = http://localhost:8088/proxy/application_1536517870700_0001/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1536517870700_0001
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-09-10 00:49:28,022 Stage-2 map = 0%, reduce = 0%
2018-09-10 00:49:40,453 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.9 sec
2018-09-10 00:49:55,042 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 5.25 sec
MapReduce Total cumulative CPU time: 5 seconds 250 msec
Ended Job = job_1536517870700_0001
MapReduce Jobs Launched:
```

```
Applications Places System acadgild@localhost:~
File Edit View Search Terminal Help
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2018-09-10 00:48:57 Starting to launch local task to process map join; maximum memory = 518979584
2018-09-10 00:49:00 Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/a4081cac-7e8b-46e8-9ecf-a0af83235fch/hive_2018-09-10_00-48-40_126_1532323502154986158-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile000--.hashtable
2018-09-10 00:49:00 Uploaded 1 File to: file:/tmp/acadgild/a4081cac-7e8b-46e8-9ecf-a0af83235fch/hive_2018-09-10_00-48-40_126_1532323502154986158-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile000--.hashtable (469 bytes)
2018-09-10 00:49:00 End of local task; Time Taken: 3.649 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_1536517870700_0001, Tracking URL = http://localhost:8088/proxy/application_1536517870700_0001/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1536517870700_0001
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-09-10 00:49:28,022 Stage-2 map = 0%, reduce = 0%
2018-09-10 00:49:40,453 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.9 sec
2018-09-10 00:49:55,042 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 5.25 sec
MapReduce Total cumulative CPU time: 5 seconds 250 msec
Ended Job = job_1536517870700_0001
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 5.25 sec HDFS Read: 13091 HDFS Write: 263 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 250 msec
OK
101 Amitabh 2
102 Sharukh 1
104 Anubhav 1
105 Pawan 1
106 Aamir 1
107 Salman 1
108 Ranbir 1
Time taken: 76.252 seconds, Fetched: 7 row(s)
```

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)

Now the query of the HBase table with SQL queries in hive using the below command :

```
INSERT OVERWRITE TABLE TRANSACTIONS_COUNT_HBASE
> select a.custid,a.fname,Count(*) from acadgilddb.CUSTOMER a join
acadgilddb.TRANSACTIONS b on a.custid = b.custno group by a.custid,a.fname;
```

Terminal Execution :

```
hive> select * from TRANSACTIONS_COUNT_HBASE;
```

```
OK
```

```
Time taken: 4.688 seconds
```

```
hive>
```

```
hive> INSERT OVERWRITE TABLE TRANSACTIONS_COUNT_HBASE
```

```
> select a.custid,a.fname,Count(*) from acadgilddb.CUSTOMER a join
acadgilddb.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 Hive 1.X releases.
```

```
Query ID = acadgild_20180910005317_07b9661c-b335-4c55-b20f-099bd3e9c234
```

```
Total jobs = 1
```

```
SLF4J: Class path contains multiple SLF4J bindings.
```

```
SLF4J: Found binding in
```

```
[jar:file:/home/acadgild/install/hive/apache-hive-2.3.3-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]
```

```
2018-09-10 00:53:33 Starting to launch local task to process map join; maximum memory = 518979584
```

```
2018-09-10 00:53:37 Dump the side-table for tag: 0 with group count: 8 into file:
```

```
file:/tmp/acadgild/a4081cac-7e8b-46e8-9ecf-a0af83235fcf/hive_2018-09-10_00-53-17_282_248955155297088385-1/-local-10002/HashTable-Stage-4/MapJoin-mapfile10--.hashtable
```

```
2018-09-10 00:53:37 Uploaded 1 File to:
```

```
file:/tmp/acadgild/a4081cac-7e8b-46e8-9ecf-a0af83235fcf/hive_2018-09-10_00-53-17_282_248955155297088385-1/-local-10002/HashTable-Stage-4/MapJoin-mapfile10--.hashtable (469 bytes)
```

```
2018-09-10 00:53:37 End of local task; Time Taken: 3.913 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_1536517870700_0002, Tracking URL =

http://localhost:8088/proxy/application_1536517870700_0002/

Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill
job_1536517870700_0002

Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 1

2018-09-10 00:54:01,006 Stage-4 map = 0%, reduce = 0%

2018-09-10 00:54:13,018 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 2.85 sec

2018-09-10 00:54:27,225 Stage-4 map = 100%, reduce = 100%, Cumulative CPU 6.69 sec

MapReduce Total cumulative CPU time: 7 seconds 420 msec

Ended Job = job_1536517870700_0002

MapReduce Jobs Launched:

Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 7.42 sec HDFS Read: 13994 HDFS Write: 0
SUCCESS

Total MapReduce CPU Time Spent: 7 seconds 420 msec

OK

Time taken: 73.464 seconds

```
hive> select * from TRANSACTIONS_COUNT_HBASE;
```

OK

101	Amitabh	2
102	Sharukh	1
104	Anubahv	1
105	Pawan	1
106	Aamir	1
107	Salman	1
108	Ranbir	1

Time taken: 0.66 seconds, Fetched: 7 row(s)

```
hbase(main):001:0> scan 'TRANSACTIONS',{NAME => 'customertransdetails',VERSIONS =>5}
```

ROW	COLUMN+CELL
101	column=customertransdetails: count, timestamp=1536521068035, value=2
101	column=customertransdetails:fname, timestamp=1536521068035,
value=Amitabh	
102	column=customertransdetails: count, timestamp=1536521068035, value=1
102	column=customertransdetails:fname, timestamp=1536521068035,
value=Sharukh	
104	column=customertransdetails: count, timestamp=1536521068035, value=1
104	column=customertransdetails:fname, timestamp=1536521068035,
value=Anubahv	
105	column=customertransdetails: count, timestamp=1536521068035, value=1
105	column=customertransdetails:fname, timestamp=1536521068035,

value=Pawan

106 column=customertransdetails: count, timestamp=1536521068035, value=1

106 column=customertransdetails:fname, timestamp=1536521068035,

value=Aamir

107 column=customertransdetails: count, timestamp=1536521068035, value=1

107 column=customertransdetails:fname, timestamp=1536521068035,

value=Salman

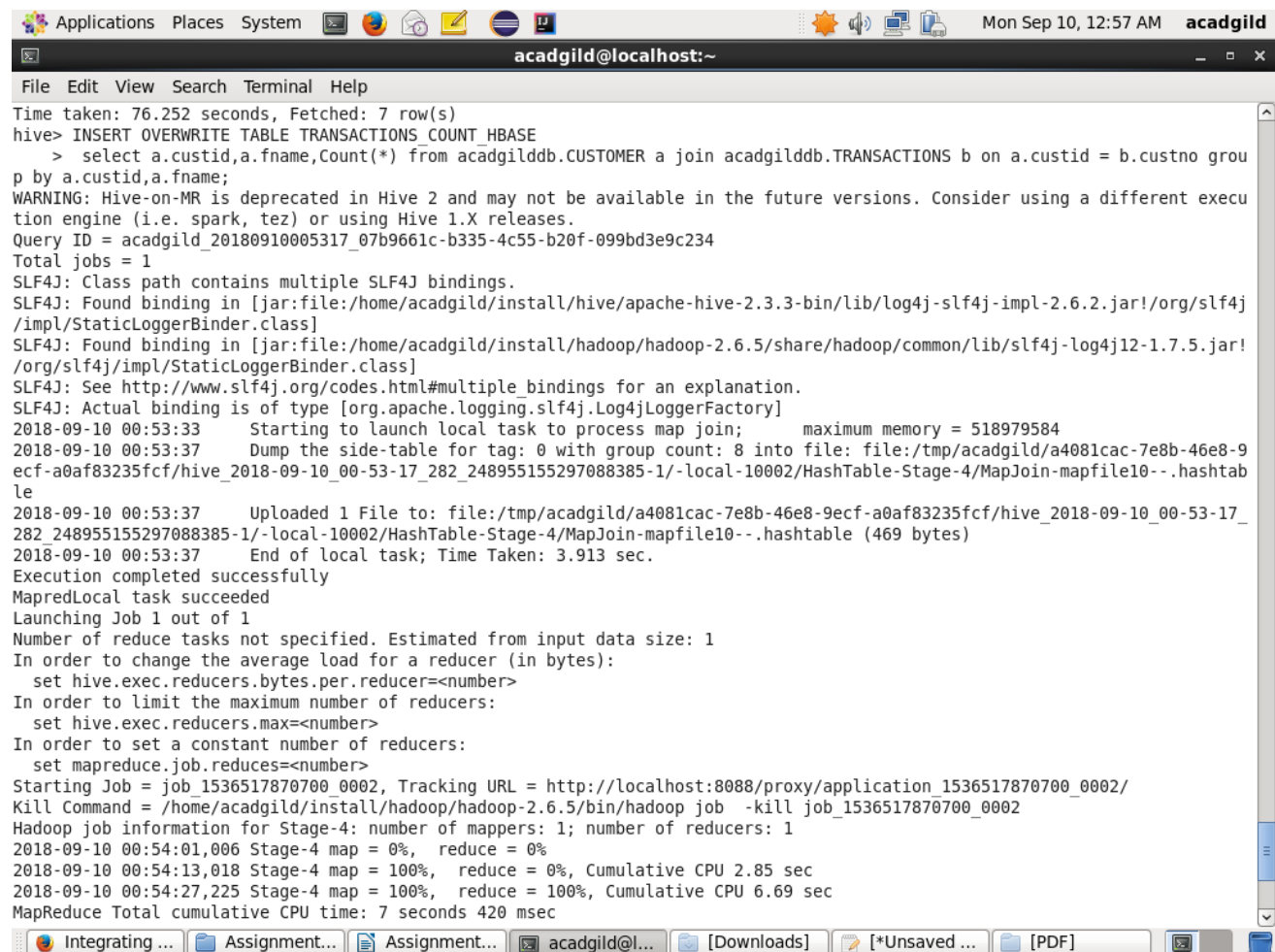
108 column=customertransdetails: count, timestamp=1536521068035, value=1

108 column=customertransdetails:fname, timestamp=1536521068035,

value=Ranbir

7 row(s) in 0.7570 seconds

OUTPUT :



```
Applications Places System [Icons] [Network] [Sound] [Volume] [Power] [Brightness] [Temperature] [Battery] [Wi-Fi] [Bluetooth] [USB] [HDMI] [DVI] [VGA] [Audio] [Video] [Mouse] [Keyboard] [Touchpad] [Trackball] [Joystick] [Gamepad] [Webcam] [Microphone] [Headset] [Speaker] [Printer] [Scanner] [Firmware] [Drivers] [Updates] [Settings] [System] [Users] [Groups] [Permissions] [Network] [Storage] [Security] [Performance] [Troubleshooting] [Help]

acadgild@localhost:~
File Edit View Search Terminal Help
Time taken: 76.252 seconds, Fetched: 7 row(s)
hive> INSERT OVERWRITE TABLE TRANSACTIONS COUNT_HBASE
> select a.custid,a.fname,Count(*) from acadgilddb.CUSTOMER a join acadgilddb.TRANSACTIONS b on a.custid = b.custno group
p 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 execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180910005317_07b9661c-b335-4c55-b20f-099bd3e9c234
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.3-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]
2018-09-10 00:53:33 Starting to launch local task to process map join; maximum memory = 518979584
2018-09-10 00:53:37 Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/a4081cac-7e8b-46e8-9
ecf-a0af83235fcf/hive_2018-09-10_00-53-17_282_248955155297088385-1/-local-10002/HashTable-Stage-4/MapJoin-mapfile10--.hashtab
le
2018-09-10 00:53:37 Uploaded 1 File to: file:/tmp/acadgild/a4081cac-7e8b-46e8-9ecf-a0af83235fcf/hive_2018-09-10_00-53-17_
282_248955155297088385-1/-local-10002/HashTable-Stage-4/MapJoin-mapfile10--.hashtable (469 bytes)
2018-09-10 00:53:37 End of local task; Time Taken: 3.913 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_1536517870700_0002, Tracking URL = http://localhost:8088/proxy/application_1536517870700_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1536517870700_0002
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 1
2018-09-10 00:54:01,006 Stage-4 map = 0%, reduce = 0%
2018-09-10 00:54:13,018 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 2.85 sec
2018-09-10 00:54:27,225 Stage-4 map = 100%, reduce = 100%, Cumulative CPU 6.69 sec
MapReduce Total cumulative CPU time: 7 seconds 420 msec
```

```
Applications Places System acadgild@localhost:~
File Edit View Search Terminal Help
ecf-a0af83235fcf/hive_2018-09-10_00-53-17_282_248955155297088385-1/-local-10002/HashTable-Stage-4/MapJoin-mapfile10--.hashtab
le
2018-09-10 00:53:37      Uploaded 1 File to: file:/tmp/acadgild/a4081cac-7e8b-46e8-9ecf-a0af83235fcf/hive_2018-09-10_00-53-17_
282_248955155297088385-1/-local-10002/HashTable-Stage-4/MapJoin-mapfile10--.hashtable (469 bytes)
2018-09-10 00:53:37      End of local task; Time Taken: 3.913 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_1536517870700_0002, Tracking URL = http://localhost:8088/proxy/application_1536517870700_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1536517870700_0002
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 1
2018-09-10 00:54:01,006 Stage-4 map = 0%, reduce = 0%
2018-09-10 00:54:13,018 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 2.85 sec
2018-09-10 00:54:27,225 Stage-4 map = 100%, reduce = 100%, Cumulative CPU 6.69 sec
MapReduce Total cumulative CPU time: 7 seconds 420 msec
Ended Job = job_1536517870700_0002
MapReduce Jobs Launched:
Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 7.42 sec HDFS Read: 13994 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 420 msec
OK
Time taken: 73.464 seconds
hive> select * from TRANSACTIONS_COUNT_HBASE;
OK
101      Amitabh 2
102      Sharukh 1
104      Anubahv 1
105      Pawan 1
106      Aamir 1
107      Salman 1
108      Ranbir 1
Time taken: 0.66 seconds, Fetched: 7 row(s)
hive>
```

```
Applications Places System acadgild@localhost:~
File Edit View Search Terminal Help
107      Salman 1
108      Ranbir 1
Time taken: 0.66 seconds, Fetched: 7 row(s)
hive> exit
>
[acadgild@localhost ~]$ hbase shell
2018-09-10 00:58:53,057 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using b
uilt-in java classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.4.4/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/Stati
cLoggerBinder.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.slf4j.impl.Log4jLoggerFactory]
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 1.4.4, rfe146eb48c24d56dbcd2f669bb5ff8197e6c918b, Sun Apr 22 20:42:02 PDT 2018

hbase(main):001:0> scan 'TRANSACTIONS',{NAME => 'customertransdetails',VERSIONS =>5}
ROW COLUMN+CELL
101      column=customertransdetails:count,timestamp=1536521068035,value=2
101      column=customertransdetails:fname,timestamp=1536521068035,value=Amitabh
102      column=customertransdetails:count,timestamp=1536521068035,value=1
102      column=customertransdetails:fname,timestamp=1536521068035,value=Sharukh
104      column=customertransdetails:count,timestamp=1536521068035,value=1
104      column=customertransdetails:fname,timestamp=1536521068035,value=Anubahv
105      column=customertransdetails:count,timestamp=1536521068035,value=1
105      column=customertransdetails:fname,timestamp=1536521068035,value=Pawan
106      column=customertransdetails:count,timestamp=1536521068035,value=1
106      column=customertransdetails:fname,timestamp=1536521068035,value=Aamir
107      column=customertransdetails:count,timestamp=1536521068035,value=1
107      column=customertransdetails:fname,timestamp=1536521068035,value=Salman
108      column=customertransdetails:count,timestamp=1536521068035,value=1
108      column=customertransdetails:fname,timestamp=1536521068035,value=Ranbir
7 row(s) in 0.7570 seconds

hbase(main):002:0>
```

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

Solution:

```
package com.hbase.api;

import java.io.IOException;
import java.util.ArrayList;
import java.util.List;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.hbase.HBaseConfiguration;
import org.apache.hadoop.hbase.client.HTable;
import org.apache.hadoop.hbase.client.Result;
import org.apache.hadoop.hbase.client.ResultScanner;
import org.apache.hadoop.hbase.client.Scan;
import org.apache.hadoop.hbase.filter.Filter;
import org.apache.hadoop.hbase.filter.FilterBase;
import org.apache.hadoop.hbase.filter.FilterList;
import org.apache.hadoop.hbase.filter.RegexStringComparator;
import org.apache.hadoop.hbase.filter.RowFilter;
import org.apache.hadoop.hbase.filter.SubstringComparator;
import org.apache.hadoop.hbase.filter.ValueFilter;
import org.apache.hadoop.hbase.filter.CompareFilter;
import org.apache.hadoop.hbase.filter.CompareFilter.CompareOp;
import org.apache.hadoop.hbase.util.Bytes;
import org.jruby.compiler.ir.operands.Array;

public class ScanData {

    public static void main(String args[]) throws IOException {
        Configuration conf = HBaseConfiguration.create();
        @SuppressWarnings("deprecation")
        HTable table = new HTable(conf, "TRANSACTIONS");
        Scan scan = new Scan();
        scan.addColumn(Bytes.toBytes("customertransdetails"), Bytes.toBytes("count"));
        scan.addColumn(Bytes.toBytes("customertransdetails"), Bytes.toBytes("fname"));
        ResultScanner result = table.getScanner(scan);
        for(Result res:result){
            byte[] val = res.getValue(Bytes.toBytes("customertransdetails"), Bytes.toBytes("count"));
            byte[] val1 = res.getValue(Bytes.toBytes("customertransdetails"), Bytes.toBytes("fname"));
            System.out.println("Row-value : "+Bytes.toString(val));
            System.out.println("Row-value : "+Bytes.toString(val1));
            System.out.println(res);
        }
        table.close();
    }
}
```

}

Explanation:

- In line 1, we are declaring a class name ScanData.
- In line 3, the *Configuration* class adds HBase configuration resources to its object *conf* with the help of *create()* method of the HBaseConfiguration class.
- In line 4, the class HTable instance “*table*” will allow to communicate with a single HBase table, it accepts configuration object and the table name as the parameters.
- In line 5, we are creating class Scan “*scan*” instance to perform Scan operations.
- In line 6, we are using addColumn method to column in the table “**TRANSACTIONS**”, where “**customertransdetails**” is the column family name and “count” is the column qualifier name of the column family “**customertransdetails**”.
- In line 7, we are using addColumn method to column in the table “**TRANSACTIONS**”, where “**customertransdetails**” is the column family name and “fname” is the column qualifier name of the column family “**customertransdetails**”.
- In line 8, we are declaring ResultScanner instance “*result*” which returns a scanner on the current table “**TRANSACTIONS**” as specified by the Scan object.
- In line 9, a foreach loop is taken, which will run each time for the rows inside the “**TRANSACTIONS**” table until the result scanner value is found.
- In line 10, we are storing entire rows, if the column family name is “**customertransdetails**” and column qualifier name is “count” found in the table “**TRANSACTIONS**” in the variable val.
- In line 11, we are storing entire rows, if the column family name is “**customertransdetails**” and column qualifier name is “fname” found in the table “**TRANSACTIONS**” in the variable val.
- In line 12, we are printing the entire variable val values with its associated column qualifier value.
- In line 13, we are closing the table operation.

Output :

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in

[jar:file:/home/acadgild/install/spark/spark-2.1.0-bin-hadoop2.6/jars/slf4j-log4j12-1.7.16.jar!/org/slf4

j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in

[jar:file:/home/acadgild/install/hbase/hbase-1.4.4/lib/slf4j-log4j12-1.7.10.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.slf4j.impl.Log4jLoggerFactory]

log4j:WARN No appenders could be found for logger (org.apache.hadoop.security.Groups).

log4j:WARN Please initialize the log4j system properly.

log4j:WARN See <http://logging.apache.org/log4j/1.2/faq.html#noconfig> for more info.

Row-value : null

Row-value : Amitabh

keyvalues={101/customertransdetails:fname/1536521068035/Put/vlen=7/seqid=0}

Row-value : null

Row-value : Sharukh

keyvalues={102/customertransdetails:fname/1536521068035/Put/vlen=7/seqid=0}

Row-value : null

Row-value : Anubahv

keyvalues={104/customertransdetails:fname/1536521068035/Put/vlen=7/seqid=0}

Row-value : null

Row-value : Pawan

keyvalues={105/customertransdetails:fname/1536521068035/Put/vlen=5/seqid=0}

Row-value : null

Row-value : Aamir

keyvalues={106/customertransdetails:fname/1536521068035/Put/vlen=5/seqid=0}

Row-value : null

Row-value : Salman

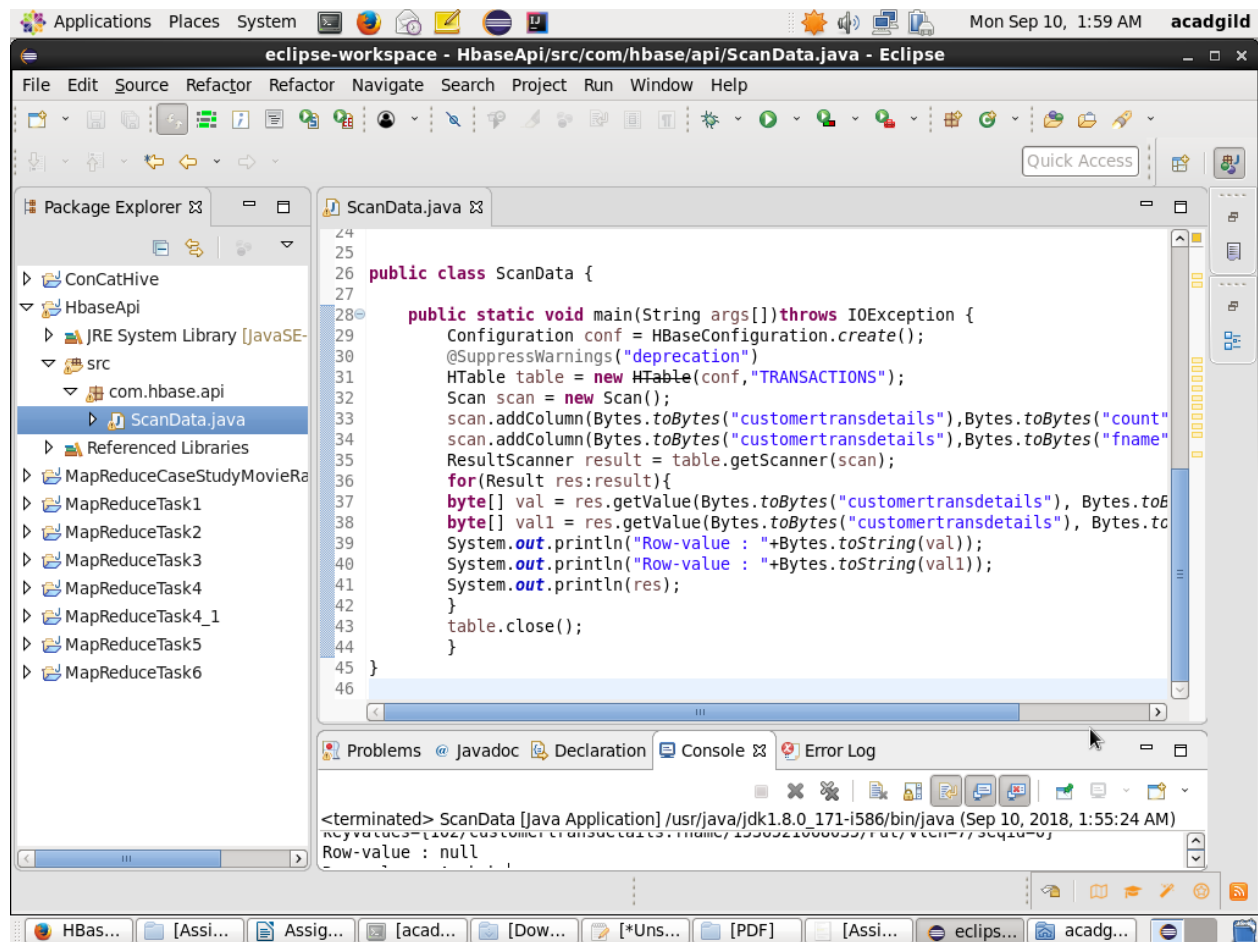
keyvalues={107/customertransdetails:fname/1536521068035/Put/vlen=6/seqid=0}

Row-value : null

Row-value : Ranbir

keyvalues={108/customertransdetails:fname/1536521068035/Put/vlen=6/seqid=0}

OUTPUT:

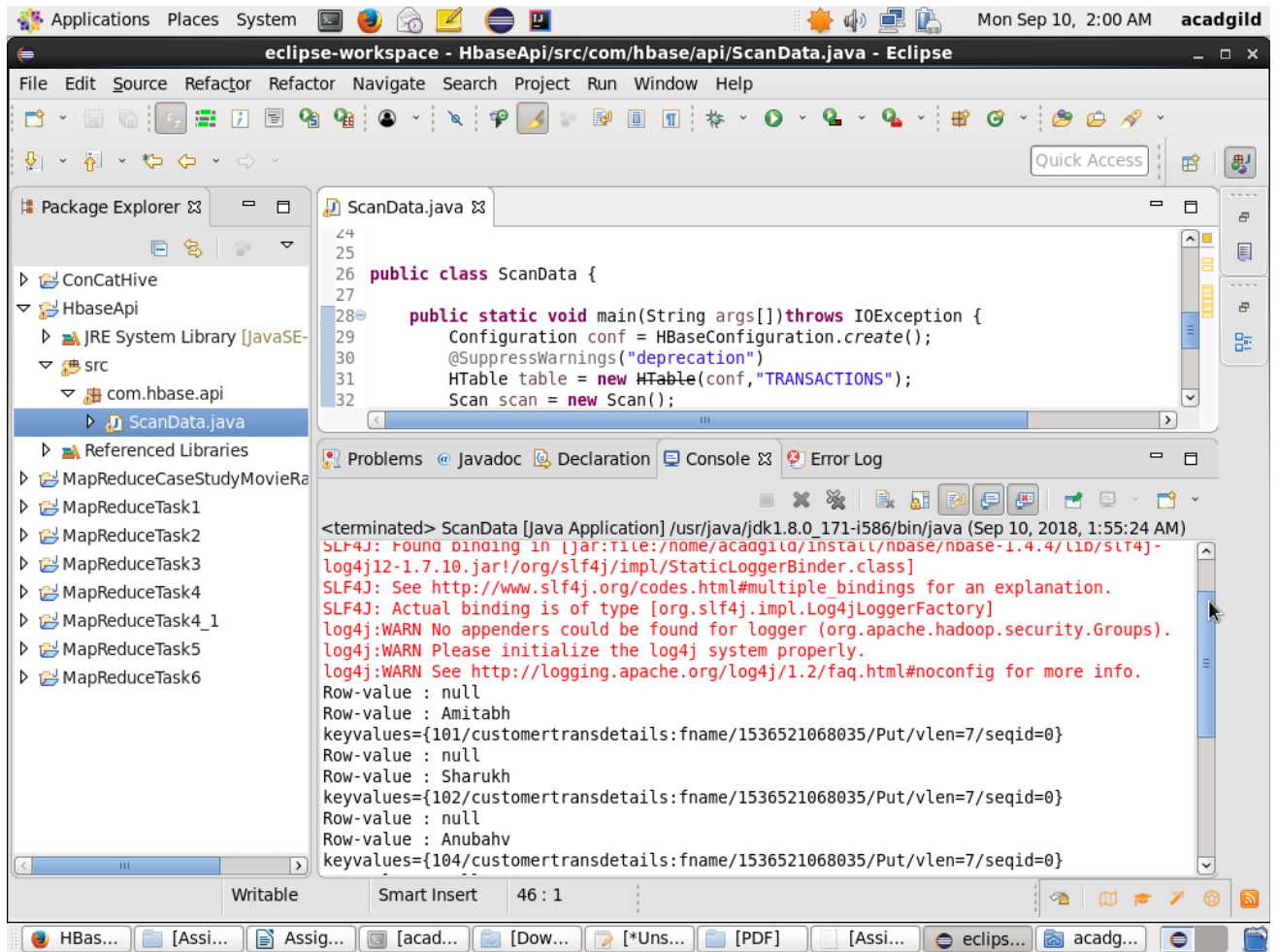


The screenshot displays the Eclipse IDE interface. The Package Explorer on the left shows the project structure, with the `ScanData.java` file selected under the `com.hbase.api` package. The main editor window shows the source code of `ScanData.java`, which is a public class with a `main` method. The code uses `HBaseConfiguration`, `HTable`, and `ResultScanner` to interact with a HBase database. The console at the bottom shows the output of the program, indicating that the row value is null.

```
public class ScanData {  
    public static void main(String args[]) throws IOException {  
        Configuration conf = HBaseConfiguration.create();  
        @SuppressWarnings("deprecation")  
        HTable table = new HTable(conf, "TRANSACTIONS");  
        Scan scan = new Scan();  
        scan.addColumn(Bytes.toBytes("customertransdetails"), Bytes.toBytes("count"));  
        scan.addColumn(Bytes.toBytes("customertransdetails"), Bytes.toBytes("fname"));  
        ResultScanner result = table.getScanner(scan);  
        for(Result res: result){  
            byte[] val = res.getValue(Bytes.toBytes("customertransdetails"), Bytes.toBytes("count"));  
            byte[] val1 = res.getValue(Bytes.toBytes("customertransdetails"), Bytes.toBytes("fname"));  
            System.out.println("Row-value : "+Bytes.toString(val));  
            System.out.println("Row-value : "+Bytes.toString(val1));  
        }  
        table.close();  
    }  
}
```

Console Output:

```
<terminated> ScanData [Java Application] /usr/java/jdk1.8.0_171-i586/bin/java (Sep 10, 2018, 1:55:24 AM)  
Row-value : null
```




```
Applications Places System Mon Sep 10, 2:02 AM acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
104 column=customertransdetails:fname, timestamp=1536521068035, value=Anubahv
105 column=customertransdetails: count, timestamp=1536521068035, value=1
105 column=customertransdetails:fname, timestamp=1536521068035, value=Pawan
106 column=customertransdetails: count, timestamp=1536521068035, value=1
106 column=customertransdetails:fname, timestamp=1536521068035, value=Aamir
107 column=customertransdetails: count, timestamp=1536521068035, value=1
107 column=customertransdetails:fname, timestamp=1536521068035, value=Salman
108 column=customertransdetails: count, timestamp=1536521068035, value=1
108 column=customertransdetails:fname, timestamp=1536521068035, value=Randhir
7 row(s) in 0.7570 seconds

hbase(main):002:0> scan 'TRANSACTIONS'
COLUMN+CELL
101 column=customertransdetails: count, timestamp=1536521068035, value=2
101 column=customertransdetails:fname, timestamp=1536521068035, value=Amitabh
102 column=customertransdetails: count, timestamp=1536521068035, value=1
102 column=customertransdetails:fname, timestamp=1536521068035, value=Sharukh
104 column=customertransdetails: count, timestamp=1536521068035, value=1
104 column=customertransdetails:fname, timestamp=1536521068035, value=Anubahv
105 column=customertransdetails: count, timestamp=1536521068035, value=1
105 column=customertransdetails:fname, timestamp=1536521068035, value=Pawan
106 column=customertransdetails: count, timestamp=1536521068035, value=1
106 column=customertransdetails:fname, timestamp=1536521068035, value=Aamir
107 column=customertransdetails: count, timestamp=1536521068035, value=1
107 column=customertransdetails:fname, timestamp=1536521068035, value=Salman
108 column=customertransdetails: count, timestamp=1536521068035, value=1
108 column=customertransdetails:fname, timestamp=1536521068035, value=Randhir
7 row(s) in 0.2200 seconds

hbase(main):003:0> describe 'TRANSACTIONS'
Table TRANSACTIONS is ENABLED
TRANSACTIONS
COLUMN FAMILIES DESCRIPTION
{NAME => 'customertransdetails', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE',
DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE
=> '65536', REPLICATION_SCOPE => '0'}
1 row(s) in 0.4930 seconds

hbase(main):004:0>
```