Initialized by Azure Data Factory!

**Introduction**

This project is a POC for testing the functionalities of Azure Data Factory in order to adapt with the requirements from the problem statement.

**What is Azure Data Factory?**

Azure Data Factory is a cloud-based data integration service for creating ETL (Extract, Transform and Load) and ETL pipelines. It allows users to create data processing workflows in the cloud, either through a graphical interface or by writing code, for orchestrating and automating data movement and data transformation. It is possible to build complex ETL processes that transform data visually with data flows or by using compute services such as Azure HDInsight Hadoop, Azure Databricks, and Azure SQL Database.

**Requirement**

• Create, run and manage pipelines for data transfer between MySql & Azure Blob Storage.

1. Line\_Number
2. Budget\_Code
3. PurchaseOrg
4. PurchaseGroup
5. CompanyCode
6. Cost\_Center
7. GeneralLedger
8. WBSElement
9. Asset
10. SubNumber
11. InternalOrder
12. ProjectNumber
13. Rank\_List
14. Budget\_Check\_Disabled

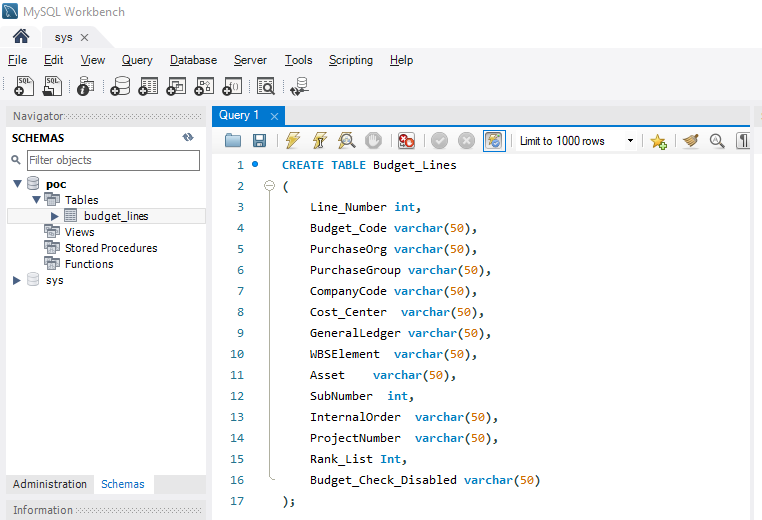
• Automation of the pipeline to be able to check the state of the last pipeline run

**Getting Started**

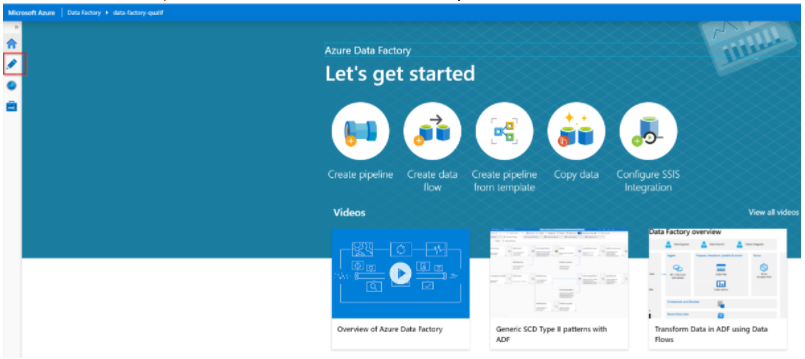
1. Required to have MySql to host Azure Data Factory.
2. On the Azure Windows, some raw data were copied into this Budget\_Lines table to simulate the scenario of raw data at the MySql . We use azcopy tool to perform the data transfer from MySql to Azure Blob.
3. Once we have the raw data on Azure, we create different pipelines to transfer these data to test container on Azure Blob Storage.

**Build and Test**

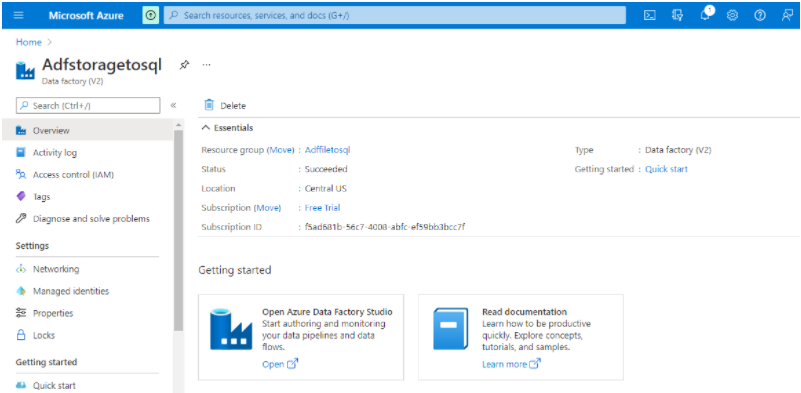
1. Login to MySql and create required table in database



2.Navigate to Azure Data Factory portal, This Azure Data Factory is running on the Windows Server. Click on Author & Monitor to open the editor mode of Azure Data Factory.

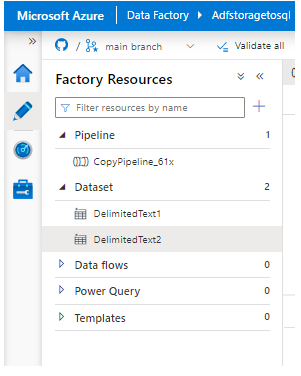


1. Click on Azure Data Factory Studio.

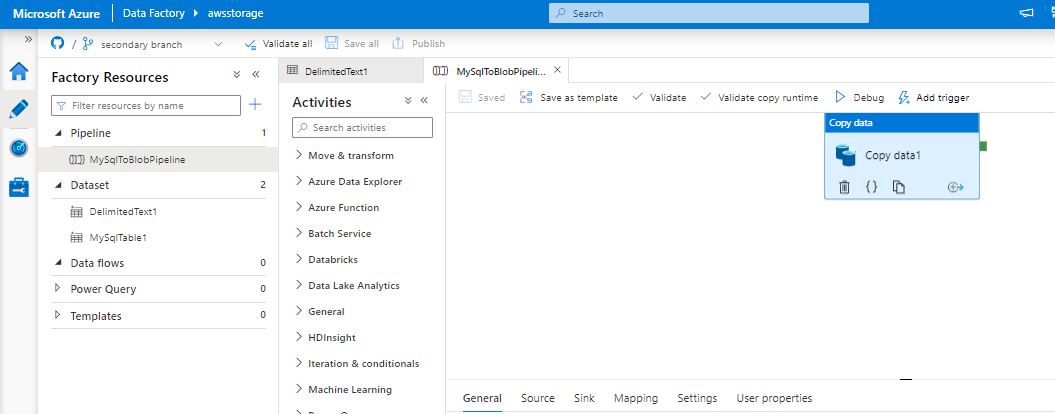


1. Under the left menu Factory Resources is where you can find all the

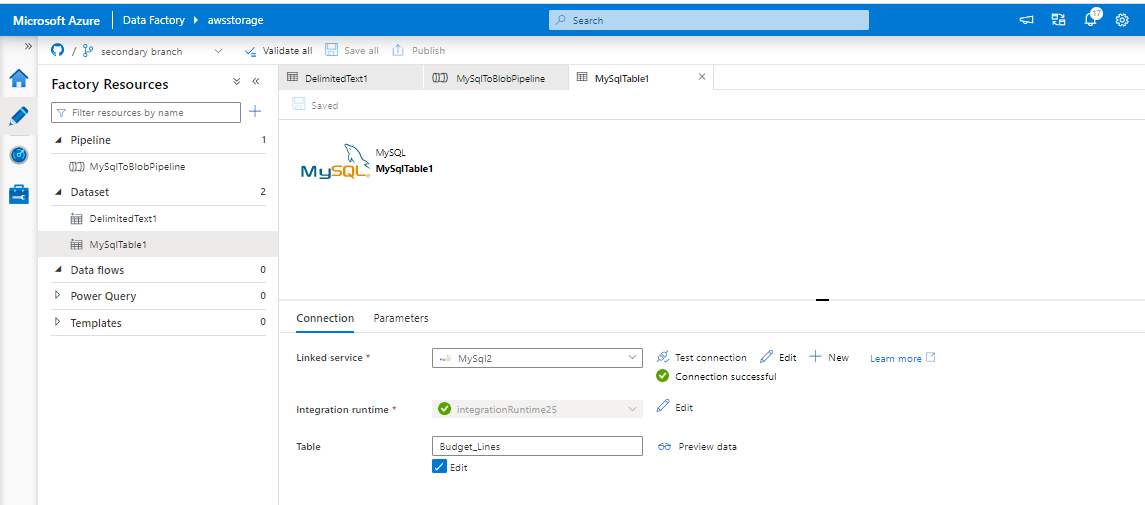
available pipelines,datasets & activities.

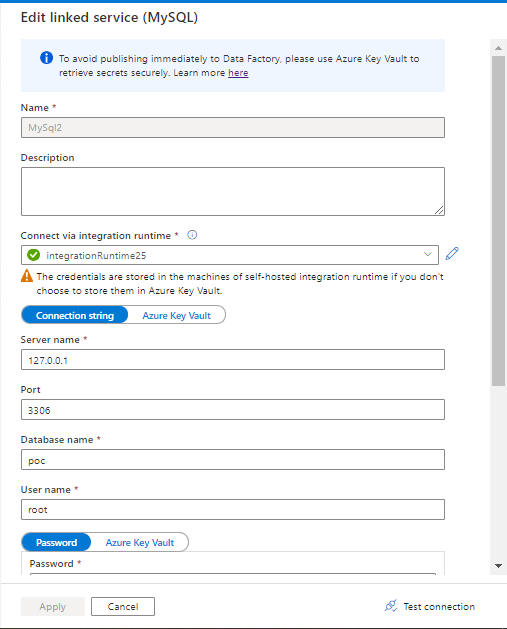


1. The Pipeline has the following workflow

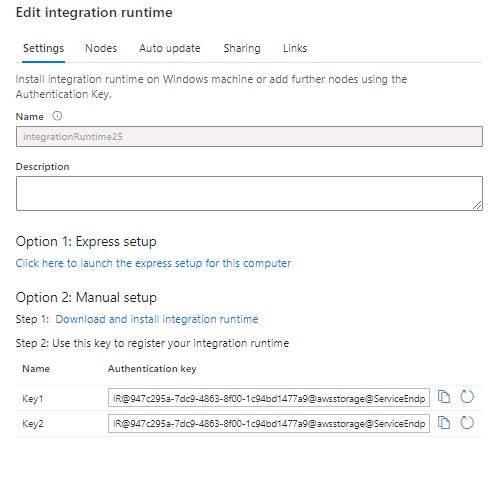


1. Two datasets we have and which defines MySql & Blob details.
   1. Dataset details using Linked service for MySql

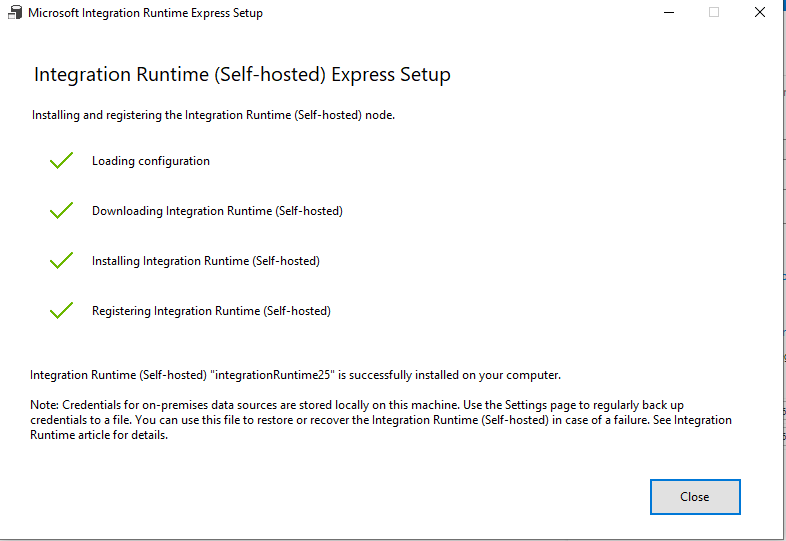




* 1. Selfhosted integration runtime details.



* 1. Click here to install self hosted integration in windows.



d.Dataset details using Linked service for Blob

