Microsoft Fabric in a Day Lab Manual – **Lab 3**

Author: Will Crayger
Email: wcrayger@lucidbi.co
LinkedIn: LinkedIn.com/in/willcrayger/
Course Material: GitHub.com/Lucid-Will/FabCon-EU-Zero-To-Hero-with-Fabric

Working with Pipelines – Extracting Source Data

Introduction

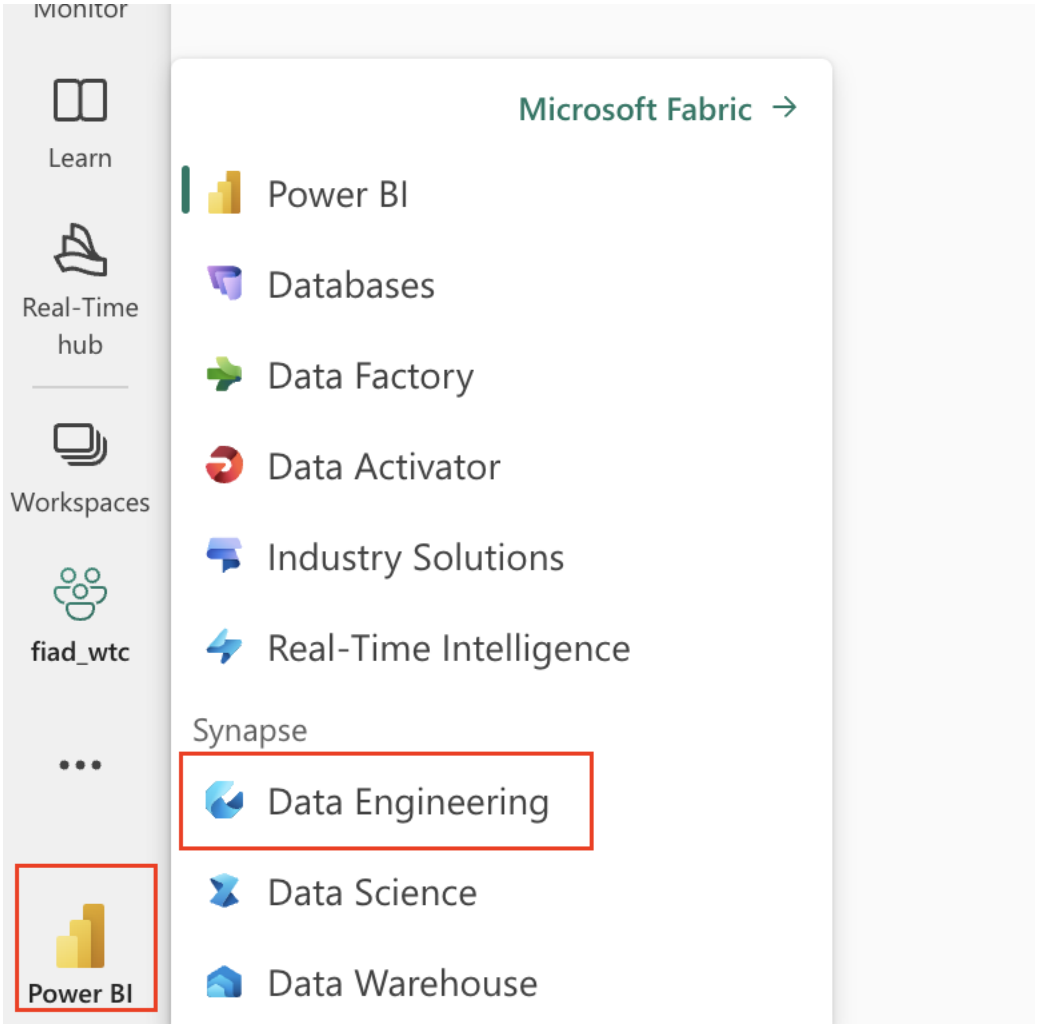
Copying data from source to target is a fundamental pattern in pipelines. To illustrate this, Lab 3 is divided into two parts, showcasing both basic and advanced applications of this approach:

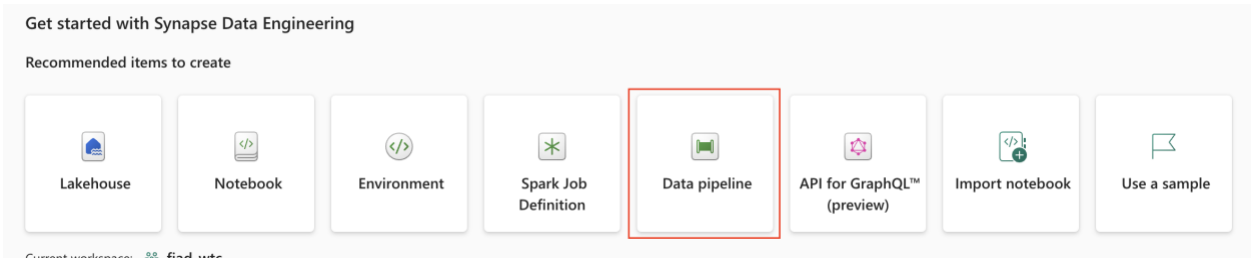
1. The first pipeline will extract data from a single table in an Azure SQL Database.
2. The second pipeline will demonstrate a more advanced, real-world "metadata-driven" method, using a lookup activity to iterate through a list of tables and extract their data.

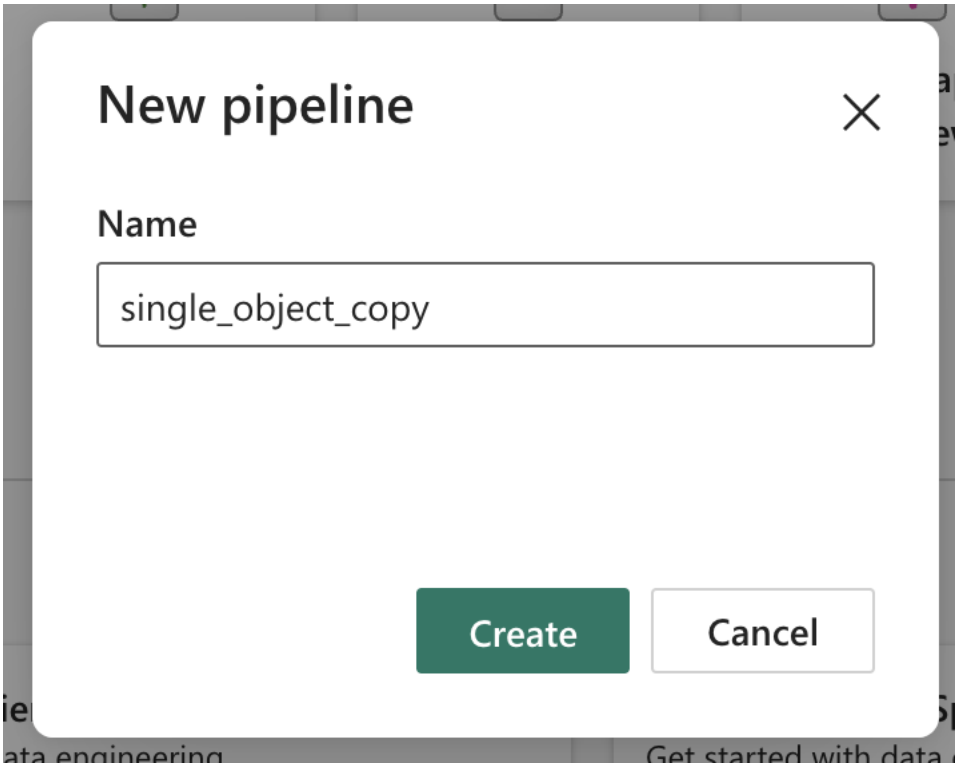
Understanding how lookups and parameters work together is key, as they enable the creation of powerful, scalable pipelines.

Single Object Copy

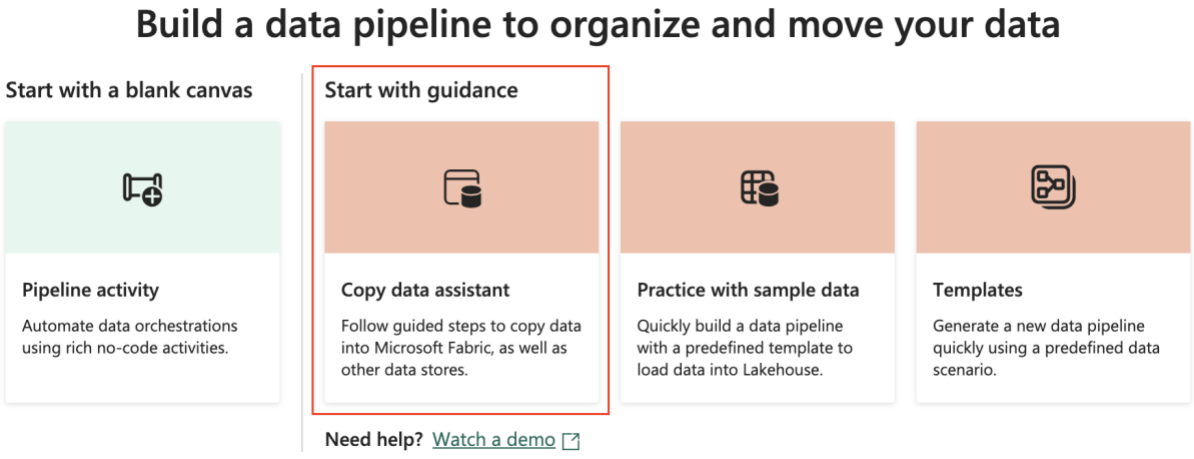
Create a Data Pipeline: To begin creating your data pipeline, navigate to the **Data Engineering Fabric Landing** page. From here, choose **Data pipeline** from the available options. Name your pipeline **single_object_copy** and click **Create**.


Monitor
Learn
Real-Time hub
Workspaces
fiad_wtc
Power BI
Microsoft Fabric
Power BI
Databases
Data Factory
Data Activator
Industry Solutions
Real-Time Intelligence
Synapse
Data Engineering
Data Science
Data Warehouse

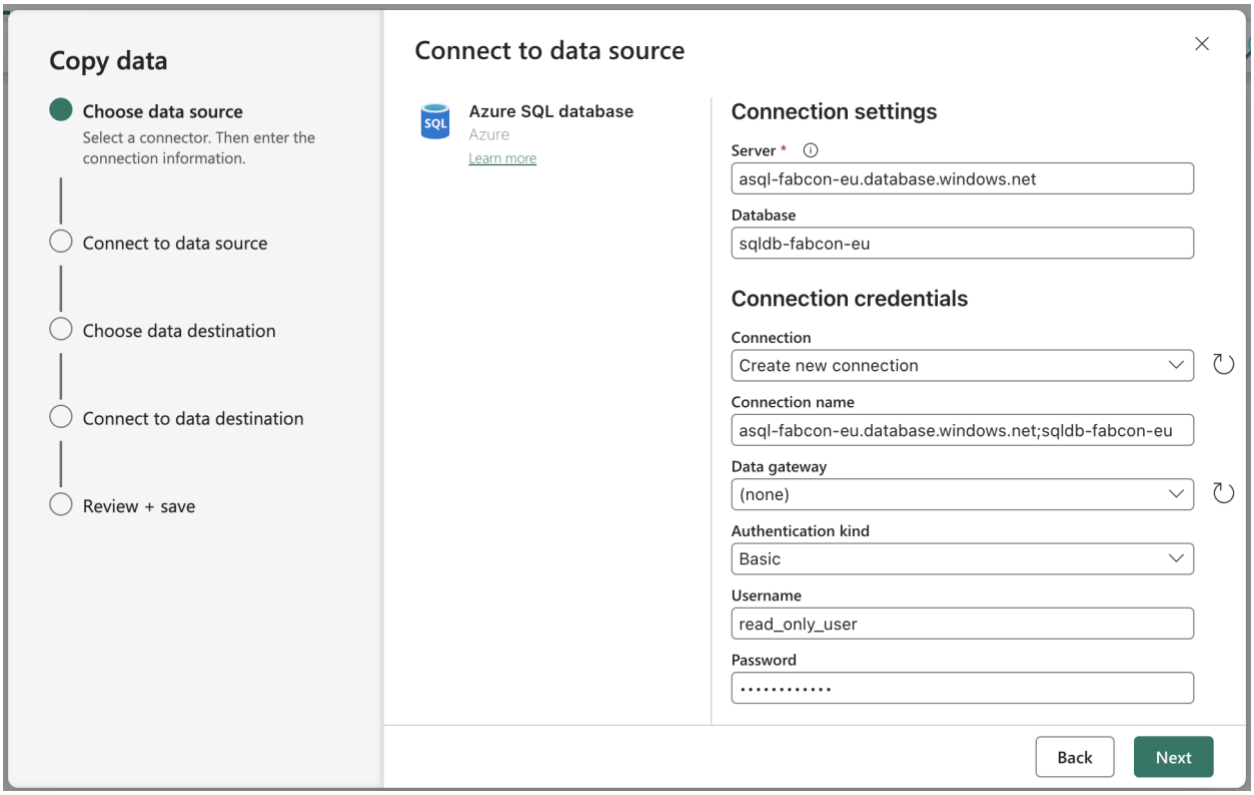

Get started with Synapse Data Engineering
Recommended items to create
Lakehouse
Notebook
Environment
Spark Job Definition
Data pipeline
API for GraphQL™ (preview)
Import notebook
Use a sample

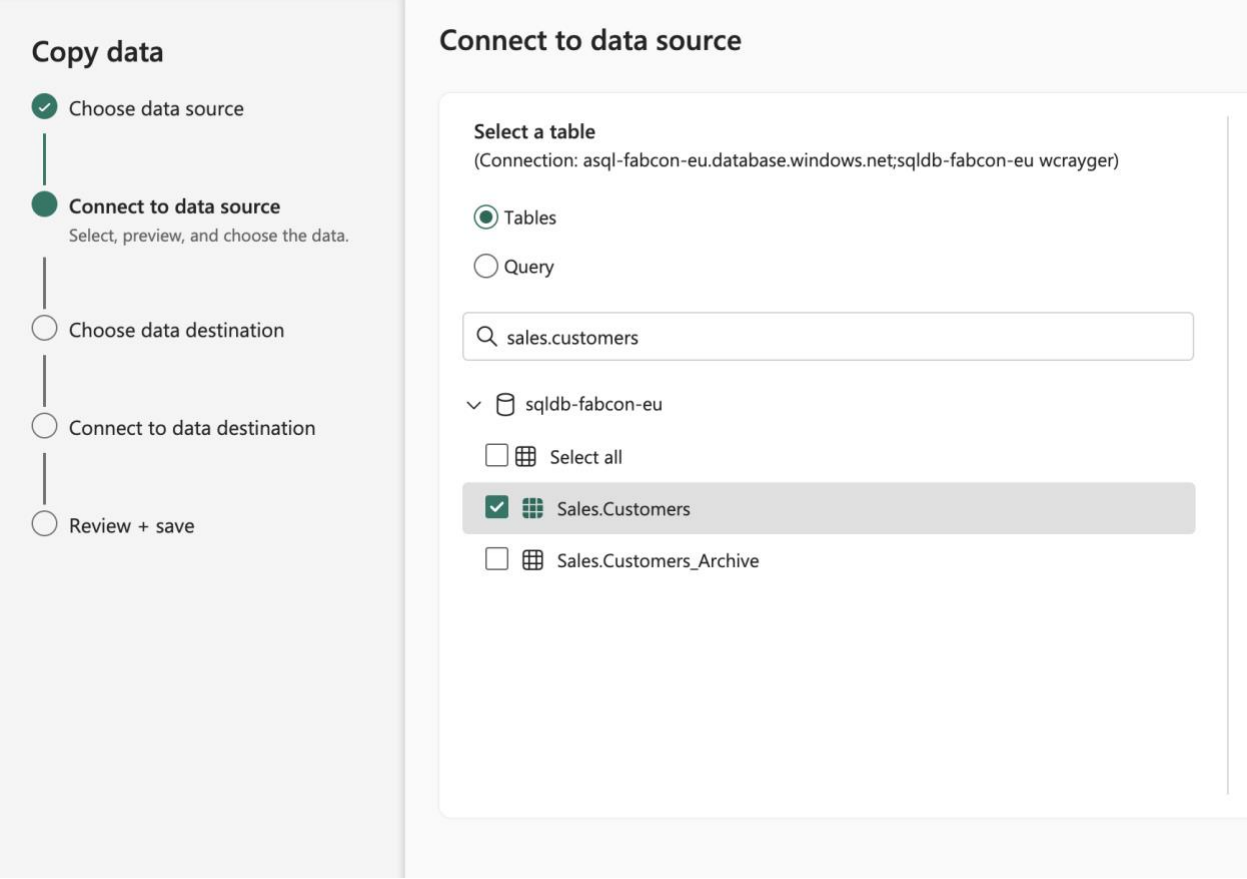


Setup the Copy Data Activity: Once your pipeline is created, you'll land on the pipeline page. Select **Copy data assistant** to start the process. You will notice a variety of available data sources listed for data extraction. Choose **Azure SQL Database** from the list and click **Next**.



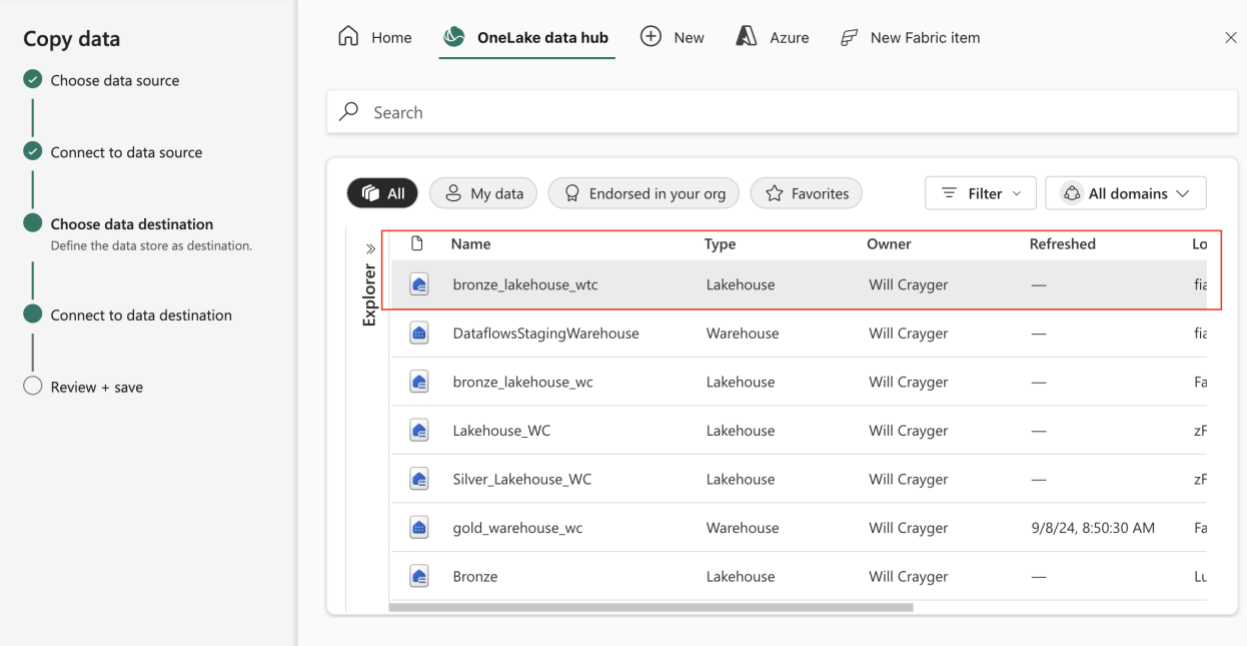
With the **Create new connection** option is selected, reauthenticate with the read-only credentials from lab 2 and click **Next**. In the table selection screen, check the box next to **Sales.Customers** and click **Next**.

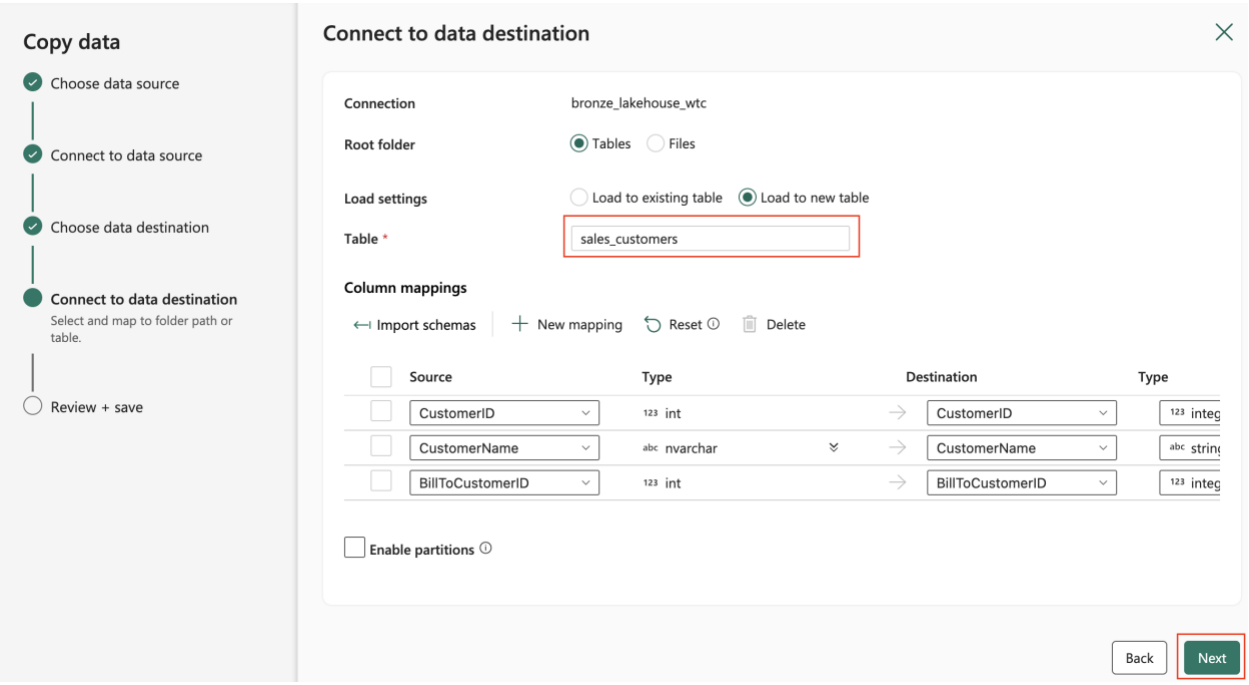

Copy data
Choose data source
Select a connector. Then enter the connection information.
Connect to data source
Choose data destination
Connect to data destination
Review + save
Connect to data source
Azure SQL database
Azure
Learn more
Connection settings
Server *
asql-fabcon-eu.database.windows.net
Database
sqldb-fabcon-eu
Connection credentials
Connection
Create new connection
Connection name
asql-fabcon-eu.database.windows.net;sqldb-fabcon-eu
Data gateway
(none)
Authentication kind
Basic
Username
read_only_user
Password
Back
Next

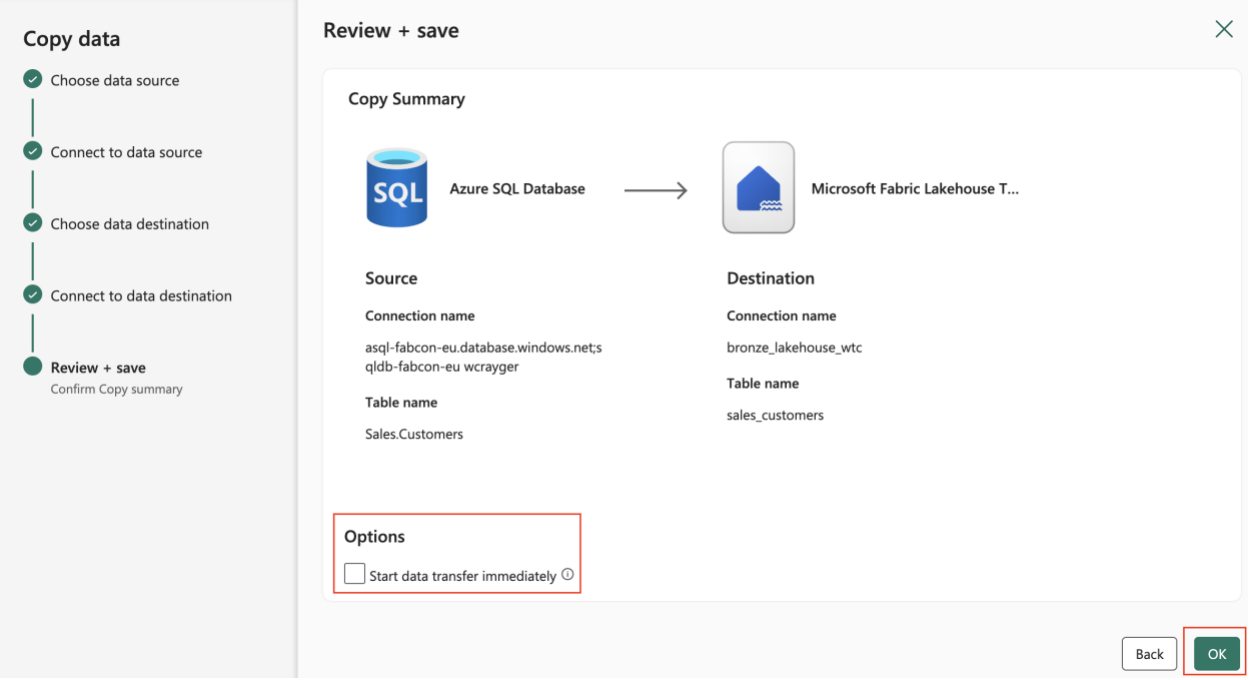


Selecting the Destination: In the **Destination** menu, switch to the **OneLake data hub** tab located at the top of the window. Choose the **bronze_lakehouse** created in Lab 1 as the **Lakehouse** destination.

On the next page, change the table name to **sales_customers.** Review the other settings to familiarize yourself but leave everything as default and then click **Next**. You'll be presented with the **Pipeline Summary**. Uncheck the **Start data transfer immediately** box. Click **OK** to save the configuration.


Copy data
Choose data source
Connect to data source
Choose data destination
Define the data store as destination.
Connect to data destination
Review + save
Home
OneLake data hub
New
Azure
New Fabric item
Search
All
My data
Endorsed in your org
Favorites
Filter
All domains
Explorer
Name
Type
Owner
Refreshed
bronze_lakehouse_wtc
Lakehouse
Will Crayger
DataflowsStagingWarehouse
Warehouse
bronze_lakehouse_wc
Lakehouse_WC
Silver_Lakehouse_WC
gold_warehouse_wc
9/8/24, 8:50:30 AM
Bronze

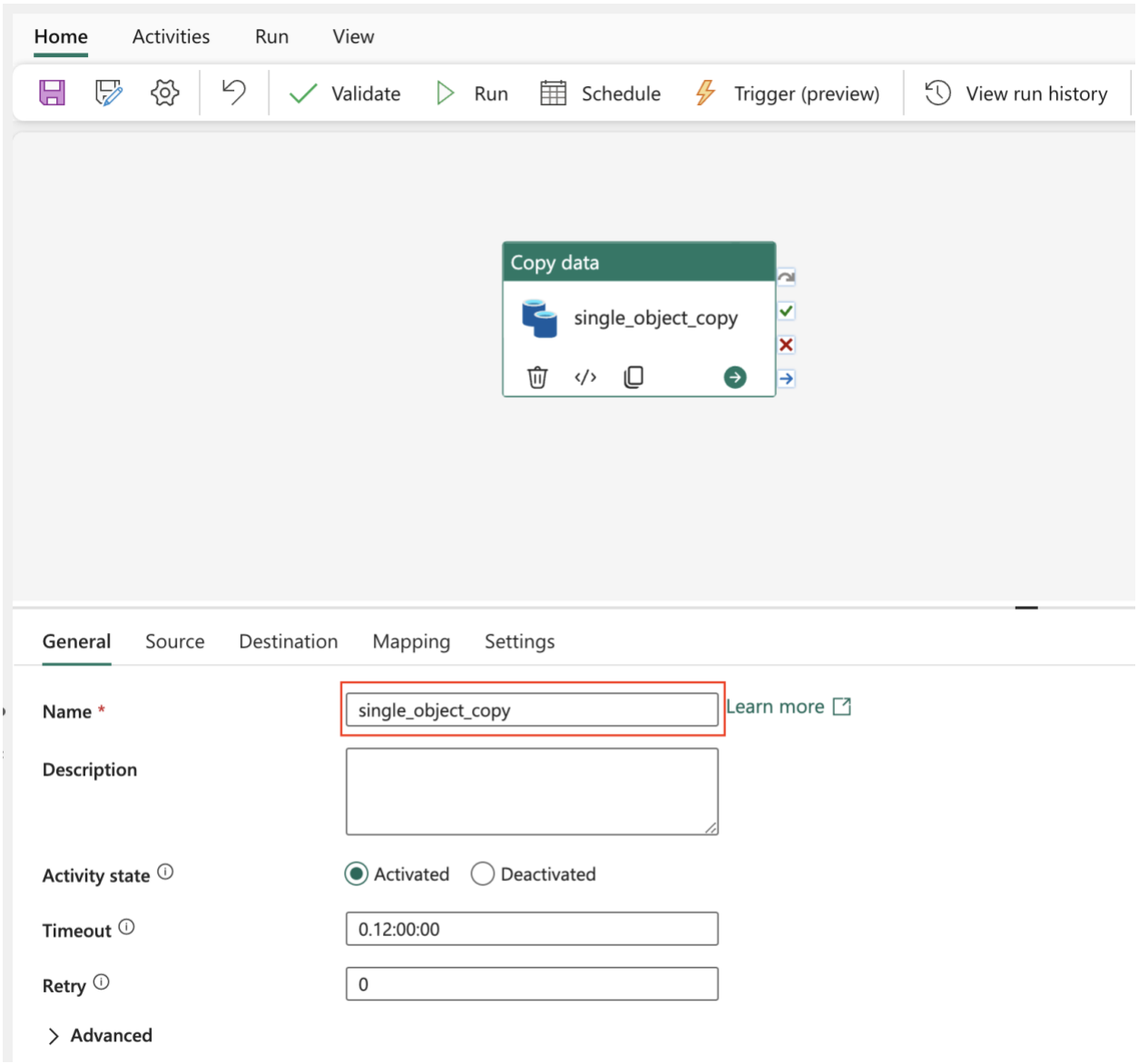

Copy data
Choose data source
Connect to data source
Choose data destination
Connect to data destination
Select and map to folder path or table.
Review + save
Connect to data destination
Connection
bronze_lakehouse_wtc
Root folder
Tables
Files
Load settings
Load to existing table
Load to new table
Table *
sales_customers
Column mappings
Import schemas
New mapping
Reset
Delete
Source
Type
Destination
CustomerID
int
CustomerName
nvarchar
BillToCustomerID
Enable partitions
Back
Next

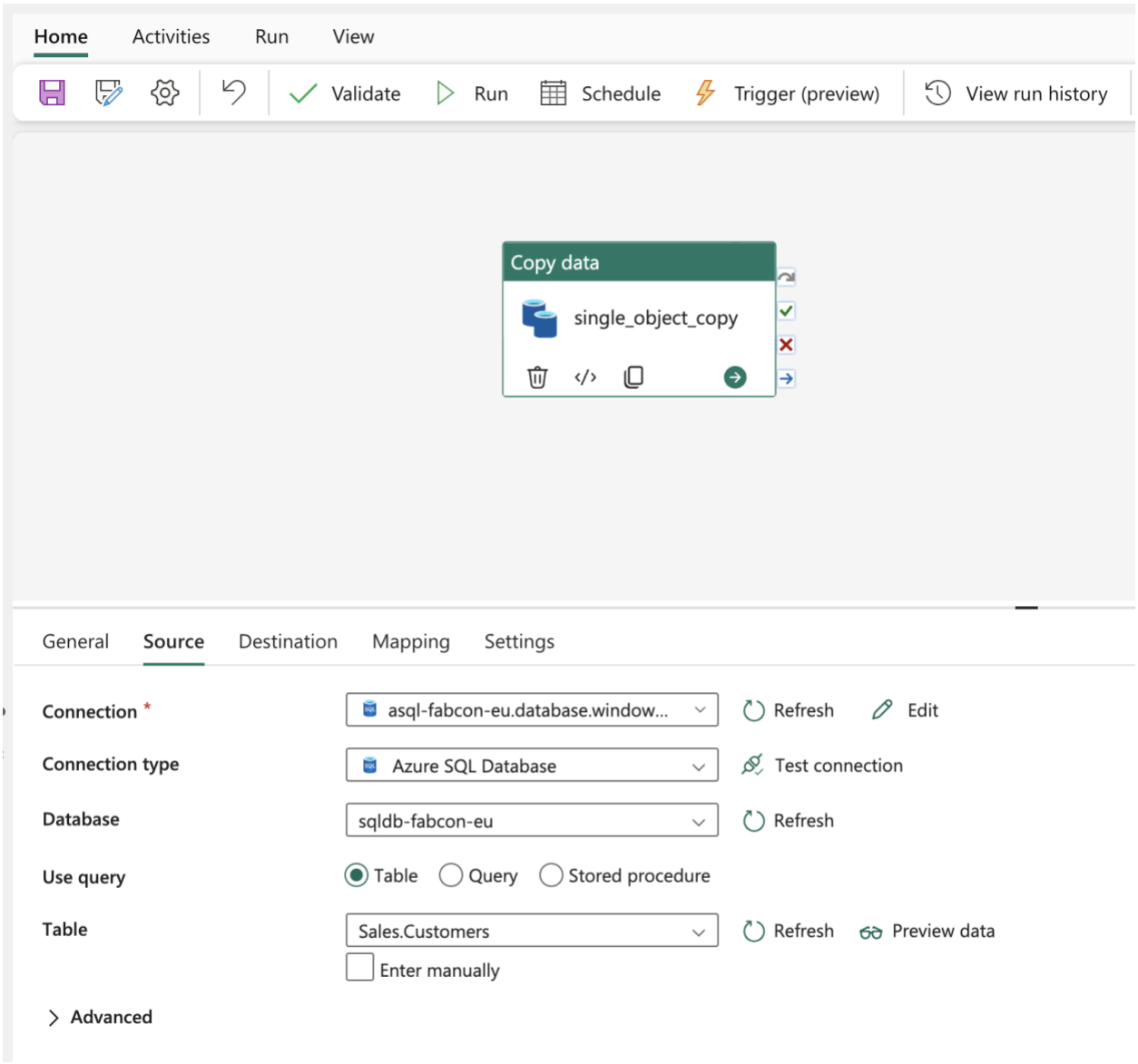


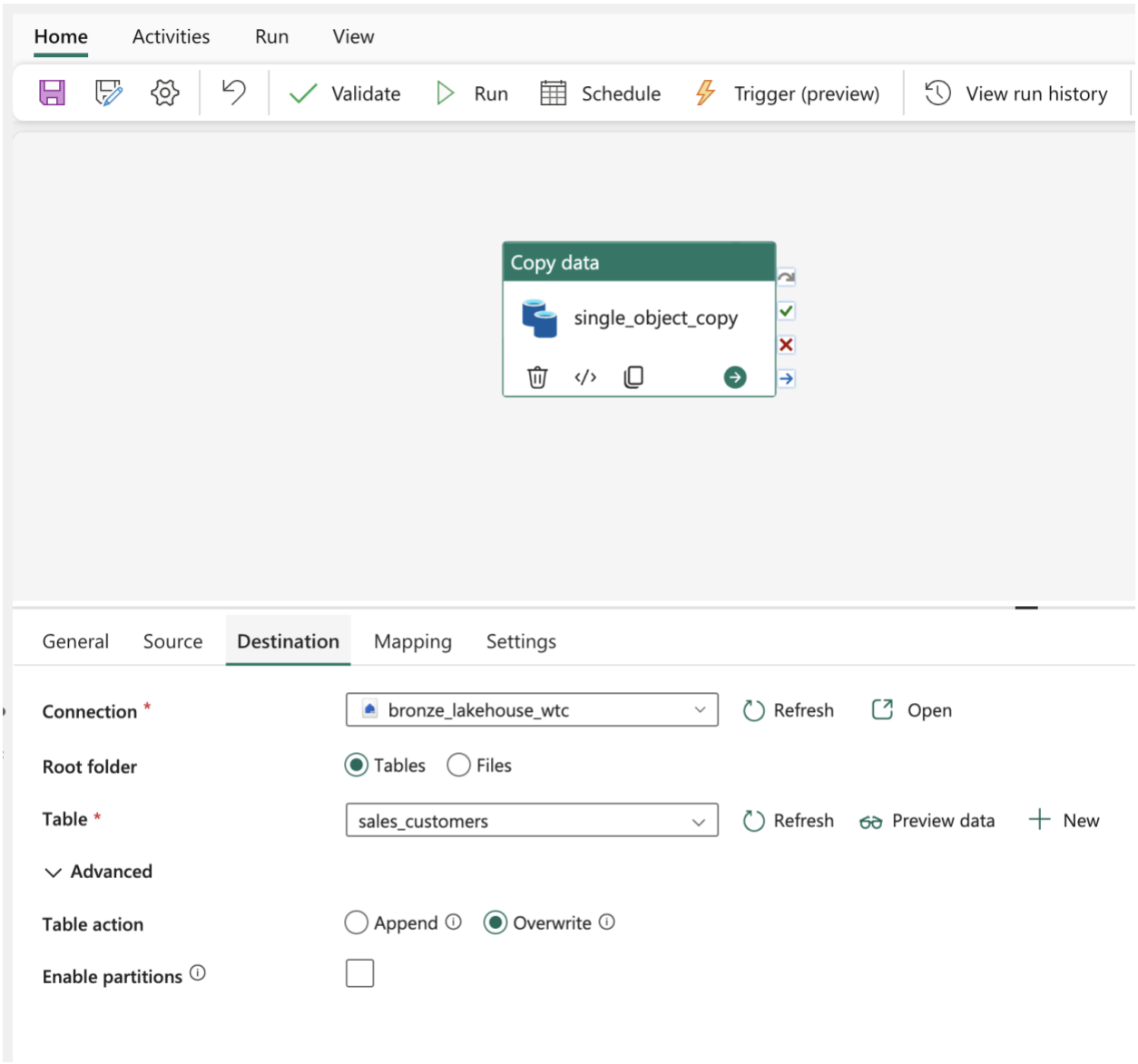
Configuring and Running the Pipeline: You should be returned to the canvas where the **Copy data** activity is visible. Select it and rename the activity to **single_object_copy** then select the **Source** tab to review the pre-populated configuration details that were set using the **Copy Data Tool**.

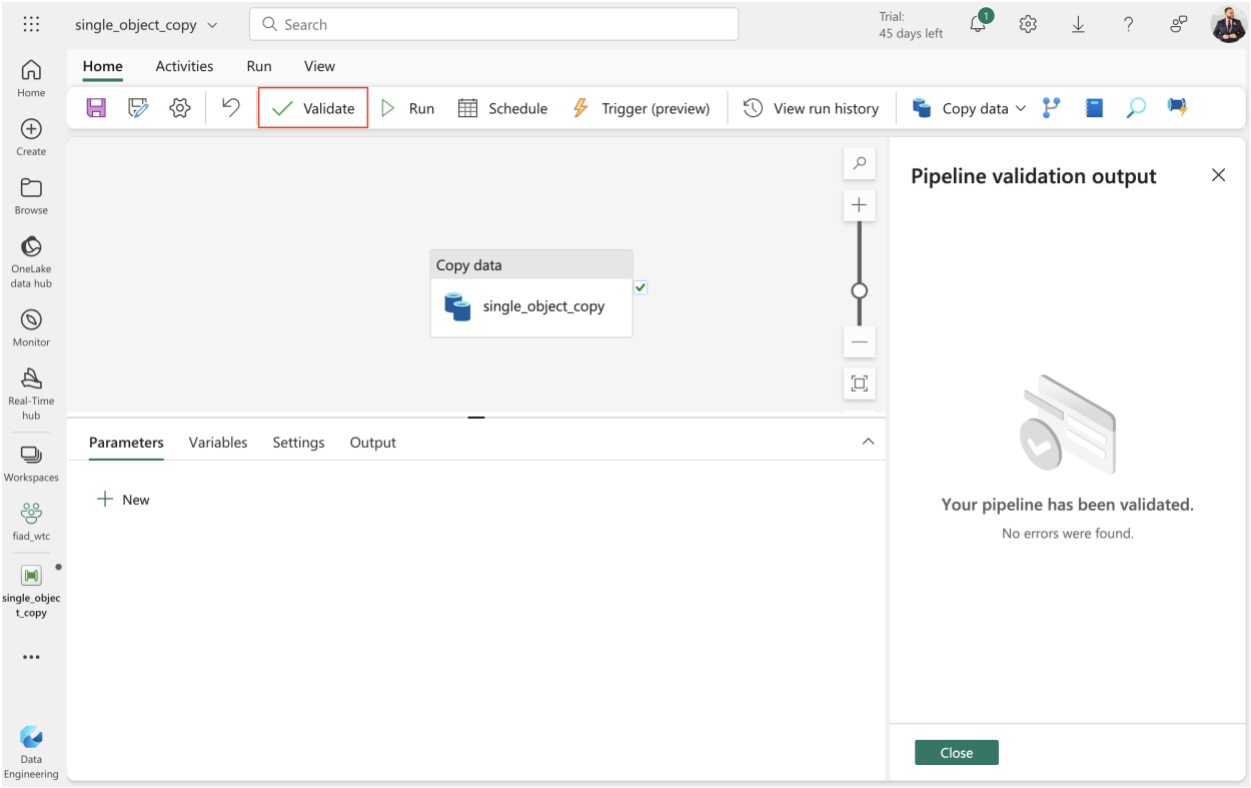
Next, select the **Destination** tab and review the settings there as well. For consistency, ensure the table name is lowercase, **sales_customers**. Expand the **Advanced** section and review the **Table action** options: **Append** and **Overwrite**. These are critically important depending on your project's requirements.

Once you've reviewed everything, return to the **Home** tab of the pipeline and click the **Validate** button. If all steps were done correctly, no errors should appear, and you can safely close the **Pipeline Validation Output**.


Home
Activities
Run
View
Validate
Run
Schedule
Trigger (preview)
View run history
Copy data
single_object_copy
General
Source
Destination
Mapping
Settings
Name *
single_object_copy
Learn more
Description
Activity state
Activated
Deactivated
Timeout
0.12:00:00
Retry
0
Advanced

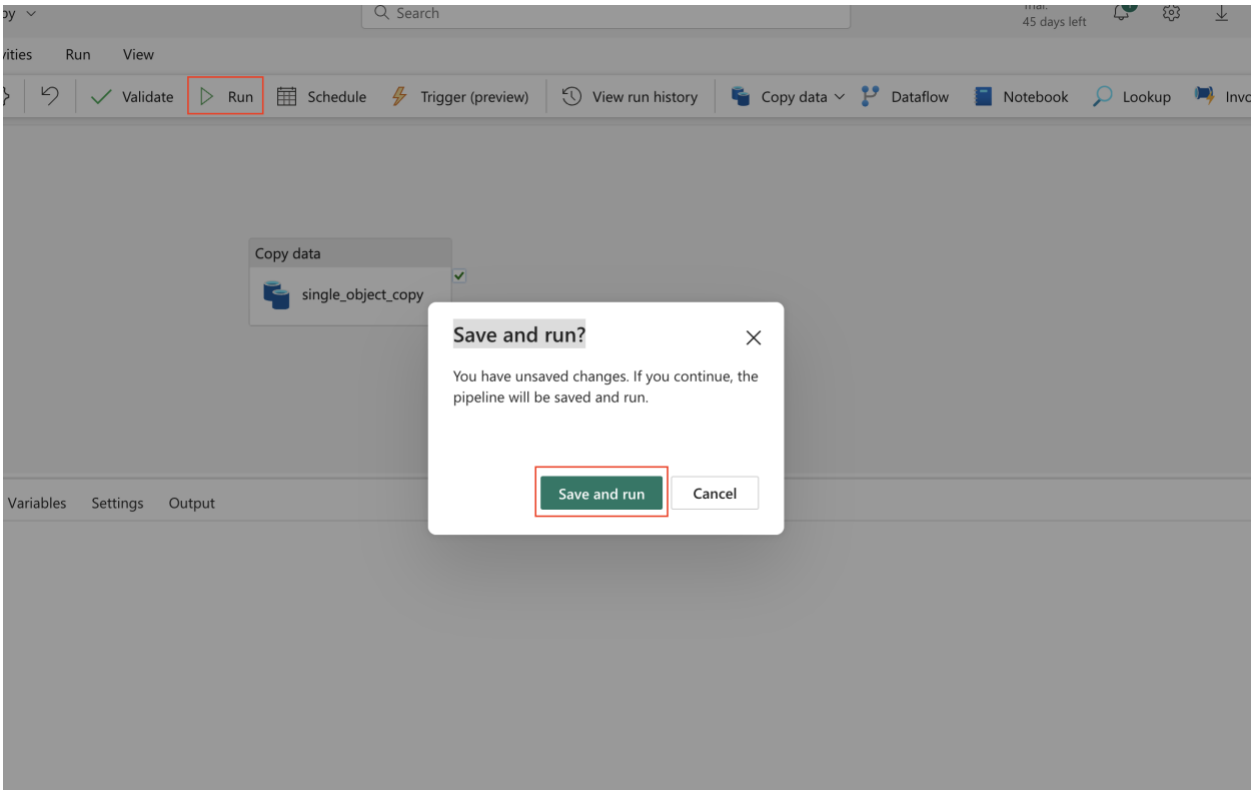

Home
Activities
Run
View
Validate
Run
Schedule
Trigger (preview)
View run history
Copy data
single_object_copy
General
Source
Destination
Mapping
Settings
Connection *
asql-fabcon-eu.database.window...
Refresh
Edit
Connection type
Azure SQL Database
Test connection
Database
sqldb-fabcon-eu
Refresh
Use query
Table
Query
Stored procedure
Table
Sales.Customers
Refresh
Preview data
Enter manually
Advanced

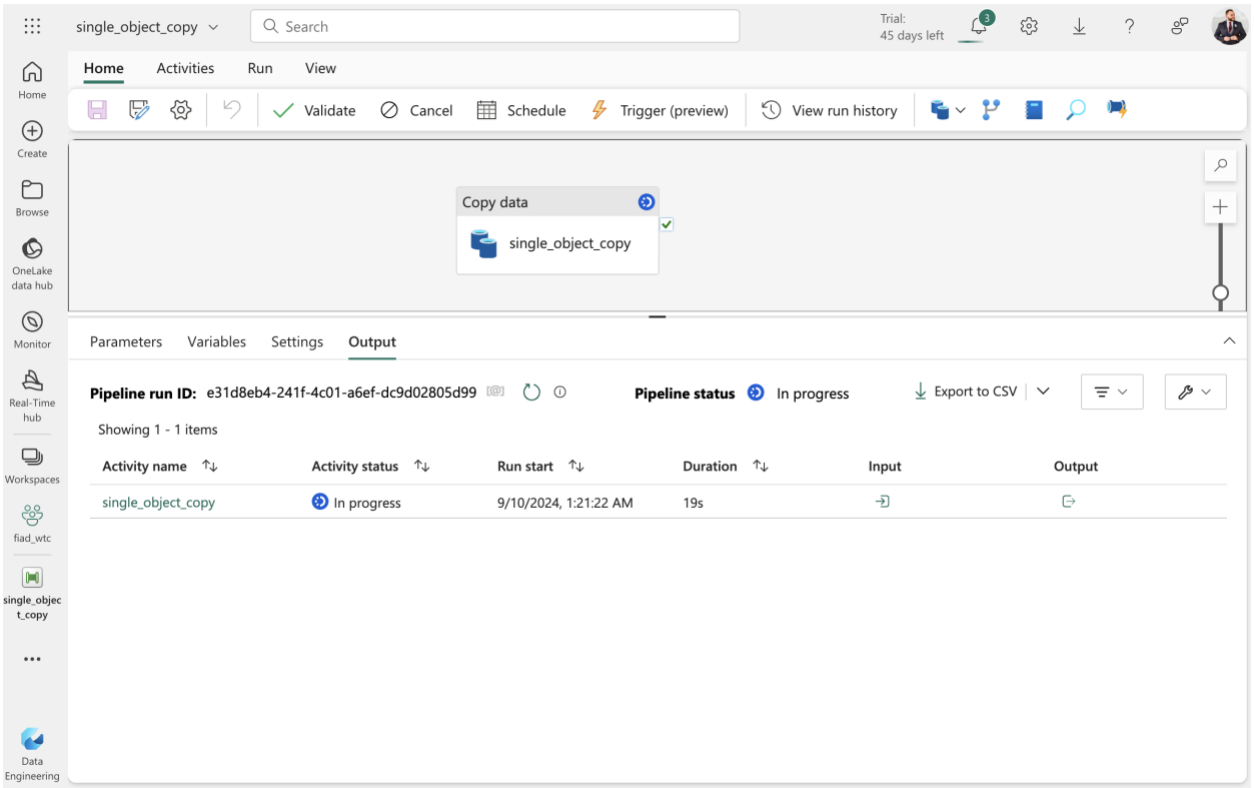

Home
Activities
Run
View
Validate
Run
Schedule
Trigger (preview)
View run history
Copy data
single_object_copy
General
Source
Destination
Mapping
Settings
Connection *
bronze_lakehouse_wtc
Refresh
Open
Root folder
Tables
Files
Table *
sales_customers
Refresh
Preview data
New
Advanced
Table action
Append
Overwrite
Enable partitions

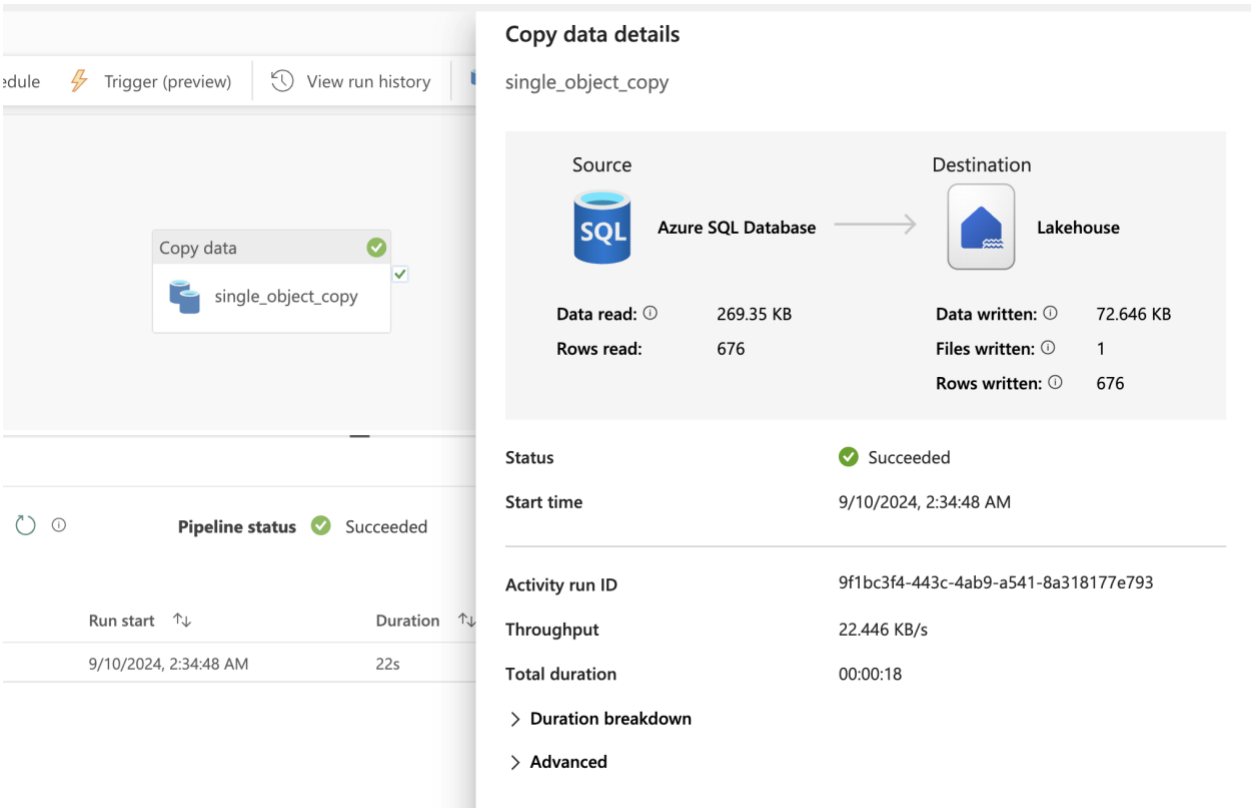


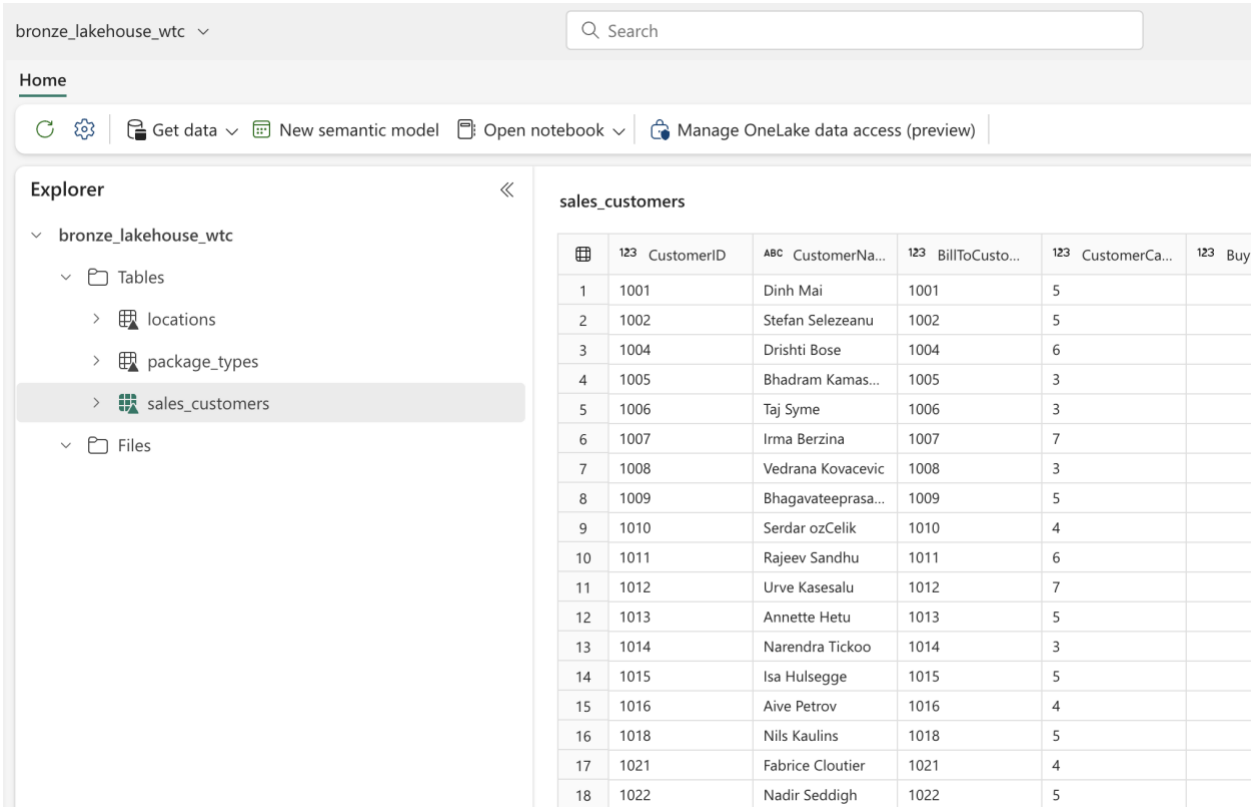
Executing and Monitoring the Pipeline: Now that your pipeline is validated, click **Run** from the activity bar. Select **Save and Run** to begin the process. The **Output** of the pipeline will automatically display, allowing you to track the execution in real-time. You can also click on the **Activity Name** to launch the **Copy data details** blade, where you can monitor progress and review various metrics related to the run.

Once the pipeline is completed, navigate back to your **Lakehouse**. There, you should see the newly created table for **Sales_Customers**. You have now successfully completed **Part 1** of the lab.


Search
Run
View
Validate
Run
Schedule
Trigger (preview)
View run history
Copy data
Dataflow
Notebook
Lookup
Copy data
single_object_copy
Save and run?
You have unsaved changes. If you continue, the pipeline will be saved and run.
Save and run
Cancel
Variables
Settings
Output


single_object_copy
Search
Trial:
45 days left
Home
Activities
Run
View
Validate
Cancel
Schedule
Trigger (preview)
View run history
Copy data
single_object_copy
Parameters
Variables
Settings
Output
Pipeline run ID: e31d8eb4-241f-4c01-a6ef-dc9d02805d99
Pipeline status
In progress
Export to CSV
Showing 1 - 1 items
Activity name
Activity status
Run start
Duration
Input
Output
single_object_copy
In progress
9/10/2024, 1:21:22 AM
19s
Home
Create
Browse
OneLake data hub
Monitor
Real-Time hub
Workspaces
fiad_wtc
single_object_copy
Data Engineering

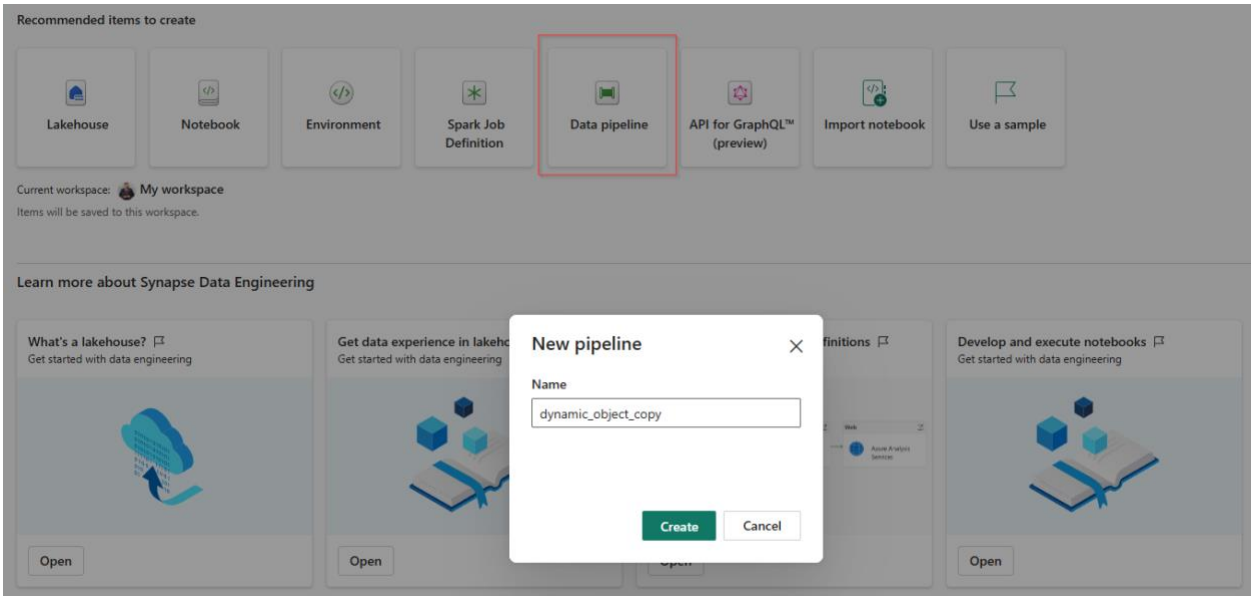


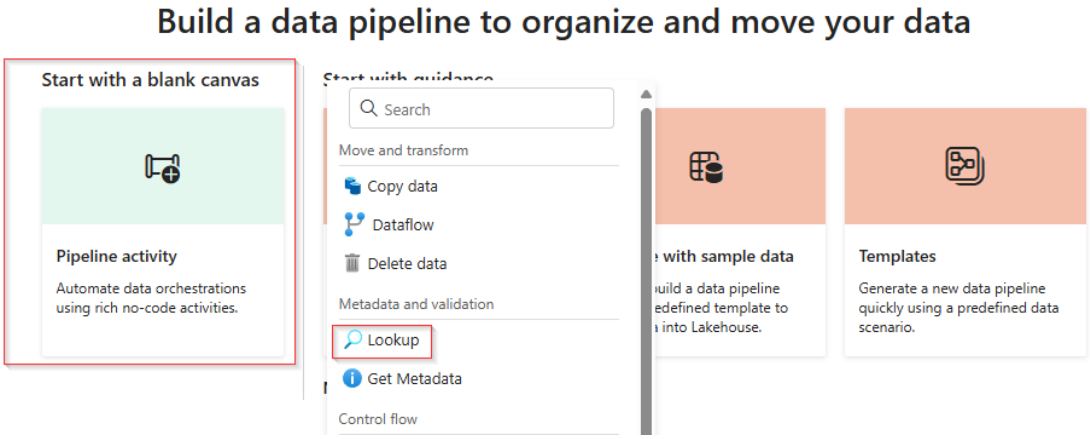


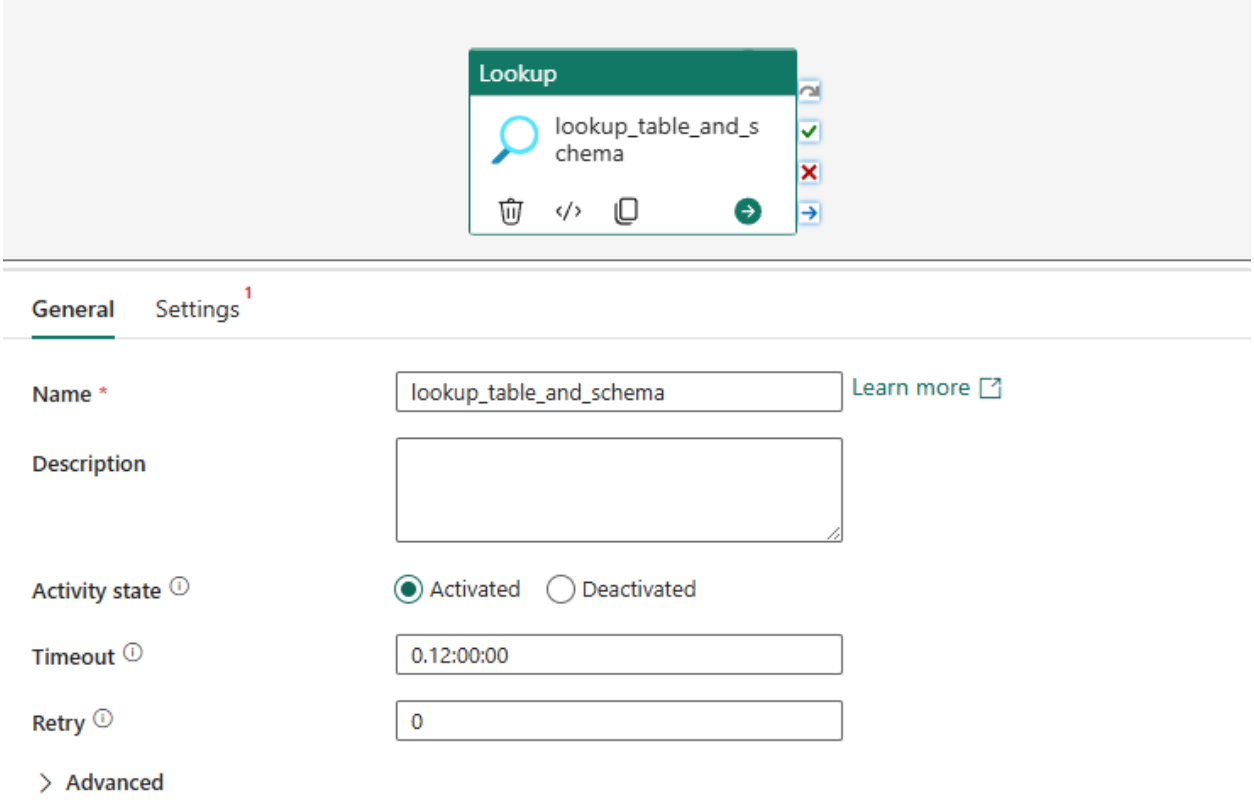
	CustomerID	CustomerNa...	BillToCusto...	CustomerCa...	Buy
1	1001	Dinh Mai	1001	5	
2	1002	Stefan Selezeanu	1002	5	
3	1004	Drishti Bose	1004	6	
4	1005	Bhadram Kamas...	1005	3	
5	1006	Taj Syme	1006	3	
6	1007	Irma Berzina	1007	7	
7	1008	Vedrana Kovacevic	1008	3	
8	1009	Bhagavateeprasa...	1009	5	
9	1010	Serdar ozCelik	1010	4	
10	1011	Rajeev Sandhu	1011	6	
11	1012	Urve Kasesalu	1012	7	
12	1013	Annette Hetu	1013	5	
13	1014	Narendra Tickoo	1014	3	
14	1015	Isa Hulsegge	1015	5	
15	1016	Aive Petrov	1016	4	
16	1018	Nils Kaulins	1018	5	
17	1021	Fabrice Cloutier	1021	4	
18	1022	Nadir Seddigh	1022	5	

Part 2: Dynamic Object Copy

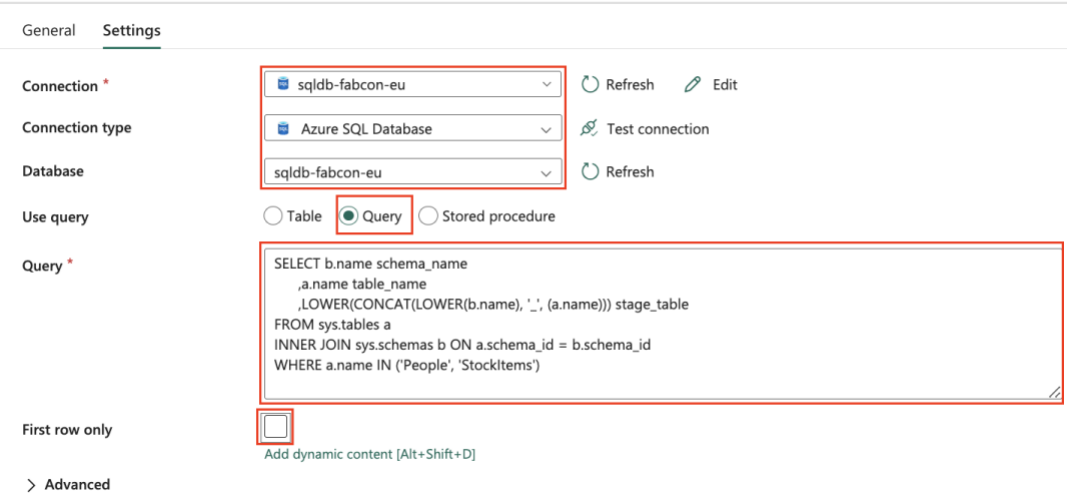
Creating the Dynamic Object Copy Pipeline: To begin, navigate to the **Data Engineering Fabric Landing** page. From there, choose **Data pipeline**. Name the pipeline **dynamic_object_copy** and click **Create**. Once the pipeline is created, click **Pipeline activity** from the landing page. We will start by executing a Lookup activity to retrieve a list of schema and table combinations. From the list of activities, choose **Lookup** and add it to the canvas. Select the **Lookup activity** and rename it to **lookup_table_and_schema**.







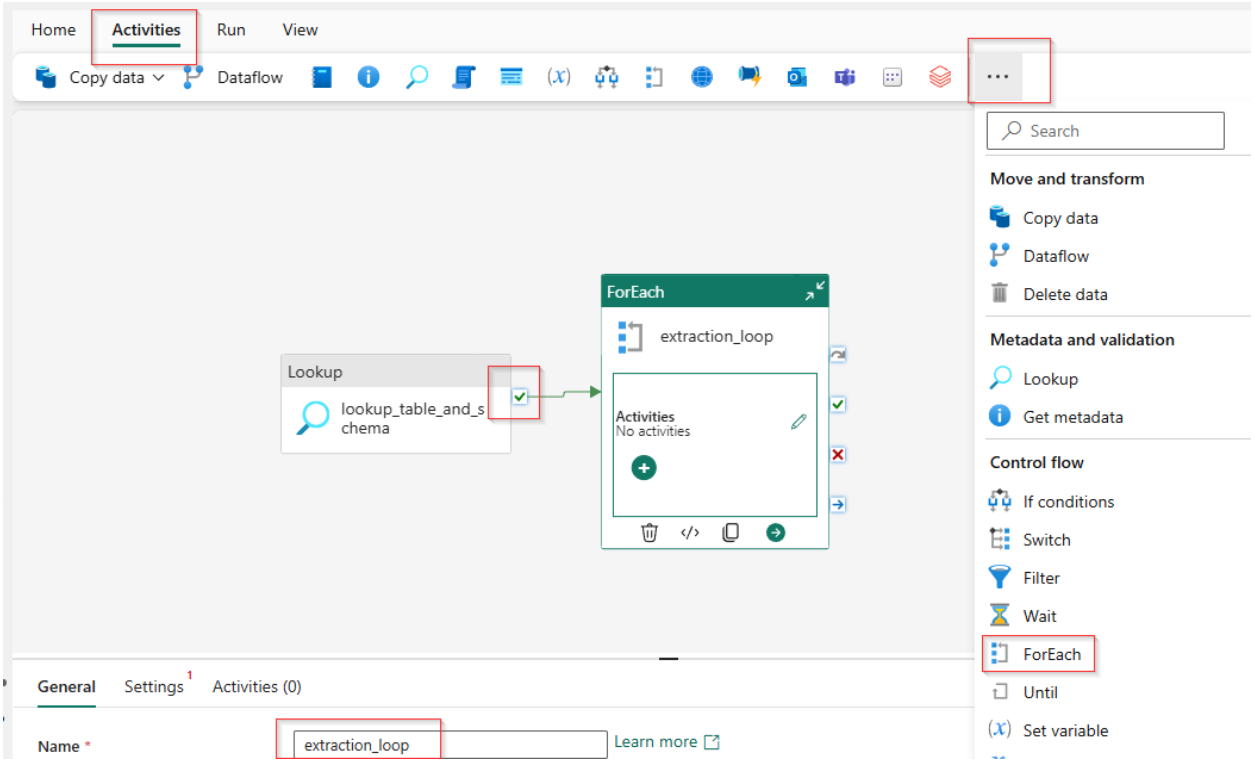
Configuring the Lookup Activity: Navigate to the **Settings** tab of the Lookup activity and select the connection created in part 1 of this lab. Change the **Connection type** to **Azure SQL Database** and click the **Query** radial button. Now, locate the **Dynamic Load Source Query** file shared as part of the lab materials. Copy the query from the file and paste it into the **Query** box of the Lookup activity. Click **Preview Data** to see the output of the Lookup query and be sure to uncheck the box labeled **First row only**.

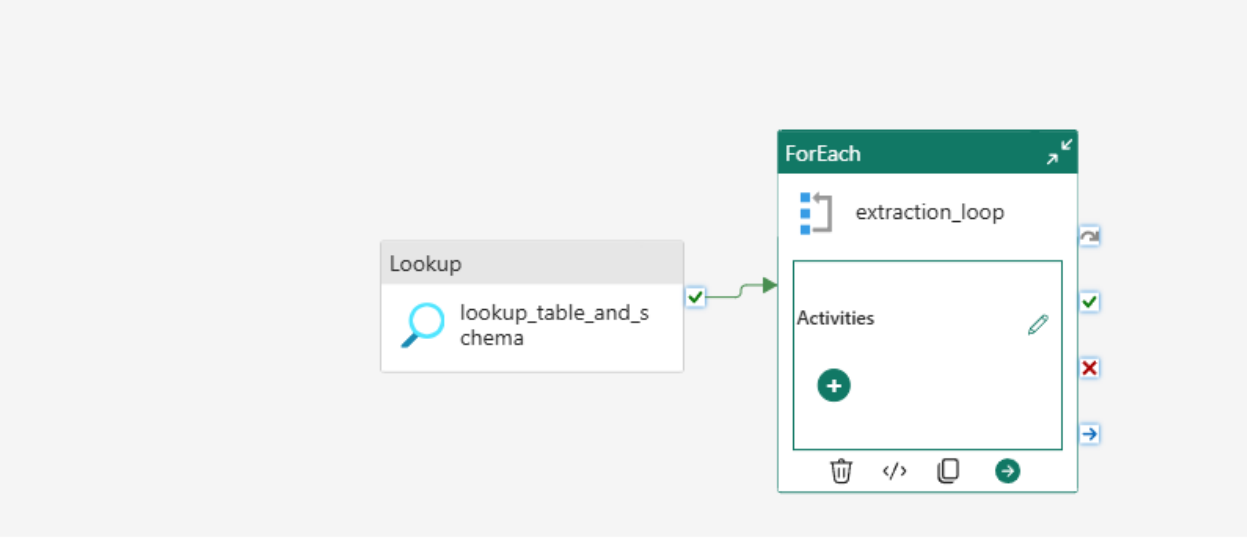


Preview data

	schema_name	table_name	stage_table
1	Application	People	application_people
2	Warehouse	StockItems	warehouse_stockitems

Setting Up the ForEach Loop: Next, open the **Activities** tab and add a **ForEach** activity to the pipeline canvas. Drag from the green checkmark on the **Lookup** activity to the **ForEach** activity to establish a connection between them. Select the ForEach activity and rename it **extraction_loop**. Go to the **Settings** tab of the ForEach activity, click inside the **Items** box, and select **Add dynamic content** to specify the items that will be looped through. Select **lookup_table_and_schema value array** from the list of activity outputs to populate the field, then click **OK**.





General | Settings 1 | Activities (0)

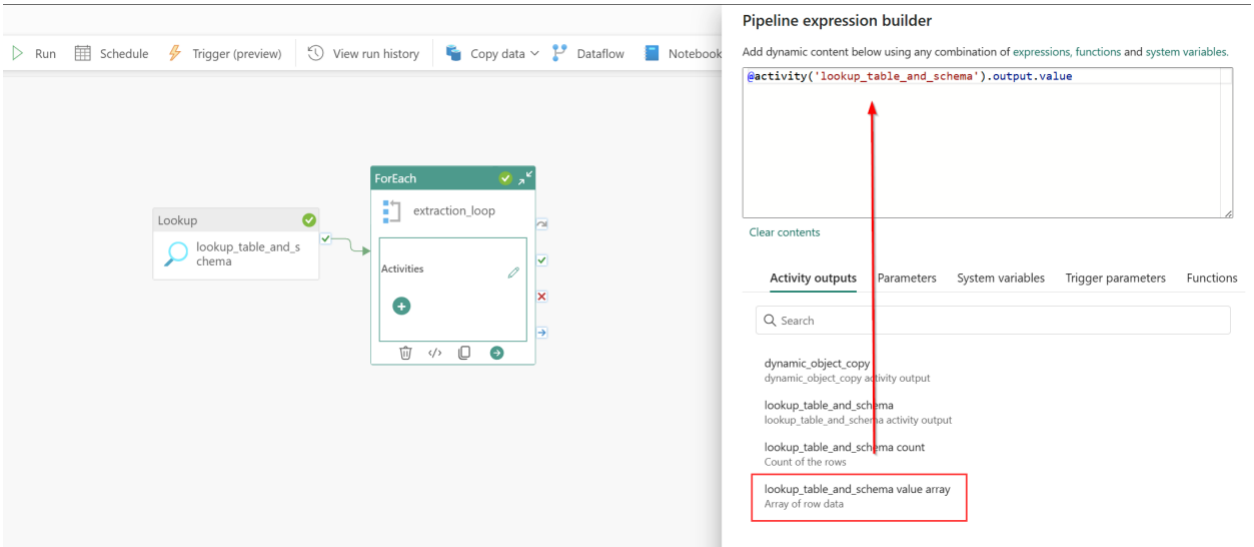
Sequential

Batch count

Items *

This property should be parameterized.

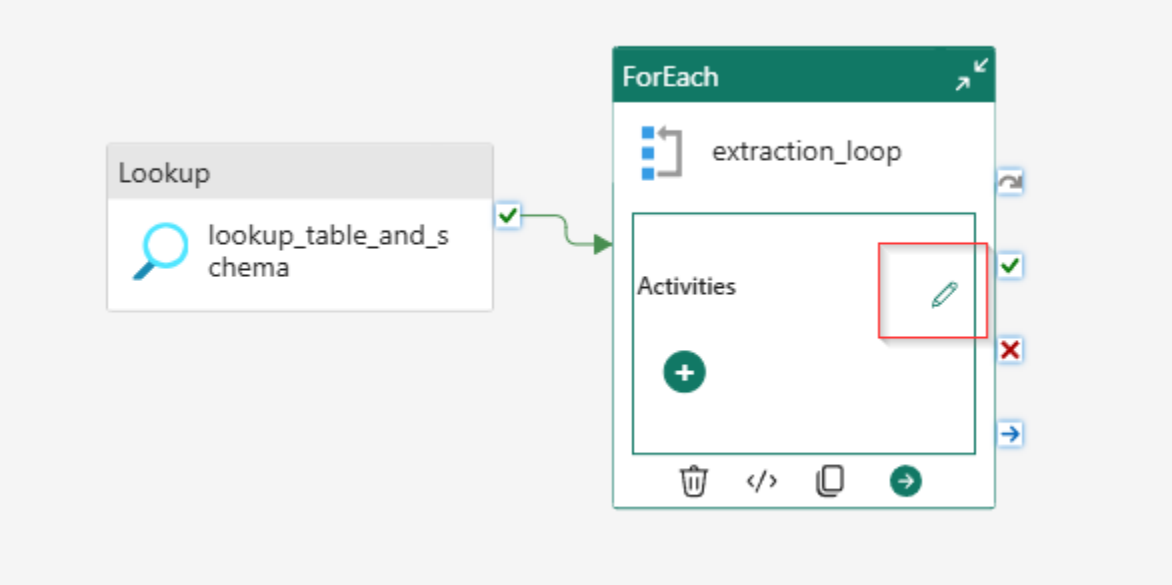
Add dynamic content [Alt+Shift+D]

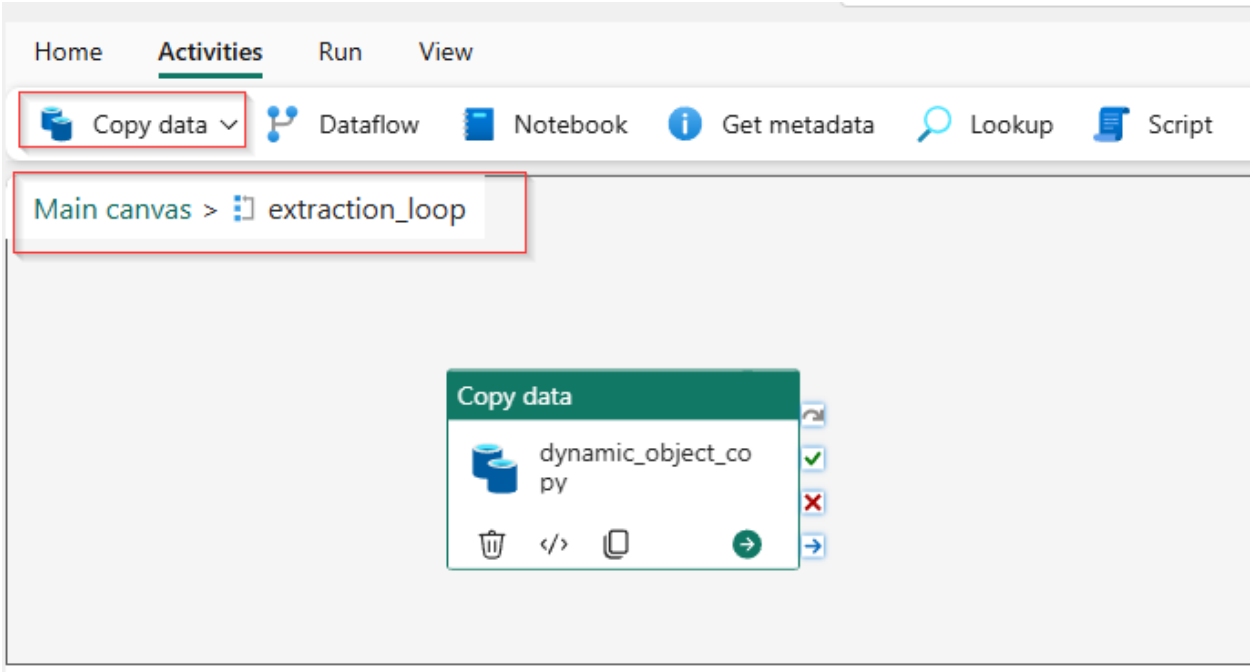


Configuring the Copy Data Activity: Once the ForEach activity is configured, click the **Pencil** icon to edit the loop's contents. In the **Activities** tab, add a **Copy Data** activity to the canvas inside the loop. Select the **Copy Data** task and rename it **dynamic_object_copy**.

In the **Source** tab, set the connection to connection created in part 1, choose **Azure SQL Database** as the connection type, and check the **Enter Manually** box for the table query.

For the **Schema Name** field, click **Add dynamic content**, then select **extraction_loop** from the **ForEach iterator** options. Specify the **schema_name** field from the Lookup activity**,** confirming the expression reads **@item().schema_name**. Click **OK**. Repeat this process for the **Table Name** field, confirming the expression reads **@item().table_name**.





General | Source [1] | Destination [1] | Mapping | Settings

Name * dynamic_object_copy Learn more

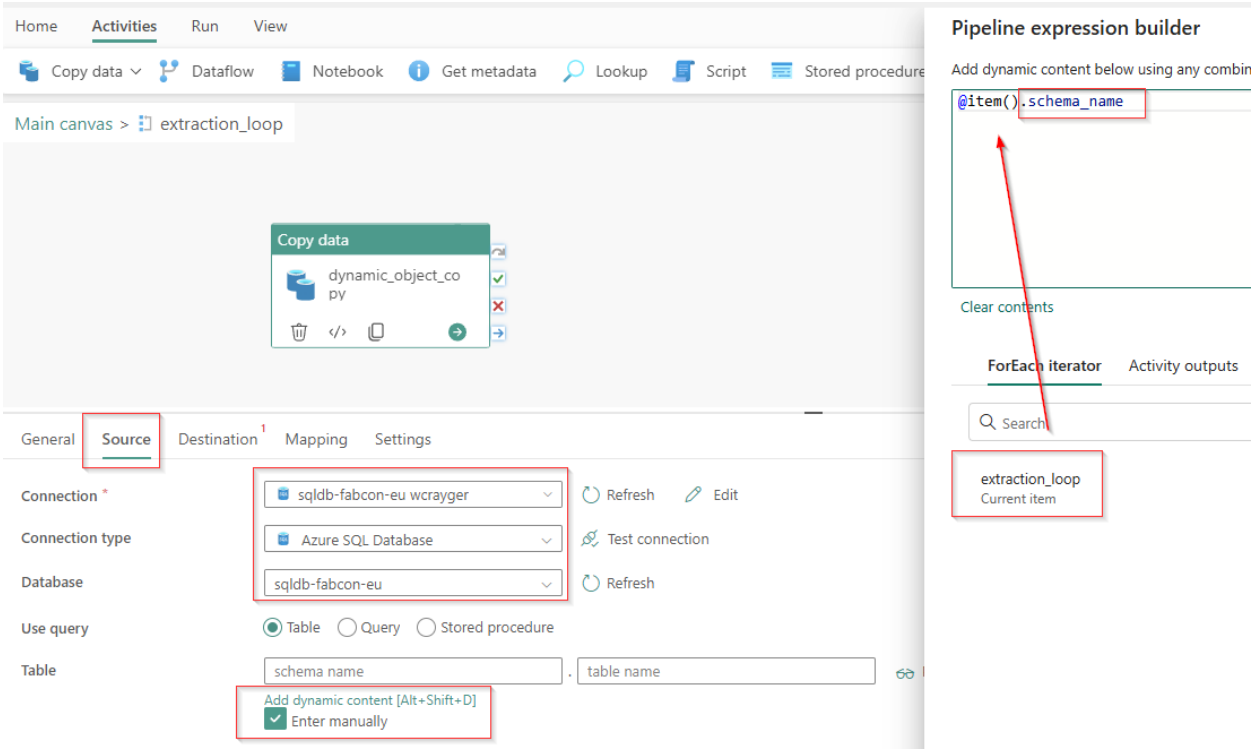
Description

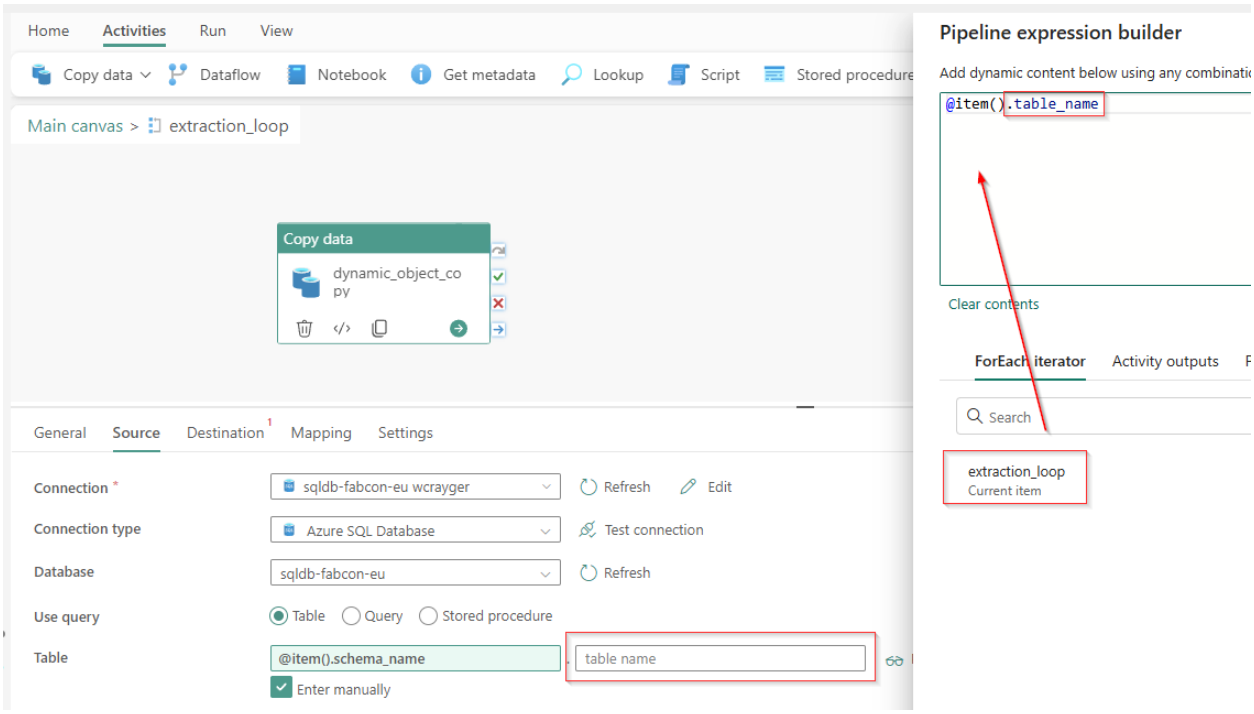
Activity state ⓘ (•) Activated ( ) Deactivated

Timeout ⓘ 0.12:00:00

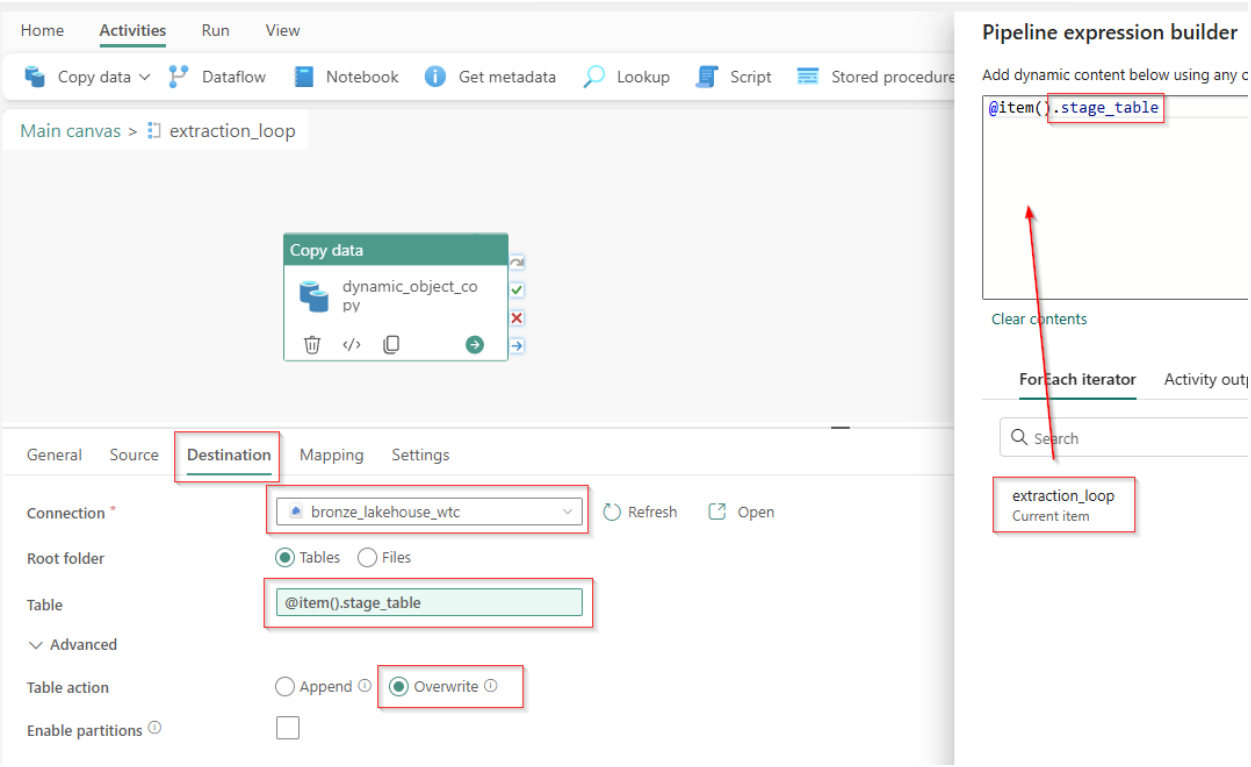
Retry ⓘ 0

> Advanced

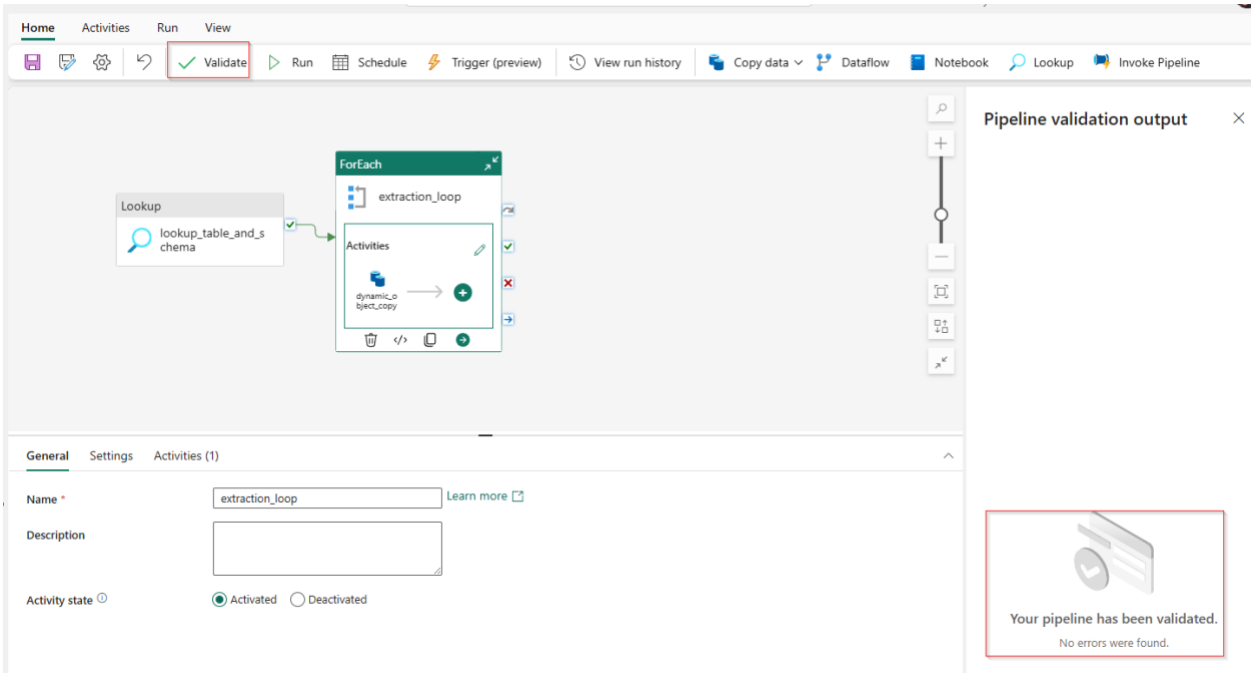

Home
Activities
Run
View
Copy data
Dataflow
Notebook
Get metadata
Lookup
Script
Stored procedure
Main canvas > extraction_loop
Copy data
dynamic_object_co
py
General
Source
Destination
Mapping
Settings
Connection *
sqldb-fabcon-eu wcrayger
Refresh
Edit
Connection type
Azure SQL Database
Test connection
Database
sqldb-fabcon-eu
Refresh
Use query
Table
Query
Stored procedure
Table
schema name
table name
Add dynamic content [Alt+Shift+D]
Enter manually
Pipeline expression builder
Add dynamic content below using any combin
@item().schema_name
Clear contents
ForEach iterator
Activity outputs
Search
extraction_loop
Current item

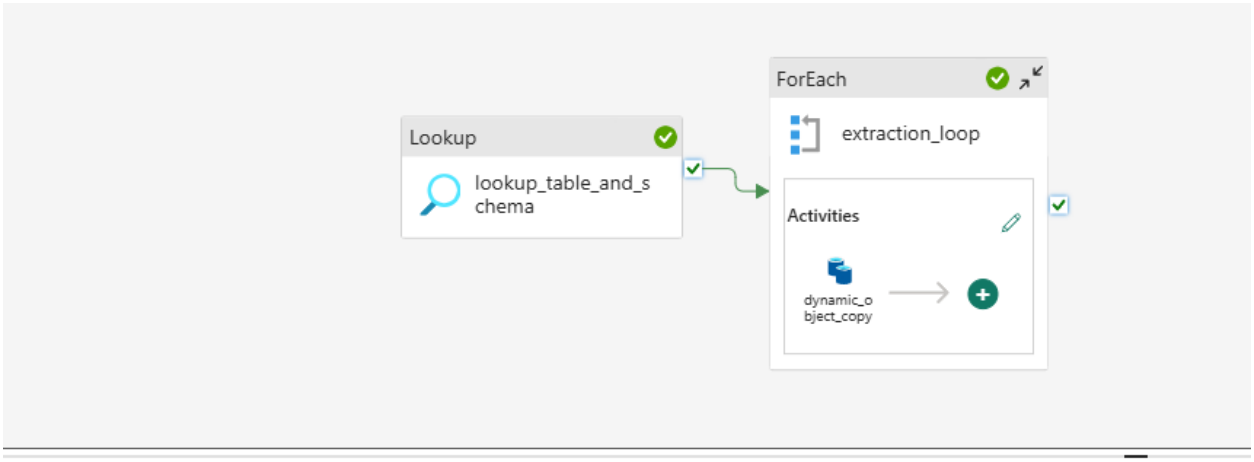

Home
Activities
Run
View
Copy data
Dataflow
Notebook
Get metadata
Lookup
Script
Stored procedure
Main canvas > extraction_loop
Copy data
dynamic_object_co
py
General
Source
Destination
Mapping
Settings
Connection *
sqldb-fabcon-eu wcrayger
Refresh
Edit
Connection type
Azure SQL Database
Test connection
Database
sqldb-fabcon-eu
Refresh
Use query
Table
Query
Stored procedure
Table
@item().schema_name
table name
Enter manually
Pipeline expression builder
Add dynamic content below using any combinati
@item().table_name
Clear contents
ForEach iterator
Activity outputs
Search
extraction_loop
Current item

Setting the Destination: Navigate to the **Destination** tab of the **dynamic_object_copy** activity and select the **Lakehouse** created in **Lab 1**. In the **Table Name** field, click **Add dynamic content** to specify your **bronze lakehouse** (e.g, **bronze_lakehouse_wtc)** as the destination, ensuring the expression reads **@item().stage_table**. Open the **Advanced** menu and change the table action to **Overwrite**.



Validating and Running the Pipeline: Return to the **Home** tab of the pipeline and click **Validate** to ensure that there are no errors in the setup. Once the validation is complete, click **Run** from the activity bar to execute the pipeline. You can monitor the pipeline run in the **Output** tab, where you'll notice multiple **Copy** activities running simultaneously.





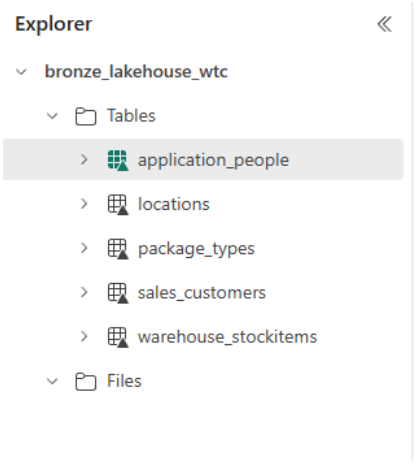
Parameters Variables Settings **Output**

Pipeline run ID: ca44fb94-a55a-4f2a-a4d6-da1e334b6466 **Pipeline status** Succeeded

Filter by keyword Showing 4 items

Activity name	Activity status	Run start
lookup_table_and_schema	Succeeded	9/11/2024, 11:10:28 AM
extraction_loop	Succeeded	9/11/2024, 11:10:43 AM
dynamic_object_copy	Succeeded	9/11/2024, 11:10:44 AM
dynamic_object_copy	Succeeded	9/11/2024, 11:10:44 AM

Confirming the Results: After the pipeline run completes, navigate back to your **Lakehouse** to confirm that the delta tables have been created. If the tables aren't immediately visible, refresh your browser or right-click on **Tables** and select **Refresh**.



application_people

	PersonID	FullName	PreferredNa...	SearchName	IsPermittedT...
1	3	Hudson Onslow	Hudson	Hudson Hudson ...	True
2	5	Eva Muirden	Eva	Eva Eva Muirden	False
3	9	Alica Fatnowna	Alica	Alica Alica Fatno...	True
4	10	Stella Rosenhain	Stella	Stella Stella Rose...	True
5	11	Ethan Onslow	Ethan	Ethan Ethan Onsl...	True
6	16	Archer Lamble	Archer	Archer Archer La...	False
7	17	Piper Koch	Piper	Piper Piper Koch	True
8	18	Katie Darwin	Katie	Katie Katie Darwin	True
9	20	Jack Potter	Jack	Jack Jack Potter	True
10	2	Kayla Woodcock	Kayla	Kayla Kayla Woo...	True
11	4	Isabella Rupp	Isabella	Isabella Isabella ...	True

You have now successfully completed **Part 2** of the lab.