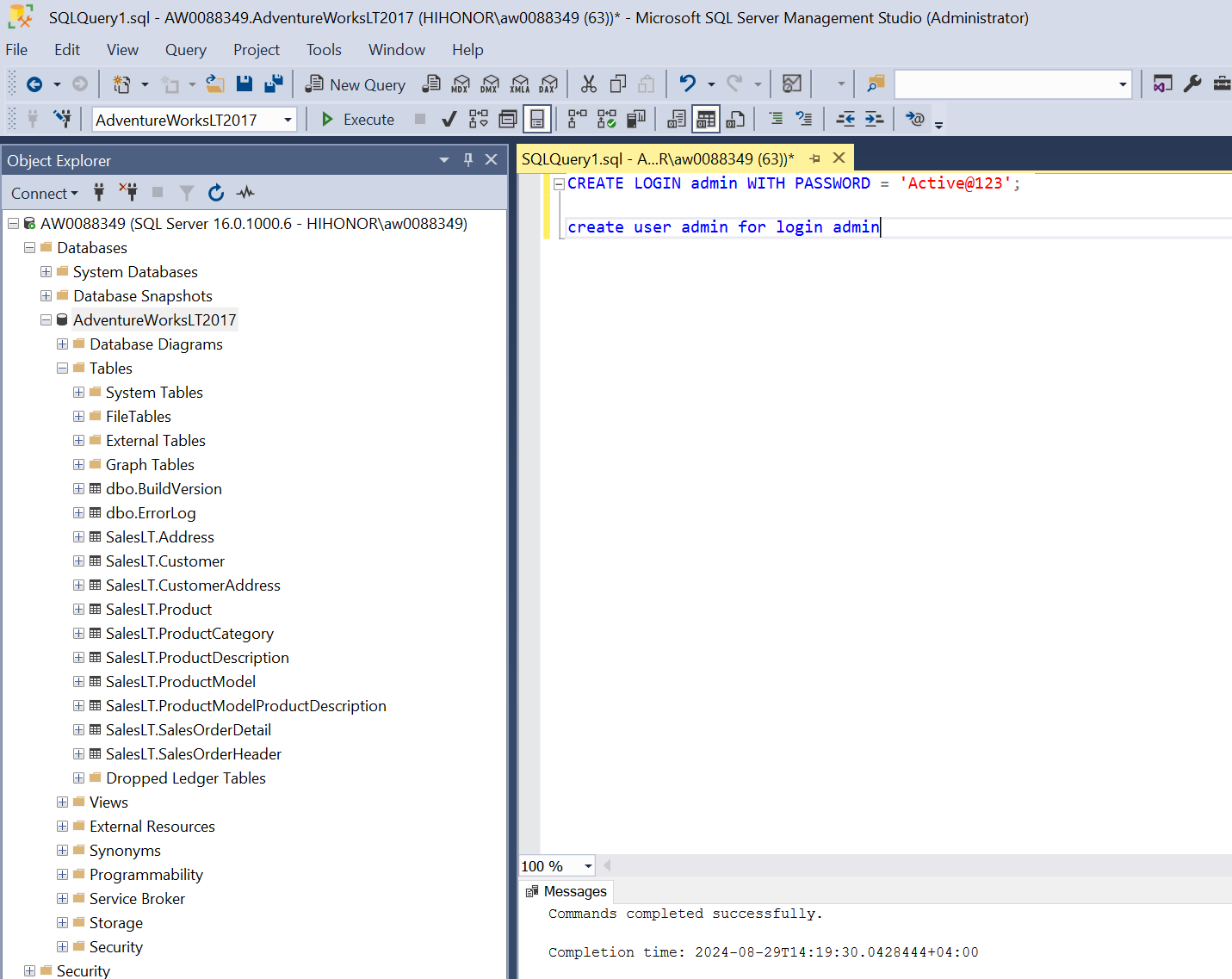
1. Install Microsoft SQl Server and also install SSMS
2. Restore AdventureWorksLT2017 databse in MSSQL server.
3. Create Login for a user:

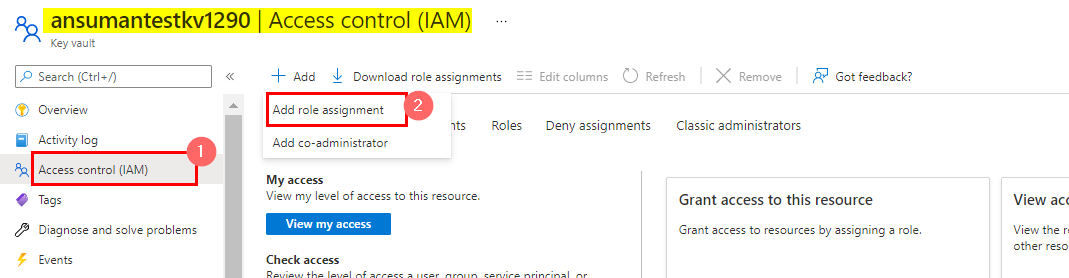


1. Go to security 🡪under user section 🡪click on user and select properties 🡪 click on membership🡪Assign **db\_datareader role**. (has reader access now)
2. Now setup Azure Resource🡪 create a resource group🡪Create a Key vault to store login credential to access on-premise SQL server

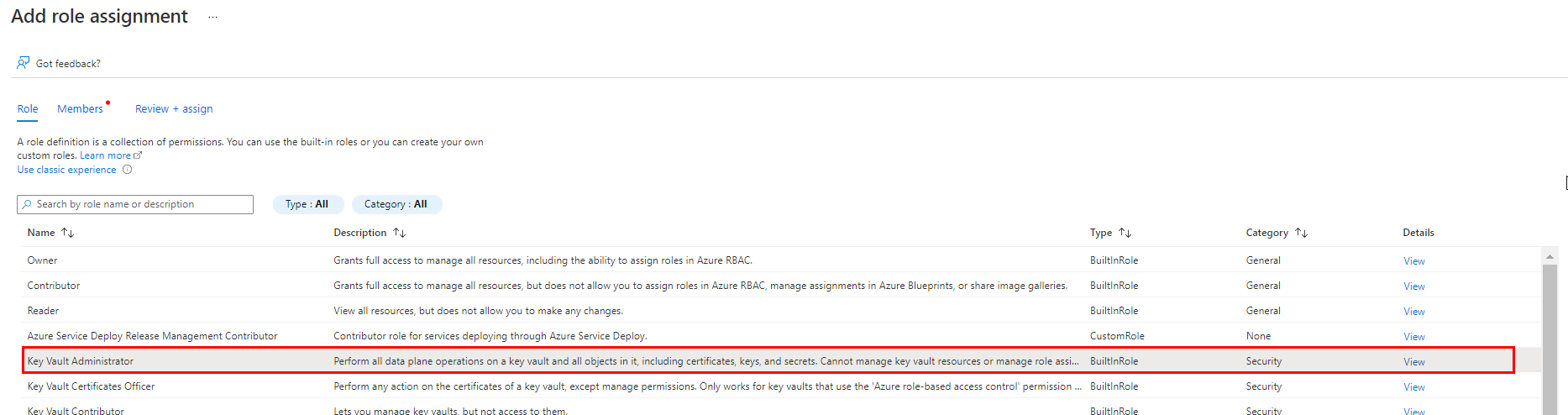
Go to Secrets🡪click on add a secret and add username and then again for password

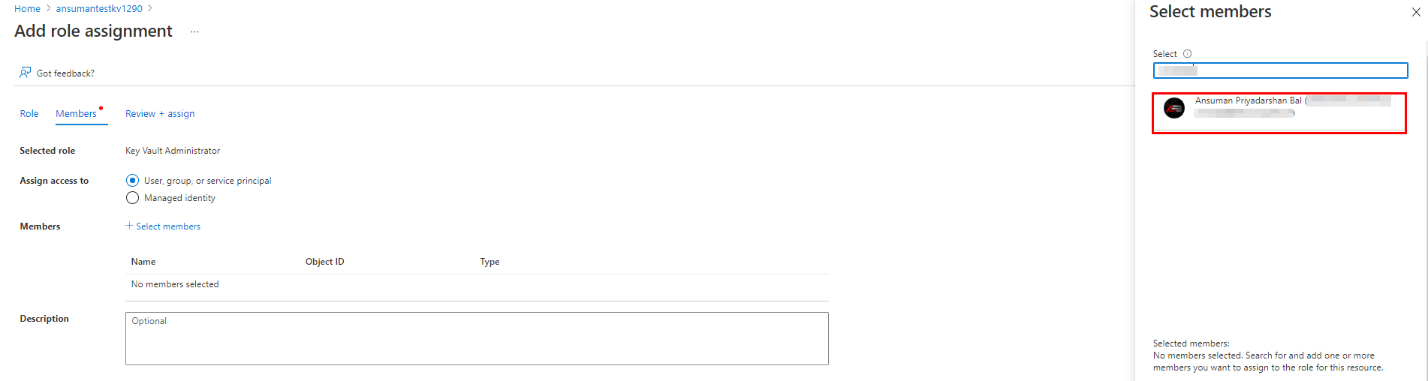
**Note if it gives error : “The operation is not allowed by RBAC. If role assignments were recently changed, please wait several minutes for role assignments to become effective.”**

1. Go to your Key vault after its created and then click on Access Control (IAM):

[](https://i.sstatic.net/xYa2e.png)

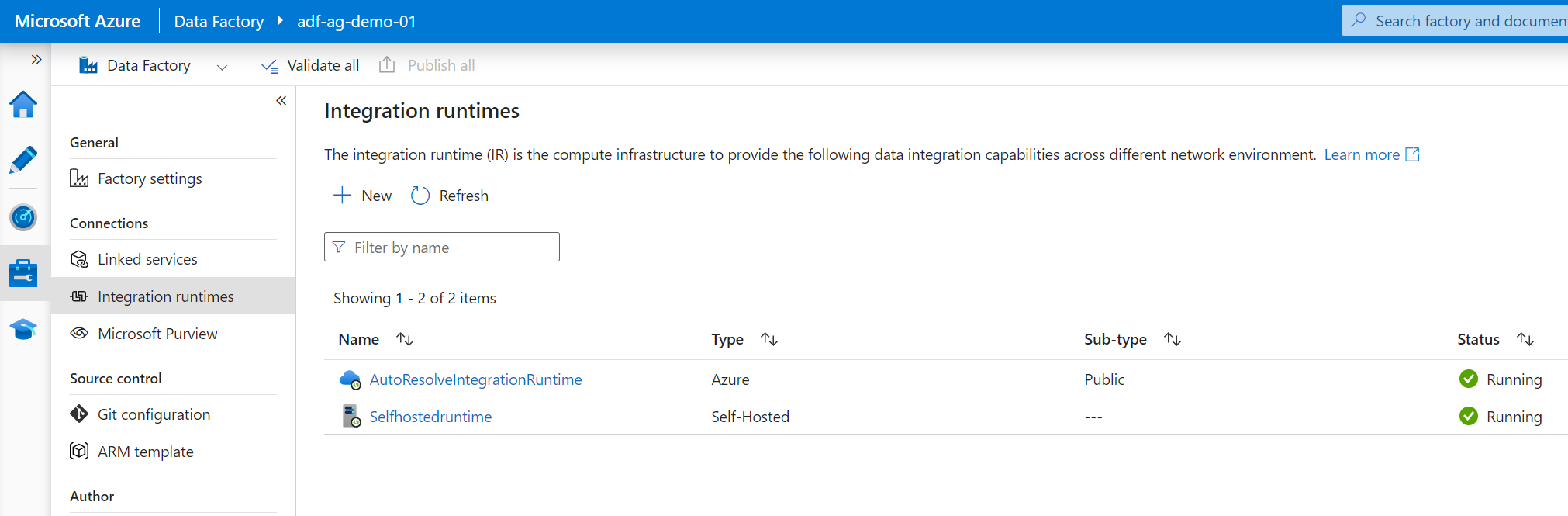
1. Then click on Add Role assignment and then add Key vault Administrator Role to your name:

[](https://i.sstatic.net/ut8yZ.png)

[](https://i.sstatic.net/954UB.png)

1. After you review and assign the role , you will be successfully able to create/manage the objects present inside the Key vault.
2. Now create a Azure data factory and then click on launch studio, we will be now copying data from on premise SQL server to Azure Data lake so for that we would need integration runtimes

Click on mange on left and then Click on Integration Runtimes🡪 Click on Add button and select Azure Self Hosted 🡪continue🡪Give name and create🡪It will download a file 🡪open the file it will download and install Microsoft Integration Runtime on your laptop.



1. Now we have to copy data from SQL server to Azure using pipeline🡪Click on Author on left 🡪Click on pipeline🡪create pipeline and the search for copy data 🡪Now we need to setup Source and Sink for this.

Select SQL server as dataset🡪create Linked service (Legacy)🡪Click on New🡪Giver Server name, Database, username and select azure key vault for password🡪Create AKV linked service🡪

Select secret name.

Test the connection.

**Note: Sometimes it’s shows Loading failed error message in that case you need to go to Key vault and then got to Access policies🡪 then create an access policy and grant all the access.**

🡪Select the respective table you want to copy.

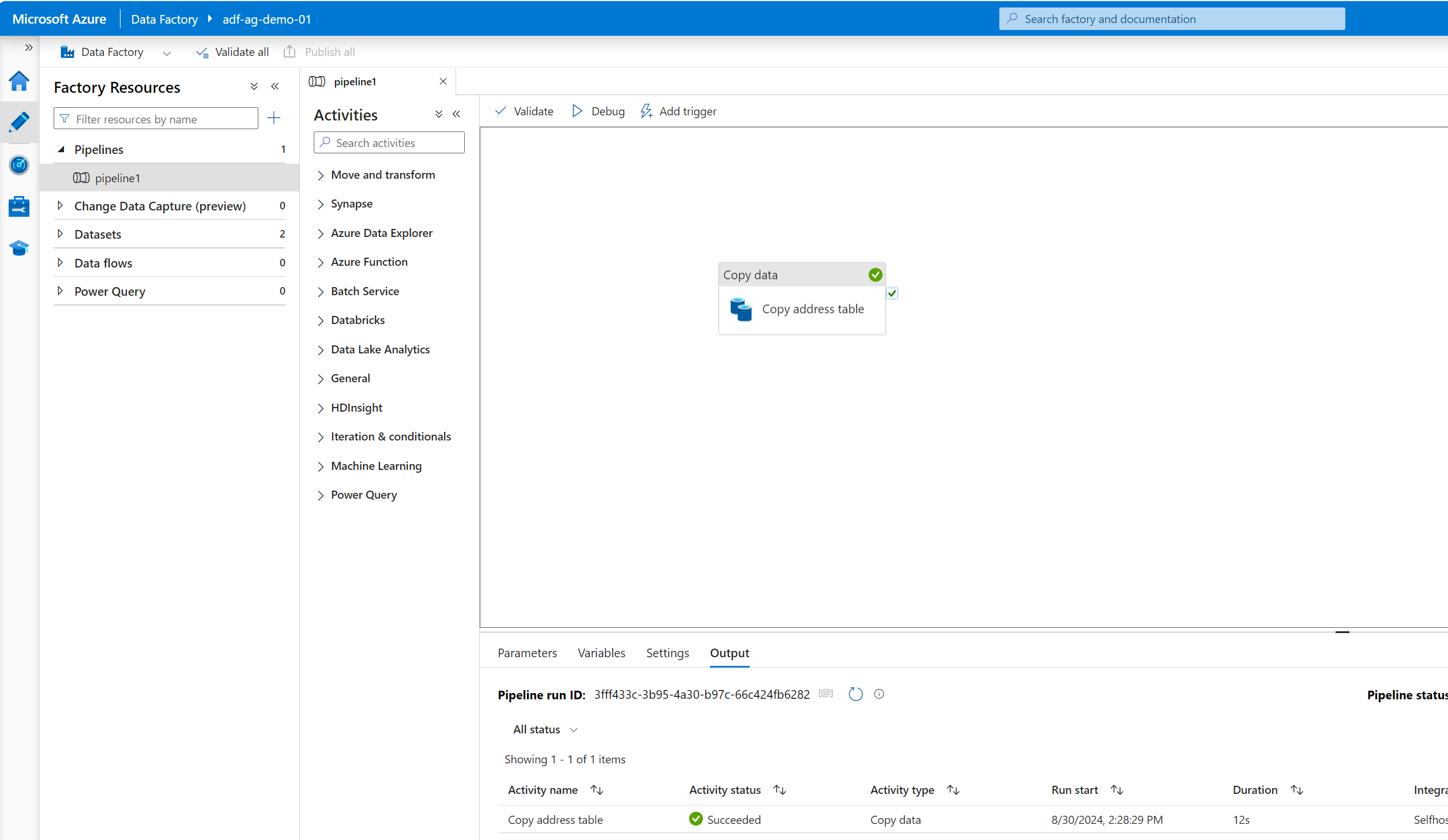
1. Now we need to configure Sink. 🡪Click on New🡪 Search for Azure gen lake2

And select format as parquet🡪Set properties🡪Configure Link Service

Test Connection

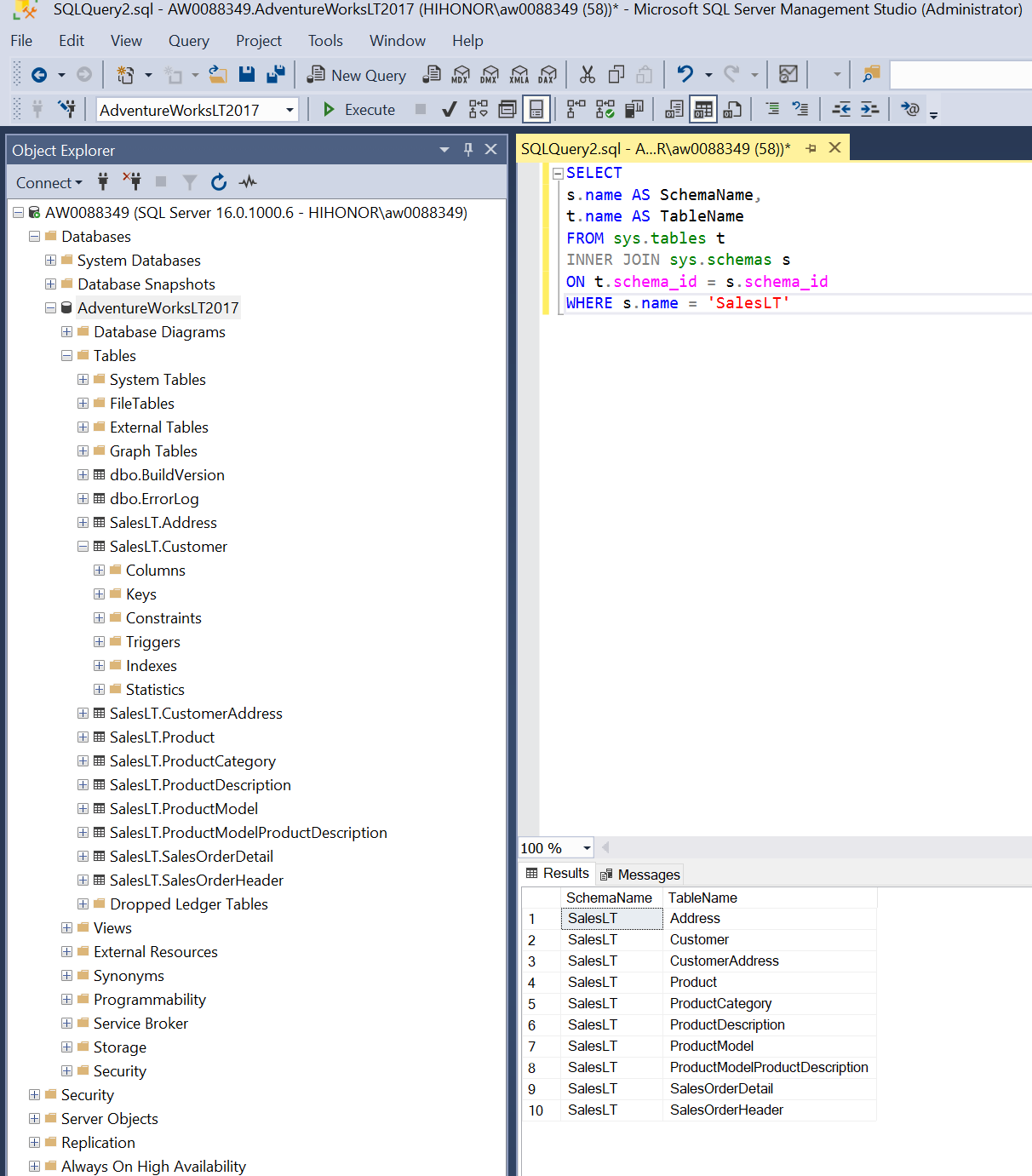
Specify the location where you need to copy files.

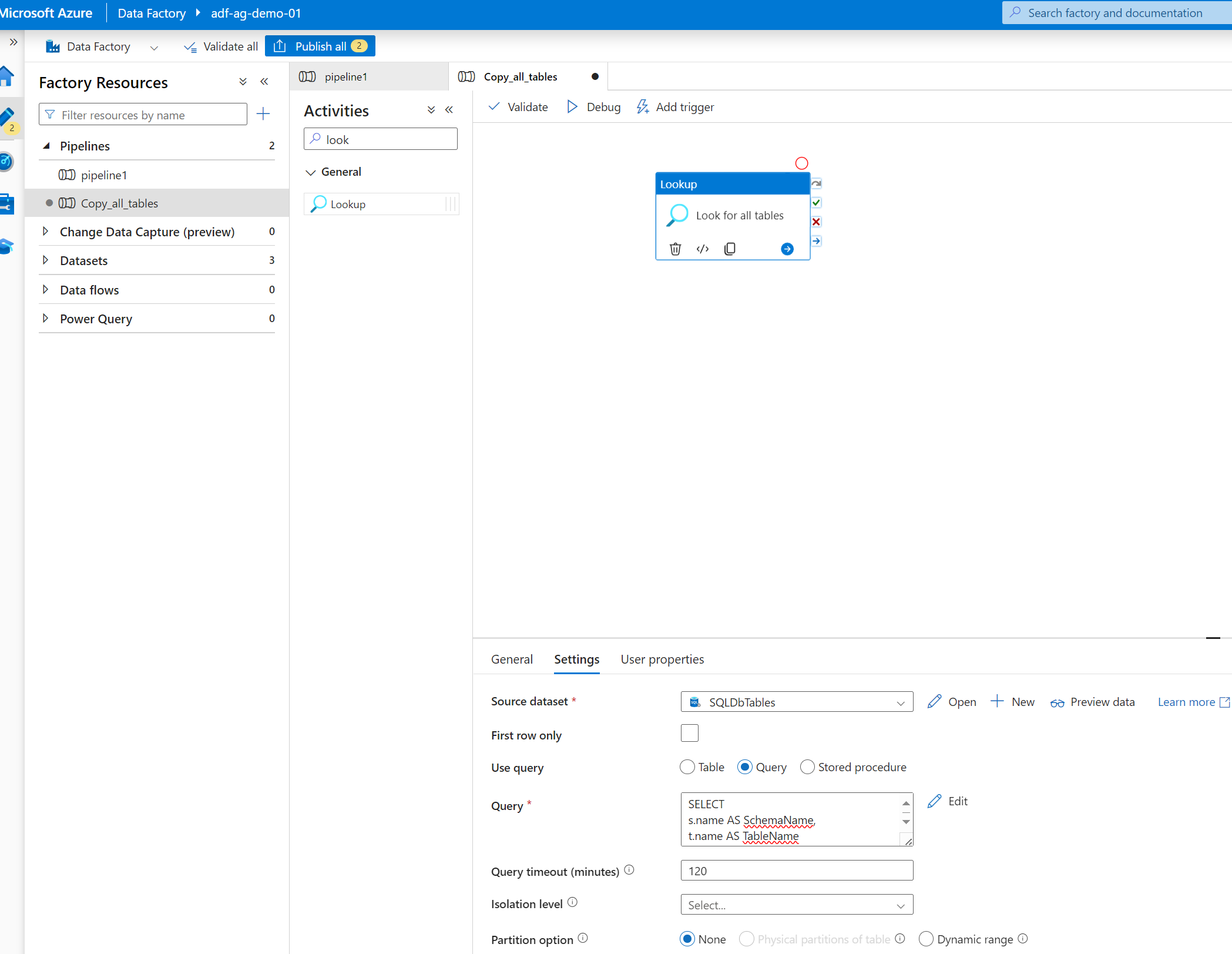
🡪Debug the copy data pipeline.



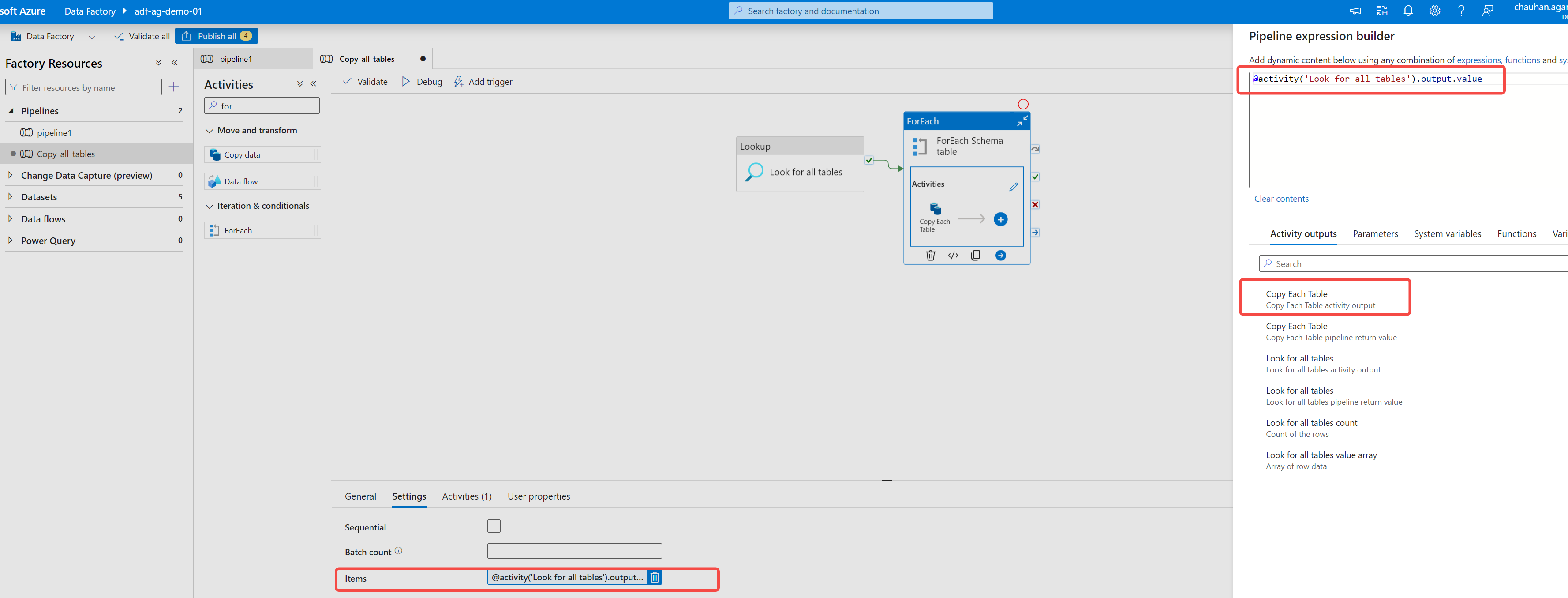
**Creating one single pipeline to copy all the tables from SQL server.**

1. First we will write query to list down all the tables to be copied.
2. Next got to ADF and select pipeline🡪give name
3. Search for lookup activity and drag it
4. Go to Settings 🡪create source dataset
5. Select use Query option🡪copy the sql query to list all the tables

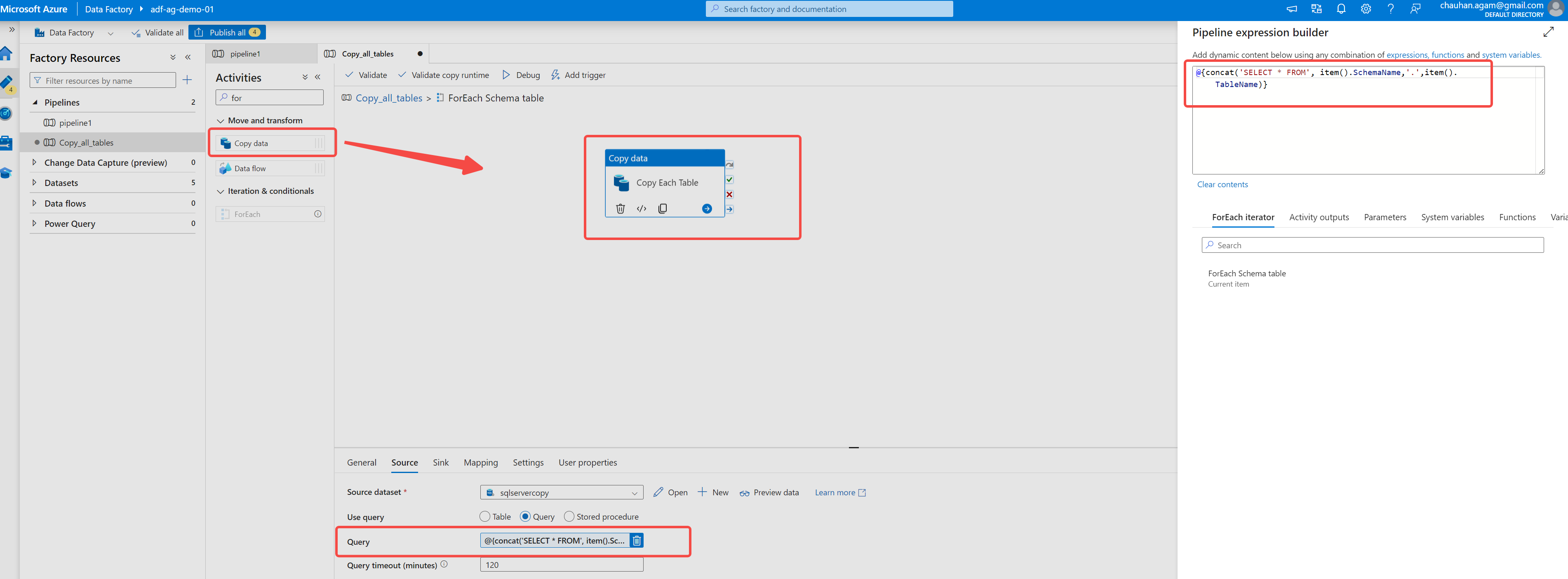




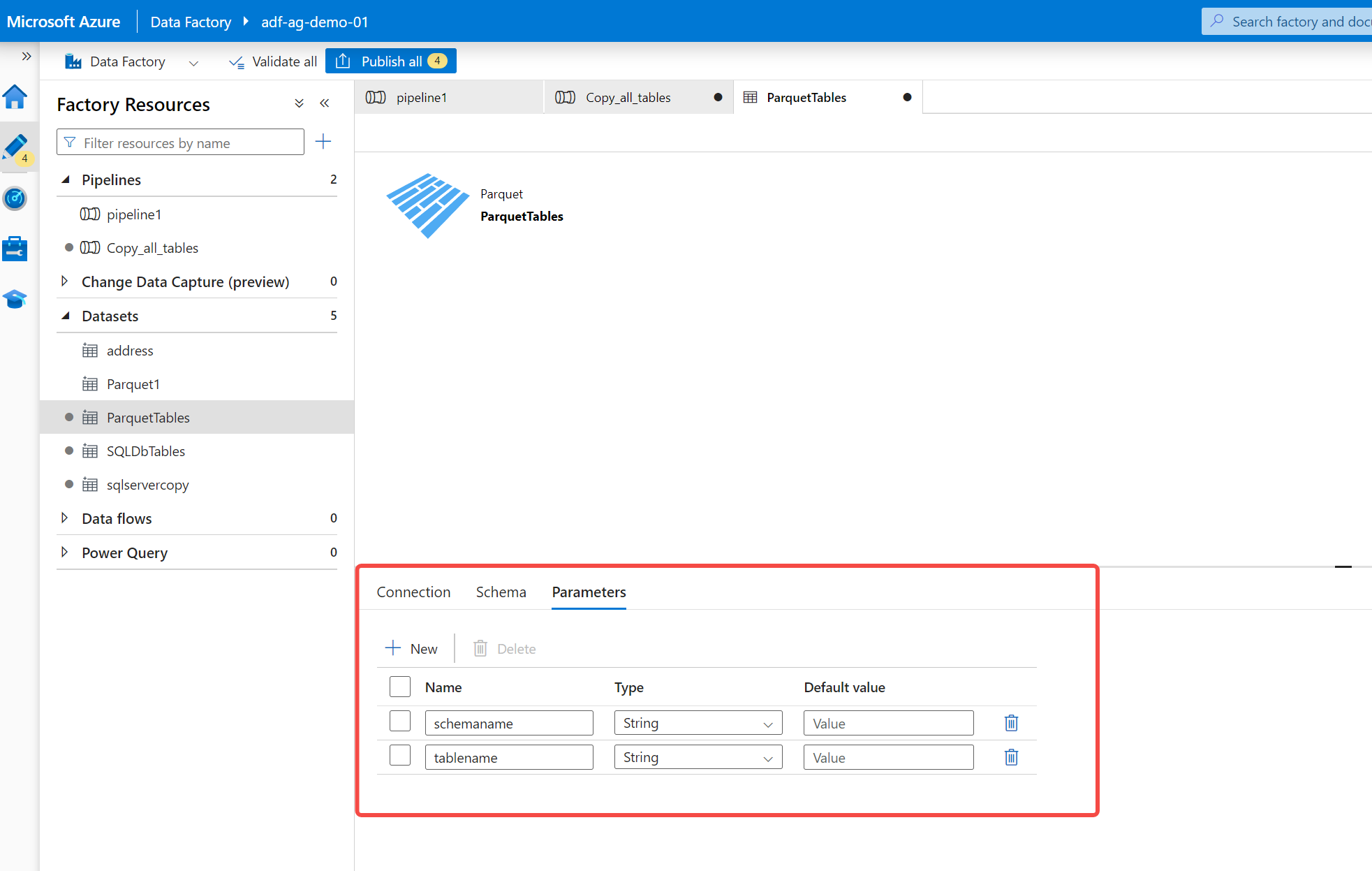
Now search for each activity and drag it 🡪click on setting 🡪select Items—then select dynamic option below.



Then click on Activity section and drag the copy data flow from left 🡪Click on Use Query option 🡪Under Query option click on dynamic field and write code in pipeline expression builder



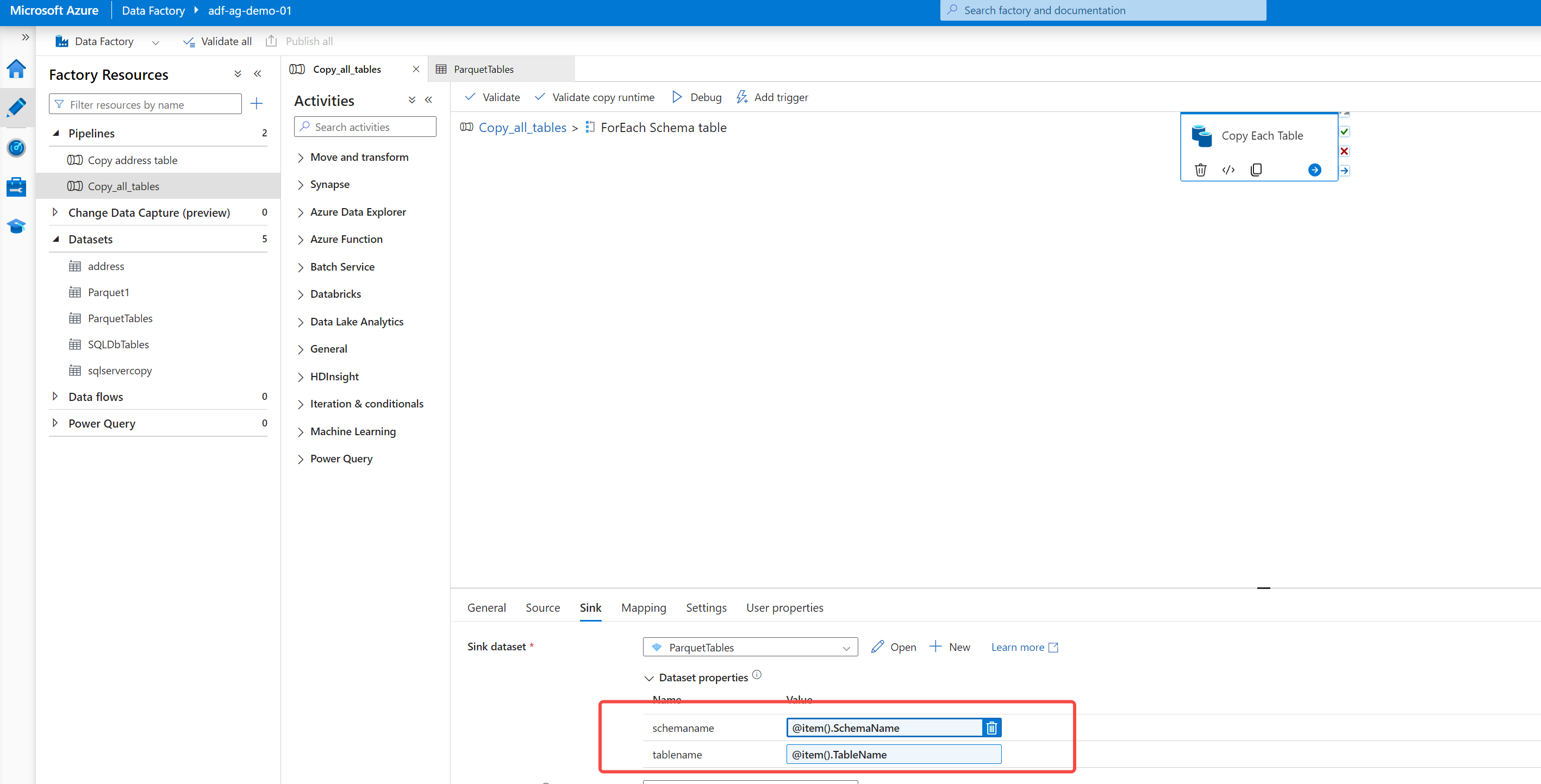
Now configure the Sink properties 🡪select azure data lake 🡪select parquet🡪go to parameters and add two parameters as below:



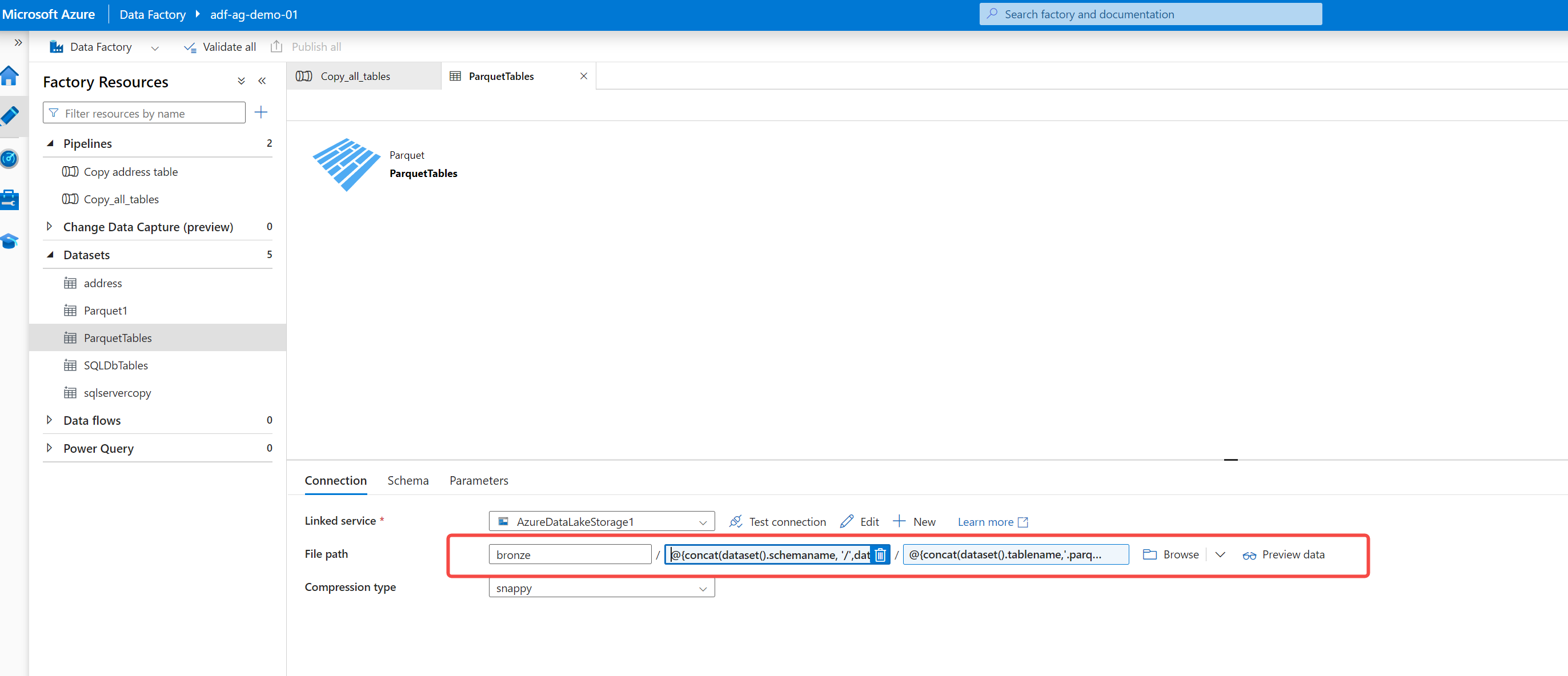
Folder structure need to be in **bronze/Schema/Tablename/tablename.parquet** format

We are getting schemaname and tablename from lookup activity so in source we already use the 2 item values

Now go to Sink dataset🡪you will see the above 2 parameter created 🡪click on values and click on dynamic section 🡪do same for other parameter as well



Now edit sink dataset and add these two values under the file path :



@{concat(dataset().schemaname, '/',dataset().tablename)}

@{concat(dataset().tablename,'.parquet')}

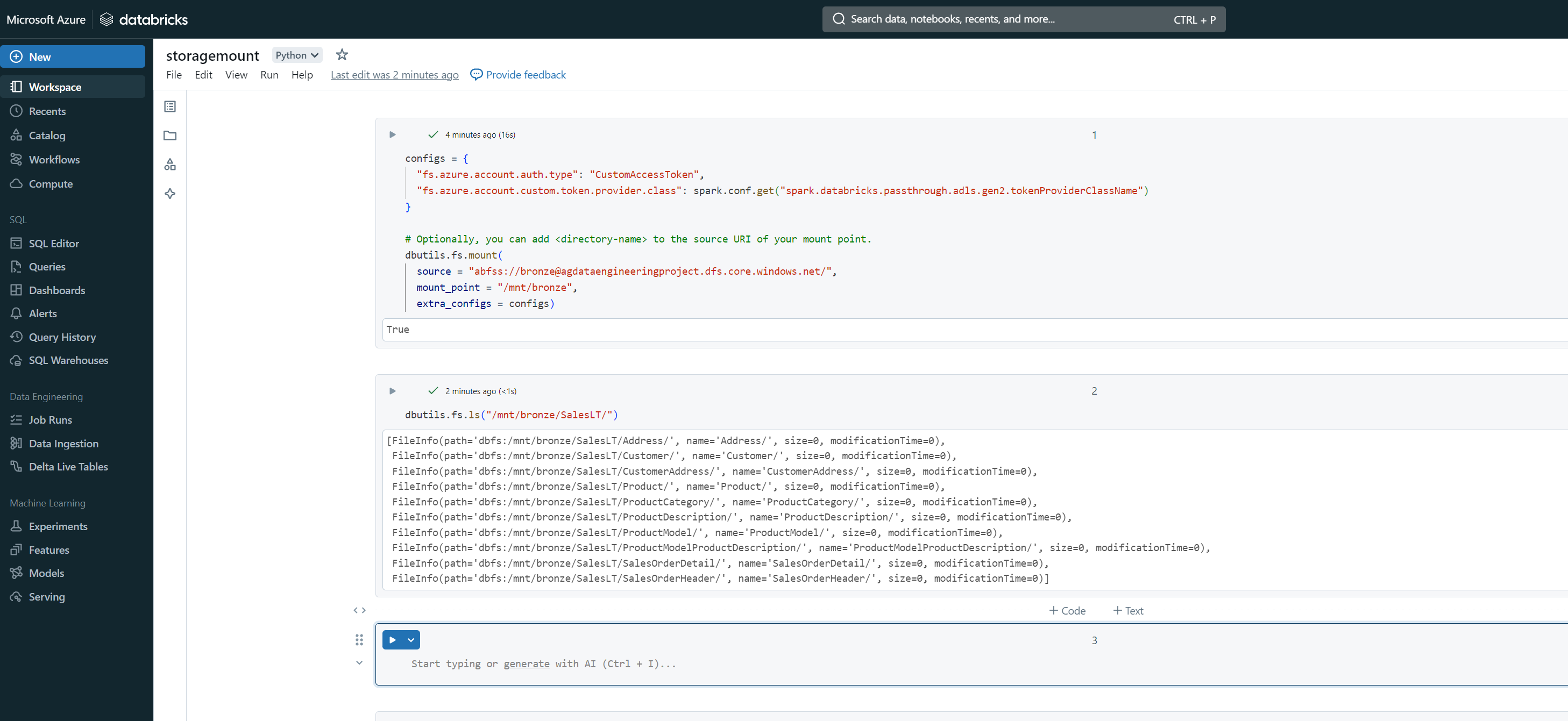
Publish all the changes🡪debug the pipeline and see if it runs successfully 🡪you can monitor the pipeline under monitor section.

After successful executing the pipeline, you will see all the tables in the container under bronze folder as per the directory.

**Creating Azure Databricks Workspace for Data transformation**

**First part: Mounting containers:**

* 1. Create databricks workspace🡪click on compute tab🡪create a cluster for compute (for now select single node, select timeout after 15 mins under advanced option)
  2. Now we will create a workbook to mount the container first



**Second part: Read data and do data transformation:**