

Azure Data Factory Integration Runtimes

Step 1

Firstly, Open Azure Data Factory go to manage tab and create a new azure integration runtime proceed with continue.

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.

Data Factory Validate all Publish all

«

Integration runtimes

The integration runtime (IR) is the compute infrastructure to provide the following data integration capabilities across different network environments.

+ New Refresh

Filter by name

Showing 1 - 1 of 1 items

Name ↑↓	Type ↑↓	Sub-type ↑↓	Status ↑↓
AutoResolveIntegrationR...	Azure	Public	Running

Integration runtime setup

Integration Runtime is the native compute used by Azure Data Factory to create based on requirements.

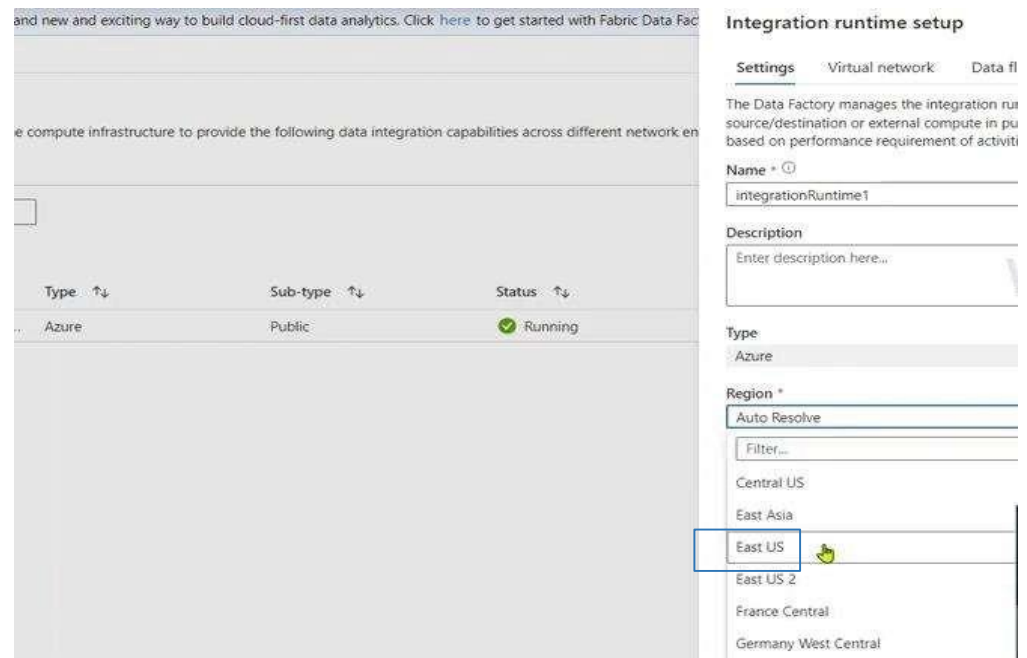
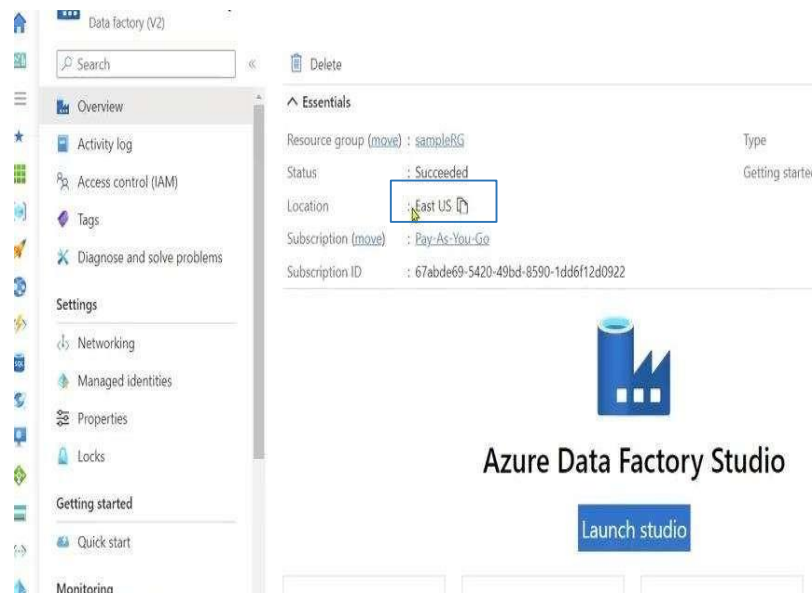
Azure, Self-Hosted
Perform data flows, data migration, and more.

Azure-SSIS
Lift-and-shift existing SSIS packages to Azure.

Airflow (Preview)
Use this for running your existing Airflow DAGs.

Step 2

Make a new IR name and description whatever region selected to create Azure Data Factory account select only that region in IR region field if you selected region East US to create Azure Data Factory account then select only East US in IR region field proceed with continue.



Step 3

In Virtual network enable virtual network configuration manage Vnet IR comes up with three option First, Interactive Authoring that is used to connect to the storages, Copy Compute scale that is used to copy the data and Pipeline and external compute scale that is used when you use other activities in a pipeline. Enable Interactive Authoring and Copy Compute scale go down and click on create button.

Search factory and documentation

Integration runtime setup

Settings Virtual network Data

Virtual network configuration ⓘ
☐ Disable ☒ Enable

Interactive authoring ⓘ
☒ Enable interactive authoring capability a
Time to live ⓘ
60 minutes

Advanced

Copy compute scale ⓘ
☒ Enable Copy compute scale
Compute size for copy ⓘ
Small (16 DLU)
Time to live ⓘ
5 minutes

Pipeline and external compute scale ⓘ
Compute size for pipeline ⓘ
Small (1 node)
Compute size for external ⓘ
Small (1 node)
Time to live ⓘ

Create Back

Step 4

Go back to Under integration runtime section you see Managed virtual network IR that is created in Virtual network.

Integration runtimes

The integration runtime (IR) is the compute infrastructure to provide the following data integration capabilities across different network environment.

+ New Refresh

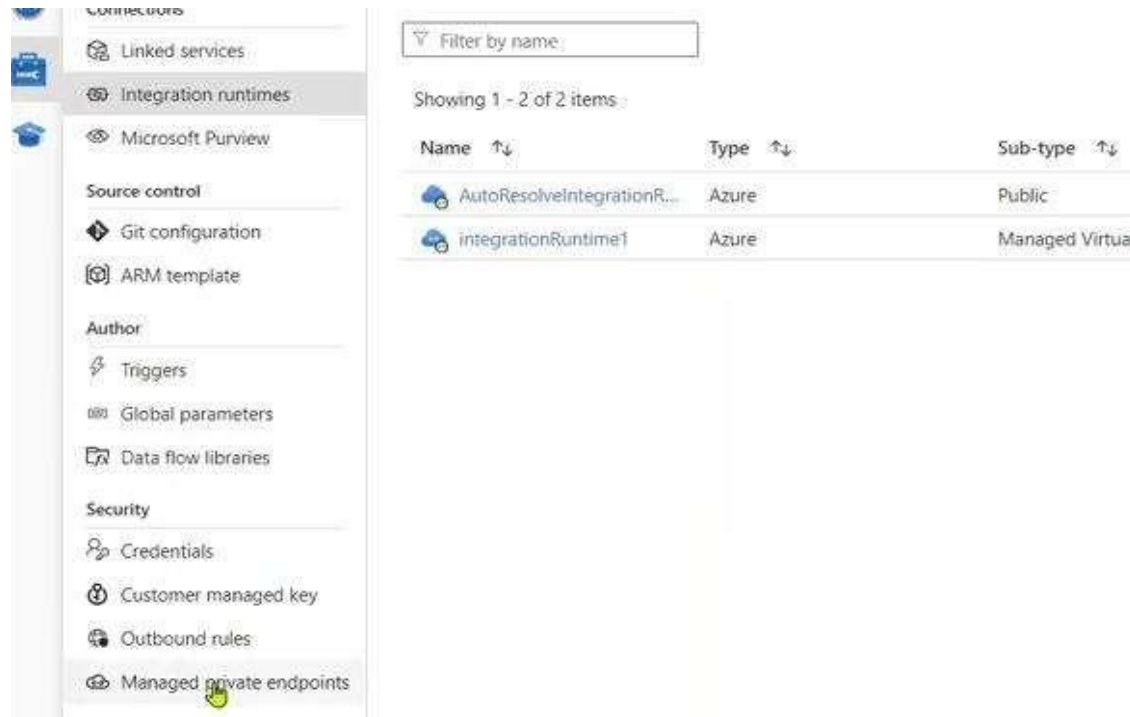
Filter by name

Showing 1 - 2 of 2 items

Name	Type	Sub-type	Status	Related
AutoResolveIntegrationR...	Azure	Public	Running	0
integrati...	Azure	Managed Virtual Network	Running	0

Step 5

Vnet Integration runtime uses private endpoints to connect to storages to transfer data. Go to Azure Data Factory manage tab and create Managed private endpoints.



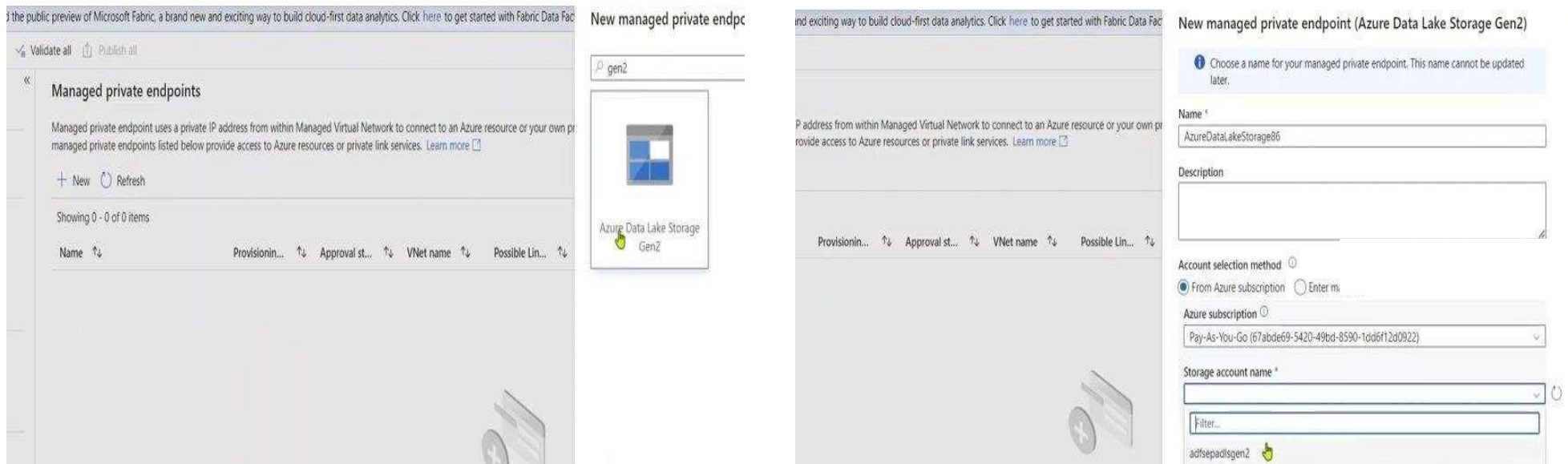
Filter by name

Showing 1 - 2 of 2 items

Name	Type	Sub-type
AutoResolveIntegrationR...	Azure	Public
integrationRuntime1	Azure	Managed Virtual

Step 6

Select storage as ADLS Gen2 proceed with continue make a new managed private endpoints name writing description is optional select account selection method as from Azure subscription select subscription as pay as you go and select ADLS Gen2 storage account name click on create button.



The screenshot displays the Azure portal interface for creating a new managed private endpoint. The left sidebar shows the 'Managed private endpoints' section with a '+ New' button. The main area shows the 'New managed private endpoint' form, which is titled 'New managed private endpoint (Azure Data Lake Storage Gen2)'. The form includes the following fields and options:

- Name:** A text input field containing 'AzureDataLakeStorage86'.
- Description:** A text input field.
- Account selection method:** Radio buttons for 'From Azure subscription' (selected) and 'Enter manually'.
- Azure subscription:** A dropdown menu showing 'Pay-As-You-Go (67abde69-5420-49bd-8590-1dd6f12d0922)'.
- Storage account name:** A text input field containing 'adfsepadi2gen2'.

The form also includes a 'Filter...' button and a 'Create' button at the bottom right.

Step 7

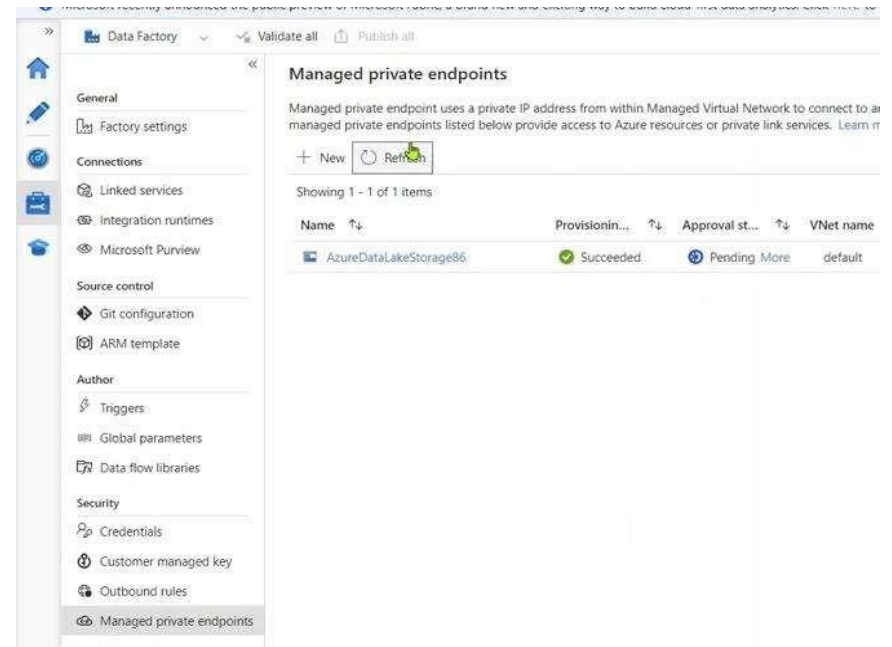
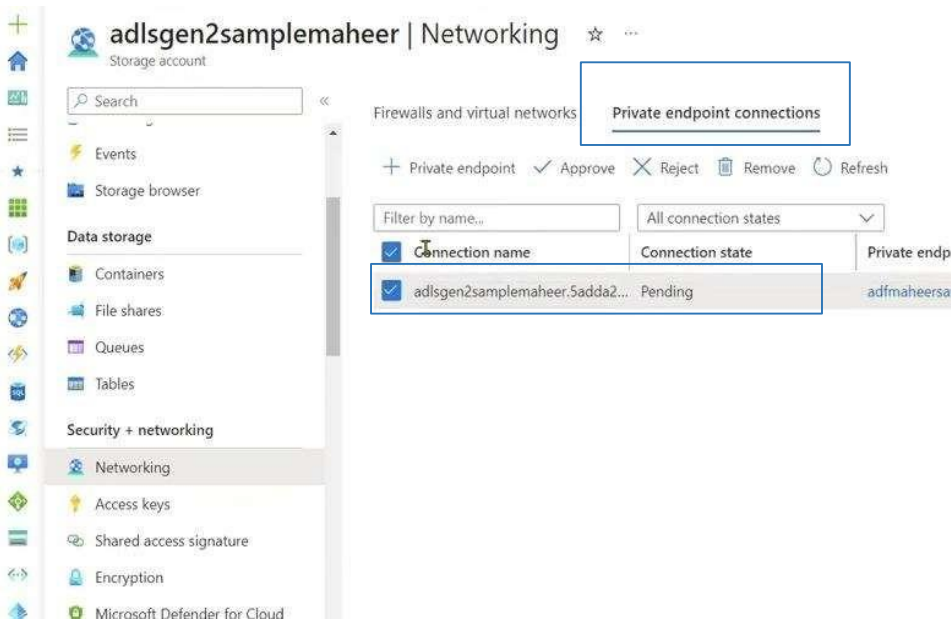
Go back to manage tab and go to Manage private endpoints you see Managed private endpoints created for ADLS Gen2 account it shows approval status pending.

The screenshot displays the 'Managed private endpoints' section in the Azure Data Factory interface. The left-hand navigation pane includes categories like General, Connections, Source control, Author, and Security, with 'Managed private endpoints' currently selected. The main content area features a table listing the endpoints. A single endpoint, 'AzureDataLakeStorage86', is shown with a 'Succeeded' provisioning status and a 'Pending' approval status. A blue rectangular box highlights the 'Approval status' column header and the 'Pending' status of the listed endpoint.

Name	Provisioning status	Approval status	VNet name
AzureDataLakeStorage86	Succeeded	Pending	default

Step 8

To Make Managed private endpoints approved successful go back to Azure portal go to storage account then go to storage account name that you select in endpoints go to networking then go to private endpoints connections enable pending connection name and click on approve. Wait for two to four until approve makes successful in managed private endpoints.



Step 9

The main agenda to create managed Vnet IR is we want to see secure data transfer from source to sink. In this step create a new linked service for ADLS Gen2 make a new linked service name select integration runtime as Managed Vnet select account method from subscription select subscription that you took and select ADLS storage account name and click on create make sure to test connection before creating linked service.

The screenshot shows the 'New linked service' configuration page in Azure Data Lake Studio. The page is divided into two main sections: a list of linked services on the left and a configuration form on the right.

Left Panel: Displays a list of linked services. The current view shows 'No linked service to show'. A message states: 'If you expected to see results, try changing your filters or clicking the refresh button.' A 'Create linked service' button is visible at the bottom.

Right Panel: The 'New linked service' configuration form for 'Azure Data Lake Storage Gen2'. The form includes the following fields and options:

- Name ***: AzureDataLakeStorage1
- Description**: (Empty text box)
- Connect via integration runtime ***:
 - ☒ integrationRuntime1 (Managed Virtu)
 - ☒ Interactive authoring enabled
- Authentication type**: Account key
- Account selection method**:
 - ☒ From Azure subscription
 - ☐ Enter manually
- Azure subscription**: Pay-As-You-Go (67abde69-5420-49bd-85...)
- Storage account name ***: adlsgen2samplemabeer
- Loading key...**: (Loading indicator)
- Managed private endpoint**: AzureDataLakeStorage86
- Test connection**:
 - ☒ To linked service
 - ☐ To file path

At the bottom of the form, there are 'Create' and 'Back' buttons.

Step 10

Create a new pipeline and add copy activity.

The screenshot displays the Azure Data Factory (ADF) interface. On the left, the 'Factory Resources' pane shows a tree view with 'Pipelines' expanded, and 'pipeline1' selected. The 'Activities' pane on the right lists various activity types, with 'Copy data' selected under the 'Move and transform' category. The main canvas shows a visual representation of the 'Copy data' activity, labeled 'Copy data1'. Below the canvas, the 'General' tab of the activity's properties is visible, showing fields for 'Name' (set to 'Copy data1'), 'Description', 'Activity state' (set to 'Activated'), and 'Timeout' (set to '0:12:00:00').

Factory Resources

- Filter resources by name
- Pipelines (1)
 - pipeline1
- Change Data Capture (preview) (0)
- Datasets (0)
- Data flows (0)
- Power Query (0)

Activities

- Search activities
- Move and transform
 - Copy data
 - Data flow
- Synapse
 - Azure Data Explorer
 - Azure Function
 - Batch Service
 - Databricks
 - Data Lake Analytics
 - General
 - HDInsight
 - Iteration & conditionals
 - Machine Learning
 - Power Query

Copy data

- Copy data1

General | Source | Sink | Mapping | Settings | User properties

Name * Copy data1 [Learn more](#)

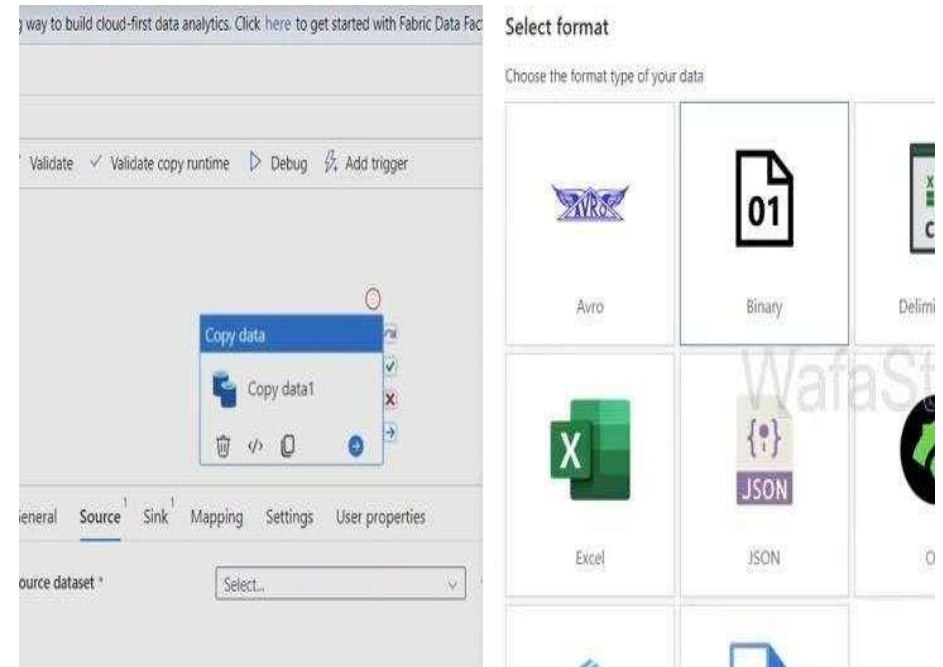
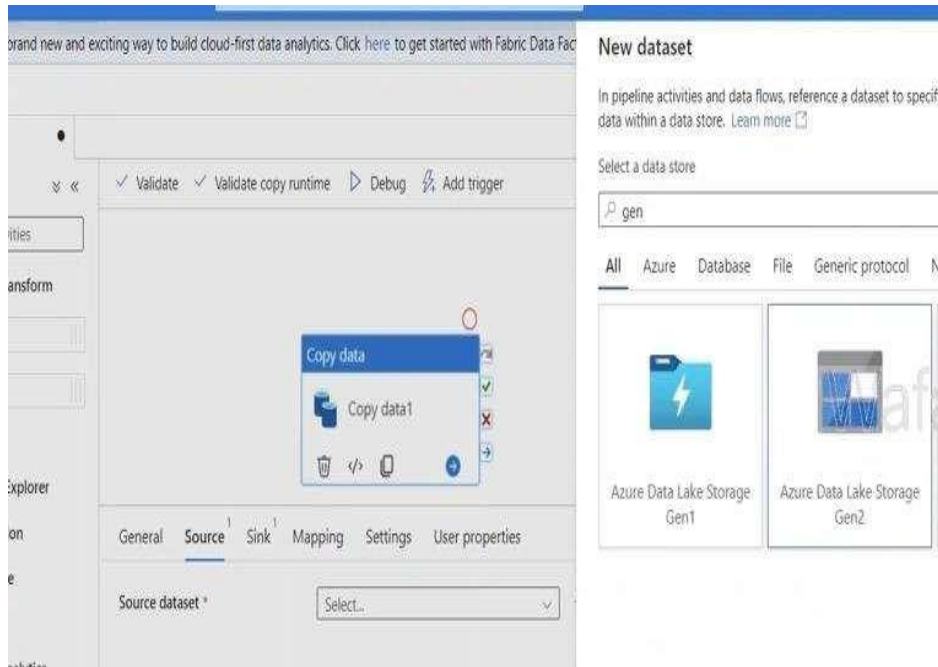
Description

Activity state ☒ Activated ☐ Deactivated

Timeout 0:12:00:00

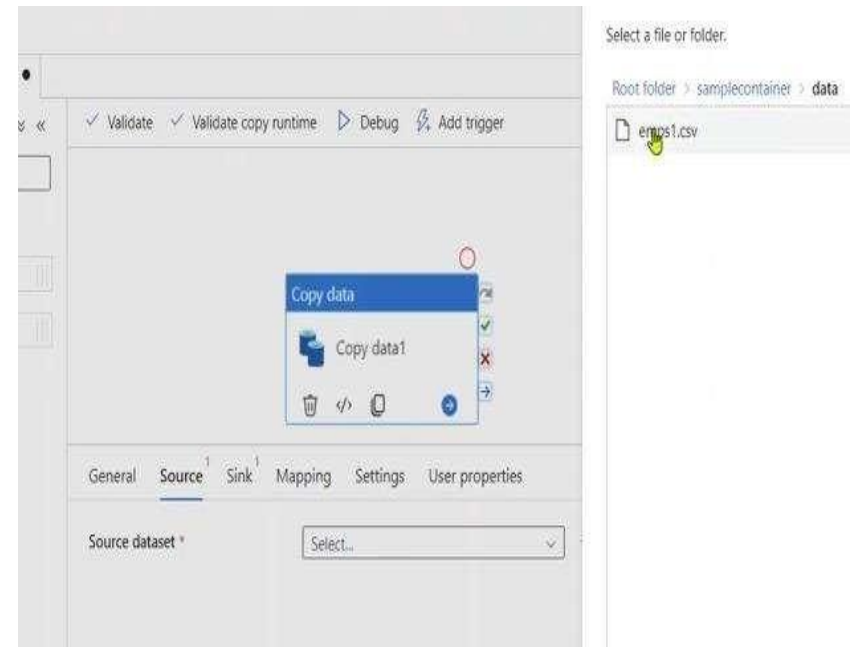
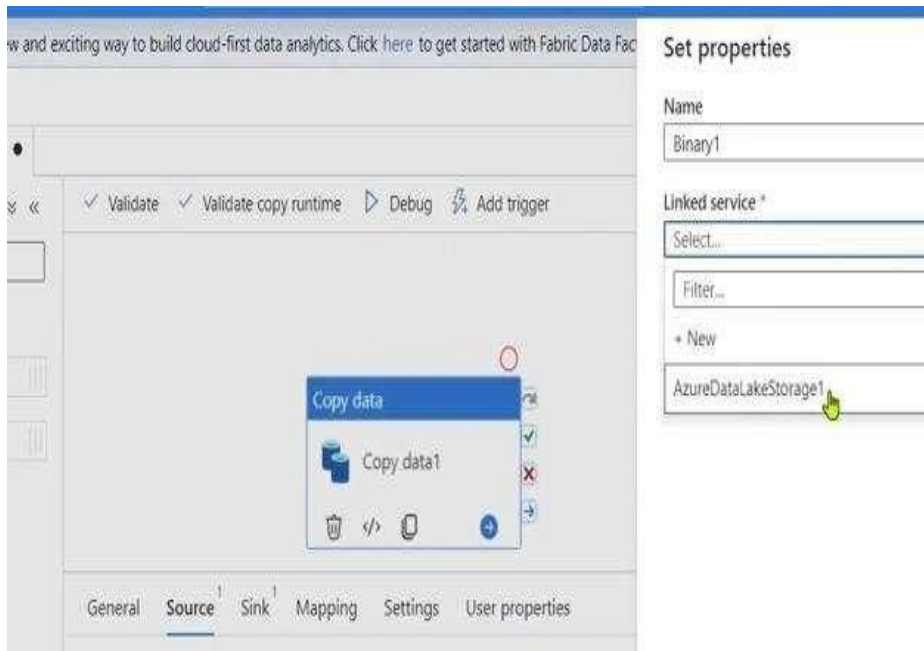
Step 11

Go to source create a new dataset for ADLS Gen2 and select format as binary proceed with continue.



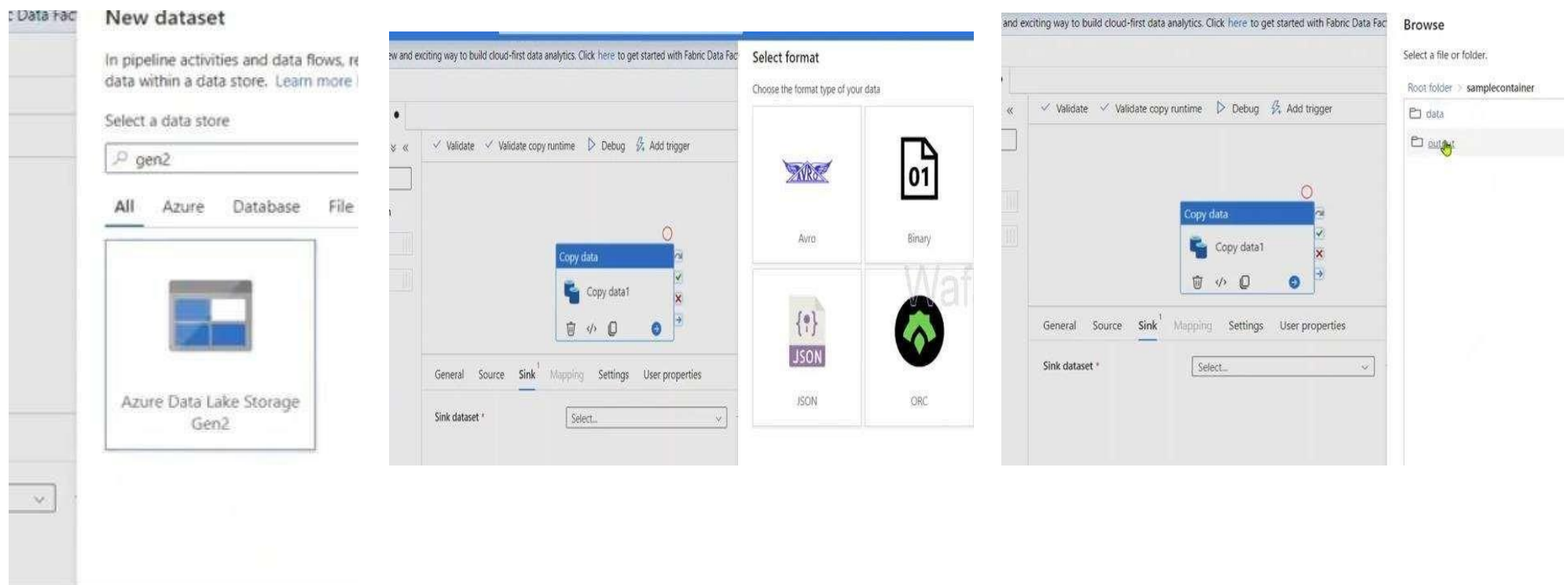
Step 12

In set properties select the linked service that you created prior select .csv file from sample container proceed with done.



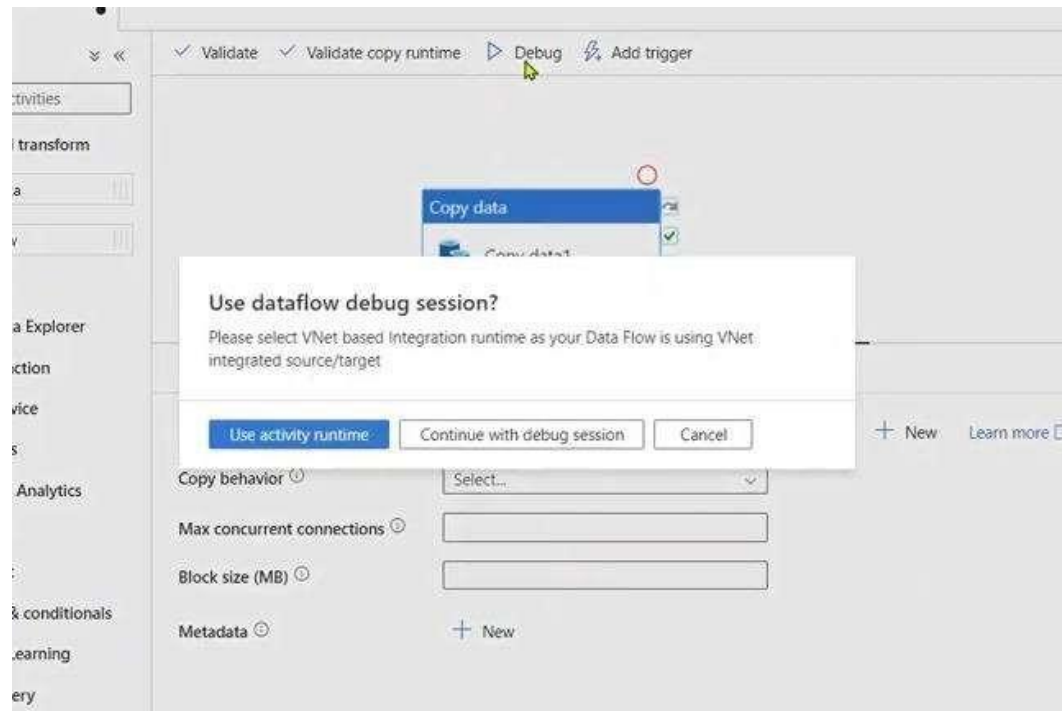
Step 13

Go to sink similarly create a new dataset for ADLS Gen2 select new dataset format as binary select the output container to store .csv file click on done.



Step 14

Prior to execute pipeline using debug mode make sure to select use activity runtime.



Step 15

Once pipeline execution successful click on info symbol ahead green tick mark you see queue time 1 min :11 sec and transfer time 3 sec it took 1 min : 11 sec because it has to get IR first in virtual network then it took 3 sec to immediately transfer data.

The screenshot shows the Azure Data Factory pipeline execution console. At the top, there are tabs for 'Parameters', 'Variables', 'Settings', and 'Output'. The 'Output' tab is selected. Below the tabs, the pipeline run ID is displayed as 'f7520bcc-42dd-4c70-94e6-268cd9579814'. The pipeline status is 'Succeeded' with a green checkmark. A table below shows the activity details:

Activity name	Activity status	Activity type	Run start	Duration
Copy data1	Succeeded	Copy data	10/23/2023, 7:21:51 PM	1m 20s

The screenshot shows the detailed execution metrics for the 'Copy data' activity. The metrics are as follows:

Metric	Value
Data read	50 bytes
Files read	1
Peak connections	1
Data written	50 b
Files written	1
Peak connections	1
Copy duration	00:01:18
Throughput	17 bytes/s
Azure Data Lake Storage Gen2 → Azure Data Lake Storage Gen2	
Start time	10/23/2023, 7:21:52 PM
Used DIs	4
Used parallel copies	1
Duration	00:01:18
Details	
Queue	00:01:11
Transfer	00:00:03
Data consistency verification	
Not verified	