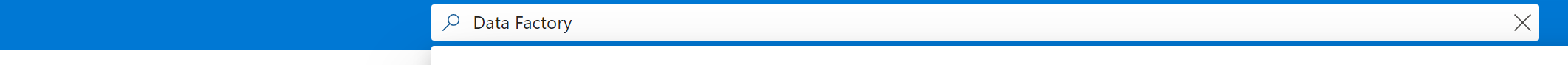
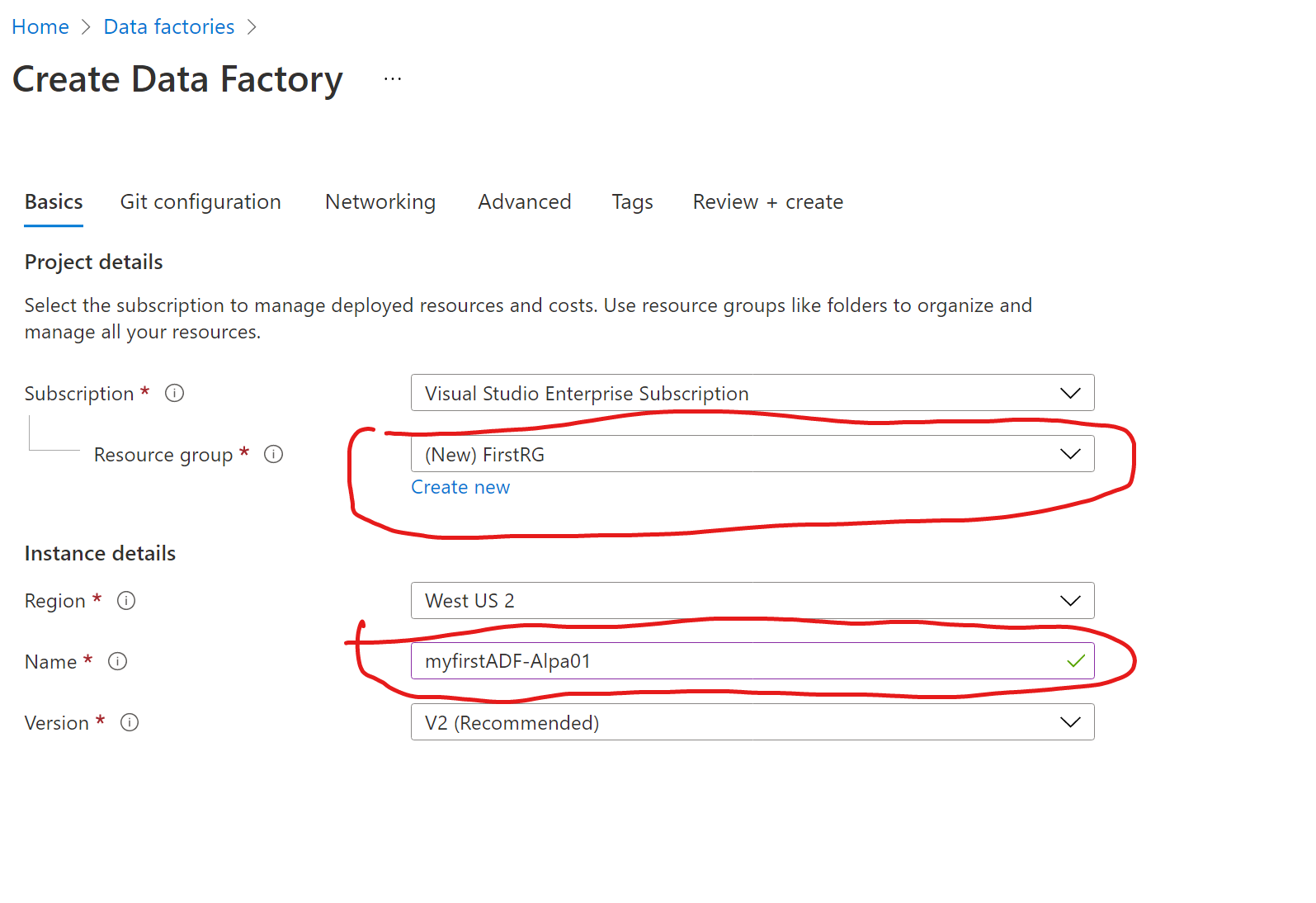
# How to Create Azure Data Factory?

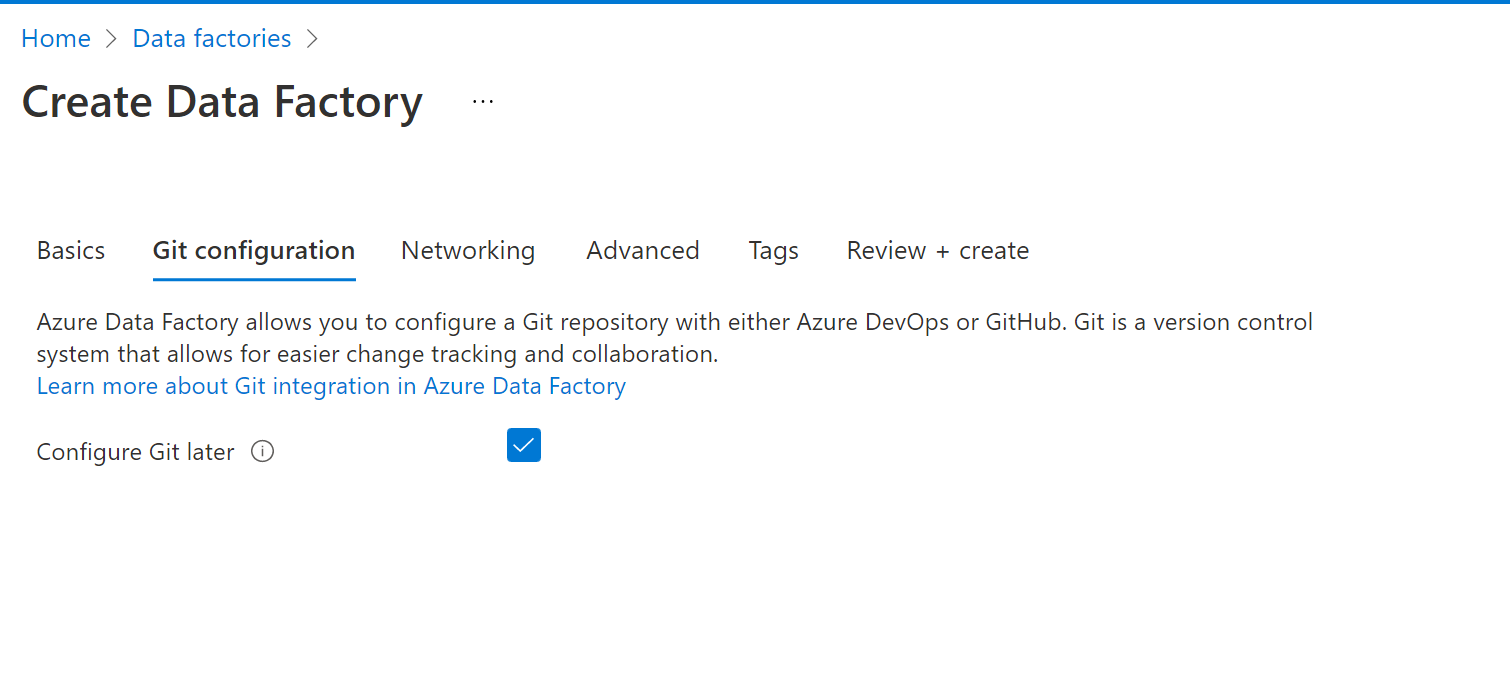
1. Go to portal.azure.com
2. Search for Data Factory



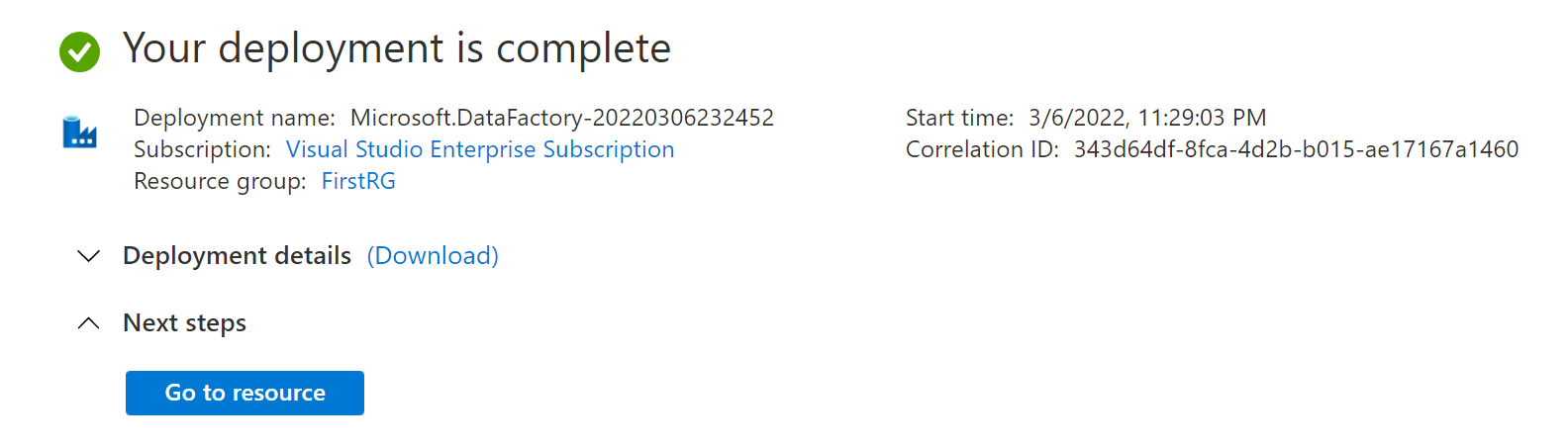
1. Click on **+ Create**
2. Create Resource Group as “FirstRG”
3. Add Name as “myfirstADF-yourname”



1. Click on **“Next : Git Configuration”**
2. Set Checkbox on



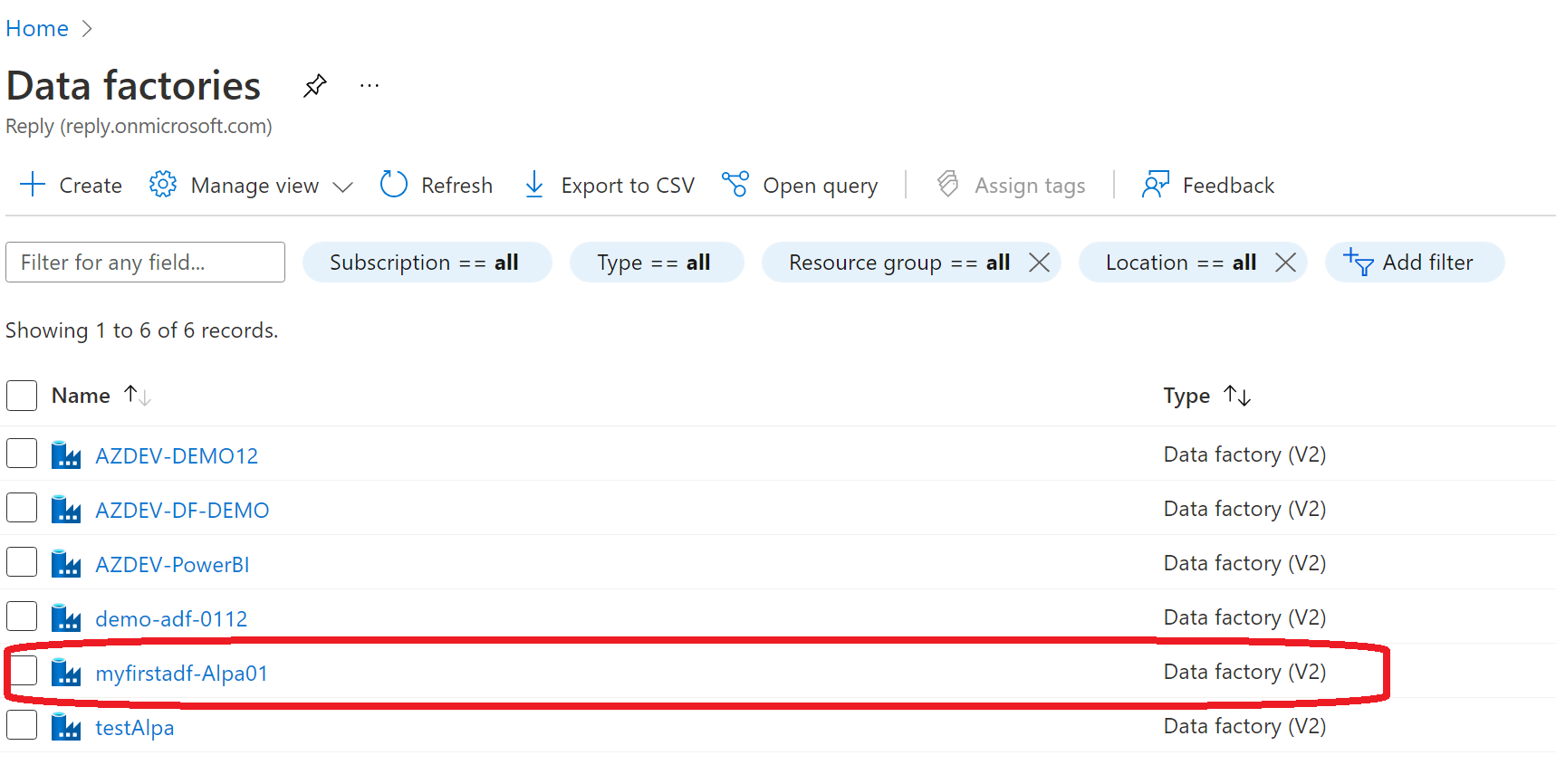
1. Click on **Create and Review (this will do validation and give error If something is not right)**
2. **Click on Create.** Wait for few minutes .You will get Green Mark like this once it has deployed successfully



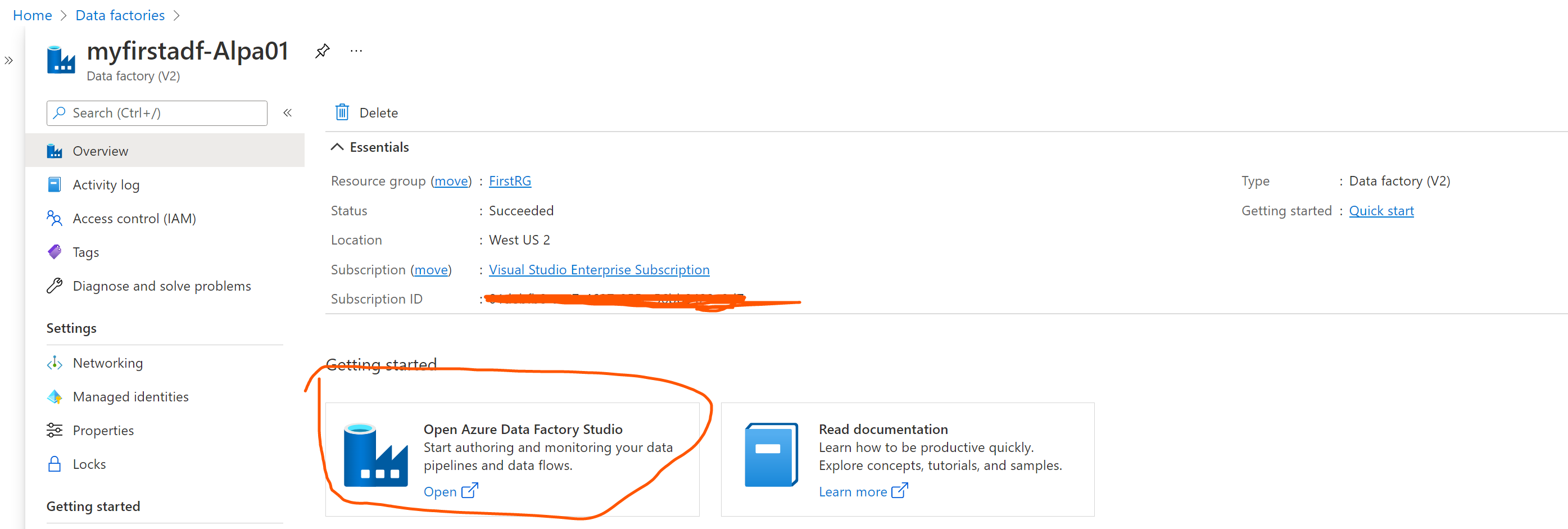
10 Congratulation!!! You have created your First ADF successfully ☺

# How to open you Azure Data Factory?

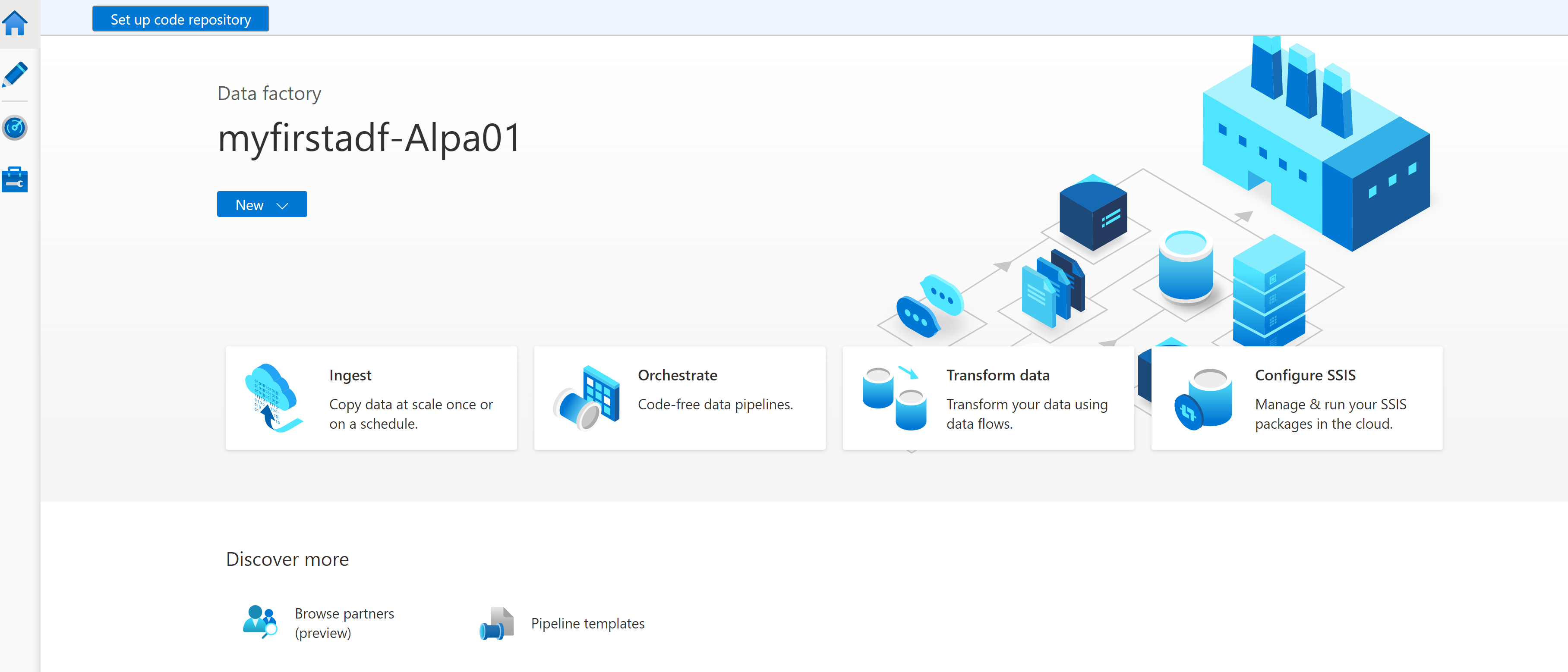
1. Go to portal.azure.com
2. Search for “**Data Factory”**
3. You will see your recent created data factory “**myfirstadf-yourname01**” as



1. Click on your data factory
2. You will see “open Azure Data Factory Studio” Click on that

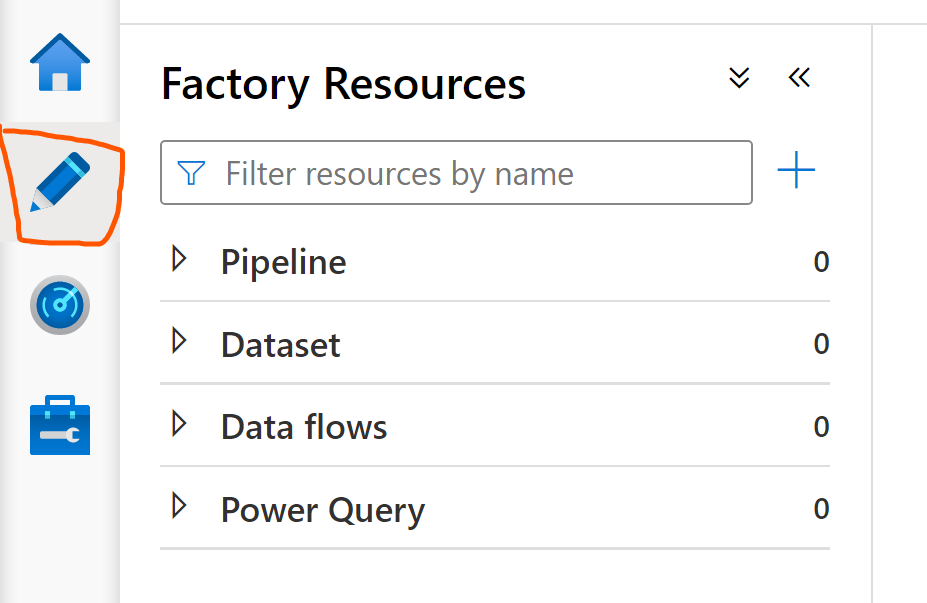


1. You will see screen as shown below



1. Left Side, you will see home Icon + three other icons for Manage, Monitor and Author ADF pipelines.
2. Click on second module which is Author Module(Pencil Icon)

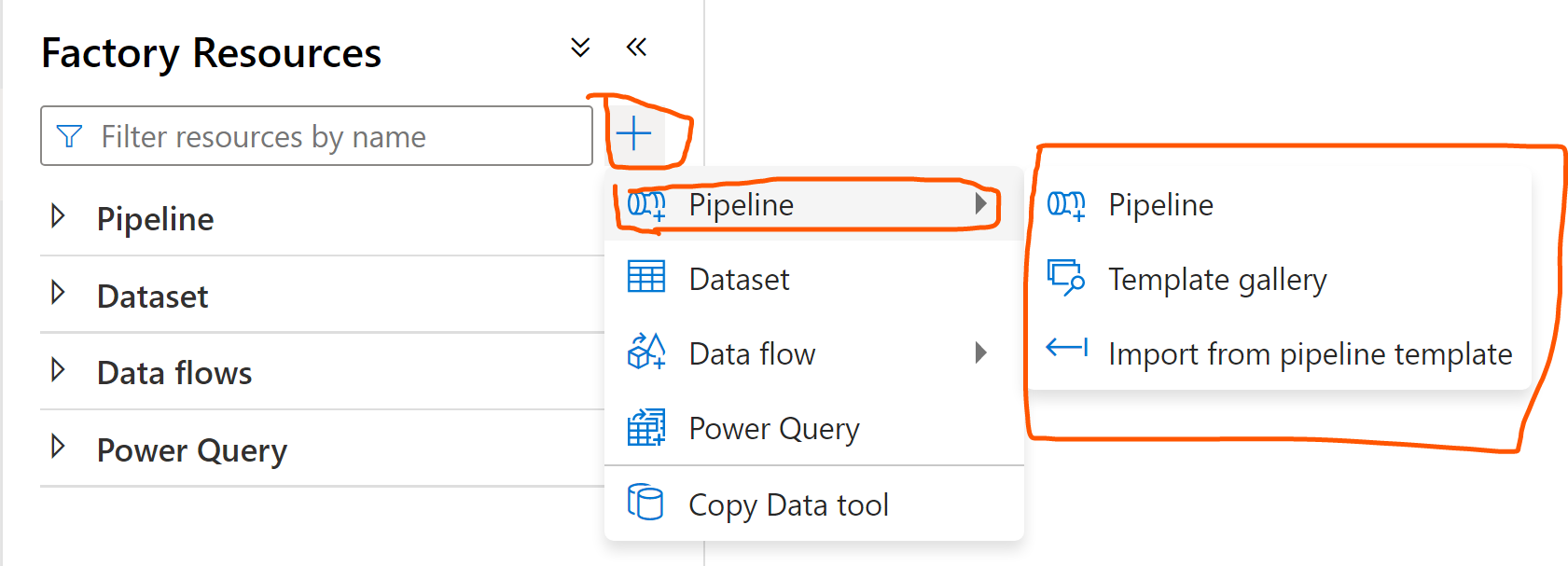
It will look like below



# How to create 1st pipeline in ADF?

You can create your pipeline in two different ways both way you get same options to create pipelines. You can create pipelines as brand new or from Template Gallery or import from pipeline template

1. Click on + icon and then click on pipeline



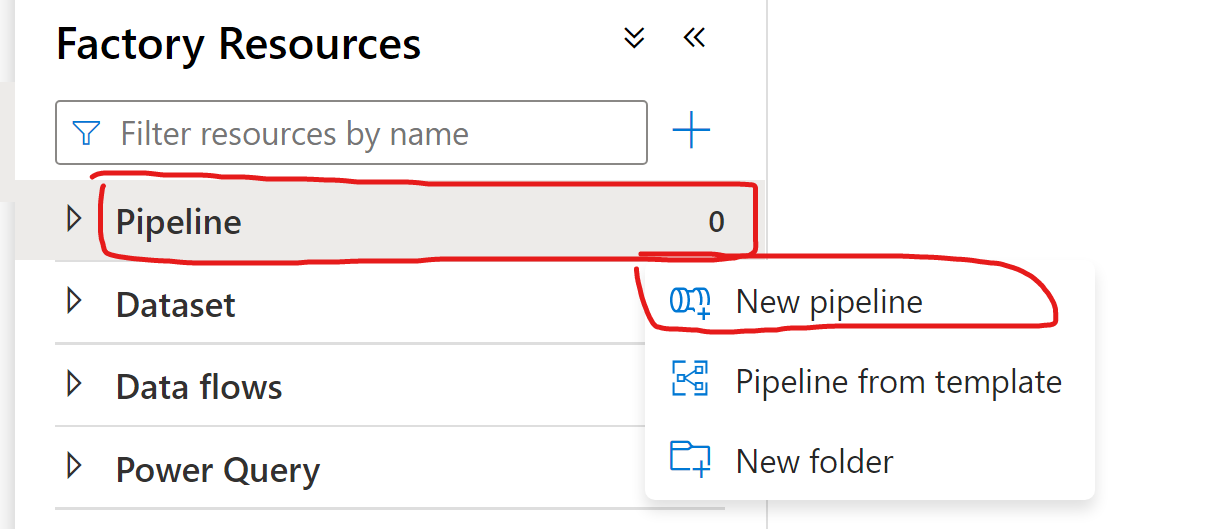
1. Click on … (three dots near pipeline)



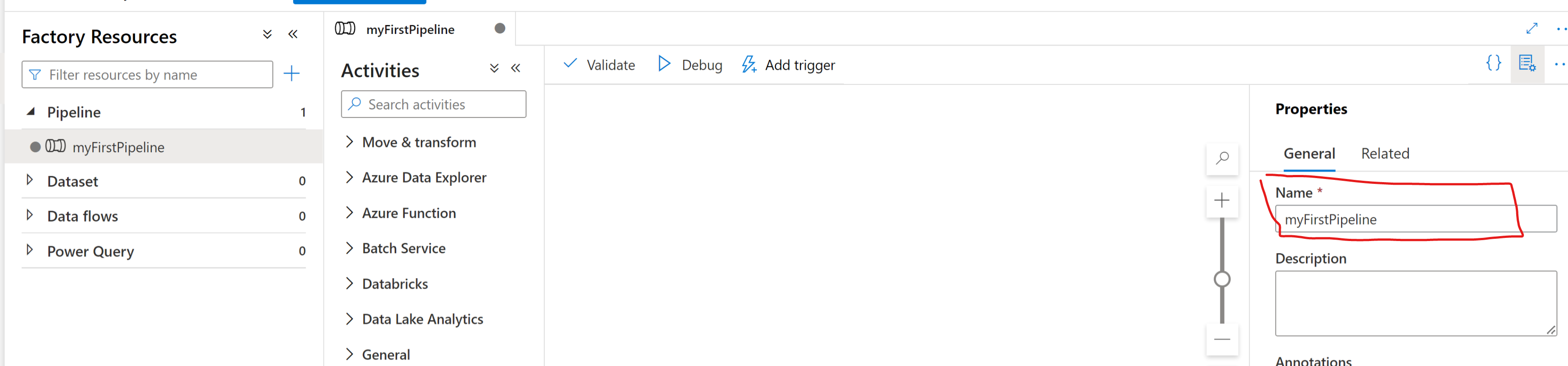
You can create brand new pipeline or you can create from existing template or from Gallery which has readymade pipeline template.

Let’s add brand new pipeline.

1. Click on New pipeline as below



1. Renamed **Pipeline 1** with suitable name

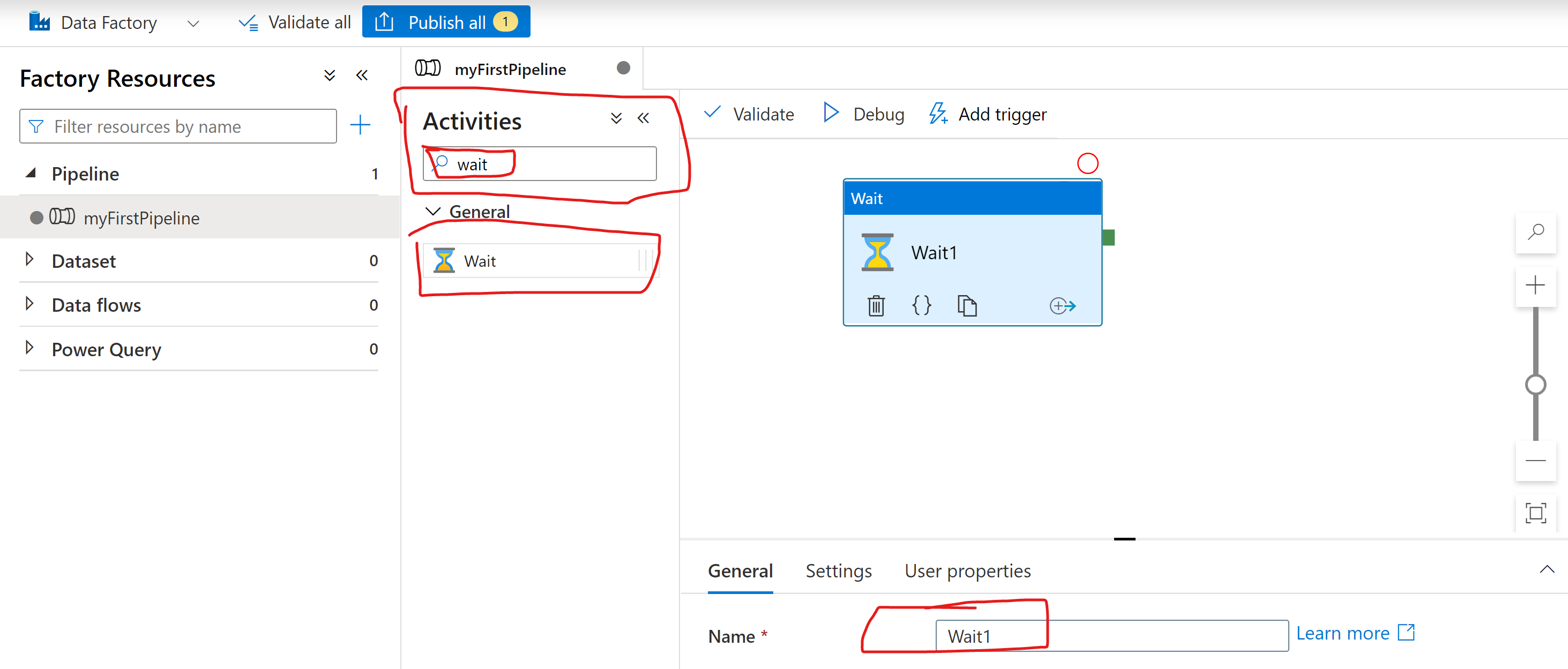


1. Add **Wait** activity

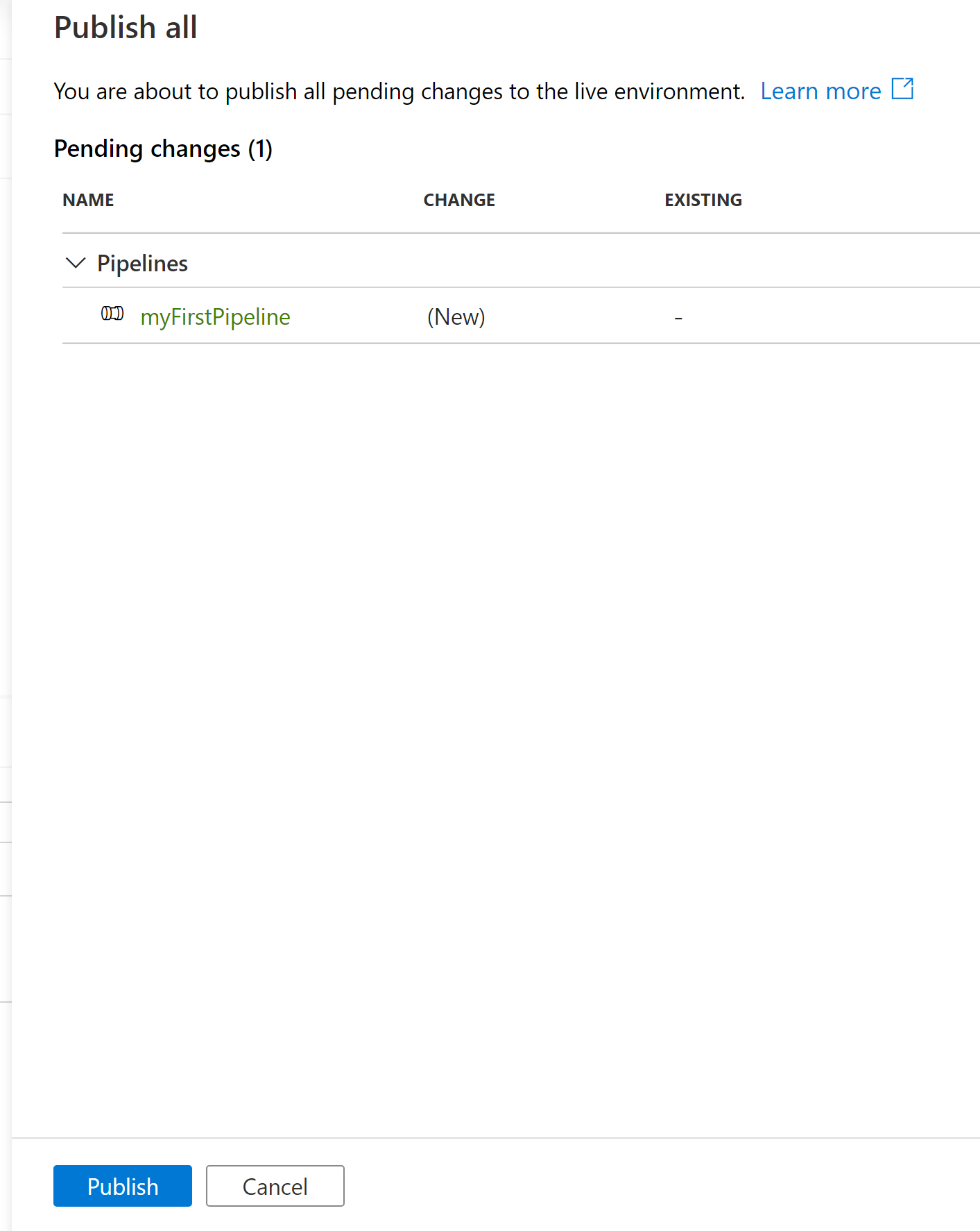
Search **wait** in activity area

Select **Wait** Activity

Drag and Drop to canvas as shown below



1. Click on **Validate** to validate pipeline
2. Click on **Debug** to run pipeline
3. Click on **Publish** all. You will see list of changes. Click on **Publish.** It will take few secs to complete it



1. Congratulation you have created your 1st pipeline

# How to set up Git Repository

To set up repository, you need following information.

1. Organization name
2. Project Name
3. Repository Name
4. Branch Name
5. Root Folder.

In case you do not have above information then, you can create them using either GitHub login or azure devops login.

1. Go to dev.azure.com OR GitHub.com
2. Create organization
3. Create project
4. Create repo
5. Initialized repo.

Now go to your Data Factory => Open Data Factory Studio

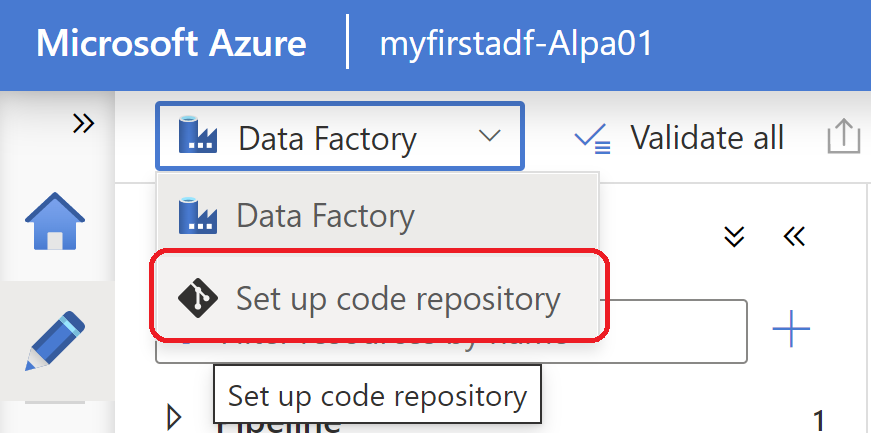
Now, Let’s Set up Repository by few button clicks away.

1. Click on **Set up code repository**

You can create it in using three different way after creating your Data Factory.



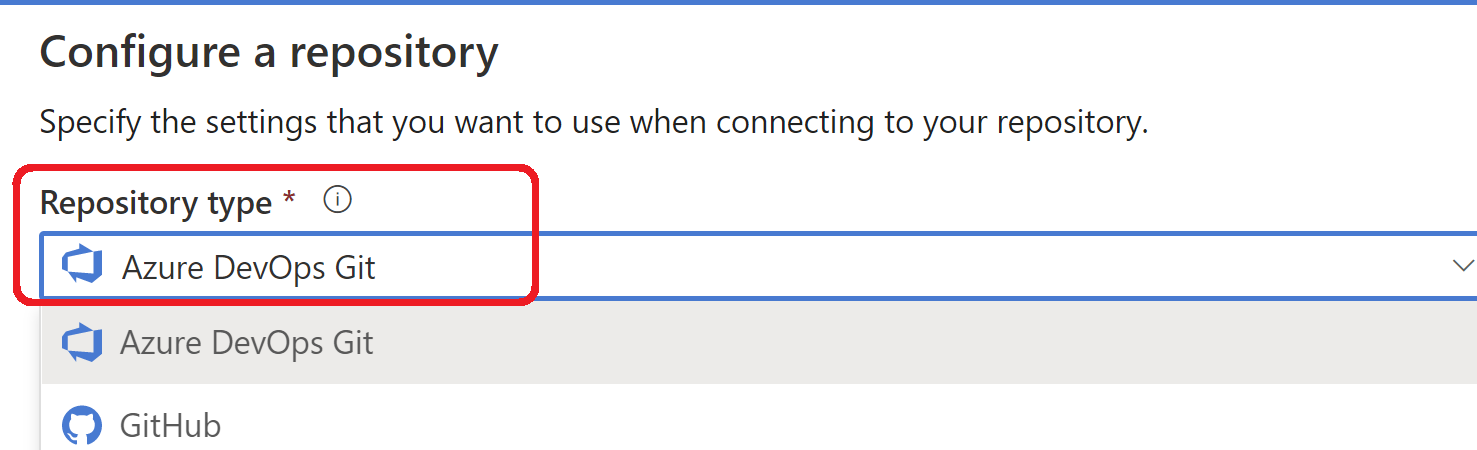
**OR**



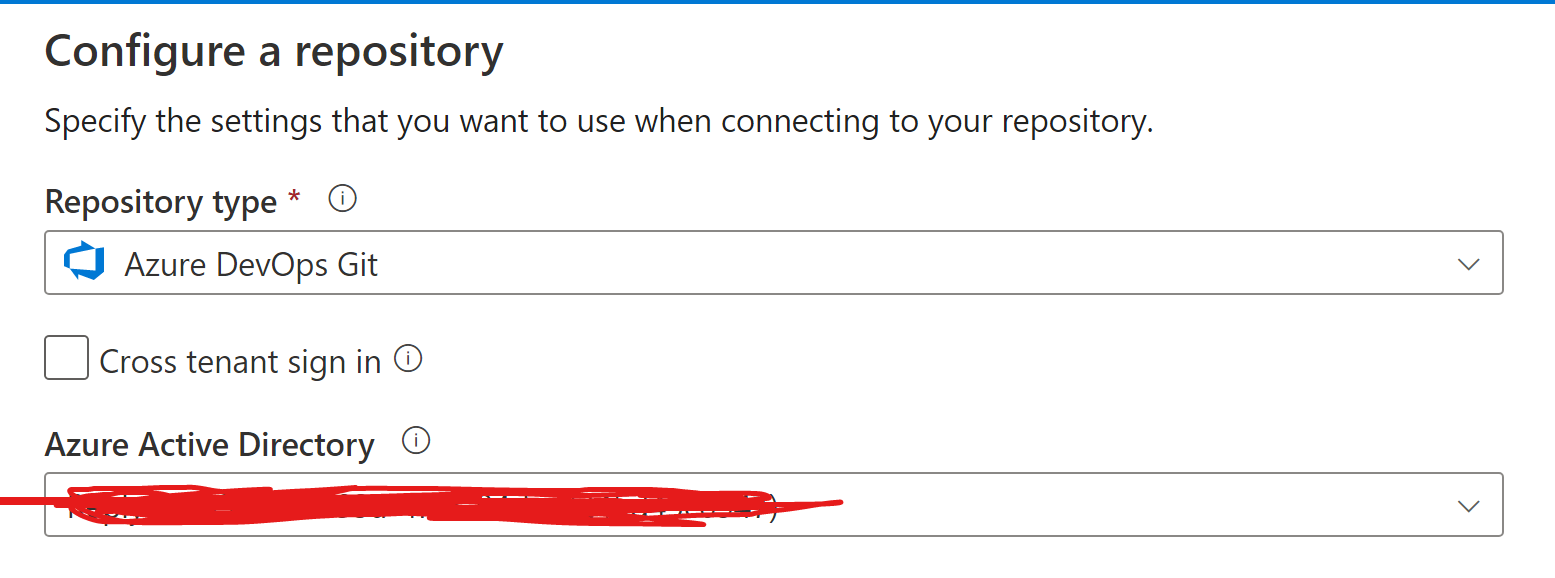
**OR**



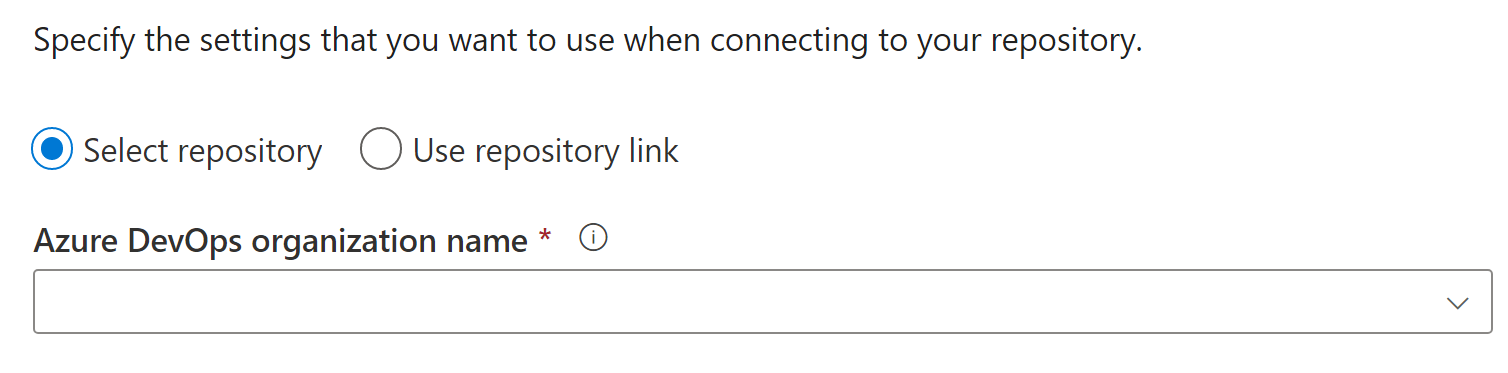
1. Select **Repository Type**



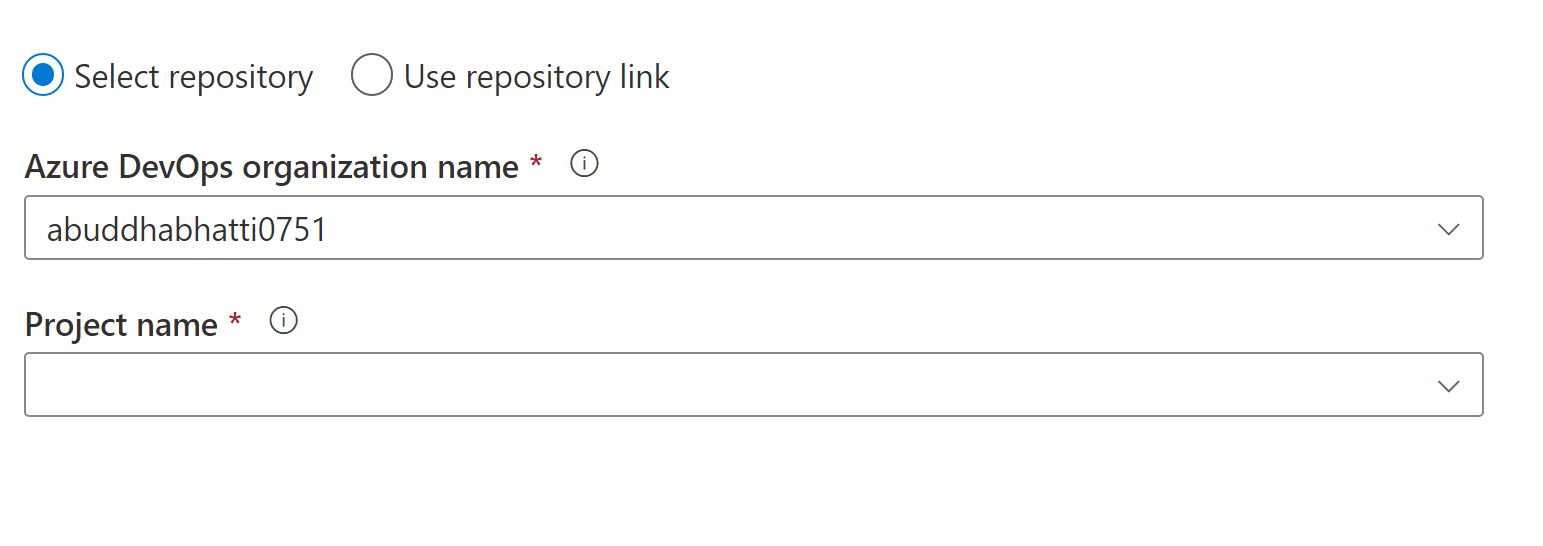
1. Select **Subscription**



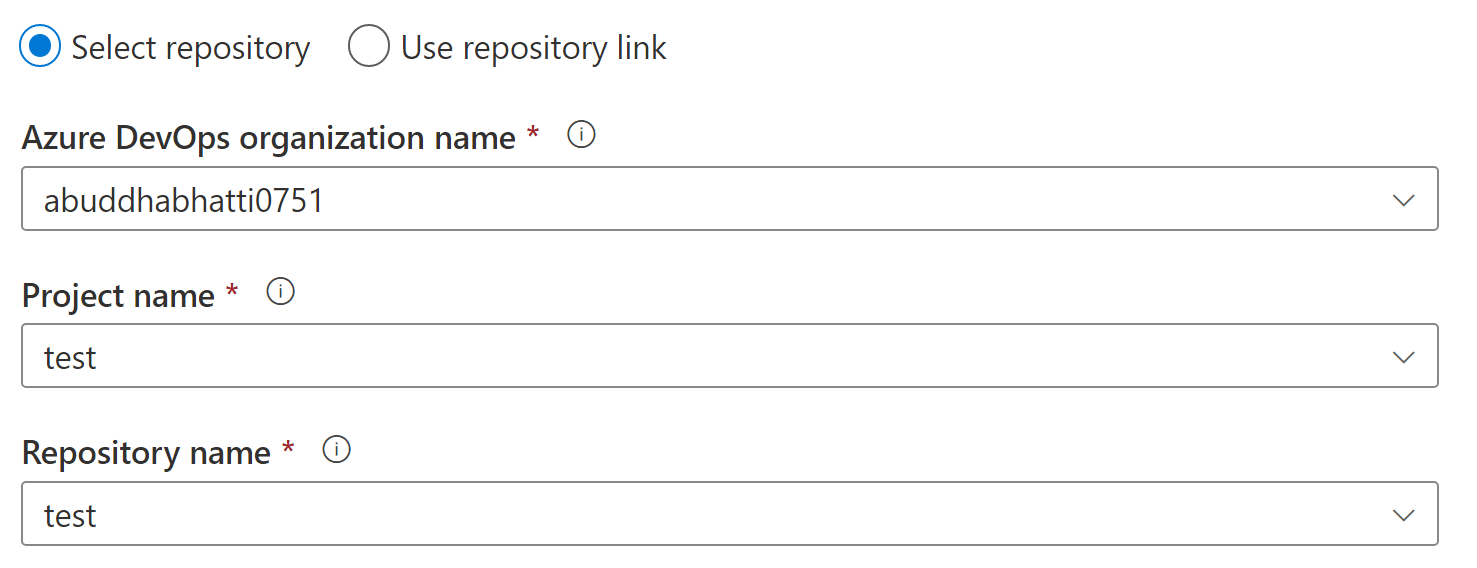
1. Click on **Continuous**
2. Select **Organization**



1. Select **Project Name**

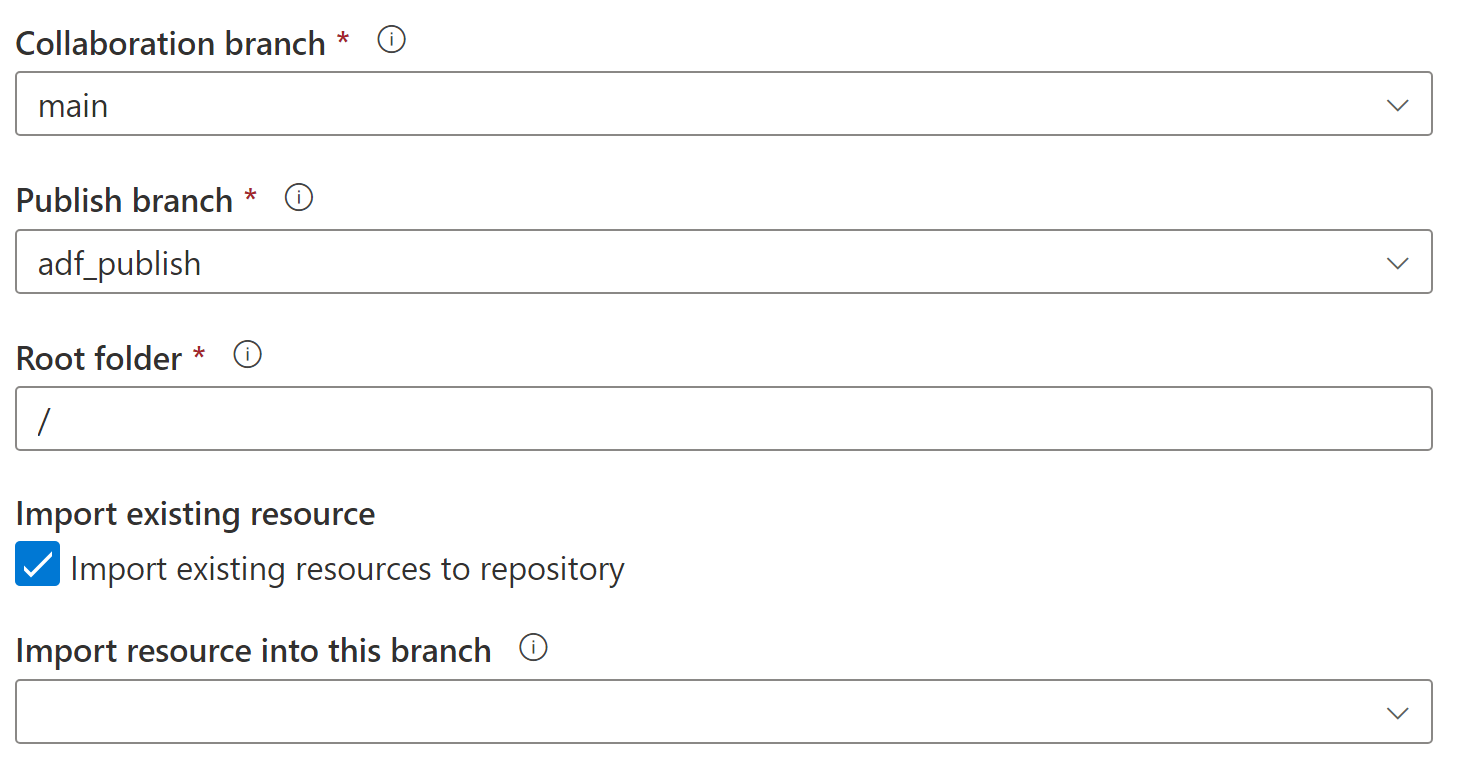


1. Select **Repository Name**

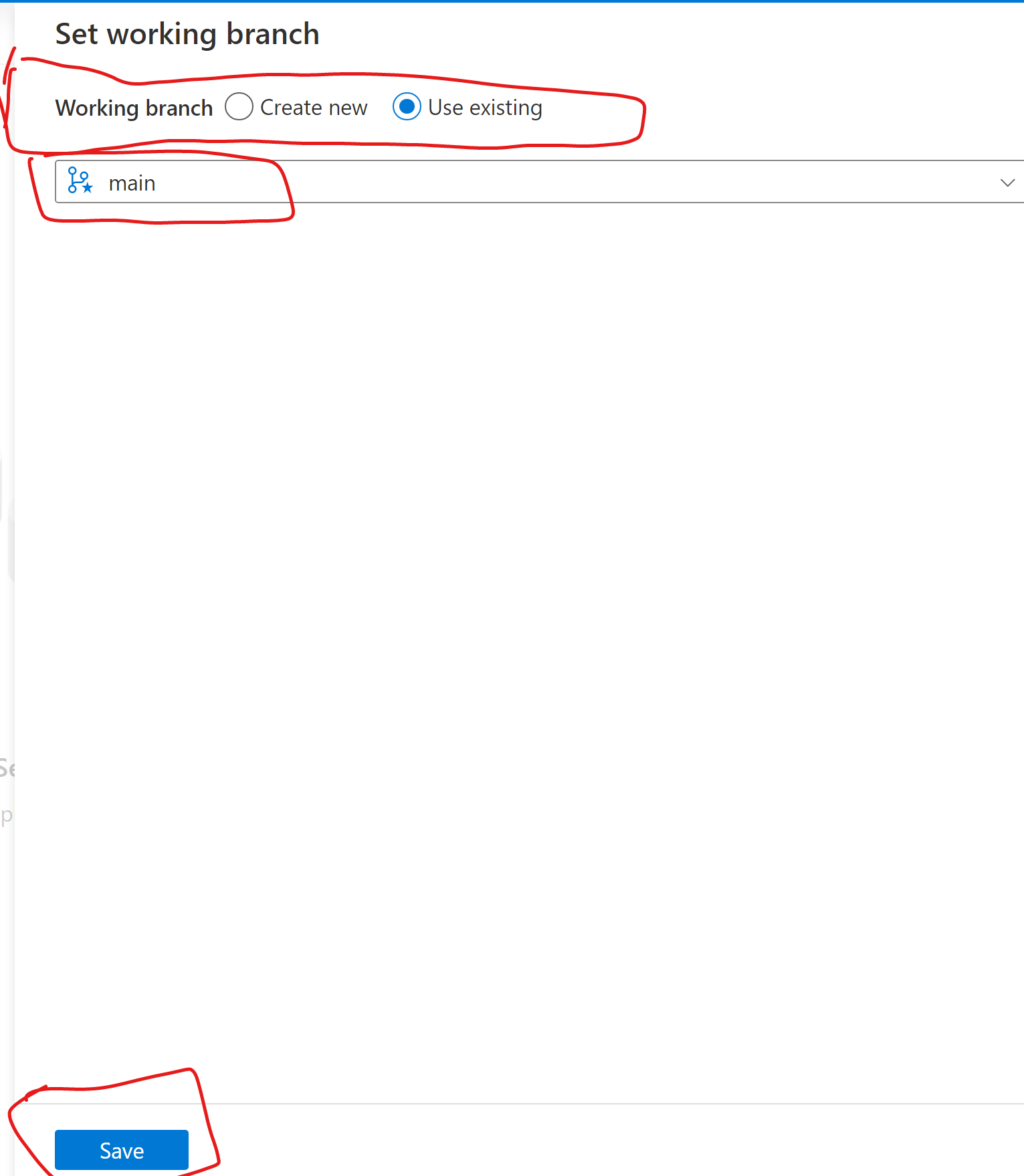


1. Select **Collaboration branch** or you can create new for you.

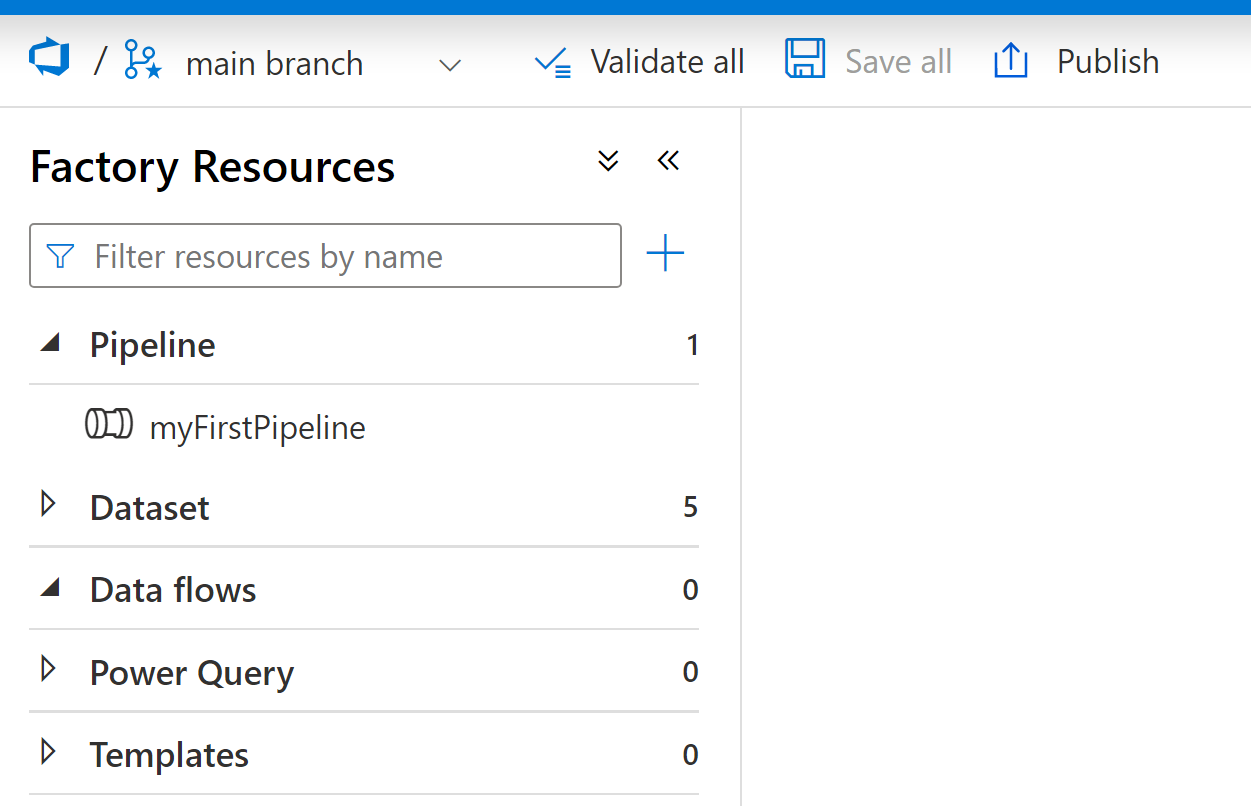
Also select **Publish\_branch**, enter / in Root **folder** and **Import existing resource checkbox**



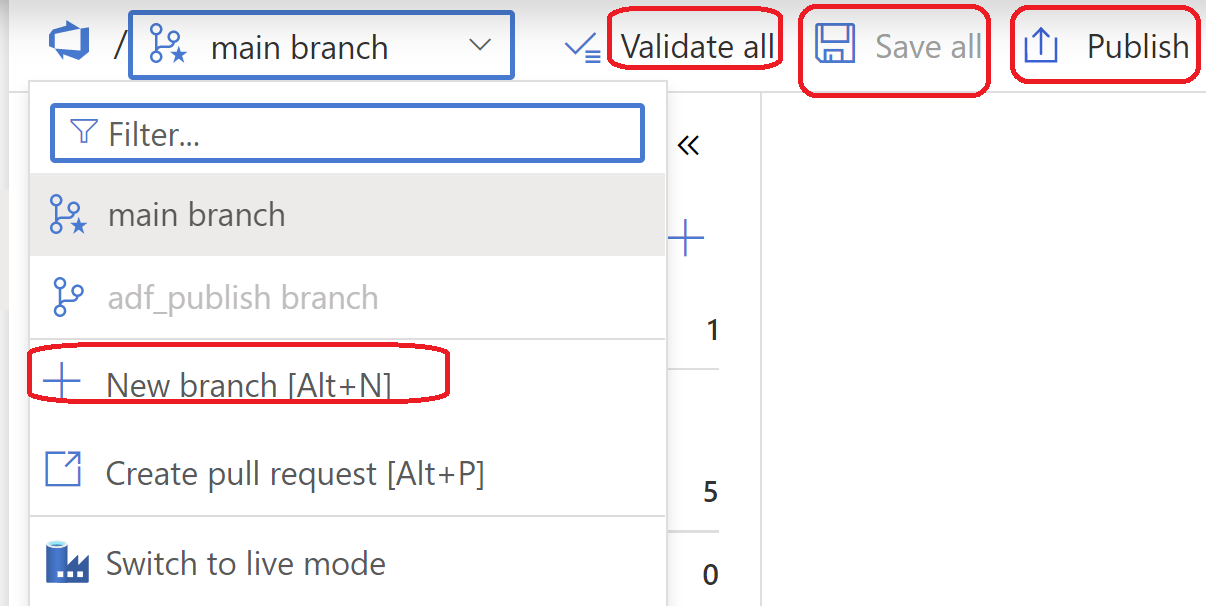
1. Click on **Apply.** Wait for few seconds
2. It will ask to set working branch. Here you can created new branch or use existing as it as shown below. Click on **Save**



1. You will see screen as below



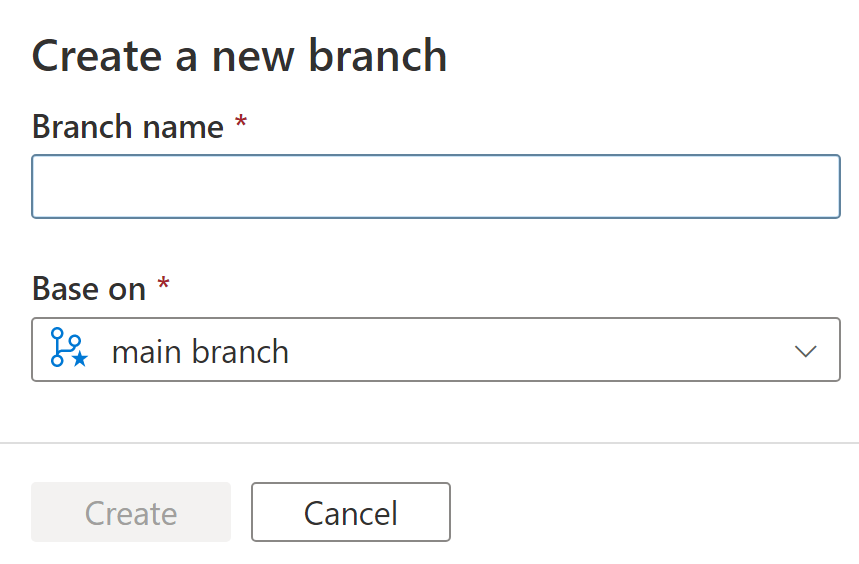
Now you can create **new branch, validate all, save all** and can **publish** them which generate ARM templates for Continuous integration and continuous delivery for CICD.



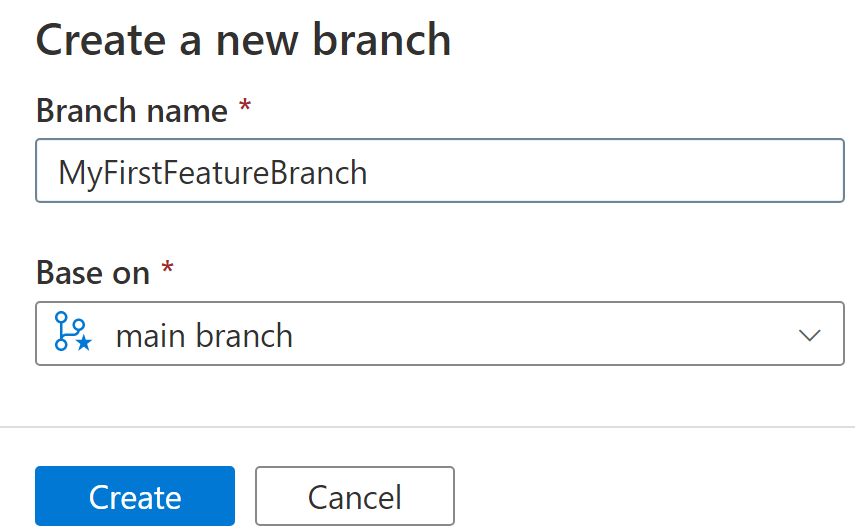
# How to create new Branch using ADF?

Now, we can create a new branch by just few button click away

1. Click on **+ New branch[Alt+N]** on above screen
2. Now, Enter branch name as you would like to give and select based on (By default its Collaboration branch)



1. Click on Create



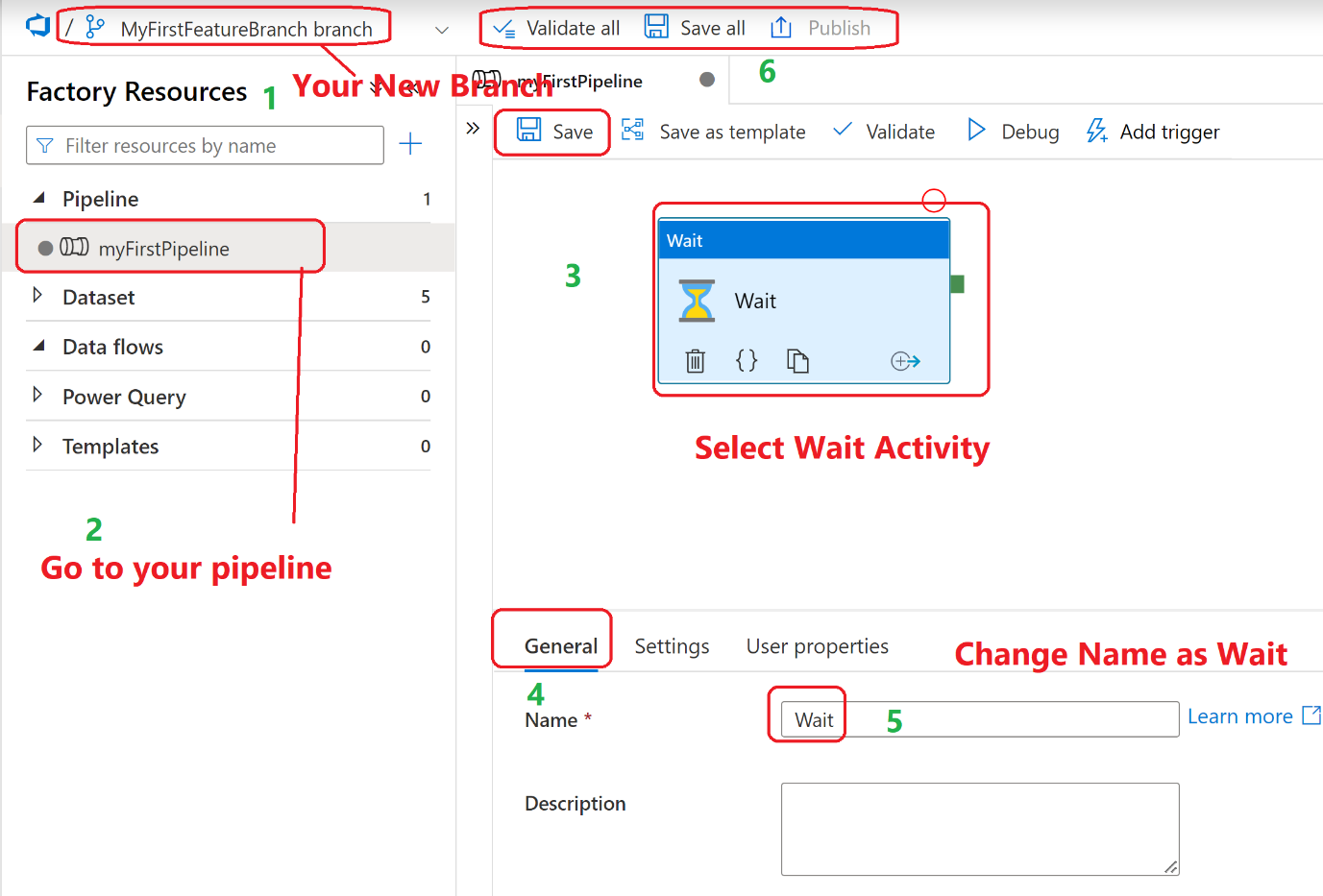
1. Now select your Pipeline which we have created earlier.

NOTE-

1 your feature branch you have just created.

2 Your first pipeline which we have created earlier

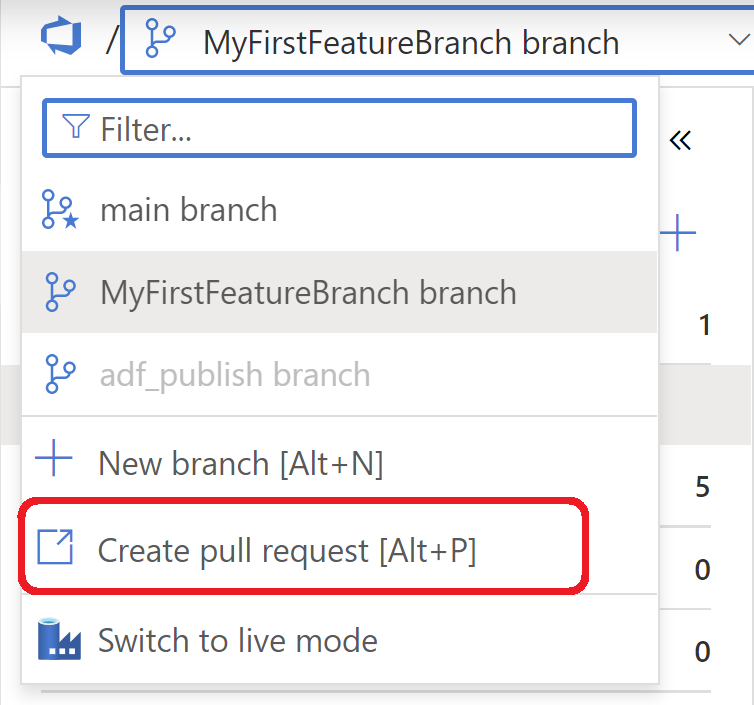
1. your pipeline has wait activity
2. Go to general Activity
3. Renamed it to Wait form Wait 1.
4. Click on Validate all . Once it happy Clicked on Save all and then Publish



# How to create Pull request using ADF?

Now in order to merge you changes into main/Collaboration branch, you need to create PULL REQUEST.

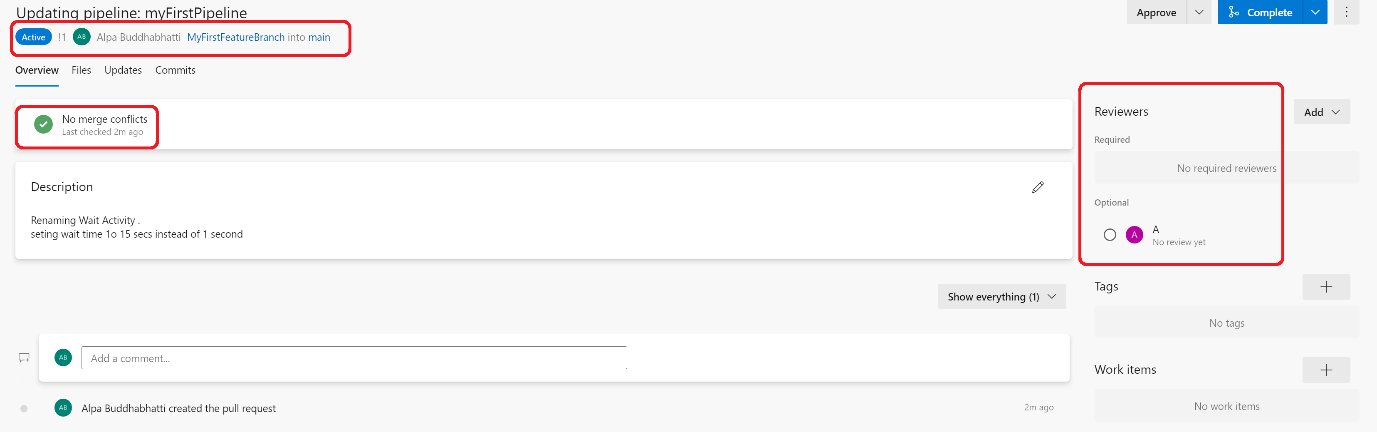
1. Click on Create pull request[All+P] as shown below



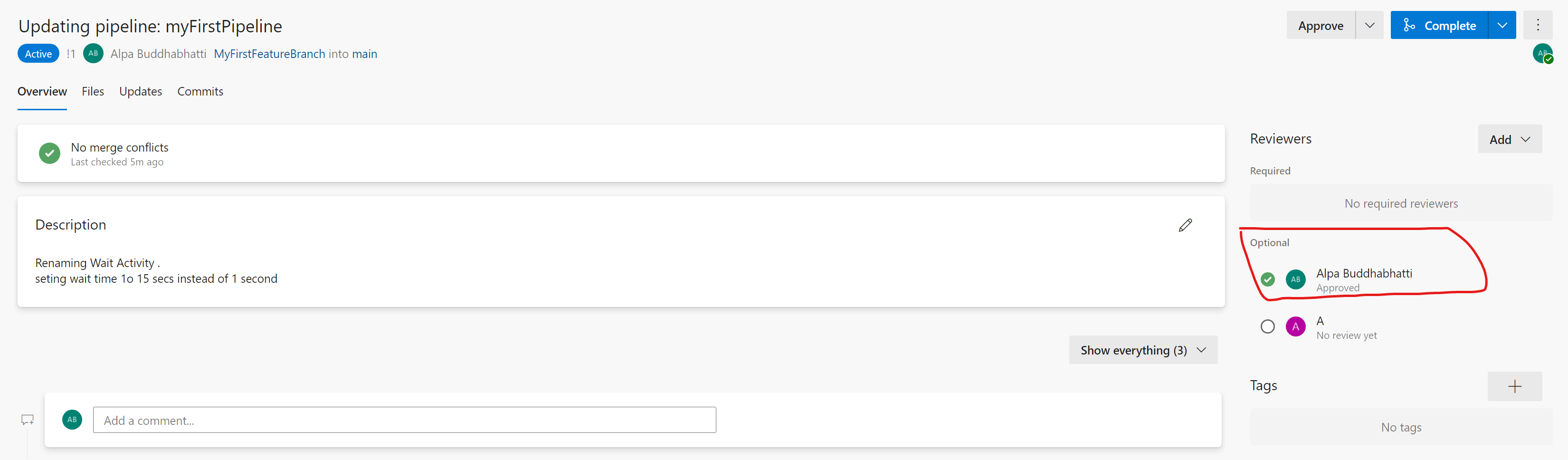
1. Enter title, Descriptions, reviewer, and click Create



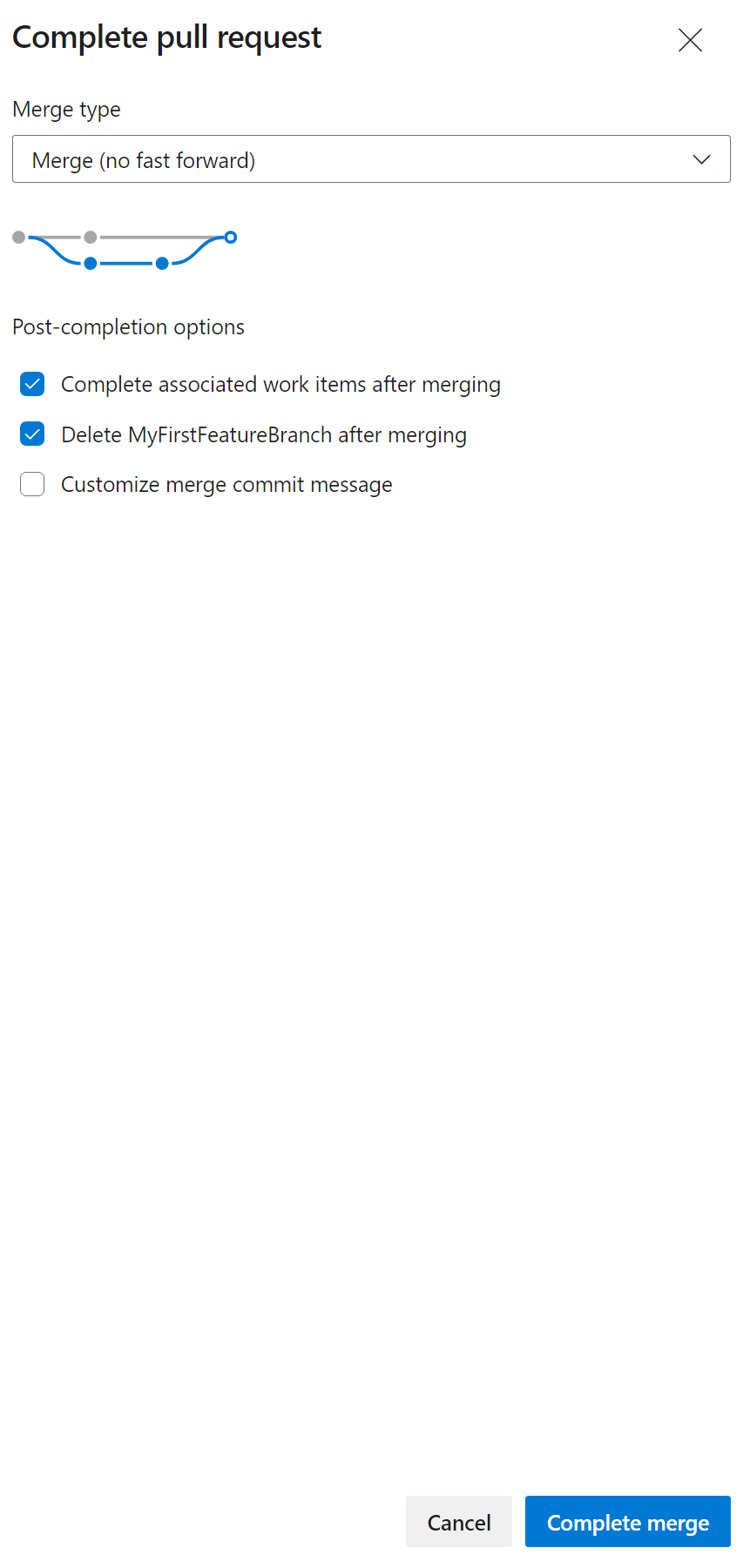
1. You will see below screen. Now your Pull requested need to approved by Reviewer and then you can click on complete



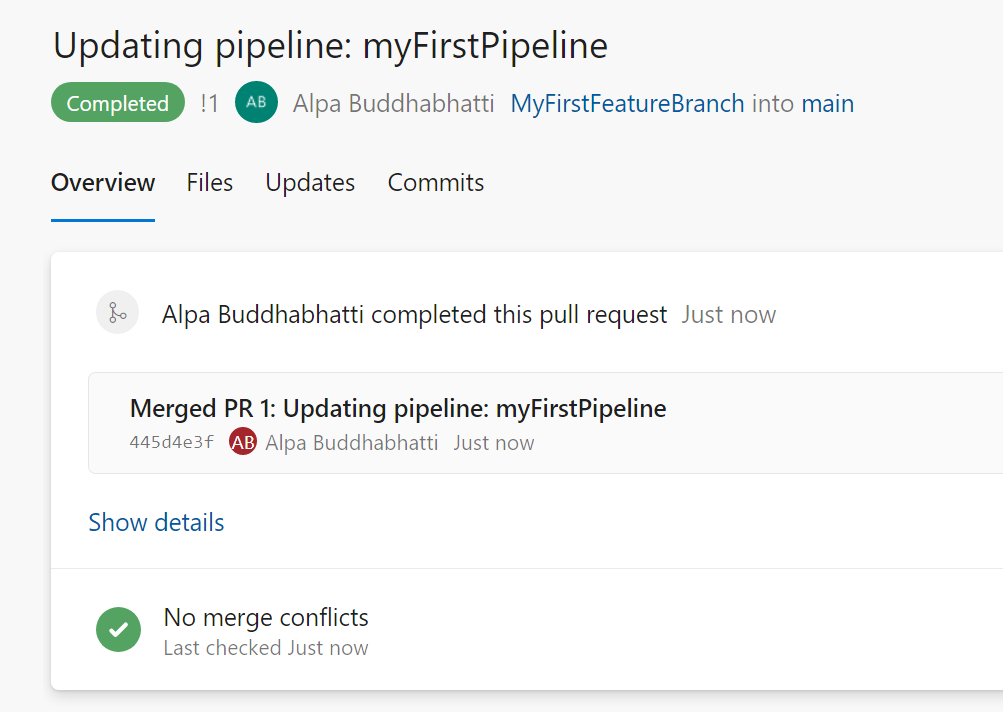
1. It will looks like below, once it approved by reviewer



1. You can complete it so code goes to Main/Collaboration branch. It will looks like as below when you click on Compete button. Click on **Complete merge button** so code goes to main branch

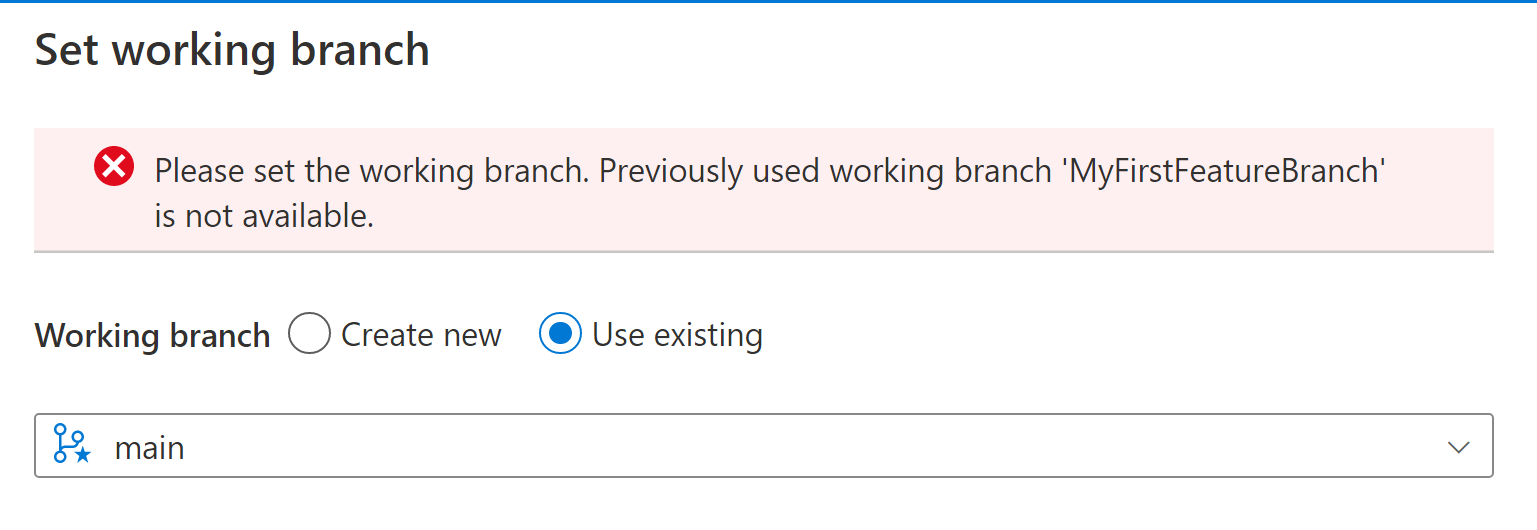


1. Your PR is now completed.

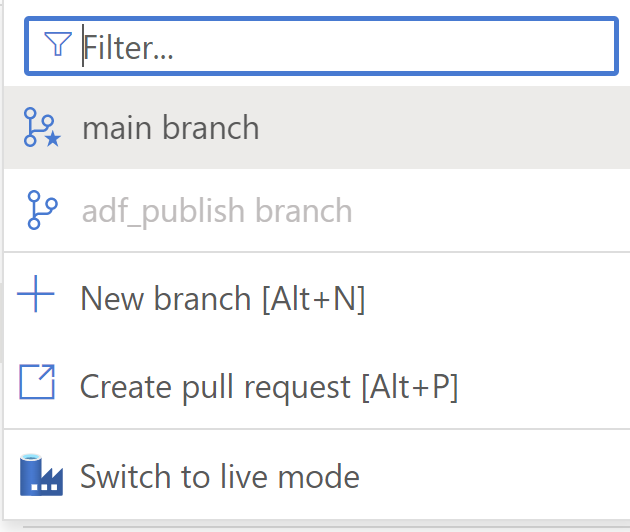


1. Go to ADF

You will see below screen. Click on **save.**



1. Now you can see, your feature branch has deleted, your come is now in main branch. Check pipeline and see Wait Activity its name has changed now





# Data Movement example

**Let’s move data from Azure Blob Storage to Azure Data Lake Gen 2**

**Here Source is Blob Storage (.csv file)**

**Sink/Destination is Azure Data Lake Gen 2 (.csv file)**

## How to create Linked Service in ADF?

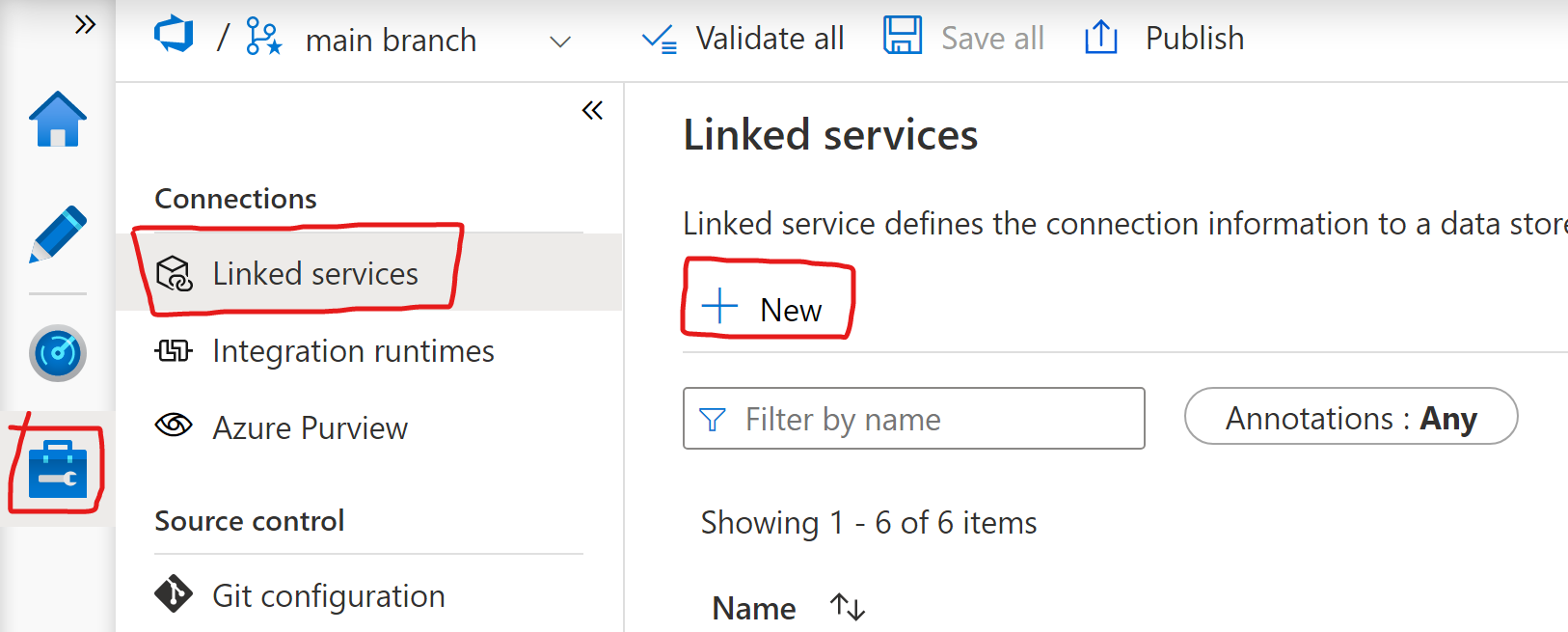
It is just like a connection manager who hold the connection string for your datasets.

It authenticate your data source so it can be used in ADF by Activity, Dataset and Dataflow. Go to Data Factory

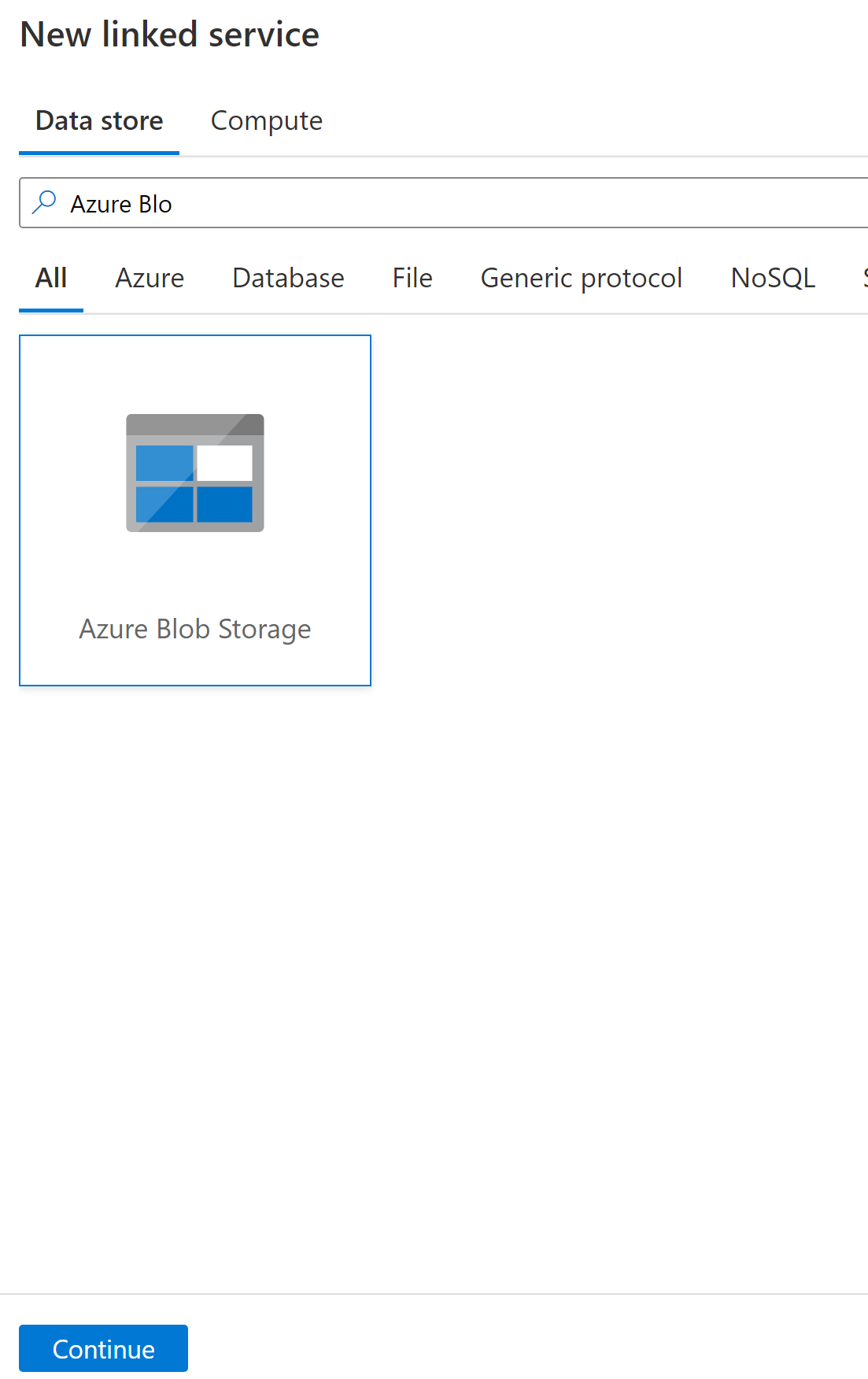
### Linked Services to pointing to Azure Blob Storage

1. Go to Manage tab and click on + New to create new linked services

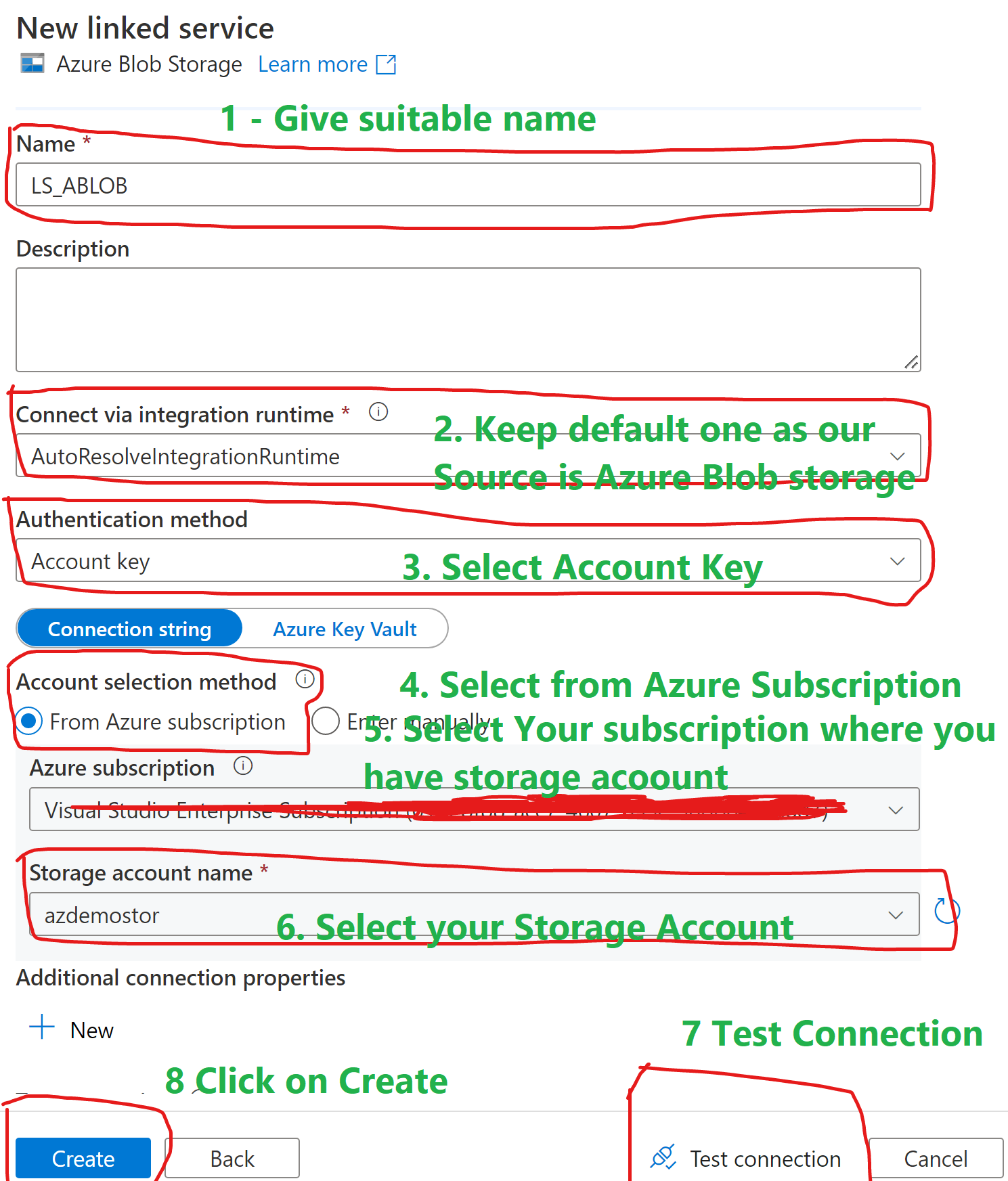
**Manage tab=> Linked Services => + New**



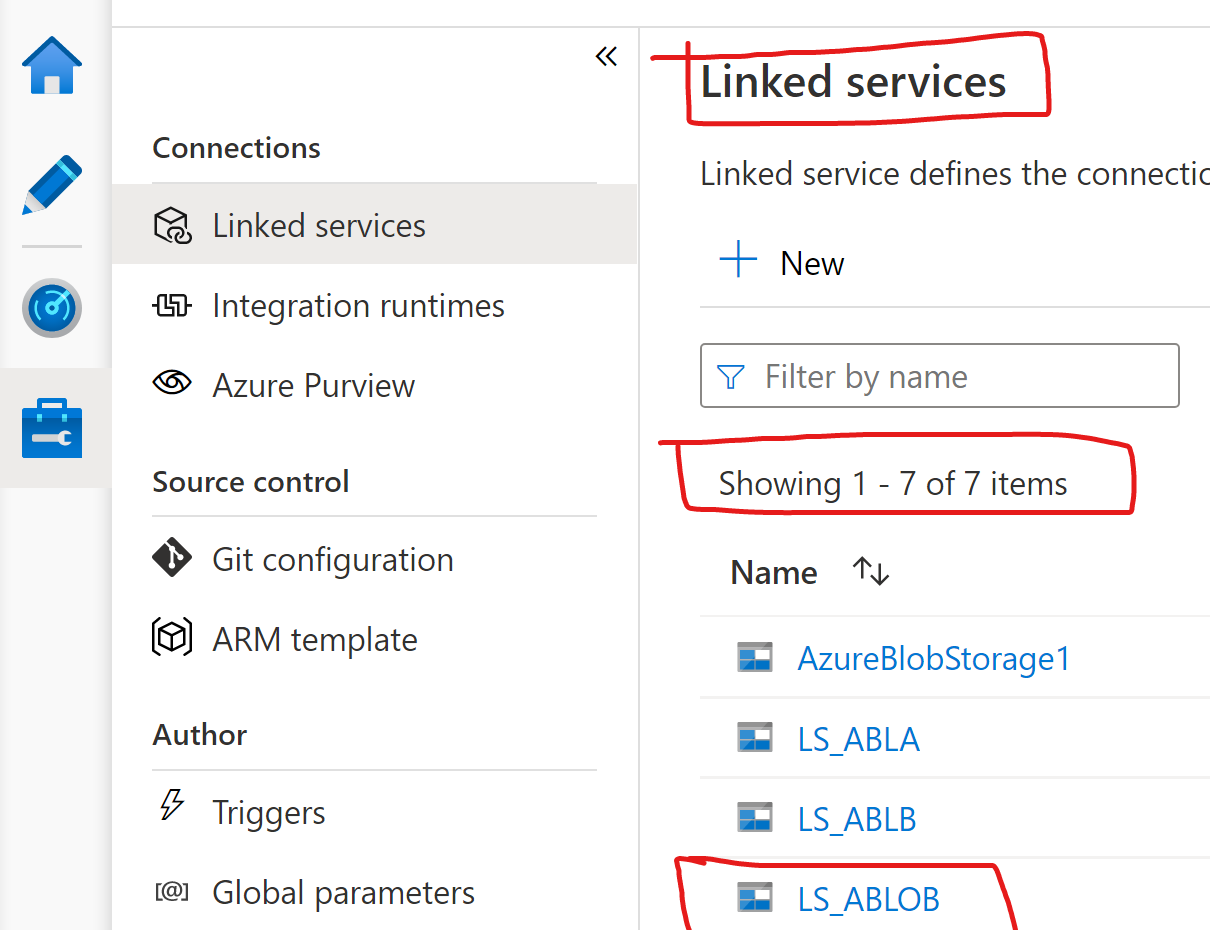
1. Search **Azure Blob Storage** in Search box as below. And Select it and click **Continuous**



1. Give suitable name
2. Select Integration
3. Select subscription
4. Select Storage account as source Linked services
5. Select Access Key
6. Click on Test to check its working
7. Click on Create



1. Once it has created, it will appears to Linked Services

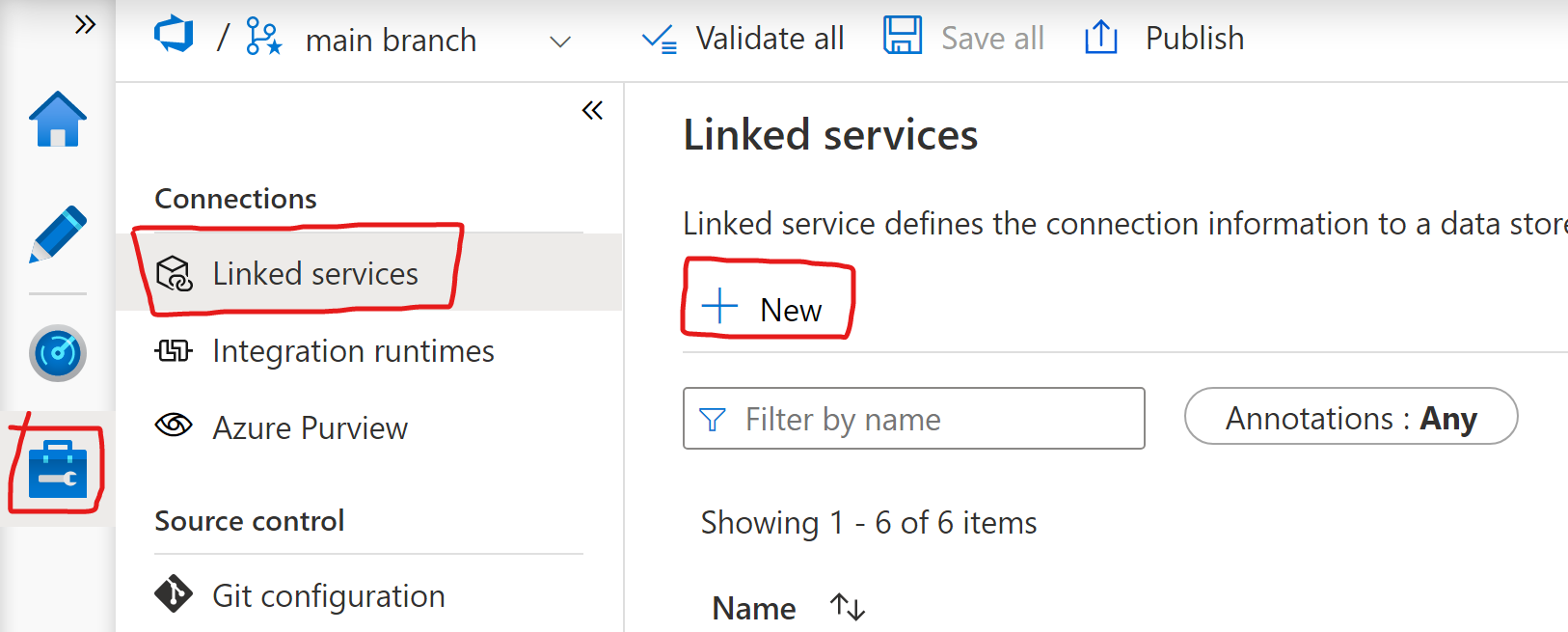


1. Congratulation you have created Linked Services which is pointing to Azure Blob Storage.

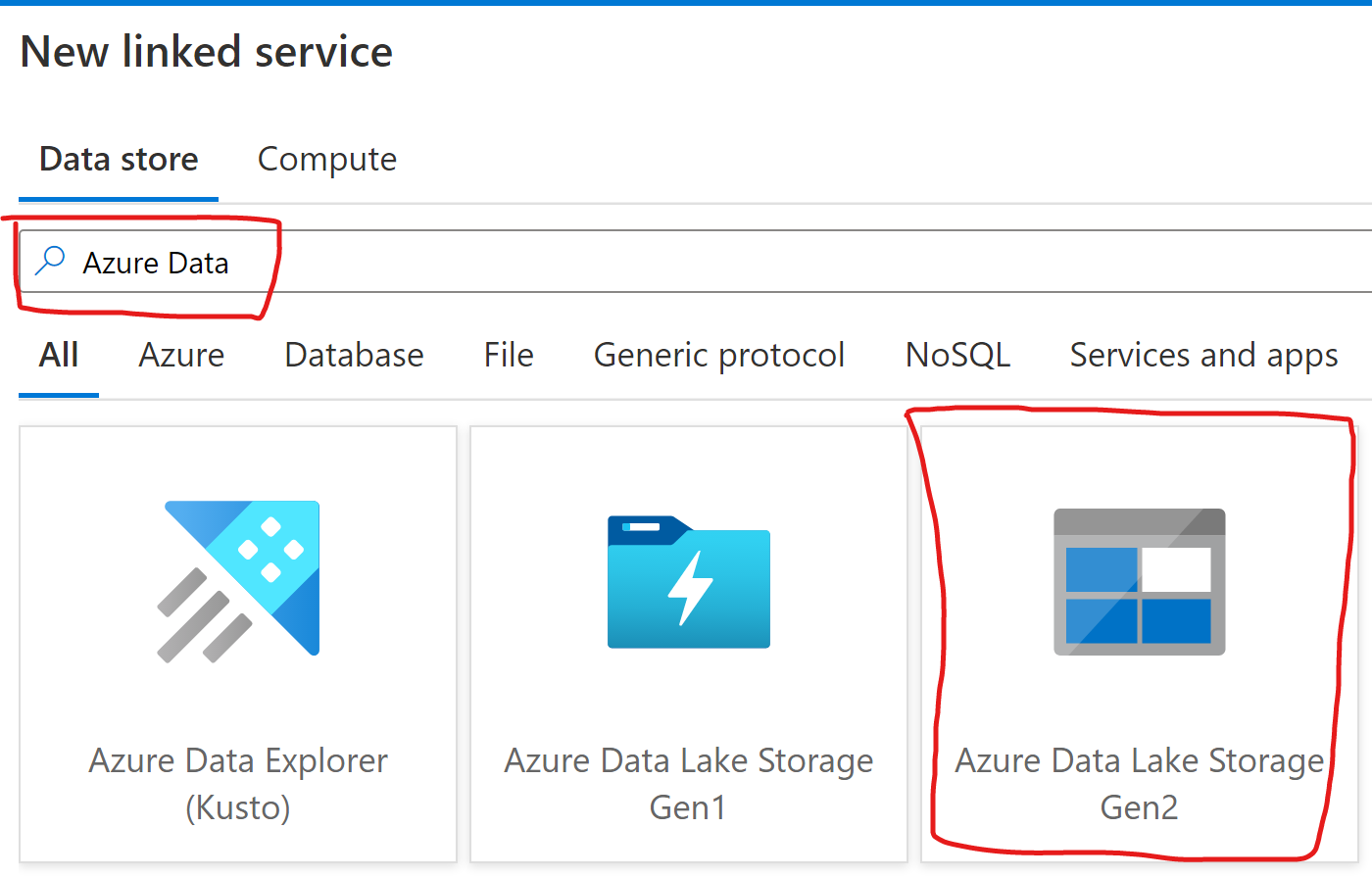
### Linked Services to pointing to Azure Data Lake Gen 2

1. Go to Manage tab and click on + New to create new linked services

**Manage tab=> Linked Services => + New**



1. Search **Azure Data Lake Gen 2** in Search box as below. And Select it and click **Continuous**



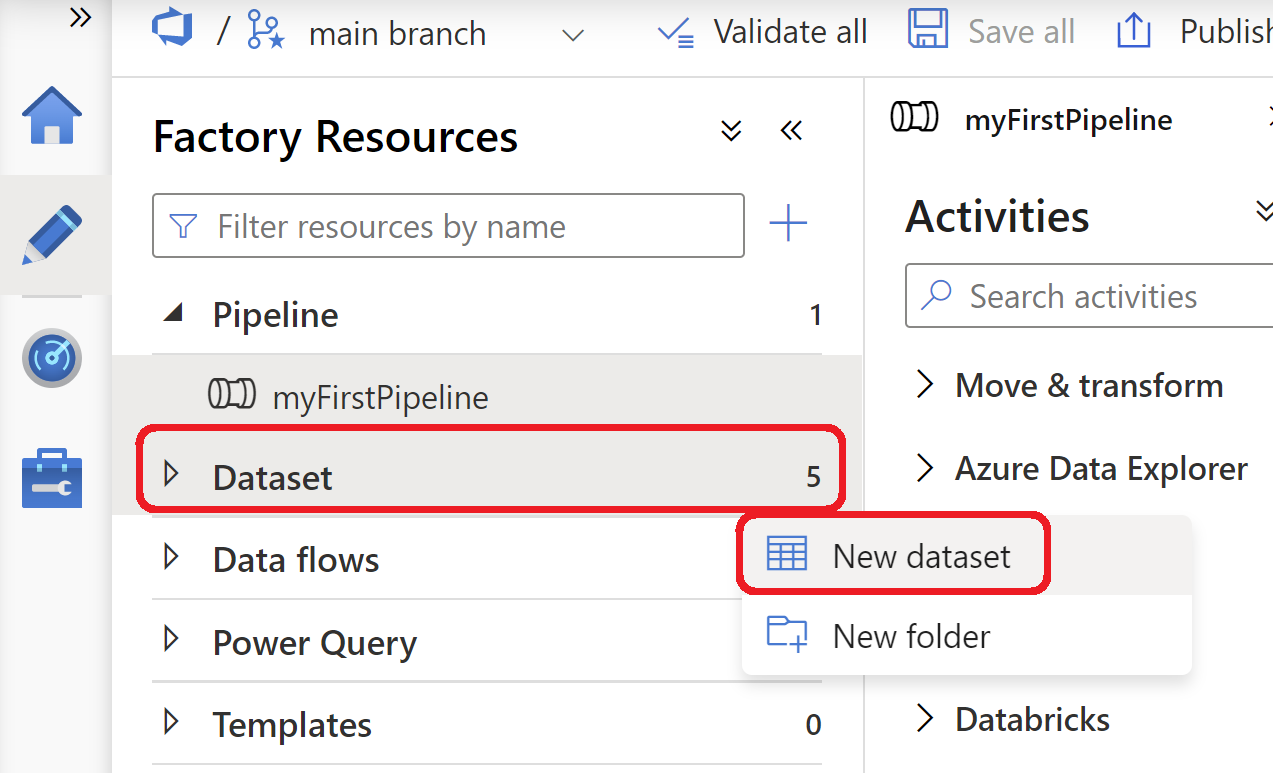
Remaining steps you can follow as above Linked services. End of this exercise you have two links services.

1. Point to Azure Blob Storage
2. Point to Azure Data Lake Gen 2

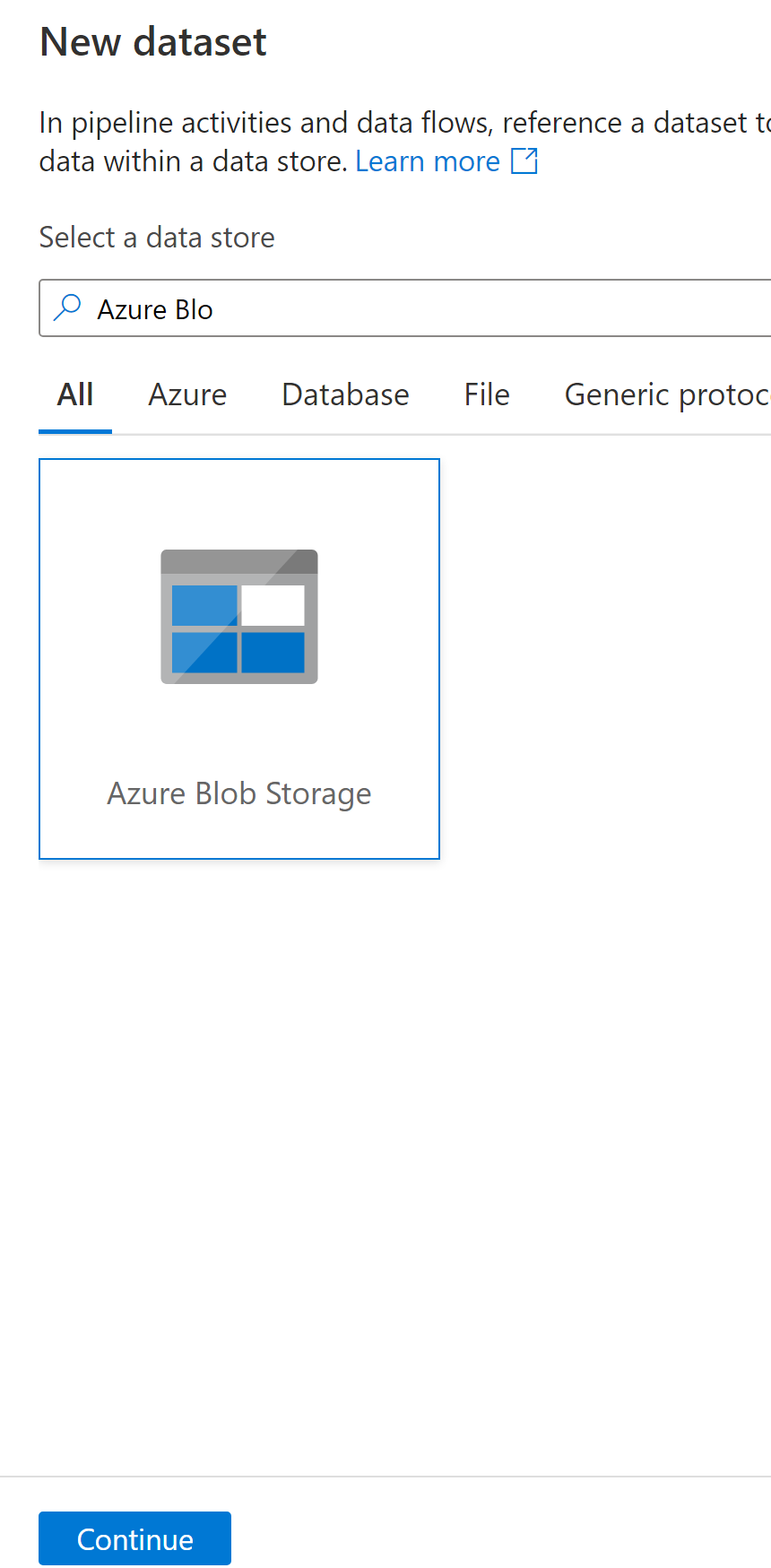
## How to create Dataset in ADF?

#### Now let’s create datasets for Source Data File (Azure Blob Storage)

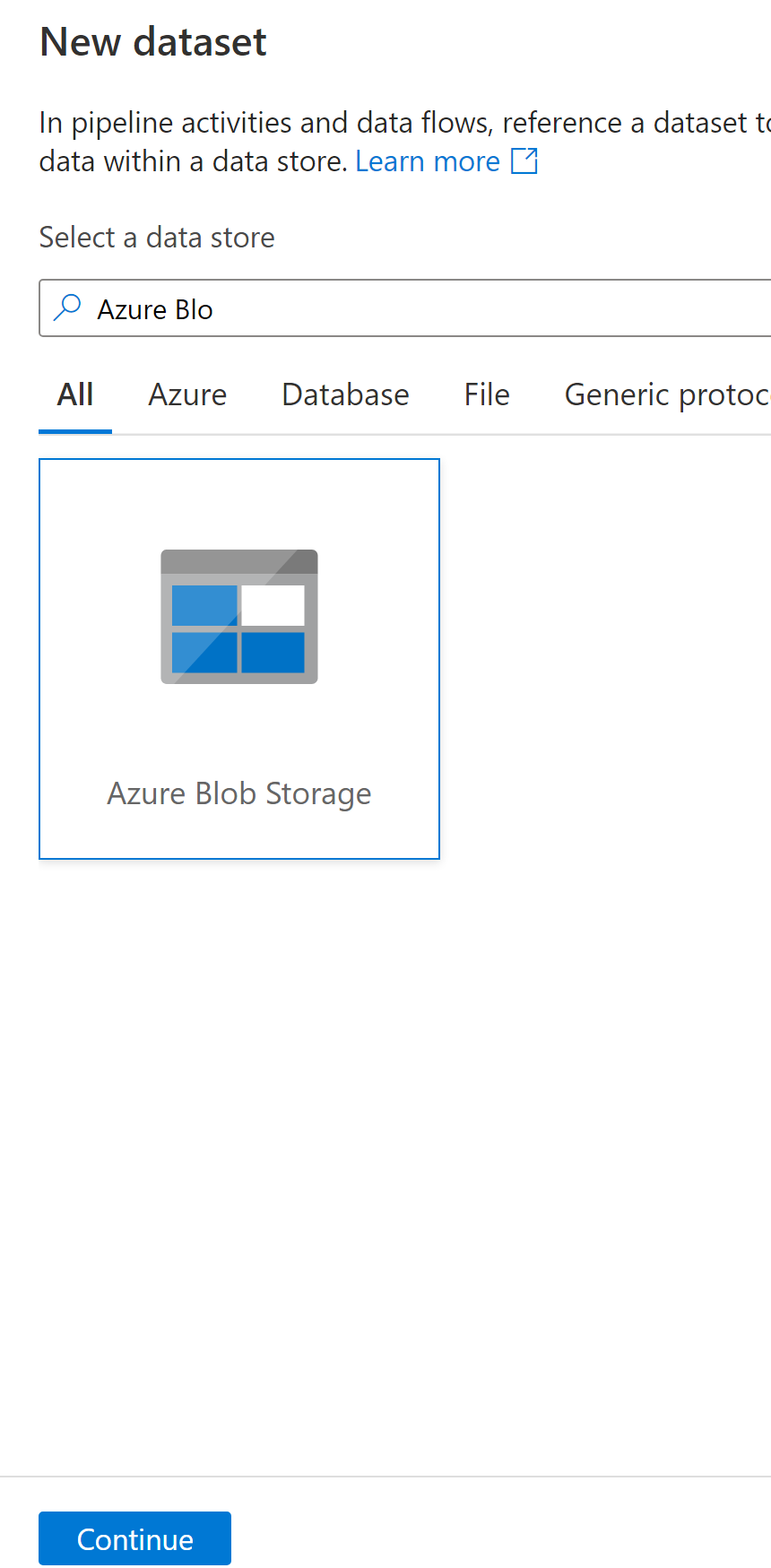
1. Right Click on Dataset => select New dataset



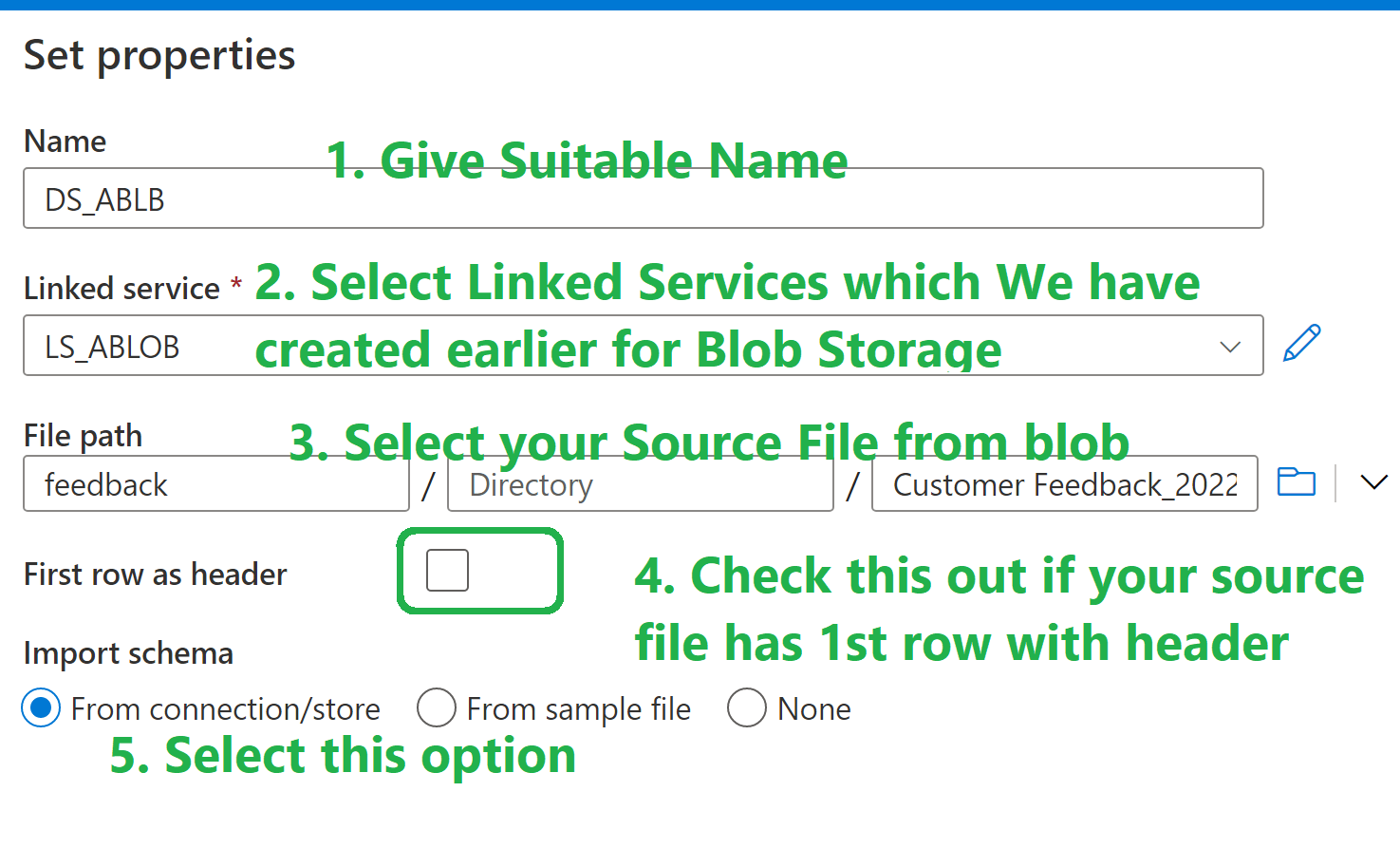
1. Now select Data Source, Search for Azure Blob and then select and Clink on **Continuous**



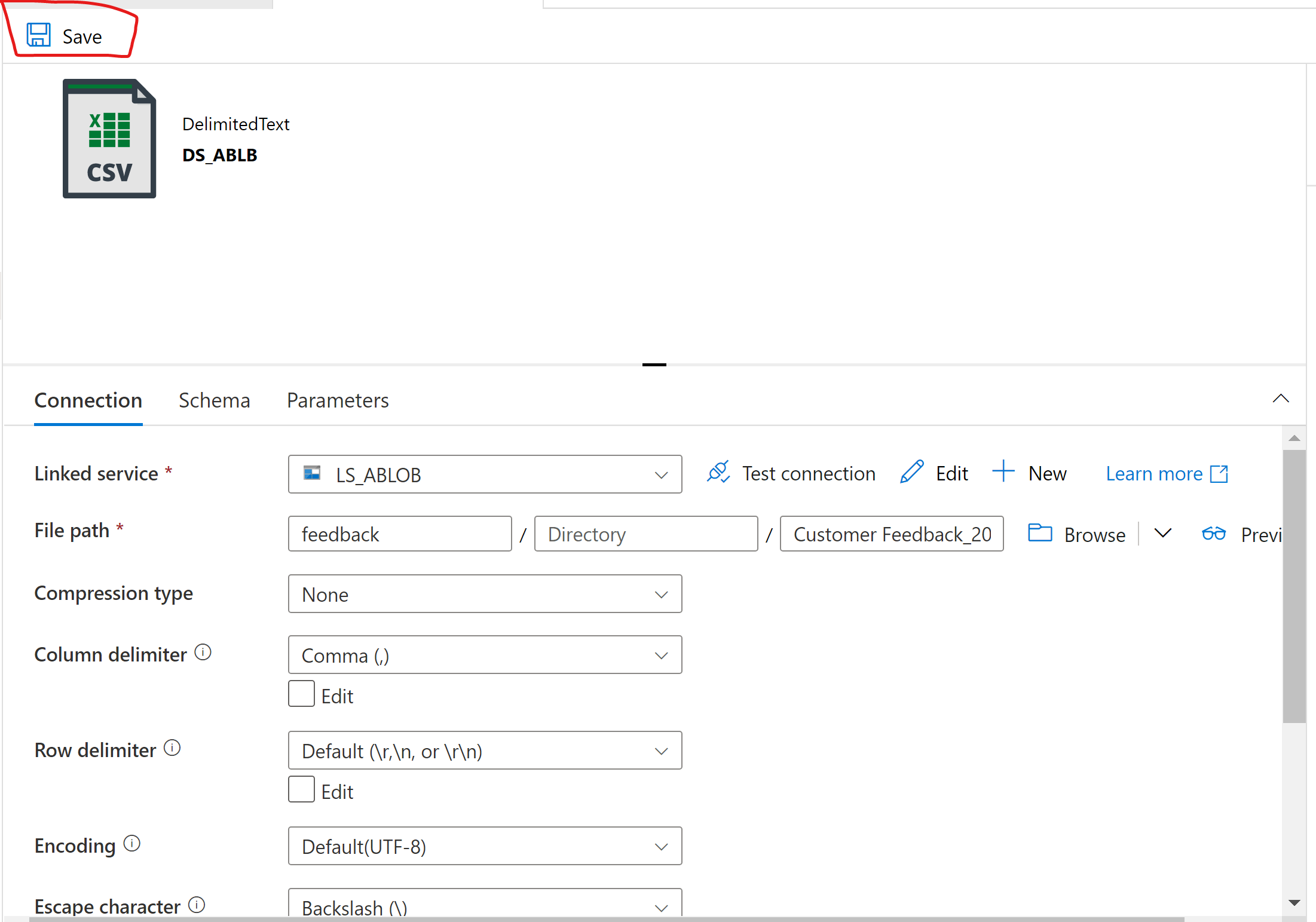
1. Select File Format .csv and **Continue**



1. Select correct file from blob Storage as shown below



1. Click Ok
2. Click Save



1. Now you have created new dataset who hold metadata in form of csv for your Source File

#### Now let’s create datasets for Sink Data File **(Azure Data Lake Gen 2)**

1. This time create a dataset for Azure Data Lake gen 2. Follow the above steps from 1 to 3. Here instead of **Azure Blob Storage**, select **Azure Data Lake Gen 2** as Source type.
2. Click ok
3. Now, give suitable name, Select Linked Service which is related to **Azure Data Lake Gen 2**, **Import Schema**

****

1. Click Ok
2. Click Save

**Now you have two datasets created**

1. **Dataset 1 have metadata for Azure Blob Storage**
2. **Dataset 2 have metadata for Azure Data Lake Gen 2.**

## How to create a pipeline with Copy Activity in ADF?

Now add Copy Activity to your Pipeline.

This activity move data from Azure Blob Storage to Azure Data Lake Gen 2

You can enhanced it by adding parameters and making it dynamic so it can support more than 1 files.

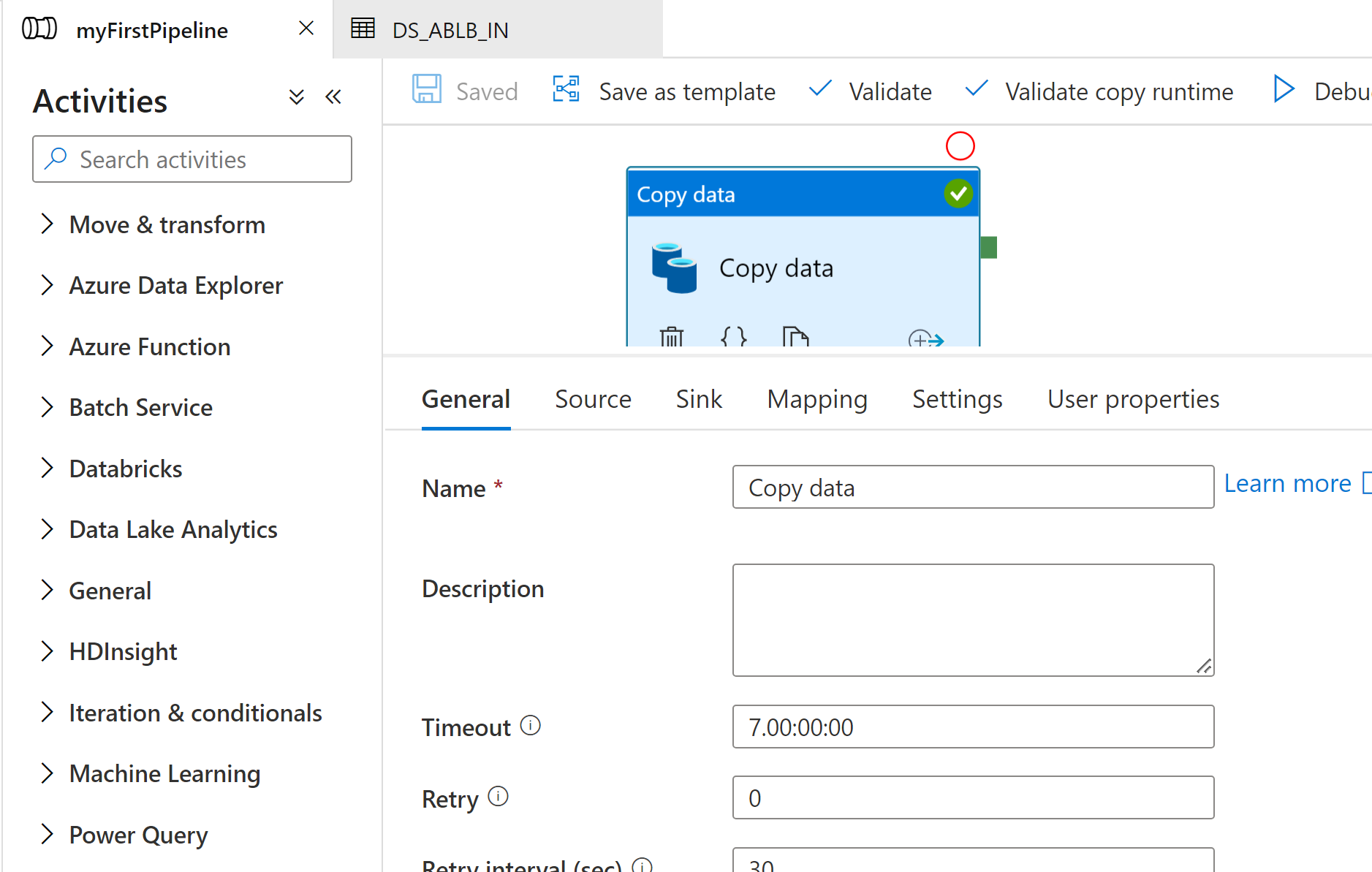
Also if you add meta data activity and for each loop activity with copy activity then one single pipeline and two datasets can supports more than 1 source file .

It can move any file from source to sink.

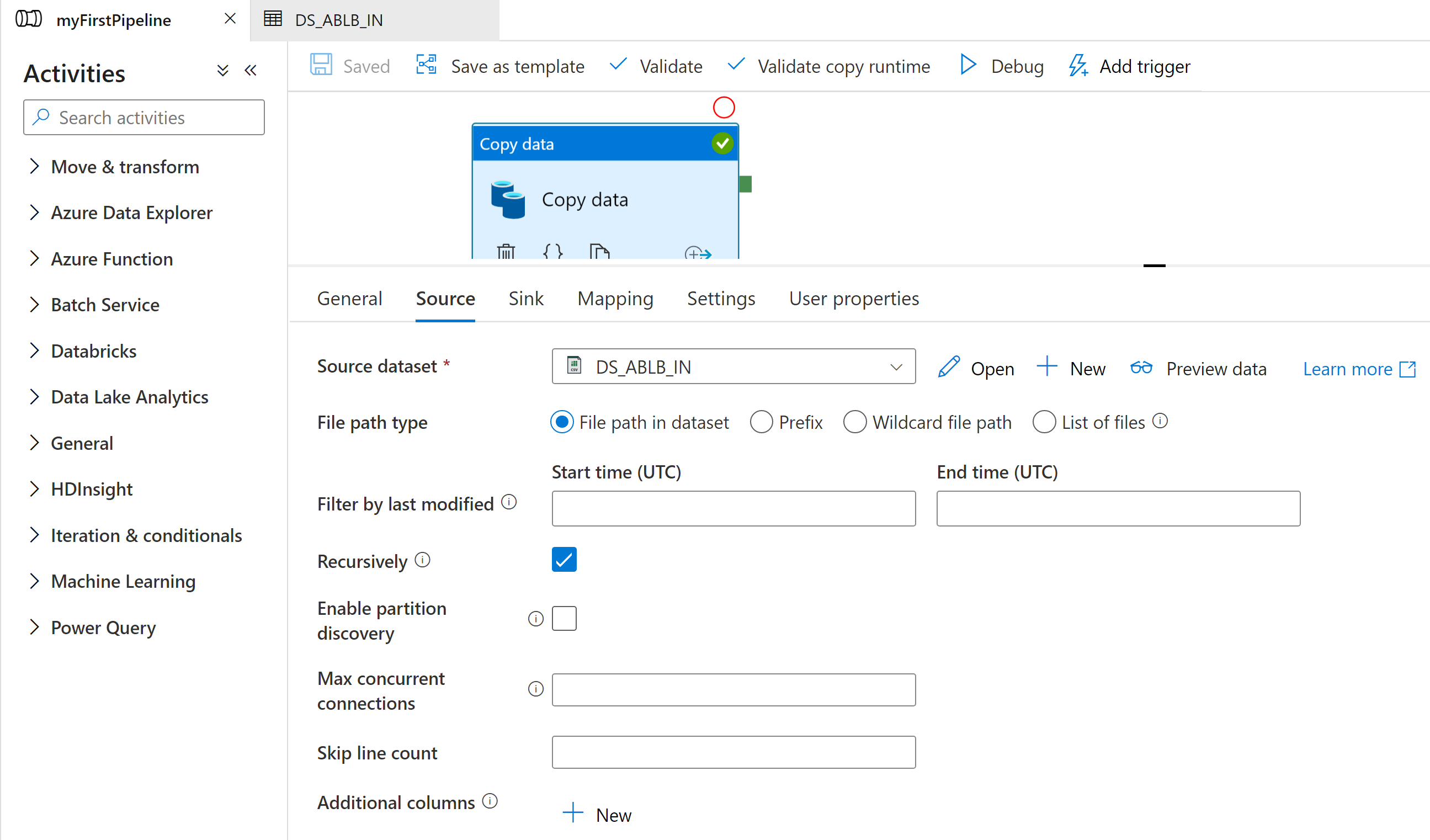
Now add Copy activity to pipelines as shown below



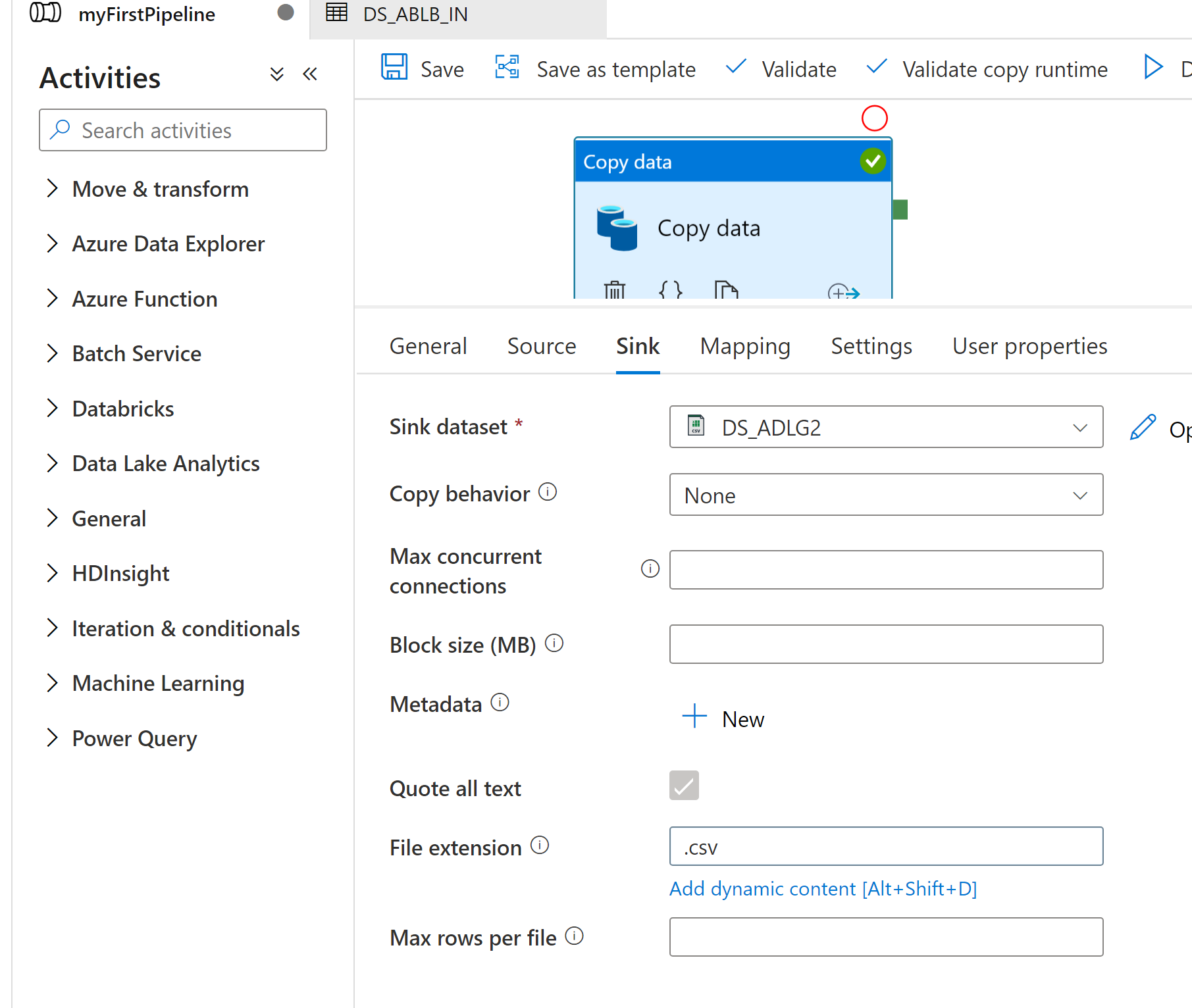
Select Copy activity, On General tab give suitable name



Select Source as below

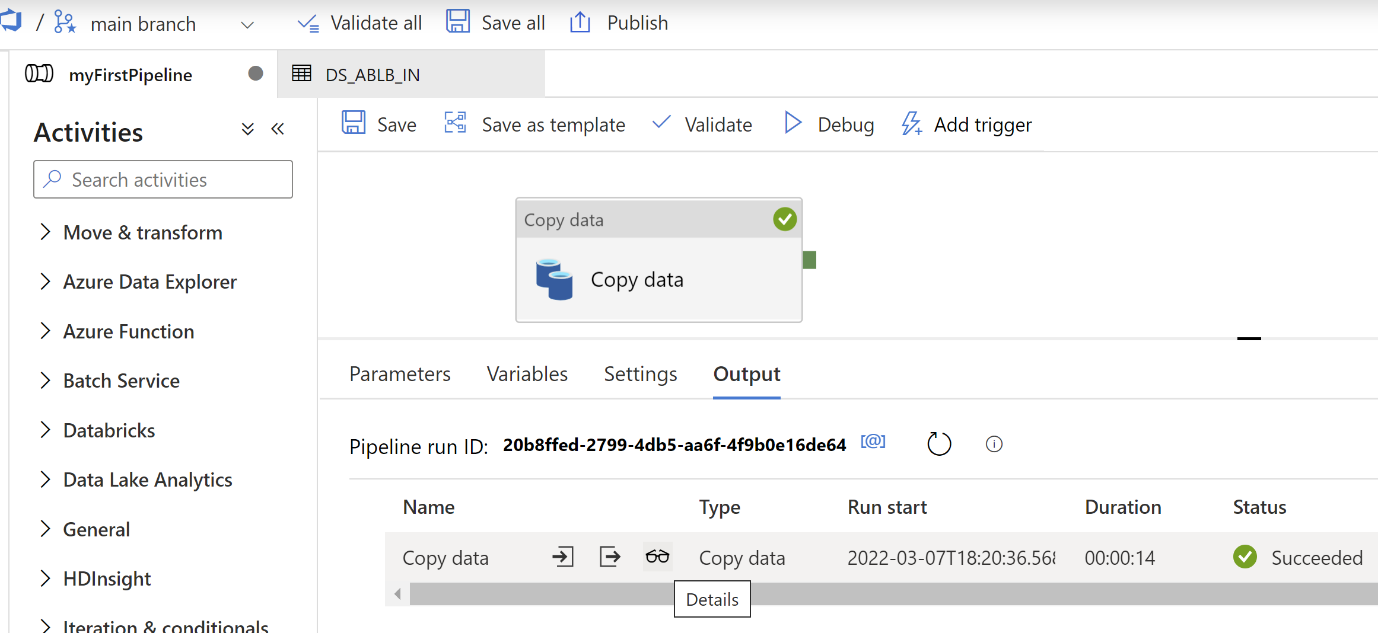


Select Sink As below

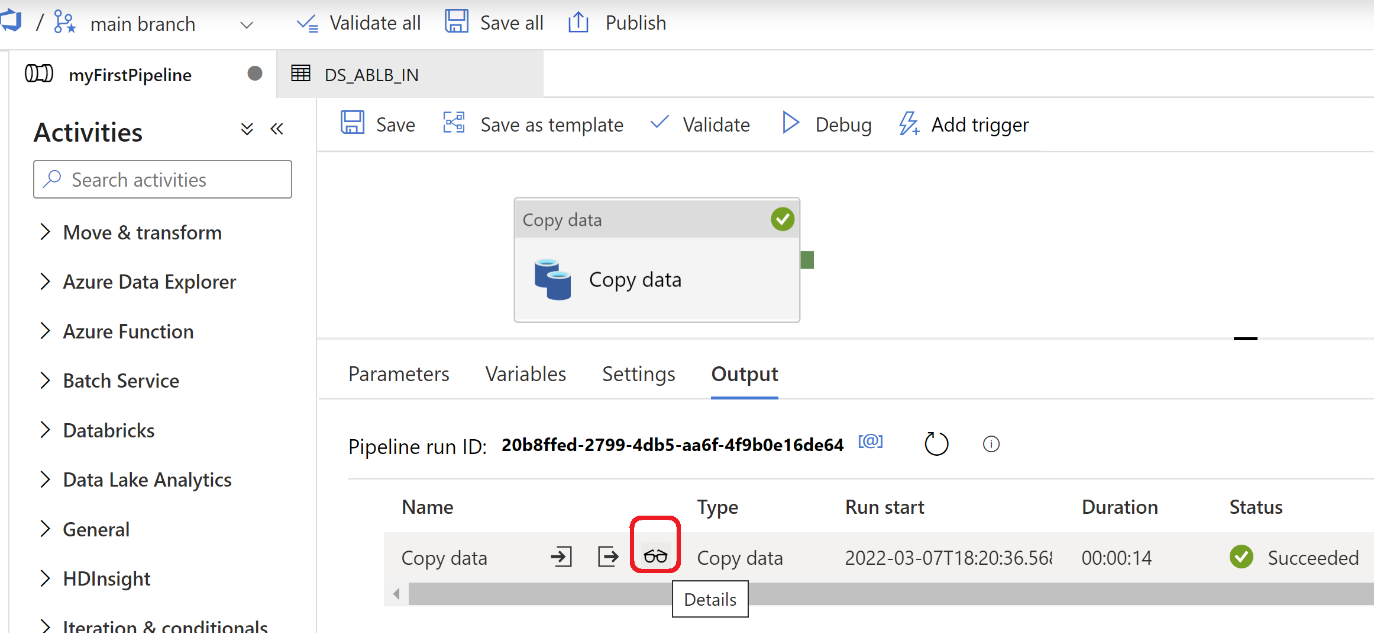


Click on Save

Click on Debug



You can monitor it by just clicking on Spec Icon



Congratulation, you have successfully move file from Azure Blob Storage to Azure Data Lake Gen 2