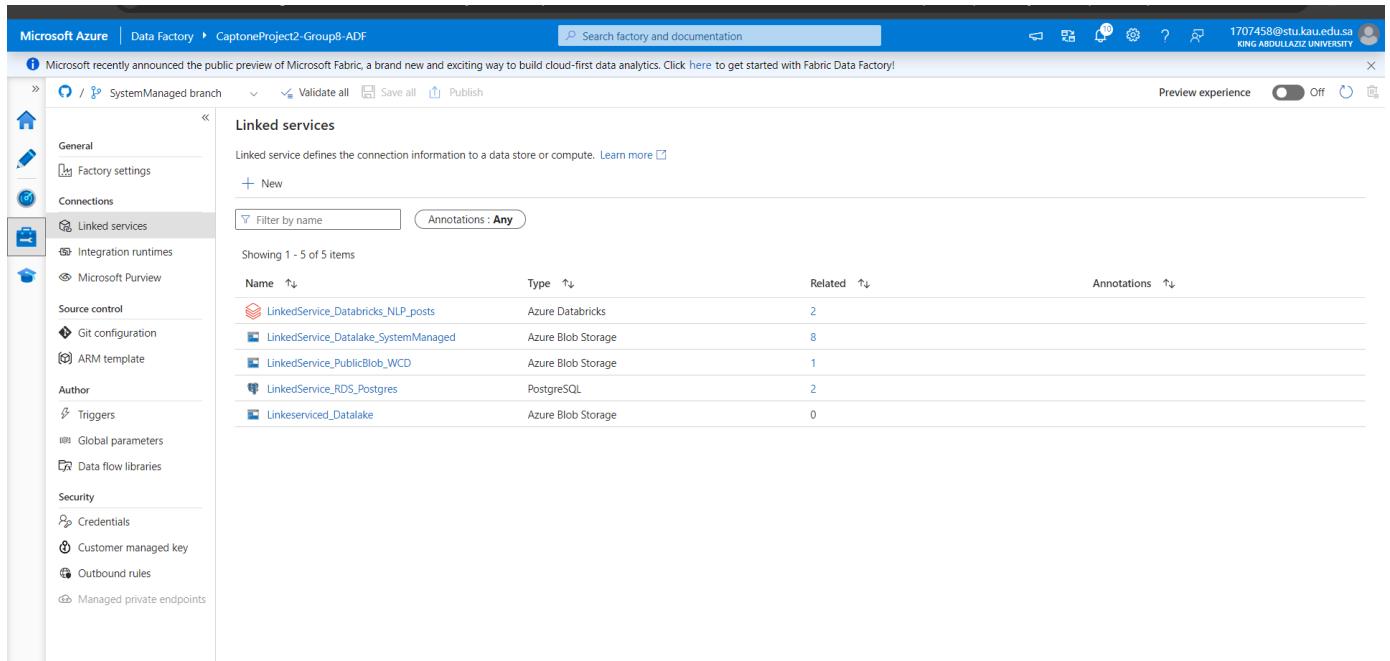


# Data Factory

## 1- Linked Services:



Microsoft Azure | Data Factory > CapstoneProject2-Group8-ADF

Search factory and documentation

1 Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!

SystemManaged branch Validate all Save all Publish Preview experience Off

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 Customer managed key Outbound rules Managed private endpoints

Linked services

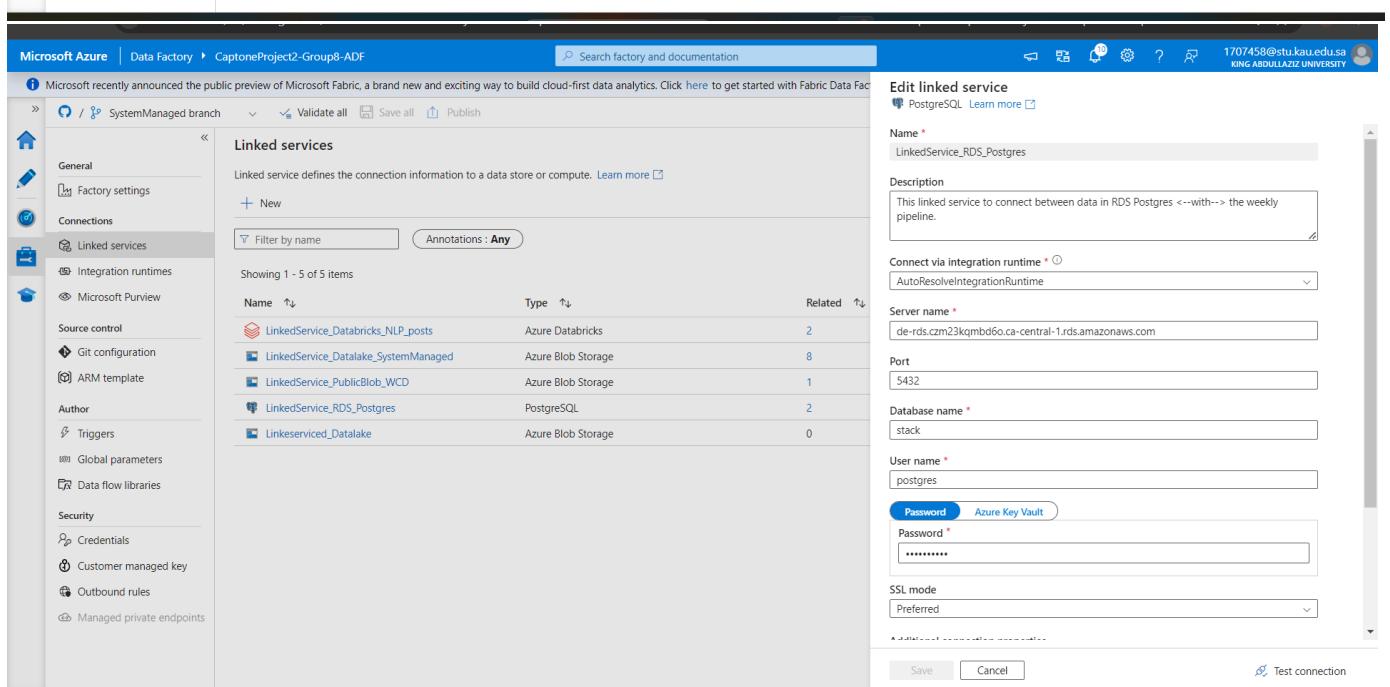
Linked service defines the connection information to a data store or compute. Learn more

+ New

Filter by name Annotations : Any

Showing 1 - 5 of 5 items

Name	Type	Related	Annotations
LinkedService_Databricks_NLP_posts	Azure Databricks	2	
LinkedService_Datalake_SystemManaged	Azure Blob Storage	8	
LinkedService_PublicBlob_WCD	Azure Blob Storage	1	
LinkedService_RDS_Postgres	PostgreSQL	2	
Linkeserviced_Datalake	Azure Blob Storage	0	

Microsoft Azure | Data Factory > CapstoneProject2-Group8-ADF

Search factory and documentation

1 Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!

SystemManaged branch Validate all Save all Publish

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 Customer managed key Outbound rules Managed private endpoints

Linked services

Linked service defines the connection information to a data store or compute. Learn more

+ New

Filter by name Annotations : Any

Showing 1 - 5 of 5 items

Name	Type	Related	Annotations
LinkedService_Databricks_NLP_posts	Azure Databricks	2	
LinkedService_Datalake_SystemManaged	Azure Blob Storage	8	
LinkedService_PublicBlob_WCD	Azure Blob Storage	1	
LinkedService_RDS_Postgres	PostgreSQL	2	
Linkeserviced_Datalake	Azure Blob Storage	0	

Edit linked service

PostgreSQL Learn more

Name \* LinkedService\_RDS\_Postgres

Description This linked service to connect between data in RDS Postgres <--with--> the weekly pipeline.

Connect via integration runtime \* AutoResolveIntegrationRuntime

Server name \* de-rds.czr23kqmbd60.ca-central-1.rds.amazonaws.com

Port 5432

Database name \* stack

User name \* postgres

Password Azure Key Vault

Password \* \*\*\*\*\*

SSL mode Preferred

Save Cancel Test connection

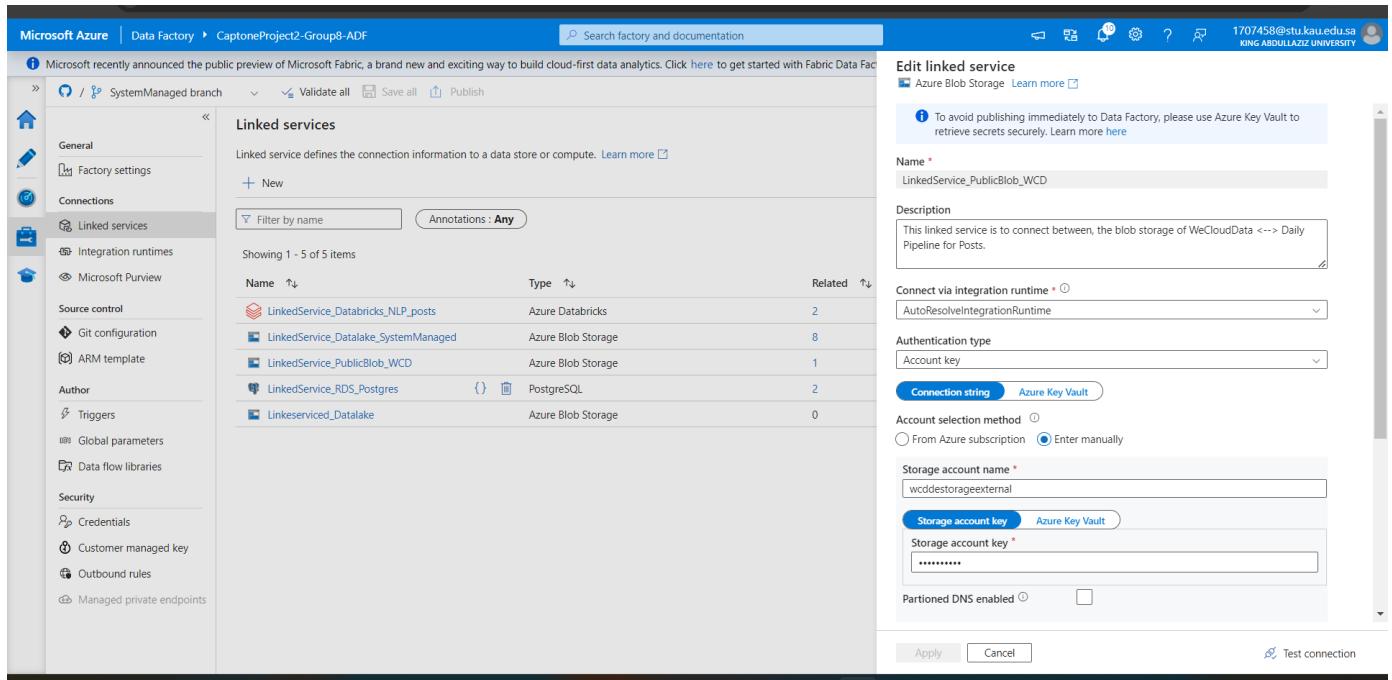
## Related

 [LinkedService\\_RDS\\_Postgres](#)

This linked service is being used by the following:

### Datasets

[Source\\_PostType\\_RDS\\_postgres](#)  
[Source\\_Users\\_RDS\\_postgres](#)



The screenshot shows the Microsoft Azure Data Factory interface. On the left, there's a navigation sidebar with various options like General, Connections, and Source control. The main area is titled 'Linked services' and shows a list of existing linked services: 'LinkedService\_Databricks\_NLP\_posts' (Azure Databricks), 'LinkedService\_Datalake\_SystemManaged' (Azure Blob Storage), 'LinkedService\_PublicBlob\_WCD' (Azure Blob Storage), 'LinkedService\_RDS\_Postgres' (PostgreSQL), and 'Linkeserviced\_Datalake' (Azure Blob Storage). To the right, a 'New' dialog is open for creating a new linked service. It's configured for 'Azure Blob Storage' with the name 'LinkedService\_PublicBlob\_WCD'. The 'Connection string' tab is selected, and it's set to 'Enter manually'. The 'Storage account name' is 'wcddestorageexternal' and the 'Storage account key' is a masked password. Other tabs like 'Azure Key Vault' and 'From Azure subscription' are also visible.

## Related

 [LinkedService\\_PublicBlob\\_WCD](#)

This linked service is being used by the following:

### Datasets

[Source\\_Posts\\_PublicBlob\\_WCD](#)

The screenshot shows the Microsoft Azure Data Factory interface. On the left, there's a sidebar with various options like General, Connections, Linked services, Integration runtimes, and Microsoft Purview. The main area shows a list of 'Linked services' with columns for Name, Type, and Related. On the right, a detailed 'Edit linked service' dialog is open for 'LinkedService\_Datalake\_SystemManaged'. The dialog includes fields for Name (set to 'LinkedService\_Datalake\_SystemManaged'), Description (a note about connecting to destination datasets), Connect via integration runtime (set to 'AutoResolveIntegrationRuntime'), Authentication type (System Assigned Managed Identity), Account selection method (set to 'Enter manually'), Service endpoint (set to '@{linkedService().ls\_prm\_service\_endpoint}'), and Storage account kind (StorageV2). There are also sections for Managed identity name, Test connection, and Annotations.

## Related

### [LinkedService\\_Datalake\\_SystemManaged](#)

This linked service is being used by the following:

#### [Datasets](#)

[Archive\\_Daily\\_Posts](#)  
[Archive\\_Weekly\\_Posttype](#)  
[Archive\\_Weekly\\_Users](#)  
[DestinationDS\\_PostType\\_Datalake](#)  
[DestinationDS\\_Posts\\_Datalake](#)  
[DestinationDS\\_Users\\_Datalake](#)

#### [Pipelines](#)

[Daily\\_Running\\_PublicBlob\\_Pipeline](#)  
[Weekly\\_Running\\_RDS\\_Pipeline](#)

The screenshot shows the Microsoft Azure Data Factory interface. On the left, there's a sidebar with various navigation options like General, Connections, Linked services, and Integration runtimes. The 'Linked services' option is selected. In the main area, there's a table listing five linked services:

Name	Type	Related
LinkedService_Databricks_NLP_posts	Azure Databricks	2
LinkedService_Datalake_SystemManaged	Azure Blob Storage	8
LinkedService_PublicBlob_WCD	Azure Blob Storage	1
LinkedService_RDS_Postgres	PostgreSQL	2
Linkeserviced_Datalake	Azure Blob Storage	0

To the right, a detailed 'Edit linked service' dialog is open for 'LinkedService\_Databricks\_NLP\_posts'. It includes fields for Name (set to 'LinkedService\_Databricks\_NLP\_posts'), Description (containing notes about linking with Databricks), Connect via integration runtime (set to 'AutoResolveIntegrationRuntime'), Account selection method (set to 'Enter manually'), Databrick Workspace URL (set to 'https://adb-560590091399003.3.azuredatabricks.net'), Authentication type (set to 'Access Token'), and a password field. It also shows options for selecting a cluster (set to 'Existing interactive cluster') and an existing cluster ID ('0510-165754-cf53jv82'). At the bottom are 'Save', 'Cancel', and 'Test connection' buttons.

## Related

### [LinkedService\\_Databricks\\_NLP\\_posts](#)

This linked service is being used by the following:

### [Pipelines](#)

[Daily\\_Running\\_PublicBlob\\_Pipeline](#)  
[Weekly\\_Running\\_RDS\\_Pipeline](#)

## 2- Datasets:

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists various components: Pipelines (2), Change Data Capture (preview) (0), Datasets (9), Data flows (0), Power Query (0), and Templates (0). The 'Datasets' section is expanded, showing datasets like Archive\_Daily\_Posts, Archive\_Weekly\_Posttype, Archive\_Weekly\_Users, DestinationDS\_Posts\_Datalake, DestinationDS\_PostType\_Datalake, DestinationDS\_Users\_Datalake, and Source\_Posts\_PublicBlob\_WCD. The 'Source\_Posts\_PublicBlob\_WCD' dataset is selected and shown in the main pane. The dataset details show it is a Parquet file type. The 'Connection' tab is active, displaying a linked service named 'LinkedService\_PublicBlob\_WCD' (selected from a dropdown menu), a file path of 'de-project-st / archive / \*parquet', and a compression type of 'snappy'. The 'Schema' and 'Parameters' tabs are also present.

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar is identical to the previous screenshot, showing the same list of resources and the expanded 'Datasets' section. The 'Datasets' section now highlights the 'Source\_PostType\_RDS\_postgres' dataset. The main pane displays the dataset details. The 'Connection' tab is active, showing a linked service named 'LinkedService\_RDS\_Postgres' (selected from a dropdown menu). The 'Table' dropdown is set to 'Loading...', and there is a checkbox for 'Enter manually'. The 'Parameters' tab is also visible.

Microsoft Azure | Data Factory > CapstoneProject2-Group8-ADF

Search factory and documentation

Preview experience Off

SystemManaged branch Validate all Save all Publish

Factory Resources

Pipelines 2

- Daily\_Running\_PublicBlob\_Pipeline
- Weekly\_Running\_RDS\_Pipeline

Change Data Capture (preview) 0

Datasets 9

- Archive\_Daily\_Posts
- Archive\_Weekly\_Posttype
- Archive\_Weekly\_Users
- DestinationDS\_Posts\_Datalake
- DestinationDS\_PostType\_Datalake
- DestinationDS\_Users\_Datalake
- Source\_Posts\_PublicBlob\_WCD
- Source\_PostType\_RDS\_postgres
- Source\_Users\_RDS\_postgres

Data flows 0

Power Query 0

Templates 0

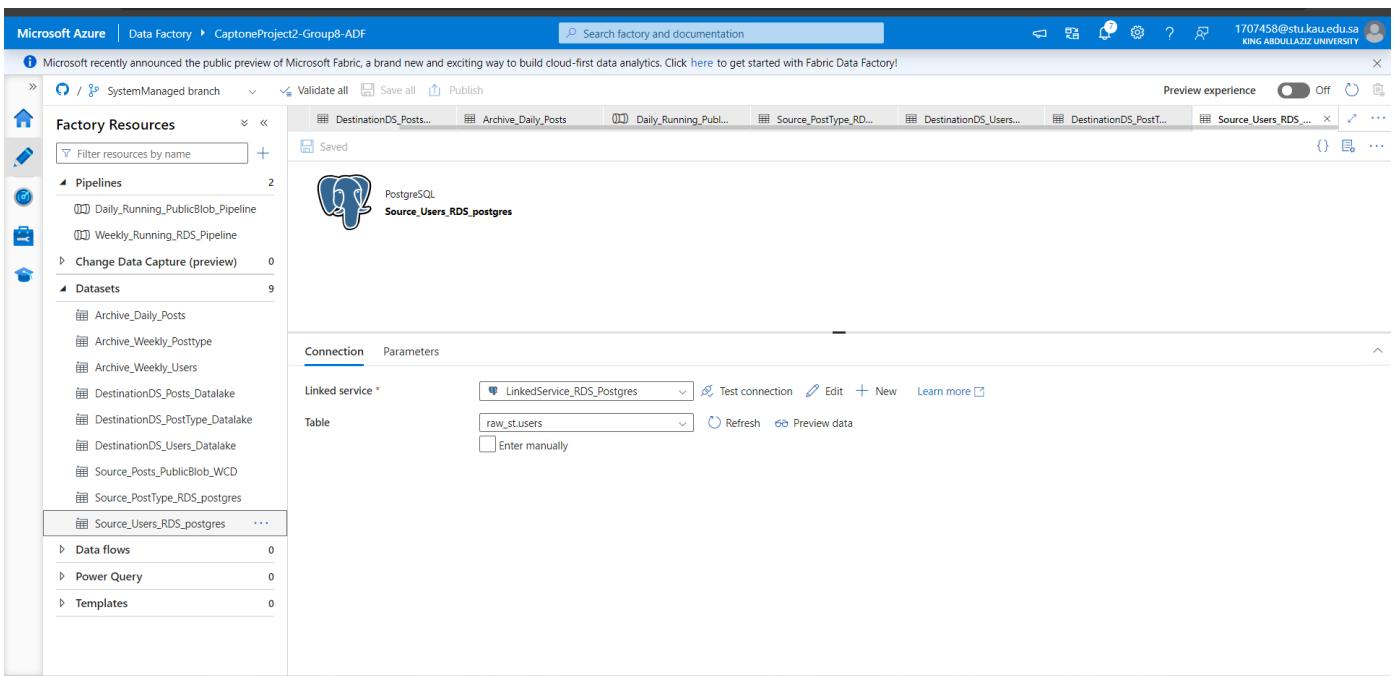
PostgreSQL

Source\_Users\_RDS\_postgres

Connection Parameters

Linked service \*  Test connection Edit New Learn more

Table  Refresh Preview data Enter manually



Microsoft Azure | Data Factory > CapstoneProject2-Group8-ADF

Search factory and documentation

Preview experience Off

SystemManaged branch Validate all Save all Publish

Factory Resources

Pipelines 2

- Daily\_Running\_PublicBlob\_Pipeline
- Weekly\_Running\_RDS\_Pipeline

Change Data Capture (preview) 0

Datasets 9

- Archive\_Daily\_Posts
- Archive\_Weekly\_Posttype
- Archive\_Weekly\_Users
- DestinationDS\_Posts\_Datalake
- DestinationDS\_PostType\_Datalake
- DestinationDS\_Users\_Datalake
- Source\_Posts\_PublicBlob\_WCD
- Source\_PostType\_RDS\_postgres
- Source\_Users\_RDS\_postgres

Data flows 0

Power Query 0

Templates 0

Parquet

DestinationDS\_Posts\_Datalake

Connection Schema Parameters

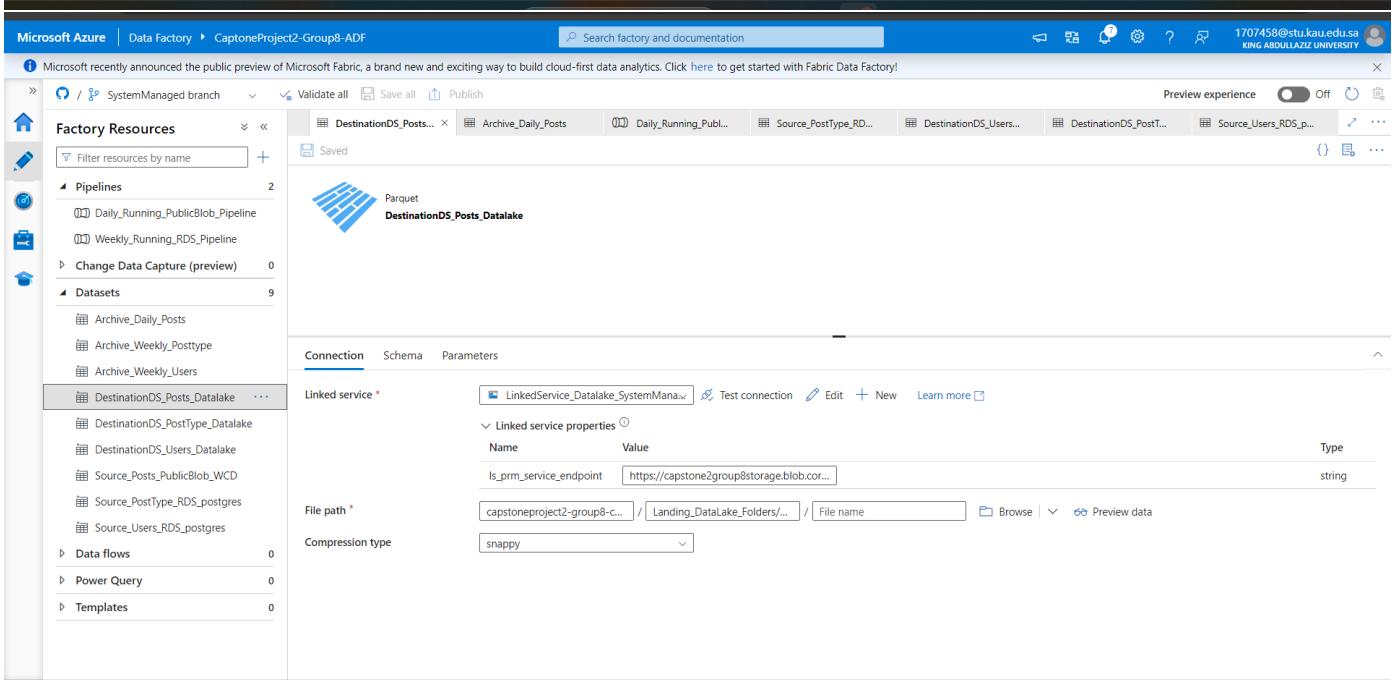
Linked service \*  Test connection Edit New Learn more

Linked service properties

Name	Type
ls_prm_service_endpoint	string

File path \*  Landing\_DataLake\_Folders/... File name Browse Preview data

Compression type snappy



Microsoft Azure | Data Factory > CapstoneProject2-Group8-ADF

Search factory and documentation

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

SystemManaged branch Validate all Save all Publish Preview experience Off

Factory Resources

Pipelines: Daily\_Running\_PublicBlob\_Pipeline, Weekly\_Running\_RDS\_Pipeline  
Change Data Capture (preview): 0  
Datasets: Archive\_Daily\_Posts, Archive\_Weekly\_Posttype, Archive\_Weekly\_Users, DestinationDS\_Posts\_Datalake, DestinationDS\_PostType\_DataLake, DestinationDS\_Users\_Datalake, Source\_Posts\_PublicBlob\_WCD, Source\_PostType\_RDS\_postgres, Source\_Users\_RDS\_postgres  
Data flows: 0  
Power Query: 0  
Templates: 0

Save

DelimitedText CSV DestinationDS\_PostType\_DataLake

Connection Schema Parameters

Linked service \* LinkedService\_Datalake\_SystemManager Test connection Edit + New Learn more Connection successful

Linked service properties

Name	Type
ls_prm_service_endpoint	string

File path \* capstoneproject2-group8-c... / Landing\_DataLake\_Folders... / File name Browse Preview data Detect format

Compression type Select...  
Column delimiter Comma ()  
Row delimiter Default (\r\n, or \n)  
Encoding Default(UTF-8)

Microsoft Azure | Data Factory > CapstoneProject2-Group8-ADF

Search factory and documentation

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

SystemManaged branch Validate all Save all Publish Preview experience Off

Factory Resources

Pipelines: Daily\_Running\_PublicBlob\_Pipeline, Weekly\_Running\_RDS\_Pipeline  
Change Data Capture (preview): 0  
Datasets: Archive\_Daily\_Posts, Archive\_Weekly\_Posttype, Archive\_Weekly\_Users, DestinationDS\_Posts\_Datalake, DestinationDS\_PostType\_DataLake, DestinationDS\_Users\_Datalake, Source\_Posts\_PublicBlob\_WCD, Source\_PostType\_RDS\_postgres, Source\_Users\_RDS\_postgres  
Data flows: 0  
Power Query: 0  
Templates: 0

Save

DelimitedText CSV DestinationDS\_Users\_DataLake

Connection Schema Parameters

Linked service \* LinkedService\_Datalake\_SystemManager Test connection Edit + New Learn more Connection successful

Linked service properties

Name	Type
ls_prm_service_endpoint	string

File path \* capstoneproject2-group8-c... / Landing\_DataLake\_Folders... / File name Browse Preview data Detect format

Compression type Select...  
Column delimiter Comma ()  
Row delimiter Default (\r\n, or \n)  
Encoding Default(UTF-8)

Microsoft Azure | Data Factory > CapstoneProject2-Group8-ADF

Search factory and documentation

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

SystemManaged branch Validate all Save all Publish Preview experience Off

Factory Resources

Pipelines: Daily\_Running\_PublicBlob\_Pipeline, Weekly\_Running\_RDS\_Pipeline, Change Data Capture (preview), Archive\_Daily\_Posts, Archive\_Weekly\_Posttype, Archive\_Weekly\_Users.

Datasets: DestinationDS\_Posts\_Datalake, DestinationDS\_PostType\_Datalake, DestinationDS\_Users\_Datalake, Source\_Posts\_PublicBlob\_WCD, Source\_PostType\_RDS\_postgres, Source\_Users\_RDS\_postgres.

Data flows: 0.

Power Query: 0.

Templates: 0.

Save

DelimitedText Archive\_Weekly\_Posttype

Connection Schema Parameters

Linked service: LinkedService\_Datalake\_SystemManager

Test connection: Connection successful

File path: capstoneproject2-group8-c... / Archive/Archive\_Posttype / File name

Compression type: Select...

Column delimiter: Comma (,)

Row delimiter: Default (\r\n, or \n)

Encoding: Default(UTF-8)

Microsoft Azure | Data Factory > CapstoneProject2-Group8-ADF

Search factory and documentation

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

SystemManaged branch Validate all Save all Publish

Factory Resources

Pipelines: Daily\_Running\_PublicBlob\_Pipeline, Weekly\_Running\_RDS\_Pipeline, Change Data Capture (preview), Archive\_Daily\_Posts, Archive\_Weekly\_Posttype, Archive\_Weekly\_Users.

Datasets: DestinationDS\_Posts\_Datalake, DestinationDS\_PostType\_Datalake, DestinationDS\_Users\_Datalake, Source\_Posts\_PublicBlob\_WCD, Source\_PostType\_RDS\_postgres, Source\_Users\_RDS\_postgres.

Data flows: 0.

Power Query: 0.

Templates: 0.

Save

DelimitedText Archive\_Weekly\_Users

Connection Schema Parameters

Linked service: LinkedService\_Datalake\_SystemManager

Test connection: Connection successful

File path: capstoneproject2-group8-c... / Archive/Archive\_Users / File name

Compression type: Select...

Column delimiter: Comma (,)

Row delimiter: Default (\r\n, or \n)

Encoding: Default(UTF-8)

Successfully saved in the Git repository  
Successfully saved all changes in the Git repository.

## 3- Pipelines:

### 3.1 Daily Pipeline:

Microsoft Azure | Data Factory > CapstoneProject2-Group8-ADF

Search factory and documentation

SystemManaged branch | Validate all | Save all | Publish

Preview experience: Off

1707458@stu.kau.edu.sa KING ABDULLAZIZ UNIVERSITY

Factory Resources

Pipelines

- Daily\_Running\_PublicBlob\_Pipeline
- Weekly\_Running\_RDS\_Pipeline
- Change Data Capture (preview)
- Datasets
- Archive\_Daily\_Posts
- Archive\_Weekly\_Posttype
- Archive\_Weekly\_Users
- DestinationDS\_Posts\_Datalake
- DestinationDS\_PostType\_Datalake
- DestinationDS\_Users\_Datalake
- Source\_Posts\_PublicBlob\_WCD
- Source\_PostType\_RDS\_postgres
- Source\_Users\_RDS\_postgres

Data flows

Power Query

Templates

Source\_Posts\_PublicBlob\_WCD

DestinationDS\_Users...

Archive\_Daily\_Posts

Daily\_Running\_Pu...

Source\_PostType\_RD...

DestinationDS\_PostT...

Archive\_Weekly\_Posttype

Archive\_Weekly\_Users

DestinationDS\_Posts\_Datalake

DestinationDS\_PostType\_Datalake

DestinationDS\_Users\_Datalake

Copy data

Delete

Copy data

Notebook

Run NLP

Convert Posts from Parquet to CSV

Pipeline run ID: 14b6afe5-b210-439b-9022-fdd329d6e64c

Pipeline status: Succeeded

All status

Showing 1 - 5 of 5 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties	Activity run ID
Run NLP	Succeeded	Notebook	5/18/2024, 1:44:51 PM	6m 5s	AutoResolveIntegrator	a9fb52b-ebd0-4de0-a1f6-d93cc4ccc3	a9fb52b-ebd0-4de0-a1f6-d93cc4ccc3
Convert Posts from Parquet t...	Succeeded	Notebook	5/18/2024, 1:44:51 PM	4m 50s	AutoResolveIntegrator	473c9377-21fc-4291-9672-777b34c9e1	473c9377-21fc-4291-9672-777b34c9e1
Copy posts data to Datalake	Succeeded	Copy data	5/18/2024, 1:44:38 PM	13s	AutoResolveIntegrator	fca99929-c0f5-49c3-8f2d-7481d2885b6	fca99929-c0f5-49c3-8f2d-7481d2885b6
Delete posts data from Datal...	Succeeded	Delete	5/18/2024, 1:44:34 PM	3s	AutoResolveIntegrator	2a4cf189-b3f4-43a3-9695-2543020830	2a4cf189-b3f4-43a3-9695-2543020830
Copy Archive Posts	Succeeded	Copy data	5/18/2024, 1:44:20 PM	15s	AutoResolveIntegrator	b97b6994-7c44-4671-ad86-6b4d87c1b	b97b6994-7c44-4671-ad86-6b4d87c1b

Microsoft Azure | Data Factory > CapstoneProject2-Group8-ADF

Search factory and documentation

SystemManaged branch | Validate all | Save all | Publish

Preview experience: Off

1707458@stu.kau.edu.sa KING ABDULLAZIZ UNIVERSITY

Factory Resources

Pipelines

- Daily\_Running\_PublicBlob\_Pipeline
- Weekly\_Running\_RDS\_Pipeline
- Change Data Capture (preview)
- Datasets
- Archive\_Daily\_Posts
- Archive\_Weekly\_Posttype
- Archive\_Weekly\_Users
- DestinationDS\_Posts\_Datalake
- DestinationDS\_PostType\_Datalake
- DestinationDS\_Users\_Datalake
- Source\_Posts\_PublicBlob\_WCD
- Source\_PostType\_RDS\_postgres
- Source\_Users\_RDS\_postgres

Data flows

Power Query

Templates

Source\_Posts\_PublicBlob\_WCD

DestinationDS\_Users...

Archive\_Daily\_Posts

Daily\_Running\_Pu...

Source\_PostType\_RD...

DestinationDS\_PostT...

Archive\_Weekly\_Posttype

Archive\_Weekly\_Users

DestinationDS\_Posts\_Datalake

DestinationDS\_PostType\_Datalake

DestinationDS\_Users\_Datalake

Copy data

Delete

Copy data

Notebook

Run NLP

Convert Posts from Parquet to CSV

General

Name: Copy Archive Posts

Description: It is copy data from the posts folder in datalake to store it into archive posts folder. That is help to have historical data.

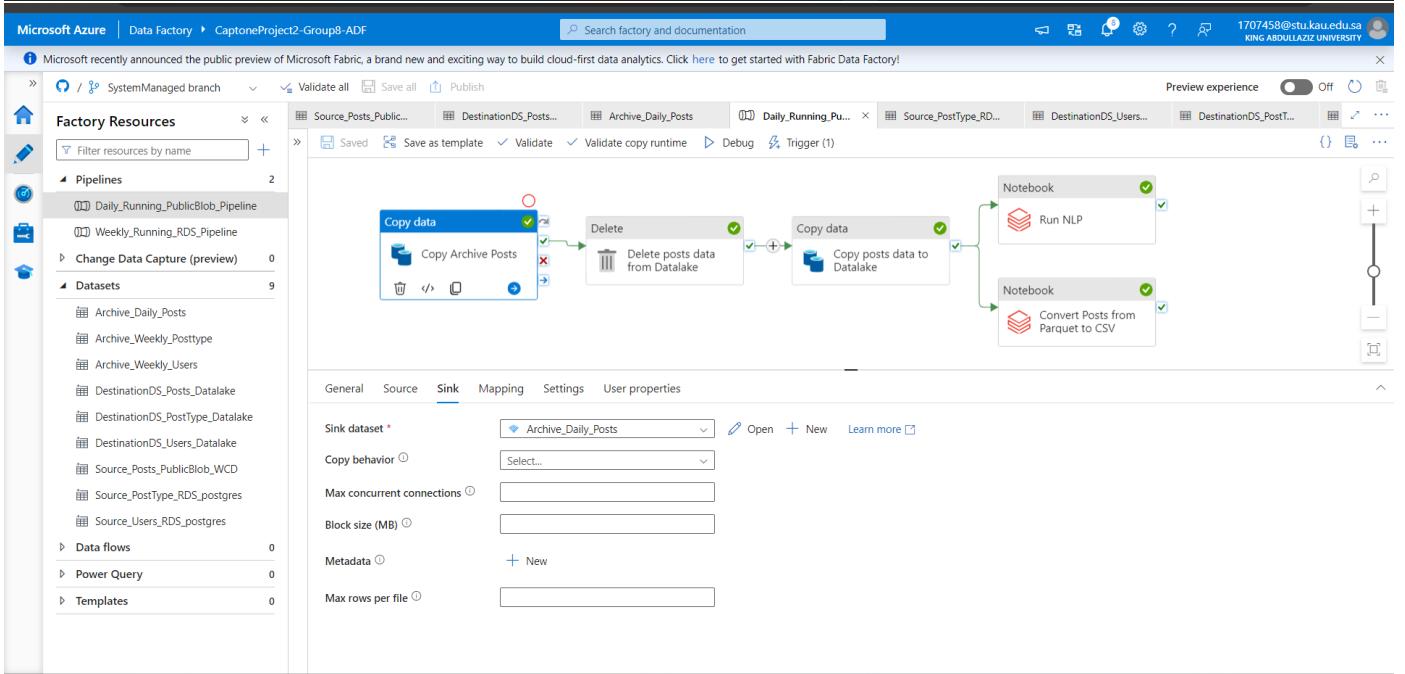
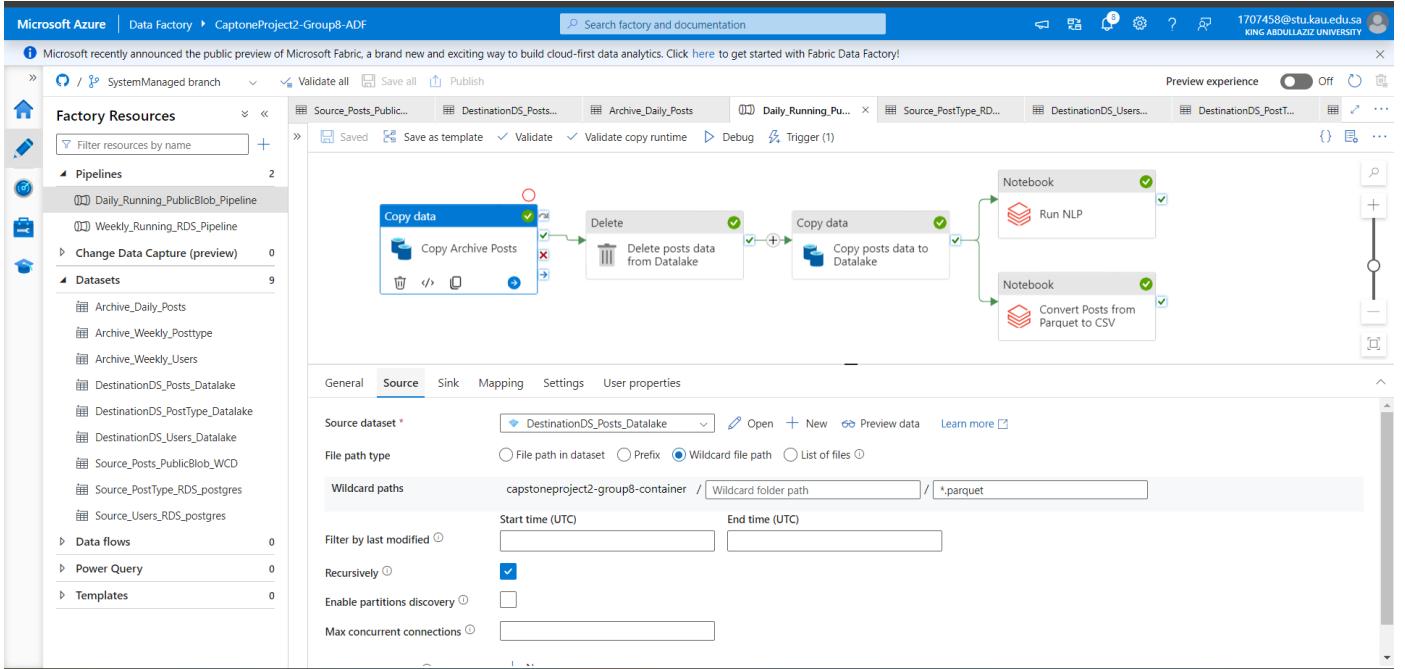
Activity state: Activated

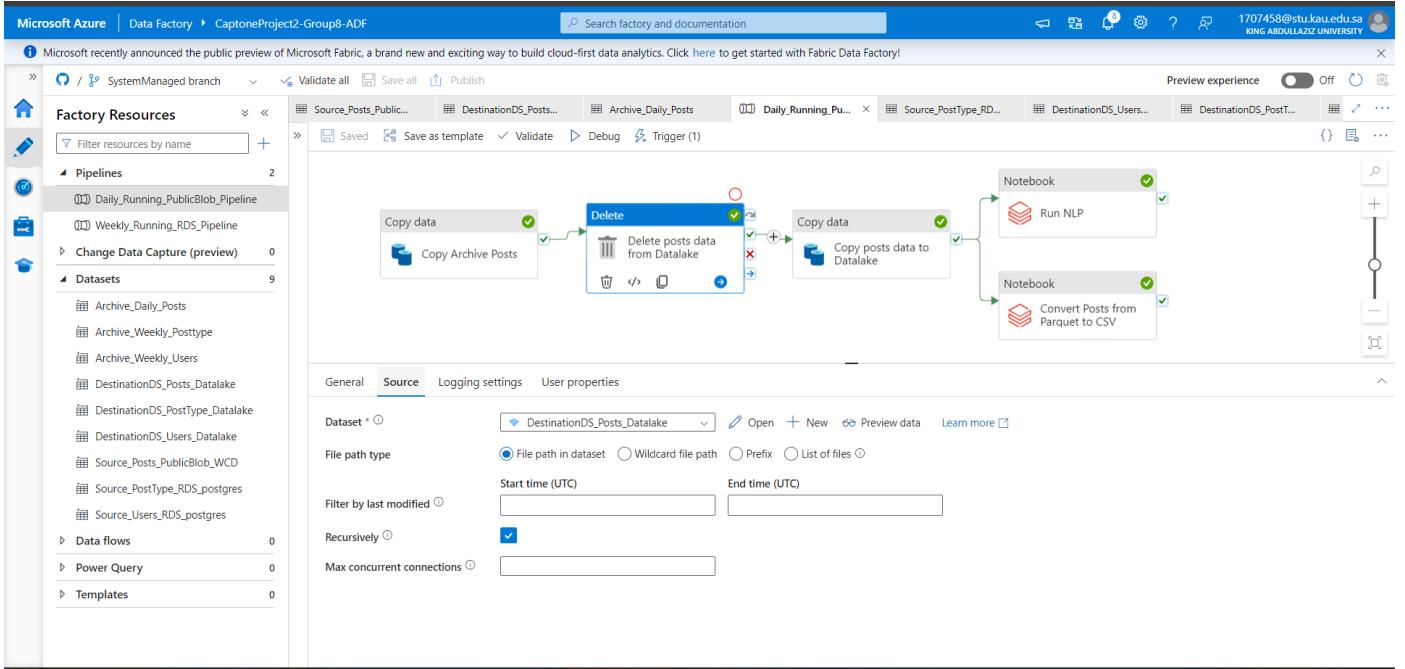
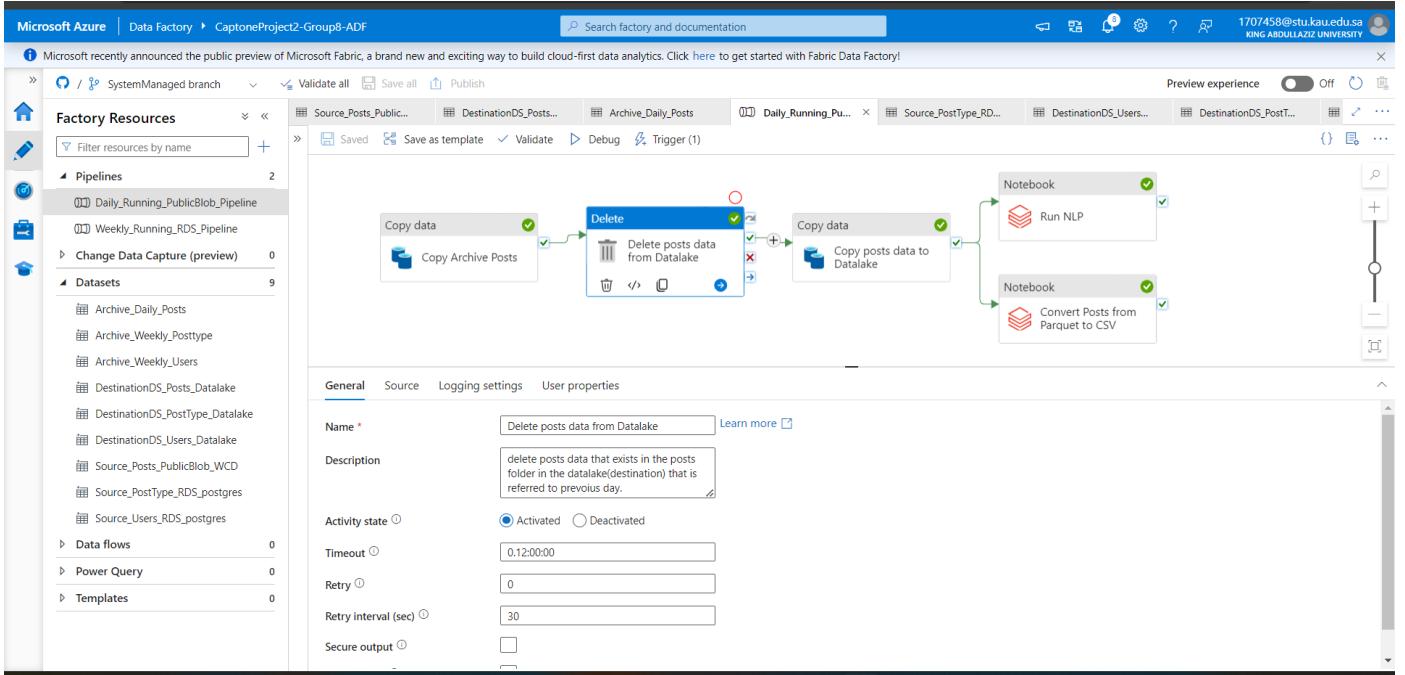
Timeout: 0:12:00

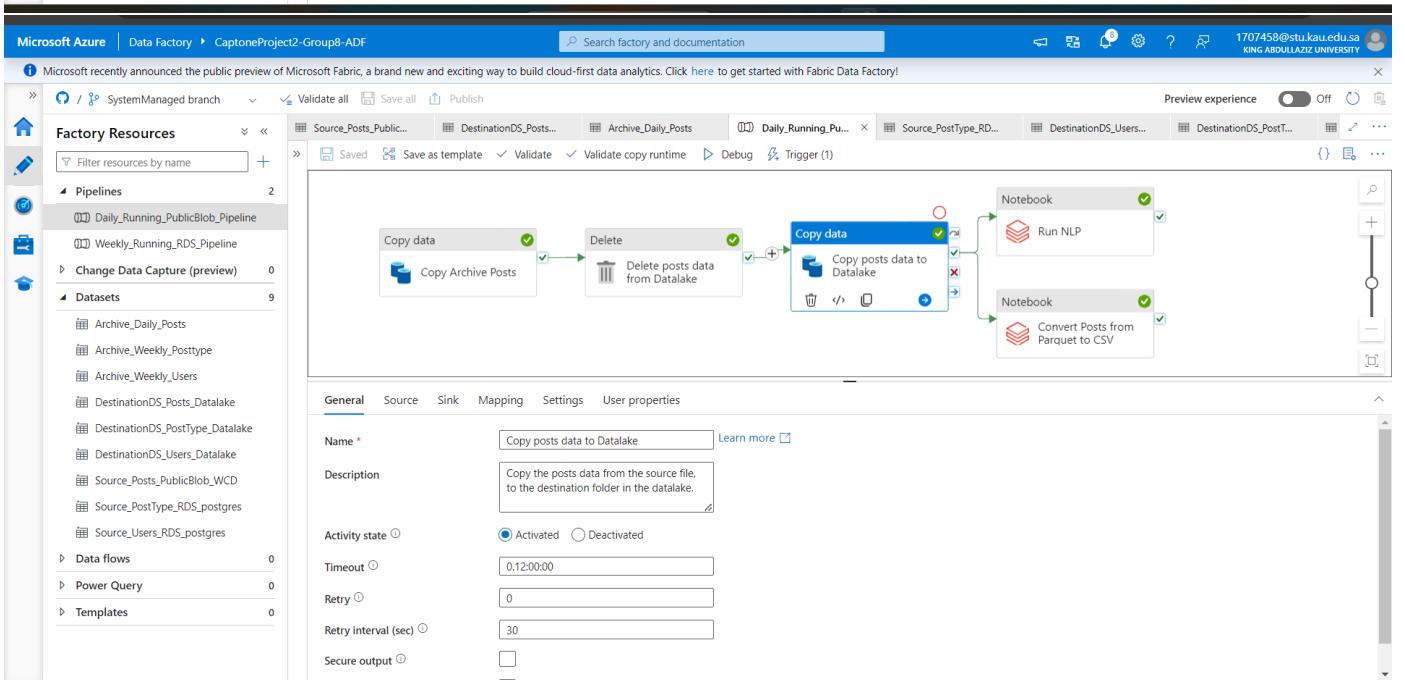
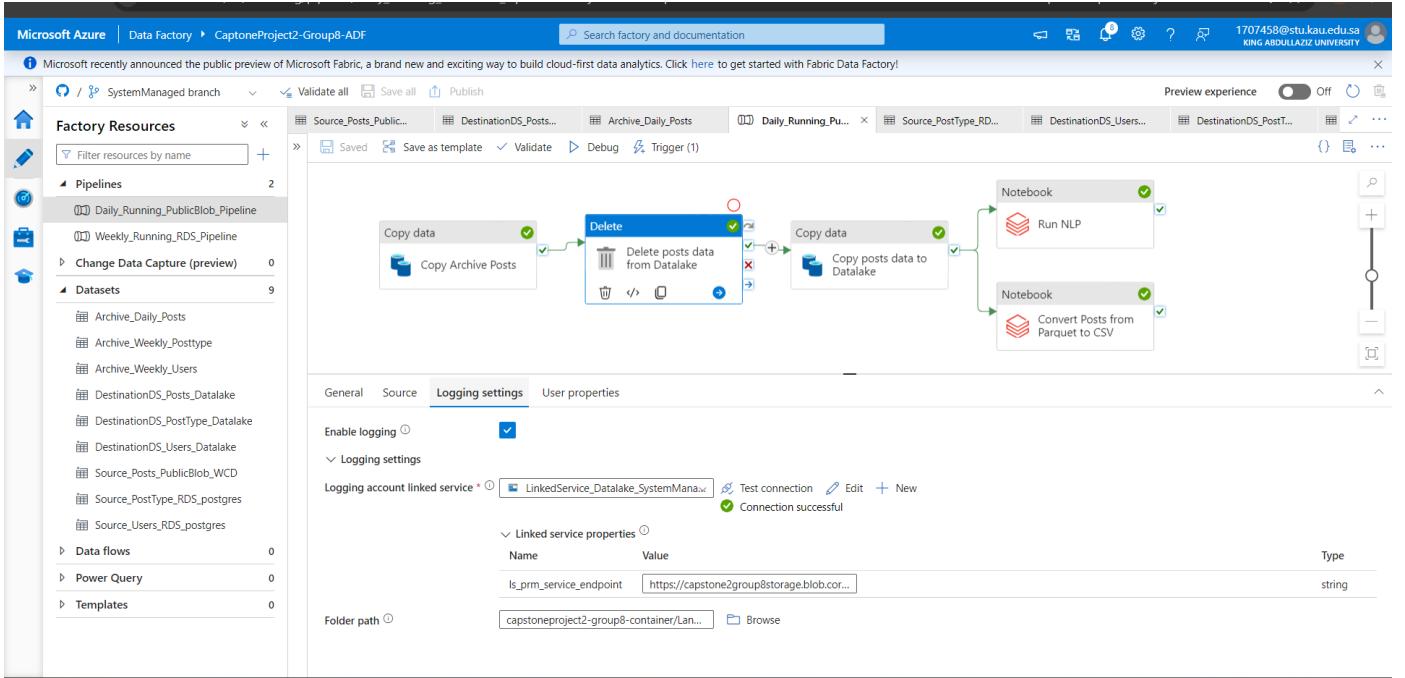
Retry: 0

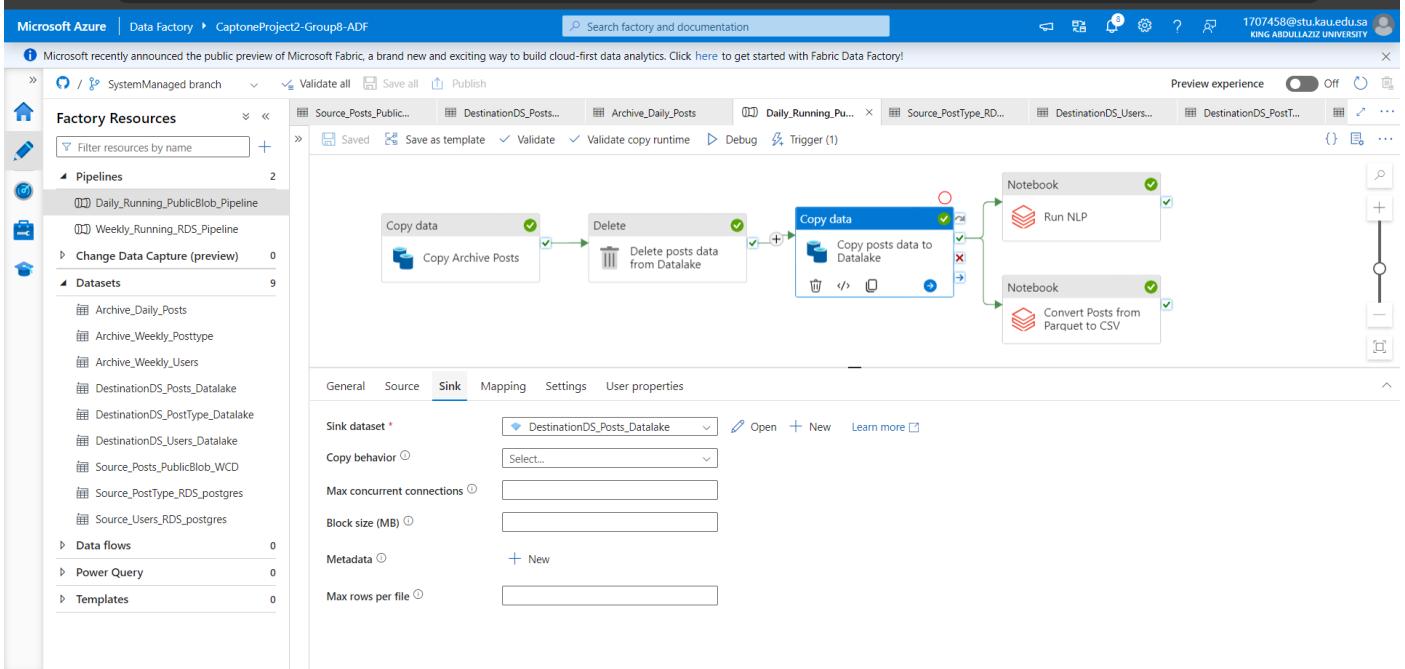
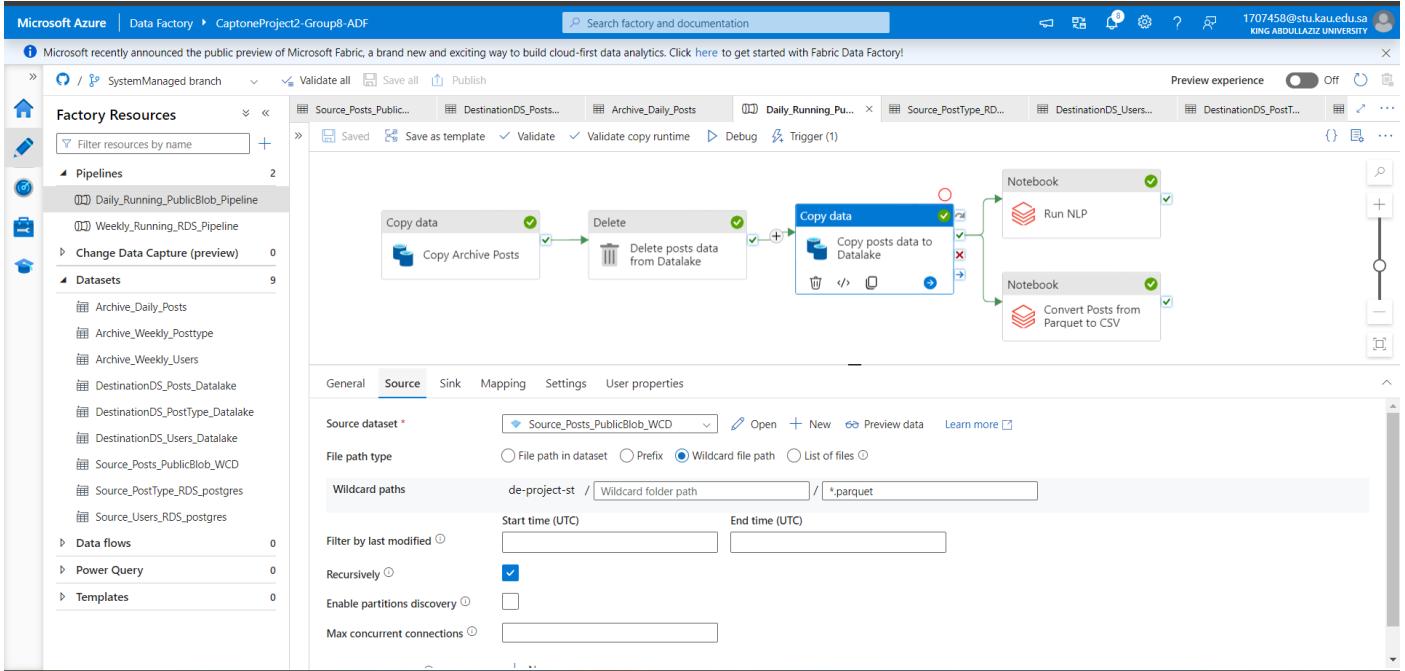
Retry interval (sec): 30

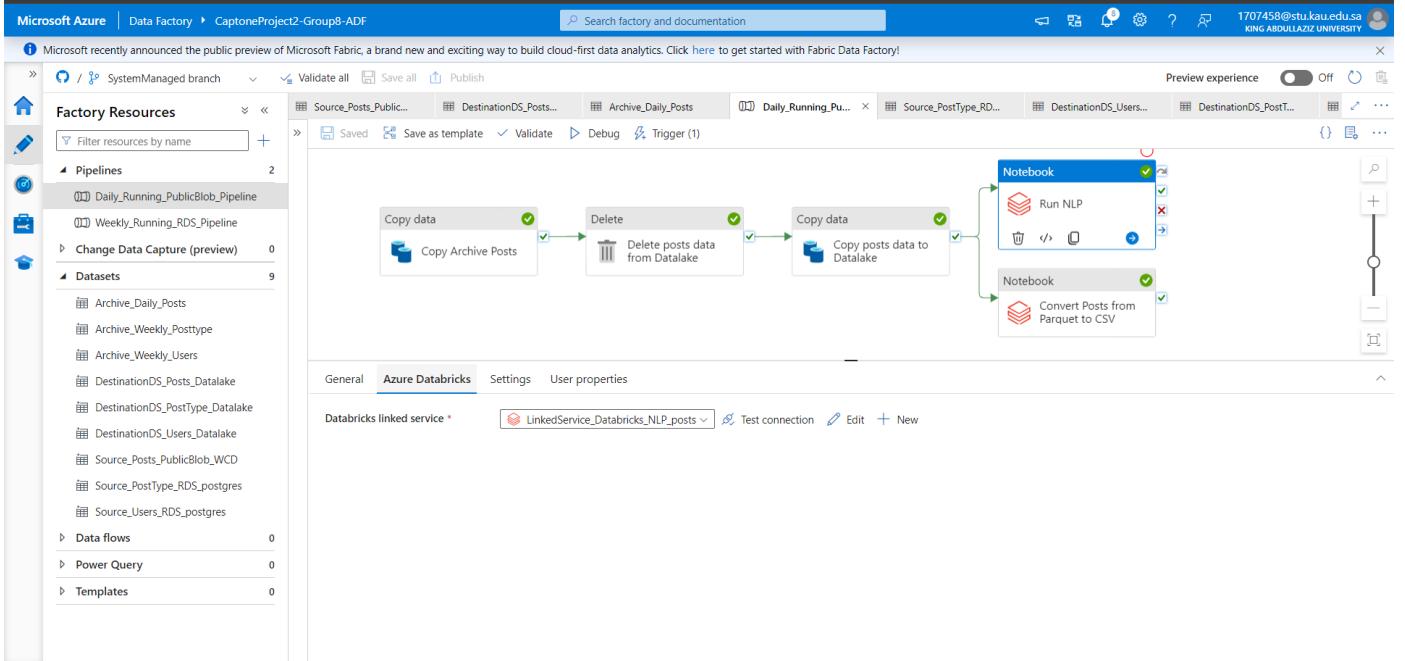
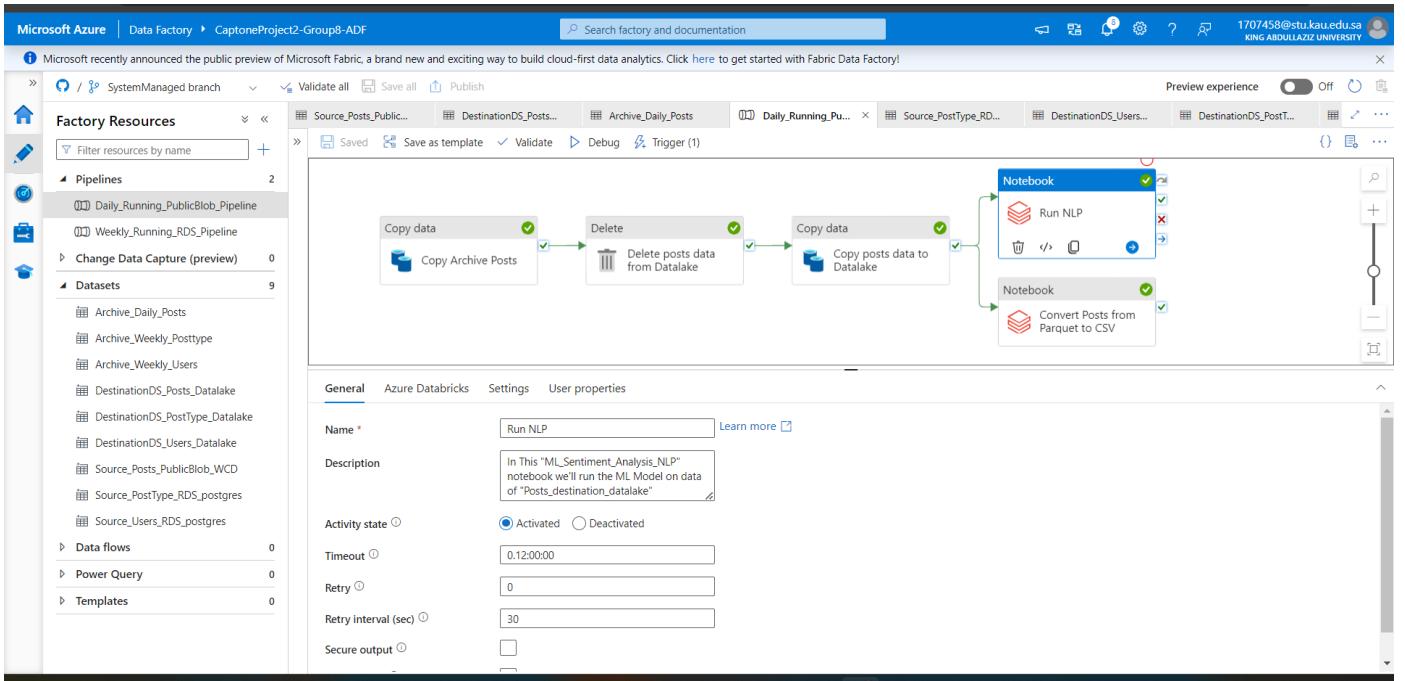
Secure output:











**Microsoft Azure | Data Factory > CaptoneProject2-Group8-ADF**

Search factory and documentation

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Preview experience: Off

**Factory Resources**

- Pipelines (2)
  - Daily\_Running\_PublicBlob\_Pipeline
  - Weekly\_Running\_RDS\_Pipeline
- Change Data Capture (preview) (0)
- Datasets (9)
  - Archive\_Daily\_Posts
  - Archive\_Weekly\_Posttype
  - Archive\_Weekly\_Users
  - DestinationDS\_Posts\_Datalake
  - DestinationDS\_PostType\_Datalake
  - DestinationDS\_Users\_Datalake
  - Source\_Posts\_PublicBlob\_WCD
  - Source\_PostType\_RDS\_postgres
  - Source\_Users\_RDS\_postgres
- Data flows (0)
- Power Query (0)
- Templates (0)

**Pipeline Details: Daily\_Running\_PublicBlob\_Pipeline**

General | Azure Databricks | **Settings** | User properties

Notebook path: du.sa/k2\_G8-ML\_Sentiment\_Analysis\_NLP

Base parameters:

Append libraries:

```

graph LR
    A[Copy data: Copy Archive Posts] --> B[Delete: Delete posts data from Datalake]
    B --> C[Copy data: Copy posts data to Datalake]
    C --> D[Notebook: Run NLP]
    C --> E[Notebook: Convert Posts from Parquet to CSV]
  
```

**Activity Settings: Convert Posts from Parquet to CSV**

Name: Convert Posts from Parquet to CSV

Description: Here we will Convert Posts from Parquet to CSV to use it in the Visualization

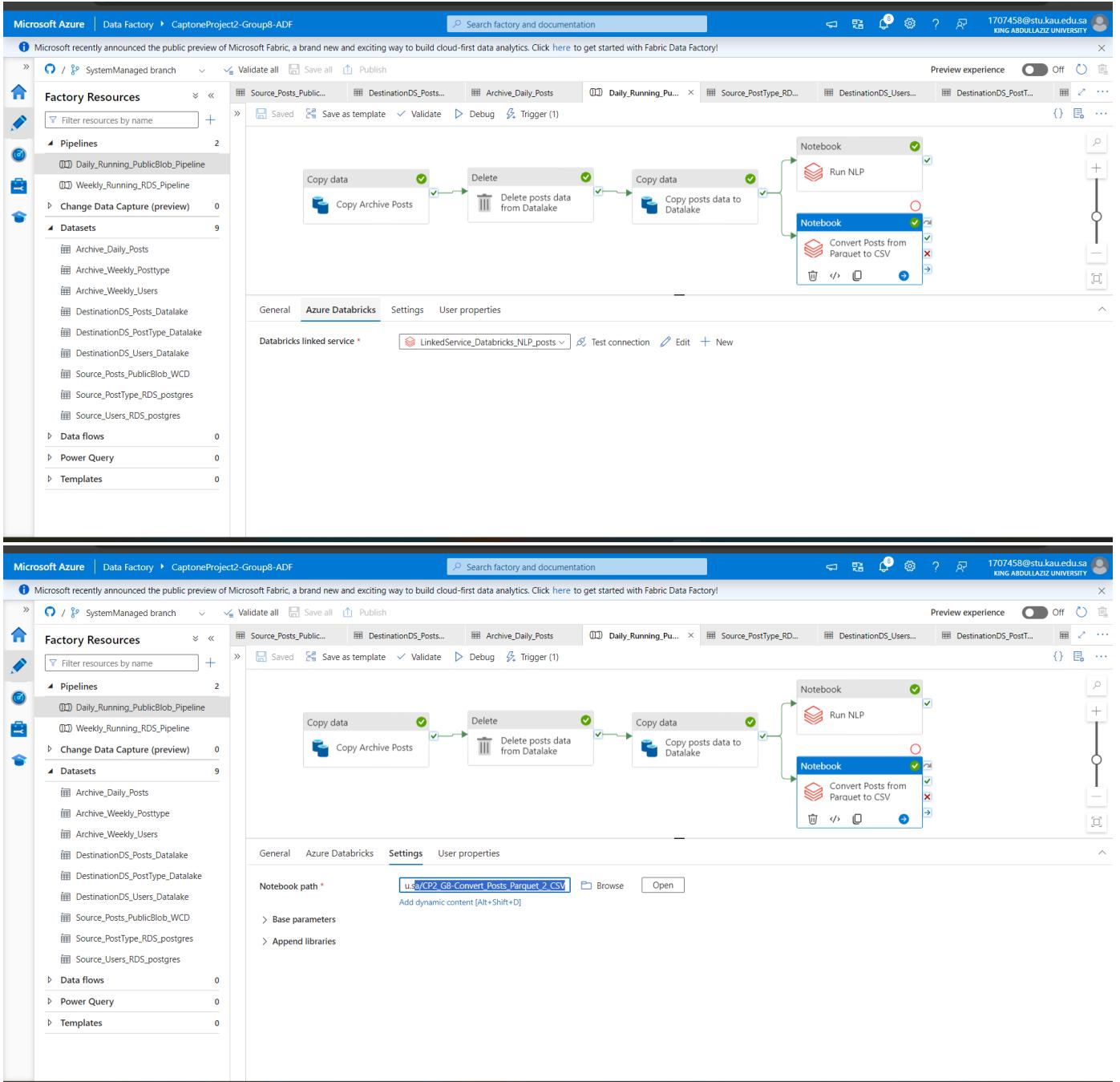
Activity state: Activated

Timeout: 0.120000

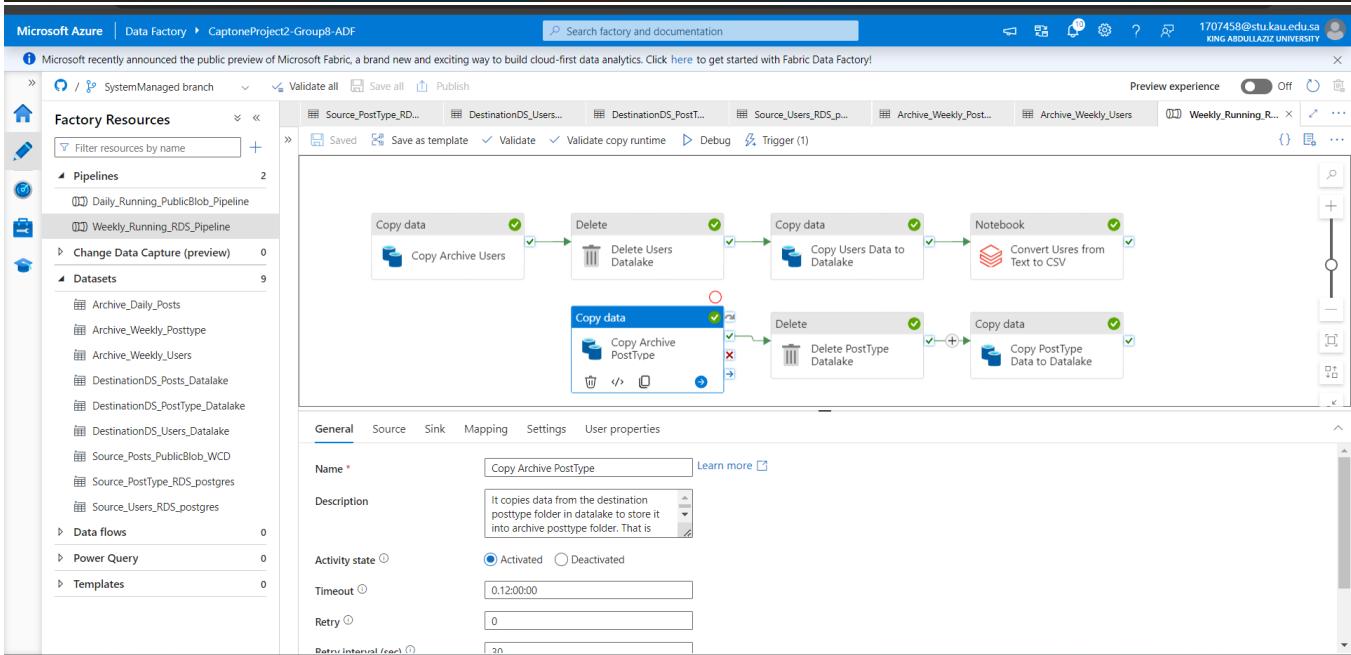
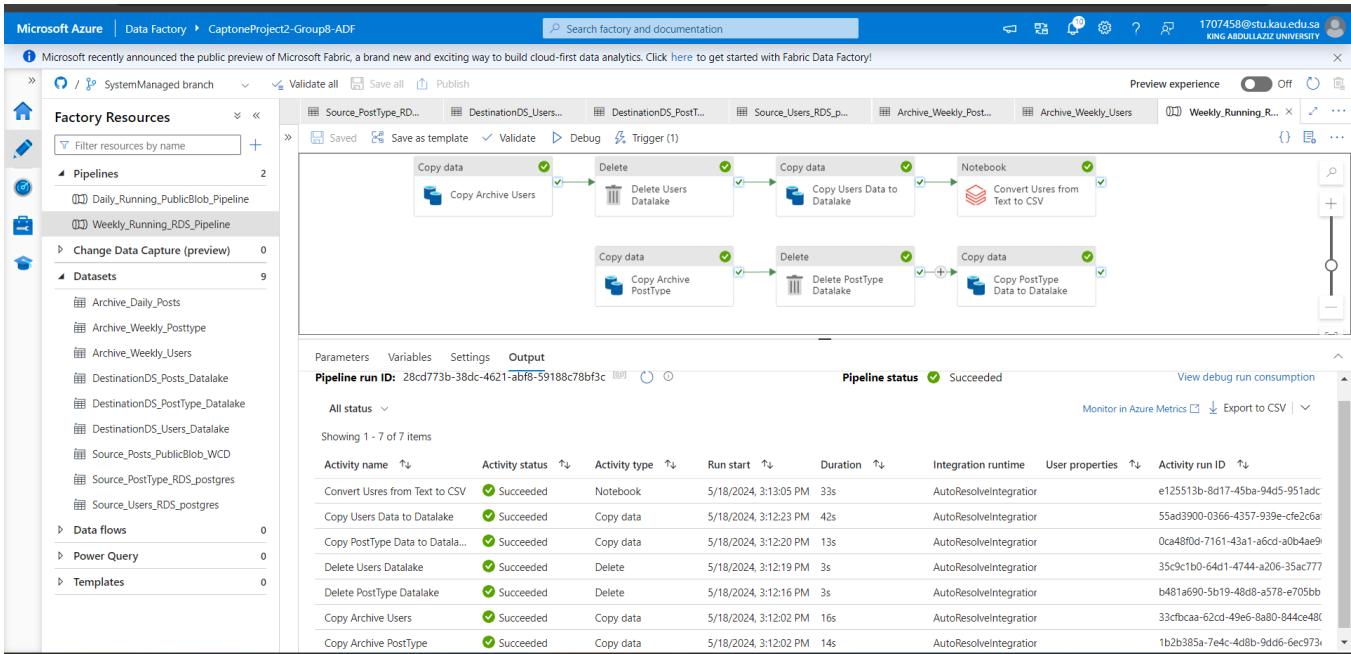
Retry: 0

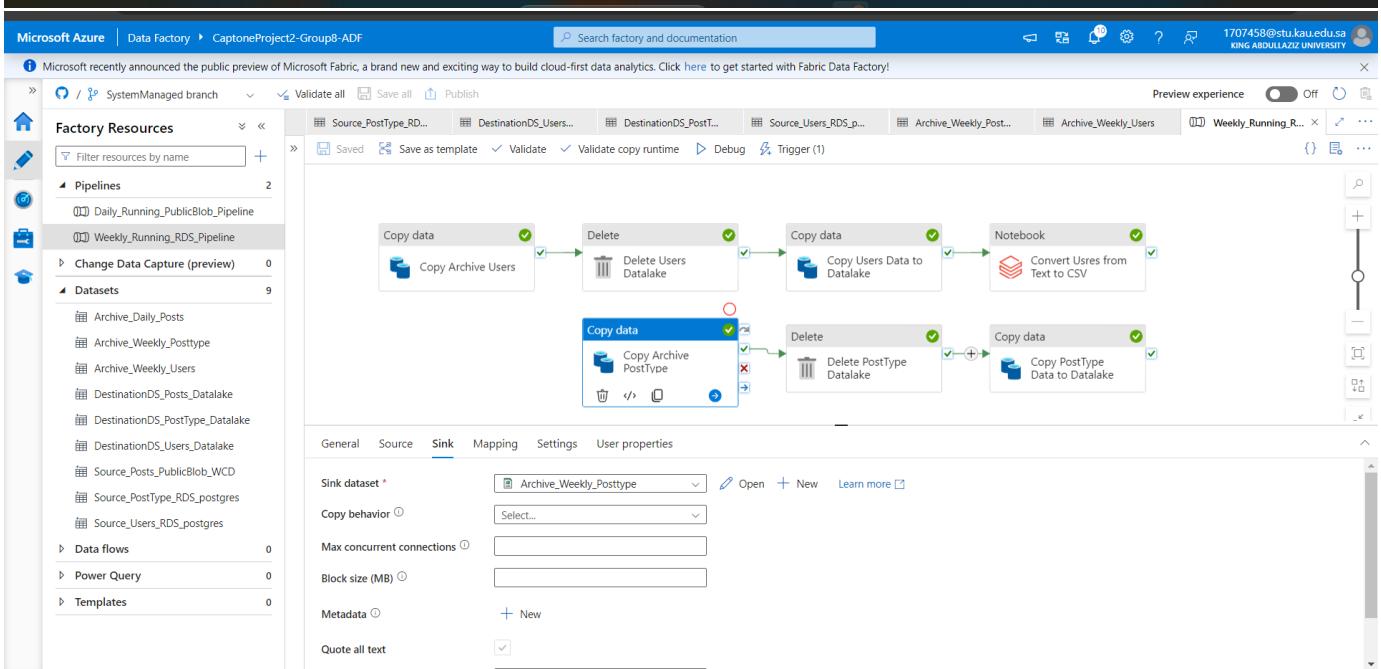
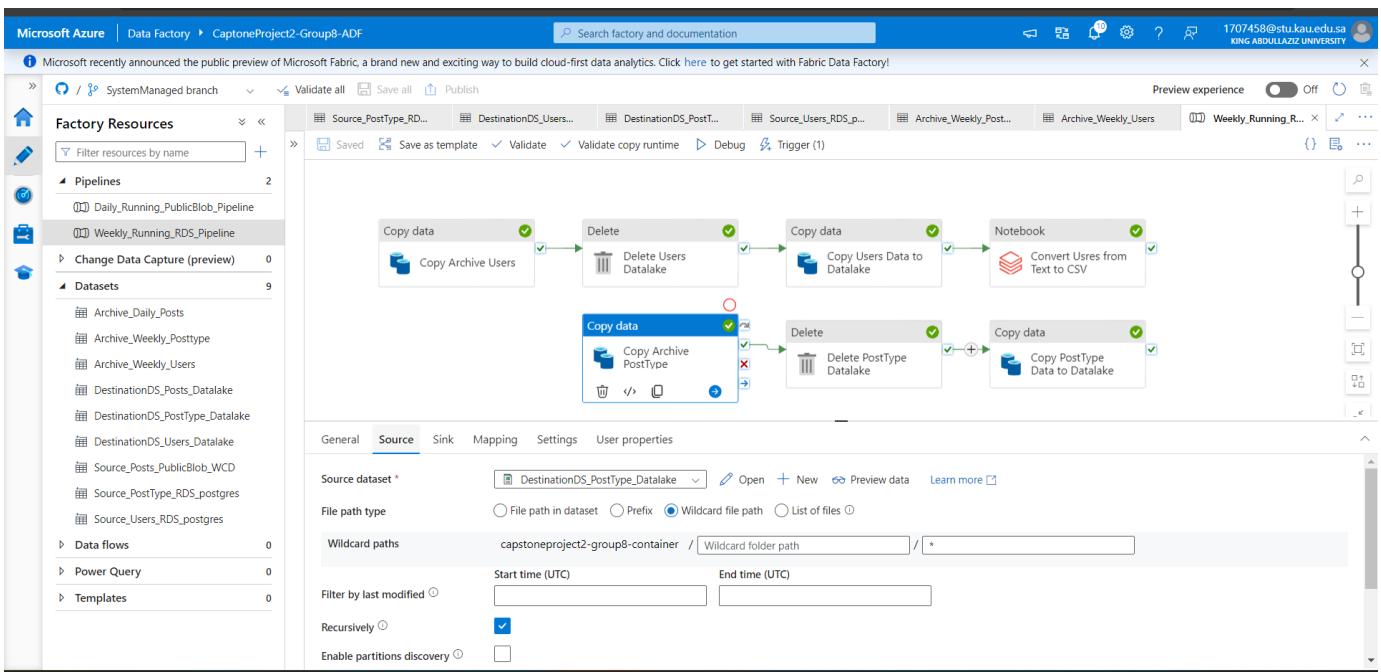
Retry interval (sec): 30

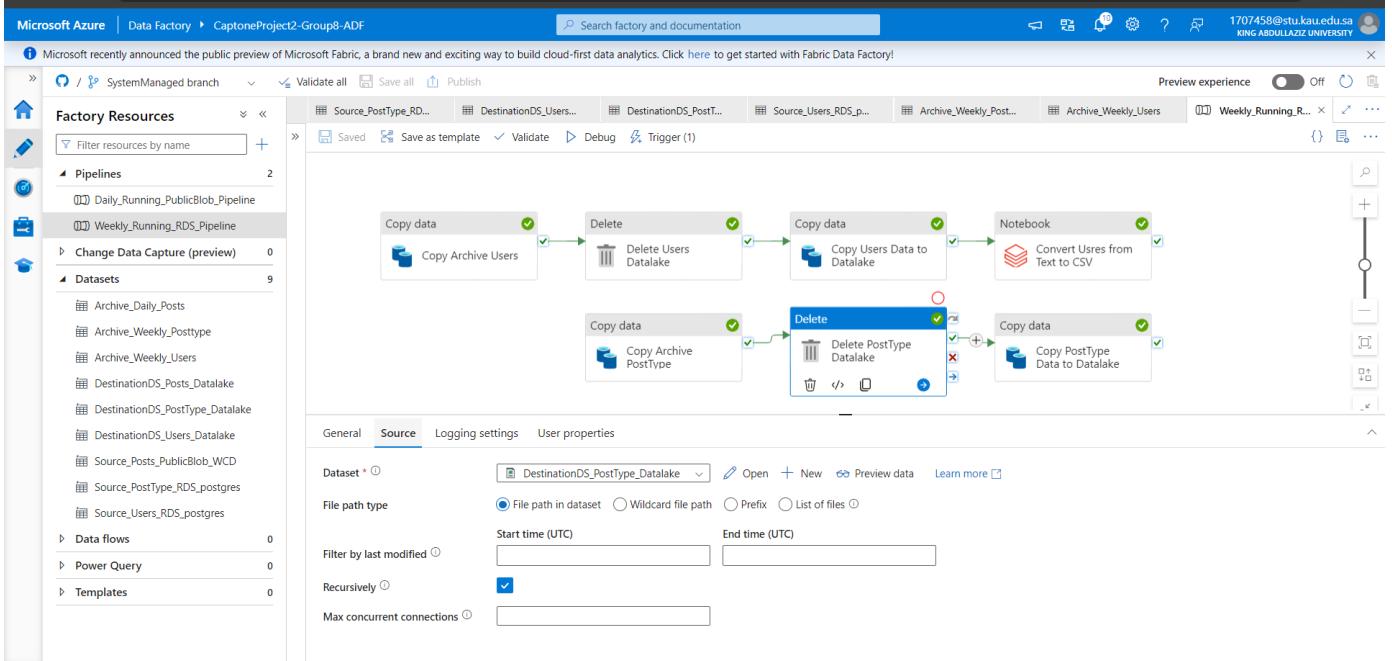
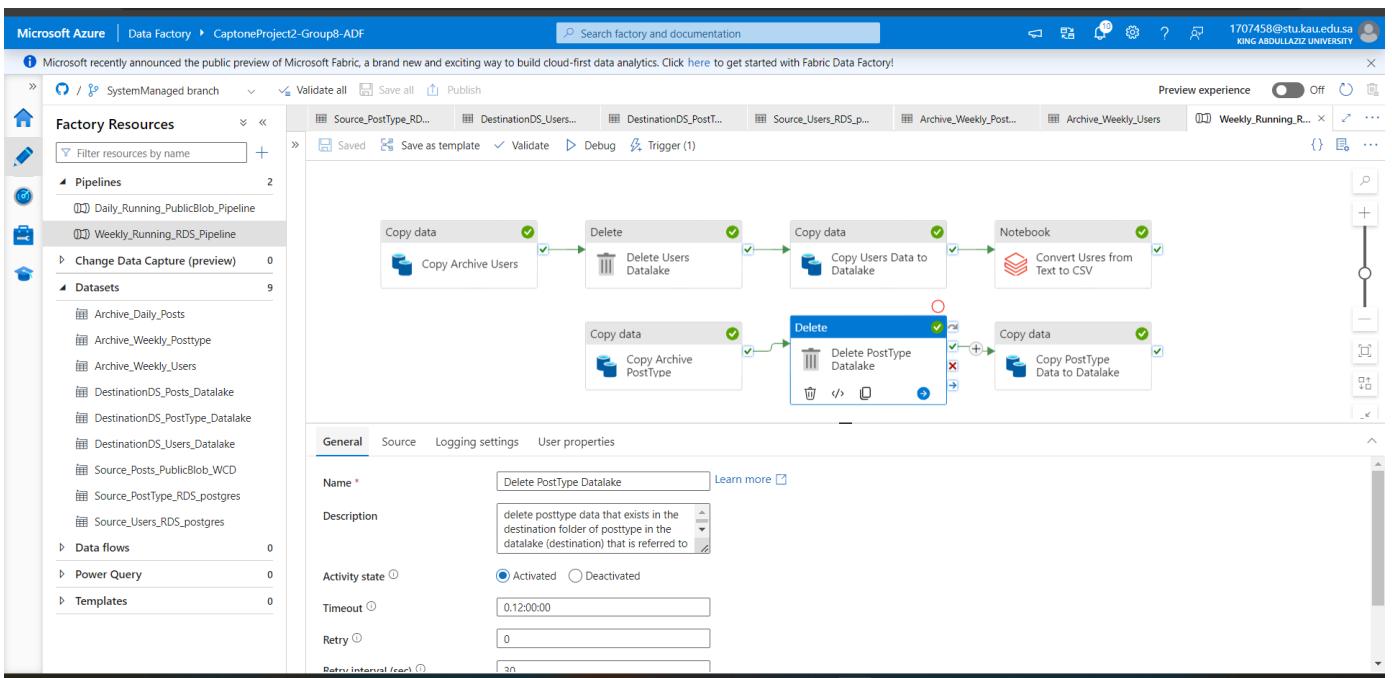
Secure output:

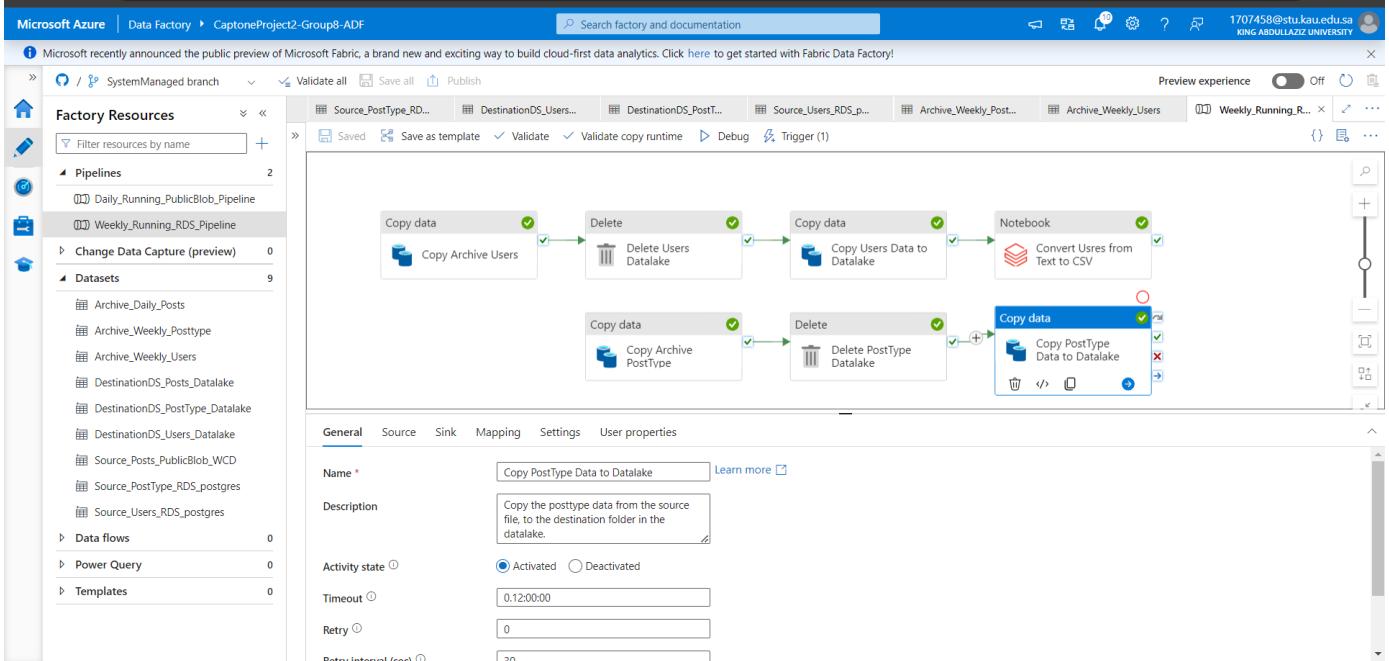
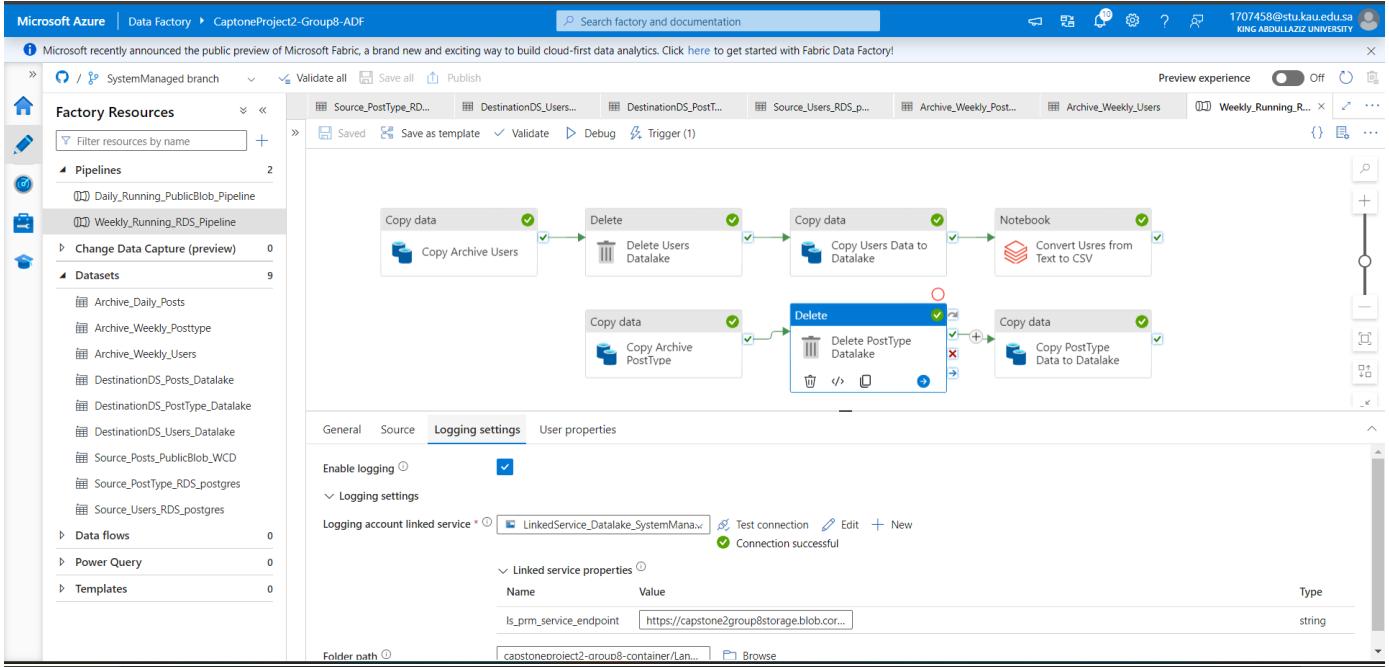


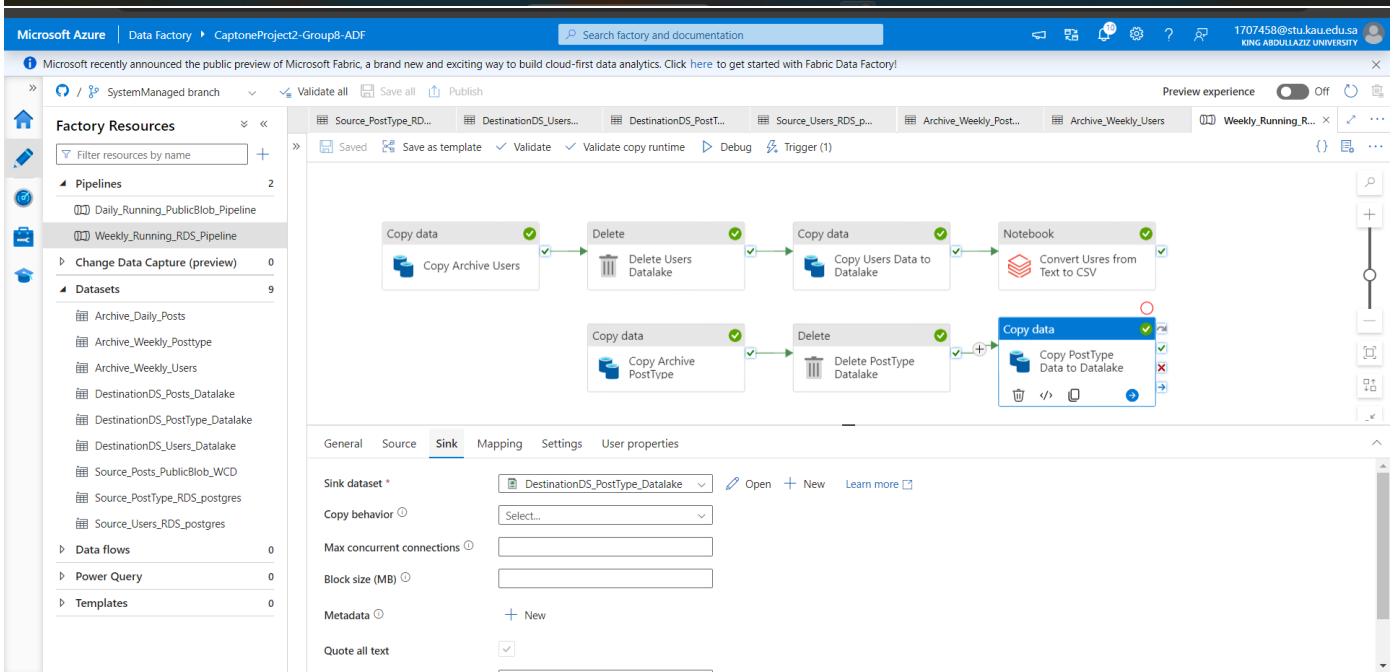
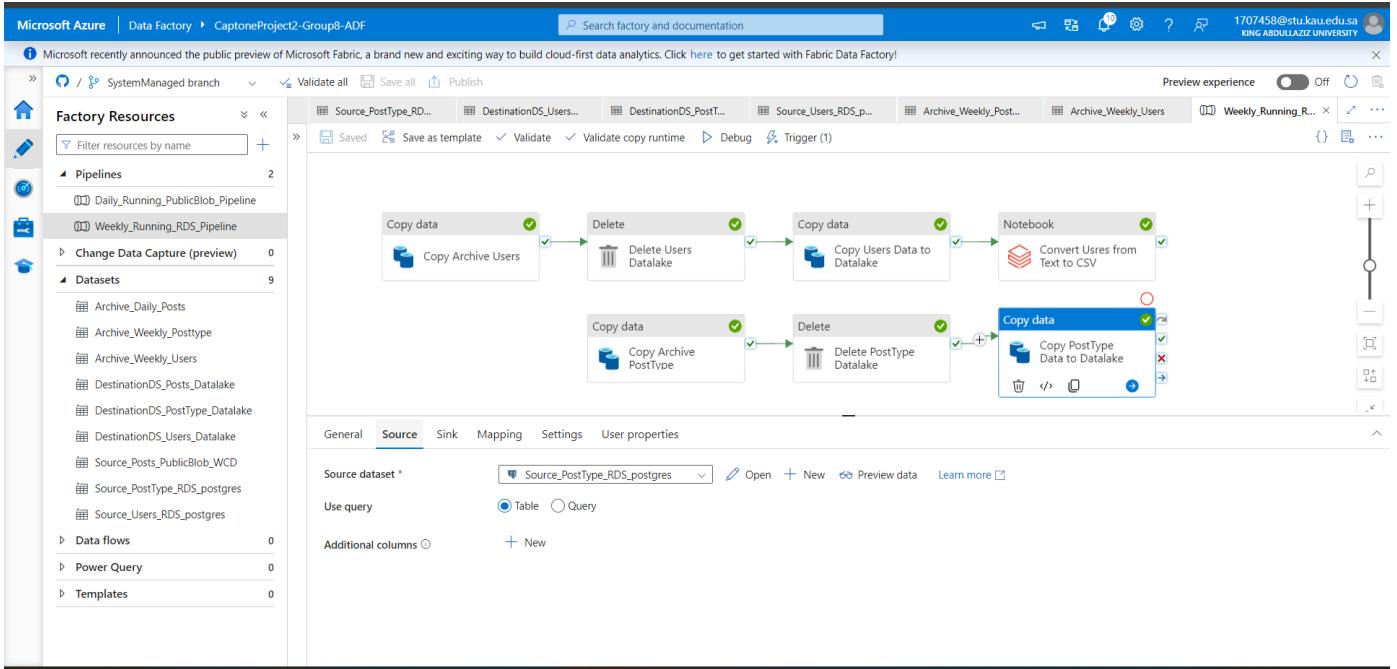
## 3.2 Weekly Pipeline:











**Microsoft Azure | Data Factory > CapstoneProject2-Group8-ADF**

Search factory and documentation

Preview experience: Off

**Factory Resources**

- Pipelines (2)
  - Daily\_Running\_PublicBlob\_Pipeline
  - Weekly\_Running\_RDS\_Pipeline
- Change Data Capture (preview) (0)
- Datasets (9)
  - Archive\_Daily\_Posts
  - Archive\_Weekly\_Posttype
  - Archive\_Weekly\_Users
  - DestinationDS\_Posts\_Datalake
  - DestinationDS\_PostType\_Datalake
  - DestinationDS\_Users\_Datalake
  - Source\_Posts\_PublicBlob\_WCD
  - Source\_PostType\_RDS\_postgres
  - Source\_Users\_RDS\_postgres
- Data flows (0)
- Power Query (0)
- Templates (0)

**Pipeline Details: Weekly\_Running\_RDS\_Pipeline**

**General Tab:**

- Name: Copy Archive Users
- Description: It copies data from the destination users folder in datalake to store it into archive users folder. That is helpful to.
- Activity state: Activated
- Timeout: 0:12:00:00
- Retry: 0
- Batch interval (min): 30

**Source Tab:**

Source dataset: DestinationDS\_Users\_Datalake

File path type: Wildcard file path

Wildcard paths: capstoneproject2-group8-container / Wildcard folder path / \*

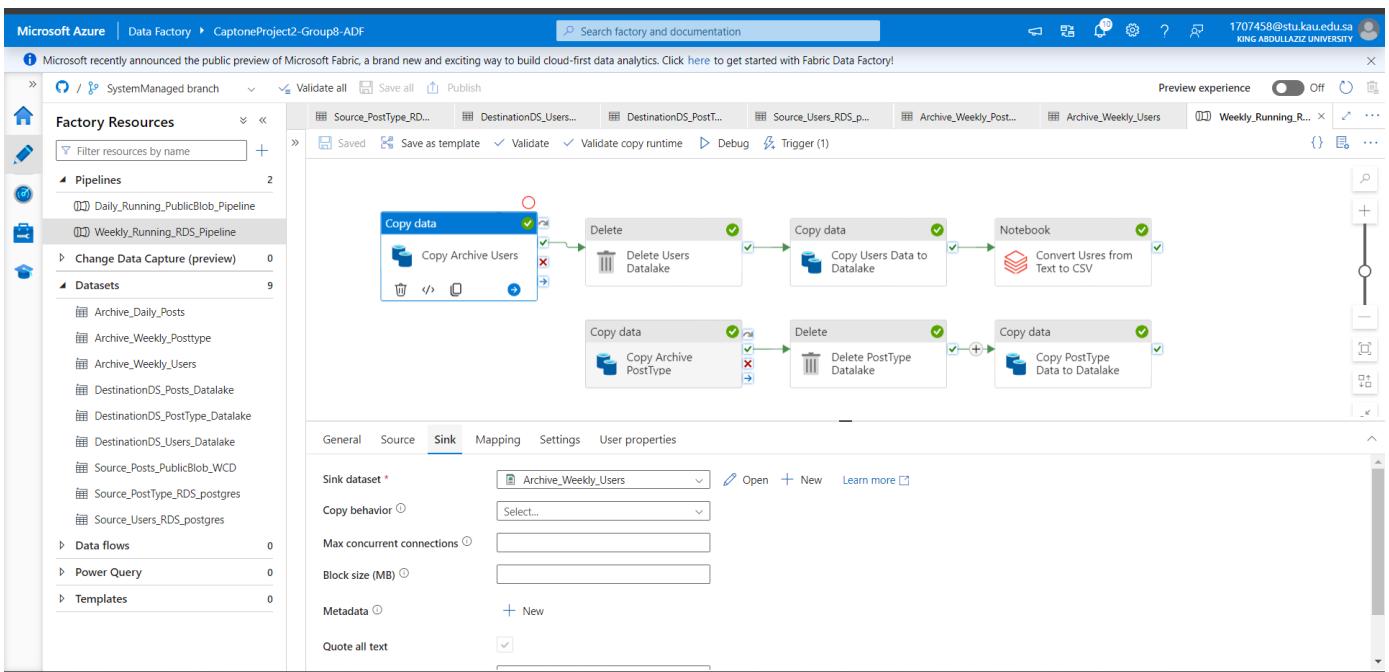
Start time (UTC): \_\_\_\_\_

End time (UTC): \_\_\_\_\_

Filter by last modified: \_\_\_\_\_

Recursively:

Enable partitions discovery:



Microsoft Azure | Data Factory > CaptoneProject2-Group8-ADF

Search factory and documentation

Preview experience Off

SystemManaged branch

Validate all Save as template Validate Debug Trigger (1)

Factory Resources

- Pipelines
  - Daily\_Running\_PublicBlob\_Pipeline
  - Weekly\_Running\_RDS\_Pipeline
- Change Data Capture (preview) 0
- Datasets 9
  - Archive\_Daily\_Posts
  - Archive\_Weekly\_Posttype
  - Archive\_Weekly\_Users
  - DestinationDS\_Posts\_Datalake
  - DestinationDS\_PostType\_Datalake
  - DestinationDS\_Users\_Datalake
  - Source\_Posts\_PublicBlob\_WCD
  - Source\_PostType\_RDS\_postgres
  - Source\_Users\_RDS\_postgres
- Data flows 0
- Power Query 0
- Templates 0

Source\_PostType\_RDS... DestinationDS\_Users... DestinationDS\_PostT... Source\_Users\_RDS\_p... Archive\_Weekly\_Post... Archive\_Weekly\_Users Weekly\_Running.R... Trigger (1)

General

Name: Delete Users Datalake

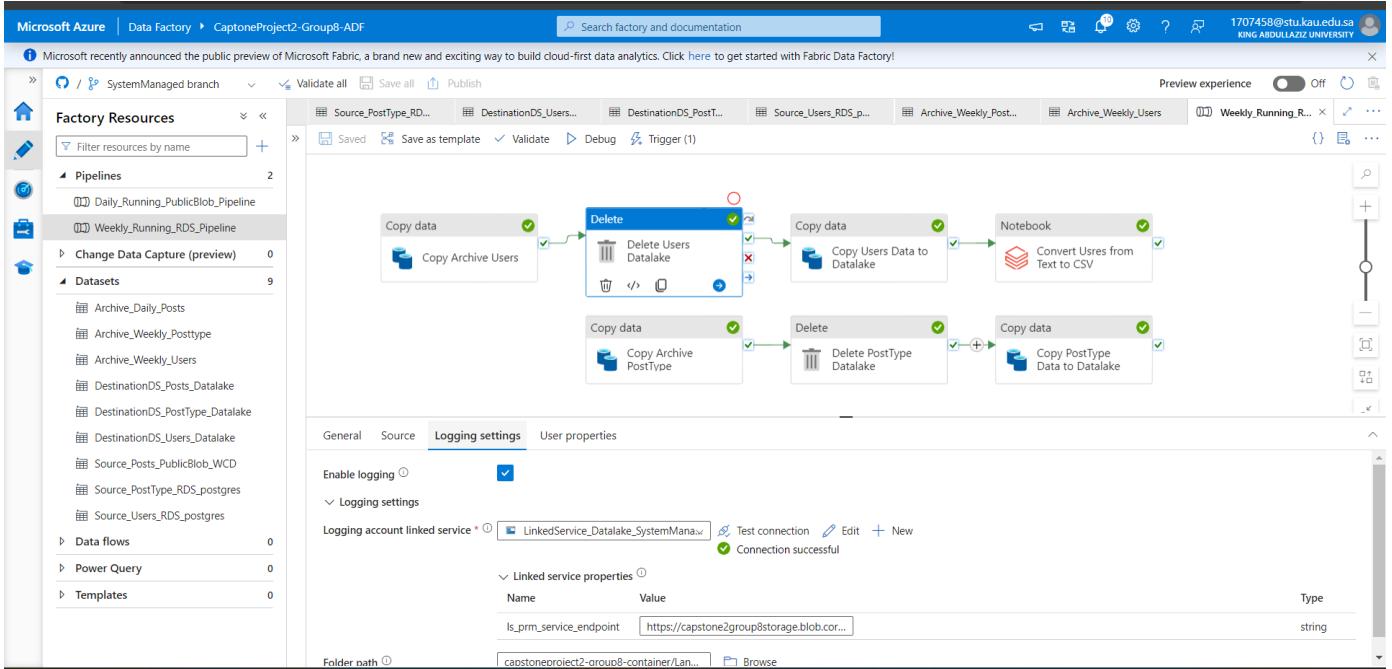
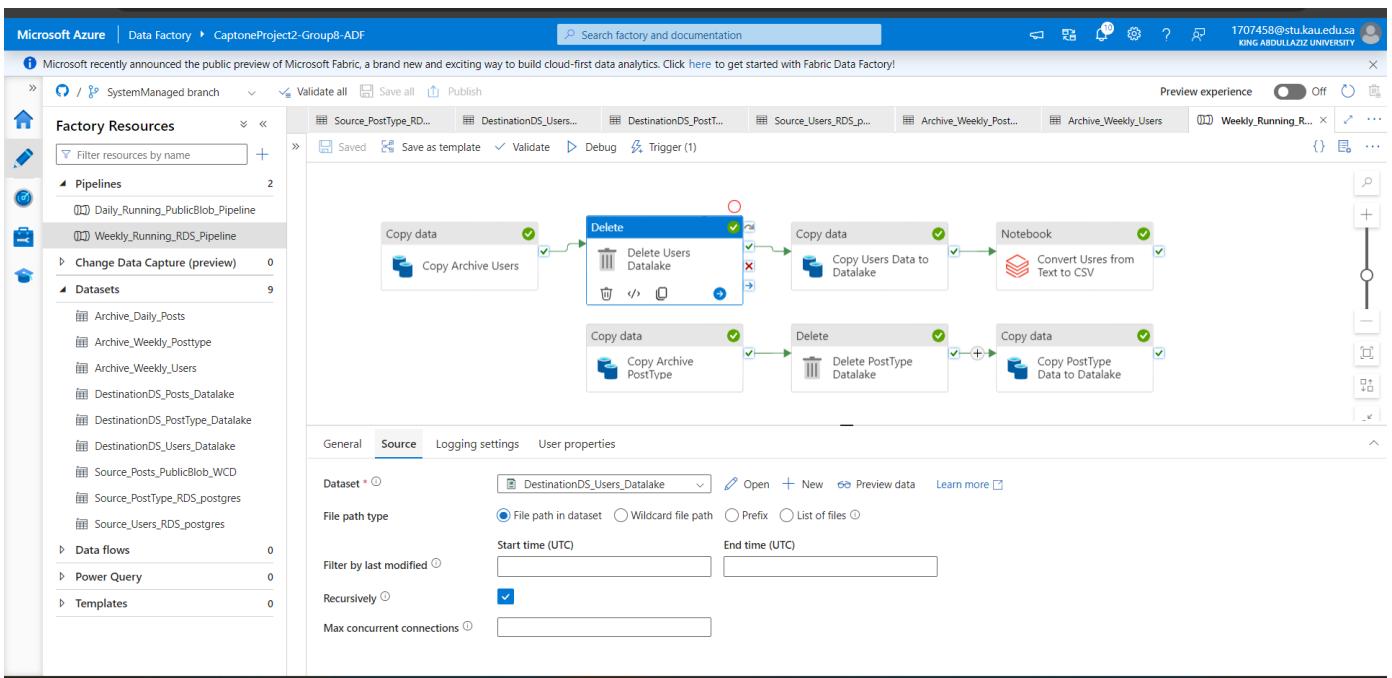
Description: delete users' data that exists in the destination folder of users in the datalake (destination) that is referred to

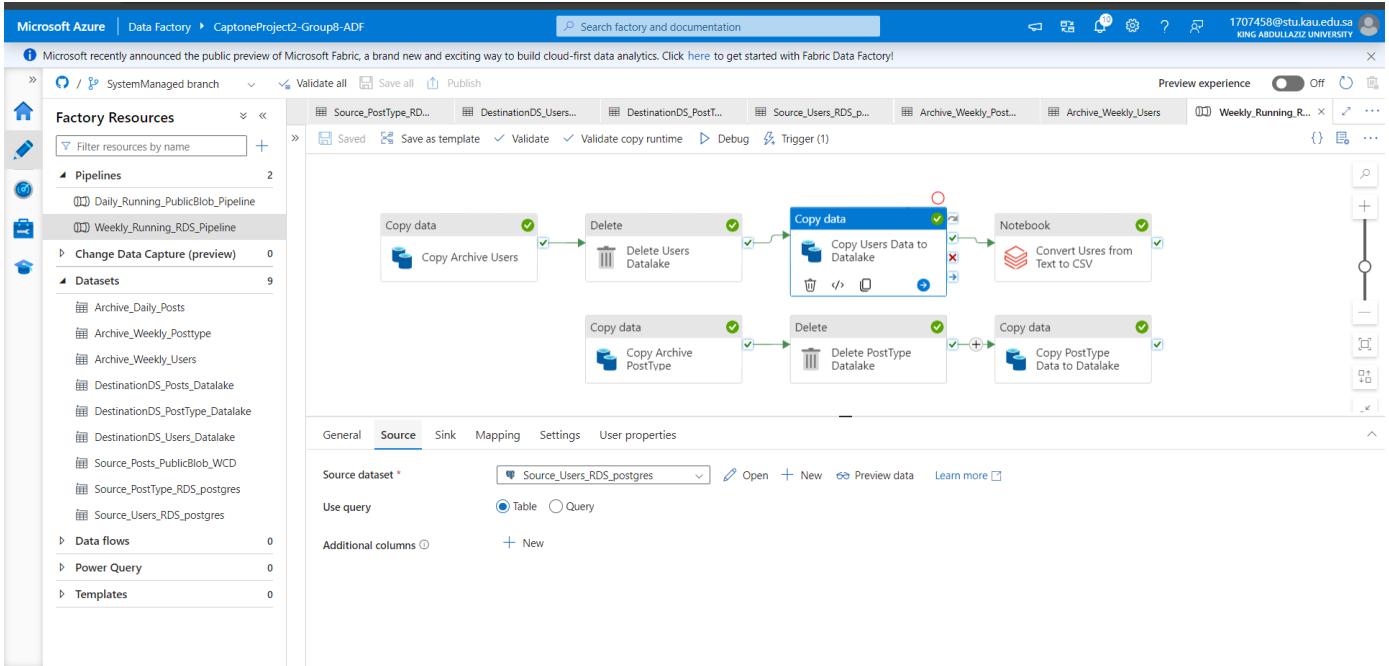
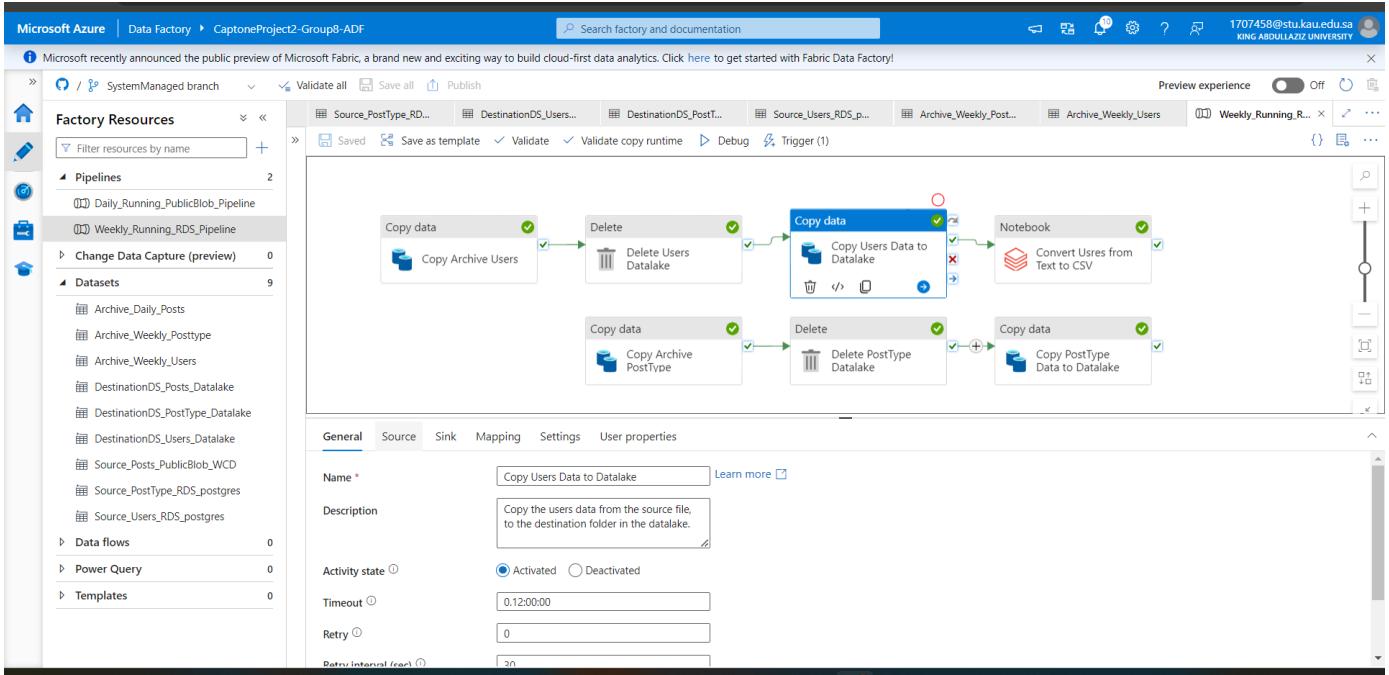
Activity state: Activated

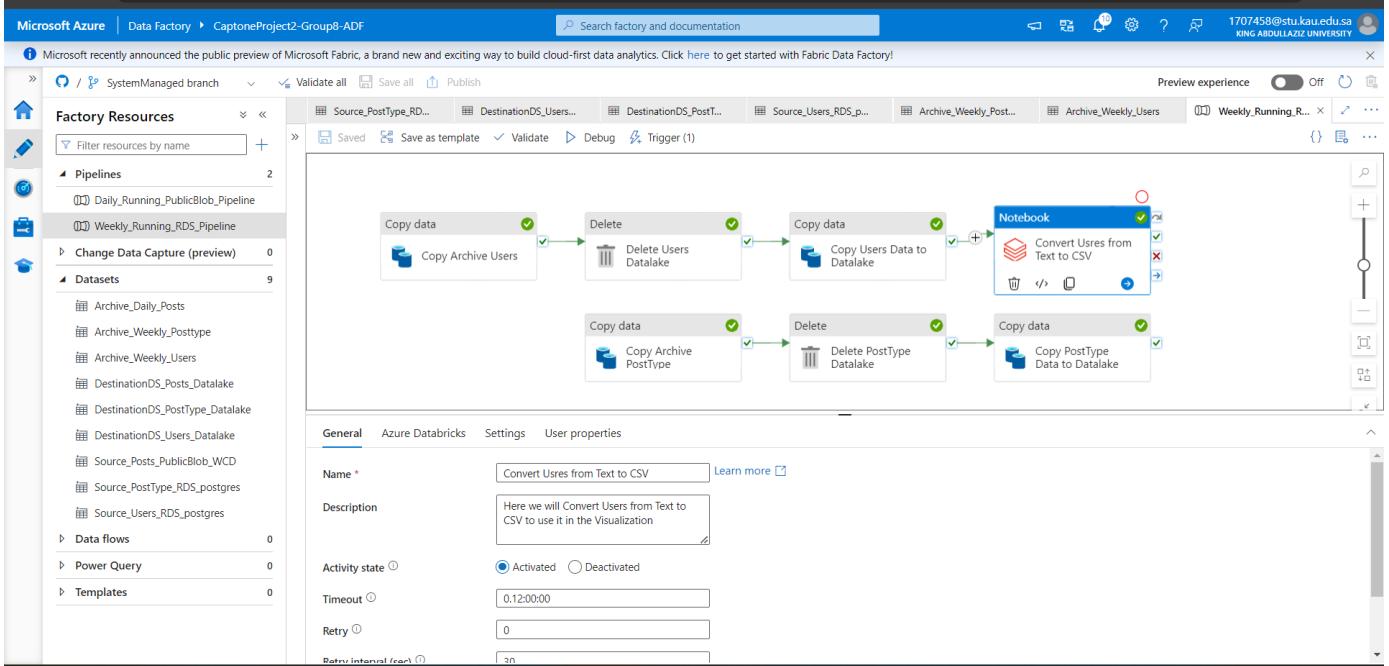
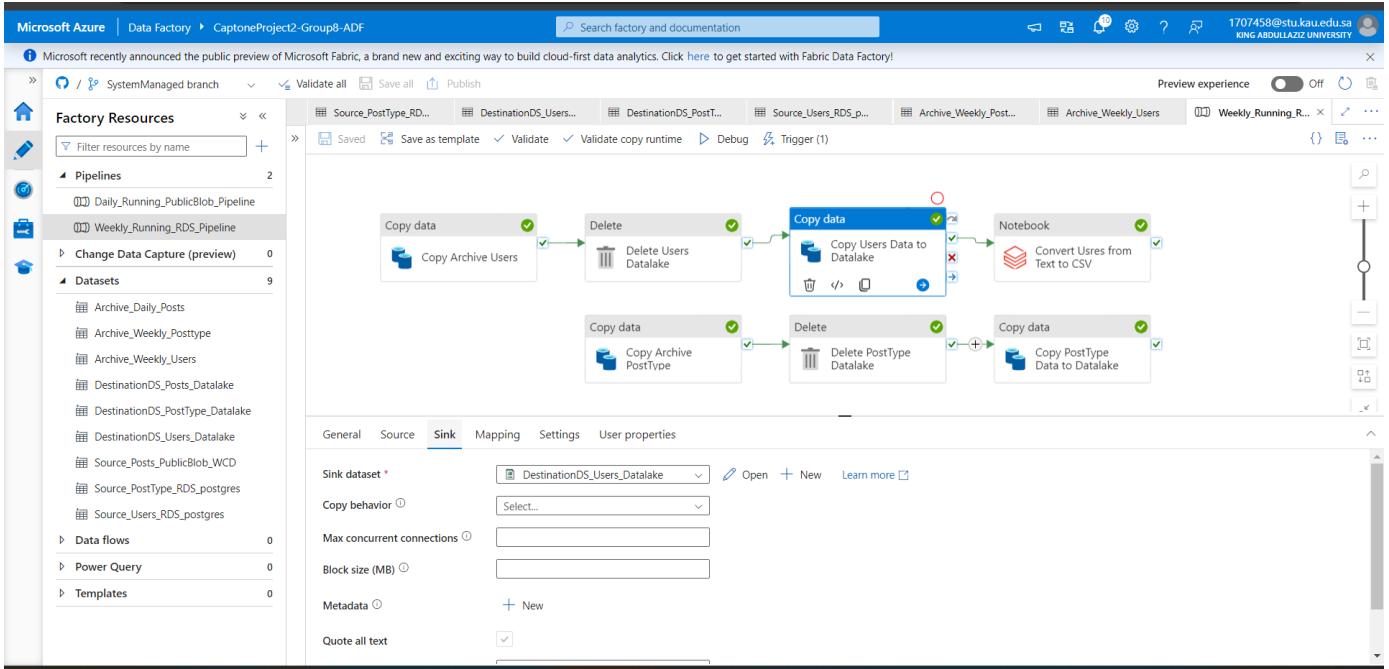
Timeout: 0:12:00:00

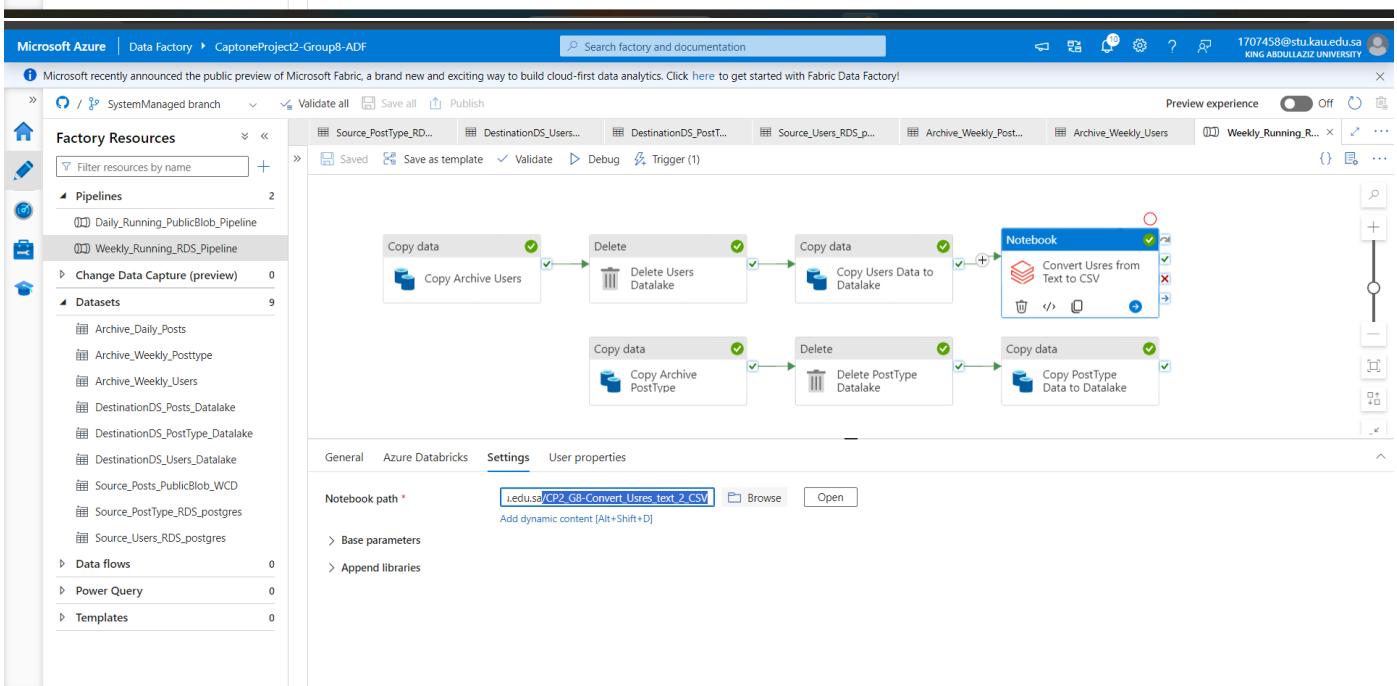
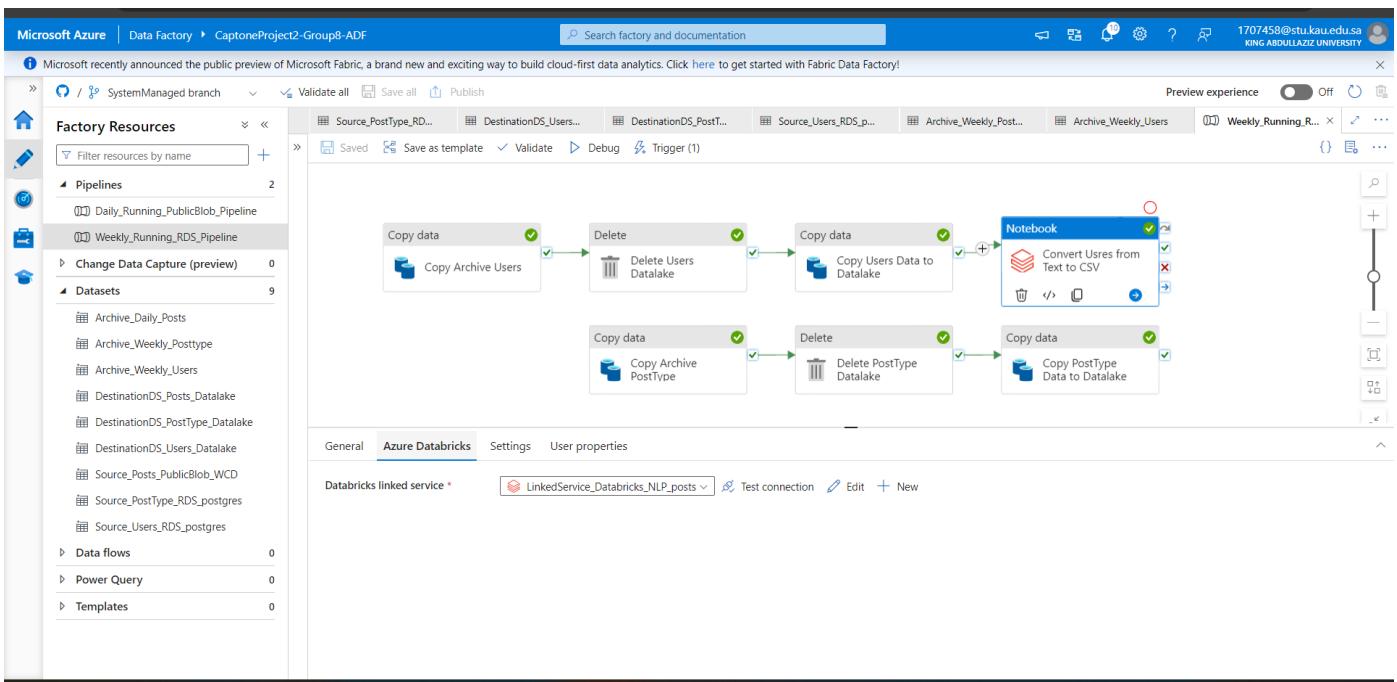
Retry: 0

Retry interval (min): 30

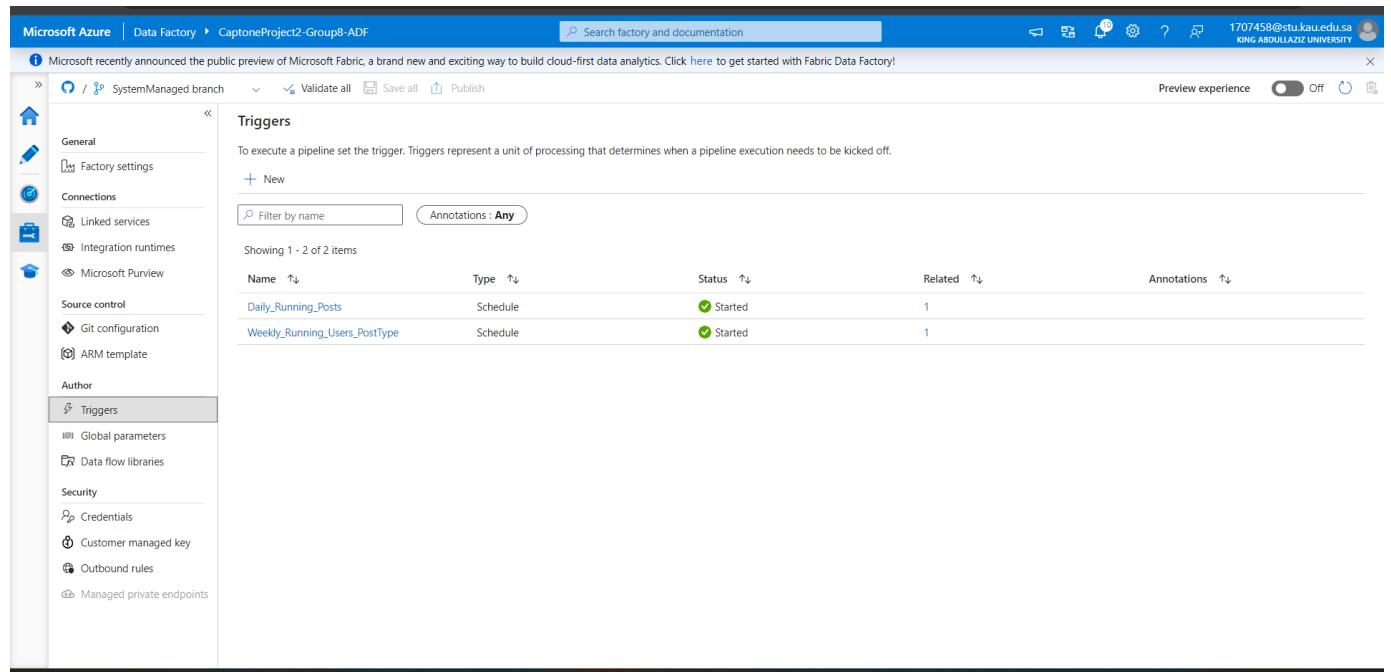




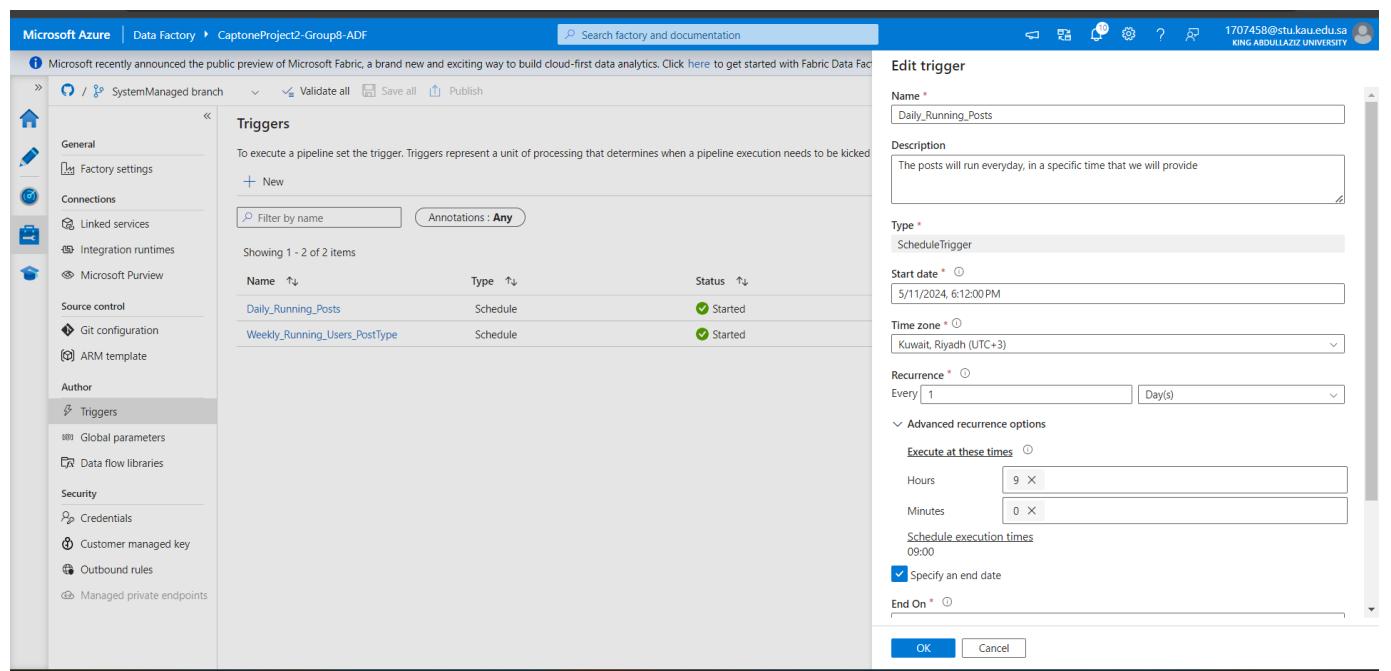




## 4- Trigger:



This screenshot shows the Microsoft Azure Data Factory interface. The left sidebar is titled 'Author' and includes sections for General, Connections, Integration runtimes, Microsoft Purview, Source control, and Author. Under Author, 'Triggers' is selected and highlighted in grey. The main content area is titled 'Triggers' and contains the following text: 'To execute a pipeline set the trigger. Triggers represent a unit of processing that determines when a pipeline execution needs to be kicked off.' Below this is a search bar with 'Filter by name' and 'Annotations: Any'. A table lists two triggers: 'Daily\_Running\_Posts' and 'Weekly\_Running\_Users\_PostType'. Both are of type 'Schedule' and are currently 'Started'. The table has columns for Name, Type, Status, Related, and Annotations.



This screenshot shows the 'Edit trigger' dialog for the 'Daily\_Running\_Posts' trigger. The left sidebar is identical to the previous screenshot. The right side of the screen is filled with the 'Edit trigger' form. The 'Name' field is populated with 'Daily\_Running\_Posts'. The 'Description' field contains the text: 'The posts will run everyday, in a specific time that we will provide'. The 'Type' field is set to 'ScheduleTrigger'. The 'Start date' is set to '5/11/2024, 6:12:00 PM'. The 'Time zone' is set to 'Kuwait, Riyadh (UTC+3)'. Under 'Recurrence', it says 'Every 1 Day(s)'. There is an 'Advanced recurrence options' section with 'Execute at these times' (Hours: 9, Minutes: 0), 'Schedule execution times' (09:00), and a checked 'Specify an end date' option. At the bottom are 'OK' and 'Cancel' buttons.

**Microsoft Azure | Data Factory > CapstoneProject2-Group8-ADF**

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory.

SystemManaged branch | Validate all | Save all | Publish

Search factory and documentation

Triggers

To execute a pipeline set the trigger. Triggers represent a unit of processing that determines when a pipeline execution needs to be kicked.

+ New

Filter by name Annotations: Any

Show 1 - 2 of 2 items

Name	Type	Status
Daily_Running_Posts	Schedule	Started
Weekly_Running_Users_PostType	Schedule	Started

**Edit trigger**

Name \* Weekly\_Running\_Users\_PostType

Description The Users and PostType will run every week, in a specific time that we will provide

Type \* ScheduleTrigger

Start date \* 5/11/2024, 6:23:00 PM

Time zone \* Nairobi (UTC+3)

Recurrence \* Every 1 Week(s)

Run on these days Sun Mon Tue Wed Thu Fri Sat

Execute at these times Hours 9 Minutes 0

Schedule execution times 09:00

OK Cancel

## 5- GitHub:

RenadBokhari

[https://github.com/RenadBokhari/CapstoneProject2\\_AzureDataFactory](https://github.com/RenadBokhari/CapstoneProject2_AzureDataFactory)

**Microsoft Azure | Data Factory > CapstoneProject2-Group8-ADF**

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

SystemManaged branch | Validate all | Save all | Publish

Search factory and documentation

Preview experience Off

Git repository

Git repository information associated with your data factory. CI/CD best practices

Edit Overwrite live mode Disconnect Import resources

Repository type	GitHub
GitHub account	RenadBokhari
Repository name	CapstoneProject2_AzureDataFactory
Collaboration branch	main
Publish branch	adf_publish
Root folder	/
Last published commit	6a21b46ce45cf0ecfaaccdcbd71ca13a6886e0d
Publish (from ADF Studio)	Enabled
Custom comment	Enabled