PowerPlatform Dataflows
PowerQuery (and more) in the Cloud

Wolfgang Strasser

Microsoft Most Valuable Professional

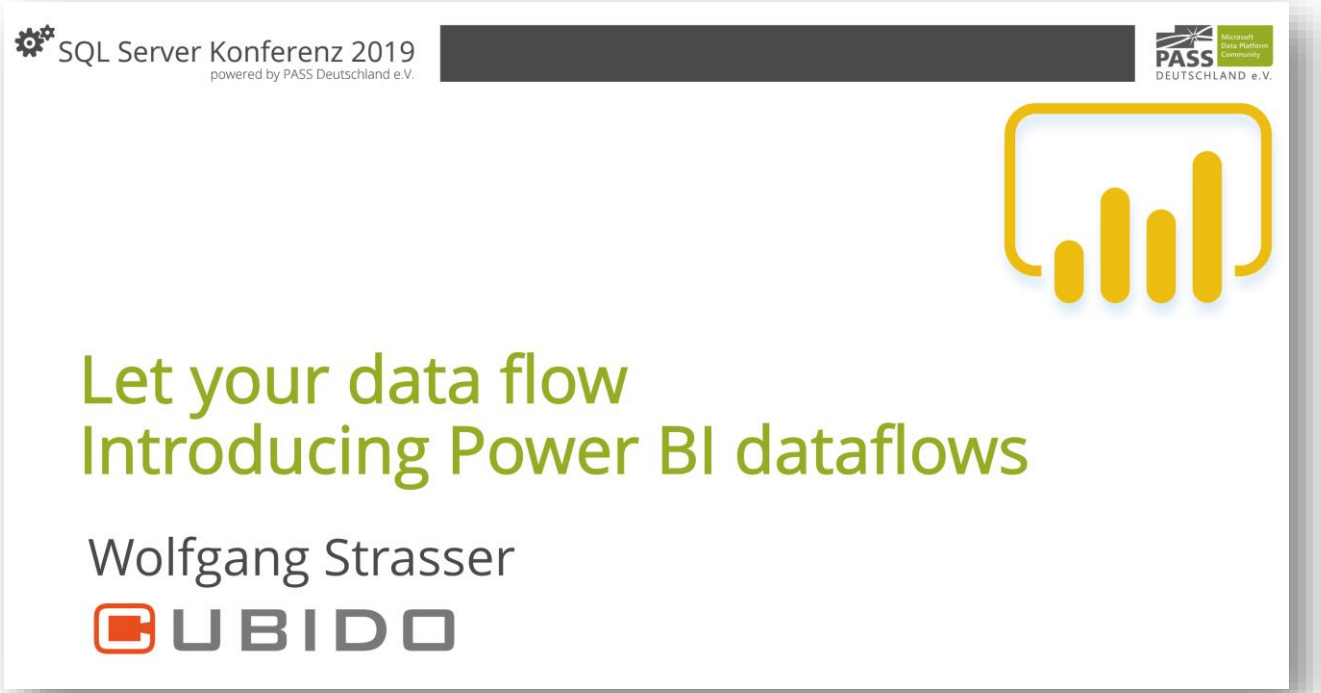
UBIDO
AN ACP GROUP COMPANY

IT for innovators.
Analytics & Development

Abstract

With PowerPlatform dataflows, users can bring PowerQuery functionality to a cloud-only life. Created in the cloud, processed in the cloud and storing results in the cloud, dataflows can be seen as cloud native part of the ETL within the PowerPlatform.
After their initial start in the Power BI universe, they now reached out into other parts of the PowerPlatform and be one of the future options of ETL pipelines to transfer data into data lakes.
In addition to data transformation, dataflows can be enriched with AI functionality to bring more ML power into your ETL processes.
Join this session if you want to learn more about the current feature set available and see what is new and possible with data preparation in the cloud with PowerPlatform dataflows.

One year ago… Power BI dataflows



What are (Power BI) dataflows?

Information Hierarchy

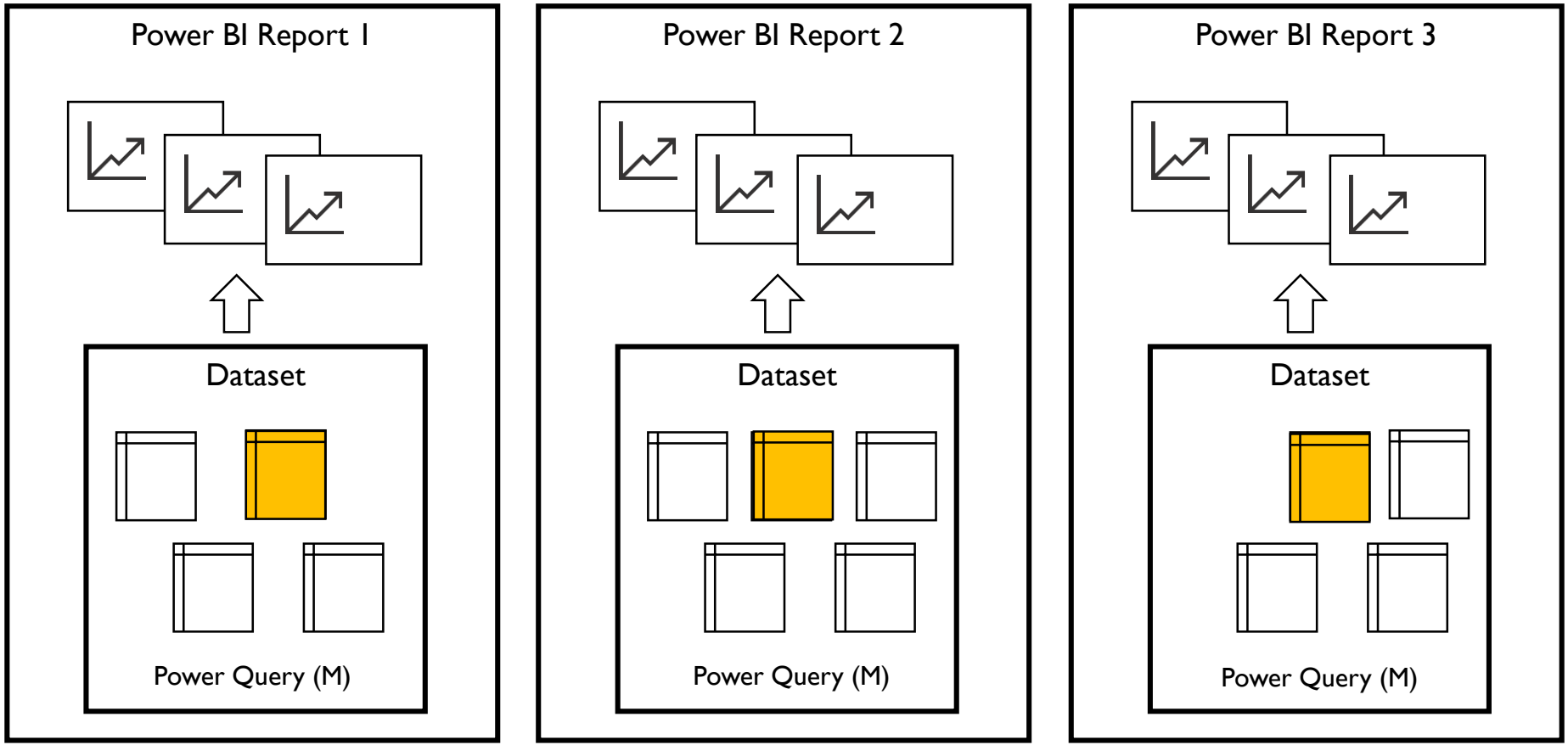
Reusability

Common Data Model (CDM)

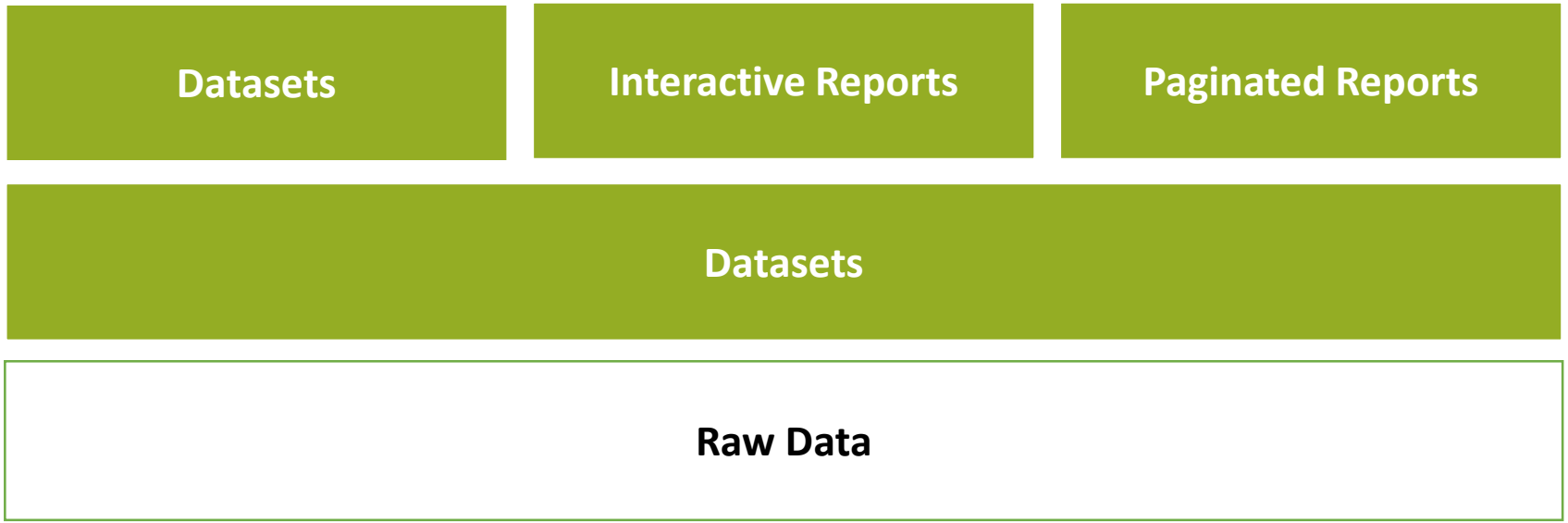
PowerQuery (Online)

Data Lake

ETL in the Context of Power BI

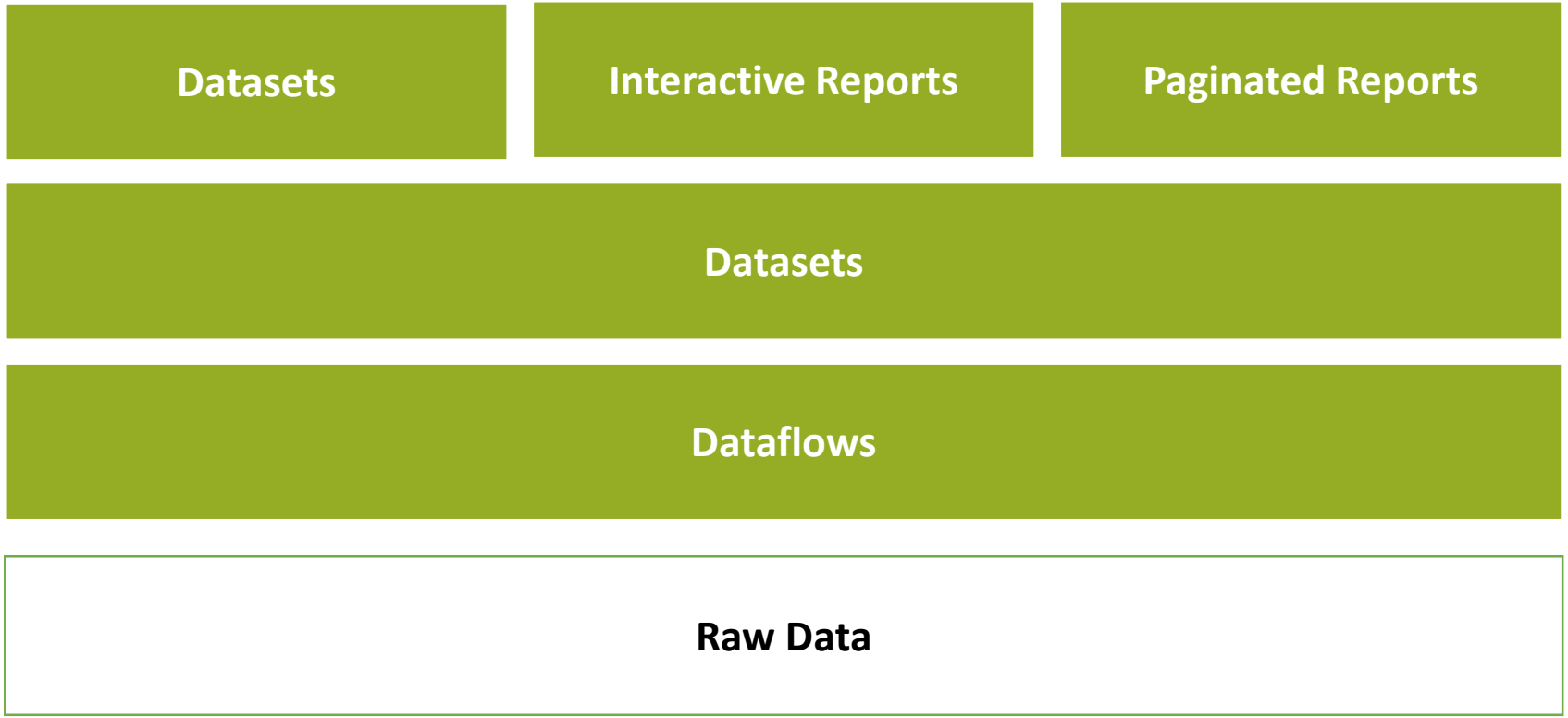


Power BI Information Hierarchy (before)



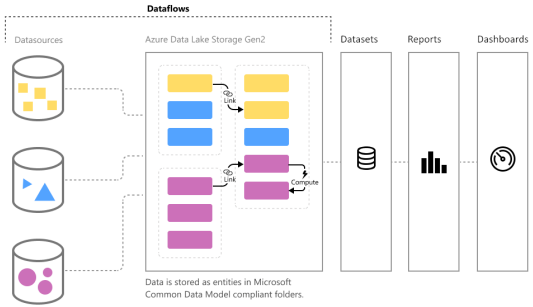
Source: Amir Netz: Power BI dataflows Whitepaper (https://go.microsoft.com/fwlink/?linkid=2034388&clcid=0x409)

Power BI Information Hierarchy (before)



Source: Amir Netz: Power BI dataflows Whitepaper (https://go.microsoft.com/fwlink/?linkid=2034388&clcid=0x409)

Fact Sheet – PBI dataflows



PowerQuery processes in the cloud

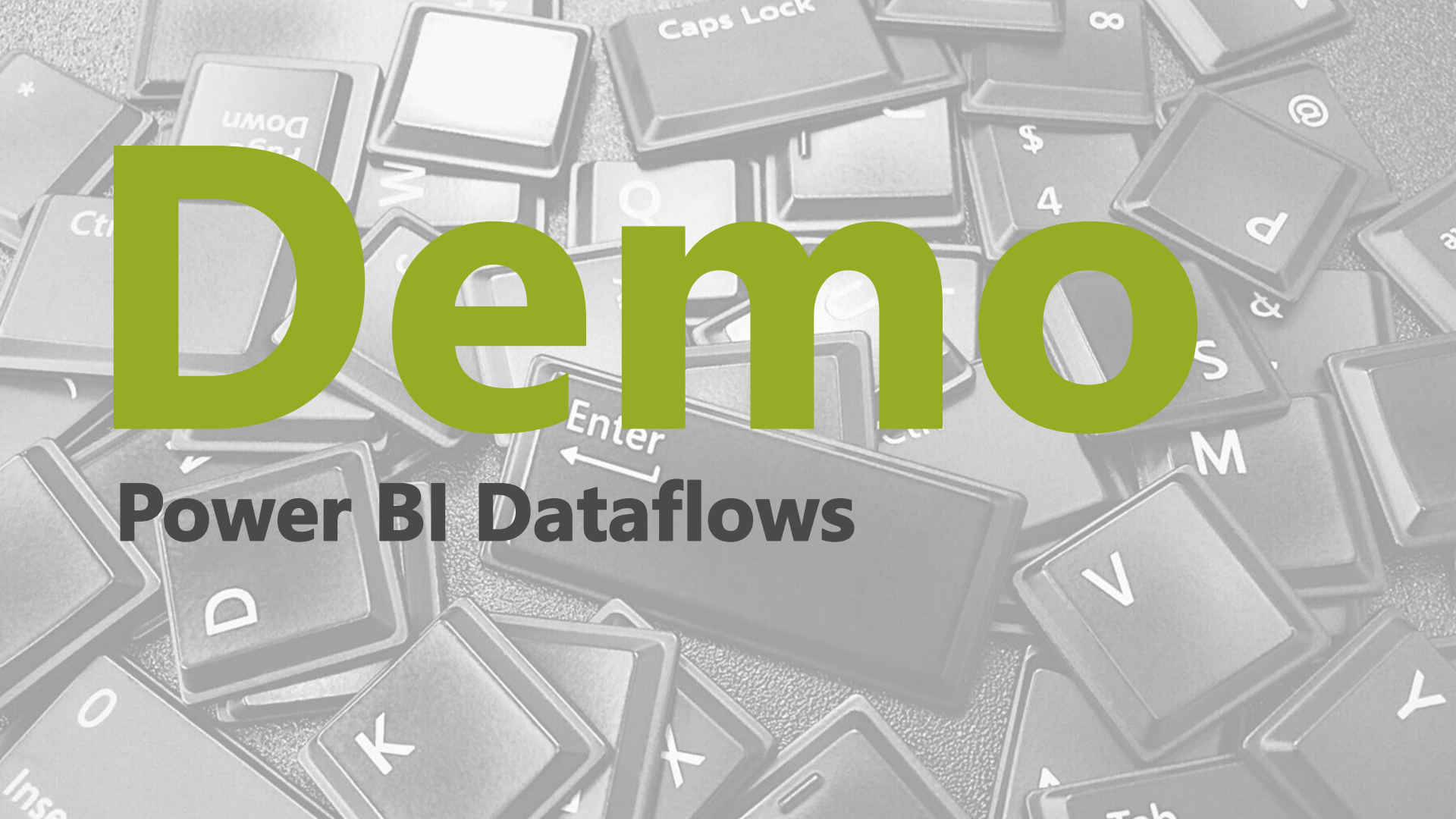
.. run independently of a Power BI report

.. store output into a data lake (Azure Data Lake Storage Gen2)

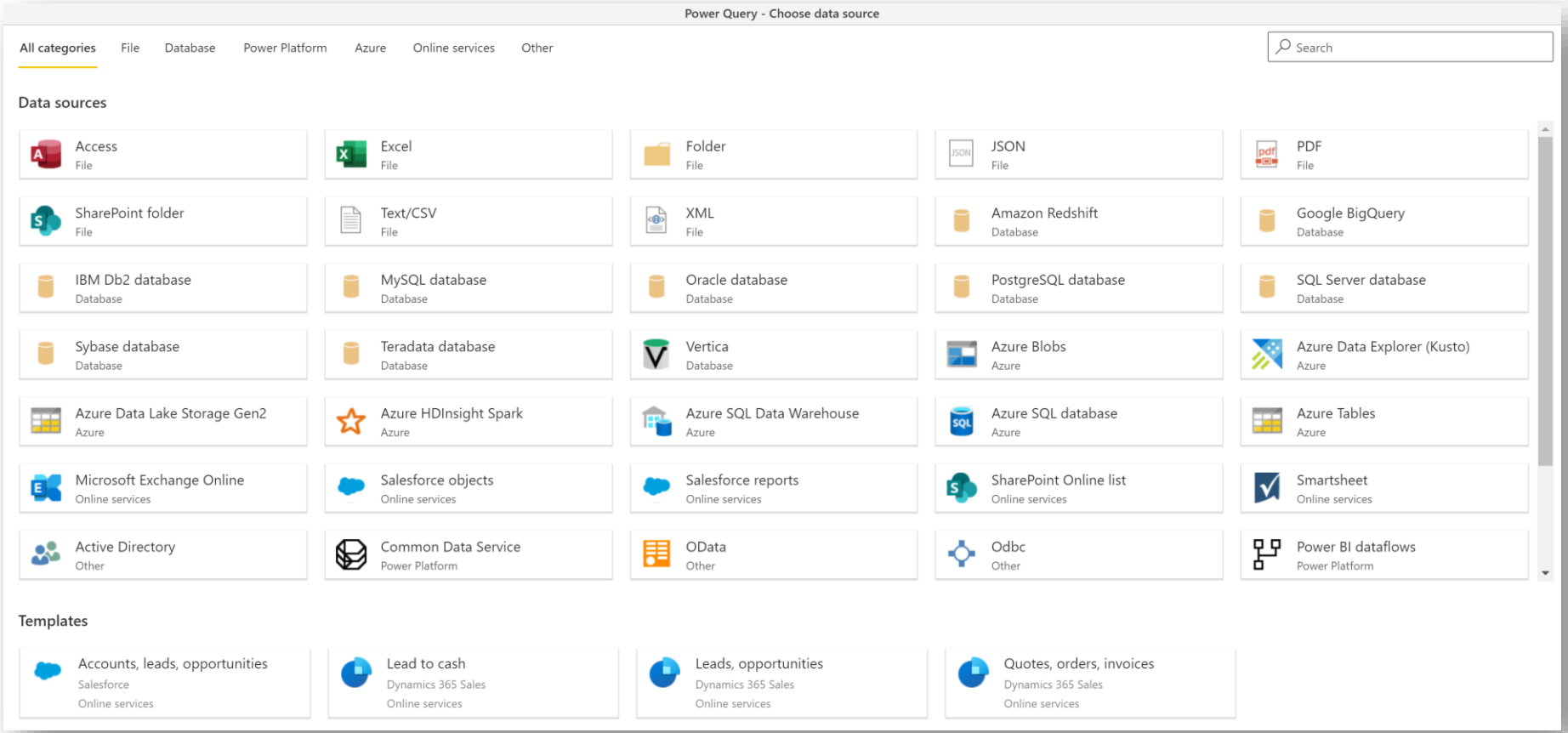
... use the Common Data Model for entity definition

... entities (result tables)

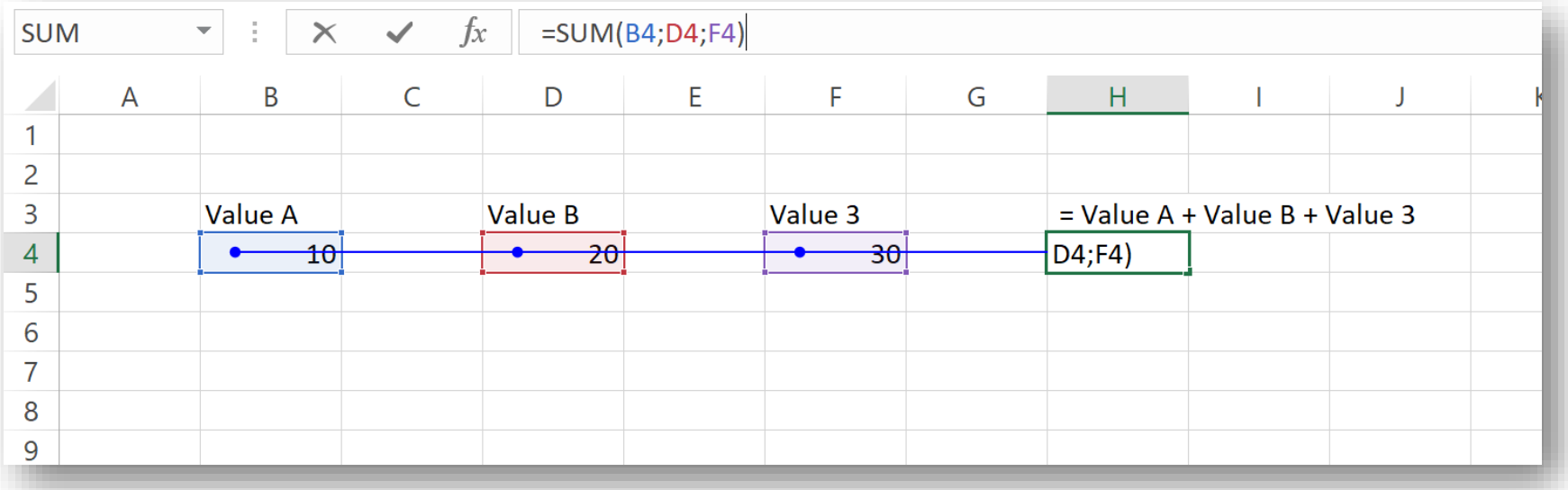
... not in “my workspace”

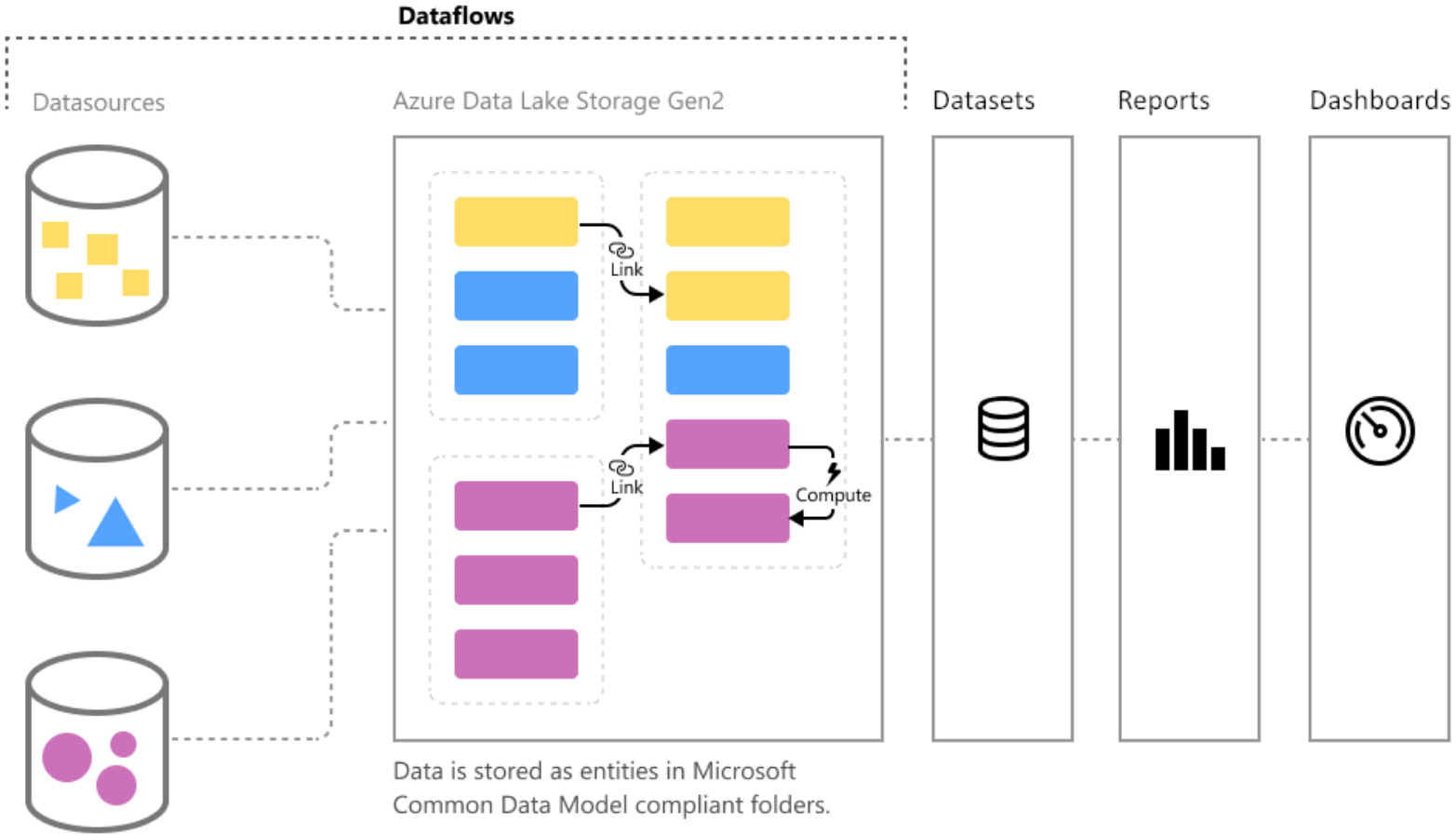

Demo
Power BI Dataflows

Power Query Online



Dataflows are like Excel





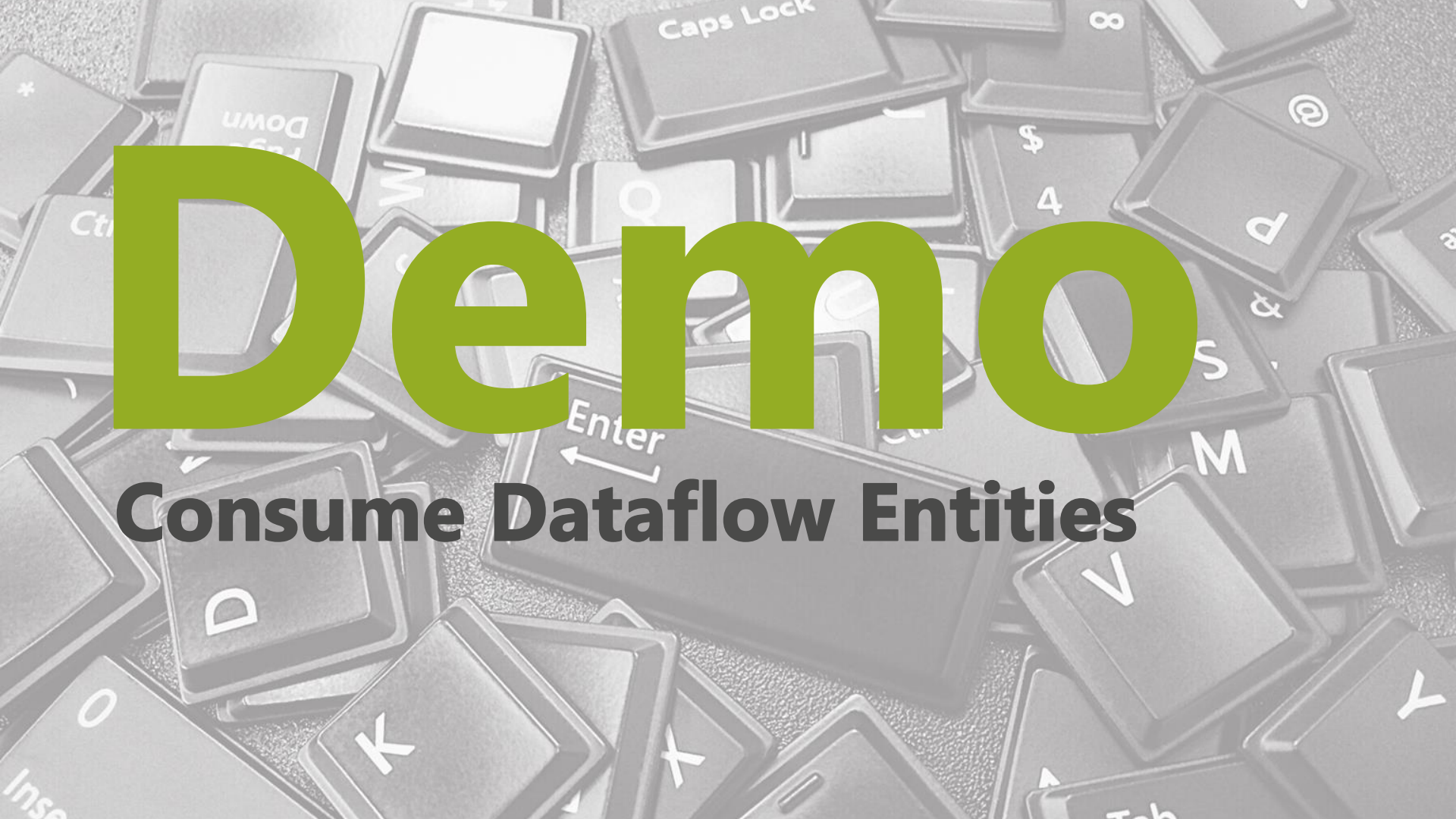
https://docs.microsoft.com/en-us/power-bi/service-dataflows-overview

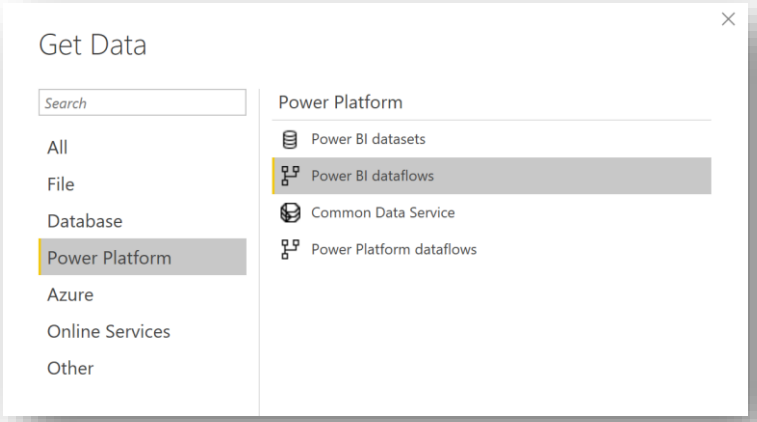
Common Data Model

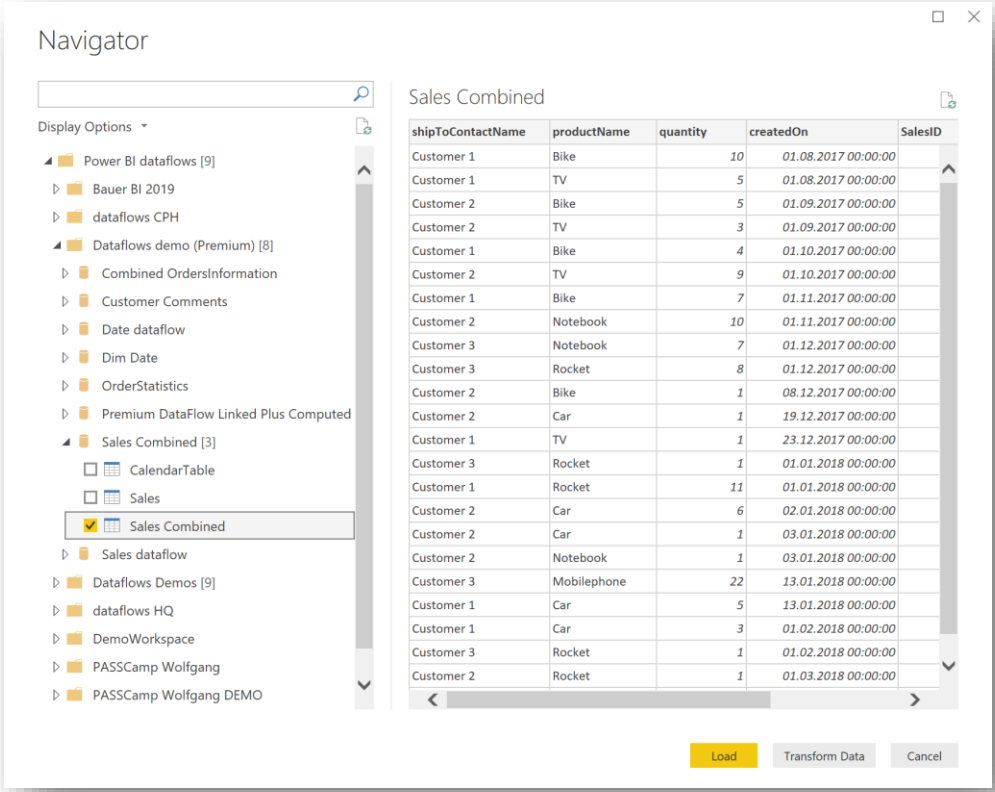
What's in Common Data Model?

In addition to the metadata system, Common Data Model includes a set of standardized, extensible data schemas that Microsoft and its partners have published. This collection of predefined schemas includes entities, attributes, semantic metadata, and relationships. The schemas represent commonly used concepts and activities, such as **Account** and **Campaign**, to simplify the creation, aggregation, and analysis of data. This graphic shows some elements of the standard entities. More information: [Common Data Model repository on GitHub]

- Set of standardized, extensible data schemas
- Entities, attributes, semantic metadata & relationships

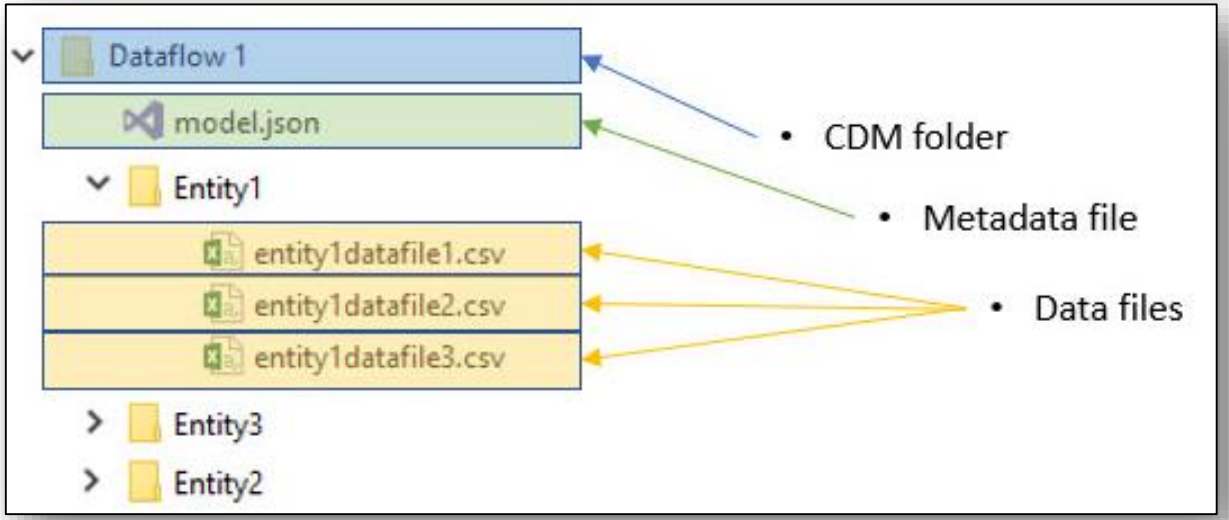

Demo
Consume Dataflow Entities





shipToContactName	productName	quantity	createdOn	SalesID
Customer 1	Bike	10	01.08.2017 00:00:00	
Customer 1	TV	5	01.08.2017 00:00:00	
Customer 2	Bike	5	01.09.2017 00:00:00	
Customer 2	TV	3	01.09.2017 00:00:00	
Customer 1	Bike	4	01.10.2017 00:00:00	
Customer 2	TV	9	01.10.2017 00:00:00	
Customer 1	Bike	7	01.11.2017 00:00:00	
Customer 2	Notebook	10	01.11.2017 00:00:00	
Customer 3	Notebook	7	01.12.2017 00:00:00	
Customer 3	Rocket	8	01.12.2017 00:00:00	
Customer 2	Bike	1	08.12.2017 00:00:00	
Customer 2	Car	1	19.12.2017 00:00:00	
Customer 1	TV	1	23.12.2017 00:00:00	
Customer 3	Rocket	1	01.01.2018 00:00:00	
Customer 1	Rocket	11	01.01.2018 00:00:00	
Customer 2	Car	6	02.01.2018 00:00:00	
Customer 2	Car	1	03.01.2018 00:00:00	
Customer 2	Notebook	1	03.01.2018 00:00:00	
Customer 3	Mobilephone	22	13.01.2018 00:00:00	
Customer 1	Car	5	13.01.2018 00:00:00	
Customer 1	Car	3	01.02.2018 00:00:00	
Customer 3	Rocket	1	01.02.2018 00:00:00	
Customer 2	Rocket	1	01.03.2018 00:00:00	

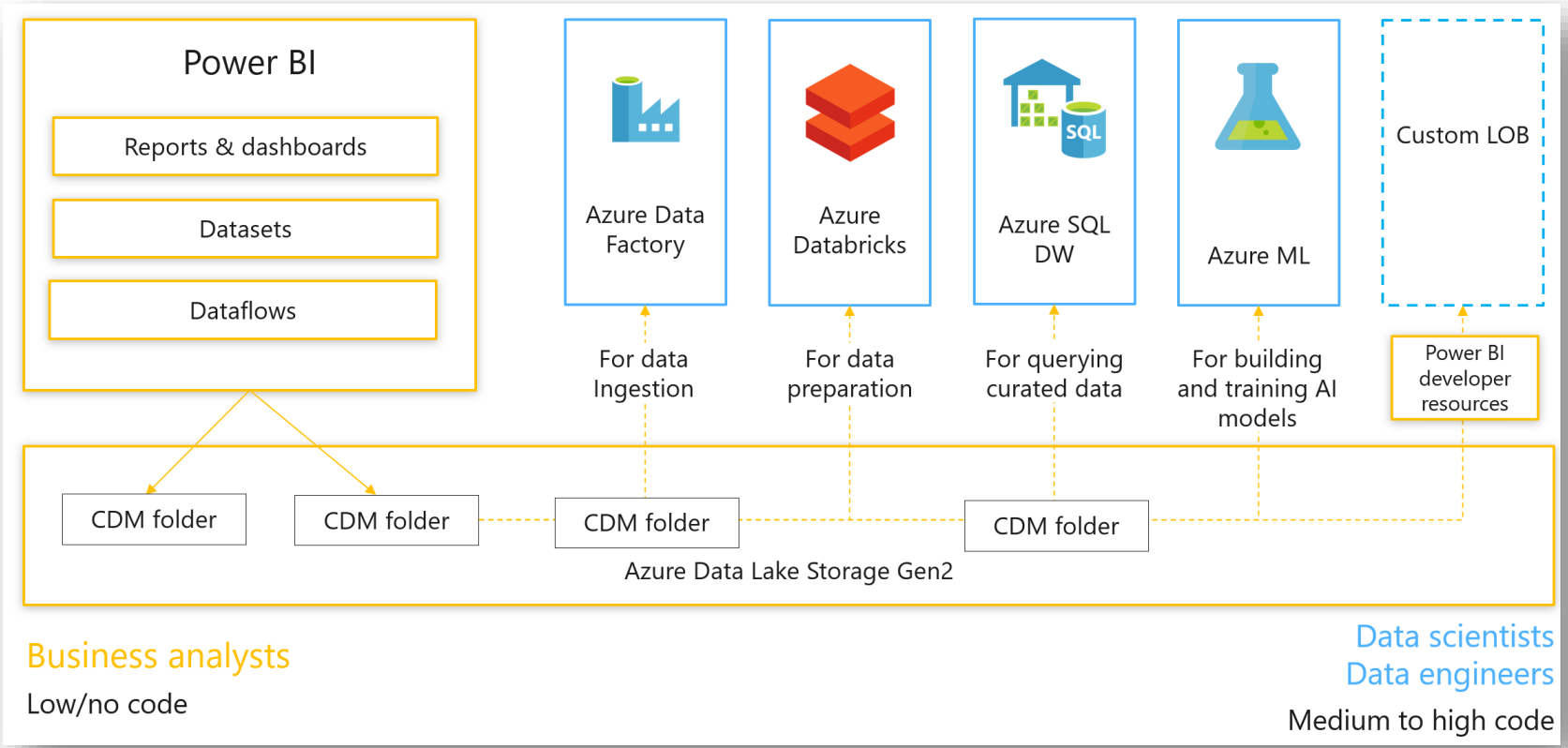
CDM Folders



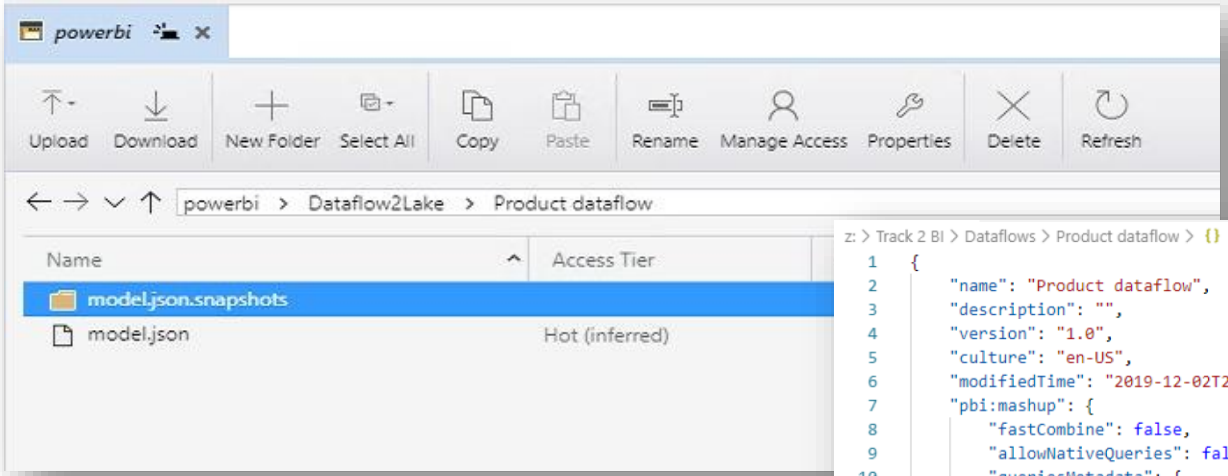
Sales From Excel.json ● First Dataflow.json ●

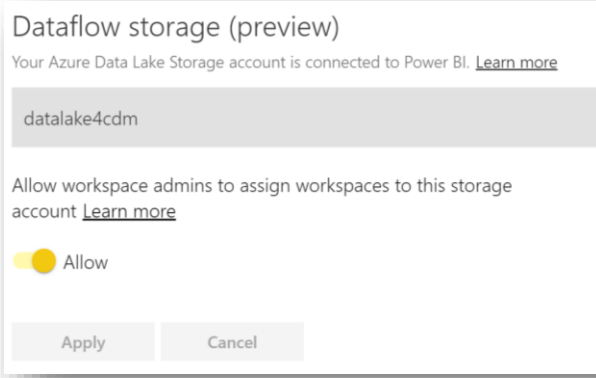
```
{
    "name": "Sales From Excel",
    "description": "",
    "version": "1.0",
    "culture": "en-US",
    "pbi:mashup": {
        "fastCombine": true,
        "allowNativeQueries": false,
        "queriesMetadata": {
            "Sales": {
                "queryId": "40b69972-542b-49ee-ab09-bff5
                "queryName": "Sales",
                "loadEnabled": true
            }
        },
        "document": "section Section1;\r\nshared Sales =
    },
    "entities": [
        {
            "$type": "LocalEntity",
            "name": "Sales",
            "description": "",
            "pbi:refreshPolicy": {
                "$type": "FullRefreshPolicy",
                "location": "Sales.csv"
            },
            "attributes": [
                {
                    "name": "SalesOrderNumber",
                    "dataType": "string"
                },
```

Power BI Dataflows and the Data Lake



Bring your Own Storage Account





z: > Track 2 BI > Dataflows > Product dataflow > {} model.json > [ ] entities > {} 0 > [ ] attributes > {} 1

```
{
    "name": "Product dataflow",
    "description": "",
    "version": "1.0",
    "culture": "en-US",
    "modifiedTime": "2019-12-02T20:34:28.5677753+00:00",
    "pbi:mashup": {
        "fastCombine": false,
        "allowNativeQueries": false,
        "queriesMetadata": {
            "Products": {
                "queryId": "eed1d3a0-543f-48fd-bd2a-e5938d922b08",
                "queryName": "Products",
                "loadEnabled": true
            }
        },
        "document": "section Section1;\r\nshared Products = let\r\n  Source = Sql.Database(\"passcam
    },
    "entities": [
        {
            "$type": "LocalEntity",
            "name": "Products",
            "description": "",
            "pbi:refreshPolicy": {
                "$type": "FullRefreshPolicy",
                "location": "Products.csv"
```


Demo
BYOSA

Power Platform dataflows

Self-Service
Data Preparation

"Data Analysts spend up to 80% of their time on data preparation delaying the time to analysis and decision making."
- Gartner

"Three-quarters of companies are not able to act on the majority of data they collect, owing in large part to disjointed systems and data integration issues."
Harvard Business Review, Closing the Customer Experience Gap

"Gartner predicts that, by 2020, due in large part to the automation of data science tasks, citizen data scientists will surpass data scientists in the amount of advanced analysis produced."
Gartner, Hype Cycle for the Digital Workplace, 2018, 18 July 2018

How can Dataflows help?

Connectors

Wide range of supported data sources of all sizes & shapes, wherever it comes from (on-premises or cloud sources)

Interactive & intuitive

Build queries over any data source and any size

Consistency of experience

PowerQuery everywhere ☺

Semantic understanding and data consistency

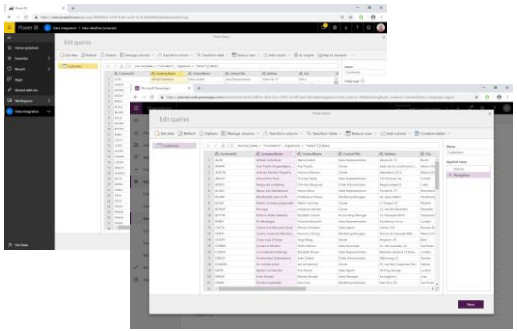
Map to standard (or extend/create new) CDM entities

Dataflows - Overview

Authoring tools

Power BI

Power Apps



Transform data
PowerQuery Online

Common Data Service

Automatic replication
("Athena")

ADLS (gen2)

Enrich data with
AI in PQ

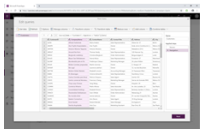
Data sources

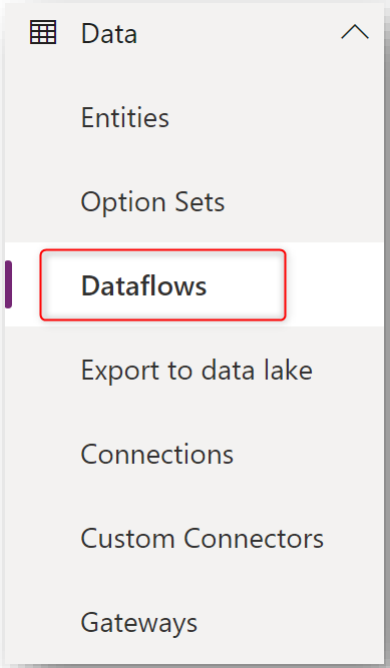
Data Transformation, Preparation Enrichment

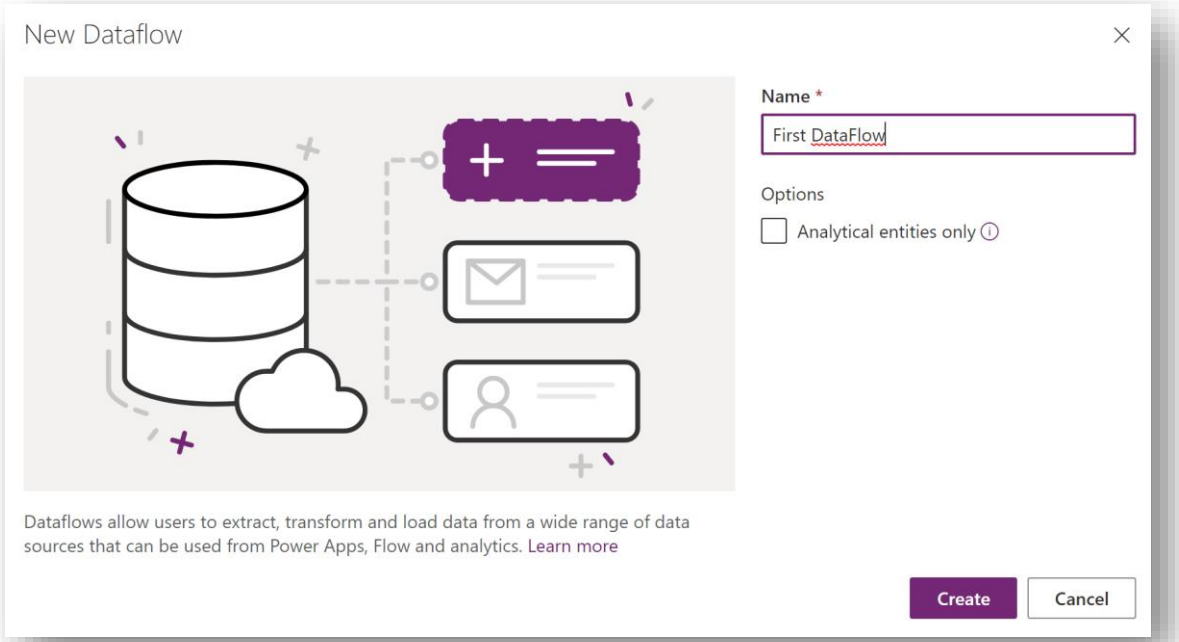
Data Storage

Data consumers

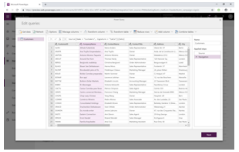
Load Data into CDS

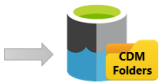






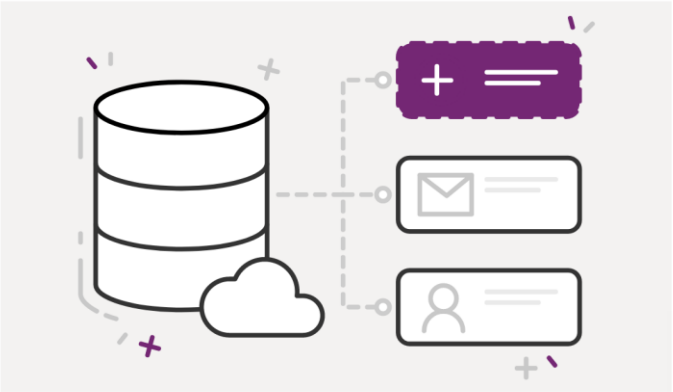
“Normal” plus Analytic entities





ADLS (gen2)

New Dataflow



Dataflows allow users to extract, transform and load data from a wide range of data sources that can be used from Power Apps, Flow and analytics. Learn more

Name *

First DataFlow

Options

Analytical entities only

Queries

Products

Load settings

Load to new entity

Load to existing entity

Do not load

Destination entity

cdmtmp_Product

Entity display name

Product

Entity description

An item that is available for sale.

Delete rows that no longer exist in the query output

Field mapping

Key fields

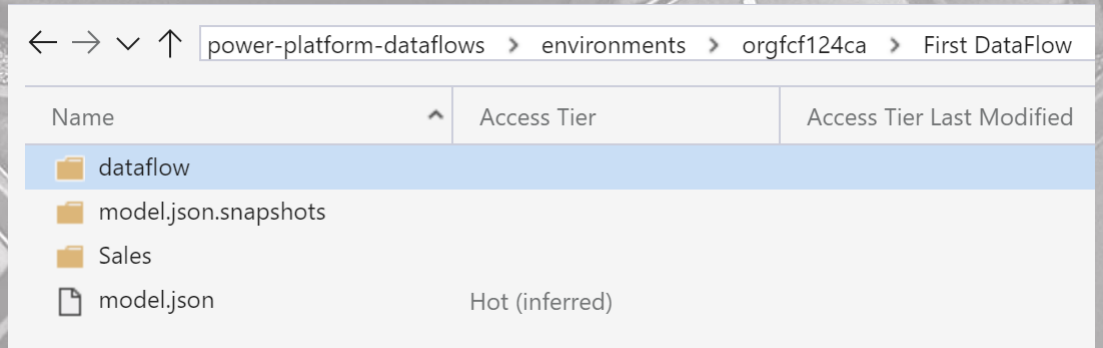
cdmtmp_ProductNumber

Auto map

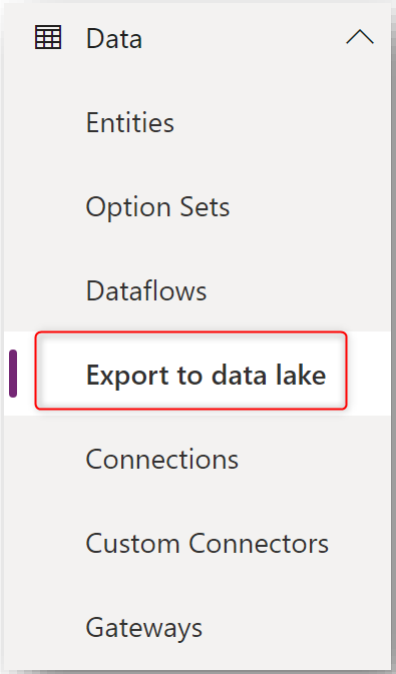
(none)	cdmtmp_DefaultBuyingUnitOfMeasure
(none)	cdmtmp_DefaultSellingQuantity
(none)	cdmtmp_DefaultSellingQuantity_UOM
(none)	cdmtmp_DefaultStockingUnitOfMeasure
(none)	cdmtmp_Description
(none)	cdmtmp_IsStocked
Name	cdmtmp_Name
(none)	cdmtmp_OrganizationId.cdmtmp_OrganizationNumber
(none)	cdmtmp_ProductCategoryId.cdmtmp_CategoryId
ArticleID	cdmtmp_ProductNumber
Category	cdmtmp_ProductType
(none)	cdmtmp_SellingUnitPrice
(none)	cdmtmp_SellingUnitPrice_CurrencyCode
(none)	cdmtmp_StandardCostAmount
(none)	cdmtmp_StandardCostAmount_CurrencyCode
(none)	cdmtmp_Status
(none)	cdmtmp_UnitOfMeasureScale
(none)	EntityImage

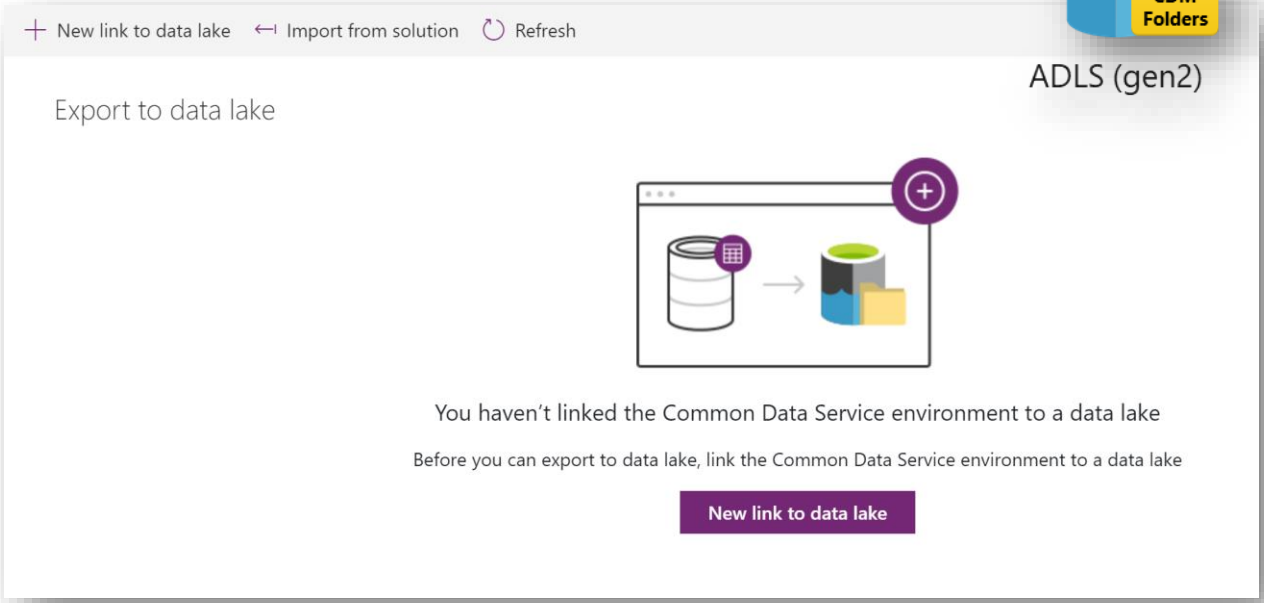
Demo

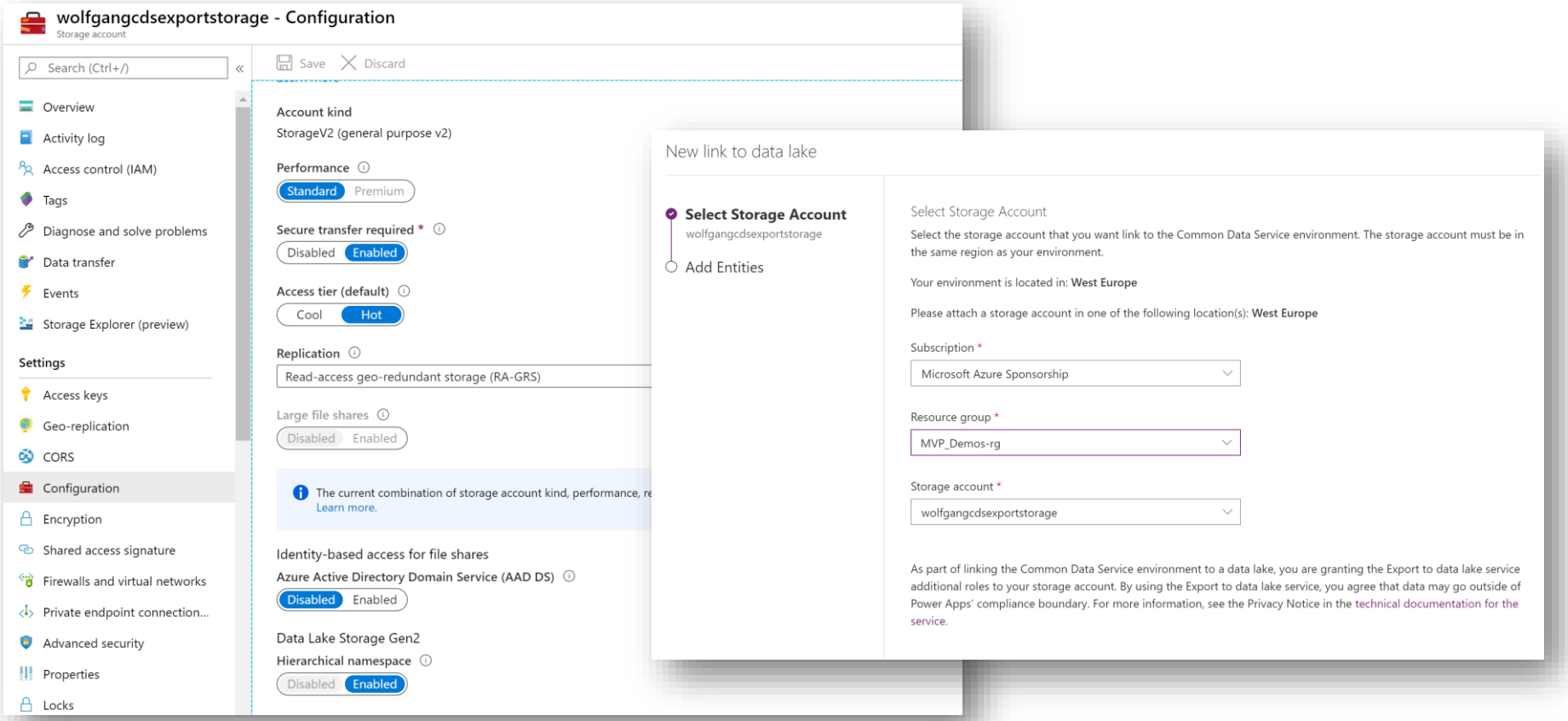
What's in the Data Lake?

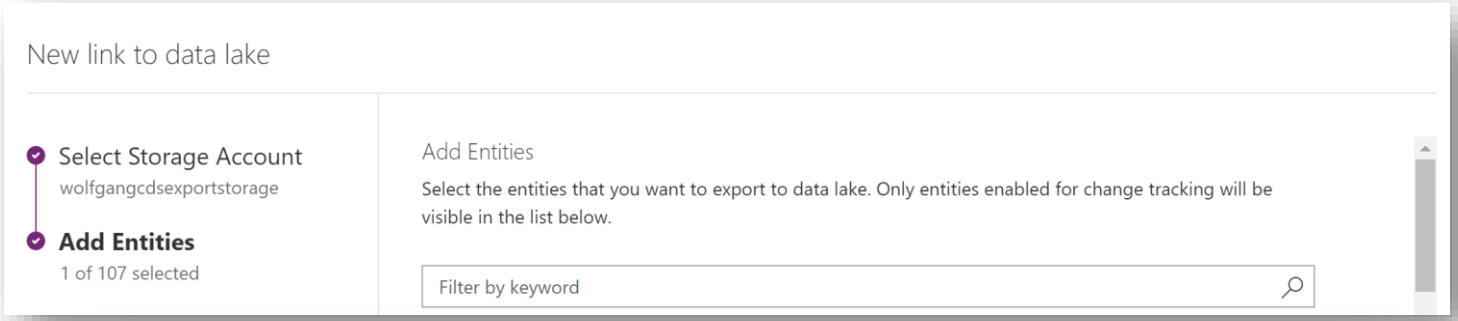


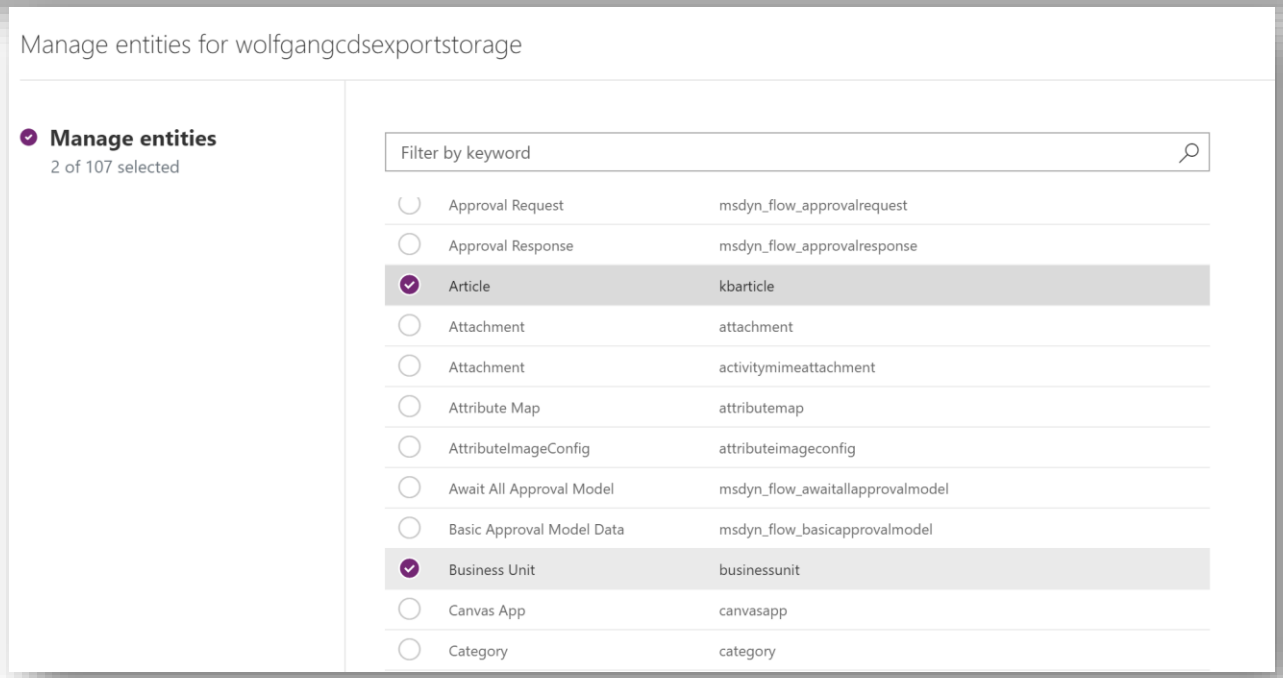
Export to Data Lake – Project “Athena”





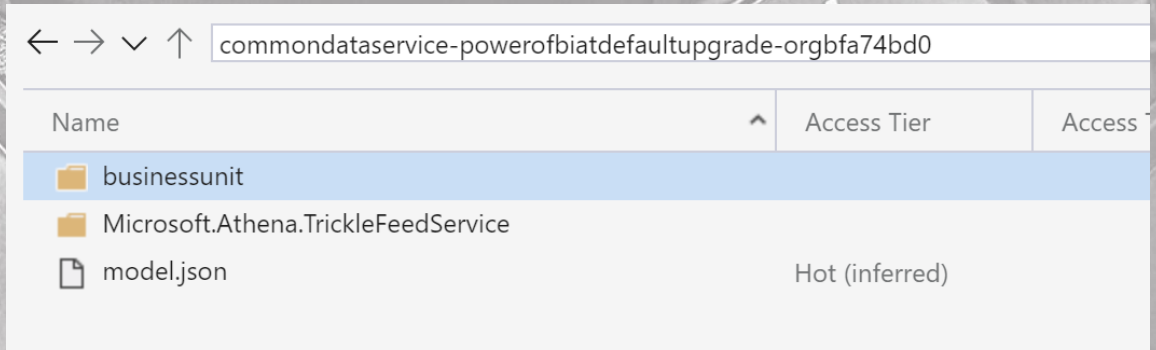

wolfgangcdsexportstorage - Configuration
Storage account
Search (Ctrl+/)
Save
Discard
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Data transfer
Events
Storage Explorer (preview)
Settings
Access keys
Geo-replication
CORS
Configuration
Encryption
Shared access signature
Firewalls and virtual networks
Private endpoint connection...
Advanced security
Properties
Locks
Account kind
StorageV2 (general purpose v2)
Performance
Standard
Premium
Secure transfer required *
Disabled
Enabled
Access tier (default)
Cool
Hot
Replication
Read-access geo-redundant storage (RA-GRS)
Large file shares
Disabled
Enabled
The current combination of storage account kind, performance, re
Learn more.
Identity-based access for file shares
Azure Active Directory Domain Service (AAD DS)
Disabled
Enabled
Data Lake Storage Gen2
Hierarchical namespace
Disabled
Enabled
New link to data lake
Select Storage Account
wolfgangcdsexportstorage
Add Entities
Select Storage Account
Select the storage account that you want link to the Common Data Service environment. The storage account must be in the same region as your environment.
Your environment is located in: West Europe
Please attach a storage account in one of the following location(s): West Europe
Subscription *
Microsoft Azure Sponsorship
Resource group *
MVP_Demos-rg
Storage account *
wolfgangcdsexportstorage
As part of linking the Common Data Service environment to a data lake, you are granting the Export to data lake service additional roles to your storage account. By using the Export to data lake service, you agree that data may go outside of Power Apps' compliance boundary. For more information, see the Privacy Notice in the technical documentation for the service.


New link to data lake
Select Storage Account
wolfgangcdsexportstorage
Add Entities
1 of 107 selected
Add Entities
Select the entities that you want to export to data lake. Only entities enabled for change tracking will be visible in the list below.
Filter by keyword

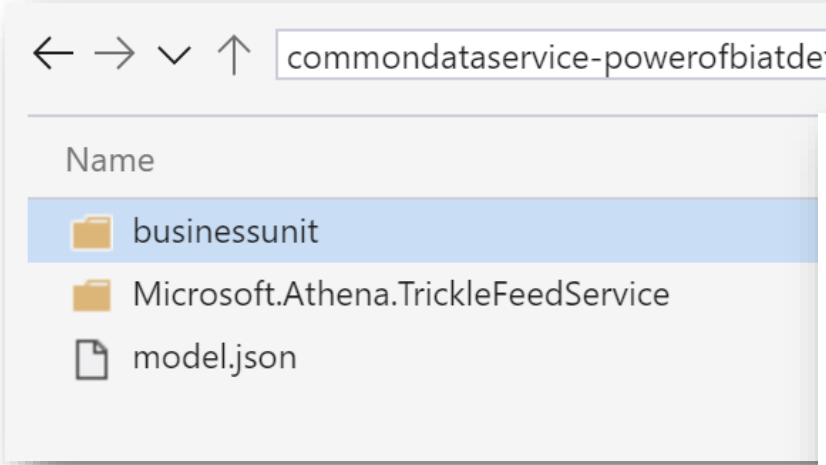

Manage entities for wolfgangcdsexportstorage
Manage entities
2 of 107 selected
Filter by keyword
Approval Request msdyn_flow_approvalrequest
Approval Response msdyn_flow_approvalresponse
Article kbarticle
Attachment attachment
Attachment activitymimeattachment
Attribute Map attributemap
AttributeImageConfig attributeimageconfig
Await All Approval Model msdyn_flow_awaitallapprovalmodel
Basic Approval Model Data msdyn_flow_basicapprovalmodel
Business Unit businessunit
Canvas App canvasapp
Category category

Demo

What's in the Data Lake?



What's in my Data Lake?

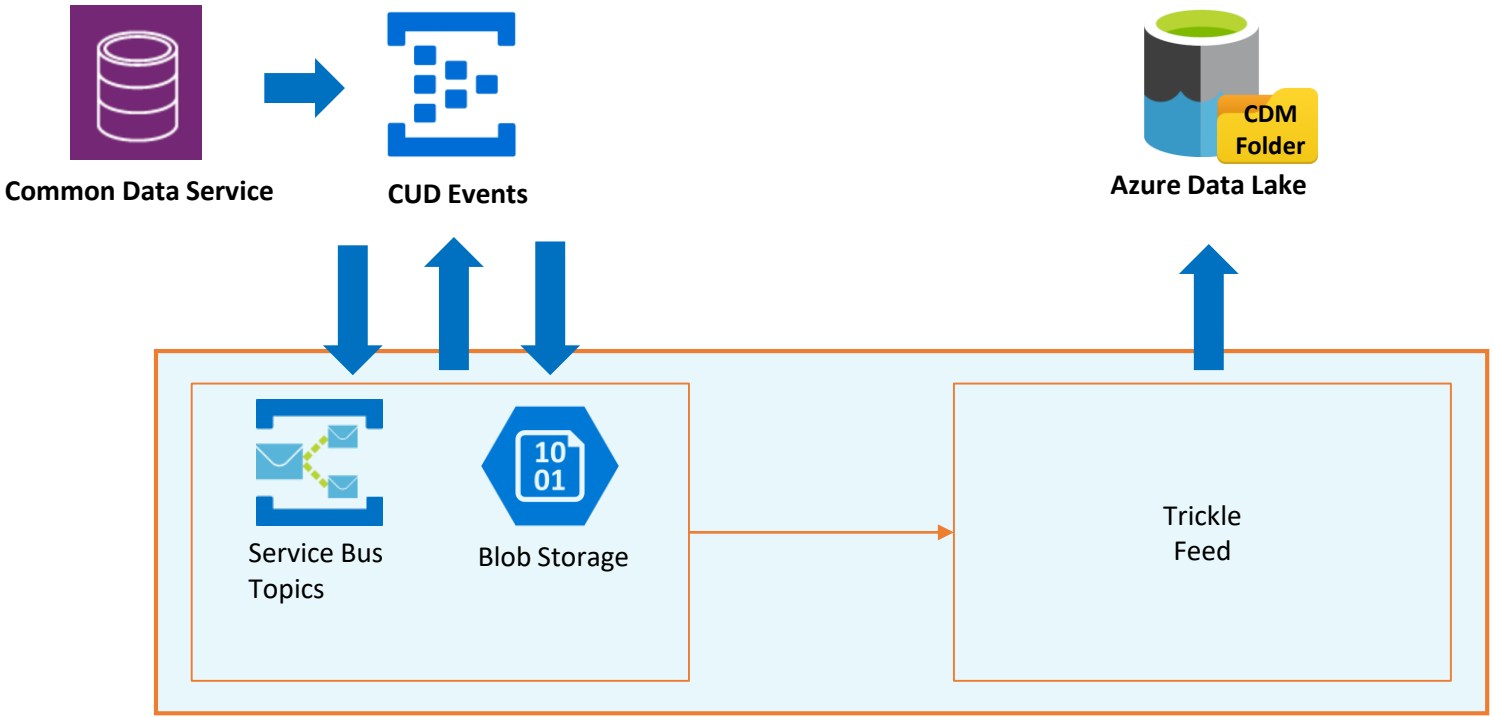



```
{
    "name": "cdm",
    "description": "cdm",
    "version": "1.0",
    "entities": [
        {
            "$type": "LocalEntity",
            "name": "businessunit",
            "description": "businessunit",
            "annotations": [
                {
                    "name": "Athena:PartitionGranularity",
                    "value": "Year"
                },
                {
                    "name": "Athena:InitialSyncState",
                    "value": "Completed"
                },
                {
                    "name": "Athena:InitialSyncDataCompletedTime",
                    "value": "3/1/2020 5:58:33 PM"
                }
```

Behind the Scenes

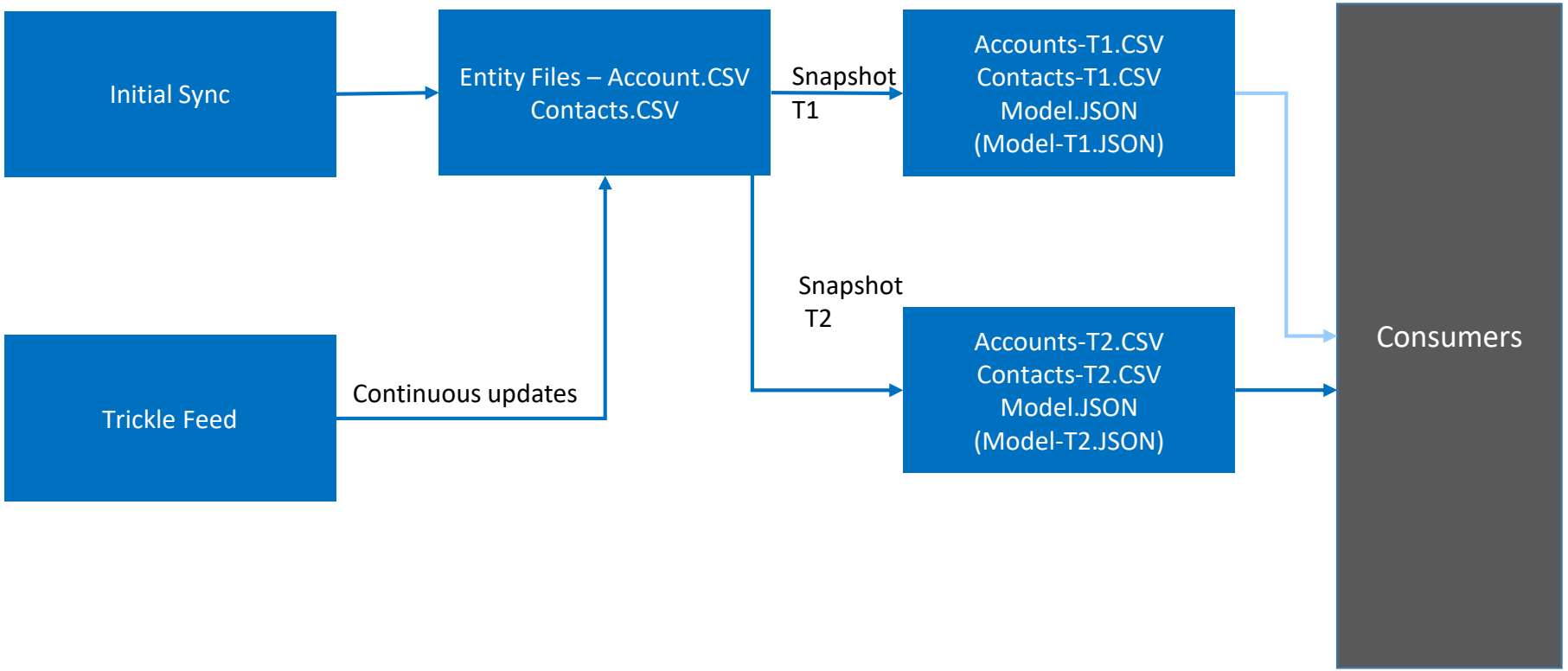
Functionality

- Support for **full** and **incremental**
- **Trickle feed** to write data to ADLS Gen 2

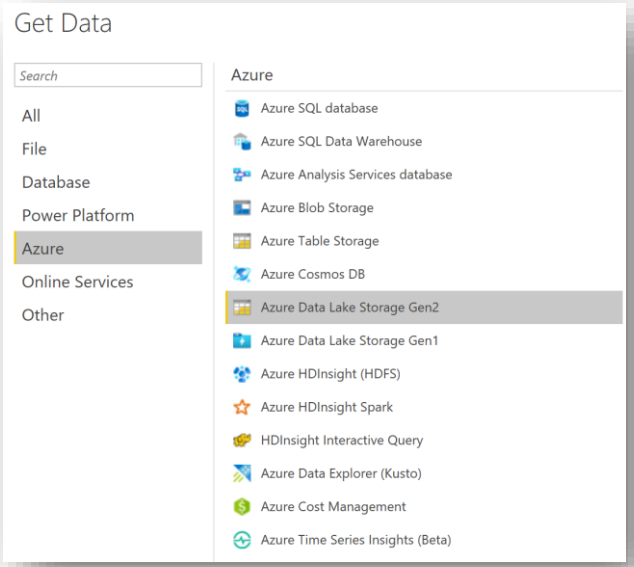


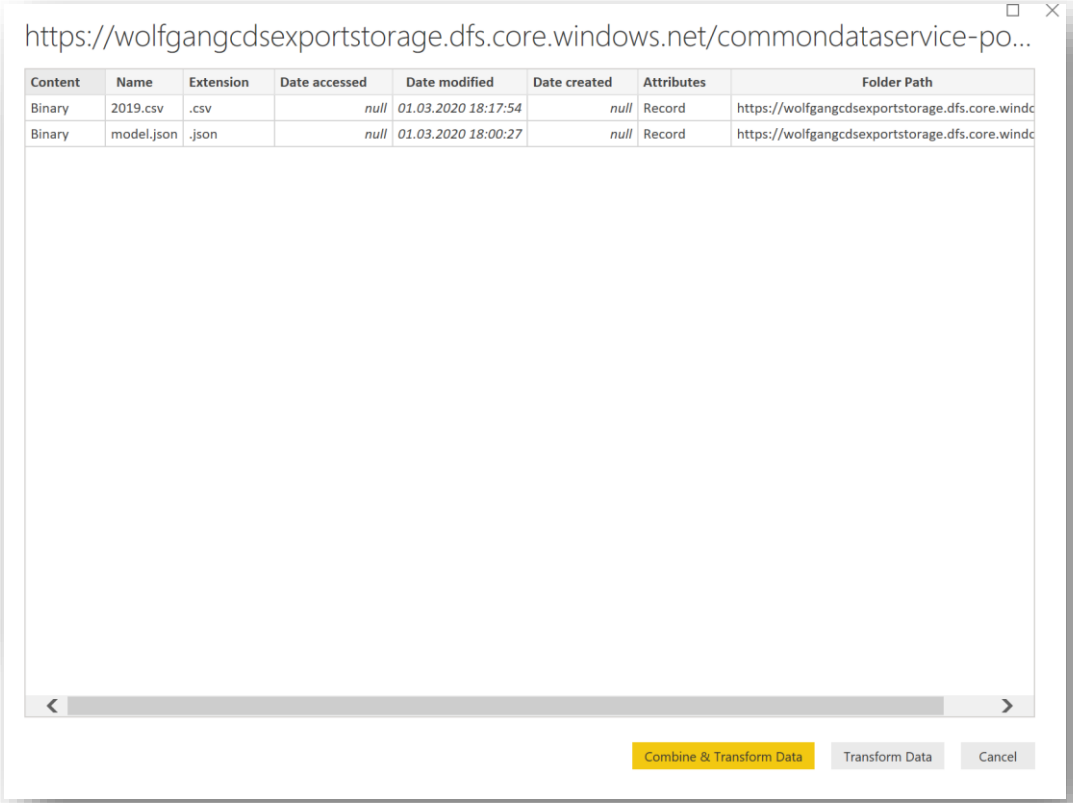
Enable analytics on CDS data

Continuous data & Metadata updates

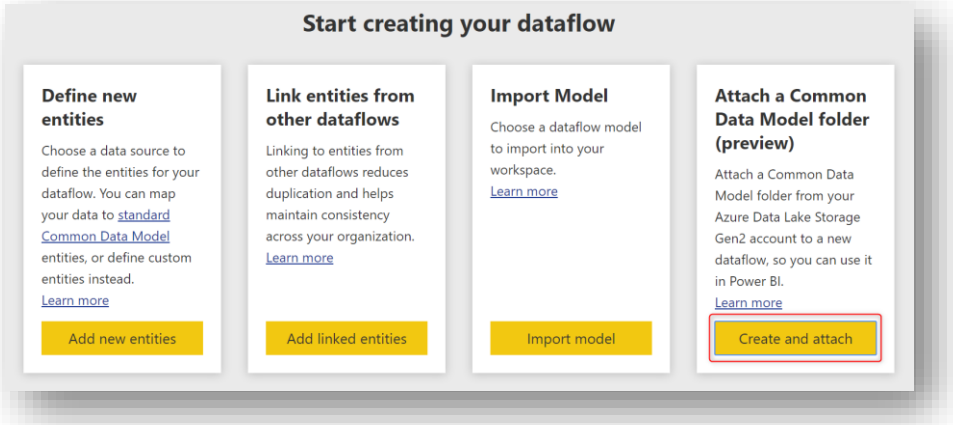


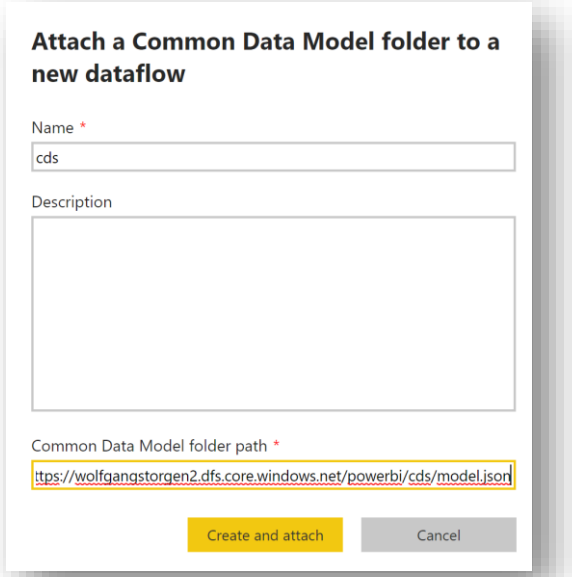
Consume in Power BI – option 1

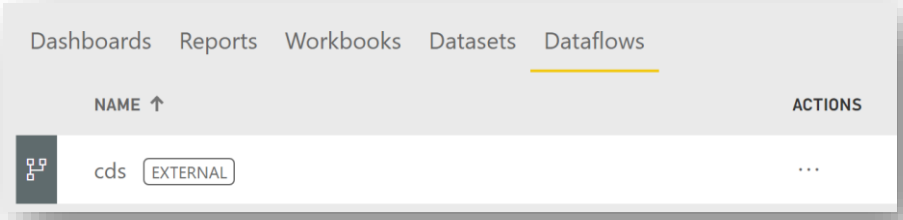


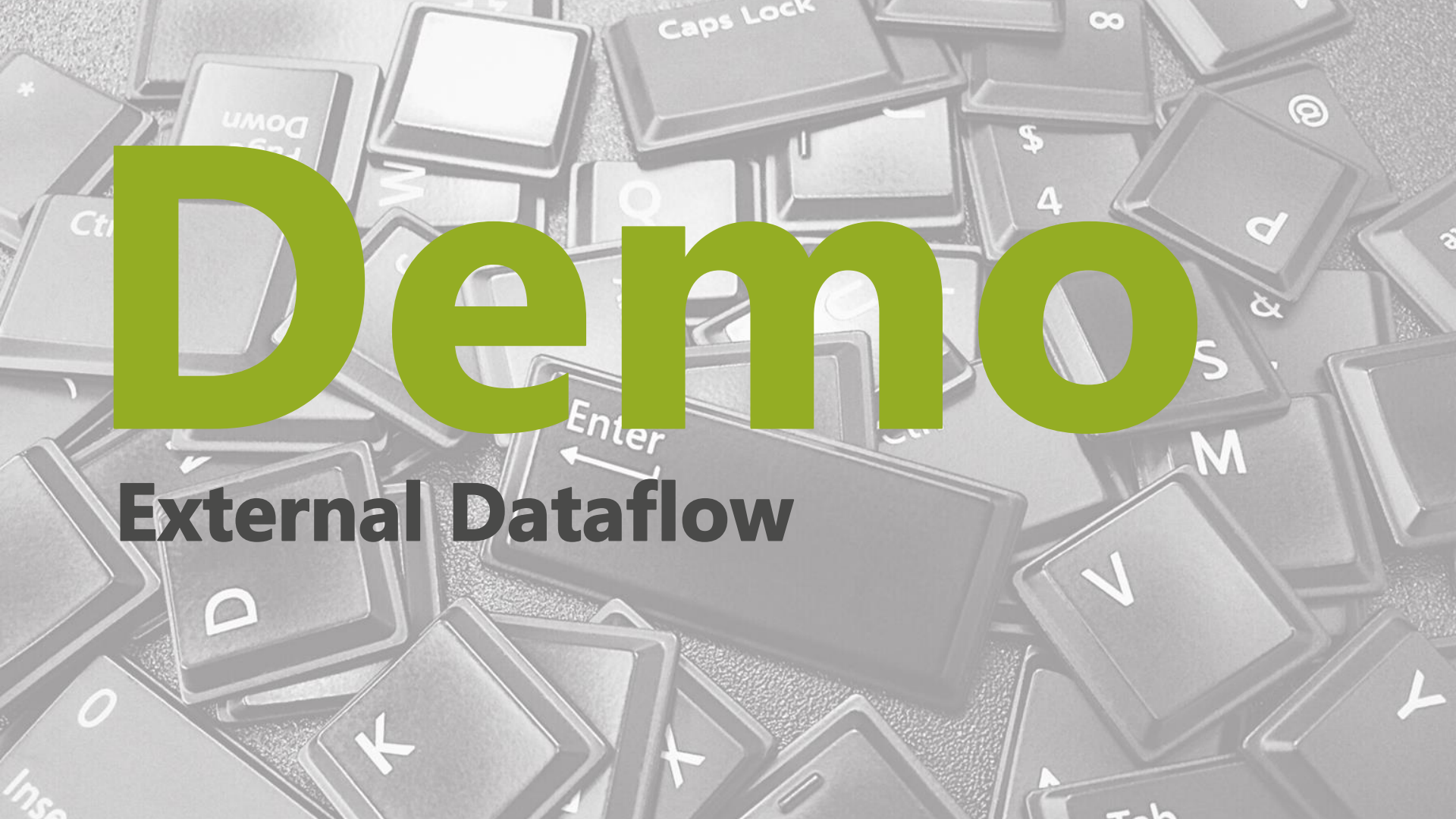


Consume in Power BI – option 2

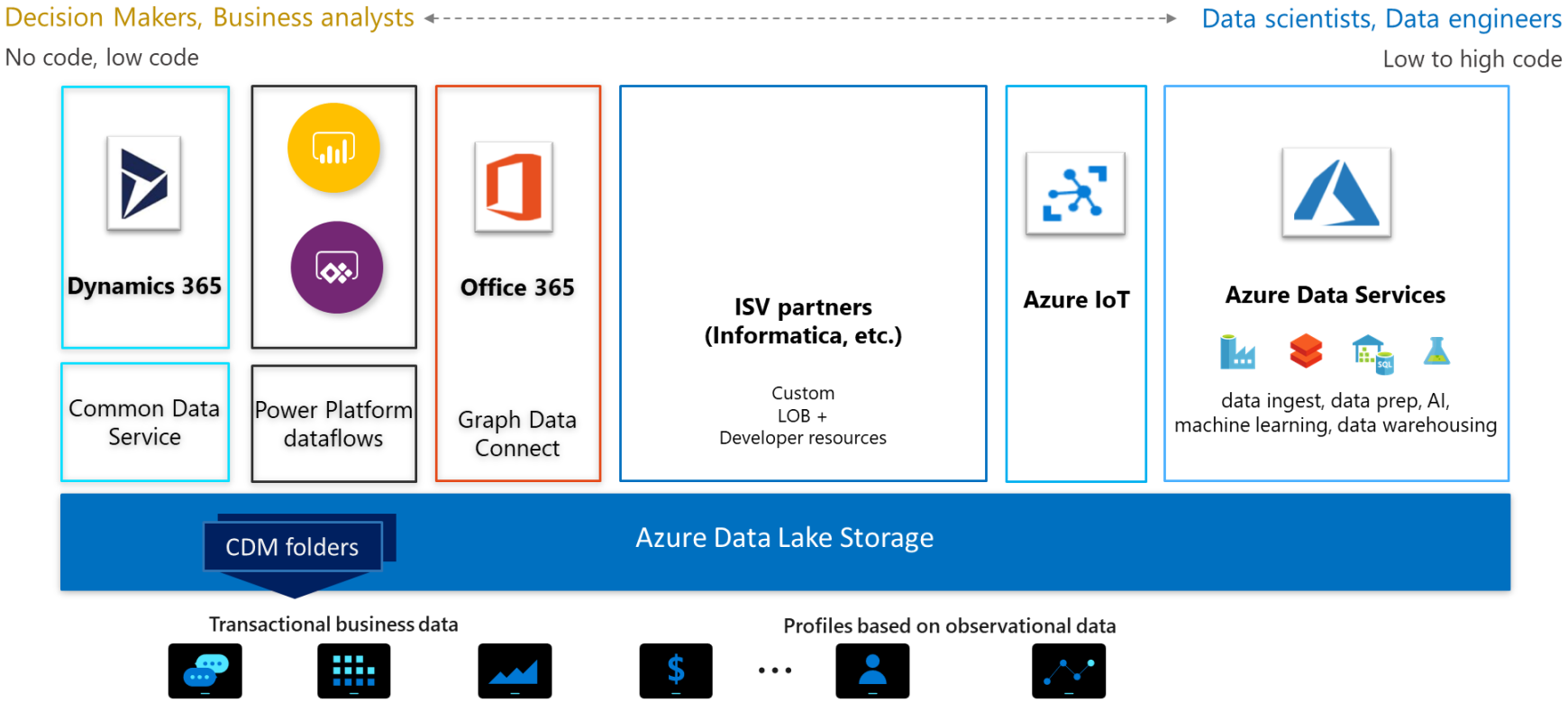







Demo
External Dataflow

Power Platform Dataflows and the Data Lake



SQL Server Konferenz 2020
powered by PASS Deutschland e.V.

Thank you very much for your attention.

Vielen Dank für Eure Aufmerksamkeit.

Wolfgang Strasser

wolfgang@powerofbi.at

@wstrasser

CUBIDO
AN ACP GROUP COMPANY

PASS DEUTSCHLAND e.V.
Microsoft Data Platform Community