

Transitioning from Azure Data Factory to Data Factory in Fabric

Riccardo Perico

BI & Power BI Engineer @ Lucient Italia



Sponsor



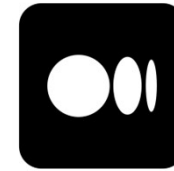
Bio



Lucient[®]
ITALIA



Riccardo Perico
BI & Power BI Engineer



[riccardo-perico](#)



[R1k91](#)



[R1k91](#)



[riccardo-perico](#)



[riccardo-perico](#)

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

Twitter Tweet

Facebook 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!

A computer monitor is shown with a white screen displaying black text. The screen is partially covered by a large, crisscrossed piece of yellow and black striped caution tape. The tape is draped over the monitor and a keyboard in front of it. The text on the screen is partially visible through the tape.

**unfortunately,
we no longer need
your services**

Agenda



What is equal?



What is it not equal?

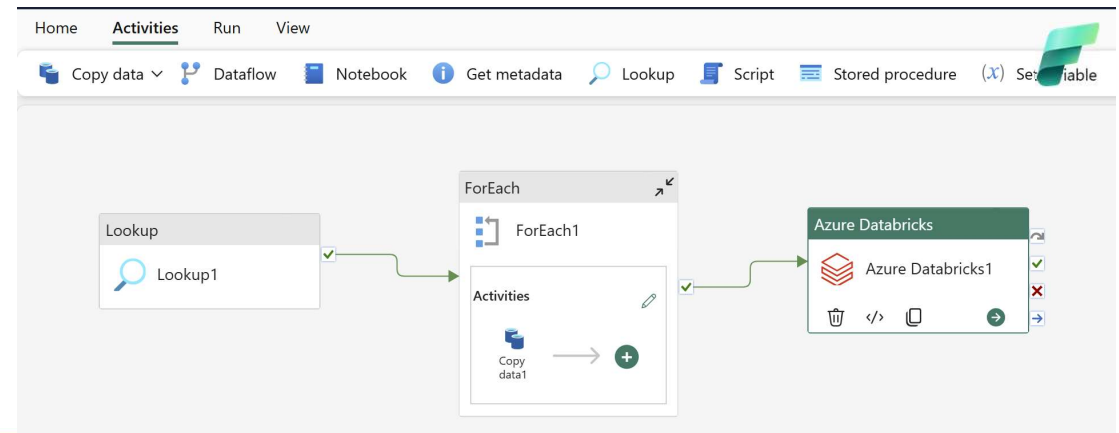
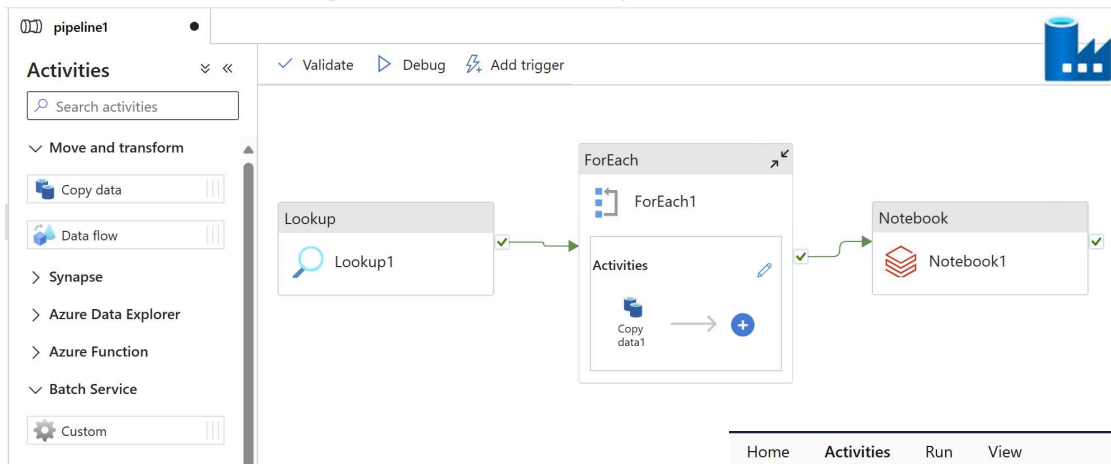


What is it improved?

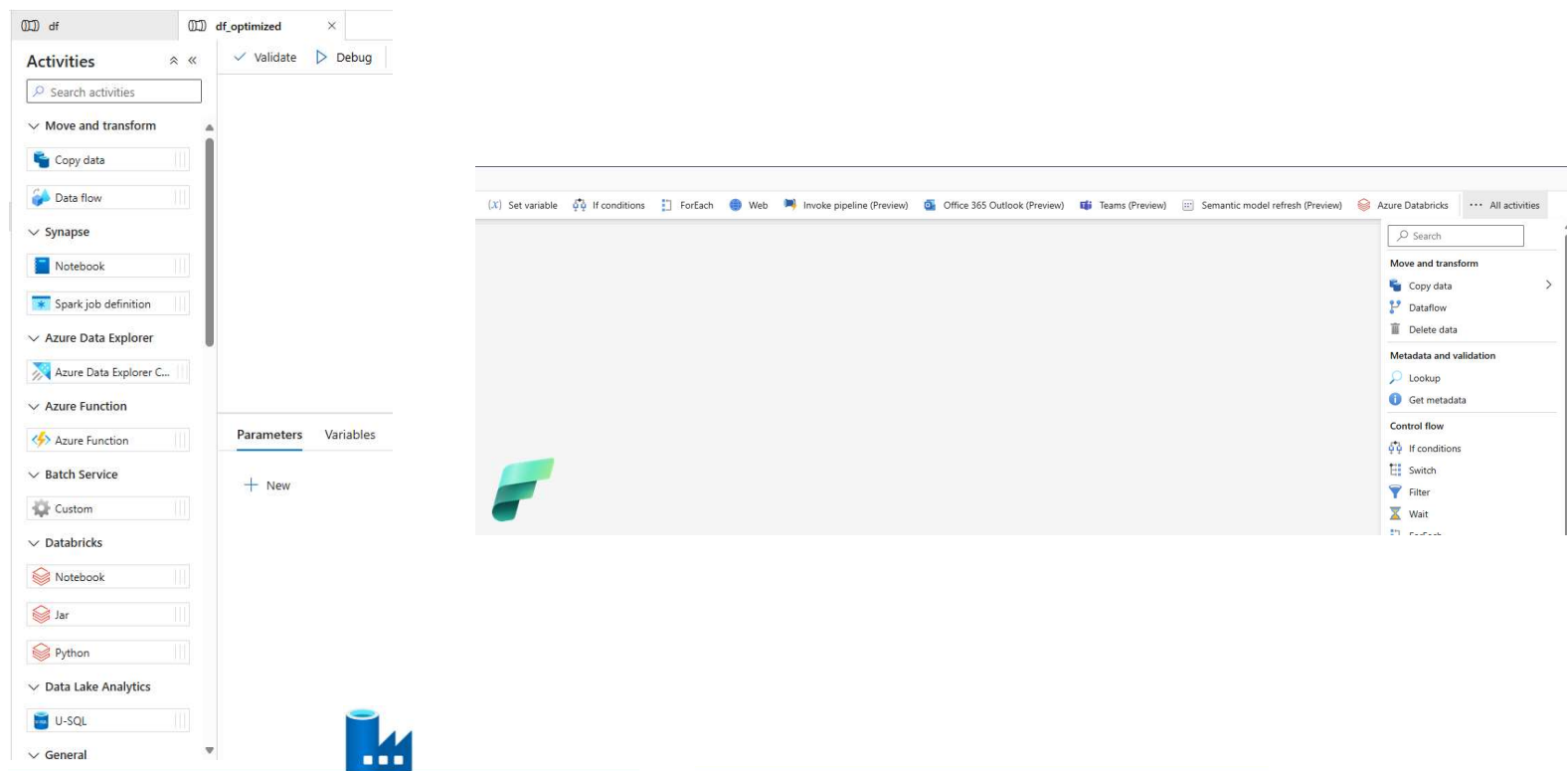


Fine delle trasmissioni

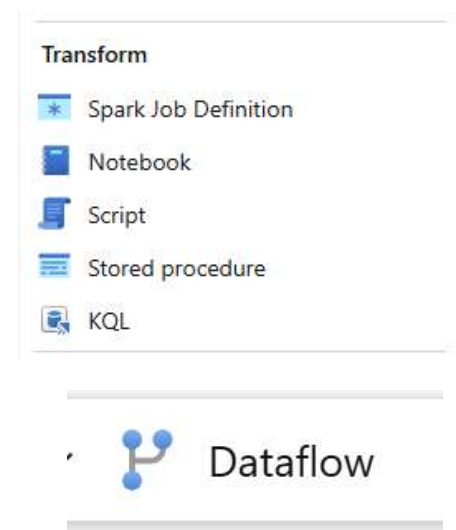
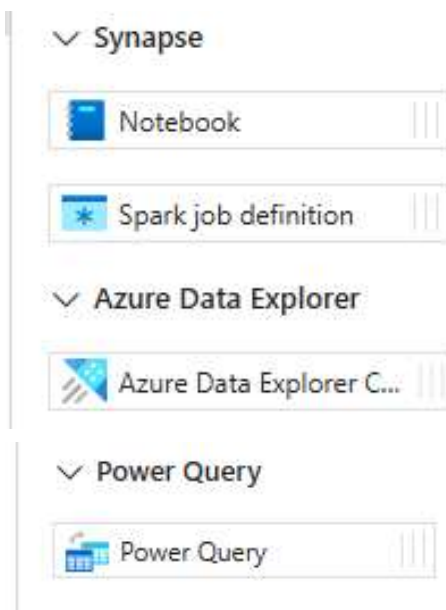
A pipeline is a logical grouping of activities that together perform a task



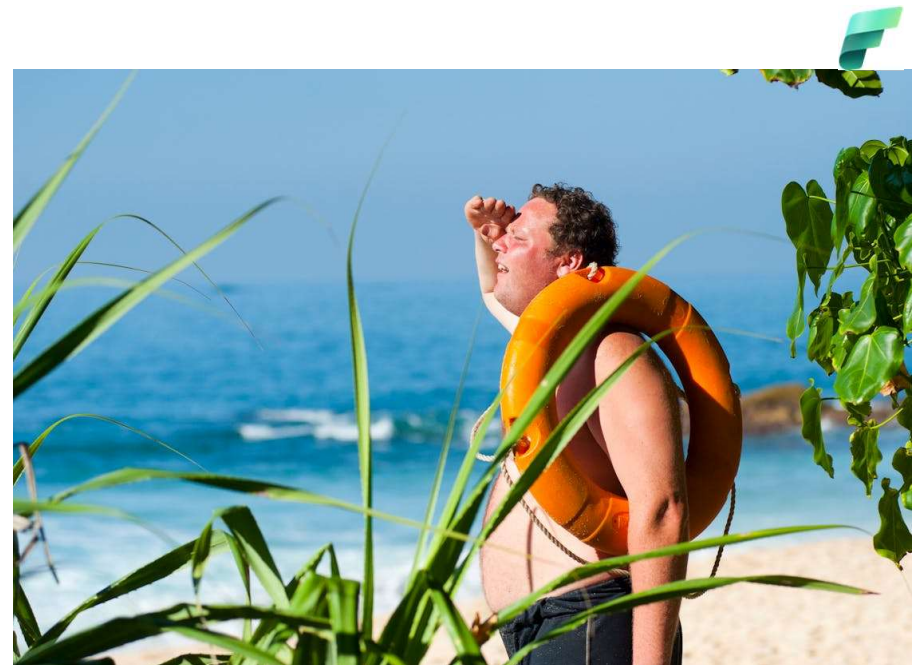
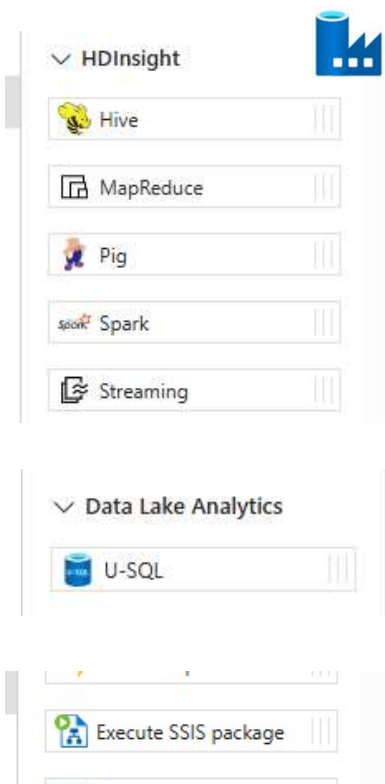
Activities are Activities



Some are different

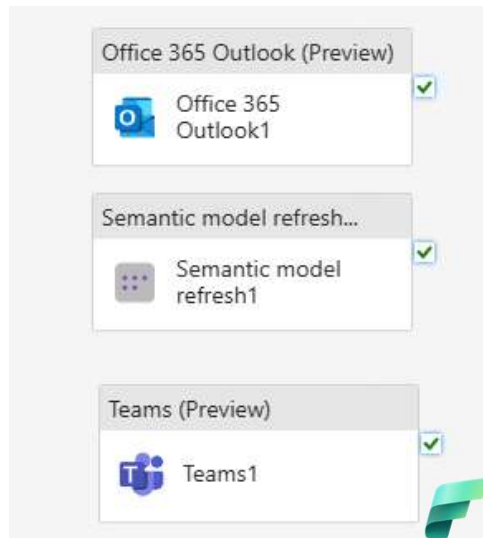


Some are missing



Some improve

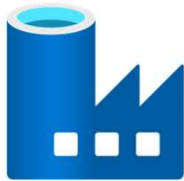
- Send an email
- Send a message in teams
- Refresh a Power BI model



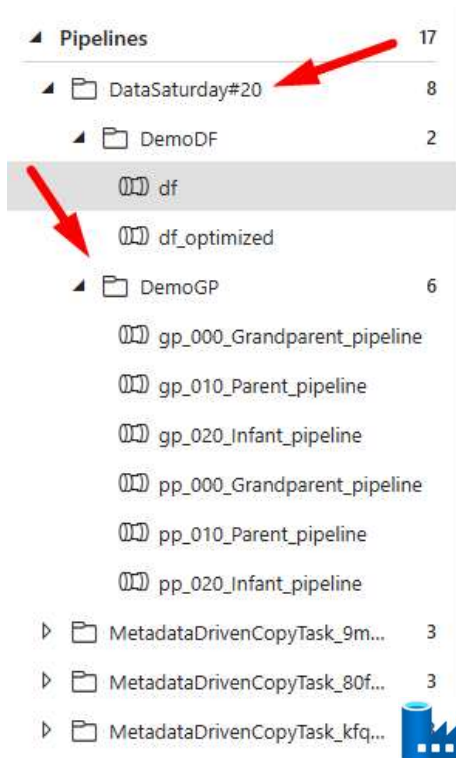
Parametrization

- Parent | Child pipeline
- Global Parameters

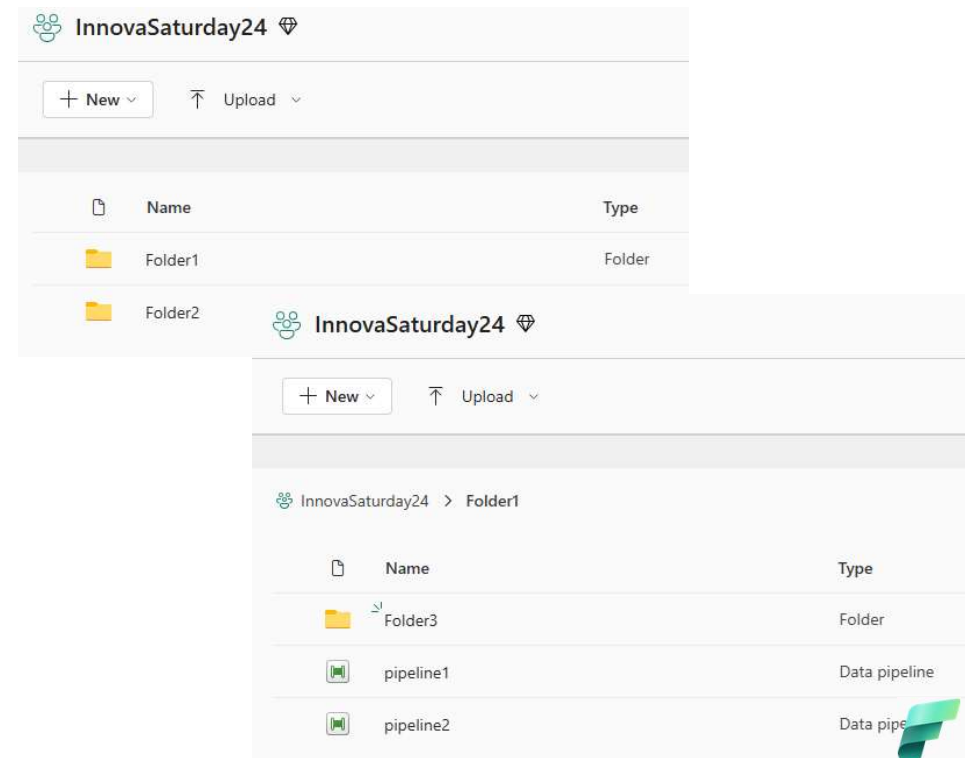
- Parent | Child pipeline



Pipeline organization



▲ Pipelines	17
▲ Folder DataSaturday#20	8
▲ Folder DemoDF	2
df	
df_optimized	
▲ Folder DemoGP	6
gp_000_Grandparent_pipeline	
gp_010_Parent_pipeline	
gp_020_Infant_pipeline	
pp_000_Grandparent_pipeline	
pp_010_Parent_pipeline	
pp_020_Infant_pipeline	
▶ Folder MetadataDrivenCopyTask_9m...	3
▶ Folder MetadataDrivenCopyTask_80f...	3
▶ Folder MetadataDrivenCopyTask_kfq...	

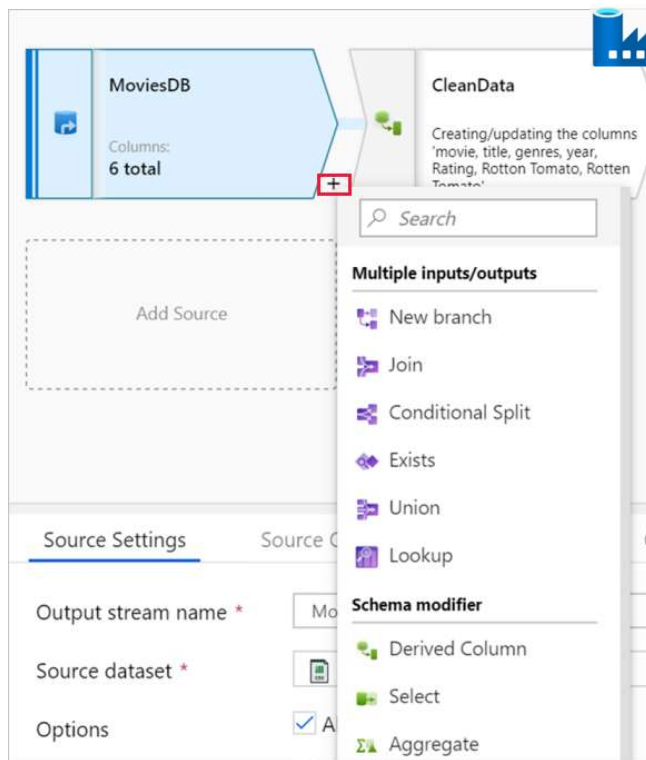


InnovaSaturday24	
+ New ↑ Upload	
Name	Type
Folder1	Folder
Folder2	

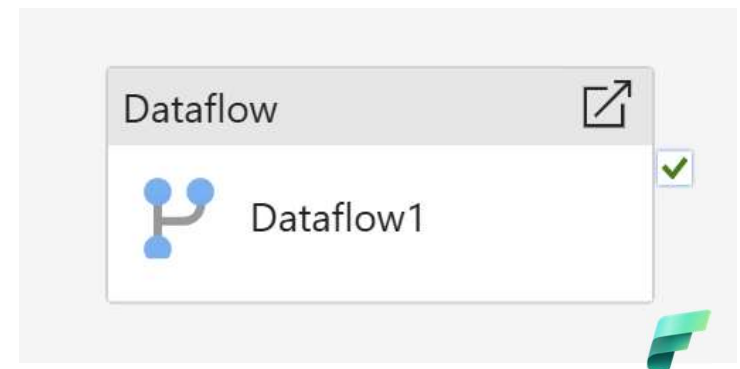
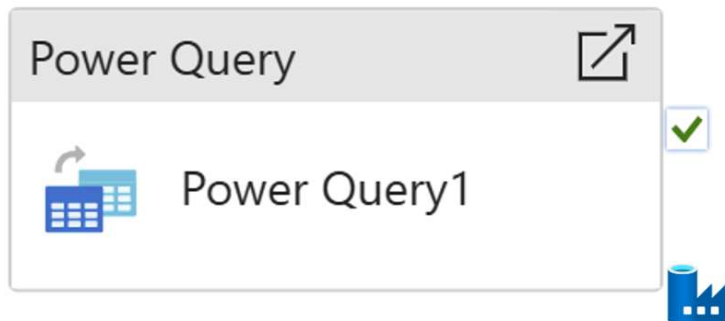
InnovaSaturday24 > Folder1	
+ New ↑ Upload	
Name	Type
Folder3	Folder
pipeline1	Data pipeline
pipeline2	Data pipe



Mapping Data Flows



Wrangling Data Flows / Power Query



Lakehouse & Warehouse Connector



	Innova_WH	Warehouse
	Innova_LH	Lakehouse



Variables

Parameters **Variables** Settings Output

+ New | Delete

<input type="checkbox"/> Name	Type	Default value
<input type="checkbox"/> myVar	String	Value

- String
- Boolean
- Array
- Integer

Parameters **Variables** Settings Output

+ New | Delete

<input type="checkbox"/> Name	Type	Default value
<input type="checkbox"/> myVar	String	Value

- String
- Boolean
- Array
- Integer



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

▲ Datasets



15

📅 archive

📅 lego_colors

▶ 📁 DemoDF

4

▶ 📁 MetadataDrivenCopyTask_9m...

3

▶ 📁 MetadataDrivenCopyTask_80f...

3

▶ 📁 MetadataDrivenCopyTask_kfq...

3

Connect to comScore Digital Analytix

Connect to Viva Insights Data

Databricks

Dataflows

Datamarts

Dataverse

DataWorld.Dataset

Delta Sharing

Dremio Cloud

Dremio Software

Dynamics 365 Business Central

Dynamics 365 Business Central (on-premises)

Dynamics 365 Customer Insights

Dynamics AX

Dynamics NAV

Erluframe

New connection



① Currently, these cloud connections are not supported by Dataflows, Dataflows Gen2, and Datamarts. To create personal cloud connections for these experiences, please use the Dataflows or Datamarts editor in "Get Data".

On-premises

Virtual network

Cloud

Connection name *

myConnection

Connection type *

Required

Create

Close



```

"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": "2894579c-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": "4c426ead-8288-4780-b44c-25ec7a87b0fd"
        }
      }
    }
  },
  "enableStaging": false
}

```

```

"source": {
  "type": "RestSource",
  "httpRequestTimeout": "00:01:40",
  "requestInterval": "00:00:00:00.010",
  "requestMethod": "GET",
  "additionalHeaders": {
    "Authorization": {
      "value": "@concat('Bearer ', activity('Get Bearer token').ou
      "type": "Expression"
    }
  },
  "paginationRules": {
    "AbsoluteUrl": {
      "value": "@pipeline().parameters.p_pagination_rule",
      "type": "Expression"
    }
  },
  "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": "4c426ead-8288-4780-b44c-25ec7a87b0fd"
        }
      }
    }
  },
  "enableStaging": false
},
"inputs": [
  {
    "type": "DatasetReference",
    "relativeUrl": {
      "value": "@pipeline().parameters.p_relative_url",
      "type": "Expression"
    }
  }
],
"outputs": [
  {
    "type": "DatasetReference",
    "timestamp": {
      "value": "@pipeline().parameters.p_timestamp",
      "type": "Expression"
    },
    "container_name": {
      "value": "@pipeline().globalParameters.gp_aad_container",
      "type": "Expression"
    }
  }
]

```



Triggering



Add trigger ☐ Data flow debug

Data flow

Data flow1

Type *

Schedule

Start date *

8/7/2024, 2:04:24 PM

Time zone *

Amsterdam, Berlin, Bern, Rome, Stockholm, Vienna (UTC+2)

☐ This time zone observes daylight savings. Trigger will auto-adjust for one hour difference.

Recurrence *

Every 15 Minute(s)

☐ Specify an end date

Annotations

+ New

Start trigger

☒ Start trigger on creation

Schedule ☐ Trigger (preview) ☐ View run history ☐ Copy

pipeline1

Data pipeline

About

Endorsement

Schedule

No previous history

The scheduled refresh is turned off

Run

Schedule

Scheduled run

☒ On ☐ Off

Repeat

By the minute

Hourly

Daily

Weekly

End date and time

mm/dd/yyyy --:--

Time zone

(UTC+01:00) Brussels, Copenhagen, Madrid



Storage Event

Type *
Storage events

Account selection method * ⓘ
☒ From Azure subscription ☐ Enter manually

Azure subscription ⓘ
Select all

Storage account name * ⓘ
/containername/

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

<input type="checkbox"/> 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



Data Factory   Validate all  Publish all

«

Connections



-  Linked services
-  Integration runtimes

Linked services

Linked service defines th

+ New

Home Activities Run View

     Validate



Managed Private Network/Endpoint



Managed private endpoints

Managed private endpoint uses a private IP address from within Managed Virtual Network to connect to an Azure resource or your own private link service. Connections using managed private endpoints listed below provide access to Azure resources or private link services. [Learn more](#)

[+ New](#) [Refresh](#)

Showing 0 - 0 of 0 items

NAME	PROVISIO...	T...	APPROVA...	T...	VNET NAME	T...	POSSIBLE ...	T...	LINKED RESOURCE ID	T...
------	-------------	------	------------	------	-----------	------	--------------	------	--------------------	------

Workspace settings

- About
- Premium
- Azure connections
- System storage
- Git integration
- Monitoring
- Other
- Power BI
- Data Engineering/Science
- Spark settings
- Network security

Network security

Private endpoints let people securely connect to an Azure resource service. [Learn more](#)

[Learn more](#)

Private endpoints

[+ Create](#) [Refresh](#) [Delete](#)

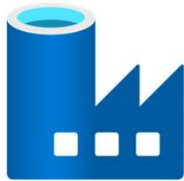
<input type="checkbox"/>	Name ↑	Activation ↓	Approval ↓
<input type="checkbox"/>	synapse-vistorage	✓ Succeeded	ⓘ Pending
<input type="checkbox"/>	synapse-sql-ws	✓ Succeeded	ⓘ Pending
<input type="checkbox"/>	synapsews	✓ Succeeded	✓ Approved
<input type="checkbox"/>	synapsews-35	✗ Failed	✓ Approved
<input type="checkbox"/>	synapsews56	✓ Succeeded	✓ Approved
<input type="checkbox"/>	SQL_PE	✓ Succeeded	ⓘ Pending

✓ Successfully created private endpoint
You can start using this endpoint when it's approved by an admin. Workspace settings



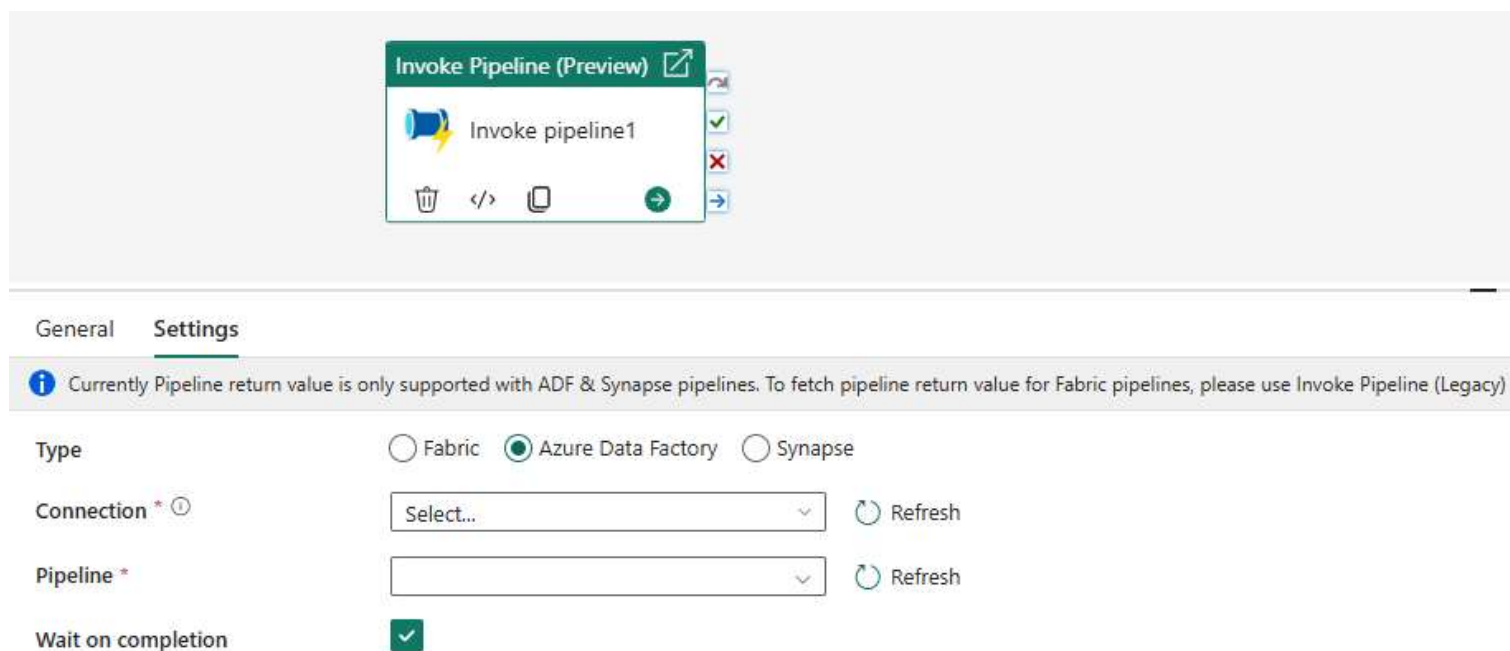
Some security stuff

- Azure Key Vault support
- Managed System Identity (MSI)
- Azure Key Vault not supported
- Workspace Identity only for ADLS Gen 2



There's good news...

ADF is not dead



Invoke Pipeline (Preview)

Invoke pipeline1

General Settings

i Currently Pipeline return value is only supported with ADF & Synapse pipelines. To fetch pipeline return value for Fabric pipelines, please use Invoke Pipeline (Legacy)

Type ☐ Fabric ☒ Azure Data Factory ☐ Synapse

Connection * Refresh

Pipeline * Refresh

Wait on completion ☒

Q&A

Thank You