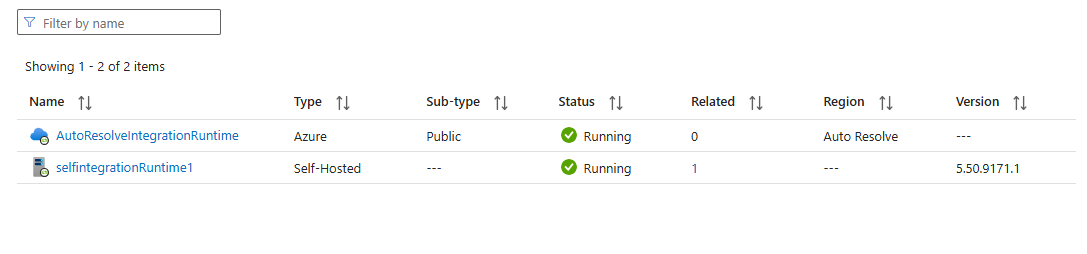
Data Pipeline for Customer Account Analysis

**Objective**

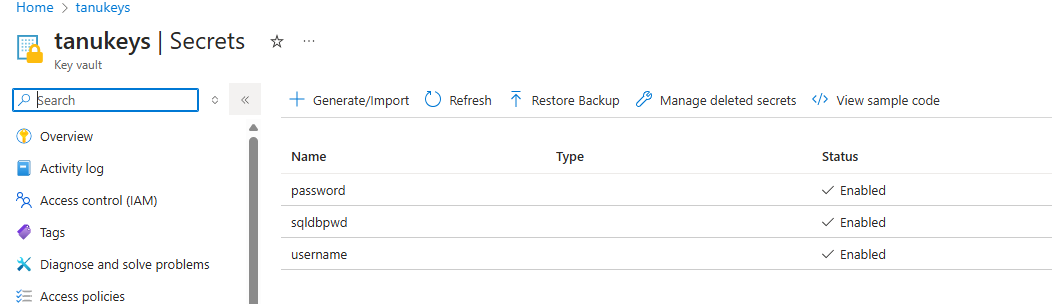
The project aims to design and implement a robust data pipeline for processing customer account data. This includes copying data from a backend team's storage account, performing necessary transformations using Databricks, and upserting (inserting or updating) data from a file stored in Azure Data Lake Storage (ADLS) GOLD Storage into a dedicated SQL pool table in Azure Synapse Analytics. The pipeline aims to ensure efficient, accurate, and scalable data processing to support downstream analytics and reporting needs.

**Step 1:****Data Ingestion (Backend Storage to Raw (Bronze)Container)**

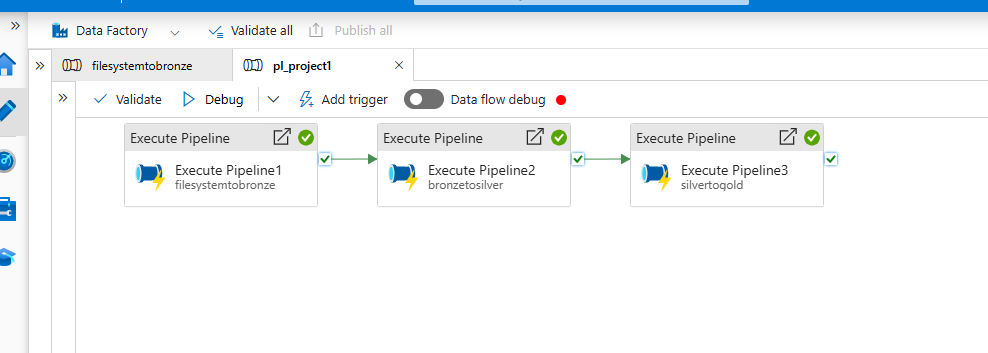
To run a self-hosted integration run time download integration runtime from Microsoft and then install it in our local system.



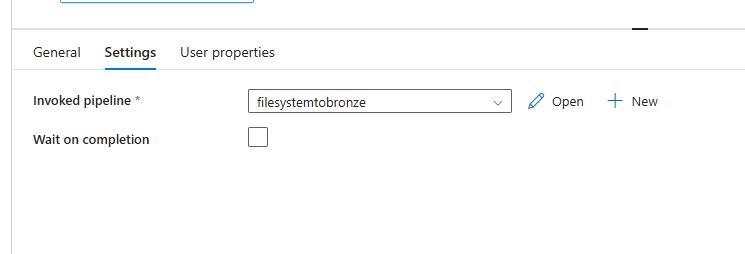
After creating integration runtime we have to create a secret in azure keyvault to store our system password and pass it to linked service.



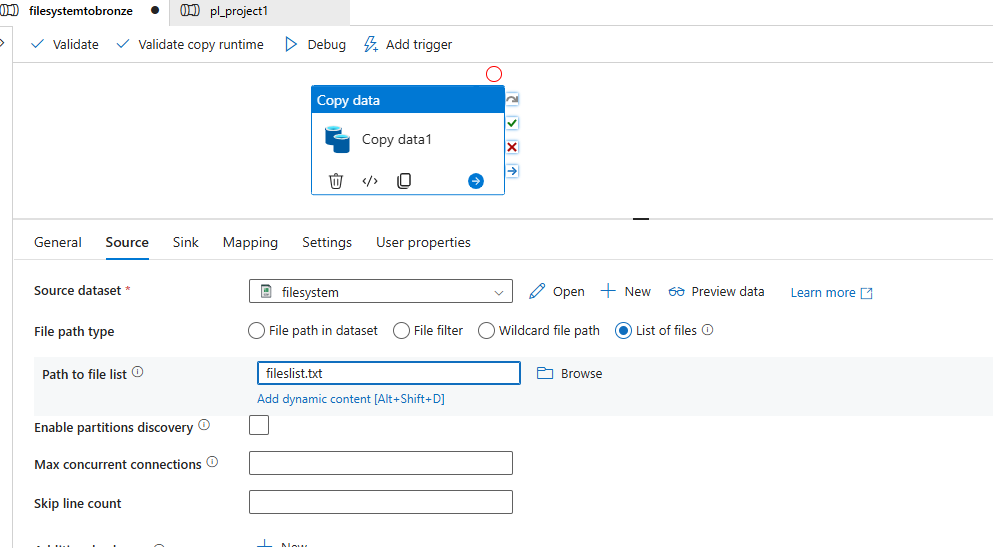
Now go to azure data factory and take a master pipeline and take execute activity task to run this copy activity from filesystem to bronze layer.



Provide pipeline name in execute pipeline activity

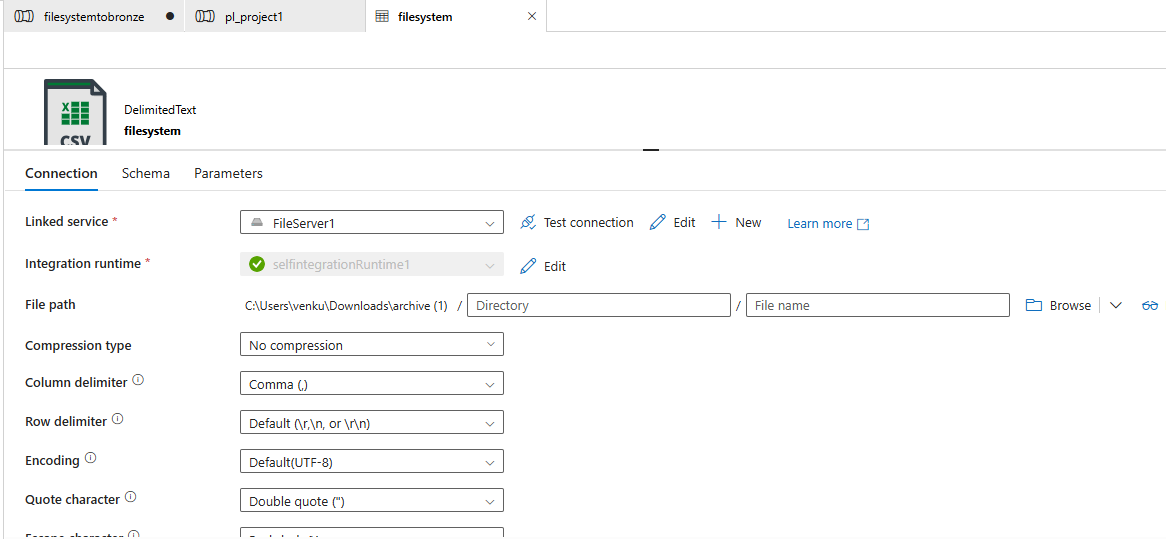


Now open this pipeline and create a copy activity as follows

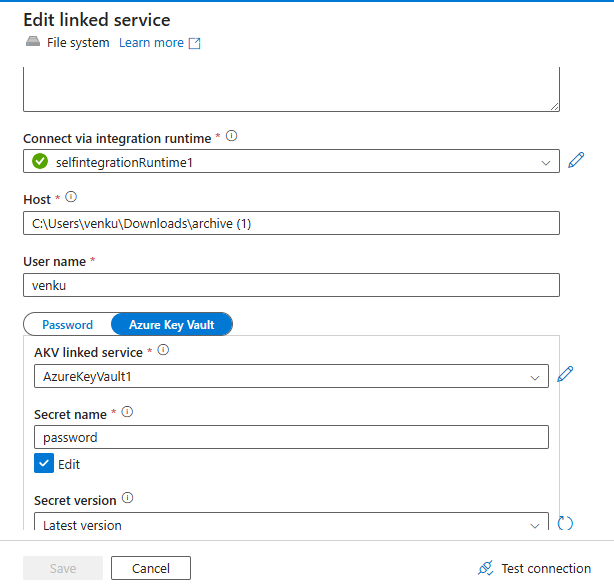


Here I am using list of files method to copy multiple files to sink at once

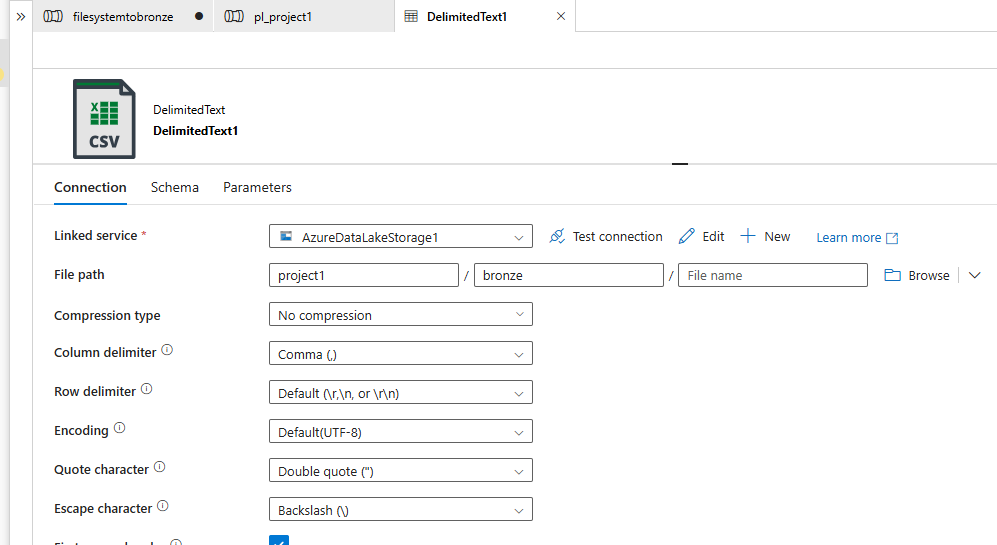
Source details



Linked service details

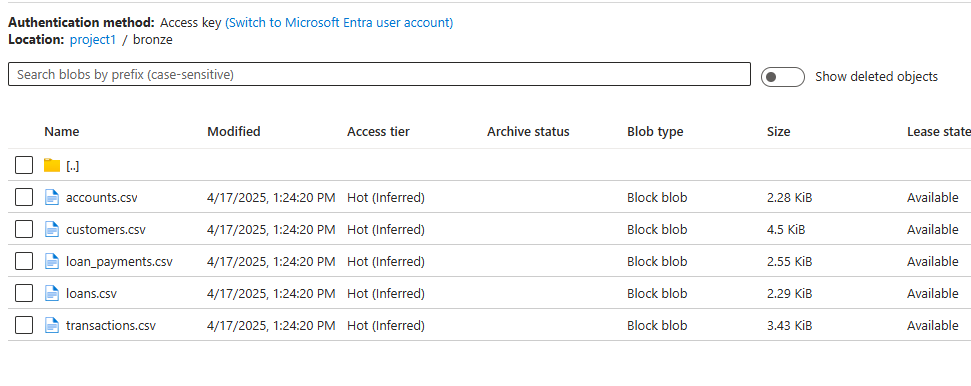


Sink details as follows



In this copy activity we are moving data from local file system to ADLS storage account

Let us look the sink location for files after pipeline run



We will repeat the same above steps for day 2 files list as well then the ADLS storage account has all files

