AZURE PROJECT

This project involved building a robust data pipeline using Azure Data factory and PySpark for advanced data transformation and Azure Synapse for Data Warehousing.  
  
 **Technologies Used :**

Azure Data factory:

Orchestrated data pipeline

Databricks & PySpark:

Transformed and processed data.

Azure Data Lake:

Stored raw and transformed data.

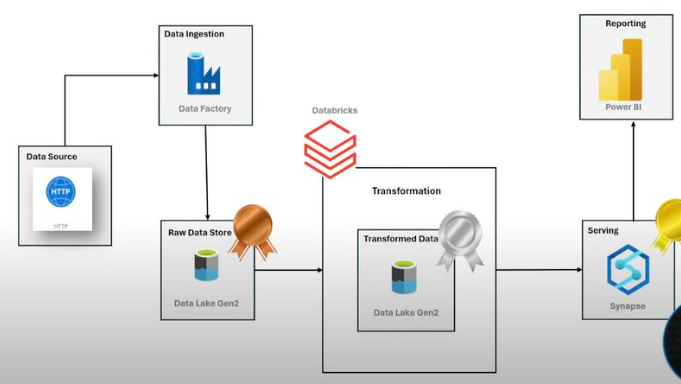
Azure Synapse Analytics:

Managed data warehousing.

Power BI:

Created interactive dashboards.

**Architecture**



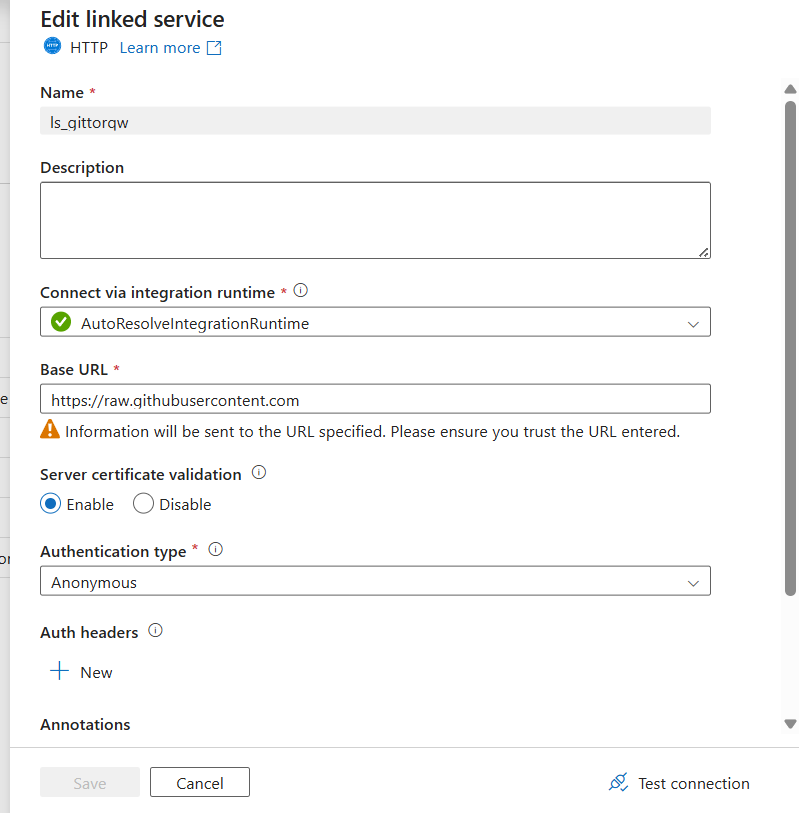
Resources:

Original Dataset: <https://lnkd.in/daJKTHkG>

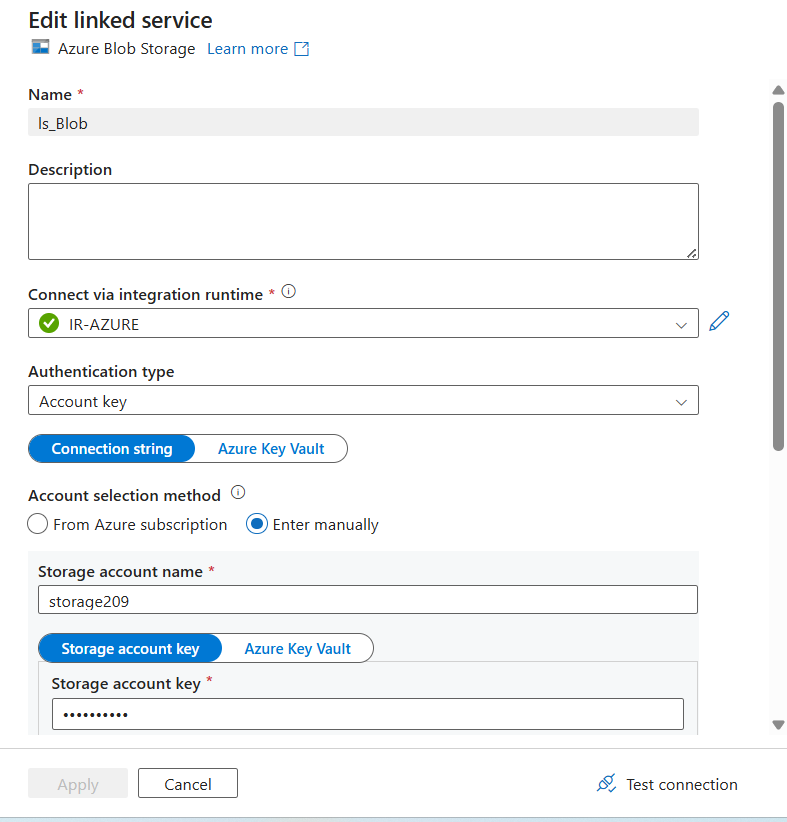
My project Source : https://github.com/anshlambaoldgit/Adventure-Works-Data-Engineering-Project/tree/main/Data

1.Copy Data from HTTP (Source) to Bronze Layer

First create link service for HTTP



Create link service for Azure Blob Storage



Copy the Data Dynamically from source to Bronze Layer:

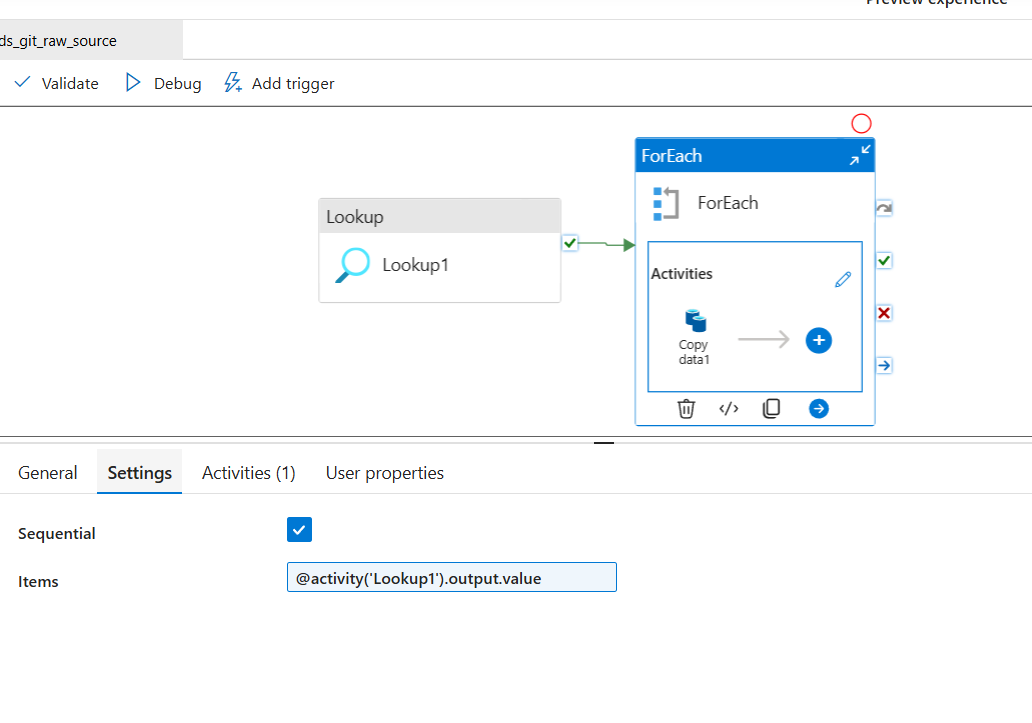
Activities for copy

1. LookUp Activity
2. For Each Activity
3. Copy Activity

LookUp Activity Source Dataset: Json script

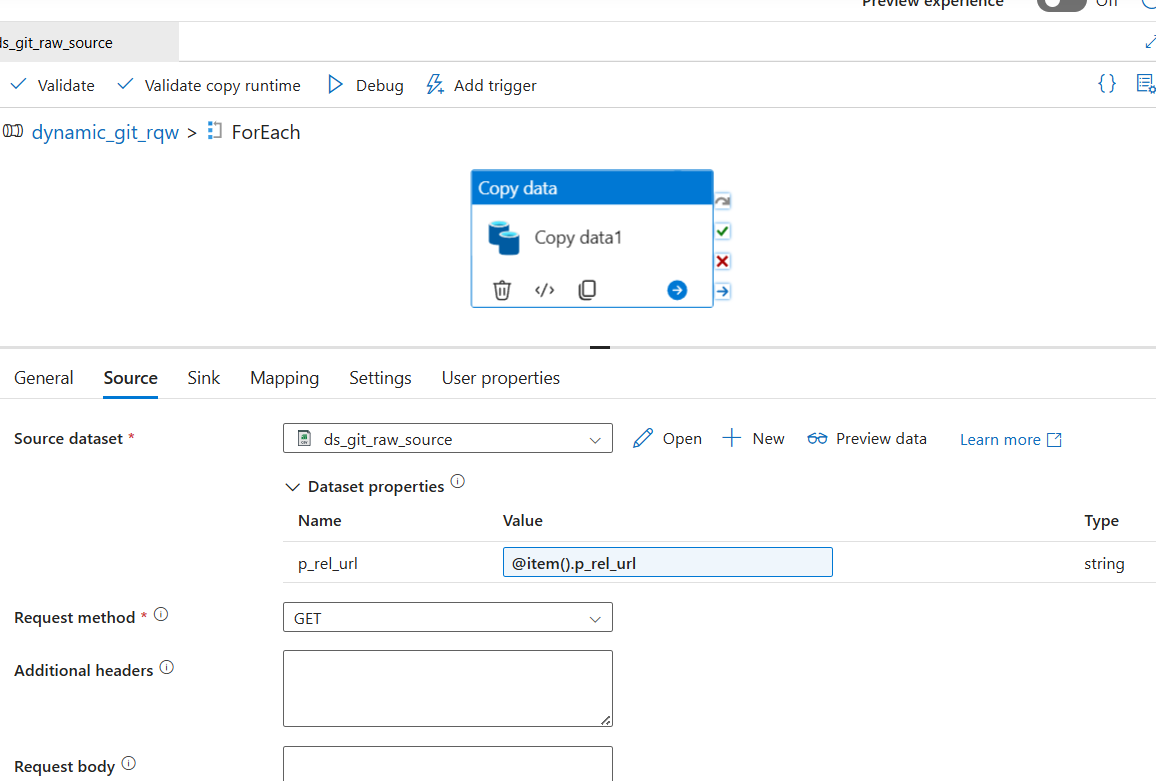


ForEach Activity:

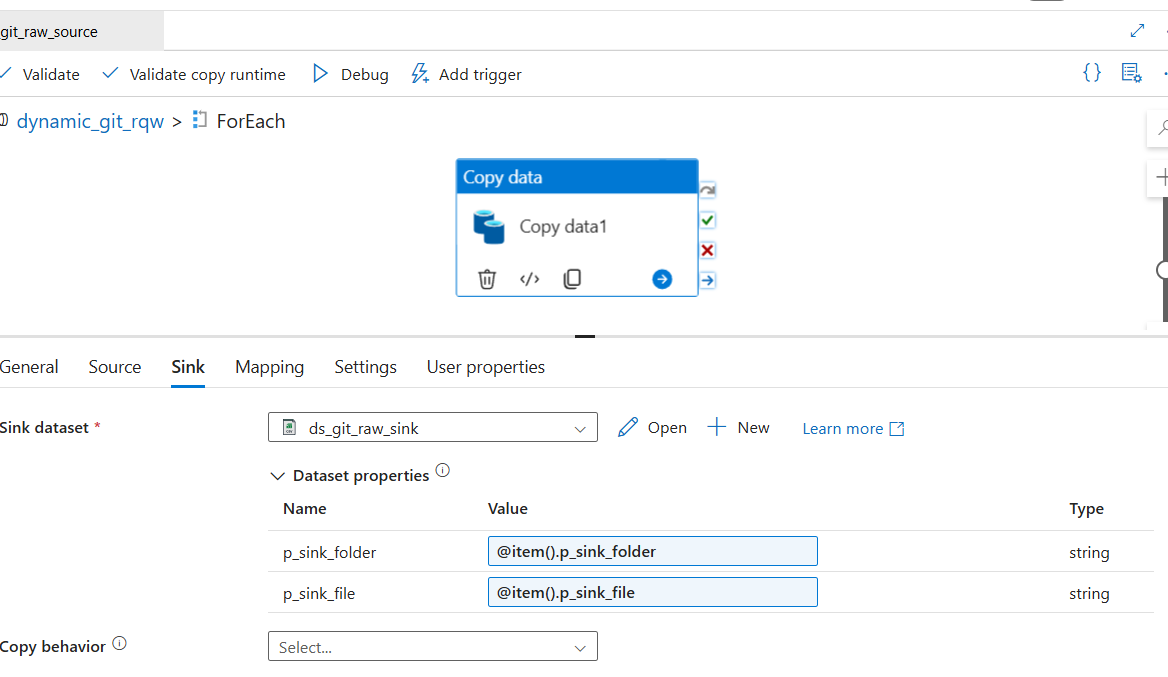


Copy Activity

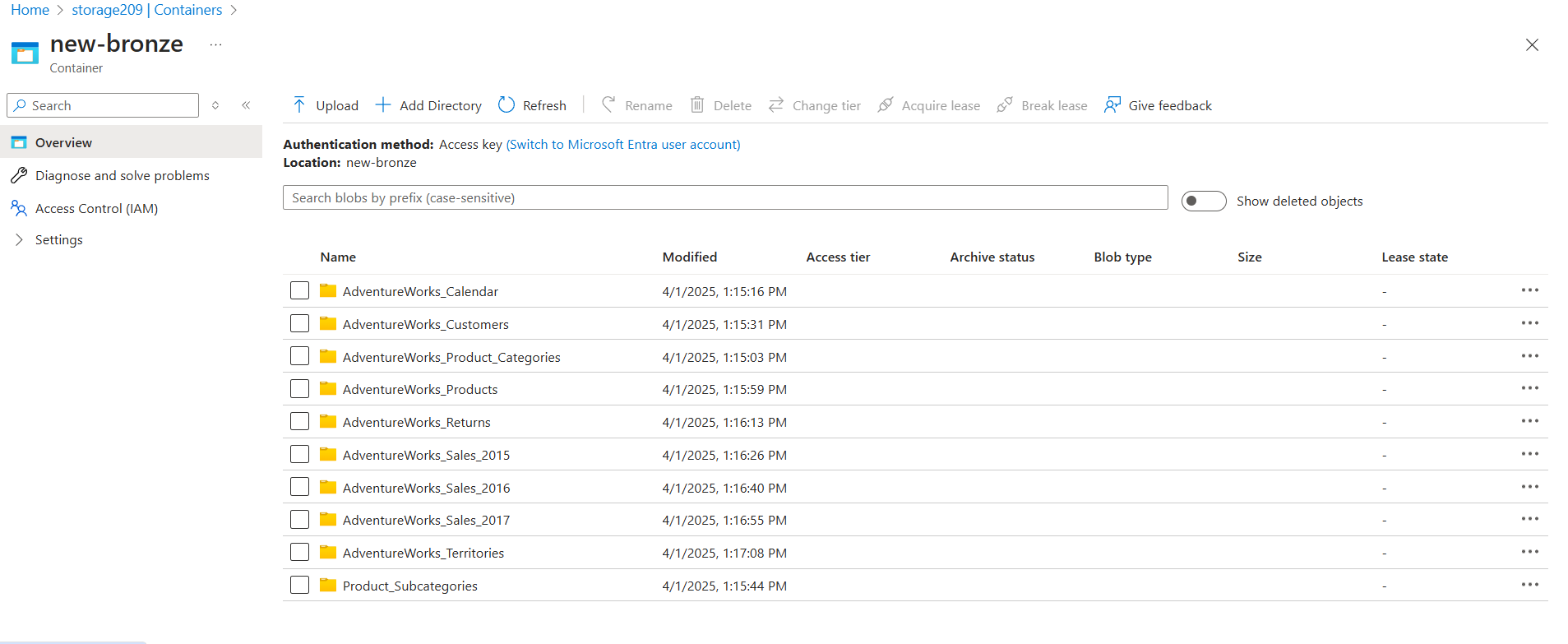
Source:



Sink:

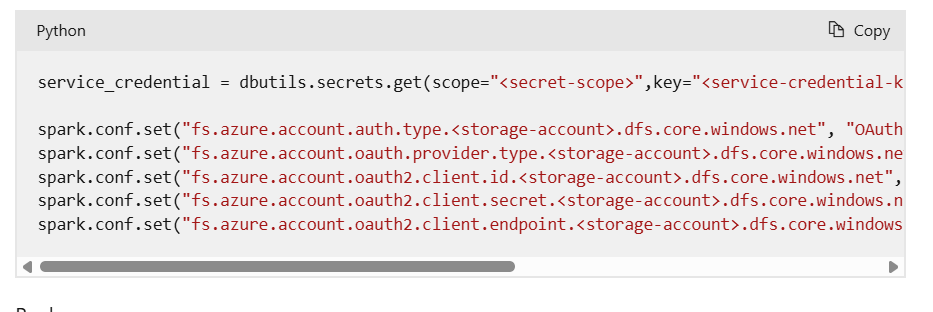


Final Output:



**Transformations Azure DataBricks:**

Access storage using a service principal and Microsoft Entra ID

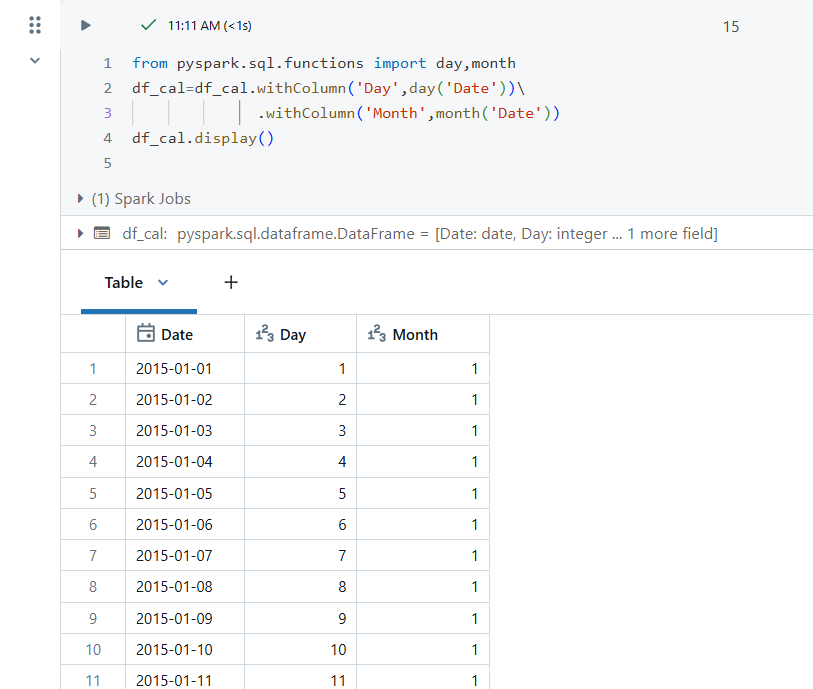


Read the Data

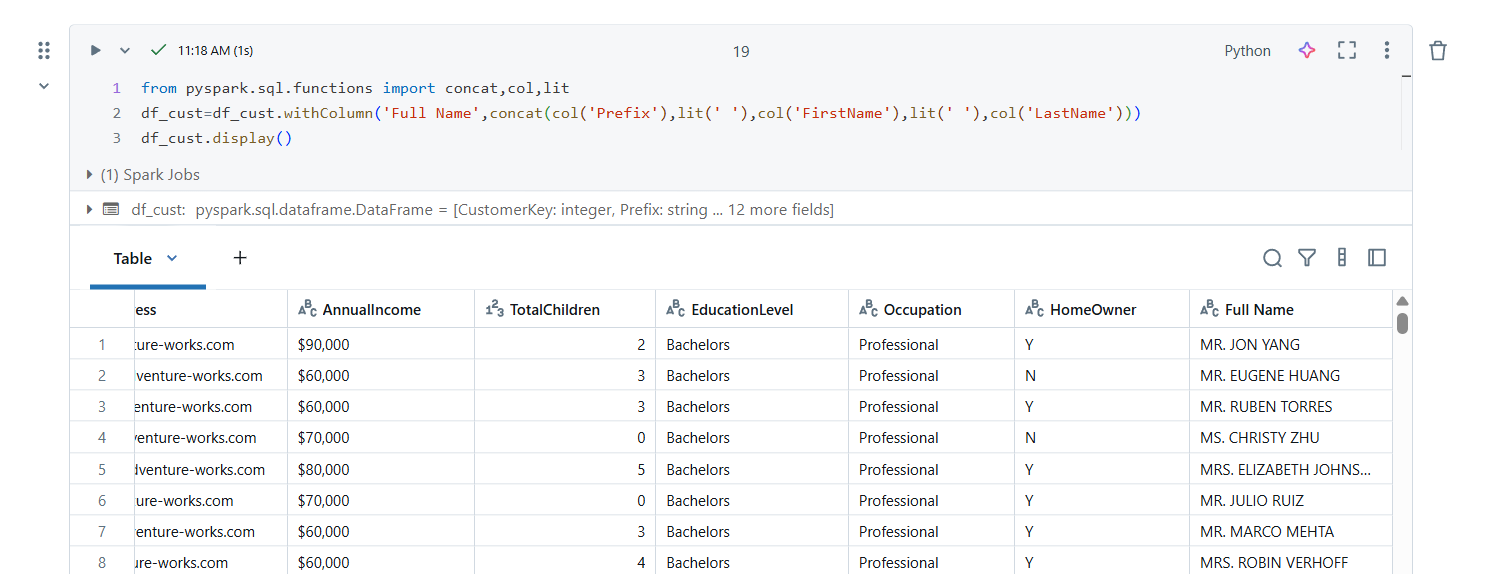


Some Transformations :

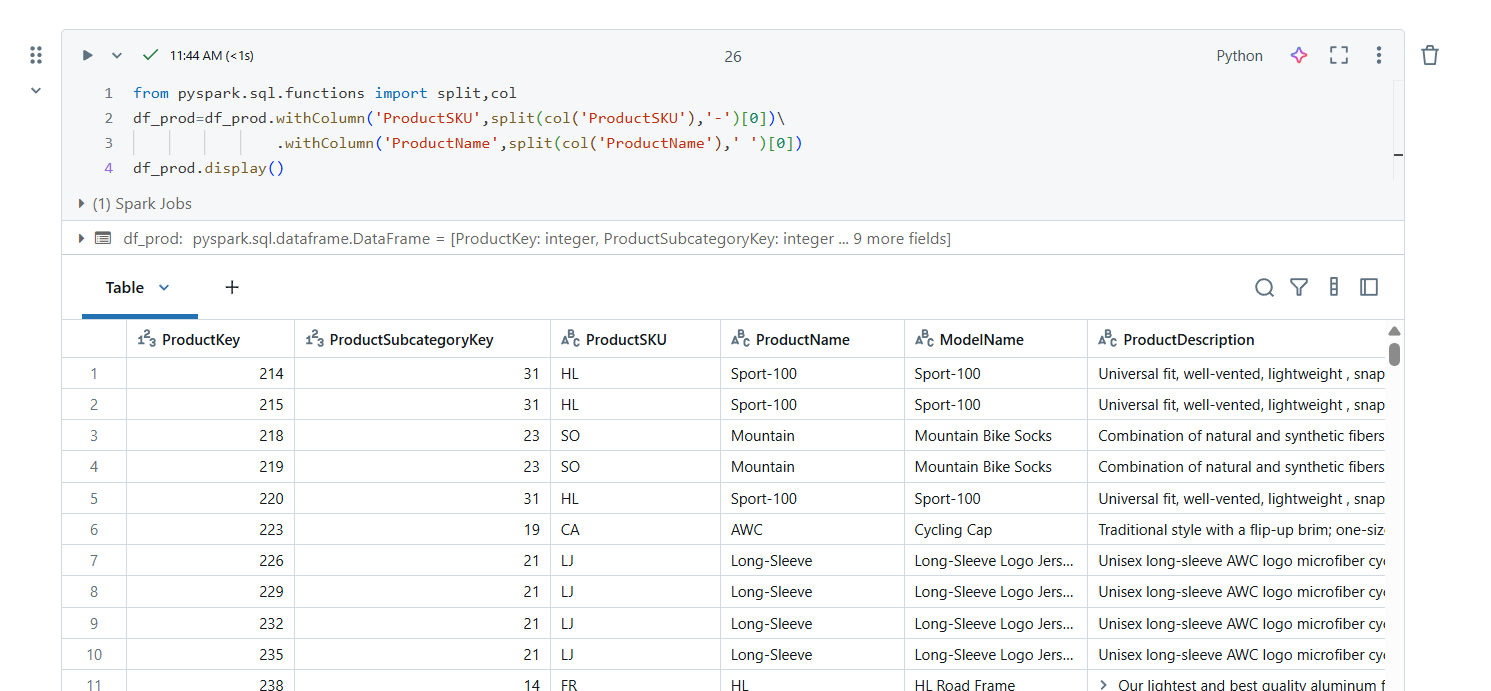
Day,month functions



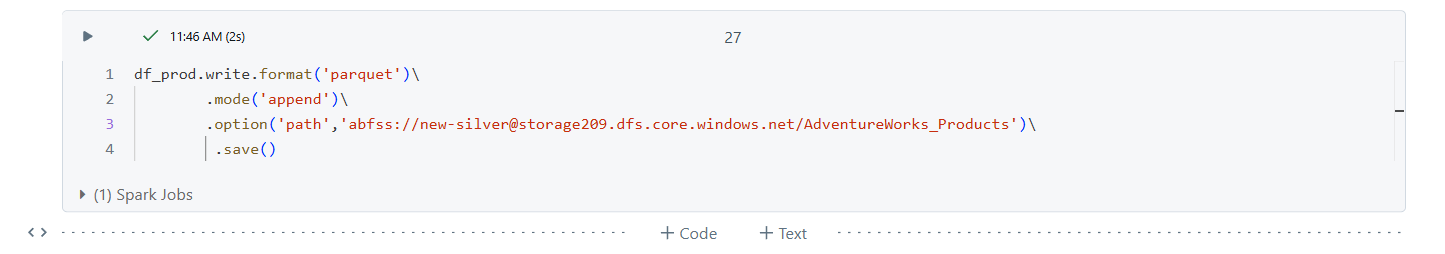
Concat,lit functions



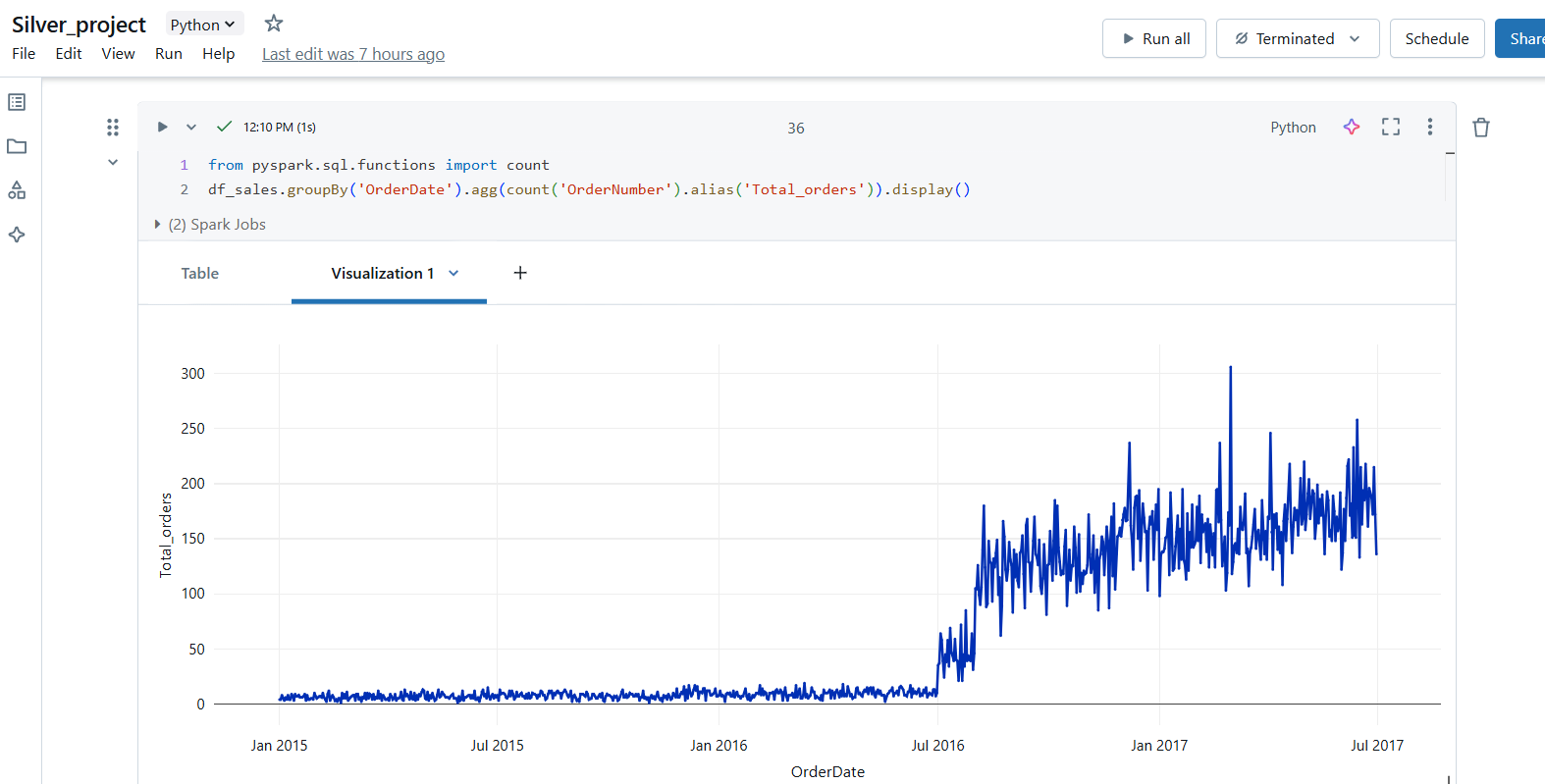
Split function

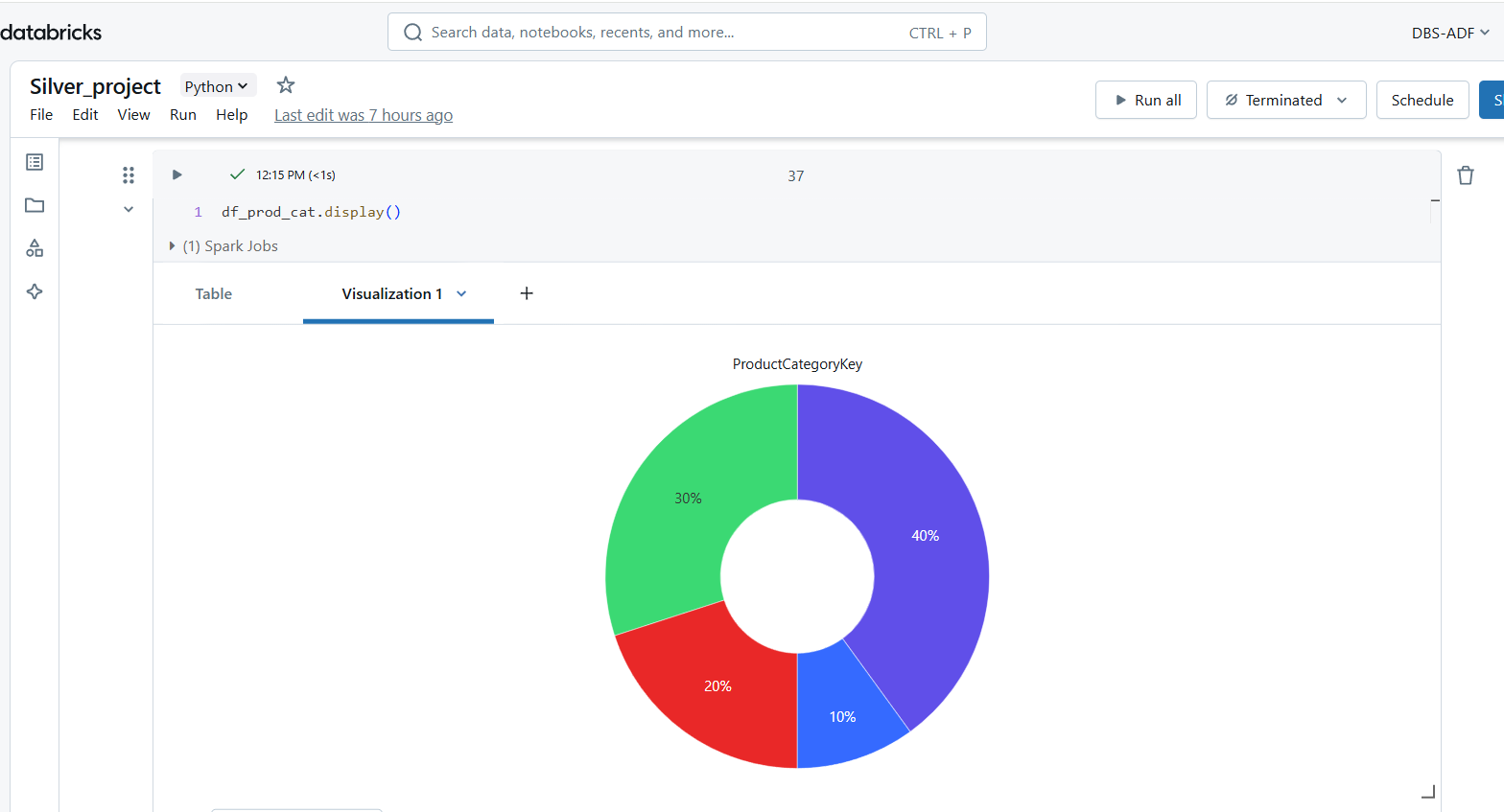


Write The Data in silver Layer:

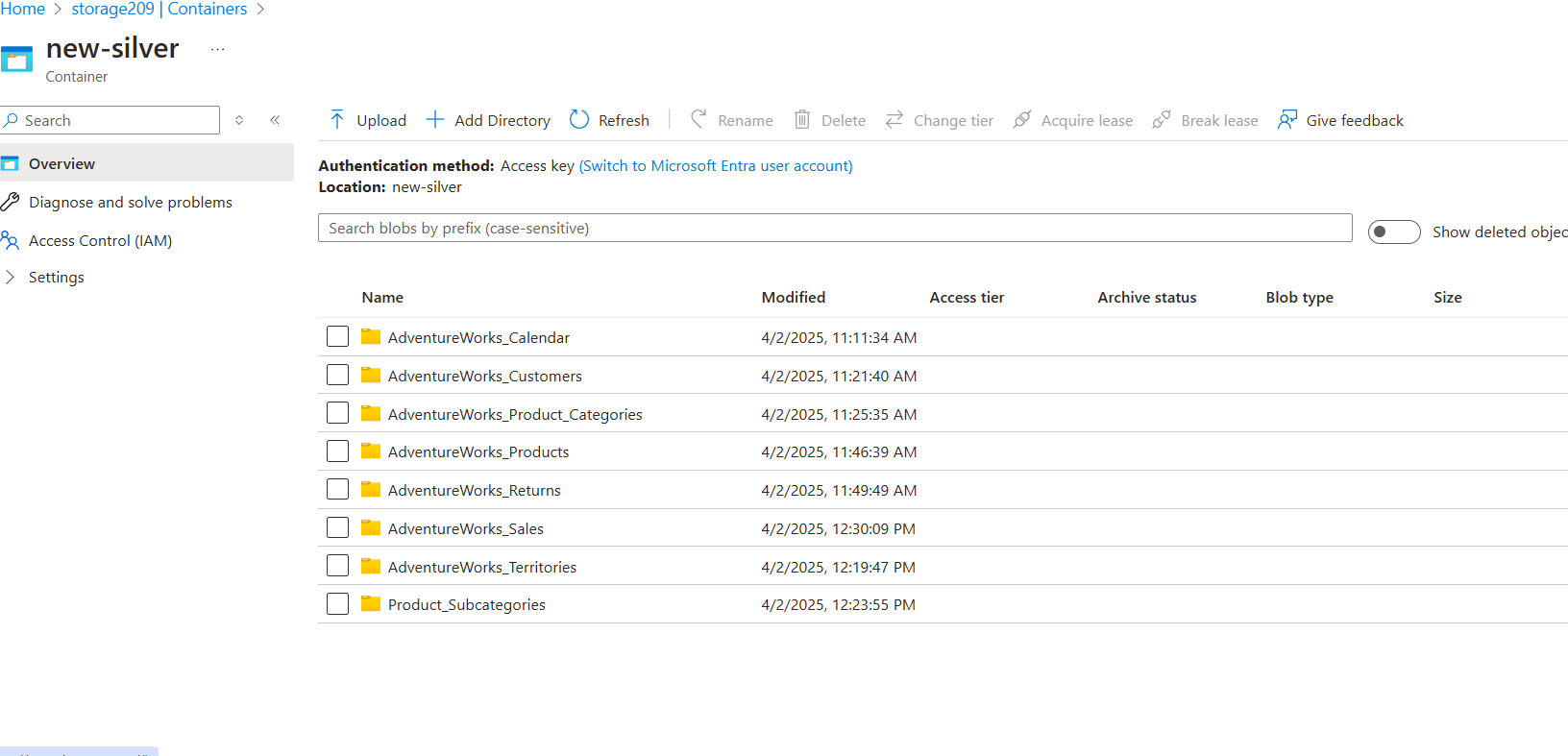


Some visualization in Silver Layer





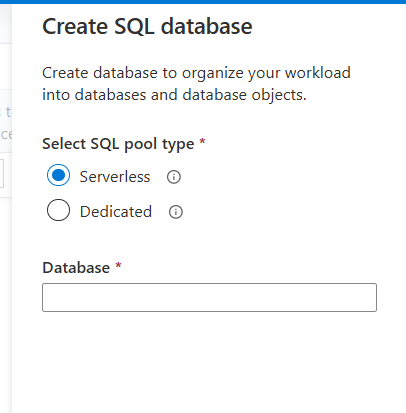
Final Data In Silver Layer:



**Azure Synapse Analytics**

Managed data warehousing.

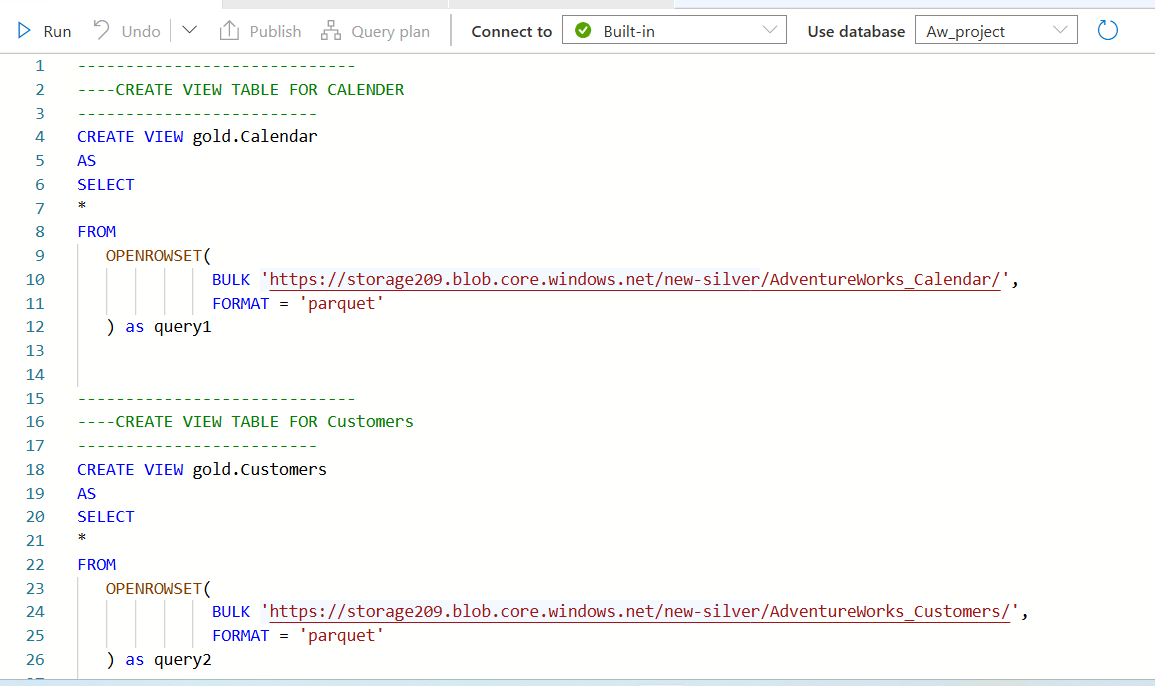
Create serverless Database



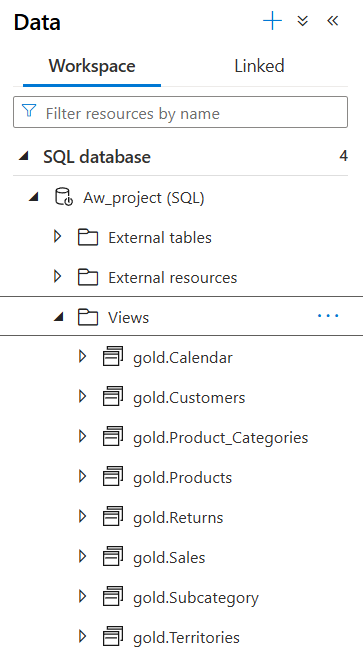
Create Schema:

Create schema <gold>

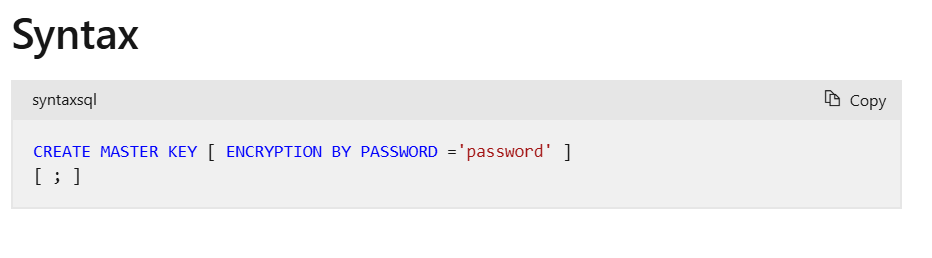
Creating View Tables



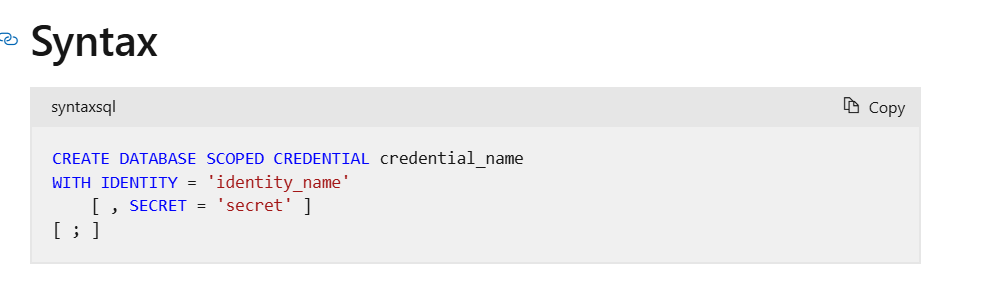
View Tables:



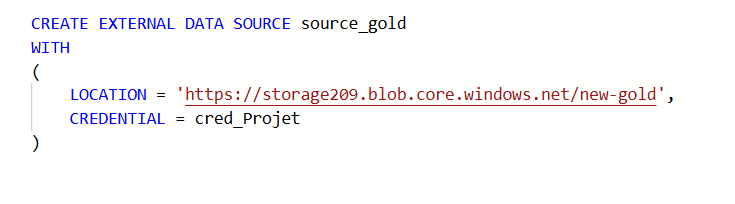
Create DataBase Master Key :



Create Database Credentials :

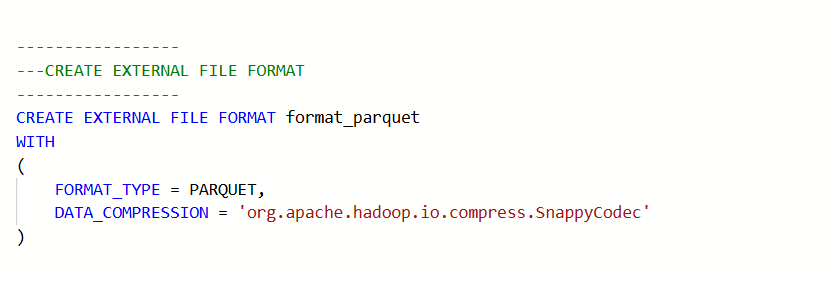


Create External Data source:

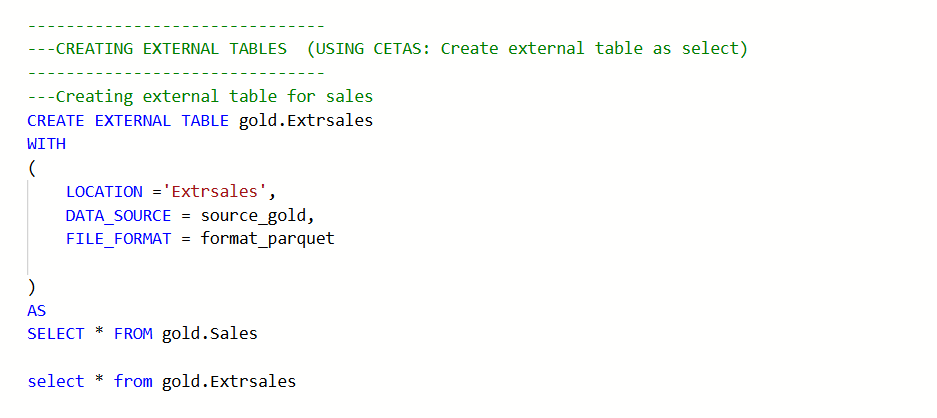


Create External File Format:

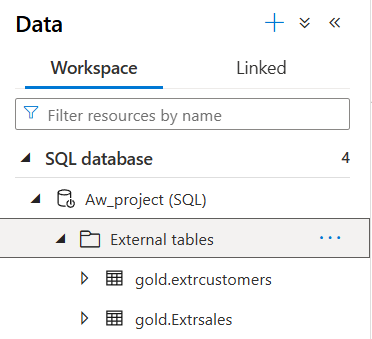
(parquet file format syntax)



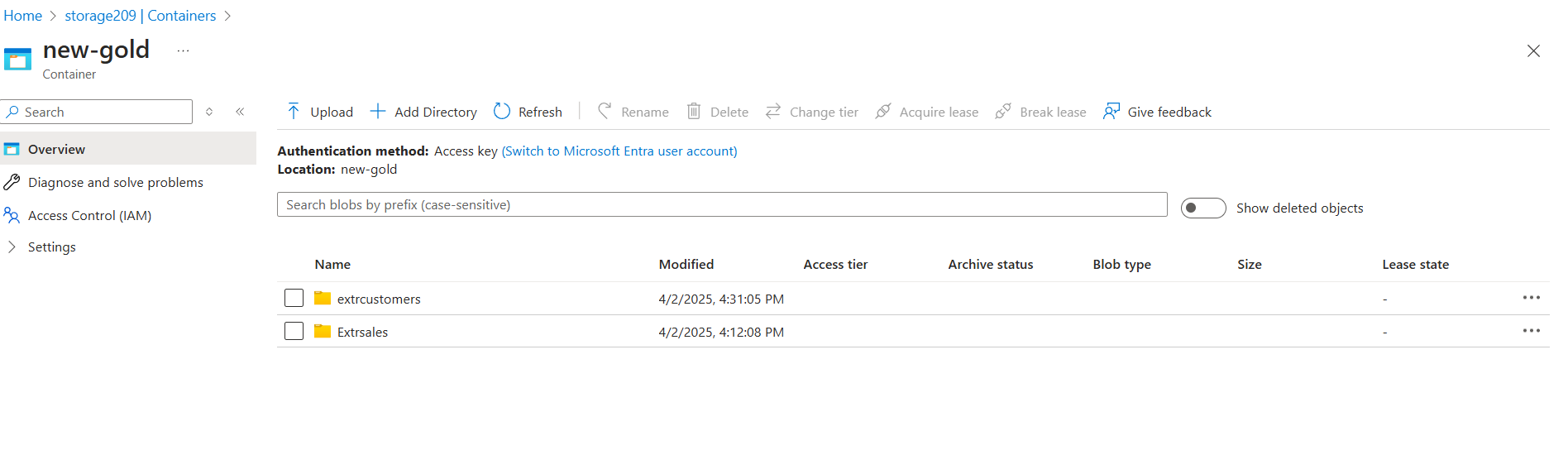
Create External Tables Using CETAS



External Tables:

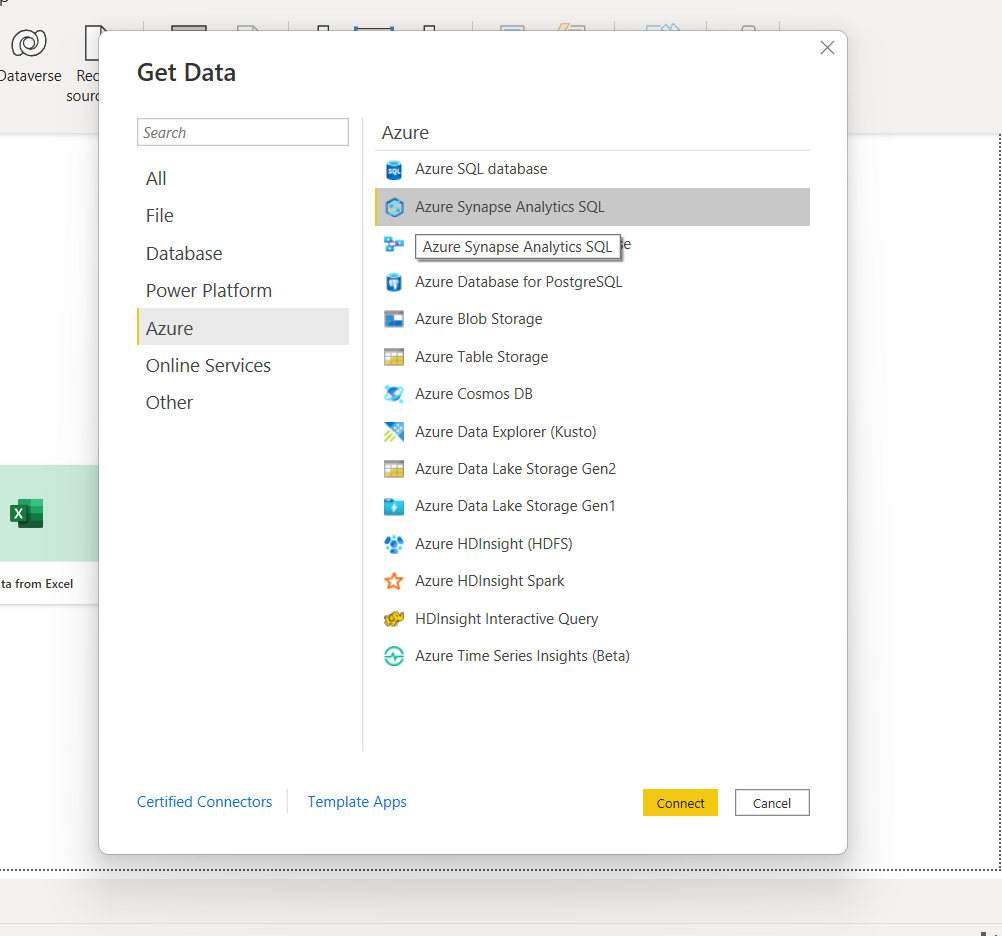


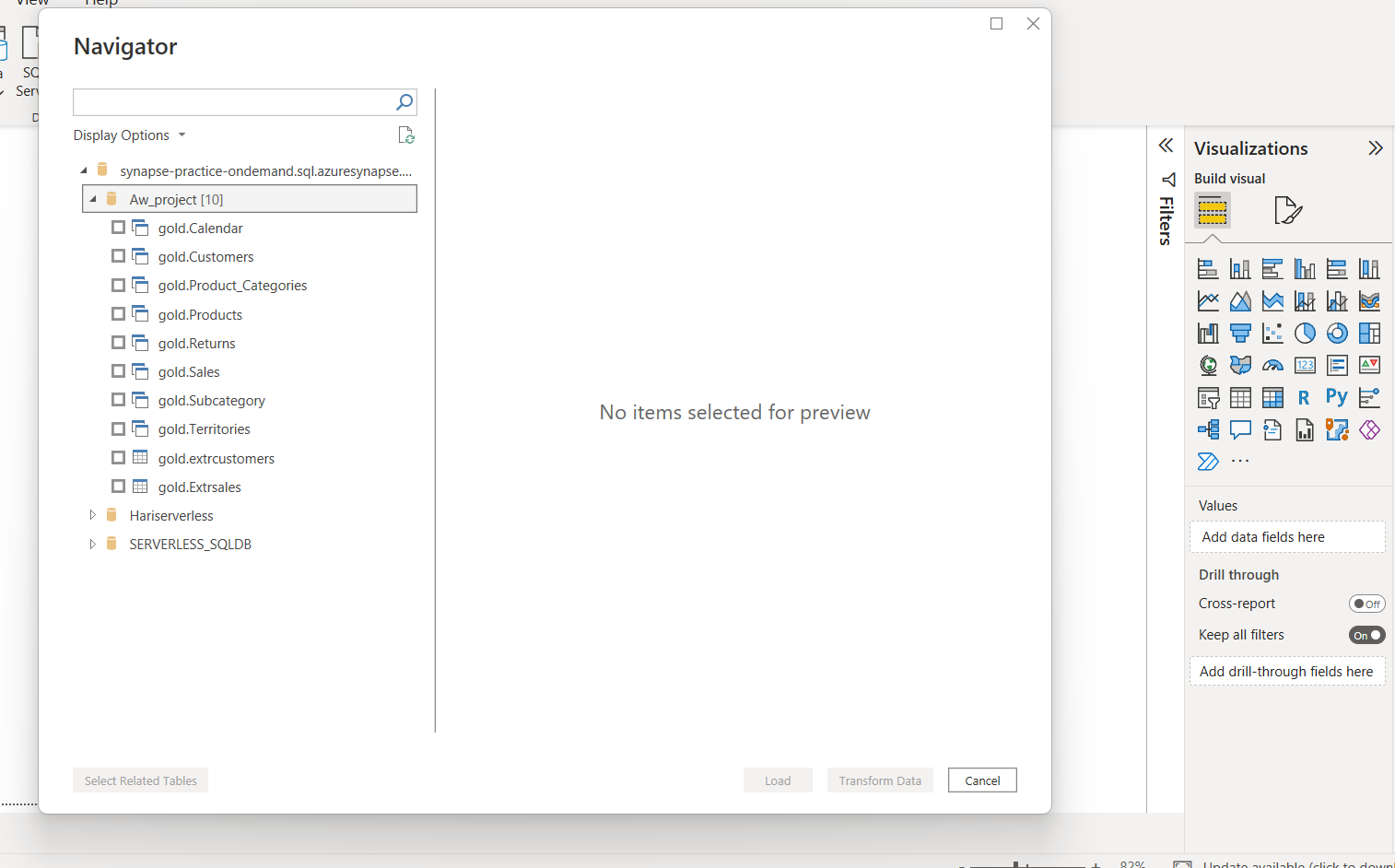
Data In Gold Layer:



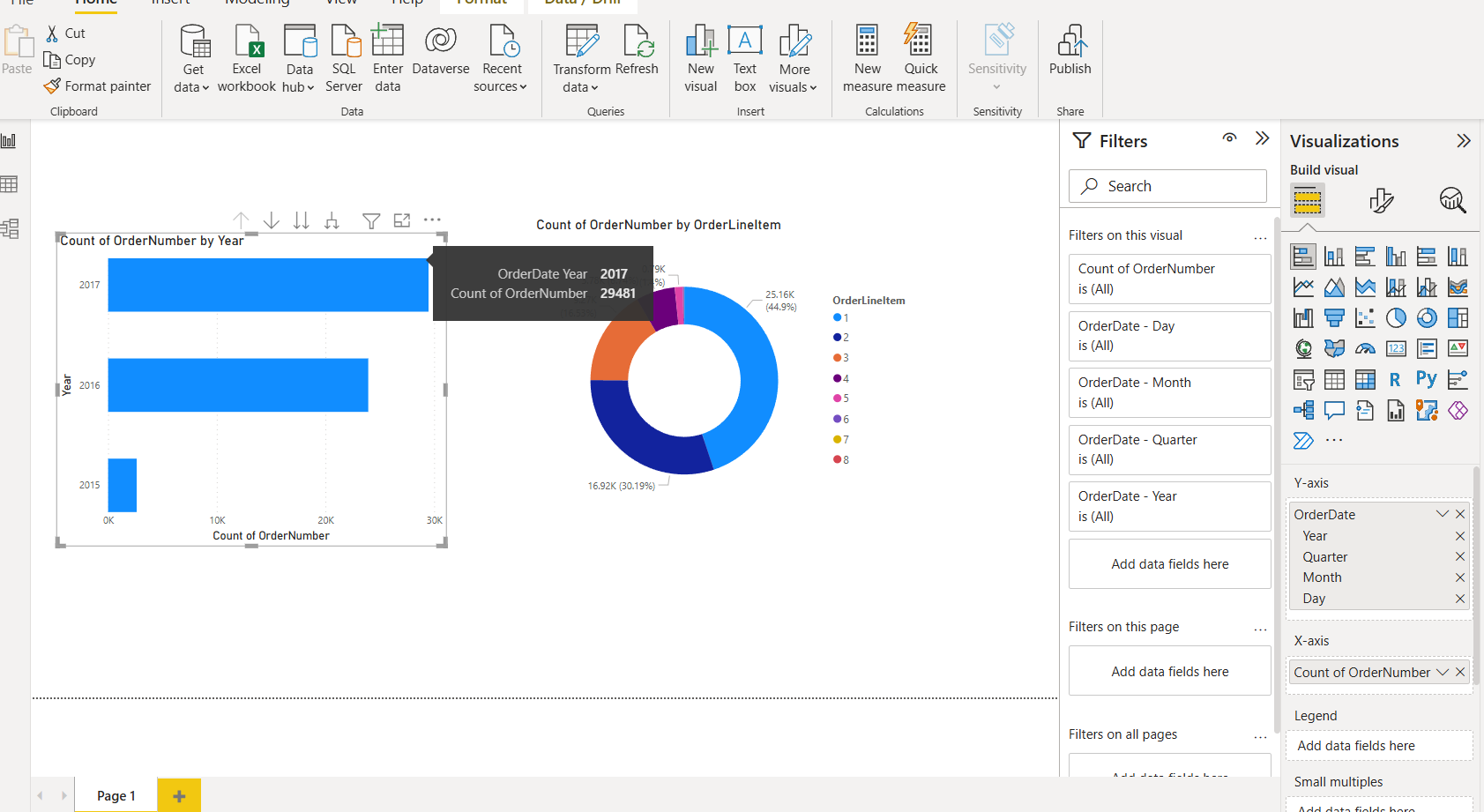
Data Visualization :

Connect to Power BI





Reports



END OF THE REPORT