

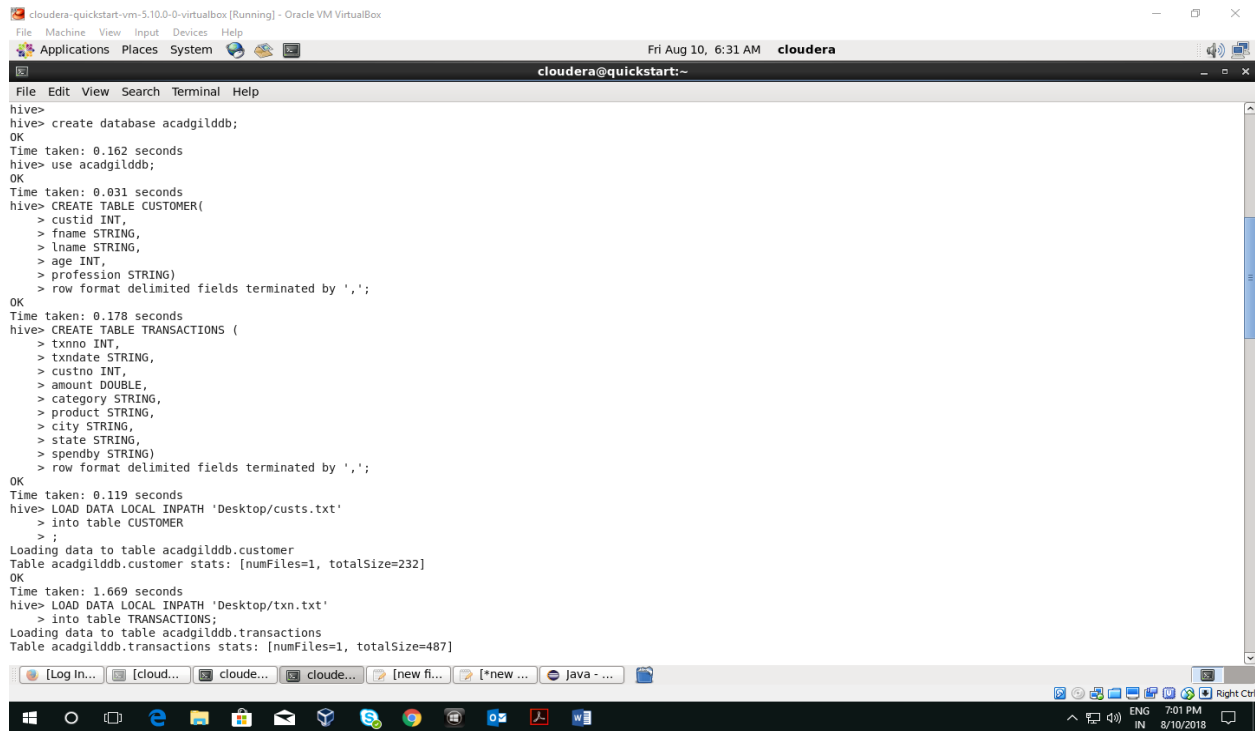
## CASE STUDY – II

### WORKING CUSTOMER & TRANSACTION DATA( HIVE , HBASE )

#### Case Study Description

Let us take up the CUSTOMER and TRANSACTIONS table we have created in the Let's Do Together section.

- Created Acadgild DB.
- Using acadgild DB as default data base
- Created CUSTOMER,TRANSACTIONS tables as below
- Loaded custs.txt,txn.txt data into tables as shown in below screenshots



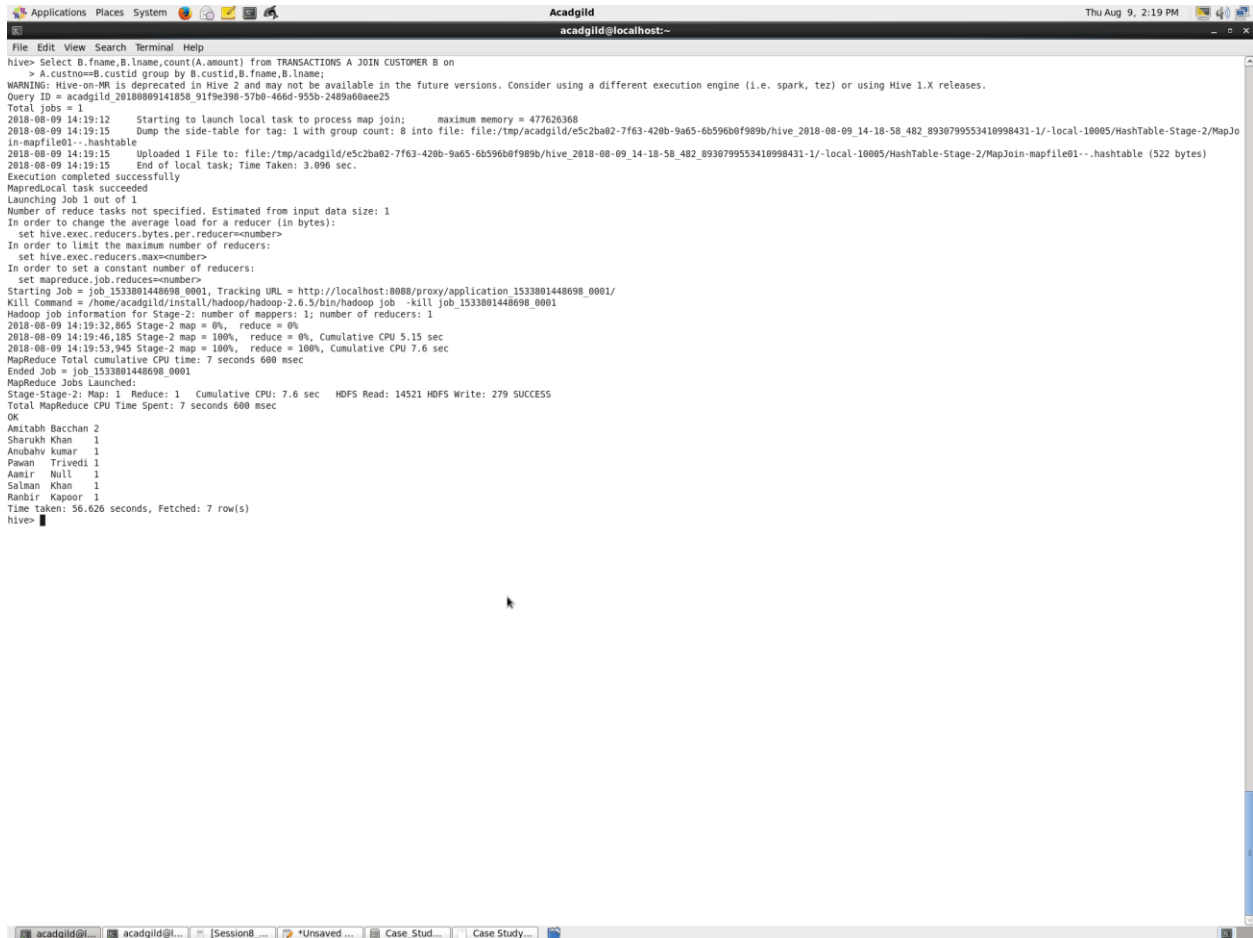
```
cloudera-quickstart-vm-5.10.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
hive>
hive> create database acadgilddb;
OK
Time taken: 0.162 seconds
hive> use acadgilddb;
OK
Time taken: 0.031 seconds
hive> CREATE TABLE CUSTOMER(
  > custid INT,
  > fname STRING,
  > lname STRING,
  > age INT,
  > profession STRING)
  > row format delimited fields terminated by ',';
OK
Time taken: 0.178 seconds
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: 0.119 seconds
hive> LOAD DATA LOCAL INPATH 'Desktop/custs.txt'
  > into table CUSTOMER
  > ;
Loading data to table acadgilddb.customer
Table acadgilddb.customer stats: [numFiles=1, totalSize=232]
OK
Time taken: 1.669 seconds
hive> LOAD DATA LOCAL INPATH 'Desktop/txn.txt'
  > into table TRANSACTIONS;
Loading data to table acadgilddb.transactions
Table acadgilddb.transactions stats: [numFiles=1, totalSize=487]
```

## CASE STUDY – II

### WORKING CUSTOMER & TRANSACTION DATA( HIVE , HBASE )

1. Find out the number of transaction done by each customer as shown below screenshot using below query.

Select C.fname,C.lname,count(T.amount) from Transactions T JOIN customer C on T.custno==C.custid group by C.custid,C.fname,C.lname;

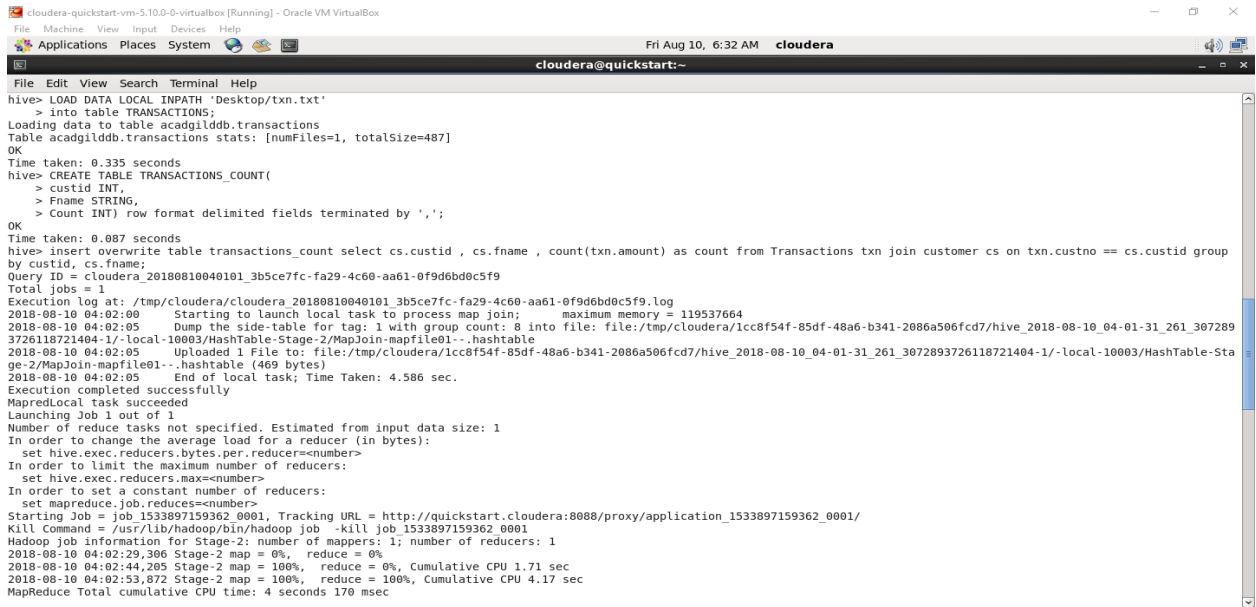


```
Acadgild
acadgild@localhost:~$
File Edit View Search Terminal Help
hive> Select C.fname,C.lname,count(T.amount) from Transactions T JOIN customer C on
> A.custno=B.custid group by B.custid,B.fname,B.lname;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180809141858_91f9e398-57b0-466d-955b-2489a60aee25
Total jobs = 1
2018-08-09 14:19:12 Starting to launch local task to process map join: maximum memory = 477626368
2018-08-09 14:19:15 Dump the side-table for tag: 1 with group count: 8 into file: file:/tmp/acadgild/esc2ba02-7f63-420b-9a65-6b596b0f989b/hive_2018-08-09_14-18-58_482_8930799553410998431-1/-local-10005/HashTable-Stage-2/MapJo
in-mapfile01...hashtable
2018-08-09 14:19:15 Uploaded 1 file to: file:/tmp/acadgild/esc2ba02-7f63-420b-9a65-6b596b0f989b/hive_2018-08-09_14-18-58_482_8930799553410998431-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile01...hashtable (522 bytes)
2018-08-09 14:19:15 End of local task; Time Taken: 3.096 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.reducers=<number>
Starting Job = job_1533801448698_0001, Tracking URL = http://localhost:8088/proxy/application_1533801448698_0001/
Kill command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1533801448698_0001
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-08-09 14:19:32,865 Stage-2 map = 0%, reduce = 0%
2018-08-09 14:19:46,185 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 5.15 sec
2018-08-09 14:19:53,945 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 7.6 sec
MapReduce Total cumulative CPU time: 7 seconds 600 msec
Ended Job = job_1533801448698_0001
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 7.6 sec HDFS Read: 14521 HDFS Write: 279 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 600 msec
OK
Amritabh Bacchan 2
Sharukh Khan 1
Anubhav kumar 1
Pawan Trivedi 1
Aamir Null 1
Salman Khan 1
Ranbir Kapoor 1
Time taken: 56.626 seconds, Fetched: 7 row(s)
hive>
```

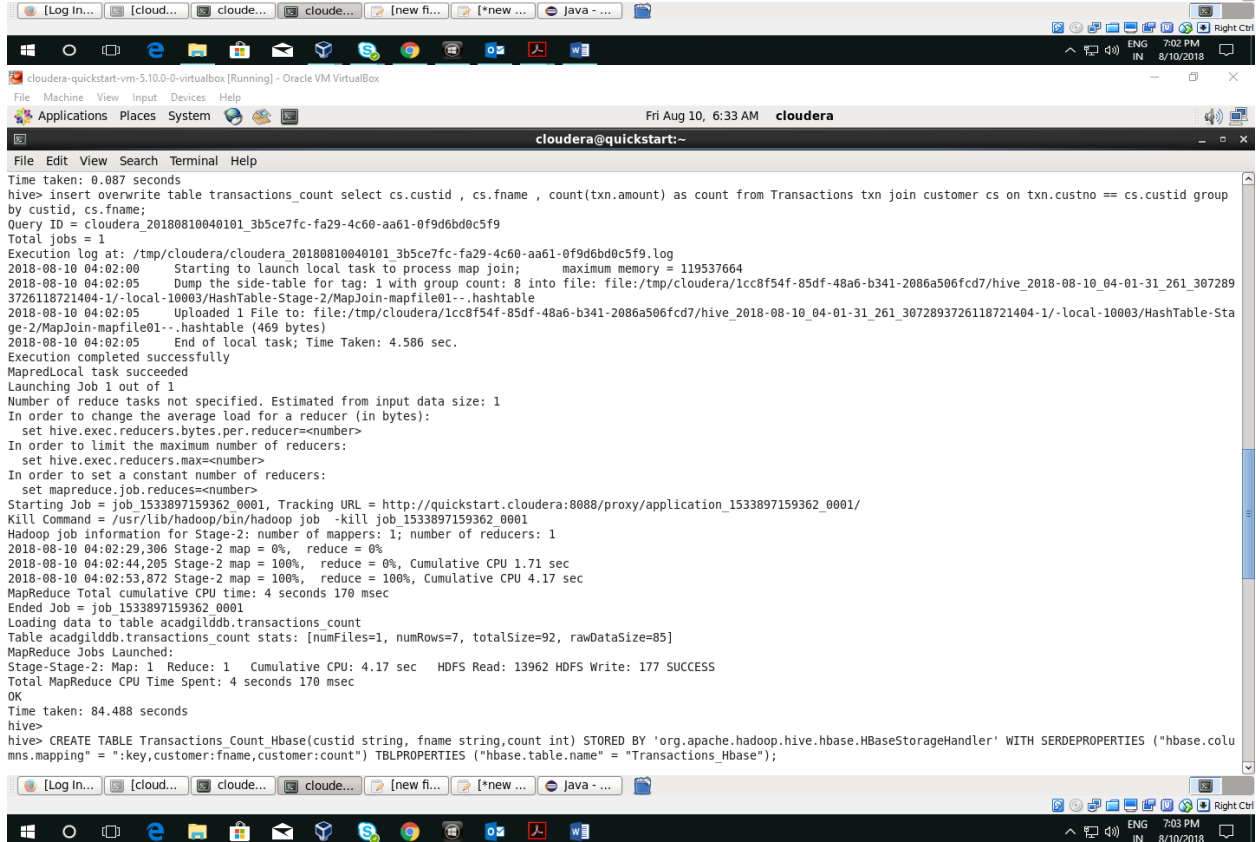
# CASE STUDY – II

## WORKING CUSTOMER & TRANSACTION DATA( HIVE , HBASE )

2. Create a new table called TRANSACTIONS\_COUNT. This table should have 3 fields - custid, fname and count.
  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.
- Please find the screenshot below for answers & I have ran query in hive



```
cloudera-quickstart-vm-5.10.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Fri Aug 10, 6:32 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
hive> LOAD DATA LOCAL INPATH 'Desktop/txn.txt'
> into table TRANSACTIONS;
Loading data to table acadgildb.transactions
Table acadgildb.transactions stats: [numFiles=1, totalSize=487]
OK
Time taken: 0.335 seconds
hive> CREATE TABLE TRANSACTIONS_COUNT(
> custid INT,
> Fname STRING,
> Count INT) row format delimited fields terminated by ',';
OK
Time taken: 0.087 seconds
hive> insert overwrite table transactions_count select cs.custid , cs.fname , count(txn.amount) as count from Transactions txn join customer cs on txn.custno == cs.custid group
by custid, cs.fname;
Query ID = cloudera_20180810040101_3b5ce7fc-fa29-4c60-aa61-0f9d6bd0c5f9
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20180810040101_3b5ce7fc-fa29-4c60-aa61-0f9d6bd0c5f9.log
2018-08-10 04:02:00 Starting to launch local task to process map join; maximum memory = 119537664
2018-08-10 04:02:05 Dump the side-table for tag: 1 with group count: 8 into file: file:/tmp/cloudera/lcc8f54f-85df-48a6-b341-2086a506fcd7/hive_2018-08-10_04-01-31_261_307289
3726118721404-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile01-..hashtable
2018-08-10 04:02:05 Uploaded 1 file to: file:/tmp/cloudera/lcc8f54f-85df-48a6-b341-2086a506fcd7/hive_2018-08-10_04-01-31_261_3072893726118721404-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile01-..hashtable (469 bytes)
2018-08-10 04:02:05 End of local task; Time Taken: 4.586 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_1533897159362_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1533897159362_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1533897159362_0001
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-08-10 04:02:29,306 Stage-2 map = 0%, reduce = 0%
2018-08-10 04:02:44,205 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.71 sec
2018-08-10 04:02:53,872 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.17 sec
MapReduce Total cumulative CPU time: 4 seconds 170 msec
MapReduce Total cumulative CPU time: 4 seconds 170 msec
```



```
cloudera-quickstart-vm-5.10.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Fri Aug 10, 6:33 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
Time taken: 0.087 seconds
hive> insert overwrite table transactions_count select cs.custid , cs.fname , count(txn.amount) as count from Transactions txn join customer cs on txn.custno == cs.custid group
by custid, cs.fname;
Query ID = cloudera_20180810040101_3b5ce7fc-fa29-4c60-aa61-0f9d6bd0c5f9
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20180810040101_3b5ce7fc-fa29-4c60-aa61-0f9d6bd0c5f9.log
2018-08-10 04:02:00 Starting to launch local task to process map join; maximum memory = 119537664
2018-08-10 04:02:05 Dump the side-table for tag: 1 with group count: 8 into file: file:/tmp/cloudera/lcc8f54f-85df-48a6-b341-2086a506fcd7/hive_2018-08-10_04-01-31_261_307289
3726118721404-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile01-..hashtable
2018-08-10 04:02:05 Uploaded 1 file to: file:/tmp/cloudera/lcc8f54f-85df-48a6-b341-2086a506fcd7/hive_2018-08-10_04-01-31_261_3072893726118721404-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile01-..hashtable (469 bytes)
2018-08-10 04:02:05 End of local task; Time Taken: 4.586 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_1533897159362_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1533897159362_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1533897159362_0001
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-08-10 04:02:29,306 Stage-2 map = 0%, reduce = 0%
2018-08-10 04:02:44,205 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.71 sec
2018-08-10 04:02:53,872 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.17 sec
MapReduce Total cumulative CPU time: 4 seconds 170 msec
Ended Job = job_1533897159362_0001
Loading data to table acadgildb.transactions_count
Table acadgildb.transactions_count stats: [numFiles=1, numRows=7, totalSize=92, rawDataSize=85]
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.17 sec HDFS Read: 13962 HDFS Write: 177 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 170 msec
OK
Time taken: 84.488 seconds
hive>
hive> CREATE TABLE Transactions_Count Hbase(custid string, fname string,count int) STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH SERDEPROPERTIES ("hbase.colu
mns.mapping" = "key,customer:fname,customer:count") TBLPROPERTIES ("hbase.table.name" = "Transactions_Hbase");
```

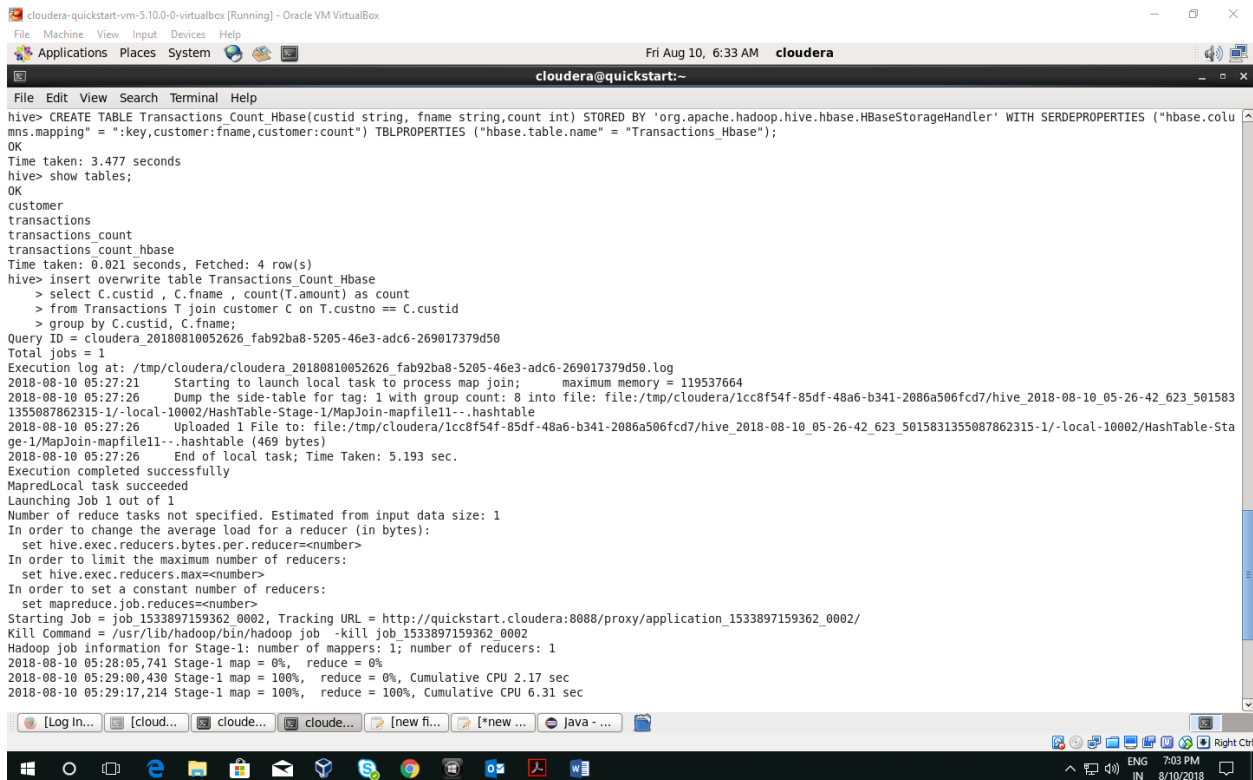
## CASE STUDY – II

### WORKING CUSTOMER & TRANSACTION DATA( HIVE , HBASE )

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.

Insert the data in TRANSACTIONS\_COUNT table using the query in step 3 again, this should populate the Hbase TRANSACTIONS table automatically

- As shown in below picture TRASACTION\_COUNT\_HBASE & TRASACTION\_\_HBASE tables are created.
- Data is inserted using below query in picture

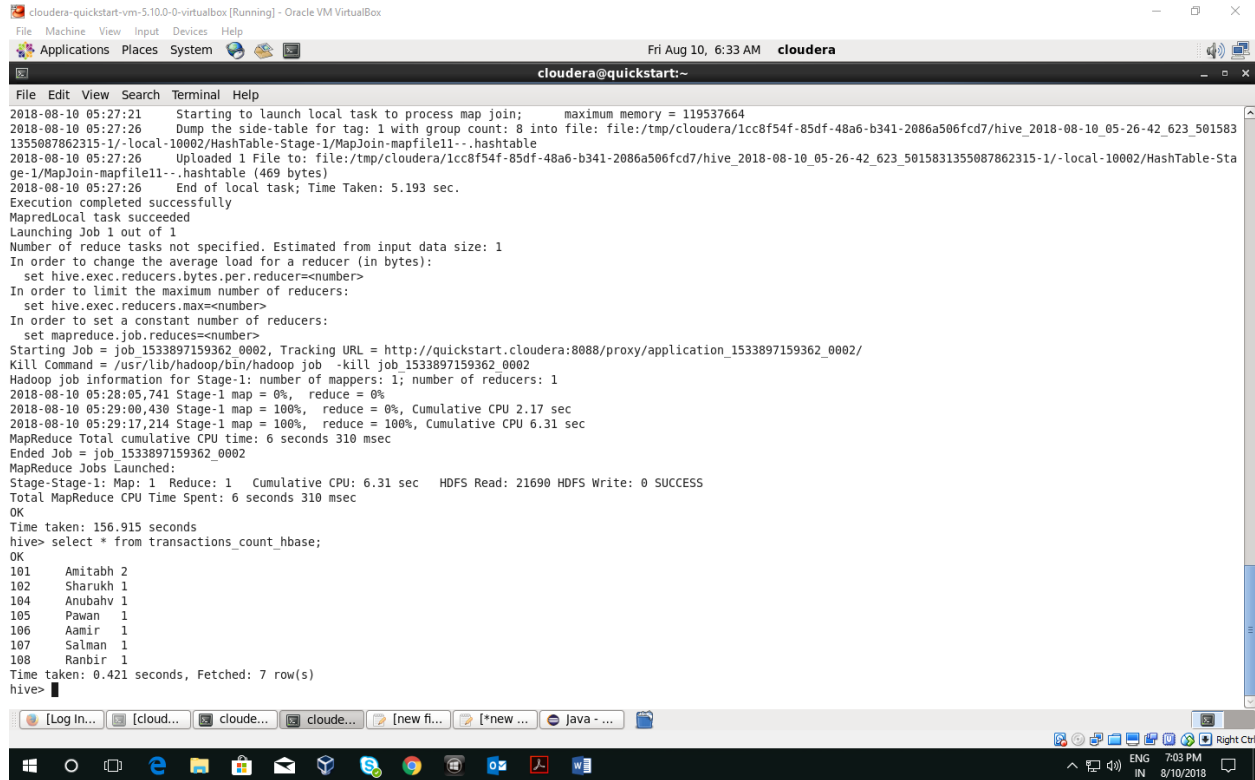


```
cloudera-quickstart-vm-5.10.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~
File Edit View Search Terminal Help
hive> CREATE TABLE Transactions_Count_Hbase(custid string, fname string, count int) STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH SERDEPROPERTIES ("hbase.columns.mapping" = ":key,customer:fname,customer:count") TBLPROPERTIES ("hbase.table.name" = "Transactions_Hbase");
OK
Time taken: 3.477 seconds
hive> show tables;
OK
customer
transactions
transactions_count
transactions_count_hbase
Time taken: 0.021 seconds, Fetched: 4 row(s)
hive> insert overwrite table Transactions_Count_Hbase
> select C.custid , C.fname , count(T.amount) as count
> from Transactions T join customer C on T.custno == C.custid
> group by C.custid, C.fname;
Query ID = cloudera_20180810052626_fab92ba8-5205-46e3-adc6-269017379d50
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20180810052626_fab92ba8-5205-46e3-adc6-269017379d50.log
2018-08-10 05:27:21 Starting to launch local task to process map join; maximum memory = 119537664
2018-08-10 05:27:26 Dump the side-table for tag: 1 with group count: 8 into file: file:/tmp/cloudera/1cc8f54f-85df-48a6-b341-2086a506fcd7/hive_2018-08-10_05-26-42_623_5015831355087862315-1/-local-10002/HashTable-Stage-1/MapJoin-mapfile11--.hashtable
2018-08-10 05:27:26 Uploaded 1 File to: file:/tmp/cloudera/1cc8f54f-85df-48a6-b341-2086a506fcd7/hive_2018-08-10_05-26-42_623_5015831355087862315-1/-local-10002/HashTable-Stage-1/MapJoin-mapfile11--.hashtable (469 bytes)
2018-08-10 05:27:26 End of local task; Time Taken: 5.193 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_1533897159362_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1533897159362_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1533897159362_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-08-10 05:28:05,741 Stage-1 map = 0%, reduce = 0%
2018-08-10 05:29:00,430 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.17 sec
2018-08-10 05:29:17,214 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.31 sec
```

# CASE STUDY – II

## WORKING CUSTOMER & TRANSACTION DATA( HIVE , HBASE )

- Please find the data inserted into transactions\_count\_hbase hive table

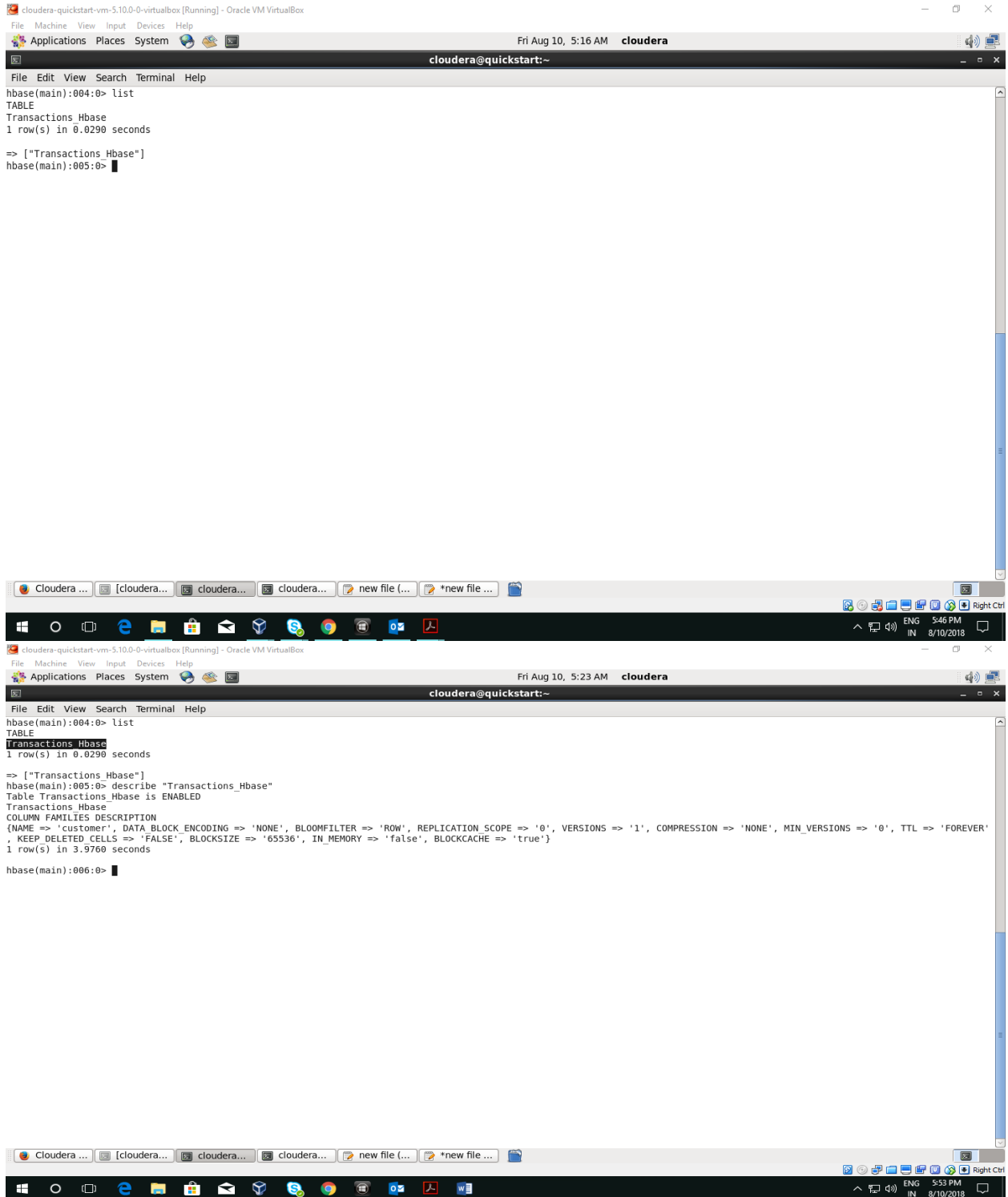


```
cloudera-quickstart-vm-5.10.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~
File Edit View Search Terminal Help
2018-08-10 05:27:21 Starting to launch local task to process map join; maximum memory = 119537664
2018-08-10 05:27:26 Dump the side-table for tag: 1 with group count: 8 into file: file:/tmp/cloudera/1cc8f54f-85df-48a6-b341-2086a506fcd7/hive_2018-08-10_05-26-42_623_501583
1355087862315-1/-local-10002/HashTable-Stage-1/MapJoin-mapfile11--.hashtable
2018-08-10 05:27:26 Uploaded 1 File to: file:/tmp/cloudera/1cc8f54f-85df-48a6-b341-2086a506fcd7/hive_2018-08-10_05-26-42_623_5015831355087862315-1/-local-10002/HashTable-Stage-1/MapJoin-mapfile11--.hashtable (469 bytes)
2018-08-10 05:27:26 End of local task; Time Taken: 5.193 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_1533897159362_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1533897159362_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1533897159362_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-08-10 05:28:05,741 Stage-1 map = 0%, reduce = 0%
2018-08-10 05:29:00,430 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.17 sec
2018-08-10 05:29:17,214 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.31 sec
MapReduce Total cumulative CPU time: 6 seconds 310 msec
Ended Job = job_1533897159362_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.31 sec HDFS Read: 21690 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 310 msec
OK
Time taken: 156.915 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: 0.421 seconds, Fetched: 7 row(s)
hive>
```

## CASE STUDY – II

### WORKING CUSTOMER & TRANSACTION DATA( HIVE , HBASE )

- Please find the hbase table created out of the query
- Please find the description of table



The screenshot shows a Cloudera Quickstart VM terminal window. The terminal displays the following commands and output:

```
hbase(main):004:0> list
TABLE
Transactions Hbase
1 row(s) in 0.0290 seconds

=> ["Transactions Hbase"]
hbase(main):005:0>
```

The second screenshot shows the same terminal window with the following commands and output:

```
hbase(main):004:0> list
TABLE
Transactions Hbase
1 row(s) in 0.0290 seconds

=> ["Transactions Hbase"]
hbase(main):005:0> describe "Transactions_Hbase"
Table Transactions_Hbase is ENABLED
Transactions_Hbase
COLUMN FAMILIES DESCRIPTION
(NAME => 'customer', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => 'FOREVER',
KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true')
1 row(s) in 3.9760 seconds

hbase(main):006:0>
```

# CASE STUDY – II

## WORKING CUSTOMER & TRANSACTION DATA( HIVE , HBASE )

- Scanned table to make sure the data is loaded properly.

```
cloudera-quickstart-vm-5.10.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Fri Aug 10, 5:32 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
hbase(main):004:0> list
TABLE
Transactions Hbase
1 row(s) in 0.0290 seconds

=> ["Transactions Hbase"]
hbase(main):005:0> describe "Transactions_Hbase"
Table Transactions_Hbase is ENABLED
Transactions_Hbase
COLUMN FAMILIES DESCRIPTION
(NAME => 'customer', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => 'FOREVER',
KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true')
1 row(s) in 3.9760 seconds

hbase(main):006:0> scan "Transactions_Hbase"
ROW COLUMN+CELL
101 column=customer:count, timestamp=1533904155835, value=2
101 column=customer:fname, timestamp=1533904155835, value=Amitabh
102 column=customer:count, timestamp=1533904155835, value=1
102 column=customer:fname, timestamp=1533904155835, value=Sharukh
104 column=customer:count, timestamp=1533904155835, value=1
104 column=customer:fname, timestamp=1533904155835, value=Anubahv
105 column=customer:count, timestamp=1533904155835, value=1
105 column=customer:fname, timestamp=1533904155835, value=Pawan
106 column=customer:count, timestamp=1533904155835, value=1
106 column=customer:fname, timestamp=1533904155835, value=Aamir
107 column=customer:count, timestamp=1533904155835, value=1
107 column=customer:fname, timestamp=1533904155835, value=Salman
108 column=customer:count, timestamp=1533904155835, value=1
108 column=customer:fname, timestamp=1533904155835, value=Randhir
7 row(s) in 2.4150 seconds

hbase(main):007:0>

Cloudera ... [cloudera... cloudera... cloudera... new file (...) *new file ...
Windows Taskbar: ENG 6:02 PM 8/10/2018
cloudera-quickstart-vm-5.10.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Fri Aug 10, 5:32 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
2018-08-10 05:27:21 Starting to launch local task to process map join; maximum memory = 119537664
2018-08-10 05:27:26 Dump the side-table for tag: 1 with group count: 8 into file: file:/tmp/cloudera/1cc8f54f-85df-48a6-b341-2086a506fcd7/hive_2018-08-10_05-26-42_623_501583
1355087862315-1/-local-10002/HashTable-Stage-1/MapJoin-mapfile11--.hashtable
2018-08-10 05:27:26 Uploaded 1 File to: file:/tmp/cloudera/1cc8f54f-85df-48a6-b341-2086a506fcd7/hive_2018-08-10_05-26-42_623_5015831355087862315-1/-local-10002/HashTable-Stage-1/MapJoin-mapfile11--.hashtable (469 bytes)
2018-08-10 05:27:26 End of local task; Time Taken: 5.193 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_1533897159362_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1533897159362_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1533897159362_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-08-10 05:28:05,741 Stage-1 map = 0%, reduce = 0%
2018-08-10 05:29:00,430 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.17 sec
2018-08-10 05:29:17,214 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.31 sec
MapReduce Total cumulative CPU time: 6 seconds 310 msec
Ended Job = job_1533897159362_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.31 sec HDFS Read: 21690 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 310 msec
OK
Time taken: 156.915 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 Randhir 1
Time taken: 0.421 seconds, Fetched: 7 row(s)
hive>
```

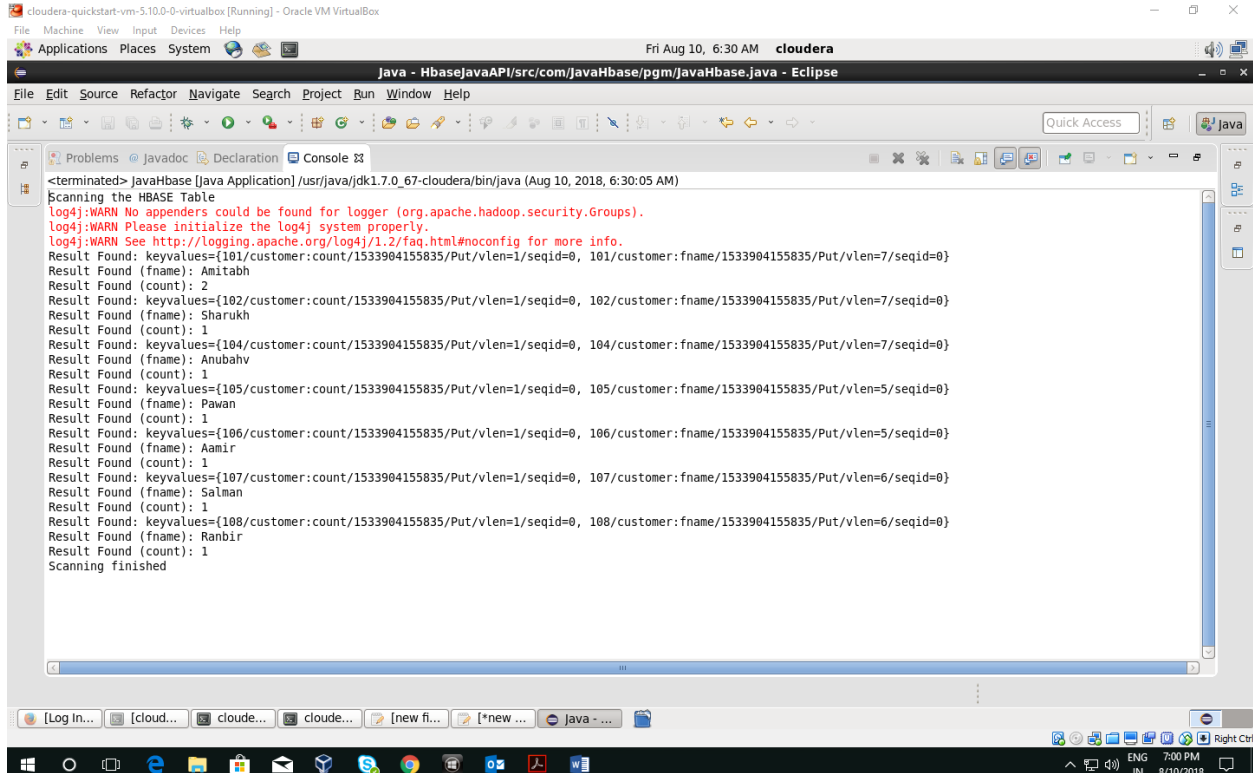
## CASE STUDY – II

### WORKING CUSTOMER & TRANSACTION DATA( HIVE , HBASE )

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

- Executed below code in java eclipse in same acadgild vm, connected to Hbase table scanned table from java code displayed code in console.

```
1 import java.io.IOException;
2 import org.apache.hadoop.conf.Configuration;
3 import org.apache.hadoop.hbase.HBaseConfiguration;
4 import org.apache.hadoop.hbase.client.HTable;
5 import org.apache.hadoop.hbase.client.Result;
6 import org.apache.hadoop.hbase.client.ResultScanner;
7 import org.apache.hadoop.hbase.client.Scan;
8 import org.apache.hadoop.hbase.util.Bytes;
9
10
11 public class HiveHbase {
12     public static void main(String args[]) throws IOException{
13         System.out.println("Scanning the HBASE Table");
14         Configuration c = HBaseConfiguration.create(); // Instantiate configuration class
15         HTable table = new HTable(c, "Transactions_Hbase"); // Instantiate HTable class
16         Scan scan = new Scan(); // Instantiate the Scan class
17         // Scanning the required columns
18         scan.addColumn(Bytes.toBytes("customer"), Bytes.toBytes("fname")); //First parameter indicates column family and second parameter is column name
19         scan.addColumn(Bytes.toBytes("customer"), Bytes.toBytes("count")); //Add all the required columns
20
21         ResultScanner scanner = table.getScanner(scan); // Get scan result
22         // Reading values from scan result
23         for (Result result = scanner.next(); result != null; result = scanner.next())
24         {
25             System.out.println("Result Found: " + result);
26             //Convert Bytes to String
27
28             System.out.println("Result Found (fname): " + Bytes.toString(result.getValue(Bytes.toBytes("customer"), Bytes.toBytes("fname"))));
29             System.out.println("Result Found (count): " + Bytes.toString(result.getValue(Bytes.toBytes("customer"), Bytes.toBytes("count"))));
30         }
31         scanner.close(); //close the scanner
32         System.out.println("Scanning finished");
33     }
34 }
```



The screenshot shows the Eclipse IDE interface. The top part displays the Java code for scanning an HBase table. The bottom part shows the console output, which includes warnings about loggers and the scanned data for the 'TRANSACTIONS' table. The console output is as follows:

```
<terminated> javaHbase [Java Application] /usr/java/jdk1.7.0_67-cloudera/bin/java (Aug 10, 2018, 6:30:05 AM)
Scanning the HBASE Table
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.
Result Found: keyValues={101/customer:count/1533904155835/Put/vlen=1/seqid=0, 101/customer:fname/1533904155835/Put/vlen=7/seqid=0}
Result Found (fname): Amitabh
Result Found (count): 2
Result Found: keyValues={102/customer:count/1533904155835/Put/vlen=1/seqid=0, 102/customer:fname/1533904155835/Put/vlen=7/seqid=0}
Result Found (fname): Sharukh
Result Found (count): 1
Result Found: keyValues={104/customer:count/1533904155835/Put/vlen=1/seqid=0, 104/customer:fname/1533904155835/Put/vlen=7/seqid=0}
Result Found (fname): Anubhav
Result Found (count): 1
Result Found: keyValues={105/customer:count/1533904155835/Put/vlen=1/seqid=0, 105/customer:fname/1533904155835/Put/vlen=5/seqid=0}
Result Found (fname): Pawan
Result Found (count): 1
Result Found: keyValues={106/customer:count/1533904155835/Put/vlen=1/seqid=0, 106/customer:fname/1533904155835/Put/vlen=5/seqid=0}
Result Found (fname): Aamir
Result Found (count): 1
Result Found: keyValues={107/customer:count/1533904155835/Put/vlen=1/seqid=0, 107/customer:fname/1533904155835/Put/vlen=6/seqid=0}
Result Found (fname): Salman
Result Found (count): 1
Result Found: keyValues={108/customer:count/1533904155835/Put/vlen=1/seqid=0, 108/customer:fname/1533904155835/Put/vlen=6/seqid=0}
Result Found (fname): Ranbir
Result Found (count): 1
Scanning finished
```