Sponsor

It seems like yesterday

Home / Updates / General availability: Azure Data Factory

■■■ NOW AVAILABLE

General availability: Azure Data Factory

Published date: August 06, 2015

Azure Data Factory is a cloud-based data integration service that automates moving and transforming data. You can quickly create, deploy, schedule, and monitor data-flow pipelines with new tools and features, including:

- A Visual Studio plug-in for authoring

- Integration with Azure Machine Learning and Azure Batch

- Enhanced Azure HDInsight integration with Apache HCatalog and on-demand cluster management

- New data connectors for cloud and on-premises stores

- Globally deployed data movement as a service

Less than 10 years later

Introducing Data Factory in Microsoft Fabric

Announcements   Data Factory

May 23, 2023 by Wee Hyong Tok   👁 54,591 Views   in Share   🐦 Tweet   f Like

See Arun Ulagaratchagan's blog post to read the full *Microsoft Fabric preview announcement*.

As organizations work through their digital transformation journeys, they are awash with data that they need to bring together to meet their analytics needs. To do this, they need to seamlessly access data from a diverse set of data sources, reshape and transform the data to the right format, and load the data into a data store like a lakehouse or data warehouse, where the data can support their analytics needs. In addition, companies are looking to stitch together different data activities (e.g., running a notebook, a stored procedure, SQL script, and more) and orchestrate specific data workflows.

Today, we are unveiling Data Factory in Microsoft Fabric, bringing together the best of Power Query and Azure Data Factory into a single, modern data integration experience which empowers both data and business professionals with capabilities to ingest and transform data as well as orchestrate data workflows. Our goal is to empower everyone to be able to connect to diverse data sources and bring that data to where it can best help you derive insights for better business decisions.

Even docs push you...

> 💡 **Tip**
>
> Try out **Data Factory in Microsoft Fabric**, an all-in-one analytics solution for enterprises. **Microsoft Fabric** covers everything from data movement to data science, real-time analytics, business intelligence, and reporting. Learn how to **start a new trial** for free!
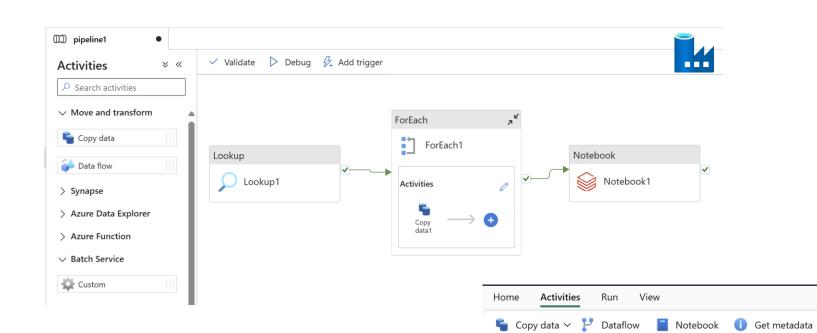
unfortunately,
we no longer need
your services

Agenda

What is equal?

What is it not equal?

What is it improved?

A **pipeline** is a logical grouping of **activities** that together perform a task

Activities are Activities
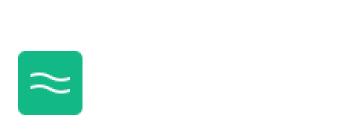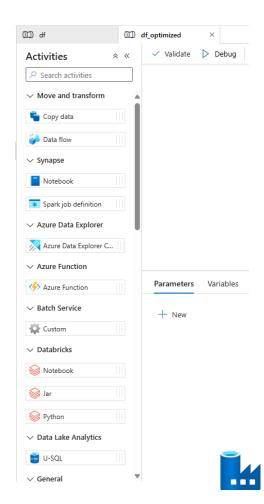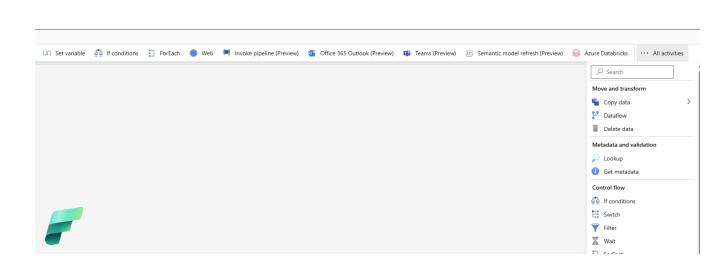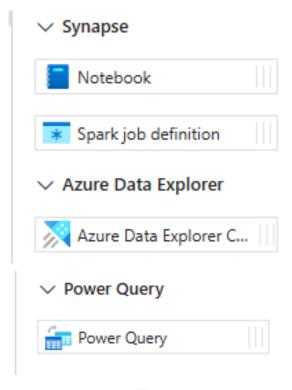
Some are different
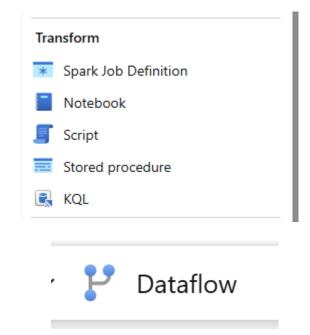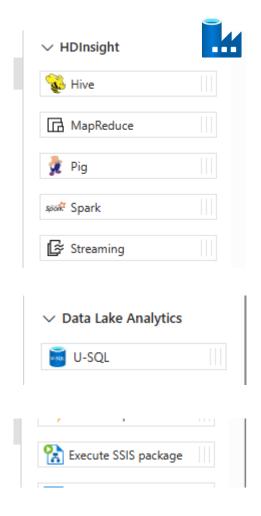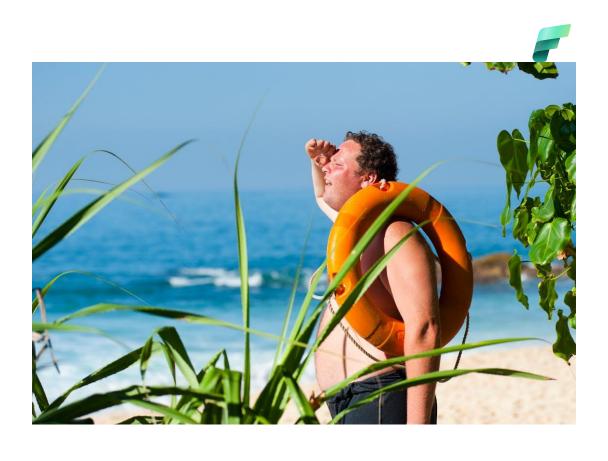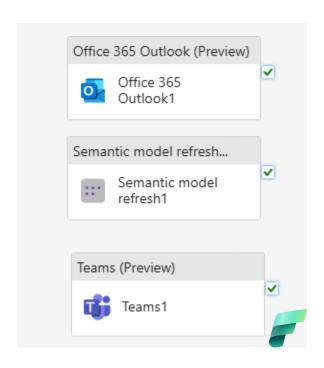
Some are missing

Some improve

Office 365 Outlook (Preview)

Office 365 Outlook1 ✓

Semantic model refresh...

Semantic model refresh1 ✓

Teams (Preview)

Teams1 ✓

- Send an email
- Send a message in teams
- Refresh a Power BI model

Parametrization

- Parent | Child pipeline
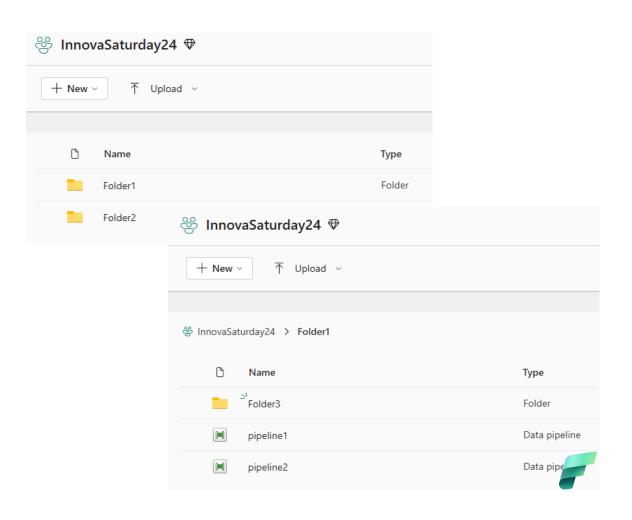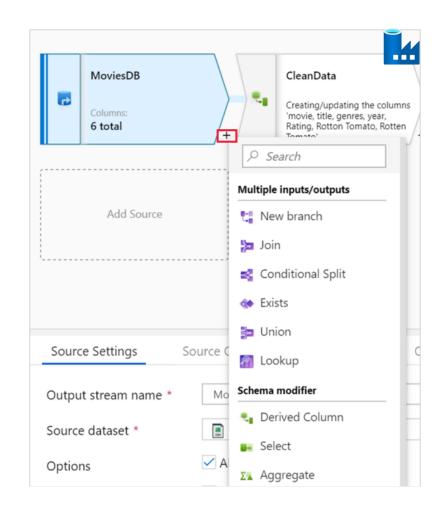- Global Parameters
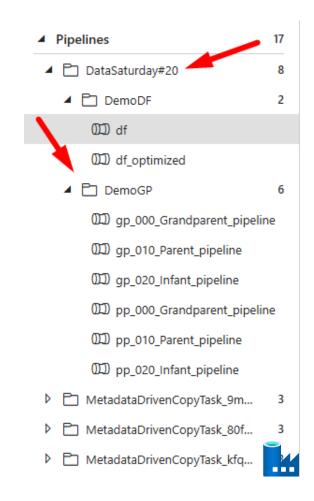
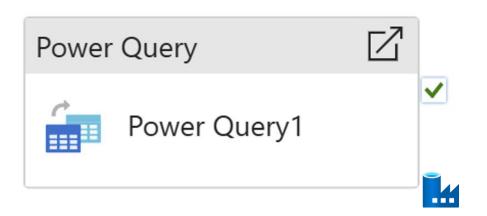- Parent | Child pipeline

Pipeline organization
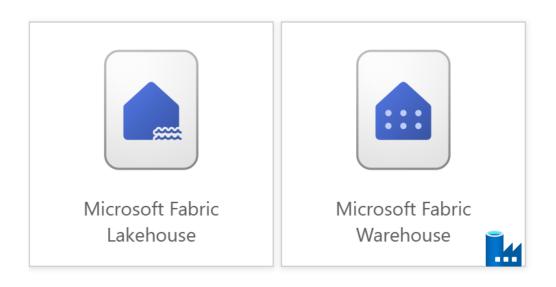
Mapping Data Flows

Wrangling Data Flows / Power Query
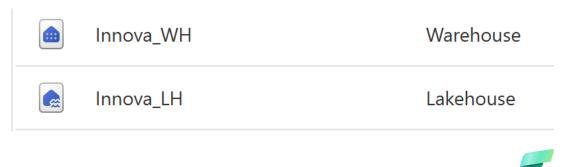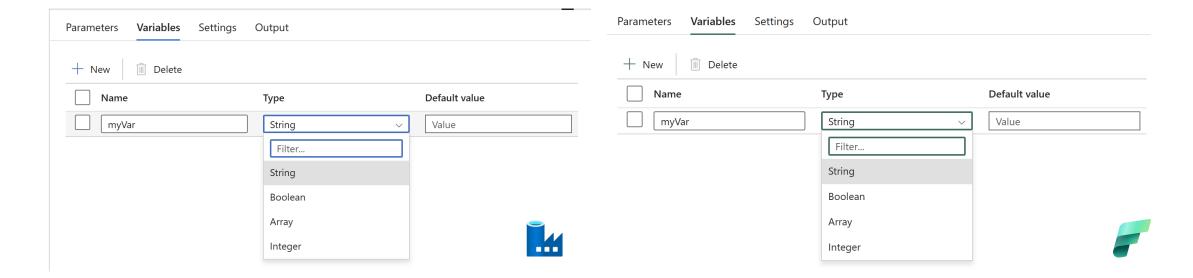
Lakehouse & Warehouse Connector

Microsoft Fabric
Lakehouse

Microsoft Fabric
Warehouse

Innova_WH   Warehouse

Innova_LH   Lakehouse

Variables

Expressions

```
@concat(
    'Test_',
    convertFromUtc(utcnow(),
    'Pacific Standard Time')
)
```

```
@concat(
    'Test_',
    convertFromUtc(utcnow(),
    'Pacific Standard Time')
)
```

Self-Hosted IR

Self-Hosted

Use this for running data movement, external and pipeline activities in an on-premises / private network by installing the integration runtime.

Note: Data flows are only supported on Azure integration runtime. You can use self-hosted integration runtime to stage the data on cloud storage and then use data flows to transform it.

View less ∧

On-premises data gateway
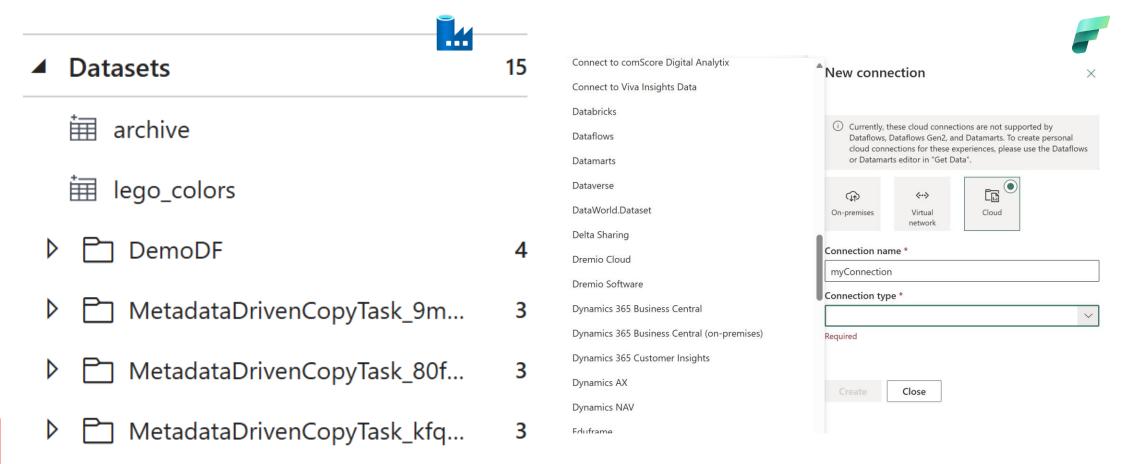
Status

Service Settings

Diagnostics

⊘ The gateway datagateway is online and ready to be used.

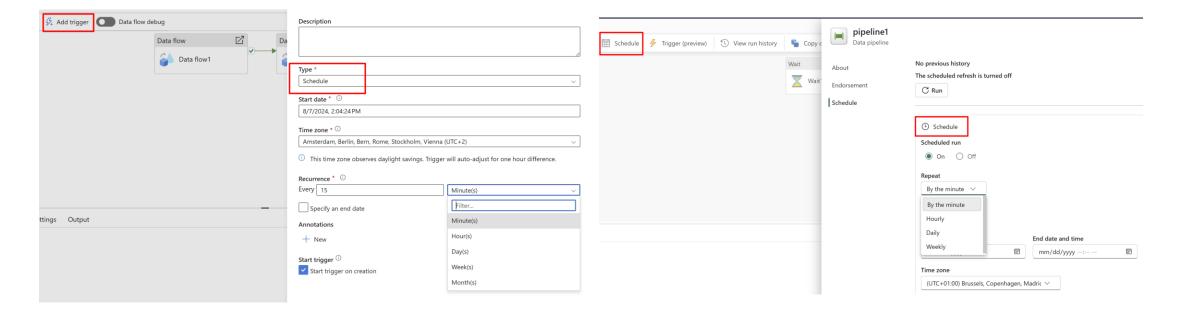Gateway version number: 3000.142.14 (September 2022)

Datasets & Linked Services

Left panel:

```
"source": {
    "type": "RestSource",
    "httpRequestTimeout": "00:01:40",
    "requestInterval": "00.00:00:00.010",
    "requestMethod": "GET",
    "additionalHeaders": {
        "Authorization": {
            "value": "@concat('Bearer ', activity('Get Bearer token').output.access_token)",
            "type": "Expression"
        }
    },
    "paginationRules": {
        "AbsoluteUrl": {
            "value": "@pipeline().parameters.p_pagination_rule",
            "type": "Expression"
        }
    },
    "datasetSettings": {
        "annotations": [],
        "type": "RestResource",
        "typeProperties": {
            "relativeUrl": "v1.0/groups?&$select=id,displayName,Description,groupTypes,MembershipType,membershipRule,mailEnabled,securityEnabled"
        },
        "schema": [],
        "externalReferences": {
            "connection": "2834579c-cba1-4ea2-a752-d7b40c6fe9a1"
        }
    }
},
"sink": {
    "type": "JsonSink",
    "storeSettings": {
        "type": "AzureBlobFSWriteSettings"
    },
    "formatSettings": {
        "type": "JsonWriteSettings"
    }
    "datasetSettings": {
        "annotations": [],
        "type": "Json",
        "typeProperties": {
            "location": {
                "type": "AzureBlobFSLocation",
                "fileName": {
                    "value": "@pipeline().globalParameters.gp_aad_groups_file_name",
                    "type": "Expression"
                },
                "folderPath": {
                    "value": "@pipeline().globalParameters.gp_aad_inbox_folder_name",
                    "type": "Expression"
                },
                "fileSystem": {
                    "value": "@pipeline().globalParameters.gp_aad_container",
                    "type": "Expression"
                }
            }
        },
        "schema": {},
        "externalReferences": {
            "connection": "4c4f6eed-8288-4780-b44c-25ec7a87b0fd"
        }
    }
},
"enableStaging": false
```

Right panel (line numbers 22-111):

```
22  "source": {
23      "type": "RestSource",
24      "httpRequestTimeout": "00:01:40",
25      "requestInterval": "00.00:00:00.010",
26      "requestMethod": "GET",
27      "additionalHeaders": {
28          "Authorization": {
29              "value": "@concat('Bearer ', activity('Get Bearer token').ou
30              "type": "Expression"
31          }
32      },
33      "paginationRules": {
34          "AbsoluteUrl": {
35              "value": "@pipeline().parameters.p_pagination_rule",
36              "type": "Expression"
37          }
38      }
39
40
41
42
43
44
45
46
47
48
49  }
50  },
51  "sink": {
52      "type": "JsonSink",
53      "storeSettings": {
54          "type": "AzureBlobFSWriteSettings"
55      },
56      "formatSettings": {
57          "type": "JsonWriteSettings"
58      }
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83  }
84  },
85  "enableStaging": false
86  },
87  "inputs": [
88      {
89          "referenceName": "ds_groups_aad_json",
90          "type": "DatasetReference",
91          "parameters": {
92              {
93              "relativeURL": {
94                  "value": "@pipeline().parameters.p_relative_url",
95                  "type": "Expression"
96              }
97          }
98      }
99  ],
100 "outputs": [
101     {
102         "referenceName": "ds_aad_groups_json",
103         "type": "DatasetReference",
104         "parameters": {
105             "timestamp": {
106                 "value": "@pipeline().parameters.p_timestamp",
107                 "type": "Expression"
108             },
109             "container_name": {
110                 "value": "@pipeline().globalParameters.gp_aad_container",
111                 "type": "Expression"
```
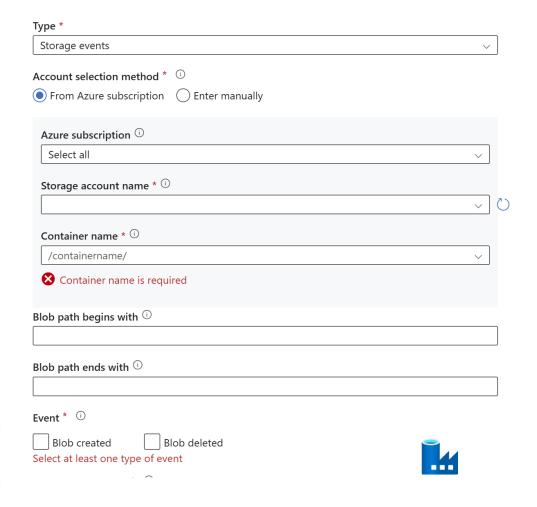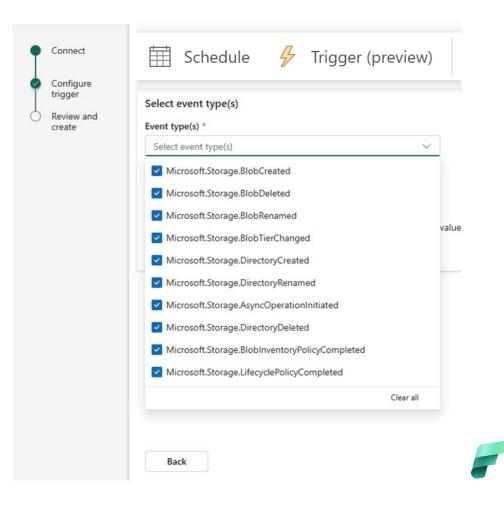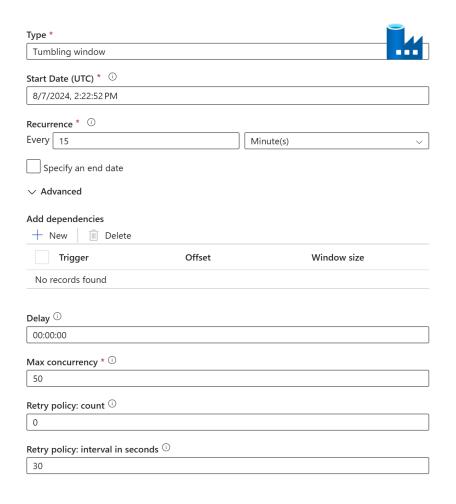
Azure

reMilano

Triggering

Storage Event

Type *

Storage events

Account selection method * ⓘ

⦿ From Azure subscription   ◯ Enter manually

Azure subscription ⓘ

Select all

Storage account name * ⓘ

Container name * ⓘ

/containername/

❌ Container name is required

Blob path begins with ⓘ

Blob path ends with ⓘ

Event * ⓘ

☐ Blob created   ☐ Blob deleted

Select at least one type of event

● Connect

✓ Configure trigger

◯ Review and create

▦ Schedule   ⚡ Trigger (preview)

Select event type(s)

Event type(s) *

Select event type(s)

☑ Microsoft.Storage.BlobCreated

☑ Microsoft.Storage.BlobDeleted

☑ Microsoft.Storage.BlobRenamed

☑ Microsoft.Storage.BlobTierChanged

☑ Microsoft.Storage.DirectoryCreated

☑ Microsoft.Storage.DirectoryRenamed

☑ Microsoft.Storage.AsyncOperationInitiated

☑ Microsoft.Storage.DirectoryDeleted

☑ Microsoft.Storage.BlobInventoryPolicyCompleted

☑ Microsoft.Storage.LifecyclePolicyCompleted

Clear all

Back

Tumbling Window Trigger

Type *

Tumbling window

Start Date (UTC) * ⓘ

8/7/2024, 2:22:52 PM

Recurrence * ⓘ

Every | 15   Minute(s) ▾

☐ Specify an end date

⌄ Advanced

Add dependencies

+ New | 🗑 Delete

☐	Trigger	Offset	Window size

No records found

Delay ⓘ

00:00:00

Max concurrency * ⓘ

50

Retry policy: count ⓘ

0

Retry policy: interval in seconds ⓘ

30
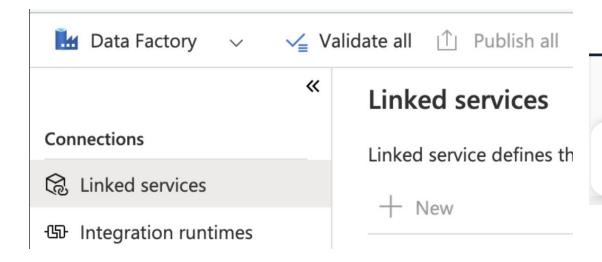
Decoupling Development

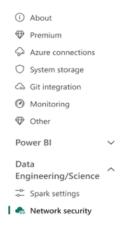# Managed Private Network/Endpoint
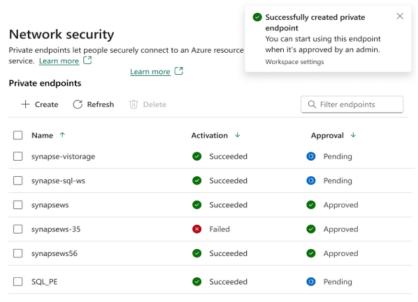
Some security stuff

- Azure Key Vault support
- Managed System Identity (MSI)

- Azure Key Vault announced in public preview
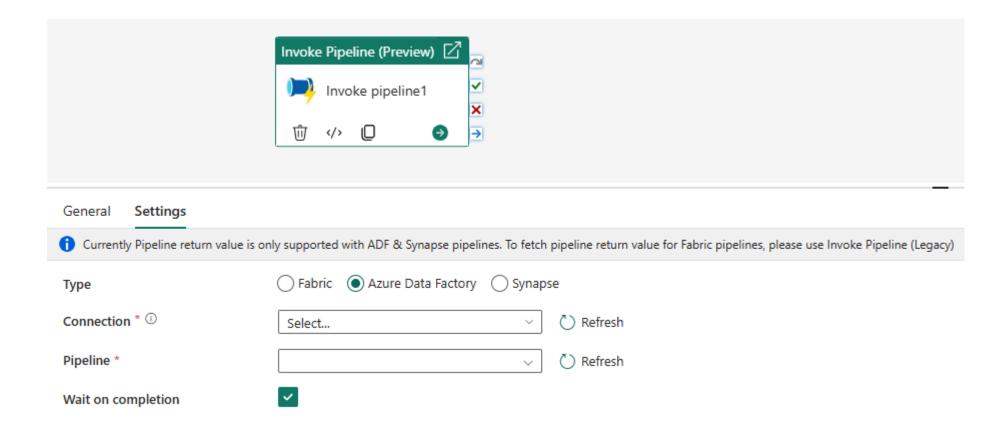- Workspace Identity only for ADLS Gen 2

ADF is not dead

Fabric is fast evolving platform, some of the content you're seeing can be changed.