

Azure Data Factory

Lab: Trigger a pipeline using storage event trigger

Pre-requisites:

- Azure Pass subscription
- Azure Data Lake Storage Gen2 storage account
- Azure SQL Database

Lab Objective:

After completing this lab, you will be able to:

- Create the trigger in Azure Data Factory.
- Create use the triggers in the pipeline

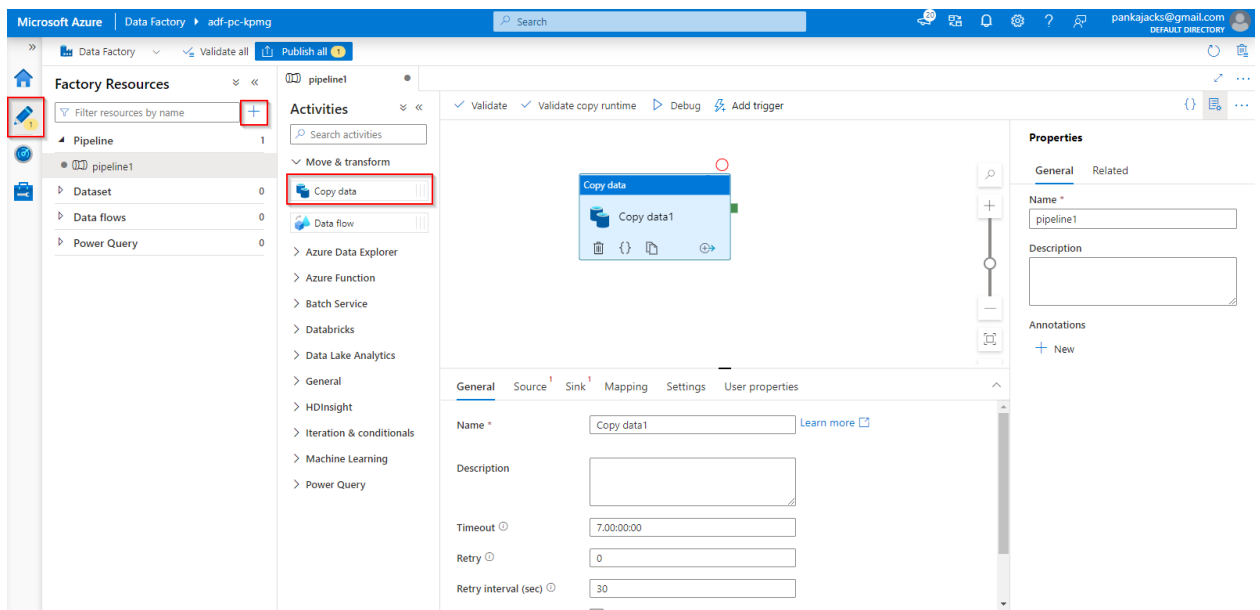
Exercise 1: Create a trigger to execute the pipeline based on storage event.

The main task for this exercise are as follows:

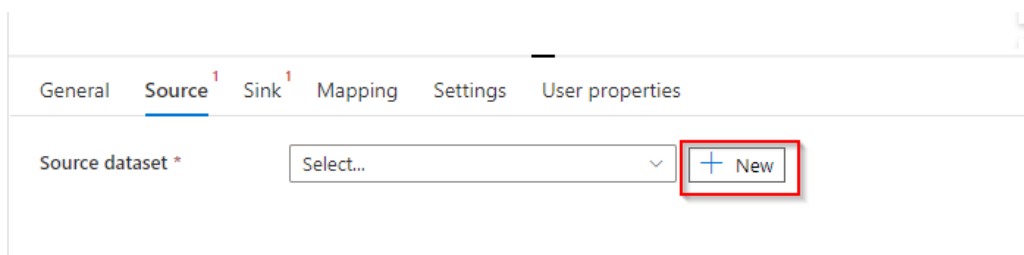
- Create a copy activity to copy data.
- Create a storage event trigger to run the pipeline.

Task 1: Create a copy activity to copy data

1. Open data factory adfxx.
2. **Open the authoring canvas** in ADF homepage, click on the **pencil icon** on the left sidebar and select the **+ pipeline button** to open the authoring canvas and create a pipeline.
3. **Add a copy activity** In the Activities pane, open the **Move and Transform** accordion and drag the **Copy data** activity onto the pipeline canvas.



4. Create a new Dataset



5. Select Azure Data Lake Storage Gen 2 and click continue and the select Delimited Text and continue

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

Select a data store

Search

All **Azure** Database File Generic protocol NoSQL Services and apps

Azure Blob Storage Azure Cosmos DB (MongoDB API) Azure Cosmos DB (SQL API)

Azure Data Explorer (Kusto) Azure Data Lake Storage Gen1 **Azure Data Lake Storage Gen2**

Continue Cancel

Select format

Choose the format type of your data

Avro Binary **DelimitedText**

Excel JSON ORC

Parquet XML

Continue Back Cancel

6. In set properties enter name ADLSGen2Source, Select Existing Linked service for the Storage account otherwise create a new.

Set properties

Name

ADLSGen2Source

Linked service *

Select... Filter...

Select...

+ New

ADLSGen2

7. Do not select the folder path or any file.

Set properties

Name
ADLSTGen2Source 1

Linked service *
ADLSTGen2

File path
File system / Directory / File

First row as header ☐ Do notSelect any file

Import schema
☐ From connection/store ☐ From sample file ☒ None

> Advanced

2

OK Back Cancel

8. Open ADLSTGen2Source and create two dataset parameters

1 Source Sink¹ Mapping Settings User properties

Source dataset * 2 ADLSTGen2Source 3 Open + New Preview data

File path type
☒ File path in dataset ☐ Wildcard file path ☐ List of files ⓘ

Filter by last modified ⓘ Start time (UTC) End time (UTC)

Recursively ⓘ ☒

Enable partition discovery ⓘ ☐

DelimitedText
ADLSGen2Source

Connection Schema **Parameters** 1

+ New | Delete

<input type="checkbox"/>	Name	Type	Default value	
<input type="checkbox"/>	FolderPath 2	String	Value	
<input type="checkbox"/>	FilePath 3	String	Value	

9. Under connection set the file path using add dynamic content as mention in the image below:

pipeline4 ADLSGen2Source

DelimitedText
ADLSGen2Source

1 **Connection** Schema Parameters

Linked service * ADLSGen2 Test connection Edit + New Learn n

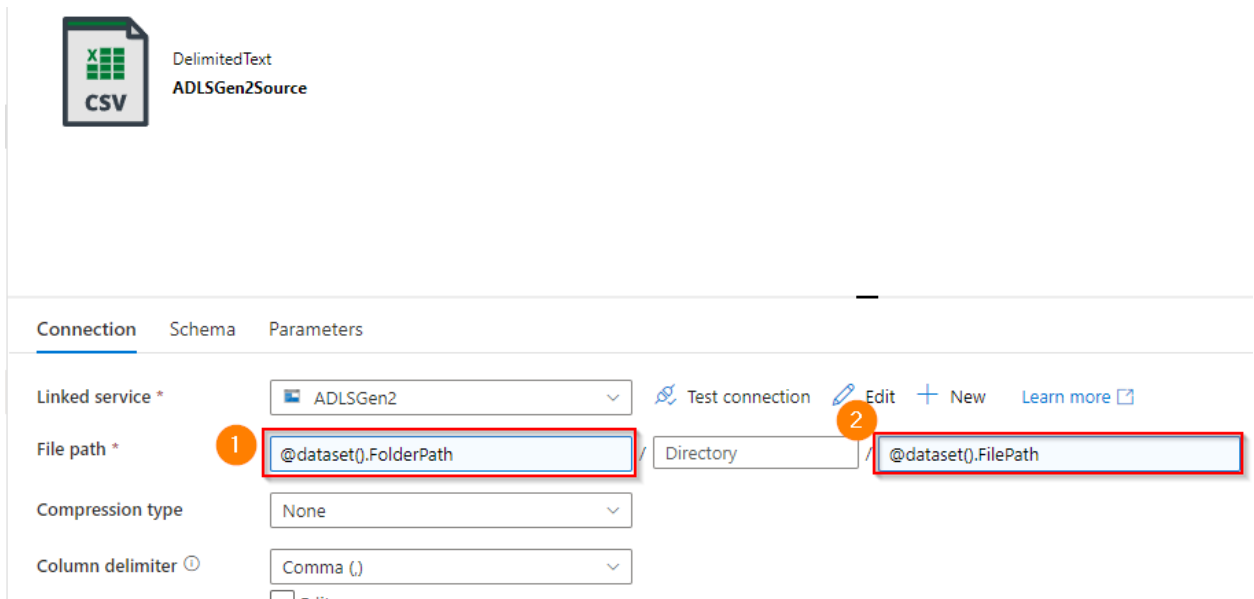
File path * File system / Directory / File Browse v

2 Add dynamic content [Alt+Shift+D]

Compression type None

Column delimiter ① Comma (,)

Set the file path as show below:



DelimitedText
ADLSGen2Source

CSV

Connection Schema Parameters

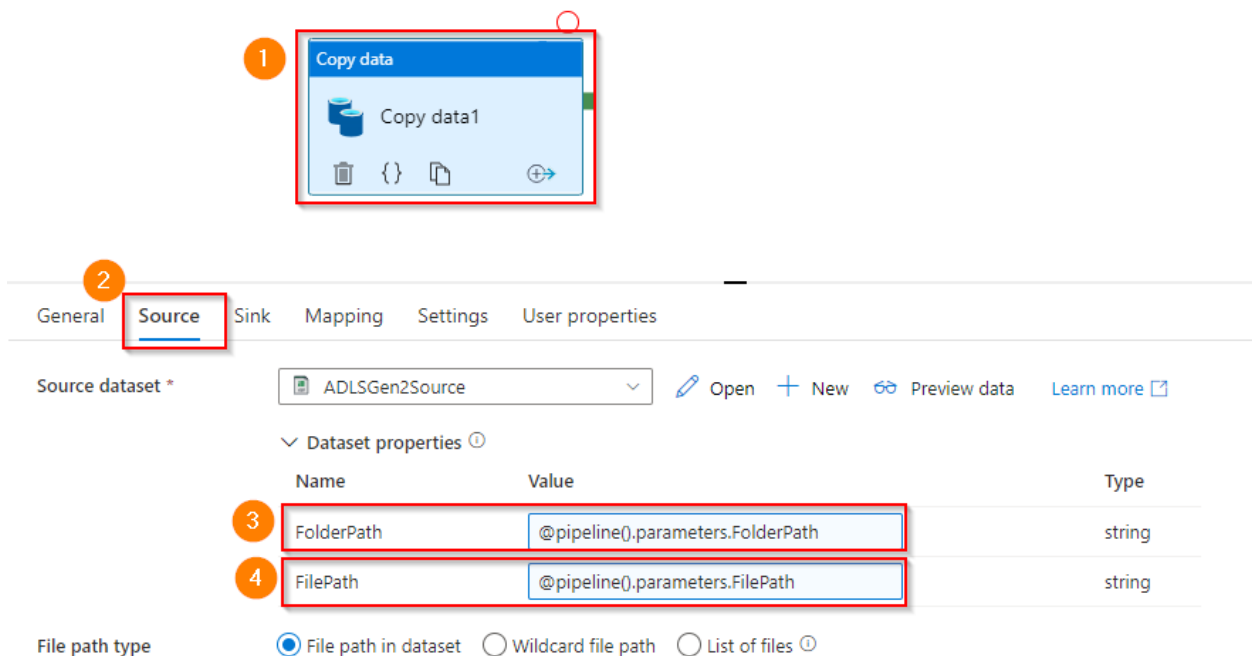
Linked service * ADLSGen2 Test connection Edit + New Learn more

File path * 1 @dataset().FolderPath Directory 2 @dataset().FilePath

Compression type None

Column delimiter ① Comma (,)

10. Select the copy activity, under source setting add dynamic content for Dataset properties using pipeline parameters.



Copy data

Copy data1

General Source Sink Mapping Settings User properties

Source dataset * ADLSGen2Source Open + New Preview data Learn more

Dataset properties ①

Name	Value	Type
3 FolderPath	@pipeline().parameters.FolderPath	string
4 FilePath	@pipeline().parameters.FilePath	string

File path type ☒ File path in dataset ☐ Wildcard file path ☐ List of files ①

11. Create another dataset for sink of type ADLSGen2 and create the parameter FileName

Name	Type	Default value
FileName	String	Value

12. Under connection setting of sink dataset browse the file path to **data->output** and for the filename select **add dynamic content** and select the Filename @dataset parameter

Linked service * ADLSGen2

File path * data / output

Compression type None

Column delimiter ① Comma (,)

Filename @dataset().FileName

Browse

13. In the sink of copy activity add the sink data set and pass the pipeline parameter value to dataset parameter filename.

The screenshot shows the 'Sink' tab of a 'Copy data' activity in Azure Data Factory. The 'Sink dataset' is 'ADLSGen2'. The 'Dataset properties' table is as follows:

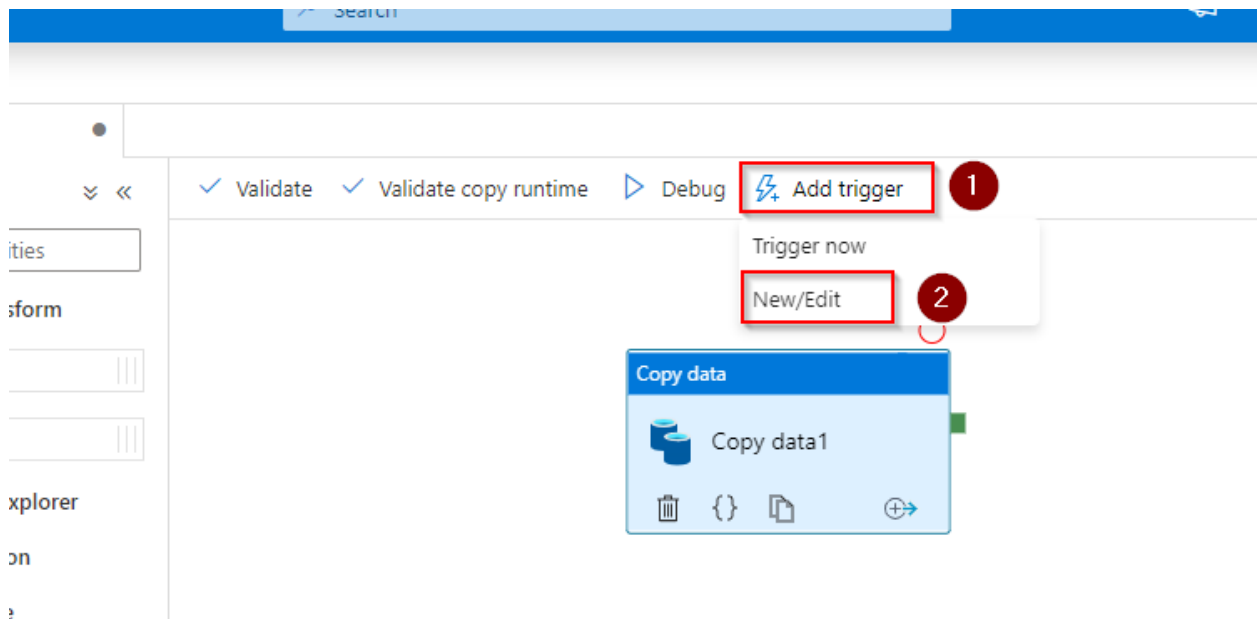
Name	Value	Type
FileName	@pipeline().parameters.FilePath	string

14. In Azure portal, enable the Microsoft.EventGrid Resource Provider from Azure Pass subscription as shown below.

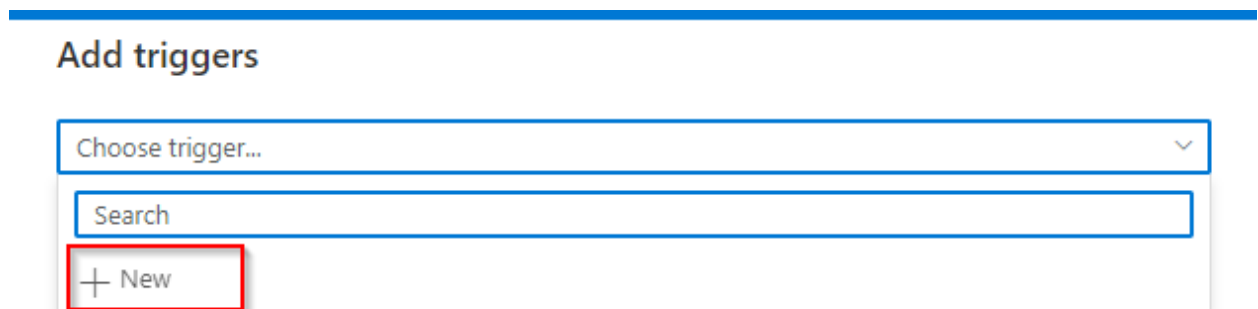
The screenshot shows the 'Resource providers' page for the 'Azure Pass - Sponsorship' subscription. The 'Microsoft.EventGrid' resource provider is highlighted with a red box and a red circle, indicating it is not yet registered. The 'Register' button is also highlighted with a red box and a red circle.

Resource provider	Registration status
Microsoft.Diagnostics	NotRegistered
Microsoft.DigitalTwins	NotRegistered
Microsoft.DocumentDB	NotRegistered
Microsoft.DomainRegistration	NotRegistered
Microsoft.EdgeOrder	NotRegistered
Microsoft.Elastic	NotRegistered
Microsoft.EnterpriseKnowledgeGraph	NotRegistered
Microsoft.EventGrid	NotRegistered
Microsoft.Experimentation	NotRegistered
Microsoft.ExtendedLocation	NotRegistered
Microsoft.Falcon	NotRegistered
Microsoft.Features	Registered
Microsoft.FidoU2F	NotRegistered
Microsoft.FluidRelay	NotRegistered
Microsoft.GuestConfiguration	NotRegistered

15. Add trigger to the pipeline:



16. Configure the Trigger as show below:



New trigger

Name *

StorageEventTrigger

Description

Type *

Storage events

Account selection method * ⓘ

☒ From Azure subscription ☐ Enter manually

Azure subscription ⓘ

Azure Pass - Sponsorship (c8fd5e96-3329-4700-b2ef-455eb9a984e4)

Storage account name * ⓘ

datalakepc1

Container name * ⓘ

data

Blob path begins with ⓘ

input/

Blob path ends with ⓘ

.csv

Event * ⓘ

☒ Blob created ☐ Blob deleted

Ignore empty blobs * ⓘ

☒ Yes ☐ No

Annotations

+ New

Start trigger ⓘ


☒ Start trigger on creation

Continue

Cancel

Click Continue:

Data preview

 Make sure you have specific filters. Configuring filters that are too broad can match a large number of files created/deleted and may significantly impact your cost.

Event Trigger Filters

Container name: **data**
Starts with: **input/**
Ends with: **.csv**

0 blobs matched in "data" Refresh

Blob name

0 - 0 of 0 items < Previous 1 Next > Go to

Continue Back Cancel

Next initialize the following:


FolderPath: @triggerBody().folderPath

FilePath: @triggerBody().fileName

Then click OK and Publish the pipeline

Edit trigger

Trigger Run Parameters

 Parameters that are not provided a value will not be included in the trigger.

Name	Type	Value
FolderPath	string	@triggerBody().folderPath
FilePath	string	@triggerBody().fileName

Make sure to "Publish" for trigger to be activated after clicking "OK"

OK Cancel

17. Create some dummy csv file and upload them from storage account, this will trigger the pipeline you can go to monitor hub under pipeline run you can see the pipeline is executing and you can also monitor the trigger run under monitor hub.

The screenshot displays the Azure Storage Explorer interface. On the left, the 'data' container is selected. The main pane shows a table of blobs with columns: Name, Modified, Access tier, Archive status, and Blob type. The table contains two entries: '[.]' and 'inputEmp.txt'. The 'Upload' button in the top toolbar is highlighted with a red box and a blue circle with the number 3. On the right, the 'Upload blob' dialog is open, showing the 'Files' section with a 'Select a file' button highlighted by a red box and a blue circle with the number 4. The 'Advanced' section is expanded, and the 'Upload' button is highlighted with a red box and a blue circle with the number 5. The 'Current uploads' section shows a list of uploads with their progress and status.

Name	Modified	Access tier	Archive status	Blob type
[.]				
inputEmp.txt	3/2/2022, 4:35:34 PM	Hot (Inferred)		Block blob

Current uploads	Dismiss	Completed	All
Data1.csv	6 B / 6 B	***	
Data1 - Copy.csv	6 B / 6 B	***	
Data1 - Copy (4).csv	6 B / 6 B	***	
Data1 - Copy (3).csv	6 B / 6 B	***	
Data1 - Copy (2).csv	6 B / 6 B	***	
Data1 - Copy (1).csv	6 B / 6 B	***	