

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>`

1_realtime Python ▾ Tabs: OFF ▾ ☆
File Edit View Run Help Last edit was now ▶ Run all

Catalog

Type to search...

For you All

- My organization
 - system
 - _databricks_internal
 - databricks_cata
 - databricksrt
 - default
 - information_schema
 - schema
 - vol0**
 - main

Creating a Volume

Just now (3s) 2

```
%sql
CREATE EXTERNAL VOLUME databricksrt.schema.vol0
LOCATION 'abfss://databricksrt@databricksete.dfs.core.windows.net/volume_folder'
```

See performance (1)

3

Start typing or generate with AI (Ctrl + I)...

Upload the file in ADLS Gen2

Home > RG_Databricks_ETE > databricksete | Containers >

databricksrt Container

Search

Upload + Add Directory Refresh | Rename Delete Change tier Acquire lease Break lease Give feedback

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: databricksrt / volume_folder

Search blobs by prefix (case-sensitive) ☐ Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[.]						...
MegaMart.csv	11/5/2025, 10:59:35 ...	Hot (Inferred)		Block blob	71.77 KB	Available

Successfully uploaded blob(s)
Successfully uploaded 1 blob(s).

Just now (5s) 3

```
%sql
SELECT * FROM csv.`/Volumes/databricksrt/schema/vol0/MegaMart.csv`
```

See performance (1) Optimize

_sqldf: pyspark.sql.connect.dataframe.DataFrame = [_c0: string, _c1: string ... 8 more fields]

	_c0	_c1	_c2	_c3	_c4	_c5	_c6	_c7	_c8
1	order_id	user_id	order_date	product_id	product_catego...	product_name	quantity	price_per_u...	payme
2	1001	U188	2025-04-20	P940	Fashion	Sneakers	2	58.53	PayPal
3	1002	U062	2025-04-16	P794	Fashion	T-Shirt	3	83.76	UPI
4	1003	U058	2025-04-18	P326	Fashion	Sunglasses	2	78.85	PayPal
5	1004	U011	2025-04-10	P574	Fashion	Sunglasses	5	46.49	PayPal

Just now (5s) 3

```
%sql
SELECT * FROM csv.`abfss://databricksrt@databricksete.dfs.core.windows.net/volume_folder`
```

See performance (1) Optimize

_sqldf: pyspark.sql.connect.dataframe.DataFrame = [_c0: string, _c1: string ... 8 more fields]

	_c0	_c1	_c2	_c3	_c4	_c5	_c6	_c7	_c8
1	order_id	user_id	order_date	product_id	product_catego...	product_name	quantity	price_per_u...	payme
2	1001	U188	2025-04-20	P940	Fashion	Sneakers	2	58.53	PayPal
3	1002	U062	2025-04-16	P794	Fashion	T-Shirt	3	83.76	UPI
4	1003	U058	2025-04-18	P326	Fashion	Sunglasses	2	78.85	PayPal

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

Scenario-2

Looping through files in folder

Home > RG_Databricks_ETE > databricksete | Containers >

databricksrt ...

Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Authentication method: Access key (Switch to Microsoft Entra user account)

Location: databricksrt / source / loopsource

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[]						...
raw_data_first.csv	11/5/2025, 11:24:36 ...	Hot (Inferred)		Block blob	10.06 KGB	Available
raw_data_second.csv	11/5/2025, 11:25:20 ...	Hot (Inferred)		Block blob	3.5 KIB	Available

2_realtime Python Tabs: OFF ☆

File Edit View Run Help

file_name
second

Workspace Databricks RealTime

- 1_realtime
- 2_realtime
- 3_realtime
- 4_realtime
- 5_realtime
- 6_realtime
- input_notebook

Data Reading Markdown

```

dbutils.widgets.text("file_name", "")

file_var = dbutils.widgets.get("file_name")

df = spark.read.format("csv")\
    .option("header", "true")\
    .load(f"abfss://databricksrt@databricksete.dfs.core.windows.net/source/loopsource/raw_data_{file_var}.csv")

df: pyspark.sql.connect.dataframe.DataFrame = [Branch_ID: string, Dealer_ID: string ... 10 more fields]

df.write.format("parquet")\
    .option("path", f"abfss://databricksrt@databricksete.dfs.core.windows.net/source/loopsink/raw_data_{file_var}")\
    .save()

```

input_notebook Python Tabs: OFF ☆

File Edit View Run Help

Workspace Databricks RealTime

- 1_realtime
- 2_realtime
- 3_realtime
- 4_realtime
- 5_realtime
- 6_realtime
- input_notebook

```

input_array = [
    {"file_name": "first"},
    {"file_name": "second"}
]

dbutils.jobs.taskValues.set(key = "my_output", value = input_array)

```

Create a job in Workflows

Microsoft Azure databricks Search data, notebooks, recent, and more... CTRL + P databricks_ete

New

- Workspace
- Recents
- Catalog
- Workflows**
- Compute
- Marketplace
- SQL
- SQL Editor
- Queries
- Dashboards
- Genie
- Alerts

Workflows

Jobs & pipelines Job runs

Filter by name or ID sub...

All Jobs Pipelines Owned by me Accessible by me Favorites Tags Run as

Create

- Job
- Orchestrate Pipelines, Notebooks, Queries, Models and more
- ETL pipeline
- Build and run an ETL pipeline with SQL and Python
- Ingestion pipeline
- Ingest data from external sources in just a few clicks

Name	Type	Tags	Run as
Gold_Products	Pipeline		Ansh Lam
test	Pipeline		Ansh Lam

Job : Normal Workflows

ETL pipeline : DLT

Ingestion pipeline : No code ingestion pipeline

New Job May 11, 2025, 11:37 AM ☆

Runs Tasks



Task name* ⓘ

DataReading

Type*

Notebook

Inputs* ⓘ

{{tasks.Input.values.my_output}}

Concurrency (optional) ⓘ

Depends on

Input X

New Job May 11, 2025, 11:37 AM ☆

Runs Tasks



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

The screenshot shows the Databricks Catalog Explorer interface. On the left sidebar, the 'Catalog' icon is circled in red. In the main panel, the 'External Data' tab is also circled in red. Below the tab, there is a search bar and a list of catalogs. The 'Quick access' section shows 'Recents', 'Favorites', and 'Catalogs' buttons. A table lists the following catalogs:

Name	Last viewed	Type
schema.databricksrt	1 hour ago	Schema
databricksrt	1 hour ago	Catalog
sparksql_cata	6 days ago	Catalog
order_man	6 days ago	Table

Connections → Create Connection

The screenshot shows the 'External Data' page with the 'Connections' tab selected and circled in red. Below the tab, there is a search bar and a table showing one connection:

Type	Name	URL	Created at	Owner	Comment
my_ms_sql_server	my_ms_sql_server	jdbc://127.0.0.1:1433/	2025-05-03 14:11:27	anshlambaaz@gmail.com	

A 'Create connection' button is circled in red in the top right corner of the main panel.

Connections >

Set up connection

1 Connection basics

2 Authentication

3 Catalog basics

4 Access

5 Metadata

Step 1

Connection basics

Connection name*

new_connection

Connection type*

MySQL

Comment

|

There are crazy amount of sources in Connection Type

1 Connection basics

2 Authentication

3 Catalog basics

4 Access

5 Metadata

Step 2

Authentication

Host*

Host name of the foreign server without scheme (i.e. no 'jdbc://' or 'https://' prefix).

host.domain.com

Port

Port of the foreign mysql instance, default to 3306.

3306

User*

User identity used to access the foreign instance.

username

Password*

Password of the foreign instance.

password123

☐ Trust server certificate

Lakehouse Federation >

Cancel

Back

Create

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

```
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]

```
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()
```

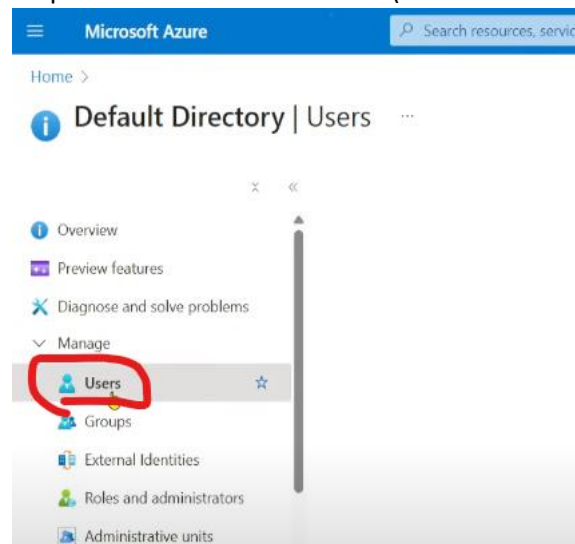
a714b212-424e-4392-af55-c63231e7c85a Last updated: 10 days ago

connect streaming query StreamingQuery at 0x7f6746c08850

Scenario – 5 (UC access control)

Granting access to new user

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



Home > Default Directory | Users > Users >

Create new user

Create a new internal user in your organization

Basics Properties Assignments Review + create

Create a new user in your organization. This user will have a user name like alice@contoso.com. [Learn more](#)

Identity

User principal name * @ [Domain not listed? Learn more](#)

Mail nickname * ☒ Derive from user principal name

Display name *

Password * ☒ Auto-generate password

Account enabled ☒

Review + create

< Previous

Next: Properties >

Home > Default Directory | Users >

Users

Default Directory

+ New user Edit (Preview) Delete Download users Bulk operations Refresh Manage view Per-user MFA

Azure Active Directory is now Microsoft Entra ID.

Search Add filter

9 users found

	Display name	User principal name	User type	On-premises sy...	Identities	Company name
<input type="checkbox"/>	Ansh Lamba	anshlambaaz_gmail.com#...	Member	No	MicrosoftAccount	
<input type="checkbox"/>	databricksunity	databricksunity@anshlam...	Member	No	anshlambaazgmail.onmicrosoft...	
<input type="checkbox"/>	fabricde	fabricde@anshlambaazg...	Member	No	anshlambaazgmail.onmicrosoft...	
<input type="checkbox"/>	fabrictut	fabrictut@anshlambaazg...	Member	No	anshlambaazgmail.onmicrosoft...	
<input type="checkbox"/>	her	her@anshlambaazgmail.o...	Member	No	anshlambaazgmail.onmicrosoft...	
<input type="checkbox"/>	newuser	newuser@anshlambaazg...	Member	No	anshlambaazgmail.onmicrosoft...	

Step2: Add user in Databricks workspace

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P databricks_ete

+ New Workspace Recents Catalog Workflows Compute Marketplace iSQL SQL Editor Queries Dashboards Genie Alerts

Catalog sql_compute Serverless 2XS Type to search...

My organization

- system
- __databricks_internal
- databricks_cata
- databricksrt
- main
- my_ms_sql_server_catalog
- sparksql_cata
- Delta Shares Received
- samples
- Legacy
- hive_metastore

Quick access

Recents Favorites Catalogs

Name

- schema databricksrt
- databricksrt
- sparksql_cata
- order_man sparksql_cata.sparksql_schema
- my_ms_sql_server_catalog

Search workspaces

- databricks_ete uksoth
- adb-ansh-unity eastus
- ADBmodern eastus
- adbolympics canadacentral
- deltalakeADBWS
- Manage account

User management

Users Service principals Groups

Add users to the account. Account users can use the account console to view and connect to their workspaces. Account admins can perform all of the management functions available in the account console. [Learn more](#)

Filter users ☐ Only account admins 4 total

Status	Name	Email	Source	Roles
✓	Ansh Lamba	anshlambaaz@gmail.com	Account ⓘ	Account admin
✓	Ansh Lamba	anshlambaaz_gmail.com#ext#@anshla...	Account ⓘ	Account admin
✓	databricksunity	databricksunity@anshlambaazgmail.on...	Account ⓘ	Account admin
✓	New User	newuser@anshlambaazgmail.onmicros...	Account ⓘ	

User management

Users Service principals Groups

Add users to the account. Account users can use the account console to view and connect to their workspaces. Account admins can perform all of the management functions available in the account console. [Learn more](#)

Filter users ☐ Only account admins 5 total

Status	Name	Email	Source	Roles
✓	Ansh Lamba	anshlambaaz@gmail.com	Account ⓘ	Account admin
✓	Ansh Lamba	anshlambaaz_gmail.com#ext#@anshla...	Account ⓘ	Account admin
✓	databricksunity	databricksunity@anshlambaazgmail.on...	Account ⓘ	Account admin
✓	her her	her@anshlambaazgmail.onmicrosoft.c...	Account ⓘ	
✓	New User	newuser@anshlambaazgmail.onmicros...	Account ⓘ	

Now login through her id

Microsoft Azure databricks

https://adb-2972879960737553.13.azuredatabricks.net/explore/data?o=2972879960737553

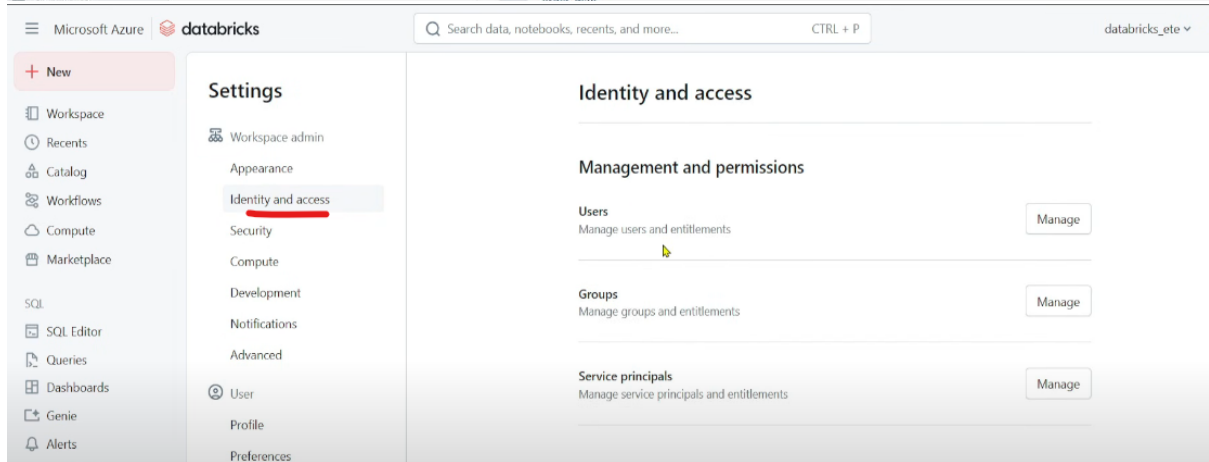
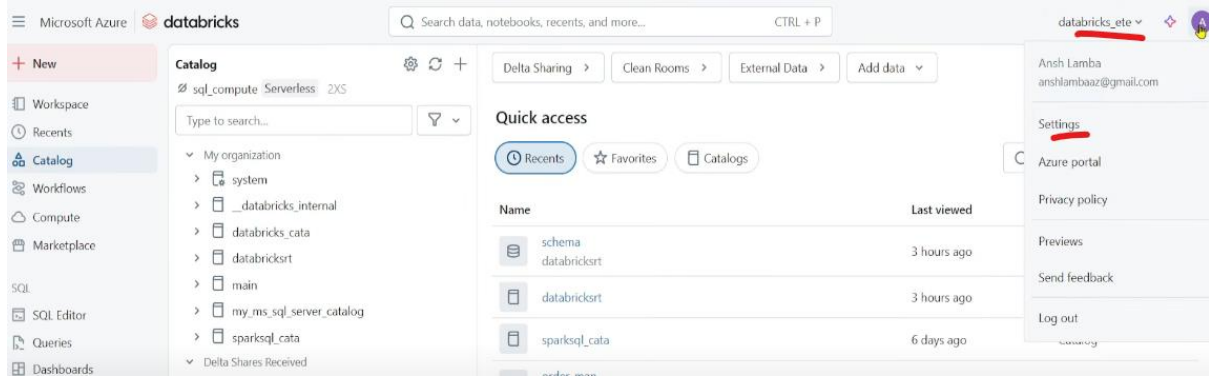
Unable to view page

You do not have permission to access this page in workspace 2972879960737553. Please contact your workspace administrator to ensure you are assigned to the workspace and have either the "Workspace access" or the "Databricks SQL access" entitlement, which is required to access this page.

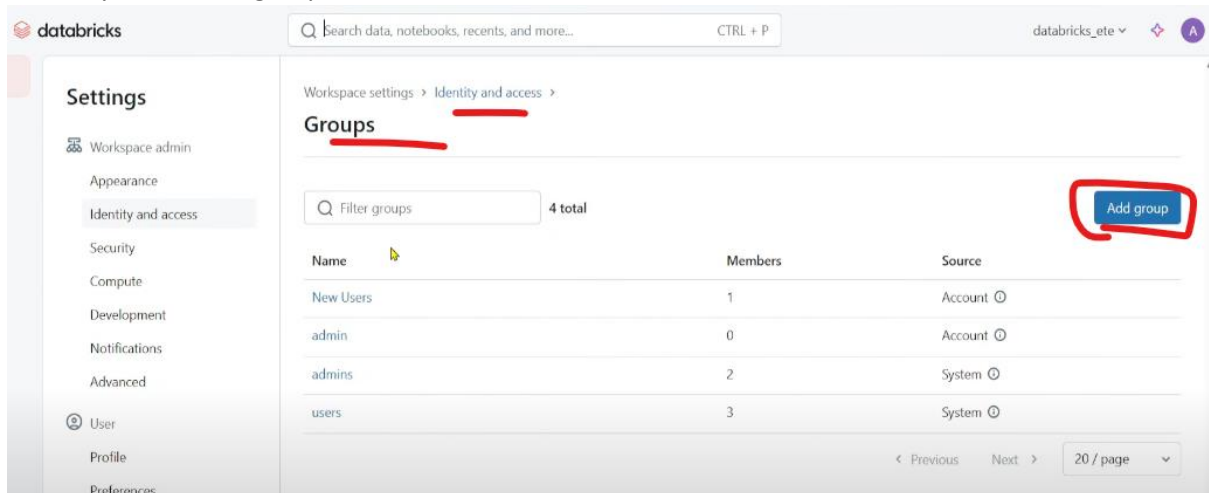
If you believe you have received this message an error, please contact Databricks support.

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



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 ▾

Settings

Workspace admin

Appearance

Identity and access

Security

Compute

Development

Notifications

Advanced

User

Profile

Preferences

Successfully added herGroup to the workspace

Group details

herGroup

Delete

Group Information Members Entitlements Permissions Preview Parent groups

Add members

Name

Email/ID



No group members found

Settings

Workspace admin

Appearance

Identity and access

Security

Compute

Development

Notifications

Advanced

User

Profile

Preferences

Successfully added her@anshlambaazgmail.onmicrosoft.com to herGroup.

Group details

herGroup

Delete

Group Information Members Entitlements Permissions Preview Parent groups

Add members

Name

Email/ID



her

her@anshlambaazgmail.onmicrosoft.com

Group details

herGroup

Delete

Group Information Members Entitlements Permissions Preview Parent groups

Name	Enabled
Allow unrestricted cluster creation ⓘ	<input type="checkbox"/>
Databricks SQL access	<input type="checkbox"/>
Workspace access	<input type="checkbox"/>

Workspace settings > Identity and access > Groups >

Group details

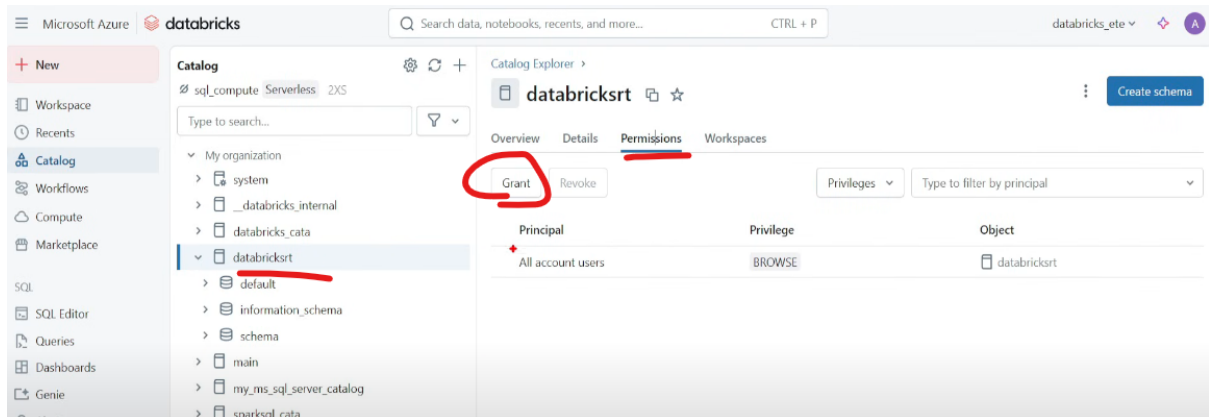
herGroup

Group Information Members Entitlements Permissions Preview Parent groups

Name	Enabled
Workspace access	<input checked="" type="checkbox"/>
Databricks SQL access	<input checked="" type="checkbox"/>
Allow unrestricted cluster creation ⓘ	<input type="checkbox"/>

Still she will not be having access to Catalogs

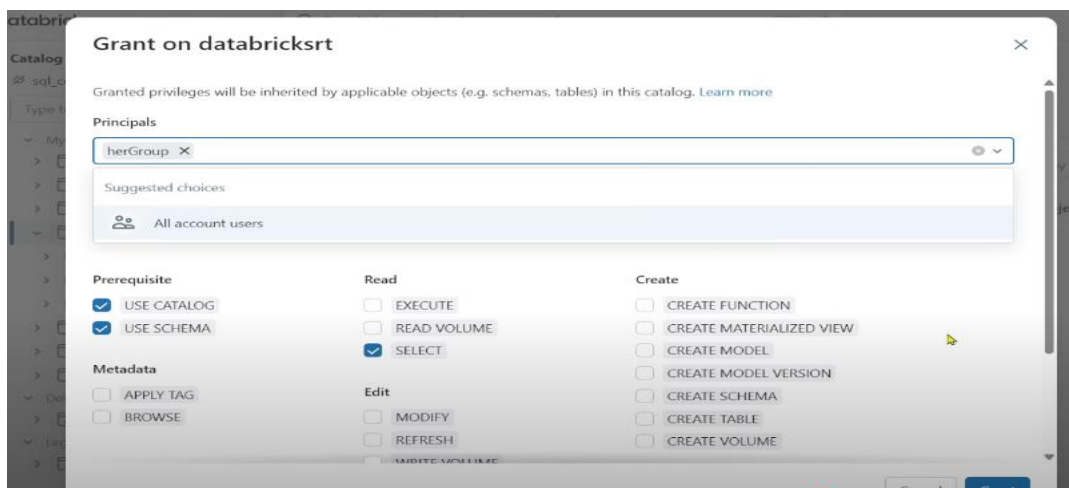
Step4: Granting access to catalogs

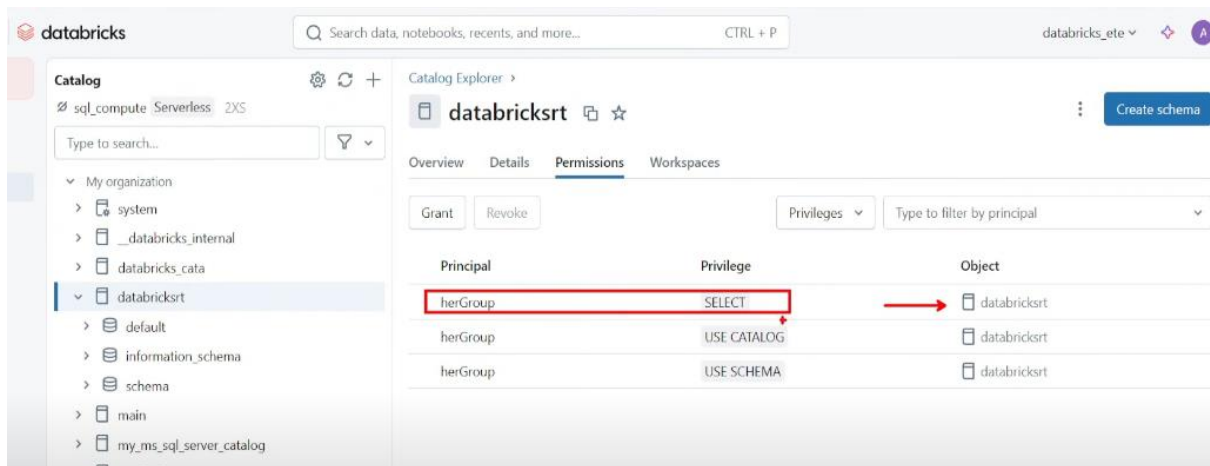


There are 2 types of access:

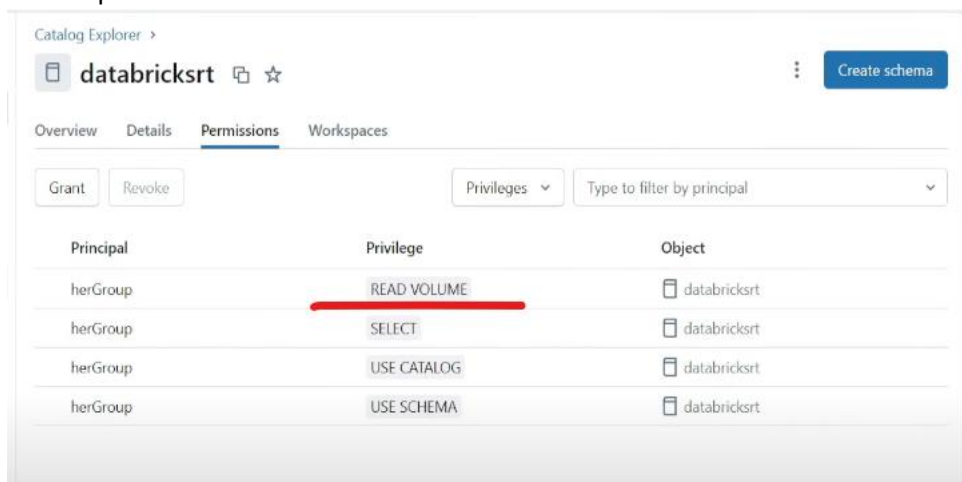
- 1) Top to Down : having access to parent container will automatically having access to all child containers
- 2) Down to Top : having access to a container means all kind of read/write permissions on the container but only usage access to its parent container (if parent container contains any other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

Top to Down



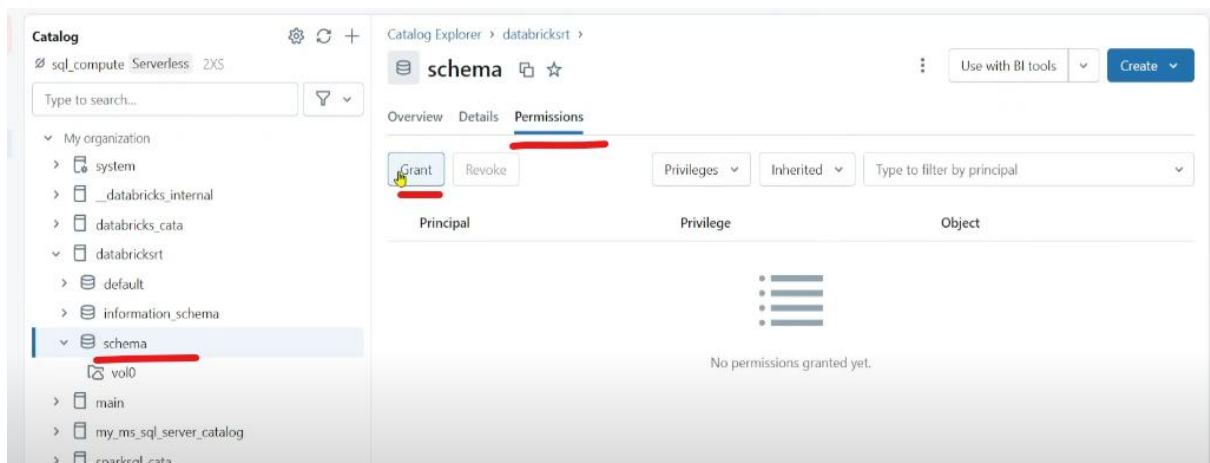


Add respective access

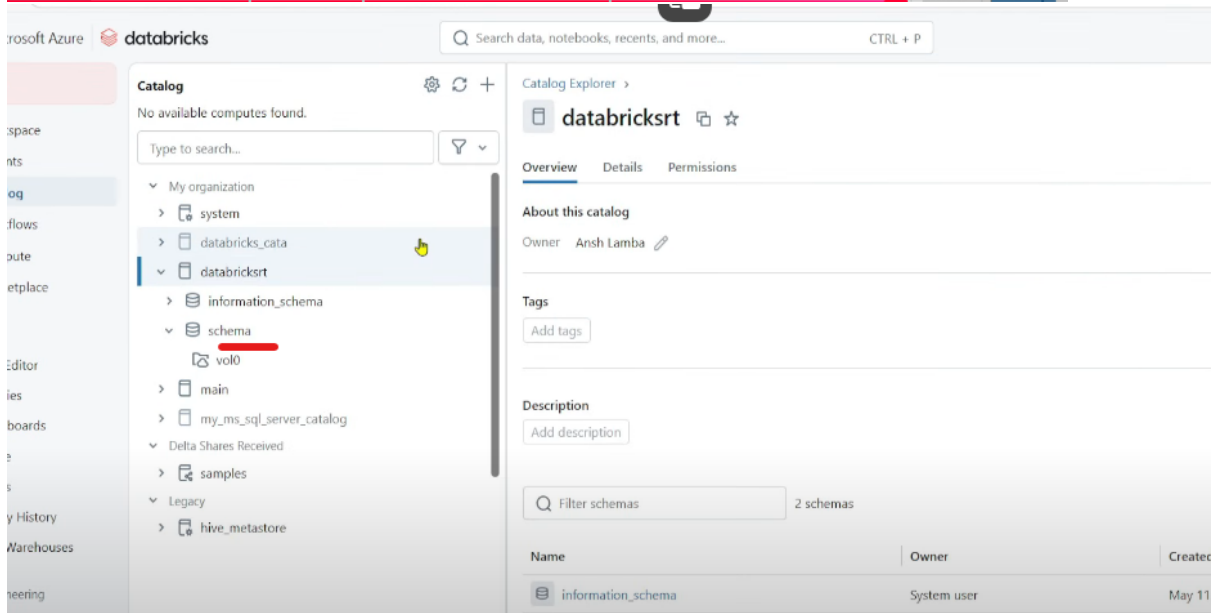
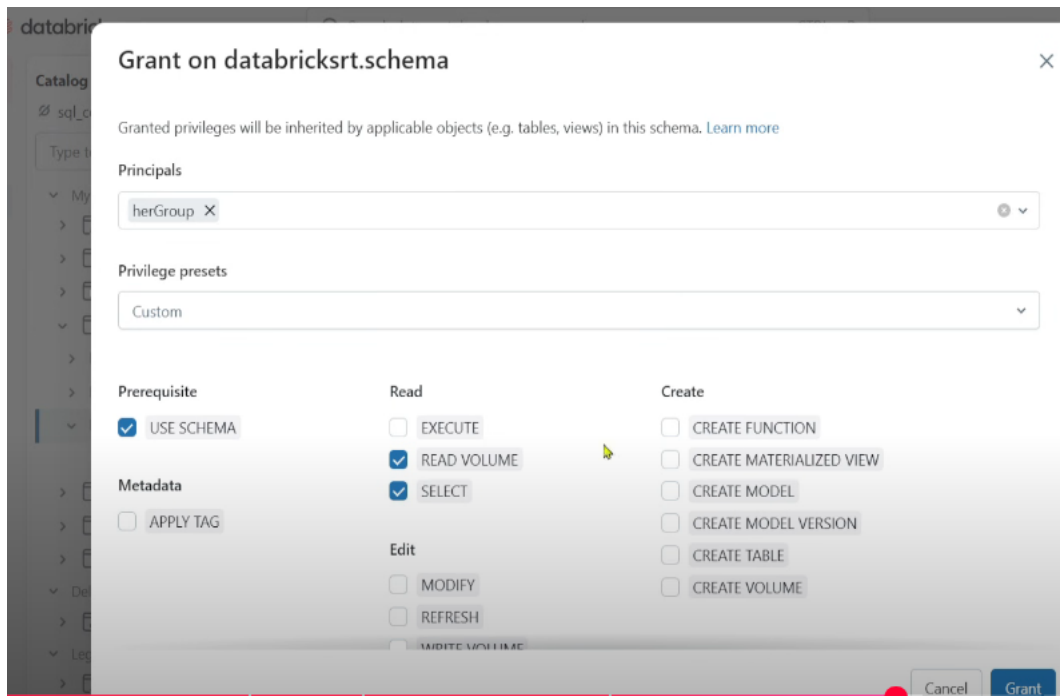


Down to Top

Lets delete all the access for now



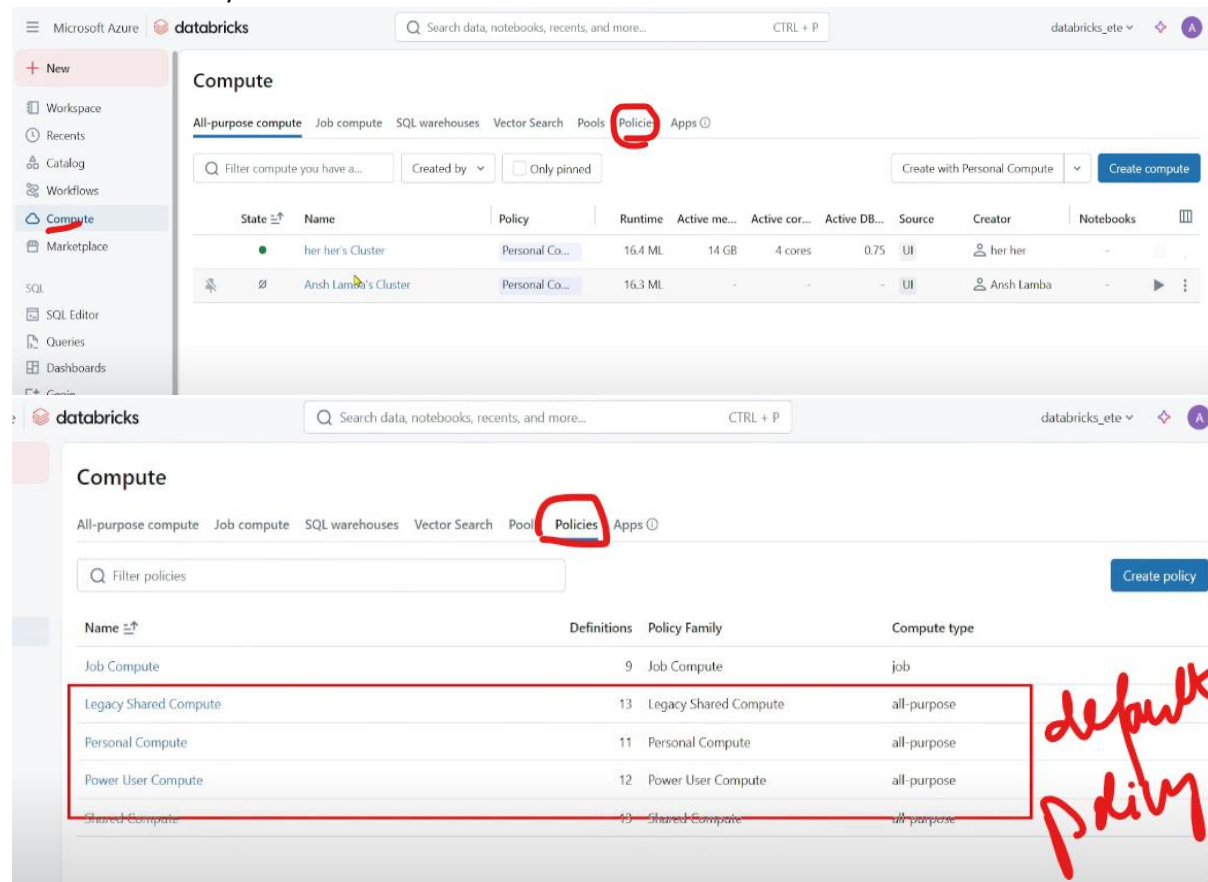
Go one level down i.e. schema and give permission



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.

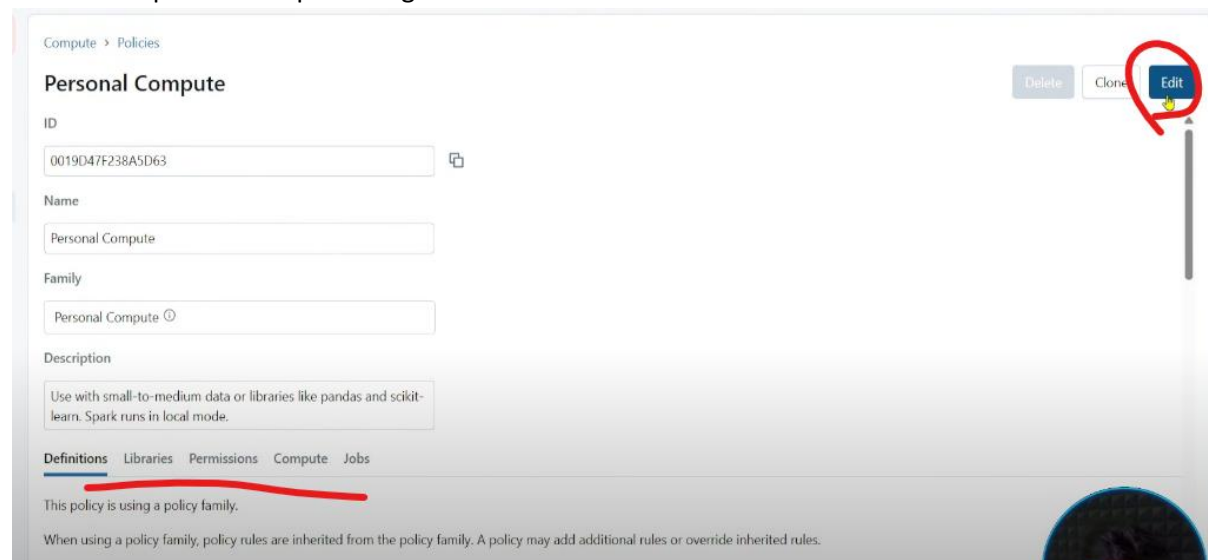


The image shows two screenshots of the Databricks interface. The top screenshot shows the 'Compute' page with the 'Policies' tab selected. The bottom screenshot shows the 'Policies' page with a table of pre-defined policies. A red box highlights the 'Legacy Shared Compute', 'Personal Compute', and 'Power User Compute' policies. A red arrow points to the 'Edit' button in the top right corner of the 'Personal Compute' policy details.

Compute Policies Table:

Name	Definitions	Policy Family	Compute type
Job Compute	9	Job Compute	job
Legacy Shared Compute	13	Legacy Shared Compute	all-purpose
Personal Compute	11	Personal Compute	all-purpose
Power User Compute	12	Power User Compute	all-purpose
Shared Compute	10	Shared Compute	all-purpose

Policies are pre defined pre configured area



The image shows the 'Personal Compute' policy details page. The 'Edit' button is circled in red. The 'Definitions' tab is selected, showing the policy's configuration.

Personal Compute Policy Details:

- ID: 0019D47F238A5D63
- Name: Personal Compute
- Family: Personal Compute
- Description: Use with small-to-medium data or libraries like pandas and scikit-learn. Spark runs in local mode.

Definitions Tab:

This policy is using a policy family.

When using a policy family, policy rules are inherited from the policy family. A policy may add additional rules or override inherited rules.

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

Compute > Policies

Create policy

Name

Ansh Policy

Family

Custom

Description

Optional

Definitions Libraries Permissions

How to define a policy?

```
1 {
2   "node_type_id" : {
3     "type" : "allowlist",
4     "values" : [ "Standard_DS3_v2" ],
5     "defaultValue" : "Standard_DS3_v2"
6   },
7   "spark_version" : {
```

Definitions Libraries Permissions

This policy is using a policy family.

When using a policy family, policy rules are inherited from the policy family. A policy may add additional rules or override inherited rules.

0 overrides Edit

```
1 {
2   "node_type_id" : {
3     "type" : "allowlist",
4     "values" : [ "Standard_DS3_v2", "Standard_DS4_v2", "Standard_DS5_v2", "Standard_NC4as_T4_v3" ],
5     "defaultValue" : "Standard_DS3_v2"
6   },
7   "spark_version" : {
8     "type" : "unlimited",
9     "defaultValue" : "auto:latest-ml"
10  },
11  "runtime_engine" : {
12    "type" : "fixed"
```

remove

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

```
41   "value" : "ON_DEMAND_AZURE",
42   "hidden" : true
43 },
44 "spark_conf.spark.databricks.cluster.profile" : {
45   "type" : "fixed",
46   "value" : "singleNode",
47   "hidden" : true
48 },
49 "autotermination_minutes" : {
50   "type" : "unlimited",
51   "defaultValue" : 4320,
52   "isOptional" : true
53 }
54 }
```

Changing 4320 mins to 10 mins

Ansh Policy

Family

Custom

Description

Optional

Definitions Libraries **Permissions**

Max compute resources per user

Unlimited

NAME	PERMISSION
Admins	Can Use
herGroup	Can Use

Now try creating a Cluster, all other policies are disabled

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

Compute

All-purpose compute Job compute SQL warehouses Vector Search Pools Apps

Filter compute you have access to Created by Only pinned

State	Name	Policy	Runtime	Active mem...	Active cores	Active DBU ...	Source	Creator
Stopped	her her's Cluster	Personal Compute	16.4 ML	-	-	0.75	UI	her her

Create with Personal Compute

Create with Ansh Policy

Create with Power User Compute

Create with Shared Compute

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

Job Runs

Data Ingestion

Compute > New compute > Simple form: OFF

her her's Ansh Policy Cluster

Single user or group access

her her

Performance

Databricks runtime version

Runtime: 16.4 LTS ML (Scala 2.12, Spark 3.5.2)

Node type

Standard_DS3_v2 14 GB Memory, 4 Cores

General purpose

Standard_DS3_v2 14 GB Memory, 4 Cores

Tags

Add tags

Key Value Add

> Automatically added tags

can't change

Performance

Databricks runtime version

Runtime: 16.4 LTS ML (Scala 2.12, Spark 3.5.2)

Node type

Standard_DS3_v2 14 GB Memory, 4 Cores

☒ Terminate after 9 minutes of inactivity Interval must be between 10 and 43200 minutes

Tags

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

Account

Workspaces

Catalog

Usage

User management

Cloud resources

Previews

Settings

Settings

User provisioning Token report App connections Feature enablement Language settings My preferences Security and compliance Account settings

Some features require you to explicitly opt-in before you can use them. Use this page to enable and configure the following account features.

Changes made to the settings on this page affect all workspaces in your Microsoft Entra ID tenant.

Personal Compute Enable for all

Enables all users to create single-machine compute using the Personal Compute default policy. Learn more

Web Terminal Enforce On

Grants all users access to the web terminal, where they can interact with the command line on their all-purpose compute resources. Learn more

Enable Admin Protection for "No Isolation Shared" Clusters Disabled

Enabling admin protection prevents automatic generation of Databricks API credentials for Databricks admins on "No Isolation Shared" clusters. Note that enabling admin protection may prevent some workloads from being run as an admin user on "No Isolation Shared" clusters. Learn more

Settings

User provisioning Token report App connections **Feature enablement** Language settings My preferences



Access control for the Personal Compute default policy is delegated to the ACLs of the Personal Compute default policy.

Some features require you to explicitly opt-in before you can use them. Use this page to enable and configure the following account features.

Changes made to the settings on this page affect all workspaces in your Microsoft Entra ID tenant.



Personal Compute

Enables all users to create single-machine compute using the Personal Compute default policy. [Learn more](#)

Delegate  

Web Terminal

Grants all users access to the web terminal, where they can interact with the command line on their all-purpose compute resources. [Learn more](#)

Enforce  On 

Microsoft Azure databricks

Search data, notebooks, recent, and more... CTRL + P databricks_ete

Compute

All-purpose compute Job compute SQL warehouses Vector Search Pools Apps

Filter compute you have access to Created by Only pinned Create with Ansh Policy

State	Name	Policy	Runtime	Active mem...	Active cores	Active DBU ...	Source	Creator	Notebooks
her her's Cluster		Personal Compute	16.4 ML	-	-	-	UI	her her	-

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

databricks

Search data, notebooks, recent, and more... CTRL + P databricks_ete

Workspace

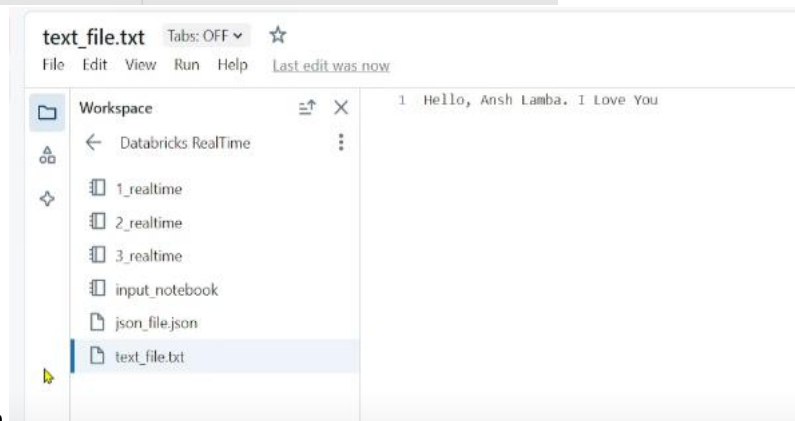
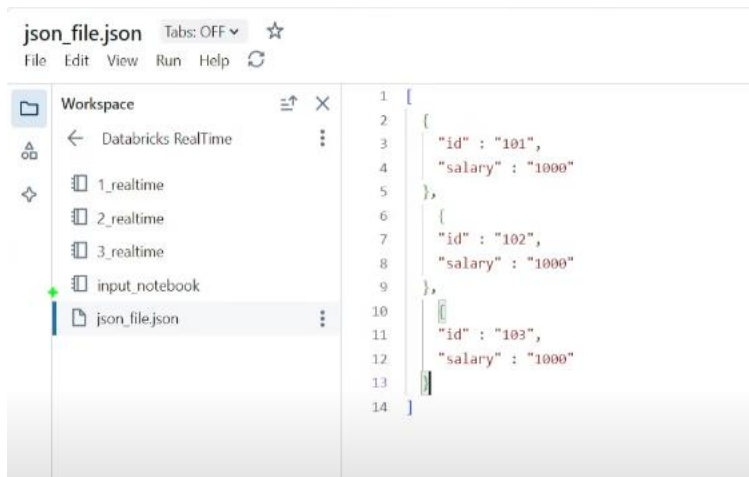
Workspace

Databricks RealTime

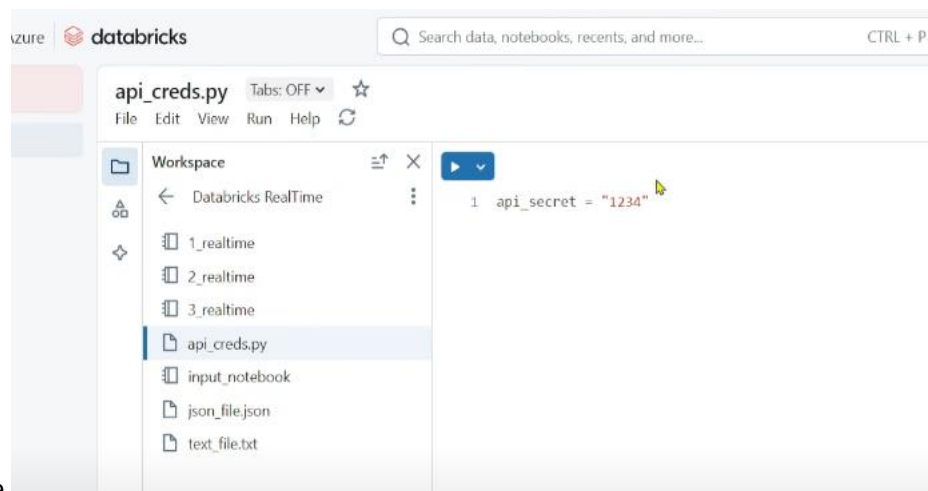
Name	Type	Owner	Created	File
1_realtime	Notebook	Ansh Lamba	May	Folder
2_realtime	Notebook	Ansh Lamba	May	Git folder
3_realtime	Notebook	Ansh Lamba	May	Notebook
input_notebook	Notebook	Ansh Lamba	May	File

Send feedback Share Create

- Folder
- Git folder
- Notebook
- File
- Query
- Dashboard
- Genie space
- Alert
- MLflow experiment



Create a text file



Create a py file

Manipulating Text File



Create a notebook
r means read mode

Writing new line in the file

```
Just now (2s) 3
f = open("/Workspace/Databricks RealTime/text_file.txt","a")
f.write("This is a new line")
18
```

a means append

```
Just now (1s) 2
f = open("/Workspace/Databricks RealTime/text_file.txt","r")
f.read()
'Hello, Ansh Lamba. I Love You This is a new line'
```

Reading Python File

```
May 12, 2025 (<1s) 5
# Importing .py file as a module
import api_creds

# Using the variable created inside the .py file
api_creds.api_secret

'1234'
```

Reading JSON Data

```
May 12, 2025 (1s) 7
import pandas as pd

df = pd.read_json("/Workspace/Databricks RealTime/json_file.json")
df
```

df: pandas.core.frame.DataFrame = [id: int64, salary: int64]

	id	salary
0	101	1000
1	102	1000
2	103	1000

But possible with Delta Sharing

The screenshot shows the Databricks Catalog Explorer interface. The top navigation bar includes 'Delta Sharing', 'Clean Rooms', 'External Data', and 'Add data'. The 'Delta Sharing' tab is highlighted with a red box. The main content area displays a list of catalogs under 'My organization' and 'Delta Shares Received'. The 'Quick access' section on the right shows a list of recent catalogs with columns for Name, Last viewed, and Type.

Name	Last viewed	Type
schema databricksrt	1 hour ago	Schema
databricksrt	1 hour ago	Catalog
sparksql_cata	6 days ago	Catalog
order_man sparksql_cata.sparksql_schema	7 days ago	Table
my_ms_sql_server_catalog	8 days ago	Catalog
databricks.cata	8 days ago	Catalog

The screenshot shows the Databricks Delta Sharing page. The 'Shared with me' tab is selected. The page displays a list of providers with columns for Name, Authentication type, Owner, Created at, and Comment. A red box highlights the 'Import data' button.

Name	Authentication type	Owner	Created at	Comment
azure:uksouthebd8150b-e6a1-4043-b2b8-f9d4b9b8d7da	DATABRICKS	anshlambaa@gmail.com#ext...	Apr 26, 2025, 03:33 PM	

Catalogs → Delta Sharing → create New recipient

The screenshot shows the Databricks Delta Sharing page with the 'Shared by me' tab selected. The 'Shares' and 'Recipients' tabs are visible. A red box highlights the 'New recipient' button. The page displays a list of shares with columns for Name, Created at, Owner, and Comment.

Name	Created at	Owner	Comment
first_share	May 03, 2025, 12:...	anshlambaa@gmail.com	

Create a new recipient

A recipient represents an organization with whom to share data. Grant the recipient access to shares after they are created.

Recipient name*

new_dlt_ws

Recipient type

Databricks

The recipient is a Databricks user and intends to use Delta Sharing on the Databricks platform.

Open

The recipient is not a Databricks user or intends to use Delta Sharing on an external platform.

Sharing identifier (Databricks recipient only)

If the recipient is a Databricks user, enter their sharing identifier. [Learn more](#)

Your current org name will appear as 'azure:uksouth:24974dbe-653b-46c2-9c29-acb936781711' on the recipient/provider side, please consider updating it [here](#).

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

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

Catalog Explorer >

Delta Sharing

Shared with me Shared by me

Create shares to share data assets across organizations, and manage access to shares with recipients.

Shares Recipients Filter shares

New recipient Share data

Name	Created at	Owner	Comment
first_share	May 03, 2025, 12...	anshlambaaz@gmail.com	

azure:uksouth:24974dbe-653b-46c2-9c29-acb936781711

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

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

Alt...

Catalog Explorer >

sql_compute Serverless 2XS

Type to search...

My organization

- system
- _databricks_internal
- databricks_cata**
- bronze
- default
- gold
- information_schema
- silver
- testschema
- databricksrt

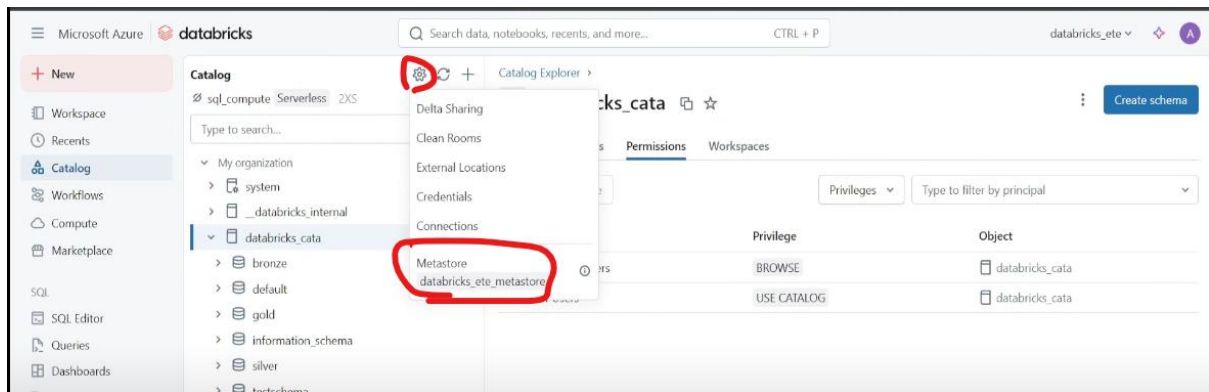
databricks_cata

Create schema

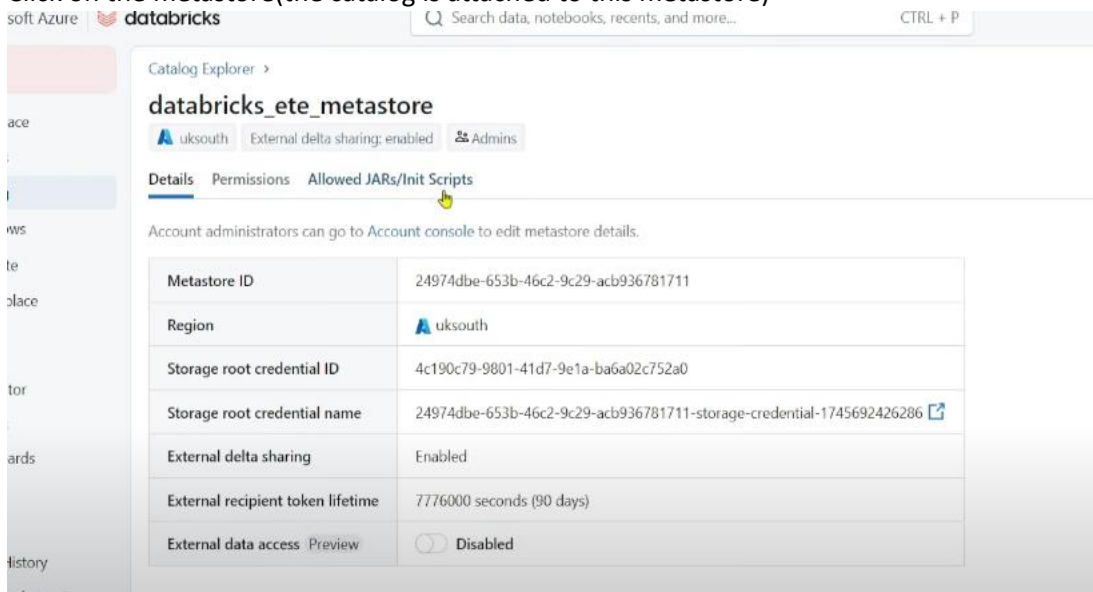
Overview Detail **Permissions** Workspaces

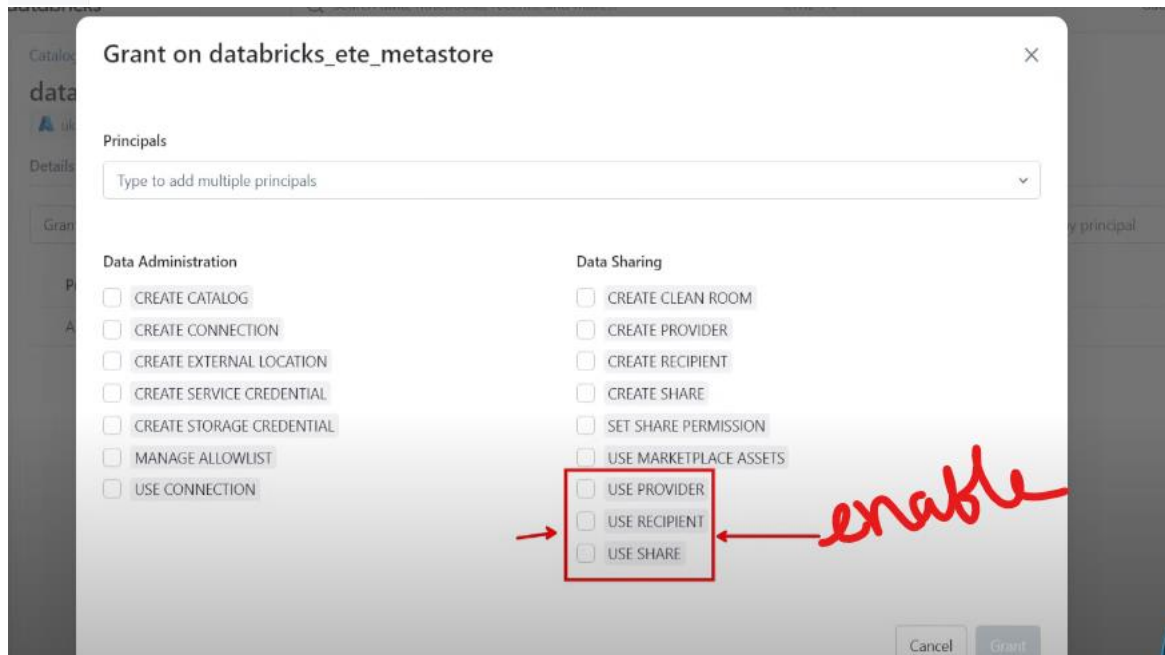
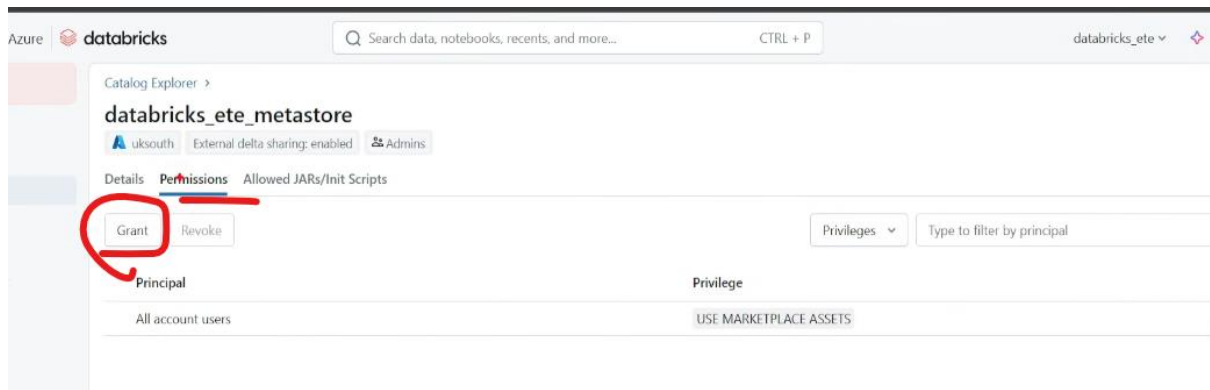
Grant Revoke Privileges Type to filter by principal

Principal	Privilege	Object
All account users	BROWSE	databricks_cata
New Users	USE CATALOG	databricks_cata



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





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)

The screenshot shows the Databricks SQL Editor interface. On the left, the 'Catalog' sidebar is open, showing a tree view of the database structure. The main editor area contains a query: `1 | SELECT SUM(salary) as total_salary FROM databricksrt.schema.dummy_tbl`. The query is executed, and the results are displayed in a table with one row: `total_salary` with a value of 300. The 'SQL Editor' option in the left sidebar is highlighted with a red line.

Save the query as alert_query_new

The screenshot shows the Databricks Alerts page. The 'Alerts' tab is selected in the left sidebar. The main area displays a table with columns: Name, Status, Last updated, Created by, and Created at. A red circle highlights the 'Create alert' button in the top right corner.

The screenshot shows the 'New alert' configuration page. The 'Alert name' field is set to 'salary_alert'. The 'Query' field is set to 'alert_query_new', with a red handwritten note 'choose query' next to it. The 'Trigger condition' section shows 'Value column' set to 'total_salary', 'Operator' set to '>', and 'Threshold value' set to 1. The 'When query result has no rows, set state to' dropdown is set to 'UNKNOWN'. The 'Preview alert' button is visible at the bottom.

Microsoft Azure databricks Search data, notebooks, recent, and more... CTRL + P

+ New

- Workspace
- Recents
- Catalog
- Workflows
- Compute
- Marketplace
- SQL
- SQL Editor
- Queries
- Dashboards
- Genie
- Alerts**
- Query History
- SQL Warehouses
- Data Engineering
- Job Runs
- Data Ingestion

Query: alert_query_new

Trigger condition

Value column: total_salary Operator: > First row

Threshold value: 3000

When query result has no rows, set state to: OK

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: He Bro, Your employees are earning more than you

Preview: ☐

Guide: Custom alert notifications

Formatting guide

ring & Alerts with Databricks >

learn.microsoft.com/en-in/azure/databricks/sql/user/alerts/#create-an-alert

- **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

☒ Send notification

Training
Module
[Configure alerts and responses - Training](#)
In this module, you learn how Azure Monitoring alerts proactively notifies you when Azure Monitor data indicates there might be a problem with you

Certification
[Microsoft Certified: Azure Database Administrator Associate - Certifications](#)
Administer an SQL Server database infrastructure cloud, on-premises and hybrid relational database 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 the Databricks SQL REST API.
[Manage notification destinations - Azure Databricks](#)
Learn how to configure notification destinations for your workspace.
[Query history - Azure Databricks](#)
Learn how to use the query history...

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, recent, 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

Last triggered at None

Query [alert_query_new](#)

Trigger condition

Value column	Operator	Threshold value
total_salary	>	3000

When query result has no rows, set state to **OK**

Notifications

When alert is triggered

Send notification just once, until back to normal.

When alert returns back to normal

☒ Send notification

Template

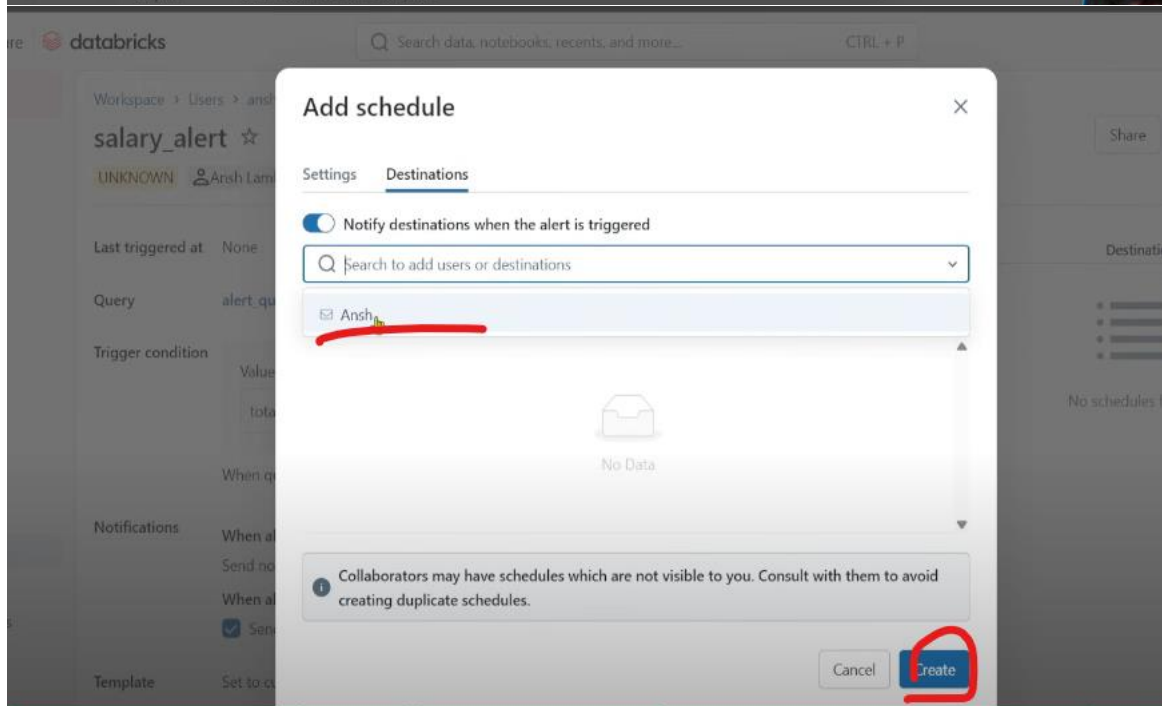
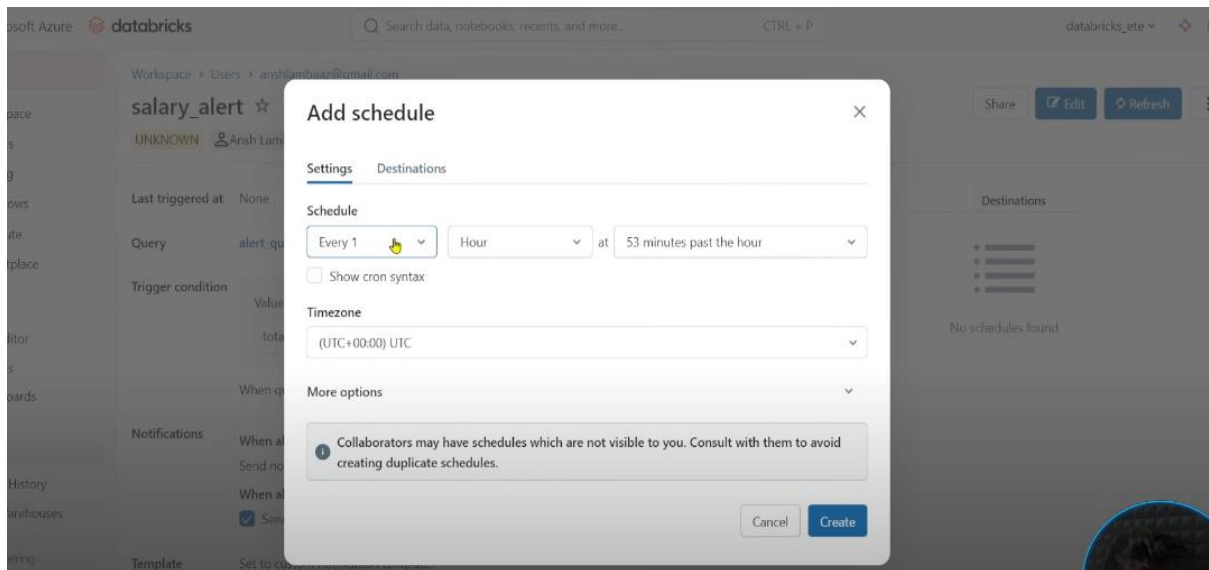
Set to custom notification template.

Schedules

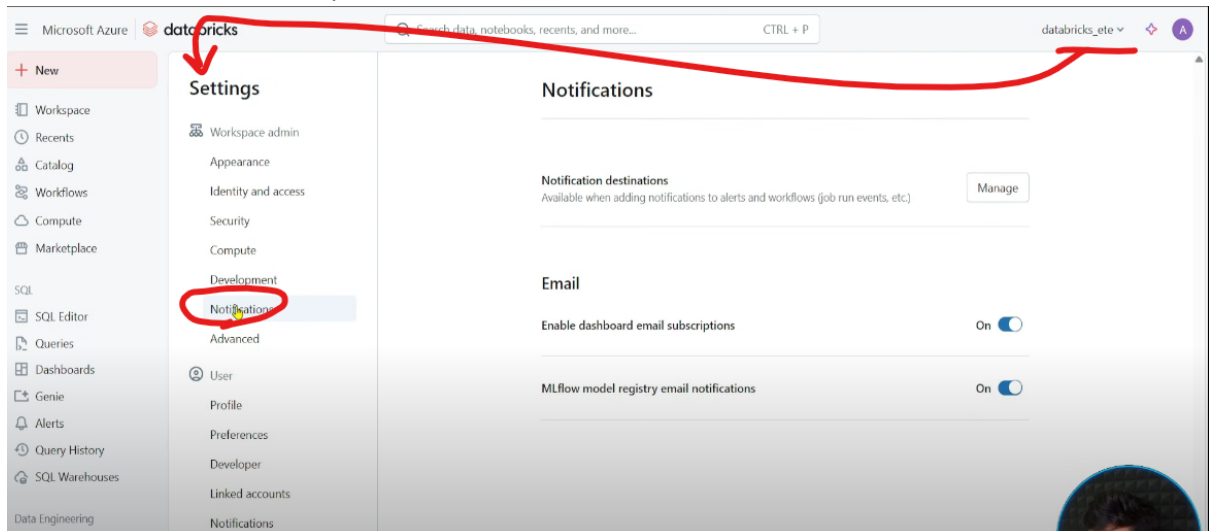
Destinations

No schedules found


[Add schedule](#)



If destination is not there by default, create one



Manage → Add destination (email id)

 databricks

CTRL + P

databricks_ete ▾

Settings

Workspace admin

Appearance

Identity and access

Security

Compute

Development

Notifications

Advanced

User

Profile

Workspace settings > Notifications >

Notification destinations

Configure the destinations that will be available for your users when adding notifications to their workflows (job run events, etc)

Add destination

Name

Ansh

1