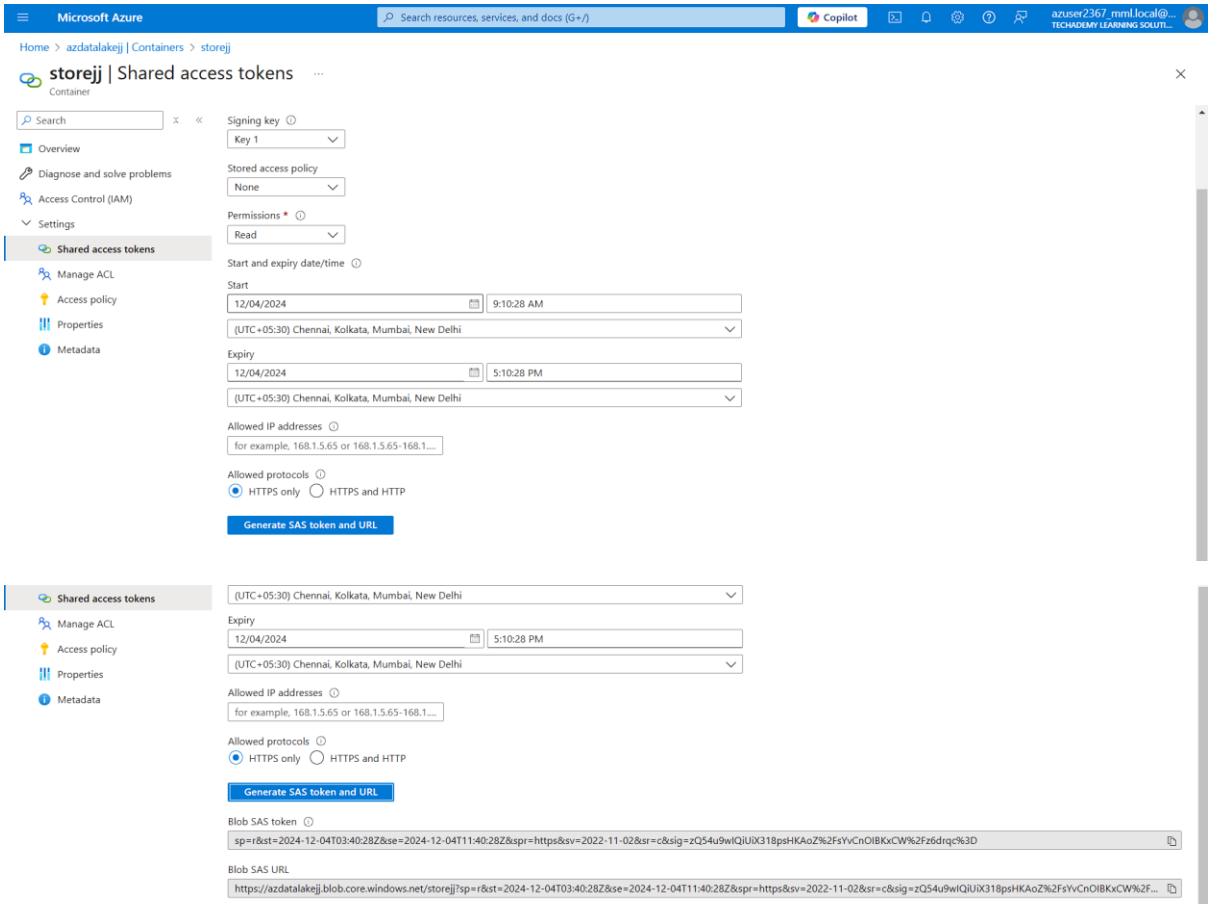
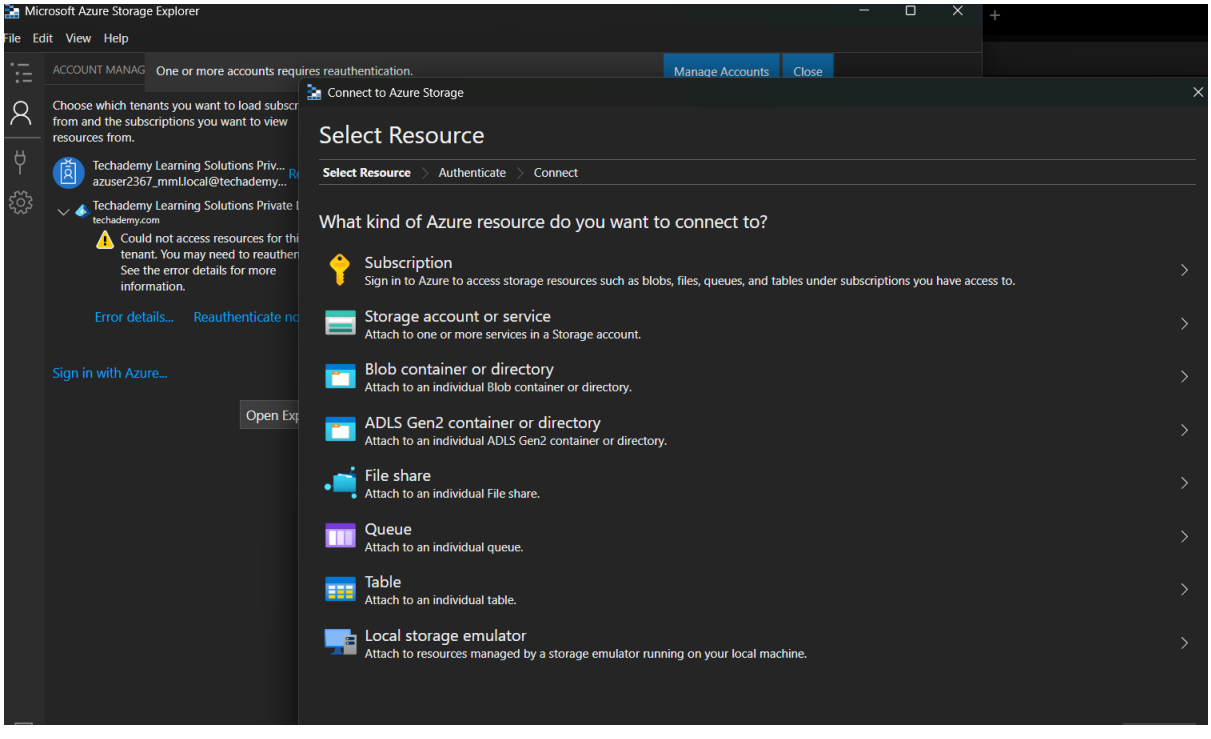
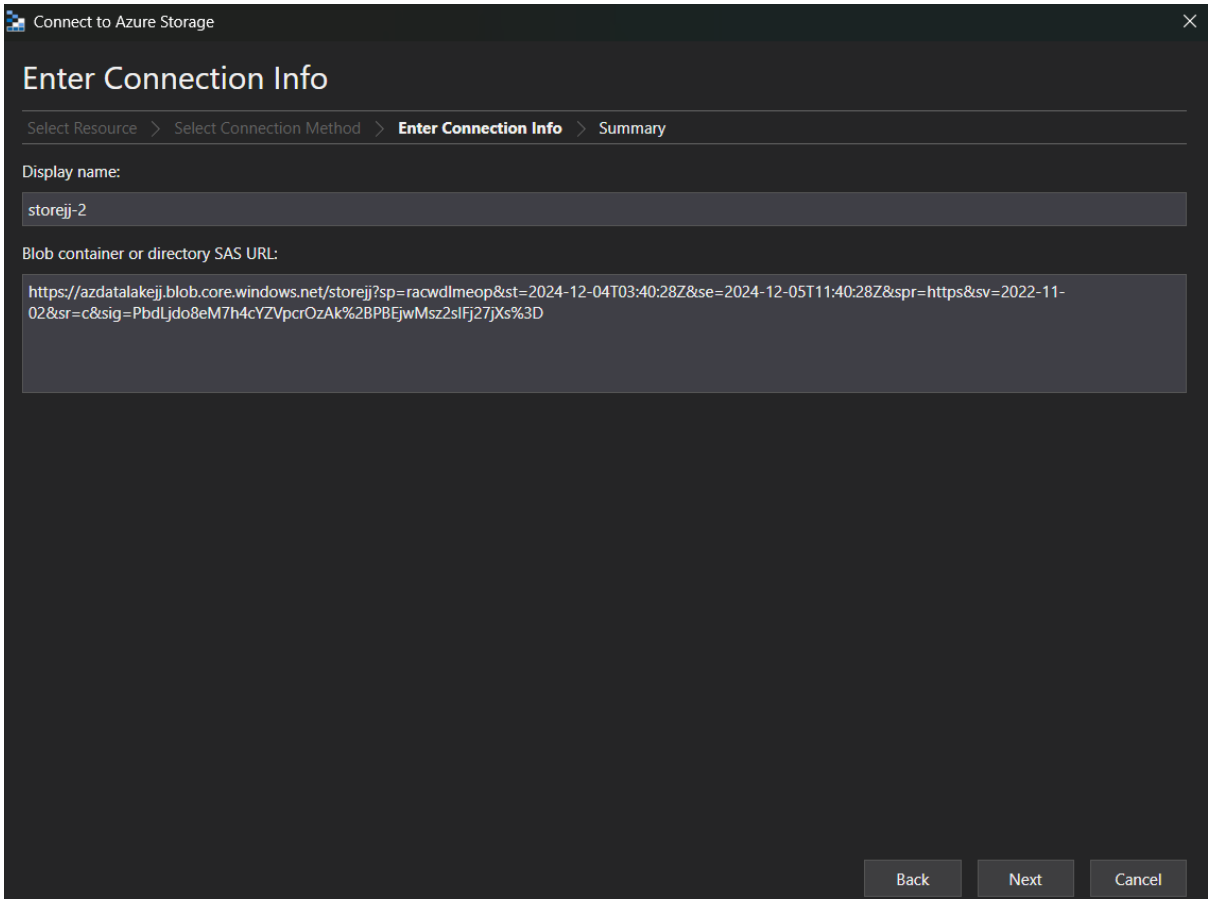
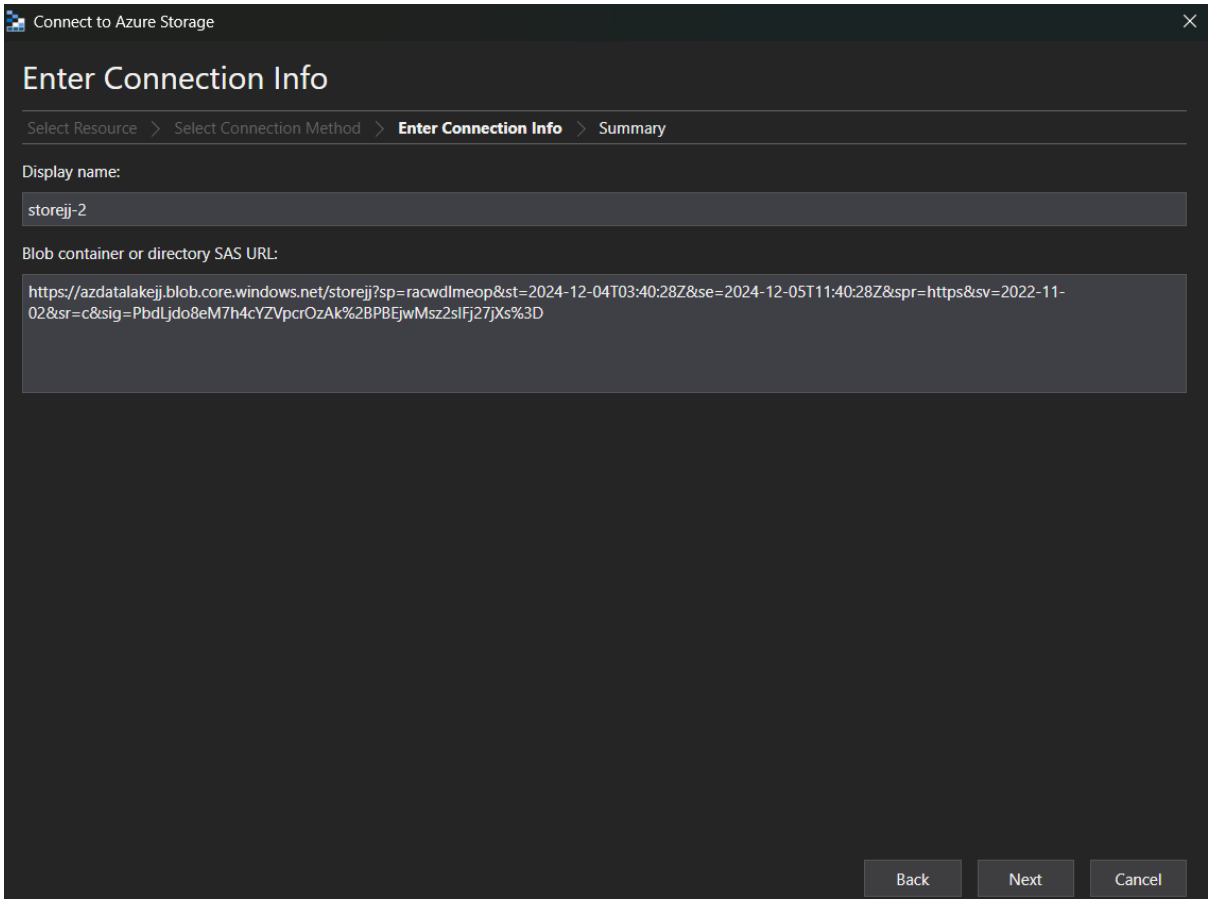
Azure Data Engineering

Connecting Azure Storage Explorer with Resource Container:

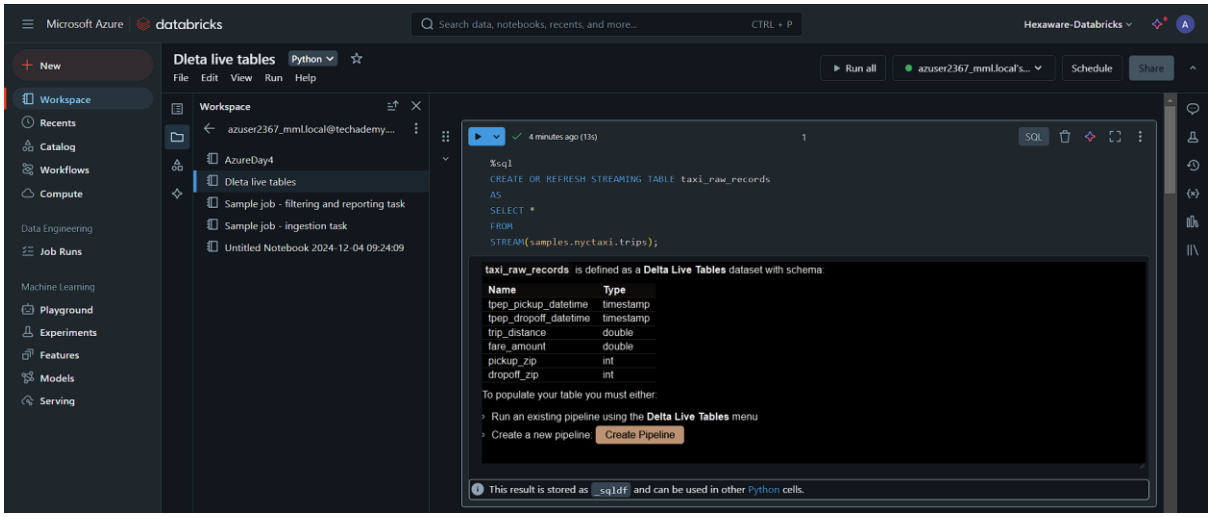


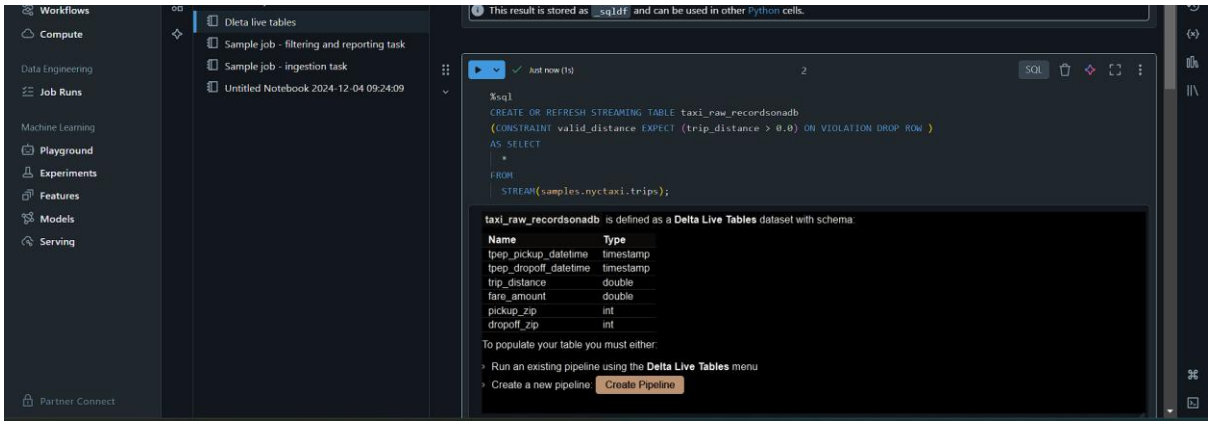


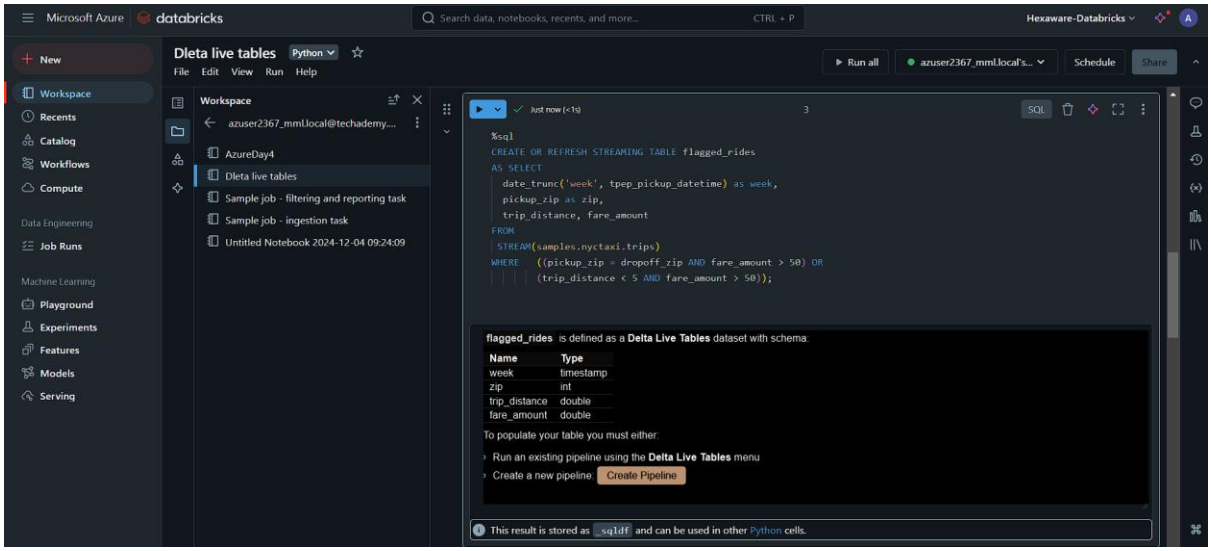

Connect to Azure Storage
Enter Connection Info
Select Resource > Select Connection Method > Enter Connection Info > Summary
Display name:
storejj-2
Blob container or directory SAS URL:
https://azdatalakejj.blob.core.windows.net/storejj?sp=racwdlmeop&st=2024-12-04T03:40:28Z&se=2024-12-05T11:40:28Z&spr=https&sv=2022-11-02&sr=c&sig=PbdLjdo8eM7h4cYZVpcrOzAk%2BPBEjwMsz2slFj27jXs%3D
Back
Next
Cancel

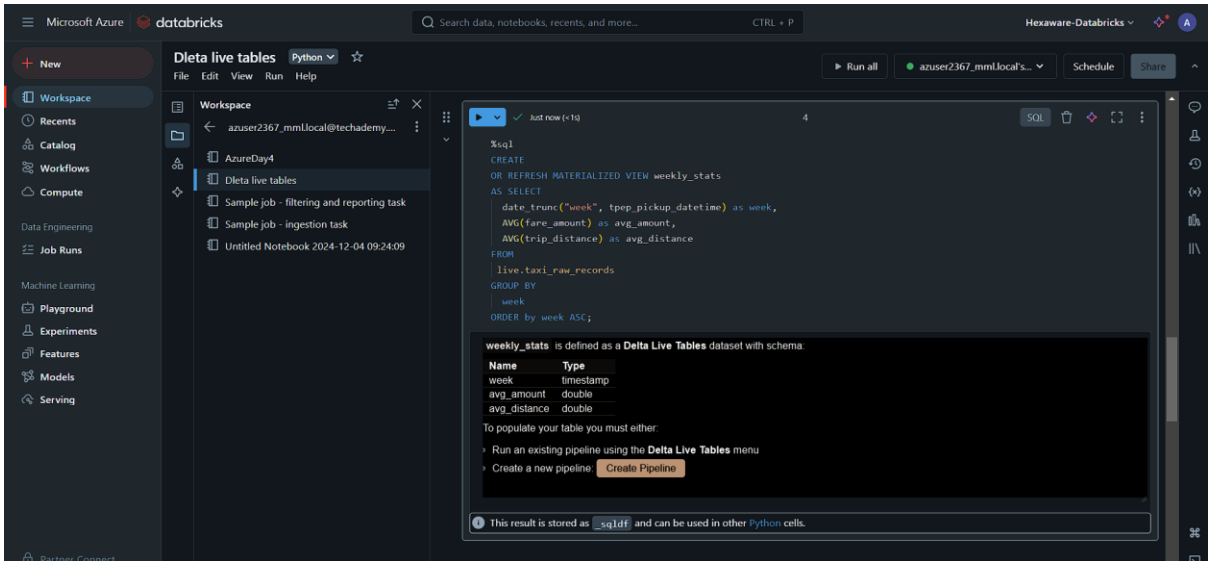

Connect to Azure Storage
Enter Connection Info
Select Resource > Select Connection Method > Enter Connection Info > Summary
Display name:
storejj-2
Blob container or directory SAS URL:
https://azdatalakejj.blob.core.windows.net/storejj?sp=racwdlmeop&st=2024-12-04T03:40:28Z&se=2024-12-05T11:40:28Z&spr=https&sv=2022-11-02&sr=c&sig=PbdLjdo8eM7h4cYZVpcrOzAk%2BPBEjwMsz2slFj27jXs%3D
Back
Next
Cancel

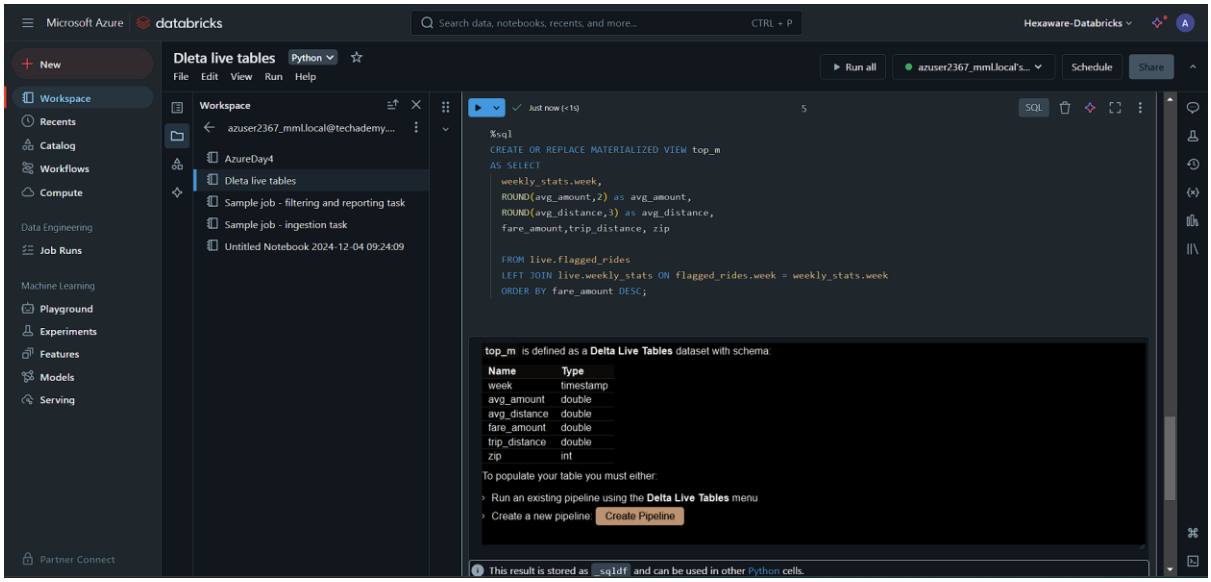
Creating Delta-Live Table:








Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
Hexaware-Databricks
New
Dleta live tables
Python
File Edit View Run Help
Run all
azuser2367_mml.local's...
Schedule
Share
Workspace
Recents
Catalog
Workflows
Compute
Data Engineering
Job Runs
Machine Learning
Playground
Experiments
Features
Models
Serving
Workspace
azuser2367_mml.local@techademy....
AzureDay4
Dleta live tables
Sample job - filtering and reporting task
Sample job - ingestion task
Untitled Notebook 2024-12-04 09:24:09
Just now (<1s)
4
SQL
%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.


Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
Hexaware-Databricks
New
Dleta live tables
Python
File Edit View Run Help
Run all
azuser2367_mml.local's...
Schedule
Share
Workspace
Recents
Catalog
Workflows
Compute
Data Engineering
Job Runs
Machine Learning
Playground
Experiments
Features
Models
Serving
Partner Connect
Workspace
azuser2367_mml.local@techademy....
AzureDay4
Dleta live tables
Sample job - filtering and reporting task
Sample job - ingestion task
Untitled Notebook 2024-12-04 09:24:09
Just now (<1s)
5
SQL
%sql
CREATE OR REPLACE MATERIALIZED VIEW top_m
AS SELECT
weekly_stats.week,
ROUND(avg_amount,2) as avg_amount,
ROUND(avg_distance,3) as avg_distance,
fare_amount,trip_distance, zip
FROM live.flagged_rides
LEFT JOIN live.weekly_stats ON flagged_rides.week = weekly_stats.week
ORDER BY fare_amount DESC;
top_m is defined as a Delta Live Tables dataset with schema:
Name Type
week timestamp
avg_amount double
avg_distance double
fare_amount double
trip_distance double
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
This result is stored as _sqldf and can be used in other Python cells.