

## PROJECT-03

### Design a dynamic pipeline that facilitates the workflow

NANDINI RATHORE

#### INDEX

1. Extract data from 5 tables from an on-premises SQL Server.....**2-8**
2. Load the extracted data into Azure Data Lake Storage Gen2 in a structured format.....**8-17**
3. Copy the data from ADLS Gen2 back to a designated file system folder on the local C: drive...**18-27**
4. Implement a metadata-driven approach using a control table that stores details such as:.....**27-32**
1. Source table names
2. Target file paths
5. Implement the email configuration task by using logic apps and send an email for any failure...**32-38**

***Design a dynamic pipeline that facilitates the following workflow:***

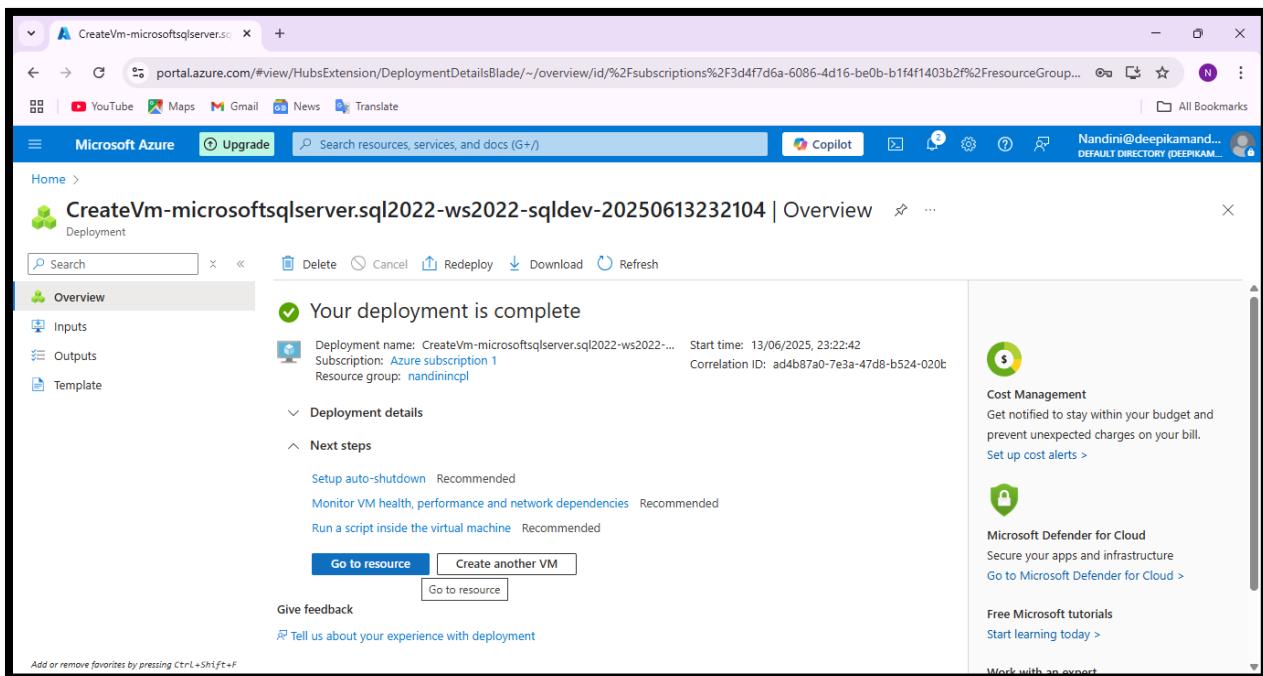
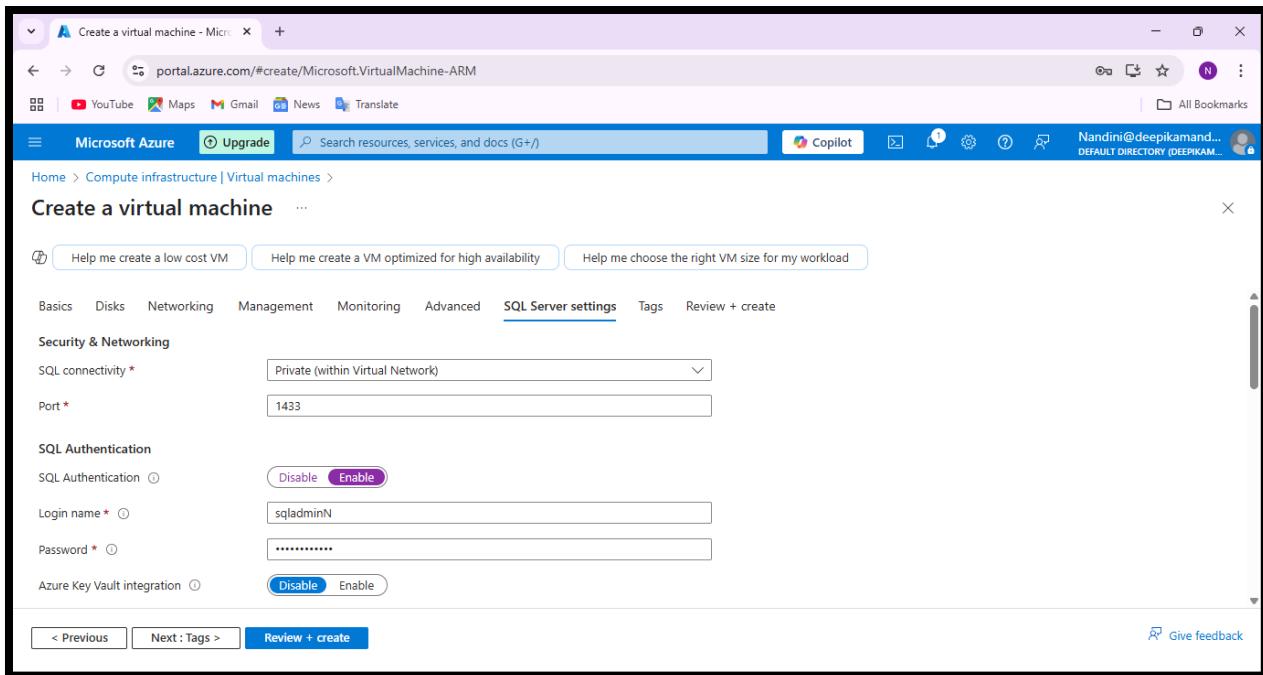
**1. Extract data from 5 tables from an on-premises SQL Server.**

The screenshot shows the Microsoft Azure portal interface. The user is on the 'Compute infrastructure | Virtual machines' page. The left sidebar has 'Virtual machines' selected. The main content area says 'No virtual machines to display' and includes a 'Create' button. There are also filter options and a note about viewing a new version of the experience.

**1. Create VM.**

The screenshot shows the 'Create a virtual machine' wizard. The user is on the first step, 'Adding details'. They have selected 'Subscription' (Azure subscription 1) and 'Resource group' (nandinincpl). Under 'Instance details', they have chosen 'Virtual machine name' (VMsqlN), 'Region' ((US) West US 2), 'Availability options' (No infrastructure redundancy required), 'Security type' (Standard), and 'Image' (Free SQL Server License: SQL Server 2022 Developer on Windows Server 2022). At the bottom, there are navigation buttons: '< Previous', 'Next : Disks >', 'Review + create', and 'Give feedback'.

**2. ADDING DETAILS.**



### 3. GO TO RESOURCES.

**VMsqliN** | Overview

**Connect**

Resource group (move) : nandinincpl

Status : Running

Location : West US 2

Subscription (move) : Azure subscription 1

Subscription ID : 3d4f7d6a-6086-4d16-be0b-b1f4f1403b2f

Operating system : Windows (Windows Server 2022 Datacenter)

Size : Standard D2s v5 (2 vcpus, 4 GiB memory)

Public IP address : 20.114.52.169

Virtual network/subnet : VMsqliN-vnet/default

DNS name : Not configured

Health state : -

Time created : 14/06/2025, 03:22 UTC

Tags (edit) : Add tags

**Properties**   **Monitoring**   **Capabilities (8)**   **Recommendations**   **Tutorials**

**Virtual machine**   **Networking**

#### 4.CONNECT VM.

**VMsqliN | Connect**

Port (change) 3389 Check access

Just-in-time policy Unsupported by plan

**Most common**

Native RDP Local machine

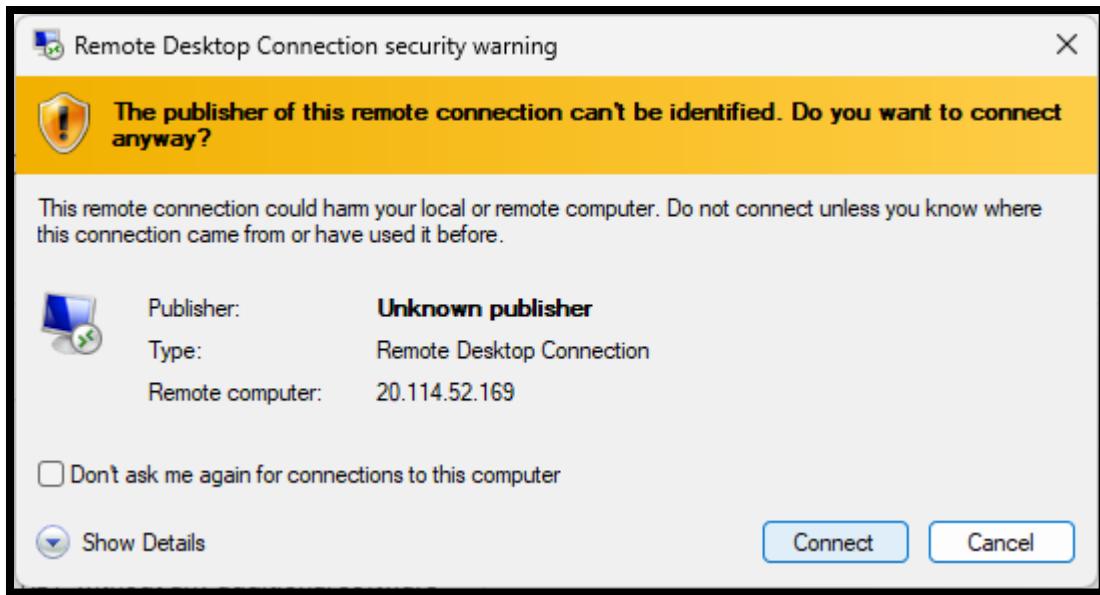
Connect via native RDP without any additional software needed. Recommended for testing only.

Public IP address (20.114.52.169)

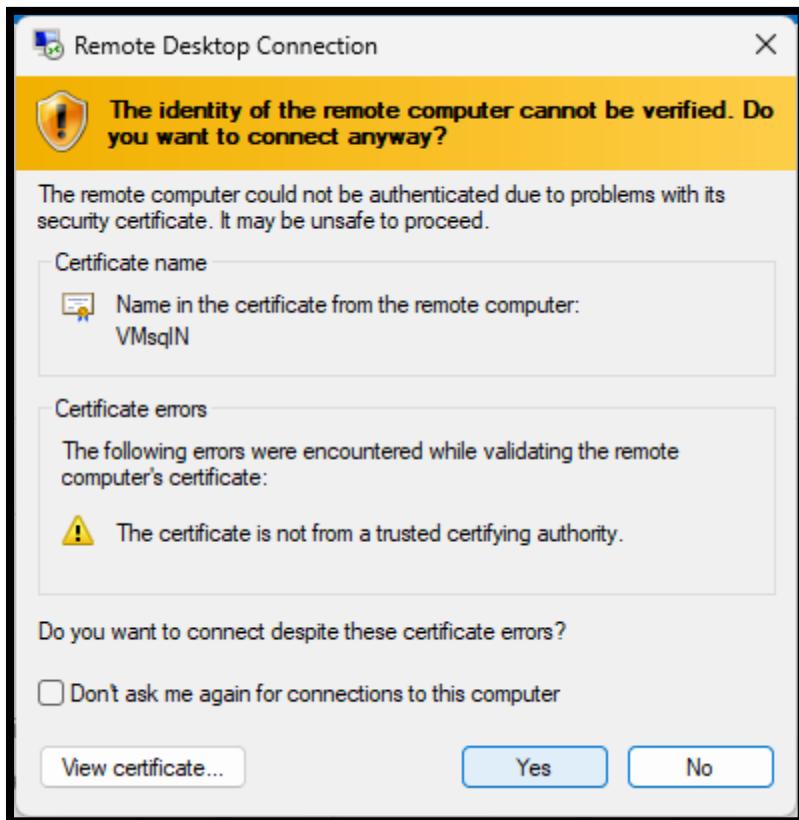
Select Download RDP file

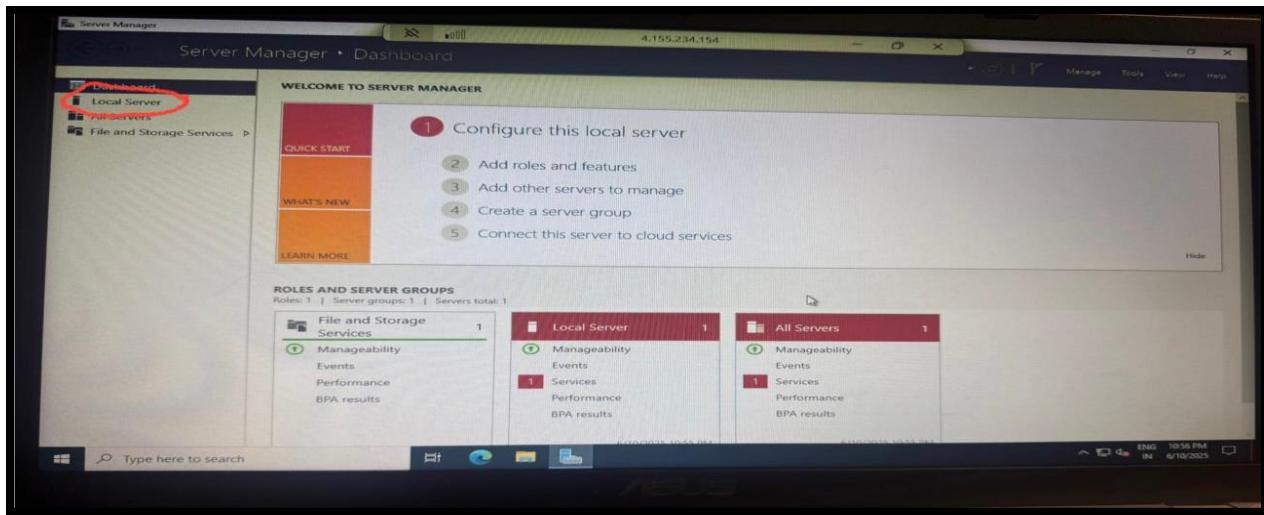
More ways to connect (4)

#### 4. DOWNLOAD RDP FILE.

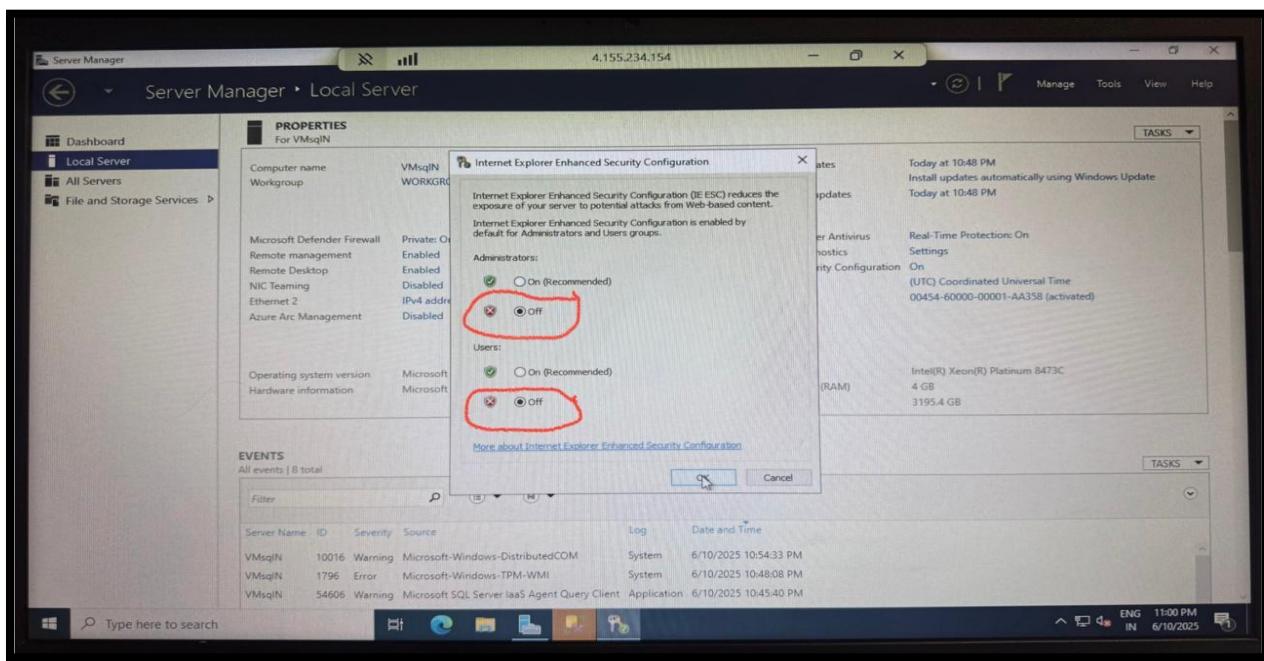


##### 5. CONNECT IT WITH SERVER.

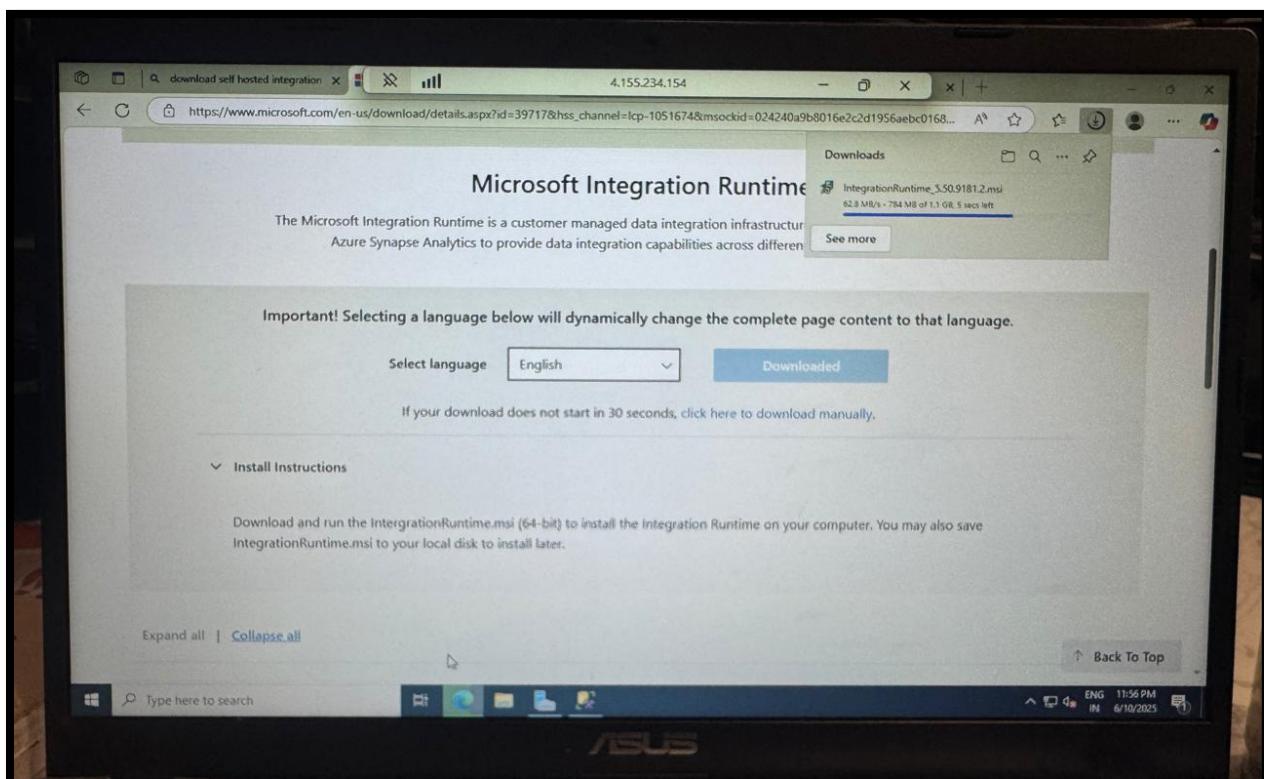
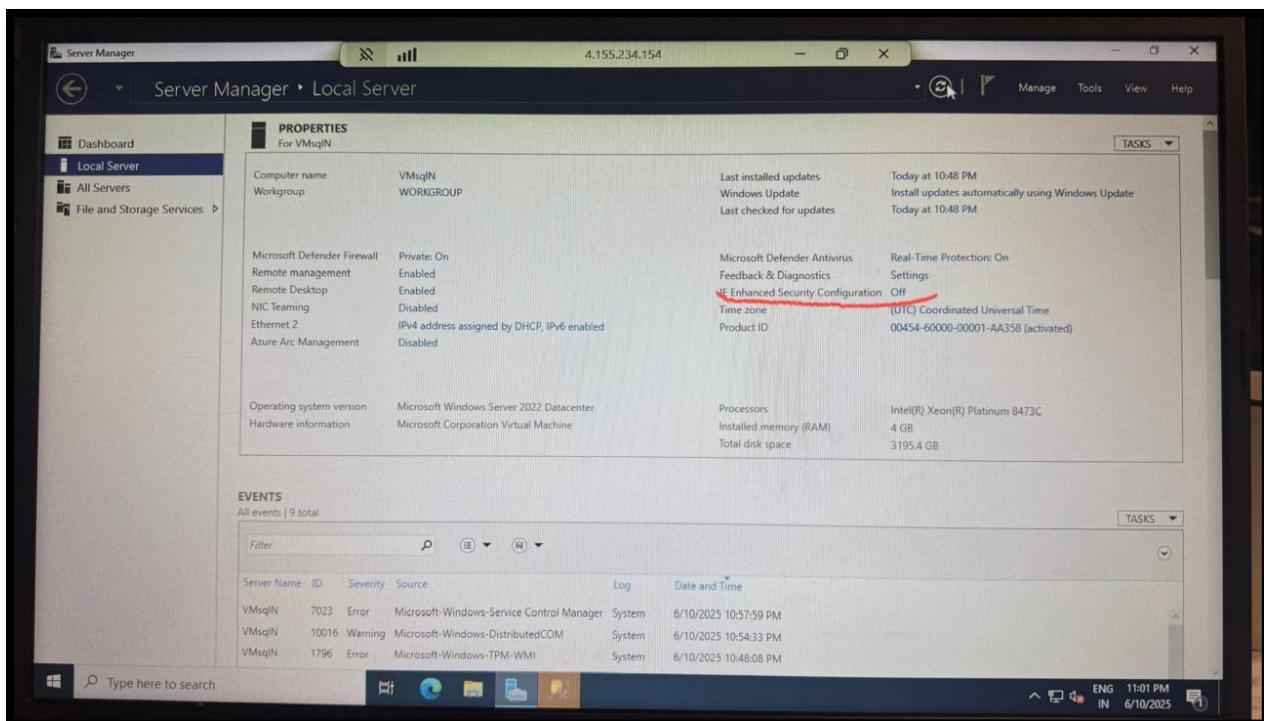




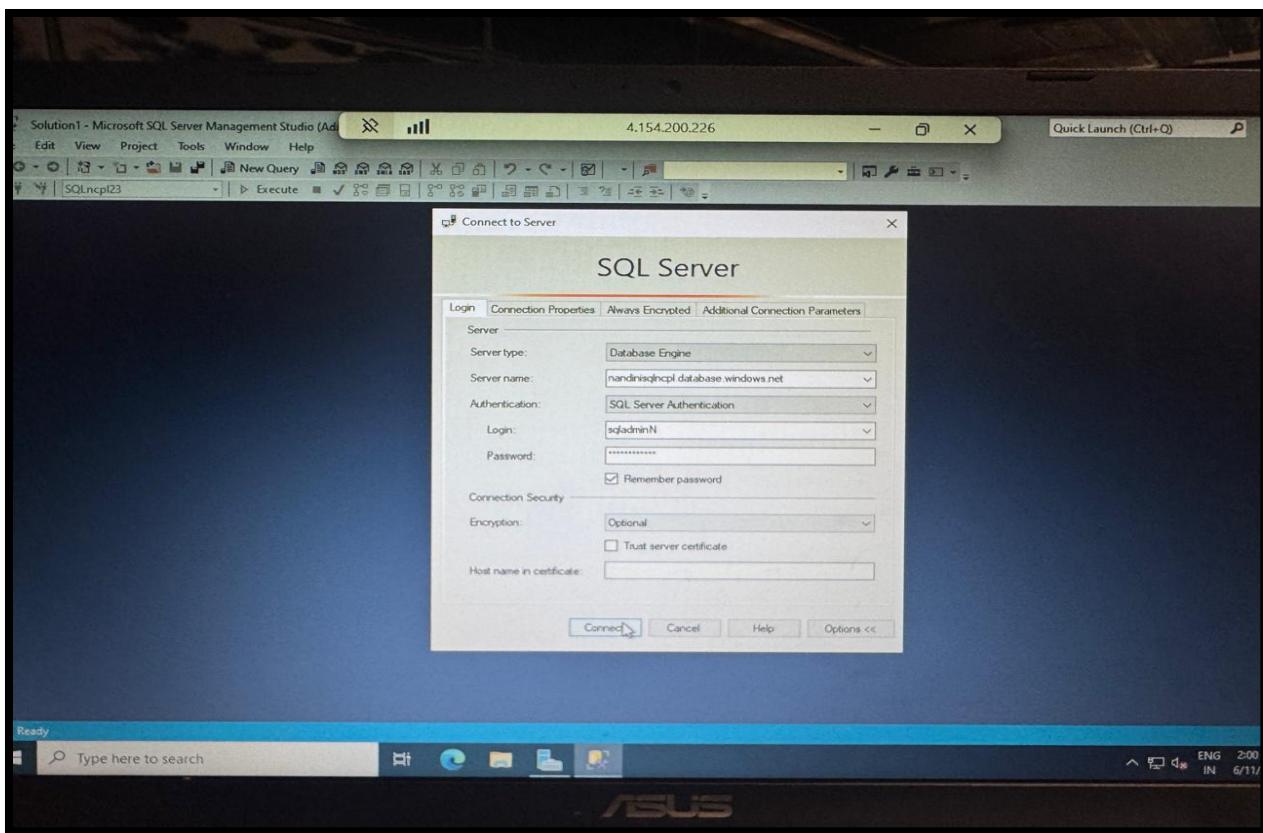
6. AFTER CONNECT WITH VM THIS SCREEN APPEARED ,SELECT LOCAL SERVER.



7. NOW SELECT 'OFF' FOR BOTH ADMINISTRATOR AND USERS.



## 7.DOWNLOAD SELF-HOSTED IR.



## 8. OPEN SSMS ,CONNECT IT WITH SQL SERVER.

```

ALTER TABLE [school].[metadata2] ADD targetlocalfilepath nvarchar(200);
UPDATE school.metadata2 SET targetlocalfilepath = 'c:\school\students2\students2.csv' WHERE tablename = 'students2';
UPDATE school.metadata2 SET targetlocalfilepath = 'c:\school\teachers2\teachers2.csv' WHERE tablename = 'teachers2';
UPDATE school.metadata2 SET targetlocalfilepath = 'c:\school\enrollment2\enrollment2.csv' WHERE tablename = 'enrollment2';
UPDATE school.metadata2 SET targetlocalfilepath = 'c:\school\attendance3\attendance3.csv' WHERE tablename = 'attendance3';
UPDATE school.metadata2 SET targetlocalfilepath = 'c:\school\fee\fee.csv' WHERE tablename = 'fee';

```

	id	tablename	schemaname	foldename	filename	TARGETFILEPATH	targetlocalfilepath
1	1	STUDENTS2	SCHOOL	SCHOOL\STUDENTS2	STUDENTS2.CSV	/RAW/SCHOOL/STUDENTS2.CSV	c:\school\students2\students2.csv
2	2	TEACHERS2	SCHOOL	SCHOOL\TEACHERS2	TEACHERS2.CSV	/RAW/SCHOOL/TEACHERS2.CSV	c:\school\teachers2\teachers2.csv
3	4	ENROLLMENT2	SCHOOL	SCHOOL\ENROLLMENT2	ENROLLMENT2.CSV	/RAW/SCHOOL/ENROLLMENT2.CSV	c:\school\enrollment2\enrollment2.csv
4	4	ATTENDANCE3	SCHOOL	SCHOOL\ATTENDANCE3	ATTENDANCE3.CSV	/RAW/SCHOOL/ATTENDANCE3.CSV	c:\school\attendance3\attendance3.csv
5	5	FEE	SCHOOL	SCHOOL\FEE	FEEL.CSV	/RAW/SCHOOL/FEE.CSV	c:\school\fee\fee.csv

## 9. CREATE A METADATA TABLE AND ALSO ADD FILE PATH FOR BOTH CLOUD AND ON-PREMISE .

**2. Load the extracted data into Azure Data Lake Storage Gen2 in a structured format.**

The screenshot shows the Microsoft Azure Synapse Analytics interface. The left sidebar has a tree view with 'Integration' expanded, showing 'Triggers', 'Integration runtimes' (which is selected and highlighted in grey), 'Security', 'Access control', and 'Credentials'. The main content area is titled 'Integration runtimes' and contains the following information:

- Description: The integration runtime (IR) is the compute infrastructure to provide the following data integration capabilities across different network environment.
- Buttons: '+ New' and 'Refresh'.
- Filter: 'Filter by name'.
- Table header: Name, Type, Sub-type, Status, Related, Region, Version.
- Table data (5 items):
 

Name	Type	Sub-type	Status	Related	Region	Version
AutoResolveIntegrationRuntime	Azure	Public	Running	11	Auto Resolve	---
IntegrationRuntime2310	Self-Hosted	---	Unavail...	More 1	---	5.50.9181.2
IntegrationRuntime2311	Self-Hosted	---	Unavail...	More 0	---	---
SELFHOSTED-IR	Self-Hosted	---	Unavail...	More 0	---	5.50.9181.2
selfhostedintegrationruntime0506	Self-Hosted	---	Unavail...	More 3	---	5.50.9181.2

**9. NOW CREATE INTEGRATION TUNTIME.**

**10.SELECT**

**SELF**

**HOSTED**

The screenshot shows the 'Integration runtime setup' page. The left sidebar is identical to the previous screenshot. The main content area has two large callout boxes:

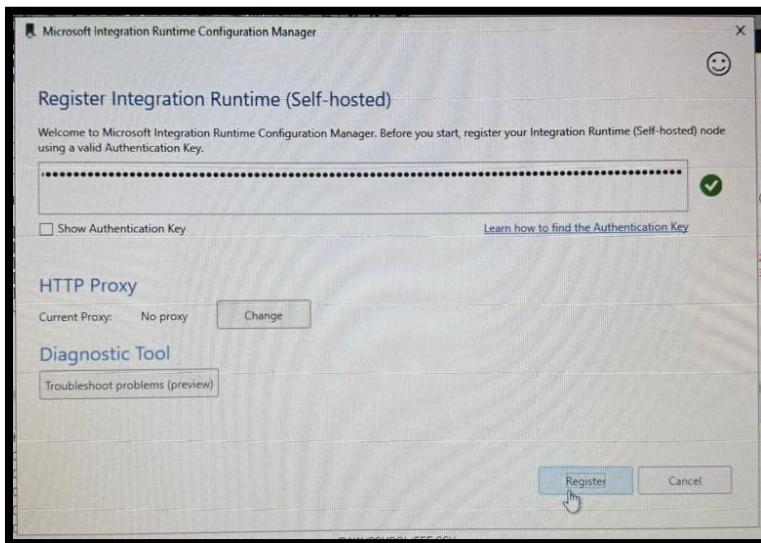
- SELF**: 'Azure, Self-Hosted' icon. Description: Perform data flows, data movement and dispatch activities to external compute.
- HOSTED**: 'Azure-SSIS' icon. Description: Lift-and-shift existing SSIS packages to execute in Azure.

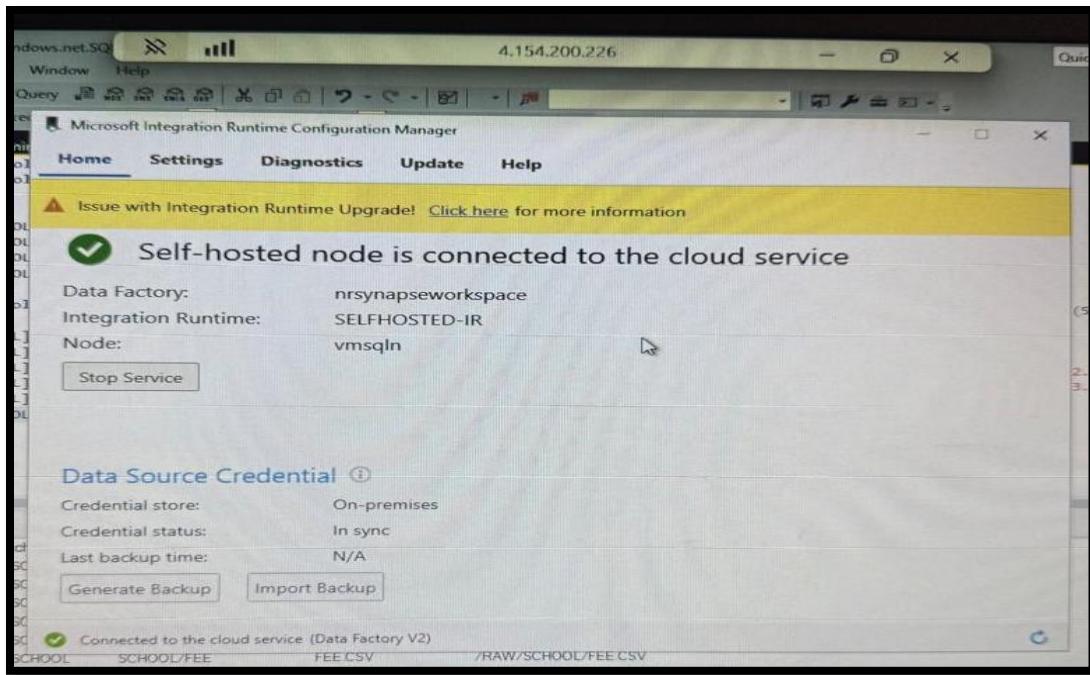
At the bottom right are 'Continue' and 'Cancel' buttons.

The screenshot shows the Microsoft Azure Synapse Analytics Integration runtime setup page. On the left, there's a sidebar with options like Synapse live, Connector upgrade advisor, Analytics pools, SQL pools, Apache Spark pools, Data Explorer pools, External connections, Linked services, Microsoft Purview, Integration, Triggers, and Integration runtimes. The Integration runtimes option is selected. The main area is titled "Integration runtime setup" and "Network environment". It describes the integration runtime as the compute infrastructure for data flows, movement, or dispatch activities. It offers two choices: "Azure" (for running data flows, external and pipeline activities in a fully managed, serverless compute) and "Self-Hosted" (for running data movement, external and pipeline activities in an on-premises or private network). The "Self-Hosted" section includes a note that data flows are only supported on Azure integration runtime and a link to view less information. At the bottom, there are "Continue", "Back", and "Cancel" buttons.

This screenshot shows the same Microsoft Azure Synapse Analytics Integration runtime setup page, but it's in the process of creating a new integration runtime. The "Name" field is filled with "shis", and the "Type" is set to "Self-Hosted". A progress bar at the bottom indicates "Creating...". The rest of the interface is identical to the previous screenshot, showing the list of existing runtimes and the "Self-Hosted" configuration details.

## 11. COPY THIS KEY1 URL AND PASTE TO ON PREMISE FOR CONNECTION OF SH-IR.



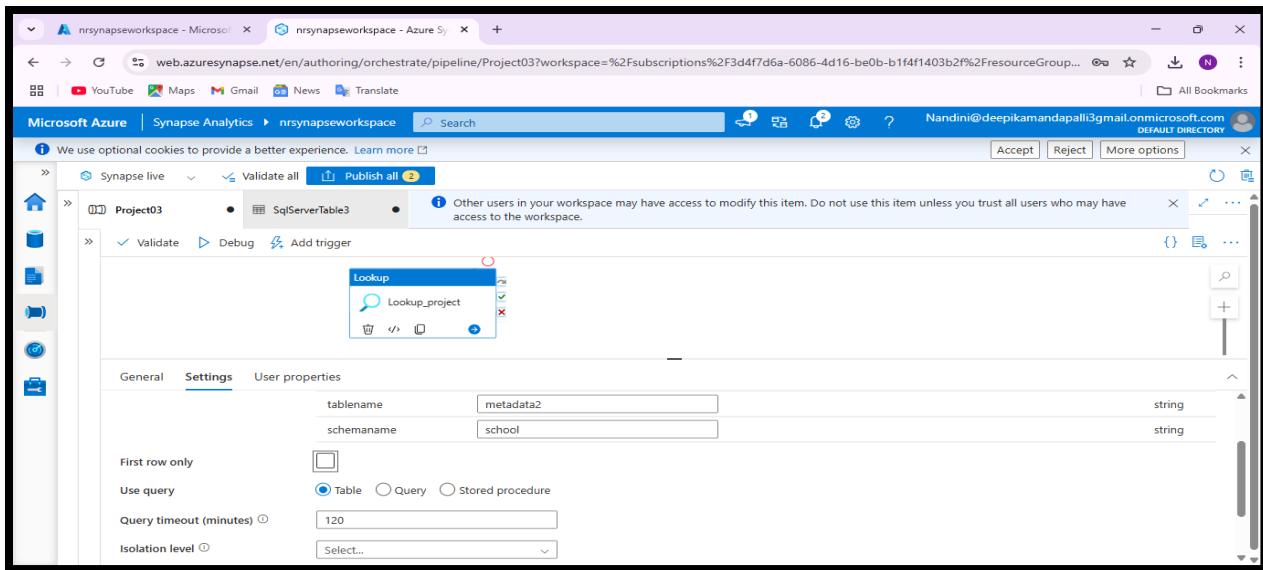


The screenshot shows the Microsoft Azure Synapse Analytics portal. The left sidebar navigation includes options like Synapse live, Validate all, Publish all, Connector upgrade advis..., Analytics pools, SQL pools, Apache Spark pools, Data Explorer pools (prev...), External connections, Linked services, Microsoft Purview, Integration, Triggers, Integration runtimes (which is currently selected), Security, Access control, and Credentials.

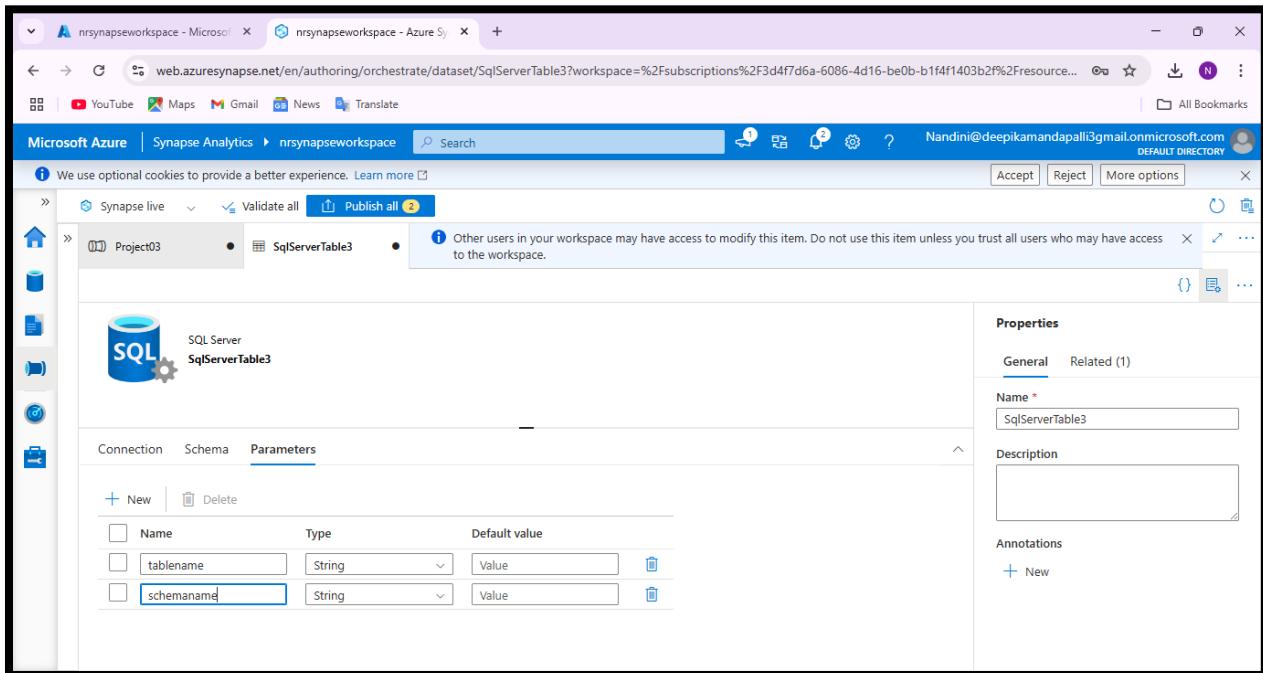
The main content area is titled "Integration runtimes" and describes them as the compute infrastructure for data integration. It includes a "New" button and a "Refresh" button. A "Filter by name" input field is available. The table below lists six items:

Name	Type	Sub-type	Status	Related	Region	Version
AutoResolveIntegrationRuntime	Azure	Public	Running	11	Auto Resolve	---
IntegrationRuntime2310	Self-Hosted	---	Unavail...	1	---	5.50.9181.2
IntegrationRuntime2311	Self-Hosted	---	Unavail...	0	---	---
SELFHOSTED-IR	Self-Hosted	---	Unavail...	0	---	5.50.9181.2
selfhostedintegrationruntime0506	Self-Hosted	---	Unavail...	3	---	5.50.9181.2
shis	Self-Hosted	---	Running	0	---	5.50.9181.2

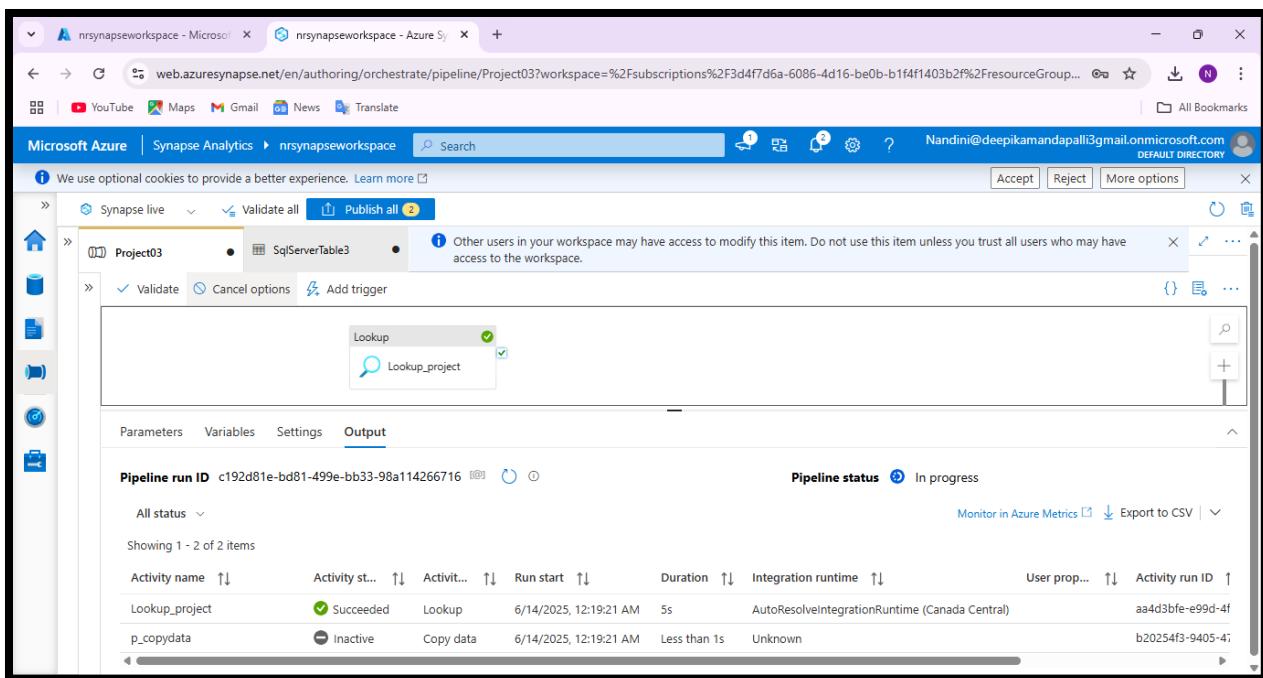
## 12. SELF HOSTED INTEGRATION RUNTIME IS RUNNING.



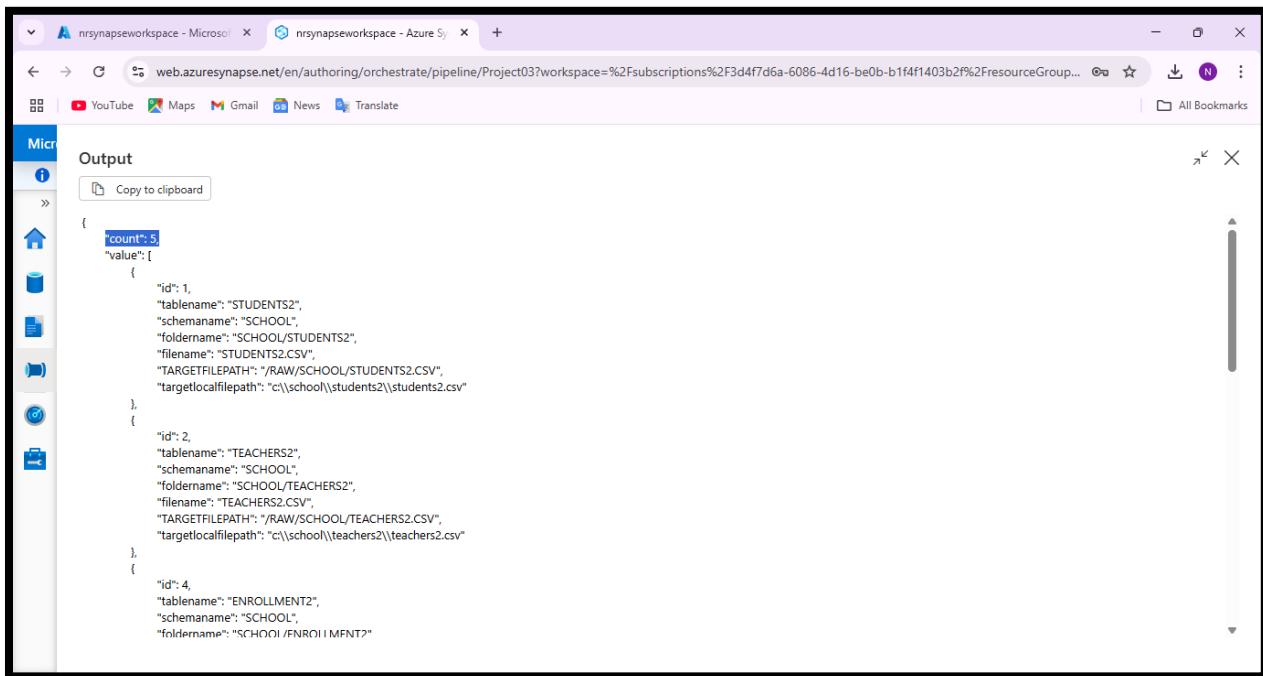
13. NOW GO TO DATA ON 'INTEGRATE' ON LEFT SIDE OF PAGE ,CREATE A PIPELINE,NAME IT AND DRAG A LOOKUP ACTIVITY TO LOOKUP FOR TABLES PRESENT IN METADATA TABLE IN SQL SERVER.



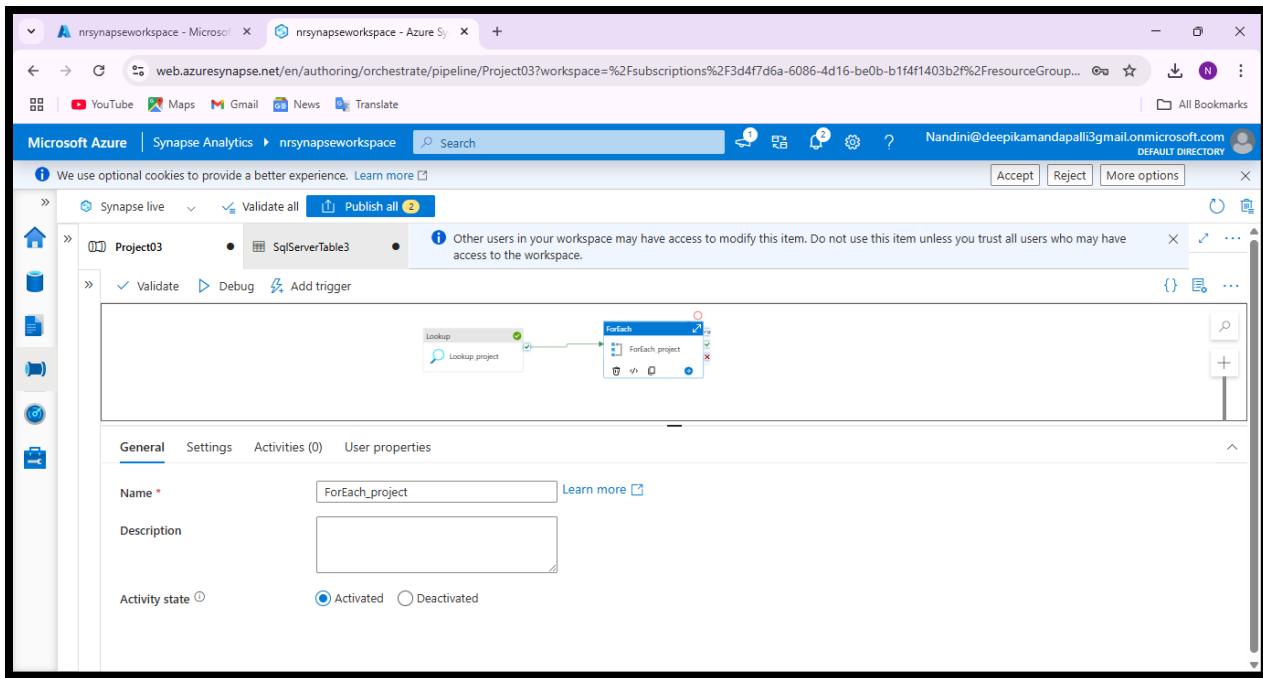
14. CREATE PARAMETERS.



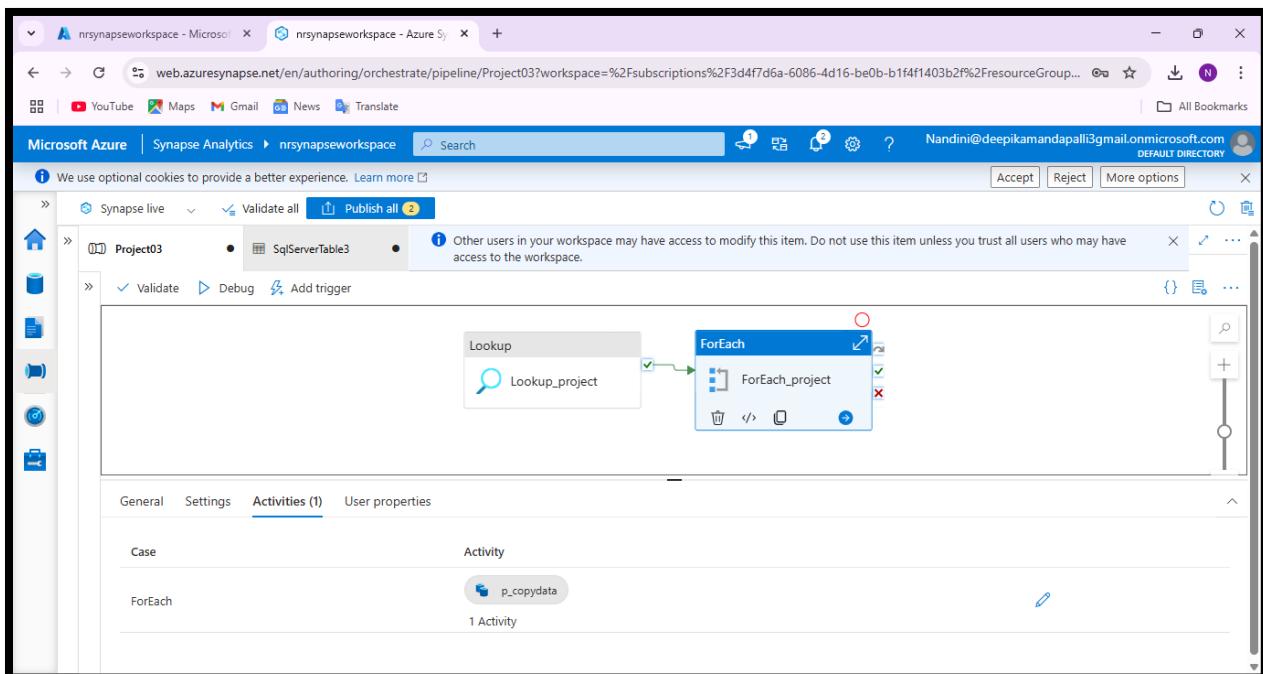
## 15. THIS ACTIVITY IS SUCCEDED.

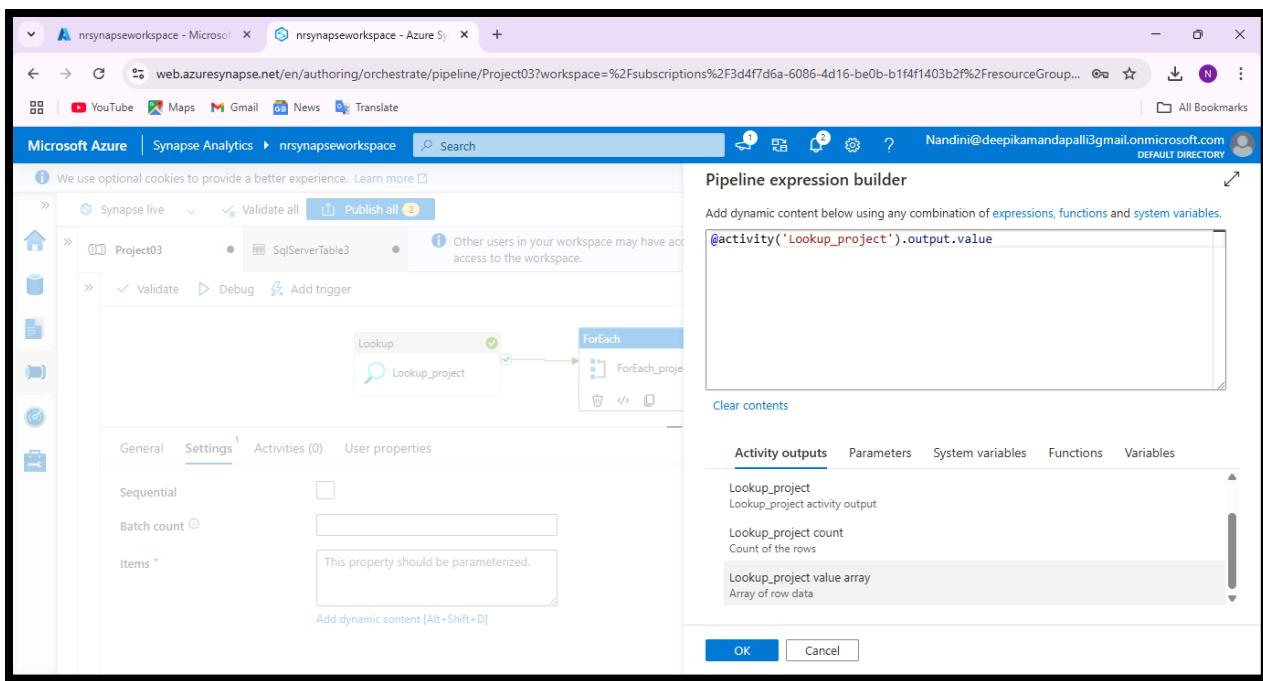


## 14. THIS SHOWS NO. OF TABLES IN METADATA TABLE,COUNT=5.

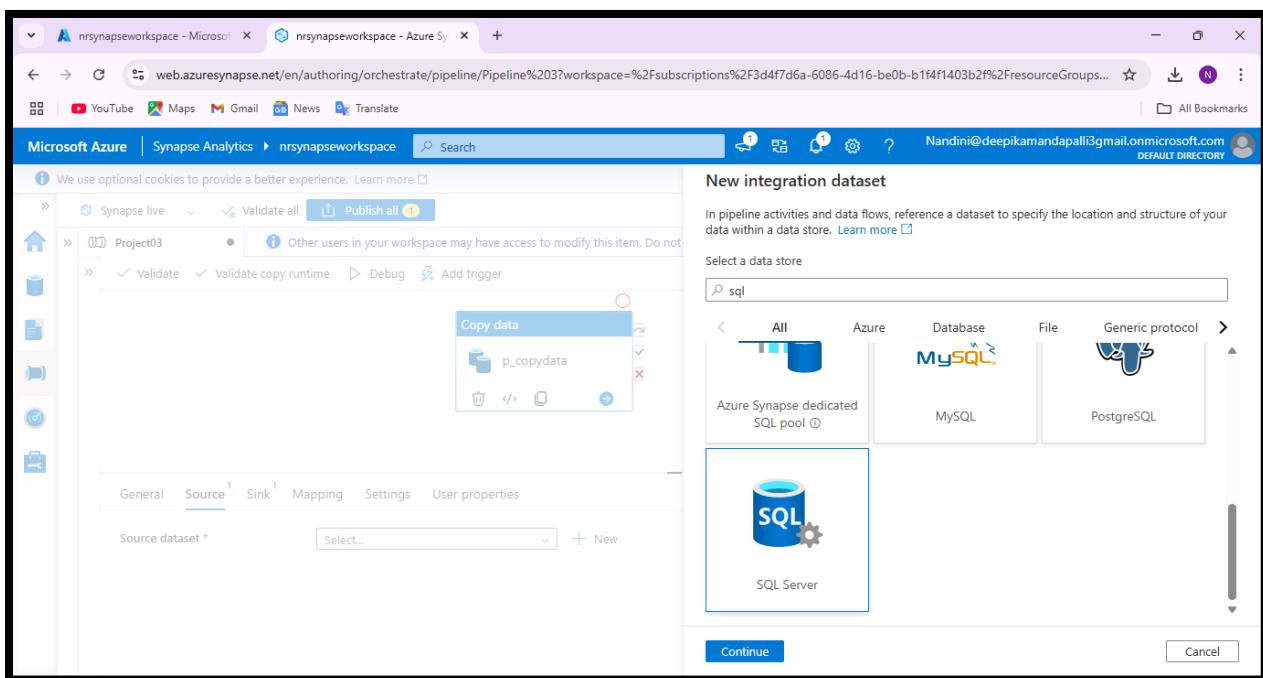


16. AFTER THAT CREATE FOR EACH LOOP ,NAME IT AND CONNECT IT WITH LOOKUP DATA SO THAT IT WILL ITERATE THE ARRAYS OF METADATA TABLES OR OUTPUT OF LOOKUP ACTIVITY.

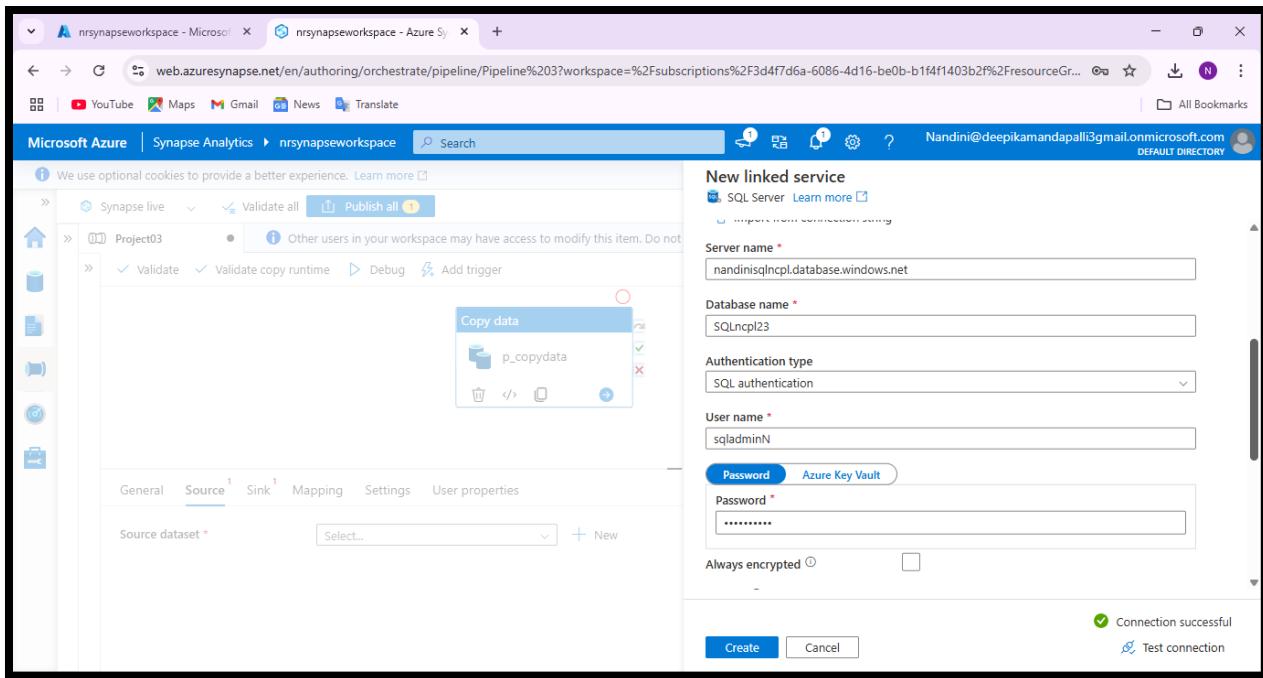




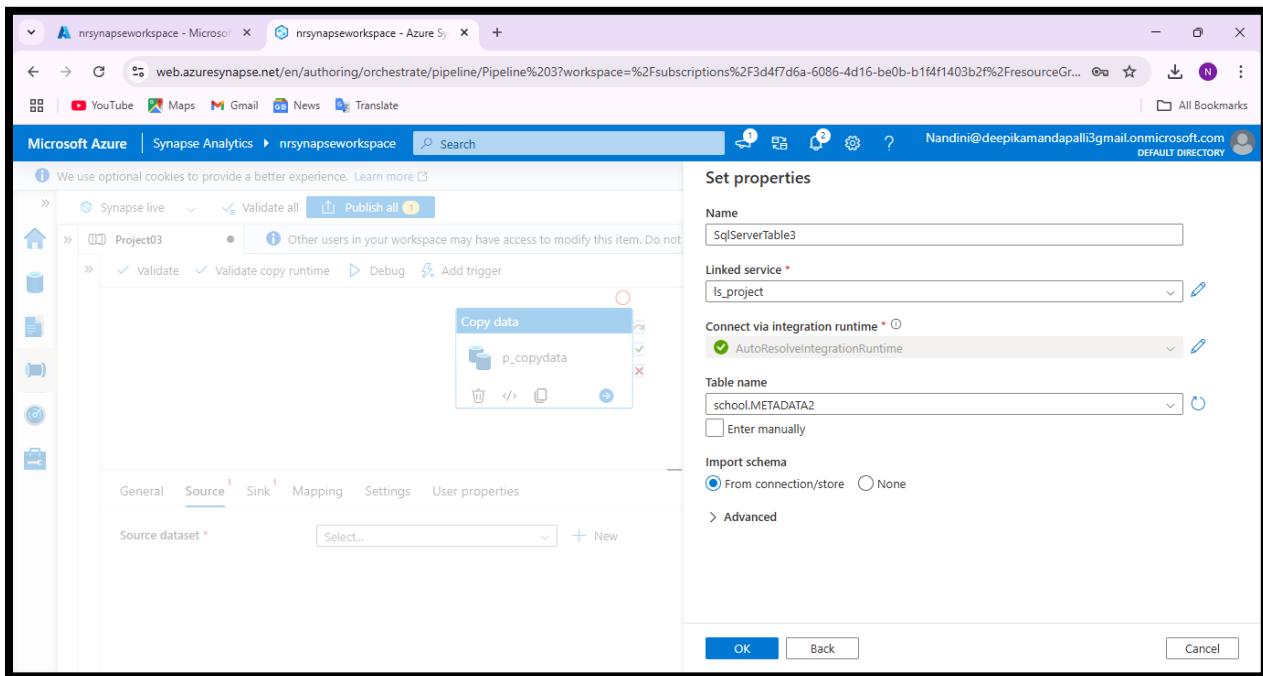
17. INSIDE FOREACH ACTIVITY CREATE COPY DATA ACTIVITY, HERE WE USE @ACTIVITY , WHICH WILL COPY OUTPUT OF LOOKUP ACTIVITY.



18. SELECT 'SQL SERVER' AS SOURCE DATASET.



## 19. CREATE NEW LINKED SERVICE.

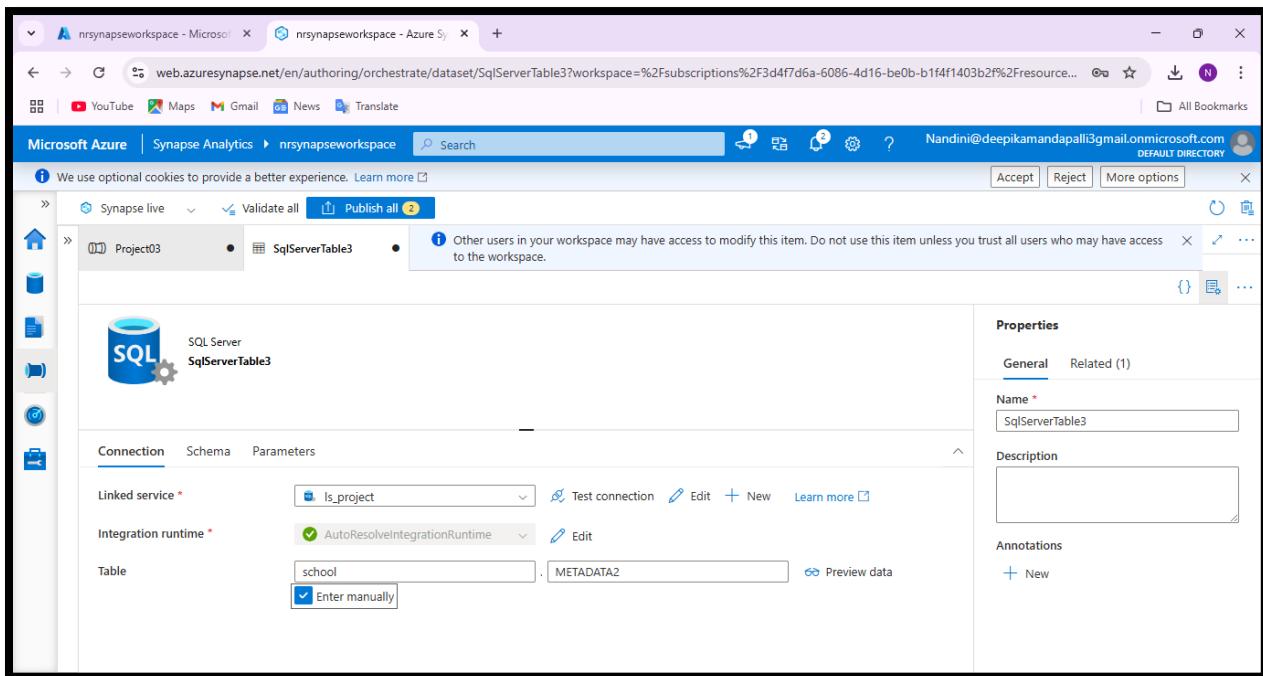


## 20. SET PROPERTIES AND SELECT TABLE NAME.'SCHOOL.METADATA2'

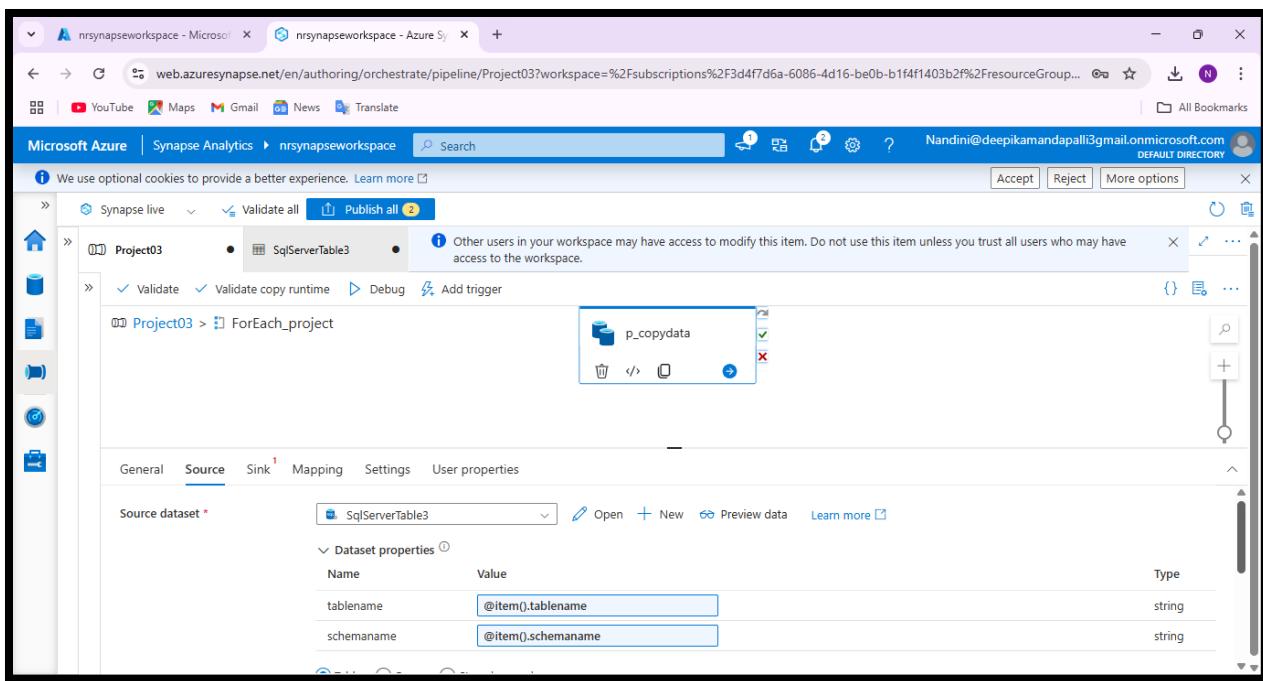
4. *Implement a metadata-driven approach using a control table that stores details such as:*

1. *Source table names*

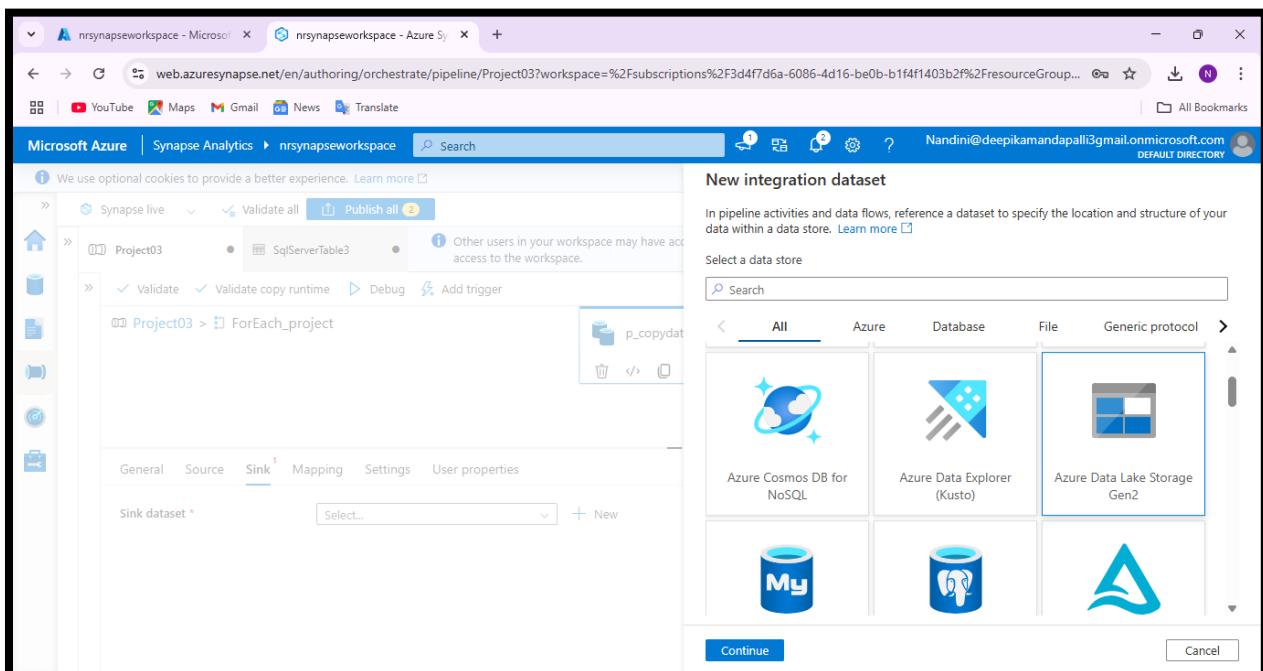
2. *Target file paths*



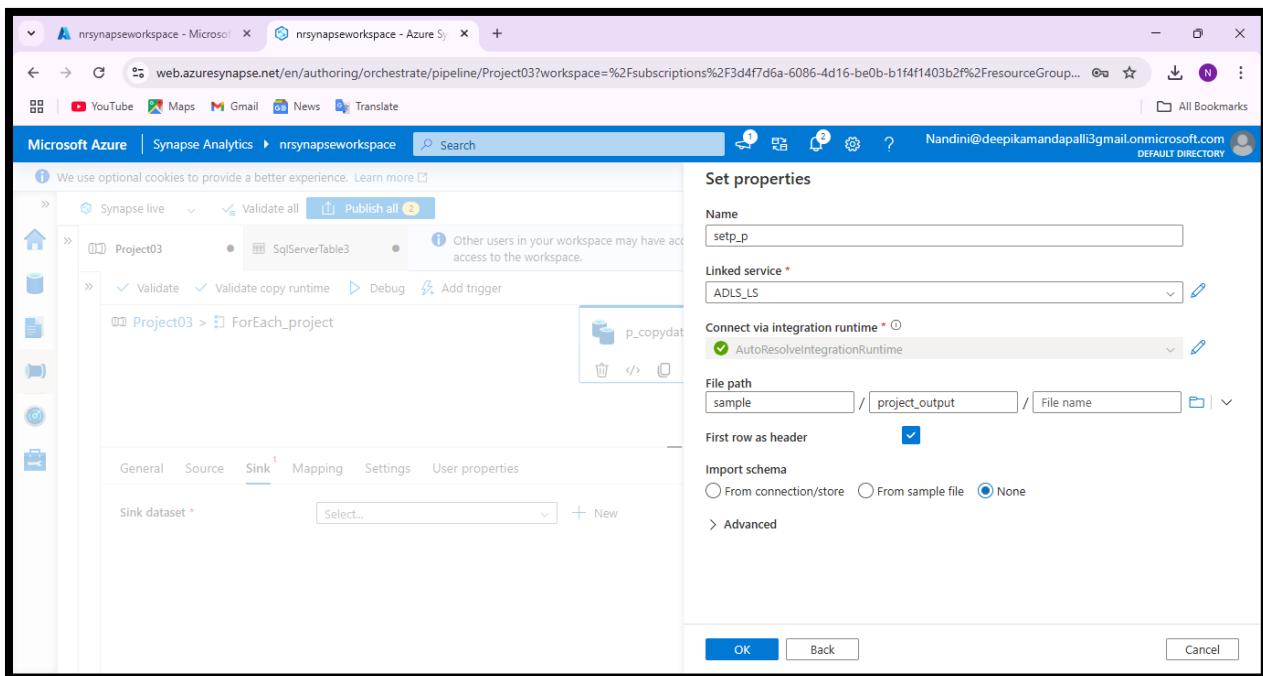
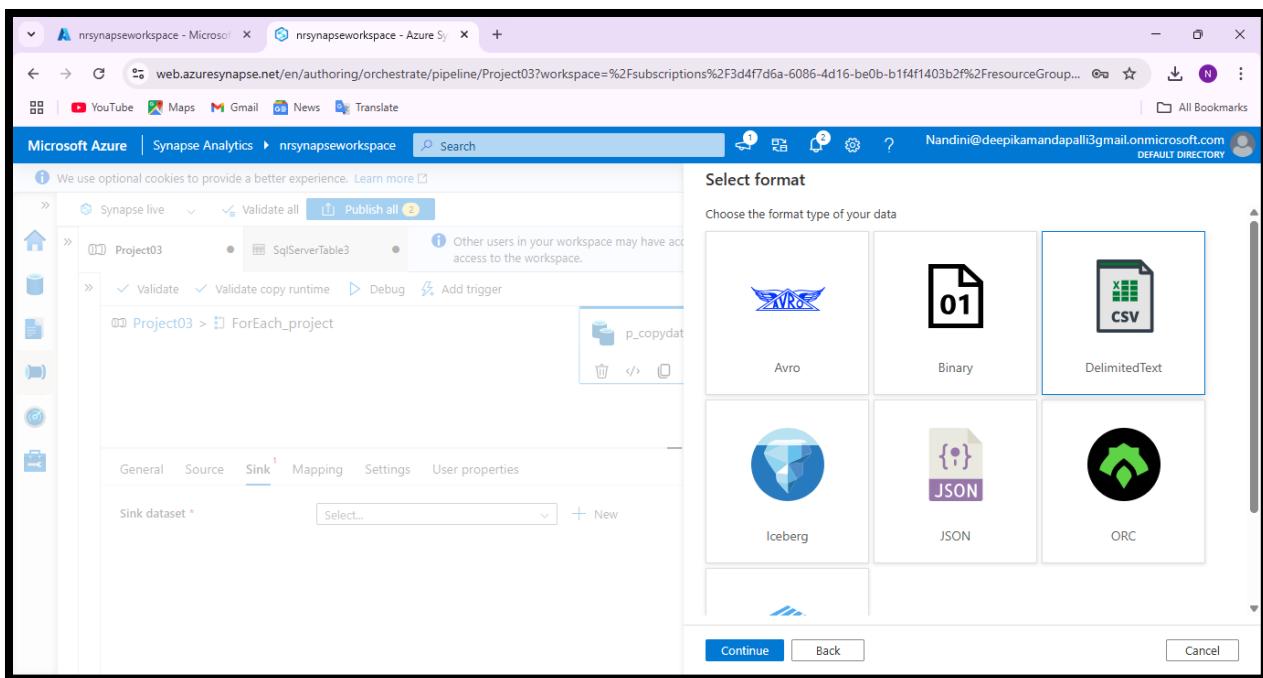
21. CREATE PARAMETERS SCHEMANAME AND TABLENAME.



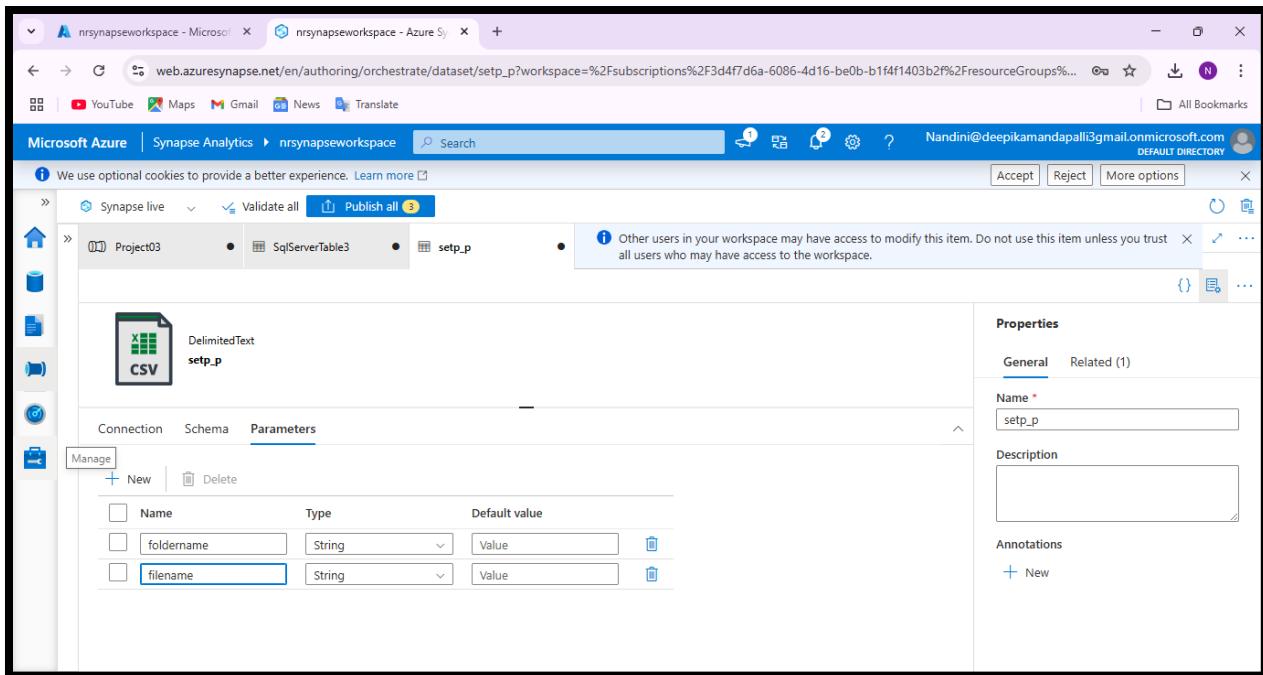
22. USE @ITEM(). FUNCTION ,THIS WILL ITERATE EVERY ITEM OF THE TABLE.



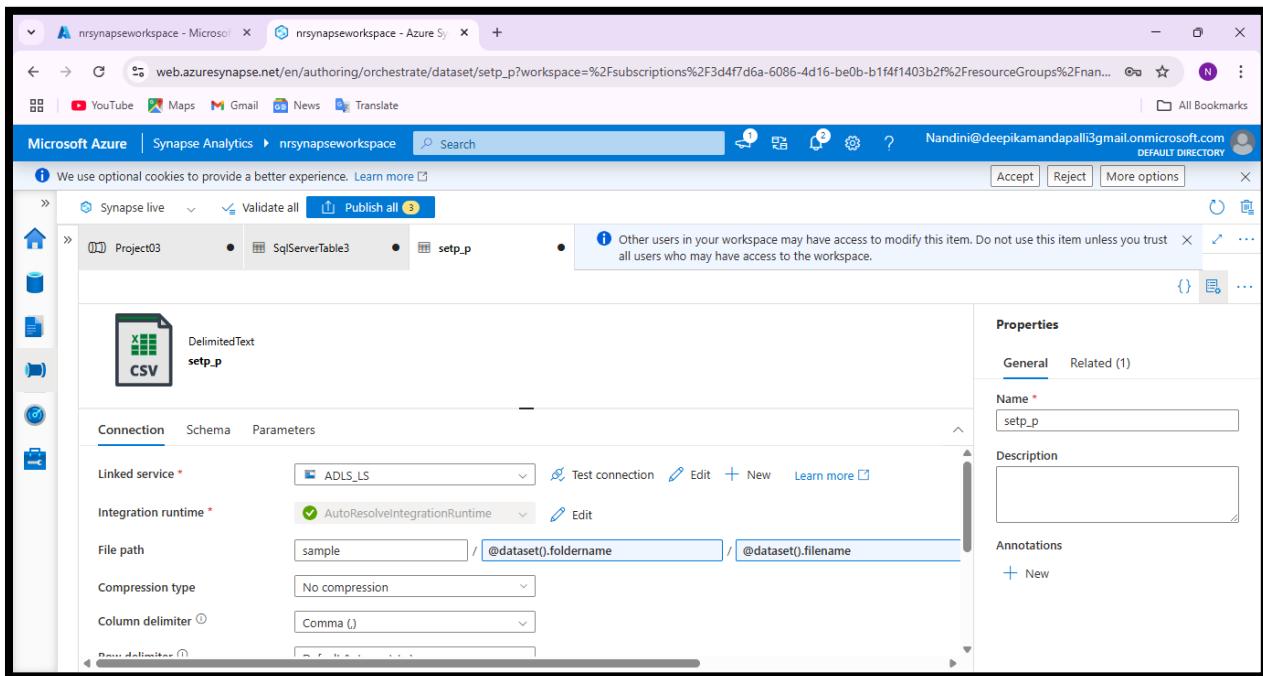
23. NOW FOR SINK SELECT ADLSGEN2,THEN CSV FORMAT.

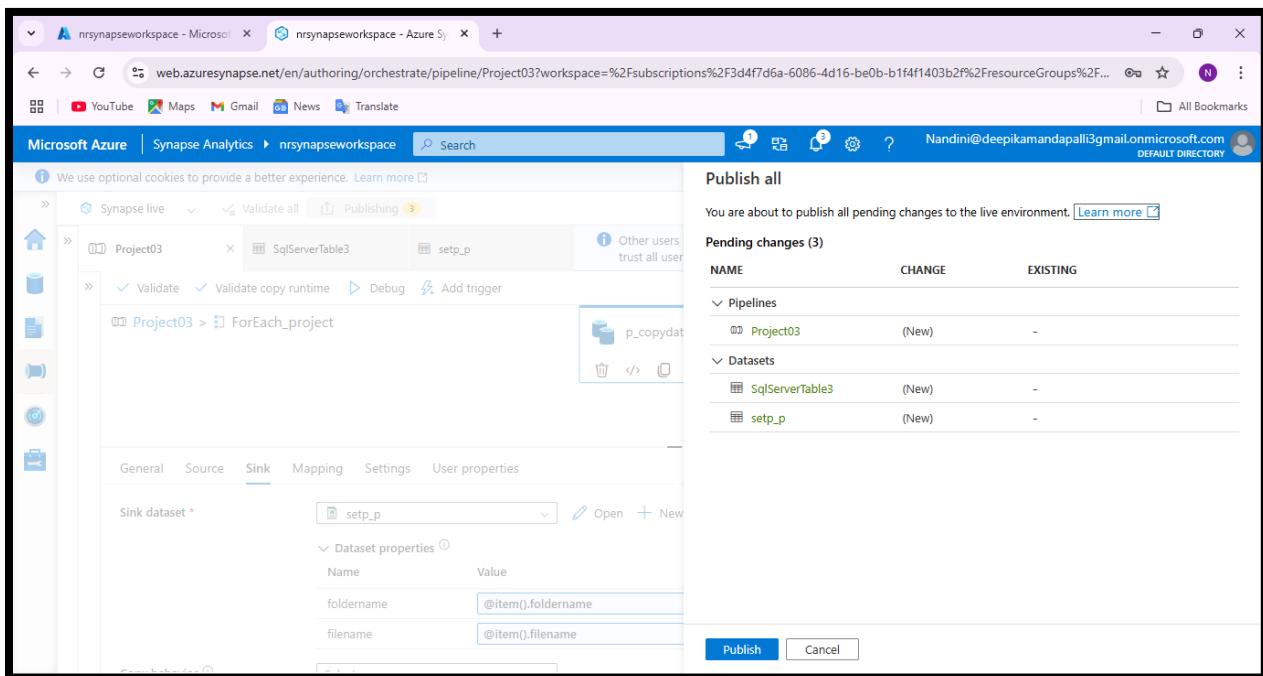
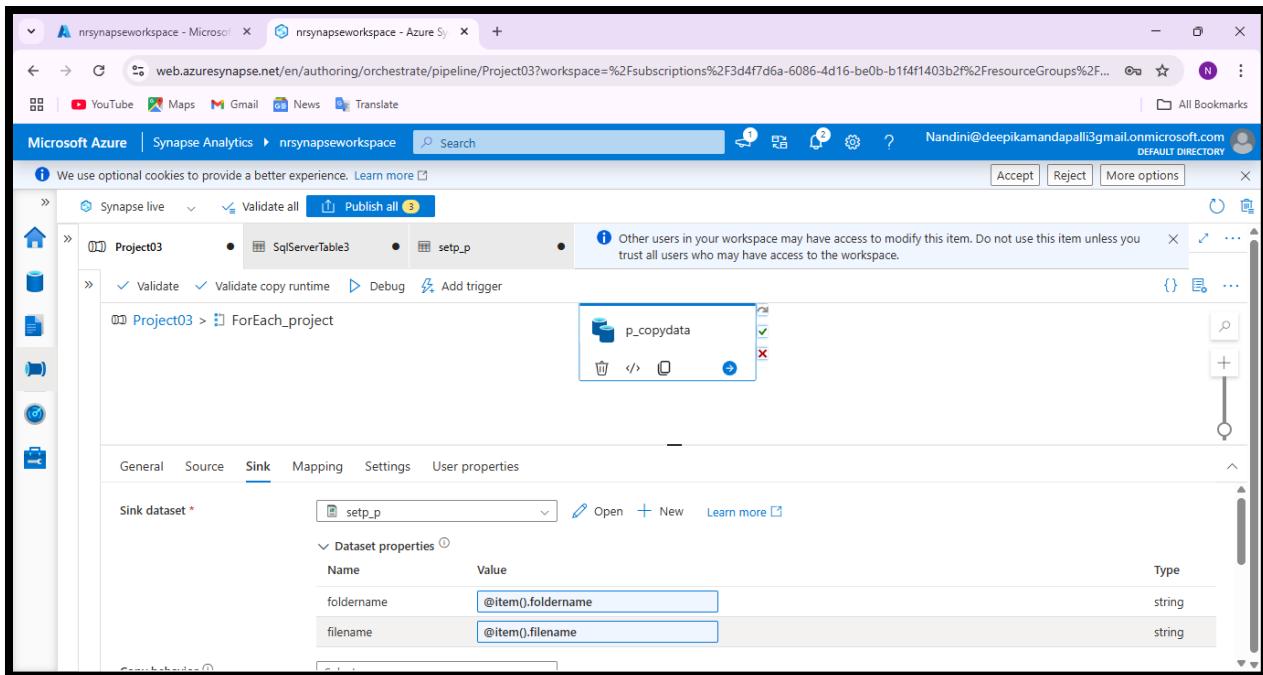


24. SET PROPERTIES AND SELECT FILE PATH WHERE WE WANT TO COPY DATA IN ADLSGEN2 ACCOUNT.

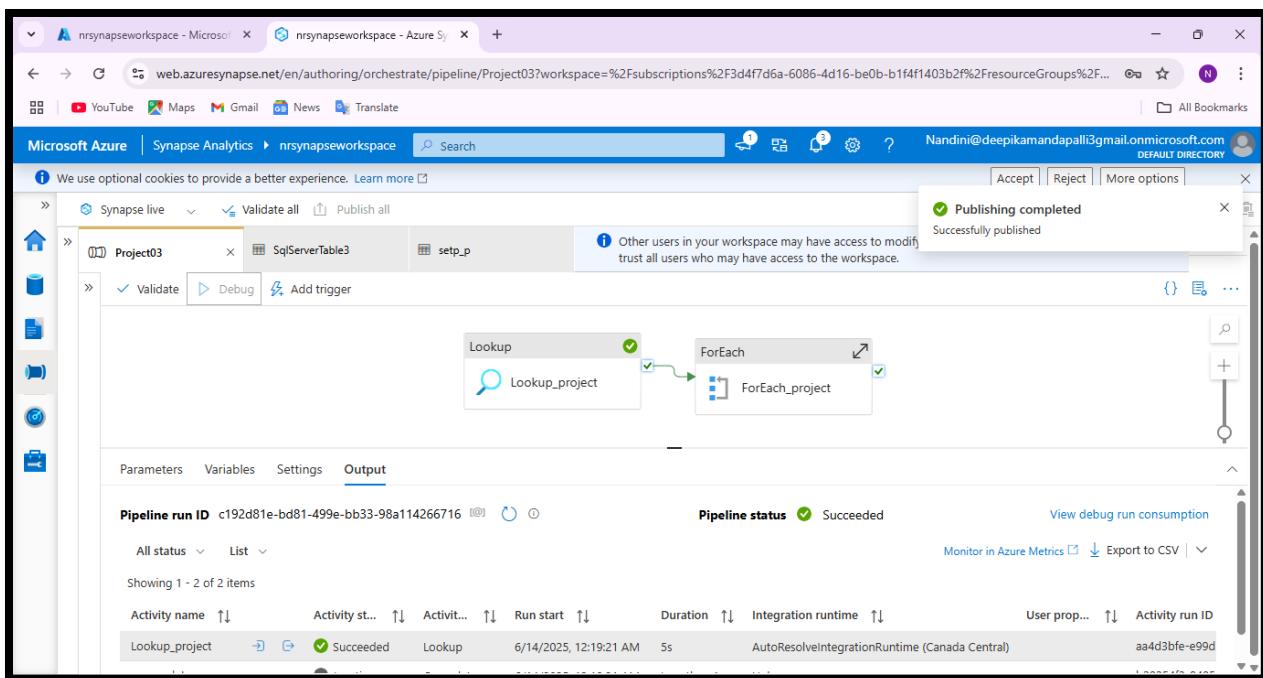


## 25.CREATE PARAMETERS FOR ADLSGEN2→FOLDERNAME,FILENAME.

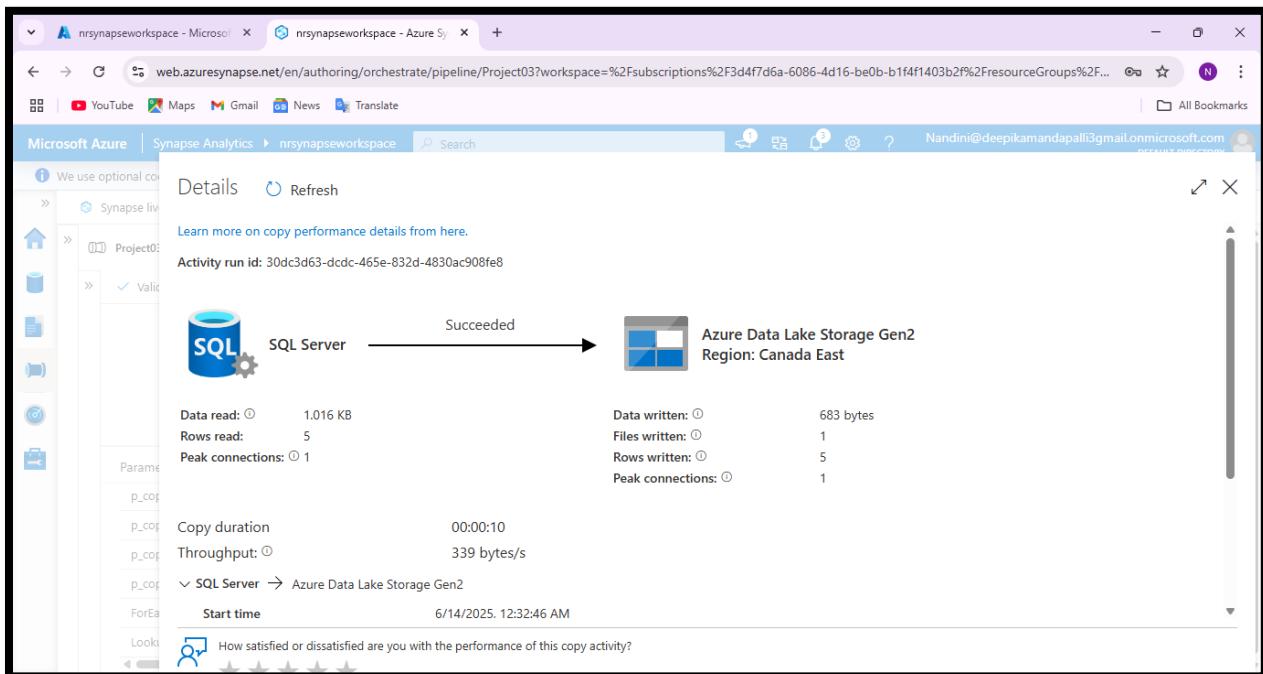




## 26.PUBLISH PIPELINE.



27.PIPELINE SUCCEEDED.



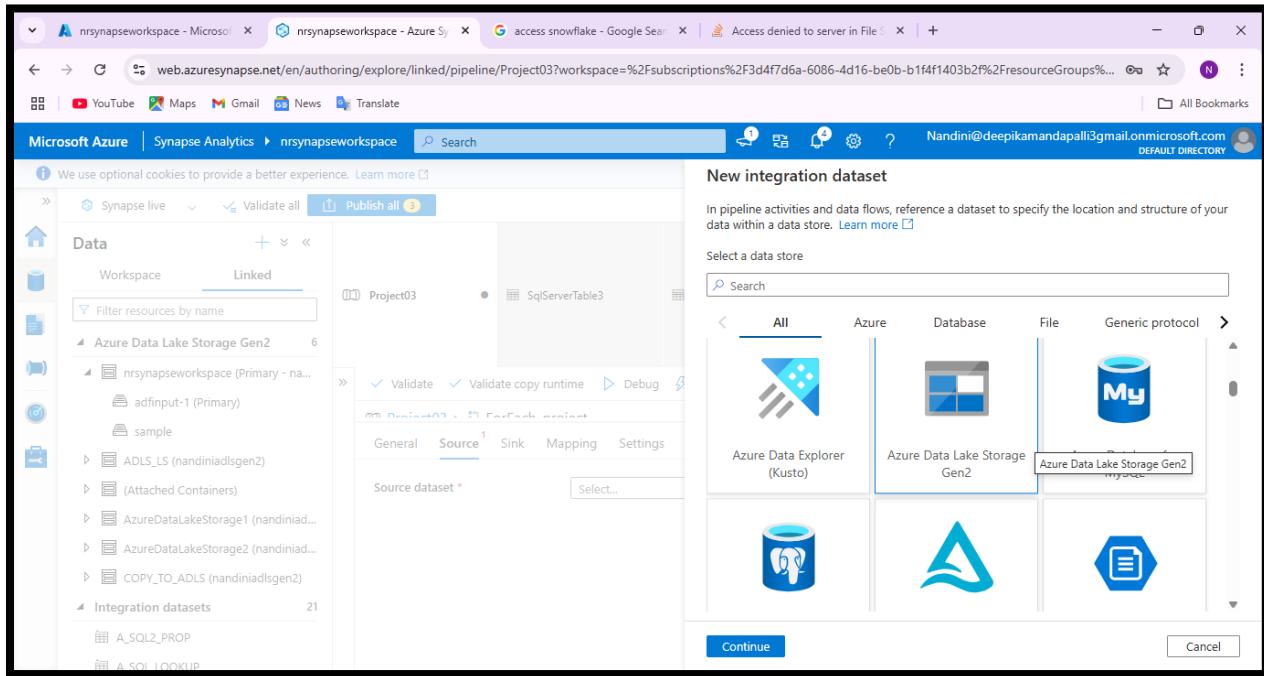
28. DATA SUCCESSFULLY LOADED FROM SQL SERVER TO ADLSGEN2.

Activity name	Activity st...	Activit...	Run start	Duration	Integration runtime	User prop...	Activity run ID
p_copydata	Succeeded	Copy data	6/14/2025, 12:32:44 AM	15s	AutoResolveIntegrationRuntime (Canada East)		599a1de4-120c
p_copydata	Succeeded	Copy data	6/14/2025, 12:32:43 AM	15s	AutoResolveIntegrationRuntime (Canada East)		30dc3d63-dcdc
p_copydata	Succeeded	Copy data	6/14/2025, 12:32:43 AM	15s	AutoResolveIntegrationRuntime (Canada East)		a1c905e0-1c0b
p_copydata	Succeeded	Copy data	6/14/2025, 12:32:43 AM	15s	AutoResolveIntegrationRuntime (Canada East)		1453add3-14f2
p_copydata	Succeeded	Copy data	6/14/2025, 12:32:43 AM	21s	AutoResolveIntegrationRuntime (Canada East)		f8c72699-7edb
ForEach_project	Succeeded	ForEach	6/14/2025, 12:32:43 AM	24s			2ad4c0d4-7aa3
Lookup_project	Succeeded	Lookup	6/14/2025, 12:32:46 AM	6s	AutoResolveIntegrationRuntime (Canada Central)		3dfb8257-64b1

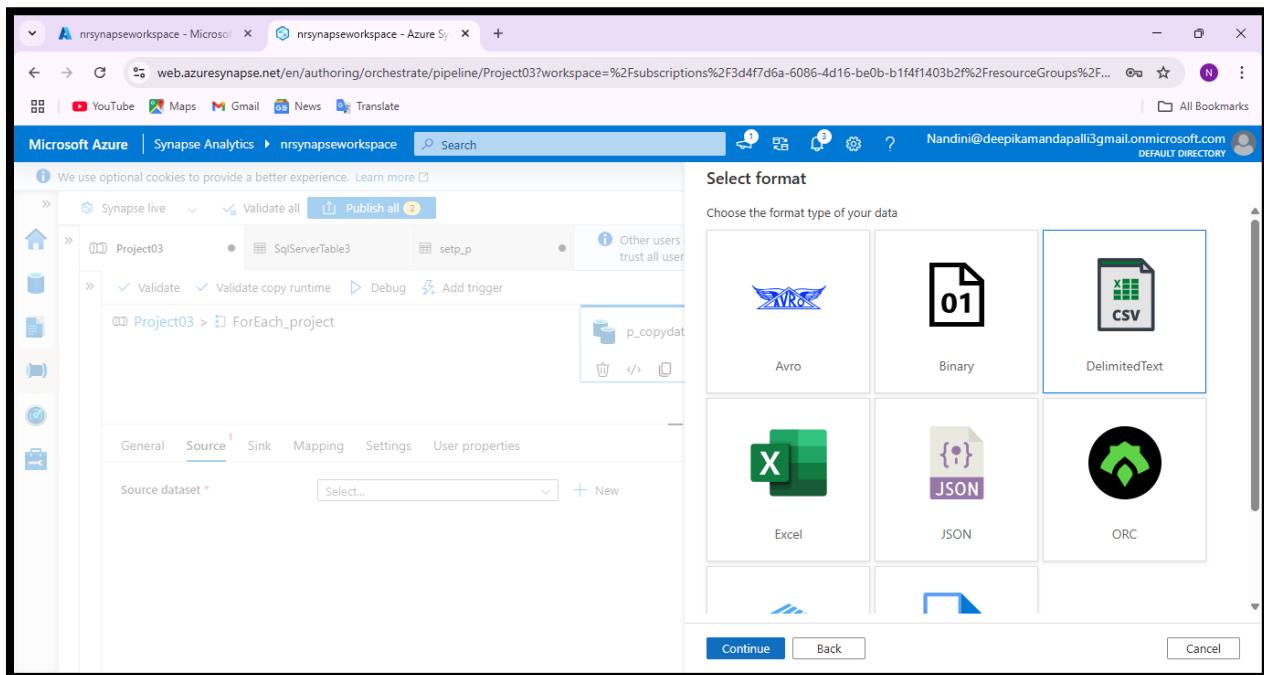
Name	Last Modified	Content Type	Size
ATTENDANCE3	14/06/2025, 00:33:01	Folder	
ENROLLMENT2	14/06/2025, 00:32:55	Folder	
FEE	14/06/2025, 00:32:56	Folder	
STUDENTS2	14/06/2025, 00:32:56	Folder	
TEACHERS2	14/06/2025, 00:32:56	Folder	

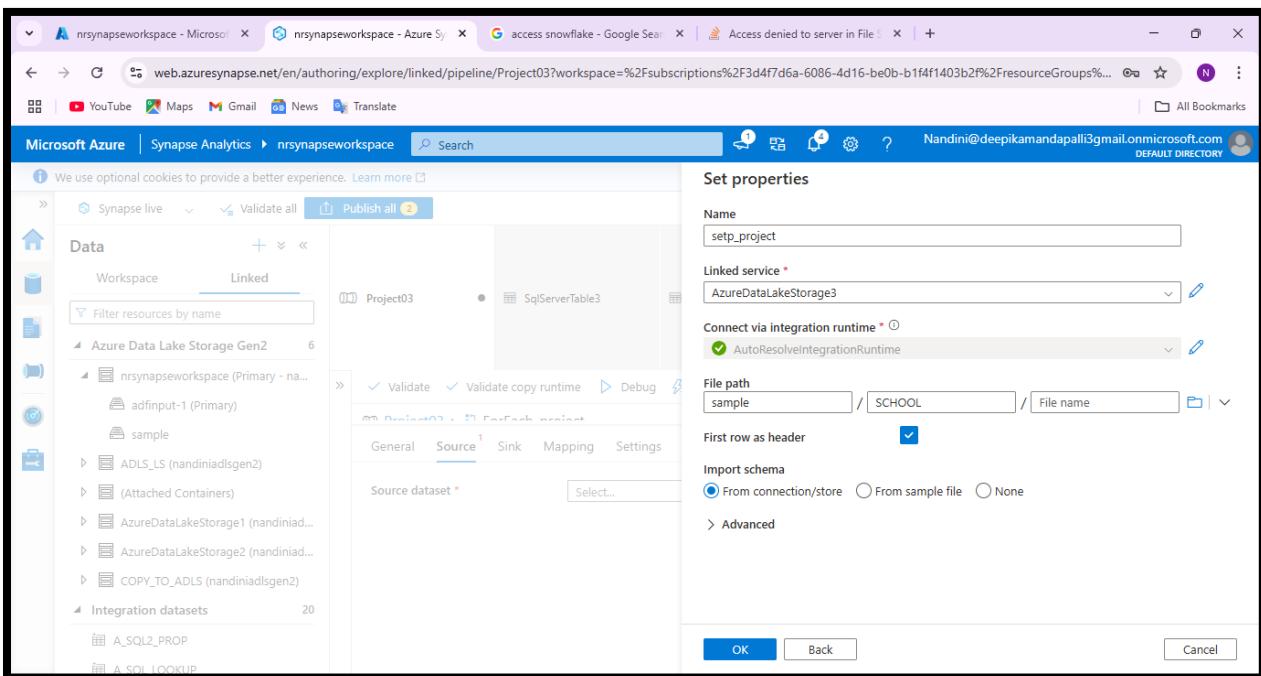
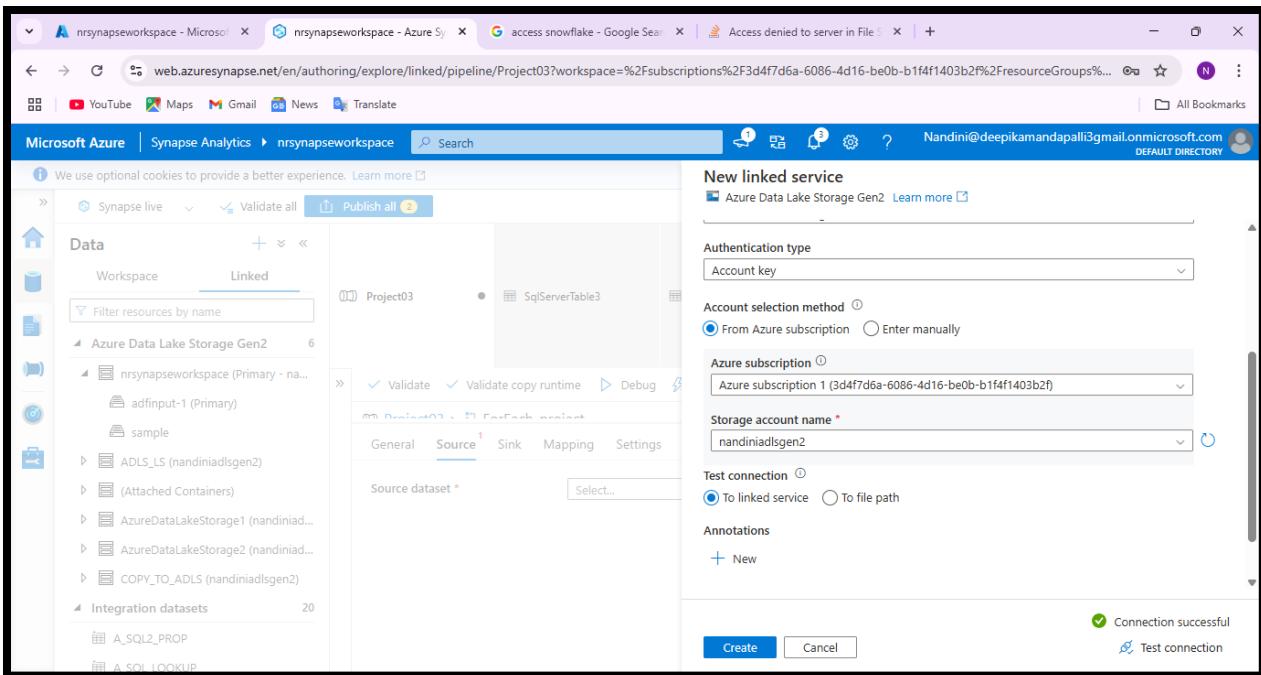
## 29. LOADED DATA PRESENT IN CLOUD.

**3. Copy the data from ADLS Gen2 back to a designated file system folder on the local C: drive.**

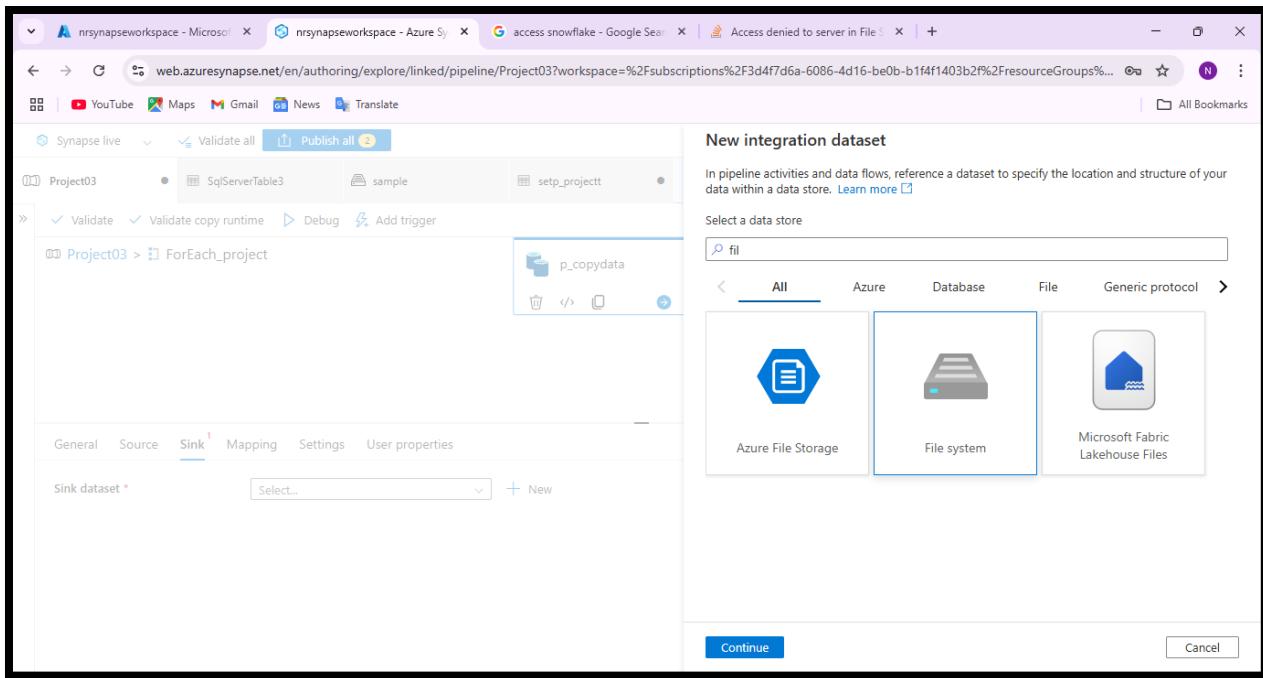


30. NOW FOR LOADING OF DATA FROM CLOUD TO SQL SERVER, USE COPY ACTIVITY IN THE SAME PIPELINE HERE FOR SOURCE DATASET → ADLSGEN2, SINK → FILE SERVER.

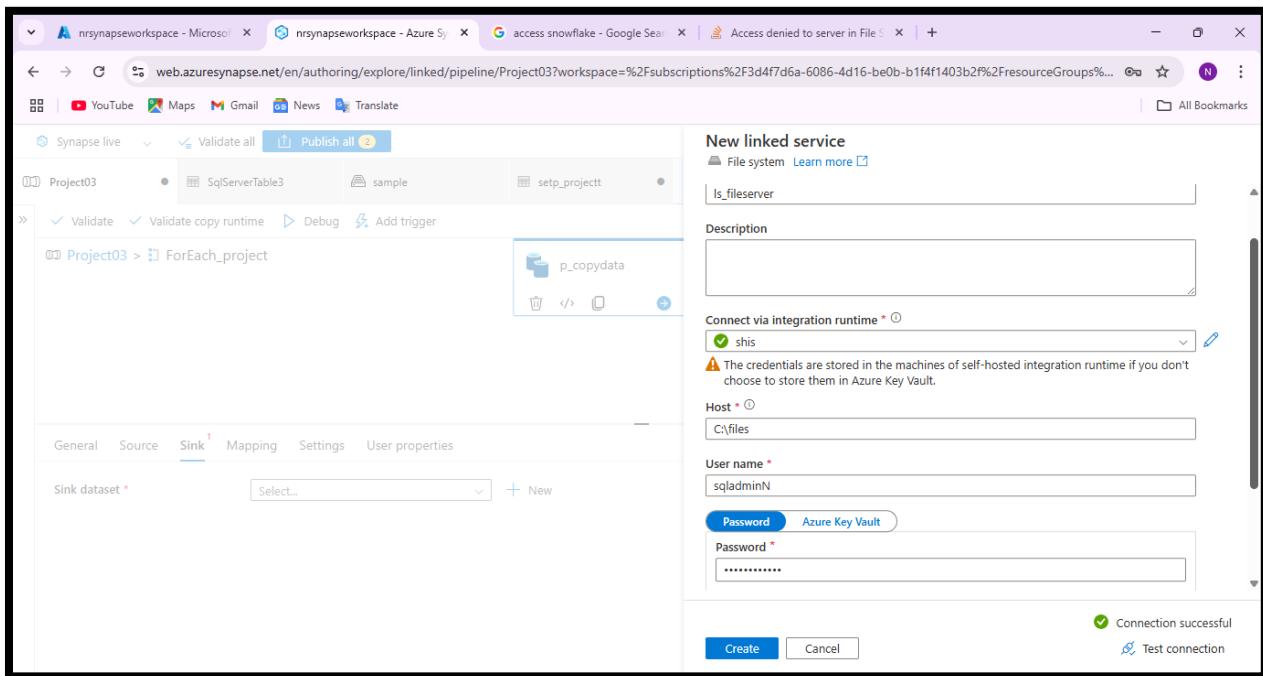




31.SELECT FOLDER WHICH YOU WANT TO LOAD ON-PREMISE.

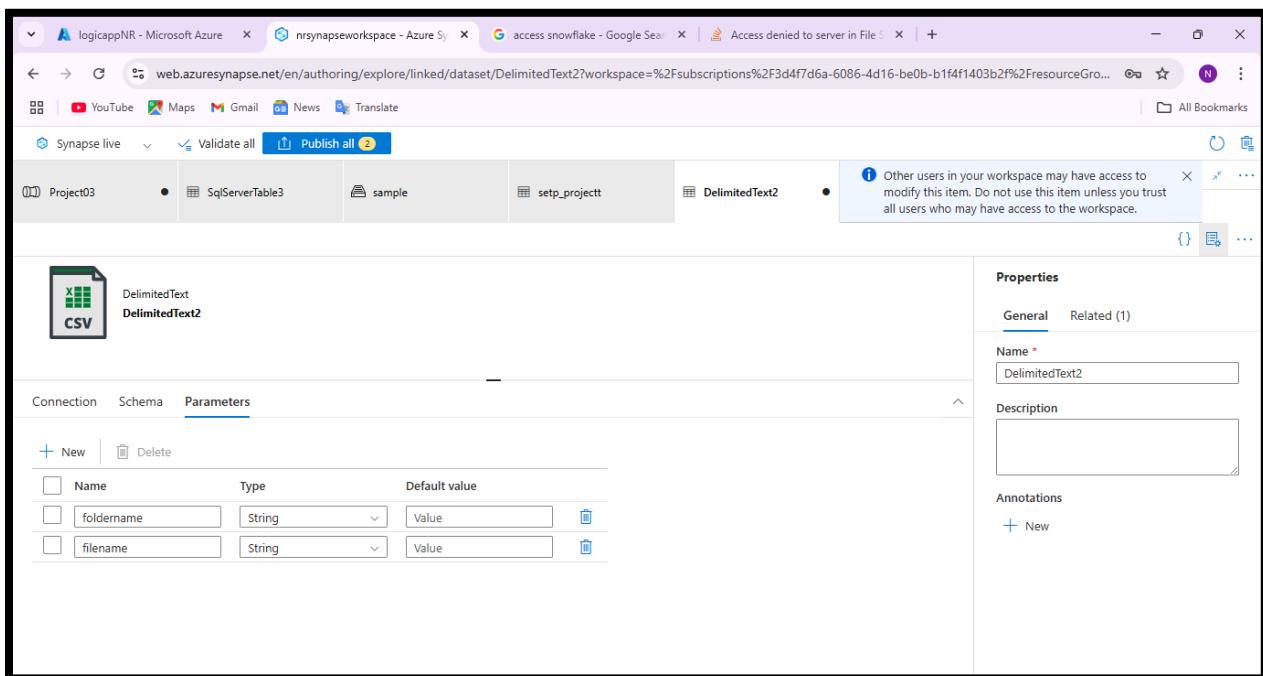
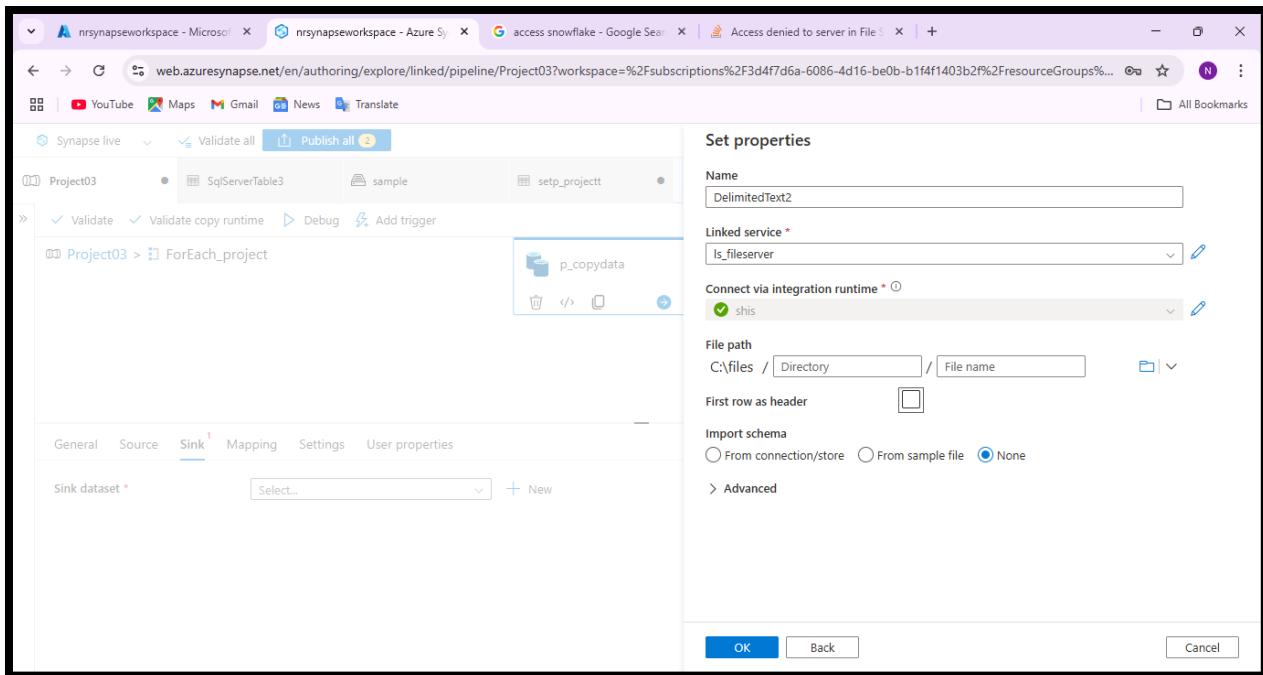


### 33.FOR SINK→FILE SYSTEM.

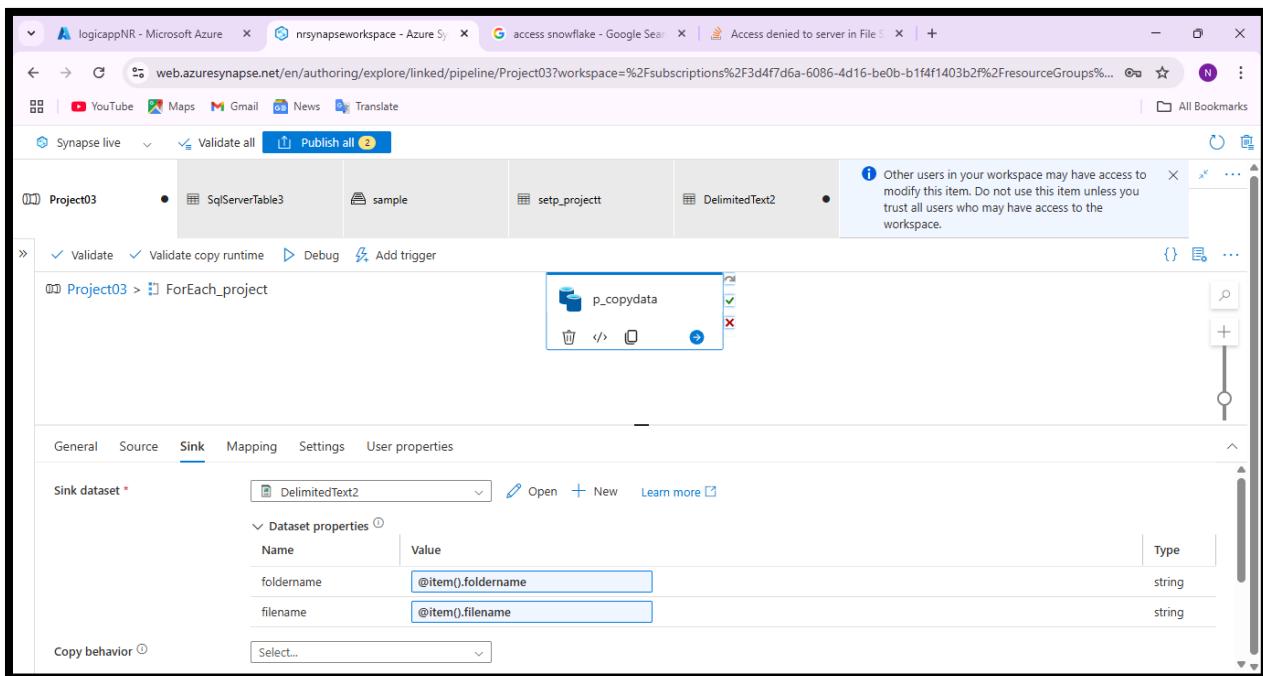
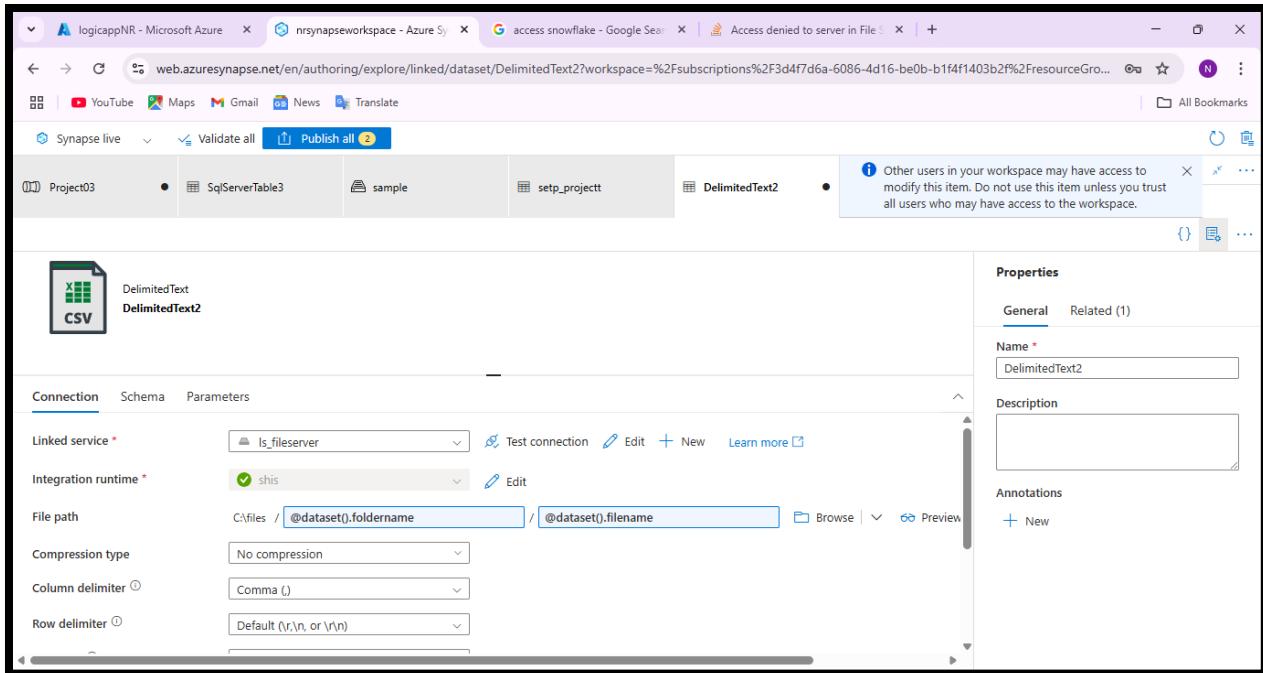


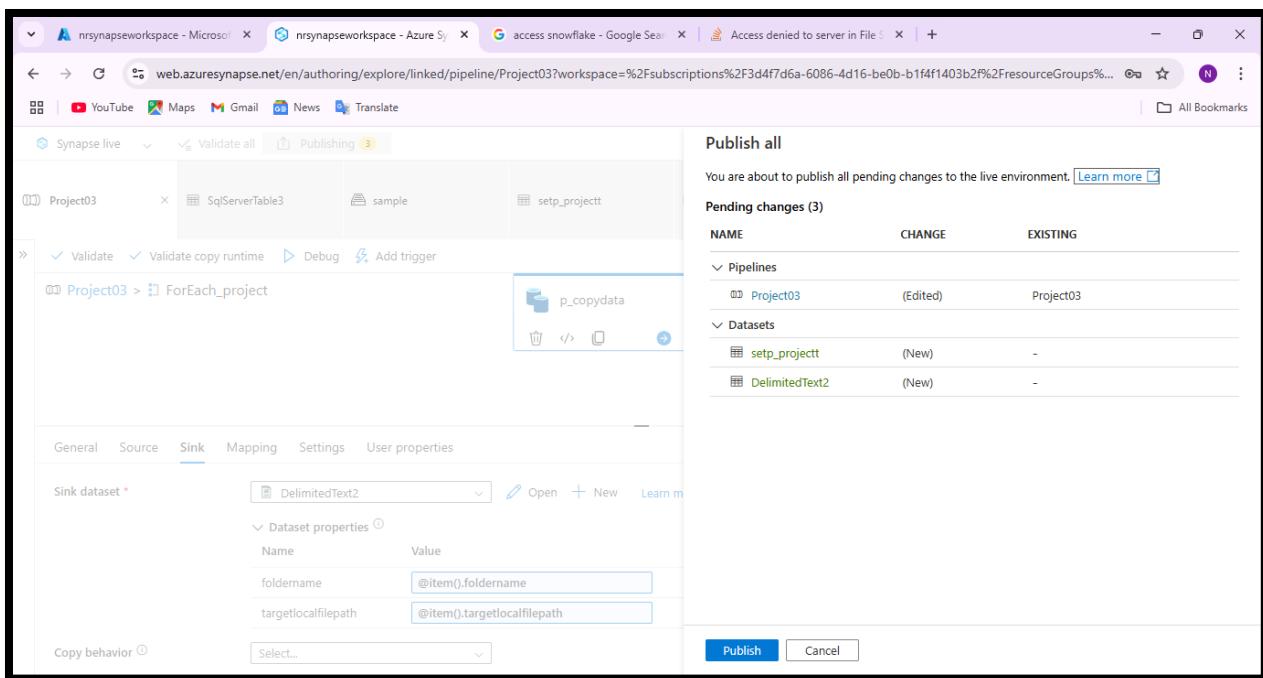
### 34.SELECT SHIR,HOST ,CREATE LINKED SERVICE.

HERE I WANT TO LOAD FROM CLOUD TO ON-PREMISE THAT'S WHY I USE "SHIR". THAT I HAVE CREATED IN VM.

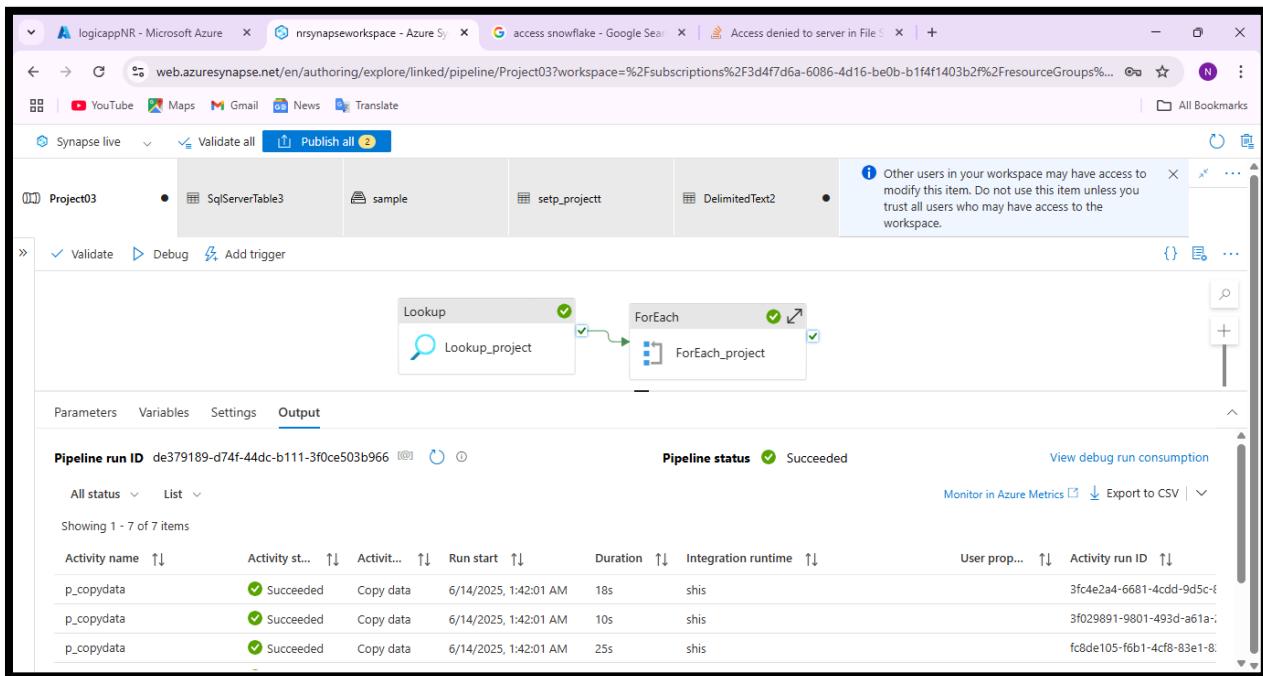


.CREATE PARAMETERS FOR FOLDERNAME AND FILENAME.

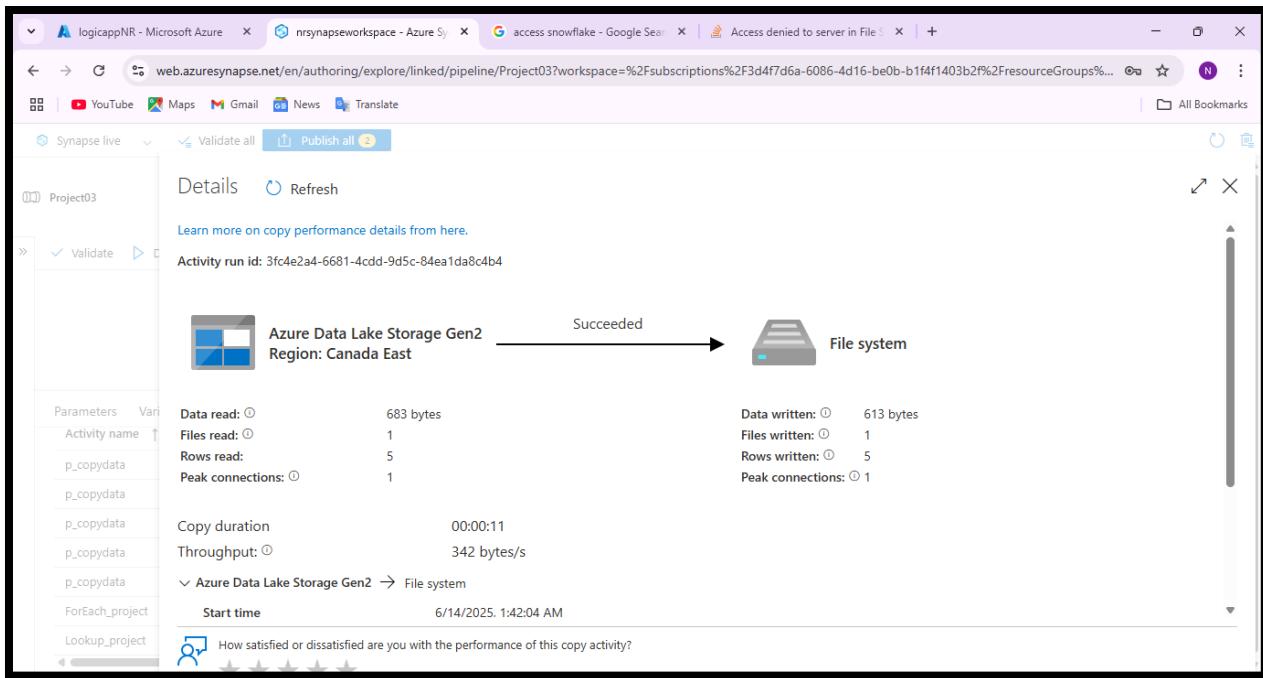




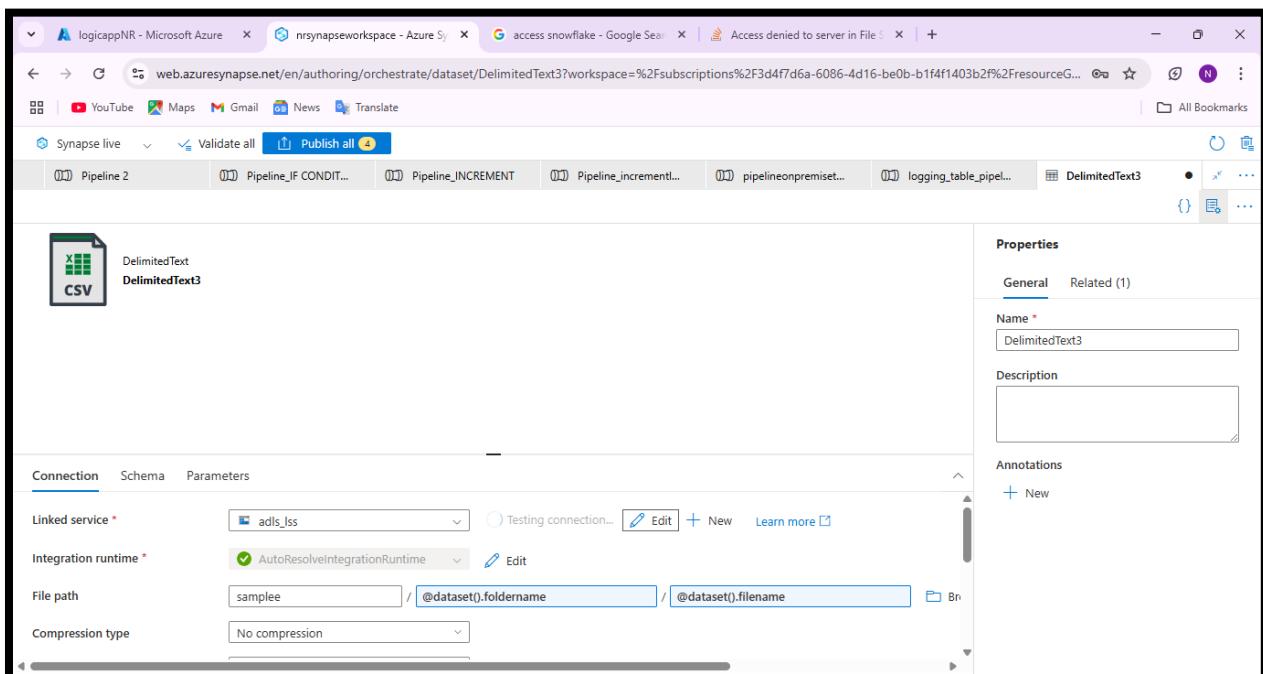
### 34.PUBLISH PIPELINE.

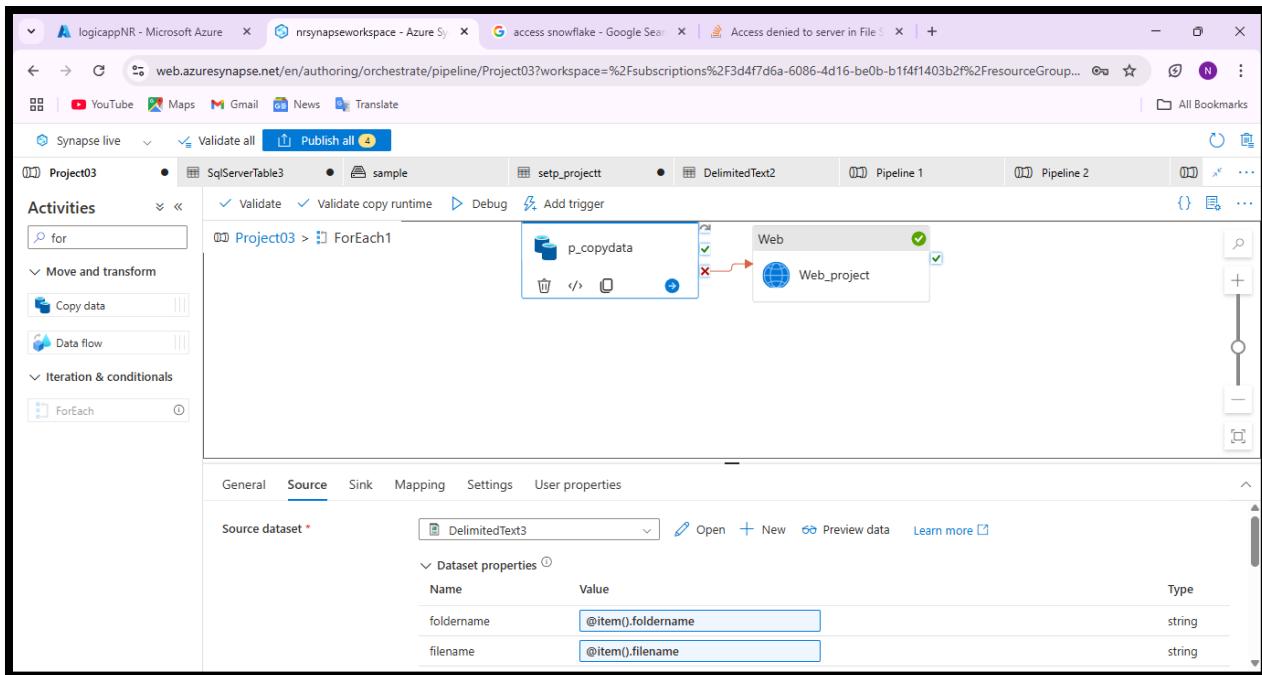


### 34.PIPELINE STATUS "SUCCEEDED".

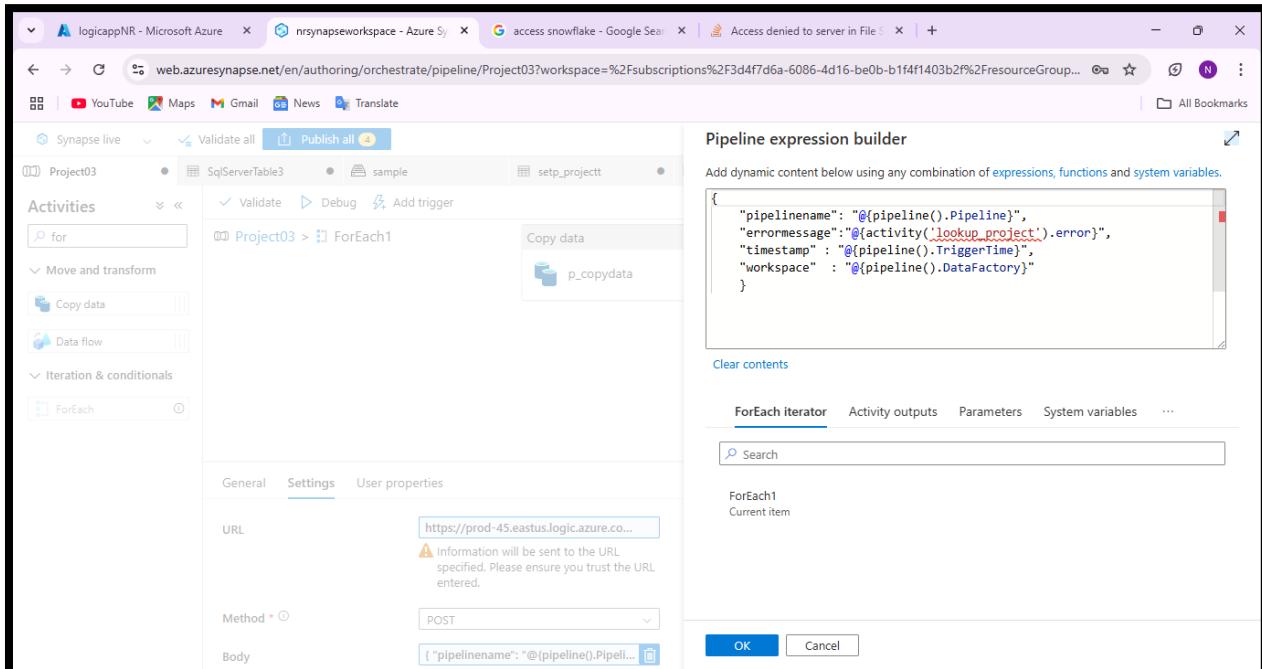


### 35. DATA LOADED FROM CLOUD TO ON-PREMISE.

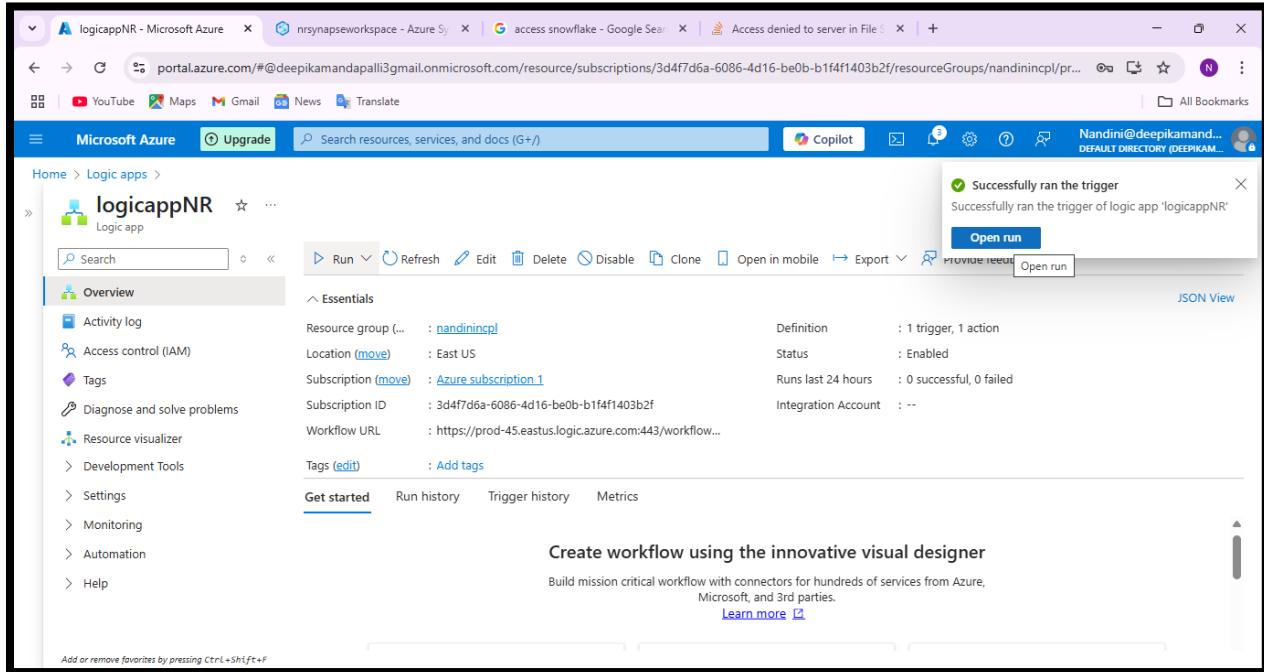




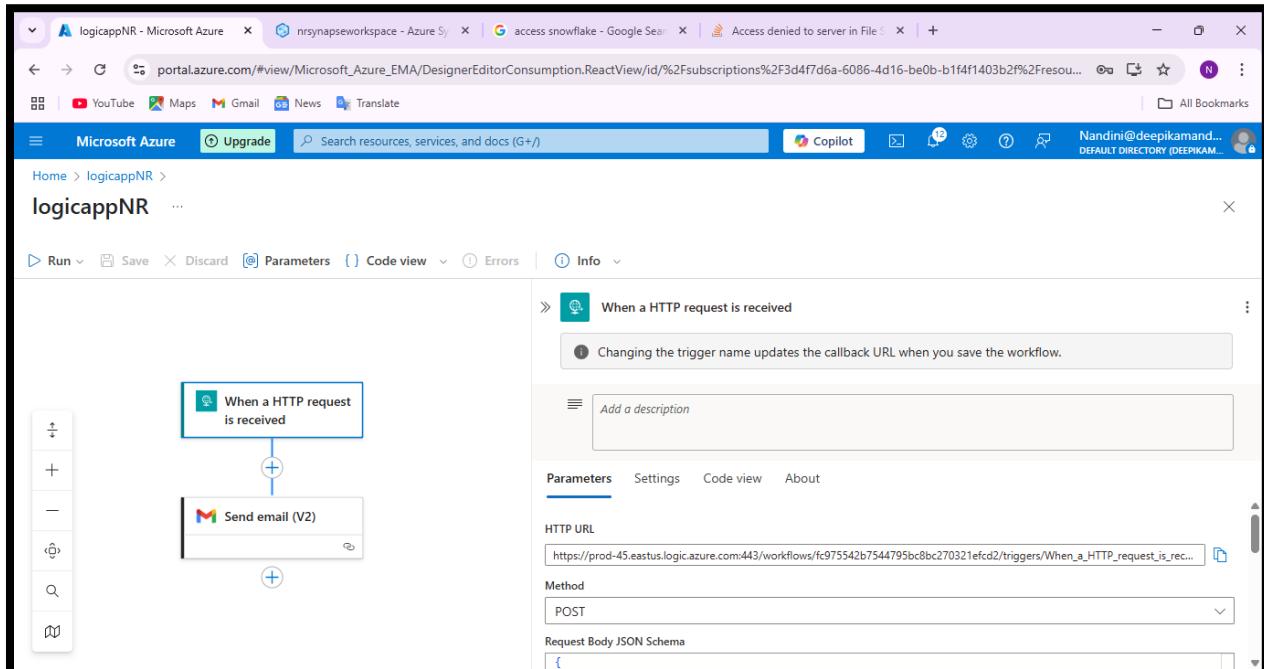
36. NOW CONNECT WEB ACTIVITY WITH COPY DATAIN FOREACH ACTIVITY. WE CONNECTED WEB ACTIVITY BECAUSE IF COPY DATA FAILS IT WILL SEND MAIL ABOUT PIPELINE STATUS BY CONNECTING SYNAPSE PIPELINE WITH LOGIC APPS.



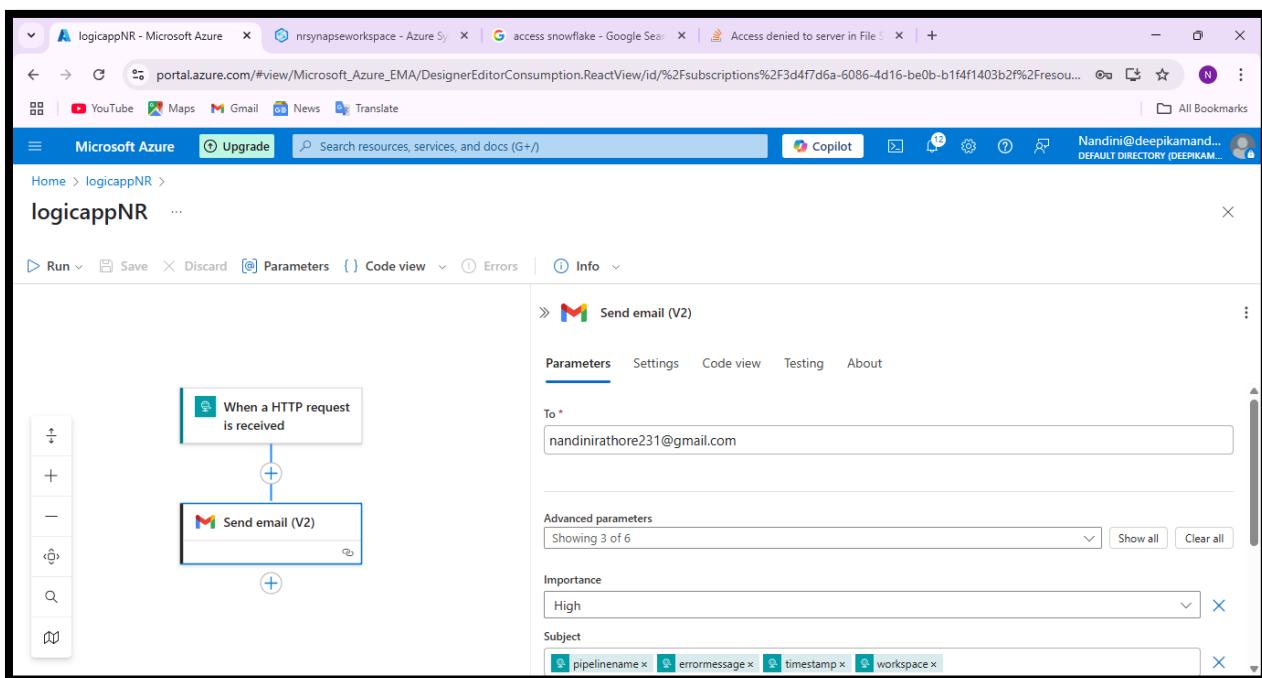
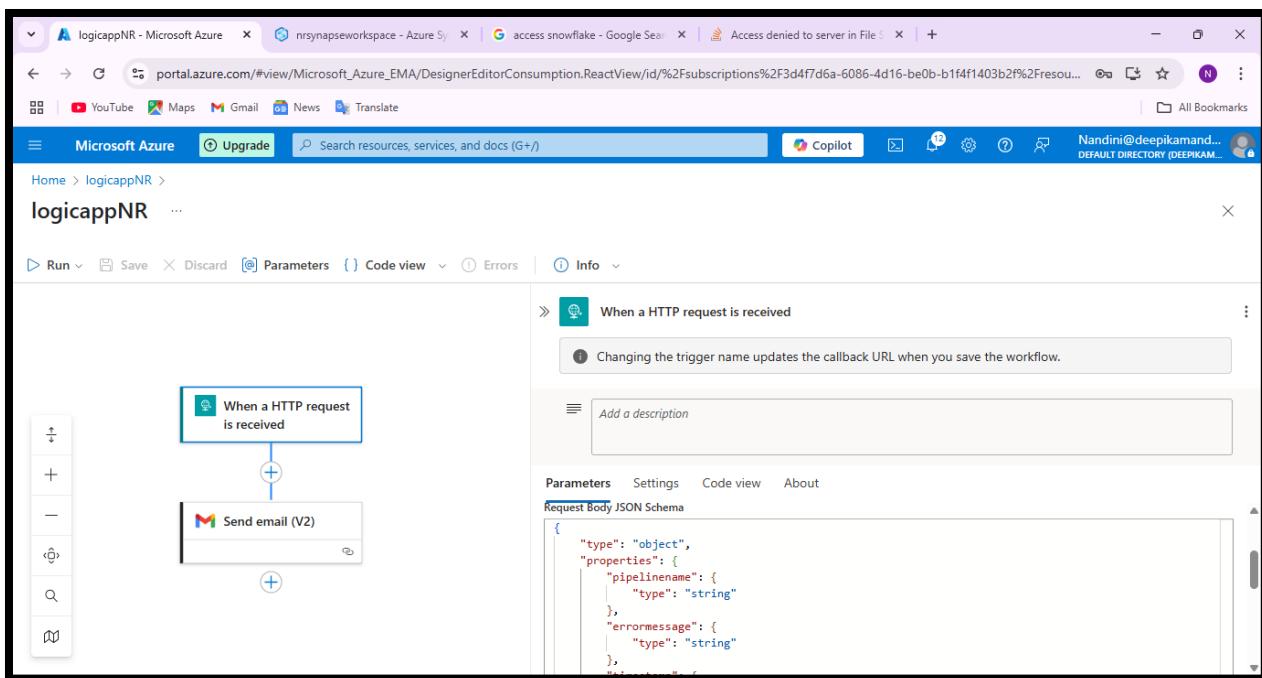
**5. Implement the email configuration task by using logic apps and send an email for any failure.**



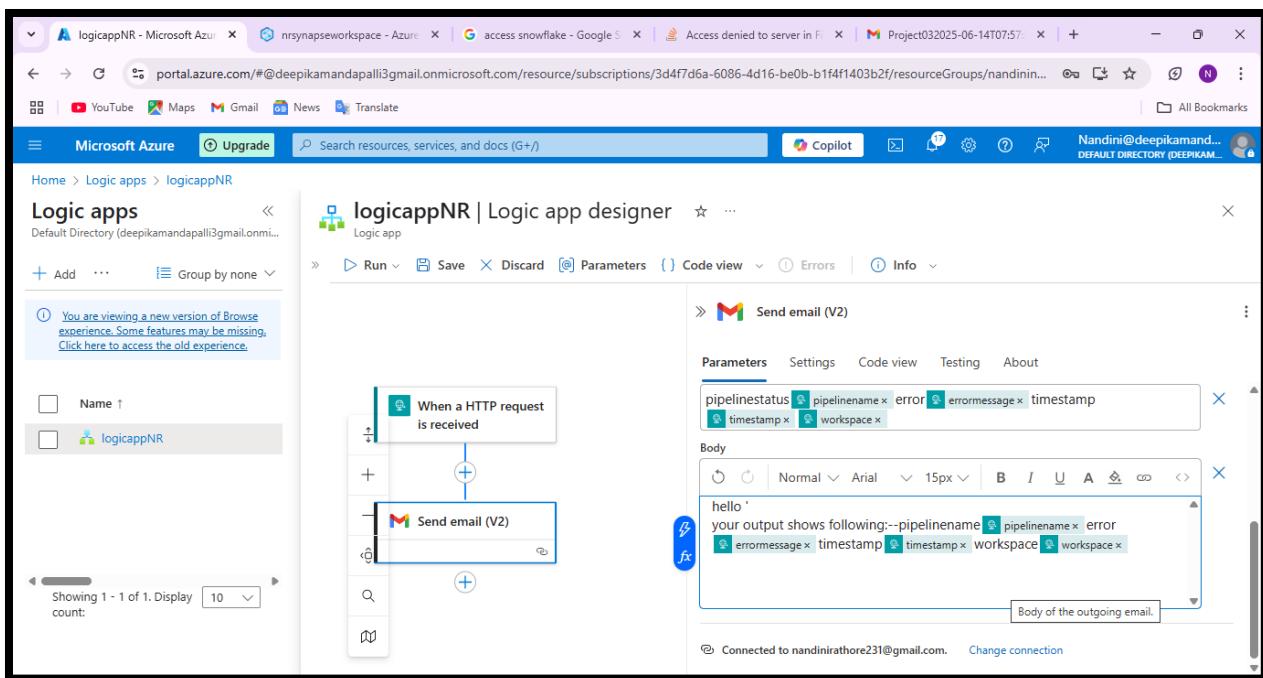
**37. OPEN LOGIC APP, RUN IT.**



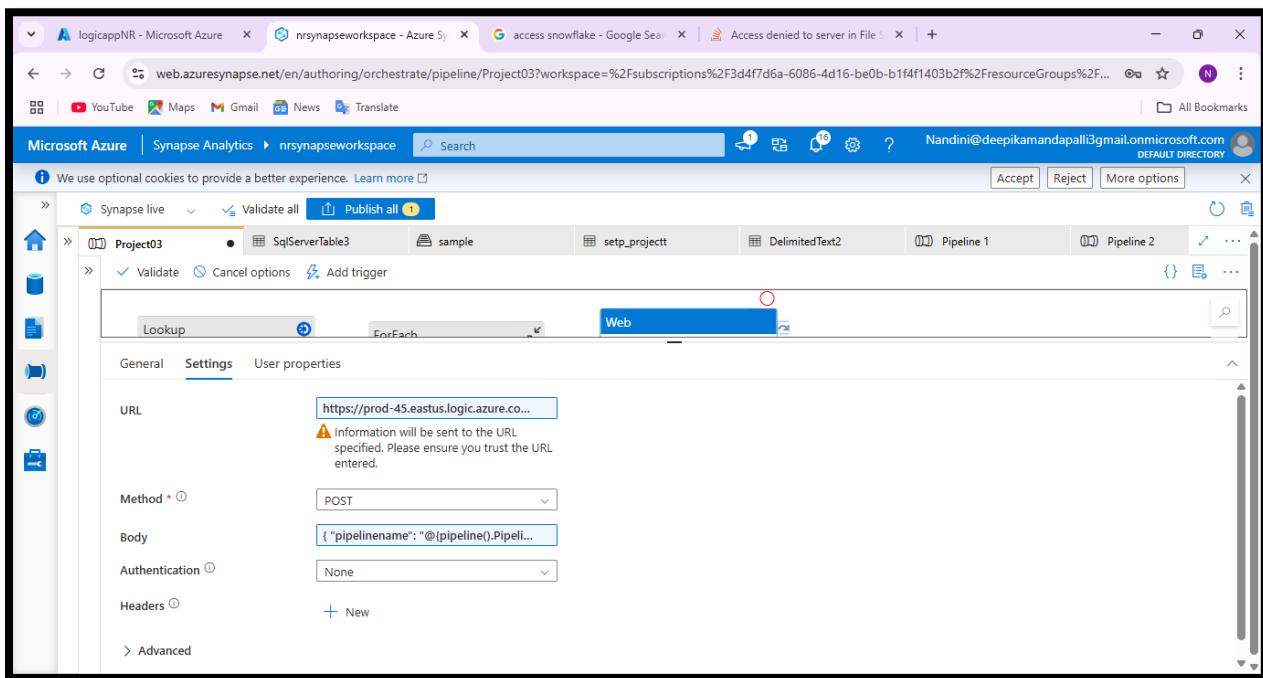
37. IN    HTTP,    ADD    POST    METHOD    ,IN    REQUEST    BODY    JSON    SCHEMA → WRITE  
PIPELINENAME,ERRORMESSAGE,TIMESTAMP,WORKSPACE IN JSON FORMAT THAT YOU WANT TO SEE IN  
MESSAGE.



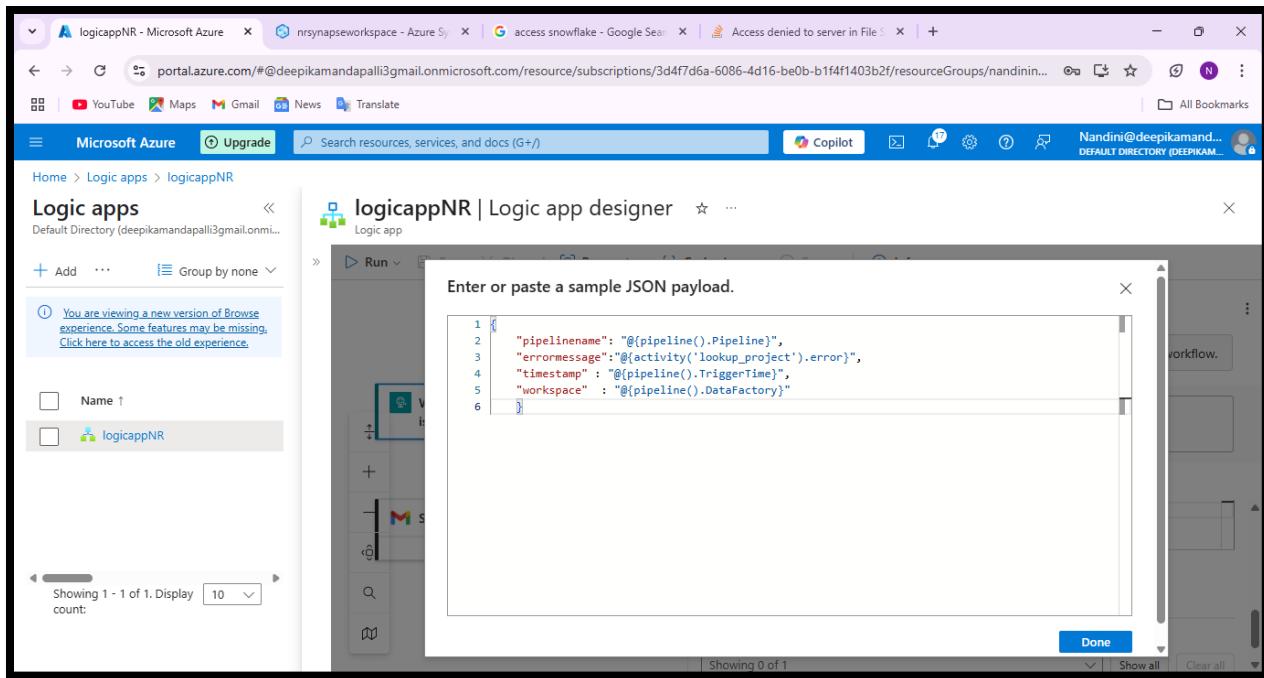
WRITE EMAIL ADDRESS TO WHOM YOU WANT WANT TO SEND A MESSAGE.SELECT PARAMETERS;IMPORTANCE,SUBJECT AND BODY.

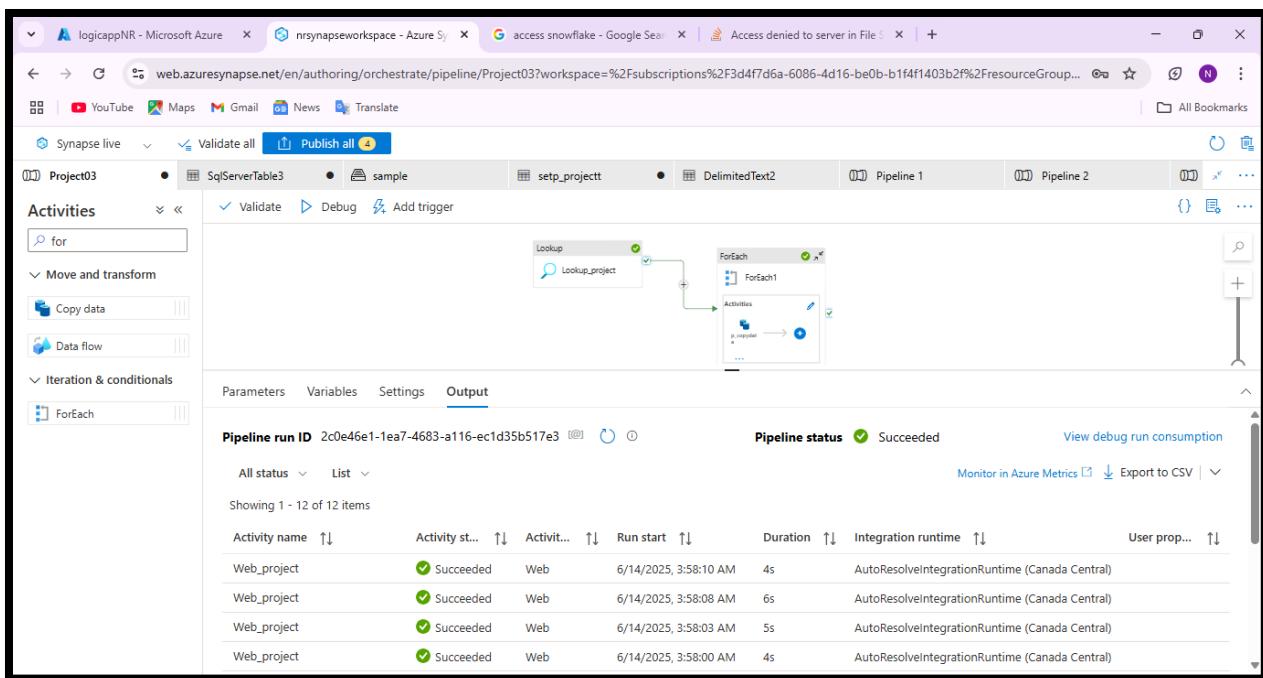


WRITE WHAT YOU WANT IN YOUR MESSAGE IN BODY. NOW SAVE THIS WORKFLOW. GO TO AZURE SYNAPSE.

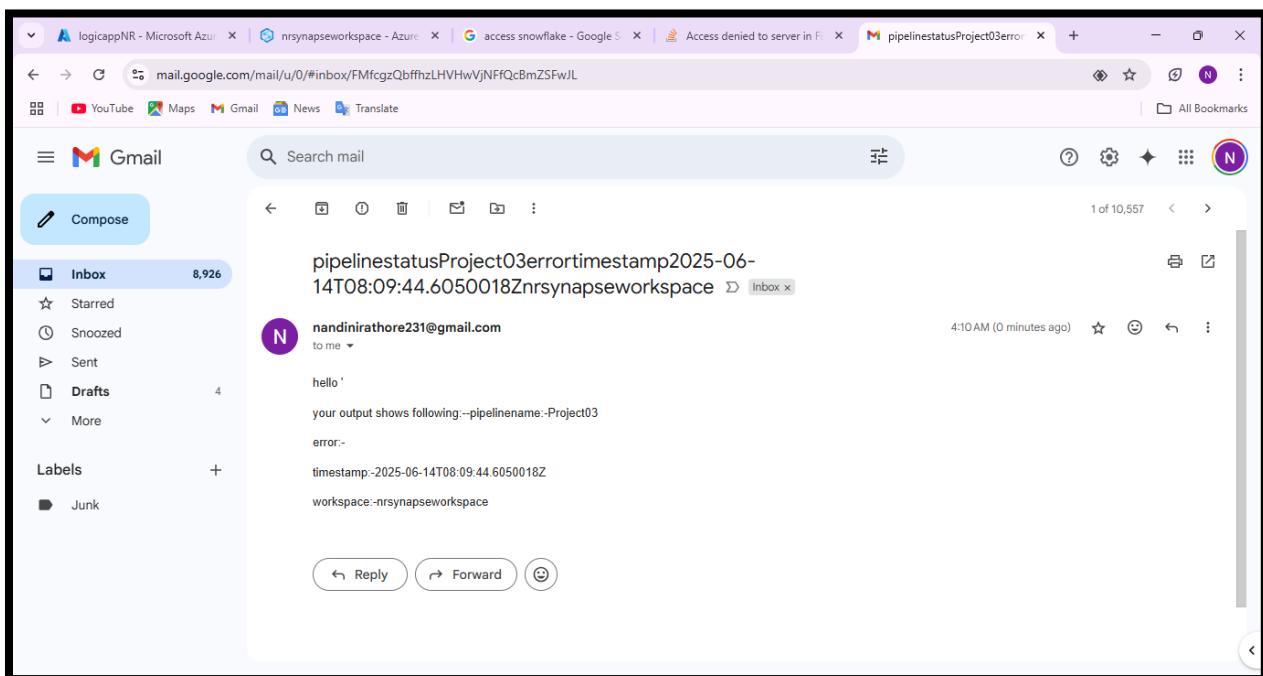


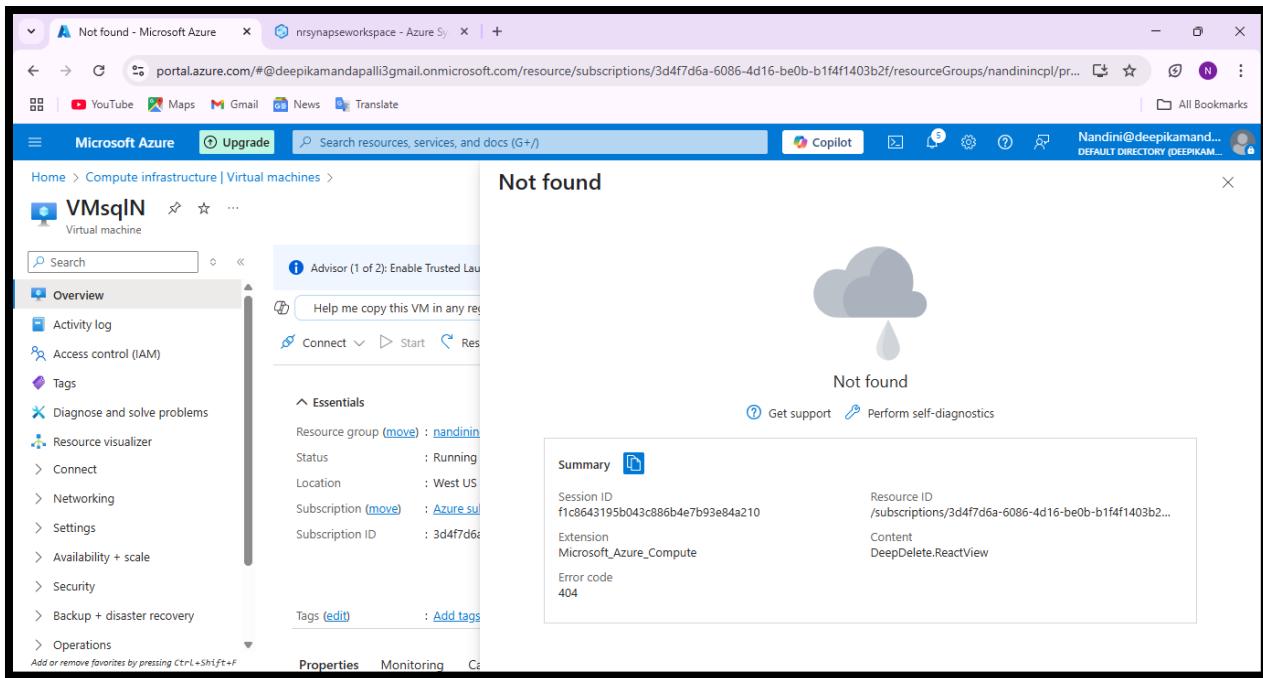
PASTE IN HTTP REQUEST WHERE JSON SCHEMA IS WRITTEN AND FROM THERE COPY URL AND PASTE THAT URL INTO URL OPTION OF WEB ACTIVITY.





I DID SOME MISTAKE MANUALLY SO THAT COPY ACTIVITY FAILS AND IT WILL SEND ME MAIL BY USING WEB ACTIVITY. SO HERE WEB ACTIVITY IS SUCCESSFUL AND SEND ME AN EMAIL OF PIPELINE FAILURE SHOWING PIPELINENAME,ERRORMESSAGE,TIMESTAMP ,WORKSPACE.





AFTER DOING ALL THE WORK DELETE VM.