**Assignment 8.3**

**Associated Data Files**

<https://acadgild.com/blog/transactions-in-hive/>

**Problem Statement:**

Refer the above given link for transactions in Hive and implement the operations given in the blog using your own sample data set and send us the screenshot.

Solution :

## What is ACID?

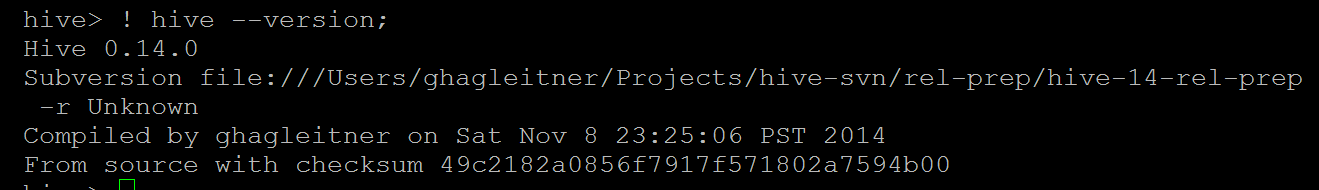
ACID stands for Atomicity, Consistency, Isolation, and Durability.

Atomicity means, a transaction should complete successfully or else it should fail completely i.e. it should not be left partially. Consistency ensures that any transaction will bring the database from one valid state to another state. Isolation states that every transaction should be independent of each other i.e. one transaction should not affect another. And Durability states that if a transaction is completed, it should be preserved in the database even if the machine state is lost or a system failure might occur.

These ACID properties are essential for a transaction and every transaction should ensure that these properties are met.

**Transactions in Hive**

Transactions in Hive are introduced in Hive 0.13, but they only partially fulfill the ACID properties like atomicity, consistency, durability, at the partition level. Here, Isolation can be provided by turning on one of the locking mechanisms available with zookeeper or in memory.



But in Hive 0.14, new API’s have been added to completely fulfill the ACID properties while performing any transaction.

Transactions are provided at the row-level in Hive 0.14. The different row-level transactions available in Hive 0.14 are as follows:

1. Insert
2. Delete
3. Update

There are numerous limitations with the present transactions available in Hive 0.14. ORC is the file format supported by Hive transaction. It is now essential to have ORC file format for performing transactions in Hive. The table needs to be bucketed in order to support transactions.

**Row-level Transactions Available in Hive 0.14**

Let’s perform some row-level transactions available in Hive 0.14. Before creating a Hive table that supports transactions, the transaction features present in Hive needs to be turned on, as by default they are turned off.

The below properties needs to be set appropriately in ***hive shell***, order-wise to work with transactions in Hive:

set hive.support.concurrency = true;

set hive.enforce.bucketing = true;

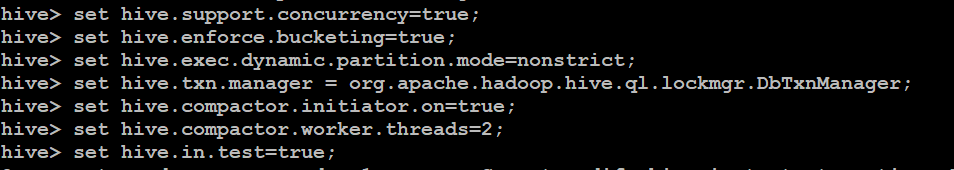
set hive.exec.dynamic.partition.mode = nonstrict;

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

set hive.compactor.initiator.on = true;

set hive.compactor.worker.threads = 2;

set hive.in.test=true;



If the above properties are not set properly, the ‘Insert’ operation will work but ‘Update’ and ‘Delete’ will not work and you will receive the following error:

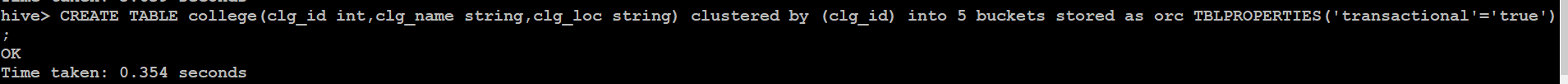
FAILED: SemanticException [Error 10294]: Attempt to do update or delete usingtransaction manager thatdoes not support these operations.

**Creating a Table That Supports Hive Transactions**

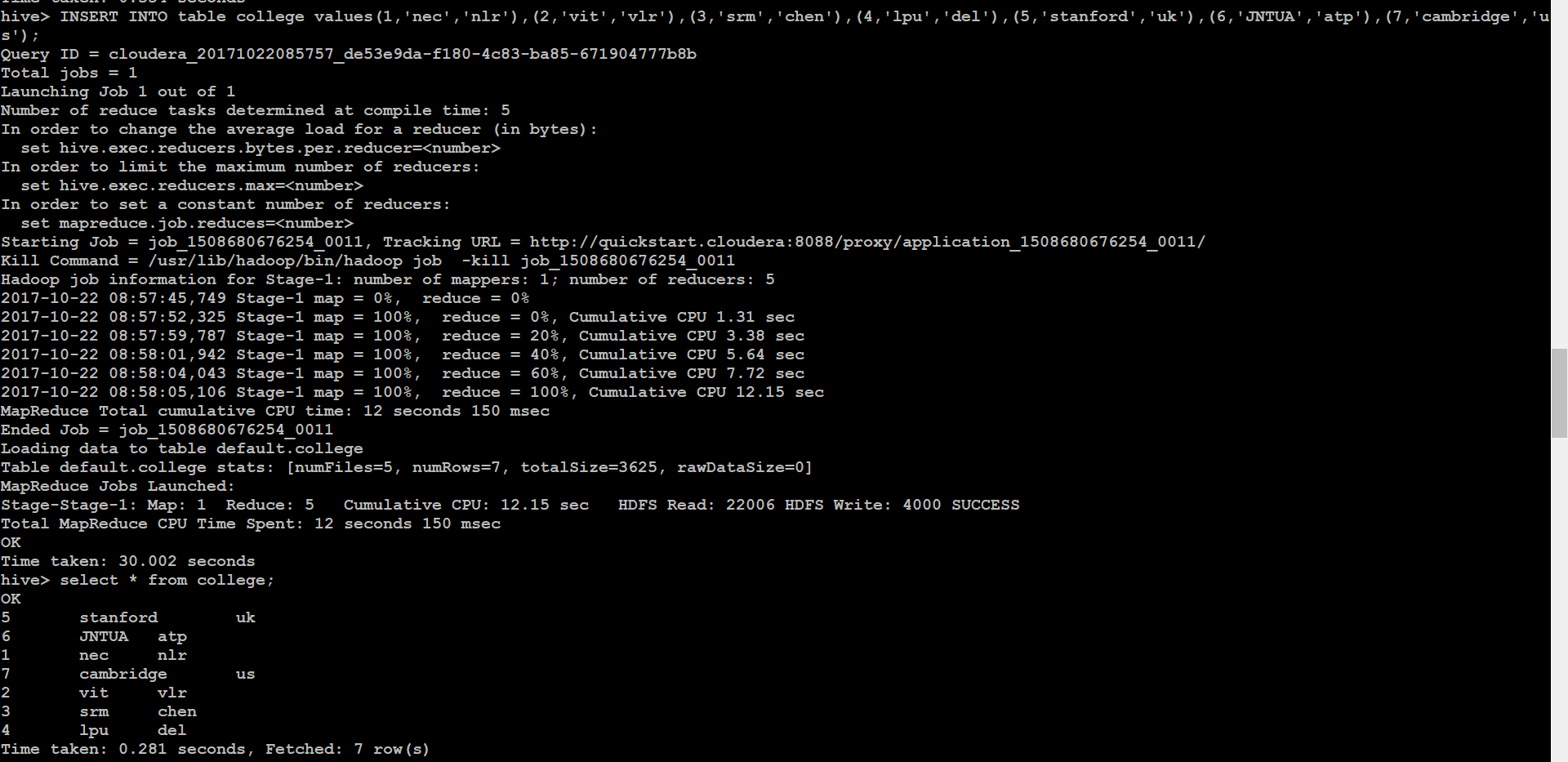
CREATE TABLE college(clg\_id int,clg\_name string,clg\_loc string)

clustered by (clg\_id) into 5 buckets stored as orc

TBLPROPERTIES('transactional'='true');



INSERT INTO table college values(1,'nec','nlr'),(2,'vit','vlr'),(3,'srm','chen'),(4,'lpu','del'),(5,'stanford','uk'),(6,'JNTUA','atp'),(7,'cambridge','us');



hive> select \* from college;

OK

5 stanforduk

6 JNTUA atp

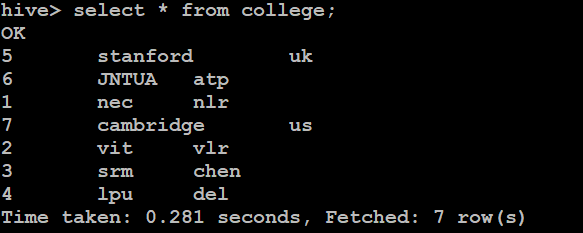
1 necnlr

7 cambridge us

2 vitvlr

3 srmchen

4 lpu del



INSERT INTO table college values(1,'nec','nlr'),(2,'vit','vlr'),(3,'srm','chen'),(4,'lpu','del'),(5,'stanford','uk'),(6,'JNTUA','atp'),(7,'cambridge','us');

Now if we try to re-insert the same data again, it will be appended to the previous data as shown below:

hive> select \* from college;

OK

5 stanforduk

5 stanforduk

6 JNTUA atp

1 necnlr

6 JNTUA atp

1 necnlr

7 cambridge us

2 vitvlr

7 cambridge us

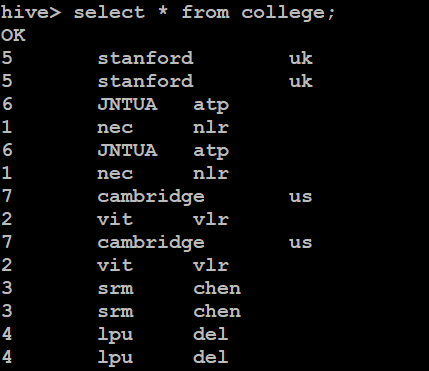
2 vitvlr

3 srmchen

3 srmchen

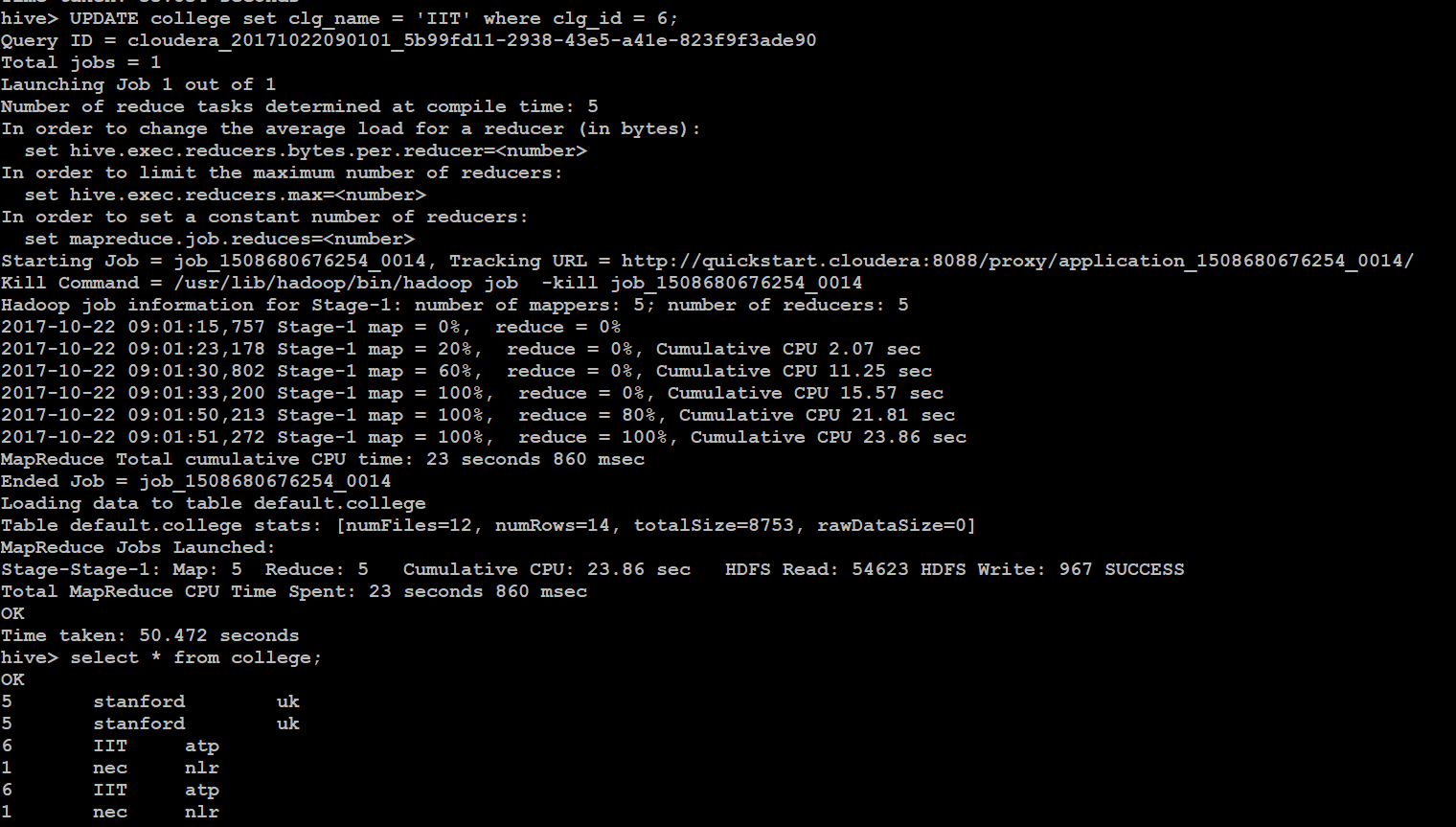
4 lpu del

4 lpu del



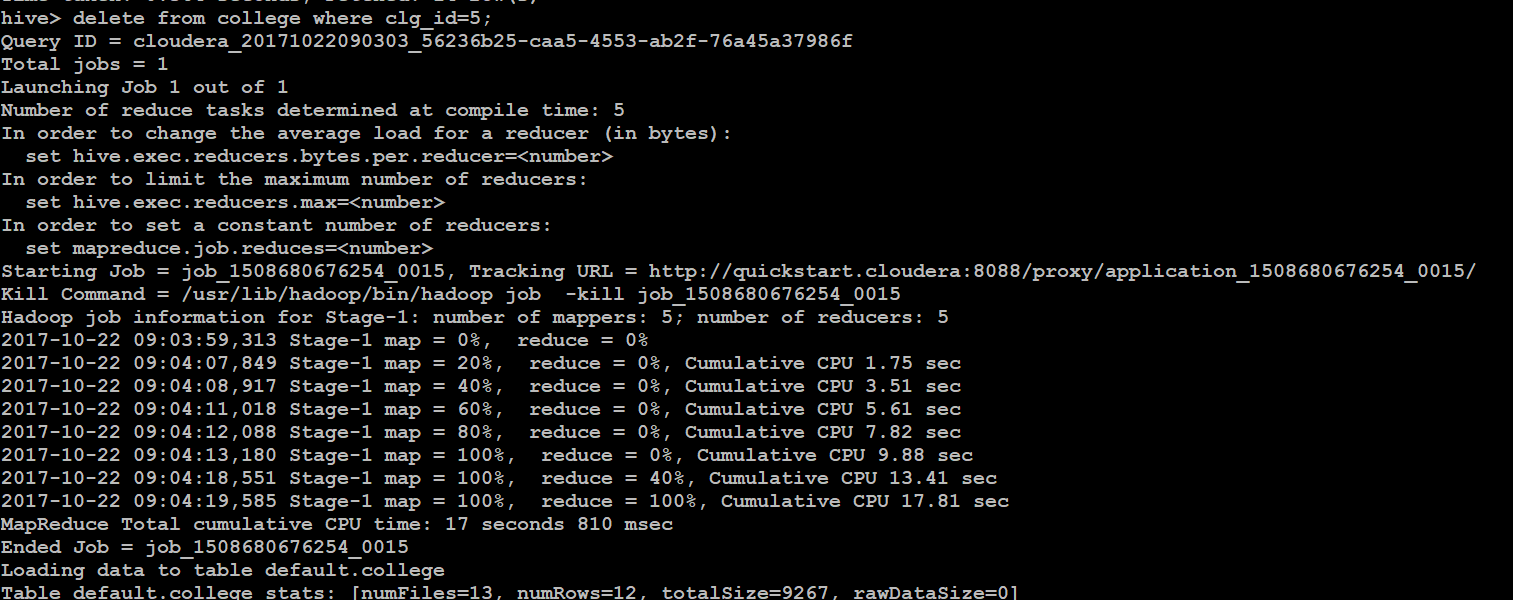
UPDATE college set clg\_name = 'IIT' where clg\_id = 6;

The above command is used to update a row in Hive table.



delete from college where clg\_id=5;

The above command will delete a single row in the Hive table.



hive> select \* from college;

OK

6 IIT atp

1 necnlr

6 IIT atp

1 necnlr

8 cambridge us

2 vitvlr

8 cambridge us

2 vitvlr

3 srmchen

3 srmchen

4 lpu del

4 lpu del

