Let your data flow

Introducing Power BI dataflows

Wolfgang Strasser, MVP
Let your data flow – Introducing Power BI dataflows

Power BI serves as a self-service BI platform with a strong focus on data preparation and interactive analysis. With the introduction of Power BI dataflows, self-service data preparation is brought to a new level. The main concepts used are:

• Usage of common and mature technologies: data is stored as entities following the Common Data Model in Azure Data Lake Storage Gen2;
• Integration: dataflows are created and managed in Power BI app workspace
• Self-Service and low-code/no-code - Power Query is used as data preparation engine
• Connectivity: dataflows will support a variety of different data sources (including cloud-based and on-premises sources)

Join this session if you would like to learn more about the basic concepts and especially see Power BI dataflows in action.
WHY Power BI dataflows?
Evolution of Business Intelligence

1st wave
Technical BI

2nd wave
Self-service BI

3rd wave
End user BI

IT to end user
Analyst to end user
Everyone
Business Intelligence needs DATA

but: Data Preparation takes time
(60-80% of a BI project)
Self-Service BI needs

More “Power” for the key users / business analysts

Faster access to new data sources

Integration

Data Models
Self-Service BI – Current State

BI ✓
ETL ×✓
ETL in the Context of Power BI
Introducing Power BI dataflows
Power BI information hierarchy

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

**App workspaces** only (Not available in “My workspace”)

**Entities ~ Tables**

**Dataflow** = data preparation pipeline

**Power Query / M**

- Multiple dataflows
- Transaction Consistency
- Dataflow Calculation engine
Good to Know

Re-usable data integration

Created / managed in the cloud

Proven technologies

- Power Query
- Azure Data Lake Storage Gen2
- Common Data Model
Common Data Model

Data Schema
Standardized, modular, extensible

Consists of
Entities
Attributes
Semantic metadata
relationships

Taken from https://aka.ms/cdmposter
https://aka.ms/cdmrepo
CDM GitHub Repo
CDS for Apps

Power BI

Data Integration

Common Data Model

ADLS v2

Enrichment and AI

Integration

Intelligence

Dynamics 365 Apps

PowerApps

Power BI Apps
Demo

Create a data flow
Power BI information hierarchy

- Dashboards
- Interactive Reports
- Paginated Reports
- Datasets
- Dataflows
- Raw Data

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

Consume dataflow entities
(=create a dataset)
Who can use PBI dataflows?
Everyone!
<table>
<thead>
<tr>
<th>Dataflow capability</th>
<th>Power BI Pro</th>
<th>Power BI Premium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scheduled refresh</td>
<td>8 per day</td>
<td>48</td>
</tr>
<tr>
<td>Total Storage</td>
<td>10 GB/user</td>
<td>100 TB/node</td>
</tr>
<tr>
<td>Dataflow Authoring with Power Query Online</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Dataflow Management within Power BI</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Dataflows Data Connector in the Power BI Desktop</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Integration with Azure</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Computed Entities (in-storage transformations via M)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>New connectors</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Dataflow incremental refresh</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Running on Power BI Premium capacity / Parallel execution of transforms</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Dataflow linked entities</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Standardized Schema / Built-In Support for the Common Data Model</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

[Link](https://docs.microsoft.com/en-us/power-bi/service-dataflows-overview#dataflow-capabilities-on-power-bi-premium)
Behind the Scenes
Behind the Scenes

Dataflows

Datasources

Azure Data Lake Storage Gen2

Datasets

Reports

Dashboards

Data is stored as entities in Microsoft Common Data Model compliant folders.

https://docs.microsoft.com/en-us/power-bi/service-dataflows-overview
Dataflows and the Data Lake

“BYOSA” – Bring Your Own Storage Account

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

- CDM folder
- Metadata file
- Data files

Dataflows Administration
Demo

Tenant settings
BYOSA
Dataflow settings
Schedule refresh
Premium – Capacity Settings
Enable dataflows

Dataflow settings (preview)

Create and use dataflows (preview)
Enabled for the entire organization

Users in the organization can create and use dataflows. Learn more

Enabled

Apply  Cancel

This setting applies to the entire organization
BYOSA – Use your own Data Lake

Admin portal

- Usage metrics
- Users
- Audit logs
- Tenant settings
- Capacity settings
- Embed Codes
- Organization visuals
- Dataflow settings (preview)
- Workspaces

Dataflow storage (preview)

Your organization's dataflow data is stored in Power BI provided storage.

Use your own Azure Data Lake Storage

When you use Azure Data Lake Storage Gen2 with Power BI, authorized users can access your organization’s dataflow data to build cloud-scale data solutions leveraging Azure services, including AI, machine learning, and more. Data Lake Storage is secured, massively scalable, and built to the open HDFS standard, allowing you to run massively-parallel analytics.

Learn more

- How to use Azure Data Lake Storage with Power BI
- About Data Lake Storage
- Azure Data Lake Storage Gen2 pricing

Connect your Azure Data Lake Storage Gen2 Preview
## Dataflow list

<table>
<thead>
<tr>
<th>Name</th>
<th>Actions</th>
<th>Last Refresh</th>
<th>Next Refresh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calendar dataflow</td>
<td><img src="edit" alt="Edit" /> <img src="download" alt="Download" /> <img src="more" alt="More" /></td>
<td>1/23/2019, 8:18:22 PM</td>
<td>N/A</td>
</tr>
<tr>
<td>First Dataflow</td>
<td><img src="edit" alt="Edit" /> <img src="download" alt="Download" /> <img src="more" alt="More" /></td>
<td>1/23/2019, 8:05:46 PM</td>
<td>1/27/2019, 2:00:00 AM</td>
</tr>
<tr>
<td>Linked Entities dataflow</td>
<td><img src="edit" alt="Edit" /> <img src="download" alt="Download" /> <img src="more" alt="More" /></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Sales From Excel</td>
<td><img src="edit" alt="Edit" /> <img src="download" alt="Download" /> <img src="more" alt="More" /></td>
<td>1/23/2019, 9:04:52 PM</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Schedule Refresh

Calendar dataflow
First Dataflow
Linked Entities dataflow
Sales From Excel

Dataflows are in preview

Settings for First Dataflow
This dataflow has been configured by wlfoggen@powerbi.at

Last refresh succeeded: Wed Jan 23 2019 20:05:46 GMT+0100 (Central European Standard Time)
Refresh history

Gateway connection
Data source credentials
Scheduled refresh

Keep your data up to date
On

Refresh frequency
Daily

Time zone
UTC Coordinated Universal Time

Time
1:00
AM

×
PBI Premium – Capacity Settings

Dataflows (Preview) - Starting
We’re preparing your workload.

Max Memory (%)
30

[Apply] [Cancel]
Resources
Resources

• Dataflows Whitepaper (by Amir Netz)
  • https://go.microsoft.com/fwlink/?linkid=2011419&clcid=0x409
• Documentation
  • https://docs.microsoft.com/en-us/power-bi/service-dataflows-overview
• Matthew Roche
• Guy in a Cube
  • https://www.youtube.com/channel/UCFp1vaKzpfoGai0vE5VJ0w
• Rezas Blog
  • http://radacad.com/blog
• My blog (#24DaysPowerPlatform)
  • https://workingondata.wordpress.com/2018/12/22/24-days-of-powerplatform-day-22-what-are-power-bi-dataflows/
Thanks for joining!

Wolfgang Strasser

- wolfgang @ powerofbi.at
- @wstrasser
- https://workingondata.wordpress.com