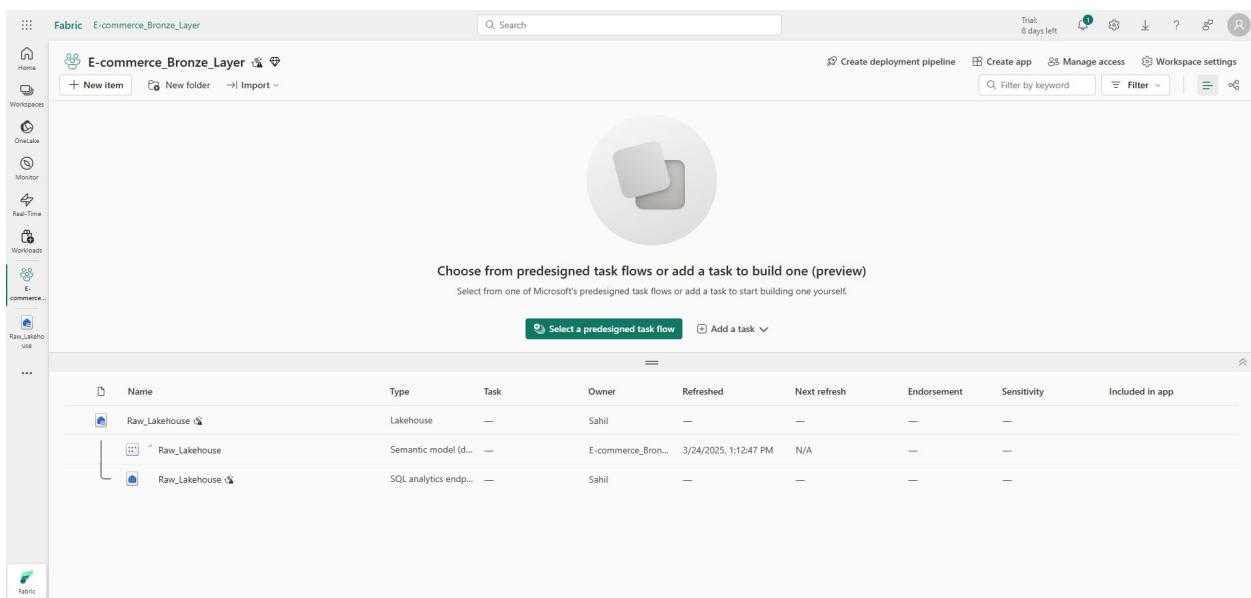


# Microsoft Fabric with Medallion Architecture (Bronze-Silver-Gold)

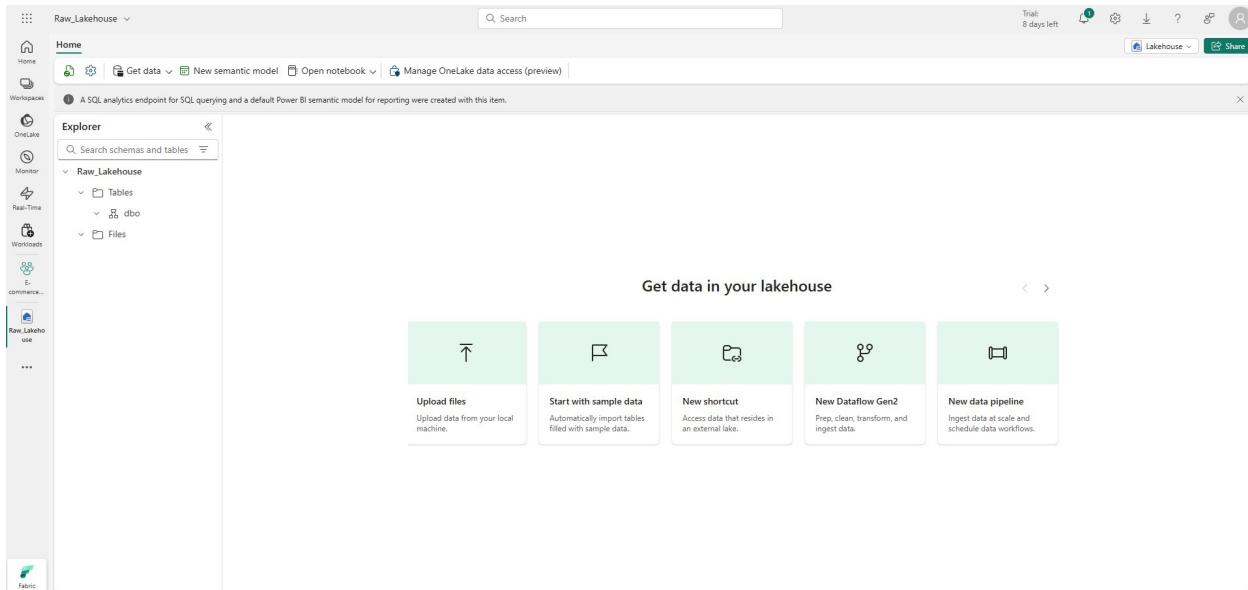
## 1. Created e-commerce bronze layer



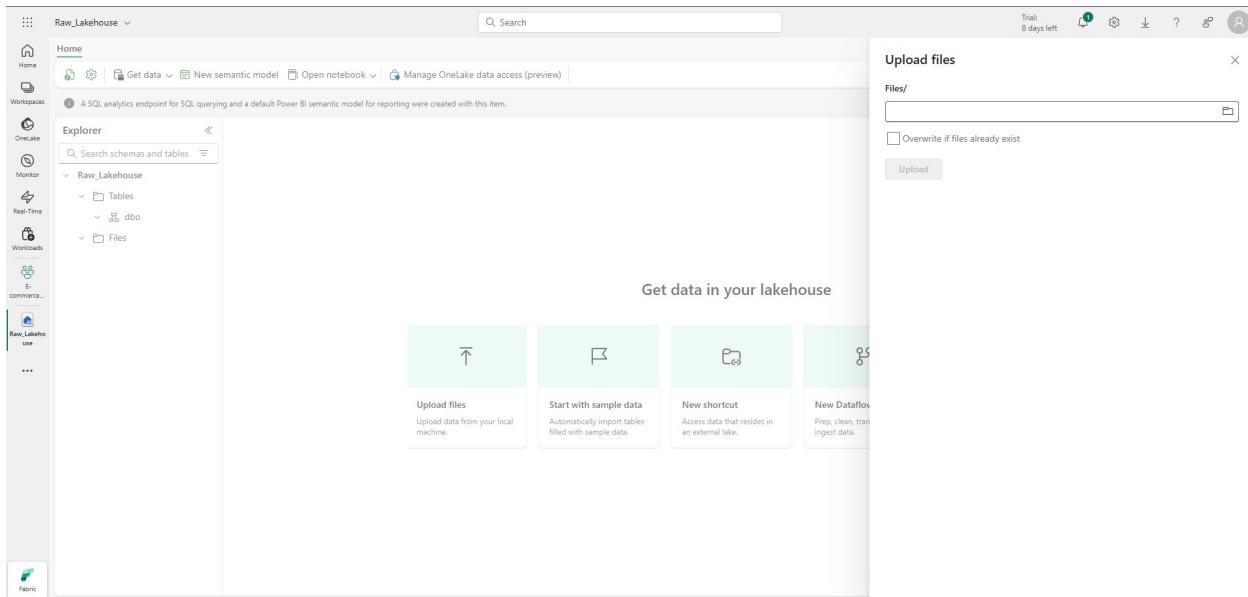
The screenshot shows the Microsoft Fabric workspace interface for the 'E-commerce\_Bronze\_Layer' workspace. The left sidebar includes links for Home, OneLake, Monitor, Real-Time, Workloads, E-commerce, and Raw\_Lakehouse. The main area displays a large circular placeholder icon with two overlapping squares. Below it, a message reads: 'Choose from predesigned task flows or add a task to build one (preview)' and 'Select from one of Microsoft's predesigned task flows or add a task to start building one yourself.' A green button labeled 'Select a predesigned task flow' is visible. At the bottom, there is a table listing three items:

Name	Type	Task	Owner	Refreshed	Next refresh	Endorsement	Sensitivity	Included in app
Raw_Lakehouse	Lakehouse	—	Sahil	—	—	—	—	—
Raw_Lakehouse	Semantic model (d...)	—	E-commerce_Bron...	3/24/2025, 1:12:47 PM	N/A	—	—	—
Raw_Lakehouse	SQL analytics endp...	—	Sahil	—	—	—	—	—

## 2. Created Lakehouse in the bronze layer



### 3. Uploading csv files to Lakehouse



A SQL analytics endpoint for SQL querying and a default Power BI semantic model for reporting were created with this item.

Get data in your lakehouse

Upload files  
Upload data from your local machine.

Start with sample data  
Automatically import tables filled with sample data.

New shortcut  
Access data that resides in an external lake.

New Dataflow  
Prep, clean, then ingest data.

Upload files

Overwrite if files already exist

Upload

Current uploads

File Name	Lakehouse Name	Size	Actions
Customers_(1).csv	Raw_Lakehouse	333 B / 333 B	<input checked="" type="checkbox"/> <input type="button" value="Delete"/>
Geographical.csv	Raw_Lakehouse	176 B / 176 B	<input checked="" type="checkbox"/> <input type="button" value="Delete"/>
Orders_(1).csv	Raw_Lakehouse	241 B / 241 B	<input checked="" type="checkbox"/> <input type="button" value="Delete"/>
Products_(1).csv	Raw_Lakehouse	219 B / 219 B	<input checked="" type="checkbox"/> <input type="button" value="Delete"/>
Transactions.csv	Raw_Lakehouse	229 B / 229 B	<input checked="" type="checkbox"/> <input type="button" value="Delete"/>

#### 4. Files are present in Lakehouse

A SQL analytics endpoint for SQL querying and a default Power BI semantic model for reporting were created with this item.

Files

Name	Date modified	Type	Size
Customers.csv	3/24/2025, 1:15:53 PM	csv	333 B
Geographical.csv	3/24/2025, 1:15:53 PM	csv	176 B
Orders.csv	3/24/2025, 1:15:53 PM	csv	241 B
Products.csv	3/24/2025, 1:15:53 PM	csv	219 B
Transactions.csv	3/24/2025, 1:15:53 PM	csv	229 B

## **5. Separate workspace for e-commerce silver layer**

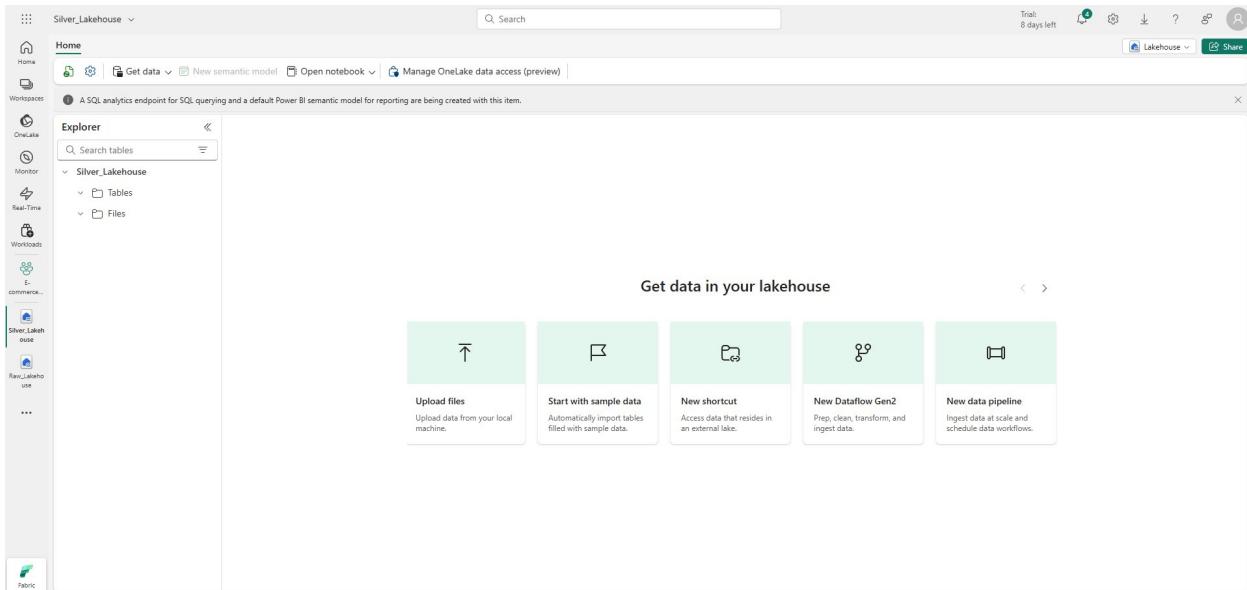
The screenshot shows two views of the Microsoft Fabric E-commerce\_Silver\_Layer workspace.

**Top View:** A large, empty circular canvas with a central paperclip icon. Below it, the text "There's nothing here yet" and "Add something new, or upload something to see them here." A green button at the top says "Select a predesigned task flow".

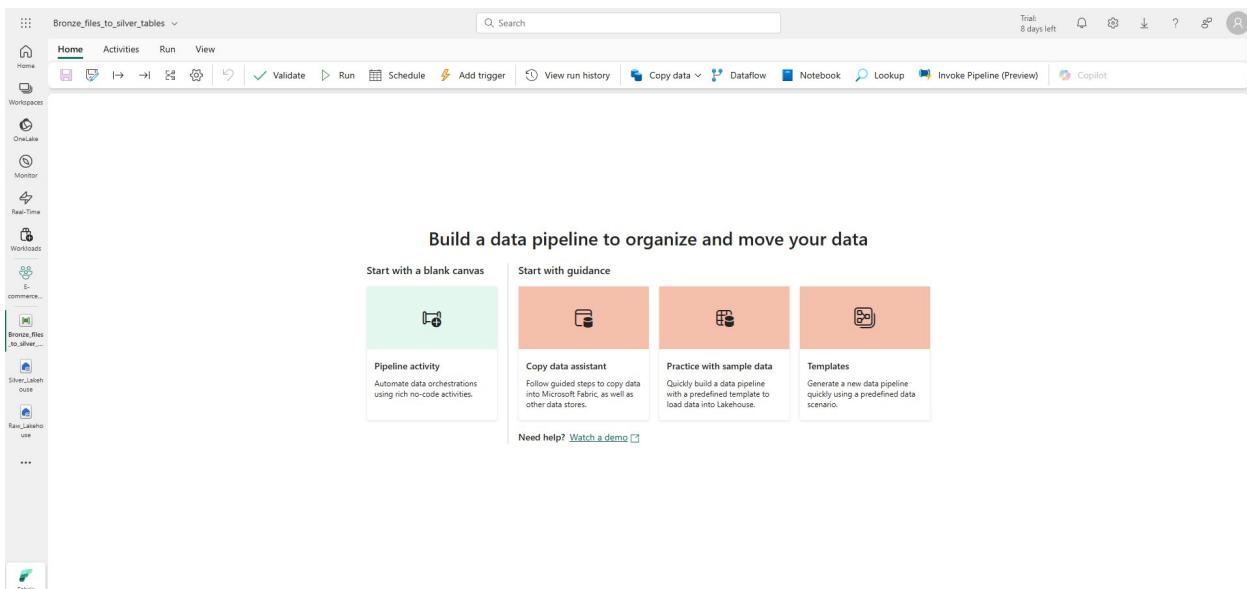
**Bottom View:** A table listing various assets:

Name	Type	Task	Owner	Refreshed	Next refresh	Endorsement	Sensitivity	Included in app
Bronze_files_to_silver_tables	Data pipeline	—	Sahil	—	—	—	—	
Notebook 2	Notebook	—	Sahil	—	—	—	—	
Silver_Lakehouse	Lakehouse	—	Sahil	—	—	—	—	
Silver_Lakehouse	Semantic model (d...)	—	E-commerce_Silver...	3/24/2025, 1:22:04 PM	N/A	—	—	
Silver_Lakehouse	SQL analytics endp...	—	Sahil	—	—	—	—	

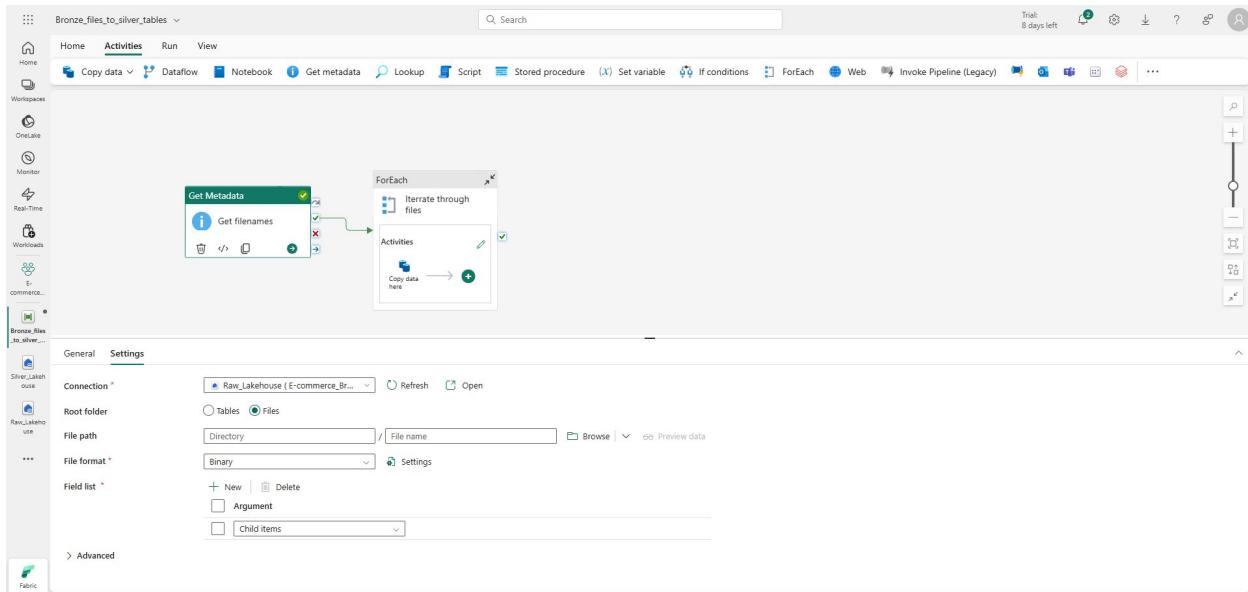
## **6. Load data from bronze to silver layer and want to load as tables not files Created a silver Lakehouse to store all the tables**



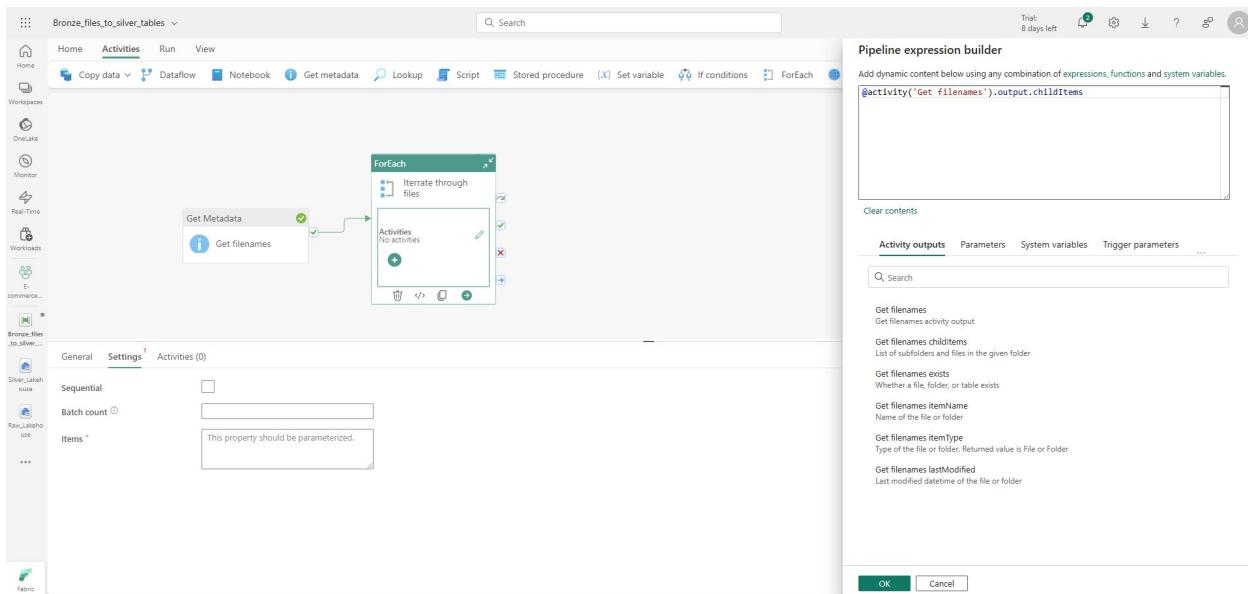
## **7. Created a data pipeline to convert all csv files into tables**



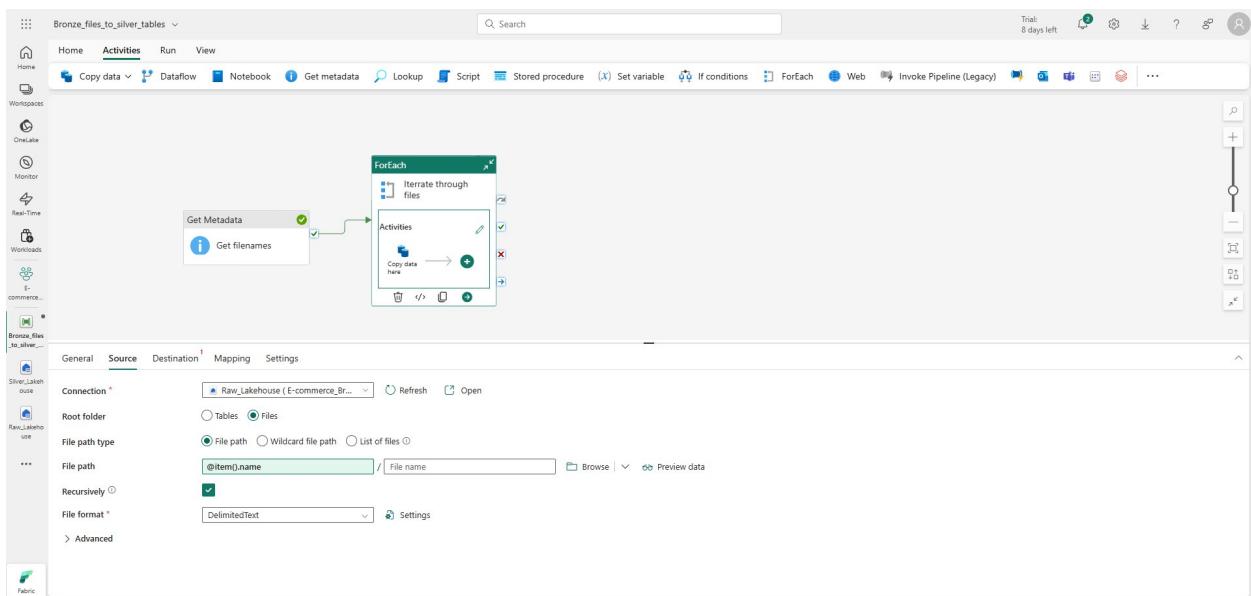
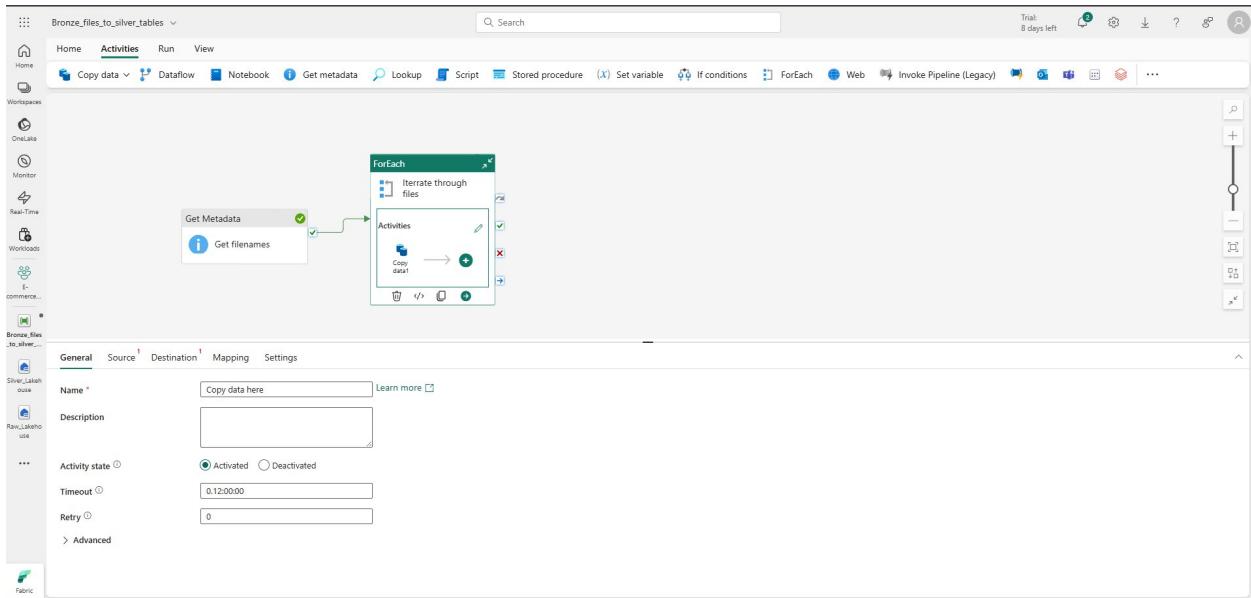
## **8. Get filenames from get meta data all files are present in array child items inside child items there is name property**

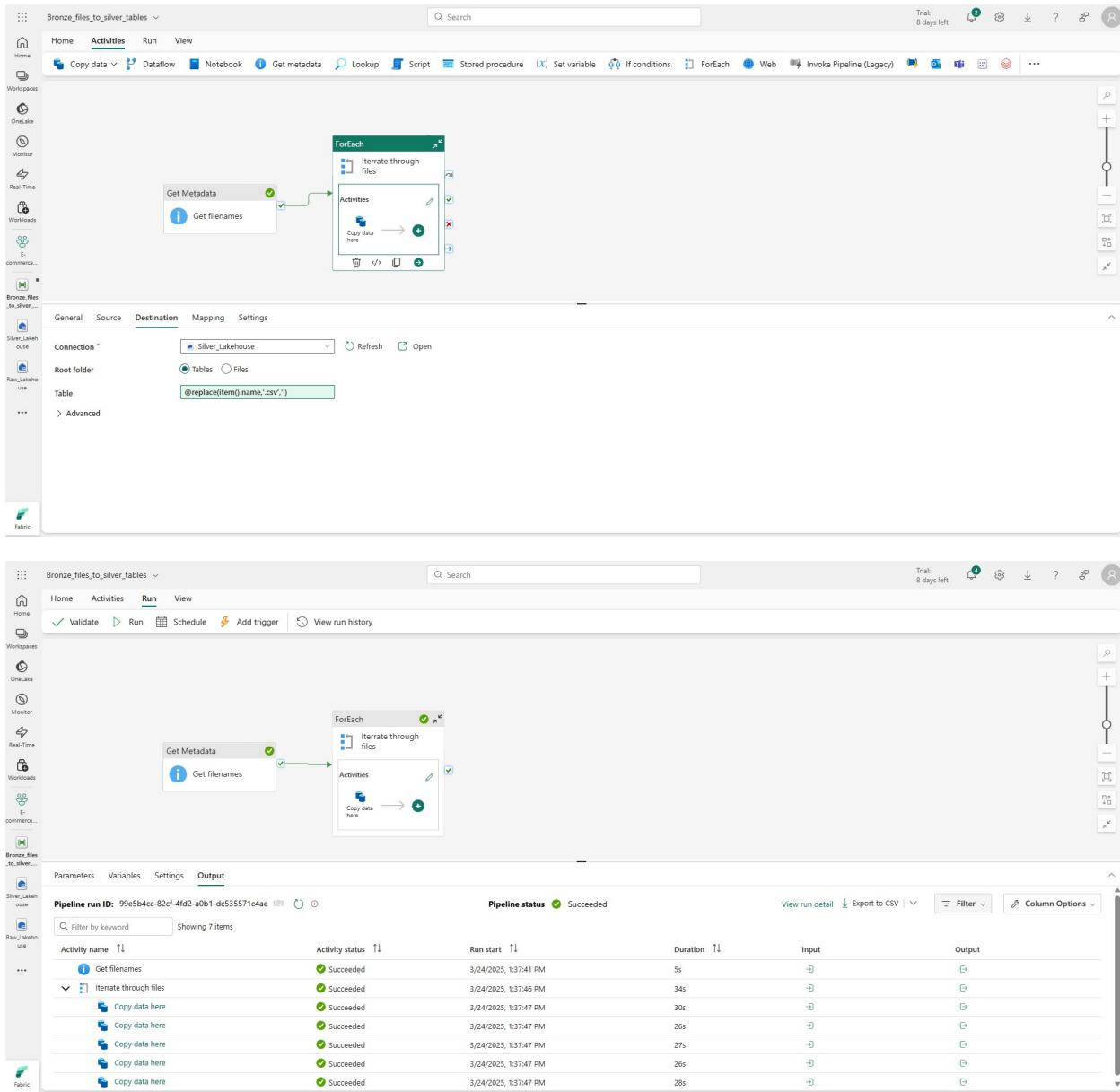


## 9. Iterate through child items used foreach

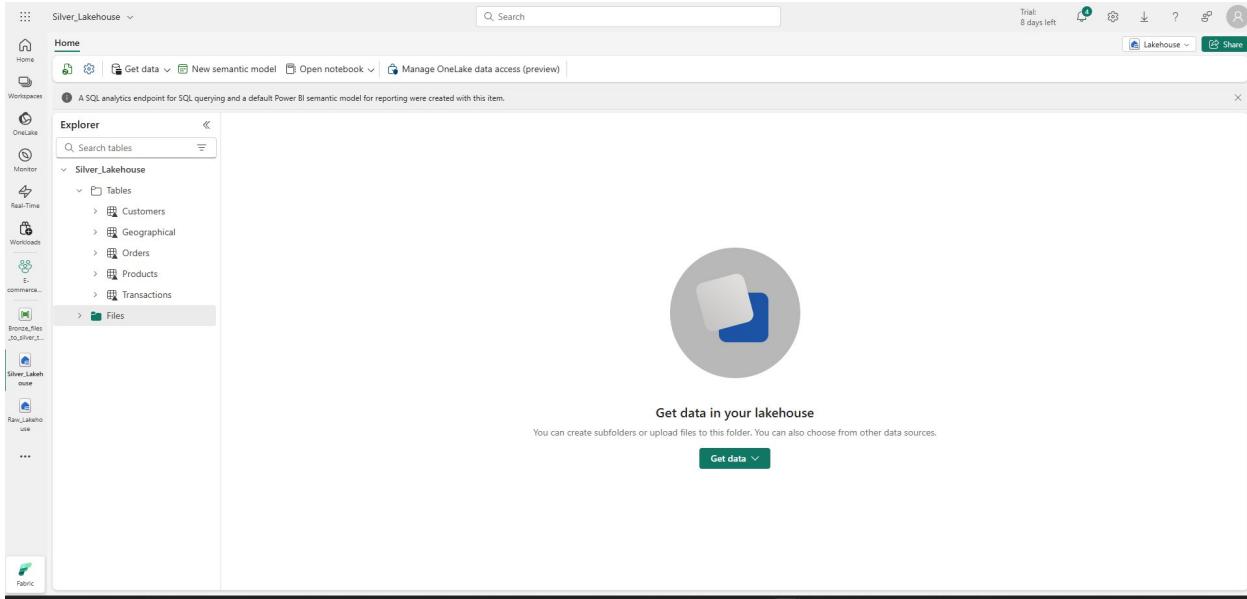


## 10. Copied entire data from source to destination





## 11. The tables are present in silver Lakehouse



## **12. Now Need to do transformation on data can do this with dataflow gen 2 or Notebook all transformations happens on silver layer**

**New item**

- Get data**  
Ingest batch and Notebook data into a single location within your Fabric workspace.
- Notebook**  
Explore, analyze, and visualize data and build ML models. Supports Apache Spark, Python, T-SQL, and more.
- Prepare data**  
Clean, transform, extract, and load your data for analysis and modeling tasks.
- Analyze and train data**  
Propose hypotheses, train models, and explore your data to make decisions and predictions.
- Develop data**  
Create and build your software, applications, and data solutions.

Name	Type	Task	Owner
Bronze_files_to_silver_tables	Data pipeline	—	Sahil
Silver_Lakehouse	Lakehouse	—	Sahil
Silver_Lakehouse	Semantic model (d...)	—	E-commerce
Silver_Lakehouse	SQL analytics endp...	—	Sahil

## **13. Connect your Lakehouse**

The screenshot shows the Databricks Notebook interface. In the center, a modal dialog box titled "Add lakehouses" is open. It contains three options:

- New lakehouse
- Existing lakehouse(s) with Schema (i)
- Existing lakehouse(s) without Schema (i)

At the bottom right of the dialog are two buttons: "Add" and "Cancel".

The background of the notebook shows a cell with the following Python code:

```
1 # Welcome to your new notebook
2 # Type here in the cell editor to add code!
3
```

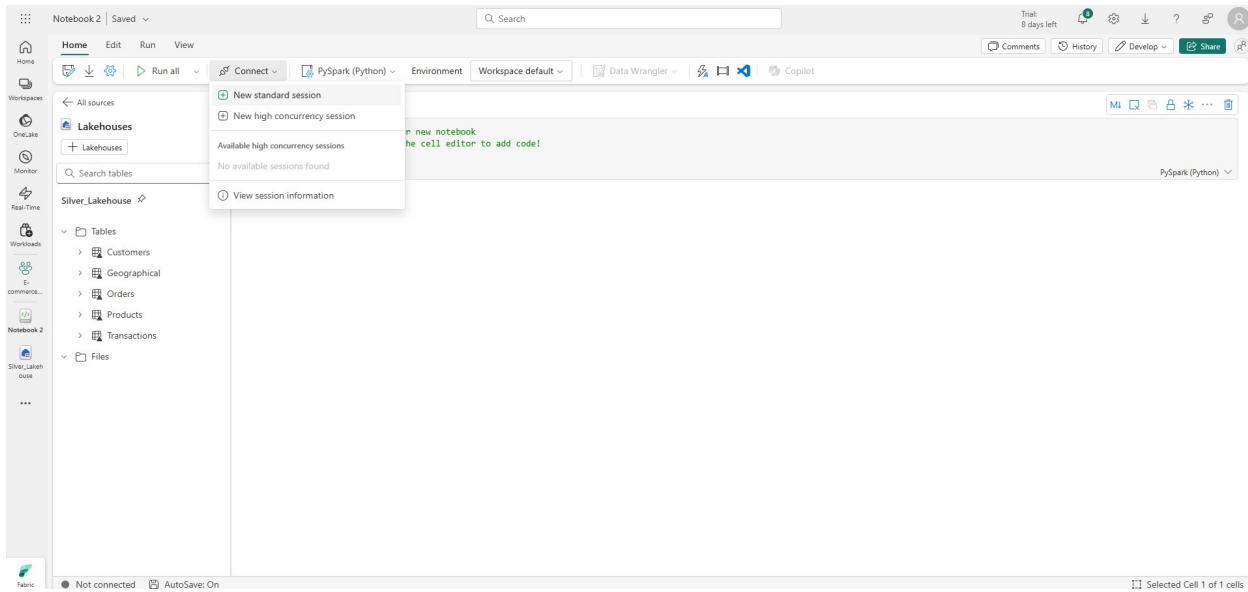
The left sidebar shows a "Lakehouses" section with a "+ Lakehouses" button. The main workspace has a "Add lakehouses" button at the bottom.

Below this, another screenshot of the Databricks Notebook interface is shown, titled "Notebook 2". It displays an "OneLake catalog" with a table of lakehouses:

Name	Owner	Refreshed	Location	Endorsement	Sensitivity
Silver_Lakehouse	Sahil	—	E-commerce_Silver...	—	—
Final_LH	Sahil	—	Dev_WS	—	—
Landing_LH	Sahil	—	Dev_WS	—	—
King_LH	Sahil	—	Test	—	—
testt	Sahil	—	Info Tech Company ...	—	—
copy_LH	Sahil	—	Info Tech Company ...	—	—
Golden_LH	Sahil	—	Info Tech Company ...	—	—
DataflowsStagingLakehouse	Sahil	—	Info Tech Company ...	—	—
DataflowsStagingLakehouse	Sahil	—	Test	—	—
DataflowsStagingLakehouse	Sahil	—	E_Comm_Golden.lay...	—	—
DataflowsStagingLakehouse	Sahil	—	Dev_WS	—	—

The "Add" and "Cancel" buttons are located at the bottom right of the catalog view.

## 14. Connect spark session



## 15. Now doing transformations

The screenshot shows the Databricks interface with a 'Standard session' selected. Five cells of PySpark code are visible, each displaying a SQL query to select all columns from a specific table:

- Cell 1: `cust\_df = spark.sql("SELECT \* FROM Silver\_Lakehouse.Customers")`
- Cell 2: `geo\_df = spark.sql("SELECT \* FROM Silver\_Lakehouse.Geographical")`
- Cell 3: `ord\_df = spark.sql("SELECT \* FROM Silver\_Lakehouse.Orders")`
- Cell 4: `prod\_df = spark.sql("SELECT \* FROM Silver\_Lakehouse.Products")`
- Cell 5: `trans\_df = spark.sql("SELECT \* FROM Silver\_Lakehouse.Transactions")`

The bottom status bar shows 'Session ready' and 'AutoSave: On'.

Notebook 2 | Saved

Home Edit Run View Standard session PySpark (Python) Environment Workspace default Data Wrangler Copilot

All sources Lakehouses + Lakehouses Search tables Silver\_Lakehouse Tables Customers Geographical Orders Products Transactions Files

```
[12] 1 transactions_df = spark.sql("SELECT * FROM Silver_Lakehouse.Transactions")
2 display(transactions_df)
✓ 1 sec - Command executed in 1 sec 501 ms by Sahil on 2:01:10 PM, 3/24/25
> Spark jobs (1 of 1 succeeded) Resources Log
```

Table Data Wrangler 6 columns, 4 rows Download Search

AMC TransactionID	AMC OrderID	AMC PaymentMethod	AMC TransactionDate	AMC Amount	AMC Currency
1 5001	1001	Credit Card	2024-01-15	250.5	USD
2 5002	1002	PayPal	2024-01-20	120.0	USD
3 5003	1003	Bank Transfer	2024-02-01	540.3	CAD
4 5004	1004	Cash	2024-02-10	75.2	MXN

```
[13] 1 from pyspark.sql.functions import col, when, to_date, regexp_replace
2 # Transformations in Transactional Table
3 # Change ID to Transaction_ID
4 # Column cleaning ( renaming )
5 # Column creation amount_flag
6 # Current flag creation
7 # remove negative/zero amounts
8 transactions_transformed = transactions_df \
9 .withColumnRenamed("TransactionID", "Transaction_ID") \
10 .withColumn("TransactionDate", to_date("TransactionDate", "yyyy-MM-dd")) \
11 .withColumn("Amount", col("Amount").cast("double")) \
12 .withColumn("Amount_flag", when(col("Amount") > 500, "HIGH").otherwise("NORMAL")) \
13 .withColumn("Currency", when(col("Currency") == "CAD", "Canadian Dollar") \
14 .when(col("Currency") == "MXN", "Mexican Peso") \
15 .otherwise(col("Currency")) \
16 .filter(col("Amount") > 0) # Removing negative or zero values
17
```

Notebook 2 | Saved

Home Edit Run View Standard session PySpark (Python) Environment Workspace default Data Wrangler Copilot

All sources Lakehouses + Lakehouses Search tables Silver\_Lakehouse Tables Customers Geographical Orders Products Transactions Files

```
[12] 1 # Current flag creation
2 # remove negative/zero amounts
3 transactions_transformed = transactions_df \
4 .withColumnRenamed("TransactionID", "Transaction_ID") \
5 .withColumn("TransactionDate", to_date("TransactionDate", "yyyy-MM-dd")) \
6 .withColumn("Amount", col("Amount").cast("double")) \
7 .withColumn("Amount_flag", when(col("Amount") > 500, "HIGH").otherwise("NORMAL")) \
8 .withColumn("Currency", when(col("Currency") == "CAD", "Canadian Dollar") \
9 .when(col("Currency") == "MXN", "Mexican Peso") \
10 .otherwise(col("Currency")) \
11 .filter(col("Amount") > 0) # Removing negative or zero values
12 ✓ <1 sec - Command executed in 264 ms by Sahil on 2:01:14 PM, 3/24/25
```

Table Data Wrangler 7 columns, 4 rows Download Search

AMC Transaction_ID	AMC Order_ID	AMC PaymentMethod	AMC TransactionDate	AMC Amount	AMC Currency	AMC Amount_Flag
1 5001	1001	Credit Card	NULL	250.5	USD	NORMAL
2 5002	1002	PayPal	NULL	120.0	USD	NORMAL
3 5003	1003	Bank Transfer	NULL	540.3	Canadian Dollar	HIGH
4 5004	1004	Cash	NULL	75.2	Mexican Peso	NORMAL

```
[13] 1 display(transactions_transformed)
✓ 4 sec - Command executed in 4 sec 599 ms by Sahil on 2:02:41 PM, 3/24/25
> Spark jobs (2 of 2 succeeded) Resources Log
```

Notebook 2 | Saved

Home Edit Run View Standard session PySpark (Python) Environment Workspace default Data Wrangler Copilot

Workspaces All sources Lakehouses Silver\_Lakehouse Tables Customers Geographical Orders Products Transactions Files

```

1 # Transformations in Geographical Data
2 # Null countries to be removed
3 # region and city
4 from pyspark import functions as F
5 geographical_transformed = geo_df \
6     .withColumn("Region", F.upper(col("Region"))) \
7     .withColumn("City", F.initcap("City")) \
8     .filter(~col("Country").isNotNull()) # Removing records with null country values
9
10 # Display
11 display(geography_transformed)
12
[16] ✓ 2 sec - Command executed in 2 sec 389 ms by Sahil on 20:12 PM, 3/24/25
> Spark jobs (2 of 2 succeeded) Resources Log

```

Table view

Region	Country	State	City	PostalCode
NORTH	USA	New York	New York City	10001
SOUTH	USA	Texas	Houston	77001
EAST	Canada	Ontario	Toronto	M5H
WEST	Mexico	Jalisco	Guadalajara	44100

1 # Data Type Transformations

Notebook 2 | Saving...

Home Edit Run View Standard session PySpark (Python) Environment Workspace default Data Wrangler Copilot

Workspaces All sources Lakehouses Silver\_Lakehouse Tables Customers Geographical Orders Products Transactions Files

```

1 # Data Type Transformations
2 from pyspark.sql.types import DecimalType, DateType
3 orders_transformed = orders \
4     .withColumn("OrderDate", F.to_date("OrderDate", "yyyy-MM-dd")) \
5     .withColumn("ShippedDate", F.to_date("ShippedDate", "yyyy-MM-dd")) \
6     .withColumn("TotalAmount", F.col("TotalAmount").cast(DecimalType(10,2)))
7
8 display(orders_transformed)
[17] ✓ 1 sec - Command executed in 1 sec 485 ms by Sahil on 2:00:06 PM, 3/24/25
> Spark jobs (1 of 1 succeeded) Resources Log

```

Table view

OrderID	CustomerID	OrderDate	ShippedDate	Status	TotalAmount
1001	201	2024-01-15	2024-01-17	Shipped	250.5
1002	202	2024-01-20	2024-01-25	Delivered	120.0
1003	203	2024-02-01	2024-02-05	Pending	540.3
1004	204	2024-02-10	2024-02-12	Shipped	75.2

1 # Data Type Transformations

The screenshot shows a PySpark session in Azure Data Studio. The sidebar on the left lists workspaces, notebooks, and other resources. The main area displays a table named 'Lakehouses' with columns: 3 EAST, Canada, Ontario, Toronto, MSH, and 4 WEST, Mexico, Jalisco, Guadalajara, 44100. Below the table, three command history entries are shown:

- [18] 1 transactions\_transformed.write.mode("overwrite").saveAsTable("transactions\_transformed") ✓ 6 sec - Command executed in 6 sec 99 ms by Sahil on 2:14:01 PM, 3/24/25
- [19] 1 orders\_transformed.write.mode("overwrite").saveAsTable("orders\_transformed") ✓ 3 sec - Command executed in 3 sec 355 ms by Sahil on 2:14:32 PM, 3/24/25
- [20] 1 geography\_transformed.write.mode("overwrite").saveAsTable("geography\_transformed") ✓ 2 sec - Command executed in 2 sec 296 ms by Sahil on 2:14:46 PM, 3/24/25

Each entry includes a link to 'Spark jobs' and 'Resources'.

## **16. In Golder layer we load data in warehouses because warehouses are more efficient for reading data as they are OLAP Systems**

The screenshot shows the Microsoft Fabric interface for the 'E-commerce\_Golden\_Layer' workspace. The sidebar includes options for Home, On-preise, Monitor, Real-Time, Workloads, and Notebooks. The main workspace area is currently empty, displaying a placeholder icon and the message "There's nothing here yet". A button at the top says "Select a predesigned task flow".

The screenshot shows the Microsoft Fabric E-commerce\_Golden\_Layer workspace. The left sidebar includes Home, Create, Monitor, Real-Time, Workloads, and E-commerce sections. The main area displays a list of pre-designed task flows under the heading "Choose from predesigned task flows or add a task to build one (preview)". A search bar at the top right says "Search". The list includes:

Name	Type	Task	Owner	Refreshed	Next refresh	Endorsement	Sensitivity	Included in app
BL_warehouse	Warehouse	—	Sahil	—	—	—	—	No
BL_warehouse	Semantic model (d...)	—	E-commerce_Gold...	3/24/2025, 2:19:32 PM	N/A	—	—	No
E-commerce_Business_analysis	Report	—	E-commerce_Gold...	3/24/2025, 2:46:33 PM	—	—	—	No
E-Commerce_Semantic_model	Semantic model	—	E-commerce_Gold...	3/24/2025, 2:46:33 PM	N/A	—	—	No
silver_tables_to_golden_warehouse	Dataflow Gen2	—	Sahil	3/24/2025, 2:55:00 PM	3/29/2025, 6:30:00 AM	—	—	No

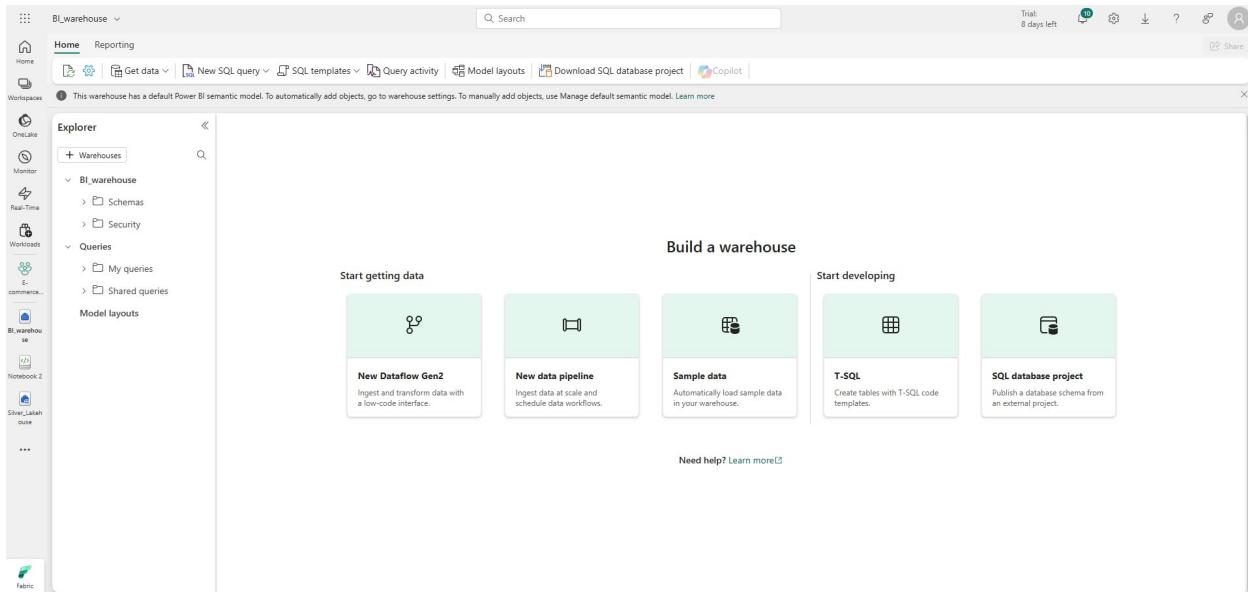
  

The screenshot shows the Microsoft Fabric E-commerce\_Golden\_Layer workspace. The left sidebar includes Home, Create, Monitor, Real-Time, Workloads, and E-commerce sections. The main area displays a "New item" dialog for a warehouse. The search bar at the top right says "Search". The dialog has tabs for "Favorites" and "All Items". It shows two options:

- Sample warehouse**: Start a new warehouse with sample data already loaded.
- Warehouse**: Provide strategic insights from multiple sources into your entire business.

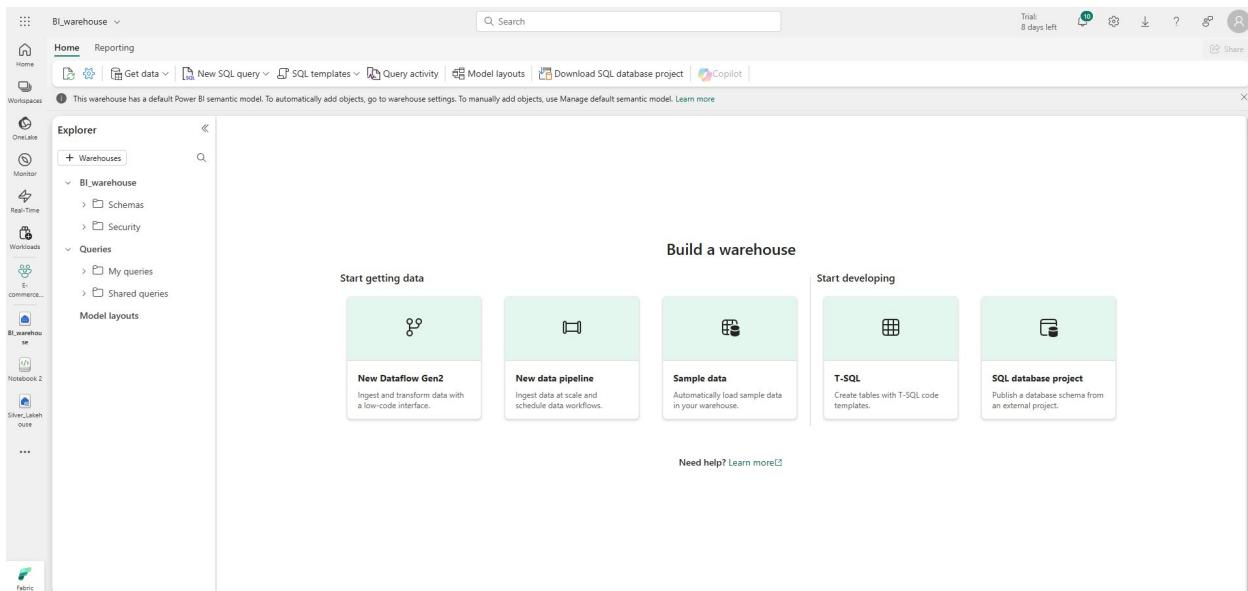
A search bar at the bottom right contains the text "ware".

## 17. Now you want to load your cleaned and transformed data to this warehouse from silver layer

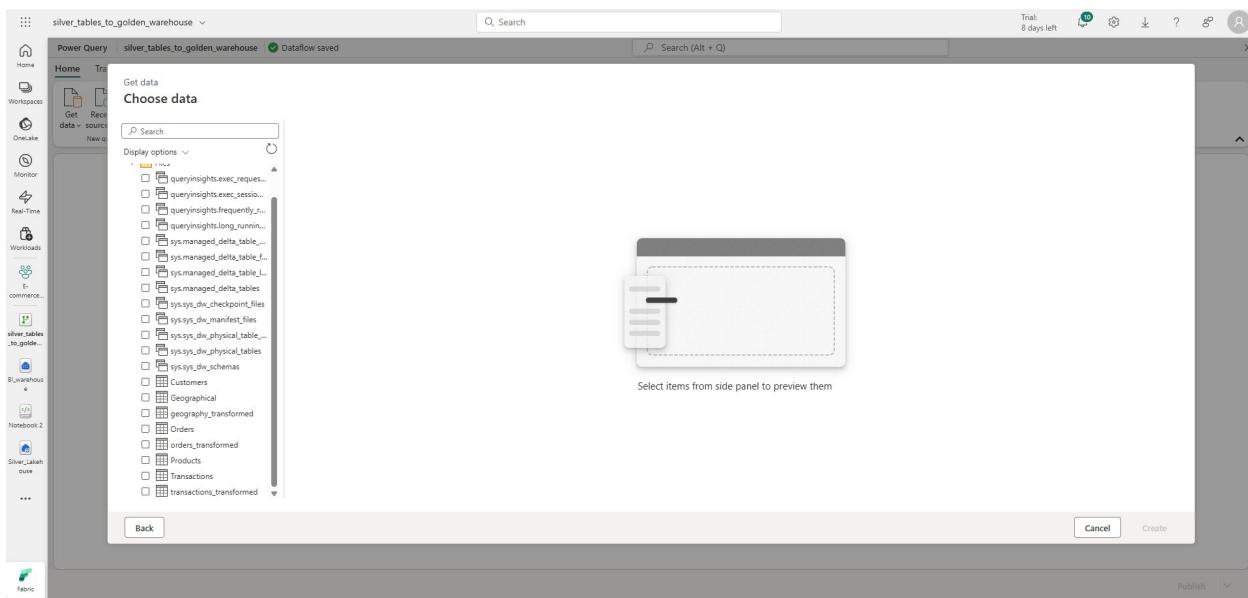
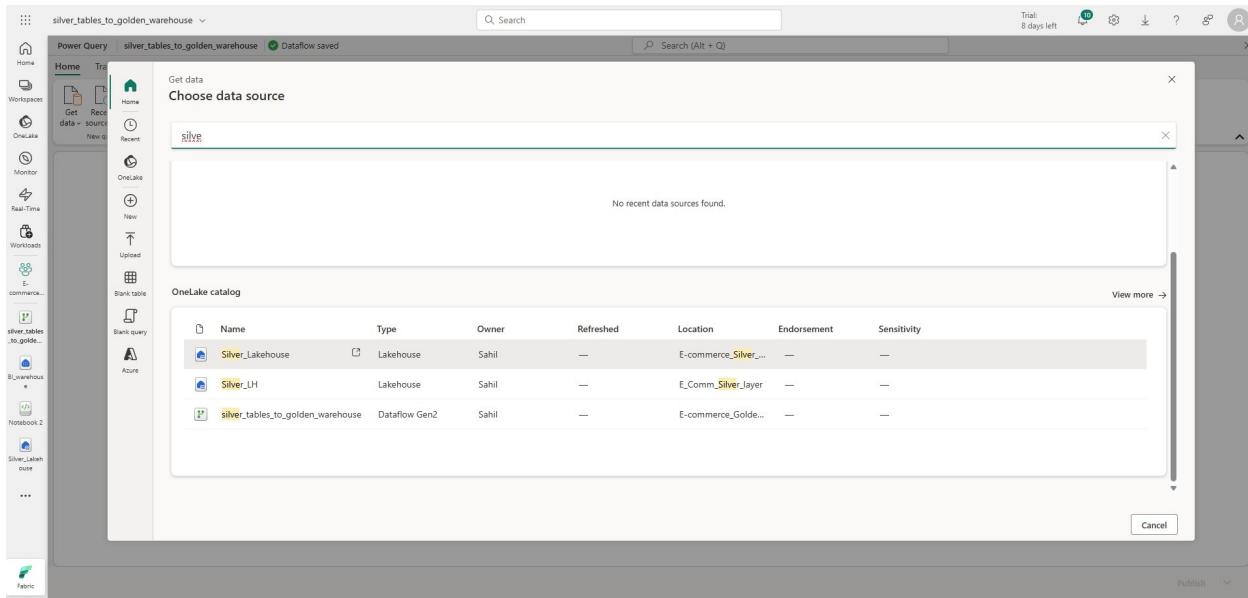


## How will I bring my data?

By dataflow gen2 why? Because of the golden layer and in this layer, we try to prefer things which are less technical so that noncoding users can also use this



## 18. Load data from silver Lake house



**Power Query** silver\_tables\_to\_golden\_warehouse Dataflow saved

**Queries [5]**

Table.TransformColumnTypes("#Navigation 2", {"CustomerID", Int64.Type})

**CustomerID** **FullName** **Email** **PhoneNumber** **Region** **Country** **JoinDate**

	Valid 100%	
1	Valid 100%	Error 0%
2	Valid 100%	Empty 0%
3	Valid 100%	
4	Valid 100%	

Completed (2.21 s) Columns: 7 Rows: 4 Column profiling based on entire data set Add default destination...

## 19. This is a silver layer and want to load into the golden layer's warehouse

**Power Query** silver\_tables\_to\_golden\_warehouse Dataflow saved

**Default destination** wa

**Choose data destination**

Getting started with a default destination

A default destination is the pre-set location where data is loaded when using Dataflow Gen2.

Only queries with table result can be written to the default destination.

Each query using the default destination gets the same destination settings (container, update method, schema options).

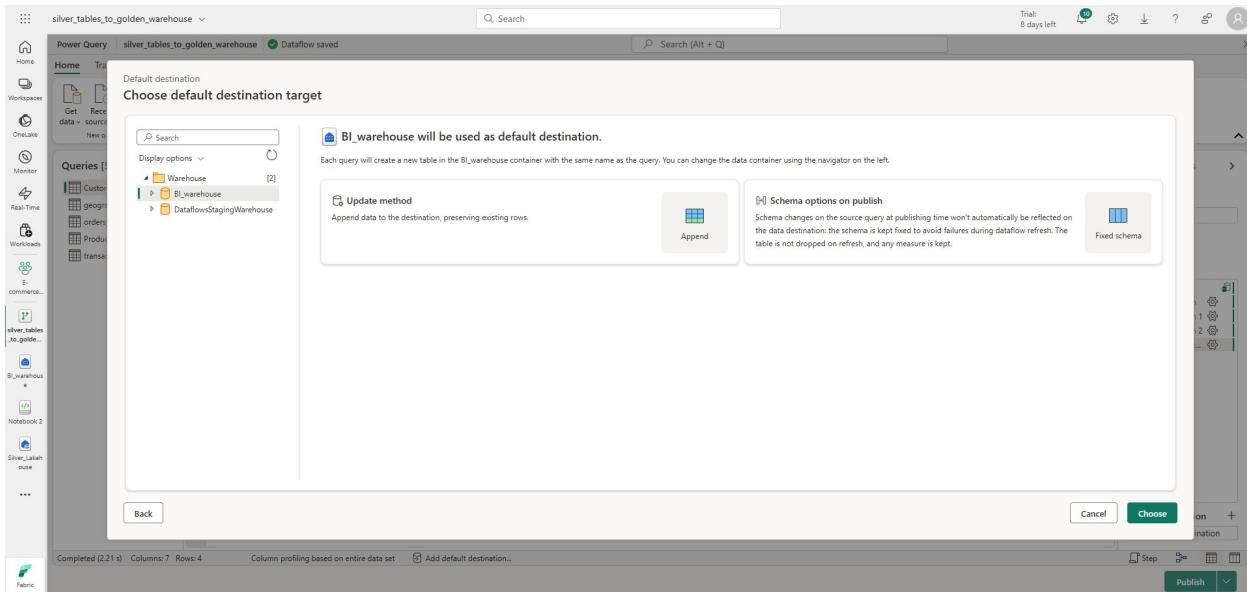
New sources

Warehouse Microsoft Fabric

OneLake catalog

Name	Type	Owner	Refreshed	Location	Endorsement	Sensitivity
DataflowsStagingWarehouse	Warehouse	Sahil	—	E-commerce_Golden...	—	—
DataflowsStagingWarehouse	Warehouse	Sahil	—	Dev_WS	—	—
test_warehouse	Warehouse	Sahil	—	Test	—	—

Completed (2.21 s) Columns: 7 Rows: 4 Column profiling based on entire data set Add default destination... Step Publish



## 20. Click on publish

The screenshot shows the Power Query Editor interface. The 'Applied steps' pane on the right lists the 'Source' and 'Navigation' steps. The 'Data destination' pane at the bottom shows the 'BI\_warehouse' destination. The main area displays a table with the following data:

	CustomerID	FullName	Email	PhoneNumber	Region	Country	JoinDate
1	201	Alice Johnson	alice@gmail.com	123-456-7890	North	USA	2023-03-10
2	202	Bob Smith	bob@gmail.com	234-567-8901	South	USA	2022-07-25
3	203	Charlie Brown	charlie@gmail.com	345-678-9012	East	Canada	2021-11-30
4	204	Diana Prince	diana@gmail.com	456-789-0123	West	Mexico	2020-09-15

## 21. Your data finally moves to the warehouse where all BI tools connect

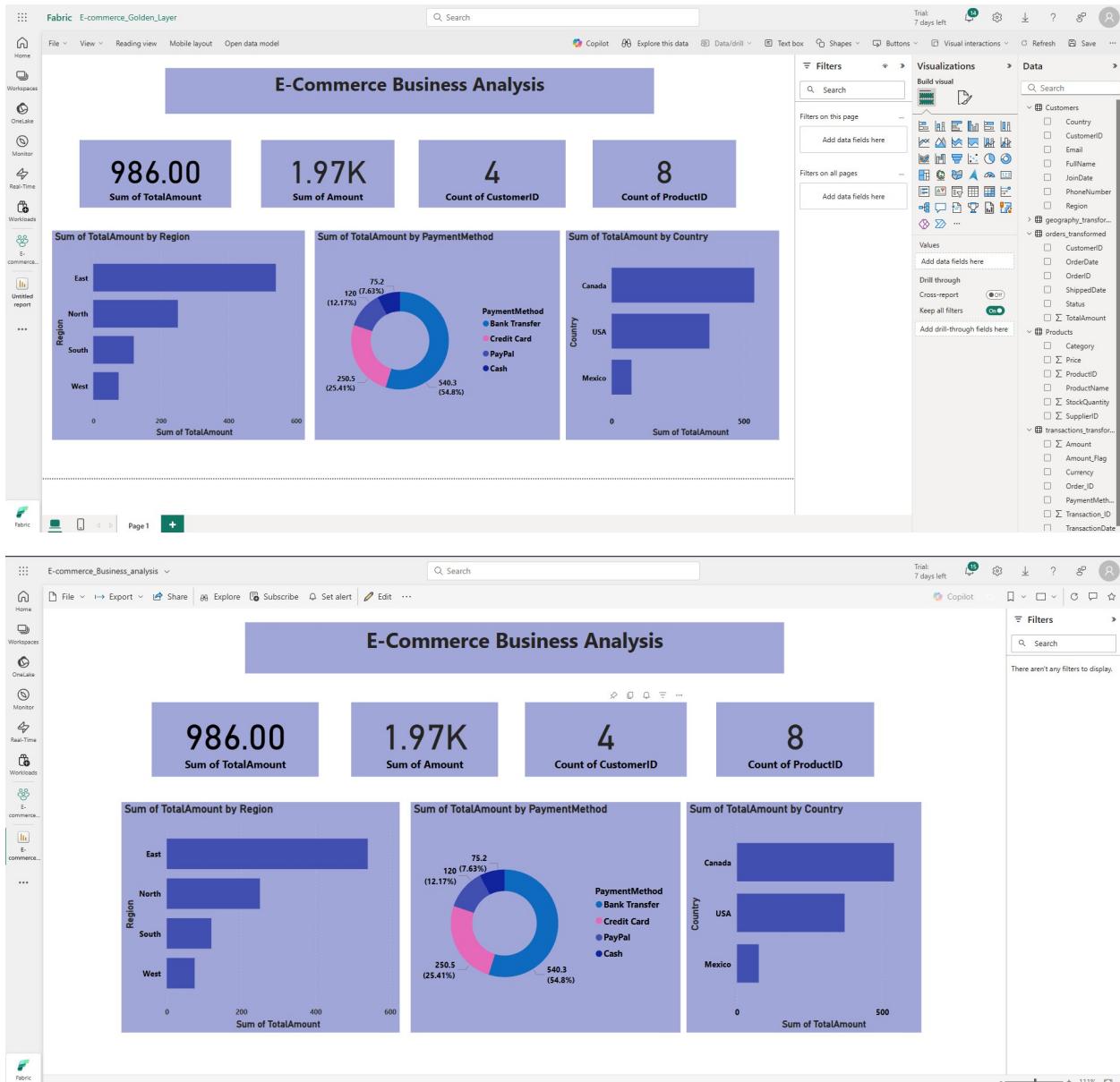
The screenshot shows the Power BI service interface. On the left, the 'Explorer' pane lists the 'BI\_warehouse' database, its schemas (Schemas, dbo), and tables (Customers, geography\_transformed, orders\_transformed, products, transactions). The 'Customers' table is selected. In the center, a 'Data preview - Customers' section displays a table with 4 rows and 7 columns: CustomerID, FullName, Email, PhoneNumber, Region, Country, and JoinDate. The data is as follows:

	CustomerID	FullName	Email	PhoneNumber	Region	Country	JoinDate
1	201	Alice Johnson	alice@gmail.com	123-456-7890	North	USA	2023-03-10
2	202	Bob Smith	bob@gmail.com	234-567-8901	South	USA	2022-07-25
3	203	Charlie Brown	charlie@gmail.com	345-678-9012	East	Canada	2021-11-30
4	204	Diana Prince	diana@gmail.com	456-789-0123	West	Mexico	2020-09-15

At the bottom, a message indicates the query succeeded (4 sec 111 ms) and shows 7 rows.

## 22. Finally, to end, users create report

The screenshot shows the Power BI service interface. The 'New semantic model' dialog is open, prompting for a 'Direct Lake semantic model name' (E-Commerce\_Semantic\_model). It explains that a semantic model will be created in the chosen workspace with the selected tables in Direct Lake storage mode. Below, it lists workspaces in a Fabric capacity (E-commerce\_Golden\_Layer) and allows selecting tables for the semantic model. The 'orders\_transformed' table is selected. The background shows the same Power BI service interface as the previous screenshot, with the 'Customers' table previewed.



## 23. To give access

The screenshot shows the Microsoft Fabric E-commerce\_Golden\_Layer workspace. On the left, there's a sidebar with icons for Home, Workspaces, OneDrive, Monitor, Real-Time, and Workbooks. The main area displays a list of items under the 'E-commerce...' category. A central placeholder says 'Choose from predesigned task flows or add a task to build one (preview)'. Below it, a table lists data assets:

Name	Type	Task	Owner	Refreshed	Next refresh	Endorsement	Sensitivity
BI_warehouse	Warehouse	—	Sahil	—	—	—	—
BI_warehouse	Semantic model (d...)	—	E-commerce_Gold...	3/24/2025, 2:19:32 PM	N/A	—	—
E-commerce_Business_analysis	Report	—	E-commerce_Gold...	3/24/2025, 2:46:33 PM	—	—	—
E-Commerce_Semantic_model	Semantic model	—	E-commerce_Gold...	3/24/2025, 2:46:33 PM	N/A	—	—
silver_tables_to_golden_warehouse	Dataflow Gen2	—	Sahil	3/24/2025, 2:55:00 ...	▲ N/A	—	—

A modal window titled 'Manage access' is open on the right, showing 'Sahil' as the owner.

## 24. To schedule the refresh for all the data pipeline, dataflows and notebooks

The screenshot shows the Microsoft Fabric E-commerce\_Golden\_Layer workspace with the 'Dataflows' tab selected. The left sidebar includes icons for Home, Dashboards, Semantic models, Workbooks, Reports, Dataflows, and App. The main area shows the settings for a dataflow named 'silver\_tables\_to\_golden\_warehouse'. It includes sections for 'Gateway Connection', 'Refresh', 'Time zone', 'Configure a refresh schedule', and 'Send refresh failure notifications to'. The 'Time zone' section is expanded, showing '(UTC) Coordinated Universal Time' selected. The 'Configure a refresh schedule' section has 'On' selected for refresh frequency and 'Daily' selected for time.