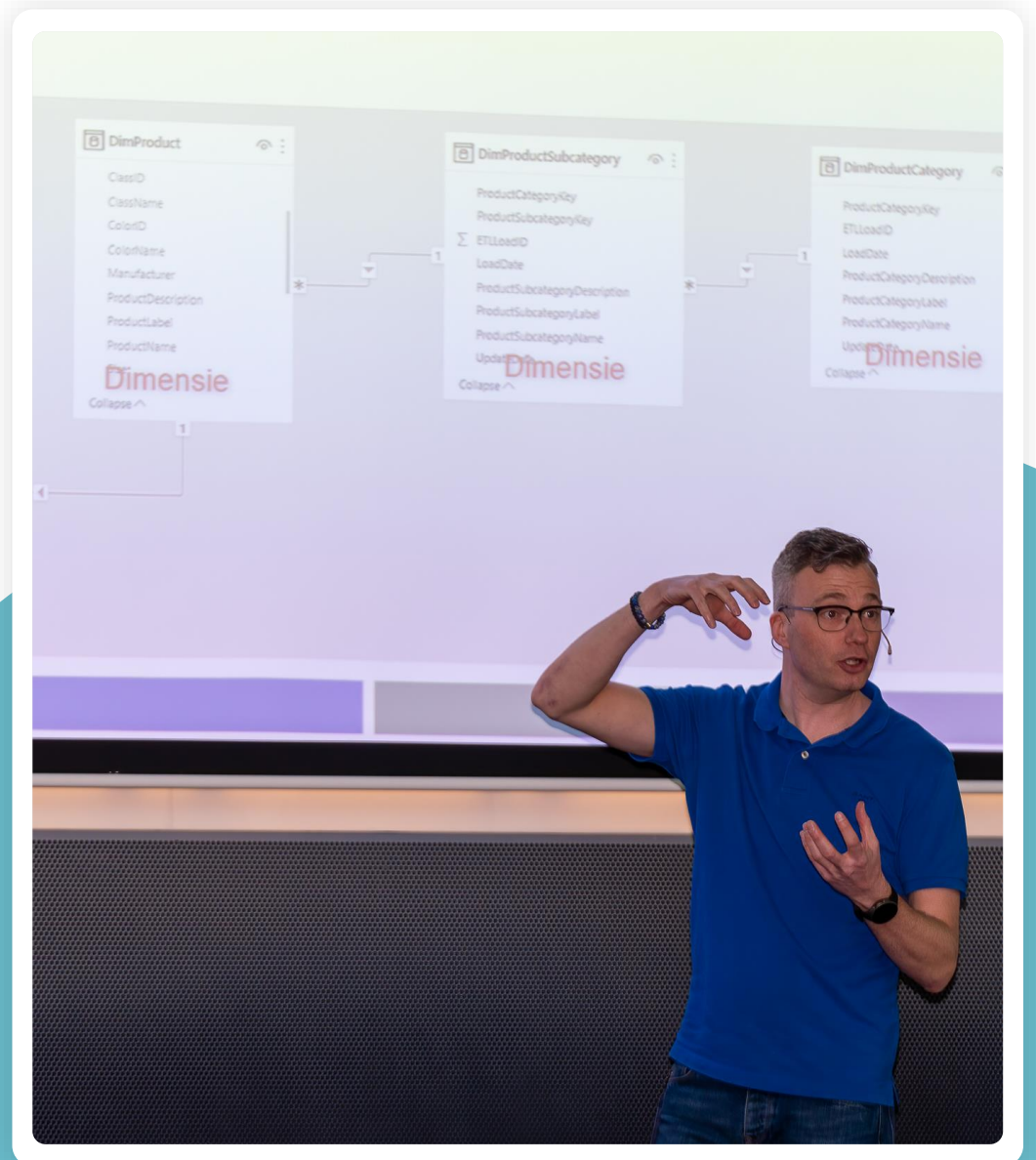


Datamodelleren in Power BI

06-3-2023



Praktische zaken

Data modeling

Duur:

Start: 9:00 uur, einde: 17:00 uur.

Pauzes:

Twee pauzes van 30 minuten.

Lunch duurt 60 minuten.

Trainingsstijl:

Luisteren en kijken en dan vooral zelf doen.

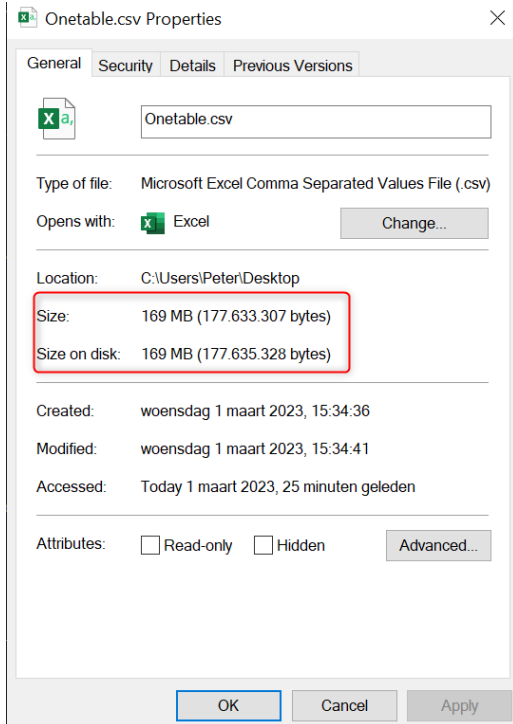
Telefoons graag geluid uit en voel je vrij om op elk ogenblik een vraag te stellen.

Waarom is het datamodel belangrijk?

- Het datamodel is de kern van Power BI.
- Power BI is geoptimaliseerd om grote databestanden te comprimeren.
- 80% van de performance problemen zit in het datamodel.
- Hoe beter je datamodel, hoe makkelijker je DAX-query's.

Voorbeeld waarom het datamodel belangrijk is

Data Modeling



Log Results History **VertiPaq Analyzer**

Tables

Columns

Relationships

Partitions

Summary

Onetable.pbix

Total Size in Memory	Last Data Refresh	Analysis Date
111,48 MB	1-3-2023 15:56:52 +01:00	1-3-2023 15:56:57 +01:00
Compatibility	Tables	Columns
1550	1	88
Server		
localhost:49880		

Log Results History **VertiPaq Analyzer**

Tables

Columns

Relationships

Partitions

Summary

Onetable_split.pbix

Total Size in Memory	Last Data Refresh	Analysis Date
19,13 MB	1-3-2023 15:50:49 +01:00	1-3-2023 15:50:57 +01:00
Compatibility	Tables	Columns
1550	4	94
Server		
localhost:50578		

Voorbeeld waarom het datamodel belangrijk is

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	Marbu
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

Waaruit bestaat een geoptimaliseerd datamodel?

Het model bestaat uit feiten (facts) en dimensies (dimensions).

Een fact-tabel bestaat uit waarden waarmee je kunt rekenen, denk aan:

- Omzet, aankoopmoment, verkochte producten, enzovoorts.

Een dimensions-tabel bestaat uit waarden waarop je wilt filteren, denk aan:

- Jaar, maand, fabrikant, klant, enzovoorts.

Voorbeeld van een geoptimaliseerd datamodel

Data Modeling

Facts-tabel

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	2020-02-17 00:00:00.000

Dimensions-tabel

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 district 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

Waarom?

- Power BI is een in-memory database
- Gebaseerd op de relationele methodologie
- Kolom georiënteerde database

Row storage lay-out

StoreKey	StoreName	ZipCode	StorePhone	AddressLine1
15	Contoso Renton Store	97001	424-555-0189	Renton, Citycenter
17	Contoso Spokane Store	97001	984-555-0185	Logan Ave Shopping mall
20	Contoso Englewood Store	87001	435-555-0113	Cherry Hills Village Shopping mall
22	Contoso Westminster Store	87001	925-555-0114	South Central Shopping mall
25	Contoso Greeley No.1 Store	80631	332-555-0150	W 29th Street, Greeley CityCenter
27	Contoso Lafayette Store	87001	113-555-0173	Lafayette Downtown
30	Contoso Fort Collins Store	87001	656-555-0119	Old Town West ST
32	Contoso Milliken Store	87001	523-555-0175	Milliken Downtown
35	Contoso Denver No.2 Store	87001	952-555-0178	West Highland Borden Ave
37	Contoso Boulder Store	87001	937-555-0137	Transit Village Broadway
40	Contoso Aurora Store	87001	794-555-0159	Willow Park Ave
42	Contoso Green Bay Store	54001	286-555-0189	Green Bay Downtown
45	Contoso Oshkosh Store	54001	786-555-0144	Oshkosh, WI Broadway
47	Contoso Milwaukee No.1 Store	54001	818-555-0192	Havenwoods St

Column storage lay-out

StoreKey	StoreName	ZipCode	StorePhone	AddressLine1
15	Contoso Renton Store	97001	424-555-0189	Renton,Citycenter
17	Contoso Spokane Store	97001	984-555-0185	Logan Ave Shopping mall
20	Contoso Englewood Store	87001	435-555-0113	Cherry Hills Village Shopping mall
22	Contoso Westminster Store	87001	925-555-0114	South Central Shopping mall
25	Contoso Greeley No.1 Store	80631	332-555-0150	W 29th Street, Greeley CityCenter
27	Contoso Lafayette Store	87001	113-555-0173	Lafayette Downtown
30	Contoso Fort Collins Store	87001	656-555-0119	Old Town West ST
32	Contoso Milliken Store	87001	523-555-0175	Milliken Downtown
35	Contoso Denver No.2 Store	87001	952-555-0178	West Highland Borden Ave
37	Contoso Boulder Store	87001	937-555-0137	Transit Village Broadway
40	Contoso Aurora Store	87001	794-555-0159	Willow Park Ave
42	Contoso Green Bay Store	54001	286-555-0189	Green Bay Downtown
45	Contoso Oshkosh Store	54001	786-555-0144	Oshkosh, WI Broadway
47	Contoso Milwaukee No.1 Store	54001	818-555-0192	Havenwoods St

Querying a columnar database

StoreKey	StoreName	ZipCode	StorePhone	AddressLine1
15	Contoso Renton Store	97001	424-555-0189	Renton,Citycenter
17	Contoso Spokane Store	97001	984-555-0185	Logan Ave Shopping mall
20	Contoso Englewood Store	87001	435-555-0113	Cherry Hills Village Shopping mall
22	Contoso Westminster Store	87001	925-555-0114	South Central Shopping mall
25	Contoso Greeley No.1 Store	80631	332-555-0150	W 29th Street, Greeley CityCenter
27	Contoso Lafayette Store	87001	113-555-0173	Lafayette Downtown
30	Contoso Fort Collins Store	87001	656-555-0119	Old Town West ST
32	Contoso Milliken Store	87001	523-555-0175	Milliken Downtown
35	Contoso Denver No.2 Store	87001	952-555-0178	West Highland Borden Ave
37	Contoso Boulder Store	87001	937-555-0137	Transit Village Broadway
40	Contoso Aurora Store	87001	794-555-0159	Willow Park Ave
42	Contoso Green Bay Store	54001	286-555-0189	Green Bay Downtown
45	Contoso Oshkosh Store	54001	786-555-0144	Oshkosh, WI Broadway
47	Contoso Milwaukee No.1 Store	54001	818-555-0192	Havenwoods St

VertiPaq compression

- **Kolom:**
 - Snel toegang tot een enkele kolom
 - Tijd nodig om rijen te materialiseren
 - Trade CPU vs. I/O
 -
- **Rij:**
 - Snel toegang tot een enkele rij
 - Geen materialisatie nodig
 - Trade I/O vs. CPU

Run Length Encoding (RLE)

Data modeling

Quarter	ProdID	Price
Q1	1	523
Q1	1	784
Q1	1	643
Q1	1	184
Q1	1	385
Q1	2	5445
Q1	2	454
Q1	2	665
Q1	1	49
Q4	3	408
Q4	3	612
Q4	3	502
Q4	3	144
Q4	1	300
Q4	1	4247
Q4	1	354
Q4	1	519
Q4	1	38

Quarter	Start	count
Q1	1	9
Q4	10	9

ProdID	Start	count	Price
1	1	5	523
2	6	3	784
1	9	1	643
3	10	4	184
1	14	5	385
			5445
			454
			665
			49
			408
			612
			502
			144
			300
			4247
			354
			519
			38

Dictionary encoding

Quarter	Quarter.id
Q1	0
Q1	0
Q1	0
Q1	0
Q1	0
Q2	1
Q2	1
Q2	1
Q2	1
Q2	1
Q3	2
Q3	2
Q3	2
Q3	2
Q4	3
Q4	3
Q4	3
Q4	3

Quarter.id	Quarter.id
0	Q1
1	Q2
2	Q3
3	Q4

Quarter.id	Start	Count
0	1	5
1	6	5
2	11	5
3	16	5

VertiPaq compression

- **Dictionary codering:**
 - Gebeurt wanneer nodig
 - Grote reeksen hele getallen worden niet gecodeerd
 - Maakt tabellen onafhankelijk van soort gegevens
- **RLE codering:**
 - Alleen als gecomprimeerde gegevens kleiner zijn dan de originele
 - Hangt sterk af van de volgorde van de gegevens
 - SASS kiest automatisch de beste sortering

VertiPaq compression




- **Compressie komt van:**
 - Kolom opslag
 - Dictionary codering
 - RLE codering
- Minder RAM nodig voor in-memory database
- Sneller kolommen scannen
- 10x is een goede gemiddelde compressie ratio, vergeleken met niet-gecomprimeerde SQL database

Oefening 1

Converteer de "sales.csv" naar aparte tabellen voor dimensies en de fact tabel. Er worden drie dimensietabellen en één facttabel verwacht.

De dimensietabellen zijn voor Customer, Product en Store.

De facttabel is voor Sales.

Queries [5]	
	<i>Bron</i>
	Sales
	Customer
	Store
	Product

**Tussen de facts en dimensions moet een relatie gemaakt worden.
Mogelijke relaties zijn:**

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

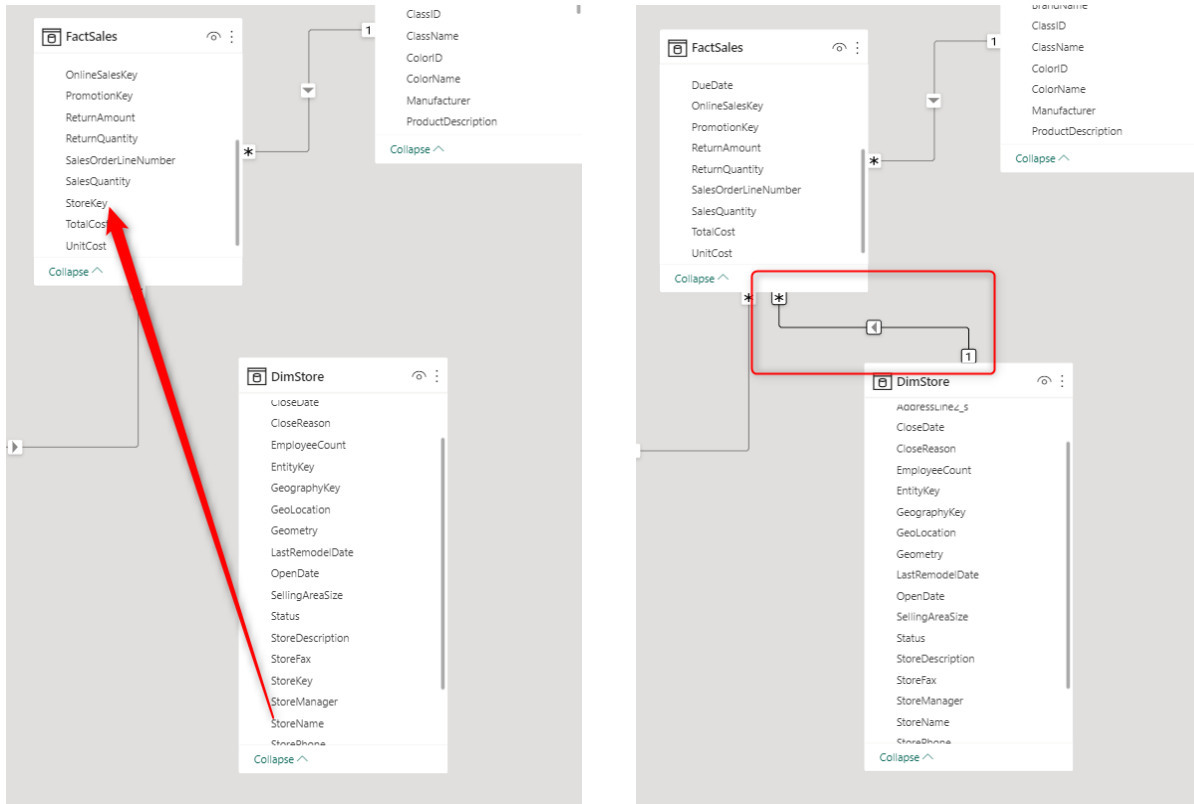
Je kunt daarnaast kiezen in welke richting je wilt filteren:

- Single
- Both

Om de relatie te leggen tussen de tabellen wordt gebruikt gemaakt van key's.

- Primary Key:
 - De Primary Key is een unieke waarde en komt ook maar een keer voor in een dimensie-tabel.
- Foreign key:
 - De Foreign key wordt gebruikt in een fact-tabel en geeft aan hoe vaak er een bijvoorbeeld een product verkocht is.
 - De Foreign key kan wel vaker voorkomen in een fact-tabel, je verkoopt een product, als het goed is, ook vaker.

Voorbeeld relaties en filteren



Edit relationship

Select tables and columns that are related.

FactSales

OnlineSalesKey	DateKey	StoreKey	ProductKey	PromotionKey	CurrencyKey	CustomerKey
20473555	2019-08-14 00:00:00.000	199	1670	1	1	1123
23353743	2020-07-23 00:00:00.000	199	1676	1	1	17042
20706373	2019-09-12 00:00:00.000	199	1676	1	1	7817

DimStore

StoreKey	GeographyKey	StoreManager	StoreType	StoreName	StoreDescription
307	710	292	Online	Contoso Asia Online Store	Contoso Asia Online S
306	586	246	Online	Contoso Europe Online Store	Contoso Europe Onlin
199	800	212	Online	Contoso North America Online Store	Contoso Online Store

Cardinality

Many to one (*:1)

Many to one (*:1)

One to one (1:1)

One to many (1:*)

Many to many (*:*)

Cross filter direction

Single

☐ Apply security filter in both directions

OK

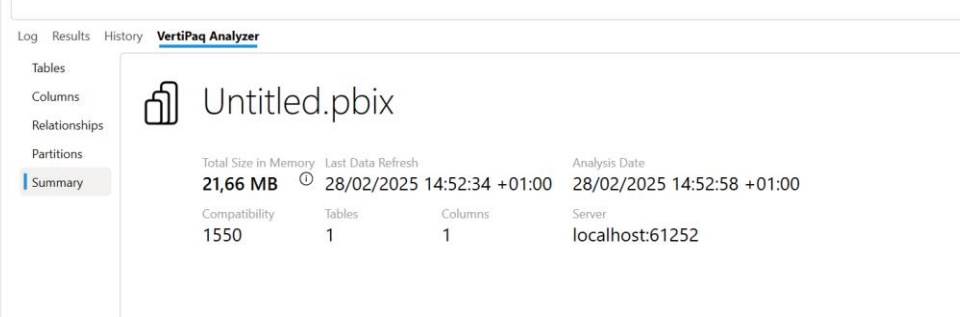
Cancel

Waarom numerieke keys cruciaal zijn

Power BI gebruikt de VertiPaq storage engine, die geoptimaliseerd is voor columnar compression. Hoe beter de compressie, hoe kleiner het model en hoe sneller de queries. Om dit inzichtelijk te maken, heb ik een test uitgevoerd op een dataset met 3,4 miljoen rijen.

Scenario 1: Numerieke key (Int64)

- De dataset is ingeladen met een Integer key
- Geheugenverbruik in Power BI: 21,66 MB



The screenshot shows the 'VertiPaq Analyzer' interface. The left sidebar has tabs for Tables, Columns, Relationships, Partitions, and Summary (selected). The main area displays the following information:

- Model Name: Untitled.pbix
- Total Size in Memory: 21,66 MB
- Last Data Refresh: 28/02/2025 14:52:34 +01:00
- Analysis Date: 28/02/2025 14:52:58 +01:00
- Compatibility: 1550
- Tables: 1
- Columns: 1
- Server: localhost:61252

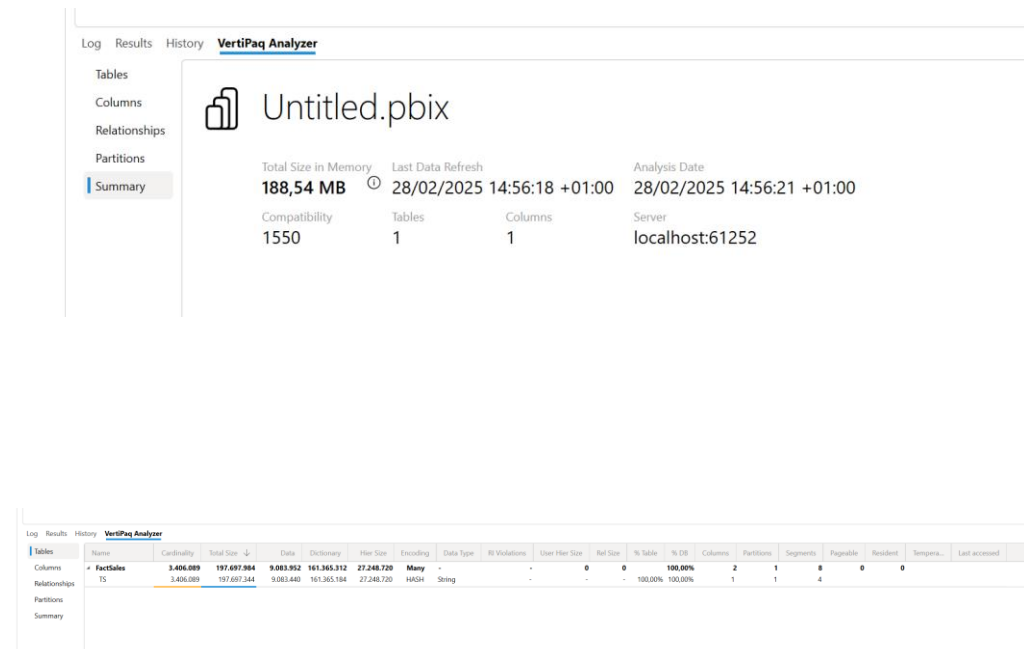
Below this, there is a detailed table with columns: Tables, Columns, Relationships, Partitions, Summary, Compatibility, Row Size, # Rows, Dictionary, Max Value, Row Count, Data Type, # of Nulls, User Hierarchy, Ref Time, % of Data, Columns, Partitions, Segments, Segments, Row Count, Temporal, Last Accessed.

Tables	Columns	Relationships	Partitions	Summary	Compatibility	Row Size	# Rows	Dictionary	Max Value	Row Count	Data Type	# of Nulls	User Hierarchy	Ref Time	% of Data	Columns	Partitions	Segments	Segments	Row Count	Temporal	Last Accessed
FactSales	3,406,089	22,707,926	9,083,952	256	13,624,360	VALUE	-	-	0	0	100.00%	2	1	1	1	1	1	1	1	1	1	1

Waarom numerieke keys cruciaal zijn

Scenario 2: Tekstuele key ('TS' + nummer als string)

- Zelfde dataset, maar met een String key
- Geheugenverbruik in Power BI: 188,54 MB



The screenshot shows the VertiPaq Analyzer interface. The top section displays the model name 'Untitled.pbix' and its total size in memory as 188,54 MB. Below this, a table provides details about the model's components.

Compatibility	Tables	Columns	Server
1550	1	1	localhost:61252

The bottom section shows a detailed table of the model's components, including tables, columns, and relationships.

Tables	Name	Cardinality	Total Size	Data	Dictionary	Size	Encoding	Data Type	Is Violations	User Hierarchy	Ref Size	% Table	% DB	Columns	Partitions	Segments	Pageable	Resident	Tempora...	Last accessed
Columns	FactSales	3,406,089	197,697,884	9,083,952	161,365,312	27,248,720	Many	-	-	0	0	100,00%	100,00%	2	1	8	0	0		
Relationships	TS	3,406,089	197,697,344	9,083,440	161,365,184	27,248,720	HASH	String	-	-	-	100,00%	100,00%	1	1	4	0	0		

Star Schema

- **Star Schema data model.**
- **Feitentabellen.**
 - Met **feiten**
 - Een gebeurtenis met dimensions
 - Een **sale** omvat een product, een klant en een datum.
 - **Metrics** die je samen kunt nemen om inzichten te verkrijgen.
- **Dimension Tabellen**
 - Informele vaststelling zoals van een product, klant, werknemer of patiënt.
 - Dimensions hebben eigenschappen zoals kleur, categorie, fabrikant of prijs.
 - Patiënten hebben eigenschappen zoals naam, adres, geboortedatum.



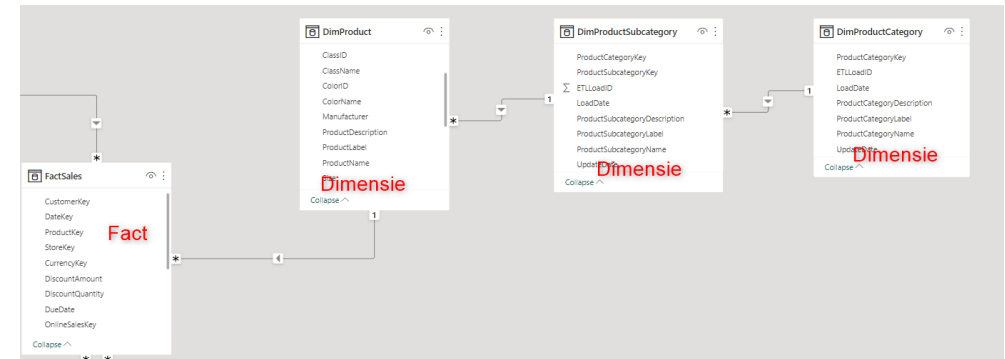
Snowflake-model

Het snowflake-model is een variant op het ster model.

Het verschil is dat dimensie-tabellen, die op elkaar aansluiten, aan elkaar gekoppeld zijn.

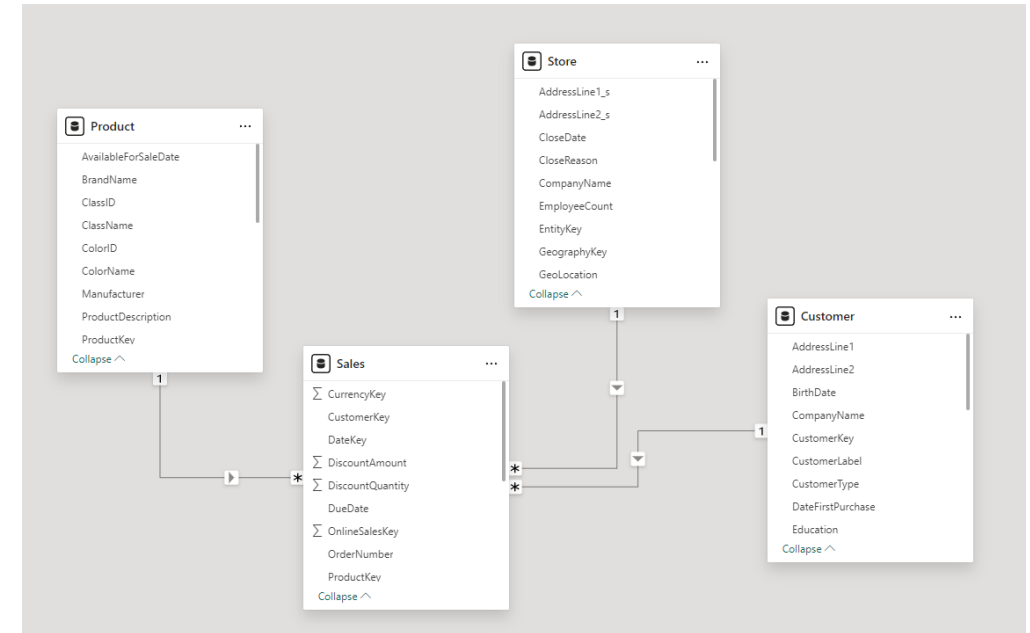
Bijvoorbeeld:

- Product
- Productsubcategorie
- Productcategorie



Oefening 2

Controleer of de tabellen correct zijn opgemaakt en controleer of de relaties juist zijn gelegd.



Meerdere fact-tabellen

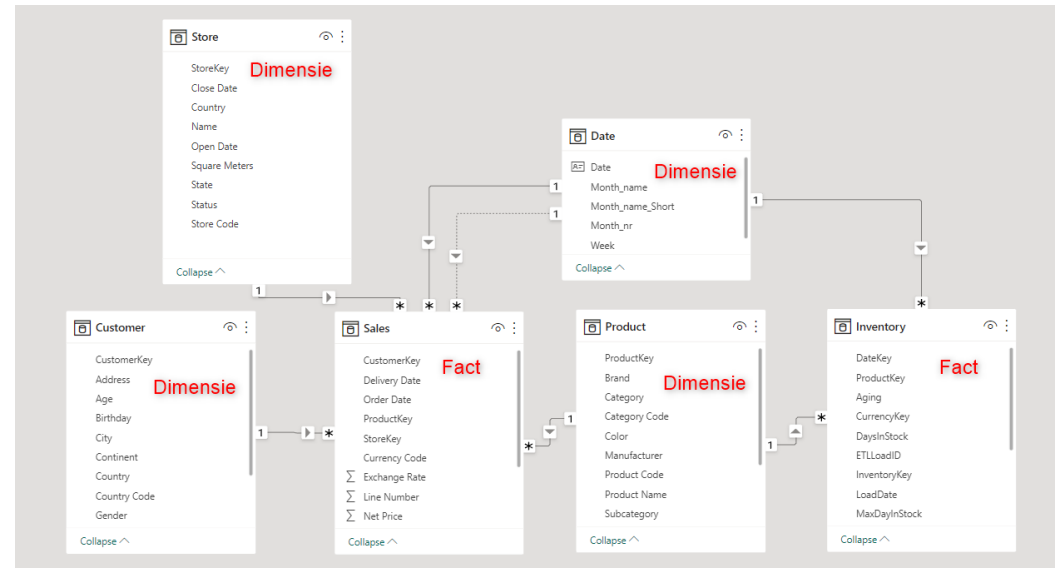
Het is mogelijk om meerdere fact-tabellen in je model te gebruiken.

Deze fact-tabellen hebben dan niet veel gemeen met elkaar.

- Bijvoorbeeld voorraad en sales.

Het is wel nodig om enkele dimensie-tabellen te hebben die gekoppeld zijn aan de fact-tabel.

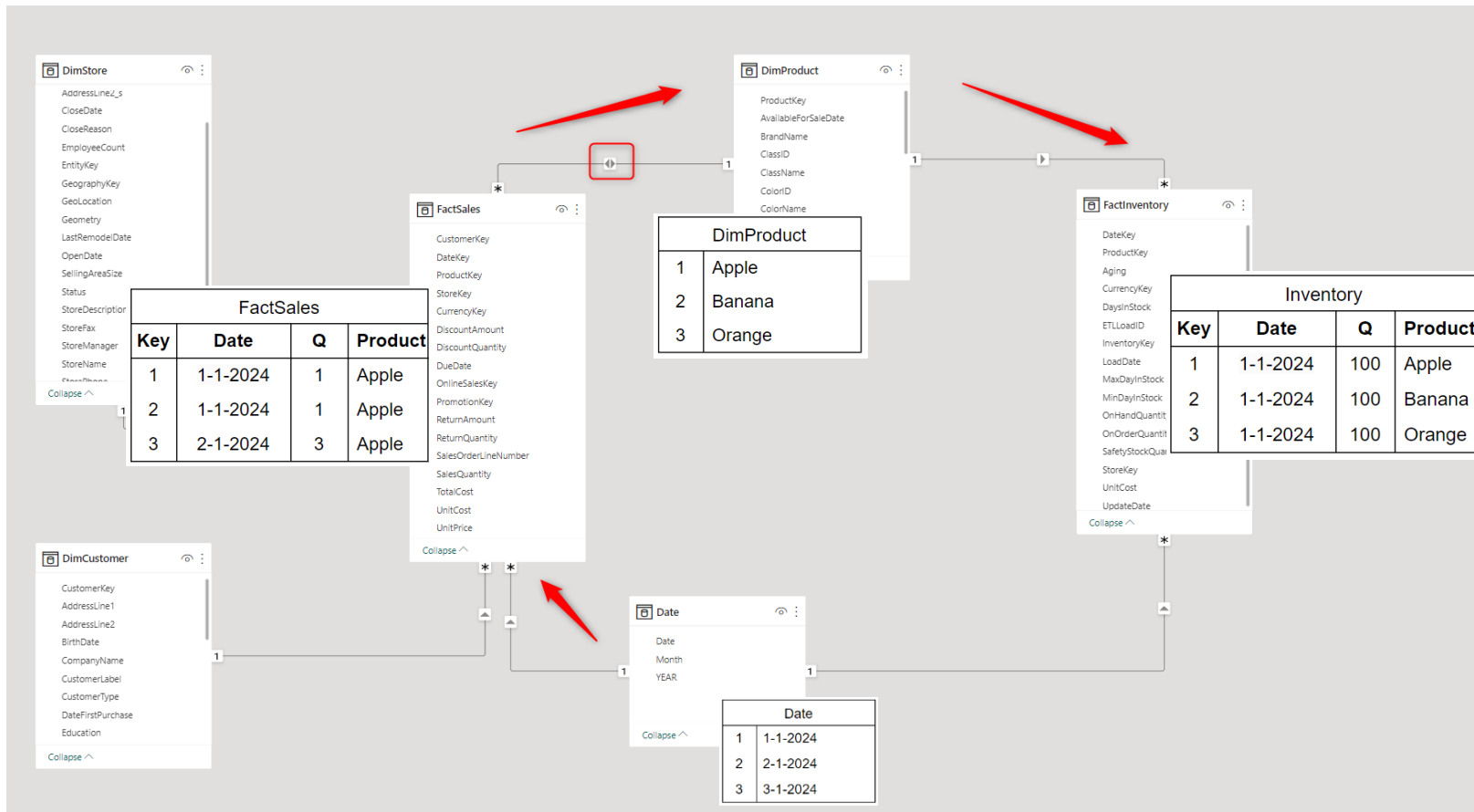
- Date en product zijn hier een voorbeeld van.



Ambiguity

- Als in een model meerdere fact-tabellen zijn opgenomen en er wordt gebruikgemaakt van bidirectionele relaties, ontstaat het risico dat het model ambigu (meerduidig) wordt.
- Dit houdt in dat het model niet weet welke route het moet volgen bij het filteren.
- Als het kan filteren via verschillende tabellen, kan het de verkeerde waarden tonen.

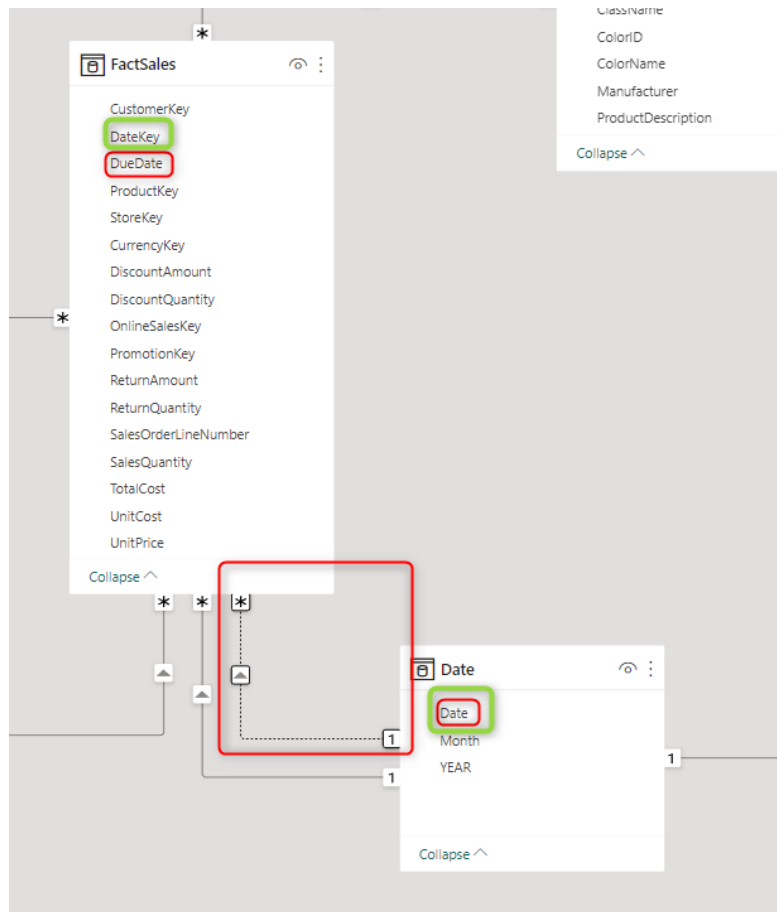
Voorbeeld ambiguity



Meerdere datum-tabellen

- Als er vergelijkingen gemaakt moeten worden tussen datums, wordt er nog wel eens gebruikgemaakt van meerdere datum-tabellen.
- De reden hiervoor is dat je tussen tabellen maar één actieve relatie kan leggen.
- Dit zorgt voor extra data in je model en dat is zonde.
- Het is beter om te kiezen voor de USERELATIONSHIP-functie in DAX.
- Hierdoor kun je gemakkelijker vergelijkingen maken en wordt de relatie alleen actief bij het gebruik van de measure.

Voorbeeld meerdere datum-tabellen



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

Gedetailleerdheid van de tabellen

De gedetailleerdheid van het laagste niveau in je tabel.

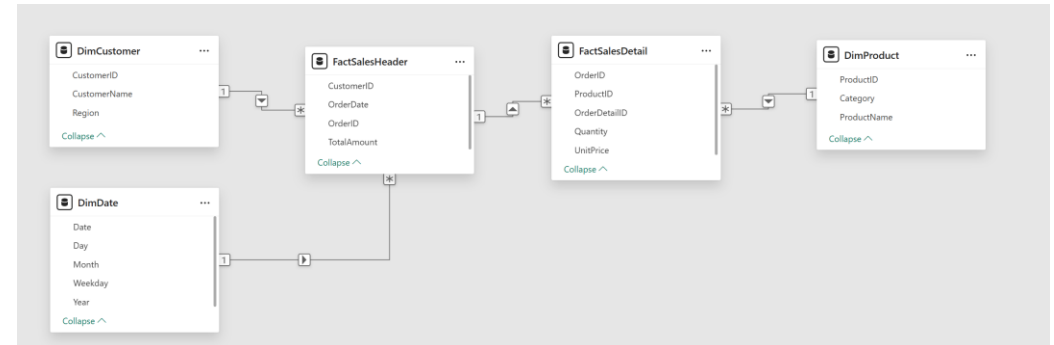
Vraag je af wat nodig is op het laagste niveau van de tabel, bijvoorbeeld:

- Wat is het laagste niveau van elke afzonderlijke aankoop?
- Is het voldoende om de omzet per dag op te slaan?

DateKey	TotalCost	Date	€ Total Sales
7-4-2019 0:00:00	52,00	7-4-2019 0:00:00	1.529.817,00
7-4-2019 0:00:00	56,00	8-4-2019 0:00:00	1.140.754,00
7-4-2019 0:00:00	66,00	9-4-2019 0:00:00	1.671.361,00
7-4-2019 0:00:00	115,00	10-4-2019 0:00:00	1.988.986,00
7-4-2019 0:00:00	131,00	11-4-2019 0:00:00	1.707.435,00
7-4-2019 0:00:00	171,00	12-4-2019 0:00:00	1.127.218,00
7-4-2019 0:00:00	242,00	13-4-2019 0:00:00	1.564.933,00
7-4-2019 0:00:00	254,00	14-4-2019 0:00:00	1.837.926,00
7-4-2019 0:00:00	285,00	15-4-2019 0:00:00	1.436.617,00
7-4-2019 0:00:00	356,00	16-4-2019 0:00:00	1.808.614,00
7-4-2019 0:00:00	408,00	17-4-2019 0:00:00	2.025.175,00
7-4-2019 0:00:00	413,00	18-4-2019 0:00:00	1.873.287,00
7-4-2019 0:00:00	436,00	19-4-2019 0:00:00	1.633.725,00
7-4-2019 0:00:00	484,00	20-4-2019 0:00:00	1.544.020,00
7-4-2019 0:00:00	509,00	21-4-2019 0:00:00	1.459.055,00
7-4-2019 0:00:00	758,00	22-4-2019 0:00:00	1.883.524,00
7-4-2019 0:00:00	827,00	23-4-2019 0:00:00	1.520.738,00
7-4-2019 0:00:00	831,00	24-4-2019 0:00:00	1.570.734,00
7-4-2019 0:00:00	1.274,00	25-4-2019 0:00:00	1.704.843,00

Header-Detail

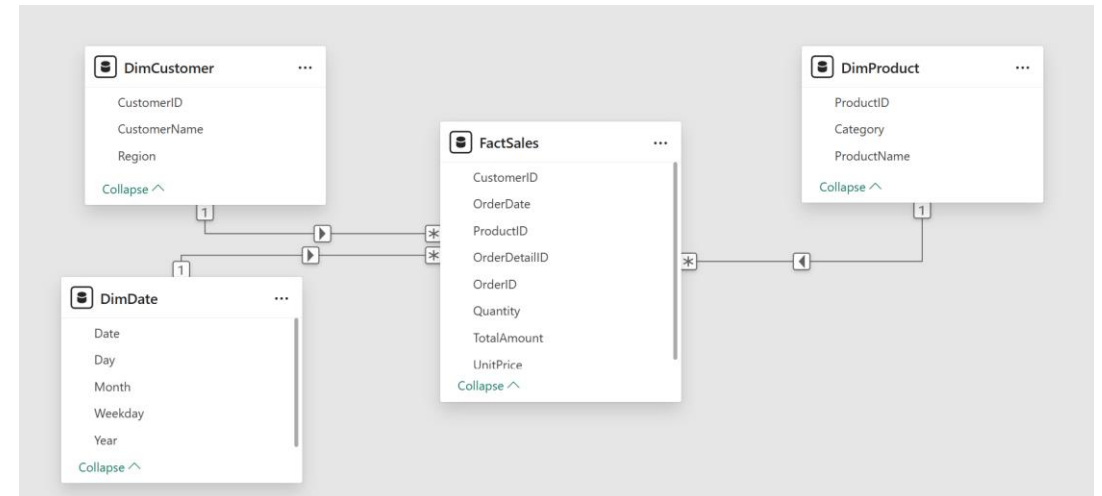
- Headertabel-Detailtabellen lijken op een aangepast star schema.
- Twee aparte star schema's.
- Je kunt van de Headertabel en Detail tabel een Fact tabel te maken.
- Met Denormalization.
- Dit verhoogt de Aggregation.



Header-Detail

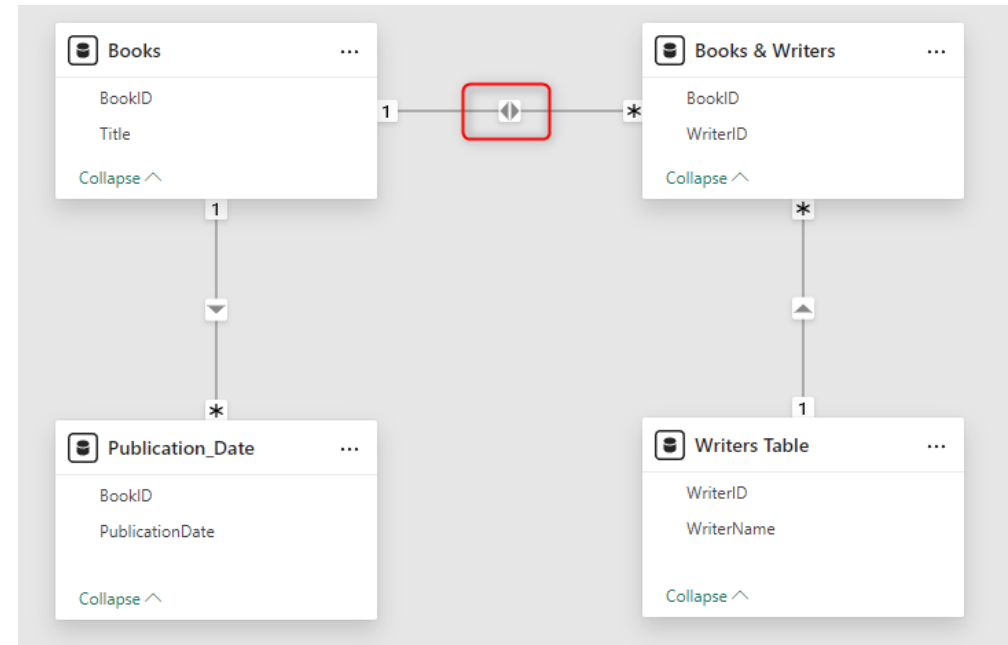
Wat is hier anders?

- De header-informatie (zoals CustomerID, OrderDate en TotalAmount) wordt op elke orderregel herhaald.
- Er is geen aparte FactSalesHeader meer, dus analyses kunnen direct op deze tabel worden uitgevoerd.
- TotalAmount wordt herhaald per orderregel, wat kan leiden tot dubbele tellingen als je niet goed filtert. Dus vervang deze door een measure.



Many-to-many relationships

- **Many(*)-to-Many(*) relationships**
 - Bijv. Books en Writers
 - CRM vs. ERP
 - Hoe los je dit op?
- **Maak a BridgeTable**
 - Maak een tabel van alle key's
- **Let op de richting van het cross filter**
- **Filteren werkt alleen van de 1-* kant**



Slowly Changing Dimensions Type 2 (SCD2)

Wat is een Slowly Changing Dimension?

- Een Slowly Changing Dimension (SCD) is een dimensietabel waarin historische wijzigingen worden bijgehouden.
- Verschil tussen SCD Type 1 (overschrijven van oude waarden) en SCD Type 2 (historie behouden door extra records toe te voegen).
- Nuttig voor trendanalyses en historische rapportages.

Hoe implementeer je SCD2 in Power BI?

- Gebruik een effectiviteitsdatum (StartDate, EndDate) of een IsActive-kolom om actieve records te onderscheiden.
- Laden van SCD2-data gebeurt meestal via een ETL-proces in Power Query of in een datawarehouse.
- Performance-optimalisatie door alleen relevante rijen te filteren in DAX-measures.

Voor- en nadelen van SCD2 in Power BI

Voordelen:

- Behoud van historische data voor nauwkeurige rapportages.
- Mogelijkheid om tijdgebaseerde analyses te doen.

Nadelen:

- Toename van de tabelgrootte, wat compressie kan beïnvloeden.
- Complexere DAX-measures voor filtering en aggregaties.

Snapshots voor tijdsgebonden data

Wat zijn snapshots?

- Een snapshot legt de toestand van data vast op een bepaald moment.
- Verschil tussen transactionele data (real-time wijzigingen) en snapshot data (periodieke metingen).
- Wordt vaak gebruikt voor trendanalyses en forecasting.

Hoe implementeer je snapshots in Power BI?

Data Modeling

- Creëer een fact snapshot table waarin de gegevens per periode worden vastgelegd.
- Gebruik een PeriodKey (bijvoorbeeld Maand of Kwartaal) om tijdsintervallen vast te leggen.
- Combineer snapshots met DAX-measures voor vergelijkingen over tijd.

Performance-overwegingen

- Volledige snapshots: Elke periode bevat een volledige dataset (hogere opslag, eenvoudigere queries).
- Delta snapshots: Alleen de wijzigingen worden opgeslagen (compact, maar complexere berekeningen nodig).

Welke data komt er in een datamodel ?

- Zorg dat er alleen data in je model staat die ook daadwerkelijk gebruikt wordt. Alles wat je niet gebruikt, verwijder je. De stelling 'dat is handig voor later' kun je beter niet gebruiken, want als het nodig is, kun je de data nog toevoegen.
- Filter zo dicht mogelijk bij de bron. Hoe verder van de bron af, hoe meer data er nodeloos wordt verplaatst.
- Gebruik geen dubbele tabellen.

Best practices

- Gebruik altijd een ster model, of als het niet anders kan, een snowflakemodel.
- Een fact-tabel bestaat uit waarden, waarmee je kunt rekenen.
- Een dimensions-tabel bestaat uit waarden, waarop je kunt filteren.
- Voorkom het gebruik van bidirectionele relaties.
- Zorg dat er alleen data in je model staat die ook daadwerkelijk gebruikt wordt.
- Filter zo dicht mogelijk bij de bron.
- Bepaal van te voren je gedetailleerdheid van de tabel.



**Heb je vragen of
interesse? Neem
contact op.**



Peter van den Bos
Business Intelligence Consultant

✉ peter@dutchbigeek.nl

☎ +31 6 13760795

dutchbigeek.nl