Azure Data Factory – Copying Data from Rest API
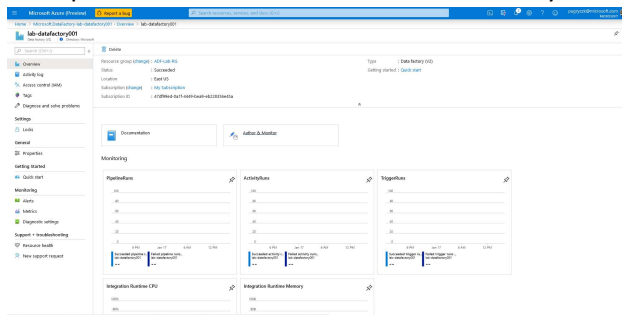
Overview
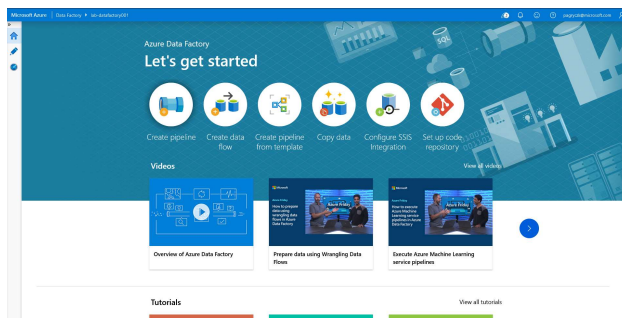
Copying Data from Rest API

Pulling Data from an API with no Authentication
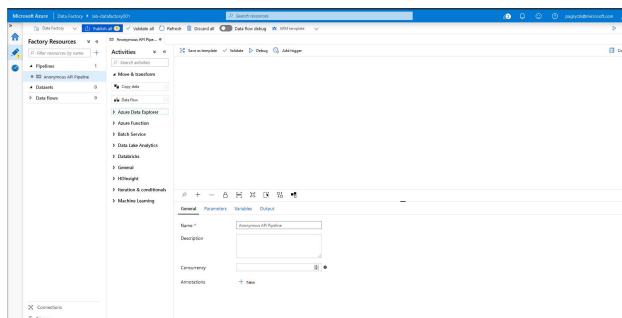
Create a Data Factory Pipeline

Navigate to the Data Factory that you created in the previous step, and click on "Author & Monitor," this will open a new window with the Data Factory UI
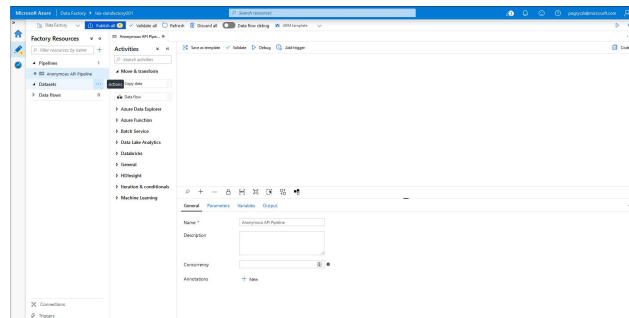

In the Data Factory UI, click on "Create pipeline"

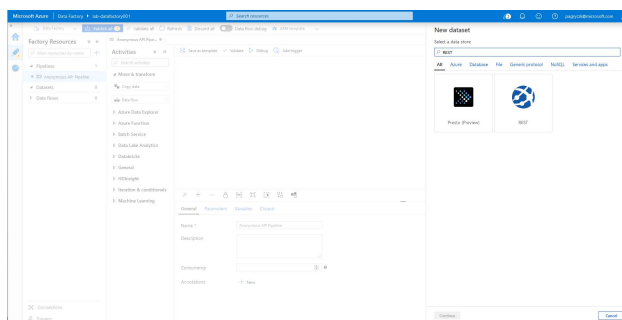Name your Pipeline "Anonymous API Pipeline"
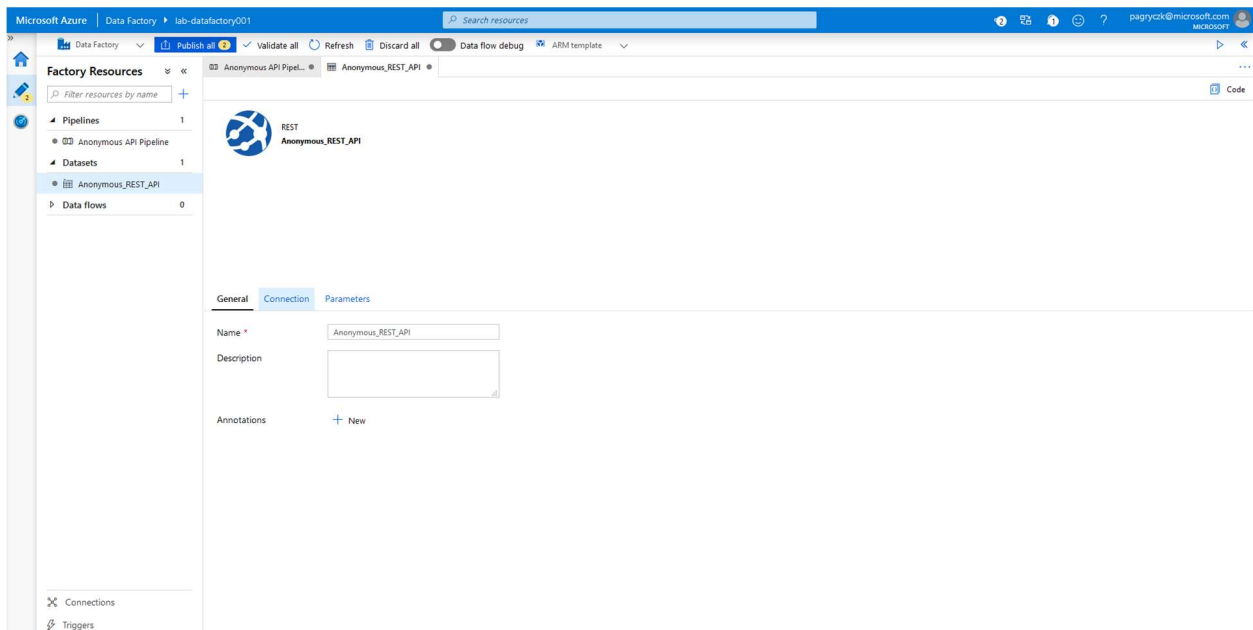
Create a Dataset

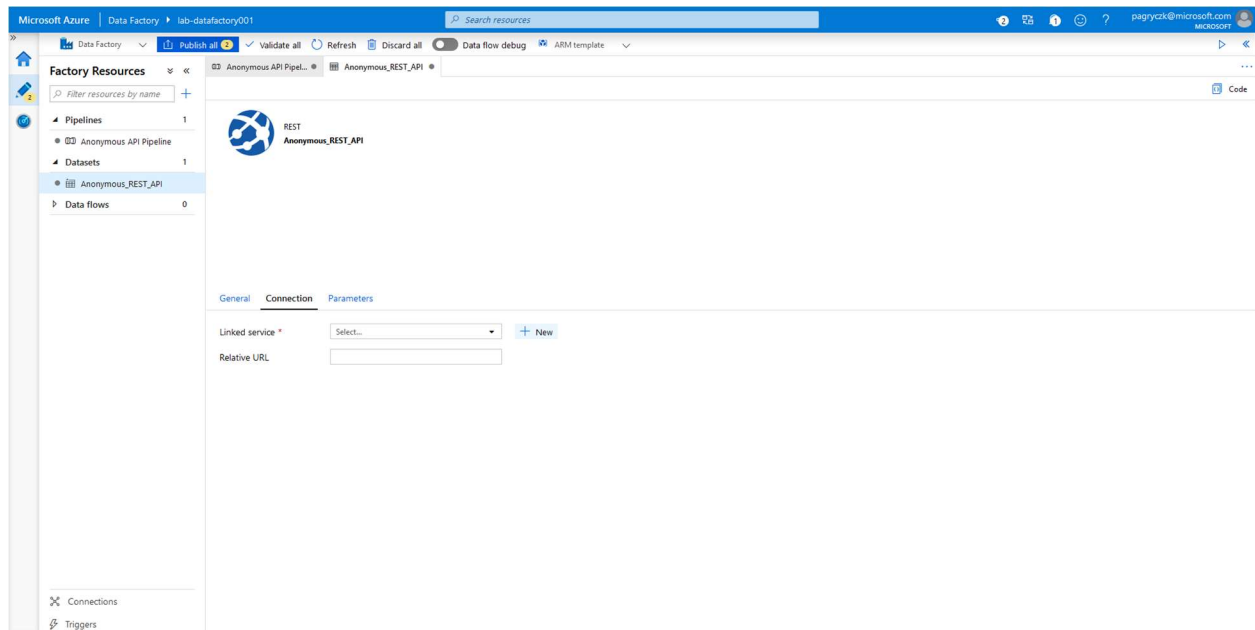Create a Dataset by clicking on the Actions Selector, and clicking New Dataset

Select REST as your data store type and continue

Rename the dataset to "Anonymous_REST_API" and then click over to the Connection Tab
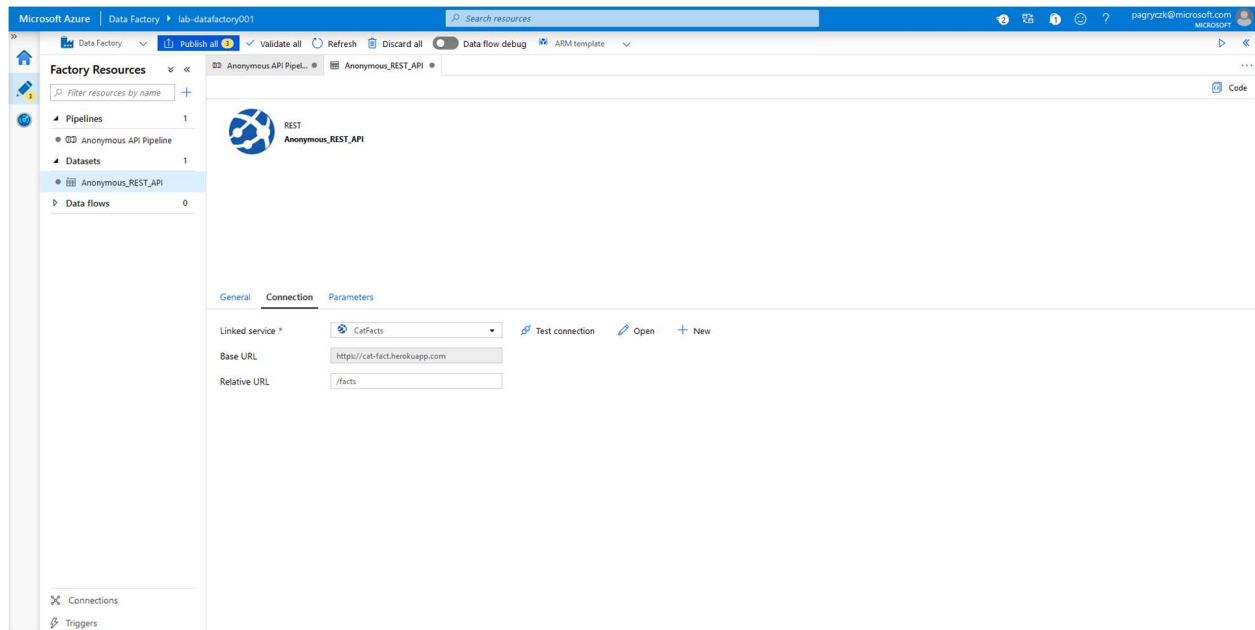

In the 'Connection' tab, for 'Linked service' select 'New'

Name the Linked service 'CatFacts', select the 'Anonymous' authentication type, and enter 'https://cat-fact.herokuapp.com' as the Base URL. At the bottom right of the page, click 'Test connection,' then click 'Create.'
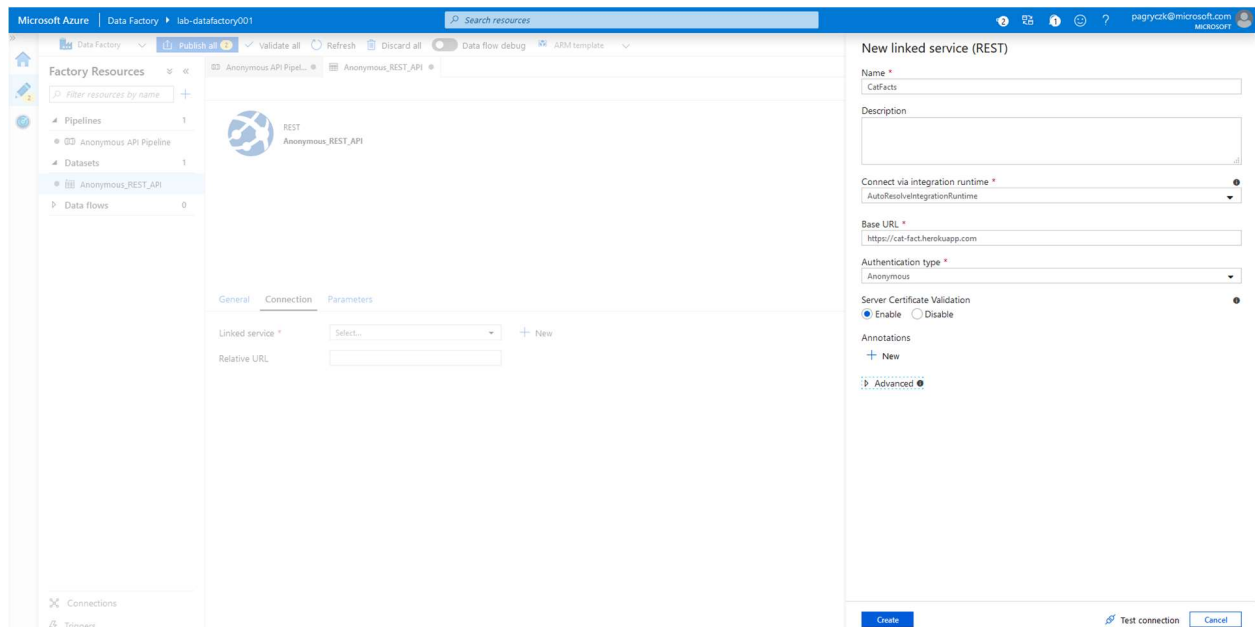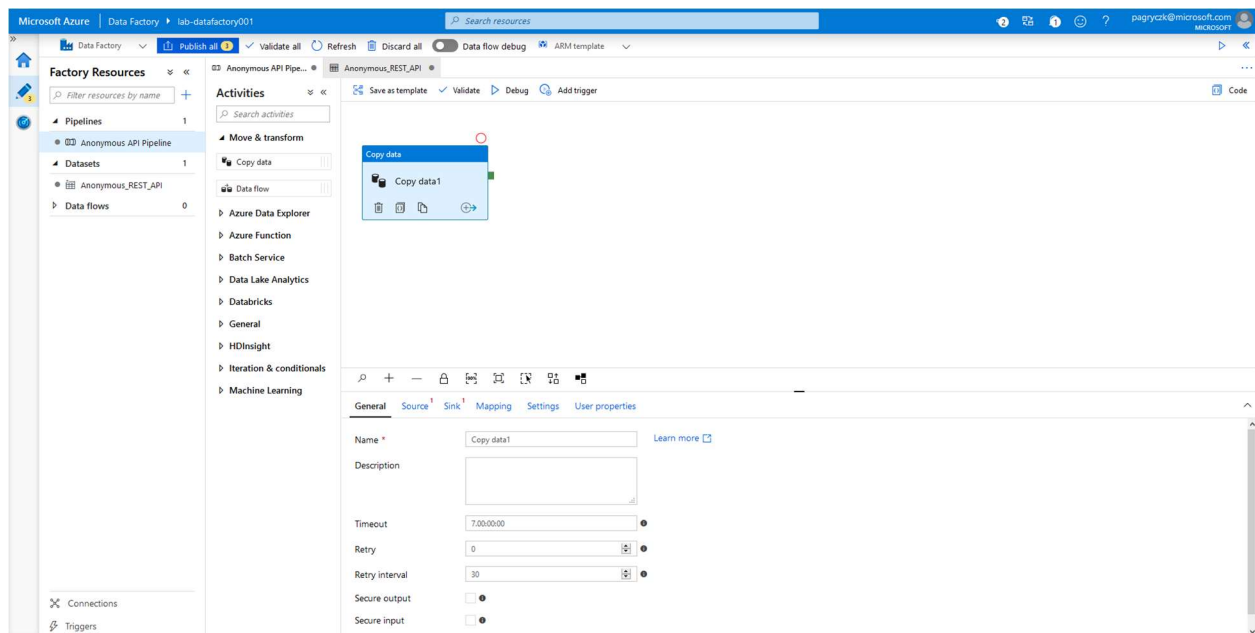

Back on the 'Anonymou_REST_API' dataset 'Connection' tab, enter '/facts' into the 'Relative URL' field.
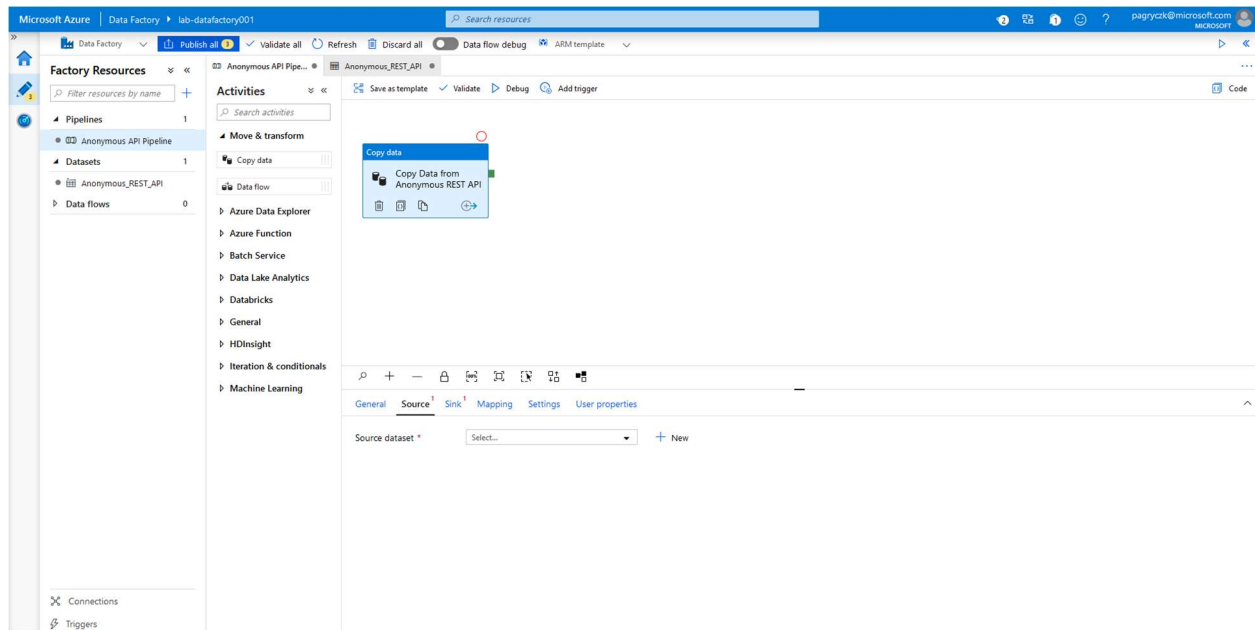
Then return to your Anonymous API Pipeline

## Create an Activity in your Pipeline

Within Activities, expand the 'Move & transform' drop down, and click and drag 'Copy data' into your pipeline.
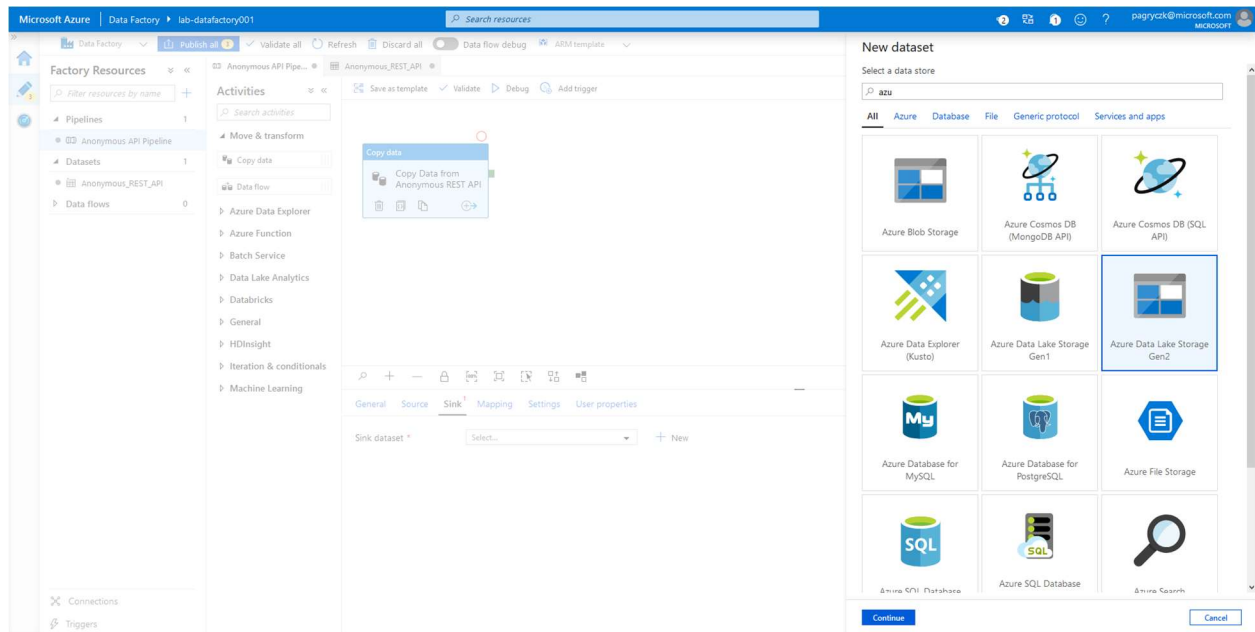

Rename your copy data activity to 'Copy Data from Anonymous REST API' and click through to the Source tab.
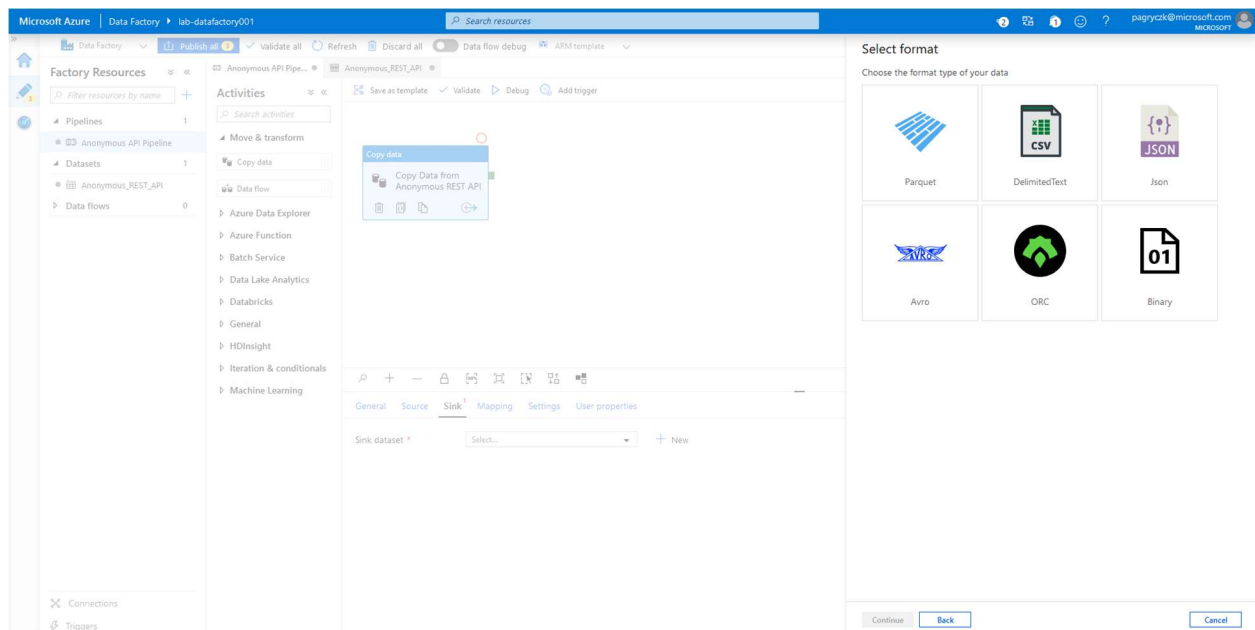
In the source tab, from the Select dropdown, pick your 'Anonymous_REST_API' and click 'Preview data' to confirm that your connection works.
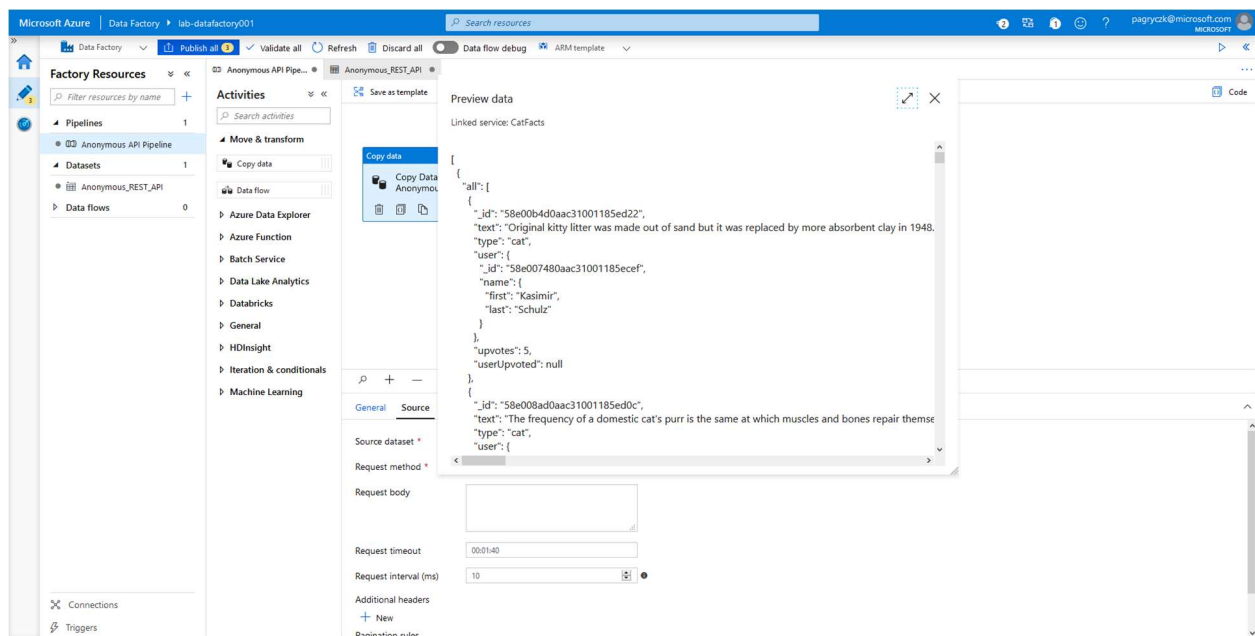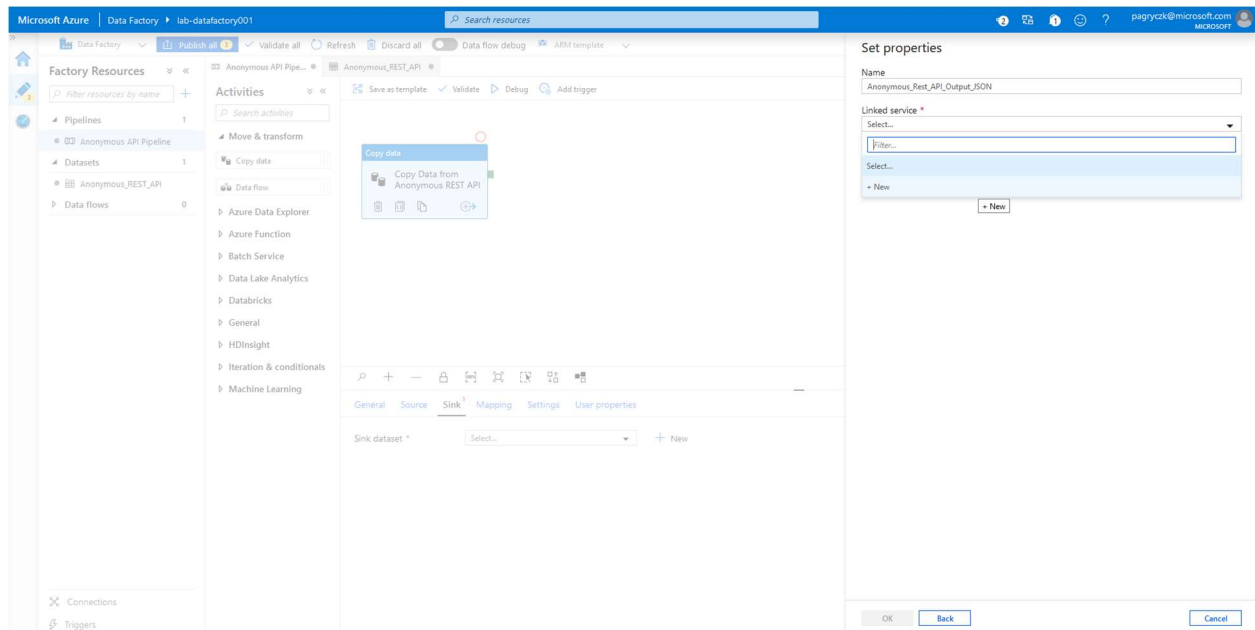

After confirming that you can preview the data, switch over to the 'Sink' tab and click 'New' and select Azure Data Lake Storage Gen2

Then continue and select JSON.

For the Properties, name the dataset 'Anonymous_Rest_API_Output_JSON' and create a new linked service.

For the linked service, select your subscription and specify the name of the storage account you created during the previous setup lab module.

Then press Create and proceed to Set Properties. Enter 'anonymousapi' as your 'File System', 'dev' as your 'Directory' and for Import Schema select 'None'
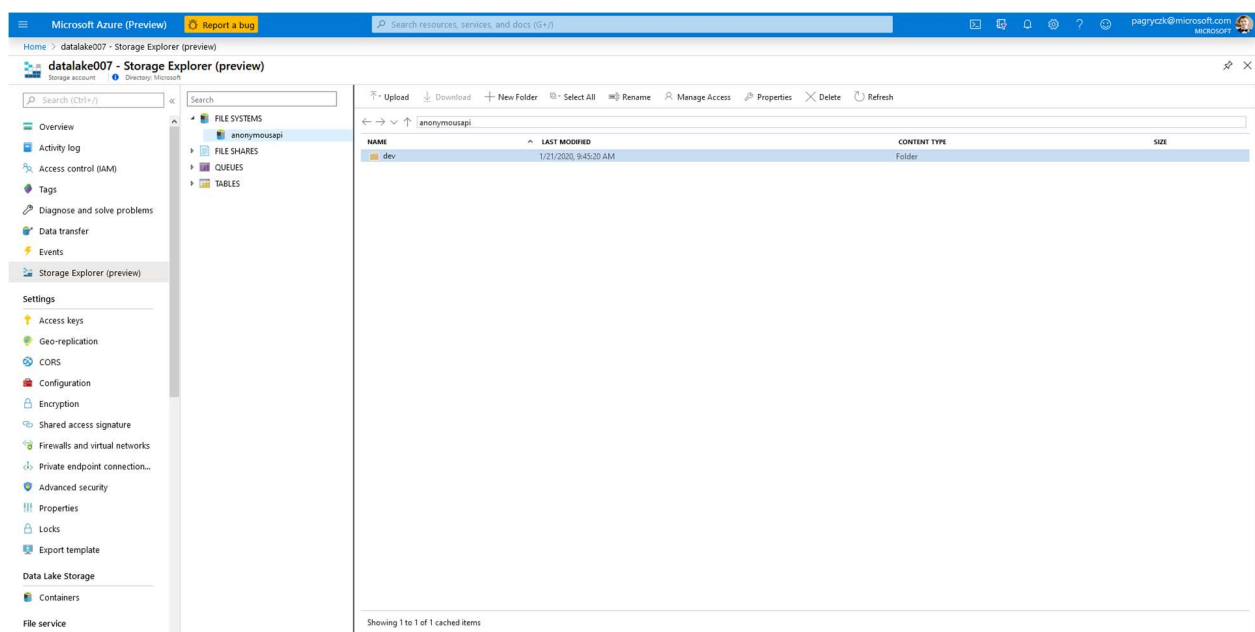
Next, go to your storage account in the portal, and in the left hand panel click on Storage Explorer (preview)
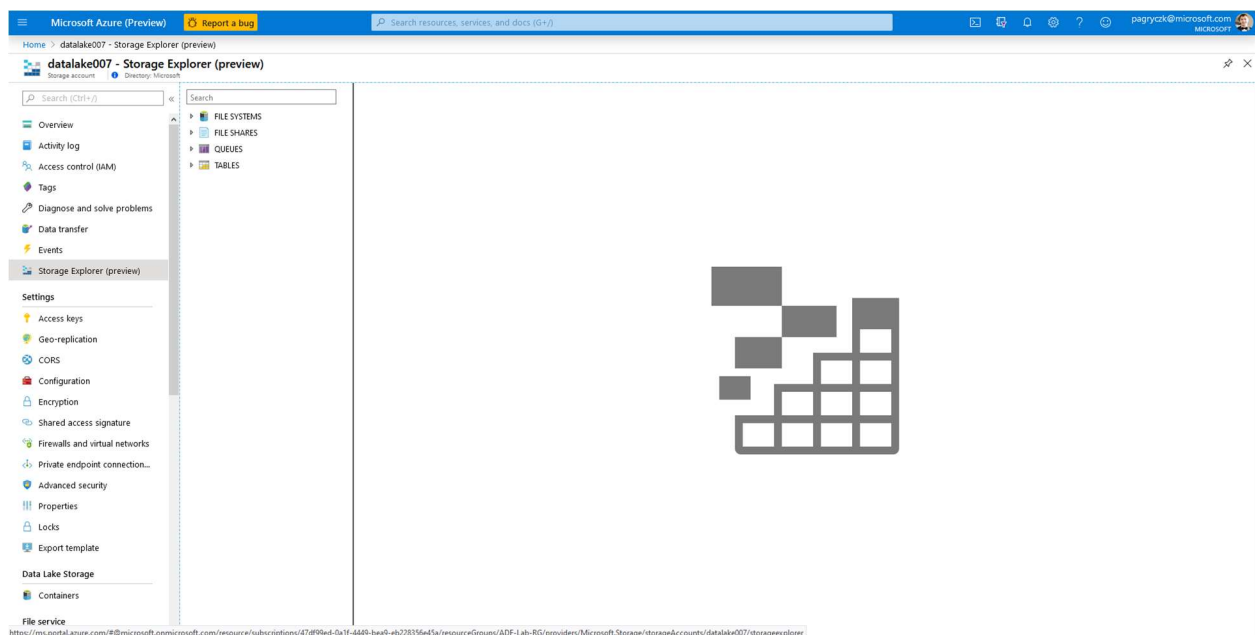
Then right click on File Systems and click 'Create file system'
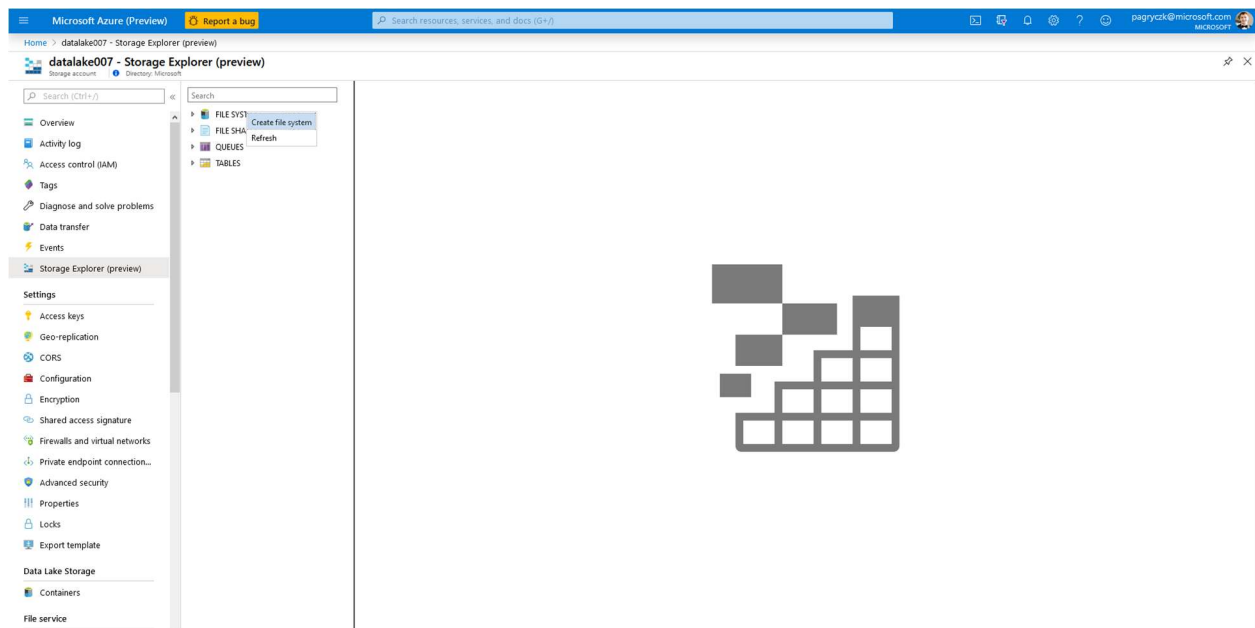
Name your file system 'anonymousapi' Then navigate to the anonymousapi file system in the storage explorer and create a New Folder named 'dev'

Next, you can return to your pipeline, click Publish All to save your Pipelines and Datasets, and click Debug to validate if your pipeline is able to run successfully.

After the pipeline run succeeds, you can return to your Storage Accounts Storage Explorer and see the file added to your File System.
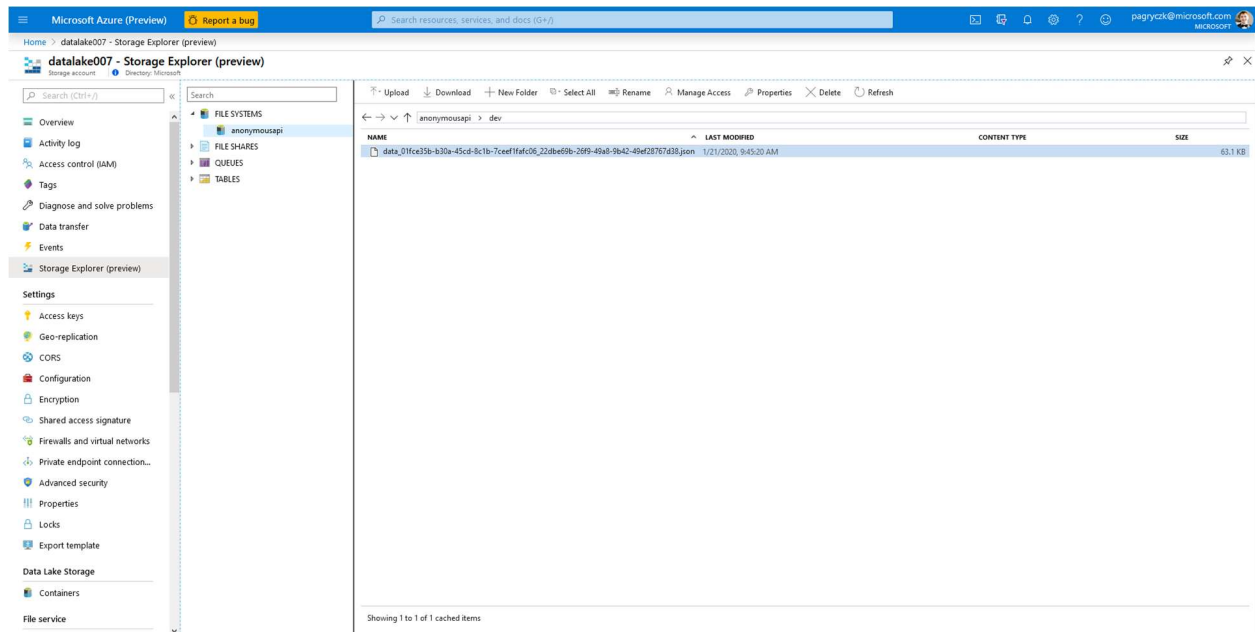
Challenge: Pulling Data from an API with Anonymous API Key Authentication

Repeat the previous steps with appropriate new values to set up a new pipeline and dataset with 'https://ecfsapi.fcc.gov/filings' as your URL, and add an api_key and q parameter to your new dataset. Your lab proctors will provide you with an API key.