**Session 26**

**Assignment 5**

**Problem Statement:**

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

**Kinds of tables present in Hive:**

There are two types of tables in Hive,

* Managed table and
* External table.

1. **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 ‘/user/hive/warehouse’ directory of HDFS which can be changed by updating the location in the ‘configure file (config file).
* If we delete a Managed table, both the table data and metadata for that table will be deleted from master-node and HDFS respectively.

1. **External Table:**

* External table is created for external use as when the data is used outside Hive.
* External table stores files on the HDFS server but tables are not linked to the source file completely.
* Whenever we want to delete the table’s metadata 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.
* Meta data is maintained on master node and deleting an external table from HIVE, only deletes the metadata not the data/file.
* If you delete an external table the file still remains on the HDFS server.
* When you drop an external table, it only drops the Meta data. That means hive is ignorant of that data now. It does not touch the data itself.

## Hive may have internal or external tables .This is a choice that affects how data is loaded, controlled, and managed.

## When to use External and Managed table:

### **Managed table:**

* The data is temporary.
* You want Hive to completely manage the life-cycle of the table and data.

### **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.
* Data needs to remain in the underlying location even after a DROP TABLE. This can apply if you are pointing multiple schema (tables or views) at a single data set or if you are iterating through various possible schema.
* Hive should not own data and control settings, directories, etc., you may have another program or process that will do those things.
* You are not creating table based on existing table (AS SELECT).