Dropping Databases and Tables

As you saw in the “Creating Databases and Tables…” readings, you can remove (drop) both databases and tables using DROP DATABASE or DROP TABLE statements. As with the CREATE statements, you can conditionally drop a database by using IF EXISTS in the statement. This will avoid an error in the case that the database or table does not exist. The syntax is:

DROP DATABASE IF EXISTS *database_name*;

DROP TABLE IF EXISTS *table_name*;

Behavior with Managed or Unmanaged (External) Tables

When you drop a *managed* table (that is, one that you created *without* the EXTERNAL keyword), the table's storage directory will be deleted. That means *you will delete all the data* for that table! This is true whether or not the table’s storage directory is under the Hive warehouse directory. (The only exception to this rule is if Hive or Impala does not have the permission required to delete the files in the storage directory, for example if they are in an S3 bucket to which Hive and Impala have only read access.)

However, when you drop an *unmanaged* (also called *externally managed*) table, the data in the table’s storage directory will *not* be deleted. In this case Hive and Impala understand that this data is intended to be managed outside of their control, it will not delete the directory that holds the data. (This is true even if that directory is under the Hive warehouse directory.)

Always exercise caution when issuing a DROP TABLE statement, and be sure you understand what data, if any, will be lost.

Dropping a Database That Contains Tables

As a safety feature, Hive and Impala will throw an error if you attempt to drop a database that contains tables. This is to prevent unintended removal of data.

You can override this safety feature by using the CASCADE keyword in the DROP DATABASE statement. The syntax is:

DROP DATABASE *database_name* CASCADE;

Use this with *great* caution! Not only will it remove the database, it will remove all tables within it, *including deleting the data* for all managed tables within the database.