Data Modeling for experts with Power BI

#PowerBIer

Marc Lelijveld

Data & Analytics consultant
Macaw Netherlands

Microsoft Power Platform
FASTTRACK RECOMMENDED
SOLUTION ARCHITECT
2021

MVP Microsoft® Most Valuable Professional

Microsoft MCT

Marc.Lelijveld@outlook.com

@MarcLelijveld

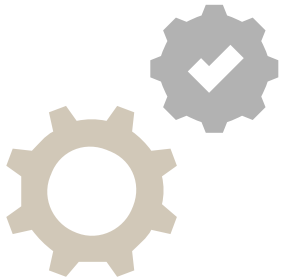
linkedin.com/in/MarcLelijveld

Data-Marc.com

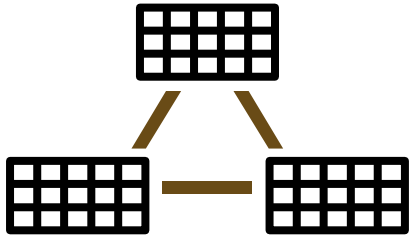
Set the stage

Gather

Clean



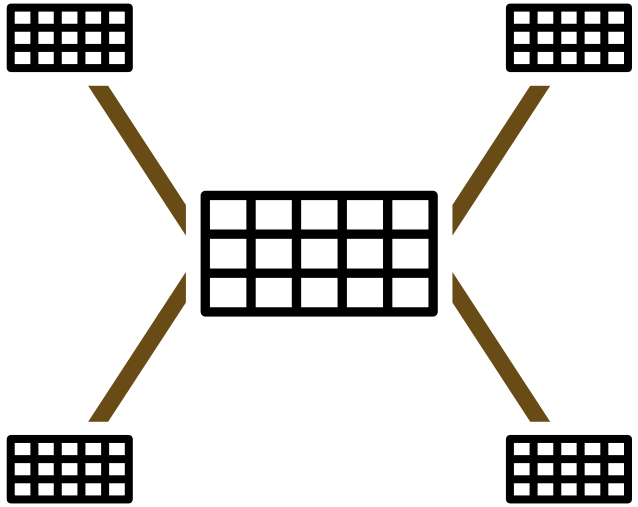
Model



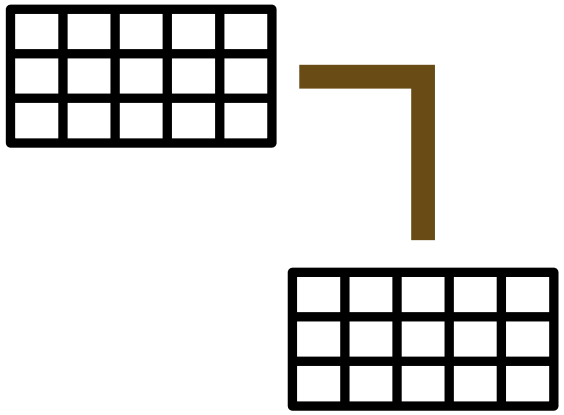
Visualize

Your basic understanding

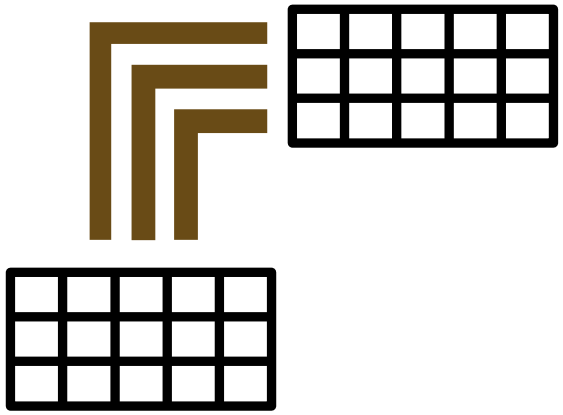
Star Schema



Relationship types



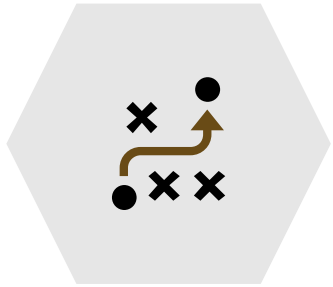
Role Playing Dimensions



Learning objectives

Leverage one and bi-directional relationships

Overcome and handle ambiguous data models



Leverage slowly changing dimensions

Relationships revisited

Relationship types

1 to 1

Every **individual record** in dataset A is mapped to one **individual record** in dataset B

Dataset A
Customer information

Customer ID	Country
AW000111024	United States
AW00019377	Germany

Dataset B
Customer information

Customer ID	Date of Birth
AW000111024	9 april 1990
AW00019377	9 april 1983

1 to many

One record of dataset A is mapped to **multiple records** in dataset B

Dataset A
Customer information

Customer ID	Country
AW000111024	United States
AW00019377	Germany

Dataset B
Sales Information

Customer ID	Product	Order Quantity
AW00011024	CA-1098	2
AW00019377	BC-M005	1
AW00019377	CA-1098	1
AW00019377	FE-6654	1
AW00019377	HL-U509-B	1
AW00019377	TI-M602	1
AW00019377	TT-M928	1
AW00019377	WB-H098	1

Many to many

Multiple records of dataset A are mapped to **multiple records** in dataset B

Dataset A
Customer information

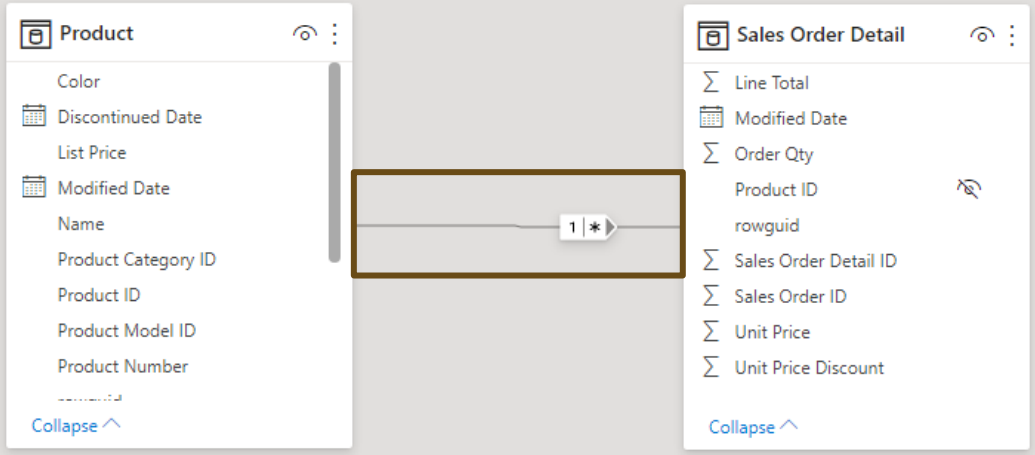
Customer ID	Store ID	Store Type
AW00011024	ON-1	Online
AW00011024	ST-1	Regional City Store
AW00011024	ST-2	Regional City Store
AW00019377	ON-1	Online
AW00019377	ST-1	Regional City Store
AW00019377	ST-2	Regional City Store

Dataset B
Sales Information

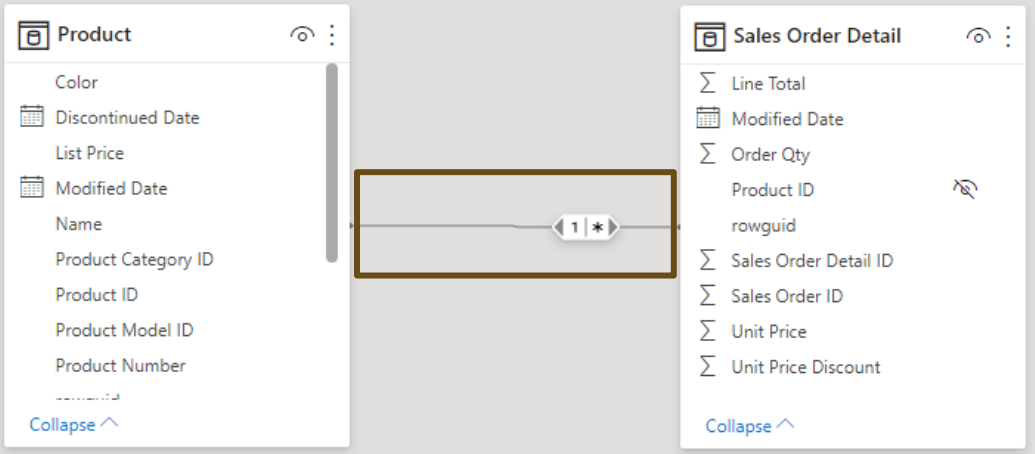
Customer ID	Product	Order Quantity
AW00011024	CA-1098	2
AW00011024	TT-M928	5
AW00019377	BC-M005	1
AW00019377	CA-1098	1
AW00019377	FE-6654	1
AW00019377	HL-U509-B	1
AW00019377	TI-M602	1
AW00019377	TT-M928	1
AW00019377	WB-H098	1

Relationship direction

Singular



Bi-directional



Relationship direction

Bi-directional relationships can result in surprising results, especially when working with multiple fact tables

→ It can have performance impacts and ‘overfilter’

→ Can lead to ambiguity

Try to avoid bi-directional relationships

You can influence the direction of a relationship for the context of a calculation by CROSSFILTER in DAX:

→ CROSSFILTER(column1, column2, direction)

Ambiguity

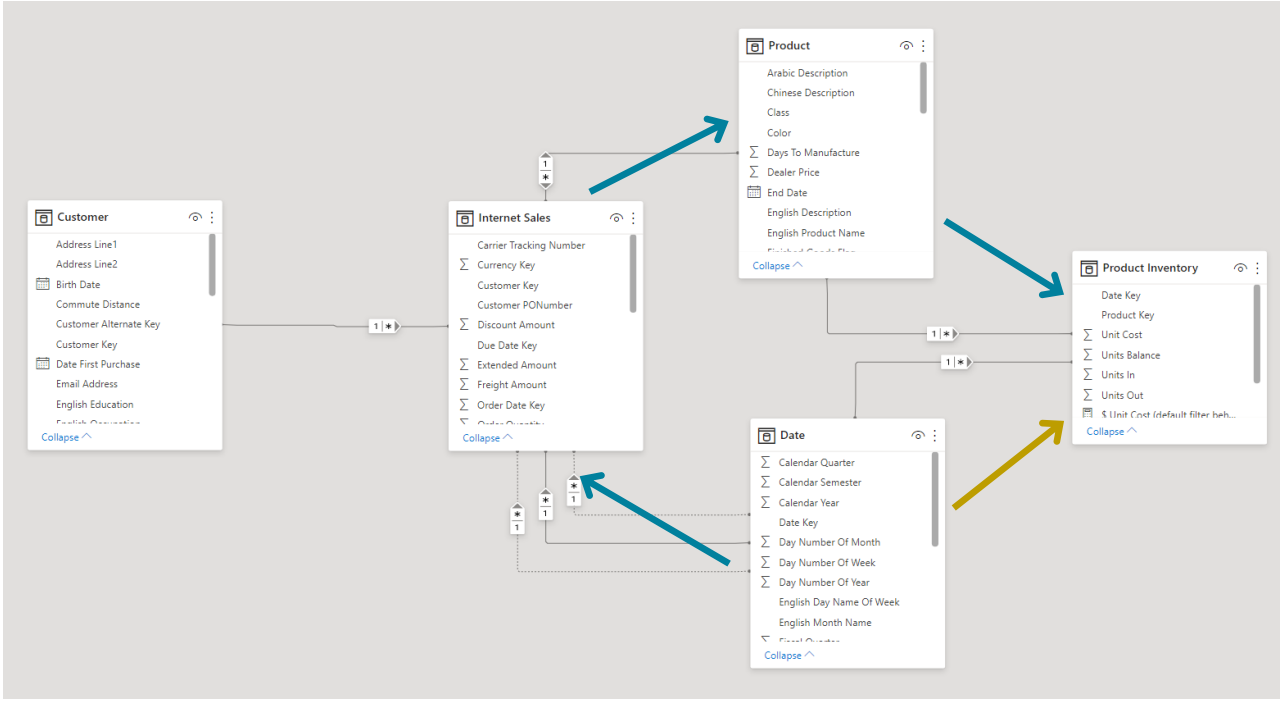
Ambiguous data models

Two filter paths to the same table

→ Leads to unexpected results

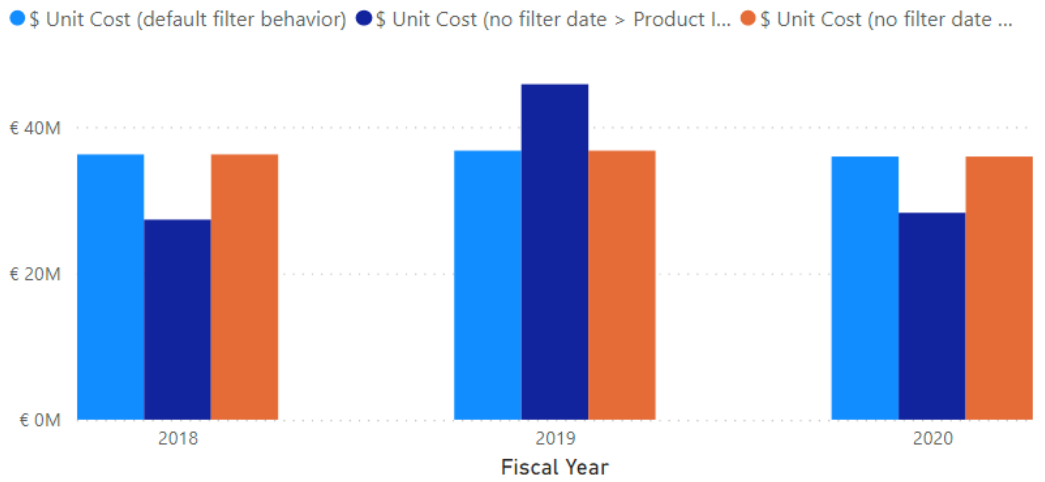
Can happen with bi-directional relationships

→ Avoid bi-directional relationships as much as possible



Ambiguous data models

- Can lead to unpredictable results
- Each of these measures calculates the same, but removes one of the relationships
- By using CROSSFILTER you can change relationships in a measure context

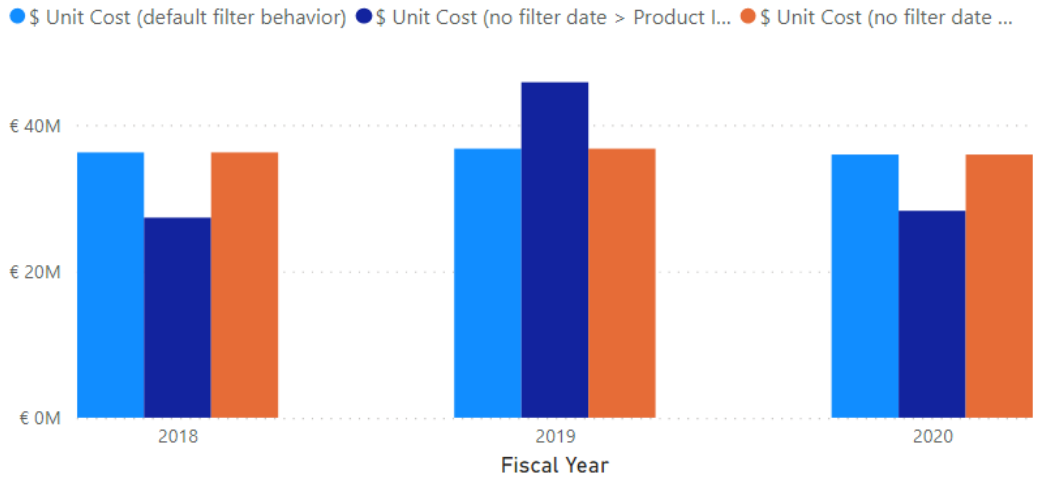


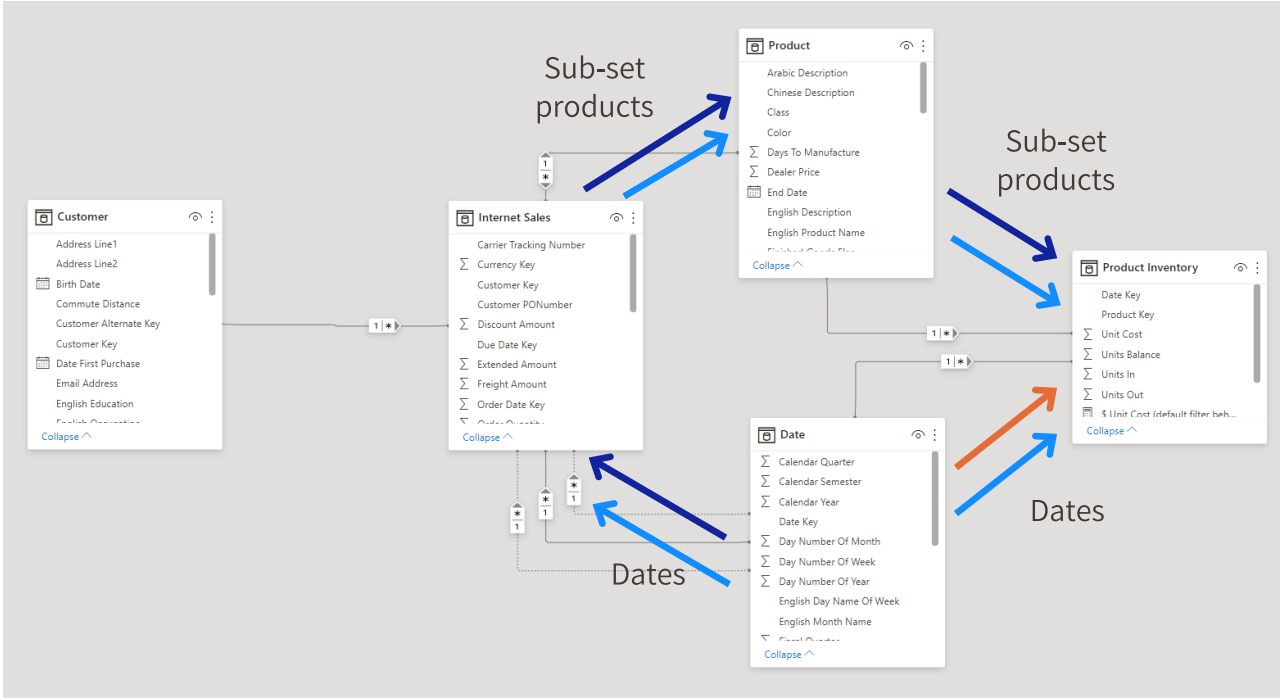
Fiscal Year	$ Unit Cost (default filter behavior)	$ Unit Cost (no filter date > Product Inventory)	$ Unit Cost (no filter date > Internet Sales)
2018	€ 36.244.707,62	€ 27.337.603,34	€ 36.244.707,62
2019	€ 36.758.094,11	€ 45.834.231,75	€ 36.758.094,11
2020	€ 35.954.533,19	€ 28.267.611,68	€ 35.954.533,19
Total	**€ 108.957.334,92**	**€ 108.957.334,92**	**€ 108.957.334,92**

Ambiguous data models

Unexpected filter behavior

- Result will be filtered by a subset of products as part of the Internet Sales (dark blue filter path)
- Result will be filtered by a subset of dates (orange filter path)





Demo – relationships direction

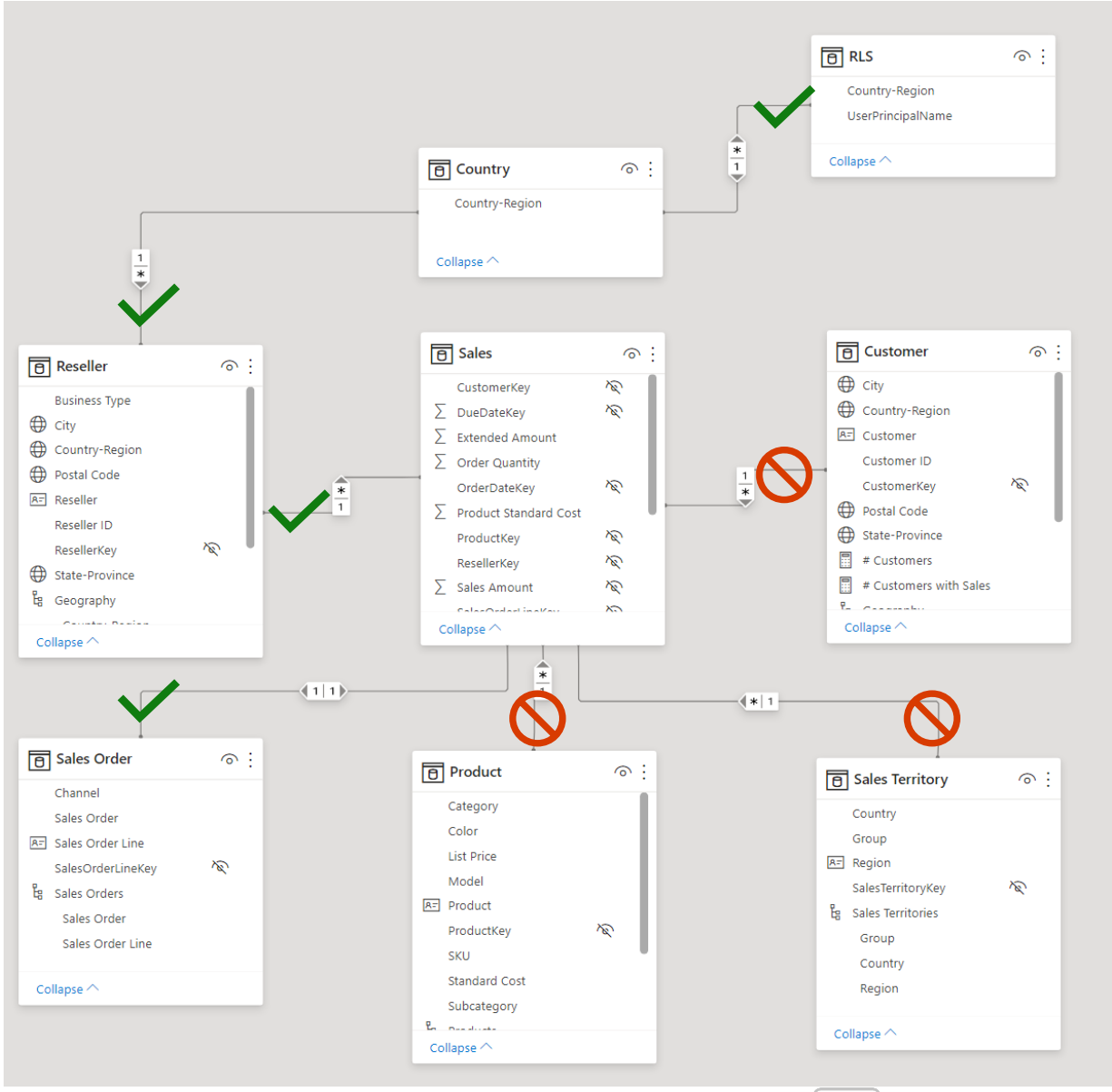
Demo time!

Because life is boring without risks

Relationships & Row level security

Does not filter dimensional tables

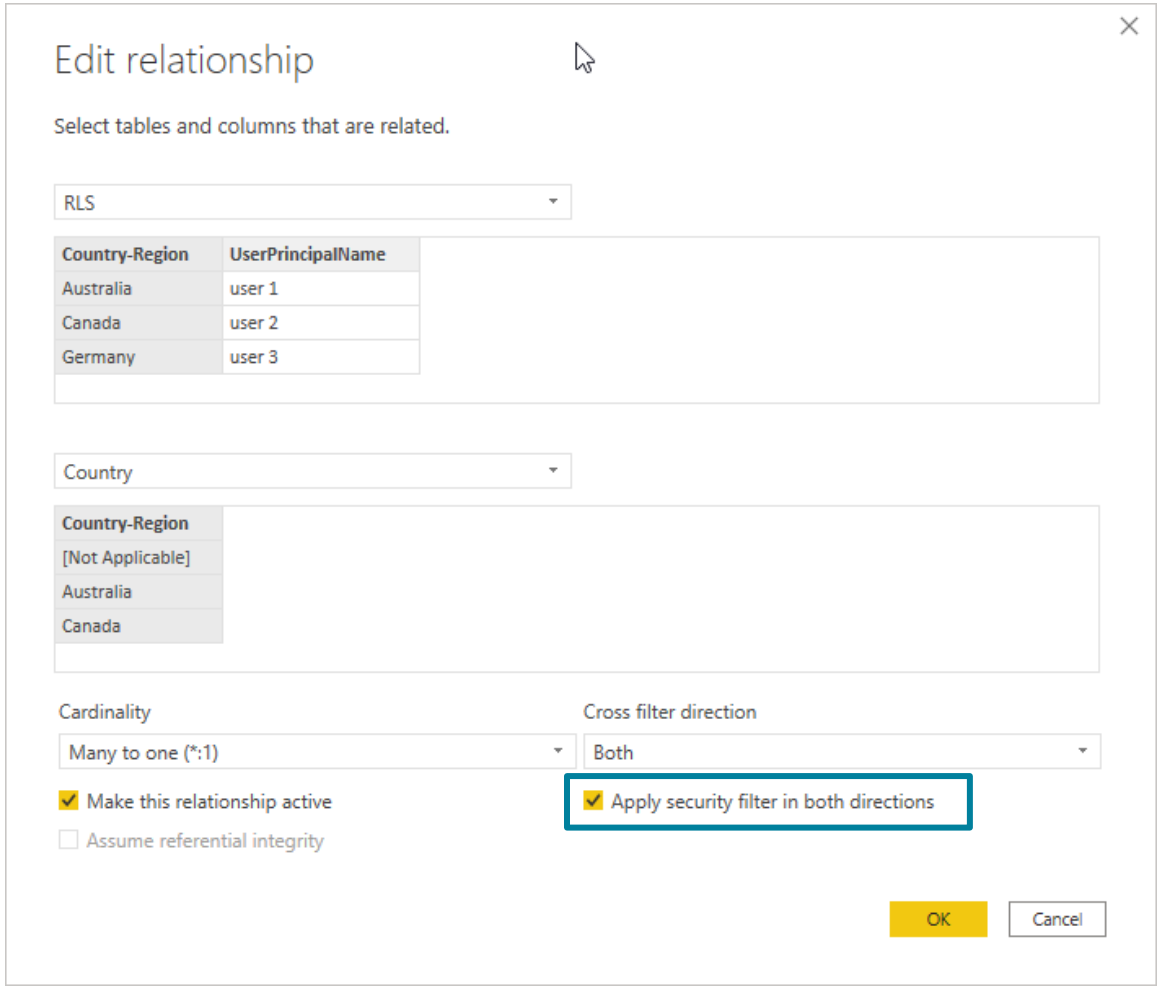
- Except over 1:1 bi-directional relationships
- Possible when adding a RLS filter to each table individually
- Requires a specific setup regarding relationships



Relationships & Row level security

Uses single directional filters by default

- ➔ you can set ‘Apply the Filter Direction when using Row Level security’ on a bi-directional relationship to force RLS to use bi-directional filtering
- ➔ Useful for dynamic row level security



Demo – Row Level Security

Demo time!

Because life is boring without risks

Slowly changing dimensions

Slowly changing dimensions

Although dimension data does not change often (hence ‘slowly’), it might change.

If it does, it’s important to consider historical reporting:

→ A manager switches stores – let’s say they move from A to B. Are all the sales in store A now counted against the manager until they switched? Or against the new manager? What about for the sales in store B?

→ A salesperson switching regions – let’s say they move from South to North.
Are all their sales in South now counted as if they happened in the North region, including the sales that happened before the Salesperson changed Regions?

Let’s talk about

Surrogate key (SK)

- Key generated in the data warehouse
- Unique identifier
- Typically auto-generated (identity)

Alternate key (AK)

Primary key from the source system that loads the data warehouse

Types of attributes on slowly changing dimensions

Type 0: Retain. No changes allowed. Once written, attribute does not change.

Type 1: Overwrite. Does not track historical data.

Customer SK	Customer AK	Name	State
1	C_1	A	WA

Customer SK	Customer AK	Name	State
1	C_1	A	**NY**

Type 2: Add new row. Tracks historical data by creating multiple records for the same alternate key **(most common)**.

Customer SK	Customer AK	Name	State
1	C_1	A	WA

optional

Customer SK	Customer AK	Name	State	Start Date	End Date	Current
1	C_1	A	WA	2000-01-01	***2020-12-31***	***0***
2	**C_1**	**A**	**NY**	**2021-01-01**	***Null***	***1***

Type 3: Limited historical data.

Customer SK	Customer AK	Name	State
1	C_1	A	WA

Customer SK	Customer AK	Name	Original State	Current State	Effective Date
1	C_1	A	WA	**NY**	***2021-01-01***

Slowly changing dimensions in Power BI

Go back to your requirements

If the Salesperson moves – are historical sales still reported in the original Region?

Yes?

Then you're in luck as Power BI already handles this when using type 2 slowly changing dimension.

No?

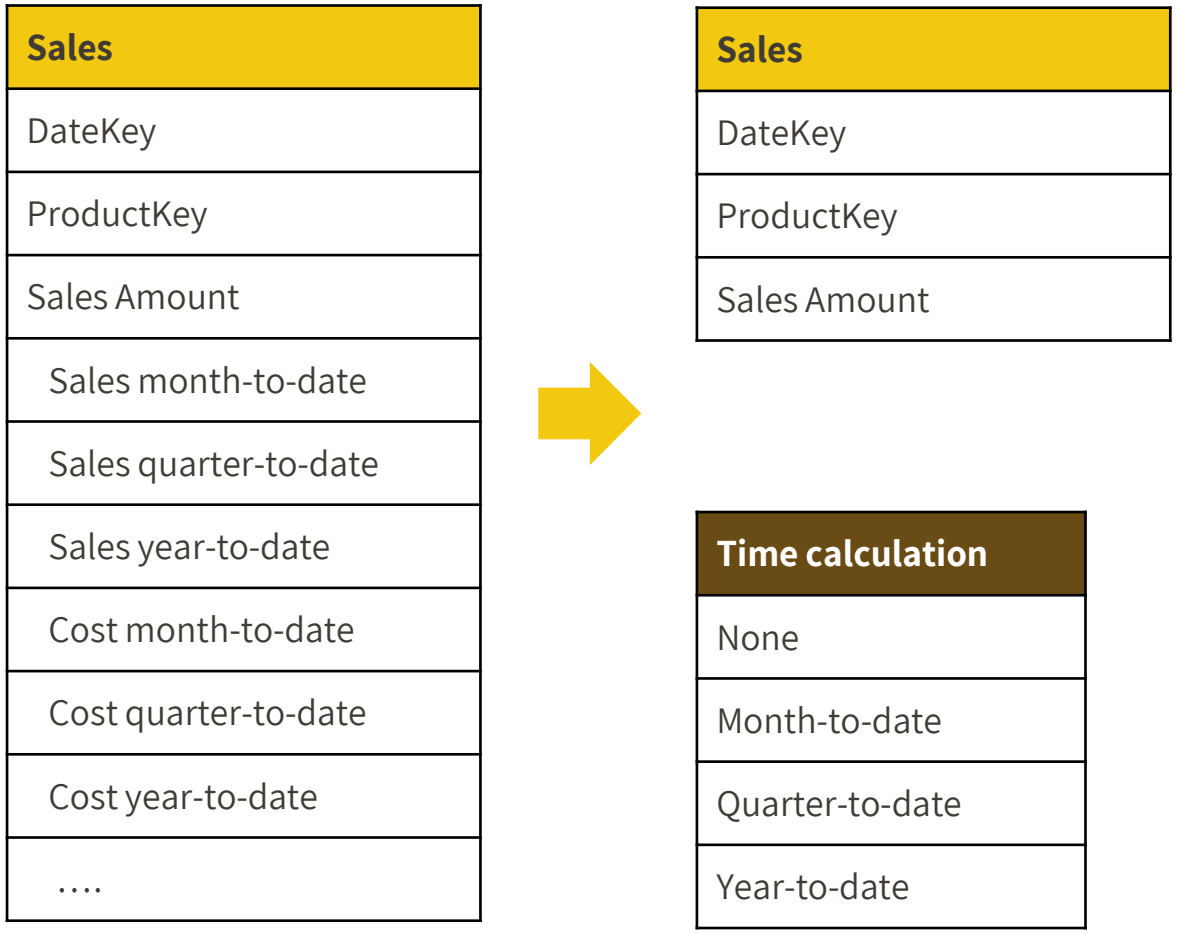
You need to do something:

- Are you sure? This can lead to confusing situations and unhappy sales managers: you might "take away" sales that happened in their region but suddenly start showing as if it always happened in another region.
- Either use type 1 slowly changing dimension (no history).
- Or handle make some changes in the source (or Data Transformation step) and use USERELATIONSHIP.
- Alternatively, you can do this with a calculation group.

Calculation groups

Address issue in complex models where there is a proliferation of redundant measures using the same calculation.

Provides a way to “change the type” of calculation without adding another measure.



Demo – Slowly changing dimensions

Demo time!

Because life is boring without risks

Wrap up

- Use a **star schema** or snowflake data model to get the best out of Power BI.
- Be careful leveraging **bi-directional** relationships
- Avoid **ambiguous** data models
- Consider **slowly changing dimensions** for historical reporting that meets your business requirements

Resources

- **Learning path: Model data in Power BI:**
  https://aka.ms/DataModelingLearningPath
- **Learning path: Use DAX in Power BI Desktop:**
  https://aka.ms/DAXLearningPath
- **DAX function: CROSSFILTER:**
  https://aka.ms/DAXCrossfilter
- **Model relationships in Power BI Desktop:**
  https://aka.ms/ModelRelationships
- **Whitepaper: Bidirectional cross-filtering for Power BI Desktop:**
  https://aka.ms/BidirectionalCrossFilteringWhitepaper
- **Slowly changing dimensions**
  aka.ms/PBISlowlyChangingDimensions
- **Calculation groups in Power BI**
  aka.ms/PowerBICalculationGroups
- **AdventureWorks 2020 demo dataset:**
  https://aka.ms/AW2020Dataset

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 and Marc Lelijveld
Watch now

Webinar Series: Mastering Data Modeling with Power BI Episode 2 - Learn Advanced Data Modeling with Power BI
by Jeroen ter Heerdt and Marc Lelijveld
Watch now

Webinar Series: Mastering Data Modeling with Power BI
Episode 3 - Data Modeling for Experts with Power BI
by Jeroen ter Heerdt and Marc Lelijveld
Register to watch on September 8th, 2021 10AM – 11AM Pacific Time

Find them all here: https://docs.microsoft.com/en-us/power-bi/fundamentals/webinars
or: aka.ms/DataModelingWebinarEpisode{1/2/3}

Thanks for attending!

Marc Lelijveld
Data & Analytics consultant
Macaw Netherlands

Marc.Lelijveld@outlook.com
@MarcLelijveld
linkedin.com/in/MarcLelijveld
Data-Marc.com