**Azure Data Pipeline – End-to-End Flow**

**1. Source**

* Data is collected from an **external source (e.g., NYC Taxi data)** via **API**.

**2. Ingestion Layer (Azure Data Factory)**

* **Azure Data Factory** fetches the source data using linked services and datasets.
* Data is ingested and stored in **Data Lake Gen2** as **Parquet files** in the **Raw Data Store**.

**3. Transformation Layer (Azure Databricks)**

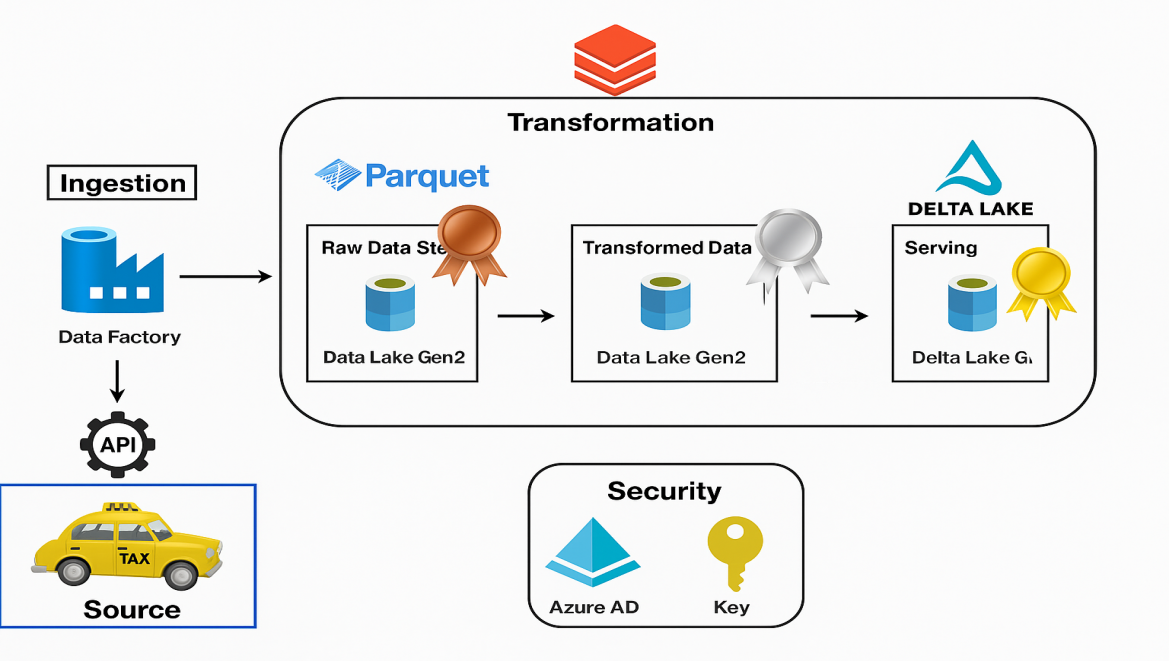
* Data is loaded from the raw zone using **Databricks notebooks**.
* It is cleaned, transformed, and written back into **Data Lake Gen2** under the **Transformed Data Zone**, again in **Parquet** format.

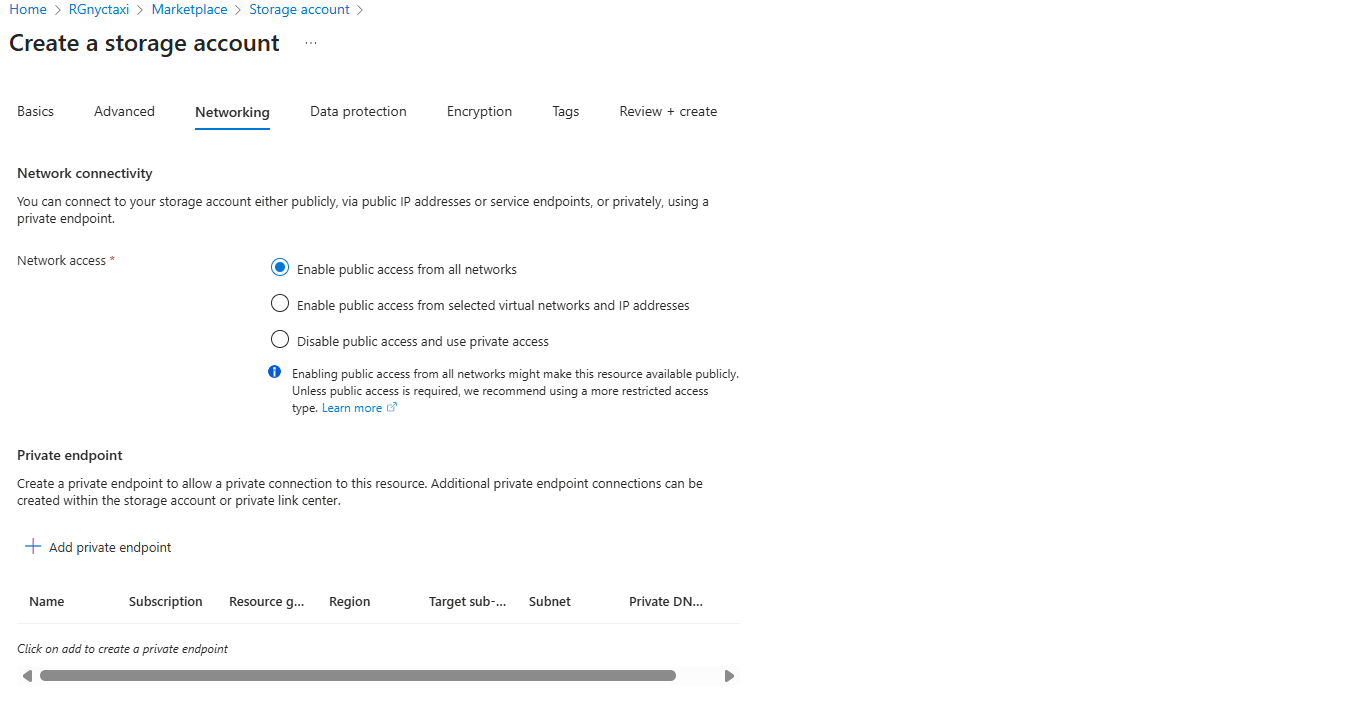
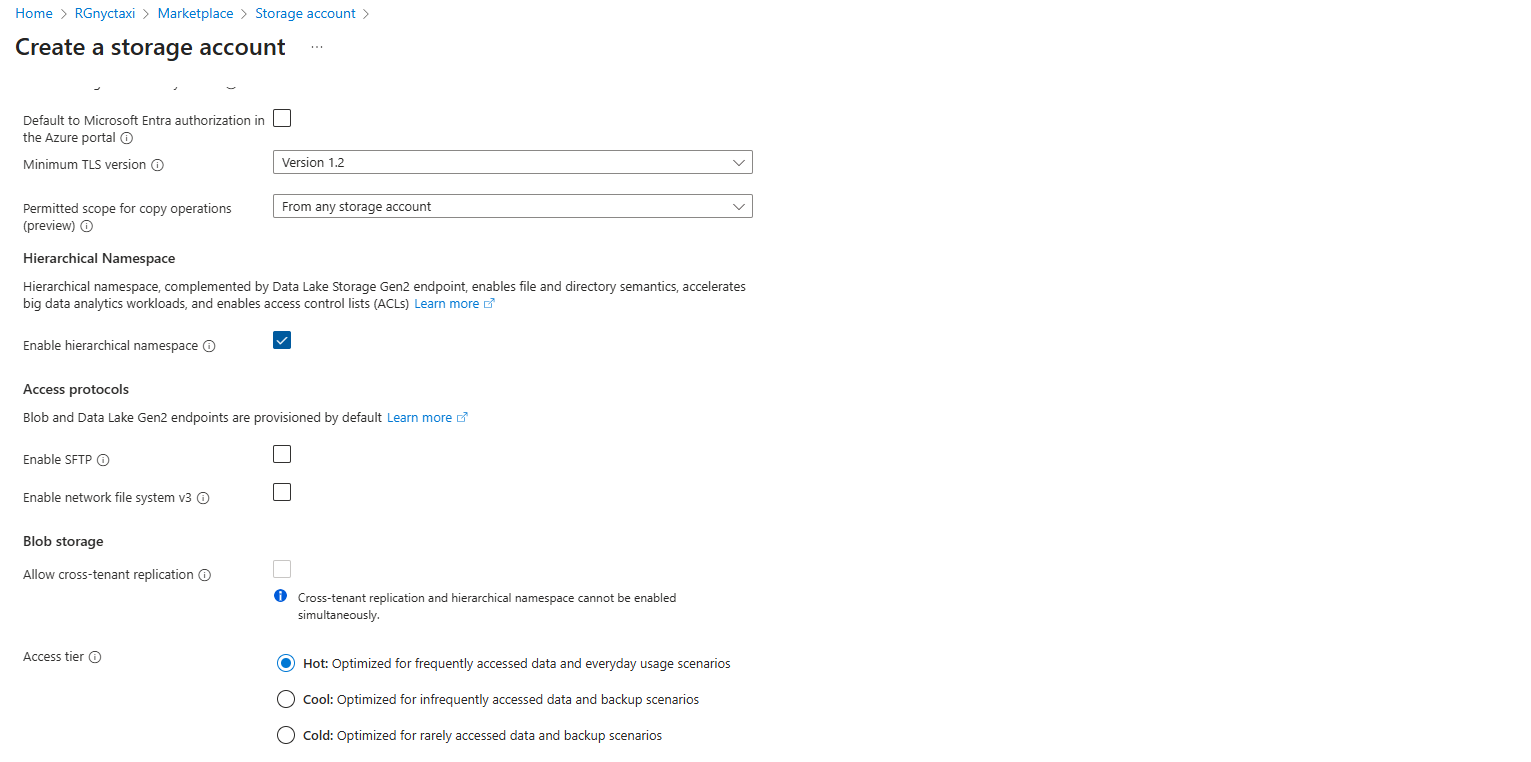
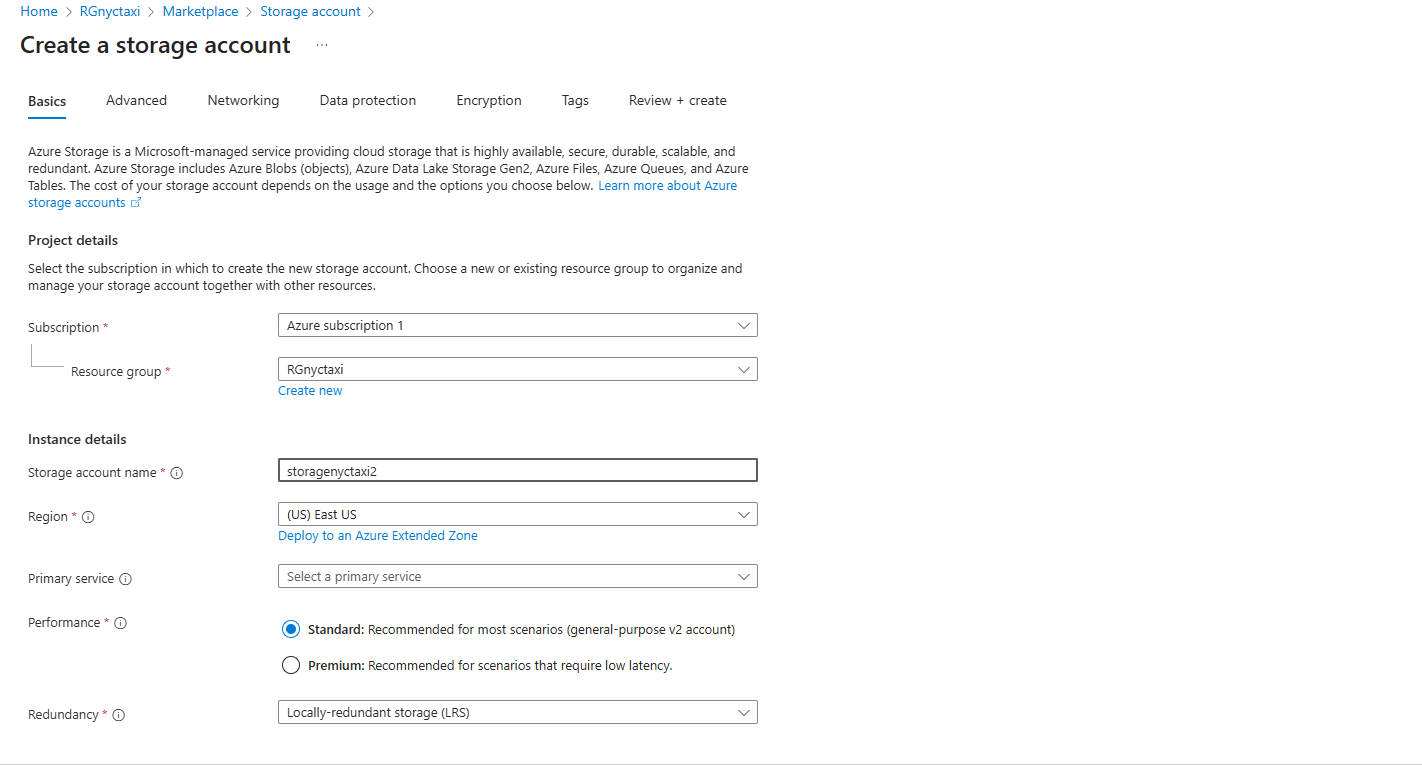
**4. Serving Layer**

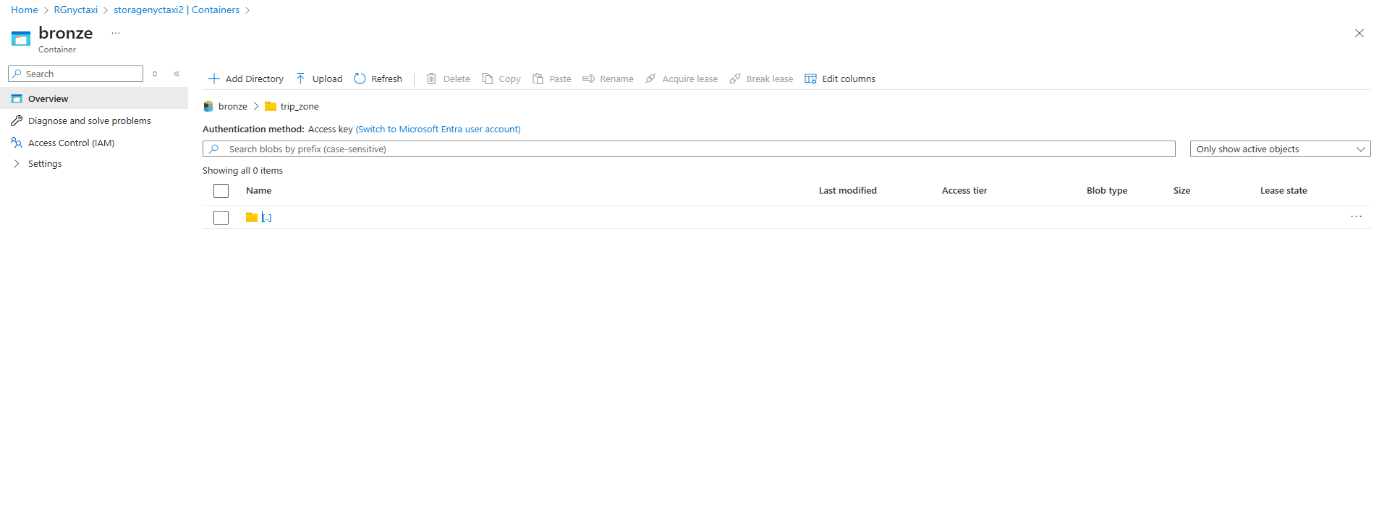
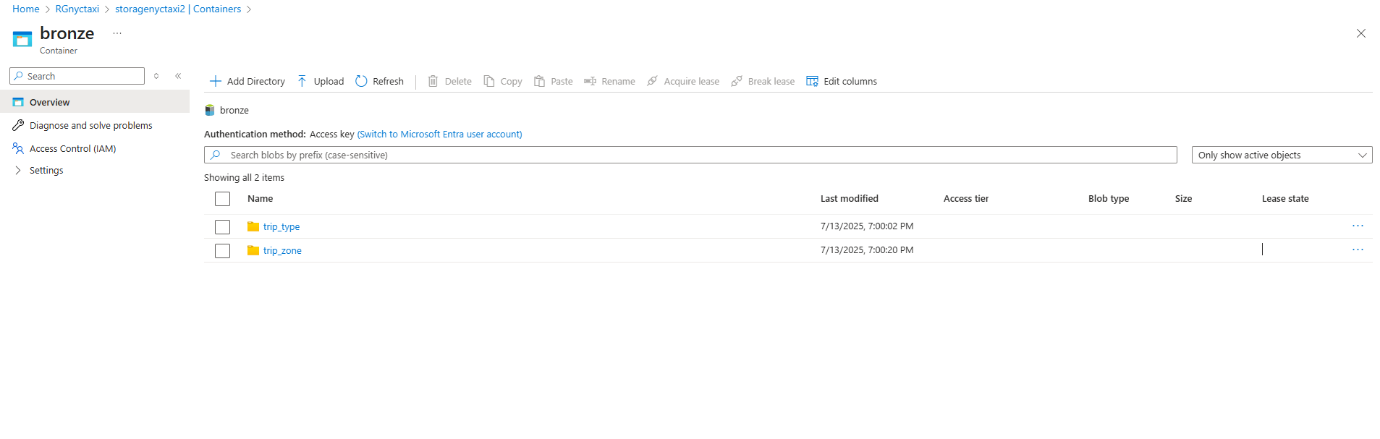
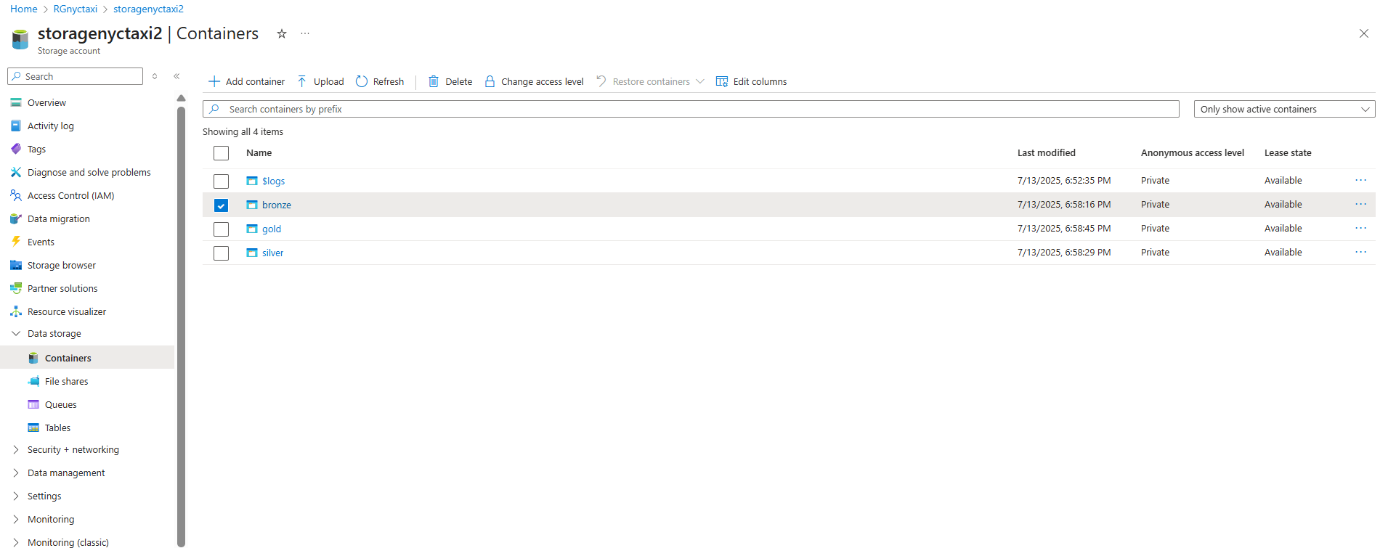
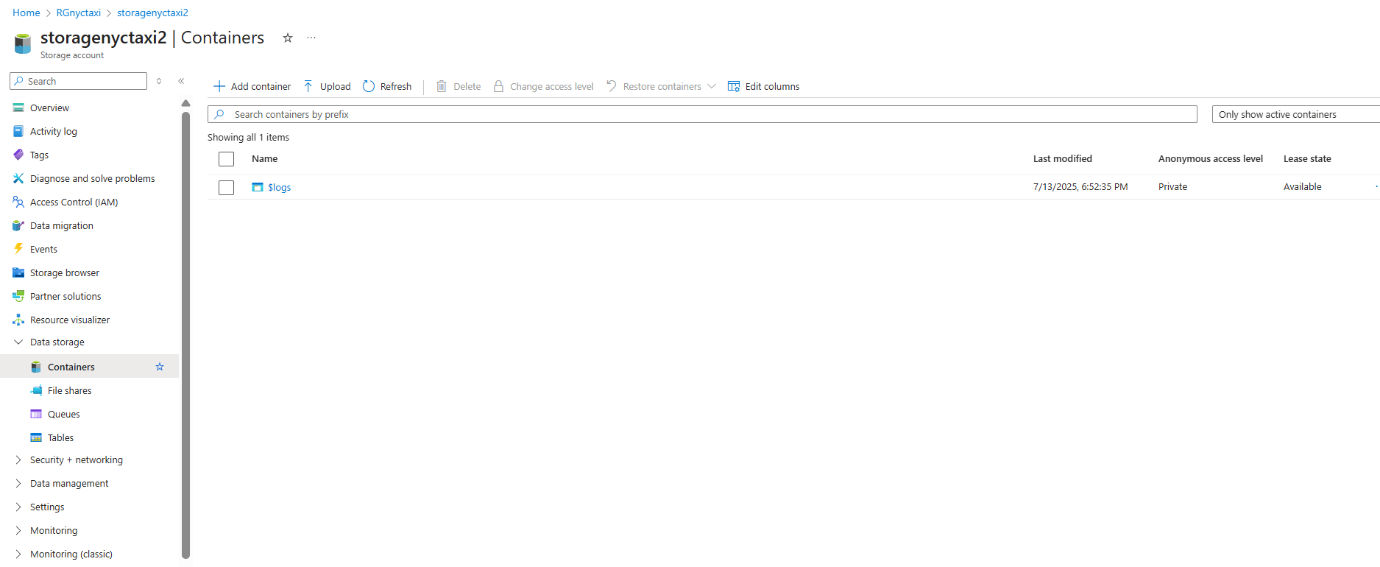
* Final processed data is saved in the **Serving Zone**, optimized for querying and analytics.
* Uses **Delta Lake** format for better performance, reliability, and ACID compliance.

**5. Security**

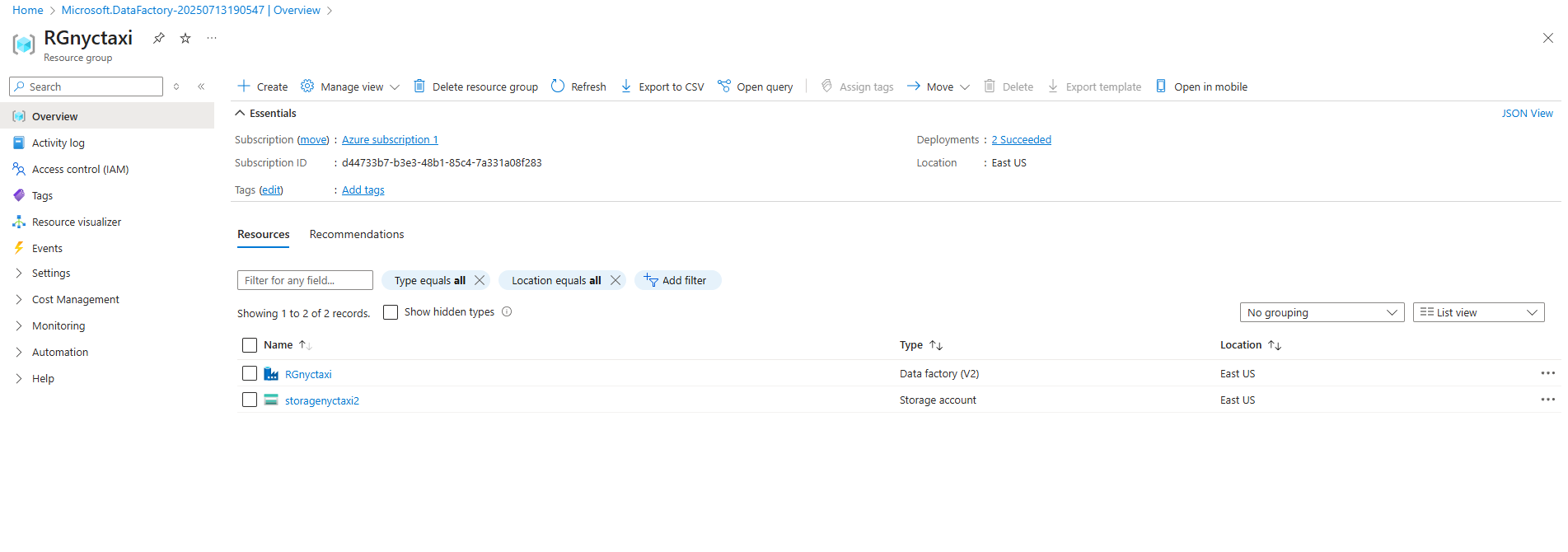
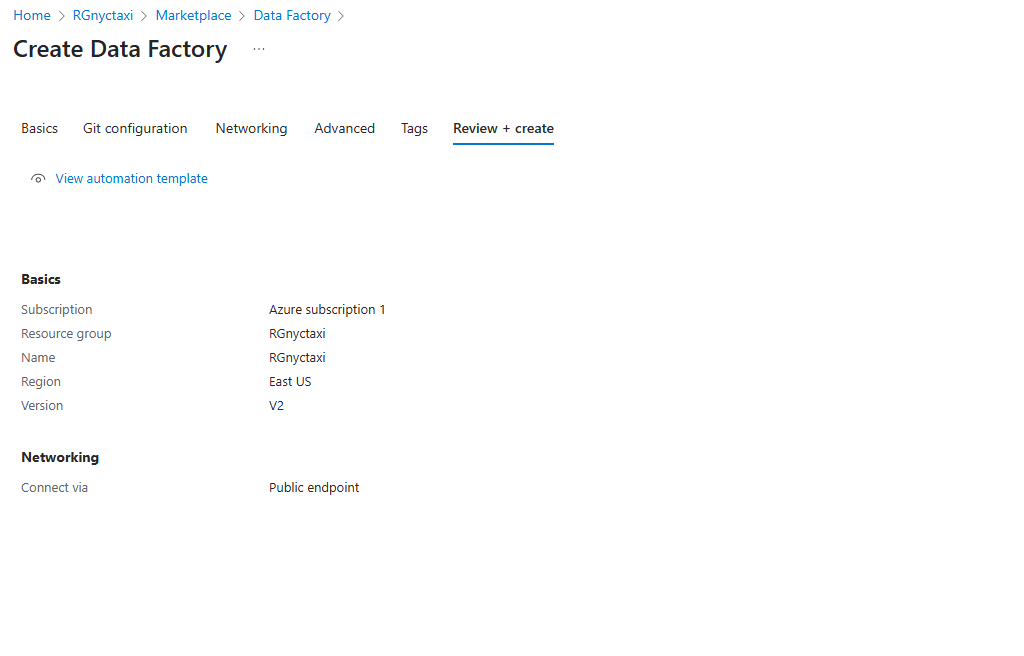
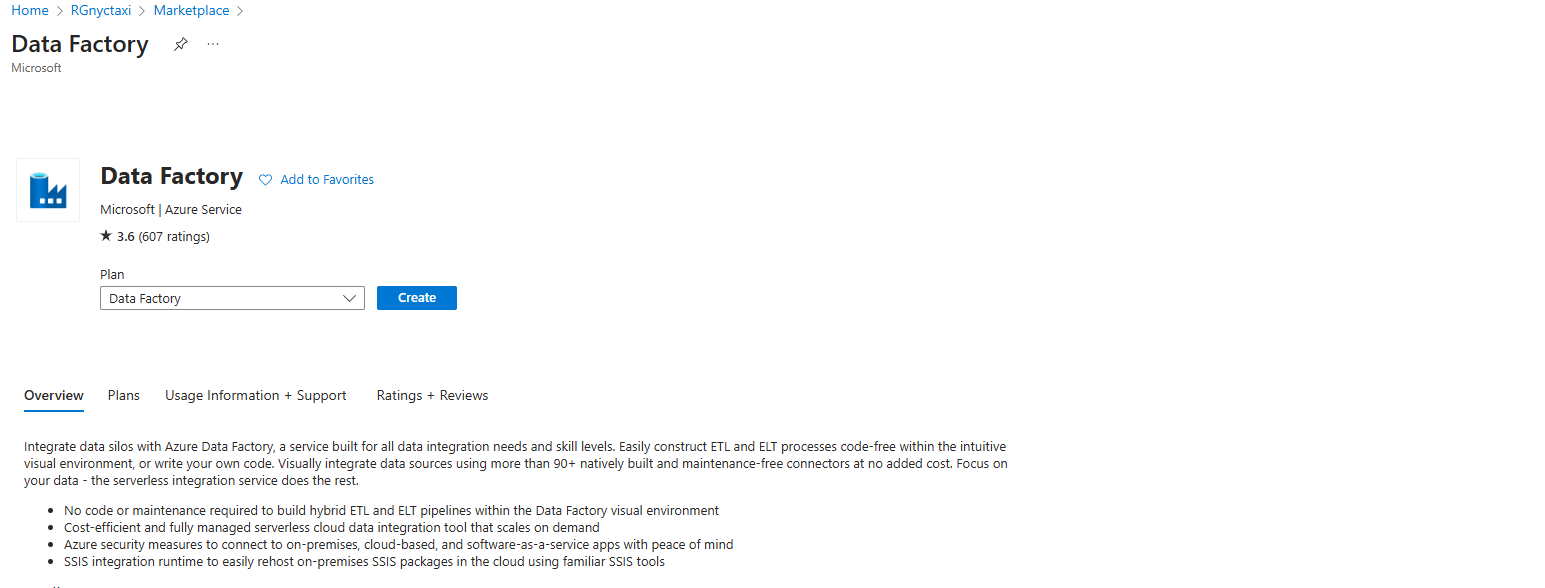
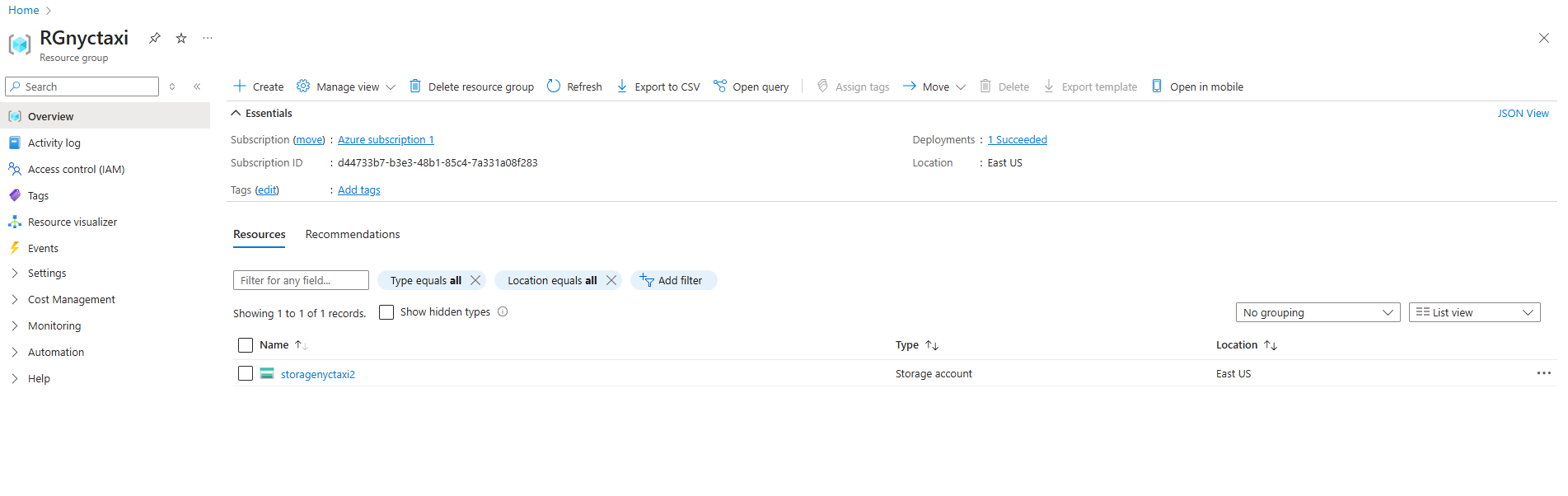
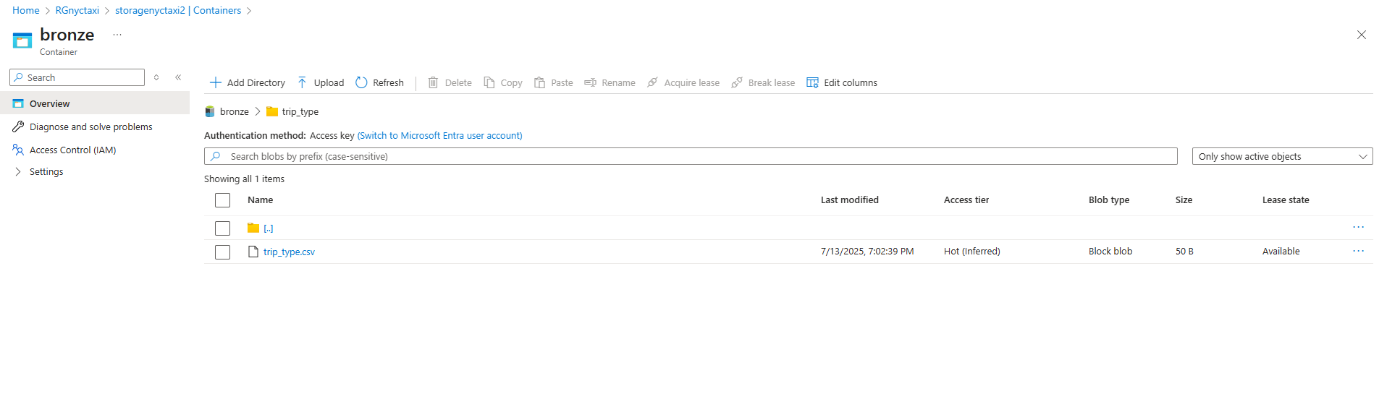
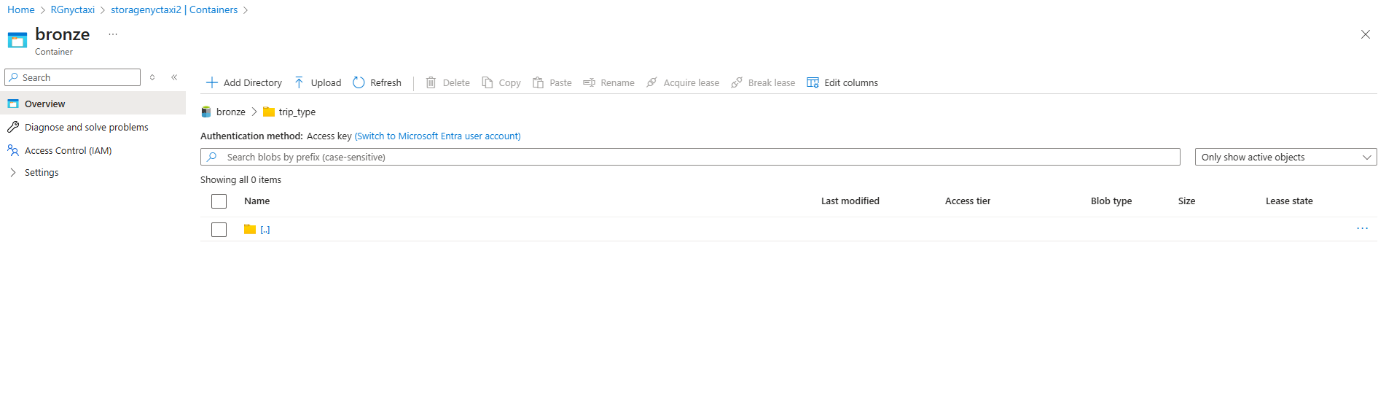
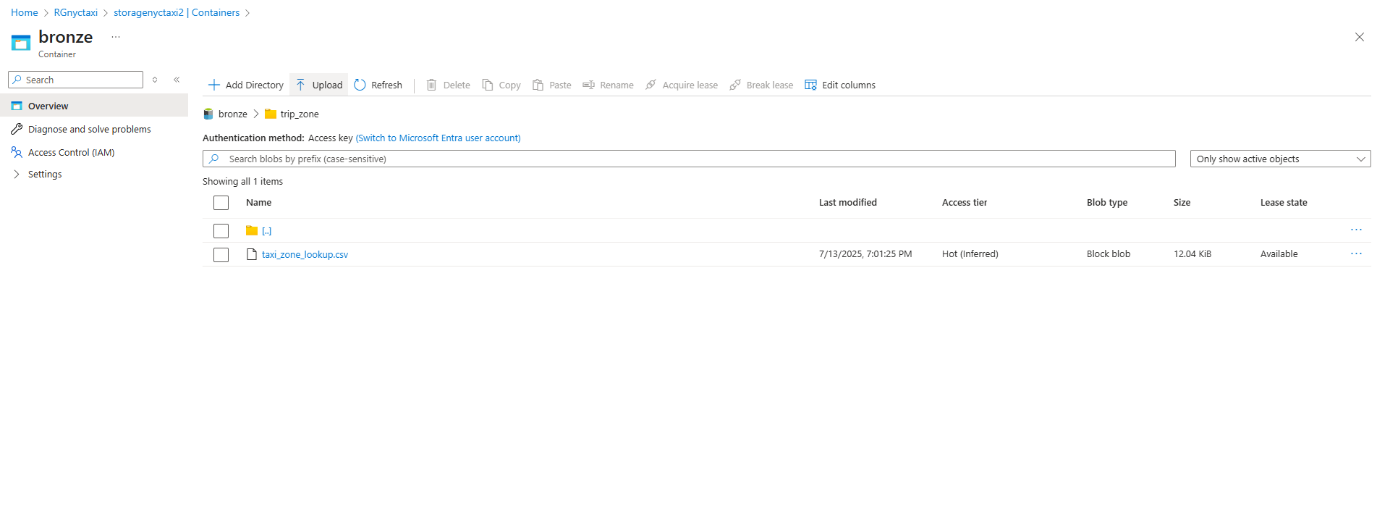
* Managed via **Azure AD** and **Key Vault**, ensuring secure access to services and secrets.



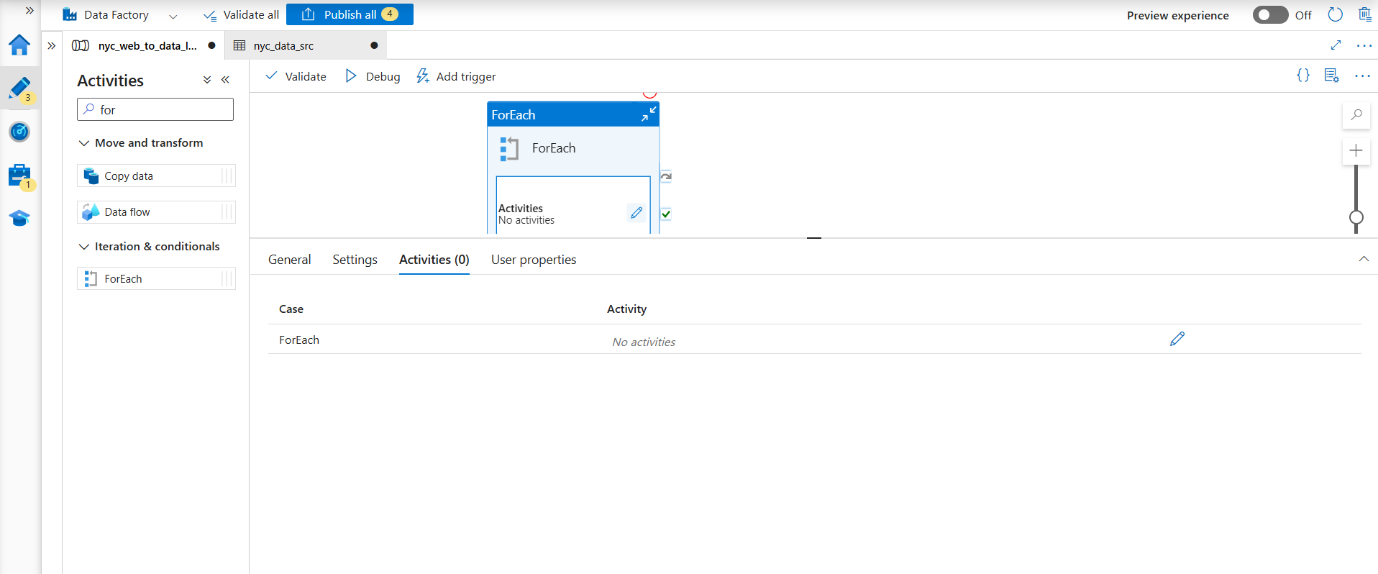
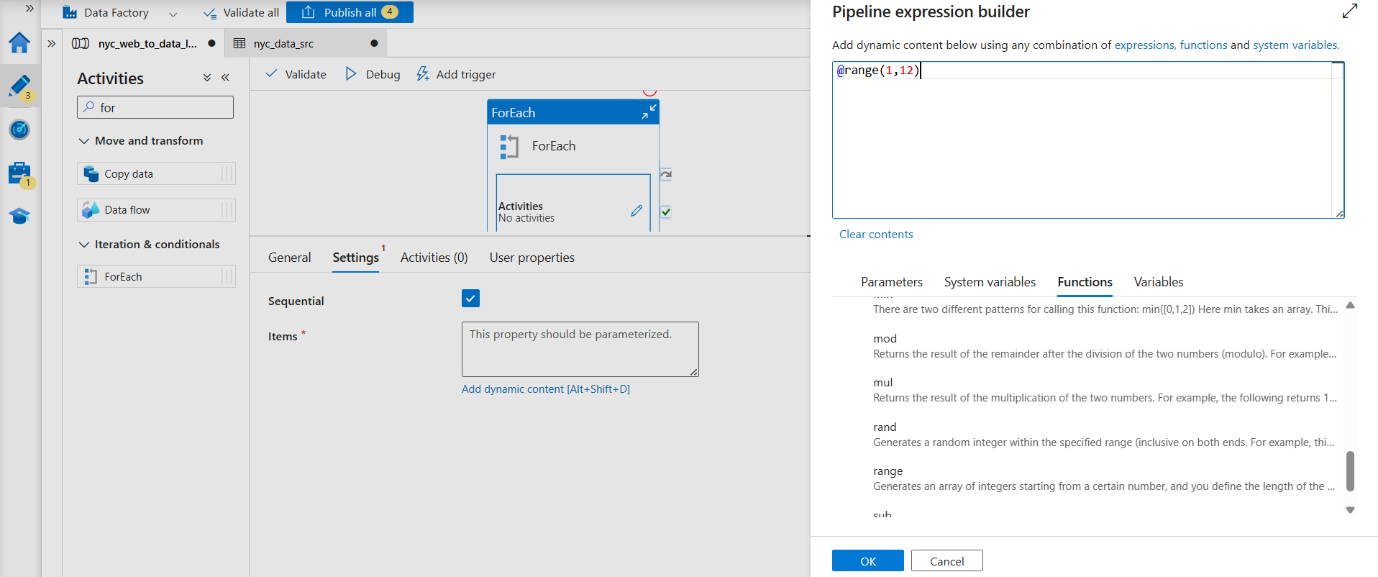
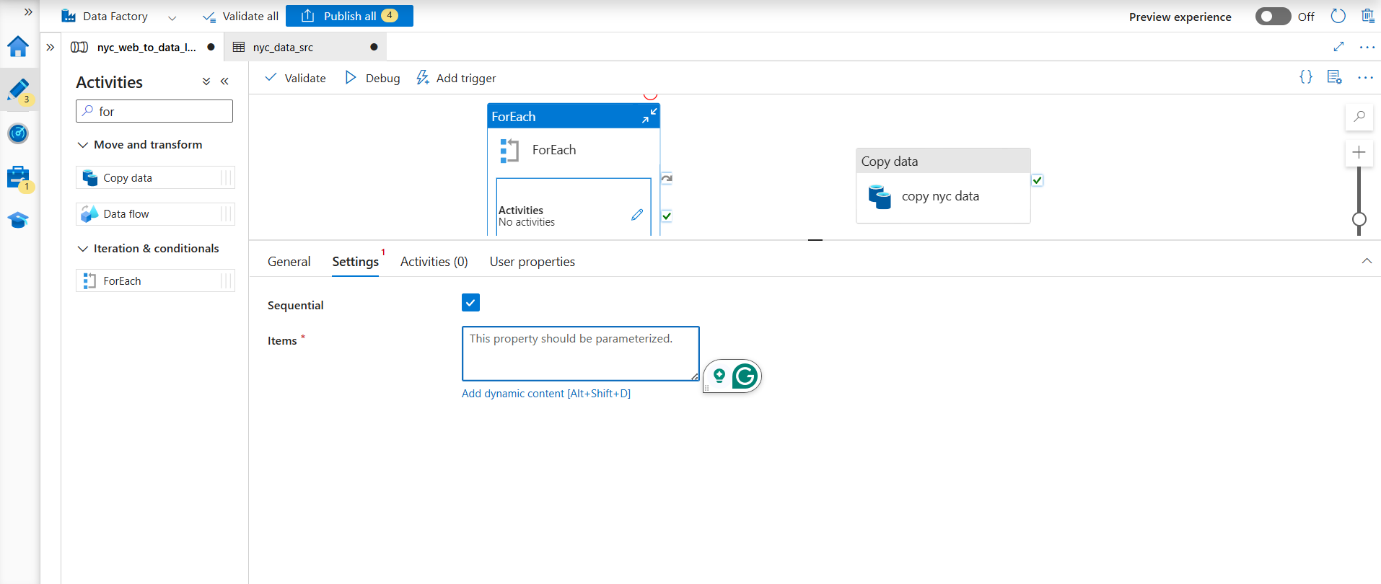
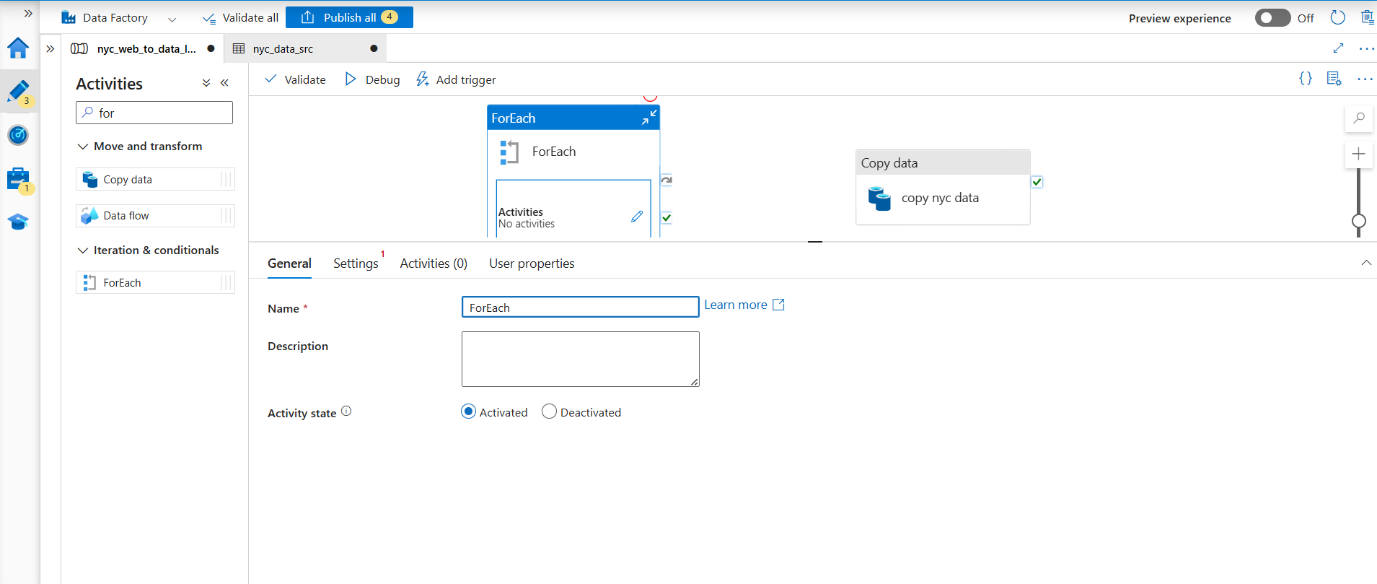
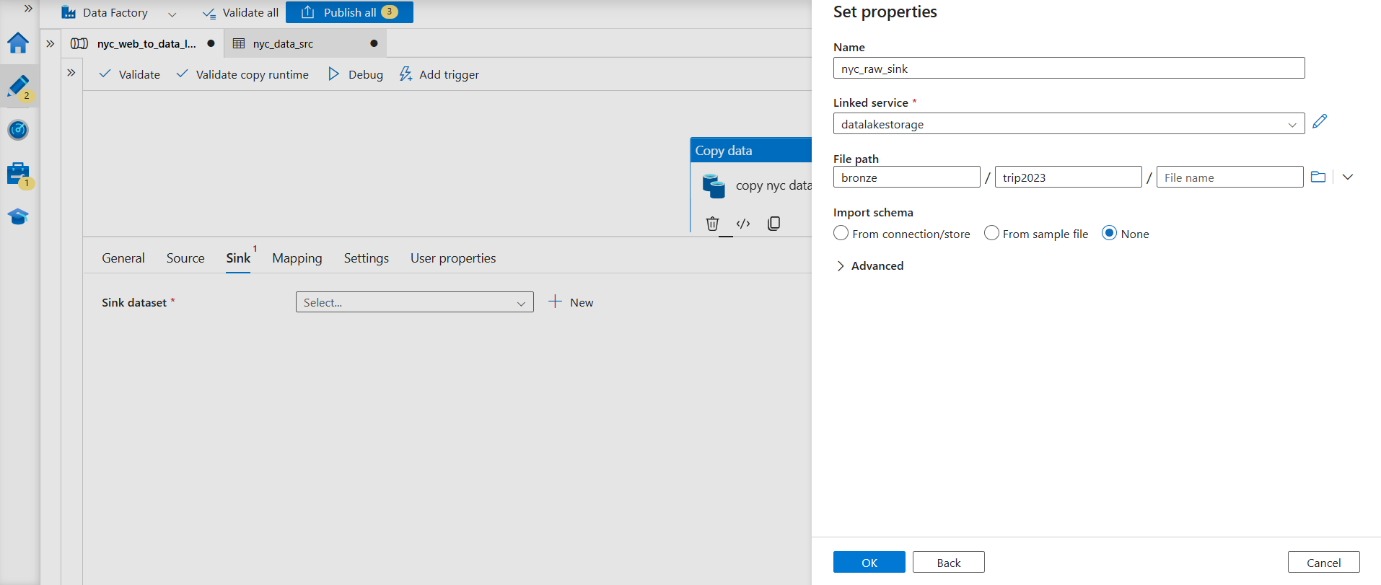
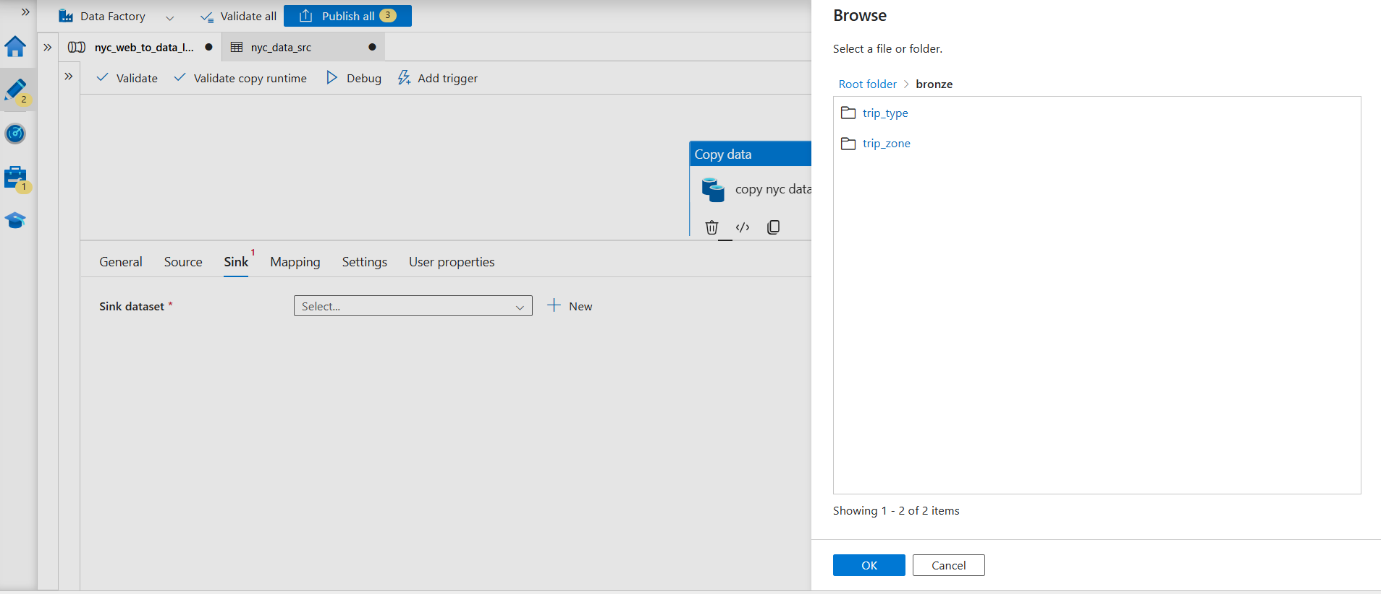
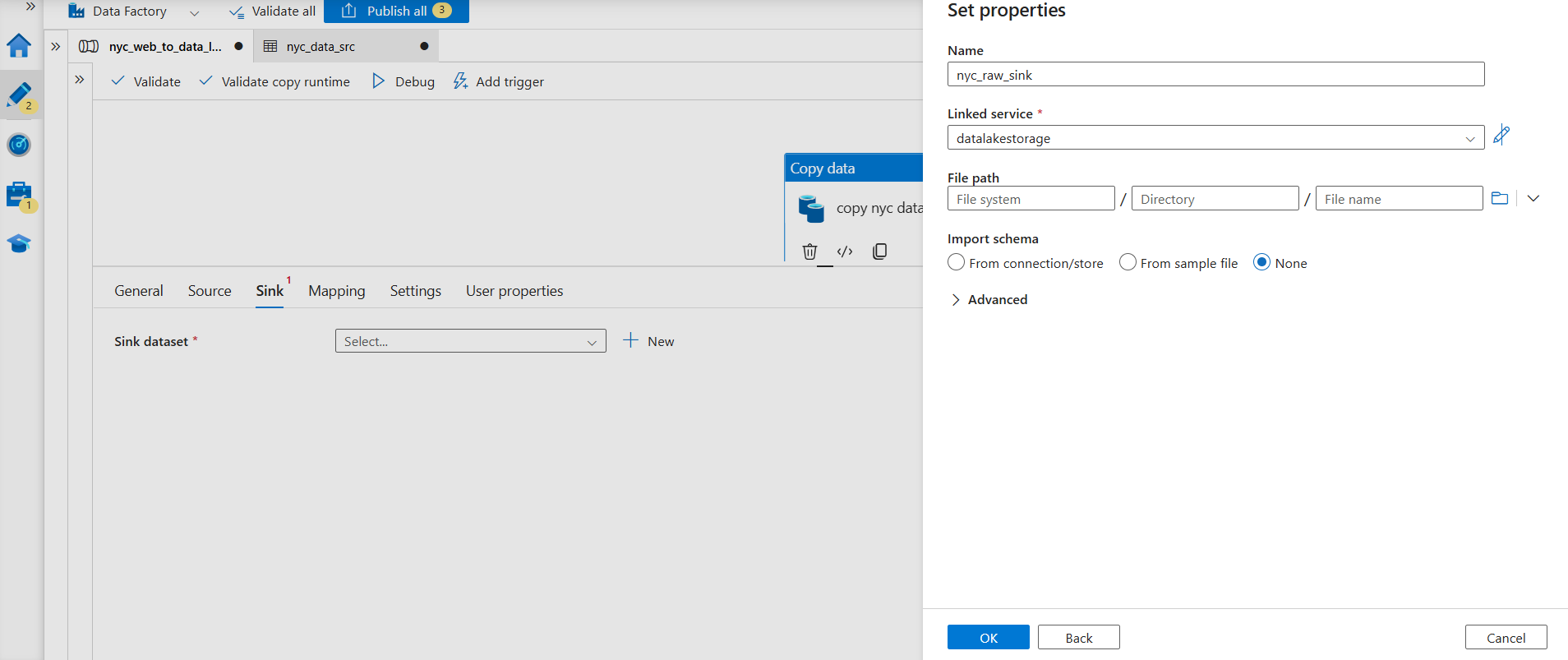
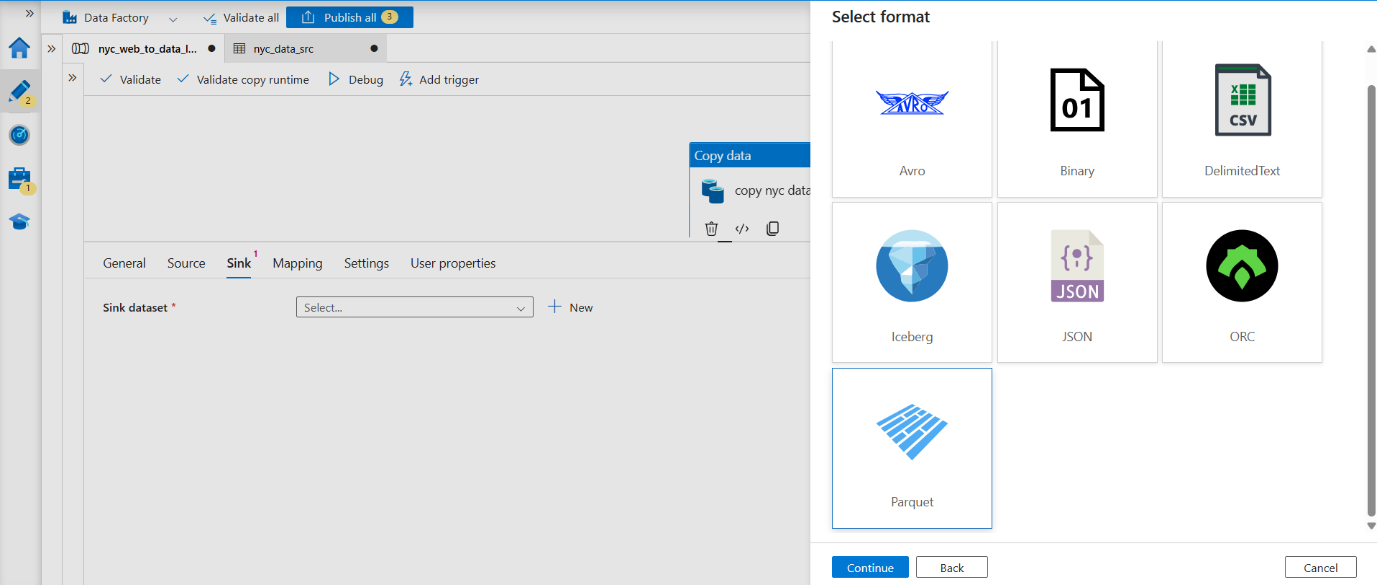
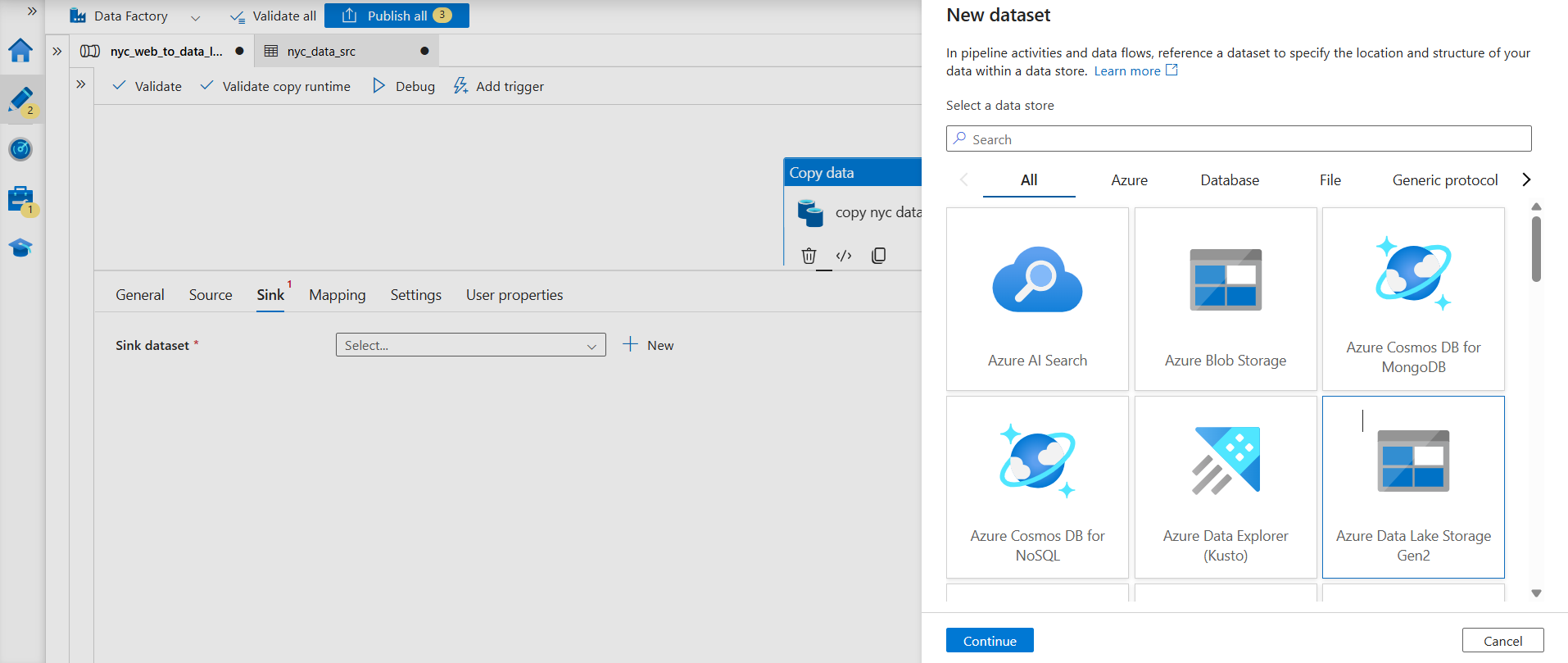
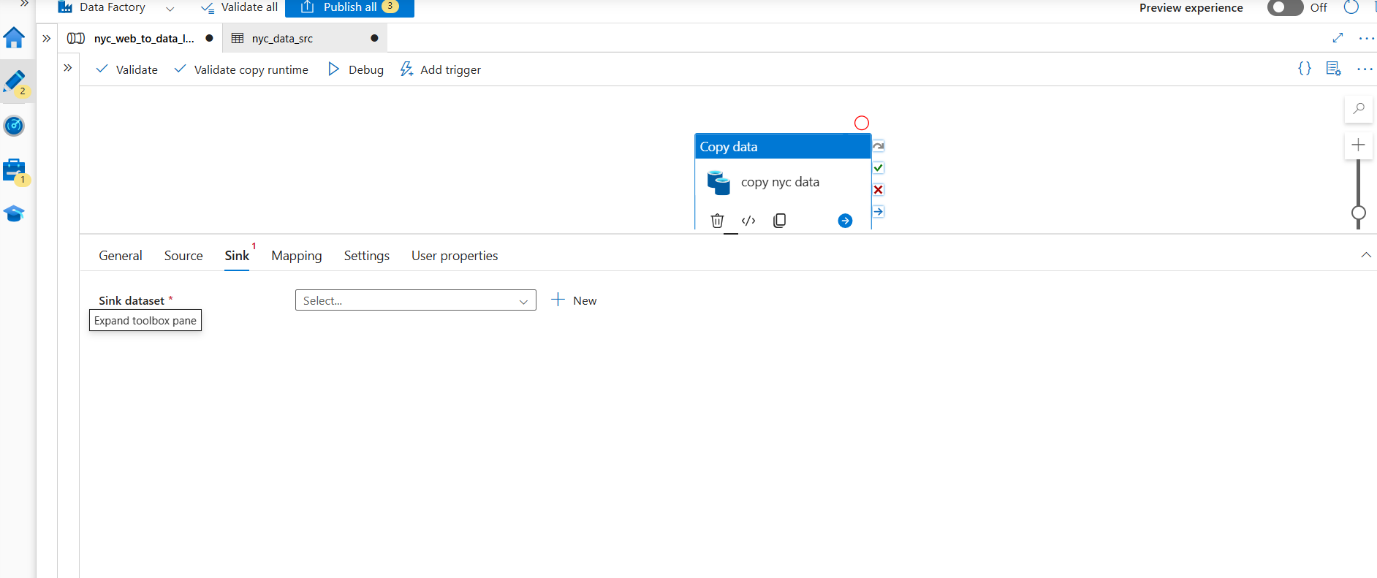
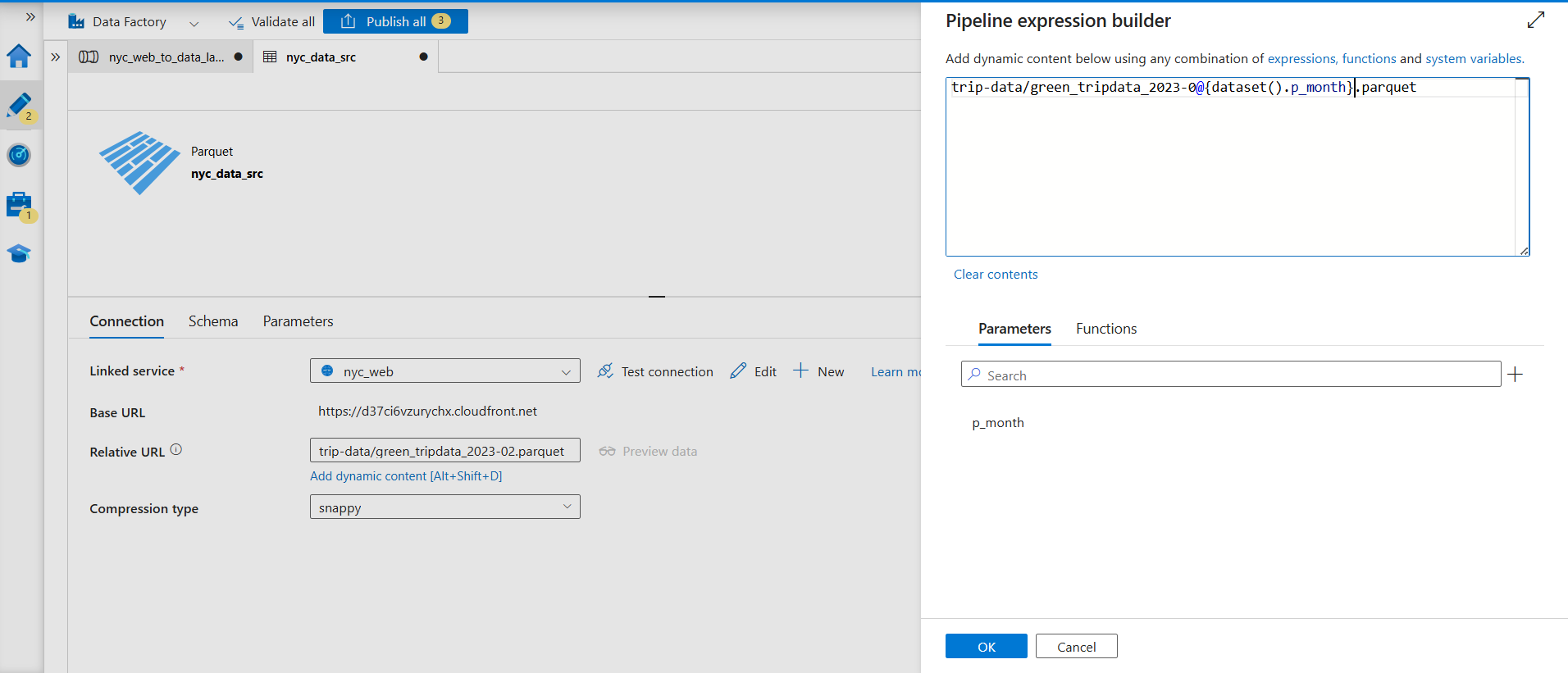
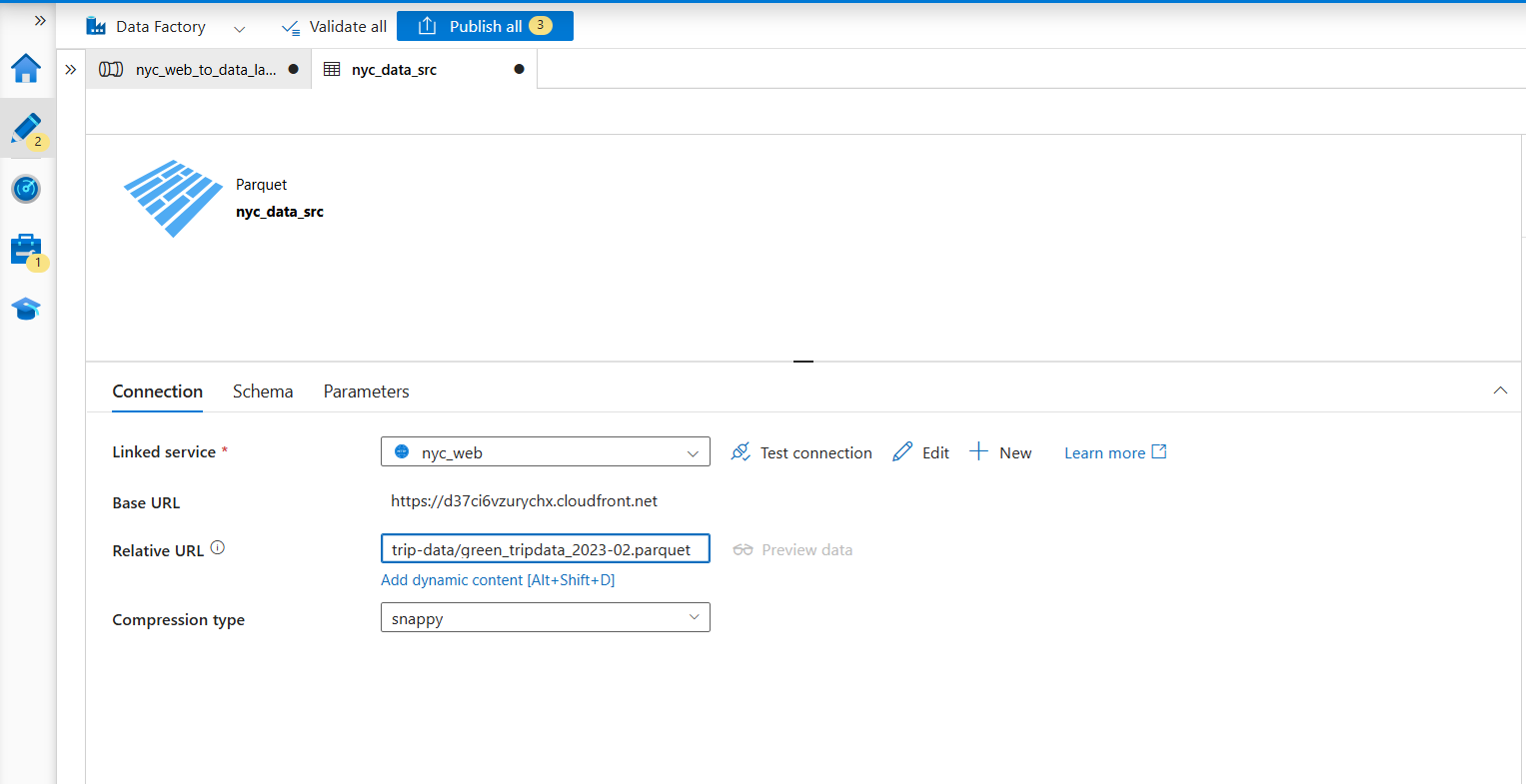
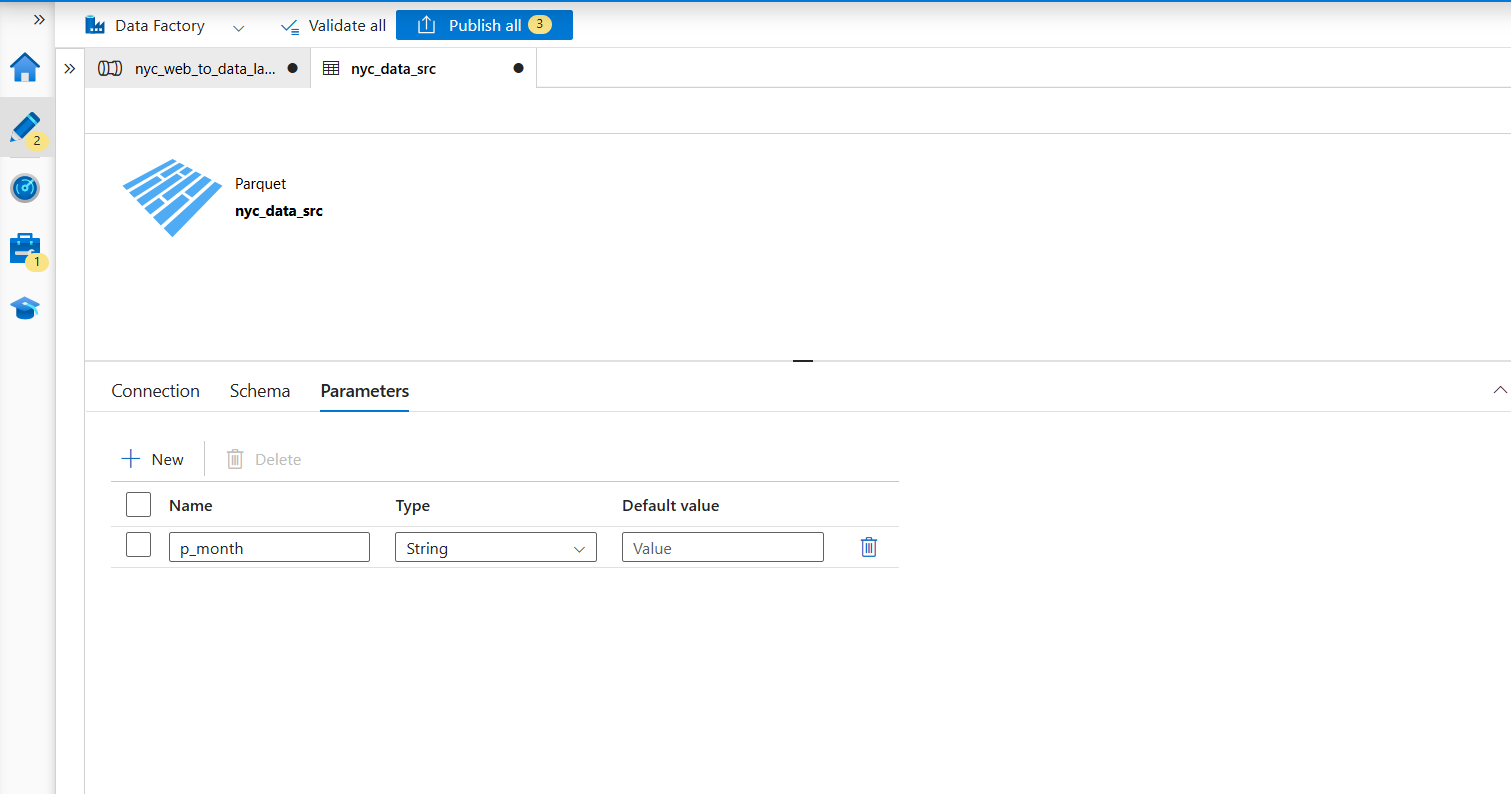
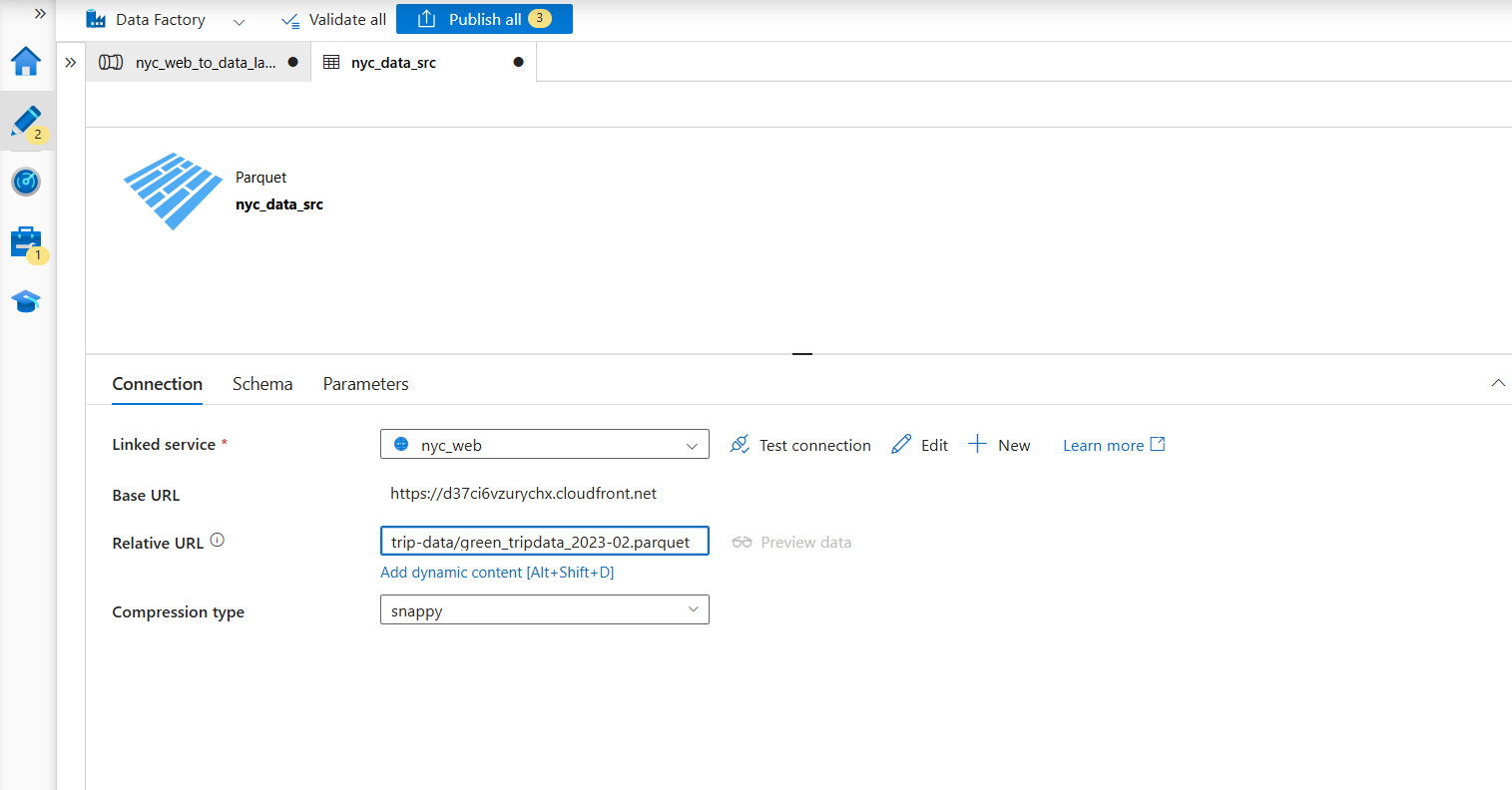
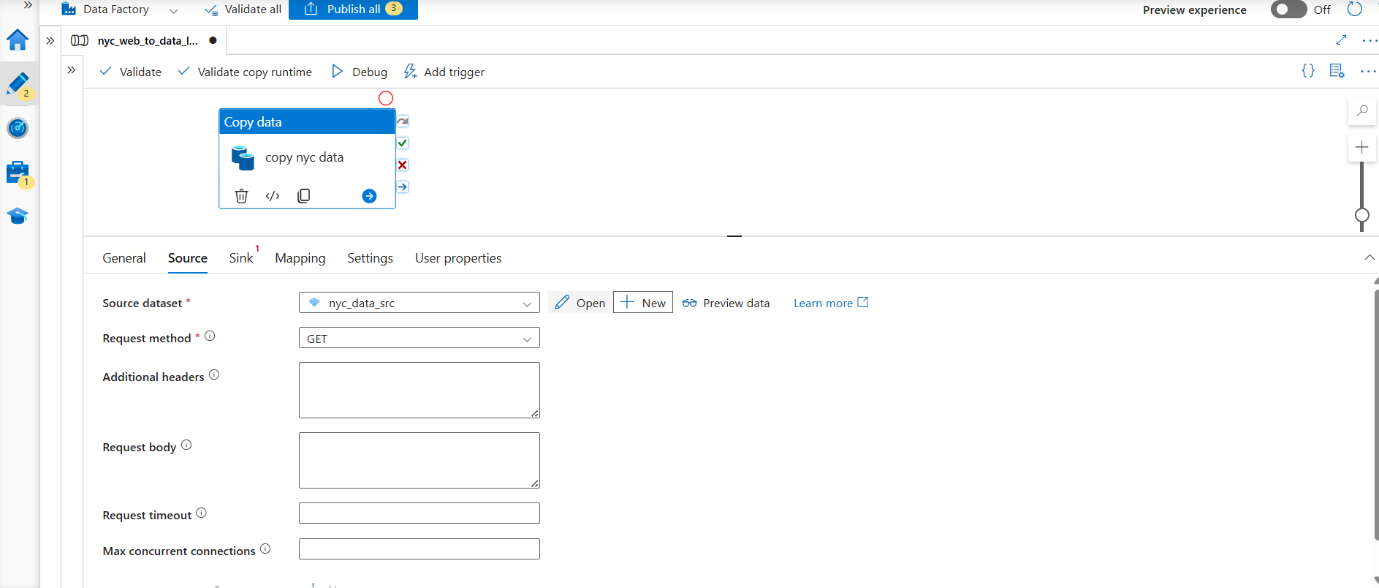
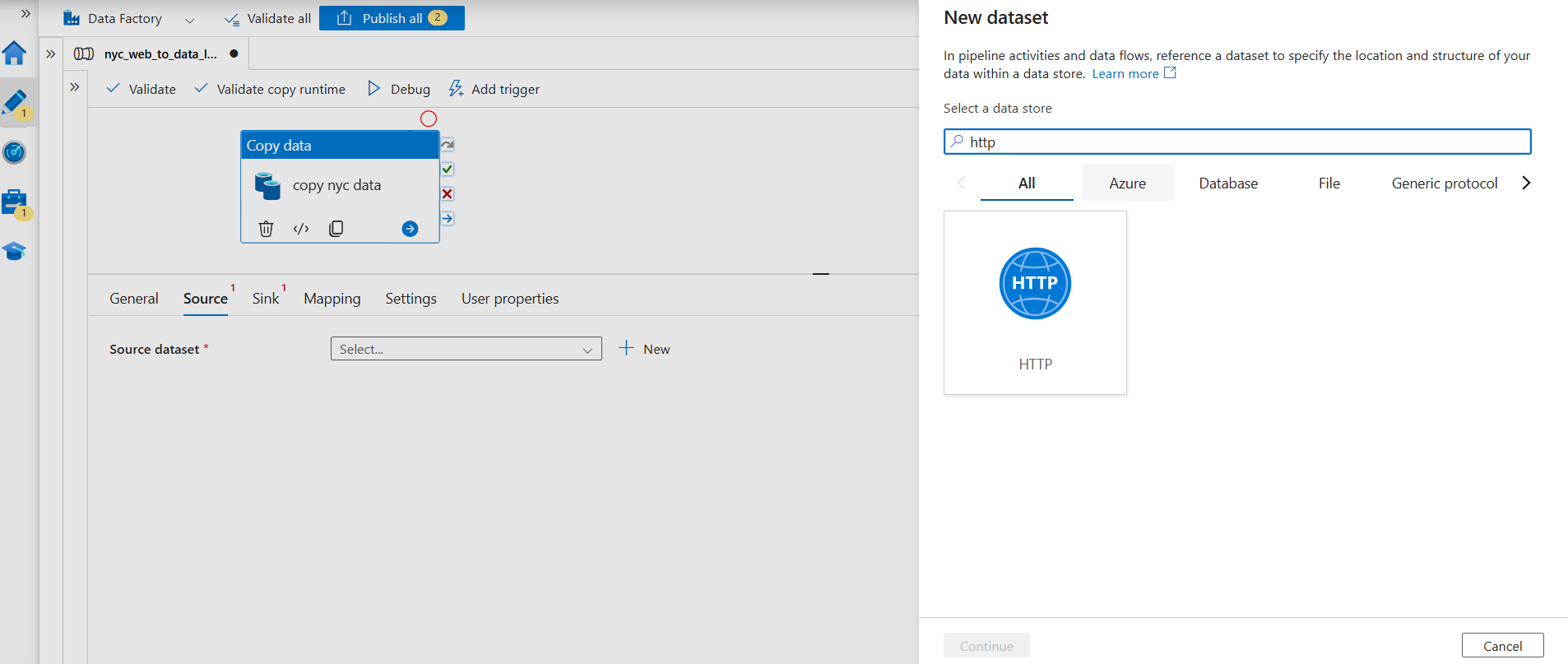
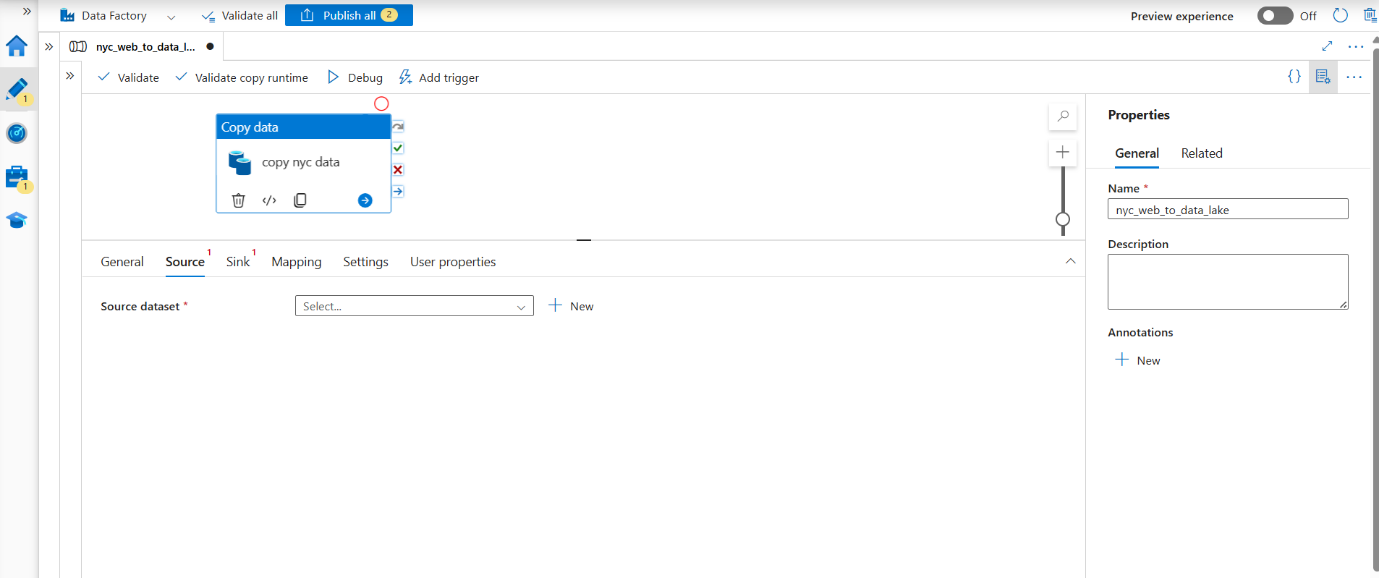
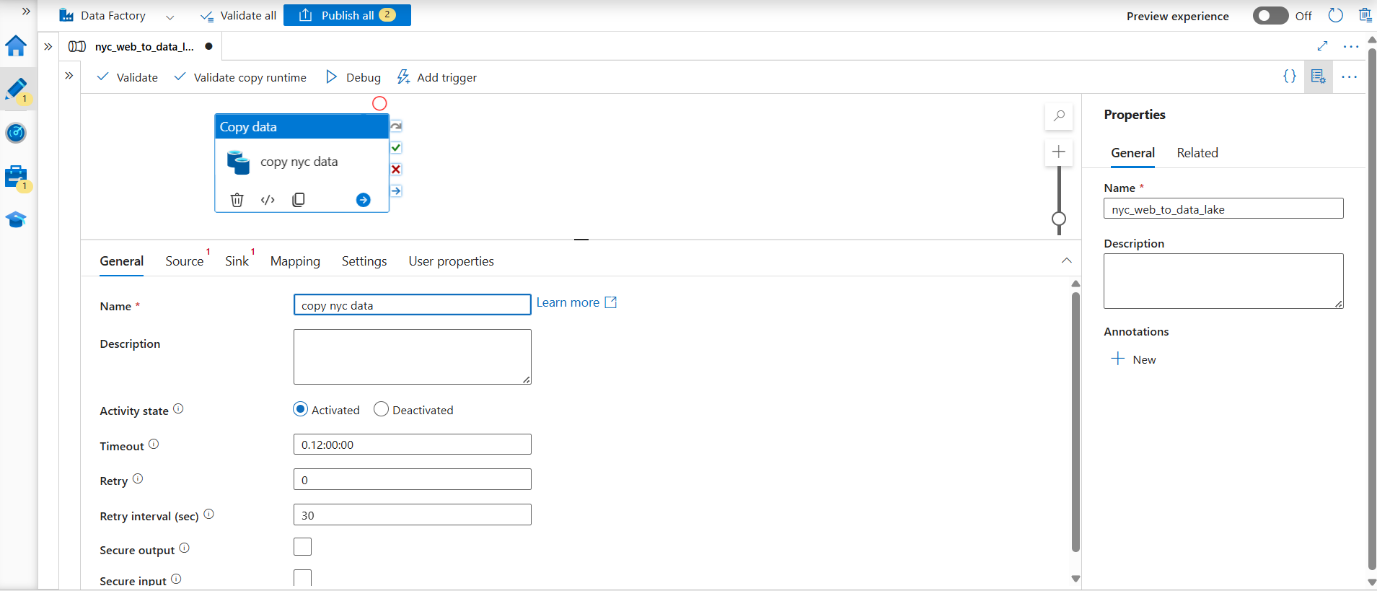
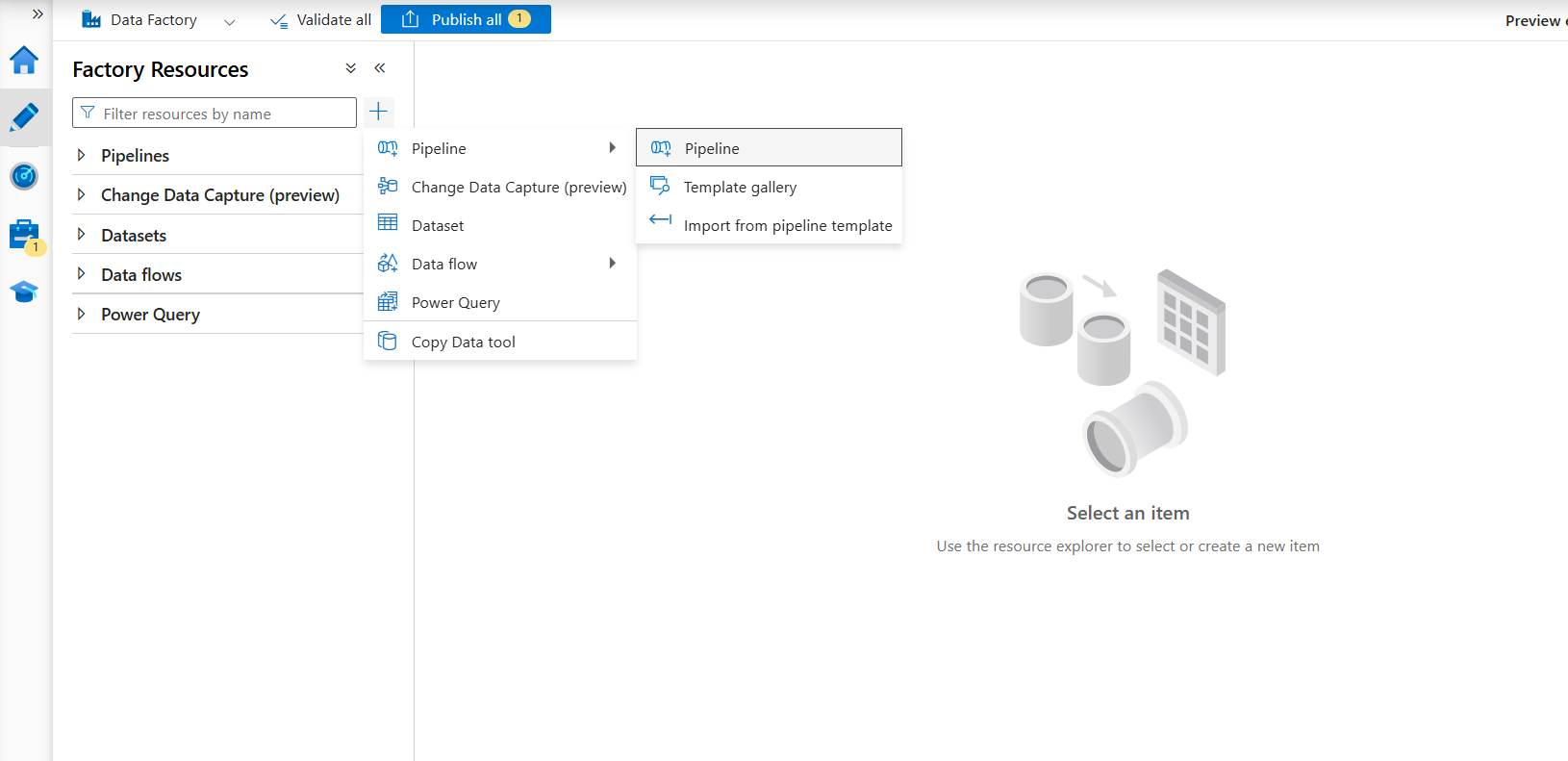
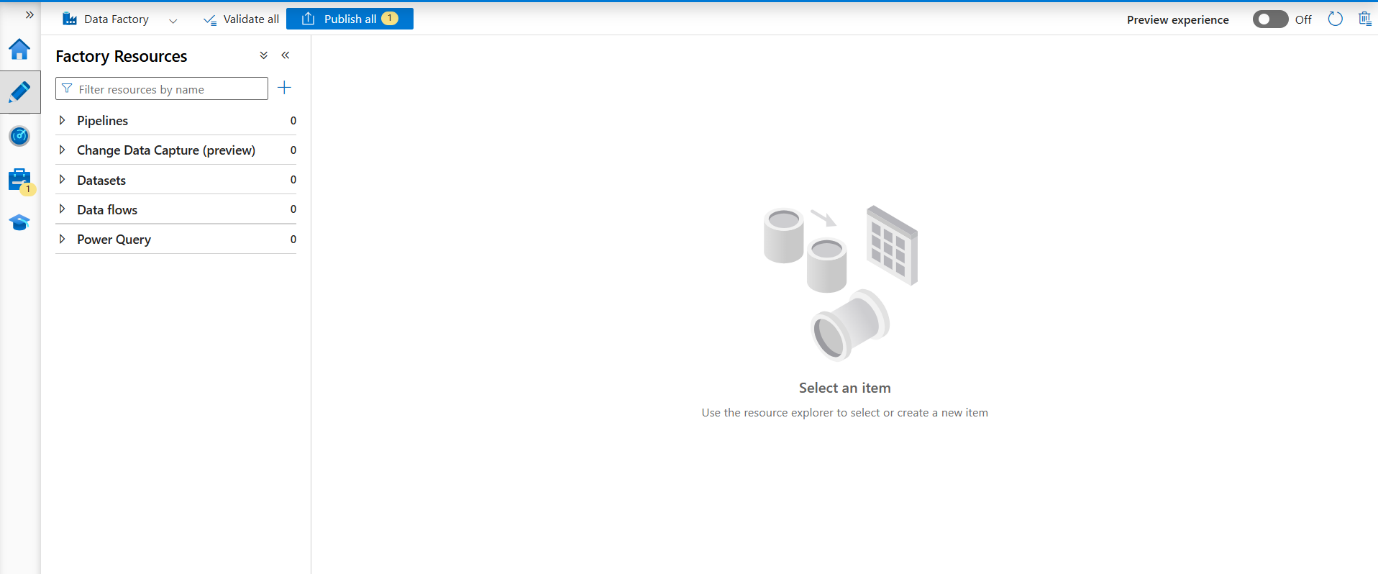
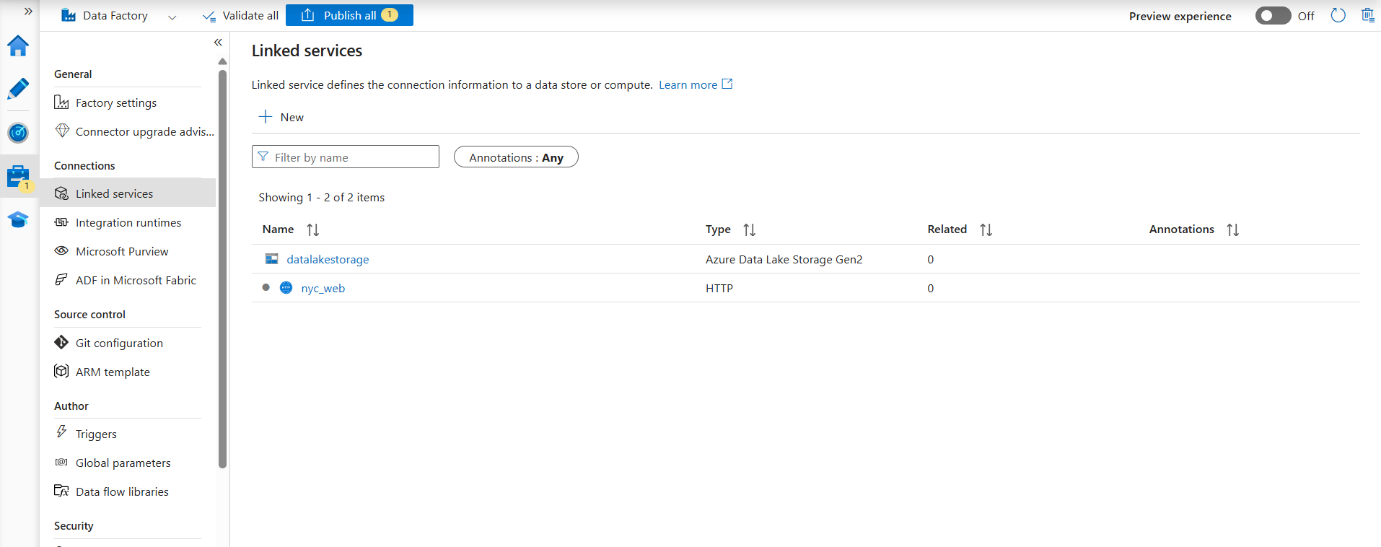
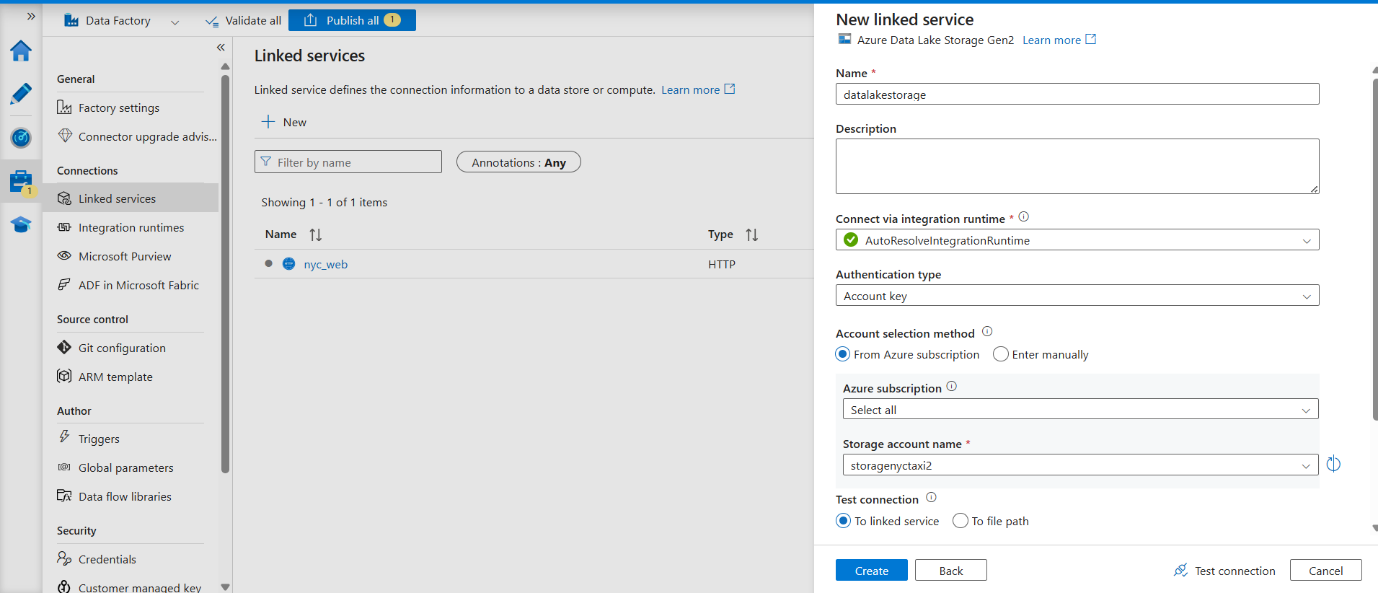
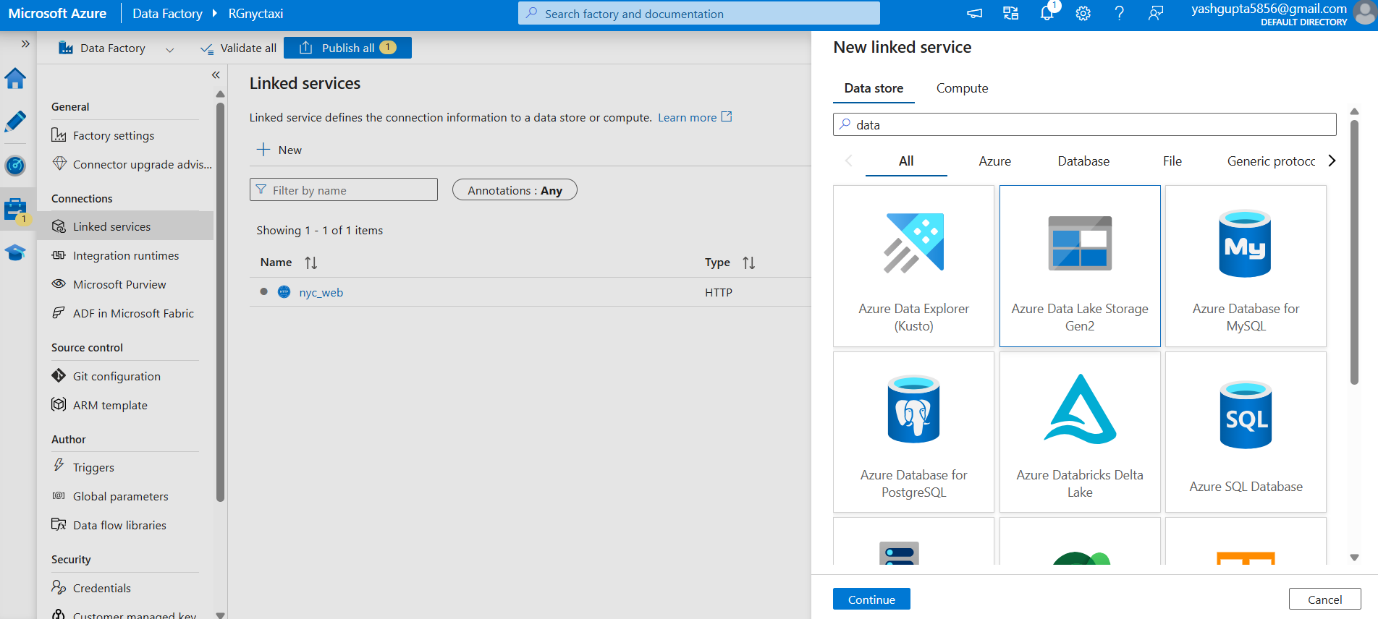
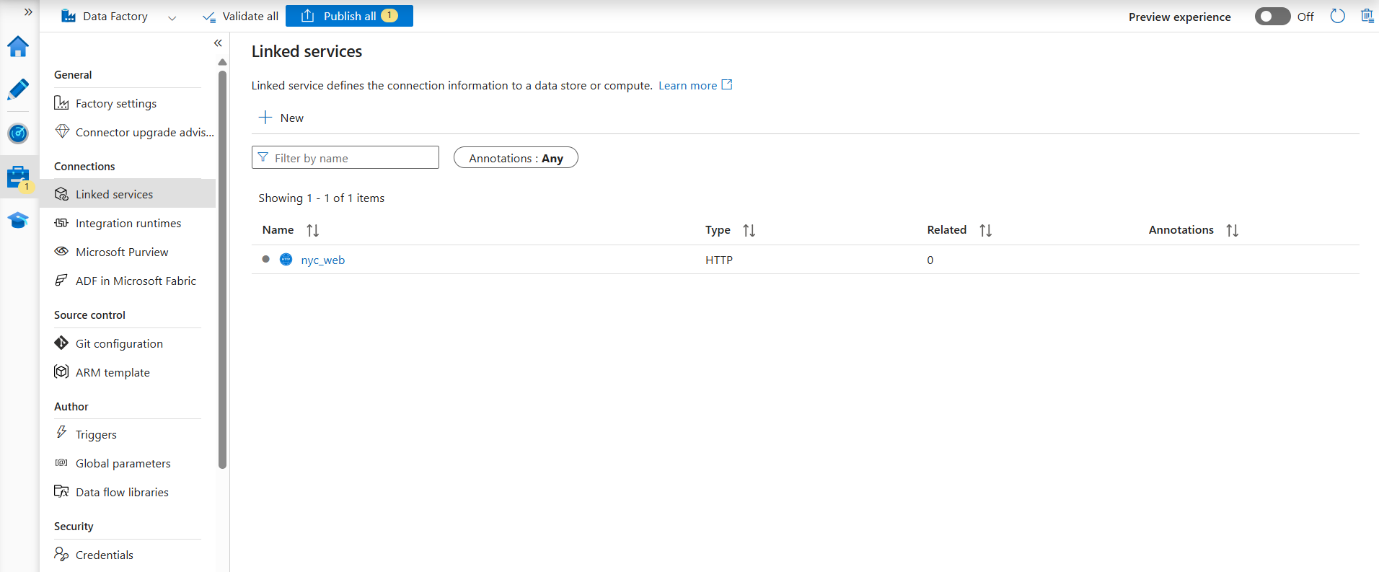
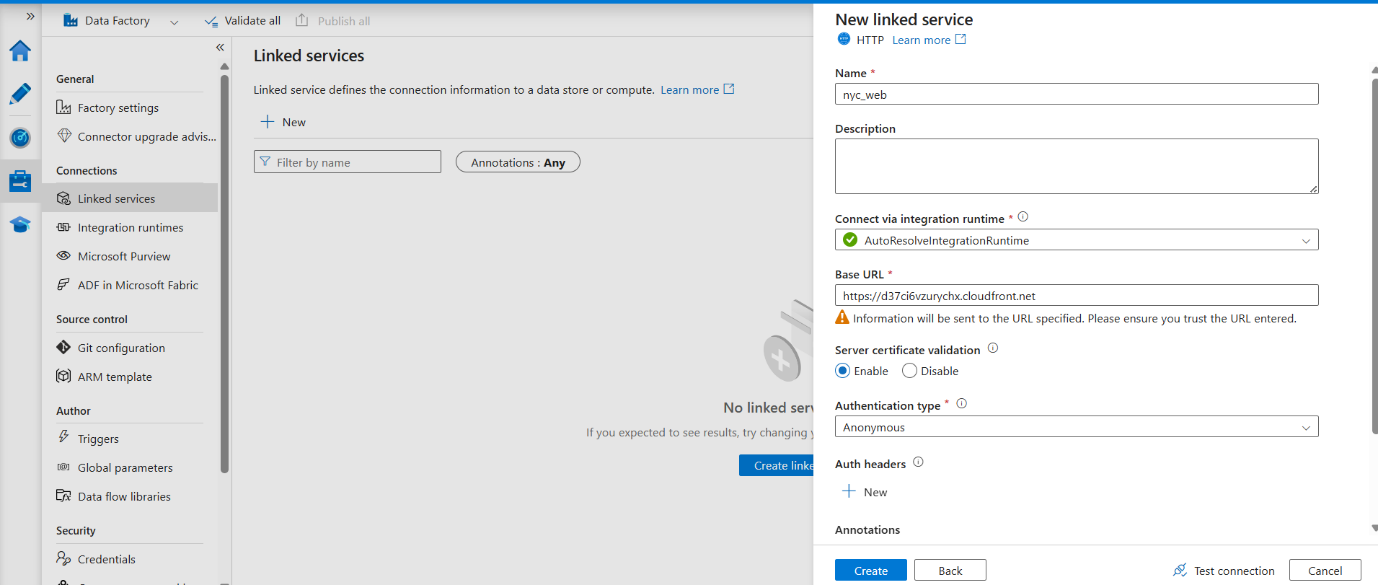
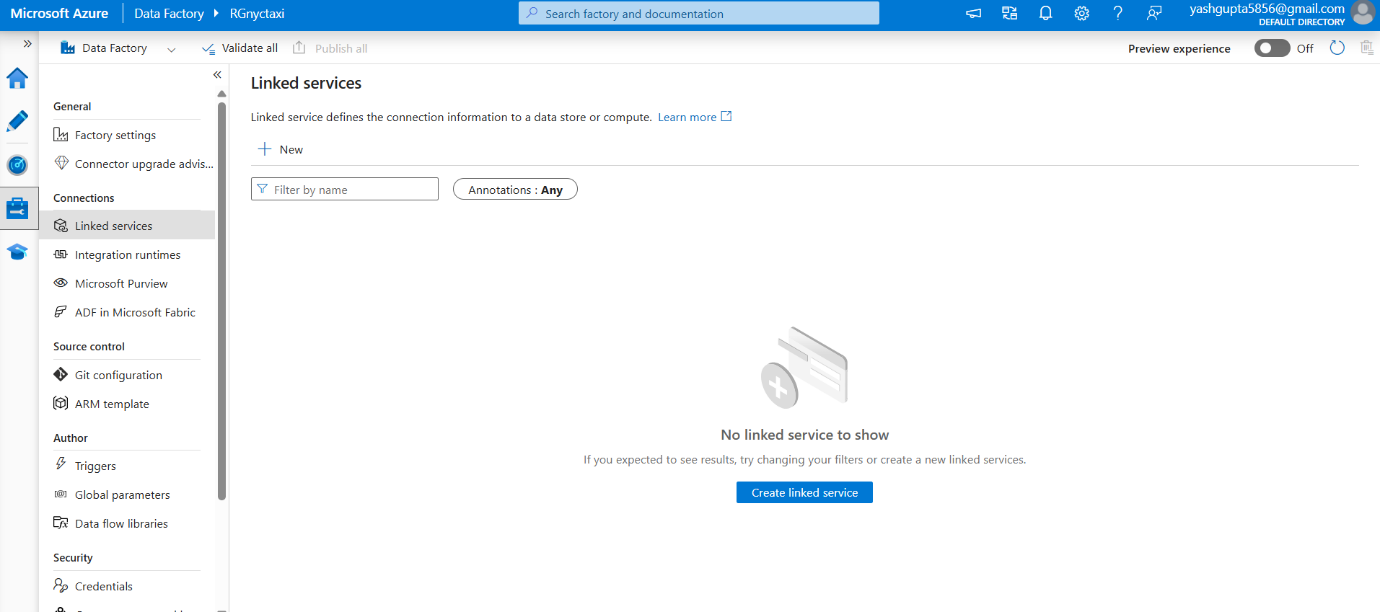
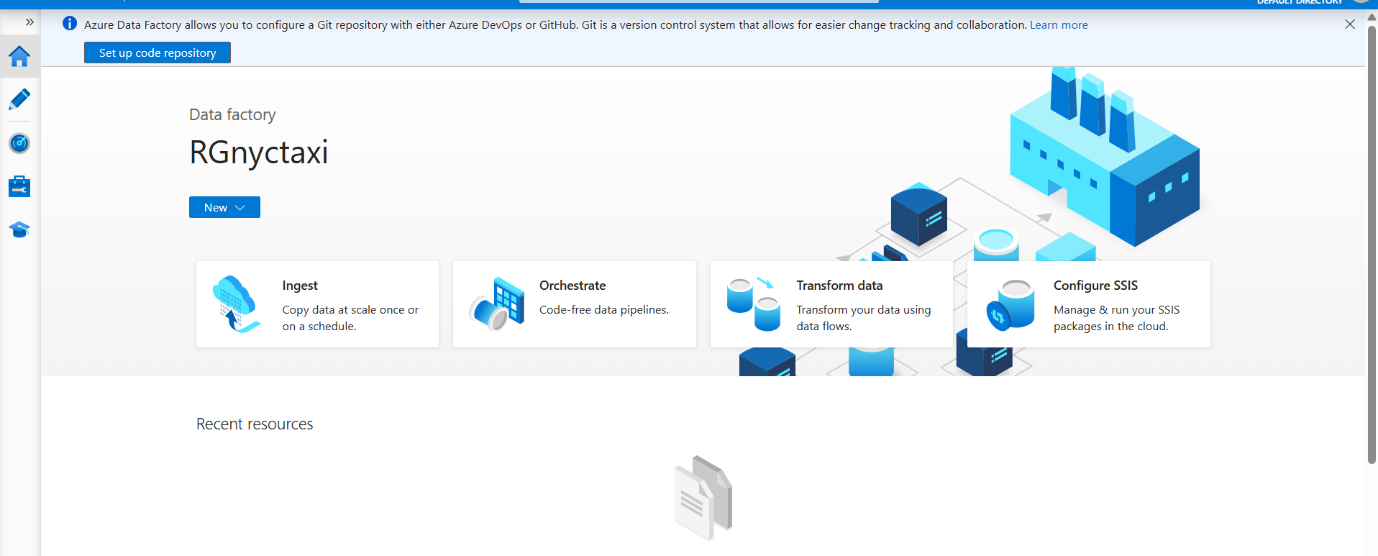
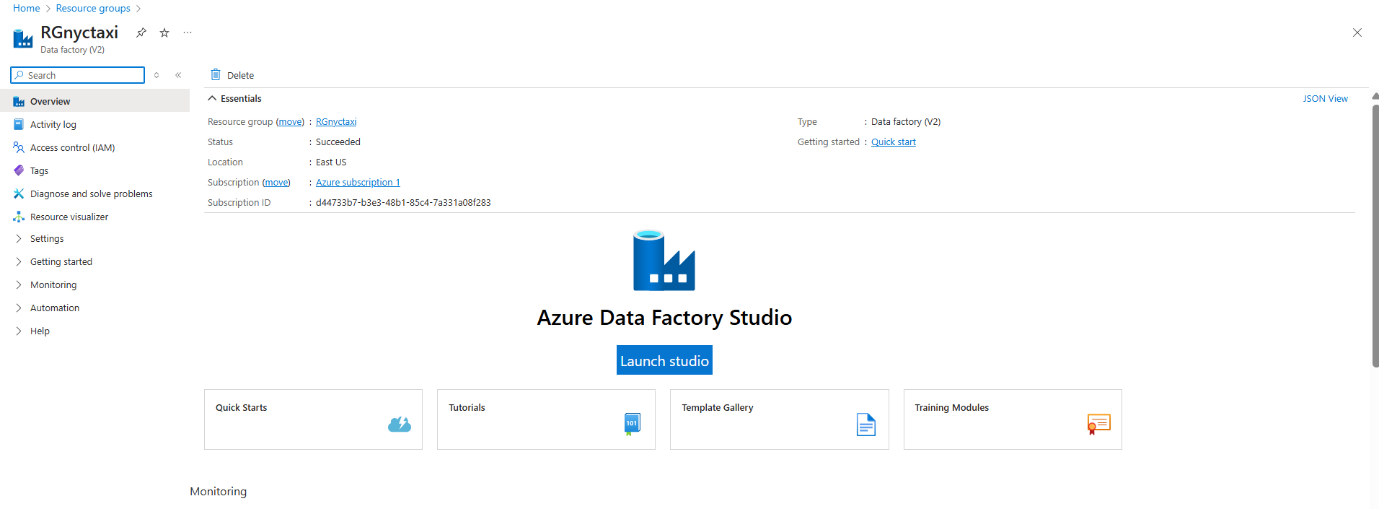
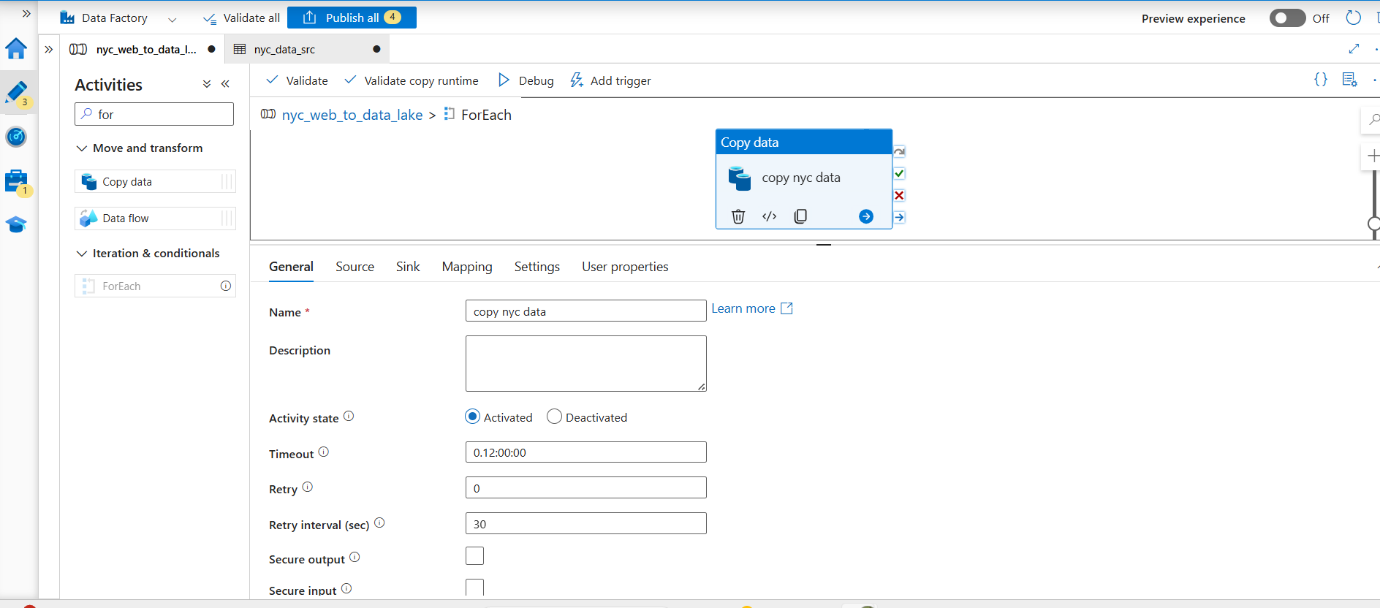
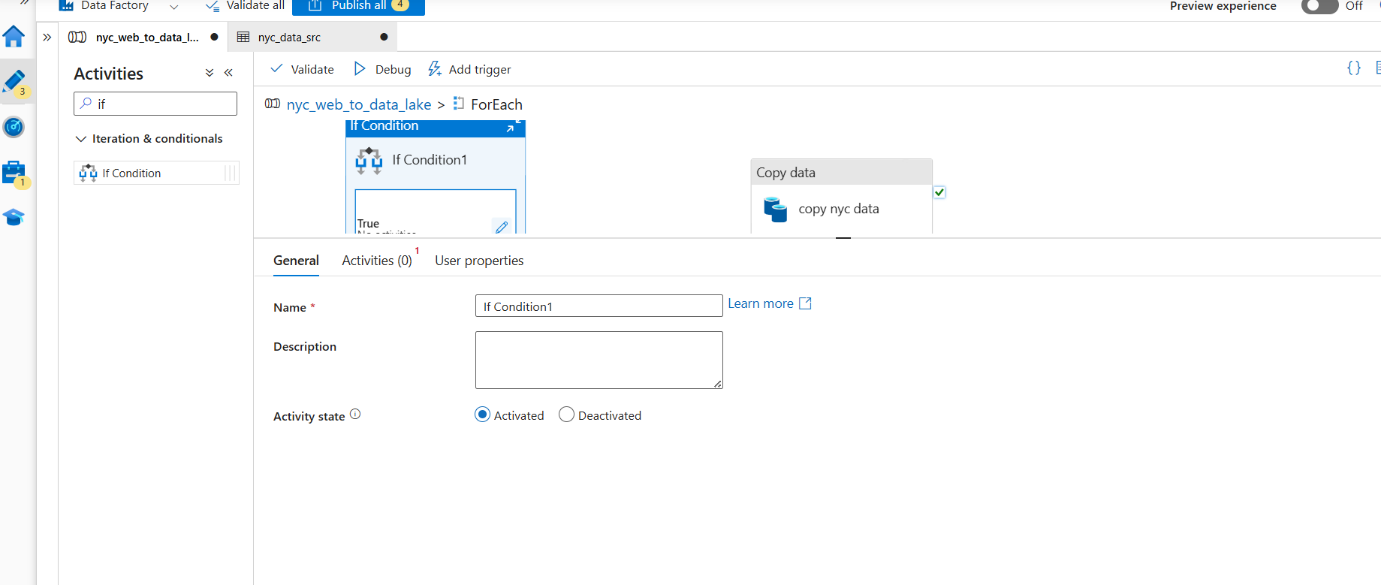
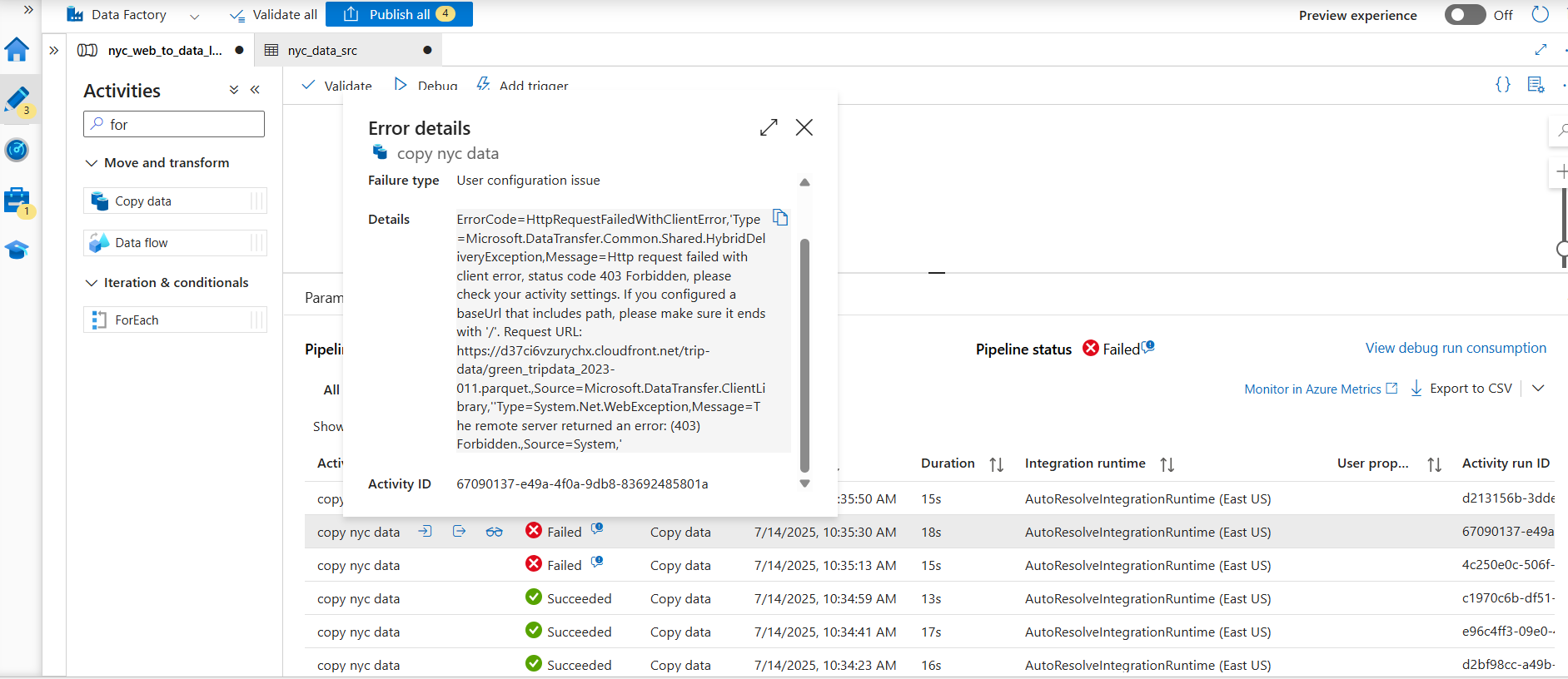
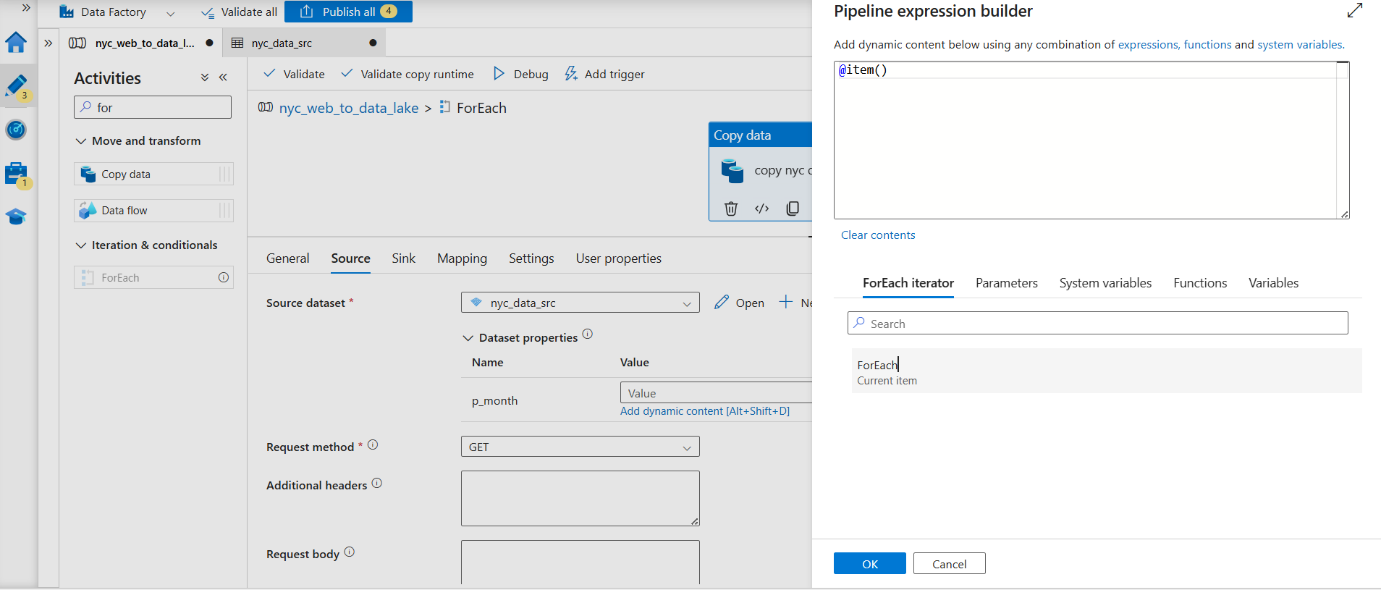
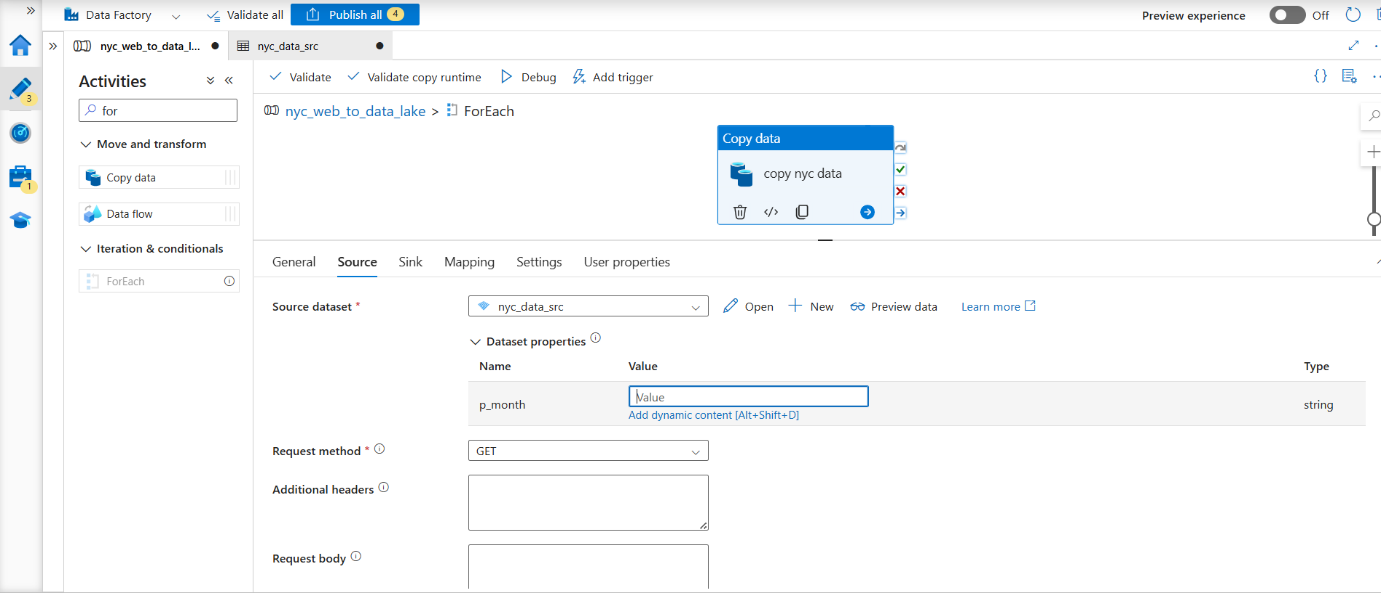
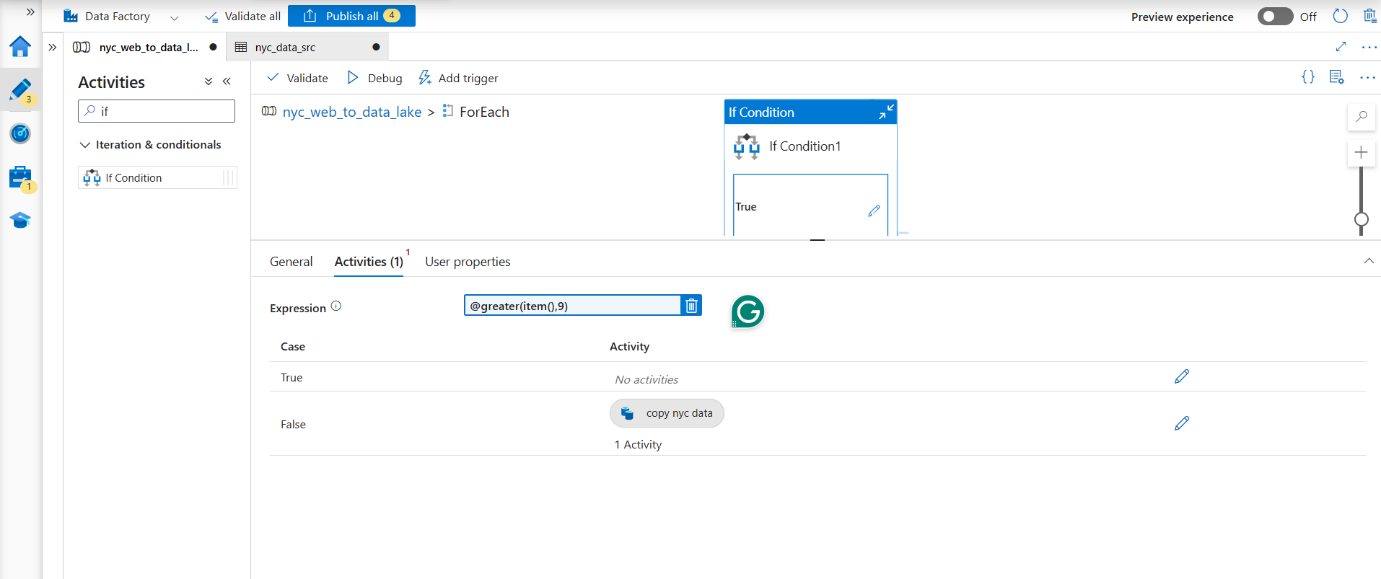
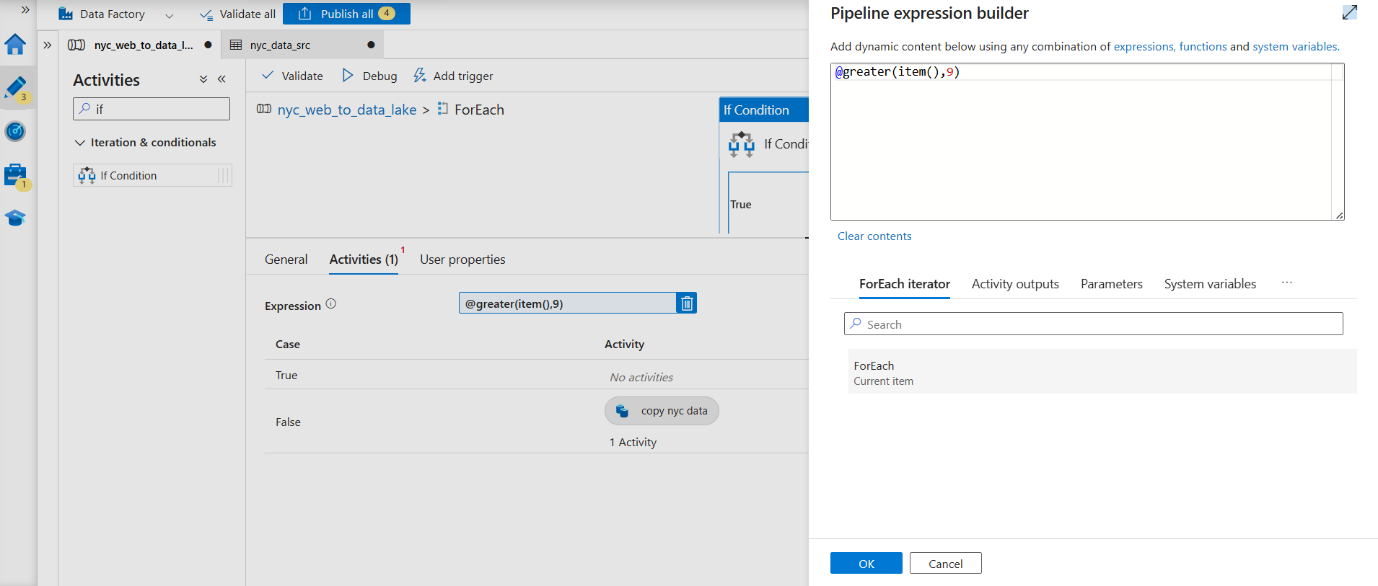
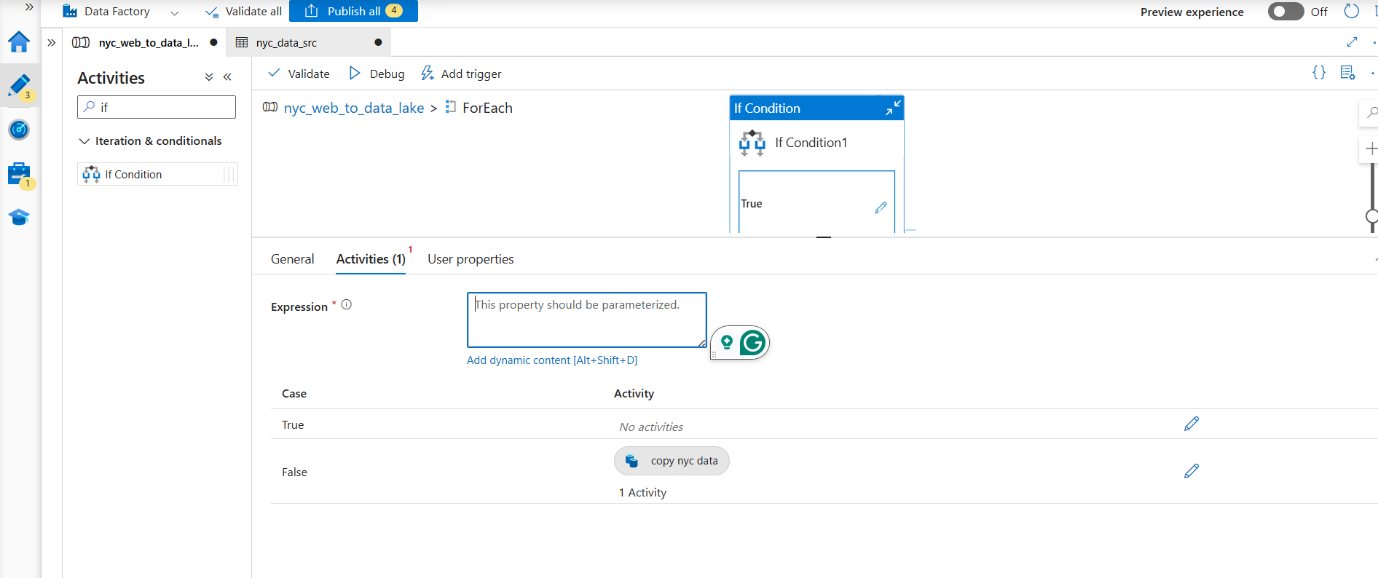
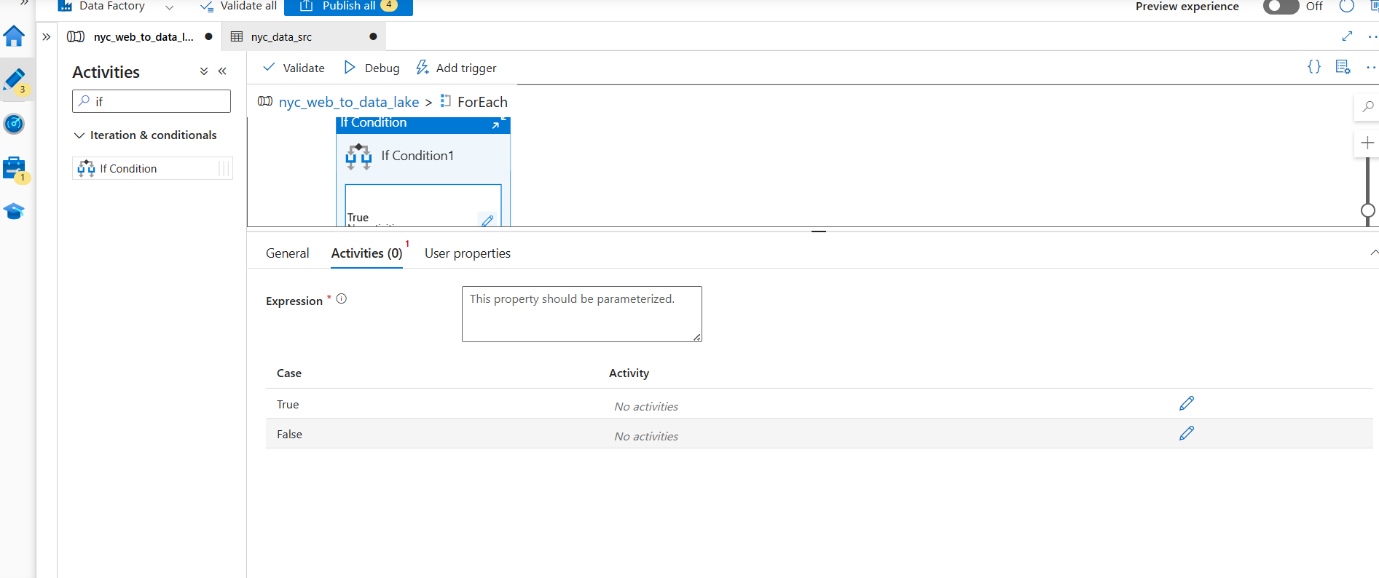










 past copy activity inside foreach activity in (activites)  past copy activity in if codition activity under falseTake copy activity inside if condition under true condition