Scenario -1 (Volumes)

Volumes

Applies to: Databricks SQL Databricks Runtime 13.3 LTS and above Unity Catalog only

Volumes are Unity Catalog objects representing a logical volume of storage in a cloud object storage location. Volumes provide capabilities for accessing, storing, governing, and organizing files. While tables provide governance over tabular datasets, volumes add governance over non-tabular datasets. You can use volumes to store and access files in any format, including structured, semi-structured, and unstructured data.

Volumes are siblings to tables, views, and other objects organized under a schema in Unity Catalog.

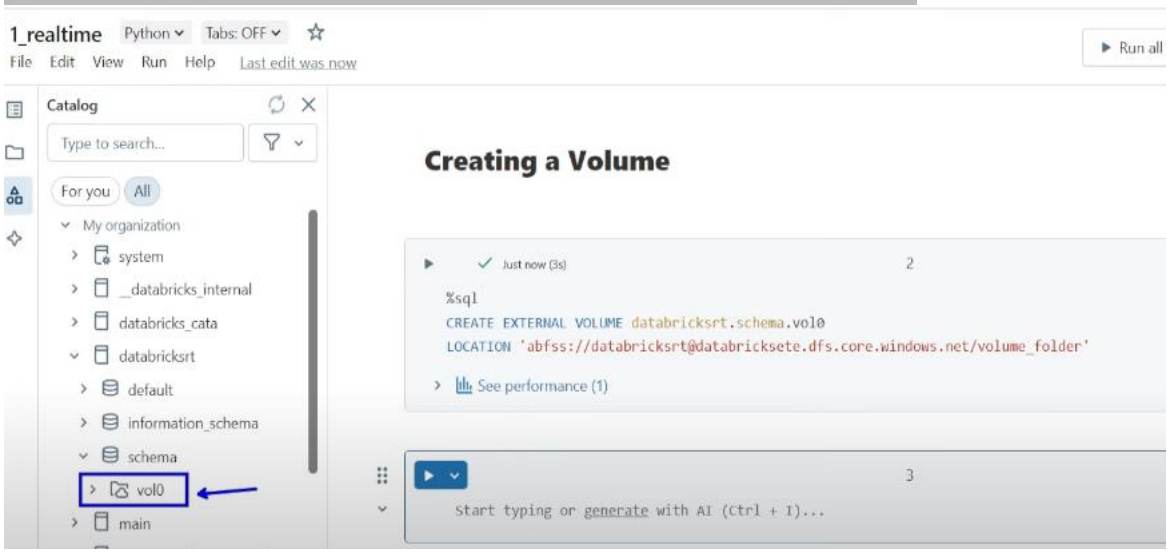
A volume can be managed or external.

Volume naming and reference

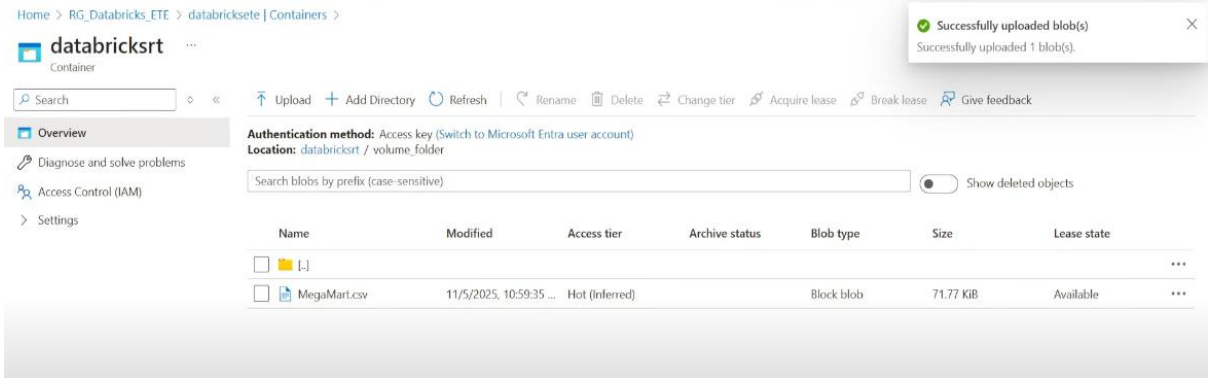
A volume name is an identifier that can be qualified with a catalog and schema name in SQL commands.

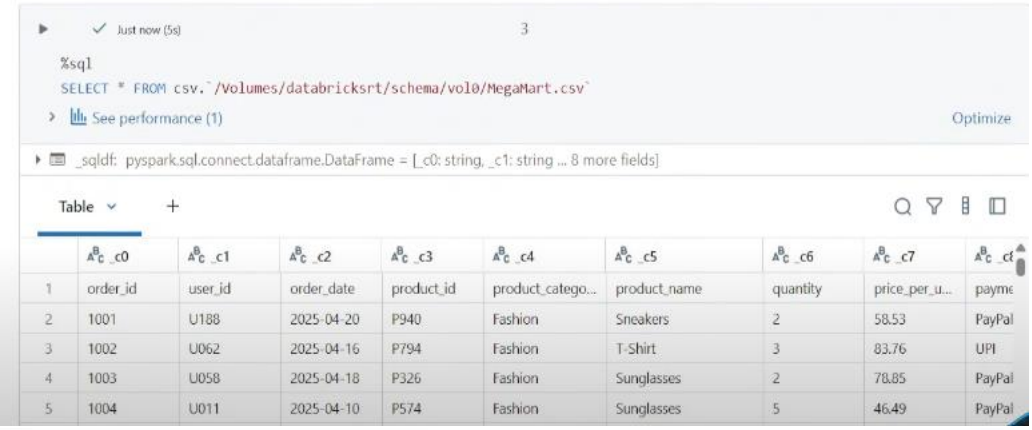
The path to access files in volumes uses the following format:

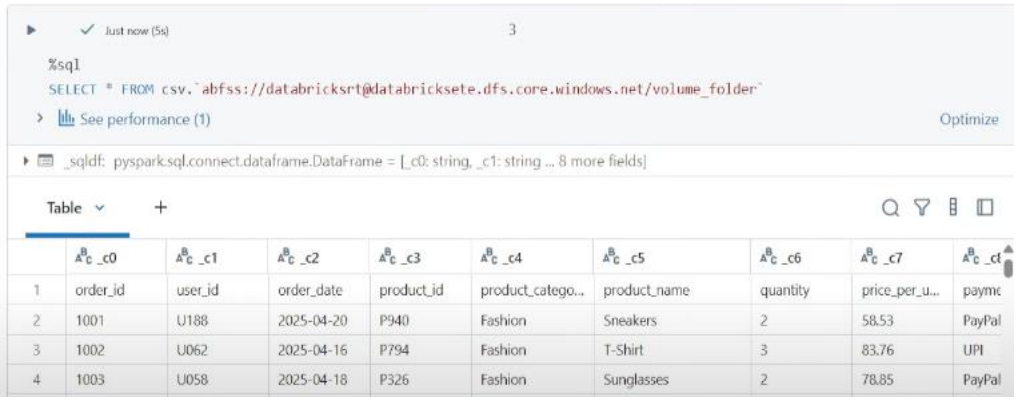
```
/Volumes/<catalog_identifier>/<schema_identifier>/<volume_identifier>/<path>/<file_name>
```



Upload the file in ADLS Gen2



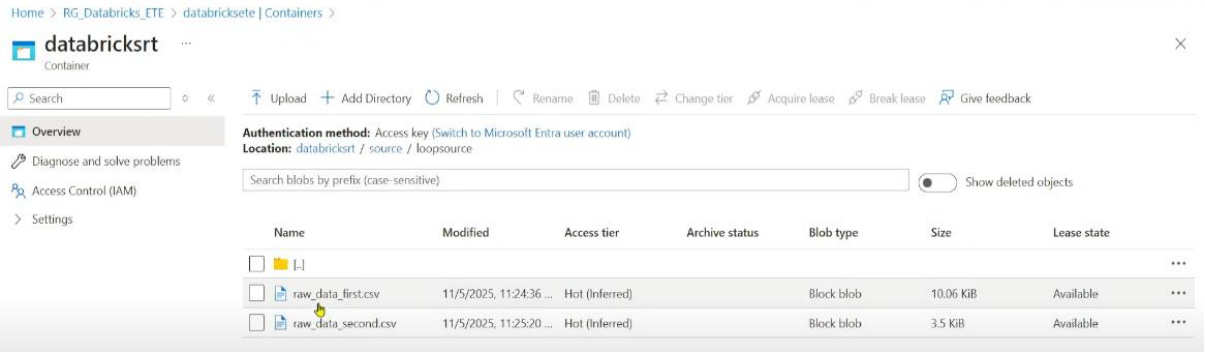


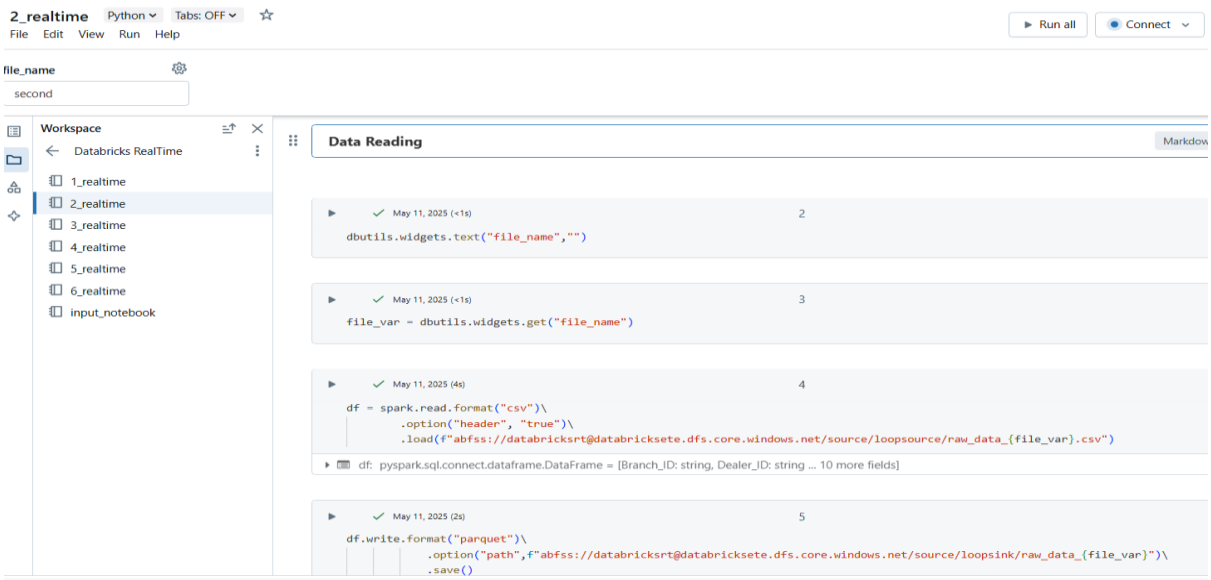


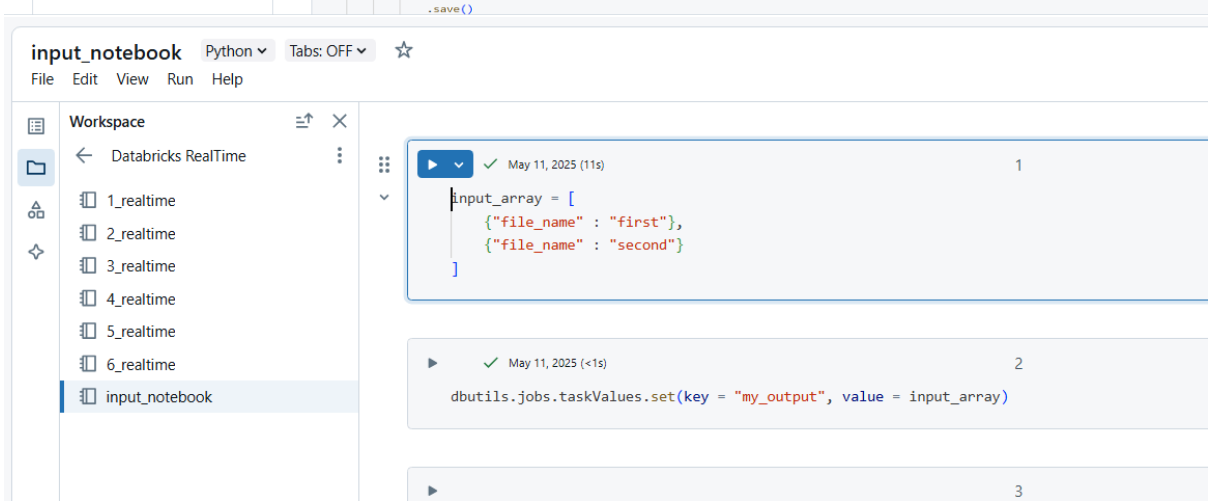
Both are same, but why do we need Volume → Volumes are governed, can be tracked using UC

Scenario-2

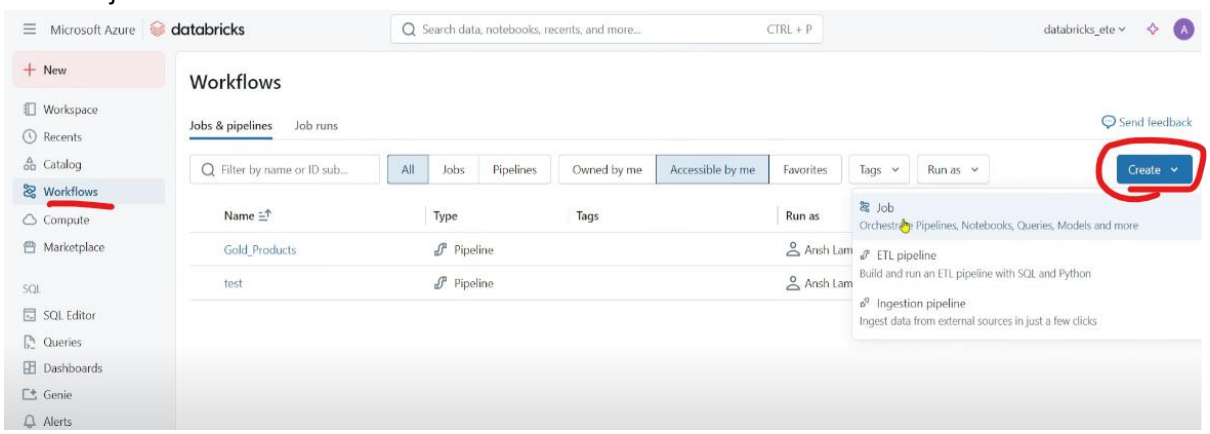
Looping through files in folder







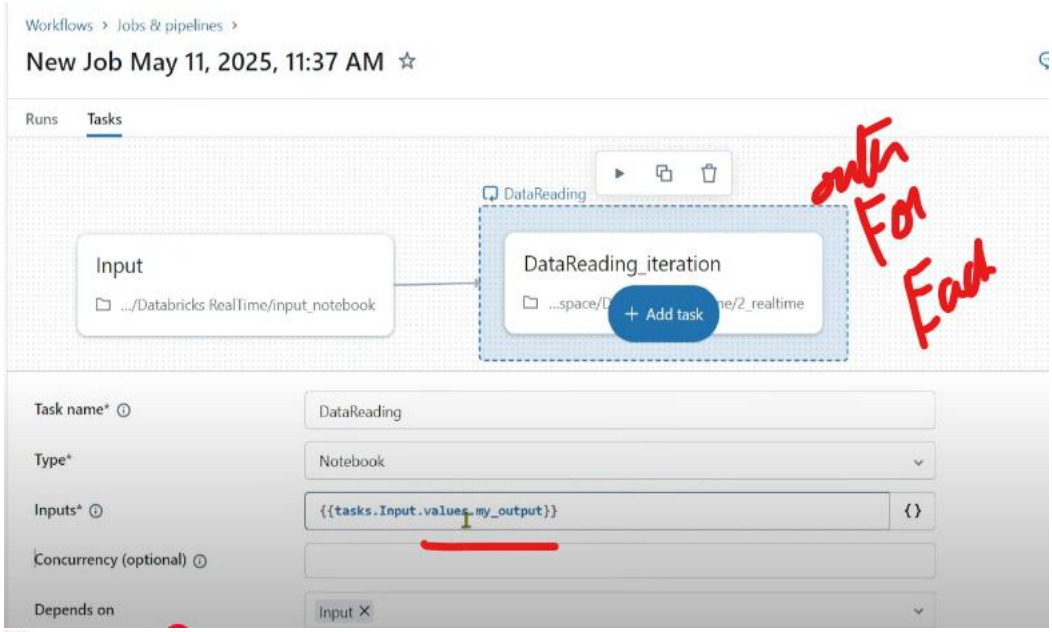
Create a job in Workflows

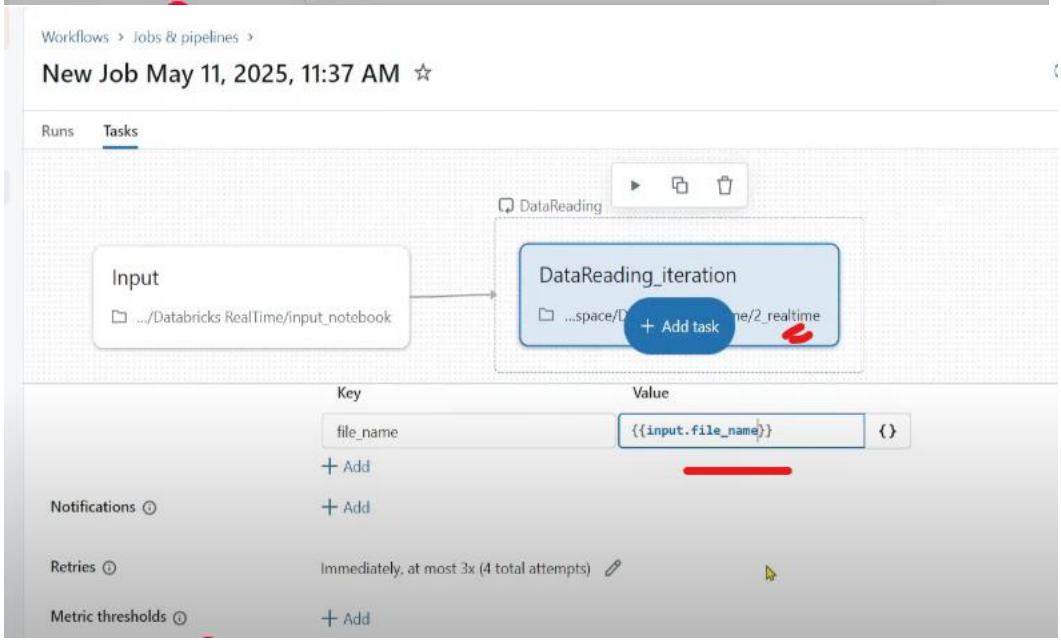


Job : Normal Workflows

ETL pipeline : DLT

Ingestion pipeline : No code ingestion pipeline


Workflows › Jobs & pipelines ›
New Job May 11, 2025, 11:37 AM
Runs
Tasks
DataReading
Input
.../Databricks RealTime/input_notebook
DataReading_iteration
+ Add task
Task name*
DataReading
Type*
Notebook
Inputs*
{{tasks.Input.values.my_output}}
Concurrency (optional)
Depends on
Input


Workflows › Jobs & pipelines ›
New Job May 11, 2025, 11:37 AM
Runs
Tasks
DataReading
Input
.../Databricks RealTime/input_notebook
DataReading_iteration
+ Add task
Key
Value
file_name
{{input.file_name}}
+ Add
Notifications
+ Add
Retries
Immediately, at most 3x (4 total attempts)
Metric thresholds
+ Add

Scenario – 3

Bring on prem data to our organisation without moving the data, query data sitting outside of Databricks/managed layer/cloud without moving the data

Concept called Lakehouse Federation

What is Lakehouse Federation?

Lakehouse Federation is the query federation platform for Azure Databricks. The term *query federation* describes a collection of features that enable users and systems to run queries against multiple data sources without needing to migrate all data to a unified system.

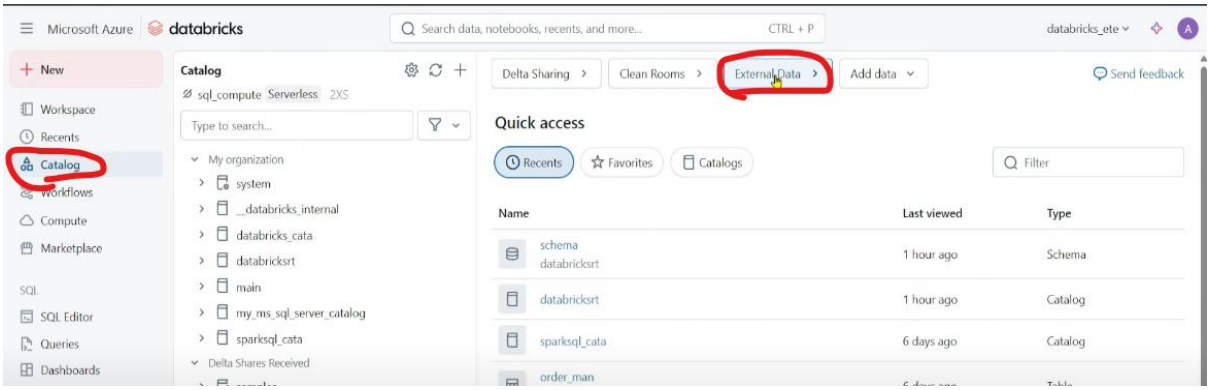
Azure Databricks uses Unity Catalog to manage query federation. You configure read-only connections to popular database solutions using drivers that are included on pro SQL warehouses, serverless SQL warehouses, and Databricks Runtime clusters. Unity Catalog's data governance and data lineage tools ensure that data access is managed and audited for all federated queries made by the users in your Azure Databricks workspaces.

Why use Lakehouse Federation?

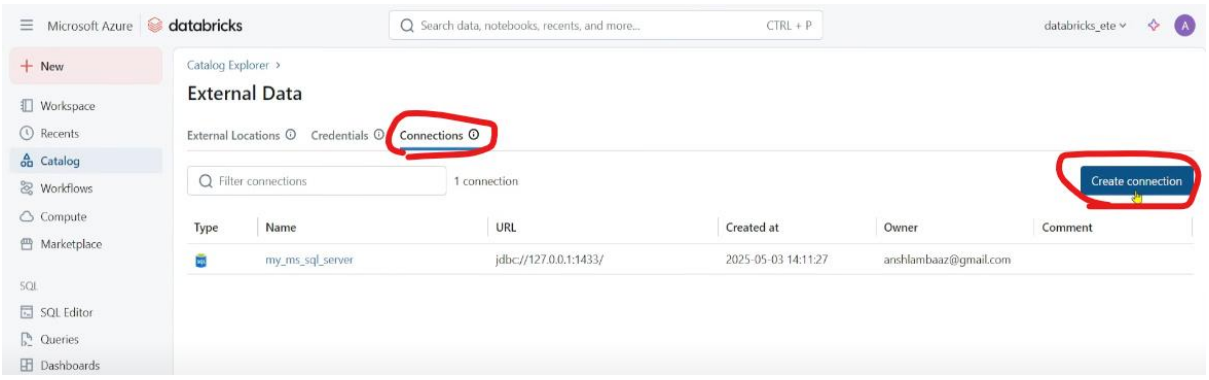
The lakehouse emphasizes central storage of data to reduce data redundancy and isolation. Your organization might have numerous data systems in production, and you might want to query data in connected systems for a number of reasons:

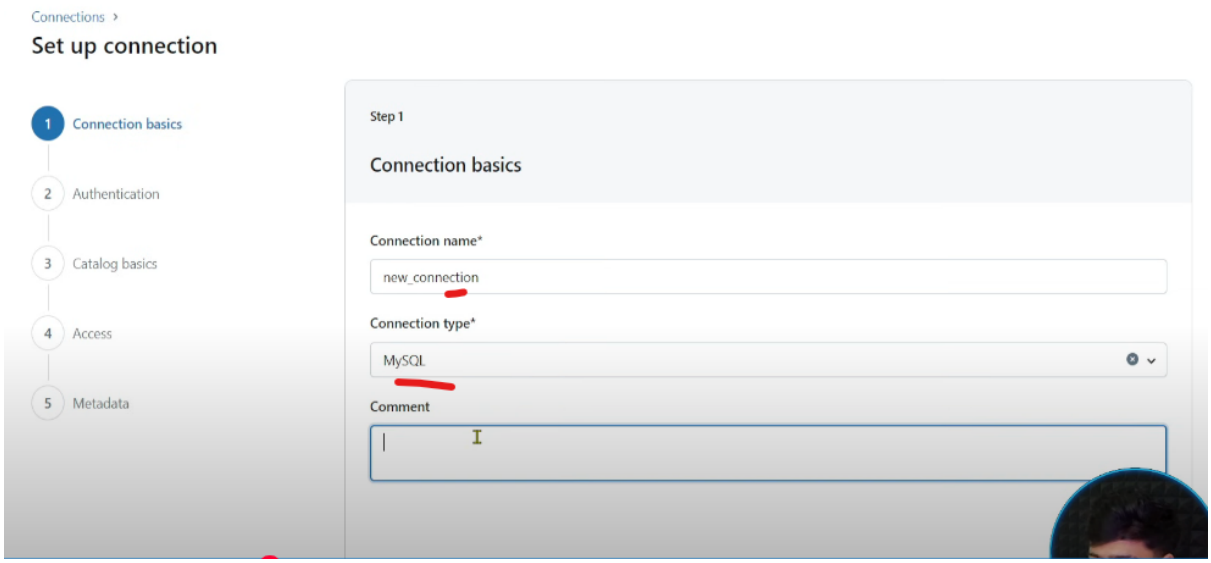
- Ad hoc reporting.
- Proof-of-concept work.
- The exploratory phase of new ETL pipelines or reports.
- Supporting workloads during incremental migration.

Catalog → External Data

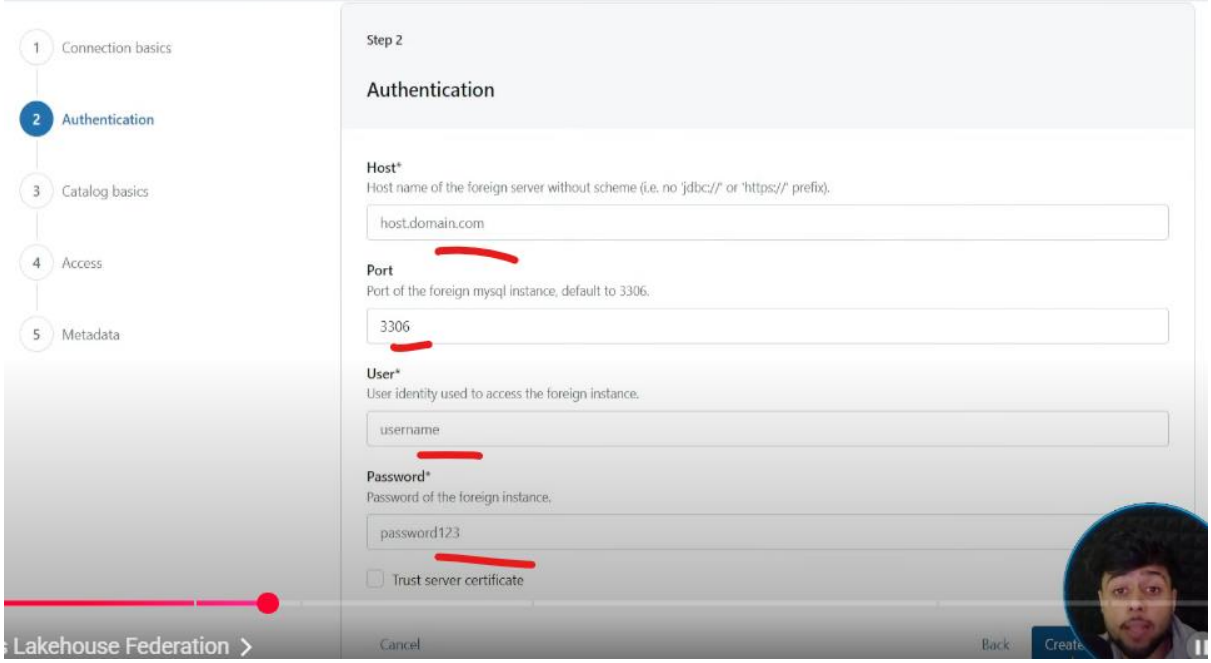


Connections → Create Connection





There are crazy amount of sources in Connection Type



Provide all these details → Create a new catalog→ access → metadata

Post that you can query the data sitting in onprem from DB workspace

Scenario – 4 (AutoLoader)

Autoloader - Incremental Loading Markdown

Streaming Query

May 11, 2025 (1s) 3

```
df = spark.readStream.format("cloudFiles")\
        .option("cloudFiles.format","csv")\
        .option("cloudFiles.schemaLocation","abfss://databricksrt@databricksete.dfs.core.windows.net/source/autosink/checkpoint")\
        .load("abfss://databricksrt@databricksete.dfs.core.windows.net/source/loopsource")
```

df: pyspark.sql.connect.dataframe.DataFrame = [Branch_ID: string, Dealer_ID: string ... 11 more fields]

May 11, 2025 4

```
df.writeStream.format("parquet")\
        .outputMode("append")\
        .option("checkpointLocation","abfss://databricksrt@databricksete.dfs.core.windows.net/source/autosink/checkpoint")\
        .trigger(once=True)\
        .option("path","abfss://databricksrt@databricksete.dfs.core.windows.net/source/autosink/data")\
        .start()
```

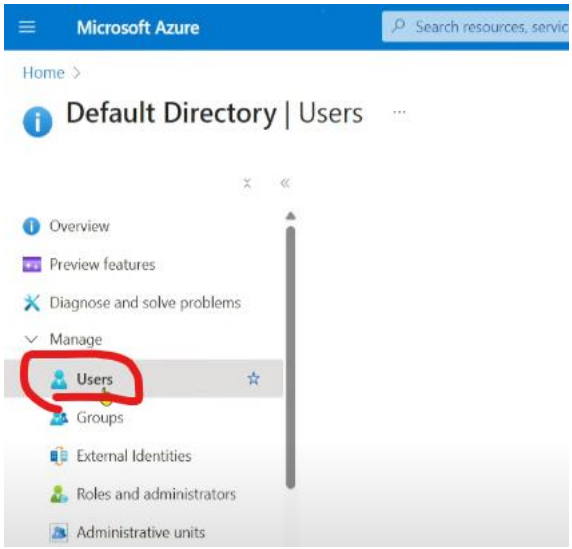
a714b212-424e-4392-af55-c63231e7c85a *Last updated: 10 days ago*

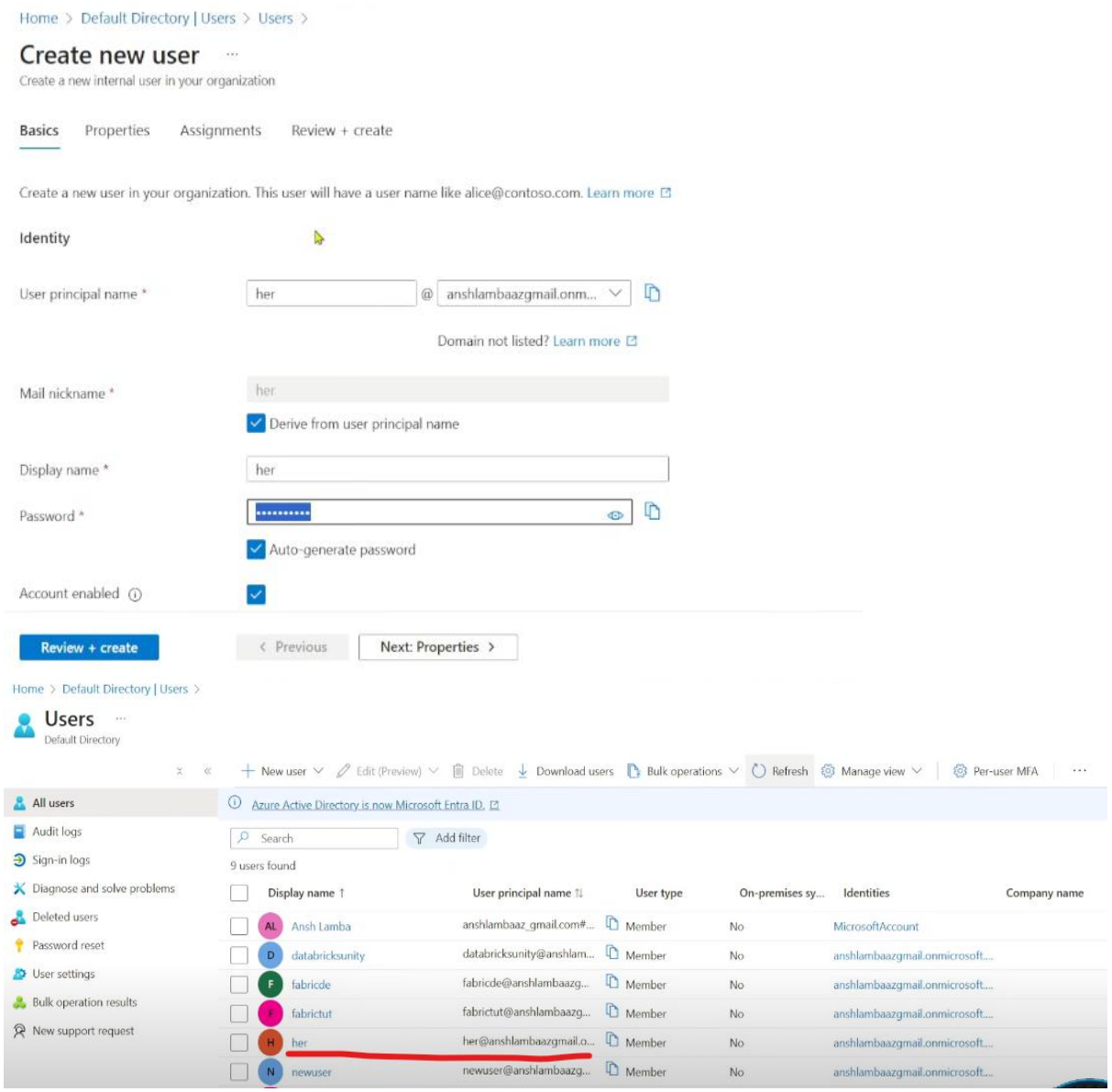
<pyspark.sql.connect.streaming.query.StreamingQuery at 0x7f674ec08850>

Scenario – 5 (UC access control)

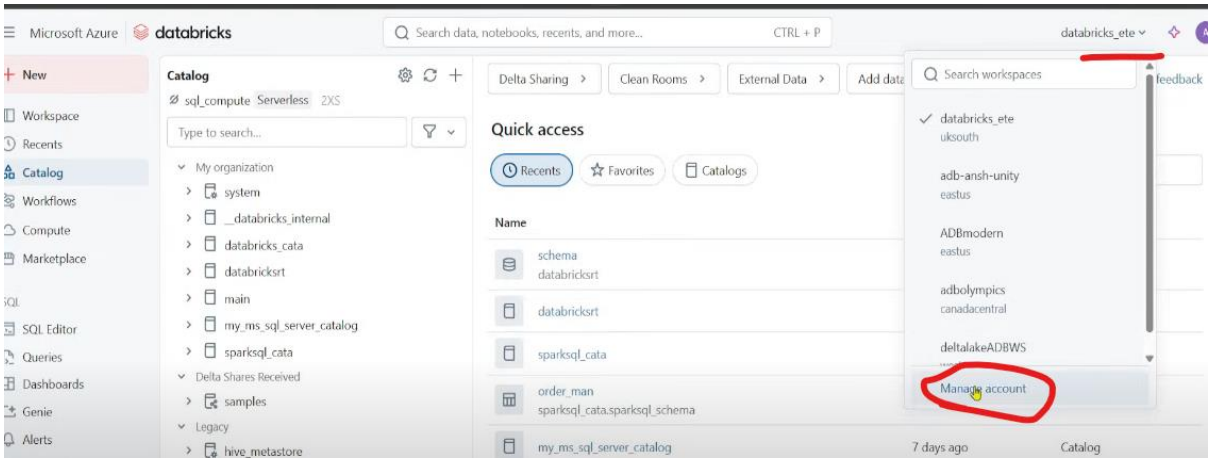
Granting access to new user

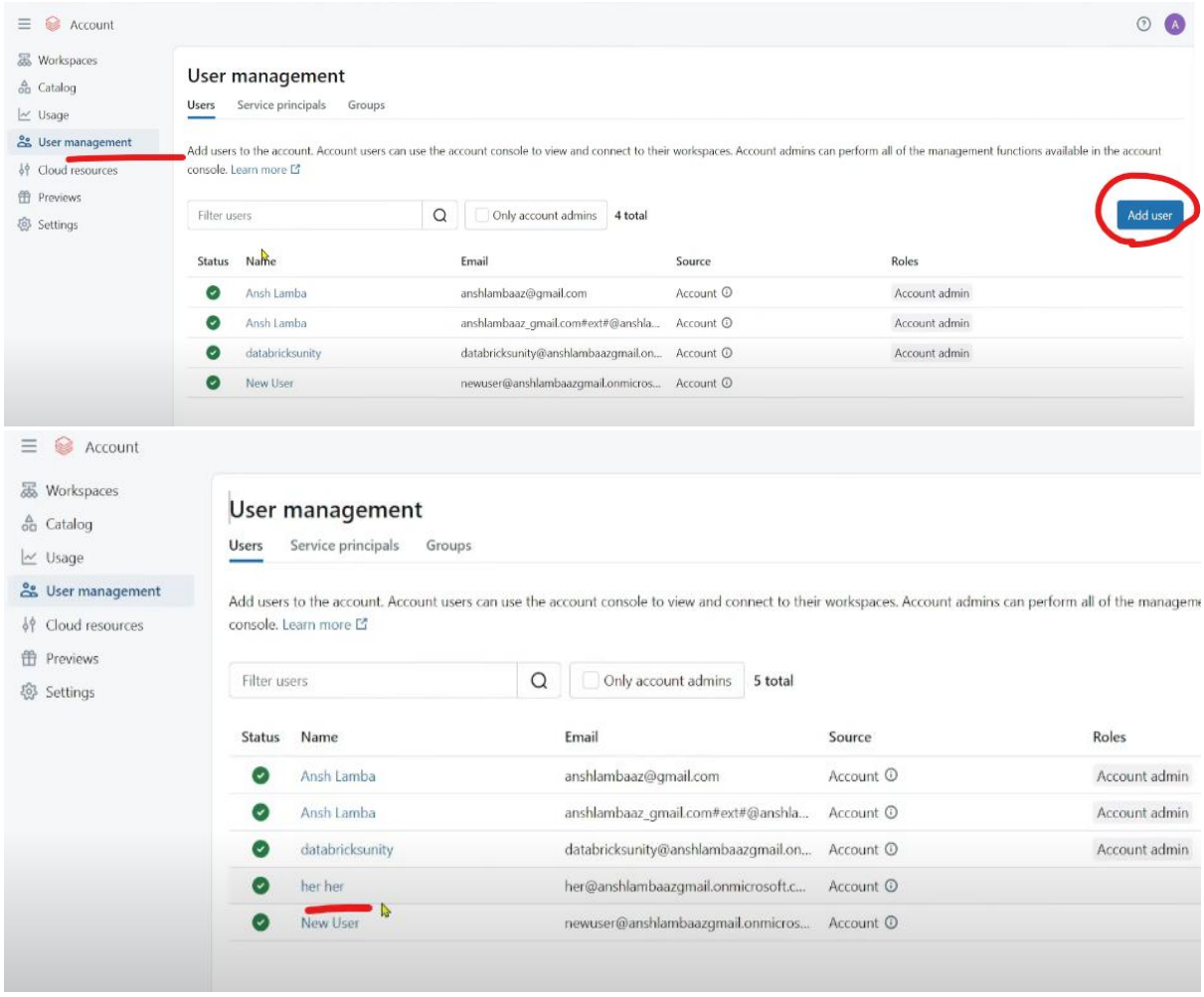
Step1 : Add user in the Entra ID (Go to Entra ID→ Manage → Users)



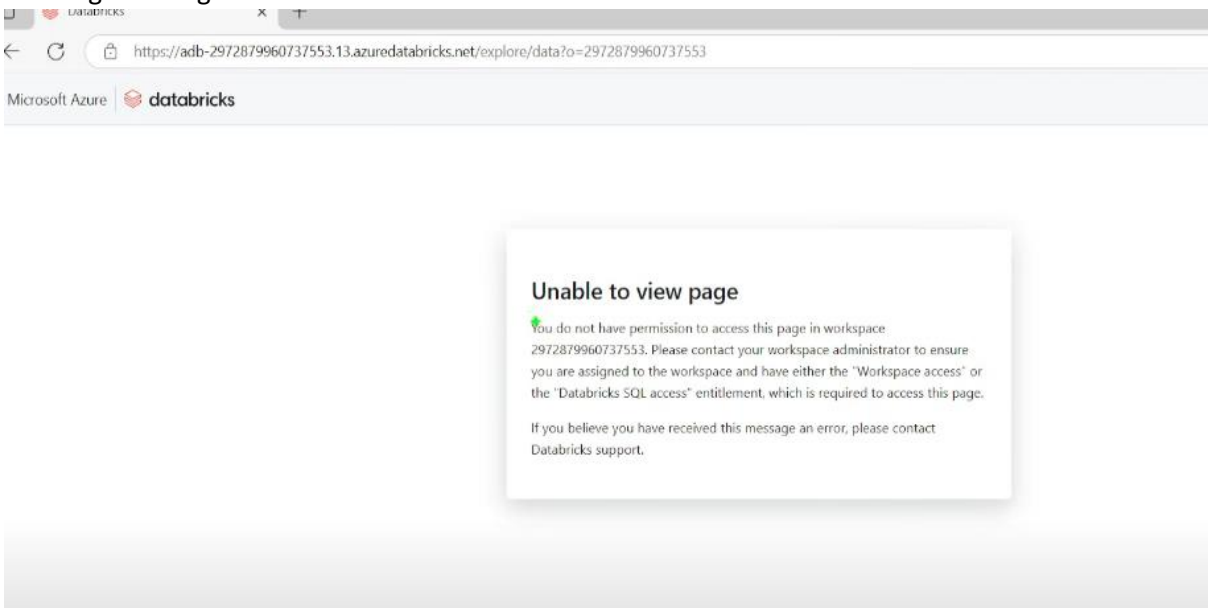


Step2: Add user in Databricks workspace



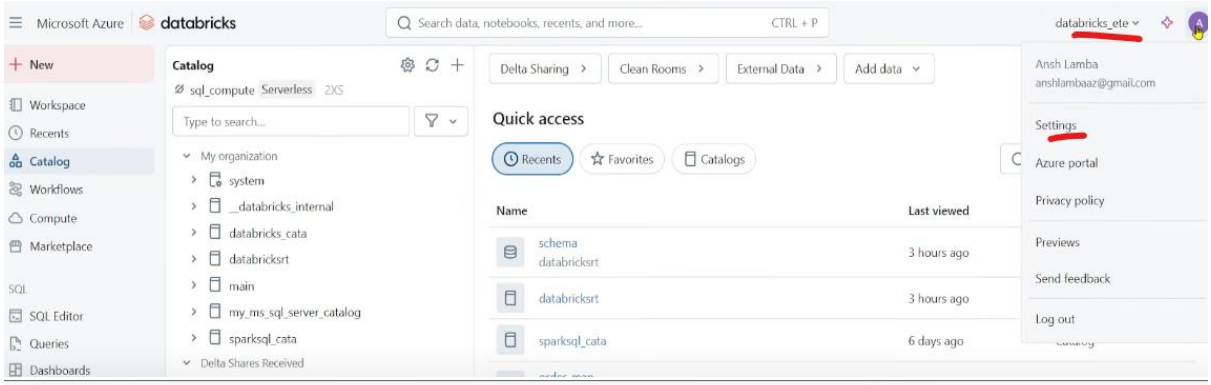


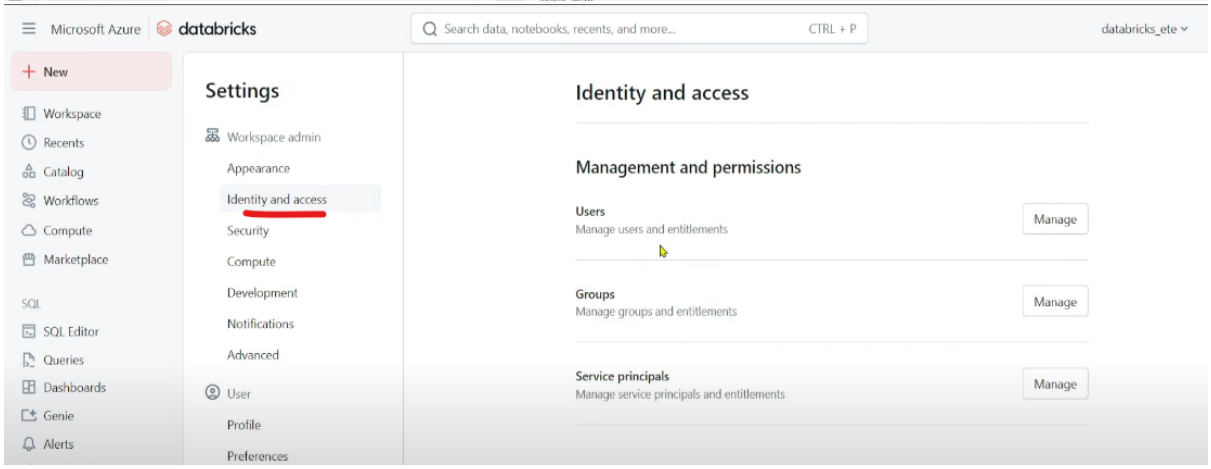
Now login through her id



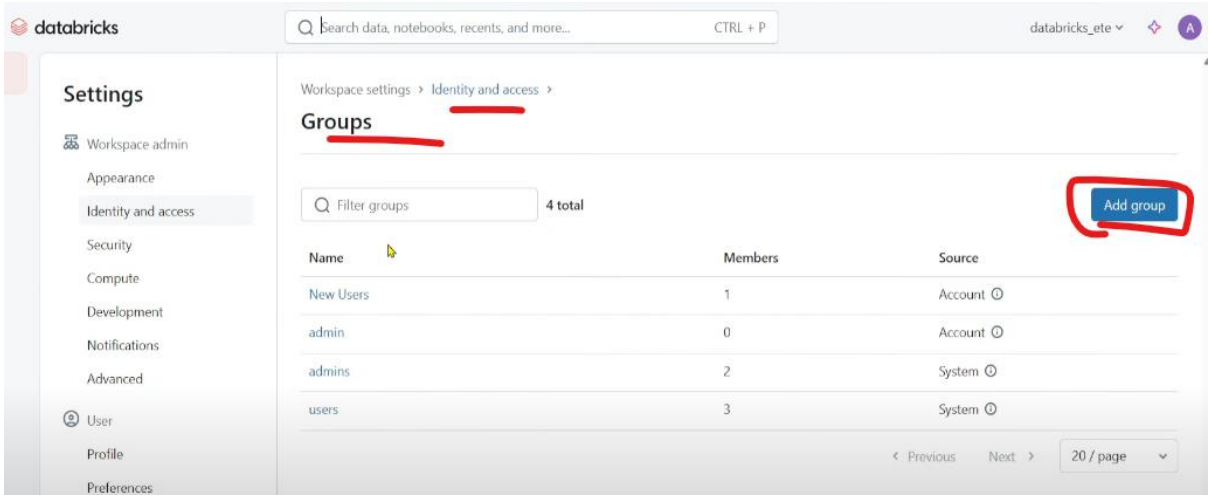
Adding user in Account Console does not mean she will have access to all the catalogs, account console manages all the workspaces but I want any specific workspace to be given access to.

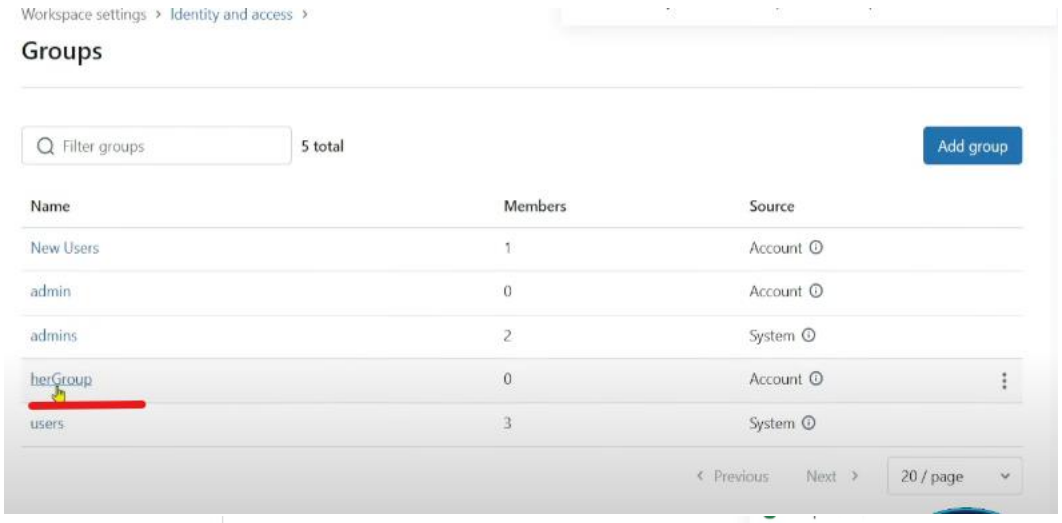
Step3 : Granting workspace specific access

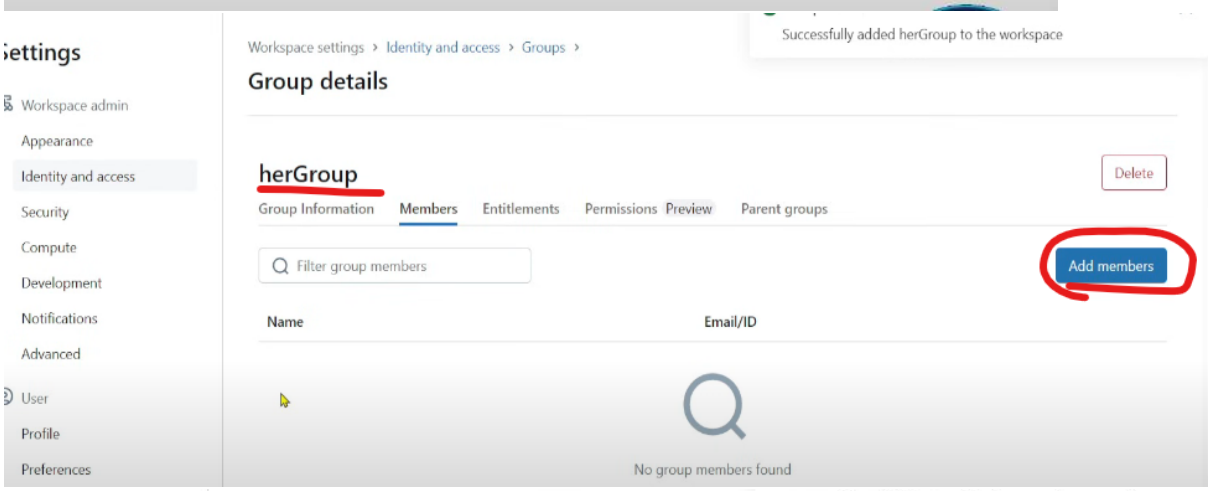


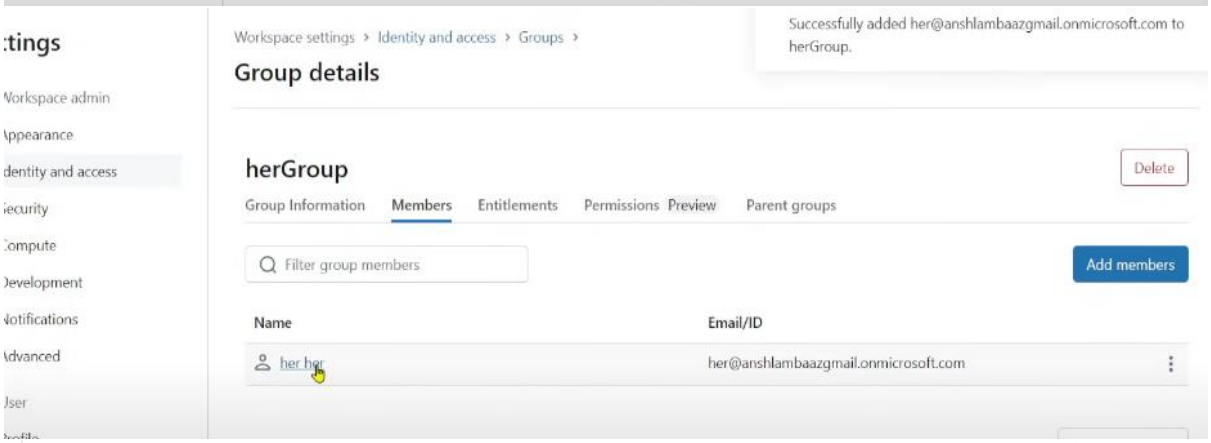


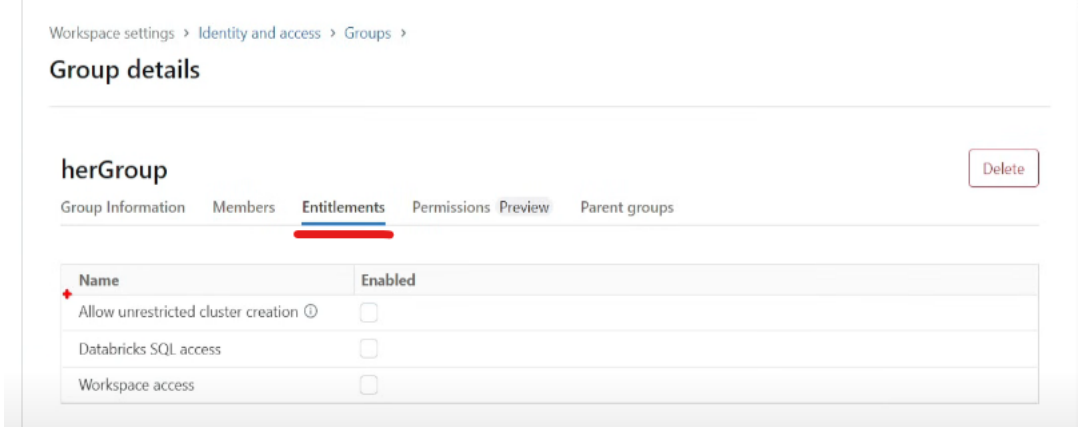
Best way to create a group

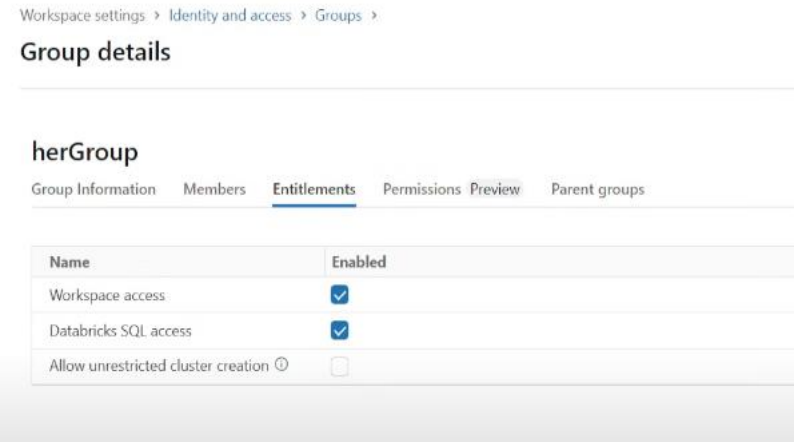



Workspace settings › Identity and access ›
Groups
Filter groups
5 total
Add group
Name
Members
Source
New Users
1
Account
admin
0
Account
admins
2
System
herGroup
0
Account
users
3
System
Previous
Next
20 / page


Successfully added herGroup to the workspace
Settings
Workspace admin
Appearance
Identity and access
Security
Compute
Development
Notifications
Advanced
User
Profile
Preferences
Workspace settings › Identity and access › Groups ›
Group details
herGroup
Delete
Group Information
Members
Entitlements
Permissions Preview
Parent groups
Filter group members
Add members
Name
Email/ID
No group members found

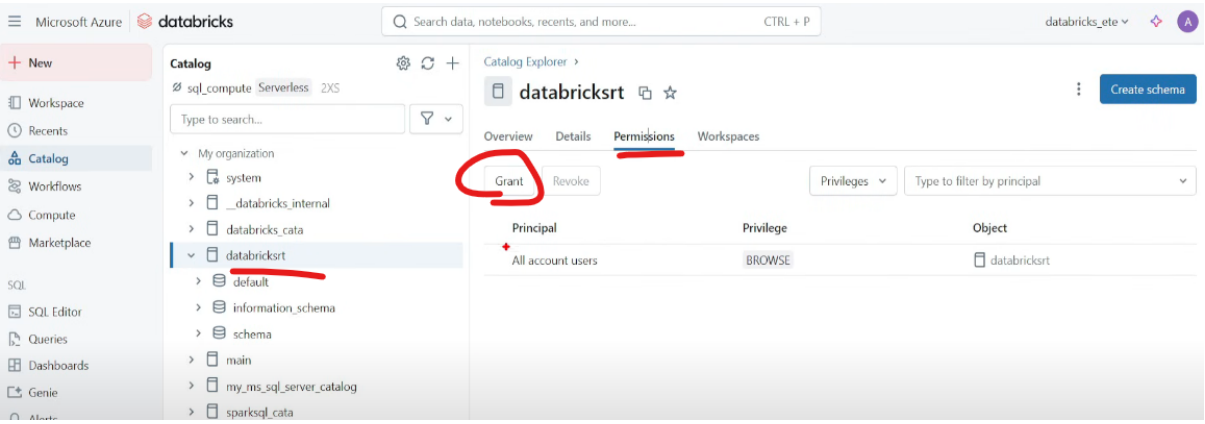

Successfully added her@anshlambaazgmail.onmicrosoft.com to herGroup.
tings
Workspace admin
Appearance
dentity and access
ecurity
ompute
evelopment
otifications
dvanced
ser
Workspace settings › Identity and access › Groups ›
Group details
herGroup
Delete
Group Information
Members
Entitlements
Permissions Preview
Parent groups
Filter group members
Add members
Name
Email/ID
her her
her@anshlambaazgmail.onmicrosoft.com


Workspace settings › Identity and access › Groups ›
Group details
herGroup
Delete
Group Information
Members
Entitlements
Permissions Preview
Parent groups
Name
Enabled
Allow unrestricted cluster creation
Databricks SQL access
Workspace access



Still she will not be having access to Catalogs

Step4: Granting access to catalogs

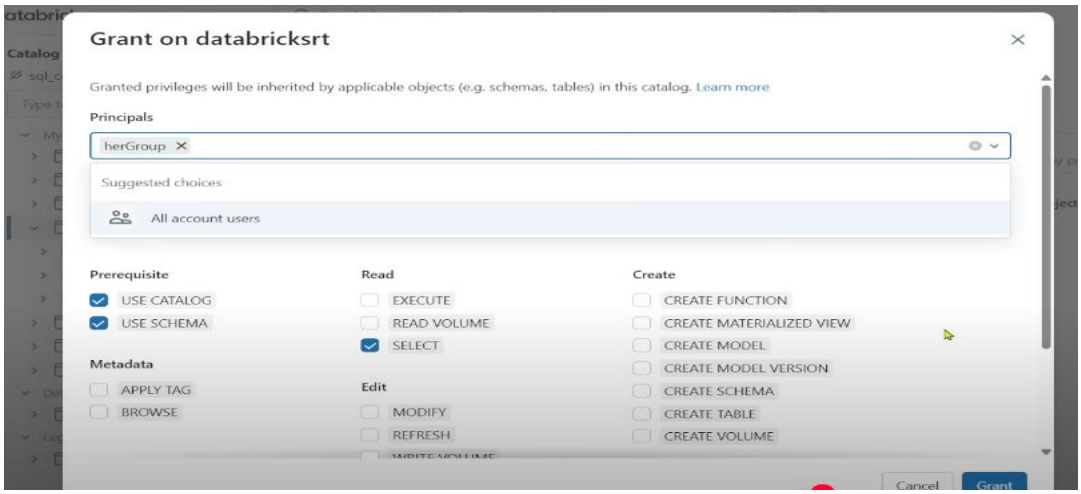


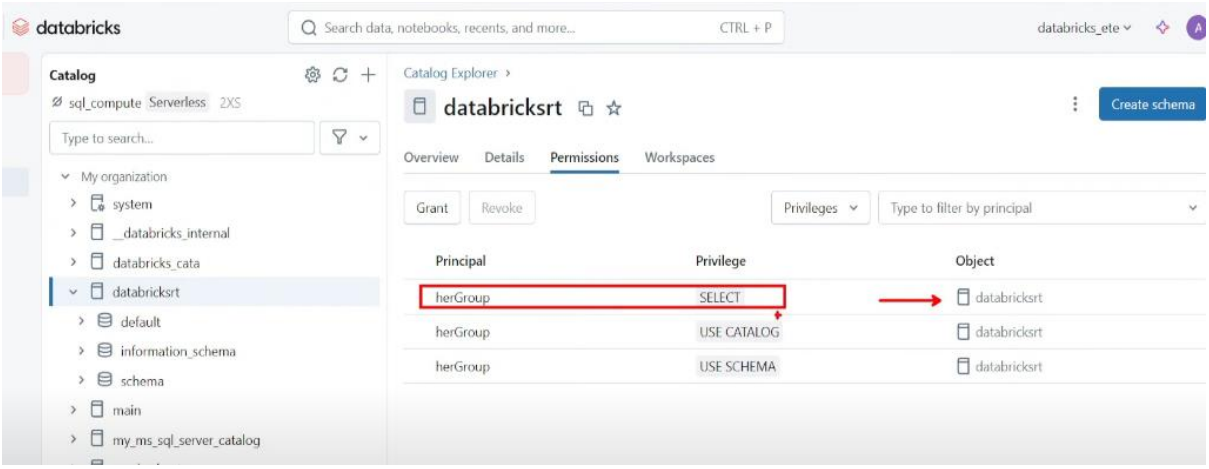
There are 2 types of access:

1) Top to Down : having access to parent container will automatically having access to all child containers

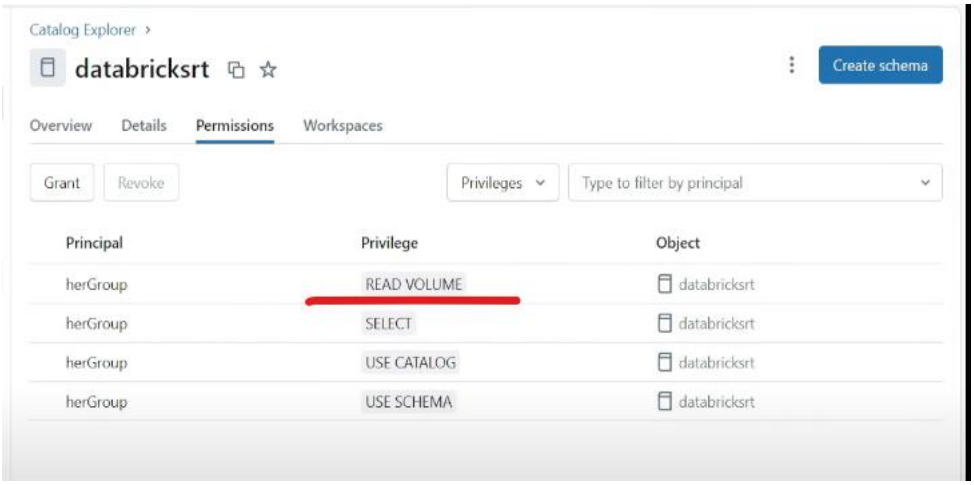
2) Down to Top : having access to a container means all kind of read/write permissions on the container but only usage access to its parent container (if parent container contains any other container, it will be hidden)

Top to Down



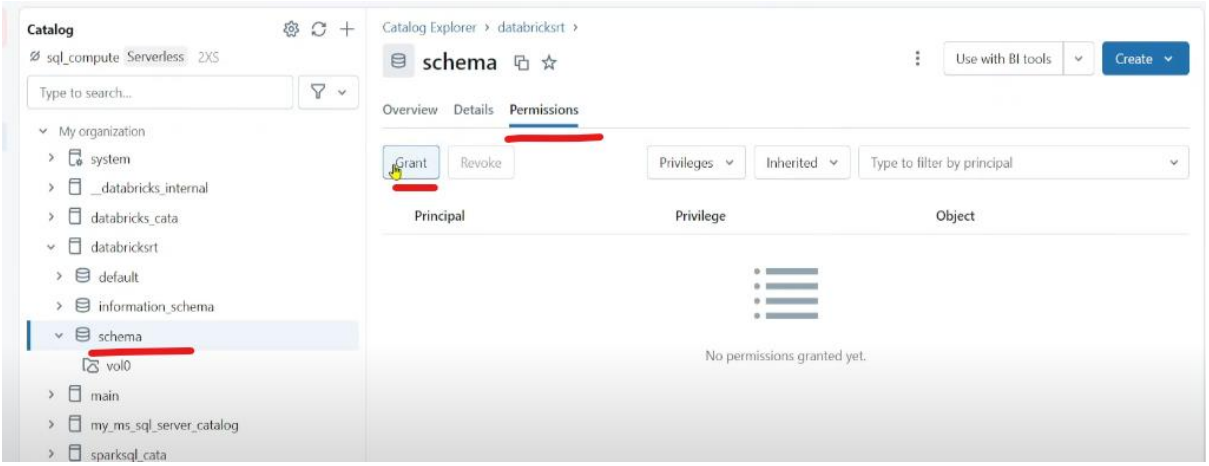


Add respective access

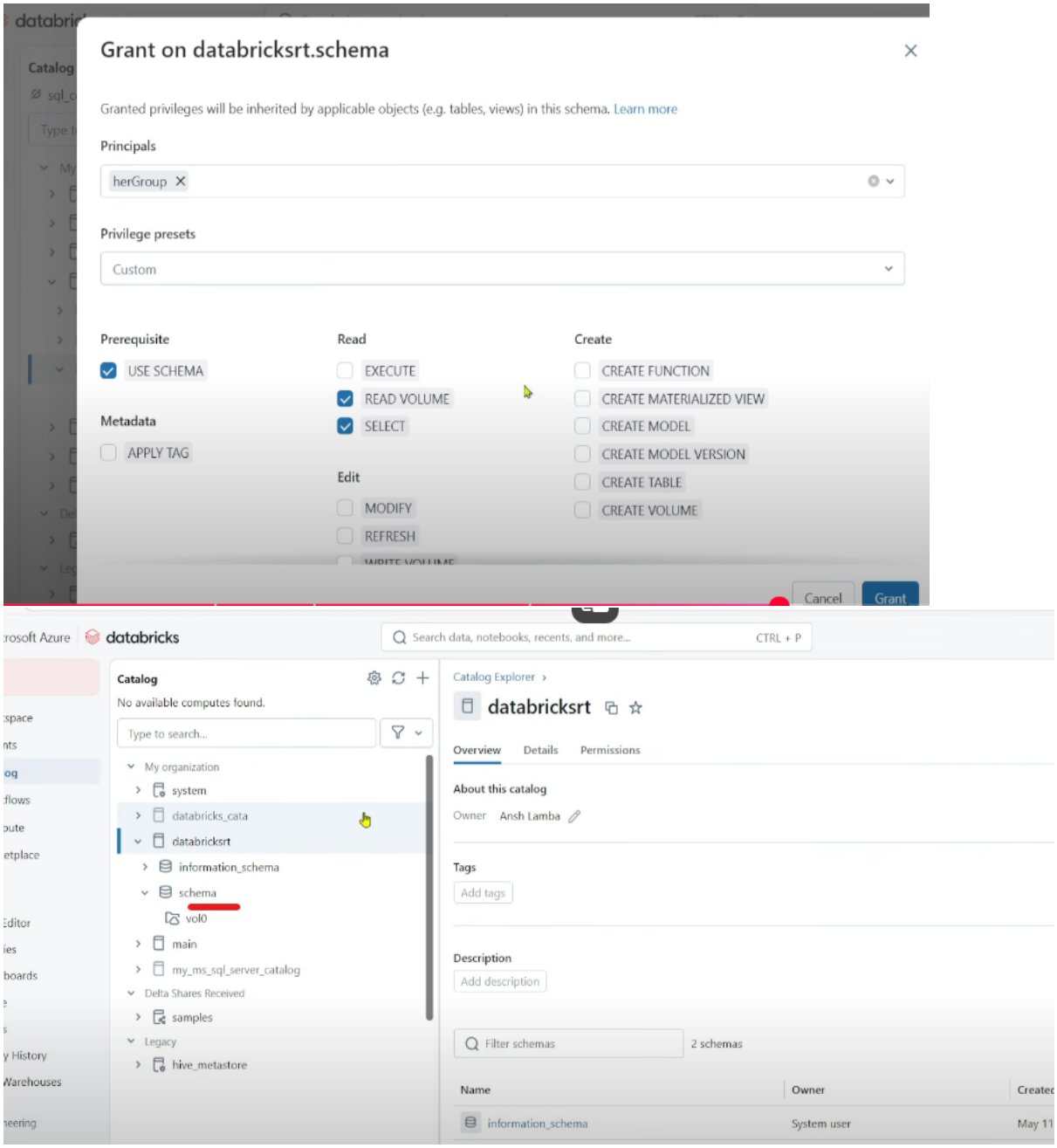


Down to Top

Lets delete all the access for now



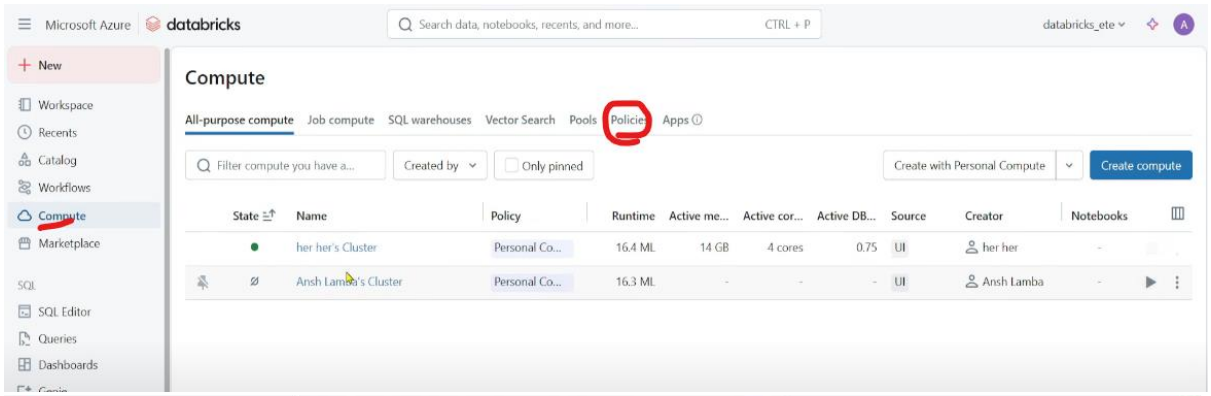
Go one level down i.e. schema and give permission

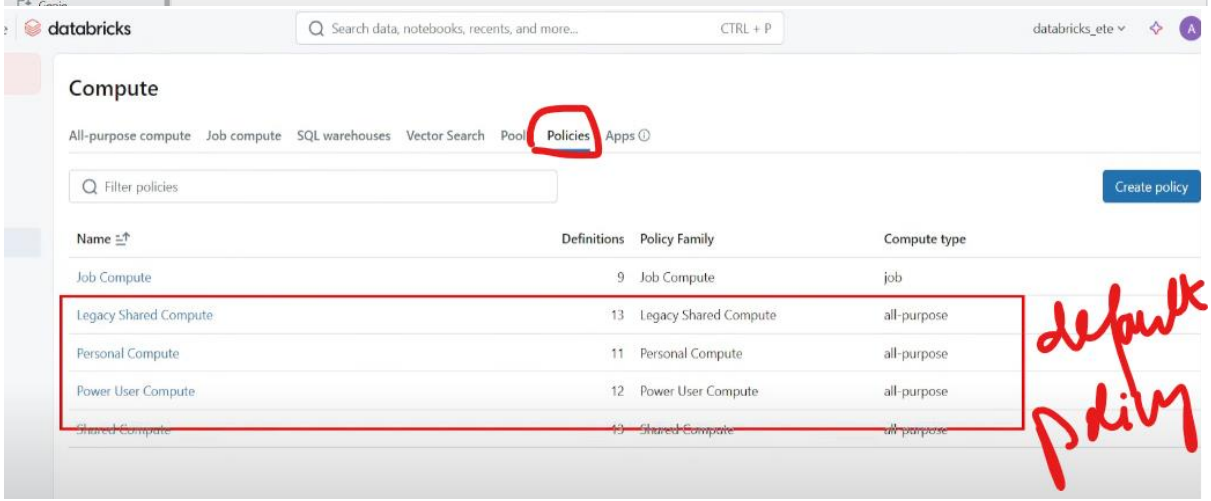


Note now she cant see the default schema, beauty of down to top approach

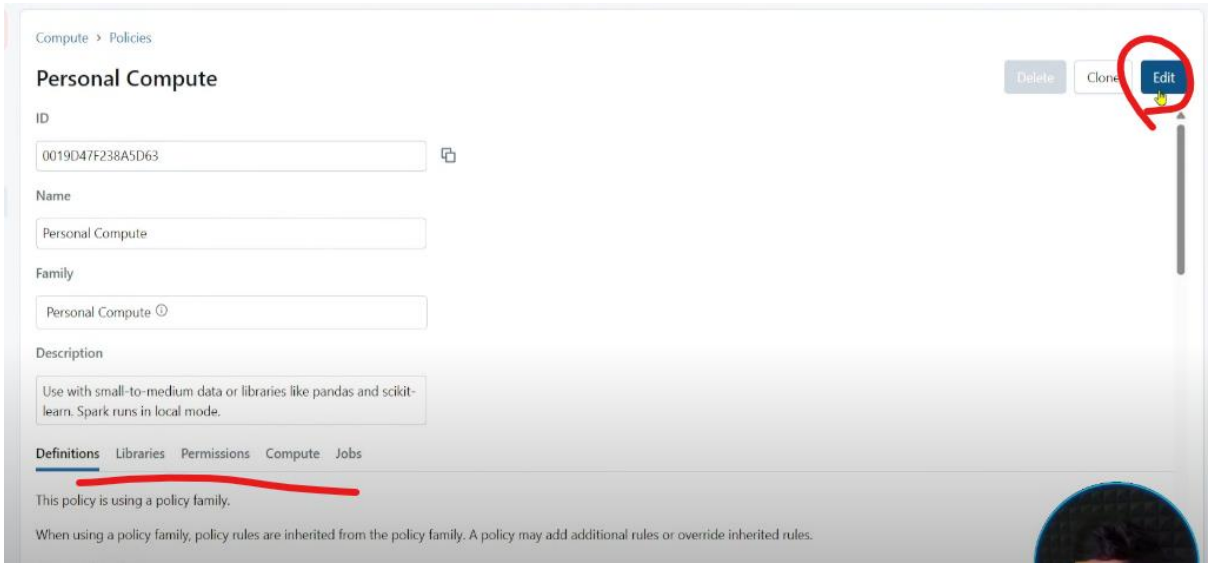
Scenario – 6 (Access on cluster)

We don't want any new user to create new cluster.

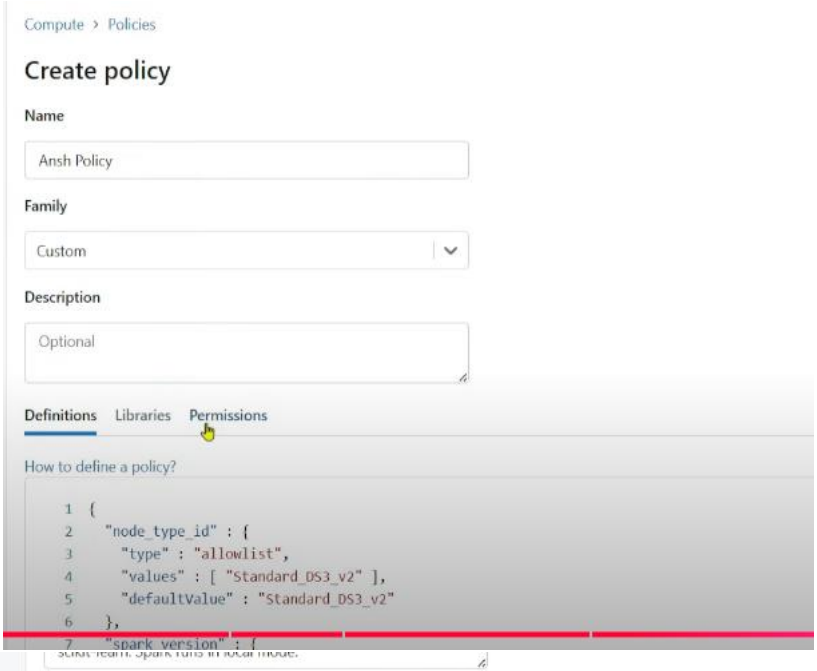


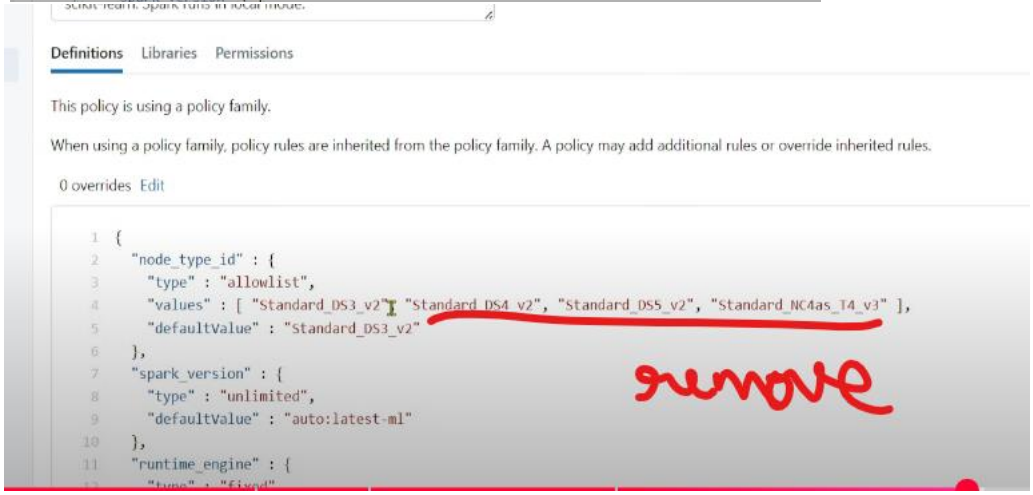


Policies are pre defined pre configured area

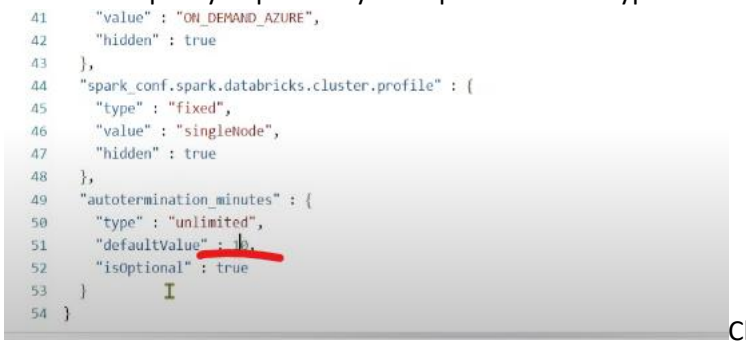


Create a new policy(Copy the code from Personal Compute)

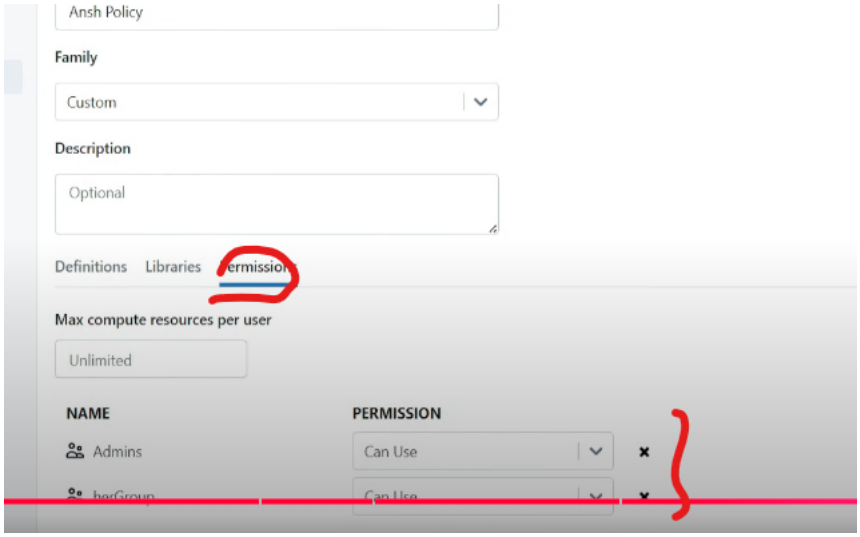




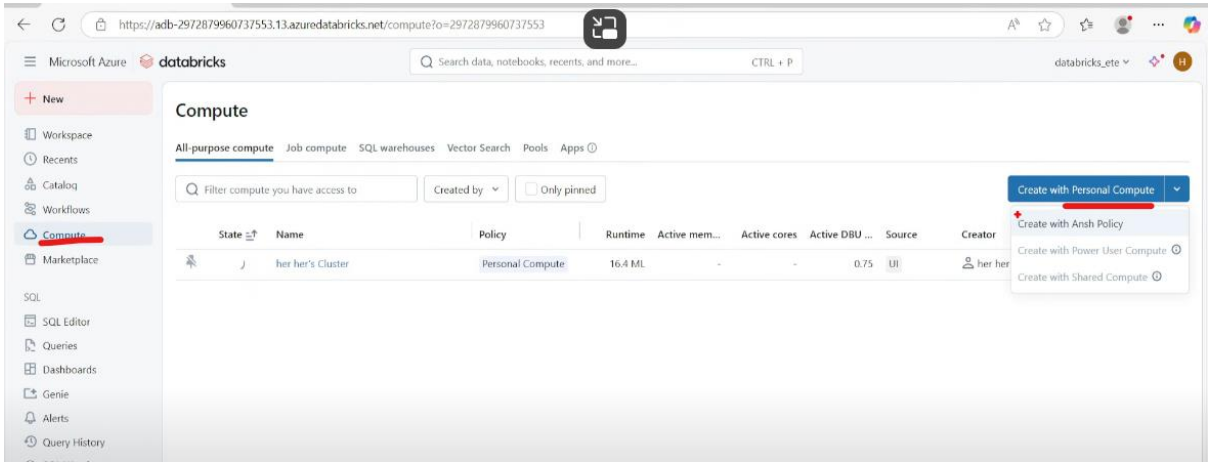
Means new policy implies only one specific cluster type to be included

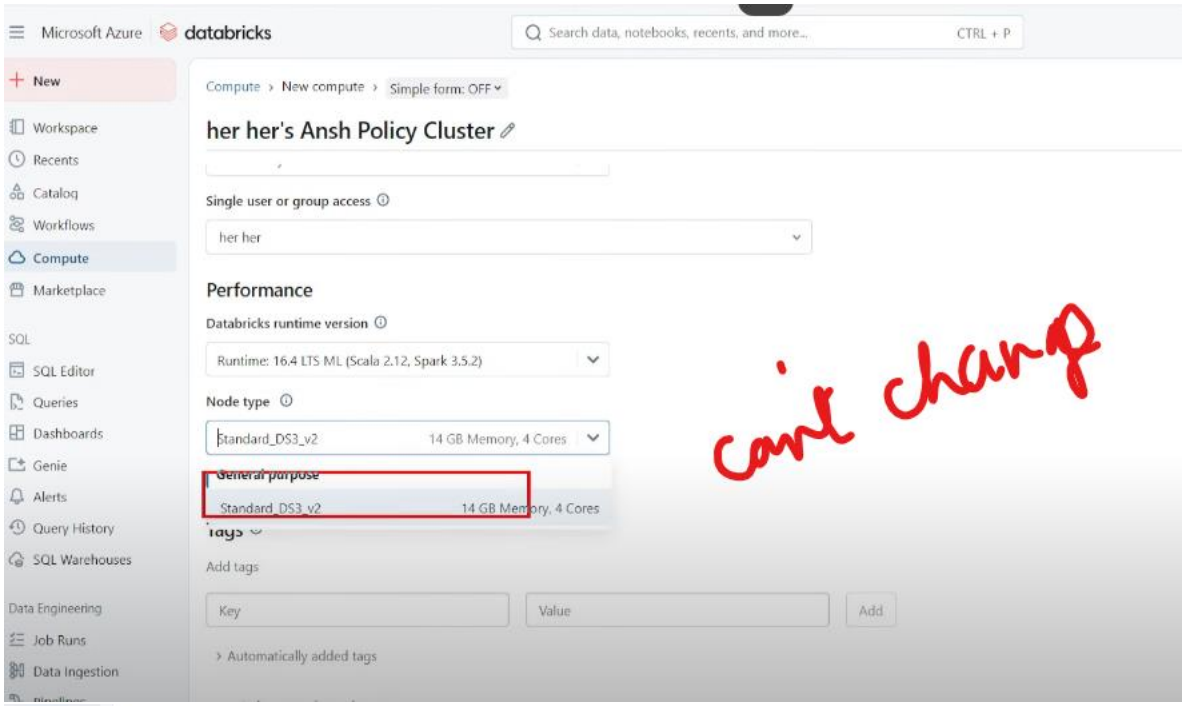


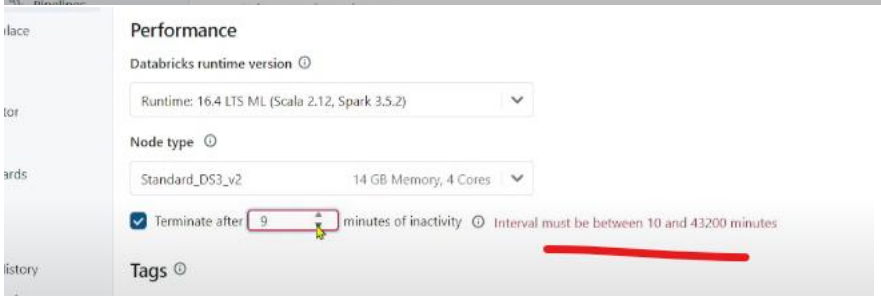
Changing 4320 mins to 10 mins



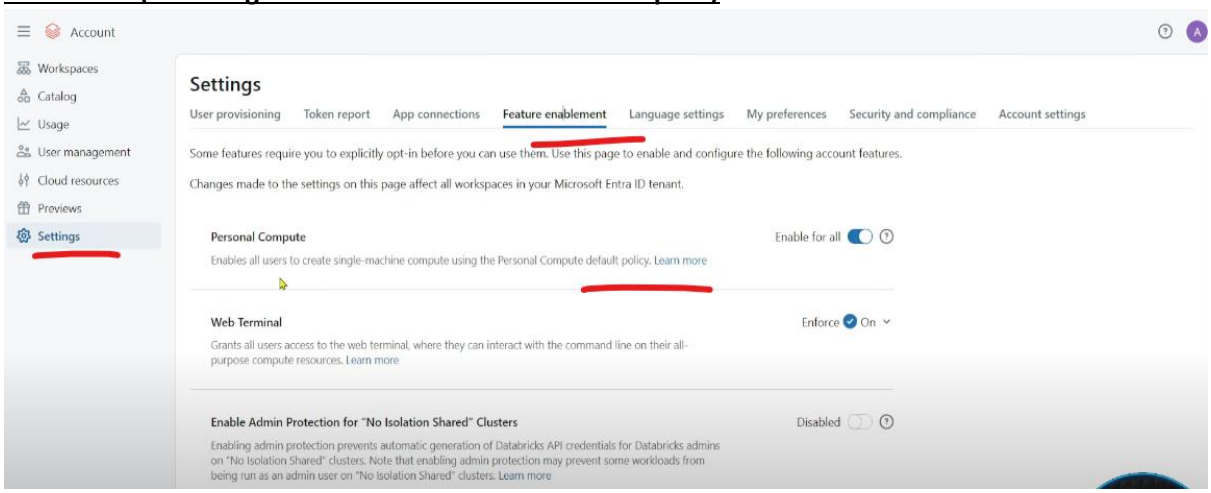
Now try creating a Cluster, all other policies are disabled

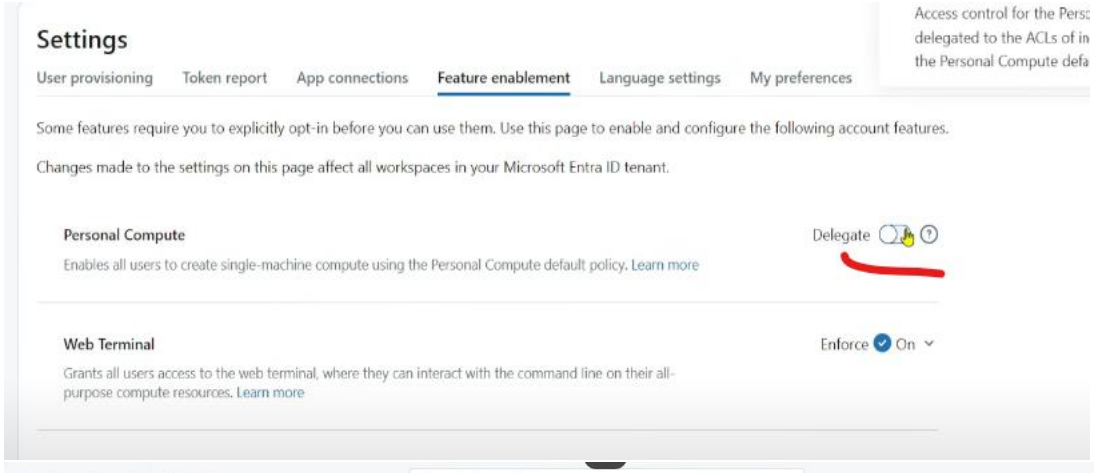


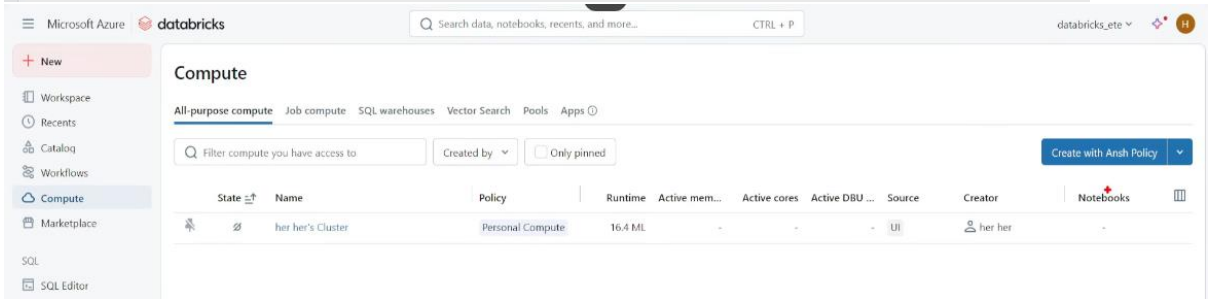




Scenario-7 (Revoking all users to create Personal Compute)





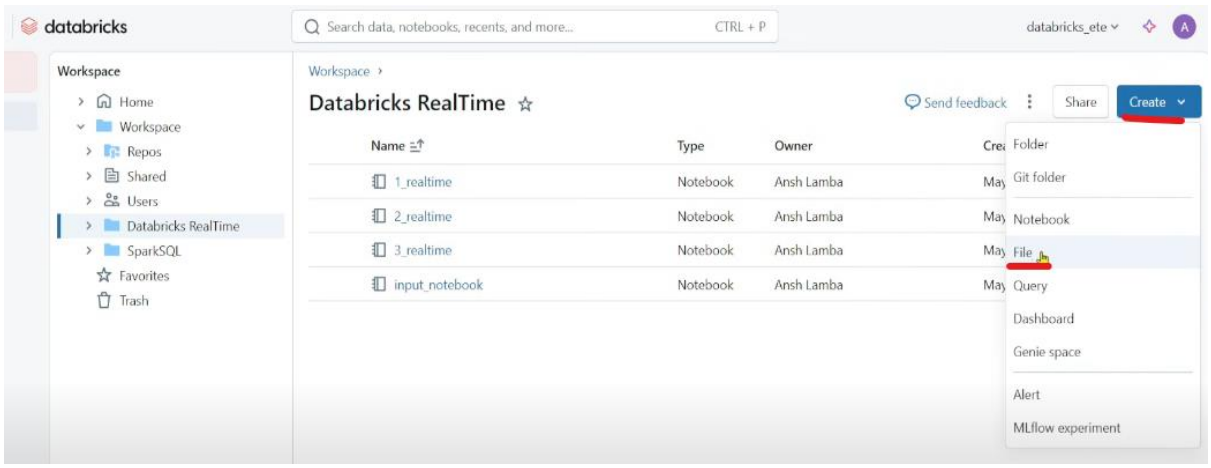


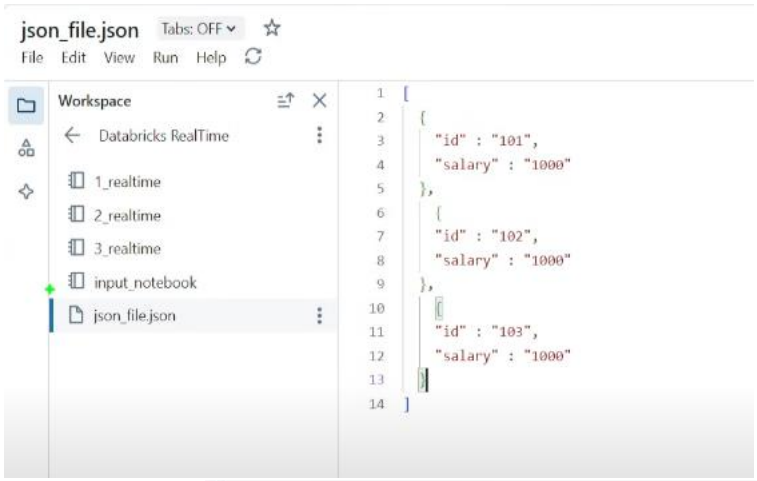
Personal compute is gone

Scenario-8 (Databricks Workspace Files)

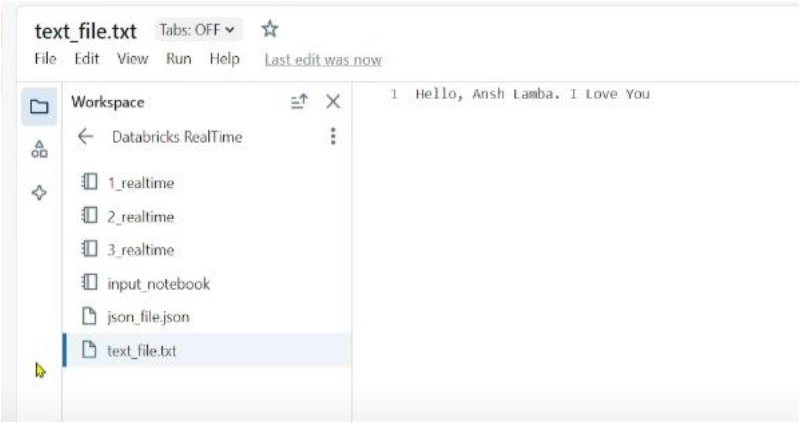
Files in Workspace Files (like json file that we use for Lookup or .py file containing API details)

Create a new file

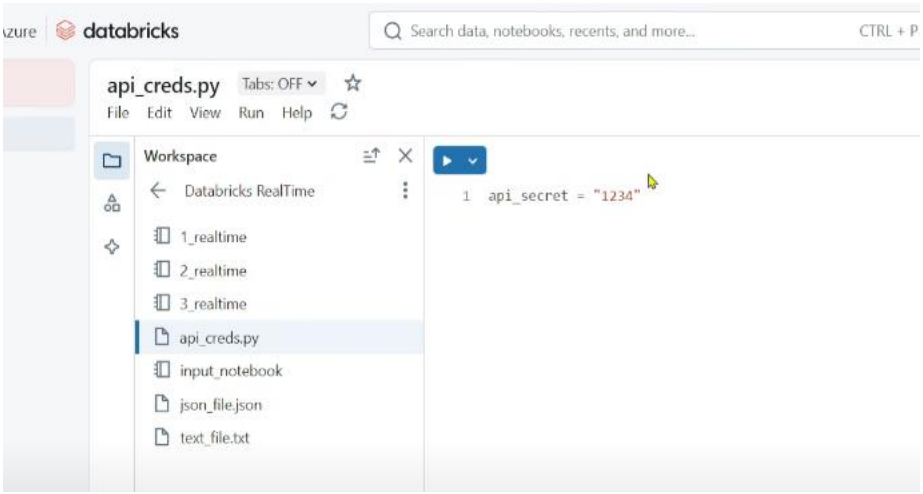




Create a text file



Create a py file

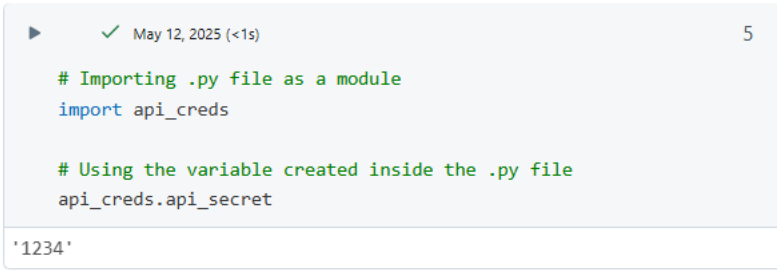


Create a notebook
r means read mode

Writing new line in the file

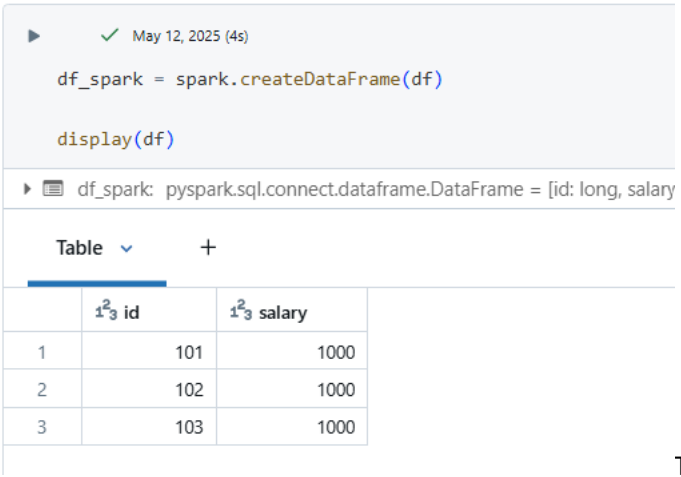
a means append

Reading Python File

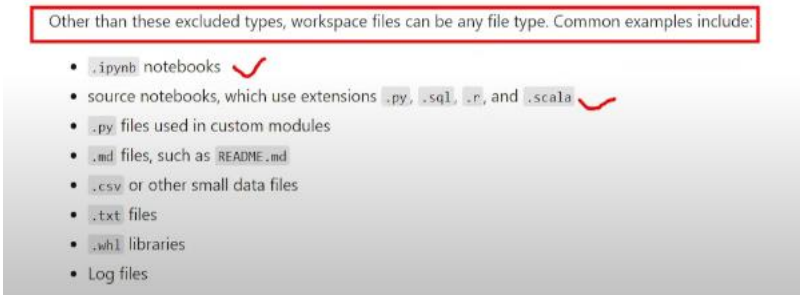


Reading JSON Data

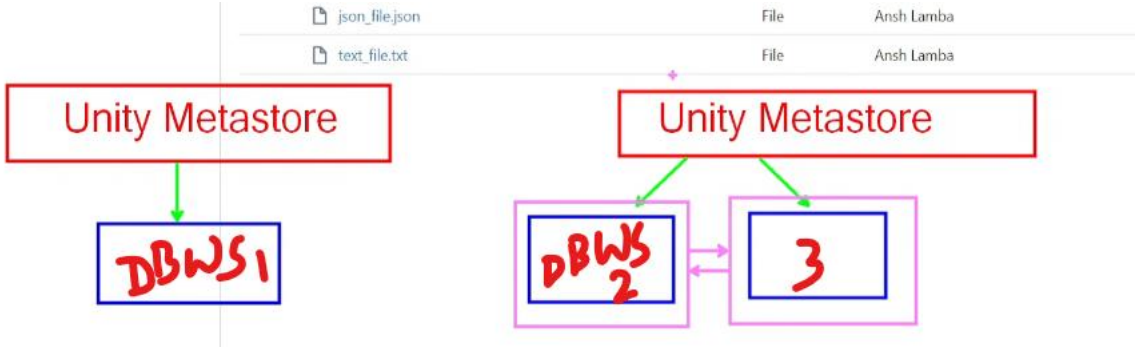
	id	salary
0	101	1000
1	102	1000
2	103	1000



To convert into a dataframe

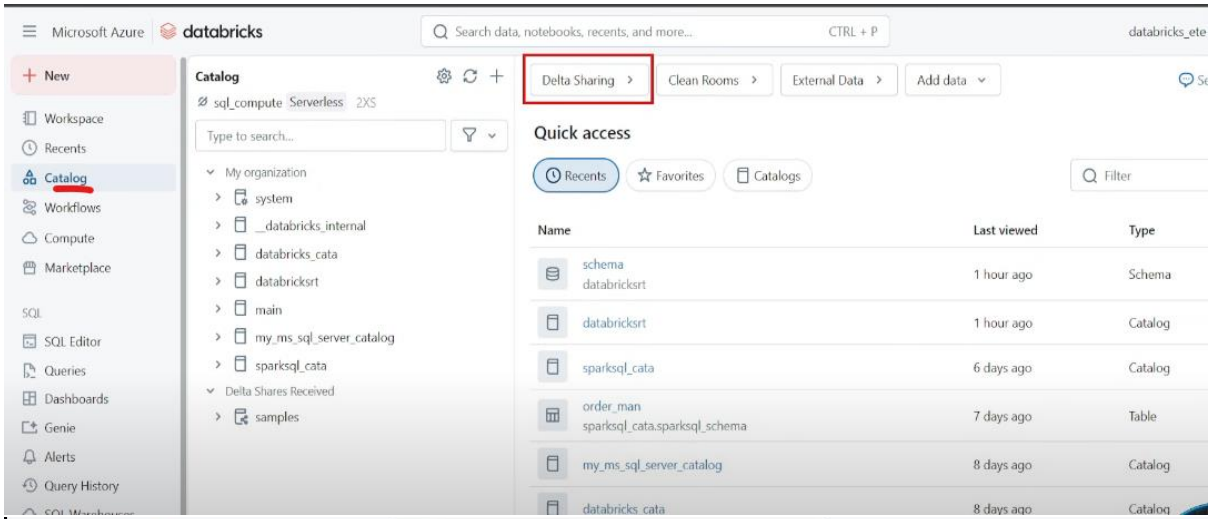


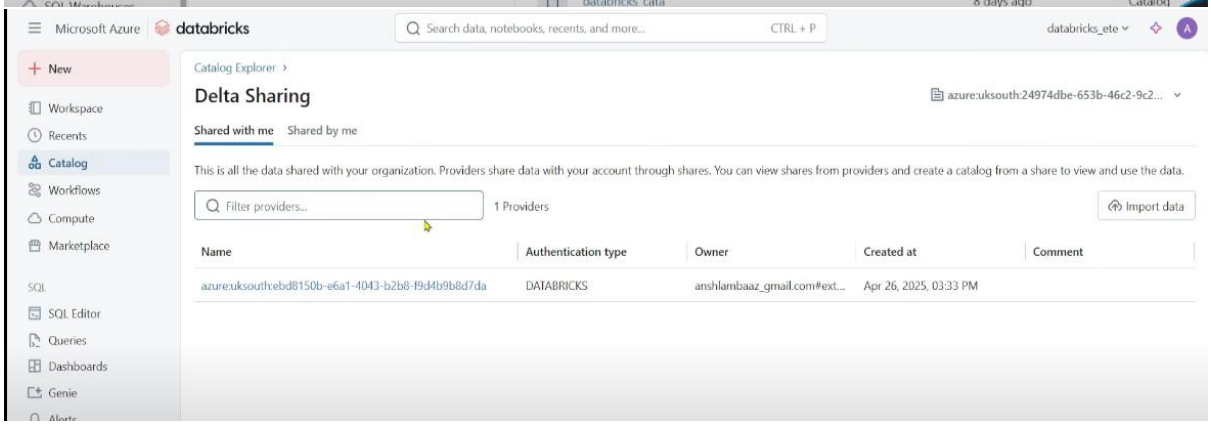
Scenario-9 (Databricks Delta Sharing)



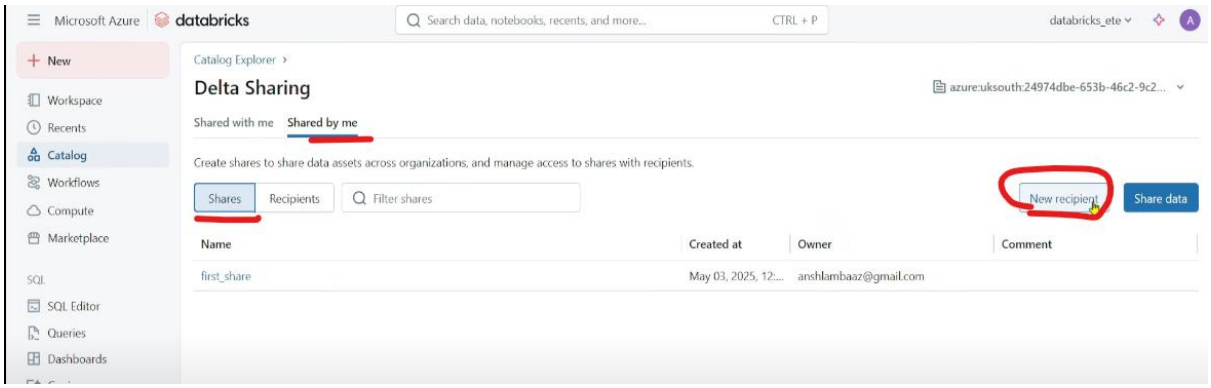
UM1 attached to WS1 while UM2 attached to WS2 and WS3, now catalogs in WS2 and WS3 can be shared among them as they are connected to same UM, now can I share catalogs of DBWS1 with DBWS 2 or 3? → Traditionally not possible.

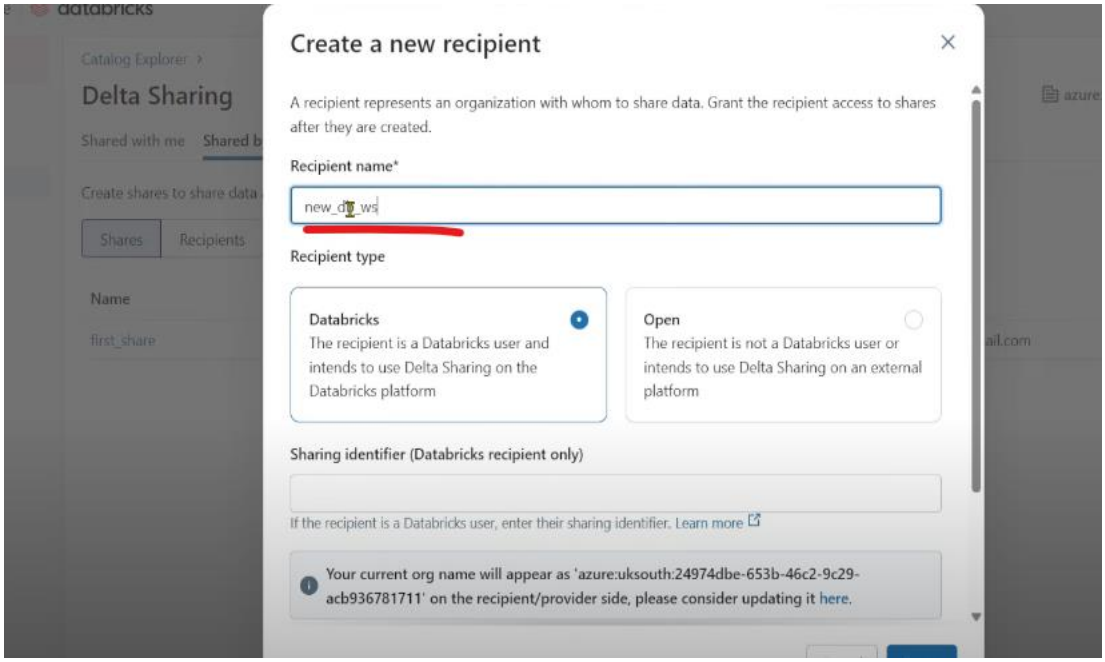
But possible with Delta Sharing



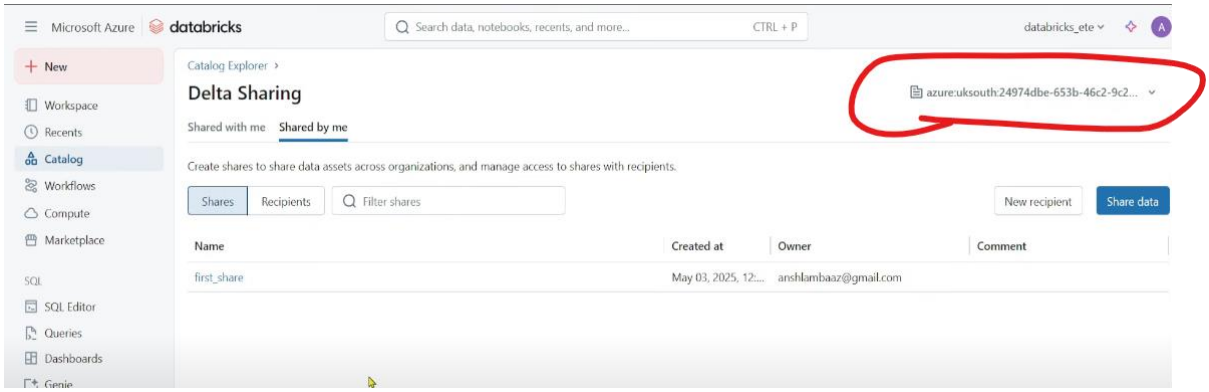


Catalogs → Delta Sharing → create New recipient

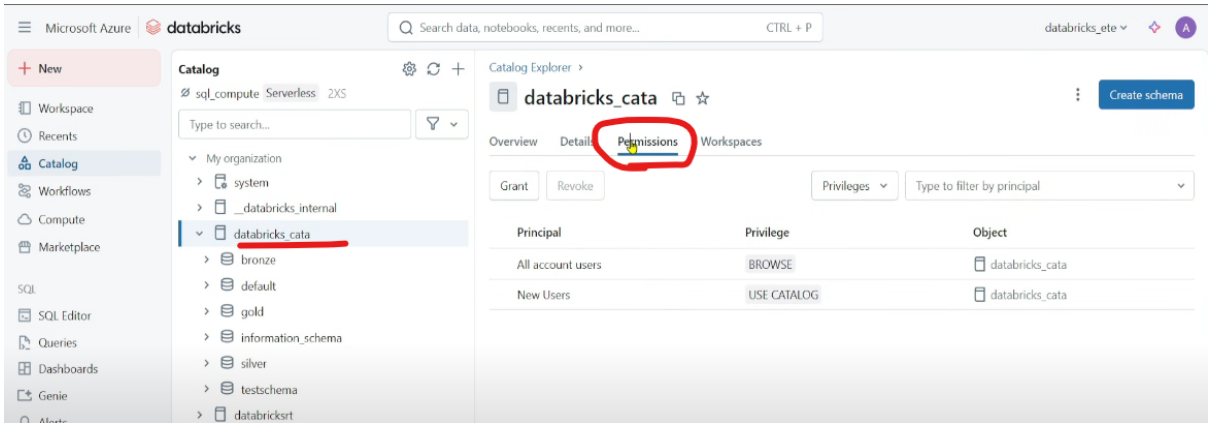


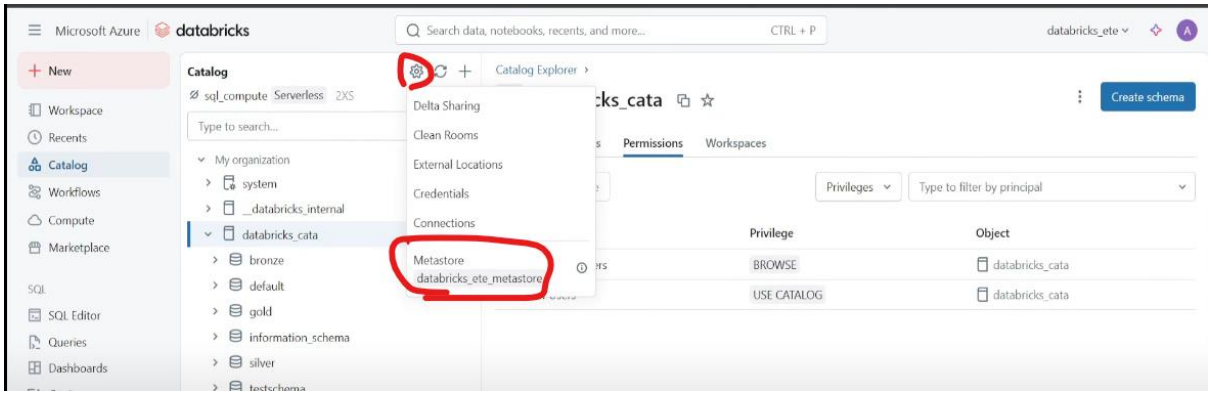


Sharing identifier: go to other DB, copy highlighted and paste.

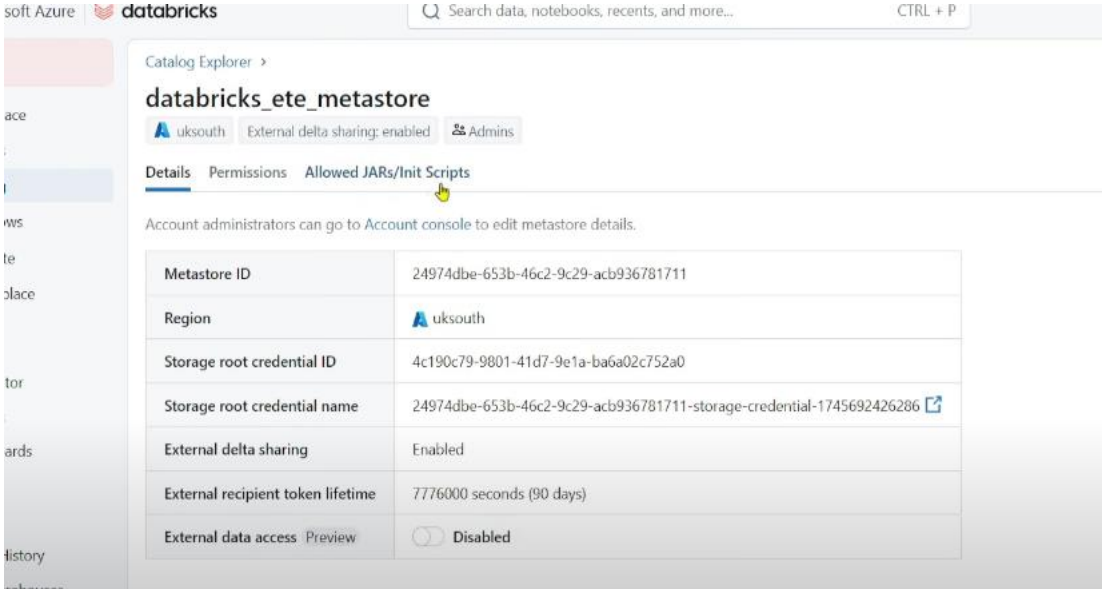


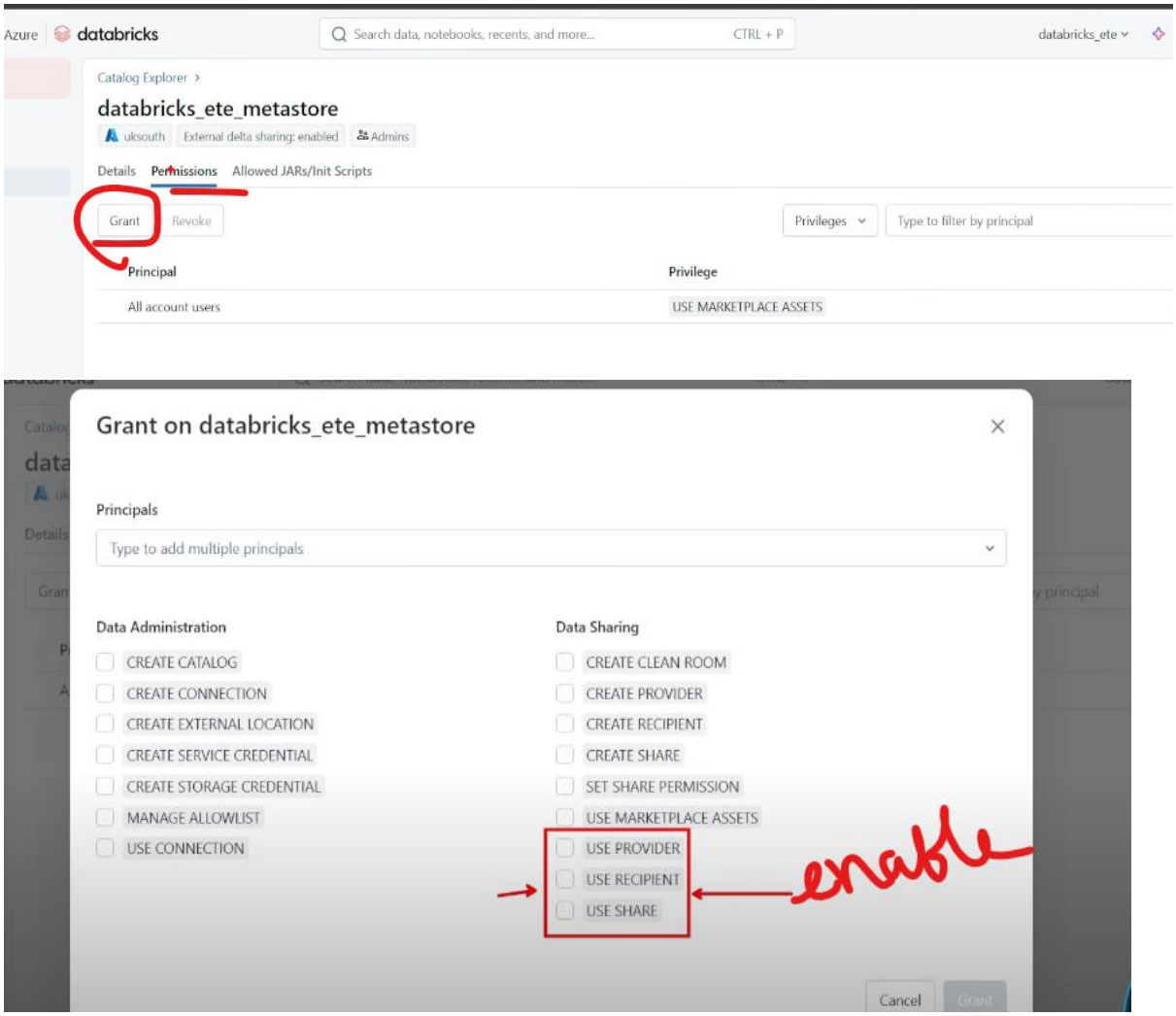
But that user cant use the catalog → have to grant permissions





Click on the metastore(the catalog is attached to this metastore)




Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_ete
Catalog Explorer
databricks_ete_metastore
uksouth
External delta sharing: enabled
Admins
Details
Permissions
Allowed JARs/Init Scripts
Grant
Revoke
Privileges
Type to filter by principal
Principal
Privilege
All account users
USE MARKETPLACE ASSETS
Grant on databricks_ete_metastore
Principals
Type to add multiple principals
Data Administration
CREATE CATALOG
CREATE CONNECTION
CREATE EXTERNAL LOCATION
CREATE SERVICE CREDENTIAL
CREATE STORAGE CREDENTIAL
MANAGE ALLOWLIST
USE CONNECTION
Data Sharing
CREATE CLEAN ROOM
CREATE PROVIDER
CREATE RECIPIENT
CREATE SHARE
SET SHARE PERMISSION
USE MARKETPLACE ASSETS
USE PROVIDER
USE RECIPIENT
USE SHARE
enable
Cancel
Grant

Scenario-10 (Functions)

May 12, 2025 (3s) 1

```
%sql
CREATE OR REPLACE FUNCTION databricksrt.schema.myfunc(p_input STRING)
RETURNS STRING
LANGUAGE SQL
RETURN concat(p_input,"hi")
```

May 12, 2025 (9s) 2

```
%sql
CREATE TABLE databricksrt.schema.dummy_tbl
(
  id INT,
  salary INT,
  name STRING
)
```

May 12, 2025 (4s) 3

```
%sql
INSERT INTO databricksrt.schema.dummy_tbl
VALUES(4,100,'aa'),
(5,100,'bb'),
(6,9,'cc')
```

_sqldf: pyspark.sql.connect.dataframe.DataFrame = [num_affected_rows: long, num_inserted_rows: long]

May 12, 2025 (13s) 4

```
%sql
SELECT id, salary, databricksrt.schema.python_func(name) as func_calling FROM databricksrt.schema.dummy_tbl
```

_sqldf: pyspark.sql.connect.dataframe.DataFrame = [id: integer, salary: integer ... 1 more field]

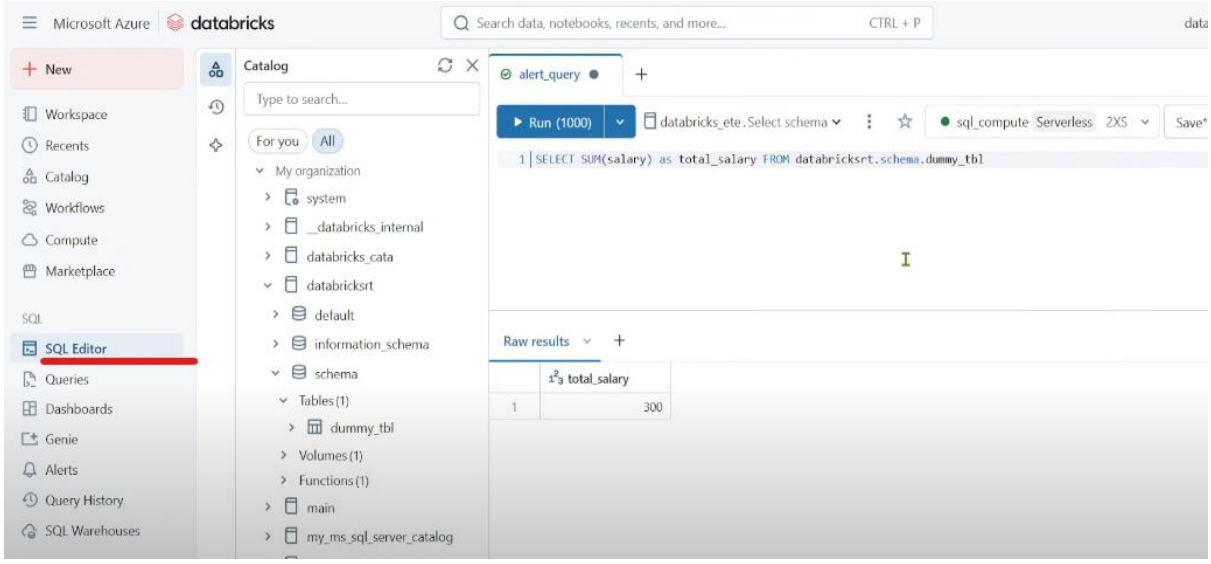
Table +

	id	salary	func_calling
1	1	100	aahi
2	2	100	bbhi
3	3	100	cchi

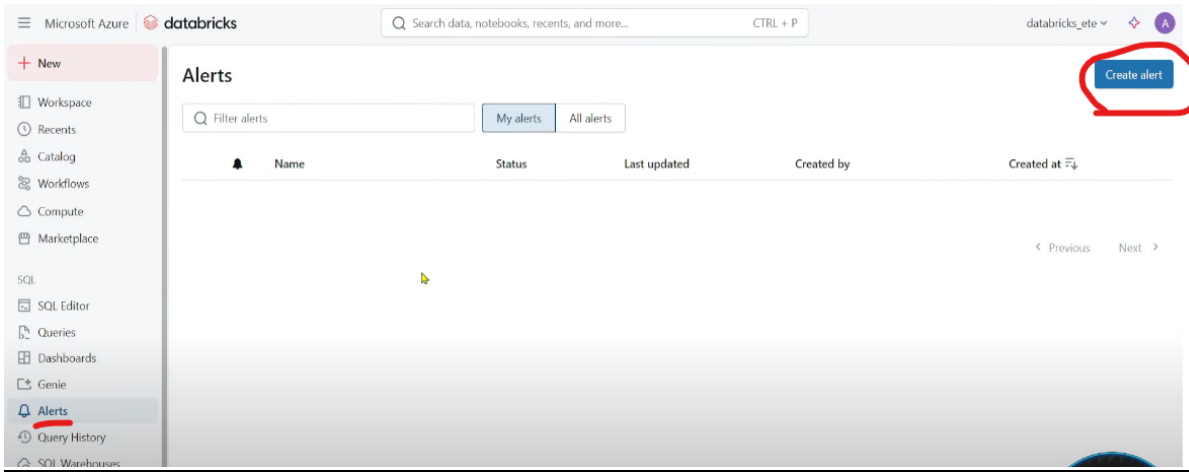
May 12, 2025 (2s) 5

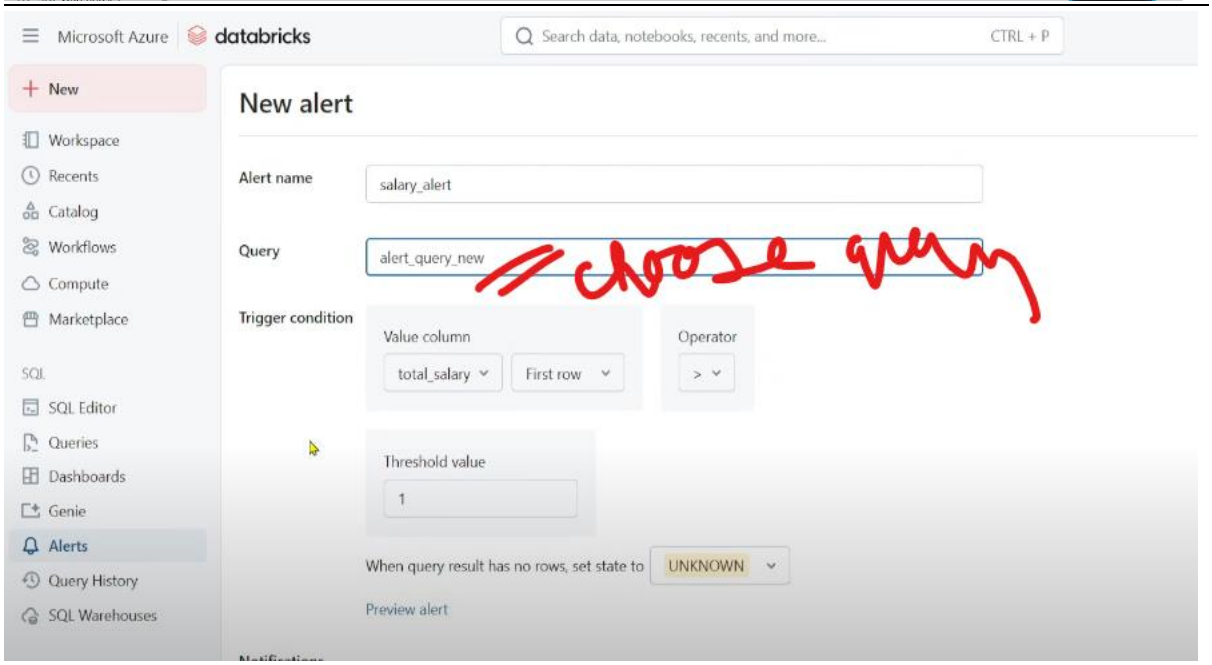
```
%sql
CREATE OR REPLACE FUNCTION databricksrt.schema.python_func(p_input STRING)
RETURNS STRING
LANGUAGE PYTHON
AS
$$
    return p_input + "hi"
$$
```

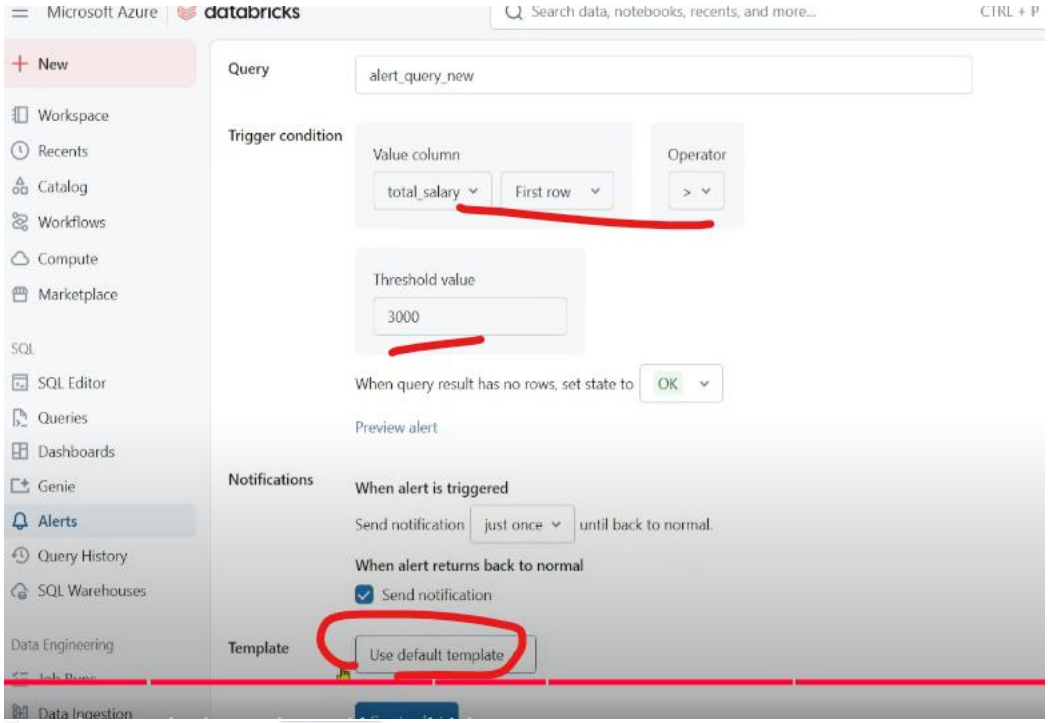
Scenario-11 (Monitoring & Alerts)

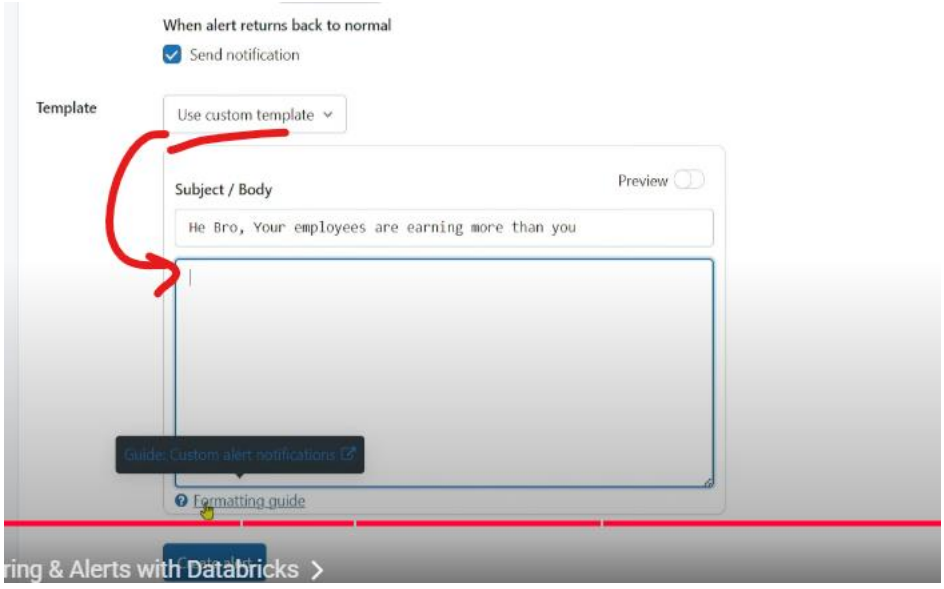


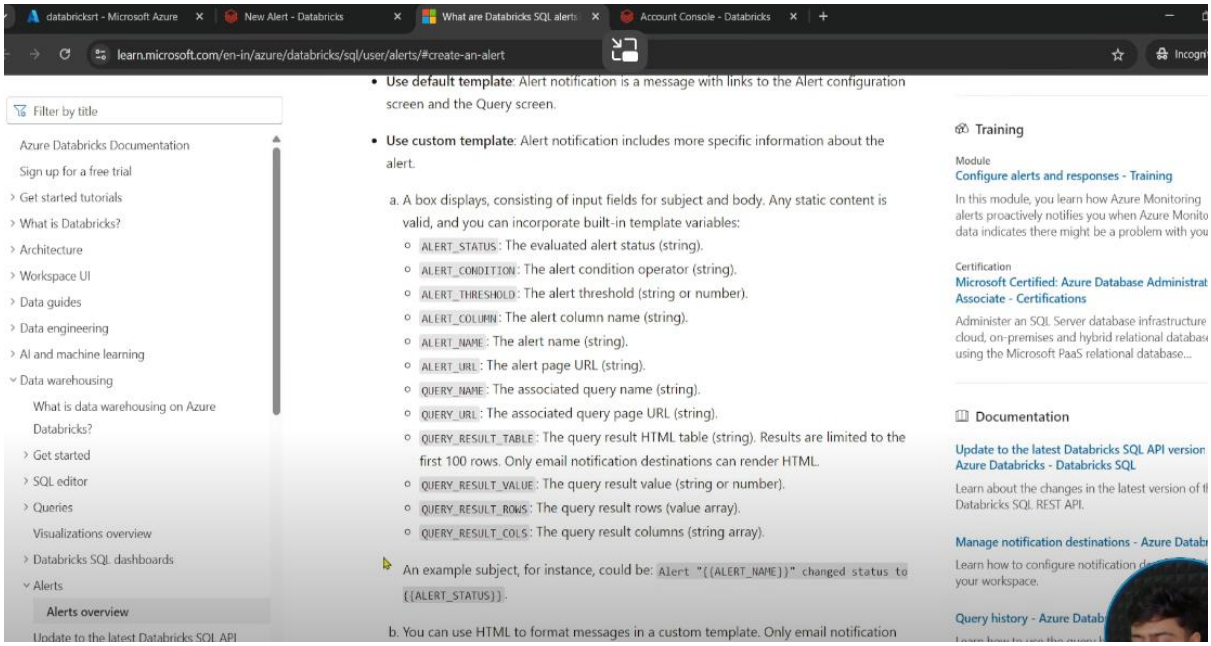
Save the query as alert_query_new



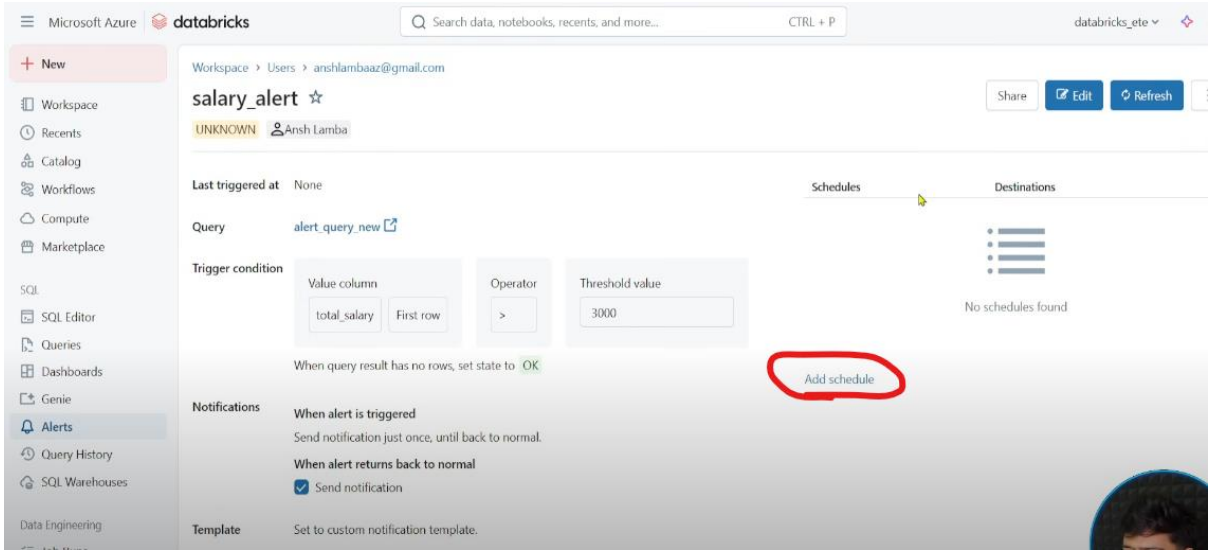


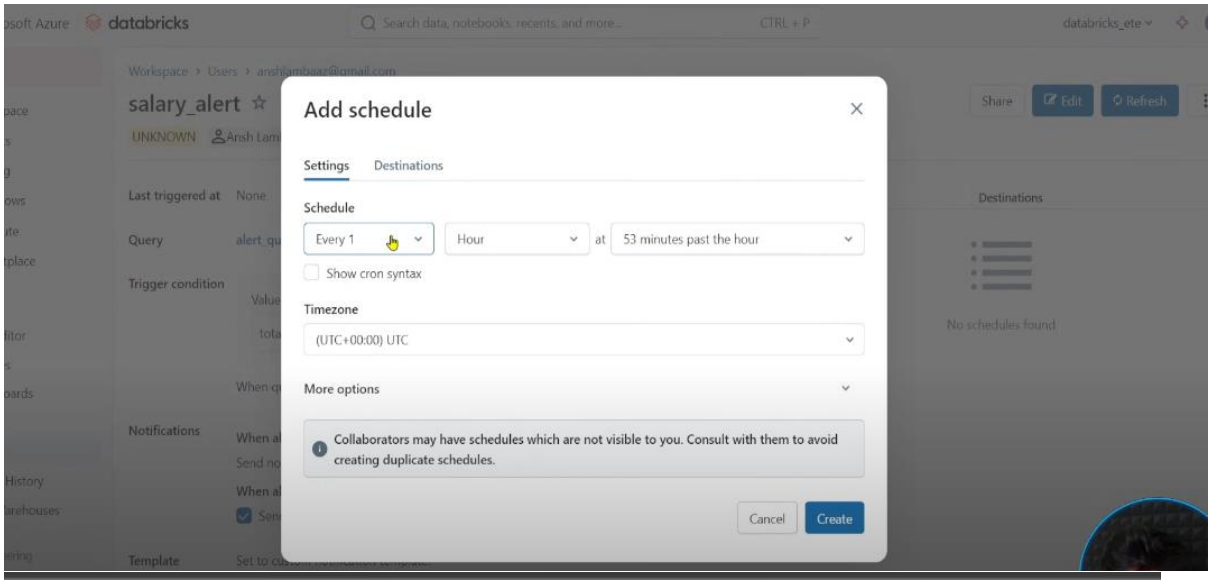

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
New
Workspace
Recents
Catalog
Workflows
Compute
Marketplace
SQL
SQL Editor
Queries
Dashboards
Genie
Alerts
Query History
SQL Warehouses
Data Engineering
Data Ingestion
Query
alert_query_new
Trigger condition
Value column
total_salary
First row
Operator
>
Threshold value
3000
When query result has no rows, set state to
OK
Preview alert
Notifications
When alert is triggered
Send notification
just once
until back to normal.
When alert returns back to normal
Send notification
Template
Use default template

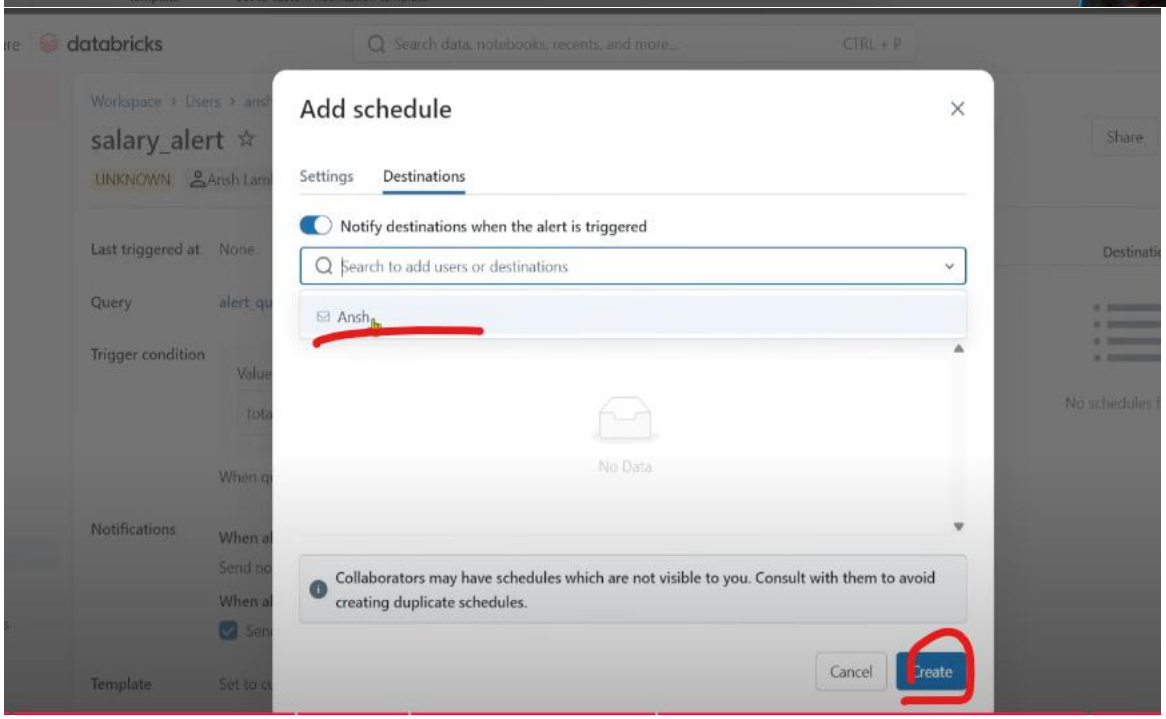

When alert returns back to normal
Send notification
Template
Use custom template
Subject / Body
Preview
He Bro, Your employees are earning more than you
Guide: Custom alert notifications
Formatting guide
ring & Alerts with Databricks


databricksrt - Microsoft Azure
New Alert - Databricks
What are Databricks SQL alerts
Account Console - Databricks
learn.microsoft.com/en-in/azure/databricks/sql/user/alerts/#create-an-alert
Filter by title
Azure Databricks Documentation
Sign up for a free trial
Get started tutorials
What is Databricks?
Architecture
Workspace UI
Data guides
Data engineering
AI and machine learning
Data warehousing
What is data warehousing on Azure Databricks?
Get started
SQL editor
Queries
Visualizations overview
Databricks SQL dashboards
Alerts
Alerts overview
Use default template: Alert notification is a message with links to the Alert configuration screen and the Query screen.
Use custom template: Alert notification includes more specific information about the alert.
a. A box displays, consisting of input fields for subject and body. Any static content is valid, and you can incorporate built-in template variables:
ALERT_STATUS: The evaluated alert status (string).
ALERT_CONDITION: The alert condition operator (string).
ALERT_THRESHOLD: The alert threshold (string or number).
ALERT_COLUMN: The alert column name (string).
ALERT_NAME: The alert name (string).
ALERT_URL: The alert page URL (string).
QUERY_NAME: The associated query name (string).
QUERY_URL: The associated query page URL (string).
QUERY_RESULT_TABLE: The query result HTML table (string). Results are limited to the first 100 rows. Only email notification destinations can render HTML.
QUERY_RESULT_VALUE: The query result value (string or number).
QUERY_RESULT_ROWS: The query result rows (value array).
QUERY_RESULT_COLS: The query result columns (string array).
An example subject, for instance, could be: Alert "{{ALERT_NAME}}" changed status to {{ALERT_STATUS}}.
b. You can use HTML to format messages in a custom template. Only email notification
Training
Module
Configure alerts and responses - Training
In this module, you learn how Azure Monitoring alerts proactively notifies you when Azure Monito data indicates there might be a problem with you
Certification
Microsoft Certified: Azure Database Administrat Associate - Certifications
Administer an SQL Server database infrastructure cloud, on-premises and hybrid relational databas using the Microsoft PaaS relational database...
Documentation
Update to the latest Databricks SQL API version Azure Databricks - Databricks SQL
Learn about the changes in the latest version of t Databricks SQL REST API.
Manage notification destinations - Azure Databr
Learn how to configure notification d your workspace.
Query history - Azure Databr

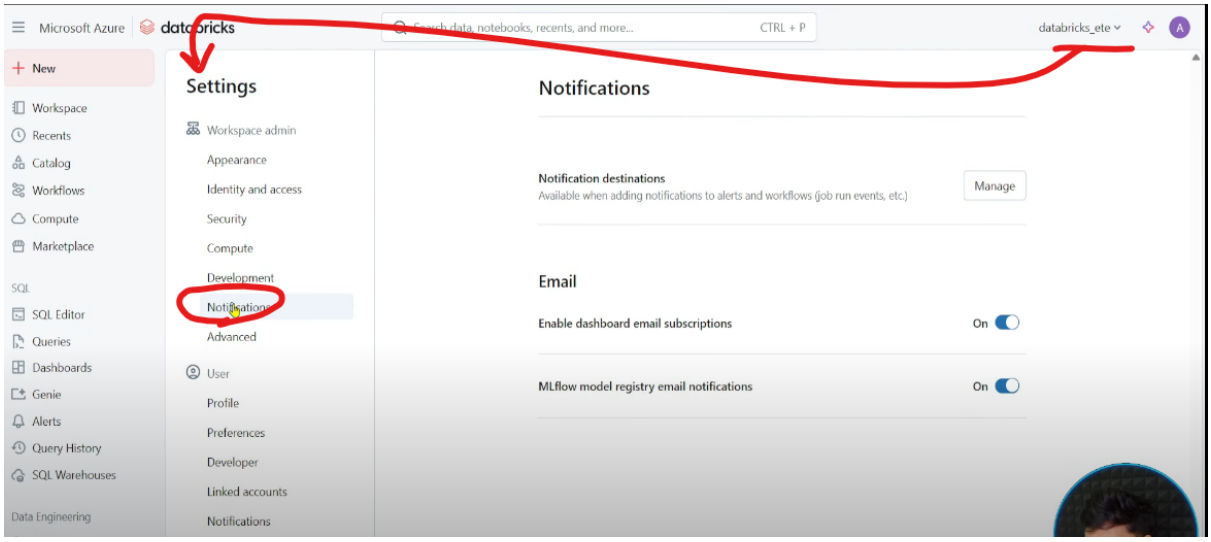
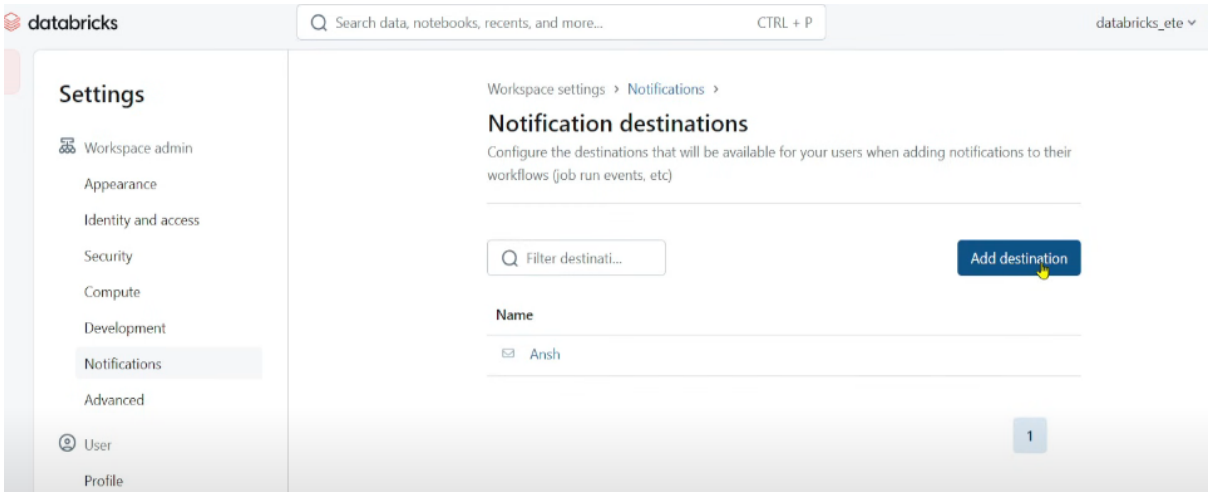
Send notification
Template
Use custom template
Subject / Body
Preview
He Bro, Your employees are earning more than you
Hey this is the Alert Status : {{ALERT_STATUS}}
Formatting guide
& Alerts with Databricks


Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_ete
New
Workspace
Recents
Catalog
Workflows
Compute
Marketplace
SQL
SQL Editor
Queries
Dashboards
Genie
Alerts
Query History
SQL Warehouses
Data Engineering
Workspace › Users › anshlambaaz@gmail.com
salary_alert
UNKNOWN
Ansh Lamba
Share
Edit
Refresh
Last triggered at
None
Query
alert_query_new
Trigger condition
Value column
total_salary
First row
Operator
>
Threshold value
3000
When query result has no rows, set state to OK
Schedules
Destinations
No schedules found
Add schedule
Notifications
When alert is triggered
Send notification just once, until back to normal.
When alert returns back to normal
Send notification
Template
Set to custom notification template.





If destination is not there by default, create one



Manage → Add destination (email id)