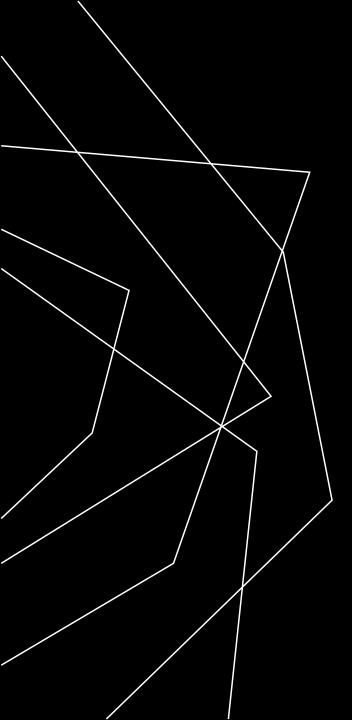# SLIDES CAN BE FOUND AT :

https://github.com/BenniDeJagere/Presentations

/{Year}/{Date}_{Event}

DATAFLOWS?

(Google) Cloud Dataflow

TensorFlow Dataflow

SSIS Data Flows

Microsoft Flow

ADFv2

Mapping Data Flows
Wrangling Data Flows

PowerApps dataflows

Power Platform dataflows

DATAFLOWS?

~~(Google) Cloud Dataflow~~

~~TensorFlow Dataflow~~

~~SSIS Data Flows~~

~~Microsoft Flow~~

~~ADFv2~~

~~Mapping Data Flows~~
~~Wrangling Data Flows~~

~~PowerApps dataflows~~

~~Power Platform dataflows~~
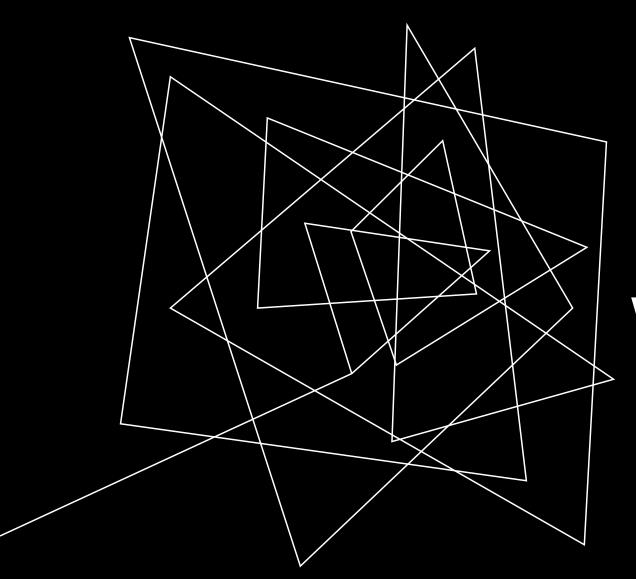
Power BI dataflows!

WHO?

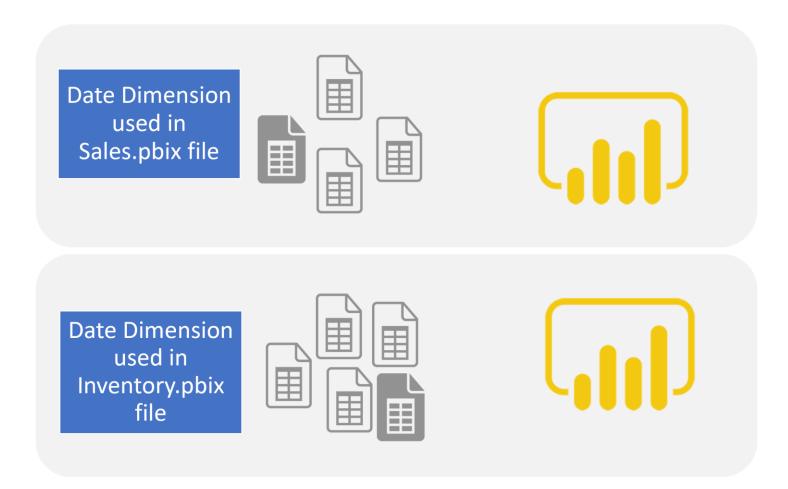# WHO SHOULD USE IT?

POWER USERS

BUSINESS ANALYSTS

TECH SAVVY USERS

EVERYONE ELSE ☺

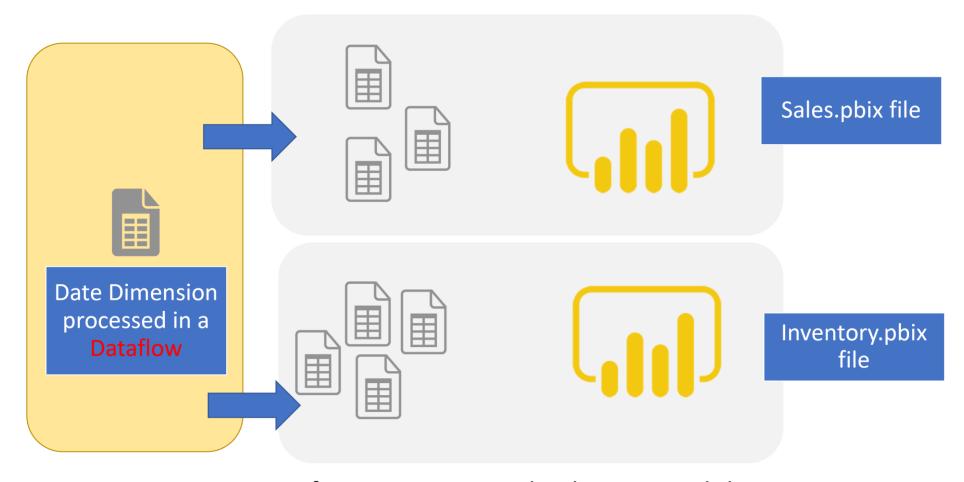WHAT?

Date Dimension transformation executed multiple times, when only once is needed

ONE ENTITY, MANY DATASETS

Date Dimension processed in a **Dataflow**

Sales.pbix file

Inventory.pbix file

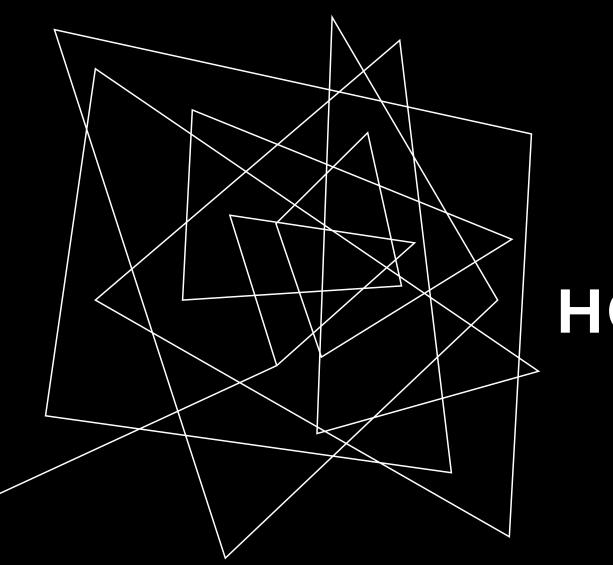Date Dimension transformation executed only once, and then used multiple times

ONE ENTITY, MANY DATASETS

ONE ENTITY, MANY DATASETS

Lack of semantic model will create many versions

Re-use of queries across datasets is limited

Shared Datasets is an all or nothing approach

Does ~~"Direct Query over Power BI datasets and Azure Analysis Services"~~ composite models change the way we work?

HOW?

ENTER, POWER BI DATAFLOWS!

Preview Late 2018, GA April 2019

Cloud based PowerQuery experience

Self-Service Data Preparation

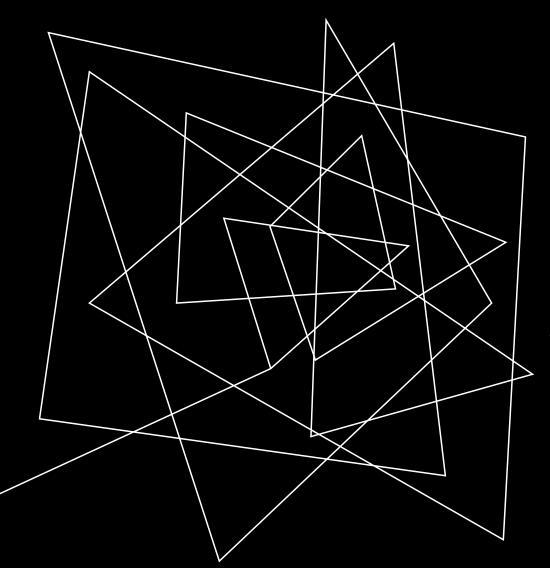Non-technical user personas

ENTER, POWER BI DATAFLOWS!

Build reusable Data Entities

Not a replacement for a DWH Solution (my 2 cents)

Manageable and securable like most Power BI artefacts

Managed CDM Folders

LET'S TAKE IT FOR A SPIN! DEMO
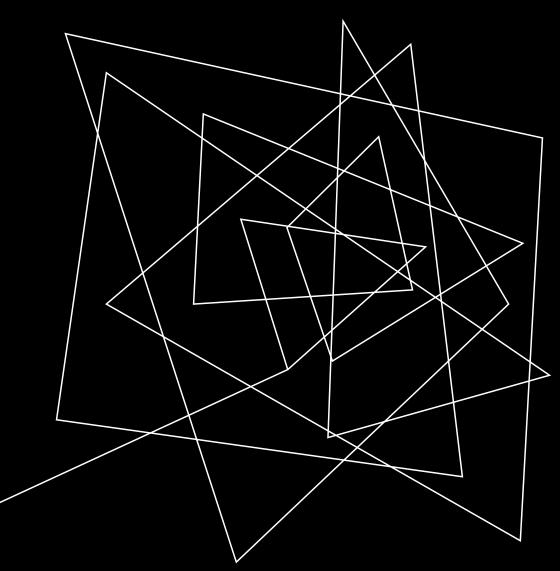
DEMO: USE CASE

www.citibikenyc.com/system-data

Public Open Data

Starts June 2013

Information about every trip

    Longer than 60 seconds

     Starts at public station

Masterdata

https://i0.wp.com/thenypost.files.wordpress.com/2013/12/citibike1.jpg

HOW:
COMMON DATA MODEL

ENTER, POWER BI DATAFLOWS!

Unified metadata layer for Cloud Apps and Services

Shared

    Data Language
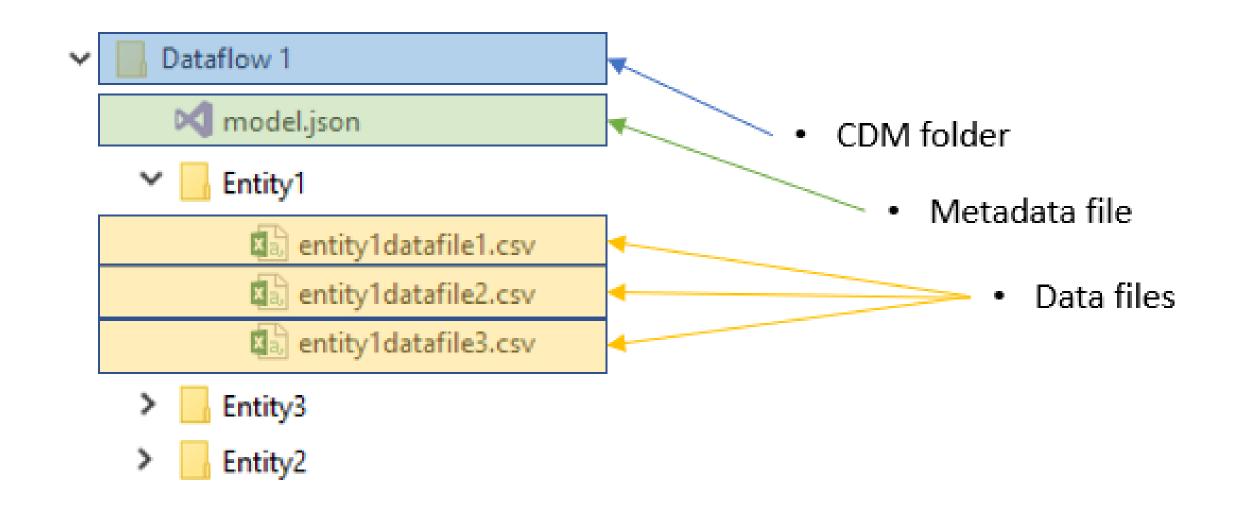
    Storage Format and Location

    Set of Default Entities
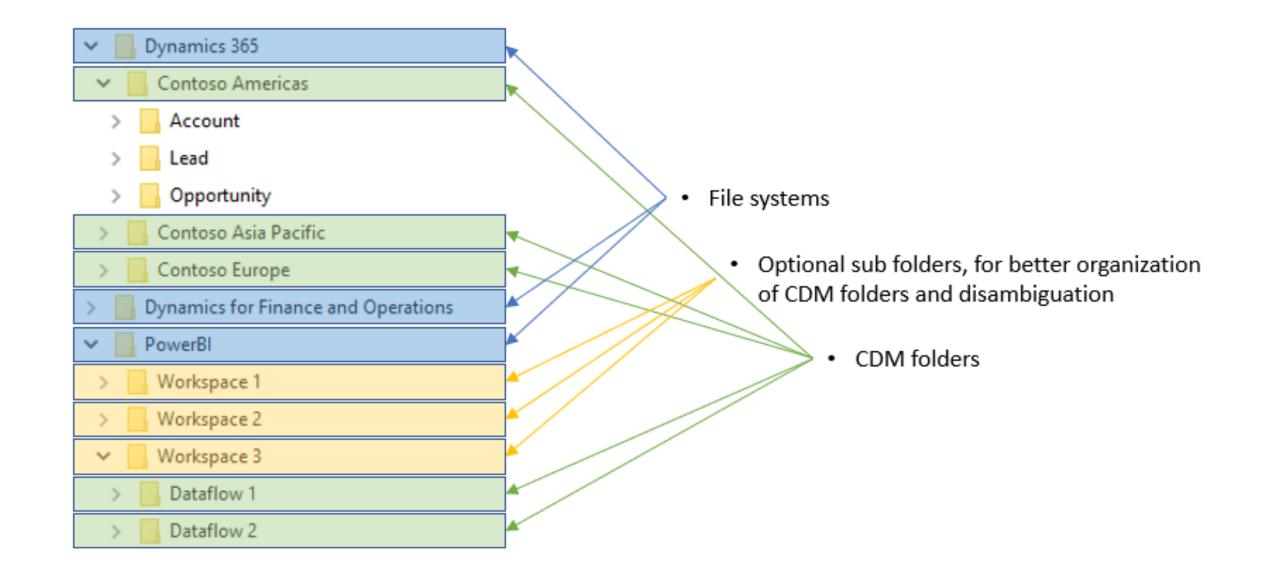
Extend Entities for all consumers
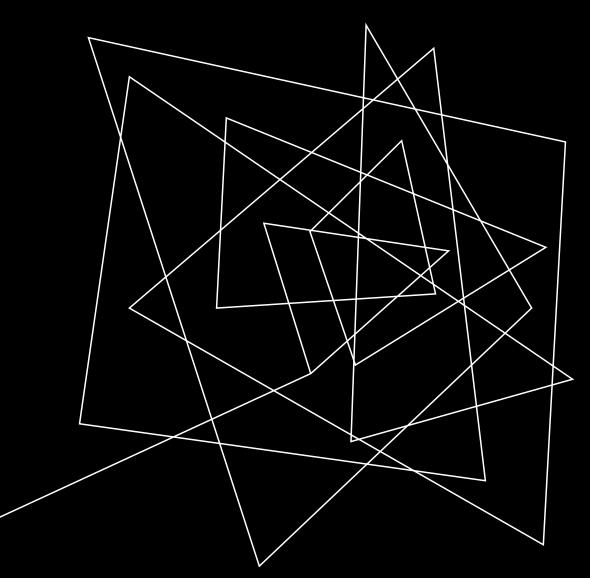
Simplified integration and disambiguation of data

COMMON DATA MODEL
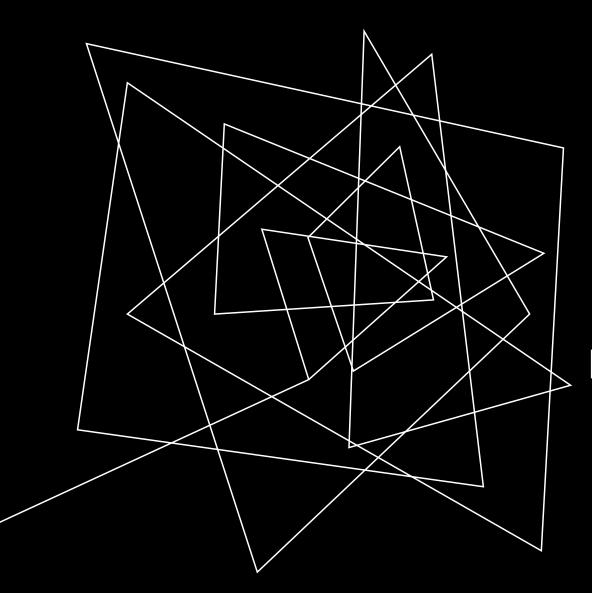
COMMON DATA MODEL
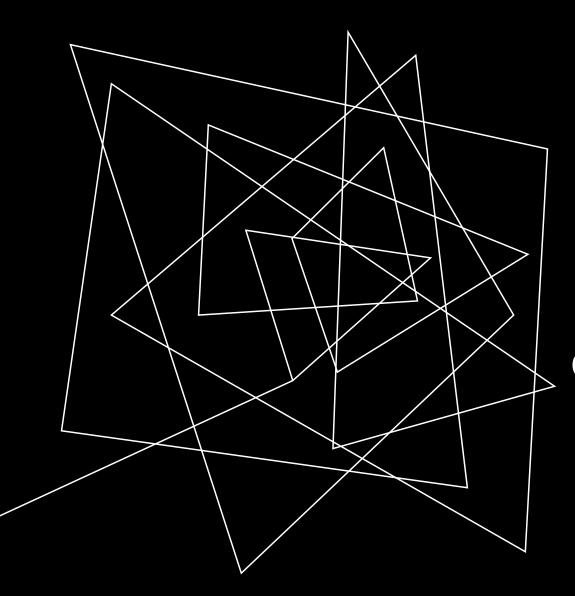
HOW:
COMMON DATA MODEL
DEMO

HOW:
DATA REFRESH

DATA REFRESH

Trigger Manual refresh in the UI

Scheduled refresh comparable to Power BI Datasets

Refresh History offers an interesting view

Automatically refreshes any related dataflows

Incremental Refresh is Premium Only

HOW:
QUERY REUSAGE

QUERY REUSABILITY (PRO)

No developer experience in Power BI Desktop

Copy/Paste queries between environments

Reference Query, Append or Merge

Create a query, and mark 'Disable for load'

Reference this query, in the same dataflow

Logic is reused, and data is duplicated every time the query is referenced
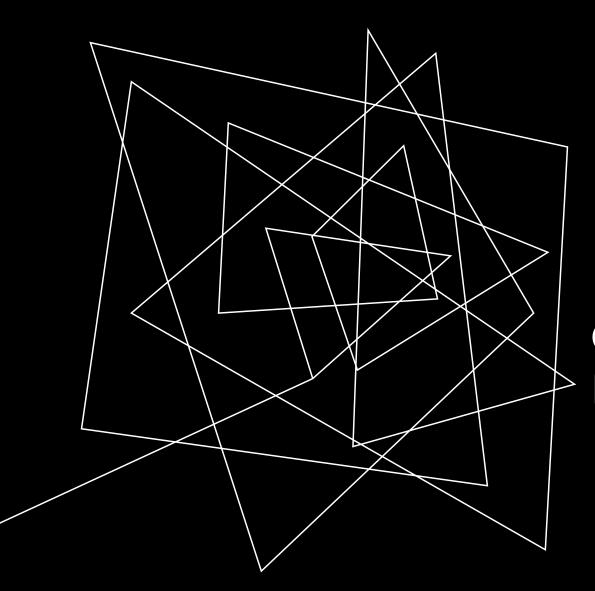
QUERY REUSABILITY (PREMIUM / PPU)

Linked/Computed Entities (Premium/PPU Only)
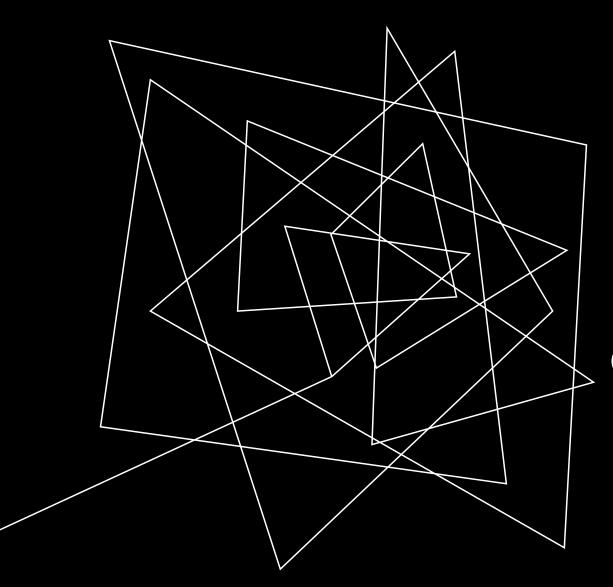
    In-lake Compute, which requires dedicated capacity

    Can Reference any Entity in any Workspace

    Data is only stored in the original Entity
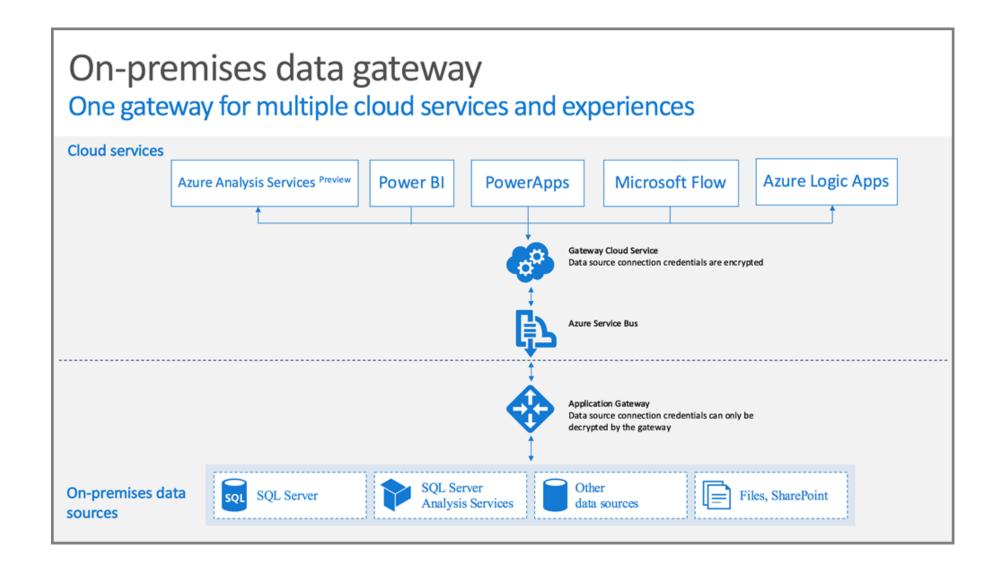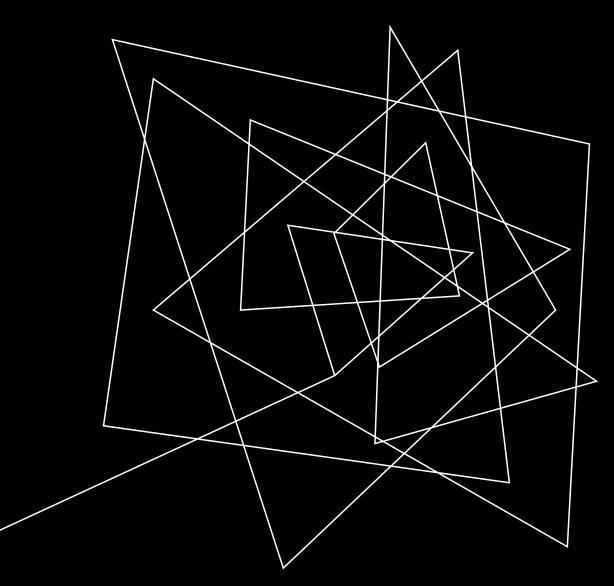
HOW:
QUERY REUSAGE DEMO

HOW:
GATEWAYS

GATEWAYS

On-premises data gateway
One gateway for multiple cloud services and experiences

Cloud services

| Azure Analysis Services Preview | Power BI | PowerApps | Microsoft Flow | Azure Logic Apps |

Gateway Cloud Service
Data source connection credentials are encrypted

Azure Service Bus

Application Gateway
Data source connection credentials can only be decrypted by the gateway

On-premises data sources

SQL Server | SQL Server Analysis Services | Other data sources | Files, SharePoint

HOW:
SECURITY

SECURING DATAFLOWS

Dataflows reside in Workspaces

"Secure"/Hide Linked Entities by disabling load

Datasets using Dataflows will likely be Import Mode

Security is the responsibility of the dataset designer!

WHY?

WHERE DO DATAFLOWS FIT IN?

Without dataflows, users might
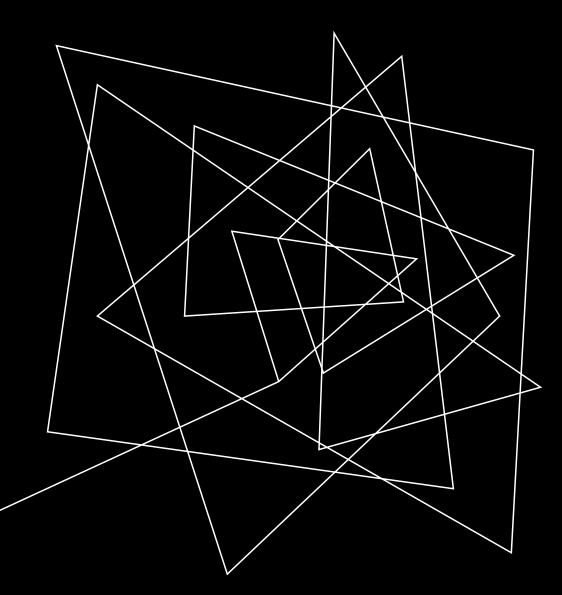
Be blocked on IT involvement

Use Excel and manual processes

Require 3rd party data preparation tools
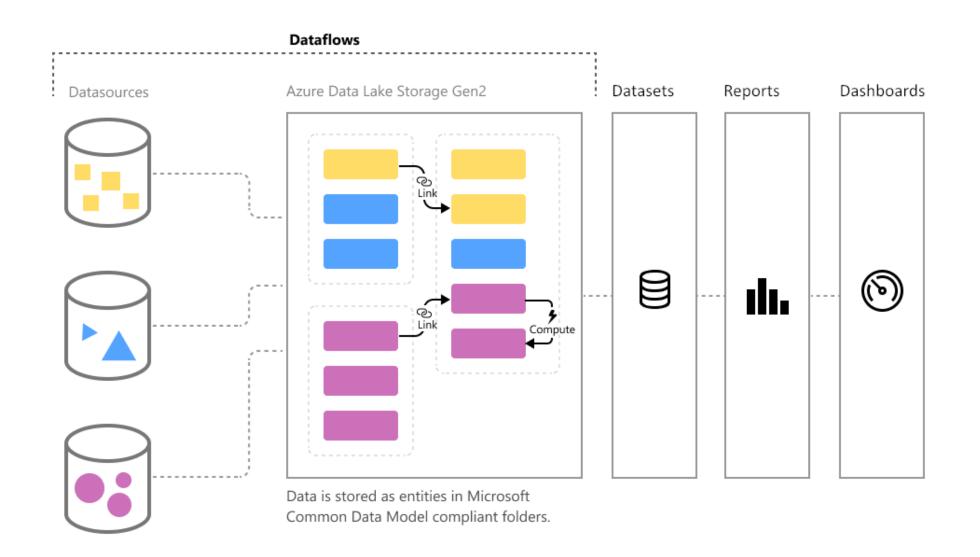
WHERE DO DATAFLOWS FIT IN?

Dataflows are for Power Query users

- Easy to build reusable data entities
- Easier to compose and orchestrate
- Multi-stage, multi-user data prep workflows
- Dataflows fill a self-service gap in the end-to-end story
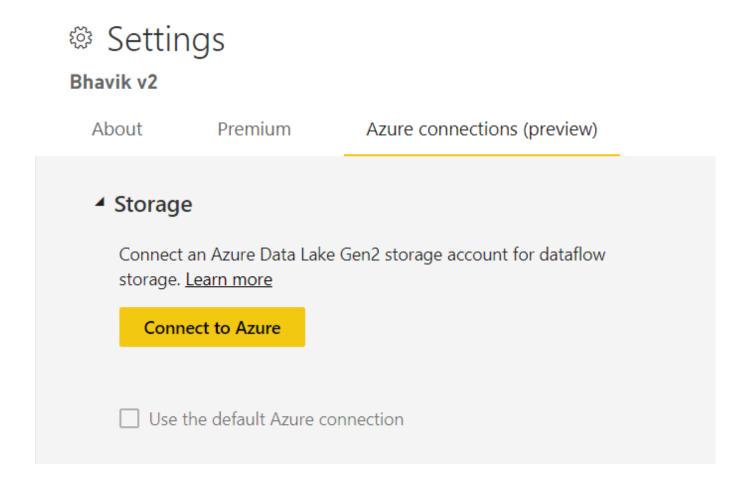
EXTENDING DATAFLOWS

EXTENDING DATAFLOWS
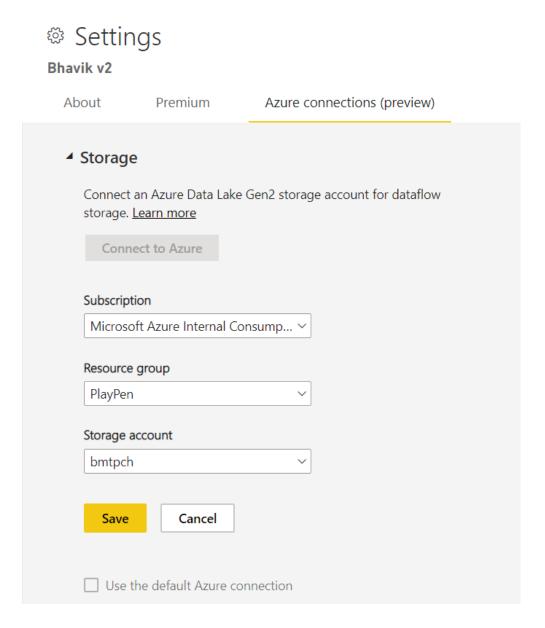
BRING YOUR OWN STORAGE ACCOUNT

https://powerbi.microsoft.com/en-us/blog/power-bi-dataflows-multiple-data-lakes-support/

BRING YOUR OWN STORAGE ACCOUNT

⚙ Settings

Bhavik v2

About   Premium   Azure connections (preview)

◢ **Storage**

Connect an Azure Data Lake Gen2 storage account for dataflow
storage. Learn more

Connect to Azure

Subscription

Microsoft Azure Internal Consump... ⌄

Resource group

PlayPen ⌄

Storage account

bmtpch ⌄

Save   Cancel

☐ Use the default Azure connection

EXTERNAL CDM FOLDERS

CDM Folders is an open standard

Any tool that can:

Access ADLSg2

Output JSON

Create the correct manifest format

Link External CDM Folder in Power BI

Done!

EXTERNAL CDM FOLDERS

Start creating your dataflow

Define new entities

Choose a data source to define the entities for your dataflow. You can map your data to standard Common Data Model entities, or define custom entities instead.
Learn more

Add new entities

Link entities from other dataflows

Linking to entities from other dataflows reduces duplication and helps maintain consistency across your organization.
Learn more

Add linked entities

Import Model

Choose a dataflow model to import into your workspace.
Learn more

Import model
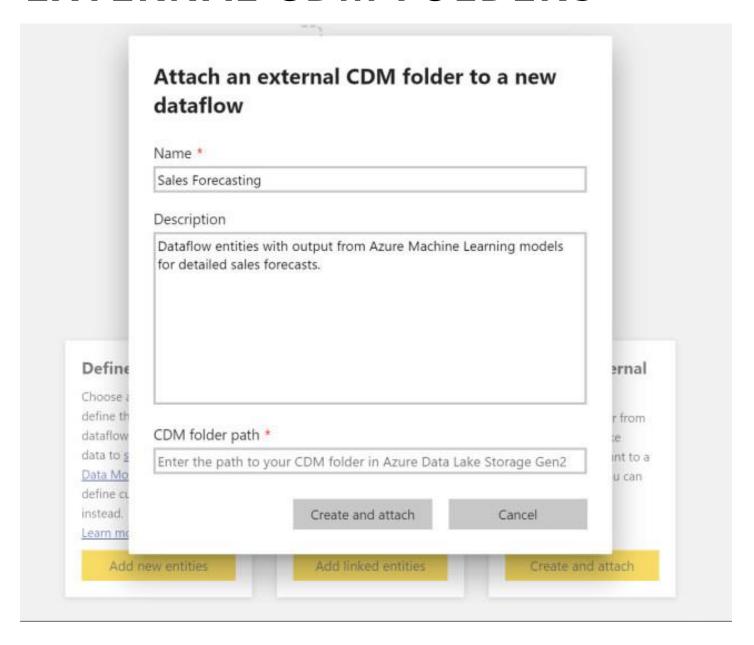
Attach a Common Data Model folder (preview)
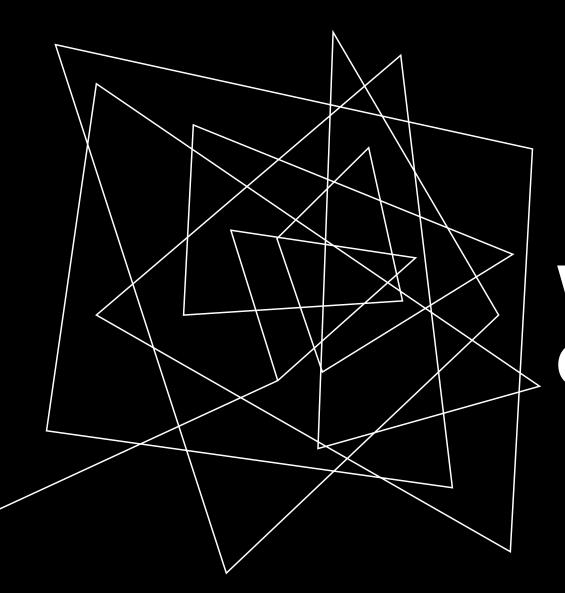
Attach a Common Data Model folder from your Azure Data Lake Storage Gen2 account to a new dataflow, so you can use it in Power BI.
Learn more

Create and attach

EXTERNAL CDM FOLDERS
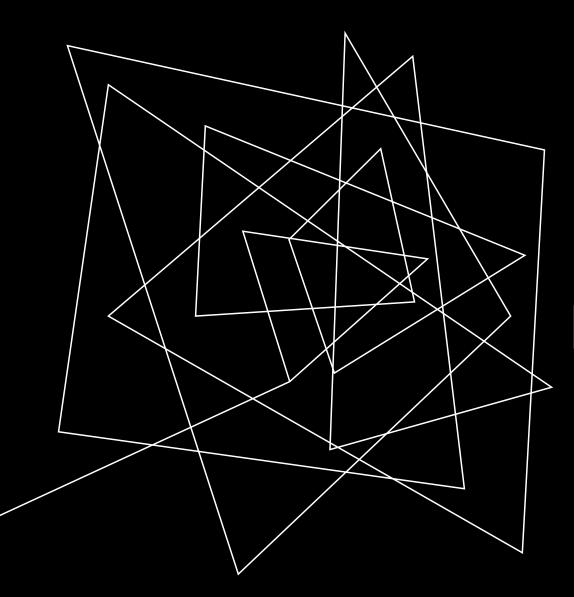
WHAT'S THE CATCH?

WHAT'S THE CATCH?

The cool stuff is Premium / PPU only

Validation of queries can take a loooooonng time

Modifying a schema will drop all data

Dataflows can only be modified by their owners

Dataflows using gateway data sources do not support multiple credentials

When working with APIs (for CI/CD) the story isn't smooth yet

Exporting/Importing creates a new ID

Importing will not fix existing references to linked entities

BONUS!

POWER BI DATAFLOWS SORT-OF-FAQ

- A capability for self-service data preparation in Power BI
- Delivered in a familiar Power Query experience
- Built on the foundation of Azure Data Lake Storage gen2
- Utilize the CDM folder format for data storage
- A tool for business users to drive data reuse without requiring IT involvement
- Enable Excel-like data lineage and orchestration

- NOT a replacement for datasets
- NOT a replacement for a data warehouse
- NOT a replacement for Azure Data Factory or SSIS
- NOT a Premium-only feature
- NOT an additional cost or fee
- NOT spelled with a space or any capital letters
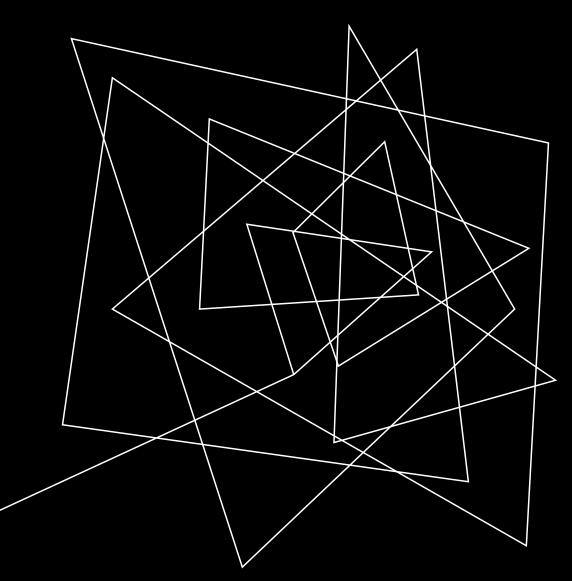
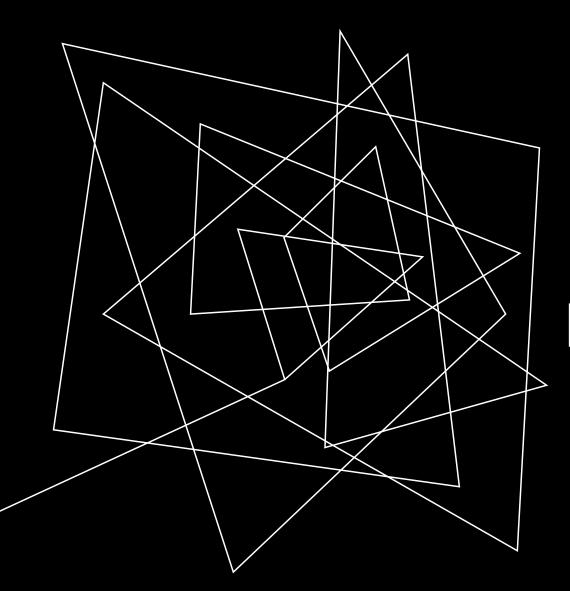@SQLAllfather – Matthew Roche

TAKEAWAYS

TAKEWAYS

Business Driven Initiatives

Fill existing gaps

Quick prototyping

Strive for IT handover to operationalize later on

Configure your own ADLSg2 Storage

BUT WAIT!

Support custom connectors in Power BI dataflows

Article • 04/19/2023 • 2 contributors

> ⓘ **Important**
>
> **Some of the functionality described in this release plan has not been released.** Delivery timelines may change and projected functionality may not be released (see Microsoft policy ⧉). Learn more: What's new and planned

Enabled for	Public preview	General availability
Admins, makers, marketers, or analysts, automatically	-	Sep 2023

Business value

You'll be able to use your custom connectors in Power BI dataflows experiences through the use of an on-premises data gateway.

Feature details

We're adding support for custom connectors in Power BI dataflows through the use of an on-premises data gateway. This support lets you validate and use your custom Power Query connector in Power Query Online and dataflows before certification.

https://learn.microsoft.com/en-us/power-platform/release-plan/2023wave1/data-integration/support-custom-connectors-power-bi-dataflows

Join us in Seattle for Microsoft Build!

President & Co-Founder Greg Brockman from OpenAI will be LIVE at Microsoft Build! Be there to explore how Microsoft and OpenAI's full-stack AI platform empowers developers and get ready to make AI work for you.

Register now

Showing 18 speakers

Satya Nadella

Chairman and CEO

Microsoft

in 🐦

Greg Brockman

President & Co-Founder

OpenAI

in 🐦

Amanda Silver

Corporate Vice President, Head of Product, Developer Division

Microsoft

in 🐦

Scott Guthrie

Executive Vice President, Cloud + AI

Microsoft

in 🐦

Kevin Scott

Chief Technology Officer and Executive Vice President, Technology & Research

Microsoft

in 🐦

Employ features from the next generation of Power BI dataflows

Article • 02/28/2023 • 2 contributors

> ⓘ **Important**
>
> **Some of the functionality described in this release plan has not been released.** Delivery timelines may change and projected functionality may not be released (see Microsoft policy ⬀). Learn more: What's new and planned

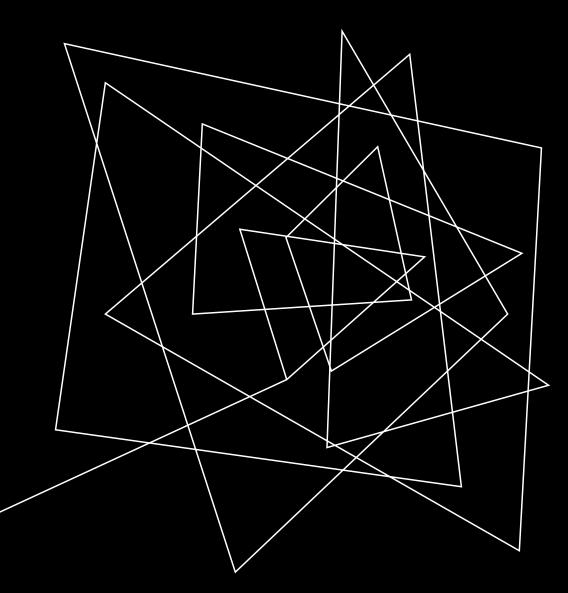| Enabled for	Public preview	General availability
Admins, makers, marketers, or analysts, automatically	Sep 2023	-

Business value

Dataflows are a self-service, cloud-based data preparation technology that helps customers bring the world's data into Power BI. Power BI creators of dataflows, for example, can promote reusability of data transformation and reduce load on data sources by centralizing data prep tasks into dataflows that other Power BI creators can take advantage of when building datasets in Power BI, Excel workbooks, and more.

Feature details

A new generation of Power BI dataflows improve on existing experiences and features in the following ways:

- The AutoSave dataflow feature automatically saves any edits to dataflows to the cloud, into a draft dataflow, so you can continue editing at a later time.
- The background publishing feature defers query validation to the background, allowing you to publish the dataflow without waiting for the potentially lengthy validation process to complete. If any validation errors are encountered, they're reported through the dataflow experience.
- An improved dataflow debugging and monitoring experience reduces the time it takes to investigate and fix refresh errors.
- We've added support for loading dataflow output to new destinations and support for loading data from data sources in VNets.

https://learn.microsoft.com/en-us/power-platform/release-plan/2023wave1/data-integration/next-generation-power-bi-dataflows

RESOURCES

RESOURCES

https://docs.microsoft.com/en-us/power-bi/service-dataflows-best-practices

https://docs.microsoft.com/en-us/power-bi/service-dataflows-overview

https://docs.microsoft.com/en-us/azure/architecture/reference-architectures/data/enterprise-bi-adf

https://docs.microsoft.com/en-us/common-data-model/

RESOURCES

https://docs.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflows-azure-data-lake-storage-integration

https://docs.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflows-premium-features

https://docs.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflows-machine-learning-integration

https://docs.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflows-features-limitations

RESOURCES

https://docs.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflows-directquery

https://docs.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflows-understand-optimize-refresh

https://docs.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflows-streaming

https://docs.microsoft.com/en-us/common-data-model/

RESOURCES

https://ssbipolar.com/2018/10/23/dataflows-in-power-bi/

https://www.sqlchick.com/entries/2019/3/28/three-ways-to-use-power-bi-dataflows

https://prologika.com/dataflows-the-good-the-bad-the-ugly/

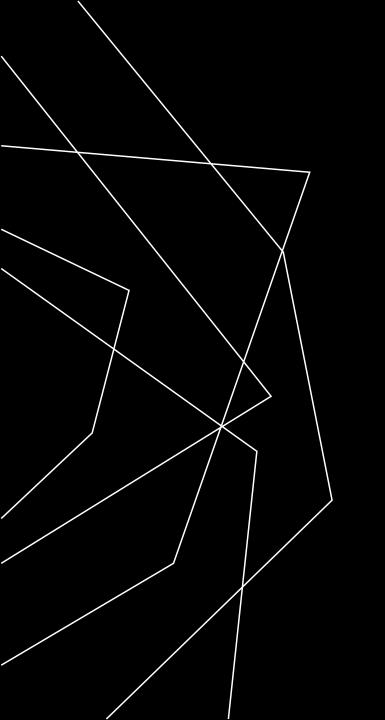https://exceleratorbi.com.au/power-bi-dataflows-and-why-you-should-care/

https://www.jamesserra.com/archive/2018/08/power-bi-dataflows/
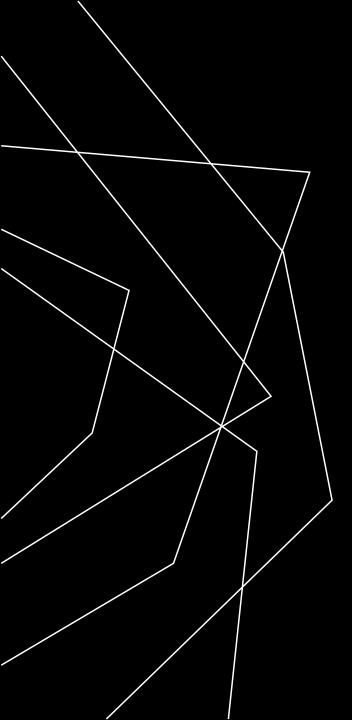https://ssbipolar.com/2019/10/14/power-bite-dataflows-vs-datasets/

https://radacad.com/what-are-the-use-cases-of-dataflow-for-you-in-power-bi

https://ssbipolar.com/2020/09/23/session-resources-patterns-for-adopting-dataflows-in-power-bi/

RESOURCES

[Dataflows Whitepaper – Amir Netz](#)

THANK YOU!

SLIDES CAN BE FOUND AT :

https://github.com/BenniDeJagere/Presentations

/{Year}/{Date}_{Event}