



Power BI dataflows deep dive

Riccardo Perico

Data Platform & BI Specialist @ SolidQ





*Own your career with interactive learning built
by community and guided by data experts.
Get involved. Get ahead.*

Explore your PASS community



 PASS
MARATHON

Free online
webinar events



PASS
LOCAL
GROUPS

Local user groups
around the world



 PASS
SUMMIT


Connect with the global
data community



PASS
VIRTUAL
GROUPS

Online special
interest user groups



 PASS CONNECTOR
INSIGHTS

Learning on-demand
and delivered to you



PASS
VOLUNTEERS

Get involved



 PASS
SQLSATURDAY
PORDENONE | 30 MAY 2020

Missed PASS Summit 2019?

Get the Recordings

Download all PASS Summit sessions on Data Management, Analytics, or Architecture for only \$399 USD

More options available at
PASSstuff.com



\$399

Content Stream
download
non-attendee option



*PASS
SUMMIT2020
NOV 10-13 | HOUSTON TX

Summit 2020 Will Launch

In-person and virtual event
planning is underway.

Register Now

We are covering all bases to ensure our community can continue reaching new and exciting heights. Plans are underway for the in-person event you all know and love along with a new venture, a new opportunity: a PASS Summit 2020 Virtual Event.

Find out more at PASS.org/summit



*PASS
SQLSATURDAY
PORDENONE | 30 MAY 2020

Thank you to our Global Sponsors and Supporters





Thank you to
our Local
Sponsors and
Supporters





This event was sponsored by Microsoft

Learn more about SQL Server 2019 today:

- Get free training: aka.ms/sqlworkshops
- Download the SQL19 eBook: aka.ms/sql19_ebook

Riccardo Perico

.....

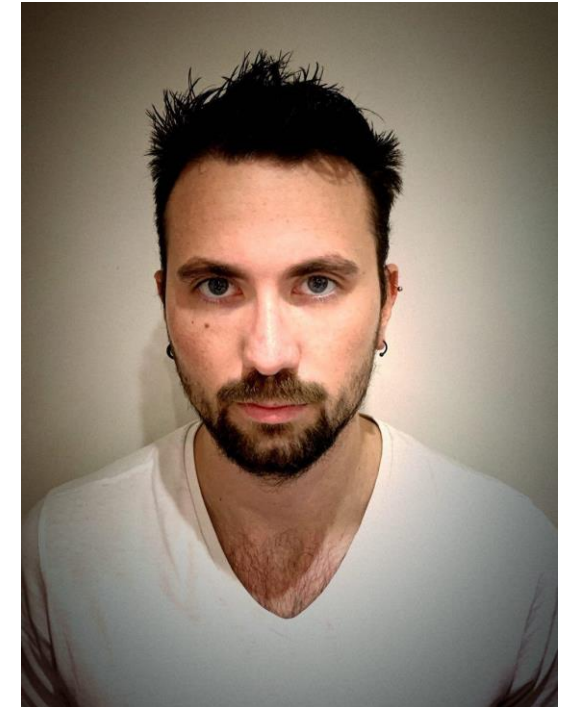
rperico@solidq.com | @R1k91  | 

Data Platform & BI Specialist @

10 years in Microsoft «Data Realm»



Speaker at Azure Saturday, SQL Saturday



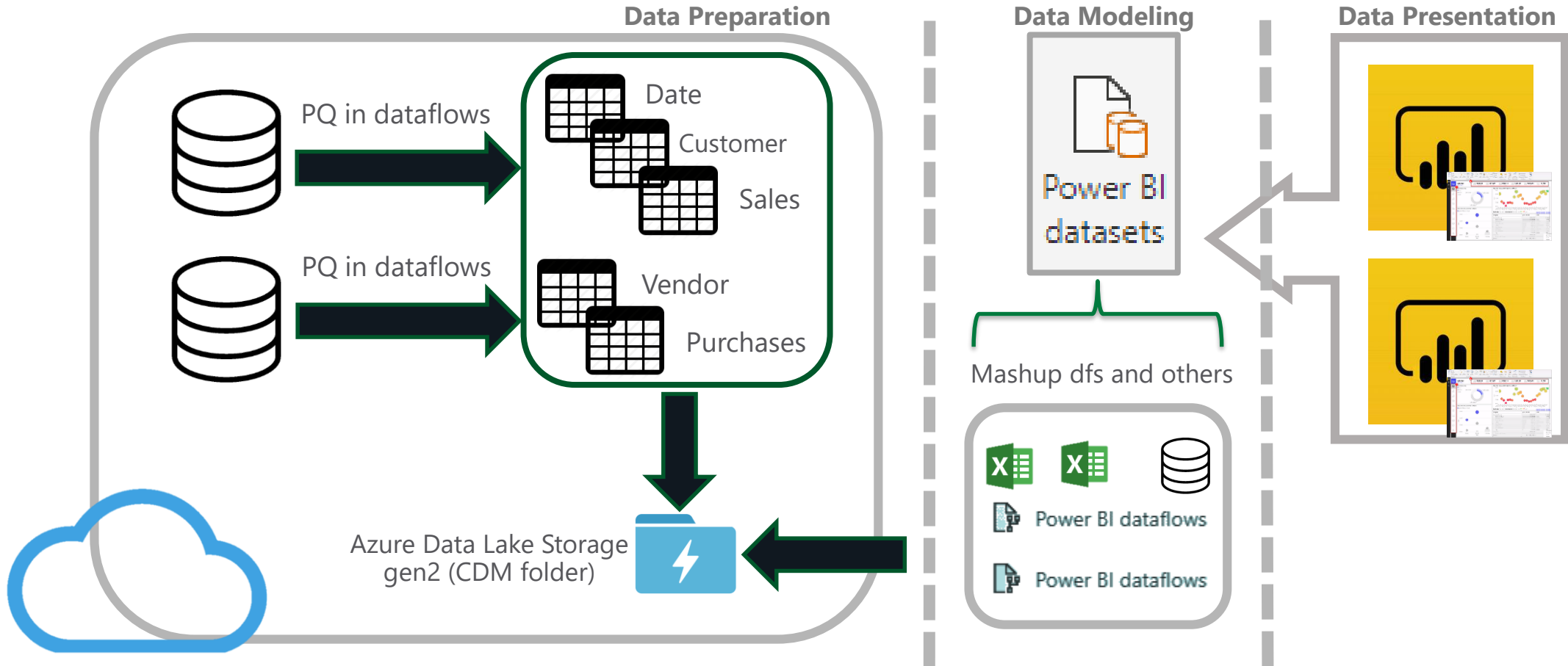
Agenda

- Power BI dataflows: let's do a recap
- ABC for CDM
- What they are and what they're not
- 4 ways to create them
- Power BI dataflows premium feature
- Use your Azure Data Lake Storage gen2
- Best practices, Tips & Tricks
- Q&A

A large, teal-colored abstract graphic on the left side of the slide. It consists of several thick, curved lines that sweep from the top left towards the bottom right, creating a sense of motion and flow.

Power BI dataflows: let's do a recap

Power BI dataflows: let's do a recap



A large, teal-colored abstract graphic on the left side of the slide. It consists of several thick, curved lines that sweep from the top left towards the bottom right, creating a sense of motion and depth. The lines are layered, with some appearing in front of others, and they all curve in a similar direction, suggesting a stylized letter 'C' or a series of connected arcs.

ABC for CDM

What you should know about CDM

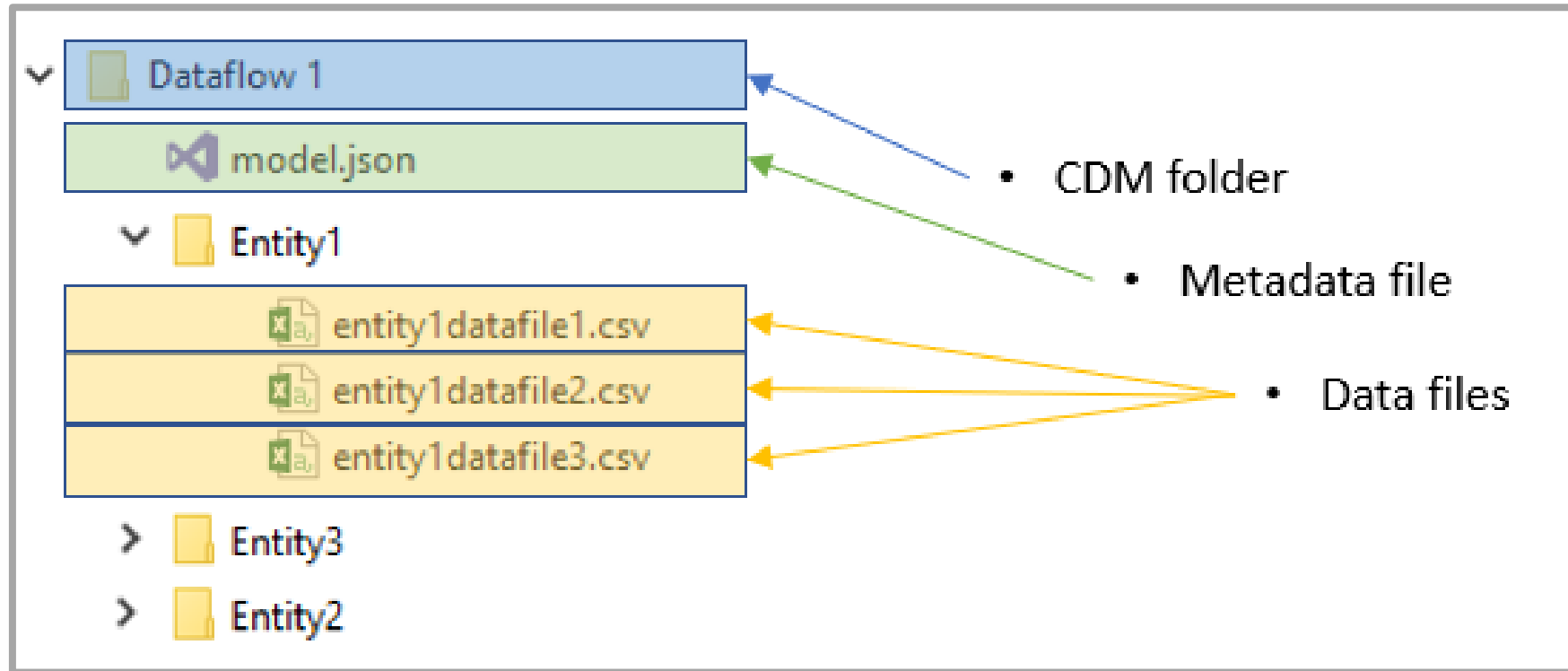
CDM is a **metadata system**

CDM provides a **shared language and consistency** among different apps (i.e. Power Platform, D365 and Azure)

Includes a set of **standard entity schemas** (you wouldn't use 9 out of 10 😊)

CDM Folder is a storage that use CDM metadata

How does CDM folder look like?



A large, teal-colored abstract graphic on the left side of the slide. It consists of several overlapping, curved, ribbon-like shapes that create a sense of depth and movement, resembling a stylized 'S' or a series of connected loops.

What they are and what they're
not

They are... They aren't

- ✓ Self-service **data preparation** tool
- ✓ It's **Power Query**
- ✓ **Store data in CDM folder** in ADLS gen2
- ✓ **Encourage data reuse**
- ✗ Don't replace **datasets**
- ✗ Don't replace **DWH**
- ✗ Don't replace **orchestrating tools** (ADF v2 or SSIS)
- ✗ **Not Premium-only** feature
- ✗ **Don't have additional costs**

dataflows Vs datasets

<https://ssbipolar.com/2019/10/14/power-bite-dataflows-vs-datasets/>

	Power BI dataflows	Power BI datasets
Implementation	CDM folder	Analysis Services Tabular model
Storage	CSV files	Vertipaq
Metadata	CDM – model.json	BISM
Development	Browser	Power BI Desktop
Primary purpose	Data preparation reuse	Data model reuse
Reuse	Acts as data source in multiple datasets	Shared dataset across workspace
Scope of reuse	Entity level reuse	Dataset level reuse
Mashup with other data source	Yes	No (it's coming...)
Reuse outside Power BI	Through ADLS gen 2	Through XMLA (only premium)
Data access methods	Import	Import, Direct Query and Live Connection
Storage Mode	Import (Direct Query in Preview)	Live Connection

Power BI dataflows Pros

Centralization of data preparation (reusability and standardization)

Isolating refresh (1 table a time)

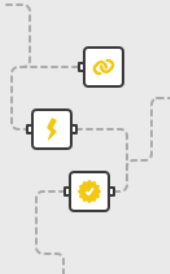
Attach an existing ADLS gen2 (preview)

Speed up dataset refresh



4 ways to create them

4 ways to create them



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](#)

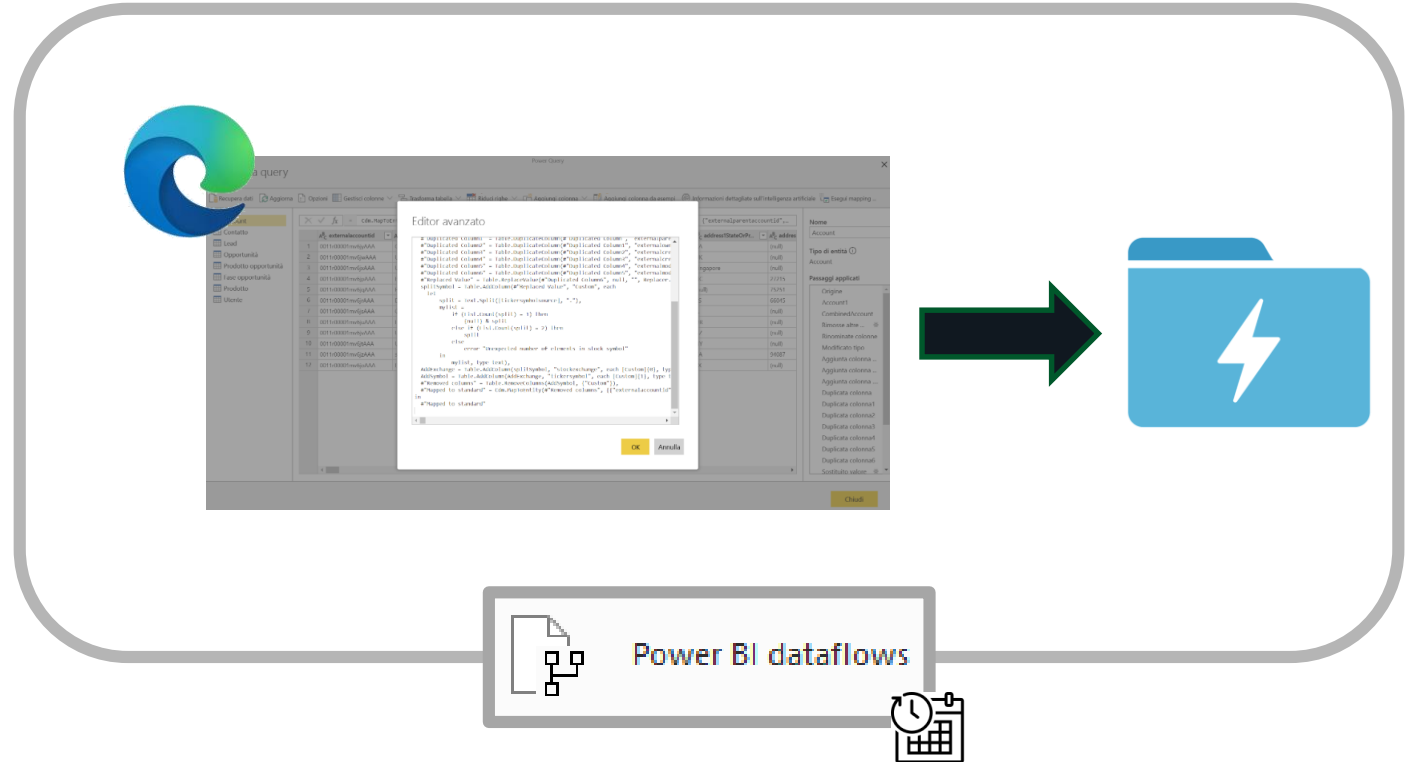
Create and attach

Method 1: Define new entities

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



A large, stylized teal graphic on the left side of the slide, consisting of several overlapping curved lines that form a shape reminiscent of a stylized 'D' or a series of nested parentheses.

Demo: Create dataflows method 1

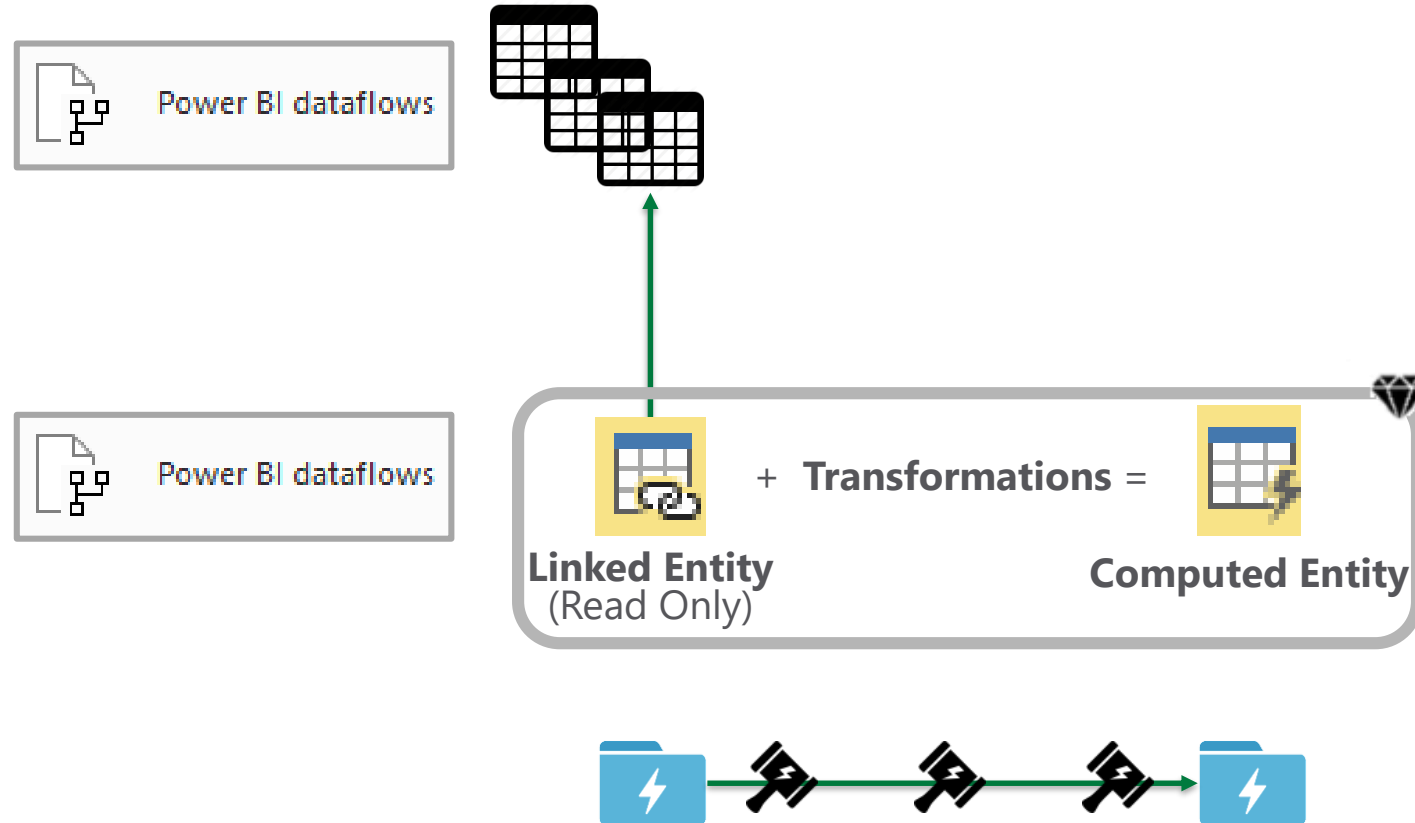
Method 2: Link Entities

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



Method 2: Refresh Matrix

	Same Workspace	Different Workspace
Refresh Source	Source and Destination	Source Only
Refresh Destination	Destination Only	Destination Only

Enable Load

Tasks Done	Resulting Entity	Enable Load on Source(s)	Premium Required	Keep in mind
Entity + Trasformation	Computed	Yes	Yes	
Entity + Trasformation	Computed	No	No	Source Entity won't be materialized in ADLS gen2
Referencing another dataflows' entity	Linked	Yes	Yes	

A large, stylized teal graphic on the left side of the slide, consisting of several overlapping curved lines that form a partial circular shape.

Demo: Create dataflows method 2

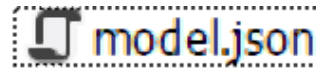
Method 3: Import Model

Import Model

Choose a dataflow model to import into your workspace.

[Learn more](#)

Import model



A large, stylized teal graphic on the left side of the slide, consisting of several overlapping curved lines that form a shape reminiscent of a stylized 'D' or a series of nested parentheses.

Demo: Create dataflows method 3

Method 4: Attach a CDM folder (preview)

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](#)

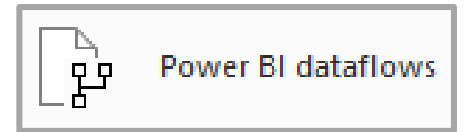
Create and attach



Create



Use



A large, teal-colored abstract graphic on the left side of the slide, consisting of several overlapping, curved, ribbon-like shapes that create a sense of movement and depth.

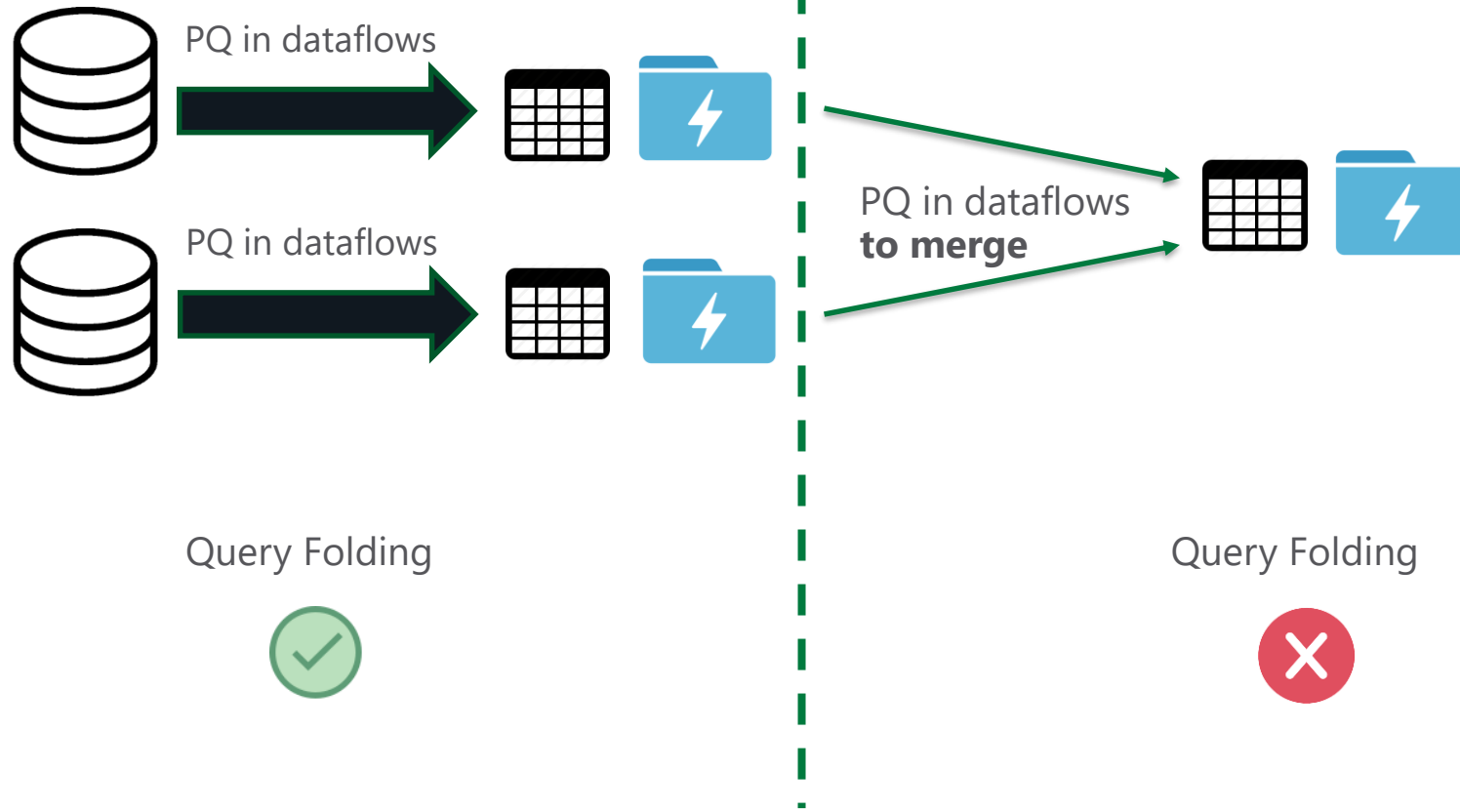
Power BI dataflows Premium Features (\$\$\$\$)

\$\$\$\$: Computed Entities & Linked Entities

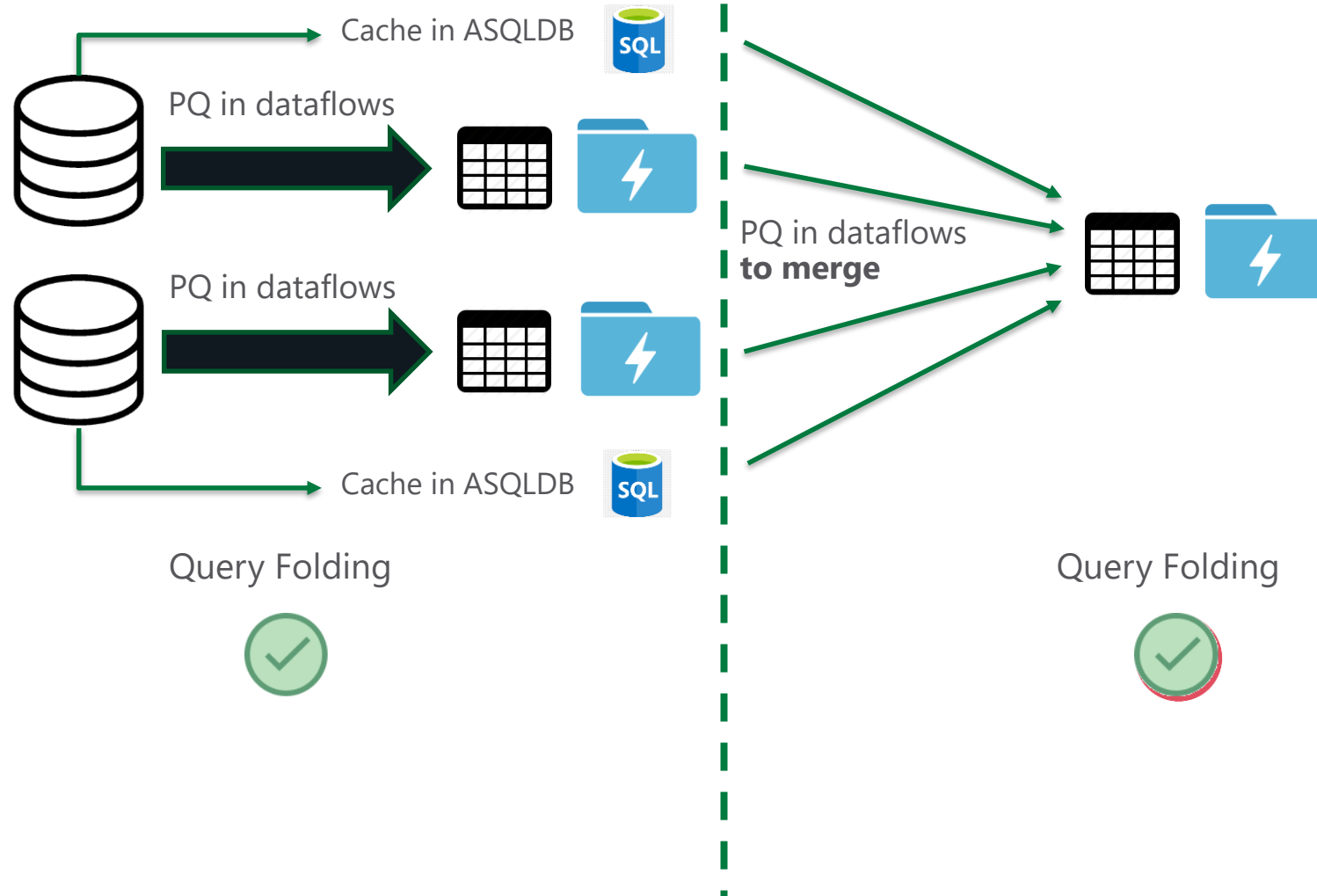
We've already talked about it

Do You remember?

\$\$\$\$: Enhanced compute engine, why?



\$\$\$\$: Enhanced compute engine, what?



\$\$\$\$: Enhanced compute engine, how?

DATAFLOWS - Active
Your workload is ready to use.

☒ On

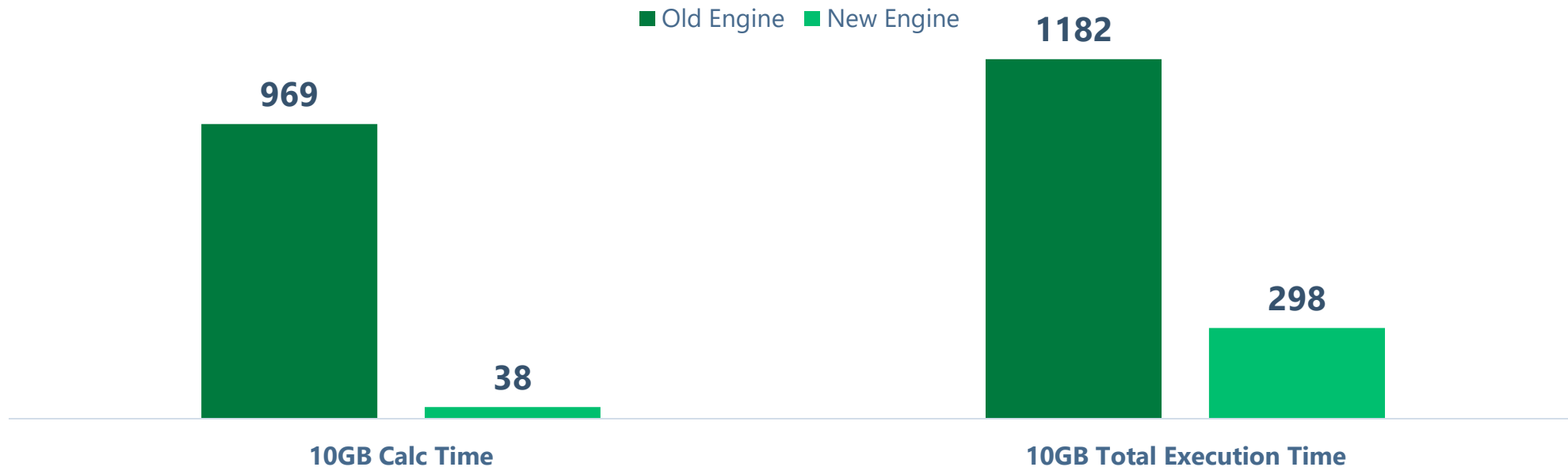
Max Memory (%)

Enhanced Dataflows Compute Engine (Preview)

☒ On

Container Size (Mb)

\$\$\$\$: Enhanced compute engine, so?



Up to 25X times faster

****Note: loading data is slower as the data is being written twice***

\$\$\$\$: AI Insights



Sentiment Analysis
Key Phrase Extraction
Language Detection
Image Tagging



Call custom **Azure ML models**
adding insights to your data

\$\$\$\$: Incremental Refresh

Incremental refresh settings

Appointment

Incremental refresh updates only data that's changed, to speed refresh, reduce capacity usage, and store historic data. [Learn more](#)

☒ On

Choose a DateTime field to filter by

createdon

Store rows from the past

5

Years

Refresh rows from the past

10

Days

☐ Detect data changes [Learn more](#)

Only refresh data if the maximum value in this field changes

Choose a field

☐ Only refresh complete [Learn more](#)

When you save these settings, data from the past [storage period] will be loaded to your dataflow storage the next time this dataflow is refreshed. Subsequent refreshes will update only data that's changed in the past [refresh period].

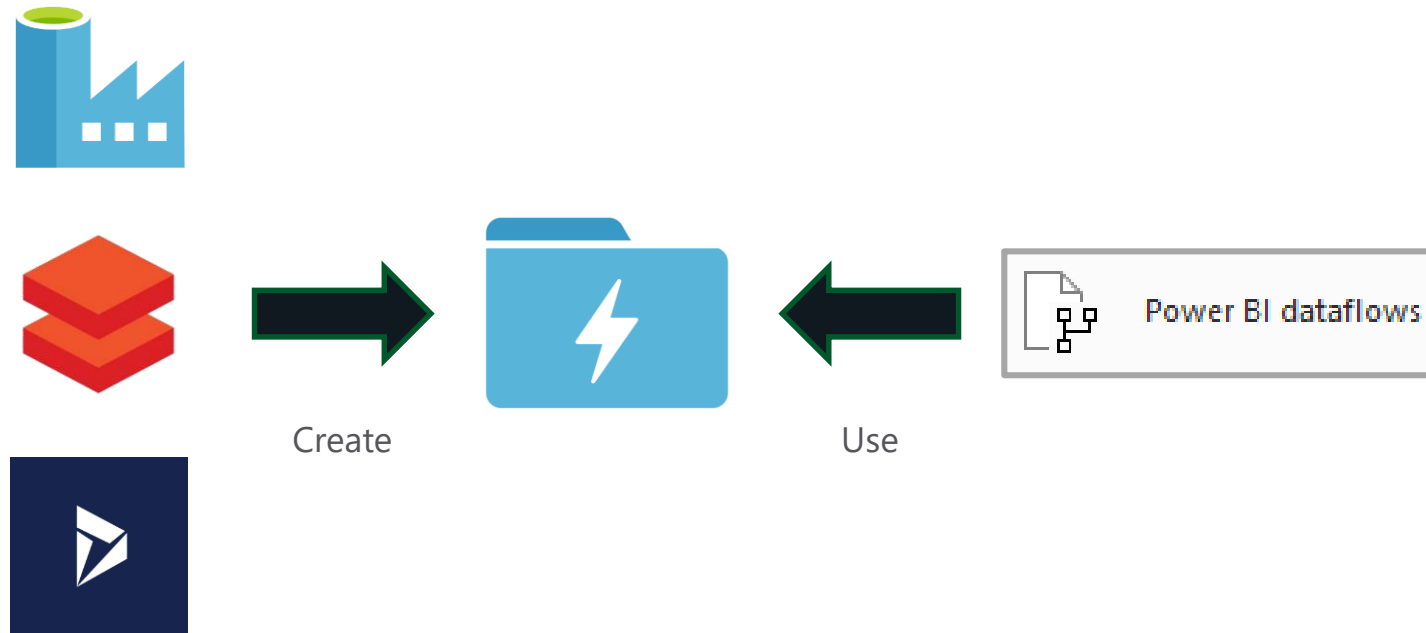
Save

Cancel

A large, teal-colored abstract graphic on the left side of the slide. It consists of several overlapping, curved, and elongated shapes that create a sense of depth and movement, resembling a stylized 'S' or a series of connected loops.

Use your Azure Data Lake
Storage gen2

Use your Azure Data Lake Storage gen2 (preview)



A large, stylized teal graphic on the left side of the slide, resembling a thick, curved arrow or a stylized letter 'C' that points towards the right.

Best Practices, Tips & Tricks

Best Practices

Split **ingestion** and **transformation** (\$\$\$\$)

Create a **separate workspace for ingestion and transformation** (\$\$\$\$)

Do not break **query folding** (\$\$\$\$)

Avoid scheduling refresh for linked entities inside the same workspace (\$\$\$\$)

Separate **on-premises** sources from **cloud sources** into different dataflows

Tips & Tricks

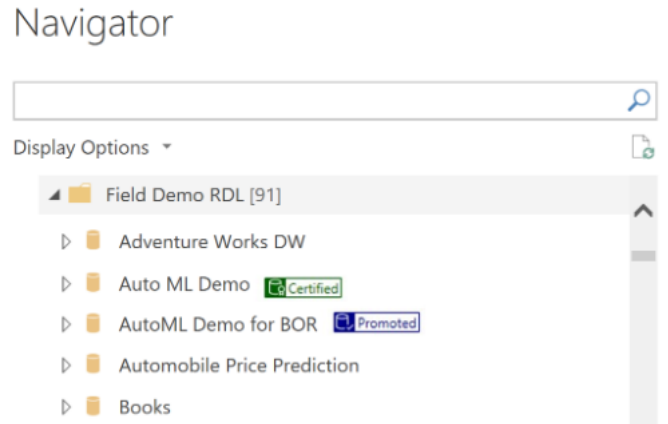
Power BI Gateway is supported for On-Prem sources

All queries with the **same Gateway**

Use "**Blank Query**" source for missing transformation Uls

Breaking News

- Endorcements



- Direct Query support (\$\$\$\$)
- Detach your ADLS gen2

Useful Links

dataflows & CDM: <http://tiny.cc/5d9uaz>

Dataflows best practices: <http://tiny.cc/dmgebi>

Shared Datasets replace SSAS: <http://tiny.cc/8imtbz>

Datasets & dataflows, my thoughts: <http://tiny.cc/pbidfdd1>

Join #PUGMilan: <https://lnkd.in/gi4TPce>



Ricordatevi di
compilate il feedback
form 😊

<https://speakerscore.com/skkl>

Thank you

#SqlSat921





PASS