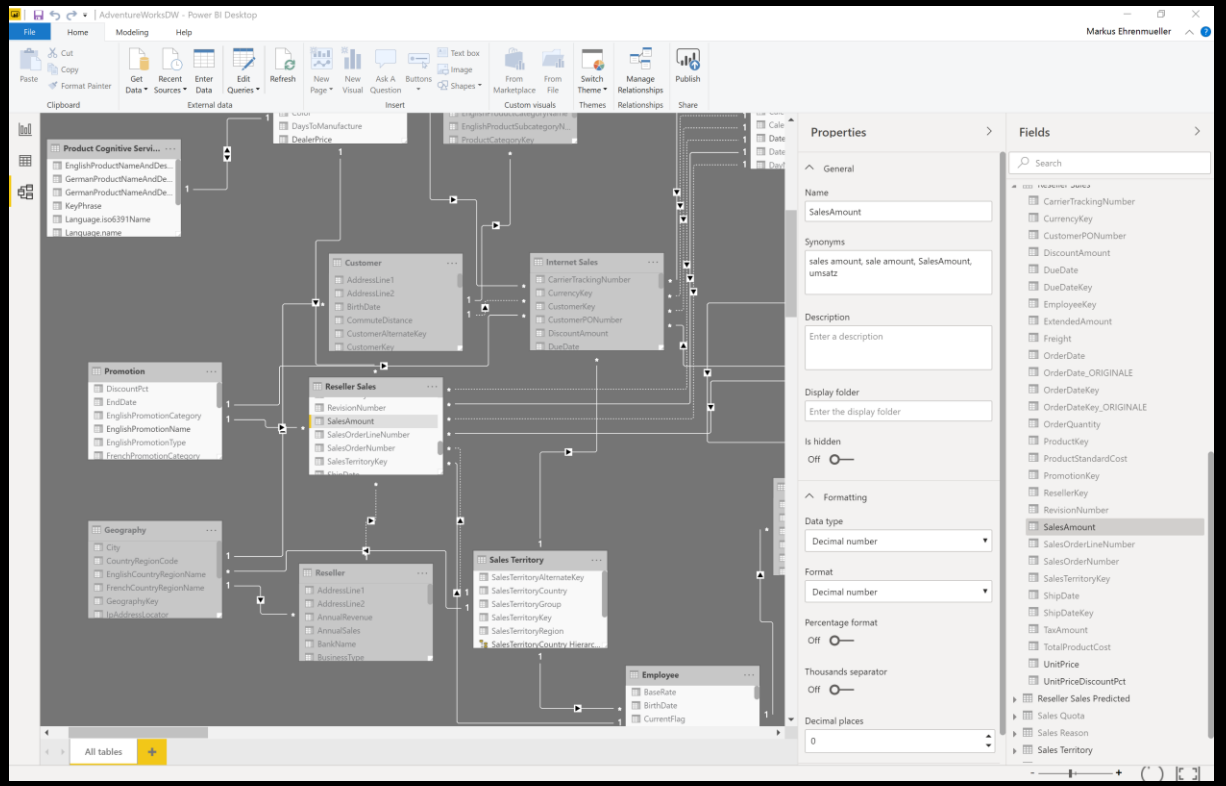
Model Your Data Like a Star

Markus Ehrenmueller-Jensen

I work with models

Other people

Me



Source: https://pxhere.com/en/photo/127022

Power BI Desktop

- Born as Self-Service Business Intelligence Tool
- Model-driven tool
- Tool does educated guesses
  - To make it comfortable to use
- In most of the cases it will work
  - In some cases, there are potential improvements
  - In rare cases guesses will be wrong
  - A good data model is crucial for good reports

Agenda

Relationships
Denormalizing

Agenda

Relationships

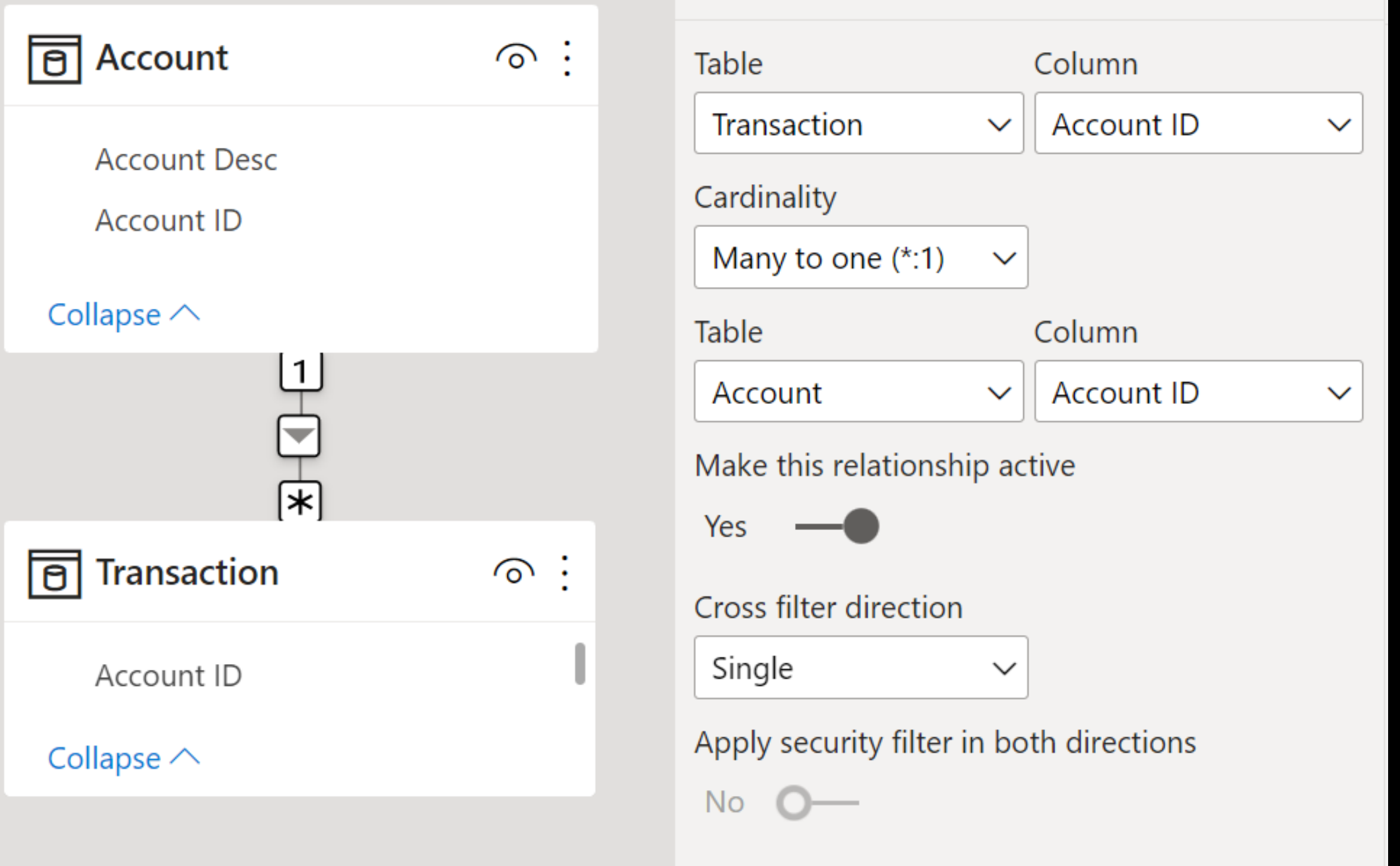
Denormalizing

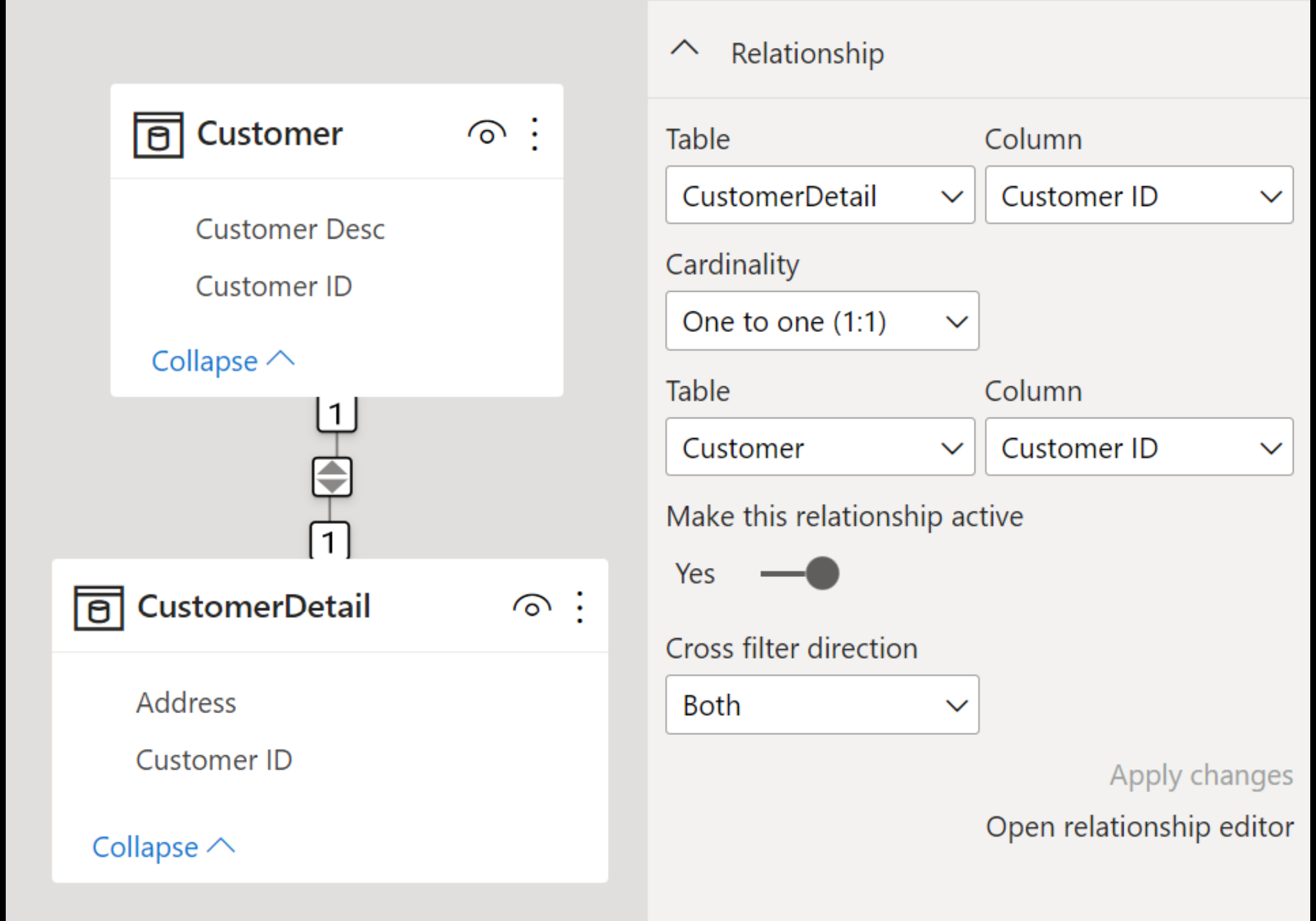
Relationship

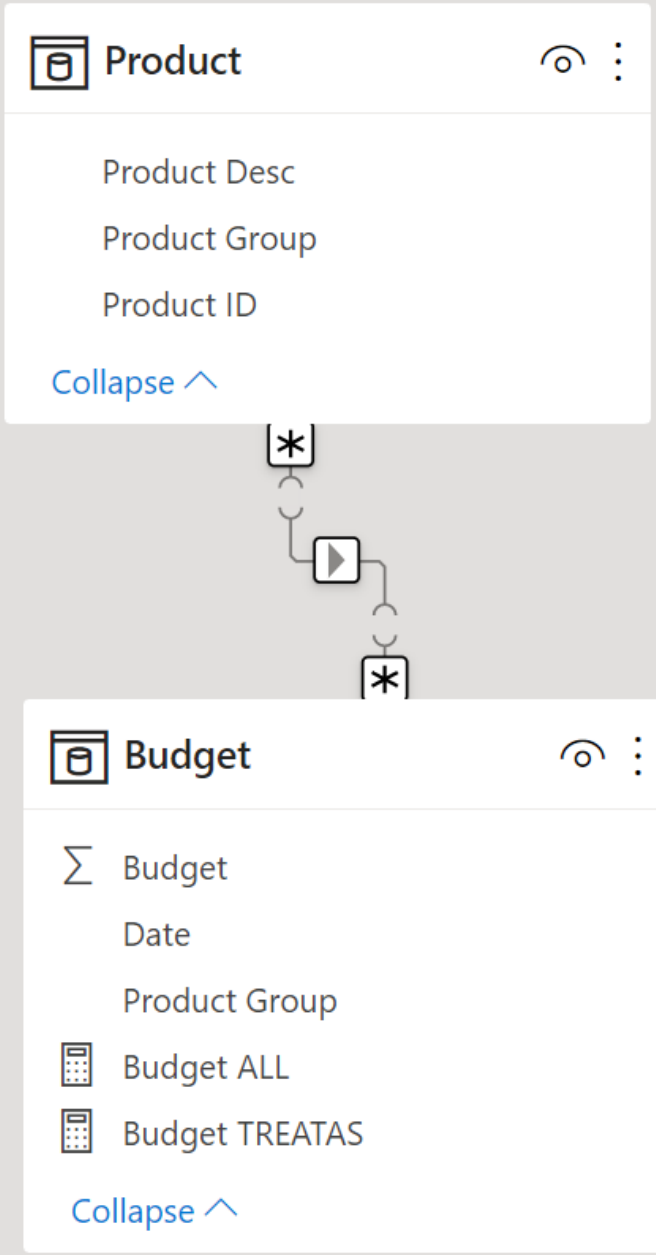
- Like foreign key constraints
  - But not a constraint
  - For filter propagation instead
- Cross filter direction
  - Single or both
- Cardinality
  - One-to-many, one-to-one, many-to-many
  - One: zero or one; column must be unique
  - Many: zero, one or many; column may contain duplicates

Demo

Relationships


Account
Account Desc
Account ID
Collapse
1
*
Transaction
Account ID
Collapse
Relationship
Table
Column
Transaction
Account ID
Cardinality
Many to one (*:1)
Table
Column
Account
Account ID
Make this relationship active
Yes
Cross filter direction
Single
Apply security filter in both directions
No


Customer
Customer Desc
Customer ID
Collapse
1
1
CustomerDetail
Address
Customer ID
Collapse
Relationship
Table
Column
CustomerDetail
Customer ID
Cardinality
One to one (1:1)
Table
Column
Customer
Customer ID
Make this relationship active
Yes
Cross filter direction
Both
Apply changes
Open relationship editor



Relationship

Table: Budget | Column: Product Group

Cardinality: Many to many (*:*)

⚠ This relationship has cardinality Many-Many. This should only be used if it is expected that neither column (and) contains unique values, and that the significantly different behavior of Many-Many relationships is understood. Learn more

Table: Product | Column: Product Group

Make this relationship active: Yes

Cross filter direction: Single

Apply security filter in both directions: No

Agenda

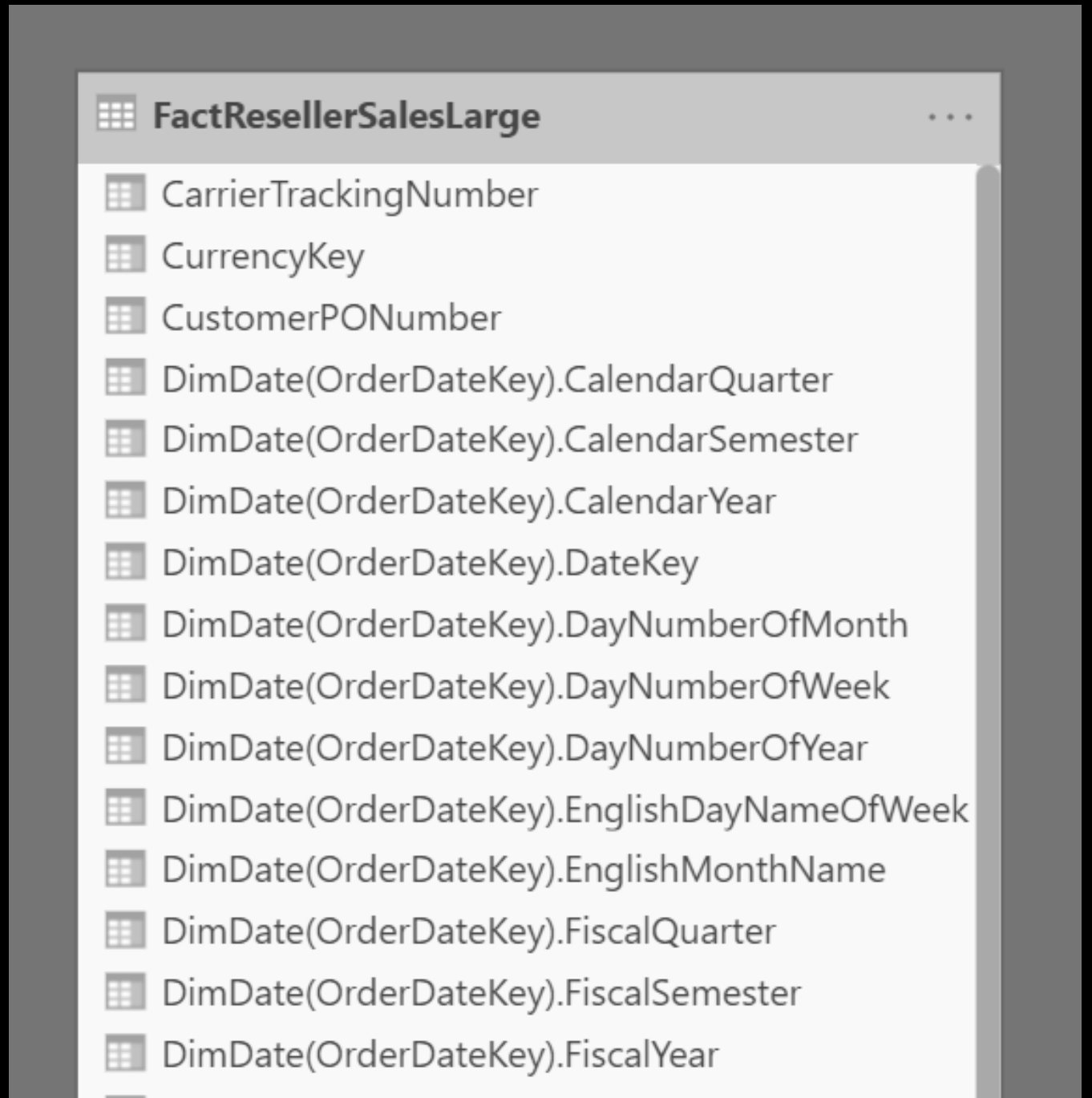
Relationships

Denormalizing

(De-)Normalizing

- Single Table
  - Redundant
- Normalized Schema
  - Avoids redundancy
  - 3NF, BCNF, …
- Denormalized Schema
  - Reintroduces some redundancy
  - Fact vs. Dimension
  - Star vs. Snowflake Schema

Single table



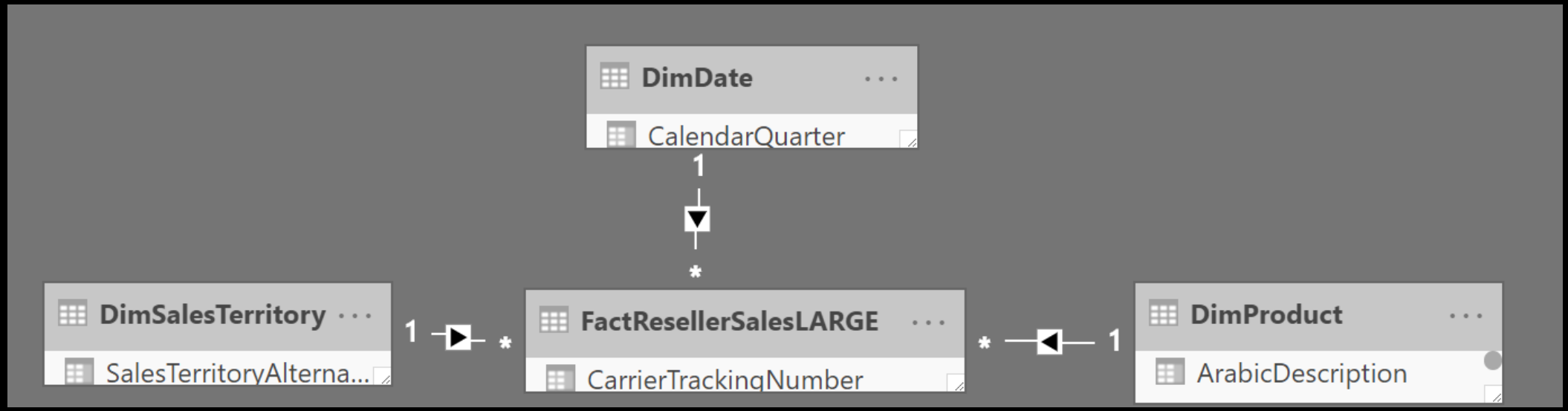
Drawbacks of Single Table

- Worse data compression
- Slower slicers
- Slower refresh
  - Slower throughput
  - Updates to dimensional data requires full refresh
- Only one fact table
  - As there are no (shared) dimension tables

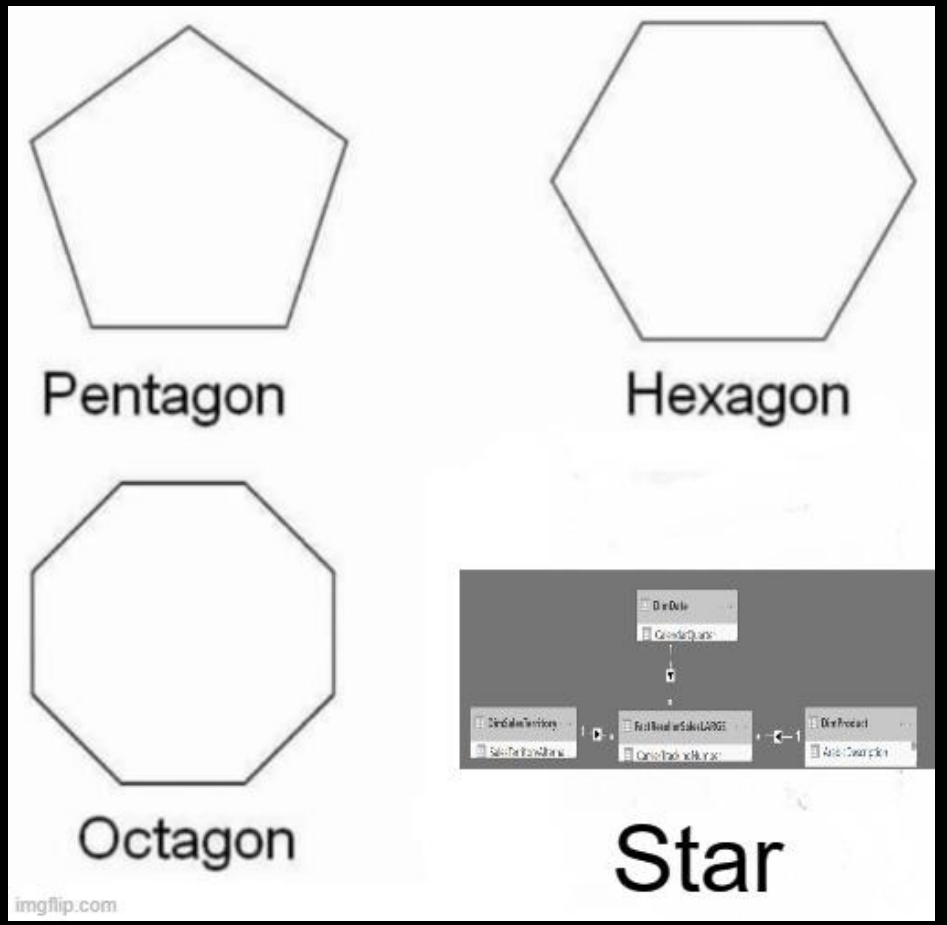
Star & Snowflake Schema

- Dimension
  - How, what, when, where, who, why
  - Scope of model: filtering & grouping
  - Relatively small
  - On the “one” side
  - Primary key + columns of all data types
- Fact
  - Transactions, details, measurements, real world events
  - Meat of model: counting & aggregating
  - Relatively big
  - On the “many” side
  - Foreign keys + numeric columns

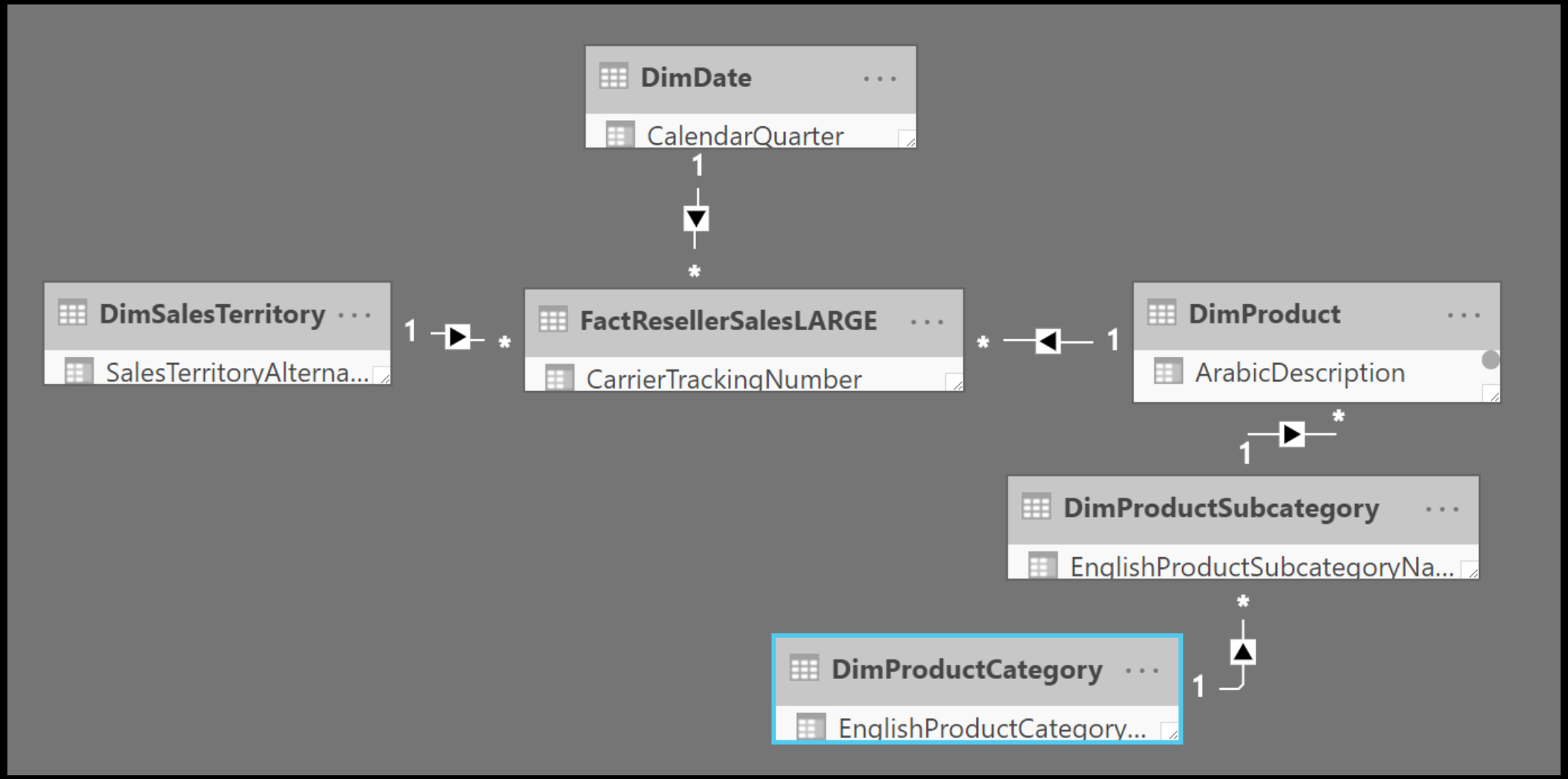
Star schema


DimDate
CalendarQuarter
1
*
DimSalesTerritory
SalesTerritoryAlterna...
1
*
FactResellerSalesLARGE
CarrierTrackingNumber
*
1
DimProduct
ArabicDescription

Star schema



Snowflake schema



Star Schema vs. Snowflake Schema

Snowflake Schema has more tables …

… take longer to load

… make filtering slower

… make fields pane less intuitive

Hierarchies not possible over columns of different tables.

Drawbacks of Star/Snowflake

- Joins on large dimensions are expensive
  - Large := 1 million rows
- Display folders cannot span multiple tables
- Slower on some DirectQuery data sources
  - Spark, Databricks, Azure Data Explorer, …

Comparison (18M facts)

Single table	Star schema	Snowflake schema
1 table	4 tables	6 tables
200 MB total size	84 MB total size	84 MB total size
0 relationships	656 KB relationships	704 KB relationships

Power Query

Filter your data
Choose friendly names
Choose appropriate data types
Flatten your dimensions
Do not flatten facts

Take-aways

Take care of your data model
- Review what Power BI did create for you
- Stay away from many-to-many cardinality
- Stay away from bidirectional cross filters
- Prefer star schema

You can exchange complexity in the model for complexity in DAX

Linklist

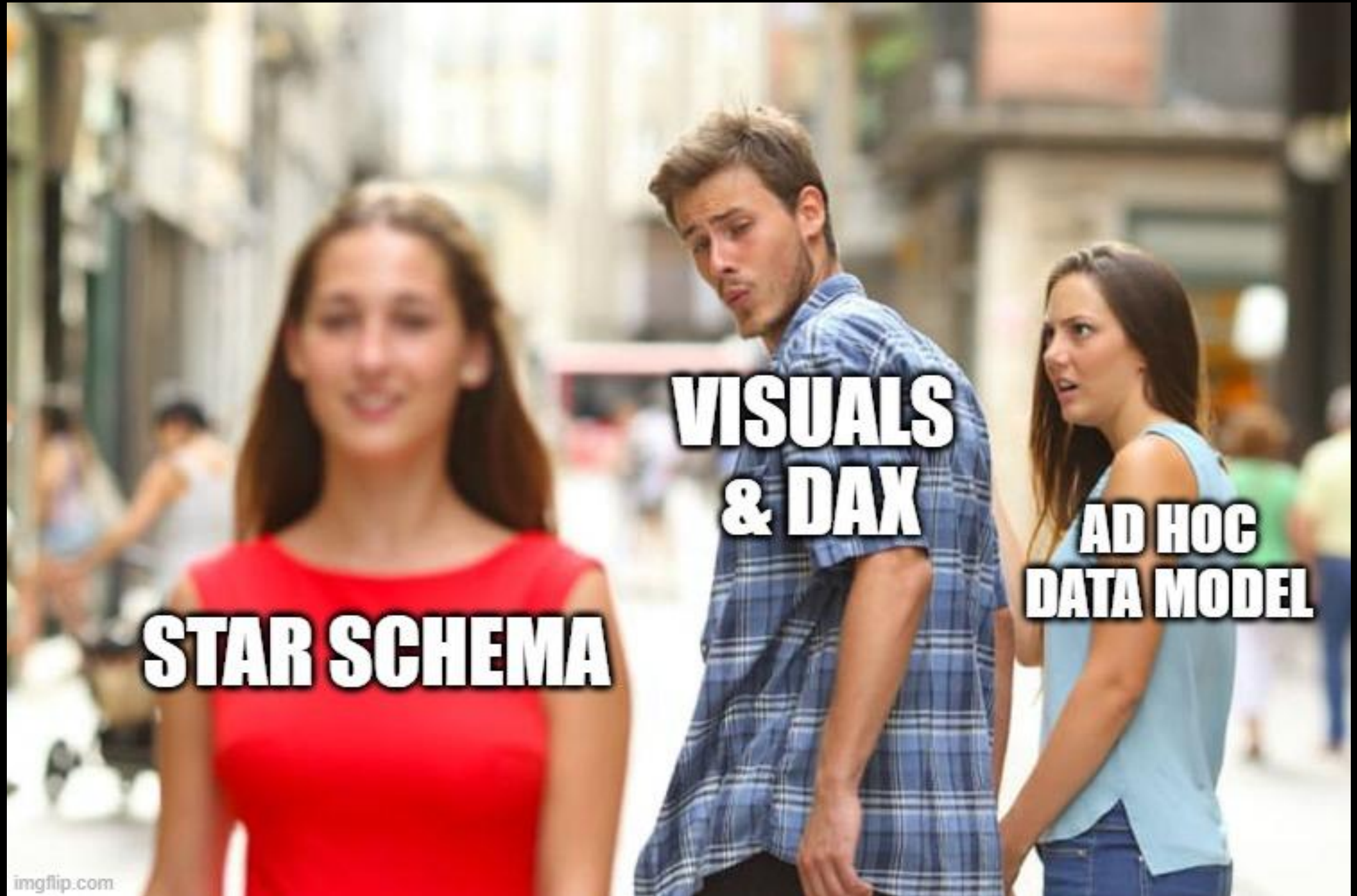
https://docs.microsoft.com/en-us/power-bi/desktop-modeling-view

https://docs.microsoft.com/en-us/power-bi/guidance/star-schema

https://docs.microsoft.com/en-us/power-bi/desktop-relationships-understand

Agenda

Relationships
Denormalizing


VISUALS
& DAX
STAR SCHEMA
AD HOC
DATA MODEL
imgflip.com

Where are the instructions?

New platform for all technical documentation

docs.microsoft.com

All content

open source, hosted on GitHub, community-enabled & in your own language

Help the community of users worldwide

https://aka.ms/intldocs & https://aka.ms/msossloc

Questions?

Savory DATA

Markus Ehrenmueller-Jensen

Founder & BI Architect
markus@savorydata.com
@MEhrenmueller
/markus-ehrenmueller
www.savorydata.com

Microsoft CERTIFIED Solutions Expert
Business Intelligence
Data Platform
Microsoft CERTIFIED Trainer

Microsoft Data Platform

BI Developer
Database Developer
Database Admin

Microsoft MCT

MVP Microsoft® Most Valuable Professional

PowerBIUG POWER BI USER GROUP

PASS AUSTRIA

PASS SQLSATURDAY VIENNA