# **Case Study – II. Customer Transaction.**

1. Start Hadoop componetns in VM using the following command and check whether all the daemons are running correctly:

\$start-all.hs

\$jps

```
[acadgild@localhost ~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/10/02 13:09:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-namenode-localhost.l
ocaldomain.out
localhost: starting datanode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-datanode-localhost.l
ocaldomain.out
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.localdomain.out
18/10/02 13:09:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable starting yarn daemons starting resourcemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-resourcemanager-localhost.
localdomain.out
localhost: starting nodemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-nodemanager-localhost.localdomain.out
[acadgild@localhost
[acadgild@localhost ~1$ ips
3649 DataNode
4101 NodeManager
3846 SecondaryNameNode
3544 NameNode
     Jps
3996 ResourceManager
[acadgild@localhost ~]$
```

2. Now start the Hive shell as follows:

```
[acadgild@localhost ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4
/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-2.3.2-bin/lib/hive-common-2.3.2.
ar!/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 eng
ne (i.e. spark, tez) or using Hive 1.X releases.
hive>
```

3. Now check what are the databases available in the hive as follows: *hive>show databases*;

```
hive> show databases;
OK
default
sports
Time taken: 0.064 seconds, Fetched: 2 row(s)
```

4. Now we have 2 databases, called *default* and *sports*. We will use *default* database for creation of tables: *hive> default* 

```
File Edit View Search Terminal Help
hive> use default;
OK
Time taken: 0.039 seconds
```

5. Now create a table called *customer* as follows:

hive > CREATE TABLE CUSTOMER (

- > Custid INT,
- > Fname string,
- > Lname string,
- > Age int,

```
> Profession string
       > Row format delimited fields terminated by '.';
hive> CREATE TABLE CUSTOMER(
    > custid INT,
    > fname STRING,
    > lname STRING,
    > age INT,
    > profession STRING
    > )
    > row format delimited fields terminated by ',';
0K
Time taken: 0.162 seconds
hive> CREATE TABLE CUSTOMER(
    > custid INT,
    > fname STRING,
    > lname STRING,
    > age INT,
    > profession STRING
    > row format delimited fields terminated by ',';
0K
Time taken: 0.162 seconds
```

6. Now load the data into the file explicitly from a file as follows: hive > LOAD DATA LOCAL INPUT '/home/acadgild/Downloads/hive/custs.txt' into table CUSTOMER;

```
File Edit View Search Terminal Help
hive> LOAD DATA LOCAL INPATH '/home/acadgild/Downloads/hive/custs.txt' into table CUSTOMER;
Loading data to table default.customer
OK
Time taken: 0.827 seconds
```

7. Now we will verify data from the table customer as follows: *hive> select \* from customer;* 

terminated by ',';

```
File Edit View Search Terminal
                                 Help
hive> select * from customer;
oĸ
4000001 Kristina
                                 55
                                          Pilot
                         Chung
4000002 Paige
                         74
                                 Teacher
                Chen
                                 Firefighter
4000003 Sherri
                Melton
                         34
4000004 Gretchen
                         Hill
                                          Computer hardware engineer
                                 66
                Puckett 74
4000005 Karen
                                 Lawyer
4000006 Patrick Song
                         42
                                 Veterinarian
4000007 Elsie
                Hamilton
                                 43
                                          Pilot
4000008 Hazel
                Bender
                         63
                                 Carpenter
4000009 Malcolm Wagner
                                 Artist
4000010 Dolores McLaughlin
                                 60
                                          Writer
Time taken: 3.16 seconds, Fetched: 10 row(s)
```

8. Now create a transaction table called TXNRECORDS as follows:
hive> CREATE TABLE TXNRECORDS (txno INT, txndate STRING, custono INT, amount DOUBLE, category STRING, product STRING, city STRING, state STRING, spendby STRING) row format delimited fields

```
File Edit View Search Terminal Help
hive> CREATE TABLE TXNRECORDS (txnno INT, txndate STRING, custno INT, amount DOUBLE, category STRING, product STRING, city S
RING, state STRING, spendby STRING)
> row format delimited fields terminated by ',';
OK
Time taken: 0.179 seconds
```

9. Now load the data into the *txnrecords* as follows:

## hive> LOAD DATA LOCAL INPATH '/home/acadgild/Downloads/hive/txns.txt' INTO TABLE TXNRECORDS;

```
File Edit View Search Terminal Help
hive> LOAD DATA LOCAL INPATH '/home/acadgild/Downloads/hive/txns.txt' into table TXNRECORDS
> ;
Loading data to table default.txnrecords
OK
Time taken: 0.848 seconds
```

10. Now display the data from the table *txnrecords* as follows:

			3 (A)				
File	Edit View Search	Terminal Hel	elp				
hive>	select * from txn	records;				[	
OK						1	
0	06-26-2011	4000001 40.	0.33 Exercise & Fitness	Cardio Machine Accessories	Clarksville	Tennessee	
redit							
1	05-26-2011	4000002 198	08.44 Exercise & Fitness	Weightlifting Gloves Lon-	g Beach Califo	rnia cred	
t							
2	06-01-2011	4000002 5.5	58 Exercise & Fitness	Weightlifting Machine Acces	sories Anaheim	m California	
redit							
3	06-05-2011	4000003 198		stics Rings Milwaukee	Wisconsin	credit	
4	12-17-2011	4000002 98.	3.81 Team Sports Field	Hockey Nashville Ten	nessee credit		
5	02-14-2011	4000004 193	3.63 Outdoor Recreation	Camping & Backpacking & Hik	ing Chicago Illino:	is cred	
3 4 5 t 6							
6	10-28-2011	4000005 27.	7.89 Puzzles Jigsaw Puzzle	es Charleston South Carol	ina credit		
7	07-14-2011	4000006 96.	6.01 Outdoor Play Equipmer	nt Sandboxes Columbus	Ohio credit		
8	01-17-2011	4000006 10.	0.44 Winter Sports Snown	nobiling Des Moines Iow	a credit		
9	05-17-2011	4000006 152	2.46 Jumping Bungee Jumpir	ig St. Petersburg Florida cre	dit		
10	05-29-2011	4000007 180	0.28 Outdoor Recreation	Archery Reno Nevada cre	dit		
11	06-18-2011	4000009 121	1.39 Outdoor Play Equipmer	nt Swing Sets Columbus	Ohio credit		
12	02-08-2011	4000009 41.	.52 Indoor Games Bowli	ng San Francisco California	credit		
13	03-13-2011	4000010 107			aii credit		
14	02-25-2011	4000010 36.	5.81 Gymnastics Vault	ing Horses Los Angeles Cal	ifornia credit		
15	10-20-2011	4000001 137	37.64 Combat Sports Fenci	ng Honolulu Hawaii cre	dit		
16	05-28-2011	4000010 35.	5.56 Exercise & Fitness	Free Weight Bars Col	umbia South (	Carolina cred	
t	t						
Triple.							

Task 1: Find the number of transactions done by each customer.

# 1. Write query as follows:

hive> select customer.custid, customer.fname, COUNT(txnrecords.custno) from customer, txnrecords where customer.custid=txnrecords.custno group by customer.custid, customer.fname;

```
File Edit View Search Terminal Help
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 7.52 sec
                                                                  HDFS Read: 18142 HDFS Write: 381 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 520 msec
4000001 Kristina
4000002 Paige
4000003 Sherri
4000004 Gretchen
                         5
4000005 Karen
4000006 Patrick 5
4000007 Elsie
4000008 Hazel
4000009 Malcolm 6
4000010 Dolores
Time taken: 51.462 seconds, Fetched: 10 row(s)
hive>
```

2. Here we can observe that all the customer names and their customer numbers and the number of transactions done by each customers.

## Task 2:

Create a new table called TRANSACTIONS COUNT. This table should have 3 fields – custid, fname and count.

1. Create the table *transactions\_count* table as follows:

```
hive> CREATE TABLE TRANSACTIONS_COUNT (

> custid INT,

> fname STRING,

> count INT

> );
```

- 2. We created the table called TRANSACTIONS COUNT.
- 3. Now lets see the table created or not inside the database *default* as follows:

hive> show tables;

```
File Edit View Search Terminal Help
hive> show tables;
OK
customer
emp_temp
emp_temp1
transactions_count
txnrecords
Time taken: 0.066 seconds, Fetched: 5 row(s)
hive>
```

## Task 3:

Now write a hive query in such way that the query populates the data obtained in Step 1 above and populates the table in step 2 above.

1. Write a query to populate the data into the table transactions\_count as follows:

hive> insert into transactions\_count select customer.custid, customer.fname, COUNT(txnrecords.custno) from customer, txnrecords where customer.custid=txnrecords.custno group by customer.custid, customer.fname;

```
hive> insert into transactions_count select customer.custid, customer.fname, COUNT(txnrecords.custno) from customer, txnrec
     rds where customer.custid=txnrecords.custno group by customer.custid, customer.fname;

WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different exection engine (i.e. spark, tez) or using Hive 1.X releases.

Query ID = acadgild_20181002163648_45d60e80-2ec5-48a8-ael0-9b7e894aclaa

Total jobs = 1

SLE41. Class path contains multiple SLE41 bindings
     SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4
/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
     Uploaded 1 File to: file:/tmp/acadgild/5298d069-6441-4e0b-8001-05ee7b082554/hive 2018-10-02 16-36-48
      2018-10-02 16:37:01
     941_840049763826257214-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile10--.hashtable (556 bytes)
2018-10-02 16:37:01     End of local task; Time Taken: 2.735 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 timit 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>
set mapreduce.job.reduces=<number>
Starting Job = job_1538465981645_0002, Tracking URL = http://localhost:8088/proxy/application_1538465981645_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1538465981645_0002/
     Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1 2018-10-02 16:37:11,578 Stage-2 map = 0%, reduce = 0% 2018-10-02 16:37:20,435 Stage-2 map = 100%, reduce = 0% Cumulative CPU 4.45 sec 2018-10-02 16:37:29,270 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 9.79 sec
      MapReduce Total cumulative CPU time: 9 seconds 790 msec
 File Edit View Search Terminal Help
2018-10-02 16:37:29,270 Stage-2 map = 100%,
                                                                                    reduce = 100%, Cumulative CPU 9.79 sec
MapReduce Total cumulative CPU time: 9 seconds 790 msec
Ended Job = job 1538465981645 0002
Loading data to table default.transactions count
Stage-Stage-2: Map: 1 Reduce: 1
                                                                 Cumulative CPU: 9.79 sec
                                                                                                                   HDFS Read: 18781 HDFS Write: 257 SUCCESS
```

```
MapReduce Jobs Launched:
Total MapReduce CPU Time Spent: 9 seconds 790 msec
0K
Time taken: 43.165 seconds
hive> select * from transactions_count;
                                                                     I
OK
4000001 Kristina
4000002 Paige
4000003 Sherri
                3
4000004 Gretchen
4000005 Karen
4000006 Patrick 5
4000007
        Elsie
                6
4000008 Hazel
                10
4000009 Malcolm 6
4000010 Dolores 6
Time taken: 0.18 seconds, Fetched: 10 row(s)
```

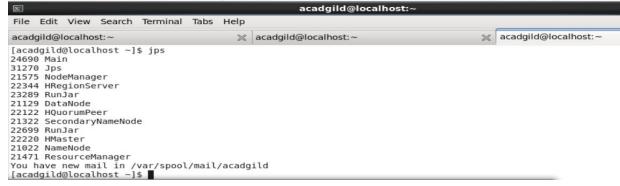
2. We can see the output of the query is populated into the table *transactions\_count* table. And we displayed the table content using the query *hive>select \* from transactions\_count* 

#### Task 4:

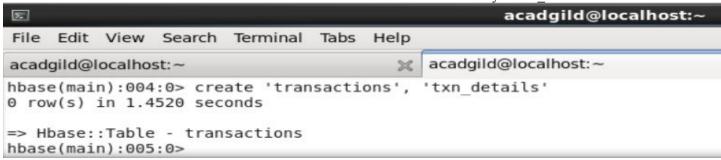
Now lets make the TRANSACTIONS\_COUNT table HBase complaint. In the sence, use Ser Des And Storage handler feature to hive to change the TRANSACTION\_COUNT table to be able to create a TRANSACTIONS table in HBase.

- 1. Now take another terminal and start HBase components in the hadoop as follows:
  - \$start-hbase.sh
- 2. Now verity HMaster is started or not using *jps* command as follows: *\$jps*





3. Create a table in the Hbase with name as *transactions* and with column family as *txn details* as follows:



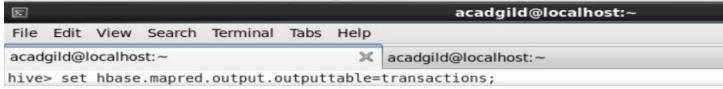
4. Create an External table in hive. We need to create an external table in Hive referring to HBase table. That can be created as shown below:

CREATE EXTERNAL TABLE transactions (



3. **Inserting Values into HBase Table Through Hive** for inserting data into the HBase table through Hive, you need to specify the HBase table name in the hive shell by using the below property before running the insert command

hive> set hbase.mapred.output.outputtable=transactions;

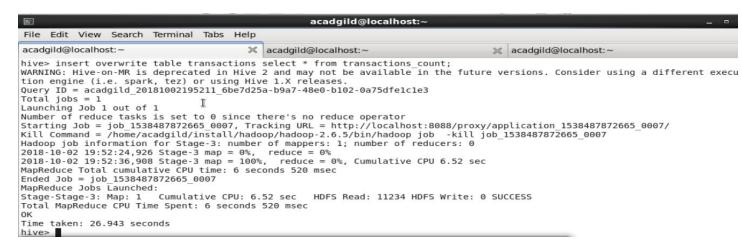


### Task 5:

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).

1. Write a hive query to write the data populated from the table *transactiosn\_count* into *transactions* table which is present in the HBase, as follows:

hive>insert overwrite table transactions select \* from transactions\_count;



2. We can see that the process ran successfully. Now we will check the table we already created in HBase, *transactions*.

hbase (main):014:0> scan 'transactions'

```
acadgild@localhost:-
     Edit View
                                      Help
File
                Search Terminal
                                Tabs
                                    acadgild@localhost:~
                                                                         acadgild@localhost:~
acadgild@localhost:~
hbase(main):014:0> scan 'transactions'
                                  COLUMN+CELL
ROW
 4000001
                                  column=txn_details:count, timestamp=1538490156137, value=8
 4000001
                                                            timestamp=1538490156137, value=Kristina
                                  column=txn_details:fname,
 4000002
                                  column=txn_details:count, timestamp=1538490156137,
                                                                                      value=6
 4000002
                                  column=txn_details:fname, timestamp=1538490156137, value=Paige
 4000003
                                  column=txn_details:count, timestamp=1538490156137,
                                                                                      value=3
 4000003
                                  column=txn details:fname,
                                                            timestamp=1538490156137, value=Sherri
 4000004
                                  column=txn_details:count, timestamp=1538490156137,
                                                                                      value=5
 4000004
                                  column=txn details:fname,
                                                            timestamp=1538490156137, value=Gretchen
 4000005
                                  column=txn_details:count, timestamp=1538490156137, value=5
                                                                                      value=Karen
 4000005
                                  column=txn details:fname, timestamp=1538490156137,
 4000006
                                  column=txn_details:count, timestamp=1538490156137,
                                                                                      value=5
 4000006
                                  column=txn details:fname, timestamp=1538490156137,
                                                                                      value=Patrick
                                  column=txn_details:count, timestamp=1538490156137, value=6
 4000007
 4000007
                                  column=txn details:fname, timestamp=1538490156137,
                                                                                      value=Elsie
                                  column=txn_details:count, timestamp=1538490156137, value=10
 4000008
 4000008
                                  column=txn details:fname, timestamp=1538490156137,
                                                                                      value=Hazel
 4000009
                                  column=txn details:count, timestamp=1538490156137, value=6
 4000009
                                  column=txn_details:fname, timestamp=1538490156137, value=Malcolm
 4000010
                                  column=txn details:count, timestamp=1538490156137, value=6
 4000010
                                  column=txn details:fname, timestamp=1538490156137, value=Dolores
10 row(s) in 0.0300 seconds
hbase(main):015:0>
```

3. We can observe that the populated data from the above table contains all the details that are populated from Hive.

#### Task 6:

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

Package myHiveProgram; import java.io.IOException; import org.apache.hadoop.conf.Configuration; import org.apache.hadoop.hbase.HBaseConfiguration; import org.apache.hadoop.hbase.util.Bytes; import org.apache.hadoop.hbase.client.HTables; import org.apache.haboop.hbase.client.Result; import org.apache.hadoop.hbase.client.ResultScanner; import org.apache.hadoop.hbase.client.Scan; public class ScanTable { public static void main(String args[]) throuws IOExceptions { Configuration config = HBaseConfiguration.create(); HTabe table = new HTable(config, "transactions"); //instantiate the Scan class Scan scan = new Scan(); //scan the columns scan.addColumn(Bytes.toBytes("txn details"), Bytes.toBytes("fname")); scan.addColumn(Bytes.toBytes("txn details"), Bytes.toBytes("count")); //Get the ResultScanner

for(Result result = scanner.next(); result != null; result=scanner.next())

ResultScanner scanner = table.getScanner(scan);

scanner.close();

System.out.println("Found row:" + result);

1. Write Java Program for scan the TRANSACTIONS table as follows:

```
table.close();
}
```

```
Package Explorer ⋈
                                    49 import java.io.IOException:
                  E 8
                                                 import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.hbase.HBaseConfiguration;
import org.apache.hadoop.hbase.util.Bytes;
 myHBaseProgram
  import org.apache.hadoop.hbase.client.HTable;
  ▶ ■ JRE System Library [JavaSE
                                                 import org.apache.hadoop.hbase.client.Result:
     src src
                                                 import org.apache.hadoop.hbase.client.ResultScanner;
     11
                                                 import org.apache.hadoop.hbase.client.Scan;
         ScanTable.java
                                            13
                                                 public class ScanTable{
   Referenced Libraries
                                                     public static void main(String args[]) throws IOException{
   Configuration config = HBaseConfiguration.create();
   HTable table = new HTable(config, "transactions");
                                            16
                                            18
                                            20
21
                                                         Scan scan = new Scan();
                                                         scan.addColumn(Bytes.toBytes("txn_details"), Bytes.toBytes("fname"));
scan.addColumn(Bytes.toBytes("txn_details"), Bytes.toBytes("count"));
                                            23
24
                                            25
26
                                                          // get the ResultScanner
                                                         for (Result result = scanner.next(); result != null; result=scanner.next())
    System.out.println("Found row : " + result);
                                            27
28
                                            30
                                            32
                                                         table.close();
```

2. When we run the above program, we will get the following output showing that the column family values, that is the customer id, customer count and its length, customer name and its length.

```
<terminated> ScanTable [Java Application] /usr/java/jdk1.8.0_151/bin/java (Oct 2, 2018, 9:49:23 PM)
keyvalues={4000001/txn details:count/1538490156137/Put/vlen=1/seqid=0,
                                                                                            4000001/txn details:fname/1538490156137/Put/vlen
keyvalues={4000002/txn_details:count/1538490156137/Put/vlen=1/seqid=0,
keyvalues={4000003/txn_details:count/1538490156137/Put/vlen=1/seqid=0,
keyvalues={4000004/txn_details:count/1538490156137/Put/vlen=1/seqid=0,
                                                                                            4000002/txn_details:fname/1538490156137/Put/vlen
                                                                                            4000003/txn details:fname/1538490156137/Put/vlen
                                                                                            4000004/txn_details:fname/1538490156137/Put/vlen
keyvalues={4000005/txn_details:count/1538490156137/Put/vlen=1/seqid=0,
keyvalues={4000006/txn_details:count/1538490156137/Put/vlen=1/seqid=0,
                                                                                            4000005/txn_details:fname/1538490156137/Put/vlen
                                                                                            4000006/txn details:fname/1538490156137/Put/vlen
keyvalues={4000007/txn_details:count/1538490156137/Put/vlen=1/seqid=0,
                                                                                            4000007/txn_details:fname/1538490156137/Put/vlen
keyvalues={4000008/txn_details:count/1538490156137/Put/vlen=2/seqid=0,
keyvalues={4000009/txn_details:count/1538490156137/Put/vlen=1/seqid=0,
                                                                                            4000008/txn_details:fname/1538490156137/Put/vlen
                                                                                            4000009/txn details:fname/1538490156137/Put/vlen
keyvalues={4000010/txn_details:count/1538490156137/Put/vlen=1/seqid=0,
                                                                                            4000010/txn_details:fname/1538490156137/Put/vlen
```