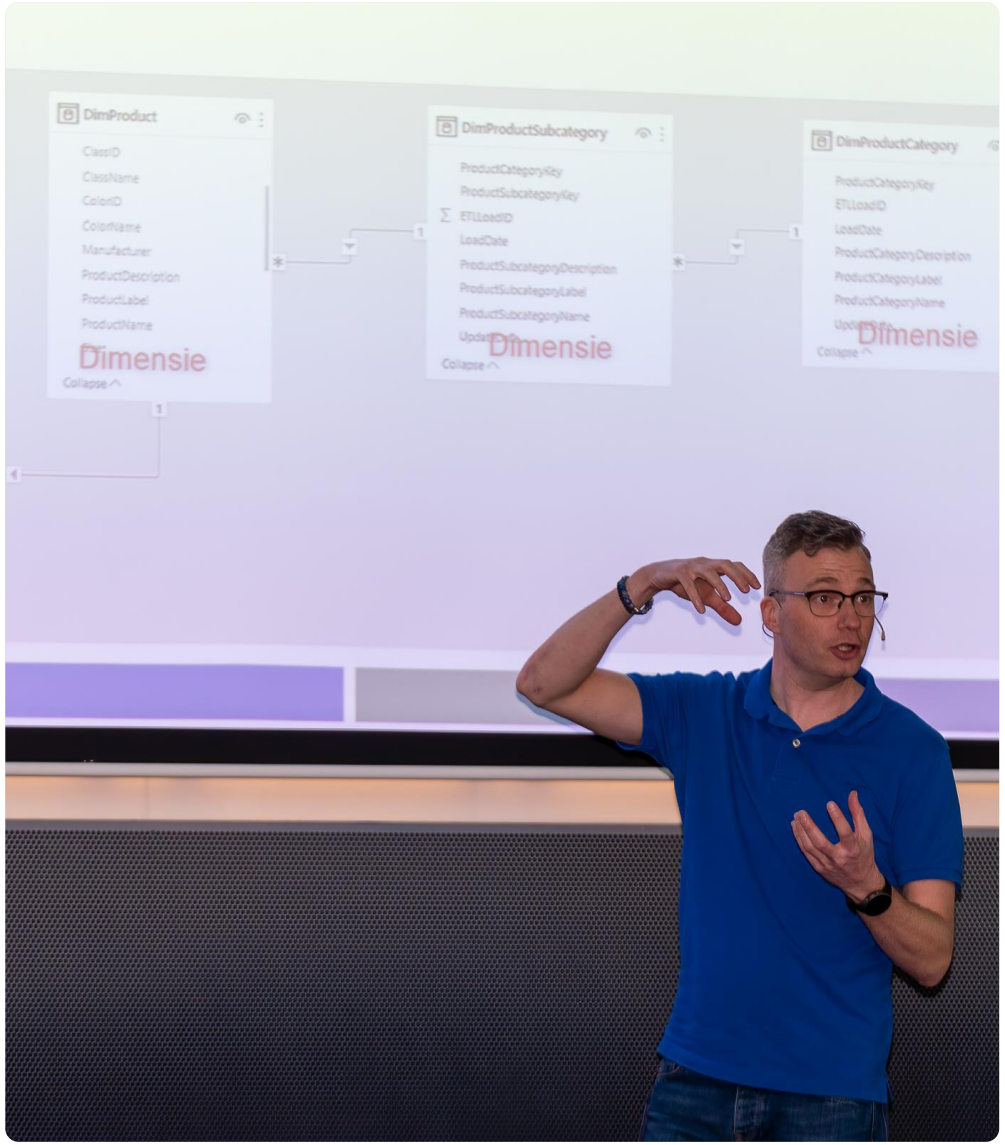
dutchBIgeek

Data modeling

08-10-2023



Thank you, partners

element 61
experience & expertise

Cloubis

Microsoft

MICROPOLE BELUX

PLAINSIGHT

Adapt and Enable

Tabular Editor

delaware

datashift

u2u

cegeka

EpicData.

LACO/

tillit data shapers

lytix

bmatix
Act informed

randstad digital

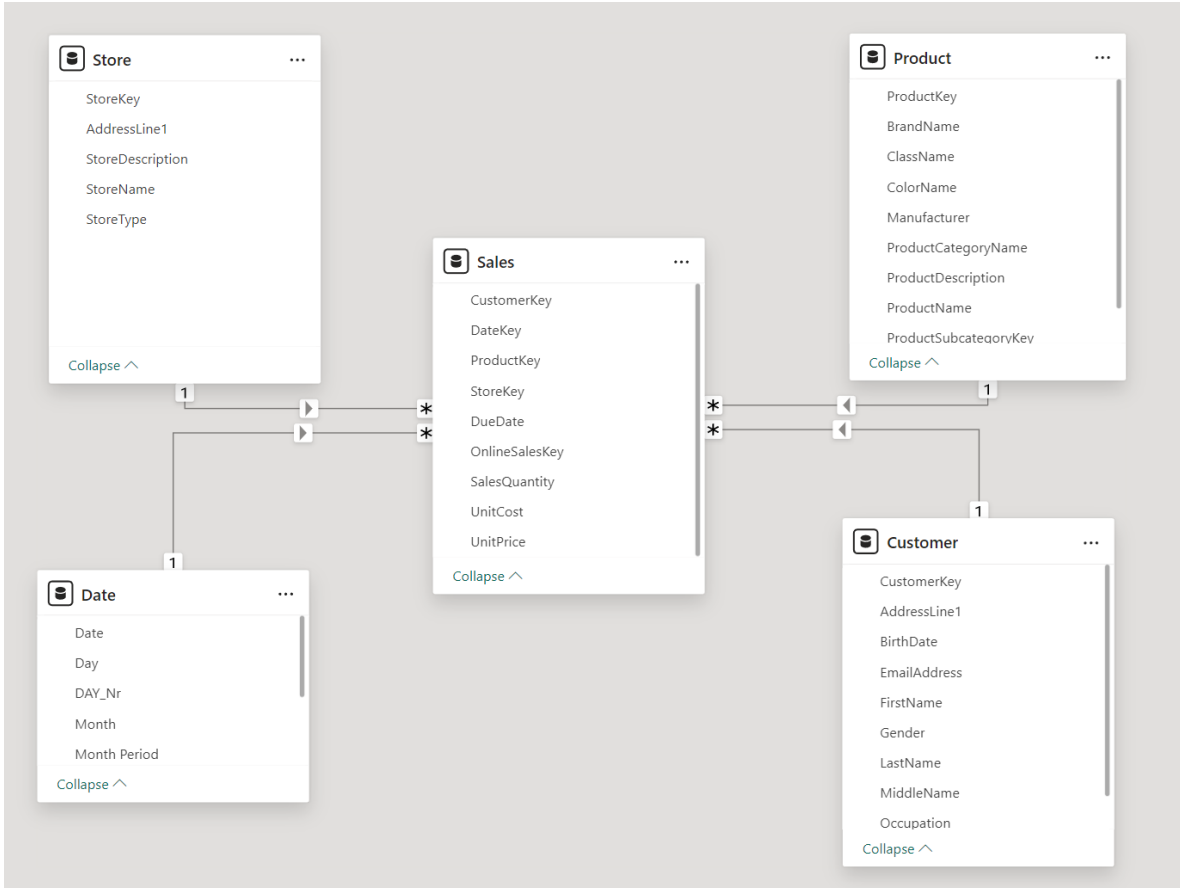
KOHERA

MONIN
Database Managed Services

inetum realdolmen
Positive digital flow

TriFinance
BEYOND ADVISORY

Data Modeling



Why is the Data Model Important?

Data Modeling

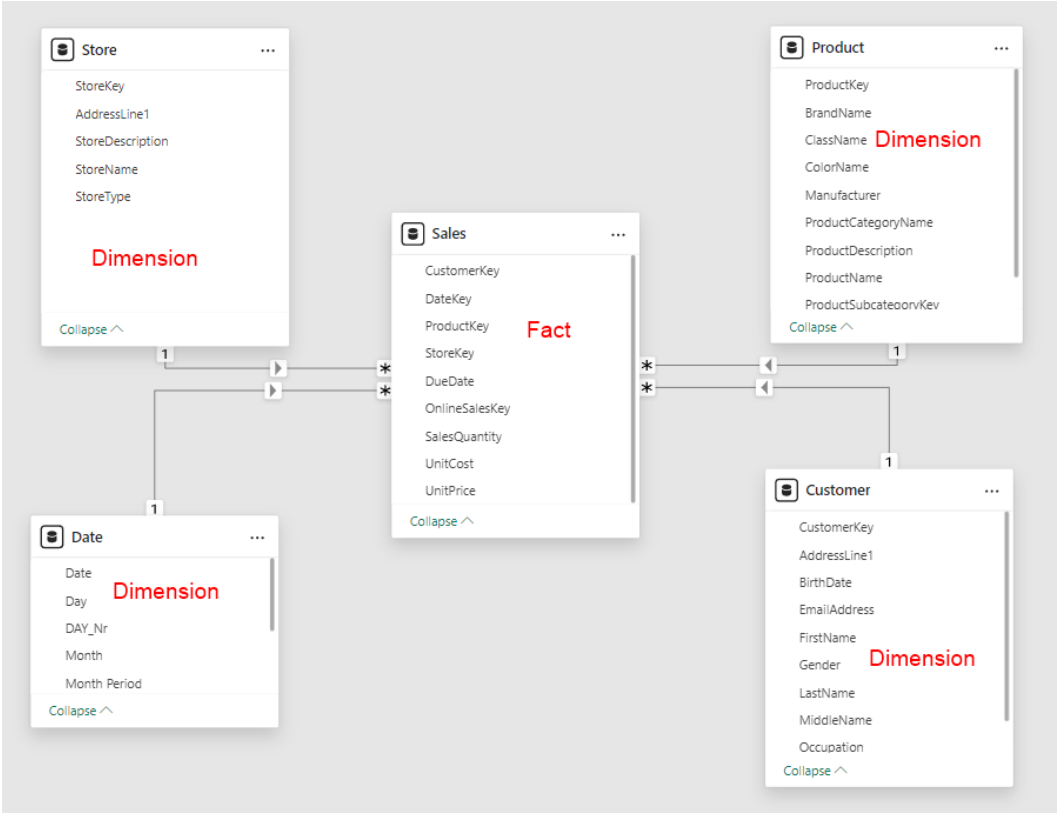
- The data model is the foundation of Power BI.
- A well-designed data model improves performance and speeds up queries.
- 80% of performance issues are related to the data model.
- A better data model makes DAX queries easier and more efficient.
- Reduces data redundancy and minimizes errors.

Star Schema

Data Modeling

- **Fact Tables:**
  - Contain facts, representing an event with dimensions.
  - A sale includes a product, a customer, and a date.
  - Metrics that can be aggregated to gain insights.

- **Dimension Tables:**
  - Descriptive attributes of entities such as a product, customer, employee, or patient.
  - Dimensions have attributes like color, category, manufacturer, or price.



dutchBIgeek

Snowflake-model

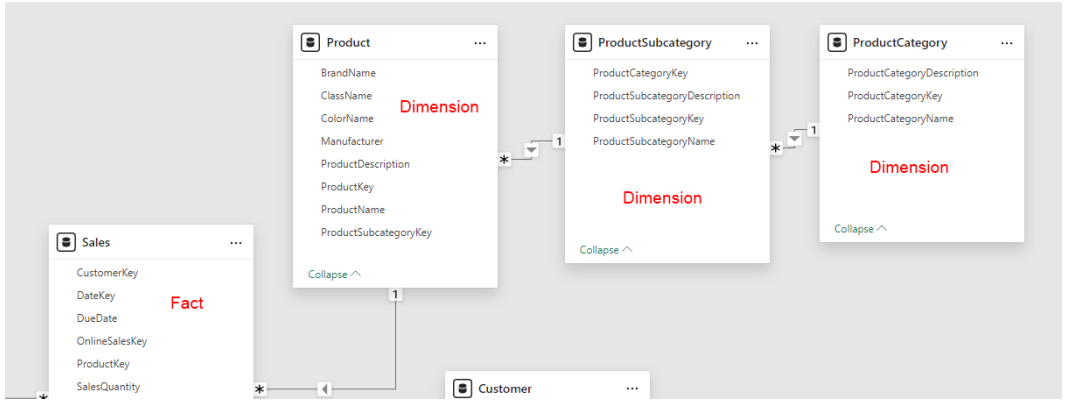
Data Modeling

The Snowflake Schema is a Variant of the Star Schema

The difference is that dimension tables that are related to each other are connected.

For example:

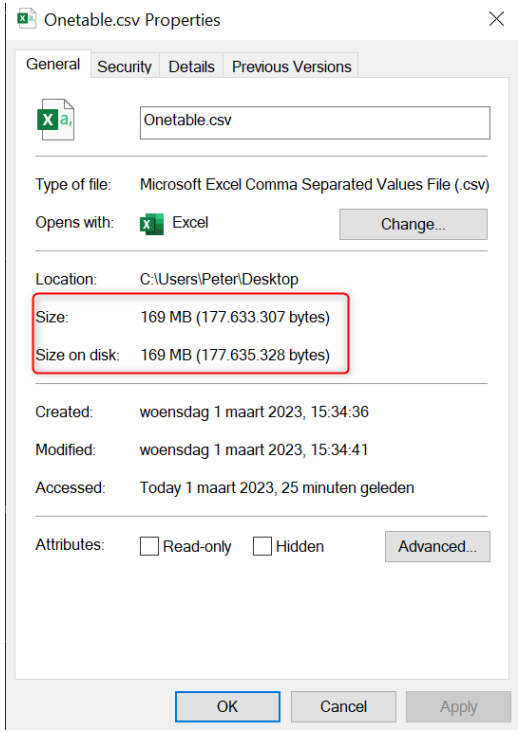
- Product
- Product Subcategory
- Product Category

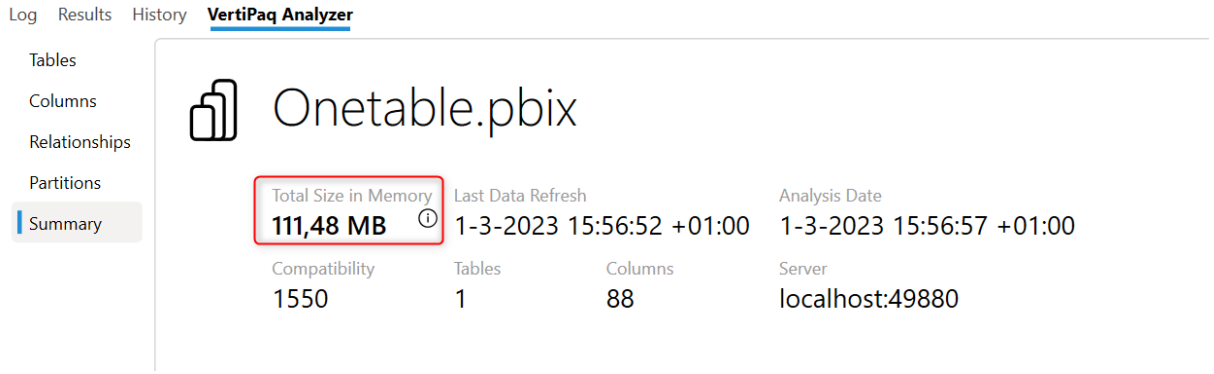


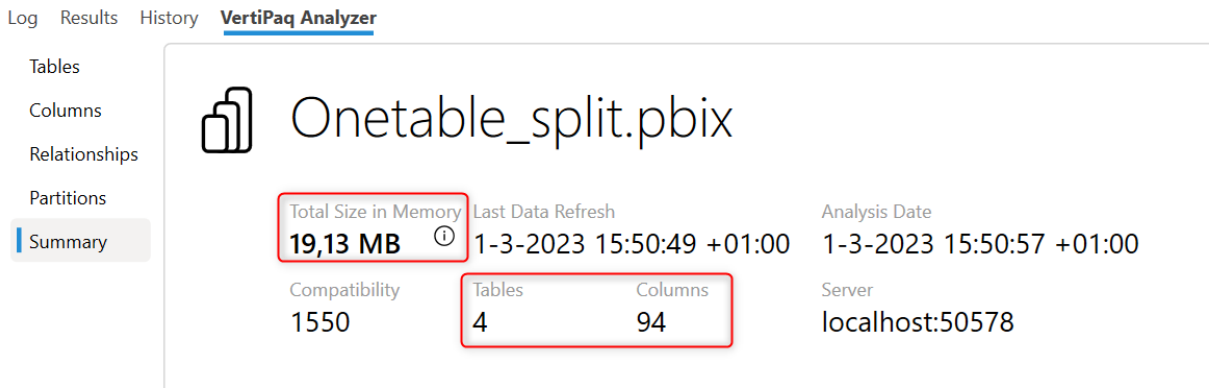
dutchBIgeek

Example of Why the Data Model is Important.

Data Modeling







dutchBIgeek

Example of Why the Data Model is Important.

Data Modeling

CustomerType	CompanyName	GeographyKey	StoreManager	StoreType	StoreName	StoreDescription	Status	OpenDate	CloseDate	EntityKey	ZipCode	ZipCodeExtension	StorePhone	StoreFax	
Company	Sapporo Company	710	292	Online	Contoso Asia Online Store	Contoso Asia Online Store	On	2004-10-11 00:00:00.000	NULL	946	10093	10093	55-555-0117	55-555-0117	China B
Company	LisbonCompany	586	246	Online	Contoso Europe Online Store	Contoso Europe Online Store	On	2004-09-03 00:00:00.000	NULL	945	10178	10178	731-555-0117	731-555-0117	Downtc
Company	Green BayCompany	800	212	Online	Contoso North America Online Store	Contoso Online Store	On	2004-08-25 00:00:00.000	NULL	59	20817	20817	450-555-0152	450-555-0152	Marbur
Company	DublinCompany	586	246	Online	Contoso Europe Online Store	Contoso Europe Online Store	On	2004-09-03 00:00:00.000	NULL	945	10178	10178	731-555-0117	731-555-0117	Downtc
Company	RidgelyCompany	800	212	Online	Contoso North America Online Store	Contoso Online Store	On	2004-08-25 00:00:00.000	NULL	59	20817	20817	450-555-0152	450-555-0152	Marbur
Company	LakelandCompany	800	212	Online	Contoso North America Online Store	Contoso Online Store	On	2004-08-25 00:00:00.000	NULL	59	20817	20817	450-555-0152	450-555-0152	Marbur
Company	RichardsonCompany	800	212	Online	Contoso North America Online Store	Contoso Online Store	On	2004-08-25 00:00:00.000	NULL	59	20817	20817	450-555-0152	450-555-0152	Marbur
Company	GlasgowCompany	586	246	Online	Contoso Europe Online Store	Contoso Europe Online Store	On	2004-09-03 00:00:00.000	NULL	945	10178	10178	731-555-0117	731-555-0117	Downtc
Company	GoulburnCompany	710	292	Online	Contoso Asia Online Store	Contoso Asia Online Store	On	2004-10-11 00:00:00.000	NULL	946	10093	10093	55-555-0117	55-555-0117	China B
Company	Nagoya Company	710	292	Online	Contoso Asia Online Store	Contoso Asia Online Store	On	2004-10-11 00:00:00.000	NULL	946	10093	10093	55-555-0117	55-555-0117	China B
Company	DublinCompany	586	246	Online	Contoso Europe Online Store	Contoso Europe Online Store	On	2004-09-03 00:00:00.000	NULL	945	10178	10178	731-555-0117	731-555-0117	Downtc
Company	TaipeiCompany	710	292	Online	Contoso Asia Online Store	Contoso Asia Online Store	On	2004-10-11 00:00:00.000	NULL	946	10093	10093	55-555-0117	55-555-0117	China B
Company	ShanghaiCompany	710	292	Online	Contoso Asia Online Store	Contoso Asia Online Store	On	2004-10-11 00:00:00.000	NULL	946	10093	10093	55-555-0117	55-555-0117	China B
Company	EdinburghCompany	586	246	Online	Contoso Europe Online Store	Contoso Europe Online Store	On	2004-09-03 00:00:00.000	NULL	945	10178	10178	731-555-0117	731-555-0117	Downtc
Company	JacksonvilleCompany	800	212	Online	Contoso North America Online Store	Contoso Online Store	On	2004-08-25 00:00:00.000	NULL	59	20817	20817	450-555-0152	450-555-0152	Marbur
Company	AppletonCompany	800	212	Online	Contoso North America Online Store	Contoso Online Store	On	2004-08-25 00:00:00.000	NULL	59	20817	20817	450-555-0152	450-555-0152	Marbur
Company	Citrus HeightsCompa	800	212	Online	Contoso North America Online Store	Contoso Online Store	On	2004-08-25 00:00:00.000	NULL	59	20817	20817	450-555-0152	450-555-0152	Marbur
Company	JacksonvilleCompany	800	212	Online	Contoso North America Online Store	Contoso Online Store	On	2004-08-25 00:00:00.000	NULL	59	20817	20817	450-555-0152	450-555-0152	Marbur
Company	LewisvilleCompany	800	212	Online	Contoso North America Online Store	Contoso Online Store	On	2004-08-25 00:00:00.000	NULL	59	20817	20817	450-555-0152	450-555-0152	Marbur
Company	EdinburghCompany	586	246	Online	Contoso Europe Online Store	Contoso Europe Online Store	On	2004-09-03 00:00:00.000	NULL	945	10178	10178	731-555-0117	731-555-0117	Downtc
Company	BerlinCompany4	586	246	Online	Contoso Europe Online Store	Contoso Europe Online Store	On	2004-09-03 00:00:00.000	NULL	945	10178	10178	731-555-0117	731-555-0117	Downtc
Company	HoustonCompany	800	212	Online	Contoso North America Online Store	Contoso Online Store	On	2004-08-25 00:00:00.000	NULL	59	20817	20817	450-555-0152	450-555-0152	Marbur
Company	ShanghaiCompany	710	292	Online	Contoso Asia Online Store	Contoso Asia Online Store	On	2004-10-11 00:00:00.000	NULL	946	10093	10093	55-555-0117	55-555-0117	China B
Company	Mumbai Company	710	292	Online	Contoso Asia Online Store	Contoso Asia Online Store	On	2004-10-11 00:00:00.000	NULL	946	10093	10093	55-555-0117	55-555-0117	China B
Company	CologneCompany	586	246	Online	Contoso Europe Online Store	Contoso Europe Online Store	On	2004-09-03 00:00:00.000	NULL	945	10178	10178	731-555-0117	731-555-0117	Downtc
Company	BerlinCompany4	586	246	Online	Contoso Europe Online Store	Contoso Europe Online Store	On	2004-09-03 00:00:00.000	NULL	945	10178	10178	731-555-0117	731-555-0117	Downtc
Company	CorvallisCompany	800	212	Online	Contoso North America Online Store	Contoso Online Store	On	2004-08-25 00:00:00.000	NULL	59	20817	20817	450-555-0152	450-555-0152	Marbur
Company	EdinburghCompany	586	246	Online	Contoso Europe Online Store	Contoso Europe Online Store	On	2004-09-03 00:00:00.000	NULL	945	10178	10178	731-555-0117	731-555-0117	Downtc
Company	Lane CoveCompany	710	292	Online	Contoso Asia Online Store	Contoso Asia Online Store	On	2004-10-11 00:00:00.000	NULL	946	10093	10093	55-555-0117	55-555-0117	China B
Company	ShanghaiCompany	710	292	Online	Contoso Asia Online Store	Contoso Asia Online Store	On	2004-10-11 00:00:00.000	NULL	946	10093	10093	55-555-0117	55-555-0117	China B
Company	YorkCompany	586	246	Online	Contoso Europe Online Store	Contoso Europe Online Store	On	2004-09-03 00:00:00.000	NULL	945	10178	10178	731-555-0117	731-555-0117	Downtc
Company	Round RockCompany	800	212	Online	Contoso North America Online Store	Contoso Online Store	On	2004-08-25 00:00:00.000	NULL	59	20817	20817	450-555-0152	450-555-0152	Marbur

Example of an Optimized Data Model.

Data Modeling

Facts-table

OnlineSalesKey	DateKey	StoreKey	ProductKey	PromotionKey	CurrencyKey	CustomerKey	SalesOrderLineNumber	SalesQuantity	ReturnQuantity	ReturnAmount	DiscountQuantity	DiscountAmount	TotalCost	UnitCost	UnitPrice	DueDate
20473555	2019-08-14 00:00:00.000	199	1670	1	1	1123	1	1	0	0	0	0	4,13	4,13	8,99	2019-08-23 00:00:00.000
23353743	2020-07-23 00:00:00.000	199	1676	1	1	17042	1	1	0	0	0	0	4,13	4,13	8,99	2020-07-28 00:00:00.000
20706373	2019-09-12 00:00:00.000	199	1676	1	1	7817	1	1	0	0	0	0	4,13	4,13	8,99	2019-09-14 00:00:00.000
20676937	2019-09-09 00:00:00.000	199	1682	1	1	6510	1	1	0	0	0	0	4,13	4,13	8,99	2019-09-10 00:00:00.000
20538889	2019-08-22 00:00:00.000	199	1670	1	1	3107	1	1	0	0	0	0	4,13	4,13	8,99	2019-09-05 00:00:00.000
23812925	2020-09-05 00:00:00.000	199	1682	1	1	4888	1	1	0	0	0	0	4,13	4,13	8,99	2020-09-10 00:00:00.000
29910972	2022-01-22 00:00:00.000	199	1664	1	1	625	1	1	0	0	0	0	4,13	4,13	8,99	2022-01-28 00:00:00.000
20708187	2019-09-13 00:00:00.000	199	1676	1	1	7952	1	1	0	0	0	0	4,13	4,13	8,99	2019-09-26 00:00:00.000
20708865	2019-09-13 00:00:00.000	199	1670	1	1	8131	1	1	0	0	0	0	4,13	4,13	8,99	2019-09-17 00:00:00.000
20878449	2019-10-04 00:00:00.000	199	1676	1	1	10712	1	1	0	0	0	0	4,13	4,13	8,99	2019-10-12 00:00:00.000
29961007	2022-01-26 00:00:00.000	199	1676	1	1	1085	1	1	0	0	0	0	4,13	4,13	8,99	2022-02-09 00:00:00.000
25165468	2021-01-17 00:00:00.000	199	1676	1	1	1079	1	1	0	0	0	0	4,13	4,13	8,99	2021-01-19 00:00:00.000
28113041	2021-09-19 00:00:00.000	199	1682	1	1	13257	1	1	0	0	0	0	4,13	4,13	8,99	2021-10-01 00:00:00.000
23445712	2020-08-01 00:00:00.000	199	1682	1	1	18199	1	1	0	0	0	0	4,13	4,13	8,99	2020-08-02 00:00:00.000
23524547	2020-08-09 00:00:00.000	199	1676	1	1	1073	1	1	0	0	0	0	4,13	4,13	8,99	2020-08-23 00:00:00.000
20635474	2019-09-04 00:00:00.000	199	1664	1	1	5246	1	1	0	0	0	0	4,13	4,13	8,99	2019-09-15 00:00:00.000
20878393	2019-10-04 00:00:00.000	199	1676	1	1	10643	1	1	0	0	0	0	4,13	4,13	8,99	2019-10-05 00:00:00.000
21843045	2020-02-03 00:00:00.000	199	1664	1	1	10939	1	1	0	0	0	0	4,13	4,13	8,99	2020-02-12 00:00:00.000

Dimensions-table

StoreKey	GeographyKey	StoreManager	StoreType	StoreName	StoreDescription	Status	OpenDate	CloseDate	EntityKey	ZipCode	ZipCodeExtension	StorePhone	StoreFax	AddressLine1_s
307	710	292	Online	Contoso Asia Online Store	Contoso Asia Online Store	On	2004-10-11 00:00:00.000	NULL	946	10093	10093	55-555-0117	55-555-0117	China Beijing Chaoyang distrct Shanlitun Rd
306	586	246	Online	Contoso Europe Online Store	Contoso Europe Online Store	On	2004-09-03 00:00:00.000	NULL	945	10178	10178	731-555-0117	731-555-0117	Downtown Berlin, Germany
199	800	212	Online	Contoso North America Online Store	Contoso Online Store	On	2004-08-25 00:00:00.000	NULL	59	20817	20817	450-555-0152	450-555-0152	Marbury St Shopping mall

What Does an Optimized Data Model Consist Of?

Data Modeling

The model consists of facts and dimensions.

- **A fact table contains values that you can calculate, such as:**
  - Revenue, purchase date, sold products, etc.
- **A dimension table contains values you want to filter by, such as:**
  - Year, month, manufacturer, customer, etc.

Keys and Relationships

To create relationships between tables, keys are used.

- **Primary Key:**
  - A unique value that appears only once in a dimension table.

- **Foreign Key:**
  - Used in a fact table to indicate how often, for example, a product has been sold.
  - The foreign key can appear multiple times in a fact table, as a product is typically sold more than once.

Relationships and Filtering

A relationship must be created between facts and dimensions. Possible relationships include:

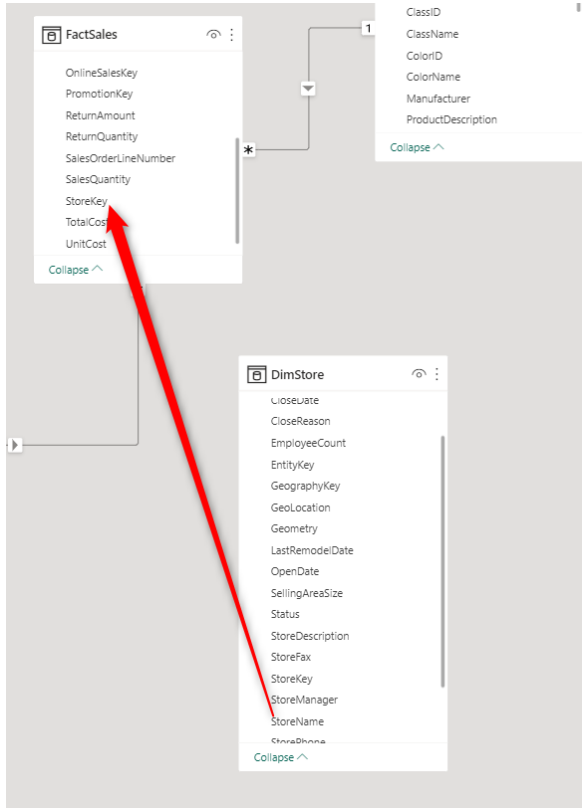
- One-to-many
- One-to-one
- Many-to-many

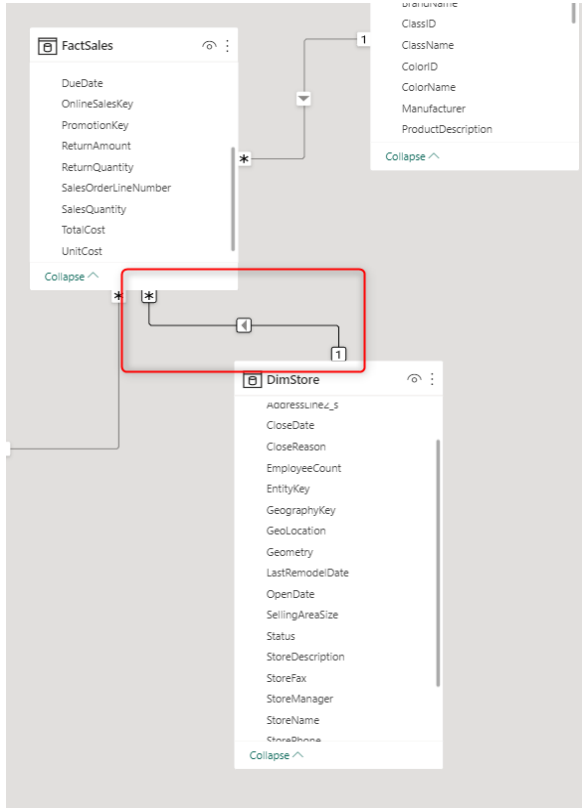
You can also choose the filter direction:

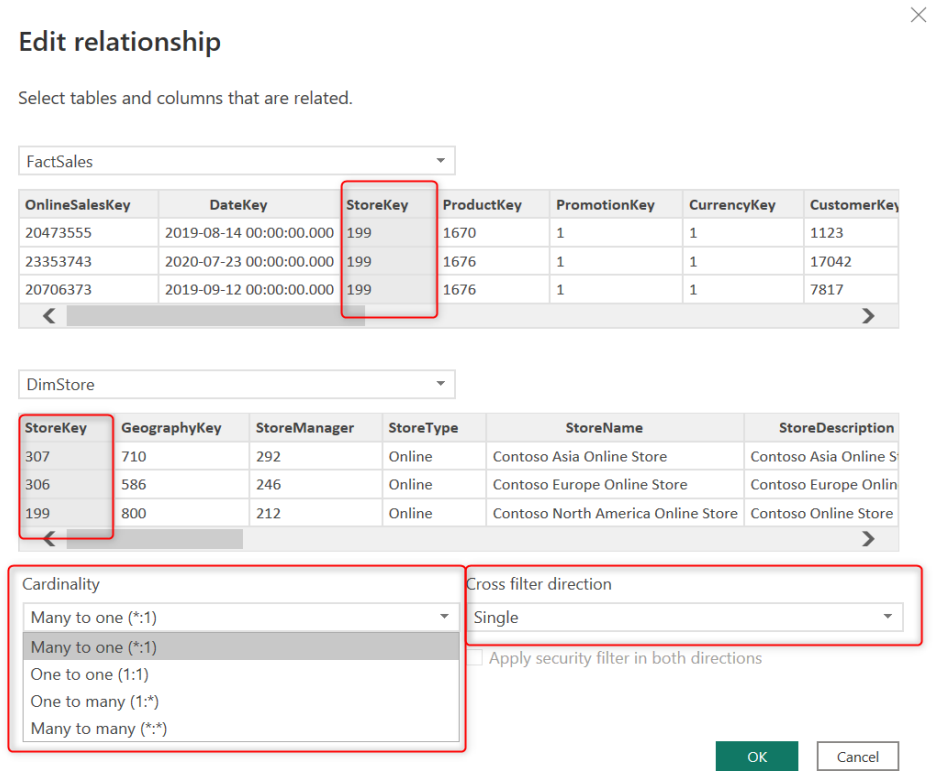
- Single
- Both

Example: Relationships and Filtering

Data Modeling







dutchBIgeek

Multiple Fact Tables

Data Modeling

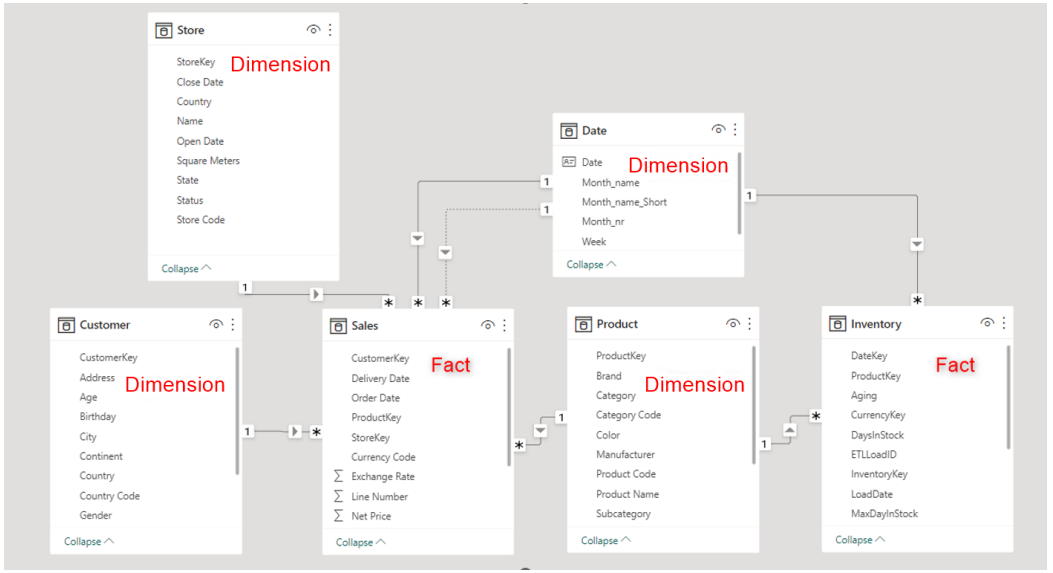
It is possible to use multiple fact tables in your model.

These fact tables may not have much in common with each other

- Inventory and sales.

However, it is necessary to have some dimension tables that are connected to the fact tables

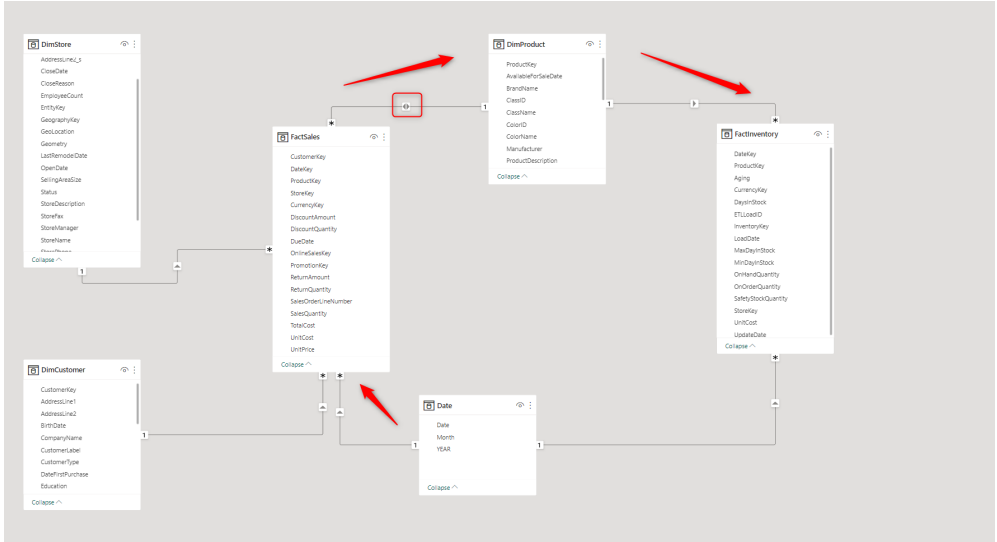
- date and product.



dutchBIgeek

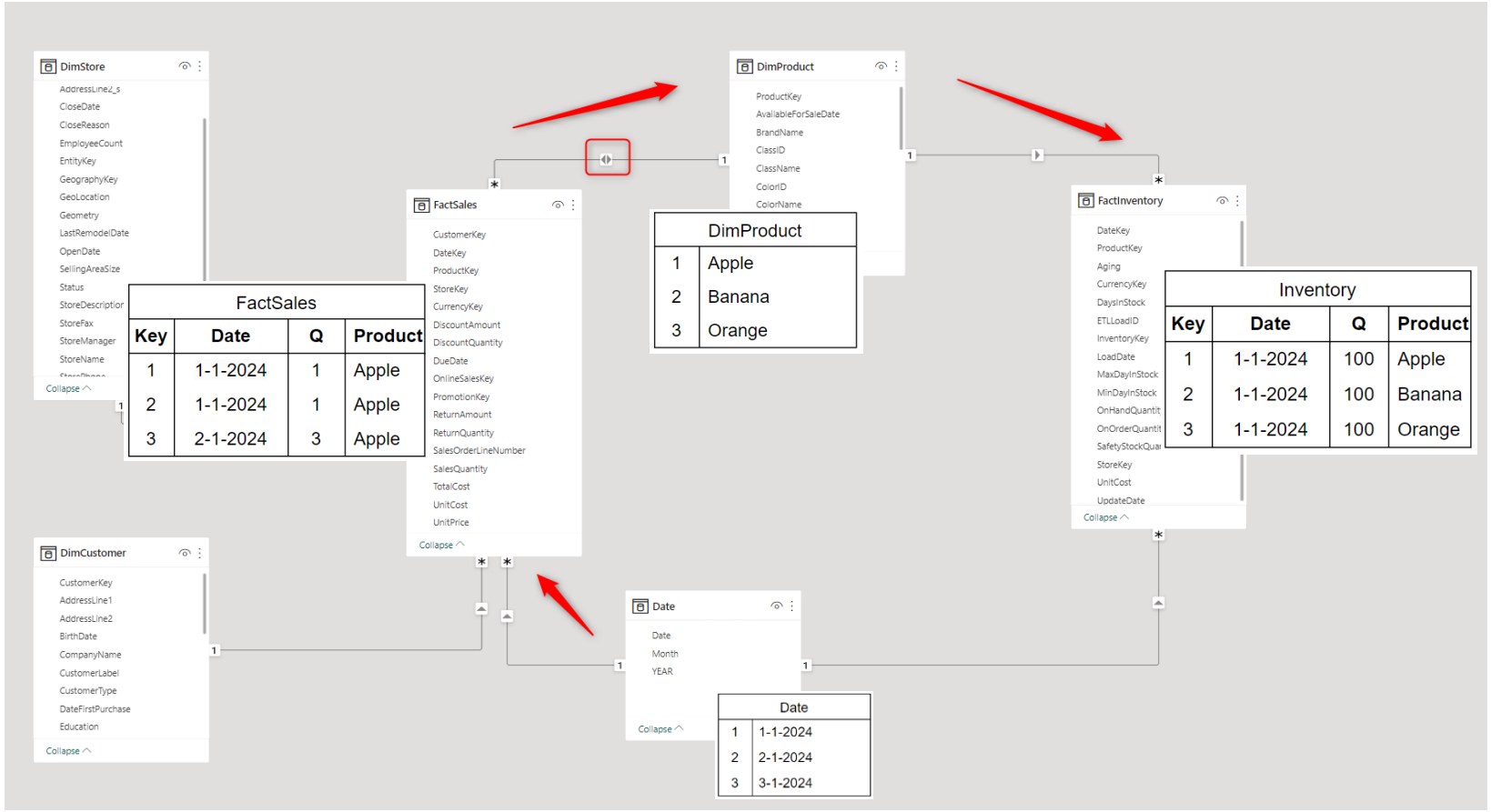
Ambiguity

- When a model includes multiple fact tables and uses bidirectional relationships, there is a risk of ambiguity.
- This means the model may not know which path to follow when filtering.
- If it can filter through multiple tables, it may display incorrect values.



Example ambiguity

Data Modeling



dutchBIgeek

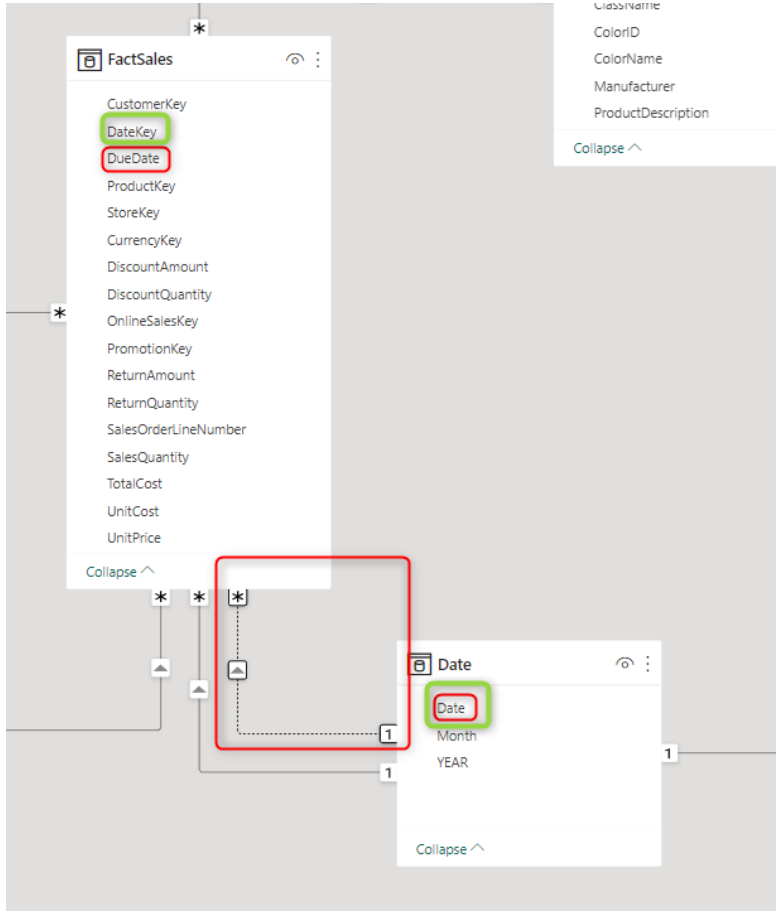
Multiple date tables

Data Modeling

- When comparisons need to be made between dates, multiple date tables are sometimes used.
- The reason for this is that you can only have one active relationship between tables.
- This adds extra data to your model, which can be inefficient.
- It is better to use the USERELATIONSHIP function in DAX.
- This allows for easier comparisons, and the relationship only becomes active when the measure is used.

Example of multiple date tables

Data Modeling




```
€ Total Sales DueDate =
CALCULATE(
    SUMX(
        FactSales,
        FactSales[SalesQuantity] * FactSales[UnitPrice]
    ), USERELATIONSHIP('Date'[Date],FactSales[DueDate])
)
```

YEAR	Month	€ Total Sales	€ Total Sales DueDate
2020	1	33.009.186,00	37.071.480,00
2020	2	35.875.667,00	32.881.523,00
2020	3	37.972.129,00	39.884.837,00
2020	4	23.246.283,00	24.885.072,00
2020	5	25.203.379,00	25.347.549,00
2020	6	22.731.355,00	22.762.326,00
2020	7	31.128.116,00	29.408.149,00
2020	8	26.639.125,00	28.419.751,00
2020	9	27.177.507,00	25.827.772,00
2020	10	29.276.942,00	30.009.064,00
2020	11	24.816.322,00	25.134.687,00
2020	12	28.318.035,00	27.652.516,00
Total		**345.394.046,00**	**349.284.726,00**

Level of detail in the tables

Data Modeling

The level of detail at the lowest level in your table.

Consider what is needed at the lowest level of the table, for example:

- What is the lowest level of each individual purchase?
- Is it sufficient to store the revenue per day?

DateKey	TotalCost
7-4-2019 0:00:00	52,00
7-4-2019 0:00:00	56,00
7-4-2019 0:00:00	66,00
7-4-2019 0:00:00	115,00
7-4-2019 0:00:00	131,00
7-4-2019 0:00:00	171,00
7-4-2019 0:00:00	242,00
7-4-2019 0:00:00	254,00
7-4-2019 0:00:00	285,00
7-4-2019 0:00:00	356,00
7-4-2019 0:00:00	408,00
7-4-2019 0:00:00	413,00
7-4-2019 0:00:00	436,00
7-4-2019 0:00:00	484,00
7-4-2019 0:00:00	509,00
7-4-2019 0:00:00	758,00
7-4-2019 0:00:00	827,00
7-4-2019 0:00:00	831,00
7-4-2019 0:00:00	1.274,00

Date	€ Total Sales
7-4-2019 0:00:00	1.529.817,00
8-4-2019 0:00:00	1.140.754,00
9-4-2019 0:00:00	1.671.361,00
10-4-2019 0:00:00	1.988.986,00
11-4-2019 0:00:00	1.707.435,00
12-4-2019 0:00:00	1.127.218,00
13-4-2019 0:00:00	1.564.933,00
14-4-2019 0:00:00	1.837.926,00
15-4-2019 0:00:00	1.436.617,00
16-4-2019 0:00:00	1.808.614,00
17-4-2019 0:00:00	2.025.175,00
18-4-2019 0:00:00	1.873.287,00
19-4-2019 0:00:00	1.633.725,00
20-4-2019 0:00:00	1.544.020,00
21-4-2019 0:00:00	1.459.055,00
22-4-2019 0:00:00	1.883.524,00
23-4-2019 0:00:00	1.520.738,00
24-4-2019 0:00:00	1.570.734,00
25-4-2019 0:00:00	1.704.843,00

Many-to-many relationships

Data Modeling

Many-to-many (M:N) relationships

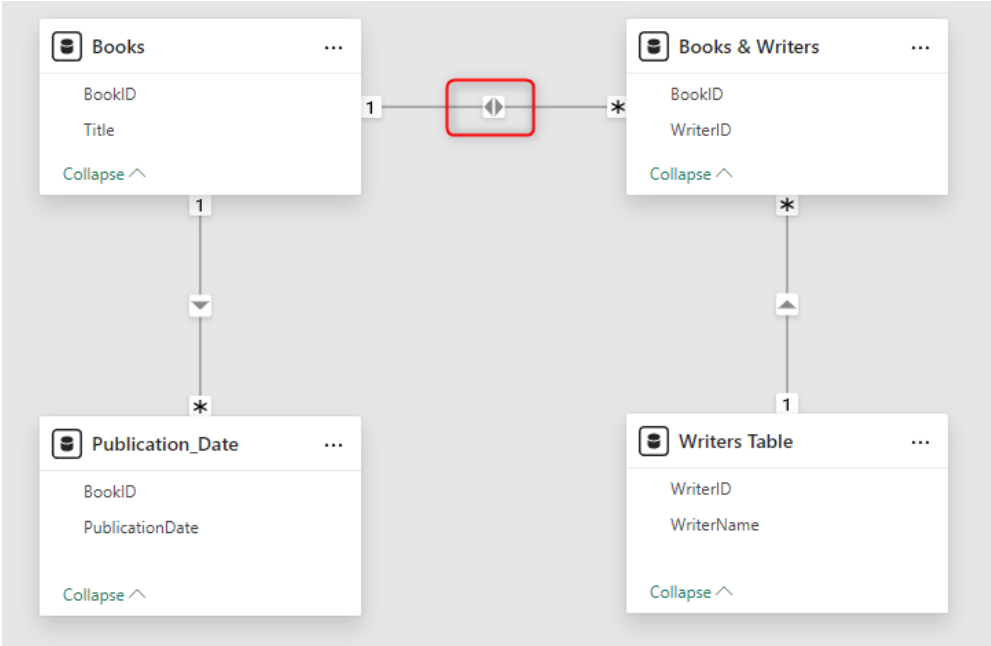
- Example: Accounts and Customers (CRM vs. ERP)

How to solve this?

- Create a bridge table:
- Make a table containing all keys.

Pay attention to the direction of the cross-filter:

- Filtering only works from the 1 to * side.



dutchBIgeek

Importance of the Star Schema for VertiPaq

Data Modeling

- **Star schema boosts performance:**
  - Faster queries by separating dimensions and facts.
  - Simplifies data for efficient VertiPaq processing.
- **Proper data preparation is key:**
  - Consistent data format ensures better compression and lower RAM usage.
- **Small tweaks, big gains:**
  - Minor adjustments can greatly improve performance.
  - Next: Three key factors for optimizing compression.

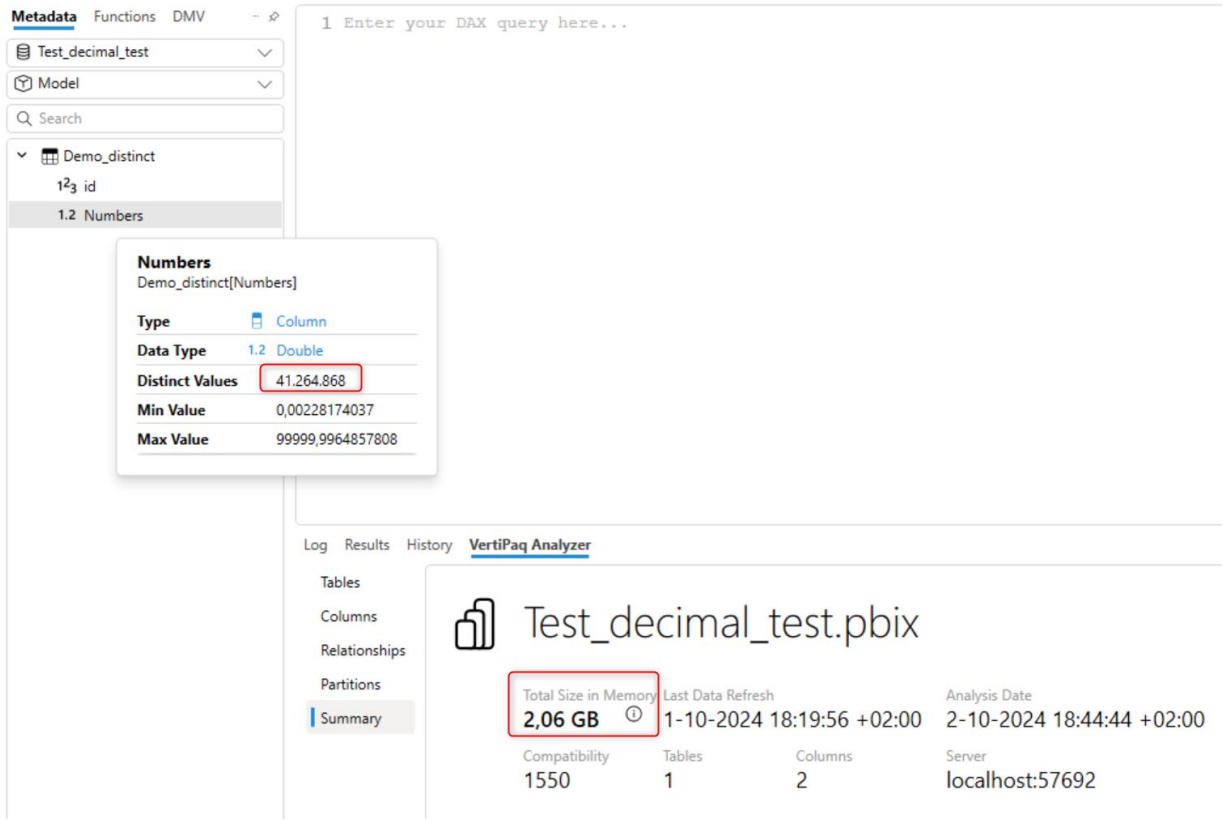
Importance of Choosing the Right Data Types

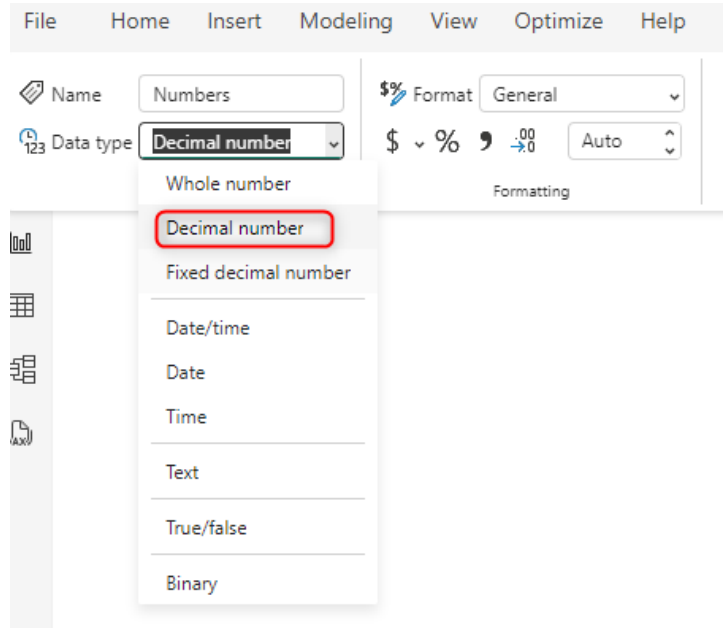
Data Modeling

- **Select the correct data types to optimize model performance**
  - Understand data requirements and report needs
- **Power Query defaults to "Decimal" for numeric columns:**
  - May lead to unnecessary digits after the decimal point
  - Evaluate if fewer decimal places or whole numbers are sufficient
- **Correct data type choice is crucial for VertiPaq compression**

Example Decimal number

Data Modeling

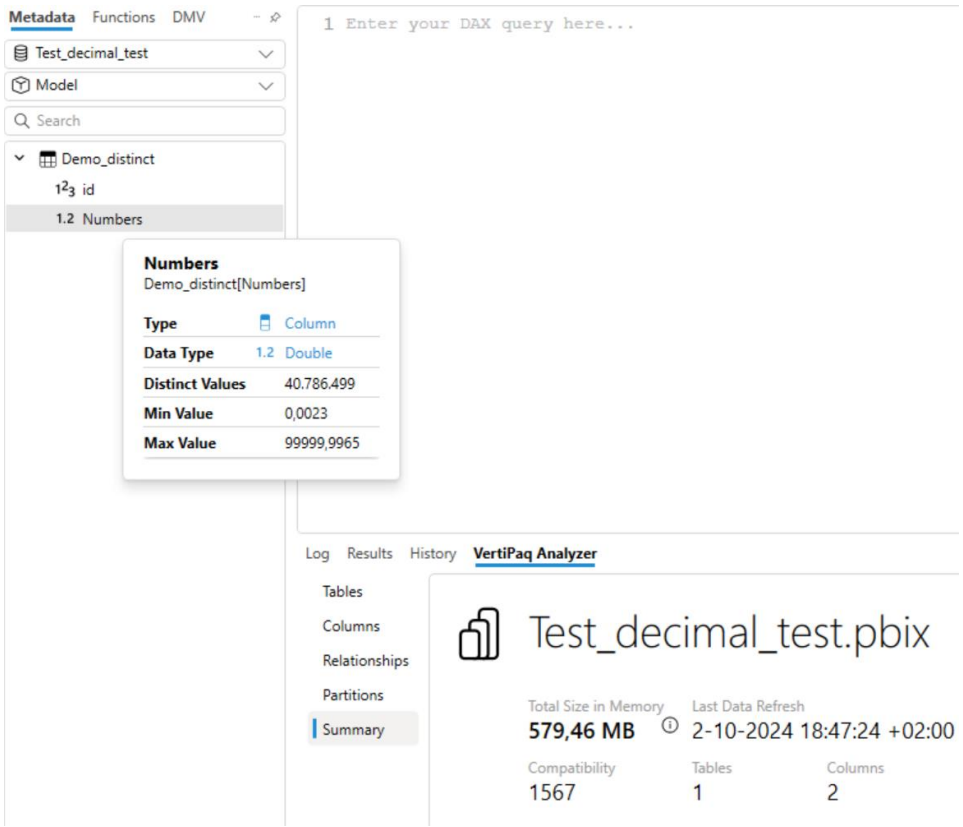


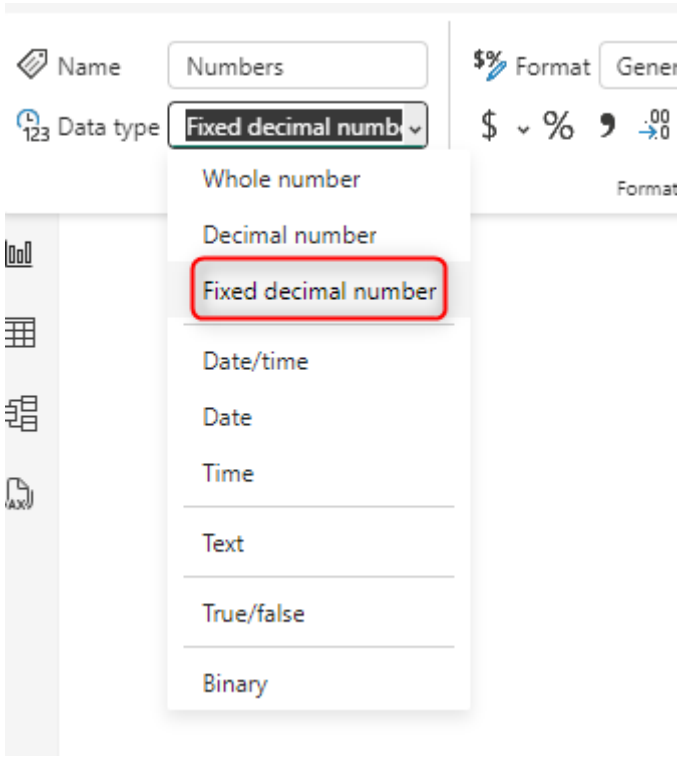


dutchBIgeek

Example Fixed Decimal number

Data Modeling





dutchBIgeek

Optimizing Data Loading for Compression

Data Modeling

- **Load only necessary data to achieve optimal compression**
- **Reduce the number of columns per table:**
  - Fewer columns lead to more effective compression
  - VertiPaq sort order technique: Stores columns with the lowest cardinality first for better compression
  - More columns, especially with higher cardinality, reduce compression efficiency
- **Limit to around 15 columns per table to maintain an efficient model**

Calculated Columns: Pros and Cons

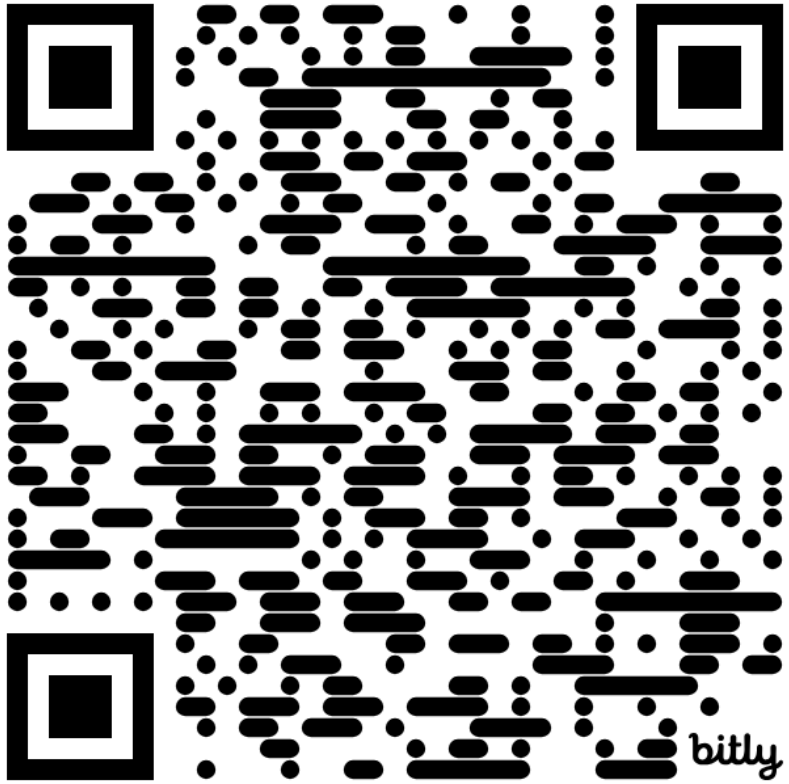
- **Calculated columns are useful during development:**
  - Allow for quick testing and validation without modifying data sources
  - Provide flexibility and speed up development

- **Limitations of calculated columns in production:**
  - Added after model compression, not compressed efficiently
  - Increase memory usage and reduce performance
  - Can undo optimization efforts by increasing model size

Best practices

- Always use a star schema, or a snowflake schema if necessary.
- A fact table contains values for calculations.
- A dimension table contains values for filtering.
- Avoid using bidirectional relationships.
- Include only the data you actually use.
- Determine the level of detail for your tables in advance.

Session Feedback 💖



https://bit.ly/dMC2024_SessionFeedback

dutchBIgeek

If you have questions or insights, please contact me !

Peter van den Bos

Business Intelligence Consultant

peter@dutchbigeek.nl

+31 6 13760795

dutchbigeek.nl