

ASSIGNMENT

TASK: Transactions in Hive

1. To work on insert, delete and update command added on hive 0.14

1.a. open hive shell and transactions features need to be turned on.

->set hive.support.concurrency = true;

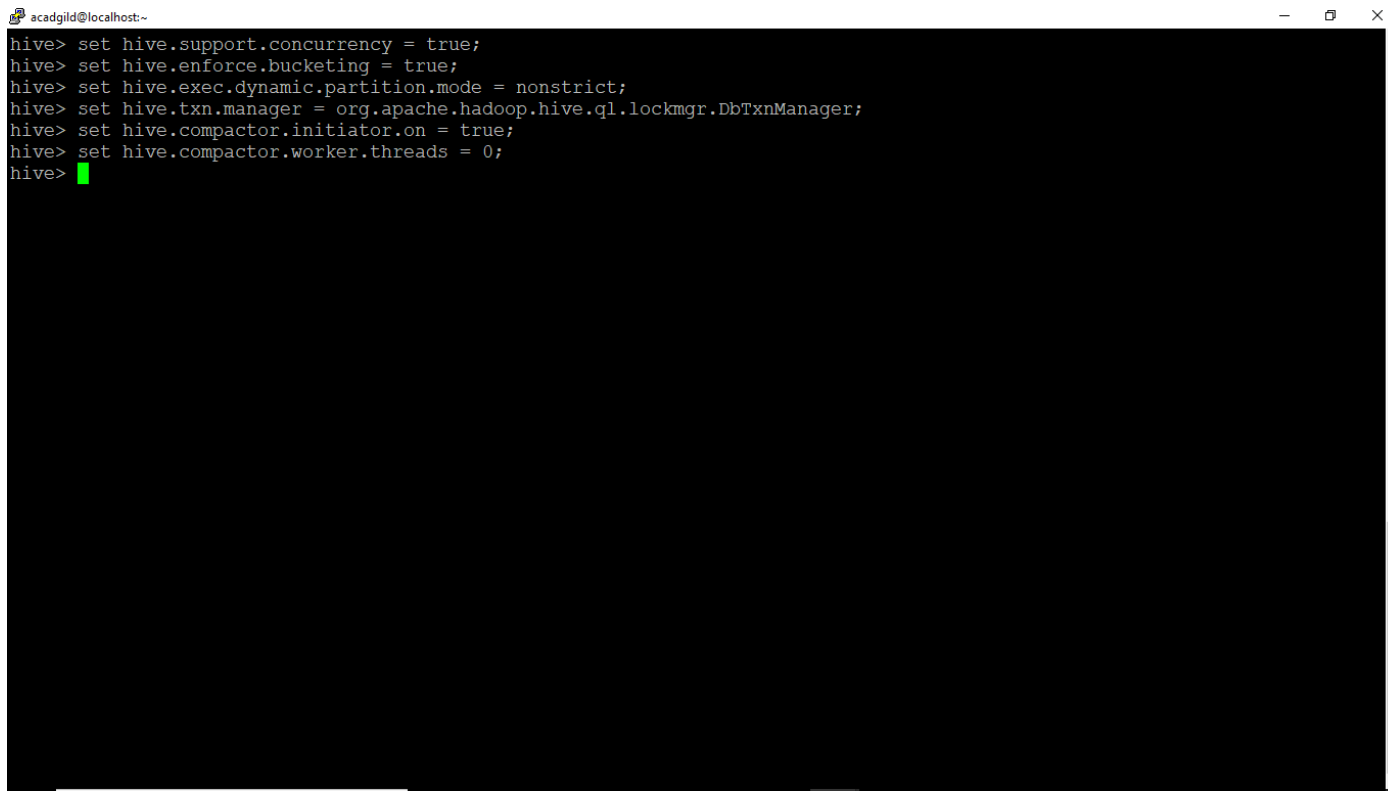
->set hive.enforce.bucketing = true;

->set hive.exec.dynamic.partition.mode = nonstrict;

->set hive.txn.manager = org.apache.hadoop.hive.ql.lockmgr.DbTxnManager;

->set hive.compactor.initiator.on = true;

->set hive.compactor.worker.threads = 0;

A screenshot of a terminal window with a black background and white text. The window title is 'acadgild@localhost:~'. The terminal shows a series of Hive configuration commands being entered at the 'hive>' prompt. The commands are: 'set hive.support.concurrency = true;', 'set hive.enforce.bucketing = true;', 'set hive.exec.dynamic.partition.mode = nonstrict;', 'set hive.txn.manager = org.apache.hadoop.hive.ql.lockmgr.DbTxnManager;', 'set hive.compactor.initiator.on = true;', and 'set hive.compactor.worker.threads = 0;'. The last line shows the prompt 'hive>' followed by a green cursor. The terminal window has standard Linux window controls (minimize, maximize, close) in the top right corner.

```
acadgild@localhost:~  
hive> set hive.support.concurrency = true;  
hive> set hive.enforce.bucketing = true;  
hive> set hive.exec.dynamic.partition.mode = nonstrict;  
hive> set hive.txn.manager = org.apache.hadoop.hive.ql.lockmgr.DbTxnManager;  
hive> set hive.compactor.initiator.on = true;  
hive> set hive.compactor.worker.threads = 0;  
hive> █
```

1.b. create a table

-> command used: `CREATE TABLE college(clg_id int,clg_name string,clg_loc string) clustered by (clg_id) into 5 buckets stored as orc TBLPROPERTIES('transactional'='true');`

```
acagild@localhost:~
hive> CREATE TABLE college(clg_id int,clg_name string,clg_loc string) clustered by (clg_id) into 5 buckets stored as orc TBLPROPERTIES('transactional'='true');
Thu May 17 22:55:51 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
Thu May 17 22:55:51 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
Thu May 17 22:55:51 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
Thu May 17 22:55:51 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
Thu May 17 22:55:51 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
Thu May 17 22:55:52 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
OK
Time taken: 5.802 seconds
hive>
```

```
acagild@localhost:~
hive> show tables
+-----+
OK
buck users
college
json table
locations
store view 1
temperature_data
train_sales_stage
users
users_partitioned
Time taken: 0.707 seconds, Fetched: 9 row(s)
hive>
```

1.c. Insert the data into the table

-> command used: INSERT INTO table college

values(1,'sskp','pun'),(2,'vit','vlr'),(3,'srm','chen'),(4,'lpu','jul'),(5,'stanford','uk'),(6,'lsr','del'),(7,'cambridge','us');

```
acadgild@localhost:~$
Time taken: 0.707 seconds, Fetched: 9 row(s)
hive> INSERT INTO table college values(1,'sskp','pun'),(2,'vit','vlr'),(3,'srm','chen'),(4,'lpu','jul'),(5,'stanford','uk')
Thu May 17 22:57:44 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended.
requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing app
icate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true an
rification.
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different ex
1.X releases.
Query ID = acadgild_20180517225737_4d2943fd-91a0-4113-9d89-7c12e90a6e2a
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 5
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>
Thu May 17 22:57:44 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended.
requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing app
icate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true an
rification.
Starting Job = job_1526577423329_0001, Tracking URL = http://localhost:8088/proxy/application_1526577423329_0001/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.9.0/bin/hadoop job -kill job_1526577423329_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2018-05-17 22:58:14,917 Stage-1 map = 0%, reduce = 0%
2018-05-17 22:58:30,923 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.71 sec
2018-05-17 22:59:19,126 Stage-1 map = 100%, reduce = 13%, Cumulative CPU 5.47 sec
2018-05-17 22:59:31,554 Stage-1 map = 100%, reduce = 27%, Cumulative CPU 8.59 sec
2018-05-17 22:59:32,819 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 10.57 sec
2018-05-17 22:59:35,602 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 12.52 sec
2018-05-17 22:59:37,009 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 14.46 sec
2018-05-17 22:59:51,487 Stage-1 map = 100%, reduce = 73%, Cumulative CPU 22.85 sec
2018-05-17 22:59:58,712 Stage-1 map = 100%, reduce = 80%, Cumulative CPU 26.61 sec
2018-05-17 22:59:59,864 Stage-1 map = 100%, reduce = 87%, Cumulative CPU 30.36 sec
```

```
acadgild@localhost:~$
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 5
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>
Thu May 17 22:57:44 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended.
requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing app
icate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true an
rification.
Starting Job = job_1526577423329_0001, Tracking URL = http://localhost:8088/proxy/application_1526577423329_0001/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.9.0/bin/hadoop job -kill job_1526577423329_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2018-05-17 22:58:14,917 Stage-1 map = 0%, reduce = 0%
2018-05-17 22:58:30,923 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.71 sec
2018-05-17 22:59:19,126 Stage-1 map = 100%, reduce = 13%, Cumulative CPU 5.47 sec
2018-05-17 22:59:31,554 Stage-1 map = 100%, reduce = 27%, Cumulative CPU 8.59 sec
2018-05-17 22:59:32,819 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 10.57 sec
2018-05-17 22:59:35,602 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 12.52 sec
2018-05-17 22:59:37,009 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 14.46 sec
2018-05-17 22:59:51,487 Stage-1 map = 100%, reduce = 73%, Cumulative CPU 22.85 sec
2018-05-17 22:59:58,712 Stage-1 map = 100%, reduce = 80%, Cumulative CPU 26.61 sec
2018-05-17 22:59:59,864 Stage-1 map = 100%, reduce = 87%, Cumulative CPU 30.36 sec
2018-05-17 23:00:01,034 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 37.66 sec
MapReduce Total cumulative CPU time: 37 seconds 660 msec
Ended Job = job_1526577423329_0001
Loading data to table default.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 5 Cumulative CPU: 37.66 sec HDFS Read: 26736 HDFS Write: 3994 SUCCESS
Total MapReduce CPU Time Spent: 37 seconds 660 msec
OK
Time taken: 147.371 seconds
hive>
```

➔ Check the data populated

- Command used: `select * from college;`

```
acadgild@localhost:~$  
hive> select * from college;  
OK  
5      stanford      uk  
6      lsr      del  
1      sskp      pun  
7      cambridge      us  
2      vit      vlr  
3      srm      chen  
4      lpu      jul  
Time taken: 0.751 seconds, Fetched: 7 row(s)  
hive>
```

1. d. Update the data:

-> command used: `UPDATE college set clg_id = 8 where clg_id = 7;`

-> ***FAILED: SemanticException[Error 10302]: Updating values of bucketing columns is not supported. Column clg_id***

```
acadgild@localhost:~$  
users_partitioned  
Time taken: 13.749 seconds, Fetched: 9 row(s)  
hive>  
hive> UPDATE college set clg_id = 8 where clg_id = 7;  
FAILED: SemanticException [Error 10302]: Updating values of bucketing columns is not supported. Column clg_id.  
hive>
```

1. E. UPDATE college set clg_name = 'IIT' where clg_id = 6;

```
acagild@localhost:~
hive>
hive> UPDATE college set clg_name = 'IIT' where clg_id = 6;
Fri May 18 07:18:58 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended.
According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit opti
on isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'fals
e'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for serve
r certificate verification.
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different ex
ecution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acagild_20180518071856_979d7561-4244-499c-a9b6-f9f9ffc407b8
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 5
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>
Fri May 18 07:18:58 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended.
According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit opti
on isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'fals
e'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for serve
r certificate verification.
Starting Job = job_1526602885203_0001, Tracking URL = http://localhost:8088/proxy/application_1526602885203_0001/
Kill Command = /home/acagild/install/hadoop/hadoop-2.9.0/bin/hadoop job -kill job_1526602885203_0001
```

```
acagild@localhost:~
Query ID = acagild_20180518071856_979d7561-4244-499c-a9b6-f9f9ffc407b8
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 5
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>
Fri May 18 07:18:58 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended.
According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit opti
on isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'fals
e'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for serve
r certificate verification.
Starting Job = job_1526602885203_0001, Tracking URL = http://localhost:8088/proxy/application_1526602885203_0001/
Kill Command = /home/acagild/install/hadoop/hadoop-2.9.0/bin/hadoop job -kill job_1526602885203_0001
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-05-18 07:19:30,175 Stage-1 map = 0%, reduce = 0%
2018-05-18 07:20:30,314 Stage-1 map = 0%, reduce = 0%
2018-05-18 07:20:39,458 Stage-1 map = 40%, reduce = 0%, Cumulative CPU 8.19 sec
2018-05-18 07:20:41,773 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.5 sec
2018-05-18 07:21:32,971 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 29.06 sec
2018-05-18 07:21:34,186 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 31.1 sec
2018-05-18 07:21:44,242 Stage-1 map = 100%, reduce = 93%, Cumulative CPU 39.7 sec
2018-05-18 07:21:45,435 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 41.2 sec
MapReduce Total cumulative CPU time: 41 seconds 200 msec
Ended Job = job_1526602885203_0001
Loading data to table default.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 41.2 sec HDFS Read: 51382 HDFS Write: 937 SUCCESS
Total MapReduce CPU Time Spent: 41 seconds 200 msec
OK
Time taken: 173.659 seconds
hive>
```

Command used to check data

➔ Select * from college;

```
acadgild@localhost:~  
hive> select * from college;  
OK  
5      stanford      uk  
6      IIT      del  
1      sskp      pun  
7      cambridge      us  
2      vit      vlr  
3      srm      chen  
4      lpu      jul  
Time taken: 4.196 seconds, Fetched: 7 row(s)  
hive>
```

1. F. To delete the data

➔ Command used: delete from college where clg_id=5;

```
acadgild@localhost:~  
Query ID = acadgild_20180518072559_67e18c4a-fac9-48b2-bb2a-428956c07d0c  
Total jobs = 1  
Launching Job 1 out of 1  
Number of reduce tasks determined at compile time: 5  
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_1526602885203_0002, Tracking URL = http://localhost:8088/proxy/application_1526602885203_0002/  
Kill Command = /home/acadgild/install/hadoop/hadoop-2.9.0/bin/hadoop job -kill job_1526602885203_0002  
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5  
2018-05-18 07:26:21,010 Stage-1 map = 0%, reduce = 0%  
2018-05-18 07:27:22,151 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 3.09 sec  
2018-05-18 07:27:23,529 Stage-1 map = 20%, reduce = 0%, Cumulative CPU 4.11 sec  
2018-05-18 07:27:25,946 Stage-1 map = 40%, reduce = 0%, Cumulative CPU 8.29 sec  
2018-05-18 07:27:28,476 Stage-1 map = 60%, reduce = 0%, Cumulative CPU 12.63 sec  
2018-05-18 07:27:29,731 Stage-1 map = 80%, reduce = 0%, Cumulative CPU 16.57 sec  
2018-05-18 07:27:33,607 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 20.71 sec  
2018-05-18 07:28:20,419 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 27.02 sec  
2018-05-18 07:28:22,791 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 29.0 sec  
2018-05-18 07:28:25,206 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 31.07 sec  
2018-05-18 07:28:30,075 Stage-1 map = 100%, reduce = 80%, Cumulative CPU 34.24 sec  
2018-05-18 07:28:31,220 Stage-1 map = 100%, reduce = 87%, Cumulative CPU 35.52 sec  
2018-05-18 07:28:32,284 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 39.27 sec  
MapReduce Total cumulative CPU time: 39 seconds 270 msec  
Ended Job = job_1526602885203_0002  
Loading data to table default.college  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 39.27 sec HDFS Read: 49866 HDFS Write: 752 SUCCESS  
Total MapReduce CPU Time Spent: 39 seconds 270 msec  
OK  
Time taken: 155.16 seconds  
hive>
```

```

acadgild@localhost:~
hive>
hive> delete from college where clg_id=5;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different ex
ecution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180518072559_67e18c4a-fac9-48b2-bb2a-428956c07d0c
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 5
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_1526602885203_0002, Tracking URL = http://localhost:8088/proxy/application_1526602885203_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.9.0/bin/hadoop job -kill job_1526602885203_0002
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-05-18 07:26:21,010 Stage-1 map = 0%, reduce = 0%
2018-05-18 07:27:22,151 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 3.09 sec
2018-05-18 07:27:23,529 Stage-1 map = 20%, reduce = 0%, Cumulative CPU 4.11 sec
2018-05-18 07:27:25,946 Stage-1 map = 40%, reduce = 0%, Cumulative CPU 8.29 sec
2018-05-18 07:27:28,476 Stage-1 map = 60%, reduce = 0%, Cumulative CPU 12.63 sec
2018-05-18 07:27:29,731 Stage-1 map = 80%, reduce = 0%, Cumulative CPU 16.57 sec
2018-05-18 07:27:33,607 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 20.71 sec
2018-05-18 07:28:20,419 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 27.02 sec
2018-05-18 07:28:22,791 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 29.0 sec
2018-05-18 07:28:25,206 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 31.07 sec
2018-05-18 07:28:30,075 Stage-1 map = 100%, reduce = 80%, Cumulative CPU 34.24 sec
2018-05-18 07:28:31,220 Stage-1 map = 100%, reduce = 87%, Cumulative CPU 35.52 sec
2018-05-18 07:28:32,284 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 39.27 sec
MapReduce Total cumulative CPU time: 39 seconds 270 msec
Ended Job = job_1526602885203_0002
Loading data to table default.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 39.27 sec HDFS Read: 49866 HDFS Write: 752 SUCCESS

```

- ➔ to check the data
- ➔ select * from college;

```

acadgild@localhost:~
> select * from college;
OK
6      IIT      del
1      sskp     pun
7      cambridge  us
2      vit      vlr
3      srm      chen
4      lpu      jul
Time taken: 0.484 seconds, Fetched: 6 row(s)
hive>

```