Power BI dataflows

The Missing Piece in Your Power BI Architecture!





Chuck Finance







RoseHuman Resources













RoseHuman Resources



















Power BI dataflows

The Missing Piece in Your Power BI Architecture!



Agenda

- What are dataflows?
- Why and how to use dataflows
- Common Data Model integration
- Azure Data Lake Storage gen2 integration
- Al features





Dave Ruijter

Data & Al Consultant Macaw Netherlands

@DaveRuijter

in daveruijter

MODELLA MARCHANIA MODELLA MARCHANIA MARCHANIA

moderndata.ai



Download for free!

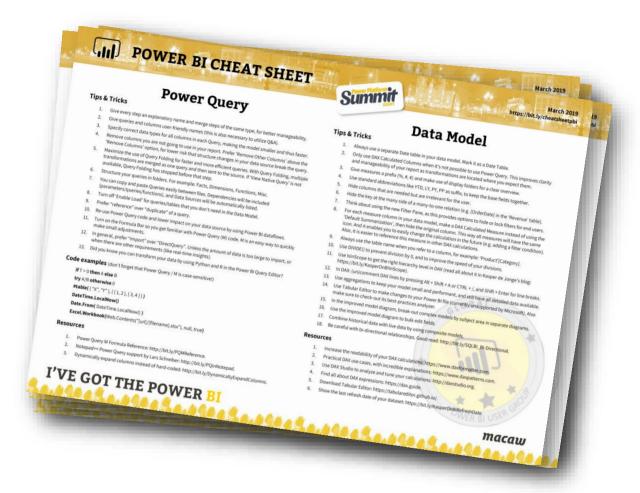
https://bit.ly/cheatsheetpbi



Dave Ruijter

Data & Al Consultant Macaw Netherlands

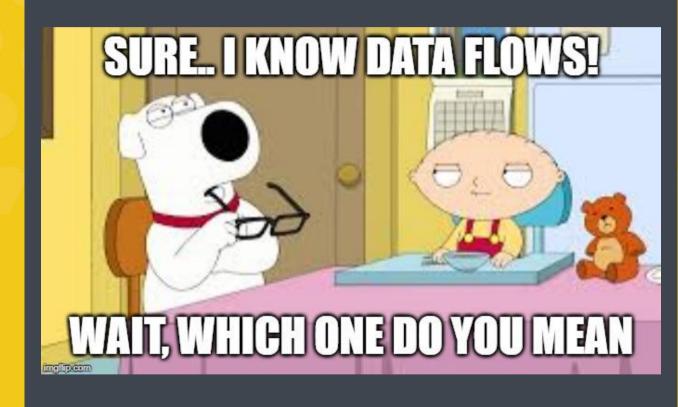
- ⊠ dave.ruijter@macaw.nl
- @DaveRuijter
- **in** daveruijter
- moderndata.ai





MORE PARTIES AND MORE PARTIES AND AND ASSESSED AND ASSESSED AS A STATE OF THE PARTIES AND ASSESS

Which 'data flow' are you talking about..?

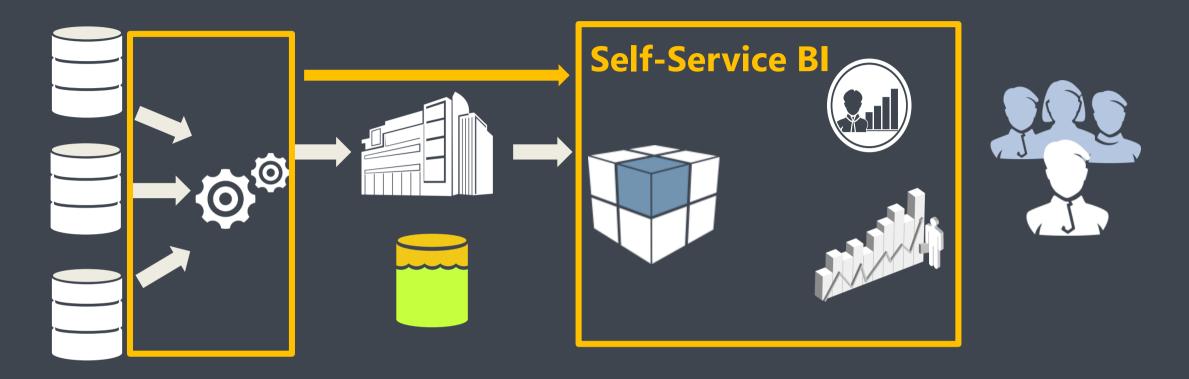




What are Power BI dataflows?

- A self-service low-code/no-code ETL tool in the Power BI Service
- A collection of entities (e.g. tables), fuelled by Power Queries
- They provide reuse of Power Queries and tables between Power BI datasets
- You create and manage them in an App Workspace





OLTP systems

Data Warehouse

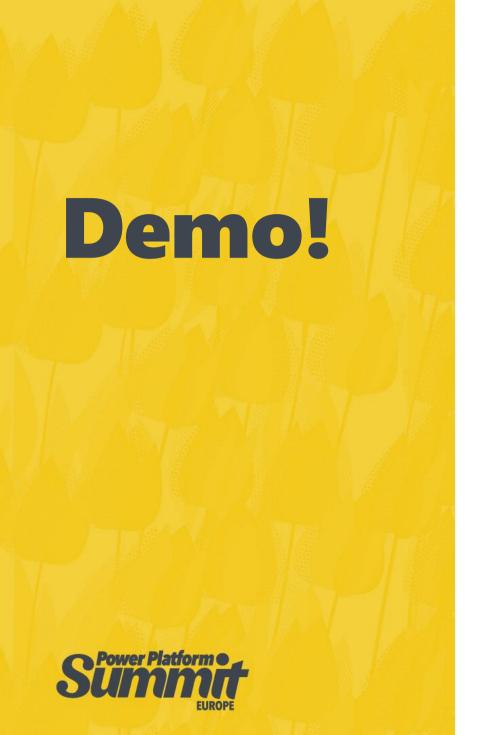
Reports and dashboards

Data Lake

Data Preparation / ETL

OLAP / Analytics Models





Intro: Power Bl dataflows

Why and when to use dataflows..



Why and when use dataflows?

- 1. When the DWH and/or Data Lake route is not an option (for now)
- 2. Reuse transformation logic
 - Share Power Queries for standard entities like Date dimension
 - Share and reuse advanced Power Queries
- 3. Reduce load on source systems
 - Less performance impact
 - Slow sources like SharePoint (throttling!)
- 4. Isolate data refreshes from different sources
- 5. Bridging the 'gap' to Azure Data Lake for the business user

AND MARIANDER MARIANDE MARIAND



Without dataflows, users will...

- Be blocked on IT involvement
- Use Excel and manual processes
- Require 3rd party data preparation tools like Alteryx, Datameer, Trifacta, etc.

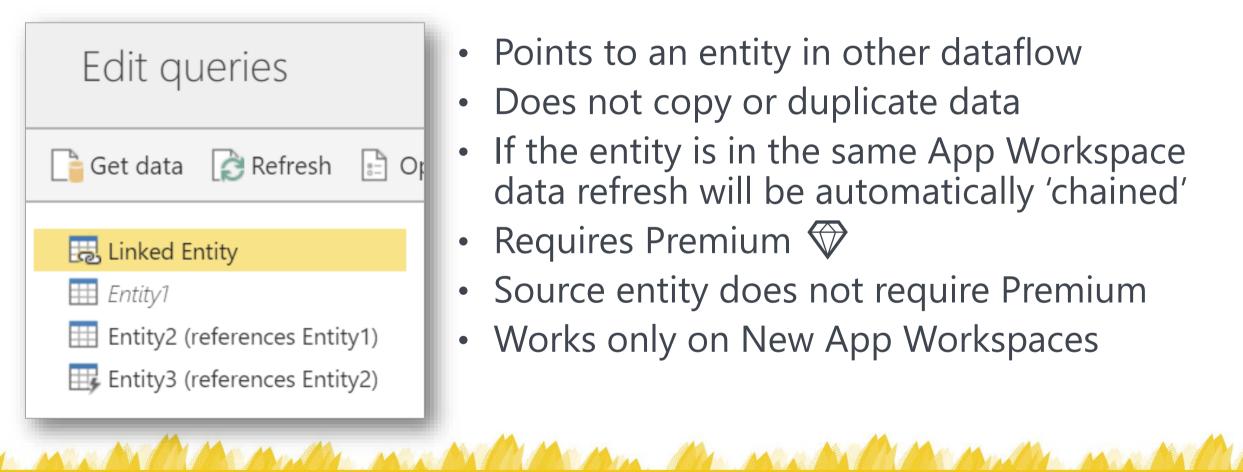


Connecting to / consuming dataflows

- From a dataflow by creating a 'Linked Entity'
- From a dataflow by creating a 'Computed Entity'
- From Power BI Desktop, just as any data source (you can save the dataset in any workspace!)



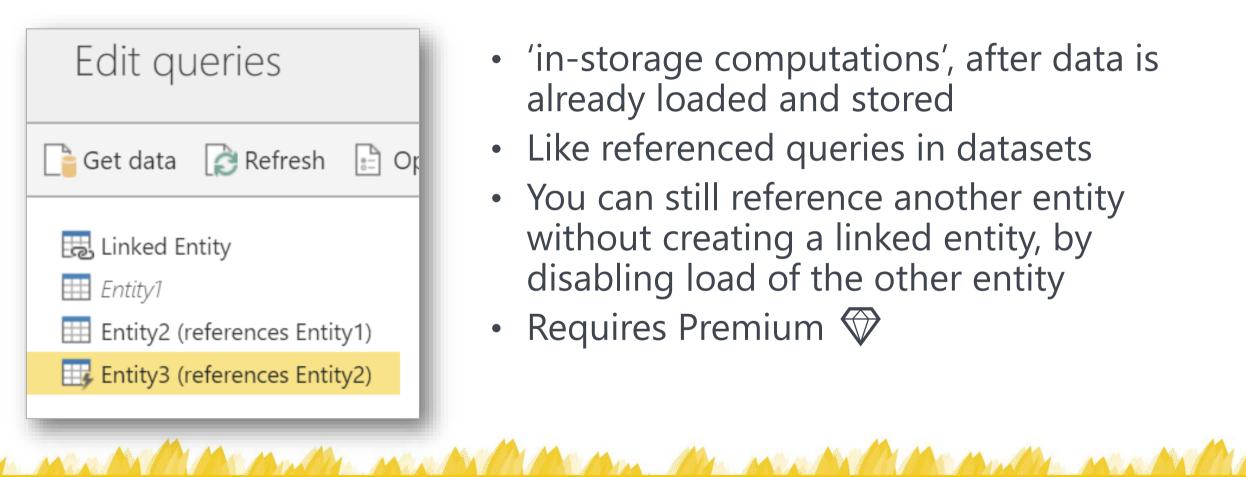
Linked Entities



- Points to an entity in other dataflow
- Does not copy or duplicate data
- If the entity is in the same App Workspace data refresh will be automatically 'chained'
- Requires Premium
- Source entity does not require Premium
- Works only on New App Workspaces

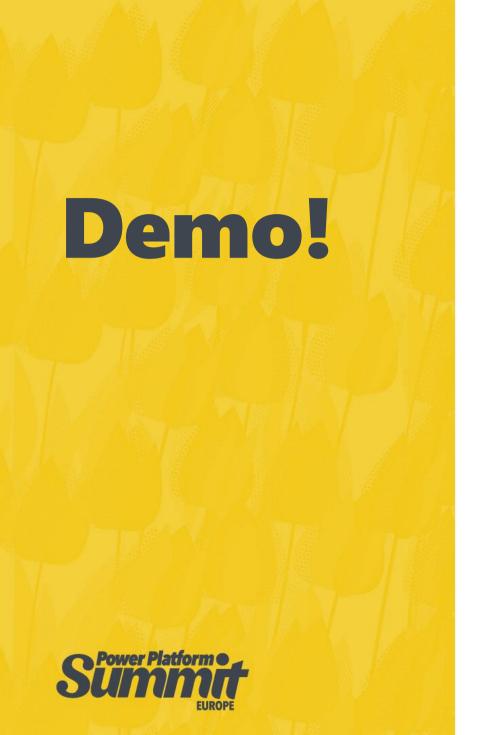


Computed Entities 4



- 'in-storage computations', after data is already loaded and stored
- Like referenced queries in datasets
- You can still reference another entity without creating a linked entity, by disabling load of the other entity
- Requires Premium





Linked Entities Computed Entities

Licensing notes

Pro

- Creating dataflows
- Consuming dataflows
- 8 refreshes
- 10 GB/user

Premium (EM3-P3, A1-A6)

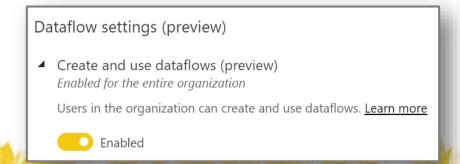


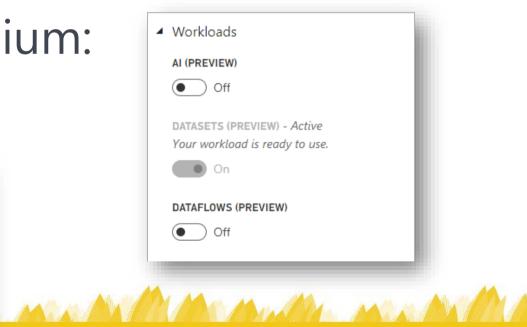
- Computed Entities
- **Linked Entities**
- External dataflow (CDM Folder)
- Incremental refresh
- 48 refreshes
- 100 TB/node



General notes (1)

- Still in preview
- Currently only creator can edit a dataflow
- Enable workload when in Premium:
- Tenant setting:







General notes (2)

- Personal Gateway is not supported
- Using Enterprise Gateway requires to be admin
- Each dataflow is limited to 1 gateway



Common Data Model integration



Common Data Model

- Provides a simple and consistent way of describing and saving data
- Streamline data exchange between solutions and organizations
- By default Power BI dataflows store data in the CDM data format
- You can map your entities to known CDM entities (known schemas)
- Benefit by quickly reusing out-of-the-box analytical solutions





Announcing the Open Data Initiative

A collection of standard data formats that make data exchange between applications and platforms a simple reality.





Azure Data Lake Storage Gen2 integration



Bring your own dataflow storage

CAROLINA MARKATAN MARKATAN MARKATAN MILITARIA DALAM MARKATAN MARKA

Still managed by Power BI

- Same as before, but now in your Storage Account in Your Azure subscription
- Power BI owns CDM folder, read-access for other services

Managed by you

- Create a CDM folder in your Storage Account, map it to a dataflow
- You own the CDM folder,
 Power BI has read-access, no
 Power Query edits possible





Datasets

Dataflows



Azure Data Factory



Azure Databricks





Azure ML

Azure Data Lake Storage Gen2

CDM folder

CDM folder

MODERAL PROPERTY MODERAL MODERAL PROPERTY MODERAL PROPERT

CDM folder

Business analysts

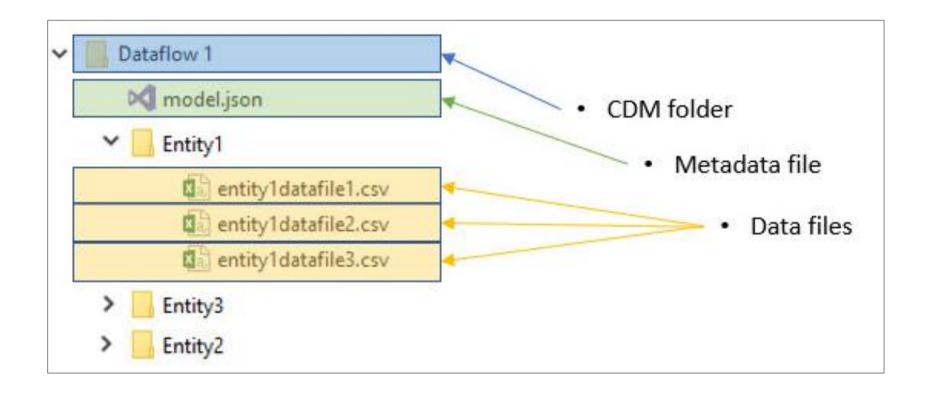
Low/no code

Data scientists Data engineers

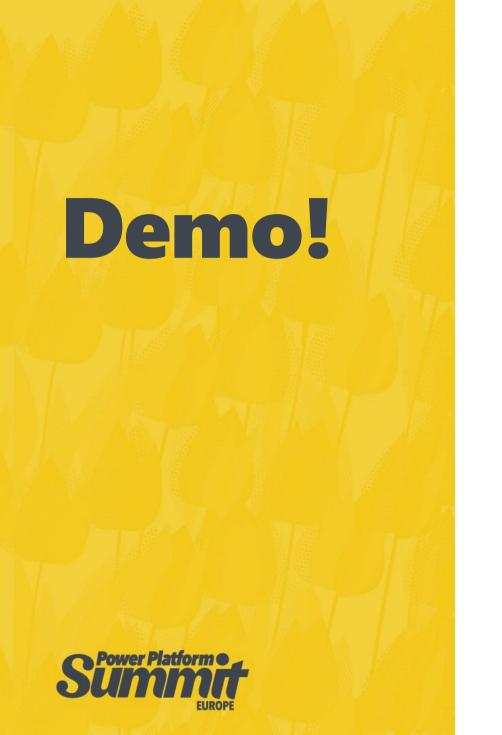
Low to high code



Dataflows in Azure Data Lake







Power BI dataflows & Azure Data Lake Storage Gen2

- Only for Workspaces v2
- Read permissions on the CDM folder, subfolders and files
- Execute permissions on the CDM folder and subfolders
- Once a dataflow storage location is configured, it cannot be changed
- When mapping the dataflow to the CDM folder, encode the uri, e.g. replace spaces with %20 etc.



- Dataflows based on CDM folders cannot be edited in Power BI
- Permissions to read a dataflow created from a CDM folder is managed by the owner of the CDM folder, and not by Power BI
- Only users who are authorized to both the workspace in which the dataflow was created, and the CDM folder, can access its data from the Power BI dataflows connector
- Linked entities are not available for dataflows created from CDM folders

MARCHANDER MARCHANDE MARCHAN BOLLAND BY MARCHANDER MARC



- When PBI manages storage: Linked Entities and Computed Entities only work when entities reside in the same storage account
- When you manage storage (and use CDM folders): Linked entities are not available



- When doing computations on data joined by on-premises and cloud data, create a new entity to perform such computations.
- This provides a better experience than using an existing entity for computations, such as an entity that is also querying data from both sources and doing in-lake transformations.



Al features

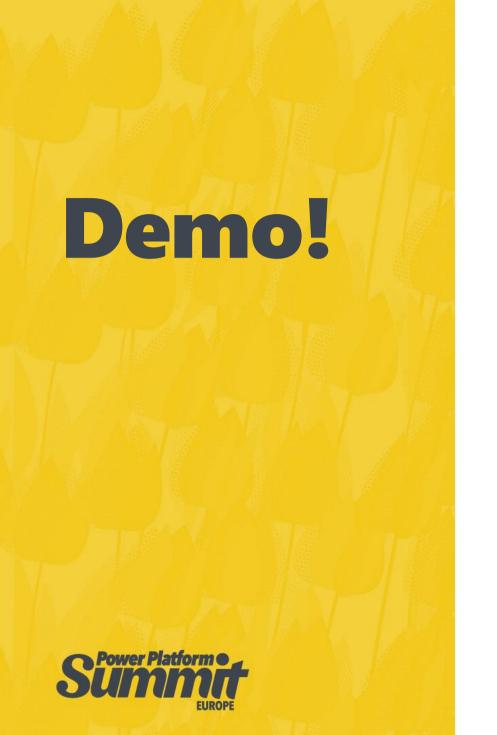


Al features in dataflows

- 1. Invoke Cognitive Services to enrich your data
- 2. Easily invoke your Azure Machine Learning models
- 3. Create new Machine Learning models with AutoML

Million 18 million mars 1861 Marian Maria Maria 1861 Maria Maria 1861 Million 18 million 1861 Mi





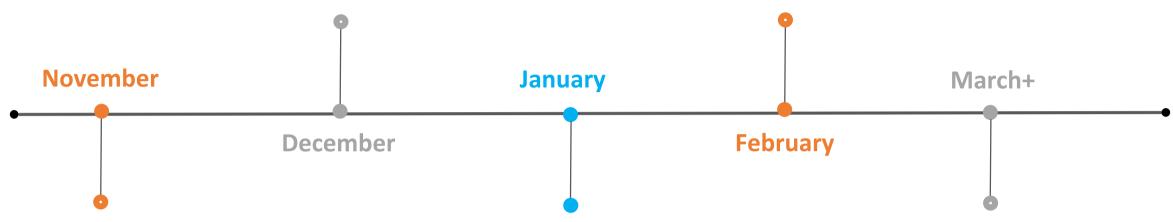
Power Bl dataflows & Al features

Roadmap for dataflows

- Over time, Azure Machine Learning, Azure Databricks, Azure Data Factory, etc. will have built-in support to read and write Common Data Model-compliant folders
- Shared ownership (edit dataflows of other authors)
- Enhancements to the dataflow calculation engine to enable high-performing large-scale data transformations (TBs)







- Dataflows Public Preview
 - ✓ Power Query Online integration + Connectors
 - ✓ Computed entities
 - ✓ Linked entities
 - ✓ Partition-based Incremental Refresh

✓ Native Queries support

CANAL MARKANIA MARKANIA

✓ New connectors support (MySQL, Teradata, Azure Data Explorer, AAD, Redshift, SmartSheet, Sybase)

Dataflows GA

- ✓ Email notifications on failed refresh
- ✓ Dataflows "Take Ownership"
- ✓ Cancel refresh of a Dataflow
- Perf Enhancement: SQL Compute engine (preview)
- Preserve Entity Relationships from data source
- ADLS dataflow access governed by workspace permissions (SAS)



Power BI dataflows FAQ

- ✓ A new 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

MARKETS JOHN MERCHEN MERCHEN MICHELLEN JOHN MERCHEN MERCHEN MICHAEL MERCHEN MICHAEL MI

NOT spelled with a space or any capital letters



Session resources

- Dataflows on public Microsoft sites:
 - Dataflows documentation: https://docs.microsoft.com/en-us/power-bi/service-dataflows-overview
 - Dataflows roadmap / release notes: https://docs.microsoft.com/en-us/business-applications-release-notes/April19/business-intelligence/power-bi-service/power-bi-dataflows/self-service-data-prep-with-dataflows
 - Dataflows on Power BI Ideas: https://ideas.powerbi.com/forums/265200-power-bi-ideas?category_id=341638
- Common Data Model on public Microsoft sites: :
 - Common Data Model documentation: https://docs.microsoft.com/en-us/common-data-model/
 - Common Data Model on GitHub: https://github.com/Microsoft/CDM
 - End to end CDM partner sample: https://aka.ms/cdmadsblog
- Al in Power Bl
 - Machine Learning in Power BI: https://docs.microsoft.com/en-us/power-bi/service-machine-learning-integration
 - Cognitive Services in Power BI: https://docs.microsoft.com/en-us/power-bi/service-cognitive-services
- Matthew Roche's blog:
 - Dataflows landing page: https://ssbipolar.com/2018/10/23/dataflows-in-power-bi/
 - Dataflows FAQ: https://ssbipolar.com/2018/11/27/power-bi-dataflows-faq/



Service Platform Service Supply Service Service Supply Service Supply Service Supply Service Service Supply Service Servi