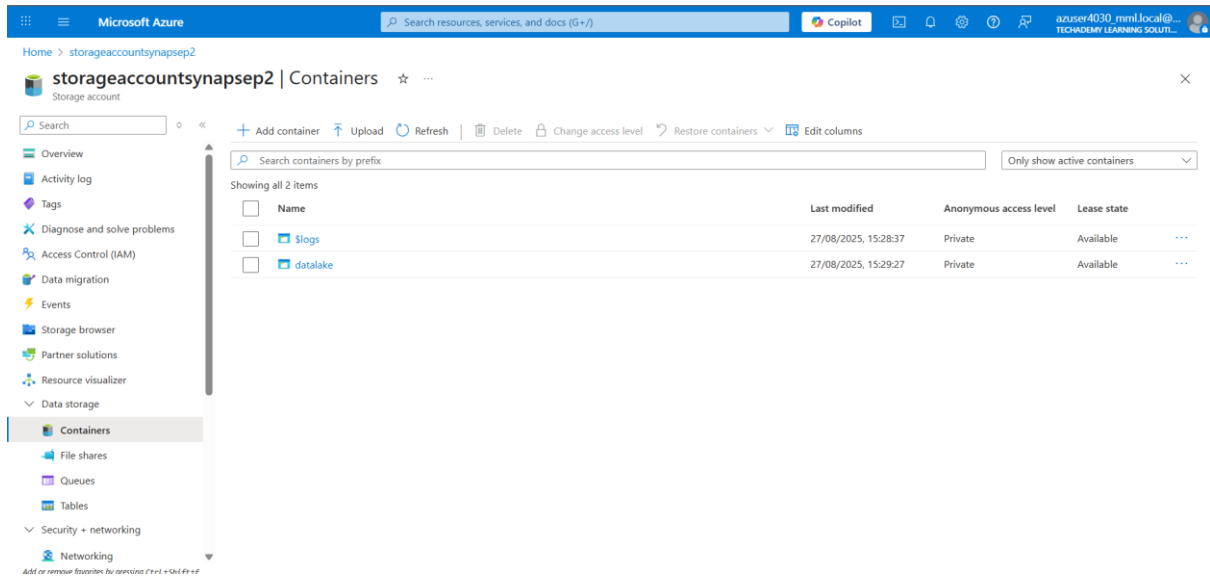


CODING CHALLENGE (AZURE DEVOPS)

TRAINEE NAME: Swathi Baskaran

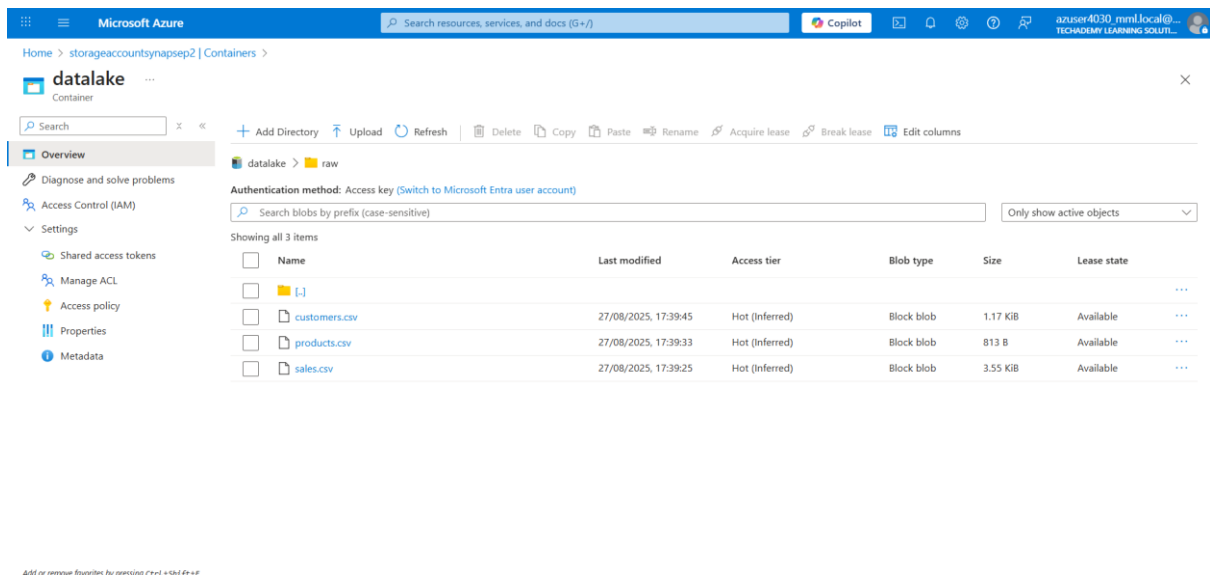
Creating a container in Azure Storage Account to store data



The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the Microsoft Azure logo, a search bar, and a user profile. The left sidebar contains a list of navigation options: Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, Storage browser, Partner solutions, Resource visualizer, Data storage, Containers (selected), File shares, Queues, Tables, Security + networking, and Networking. The main pane displays the 'Containers' view for the storage account 'storageaccountsynapsep2'. It shows a list of containers with columns: Name, Last modified, Anonymous access level, and Lease state. The containers listed are 'slogs' and 'datalake'. The 'datalake' container is selected, and its details are shown in the main pane.

Name	Last modified	Anonymous access level	Lease state
slogs	27/08/2025, 15:28:37	Private	Available
datalake	27/08/2025, 15:29:27	Private	Available

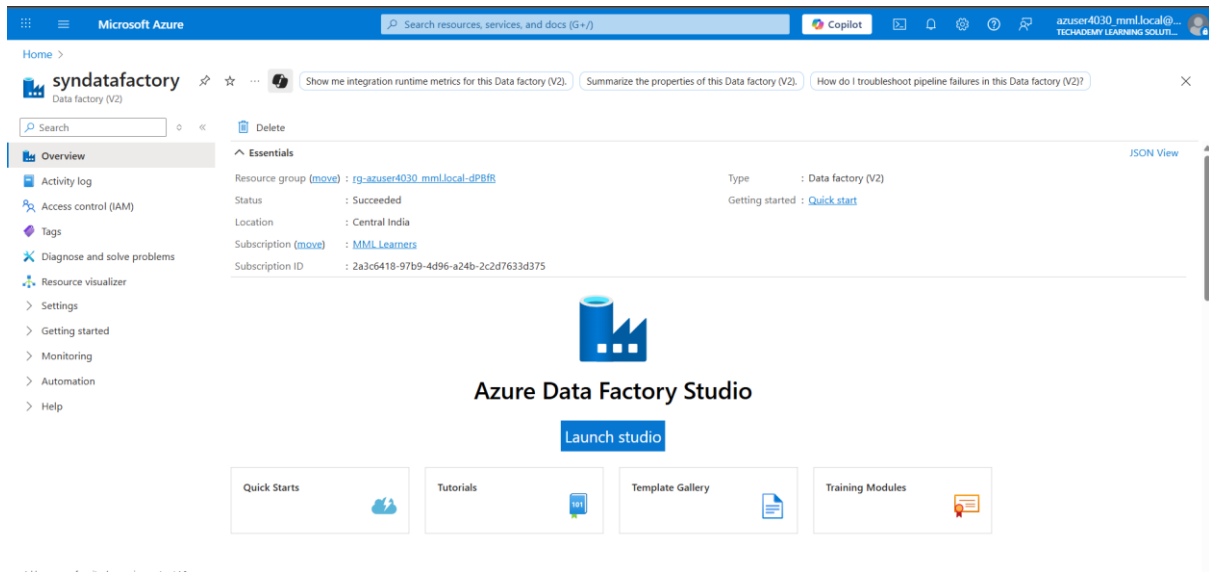
Uploading the csv files into the container



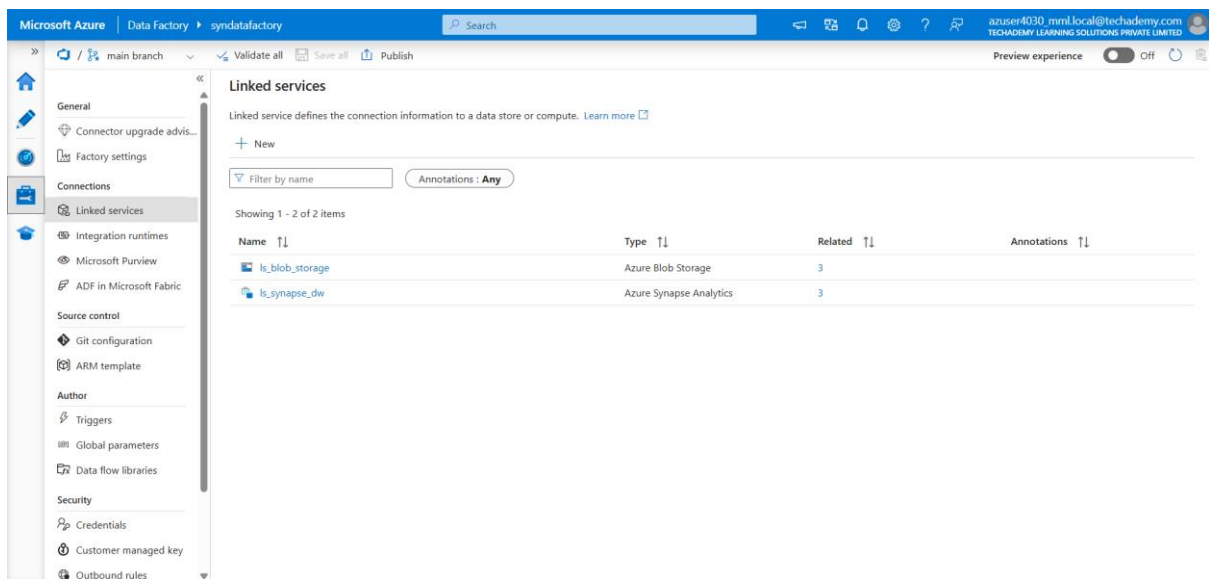
The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the Microsoft Azure logo, a search bar, and a user profile. The left sidebar contains a list of navigation options: Overview, Diagnose and solve problems, Access Control (IAM), Settings, Shared access tokens, Manage ACL, Access policy, Properties, and Metadata. The main pane displays the 'Overview' view for the 'datalake' container. It shows a list of blobs with columns: Name, Last modified, Access tier, Blob type, Size, and Lease state. The blobs listed are 'customers.csv', 'products.csv', and 'sales.csv'. The 'customers.csv' blob is selected, and its details are shown in the main pane.

Name	Last modified	Access tier	Blob type	Size	Lease state
customers.csv	27/08/2025, 17:39:45	Hot (Inferred)	Block blob	1.17 KiB	Available
products.csv	27/08/2025, 17:39:33	Hot (Inferred)	Block blob	813 B	Available
sales.csv	27/08/2025, 17:39:25	Hot (Inferred)	Block blob	3.55 KiB	Available


Creating a Data Factory Studio



Creating Linked Services for both Azure Synapse Analytics and Azure Blob Storage



Azure Blob Storage

Edit linked service
 Azure Blob Storage [Learn more](#)

Name *

ls_blob_storage

Description

Connect via integration runtime *

☒ AutoResolveIntegrationRuntime

Authentication type

Account key


Account selection method

☐ From Azure subscription ☒ Enter manually


Storage account name *

storageaccountsynapse2

Storage account key *

Apply **Cancel**  Test connection

Azure Synapse Analytics

Edit linked service
 Azure Synapse Analytics [Learn more](#)

Name *

ls_synapse_dw

Description

Connect via integration runtime *

☒ AutoResolveIntegrationRuntime

Version

☒ 2.0 (Recommended) ☐ 1.0

Account selection method

☐ From Azure subscription ☒ Enter manually

Fully qualified domain name *

synapseanalprojectws.sqlazuresynapse.net


Database name *

dwpool

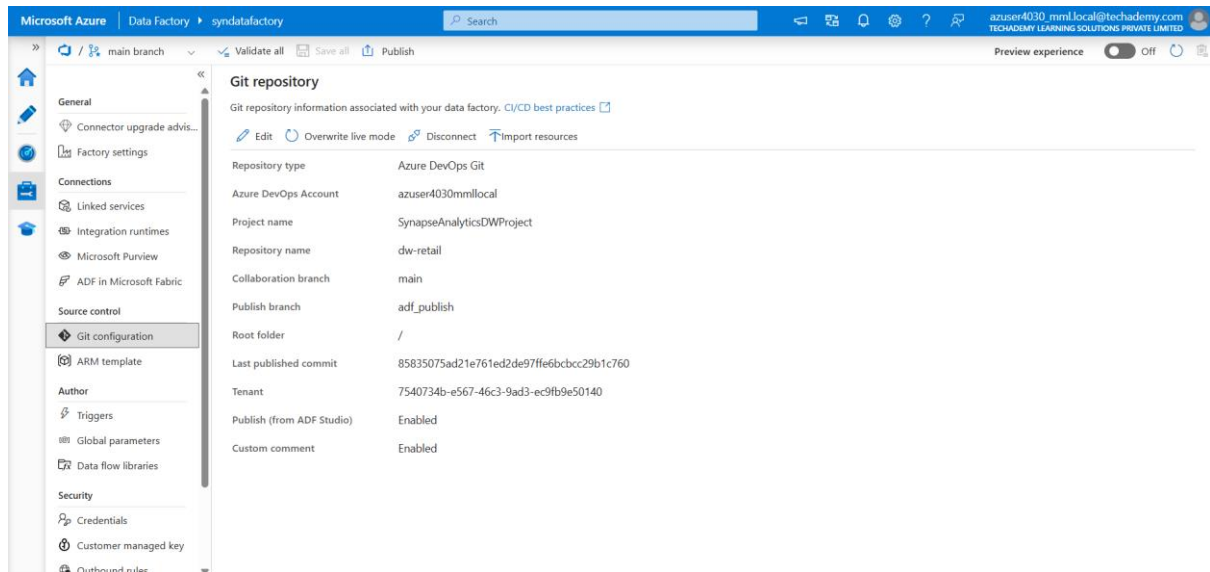
Authentication type *

SQL authentication

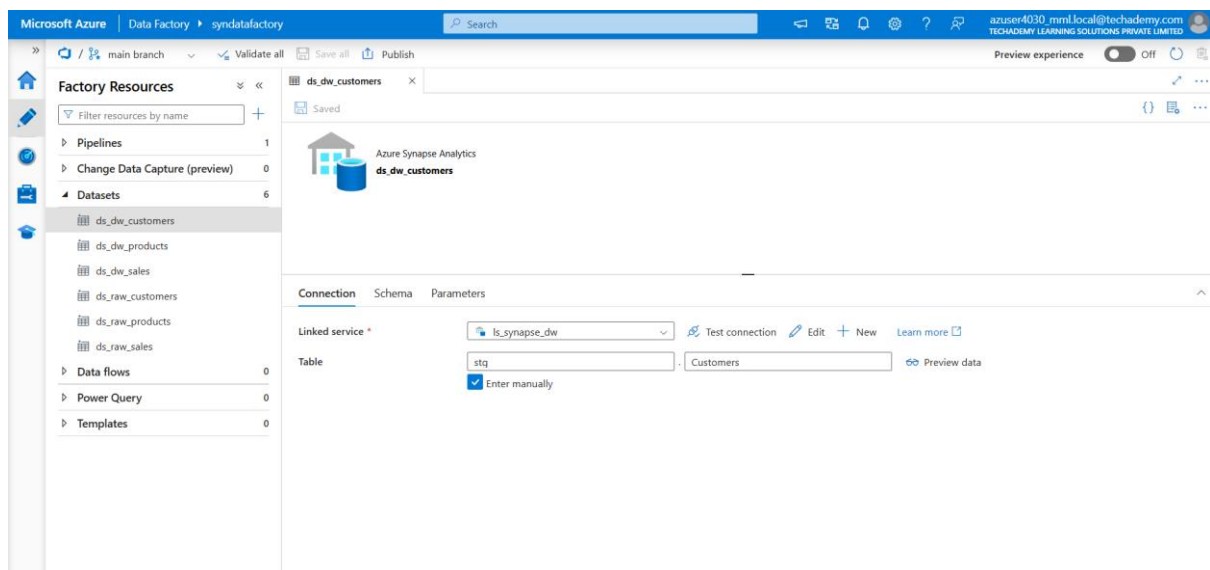
User name *

Save **Cancel**  Test connection

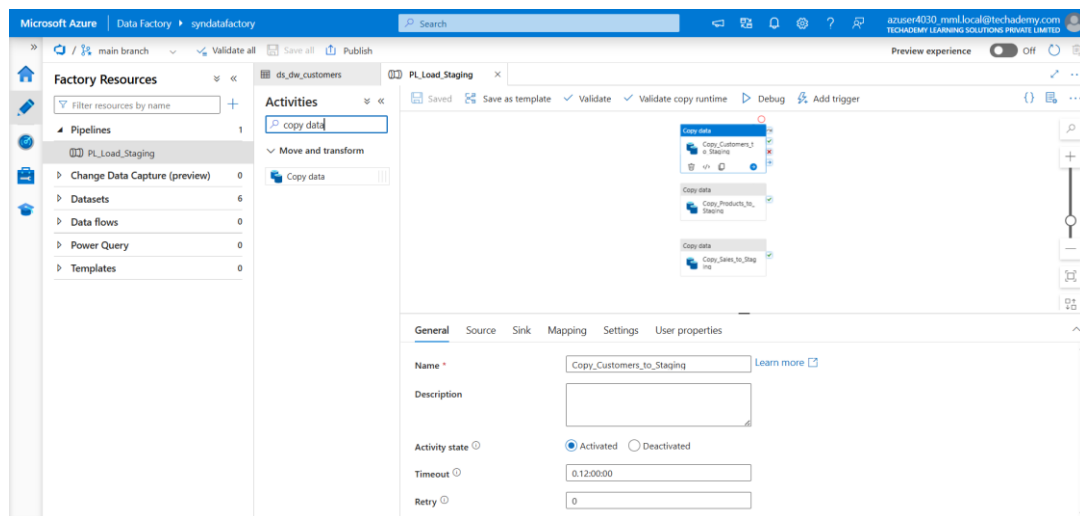
Integrating ADF with Git (Azure DevOps)



Creating relevant Datasets in Azure Data Factory



Creating a pipeline by using copy activity



Setting the source for the Copy Activity

General **Source** Sink Mapping Settings User properties

Source dataset * ds_raw_customers [Open](#) [New](#) [Preview data](#) [Learn more](#)

File path type ☒ File path in dataset ☐ Prefix ☐ Wildcard file path ☐ List of files

Filter by last modified Start time (UTC) End time (UTC)

Recursively ☒

Enable partitions discovery ☐

Max concurrent connections

Skip line count

Additional columns [New](#)

Setting the sink for the Copy Activity

General Source **Sink** Mapping Settings User properties

Sink dataset * ds_dw_customers [Open](#) [New](#) [Learn more](#)

Copy method ☒ Copy command ☐ Bulk insert ☐ Upsert

Allow copy command ☒

Default values [New](#)

Additional options [New](#)

Table option ☒ Use existing ☐ Auto create table

Pre-copy script

Write batch timeout e.g. 00:30:00

Creating a sql Pool in Azure Synapse Analytics

Properties

Name

dwpool

Workspace SQL endpoint

synapseanalprojectws.sql.azuresynapse.net

Status

Online

Creation date

08/27/2025, 3:46:11 PM +05:30

Connection strings

ADO.NET (SQL authentication)

Server=tcp:synapseanalprojectws.sql.azuresynapse.net,1433;Initial Catalog=dwpool;Persist S...

Configuration

Workspace

Close

The SQL pool is running

SQL pools

The serverless SQL pool, Built-in, is immediately available for your workspace. Dedicated SQL pools can be configured to adapt to team or organizational requirements and constraints. [Learn more](#)

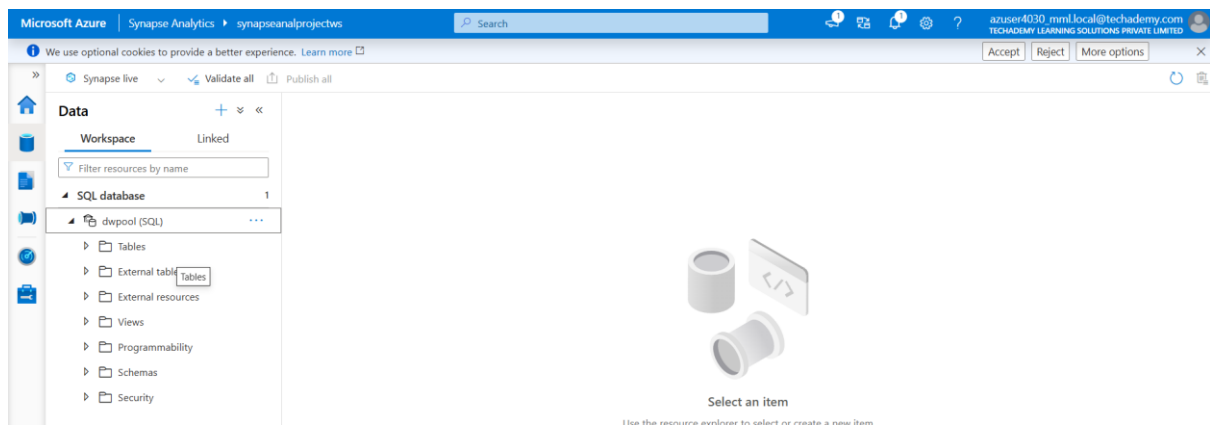
+ New Refresh

Filter by name

Showing 1-2 of 2 items (1 Serverless, 1 Dedicated)

Name	Type	Status	Size
Built-in	Serverless	Online	Auto
dwpool	Dedicated	Online	DW100c

Connect it to the data



Creating tables

```
Creating ETL Log | Creating_Fact_Tables | Other users in your workspace may have access to modify this item. Do not use this item may have access to the workspace.

Run | Undo | Publish | Query plan | Connect to: dwpool | Use database: dwpool

1 | -----
2 | -- 01. Drop FactSales if it already exists
3 | -----
4 | IF OBJECT_ID('dw.FactSales') IS NOT NULL
5 |     DROP TABLE dw.FactSales;
6 | GO
7 |
8 | -----
9 | -- 02. Create FactSales table
10 | -----
11 | CREATE TABLE dw.FactSales
12 | (
13 |     SaleID INT NOT NULL,
14 |     SaleDate DATETIME2 NOT NULL,
15 |     CustomerSK INT NOT NULL,
16 |     ProductSK INT NOT NULL,
17 |     Quantity INT NOT NULL,
18 |     UnitPrice DECIMAL(18,2) NOT NULL,
19 |     Amount DECIMAL(18,2) NOT NULL,
20 |     LoadDate DATETIME2 NULL -- no default, we set explicitly
21 | )
22 | WITH
23 | (
24 |     DISTRIBUTION = HASH(SaleID),
25 |     CLUSTERED COLUMNSTORE INDEX
26 | );
27 | GO
28 |

Run | Undo | Publish | Query plan | Connect to: dwpool | Use database: dwpool

1 | IF NOT EXISTS (SELECT * FROM sys.schemas WHERE name = 'staging')
2 | BEGIN
3 |     EXEC('CREATE SCHEMA staging');
4 | END;
5 |
6 | TRUNCATE TABLE staging.Sales;
7 |
8 | CREATE TABLE staging.Sales
9 | (
10 |     Item_Identifier NVARCHAR(50),
11 |     Item_Weight FLOAT,
12 |     Item_Fat_Content NVARCHAR(50),
13 |     Item_Visibility FLOAT,
14 |     Item_Type NVARCHAR(100),
15 |     Item_MRP FLOAT,
16 |     Outlet_Identifier NVARCHAR(50),
17 |     Outlet_Establishment_Year INT,
18 |     Outlet_Size NVARCHAR(50),
19 |     Outlet_Location_Type NVARCHAR(50),
20 |     Outlet_Type NVARCHAR(50),
21 |     Item_Outlet_Sales FLOAT
22 | );
23 |
24 | SELECT TOP 100 *
25 | FROM staging.Sales;
26 |
```

Running the Pipeline in Azure Data Factory

The screenshot shows the Azure Data Factory interface for a pipeline named **PL_Load_Staging**. The pipeline is currently in an **In progress** state. The left sidebar shows the **Activities** pane with a search for **copy data** and a list of activities under **Move and transform**. The main canvas displays the pipeline flow with three sequential **Copy data** activities: **Copy_Customers_to_Staging**, **Copy_Products_to_Staging**, and **Copy_Sales_to_Staging**. The bottom pane shows the **Output** tab with the **Pipeline run ID** 29e4a21d-3421-4b5a-9d8c-2850ac0ecc5c. The status is **In progress**. Below the status, it says "Showing 0 - 0 of 0 items".

Successfully ran the pipeline

The screenshot shows the Azure Data Factory interface for the same pipeline **PL_Load_Staging**, now in a **Succeeded** state. The left sidebar shows the **Factory Resources** pane with a list of resources: **Pipelines** (1), **Change Data Capture (preview)** (0), **Datasets** (6), **Data flows** (0), **Power Query** (0), and **Templates** (0). The main canvas displays the pipeline flow with three sequential **Copy data** activities: **Copy_Customers_to_Staging**, **Copy_Products_to_Staging**, and **Copy_Sales_to_Staging**. The bottom pane shows the **Output** tab with the **Pipeline run ID** 4855393a-b3ad-487e-8376-20783a89a83e. The status is **Succeeded**. Below the status, it says "Showing 1 - 3 of 3 items".

Activity name	Activity status	Activity name	Run start	Duration	Integration runtime
Copy_Products_to_Staging	Succeeded	Copy data	8/29/2025, 11:51:10 AM	16s	AutoResolveIntegrationRuntime (Central In
Copy_Sales_to_Staging	Succeeded	Copy data	8/29/2025, 11:51:10 AM	16s	AutoResolveIntegrationRuntime (Central In
Copy_Customers_to_Staging	Succeeded	Copy data	8/29/2025, 11:51:10 AM	16s	AutoResolveIntegrationRuntime (Central In

Running the relevant SQL queries to validate the pipeline execution

BusinessQueries

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

Run

Undo

Publish

Query plan

Connect to dwpool

Use database dwpool

1

-- Build Some Business Queries

2

-- Total Sales by Product Category

3

SELECT dp.Category, SUM(f.Amount) AS TotalSales

4

FROM dw.FactSales f

5

JOIN dw.DimProduct dp ON f.ProductSK = dp.ProductSK

6

GROUP BY dp.Category

7

ORDER BY TotalSales DESC;

Results

Messages

View

Table

Chart

Export results

Search

Category	TotalSales
Electronics	1400867.40
Grocery	1187857.20
Home	882835.80
Clothing	507745.20
Sports	208106.40

00:00:03

Query executed successfully.