Transitioning from Azure Data Factory to Data Factory in Fabric

Riccardo Perico

BI & Power BI Engineer @ Lucient Italia



Sponsor











Bio





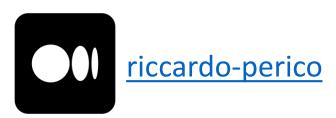








Riccardo Perico
BI & Power BI Engineer







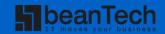




riccardo-perico











It seems like yesterday

Home / Updates / General availability: Azure Data Factory



General availability: Azure Data Factory

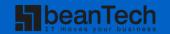
Published date: August 06, 2015

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



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 <u>Data Factory in Microsoft Fabric</u>, an all-in-one analytics solution for enterprises. <u>Microsoft Fabric</u> covers everything from data movement to data science, real-time analytics, business intelligence, and reporting. Learn how to <u>start a new trial</u> for free!













Agenda



What is equal?



What is it not equal?



What is it improved?



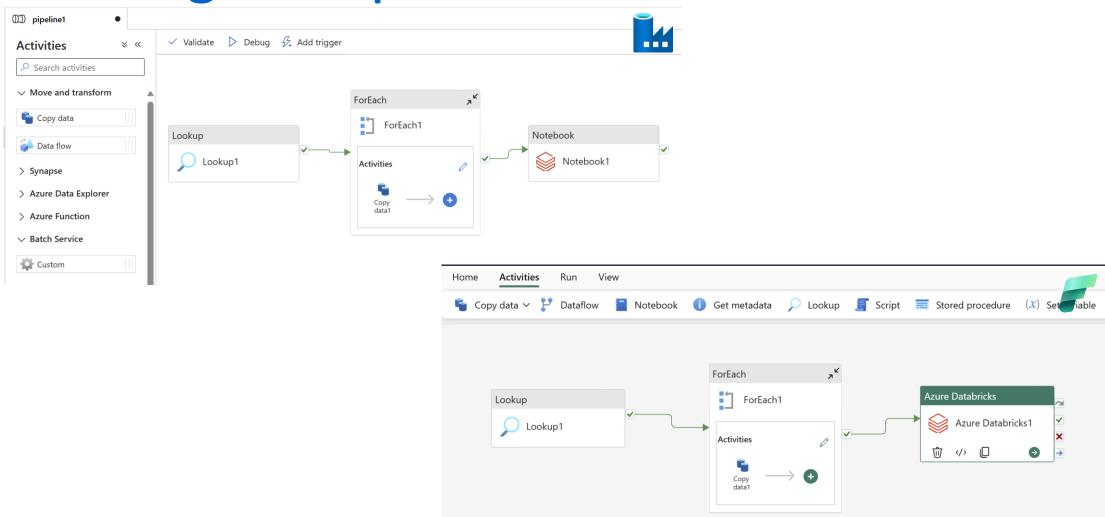








A pipeline is a logical grouping of activities that together perform a task







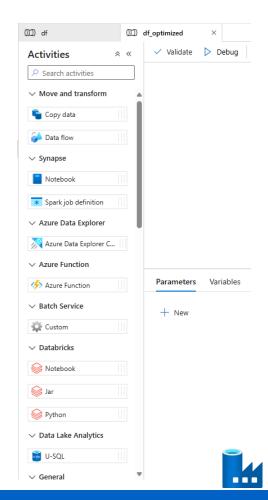


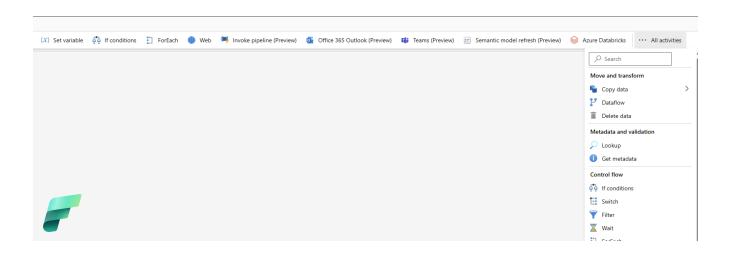






Activities are Activities









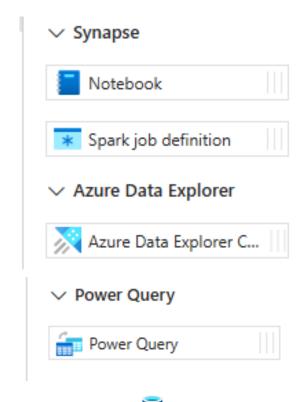


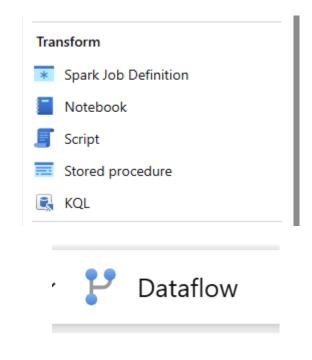






Some are different



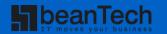








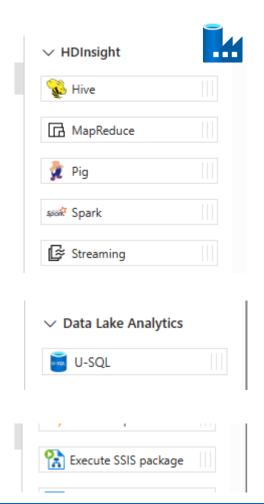


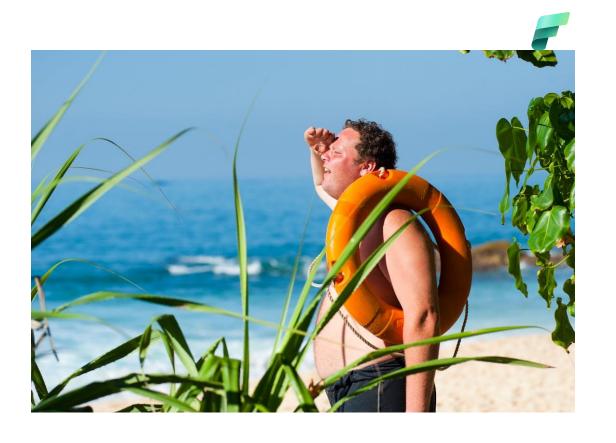






Some are missing









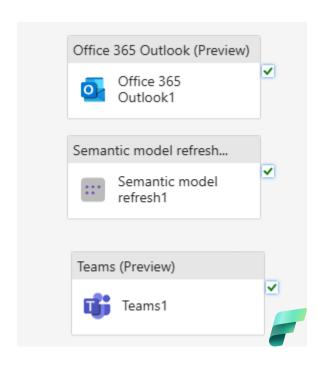








Some improve

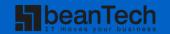


- Send an email
- Send a message in teams
- Refresh a Power BI model













Parametrization

- Parent | Child pipeline
- Global Parameters

Parent | Child pipeline









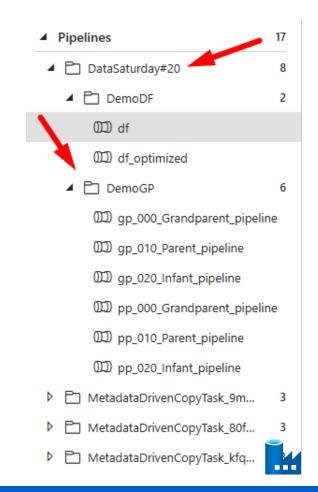


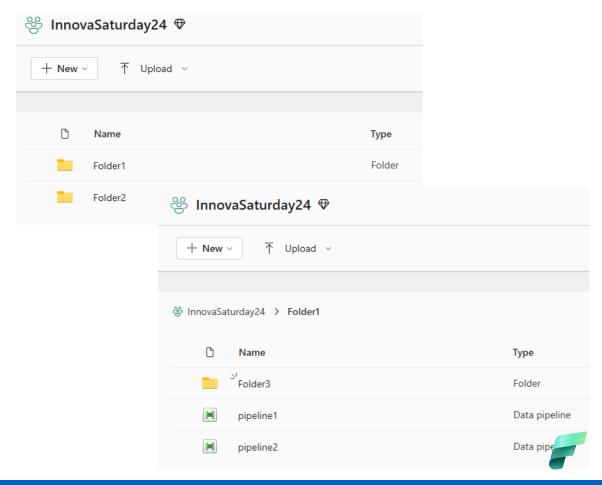






Pipeline organization









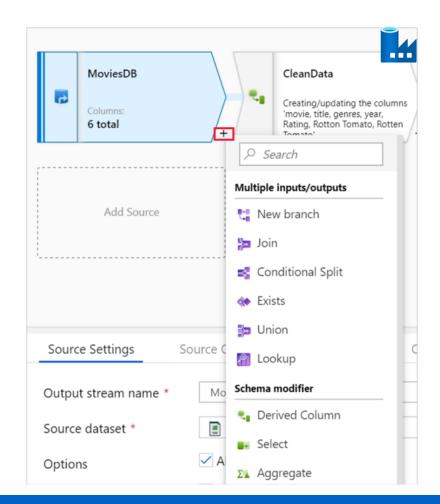








Mapping Data Flows









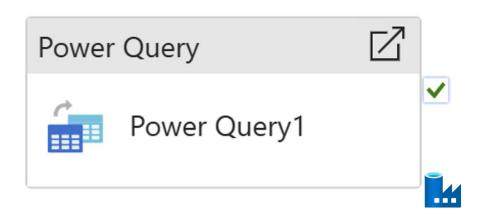


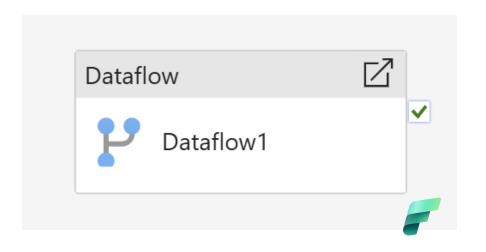






Wrangling Data Flows / Power Query













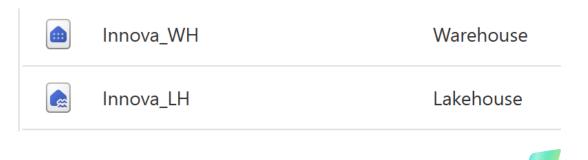




Lakehouse & Warehouse Connector











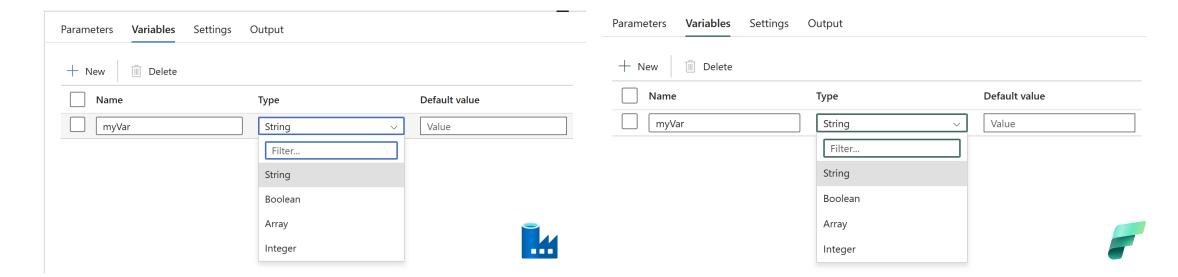








Variables















Expressions

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

















Self-Hosted IR



Self-Hosted

Use this for running data movement, external and pipeline activities in an onpremises / 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

used.

Gateway version number: 3000.142.14 (September 2022)







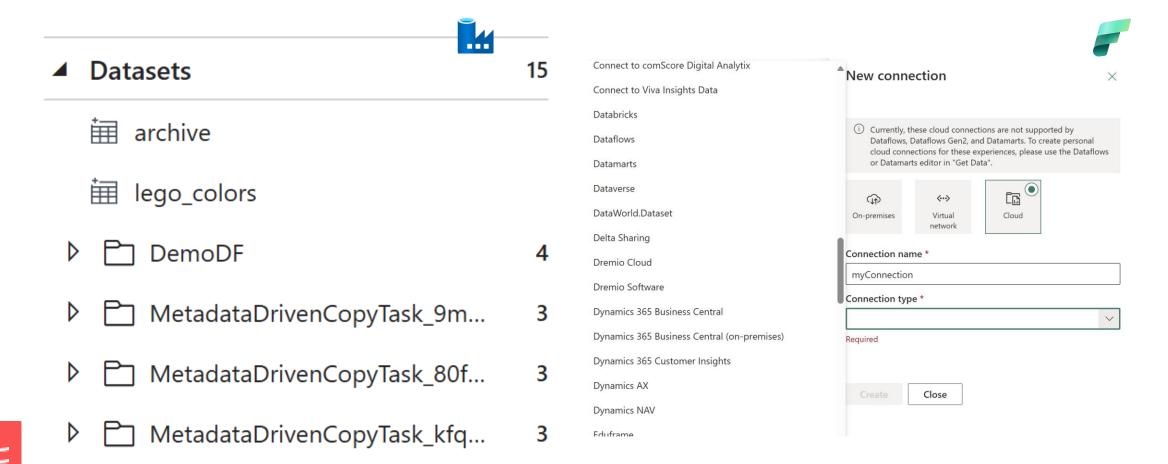








Datasets & Linked Services











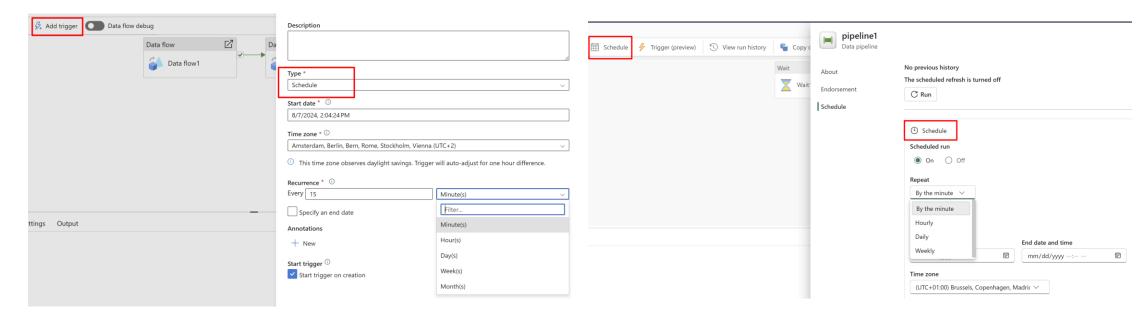


```
"source": {
    "type": "RestSource",
                                                                                                                                                                                   "type": "RestSource",
    "httpRequestTimeout": "00:01:40",
"requestInterval": "00.00:00:00.010",
                                                                                                                                                                                   "httpRequestTimeout": "00:01:40",
"requestInterval": "00.00:00:00.010",
    "requestMethod": "GET",
                                                                                                                                                                                   "requestMethod": "GET",
    "additionalHeaders":
                                                                                                                                                                                   "additionalHeaders":
        "Authorization":
                                                                                                                                                                                       "Authorization":
            "value": "@concat('Bearer ', activity('Get Bearer token').output.access_token)",
                                                                                                                                                                                           "value": "@concat('Bearer ', activity('Get Bearer token').ou
            "type": "Expression"
                                                                                                                                                                                           "type": "Expression"
    "paginationRules":
                                                                                                                                                                                   "paginationRules": {
        "AbsoluteUrl":
                                                                                                                                                                                       "AbsoluteUrl":
            "value": "@pipeline().parameters.p_pagination_rule",
                                                                                                                                                                                           "value": "@pipeline().parameters.p_pagination_rule",
            "type": "Expression"
                                                                                                                                                                                           "type": "Expression"
     datasetSettings":
        "annotations": [],
        "type": "RestResource",
        "typeProperties":
            "relativeUrl": "v1.0/groups?&$select=id,displayName,Descr.ption,groupTypes,MembershipType,membershipRule,mailEnabled,securityEnabled"
        "schema": [],
        "externalReferences": {
            "connection": "2834579c-cba1-4ea2-a752-d7b40c6fe9a1"
                                                                                                                                                              49
"sink": {
                                                                                                                                                                               "sink":
    "type": "JsonSink",
                                                                                                                                                                                   "type": "JsonSink",
    "storeSettings":
                                                                                                                                                                                   "storeSettings": {
        "type": "AzureBlobFSWriteSettings"
                                                                                                                                                                                       "type": "AzureBlobFSWriteSettings"
    "formatSettings": {
                                                                                                                                                                                   "formatSettings": {
        "type": "JsonWriteSettings"
                                                                                                                                                                                       "type": "JsonWriteSettings"
     "datasetSettings":
        "annotations": [],
        "type": "Json",
        "typeProperties": {
            "location":
                "type": "AzureBlobFSLocation",
                "fileName": {
                    "value": "@pipeline().globalParameters.gp_aad_groups_file_name",
                    "type": "Expression"
                    "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
                                                                                                                                                                               "enableStaging": false
                                                                                                                                                                           "inputs": [
                                                                                                                                                                                   "type": "DatasetReference",
                                                                                                                                                                                       "relativeURL":
                                                                                                                                                                                           "value": "@pipeline().parameters.p_relative_url",
                                                                                                                                                                                           "type": "Expression"
                                                                                                                                                                           "outputs": [
                                                                                                                                                                                   "type": "DatasetReference",
                                                                                                                                                            103
                                                                                                                                                            105
                                                                                                                                                                                           "value": "@pipeline().parameters.p_timestamp",
                                                                                                                                                                                           "type": "Expression"
                                                                                                                                                             108
                                                                                                                                                                                       "container_name": {
                                                                                                                                                                                           "value": "@pipeline().globalParameters.gp_aad_container",
                                                                                                                                                                                           "type": "Expression"
```

Triggering











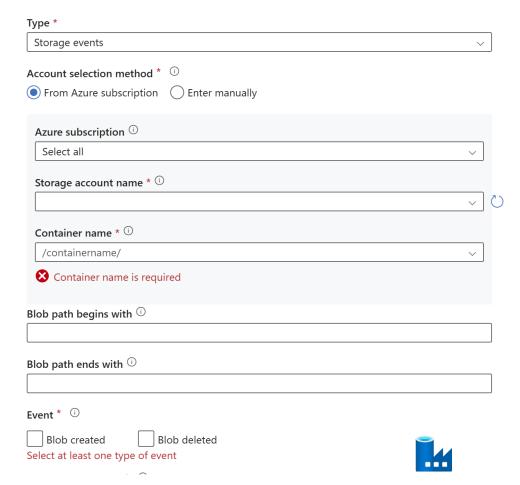


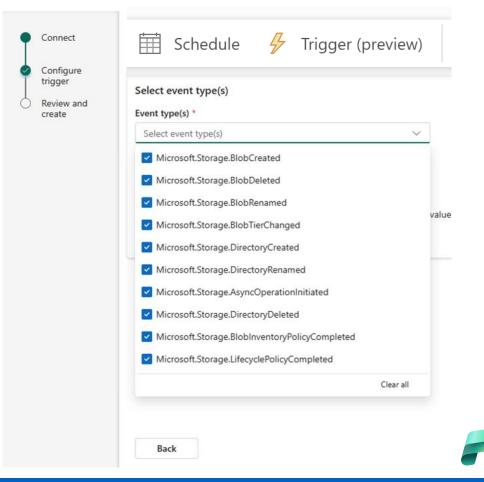






Storage Event









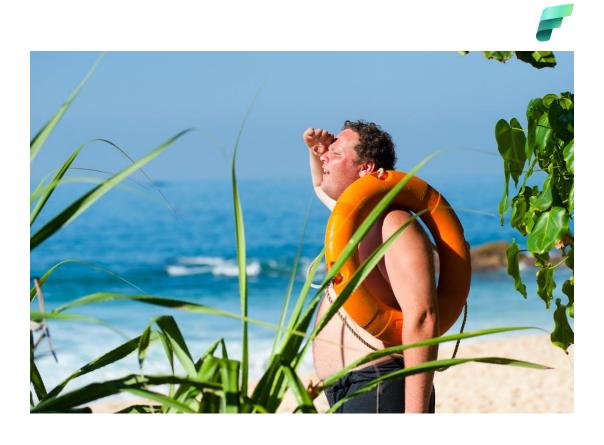






Tumbling Window Trigger

Type *				
Tumbling window				
Start Date (UTC) * (i)				
8/7/2024, 2:22:52 PM				
- + 0				
Recurrence * ①				
Every 15		Minute(s)		~
Specify an end date				
√ Advanced				
Add dependencies				
+ New Delete				
Trigger	Offset		Window size	
No records found				
Delay ①				
00:00:00				
Max concurrency * ①				
50				
D-4				
Retry policy: count ①				
0				
Retry policy: interval in seconds (i)				











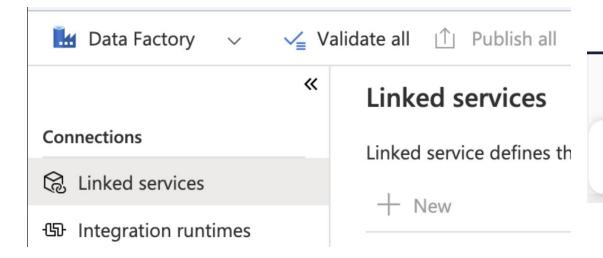


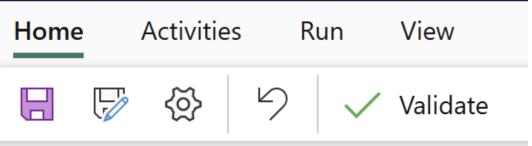


Decoupling Development

















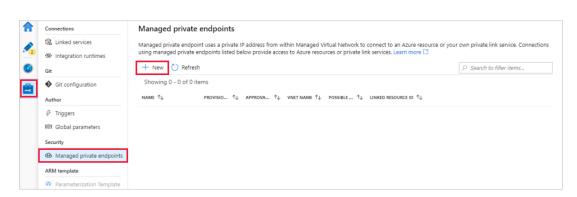


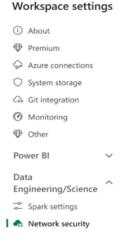


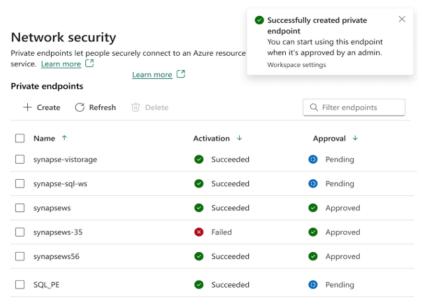
Managed Private Network/Endpoint























Some security stuff

- Azure Key Vault support
- Managed System Identity (MSI)
- Azure Key Vault no supported
- Workspace Identity only for ADLS Gen 2













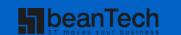




There's good news...



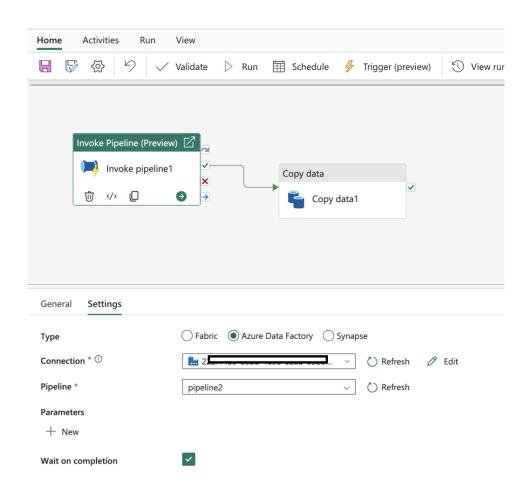








ADF is not dead



New item

Select an item type



Prepare data

Clean, transform, extract, and load your data for analysis and modeling tasks.













Q&A



Thank You