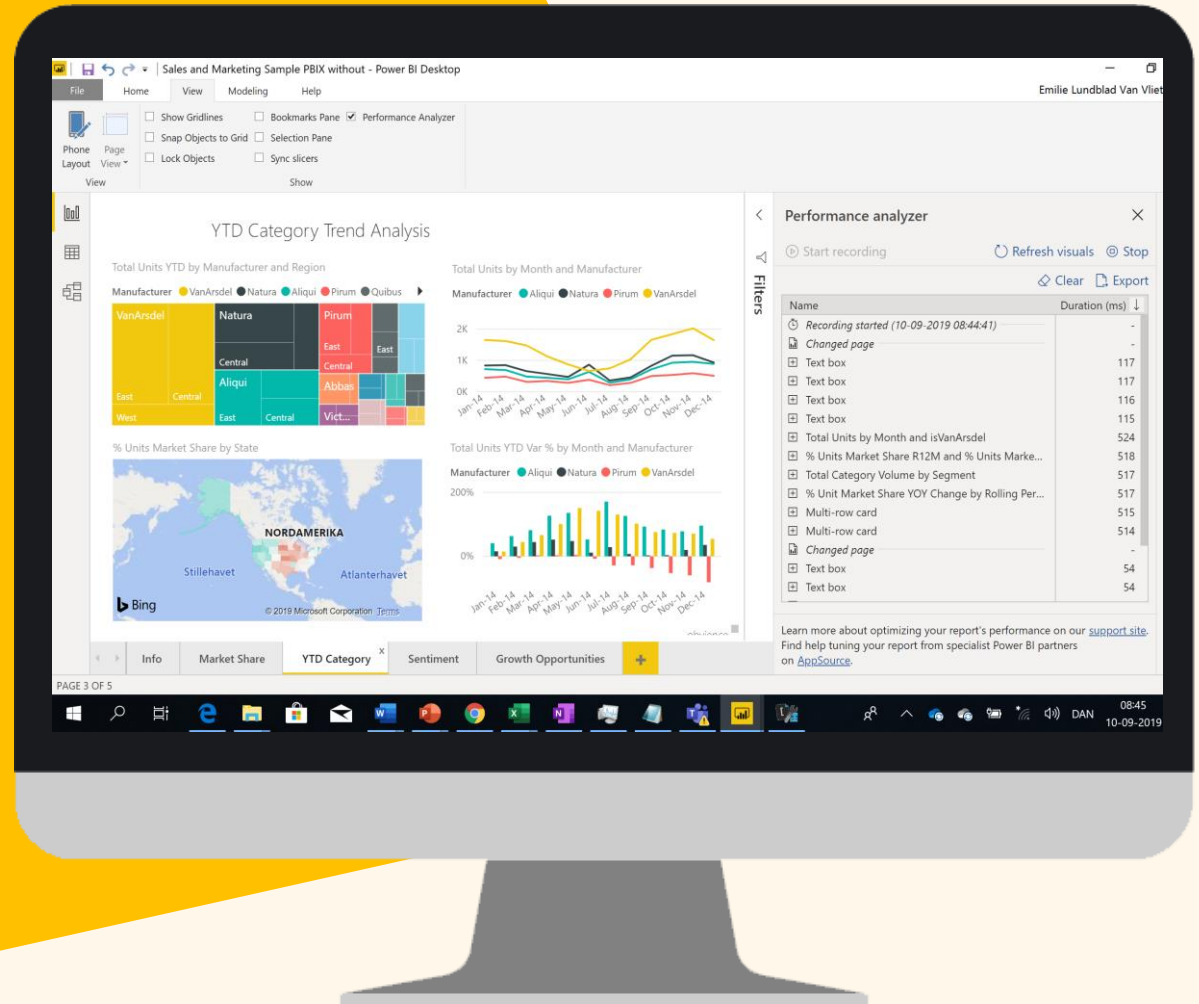


Power BI: 10 tricks til god datamodellering

Emilie Lundblad van Vliet

Data & AI



Agenda

Power BI

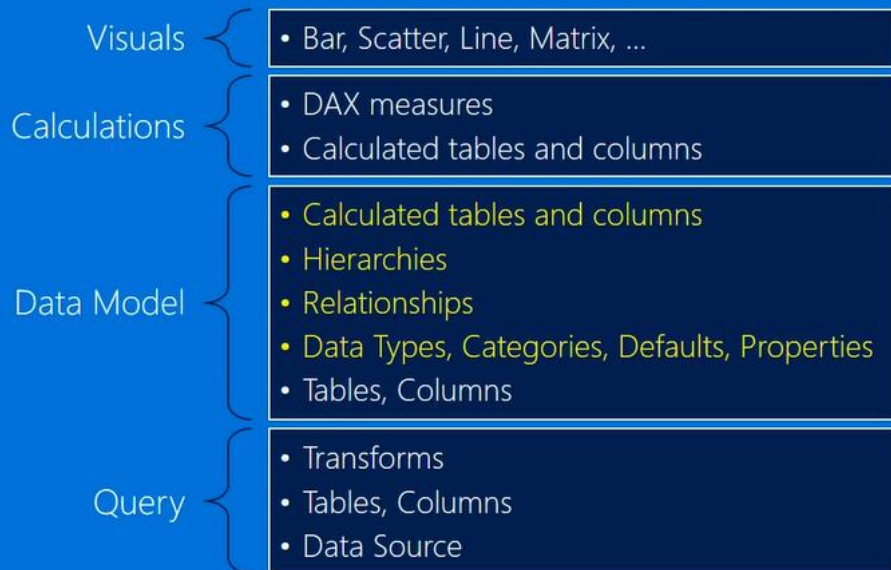
10 tricks til lækre rapporter

1. Datamodel – Hvad fikses hvor?
2. Power Query – Data Load, kvalitet & størrelse
3. Power Query – Tabel format
4. Power Query - M
5. Data modeller – Best practice
6. Star schema – Facts & Dimensioner
7. Forbedring af userbility & Performance
8. Auto date time
9. Dax
10. Performance analyzer

Data modeller – hvad fikses hvor?

Hvilke værktøjer til hvilke lag?

Data Model



A Data Model is like the structural foundation for your house.

Get it right and everything else goes much better.

Power BI Desktop
(Data modelling)

Power Query

Power Query

Query – ETL lag

The screenshot displays the Microsoft Power Query interface. The top ribbon is shown with the 'Edit Queries' button highlighted by a yellow box and a yellow arrow pointing to the 'Transform' ribbon. The 'Transform' ribbon includes various options like 'Close & Apply', 'New Source', 'Recent Sources', 'Data source settings', 'Manage Parameters', 'Refresh Preview', 'Properties', 'Advanced Editor', 'Manage', 'Choose Columns', 'Remove Columns', 'Keep Rows', 'Remove Rows', 'Sort', 'Split Column', 'Group By', 'Data Type: Text', 'Use First Row as Headers', 'Replace Values', 'Merge Queries', 'Append Queries', and 'Combine Files'.

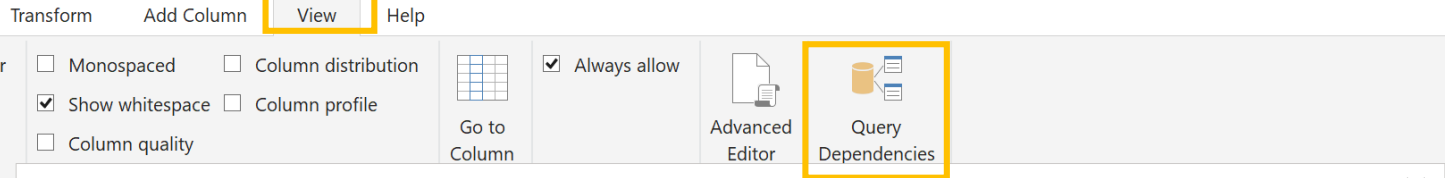
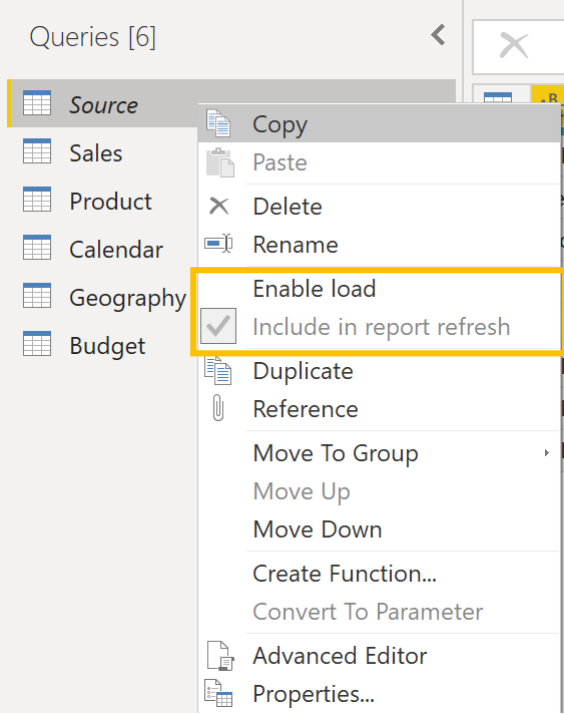
The 'Queries [11]' pane on the left shows a list of queries, with 'Source' selected. The main area displays a table with the following data:

	A ^B C Name	Data	A ^B C Item	A ^B C Kind	Hidden
1	Sales	Table	Sales	Sheet	
2	Geography	Table	Geography	Sheet	
3	Product	Table	Product	Sheet	
4	Calendar	Table	Calendar	Sheet	
5	_xlnm...FilterDatabase	Table	Geography!_xlnm...FilterDatabase	DefinedName	

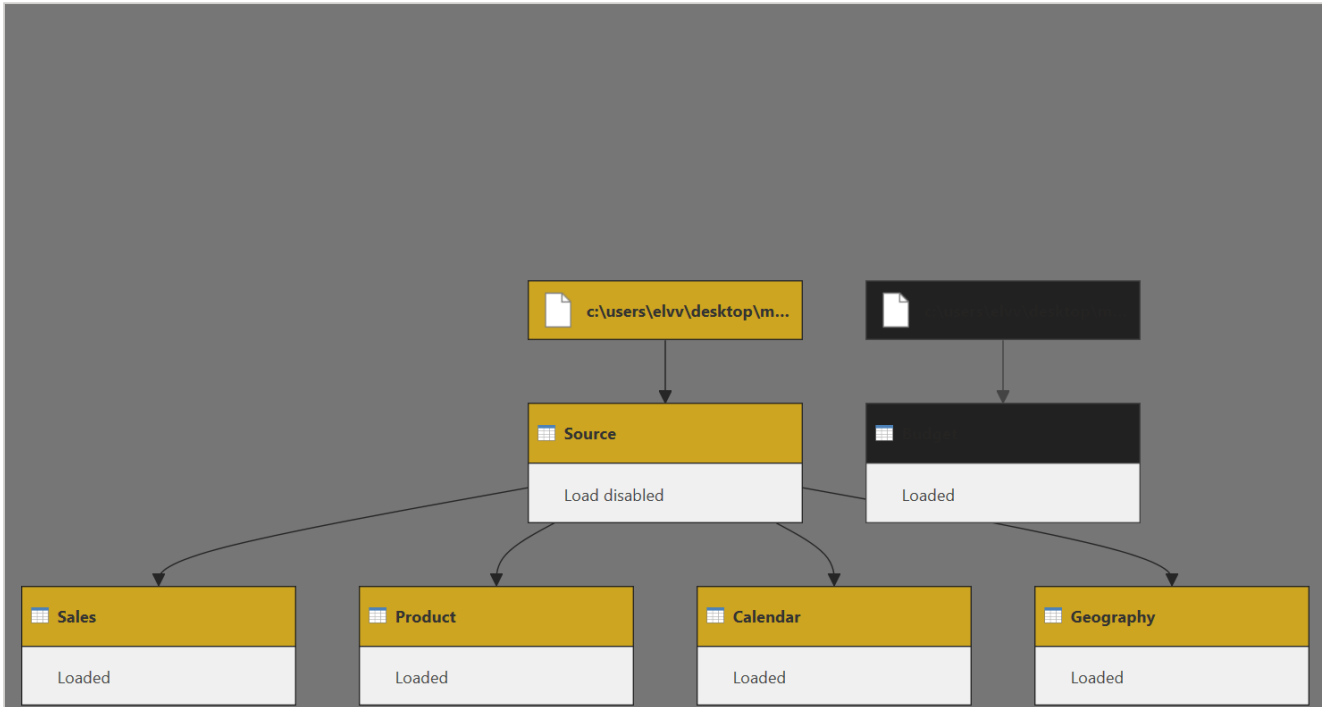
The 'Query Settings' pane on the right shows the 'PROPERTIES' section with the 'Name' field set to 'Source' and the 'APPLIED STEPS' section with a single step named 'Source'.

Power Query

Hurtigere Load i power Query



Query Dependencies



Power Query

Dårlig data kvalitet?

The screenshot displays the Microsoft Power Query interface. The 'Data Preview' tab is active, showing a table with columns: ProductID, Date, Zip, Units, and Revenue. A yellow box highlights the 'Data Preview' ribbon, which includes options for 'Monospaced', 'Column distribution', 'Show whitespace', 'Column profile', 'Column quality', 'Always allow', 'Go to Column', 'Columns', 'Parameters', 'Advanced Editor', and 'Query Dependencies'. The 'Column quality' section is expanded, showing quality metrics for each column. The 'ProductID' column has 100% valid data. The 'Date' column has 100% valid data. The 'Zip' column has 100% valid data. The 'Units' column has 100% valid data. The 'Revenue' column has 100% valid data. The 'Data Preview' ribbon also shows a 'Query Settings' icon and a 'Layout' button.

Column	Valid	Error	Empty	Distinct	Unique
ProductID	100 %	0 %	0 %	1	0
Date	100 %	0 %	0 %	5	0
Zip	100 %	0 %	0 %	946	895
Units	100 %	0 %	0 %	1	0
Revenue	100 %	0 %	0 %	1	0

1 distinct, 0 unique

5 distinct, 0 unique

946 distinct, 895 unique

1 distinct, 0 unique

1 distinct, 0 unique

Power Query

Dårlig data kvalitet?

Formula bar: = Source{[Item="Budget",Kind="Sheet"]}[Data]

Column1
1 Please manually input budget revenue numbers in this sh
2 Author
3 Budget version
4
5 GeolD
6 Canada, R2C
7 Germany, 12101
8 France, 6500
9 Mexico, 2000
10 USA, 8052

Transform Add Column View Help

Any Column: Transpose, Reverse Rows, Count Rows, Detect Data Type, Rename, Pivot Column, Convert to List, Replace Values, Fill, Unpivot Columns, Move, Merge Columns

Text Column: Split Column, Format, Parse, Extract

- NULL værdier – erstat med 0, Blank eller filtre ud
- Remove errors – god quick & dirty løsning
- Replace errors – bedre mere rigtig løsning
- Remove duplicates – god når man skal skabe dimensioner
- Split er god til dato tid eller kolonner med flere data i
- Group by er god til tidsserier, hvor man skal kunne finde første og sidste record for en kunde eller lignende

Power Query

Reducér data Power Query

Power Query

Fjern kolonner som ikke bruges;

- ID
- Forretningsnøgler / Businesskeys /Key
- Tekst beskrivelser (Strings)
- Date Created, Changed ect.

Reducér data størrelse;

- Split dato / tid
- Fjern tid, hvis ikke nødvendig

	Date
1.2	Decimal Number
\$	Fixed decimal number
123	Whole Number
%	Percentage
📅	Date/Time
📅	Date
🕒	Time
🌐🕒	Date/Time/Timezone
🕒	Duration
ABC	Text
☒	True/False
☐	Binary
	Using Locale...
7	07-01-2012 00:00:00
8	08-01-2012 00:00:00
9	09-01-2012 00:00:00
10	10-01-2012 00:00:00

	Choose Columns...
1	Choose Columns...
2	Keep Top Rows...
3	Keep Bottom Rows...
4	Keep Range of Rows...
5	Keep Duplicates
6	Keep Errors
7	Remove Top Rows...
8	Remove Bottom Rows...
9	Remove Alternate Rows...
10	Remove Duplicates
11	Remove Errors
12	Merge Queries...
13	Append Queries...

Choose Columns

Choose the columns to keep

- (Select All Columns)
- Source.Name
- Transform File from International.ProductID
- Transform File from International.Date
- Transform File from International.Zip
- Transform File from International.Units
- Transform File from International.Revenue
- Transform File from International.Country

OK

Cancel

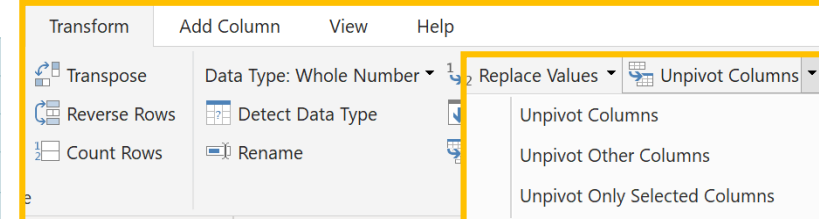
Power Query

Tabel format – Power BI er IKKE Excel

Power BI er ikke Excel, undgå derfor brede tabeller:

- Dårlig for performance og tager hukommelse
- Giver problemer med at slice og dice
- Dårlig til Dax beregninger
- Problemer med relationer (mange til mange & tovejs filtrering)
- Typisk eksempel Excel budget
- Løsningen er ofte pivotering

GeoID	01-01-2017	01-02-2017	01-03-2017	01-04-2017	01-05-2017	01-06-2017
Canada, R2C	128000	112000	144000	80000	96000	176000
Germany, 12101	300000	180000	480000	540000	780000	600000
France, 6500	248000	682000	248000	310000	372000	558000
Mexico, 2000	660000	165000	275000	495000	440000	440000
USA, 8052	20000000	8000000	14000000	18000000		



	GeoID	Date	BudgetRevenue
1	Canada, R2C	01-01-2017	128000
2	Canada, R2C	01-02-2017	112000
3	Canada, R2C	01-03-2017	144000
4	Canada, R2C	01-04-2017	80000
5	Canada, R2C	01-05-2017	96000
6	Canada, R2C	01-06-2017	176000
7	Canada, R2C	01-07-2017	48000

Power Query

M – hvad er det

```
let
Source = Excel.Workbook(File.Contents("C:\Users\elvv\Desktop\Masterclass\Medium Masterclass Power BI Masterclass USB Files\Data & Assets\S
Budget_Sheet = Source[Item="Budget",Kind="Sheet"])[Data],
#"Promoted Headers" = Table.PromoteHeaders(Budget_Sheet, [PromoteAllScalars=true]),
#"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"Please manually input budget revenue numbers in this sheet..", type te
#"Removed Top Rows" = Table.Skip(#"Changed Type",3),
#"Promoted Headers1" = Table.PromoteHeaders(#"Removed Top Rows", [PromoteAllScalars=true]),
#"Changed Type1" = Table.TransformColumnTypes(#"Promoted Headers1",{{"GeoID", type text}, {"1/1/2012", type number}, {"2/1/2012", type num
#"Unpivoted Other Columns" = Table.UnpivotOtherColumns(#"Changed Type1", {"GeoID"}, "Attribute", "Value"),
#"Changed Type2" = Table.TransformColumnTypes(#"Unpivoted Other Columns",{{"Attribute", type date}}),
#"Renamed Columns" = Table.RenameColumns(#"Changed Type2",{{"Attribute", "Date"}, {"Value", "Budget"}})
in
#"Renamed Columns"
```

Husk Advanced editor har ingen ctrl + z
Dupliker derfor altid query før du
begynder at rette direkte i M koden.

Query Settings

PROPERTIES

Name
Budget

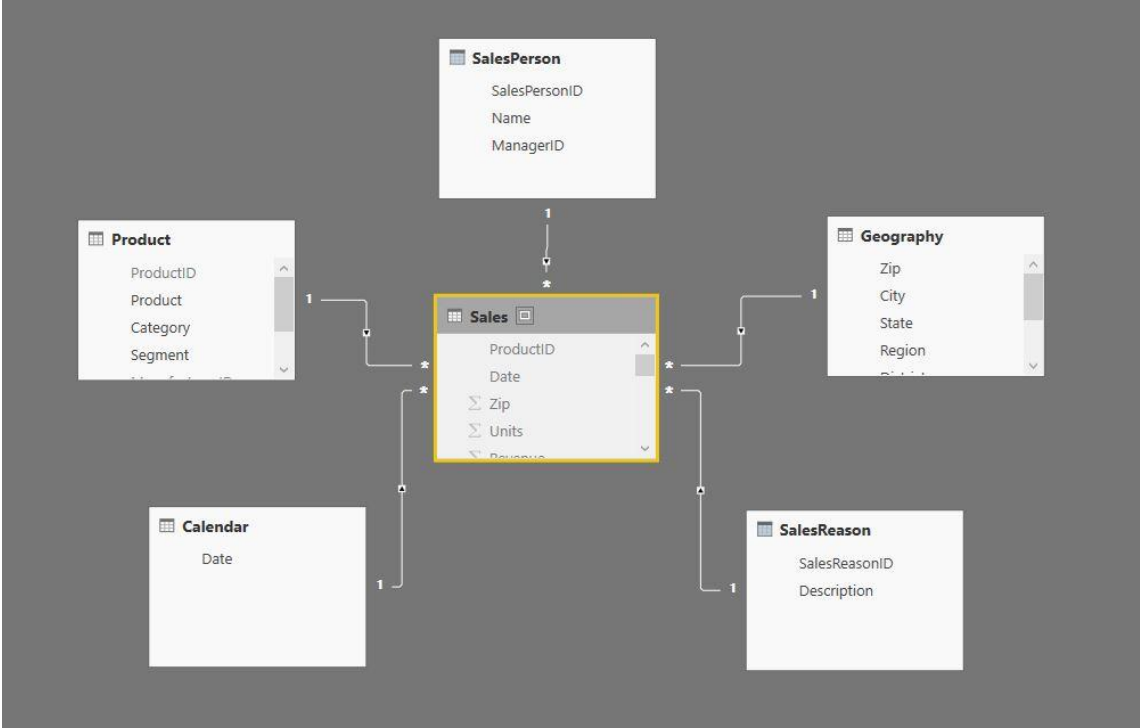
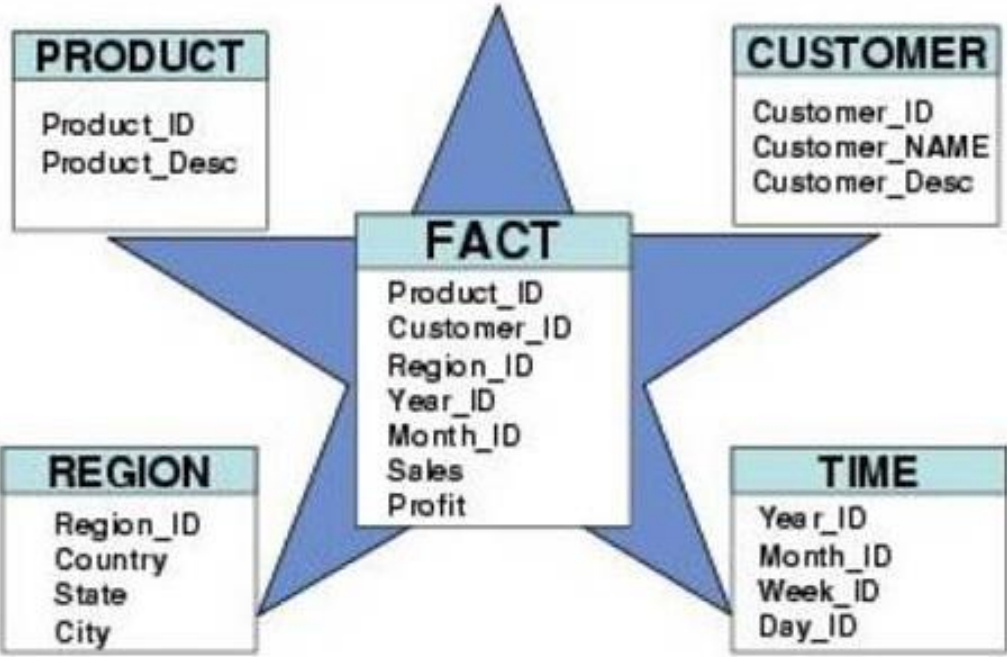
[All Properties](#)

APPLIED STEPS

- Source
- Navigation
- Promoted Headers
- Changed Type
- Removed Top Rows
- Promoted Headers1
- Changed Type1
- Unpivoted Other Columns
- Changed Type2
- Renamed Columns**

Power BI Desktop – model lag

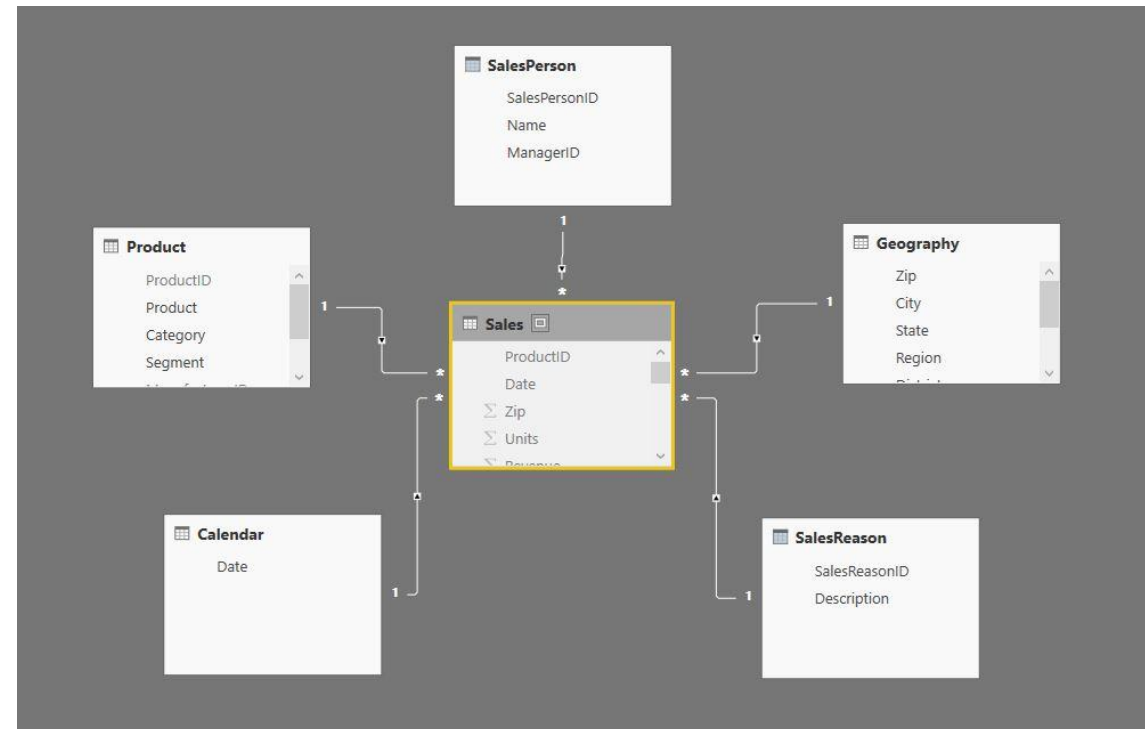
Data modelling – star schema



Power BI Desktop - Data model lag

Dimensions modellering

- “Hvem, hvad, hvor, hvornår, hvorfor og hvordan”
- En fact er fakta om processer, oftest forretningsprocesser.
For økonomer; tidsserier som ændre sig over tid.
Hvordan = Salg / Sales
- Dimensionerne er de dimensioner som danner rammen omkring fact'en / tidsserien, som består gennem tiden;
Hvor = geografi el. Location
Hvem = Salesperson / Kunde
Hvorfor = Salesreason
Hvornår = Calendar / Kalender
Hvad = Product



Power BI Desktop

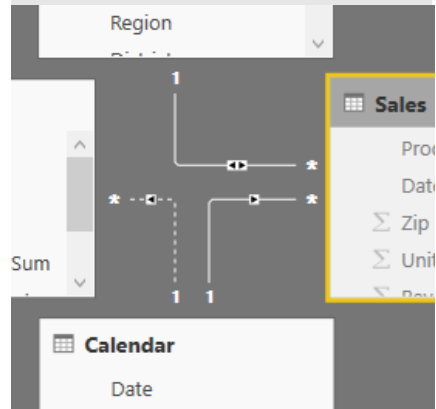
Værktøjer som kan forbedre model

Data Typer

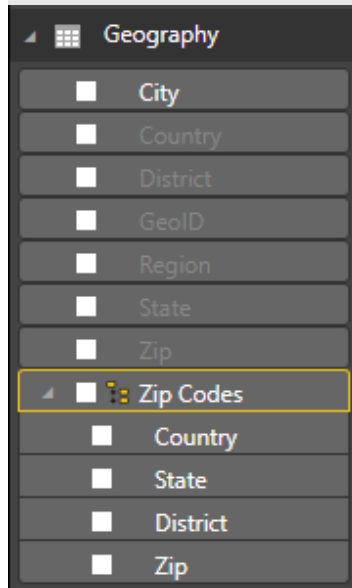
Data type: Whole Number ▾

- Decimal Number
- Fixed decimal number
- Whole Number
- Date/Time
- Date
- Time
- Text
- True/False
- Binary

Relationer



Hierarkier



Beregninger (Measures)

```
1 Protected Budget = SUMX (
2   KEEPFILTERS (
3     CROSSJOIN (
4       VALUES ( Customer[CountryRegion] );
5       VALUES ( 'Product'[Brand] )
6     )
7   );
8   CALCULATE (
9     SUM ( 'Allocated Budget'[Budget] ) * [Perc]
10  )
11 )
```

Power BI Desktop

Værktøjer som kan forbedre model

Data Typer

- > Performance, størrelse og hukommelse skal overvejes.
- > Formatering
- > Kategorier
 - > Bedre funktionalitet
- > Summarization
 - > Usability

Relationer

- > Typer af relationer
 - Two-way relationer pros/cons
 - > God til Sikkerhed & pris/valuta omregninger
- > Role-playing
 - > Indaktive relationer (Lookup)
- > Mange til Mange
 - > Lad vær!

Hierarkier

- > Giver viden om data strukturer og forretnings logik
- > Forbedre Navigation
- > Usability
- > Parent-Child relationer
 - > The Good, Bad and Ugly

Beregninger (Measures)

- > Forretnings behov
 - > Filtreringer
 - > Fikse "dårlig" modellering?
 - > Iterator funktioner
 - > SUMX vs SUM
 - > Performance!!!
 - > iterativ proces

Power BI Desktop - Indstillinger

Auto date time Pro's & Con's

Options

GLOBAL

- Data Load
- Power Query Editor
- DirectQuery
- R scripting
- Python scripting
- Security
- Privacy
- Regional Settings
- Updates
- Usage Data
- Diagnostics
- Preview features
- Auto recovery
- Report settings

CURRENT FILE

- Data Load
- Regional Settings
- Privacy
- Auto recovery
- DirectQuery
- Query reduction

Type Detection

Automatically detect column types and headers for unstructured sources

Relationships

- Import relationships from data sources on first load
- Update or delete relationships when refreshing data
- Autodetect new relationships after data is loaded

[Learn more](#)

Time intelligence

Auto date/time

Background Data

Allow data preview to download in the background

Parallel loading of tables

Enable parallel loading of tables

Q&A

Turn on Q&A to ask natural language questions about your data

OK Cancel

Pro's:

- Nem måde at få dato hierarki, især som ny til Power BI
- Hurtig løsning

Con's:

- Forøger modelstørrelse markant (op til 50%)
- Forvirrende for bruger med flere dato hierkier
- Giver dårlig performance på større modeller

Calendar

- Date
- Date Hierarc...
- Year
- Quarter
- Month
- Day

	Masterclass model med auto date time	09-09-2019	Microsoft.Mi...	6.326 KB
	Masterclass model	09-09-2019	Microsoft.Mi...	5.543 KB

DAX

Dax – do's & don'ts

Columns vs. Measures

Brug kolonner når;

- Brug for at kunne slice og filtrere

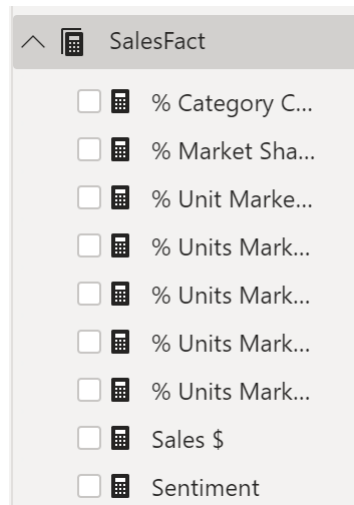
Brug measures

- Procent beregninger
- Brøk beregninger
- Komplekse beregninger (Valuta, calculate, periode beregninger)

Hukommelse og CPU

- Kolonner bruger hukommelse
- Measures bruger CPU

- Measure tabeller?
- Pro pænt måde at holde styr på beregninger
- Con virker dårligt med nye PBI ML features som key influencer og quick insights



Data model best practice

Hvad er bedst hvor? DAX vs. M (Power Query)

Brug DAX

- DAX kan godt "Wrangle" efter load til Power BI Desktop
- DAX laver ændringer/beregninger når de bruges, dette burde hukommelse, hvilket rammer model performance
- DAX kan referer til data overalt i modellen.
- DAX er designet til analytics

Brug M

- M er til ETL / "wrangle" data ved load
- M skubber "wrangling" så lang ned mod kilden som muligt (data folding).
Kan ramme performance på data kilde, (Derfor source query med ref.)
- M kan kun referer til kolonner i same tabel (Ellers skal man flette eller tilføje)
- M is designet til ETL/ wrangling af data

Reducér model og data

Performance analyzer

Sales and Marketing Sample PBIX without - Power BI Desktop

File Home View Modeling Help

Phone Layout Page View View Show

Show Gridlines Show Bookmarks Pane Performance Analyzer Snap Objects to Grid Selection Pane Lock Objects Sync slicers

YTD Category Trend Analysis

Total Units YTD by Manufacturer and Region

Manufacturer VanArsdel Natura Aliqui Pirum Quibus

Total Units by Month and Manufacturer

Manufacturer Aliqui Natura Pirum VanArsdel

% Units Market Share by State

Total Units YTD Var % by Month and Manufacturer

Manufacturer Aliqui Natura Pirum VanArsdel

Performance analyzer

Start recording Refresh visuals Stop

Clear Export

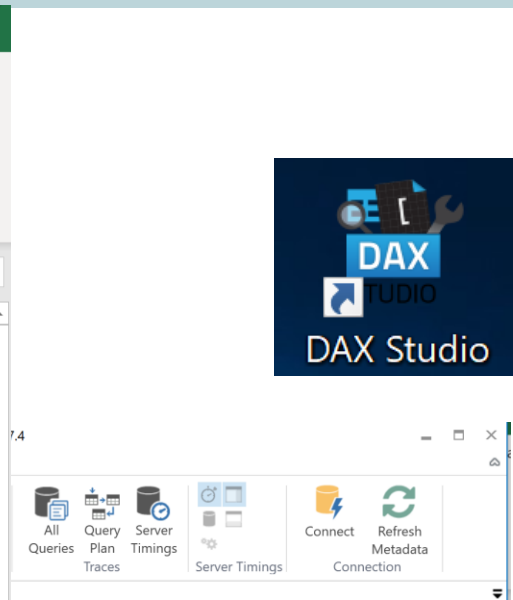
Name	Duration (ms)
Recording started (10-09-2019 08:44:41)	-
Changed page	-
Text box	117
Text box	117
Text box	116
Text box	115
Total Units by Month and isVanArsdel	524
% Units Market Share R12M and % Units Marke...	518
Total Category Volume by Segment	517
% Unit Market Share YOY Change by Rolling Per...	517
Multi-row card	515
Multi-row card	514
Changed page	-
Text box	54
Text box	54

Learn more about optimizing your report's performance on our [support site](#). Find help tuning your report from specialist Power BI partners on [AppSource](#).

Reducér model og data

Dax studio & Vertipac analyzer

Rækkemærkater	Cardinality	Table Size	Columns	Total Size	Data Size	Dictionary Size	Columns Hierarchies Size	Encoding	User Hierarchies Size	Relationships Size	Table Size %	Database Size %	Segn
Date	6.209	675.980		571.980	45.920	468.540	57.520	Many	104.000			5,63 %	
Geo	39.948	3.793.852		2.696.668	156.128	1.991.420	549.120	Many	1.097.184			31,60 %	
Manufacturer	14	34.992		34.992	24	34.648	320	Many				0,29 %	
Product	2.412	330.659		295.667	9.320	248.827	37.520	Many	34.976	16		2,75 %	
SalesFact	1.260.752	7.030.285		6.951.733	5.195.320	1.408.621	347.792	Many		78.552		58,56 %	
Sentiment	21.473	139.532		139.356	71.080	65.428	2.848	Many		176		1,16 %	
Hovedtotal	1.330.808	12.005.300		10.690.396	5.477.792	4.217.484	995.120	Many	1.236.160	78.744		100,00 %	



<https://www.sqlbi.com/tools/vertipaq-analyzer/>
<https://daxstudio.org/>