

PowerPlatform Dataflows PowerQuery (and more) in the Cloud

Wolfgang Strasser 





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



Let your data flow
Introducing Power BI dataflows

Wolfgang Strasser



What are (Power BI) dataflows?



Information Hierarchy

Reusability

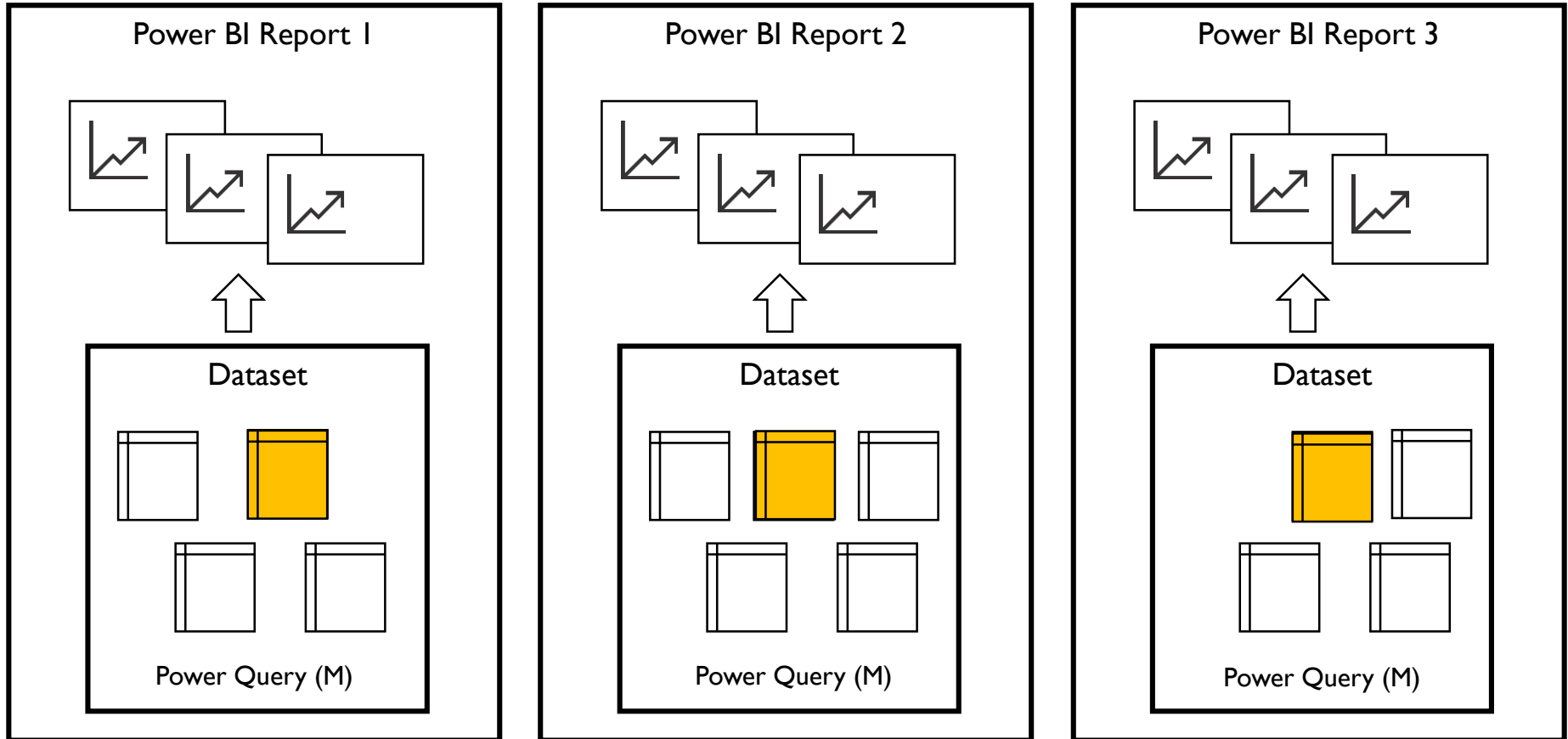
Common Data Model (CDM)

PowerQuery (Online)

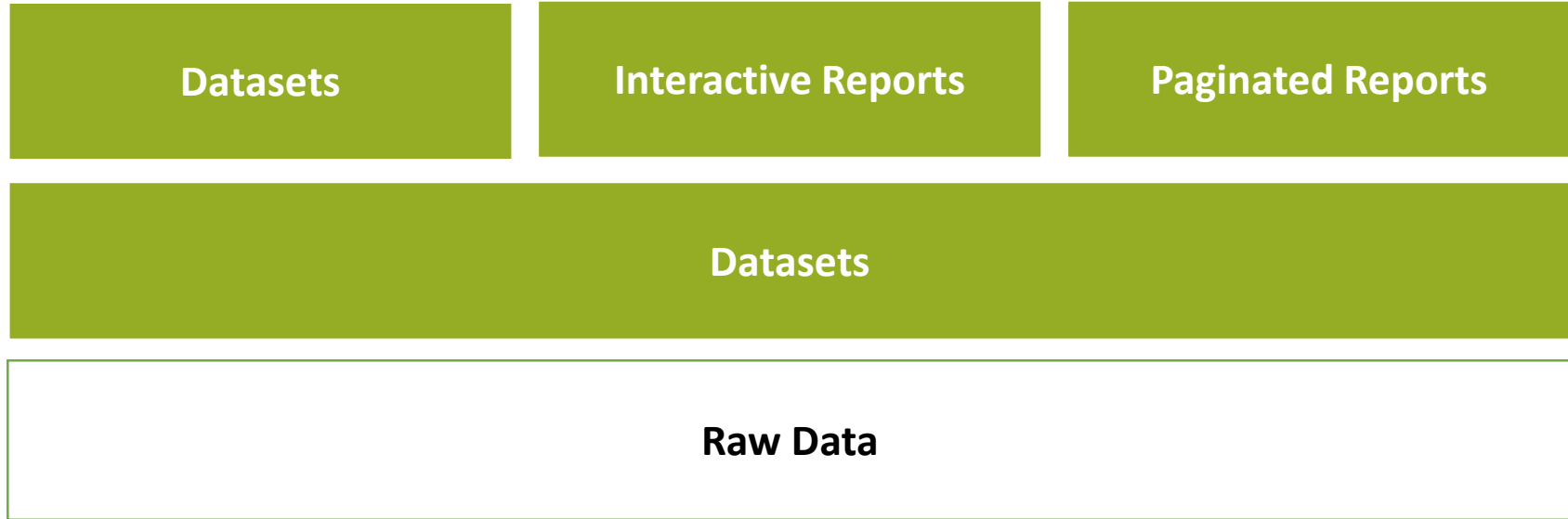
Data Lake



ETL in the Context of Power BI

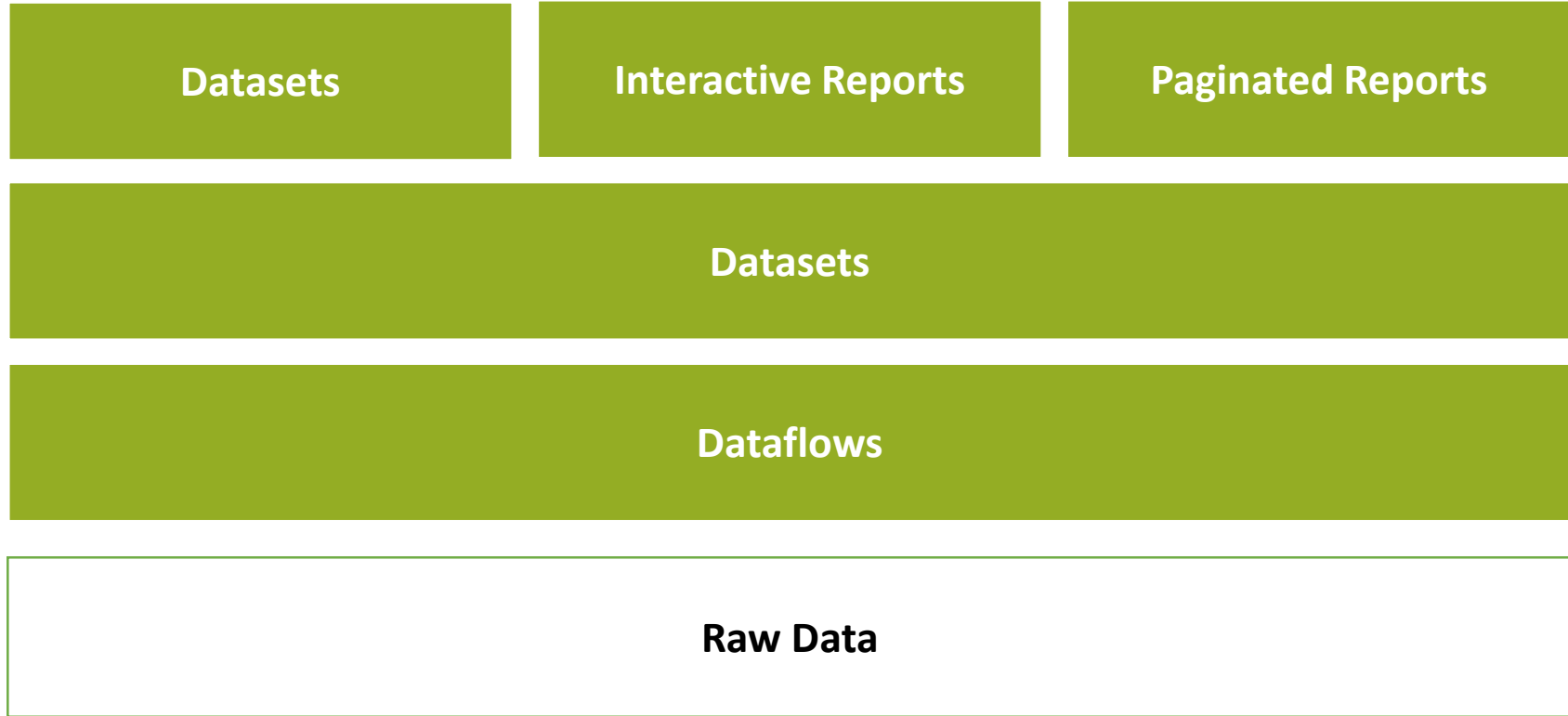


Power BI Information Hierarchy (before)



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

Power BI Information Hierarchy (before)



Fact Sheet – PBI dataflows

PowerQuery processes in the cloud

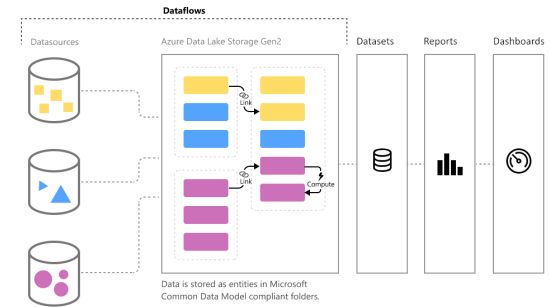
.. run independently of a Power BI report

.. 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

Power Query - Choose data source

All categories File Database Power Platform Azure Online services Other

 Search

Data sources

 Access File	 Excel File	 Folder File	 JSON File	 PDF File
 SharePoint folder File	 Text/CSV File	 XML File	 Amazon Redshift Database	 Google BigQuery Database
 IBM Db2 database Database	 MySQL database Database	 Oracle database Database	 PostgreSQL database Database	 SQL Server database Database
 Sybase database Database	 Teradata database Database	 Vertica Database	 Azure Blobs Azure	 Azure Data Explorer (Kusto) Azure
 Azure Data Lake Storage Gen2 Azure	 Azure HDInsight Spark Azure	 Azure SQL Data Warehouse Azure	 Azure SQL database Azure	 Azure Tables Azure
 Microsoft Exchange Online Online services	 Salesforce objects Online services	 Salesforce reports Online services	 SharePoint Online list Online services	 Smartsheet Online services
 Active Directory Other	 Common Data Service Power Platform	 OData Other	 Odbc Other	 Power BI dataflows Power Platform

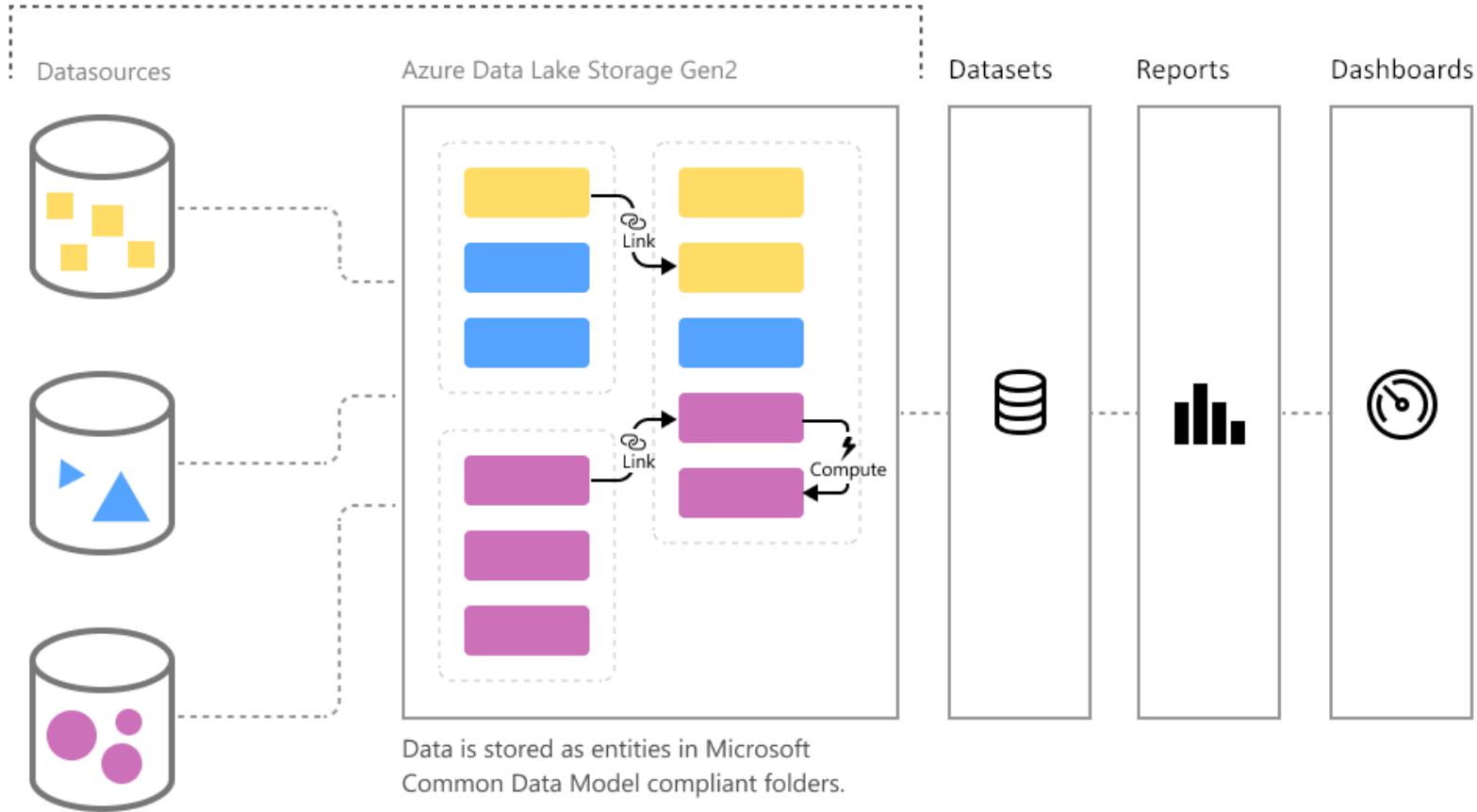
Templates

 Accounts, leads, opportunities Salesforce Online services	 Lead to cash Dynamics 365 Sales Online services	 Leads, opportunities Dynamics 365 Sales Online services	 Quotes, orders, invoices Dynamics 365 Sales Online services
---	--	--	--

Dataflows are like Excel

[illegible]

Dataflows

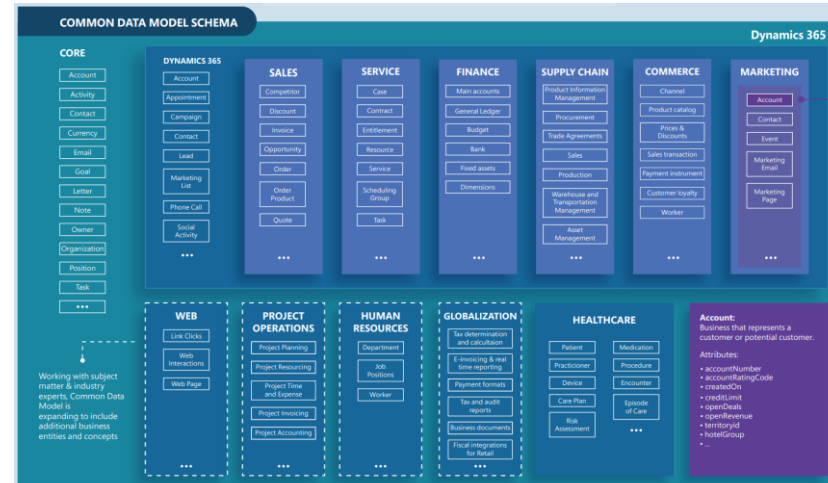


Common Data Model

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

Get Data

All
File
Database
Power Platform
Azure
Online Services
Other

Power Platform

Power BI datasets
Power BI dataflows
Common Data Service
Power Platform dataflows

Navigator

Display Options

- Power BI dataflows [9]
 - Bauer BI 2019
 - dataflows CPH
- Dataflows demo (Premium) [8]
 - Combined OrdersInformation
 - Customer Comments
 - Date dataflow
 - Dim Date
 - OrderStatistics
 - Premium DataFlow Linked Plus Computed
- Sales Combined [3]
 - ☐ CalendarTable
 - ☐ Sales
 - ☒ Sales Combined
- Sales dataflow
- Dataflows Demos [9]
- dataflows HQ
- DemoWorkspace
- PASSCamp Wolfgang
- PASSCamp Wolfgang DEMO

Sales Combined

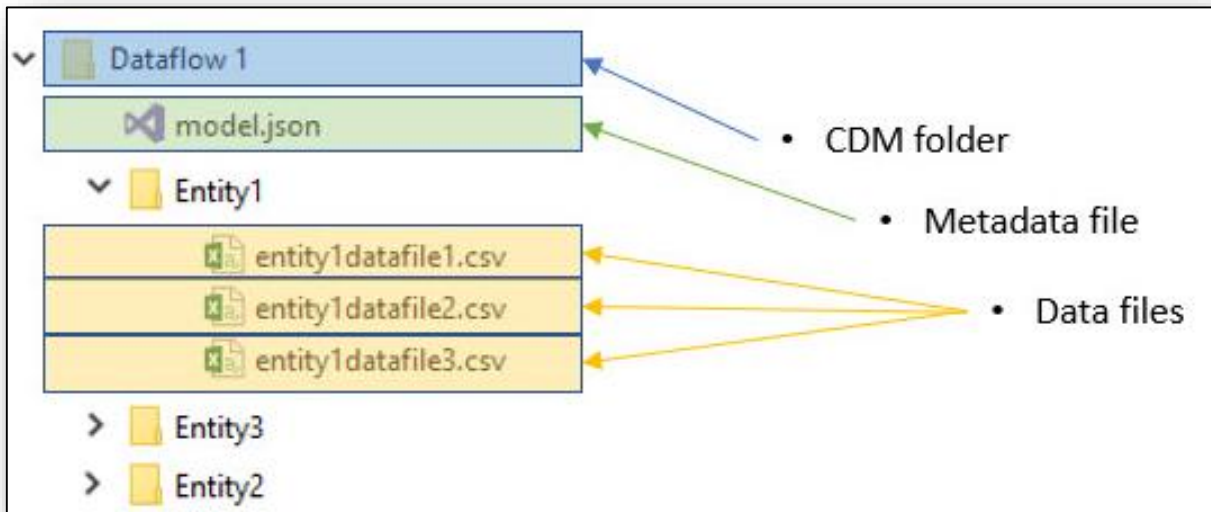
shipToContactName	productName	quantity	createdOn	SalesID
Customer 1	Bike	10	01.08.2017 00:00:00	
Customer 1	TV	5	01.08.2017 00:00:00	
Customer 2	Bike	5	01.09.2017 00:00:00	
Customer 2	TV	3	01.09.2017 00:00:00	
Customer 1	Bike	4	01.10.2017 00:00:00	
Customer 2	TV	9	01.10.2017 00:00:00	
Customer 1	Bike	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	

Load

Transform Data

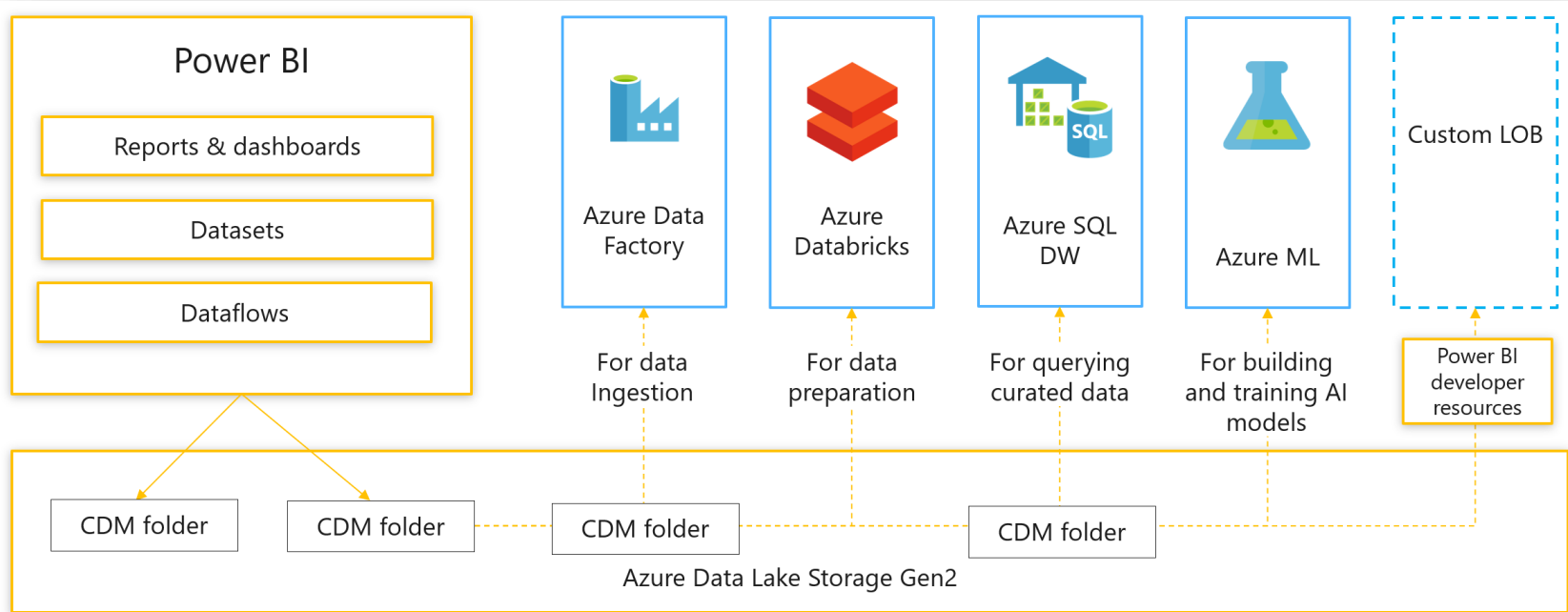
Cancel

CDM Folders



```
{ Sales From Excel.json • First Dataflow.json •  
1 {  
2   "name": "Sales From Excel",  
3   "description": "",  
4   "version": "1.0",  
5   "culture": "en-US",  
6   "pbi:mashup": {  
7     "fastCombine": true,  
8     "allowNativeQueries": false,  
9     "queriesMetadata": {  
10      "Sales": {  
11        "queryId": "40b69972-542b-49ee-ab09-bff5",  
12        "queryName": "Sales",  
13        "loadEnabled": true  
14      }  
15    },  
16    "document": "section Section1;\r\n\nshared Sales =  
17  },  
18  "entities": [  
19    {  
20      "$type": "LocalEntity",  
21      "name": "Sales",  
22      "description": "",  
23      "pbi:refreshPolicy": {  
24        "$type": "FullRefreshPolicy",  
25        "location": "Sales.csv"  
26      },  
27      "attributes": [  
28        {  
29          "name": "SalesOrderNumber",  
30          "dataType": "string"  
31        },  
32      ]  
33    }  
34  ]  
35 }
```

Power BI Dataflows and the Data Lake



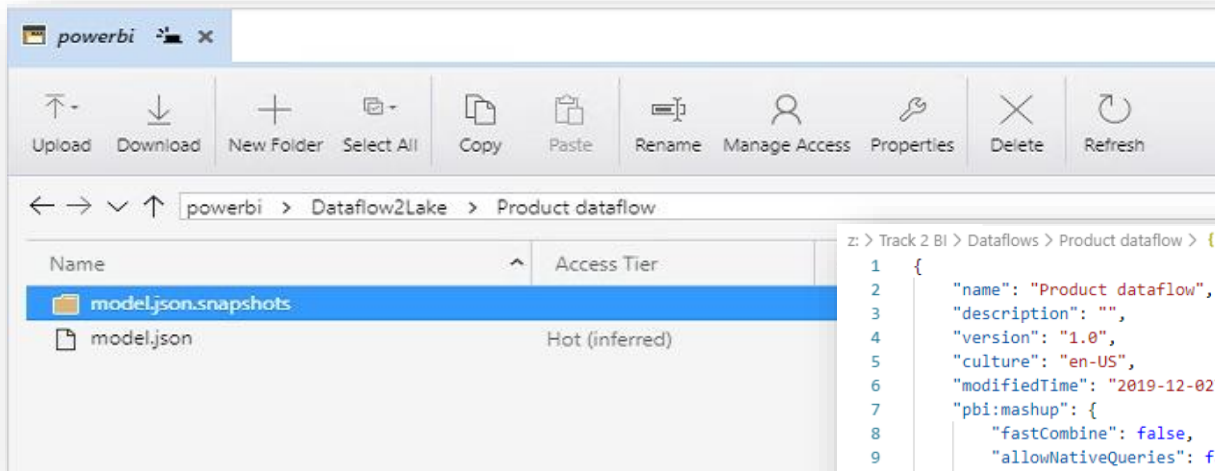
Business analysts

Low/no code

Data scientists
Data engineers

Medium to high code

Bring your Own Storage Account



Dataflow storage (preview)

Your Azure Data Lake Storage account is connected to Power BI. [Learn more](#)

datalake4cdm

Allow workspace admins to assign workspaces to this storage account [Learn more](#)

☒ Allow

Apply

Cancel

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

```
1 {
2   "name": "Product dataflow",
3   "description": "",
4   "version": "1.0",
5   "culture": "en-US",
6   "modifiedTime": "2019-12-02T20:34:28.5677753+00:00",
7   "pbi:mashup": {
8     "fastCombine": false,
9     "allowNativeQueries": false,
10    "queriesMetadata": {
11      "Products": {
12        "queryId": "eed1d3a0-543f-48fd-bd2a-e5938d922b08",
13        "queryName": "Products",
14        "loadEnabled": true
15      }
16    },
17    "document": "section Section1;\r\n\nshared Products = let\r\n\n  Source = Sql.Database(\"passcan
18  },
19  "entities": [
20    {
21      "$type": "LocalEntity",
22      "name": "Products",
23      "description": "",
24      "pbi:refreshPolicy": {
25        "$type": "FullRefreshPolicy",
26        "location": "Products.csv"
27      }
28    }
29  ]
30 }
```



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

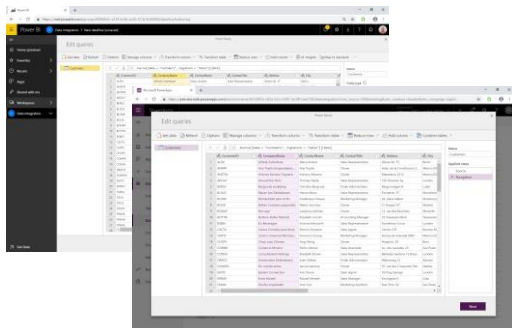
Authoring tools



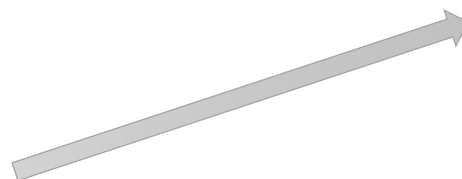
Power BI



Power Apps



Transform data
PowerQuery Online



Common Data Service



Automatic replication
("Athena")



Enrich data with
AI in PQ



ADLS (gen2)



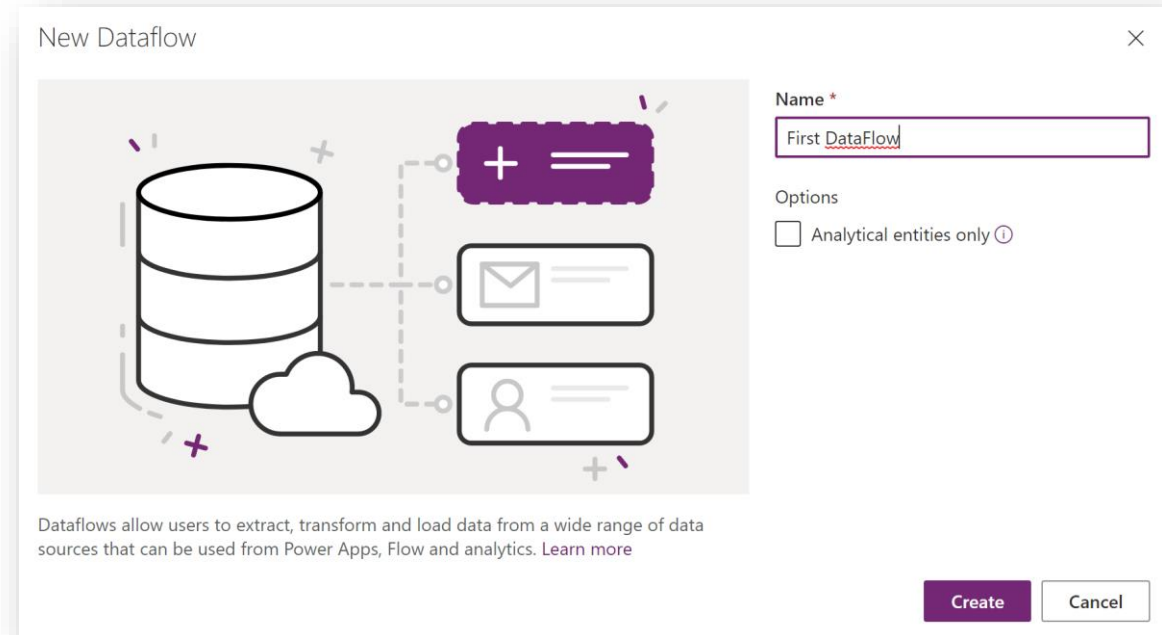
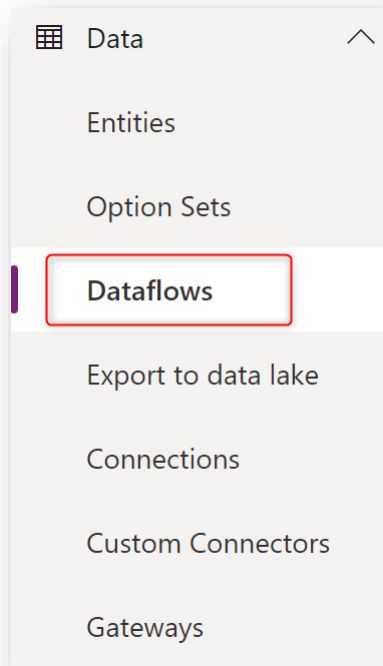
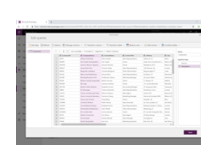
**Data
sources**

**Data Transformation, Preparation
Enrichment**

**Data
Storage**

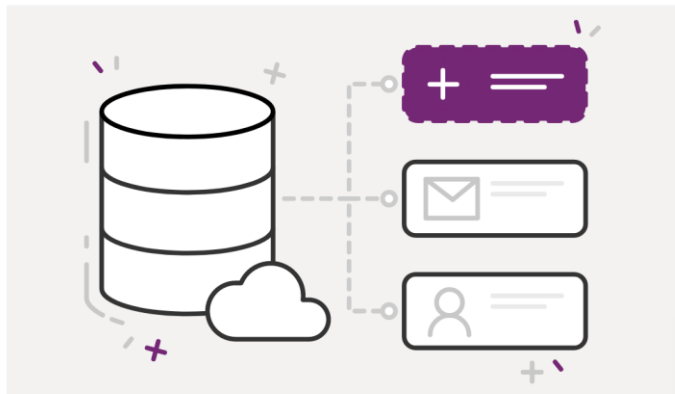
**Data
consumers**

Load Data into CDS



“Normal” plus Analytic entities

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.

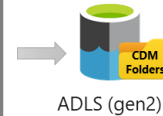
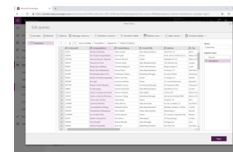
☐ Delete rows that no longer exist in the query output ⓘ

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?

← → ∨ ↑ power-platform-dataflows > environments > orgfcf124ca > First DataFlow

Name	Access Tier	Access Tier Last Modified
dataflow		
model.json.snapshots		
Sales		
model.json	Hot (inferred)	

Export to Data Lake – Project “Athena”



Automatic replication
("Athena")

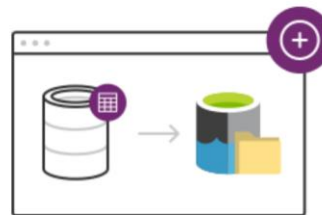


ADLS (gen2)

- Data
- Entities
- Option Sets
- Dataflows
- Export to data lake**
- Connections
- Custom Connectors
- Gateways

+ New link to data lake ← Import from solution ↻ Refresh

Export to data lake



You haven't linked the Common Data Service environment to a data lake
Before you can export to data lake, link the Common Data Service environment to a data lake

[New link to data lake](#)

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

Identity-based access for file shares
Azure Active Directory Domain Service (AAD DS) ⓘ
Disabled Enabled

Data Lake Storage Gen2
Hierarchical namespace ⓘ
Disabled **Enabled**

Information: The current combination of storage account kind, performance, secure transfer required, access tier, replication, large file shares, and hierarchical namespace is not supported. [Learn more.](#)

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.



Manage entities for wolfgangcdsexportstorage

Manage entities

2 of 107 selected



<input type="radio"/>	Approval Request	msdyn_flow_approvalrequest
<input type="radio"/>	Approval Response	msdyn_flow_approvalresponse
<input checked="" type="radio"/>	Article	kbarticle
<input type="radio"/>	Attachment	attachment
<input type="radio"/>	Attachment	activymimeattachment
<input type="radio"/>	Attribute Map	attributemap
<input type="radio"/>	AttributImageConfig	attributeimageconfig
<input type="radio"/>	Await All Approval Model	msdyn_flow_awaitallapprovalmodel
<input type="radio"/>	Basic Approval Model Data	msdyn_flow_basicapprovalmodel
<input checked="" type="radio"/>	Business Unit	businessunit
<input type="radio"/>	Canvas App	canvasapp
<input type="radio"/>	Category	category

Demo

What's in the Data Lake?




← → ∨ ↑ <input type="text" value="commondataservice-powerofbiatdefaultupgrade-orgbfa74bd0"/>		
Name	Access Tier	Access
businessunit		
Microsoft.Athena.TrickleFeedService		
model.json		Hot (inferred)

What's in my Data Lake?



commondataservice-powerofbiatdefaultupgrade-orgbfa74bd0

Name

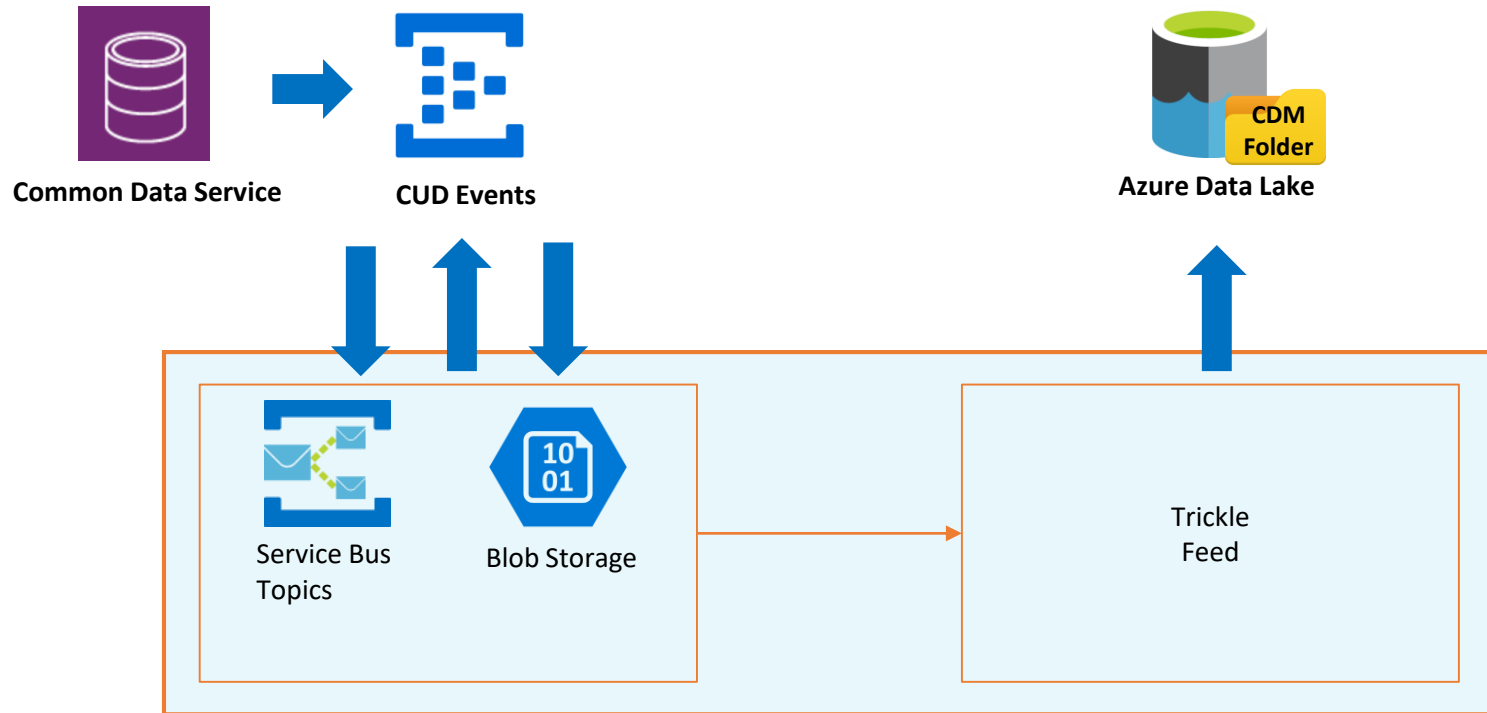
-  businessunit
-  Microsoft.Athena.TrickleFeedService
-  model.json

```
{
  "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

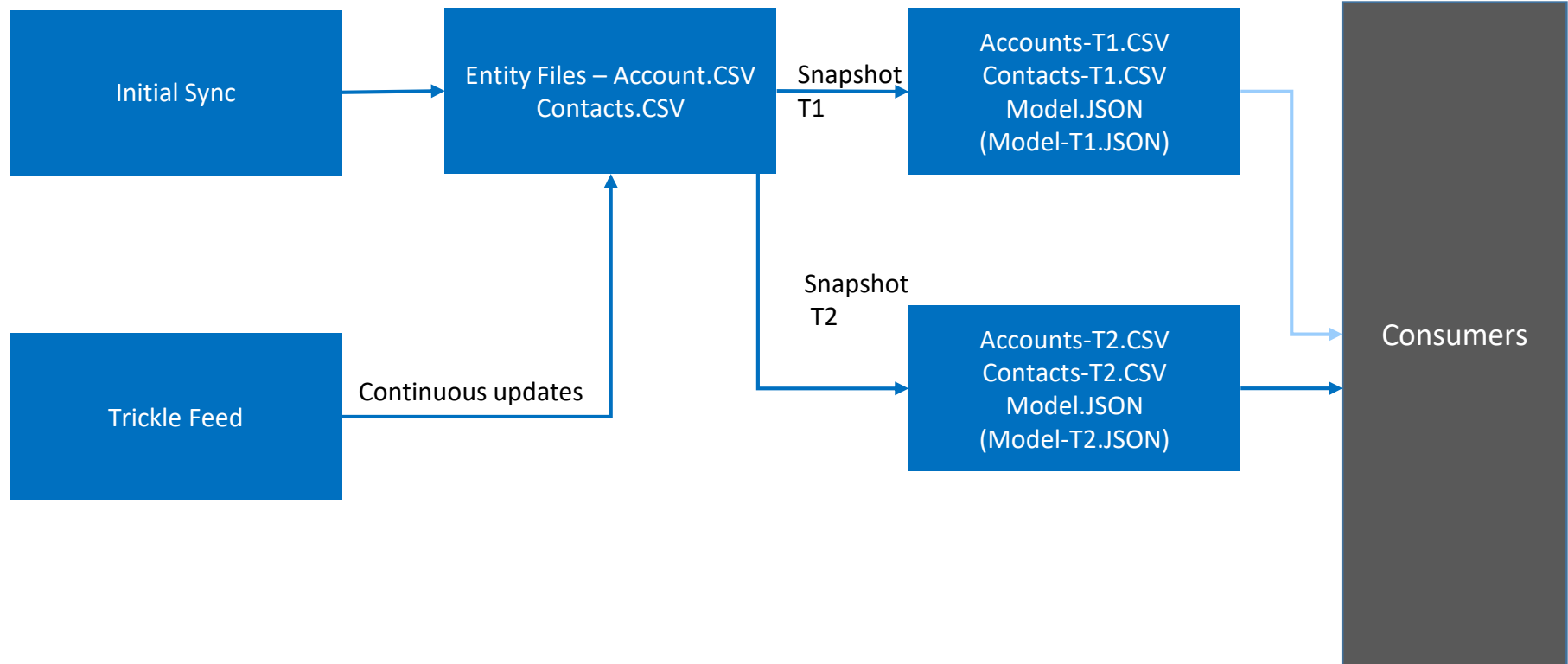
Functionality

- 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

Get Data

All

File

Database















Power Platform

Azure

Online Services

Other

Azure

-  Azure SQL database
-  Azure SQL Data Warehouse
-  Azure Analysis Services database
-  Azure Blob Storage
-  Azure Table Storage
-  Azure Cosmos DB
-  Azure Data Lake Storage Gen2
-  Azure Data Lake Storage Gen1
-  Azure HDInsight (HDFS)
-  Azure HDInsight Spark
-  HDInsight Interactive Query
-  Azure Data Explorer (Kusto)
-  Azure Cost Management
-  Azure Time Series Insights (Beta)

https://wolfgangcdsexportstorage.dfs.core.windows.net/commondataservice-po...

Content	Name	Extension	Date accessed	Date modified	Date created	Attributes	Folder Path
Binary	2019.csv	.csv	null	01.03.2020 18:17:54	null	Record	https://wolfgangcdsexportstorage.dfs.core.windc
Binary	model.json	.json	null	01.03.2020 18:00:27	null	Record	https://wolfgangcdsexportstorage.dfs.core.windc

Combine & Transform Data

Transform Data

Cancel

Consume in Power BI – option 2

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

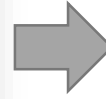
Choose a dataflow model to import into your workspace. [Learn more](#)

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



Attach a Common Data Model folder to a new dataflow

Name *

cds

Description

Common Data Model folder path *

<https://wolfgangstorgen2.dfs.core.windows.net/powerbi/cds/model.json>

Create and attach

Cancel



Dashboards Reports Workbooks Datasets Dataflows

NAME ↑

ACTIONS



cds

EXTERNAL

...



Demo

External Dataflow

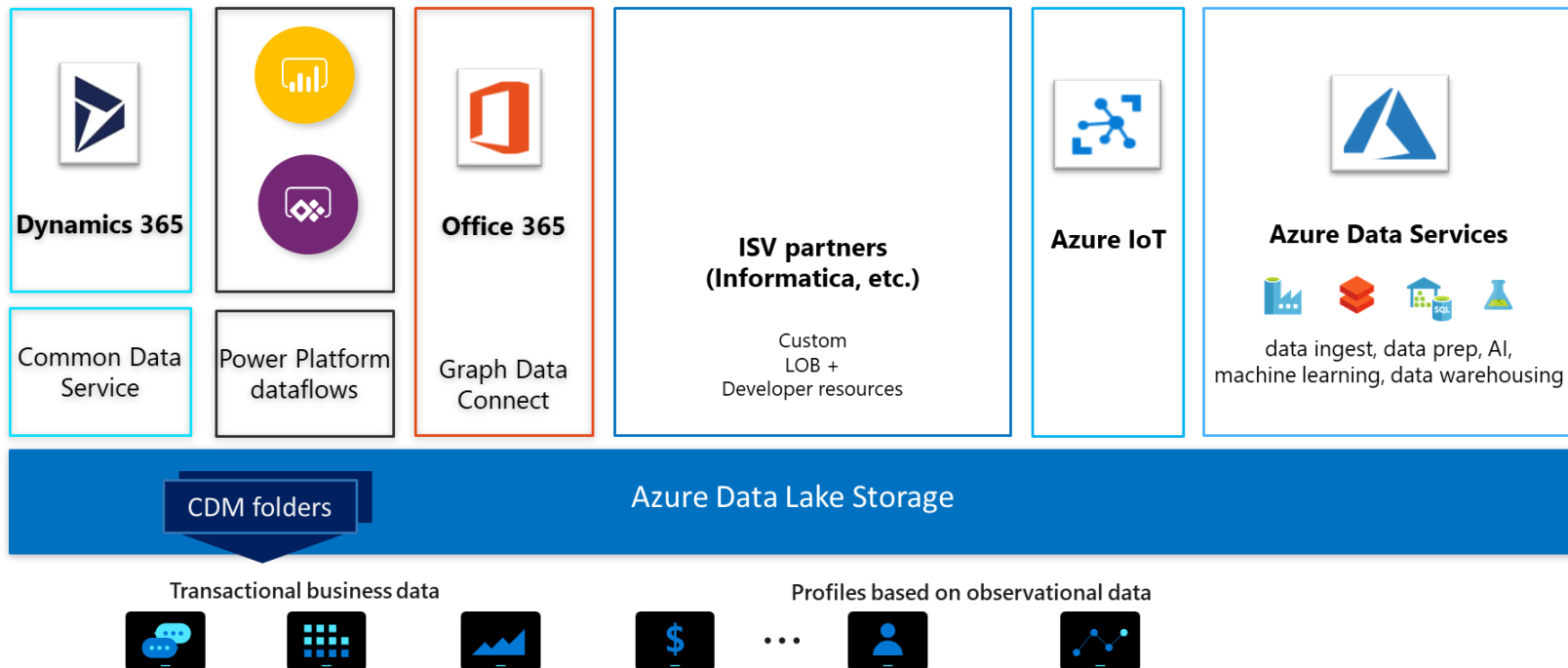
Power Platform Dataflows and the Data Lake

Decision Makers, Business analysts

Data scientists, Data engineers

No code, low code

Low to high code





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



UBIDO

AN ACP GROUP COMPANY



Microsoft
Data Platform
Community

PASS
DEUTSCHLAND e.V.