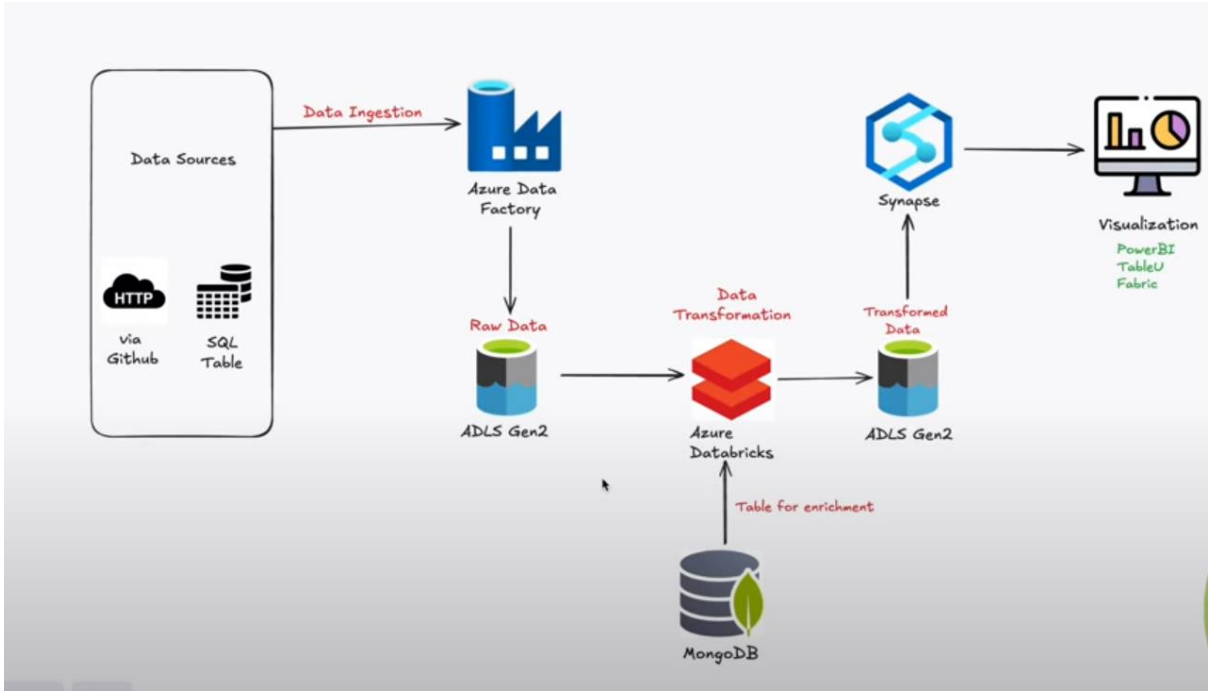
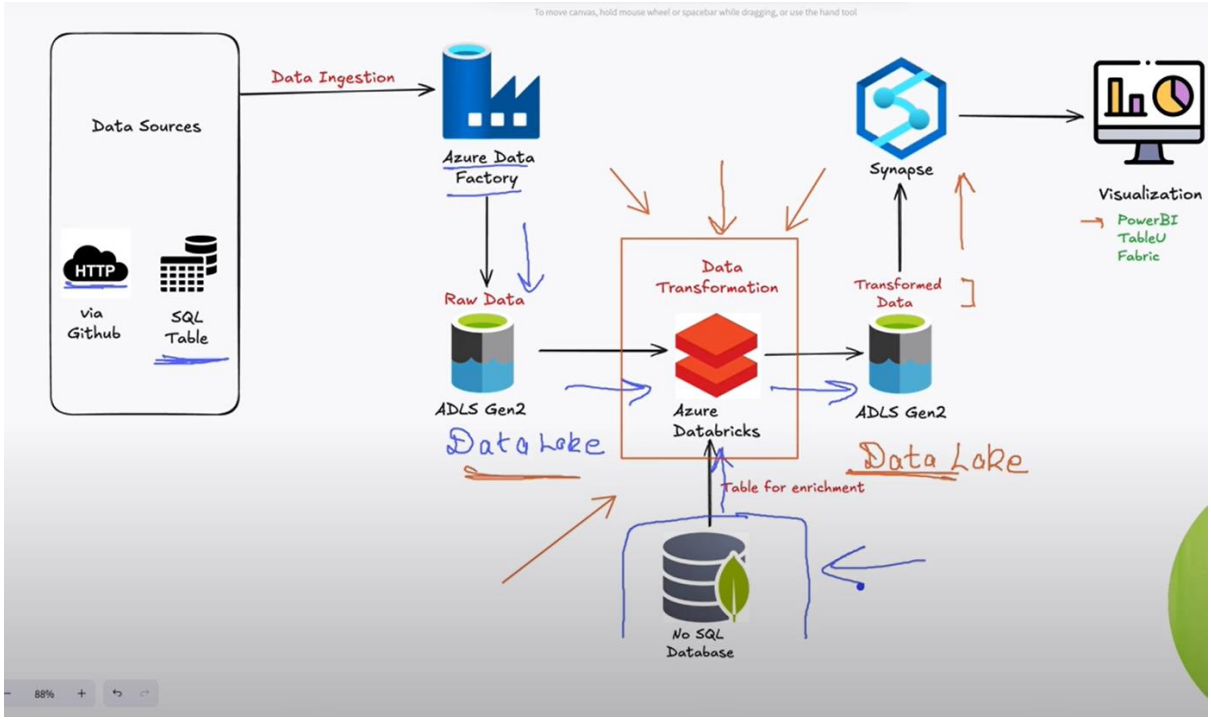
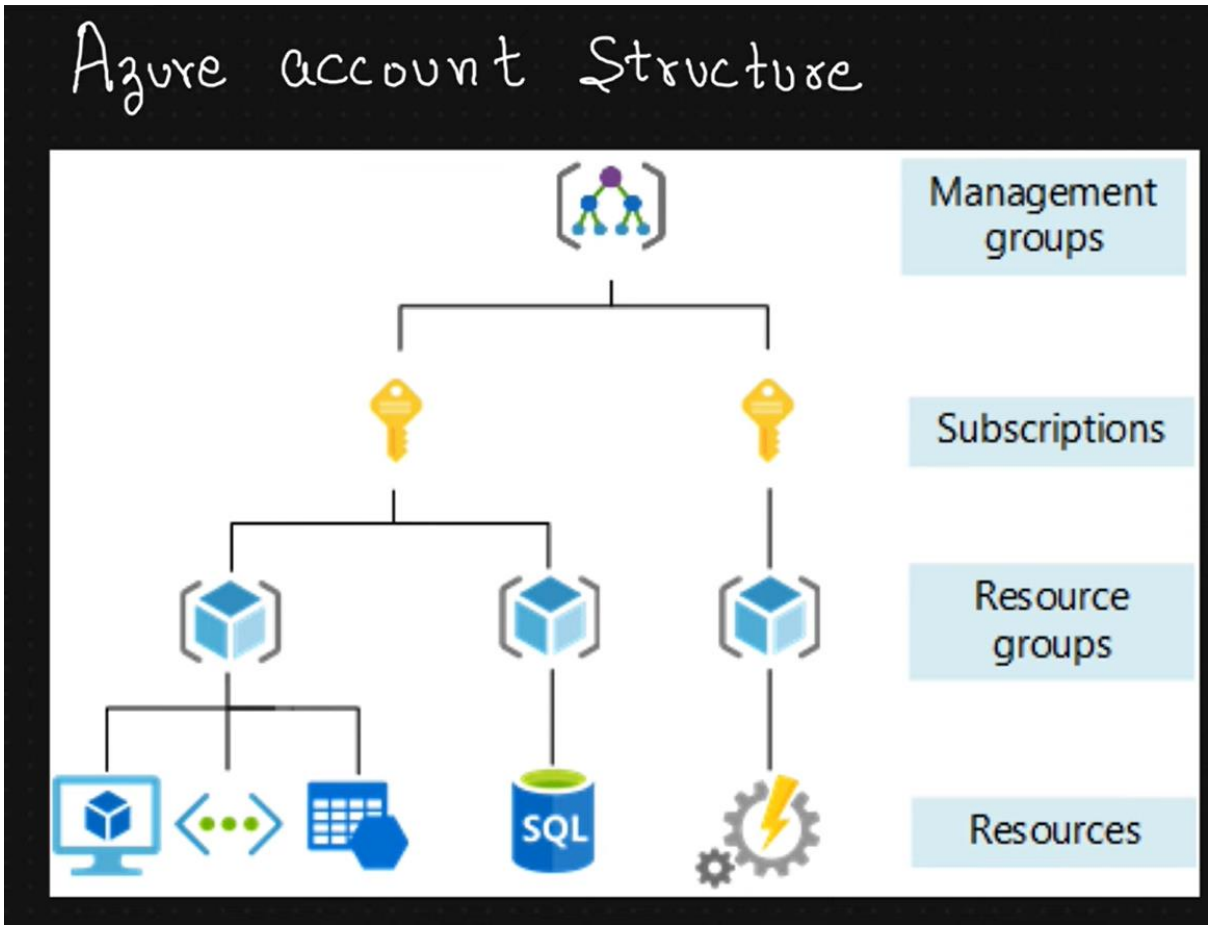
Domain – Ecommerce Data Engineering

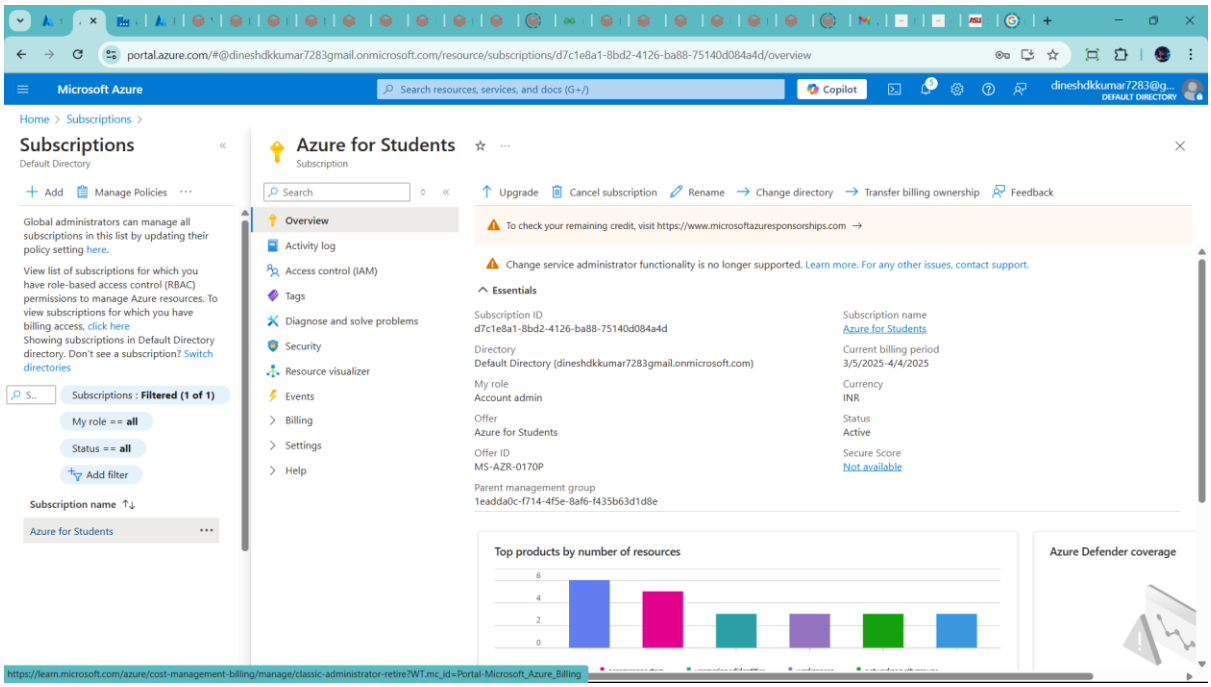




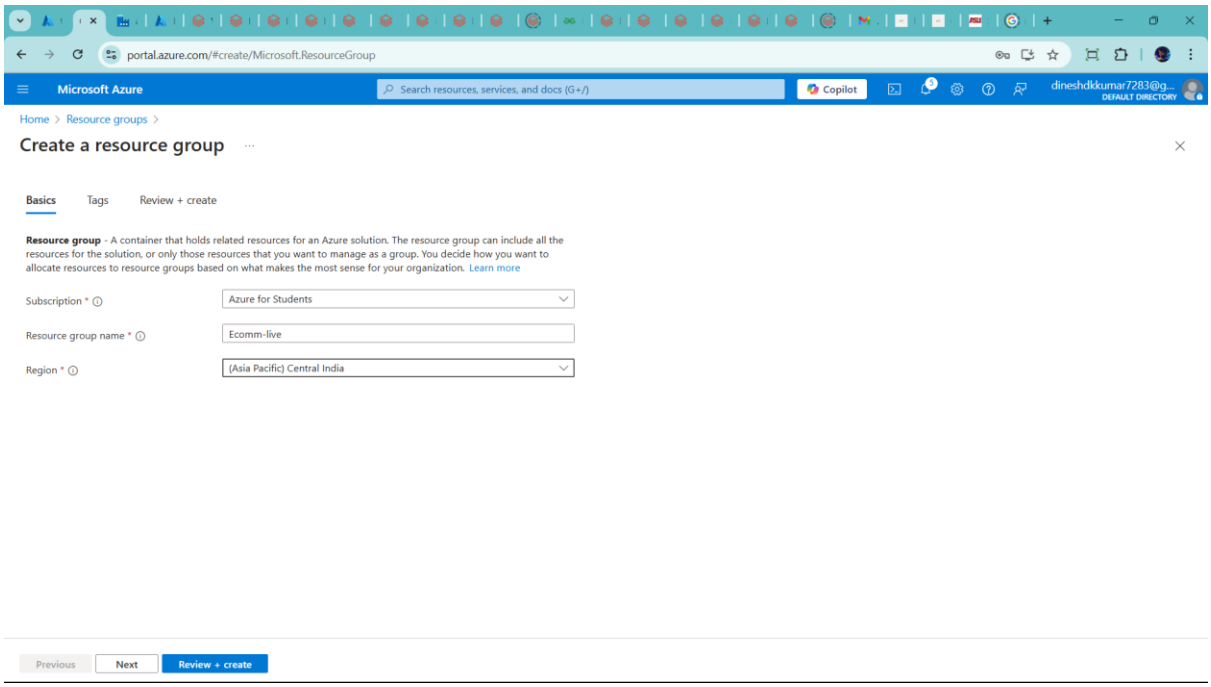
Azure Account Structure

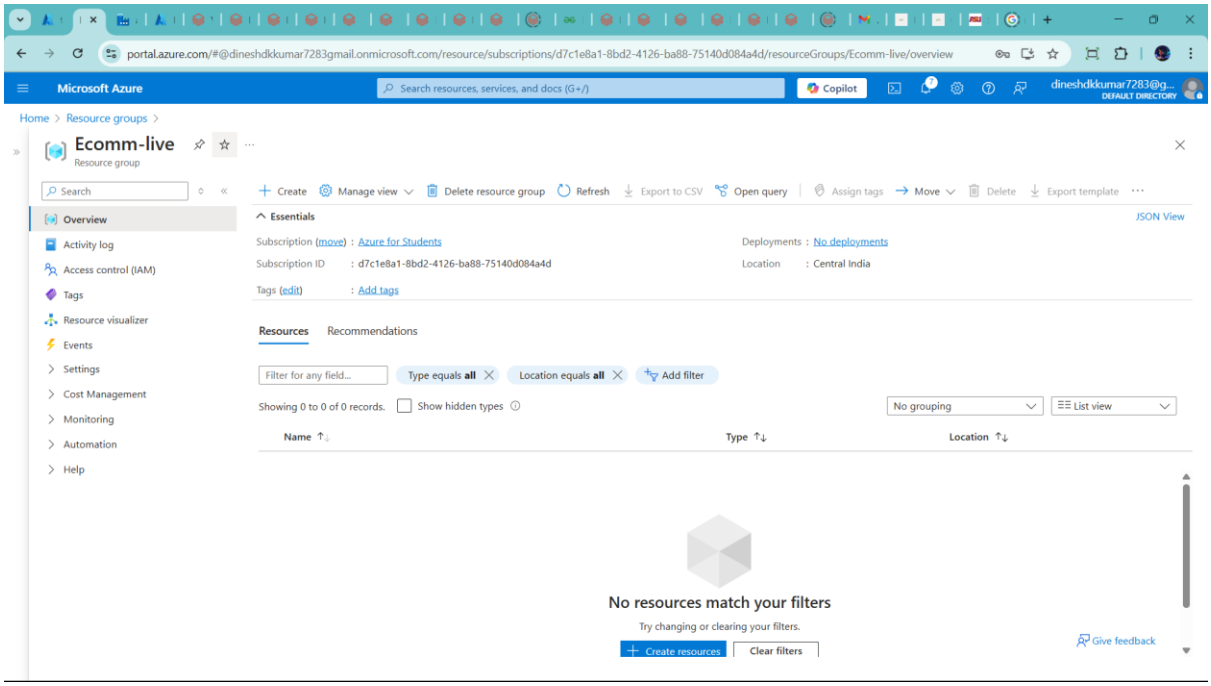


→it already free train in student id

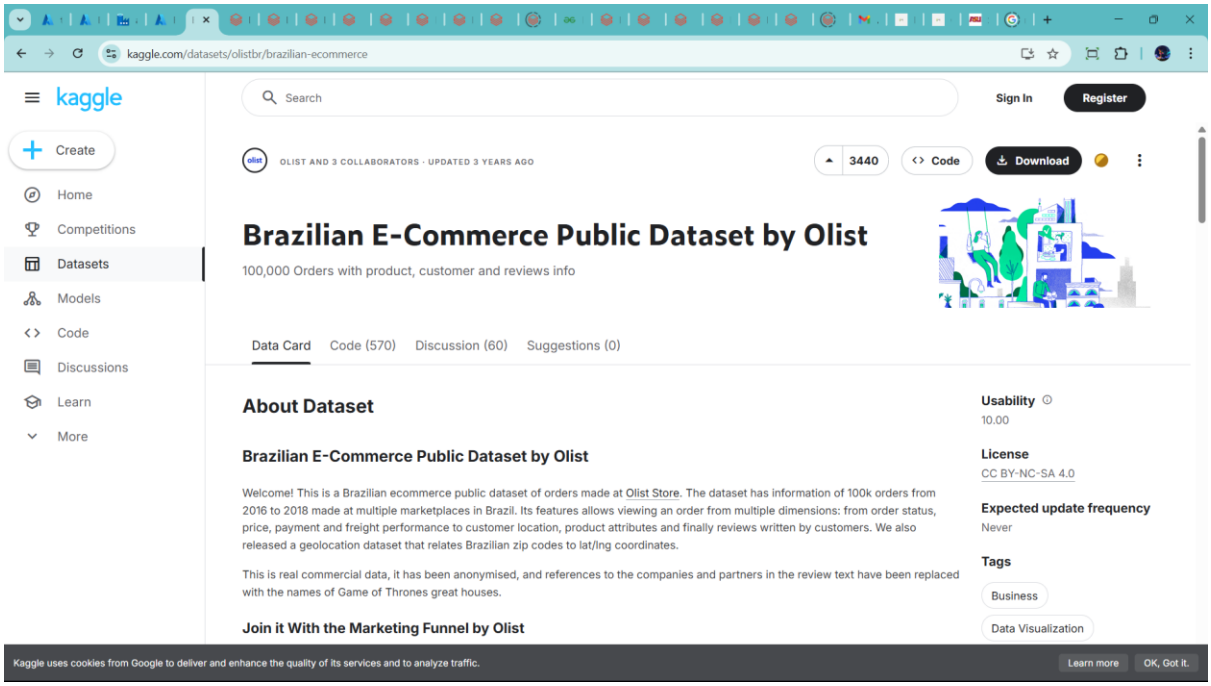


→Create Azure Resource Group

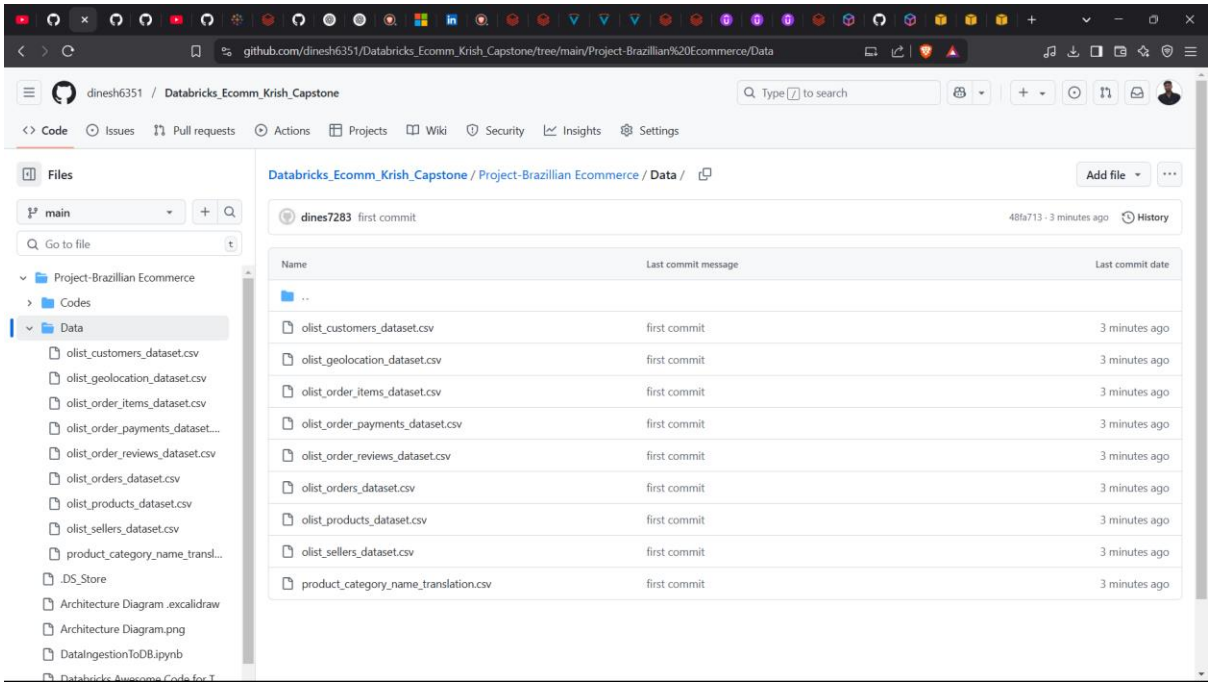




→Dataset used

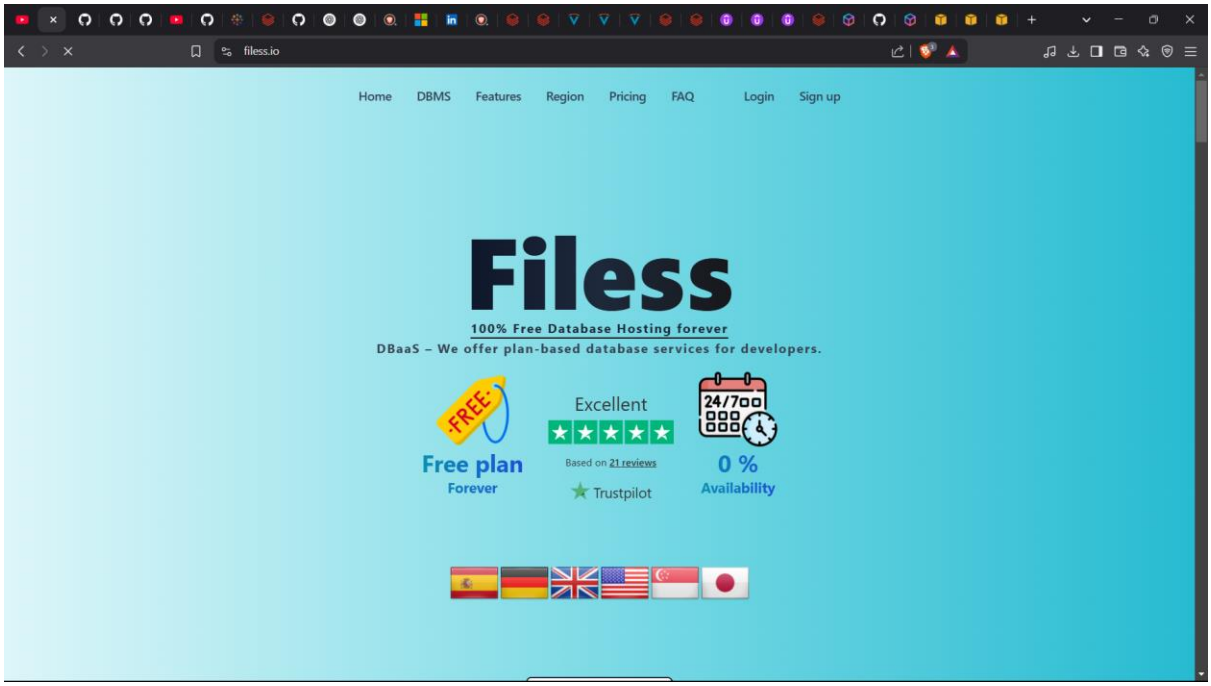


--

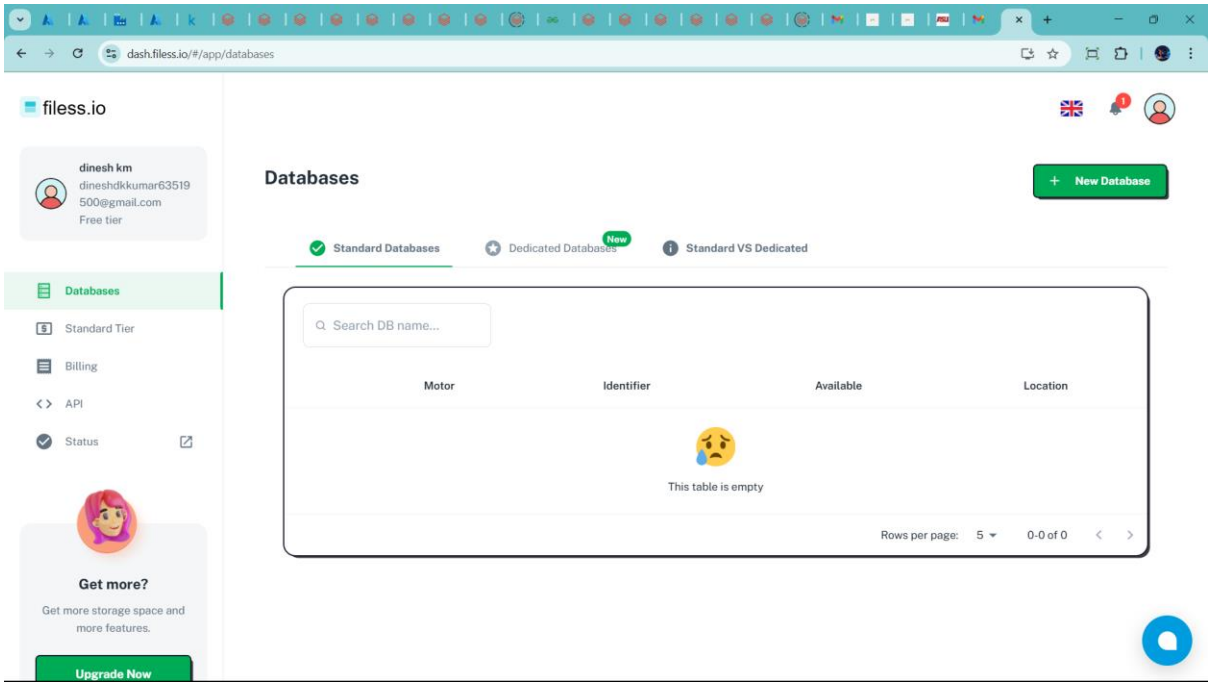


→use one dataset in SQL as well

→use this website

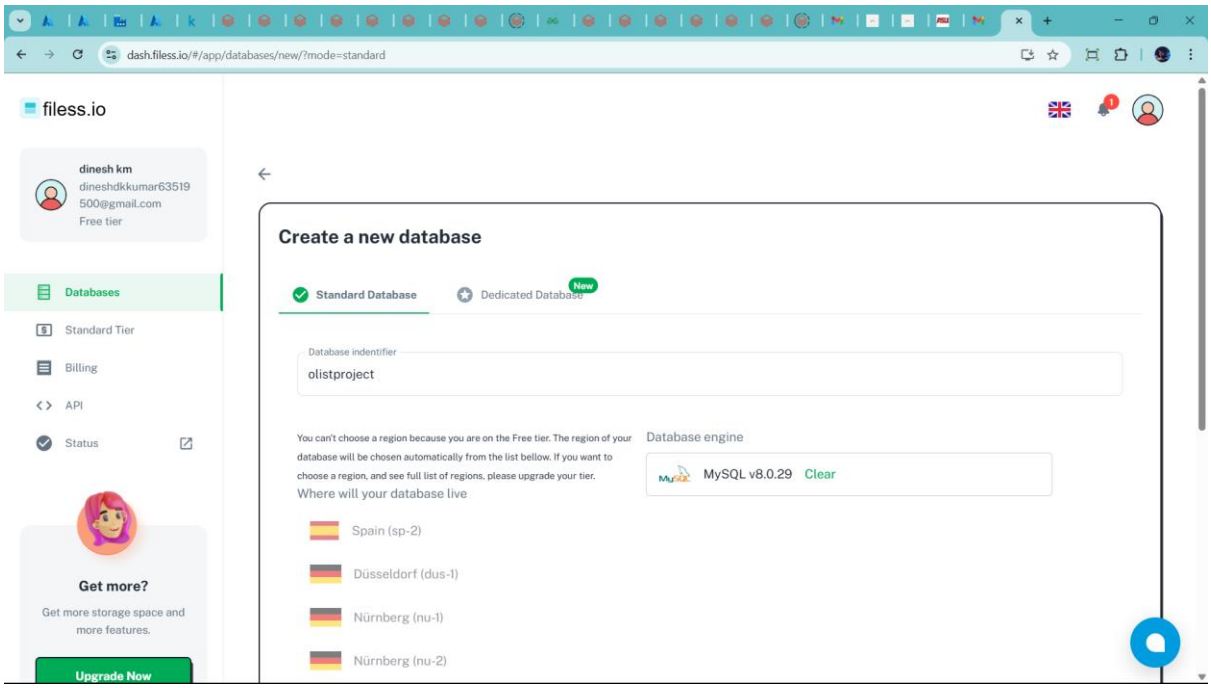


→Register and login



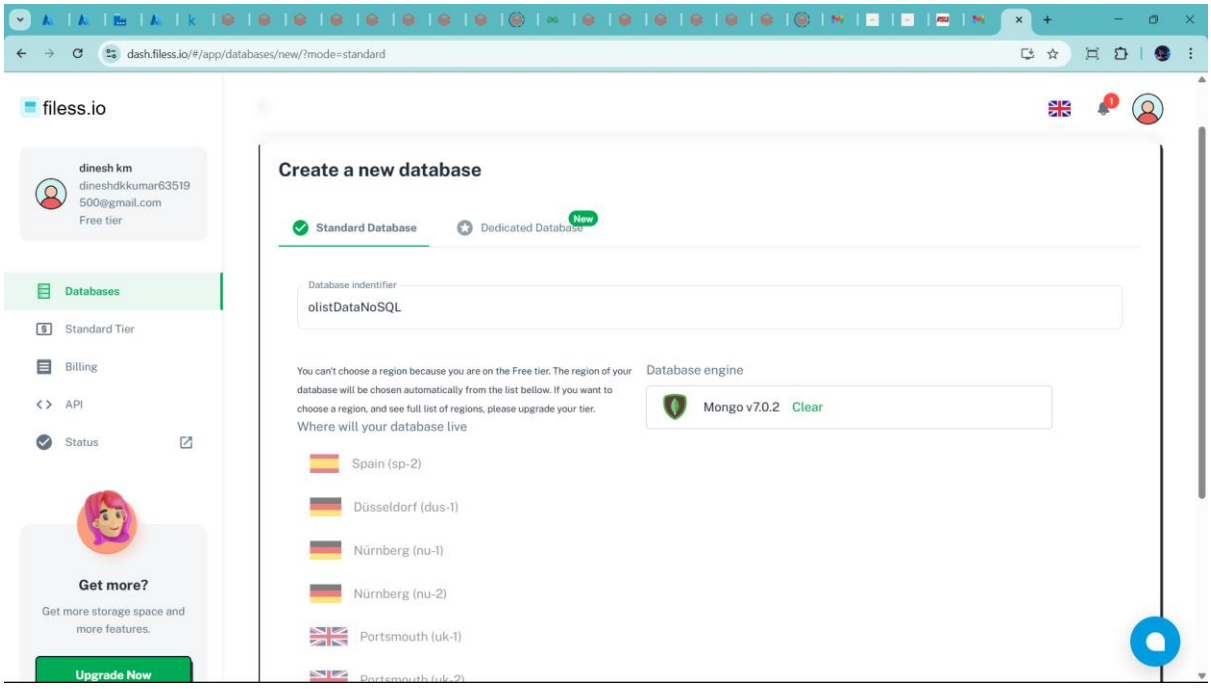
→Databases

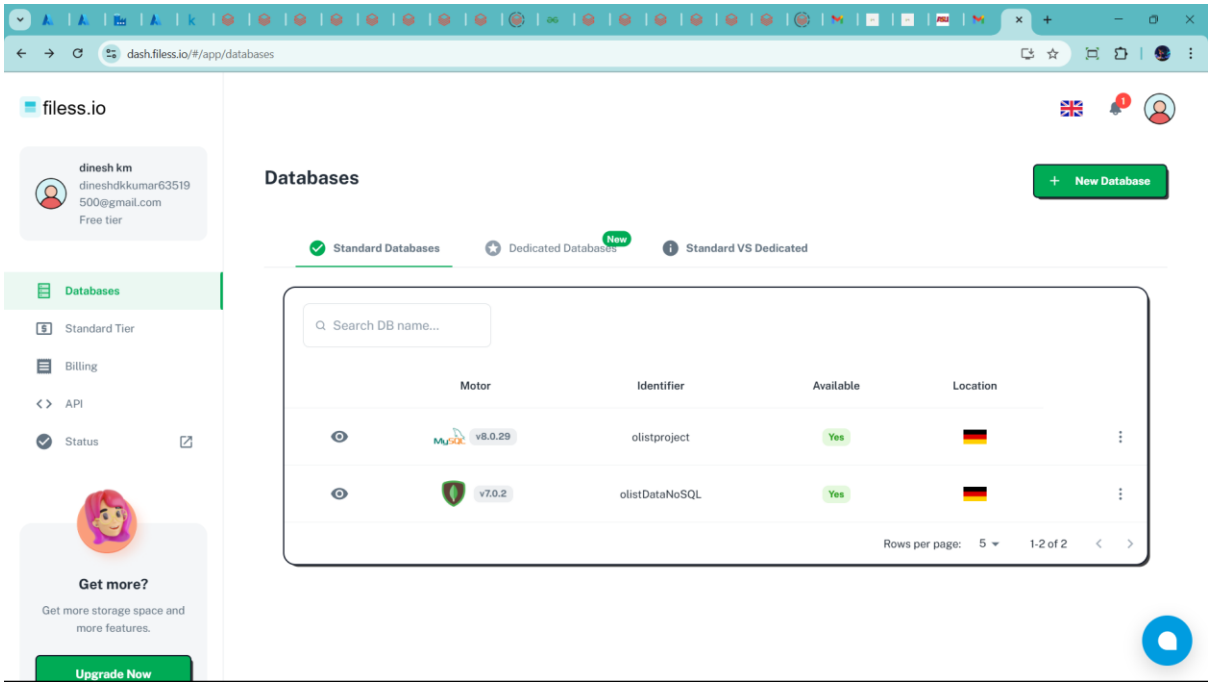
1) SQL 2)NoSQL 3)Https



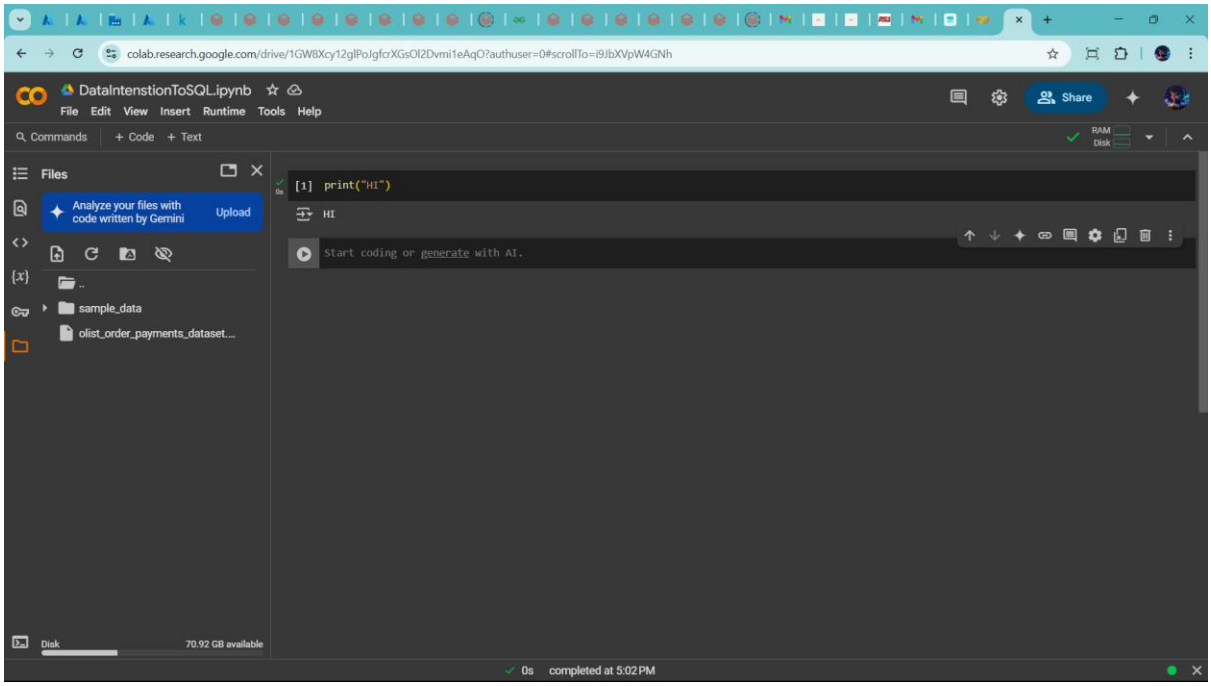
→Click create

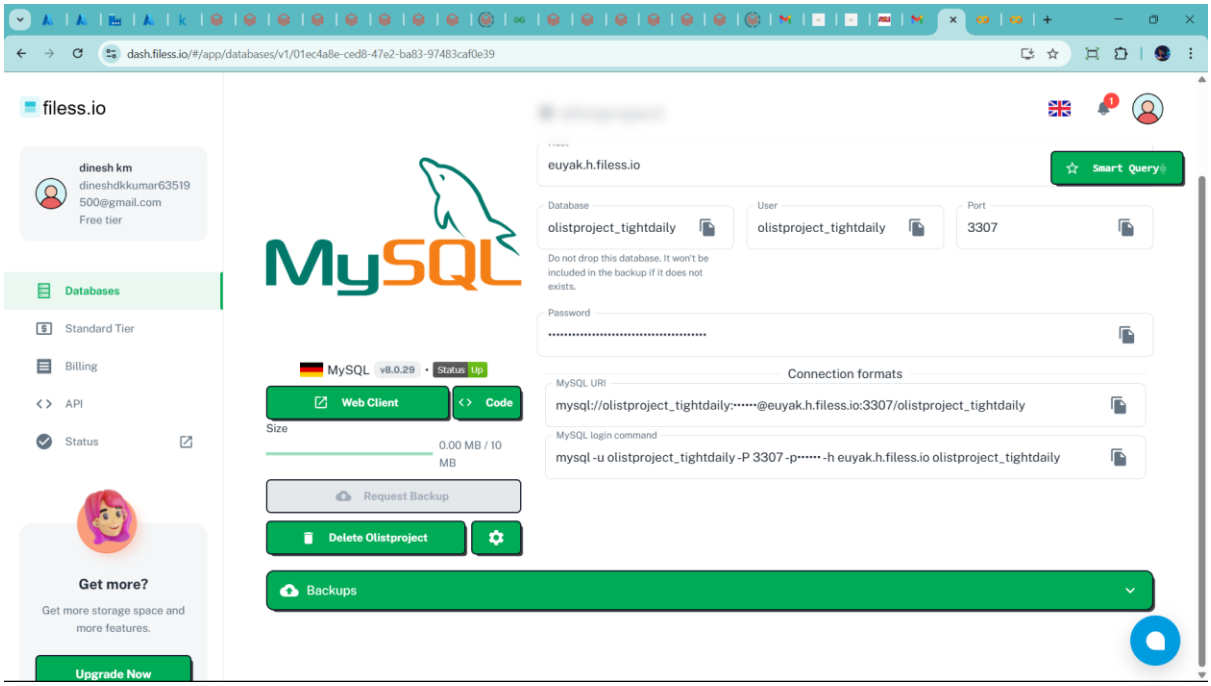
→Create same as mango DB



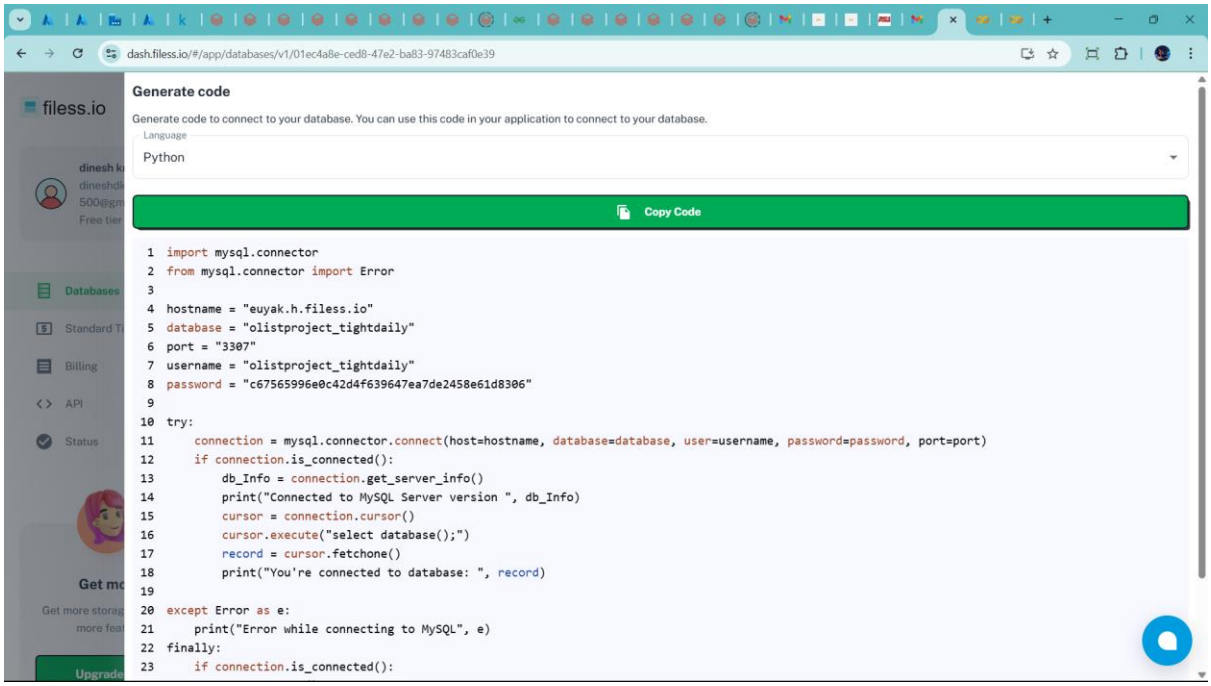


→use googlecolab to read data and write in SQL

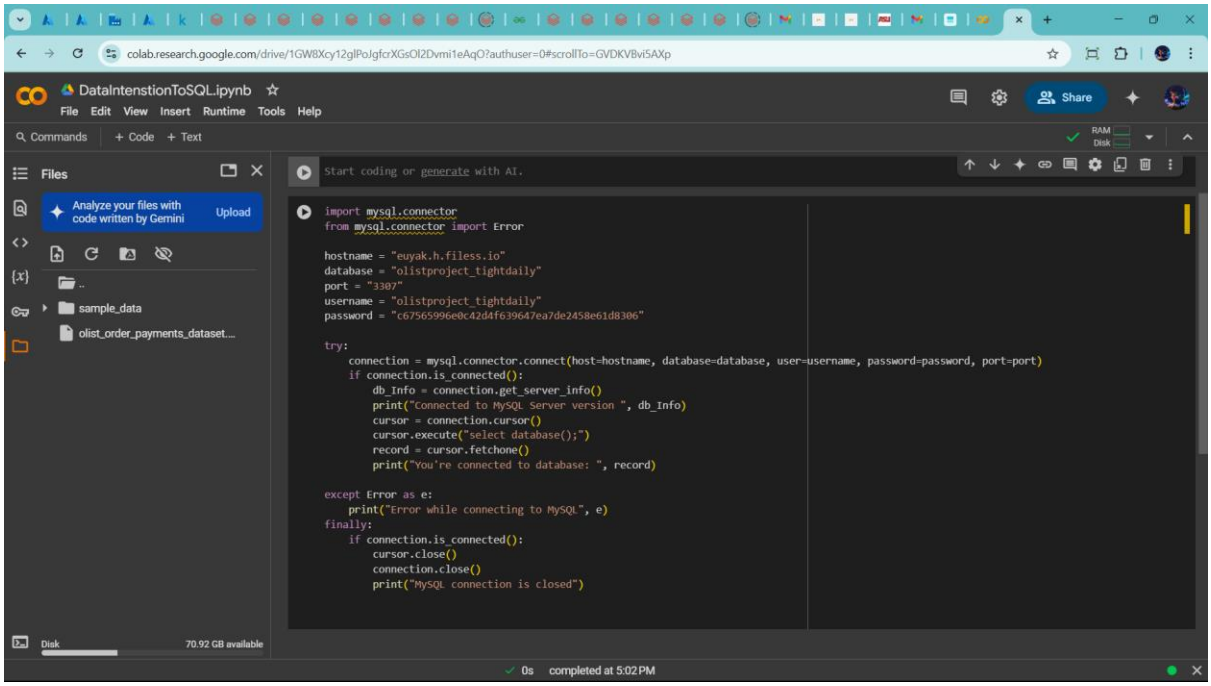


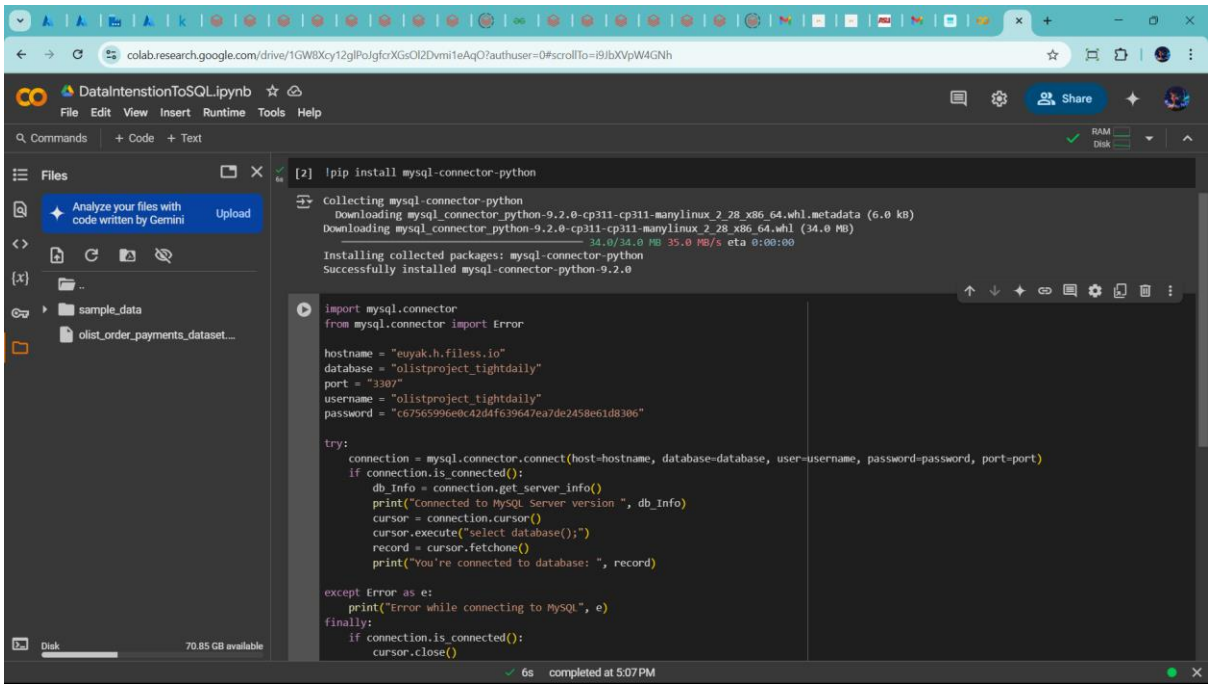


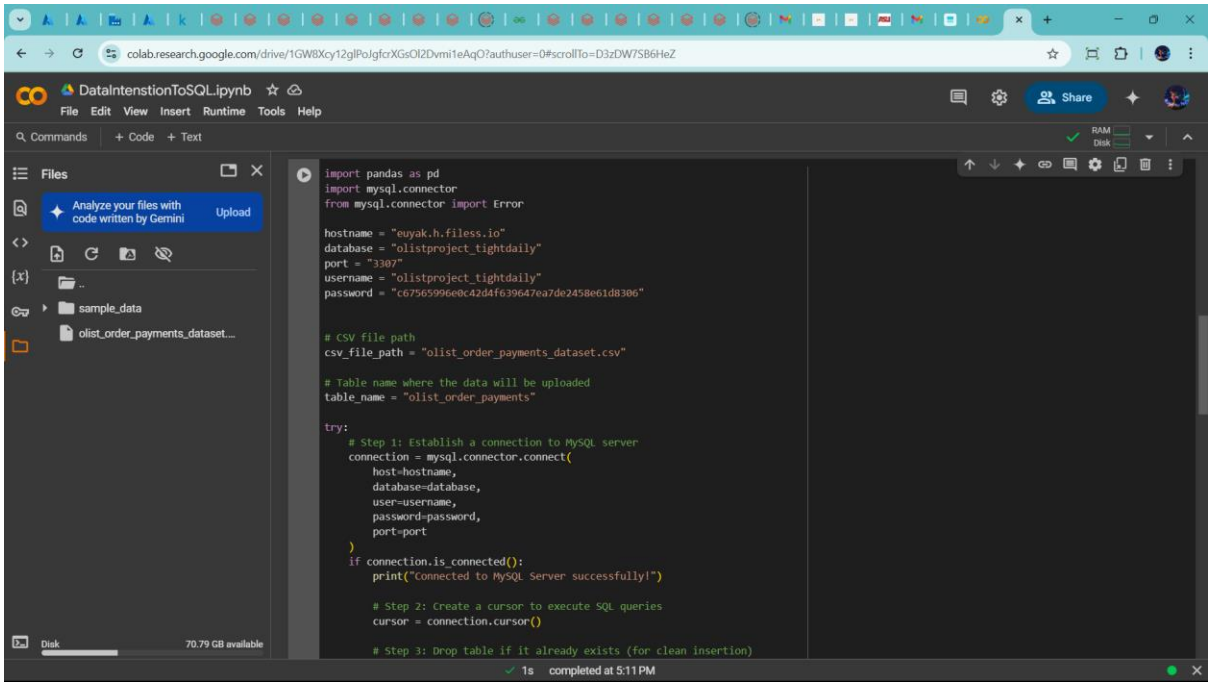
→click code and change python

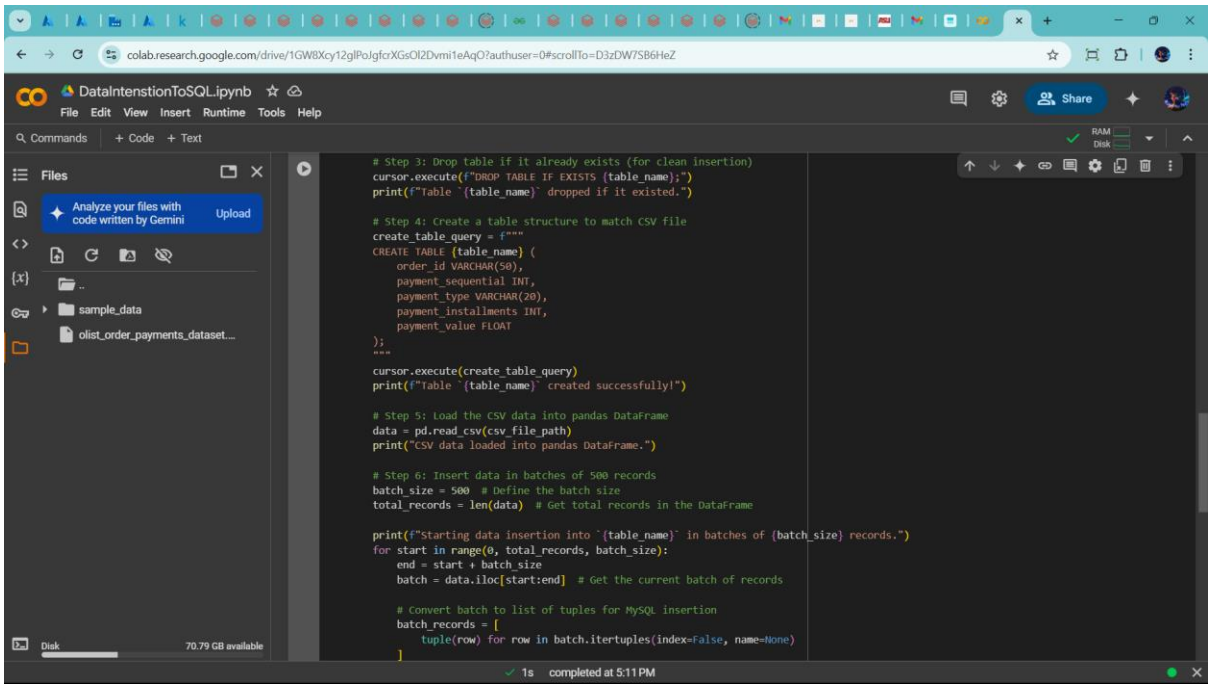


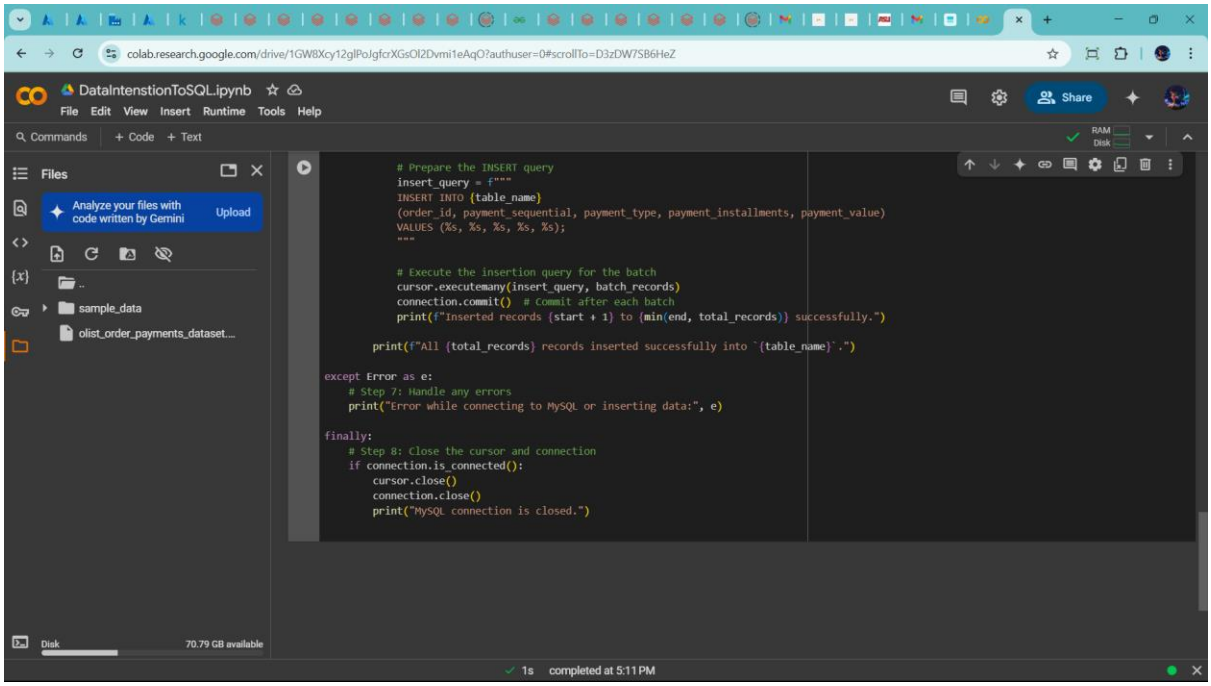
→Paste in colab

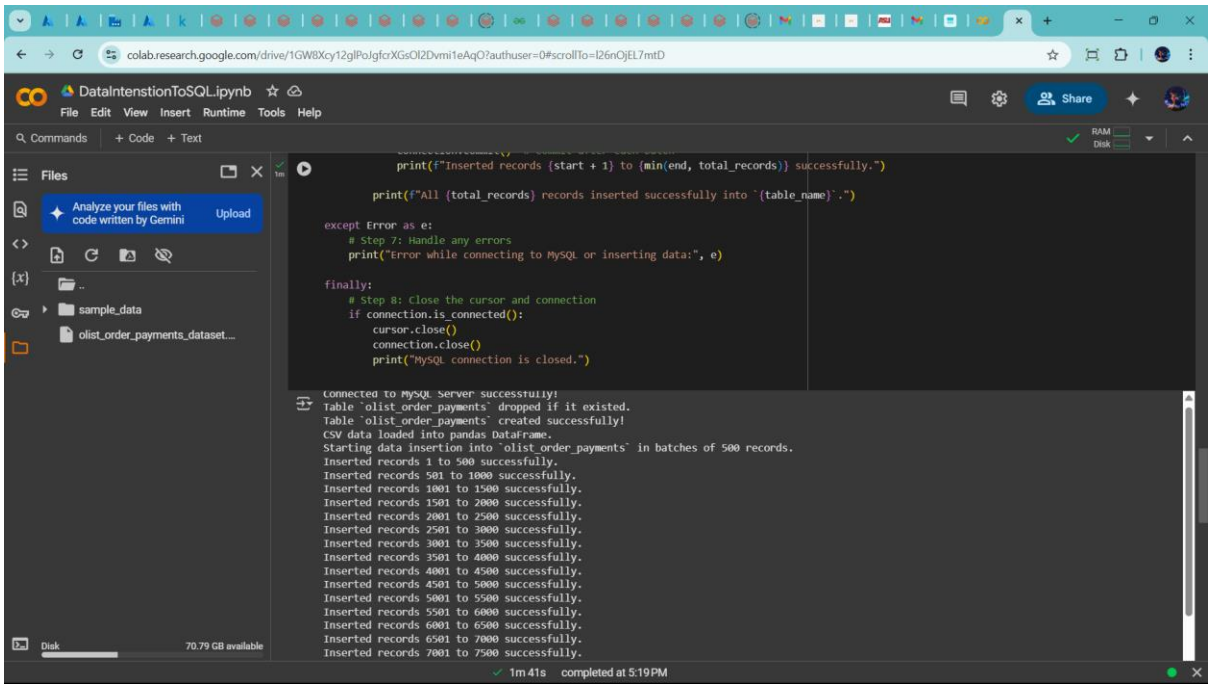

colab.research.google.com/drive/1GW8Xcy12glPoJgfcrXGsOl2Dvmi1eAqO?authuser=0#scrollTo=GVDKVBvi5AXp
DataIntenstionToSQL.ipynb
File Edit View Insert Runtime Tools Help
Share
Commands + Code + Text
Files
Analyze your files with code written by Gemini
Upload
sample_data
olist_order_payments_dataset....
Disk 70.92 GB available
Start coding or generate with AI.
import mysql.connector
from mysql.connector import Error
hostname = "euyak.h.filess.io"
database = "olistproject_tightdaily"
port = "3307"
username = "olistproject_tightdaily"
password = "c67565996e0c42d4f639647ea7de2458e61d8306"
try:
connection = mysql.connector.connect(host=hostname, database=database, user=username, password=password, port=port)
if connection.is_connected():
db_Info = connection.get_server_info()
print("Connected to MySQL Server version ", db_Info)
cursor = connection.cursor()
cursor.execute("select database();")
record = cursor.fetchone()
print("You're connected to database: ", record)
except Error as e:
print("Error while connecting to MySQL", e)
finally:
if connection.is_connected():
cursor.close()
connection.close()
print("MySQL connection is closed")
0s completed at 5:02 PM

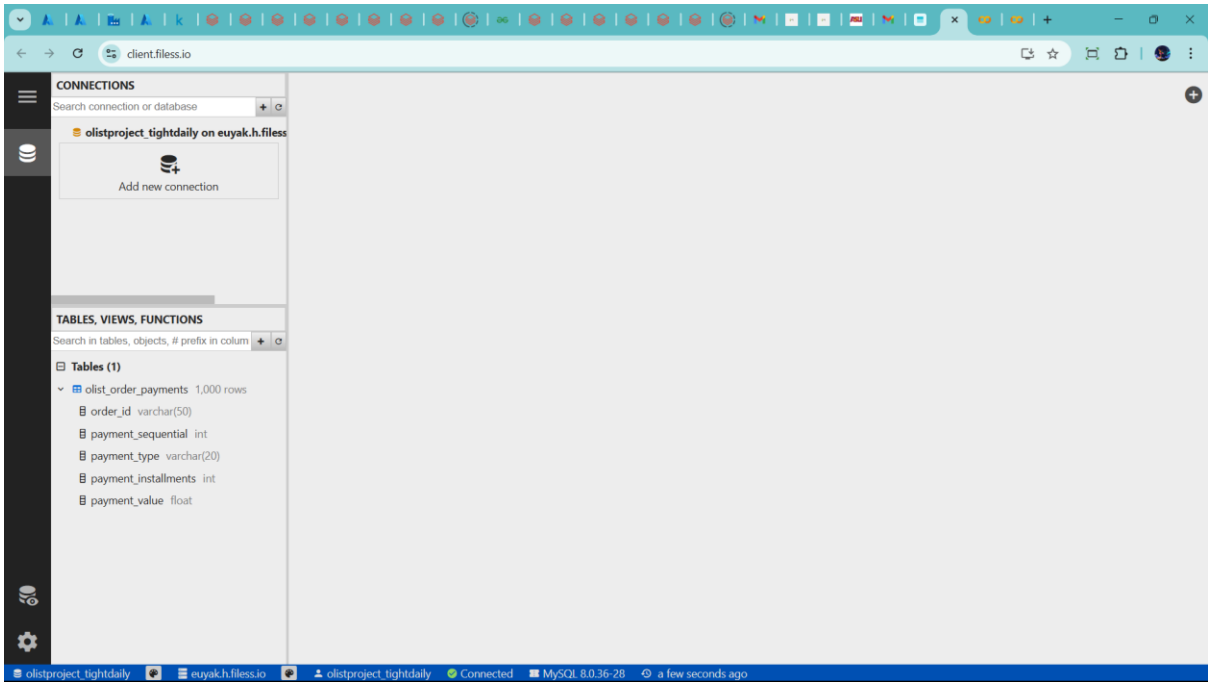

colab.research.google.com/drive/1GW8Xcy12glPoJgfcrXGsOl2Dvmi1eAqO?authuser=0#scrollTo=i9JbXVpW4GNh
DataIntenstionToSQL.ipynb
File Edit View Insert Runtime Tools Help
Share
Commands + Code + Text
Files
Analyze your files with code written by Gemini
Upload
sample_data
olist_order_payments_dataset....
Disk 70.85 GB available
[2] !pip install mysql-connector-python
Collecting mysql-connector-python
Downloading mysql_connector_python-9.2.0-cp311-cp311-manylinux_2_28_x86_64.whl.metadata (6.0 kB)
Downloading mysql_connector_python-9.2.0-cp311-cp311-manylinux_2_28_x86_64.whl (34.0 MB)
34.0/34.0 MB 35.0 MB/s eta 0:00:00
Installing collected packages: mysql-connector-python
Successfully installed mysql-connector-python-9.2.0
import mysql.connector
from mysql.connector import Error
hostname = "euyak.h.filess.io"
database = "olistproject_tightdaily"
port = "3307"
username = "olistproject_tightdaily"
password = "c67565996e0c42d4f639647ea7de2458e61d8306"
try:
connection = mysql.connector.connect(host=hostname, database=database, user=username, password=password, port=port)
if connection.is_connected():
db_Info = connection.get_server_info()
print("Connected to MySQL Server version ", db_Info)
cursor = connection.cursor()
cursor.execute("select database();")
record = cursor.fetchone()
print("You're connected to database: ", record)
except Error as e:
print("Error while connecting to MySQL", e)
finally:
if connection.is_connected():
cursor.close()
6s completed at 5:07 PM

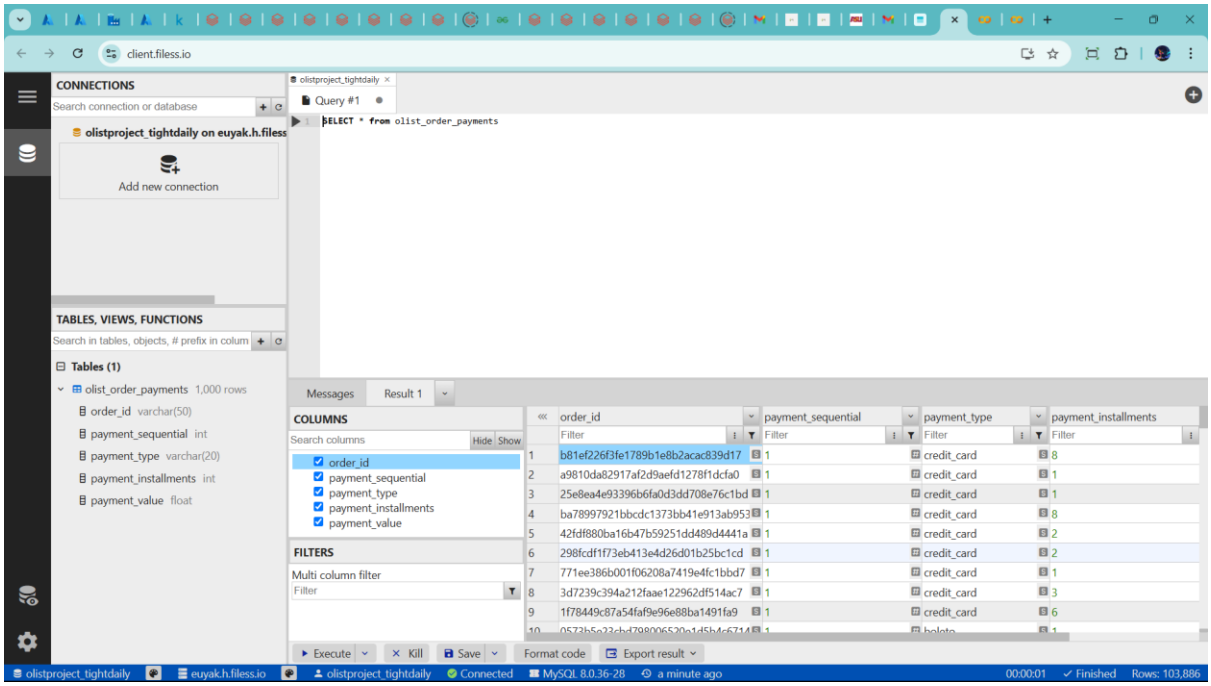

colab.research.google.com/drive/1GW8Xcy12glPoJgfcrXGsOl2Dvmi1eAqO?authuser=0#scrollTo=D3zDW7SB6HeZ
DataIntenstionToSQL.ipynb
File Edit View Insert Runtime Tools Help
Share
Commands + Code + Text
RAM
Disk
Files
Analyze your files with code written by Gemini
Upload
sample_data
olist_order_payments_dataset....
Disk 70.79 GB available
import pandas as pd
import mysql.connector
from mysql.connector import Error
hostname = "euyak.h.filess.io"
database = "olistproject_tightdaily"
port = "3307"
username = "olistproject_tightdaily"
password = "c67565996e0c42d4f639647ea7de2458e61d8306"
# CSV file path
csv_file_path = "olist_order_payments_dataset.csv"
# Table name where the data will be uploaded
table_name = "olist_order_payments"
try:
# Step 1: Establish a connection to MySQL server
connection = mysql.connector.connect(
host=hostname,
database=database,
user=username,
password=password,
port=port
)
if connection.is_connected():
print("Connected to MySQL Server successfully!")
# Step 2: Create a cursor to execute SQL queries
cursor = connection.cursor()
# Step 3: Drop table if it already exists (for clean insertion)
1s completed at 5:11 PM


colab.research.google.com/drive/1GW8Xcy12glPoJgfcrXGsOl2Dvmi1eAqO?authuser=0#scrollTo=D3zDW7SB6HeZ
DataIntenstionToSQL.ipynb
File Edit View Insert Runtime Tools Help
Share
Commands + Code + Text
RAM
Disk
Files
Analyze your files with code written by Gemini
Upload
sample_data
olist_order_payments_dataset....
Disk 70.79 GB available
# Step 3: Drop table if it already exists (for clean insertion)
cursor.execute(f"DROP TABLE IF EXISTS {table_name};")
print(f"Table `{table_name}` dropped if it existed.")
# Step 4: Create a table structure to match CSV file
create_table_query = f"""
CREATE TABLE {table_name} (
order_id VARCHAR(50),
payment_sequential INT,
payment_type VARCHAR(20),
payment_installments INT,
payment_value FLOAT
);
"""
cursor.execute(create_table_query)
print(f"Table `{table_name}` created successfully!")
# Step 5: Load the CSV data into pandas DataFrame
data = pd.read_csv(csv_file_path)
print("CSV data loaded into pandas DataFrame.")
# Step 6: Insert data in batches of 500 records
batch_size = 500 # Define the batch size
total_records = len(data) # Get total records in the DataFrame
print(f"Starting data insertion into `{table_name}` in batches of {batch_size} records.")
for start in range(0, total_records, batch_size):
end = start + batch_size
batch = data.iloc[start:end] # Get the current batch of records
# Convert batch to list of tuples for MySQL insertion
batch_records = [
tuple(row) for row in batch.itertuples(index=False, name=None)
]
1s completed at 5:11 PM

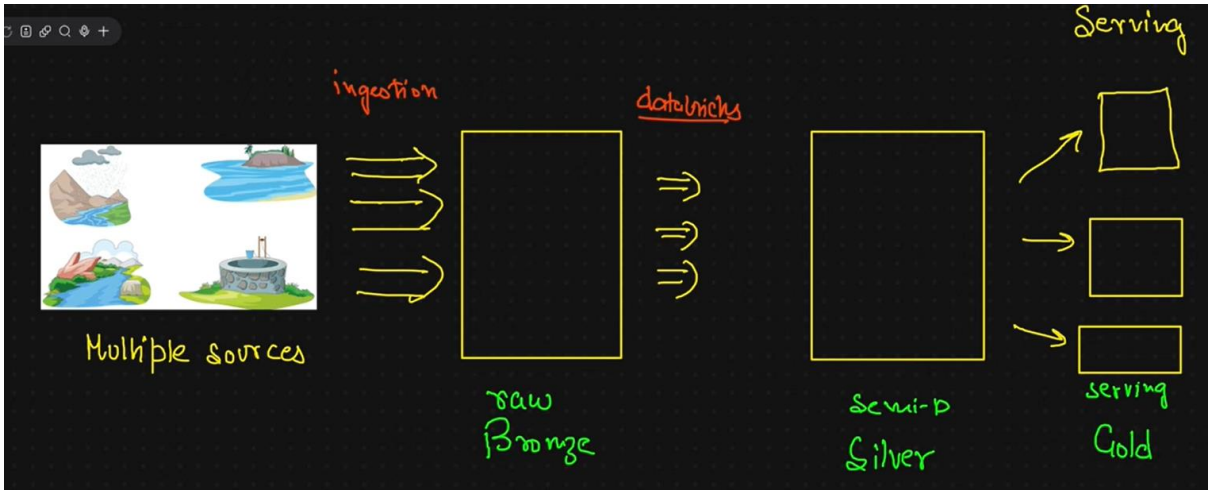

colab.research.google.com/drive/1GW8Xcy12glPoJgfcrXGsOl2Dvmi1eAqO?authuser=0#scrollTo=D3zDW7SB6HeZ
DataIntenstionToSQL.ipynb
File Edit View Insert Runtime Tools Help
Share
Commands + Code + Text
Files
Analyze your files with code written by Gemini
Upload
sample_data
olist_order_payments_dataset....
Disk 70.79 GB available
# Prepare the INSERT query
insert_query = f"""
INSERT INTO {table_name}
(order_id, payment_sequential, payment_type, payment_installments, payment_value)
VALUES (%s, %s, %s, %s, %s);
"""
# Execute the insertion query for the batch
cursor.executemany(insert_query, batch_records)
connection.commit() # Commit after each batch
print(f"Inserted records {start + 1} to {min(end, total_records)} successfully.")
print(f"All {total_records} records inserted successfully into `{table_name}`.")
except Error as e:
# Step 7: Handle any errors
print("Error while connecting to MySQL or inserting data:", e)
finally:
# Step 8: Close the cursor and connection
if connection.is_connected():
cursor.close()
connection.close()
print("MySQL connection is closed.")
1s completed at 5:11 PM


colab.research.google.com/drive/1GW8Xcy12glPoJgfcrXGsOl2Dvmi1eAqO?authuser=0#scrollTo=I26nOjEL7mtD
DataIntenstionToSQL.ipynb
File Edit View Insert Runtime Tools Help
Share
Commands + Code + Text
Files
Analyze your files with code written by Gemini
Upload
sample_data
olist_order_payments_dataset....
Disk 70.79 GB available
print(f"Inserted records {start + 1} to {min(end, total_records)} successfully.")
print(f"All {total_records} records inserted successfully into `{table_name}`.")
except Error as e:
# Step 7: Handle any errors
print("Error while connecting to MySQL or inserting data:", e)
finally:
# Step 8: Close the cursor and connection
if connection.is_connected():
cursor.close()
connection.close()
print("MySQL connection is closed.")
Connected to MySQL Server successfully!
Table `olist_order_payments` dropped if it existed.
Table `olist_order_payments` created successfully!
CSV data loaded into pandas DataFrame.
Starting data insertion into `olist_order_payments` in batches of 500 records.
Inserted records 1 to 500 successfully.
Inserted records 501 to 1000 successfully.
Inserted records 1001 to 1500 successfully.
Inserted records 1501 to 2000 successfully.
Inserted records 2001 to 2500 successfully.
Inserted records 2501 to 3000 successfully.
Inserted records 3001 to 3500 successfully.
Inserted records 3501 to 4000 successfully.
Inserted records 4001 to 4500 successfully.
Inserted records 4501 to 5000 successfully.
Inserted records 5001 to 5500 successfully.
Inserted records 5501 to 6000 successfully.
Inserted records 6001 to 6500 successfully.
Inserted records 6501 to 7000 successfully.
Inserted records 7001 to 7500 successfully.
1m 41s completed at 5:19 PM

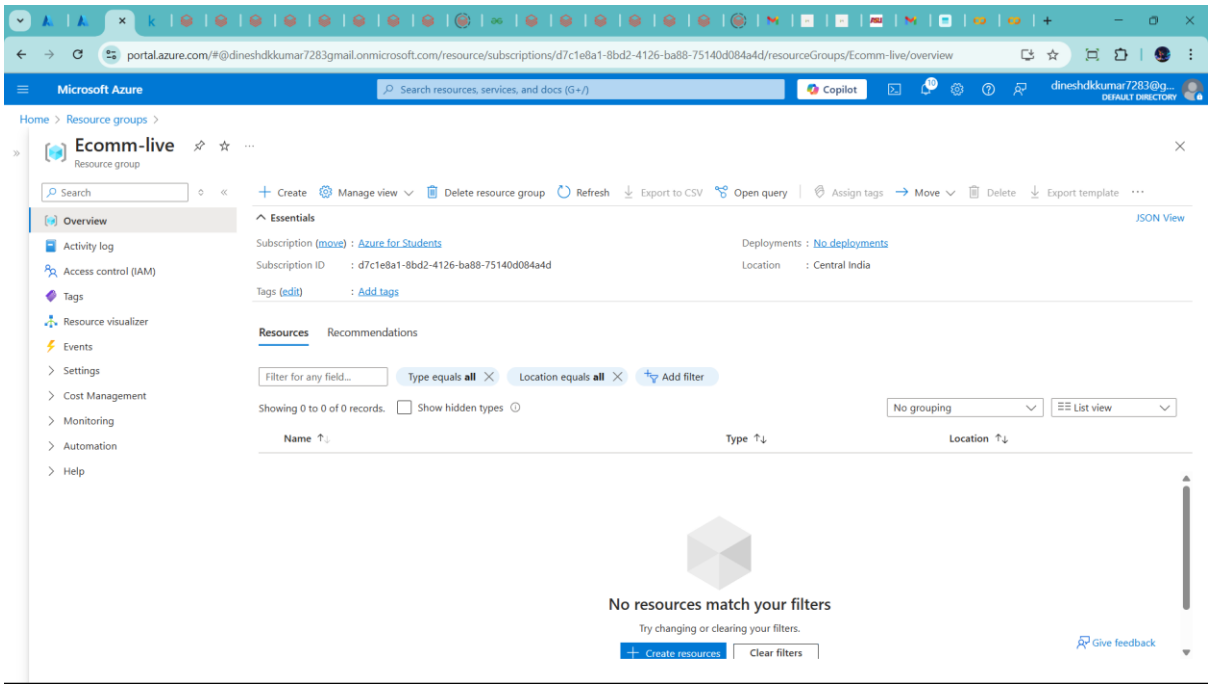




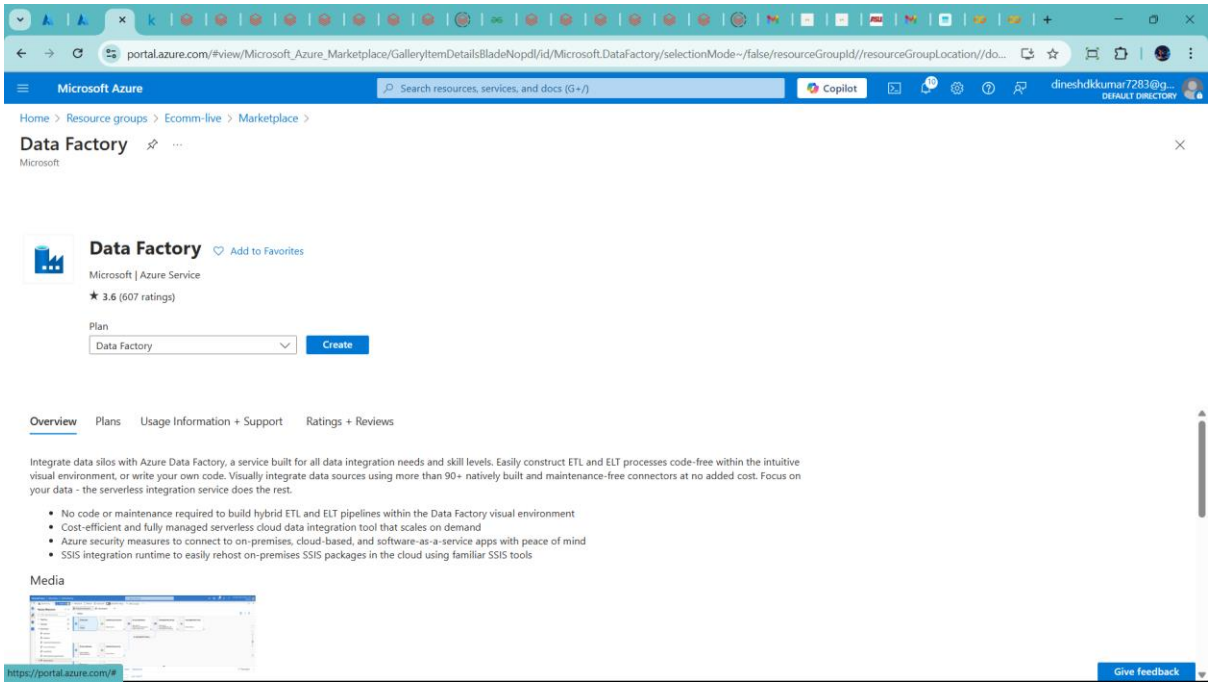
Create pipeline form different lake to store some where

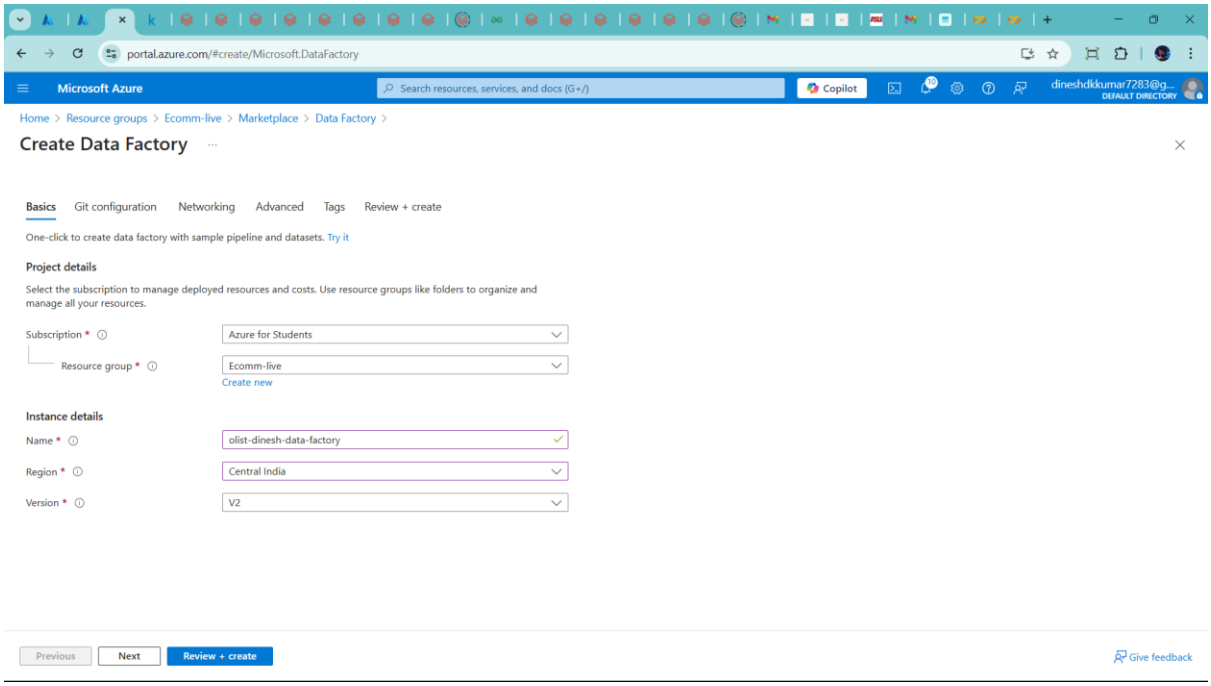


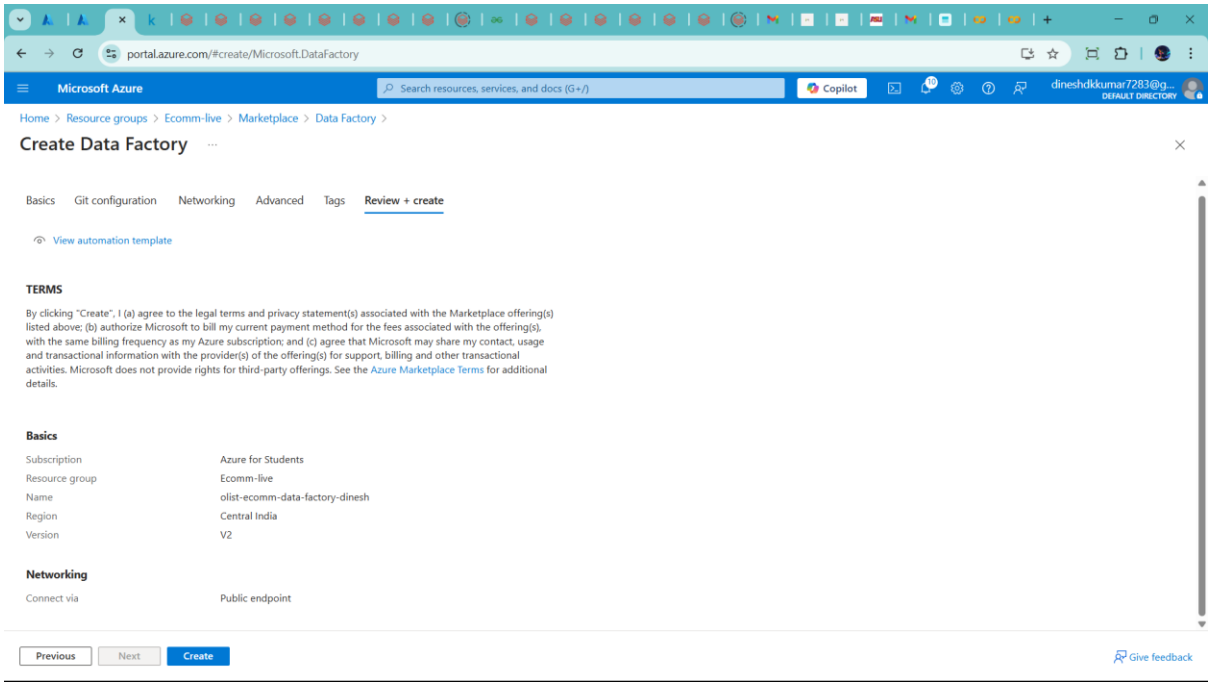
Go to Azure

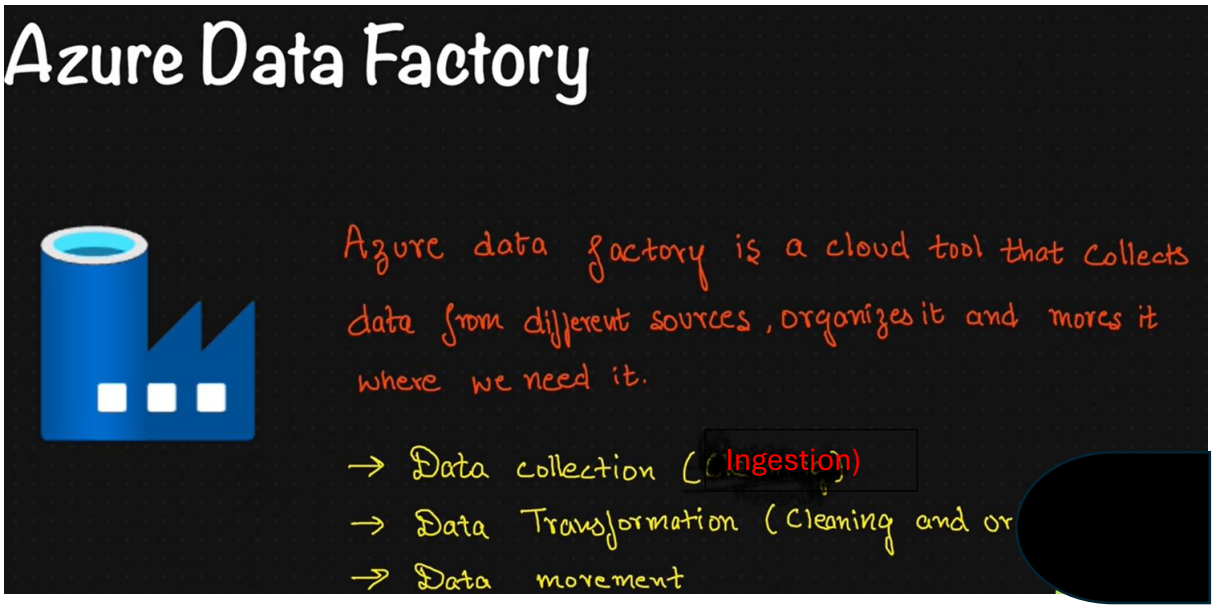


→create datafactory

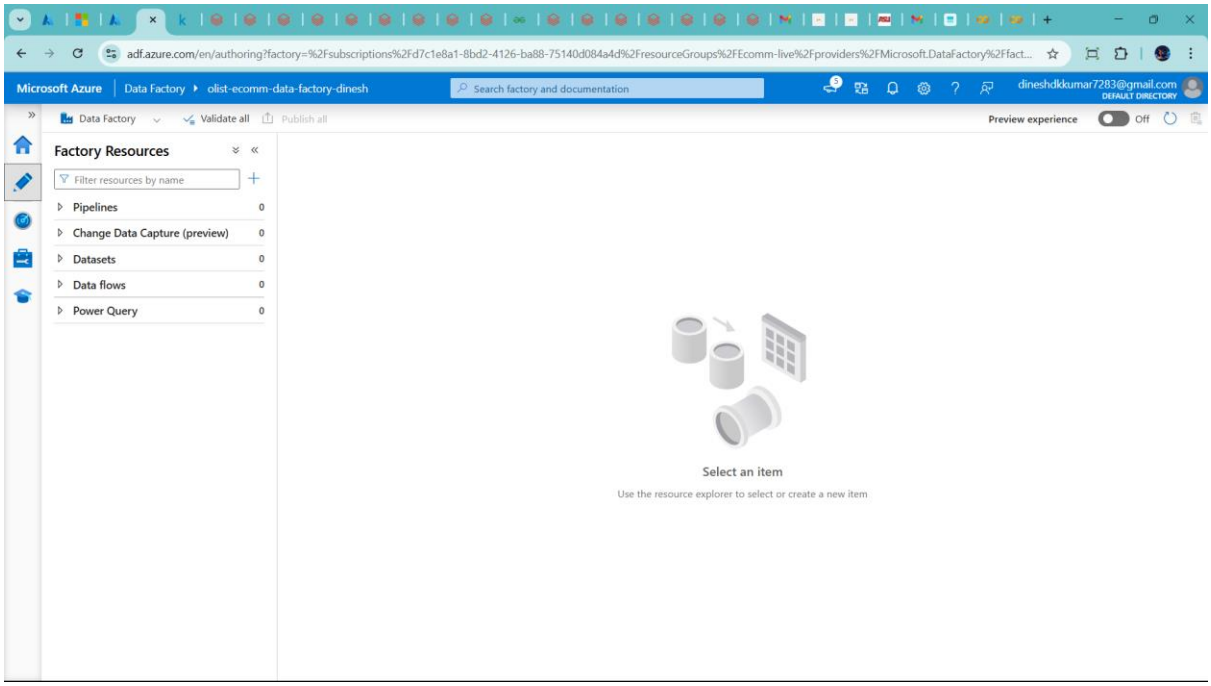

portal.azure.com/#view/Microsoft_Azure_Marketplace/GalleryItemDetailsBladeNopdl/id/Microsoft.DataFactory/selectionMode~/false/resourceGroupId//resourceGroupLocation//do...
Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
dineshdkkumar7283@g...
DEFAULT DIRECTORY
Home > Resource groups > Ecomm-live > Marketplace >
Data Factory
Microsoft
Data Factory
Add to Favorites
Microsoft | Azure Service
★ 3.6 (607 ratings)
Plan
Data Factory
Create
Overview
Plans
Usage Information + Support
Ratings + Reviews
Integrate data silos with Azure Data Factory, a service built for all data integration needs and skill levels. Easily construct ETL and ELT processes code-free within the intuitive visual environment, or write your own code. Visually integrate data sources using more than 90+ natively built and maintenance-free connectors at no added cost. Focus on your data - the serverless integration service does the rest.
No code or maintenance required to build hybrid ETL and ELT pipelines within the Data Factory visual environment
Cost-efficient and fully managed serverless cloud data integration tool that scales on demand
Azure security measures to connect to on-premises, cloud-based, and software-as-a-service apps with peace of mind
SSIS integration runtime to easily rehost on-premises SSIS packages in the cloud using familiar SSIS tools
Media
https://portal.azure.com/#
Give feedback

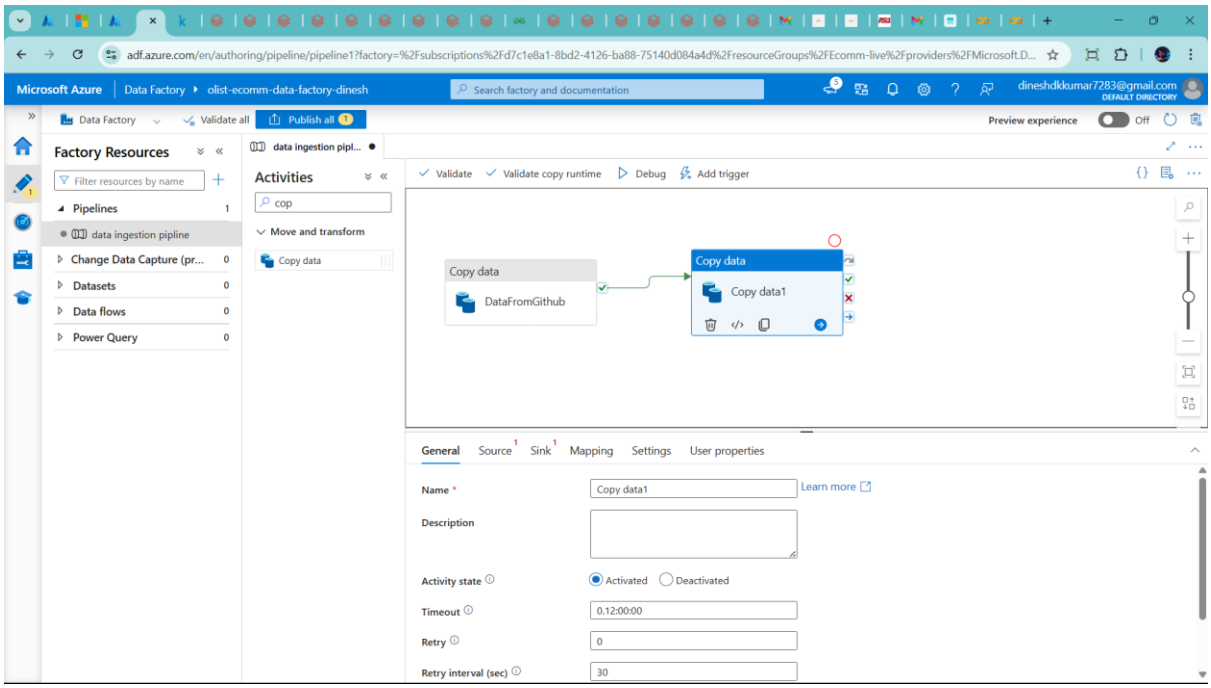

portal.azure.com/#create/Microsoft.DataFactory
Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
dineshdkkumar7283@g...
DEFAULT DIRECTORY
Home > Resource groups > Ecomm-live > Marketplace > Data Factory >
Create Data Factory
Basics
Git configuration
Networking
Advanced
Tags
Review + create
One-click to create data factory with sample pipeline and datasets. Try it
Project details
Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.
Subscription *
Azure for Students
Resource group *
Ecomm-live
Create new
Instance details
Name *
olist-dinesh-data-factory
Region *
Central India
Version *
V2
Previous
Next
Review + create
Give feedback





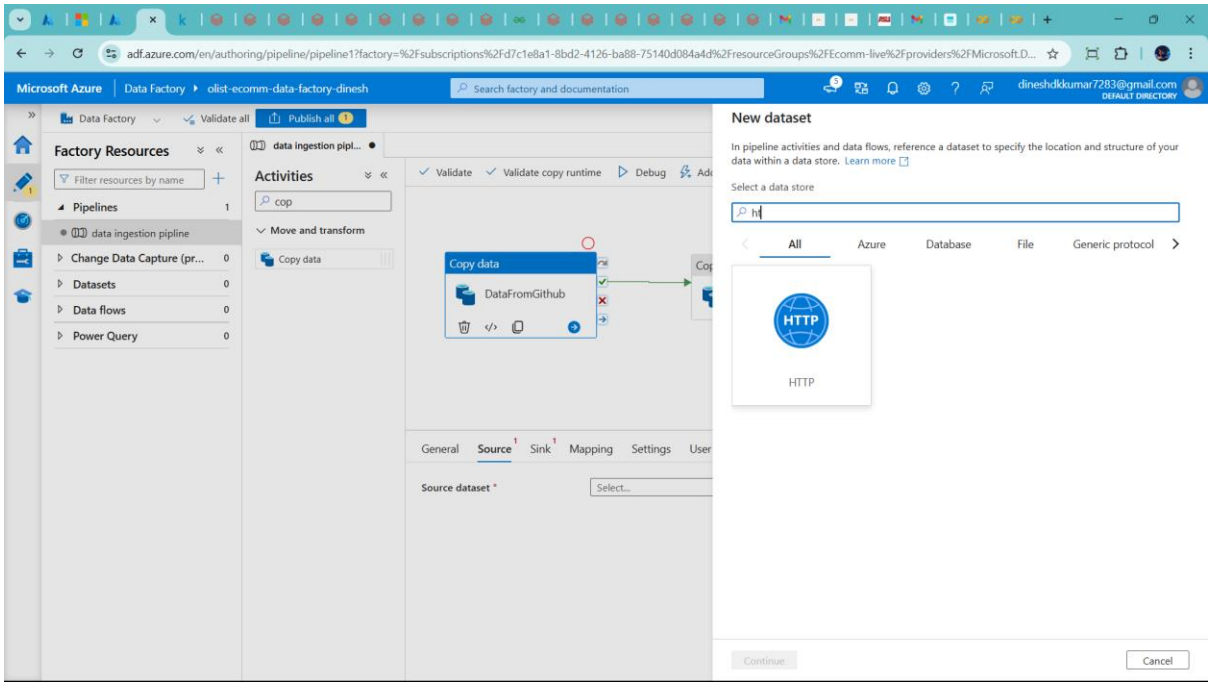
-→Launch Data Factory

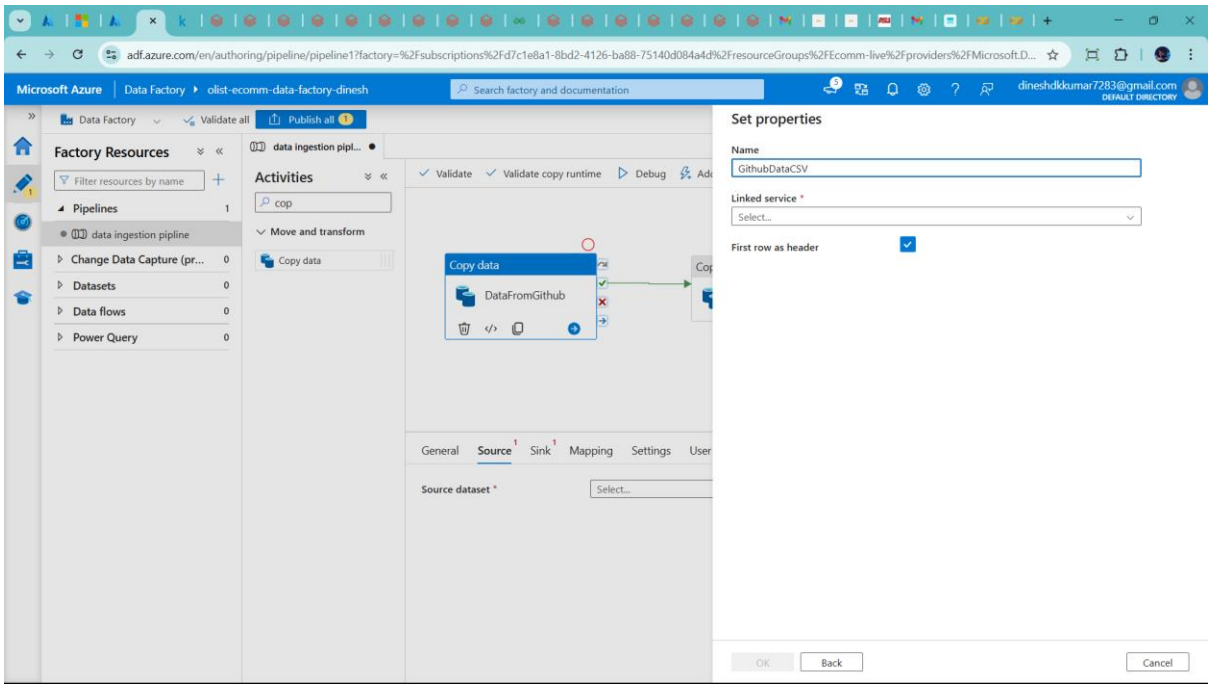




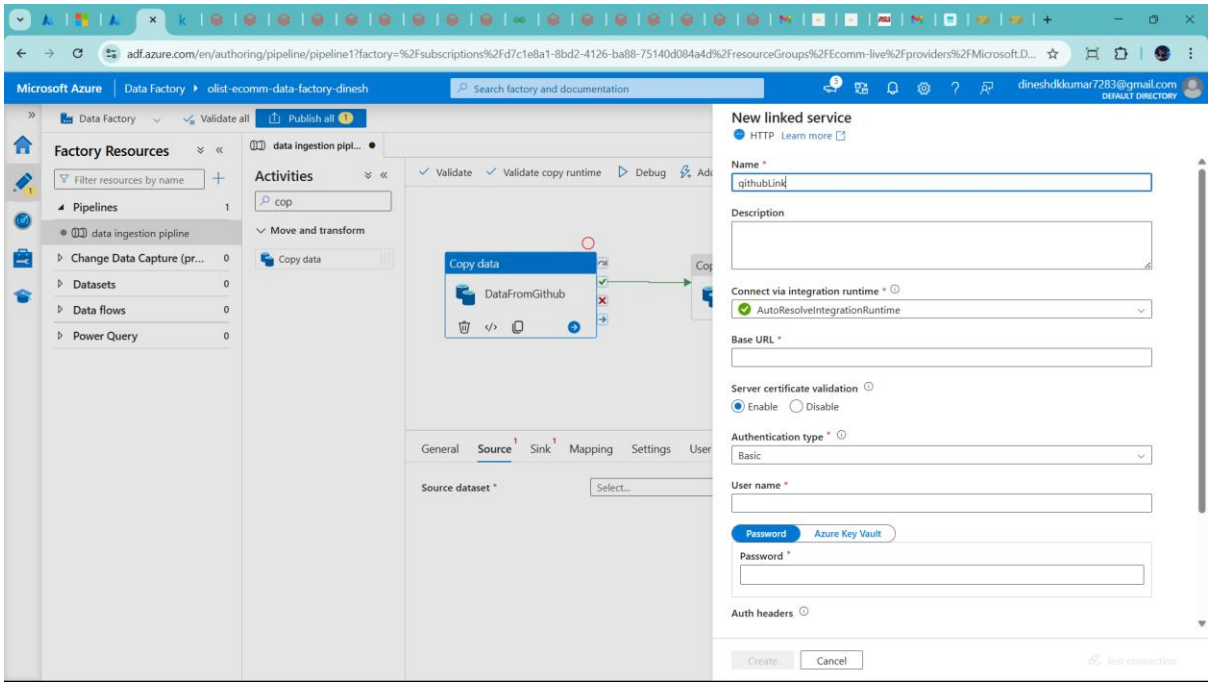
→Click Source →click + new

→choose https →csv

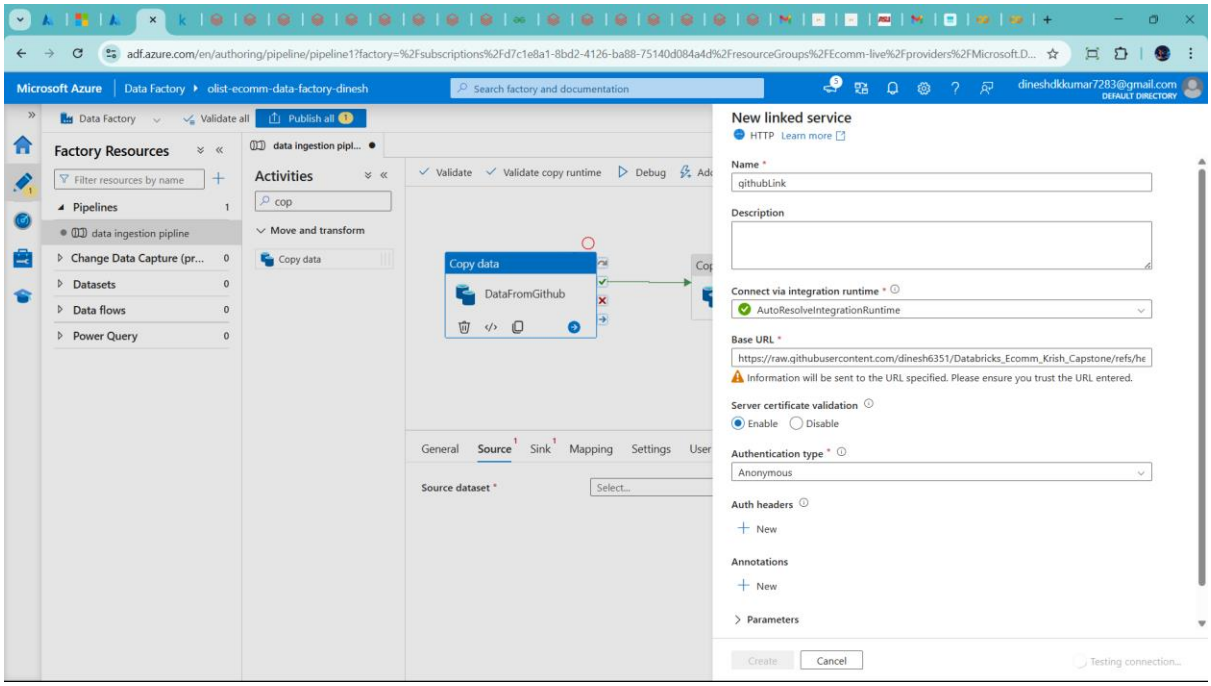


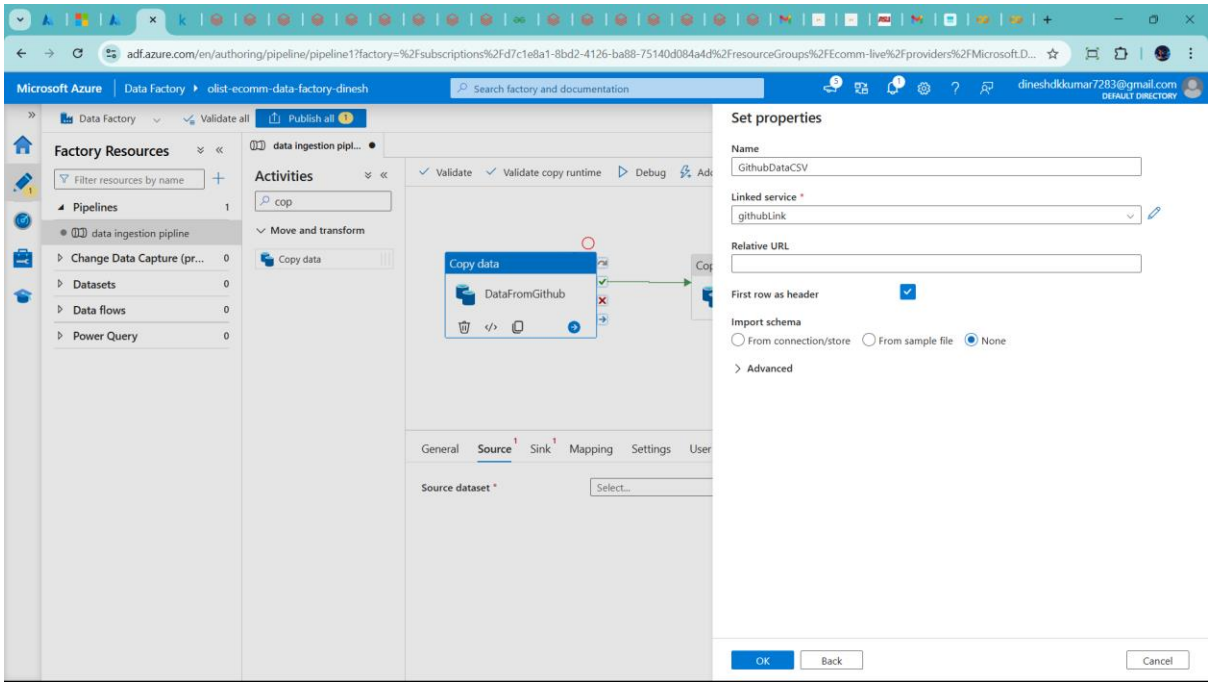


→click linked service -> +New

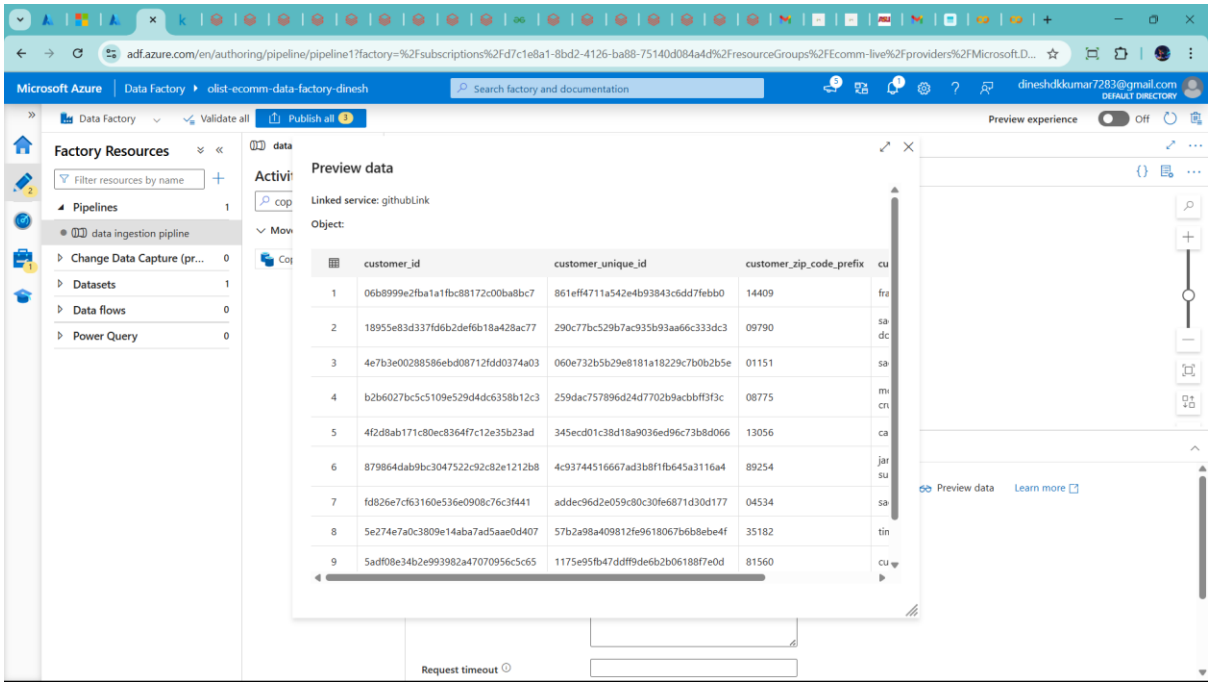


→in this base url add full url link

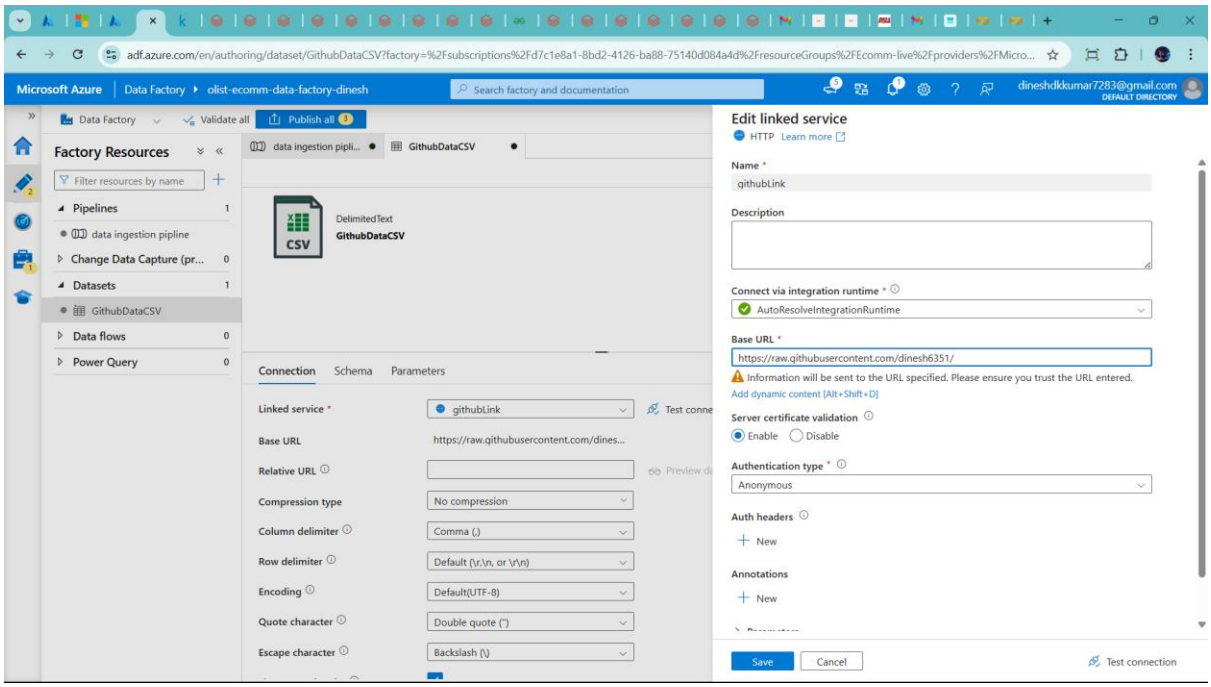


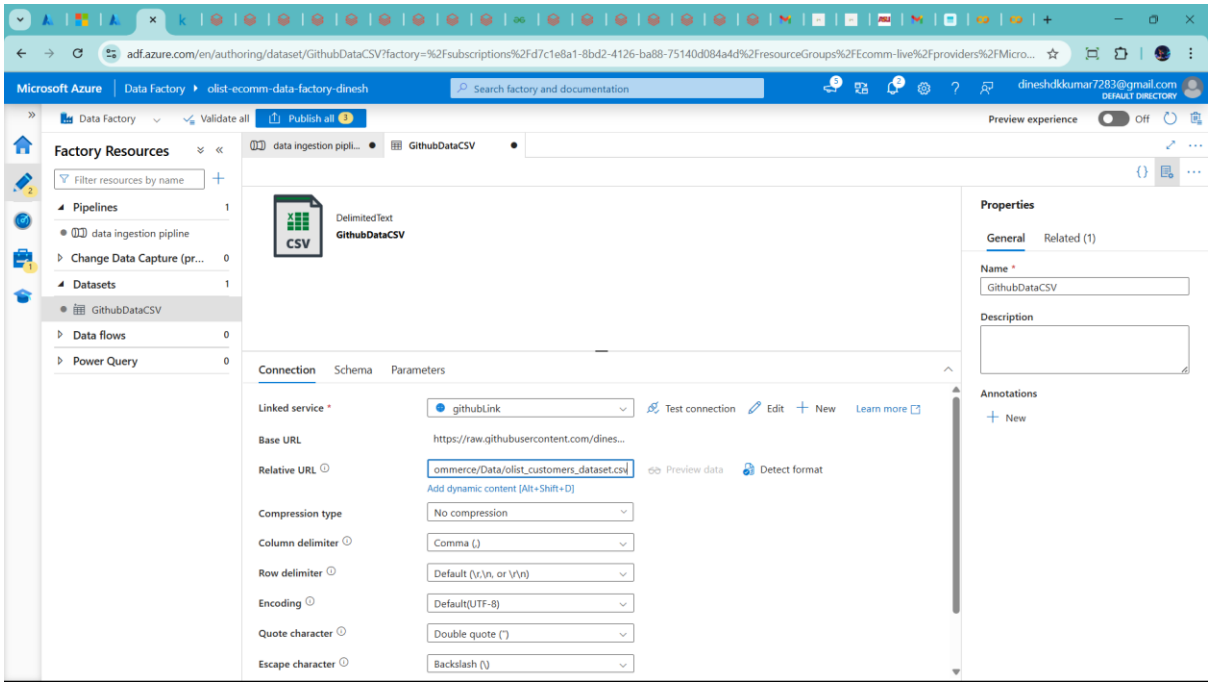

Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (pr...
Datasets
Data flows
Power Query
Activities
cop
Move and transform
Copy data
Validate
Validate copy runtime
Debug
Copy data
DataFromGithub
General
Source
Sink
Mapping
Settings
Source dataset *
Select...
Set properties
Name
GithubDataCSV
Linked service *
githubLink
Relative URL
First row as header
Import schema
From connection/store
From sample file
None
Advanced
OK
Back
Cancel

→Click ok


Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (pr...
Datasets
Data flows
Power Query
Preview data
Linked service: githubLink
Object:
customer_id
customer_unique_id
customer_zip_code_prefix
1 06b8999e2fba1a1fbc88172c00ba8bc7 861eff4711a542e4b93843c6dd7febb0 14409
2 18955e83d337fd6b2def6b18a428ac77 290c77bc529b7ac935b93aa66c333dc3 09790
3 4e7b3e00288586ebd08712fdd0374a03 060e732b5b29e8181a18229c7b0b2b5e 01151
4 b2b6027bc5c5109e529d4dc6358b12c3 259dac757896d24d7702b9acbbff3f3c 08775
5 4f2d8ab171c80ec8364f7c12e35b23ad 345ecd01c38d18a9036ed96c73b8d066 13056
6 879864dab9bc3047522c92c82e1212b8 4c93744516667ad3b8f1fb645a3116a4 89254
7 fd826e7cf63160e536e0908c76c3f441 addec96d2e059c80c30fe6871d30d177 04534
8 5e274e7a0c3809e14aba7ad5aae0d407 57b2a98a409812fe9618067b6b8ebe4f 35182
9 5adf08e34b2e993982a47070956c5c65 1175e95fb47ddff9de6b2b06188f7e0d 81560
Preview data
Learn more
Request timeout

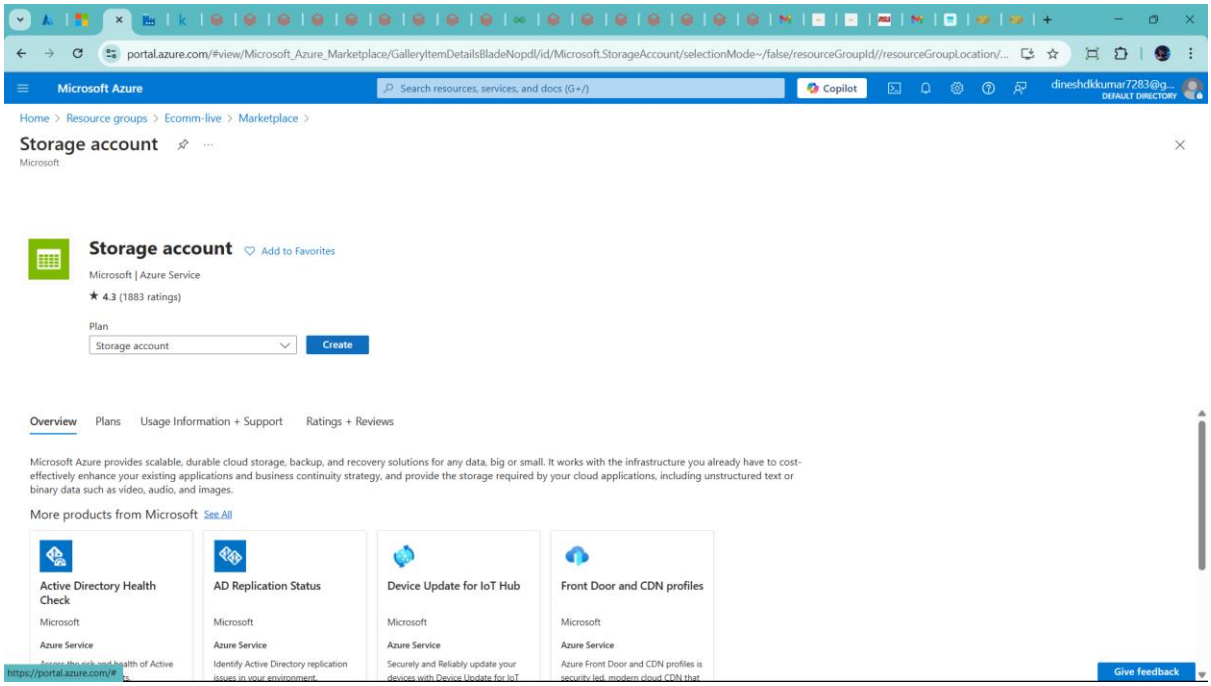
→Change it Base URL

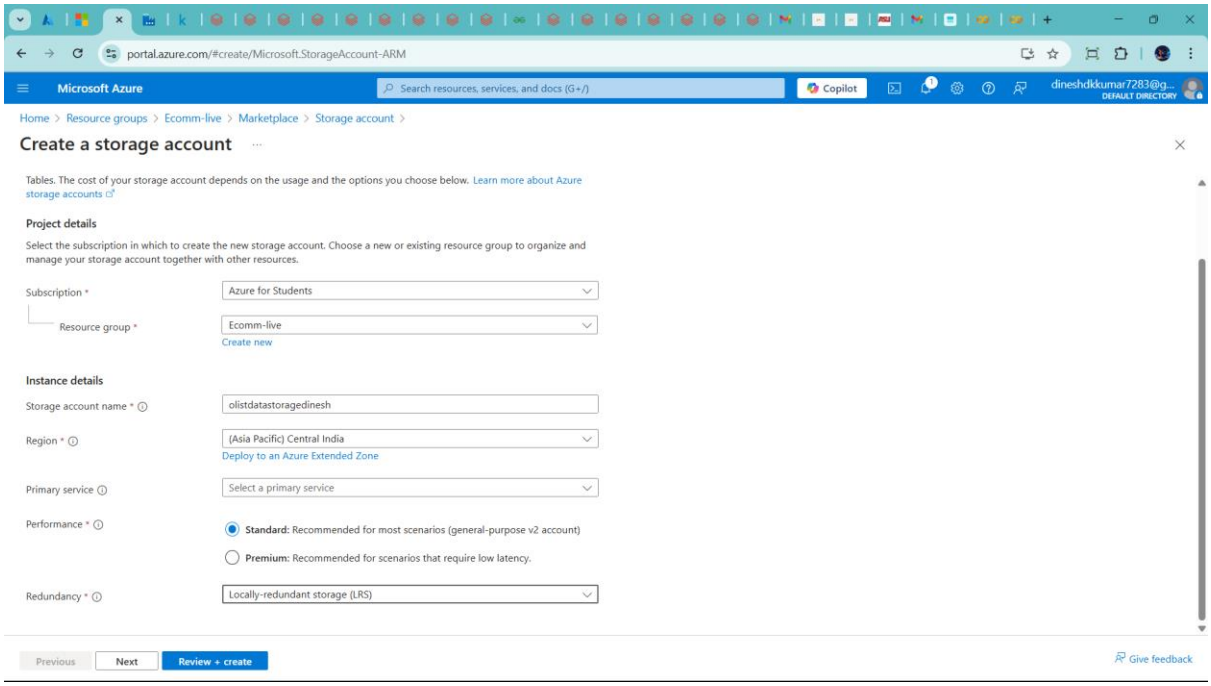




→where you to storage destination storage →Sink

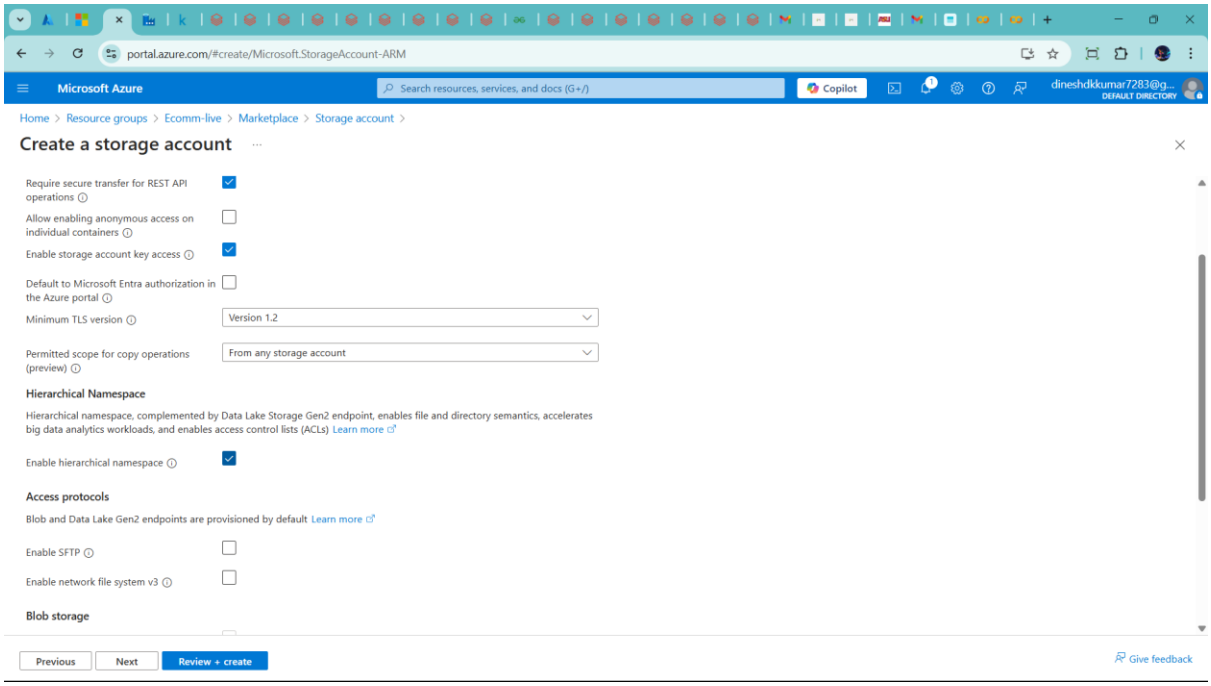
→Now create storage account



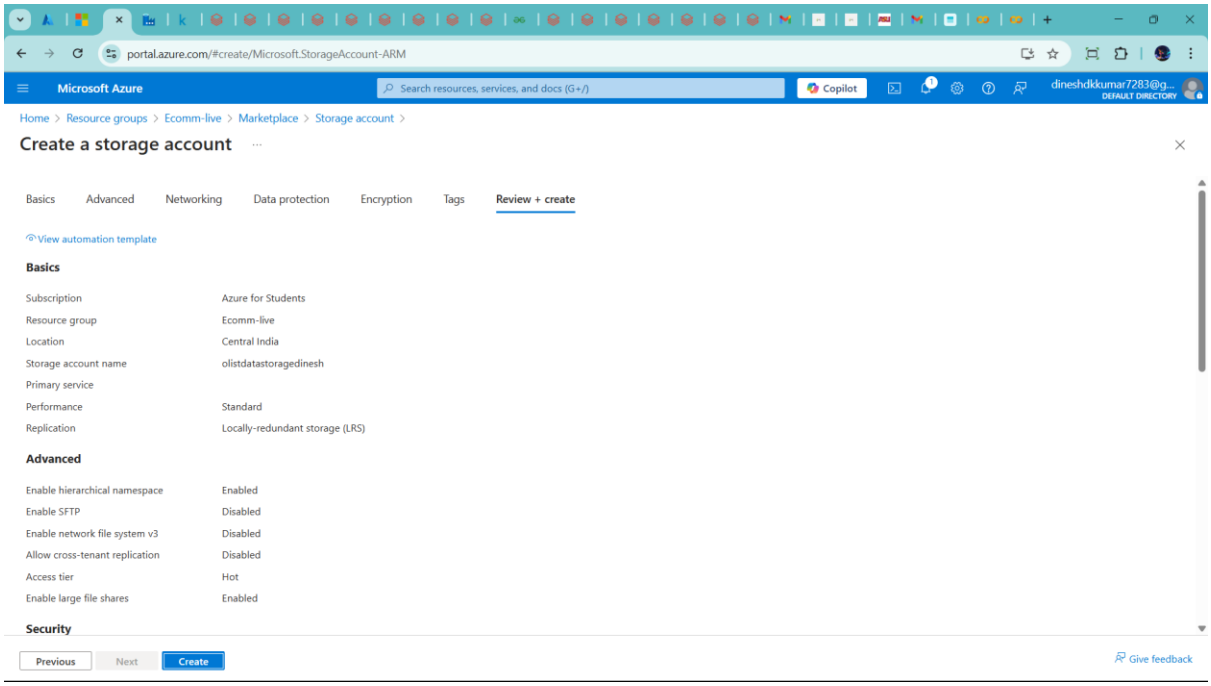


→Redundancy →how duplicate copy store different location

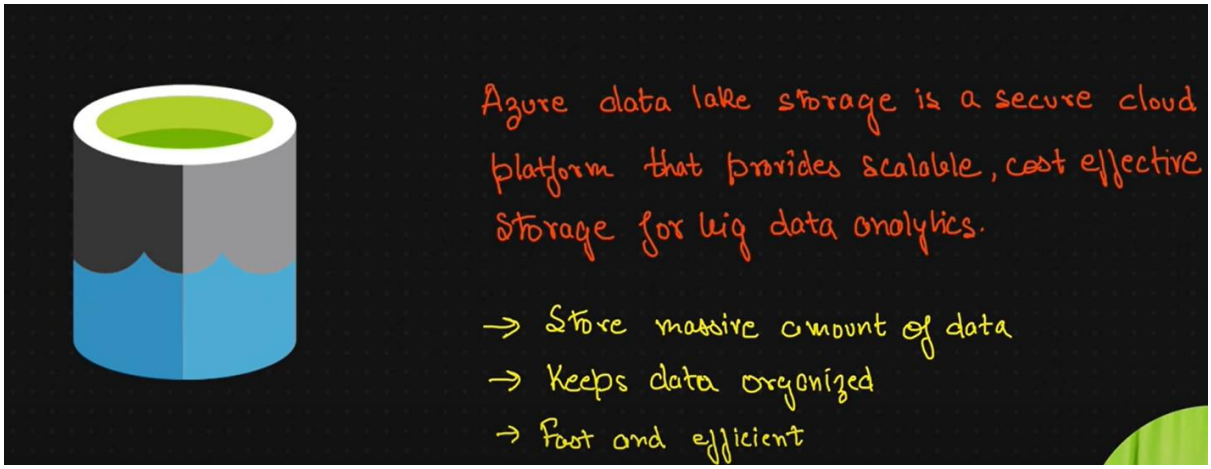
→Enable Hierachical →it possible to create folder

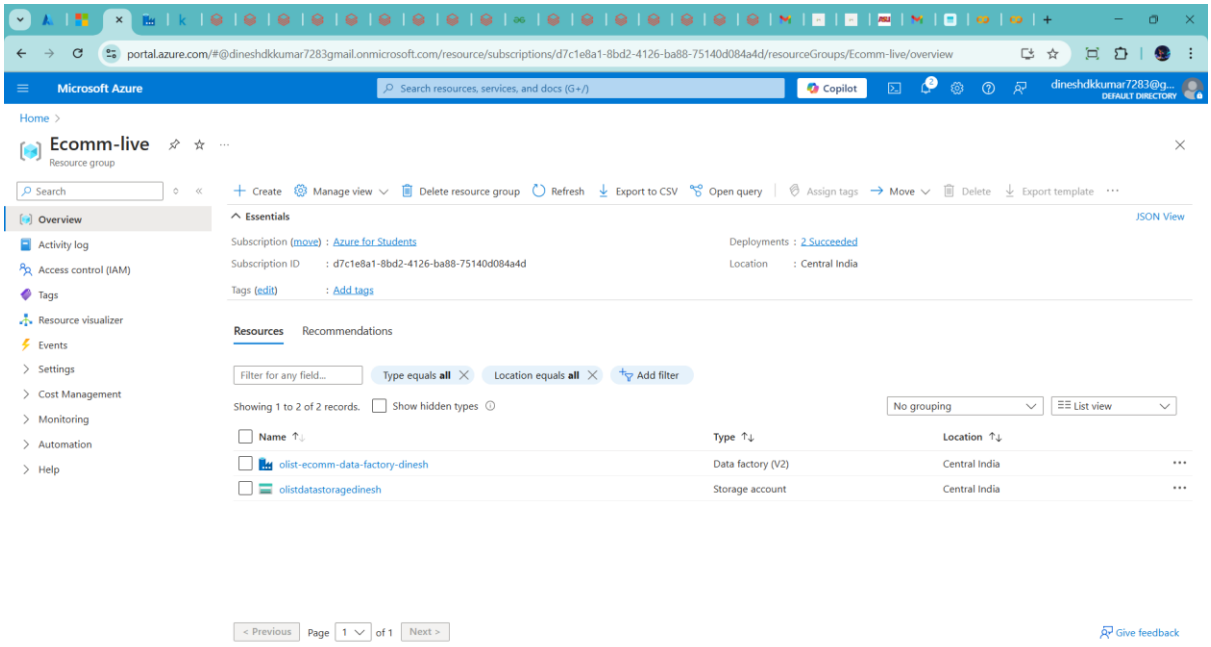


→Click review +create

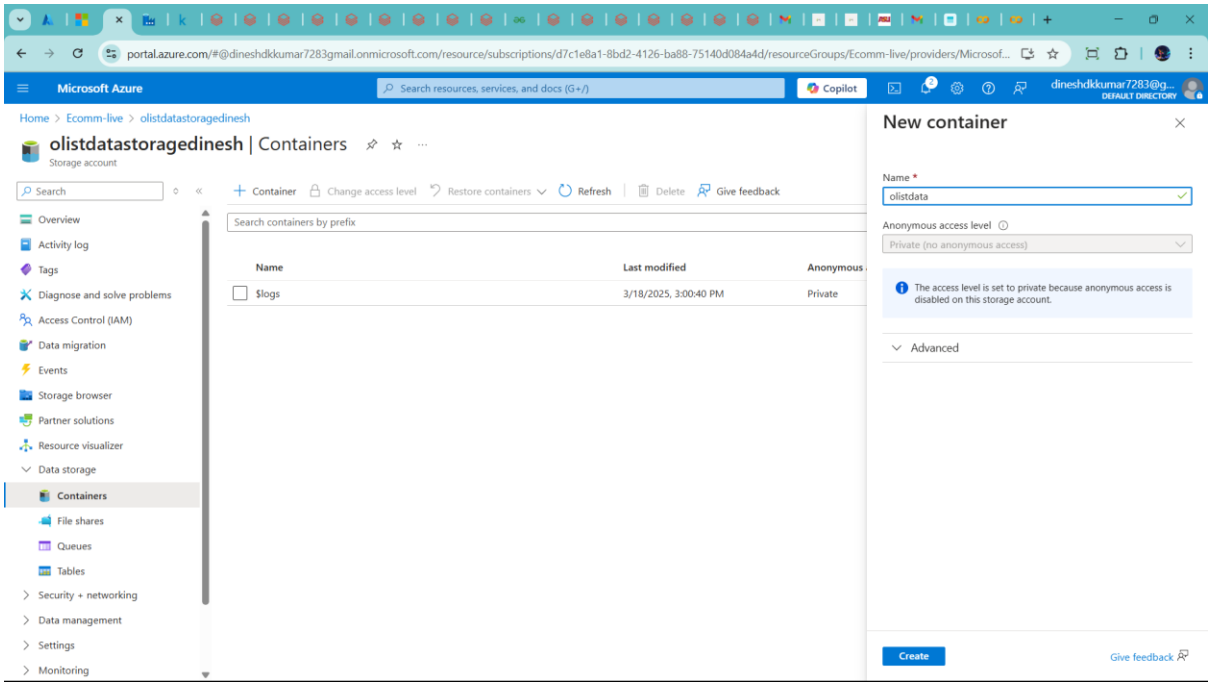


→Create

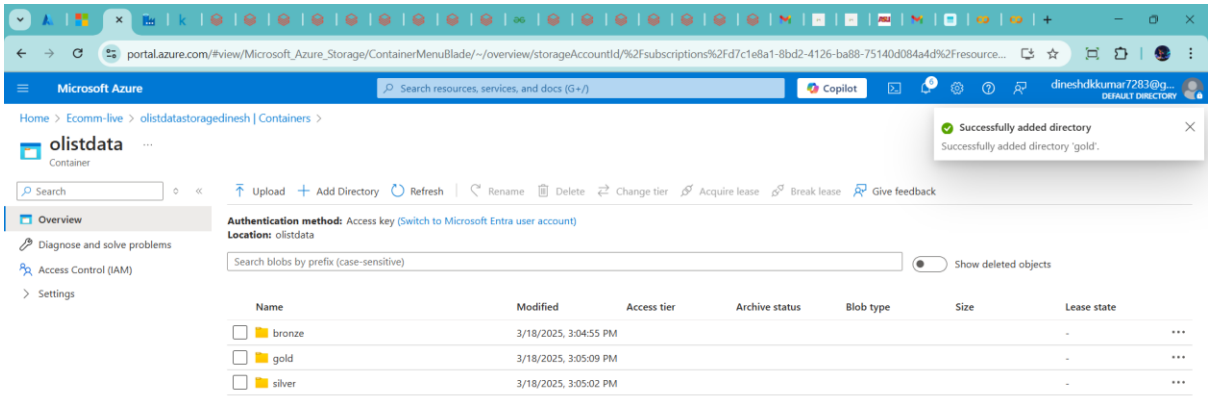


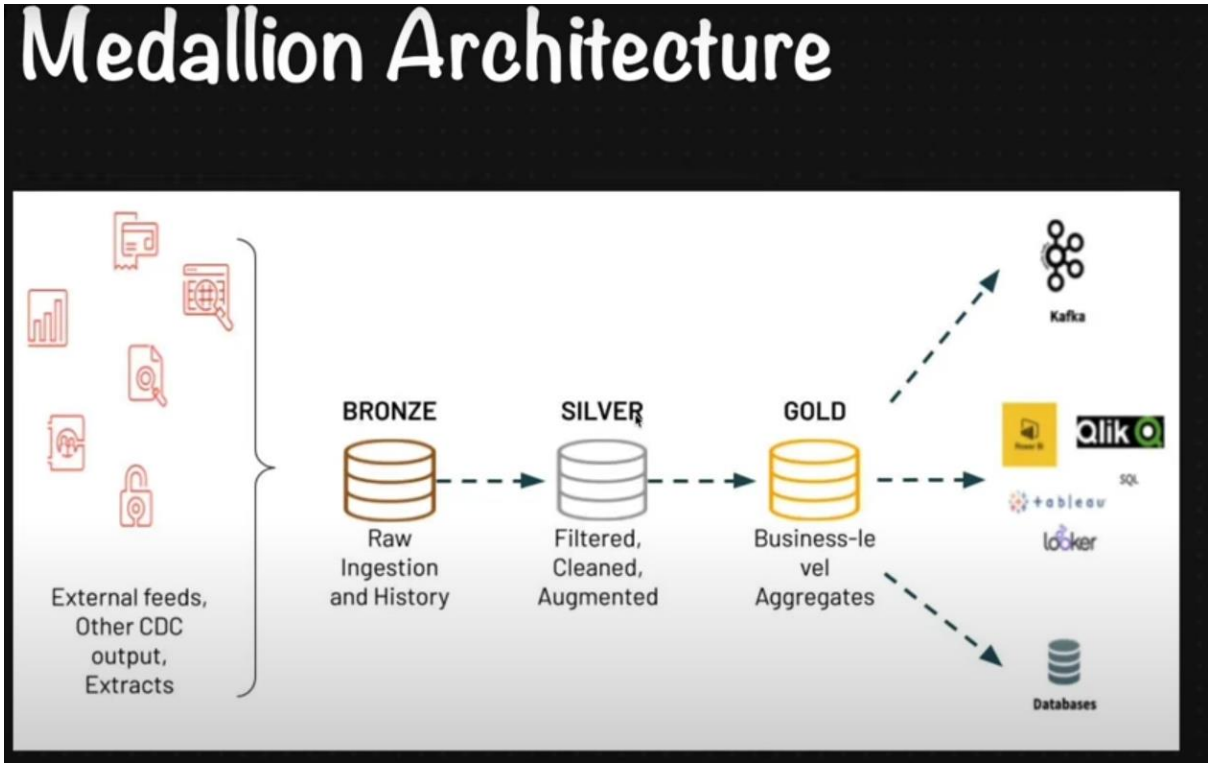


→create container

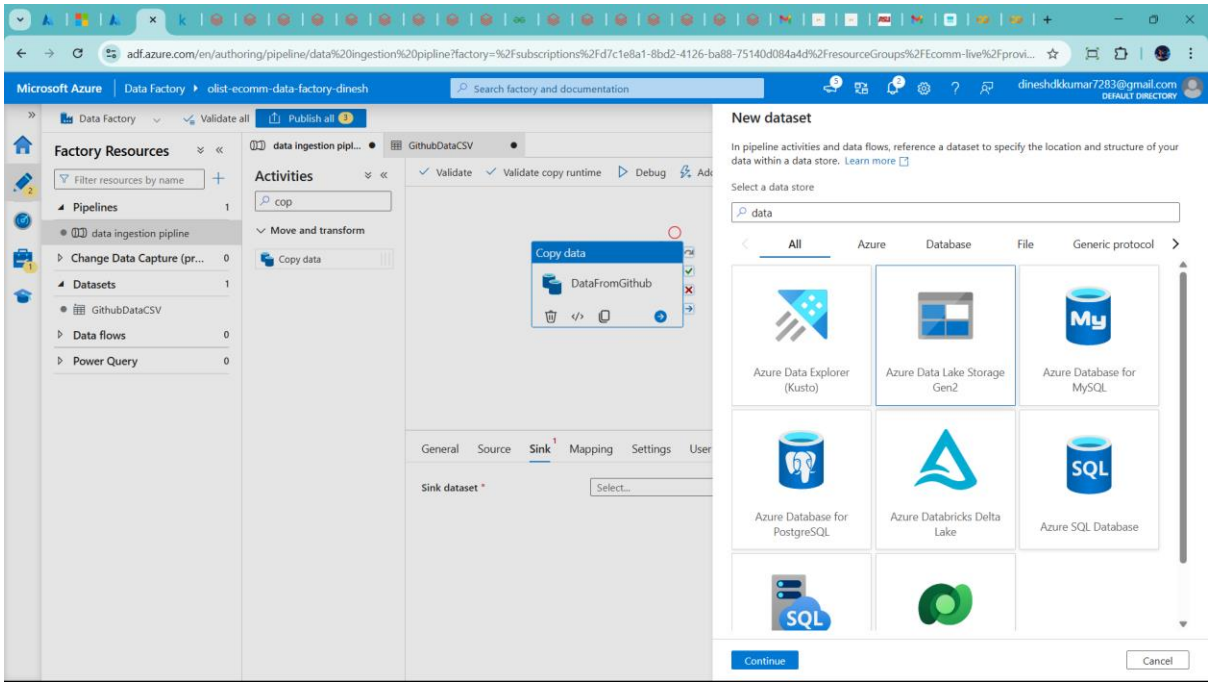


→add directory (bronze,silver,gold)

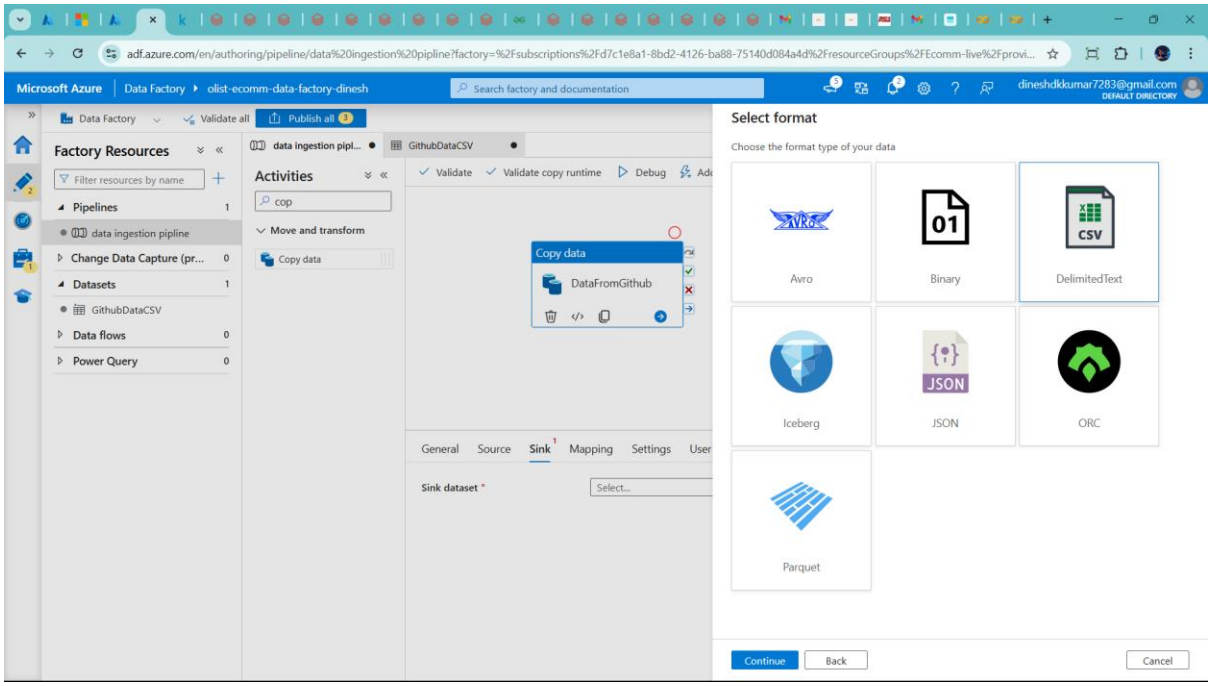




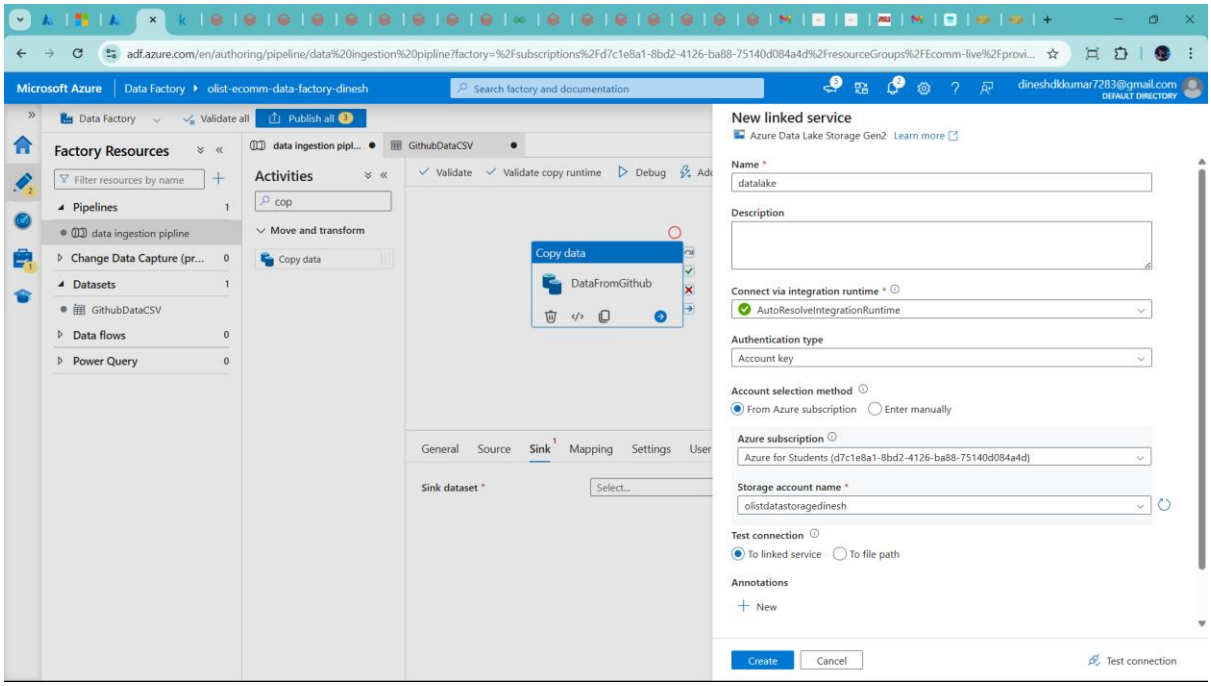
→go to sink → click new →click data lake gen 2 →Continue



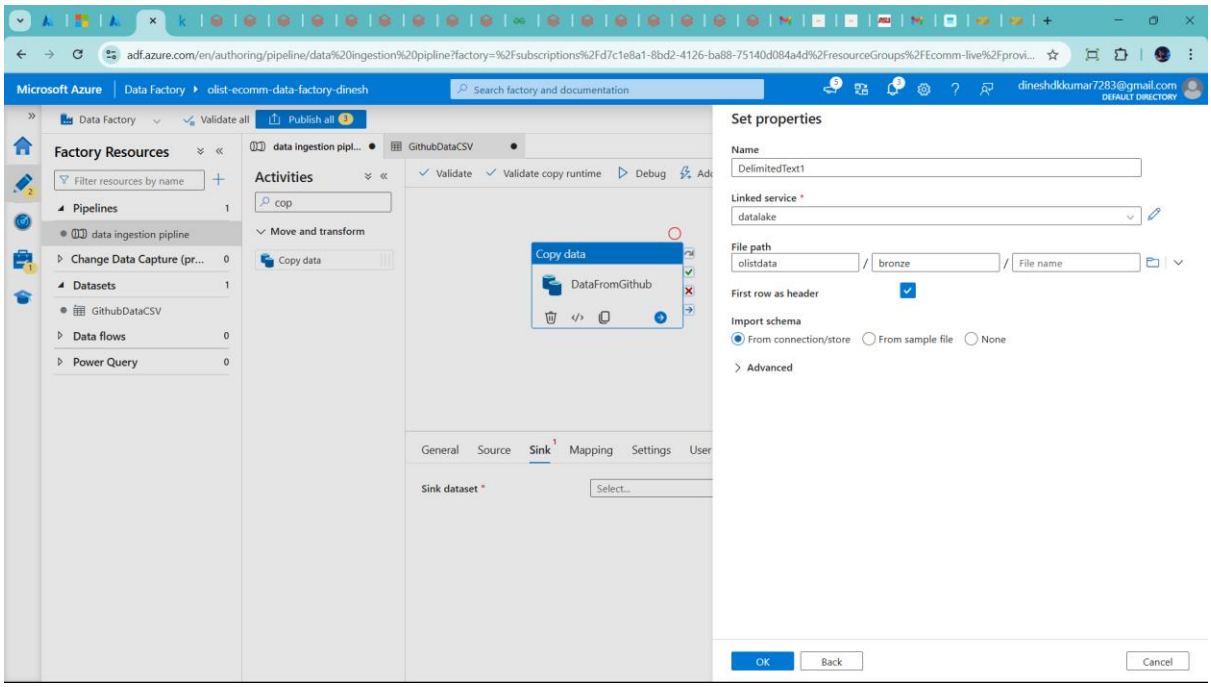
→send to csv



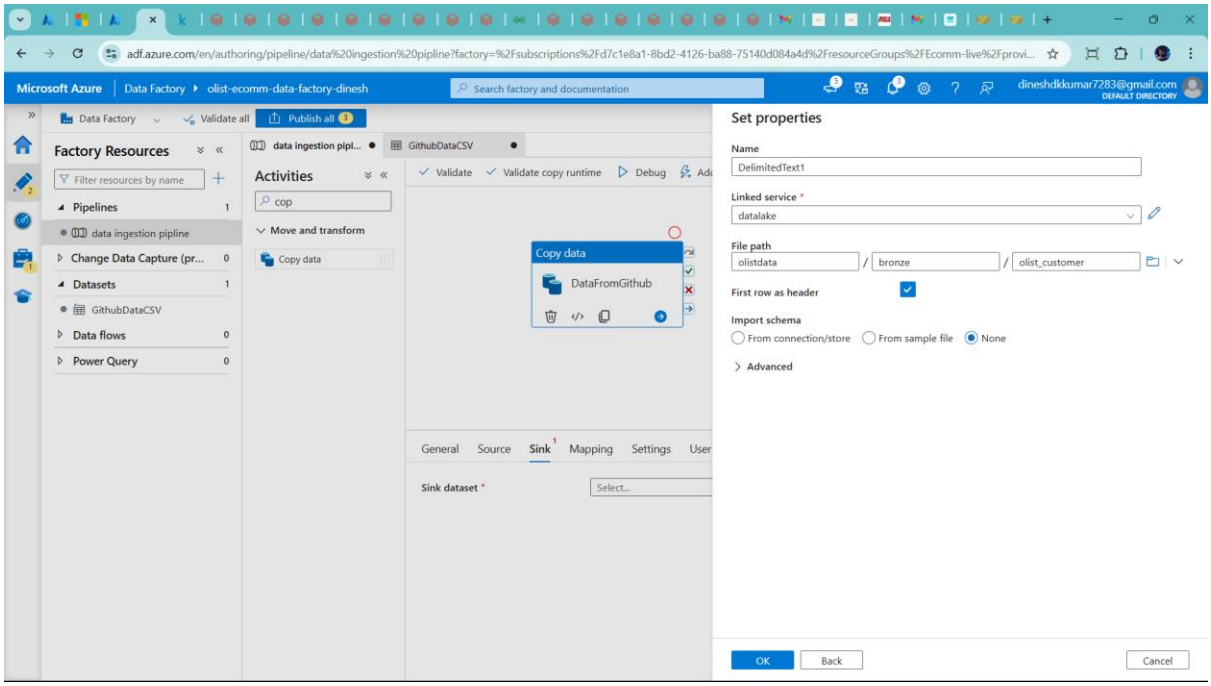
→Click add new link service

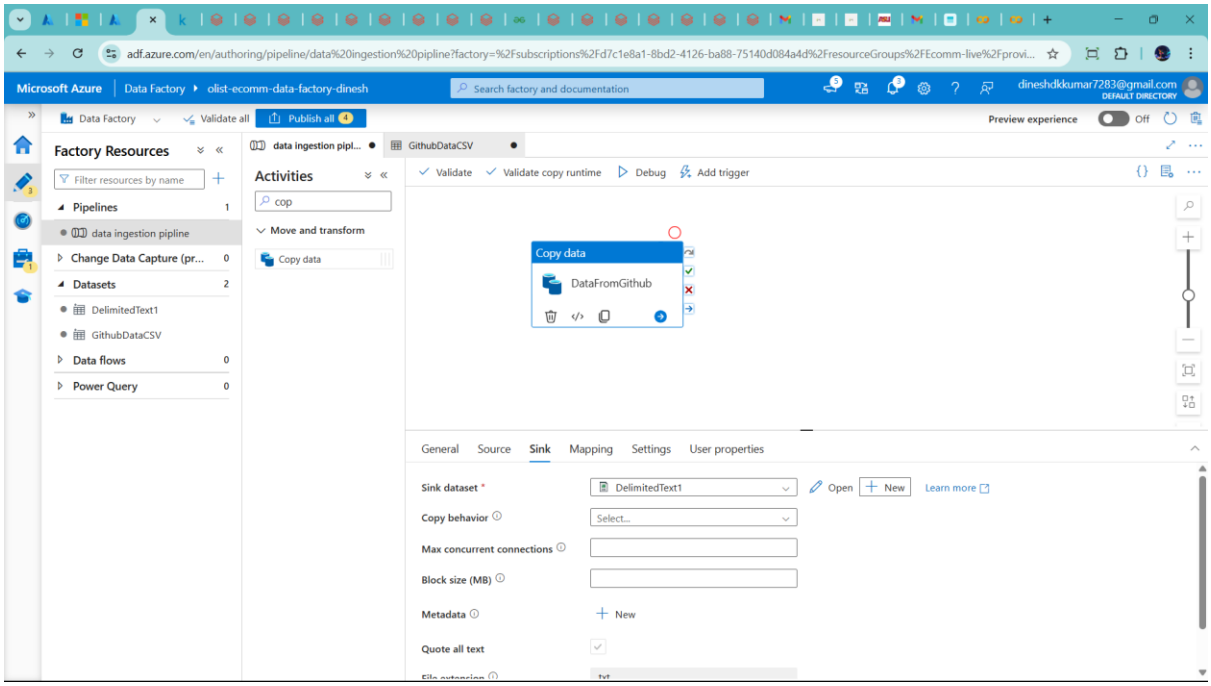


→Click file path icon →choose bronze folder →click ok

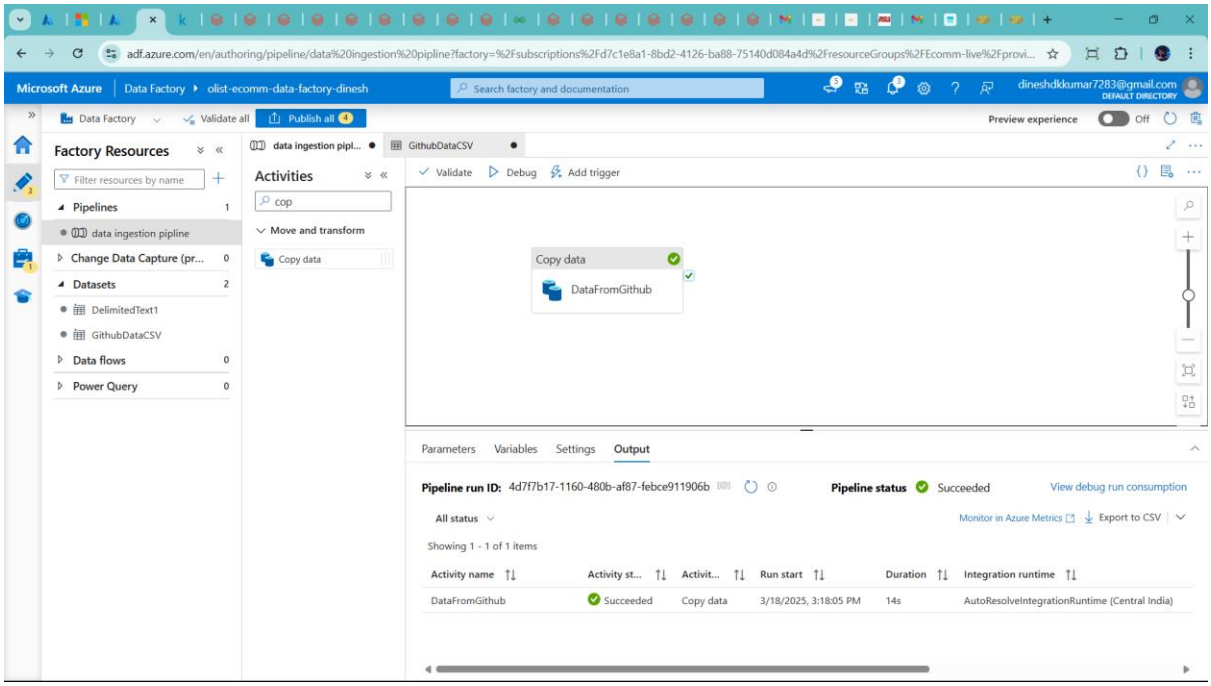


→final add file name





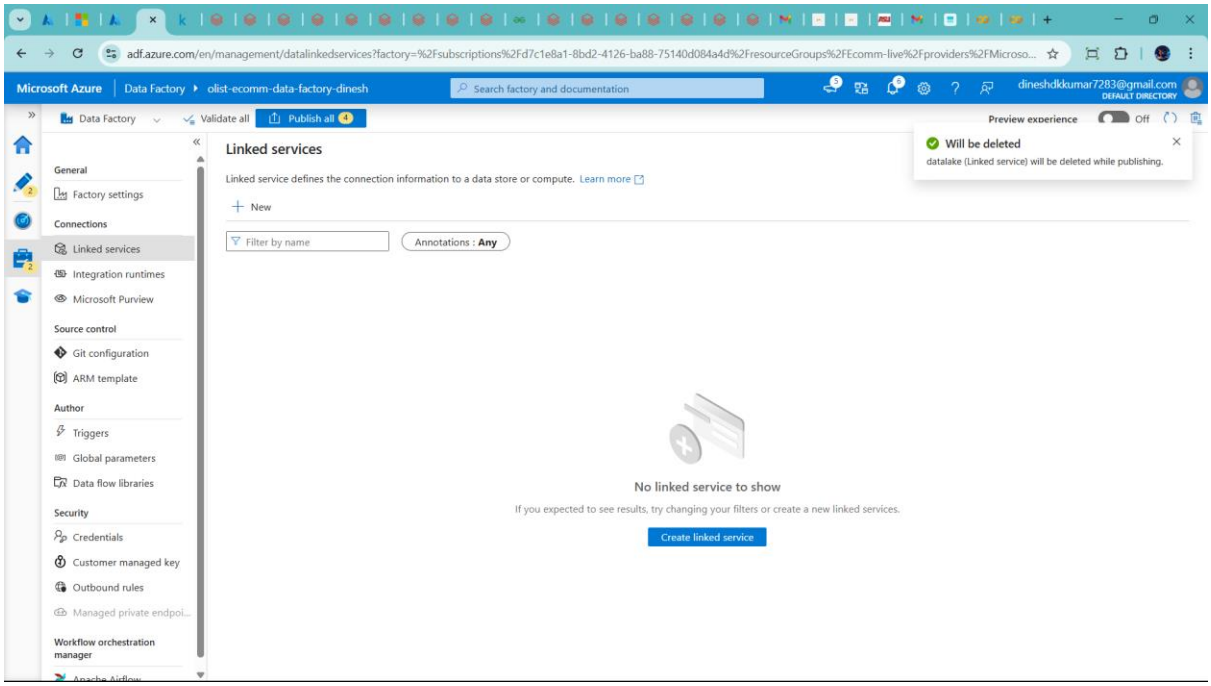
→now validated and debug



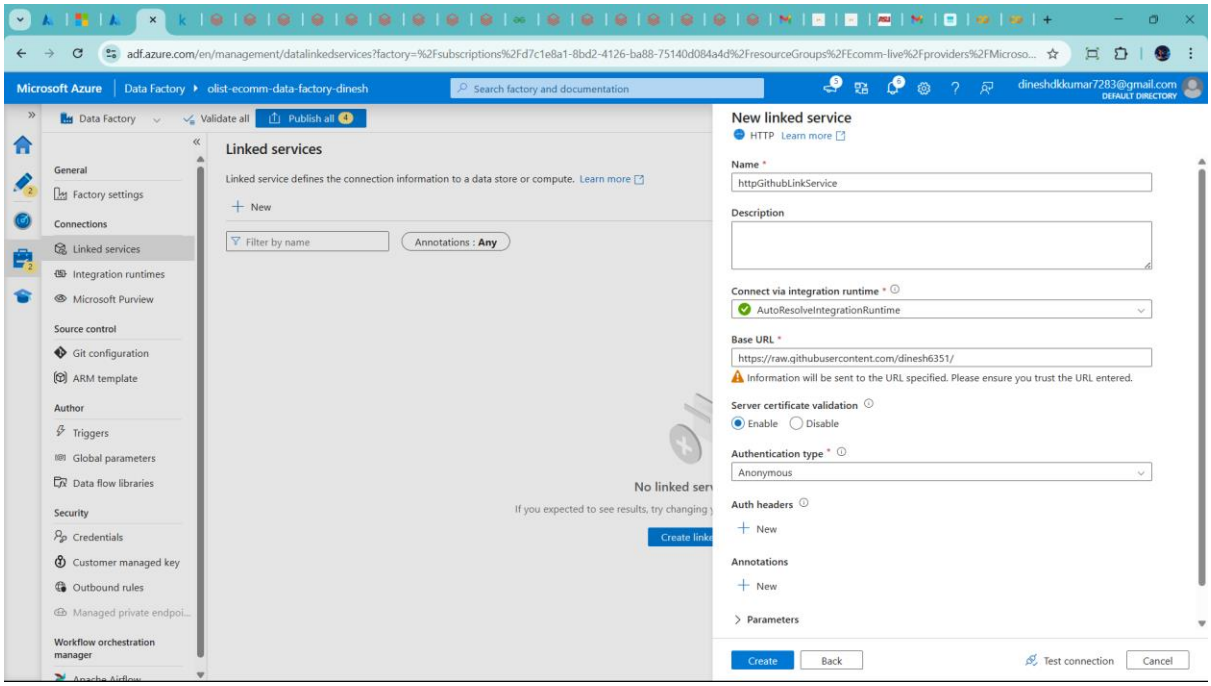
→hot (high cost)

Now to create parameter

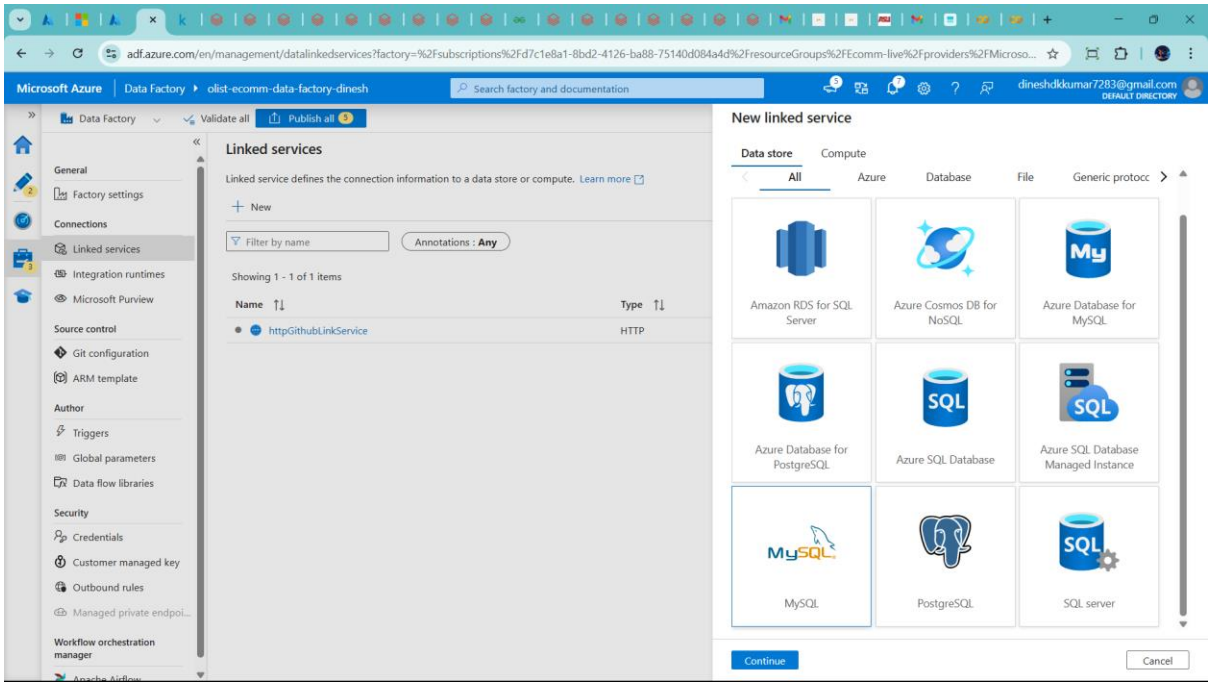
→before create delete all link service

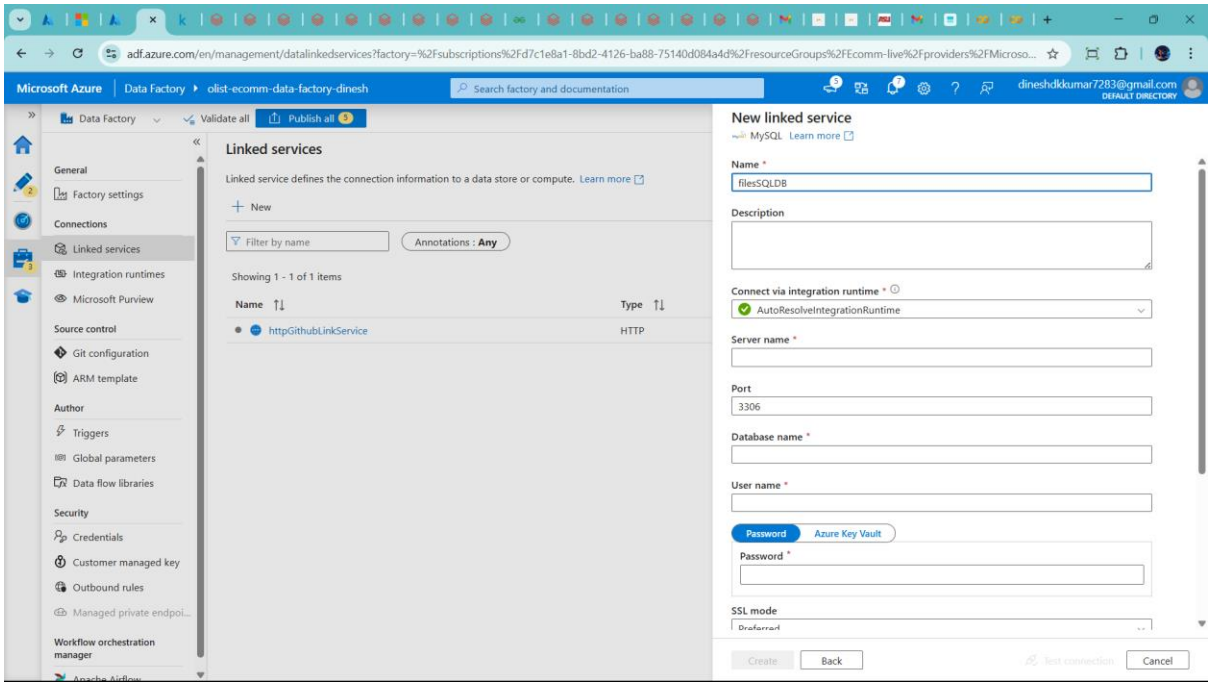


→create new link service →https (github)

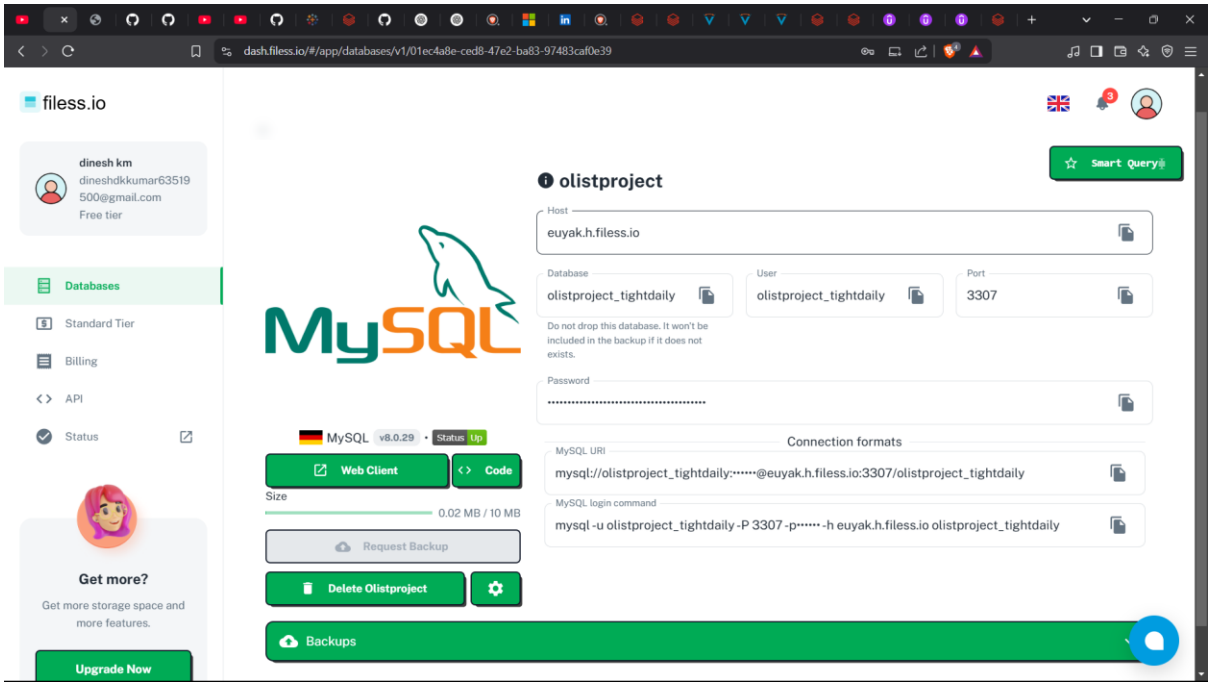


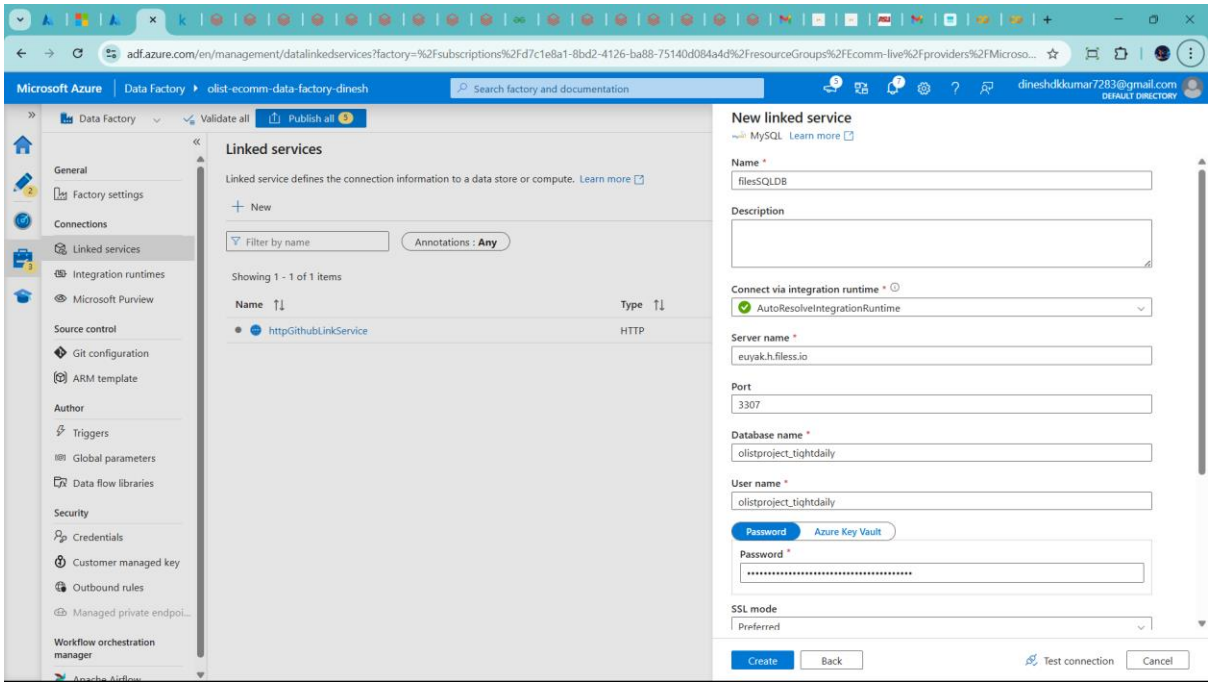
→create another linked service sql



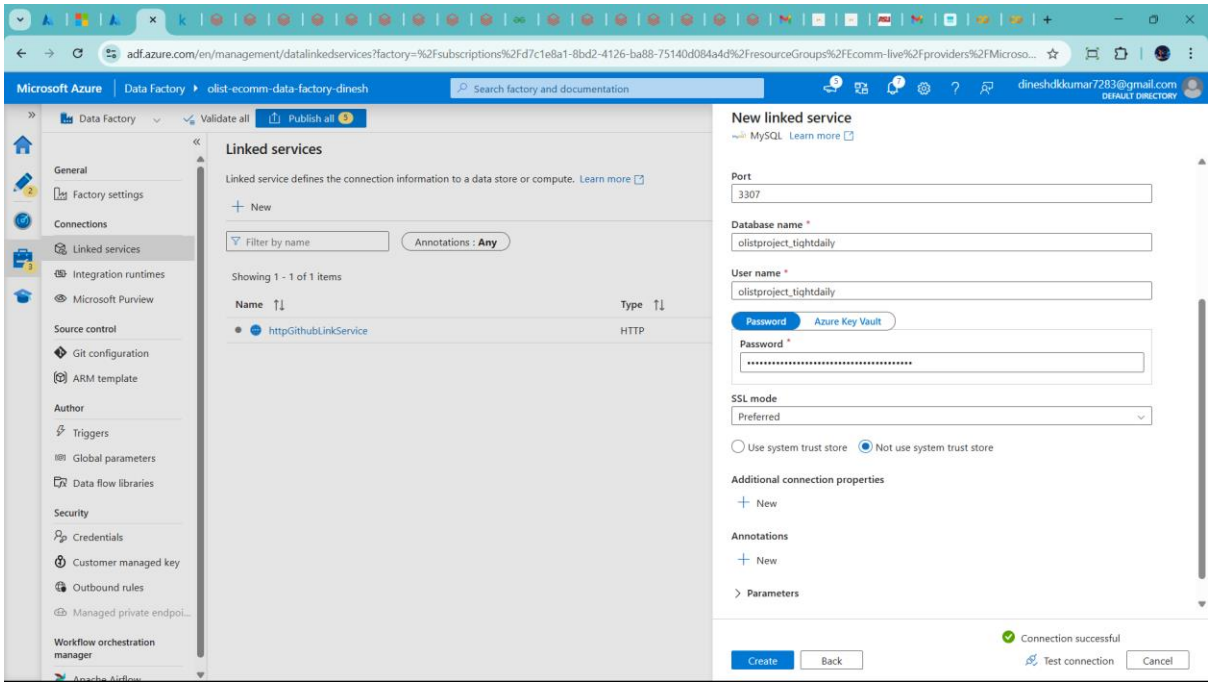


→fill all according in file.io

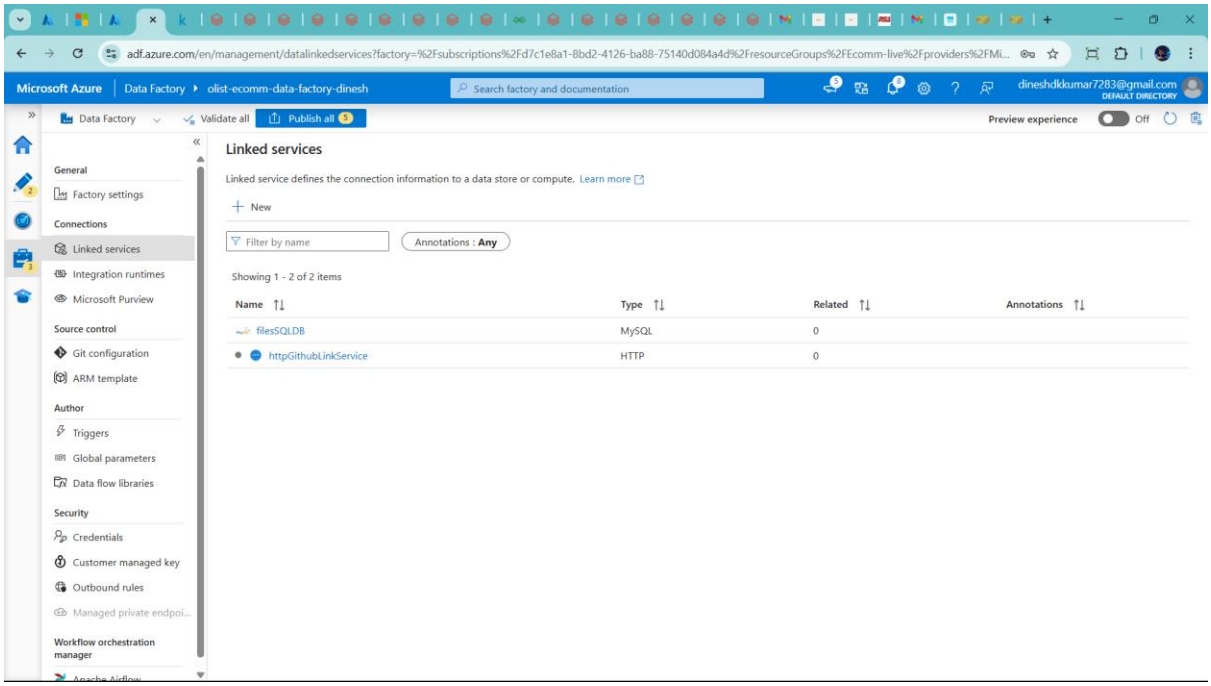




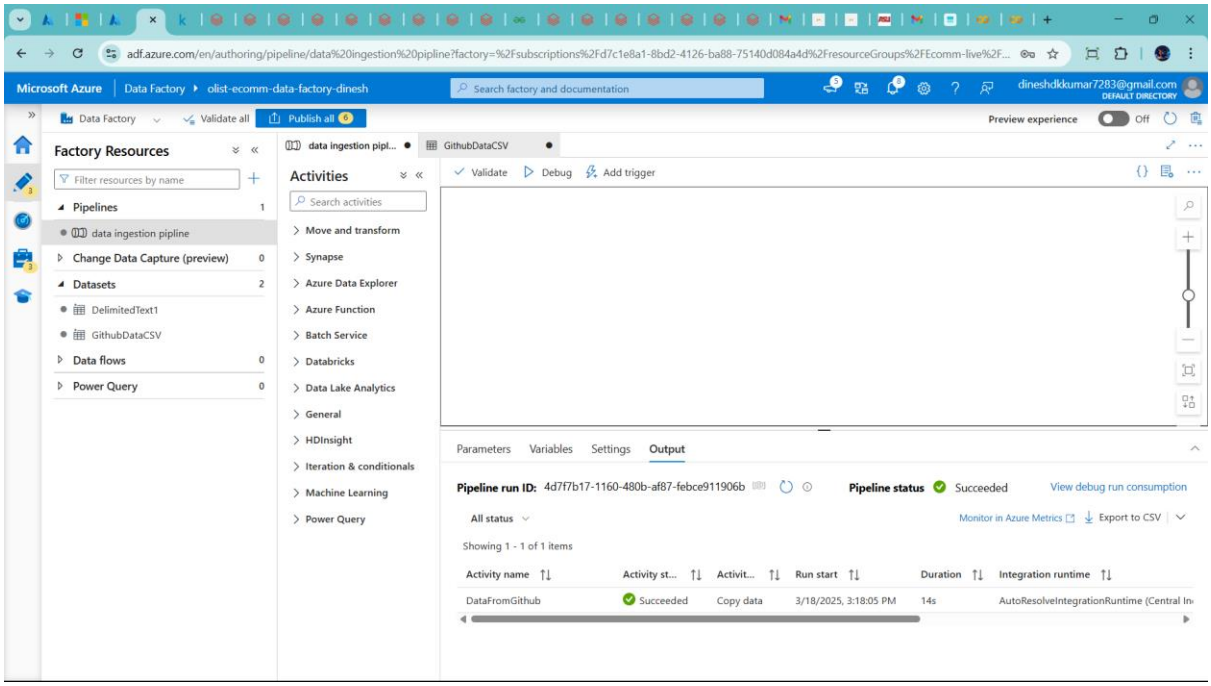
→test connection



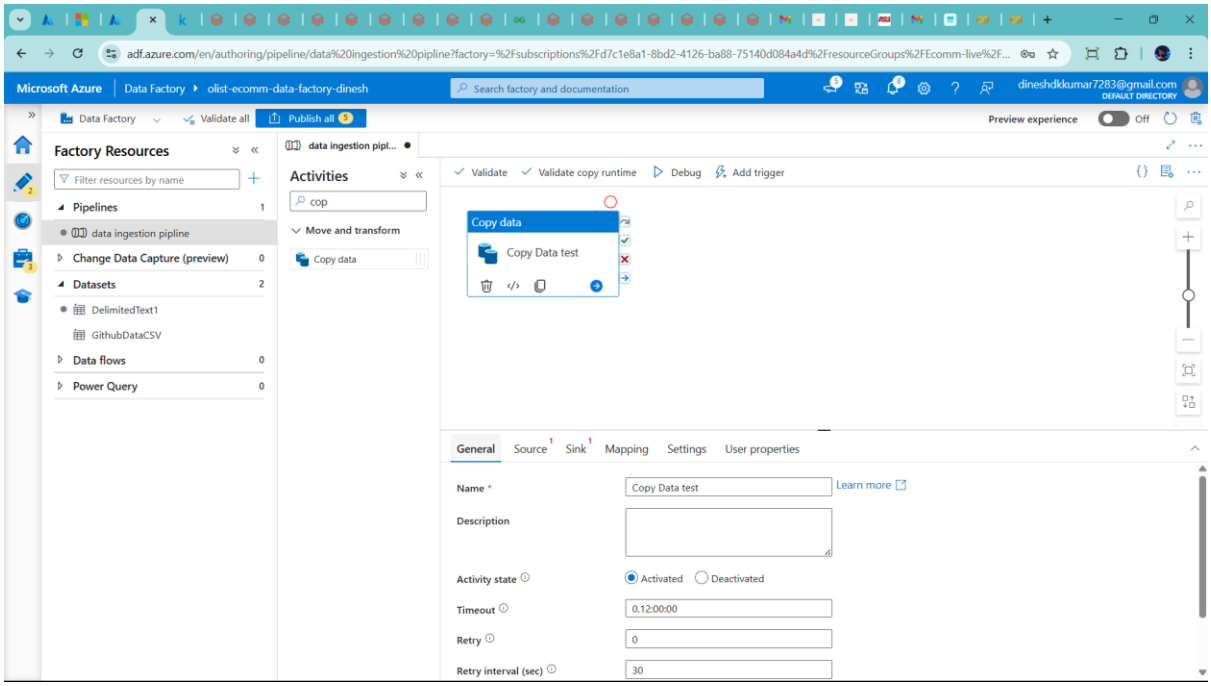
→Create



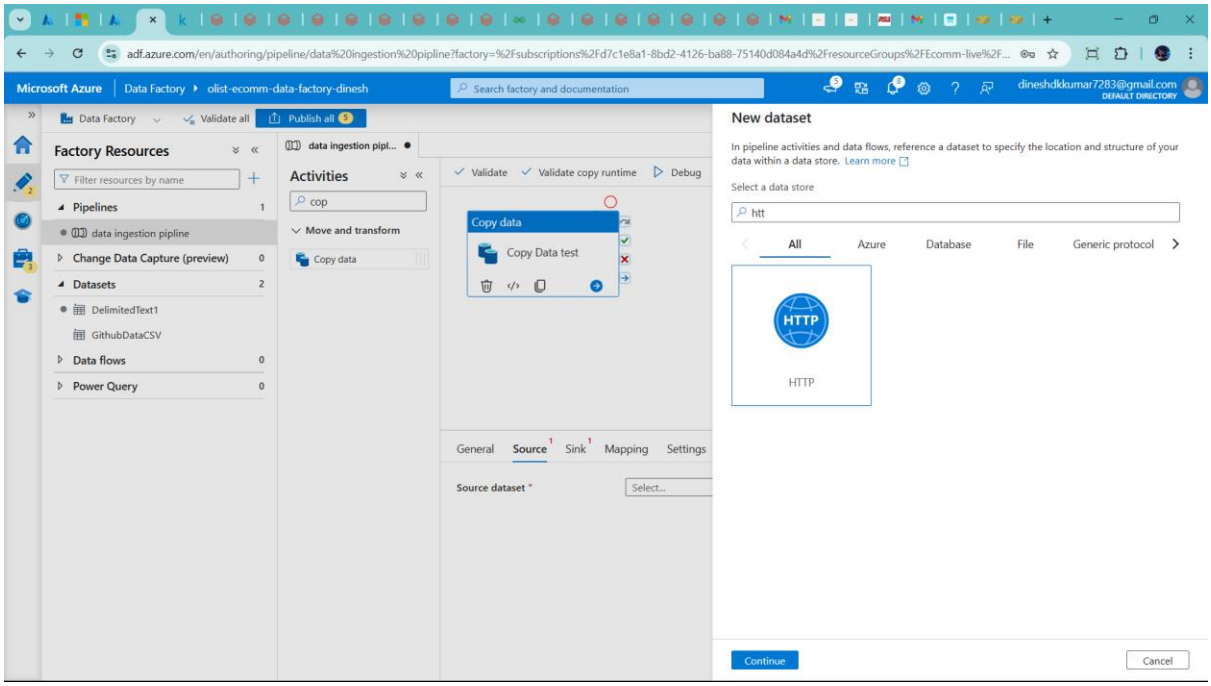
→delete pipeline copy activity

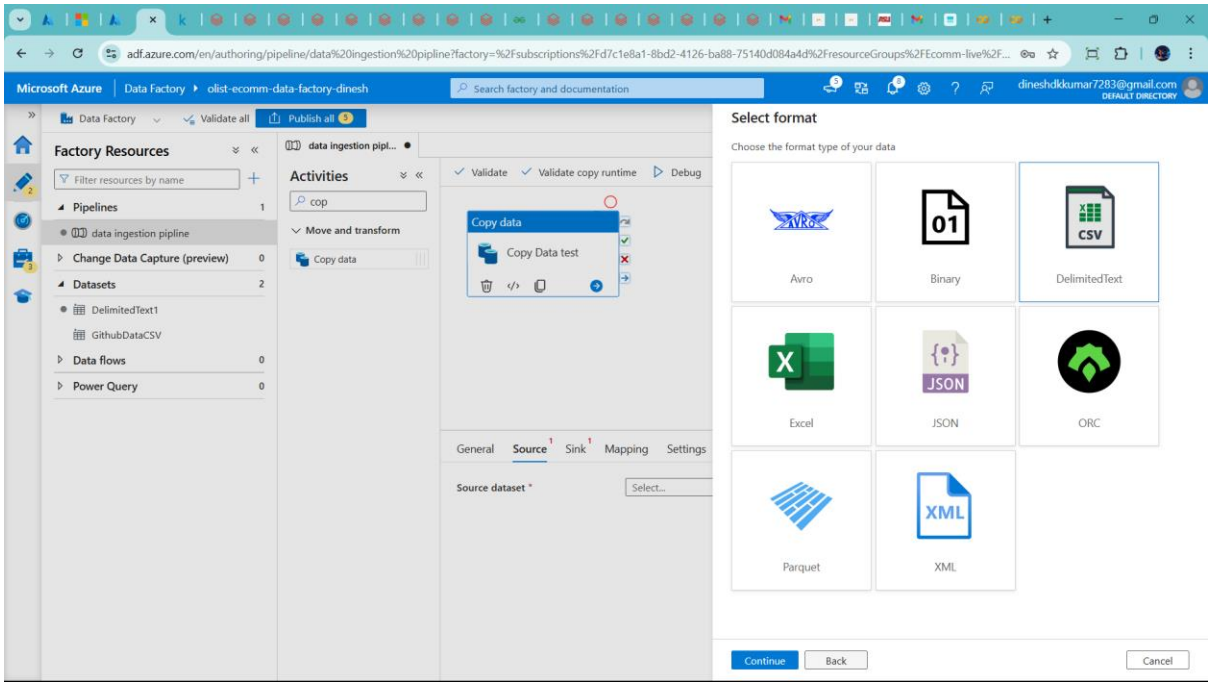


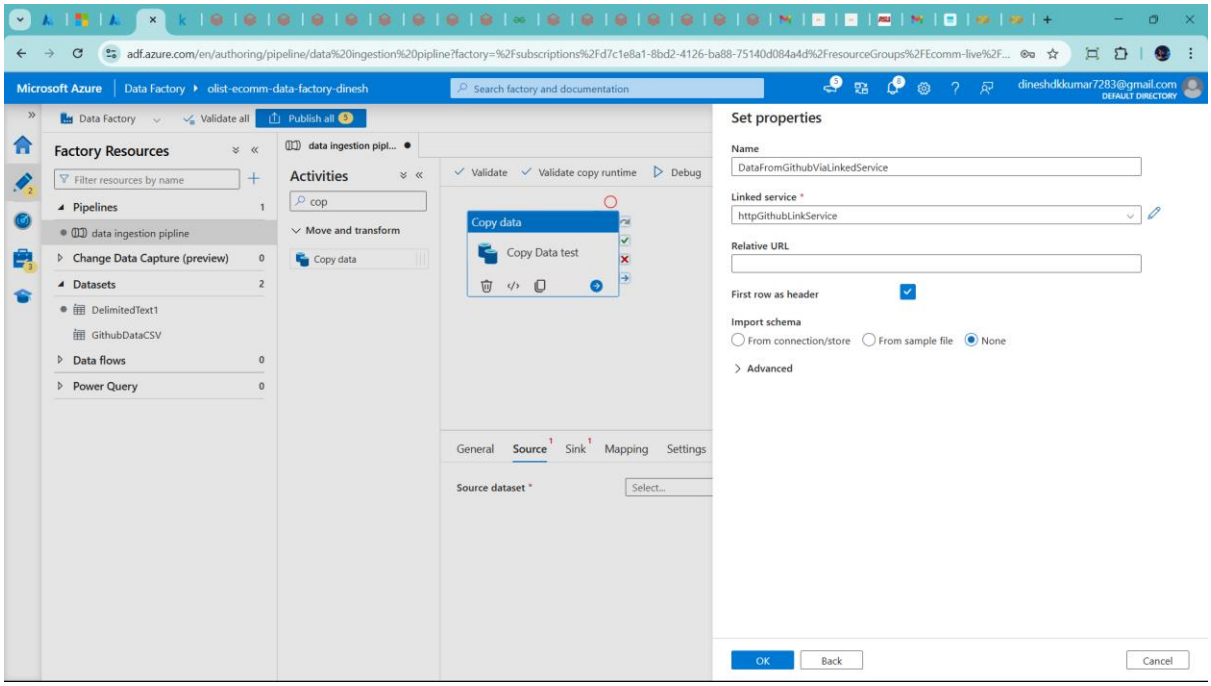
→drap again new one



→click source and add + New


adf.azure.com/en/authoring/pipeline/data%20ingestion%20pipline?factory=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2F...
Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
DelimitedText1
GithubDataCSV
Data flows
Power Query
data ingestion pipl...
Activities
cop
Move and transform
Copy data
Validate
Validate copy runtime
Debug
Copy data
Copy Data test
General
Source
Sink
Mapping
Settings
Source dataset *
Select...
New dataset
In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. Learn more
Select a data store
htt
All
Azure
Database
File
Generic protocol
HTTP
HTTP
Continue
Cancel

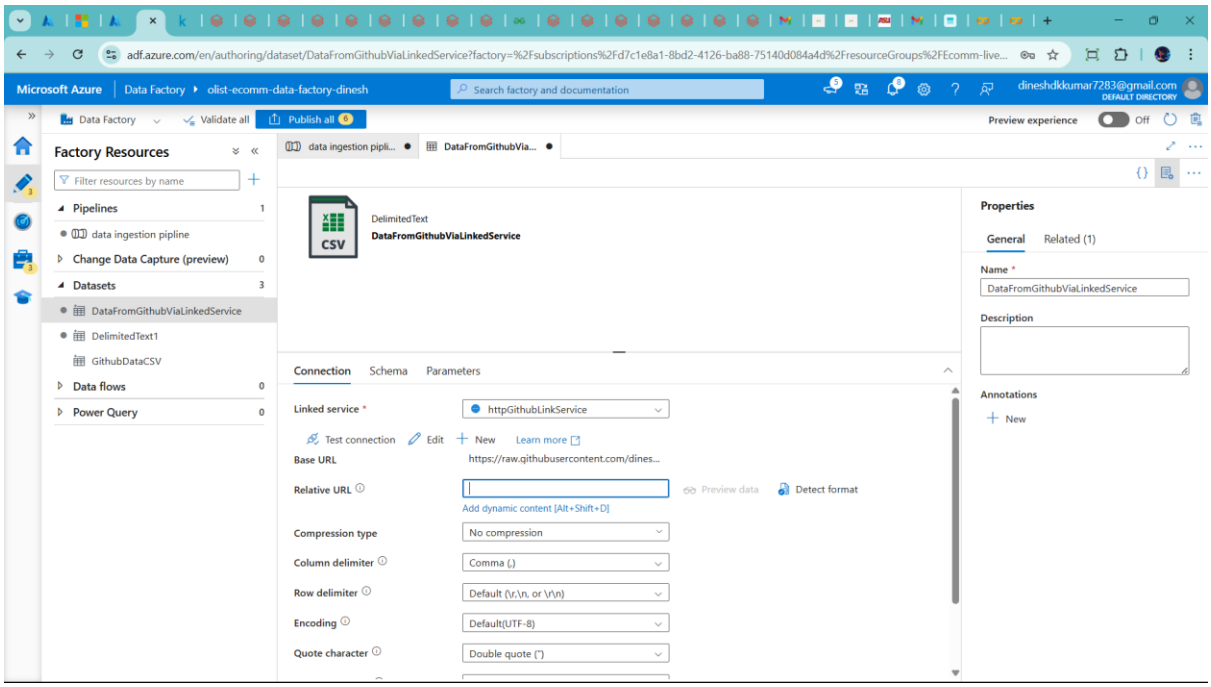

adf.azure.com/en/authoring/pipeline/data%20ingestion%20pipline?factory=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2F...
Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
DelimitedText1
GithubDataCSV
Data flows
Power Query
data ingestion pipl...
Activities
cop
Move and transform
Copy data
Validate
Validate copy runtime
Debug
Copy data
Copy Data test
General
Source
Sink
Mapping
Settings
Source dataset *
Select...
Select format
Choose the format type of your data
Avro
Binary
DelimitedText
Excel
JSON
ORC
Parquet
XML
Continue
Back
Cancel



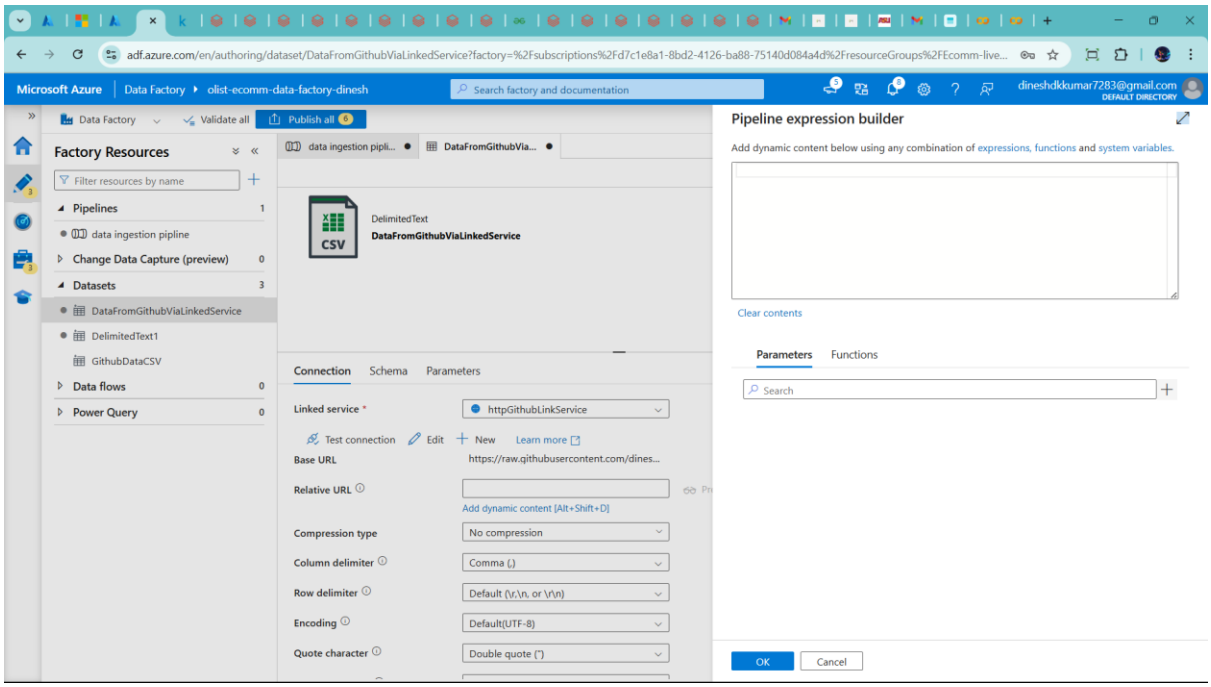
→clicl ok

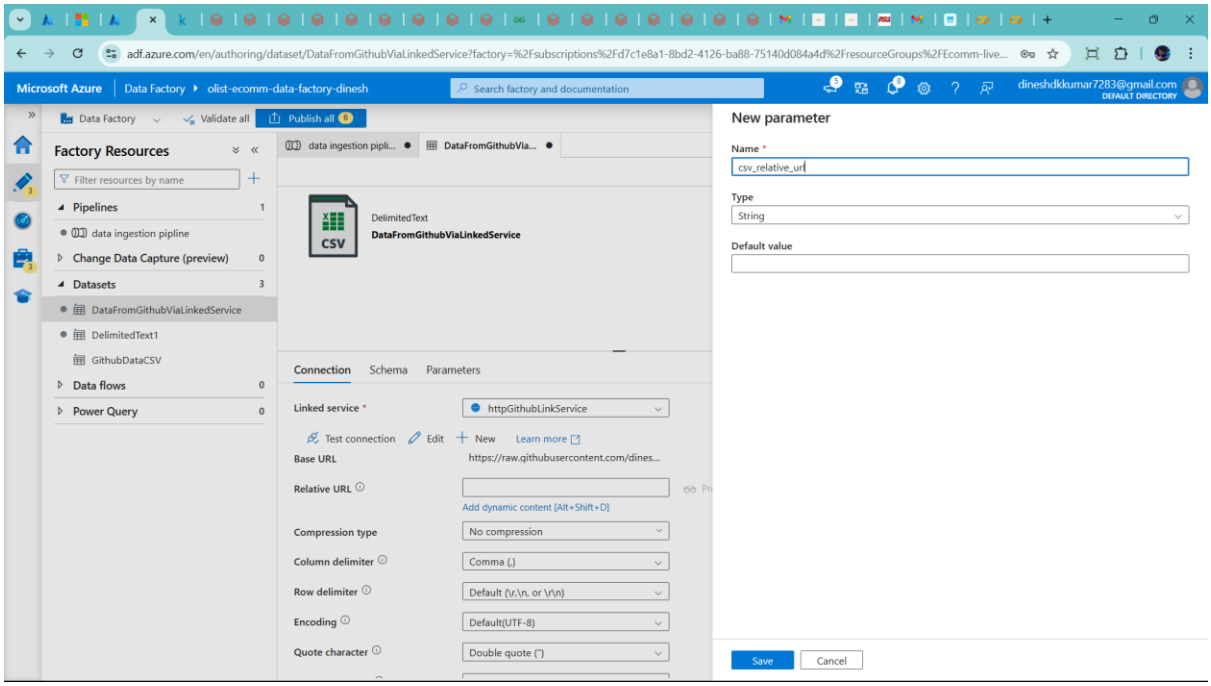
→now add relative URL -in dynamic

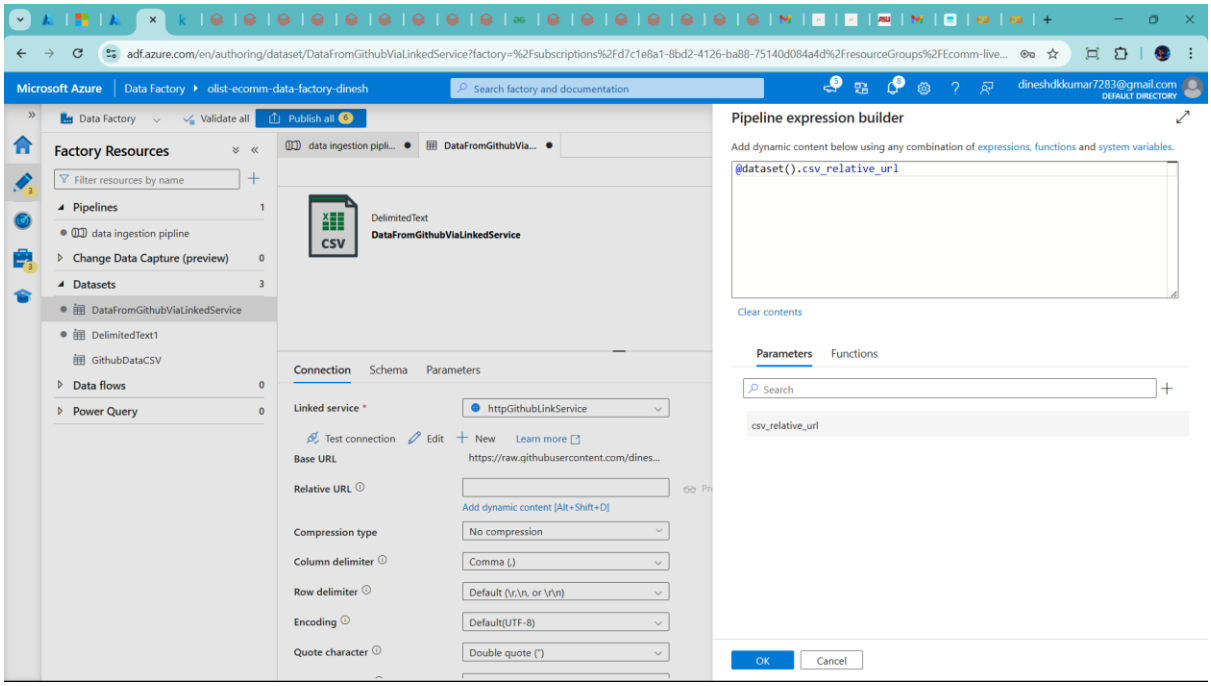
→add dynamin url



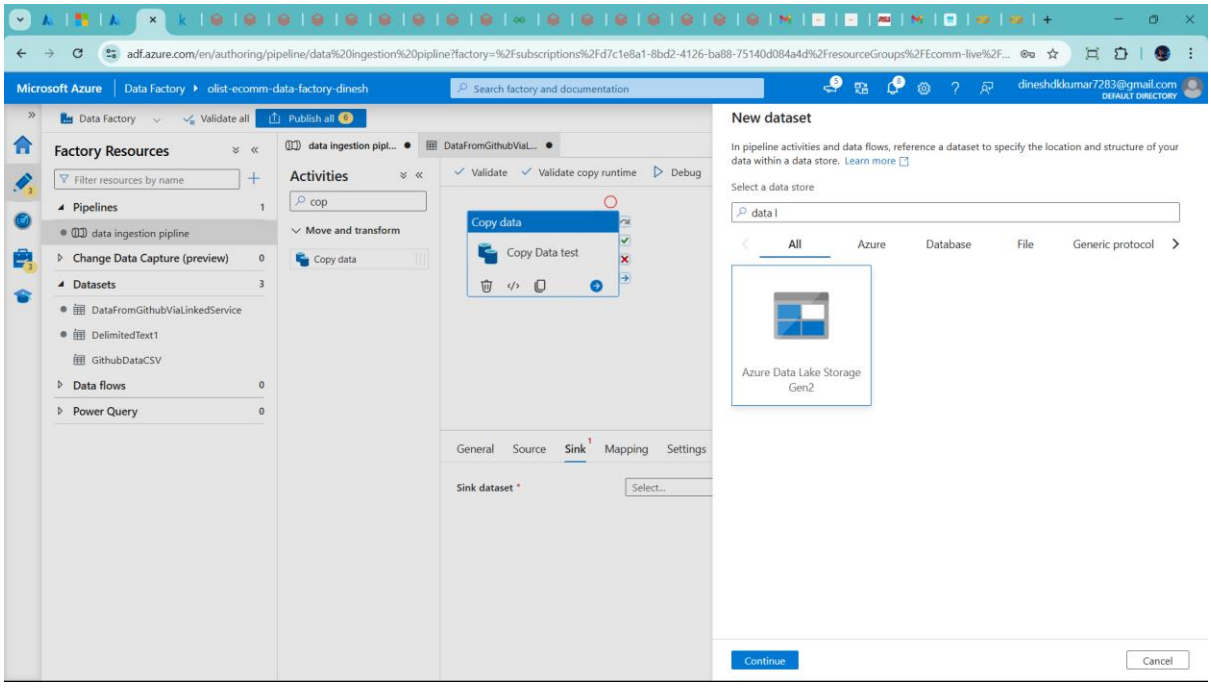
→click parameter

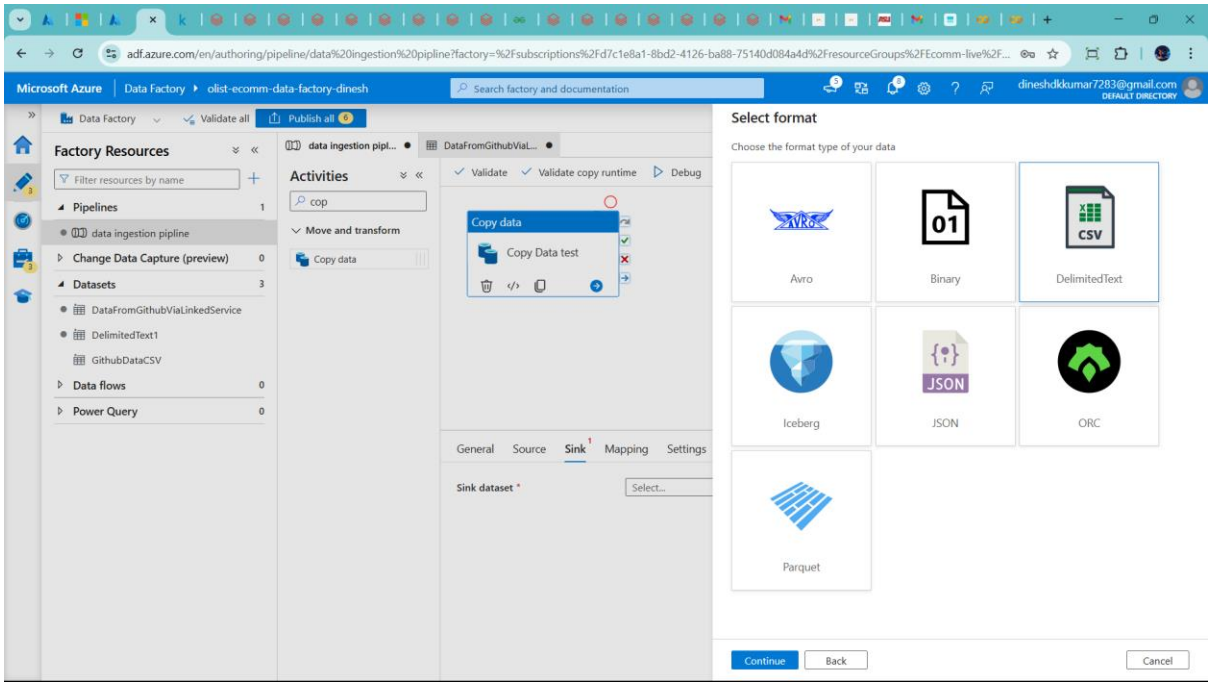


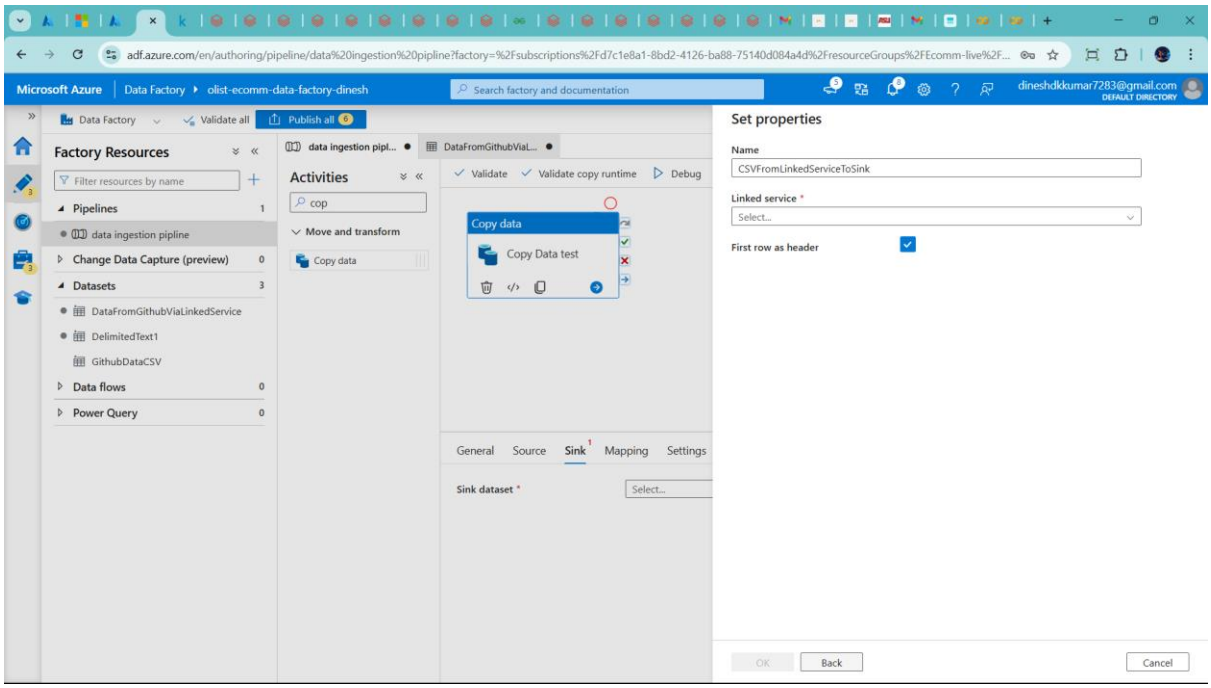




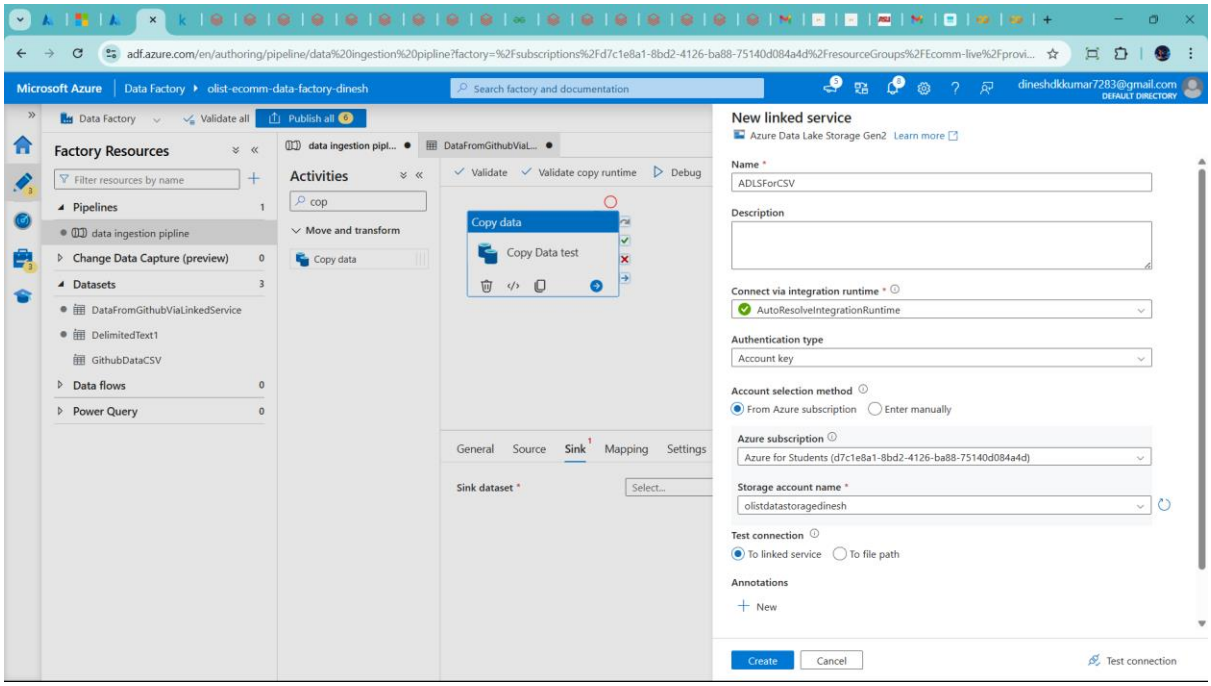
→Same as Sink →click New

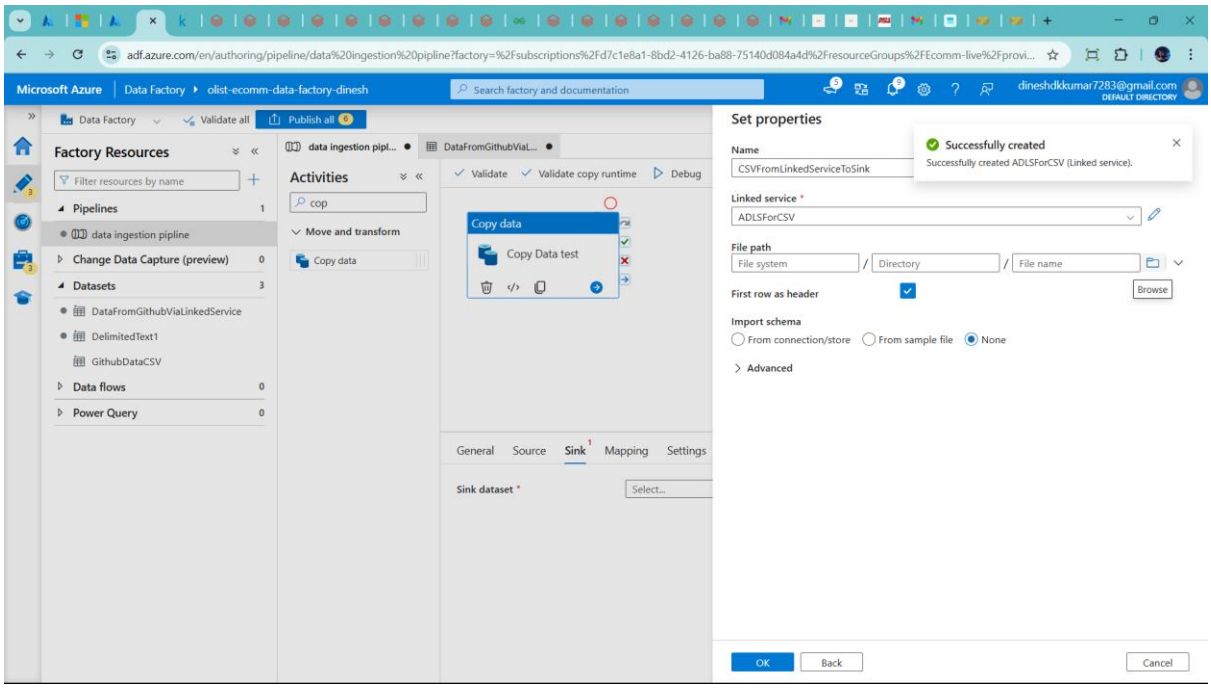




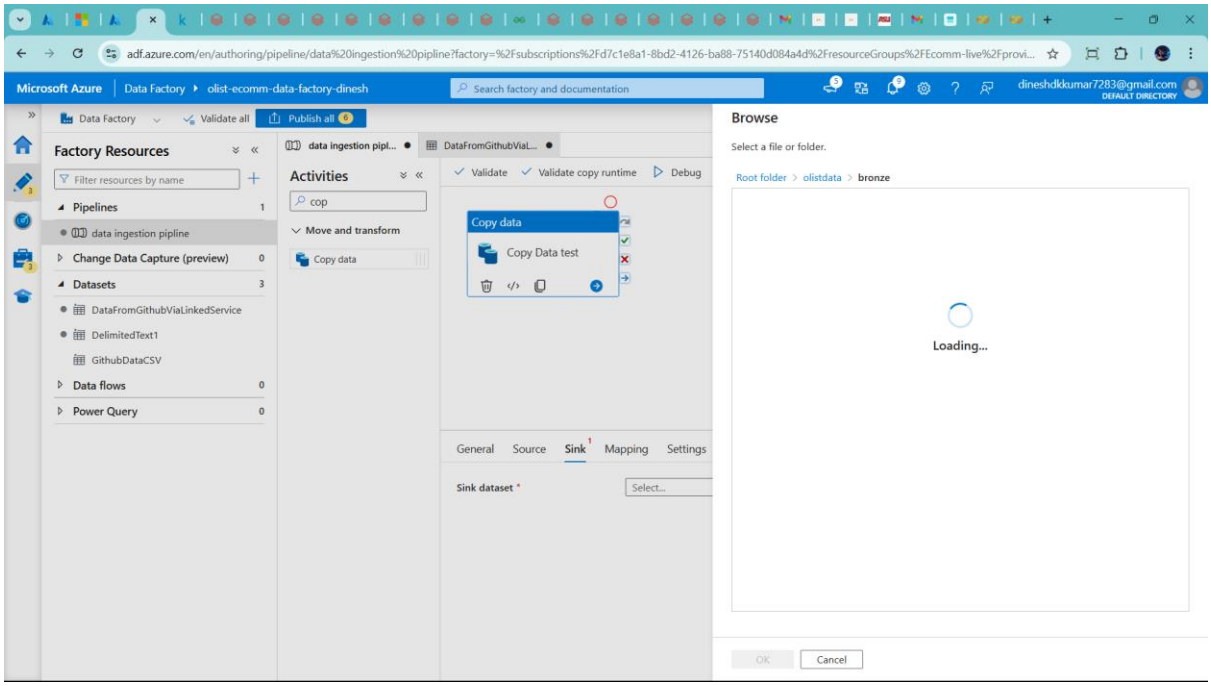


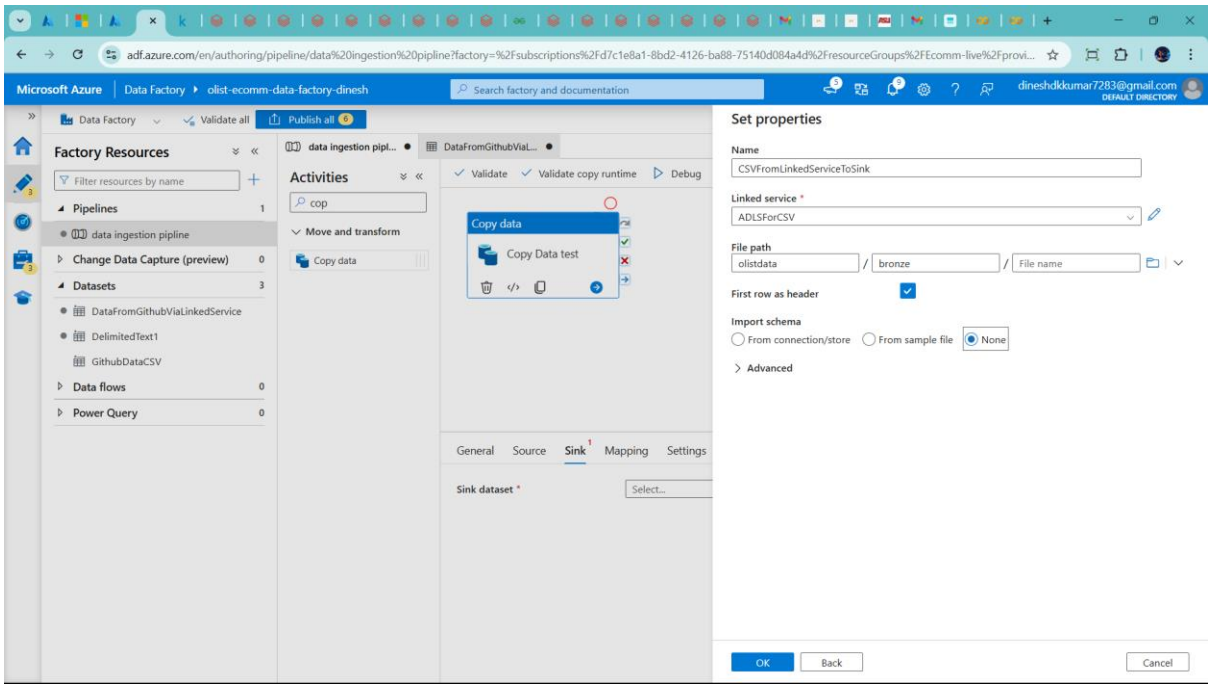
→add linked service new

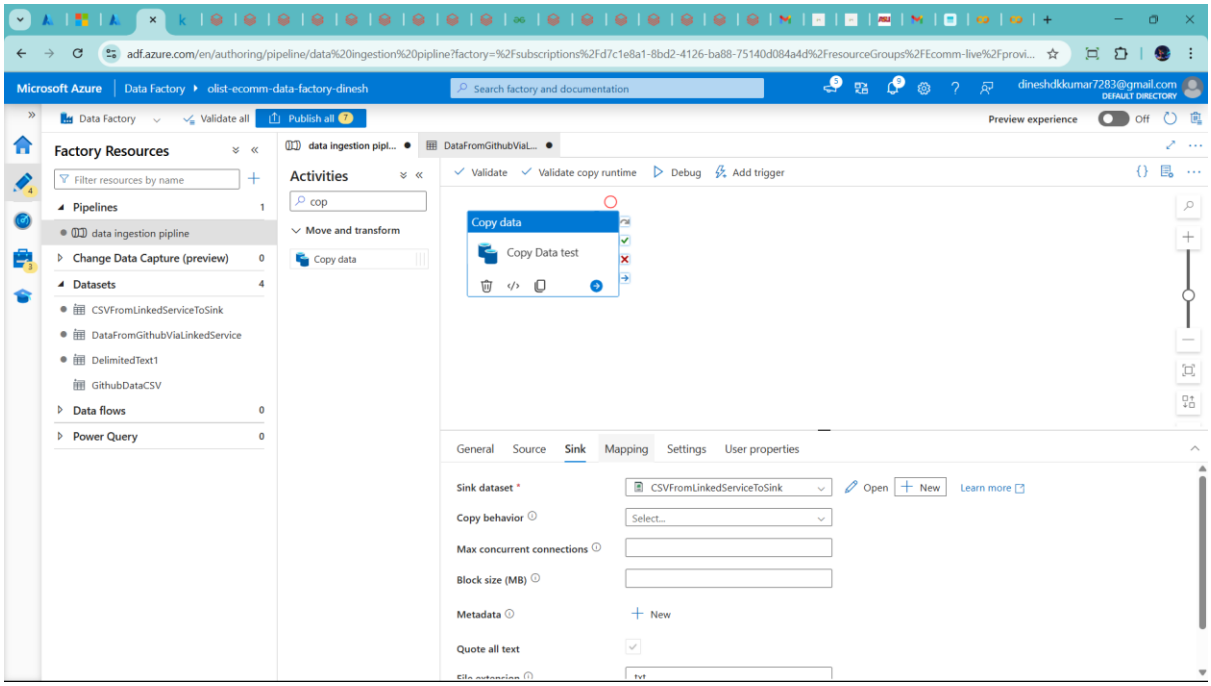




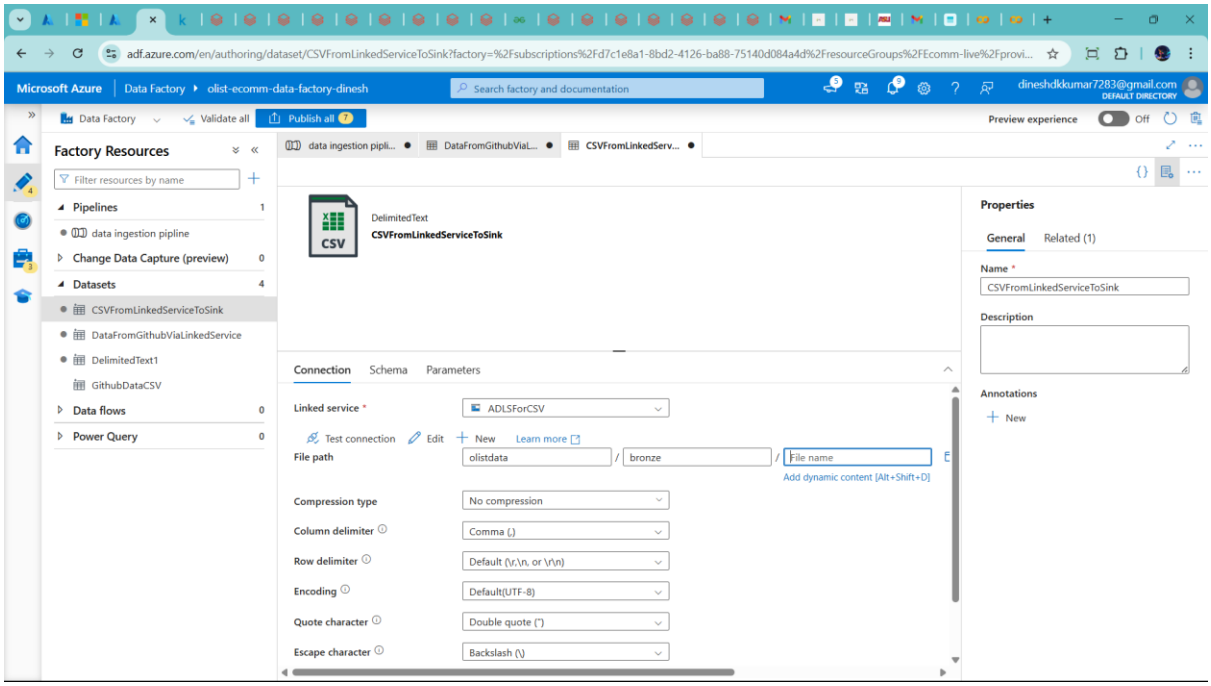
→choose bronze folder


adf.azure.com/en/authoring/pipeline/data%20ingestion%20pipline?factory=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2Fprovi...
Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
Data flows
Power Query
data ingestion pipl...
DataFromGithubViaL...
Activities
cop
Move and transform
Copy data
Validate
Validate copy runtime
Debug
Copy data
Copy Data test
General
Source
Sink
Mapping
Settings
Sink dataset *
Select...
Browse
Select a file or folder.
Root folder > olistdata > bronze
Loading...
OK
Cancel

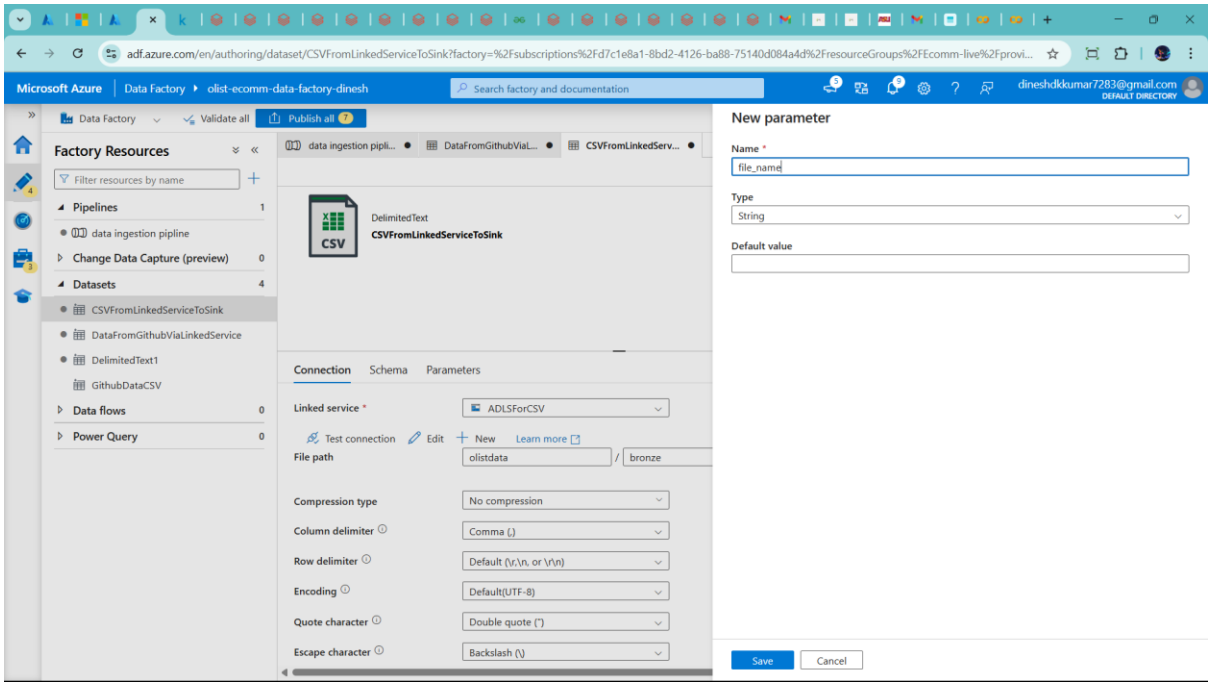

adf.azure.com/en/authoring/pipeline/data%20ingestion%20pipline?factory=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2Fprovi...
Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
Data flows
Power Query
data ingestion pipl...
DataFromGithubViaL...
Activities
cop
Move and transform
Copy data
Validate
Validate copy runtime
Debug
Copy data
Copy Data test
General
Source
Sink
Mapping
Settings
Sink dataset *
Select...
Set properties
Name
CSVFromLinkedServiceToSink
Linked service *
ADLSForCSV
File path
olistdata
bronze
File name
First row as header
Import schema
From connection/store
From sample file
None
Advanced
OK
Back
Cancel

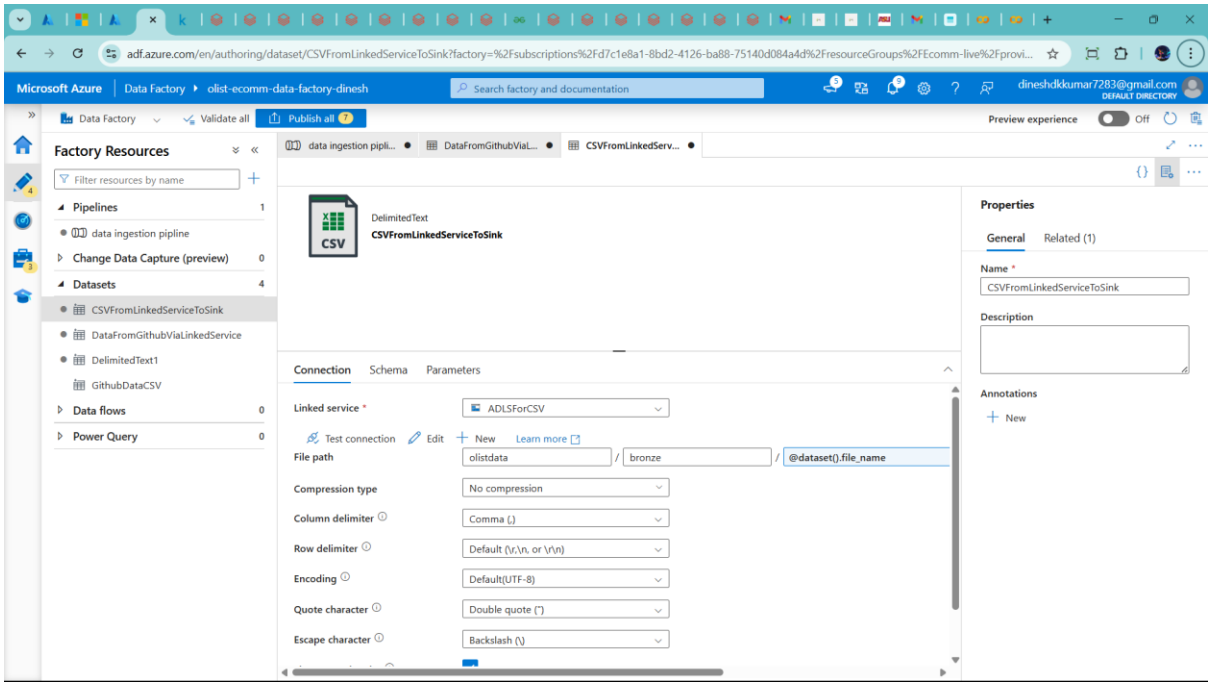


→add filename as dynamic

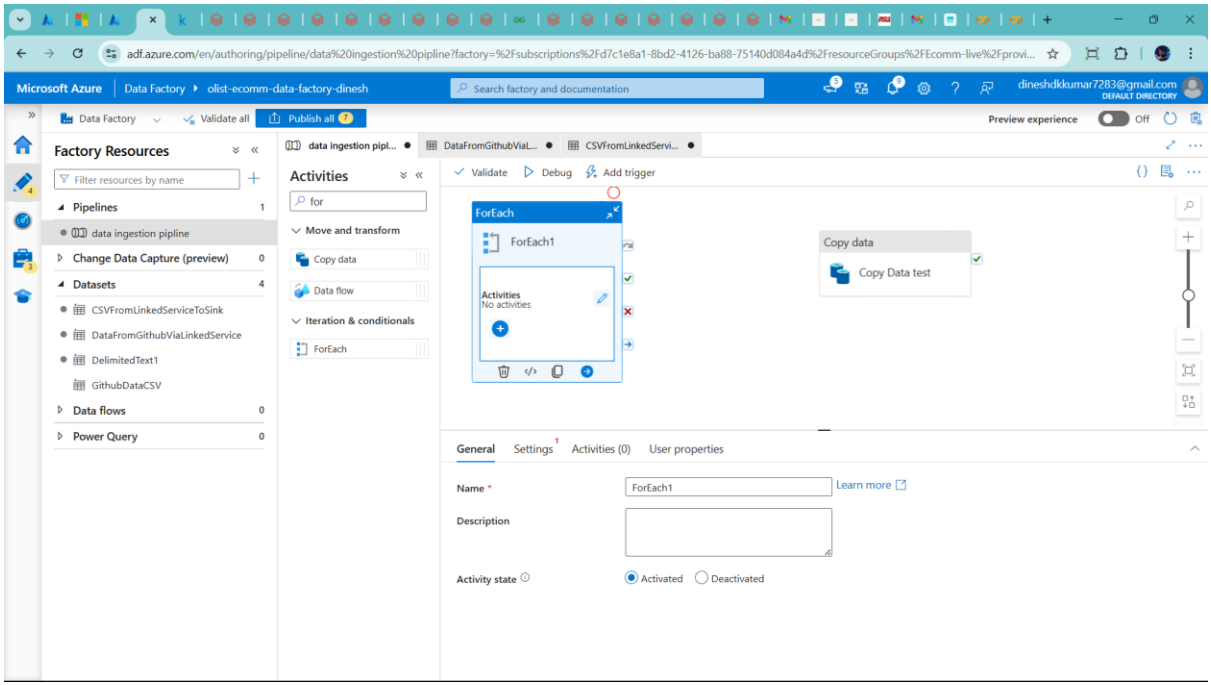


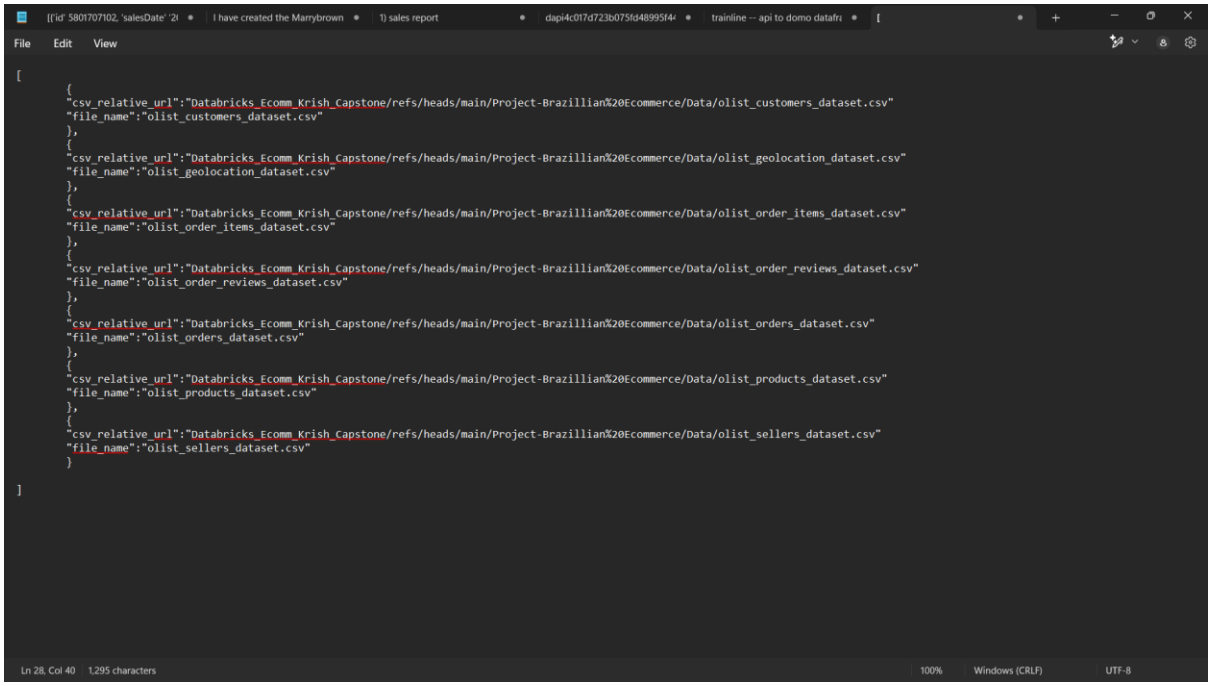
→click add dynamin and create new parameter





Now we For Each loop

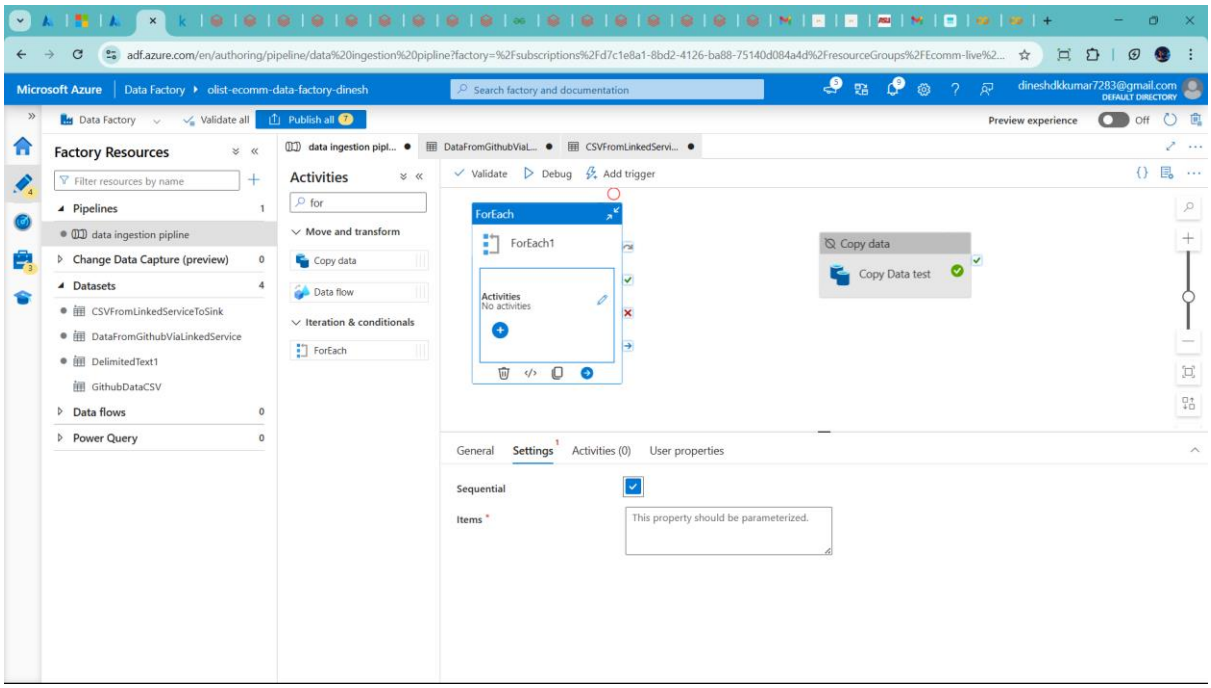

adf.azure.com/en/authoring/pipeline/data%20ingestion%20pipline?factory=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2Fprovi...
Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all 7
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 1
data ingestion pipline
Change Data Capture (preview) 0
Datasets 4
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
Data flows 0
Power Query 0
data ingestion pipl...
DataFromGithubViaL...
CSVFromLinkedServi...
Activities
for
Move and transform
Copy data
Data flow
Iteration & conditionals
ForEach
Validate
Debug
Add trigger
ForEach
ForEach1
Activities
No activities
Copy data
Copy Data test
General
Settings 1
Activities (0)
User properties
Name *
ForEach1
Learn more
Description
Activity state
Activated
Deactivated


File
Edit
View
[
{
"csv_relative_url":"Databricks_Ecomm_Krish_Capstone/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_customers_dataset.csv"
"file_name":"olist_customers_dataset.csv"
},
{
"csv_relative_url":"Databricks_Ecomm_Krish_Capstone/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_geolocation_dataset.csv"
"file_name":"olist_geolocation_dataset.csv"
},
{
"csv_relative_url":"Databricks_Ecomm_Krish_Capstone/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_order_items_dataset.csv"
"file_name":"olist_order_items_dataset.csv"
},
{
"csv_relative_url":"Databricks_Ecomm_Krish_Capstone/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_order_reviews_dataset.csv"
"file_name":"olist_order_reviews_dataset.csv"
},
{
"csv_relative_url":"Databricks_Ecomm_Krish_Capstone/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_orders_dataset.csv"
"file_name":"olist_orders_dataset.csv"
},
{
"csv_relative_url":"Databricks_Ecomm_Krish_Capstone/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_products_dataset.csv"
"file_name":"olist_products_dataset.csv"
},
{
"csv_relative_url":"Databricks_Ecomm_Krish_Capstone/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_sellers_dataset.csv"
"file_name":"olist_sellers_dataset.csv"
}
]
Ln 28, Col 40
1,295 characters
100%
Windows (CRLF)
UTF-8

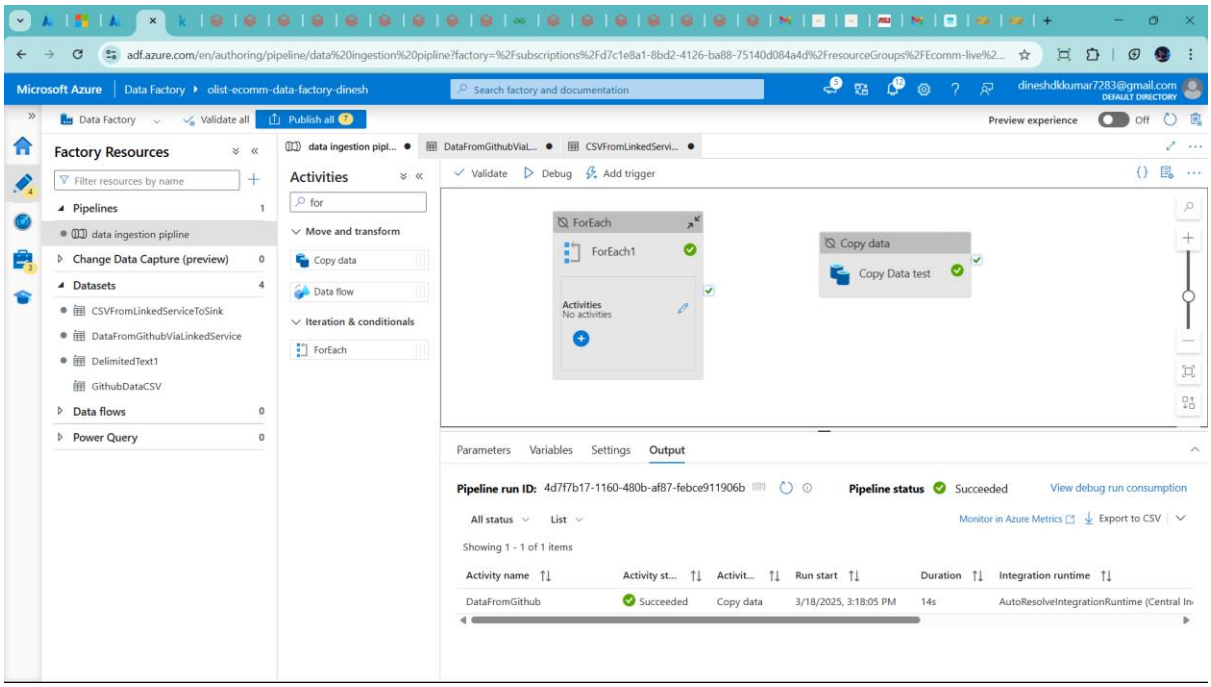
→enable sequincal

Sequencial →line by line

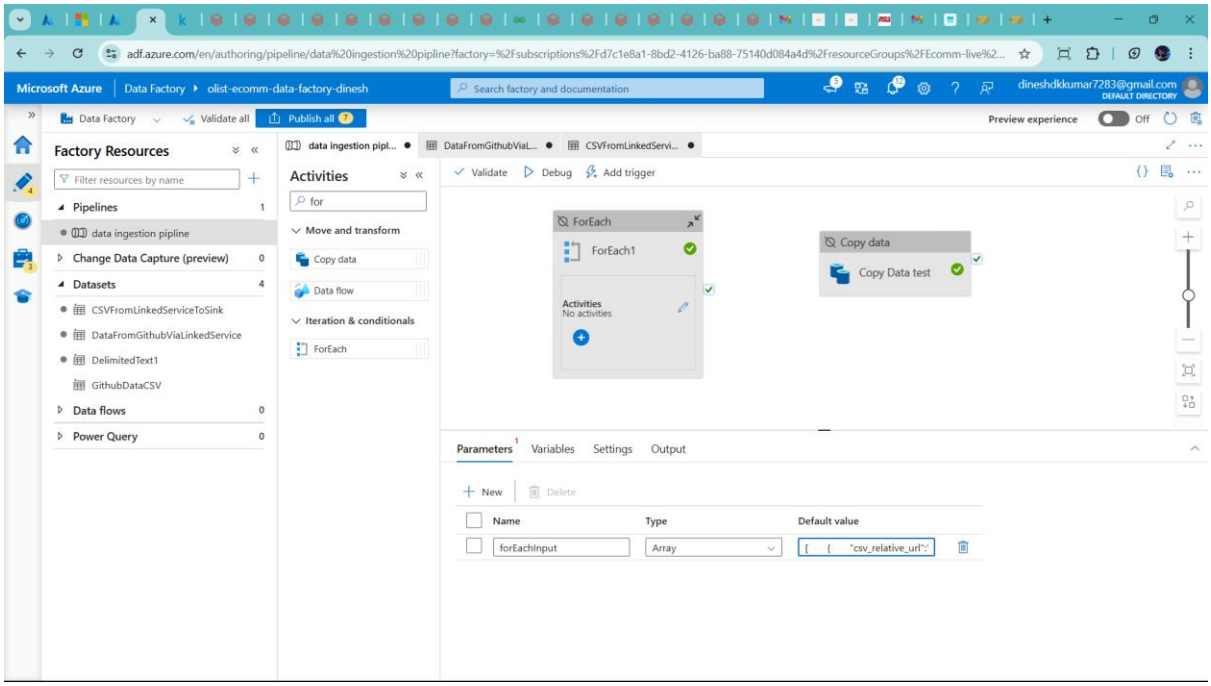
Batch count → splilt to run paraelle



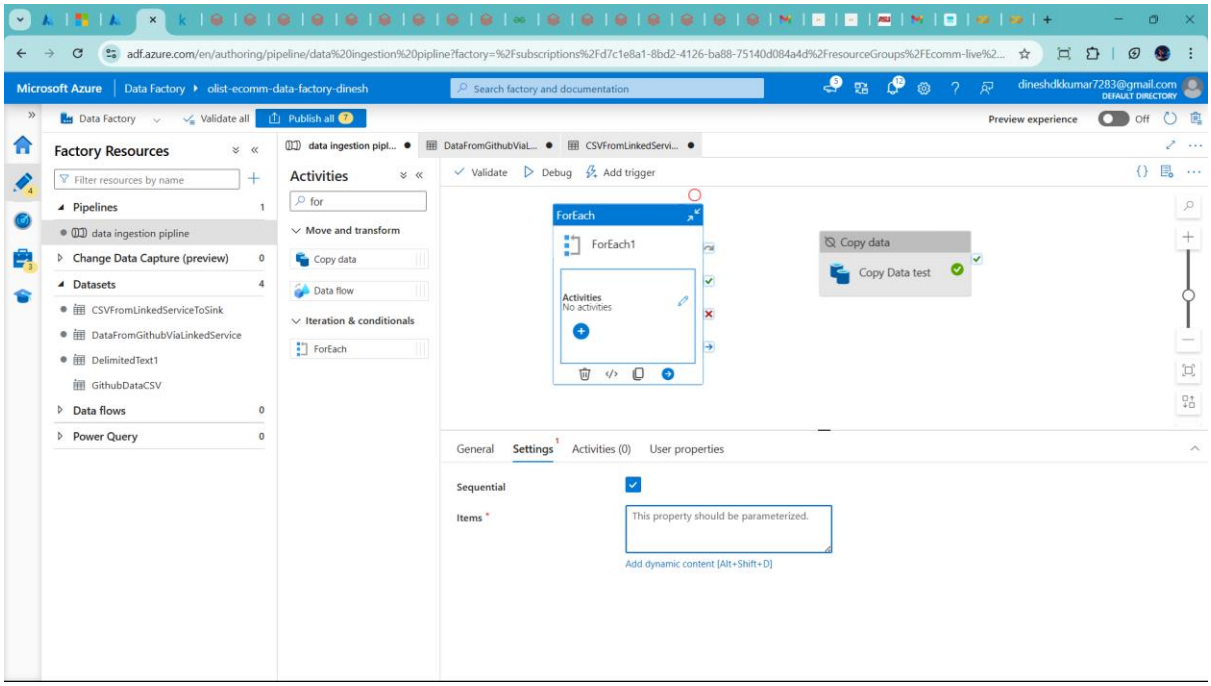
→click outside of activity

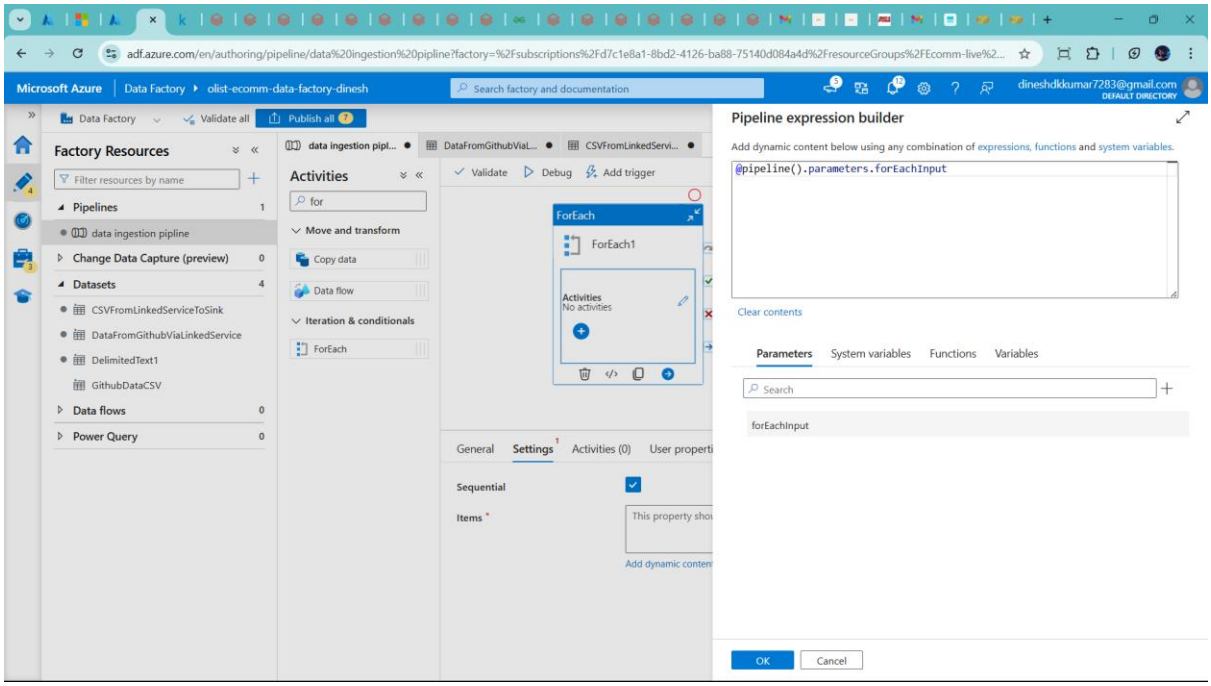


Click paramenter

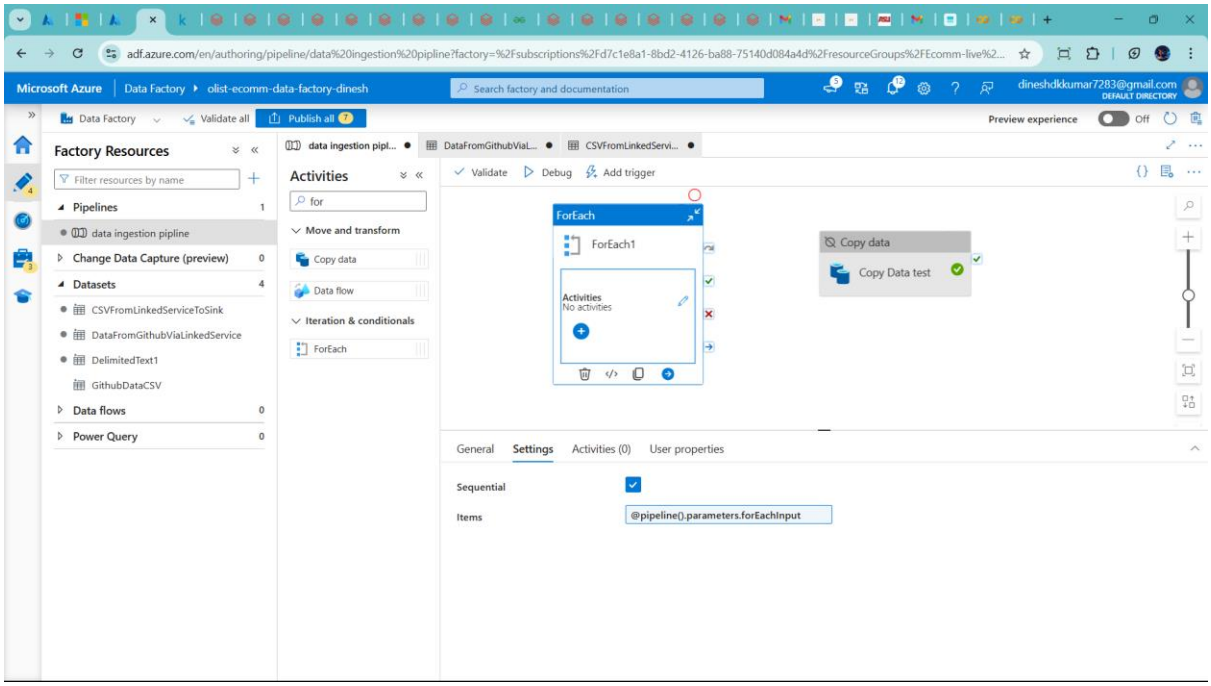


→click setting →add dynamic content



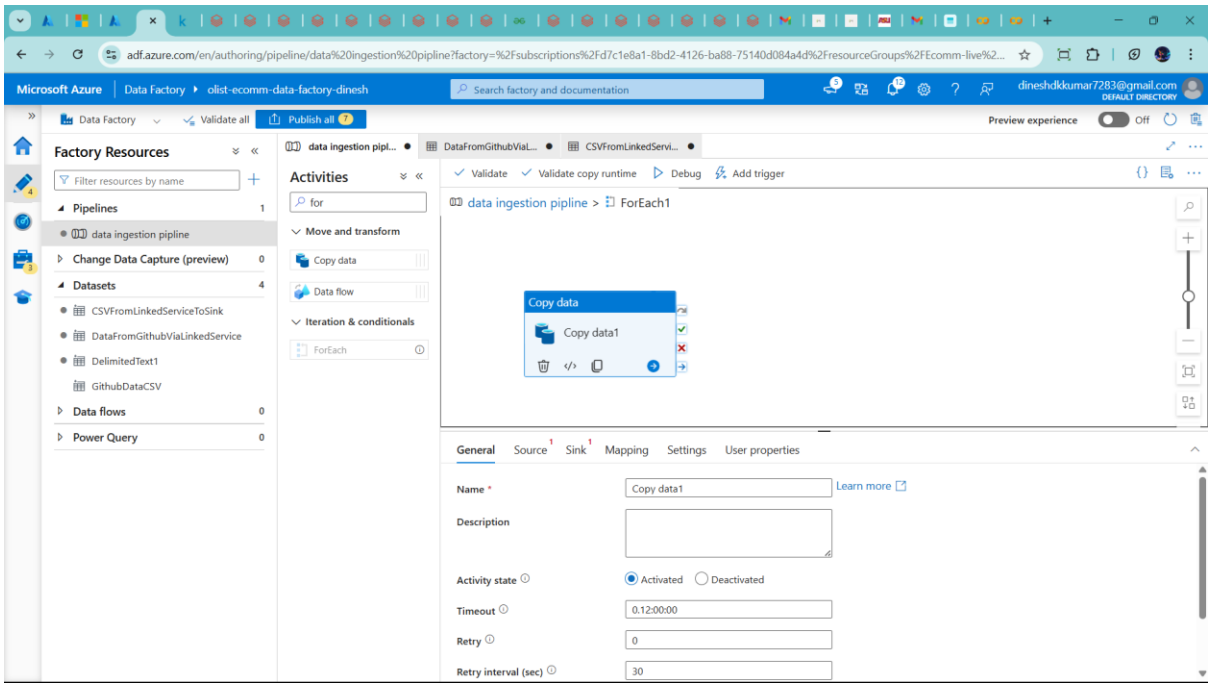

Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
Data flows
Power Query
Activities
for
Move and transform
Copy data
Data flow
Iteration & conditionals
ForEach
Validate
Debug
Add trigger
ForEach1
Activities
No activities
General
Settings
Activities (0)
Sequential
Items
Pipeline expression builder
Add dynamic content below using any combination of expressions, functions and system variables.
@pipeline().parameters.forEachInput
Clear contents
Parameters
System variables
Functions
Variables
Search
forEachInput
OK
Cancel

→click ok

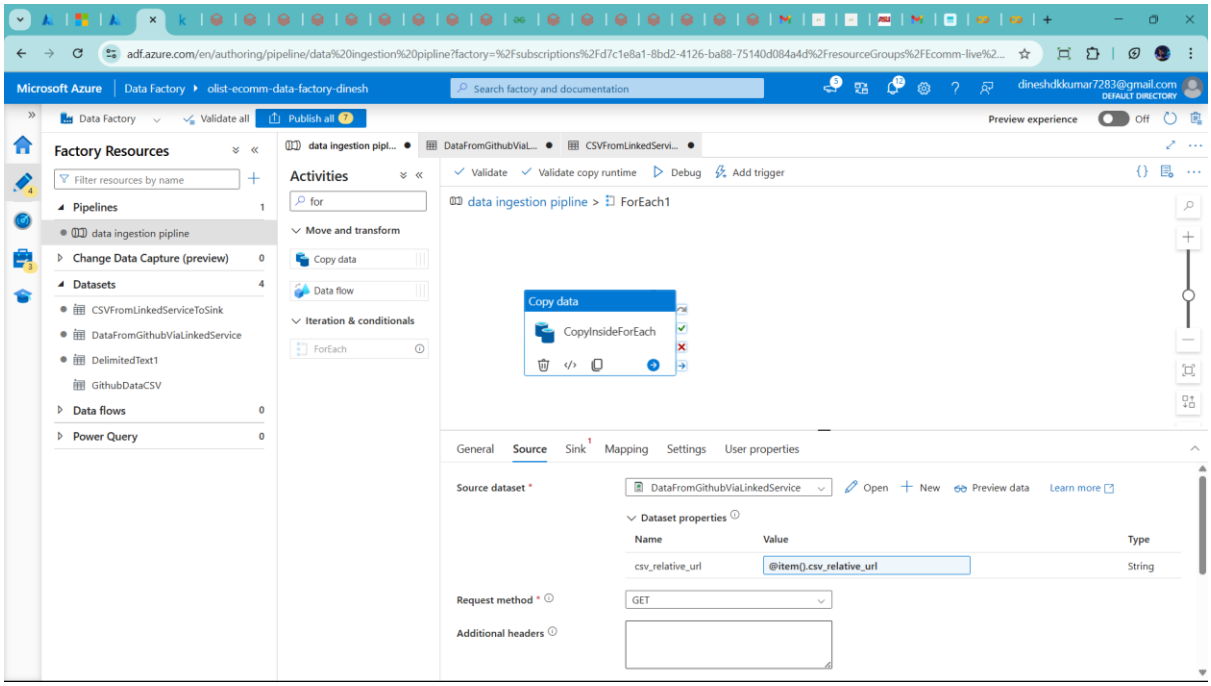

Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
Data flows
Power Query
Activities
for
Move and transform
Copy data
Data flow
Iteration & conditionals
ForEach
Validate
Debug
Add trigger
ForEach1
Activities
No activities
Copy data
Copy Data test
General
Settings
Activities (0)
User properties
Sequential
Items
@pipeline().parameters.forEachInput

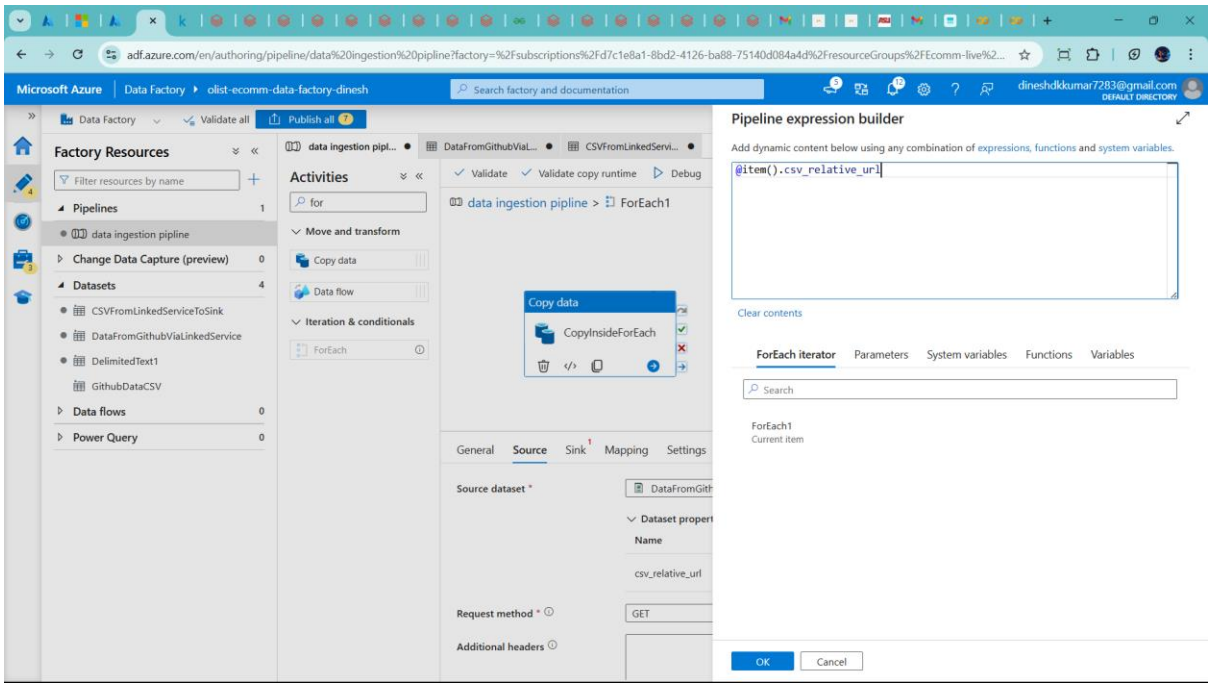
→

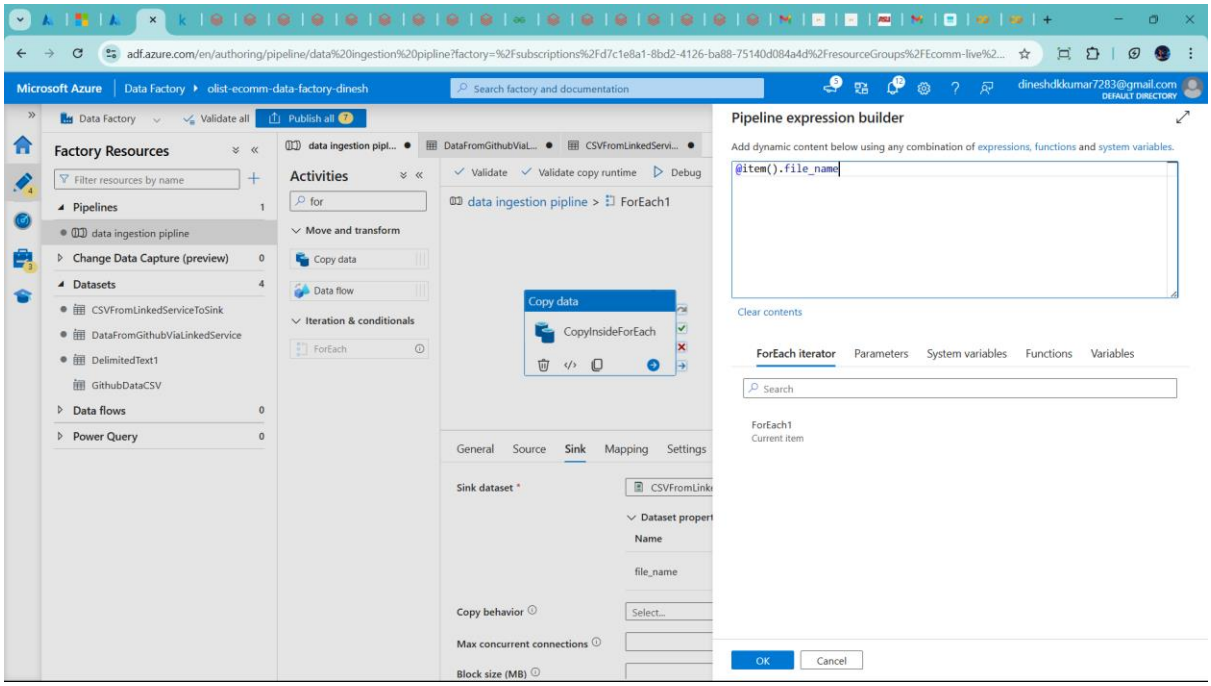
→Click for each edit pencile →drag copy data



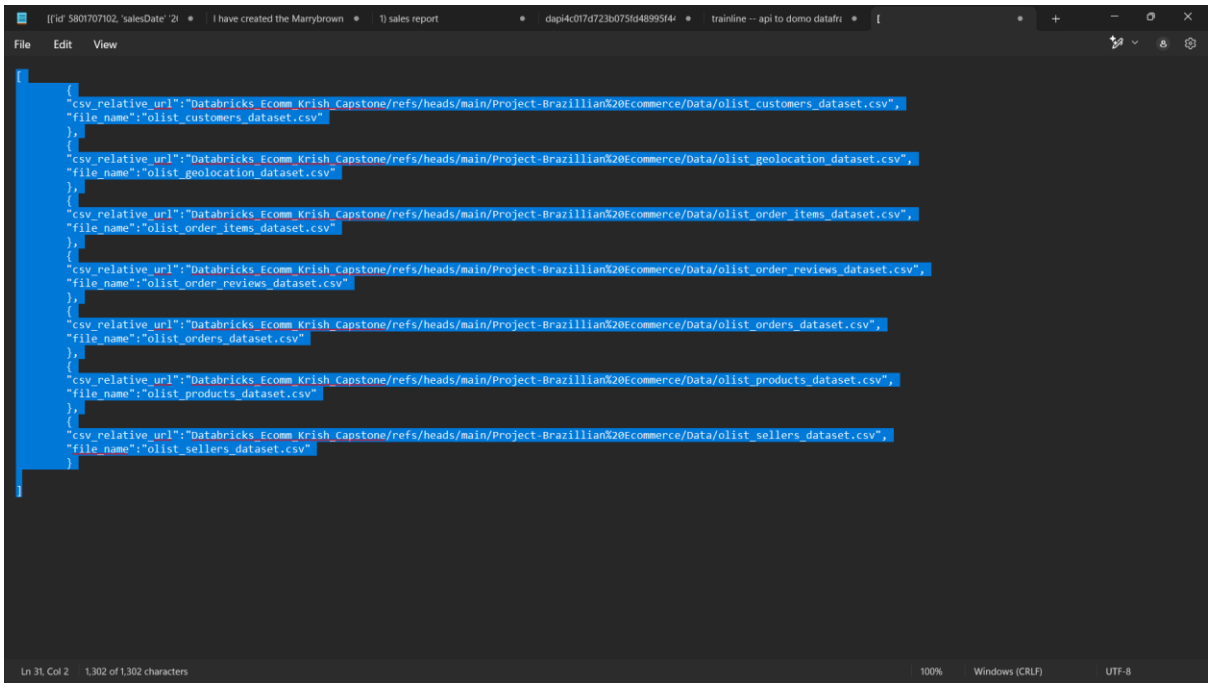
→add source and sink link same as above



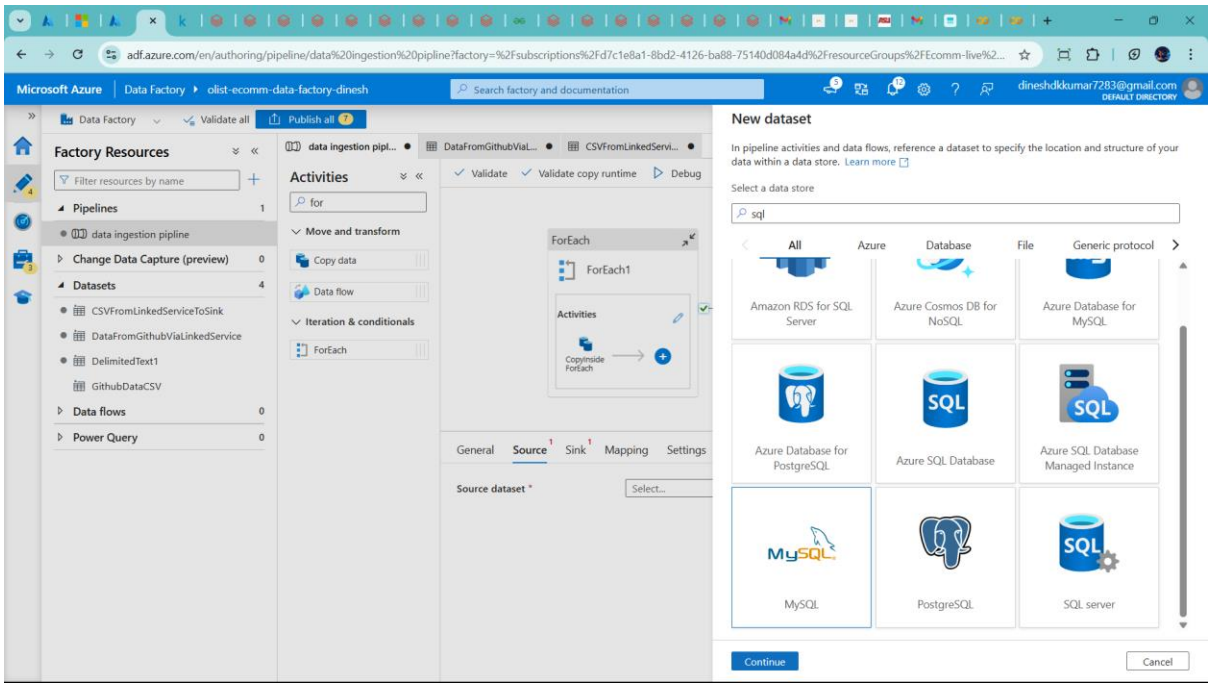

adf.azure.com/en/authoring/pipeline/data%20ingestion%20pipline?factory=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2...
Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
Data flows
Power Query
data ingestion pipl...
DataFromGithubViaL...
CSVFromLinkedServi...
Activities
for
Move and transform
Copy data
Data flow
Iteration & conditionals
ForEach
Validate
Validate copy runtime
Debug
data ingestion pipline > ForEach1
Copy data
CopyInsideForEach
General
Source
Sink
Mapping
Settings
Source dataset *
DataFromGith
Dataset propert
Name
csv_relative_url
Request method *
GET
Additional headers
Pipeline expression builder
Add dynamic content below using any combination of expressions, functions and system variables.
@item().csv_relative_url
Clear contents
ForEach iterator
Parameters
System variables
Functions
Variables
Search
ForEach1
Current item
OK
Cancel

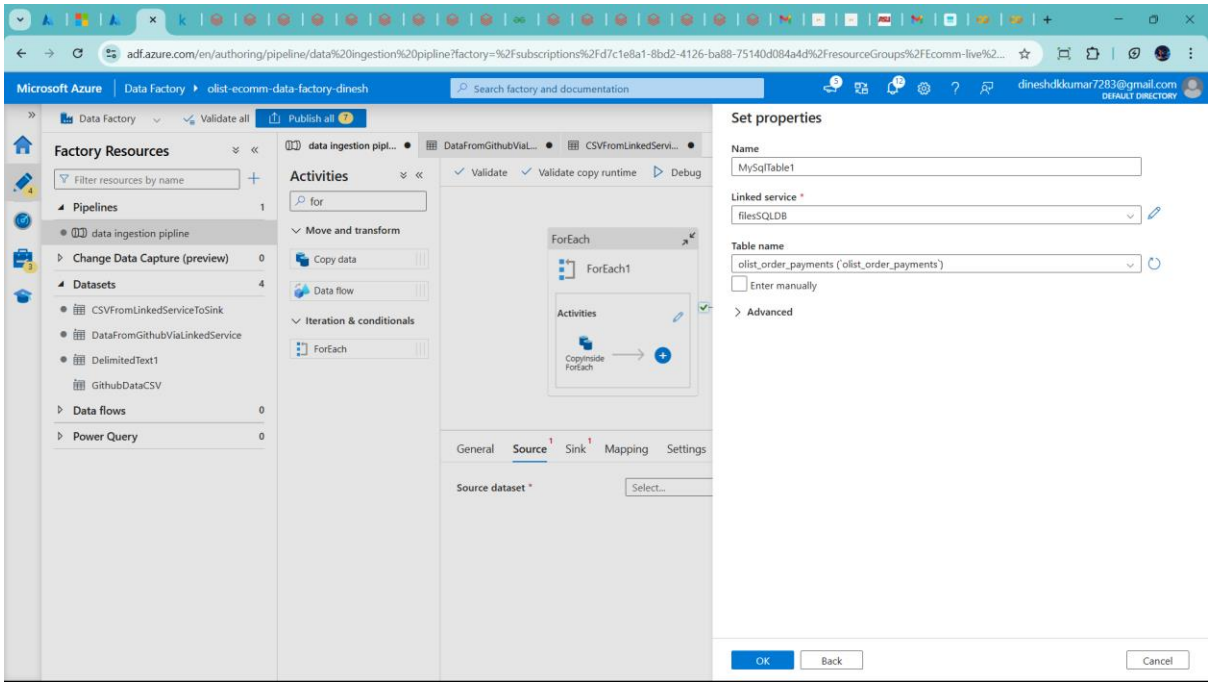

adf.azure.com/en/authoring/pipeline/data%20ingestion%20pipline?factory=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2...
Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
Data flows
Power Query
data ingestion pipl...
DataFromGithubViaL...
CSVFromLinkedServi...
Activities
for
Move and transform
Copy data
Data flow
Iteration & conditionals
ForEach
Validate
Validate copy runtime
Debug
data ingestion pipline > ForEach1
Copy data
CopyInsideForEach
General
Source
Sink
Mapping
Settings
Sink dataset *
CSVFromLinke
Dataset propert
Name
file_name
Copy behavior
Select...
Max concurrent connections
Block size (MB)
Pipeline expression builder
Add dynamic content below using any combination of expressions, functions and system variables.
@item().file_name
Clear contents
ForEach iterator
Parameters
System variables
Functions
Variables
Search
ForEach1
Current item
OK
Cancel

→correct

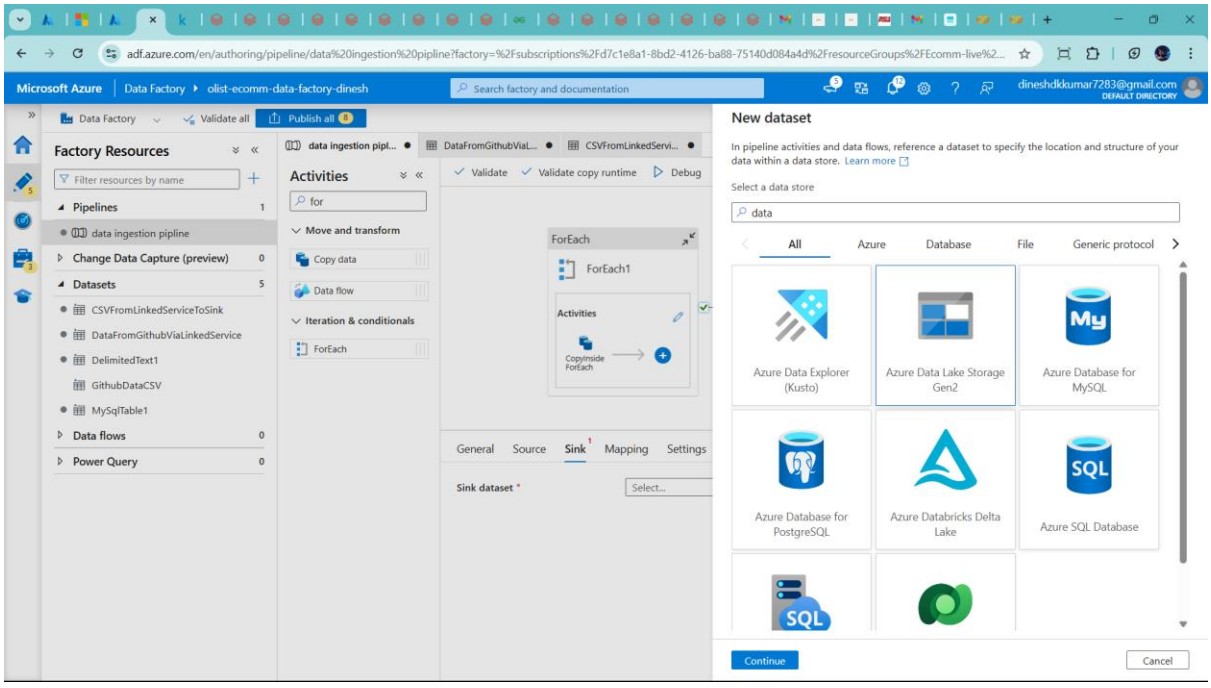


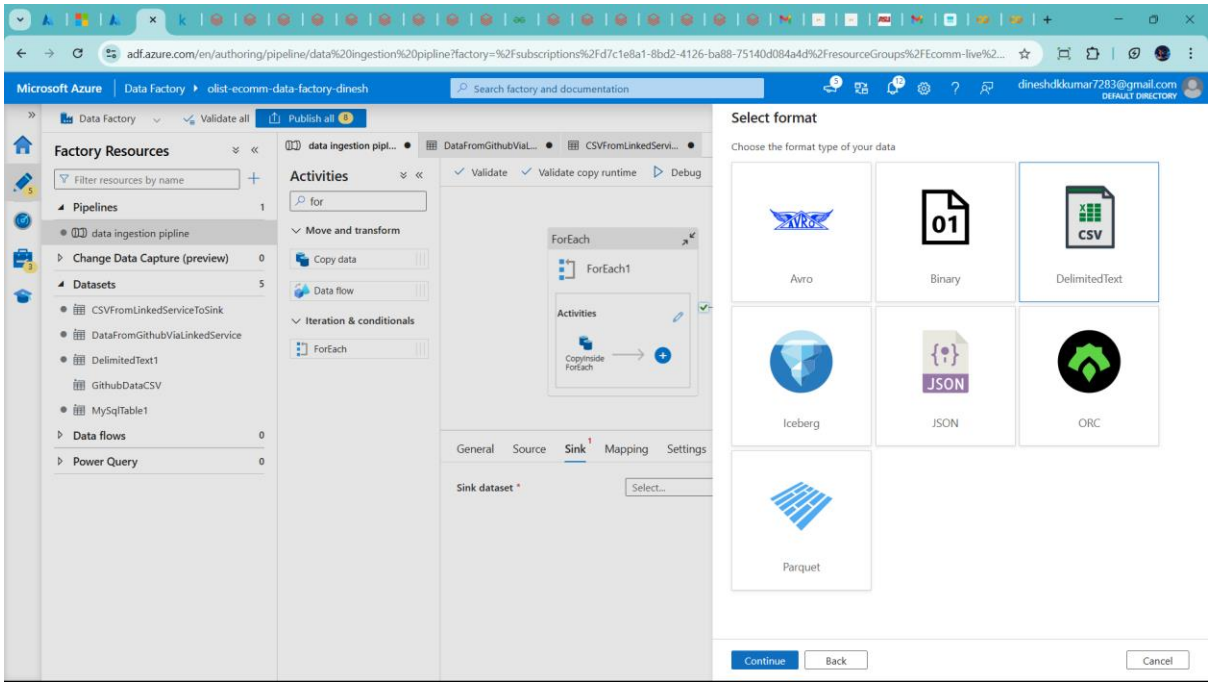
→Now add SQL source in this copy activity

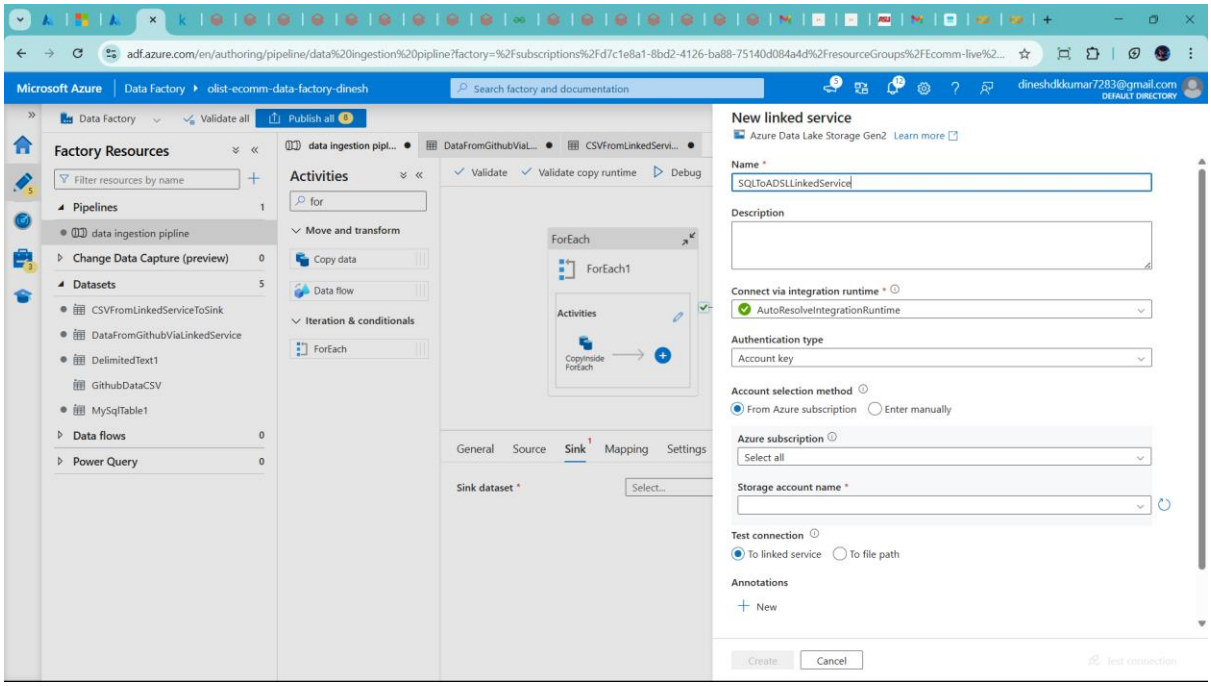


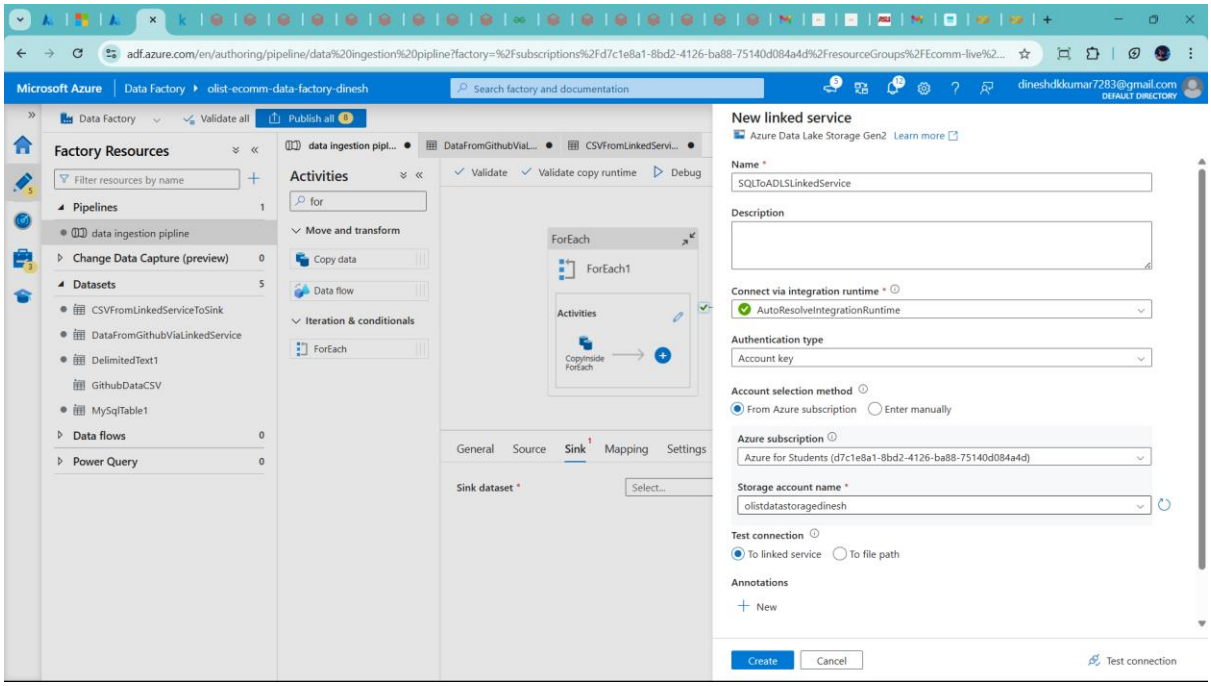

Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
Data flows
Power Query
Activities
for
Move and transform
Copy data
Data flow
Iteration & conditionals
ForEach
Validate
Validate copy runtime
Debug
ForEach1
Activities
CopyInside ForEach
General
Source
Sink
Mapping
Settings
Source dataset *
Select...
Set properties
Name
MySqlTable1
Linked service *
filesSQLDB
Table name
olist_order_payments ('olist_order_payments')
Enter manually
Advanced
OK
Back
Cancel

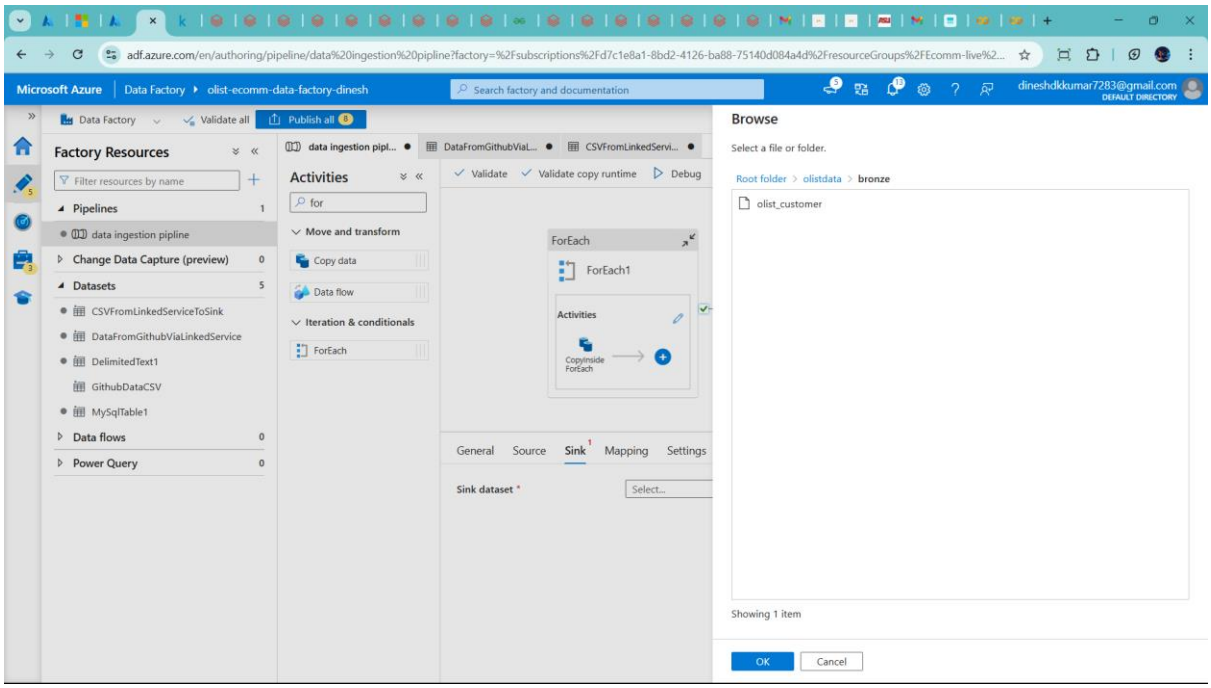
→in sink

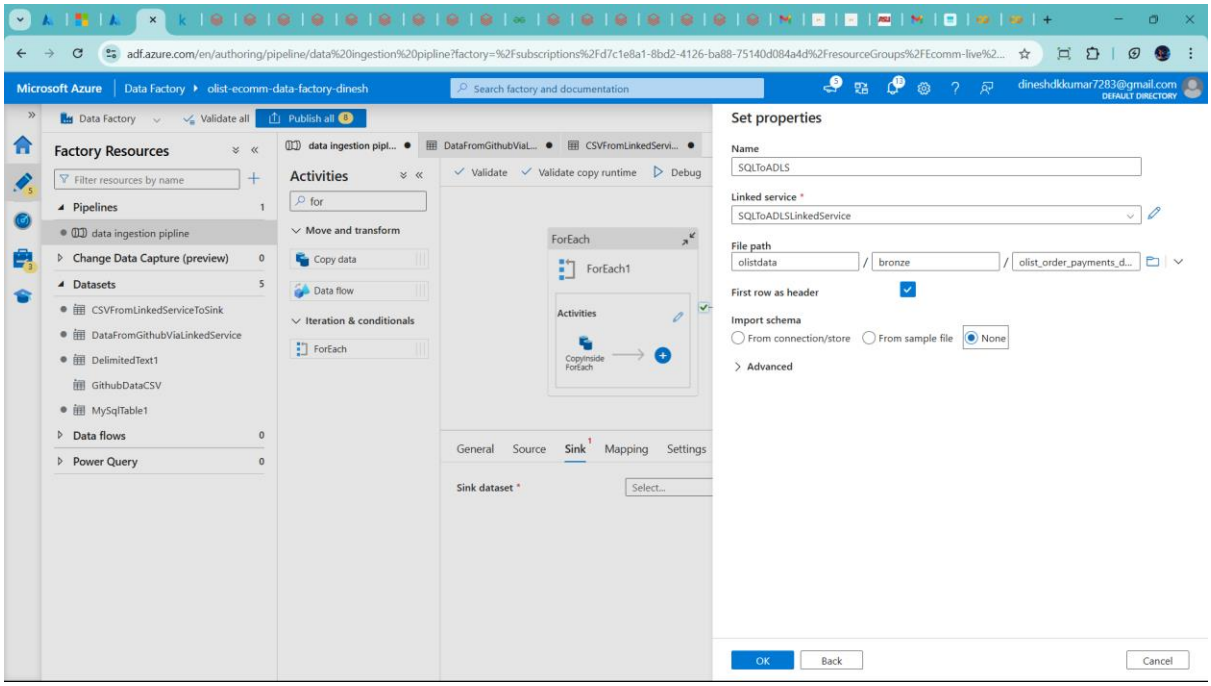

Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
MySqlTable1
Data flows
Power Query
Activities
for
Move and transform
Copy data
Data flow
Iteration & conditionals
ForEach
Validate
Validate copy runtime
Debug
ForEach1
Activities
CopyInside ForEach
General
Source
Sink
Mapping
Settings
Sink dataset *
Select...
New dataset
In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. Learn more
Select a data store
data
All
Azure
Database
File
Generic protocol
Azure Data Explorer (Kusto)
Azure Data Lake Storage Gen2
Azure Database for MySQL
Azure Database for PostgreSQL
Azure Databricks Delta Lake
Azure SQL Database
Continue
Cancel

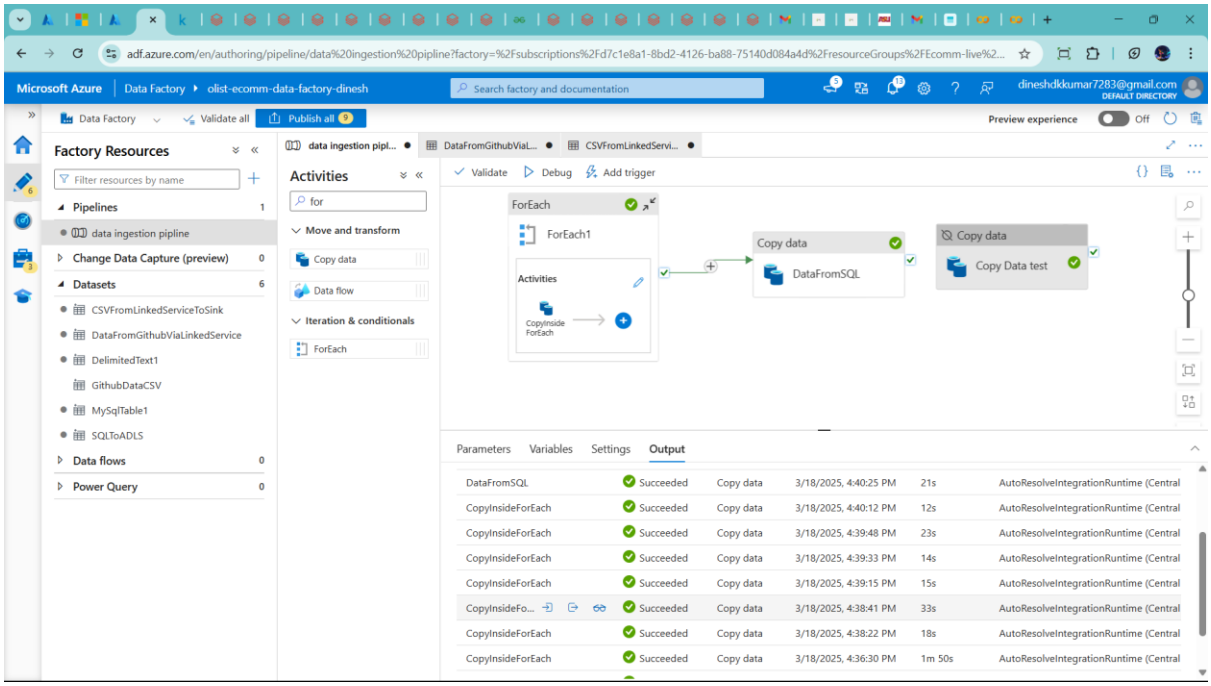

Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
MySqlTable1
Data flows
Power Query
Activities
for
Move and transform
Copy data
Data flow
Iteration & conditionals
ForEach
Validate
Validate copy runtime
Debug
ForEach1
Activities
CopyInside ForEach
General
Source
Sink
Mapping
Settings
Sink dataset *
Select...
Select format
Choose the format type of your data
Avro
Binary
DelimitedText
Iceberg
JSON
ORC
Parquet
Continue
Back
Cancel


Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
MySqlTable1
Data flows
Power Query
Activities
for
Move and transform
Copy data
Data flow
Iteration & conditionals
ForEach
Validate
Validate copy runtime
Debug
ForEach1
Activities
CopyInside ForEach
General
Source
Sink
Mapping
Settings
Sink dataset *
Select...
New linked service
Azure Data Lake Storage Gen2
Learn more
Name *
SQLToADSLLinkedService
Description
Connect via integration runtime *
AutoResolveIntegrationRuntime
Authentication type
Account key
Account selection method
From Azure subscription
Enter manually
Azure subscription
Select all
Storage account name *
Test connection
To linked service
To file path
Annotations
New
Create
Cancel
Test connection

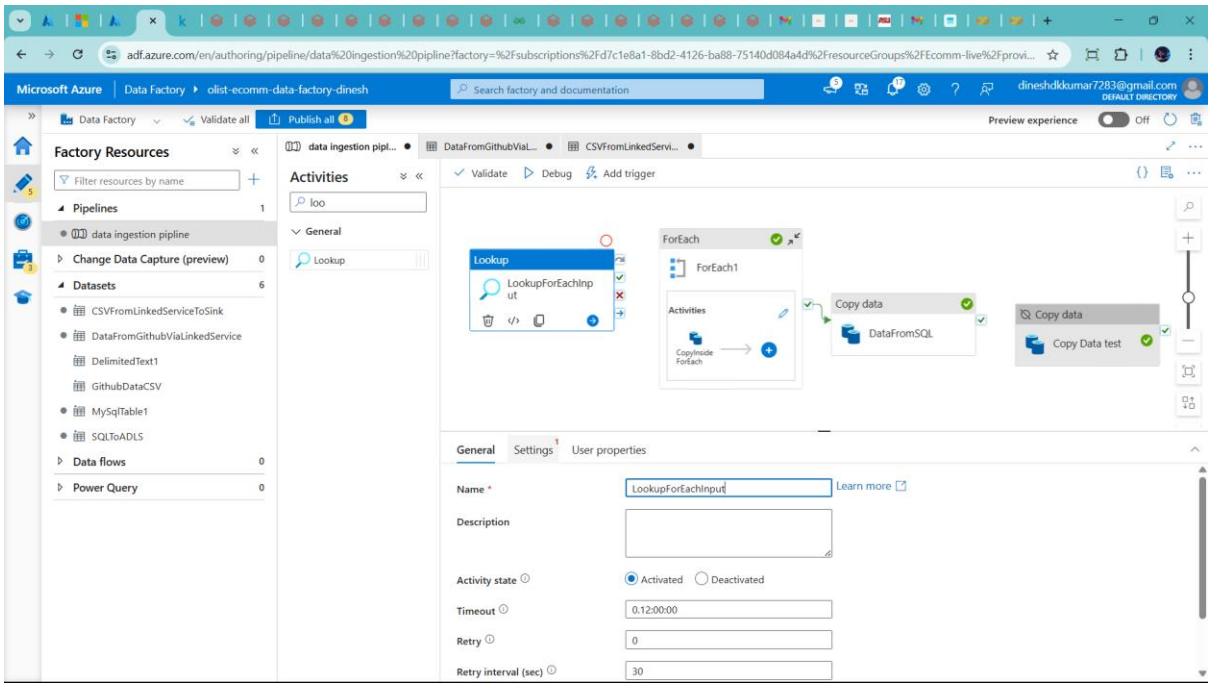

adf.azure.com/en/authoring/pipeline/data%20ingestion%20pipline?factory=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2...
Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
MySqlTable1
Data flows
Power Query
data ingestion pipl...
DataFromGithubViaL...
CSVFromLinkedServi...
Activities
for
Move and transform
Copy data
Data flow
Iteration & conditionals
ForEach
Validate
Validate copy runtime
Debug
ForEach
ForEach1
Activities
CopyInside ForEach
General
Source
Sink
Mapping
Settings
Sink dataset *
Select...
New linked service
Azure Data Lake Storage Gen2
Learn more
Name *
SQLToADLSLinkedService
Description
Connect via integration runtime *
AutoResolveIntegrationRuntime
Authentication type
Account key
Account selection method
From Azure subscription
Enter manually
Azure subscription
Azure for Students (d7c1e8a1-8bd2-4126-ba88-75140d084a4d)
Storage account name *
olistdatastoragedinesh
Test connection
To linked service
To file path
Annotations
New
Create
Cancel
Test connection

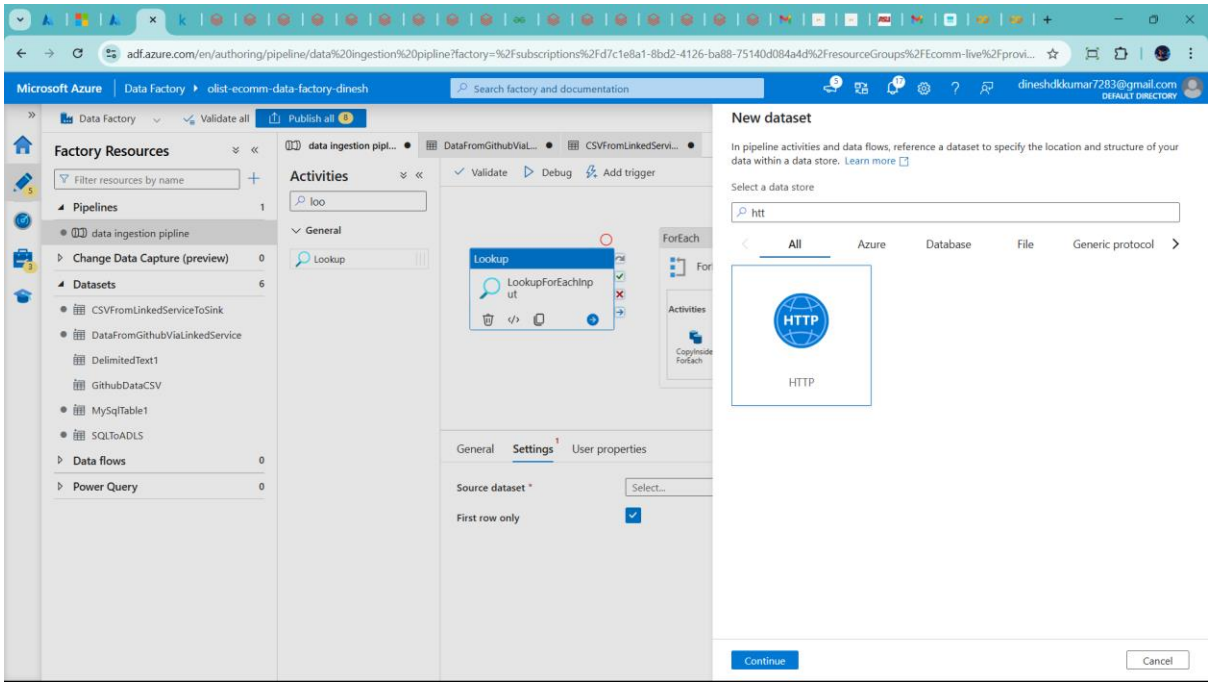

adf.azure.com/en/authoring/pipeline/data%20ingestion%20pipline?factory=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2...
Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
MySqlTable1
Data flows
Power Query
data ingestion pipl...
DataFromGithubViaL...
CSVFromLinkedServi...
Activities
for
Move and transform
Copy data
Data flow
Iteration & conditionals
ForEach
Validate
Validate copy runtime
Debug
ForEach
ForEach1
Activities
CopyInside ForEach
General
Source
Sink
Mapping
Settings
Sink dataset *
Select...
Browse
Select a file or folder.
Root folder
olistdata
bronze
olist_customer
Showing 1 item
OK
Cancel

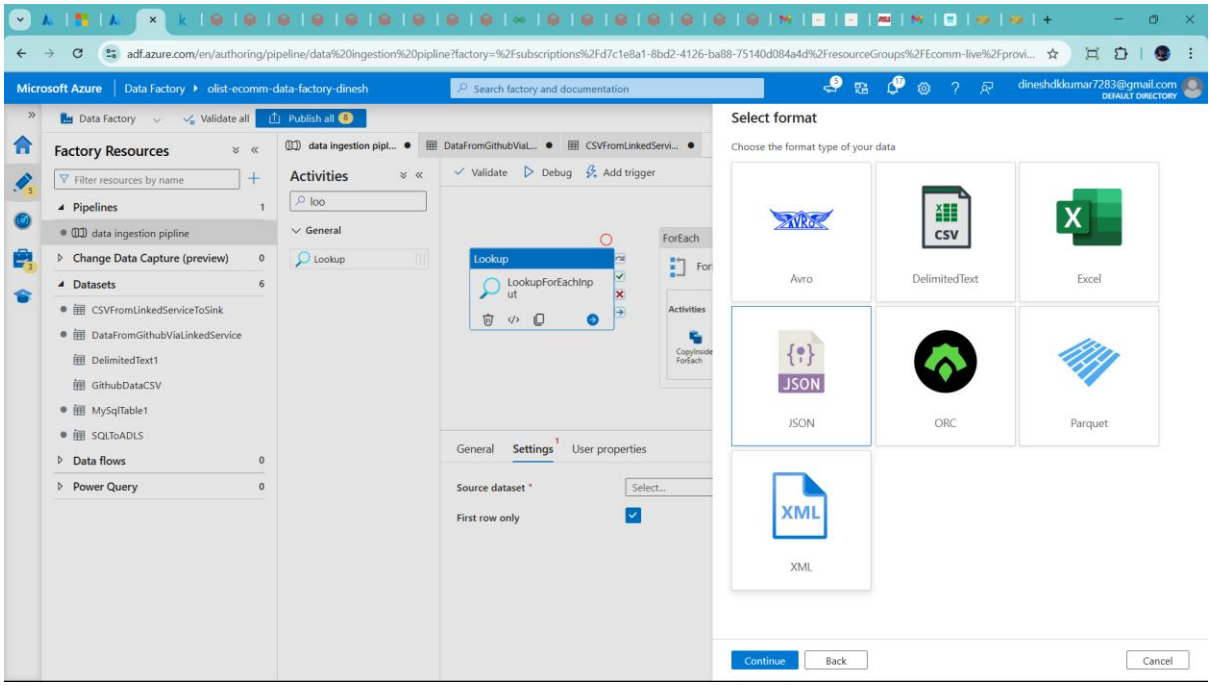

adf.azure.com/en/authoring/pipeline/data%20ingestion%20pipline?factory=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2...
Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines 1
data ingestion pipline
Change Data Capture (preview) 0
Datasets 5
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
MySqlTable1
Data flows 0
Power Query 0
data ingestion pipl...
DataFromGithubViaL...
CSVFromLinkedServi...
Activities
for
Move and transform
Copy data
Data flow
Iteration & conditionals
ForEach
Validate
Validate copy runtime
Debug
ForEach
ForEach1
Activities
CopyInside ForEach
General
Source
Sink 1
Mapping
Settings
Sink dataset *
Select...
Set properties
Name
SQLToADLS
Linked service *
SQLToADLSLinkedService
File path
olistdata
bronze
olist_order_payments_d...
First row as header
Import schema
From connection/store
From sample file
None
Advanced
OK
Back
Cancel

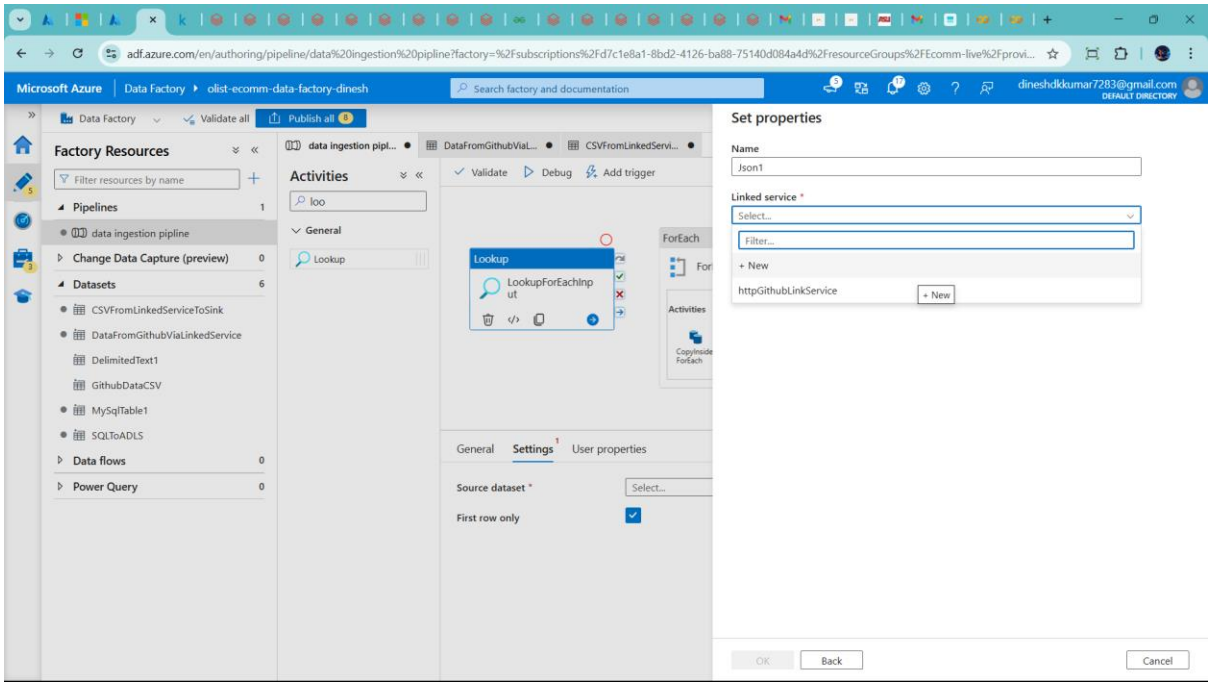

adf.azure.com/en/authoring/pipeline/data%20ingestion%20pipline?factory=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2...
Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all 9
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 1
data ingestion pipline
Change Data Capture (preview) 0
Datasets 6
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
MySqlTable1
SQLToADLS
Data flows 0
Power Query 0
Activities
for
Move and transform
Copy data
Data flow
Iteration & conditionals
ForEach
Validate
Debug
Add trigger
ForEach
ForEach1
Activities
CopyInside ForEach
Copy data
DataFromSQL
Copy data
Copy Data test
Parameters
Variables
Settings
Output
DataFromSQL Succeeded Copy data 3/18/2025, 4:40:25 PM 21s AutoResolveIntegrationRuntime (Central
CopyInsideForEach Succeeded Copy data 3/18/2025, 4:40:12 PM 12s AutoResolveIntegrationRuntime (Central
CopyInsideForEach Succeeded Copy data 3/18/2025, 4:39:48 PM 23s AutoResolveIntegrationRuntime (Central
CopyInsideForEach Succeeded Copy data 3/18/2025, 4:39:33 PM 14s AutoResolveIntegrationRuntime (Central
CopyInsideForEach Succeeded Copy data 3/18/2025, 4:39:15 PM 15s AutoResolveIntegrationRuntime (Central
CopyInsideFo... Succeeded Copy data 3/18/2025, 4:38:41 PM 33s AutoResolveIntegrationRuntime (Central
CopyInsideForEach Succeeded Copy data 3/18/2025, 4:38:22 PM 18s AutoResolveIntegrationRuntime (Central
CopyInsideForEach Succeeded Copy data 3/18/2025, 4:36:30 PM 1m 50s AutoResolveIntegrationRuntime (Central

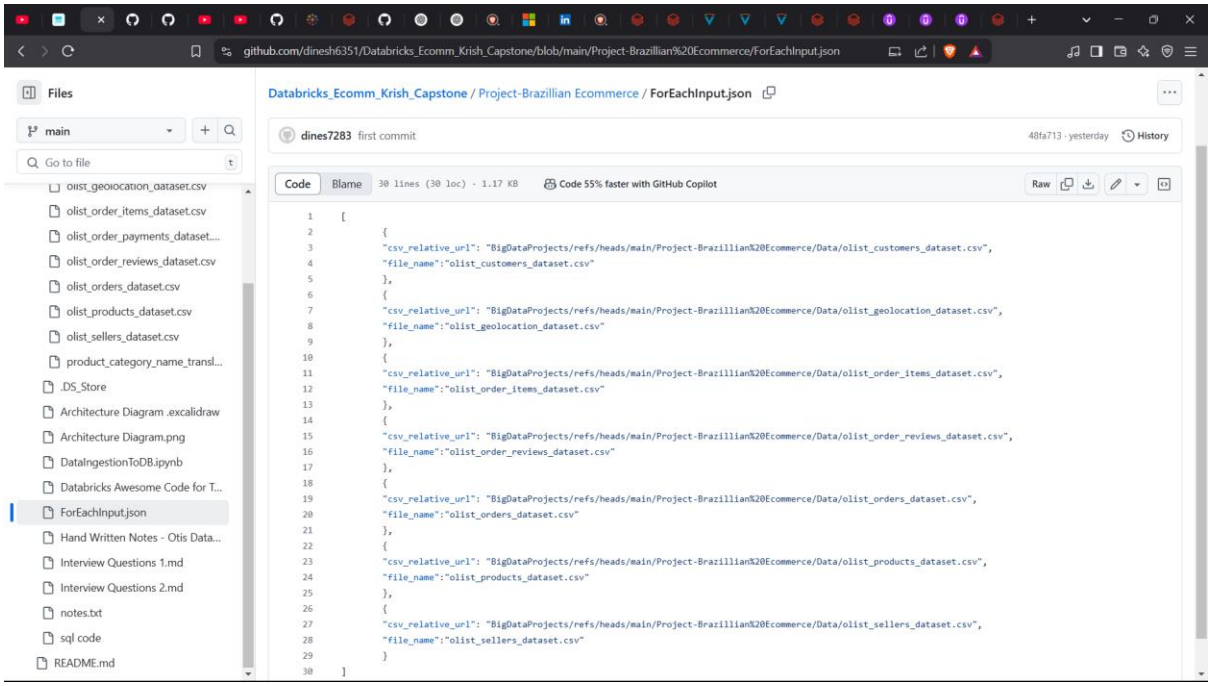
→Add Lookup

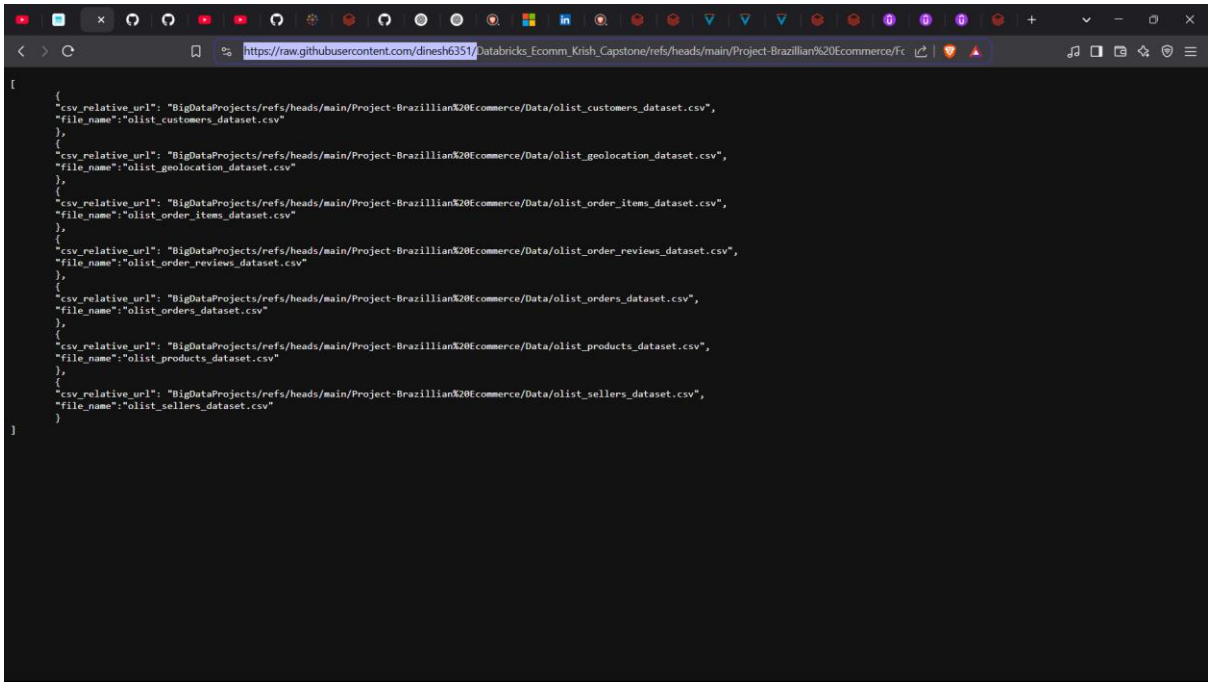

Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
Factory Resources
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
MySqlTable1
SQLToADLS
Data flows
Power Query
Activities
loo
General
Lookup
Validate
Debug
Add trigger
ForEach
ForEach1
Copy data
DataFromSQL
Copy Data test
General
Settings
User properties
Name *
LookupForEachInput
Learn more
Description
Activity state
Activated
Deactivated
Timeout
0.12:00:00
Retry
0
Retry interval (sec)
30

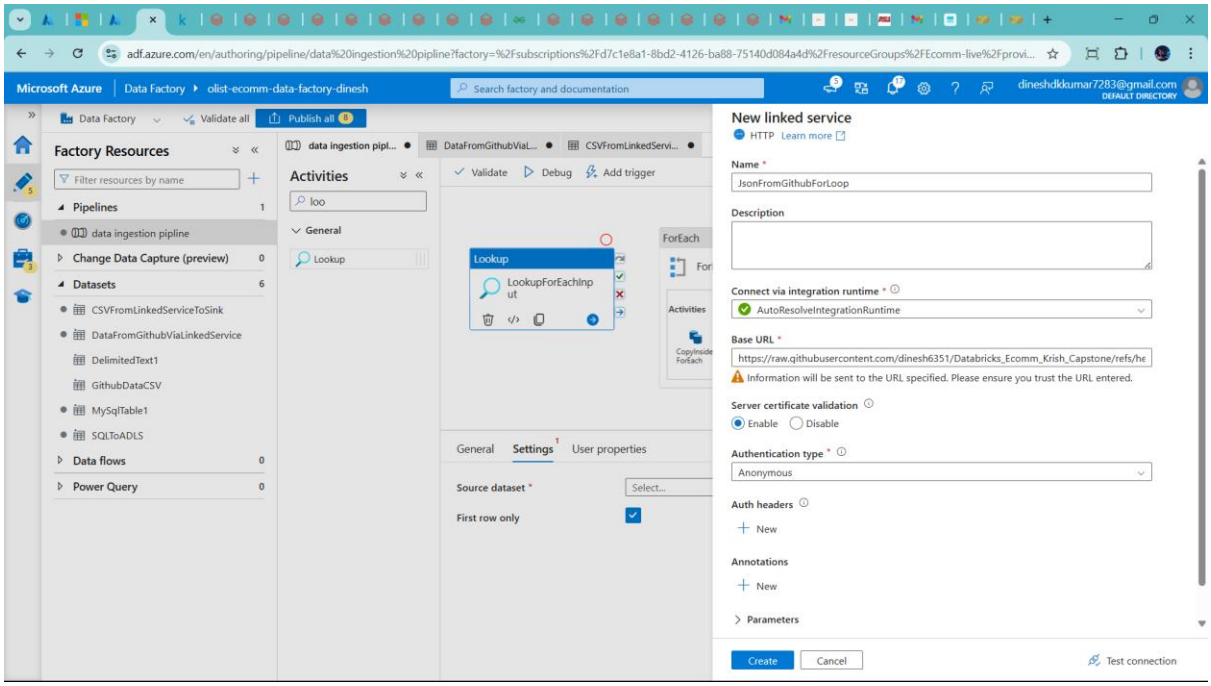

Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
Factory Resources
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
MySqlTable1
SQLToADLS
Data flows
Power Query
Activities
loo
General
Lookup
Validate
Debug
Add trigger
General
Settings
User properties
Source dataset *
Select...
First row only
New dataset
In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. Learn more
Select a data store
htt
All
Azure
Database
File
Generic protocol
HTTP
Continue
Cancel

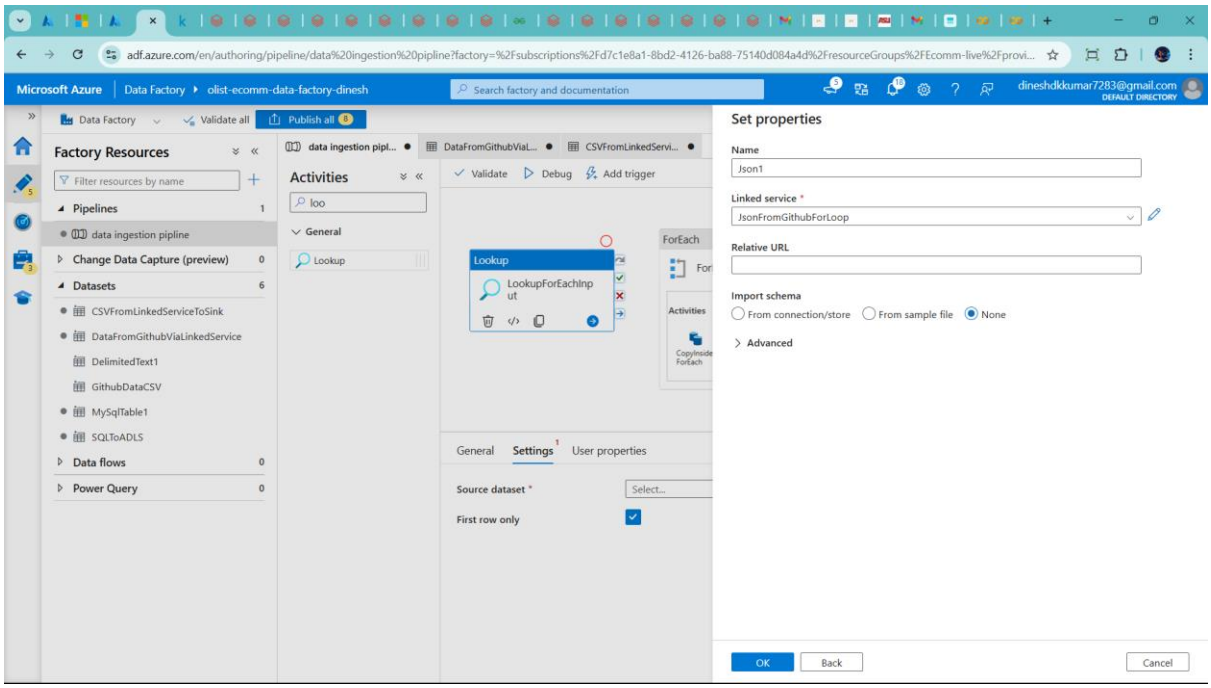

adf.azure.com/en/authoring/pipeline/data%20ingestion%20pipline?factory=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2Fprovi...
Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
MySqlTable1
SQLToADLS
Data flows
Power Query
Activities
loo
General
Lookup
Validate
Debug
Add trigger
ForEach
Lookup
LookupForEachInput
General
Settings
User properties
Source dataset *
Select...
First row only
Select format
Choose the format type of your data
Avro
DelimitedText
Excel
JSON
ORC
Parquet
XML
Continue
Back
Cancel


adf.azure.com/en/authoring/pipeline/data%20ingestion%20pipline?factory=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2Fprovi...
Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
MySqlTable1
SQLToADLS
Data flows
Power Query
Activities
loo
General
Lookup
Validate
Debug
Add trigger
ForEach
Lookup
LookupForEachInput
General
Settings
User properties
Source dataset *
Select...
First row only
Set properties
Name
Json1
Linked service *
Select...
Filter...
+ New
httpGithubLinkService
+ New
OK
Back
Cancel

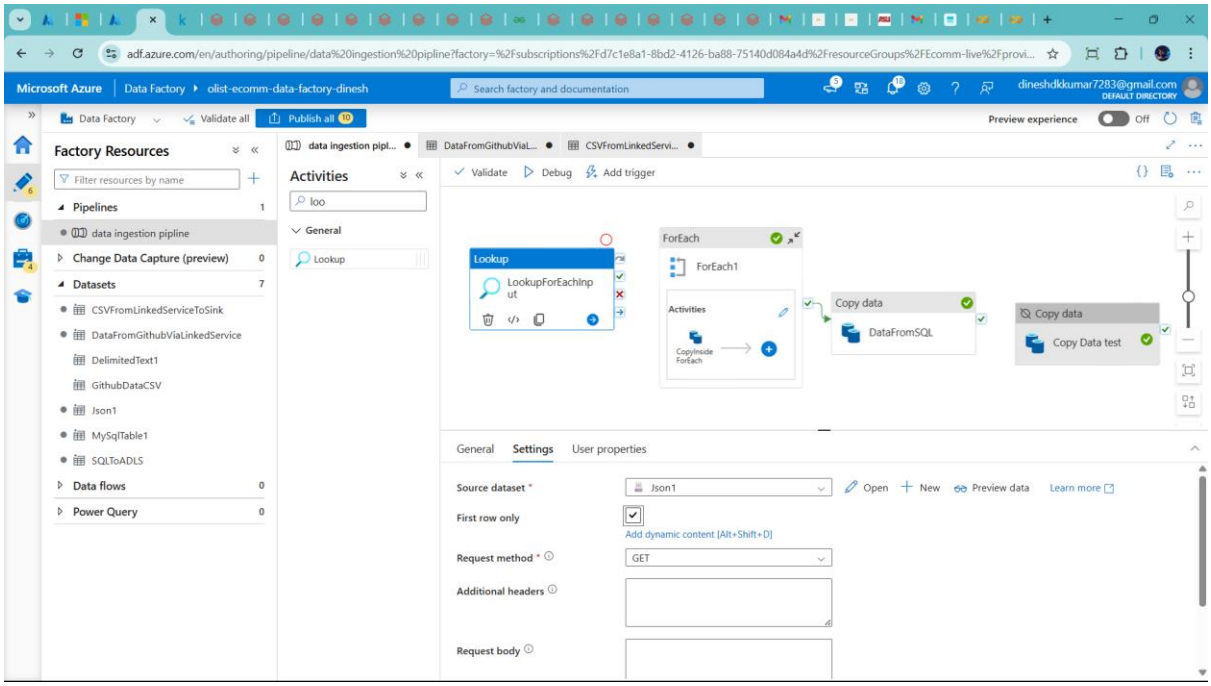

github.com/dinesh6351/Databricks_Ecomm_Krish_Capstone/blob/main/Project-Brazillian%20Ecommerce/ForEachInput.json
Files
main
Go to file
olist_order_items_dataset.csv
olist_order_payments_dataset....
olist_order_reviews_dataset.csv
olist_orders_dataset.csv
olist_products_dataset.csv
olist_sellers_dataset.csv
product_category_name_transl...
.DS_Store
Architecture Diagram .excalidraw
Architecture Diagram.png
DataIngestionToDB.ipynb
Databricks Awesome Code for T...
ForEachInput.json
Hand Written Notes - Otis Data...
Interview Questions 1.md
Interview Questions 2.md
notes.txt
sql code
README.md
Databricks_Ecomm_Krish_Capstone / Project-Brazillian Ecommerce / ForEachInput.json
dines7283 first commit
48fa713 · yesterday
History
Code
Blame
30 lines (30 loc) · 1.17 KB
Code 55% faster with GitHub Copilot
Raw
[
{
"csv_relative_url": "BigDataProjects/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_customers_dataset.csv",
"file_name":"olist_customers_dataset.csv"
},
{
"csv_relative_url": "BigDataProjects/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_geolocation_dataset.csv",
"file_name":"olist_geolocation_dataset.csv"
},
{
"csv_relative_url": "BigDataProjects/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_order_items_dataset.csv",
"file_name":"olist_order_items_dataset.csv"
},
{
"csv_relative_url": "BigDataProjects/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_order_reviews_dataset.csv",
"file_name":"olist_order_reviews_dataset.csv"
},
{
"csv_relative_url": "BigDataProjects/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_orders_dataset.csv",
"file_name":"olist_orders_dataset.csv"
},
{
"csv_relative_url": "BigDataProjects/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_products_dataset.csv",
"file_name":"olist_products_dataset.csv"
},
{
"csv_relative_url": "BigDataProjects/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_sellers_dataset.csv",
"file_name":"olist_sellers_dataset.csv"
}
]


https://raw.githubusercontent.com/dinesh6351/Databricks_Ecomm_Krish_Capstone/refs/heads/main/Project-Brazillian%20Ecommerce/Fc
[
{
"csv_relative_url": "BigDataProjects/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_customers_dataset.csv",
"file_name":"olist_customers_dataset.csv"
},
{
"csv_relative_url": "BigDataProjects/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_geolocation_dataset.csv",
"file_name":"olist_geolocation_dataset.csv"
},
{
"csv_relative_url": "BigDataProjects/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_order_items_dataset.csv",
"file_name":"olist_order_items_dataset.csv"
},
{
"csv_relative_url": "BigDataProjects/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_order_reviews_dataset.csv",
"file_name":"olist_order_reviews_dataset.csv"
},
{
"csv_relative_url": "BigDataProjects/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_orders_dataset.csv",
"file_name":"olist_orders_dataset.csv"
},
{
"csv_relative_url": "BigDataProjects/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_products_dataset.csv",
"file_name":"olist_products_dataset.csv"
},
{
"csv_relative_url": "BigDataProjects/refs/heads/main/Project-Brazillian%20Ecommerce/Data/olist_sellers_dataset.csv",
"file_name":"olist_sellers_dataset.csv"
}
]



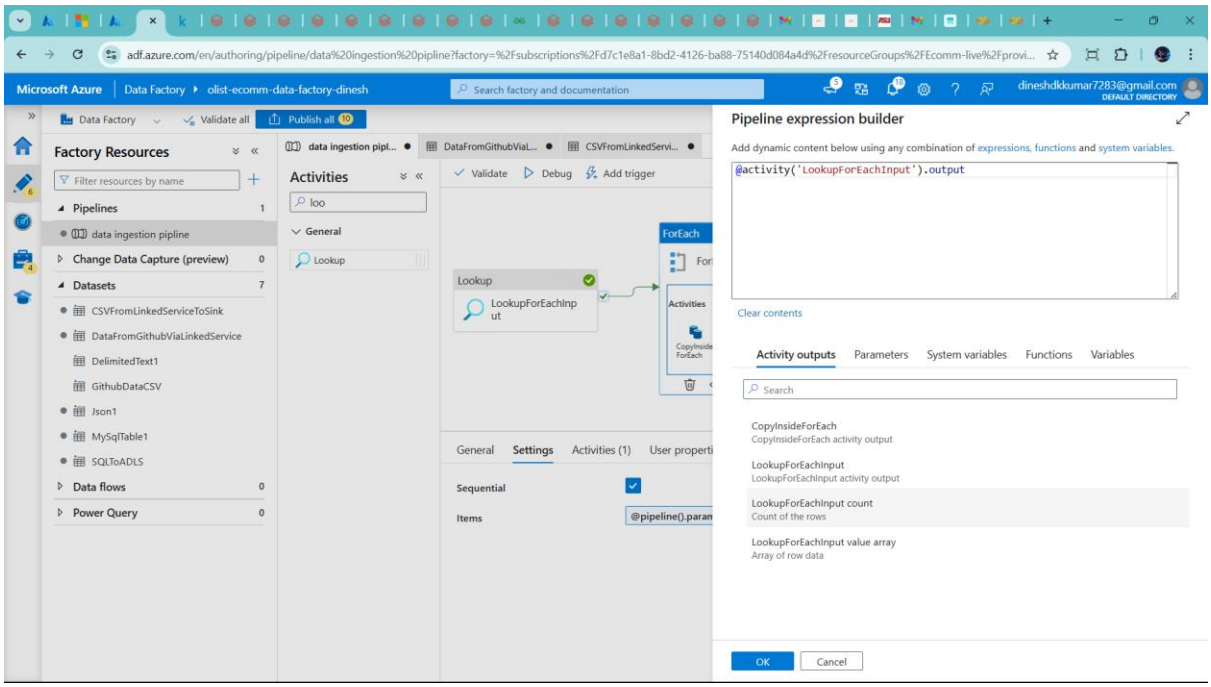


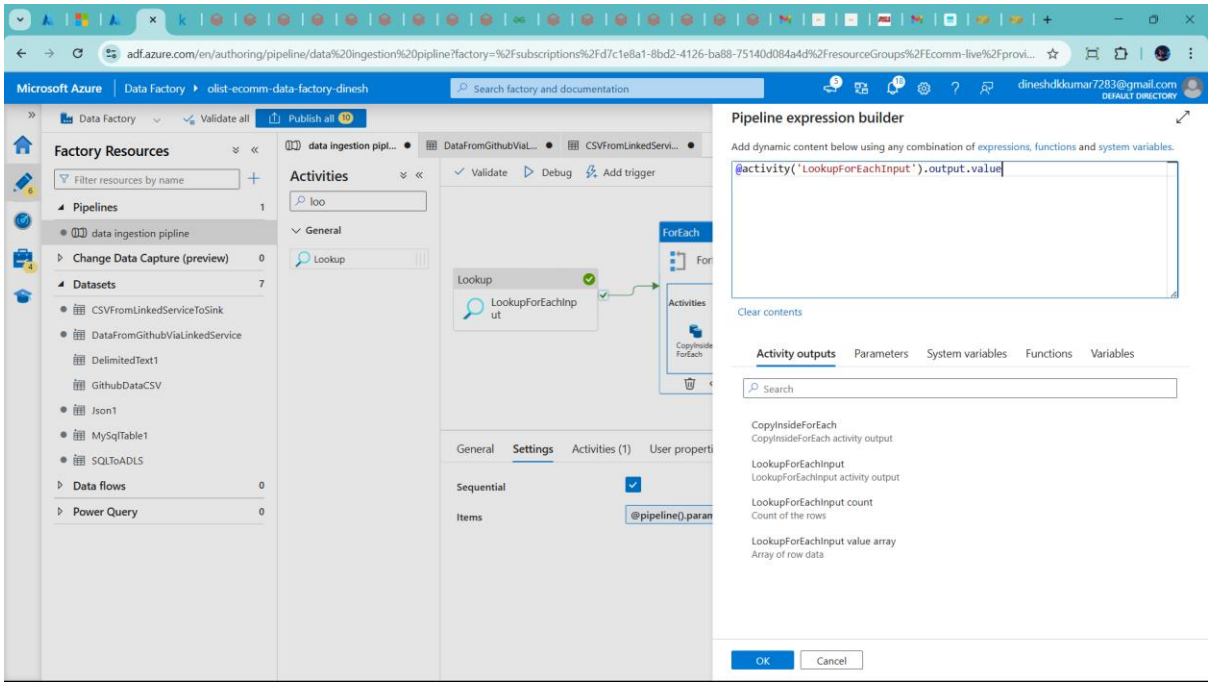
→uncheck first row only

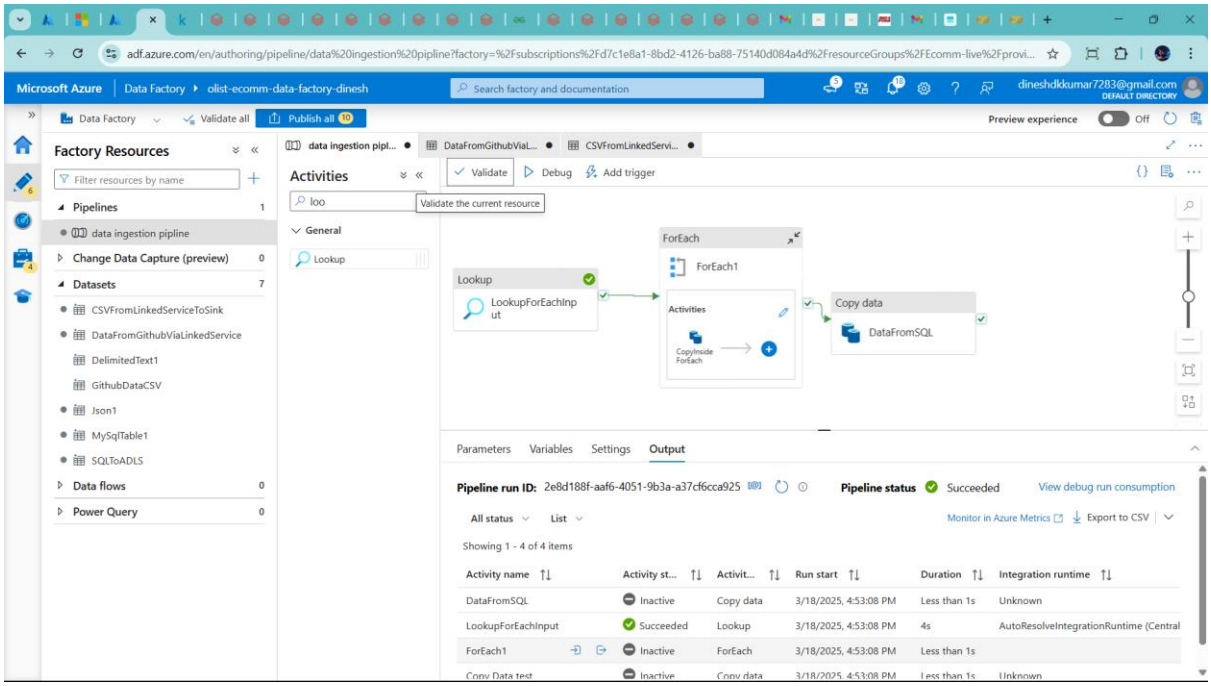


Change foreach as well

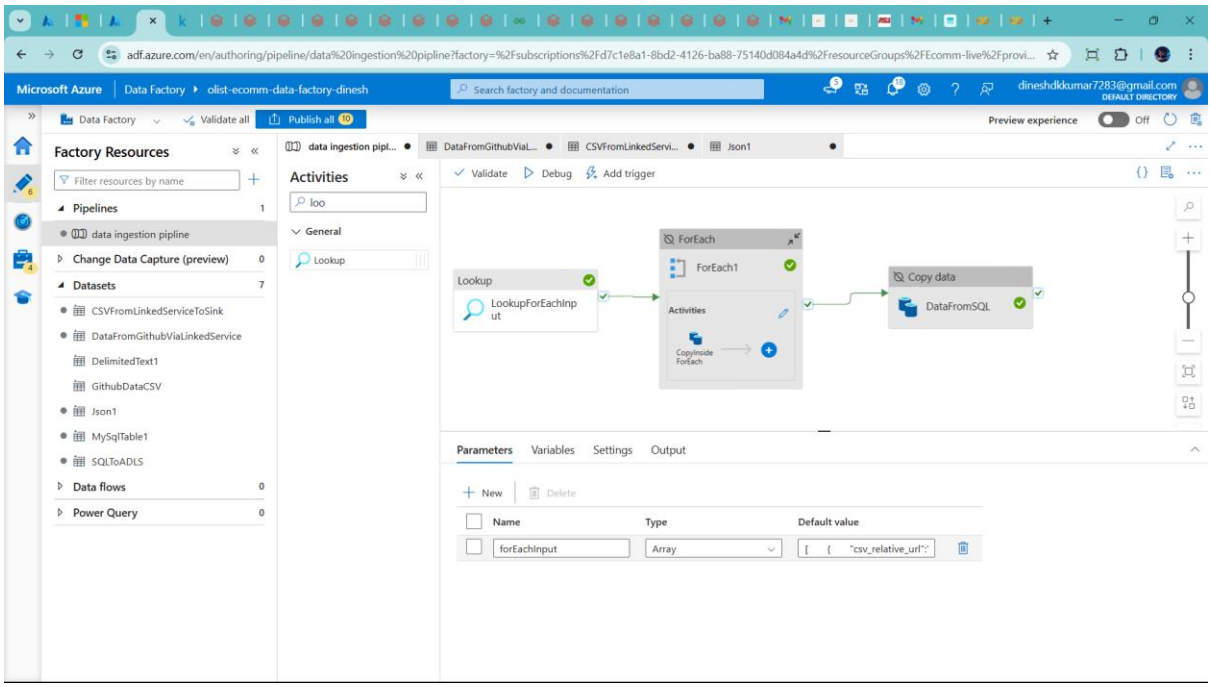
→click setting



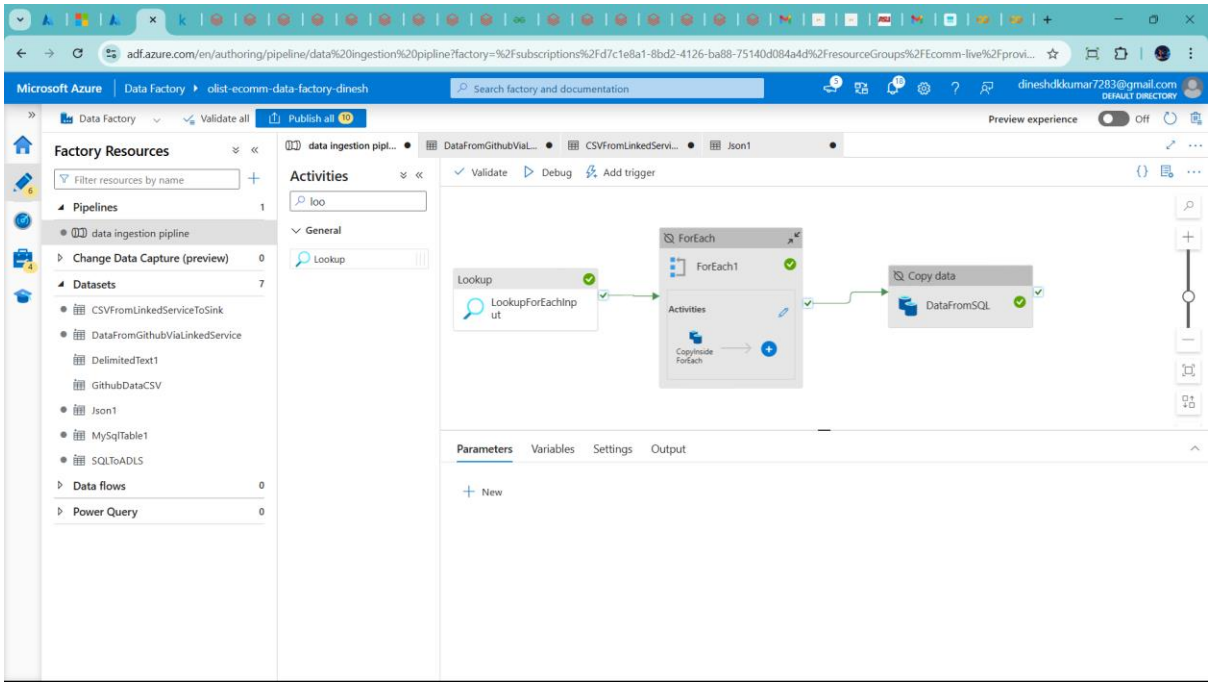

Pipeline expression builder
Add dynamic content below using any combination of expressions, functions and system variables.
@activity('LookupForEachInput').output.value
Clear contents
Activity outputs
Parameters
System variables
Functions
Variables
CopyInsideForEach
CopyInsideForEach activity output
LookupForEachInput
LookupForEachInput activity output
LookupForEachInput count
Count of the rows
LookupForEachInput value array
Array of row data
OK
Cancel

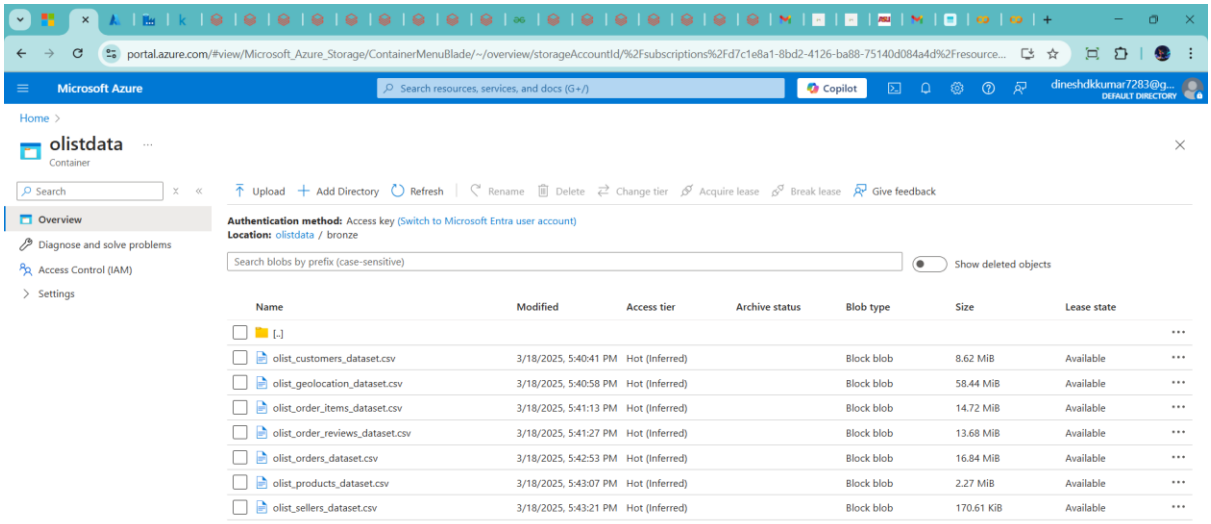

Pipeline run ID: 2e8d188f-aaf6-4051-9b3a-a37cf6cca925
Pipeline status Succeeded
View debug run consumption
Parameters
Variables
Settings
Output
Showing 1 - 4 of 4 items
DataFromSQL
LookupForEachInput
ForEach1
Copy Data test

→Now Debug

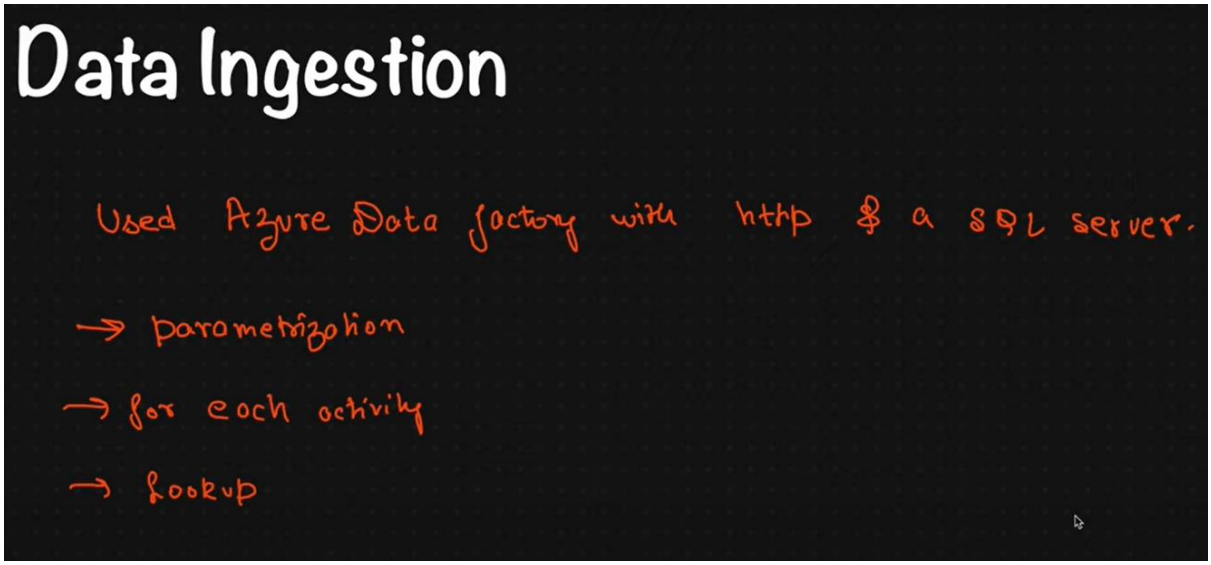

Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
Validate all
Publish all
Preview experience
Factory Resources
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
Json1
MySqlTable1
SQLToADLS
Data flows
Power Query
Activities
General
Lookup
Validate
Debug
Add trigger
ForEach
ForEach1
Activities
Copy data
DataFromSQL
Parameters
Variables
Settings
Output
New
Delete
Name
Type
Default value
forEachInput
Array

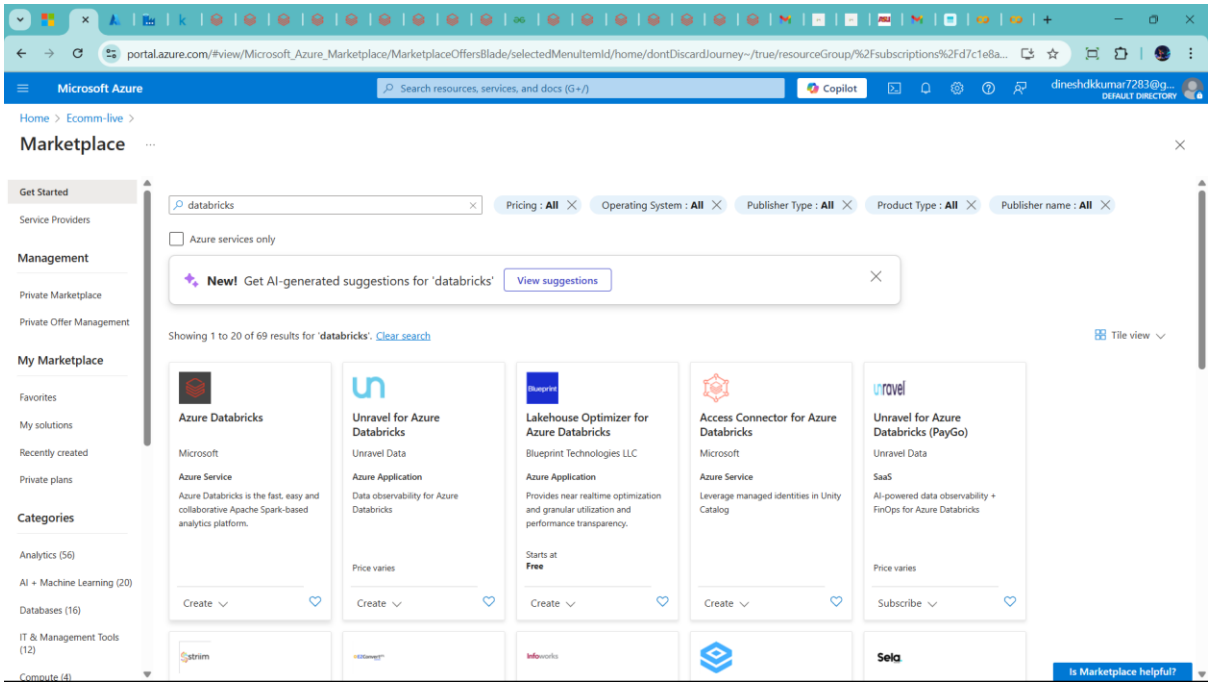
→Remove


Microsoft Azure
Data Factory
olist-ecomm-data-factory-dinesh
Search factory and documentation
Validate all
Publish all
Preview experience
Factory Resources
Pipelines
data ingestion pipline
Change Data Capture (preview)
Datasets
CSVFromLinkedServiceToSink
DataFromGithubViaLinkedService
DelimitedText1
GithubDataCSV
Json1
MySqlTable1
SQLToADLS
Data flows
Power Query
Activities
General
Lookup
Validate
Debug
Add trigger
ForEach
ForEach1
Activities
Copy data
DataFromSQL
Parameters
Variables
Settings
Output
New

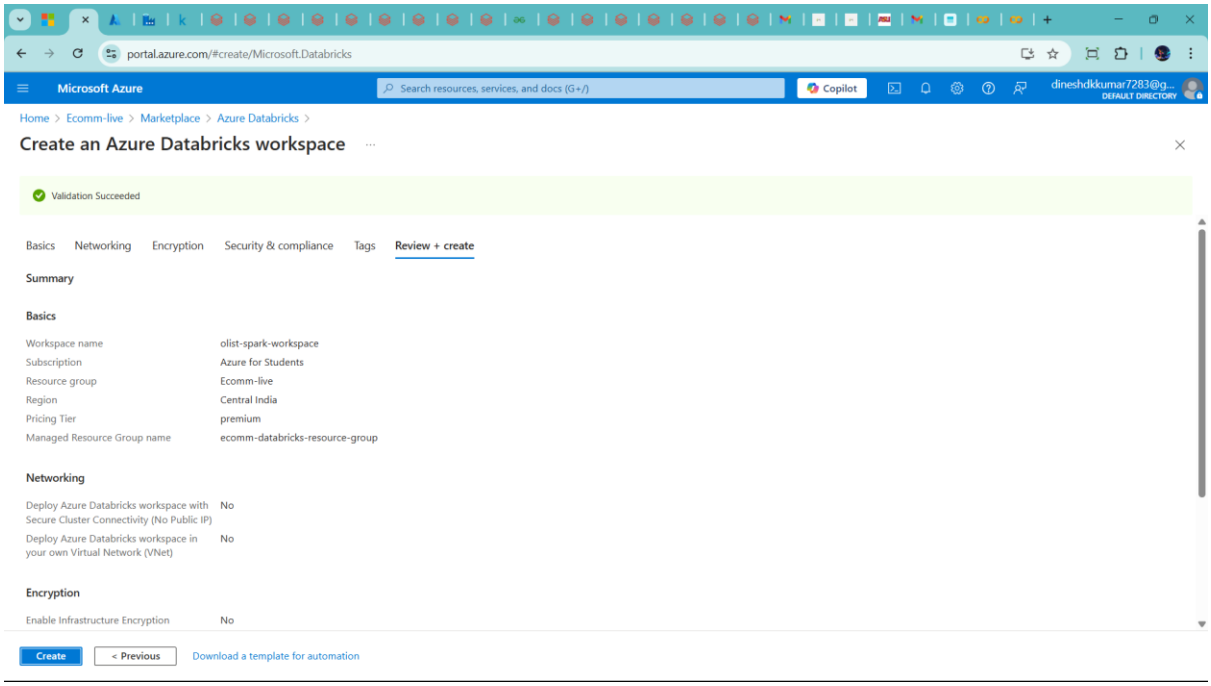

portal.azure.com/#view/Microsoft_Azure_Storage/ContainerMenuBlade/~/overview/storageAccountId/%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2Fresource...
Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
dineshdkkumar7283@g...
DEFAULT DIRECTORY
Home
olistdata
Container
Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: olistdata / bronze
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
Size
Lease state
[..]
olist_customers_dataset.csv 3/18/2025, 5:40:41 PM Hot (Inferred) Block blob 8.62 MiB Available
olist_geolocation_dataset.csv 3/18/2025, 5:40:58 PM Hot (Inferred) Block blob 58.44 MiB Available
olist_order_items_dataset.csv 3/18/2025, 5:41:13 PM Hot (Inferred) Block blob 14.72 MiB Available
olist_order_reviews_dataset.csv 3/18/2025, 5:41:27 PM Hot (Inferred) Block blob 13.68 MiB Available
olist_orders_dataset.csv 3/18/2025, 5:42:53 PM Hot (Inferred) Block blob 16.84 MiB Available
olist_products_dataset.csv 3/18/2025, 5:43:07 PM Hot (Inferred) Block blob 2.27 MiB Available
olist_sellers_dataset.csv 3/18/2025, 5:43:21 PM Hot (Inferred) Block blob 170.61 KiB Available

---


Data Ingestion
Used Azure Data factory with http & a SQL server.
→ parametrization
→ for each activity
→ lookup



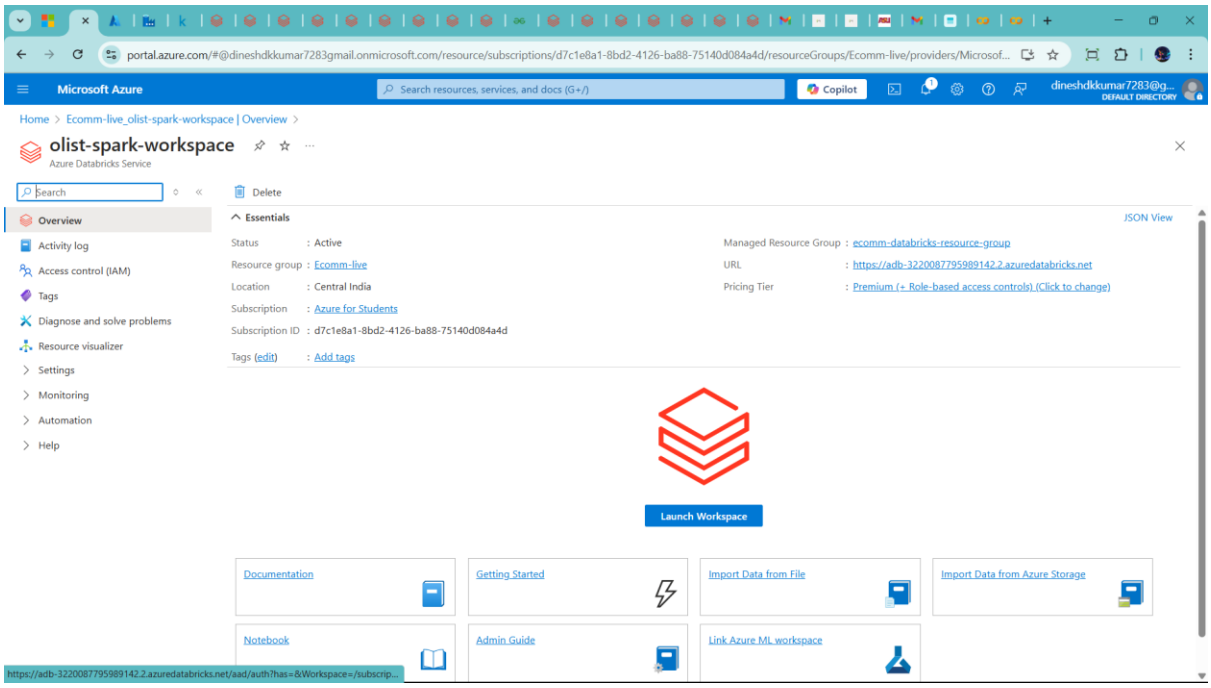
→I choose premium



Azure Databricks

Azure Databricks

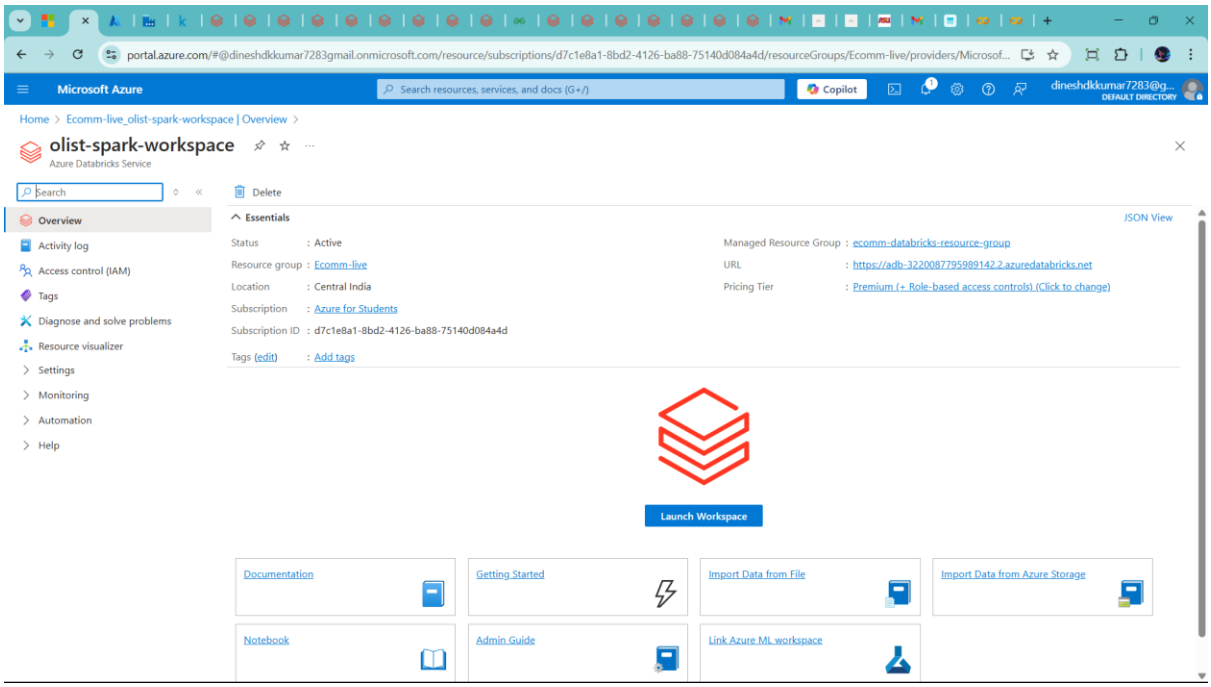
→ Spark powered
→ Integrated with azure
→ Handles big Data easily
→ Great for machine learning



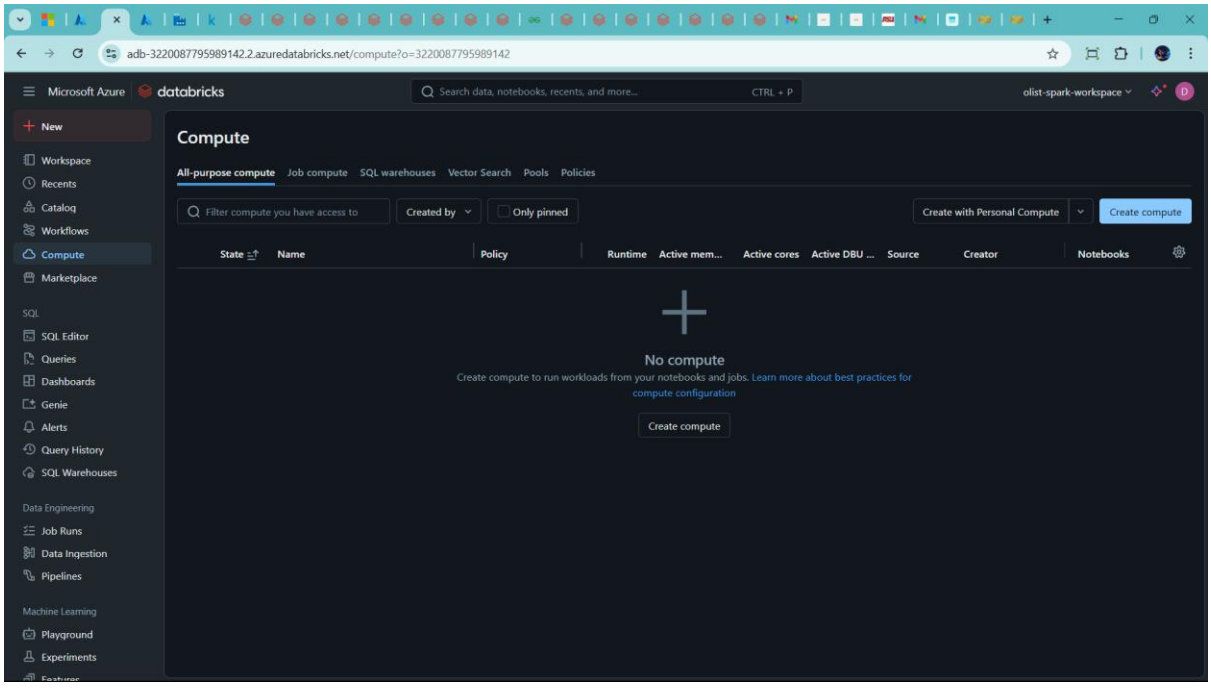
Azure Databricks Workflow

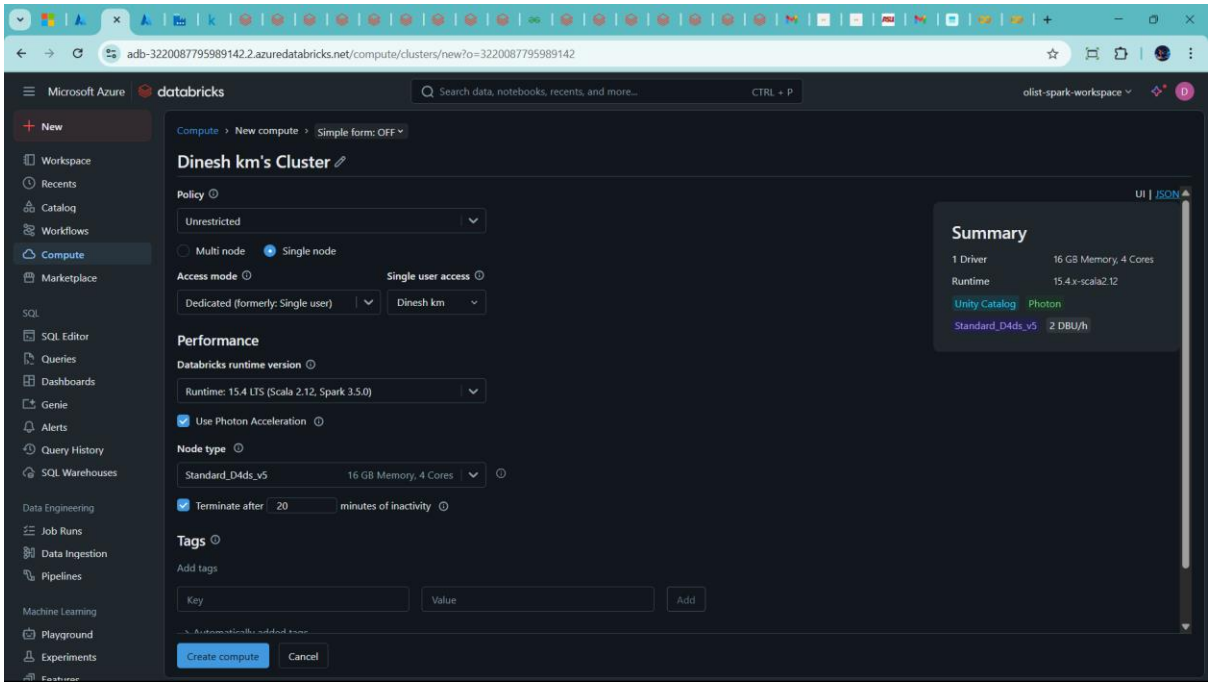
1. Read data from ADLS gen 2
2. Perform basic transformation like cleaning, renaming & filtering
3. Use join operation to integrate multiple dataset.
4. Enrich the data via mongoDB.
5. Perform aggregation & derive insights
6. Write final data back to ADLS gen 2.

→Launch Databricks workspace

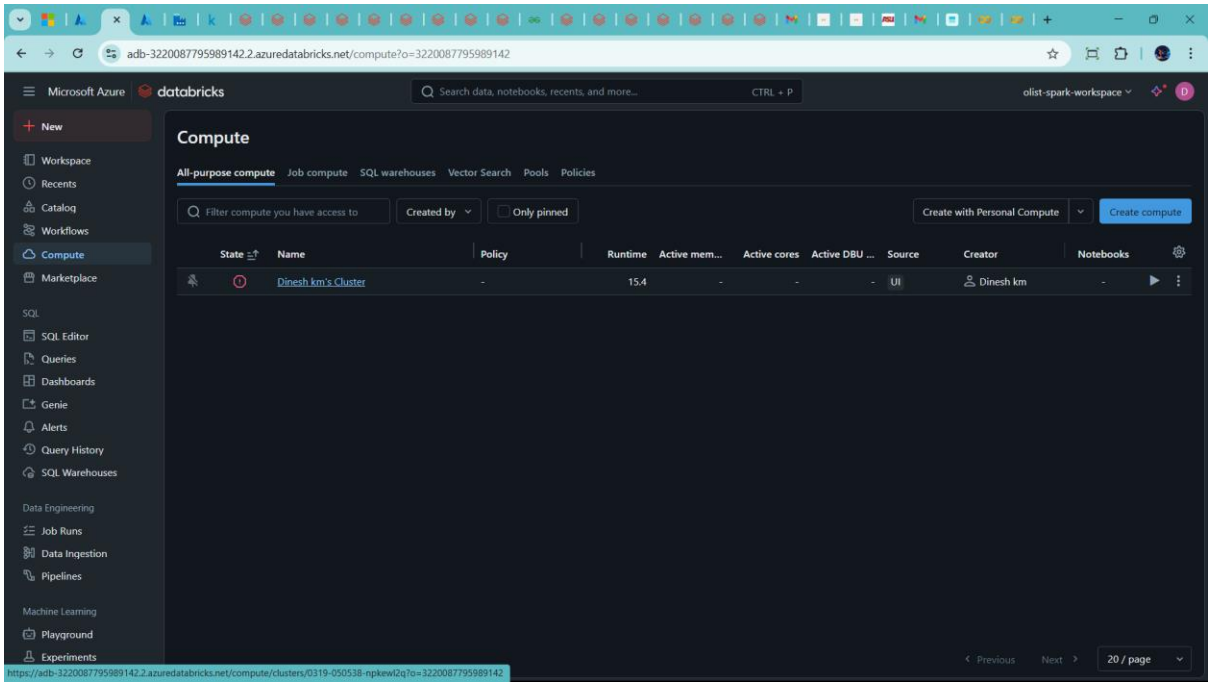


→Create comupte →it is like hardware



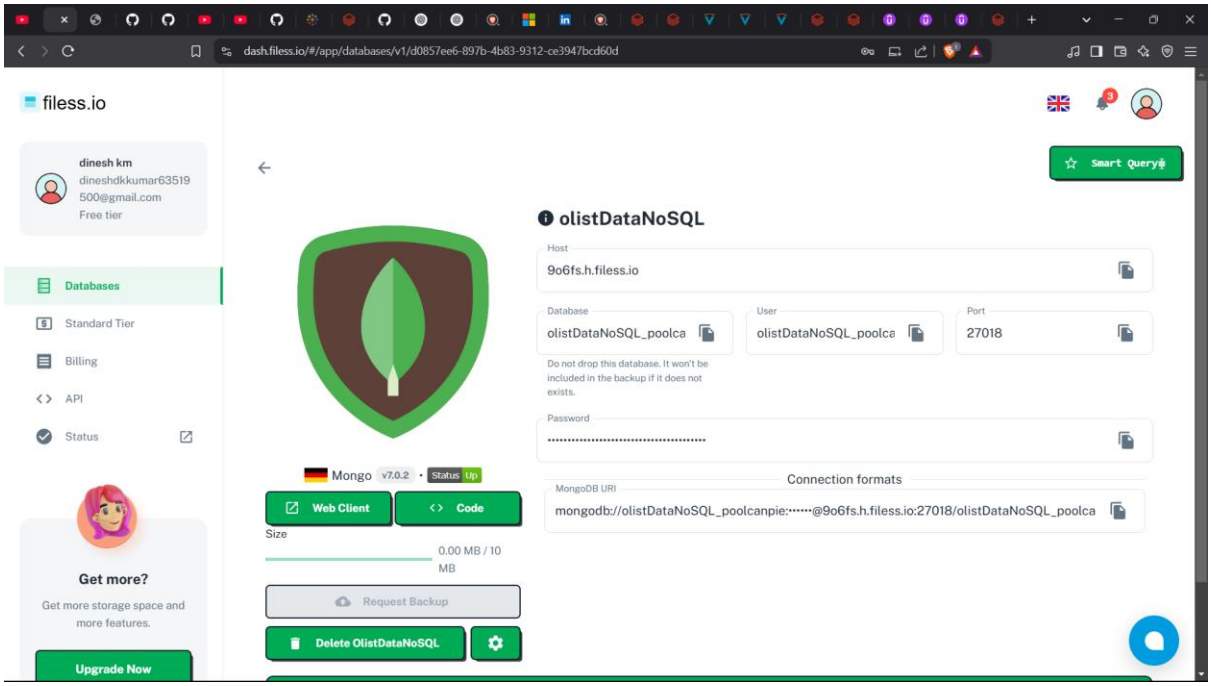


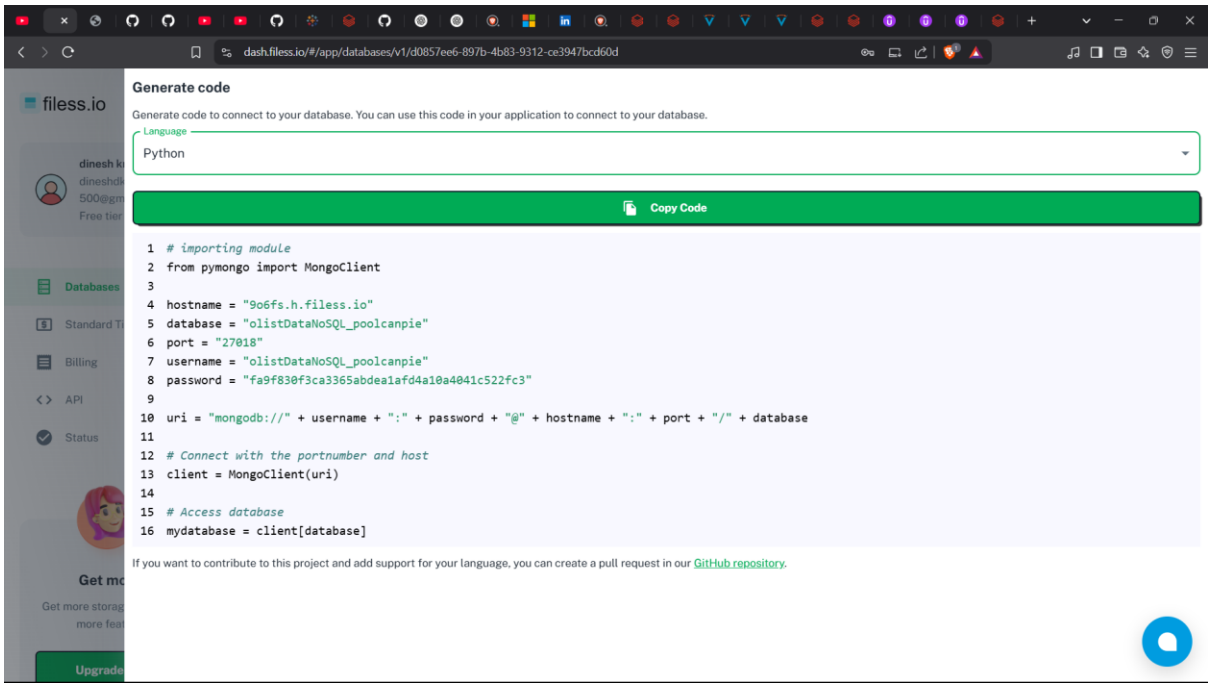
→click compute



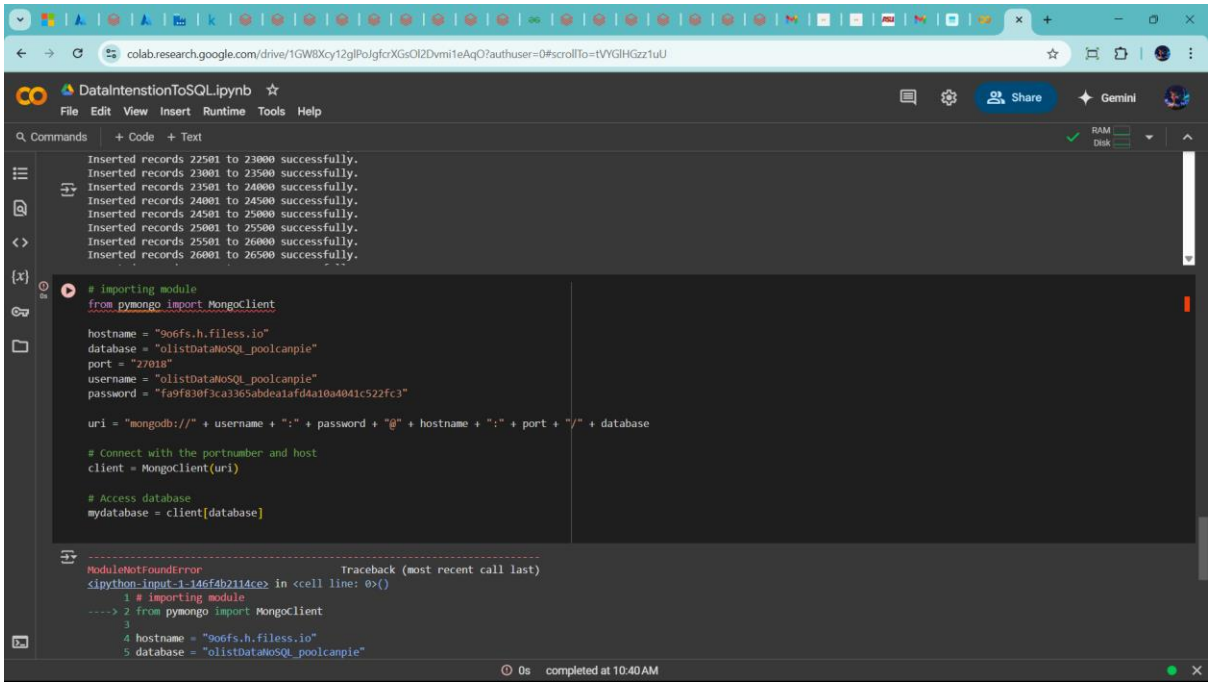
→Create to add data in MangoDB

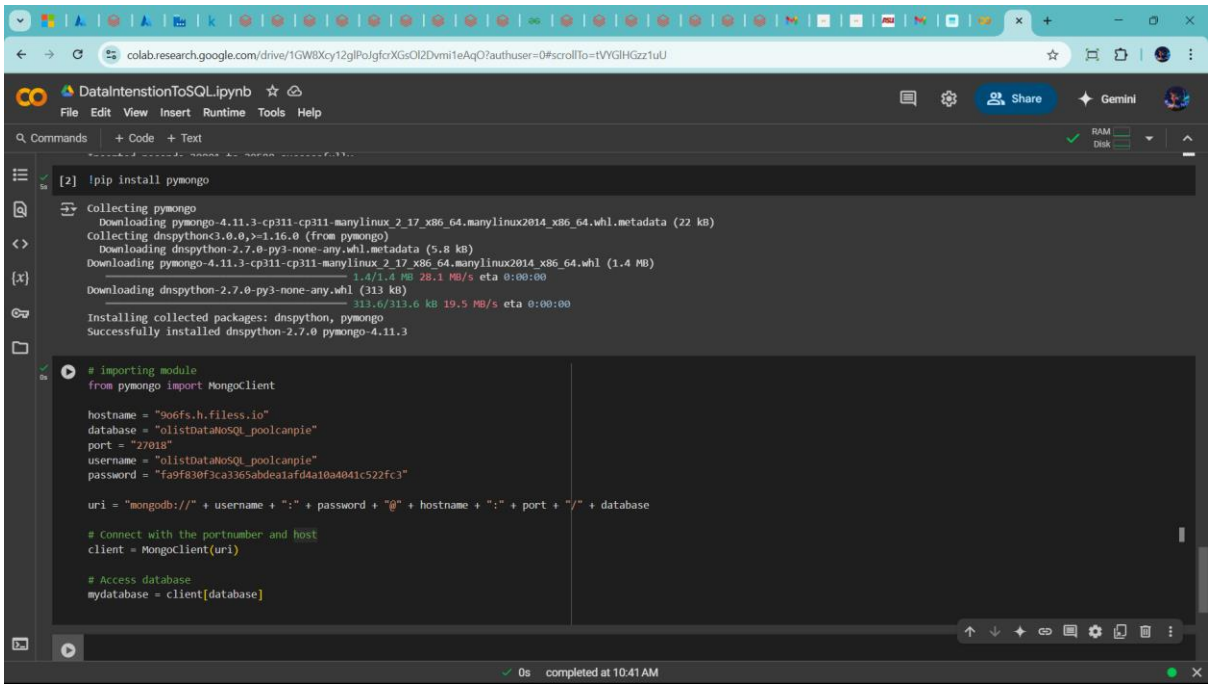
→Go to file.io



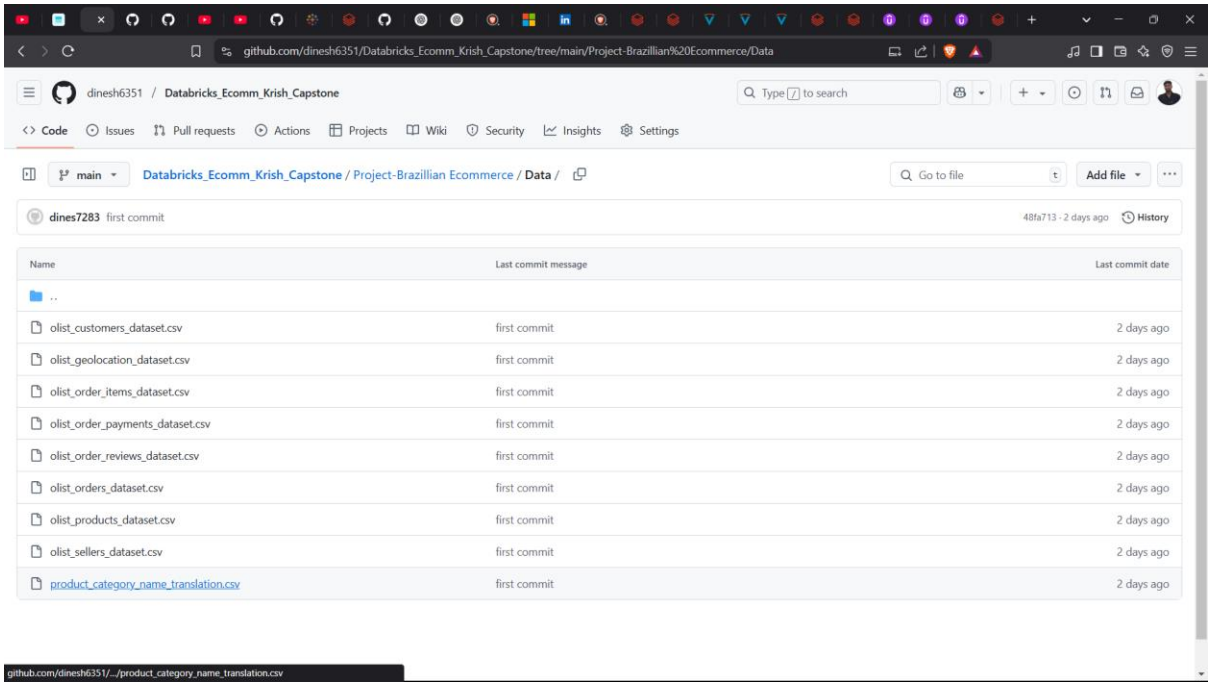


→Copy the code and add in googleColab

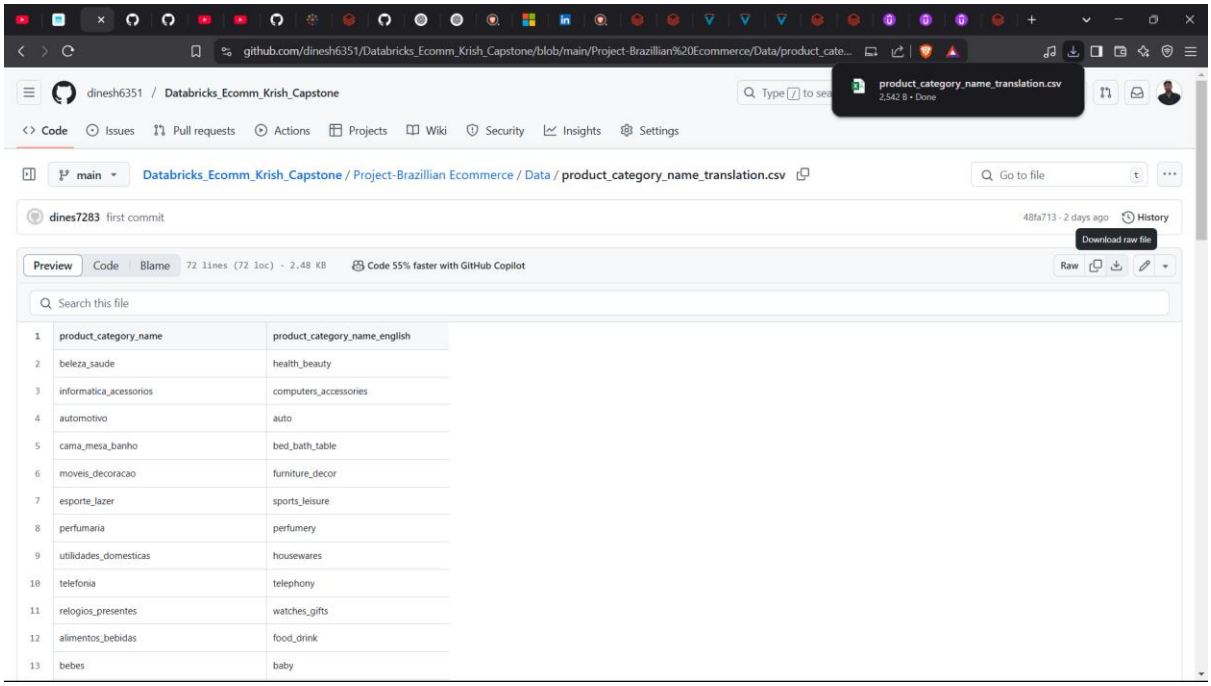




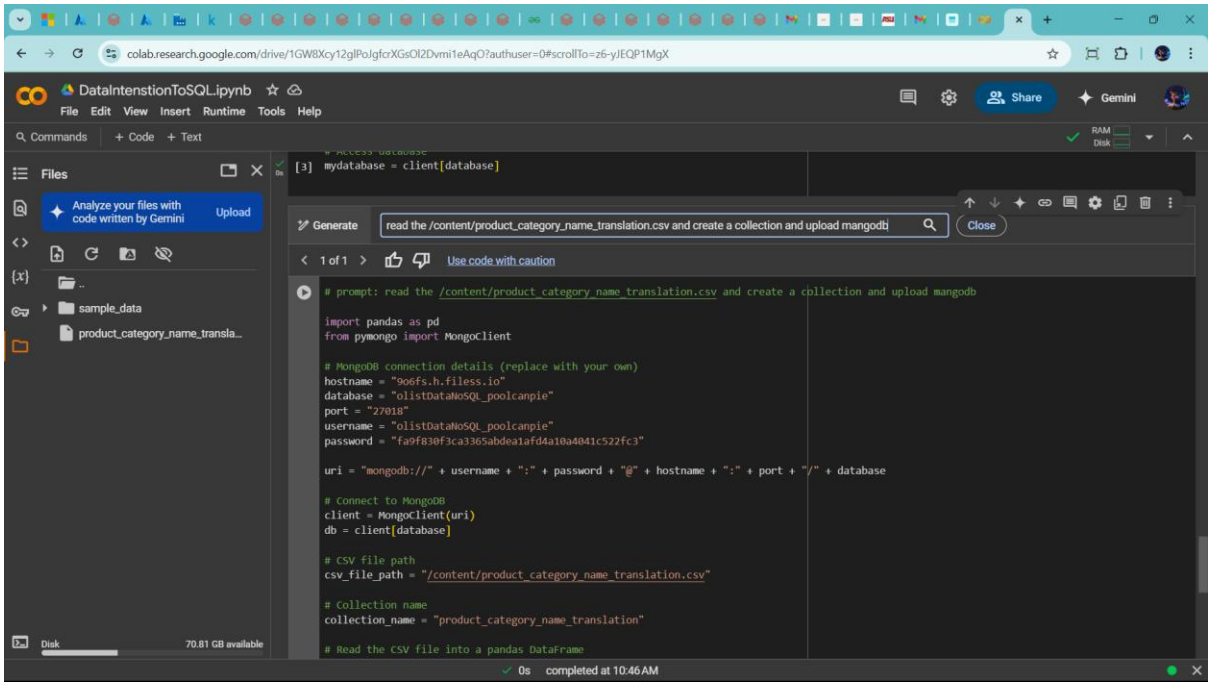
→add file data →go to github

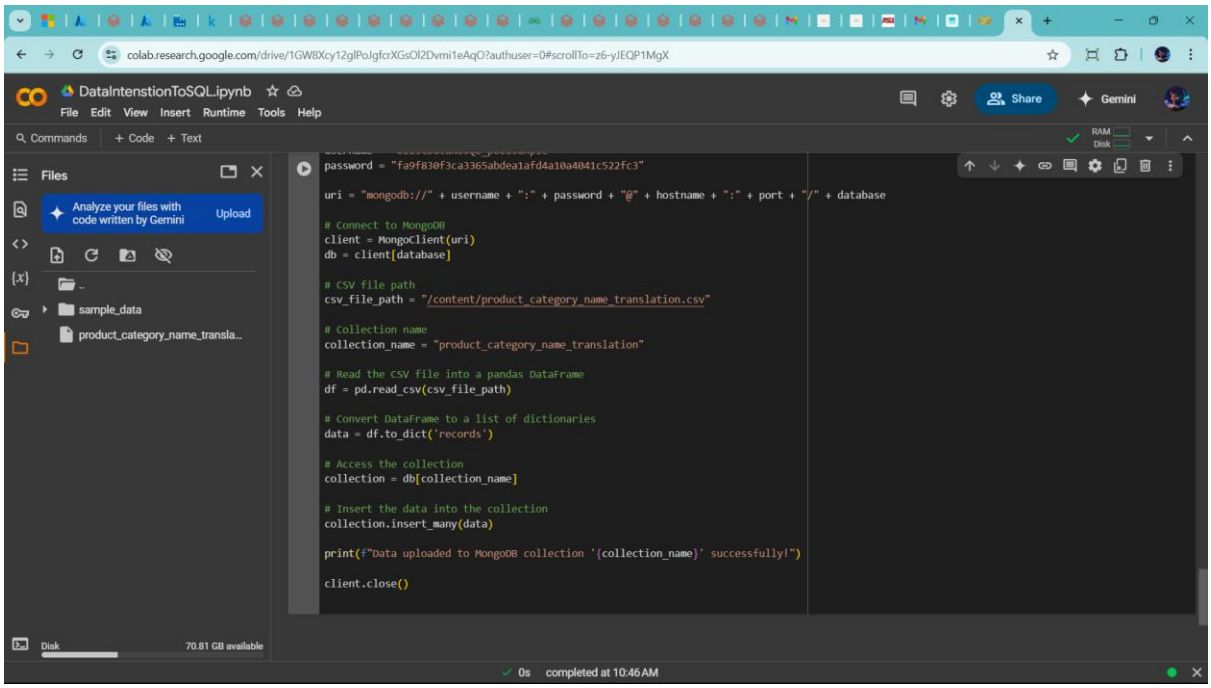


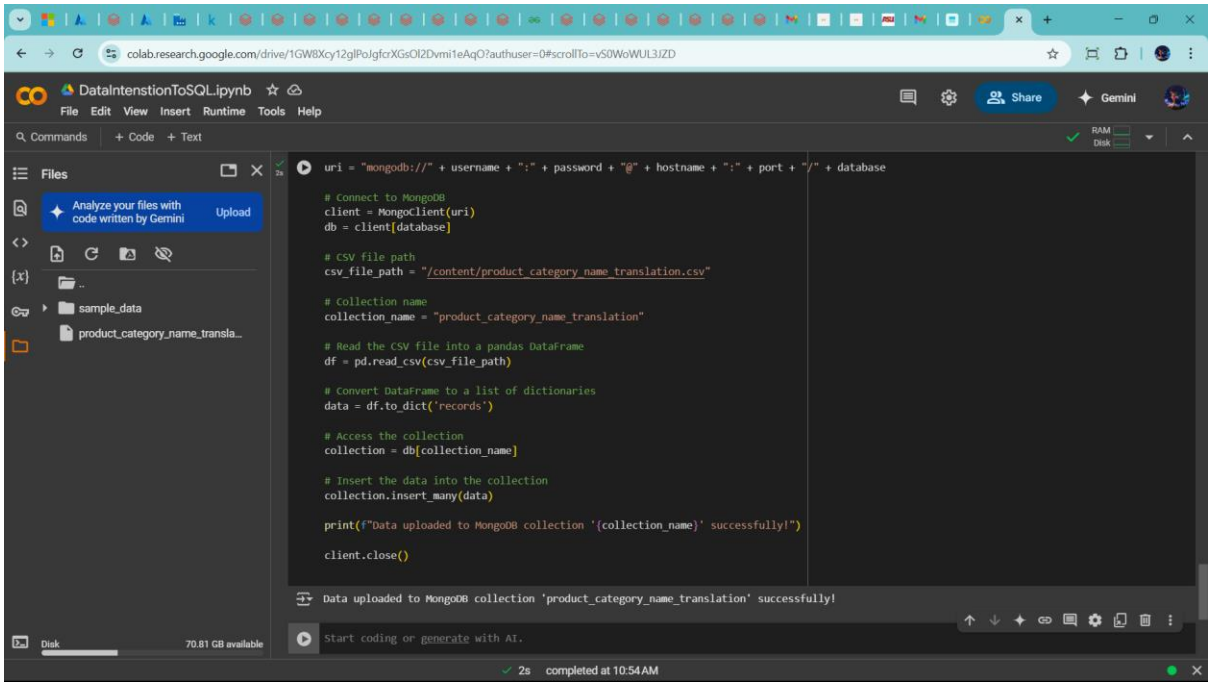
→download it

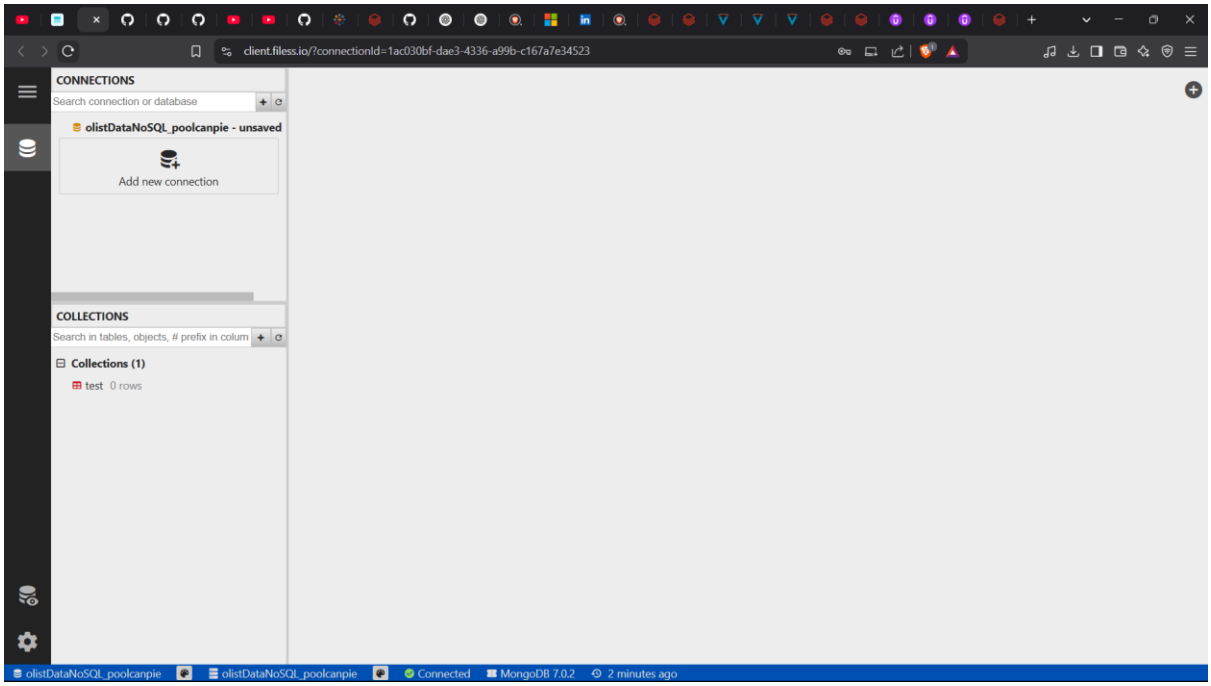


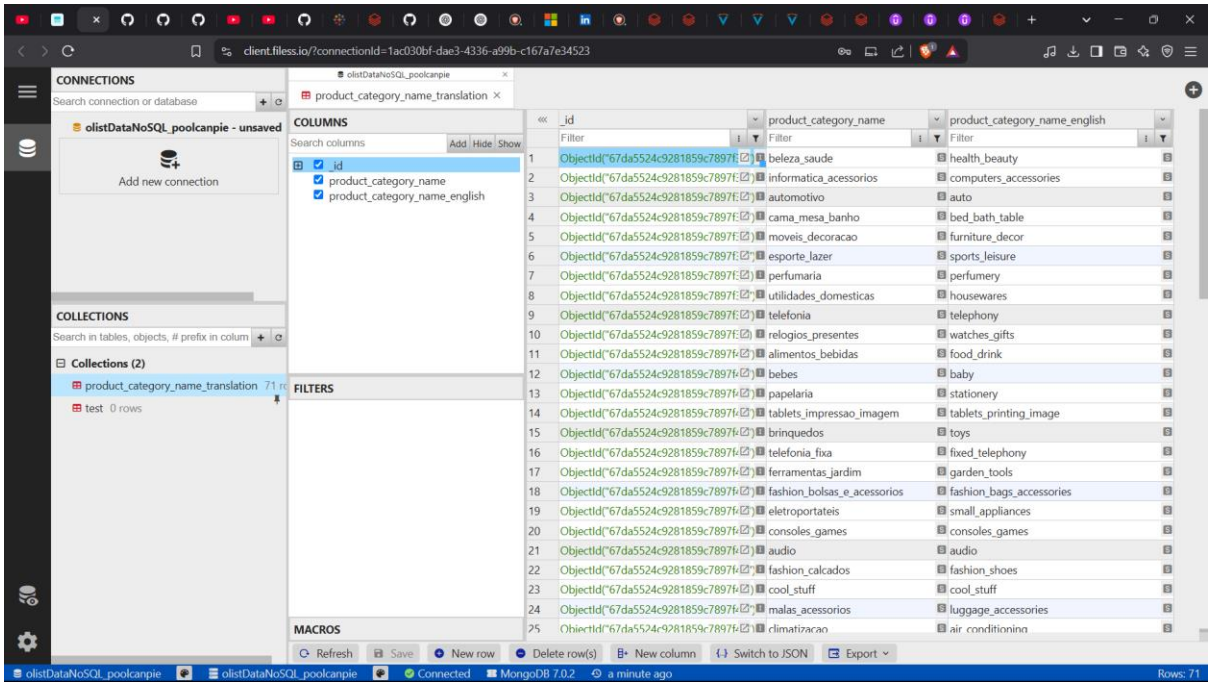
→add in google colab


colab.research.google.com/drive/1GW8Xcy12glPoJgfcrXGsOl2Dvmi1eAqO?authuser=0#scrollTo=z6-yJEQP1MgX
DataIntenstionToSQL.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands
+ Code
+ Text
Files
Analyze your files with code written by Gemini
Upload
sample_data
product_category_name_transla...
Disk
70.81 GB available
[3] mydatabase = client[database]
Generate
read the /content/product_category_name_translation.csv and create a collection and upload mangodb
Close
1 of 1
Use code with caution
# prompt: read the /content/product_category_name_translation.csv and create a collection and upload mangodb
import pandas as pd
from pymongo import MongoClient
# MongoDB connection details (replace with your own)
hostname = "9o6fs.h.filess.io"
database = "olistDataNoSQL_poolcanpie"
port = "27018"
username = "olistDataNoSQL_poolcanpie"
password = "fa9f830f3ca3365abdea1afd4a10a4041c522fc3"
uri = "mongodb://" + username + ":" + password + "@" + hostname + ":" + port + "/" + database
# Connect to MongoDB
client = MongoClient(uri)
db = client[database]
# CSV file path
csv_file_path = "/content/product_category_name_translation.csv"
# Collection name
collection_name = "product_category_name_translation"
# Read the CSV file into a pandas DataFrame
0s completed at 10:46 AM

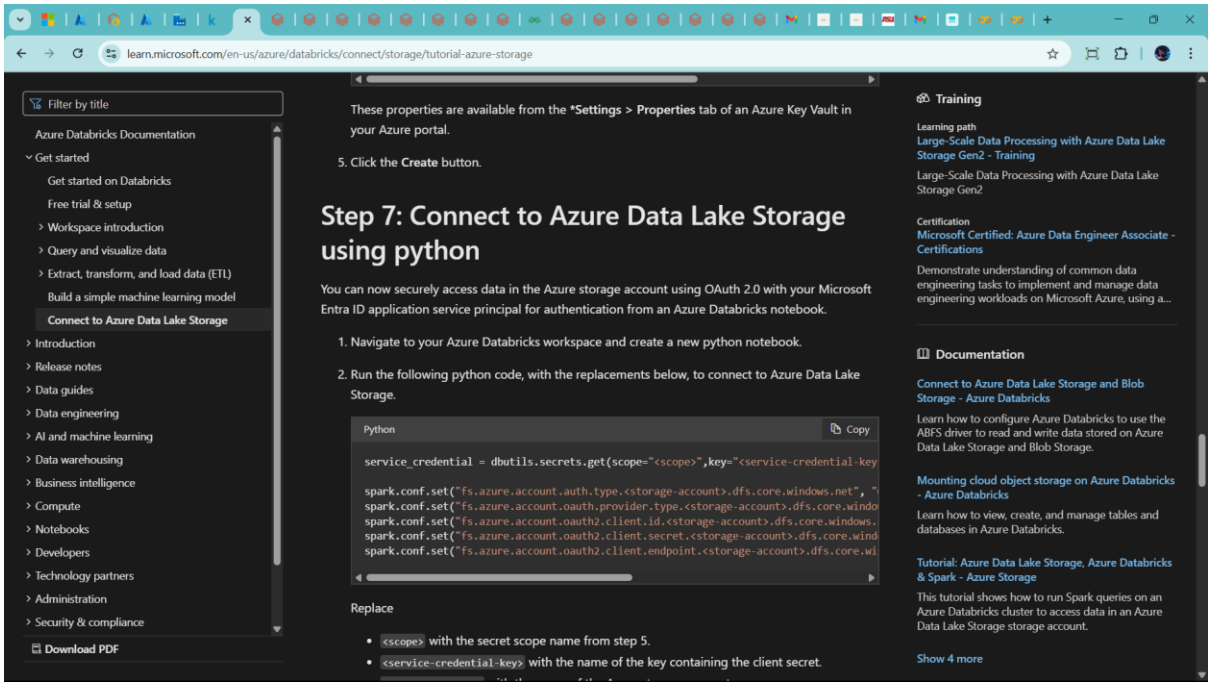

colab.research.google.com/drive/1GW8Xcy12glPoJgfcrXGsOl2Dvmi1eAqO?authuser=0#scrollTo=z6-yJEQP1MgX
DataIntenstionToSQL.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands
+ Code
+ Text
Files
Analyze your files with code written by Gemini
Upload
sample_data
product_category_name_transla...
Disk
70.81 GB available
password = "fa9f830f3ca3365abdea1afd4a10a4041c522fc3"
uri = "mongodb://" + username + ":" + password + "@" + hostname + ":" + port + "/" + database
# Connect to MongoDB
client = MongoClient(uri)
db = client[database]
# CSV file path
csv_file_path = "/content/product_category_name_translation.csv"
# Collection name
collection_name = "product_category_name_translation"
# Read the CSV file into a pandas DataFrame
df = pd.read_csv(csv_file_path)
# Convert DataFrame to a list of dictionaries
data = df.to_dict('records')
# Access the collection
collection = db[collection_name]
# Insert the data into the collection
collection.insert_many(data)
print(f"Data uploaded to MongoDB collection '{collection_name}' successfully!")
client.close()
0s completed at 10:46 AM


colab.research.google.com/drive/1GW8Xcy12glPoJgfcrXGsOI2Dvmi1eAqO?authuser=0#scrollTo=vS0WoWUL3JZD
DataIntenstionToSQL.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands
+ Code + Text
Files
Analyze your files with code written by Gemini
Upload
sample_data
product_category_name_transla...
Disk
70.81 GB available
uri = "mongodb://" + username + ":" + password + "@" + hostname + ":" + port + "/" + database
# Connect to MongoDB
client = MongoClient(uri)
db = client[database]
# CSV file path
csv_file_path = "/content/product_category_name_translation.csv"
# Collection name
collection_name = "product_category_name_translation"
# Read the CSV file into a pandas DataFrame
df = pd.read_csv(csv_file_path)
# Convert DataFrame to a list of dictionaries
data = df.to_dict('records')
# Access the collection
collection = db[collection_name]
# Insert the data into the collection
collection.insert_many(data)
print(f"Data uploaded to MongoDB collection '{collection_name}' successfully!")
client.close()
Data uploaded to MongoDB collection 'product_category_name_translation' successfully!
Start coding or generate with AI.
2s completed at 10:54 AM

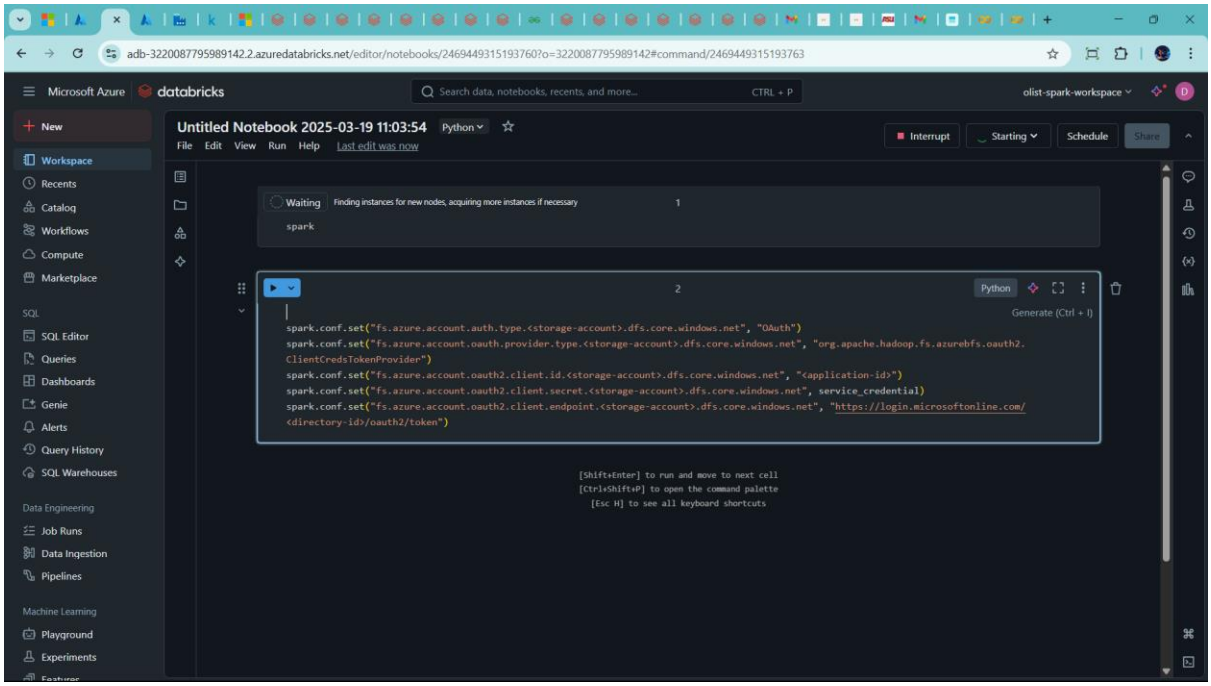

client.filess.io/?connectionId=1ac030bf-dae3-4336-a99b-c167a7e34523
CONNECTIONS
Search connection or database
olistDataNoSQL_poolcanpie - unsaved
Add new connection
COLLECTIONS
Search in tables, objects, # prefix in colum
Collections (1)
test 0 rows
olistDataNoSQL_poolcanpie
olistDataNoSQL_poolcanpie
Connected
MongoDB 7.0.2
2 minutes ago

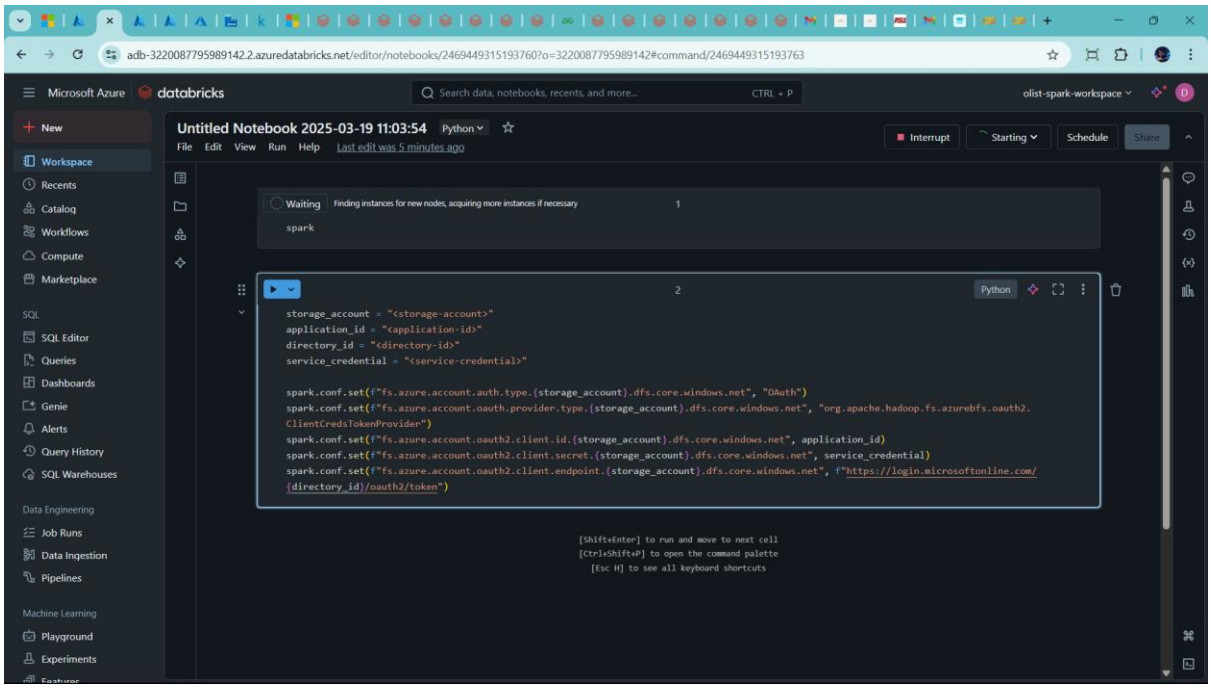


→connect ADLS gen 2 to databricks



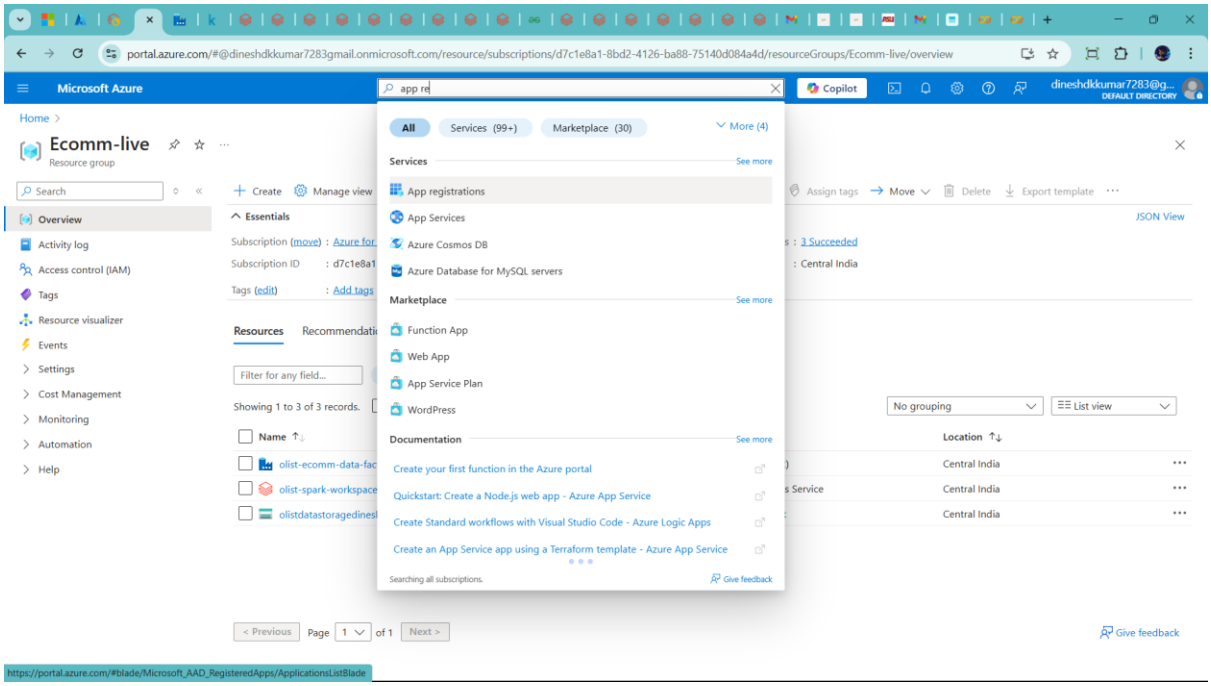
→copy and add in datariicks



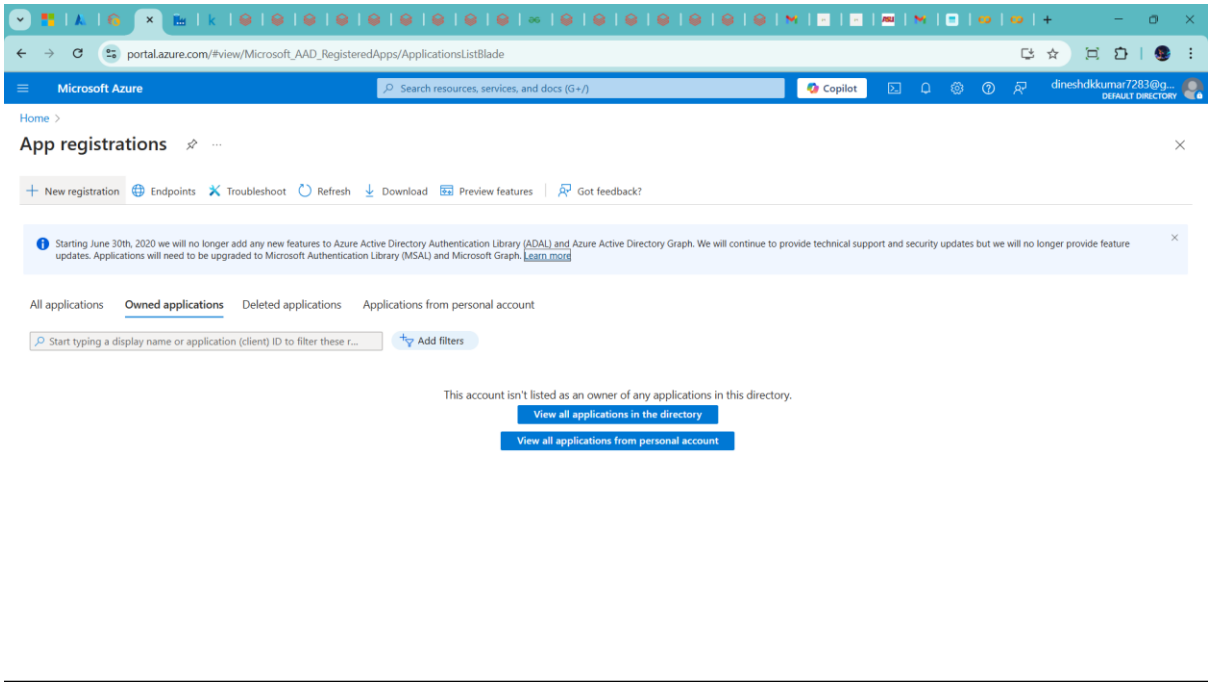


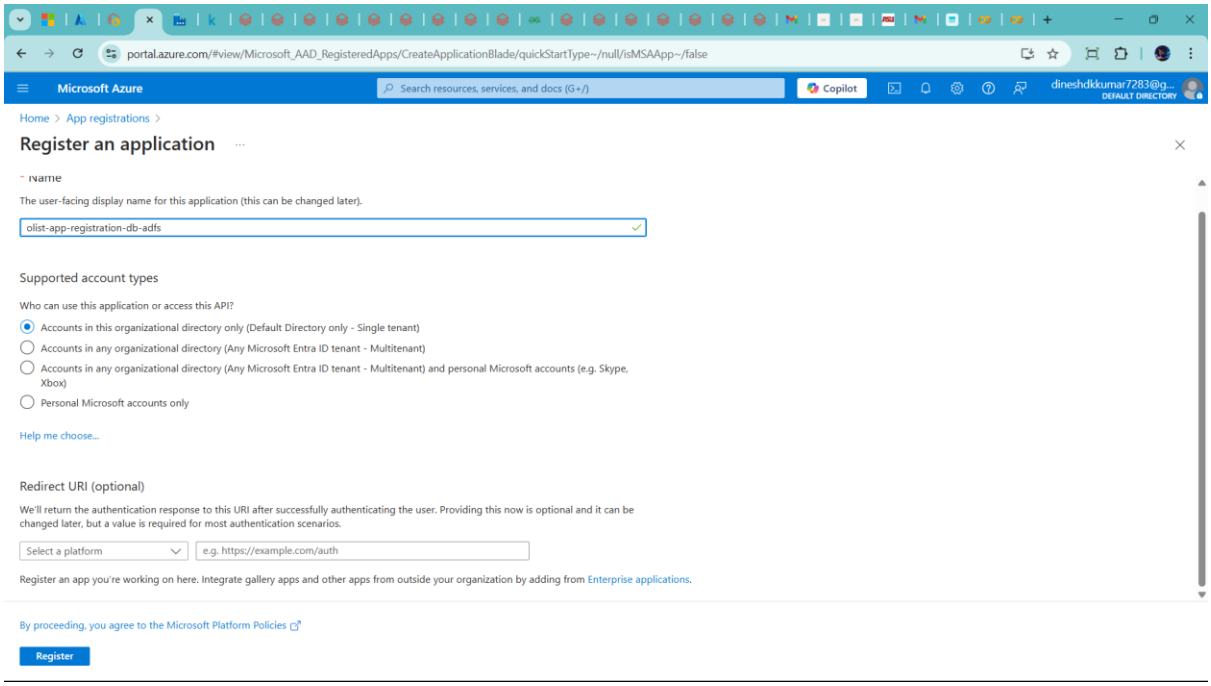
-->change bases on Ai

→search app resgisteration in azure

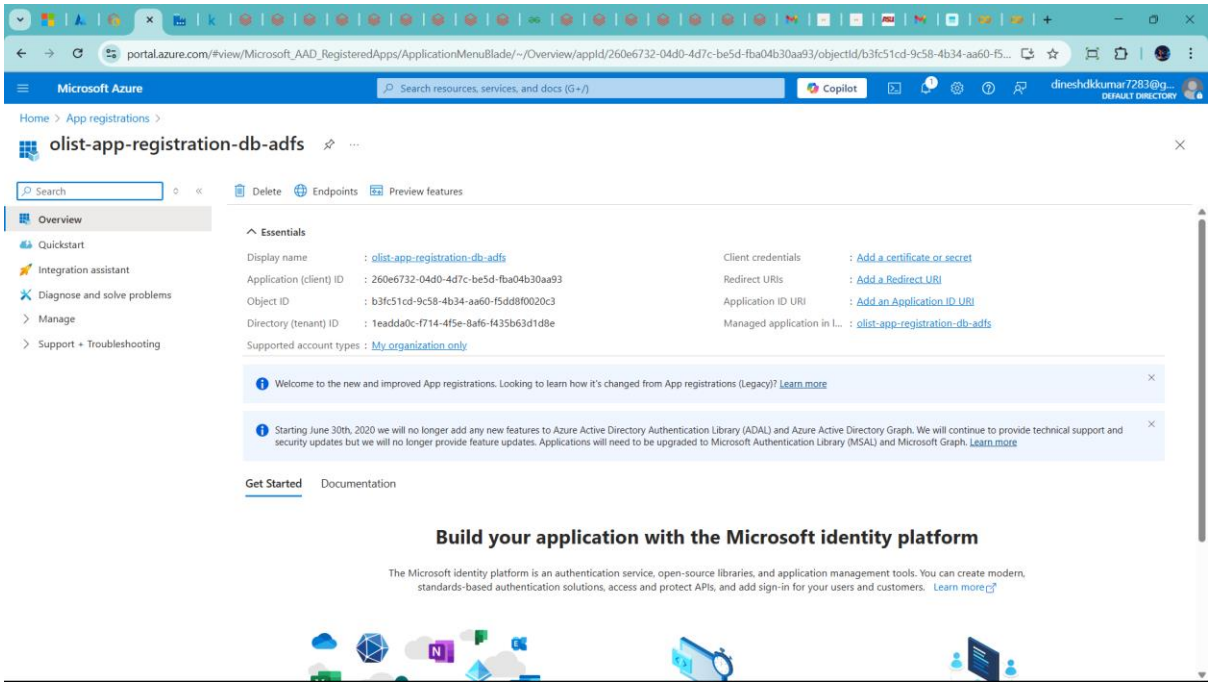


→Click new registeration

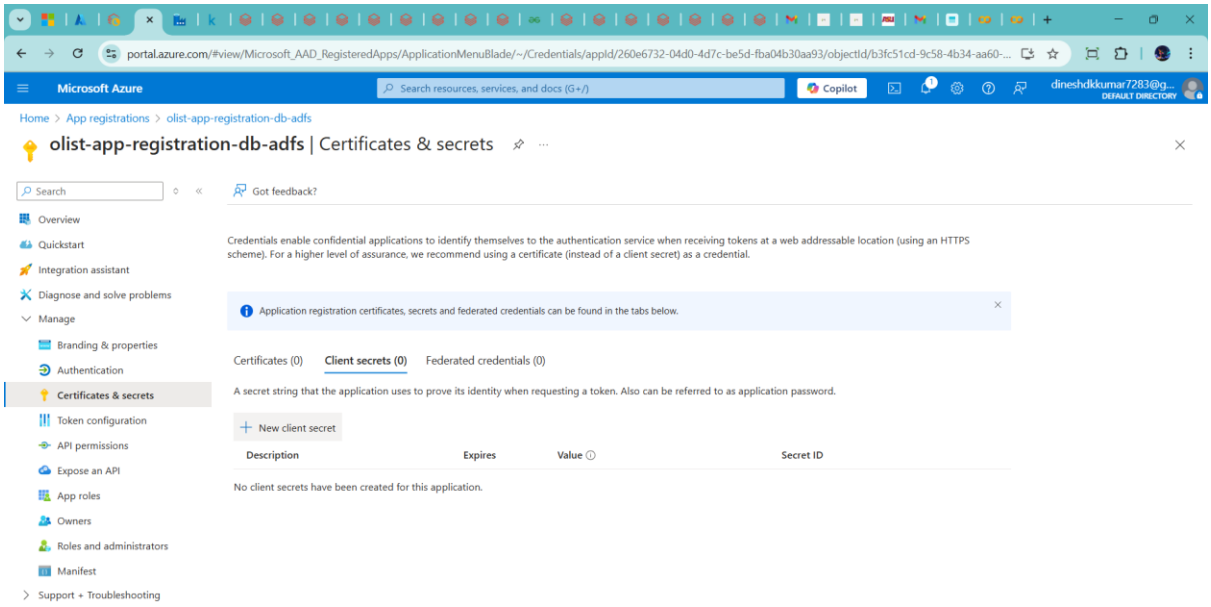




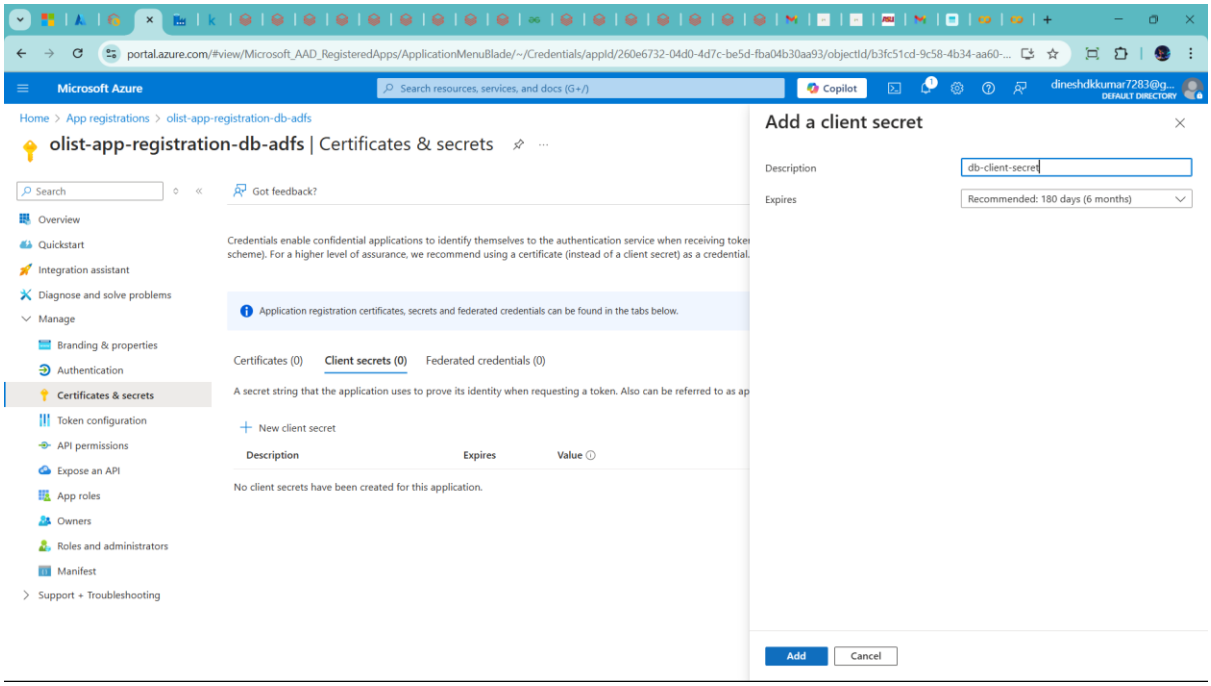
→fill and click registeration



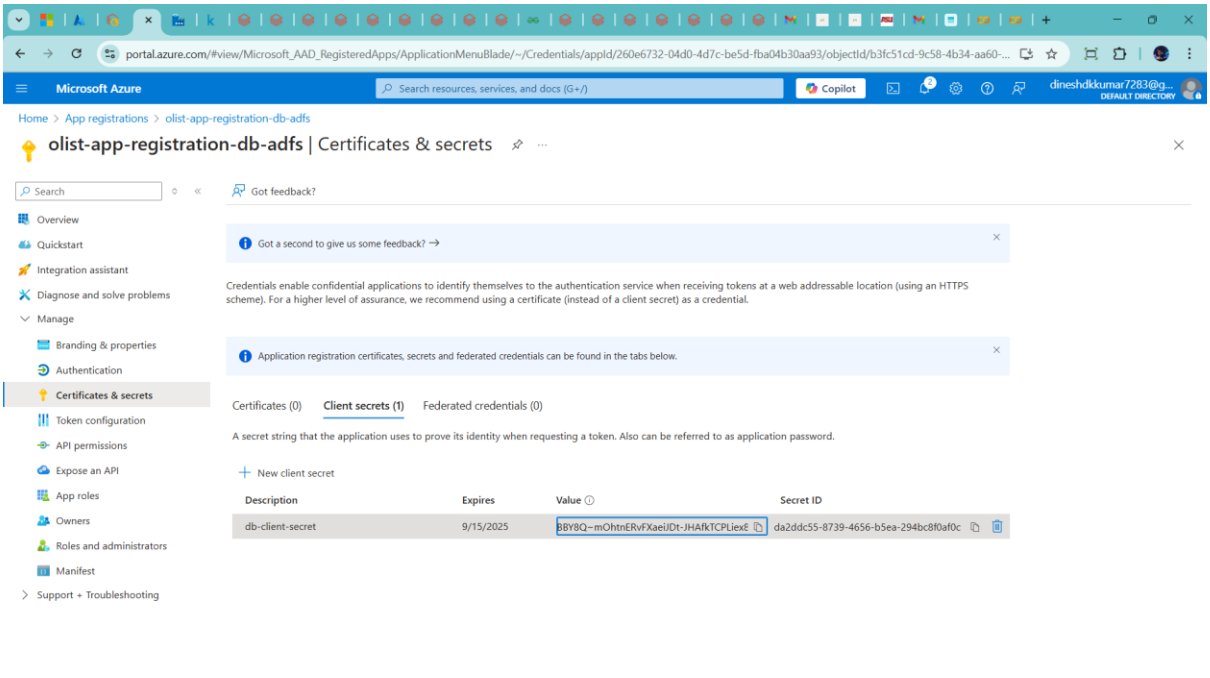
→click manage→certificate and secrets →click client secrets



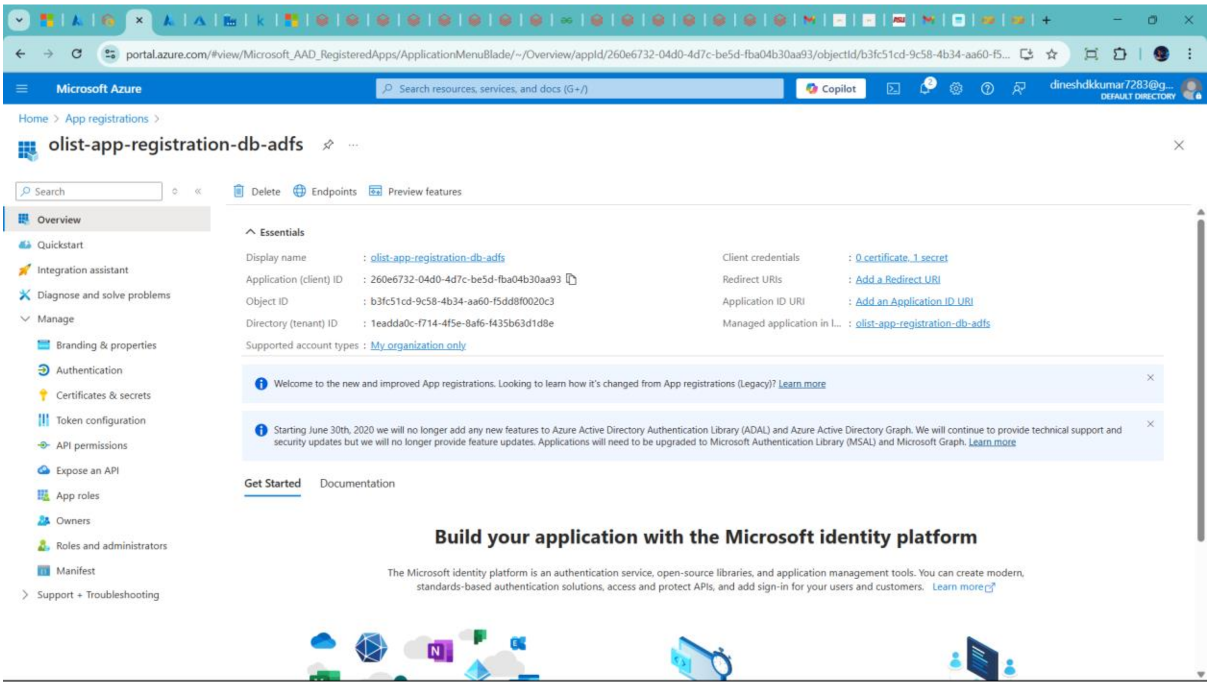
→click add



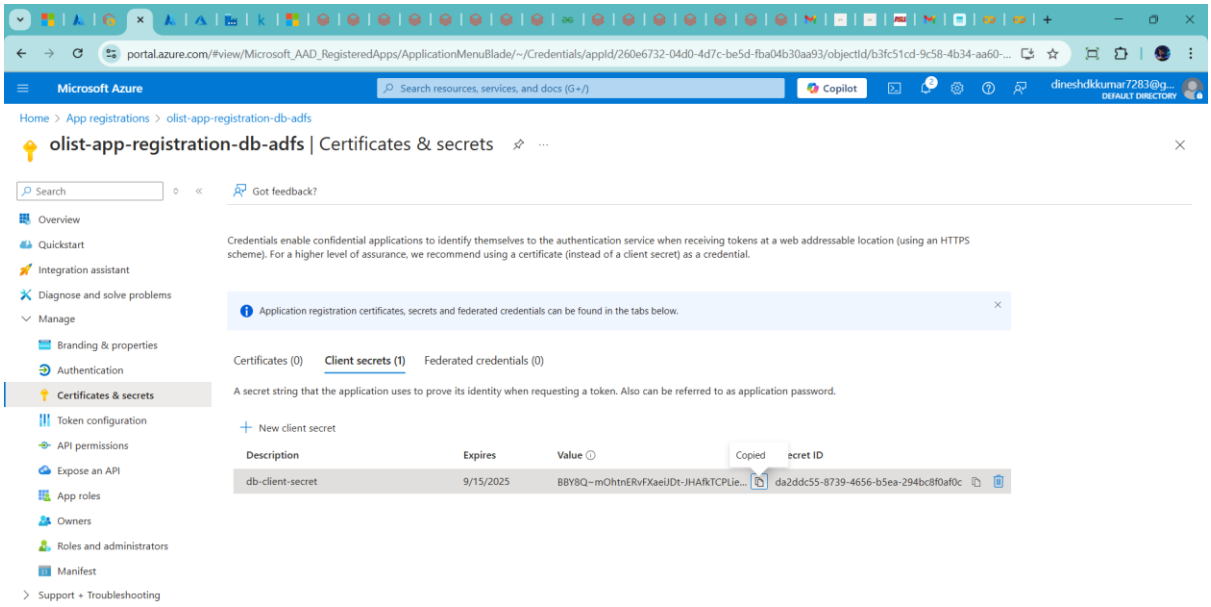
→value will be use →so copy the value



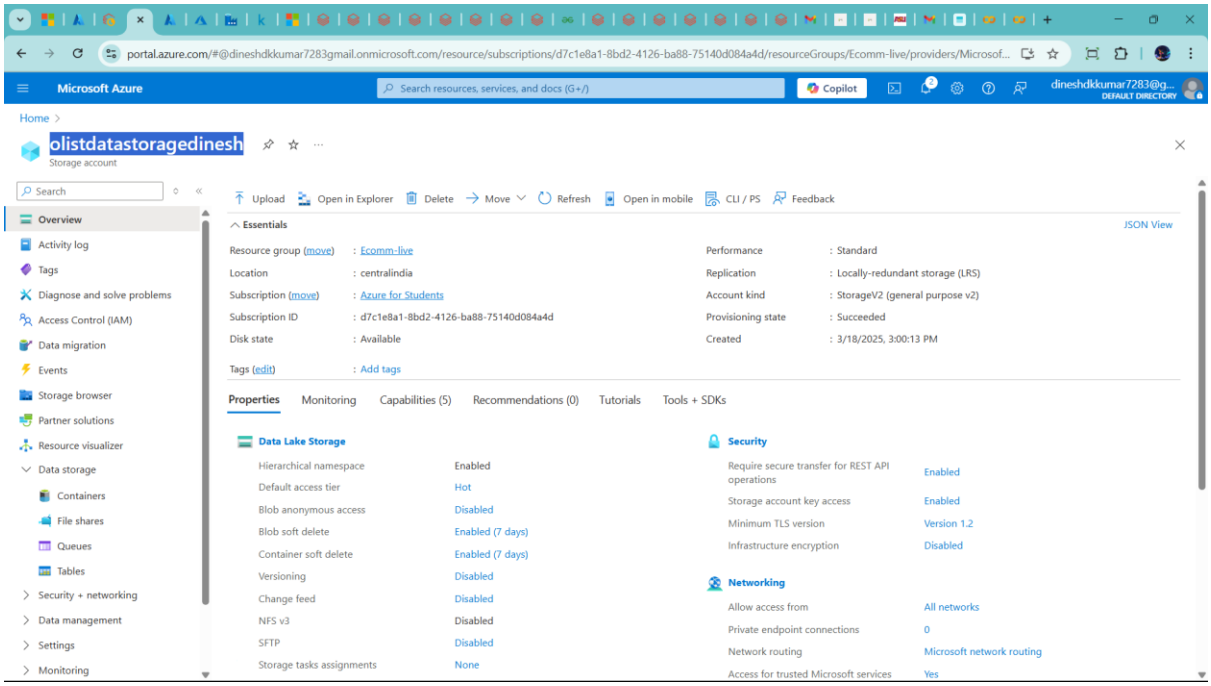
→now add application in this databricks

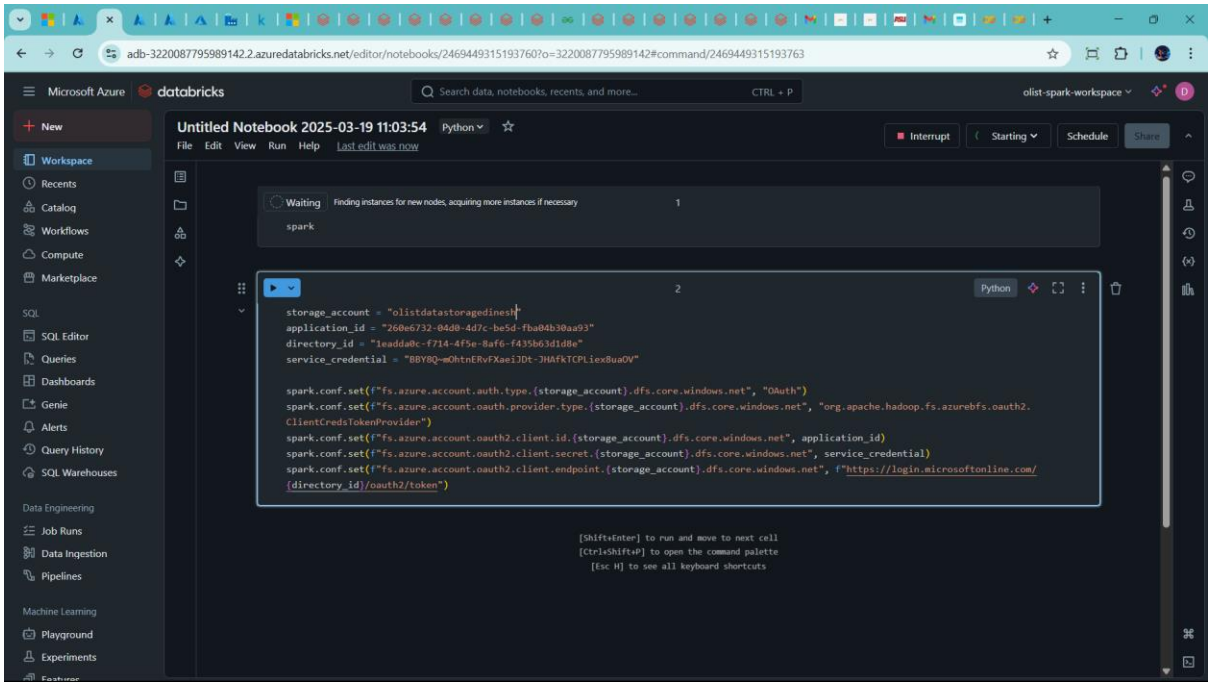


→copy as screted

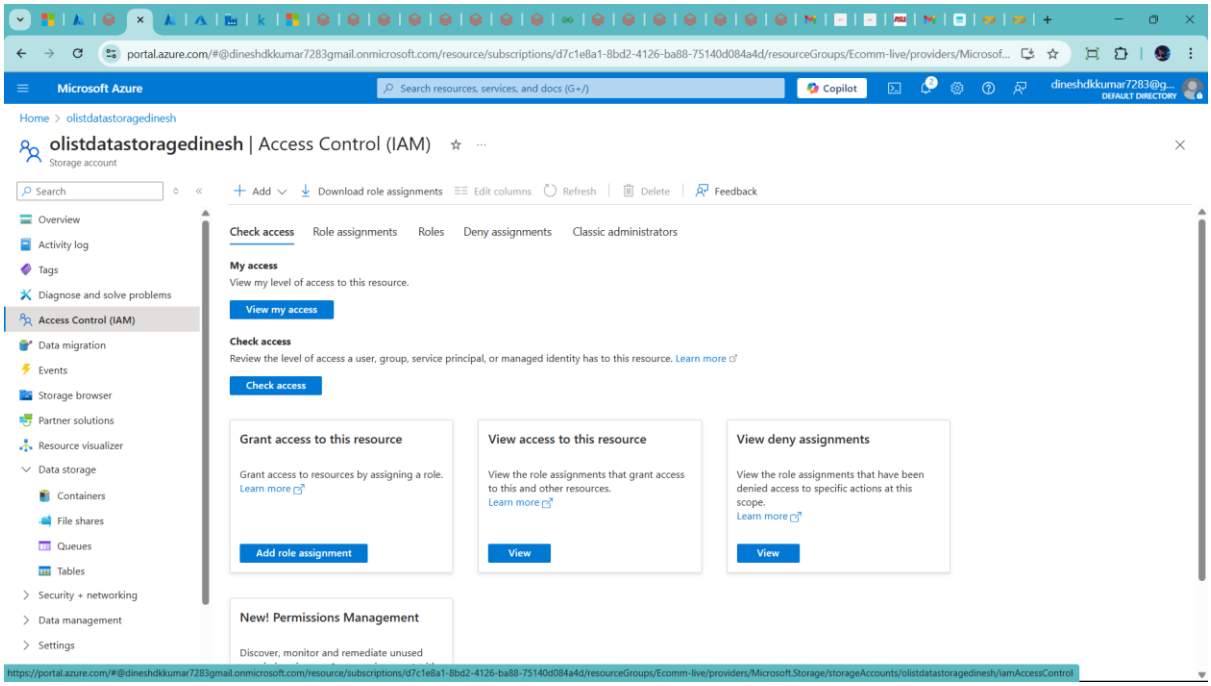


→copy name as storange account

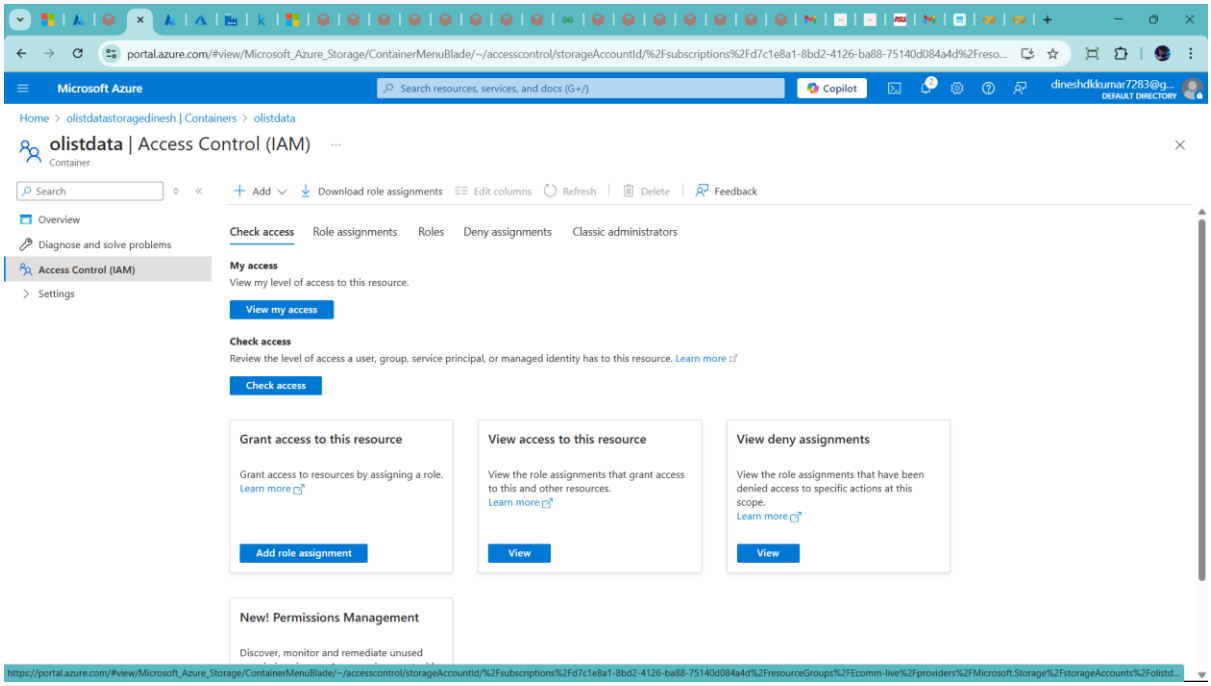




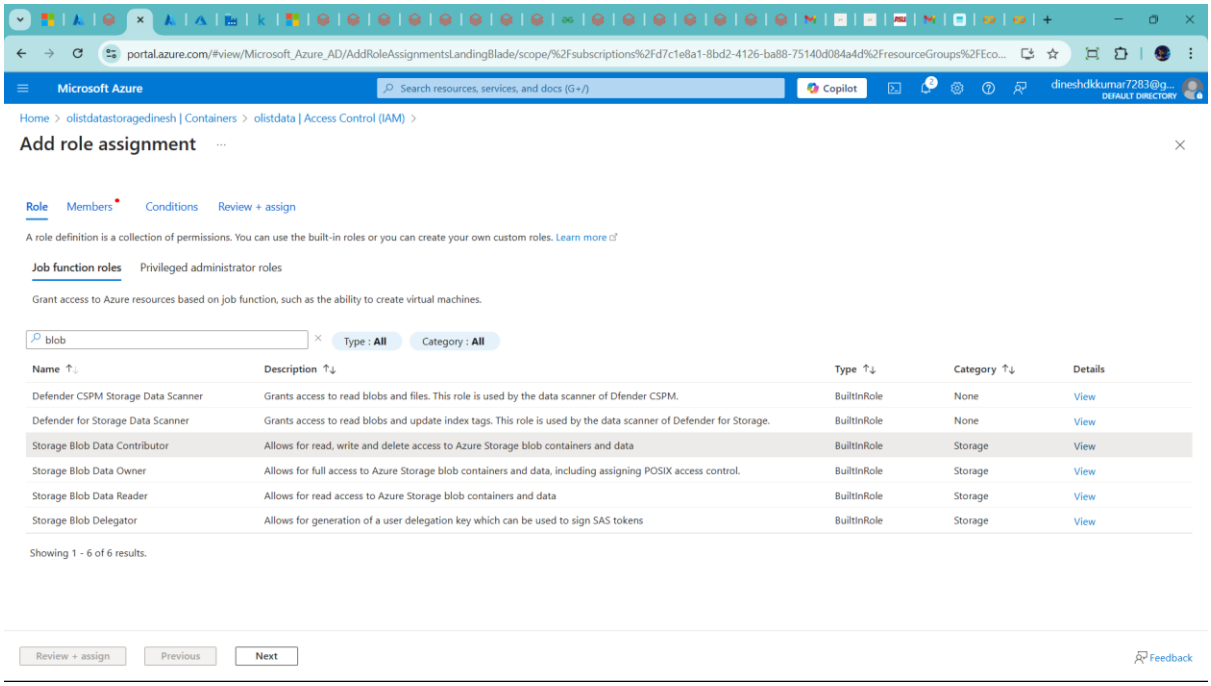
→go to container →IAM



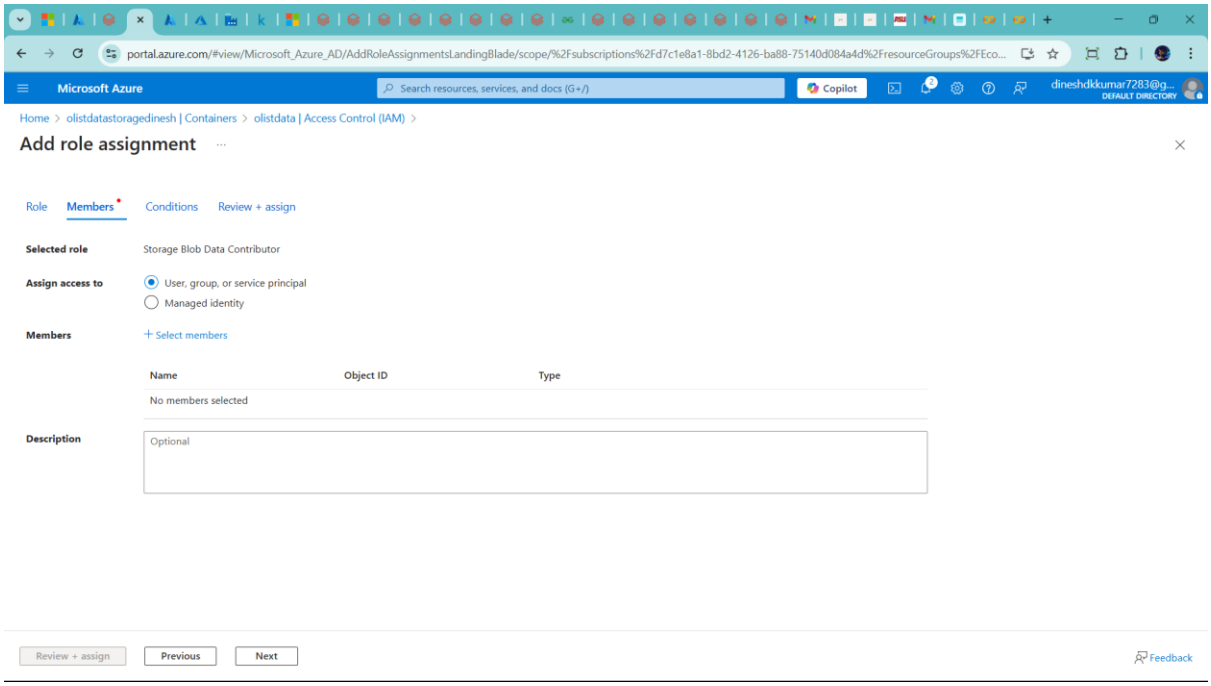
→not above

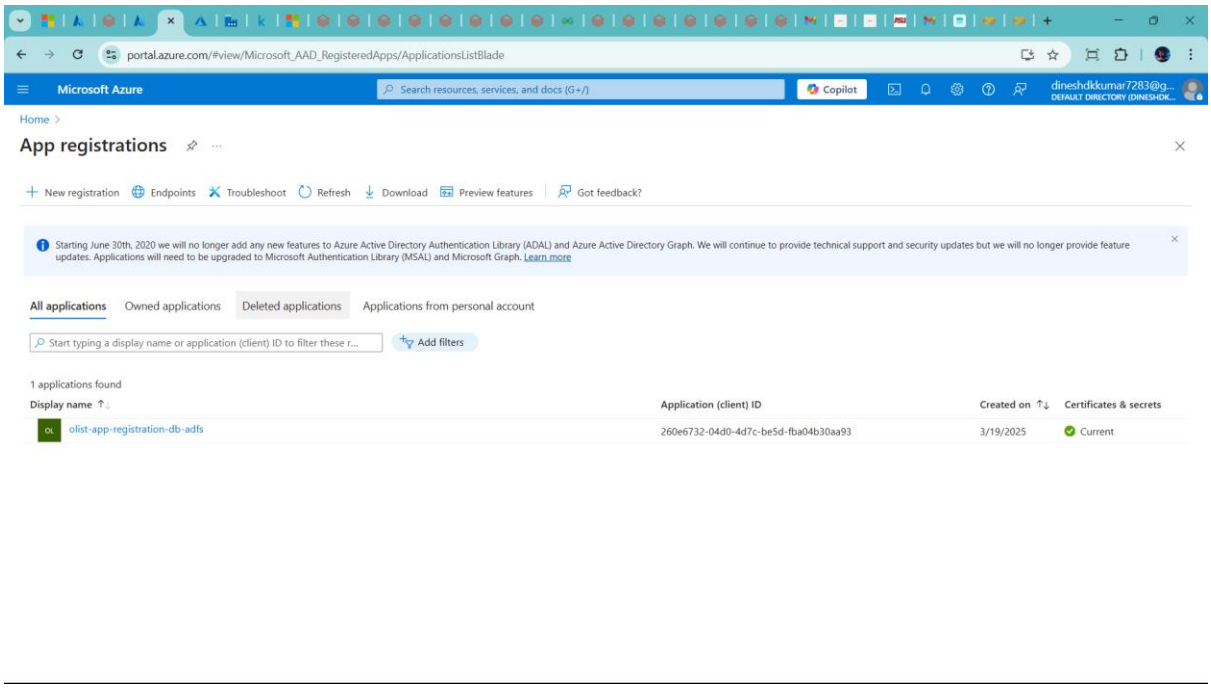


→click Add →Add assign role

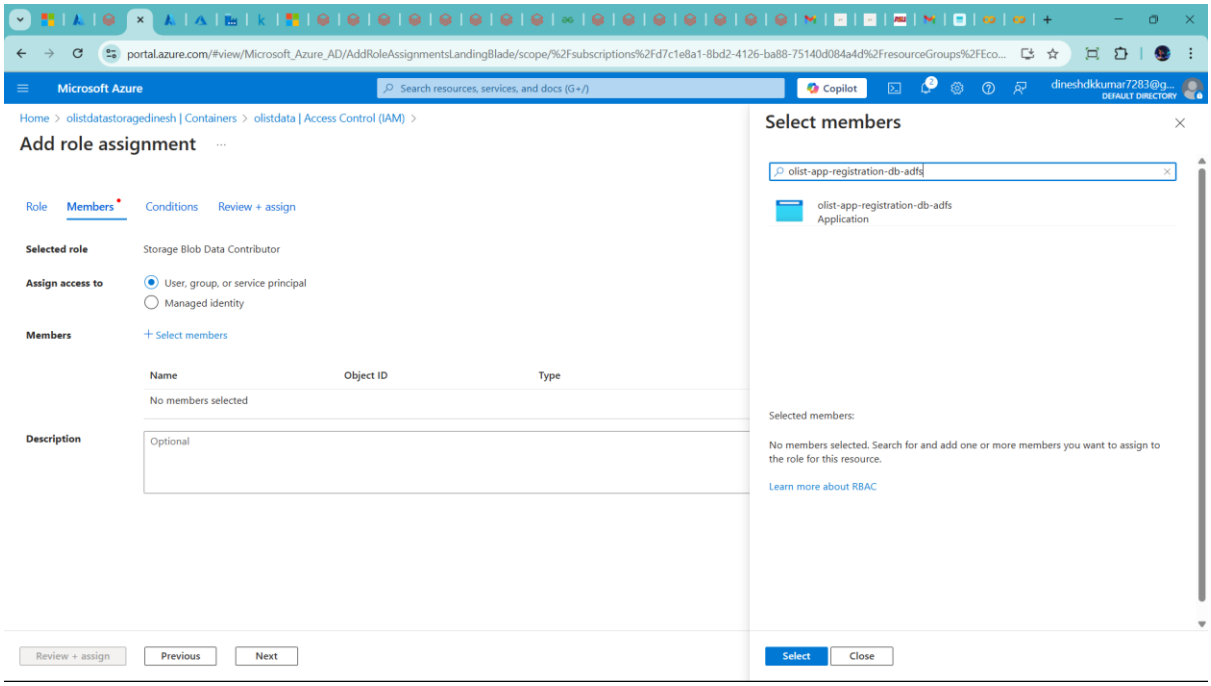


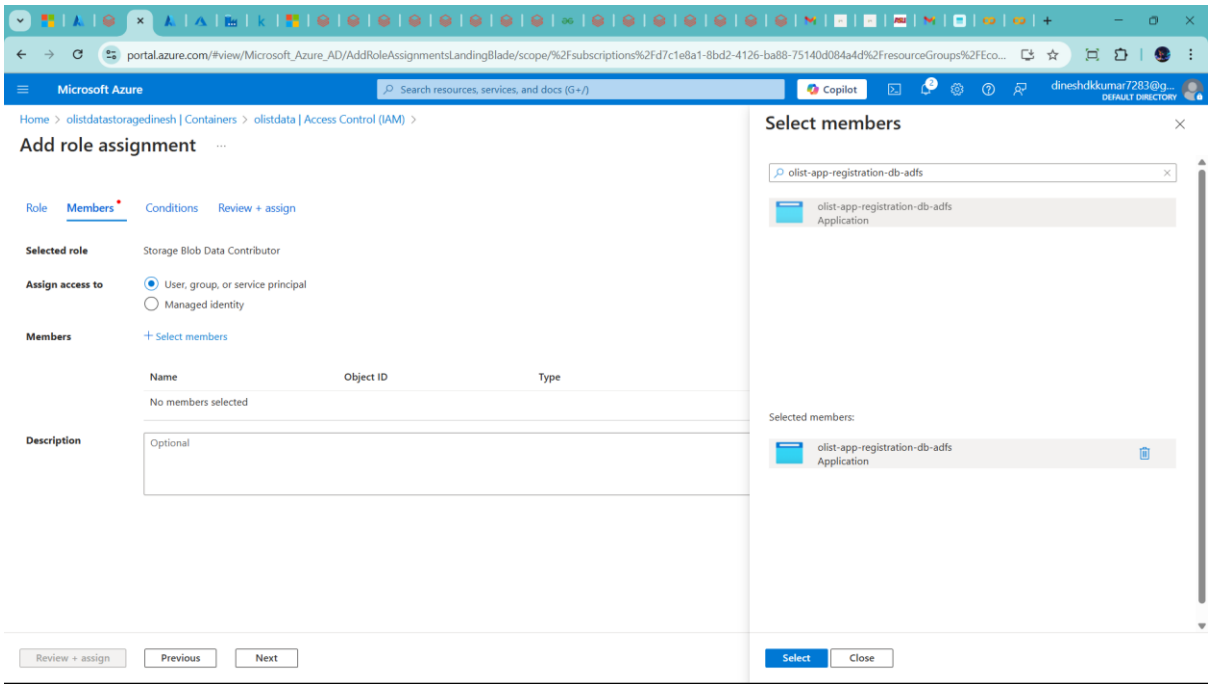
→click user group or principle → identity and select members

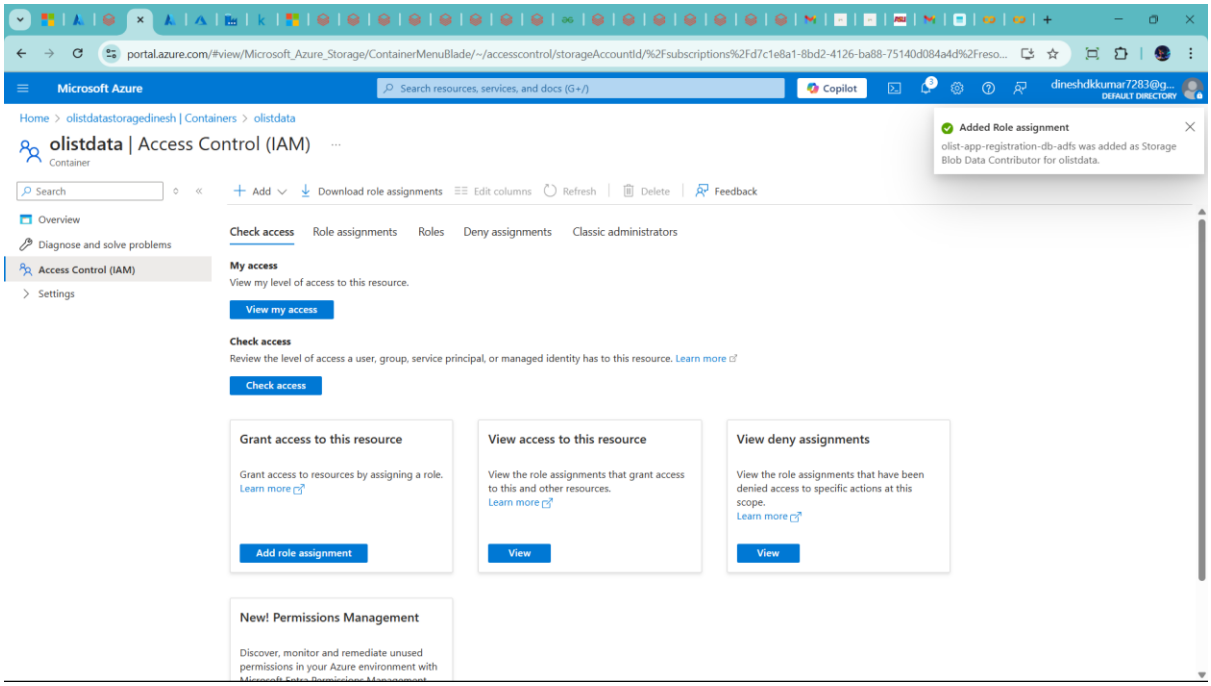




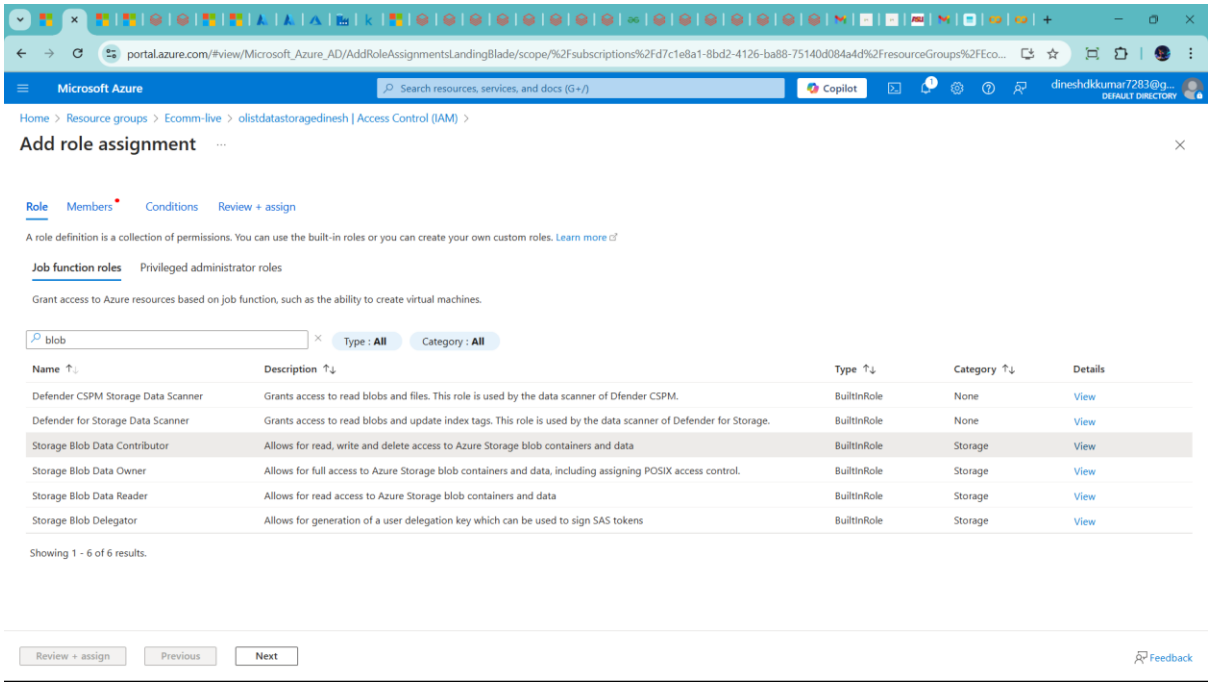
→same name in select members

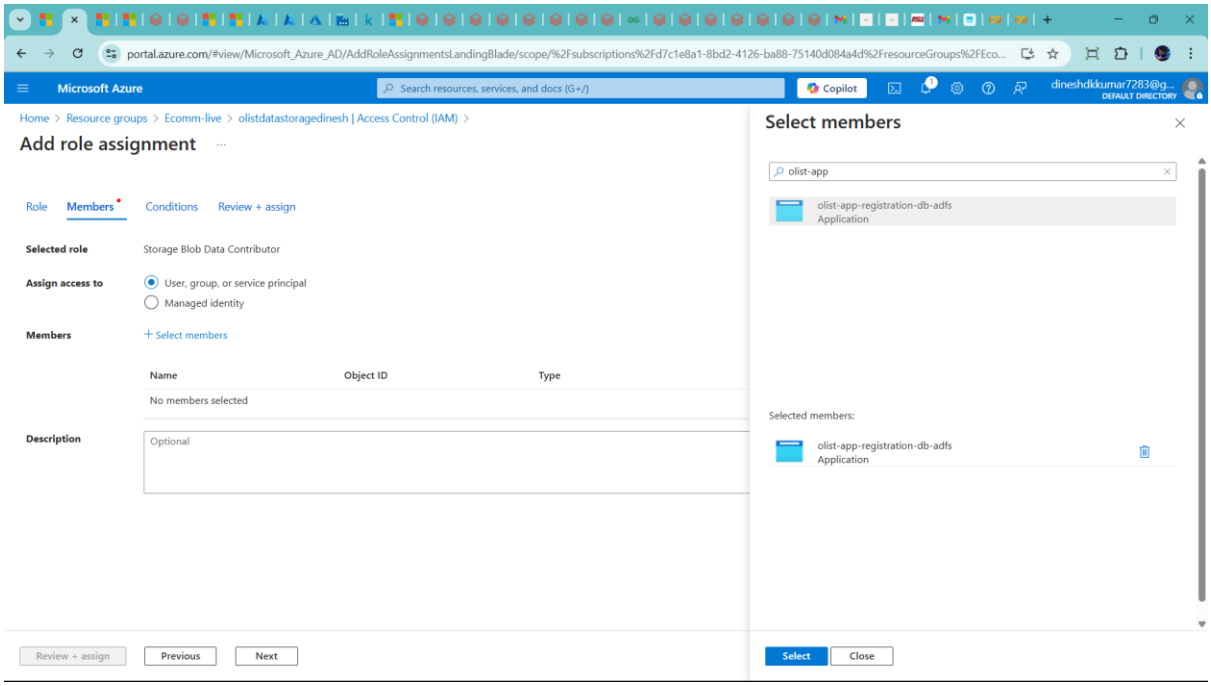

portal.azure.com/#view/Microsoft_Azure_AD/AddRoleAssignmentsLandingBlade/scope/%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEco...
Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
dineshdkkumar7283@g...
DEFAULT DIRECTORY
Home > olistdatastoragedinesh | Containers > olistdata | Access Control (IAM) >
Add role assignment
Role
Members
Conditions
Review + assign
Selected role
Storage Blob Data Contributor
Assign access to
User, group, or service principal
Managed identity
Members
+ Select members
Name
Object ID
Type
No members selected
Description
Optional
Review + assign
Previous
Next
Select members
olist-app-registration-db-adfs
olist-app-registration-db-adfs
Application
Selected members:
No members selected. Search for and add one or more members you want to assign to the role for this resource.
Learn more about RBAC
Select
Close

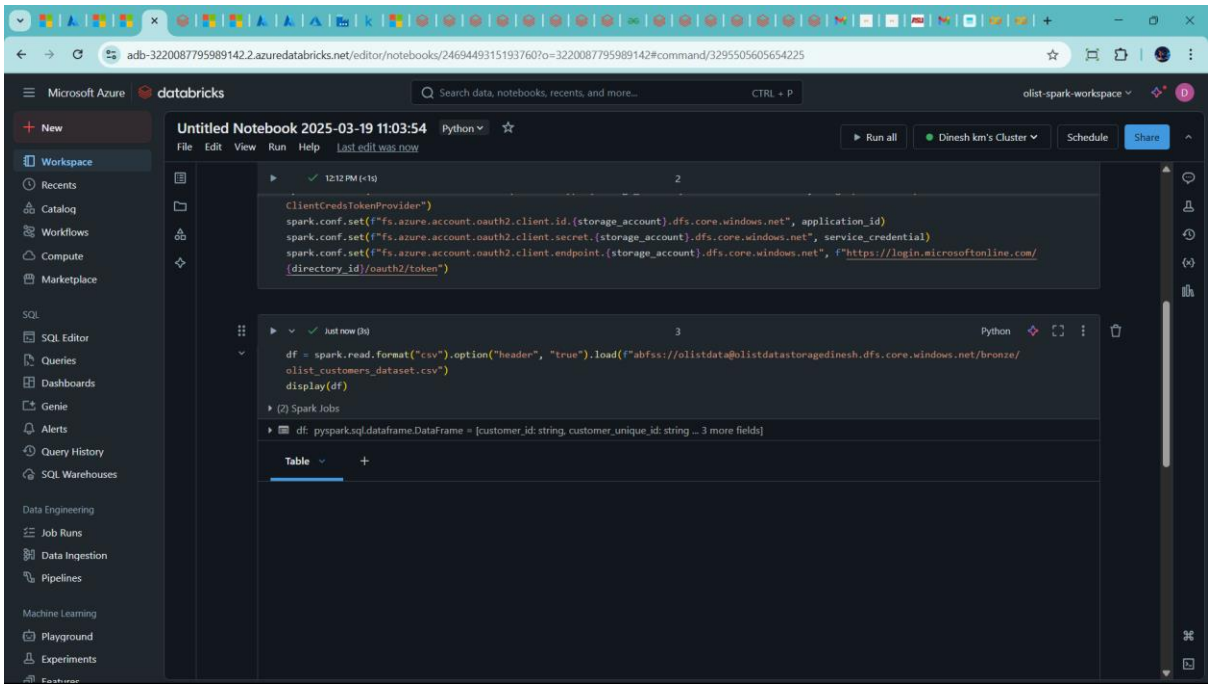

portal.azure.com/#view/Microsoft_Azure_AD/AddRoleAssignmentsLandingBlade/scope/%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEco...
Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
dineshdkkumar7283@g...
DEFAULT DIRECTORY
Home > olistdatastoragedinesh | Containers > olistdata | Access Control (IAM) >
Add role assignment
Role
Members
Conditions
Review + assign
Selected role
Storage Blob Data Contributor
Assign access to
User, group, or service principal
Managed identity
Members
+ Select members
Name
Object ID
Type
No members selected
Description
Optional
Review + assign
Previous
Next
Select members
olist-app-registration-db-adfs
olist-app-registration-db-adfs
Application
Selected members:
olist-app-registration-db-adfs
Application
Select
Close


Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
Home > olistdatastoragedinesh | Containers > olistdata
olistdata | Access Control (IAM)
Container
Added Role assignment
olist-app-registration-db-adfs was added as Storage Blob Data Contributor for olistdata.
Search
Add
Download role assignments
Edit columns
Refresh
Delete
Feedback
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Check access
Role assignments
Roles
Deny assignments
Classic administrators
My access
View my level of access to this resource.
View my access
Check access
Review the level of access a user, group, service principal, or managed identity has to this resource. Learn more
Check access
Grant access to this resource
Grant access to resources by assigning a role.
Learn more
Add role assignment
View access to this resource
View the role assignments that grant access to this and other resources.
Learn more
View
View deny assignments
View the role assignments that have been denied access to specific actions at this scope.
Learn more
View
New! Permissions Management
Discover, monitor and remediate unused permissions in your Azure environment with

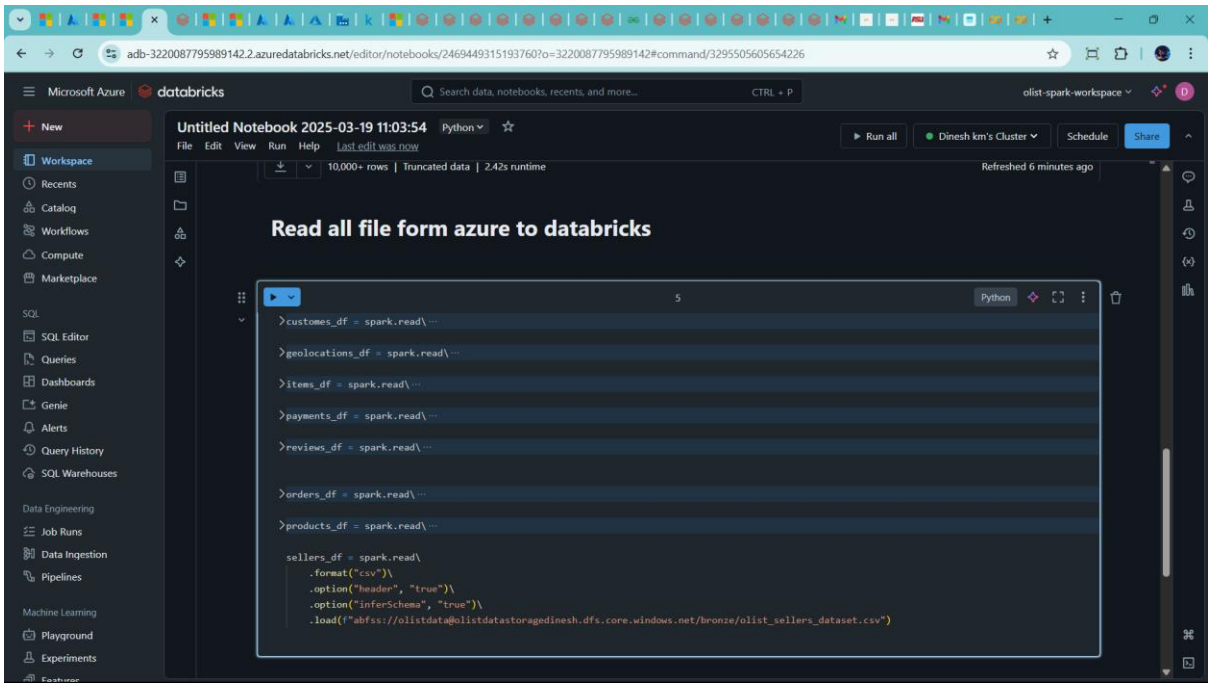
→verfy again


Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
Home > Resource groups > Ecomm-live > olistdatastoragedinesh | Access Control (IAM) >
Add role assignment
Role
Members
Conditions
Review + assign
A role definition is a collection of permissions. You can use the built-in roles or you can create your own custom roles. Learn more
Job function roles
Privileged administrator roles
Grant access to Azure resources based on job function, such as the ability to create virtual machines.
blob
Type : All
Category : All
Name
Description
Type
Category
Details
Defender CSPM Storage Data Scanner
Grants access to read blobs and files. This role is used by the data scanner of Dfender CSPM.
BuiltInRole
None
View
Defender for Storage Data Scanner
Grants access to read blobs and update index tags. This role is used by the data scanner of Defender for Storage.
BuiltInRole
None
View
Storage Blob Data Contributor
Allows for read, write and delete access to Azure Storage blob containers and data
BuiltInRole
Storage
View
Storage Blob Data Owner
Allows for full access to Azure Storage blob containers and data, including assigning POSIX access control.
BuiltInRole
Storage
View
Storage Blob Data Reader
Allows for read access to Azure Storage blob containers and data
BuiltInRole
Storage
View
Storage Blob Delegator
Allows for generation of a user delegation key which can be used to sign SAS tokens
BuiltInRole
Storage
View
Showing 1 - 6 of 6 results.
Review + assign
Previous
Next
Feedback

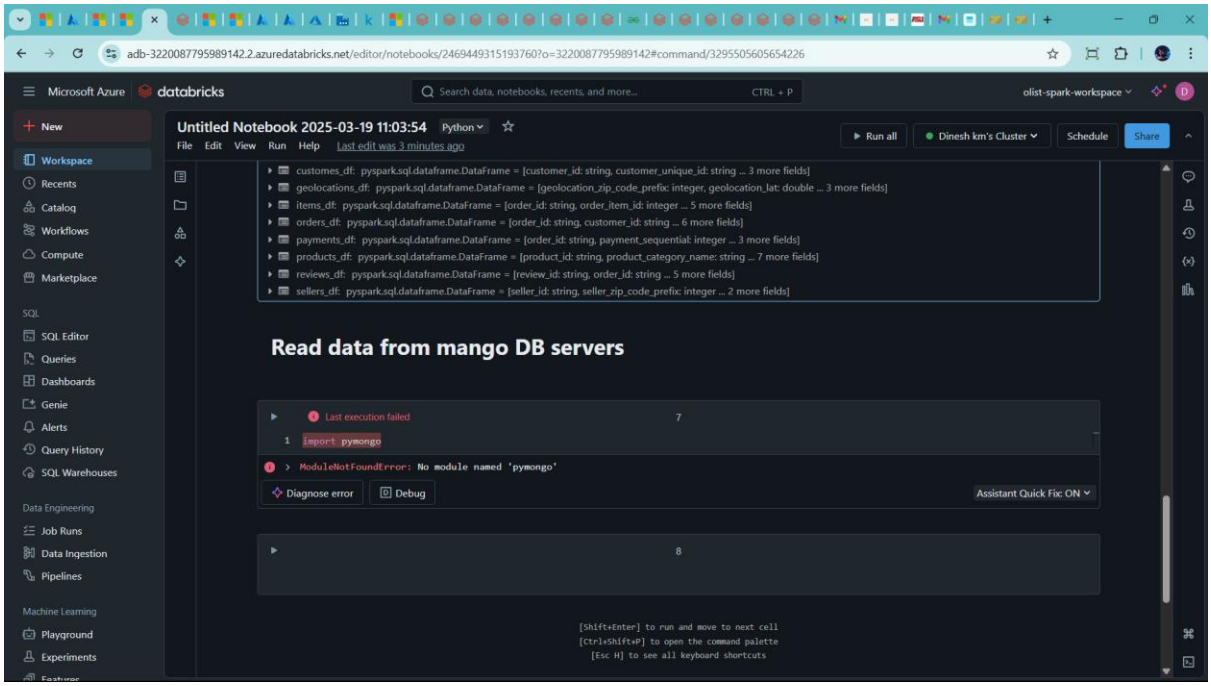

Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
Home > Resource groups > Ecomm-live > olistdatastoragedinesh | Access Control (IAM) >
Add role assignment
Role
Members
Conditions
Review + assign
Selected role
Storage Blob Data Contributor
Assign access to
User, group, or service principal
Managed identity
Members
+ Select members
Name
Object ID
Type
No members selected
Description
Optional
Review + assign
Previous
Next
Select members
olist-app
olist-app-registration-db-adfs
Application
Selected members:
olist-app-registration-db-adfs
Application
Select
Close


Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
olist-spark-workspace
New
Workspace
Recents
Catalog
Workflows
Compute
Marketplace
SQL
SQL Editor
Queries
Dashboards
Genie
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Pipelines
Machine Learning
Playground
Experiments
Untitled Notebook 2025-03-19 11:03:54
Python
File Edit View Run Help Last edit was now
Run all
Dinesh km's Cluster
Schedule
Share
ClientCredsTokenProvider")
spark.conf.set(f"fs.azure.account.oauth2.client.id.{storage_account}.dfs.core.windows.net", application_id)
spark.conf.set(f"fs.azure.account.oauth2.client.secret.{storage_account}.dfs.core.windows.net", service_credential)
spark.conf.set(f"fs.azure.account.oauth2.client.endpoint.{storage_account}.dfs.core.windows.net", f"https://login.microsoftonline.com/{directory_id}/oauth2/token")
df = spark.read.format("csv").option("header", "true").load(f"abfss://olistdata@olistdatastoragedinesh.dfs.core.windows.net/bronze/olist_customers_dataset.csv")
display(df)
(2) Spark Jobs
df: pyspark.sql.dataframe.DataFrame = [customer_id: string, customer_unique_id: string ... 3 more fields]
Table

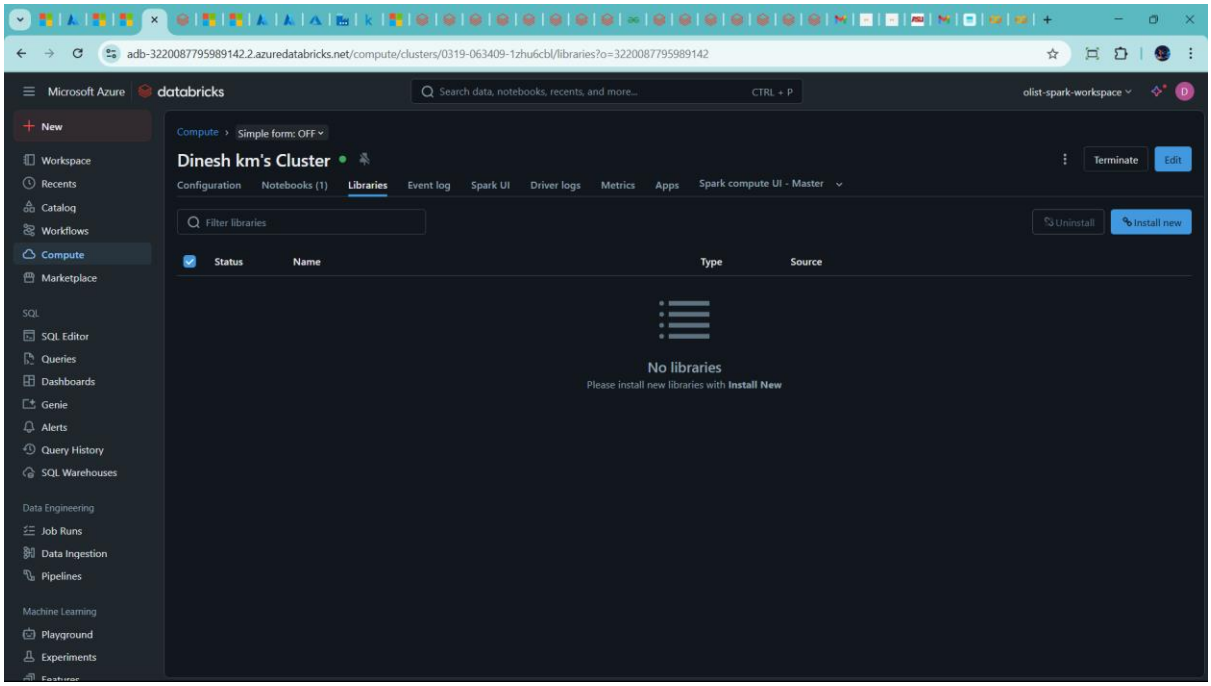
Part--2



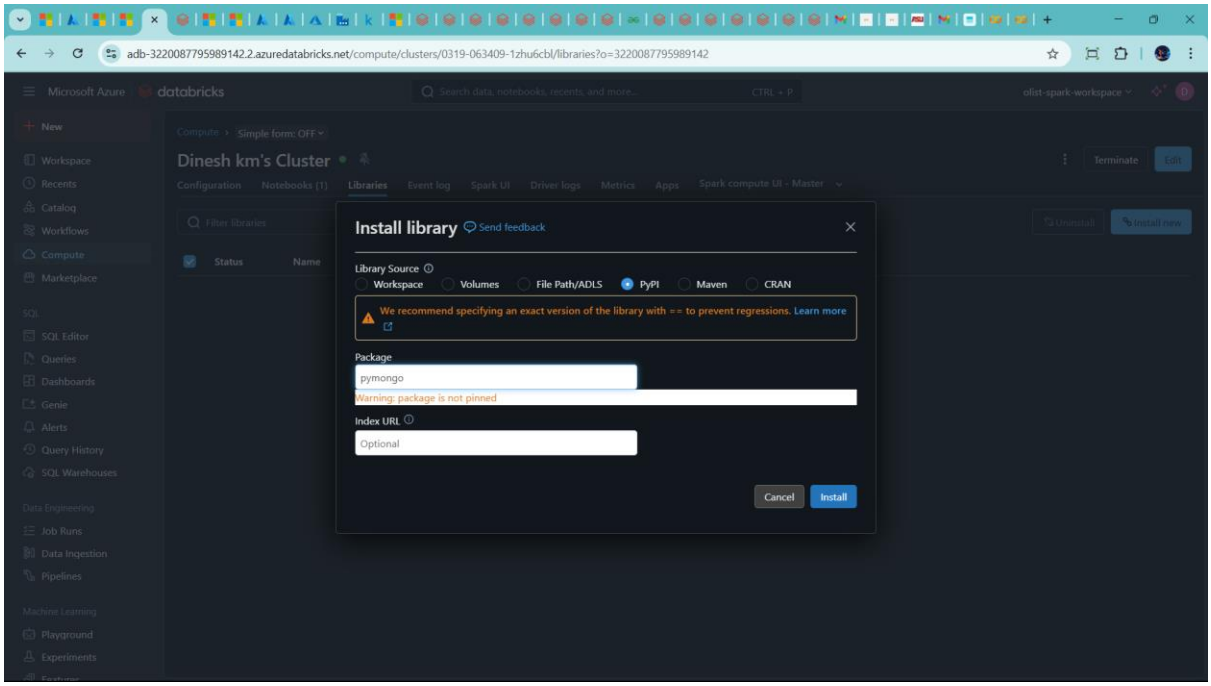
→they show error

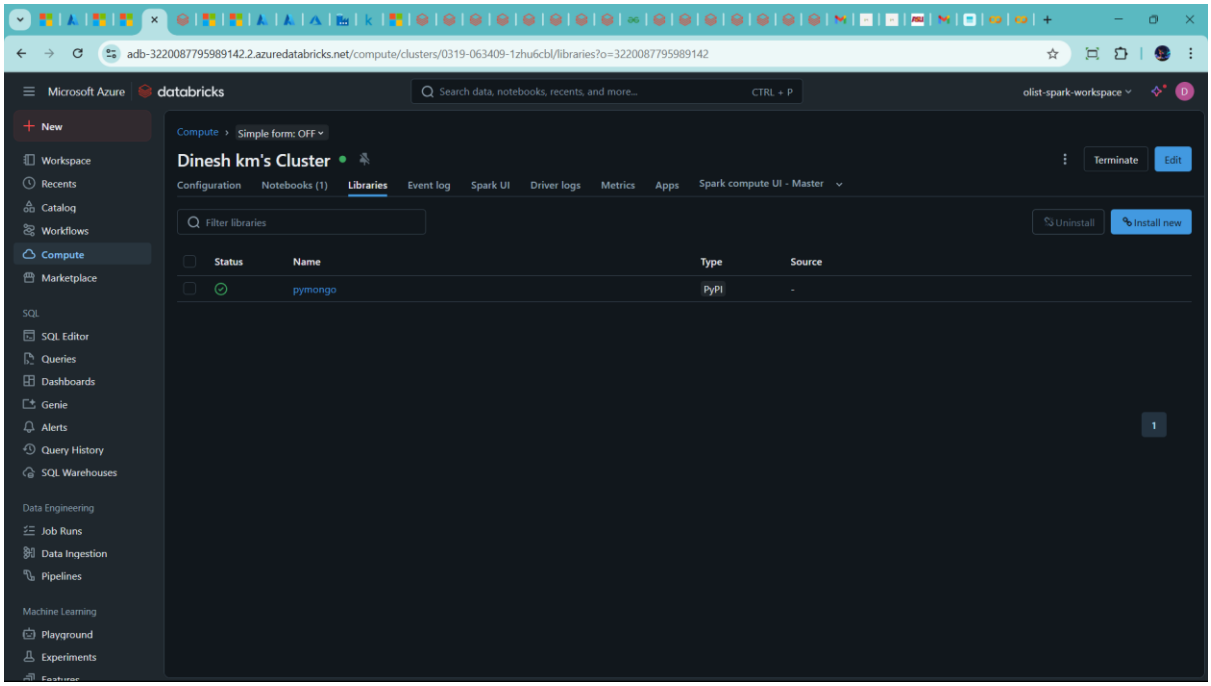


→go to compute→click libries

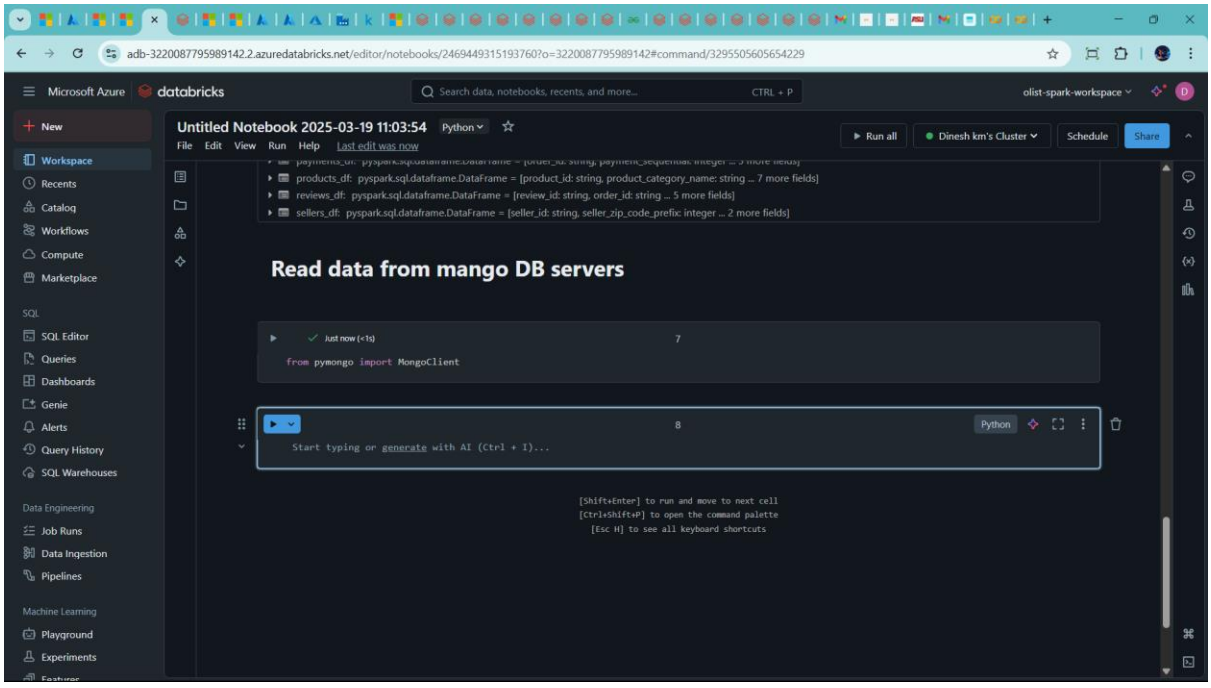


→click install new →click pypi→type pymongo →click install

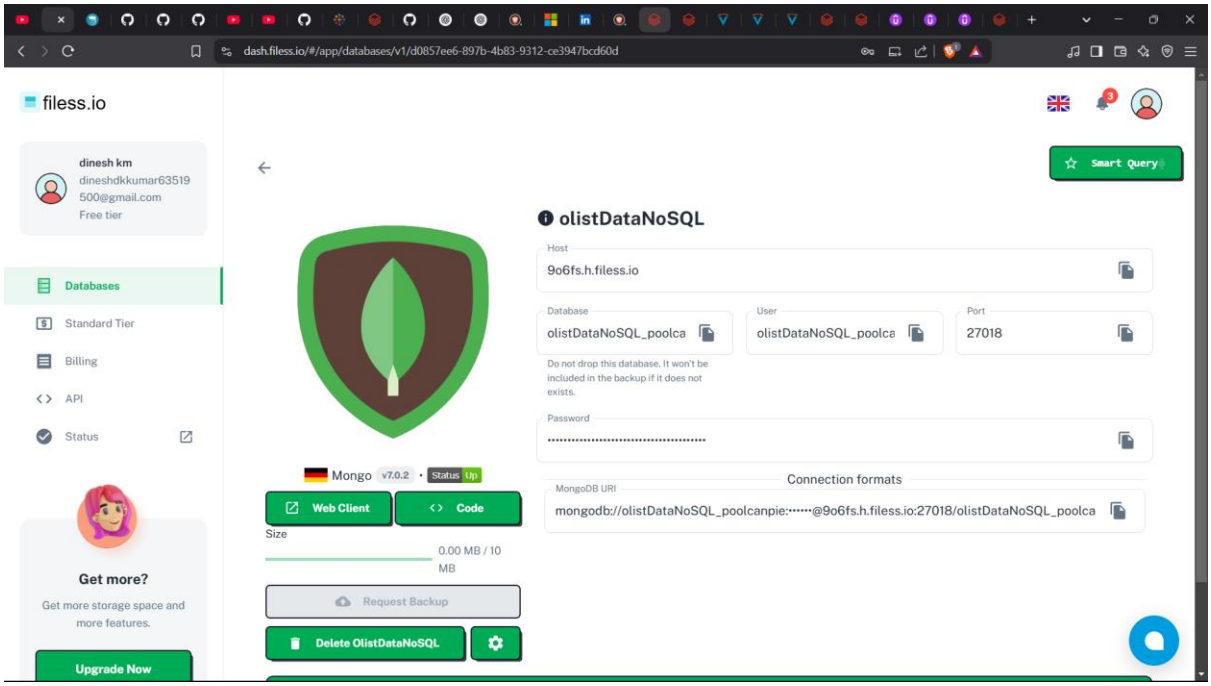




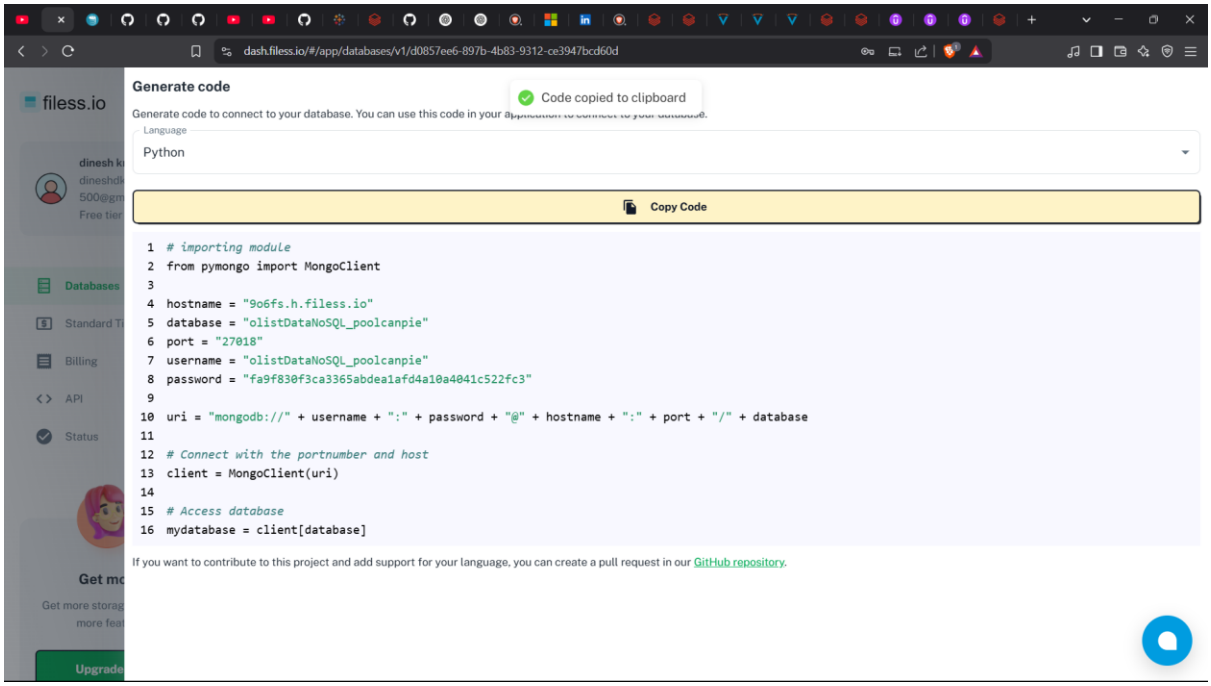
→go workspace in notebook run again



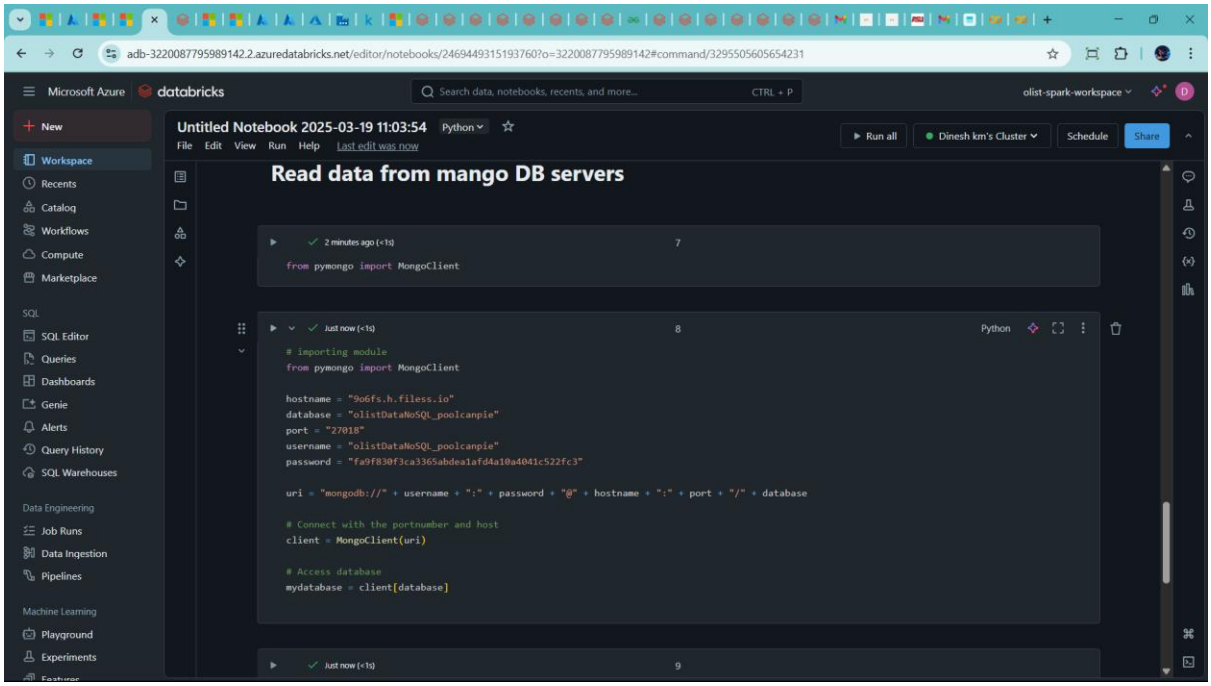
→go to files.io in mongodb

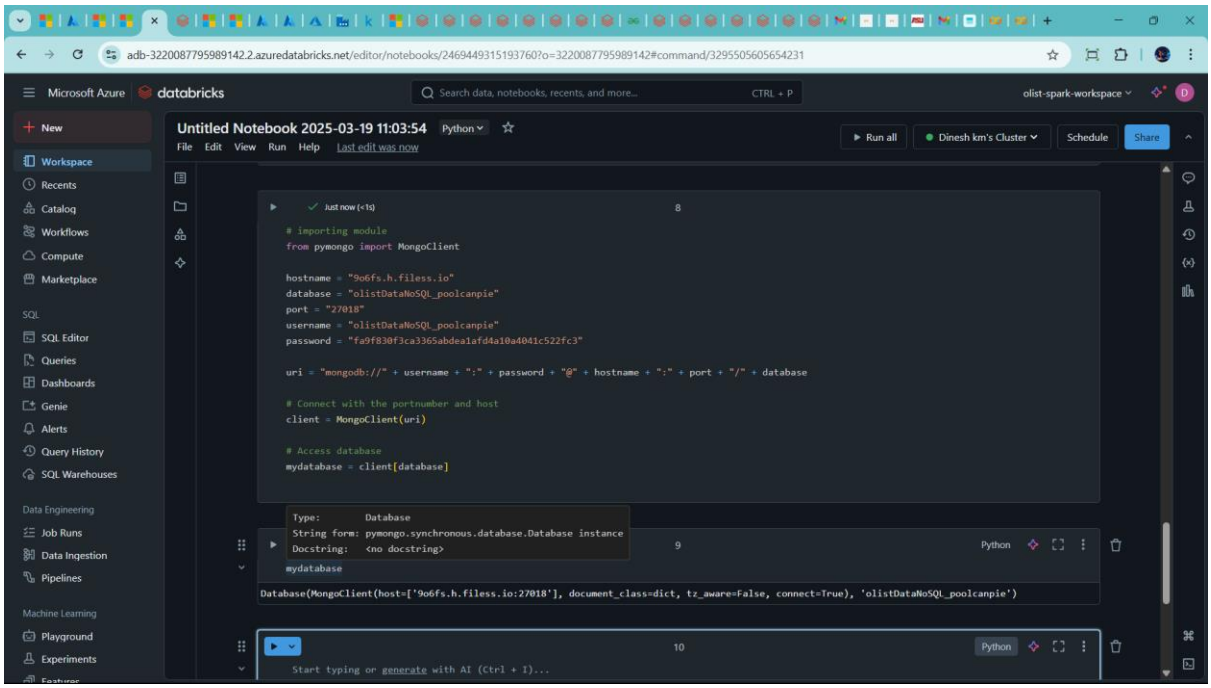


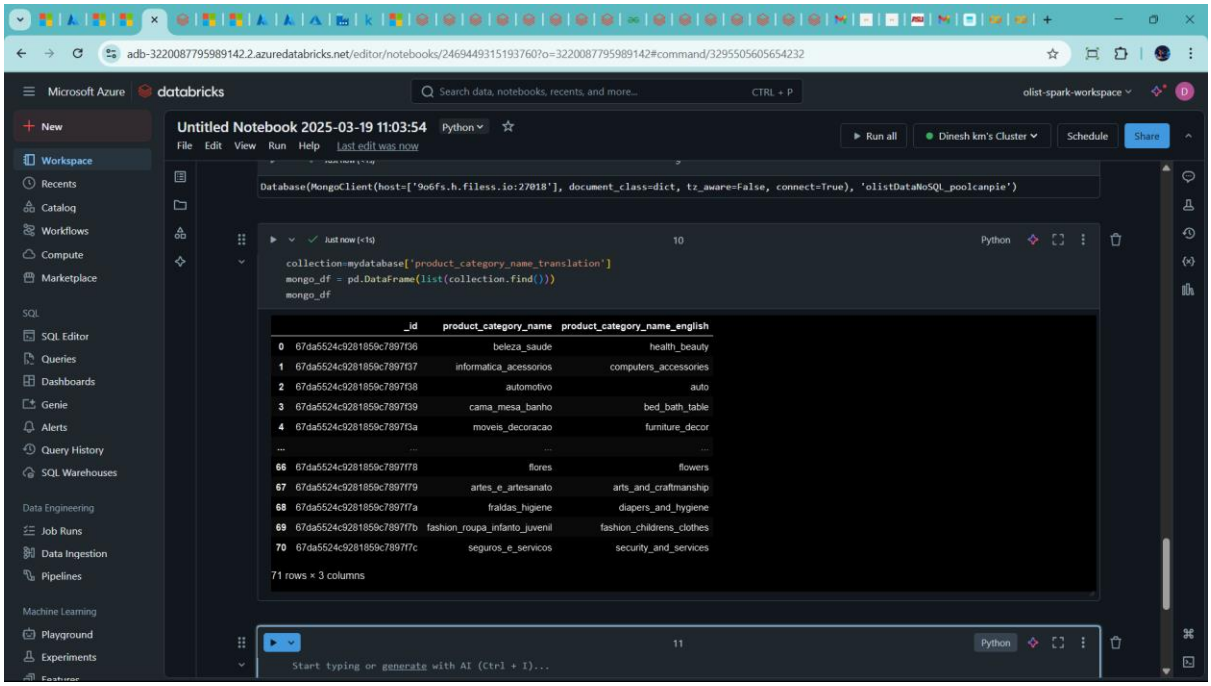
→click code→seach python and copy

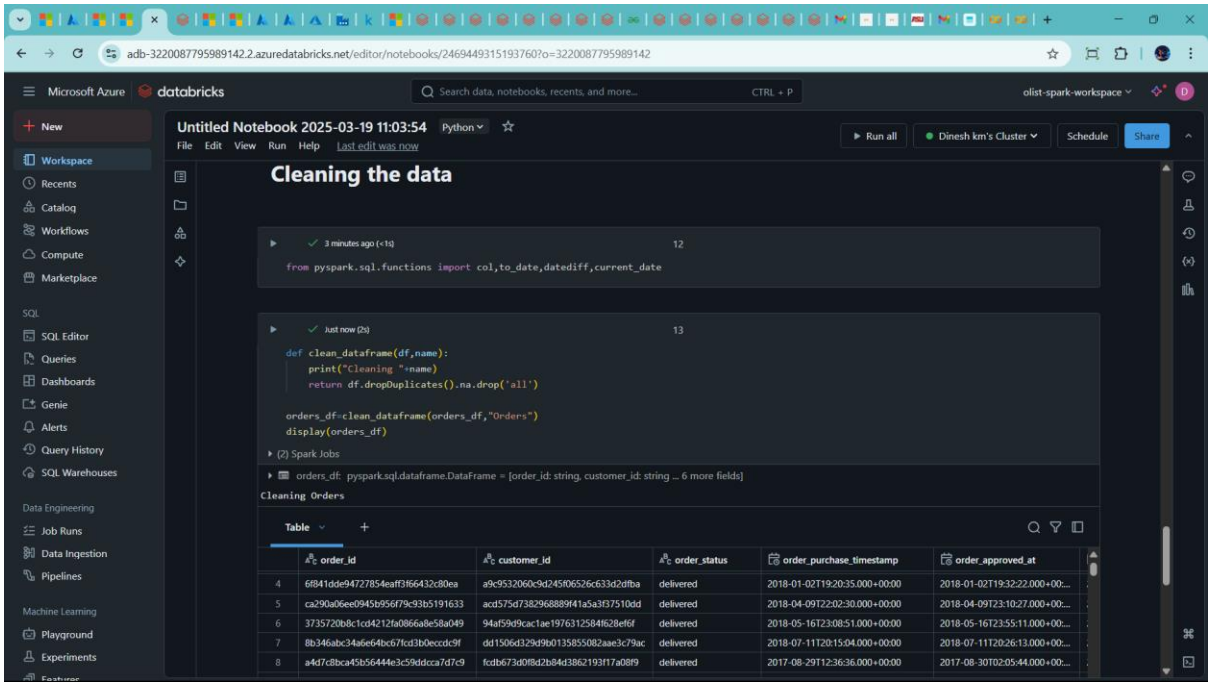


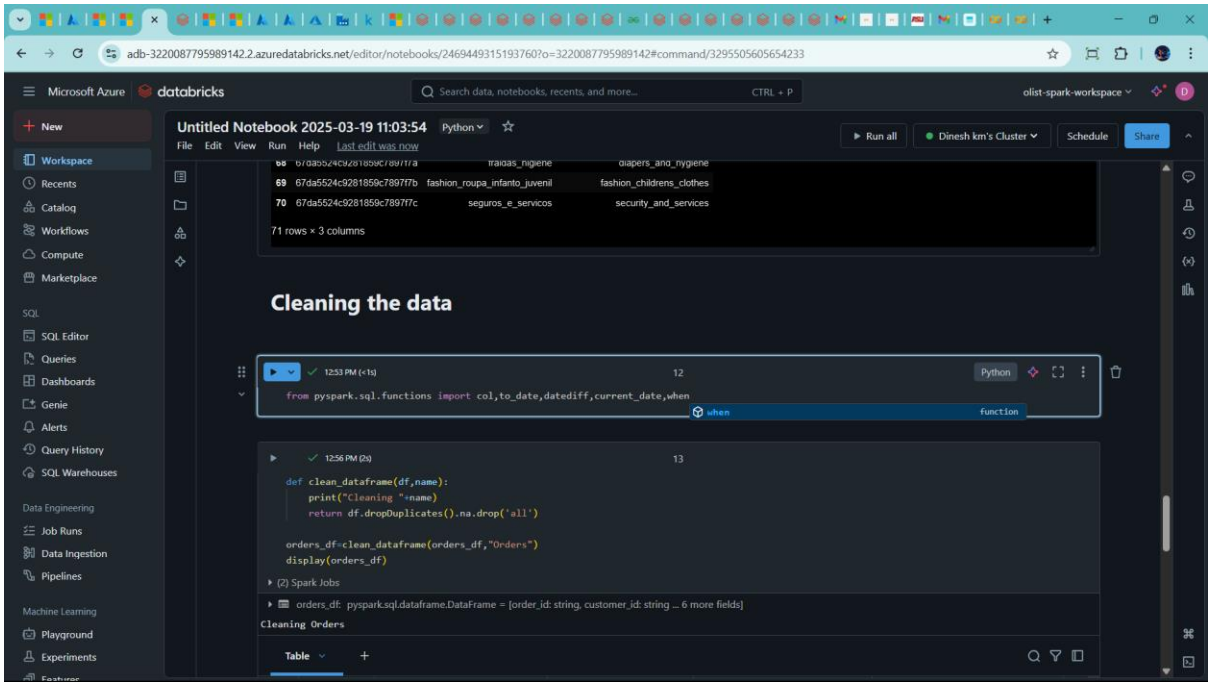
→add in notebook

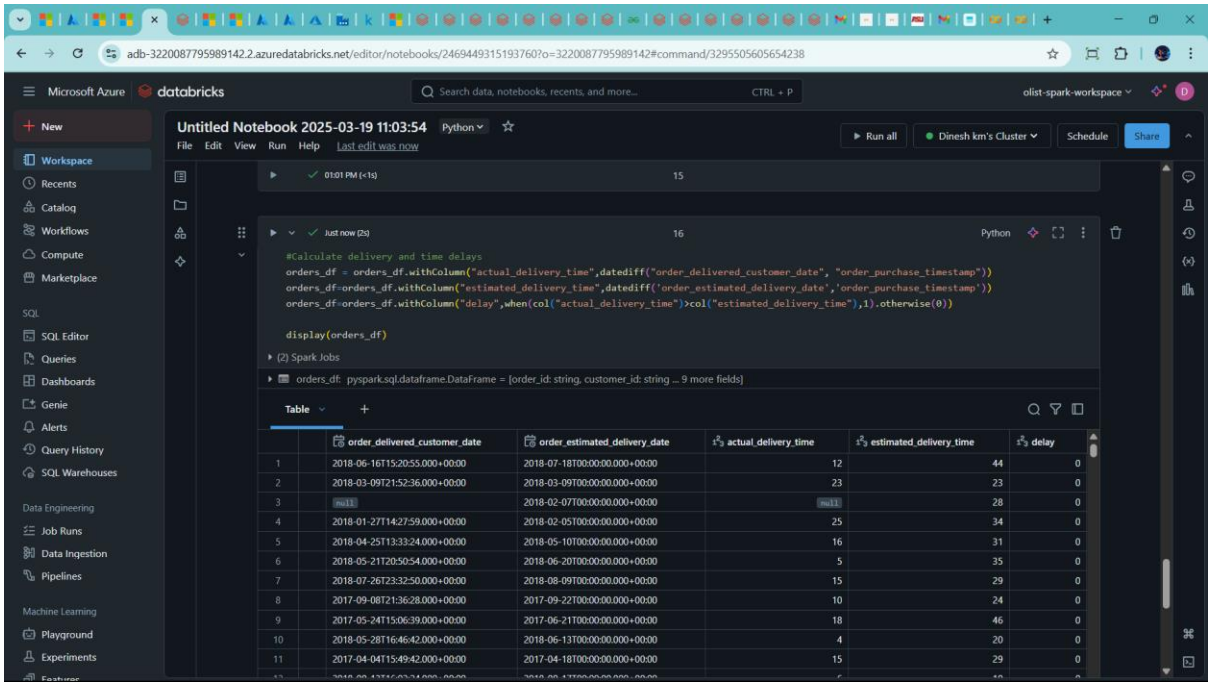

adb-3220087795989142.2.azuredatabricks.net/editor/notebooks/2469449315193760?o=3220087795989142#command/3295505605654231
Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
olist-spark-workspace
New
Workspace
Recents
Catalog
Workflows
Compute
Marketplace
SQL
SQL Editor
Queries
Dashboards
Genie
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Pipelines
Machine Learning
Playground
Experiments
Untitled Notebook 2025-03-19 11:03:54
Python
File Edit View Run Help
Last edit was now
Run all
Dinesh km's Cluster
Schedule
Share
Read data from mango DB servers
2 minutes ago (<1s)
7
from pymongo import MongoClient
Just now (<1s)
8
Python
# importing module
from pymongo import MongoClient
hostname = "9o6fs.h.filess.io"
database = "olistDataNoSQL_poolcanpie"
port = "27018"
username = "olistDataNoSQL_poolcanpie"
password = "fa9f830f3ca3365abdea1afd4a10a4041c522fc3"
uri = "mongodb://" + username + ":" + password + "@" + hostname + ":" + port + "/" + database
# Connect with the portnumber and host
client = MongoClient(uri)
# Access database
mydatabase = client[database]
Just now (<1s)
9


adb-3220087795989142.2.azuredatabricks.net/editor/notebooks/2469449315193760?o=3220087795989142#command/3295505605654231
Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
olist-spark-workspace
New
Workspace
Recents
Catalog
Workflows
Compute
Marketplace
SQL
SQL Editor
Queries
Dashboards
Genie
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Pipelines
Machine Learning
Playground
Experiments
Untitled Notebook 2025-03-19 11:03:54
Python
File Edit View Run Help
Last edit was now
Run all
Dinesh km's Cluster
Schedule
Share
Just now (<1s)
8
# importing module
from pymongo import MongoClient
hostname = "9o6fs.h.filess.io"
database = "olistDataNoSQL_poolcanpie"
port = "27018"
username = "olistDataNoSQL_poolcanpie"
password = "fa9f830f3ca3365abdea1afd4a10a4041c522fc3"
uri = "mongodb://" + username + ":" + password + "@" + hostname + ":" + port + "/" + database
# Connect with the portnumber and host
client = MongoClient(uri)
# Access database
mydatabase = client[database]
Type: Database
String form: pymongo.synchronous.database.Database instance
Docstring: <no docstring>
9
Python
mydatabase
Database(MongoClient(host=['9o6fs.h.filess.io:27018'], document_class=dict, tz_aware=False, connect=True), 'olistDataNoSQL_poolcanpie')
10
Python
Start typing or generate with AI (Ctrl + I)...

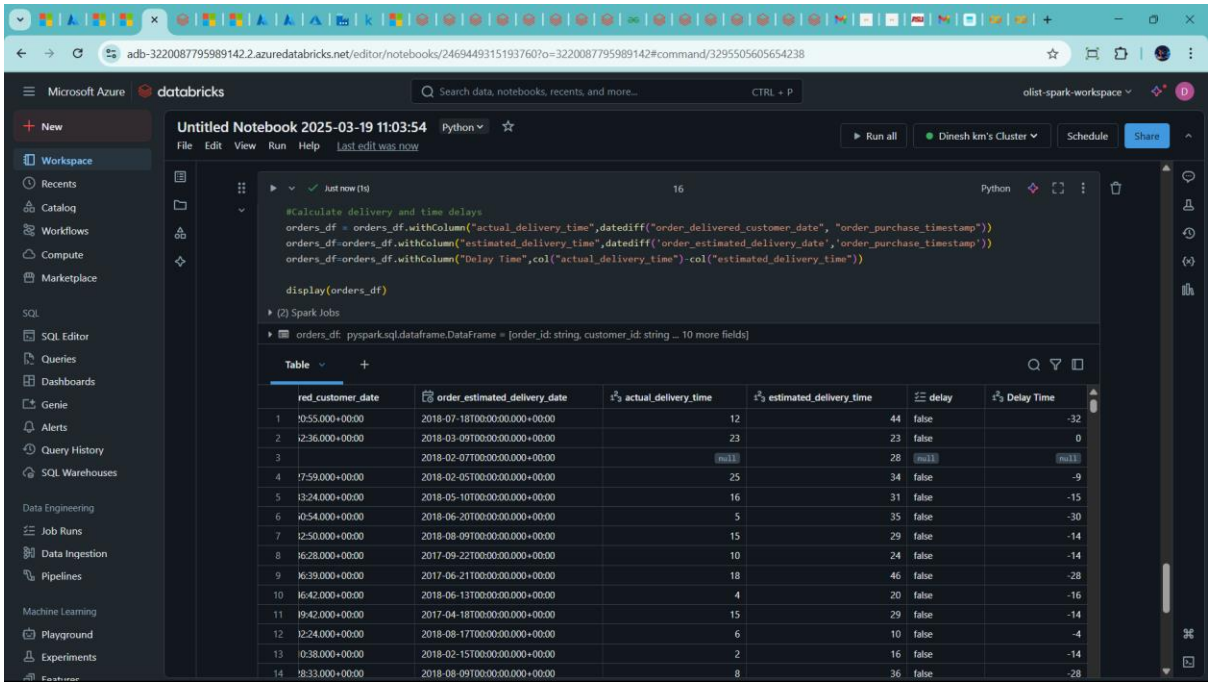

adb-3220087795989142.2.azuredatabricks.net/editor/notebooks/2469449315193760?o=3220087795989142#command/3295505605654232
Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
olist-spark-workspace
New
Workspace
Recents
Catalog
Workflows
Compute
Marketplace
SQL
SQL Editor
Queries
Dashboards
Genie
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Pipelines
Machine Learning
Playground
Experiments
Untitled Notebook 2025-03-19 11:03:54
Python
File Edit View Run Help Last edit was now
Run all
Dinesh km's Cluster
Schedule
Share
Database(MongoClient(host=['9o6fs.h.filess.io:27018'], document_class=dict, tz_aware=False, connect=True), 'olistDataNoSQL_poolcanpie')
Just now (<1s)
10
Python
collection=mydatabase['product_category_name_translation']
mongo_df = pd.DataFrame(list(collection.find()))
mongo_df
_id
product_category_name
product_category_name_english
0 67da5524c9281859c7897f36 beleza_saude health_beauty
1 67da5524c9281859c7897f37 informatica_acessorios computers_accessories
2 67da5524c9281859c7897f38 automotivo auto
3 67da5524c9281859c7897f39 cama_mesa_banho bed_bath_table
4 67da5524c9281859c7897f3a moveis_decoracao furniture_decor
66 67da5524c9281859c7897f78 flores flowers
67 67da5524c9281859c7897f79 artes_e_artesanato arts_and_craftmanship
68 67da5524c9281859c7897f7a fraldas_higiene diapers_and_hygiene
69 67da5524c9281859c7897f7b fashion_roupa_infanto_juvenil fashion_childrens_clothes
70 67da5524c9281859c7897f7c seguros_e_servicos security_and_services
71 rows × 3 columns
11
Python
Start typing or generate with AI (Ctrl + I)...


adb-3220087795989142.2.azuredatabricks.net/editor/notebooks/2469449315193760?o=3220087795989142
Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
olist-spark-workspace
Untitled Notebook 2025-03-19 11:03:54
Python
File Edit View Run Help Last edit was now
Run all
Dinesh km's Cluster
Schedule
Share
Cleaning the data
3 minutes ago (<1s)
12
from pyspark.sql.functions import col,to_date,datediff,current_date
Just now (2s)
13
def clean_dataframe(df,name):
print("Cleaning "+name)
return df.dropDuplicates().na.drop('all')
orders_df=clean_dataframe(orders_df,"Orders")
display(orders_df)
(2) Spark Jobs
orders_df: pyspark.sql.dataframe.DataFrame = [order_id: string, customer_id: string ... 6 more fields]
Cleaning Orders
Table
order_id
customer_id
order_status
order_purchase_timestamp
order_approved_at
4 6f841dde94727854eaff3f66432c80ea a9c9532060c9d245f06526c633d2dfba delivered 2018-01-02T19:20:35.000+00:00 2018-01-02T19:32:22.000+00:...
5 ca290a06ee0945b956f79c93b5191633 acd575d738296889f41a5a3f37510dd delivered 2018-04-09T22:02:30.000+00:00 2018-04-09T23:10:27.000+00:...
6 3735720b8c1cd4212fa0866a8e58a049 94af59d9cac1ae19763125 84f628ef6f delivered 2018-05-16T23:08:51.000+00:00 2018-05-16T23:55:11.000+00:...
7 8b346abc34a6e64bc67fcd3b0eccdc9f dd1506d329d9b0135855082aae3c79ac delivered 2018-07-11T20:15:04.000+00:00 2018-07-11T20:26:13.000+00:...
8 a4d7c8bca45b56444e3c59ddcca7d7c9 fcdb673d0f8d2b84d3862193f17a08f9 delivered 2017-08-29T12:36:36.000+00:00 2017-08-30T02:05:44.000+00:...

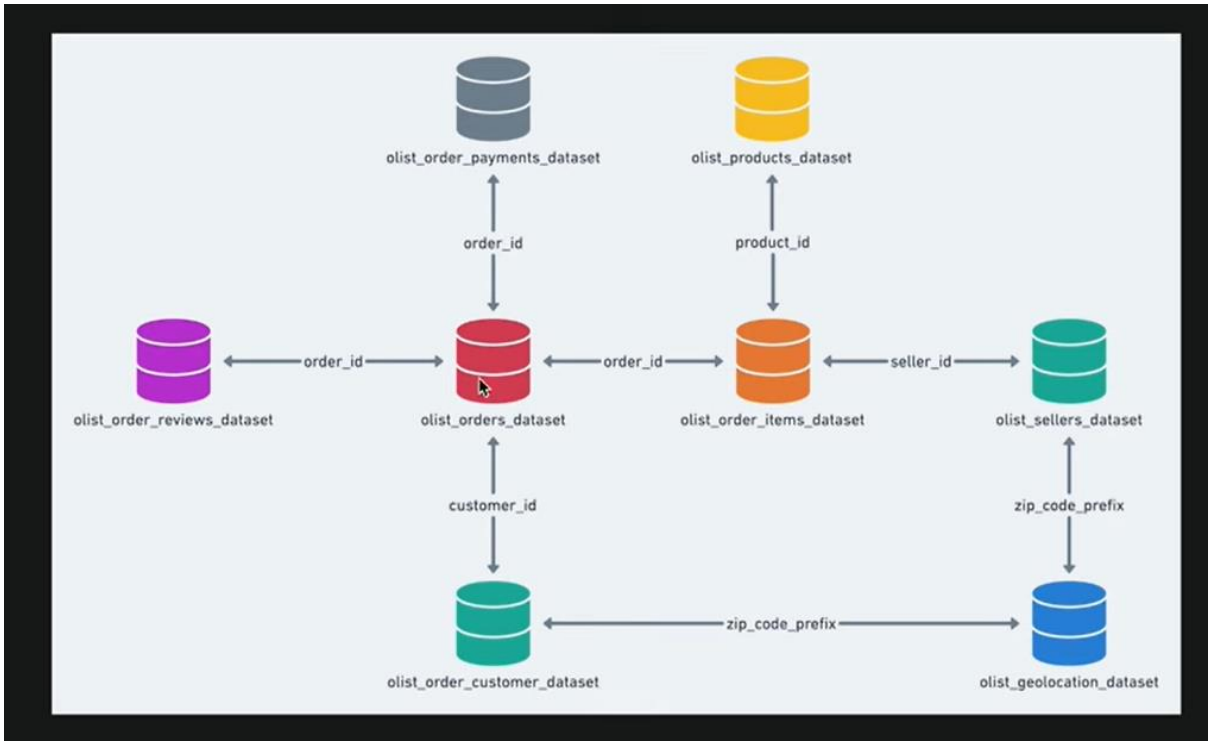


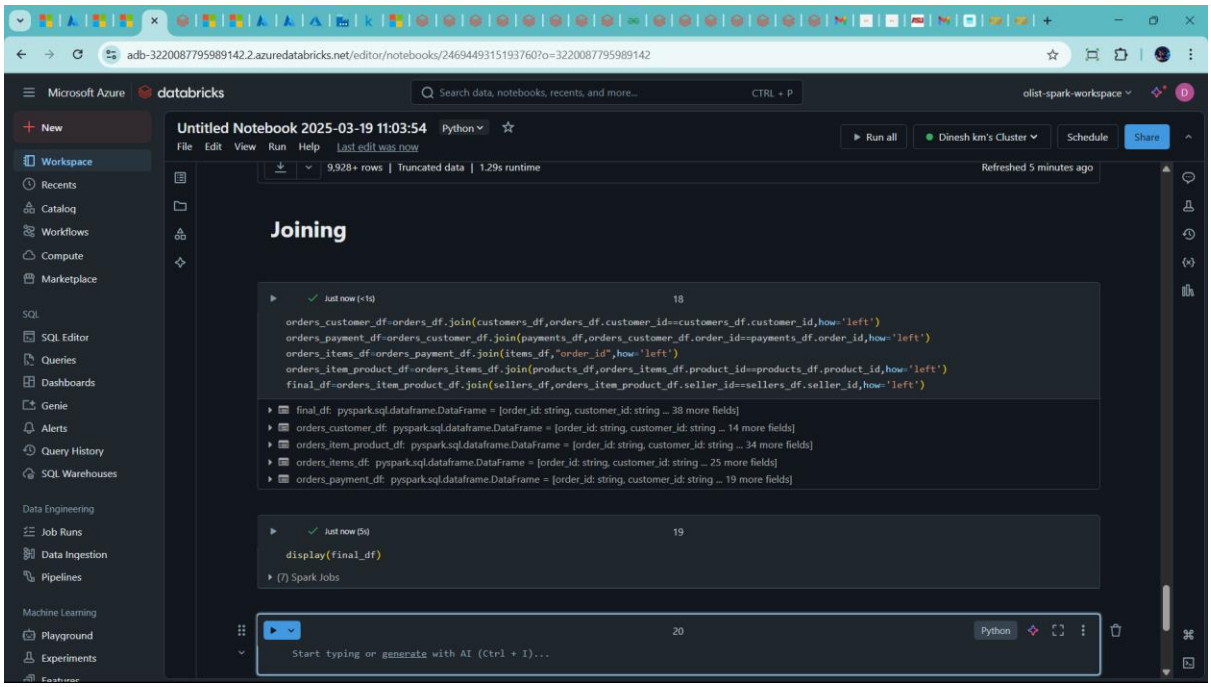


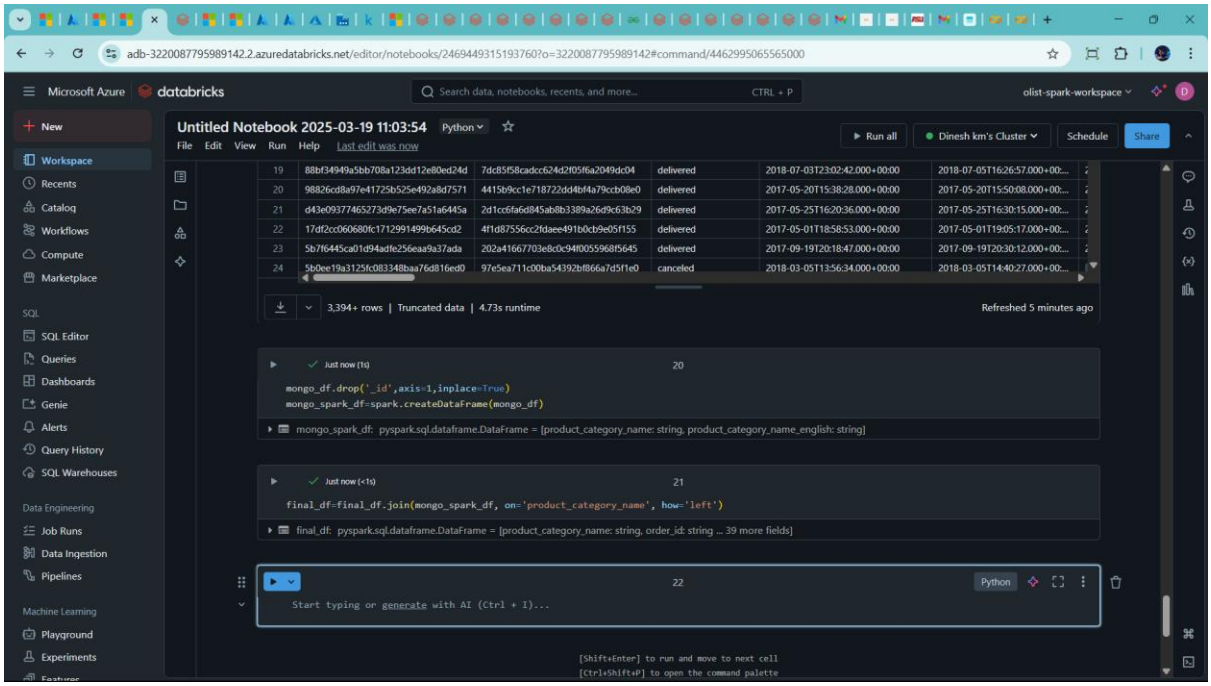
	order_delivered_customer_date	order_estimated_delivery_date	actual_delivery_time	estimated_delivery_time	delay
1	2018-06-16T15:20:55.000+00:00	2018-07-18T00:00:00.000+00:00	12	44	0
2	2018-03-09T21:52:36.000+00:00	2018-03-09T00:00:00.000+00:00	23	23	0
3	null	2018-02-07T00:00:00.000+00:00	null	28	0
4	2018-01-27T14:27:59.000+00:00	2018-02-05T00:00:00.000+00:00	25	34	0
5	2018-04-25T13:33:24.000+00:00	2018-05-10T00:00:00.000+00:00	16	31	0
6	2018-05-21T20:50:54.000+00:00	2018-06-20T00:00:00.000+00:00	5	35	0
7	2018-07-26T23:32:50.000+00:00	2018-08-09T00:00:00.000+00:00	15	29	0
8	2017-09-08T21:36:28.000+00:00	2017-09-22T00:00:00.000+00:00	10	24	0
9	2017-05-24T15:06:39.000+00:00	2017-06-21T00:00:00.000+00:00	18	46	0
10	2018-05-28T16:46:42.000+00:00	2018-06-13T00:00:00.000+00:00	4	20	0
11	2017-04-04T15:49:42.000+00:00	2017-04-18T00:00:00.000+00:00	15	29	0


Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
olist-spark-workspace
Untitled Notebook 2025-03-19 11:03:54
Python
File Edit View Run Help Last edit was now
Run all
Dinesh km's Cluster
Schedule
Share
New
Workspace
Recents
Catalog
Workflows
Compute
Marketplace
SQL
SQL Editor
Queries
Dashboards
Genie
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Pipelines
Machine Learning
Playground
Experiments
Just now (1s)
16
Python
#Calculate delivery and time delays
orders_df = orders_df.withColumn("actual_delivery_time",datediff("order_delivered_customer_date", "order_purchase_timestamp"))
orders_df=orders_df.withColumn("estimated_delivery_time",datediff('order_estimated_delivery_date','order_purchase_timestamp'))
orders_df=orders_df.withColumn("Delay Time",col("actual_delivery_time")-col("estimated_delivery_time"))
display(orders_df)
(2) Spark Jobs
orders_df: pyspark.sql.dataframe.DataFrame = [order_id: string, customer_id: string ... 10 more fields]
Table
order_estimated_delivery_date
actual_delivery_time
estimated_delivery_time
delay
Delay Time
2018-07-18T00:00:00.000+00:00 12 44 false -32
2018-03-09T00:00:00.000+00:00 23 23 false 0
2018-02-07T00:00:00.000+00:00 null 28 null null
2018-02-05T00:00:00.000+00:00 25 34 false -9
2018-05-10T00:00:00.000+00:00 16 31 false -15
2018-06-20T00:00:00.000+00:00 5 35 false -30
2018-08-09T00:00:00.000+00:00 15 29 false -14
2017-09-22T00:00:00.000+00:00 10 24 false -14
2017-06-21T00:00:00.000+00:00 18 46 false -28
2018-06-13T00:00:00.000+00:00 4 20 false -16
2017-04-18T00:00:00.000+00:00 15 29 false -14
2018-08-17T00:00:00.000+00:00 6 10 false -4
2018-02-15T00:00:00.000+00:00 2 16 false -14
2018-08-09T00:00:00.000+00:00 8 36 false -28

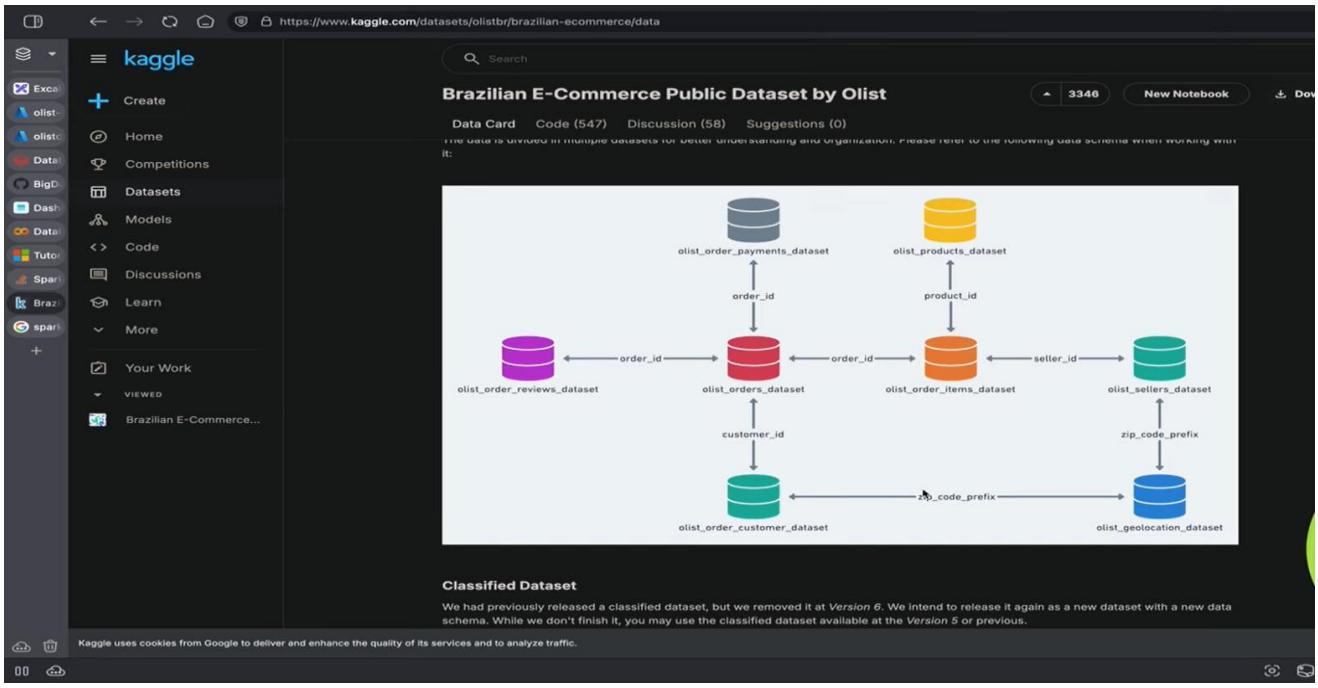
→Joining

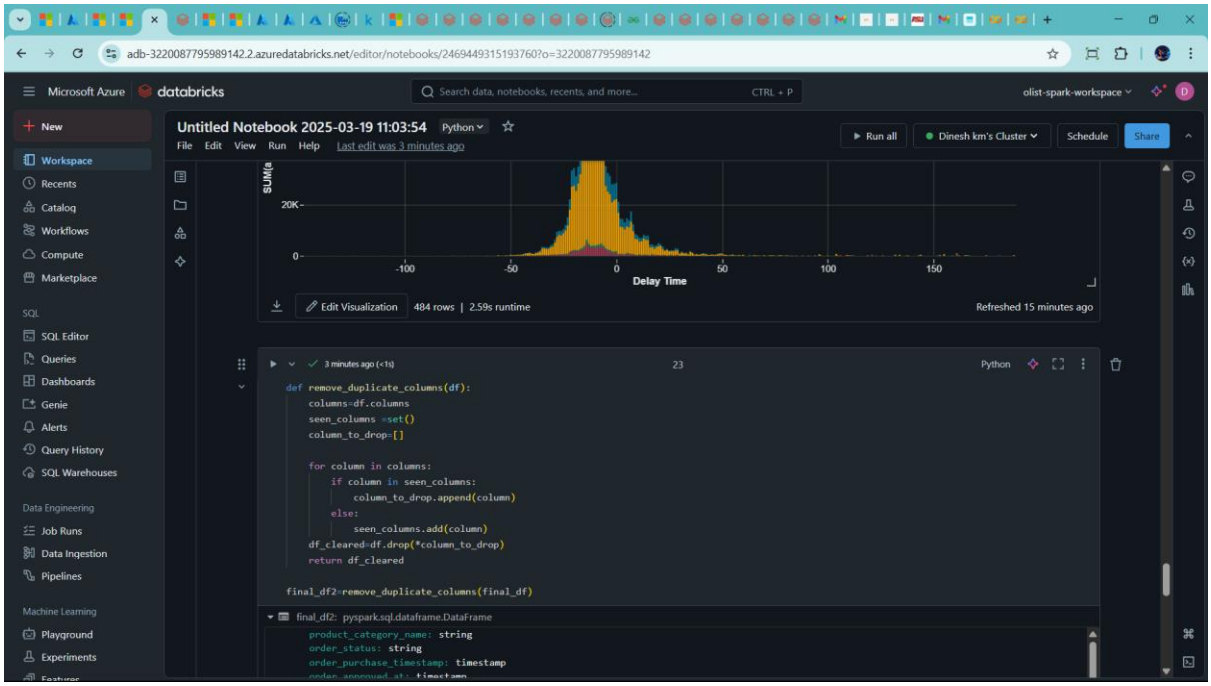

olist_order_payments_dataset
olist_products_dataset
order_id
product_id
order_id
order_id
seller_id
olist_order_reviews_dataset
olist_orders_dataset
olist_order_items_dataset
olist_sellers_dataset
customer_id
zip_code_prefix
zip_code_prefix
olist_order_customer_dataset
olist_geolocation_dataset

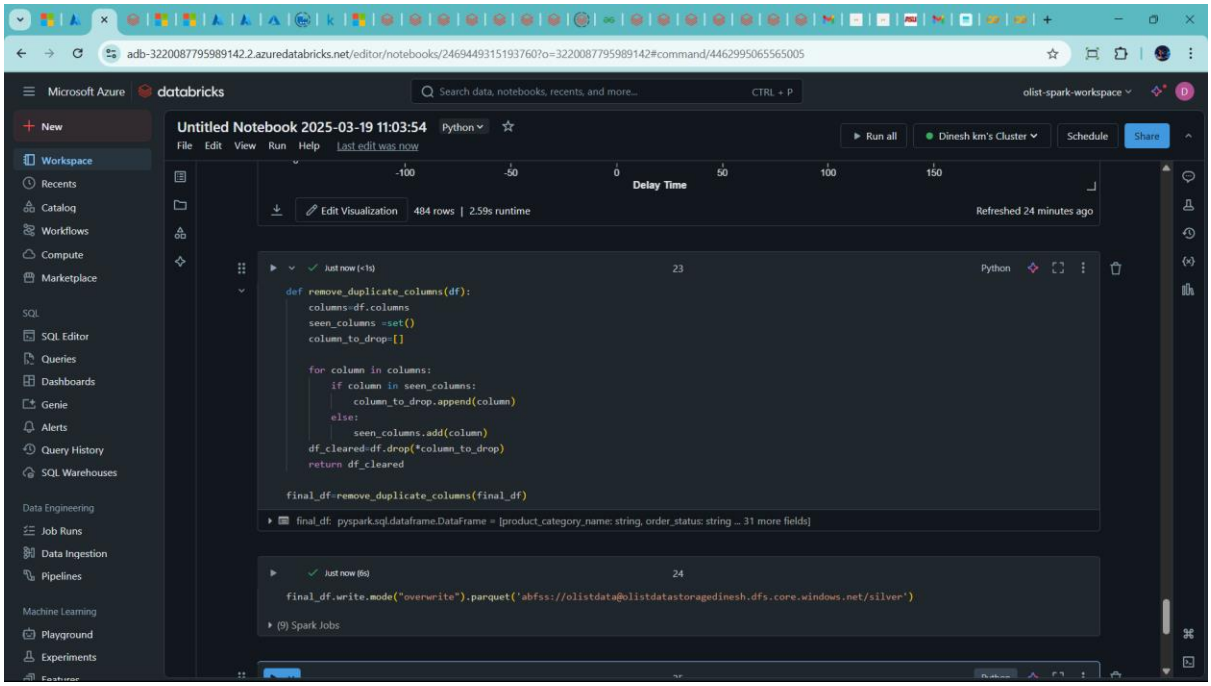

adb-3220087795989142.2.azuredatabricks.net/editor/notebooks/2469449315193760?o=3220087795989142
Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
olist-spark-workspace
New
Workspace
Recents
Catalog
Workflows
Compute
Marketplace
SQL
SQL Editor
Queries
Dashboards
Genie
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Pipelines
Machine Learning
Playground
Experiments
Untitled Notebook 2025-03-19 11:03:54
Python
File Edit View Run Help Last edit was now
Run all
Dinesh km's Cluster
Schedule
Share
9,928+ rows | Truncated data | 1.29s runtime
Refreshed 5 minutes ago
Joining
Just now (<1s)
18
orders_customer_df=orders_df.join(customers_df,orders_df.customer_id==customers_df.customer_id,how='left')
orders_payment_df=orders_customer_df.join(payments_df,orders_customer_df.order_id==payments_df.order_id,how='left')
orders_items_df=orders_payment_df.join(items_df,"order_id",how='left')
orders_item_product_df=orders_items_df.join(products_df,orders_items_df.product_id==products_df.product_id,how='left')
final_df=orders_item_product_df.join(sellers_df,orders_item_product_df.seller_id==sellers_df.seller_id,how='left')
final_df: pyspark.sql.dataframe.DataFrame = [order_id: string, customer_id: string ... 38 more fields]
orders_customer_df: pyspark.sql.dataframe.DataFrame = [order_id: string, customer_id: string ... 14 more fields]
orders_item_product_df: pyspark.sql.dataframe.DataFrame = [order_id: string, customer_id: string ... 34 more fields]
orders_items_df: pyspark.sql.dataframe.DataFrame = [order_id: string, customer_id: string ... 25 more fields]
orders_payment_df: pyspark.sql.dataframe.DataFrame = [order_id: string, customer_id: string ... 19 more fields]
Just now (5s)
19
display(final_df)
(7) Spark Jobs
20
Python
Start typing or generate with AI (Ctrl + I)...


adb-3220087795989142.2.azuredatabricks.net/editor/notebooks/2469449315193760?o=3220087795989142#command/4462995065565000
Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
olist-spark-workspace
Untitled Notebook 2025-03-19 11:03:54
Python
File Edit View Run Help Last edit was now
Run all
Dinesh km's Cluster
Schedule
Share
19 88bf34949a5bb708a123dd12e80ed24d 7dc85f58cadcc624d2f05f6a2049dc04 delivered 2018-07-03T23:02:42.000+00:00 2018-07-05T16:26:57.000+00:...
20 98826cd8a97e41725b525e492a8d7571 4415b9cc1e718722dd4bf4a79ccb08e0 delivered 2017-05-20T15:38:28.000+00:00 2017-05-20T15:50:08.000+00:...
21 d43e09377465273d9e75ee7a51a6445a 2d1cc6fa6d845ab8b3389a26d9c63b29 delivered 2017-05-25T16:20:36.000+00:00 2017-05-25T16:30:15.000+00:...
22 17df2cc060680fc1712991499b645cd2 4f1d87556cc2fdaee491b0cb9e05f155 delivered 2017-05-01T18:58:53.000+00:00 2017-05-01T19:05:17.000+00:...
23 5b7f6445ca01d94adfe256eaa9a37ada 202a41667703e8c0c94f0055968f5645 delivered 2017-09-19T20:18:47.000+00:00 2017-09-19T20:30:12.000+00:...
24 5b0ee19a3125fc083348baa76d816ed0 97e5ea711c00ba54392bf866a7d5f1e0 canceled 2018-03-05T13:56:34.000+00:00 2018-03-05T14:40:27.000+00:...
3,394+ rows | Truncated data | 4.73s runtime
Refreshed 5 minutes ago
Just now (1s)
20
mongo_df.drop('_id',axis=1,inplace=True)
mongo_spark_df=spark.createDataFrame(mongo_df)
mongo_spark_df: pyspark.sql.dataframe.DataFrame = [product_category_name: string, product_category_name_english: string]
Just now (<1s)
21
final_df=final_df.join(mongo_spark_df, on='product_category_name', how='left')
final_df: pyspark.sql.dataframe.DataFrame = [product_category_name: string, order_id: string ... 39 more fields]
22
Python
Start typing or generate with AI (Ctrl + I)...
[Shift+Enter] to run and move to next cell
[Ctrl+Shift+P] to open the command palette

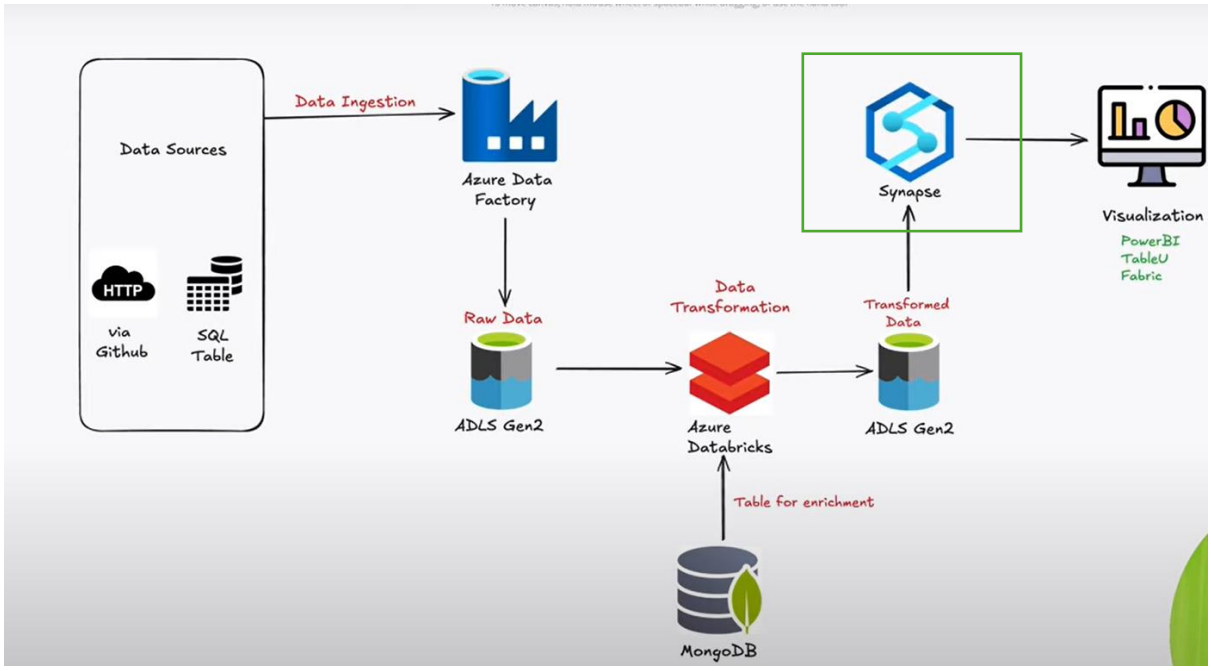
This is over dataset



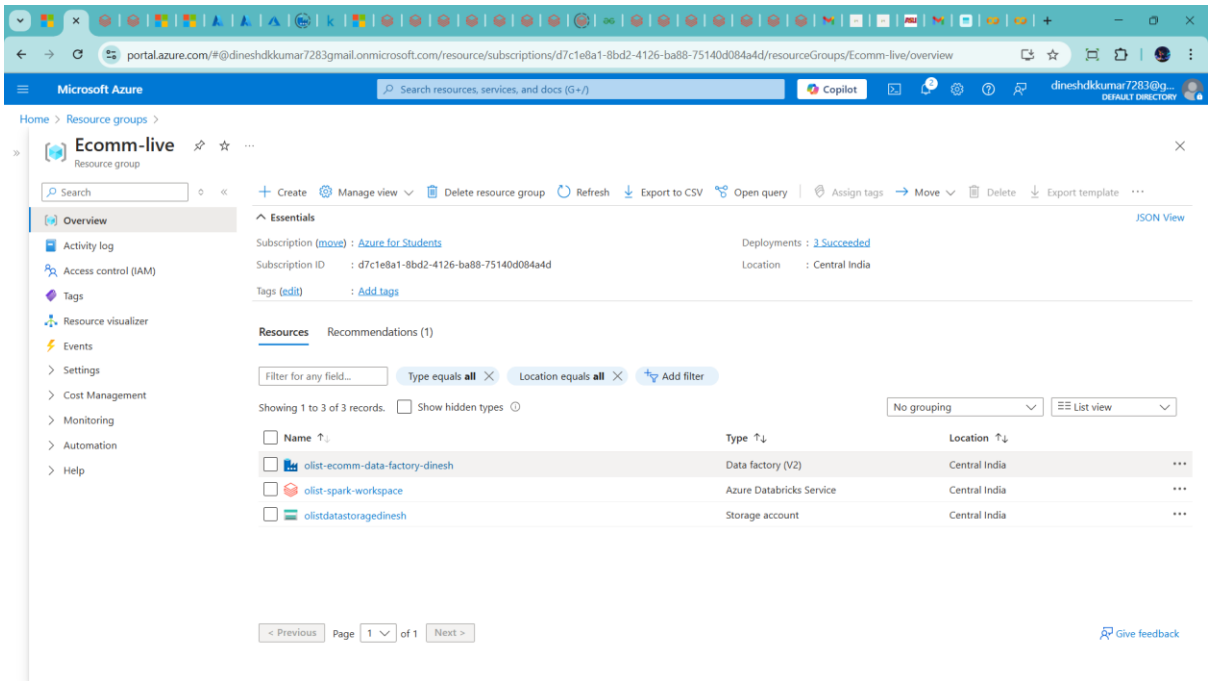

adb-3220087795989142.2.azuredatabricks.net/editor/notebooks/2469449315193760?o=3220087795989142
Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
olist-spark-workspace
New
Workspace
Recents
Catalog
Workflows
Compute
Marketplace
SQL
SQL Editor
Queries
Dashboards
Genie
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Pipelines
Machine Learning
Playground
Experiments
Untitled Notebook 2025-03-19 11:03:54
Python
File Edit View Run Help
Last edit was 3 minutes ago
Run all
Dinesh km's Cluster
Schedule
Share
SUM(a
20K
0
-100
-50
0
50
100
150
Delay Time
Edit Visualization
484 rows | 2.59s runtime
Refreshed 15 minutes ago
3 minutes ago (<1s)
23
Python
def remove_duplicate_columns(df):
columns=df.columns
seen_columns =set()
column_to_drop=[]
for column in columns:
if column in seen_columns:
column_to_drop.append(column)
else:
seen_columns.add(column)
df_cleared=df.drop(*column_to_drop)
return df_cleared
final_df2=remove_duplicate_columns(final_df)
final_df2: pyspark.sql.dataframe.DataFrame
product_category_name: string
order_status: string
order_purchase_timestamp: timestamp

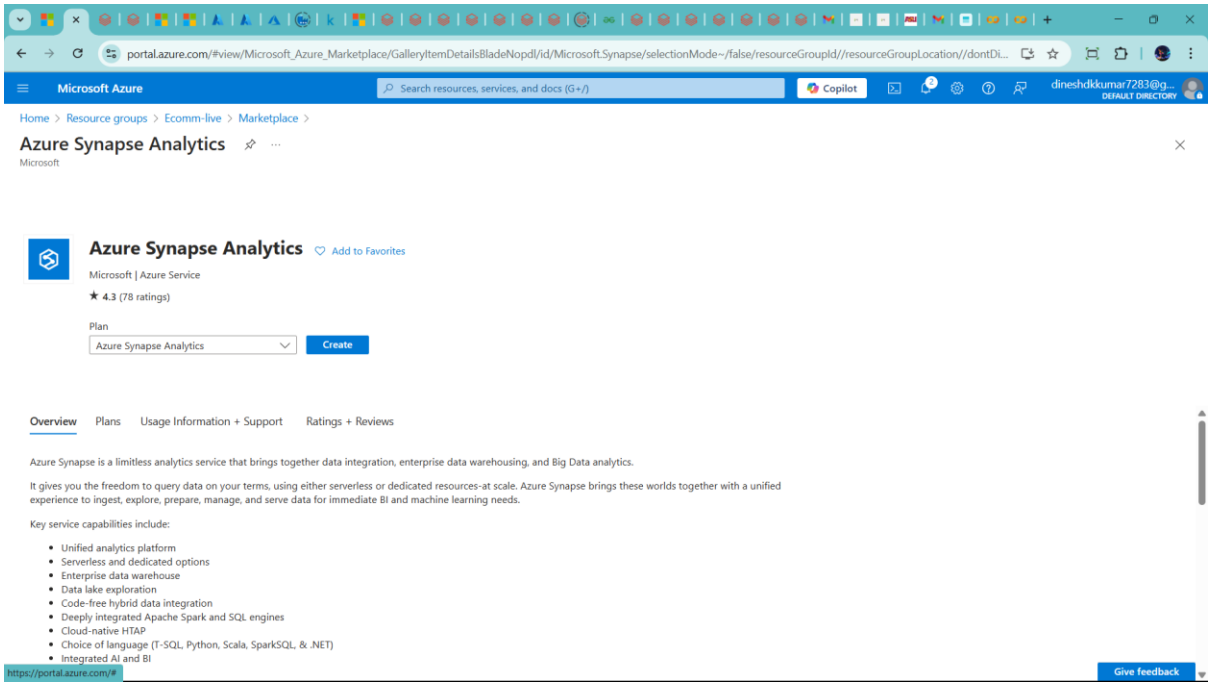

adb-3220087795989142.2.azuredatabricks.net/editor/notebooks/2469449315193760?o=3220087795989142#command/4462995065565005
Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
olist-spark-workspace
Untitled Notebook 2025-03-19 11:03:54
Python
File Edit View Run Help
Last edit was now
Run all
Dinesh km's Cluster
Schedule
Share
-100
-50
0
50
100
150
Delay Time
Edit Visualization
484 rows | 2.59s runtime
Refreshed 24 minutes ago
Just now (<1s)
23
Python
def remove_duplicate_columns(df):
columns=df.columns
seen_columns =set()
column_to_drop=[]
for column in columns:
if column in seen_columns:
column_to_drop.append(column)
else:
seen_columns.add(column)
df_cleared=df.drop(*column_to_drop)
return df_cleared
final_df=remove_duplicate_columns(final_df)
final_df: pyspark.sql.dataframe.DataFrame = [product_category_name: string, order_status: string ... 31 more fields]
Just now (6s)
24
final_df.write.mode("overwrite").parquet('abfss://olistdata@olistdatastoragedinesh.dfs.core.windows.net/silver')
(9) Spark Jobs

Synapse

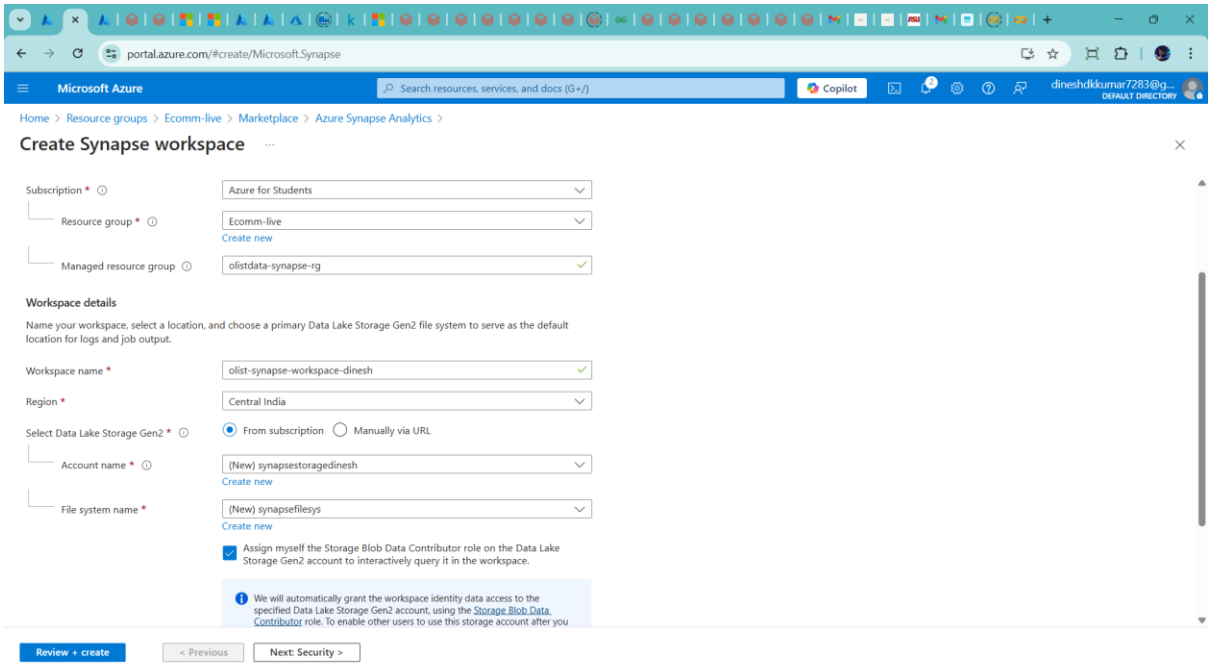


→create workspace synapse in azure

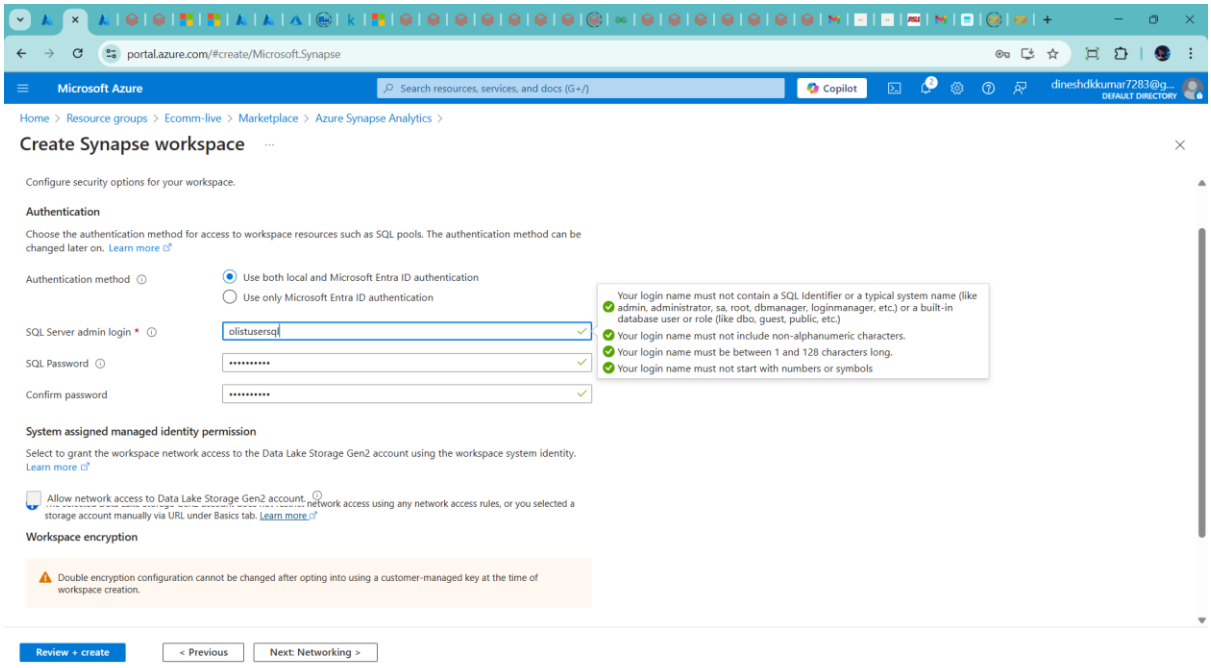




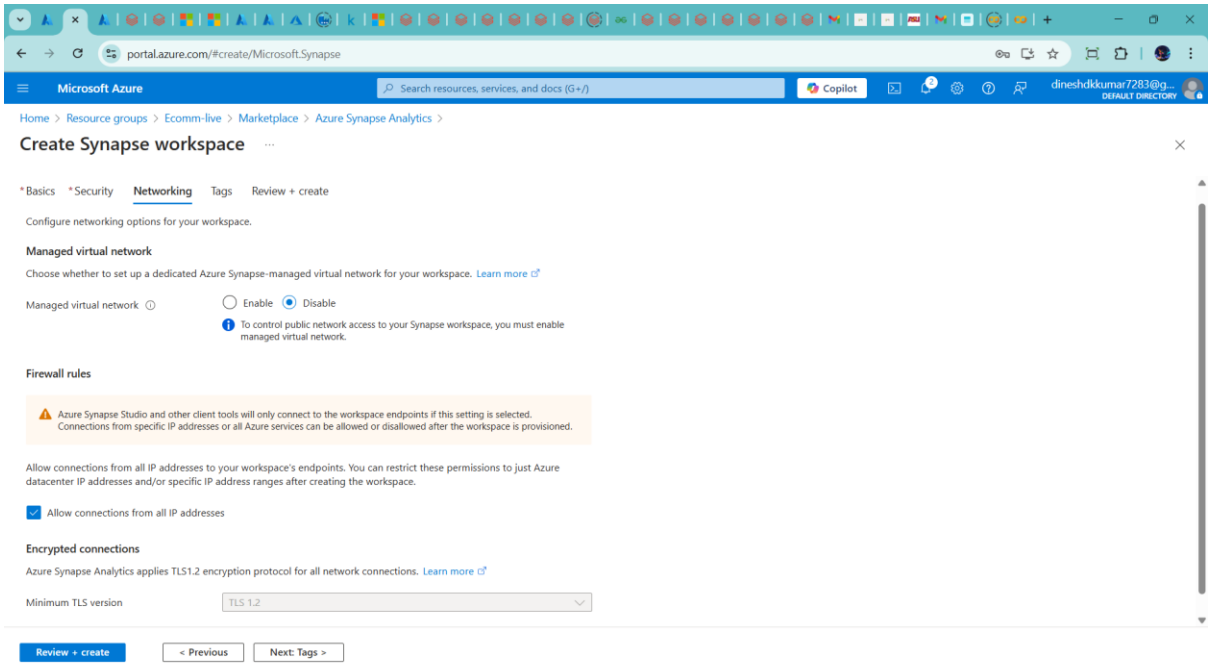
➔ synapse manage different storage taken



→click new

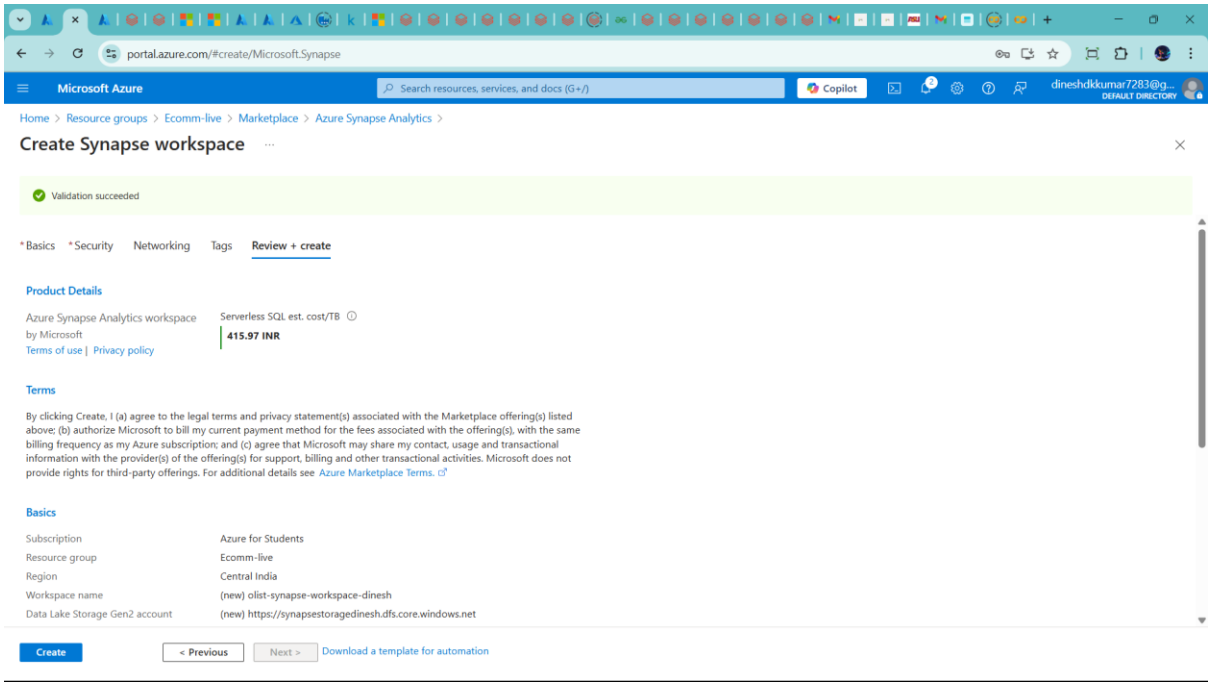


→click new

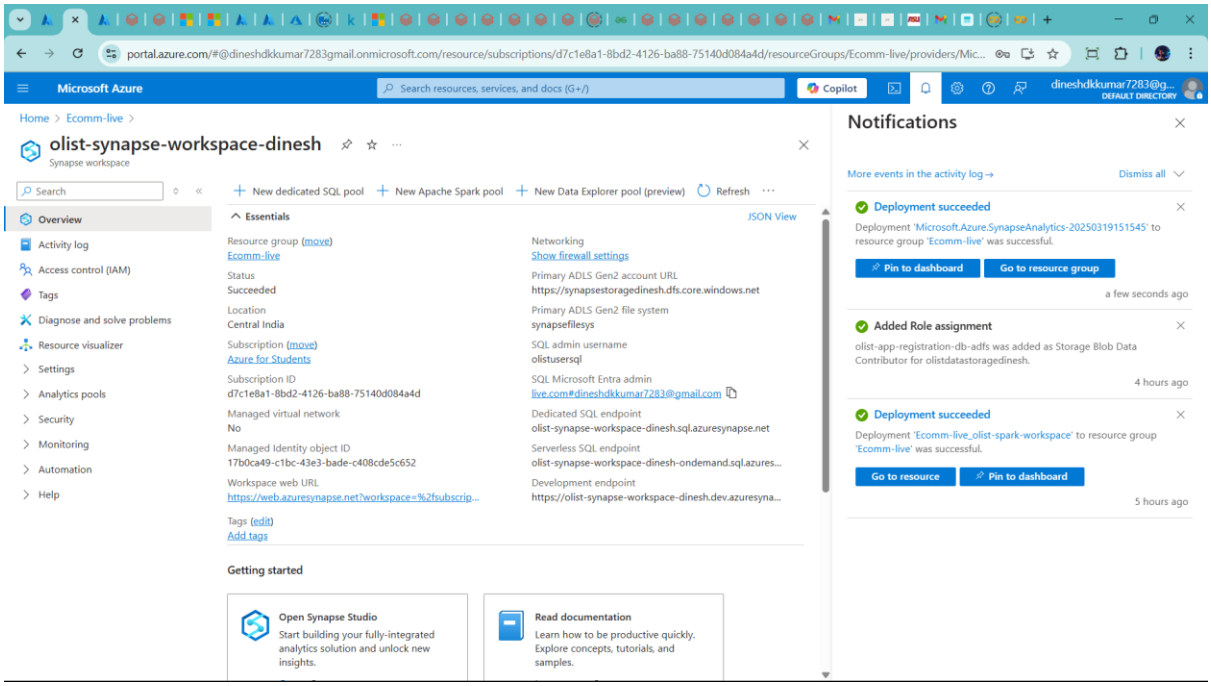


→if enable you have sensitive data

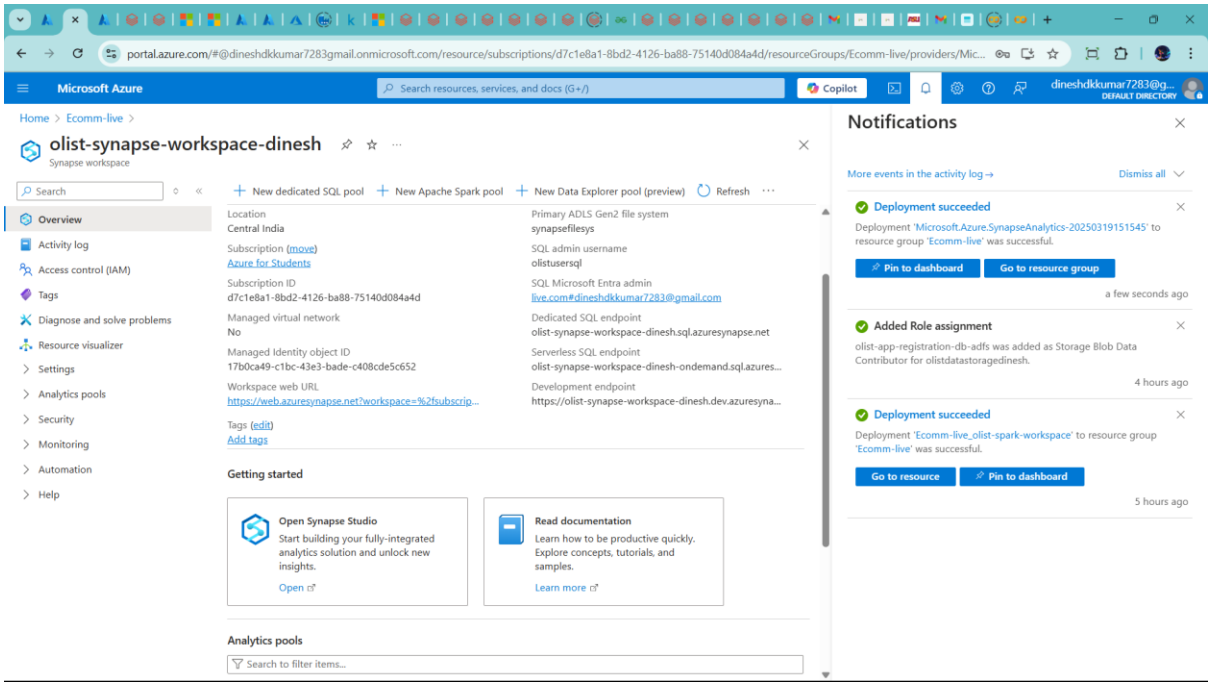
So by default →click new and review + create

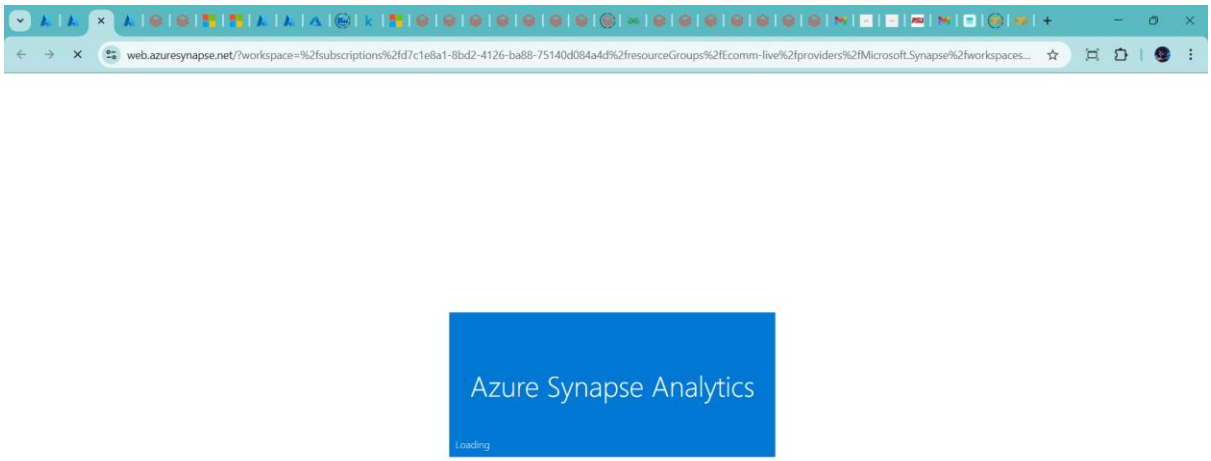


→complete

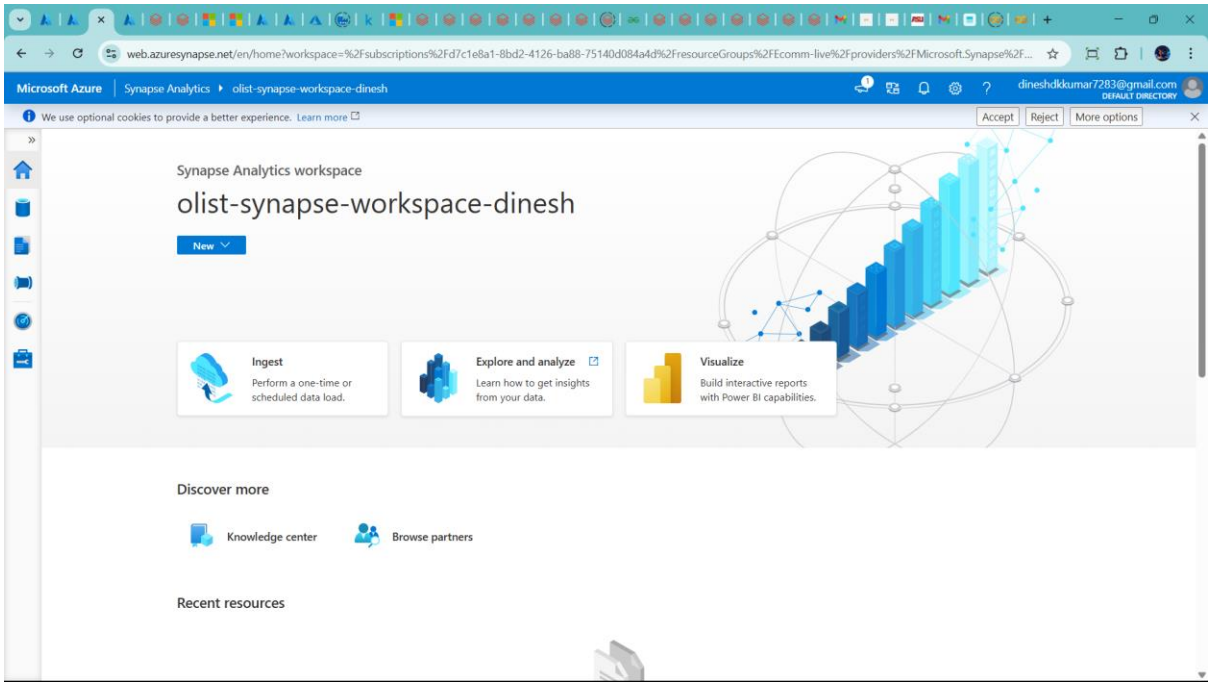


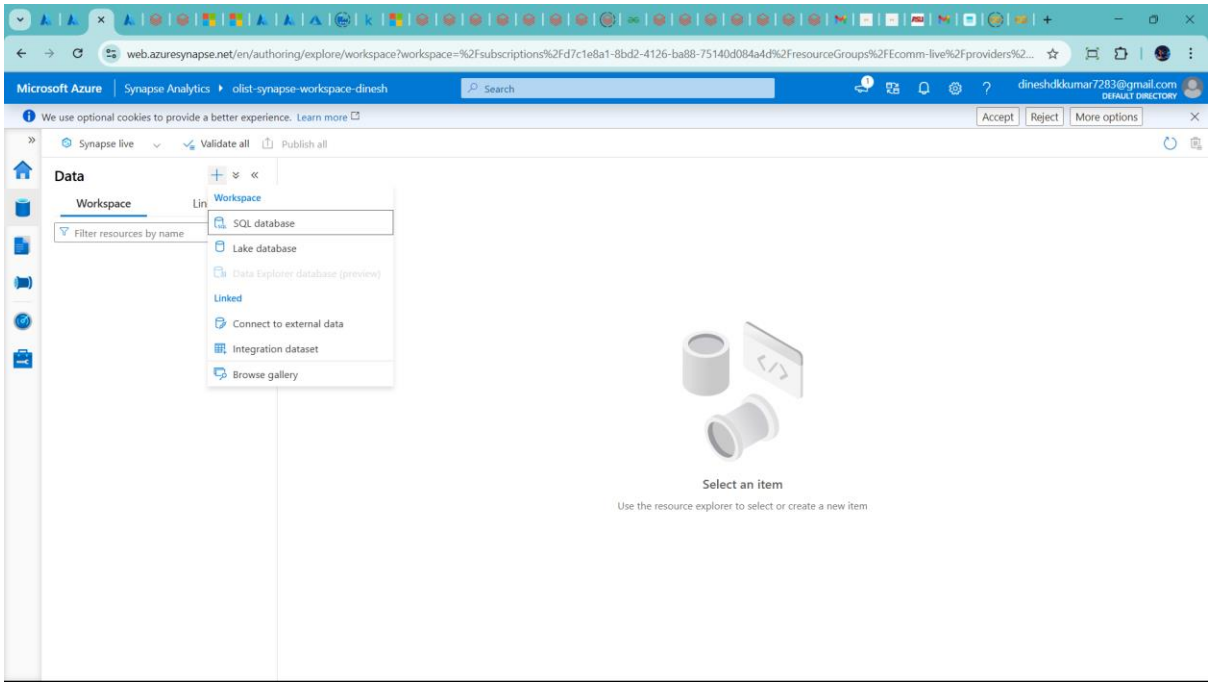
→click open



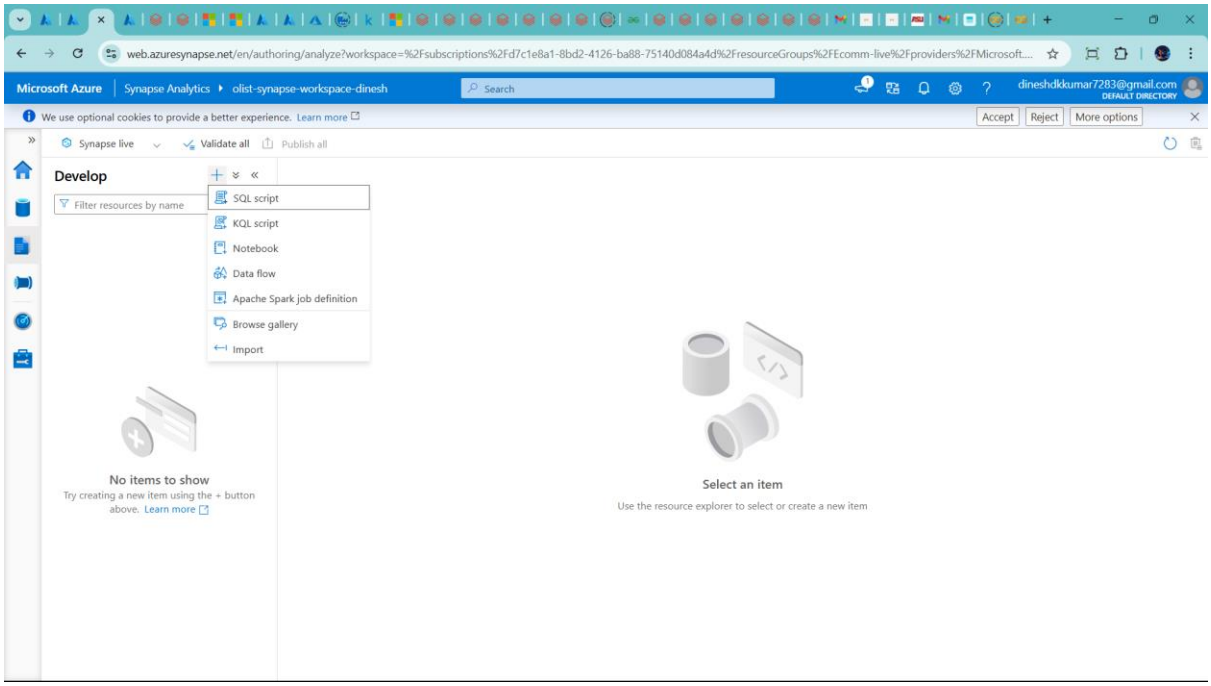


→it is synapse



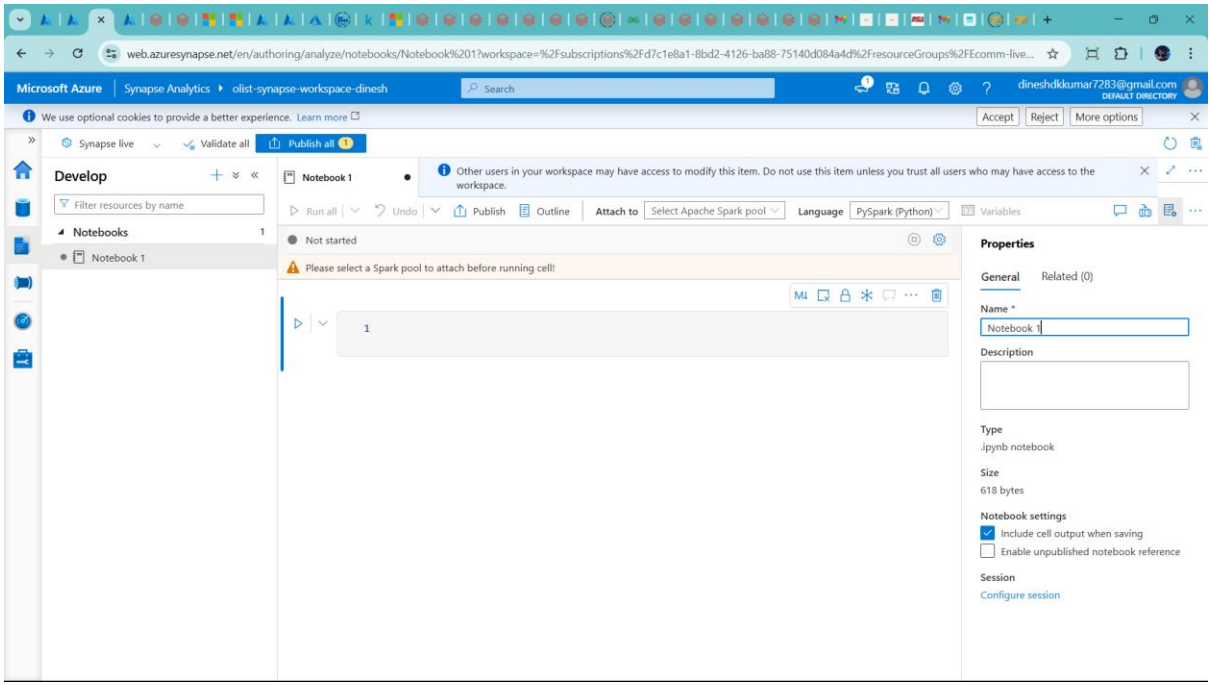


→lakedatabse → it is create by spark

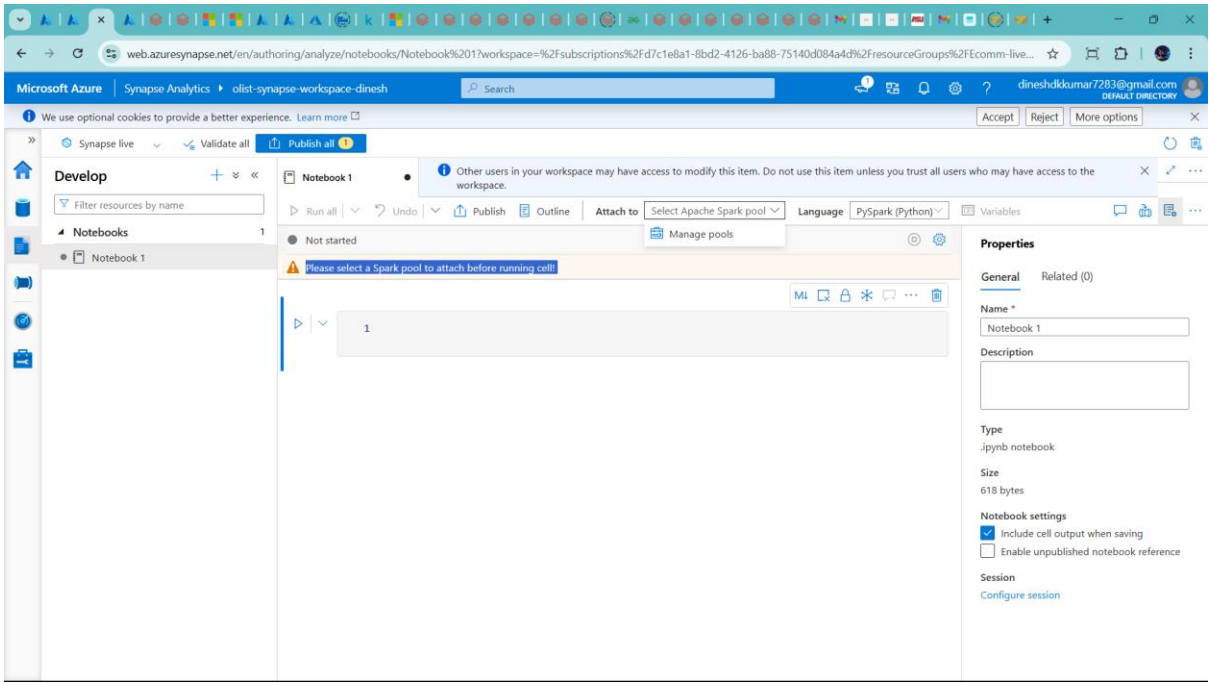


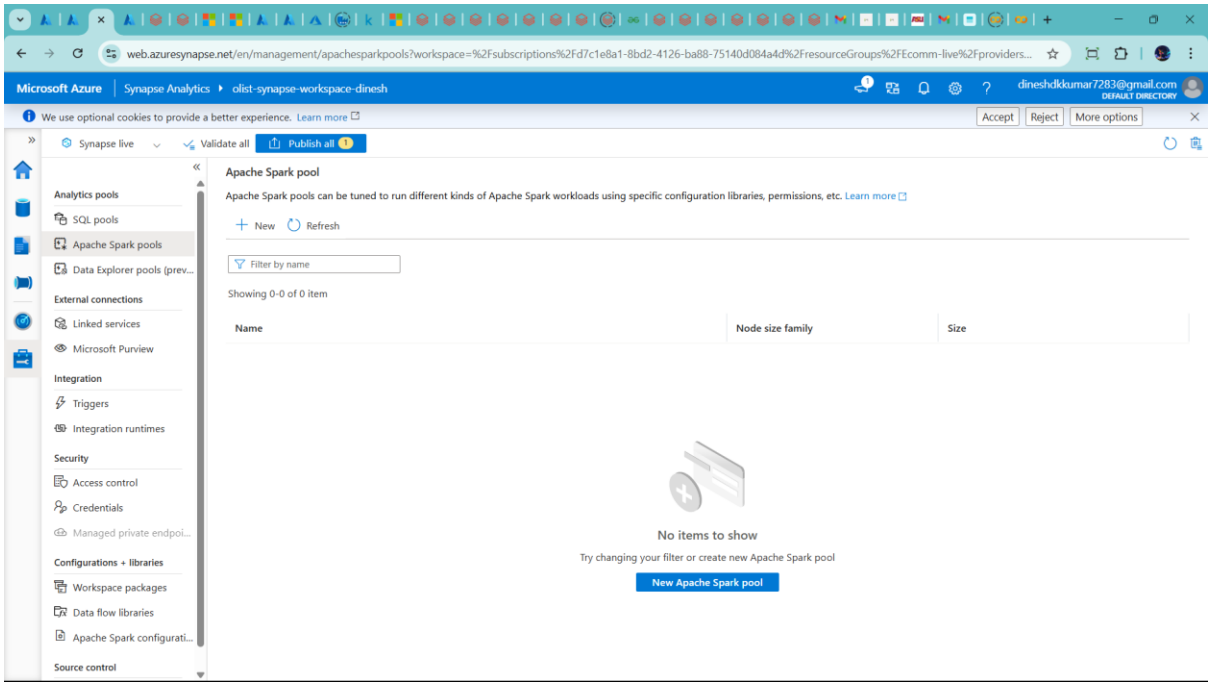
→develop →sql script,notebook

→click notebook



→click manage pool

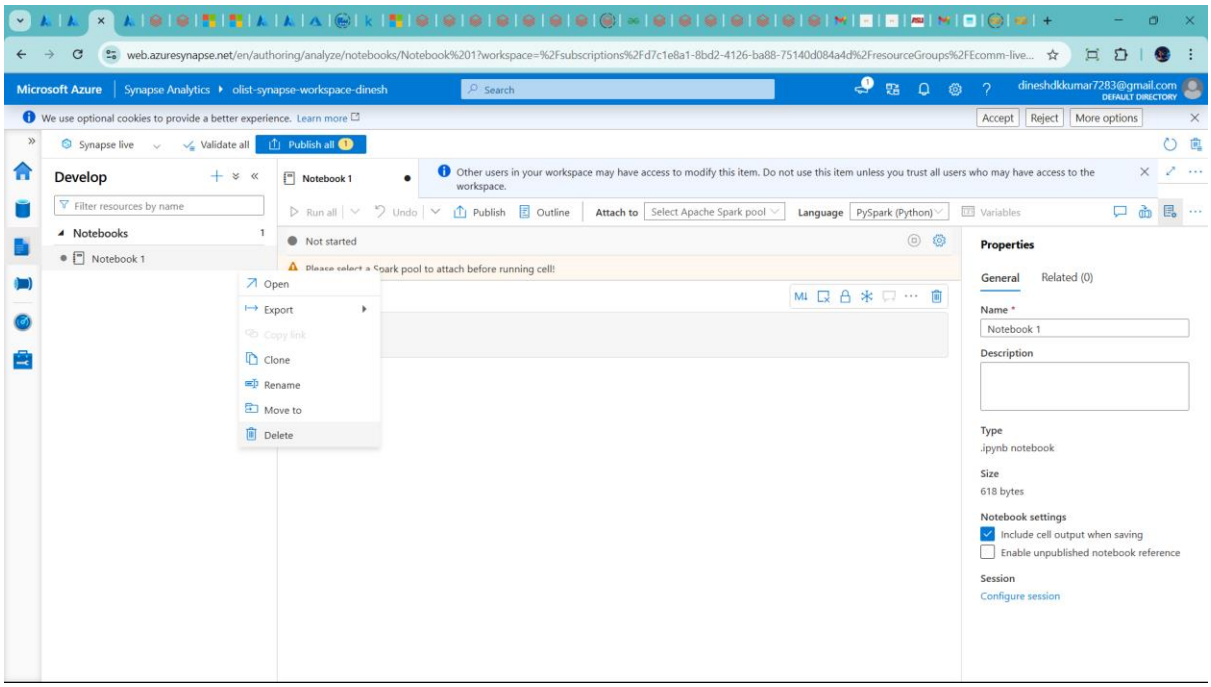


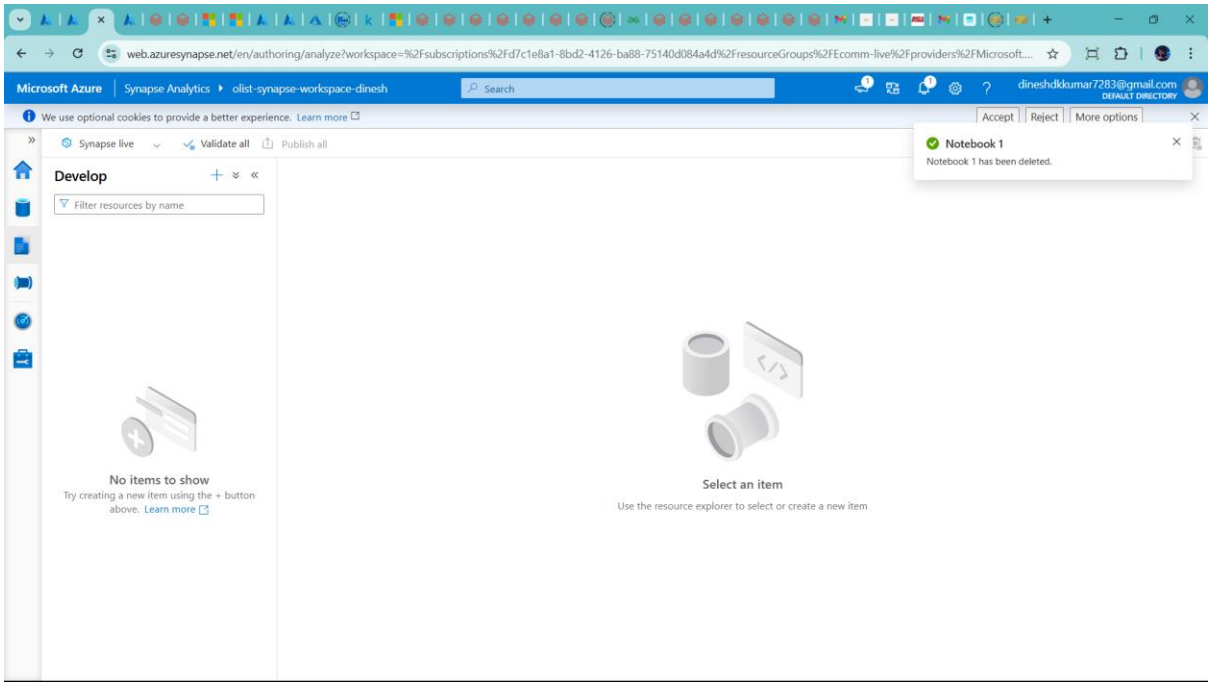


→click new apache spark pool

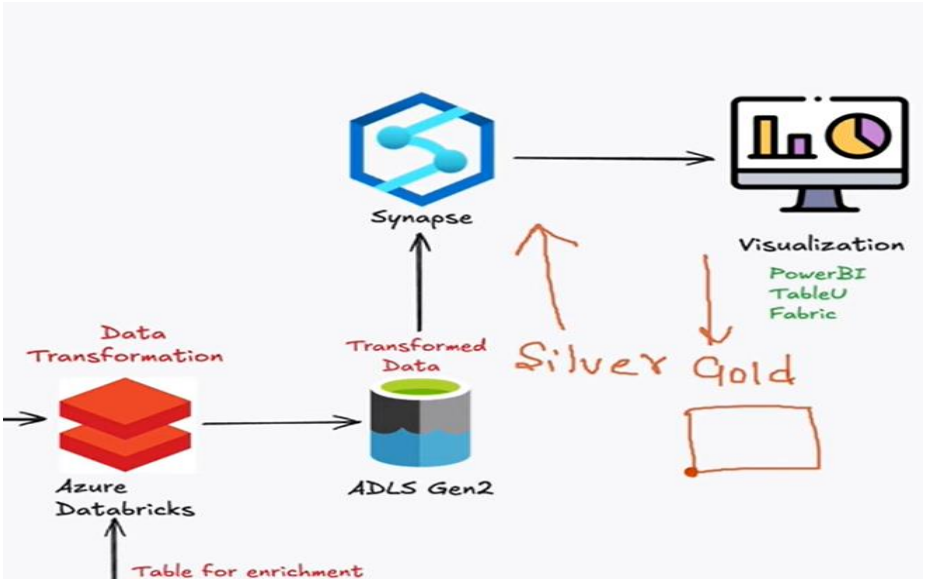
→explore concept only and cannot create anythink

→go back and delete notebook



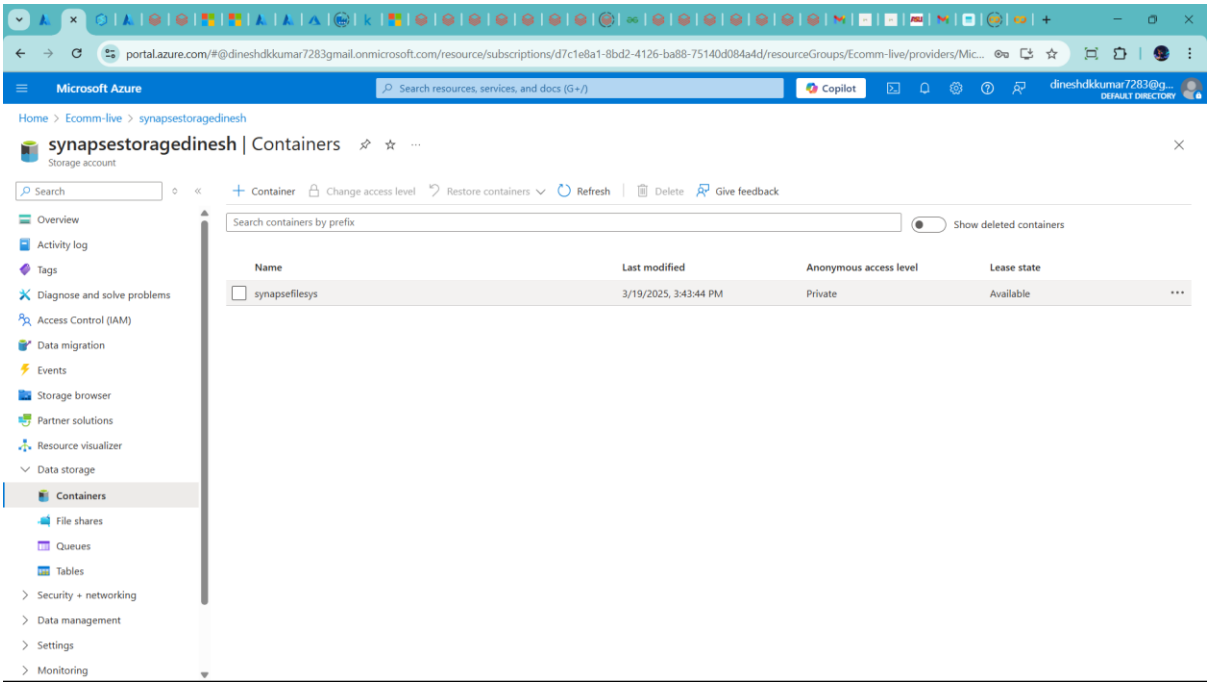


→buid gold layer using synapse

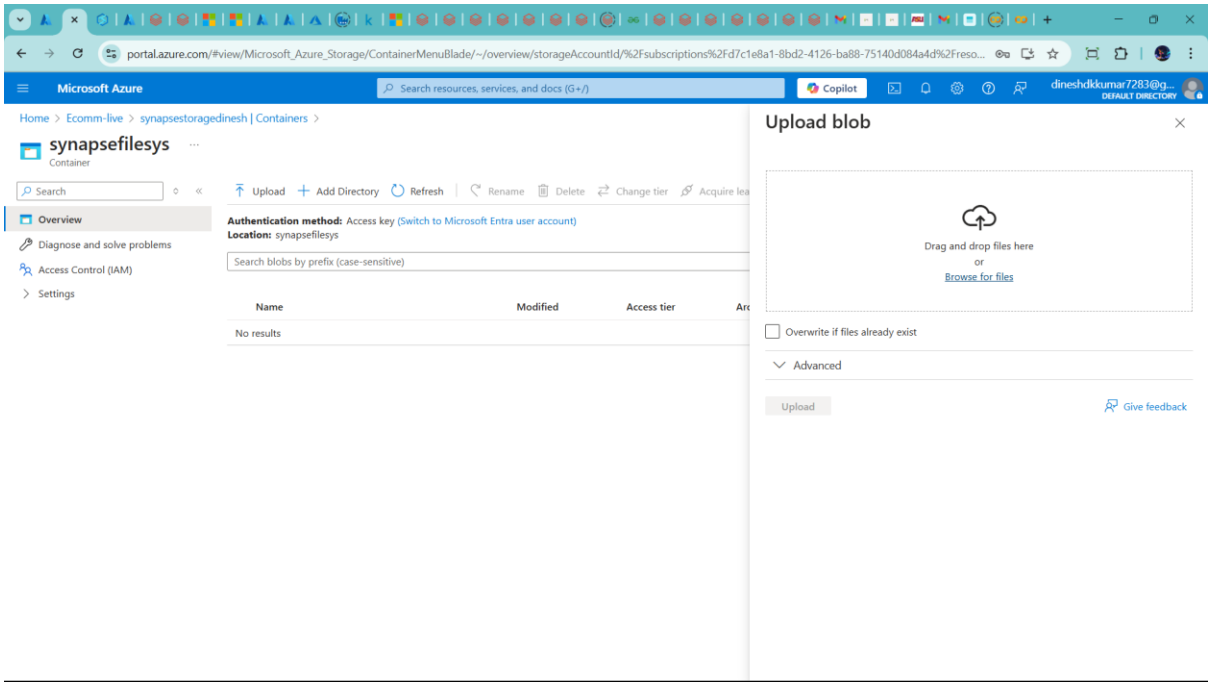


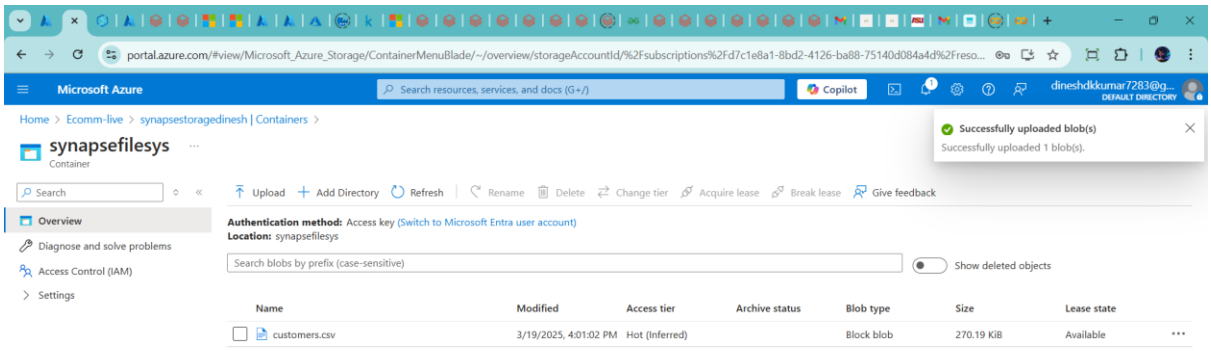
→validate

→go TO synapsetostorage container

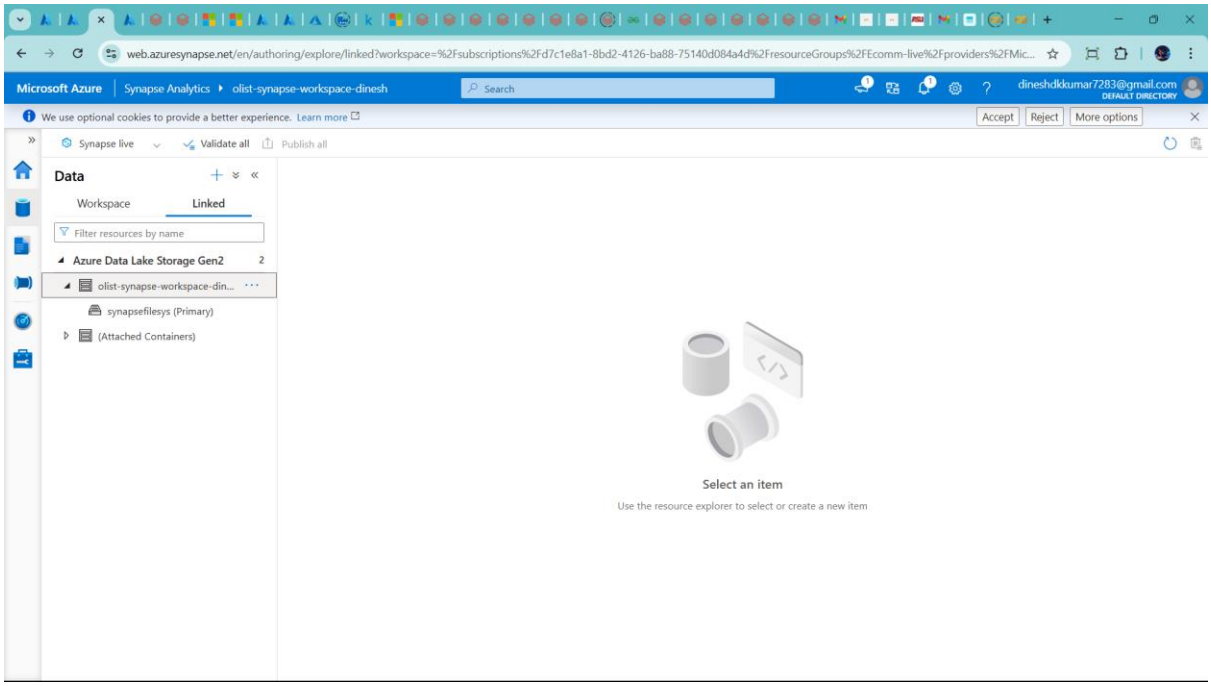


Click dynapsefilesys →click upload →add customer.csv file

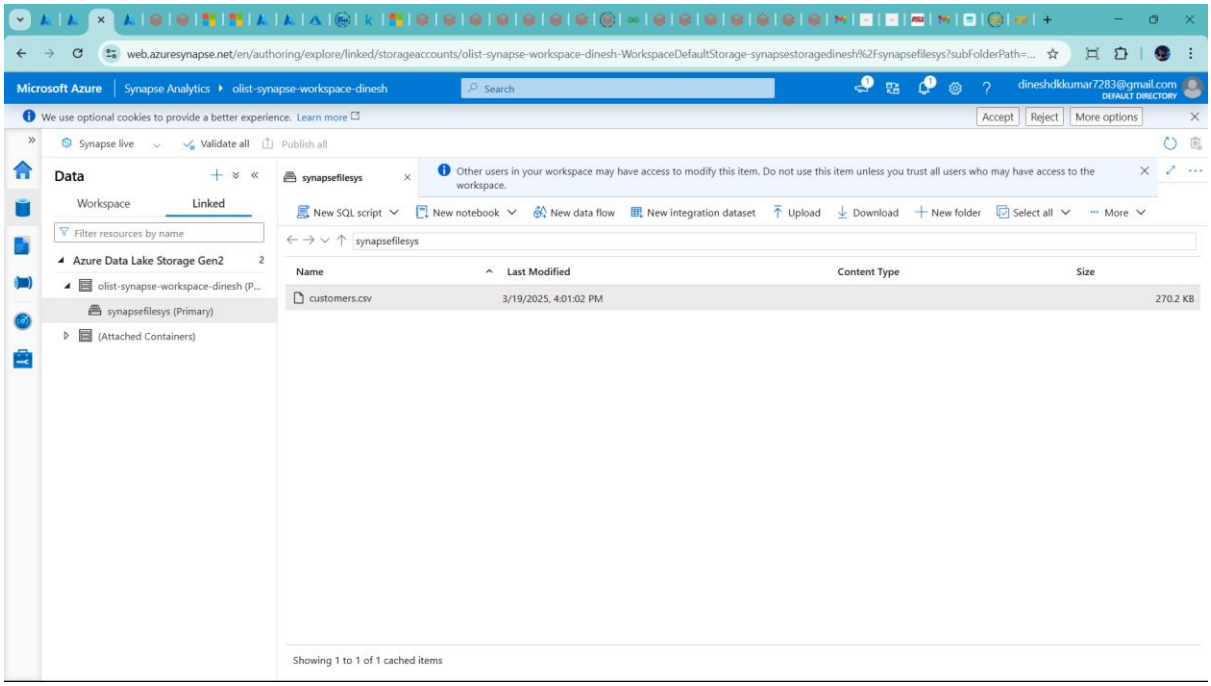




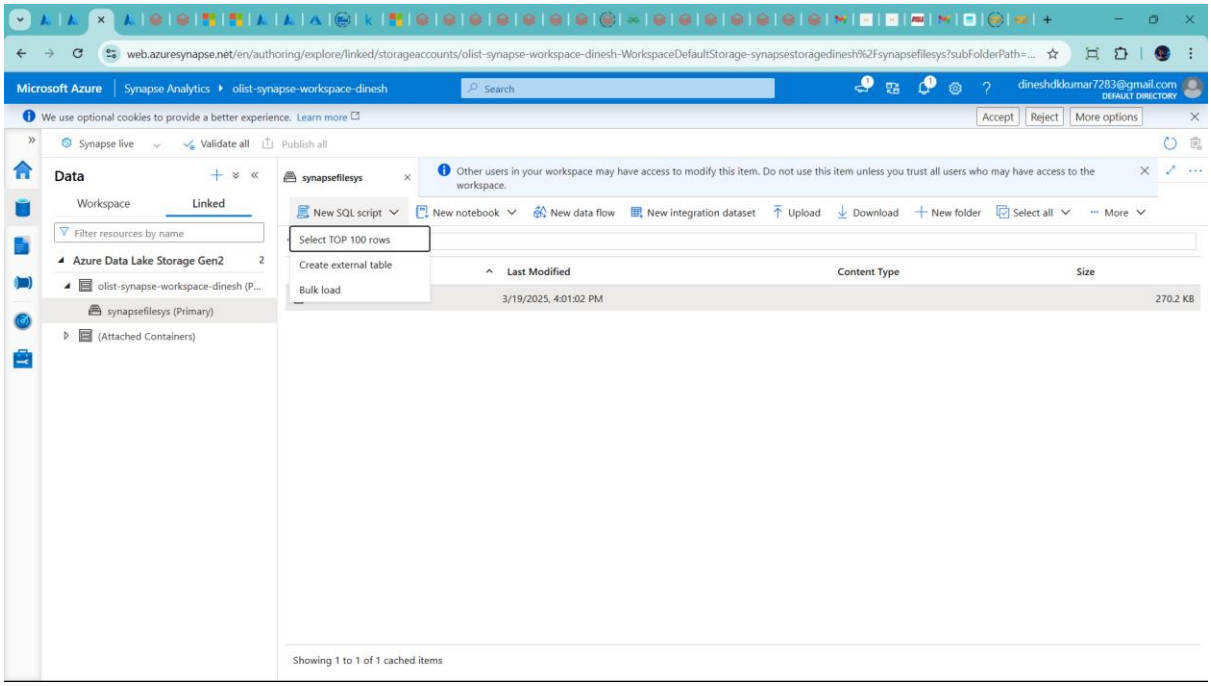
→go to check auzre synapse side→click data →click Linked



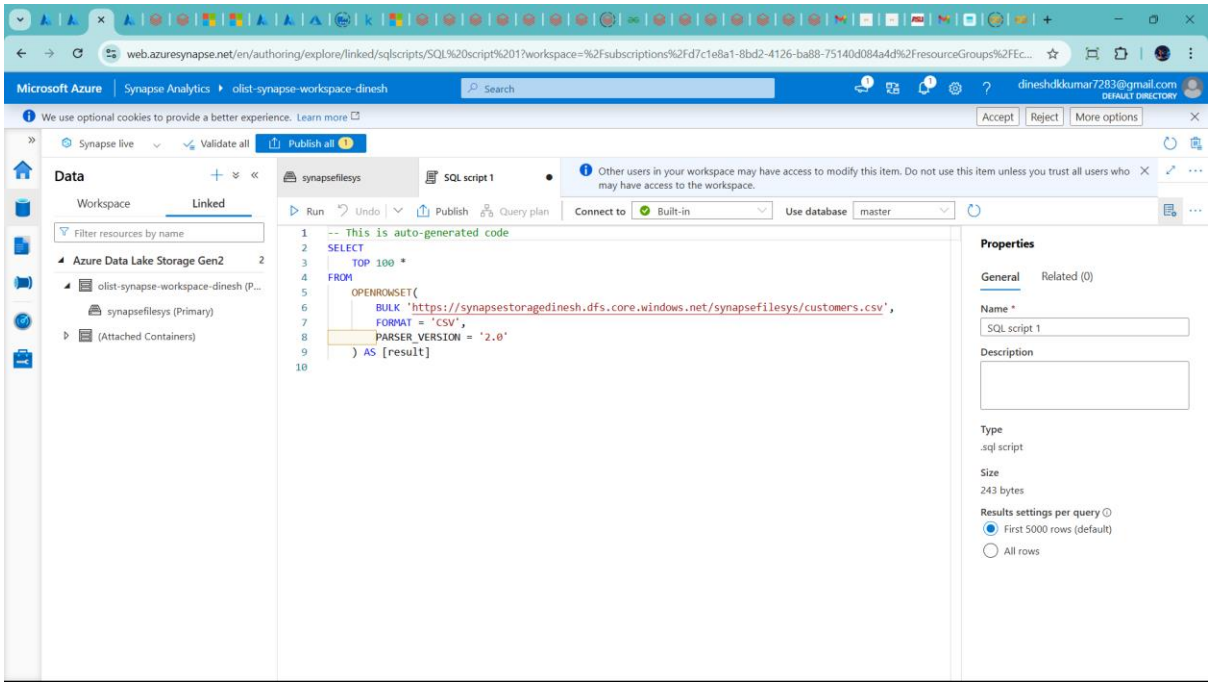
→click synapsefilessys



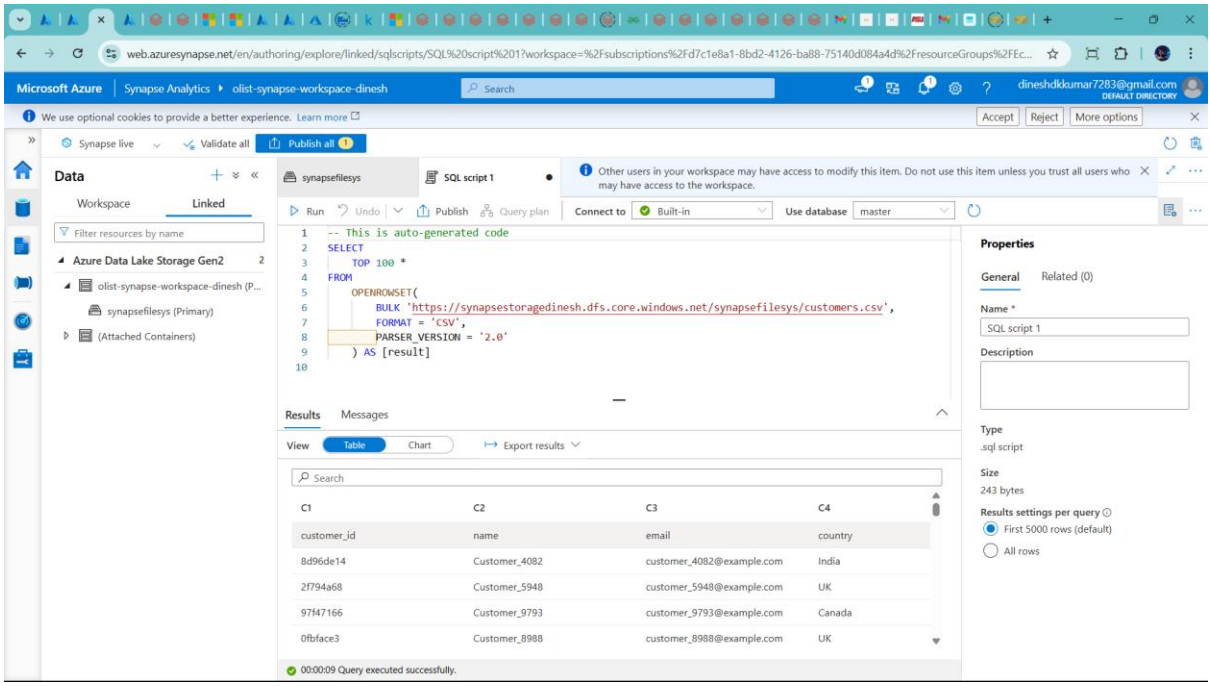
→click new SQL script →arrow



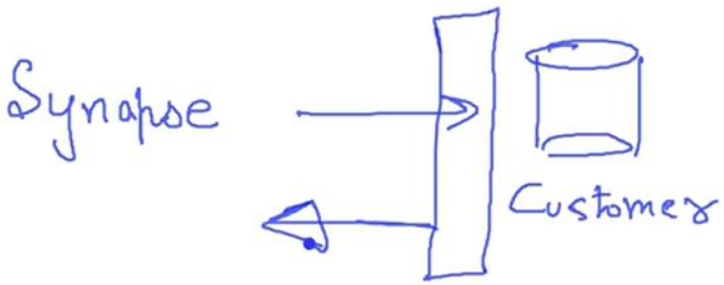
→Top 100 rows



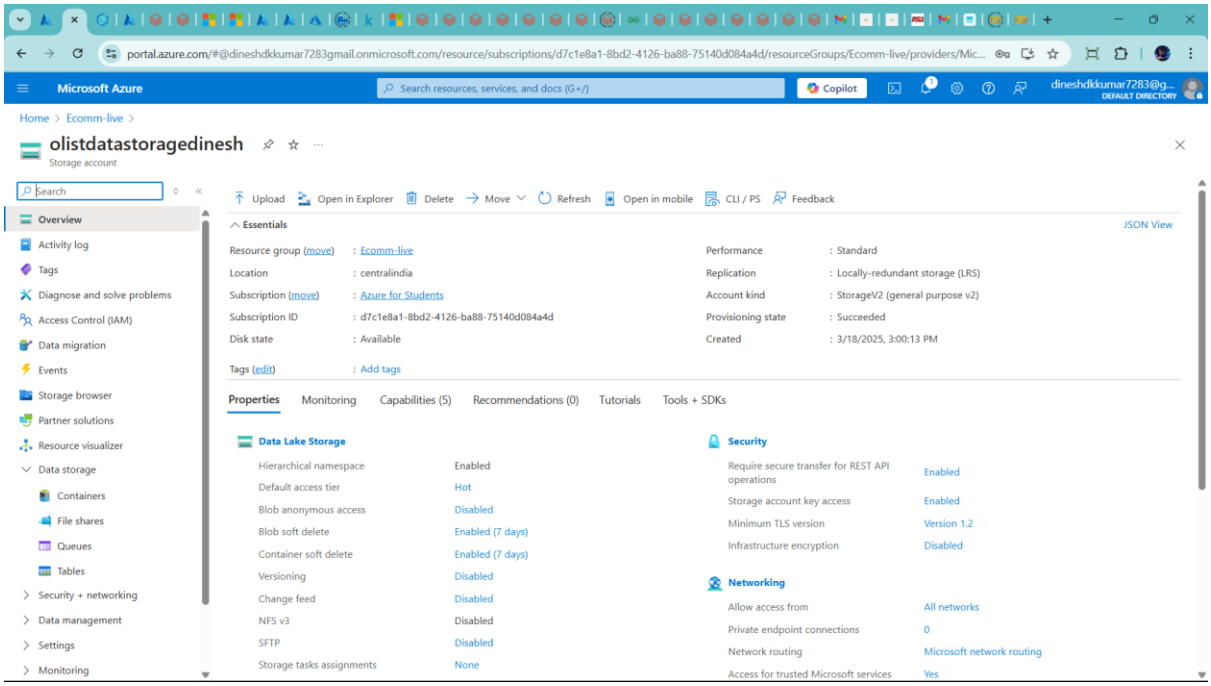
→run

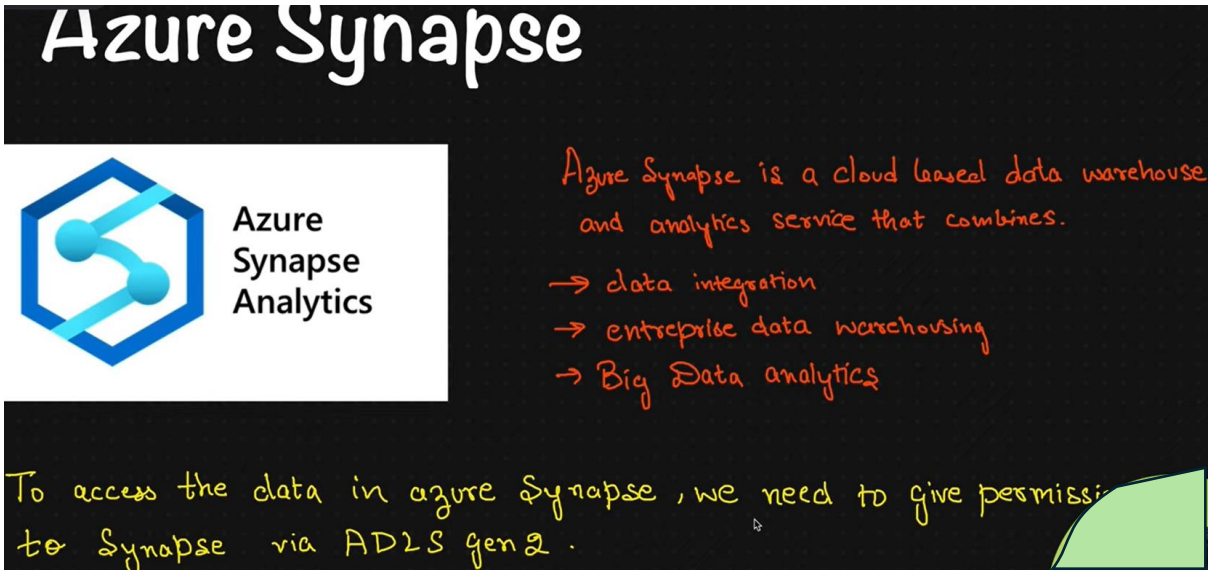


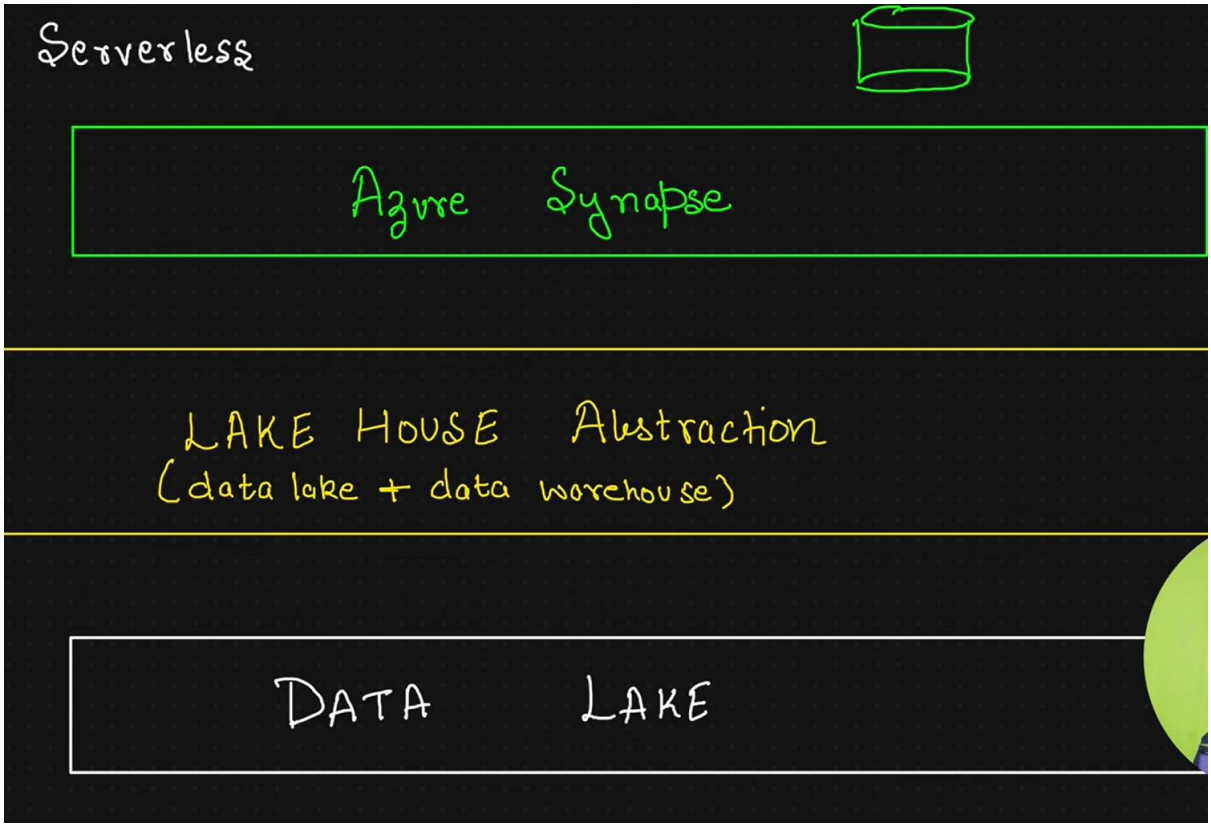
→openrowset →it is retrive data form ADLS Gen 2 . but they cannot store daa in synapse. It data present in ADLS gen 2 storage only they share only meta data



-→Now I want to read a data from olistdatastoragedinesh

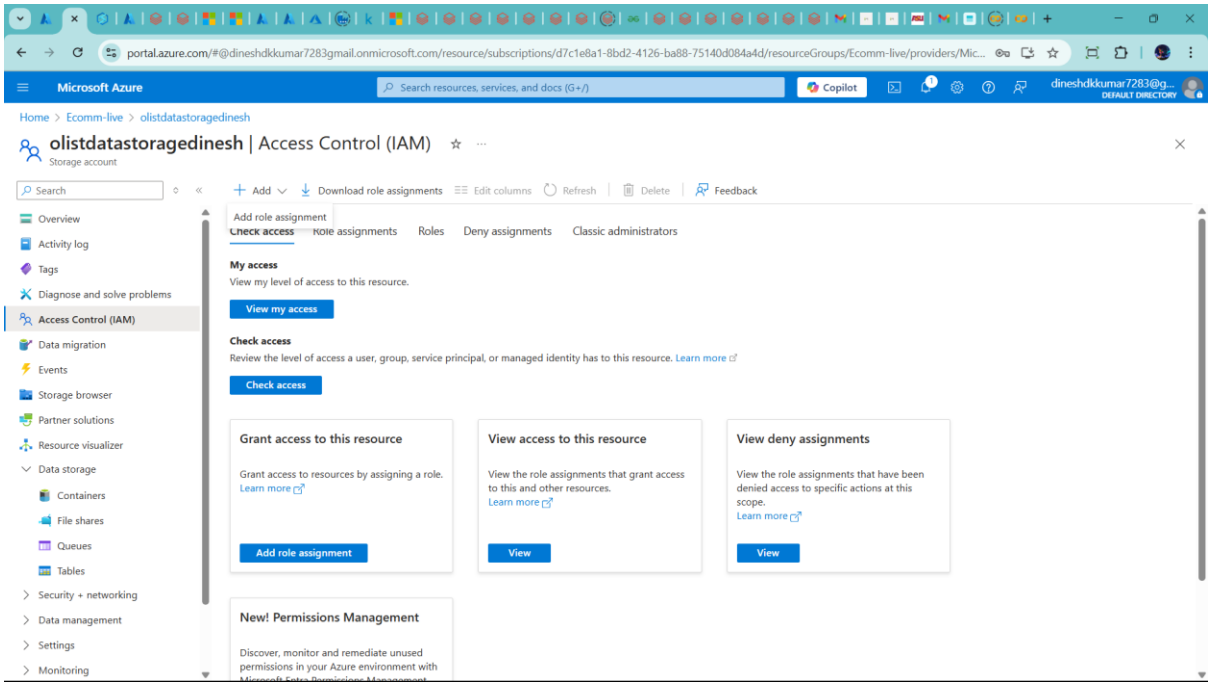

portal.azure.com/#@dineshdkkumar7283gmail.onmicrosoft.com/resource/subscriptions/d7c1e8a1-8bd2-4126-ba88-75140d084a4d/resourceGroups/Ecomm-live/providers/Mic...
Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
dineshdkkumar7283@g...
DEFAULT DIRECTORY
Home > Ecomm-live >
olistdatastoragedinesh
Storage account
Search
Overview
Activity log
Tags
Diagnose and solve problems
Access Control (IAM)
Data migration
Events
Storage browser
Partner solutions
Resource visualizer
Data storage
Containers
File shares
Queues
Tables
Security + networking
Data management
Settings
Monitoring
Upload
Open in Explorer
Delete
Move
Refresh
Open in mobile
CLI / PS
Feedback
Essentials
JSON View
Resource group (move) : Ecomm-live
Location : centralindia
Subscription (move) : Azure for Students
Subscription ID : d7c1e8a1-8bd2-4126-ba88-75140d084a4d
Disk state : Available
Performance : Standard
Replication : Locally-redundant storage (LRS)
Account kind : StorageV2 (general purpose v2)
Provisioning state : Succeeded
Created : 3/18/2025, 3:00:13 PM
Tags (edit) : Add tags
Properties
Monitoring
Capabilities (5)
Recommendations (0)
Tutorials
Tools + SDKs
Data Lake Storage
Hierarchical namespace Enabled
Default access tier Hot
Blob anonymous access Disabled
Blob soft delete Enabled (7 days)
Container soft delete Enabled (7 days)
Versioning Disabled
Change feed Disabled
NFS v3 Disabled
SFTP Disabled
Storage tasks assignments None
Security
Require secure transfer for REST API operations Enabled
Storage account key access Enabled
Minimum TLS version Version 1.2
Infrastructure encryption Disabled
Networking
Allow access from All networks
Private endpoint connections 0
Network routing Microsoft network routing
Access for trusted Microsoft services Yes


Azure Synapse
Azure
Synapse
Analytics
Azure Synapse is a cloud based data warehouse and analytics service that combines.
→ data integration
→ enterprise data warehousing
→ Big Data analytics
To access the data in azure Synapse, we need to give permissi
to Synapse via AD2S gen2.

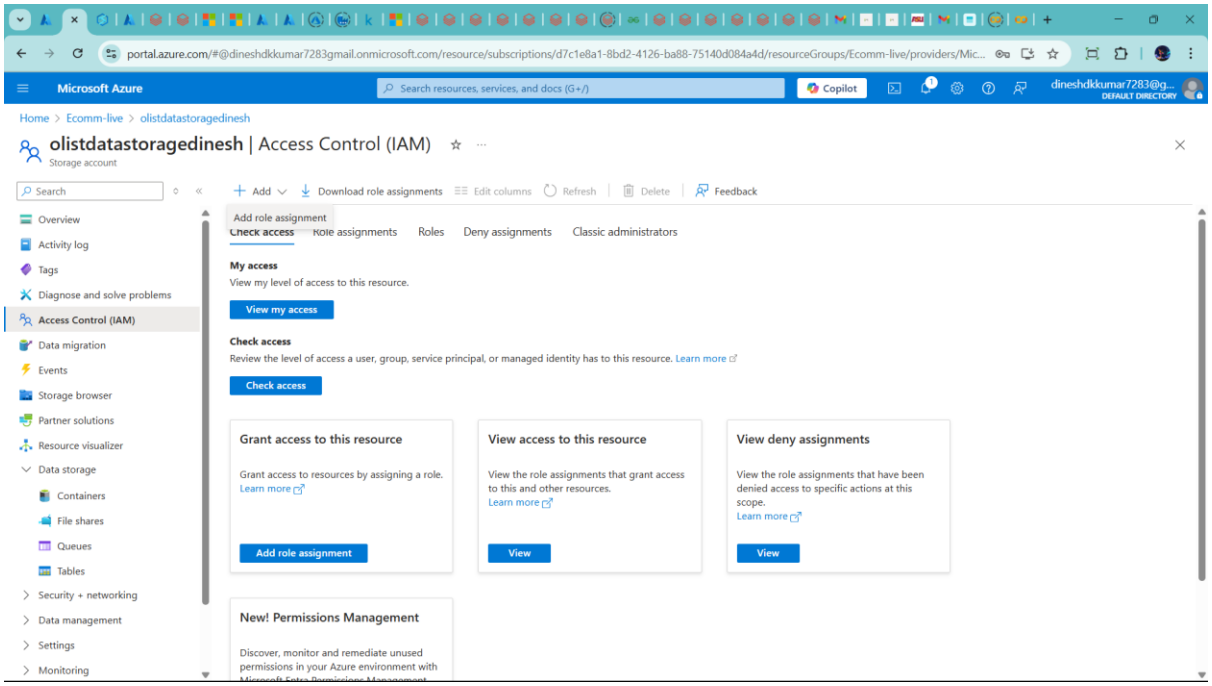


→data present in data lake

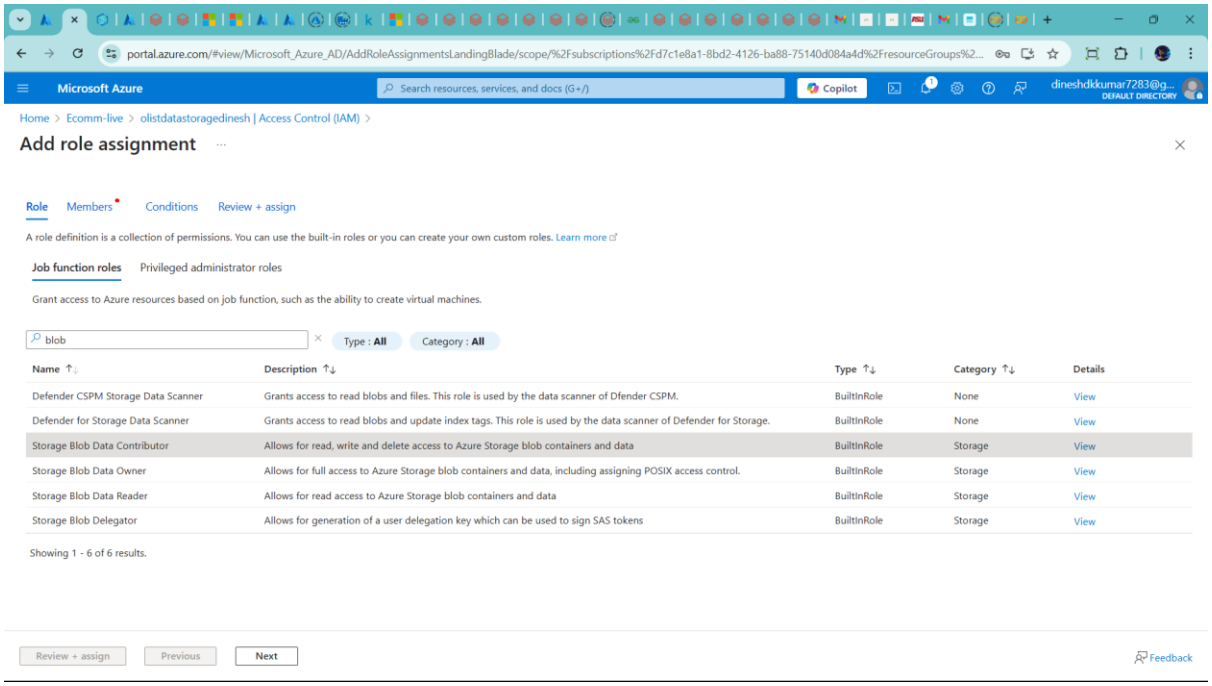
→go to stoage contaniner →give permisioin to get data from sysnapse

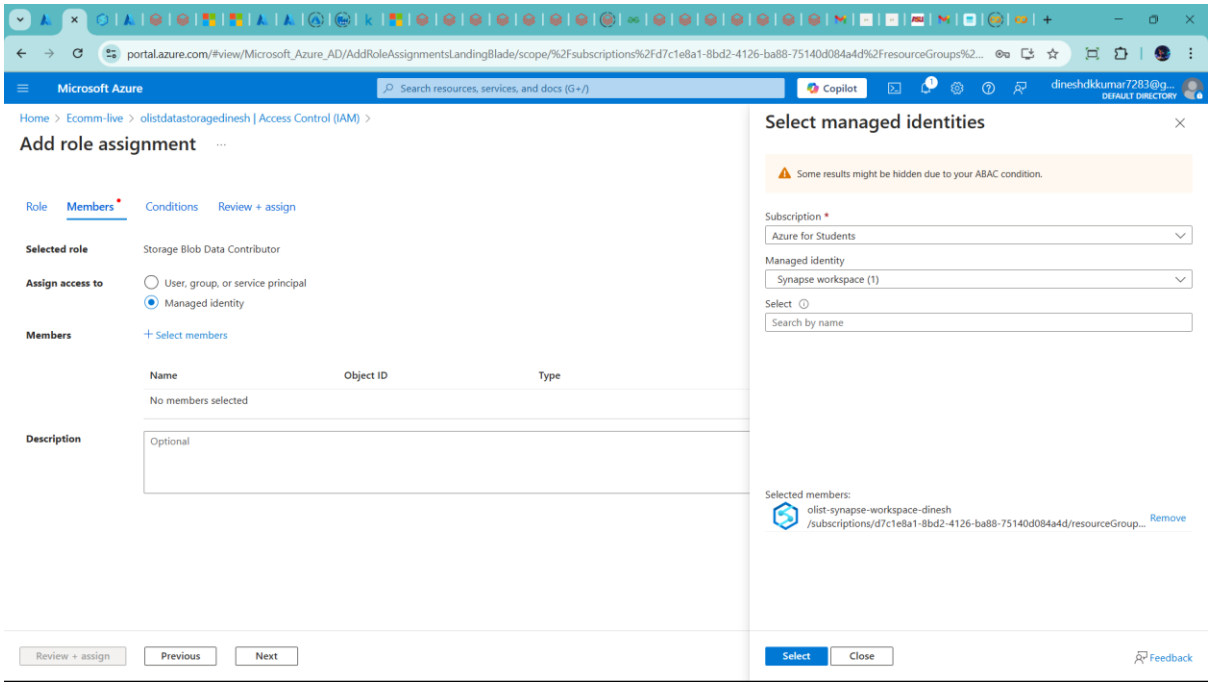


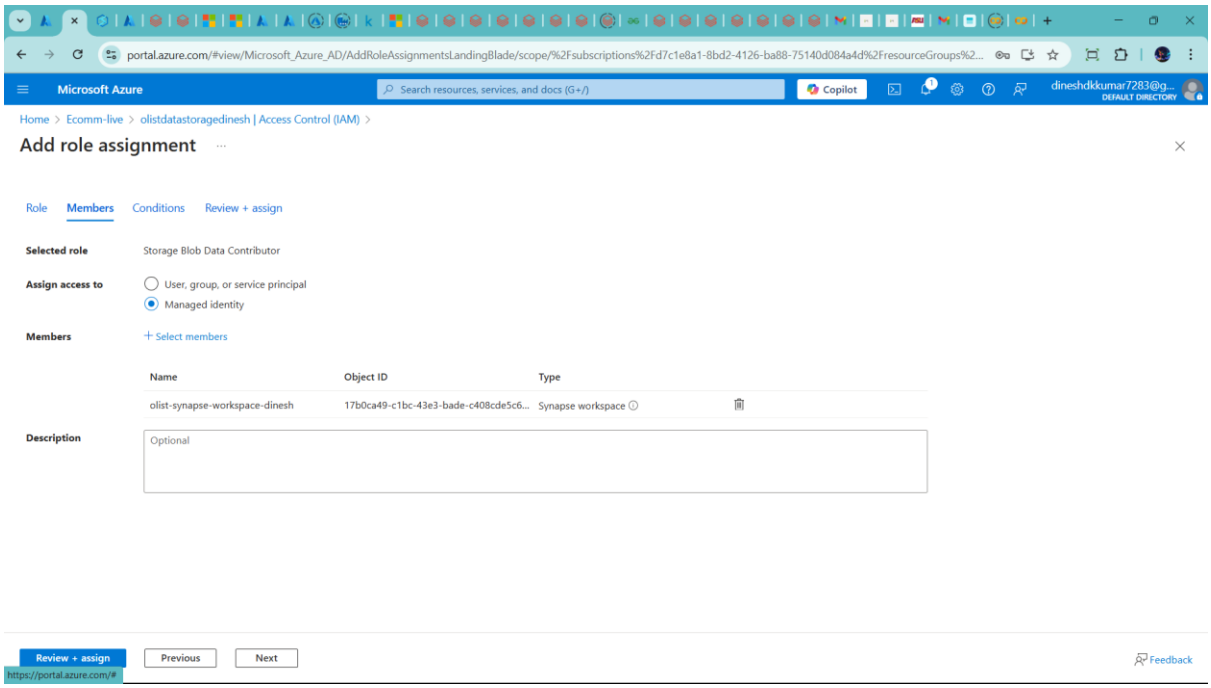
→click olistdatastoragedinesh →click Add roleassigment



→click blob container



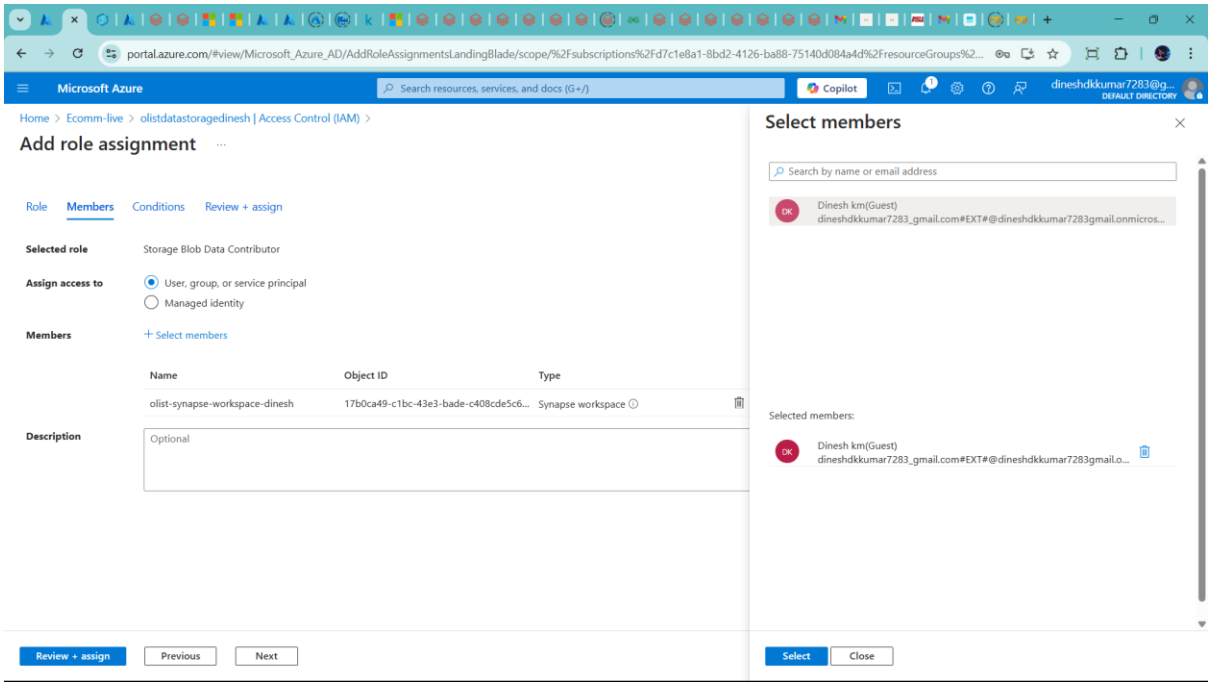




→you can muple user want to access try this below steps

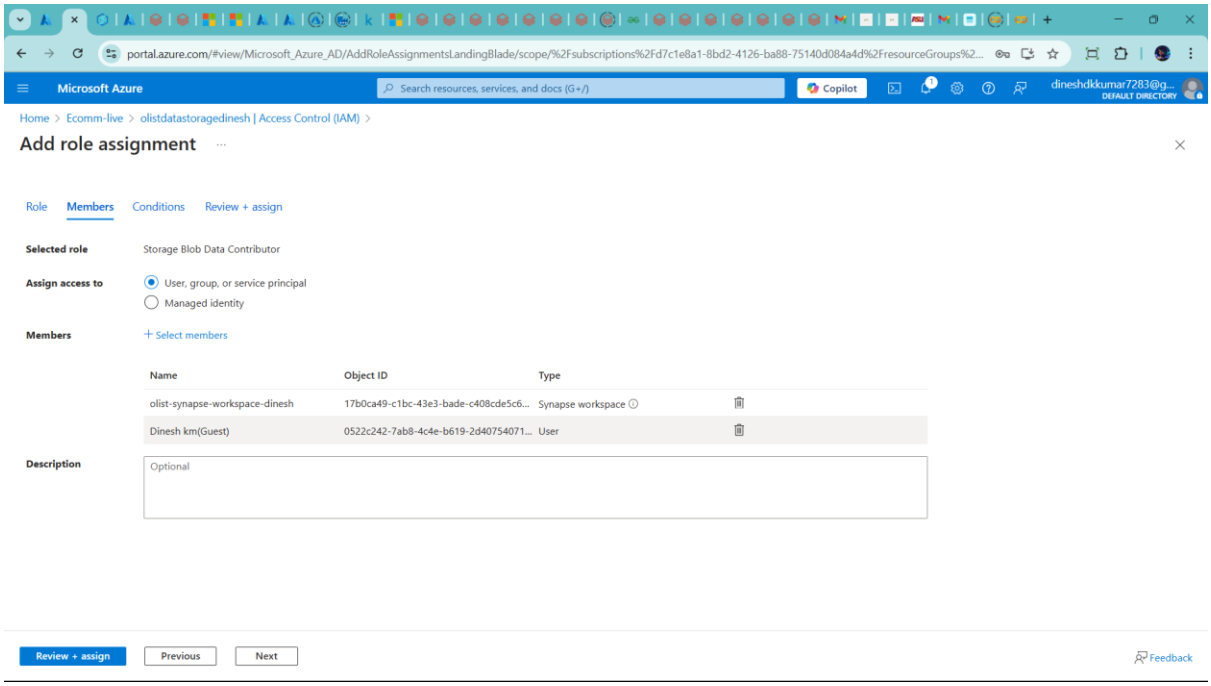
→click user,group or service principal

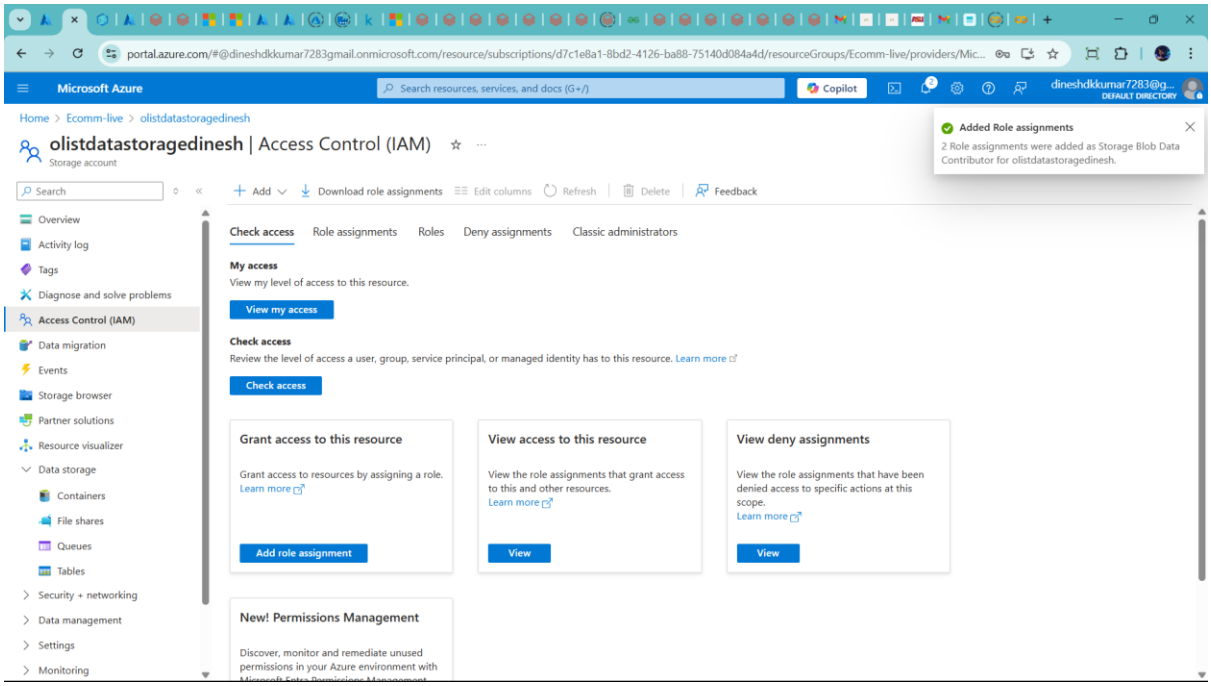
→click select member



→click select

→click review +assign





Transact-SQL syntax conventions

Syntax

`OPENROWSET` syntax is used to query external data sources:

syntaxsql

```
OPENROWSET
(  'provider_name'
    , { 'datasource' ; 'user_id' ; 'password' | 'provider_string' }
    , {  [ catalog. ] [ schema. ] object | 'query' }
)
```

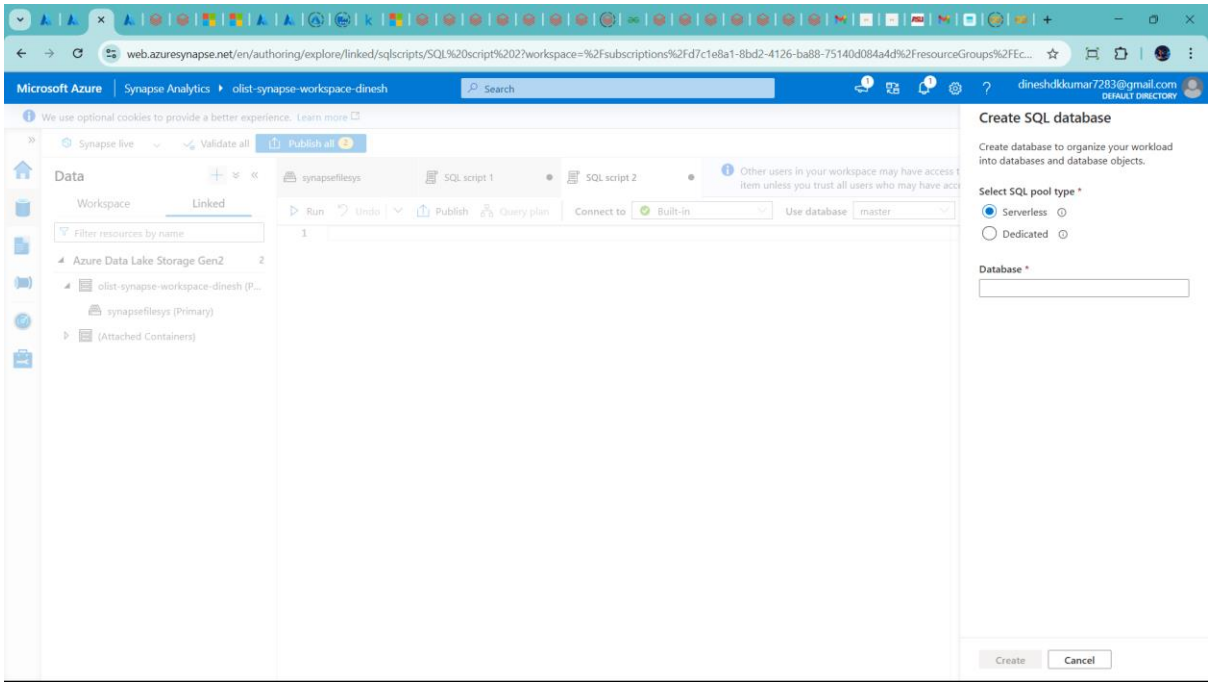
`OPENROWSET(BULK)` syntax is used to read external files:

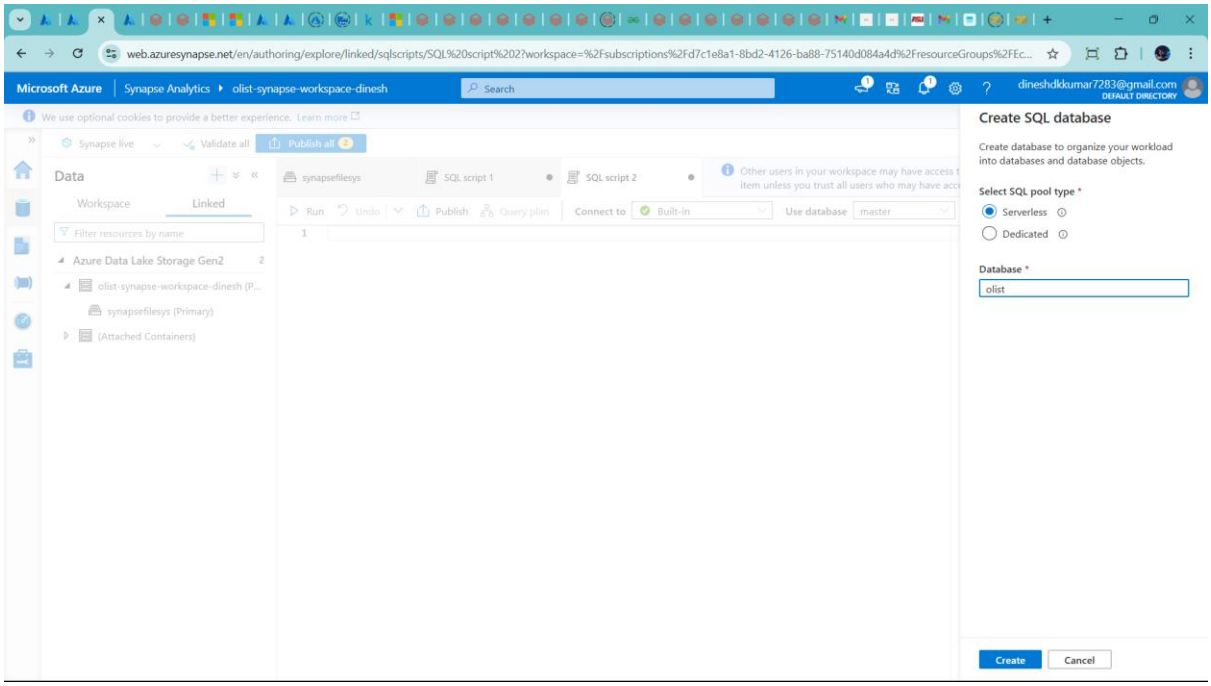
SQL Pool :- Serverless vs Dedicated

Feature	Serverless SQL Pool (Pay-as-you-go)	Dedicated SQL Pool (Provisioned)
How it works?	You don't set up or manage servers. It is **on-demand**. Synapse spins up resources when you need them.	You reserve fixed resources (compute power) for your data, which are always available.
Payment Model	Pay only for the queries you run.	Pay for the reserved compute (fixed cost).
Use Case	Good for **exploring large datasets** or running queries occasionally.	Best for **big workloads** that need fast performance and predictable speed.
Data Location	Data stays in the **Data Lake** (e.g., ADLS Gen2).	Data is loaded into a **dedicated SQL database** in Synapse.
Setup Complexity	Very **easy** to set up—no servers to manage.	More setup needed, but provides **high performance**.

Bus own Car

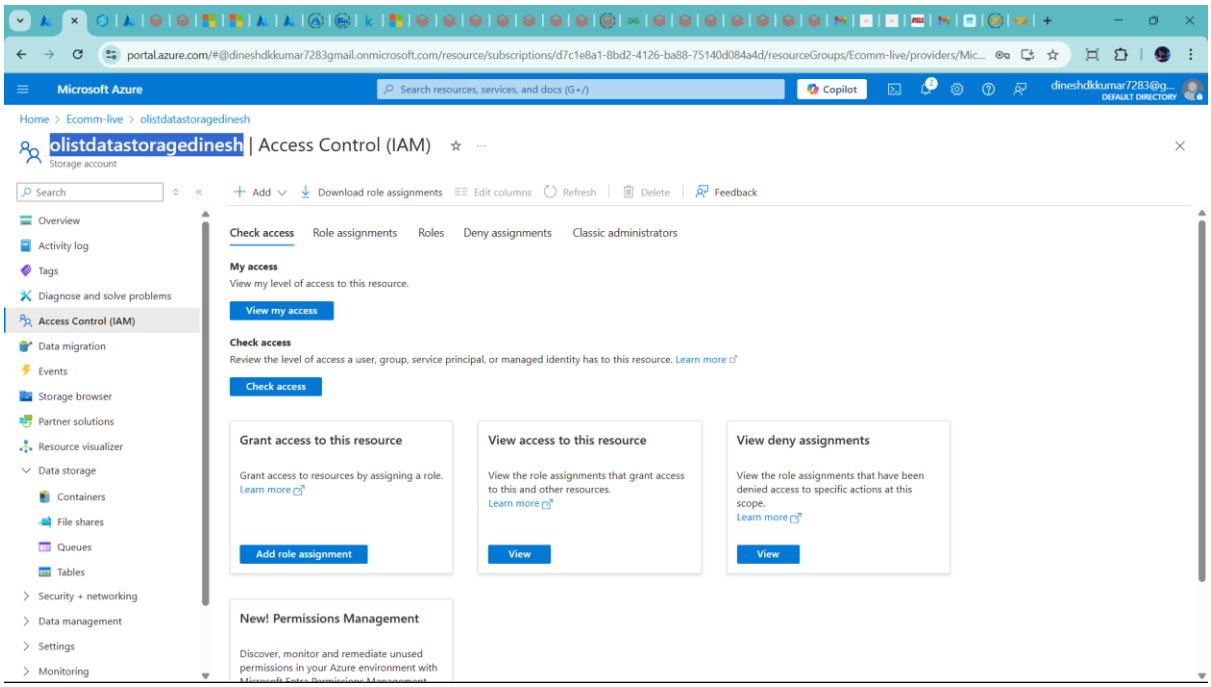
→go to data →click Sql databases



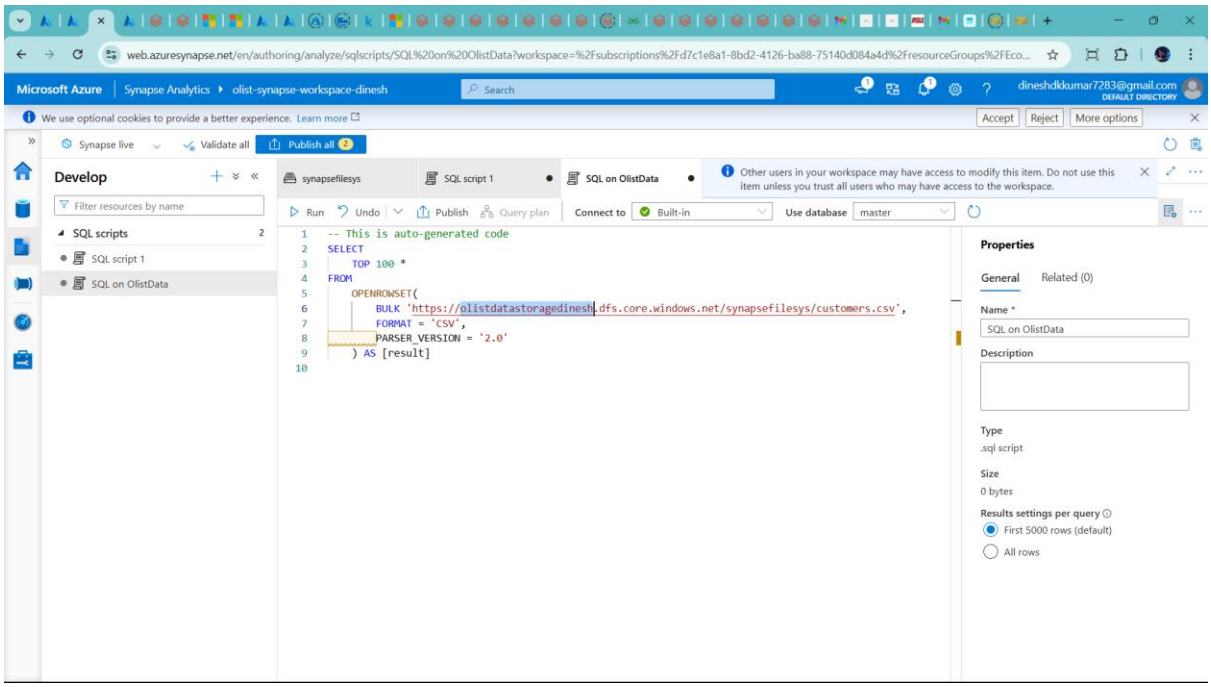


→create

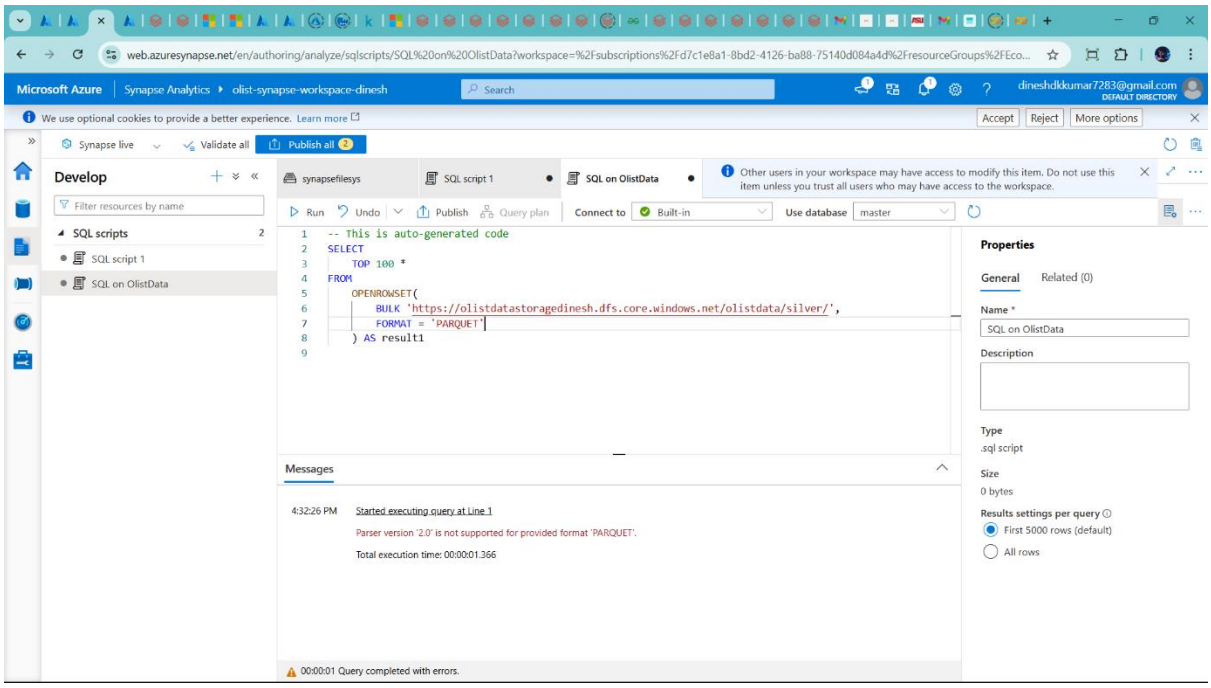
→copy that name



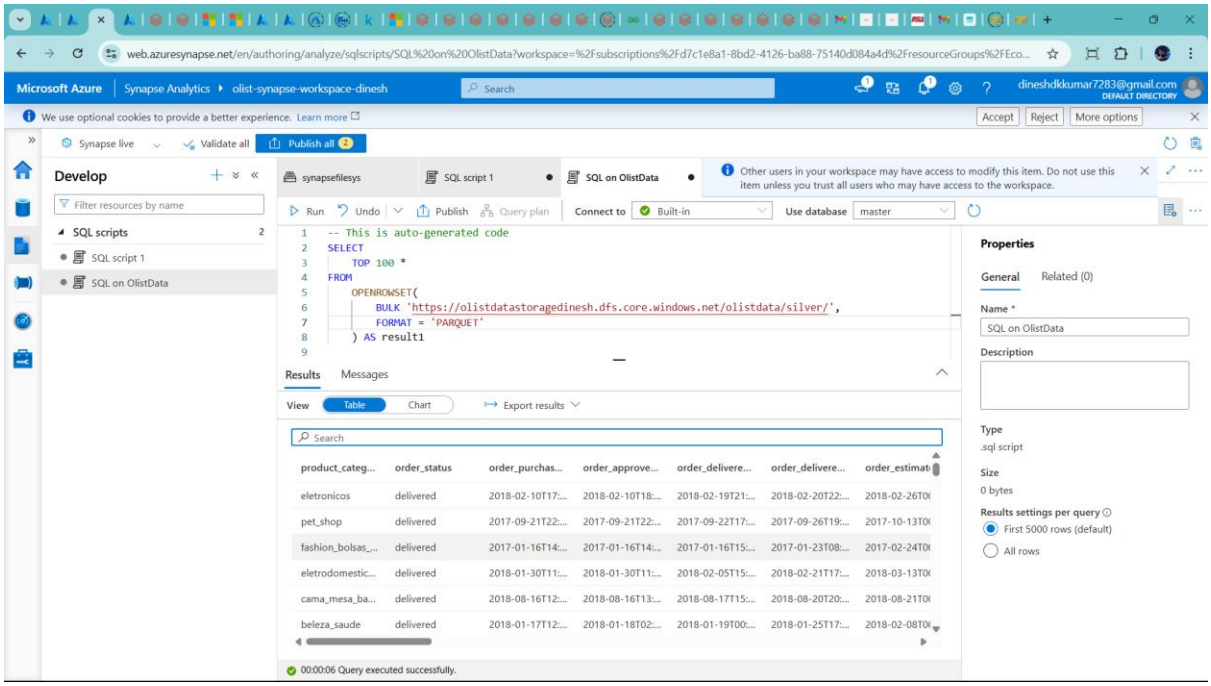
→change it



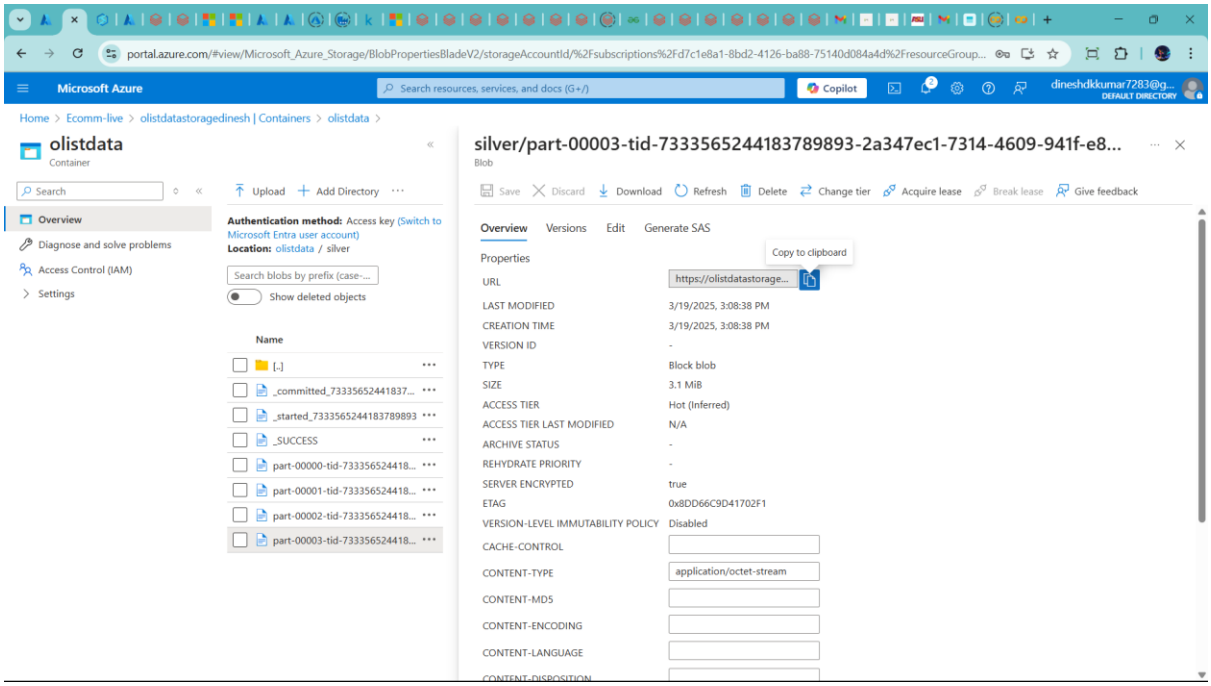
→



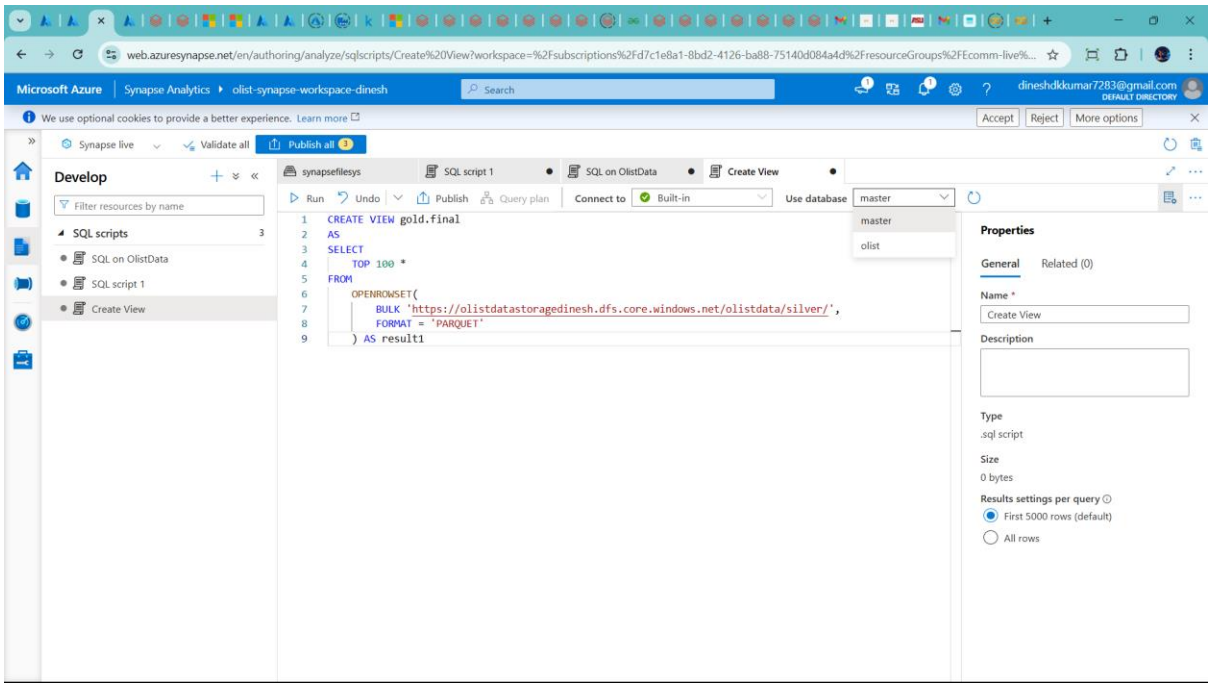
→click run



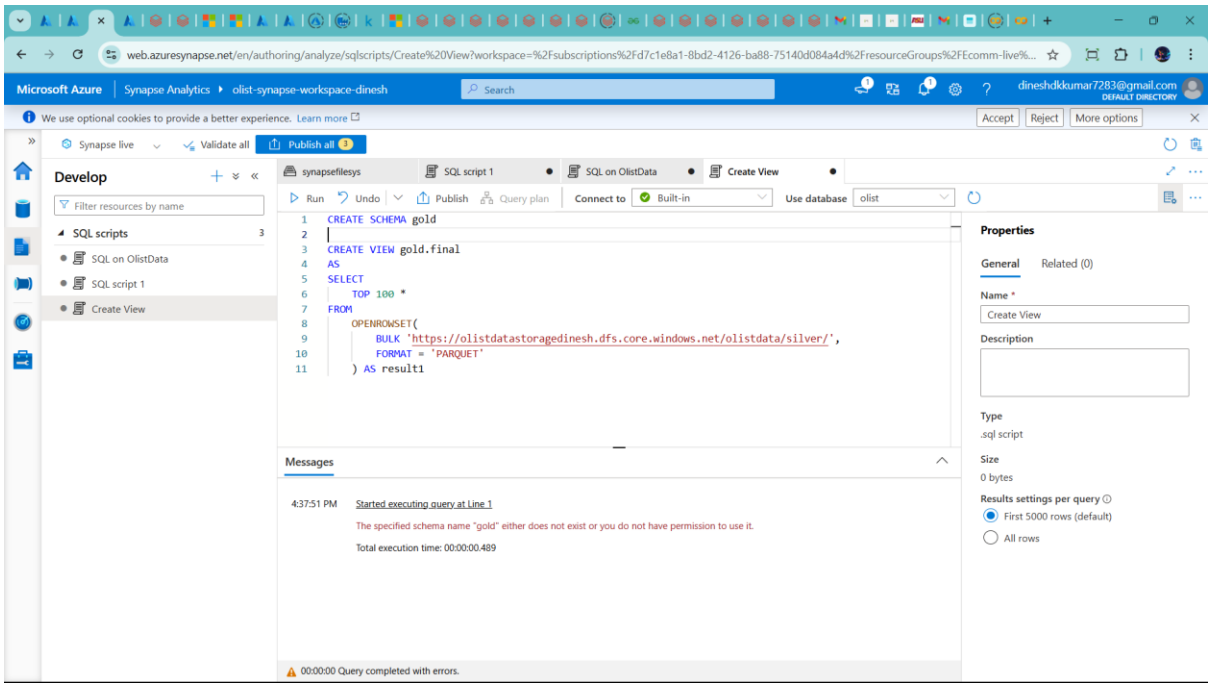
→path name already present in storage container

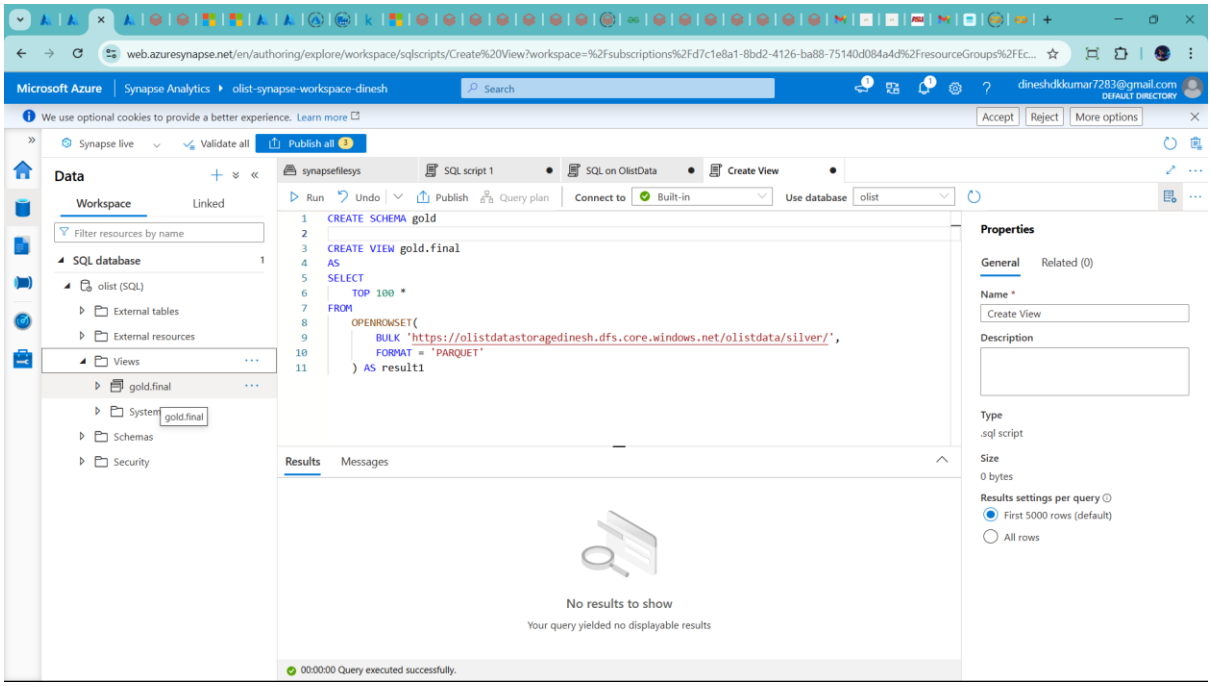


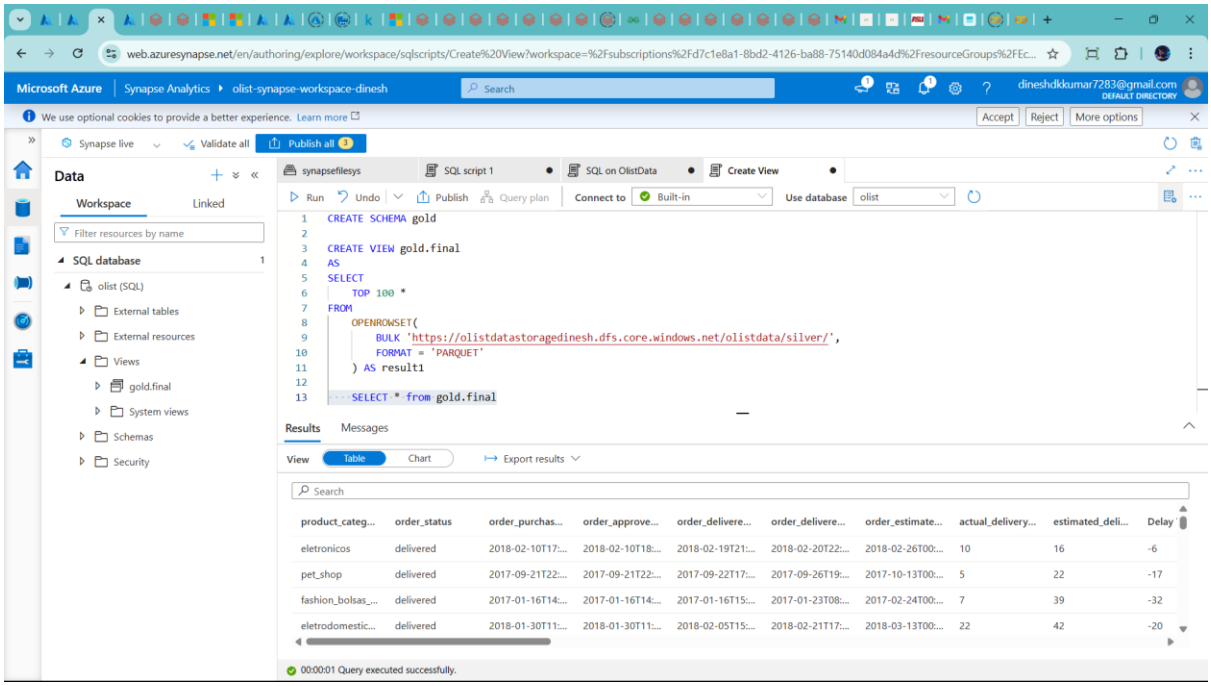
→chage database master to olist



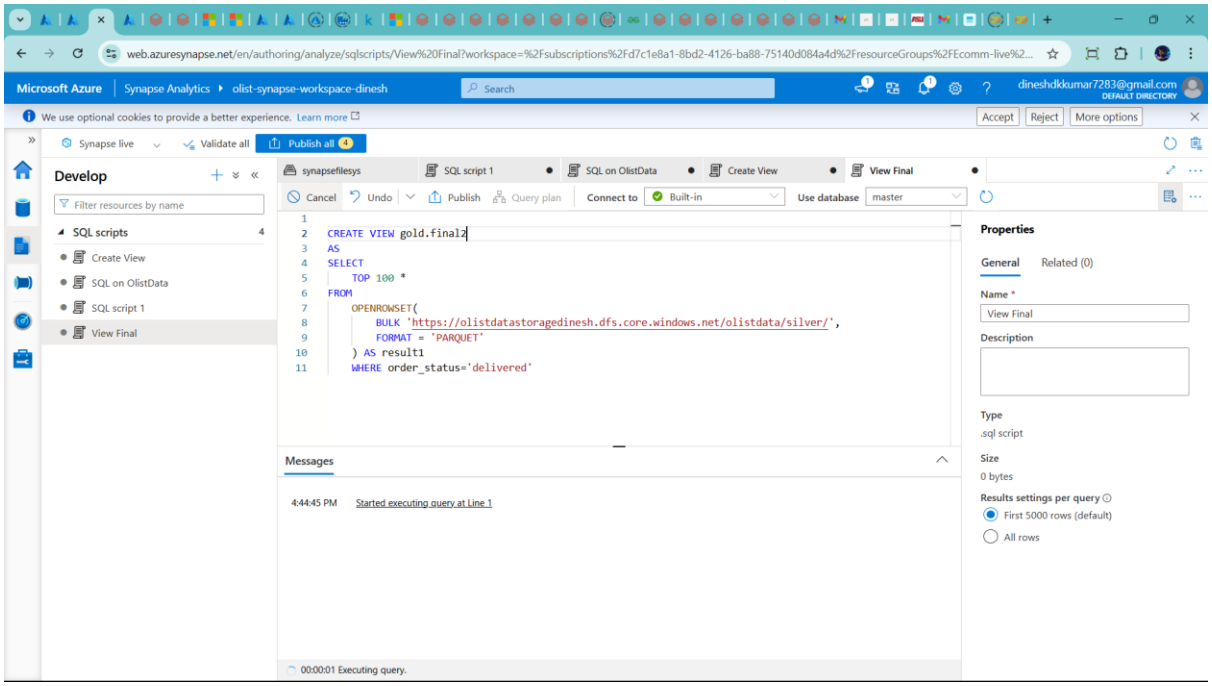
and run→successful create view

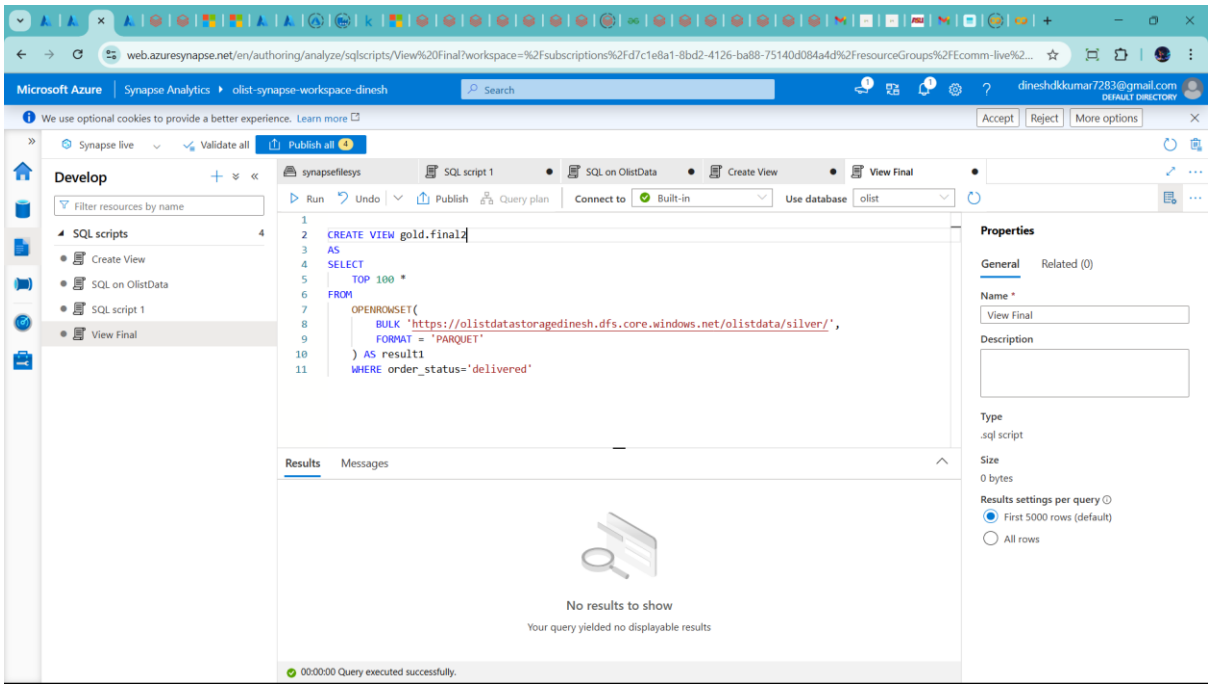






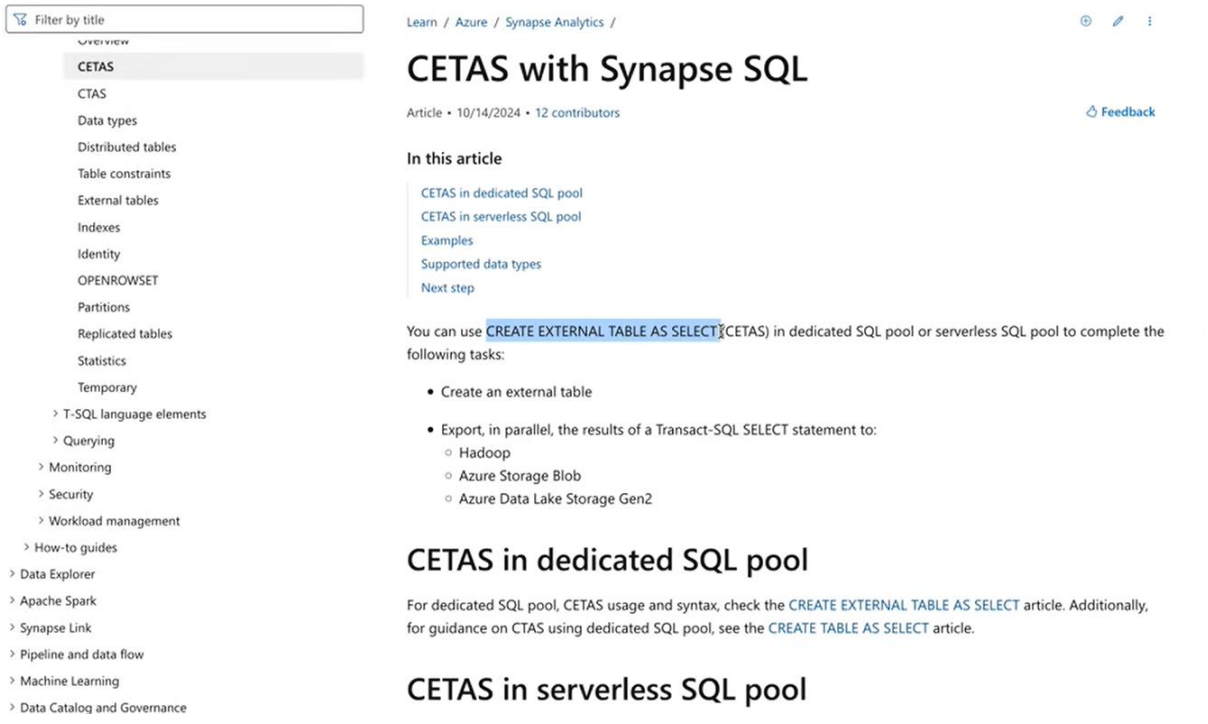
→create another view

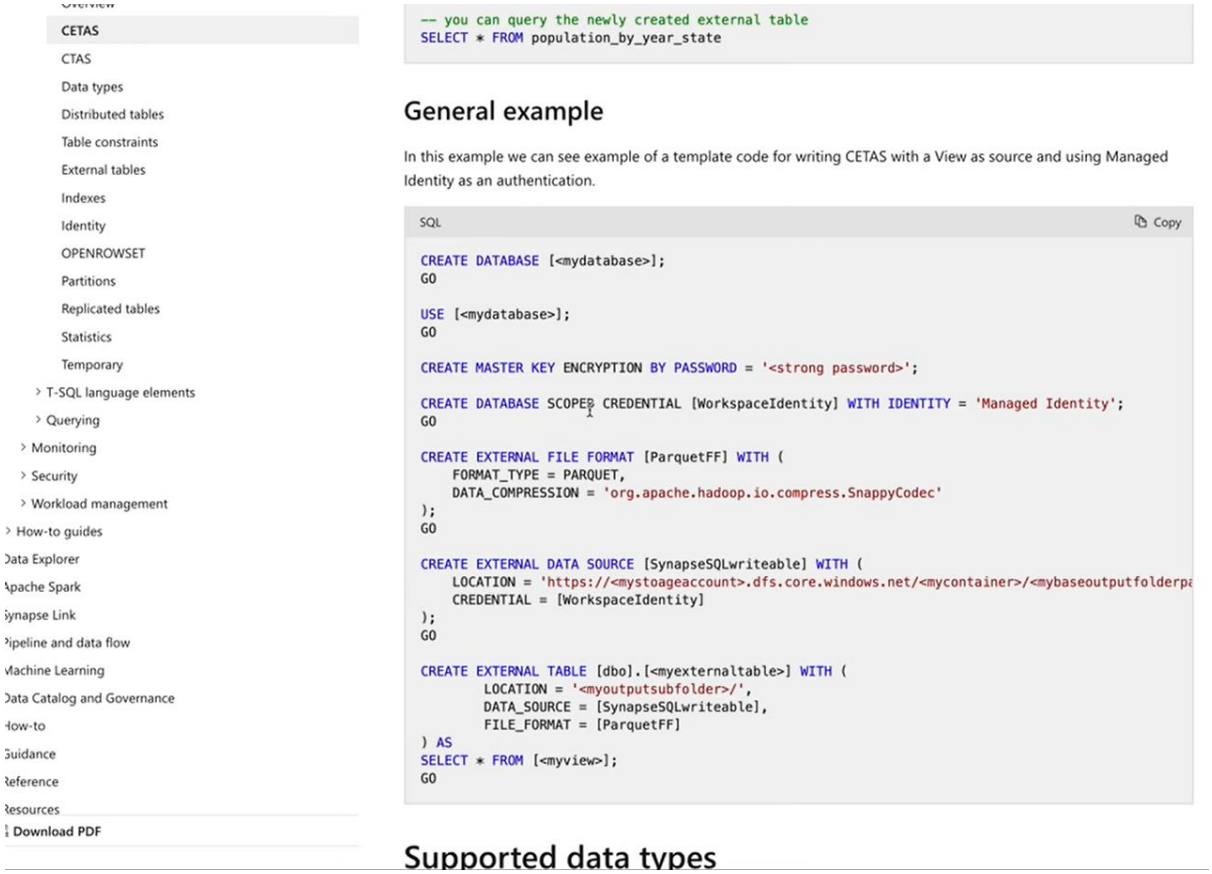

web.azuresynapse.net/en/authoring/analyze/sqlscripts/View%20Final?workspace=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2...
Microsoft Azure
Synapse Analytics
olist-synapse-workspace-dinesh
Search
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
We use optional cookies to provide a better experience. Learn more
Accept
Reject
More options
Synapse live
Validate all
Publish all 4
Develop
Filter resources by name
SQL scripts 4
Create View
SQL on OlistData
SQL script 1
View Final
synapsefilesys
SQL script 1
SQL on OlistData
Create View
View Final
Cancel
Undo
Publish
Query plan
Connect to
Built-in
Use database
master
1 CREATE VIEW gold.final2
2 AS
3 SELECT
4 TOP 100 *
5 FROM
6 OPENROWSET(
7 BULK 'https://olistdatastoragedinesh.dfs.core.windows.net/olistdata/silver/',
8 FORMAT = 'PARQUET'
9) AS result1
10 WHERE order_status='delivered'
Messages
4:44:45 PM Started executing query at Line 1
00:00:01 Executing query.
Properties
General
Related (0)
Name *
View Final
Description
Type
.sql script
Size
0 bytes
Results settings per query
First 5000 rows (default)
All rows


web.azuresynapse.net/en/authoring/analyze/sqlscripts/View%20Final?workspace=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-live%2...
Microsoft Azure
Synapse Analytics
olist-synapse-workspace-dinesh
Search
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
We use optional cookies to provide a better experience. Learn more
Accept
Reject
More options
Synapse live
Validate all
Publish all 4
Develop
Filter resources by name
SQL scripts 4
Create View
SQL on OlistData
SQL script 1
View Final
synapsefilesys
SQL script 1
SQL on OlistData
Create View
View Final
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
olist
1 CREATE VIEW gold.final2
2 AS
3 SELECT
4 TOP 100 *
5 FROM
6 OPENROWSET(
7 BULK 'https://olistdatastoragedinesh.dfs.core.windows.net/olistdata/silver/',
8 FORMAT = 'PARQUET'
9) AS result1
10 WHERE order_status='delivered'
Results
Messages
No results to show
Your query yielded no displayable results
00:00:00 Query executed successfully.
Properties
General
Related (0)
Name *
View Final
Description
Type
.sql script
Size
0 bytes
Results settings per query
First 5000 rows (default)
All rows

Azure Synapse Workflow

0. Medallin architecture and Lakehouse (Pre-requisite)
1. Create a serverless SQL pool. (for just reading/Querying)
2. Create Schema & then User / password
3. Use OPENROWSET to read data from silver layer.
   (here data is not stored, just referenced to silver layer)
4. Create a view to easily Query the data.



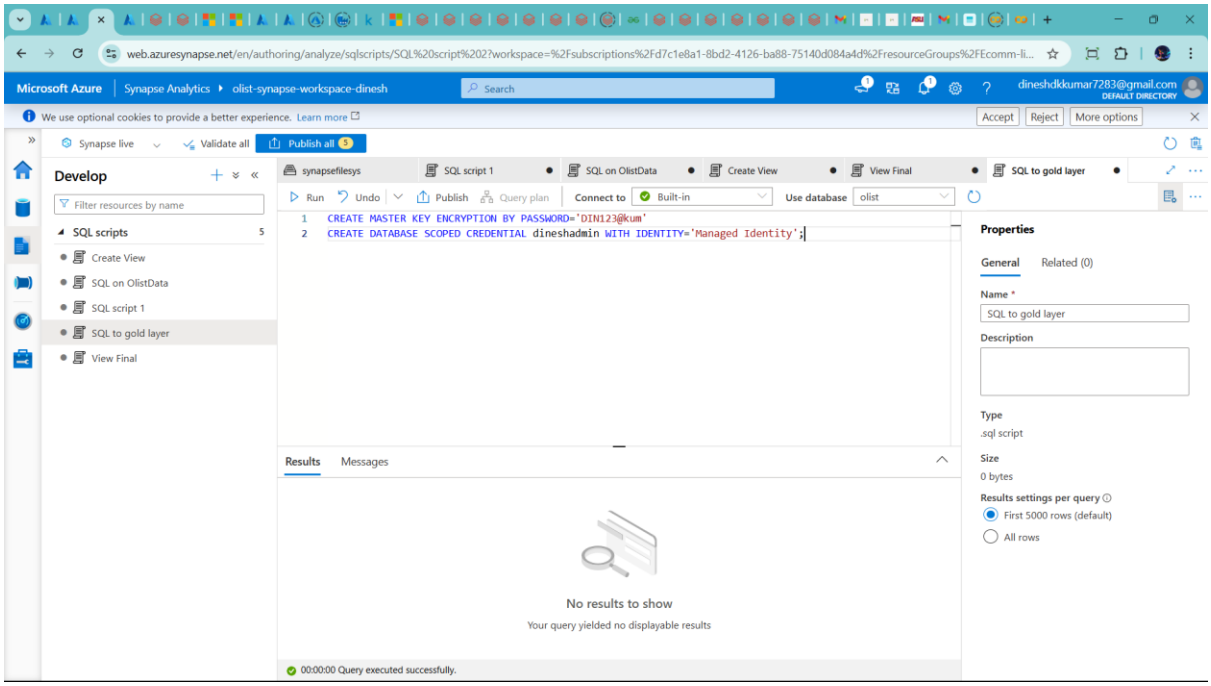


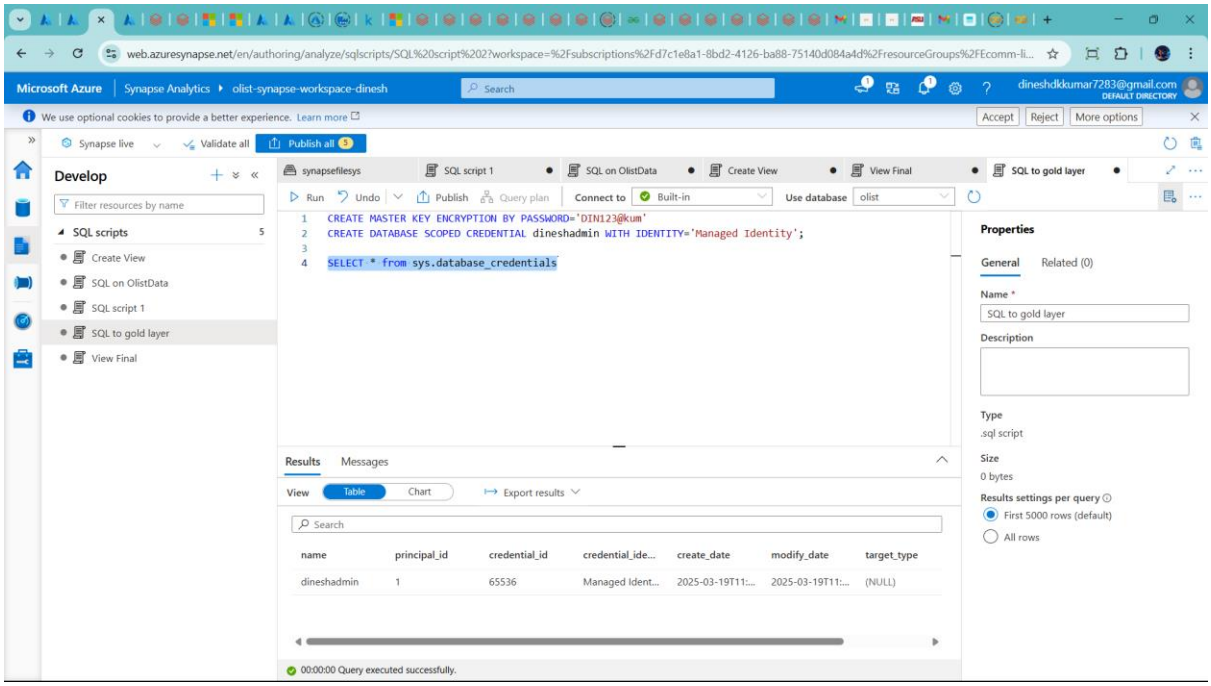
→I want to create external table to store data in gold layer

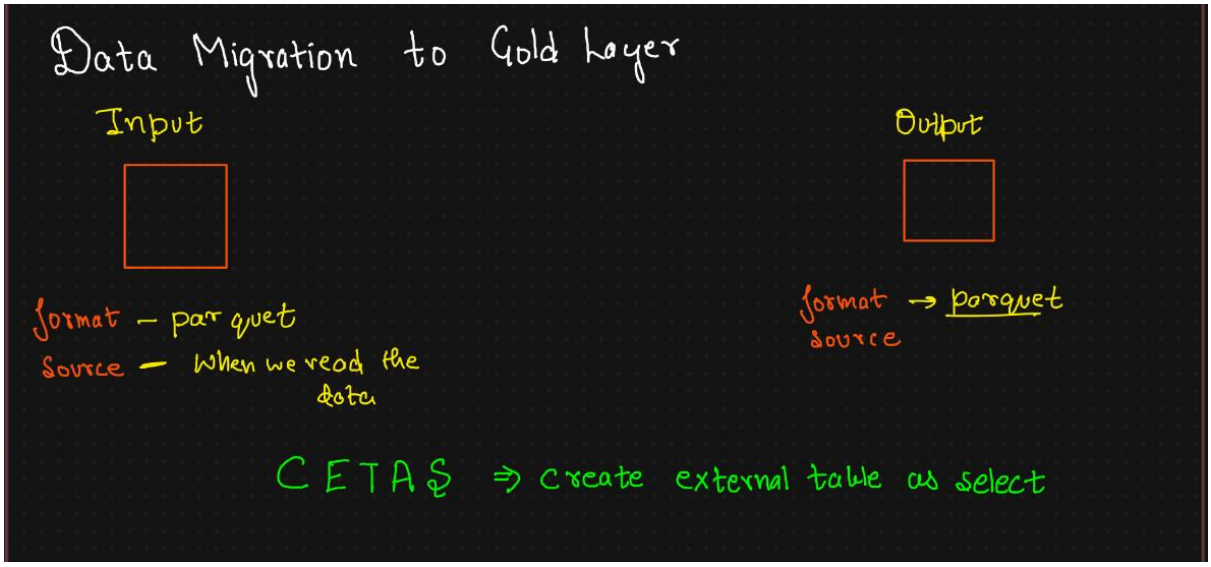
→now create master key password

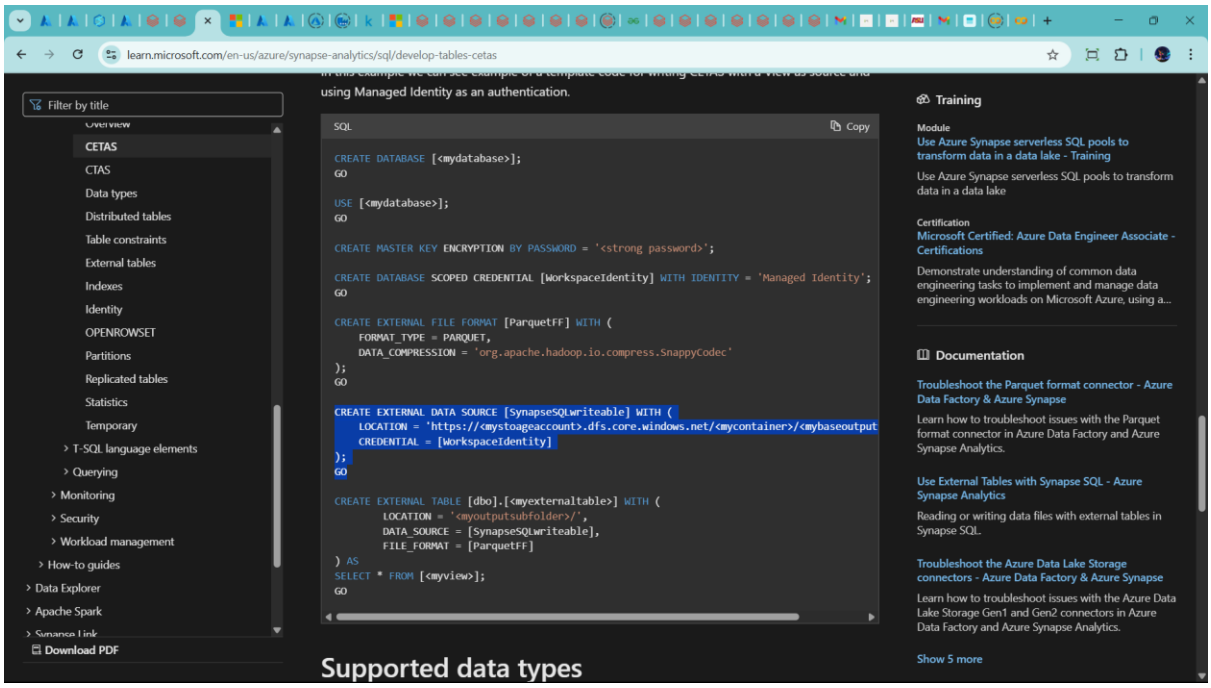
→first click new sql script → name as SQL to gold layer

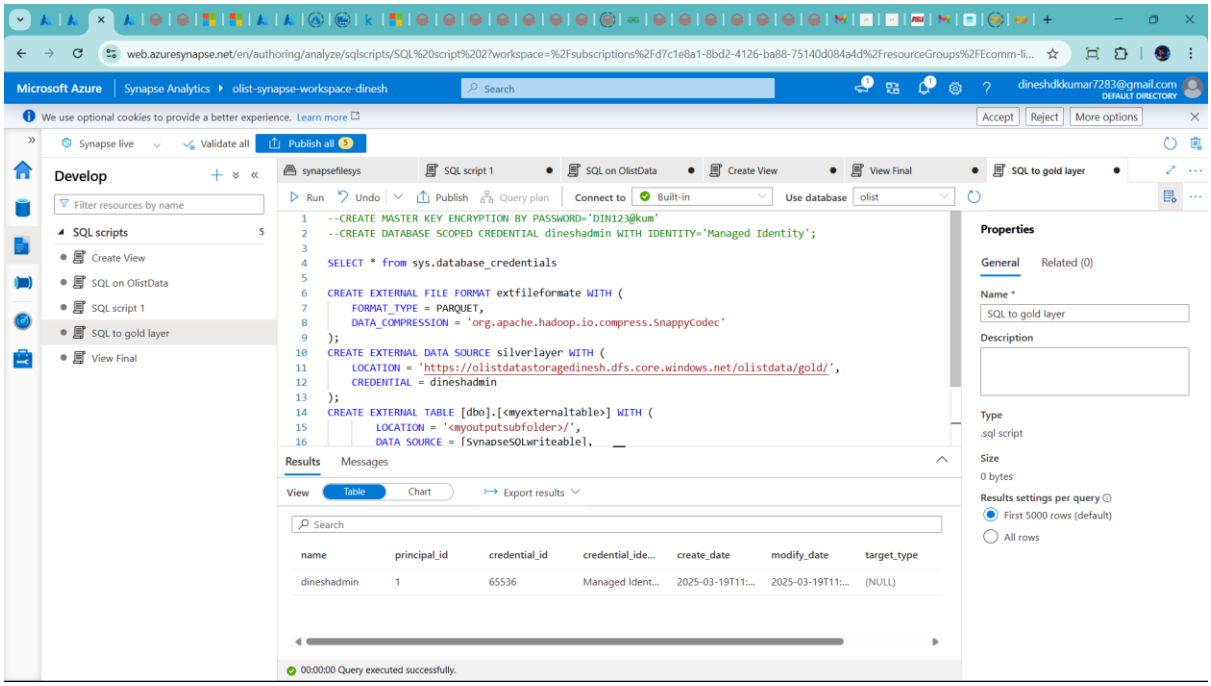
```
1
2   use olist
3   drop master key
4
5
6   select * from sys.database_credentials
7   drop database scoped CREDENTIAL mayankadmin
```


web.azuresynapse.net/en/authoring/analyze/sqlscripts/SQL%20script%202?workspace=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-li...
Microsoft Azure
Synapse Analytics
olist-synapse-workspace-dinesh
Search
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
We use optional cookies to provide a better experience. Learn more
Accept
Reject
More options
Synapse live
Validate all
Publish all 5
Develop
Filter resources by name
SQL scripts 5
Create View
SQL on OlistData
SQL script 1
SQL to gold layer
View Final
synapsefilesys
SQL script 1
SQL on OlistData
Create View
View Final
SQL to gold layer
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
olist
1 CREATE MASTER KEY ENCRYPTION BY PASSWORD='DIN123@kum'
2 CREATE DATABASE SCOPED CREDENTIAL dineshadmin WITH IDENTITY='Managed Identity';
Properties
General
Related (0)
Name *
SQL to gold layer
Description
Type
.sql script
Size
0 bytes
Results settings per query
First 5000 rows (default)
All rows
Results
Messages
No results to show
Your query yielded no displayable results
00:00:00 Query executed successfully.

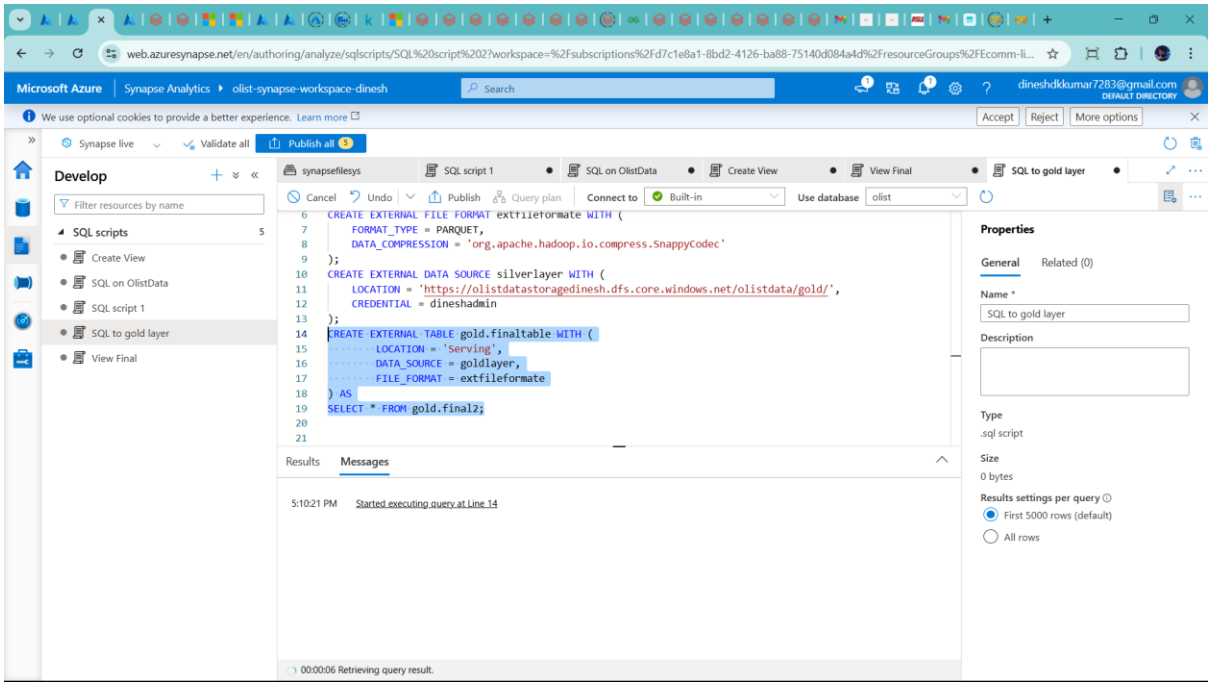

web.azuresynapse.net/en/authoring/analyze/sqlscripts/SQL%20script%202?workspace=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-li...
Microsoft Azure
Synapse Analytics
olist-synapse-workspace-dinesh
Search
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
We use optional cookies to provide a better experience. Learn more
Accept
Reject
More options
Synapse live
Validate all
Publish all 5
Develop
Filter resources by name
SQL scripts 5
Create View
SQL on OlistData
SQL script 1
SQL to gold layer
View Final
synapsefilesys
SQL script 1
SQL on OlistData
Create View
View Final
SQL to gold layer
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
olist
1 CREATE MASTER KEY ENCRYPTION BY PASSWORD='DIN123@kum'
2 CREATE DATABASE SCOPED CREDENTIAL dineshadmin WITH IDENTITY='Managed Identity';
3
4 SELECT * from sys.database_credentials
Properties
General
Related (0)
Name *
SQL to gold layer
Description
Type
.sql script
Size
0 bytes
Results settings per query
First 5000 rows (default)
All rows
Results
Messages
View
Table
Chart
Export results
Search
name
principal_id
credential_id
credential_ide...
create_date
modify_date
target_type
dineshadmin
1
65536
Managed Ident...
2025-03-19T11:...
2025-03-19T11:...
(NULL)
00:00:00 Query executed successfully.


Data Migration to Gold Layer
Input
Output
format – parquet
source – When we read the data
format → parquet
source
CETAS ⇒ create external table as select

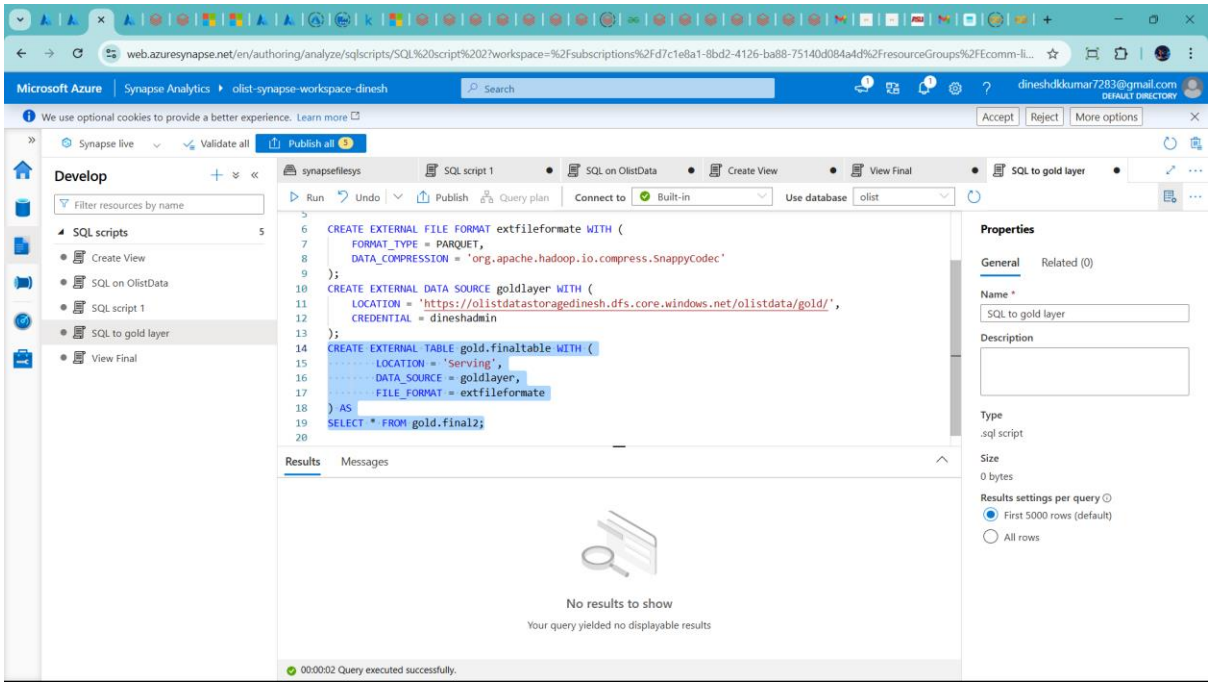

learn.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-cetas
Filter by title
CETAS
CTAS
Data types
Distributed tables
Table constraints
External tables
Indexes
Identity
OPENROWSET
Partitions
Replicated tables
Statistics
Temporary
T-SQL language elements
Querying
Monitoring
Security
Workload management
How-to guides
Data Explorer
Apache Spark
Download PDF
using Managed Identity as an authentication.
SQL
Copy
CREATE DATABASE [<mydatabase>];
GO
USE [<mydatabase>];
GO
CREATE MASTER KEY ENCRYPTION BY PASSWORD = '<strong password>';
CREATE DATABASE SCOPED CREDENTIAL [WorkspaceIdentity] WITH IDENTITY = 'Managed Identity';
GO
CREATE EXTERNAL FILE FORMAT [ParquetFF] WITH (
FORMAT_TYPE = PARQUET,
DATA_COMPRESSION = 'org.apache.hadoop.io.compress.SnappyCodec'
);
GO
CREATE EXTERNAL DATA SOURCE [SynapseSQLwriteable] WITH (
LOCATION = 'https://<mystoageaccount>.dfs.core.windows.net/<mycontainer>/<mybaseoutput
CREDENTIAL = [WorkspaceIdentity]
);
GO
CREATE EXTERNAL TABLE [dbo].[<myexternaltable>] WITH (
LOCATION = '<myoutputsubfolder>/',
DATA_SOURCE = [SynapseSQLwriteable],
FILE_FORMAT = [ParquetFF]
) AS
SELECT * FROM [<myview>];
GO
Supported data types
Training
Module
Use Azure Synapse serverless SQL pools to transform data in a data lake - Training
Use Azure Synapse serverless SQL pools to transform data in a data lake
Certification
Microsoft Certified: Azure Data Engineer Associate - Certifications
Demonstrate understanding of common data engineering tasks to implement and manage data engineering workloads on Microsoft Azure, using a...
Documentation
Troubleshoot the Parquet format connector - Azure Data Factory & Azure Synapse
Learn how to troubleshoot issues with the Parquet format connector in Azure Data Factory and Azure Synapse Analytics.
Use External Tables with Synapse SQL - Azure Synapse Analytics
Reading or writing data files with external tables in Synapse SQL.
Troubleshoot the Azure Data Lake Storage connectors - Azure Data Factory & Azure Synapse
Learn how to troubleshoot issues with the Azure Data Lake Storage Gen1 and Gen2 connectors in Azure Data Factory and Azure Synapse Analytics.
Show 5 more

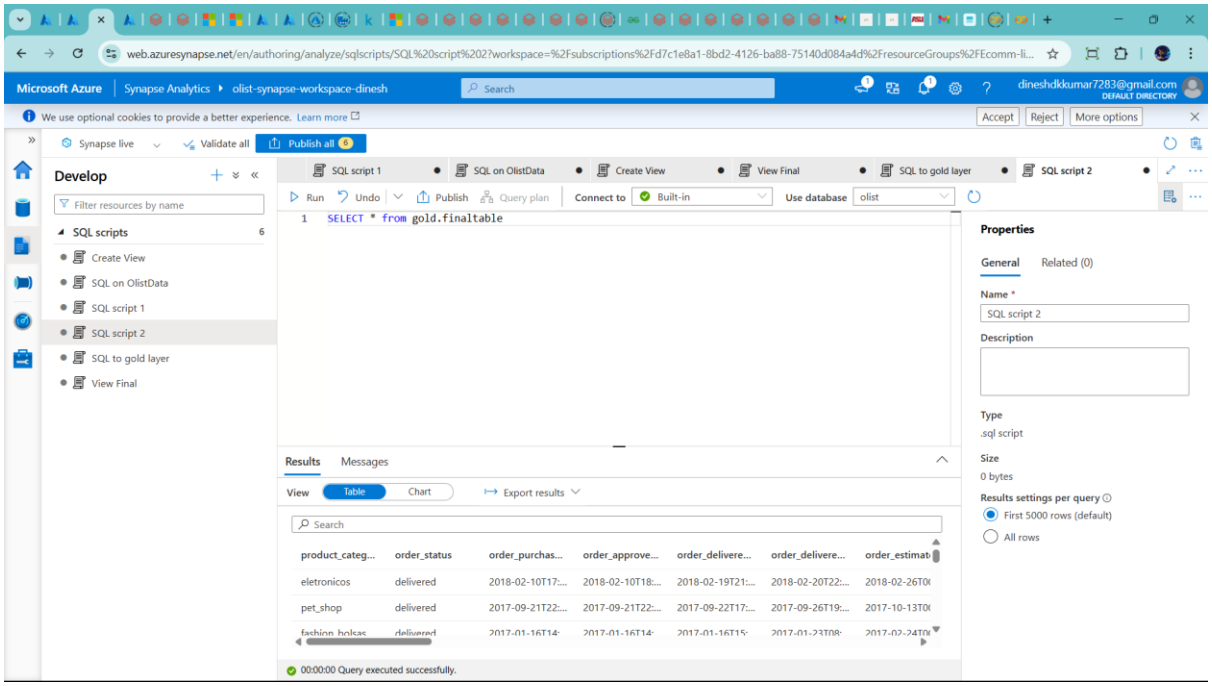


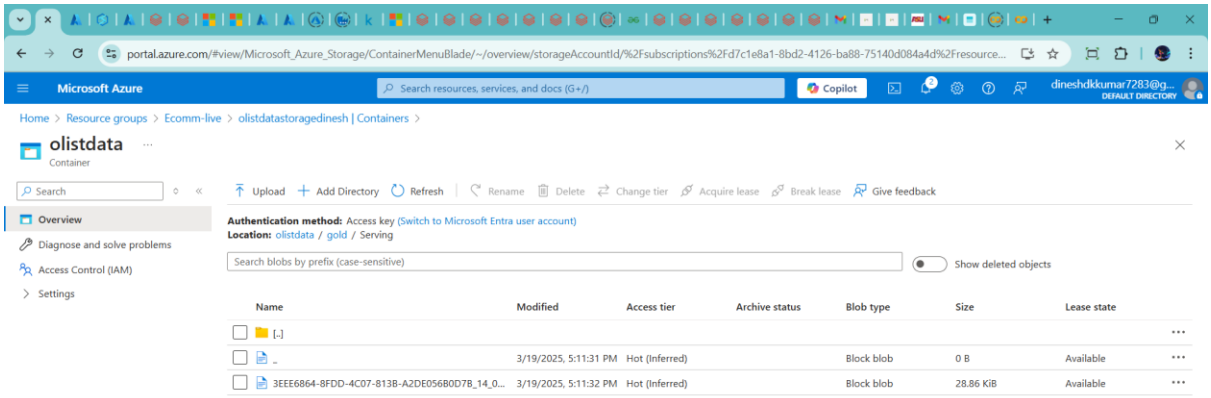
--run all one by one



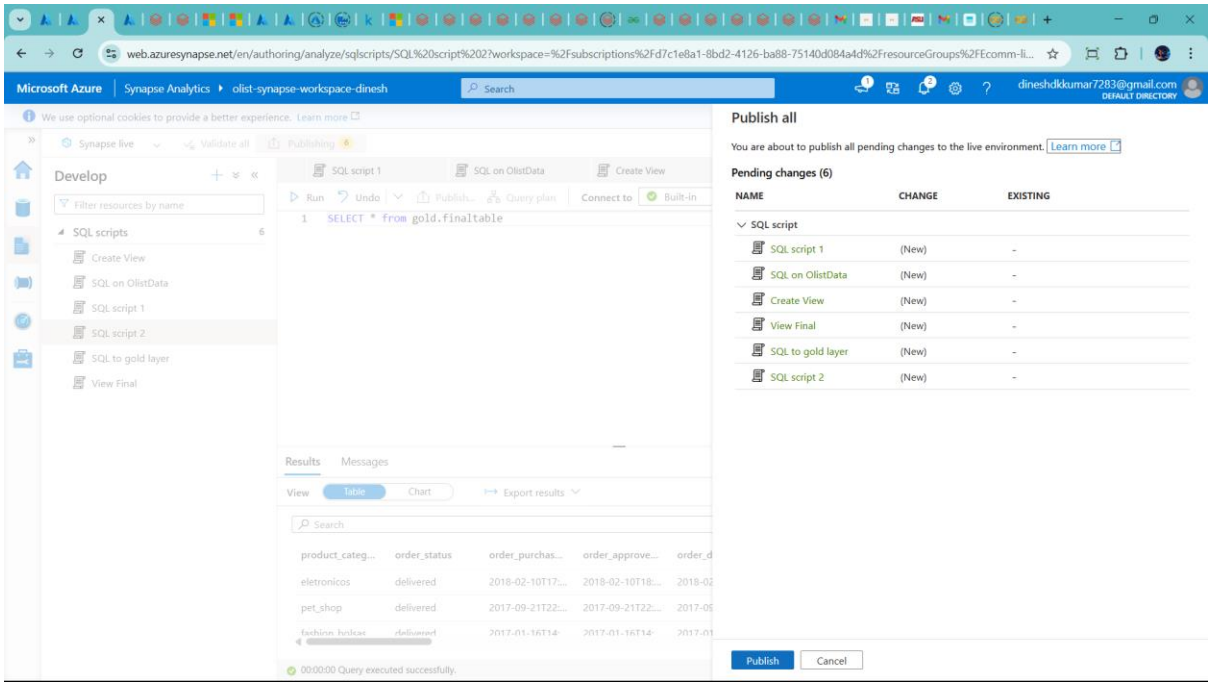
→chage mistake


web.azuresynapse.net/en/authoring/analyze/sqlscripts/SQL%20script%202?workspace=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-li...
Microsoft Azure
Synapse Analytics
olist-synapse-workspace-dinesh
Search
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
We use optional cookies to provide a better experience. Learn more
Accept
Reject
More options
Synapse live
Validate all
Publish all 5
Develop
Filter resources by name
SQL scripts 5
Create View
SQL on OlistData
SQL script 1
SQL to gold layer
View Final
synapsefilesys
SQL script 1
SQL on OlistData
Create View
View Final
SQL to gold layer
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
olist
CREATE EXTERNAL FILE FORMAT extfileformate WITH (
FORMAT_TYPE = PARQUET,
DATA_COMPRESSION = 'org.apache.hadoop.io.compress.SnappyCodec'
);
CREATE EXTERNAL DATA SOURCE goldlayer WITH (
LOCATION = 'https://olistdatastoragedinesh.dfs.core.windows.net/olistdata/gold/',
CREDENTIAL = dineshadmin
);
CREATE EXTERNAL TABLE gold.finaltable WITH (
LOCATION = 'Serving',
DATA_SOURCE = goldlayer,
FILE_FORMAT = extfileformate
) AS
SELECT * FROM gold.final2;
Results
Messages
No results to show
Your query yielded no displayable results
00:00:02 Query executed successfully.
Properties
General
Related (0)
Name *
SQL to gold layer
Description
Type
.sql script
Size
0 bytes
Results settings per query
First 5000 rows (default)
All rows

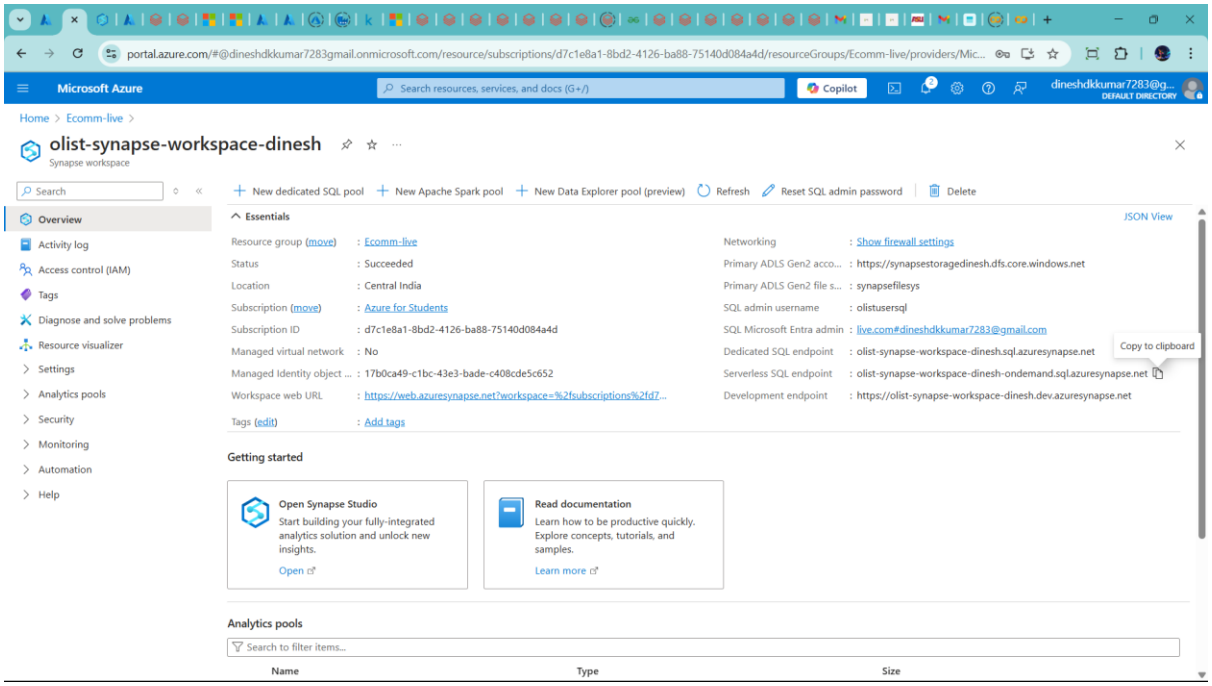

web.azuresynapse.net/en/authoring/analyze/sqlscripts/SQL%20script%202?workspace=%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FEcomm-li...
Microsoft Azure
Synapse Analytics
olist-synapse-workspace-dinesh
Search
dineshdkkumar7283@gmail.com
DEFAULT DIRECTORY
We use optional cookies to provide a better experience. Learn more
Accept
Reject
More options
Synapse live
Validate all
Publish all 6
Develop
Filter resources by name
SQL scripts 6
Create View
SQL on OlistData
SQL script 1
SQL script 2
SQL to gold layer
View Final
SQL script 1
SQL on OlistData
Create View
View Final
SQL to gold layer
SQL script 2
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
olist
SELECT * from gold.finaltable
Results
Messages
View
Table
Chart
Export results
Search
product_categ...
order_status
order_purchas...
order_approve...
order_delivere...
order_delivere...
order_estimat
eletronicos
delivered
2018-02-10T17:...
2018-02-10T18:...
2018-02-19T21:...
2018-02-20T22:...
2018-02-26T0
pet_shop
delivered
2017-09-21T22:...
2017-09-21T22:...
2017-09-22T17:...
2017-09-26T19:...
2017-10-13T0
fashion_bolsas
delivered
2017-01-16T14:
2017-01-16T14:
2017-01-16T15:
2017-01-23T08:
2017-02-24T0
00:00:00 Query executed successfully.
Properties
General
Related (0)
Name *
SQL script 2
Description
Type
.sql script
Size
0 bytes
Results settings per query
First 5000 rows (default)
All rows

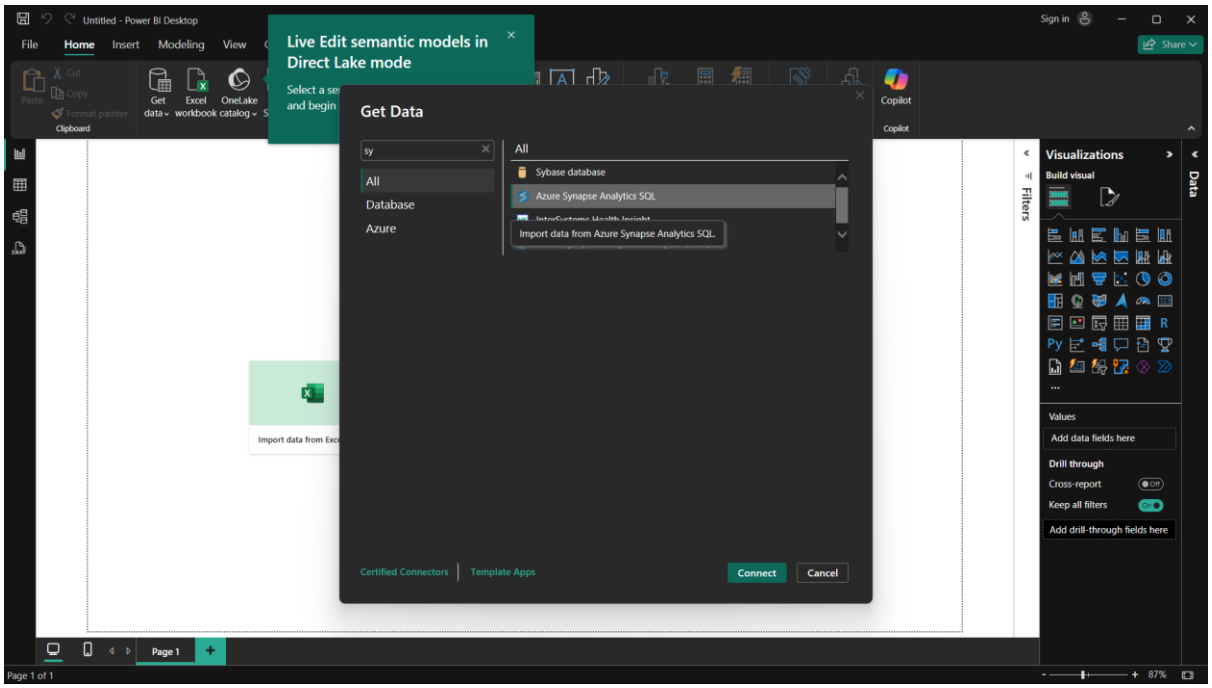


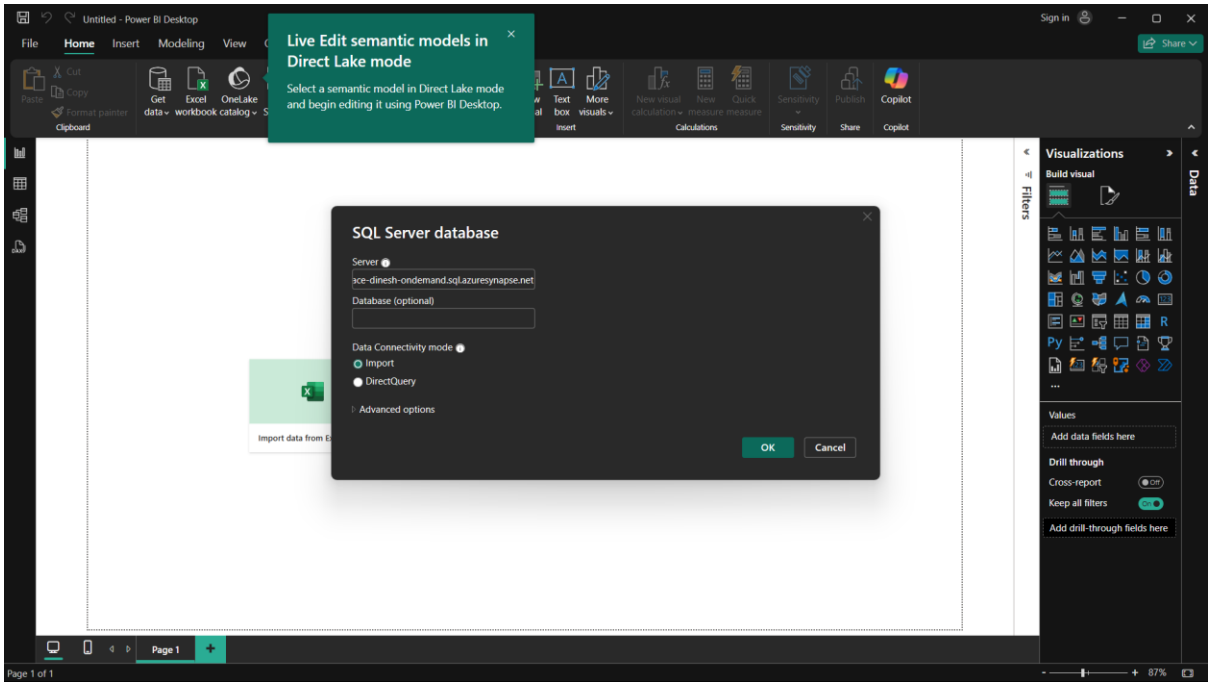
→publish all

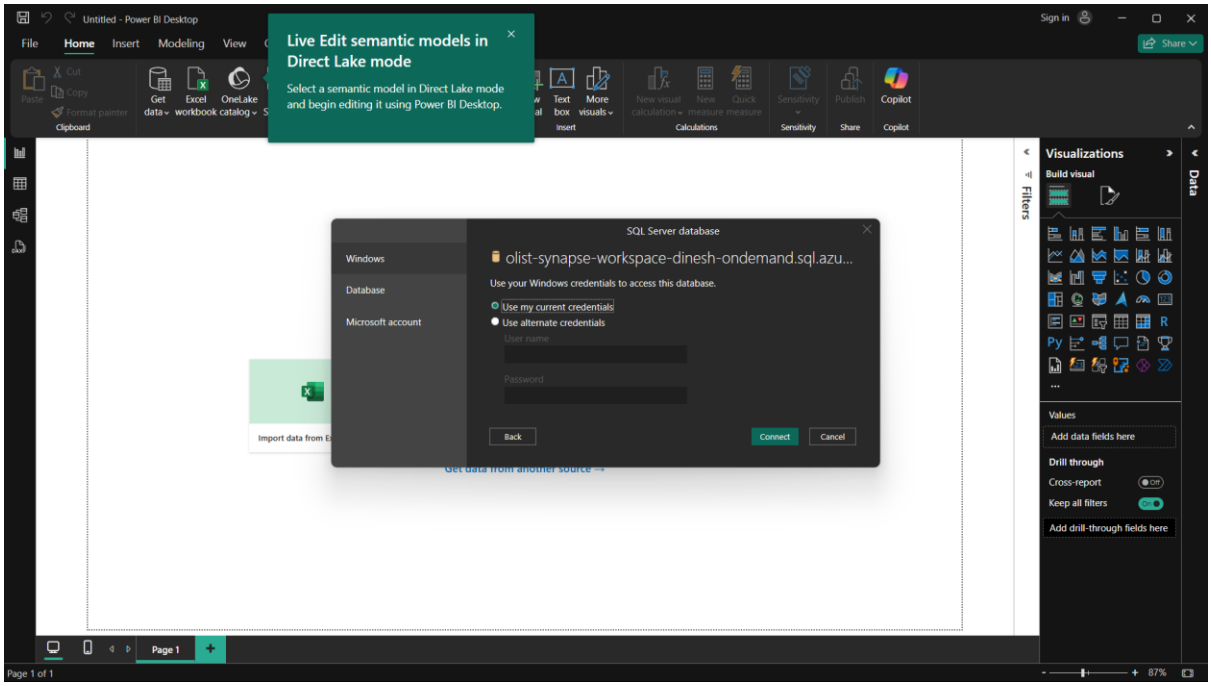


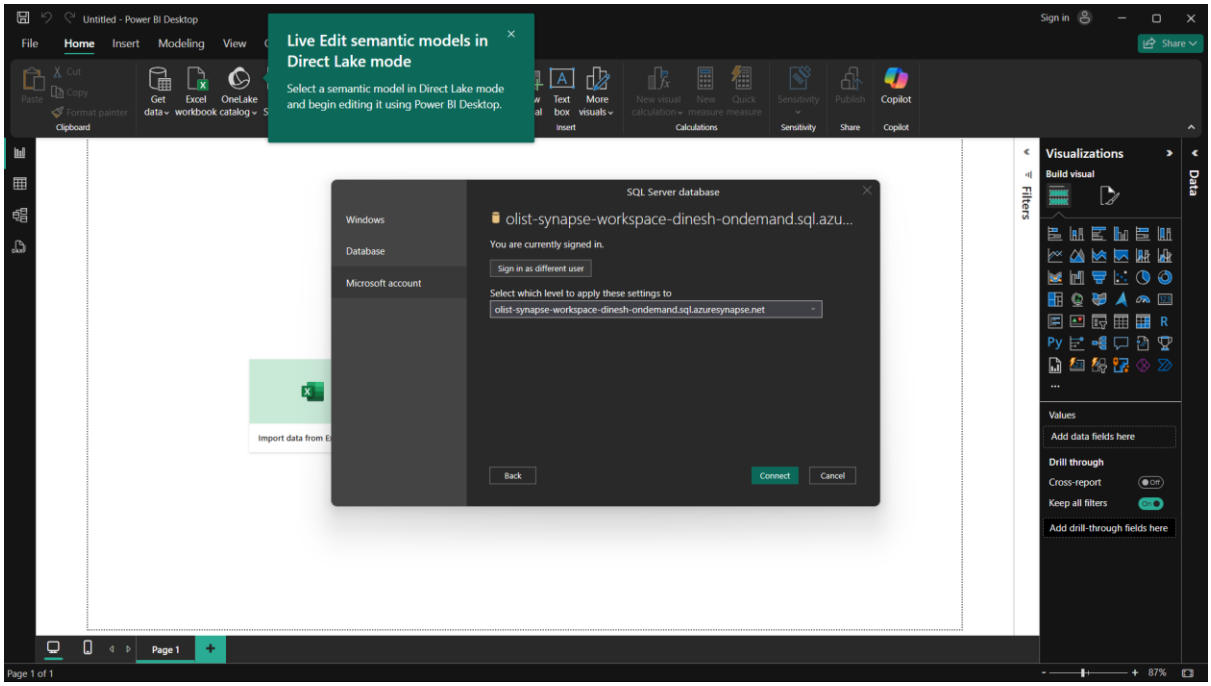
→use this to connect any Bi tools use serverless endpoint

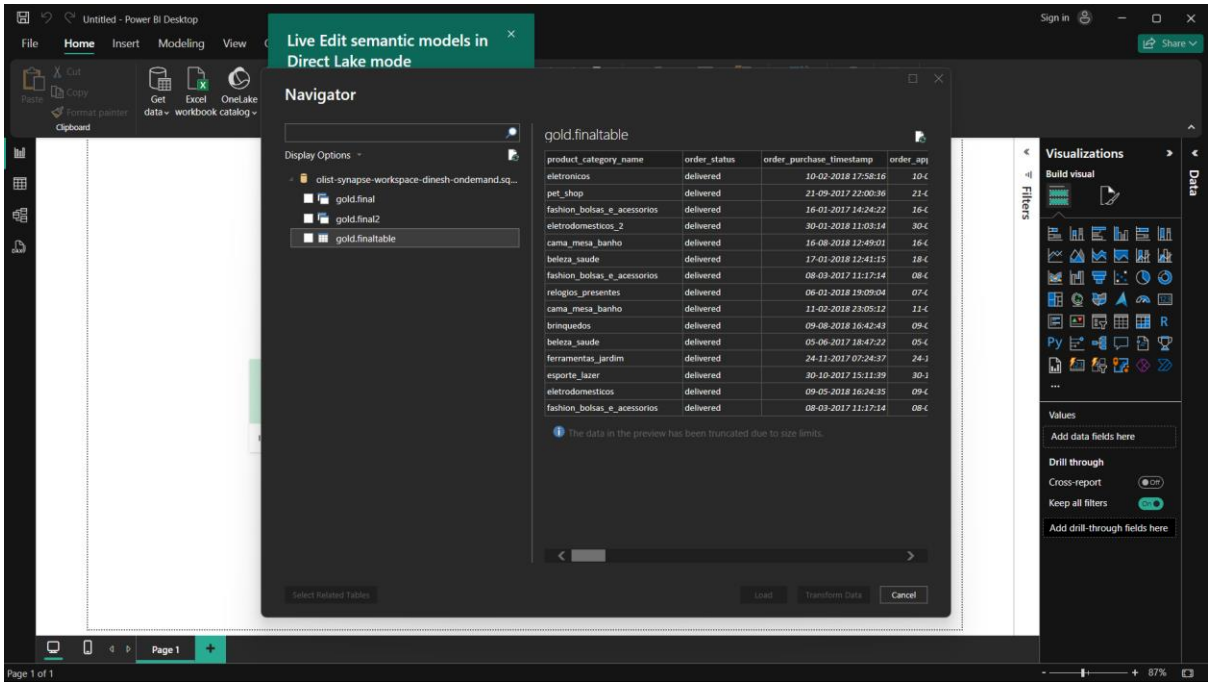

Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
dineshdkkumar7283@g...
DEFAULT DIRECTORY
Home > Ecomm-live >
olist-synapse-workspace-dinesh
Synapse workspace
Search
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Resource visualizer
Settings
Analytics pools
Security
Monitoring
Automation
Help
New dedicated SQL pool
New Apache Spark pool
New Data Explorer pool (preview)
Refresh
Reset SQL admin password
Delete
Essentials
JSON View
Resource group (move) : Ecomm-live
Status : Succeeded
Location : Central India
Subscription (move) : Azure for Students
Subscription ID : d7c1e8a1-8bd2-4126-ba88-75140d084a4d
Managed virtual network : No
Managed Identity object ... : 17b0ca49-c1bc-43e3-bade-c408cde5c652
Workspace web URL : https://web.azuresynapse.net?workspace=%2fsubscriptions%2fd7...
Tags (edit) : Add tags
Networking : Show firewall settings
Primary ADLS Gen2 acco... : https://synapsestoragedinesh.dfs.core.windows.net
Primary ADLS Gen2 file s... : synapsefilesys
SQL admin username : olistusersql
SQL Microsoft Entra admin : live.com#dineshdkkumar7283@gmail.com
Dedicated SQL endpoint : olist-synapse-workspace-dinesh.sql.azuresynapse.net
Serverless SQL endpoint : olist-synapse-workspace-dinesh-ondemand.sql.azuresynapse.net
Development endpoint : https://olist-synapse-workspace-dinesh.dev.azuresynapse.net
Copy to clipboard
Getting started
Open Synapse Studio
Start building your fully-integrated analytics solution and unlock new insights.
Open
Read documentation
Learn how to be productive quickly. Explore concepts, tutorials, and samples.
Learn more
Analytics pools
Search to filter items...
Name
Type
Size

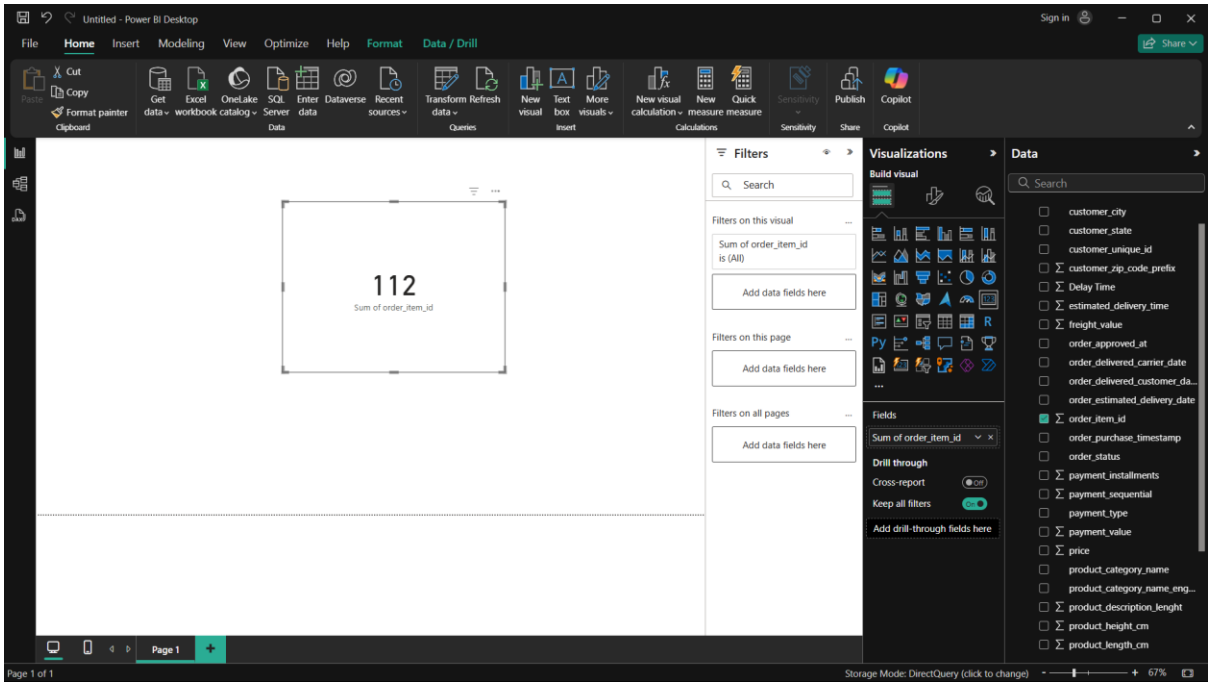

Untitled - Power BI Desktop
Sign in
File
Home
Insert
Modeling
View
Share
Live Edit semantic models in Direct Lake mode
Get Data
sy
All
Database
Azure
Sybase database
Azure Synapse Analytics SQL
Import data from Azure Synapse Analytics SQL
Certified Connectors
Template Apps
Connect
Cancel
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Filters
Data
Page 1
Page 1 of 1
87%


Untitled - Power BI Desktop
Live Edit semantic models in Direct Lake mode
Select a semantic model in Direct Lake mode and begin editing it using Power BI Desktop.
SQL Server database
Server
ace-dinesh-ondemand.sql.azuresynapse.net
Database (optional)
Data Connectivity mode
Import
DirectQuery
Advanced options
OK
Cancel
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Page 1
Page 1 of 1

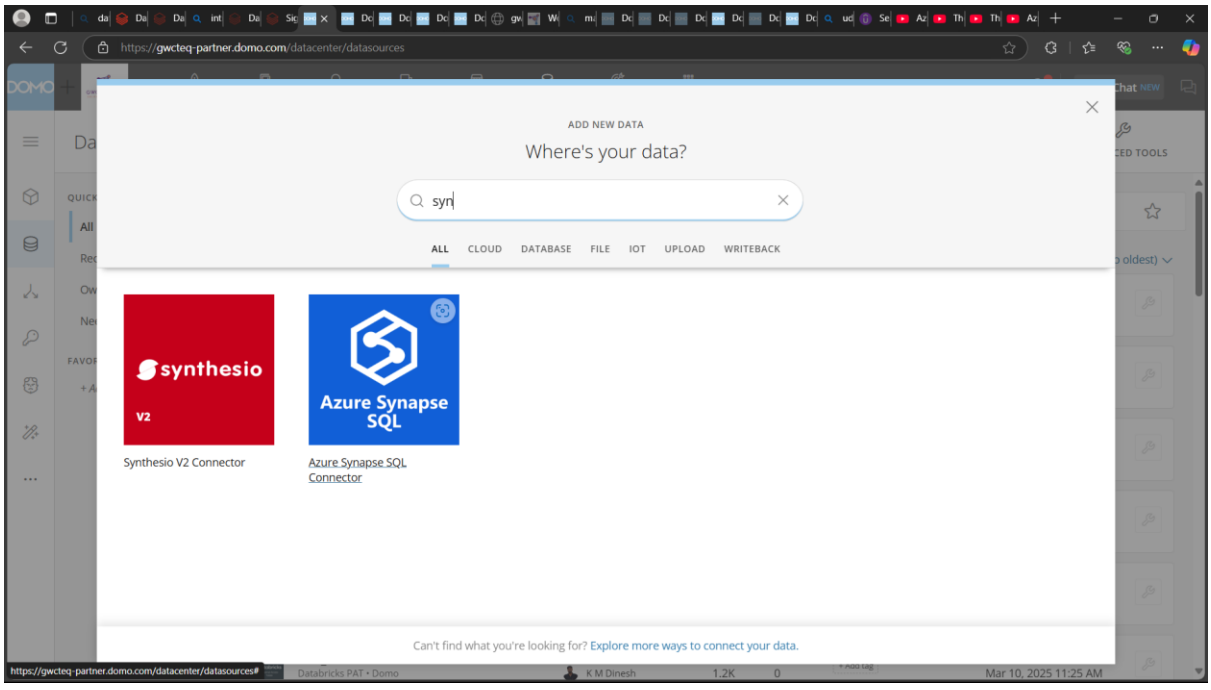

Untitled - Power BI Desktop
Live Edit semantic models in Direct Lake mode
Select a semantic model in Direct Lake mode and begin editing it using Power BI Desktop.
SQL Server database
Windows
Database
Microsoft account
olist-synapse-workspace-dinesh-ondemand.sql.azu...
Use your Windows credentials to access this database.
Use my current credentials
Use alternate credentials
User name
Password
Back
Connect
Cancel
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Page 1
Page 1 of 1

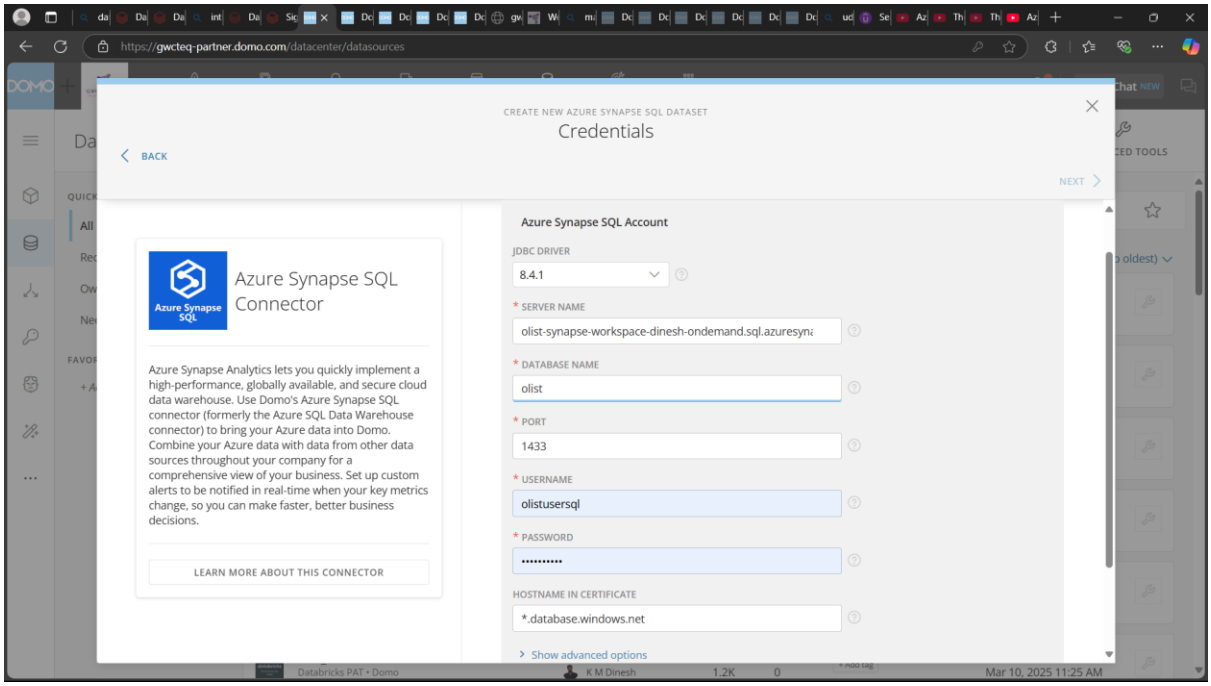

Untitled - Power BI Desktop
Live Edit semantic models in Direct Lake mode
Select a semantic model in Direct Lake mode and begin editing it using Power BI Desktop.
SQL Server database
Windows
Database
Microsoft account
olist-synapse-workspace-dinesh-ondemand.sql.azu...
You are currently signed in.
Sign in as different user
Select which level to apply these settings to
olist-synapse-workspace-dinesh-ondemand.sql.azuresynapse.net
Back
Connect
Cancel
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Page 1
Page 1 of 1

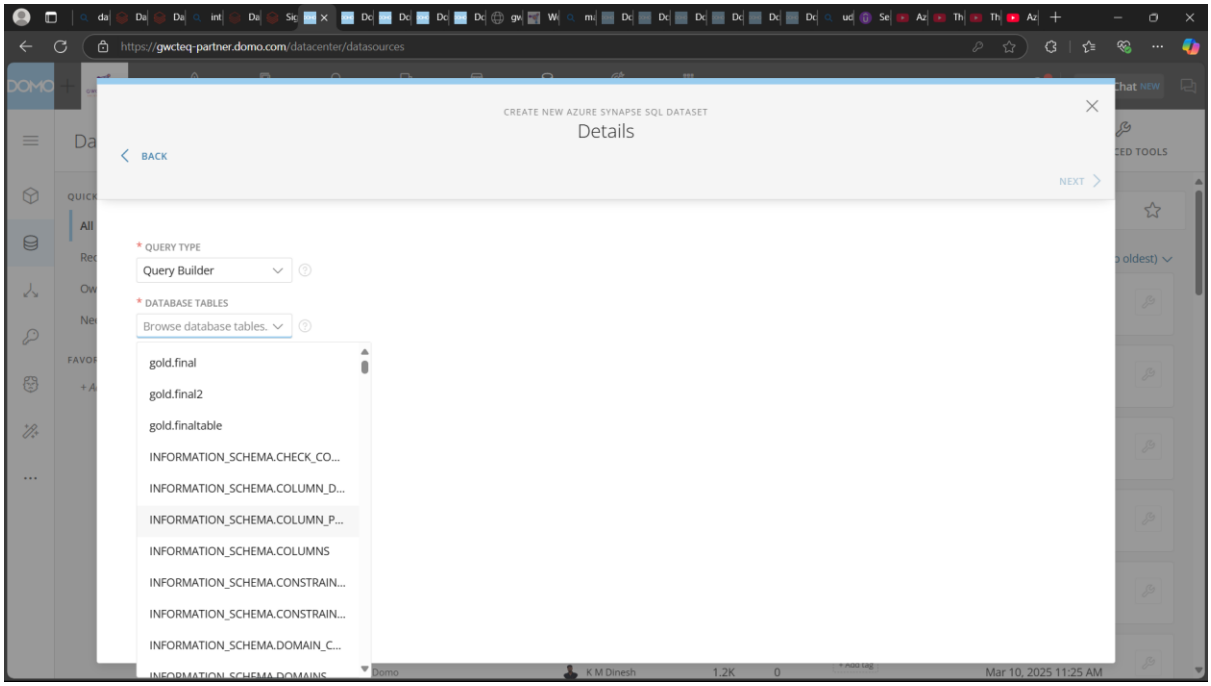

Untitled - Power BI Desktop
Live Edit semantic models in Direct Lake mode
Navigator
Display Options
olist-synapse-workspace-dinesh-ondemand.sq...
gold.final
gold.final2
gold.finaltable
gold.finaltable
product_category_name
order_status
order_purchase_timestamp
eletronicos
delivered
pet_shop
delivered
fashion_bolsas_e_acessorios
delivered
eletrodomesticos_2
delivered
cama_mesa_banho
delivered
beleza_saude
delivered
fashion_bolsas_e_acessorios
delivered
relogios_presentes
delivered
cama_mesa_banho
delivered
brinquedos
delivered
beleza_saude
delivered
ferramentas_jardim
delivered
esporte_lazer
delivered
eletrodomesticos
delivered
fashion_bolsas_e_acessorios
delivered
The data in the preview has been truncated due to size limits.
Select Related Tables
Load
Transform Data
Cancel
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Page 1
Page 1 of 1

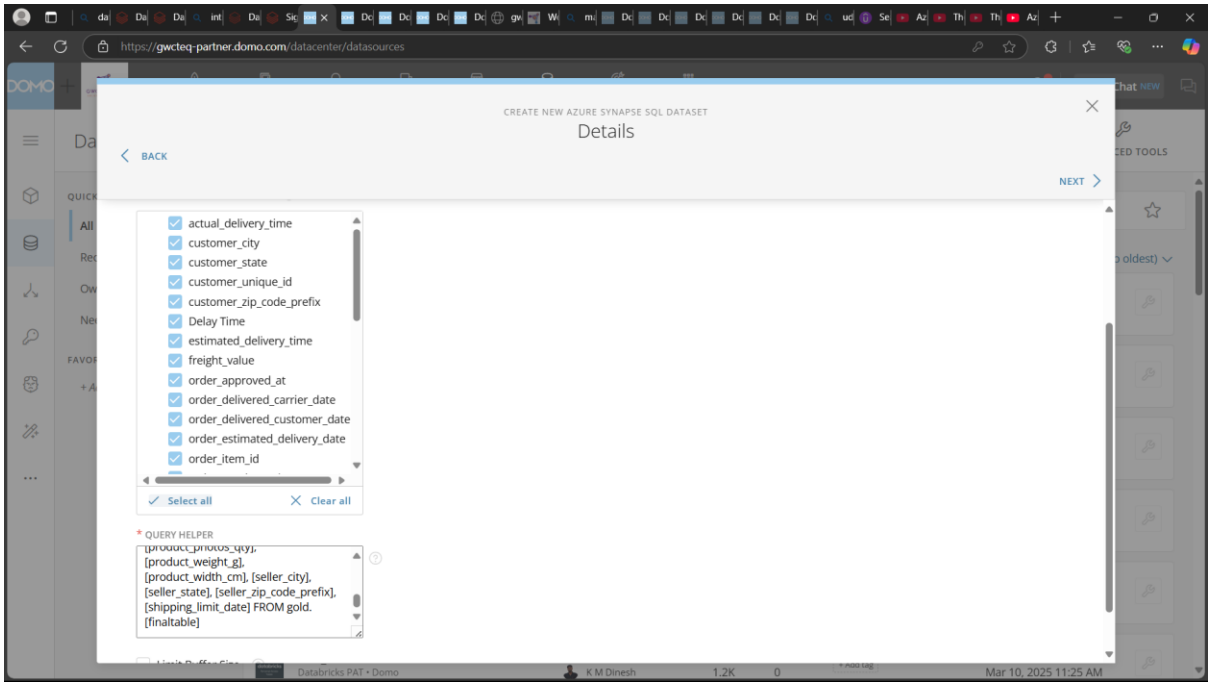


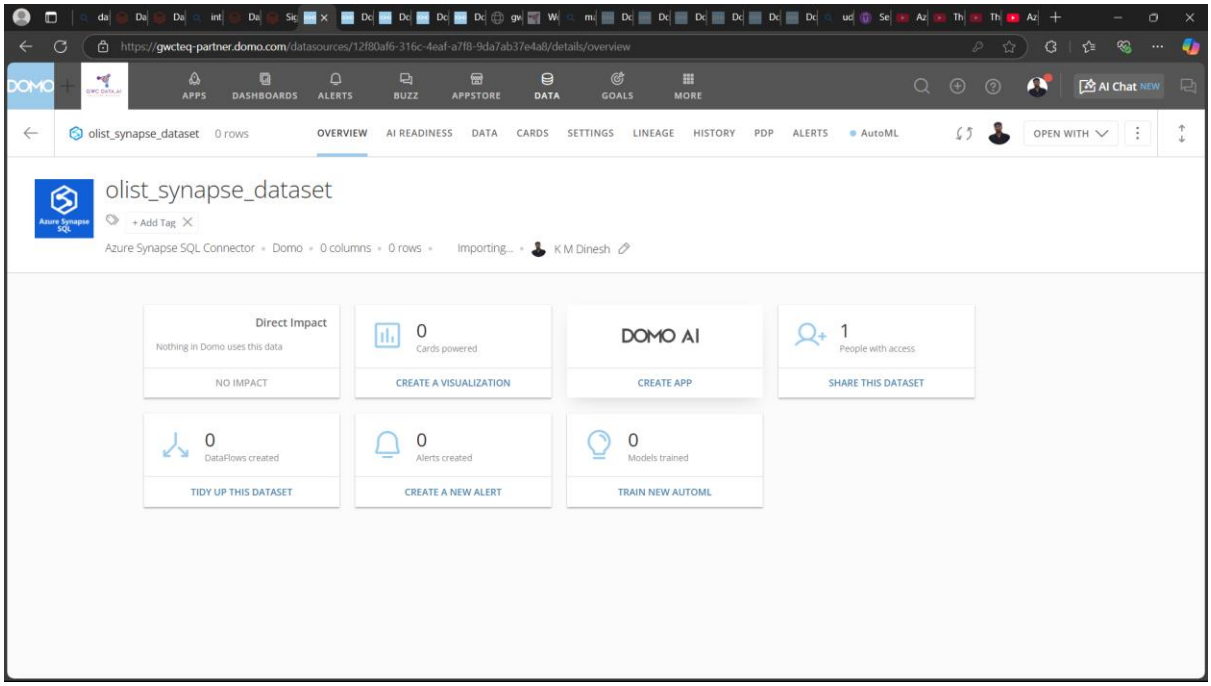
Will connect with DOMO

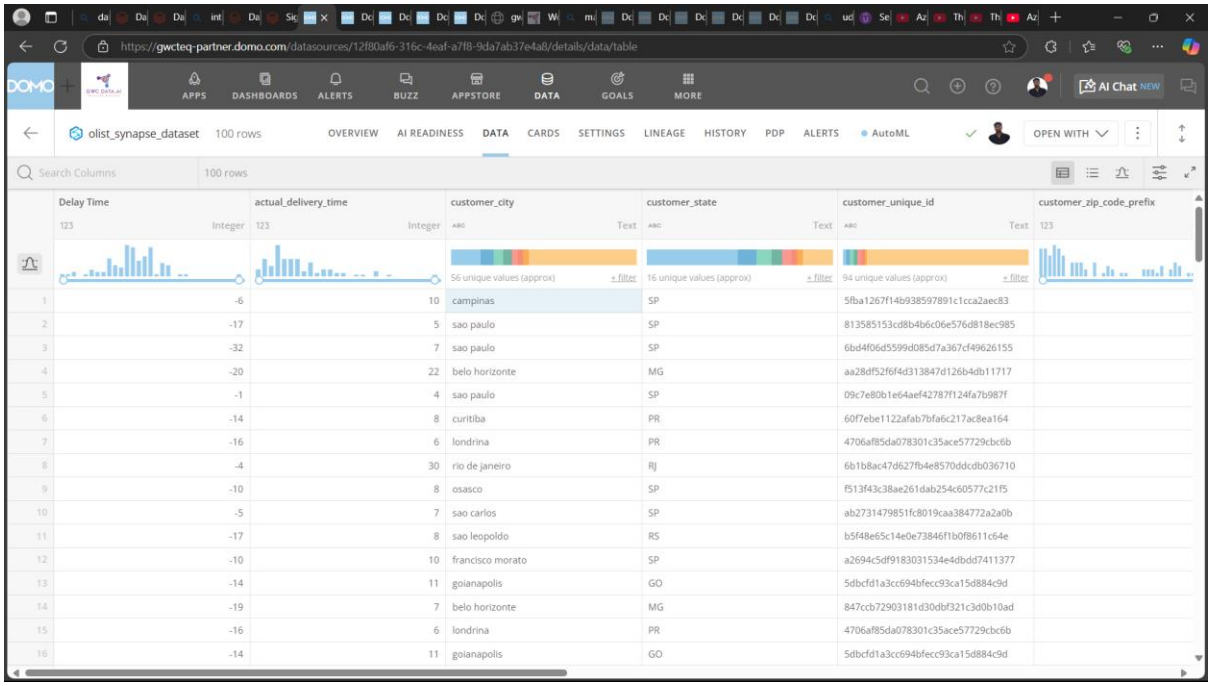



https://gwcteq-partner.domo.com/datacenter/datasources
CREATE NEW AZURE SYNAPSE SQL DATASET
Credentials
BACK
NEXT
Azure Synapse SQL Connector
Azure Synapse Analytics lets you quickly implement a high-performance, globally available, and secure cloud data warehouse. Use Domo's Azure Synapse SQL connector (formerly the Azure SQL Data Warehouse connector) to bring your Azure data into Domo. Combine your Azure data with data from other data sources throughout your company for a comprehensive view of your business. Set up custom alerts to be notified in real-time when your key metrics change, so you can make faster, better business decisions.
LEARN MORE ABOUT THIS CONNECTOR
Azure Synapse SQL Account
JDBC DRIVER
8.4.1
* SERVER NAME
olist-synapse-workspace-dinesh-ondemand.sql.azuresyn
* DATABASE NAME
olist
* PORT
1433
* USERNAME
olistusersql
* PASSWORD
HOSTNAME IN CERTIFICATE
*.database.windows.net
Show advanced options


https://gwcteq-partner.domo.com/datacenter/datasources
CREATE NEW AZURE SYNAPSE SQL DATASET
Details
BACK
NEXT
* QUERY TYPE
Query Builder
* DATABASE TABLES
Browse database tables.
gold.final
gold.final2
gold.finaltable
INFORMATION_SCHEMA.CHECK_CO...
INFORMATION_SCHEMA.COLUMN_D...
INFORMATION_SCHEMA.COLUMN_P...
INFORMATION_SCHEMA.COLUMNS
INFORMATION_SCHEMA.CONSTRAIN...
INFORMATION_SCHEMA.CONSTRAIN...
INFORMATION_SCHEMA.DOMAIN_C...


https://gwcteq-partner.domo.com/datacenter/datasources
CREATE NEW AZURE SYNAPSE SQL DATASET
Details
BACK
NEXT
actual_delivery_time
customer_city
customer_state
customer_unique_id
customer_zip_code_prefix
Delay Time
estimated_delivery_time
freight_value
order_approved_at
order_delivered_carrier_date
order_delivered_customer_date
order_estimated_delivery_date
order_item_id
Select all
Clear all
QUERY HELPER
[product_weight_g],
[product_width_cm], [seller_city],
[seller_state], [seller_zip_code_prefix],
[shipping_limit_date] FROM gold.
[finaltable]
Databricks PAT • Domo
K M Dinesh
1.2K
0
Mar 10, 2025 11:25 AM


https://gwcteq-partner.domo.com/datasources/12f80af6-316c-4eaf-a7f8-9da7ab37e4a8/details/overview
APPS
DASHBOARDS
ALERTS
BUZZ
APPSTORE
DATA
GOALS
MORE
AI Chat
olist_synapse_dataset 0 rows
OVERVIEW
AI READINESS
DATA
CARDS
SETTINGS
LINEAGE
HISTORY
PDP
ALERTS
AutoML
OPEN WITH
olist_synapse_dataset
+ Add Tag
Azure Synapse SQL Connector • Domo • 0 columns • 0 rows • Importing... • K M Dinesh
Direct Impact
Nothing in Domo uses this data
NO IMPACT
0
Cards powered
CREATE A VISUALIZATION
DOMO AI
CREATE APP
1
People with access
SHARE THIS DATASET
0
DataFlows created
TIDY UP THIS DATASET
0
Alerts created
CREATE A NEW ALERT
0
Models trained
TRAIN NEW AUTOML



Use this dataset to create dashboard