

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

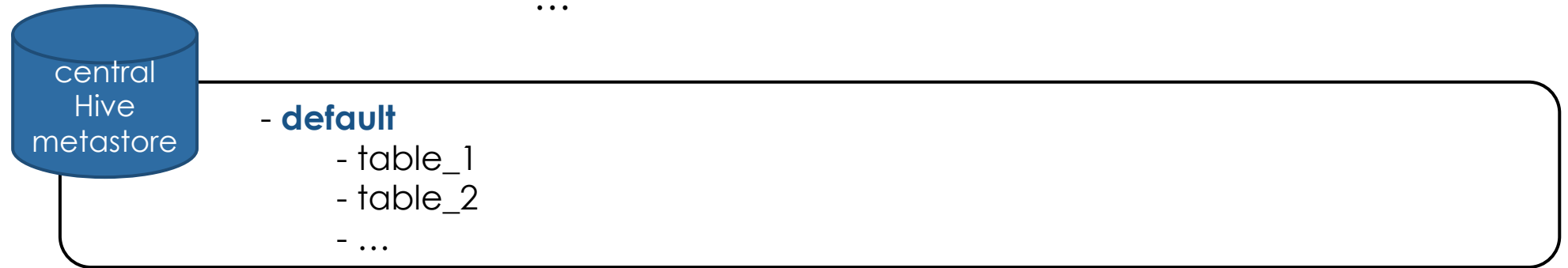
- ▶ repository of metadata
 - ▶ Databases
 - ▶ Tables
 - ▶ ...

```
CREATE TABLE table1;
```

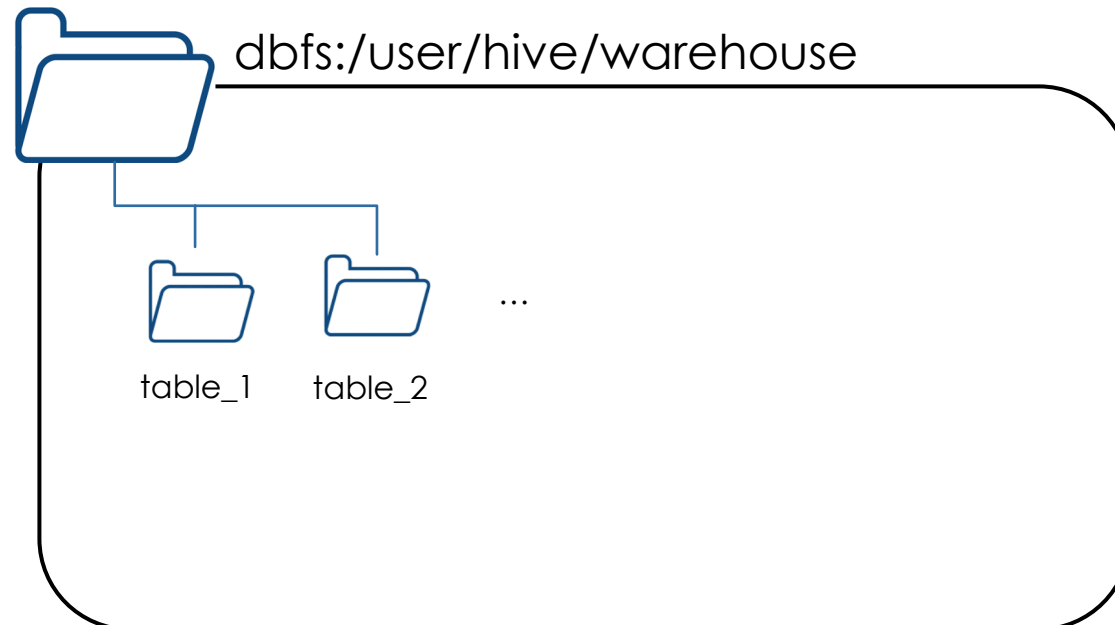
```
CREATE TABLE table2;
```

...

Workspace

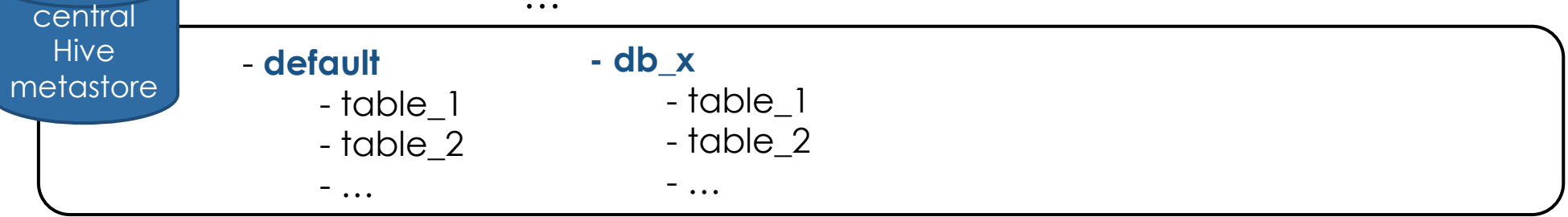
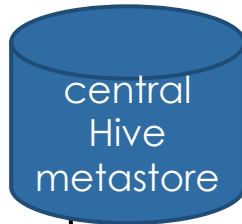


Storage



```
CREATE SCHEMA db_x
```

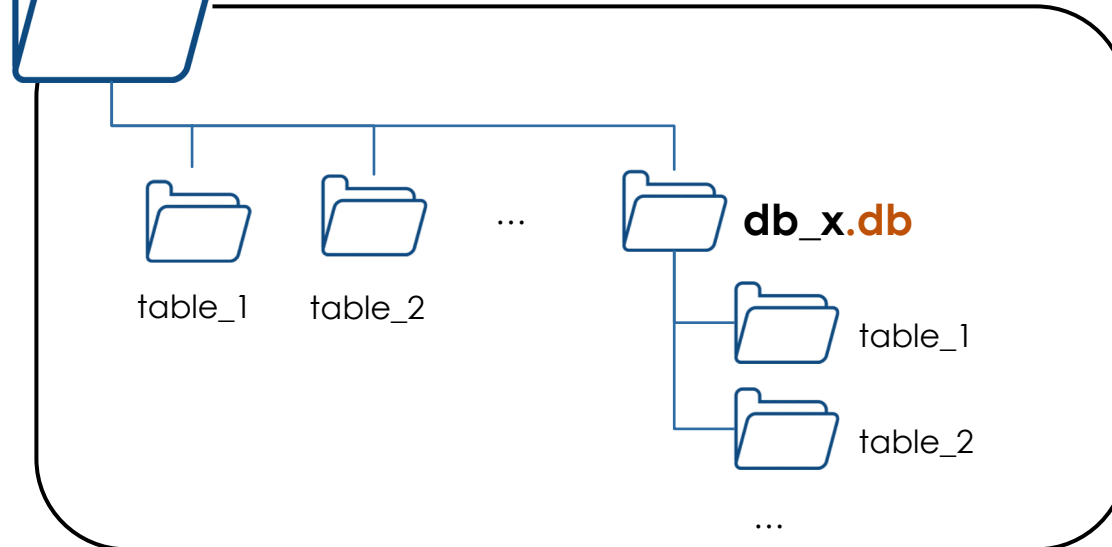
```
USE db_x;  
CREATE TABLE table1;  
CREATE TABLE table2;  
...
```



Workspace



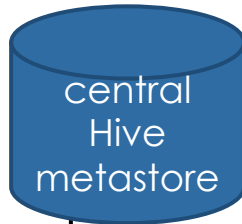
dbfs:/user/hive/warehouse



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

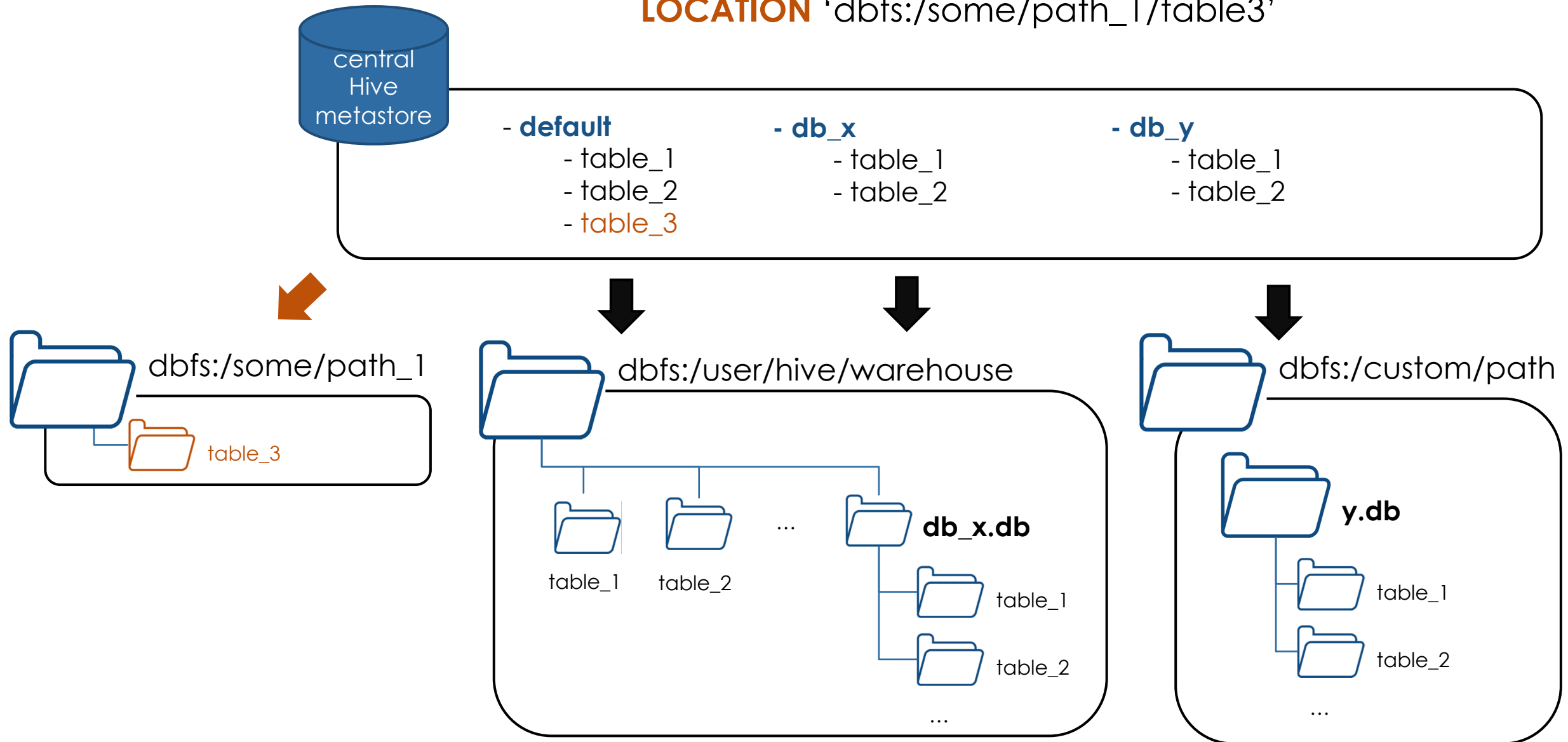
- ▶ Created under the database directory
 - ▶ CREATE TABLE table_name
- ▶ Dropping the table, delete the underlying data files

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

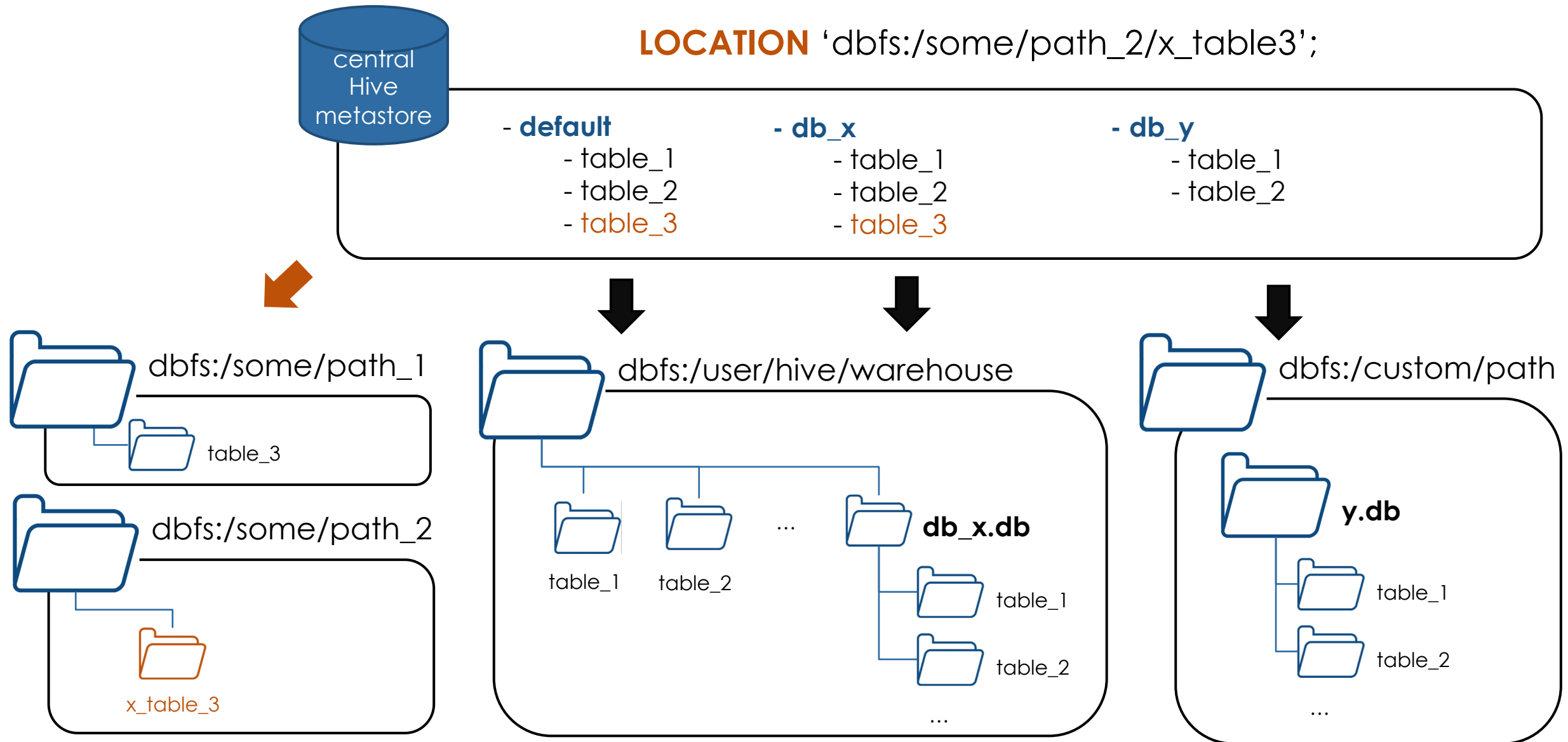
LOCATION 'dbfs:/some/path_1/table3'



USE db_x;

CREATE TABLE table3

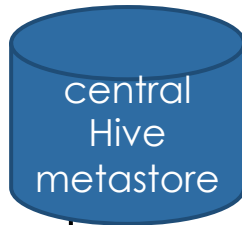
LOCATION 'dbfs:/some/path_2/x_table3';



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**

