**Name:** V Venkata Sri Prasad

**Batch:** Data Engineering

**Date:** 20/02/2024 – (Day 24)

Topics:

1. What is Azure Data Factory
2. Creating Axure Data Factory
3. Copying data from one resource to another using pipeline



**What is Azure Data Factory?**

Azure Data Factory is a cloud-based data integration service that allows you to create data-driven workflows in the cloud for orchestrating and automating data movement and data transformation.

ADF does not store any data itself. It allows you to create data-driven workflows to orchestrate the movement of data between supported data stores and then process the data using compute services in other regions or in an on-premise environment. It also allows you to monitor and manage workflows using both programmatic and UI mechanisms.

**Azure Data Factory use cases**

**ADF can be used for:**

Supporting data migrations

Getting data from a client’s server or online data to an Azure Data Lake

Carrying out various data integration processes

Integrating data from different ERP systems and loading it into Azure Synapse for reporting

**How does Azure Data Factory work?**

The Data Factory service allows you to create data pipelines that move and transform data and then run the pipelines on a specified schedule (hourly, daily, weekly, etc.). This means the data that is consumed and produced by workflows is time-sliced data, and we can specify the pipeline mode as scheduled (once a day) or one time.

**Azure Data Factory pipelines (data-driven workflows) typically perform three steps.**

**Step 1: Connect and Collect**

Connect to all the required sources of data and processing such as SaaS services, file shares, FTP, and web services. Then, move the data as needed to a centralized location for subsequent processing by using the Copy Activity in a data pipeline to move data from both on-premise and cloud source data stores to a centralization data store in the cloud for further analysis.

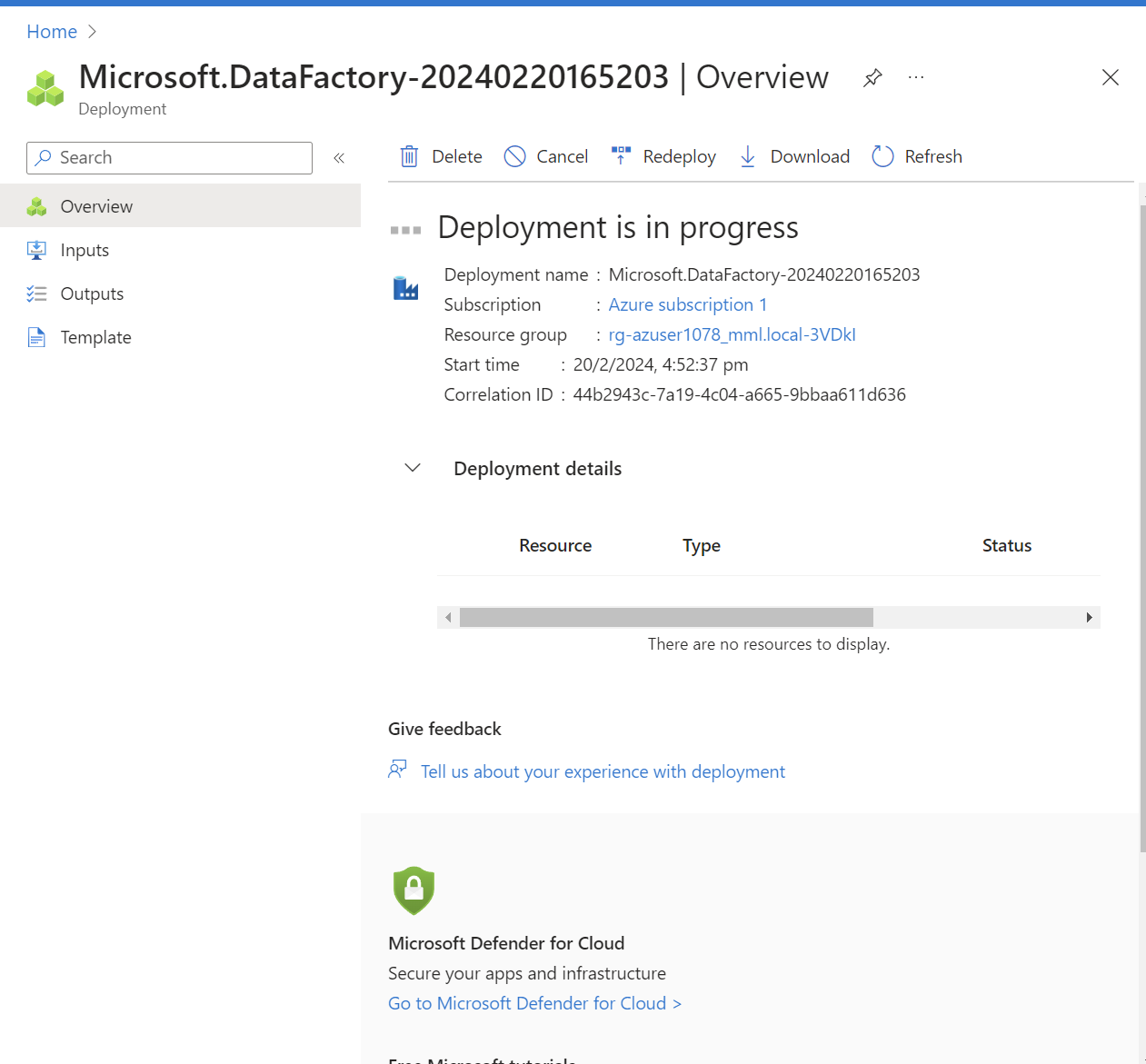
**Step 2: Transform and Enrich**

Once data is present in a centralized data store in the cloud, it is transformed using compute services such as HDInsight Hadoop, Spark, Azure Data Lake Analytics, and Machine Learning.

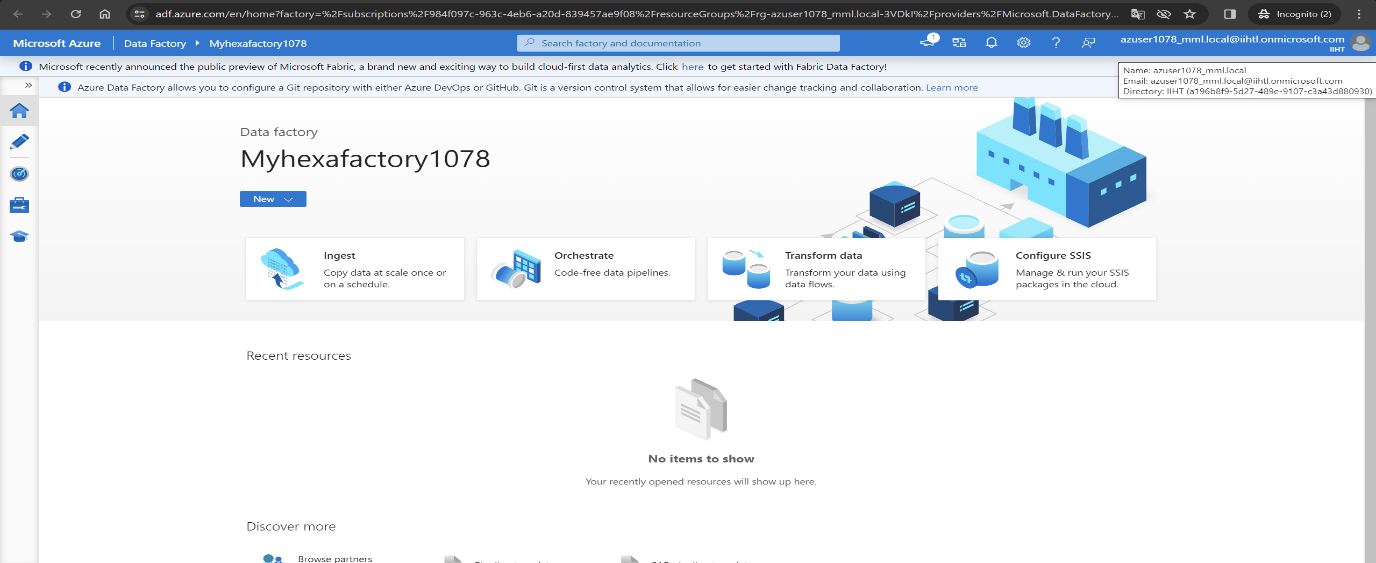
**Step 3: Publish**

Deliver transformed data from the cloud to on-premise sources like SQL Server or keep it in your cloud storage sources for consumption by BI and analytics tools and other applications.

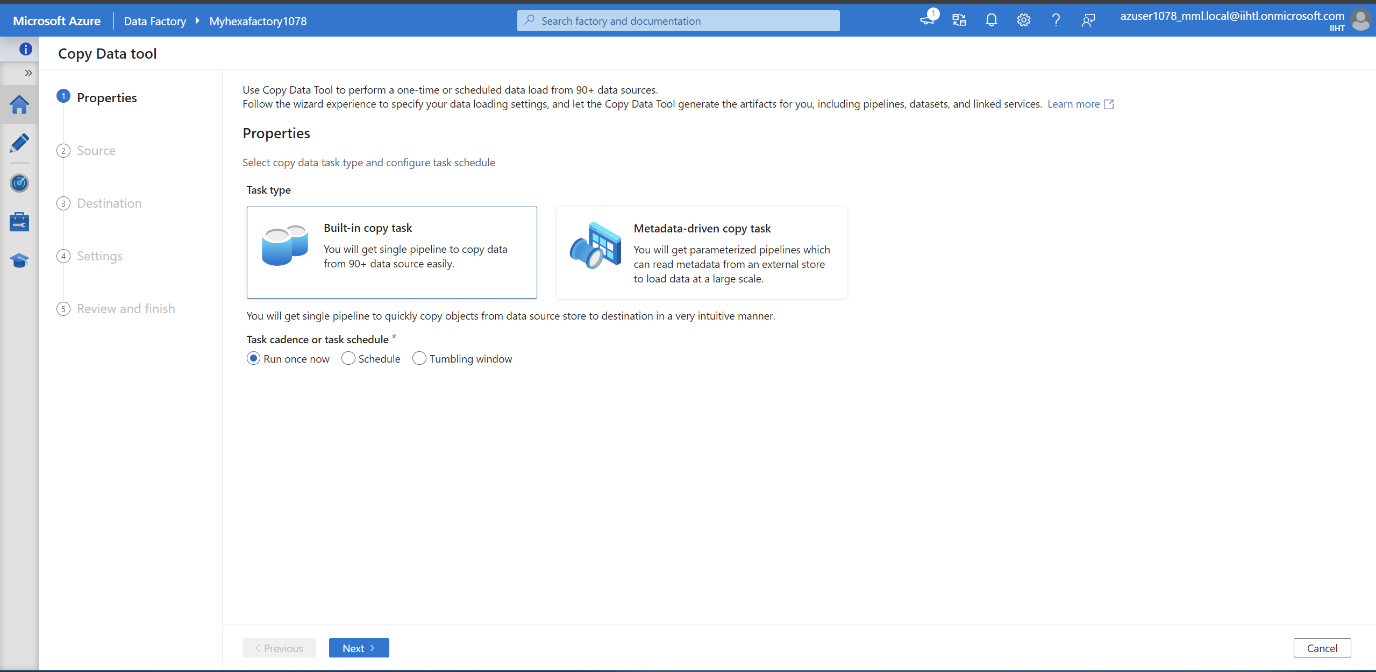
Create a Data Factory account in azure portal and open the resource.



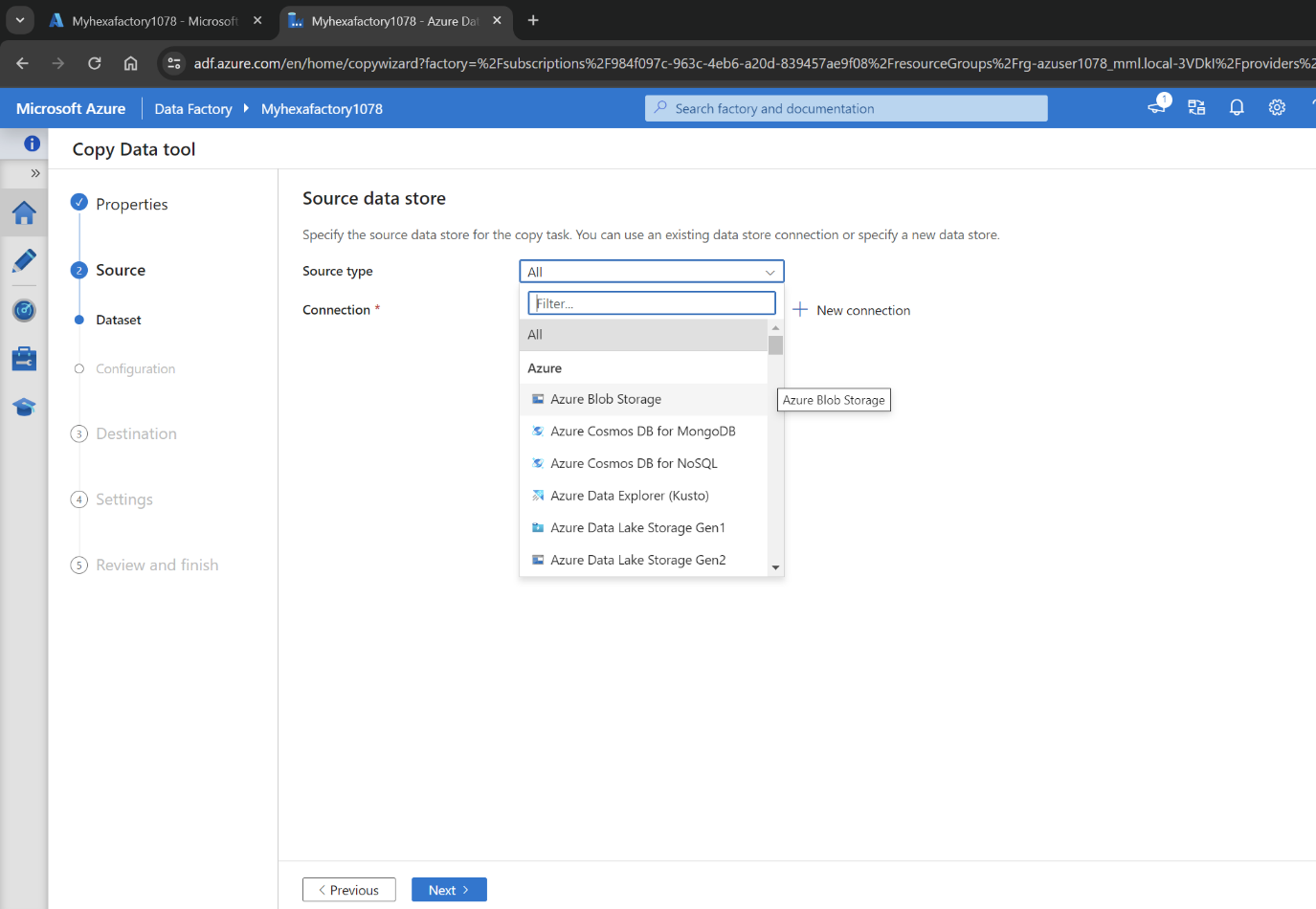
Go to home page and click on Ingest



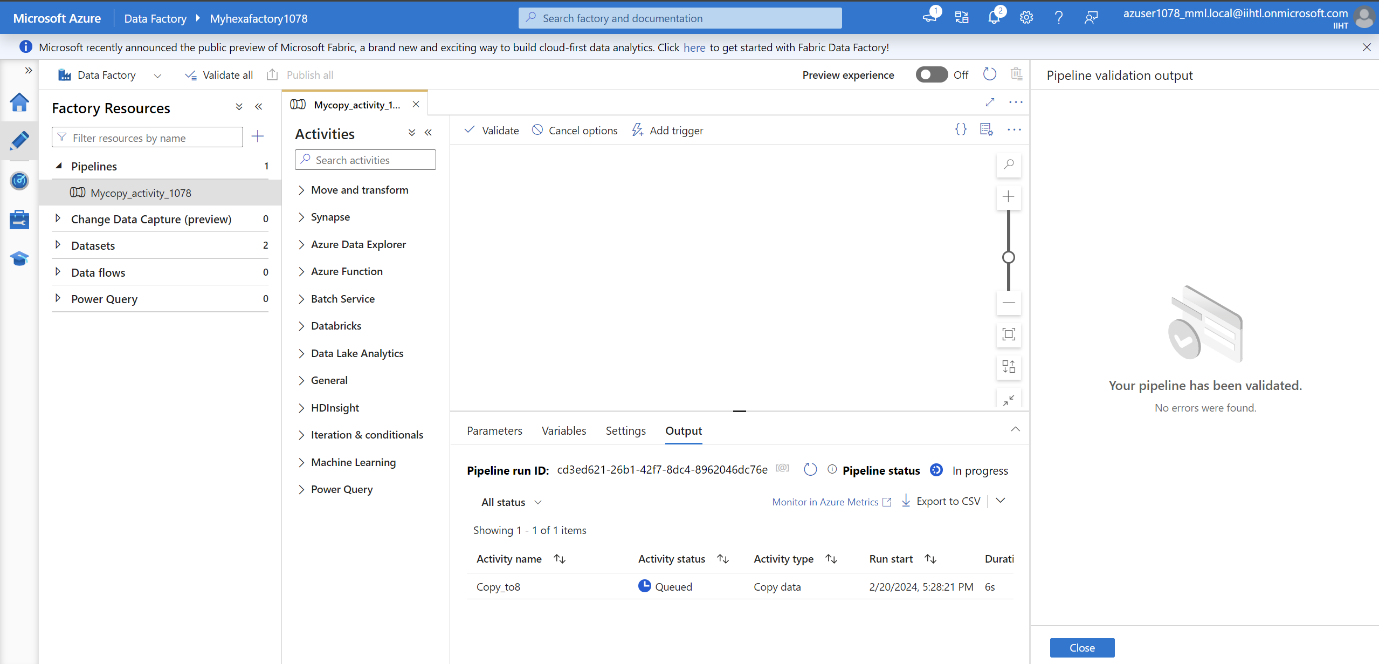
Select Built a copy task and run once now and click on next



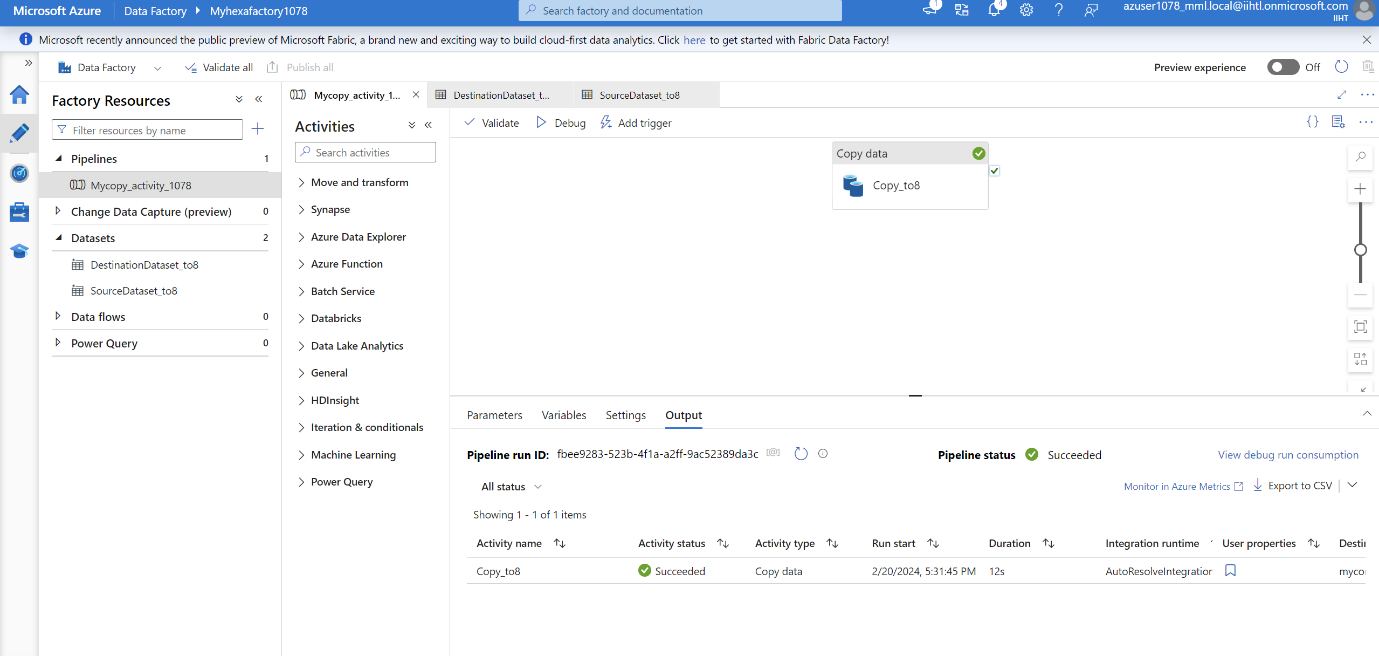
Select the source dataset and give the path with access key



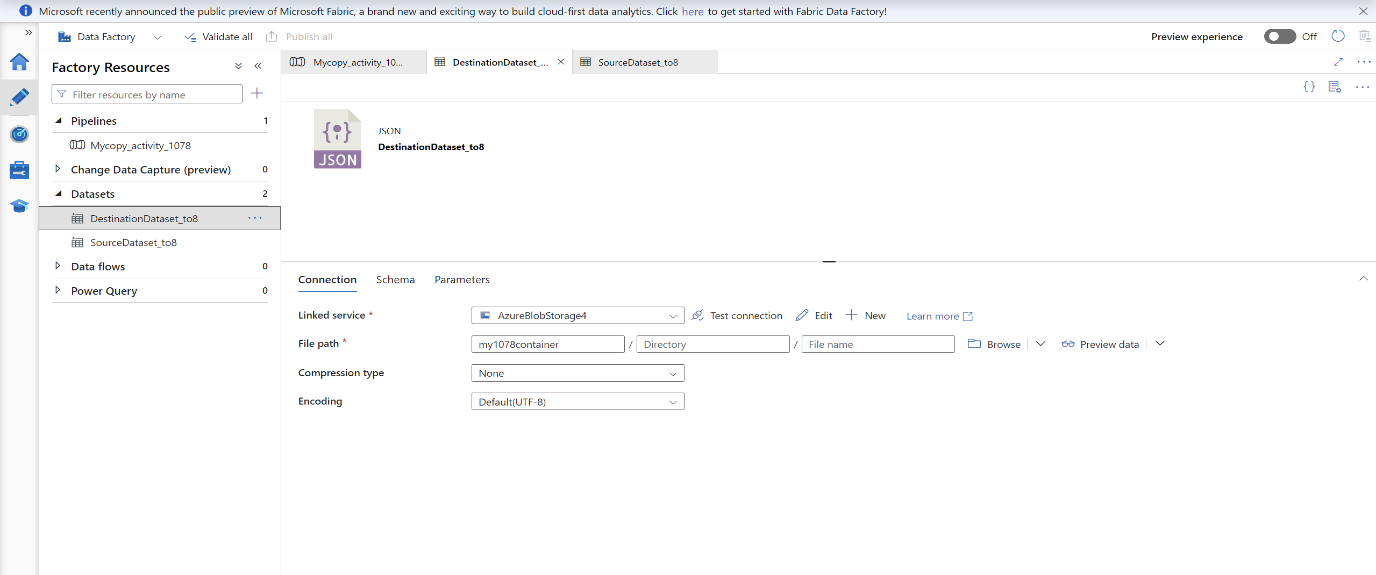
After creating we can see here the activity is created



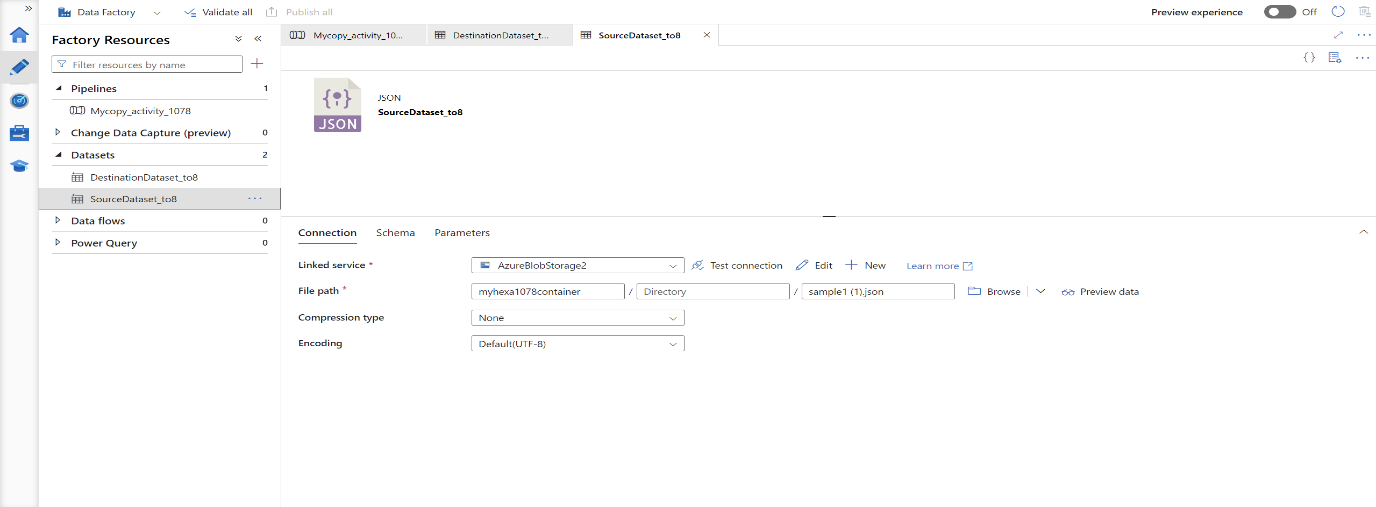
Clicking on validate and Debug to run the testcase and we can see it is successfull



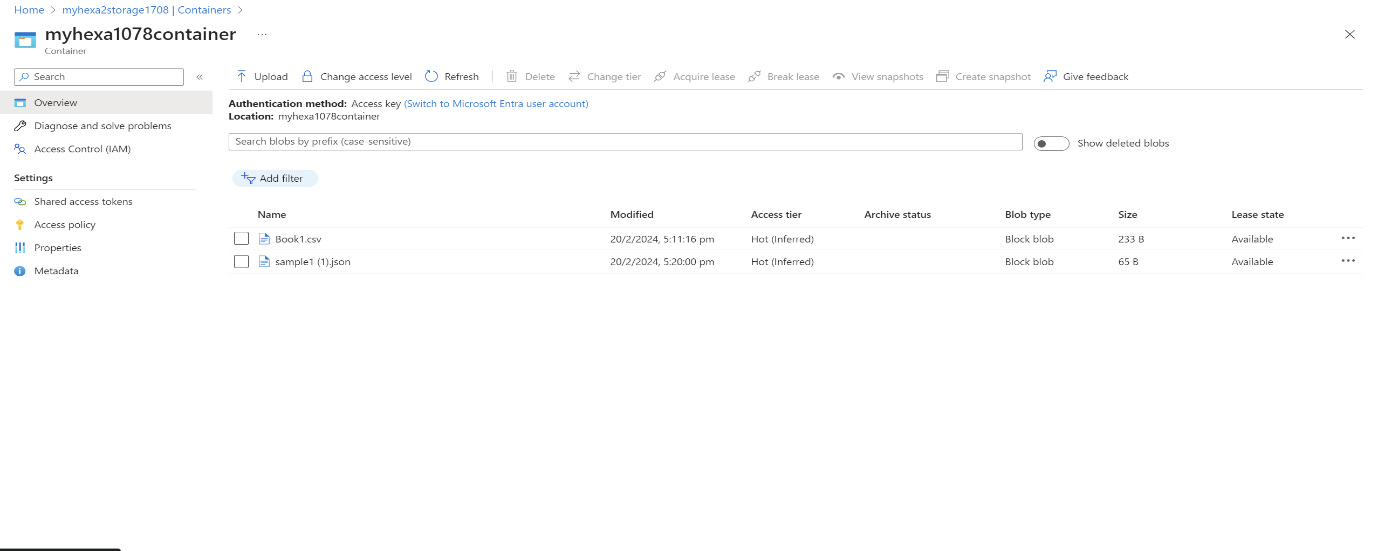
Here is the destination json file now let us check in storage account



This is my file in the souce



So here the data is copyed into new container from source container



Notes:

