**Assignment 28.1**

**Problem Statement:**

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

2. Refer our below blog and perform Hive Transaction Operations.

Share and explain the commands used with their results.

**Solution:**

* **ACID properties in Hive:**

ACID stands for:

* **A**tomicity
* **C**onsistency
* **I**solation and
* **D**urability

**Atomicity:**

Atomicity means, a transaction should complete successfully or else it should fail completely i.e. it should not be left partially.

**Consistency:**

Consistency ensures that any transaction will bring the database from one valid state to another state.

**Isolation:**

Isolation states that every transaction should be independent of each other i.e. one transaction should not affect another.

**Durability:**

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.

* **Hive Transaction Operations:**

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:

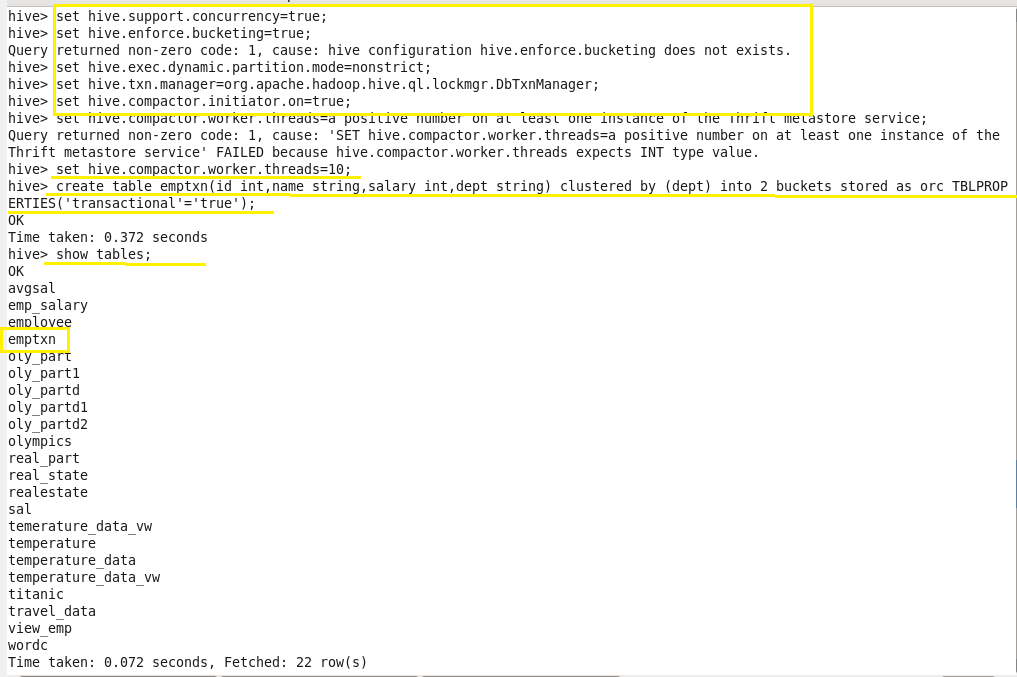
If these properties are not set properly, the ‘Insert’ operation will work but ‘Update’ and ‘Delete’ will not work and you will receive the following error:

|  |  |
| --- | --- |
| 1 | FAILED: SemanticException [Error 10294]: Attempt to do update or delete using transaction manage. |

* **Creating a Table That Supports Hive Transactions:**

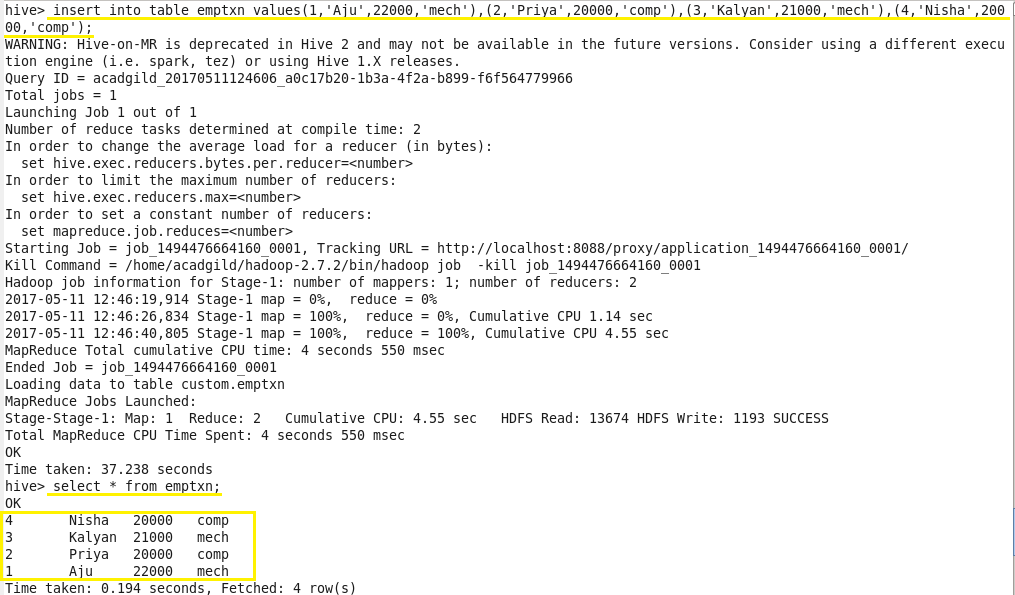
Here we create a table with name ‘emptxn’ and the columns present in the table are ‘id, name, salary,dept’.

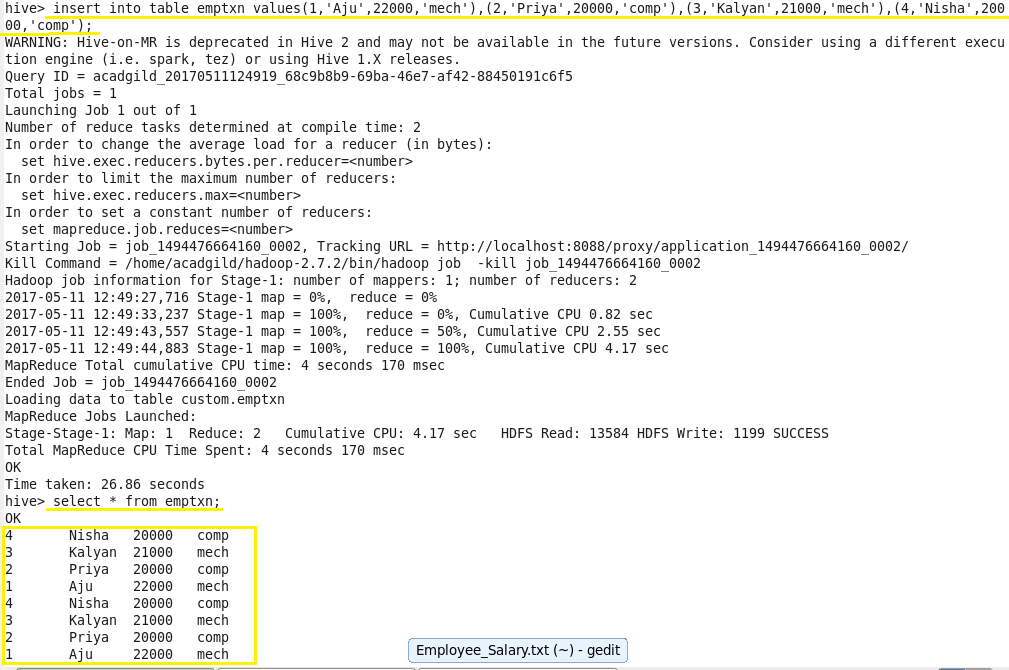
We are bucketing the table by ‘dept’ and the table format is ‘orc’, also we are enabling the transactions in the table by specifying it inside the TBLPROPERTIES as ‘transactional’=’true’.



* **Inserting values into table:**

Here we insert records into table row wise. Here, each row is separated by ‘( )’ brackets.





* **Updating Data in Hive Table:**

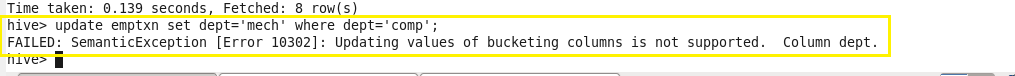
**Bucketed columns:**

The Update command is not supported on the columns that are bucketed.

In this table, we have bucketed the **‘dept’** column and performing the Update operation on the same column, so we have go the error:

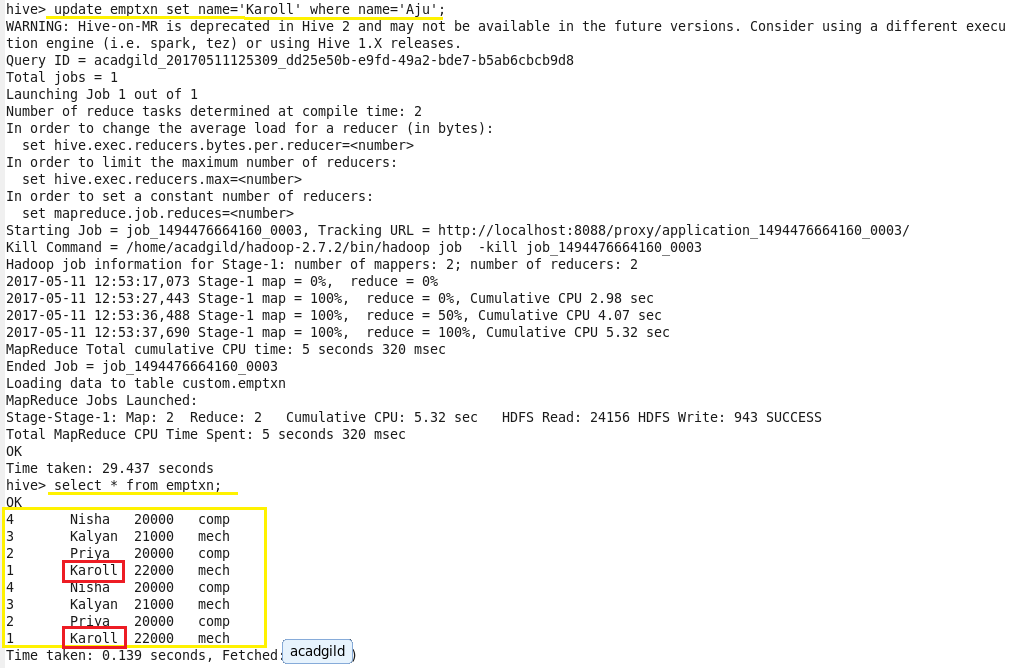
**FAILED: SemanticException[Error 10302]:**

**Updating values of bucketing columns is not supported. Column dept.**



**Non-bucketed columns:**

**When we perform the update operation on** **non-bucketed** **column the data is updated successfully.**



* **Deleting a Row from Hive Table:**

Here we delete the records which contains the name Nisha. All the records for Nisha will be deleted after executing this. We can check that using “select \* from emptxn” as below:

