Problem Statement:

How many kinds of tables are present in hive and explain the difference between them with a demo.

Managed and External tables are the two different types of tables in hive used to improve how data is loaded, managed and controlled. In this blog, we will be discussing the types of tables in Hive and the difference between them and how to create those tables and when to use those tables for a particular dataset.

**Tables in Hive**

As mentioned above, Hive has two types of tables:

* Managed table
* External table

Let us see about the above tables in detail

**Managed table**

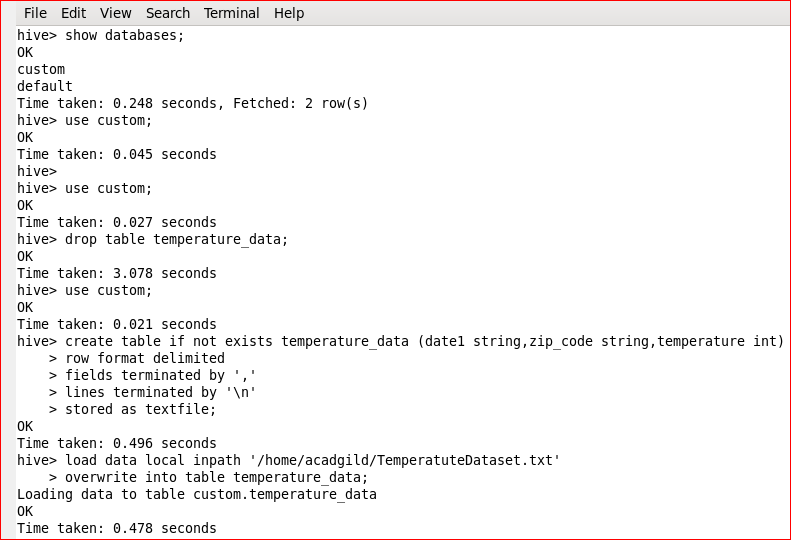
Managed table is also called as internal table. This is the default table in Hive. When we create a table in Hive without specifying it as external, by default we will get a Managed table.

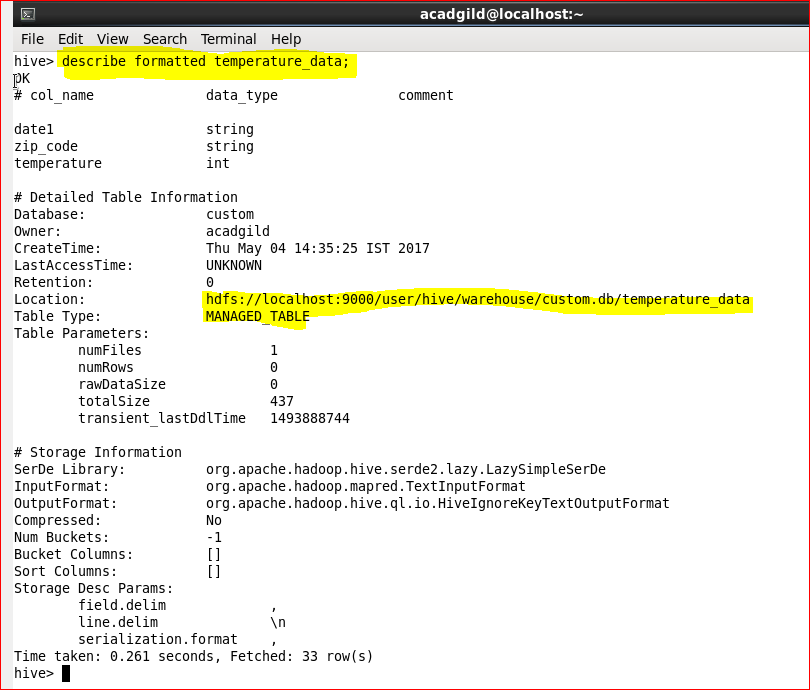
If we create a table as a managed table, the table will be created in a specific location in HDFS.

By default, the table data will be created in **/usr/hive/warehouse**directory of HDFS.

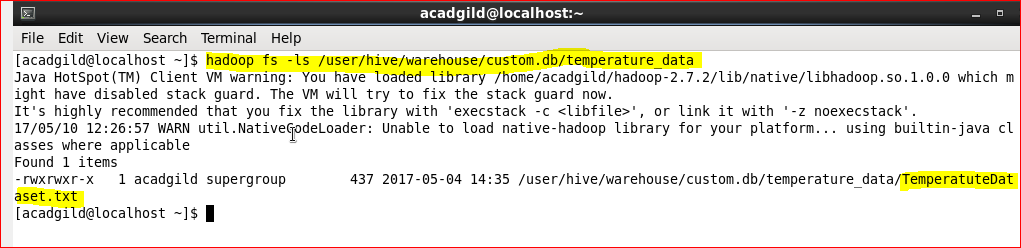
If we delete a Managed table, both the table data and meta data for that table will be deleted from the HDFS.

Let us create a managed table with the below command.

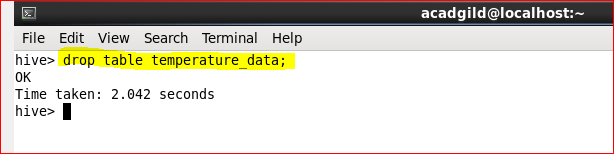


We have successfully created the table and to check the details of the table type the below command:

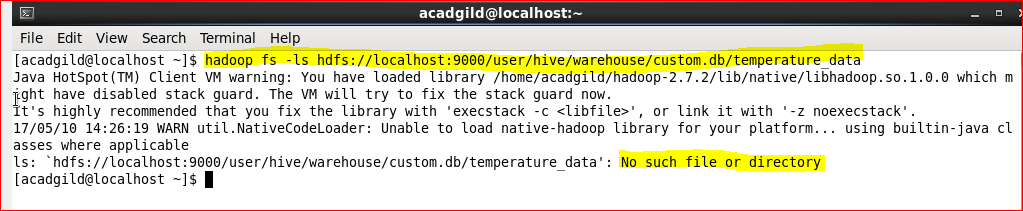
Check the contents of the table in HDFS by using the below command:



Now let us delete the above created table by using the command



Now let us try to check the contents of the table in HDFS using the below command:

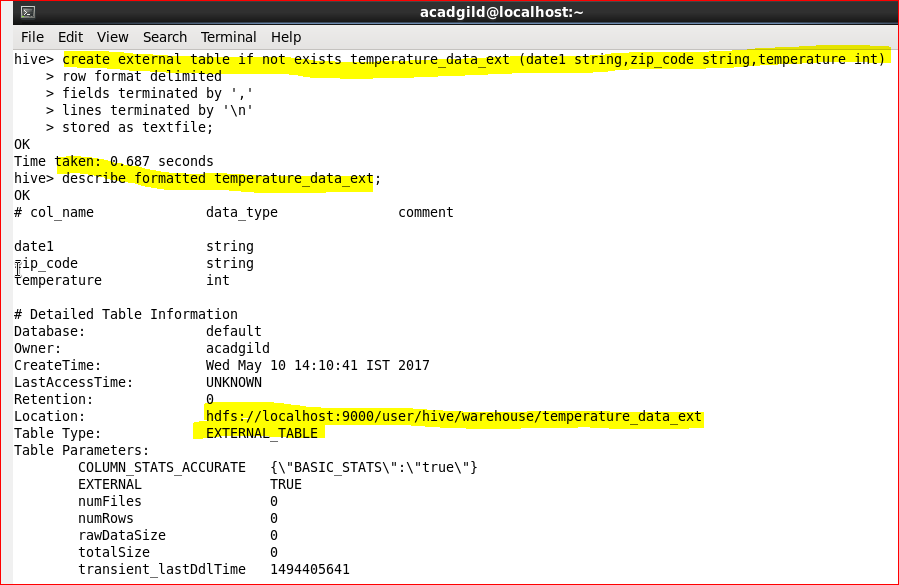


In the above image, you can see that it is displaying like **No such file or directory**because both the table and its contents are deleted from the HDFS location.

**EXTERNAL TABLE**

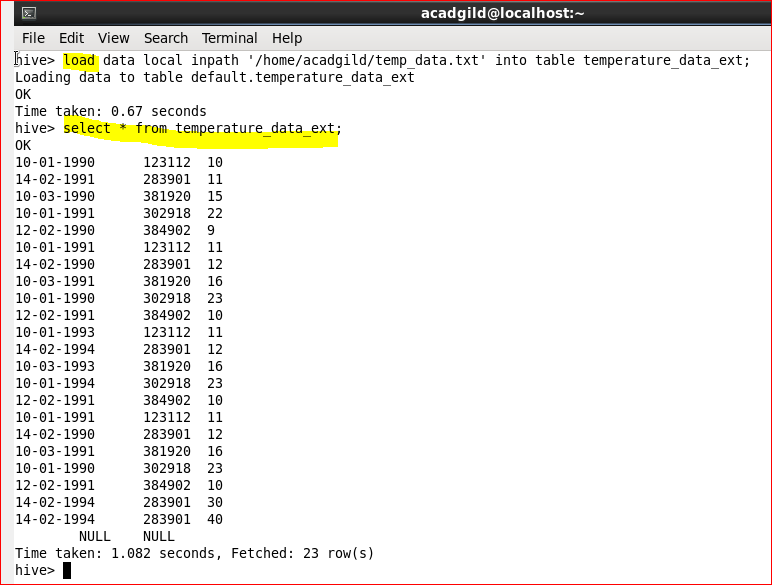
External table is created for external use as when the data is used outside Hive. Whenever we want to delete the table’s meta data and we want to keep the table’s data as it is, we use External table. External table only deletes the schema of the table.

Let us create an external table by using the below command:

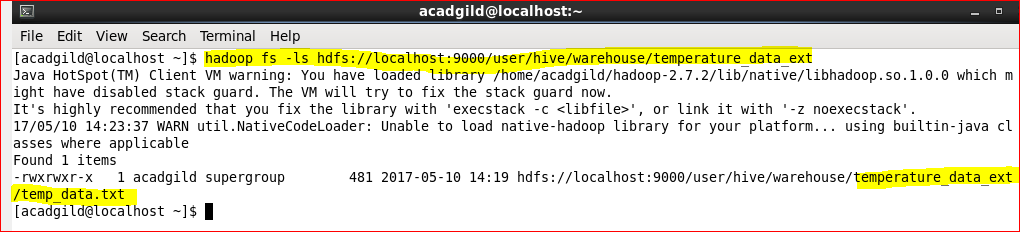


In the above image we can see the **EXTERNAL\_TABLE as the entry for the option Table type** which says that the above table is an External table.

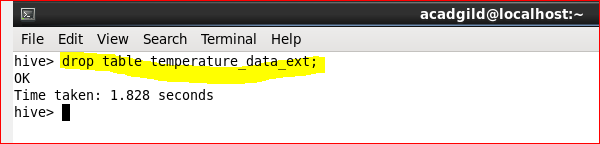
Now let us load some data into the table using the below command:



Let us check the contents in HDFS by using the below command:

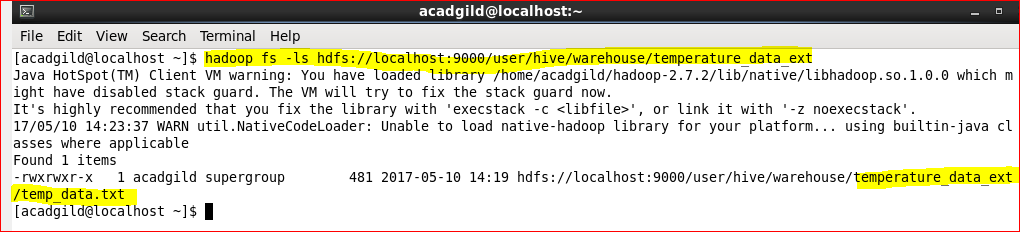


Let’s now delete the table using the below command:



We have successfully deleted the table.

Now let us check the HDFS location of the table using the below command:



You can see that the contents of the table are still present in the HDFS location.

If we create an External table, after deleting the table only the meta data related to table is deleted but not the contents of the table.

The above approach will work only if your data is in **/user/hive/warehouse**directory. But if your data is in another location, if you delete the table the data will also get deleted.

**When to use External and Managed table**

**Managed table**

* Data is temporary
* Hive to Manage the table data completely not allowing any external source to use the table
* Don’t want data after deletion

**External table**

* The data is also used outside of Hive. For example, the data files are read and processed by an existing program that doesn’t lock the files
* Hive should not own data and control settings, dirs, etc., you have another program or process that will do those things
* You are not creating table based on existing table (AS SELECT)
* Can create table back and with the same schema and point the location of the data