**ASSIGNMENT 28.1**

**What are ACID properties and Explain Transactions in Hive.**

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 with ACID semantics have been added to Hive to address the following use cases:

1. Streaming ingest of data
2. Slow changing dimensionsData restatement.  Sometimes collected data is found to be incorrect and needs correction.
3. Starting with Hive 0.14 these use cases can be supported via [*INSERT*](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML)*,*[*UPDATE*](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Update), and [*DELETE*](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Delete).
4. Bulk updates using [SQL MERGE](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Merge) statement.

**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**

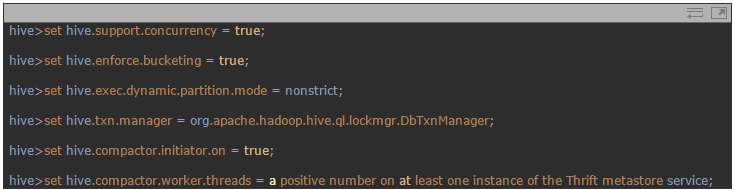
1. **Update**

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:

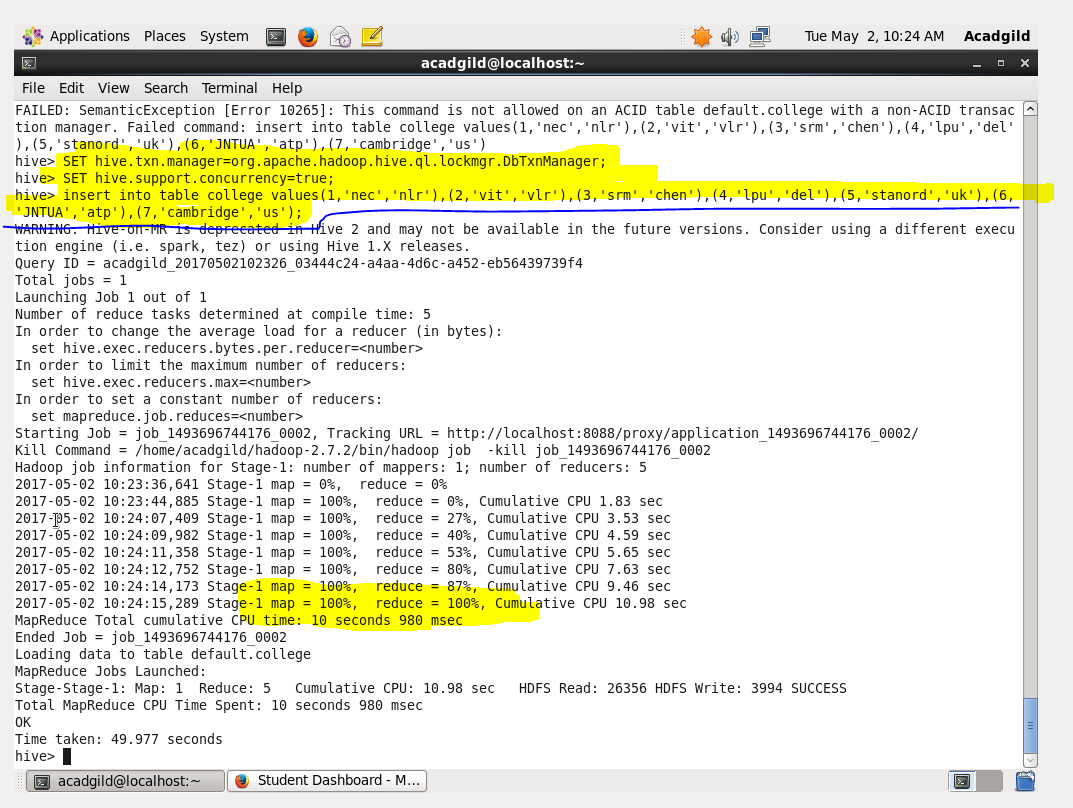


**Hive Transaction Operations**

## Creating a Table That Supports Hive Transactions

****

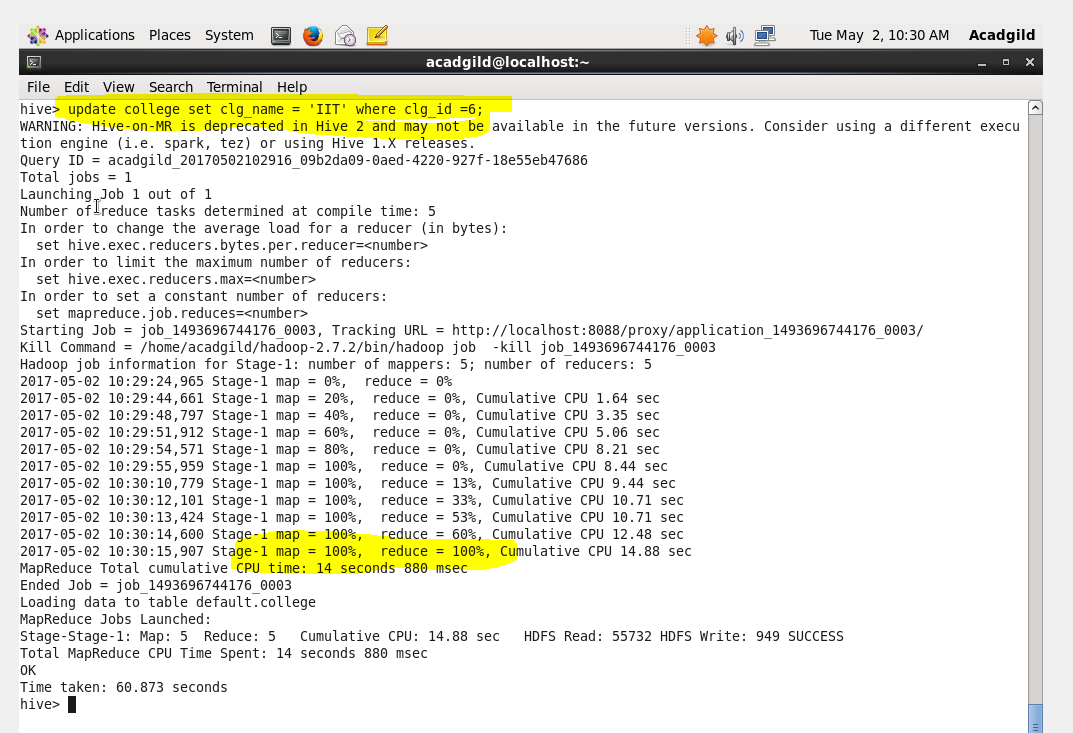
## Inserting Data into a Hive Table

****

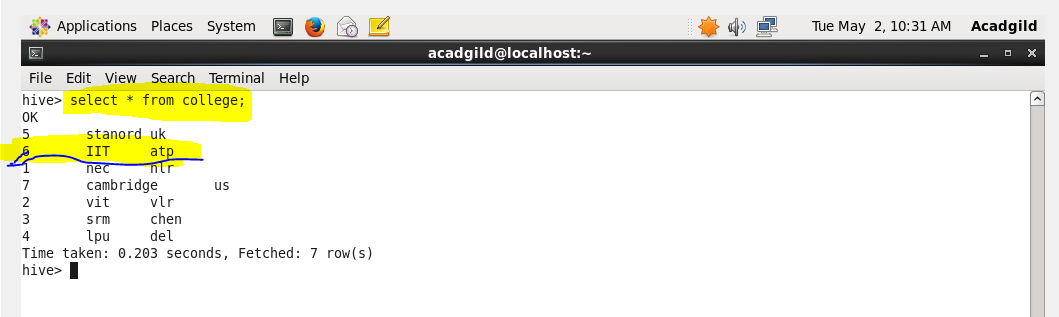
**Output**

****

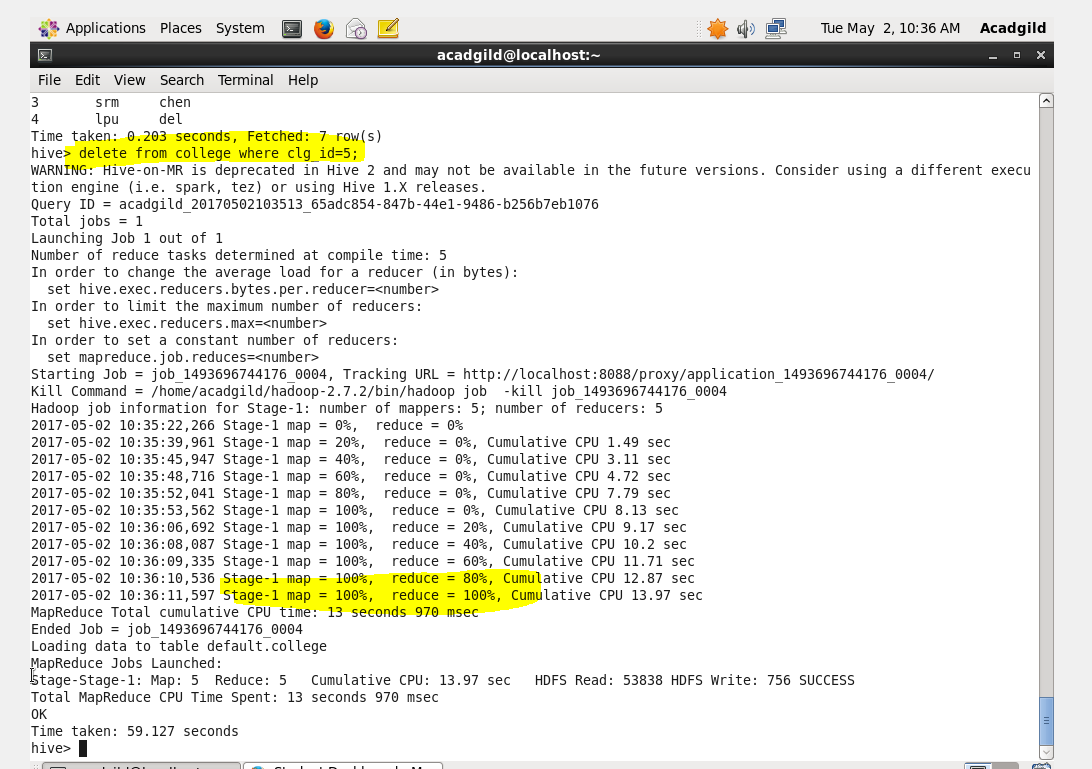
## Updating the Data in Hive Table

****

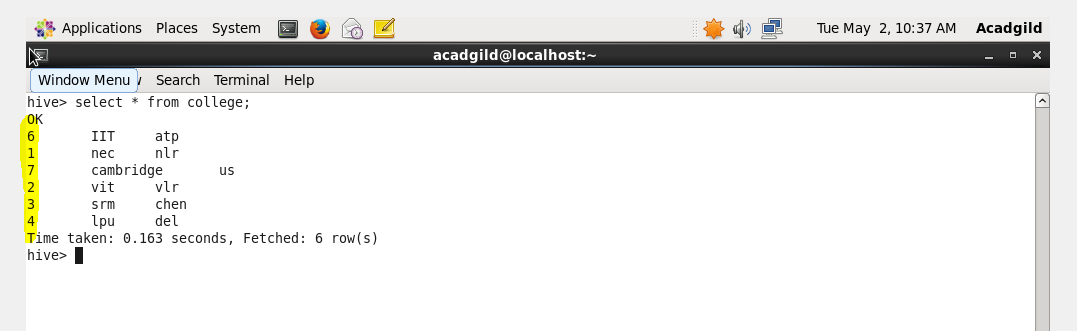
**Output**

****

## Deleting a Row from Hive Table

****

**Output**

****