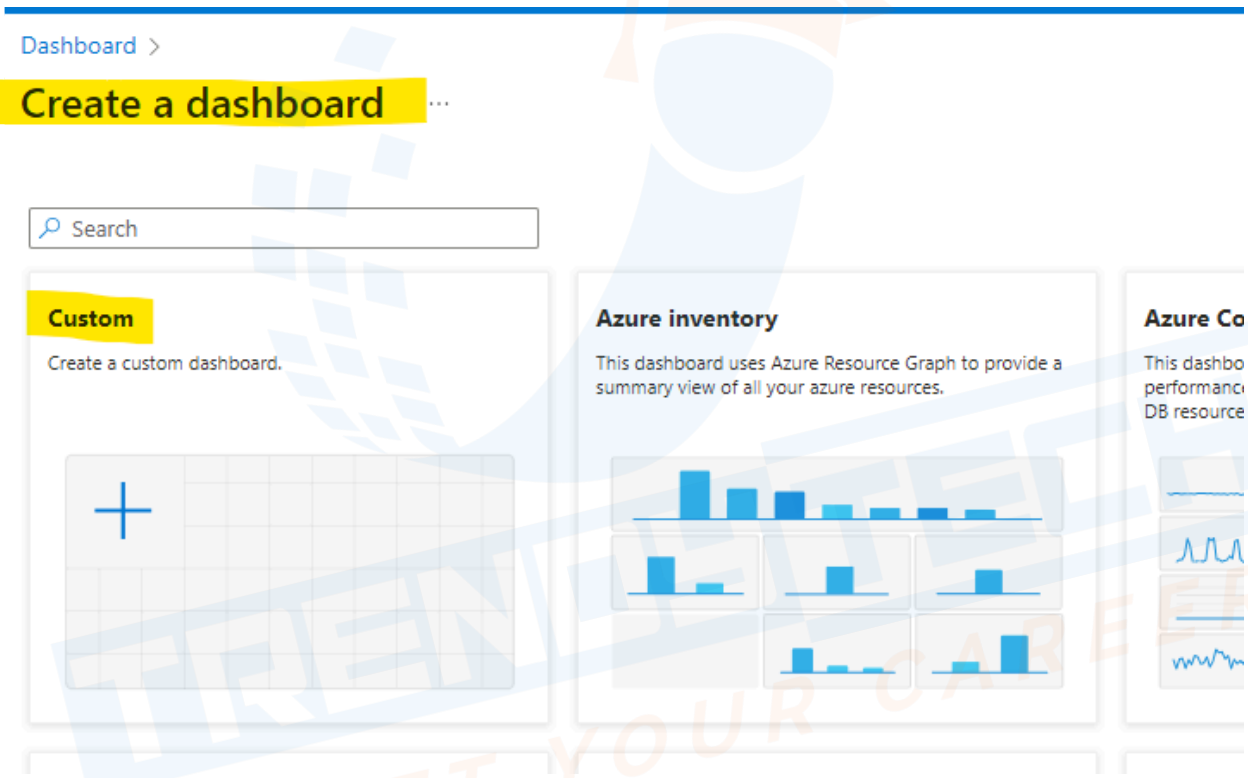
Use-case 1 : Ingesting the data from Source(RDBMS Table), transferring and loading the data to Sink(ADLS GEN2)

Pre-Steps -

Create a Dashboard and a Resource Group for the project to organize the resources related to the project at one place.

Click on Dashboard => Create => Custom =>



Note : Also you can create a resource group while creating the resource. Here we have created resource group while creating SQL Database as highlighted below

Creation of Source System - Azure SQL DB

- Create an Azure SQL Database and pin the resource to a dashboard for better organization of resources.

Home > SQL databases >

Create SQL Database

Microsoft

Subscription * : Pay-As-You-Go

Resource group * : (New) trendytech-rg

Create new

Database details

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

Database name * : trendytech-db

Server * : (new) customer360-server1 (Central India)

Create new

Compute + storage * : **General Purpose - Serverless**

Standard-series (Gen5), 2 vCores, 32 GB storage, zone redundant disabled

Configure database

Home > SQL databases > Create SQL Database >

Create SQL Database Server

Microsoft

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.

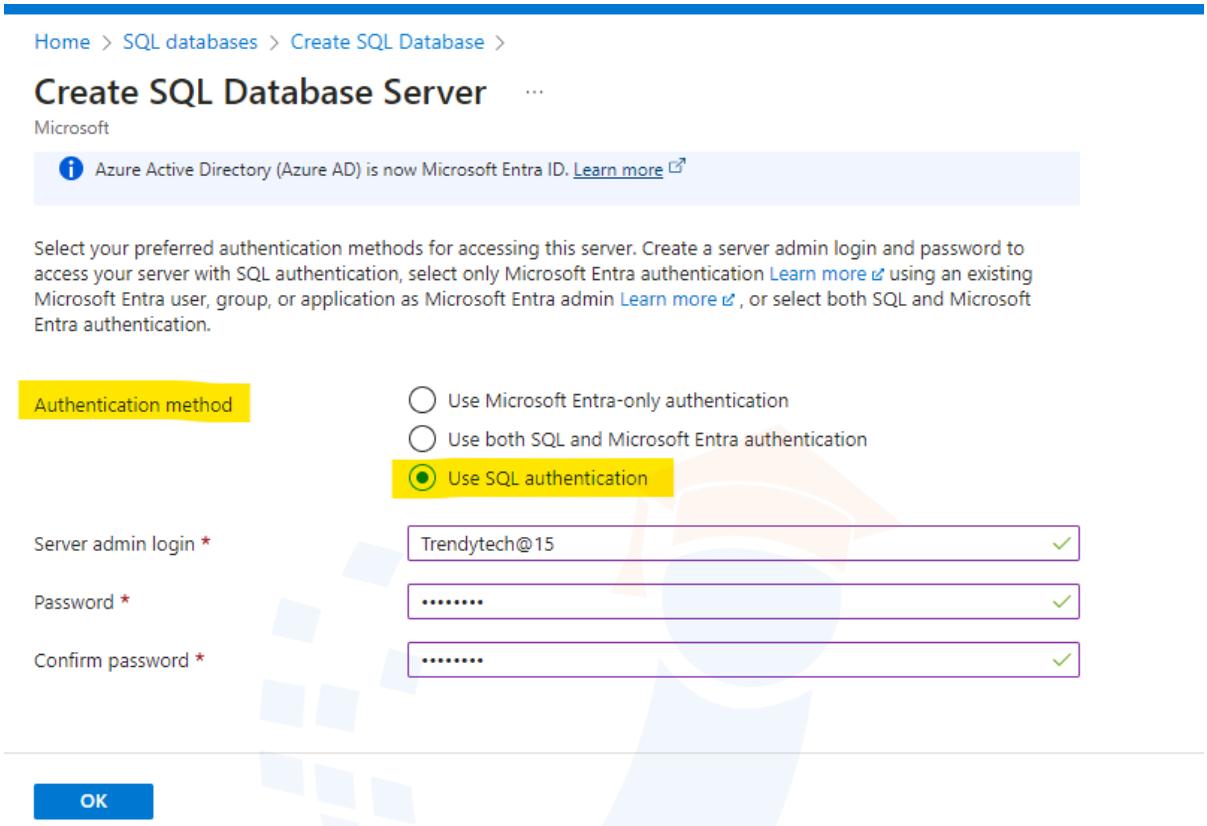
Server name * : customer360-server1 .database.windows.net

Location * : (Asia Pacific) Central India

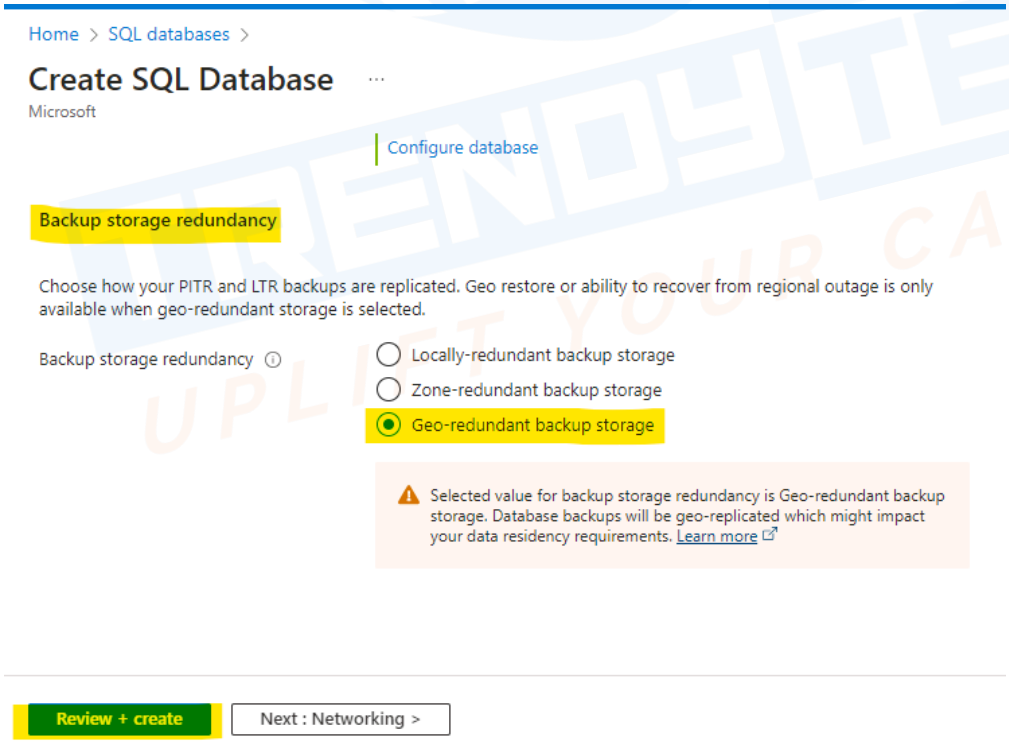
Authentication

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

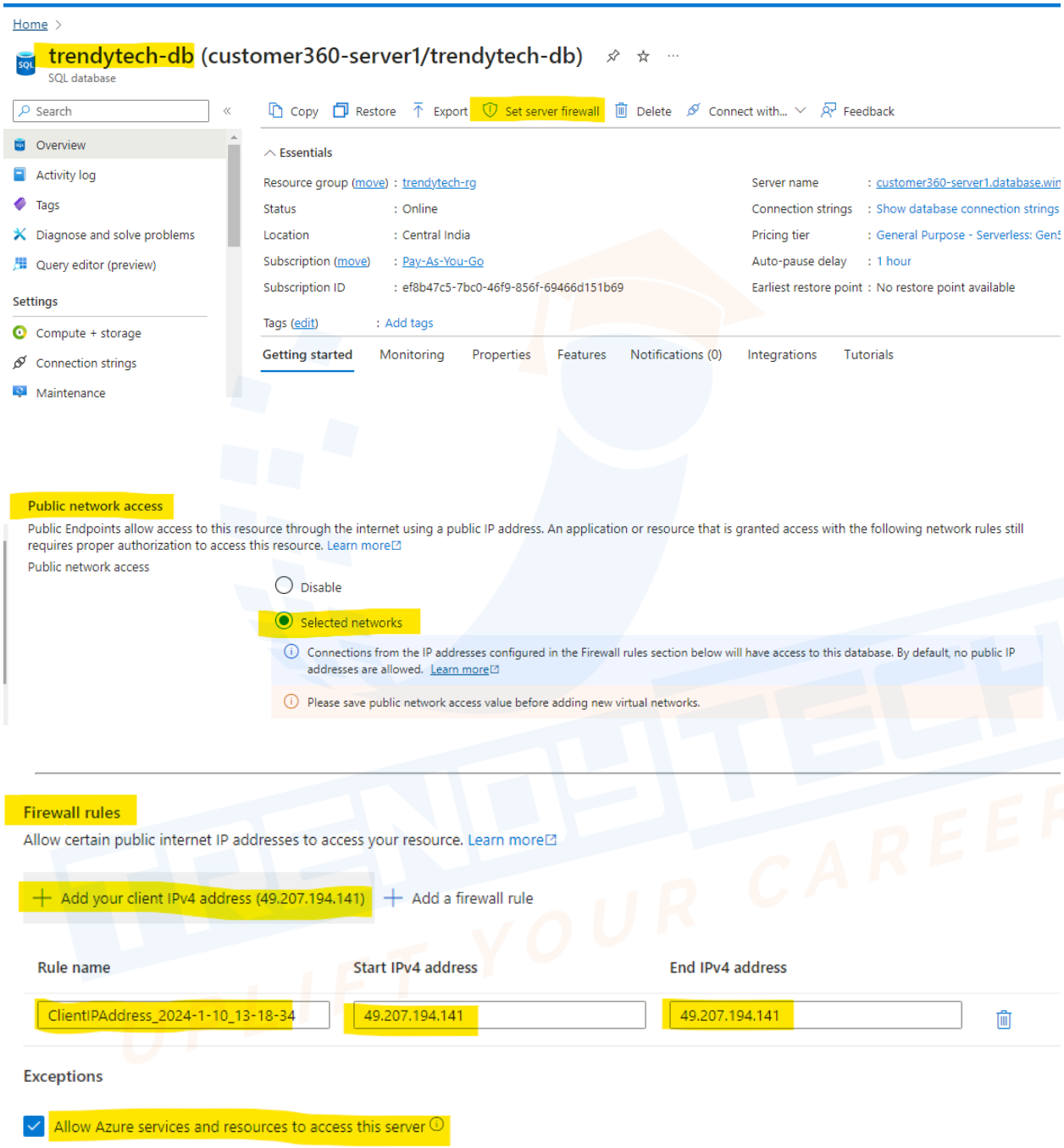
OK



Note: Please note down your username and password for Future use.

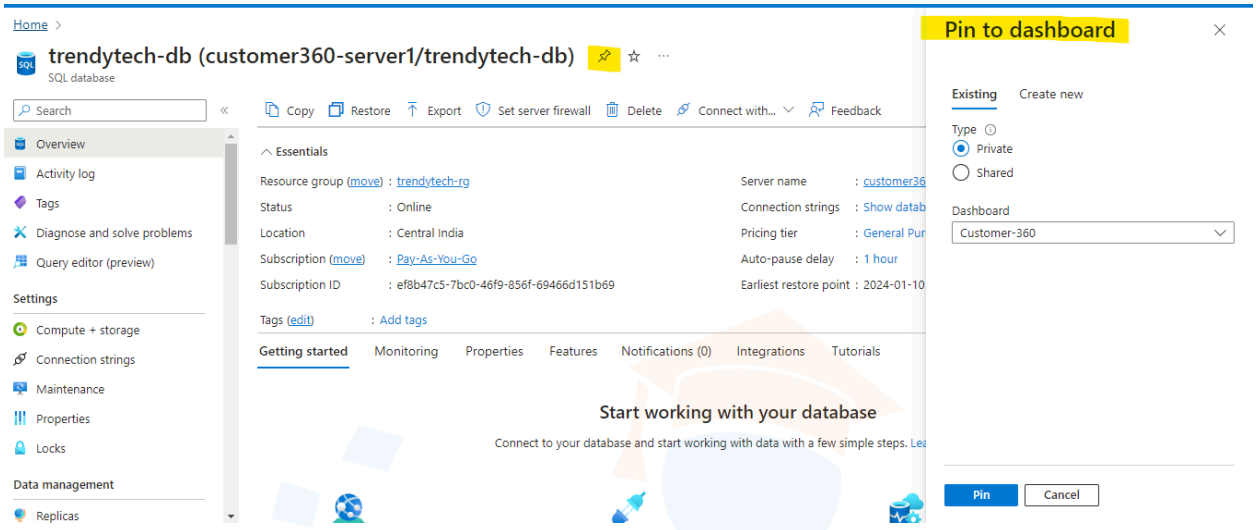


- Once the SQL DB resource is deployed, set the server firewall rule under overview tab by adding your respective IP-address



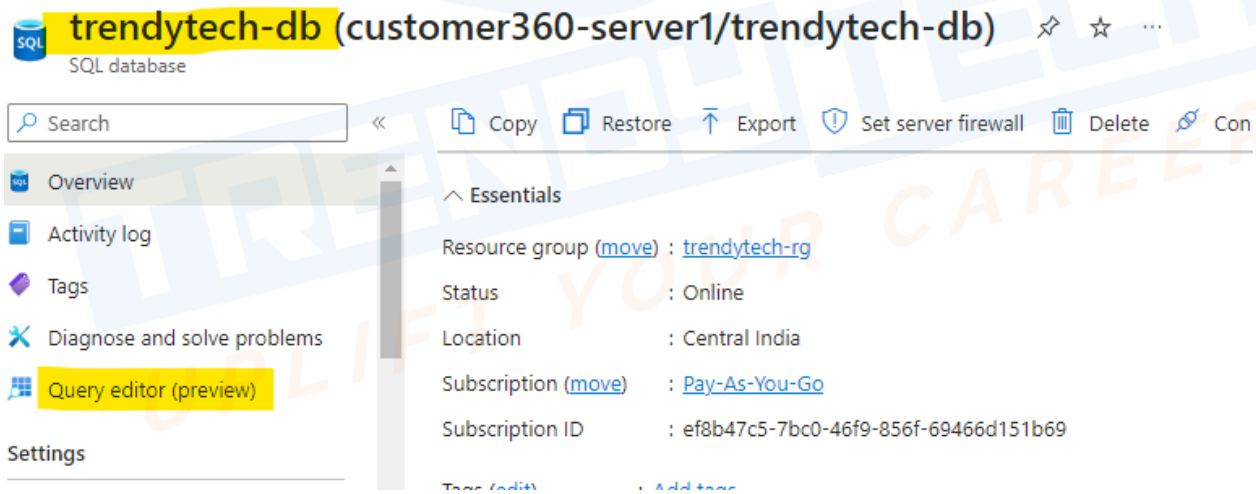
Note: You can pin this database to the dashboard that you have created

Click on the pin option as shown in the screenshot and pin it. Similarly you can pin all the other resources also.

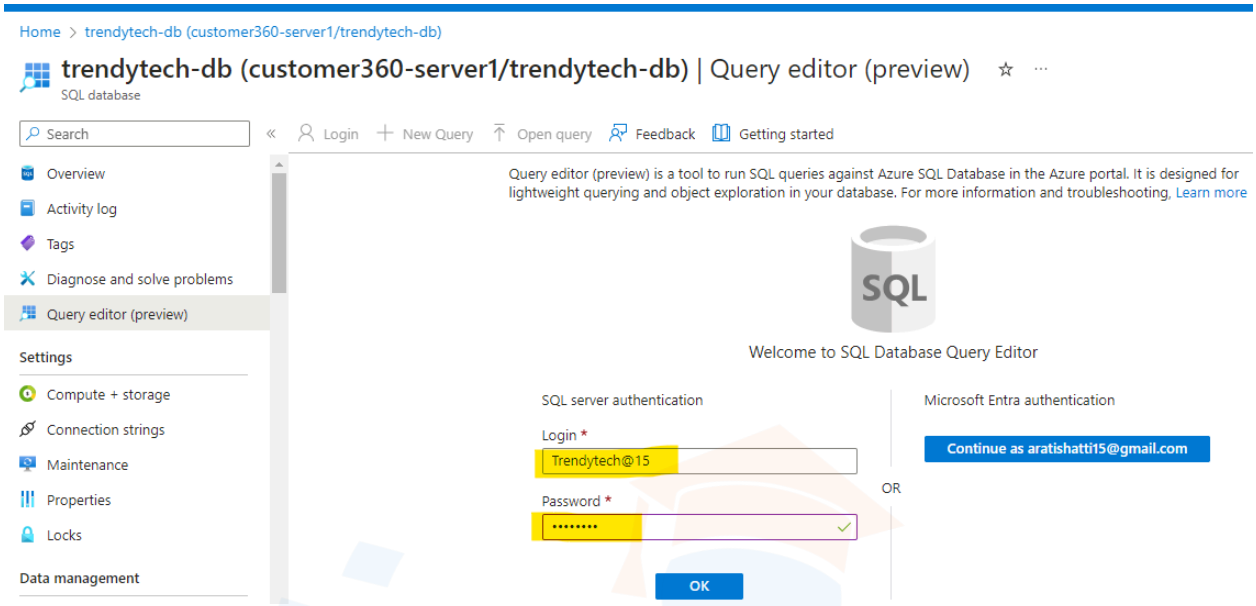


- Login to the SQL terminal in your azure account (You could also access the SQL server from your local system by installing and remotely connecting to Azure Data Studio or SQL server management studio)

Note: Here Query editor is used refer attached screenshot you can use this option also.



Mention the username and password that you used for SQL authentication



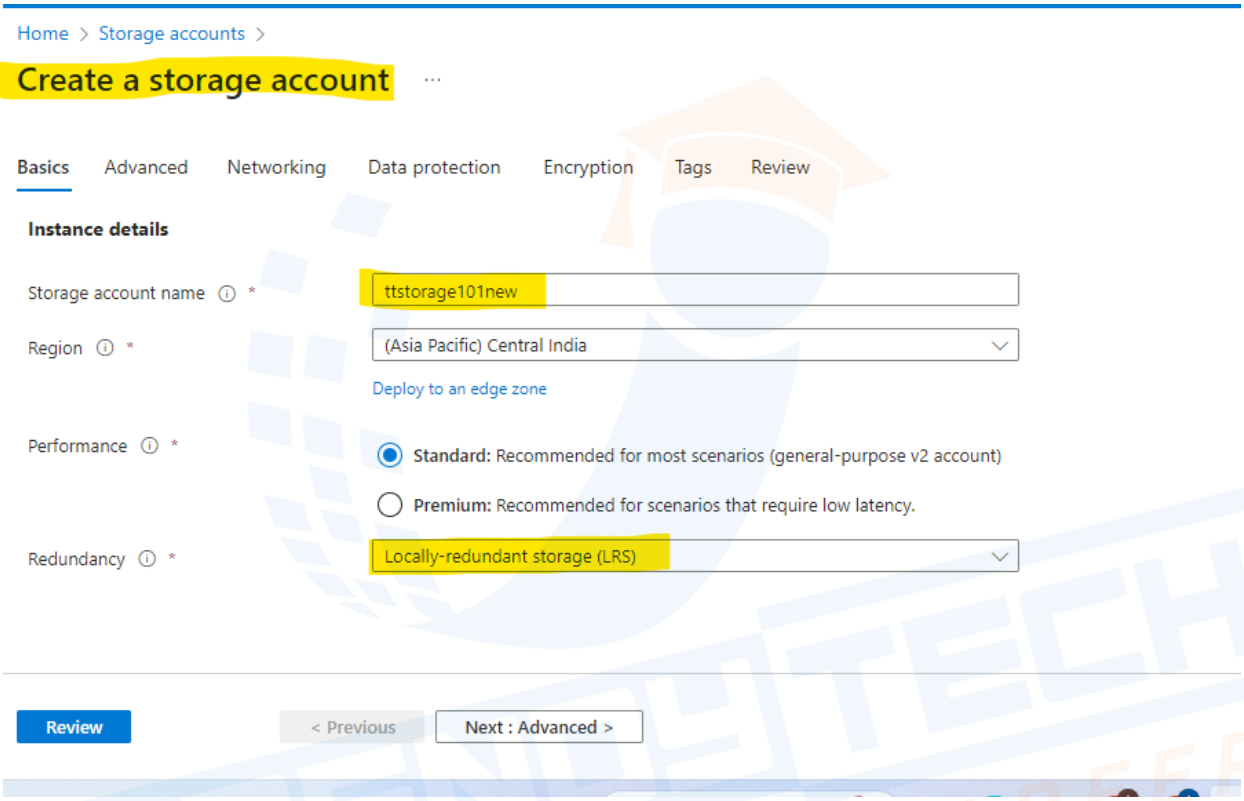
- Create Table (Courses Table) - Insert the Data into the table

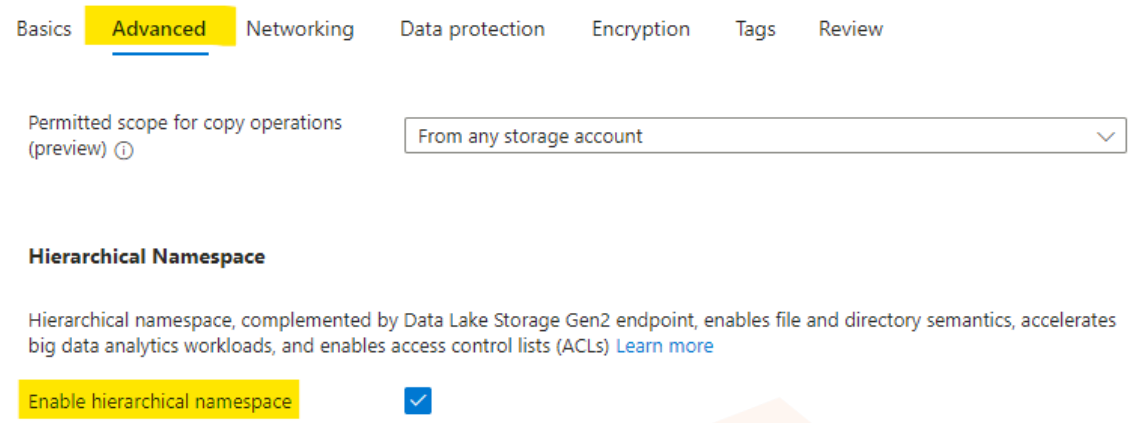
```
create table courses(course_id int NOT NULL, course_name varchar(30) NOT NULL, course_duration_months int NOT NULL, course_fee int NOT NULL, PRIMARY KEY (course_id)) ;

insert into courses values (1, 'bigdata', 6, 50000);
insert into courses values (2, 'webdevelopment', 3, 20000);
insert into courses values (3, 'datascience', 6, 40000);
insert into courses values (4, 'devops', 1, 10000);
```

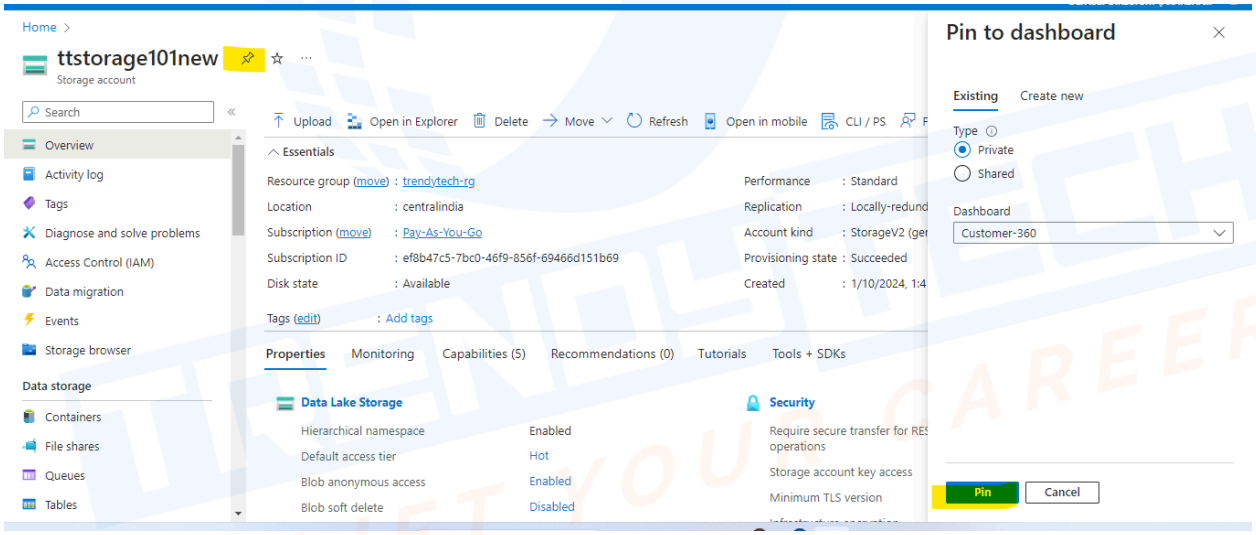
Creation of Sink System - ADLS Gen2

- Create a Storage Account (Enable Hierarchical namespace to make it as data lake storage and not just the blob storage)

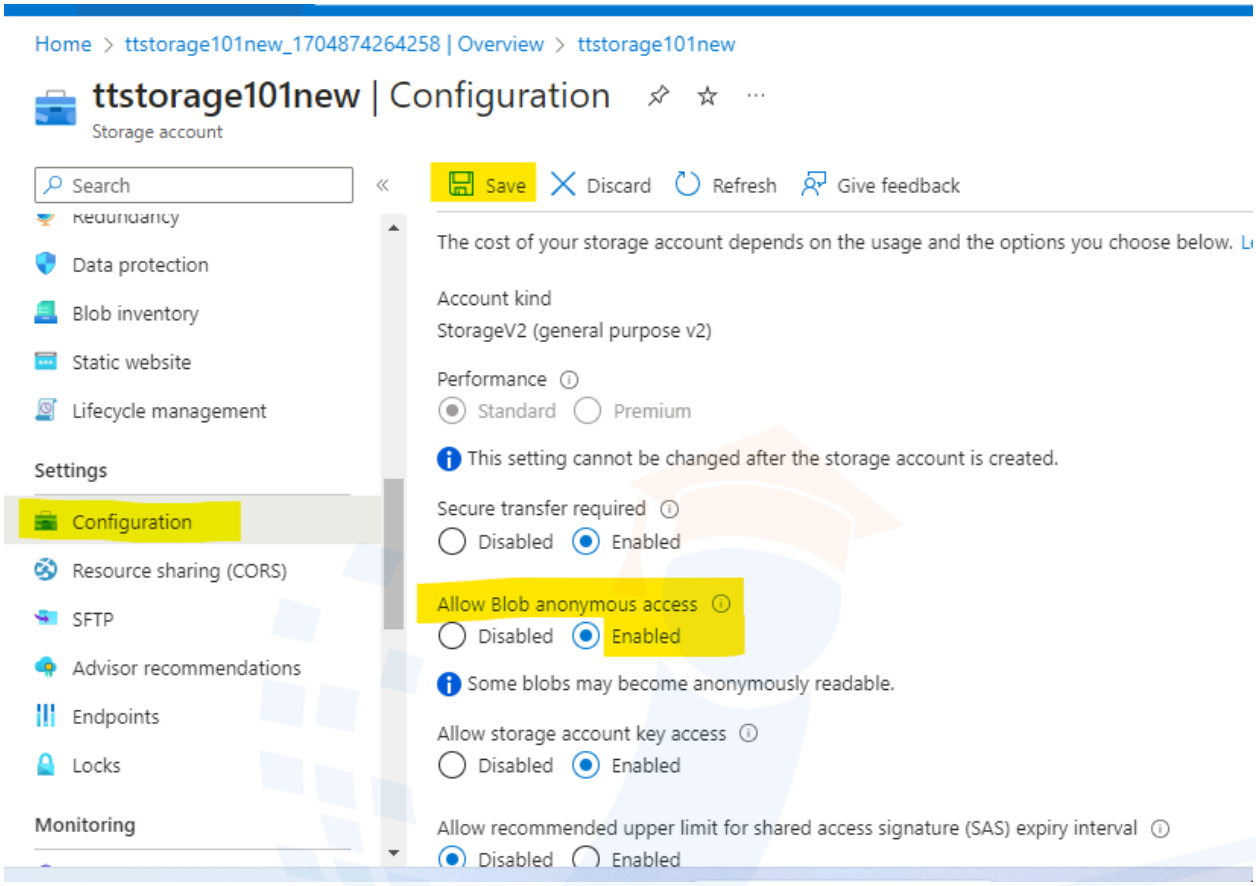




Note: Once a resource has been created, you have the option to add it to the dashboard for quick access. Refer attached screenshot.

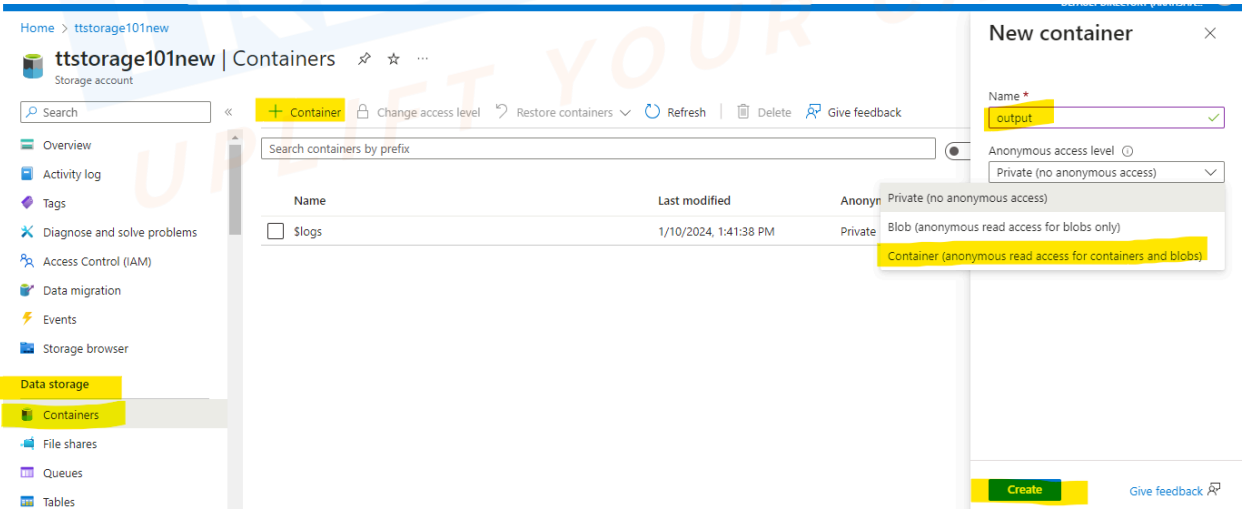


Note : After creating resources go to Setting => Configuration => Enable (Allow Blob anonymous access) => click on “Save”



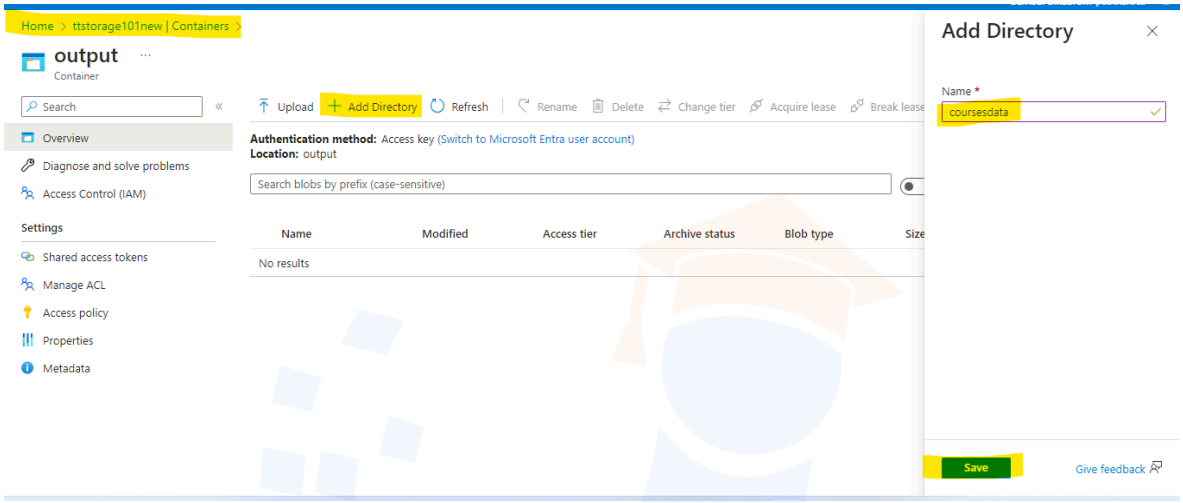
- Create a Container inside the Storage Account. First to go your resource

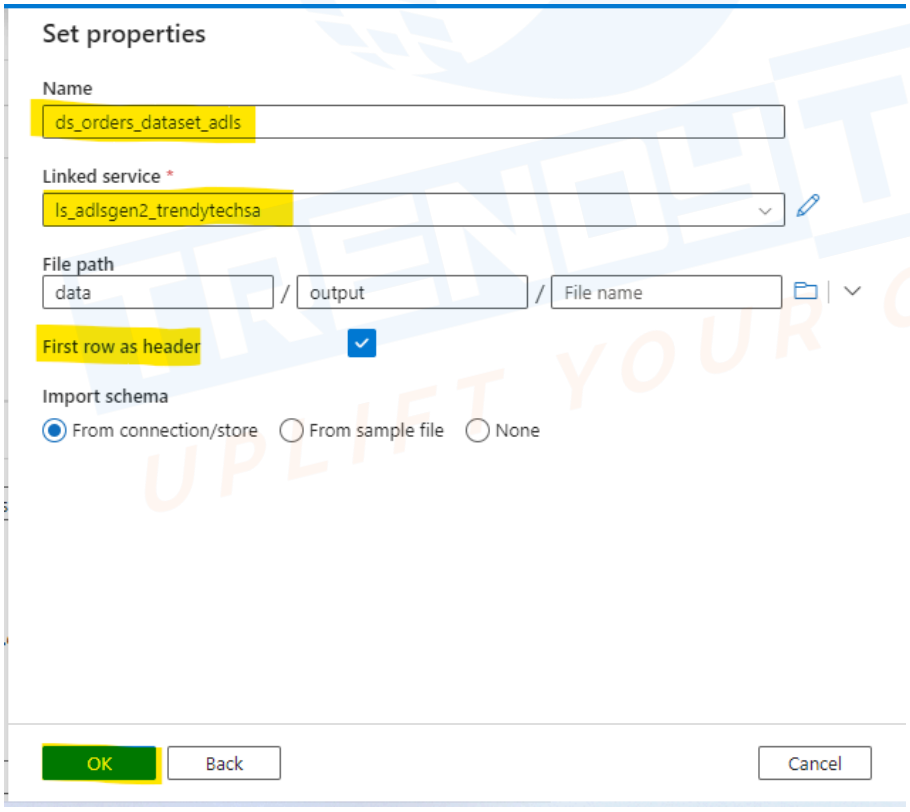
Click on Container option in Data Storage => new container => set Anonymous access level to “Container (anonymous access for container and blob)” as shown in below screenshot and create it.



- Create a Directory under the Container.

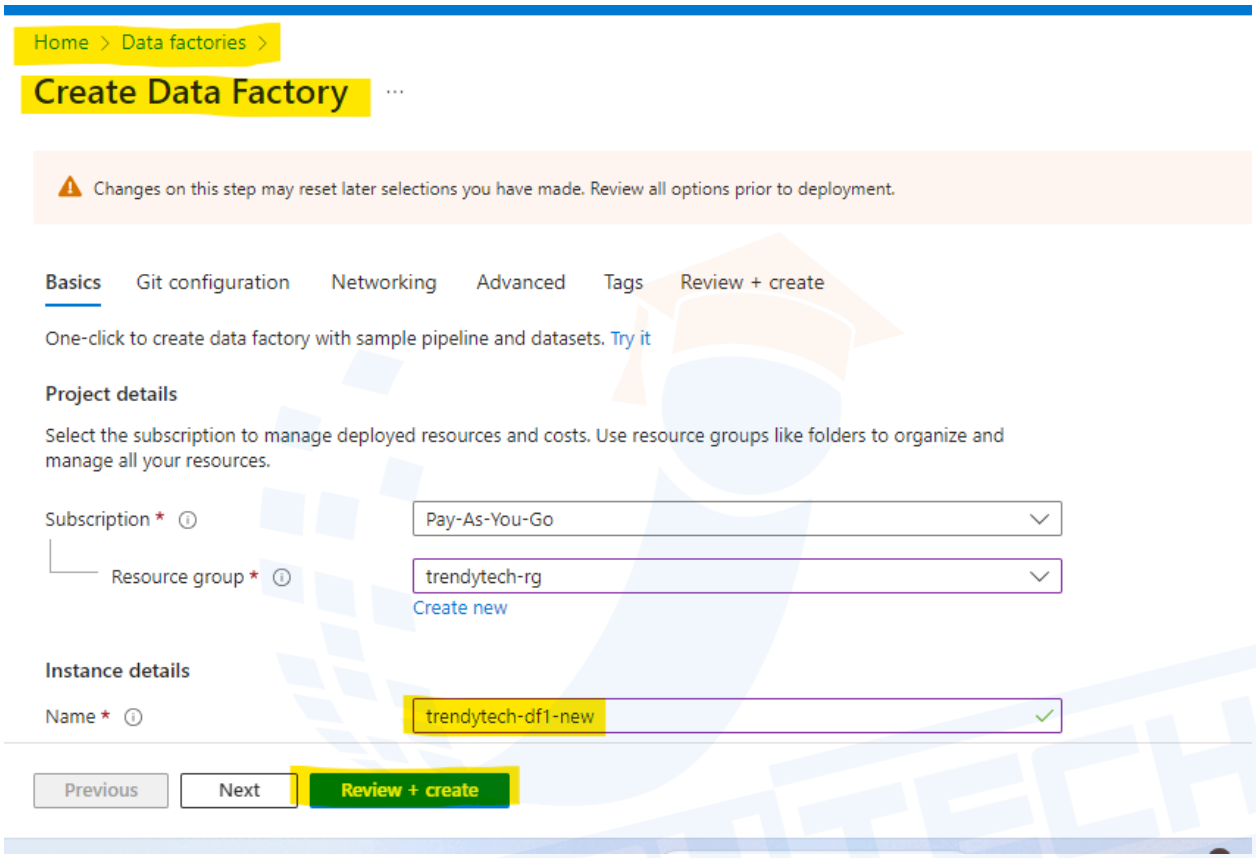
Select the container (here “output”) => click on Add Directory option.





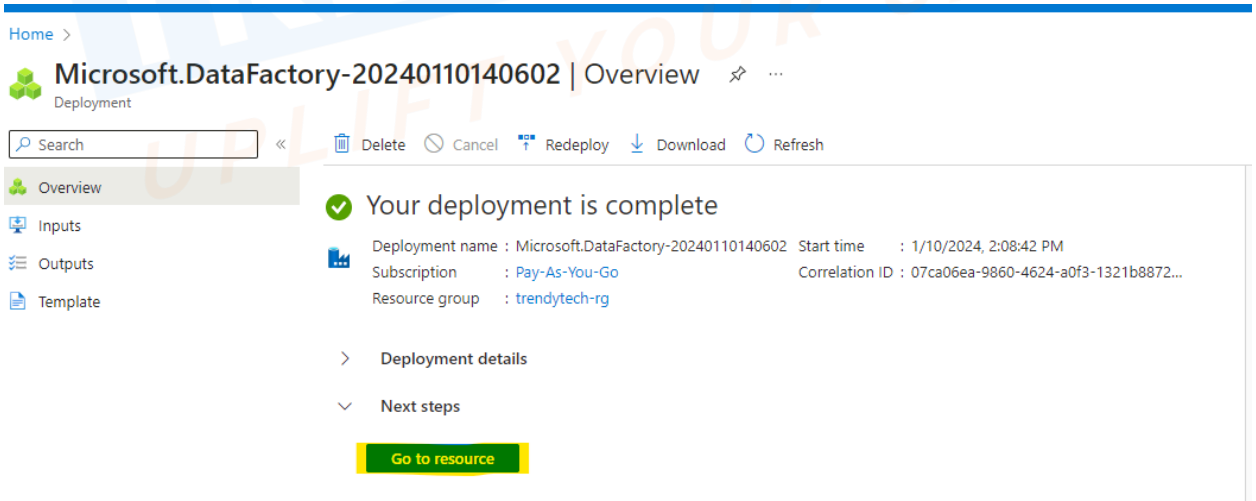
Creation of Data Integration Service - Azure Data Factory (To ingest data from source and load it to sink)

- Create Azure Data Factory resource.



- Open the Azure Data Factory Studio after the resource is deployed.

Click on Go to resource => launch

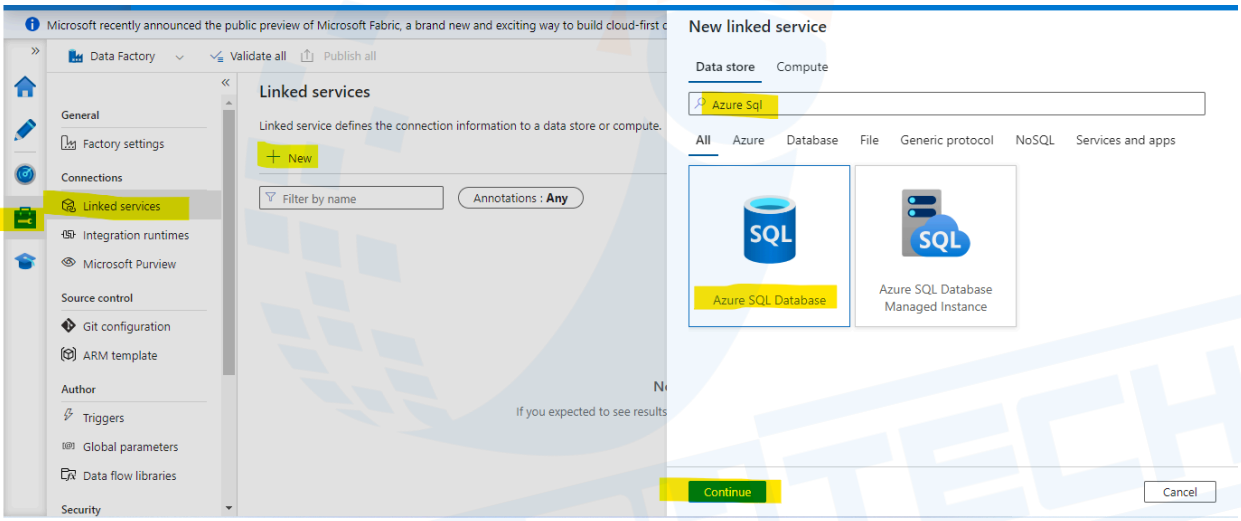


- Firstly, connect to the source and sink using the Linked Service.

- Connect to the Source using the Linked Service Azure SQL Database. (Authentication to connect to SQL server and enable this service to be accessible by other services under firewall settings).

Check below screenshot for **creating linked service** for our source i.e Azure SQL Database

Go to Monitor => Linked Service => New => search Azure SQL and select Azure SQL Database => Continue



New linked service

Azure SQL Database Learn more

Name *

ls_sql_db

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Connection string | Azure Key Vault

Account selection method

From Azure subscription | Enter manually

Azure subscription

Pay-As-You-Go (ef8b47c5-7bc0-46f9-856f-69466d151b69)

New linked service

Azure SQL Database Learn more

Azure subscription

Pay-As-You-Go (ef8b47c5-7bc0-46f9-856f-69466d151b69)

Server name *

customer360-server1

Database name *

trendytech-db

Authentication type *

SQL authentication

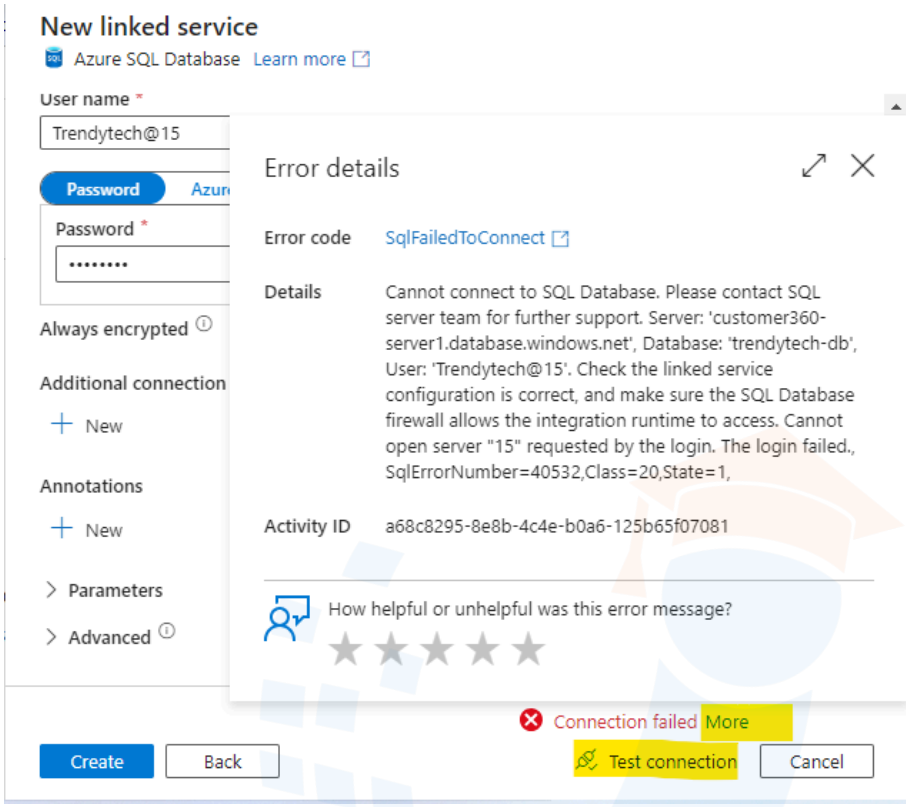
User name *

Trendytech@15

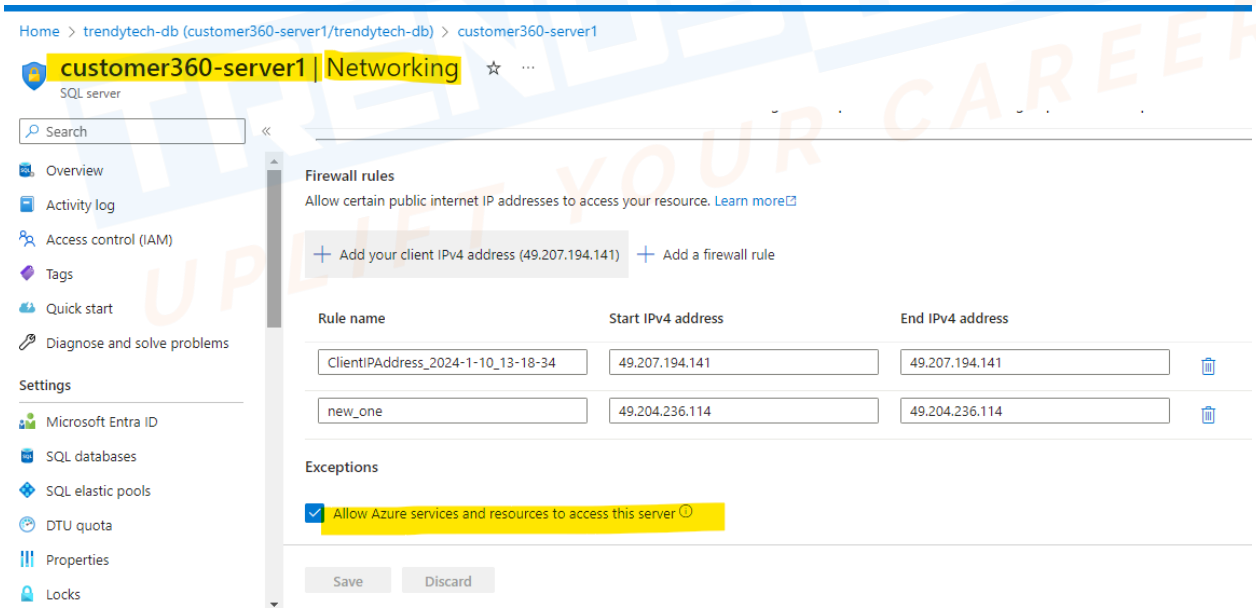
Password | Azure Key Vault

Password *

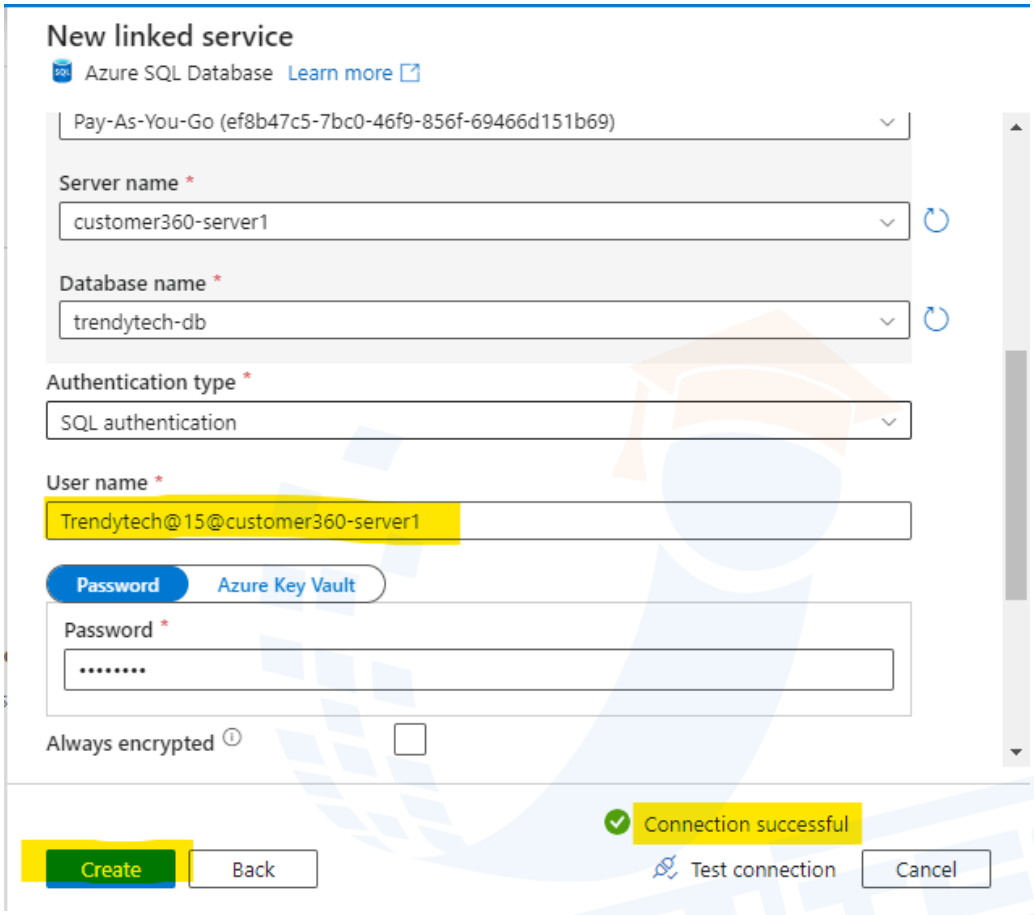
••••••••



If you get this error go to the server And allow in firewall rule Allow Azure services and resources to access this server

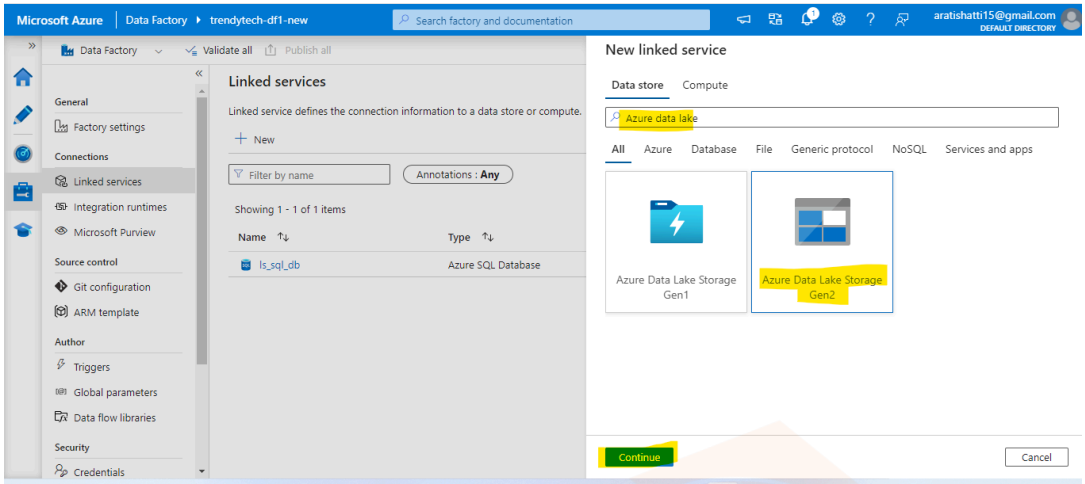


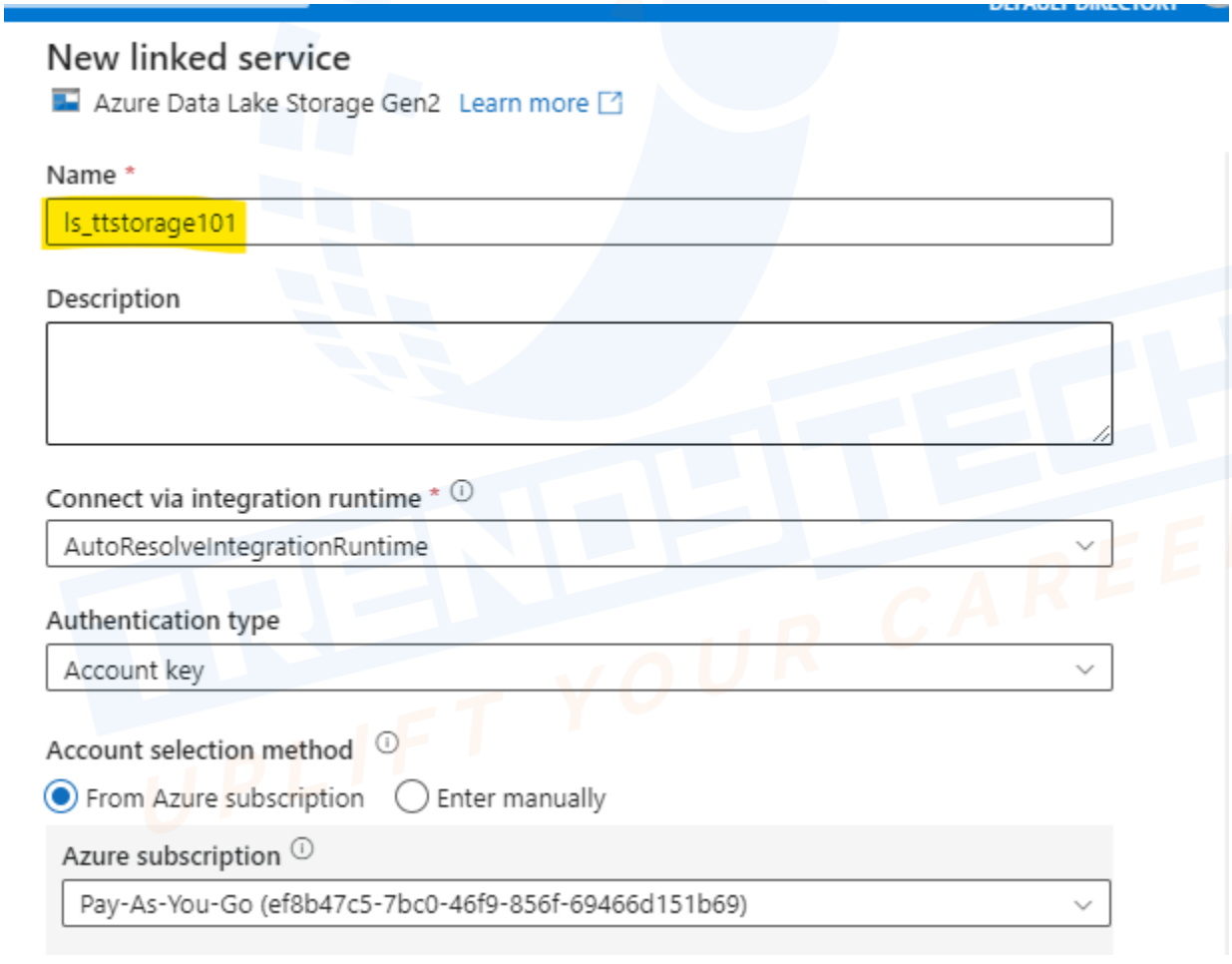
If you still get this error in username mention <username>@<servername> here Trendytech@15@customer360-server1. Refer below screenshot.

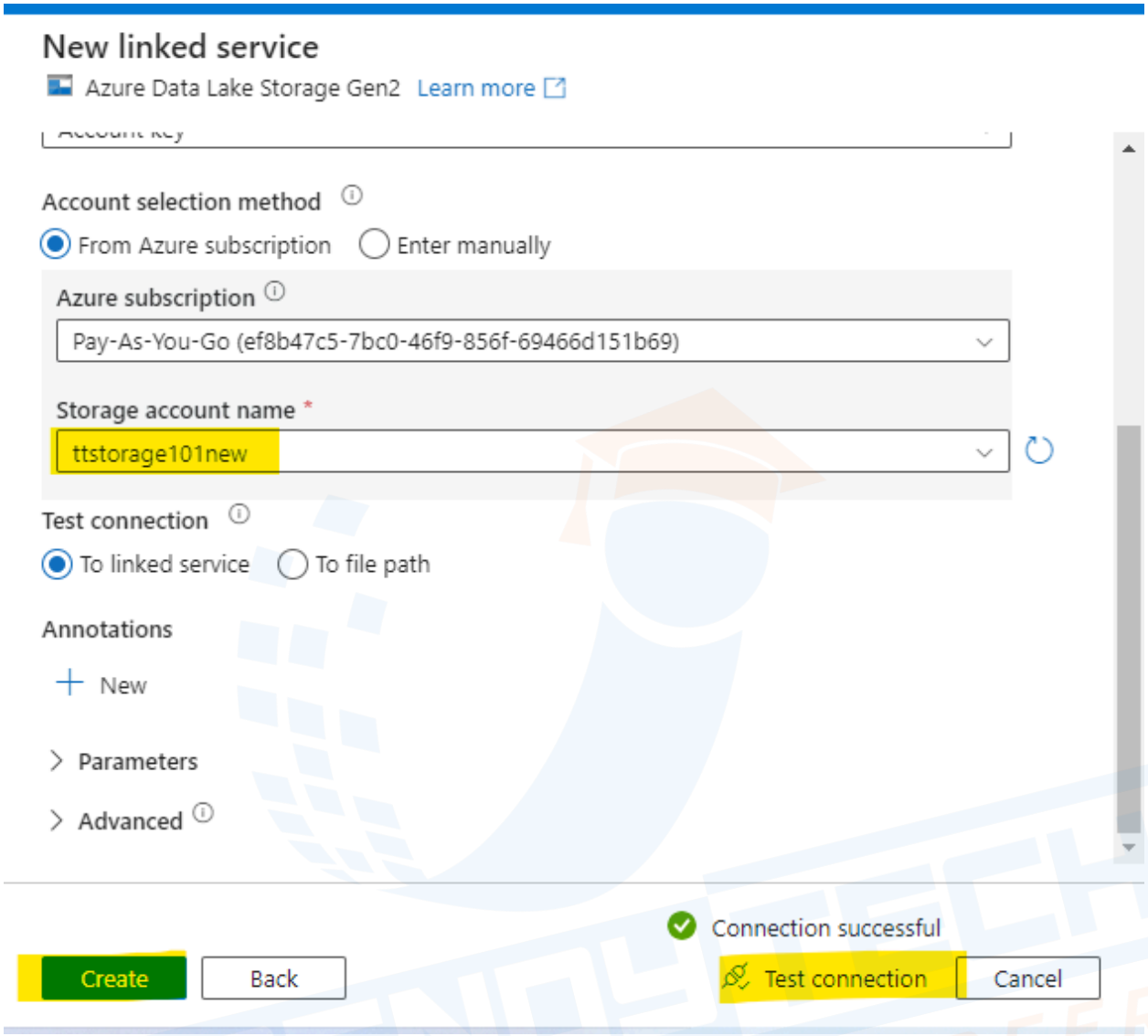


- Connect to the Sink using the Linked Service ADLS Gen2.

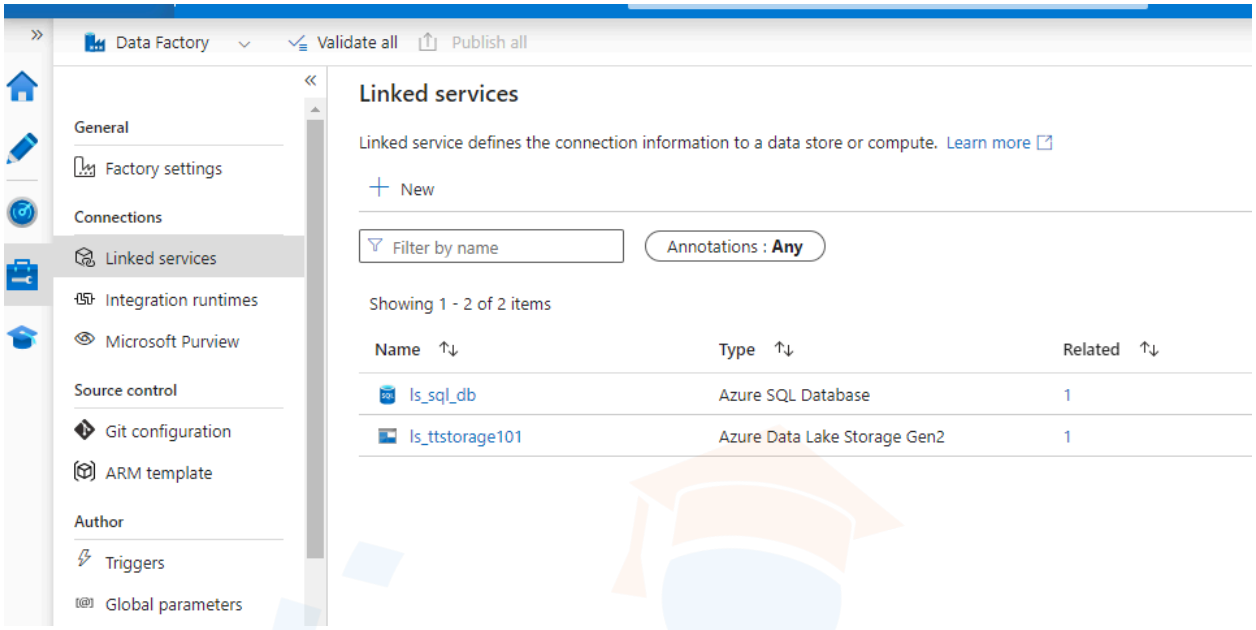
Check below screenshot for creating linked service for our sink i.e ADLS gen storage


Microsoft Azure
Data Factory
trendytech-df1-new
Search factory and documentation
aratishatti15@gmail.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all
General
Factory settings
Connections
Linked services
Integration runtimes
Microsoft Purview
Source control
Git configuration
ARM template
Author
Triggers
Global parameters
Data flow libraries
Security
Credentials
Linked services
Linked service defines the connection information to a data store or compute.
New
Filter by name
Annotations : Any
Showing 1 - 1 of 1 items
Name
Type
ls_sql_db
Azure SQL Database
New linked service
Data store
Compute
Azure data lake
All
Azure
Database
File
Generic protocol
NoSQL
Services and apps
Azure Data Lake Storage Gen1
Azure Data Lake Storage Gen2
Continue
Cancel


New linked service
Azure Data Lake Storage Gen2
Learn more
Name *
ls_ttstorage101
Description
Connect via integration runtime *
AutoResolveIntegrationRuntime
Authentication type
Account key
Account selection method
From Azure subscription
Enter manually
Azure subscription
Pay-As-You-Go (ef8b47c5-7bc0-46f9-856f-69466d151b69)

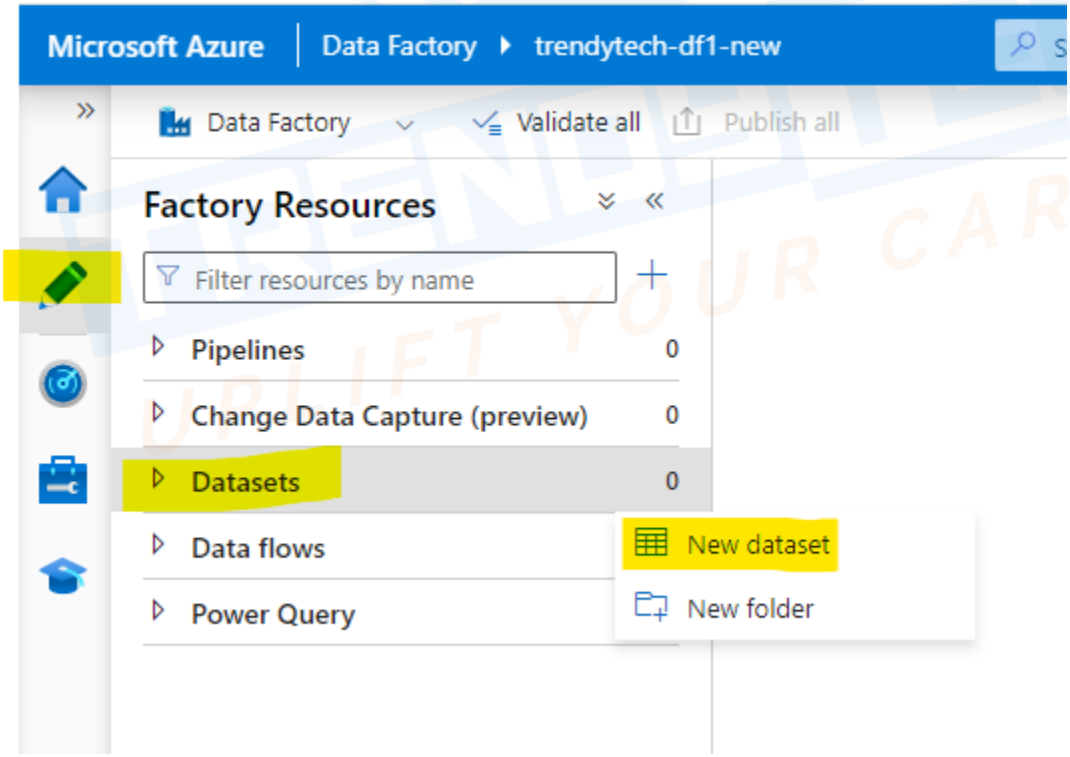


Note: After creating new linked services in Azure Data Factory, be sure to publish these changes to make them active and available for use in your data workflows.

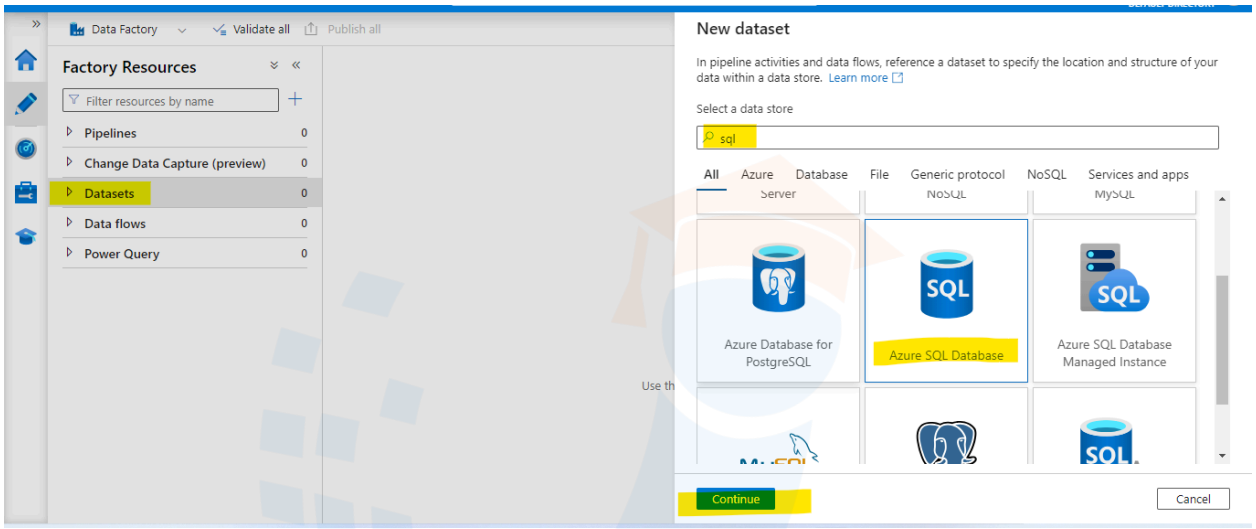


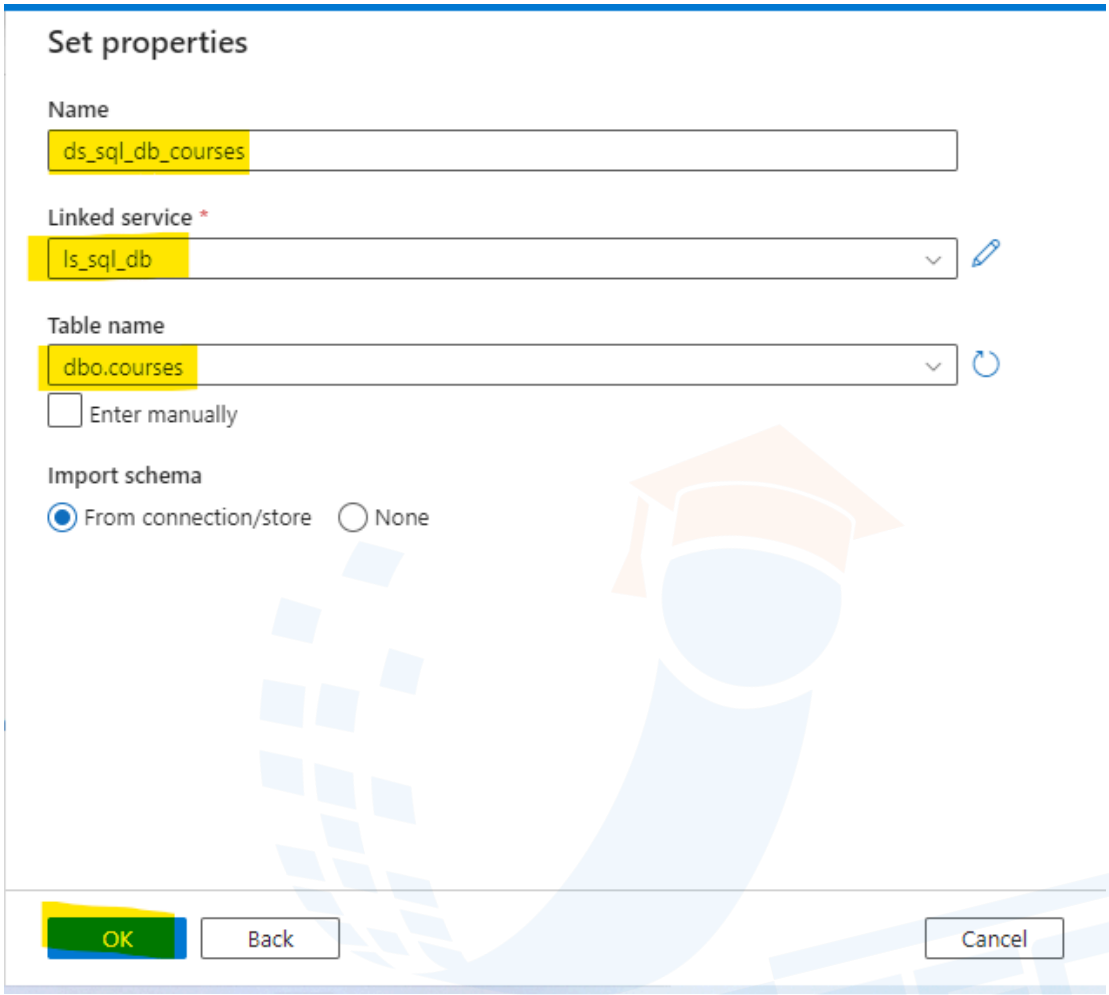
- Create and Publish Datasets on-top of Source and Sink indicating the format type of data to be stored underneath.

To create dataset click on Author => Datasets => New datasets

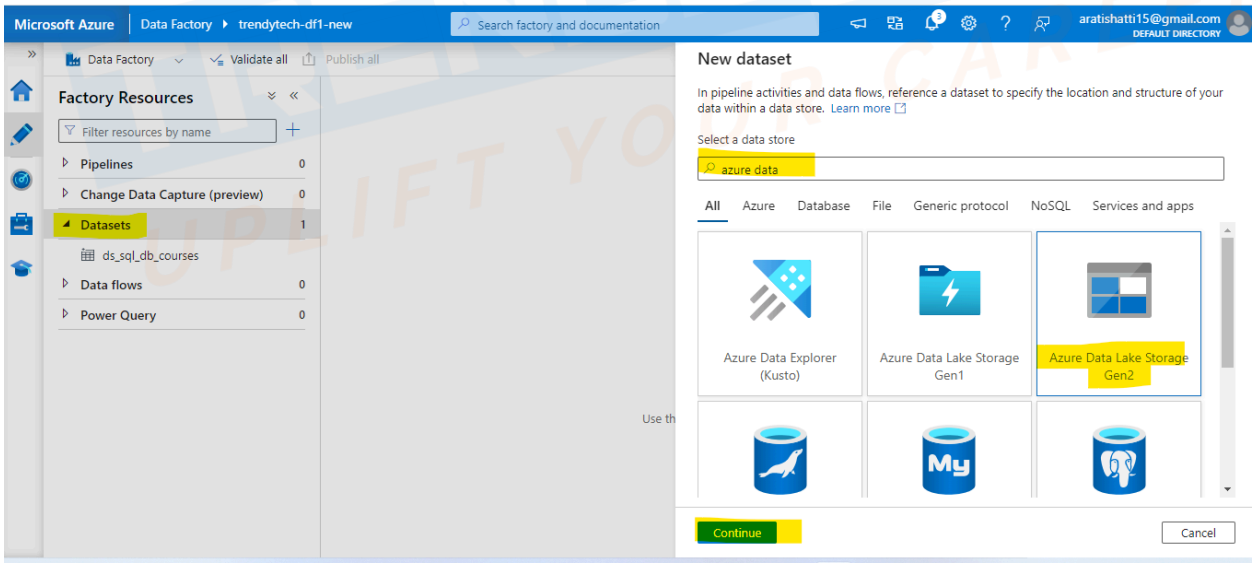


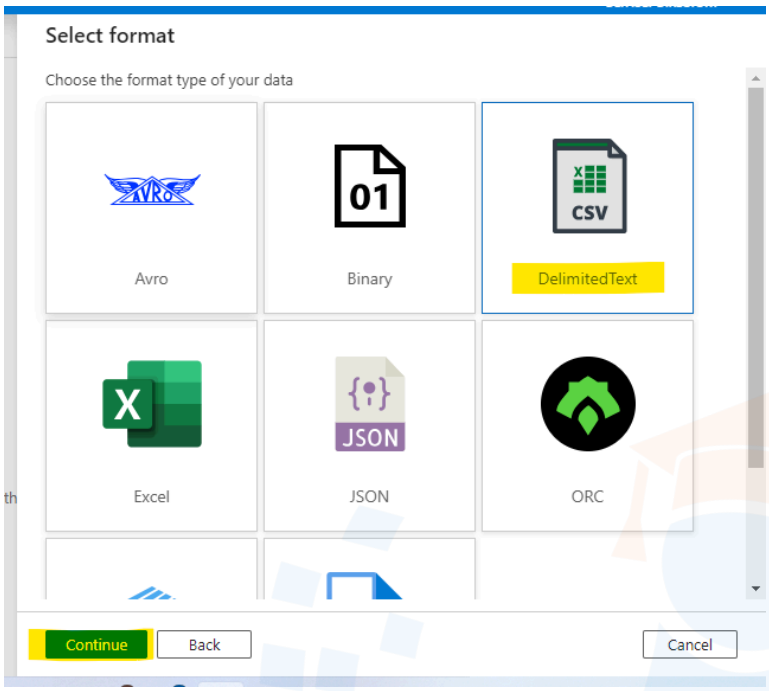
Dataset for source (courses table)

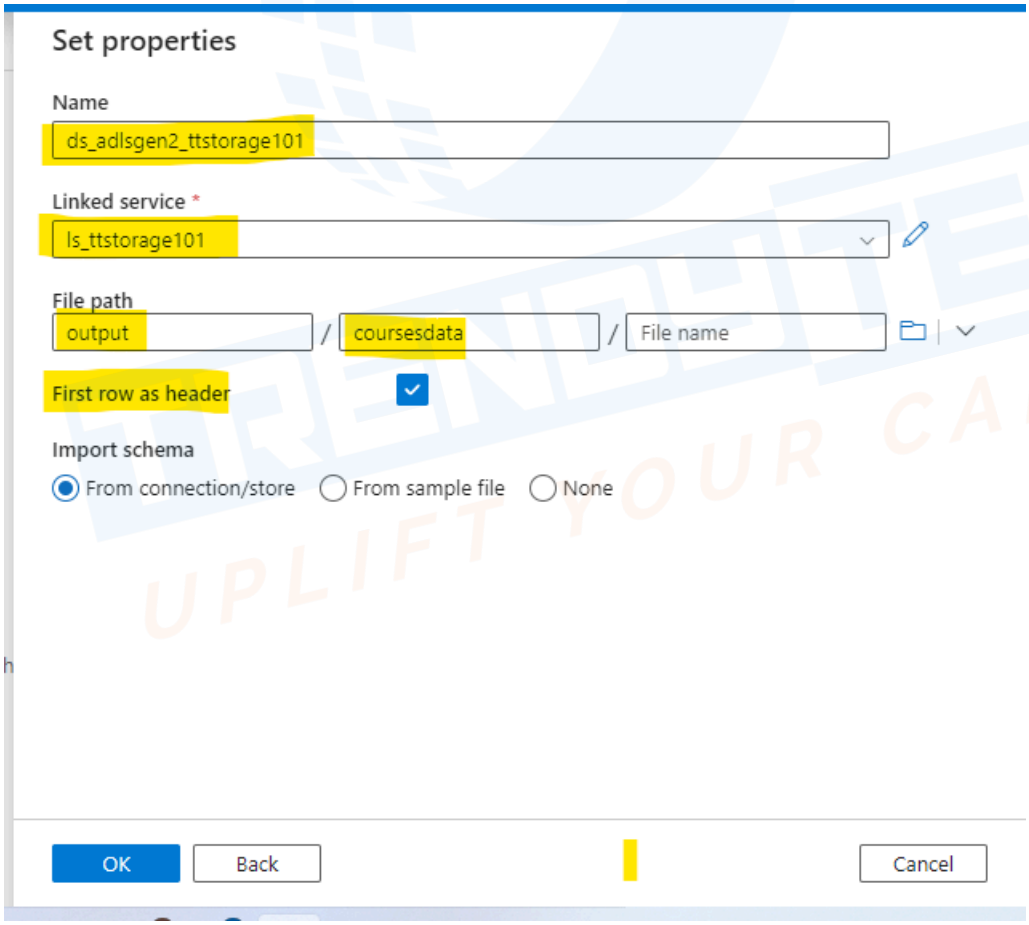




Dataset creation for sink please check below steps:




Select format
Choose the format type of your data
Avro
Binary
DelimitedText
Excel
JSON
ORC
Continue
Back
Cancel

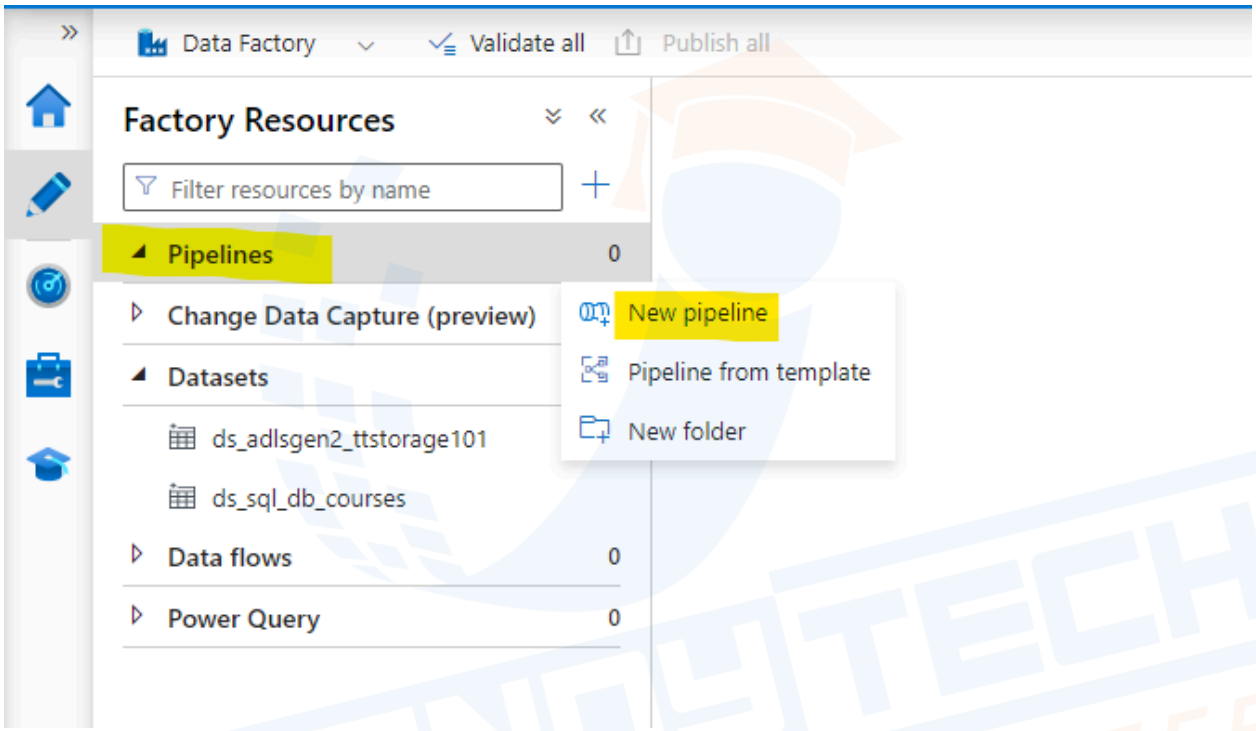

Set properties
Name
ds_adlsgen2_ttstorage101
Linked service *
ls_ttstorage101
File path
output
coursesdata
File name
First row as header
Import schema
From connection/store
From sample file
None
OK
Back
Cancel

Note: After creating new datasets in Azure Data Factory, be sure to publish these changes to make them active and available for use in your data workflows.

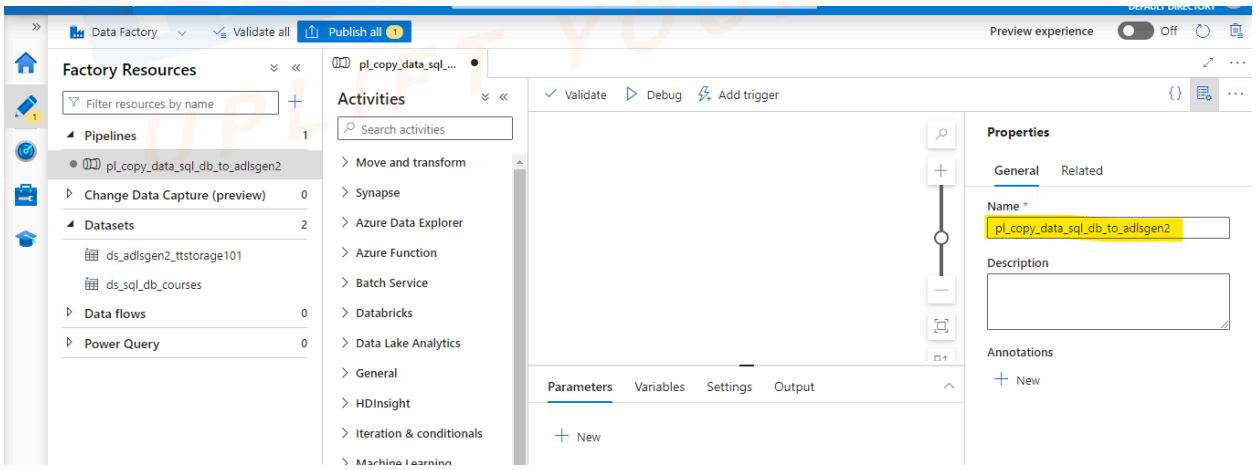
- Create a Pipeline and a Copy activity within the pipeline. Select the Source as Azure SQL DB and the Sink as ADLS Gen2 as created in the previous steps.

Now we will create the pipeline and we will keep the copy activity in it

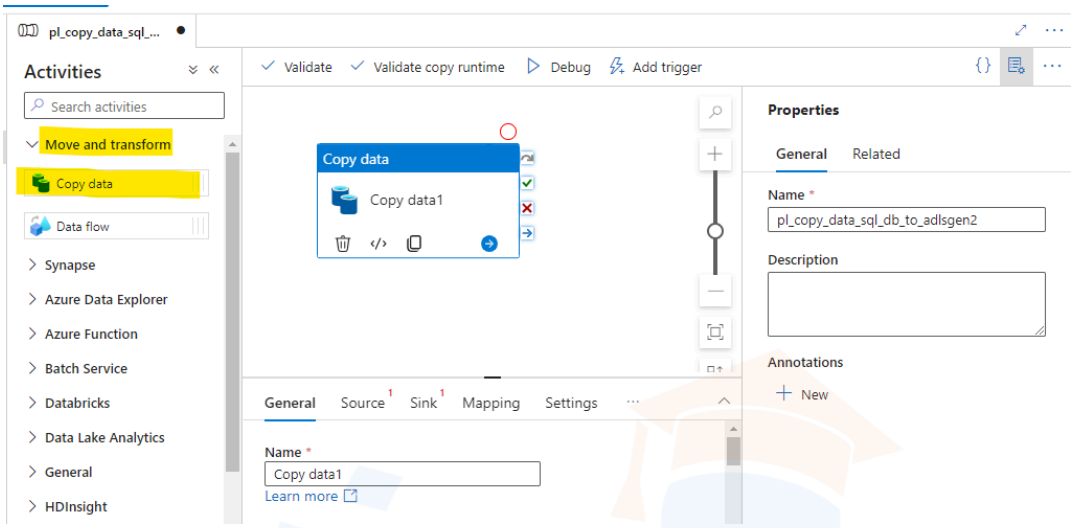
Click on three dots for pipelines option you will get below screen



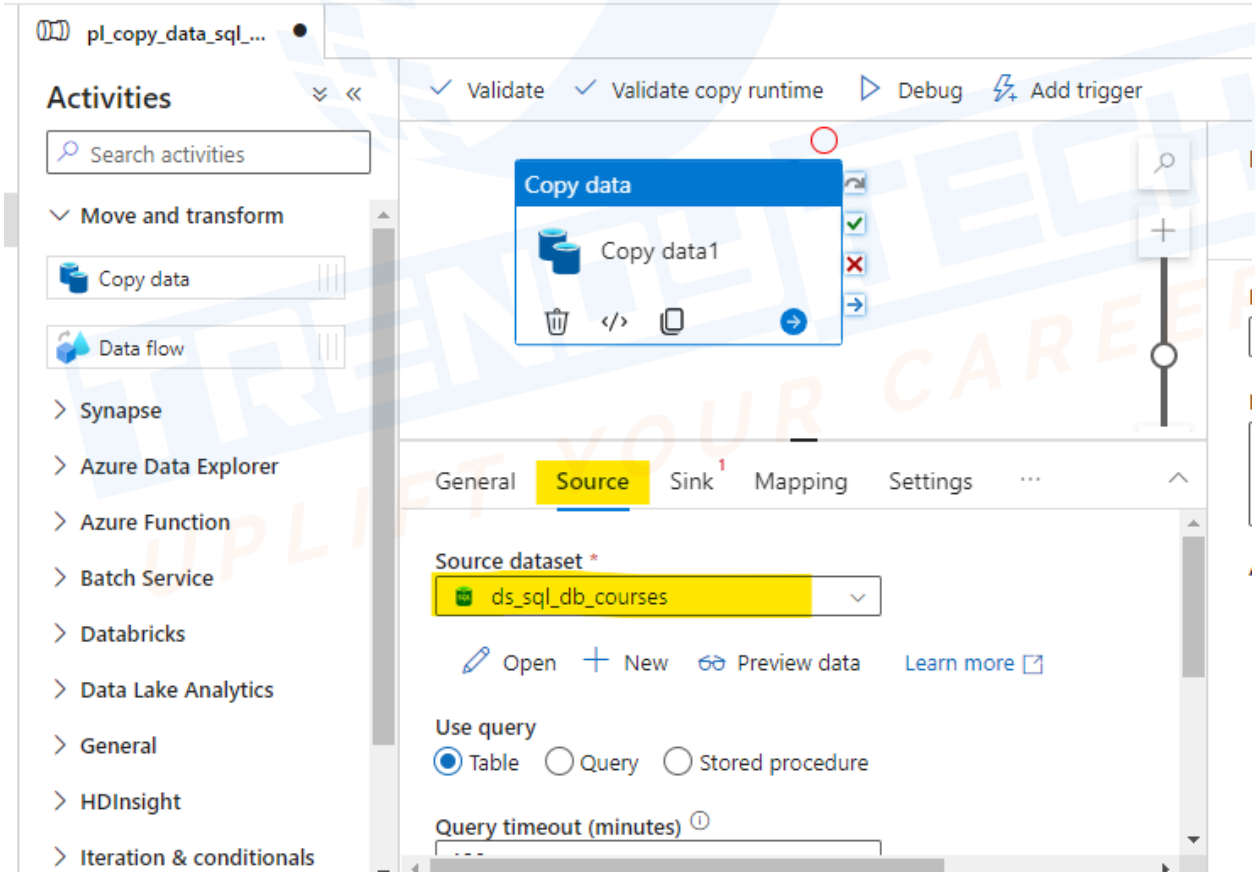
Give name as per your choice highlighted as below



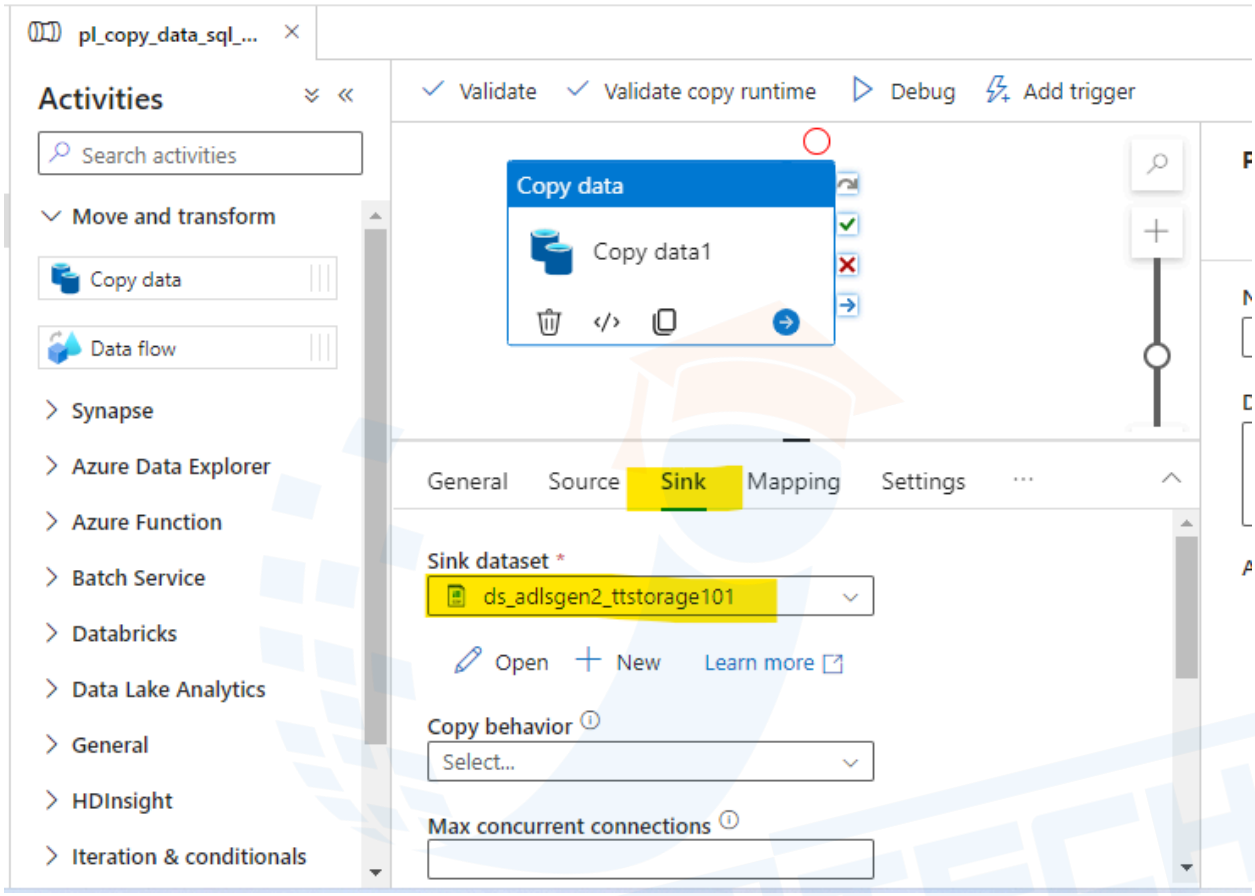
Now click on “Move and transform” and drag copy activity in the pipeline as shown below.



Now in source select dataset for course table in sql db here “ds_sql_db_courses”

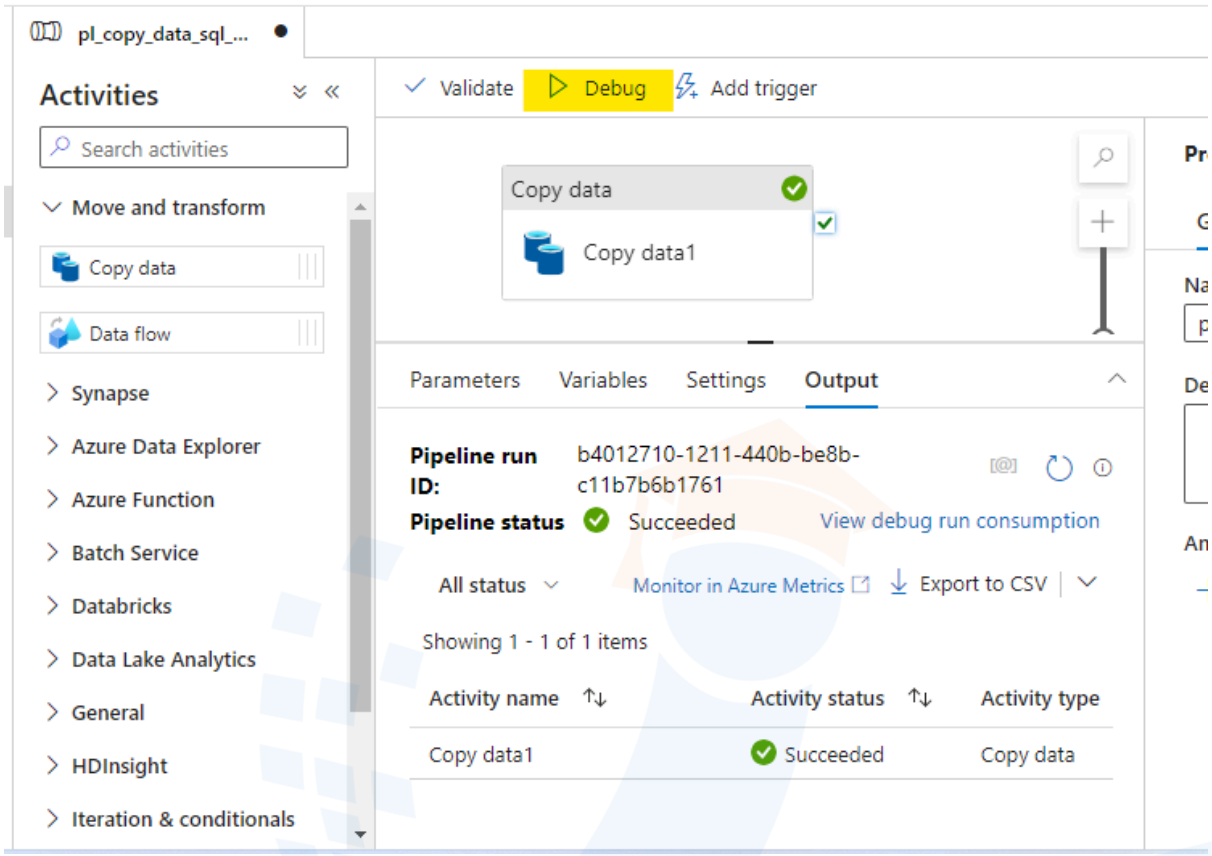


Now in sink select dataset for sink here “ds_adlsgen2_ttstorage101”

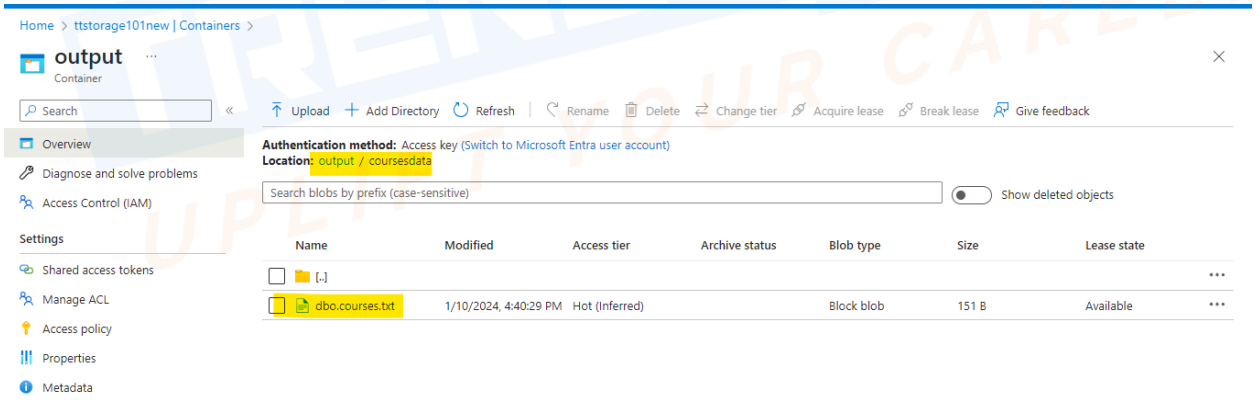


- Debug the pipeline and validate it.

Note: After creating new pipelines in Azure Data Factory, be sure to publish these changes to make them active and available for use in your data workflows.

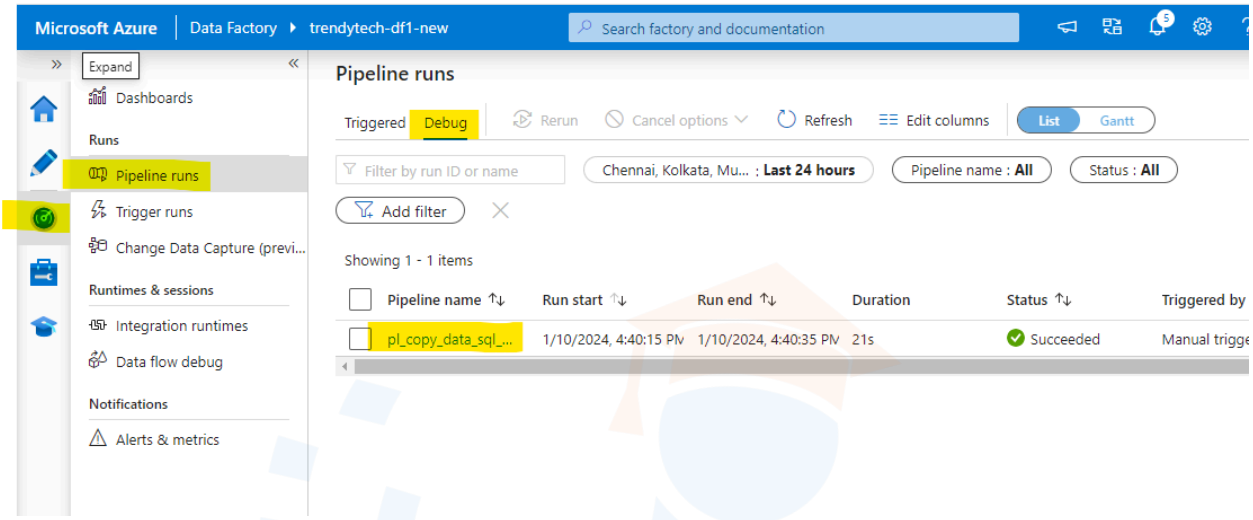


Also in your ADLS gen2 storage you will see the data copied after successfully running the pipeline. Refer below screenshot.



- Monitor the pipeline under the monitor tab.

Note : here after the Debug option it will be listed in the debug option in “pipeline runs”. If you click on the Trigger option it will be listed in the Trigger option.



Note: At the end please delete all the resources that you have created.