

# 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 table1;
```

```
CREATE TABLE table2;
```

```
...
```



- **default**
- table\_1
- table\_2
- ...

**Workspace**



dbfs:/user/hive/warehouse



table\_1



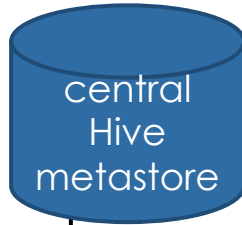
table\_2

...

**Storage**

```
CREATE SCHEMA db_x
```

```
USE db_x;  
CREATE TABLE table1;  
CREATE TABLE table2;  
...
```



- **default**

- table\_1
- table\_2
- ...

- **db\_x**

- table\_1
- table\_2
- ...

**Workspace**



dbfs:/user/hive/warehouse



table\_1



table\_2

...



**db\_x.db**



table\_1



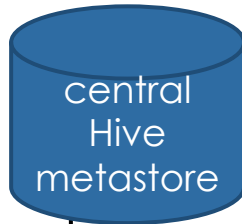
table\_2

...

**Storage**

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

```
USE db_y;  
CREATE TABLE table1;  
CREATE TABLE table2;  
...
```



**Workspace**

- **default**

- table\_1
- table\_2
- ...

- **db\_x**

- table\_1
- table\_2
- ...

- **db\_y**

- table\_1
- table\_2
- ...



dbfs:/user/hive/warehouse



table\_1



table\_2

...



**db\_x.db**



table\_1



table\_2

...



dbfs:/custom/path



**db\_y.db**



table\_1



table\_2

...

**Storage**

# Tables

## Manged tables

- ▶ Created under the database directory
  - ▶ CREATE TABLE table\_name
- ▶ Dropping the table, delete the underlying data files

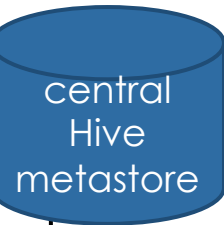
## External tables

- ▶ 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'



- **default**

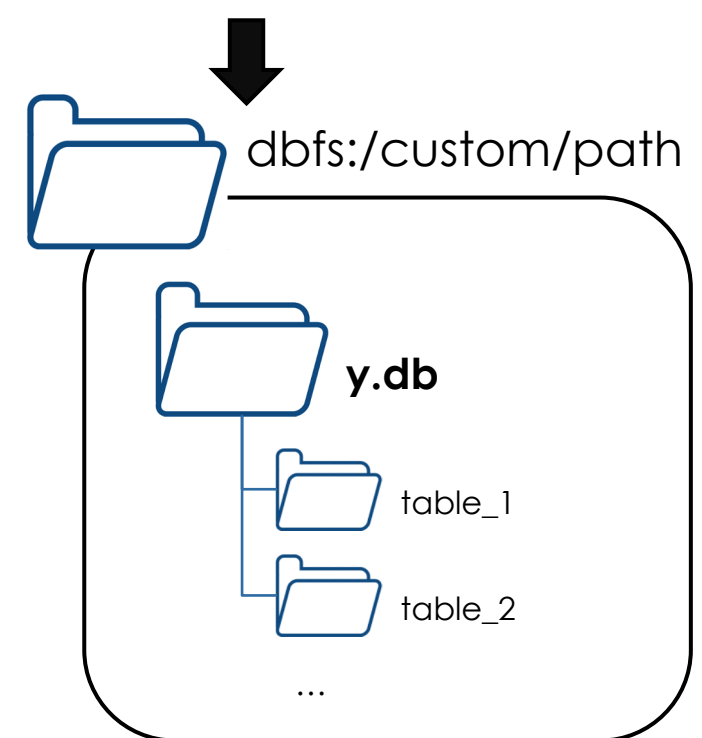
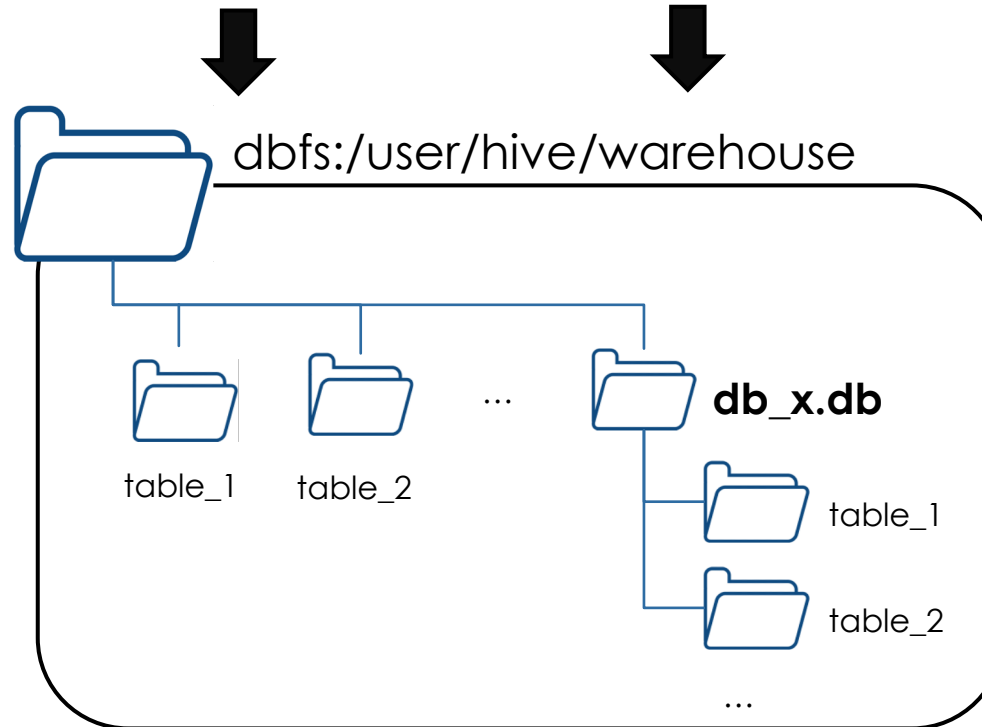
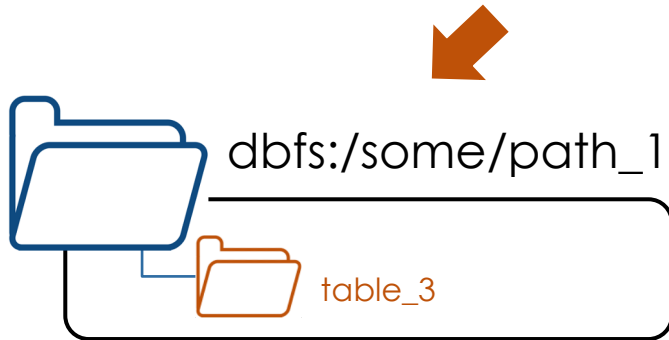
- table\_1
- table\_2
- **table\_3**

- **db\_x**

- table\_1
- table\_2

- **db\_y**

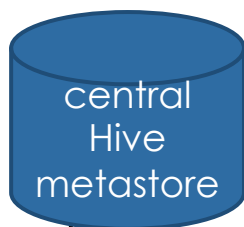
- table\_1
- table\_2



USE db\_x;

CREATE TABLE table3

**LOCATION** 'dbfs:/some/path\_2/x\_table3';



- **default**

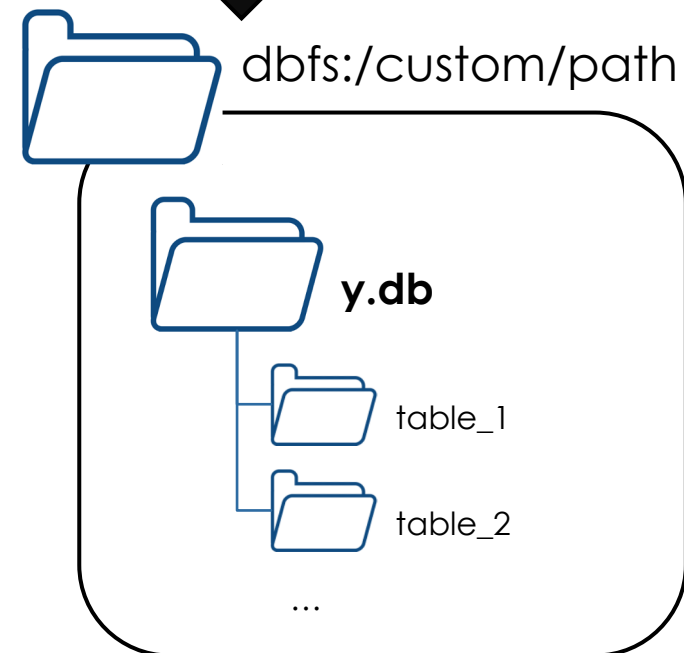
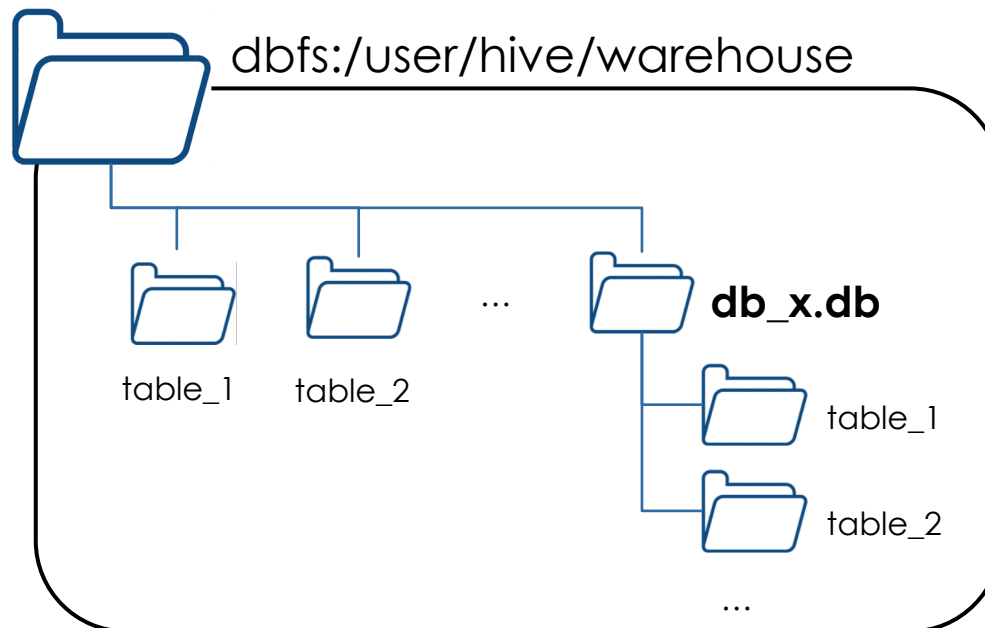
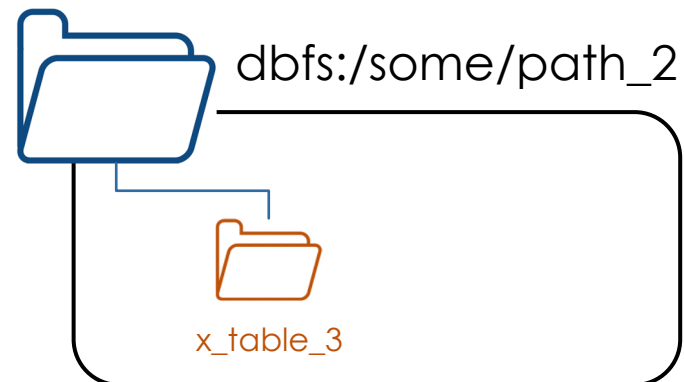
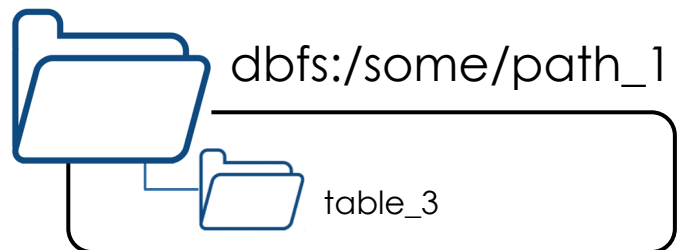
- table\_1
- table\_2
- **table\_3**

- **db\_x**

- table\_1
- table\_2
- **table\_3**

- **db\_y**

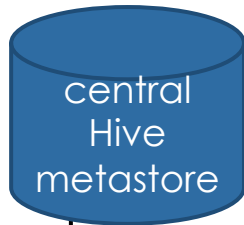
- table\_1
- table\_2



USE db\_y;

CREATE TABLE table3

**LOCATION** 'dbfs:/some/path\_2/y\_table3';



- **default**

- table\_1
- table\_2
- **table\_3**

- **db\_x**

- table\_1
- table\_2
- **table\_3**

- **db\_y**

- table\_1
- table\_2
- **table\_3**

