

To-do

Please perform the following steps:

1. Start the hive shell

```

[acadgild@localhost ~]$ sudo service mysql start
Starting mysqld:
[acadgild@localhost ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/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.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>

```

2. Create the data in hive shell using **create database acadgilddb;** command.

```

[acadgild@localhost ~]$ hive
hive> create database acadgilddb;
OK
Time taken: 0.093 seconds
hive>

```

3. Execute **use acadgilddb;** command to get into the database

```

[acadgild@localhost ~]$ hive
hive> use acadgilddb;
OK
Time taken: 0.069 seconds
hive>

```

4. Now create an internal table by the name customer as follows :-

```

CREATE TABLE CUSTOMER(
    custid INT,
    fname STRING,
    lname STRING,
    age INT,
    profession STRING)
row format delimited fields terminated by ',';

```

```

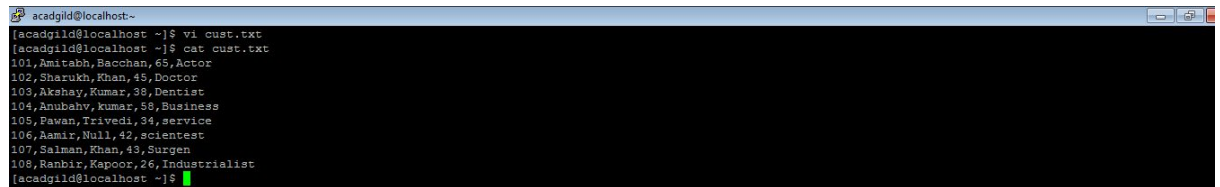
[acadgild@localhost ~]$ hive
hive> CREATE TABLE CUSTOMER(custid INT,fname STRING,lname STRING,age INT,profession STRING)row format delimited fields terminated by ',';
OK
Time taken: 1.27 seconds
hive>

```

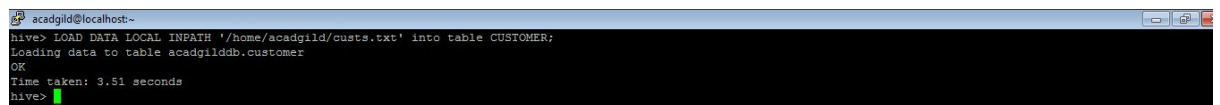
5. Load this table with the following command

LOAD DATA LOCAL INPATH '<your location of attached custs.txt file>/custs.txt'
into table CUSTOMER;

```
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
```



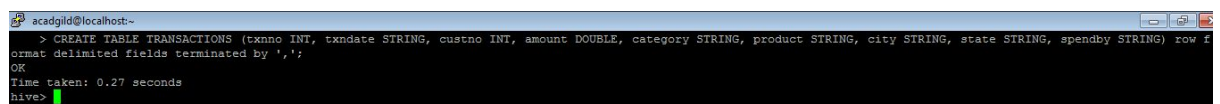
```
acadgild@localhost:~$ vi cust.txt
acadgild@localhost:~$ cat cust.txt
101,Amitabh,Bacchan,65,Actor
102,Sharukh,Khan,45,Doctor
103,Akshay,Kumar,38,Dentist
104,Anubahv,kumar,58,Business
105,Pawan,Trivedi,34,service
106,Aamir,Null,42,scientest
107,Salman,Khan,43,Surgen
108,Ranbir,Kapoor,26,Industrialist
acadgild@localhost:~$
```



```
hive> LOAD DATA LOCAL INPATH '/home/acadgild/custs.txt' into table CUSTOMER;
Loading data to table acadgild.db.customer
OK
Time taken: 3.51 seconds
hive>
```

6. Now create the another internal table by the name transaction as follows :

```
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 ',';
```

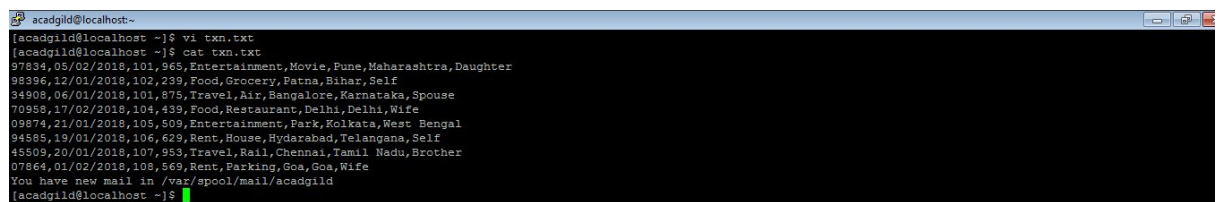


```
acadgild@localhost:~$
> CREATE TABLE TRANSACTIONS (txnno INT, txndate STRING, custno INT, amount DOUBLE, category STRING, product STRING, city STRING, state STRING, spendby STRING) row f
ormat delimited fields terminated by ',';
OK
Time taken: 0.27 seconds
hive>
```

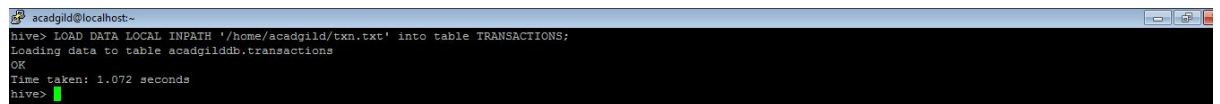
7. Load this table with the following command

LOAD DATA LOCAL INPATH '<you location of attached custs.txt file>/txn.txt'
into table TRANSACTIONS;

```
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
```



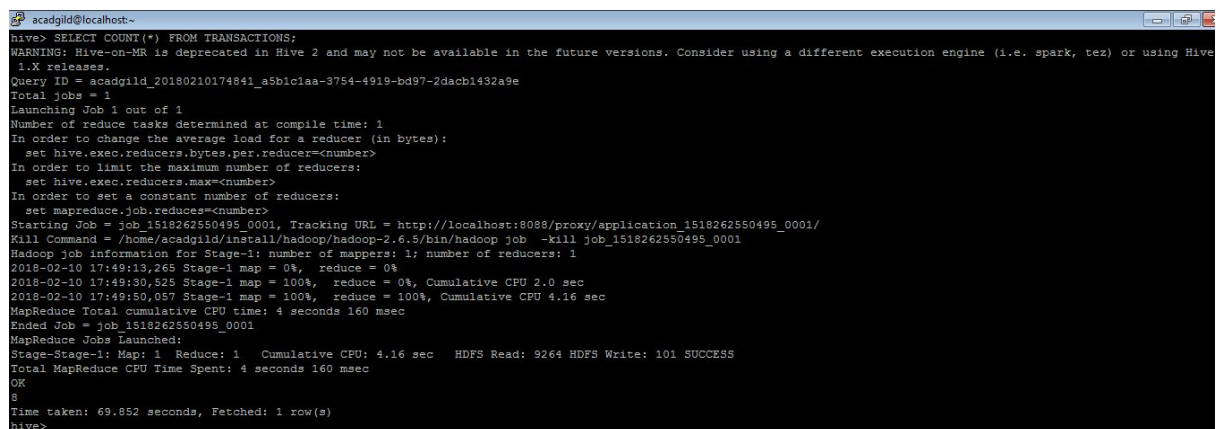
```
acadgild@localhost:~$ vi txn.txt
acacgild@localhost:~$ cat txn.txt
97834,05/02/2018,101,965,Entertainment,Movie,Pune,Maharashtra,Daughter
98396,12/01/2018,102,239,Food,Grocery,Patna,Bihar,Self
34908,06/01/2018,101,875,Travel,Air,Bangalore,Karnataka,Spouse
70958,17/02/2018,104,439,Food,Restaurant,Delhi,Delhi,Wife
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
You have new mail in /var/spool/mail/acacgild
acacgild@localhost:~$
```



```
hive> LOAD DATA LOCAL INPATH '/home/acacgild/txn.txt' into table TRANSACTIONS;
Loading data to table acacgild.db.transactions
OK
Time taken: 1.072 seconds
hive>
```

8. Now perform the following queries on the above two tables :-

SELECT COUNT(*) FROM TRANSACTIONS;



```
acacgild@localhost:~$
hive> SELECT COUNT(*) FROM TRANSACTIONS;
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 = acacgild_20180210174841_a5b1c1aa-3754-4919-bd97-2dadb1432a9e
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 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_1518262550495_0001, Tracking URL = http://localhost:8088/proxy/application_1518262550495_0001/
Kill Command = /home/acacgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1518262550495_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-02-10 17:49:13,265 Stage-1 map = 0%, reduce = 0%
2018-02-10 17:49:30,525 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.0 sec
2018-02-10 17:49:50,057 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.16 sec
MapReduce Total cumulative CPU time: 4 seconds 160 msec
Ended Job = job_1518262550495_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.16 sec HDFS Read: 9264 HDFS Write: 101 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 160 msec
OK
8
Time taken: 69.852 seconds, Fetched: 1 row(s)
hive>
```

select a.fname, b.amount from CUSTOMER a join TRANSACTIONS b on a.custid
=b.custno;

ACADGILD – SESSION8(HIVE BASICS)

```
acadgild@localhost:~  
> select a.fname, b.amount from CUSTOMER a join TRANSACTIONS b on a.custid =b.custno;  
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_20180210175028_5a61c021-9ae6-4f3c-9653-15e9222f6a7b  
Total jobs = 1  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.  
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]  
2018-02-10 17:50:45 Starting to launch local task to process map join: maximum memory = 518979584  
2018-02-10 17:50:53 Dump the side-table for tag: 0 with group count: 8 into file: file:/tmp/acadgild/27a0e4b0-a8da-4e14-894c-c01762022e9e/hive_2018-02-10_17-50-28_2  
33_828298095125264230-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile00--.hashtable  
2018-02-10 17:50:53 Uploaded 1 File to: file:/tmp/acadgild/27a0e4b0-a8da-4e14-894c-c01762022e9e/hive_2018-02-10_17-50-28_233_828298095125264230-1/-local-10004/Hash  
table-Stage-3/MapJoin-mapfile00--.hashtable (469 bytes)  
2018-02-10 17:50:53 End of local task; Time Taken: 7.465 sec.  
Execution completed successfully  
MapredLocal task succeeded  
Launching Job 1 out of 1  
Number of reduce tasks is set to 0 since there's no reduce operator  
Starting Job = job_1518262550495_0002, Tracking URL = http://localhost:8088/proxy/application_1518262550495_0002/  
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1518262550495_0002  
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0  
2018-02-10 17:51:07,810 Stage-3 map = 0%, reduce = 0%  
2018-02-10 17:51:20,977 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.81 sec  
MapReduce Total cumulative CPU time: 3 seconds 50 msec  
Ended Job = job_1518262550495_0002  
MapReduce Jobs Launched:  
Stage-Stage-3: Map: 1 Cumulative CPU: 3.05 sec HDFS Read: 7352 HDFS Write: 289 SUCCESS  
Total MapReduce CPU Time Spent: 3 seconds 50 msec  
OK  
Amitabh 965.0  
Sharukh 239.0  
Amitabh 875.0  
Anubhav 439.0  
Pavan 509.0  
Aamir 629.0  
Salman 953.0  
Ranbir 569.0  
Time taken: 57.167 seconds, Fetched: 8 row(s)  
hive>
```

SELECT CATEGORY, SUM(amount) FROM TRANSACTIONS group by category;

```
acadgild@localhost:~  
hive> SELECT CATEGORY, SUM(amount) FROM TRANSACTIONS group by category;  
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_20180210175226_27eedb6f-a701-43d3-b09f-cac1e8caa479  
Total jobs = 1  
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_1518262550495_0003, Tracking URL = http://localhost:8088/proxy/application_1518262550495_0003/  
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1518262550495_0003  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
2018-02-10 17:52:41,883 Stage-1 map = 0%, reduce = 0%  
2018-02-10 17:52:57,194 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.99 sec  
2018-02-10 17:53:19,421 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.78 sec  
MapReduce Total cumulative CPU time: 3 seconds 780 msec  
Ended Job = job_1518262550495_0003  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.78 sec HDFS Read: 9761 HDFS Write: 193 SUCCESS  
Total MapReduce CPU Time Spent: 3 seconds 780 msec  
OK  
Entertainment 1474.0  
Food 678.0  
Rent 1199.0  
Travel 1828.0  
Time taken: 52.671 seconds, Fetched: 4 row(s)  
hive>
```

SELECT CATEGORY, SUM(amount) FROM TRANSACTIONS GROUP BY category
SORT BY category DESC;

ACADGILD – SESSION8(HIVE BASICS)

```
acdgild@localhost:~$
hive> SELECT CATEGORY, SUM(amount) FROM TRANSACTIONS GROUP BY category SORT BY category DESC;
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 = acdgild_20180210175354_3d9b0661-4bc7-4df2-b763-e6aca8e24e02
Total jobs = 1
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_1518262550495_0004, Tracking URL = http://localhost:8080/proxy/application_1518262550495_0004/
Kill Command = /home/acdgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1518262550495_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-02-10 17:54:06,943 Stage-1 map = 0%, reduce = 0%
2018-02-10 17:54:15,875 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.25 sec
2018-02-10 17:54:24,498 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.92 sec
MapReduce Total cumulative CPU time: 2 seconds 920 msec
Ended Job = job_1518262550495_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.92 sec HDFS Read: 9672 HDFS Write: 193 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 920 msec
OK
Travel 1628.0
Rent 1198.0
Food 678.0
Entertainment 1474.0
Time taken: 30.684 seconds, Fetched: 4 row(s)
hive>
```