	Global Influenza Branch		
Job Aid: Data Factory			
Doc. No.	Rev. No. 01	Effective Date:	Page 1 of 9

Purpose:

This document is to set up Data Factory, designed to deliver extraction, transformation, and loading (ETL) processes within the cloud.

Scope:

This document applies to all data analyst and epidemiologist working collaboratively using Data Factory to create, schedule and manage data pipelines that can move and transform data from various sources to different destinations.

Materials:

File Name	Location
Data Factory	Password Vault (cdc.gov)
ADF	ncird-ndlp-id-gib-dev-adf
Databricks	https://adb-927644022200234.14.azuredatabricks.net/
Workspace Common Folder	id-gib
Azure DevOps Git Repo	https://dev.azure.com/CDC-Org1/NDSP/_git/id-gib-adf

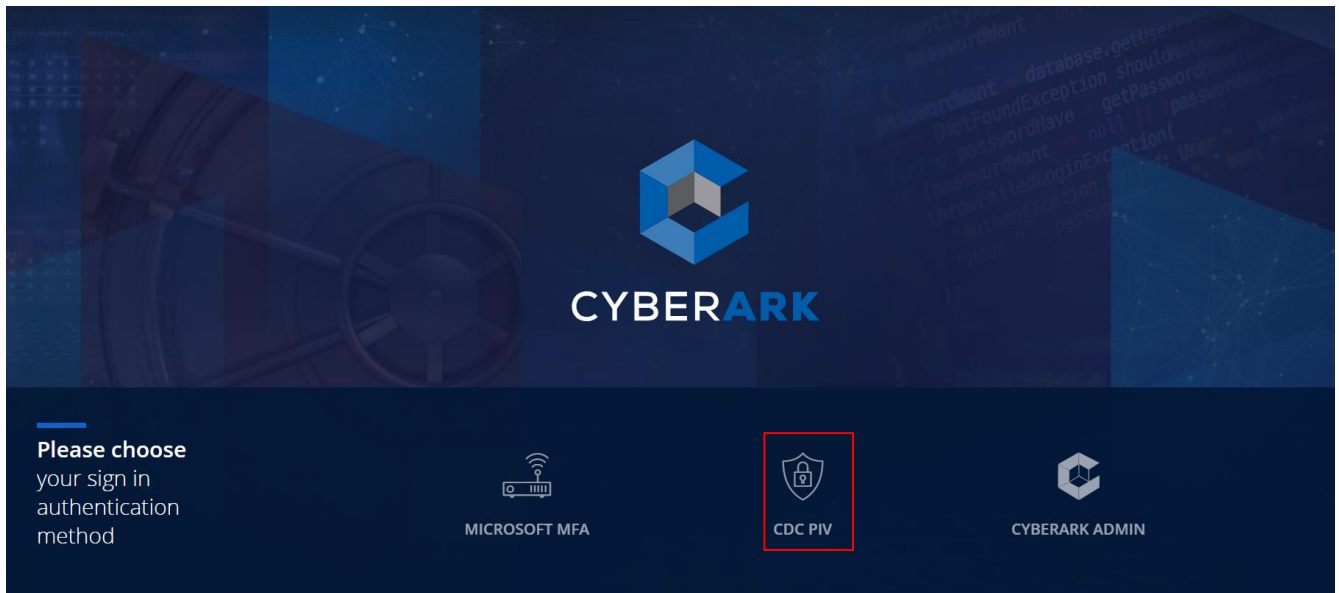
Prerequisites:

- Need SU account created.
- Basic understanding of data exchange.
- R/ SAS Viya programming language

Process:

1. Log in using CDC PIV credentials (this should show your su account as cdcicid-su@cdc.gov)

	Global Influenza Branch		
Job Aid: Data Factory			
Doc. No.	Rev. No. 01	Effective Date:	Page 2 of 9



2. Authenticate your user account and click ok.

Select a certificate for authentication

Site pvwacl.cdc.gov:443 needs your credentials:

	126e7e62-1003-4a06-a227-1944408176c6 MS-Organization-Access 1/10/2022
	Kimberly Y. Jones -A HHS-FPKI-Intermediate-CA-E1 Authentication - Kimberly Y. Jones -A 8/22/2022

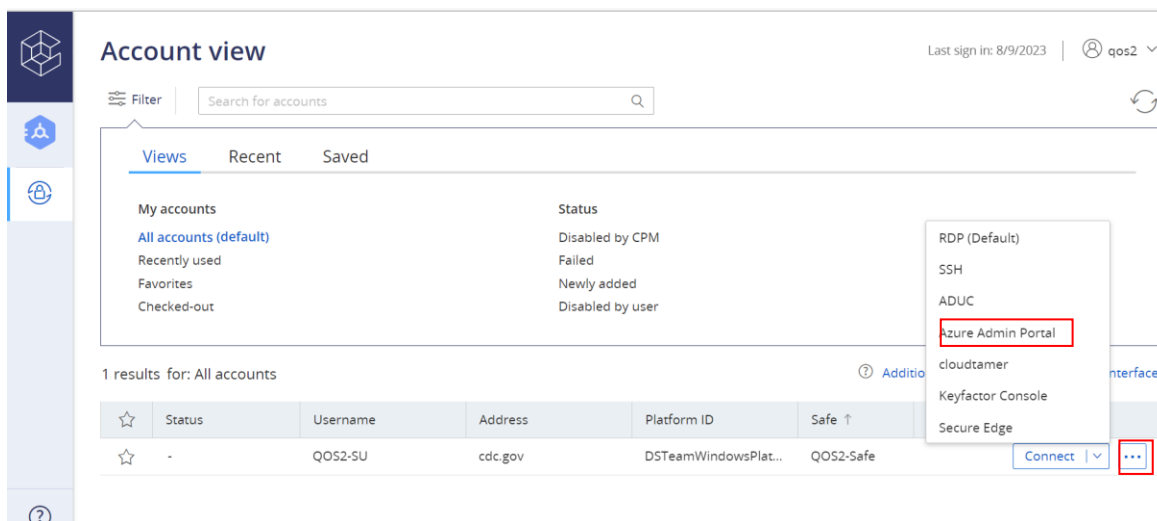
[Certificate information](#)

OK

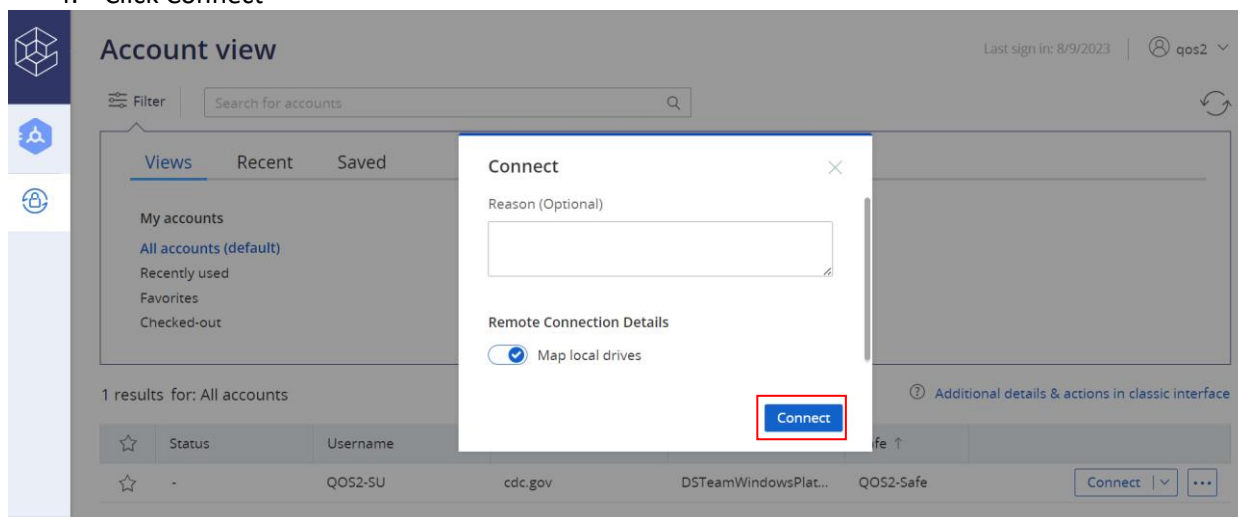
Cancel

3. Connect with "Azure Admin Portal"

- a. You will need to select the 3 dots and copy your new ssh key to enter databricks with your SU account.
- b. SSH key is only available for same day usage.



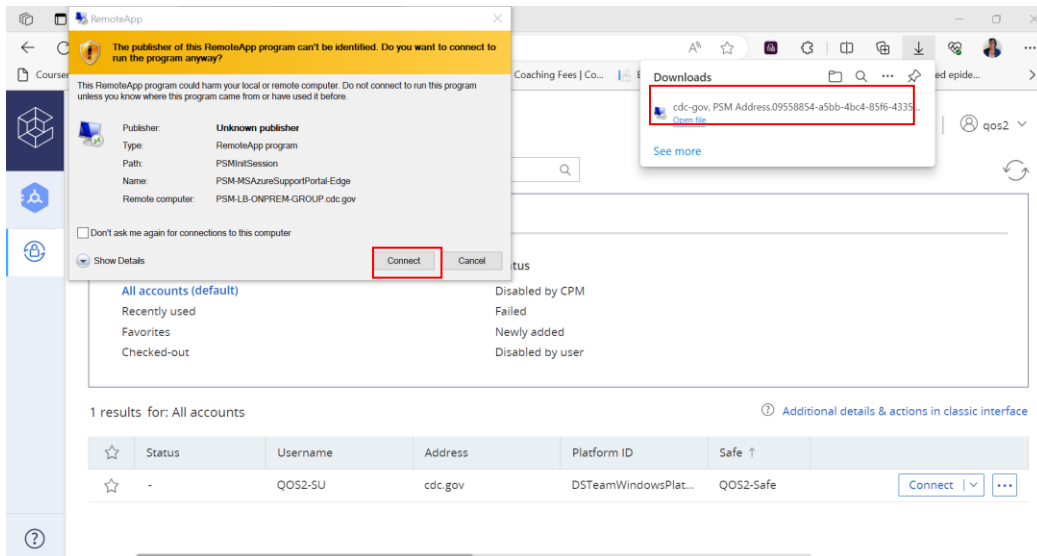
4. Click Connect



5. Click on "openfile" to download the Azure Data Factory

a. Click on connect

	Global Influenza Branch		
Job Aid: Data Factory			
Doc. No.	Rev. No. 01	Effective Date:	Page 4 of 9

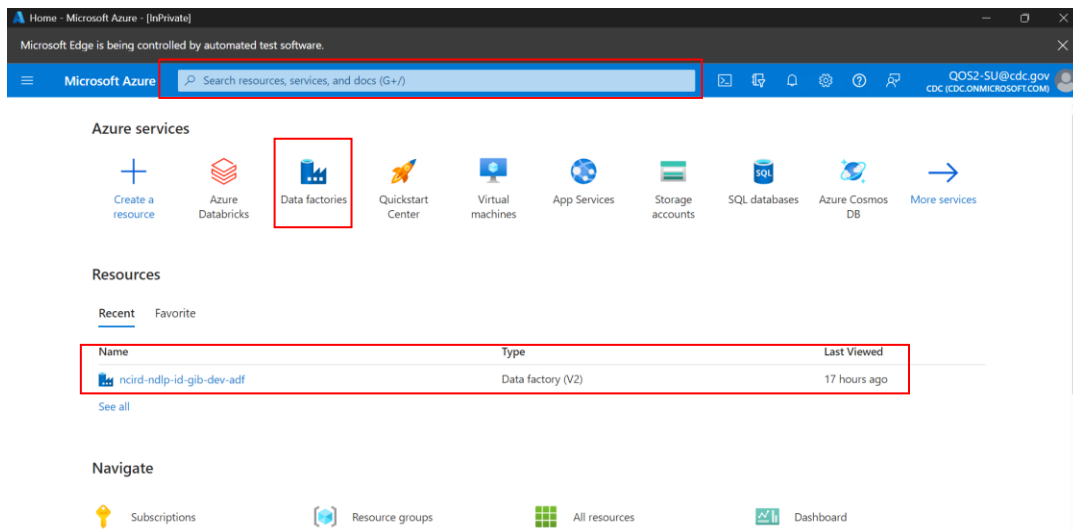


6. Log into your SU account
 - a. Use the password from step 3a.

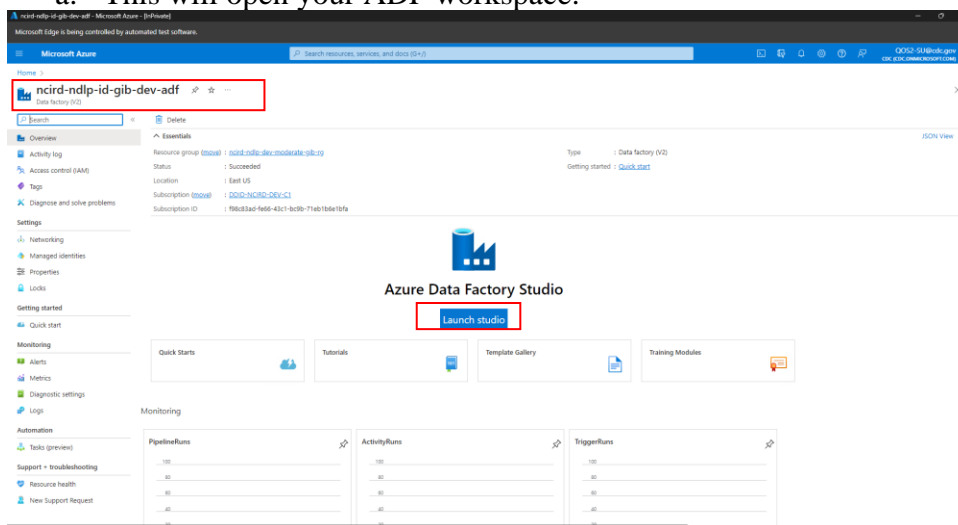


7. In the search bar "Data Factories" to add as a feature.
8. Click on the ADF resource link, "[ncird-ndlp-id-gib-dev-adf](#)".

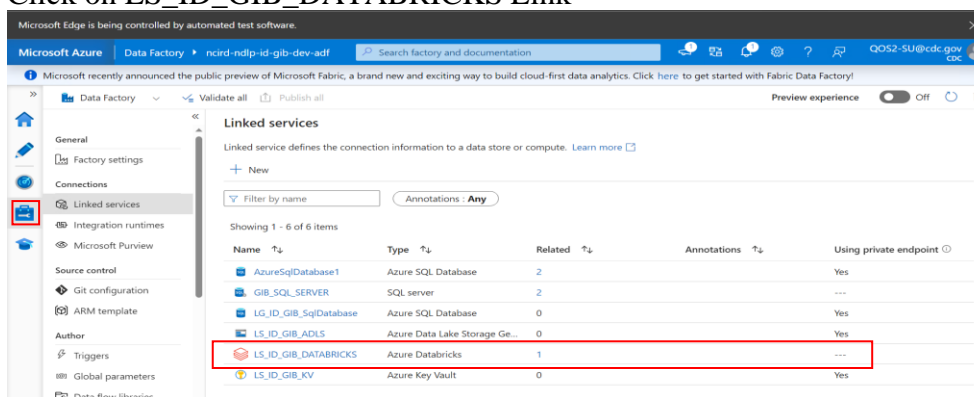
	Global Influenza Branch		
Job Aid: Data Factory			
Doc. No.	Rev. No. 01	Effective Date:	Page 5 of 9



9. Click “Launch studio”
 - a. This will open your ADF workspace.



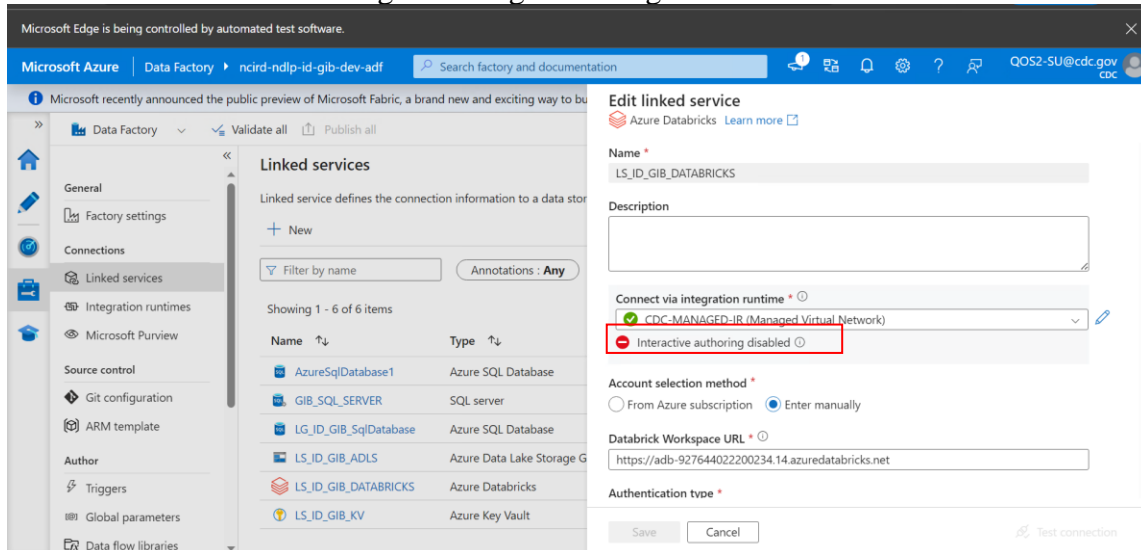
10. Click on the folder icon to get to the linked services.
11. Click on LS_ID_GIB_DATABRICKS Link



	Global Influenza Branch		
Job Aid: Data Factory			
Doc. No.	Rev. No. 01	Effective Date:	Page 6 of 9

12. Click on the pencil icon.

a. Interactive authoring needs to get enable.



13. Click on the words “Virtual network”.

a. Enable

b. Select time to live (30-60 minutes)

14. Click Apply to save the enabled feature.

Edit integration runtime

Settings
Virtual network
Data flow runtime

Virtual network configuration ⓘ

☐ Disable
☒ Enable

Interactive authoring ⓘ

☐ Disable
☒ Enable

Time to live ⓘ

60 minutes



> Advanced

Apply
Cancel

a. You will now see the Interactive authoring enabled with a green check mark.



b. Cancel

	Global Influenza Branch		
Job Aid: Data Factory			
Doc. No.	Rev. No. 01	Effective Date:	Page 7 of 9

Edit linked service
 Azure Databricks [Learn more](#) 

Name *
 LS_ID_GIB_DATABRICKS


Description

Connect via integration runtime * ⓘ
☒ CDC-MANAGED-IR (Managed Virtual Network)  
☒ Interactive authoring enabled ⓘ

Account selection method *
☐ From Azure subscription ☒ Enter manually

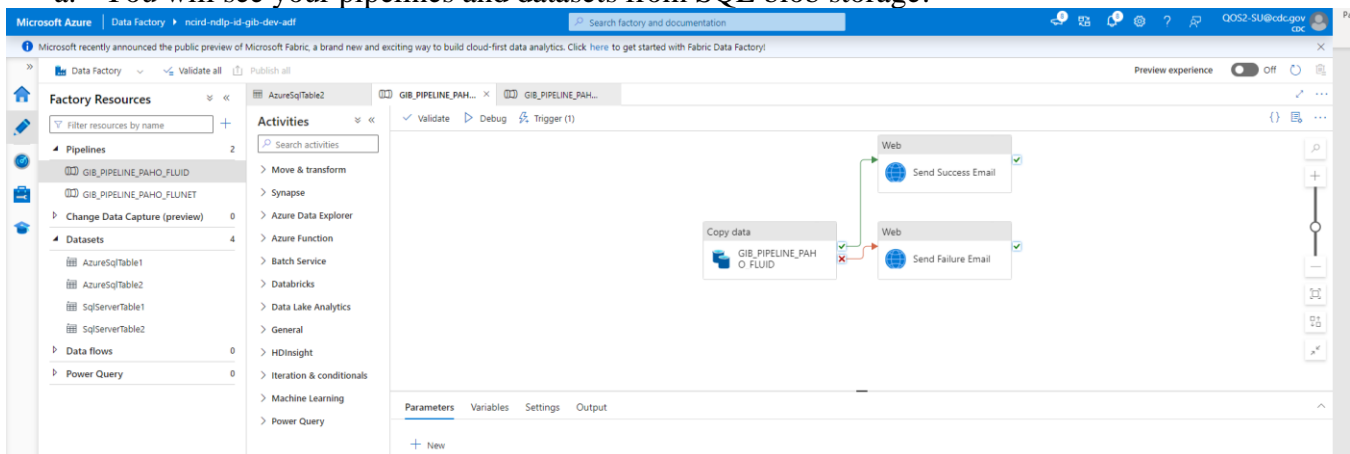
Databricks Workspace URL * ⓘ

Authentication type *

Save **Cancel**  Test connection

15. Click on the Pencil icon.

a. You will see your pipelines and datasets from SQL blob storage.

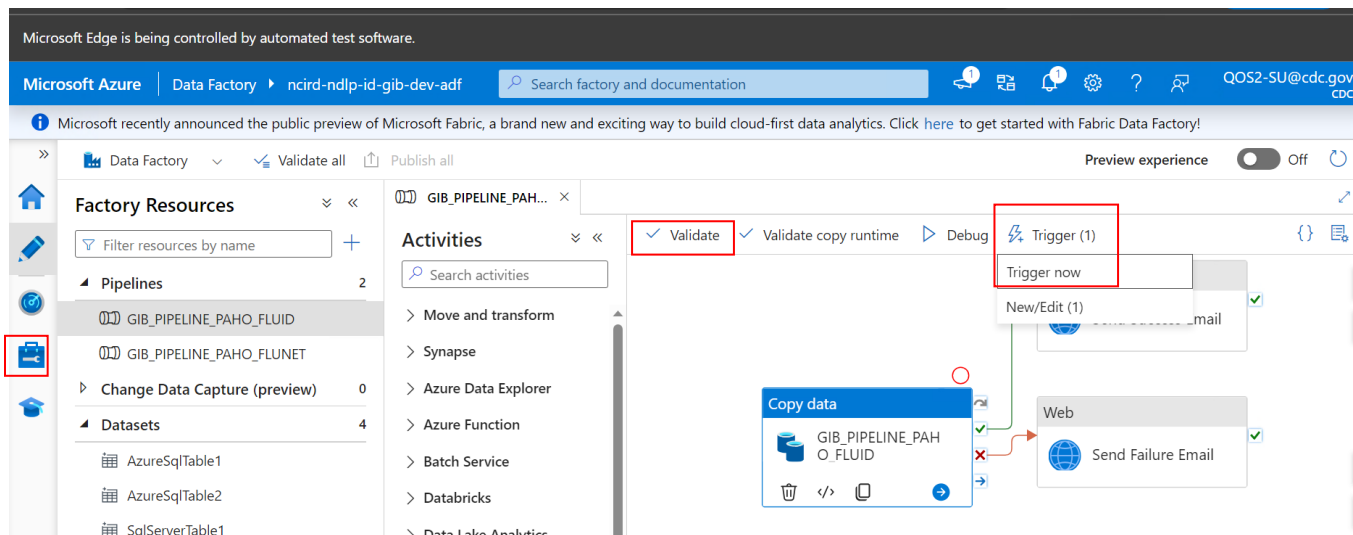


16. Click on pipeline “GIB_PIPELINES_PAHO_FLUID”.

17. Click on the GIB_PIPELINES_PAHO_FLUID picture.

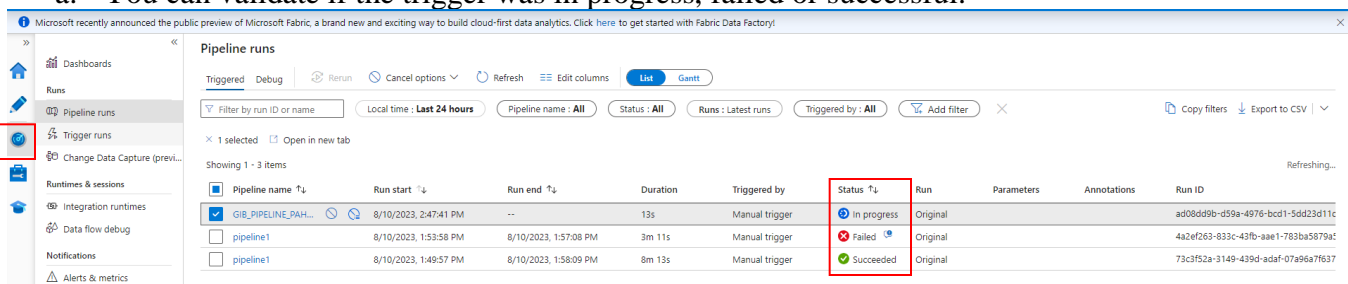
- 1st Validation Activity- confirms the data pipeline control flow based on conditions or rule.
- 2nd Publish- storing the pipeline, data flows and setting the parameters for Source and Sink.
- 3rd Trigger Now- provides the ability to choose and process data from the past.

	Global Influenza Branch		
Job Aid: Data Factory			
Doc. No.	Rev. No. 01	Effective Date:	Page 8 of 9



18. Click on the monitor icon.

a. You can validate if the trigger was in progress, failed or successful.



Pipeline name	Run start	Run end	Duration	Triggered by	Status	Run	Parameters	Annotations	Run ID
GIB_PIPELINE_PAHO_FLUID	8/10/2023, 2:47:41 PM	--	13s	Manual trigger	In progress	Original			ad08dd9b-d59a-4976-bcd1-5dd23d11c
pipeline1	8/10/2023, 1:53:58 PM	8/10/2023, 1:57:08 PM	3m 11s	Manual trigger	Failed	Original			4a2ef263-933c-43fb-aae1-783ba5879a5
pipeline1	8/10/2023, 1:49:57 PM	8/10/2023, 1:58:09 PM	8m 13s	Manual trigger	Succeeded	Original			73c3f52a-3149-439d-adaf-07a96a7f637

19. Email notification will be sent if the pipelines has failed or ran successfully on biweekly schedule.

Pipeline Run Success



Pipeline Run Success

Data Factory Name: ncird-ndlp-id-gib-dev-adf
 Pipeline Name: GIB_PIPELINE_PAHO_FLUID
 Pipeline Run Id: c823299e-0ec1-41b7-b846-efc8413933e0
 Time: 2023-08-11T15:55:41.5192379Z

Information

Pipeline run finished successfully!

20. Click Alert bell icon > Close.

a. Notifications appear with explanations of each status.

	Global Influenza Branch		
Job Aid: Data Factory			
Doc. No.	Rev. No. 01	Effective Date:	Page 9 of 9

