

Power BI dataflows?

Why you need to use them!

Benni De Jagere, Realdolmen



Explore your PASS community

*Own your career with interactive learning built
by community and guided by data experts.
Get involved. Get ahead.*



 PASS
MARATHON

Free online
webinar events



PASS
LOCAL
GROUPS

Local user groups
around the world



 PASS
SUMMIT

Connect with the
global data community



PASS
VIRTUAL
GROUPS

Online special
interest user groups

 PASS.org

 PASS CONNECTOR
INSIGHTS

Learning on-demand
and delivered to you



PASS
VOLUNTEERS

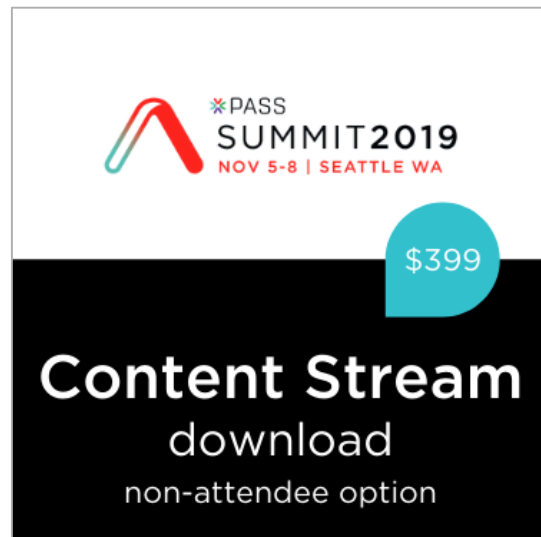
Get involved

Missed PASS Summit 2019?

Get the Recordings

Download all PASS Summit sessions on Data Management, Analytics, or Architecture for only \$399 USD

More options available at
PASSstuff.com





The First PASS Virtual Summit

You don't want to miss this.

Register Now

We are thrilled to say the first ever PASS Virtual Summit will be taking place this November!

We will be launching PASS Summit into a whole new world, and with that comes exciting opportunities with what we can offer you. We will go beyond simply moving sessions online to fully embracing everything a virtual environment can offer.

Find out more at PASS.org/summit





Thank you to
our Global
Sponsors and
Supporters





And also to our
Local Sponsors



Hello, who dis?



Benni De Jagere?

Senior Data Insights Consultant  REALDOLMEN
a Gft Group company

dataMinds.be, Co-leader

 @BenniDeJagere

 /bennidejagere

#SayNoToPieCharts



Data Platform, March 2020



Dataflows?

~~(Google) Cloud Dataflow~~

~~TensorFlow Dataflow~~

~~SSIS Data Flows~~

~~Microsoft Flow~~

~~ADFe2~~

~~Mapping Data Flows~~

~~Wrangling Data Flows~~

~~PowerApps dataflows~~

Power BI dataflows!



Who?



Who should be using Dataflows

- (Financial) Power Users
- Business Analysts
- Tech savvy Users
- Everyone else 😊



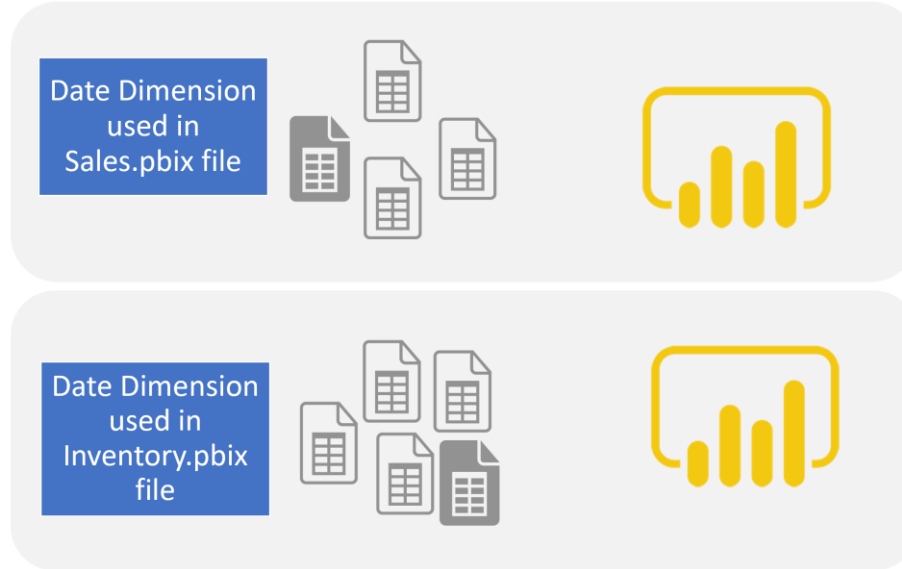
What?



Only Entity, Many Datasets

- Lack of semantic model will create many versions
- Re-use of queries across datasets is limited
- Shared Datasets is all or nothing approach

Only Entity, Many Datasets



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



Only Entity, Many Datasets



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



Why?



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**
- Manageable and securable like all Power BI Objects
- Managed CDM Folders



Demo - Dataflows



New York Citibikes

- 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?quality=90&strip=all&ssl=1>



*PASS
SQLSATURDAY

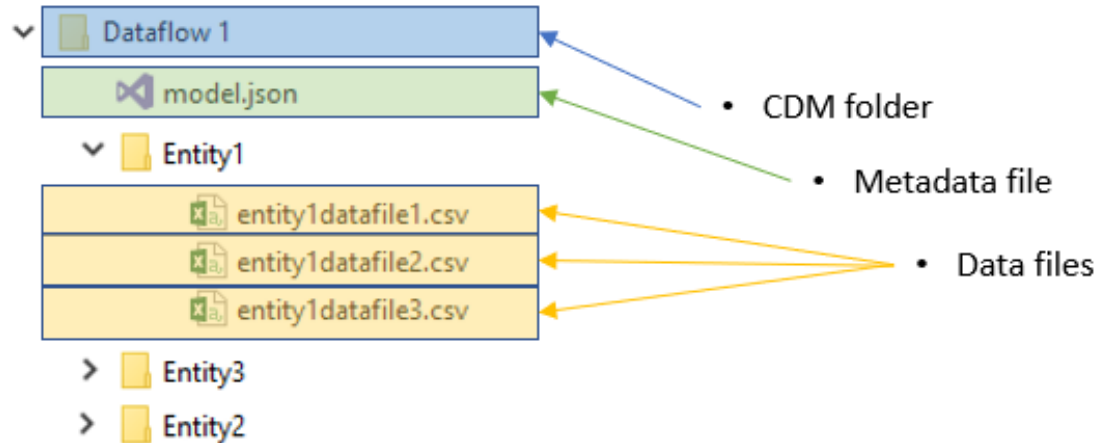
How? Common Data Model



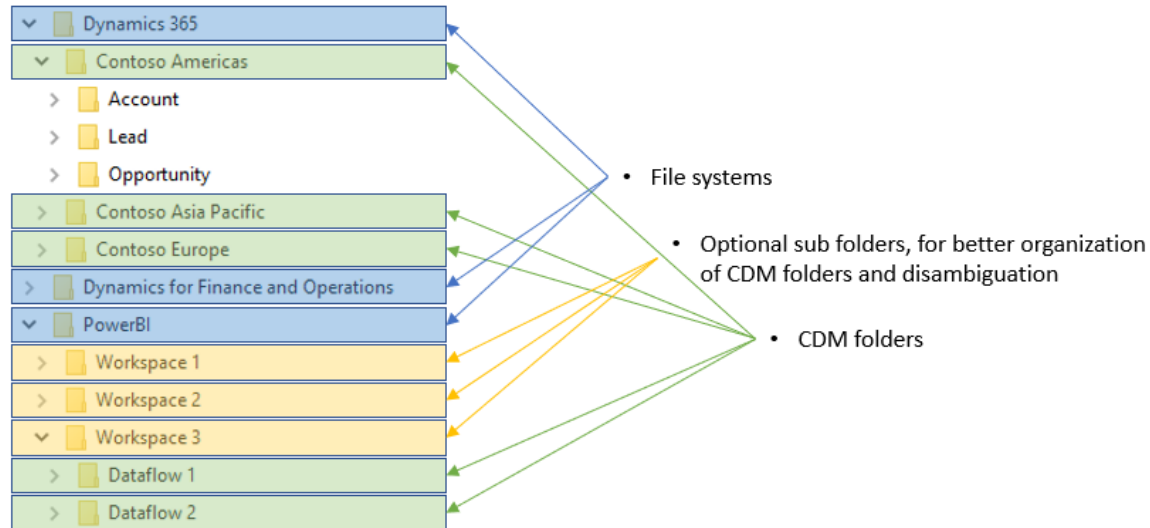
Common Data Model

- 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





*PASS
SQLSATURDAY

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 dataflows it is dependent on
- Incremental Refresh is Premium Only (for now?)



Demo - Scheduled Refresh





*PASS
SQLSATURDAY

How? Query Reusage



Query Reusability

- No developer experience in Power BI Desktop
 - Copy/Paste queries between environments
- Linked/Computed Entities (Premium Only)
 - In-lake Compute, which requires dedicated capacity
 - Can Reference any Entity in any Workspace
 - Data is only stored in the original Entity
- Reference Query, Append or Merge
 - Create a query, and mark 'Disable for load'
 - Reference this query, in the same workspace
 - Logic is reused, and data is duplicated every time the query is referenced

Demo – Query Reusage



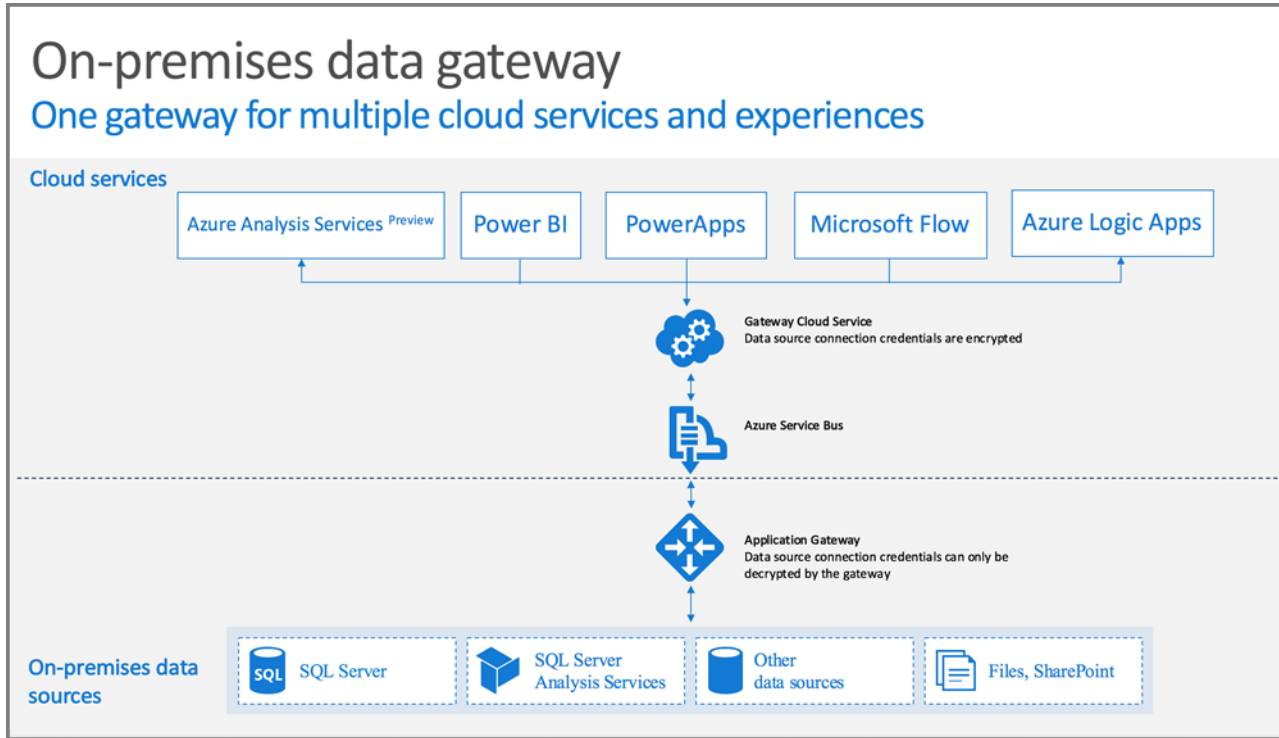


*PASS
SQLSATURDAY

Gateway & Security



On-Premises Datasources



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!



Demo: Gateway & Security





*PASS
SQLSATURDAY

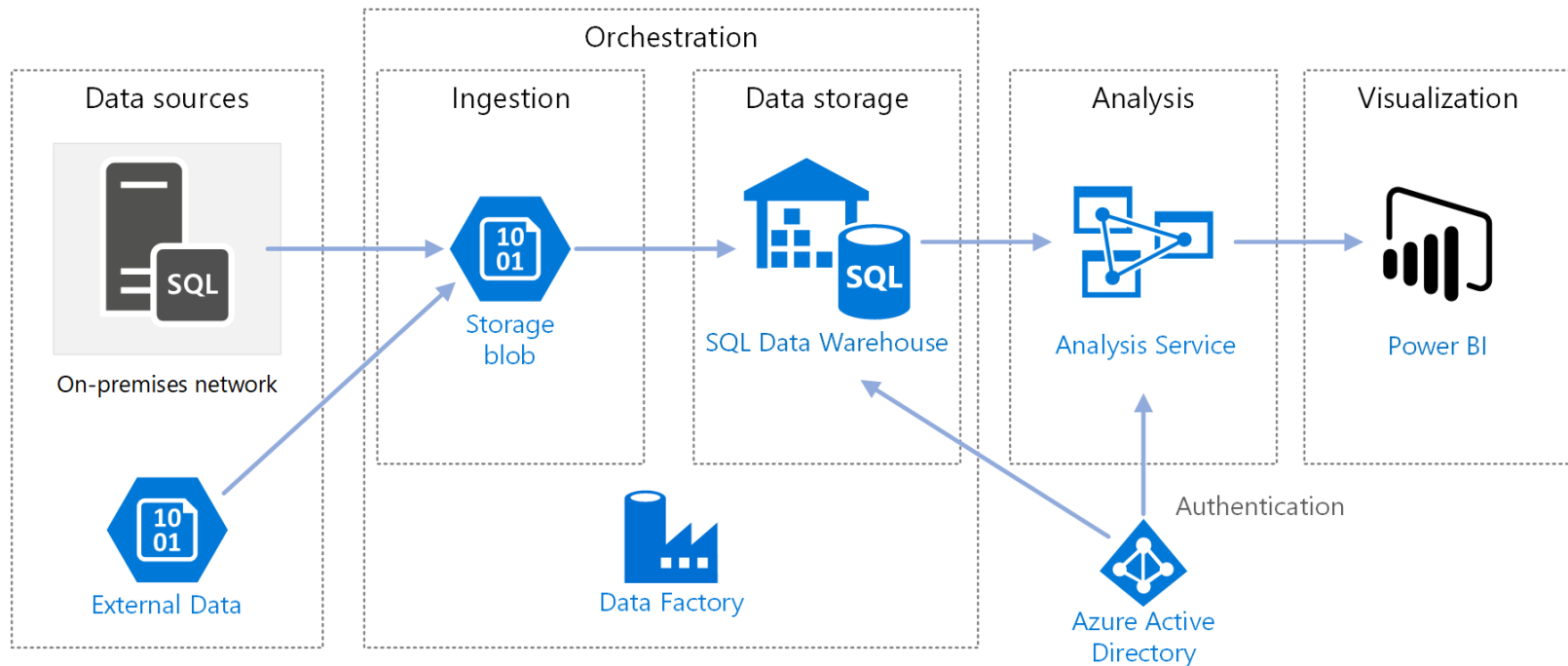
The Big Picture



The Big Picture



The Big Picture





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
- 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



*PASS
SQLSATURDAY

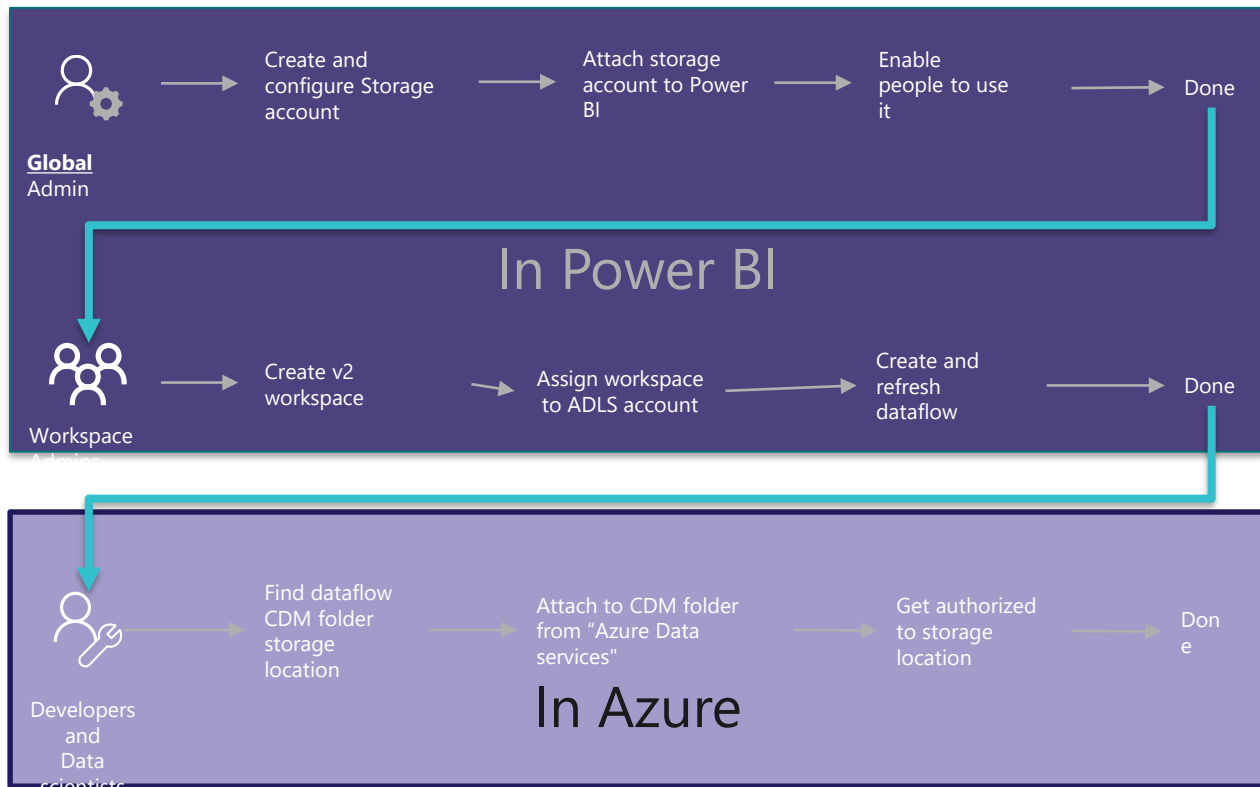
Extending Dataflows



Bring your own Storage Account (OLD)

- Use your organizational ADLSg2 resource and attach it to Power BI
- A Global Admin and an Azure subscription is required
- Once a dataflow storage location is configured, it cannot be changed
- By default, only owners of a dataflow stored in Azure Data Lake Storage Gen2 can access its data
- To authorize additional users to the dataflows stored in Azure, you must add them to the dataflow's CDM folder
- Creating dataflows with linked entities is only possible when they are stored in the same storage account

Bring your own Storage Account (OLD)



What "done" means:

Workspace Admins can now configure workspace to store its dataflows in ADLS account Global Admin configured

Dataflow CDM folder is stored in ADLS and is accessible by its creator ONLY.

Other people in workspace can get data if they are authorized to the CDM folder outside of Power BI

Now data can be read from CDM Folder via Azure Data Services or LOB solutions that are CDM folder aware

Bring your own Storage Account (OLD)

- Use your organizational ADLSg2 resource and attach it to Power BI
- A Global Admin and an Azure subscription is required
- Once a dataflow storage location is configured, it cannot be changed
- By default, only owners of a dataflow stored in Azure Data Lake Storage Gen2 can access its data
- To authorize additional users to the dataflows stored in Azure, you must add them to the dataflow's CDM folder
- Creating dataflows with linked entities is only possible when they are stored in the same storage account

Bring your own Storage Account

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

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

☐ Use the default Azure connection

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

bmpch ▾

Save

Cancel

☐ Use the default Azure connection

External CDM Folders

- CDM Folders is an open standard
- Any tool that can access ADLSg2, and output JSON
- Link External CDM Folder in Power BI
- Done!

☆ Favorites >

🕒 Recent >

🗑 Apps

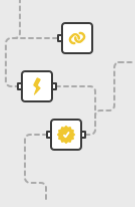
👤 Shared with me

📁 Workspaces >

👤 Workspace

🔒

⌵



Add entities to 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

Attach an external CDM folder

Attach a CDM 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



Attach an external CDM folder to a new dataflow

Name *

Sales Forecasting

Description

Dataflow entities with output from Azure Machine Learning models for detailed sales forecasts.

CDM folder path *

Enter the path to your CDM folder in Azure Data Lake Storage Gen2

Create and attach

Cancel

Define

Choose to define the dataflow entities by attaching data to an existing CDM folder or by defining new CDM entities instead. [Learn more](#)

External

or from
ce
int to a
u can

Add new entities

Add linked entities

Create and attach



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



Recap



Why you should use it!

- Business Driven Initiatives
- Fill existing gaps
- Quick prototyping
- Strive for IT handover to operationalize later on
- Configure your own ADLSg2 Storage

The Self-Service BI Hoax

“Products and technology don’t enable Self-Service BI,
people and methodologies do”

[Nimrod Avissar on LinkedIn Pulse](#)

Resources

- <https://docs.microsoft.com/en-us/power-bi/service-dataflows-best-practices>
- <https://powerbi.microsoft.com/en-us/blog/power-bi-dataflows-and-azure-data-lake-storage-gen2-integration-preview/>
- <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://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>

Resources

- [Dataflows Whitepaper – Amir Netz](#)
- [Release Notes & Roadmap](#)
- [CDM Tutorial](#)
- [CDM Documentation](#)
- [Dataflows and CDM Sessions at MBAS](#)

Benni De Jagere?

Senior Data Insights Consultant

dataMindse, Co-leader

 @BenniDeJagere

 /bennidejagere

#SayNoToPieCharts



ata Platform, March 2020





This event was sponsored by Microsoft

Learn more about SQL Server 2019 today:

- Get free training: aka.ms/sqlworkshops
- Download the SQL19 eBook: aka.ms/sql19_ebook

Exit

