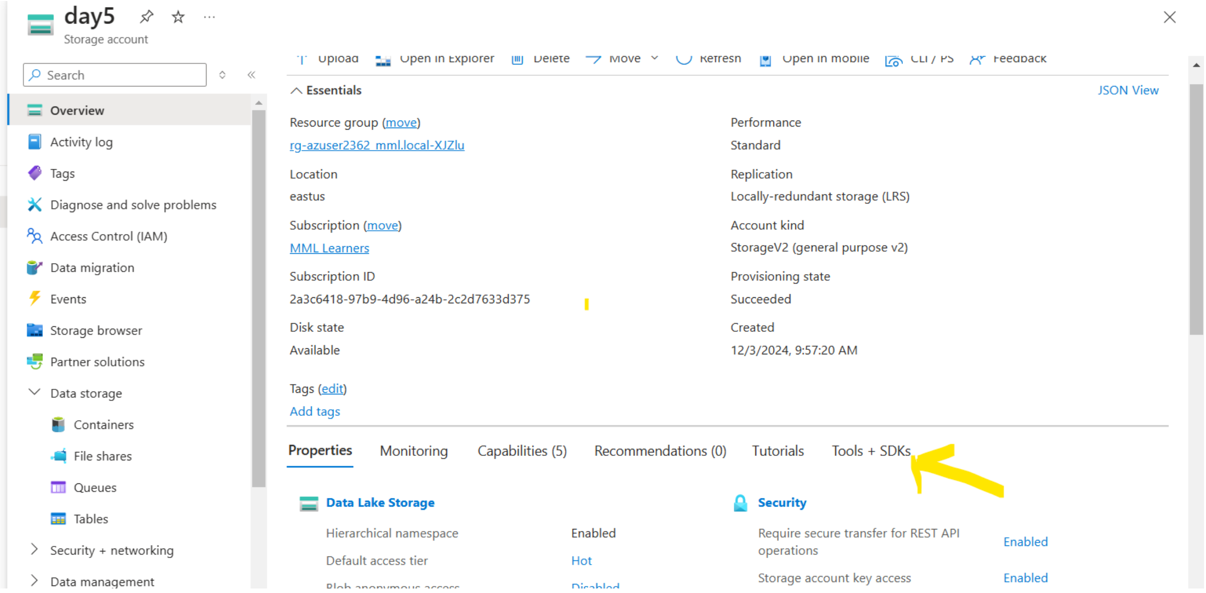
Day 5 Assignment | Dec 03 2024

Steps to Download Azure blob storage account on desktop

Step1. Create a storage account and in overview select “Tools/SDK”



Resource group (move)
rg-azuser2362_mml.local-XJZlu

Location
eastus

Subscription (move)
MML Learners

Subscription ID
2a3c6418-97b9-4d96-a24b-2c2d7633d375

Disk state
Available

Tags (edit)
Add tags

Performance
Standard

Replication
Locally-redundant storage (LRS)

Account kind
StorageV2 (general purpose v2)

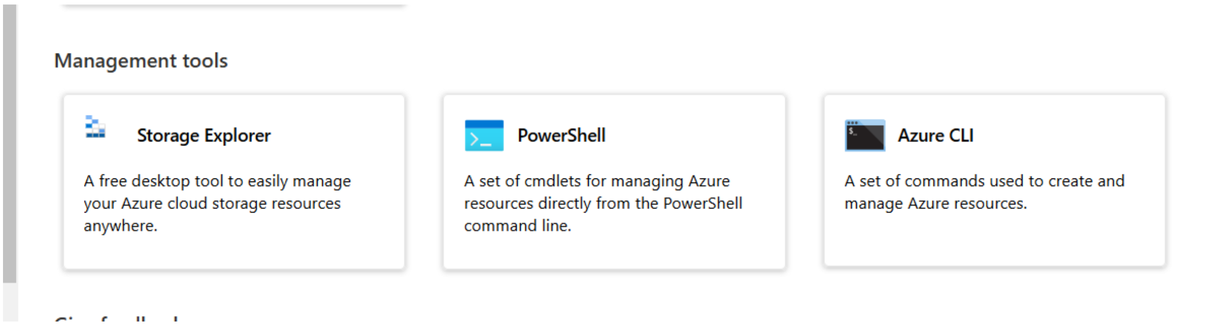
Provisioning state
Succeeded

Created
12/3/2024, 9:57:20 AM

Hierarchical namespace Enabled
Default access tier Hot
Blob anonymous access Disabled

Require secure transfer for REST API operations Enabled
Storage account key access Enabled

Step2. Select storage explorer



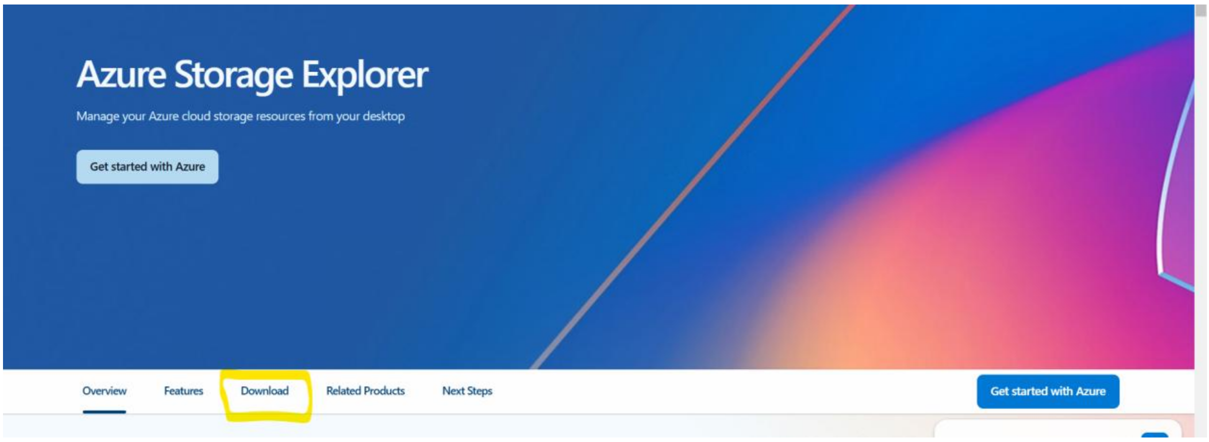
Management tools

Storage Explorer
A free desktop tool to easily manage your Azure cloud storage resources anywhere.

PowerShell
A set of cmdlets for managing Azure resources directly from the PowerShell command line.

Azure CLI
A set of commands used to create and manage Azure resources.

Step3: Go to download and then download “windows*64” version and then install it on desktop



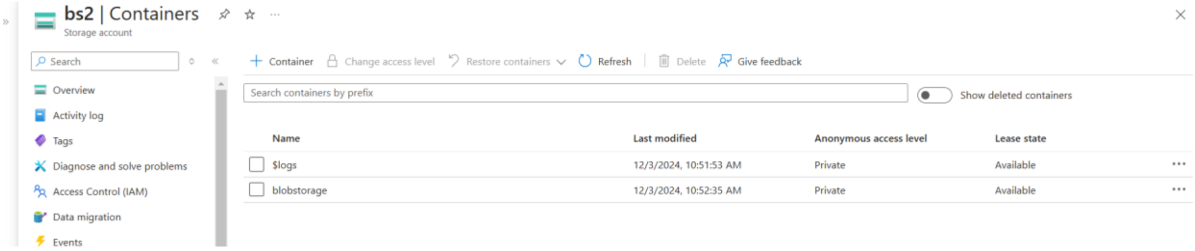
Azure Storage Explorer

Manage your Azure cloud storage resources from your desktop

Create 2 storage accounts one for blob storage and other for Datalake gen2

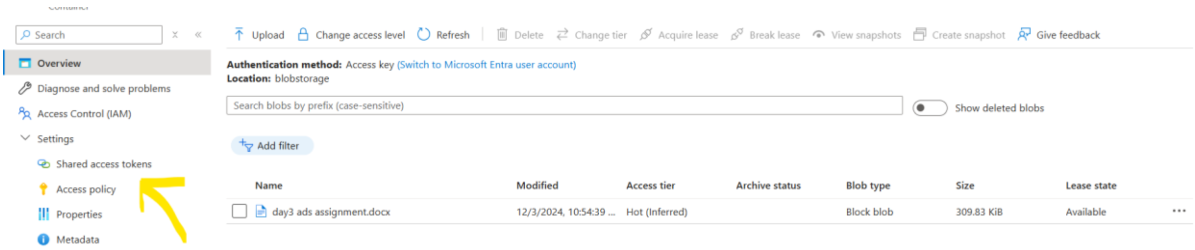
1.Blob storage

Step1: Create container



Name	Last modified	Anonymous access level	Lease state
$logs	12/3/2024, 10:51:53 AM	Private	Available
blobstorage	12/3/2024, 10:52:35 AM	Private	Available

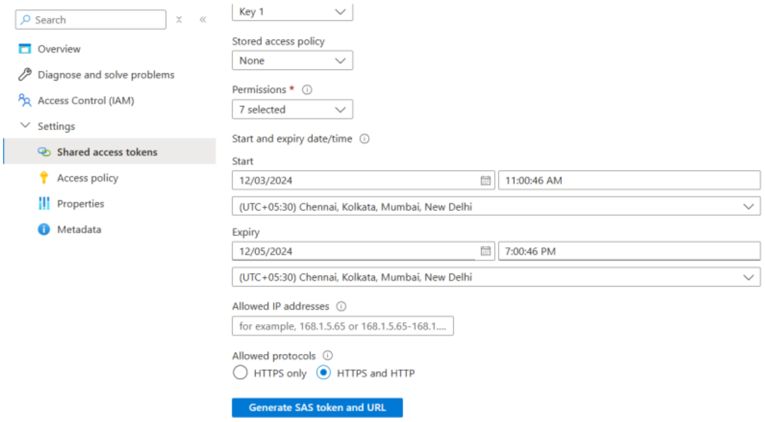
Step2: select settings/shared token access



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

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
day3 ads assignment.docx	12/3/2024, 10:54:39 ...	Hot (Inferred)		Block blob	309.83 KiB	Available

Step3:Generate SAS URL by giving following specifications



Key 1

Stored access policy: None

Permissions *: 7 selected

Start and expiry date/time

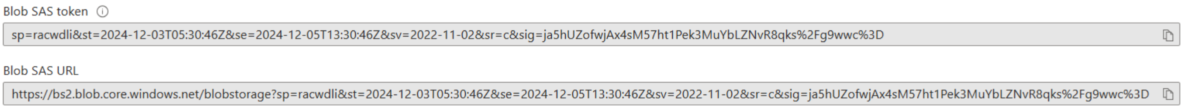
Start: 12/03/2024 11:00:46 AM
(UTC+05:30) Chennai, Kolkata, Mumbai, New Delhi

Expiry: 12/05/2024 7:00:46 PM
(UTC+05:30) Chennai, Kolkata, Mumbai, New Delhi

Allowed IP addresses: for example, 168.1.5.65 or 168.1.5.65-168.1....

Allowed protocols: ○ HTTPS only ● HTTPS and HTTP

Step4: Copy blob SAS URL



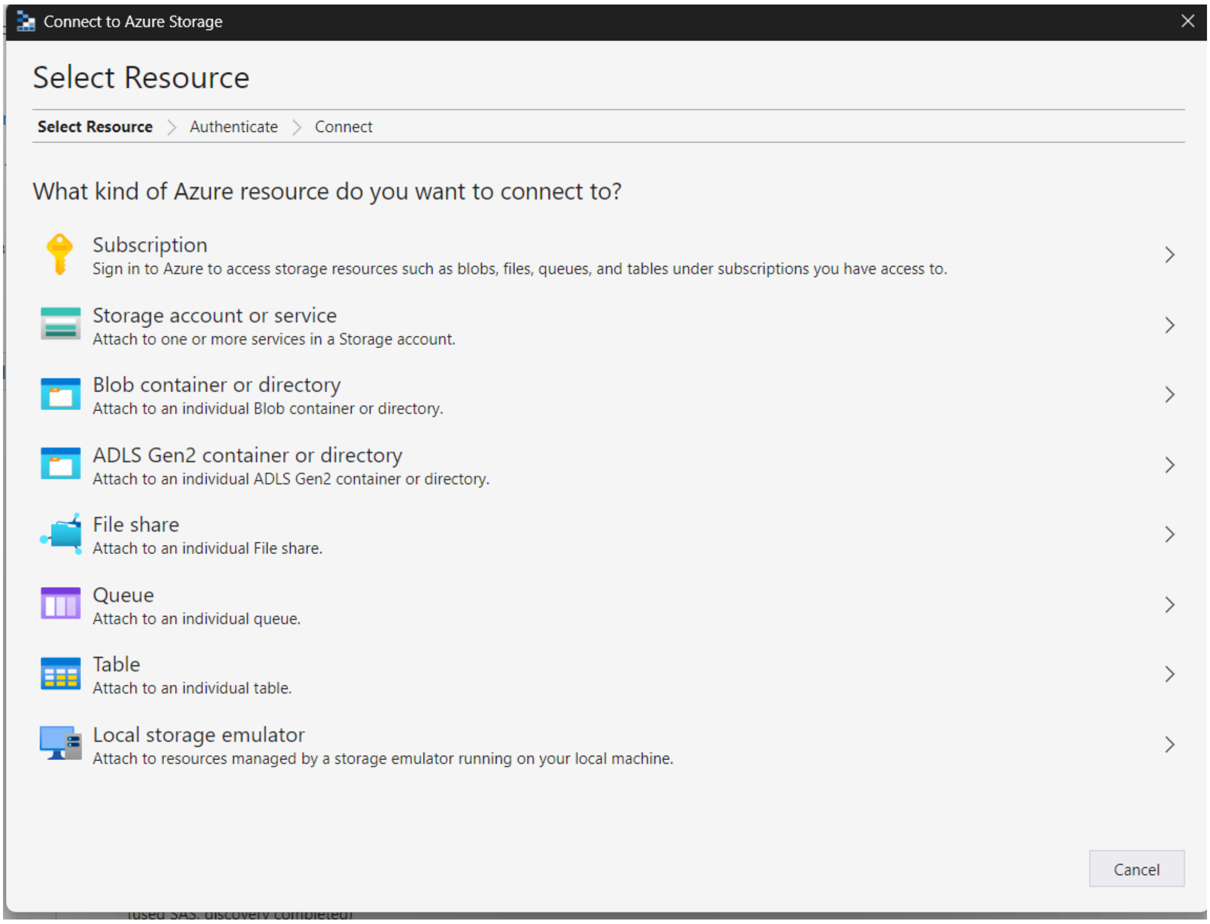
Blob SAS token

sp=racwdli&st=2024-12-03T05:30:46Z&se=2024-12-05T13:30:46Z&sv=2022-11-02&sr=c&sig=ja5hUZofwjAx4sM57ht1Pek3MuYbLZNvR8qks%2Fg9wwc%3D

Blob SAS URL

https://bs2.blob.core.windows.net/blobstorage?sp=racwdli&st=2024-12-03T05:30:46Z&se=2024-12-05T13:30:46Z&sv=2022-11-02&sr=c&sig=ja5hUZofwjAx4sM57ht1Pek3MuYbLZNvR8qks%2Fg9wwc%3D

Step5:In azure blob storage account, go to plugin.Select blob container or directory



Connect to Azure Storage

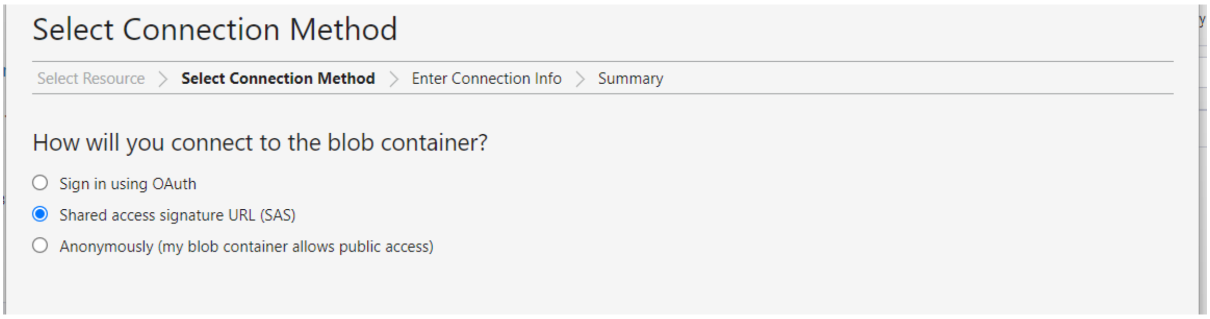
Select Resource

Select Resource > Authenticate > Connect

What kind of Azure resource do you want to connect to?

- Subscription
  Sign in to Azure to access storage resources such as blobs, files, queues, and tables under subscriptions you have access to.
- Storage account or service
  Attach to one or more services in a Storage account.
- Blob container or directory
  Attach to an individual Blob container or directory.
- ADLS Gen2 container or directory
  Attach to an individual ADLS Gen2 container or directory.
- File share
  Attach to an individual File share.
- Queue
  Attach to an individual queue.
- Table
  Attach to an individual table.
- Local storage emulator
  Attach to resources managed by a storage emulator running on your local machine.

Step 6:select shared access signature URL(SAS)



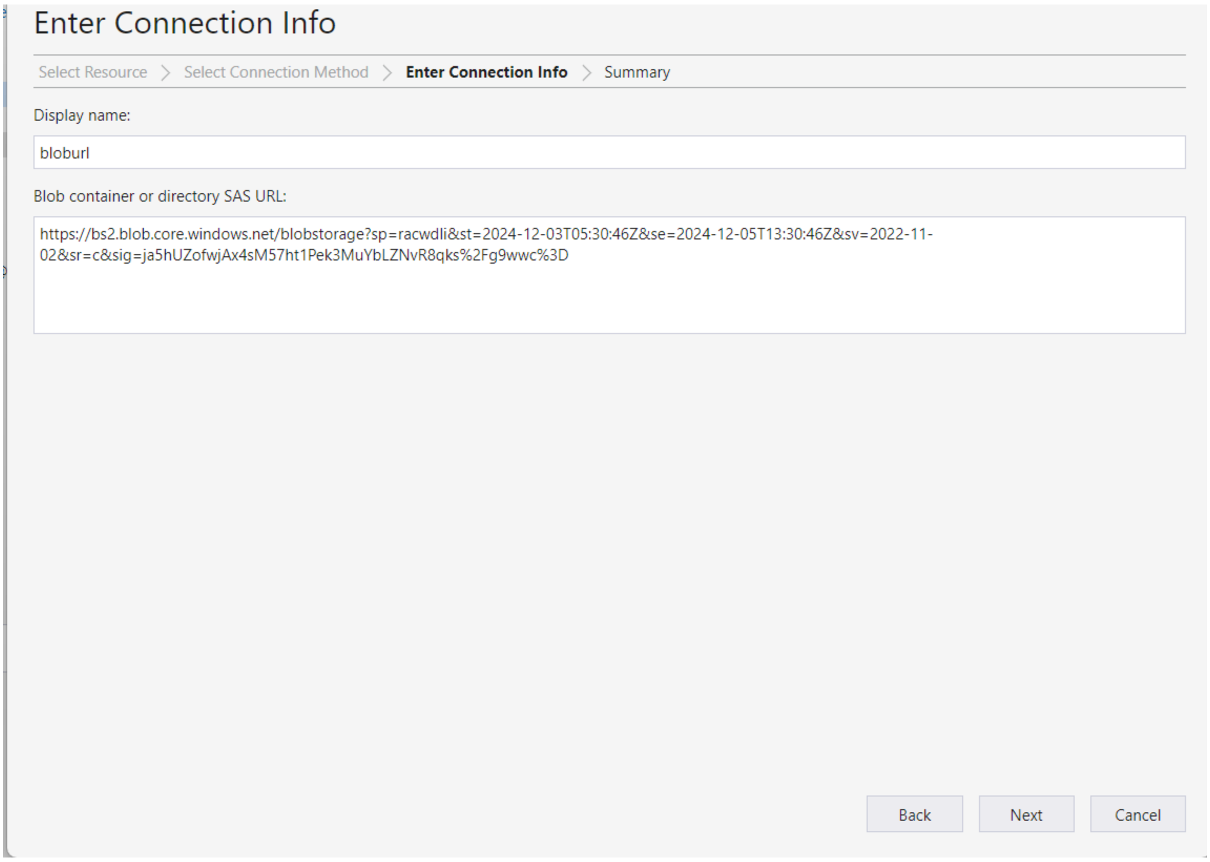
Select Connection Method

Select Resource > **Select Connection Method** > Enter Connection Info > Summary

How will you connect to the blob container?

○ Sign in using OAuth
● Shared access signature URL (SAS)
○ Anonymously (my blob container allows public access)

Step7:Paste the URL and click on “connect”



Enter Connection Info

Display name:

bloburl

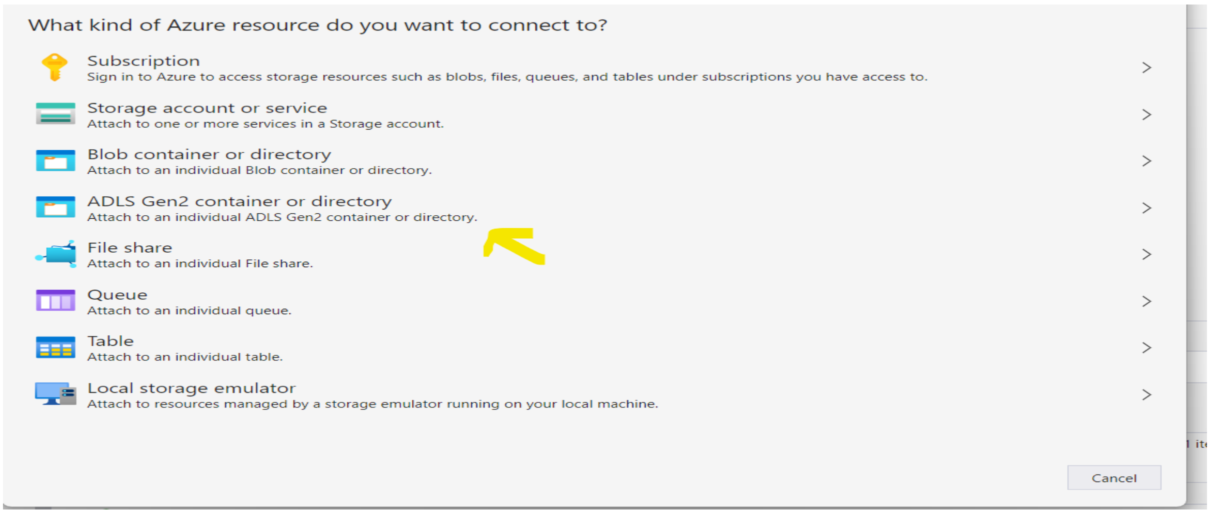
Blob container or directory SAS URL:

https://bs2.blob.core.windows.net/blobstorage?sp=racwdli&st=2024-12-03T05:30:46Z&se=2024-12-05T13:30:46Z&sv=2022-11-02&sr=c&sig=ja5hUZofwjAx4sM57ht1Pek3MuYbLZNvR8qks%2Fg9wwc%3D

2.Blob storage data lake gen2

The steps followed are same except the 5th step

Step5 Since it is data lake account select ADLS Gen2 container under “plug in”



What kind of Azure resource do you want to connect to?

Subscription
Sign in to Azure to access storage resources such as blobs, files, queues, and tables under subscriptions you have access to.

Storage account or service
Attach to one or more services in a Storage account.

Blob container or directory
Attach to an individual Blob container or directory.

ADLS Gen2 container or directory
Attach to an individual ADLS Gen2 container or directory.

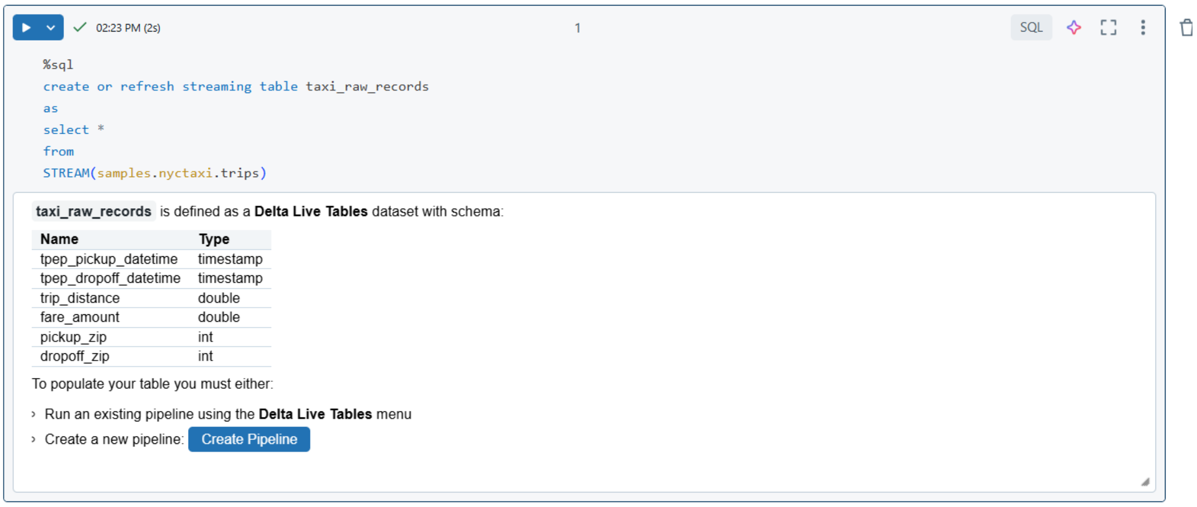
File share
Attach to an individual File share.

Queue
Attach to an individual queue.

Table
Attach to an individual table.

Local storage emulator
Attach to resources managed by a storage emulator running on your local machine.

Delta Live Tables



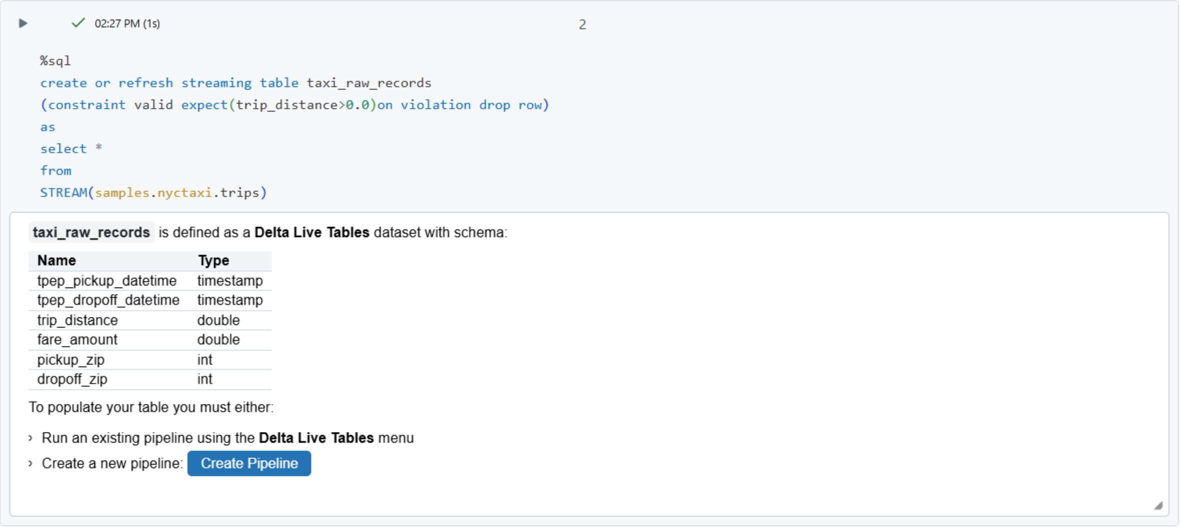
```
%sql
create or refresh streaming table taxi_raw_records
as
select *
from
STREAM(samples.nyctaxi.trips)
```

taxi_raw_records is defined as a **Delta Live Tables** dataset with schema:

Name	Type
tpep_pickup_datetime	timestamp
tpep_dropoff_datetime	timestamp
trip_distance	double
fare_amount	double
pickup_zip	int
dropoff_zip	int

To populate your table you must either:

- Run an existing pipeline using the **Delta Live Tables** menu
- Create a new pipeline: Create Pipeline



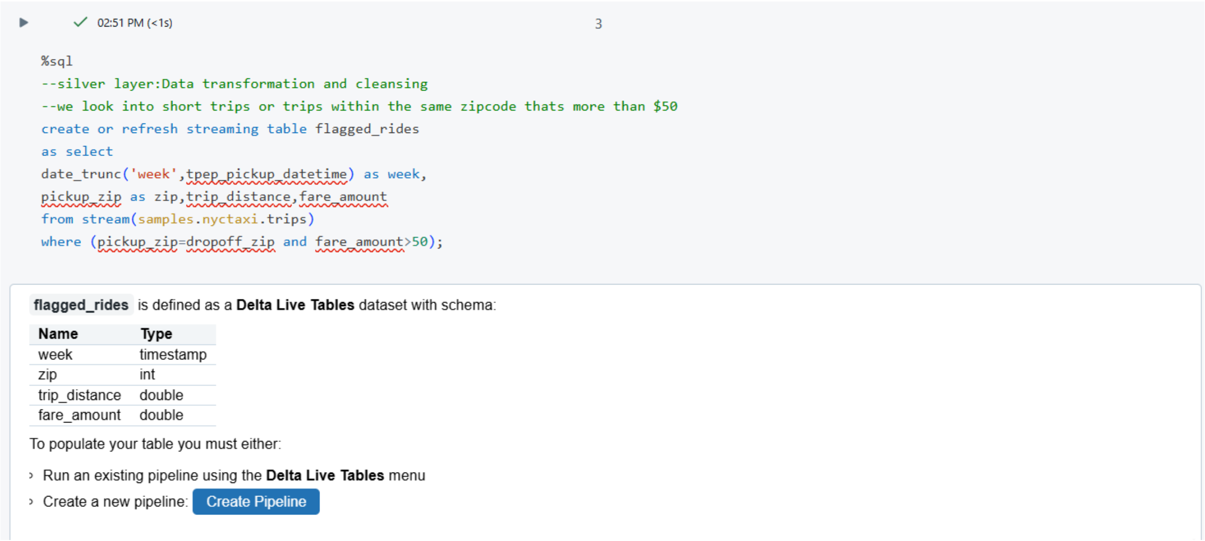
```
%sql
create or refresh streaming table taxi_raw_records
(constraint valid expect(trip_distance>0.0)on violation drop row)
as
select *
from
STREAM(samples.nyctaxi.trips)
```

taxi_raw_records is defined as a **Delta Live Tables** dataset with schema:

Name	Type
tpep_pickup_datetime	timestamp
tpep_dropoff_datetime	timestamp
trip_distance	double
fare_amount	double
pickup_zip	int
dropoff_zip	int

To populate your table you must either:

- Run an existing pipeline using the **Delta Live Tables** menu
- Create a new pipeline: Create Pipeline



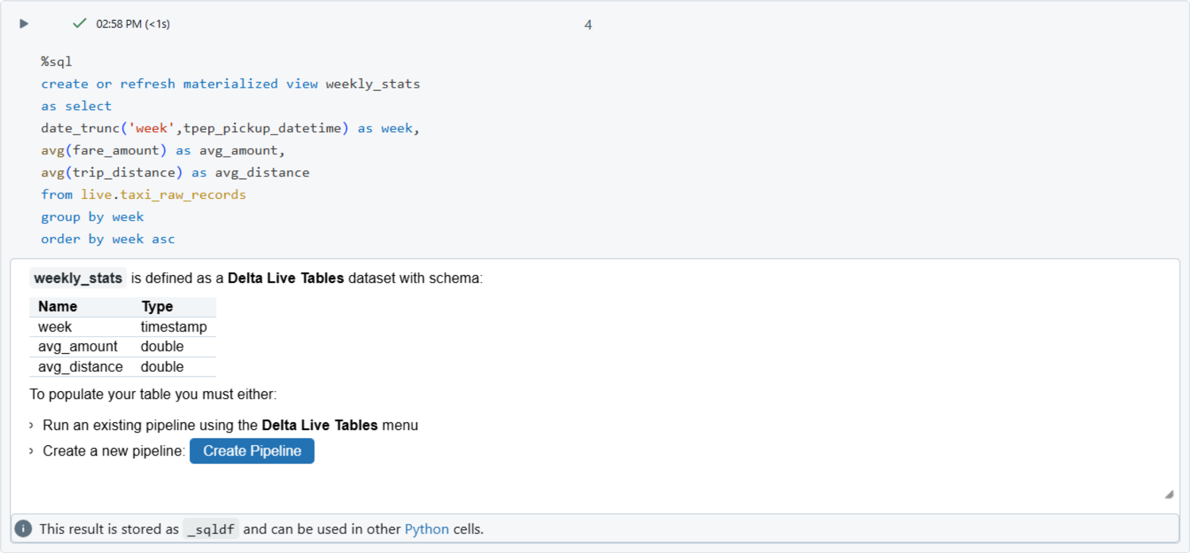
```
%sql
--silver layer:Data transformation and cleansing
--we look into short trips or trips within the same zipcode thats more than $50
create or refresh streaming table flagged_rides
as select
date_trunc('week',tpep_pickup_datetime) as week,
pickup_zip as zip,trip_distance,fare_amount
from stream(samples.nyctaxi.trips)
where (pickup_zip=dropoff_zip and fare_amount>50);
```

flagged_rides is defined as a **Delta Live Tables** dataset with schema:

Name	Type
week	timestamp
zip	int
trip_distance	double
fare_amount	double

To populate your table you must either:

- Run an existing pipeline using the **Delta Live Tables** menu
- Create a new pipeline: Create Pipeline



```
%sql
create or refresh materialized view weekly_stats
as select
date_trunc('week',tpep_pickup_datetime) as week,
avg(fare_amount) as avg_amount,
avg(trip_distance) as avg_distance
from live.taxi_raw_records
group by week
order by week asc
```

`weekly_stats` is defined as a **Delta Live Tables** dataset with schema:

Name	Type
week	timestamp
avg_amount	double
avg_distance	double

To populate your table you must either:

› Run an existing pipeline using the **Delta Live Tables** menu

› Create a new pipeline: Create Pipeline

This result is stored as `_sqldf` and can be used in other Python cells.