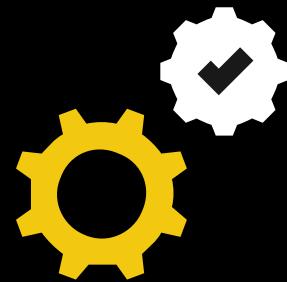


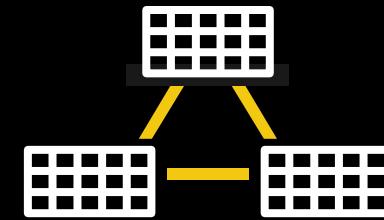
Gather



Clean



Model



Visualize



Data Modeling for Experts

It is time to take the next step with your data modeling skills!

Thank you, partners



KOHERA

element
experience & expertise
61

lytix

bmatix
Act informed

inetum.
realdolmen
Positive digital flow

datasense

MICROPOLIS
BELUX

LACO/



AKABI

Cloubis

datashift

EpicData.

Sparkle

Tabular Editor

solarwinds

u2u

ae
Adapt
and
Enable

MONIN
Database Managed Services

proximus NXT
tech. bizz. people.

tillit
data
shapers

ORDINA
Ahead of change



Marc Lelijveld

Technical Evangelist | Solution Architect
Macaw Netherlands



@MarcLelijveld



linkedin.com/in/MarcLelijveld



Data-Marc.com

FAVORITE STUFF:



After this session you will be able to:



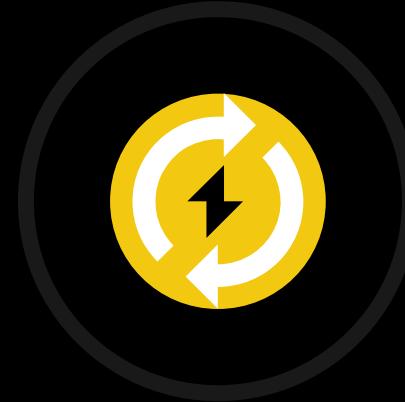
Successfully implement
calculation groups



Build and optimize
composite models



Combine DirectQuery
and import storage
modes in a single
data model



Optimize query
performance in
complex situations

What are we talking about?

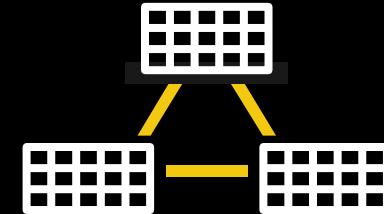
Gather



Clean



Model

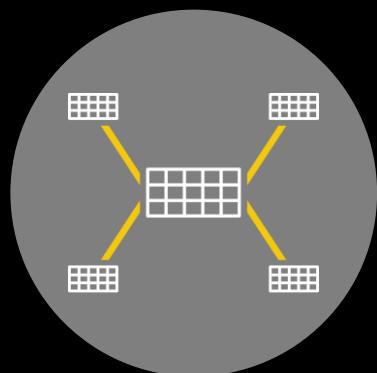
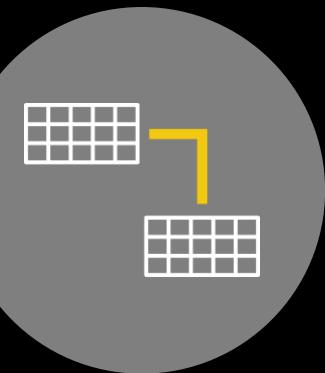


Visualize



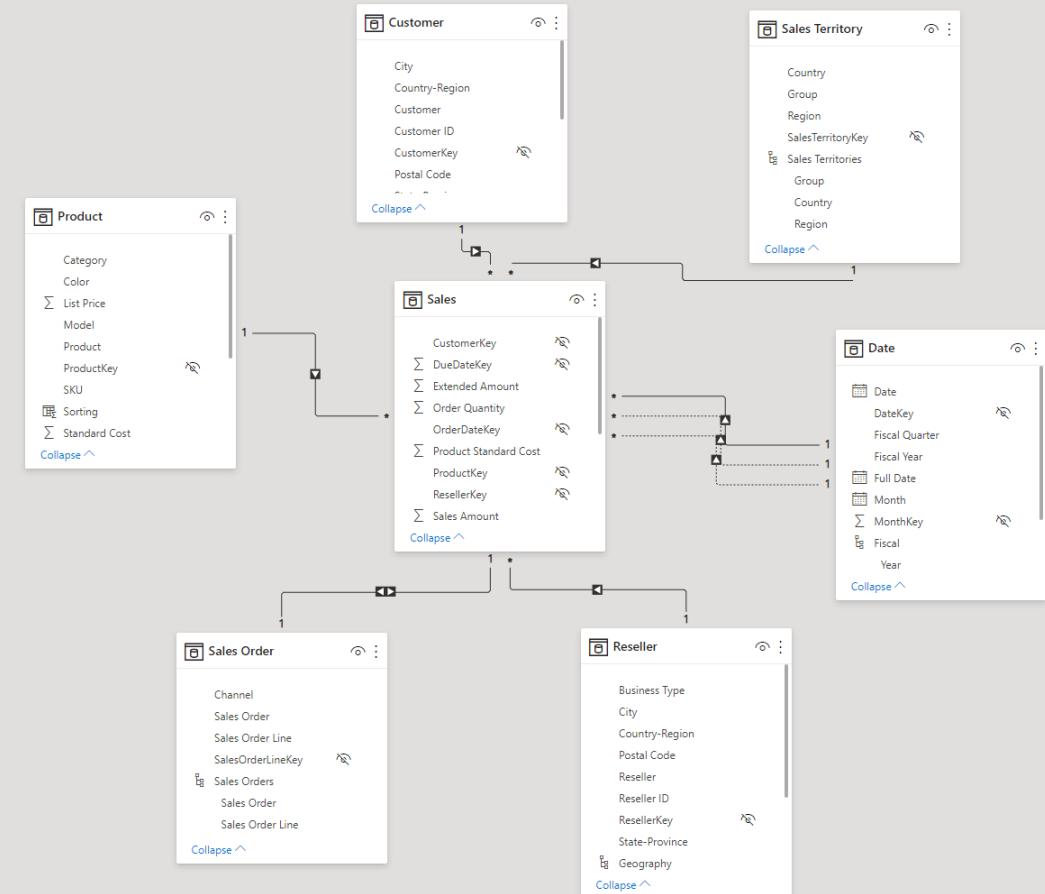
We assume you know...

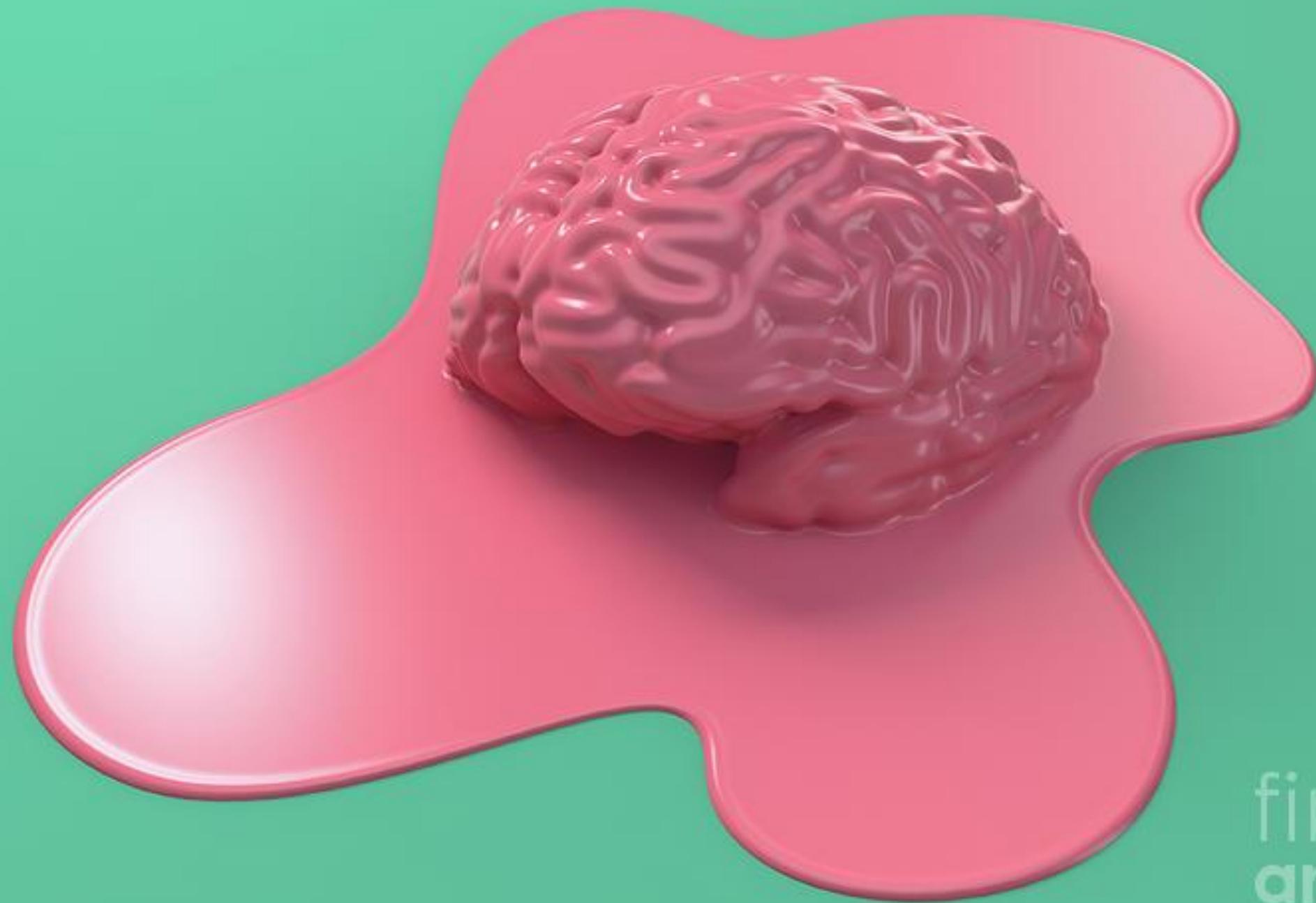
- The importance of a star schema
- Different data model types
- Modeling best practices, including naming conventions for tables and columns
- Auto date/time
- Relationship direction
- Limited and regular relationships
- Ambiguous data models
- Aggregations
- Dealing with multiple fact tables
- Different types of fact tables
- Slowly changing dimensions
- Data groups and hierarchies



AdventureWorks

Our data model





fineart
america

Storage modes

Different types of storage modes

Three storage modes

- **Import** – data cached in the model
- **DirectQuery** – queries are submitted to the back-end data source
- **Dual** – can act in both above storage modes, depending on query context

Configuring storage modes

- Storage modes are set on table level
- Setting storage mode to Import is an irreversible operation
- Data in DirectQuery mode cannot be displayed in the data tab

The screenshot shows the Power BI Desktop interface with the 'Data' tab selected. On the left, the Data Model view displays three tables: 'Product Subcategory', 'Product', and 'Internet Sales - Agg'. The 'Internet Sales - Agg' table is currently selected, as indicated by the blue border around its preview and the yellow selection bar on the far left.

The 'Properties' pane on the right provides configuration options for the selected table:

- Name:** Internet Sales
- Description:** Enter a description
- Synonyms:** Enter a comma-separated list of synonyms for Q&A
- Row label:** Select a row label
- Key column:** Select a column with unique values
- Is hidden:** No (radio button)
- Is featured table:** No (radio button)
- Advanced:** Storage mode (dropdown menu)

The 'Storage mode' dropdown in the 'Advanced' section is highlighted with a red box and set to 'DirectQuery'.

Understanding Dual mode

Example

- Internet Sales → DirectQuery
- Internet Sales - Agg → Import
- Product Category → Dual

Screenshot of Power BI Desktop showing the Data Model view. The ribbon is at the top, and the Data Model view shows four tables: Product Subcategory, Product, Internet Sales - Agg, and Internet Sales. The Internet Sales table is selected and highlighted with a yellow border. A red box highlights the 'Tables' icon in the ribbon's Home tab. A red box also highlights the 'Storage mode' dropdown in the Properties pane for the Internet Sales table, which is set to 'DirectQuery'. The Properties pane on the right contains fields for General, Name (Internet Sales), Description (Enter a description), Synonyms (Enter a comma-separated list of synonyms for Q&A), Row label (Select a row label), Key column (Select a column with unique values), Is hidden (No), Is featured table (No), and Advanced (Storage mode set to DirectQuery).

Benefits

Benefits to choose one or the other

- Improve query performance – cache data for faster end-user performance
- Data refresh optimization – no need to refresh for non-cached data
- Near-real time requirements – reduce query latency when in DirectQuery mode
- Large datasets – choose to not import certain data



Caches and DirectQuery

Risks of mixing storage modes

- Avoid mismatch in results when query bits DirectQuery compared to Import
- Data cached (import) could be behind compared to DirectQuery data
- Make sure cached data is kept **in sync** – regularly refresh!



Behavior of Dual mode

Query context defines storage mode used

- **Sales per product subcategory**, both tables will use Import mode
- **Sales per product**, both tables will hit DirectQuery
- **Extended Amount per product subcategory**, both tables will hit DirectQuery

The screenshot shows the Power BI Desktop interface with the Data Model view open. Three tables are present: Product Subcategory, Product, and Internet Sales - Agg. The Internet Sales - Agg table is highlighted with a yellow border. The Product table has its Storage mode set to DirectQuery in the Properties pane.

Product Subcategory table:

- English Product Subcategory N...
- French Product Subcategory Na...
- ProductCategoryKey
- Σ ProductSubcategoryAlt...
- ProductSubcategoryKey
- Spanish Product Subcategory N...

Product table:

- Arabic Description
- Chinese Description
- Class
- Color
- Σ Days To Manufacture
- Σ Dealer Price

Internet Sales - Agg table:

- Carrier Tracking Number
- Σ CurrencyKey
- Customer PO Number
- CustomerKey
- Σ Discount Amount
- DueDateKey
- Σ Extended Amount
- Σ Freight Amount
- Σ Order Calendar Year

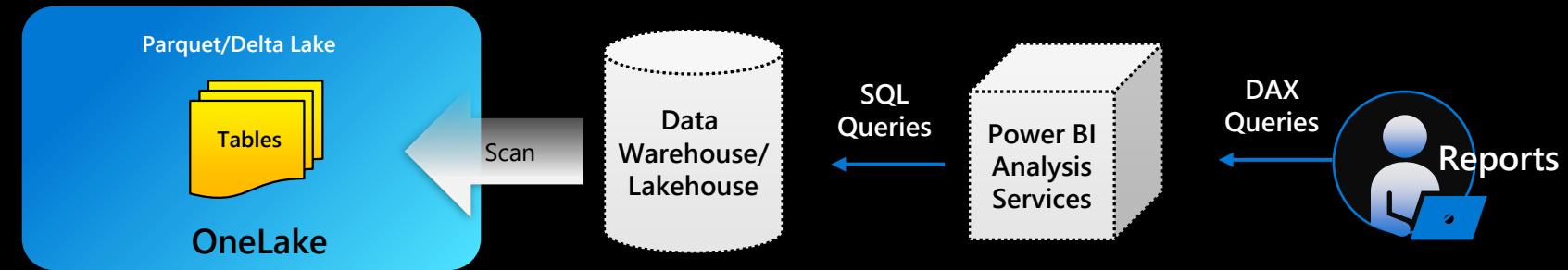
Properties pane (for Product table):

- Name: Internet Sales
- Description: Enter a description
- Synonyms: Enter a comma-separated list of synonyms for Q&A
- Row label: Select a row label
- Key column: Select a column with unique values
- Is hidden: No
- Is featured table: No
- Advanced settings
- Storage mode: DirectQuery

And then there is Direct Lake

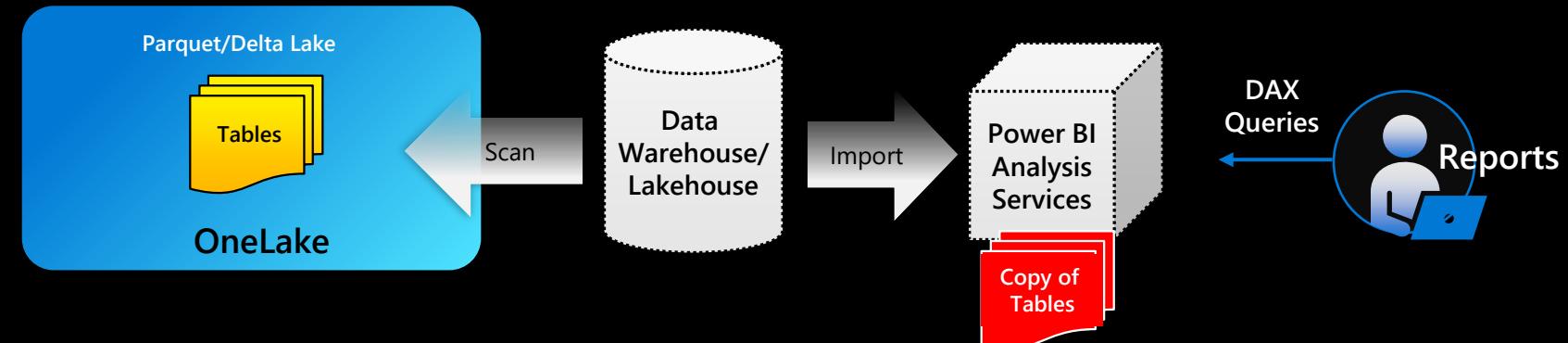
"Direct Query Mode"

Slow, but real time



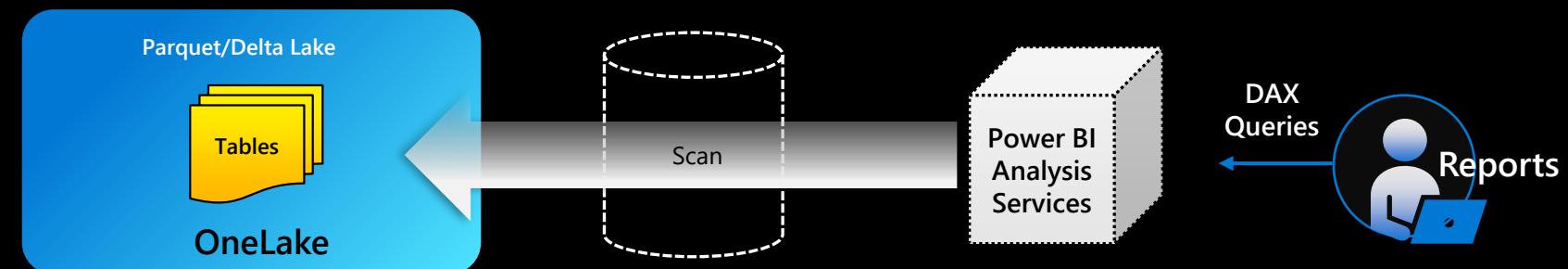
"Import Mode"

Latent & duplicative but fast



"Direct Lake Mode"

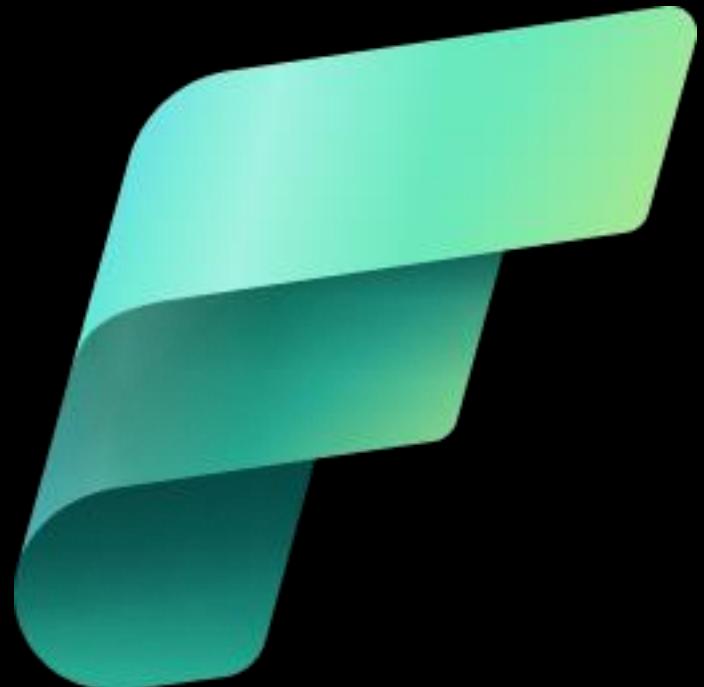
Best of both worlds



Direct Lake

Keep in mind that:

- No data is imported
- On-demand load and paging
- Reads data from Lake / Parquet format
- Only applicable when using MS Fabric
- Performance dependent on capacity size/utilization
- Falls back to DirectQuery when limitations are hit!



Composite models

Introducing composite models

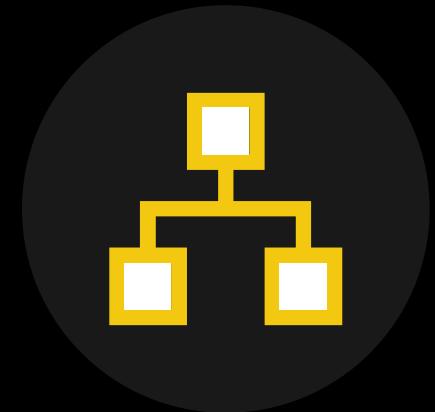
Composite models do either or both of the following actions

- Combine data from one or more DirectQuery sources
- Combines data from DirectQuery and import data

Three related features in Power BI

- Composite models, to allow different connections and storage models in a single data model
- Many-to-many relationships
- Storage modes

Limited relationships might occur, when relationships cross source group



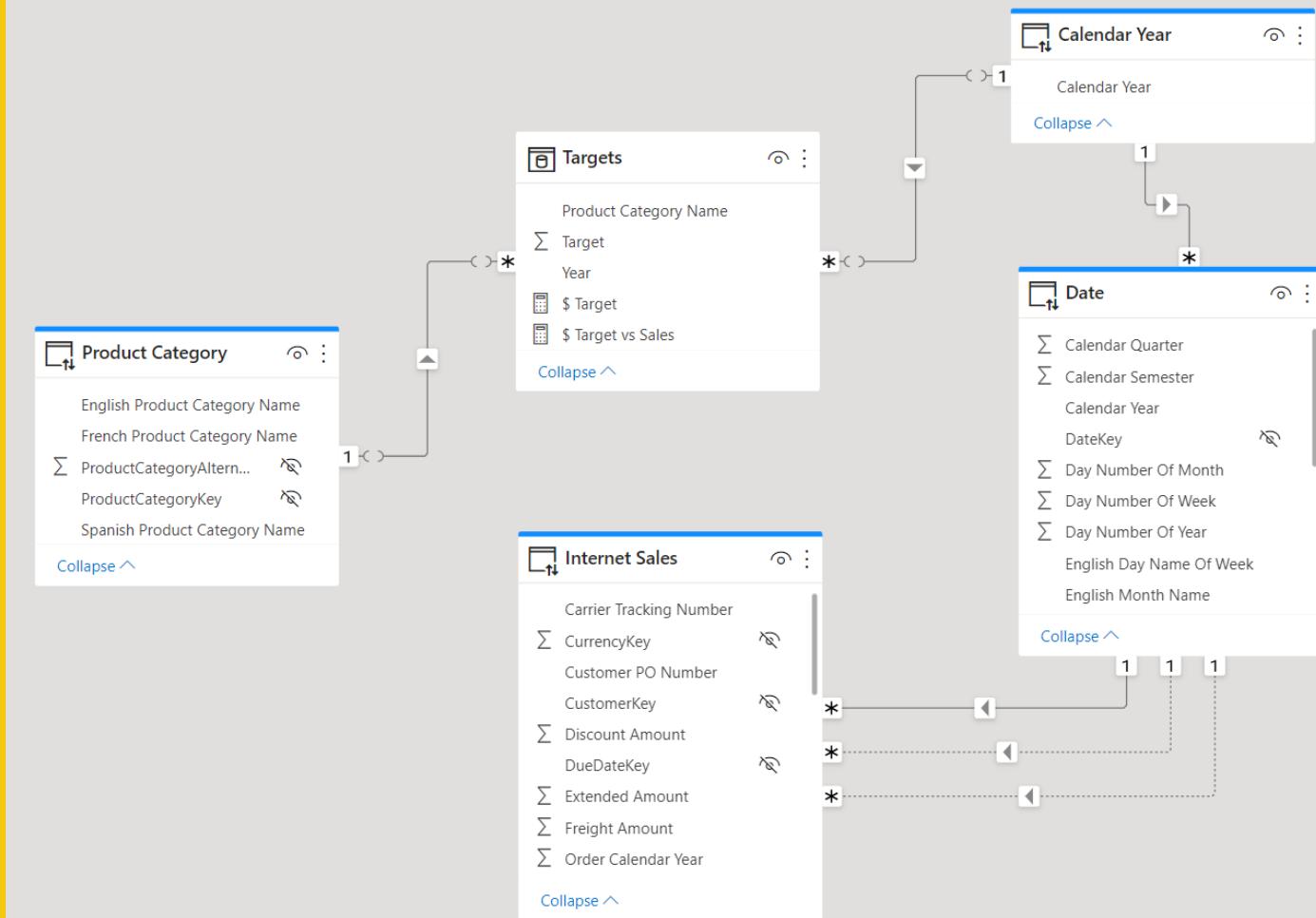
Example, sales model with local target sheet

Example use case for composite models in Power BI

- Centrally maintained data model, containing all sales
- Region has their own sheet with targets in Excel, which they can blend with the central data model

Source group – everything loaded from one DirectQuery source

All import tables and calculated tables are in one source group

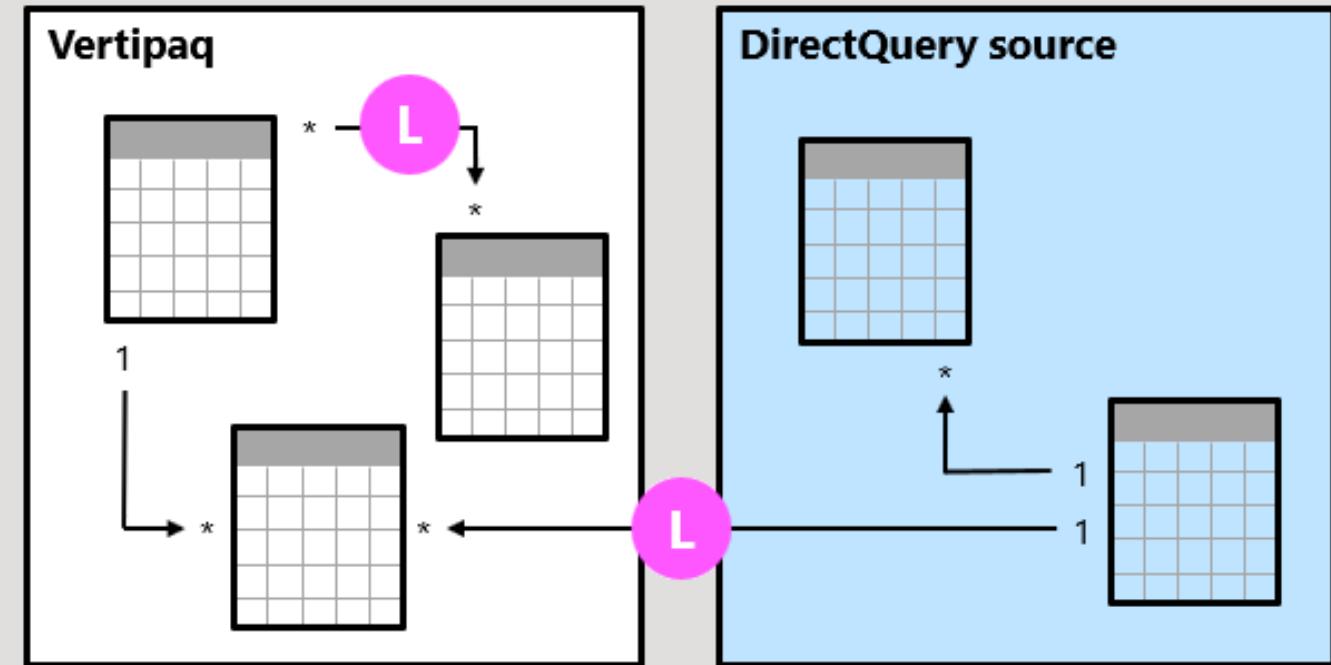


Limited relationships

A model relationship is *limited* when there's no guaranteed "one" side

It can be the case for two reasons:

- The relationship uses a Many-to-many cardinality type (even if one or both columns contain unique values)
- The relationship is cross source group (which can only ever be the case for Composite models)
- It can be a cross-storage mode but does not have to be



Composite models

Demo

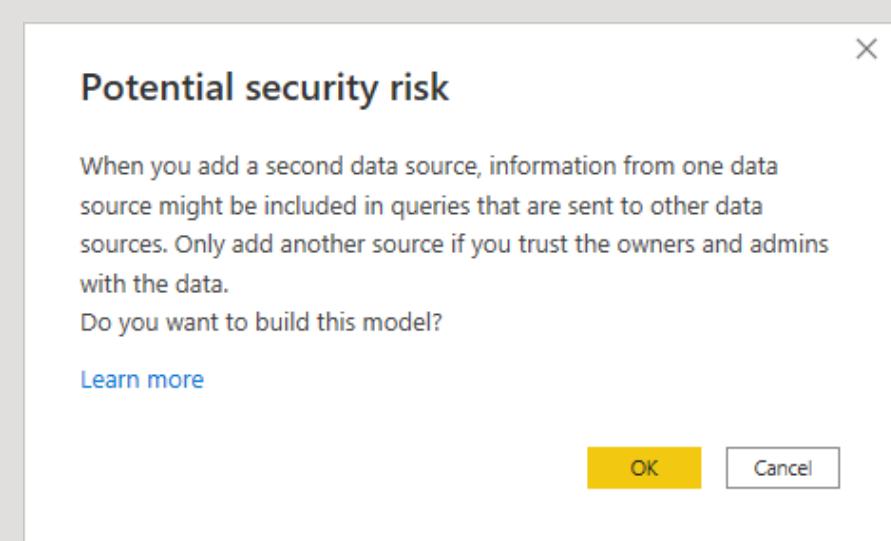
Security implications

Composite models do either or both of the following actions

- A query sent to one data source can include data values that have been retrieved from another data source

Potentially

- Any administrator of the data sources, who can view the traces or audit logs could view this information, without permissions to the original data source
- The encryption setting for each source should be considered, to avoid breaching encryption while the data is included in a query sent to another unencrypted source



Performance implications

Composite models might have slower performance

- **General limitations of DirectQuery** apply, make sure the back-end data source has sufficient resources to provide best performance as possible
- Query execution potentially needs to **cross sources**, where values from one source needs to be evaluated in another source (rich set of values in the WHERE clause)
- **Difference in granularity** of data across sources, where on-demand aggregation is required before displaying data in a visual
- **Multiple queries** must be executed to retrieve results (typically with DistinctCount)



Composite models using Power BI datasets and Analysis Services

With the ability to connect to Power BI datasets and Analysis Services models using DirectQuery (in preview since December 2020) you can now build a composite model using these sources.

Note the following design goals / non-goals for this feature:

Goals:

- **Extend:** Add data to enterprise semantic model
- **Enrich:** Add measure, calculation, formatting to enrich enterprise semantic model without adding any data
- **Overview:** disjoined datasets to give business overview ('dashboard')

Non goals:

- Combine two or more sources by joining large shared dimension tables
- Combine large dimension table and large fact table from different sources

Composite models 2

Demo

Calculation groups

Introducing calculation groups

Benefits

- Reduce the number of redundant measures and grouping common measure expressions as calculation items
- Avoids duplicating logic in different measures
- Typical use cases are
 - Time-intelligence calculations (YTD / QTD / MTD / ...)
 - Format string change, like currency conversions



Limitations

- Can only be created from external tools in Power BI (Any tool using the XMLA endpoint such as Tabular Editor) – **but stay tuned....**
- Object level security on Calculation group items is not supported
- Smart narrative visuals in Power BI are not supported with Calculation Groups
- Be aware of potential unexpected behavior in combination with Composite Models

How calculation groups work

Specific DAX expressions for Calculation Groups

- SELECTEDMEASURE()
- SELECTEDMEASURENAME()
- ISSELECTEDMEASURE()
- SELECTEDMEASUREFORMATSTRING()

Classic measure:

```
MTD =  
CALCULATE (  
    SUM ( Sales[SalesAmount] ),  
    DATESMTD ( DimDate[Date] )  
)
```

Dynamic measure context MTD with Calculation Group:

```
MTD =  
CALCULATE (  
    SELECTEDMEASURE (),  
    DATESMTD ( DimDate[Date] )  
)
```

Create and use calculation groups

Demo

Wholesale and non-wholesale queries

Remote vs local model

Remote model – a model that serves as a data source for the current (local) model

- Remote / vs local table – a remote table is a table that is sourced from a remote model.
Any table that is not from a remote model is considered local
- Remote measures are treated as a black box to the composite model: a calc groups might
be applied, or not, depends on wholesale vs non-wholesale



Wholesale vs non-wholesale

- **Wholesale:** depends on objects in single remote source group
 - Calculated columns defined on a table from a remote source group must be wholesale
- **Non-wholesale:** depends on objects in more than one source group



Calculation groups and measures in a composite model

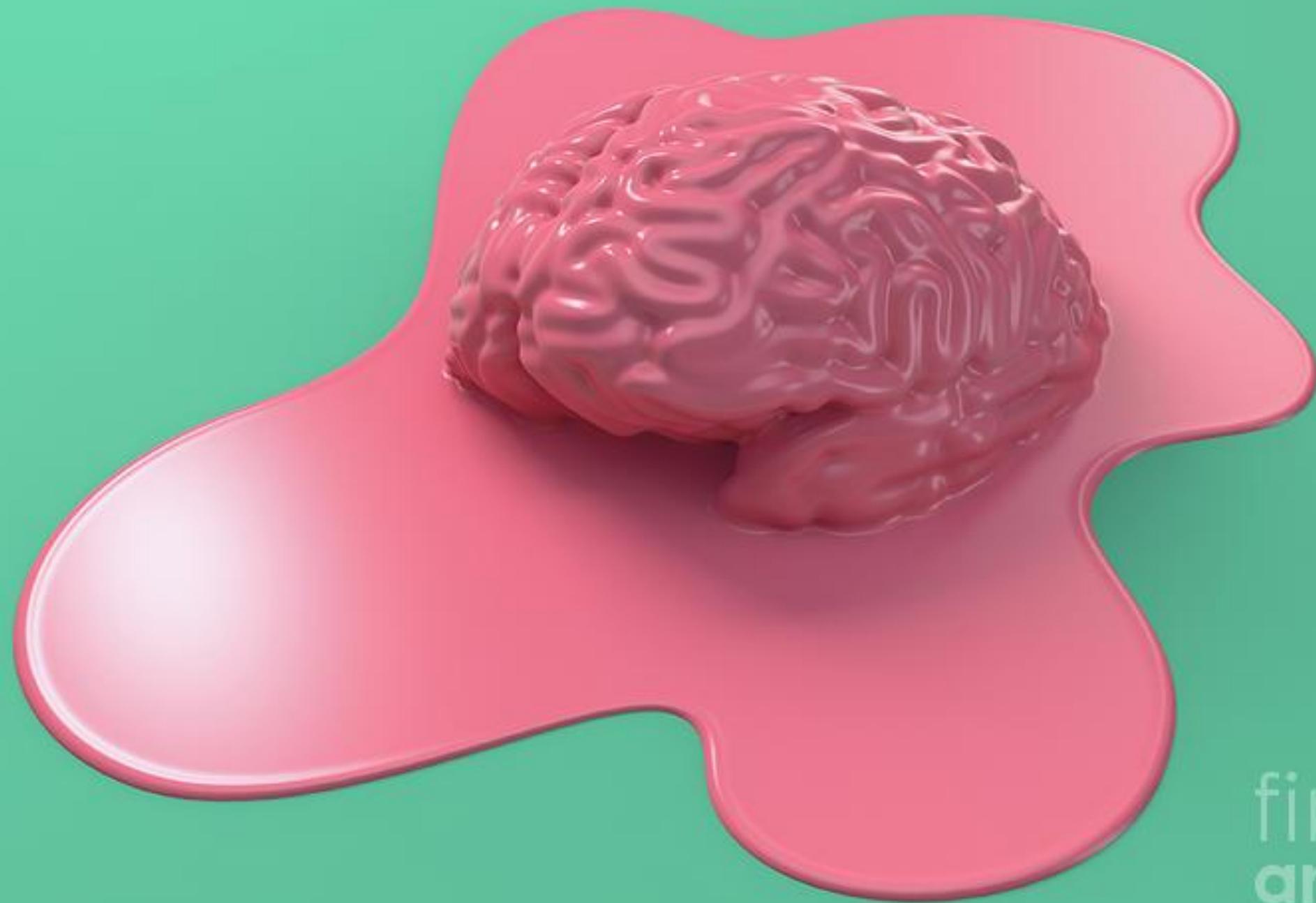
Depending on the configuration, calculation groups and measures have interesting interactions

| Measure location | Is measure wholesaleable | Remote calc group applied |
|------------------|---|----------------------------|
| Remote | Yes | Yes |
| Local | Yes | Yes |
| Local | No (either because it depends on more than one source group or just the local source group) | No – but can be indirectly |

Remember: measures might be impacted by remote calculation groups, when applied

Calculation groups and wholesale / non-wholesale measures

Demo



fineart
america

Key take-aways

Consider picking the right storage mode for your need

Be aware of the security and performance implications of composite models

Reduce the number of redundant measures and avoid duplication calculation logic by successfully leveraging calculation groups

When working with composite models using Power BI datasets or Azure Analysis Services and calculation groups, keep in mind wholesale vs non-wholesale and the interaction with calculation groups



Resources

Calculation groups

aka.ms/PowerBICalculationGroups

Different storage modes in Power BI

aka.ms/PowerBIStorageModes

Using Direct Query in Power BI

aka.ms/DirectQueryPBI

Guidance for Composite Models

aka.ms/CompositeModelsPBI

Relationship evaluation

aka.ms/DirectQueryRealtionshipEvalPBI

DirectQuery for Power BI datasets and Azure Analysis Services (preview)

aka.ms/PBIDirectQueryDatasets

Model relationships in Power BI

aka.ms/ModelRelationshipsPBI

Want to review all this content?

Webinar series: Mastering data modeling with Power BI

Episode 1 - Data modeling 101: Increasing the impact of Power BI

by Jeroen ter Heerdt, Microsoft and Marc Lelijveld, Macaw Netherlands

[Watch now ↗](#)

Webinar series: Mastering data modeling with Power BI

Episode 2 - Learn advanced data modeling with Power BI

by Jeroen ter Heerdt, Microsoft and Marc Lelijveld, Macaw Netherlands

[Watch now ↗](#)

Webinar series: Mastering data modeling with Power BI

Episode 3 - Data modeling for experts with Power BI

by Jeroen ter Heerdt, Microsoft and Marc Lelijveld, Macaw Netherlands

[Watch now ↗](#)

Webinar series: Data modeling with Power BI

Episode 4 - Calculation groups and composite models

By Jeroen ter Heerdt, Microsoft and Marc Lelijveld, Macaw Netherlands

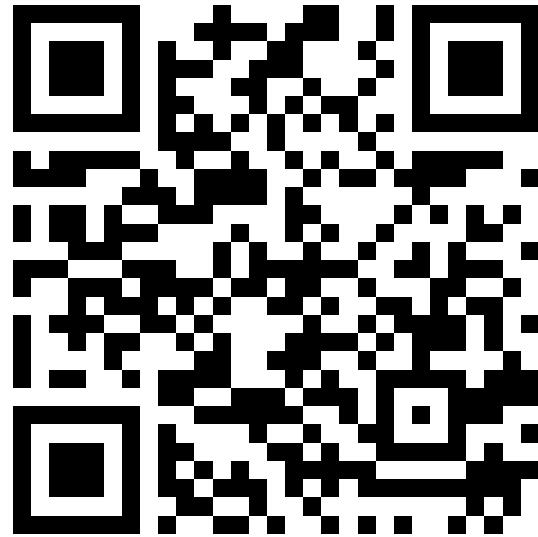
[Watch now ↗](#)

Find them all here:

docs.microsoft.com/en-us/power-bi/fundamentals/webinars



Thank you



https://bit.ly/dMC2023_SessionFeedback