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

There are two types of table

* External table
* Internal table

For External Tables ,Hive does not move the data into its warehouse directory. If the external table is dropped, then the table metadata is deleted but not the data.

For Internal tables , Hive moves data into its warehouse directory. If the table is dropped, then the table metadata and the data will be deleted.

**For External Tables -**

* External table stores files on the HDFS server but tables are not linked to the source file completely.
* If you delete an external table the file still remains on the HDFS server.

As an example if you create an **external table** called **“table\_test”** in HIVE using HIVE-QL and link the table to file **“file”**, ***then deleting “table\_test” from HIVE will not delete “file” from HDFS***.

* External table files are accessible to anyone who has access to HDFS file structure and therefore security needs to be managed at the HDFS file/folder level.
* **Meta data is maintained on master node and deleting an external table from HIVE, only deletes the metadata not the data/file.**

**For Internal Tables-**

* Stored in a directory based on settings in hive.metastore.warehouse.dir, ***by default internal tables are stored*** in the following directory ***“/user/hive/warehouse”*** you can change it by updating the location in the config file .
* Deleting the table deletes the metadata and data from master-node and HDFS respectively.
* Internal table file security is controlled solely via HIVE. Security needs to be managed within HIVE, probably at the schema level (depends on organization).

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

**Use EXTERNAL tables when:**

* 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).**

**Use INTERNAL tables when:**

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