Relational Entities on Databricks

Learning Objectives

- Databases
- ▶ Tables
- ▶ The impact of the LOCATION keyword

Database

Databases = Schemas in Hive metastore

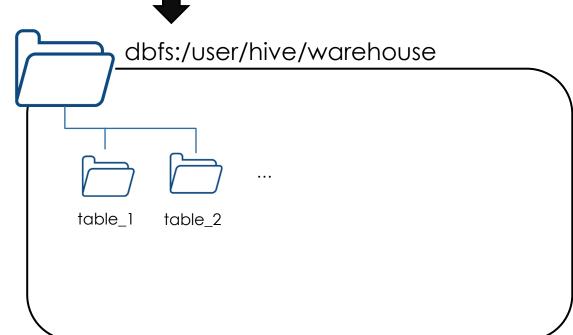
- CREATE DATABASE db_name
- CREATE SCHEMA db_name

Hive metastore

- repository of metadata
 - Databases
 - ▶ Tables
 - **...**

CREATE TABLE table 1; CREATE TABLE table2; • • • central Hive - default metastore -table_1 Workspace -table_2 - ... dbfs:/user/hive/warehouse

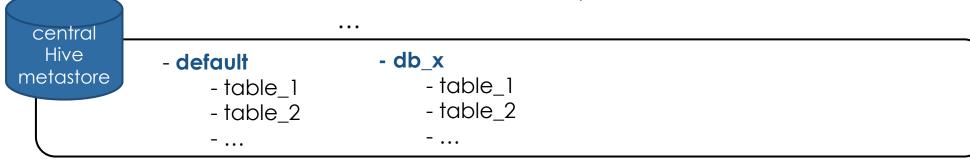
Storage



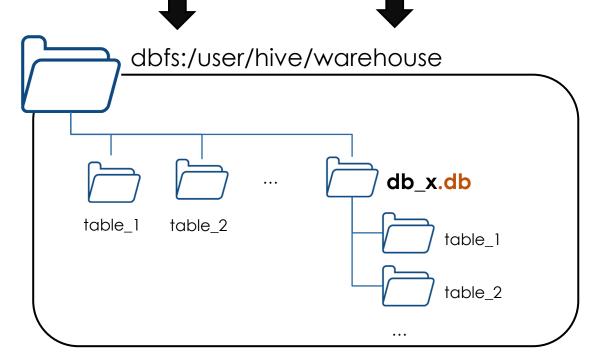
CREATE SCHEMA db_x

USE db_x; CREATE TABLE table1; CREATE TABLE table2;

Workspace



Storage



CREATE SCHEMA db_y
LOCATION 'dbfs:/custom/path/db_y.db'

USE db_y; CREATE TABLE table1; CREATE TABLE table2;

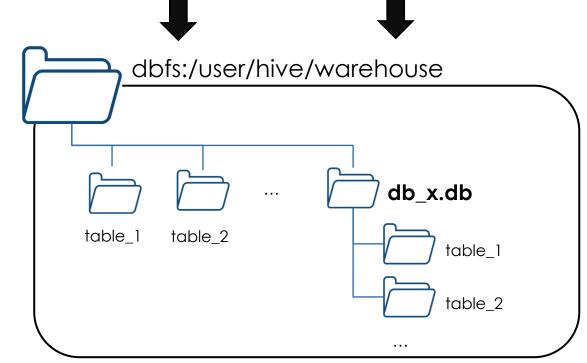
Workspace

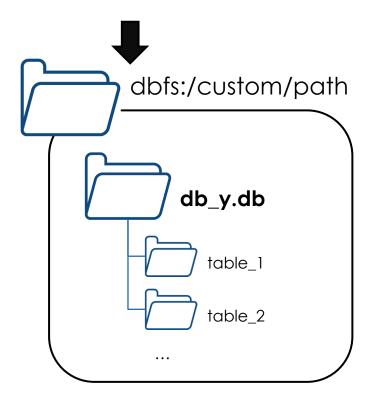
CREATE TABLE table2;
....

- default
- table_1
- table_2
- ...
- table_2
- ...
- central
Hive
metastore

- db_x
- table_1
- table_1
- table_2
- ...
- table_2
- ...
- ...

Storage





Tables

Manged tables

- Created under the database directory
 - CREATE TABLE table_name

Dropping the table, delete the underlying data files

External tables

- Created outside the database directory
 - CREATE TABLE table_name LOCATION 'path'
- Dropping the table, will Not delete the underlying data files

CREATE TABLE table3

LOCATION 'dbfs:/some/path_1/table3' central Hive metastore - default - db_x - db_y - table_1 - table_1 - table_1 - table_2 - table_2 -table 2 - table_3 dbfs:/some/path_1 dbfs:/custom/path dbfs:/user/hive/warehouse table_3 y.db $db_x.db$ table_1 table_2 table_1 table_1 table_2 table_2

USE db_x; CREATE TABLE table3 **LOCATION** 'dbfs:/some/path_2/x_table3'; central Hive metastore - default - db_x - db_y - table_1 - table_1 - table_1 - table_2 - table_2 -table 2 - table_3 - table_3 dbfs:/some/path_1 dbfs:/user/hive/warehouse dbfs:/custom/path table_3 y.db dbfs:/some/path_2 $db_x.db$ table_1 table_2 table_1 table_1 table 2 table_2 x_table_3

Derar Alhussein © Udemy | Databricks Certified Data Engineer Associate - Preparation

USE db_y; **CREATE TABLE table3 LOCATION** 'dbfs:/some/path_2/y_table3'; central Hive metastore - default - db_x - db_y - table_1 - table_1 - table_1 - table_2 - table_2 -table 2 - table_3 - table_3 - table_3 dbfs:/some/path_1 dbfs:/user/hive/warehouse dbfs:/custom/path table_3 y.db $db_x.db$ dbfs:/some/path_2 table_1 table_2 table_1 table_1 table 2 table_2 y_table_3 x table 3

Derar Alhussein © Udemy | Databricks Certified Data Engineer Associate - Preparation