

Introduction to Fabric Data Pipelines

Koen Verbeeck – He/him

ae

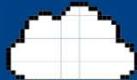
Contact



Koen Verbeeck



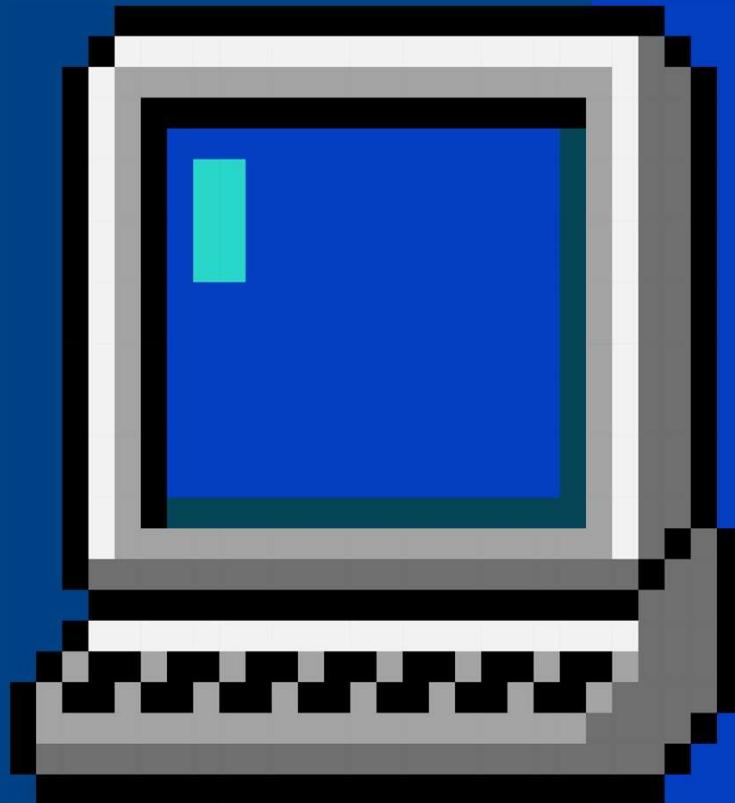
@Ko_Ver



@koenv.bsky.social



SQLkover.com

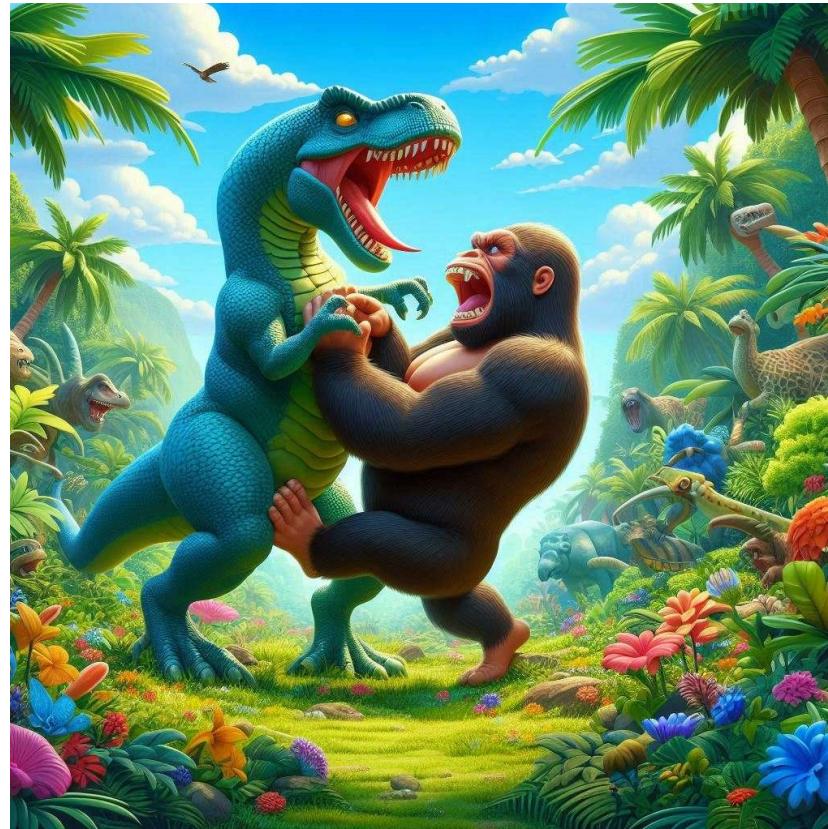


ae

Outline

- 
- 01** What are Fabric Data Pipelines?
 - 02** What are the building blocks?
 - 03** Fabric Pipelines vs ADF
 - 04** Tips & Tricks

Quick Poll



What are Fabric Data Pipelines?

de



Microsoft Fabric

ae



Data
Factory



Real-Time
Intelligence



Databases



Analytics



Industry
Solutions



Power BI



Partner
solutions



Copilot in Fabric



OneLake



Microsoft Purview

A *fully managed, serverless data integration* service. **Visually** integrate data sources with more than 170 **built-in connectors** at no added cost. Easily construct **ETL** and **ELT** processes code-free in an *intuitive environment or write your own code.*

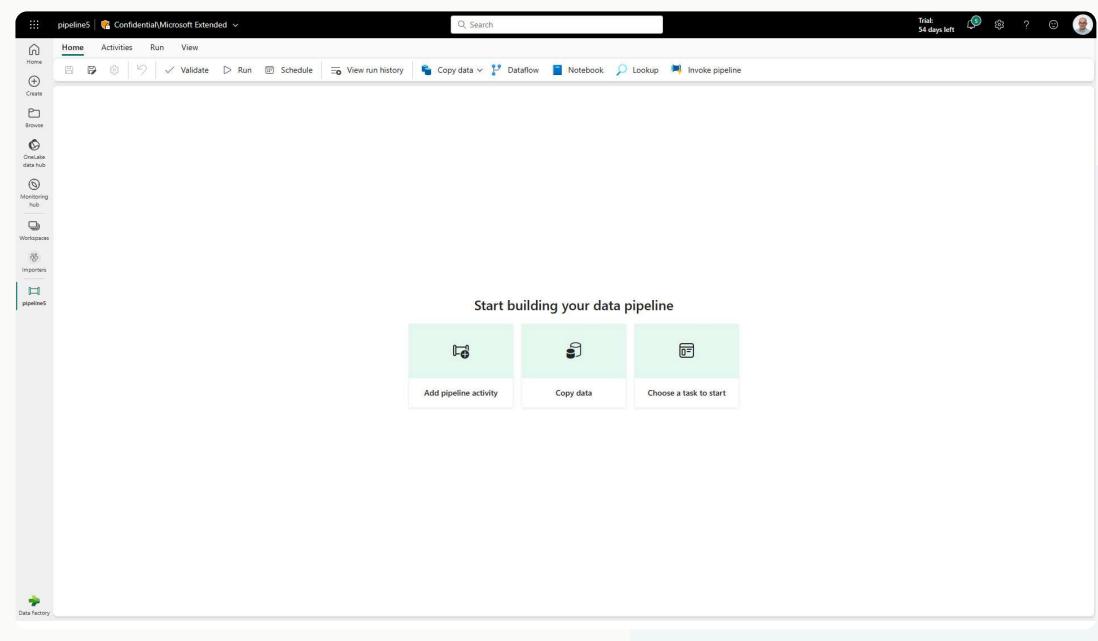
Data Factory

Data Factory in Microsoft Fabric provides cloud-scale data movement and data transformation services that allows you to solve the most complex ETL scenarios

Core to Data Factory are Data Pipelines and Dataflows to give users the option to a low-code, collaborative and enterprise scale approach for their ETL process

Key Capabilities:

- Latest capabilities:
- Output destination to Lakehouse
- 170+ connectors available in Data Factory
- Pipeline Lakehouse copy assist
- Create data pipeline in Lakehouse portal
- Pipeline templates
- Pipeline support for Spark notebooks
- Service principle auth support

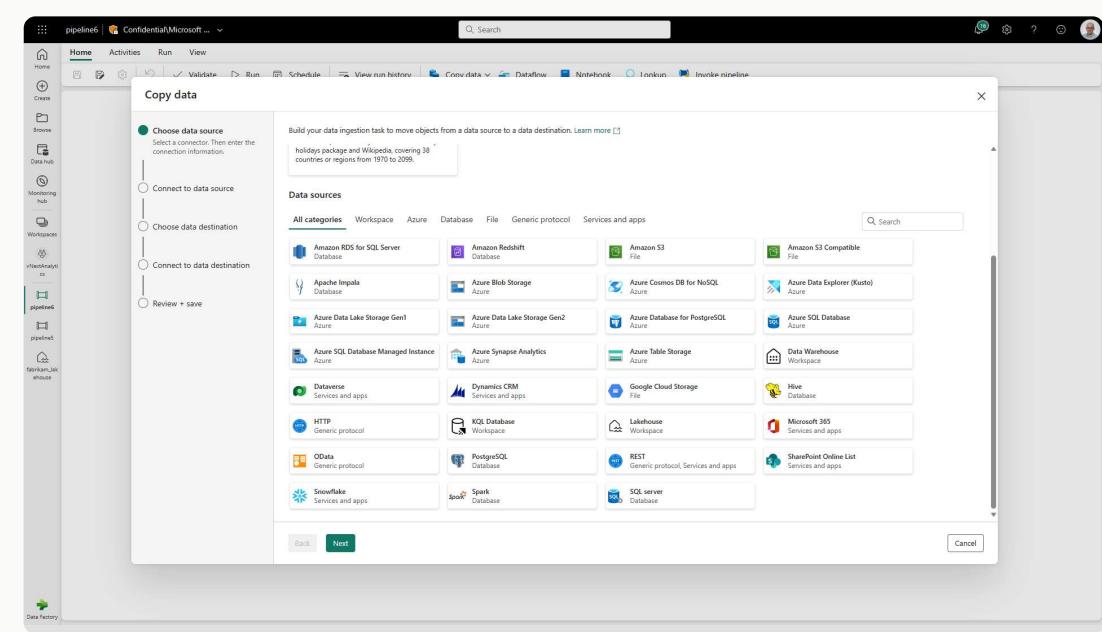


Autonomous ETL can unlock operational efficiencies and help orchestrate, monitor and manage pipeline performance.

Data Pipelines

Data Pipelines enable powerful workflow capabilities at cloud-scale like building complex workflows, moving PB-size data, and defining sophisticated control flow pipelines

Data pipelines can be used to build complex ETL and data factory workflows that can perform a number of different tasks at scale. Additionally, control flow capabilities are built into pipelines so you can build workflow logic which provide loops and conditional

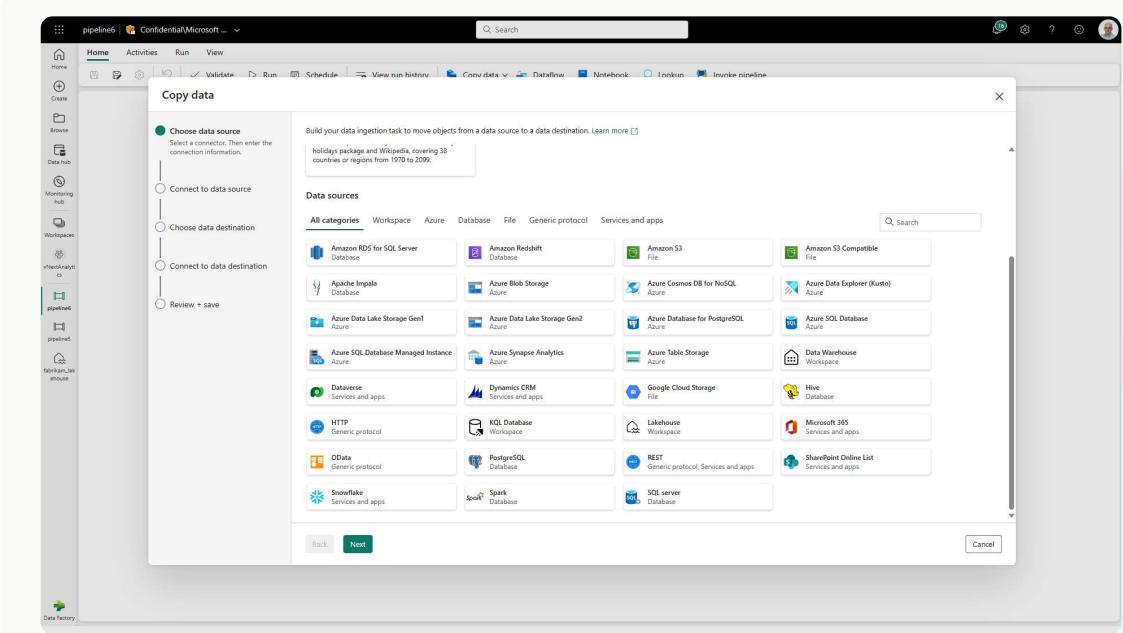


Data Pipelines | Connectors

New Connectors provide a low-code interface for ingesting data from a variety of data sources

Connectors:

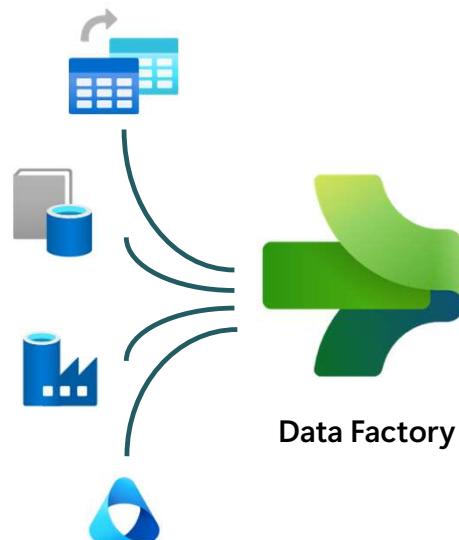
- Warehouse Connector; connect to existing Azure
- Lakehouse connector
- 100+ connectors in the copy activity
- Access to on-premises data
- Access protected data inside of a VNET



Unifying data in OneLake

Data Factory

Seamlessly connect to more than 170+ data stores



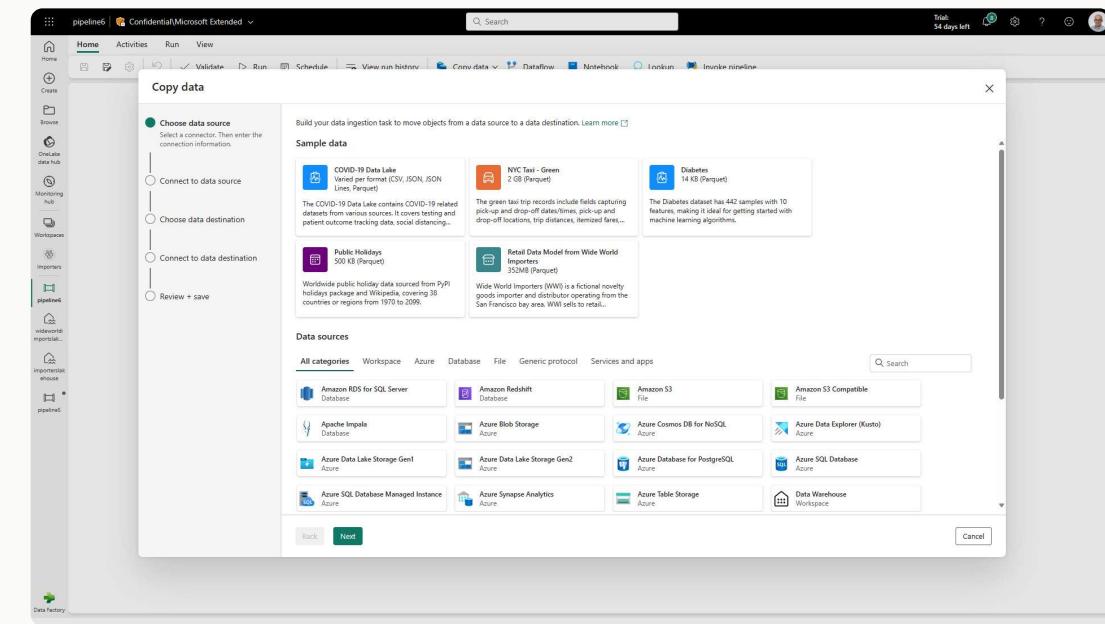
Azure Database for PostgreSQL	Azure Databricks Delta Lake	Amazon RDS for Oracle	Amazon RDS for SQL Server	Amazon Redshift	Phoenix	PostgreSQL	Presto	Magento (Preview)
Azure SQL Database	Azure SQL Database Managed Instance	Apache Impala	Azure SQL Database Managed Instance	DB2	SAP BW	SAP BW	SAP HANA	Oracle Eloqua (Preview)
Azure Table Storage	MongoDB Atlas	Drill	Google AdWords	Google BigQuery	SAP Table	SQL server	Spark	PayPal (Preview)
Azure Cosmos DB (MongoDB API)	Azure Cosmos DB (SQLAPI)	Greenplum	HBase	Hive	Amazon S3	Amazon S3 Compatible	FTP	SAP Cloud For Customer
Azure Data Lake Storage Gen1	Azure Data Lake Storage Gen1 for Cosmos Structured Stream	IBM	Informix	MariaDB	Microsoft Access	File system	Google Cloud Storage (S3APD)	HDFS
Azure Data Lake Storage Gen1 for Cosmos Structured Stream	Azure Database for MariaDB	MySQL	Netezza	Oracle	HTTP	Oracle Cloud Storage (S3AP)	SFTP	Salesforce Marketing Cloud
teradata	VERTICA	ODBC	OData	REST	Amazon Marketplace Web Service	Concur (Preview)	Dataverse (Common Data Service for App)	Web Table
Jira	Kusto	SharePoint Online List	Dynamics 365	Dynamics AX	Dynamics CRM	Cassandra	Couchbase (Preview)	MongoDB

Data Pipelines | Sample data

Sample Datasets helps new users get started quickly, building out their ELT processes using Data Pipelines

Sample datasets:

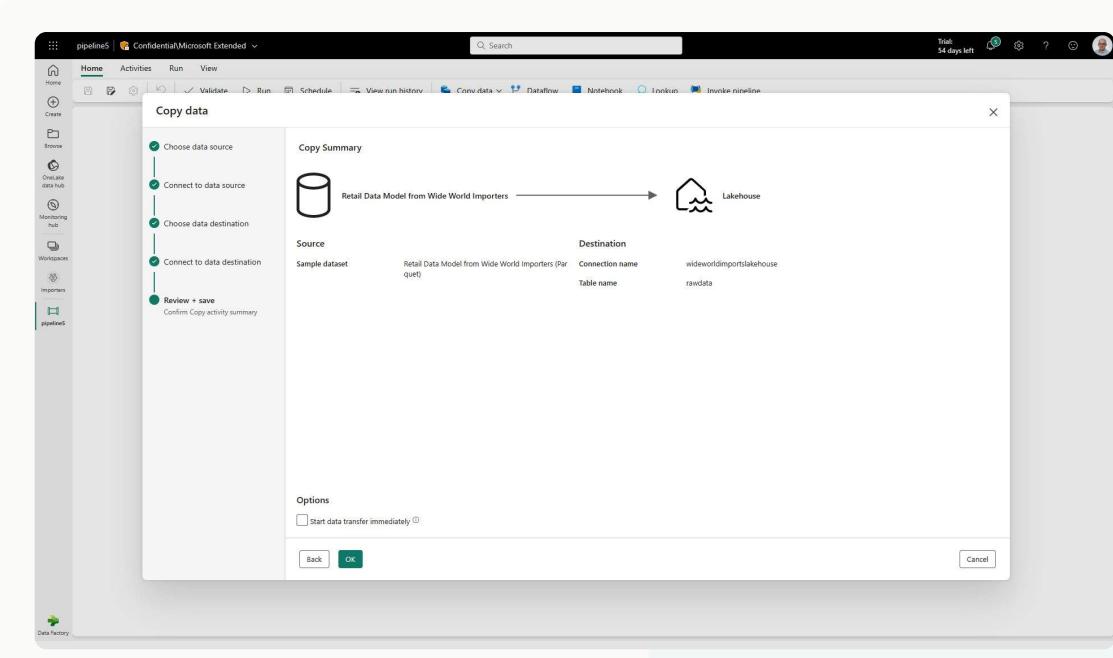
- COVID-19 Data Lake (CSV, JSON, JSON Lines, Parquet)
- NYC Tax – Green (2GB Parquet)
- Diabetes (14K Parquet)
- Public Holidays (500KB Parquet)
- Retail Data Model from Wide World Importers (352MB Parquet)



Data Pipelines | Lakehouse copy assist

Simply copying data to a Lakehouse with copy assist capabilities within the Data Pipeline

Additionally, users can create a Data Pipeline without having to leave the Lakehouse portal



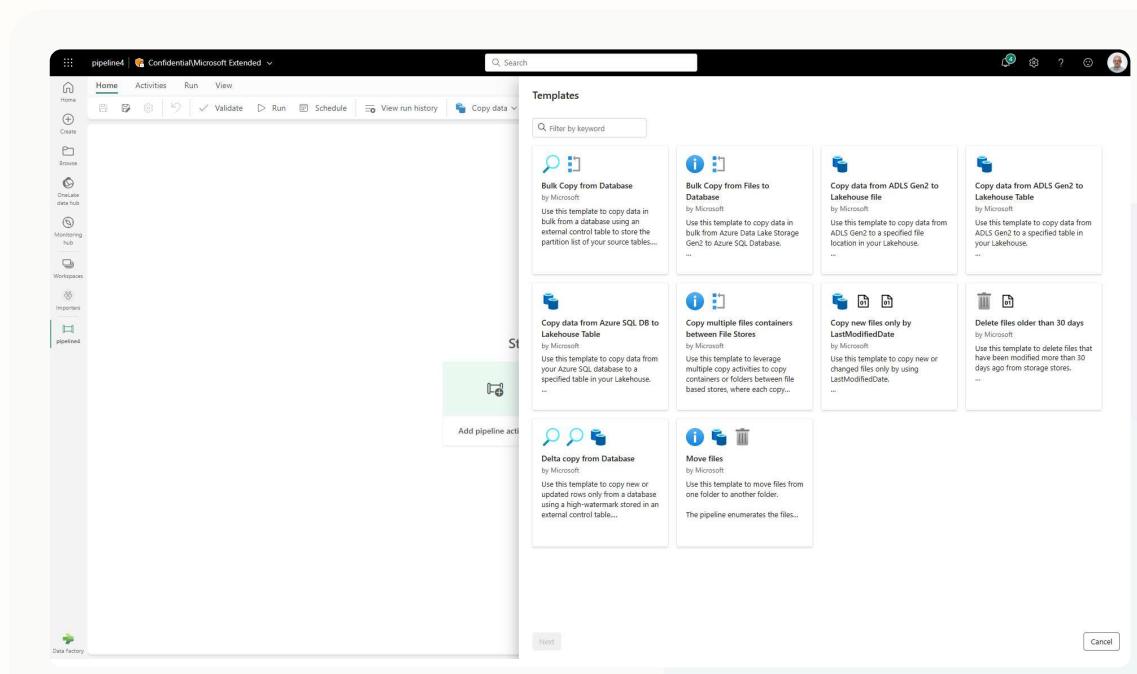
Data Pipelines | Templates

Quickly get started with data integration

Template help reduce development time by providing an easy way to create pipeline for common data integration scenarios

Available Data Pipeline Templates:

- Bulk copy from Database
- Bulk copy from File to Database
- Copy data from ADLS Gen2 to Lakehouse file
- Copy from ADLS Gen2 to Lakehouse Table
- Copy data from Azure AQL DB to Lakehouse Table
- Copy multiple files containers between File Stores
- Copy new files only by Last Modified Date
- Delete files older than 30 days
- Delta copy from Database
- Move files





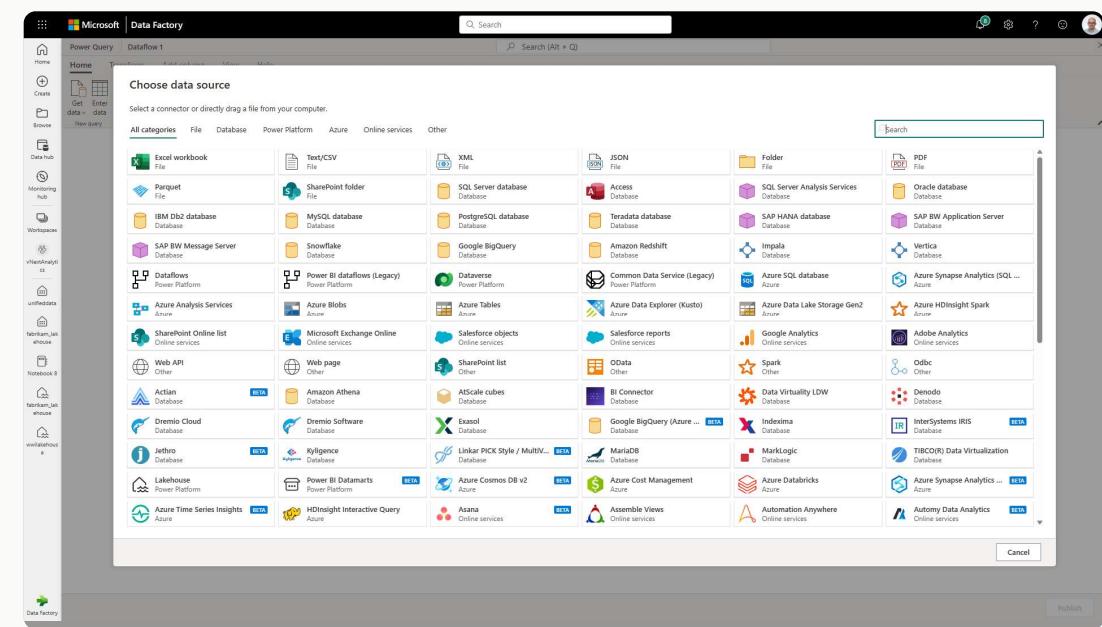
Dataflow

Dataflow provides a low-code interface for ingesting data from hundreds of data sources

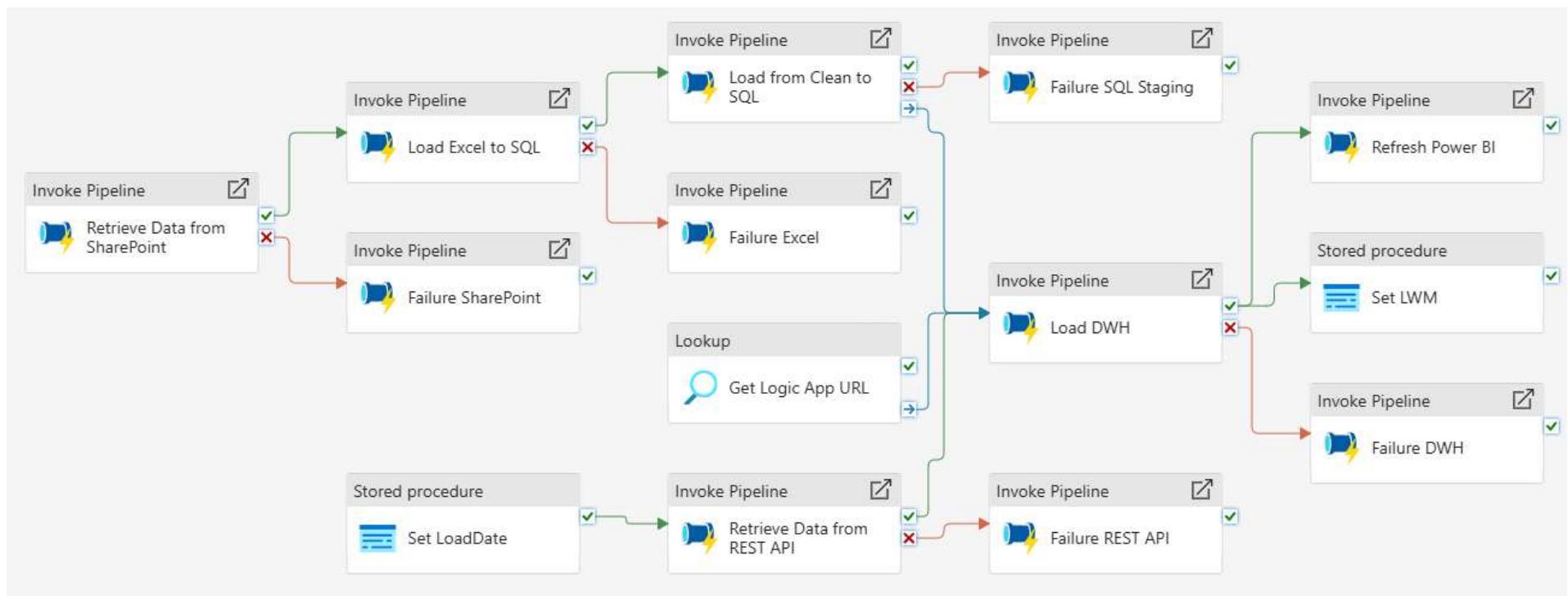
Dataflow quickly and easily unify disparate data sources, establish a more collaborative analytics approach, and promote more informed, agile decision making.

Key Capabilities:

- Accelerate data transformation with code-free data flows
 - Scale out using Fabric compute and Data Factory fast copy
 - Load results of data transformations into multiple destinations (Azure SQL Databases, Lakehouse, etc.)



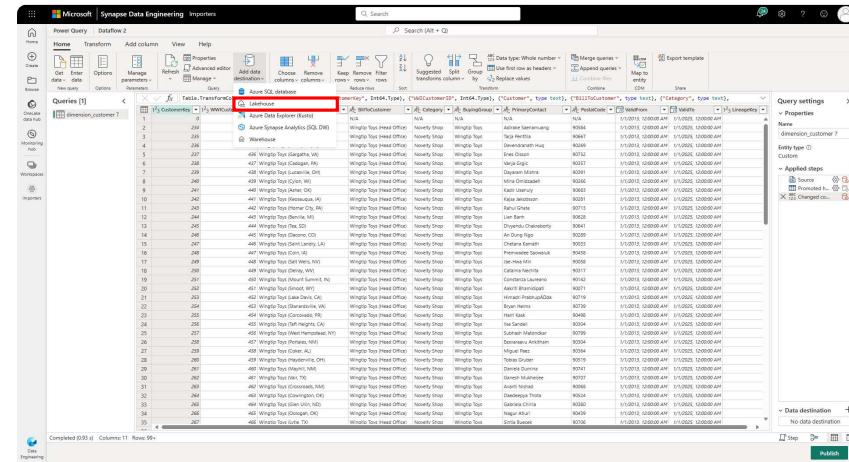
Basically a cloud ELT orchestrator



What about ETL?

- Transformations are done in other Fabric Compute
 - SQL in Warehouse
 - Spark/SQL in Notebook (or Spark Job)
 - KQL in Eventhouse
 - AI stuff in shortcuts?

- Dataflows Gen 2
 - Aka Power Query on steroids
 - Many improvements w.r.t. Power BI Dataflows
 - But not really suited for scale?



demo

What are the building blocks?

de

Pipeline

New item

 Favorites

 All items

Get data

Ingest batch and real-time data into a single location within your Fabric workspace.

Pipeline

Ingest data at scale and schedule data workflows.



Prepare data

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

Pipeline

Ingest data at scale and schedule data workflows.



Build a pipeline to organize and move your data

Start with a blank canvas



Pipeline activity

Automate data orchestrations using rich no-code activities.

Start with guidance



Copy data assistant

Follow guided steps to copy data into Microsoft Fabric, as well as other data stores.



Practice with sample data

Quickly build a pipeline with a predefined template to load data into Lakehouse.



Templates

Generate a new pipeline quickly using a predefined data scenario.

Need help? Ask Copilot 

Activities

Move and transform

 Copy data >

 Copy job

 Dataflow

 Delete data

Metadata and validation

 Lookup

 Get metadata

Control flow

 If conditions

 Switch

 Filter

 Wait

 ForEach

 Until

 Set variable

 Append variable

 Fail

Orchestrate

 Invoke Pipeline

 Invoke Pipeline (Legacy)

 Web

 WebHook

 Semantic model refresh

 Azure Databricks

 Functions

 Azure HDInsight

 Azure Batch

Notifications

 Office 365 Email

 Microsoft Teams

 Office 365 Outlook (Legacy)

 Teams (Legacy)

Transform

 Spark Job Definition

 Notebook

 Script

 Stored procedure

 KQL

Machine Learning

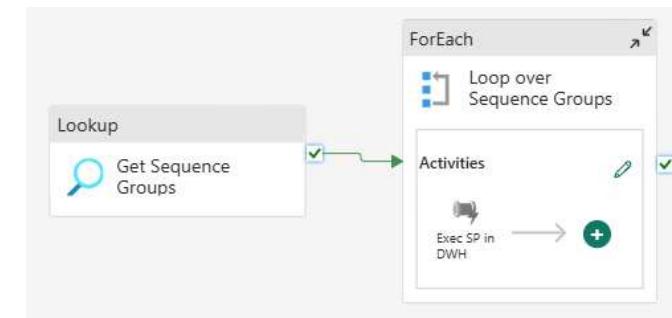
 Azure Machine Learning

Three types of activity

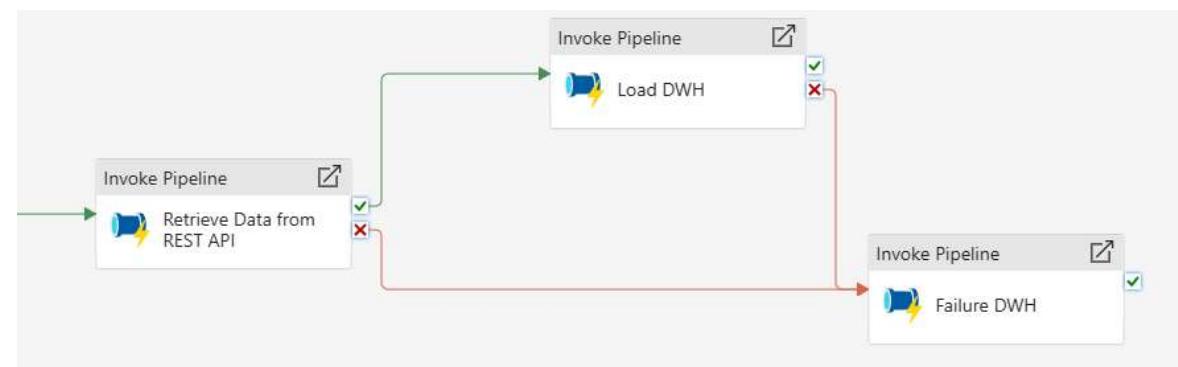
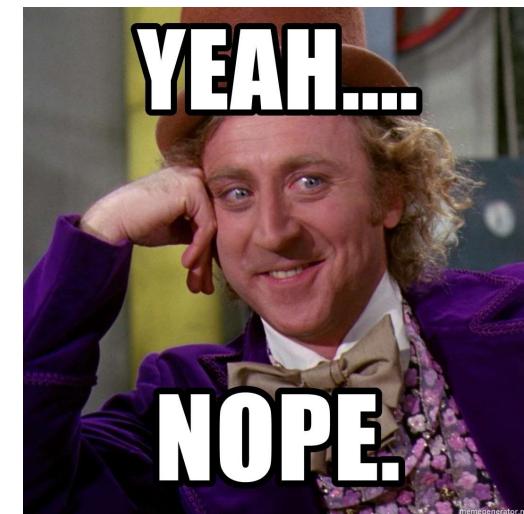
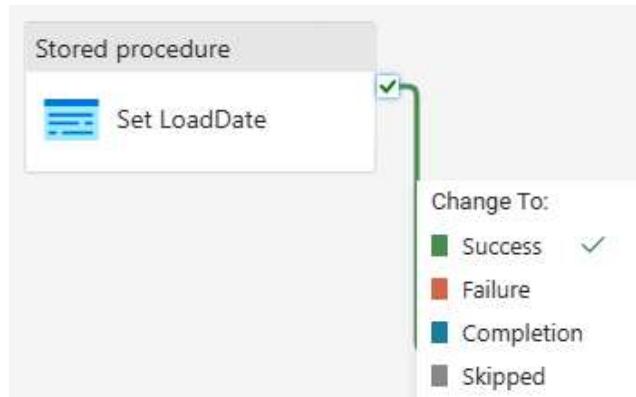
- **Data Movement** activities
 - Copy Data activity and alternatives
 - Data is copied using Fabric compute (CU)

- **External** (transformation) activities
 - Compute is not in the Pipeline
 - E.g. stored procedure, Notebook, Azure/Fabric Function ...

- **Pipeline** (control flow) activities
 - Lightweight “internal” activities
 - E.g. Lookup, Set Variable, For Each, If condition ...



Activity dependency



Copy data activity

The screenshot shows the 'Copy data' activity configuration screen. At the top, there's a toolbar with a 'Copy data' button, a 'Dummy Copy' name, and several icons for managing the activity. Below the toolbar, there are tabs: General, Source (which is selected), Destination, Mapping, and Settings.

Source Tab Configuration:

- Connection:** blobstorage koentest
- File path type:** Wildcard file path (selected)
- Container:** incrementaltest
- Wildcard paths:** Wildcard folder path / *
- Recursively:** checked
- File format:** DelimitedText

Destination Tab Configuration:

The destination tab is not visible in the screenshot.

Copy data activity

General Source **Destination** Mapping Settings

Connection *

Connection type

Database

Table option Use existing Auto create table ①

Table * .

General Source Destination **Mapping** Settings

<input type="checkbox"/> Source	Type	<input type="checkbox"/> Destination	Type
<input type="checkbox"/> <input type="button" value="Index"/>	abc String	<input type="checkbox"/> <input type="button" value="Index"/>	ANY nvarchar
<input type="checkbox"/> <input type="button" value="MovieTitle"/>	abc String	<input type="checkbox"/> <input type="button" value="MovieTitle"/>	ANY nvarchar

Dynamic Content

- Pipelines support **parameters & variables**
- On many fields you can define **dynamic content**
- You can use these to make your pipelines more flexible and dynamic

Invoked pipeline *

Wait on completion

Parameters

Name	Type	Value	Default value
loaddate	string	<code>@pipeline().parameters.loaddate</code>	20250603

Pipeline expression builder

Add dynamic content below using any combination of expressions, functions and system variables.

```
@concat('SELECT TableName, SQLStatement FROM etl.ProList WHERE  
ProcGroup = ''DWH'' AND ToLoad = 1 AND SequenceGroup = ',  
string(pipeline().parameters.SequenceGroup))
```

[Clear contents](#)

[Evaluate expression](#)

[Parameters](#) [System variables](#) [Trigger parameters](#) [Functions](#) [...](#)

Search

[Expand all](#)

[Collection Functions](#)

[Conversion Functions](#)

[Date Functions](#)

[Logical Functions](#)

[Math Functions](#)

[String Functions](#)

[concat](#)

Combines any number of strings together. For example, if parameter1 is foo, the following e...

[endswith](#)

Checks if the string ends with a value case insensitively. For example, the following expressio...

[guid](#)

Generates a globally unique string (aka. guid). For example, the following output could be g...

[indexof](#)

Find the index of a value within a string case insensitively. For example, the following express...

[lastindexof](#)

Find the last index of a value within a string case insensitively. For example, the following ex...

[replace](#)

[OK](#)

[Cancel](#)

Copy Job

- New item type
- Simplified data movement, like Copy Data activity
- Supports
 - Full copy / Incremental copy / CDC replication
 - Truncate destination
 - Append / overwrite / merge
- Copy Job activity in Fabric Pipeline for orchestration

Copy Job

	Pipeline copy activity	Copy job
Use case	Data lake and data warehouse migration, data ingestion,	Data Ingestion, Incremental copy, Application, Data Warehouse migration, Data transformation
Primary developer persona		Business Analyst, Integrator, Data Engineer
Primary developer skill set	ETL, SQL, JSON	No code, Low code
Code written		Very high
Data volume		
Development interface	canvas	Wizard, canvas
Sources	50+ connectors	50+ connectors

<https://learn.microsoft.com/en-us/fabric/fundamentals/decision-guide-pipeline-dataflow-spark>

demo

Fabric Pipelines vs ADF

ae

ADF vs Fabric Pipelines

- They are basically the same
- ... aside from some differences
- ADF and Synapse Pipelines are almost equal
 - Different Devops task for deploying
 - No global parameters in Synapse
 - UI is slightly different
- There are more discrepancies between Fabric and ADF/Synapse



Global parameters		
Global parameters are constants across a Data Factory that can be consumed by a pipeline.		
	New	Edit all
Name	Type	Value
environment	String	dev
subscription	String	a6af8505-790d-4
resourcegroup	String	dwh-dev
sqlservername	String	ae-dwh-dev

Linked service ≈ Fabric connection

- points to a certain data location
- contains info on how to authenticate
- can integrate with Azure Key Vault

Edit linked service

Azure Blob Storage [Learn more](#)

Name *

Description

Connect via integration runtime * [①](#)

Authentication method

Account selection method [①](#)
 From Azure subscription Enter manually

Azure subscription [①](#)

Storage account name *

Storage account kind

Managed identity name: **mssqltips-df**
Managed identity object ID:

Grant Data Factory service managed identity access to your Azure Blob Storage.
[Learn more](#)

Test connection [①](#)
 To linked service To file path

[Save](#) [Cancel](#) [Test connection](#)

Linked service ≈ Fabric connection

- Fabric connections have an owner 
- By default, Fabric adds your username 
- After creation, only authentication can be edited 

Manage Connections and Gateways

Connections On-premises data gateways Virtual network data gateways Azure Key Vault references

Cloud and data gateway connections for artifacts. [Learn more about supported connections.](#)

Name ↑	Connection type	Users
blobstorage koentest	Azure Blob Storage	Koen
CapacityMetricsCES	Capacity Metrics	Koen
eventhubs_fueltype koentest	EventHub	Koen
FabricSql koentest	ase	Koen
https://learn.microsoft.com	mycheapdwh-workspaceidentity	Koen
https://raw.githubusercontent.com	Connection ID	Koen
koensql.database.windows	97669a6c-d957-4c61-8310-204756755605	Koen
Lakehouse	Connection type	Koen
	SQL Server	
	Server	mssqltips.database.windows.net
	Database	mycheapdwh
Authentication		
ⓘ Workspace identity is currently only supported for Dataflows Gen2 with CI/CD, Data pipelines, OneLake shortcuts, Semantic models.		
Authentication method *		
Workspace identity		

Datasets

Don't exist in Fabric 😱



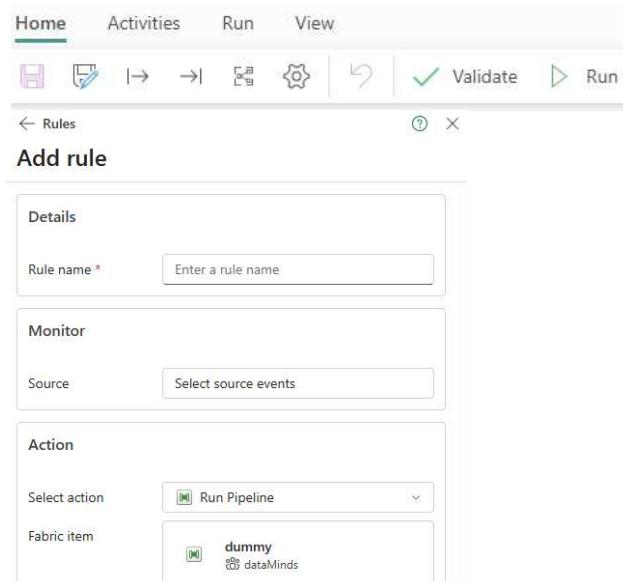
DelimitedText
CSV_Badges

Connection Schema Parameters

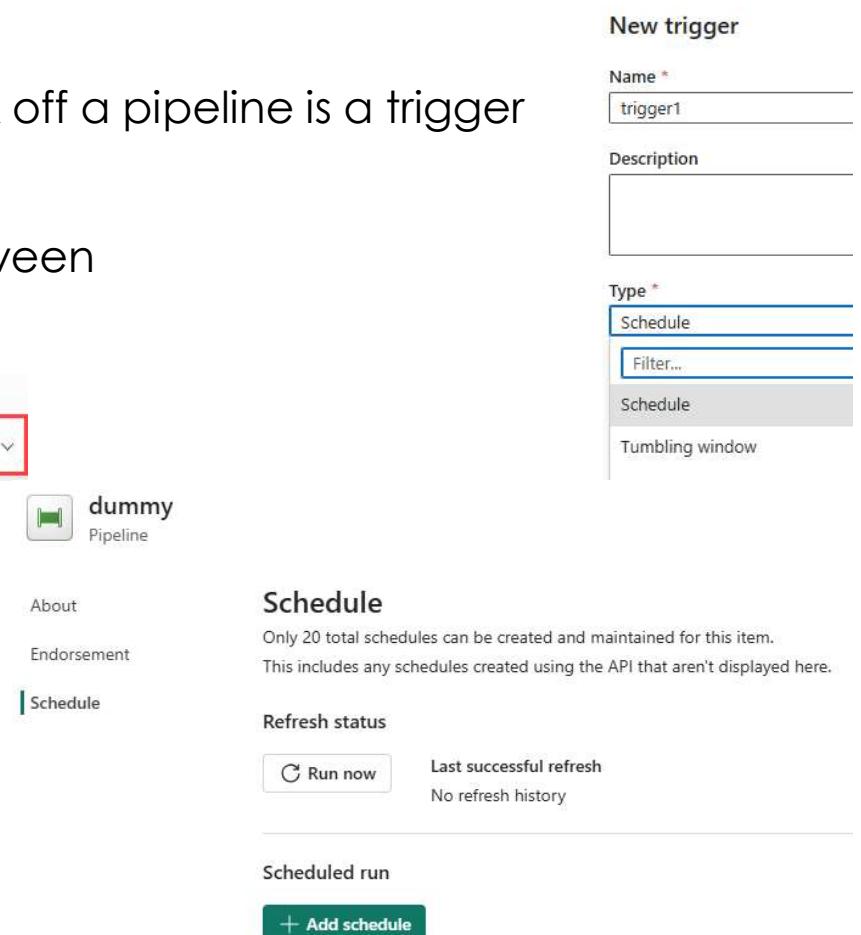
Linked service *	<input type="text" value="Blob_MSSQLTips"/>	Test connection	Edit	New	Learn more
File path *	<input type="text" value="mssqltips"/>	/	<input type="text" value="Directory"/>	/	<input type="text" value="badges.csv.gz"/>
Compression type	<input type="text" value="gzip (.gz)"/>				
Compression level	<input type="text" value="Fastest"/>				
Column delimiter ⓘ	<input type="text" value="Comma ()"/>				
	<input type="checkbox"/> Edit				
Row delimiter ⓘ	<input type="text"/>				
	<input checked="" type="checkbox"/> Edit				
Encoding ⓘ	<input type="text" value="Default(UTF-8)"/>				
Escape character ⓘ	<input)"="" type="text" value="Backslash (\"/>				
	<input type="checkbox"/> Edit				
Quote character ⓘ	<input)"="" type="text" value="Double quote ("/>				
	<input type="checkbox"/> Edit				

Triggers

- In ADF/Synapse, everything that can kick off a pipeline is a trigger
- In Fabric, there's a distinction made between schedules and triggers



The screenshot shows the 'Add rule' interface in Microsoft Fabric. At the top, there is a navigation bar with 'Home', 'Activities', 'Run', 'View', and other icons. Below the navigation bar, there are two tabs: 'Schedule' and 'Trigger'. The 'Trigger' tab is highlighted with a red box. On the left side, there is a sidebar with 'Add rule' and three sections: 'Details', 'Monitor', and 'Action'. The 'Details' section has a 'Rule name' field. The 'Monitor' section has a 'Source' field. The 'Action' section has a 'Select action' dropdown set to 'Run Pipeline' and a 'Fabric item' dropdown containing 'dummy' and 'dataMinds'.



The screenshot shows the 'New trigger' configuration page. At the top, it says 'New trigger'. Below that, there are fields for 'Name *' (containing 'trigger1') and 'Description'. A 'Type *' dropdown is open, showing 'Schedule' as the selected option. Other options include 'Filter...', 'Schedule', and 'Tumbling window'. On the right, there is a 'dummy Pipeline' card with sections for 'About', 'Endorsement', 'Schedule', 'Refresh status', and 'Scheduled run'. The 'Schedule' section contains a 'Run now' button and a note about refresh history. The 'Scheduled run' section has a '+ Add schedule' button.

Integration runtimes



Azure

Use this for running data flows, data movement, external and pipeline activities in a fully managed, serverless compute in Azure.

Name ↑↓	Type ↑↓	Sub-type ↑↓	Status ↑↓	Related ↑↓	Region ↑↓
AutoResolveIntegrationRuntime	Azure	Public	Running	0	Auto Resolve

On-premises data gateway



Self-Hosted

Use this for running activities in an on-premises / private network
[View more](#) ▾

Status
Service Settings
Diagnostics
Network
Connectors

Your gateway is all set up.

[Sign in](#) for more information on your gateway.

Gateway version number: 3000.176.108 (June 2023)

A new version is available.

[Download](#)

Help us improve the on-premises data gateway by sending usage information to Microsoft.

[Read the privacy statement online](#)



Azure-SSIS

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

[Sign in](#)

[Close](#)

Only in Fabric...

- Semantic model refresh
- Azure Databricks
- Functions**
- Azure HDInsight
- Azure Batch

Notifications

- Office 365 Email
- Microsoft Teams
- Office 365 Outlook (Legacy)
- Teams (Legacy)

- Transform**
- Spark Job Definition
 - Notebook
 - Script
 - Stored procedure
 - KQL**

Get data

Ingest batch and real-time data into a single location within your Fabric workspace.

Copy job

Makes it easy to copy data in Fabric. Includes full copy, incremental copy, and event-based copy modes.



Invoke Pipeline

Runny McRunFace

General **Settings** 2

Type Fabric Azure Data Factory Azure Synapse Analytics

Connection * ?

Workspace *

Pipeline *

Wait on completion

Home

Prediction model variables

Variables

Name *	Note	Value type	Default value set *	...
StudiedHoursForPredictedScore	<input type="button" value=""/>	Integer	6	
Model training dataset	<input type="button" value=""/>	String	('Hours': [2, 3, 4, 5, 6, 7, 8, 9, 10])	
NYC pipeline source	<input type="button" value=""/>	String	S:\\Data\\Clean\\CountryScoresDat...	
Refresh data	<input type="button" value=""/>	Boolean	True	



<https://youtube.com/watch?v=Fct1dcZMyWs>

Costs in ADF

ae



Lessons Learned:

**Understanding Pipeline Pricing
in Azure Data Factory and Azure Synapse Analytics**

Cathrine Wilhelmsen

DataMinutes #2 • January 21st, 2022



© 2022 Cathrine Wilhelmsen (hi@cathrinew.net)



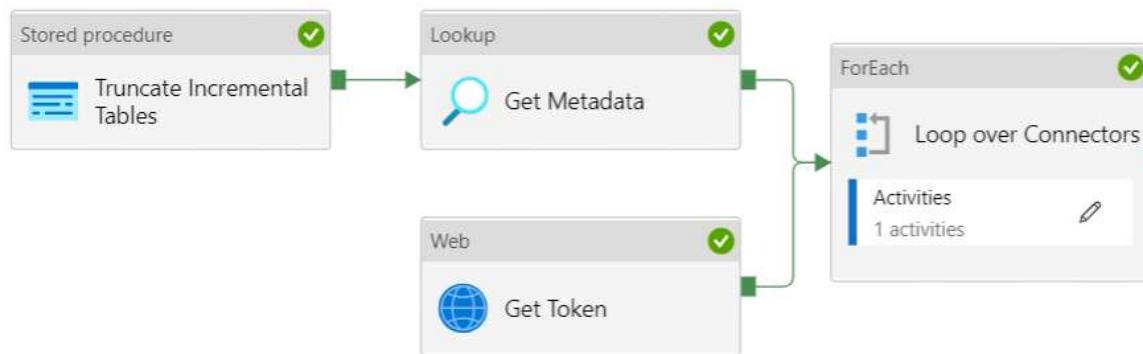
Important to remember

ae

**ALL ACTIVITIES ARE
PRO-RATED BY THE
MINUTE
AND ROUNDED UP**

Example

ae



2 minutes and 23 seconds runtime

Parameters Variables Settings Output

Showing 101 - 104 of 104 items

Name	Type	Run start	Duration	Status	Integr
Loop over Connectors	ForEach	2022-03-27T09:08:02.639Z	00:02:14	✓ Succeeded	
Get Metadata	Lookup	2022-03-27T09:07:58.710Z	00:00:03	✓ Succeeded	AutoR
Truncate Incremental Tables	Stored proc...	2022-03-27T09:07:55.258Z	00:00:03	✓ Succeeded	AutoR
Get Token	Web	2022-03-27T09:07:55.258Z	00:00:03	✓ Succeeded	AutoR

Example

Debug run consumption

	Quantity	Unit
Pipeline orchestration		
Activity runs	104	Activity runs
Pipeline execution		
Azure integration runtime		
Data movement activities	1.6667	DIU-hour
Pipeline activities	0.0167	Execution hours
External activities	0.0333	Execution hours

1 minute
2 minutes → 1.6667
100 minutes!

Orchestration (Azure-IR): \$0.001/run

Pipeline (Azure-IR): \$0.005/hour
→ \$0,0000835/run

External (Azure-IR): \$0.00025/hour
→ \$0,000008325/run



Data Movement (Azure-IR): \$0,25/hour
→ $2 * 1.6667 * \$0.25 = \$0,83335/\text{run}$

\$0,834441825 / run
once every day: \$25.03
once every hour: \$600.80
once every 5 min: \$7209.58

50% cost saving with one single config

General Source Sink Mapping Settings User properties

i You will be charged # of used DIUs * copy duration * \$0.25/DIU-hour. Local currency and separate discounting may apply per subscription type. [Learn more](#)

Maximum data integration unit ⓘ

Auto

Filter...

Degree of copy parallelism ⓘ

Auto

2

Data consistency verification ⓘ

4

Fault tolerance ⓘ

8

Enable logging ⓘ

16

Enable staging ⓘ

32

The screenshot displays a configuration interface for a data integration task. The 'Settings' tab is active. A dropdown for 'Maximum data integration unit' is set to 'Auto'. A second dropdown for 'Degree of copy parallelism' is set to 'Auto' and has '2' selected, which is highlighted with a red box. Other options in this dropdown are 4, 8, 16, and 32. The background shows other configuration items like 'Data consistency verification', 'Fault tolerance', 'Enable logging', and 'Enable staging'.

Fabric uses CU consumption

- Only two types of activities:
 - Data movement – **1,5** CU hours
 - Data orchestration – 0,0056 CU hours per activity run
 - In [docs](#), 1 CU hour = **\$0,18**
- Copying 1TB of data to a warehouse
 - Takes about 11 minutes (662 seconds)
 - “intelligent throughput optimization” was 4
 - Utilized CU hours = $4 * 1,5 * (11/60) = 1,1$ CU hours (or 3960 CU seconds)
 - $1,1 \text{ CU hours} * (\$0,18 / \text{CU hours}) = \$0,20$
- Check out the [pricing examples](#) (they are misleading though as they deal with only 1 copy activity)

ADF vs Fabric

- In ADF, duration matters for all activities
- In Fabric, duration only matters for copy data activity
- Cost of the same pipeline
 - ADF = **\$0,83335 per run** (with default settings: **\$1,66 per run**)
 - Fabric = 36000 CU = 10 CU hour

Operation name	CU (s)	Duration (s)	Users	Billing type
DataMovement	36,000.0000	1,981.9620	1	Billable
ActivityRun	80.6400	178.2830	1	Billable
Total	36,080.6400	2,160.2450	1	Billable

=> 10 CU hour / (4 * 1.5) = 1,666667 hours (same as in ADF, thus 100 minutes)

=> 10 CU hour * (\$0,18/CU hour) = **\$1,8 per run**

- Fabric is about 8% more expensive (keep in mind you pay for the capacity running)

Tips & tricks

de

Ownership!!!

Manage Connections and Gateways

Connections On-premises data gateways Virtual network data gateways Azure Key Vault references

Cloud and data gateway connections for artifacts. [Learn more about supported connections.](#)

Name ↑	Connection type	Users
blobstorage koentest	Azure Blob Storage	Koen
CapacityMetricsCES	Capacity Metrics	Koen
eventhubs_fueltype koentest	Eventhub	Koen
FabricSql koentest	Fabric SQL database	Koen
https://learn.microsoft.com/en-us/fabric/release-plan/	Web	Koen
https://raw.githubusercontent.com/MicrosoftLearning/dp-data/main/orders.csv	Web	Koen
koensql.database.windows.net:mycheapdwh	SQL Server	Koen
Lakehouse	Lakehouse	Koen

 dataMinds 
Demo workspace for Fabric Data Pipelines

+ New item New folder → Import Migrate

	Name	Status	Type	Task	Owner
	copyjob1		Copy job	—	Koen Test
	dummy		Pipeline	—	Koen Test

Capacity Consumption

- In general, the less code, the more CU usage ;)
- Follow the same guidelines as for ADF cost
 - Minimize (Copy Data) activities
 - Use Fabric Data Pipelines as an ELT orchestrator
 - Use shortcuts to minimize copying
- When you schedule notebooks, allow the session to be reused

Spark settings

Configure and manage settings for Spark workloads and the default environment for the workspace.

Pool Environment Jobs **High concurrency** Automatic log

For notebooks

When high concurrency for notebooks is on, multiple notebooks can use the same Spark application to reduce the start time for each session. [Learn more about running notebooks in high concurrency mode](#)

On

For pipeline running multiple notebooks

When high concurrency for pipelines is on, multiple notebooks can use the same Spark application to reduce the start time for each session. [Learn more about running pipelines in high concurrency mode](#)

On

Library Variables

ae

WS Variables

Variables

Name *	Note	Type ⓘ
Source_LH		String
Source_WSID		String
Destination_LH		String
Destination_WSID		String
SourceTable_Name		String
DestinationTable_Name		String

Default value set * Active

- 1f61c499-89cd-4df5-92c7-d110b...
- 71bc08cb-a3dd-4d72-b101-1b3af...
- 4fe228d3-a363-4b7f-a5d4-fae9d2...
- dfdf8621-3a7f-44ed-a44d-64ae48...
- Processed
- DevCopiedData

Alternative value sets

Test VS *
4fe228d3-a363-4b7f-a5d4-fae9d2...
dfdf8621-3a7f-44ed-a44d-64ae48...
c0f13027-9bf4-4e8c-8f57-ec5c18...
dfdf8621-3a7f-44ed-a44d-64ae48...
DevCopiedData
TestCopiedData

Prod VS *
c0f13027-9bf4-4e8c-8f57-ec5c18...
dfdf8621-3a7f-44ed-a44d-64ae48...
084e2ac5-386d-4c13-8f9a-d32fc5...
dfdf8621-3a7f-44ed-a44d-64ae48...
TestCopiedData
ProdCopiedData

Search Filter Focus on Active Validations

Metadata FTW

- Parameterize everything
- Build metadata-driven pipelines
 - Create a pipeline once, add metadata later

General Settings

Connection * ⓘ @pipeline().libraryVariables.warehouse_id

Connection type * Warehouse

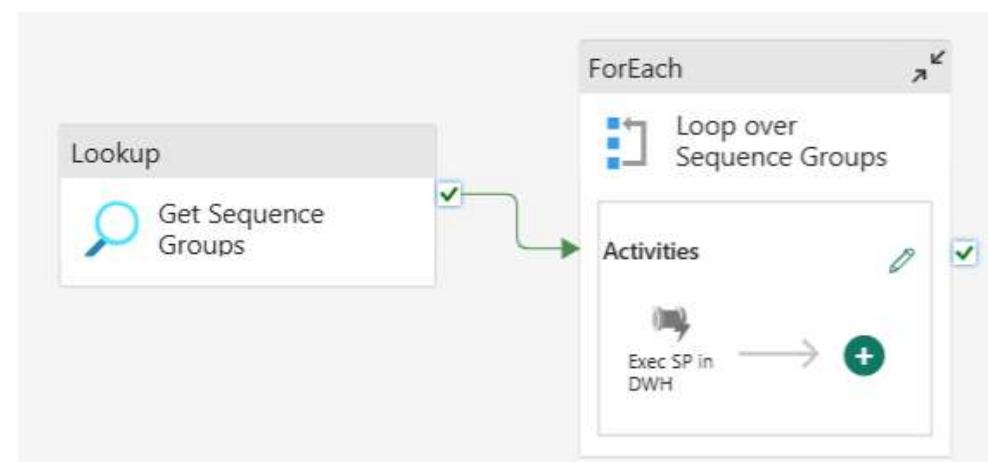
Workspace ID @pipeline().DataFactory

SQL connection string @pipeline().libraryVariables.warehouse_id

Script * Query ⓘ NonQuery ⓘ

+

@item().SQLStatement



Resiliency

- Create resilient pipelines by implementing restartability & retries
 - Change the default time-out (no task should take 12 hours)
 - Prevent errors e.g. check for existence

The screenshot shows a software interface for configuring a pipeline. It has two tabs at the top: 'General' (underlined in green) and 'Settings'. The 'General' tab displays the following fields:

- Name *: my lookup
- Description: (empty text area)
- Activity state ①: Activated (radio button selected)
- Timeout ①: 0.12:00:00
- Retry ①: 0
- > Advanced: (link)

The 'Settings' tab displays the following fields:

- Timeout ①: 0.00:10:00
- Retry ①: 2
- Advanced: (link)
- Retry interval (sec) ①: 30

A large orange arrow points from the 'General' tab towards the 'Settings' tab, indicating the progression from basic pipeline setup to advanced resiliency configurations.

- Build modular and **idempotent** pipelines
 - Don't build one monolithic pipeline that does it all
- Don't forget about monitoring and alerting

Various

- Use git and CI/CD, obviously (and multiple environments!)
- Use clear naming conventions and use folders to organize
- Use Key Vault if possible
- Use managed identities / workspace identity where possible
- Typically I don't parse JSON in a Copy Data activity

conclusion

Conclusion

- There's a big overlap between ADF and Fabric Pipelines
- Biggest differences:
 - how cost is calculated due to Fabric capacity
 - invoking child pipelines
 - CI/CD and parameterization
 - connection configuration
- Most ADF best practices & design patterns apply to Fabric Pipelines
- Use Fabric Pipelines as an ELT orchestrator
 - definitively not for streaming data

thank
you