**ADF Mini Project**

**Pre Requirements:**

* Azure Data Factory
* Azure Data Lake Storage
* Azure SQL Data Base

**Things need to be done before starting work on the Azure data factory**

**1. Azure Data Lake Storage (ADLS) Part:**

1. Create an ADLS Account
2. Inside that create a container name “data”
3. Inside the “data” container create two folders

* Input
* Output

1. Inside the Input folder and output folder upload the files given in the “Raw Data” folder

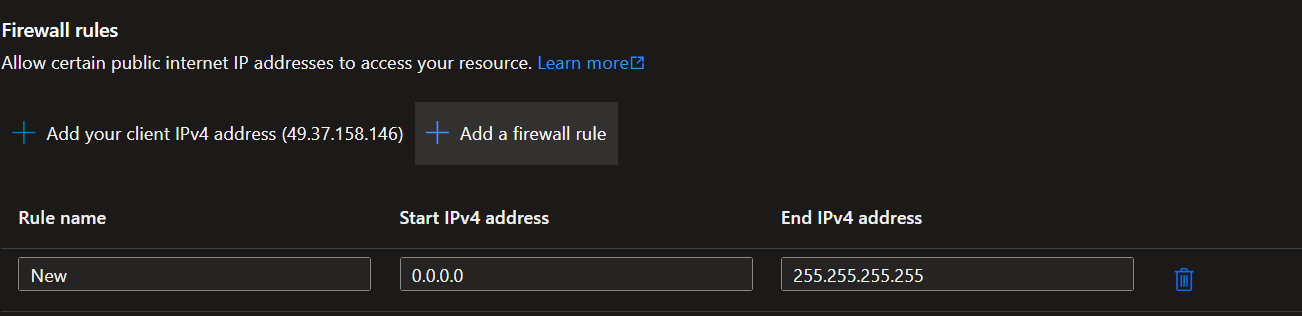
**2. Azure SQL Database Part:**

1. Create a SQL Database account
2. Set up the firewall for SQL DB Server so that ADF can access the tables

By mentioning the below :

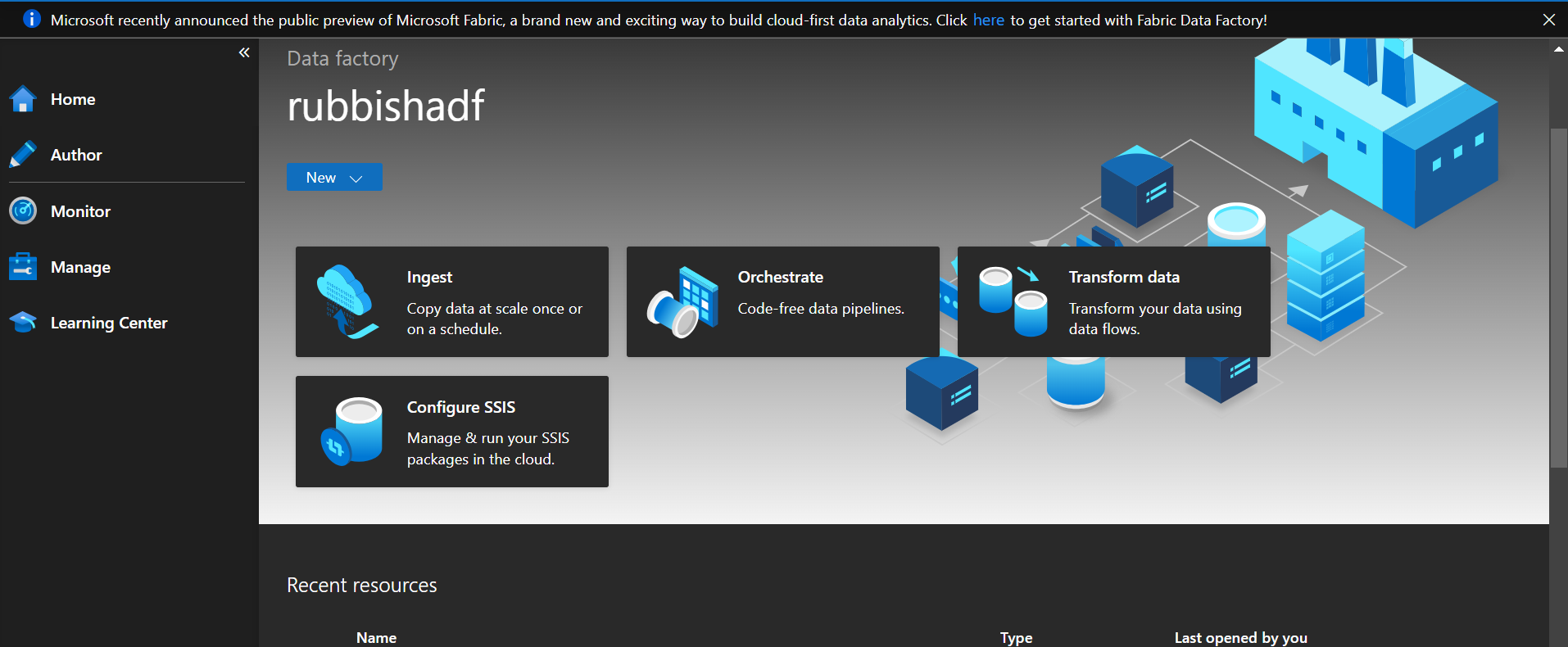
Click on “Set server Firewall” and inside that you will be able to see “Firewall rules”

Create a new rule like in the screenshot and save it



**3. Azure Data Factory Part :**

1. Create an Azure Data Factory in azure resource group
2. Open the ADF You will be able to see the below screen

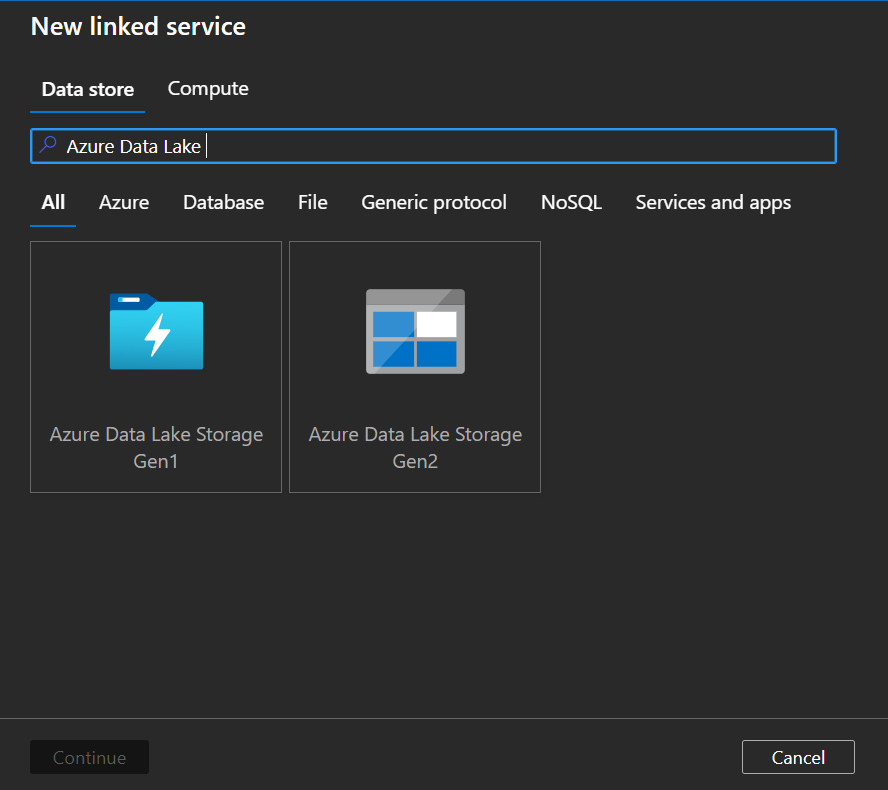


**Creating the Linked services:**

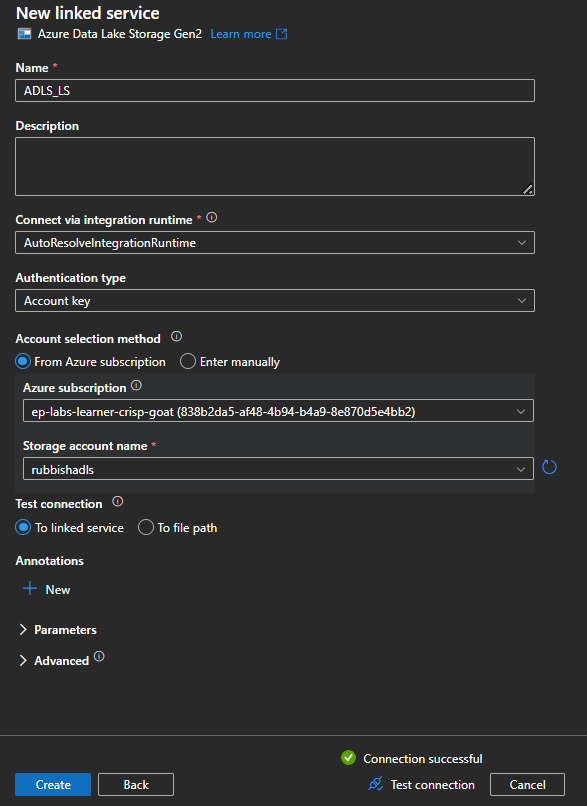
* 1. Here in the Manage tab you will be able to see Linked Service there we have to create 2 linked services
  2. One for ADF- ADLS
  3. One for ADF-SQL DB

**What is a linked service**: A linked service is like a connection between two things, allowing them to talk to each other. It helps data move between different systems or programs easily.

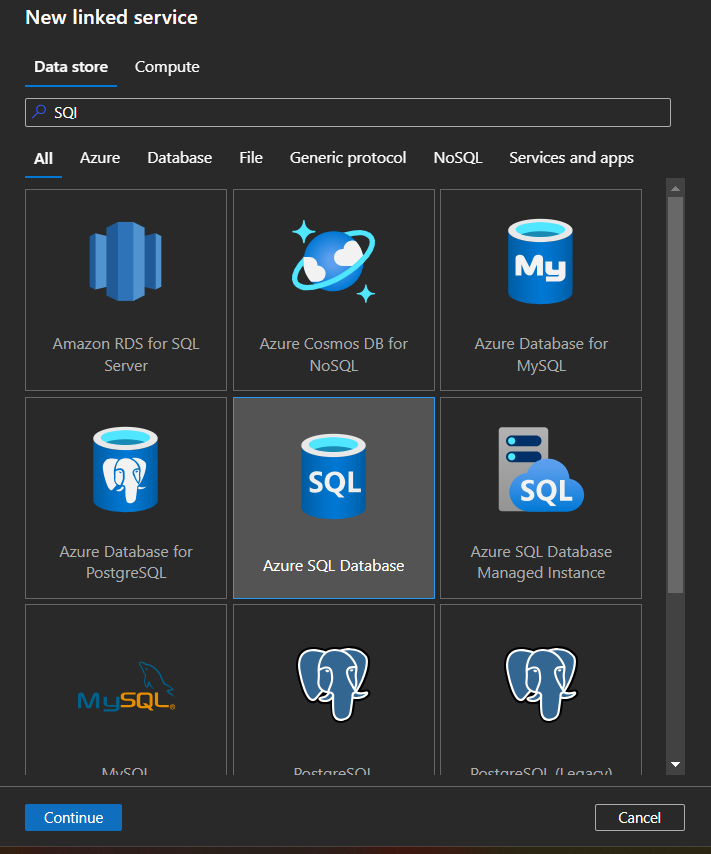
* 1. To create a Linked service form ADF to ADLS click on ‘+ New ’ icon on top search for Azure Data Lake there you will see ADLS gen2 account and then select it

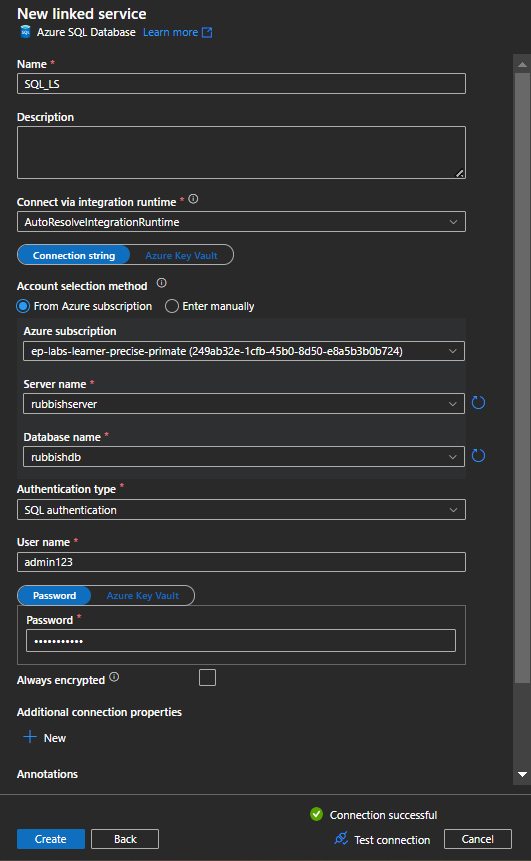


* 1. Give the name to the Linked service in our case we have to give “ADLS\_LS” and select the Subscription and Storage account name. Once given all the things test the connection to check if the Linked Service is working properly or not y



* 1. Same way create the Dataset for SQL DB also with the help of the below screen shots
  2. selecting the SQL DB below page will come up

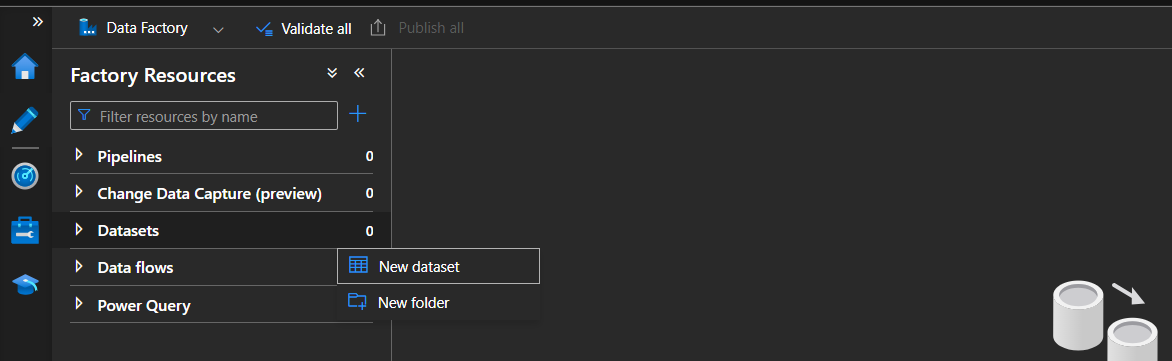




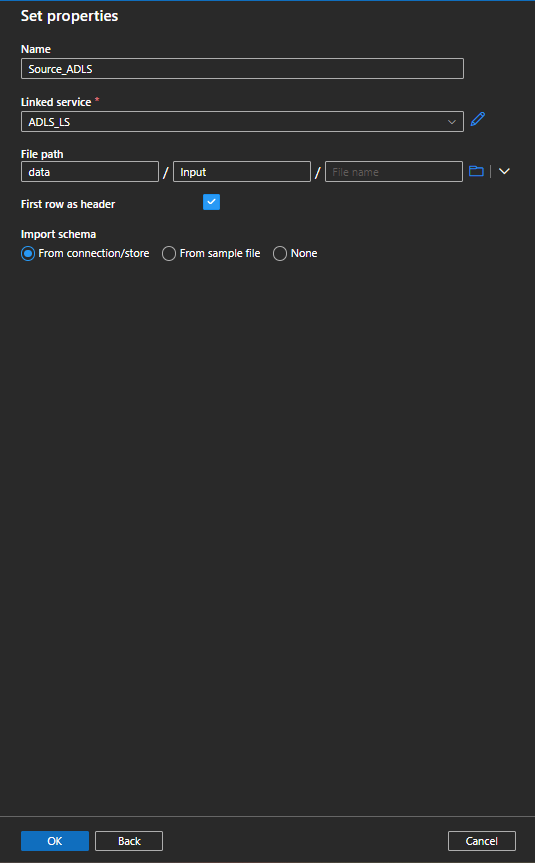
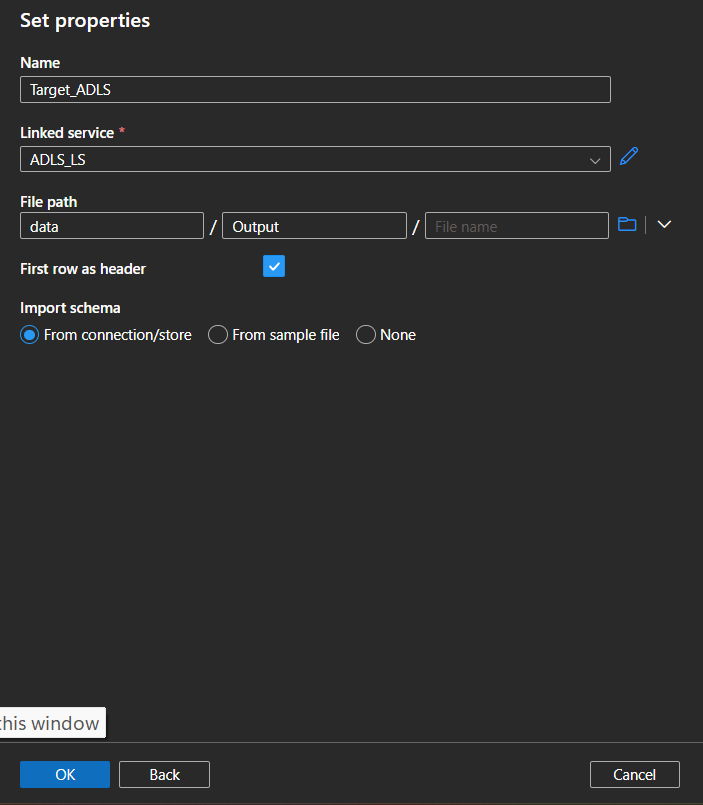
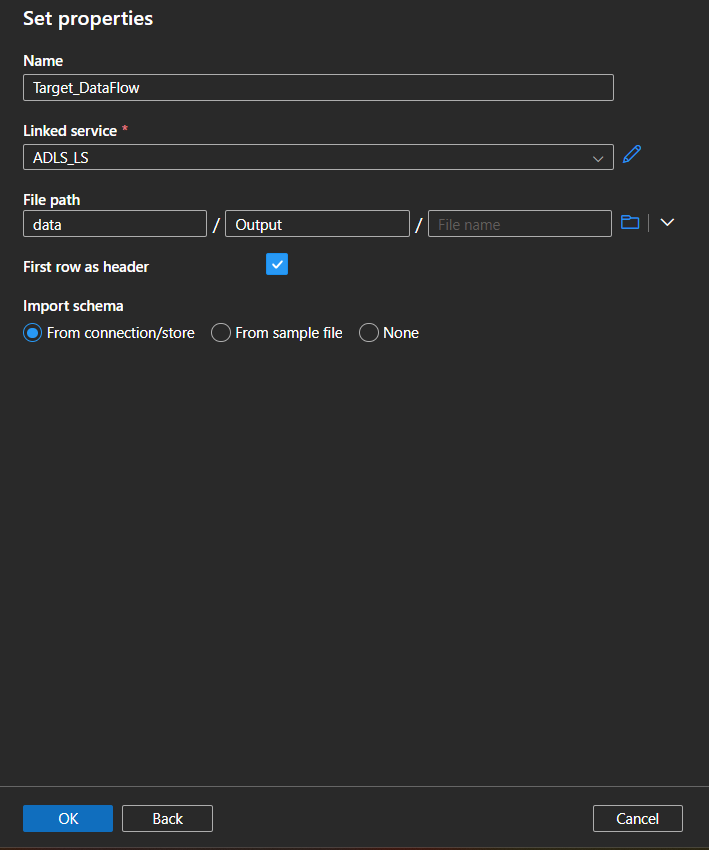
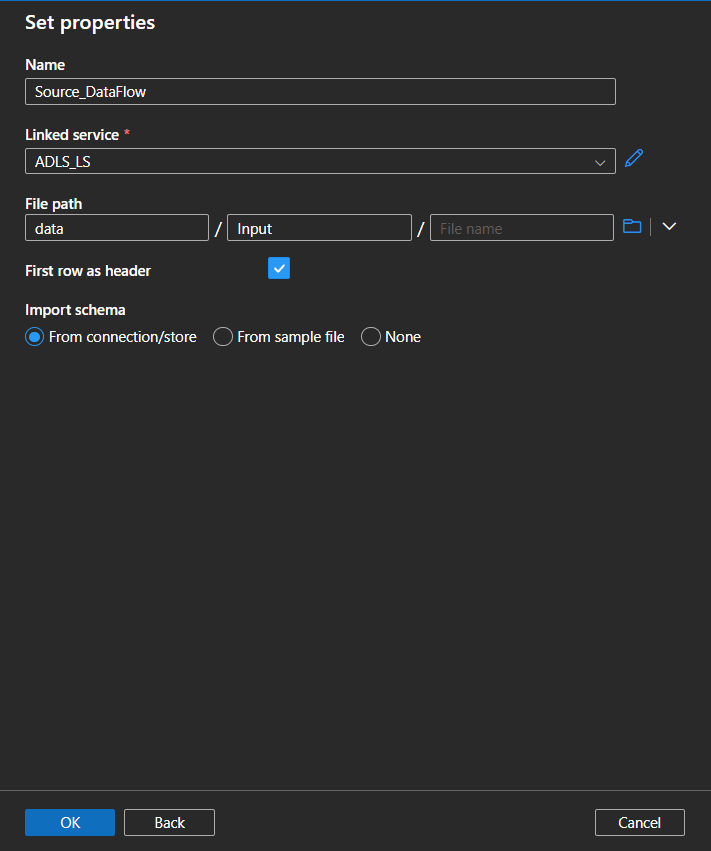
* 1. Here give the Linked service name as “SQL\_LS”, and select the subscription that you have next database name, authentication should be “SQL authentication”
  2. Give the “User name” and “Password” that you have mentioned while creating the database and give the
  3. Next check the connection

**Creating the Datasets:**

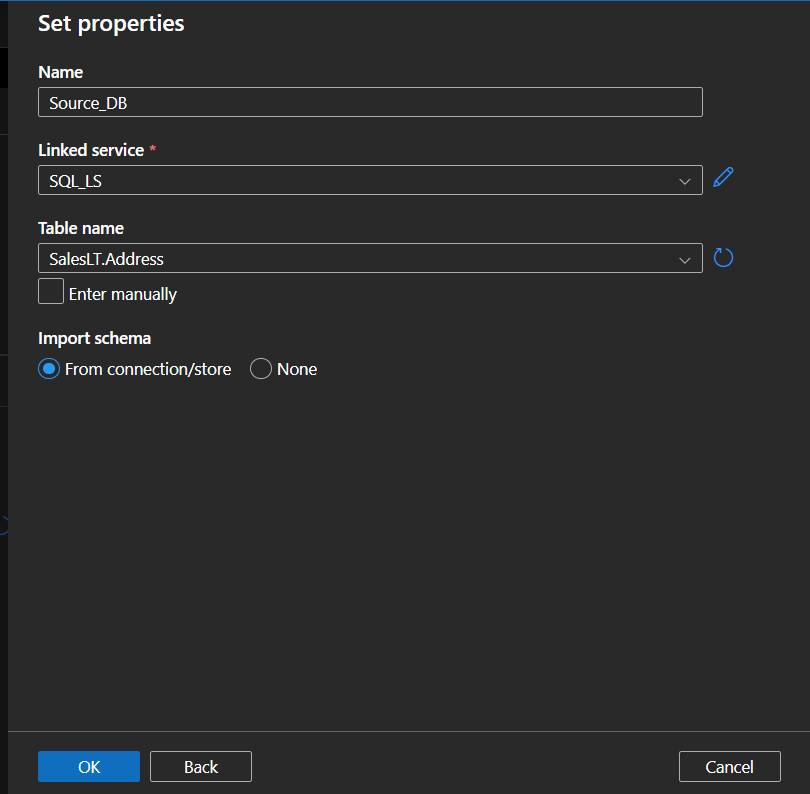
1. Creating of Source and Target Datasets are the same, so by clicking on New Dataset in the Author tab



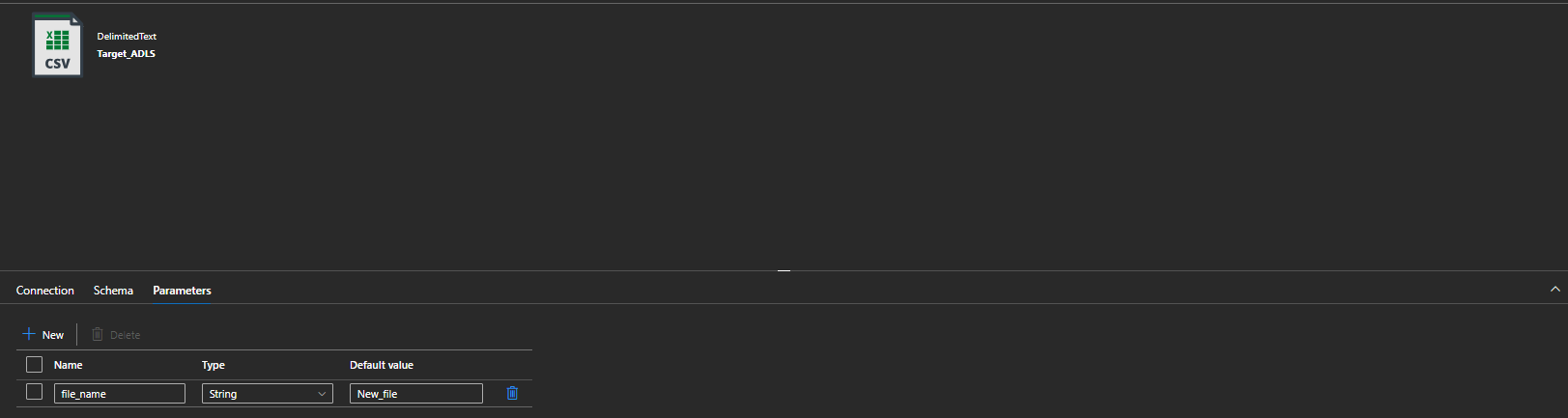
1. You will be prompted to the below slide Post that selects “Azure data lake gen2” and there select “DelimitedText”s file and give the names and parameters as mentioned below.

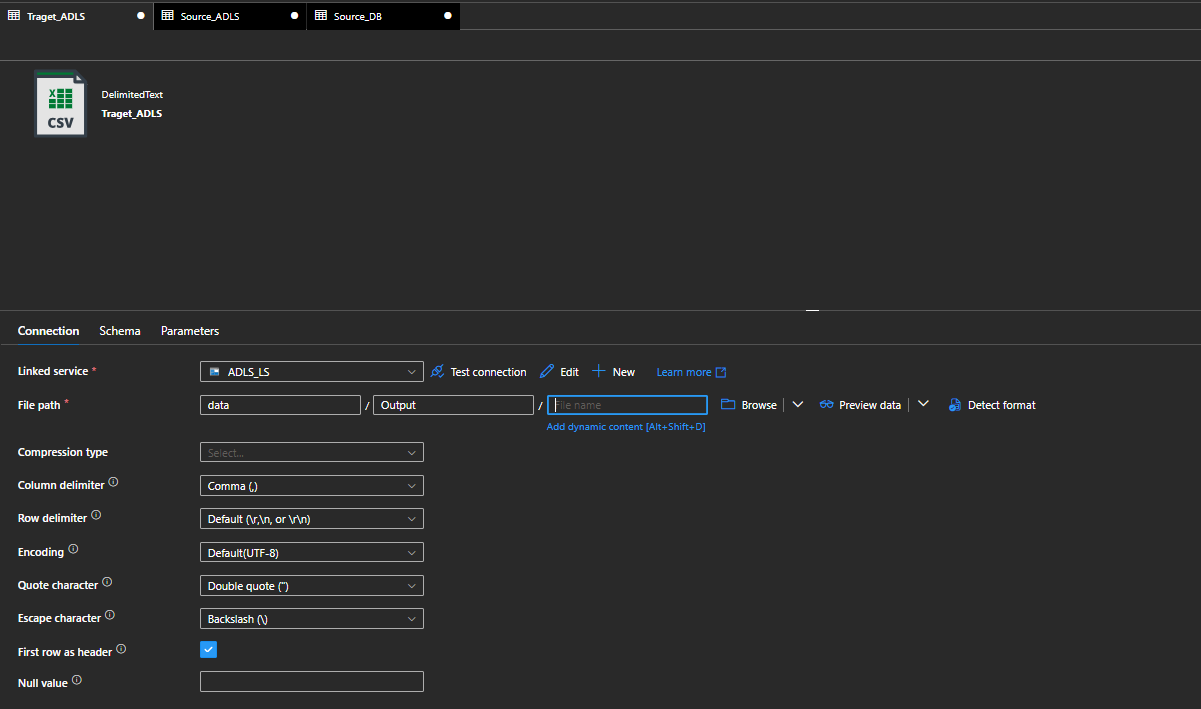
1. In the same way create a dataset for SQL DB by the below images



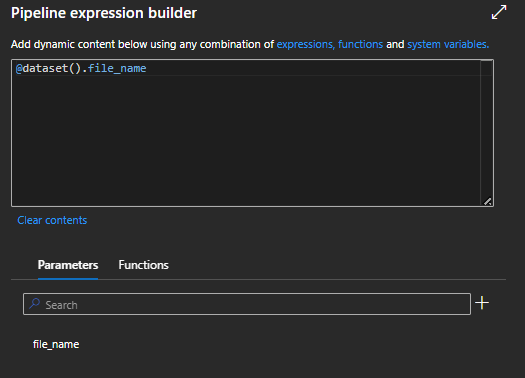
1. you are done with the creation of Datasets select the “Target\_ADLS” and click on “parameters” tab. Then declare a parameter as stated below



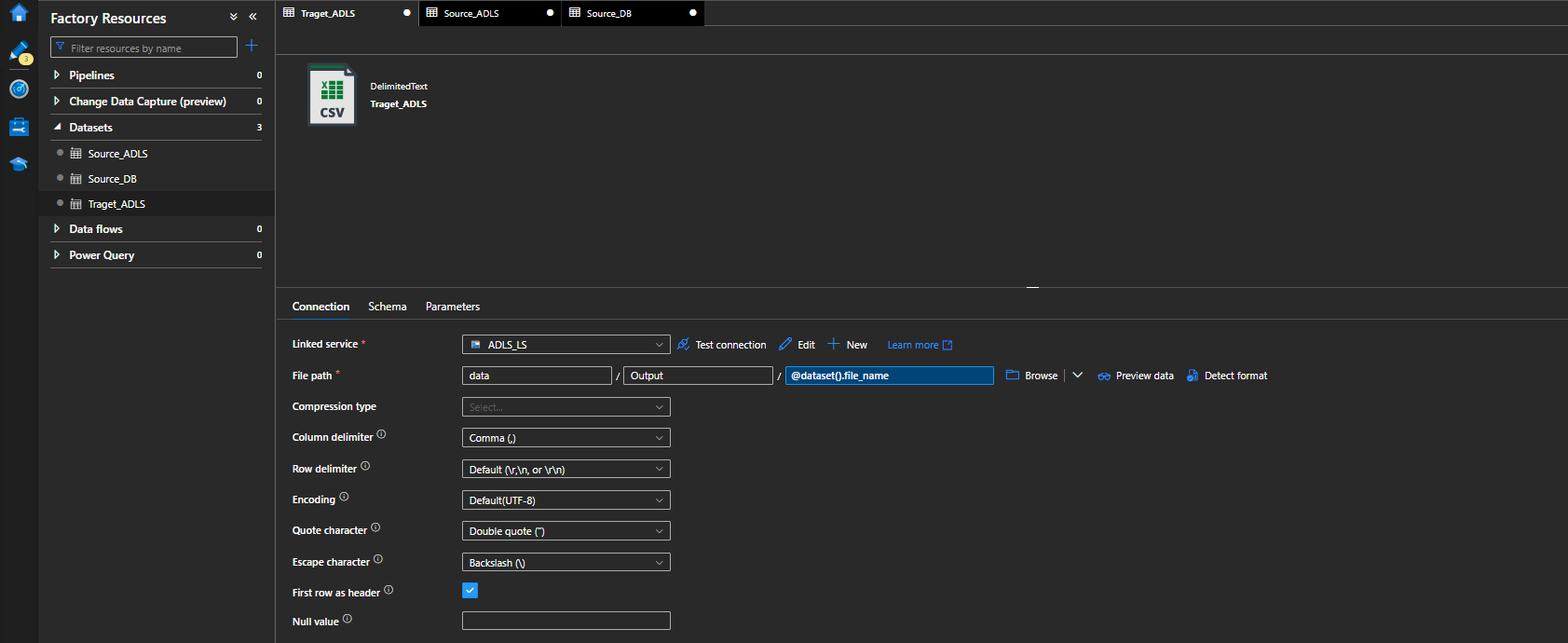
1. Once that is done select the “Connection” tab again and click on the space in the “filename“ input box. Then click on “add dynamic content”.



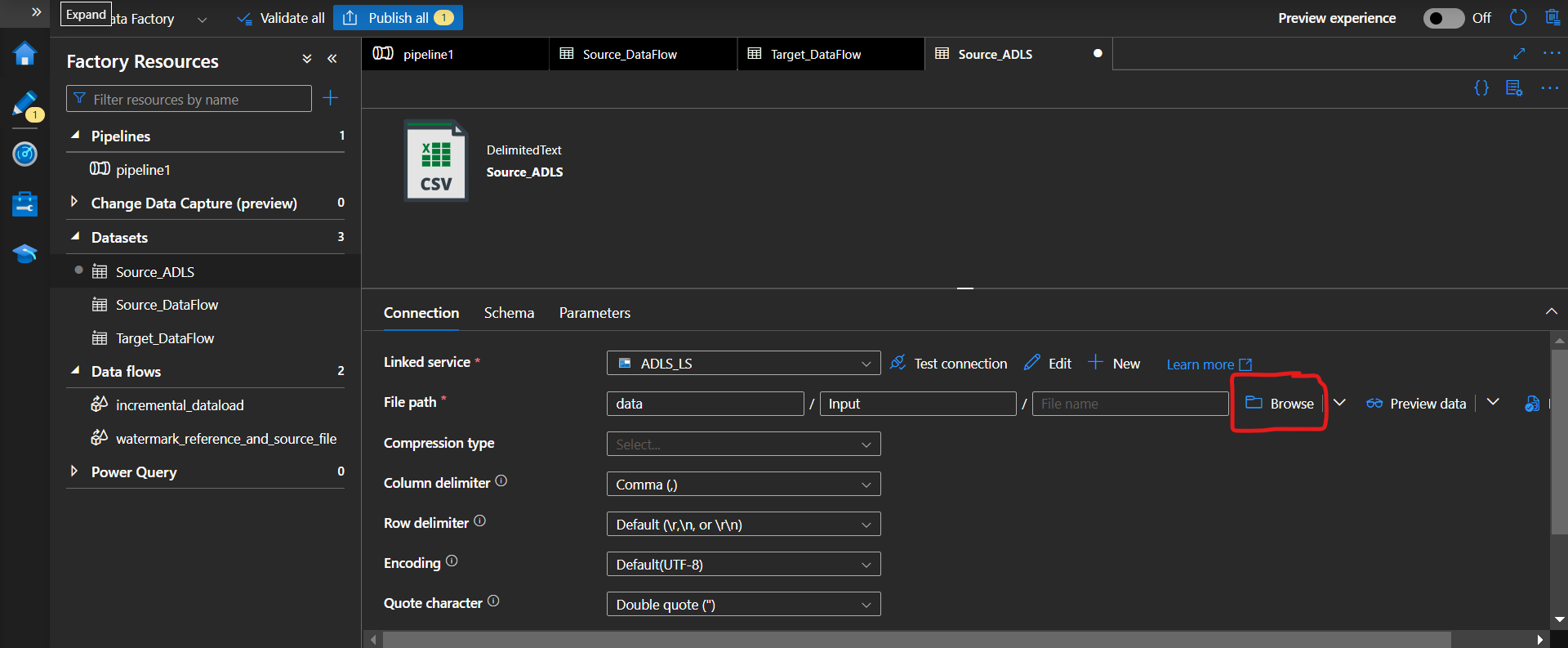
1. Inside the popup select the “file\_name ” form Parameters tab and click on “OK”



1. You will see the below page for “Target\_ADLS”



1. Once the target dataset is done for the remaining 4 out of 5 datasets i.e. for” Source\_ADLS”,” Source\_DataFlow”,” Source\_DataFlow\_ADLSss” and ”Target\_DataFlow” click on the Browse button mentioned in the image



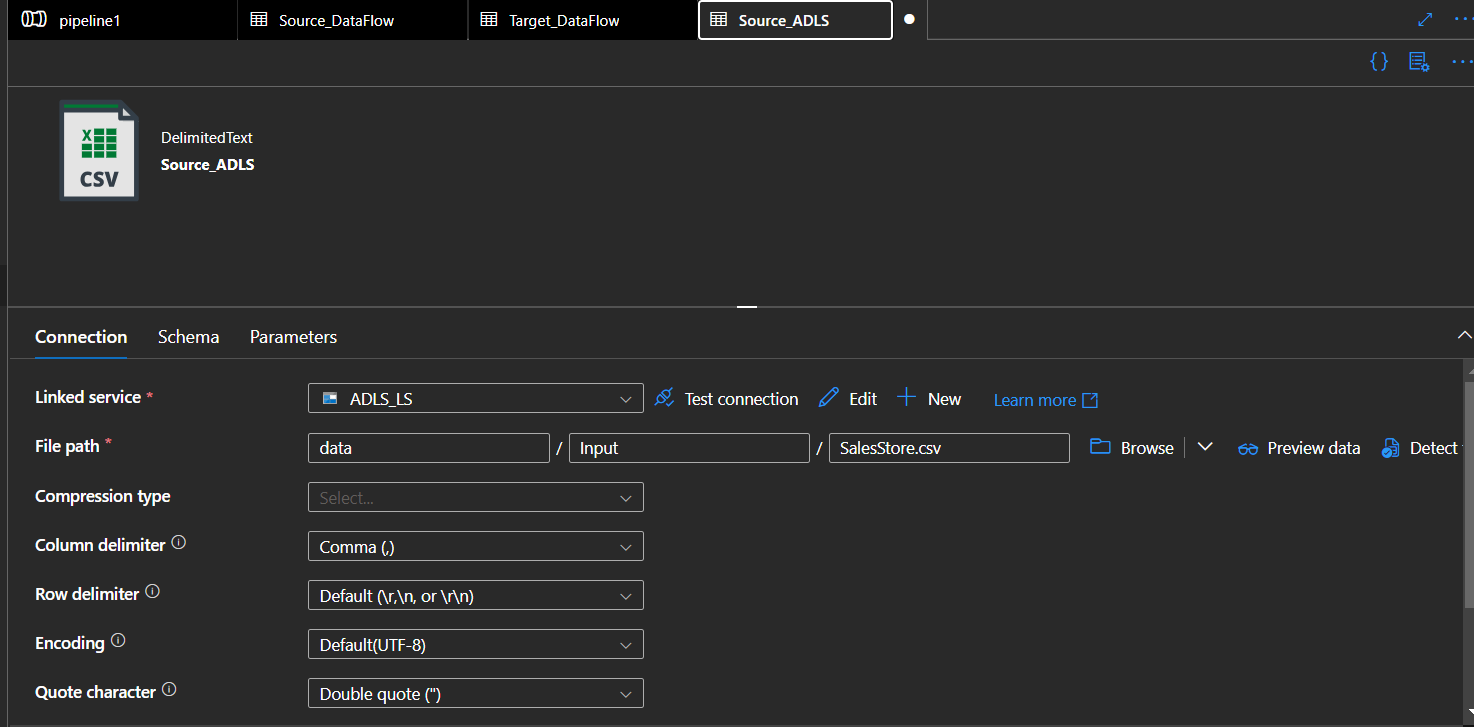
You will be directed to the root folder select data/Input/SalesStore.csv and click on ok

Follow the same steps for the remaining four datasets also

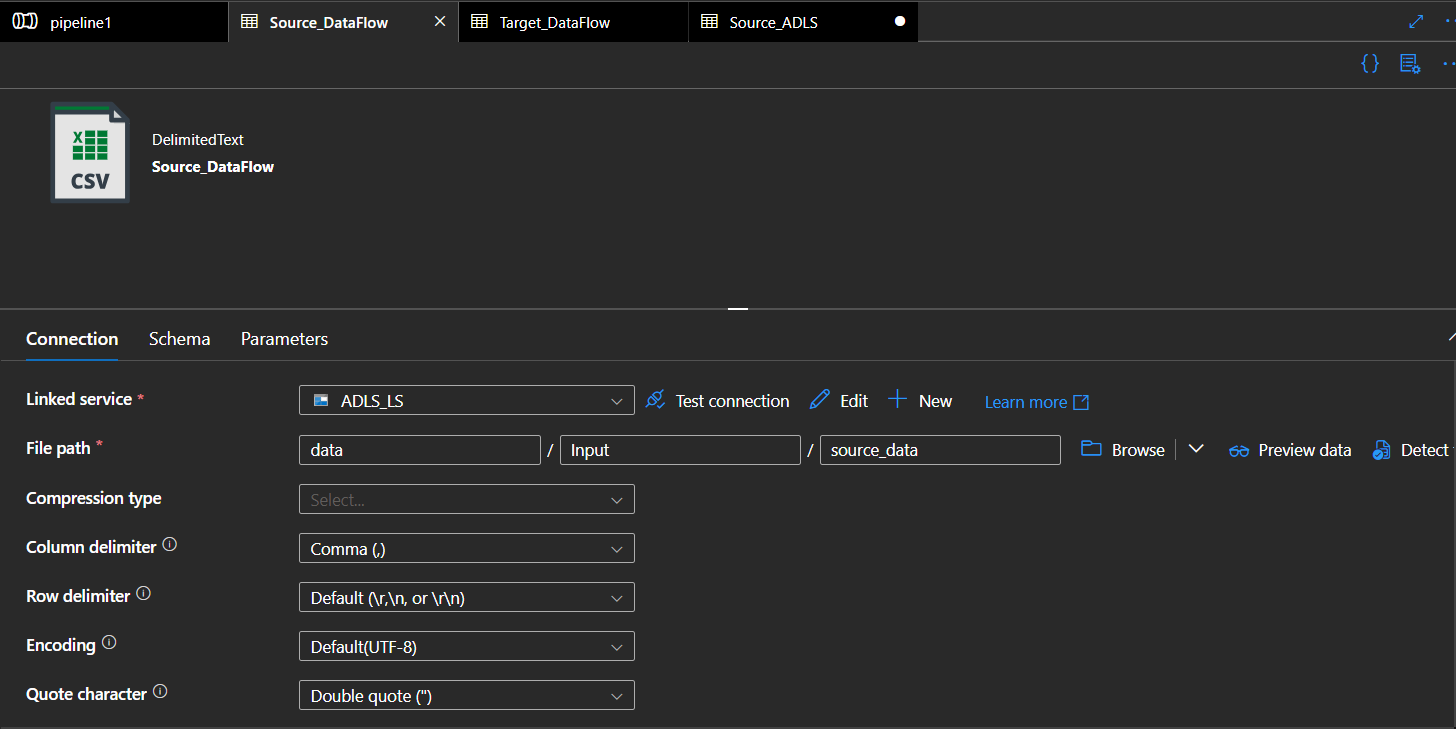
* Source\_ADLS : data/Input/
* Source\_DataFlow: data/Input/ source\_data
* Target\_DataFlow: data/Output/ New\_file\_Dataflow
* Source\_DataFlow\_ADLS : data/Input/SalesStore.csv

1. Once that is done you will have the below values for all three data sets attached over the image

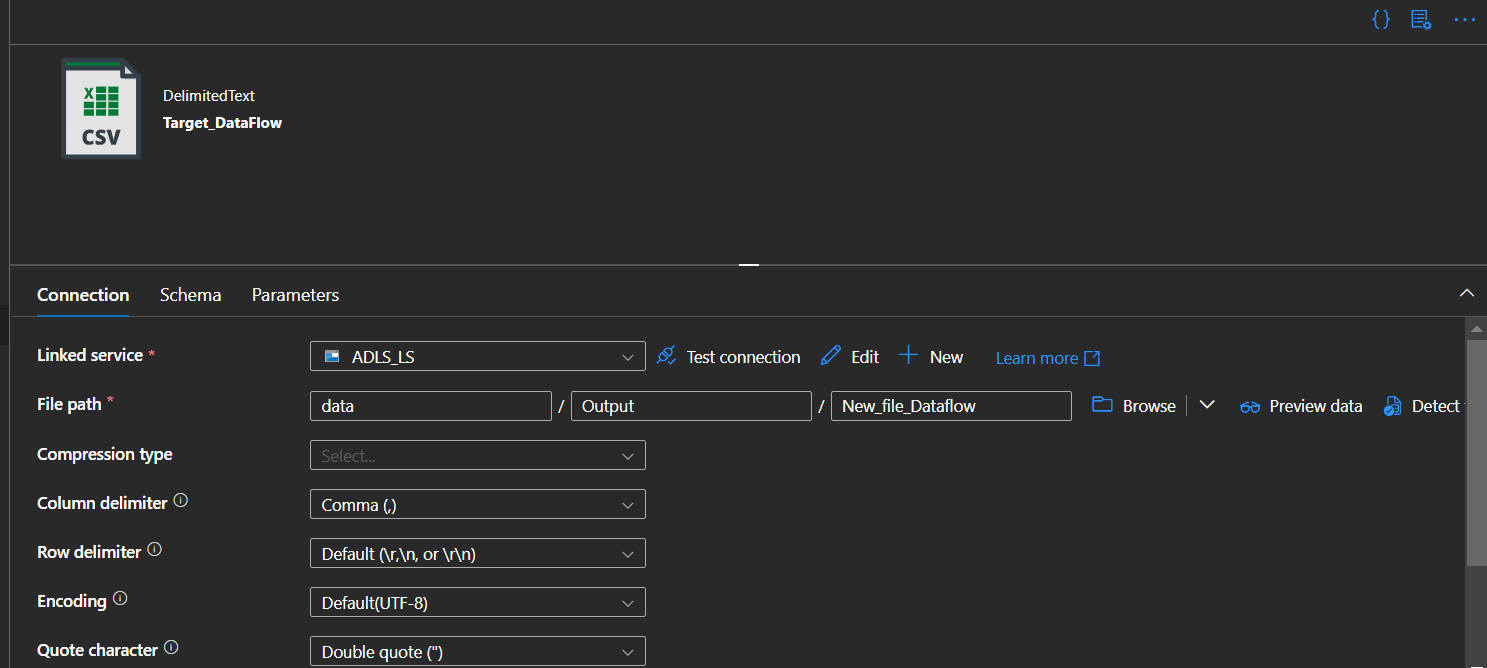
**Source\_ADLS:**



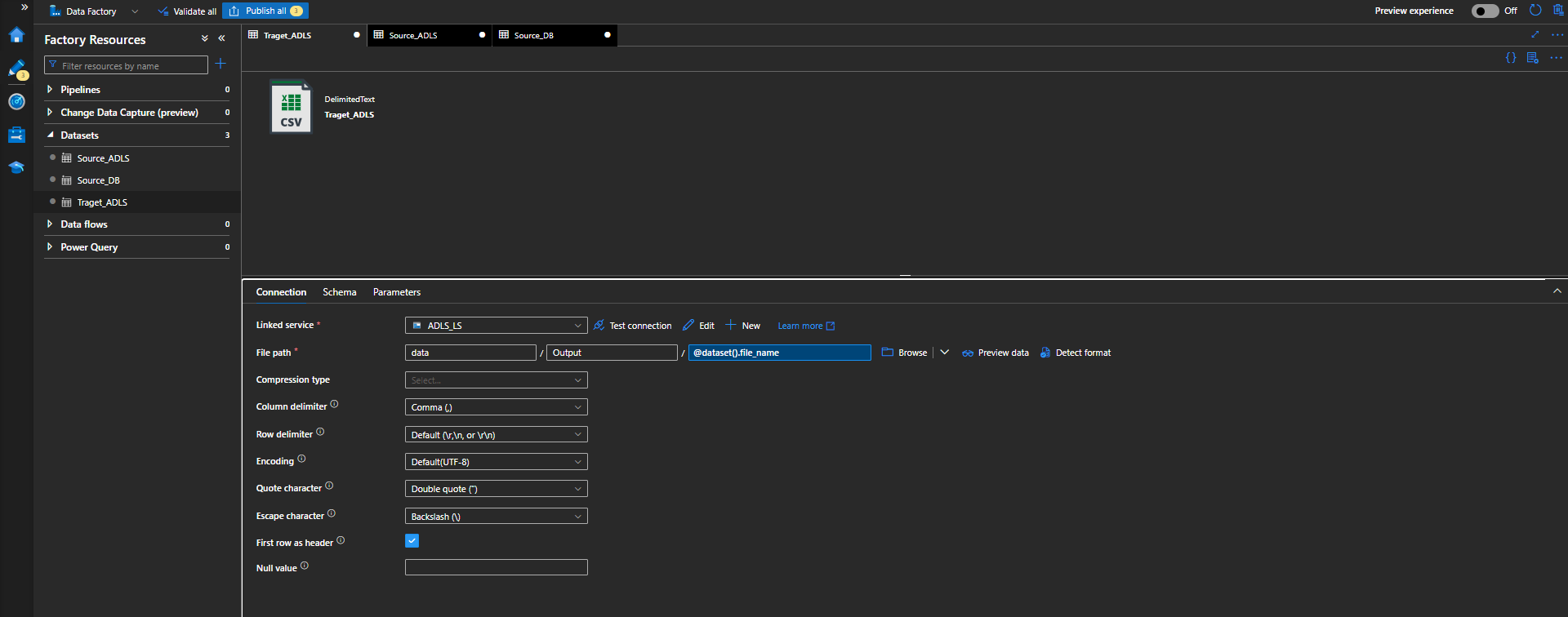
**Source\_DataFlow:**



**Target\_DataFlow:**

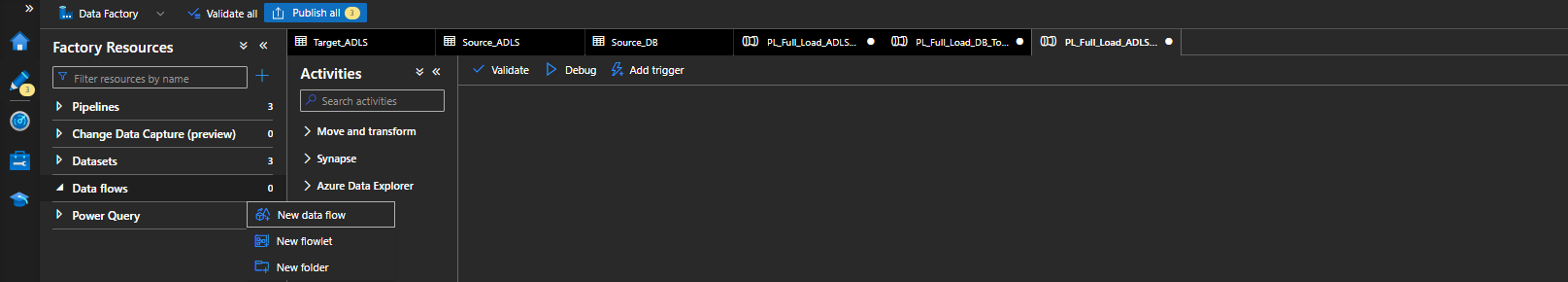


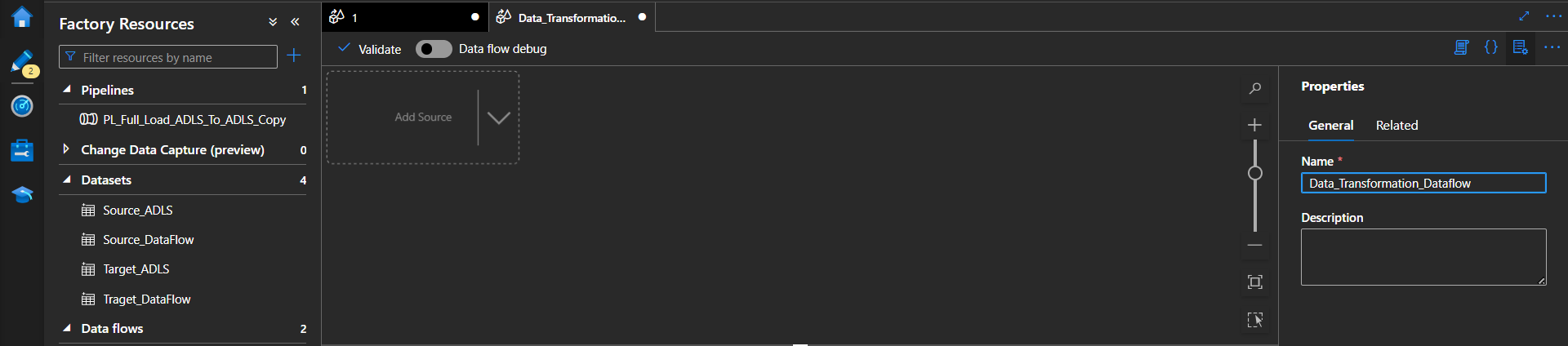
1. click on publish on the top to save all 5 datasets which was



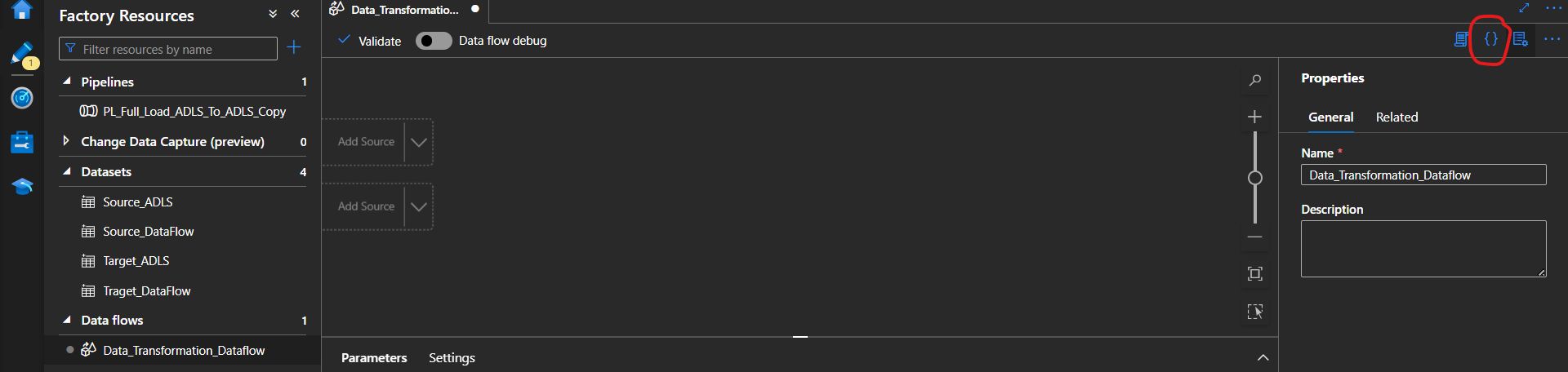
**Creating Dataflows:**

1. Click the “New Datflow” and rename the Dataflow with the name “Data\_Transformation\_Dataflow”



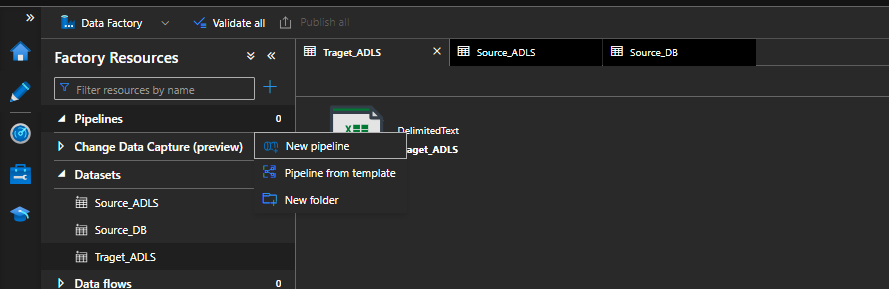


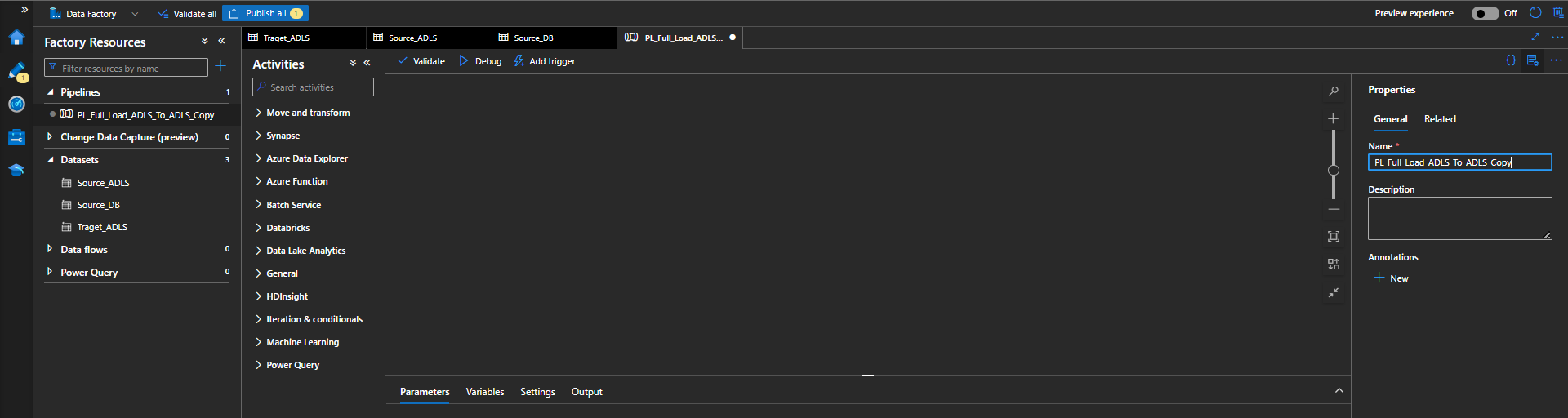
1. Once that is done click on the “{}“on top of the “name” and copy the script available in the “dataflow” Json of the pipeline folder attached in the zip file

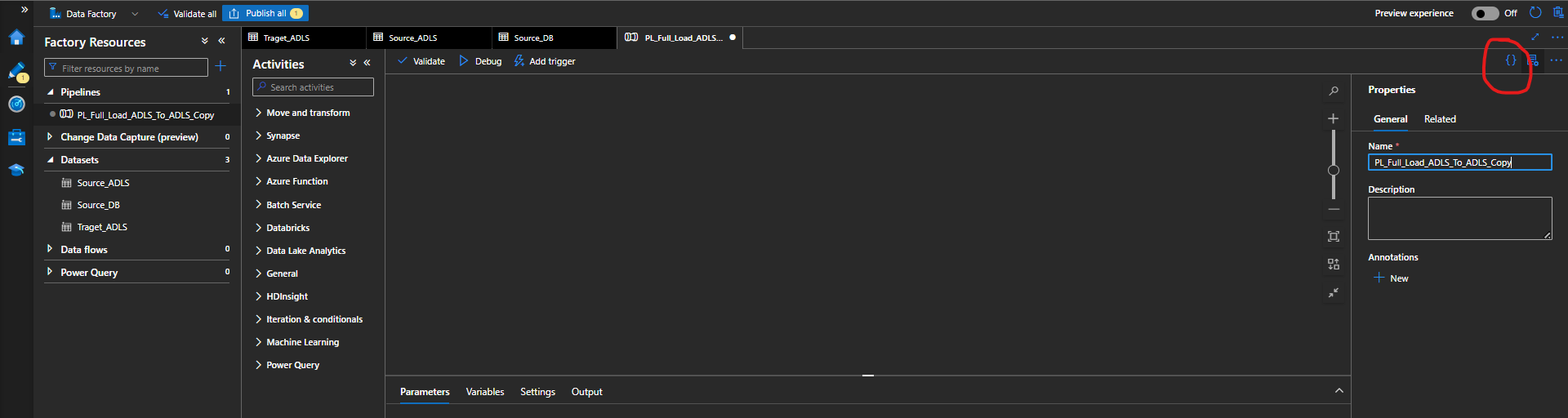


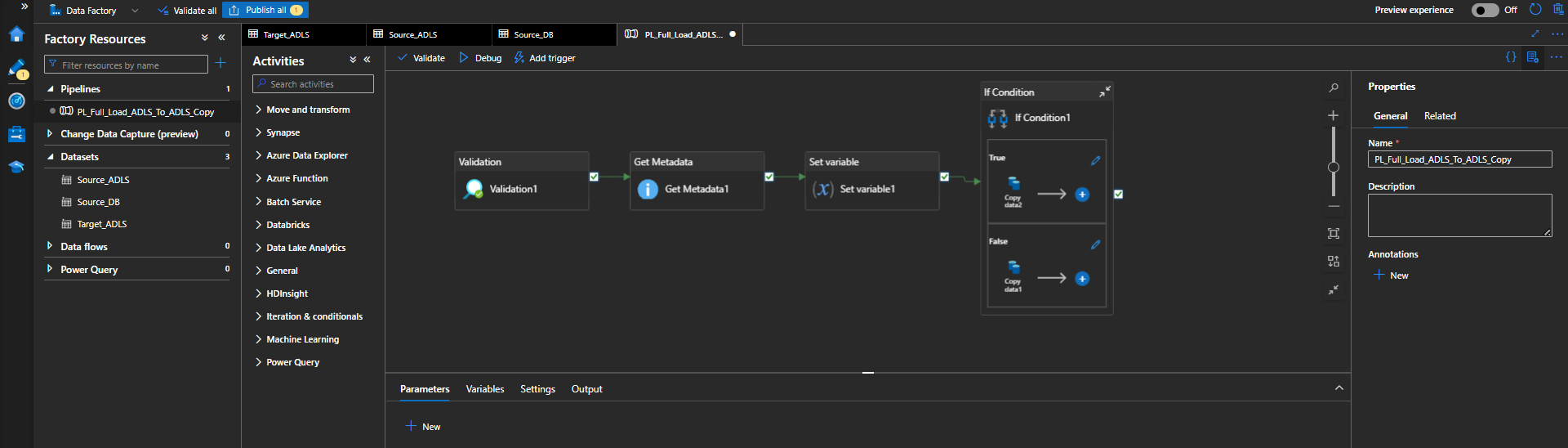
1. Once it is done click on ok you will be able to see the dataflow like below
2. Do the same thing with other workflow post that we can “Publish” it to save our changes.

**PIPELINE CREATION**

1. Click the “New Pipeline” and rename the pipeline with the name “PL\_Full\_Load\_ADLS\_To\_ADLS\_Copy”



1. Once that is done click on the “{}“on top of the “name” and copy all the script available in the “PL\_Full\_Load\_ADLS\_To\_ADLS\_Copy” Json of the pipeline folder attached in the zip file.
2. Once clicking on Ok you will be getting the below screen:



1. Follow the same above steps remaining pipelines once done creation click on publish and publish them.