

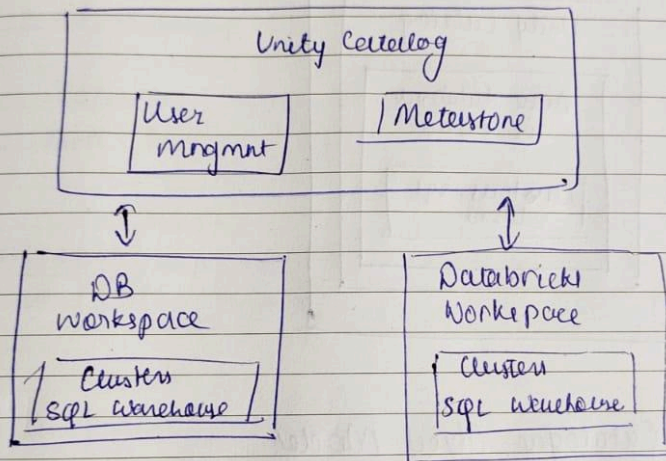
Azure Databricks Assignment - 5

Mitushi Vishwakarma

Azure Databricks Day-5.

Unity Catalog.

- provides centralized access control, auditing, lineage and data discovery capabilities



Key features :-

- Define once, secure everywhere. :
- Standards-compliant security model based on ANSI SQL, allows administrators to grant permissions in their existing data lake.

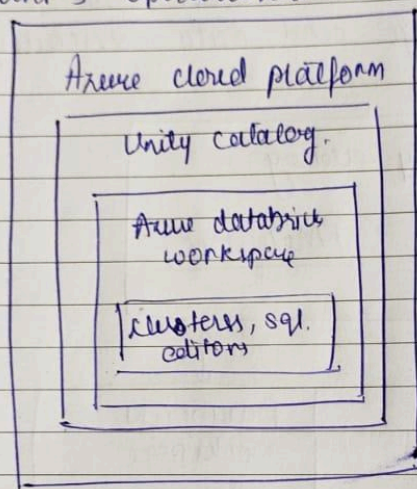
2023
SATURDAY
WK 46 • DAY 322-043
NOVEMBER

18-11-2023

18

NOVEMBER 2023
WK M T W T F S S
48 1 2 3 4 5
49 6 7 8 9 10 11
50 12 13 14 15 16 17
51 18 19 20 21 22 23
52 24 25 26 27 28 29 30

- built-in auditing and lineage → captures user level audit logs.
- data discovery
- System tables → easily access and query accounts operational data.



Unity Catalogue Object Model:

Metastore :- top level container for metadata

Each metastore exposes a three-level namespace (catalog, schema, table)

Catalog :- first layer of object hierarchy used to organize data assets.

NOVEMBER
WK 46 • DAY 321-044
FRIDAY
2023
17

DECEMBER 2023
S S F T W T M
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

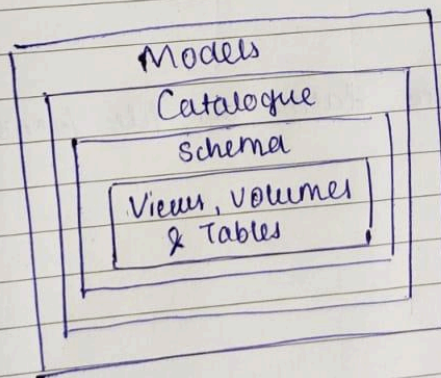
Schema :- databases / Schemas are second layer of object hierarchy; contains tables and views.

Tables, views and volumes :-

At the lowest level are tables, views and volumes.

Volumes provide governance for non-tabular data.

Models : registered models can be managed in Unity Catalog. (lowest level in hierarchy)



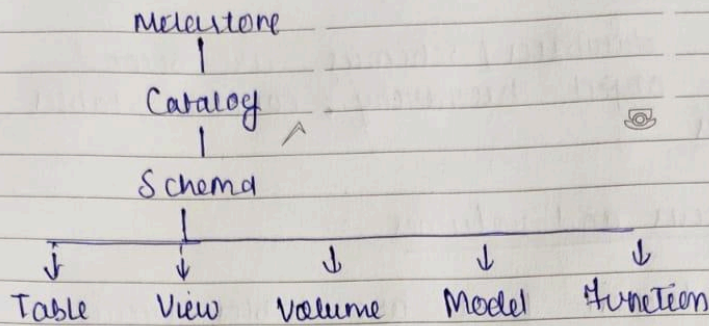
Hierarchical namespace

2023
THURSDAY
WK 46 • DAY 320-045
NOVEMBER

16-11-2023

16

Always Aim to be the best in what you do and give 100%



Azure databricks account admin

- ↳ create one metastore for each region
- ↳ in which they operate and assign it to DB workspace

Also, a metastore can be configured in ADLS Gen2 containers

externally managed data has file formats:

DELTA

CSV

JSON

AVRO

PARQUET

ORC

TEXT

15 - 11 - 2023

15

NOVEMBER

WK 46 • DAY 319-046

WEDNESDAY

2023

26 25 18 11 4
26 27 28 29 30 31
19 20 21 22 23 24
12 13 14 15 16 17
5 6 7 8 9 10
1 2 3

To create a catalog, use the `CREATE CATALOG` command with `spark.sql`. You must be a metastore admin or user with the `CREATE CATALOG` privilege on the metastore to create a catalog. If your workspace was enabled for Unity Catalog by default, then workspace admins have the `CREATE CATALOG` privilege by default.

The following commands show how to:

1. Create a catalog.

Create a catalog.

```
spark.sql("CREATE CATALOG IF NOT EXISTS quickstart_catalog")
```

2. Select a catalog.

Set the current catalog.

```
spark.sql("USE CATALOG quickstart_catalog")
```

3. Show all catalogs.

Show all catalogs in the metastore.

```
display(spark.sql("SHOW CATALOGS"))
```

4. Grant permissions on a catalog.

```
spark.sql("""
GRANT CREATE, USE CATALOG
ON CATALOG quickstart_catalog
TO `account users`""")
```

5. Show all grants on a catalog.

```
display(spark.sql("SHOW GRANT ON CATALOG quickstart_catalog"))
```

Scheduling a job :

The screenshot shows the Databricks workspace with a Python notebook open. A 'Scheduling' dialog box is displayed over the notebook code. The dialog box contains the following fields:

- Job name:** Scheduling
- Schedule:** Manual (selected), Scheduled (radio button)
- Every:** Day, at 23:49, Asia/Calcutta
- Cluster:** Add new job cluster, 126 GB - 36 Cores - DBR 13.3 LTS - Photon - Spark 3.4.1 - Scala 2.12
- Parameters:** + Add
- Alerts:** mitushi.vishjs@gmail.com, Start, Success, Failure (checked), X

The dialog box has 'Cancel' and 'Create' buttons. Below the dialog box, the notebook code is visible, showing a PySpark script that creates a DataFrame with sample data and prints it.

Viewing Scheduled job in workflow :

The screenshot shows the Databricks 'Workflows' page. The 'Jobs' tab is selected. A table lists the scheduled jobs:

Name	Tags	Created by	Trigger	Last run
Scheduling		Mitushi Vishwakarma	Scheduled	

The table has a 'Filter jobs' search bar and a 'Only jobs owned by me' checkbox. A 'Create job' button is in the top right corner. The bottom right corner shows '1-1 of 1 items' and a page number '1'.