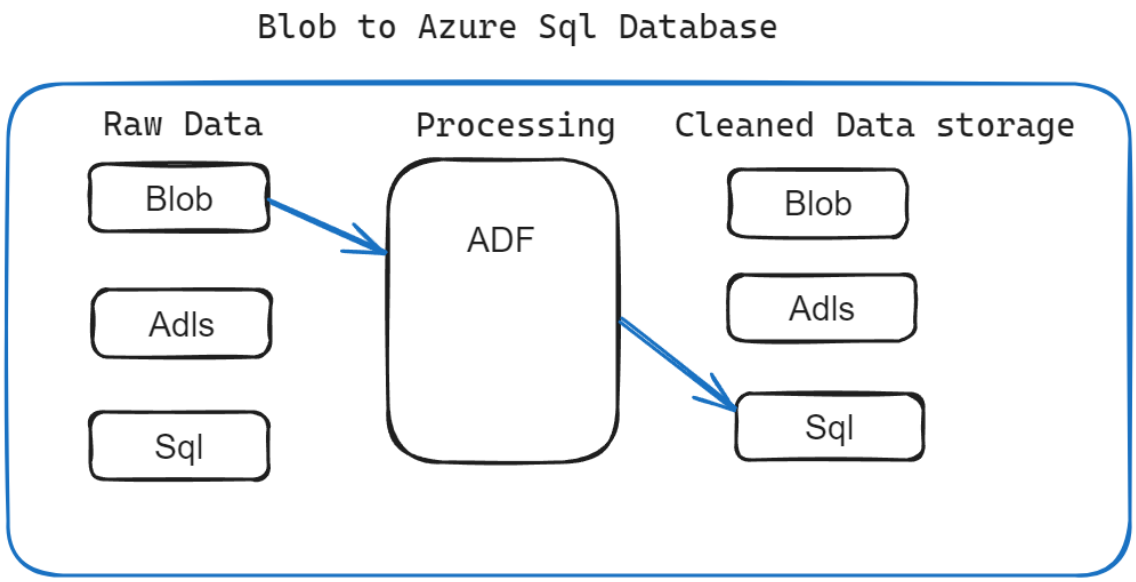
Azure Data Factory pipeline1

Architecture of pipeline1:



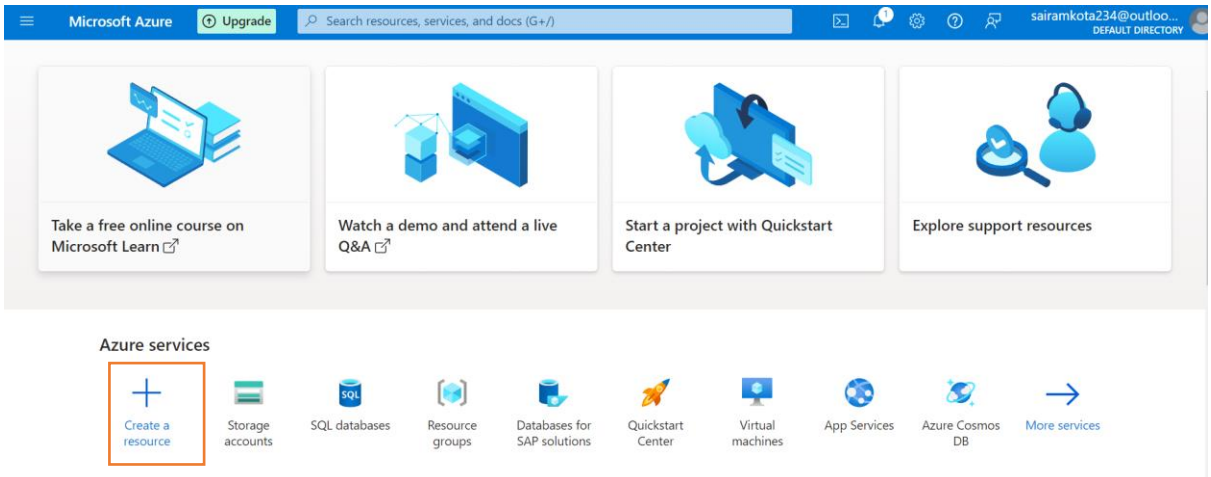
Description:

Etl process to transfer data from Blob storage to Sql Data Bases Using Azure Data Factory (ADF).

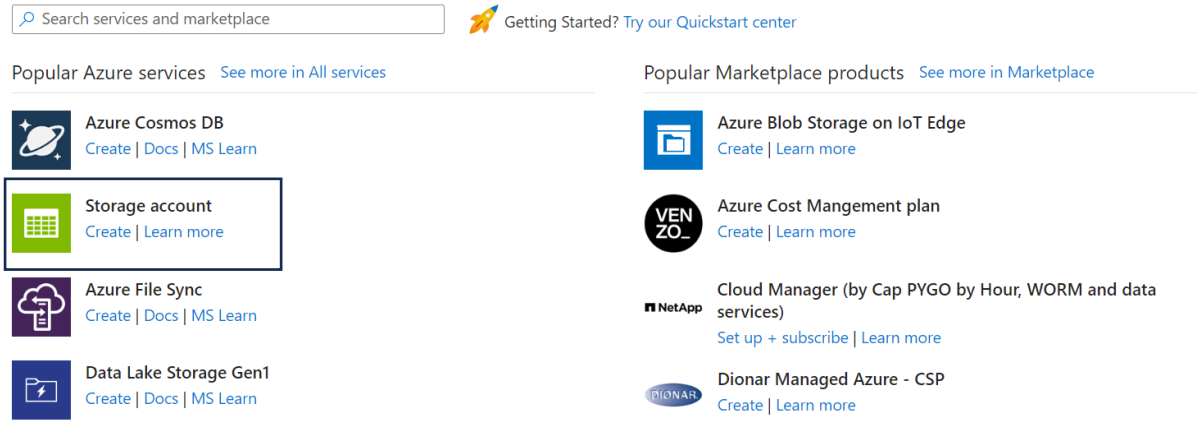
Azure Services Used:

- Azure Blob Storage
- Azure Sql Database
- Azure Data Factory

Creating Azure Bob Storage:

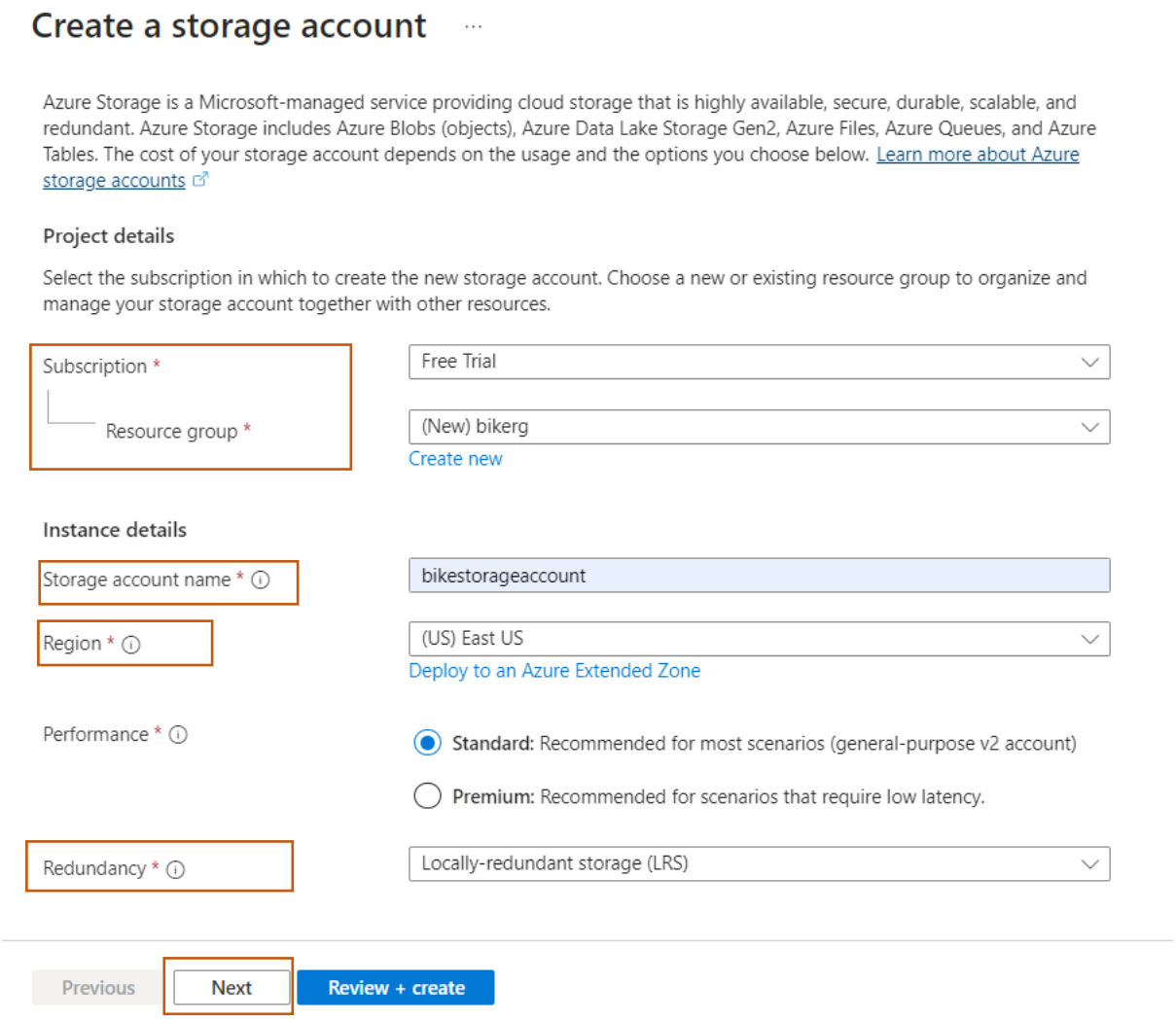


Click on Create Resource and click on storage and create Storage account name should be unique.

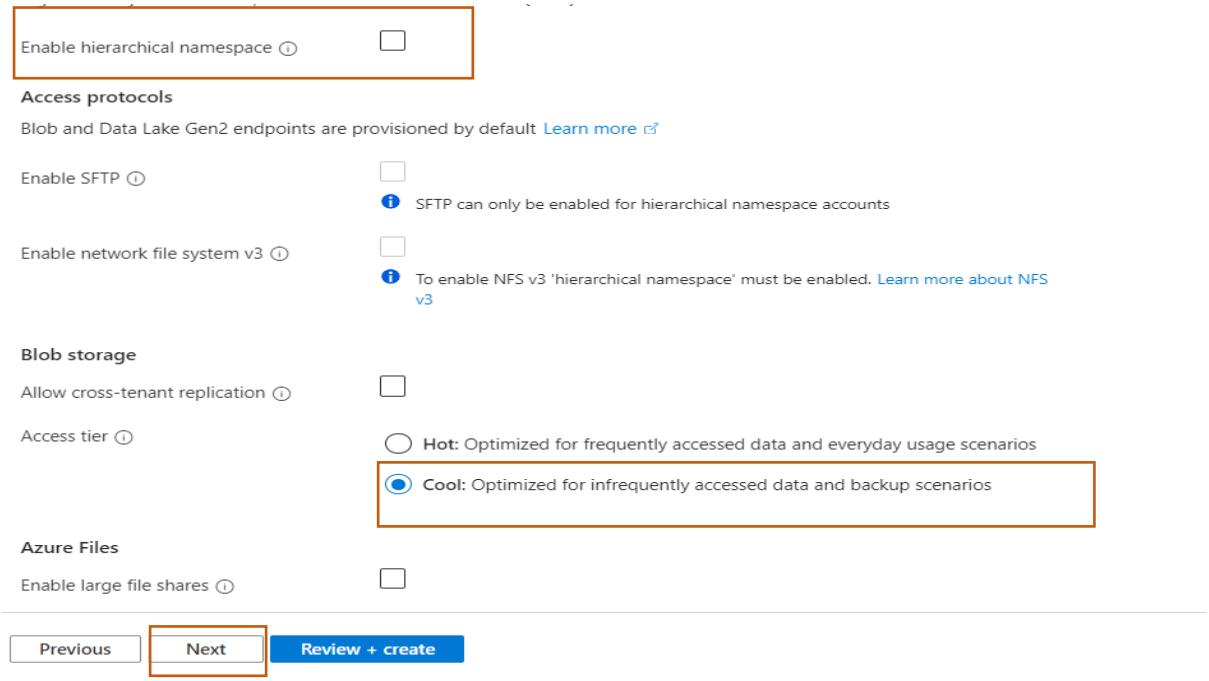


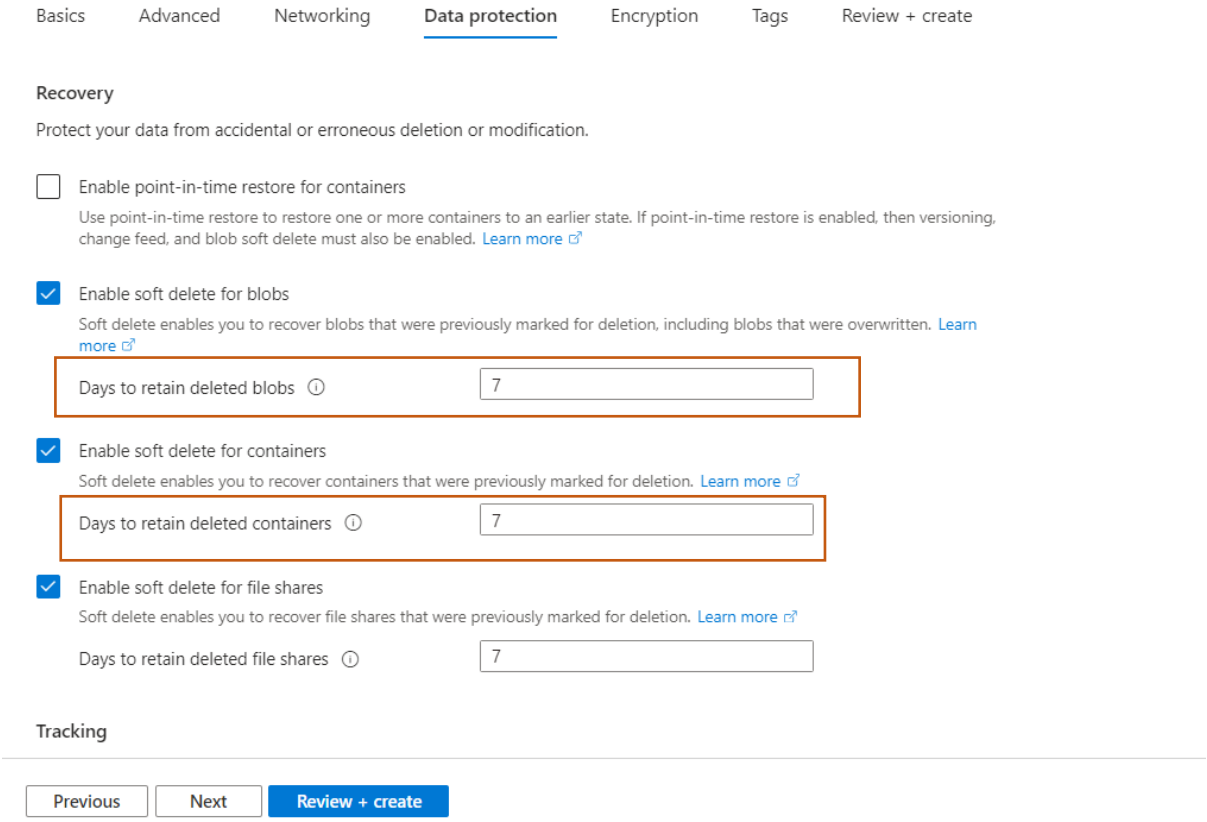
Click on Storage Account and fill all the mandatory details.

- Click on the type of Subscription.
- Create a resource group if already exists select one.
- Give Storage account name.
- Choose the Region and Performance.
- Select type of Redundancy
- Need to create a type of File systems then click on Enable hierarchical namespace.
- Choose type of Access tier
- If want to change the type of delete data protection of days for **soft delete blobs or soft delete containers gives the required days**.



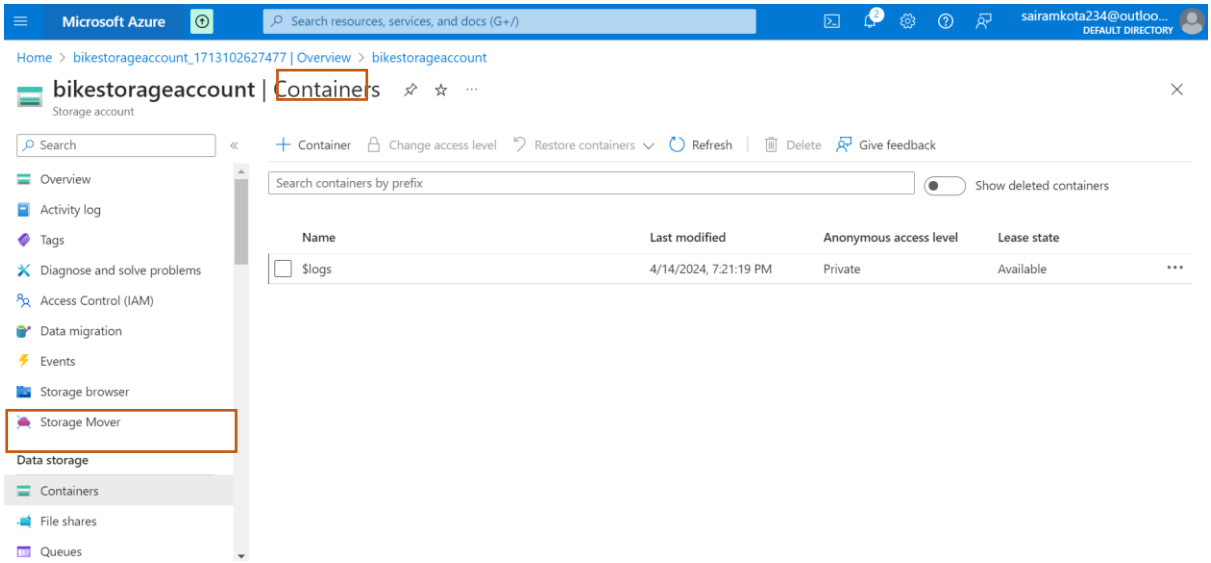
If you choose Enable hierarchical namespace it will create Azure Data Lake Storage.



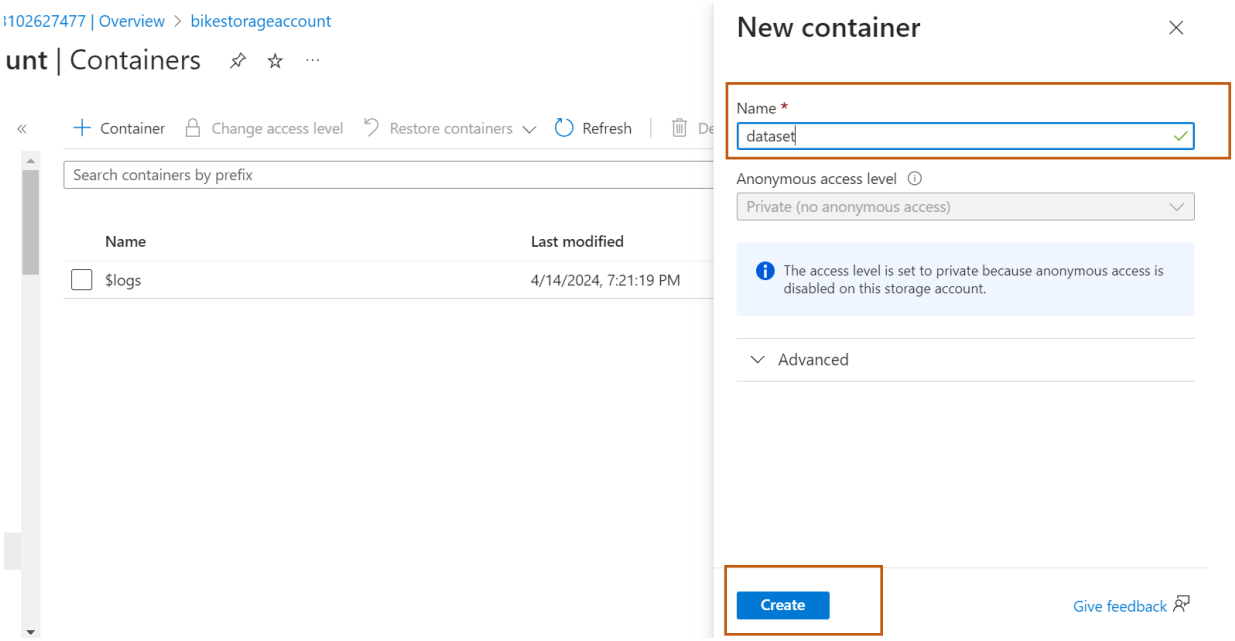


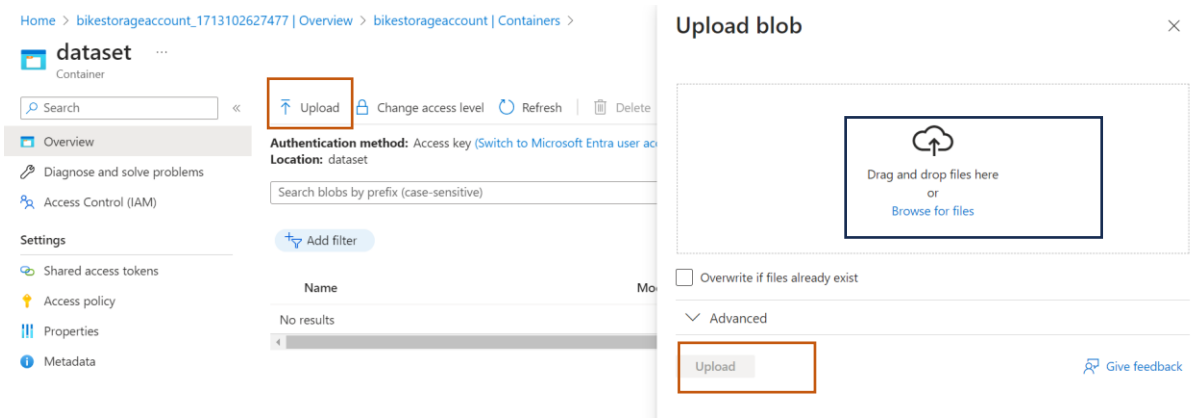
If we have tags, please add it, and create it.

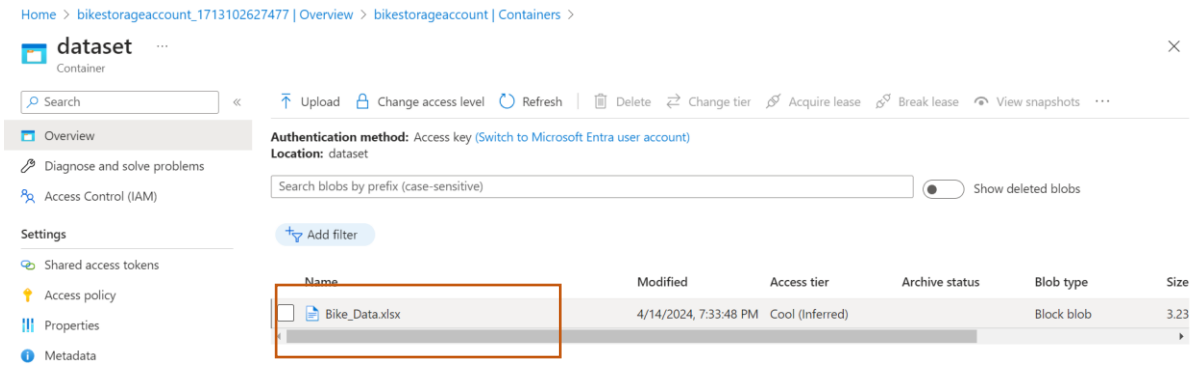
Once completed Azure blob storage click on containers and container and drops the files.



Give the name of the container and create. Once created dataset add files to the dataset using upload option.

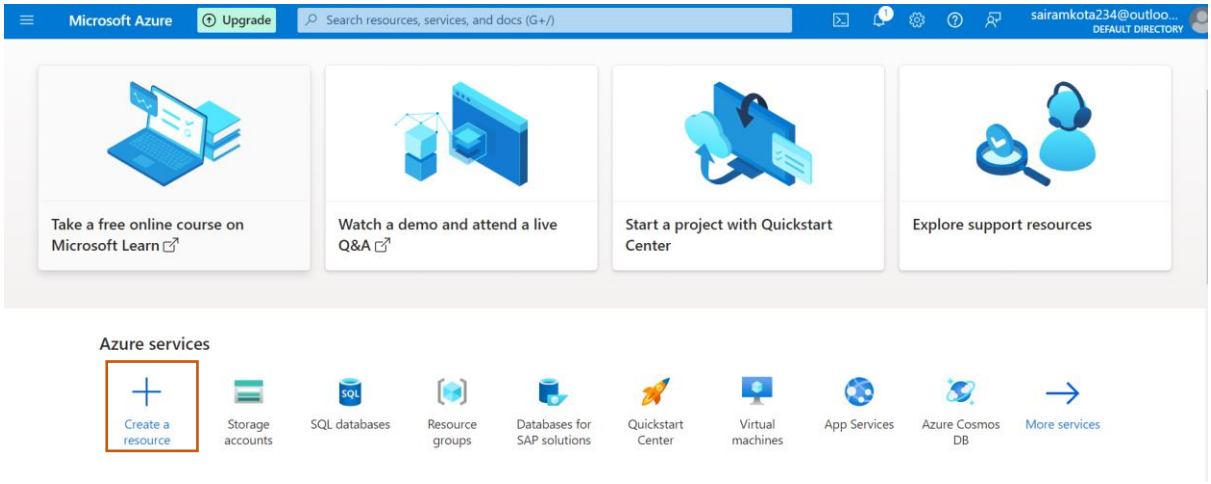




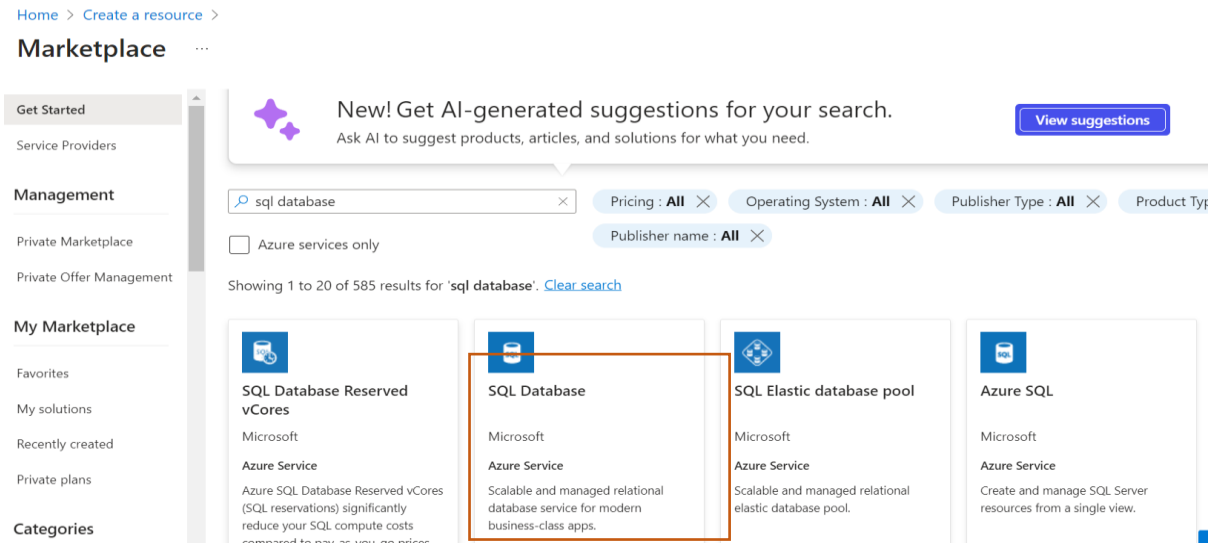


Bike Data successfully added into blob.

Azure Sql Database:



Search for Sql Database → Select Azure Sql Database



Create Sql Database, and fill required details.

- Select Subscription
- Select resource if not exists create a resource group.
- Give Database Name
- If we have server select or click on create a server
    - Create new server.
    - Give server name.
    - Choose location.
    - Use Sql Authentication
    - Give Server admin login.
    - Give password and Confirm Password should match and click on ok.
- Choose Compute Storage type.
    - Choose Dtu – based purchasing model.
        - Choose Basic (For less demanding workloads) and click on ok.
- Choose the Back storage and redundancy based on the requirement.

Home > Create a resource > Marketplace > SQL Database >

Create SQL Database

Microsoft

SQL Database Hyperscale: Low price, high scalability, and best feature set. Learn more

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * : Free Trial

Resource group * : bikerg

Create new

Database details

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name * : lakshmandb

Server * : Select a server

Create new

The value must not be empty.

Want to use SQL elastic pool? : Yes / No

Server details

Enter required settings for this server, including providing a name and location. This server will be created in the same subscription and resource group as your database.

Server name * : lakshmanserver .database.windows.net

Location * : (US) East US

Authentication

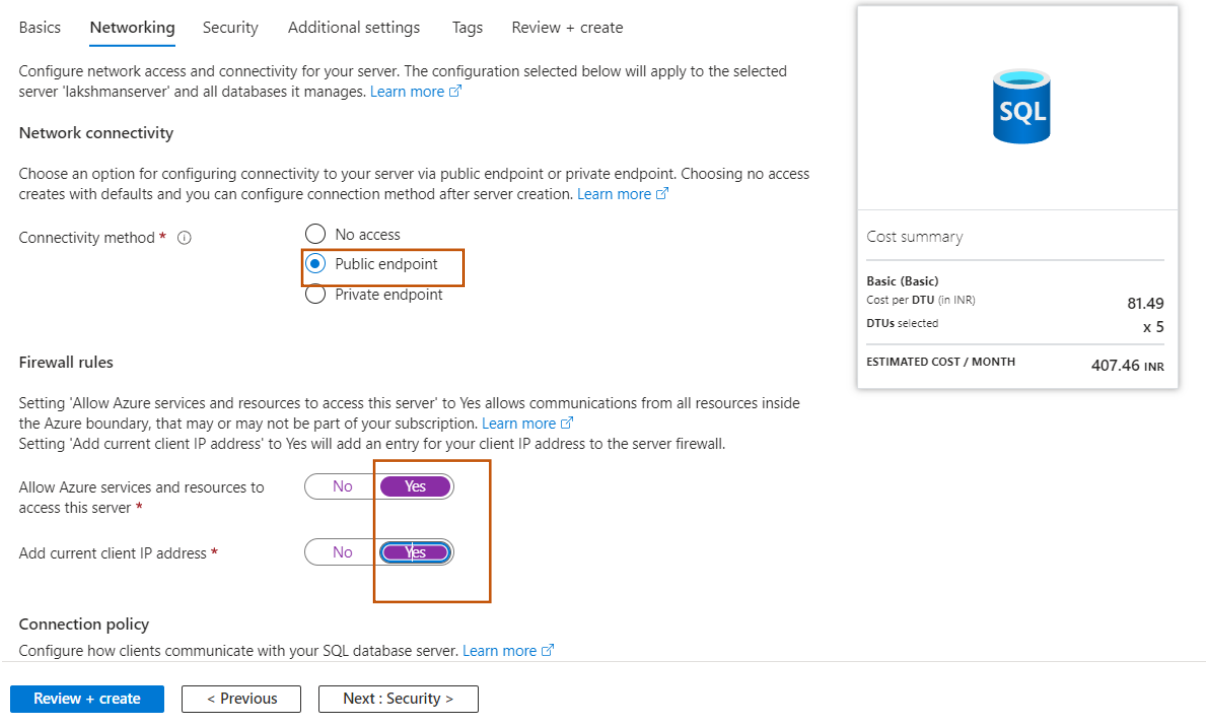
Azure Active Directory (Azure AD) is now Microsoft Entra ID. Learn more

Select your preferred authentication methods for accessing this server. Create a server admin login and password to access your server with SQL authentication, select only Microsoft Entra authentication Learn more using an existing Microsoft Entra user, group, or application as Microsoft Entra admin Learn more , or select both SQL and Microsoft Entra authentication.

Authentication method

- Use Microsoft Entra-only authentication
- Use both SQL and Microsoft Entra authentication
- Use SQL authentication

Next:

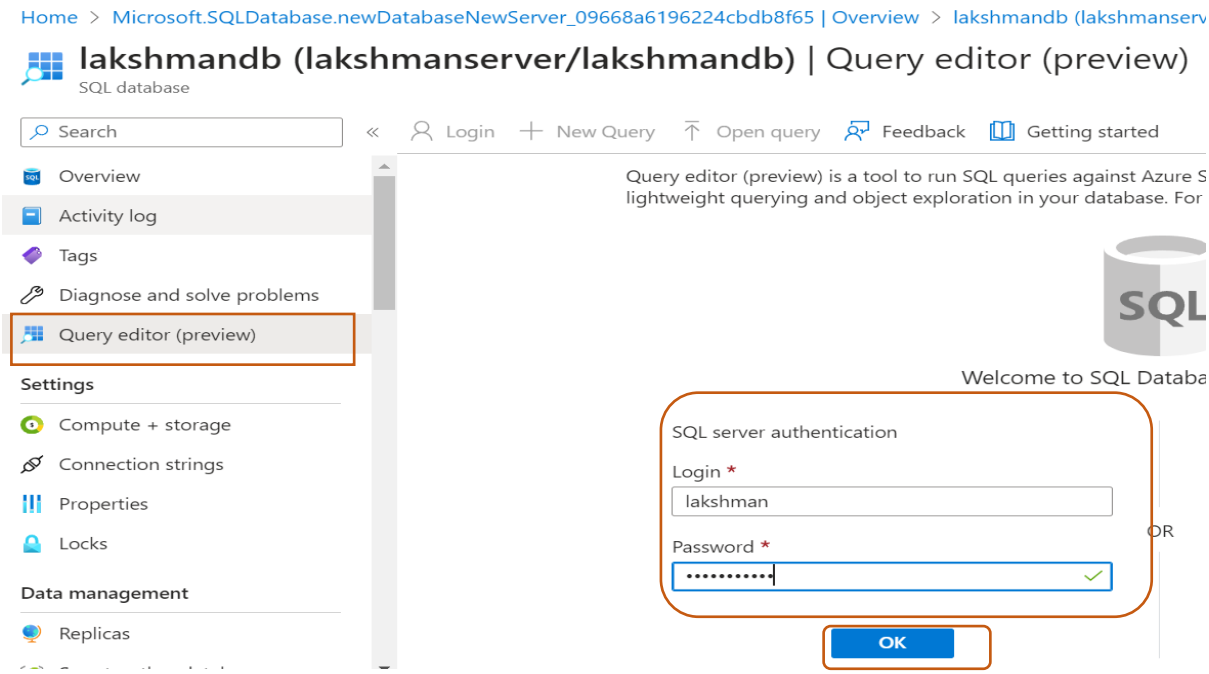


Choose Public end point and give firewalls to Yes for both.

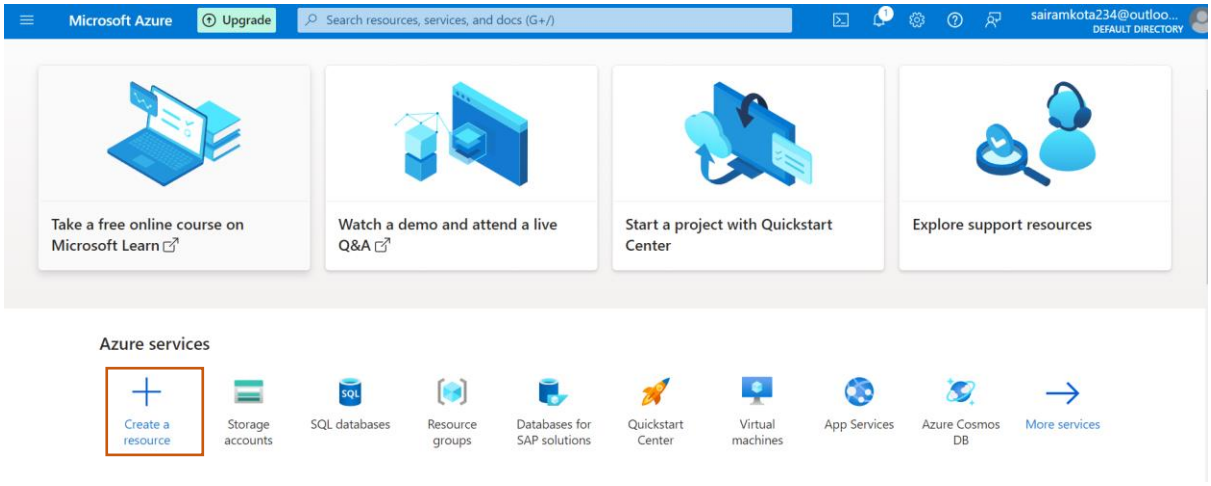
1. Allow Azure services and resources to access this server.
2. Add current client IP address.

Remaining tabs all are default and create it will deploy.

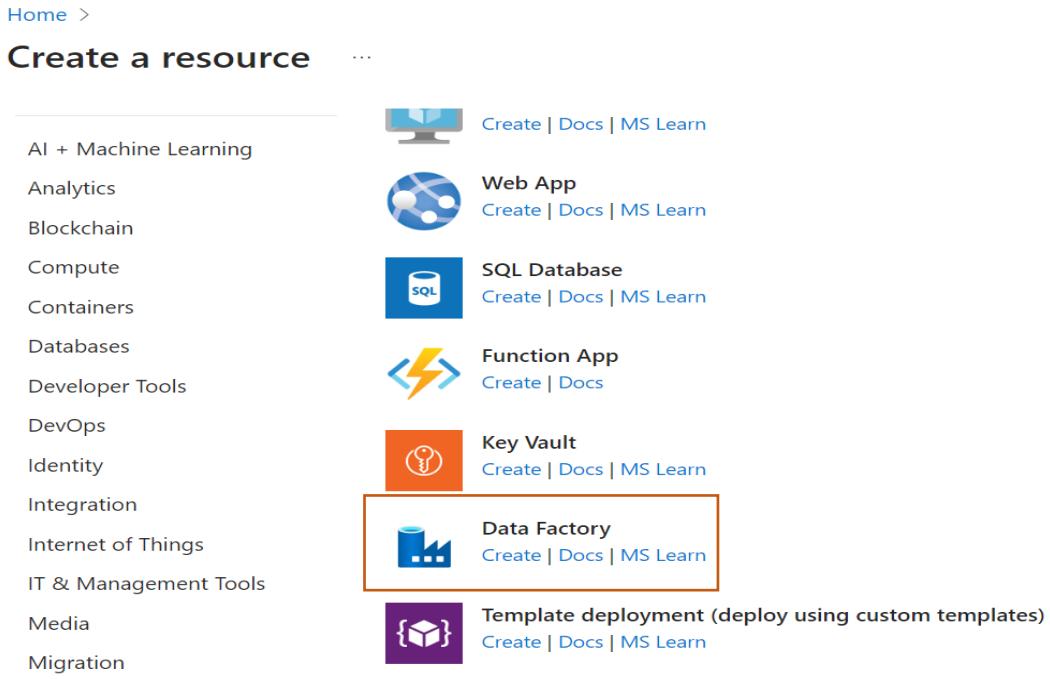
Once completed deployment using power Query editor connect to the server giving server name & password and click on ok. After Successful login will see the Azure Sql.



Azure Data Factory



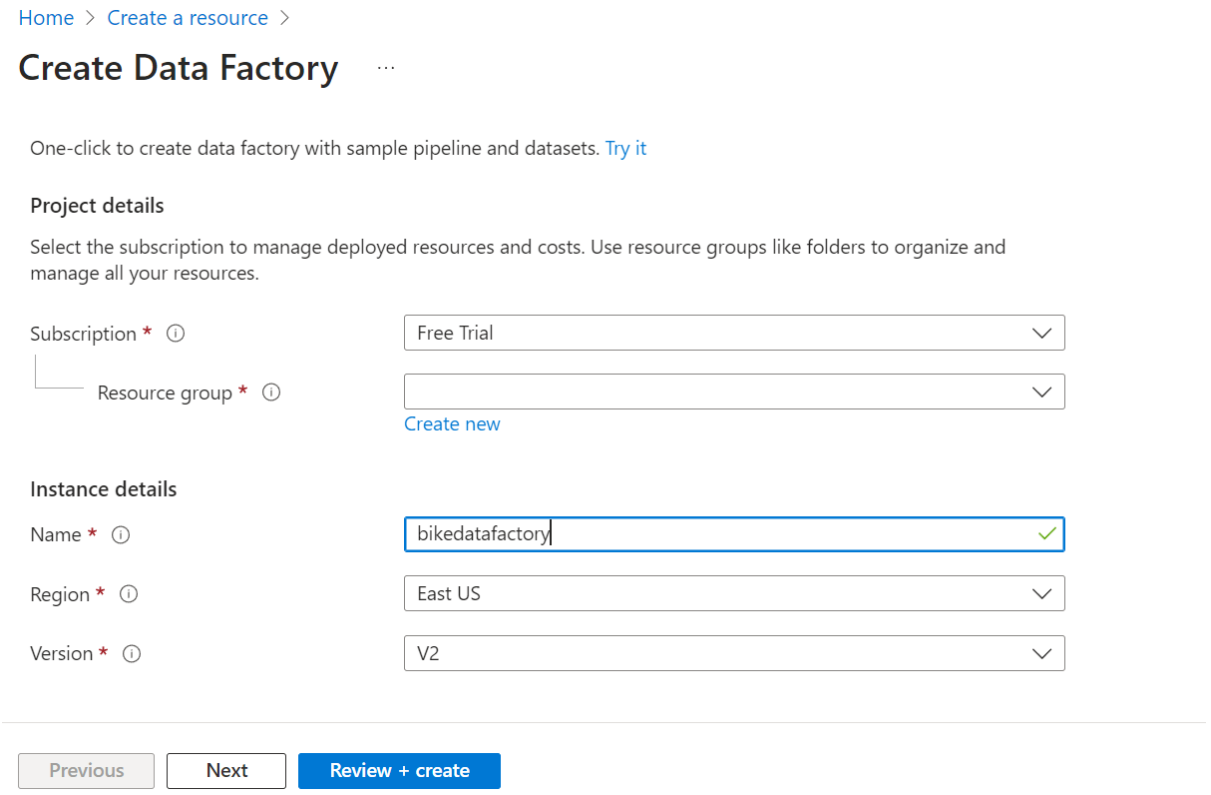
Click on create resource → Search for Azure Data Factory



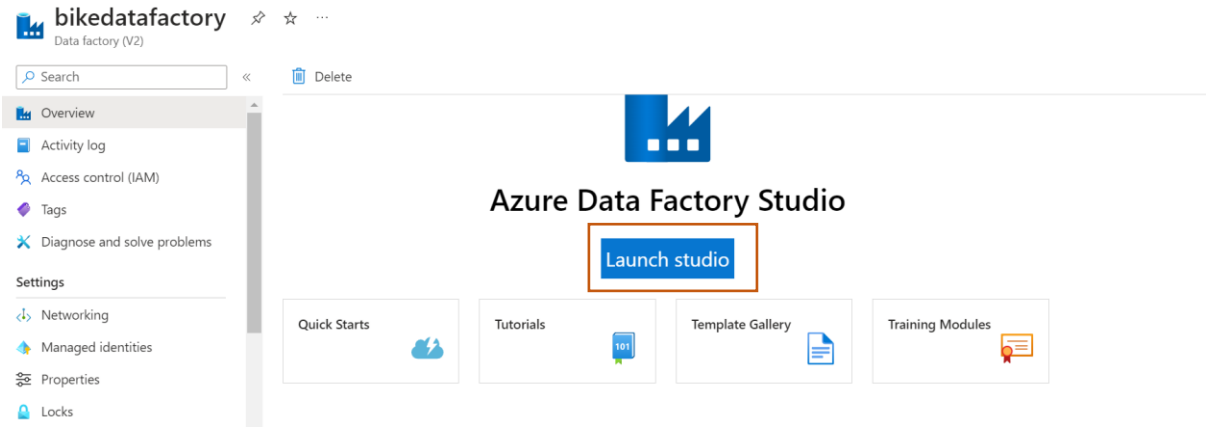
Click on Azure Data Factory and fill the required details.

- Choose Subscription
- If we have resource group, please select otherwise create it.
- Give the instance details names
- Select region.
- By default, version is v2 only.

Remaining all tabs are default go and create the azure data factory.

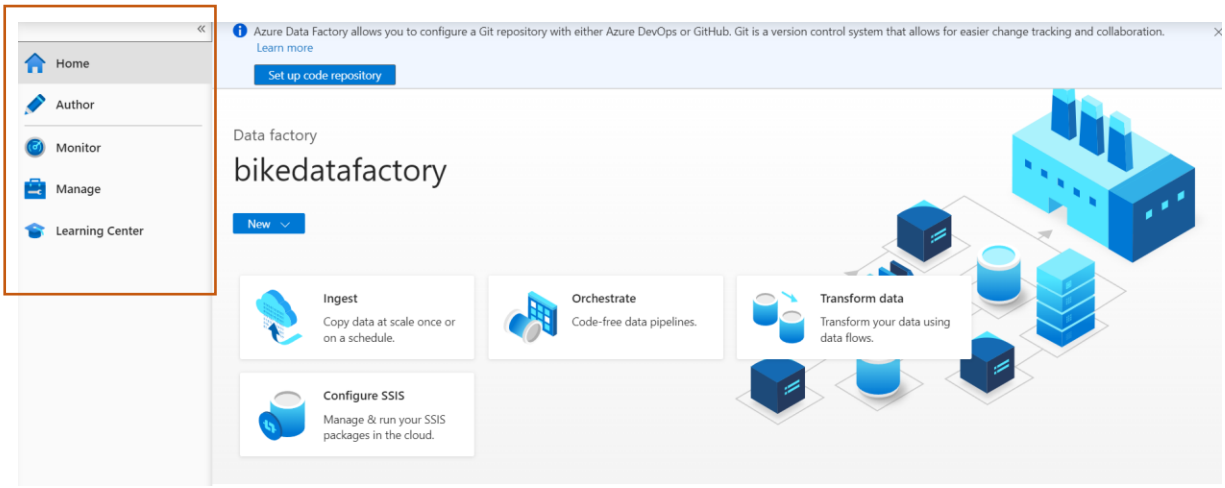


After Deployed of Azure Data Factory launch studio.

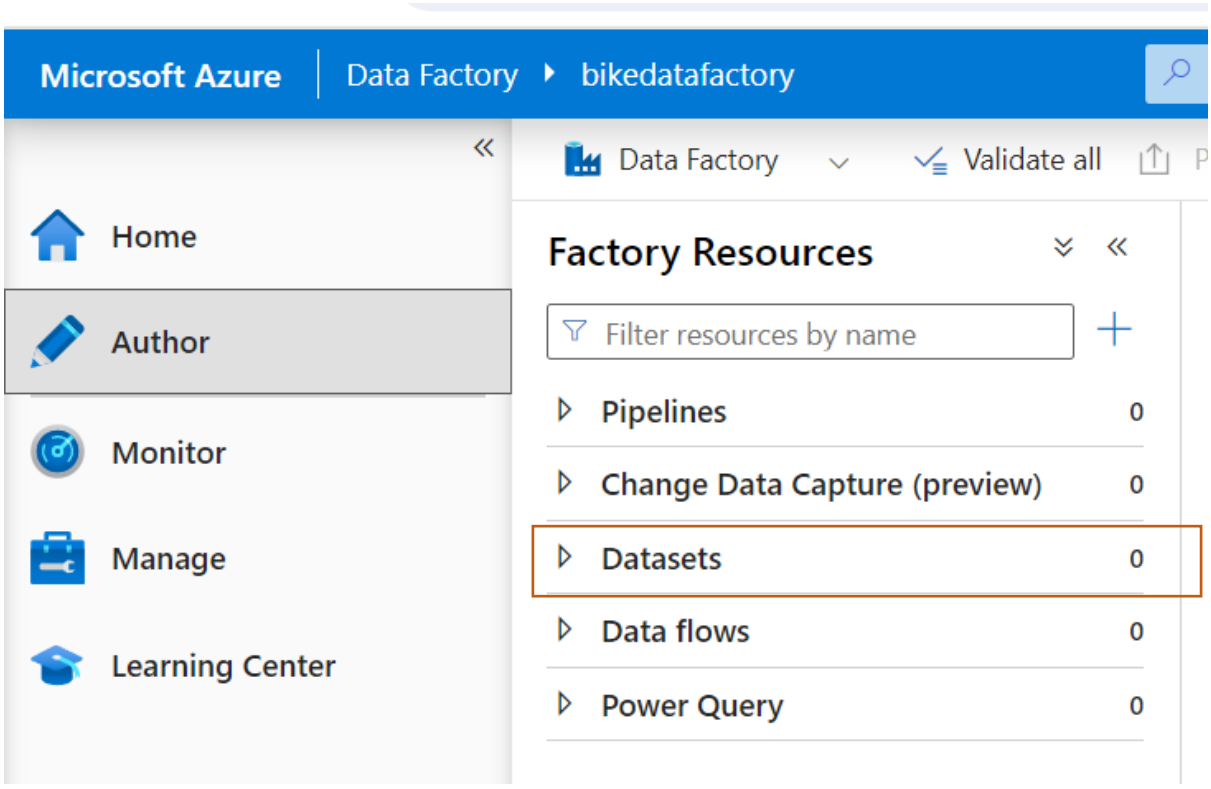


After launching Azure studio will see the below things.

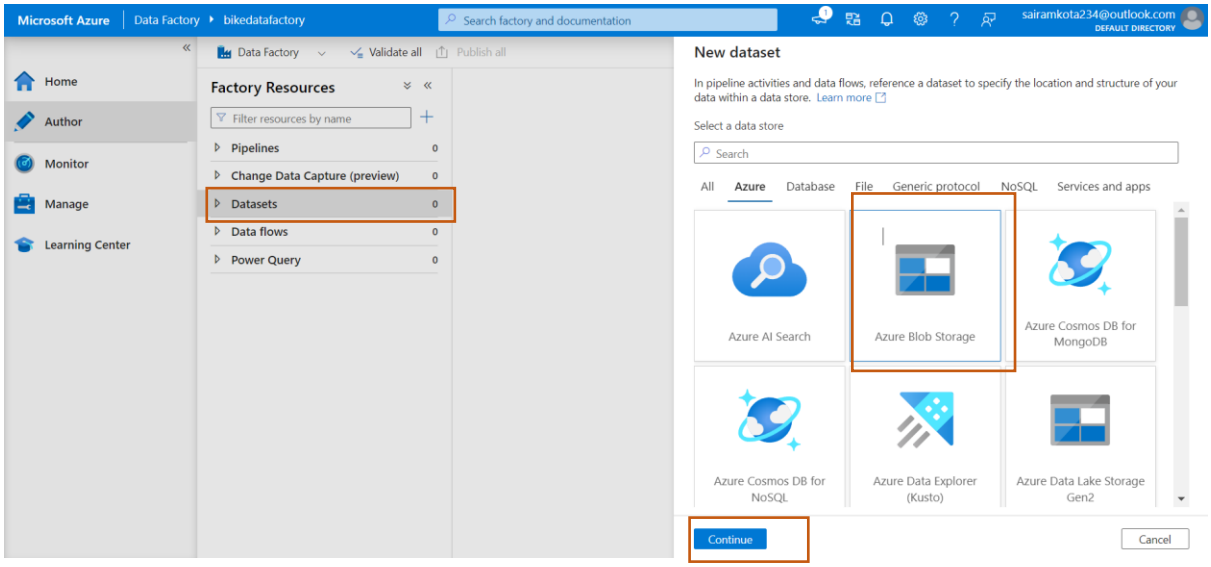
1. Home tab
2. Author tab
3. Monitor tab
4. Management Hub tab

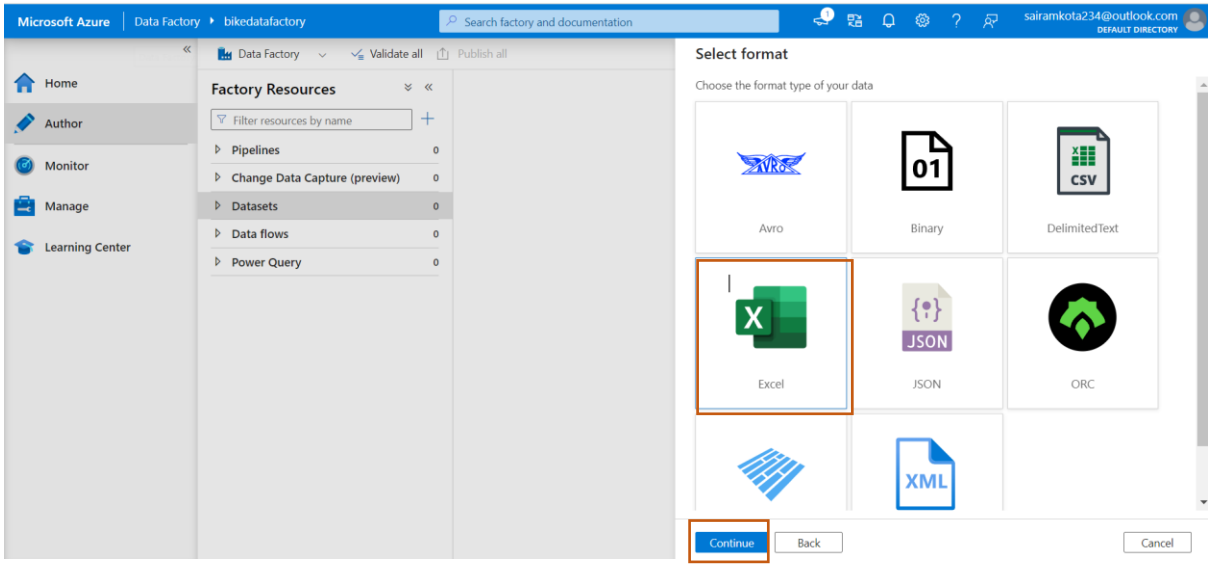


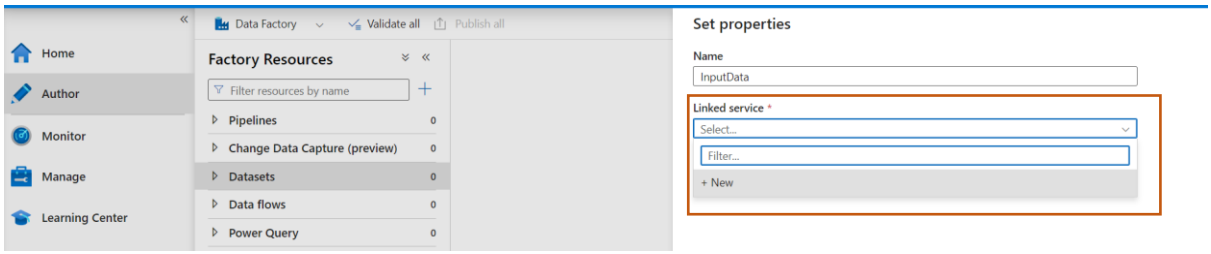
Click on Author Tab and will see the factory resources.



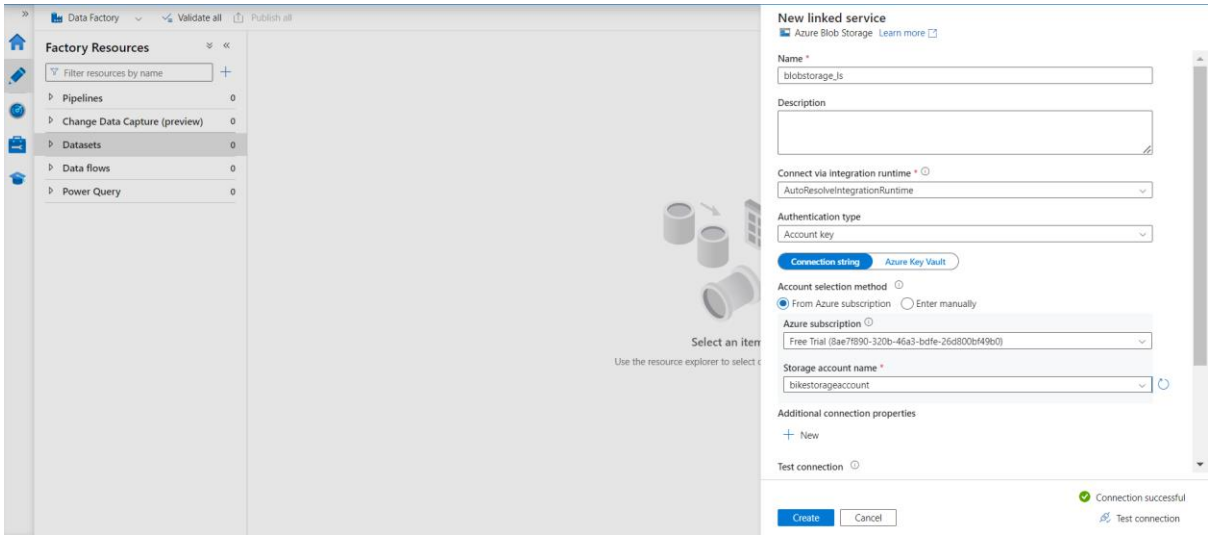
- Click on Datasets
- Select New Data set.
- Choose Azure
- Select Azure Blob storage because our blob available here only and click on continue.
- Choose type of file here (here file is Excel) and click on continue
- Select on the link service and click new
- Need to give link service connection (It is a connection b/w blob and ADF)



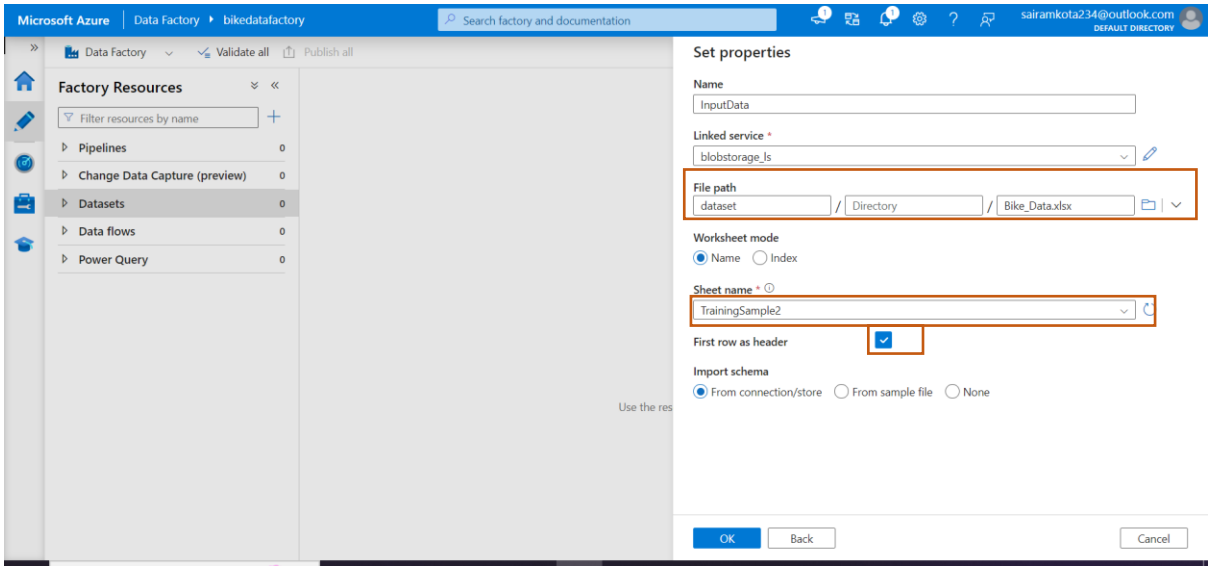




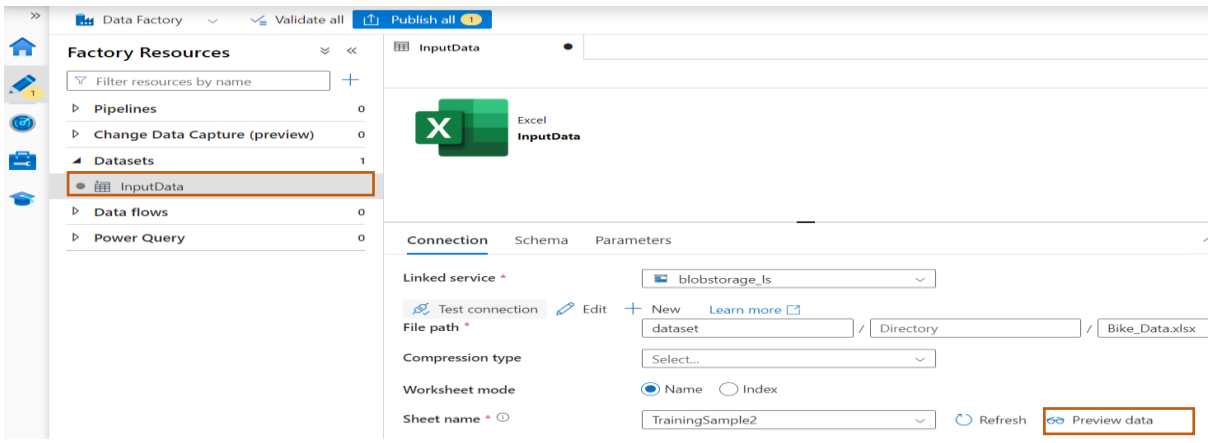
- After clicking on new it opens the new linked service need to fill the required details.
- Select Azure Subscription
- Select storage account name
- Before creating check the test connection once successful click on create.



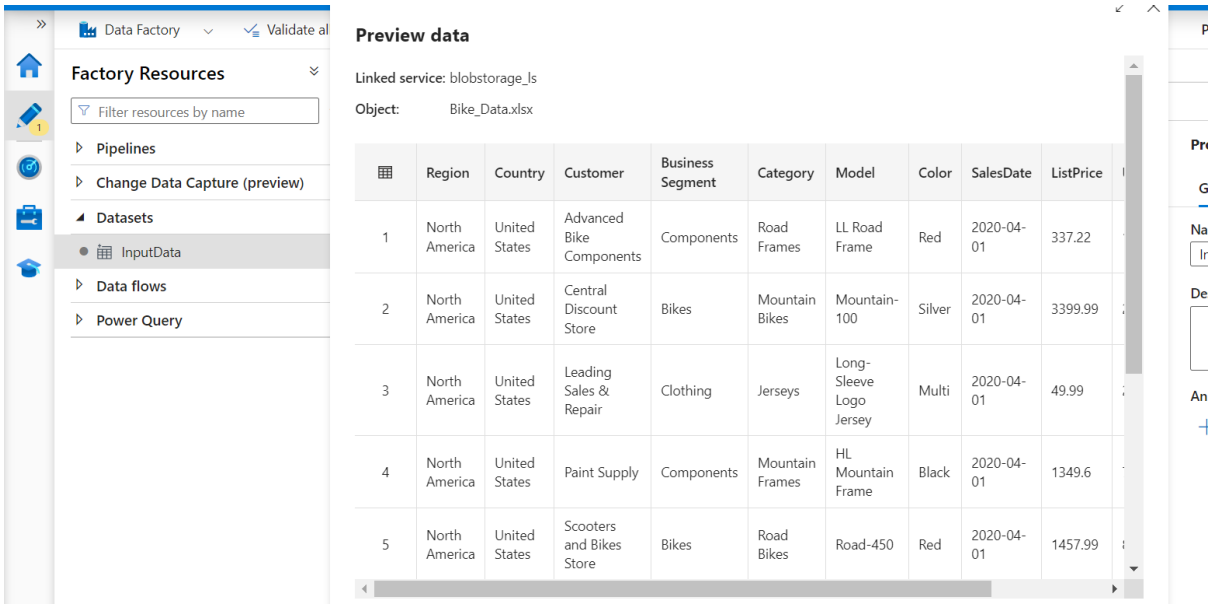
- After this need to give set properties
- Select the location and correct data
- It comes as File path
- Select sheet name
- Select First row as header and click on ok

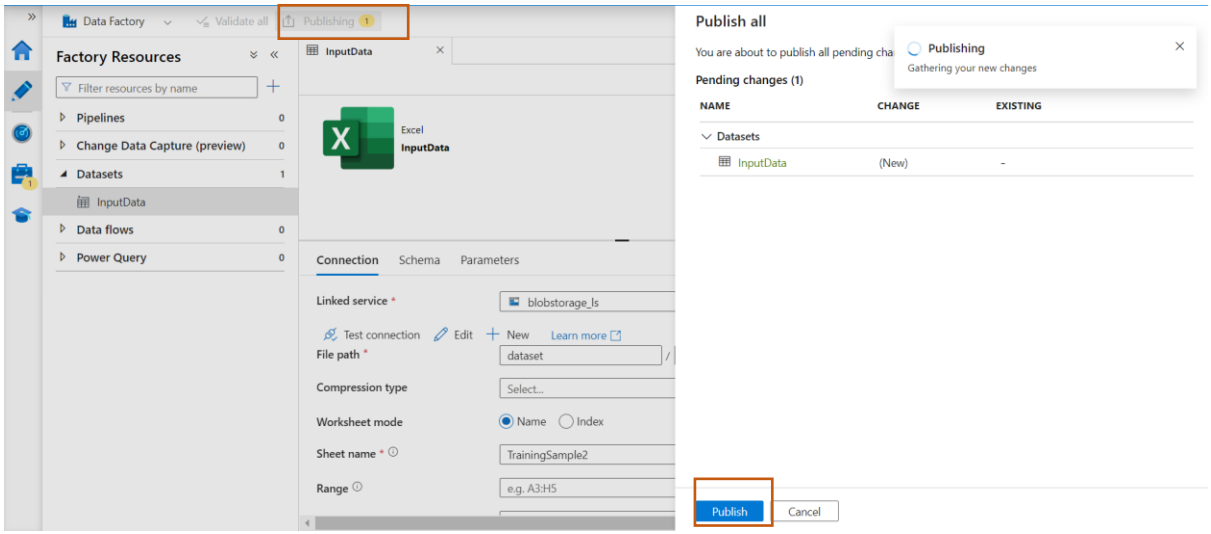


- Finally we are seeing data in Input Data in Datasets before publish once see the preview data



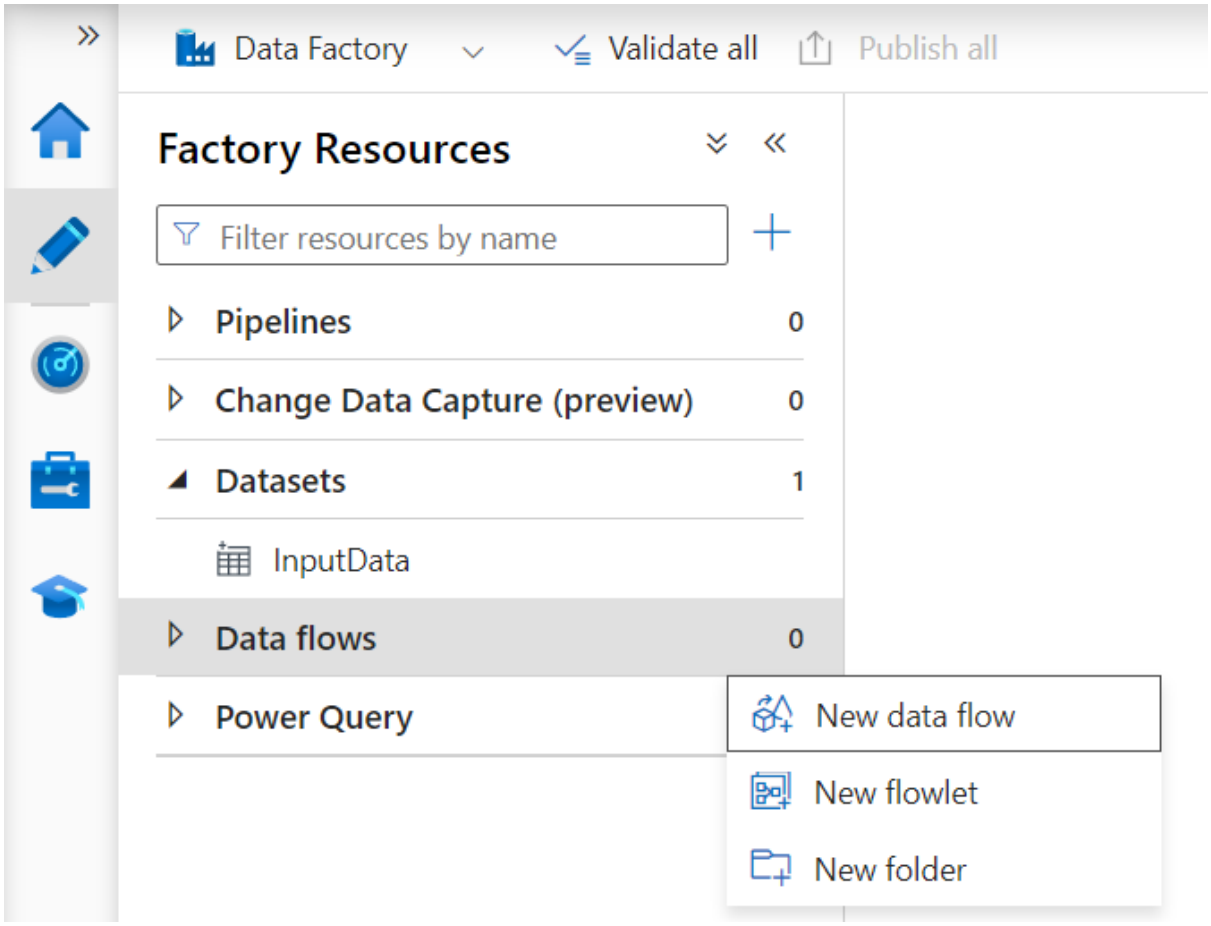
- Seeing preview here
- Here there is no save we need to publish



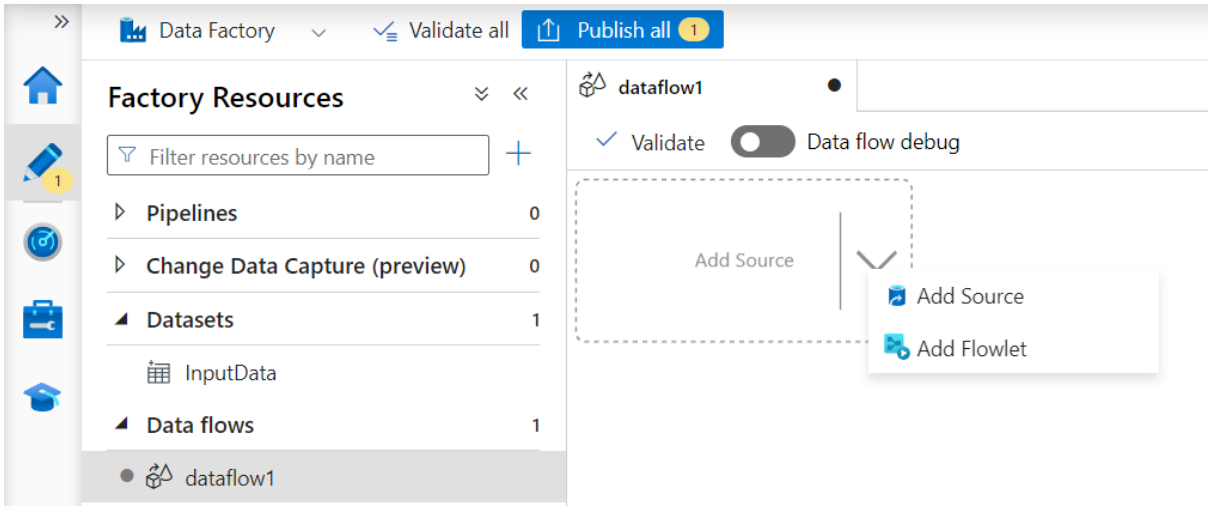


Once published go to the transformations (Data Flows)

- Click on New data flow

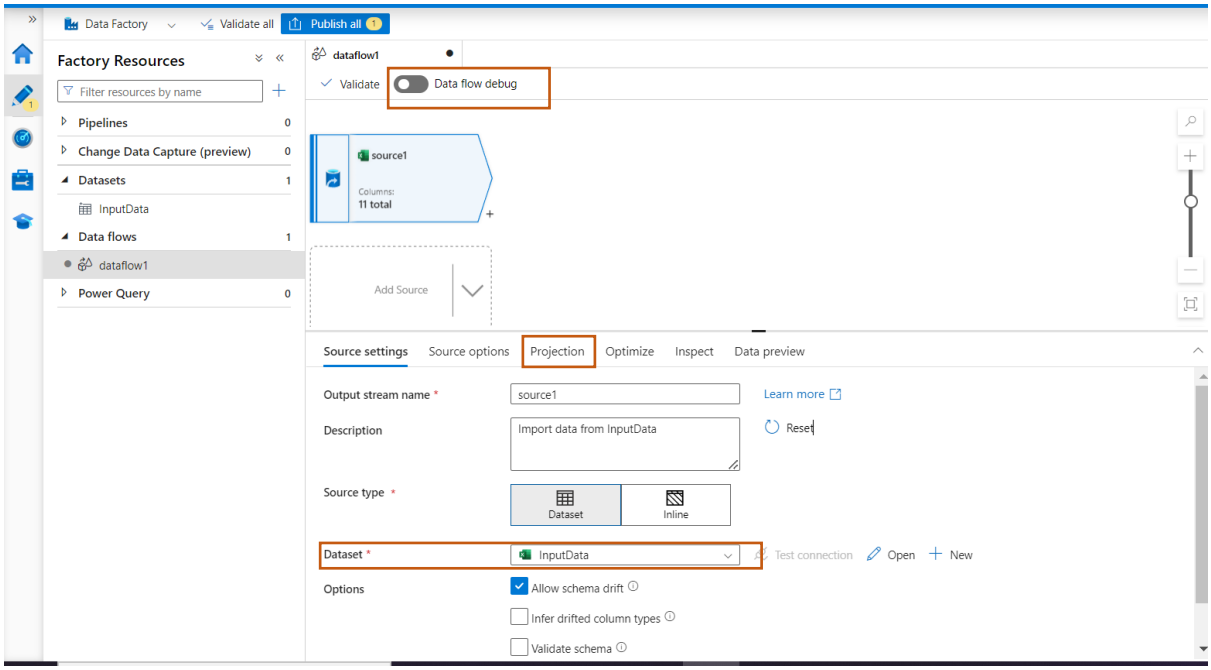


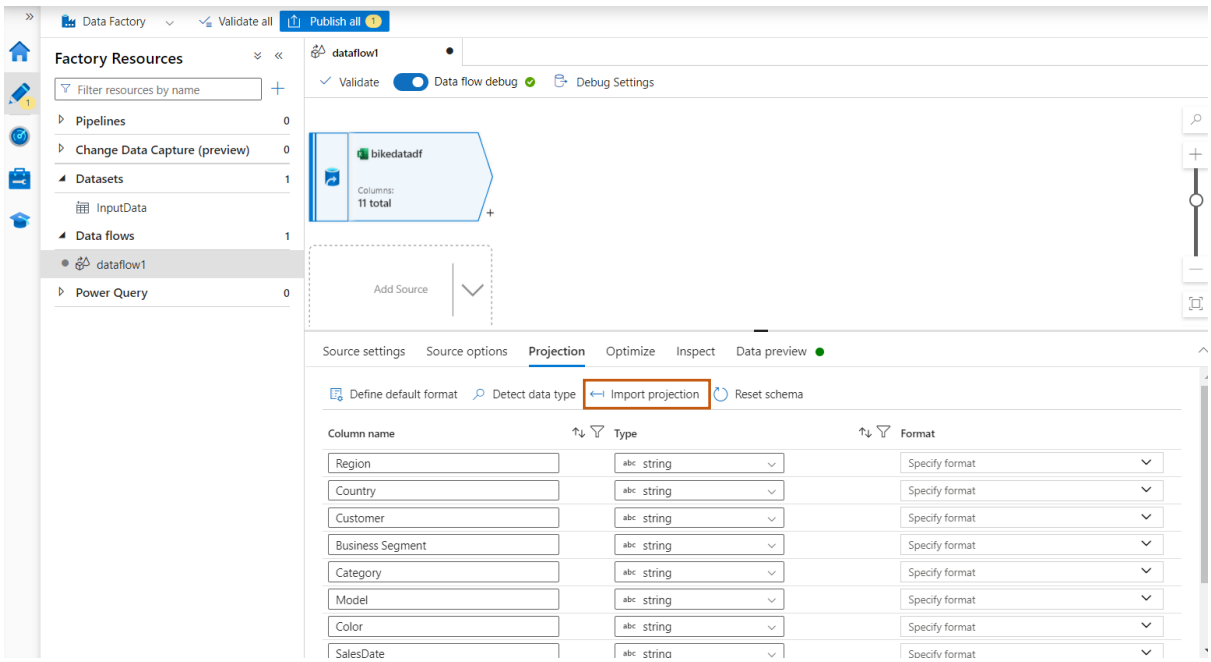
- Click on Add Source



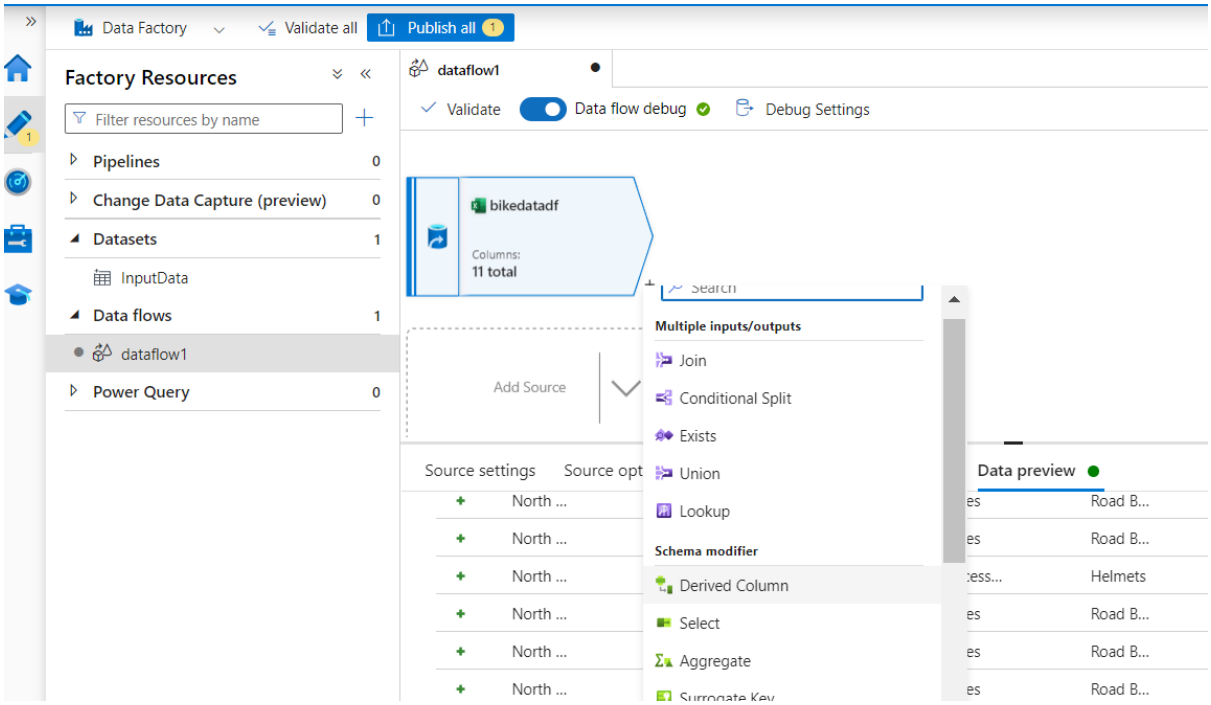
,

- Select the Data set
- Click on Data flow debug and select 1 hour to check the data preview once data is Data flow debug enabled check the preview once.
- Click on projection and click on Import projection

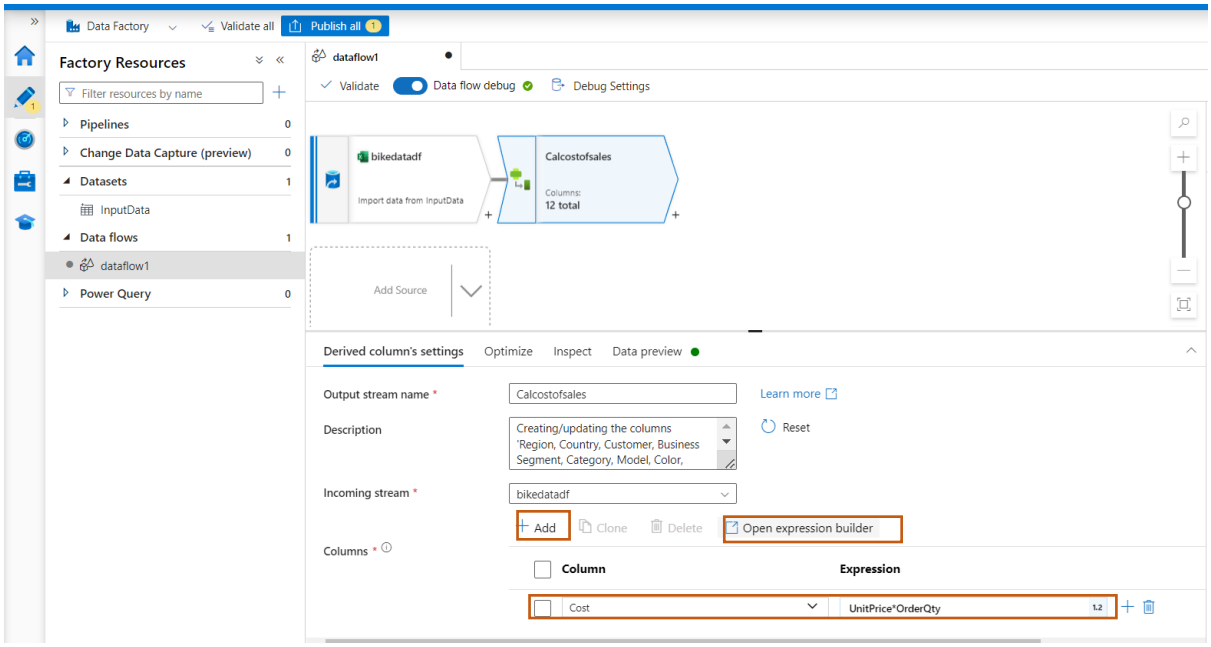


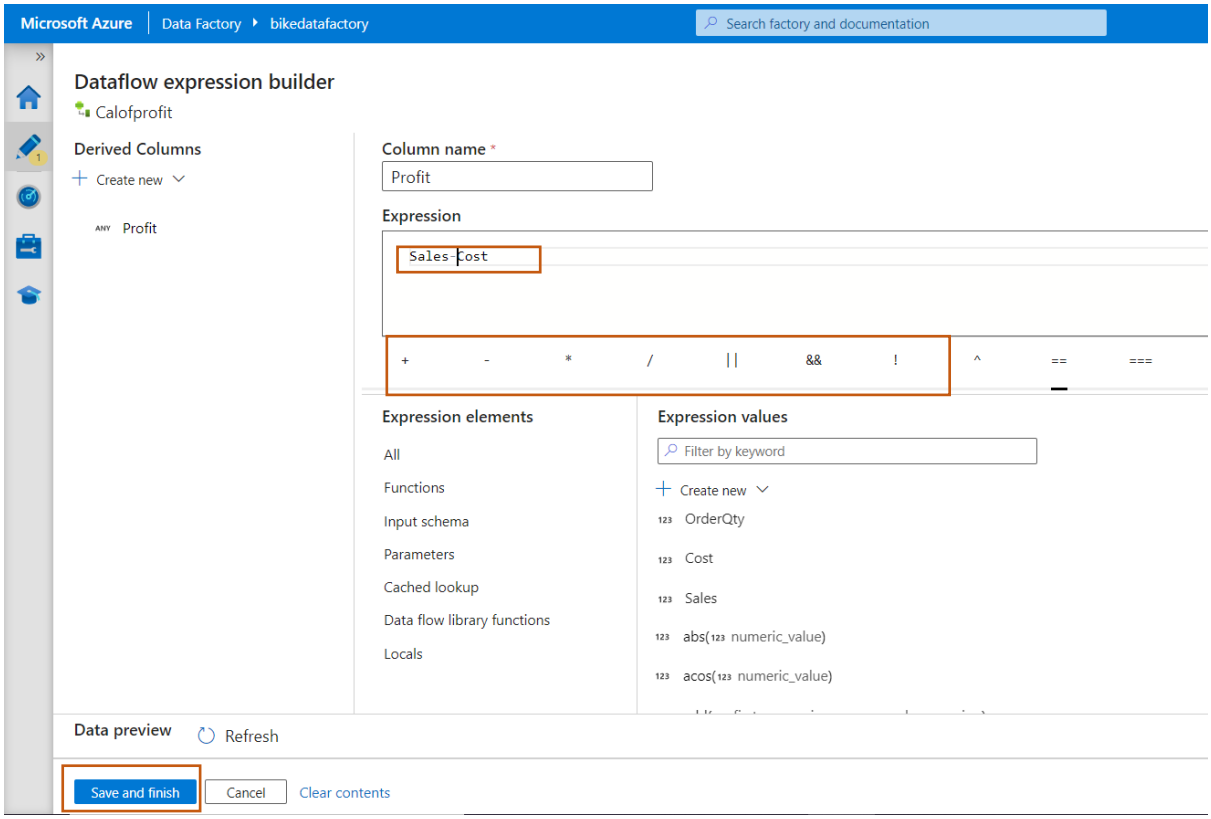


- We need to calculate cost, sales, and profit
- By click on derived column we can multiply the columns

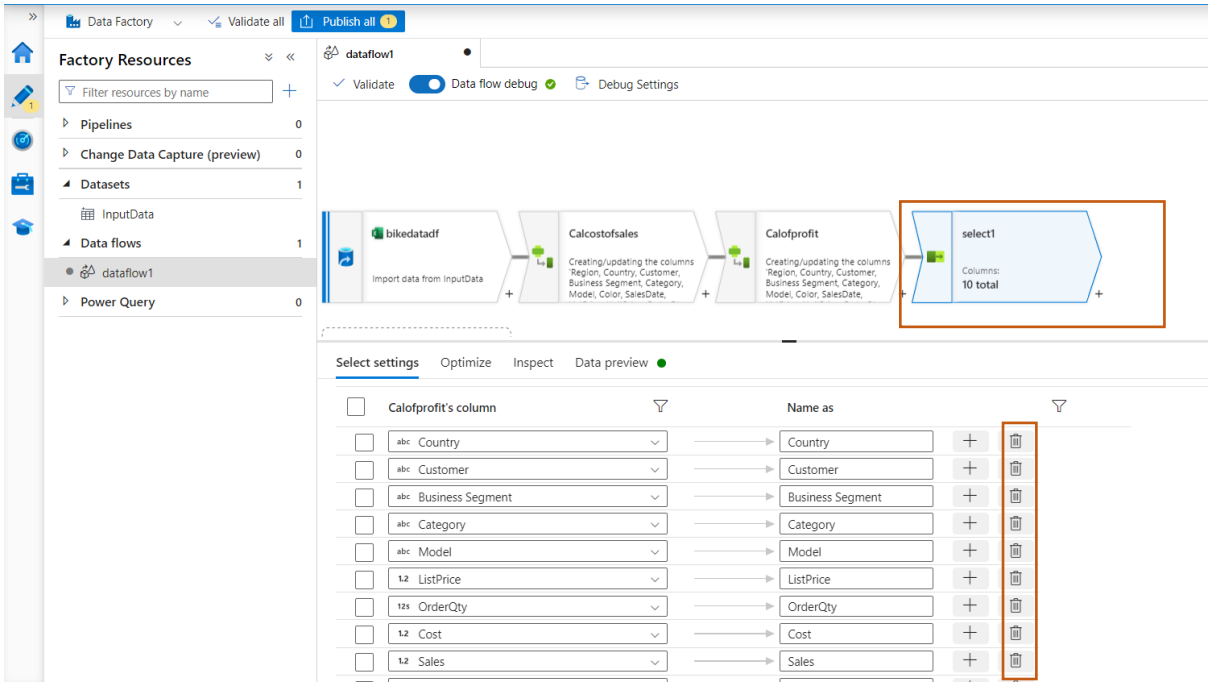


- After adding derived column open expression builder add expression
- Need multiple columns click on Add condition we can add one more column
- We can check how the data populating using data preview
    - Cost
    - Sales
- Based on the above two derived columns we can't create new column called profit so that we add one more derived column add profit and we can see the preview.

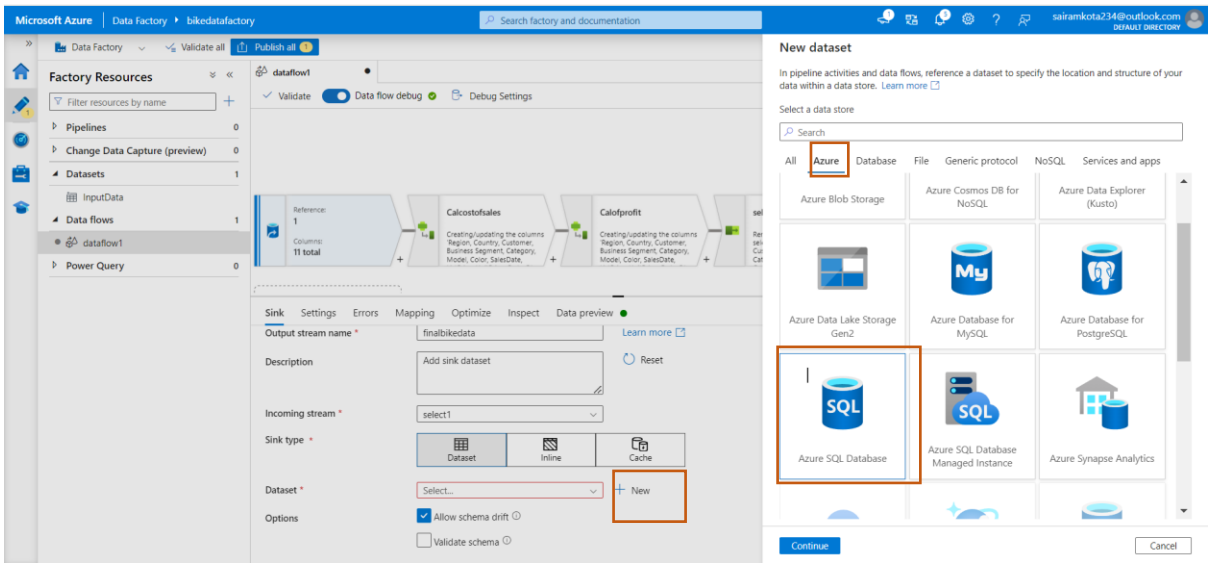




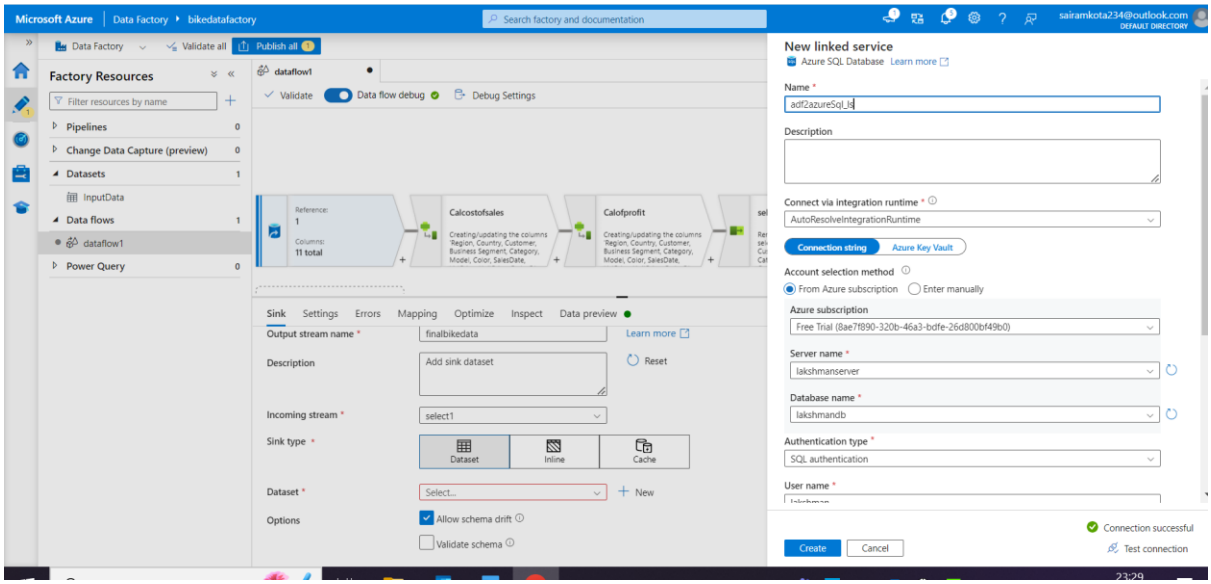
- After this we need to select only the select columns click on the + symbol click only select
- Delete the unwanted columns



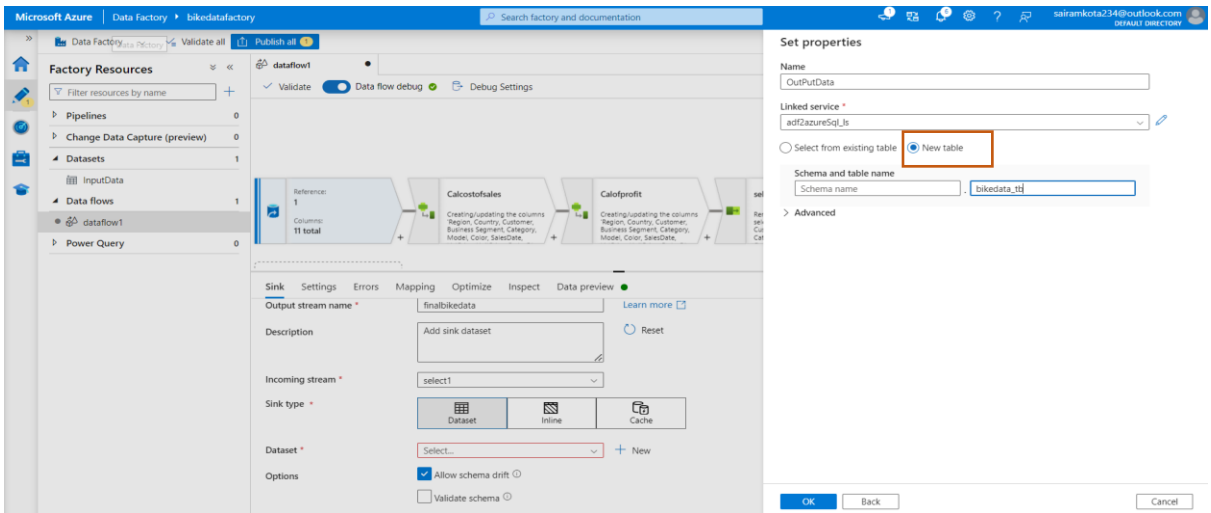
After all this click on + symbol click on Sink we are creating temporary data set to Sql data base.



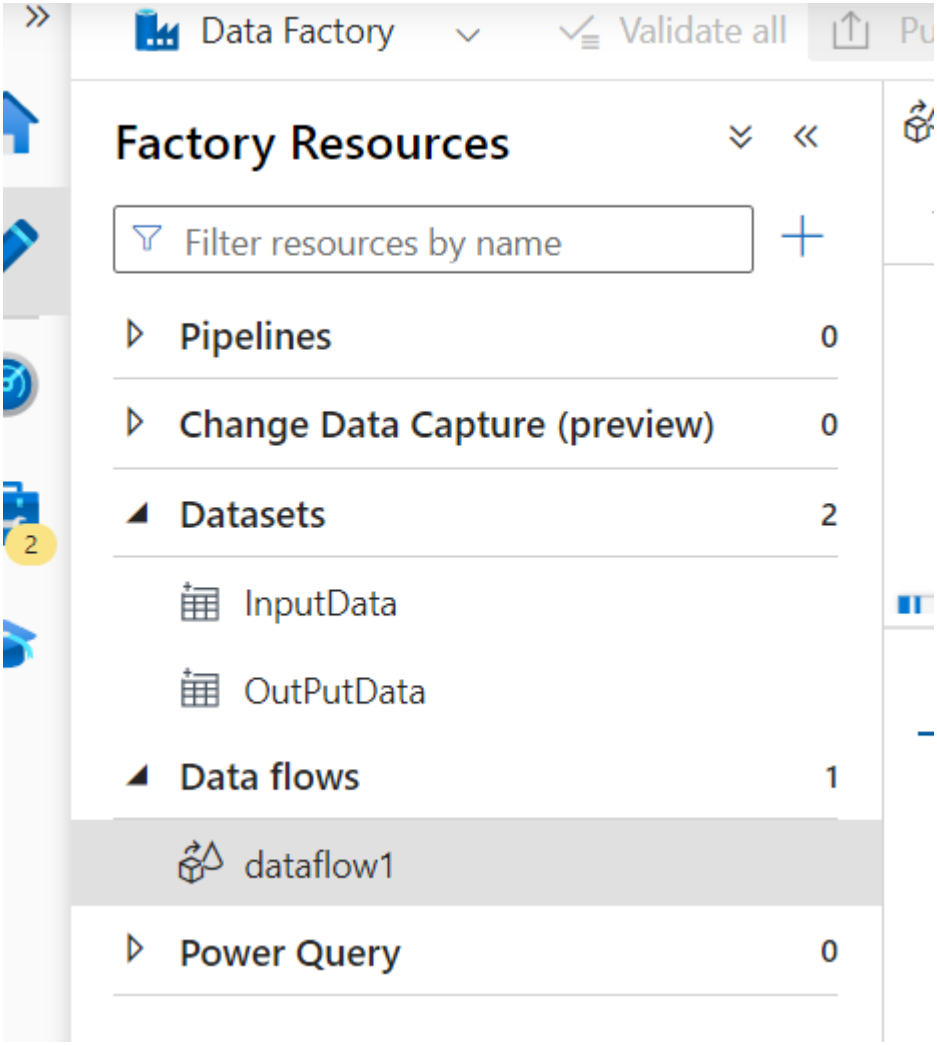
- After this will get a new linked service click on new
- Fill all the details and test the connection and click on create



- After that click on new table and write table name and click on ok.

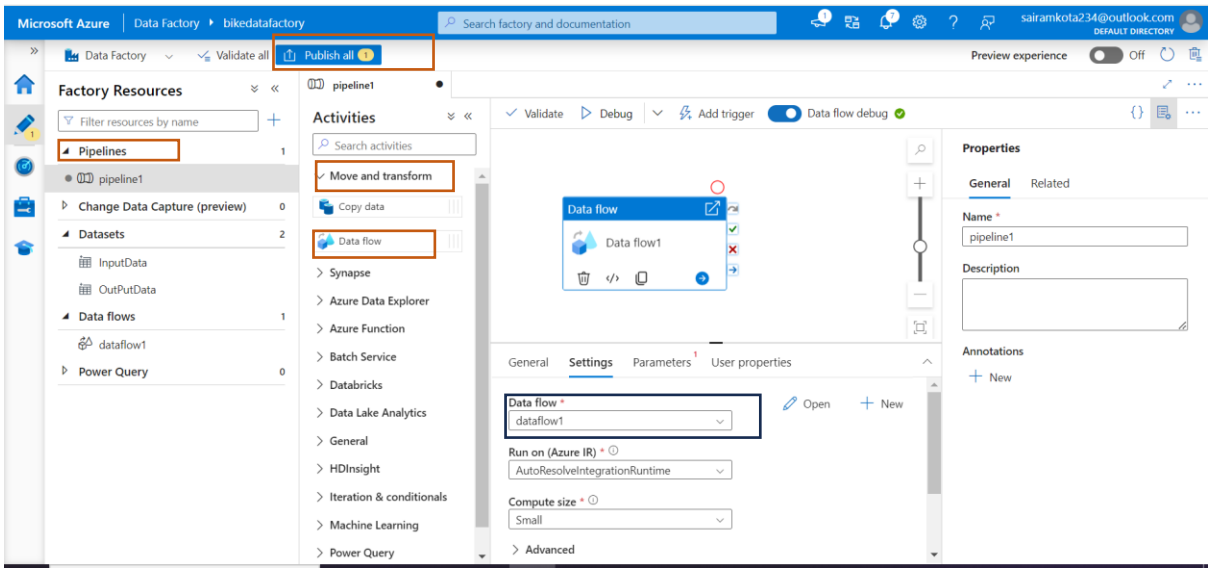


- After that publish the data

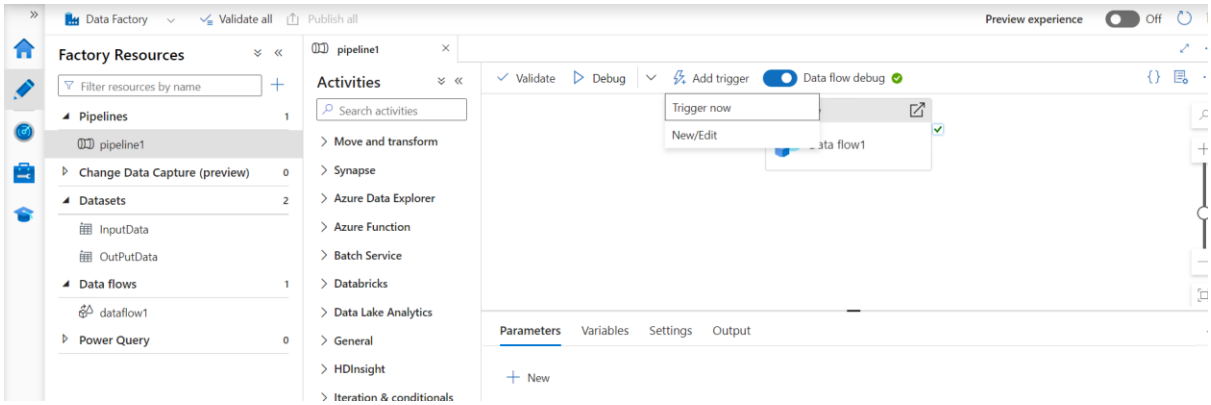


In Data Sets we can see 2 now.

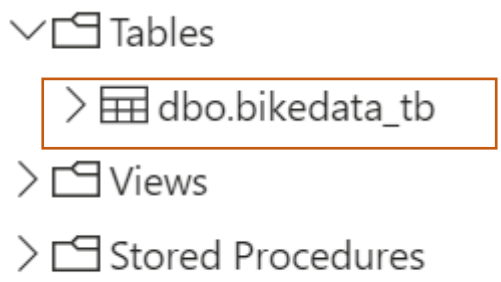
- After that click on pipelines New pipeline
  - Move and Transform
    - Drag and drop Data Flow
- In settings select the dataflow
- After that publish



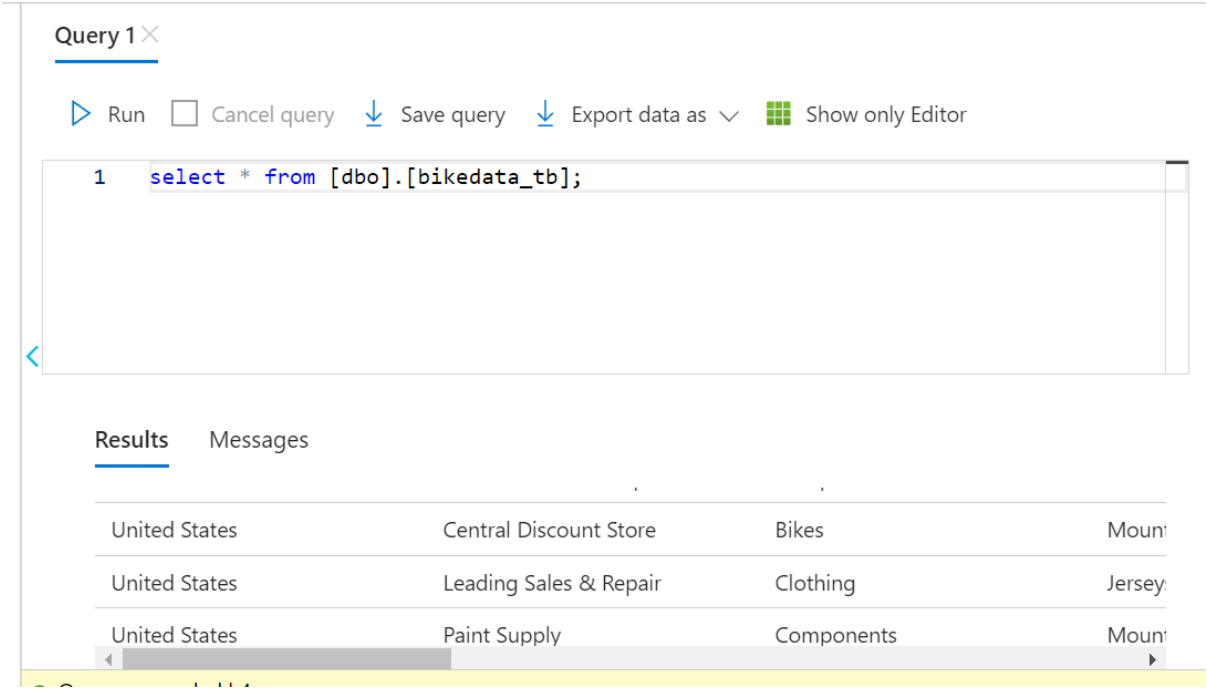
- After Publish click on the Trigger now and click on ok
- It will store in Sql DataBase



Once Pipeline completed successfully, we can see the table Azure Sql Database.



We are seeing the data


Query 1
Run
Cancel query
Save query
Export data as
Show only Editor
1 select * from [dbo].[bikedata_tb];
Results
Messages
United States Central Discount Store Bikes Mount
United States Leading Sales & Repair Clothing Jersey
United States Paint Supply Components Mount