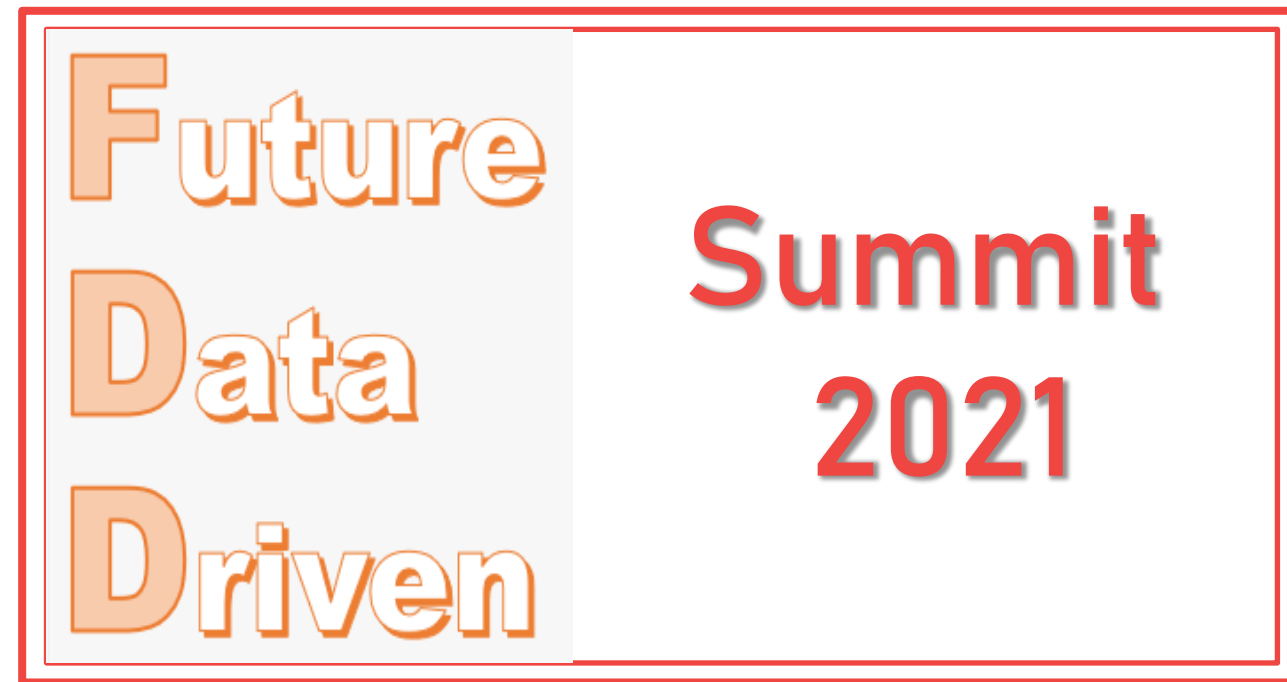


Power BI dataflows? Why you need to implement it!

Benni De Jagere





Power BI dataflows? Why you need to implement it!



Benni De Jagere

Enterprise Data Architect, Data Platform MVP

Inetum-Realdolmen **inetum.**
realdolmen
Positive digital flow



About me



dataMinds.be Co-Leader



@BenniDeJagere



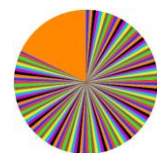
/bennidejagere



sessionize /bennidejagere



Data Platform



#SayNoToPieCharts



dataflows?

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 be using dataflows?

(Financial) Power Users

Business Analysts

Tech savvy Users

Everyone else 😊



What?

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

Will “Direct Query over Power BI datasets and Azure Analysis Services” change the way we work?

One 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

One Entity, Many Datasets



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



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



Power BI dataflows

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

Common Data Model

Unified metadata layer for Cloud Apps and Services

Shared

- Data Language

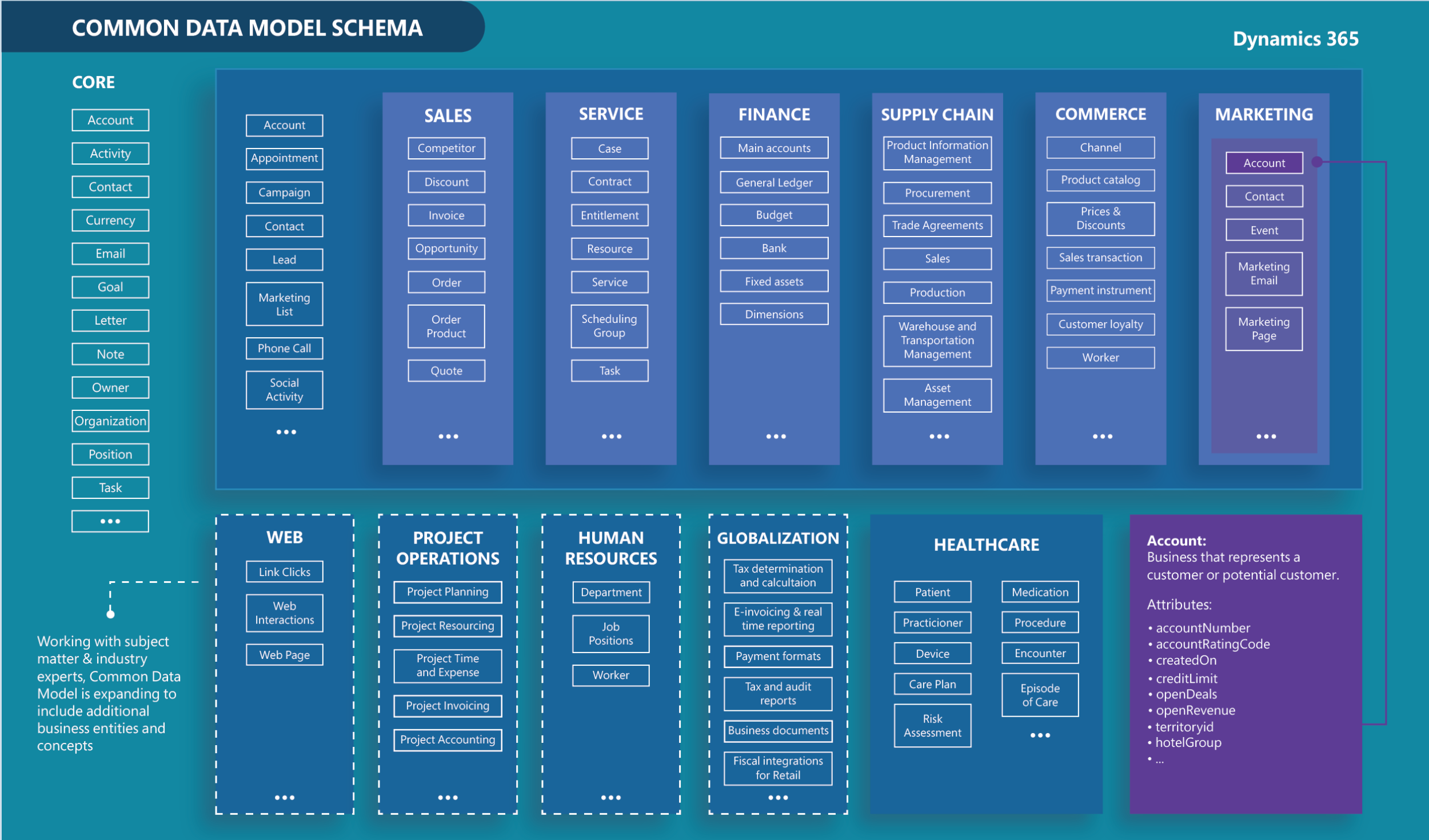
- Storage Format and Location

- Set of Default Entities

Extend Entities for all consumers

Simplified integration and disambiguation of data

INTEGRATE & DISAMBIGUATE DATA WITH
THE COMMON DATA MODEL

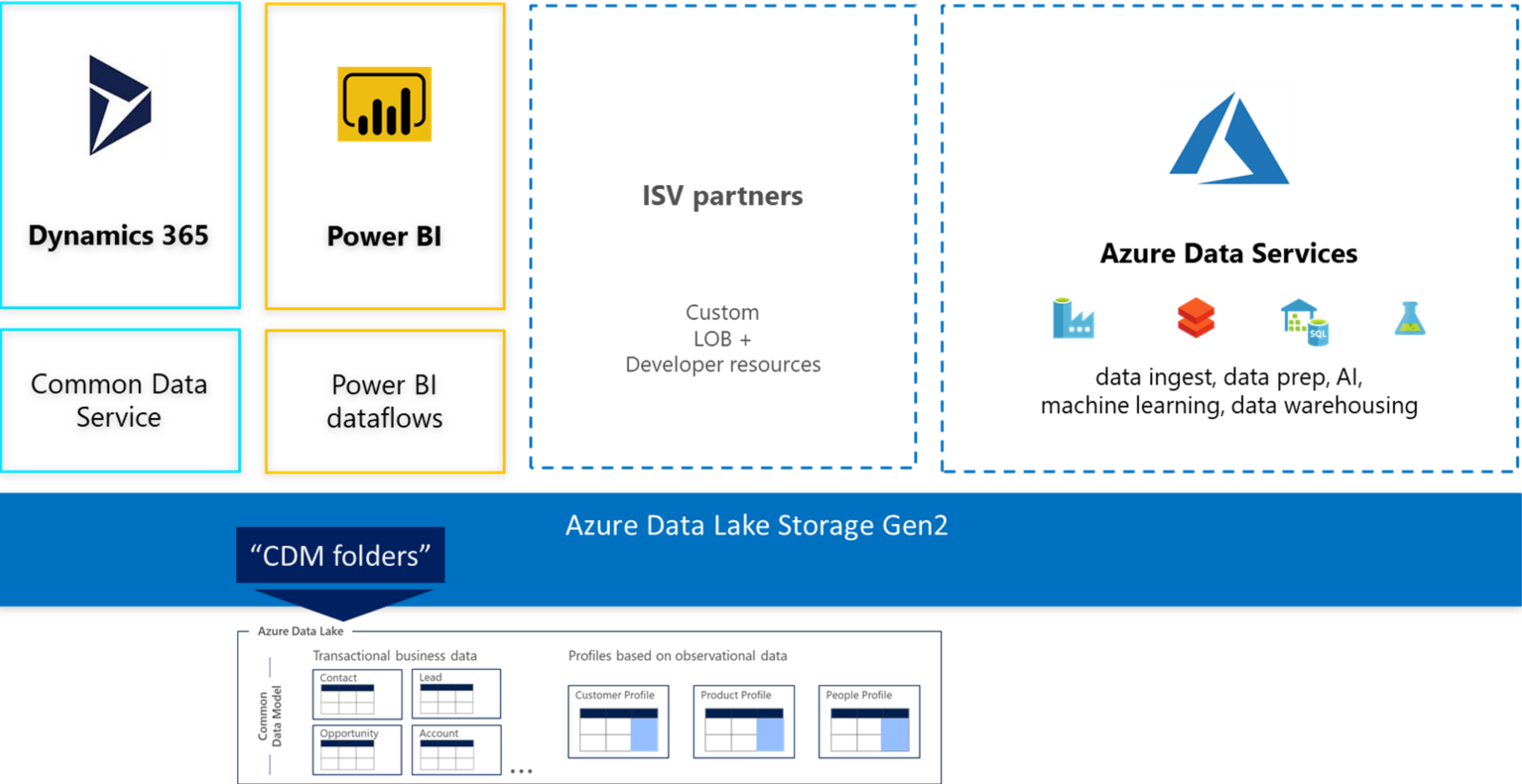


Decision Makers, Business analysts

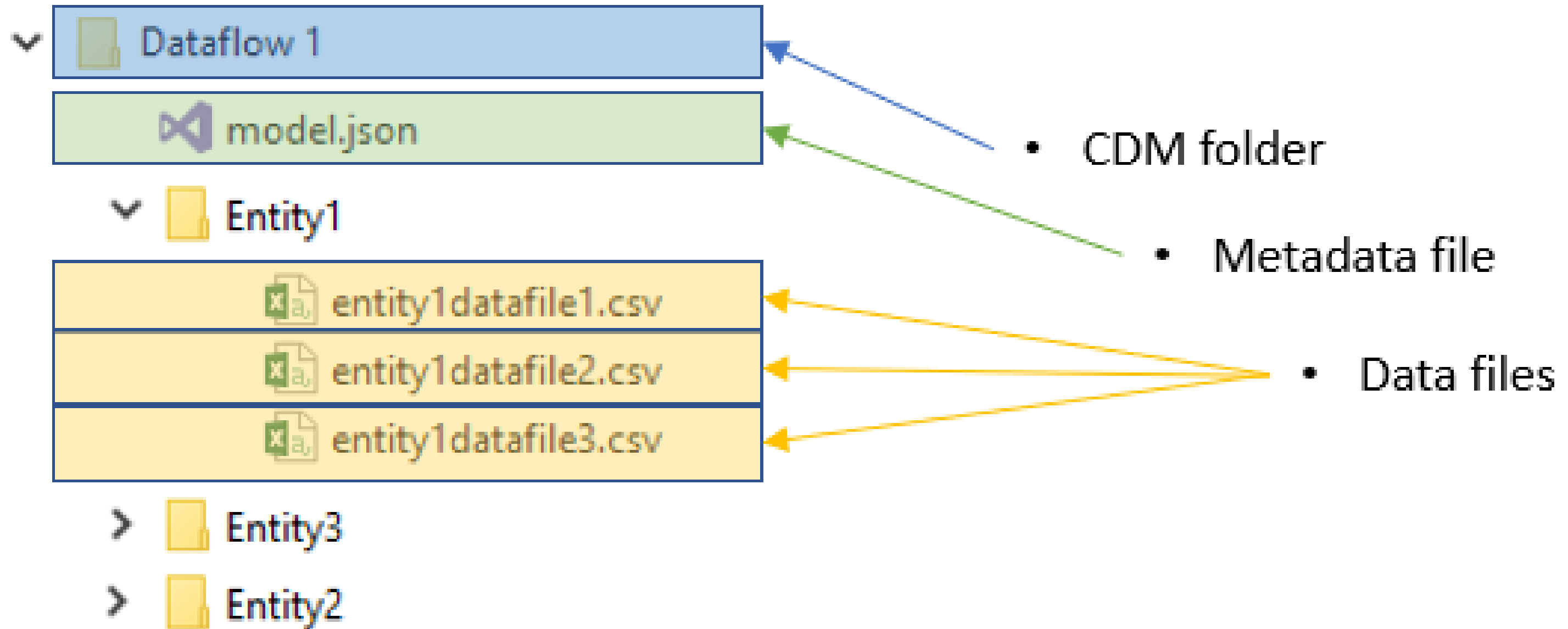
No code, low code

Data scientists, Data engineers

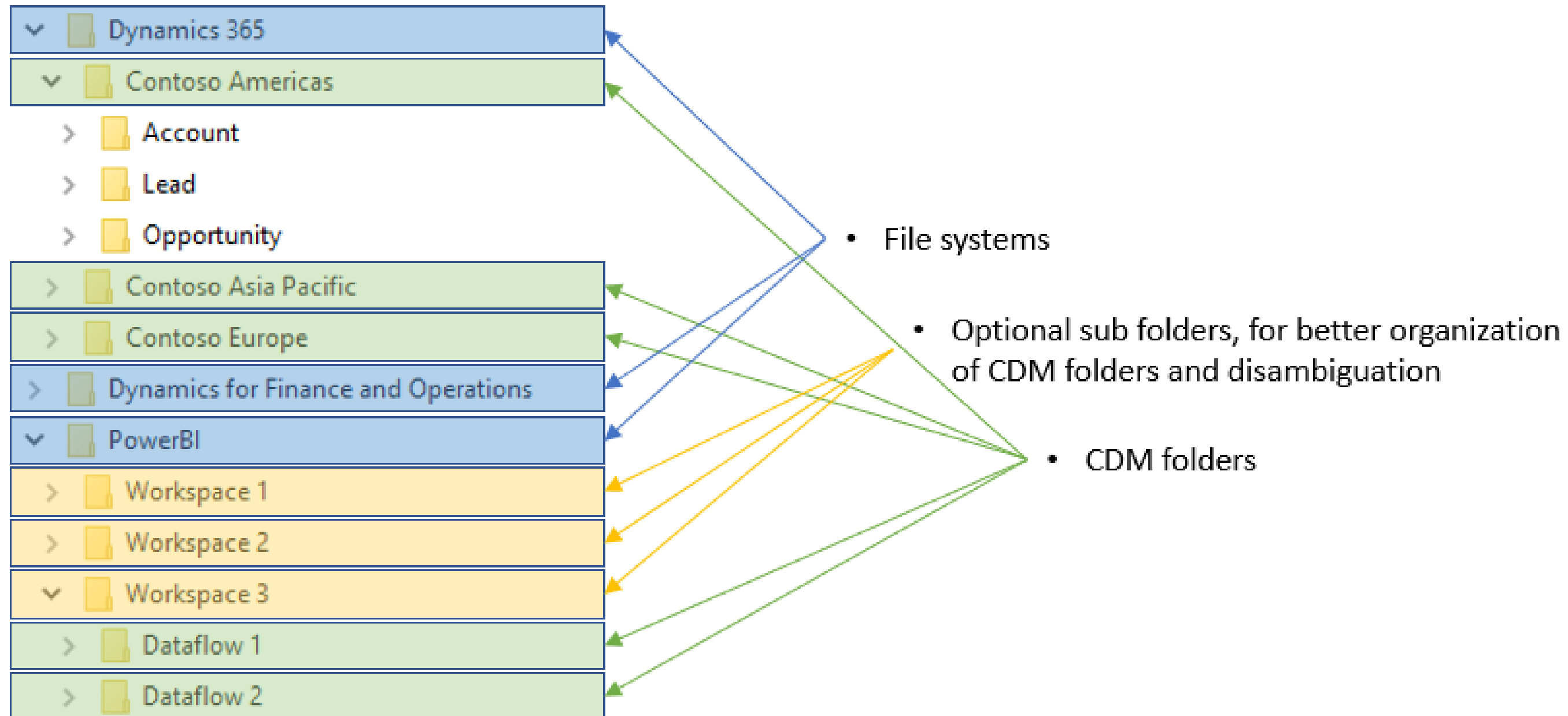
Low to high code



Common Data Model



Common Data Model





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 dataflows it is dependent on

Incremental Refresh is Premium Only (for now?)



How?

Query Reusage

Query Reusability

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 workspace

Logic is reused, and data is duplicated every time the query is referenced

Query Reusability

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



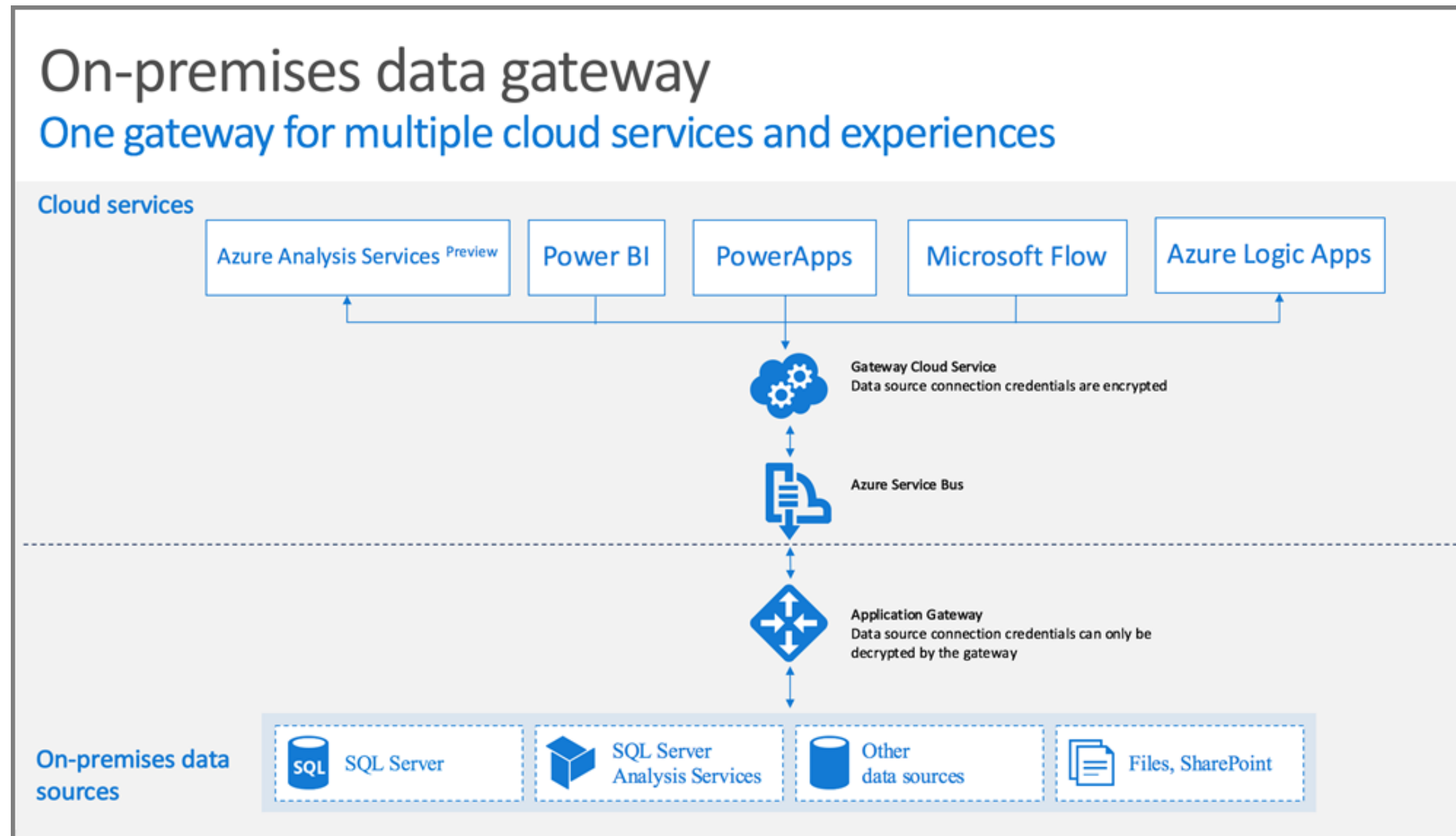
Query Reusability

Demo



Gateways & 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!



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

Bring your own Storage Account



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

<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

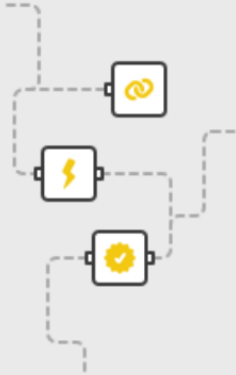
- Output JSON

- 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

The screenshot shows a modal dialog box titled "Attach an external CDM folder to a new dataflow". The dialog has three main input sections: "Name" with a text box containing "Sales Forecasting", "Description" with a text area containing "Dataflow entities with output from Azure Machine Learning models for detailed sales forecasts.", and "CDM folder path" with a text box containing the placeholder text "Enter the path to your CDM folder in Azure Data Lake Storage Gen2". At the bottom of the dialog are two buttons: "Create and attach" and "Cancel". In the background, parts of other UI elements are visible, including a "Define" panel on the left and a "External" panel on the right, both with yellow buttons labeled "Add new entities", "Add linked entities", and "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

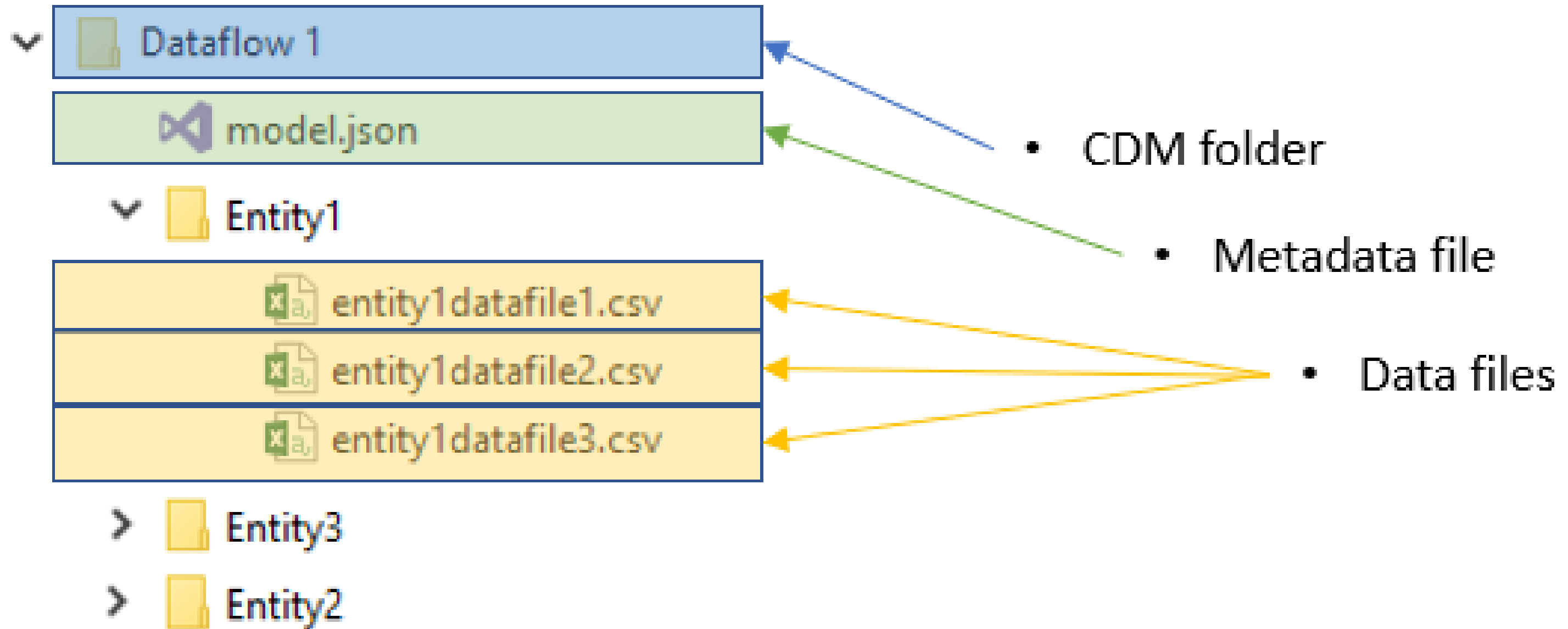
Choose a CDM folder to attach to the dataflow. You can define the dataflow entities to be added to the dataflow, or you can define the CDM folder path instead. [Data Model](#) [Learn more](#)

External

Choose a CDM folder to attach to the dataflow. You can define the dataflow entities to be added to the dataflow, or you can define the CDM folder path instead. [Data Model](#) [Learn more](#)

Add new entities Add linked entities Create and attach

External CDM Folders





What's the catch?

Considerations & Limitations

The cool stuff is Premium / PPU only

Validation of queries can take a looooooonng 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



Takeaways

Takeaways

Business Driven Initiatives

Fill existing gaps

Quick prototyping

Strive for IT handover to operationalize later on

Configure your own ADLSg2 Storage

Power BI dataflows – April Update

Coming Soon

Careful reading might have noticed a few teasers. Indeed, we have some exciting features that will be available in the coming months ahead. Currently, we are laser focused on **performance and reliability improvements**, and **delivering on YOUR top asks, including**:

- **Background Validation**
 - With this update, users will be able to Save a dataflow, leave the screen, and continue working. Validation now takes place in the background so users can save the dataflow without friction.
- **Multiple Contributors to Dataflows**
 - We're working to allow non-owners to edit Dataflows, see the applied steps, and manage them – making Dataflows more collaborative so that teams can be productive.
- **Notifications to Non-Owners**
 - Sending refresh failure emails to users or groups, just like Datasets
- **Deployment Pipelines**
 - We'll bring support for Deployment Pipelines to Dataflows, allowing you to deploy and update Dataflows across the lifecycle stages and make them more enterprise grade.
- **Updated Guidance**
 - Advice for common scenarios and prescriptive how-to's

<https://powerbi.microsoft.com/en-us/blog/whats-new-in-dataflows-april-2021/>

The Self-Service BI Hoax

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

[Nimrod Avissar on LinkedIn Pulse](#)

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-bi-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](#)

Feedback

Your feedback is important to us.

Don't forget to rate and review the sessions.