

Managed vs. External Tables

Hive fundamentally knows two different types of tables:

- Managed (Internal)
- External

Introduction

This document lists some of the differences between the two but the fundamental difference is that Hive assumes that it *owns* the data for managed tables. That means that the data, its properties and data layout will and can only be changed via Hive command. The data still lives in a normal file system and nothing is stopping you from changing it without telling Hive about it. If you do though it violates invariants and expectations of Hive and you might see undefined behavior.

Another consequence is that data is attached to the Hive entities. So, whenever you change an entity (e.g. drop a table) the data is also changed (in this case the data is deleted). This is very much like with traditional RDBMS where you would also not manage the data files on your own but use a SQL-based access to do so.

For external tables Hive assumes that it does *not* manage the data.

Managed or external tables can be identified using the [DESCRIBE FORMATTED table_name](#) command, which will display either MANAGED_TABLE or EXTERNAL_TABLE depending on table type.

[Statistics](#) can be managed on internal and external tables and partitions for query optimization.

Feature comparison

This means that there are lots of features which are only available for one of the two table types but not the other. This is an incomplete list of things:

- ARCHIVE/UNARCHIVE/TRUNCATE/MERGE/CONCATENATE only work for managed tables
- DROP deletes data for managed tables while it only deletes metadata for external ones
- ACID/Transactional only works for managed tables
- [Query Results Caching](#) only works for managed tables
- Only the RELY constraint is allowed on external tables
- Some Materialized View features only work on managed tables

Managed tables

A managed table is stored under the [hive.metastore.warehouse.dir](#) path property, by default in a folder path similar to `/user/hive/warehouse/databasename.db/tablename/`. The default location can be overridden by the `location` property during table creation. If a managed table or partition is dropped, the data and metadata associated with that table or partition are deleted. If the PURGE option is not specified, the data is moved to a trash folder for a defined duration.

Use managed tables when Hive should manage the lifecycle of the table, or when generating temporary tables.

External tables

An external table describes the metadata / schema on external files. External table files can be accessed and managed by processes outside of Hive. External tables can access data stored in sources such as Azure Storage Volumes (ASV) or remote HDFS locations. If the structure or partitioning of an external table is changed, an [MSCK REPAIR TABLE table_name](#) statement can be used to refresh metadata information.

Use external tables when files are already present or in remote locations, and the files should remain even if the table is dropped.