



COMMUNITY SUMMIT

POWERED BY DYNAMIC COMMUNITIES

Power BI dataflows?

Why you need to use them!

Benni De Jagere

Benni De Jagere?

Senior Data Insights Consultant

dataMinds.be, Co-leader

 @BenniDeJagere

 /bennidejagere

#SayNoToPieCharts



Dataflows?

- ~~(Google) Cloud Dataflow~~
- ~~TensorFlow Dataflow~~
- ~~SSIS Data Flows~~
- ~~Microsoft Flow~~
- ~~ADFv2~~
 - ~~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 😊

Why?

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
used in
Sales.pbix file



Date Dimension
used in
Inventory.pbix
file



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



What?

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

The Demo Case

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>

Demo: Dataflows

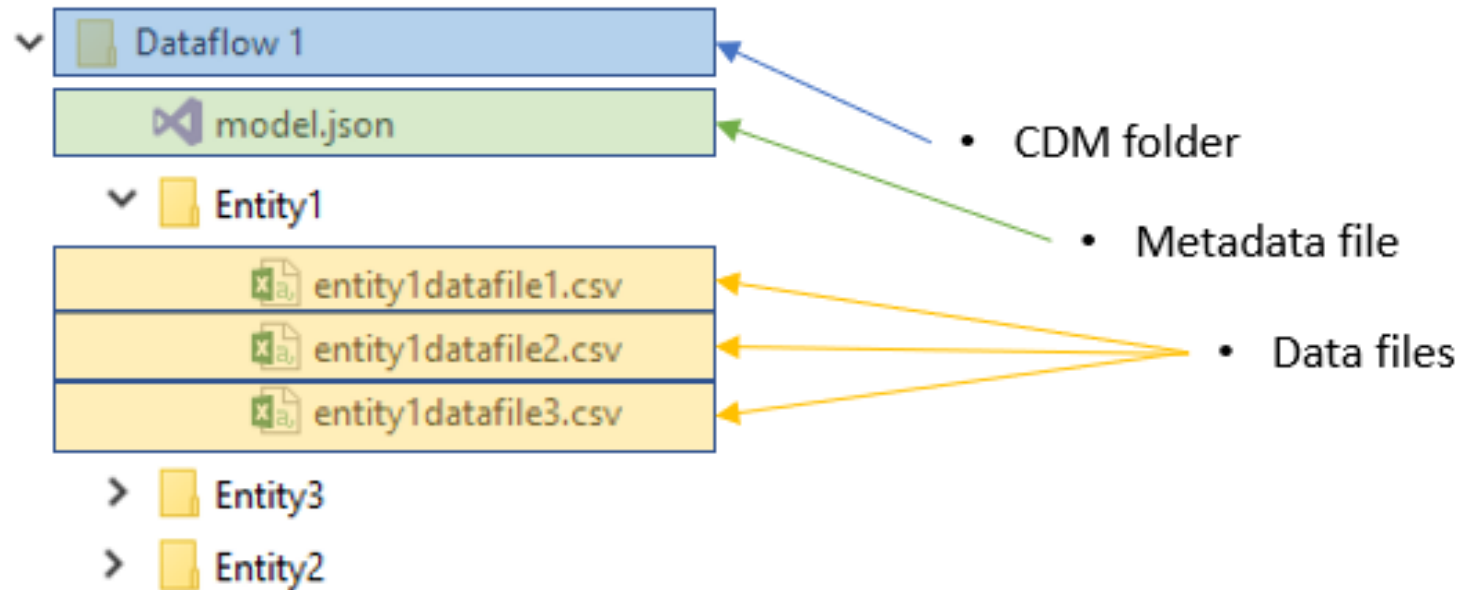
How?

Common Data Model

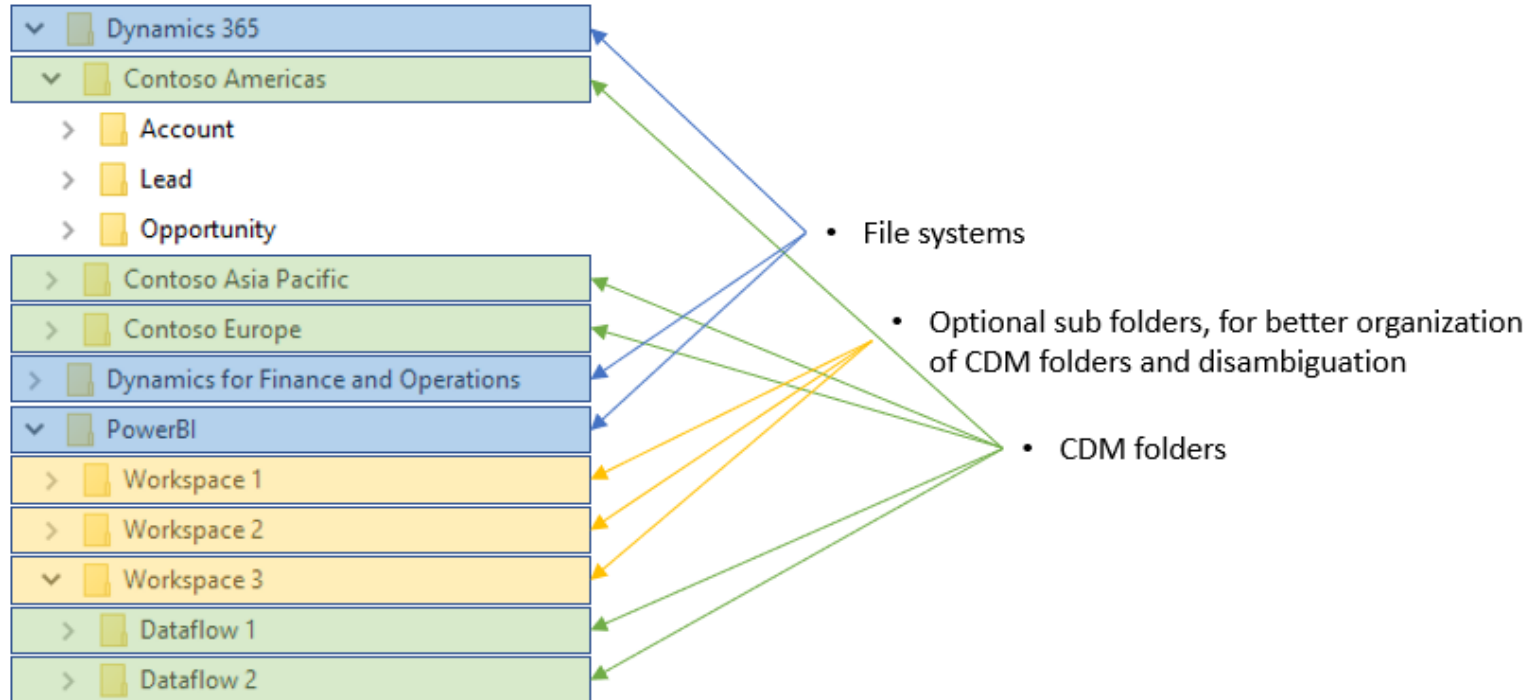
Common Data Model

- 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



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

How?

Query Reusage

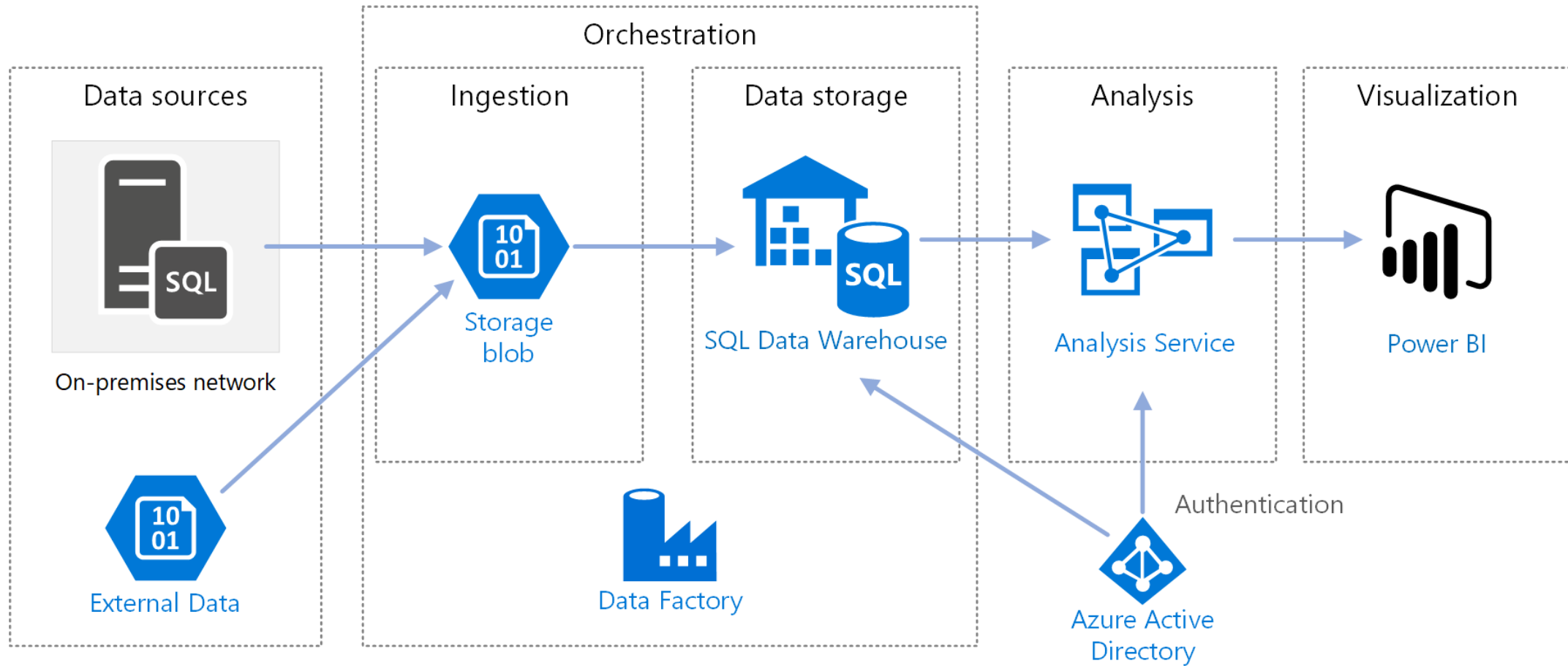
Query Reusability

- 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

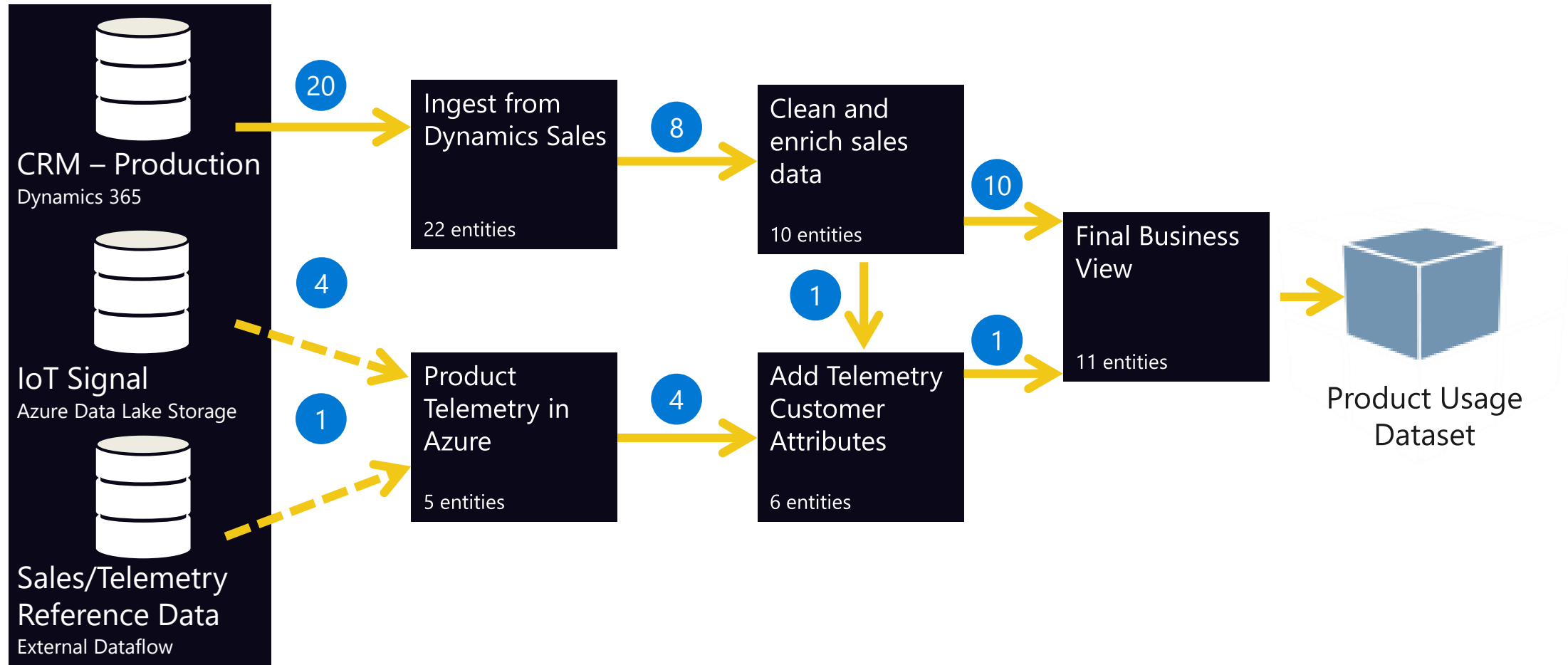
Demo: Query Reusage

The Big Picture

The Big Picture



The Big Picture



Why?

Where do Dataflows fit in?

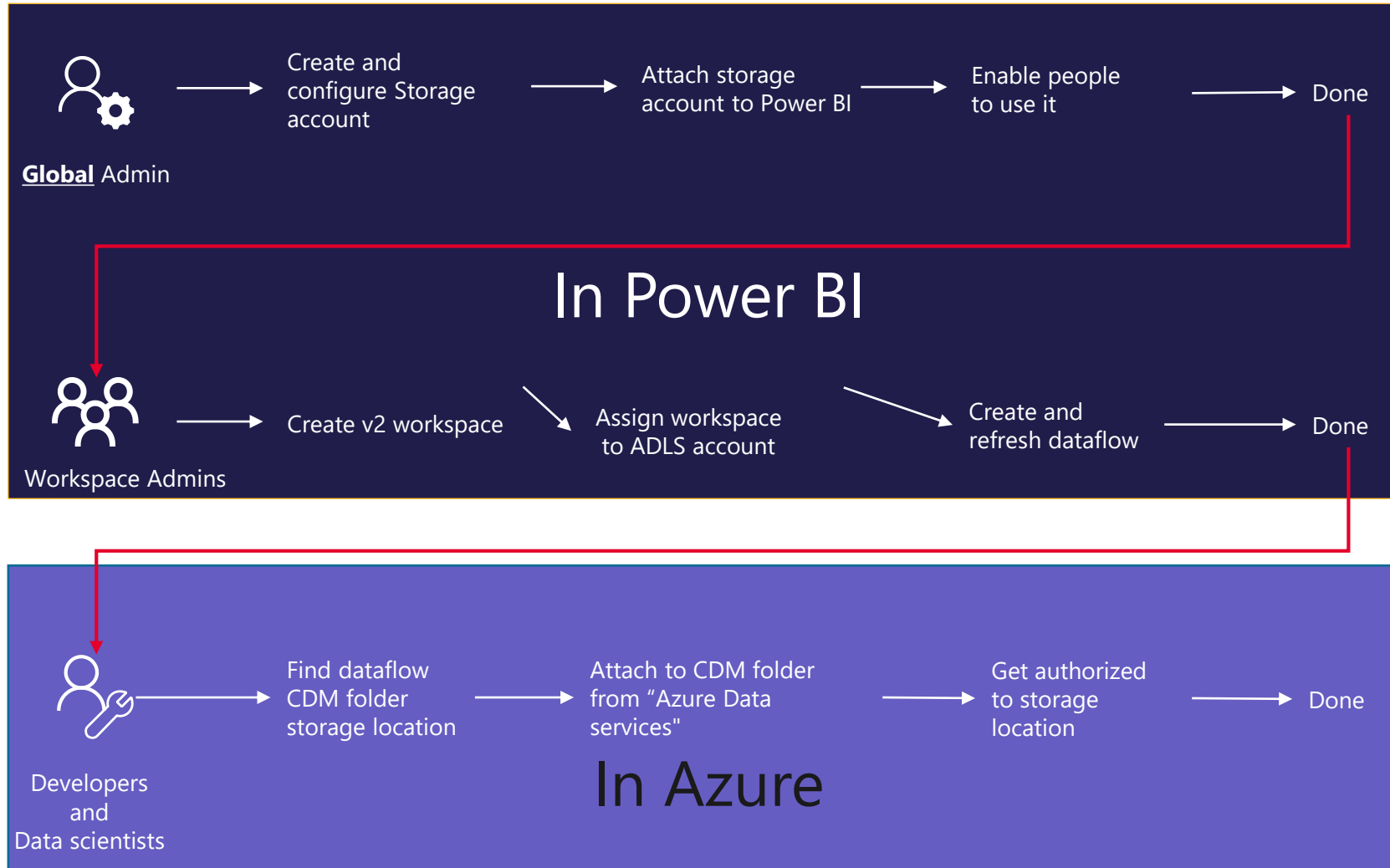
- Without dataflows, users might
 - Be blocked on IT involvement
 - Use Excel and manual processes
 - Require 3rd party data preparation tools
- Dataflows are for Power Query users
 - Easy to build reusable data entities
 - Easier to compose and orchestrate
 - Multi-stage, multi-user data prep workflows
 - Dataflows fill a self-service gap in the end-to-end story

Extending Dataflows

Bring your own Storage Account

- 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



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

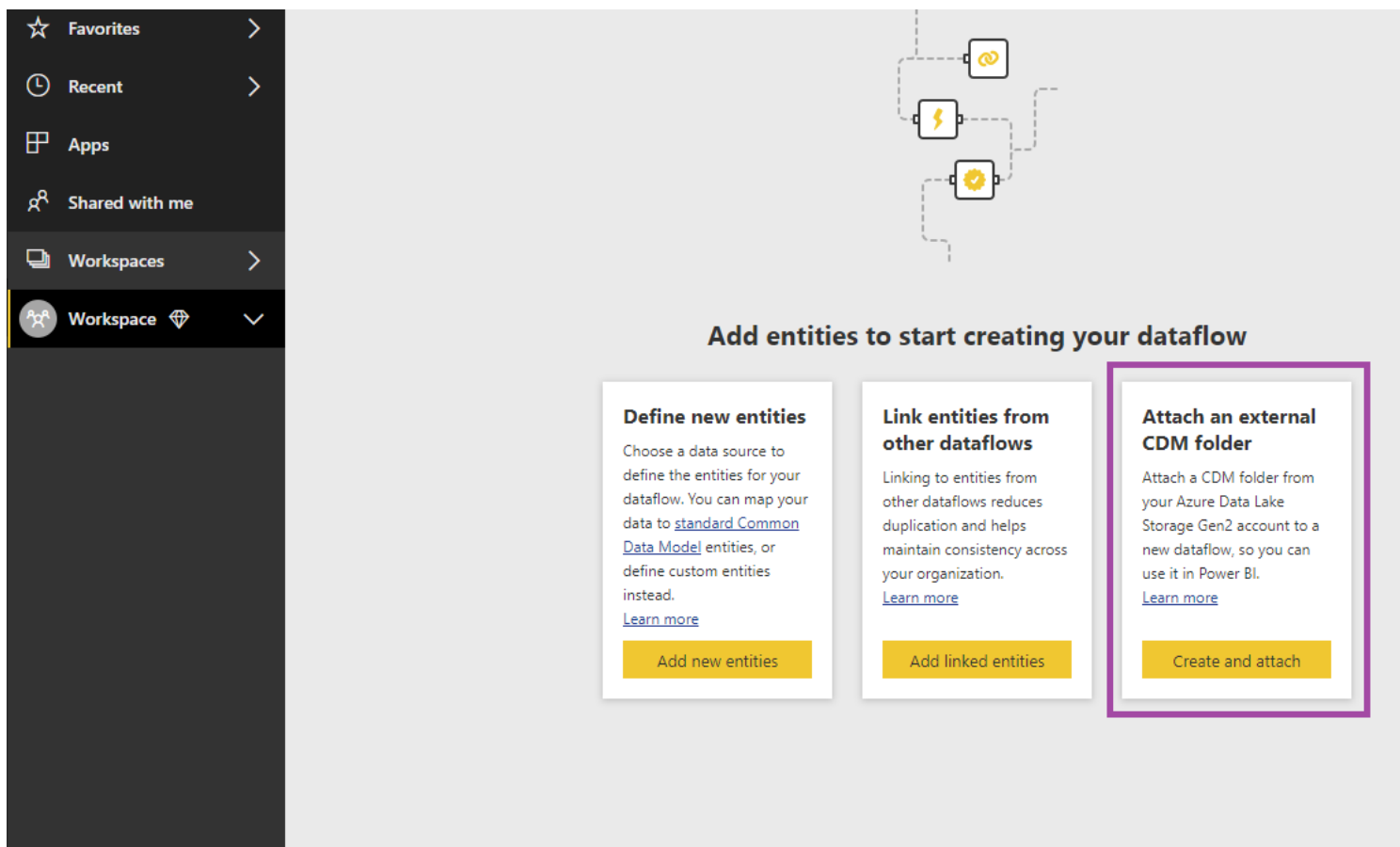
Bring your own Storage Account

- 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

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!

External CDM Folders



The screenshot displays the Databricks workspace interface. On the left is a dark sidebar with navigation options: Favorites, Recent, Apps, Shared with me, Workspaces, and Workspace (selected). The main area has a light gray background. At the top, there's a diagram showing three interconnected nodes with icons: a circle with a dot, a lightning bolt, and a gear. Below this, the heading 'Add entities to start creating your dataflow' is centered. Three white cards are arranged horizontally, each with a yellow button at the bottom. The third card, 'Attach an external CDM folder', is highlighted with a purple border. It contains text about attaching a CDM folder from an Azure Data Lake Storage Gen2 account to a new dataflow for use in Power BI, with a 'Learn more' link.

Workspace

- 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

External CDM Folders

Attach an external CDM folder to a new dataflow

Name *

Description

CDM folder path *

Create and attach Cancel

Define

Choose to define the dataflow data to generate from a data model or from a data source. If you choose a data model, you can define the dataflow entities instead. [Learn more](#)

Add new entities

External

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

Create and attach

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://excleratorbi.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-bi-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

dataMinds.be, Co-leader

 @BenniDeJagere

 /bennidejagere

#SayNoToPieCharts



THANK YOU!

