

Basics of building a data model in Power BI

Jason Romans

Thank you, Sponsors!



Quest



I D E R A



Join Our
Local User
Groups:

SQLOrlando



Jason Romans

Builder of Models

Senior BI Engineer

thedaxshepherd@gmail.com

<https://thedaxshepherd.net/>

<https://www.linkedin.com/in/jason-r-sql-jar>



- *Live in Nashville, TN*
- *Started as a SQL Server DBA*
- *Transitioned to the Microsoft BI Stack*
- *Work on everything from SQL Server Integration Services, SQL Server Database, Analysis Services, and Power BI*

Fabric Care



What is a model?

A Tabular model is a set of metadata like tables, relationships, measures, KPIs, calculation groups, hierarchies, translations, security roles, and many other elements that form the “semantic model” used to provide a navigation system in client tools like Power BI and Excel reports.

-- [SQLBI](#)

Daunting Task

- How do I bring in the data
- What should the model look like
- Get Data – Model Data – Visualize Data

What you
picture



Reality

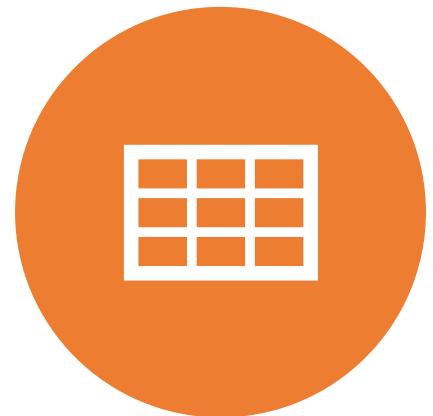


Instructions

- Unfortunately, no step-by-step instructions
- Bulk Lego blocks
- Guidelines

Get Data

Where is the data



EXCEL SPREADSHEET

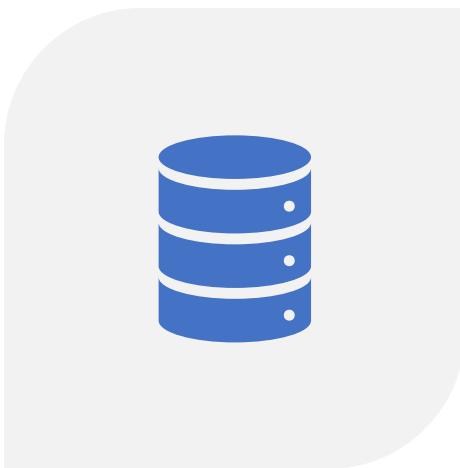


SQL SERVER

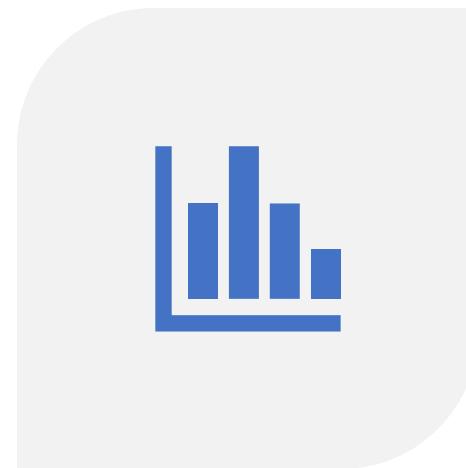


CSV

What form is the data in



SOURCE DATA THAT IS IN
SEPARATE TABLES

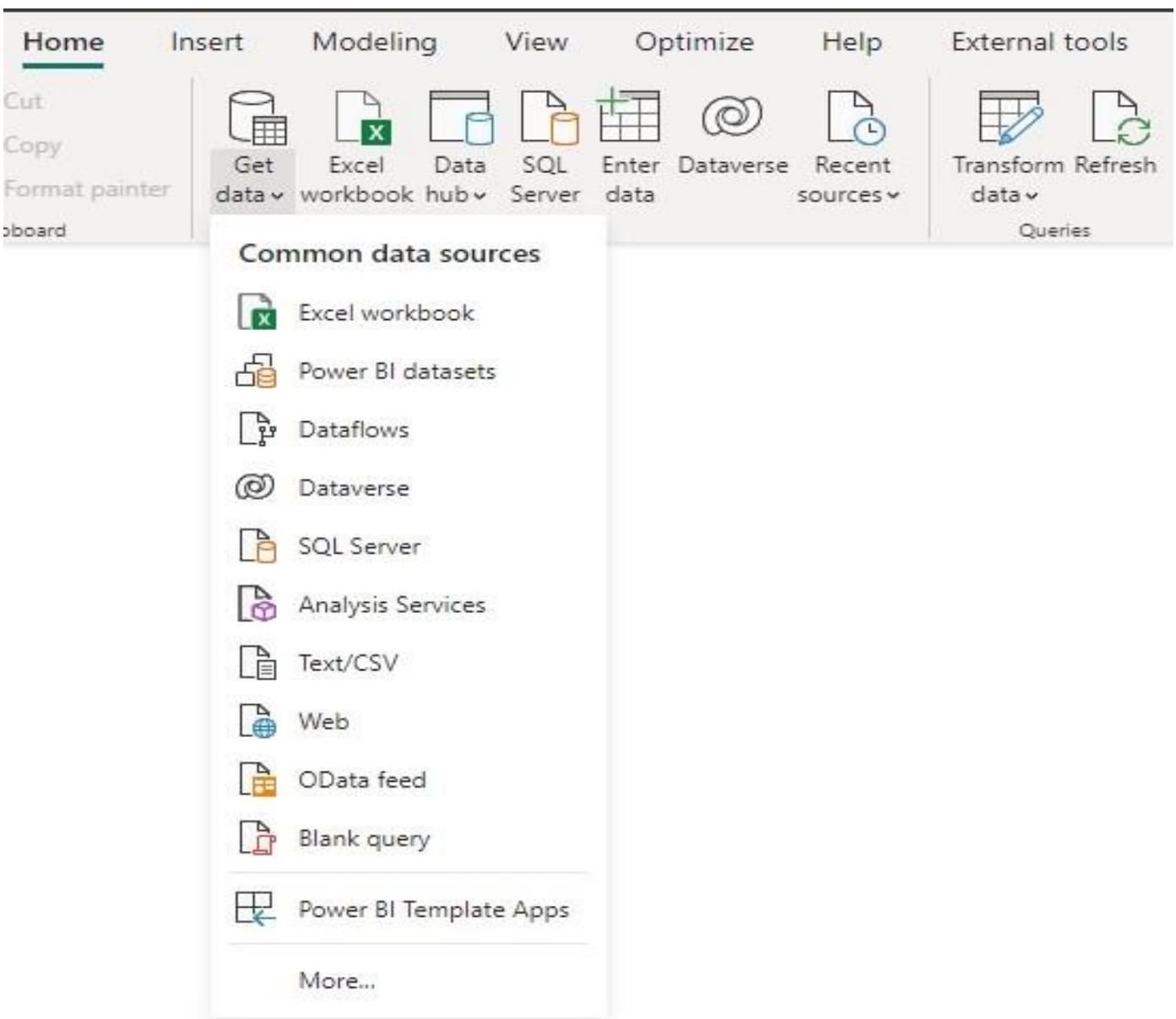


REPORT FORM WITH ALL
THE DATA COMBINED

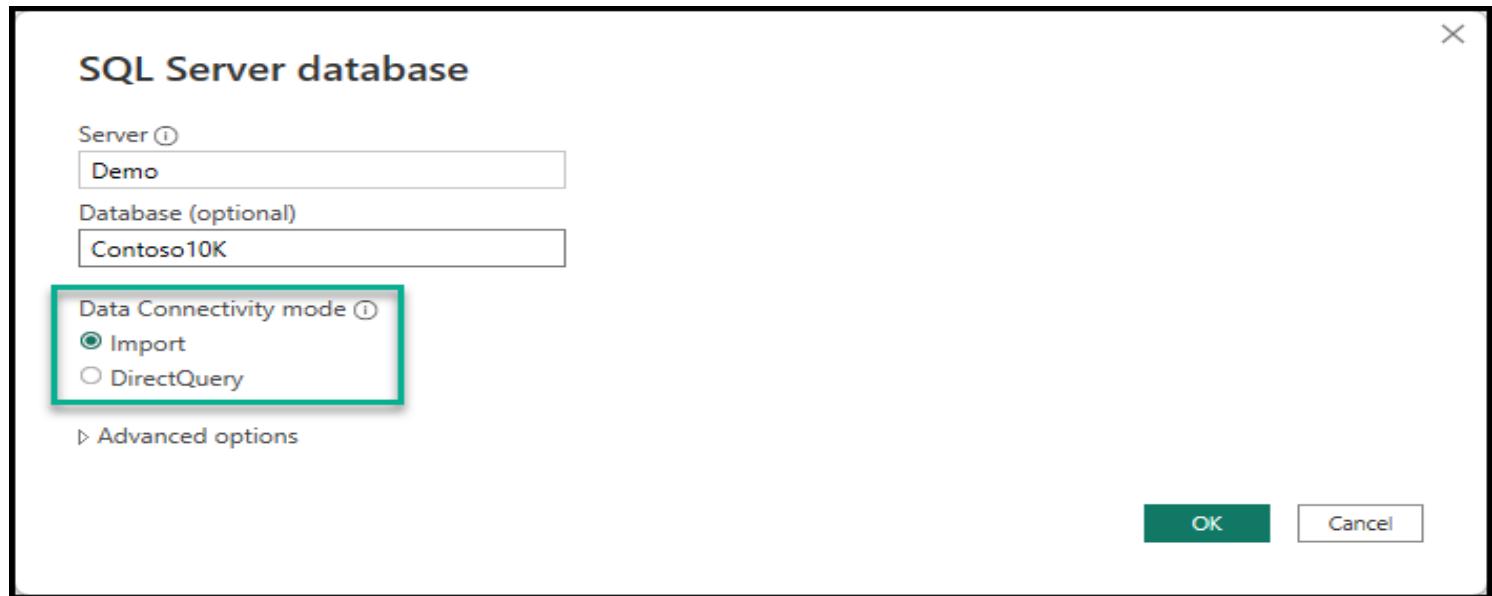
Toy Model Builder

- Start - picture of finished product
- Proceed through instructions step by step
- Unlike toy model – changes can be made – no glue

Get Data



Import or DirectQuery



Select Tables or Views

Navigator

Display Options ▾

- demo: Contoso10K [20]
 - Currency Exchange
 - Customer
 - Date
 - DIM_Customer
 - DIM_Date
 - DIM_Product
 - DIM_Store
 - FACT_Sales
 - Product
 - Sales
 - Store
 - Data.CurrencyExchange
 - Data.Customer
 - Data.Date
 - Data.GeoLocations
 - Data.OrderRows
 - Data.Orders
 - Data.Product
 - Data.Store
 - fn_IsHoliday

Sales

Order Number	Line Number	Order Date	Delivery Date	CustomerKey	StoreKey	ProductKey
36600	0	1/1/2018	1/1/2018	1620763	540	
36700	0	1/2/2018	1/2/2018	1320351	480	
36700	1	1/2/2018	1/2/2018	1320351	480	
36700	2	1/2/2018	1/2/2018	1320351	480	
36700	3	1/2/2018	1/2/2018	1320351	480	
36700	4	1/2/2018	1/2/2018	1320351	480	
36701	0	1/2/2018	1/2/2018	2077250	585	
36800	0	1/3/2018	1/10/2018	1743422	999999	
36800	1	1/3/2018	1/10/2018	1743422	999999	
36800	2	1/3/2018	1/10/2018	1743422	999999	
36800	3	1/3/2018	1/10/2018	1743422	999999	
36801	0	1/3/2018	1/3/2018	1013948	390	
36801	1	1/3/2018	1/3/2018	1013948	390	
36801	2	1/3/2018	1/3/2018	1013948	390	
36801	3	1/3/2018	1/3/2018	1013948	390	
36900	0	1/4/2018	1/4/2018	544050	220	
36901	0	1/4/2018	1/14/2018	485447	999999	
36902	0	1/4/2018	1/17/2018	1075463	999999	
37000	0	1/5/2018	1/5/2018	1319917	550	
37000	1	1/5/2018	1/5/2018	1319917	550	
37000	2	1/5/2018	1/5/2018	1319917	550	
37000	3	1/5/2018	1/5/2018	1319917	550	
37100	0	1/6/2018	1/6/2018	1107123	380	
37100	1	1/6/2018	1/6/2018	1107123	380	

Select Related Tables Load Transform Data Cancel

Load or Transform Data

Navigator

Display Options ▾

- demo: Contoso10K [20]
 - Currency Exchange
 - Customer
 - Date
 - DIM_Customer
 - DIM_Date
 - DIM_Product
 - DIM_Store
 - FACT_Sales
 - Product
 - Sales
 - Store
 - Data.CurrencyExchange
 - Data.Customer
 - Data.Date
 - Data.GeoLocations
 - Data.OrderRows
 - Data.Orders
 - Data.Product
 - Data.Store
 - fn IsHolidav

Sales

Order Number	Line Number	Order Date	Delivery Date	CustomerKey	StoreKey	ProductKey
36600	0	1/1/2018	1/1/2018	1620763	540	
36700	0	1/2/2018	1/2/2018	1320351	480	
36700	1	1/2/2018	1/2/2018	1320351	480	
36700	2	1/2/2018	1/2/2018	1320351	480	
36700	3	1/2/2018	1/2/2018	1320351	480	
36700	4	1/2/2018	1/2/2018	1320351	480	
36701	0	1/2/2018	1/2/2018	2077250	585	
36800	0	1/3/2018	1/10/2018	1743422	999999	
36800	1	1/3/2018	1/10/2018	1743422	999999	
36800	2	1/3/2018	1/10/2018	1743422	999999	
36800	3	1/3/2018	1/10/2018	1743422	999999	
36801	0	1/3/2018	1/3/2018	1013948	390	
36801	1	1/3/2018	1/3/2018	1013948	390	
36801	2	1/3/2018	1/3/2018	1013948	390	
36801	3	1/3/2018	1/3/2018	1013948	390	
36900	0	1/4/2018	1/4/2018	544050	220	
36901	0	1/4/2018	1/14/2018	485447	999999	
36902	0	1/4/2018	1/17/2018	1075463	999999	
37000	0	1/5/2018	1/5/2018	1319917	550	
37000	1	1/5/2018	1/5/2018	1319917	550	
37000	2	1/5/2018	1/5/2018	1319917	550	
37000	3	1/5/2018	1/5/2018	1319917	550	
37100	0	1/6/2018	1/6/2018	1107123	380	
37100	1	1/6/2018	1/6/2018	1107123	380	

Select Related Tables

Load Transform Data Cancel

Power Query Editor

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

New Recent Enter Data source settings Manage Parameters Refresh Advanced Editor Properties Close & Apply Data Sources Parameters Query Choose Columns Remove Columns Keep Rows Remove Rows Sort Split Column Group By Data Type: Whole Number Use First Row as Headers Replace Values Close New Query Manage Columns Reduce Rows Transform Text Analytics Vision Azure Machine Learning AI Insights

Queries [4]

Store Product Sales Customer

= Source{[Schema="dbo",Item="Sales"]}[Data]

	Order Number	Line Number	Order Date	Delivery Date	CustomerKey
1	36600	0	1/1/2018	1/1/2018	
2	36700	0	1/2/2018	1/2/2018	
3	36700	1	1/2/2018	1/2/2018	
4	36700	2	1/2/2018	1/2/2018	
5	36700	3	1/2/2018	1/2/2018	
6	36700	4	1/2/2018	1/2/2018	
7	36701	0	1/2/2018	1/2/2018	
8	36800	0	1/3/2018	1/10/2018	
9	36800	1	1/3/2018	1/10/2018	
10	36800	2	1/3/2018	1/10/2018	
11	36800	3	1/3/2018	1/10/2018	
12	36801	0	1/3/2018	1/3/2018	
13	36801	1	1/3/2018	1/3/2018	
14	36801	2	1/3/2018	1/3/2018	
15	36801	3	1/3/2018	1/3/2018	
16	36900	0	1/4/2018	1/4/2018	
17	36901	0	1/4/2018	1/14/2018	
18	36902	0	1/4/2018	1/17/2018	
19	37000	0	1/5/2018	1/5/2018	
20	37000	1	1/5/2018	1/5/2018	
21	37000	2	1/5/2018	1/5/2018	
22	37000	3	1/5/2018	1/5/2018	
23	37100	0	1/6/2018	1/6/2018	
24	37100	1	1/6/2018	1/6/2018	
25					

13 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

Query Settings

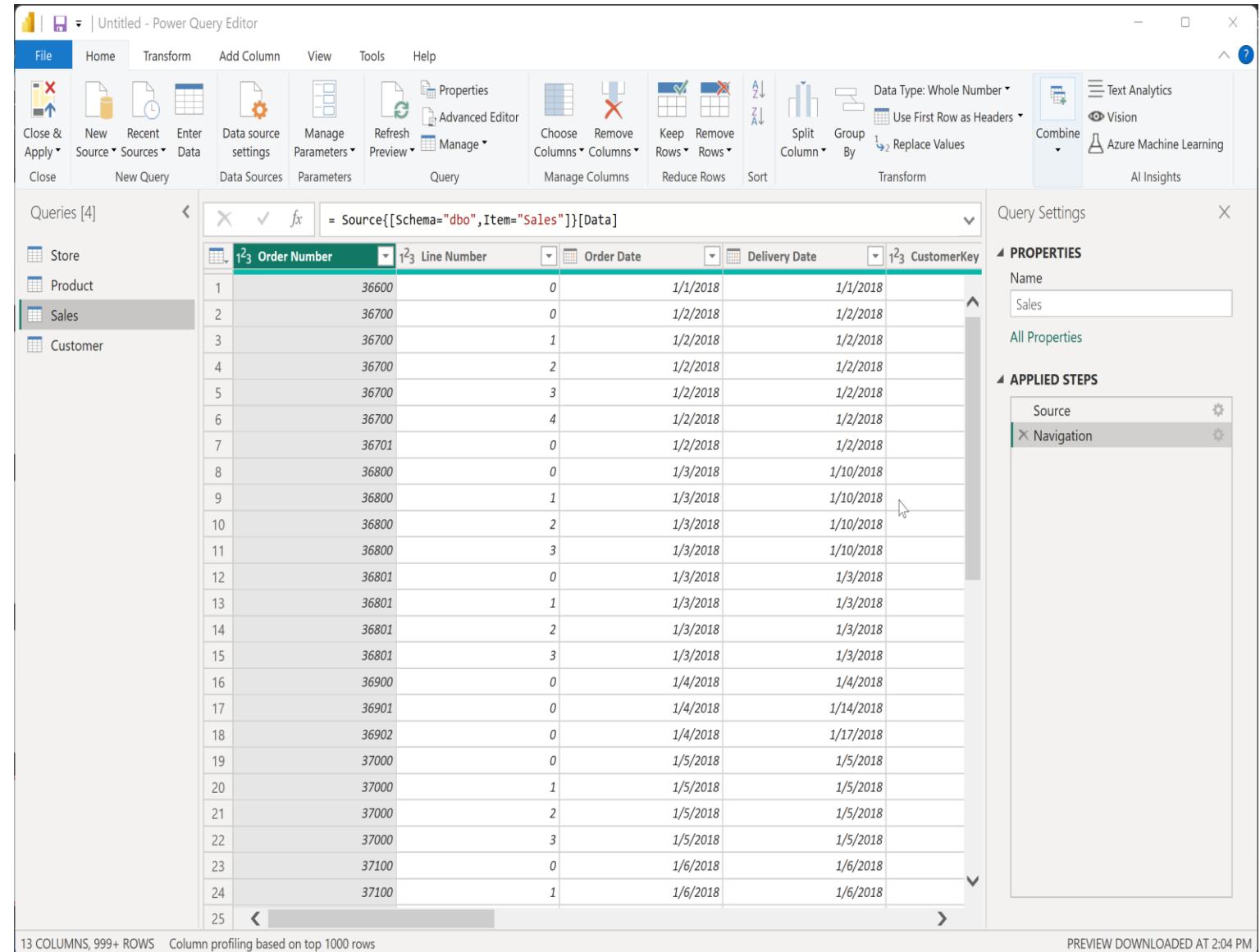
PROPERTIES

Name: Sales

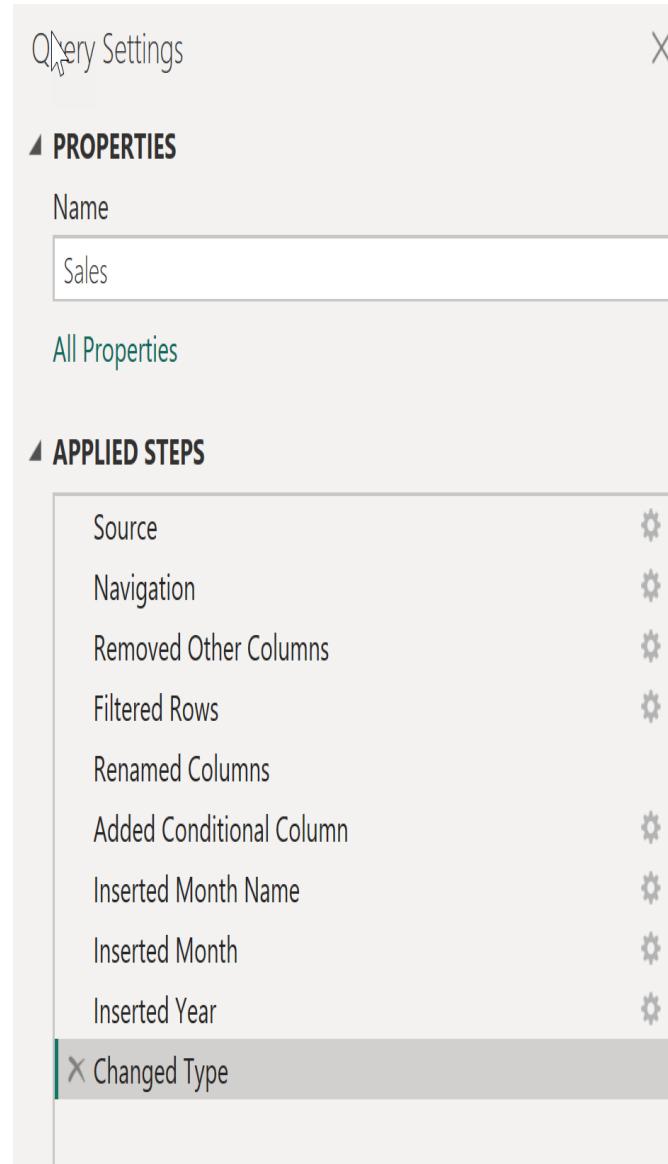
APPLIED STEPS

Source Navigation

PREVIEW DOWNLOADED AT 2:04 PM



Power Query Steps like a Recipe



Promote First Row to Headers (CSV)

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Group By Use First Row as Headers ▾

Queri Use First Row as Headers Use Headers as First Row

Transpose Reverse Rows Count Rows

Data Type: Text ▾ Replace Values ▾ Unpivot Columns

Detect Data Type Fill ▾ Move ▾

Rename Pivot Column Convert to List

Any Column

Promote the first row of this table into column headers.

	A ^B _C Column1	A ^B _C Column2
1	Order Number	Line Number
2	36600	0
3	36700	0
4	36700	1
5	36700	2

Promote First Row to Headers (CSV)

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

New Recent Enter Data source settings Manage Parameters Refresh Preview Manage

Close & Apply Close New Query Data Source... Parameters Query Sort Transform

Manage Columns Reduce Rows Sort

Data Type: Text ▾ Use First Row as Headers ▾ Split Column By Replace Values

Text Analytics Vision Azure Machine Learning AI Insights

Queries [1] Sales

= Csv.Document(File.Contents("C:\ContosoDataFiles\Sales.csv"),[Delimiter=",",

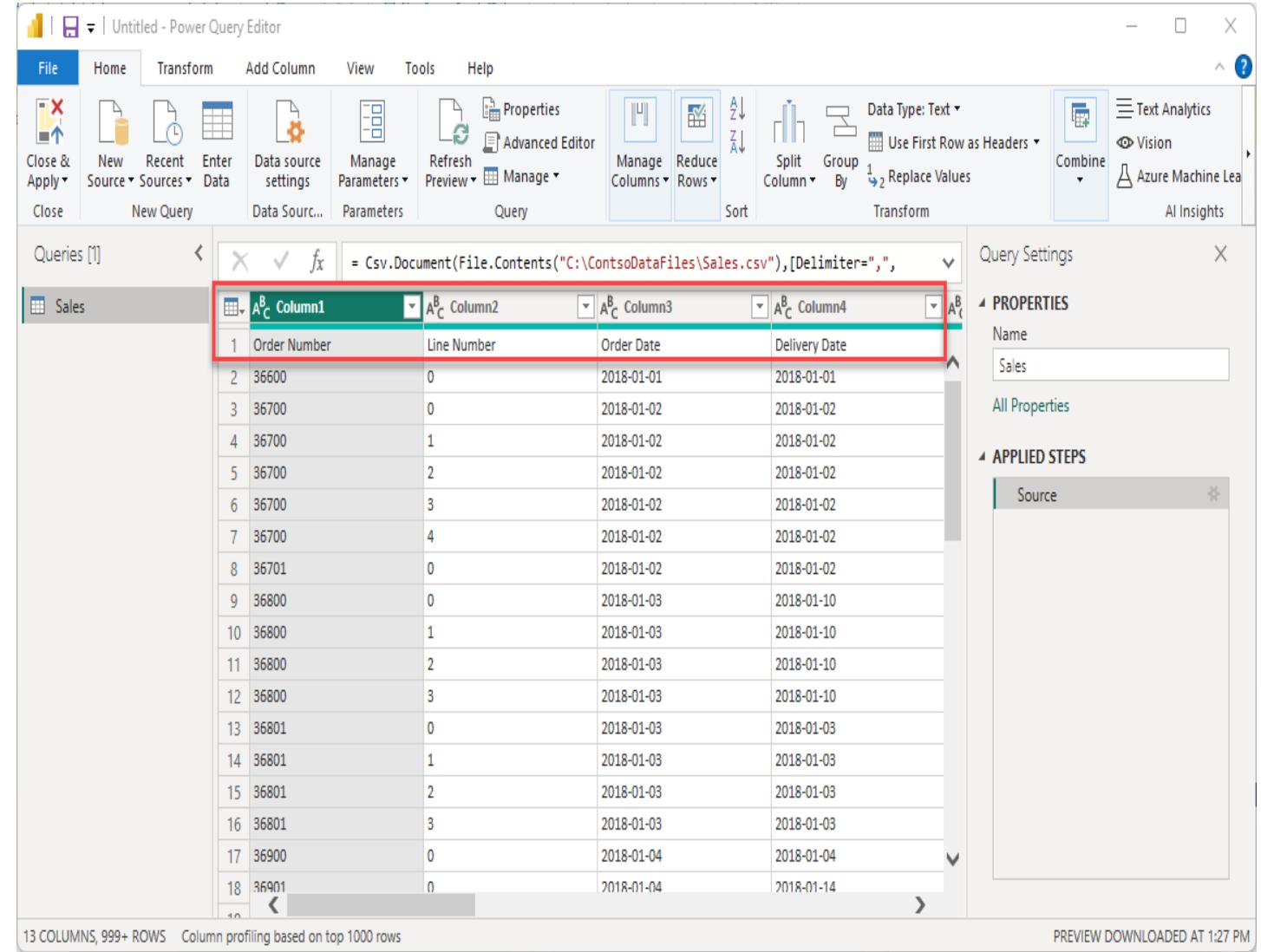
	Column1	Column2	Column3	Column4
1	Order Number	Line Number	Order Date	Delivery Date
2	36600	0	2018-01-01	2018-01-01
3	36700	0	2018-01-02	2018-01-02
4	36700	1	2018-01-02	2018-01-02
5	36700	2	2018-01-02	2018-01-02
6	36700	3	2018-01-02	2018-01-02
7	36700	4	2018-01-02	2018-01-02
8	36701	0	2018-01-02	2018-01-02
9	36800	0	2018-01-03	2018-01-10
10	36800	1	2018-01-03	2018-01-10
11	36800	2	2018-01-03	2018-01-10
12	36800	3	2018-01-03	2018-01-10
13	36801	0	2018-01-03	2018-01-03
14	36801	1	2018-01-03	2018-01-03
15	36801	2	2018-01-03	2018-01-03
16	36801	3	2018-01-03	2018-01-03
17	36900	0	2018-01-04	2018-01-04
18	36901	0	2018-01-04	2018-01-14

13 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 1:27 PM

Query Settings

Properties Name Sales All Properties

Applied Steps Source



Promote First Row to Headers (CSV)

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

New Recent Enter Data source settings Manage Parameters Refresh Preview Manage

Close & Apply Close New Query Data Source... Parameters Query Sort Transform

Manage Columns Reduce Rows Sort

Data Type: Text Use First Row as Headers

Split Group By Replace Values

Text Analytics Vision Azure Machine Lea

Combine AI Insights

Queries [1] Sales

= Table.PromoteHeaders(Source, [PromoteAllScalars=true])

	Order Number	Line Number	Order Date	Delivery Date
1	36600	0	2018-01-01	2018-01-01
2	36700	0	2018-01-02	2018-01-02
3	36700	1	2018-01-02	2018-01-02
4	36700	2	2018-01-02	2018-01-02
5	36700	3	2018-01-02	2018-01-02
6	36700	4	2018-01-02	2018-01-02
7	36701	0	2018-01-02	2018-01-02
8	36800	0	2018-01-03	2018-01-10
9	36800	1	2018-01-03	2018-01-10
10	36800	2	2018-01-03	2018-01-10
11	36800	3	2018-01-03	2018-01-10
12	36801	0	2018-01-03	2018-01-03
13	36801	1	2018-01-03	2018-01-03
14	36801	2	2018-01-03	2018-01-03
15	36801	3	2018-01-03	2018-01-03
16	36900	0	2018-01-04	2018-01-04
17	36901	0	2018-01-04	2018-01-14
18	36902	0	2018-01-04	2018-01-17

13 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 1:28 PM

Query Settings

PROPERTIES

Name: Sales

All Properties

APPLIED STEPS

Source Promoted Headers

Remove Other Columns

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Close & Apply New Source Recent Sources Enter Data Data source settings Manage Parameters Refresh Preview Advanced Editor Properties Choose Columns Remove Columns Keep Rows Remove Rows Split Column Group By Replace Values Data Type: Any Use First Row as Headers

Queries [1] Sales

= Contoso10K{[Schema="dbo",Item="Sales"]}[Data]

Order Number	Line Number	Order Date	Delivery Date	CustomerKey
36600	0	1/1/2018	1/1/2018	1620763
36700	0	1/2/2018	1/2/2018	1320351
36700	1	1/2/2018	1/2/2018	1320351
36700	2	1/2/2018	1/2/2018	1320351
36700	3	1/2/2018	1/2/2018	1320351
36700	4	1/2/2018	1/2/2018	1320351
36701	0	1/2/2018	1/2/2018	2077250
36800	0	1/3/2018	1/10/2018	1743422
36800	1	1/3/2018	1/10/2018	1743422
36800	2	1/3/2018	1/10/2018	1743422
36800	3	1/3/2018	1/10/2018	1743422
36801	0	1/3/2018	1/3/2018	1013948
36801	1	1/3/2018	1/3/2018	1013948
36801	2	1/3/2018	1/3/2018	1013948
36801	3	1/3/2018	1/3/2018	1013948
36900	0	1/4/2018	1/4/2018	544050
36901	0	1/4/2018	1/14/2018	485447
36902	0	1/4/2018	1/17/2018	1075463
37000	0	1/5/2018	1/5/2018	1319917
37000	1	1/5/2018	1/5/2018	1319917
37000	2	1/5/2018	1/5/2018	1319917
37000	3	1/5/2018	1/5/2018	1319917
37100	0	1/6/2018	1/6/2018	1107123
37100	1	1/6/2018	1/6/2018	1107123
37100	2	1/6/2018	1/6/2018	1107123
37100	3	1/6/2018	1/6/2018	1107123
37100	4	1/6/2018	1/6/2018	1107123
37100	5	1/6/2018	1/6/2018	1107123
37100	6	1/6/2018	1/6/2018	1107123
37100	7	1/6/2018	1/6/2018	1107123
37100	8	1/6/2018	1/6/2018	1107123
37100	9	1/6/2018	1/6/2018	1107123
37100	10	1/6/2018	1/6/2018	1107123
37100	11	1/6/2018	1/6/2018	1107123
37100	12	1/6/2018	1/6/2018	1107123
37100	13	1/6/2018	1/6/2018	1107123
37100	14	1/6/2018	1/6/2018	1107123
37100	15	1/6/2018	1/6/2018	1107123
37100	16	1/6/2018	1/6/2018	1107123
37100	17	1/6/2018	1/6/2018	1107123
37100	18	1/6/2018	1/6/2018	1107123
37100	19	1/6/2018	1/6/2018	1107123
37100	20	1/6/2018	1/6/2018	1107123
37100	21	1/6/2018	1/6/2018	1107123
37100	22	1/6/2018	1/6/2018	1107123
37100	23	1/6/2018	1/6/2018	1107123
37100	24	1/6/2018	1/6/2018	1107123
37100	25	1/6/2018	1/6/2018	1107123

13 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED ON TUESDAY

Text Analytics Vision Azure Machine Learning AI Insights

Query Settings

Properties Name Sales All Properties

Applied Steps Source Navigation

Filter Rows

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

New Recent Enter Data source settings Manage Parameters Refresh Advanced Editor Properties Preview Manage Choose Columns Remove Keep Rows Remove Sort Split Column Group By Replace Values Data Type: Date Use First Row as Headers Close & Apply Close New Query Data Sources Parameters Query Manage Columns Reduce Rows Transform

Queries [1] Sales = Contoso10K[[Schema="dbo",Item="Sales"]][Data]

	Order Number	Line Number	Order Date	Delivery Date	CustomerKey
1	36600	0	1/1/2018	1/1/2018	1620763
2	36700	0	1/2/2018	1/2/2018	1320351
3	36700	1	1/2/2018	1/2/2018	1320351
4	36700	2	1/2/2018	1/2/2018	1320351
5					1320351
6					1320351
7					2077250
8					1743422
9					1743422
10					1743422
11					1743422
12					1013948
13					1013948
14					1013948
15					1013948
16					544050
17					485447
18					1075463
19					1319917
20					1319917
21					1319917
22					1319917
23					1107123
24					1107123
25					

Filter Rows

Apply one or more filter conditions to the rows in this table.

Basic Advanced

Keep rows where 'Order Date'

And Or

OK Cancel

13 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED ON TUESDAY

Text Analytics Vision Azure Machine Learning All Insights

Combine

Properties Name Sales All Properties

Applied Steps Source Navigation

Set Data Types

- Correct types matter
 - Same type for relationships
- Performance
- Calculations

ProductKey as a String

Name	Cardinality	Total Size ↓	Data	Dictionary	Hier Size	Encoding	Data Type	RI Violations	User Hier Size	Rel Size	% Table
Product	2,517	15,142,472	27,520	15,043,768	71,184	Many	-	-	0	0	
Product Name	2,517	1,128,480	4,160	1,104,176	20,144	HASH	String	-	-	-	7.45%
Product Code	2,517	1,127,544	4,160	1,103,240	20,144	HASH	String	-	-	-	7.45%
ProductKey	2,517	1,127,416	4,160	1,103,112	20,144	HASH	String	-	-	-	7.45%
Unit Cost	480	1,076,408	3,008	1,069,544	3,856	HASH	String	-	-	-	7.11%
Unit Price	426	1,075,520	3,008	1,069,088	3,424	HASH	String	-	-	-	7.10%
Weight	297	1,073,176	2,736	1,068,056	2,384	HASH	String	-	-	-	7.09%
Subcategory	32	1,067,560	1,352	1,065,936	272	HASH	String	-	-	-	7.05%
Subcategory Code	32	1,067,560	1,352	1,065,936	272	HASH	String	-	-	-	7.05%
Color	16	1,067,144	1,192	1,065,808	144	HASH	String	-	-	-	7.05%
Brand	11	1,066,464	600	1,065,768	96	HASH	String	-	-	-	7.04%
Manufacturer	11	1,066,464	600	1,065,768	96	HASH	String	-	-	-	7.04%
Category	8	1,066,288	464	1,065,744	80	HASH	String	-	-	-	7.04%
Category Code	8	1,066,288	464	1,065,744	80	HASH	String	-	-	-	7.04%
Weight Unit Measure	4	1,065,896	136	1,065,712	48	HASH	String	-	-	-	7.04%

ProductKey as a Whole Number

Name	Cardinality	Total Size ↓	Data	Dictionary	Hier Size	Encoding	Data Type	RI Violations	User Hier Size	Rel Size	% Table
Product	2,517	14,115,268	27,520	14,016,564	71,184	Many	-	-	0	0	
Product Name	2,517	1,128,480	4,160	1,104,176	20,144	HASH	String	-	-	-	7.99%
Product Code	2,517	1,127,544	4,160	1,103,240	20,144	HASH	String	-	-	-	7.99%
Unit Cost	480	1,076,408	3,008	1,069,544	3,856	HASH	String	-	-	-	7.63%
Unit Price	426	1,075,520	3,008	1,069,088	3,424	HASH	String	-	-	-	7.62%
Weight	297	1,073,176	2,736	1,068,056	2,384	HASH	String	-	-	-	7.60%
Subcategory	32	1,067,560	1,352	1,065,936	272	HASH	String	-	-	-	7.56%
Subcategory Code	32	1,067,560	1,352	1,065,936	272	HASH	String	-	-	-	7.56%
Color	16	1,067,144	1,192	1,065,808	144	HASH	String	-	-	-	7.56%
Brand	11	1,066,464	600	1,065,768	96	HASH	String	-	-	-	7.56%
Manufacturer	11	1,066,464	600	1,065,768	96	HASH	String	-	-	-	7.56%
Category	8	1,066,288	464	1,065,744	80	HASH	String	-	-	-	7.55%
Category Code	8	1,066,288	464	1,065,744	80	HASH	String	-	-	-	7.55%
Weight Unit Measure	4	1,065,896	136	1,065,712	48	HASH	String	-	-	-	7.55%
ProductKey	2,517	100,212	4,160	75,908	20,144	HASH	Int64	-	-	-	0.71%

Cardinality

Name	Cardinality	Total Size ↓	Data	Dictionary	Hier Size	Encoding	Data Type	RI Violations	User Hier Size	Rel Size	% Table
Product	2,517	15,142,472	27,520	15,043,768	71,184	Many	-	-	0	0	
Product Name	2,517	1,128,480	4,160	1,104,176	20,144	HASH	String	-	-	-	7.45%
Product Code	2,517	1,127,544	4,160	1,103,240	20,144	HASH	String	-	-	-	7.45%
ProductKey	2,517	1,127,416	4,160	1,103,112	20,144	HASH	String	-	-	-	7.45%
Unit Cost	480	1,076,408	3,008	1,069,544	3,856	HASH	String	-	-	-	7.11%
Unit Price	426	1,075,520	3,008	1,069,088	3,424	HASH	String	-	-	-	7.10%
Weight	297	1,073,176	2,736	1,068,056	2,384	HASH	String	-	-	-	7.09%
Subcategory	32	1,067,560	1,352	1,065,936	272	HASH	String	-	-	-	7.05%
Subcategory Code	32	1,067,560	1,352	1,065,936	272	HASH	String	-	-	-	7.05%
Color	16	1,067,144	1,192	1,065,808	144	HASH	String	-	-	-	7.05%
Brand	11	1,066,464	600	1,065,768	96	HASH	String	-	-	-	7.04%
Manufacturer	11	1,066,464	600	1,065,768	96	HASH	String	-	-	-	7.04%
Category	8	1,066,288	464	1,065,744	80	HASH	String	-	-	-	7.04%
Category Code	8	1,066,288	464	1,065,744	80	HASH	String	-	-	-	7.04%
Weight Unit Measure	4	1,065,896	136	1,065,712	48	HASH	String	-	-	-	7.04%

Cardinality – Unique Values

- The less unique values the better
- Do you really need Date and Time
- Does the time need to include seconds
 - What about morning, afternoon, night

Custom Columns

Add Conditional Column

Add a conditional column that is computed from the other columns or values.

New column name
Store Size

	Column Name	Operator	Value ⓘ	Output ⓘ	
If	Square Meters	equals	ABC 123 null	Then ABC 123 N/A	...
Else If	Square Meters	is less than or equ...	ABC 123 500	Then ABC 123 Small	...
Else If	Square Meters	is less than or equ...	ABC 123 1000	Then ABC 123 Medium	

Add Clause

Else ⓘ ABC 123 Large

OK Cancel

Add Column from Example

Add Column From Examples
Enter sample values to create a new column (Ctrl+Enter to apply).

	A ^b _c Zip Code	A ^b _c Country Code	A ^b _c Country	A ^b _c Continent	A ^b _c Birthday	Column1
1	13901	US	United States	North America	5/14/1982 0:00	
2	99362	US	United States	North America	6/25/1947 0:00	
3	47274	US	United States	North America	8/29/1942 0:00	
4	48607	US	United States	North America	8/16/1946 0:00	
5	36330	US	United States	North America	12/19/1992 0:00	
6	76102	US	United States	North America	3/3/1946 0:00	
7	91762	US	United States	North America	10/3/2001 0:00	
8	20036	US	United States	North America	3/15/1976 0:00	
9	07501	US	United States	North America	4/9/1979 0:00	
10	52501	US	United States	North America	9/3/1947 0:00	
11	78040	US	United States	North America	10/3/1995 0:00	
12	71201	US	United States	North America	2/13/1997 0:00	
13	53916	US	United States	North America	4/28/1964 0:00	
14	85283	US	United States	North America	1/2/1953 0:00	
15	63011	US	United States	North America	1/13/1956 0:00	
16	97205	US	United States	North America	12/21/1971 0:00	
17	39211	US	United States	North America	11/24/1998 0:00	
18	85364	US	United States	North America	10/12/1946 0:00	
19	03276	US	United States	North America	5/29/1938 0:00	
20	95204	US	United States	North America	2/13/1982 0:00	
21	85724	US	United States	North America	11/19/1943 0:00	
22	97205	US	United States	North America	10/19/1975 0:00	

Column1

- 5/14/1982 0:00 (Birthday)
- 14 (Day from Birthday)
- 5 (Day of Week from Birthday)
- Friday (Day of Week Name from Birthday)
- 134 (Day of Year from Birthday)
- 31 (Days in Month from Birthday)
- 1982-05-14T23:59:59.9999999 (End of Day from Birthday)
- 1982-05-31T23:59:59.9999999 (End of Month from Birthday)
- 1982-06-30T23:59:59.9999999 (End of Quarter from Birthday)
- 1982-05-15T23:59:59.9999999 (End of Week from Birthday)
- 1982-12-31T23:59:59.9999999 (End of Year from Birthday)
- 5 (Month from Birthday)
- May (Month Name from Birthday)
- 2 (Quarter of Year from Birthday)
- 5/1/1982 12:00:00 AM (Start of Month from Birthday)
- 4/1/1982 12:00:00 AM (Start of Quarter from Birthday)
- 5/9/1982 12:00:00 AM (Start of Week from Birthday)
- 1/1/1982 12:00:00 AM (Start of Year from Birthday)
- 3 (Week of Month from Birthday)
- 20 (Week of Year from Birthday)

Query Settings

PROPERTIES

Name
Customer

All Properties

OK Cancel

Add Column from Example

Add Column From Examples
Enter sample values to create a new column (Ctrl+Enter to apply).
Transform: Date.Year(DateTime.From([Birthday]))

	A ^B _C Zip Code	A ^B _C Country Code	A ^B _C Country	A ^B _C Continent	A ^B _C Birthday	Year
1	13901	US	United States	North America	5/14/1982 0:00	1982
2	99362	US	United States	North America	6/25/1947 0:00	1947
3	47274	US	United States	North America	8/29/1942 0:00	1942
4	48607	US	United States	North America	8/16/1946 0:00	1946
5	36330	US	United States	North America	12/19/1992 0:00	1992
6	76102	US	United States	North America	3/3/1946 0:00	1946
7	91762	US	United States	North America	10/3/2001 0:00	2001
8	20036	US	United States	North America	3/15/1976 0:00	1976
9	07501	US	United States	North America	4/9/1979 0:00	1979
10	52501	US	United States	North America	9/3/1947 0:00	1947
11	78040	US	United States	North America	10/3/1995 0:00	1995
12	71201	US	United States	North America	2/13/1997 0:00	1997
13	53916	US	United States	North America	4/28/1964 0:00	1964
14	85283	US	United States	North America	1/2/1953 0:00	1953
15	63011	US	United States	North America	1/13/1956 0:00	1956
16	97205	US	United States	North America	12/21/1971 0:00	1971
17	39211	US	United States	North America	11/24/1998 0:00	1998
18	85364	US	United States	North America	10/12/1946 0:00	1946
19	03276	US	United States	North America	5/29/1938 0:00	1938
20	95204	US	United States	North America	2/13/1982 0:00	1982
21	85724	US	United States	North America	11/19/1943 0:00	1943
22	97205	US	United States	North America	10/19/1975 0:00	1975
23	48066	US	United States	North America	8/13/1995 0:00	1995
24	38049	US	United States	North America	11/13/1939 0:00	1939
25	13202	US	United States	North America	11/24/1961 0:00	1961
26	49546	US	United States	North America	8/24/1940 0:00	1940
27	60014	US	United States	North America	6/17/1985 0:00	1985
28	32501	US	United States	North America	4/16/1951 0:00	1951
29	00336	US	United States	North America	8/23/1990 0:00	1990

OK Cancel

Get Data Demo

Query Folding

The screenshot shows the Power Query Editor interface. The main area displays a table titled "Sales" with columns: Order Number, Line Number, Order Date, Delivery Date, and CustomerKey. The data consists of 25 rows of sales records. The "Applied Steps" pane on the right lists a step named "Removed Other Columns". A context menu is open over this step, with the "View Native Query" option highlighted with a red box.

	Order Number	Line Number	Order Date	Delivery Date	CustomerKey
1	53000	0	6/14/2018	6/14/2018	1365902
2	53000	1	6/14/2018	6/14/2018	1365902
3	53200	0	6/16/2018	6/16/2018	1656264
4	53200	1	6/16/2018	6/16/2018	1656264
5	53200	2	6/16/2018	6/16/2018	1656264
6	53200	3	6/16/2018	6/16/2018	1656264
7	53600	0	6/20/2018	7/2/2018	1799984
8	53700	0	6/21/2018	6/21/2018	1373034
9	53700	1	6/21/2018	6/21/2018	1373034
10	53700	2	6/21/2018	6/21/2018	1373034
11	53900	0	6/23/2018	6/23/2018	326607
12	53900	1	6/23/2018	6/23/2018	326607
13	53900	2	6/23/2018	6/23/2018	326607
14	54300	0	6/27/2018	6/27/2018	687715
15	54400	0	6/28/2018	6/28/2018	1734949
16	54400	1	6/28/2018	6/28/2018	1734949
17	54600	0	6/30/2018	6/30/2018	1043780
18	54600	1	6/30/2018	6/30/2018	1043780
19	55000	0	7/4/2018	7/4/2018	98175
20	55000	1	7/4/2018	7/4/2018	98175
21	55000	2	7/4/2018	7/4/2018	98175
22	55000	3	7/4/2018	7/4/2018	98175
23	55000	4	7/4/2018	7/4/2018	98175
24	55000	5	7/4/2018	7/4/2018	98175
25	55000	6	7/4/2018	7/4/2018	98175

5 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 2:06 PM

Native Query

Native Query

```
select [__].[Order Number],  
       [__].[Line Number],  
       [__].[Order Date],  
       [__].[Delivery Date],  
       [__].[CustomerKey]  
  from  
(  
    select [Order Number],  
          [Line Number],  
          [Order Date],  
          [Delivery Date],  
          [CustomerKey]  
     from [dbo].[Sales] as [$Table]  
) as [__]  
 where [__].[Order Date] > convert(date, '2018-06-13')
```

OK

Breaking (Bad) Query Folding

The screenshot shows the Microsoft Power Query Editor interface. A query named "Sales" is selected. The formula bar at the top displays the formula: `= Table.LastN(#"Removed Other Columns", 20)`. The main area shows a table with 20 rows of sales data. The columns are labeled: Order Number, Line Number, Order Date, Delivery Date, and CustomerKey. The data includes various order numbers like 115302, 115400, and 115501, along with their corresponding dates and keys. On the right side, there is a "Query Settings" pane and an "APPLIED STEPS" pane. The "APPLIED STEPS" pane is expanded, showing a list of steps: Source, Navigation, Removed Other Columns, and Keep bottom 20 rows. The "Keep bottom 20 rows" step has a context menu open, with the option "View Native Query" highlighted with a red box. The status bar at the bottom indicates "5 COLUMNS, 20 ROWS" and "Column profiling based on top 1000 rows".

Keep bottom 20 rows breaks query folding
SQL Server does not have a bottom clause

Query Folding

The screenshot shows the Microsoft Power Query Editor interface. On the left, there's a preview pane displaying a table titled "Sales" with 20 rows of data, each containing an "Order Number". To the right of the preview is a "Native Query" dialog box. Inside the dialog, the following SQL query is displayed:

```
select top 20
    [Order Number],
    [Line Number],
    [Order Date],
    [Delivery Date],
    [CustomerKey]
from [dbo].[Sales] as [$Table]
```

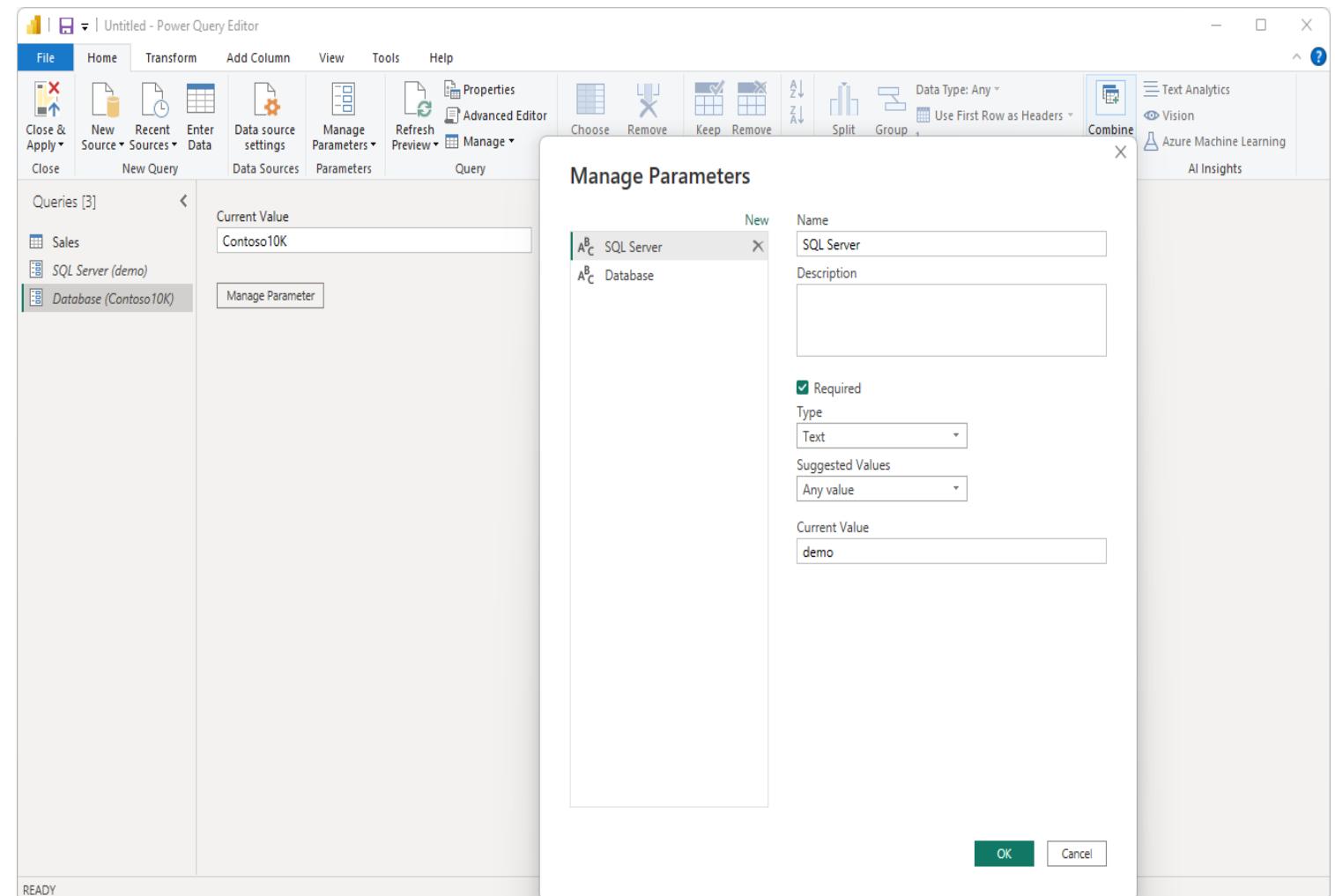
Below the preview pane, a status bar indicates "5 COLUMNS, 20 ROWS" and "Column profiling based on top 1000 rows". On the far right, a "PREVIEW DOWNLOADED AT 2:53 PM" message is visible. The Power Query ribbon at the top includes tabs like File, Home, Transform, Add Column, View, Tools, and Help, along with various toolbar icons for data manipulation.

Keep Top 20 Rows does not
SQL Server has a top N clause

Query Folding

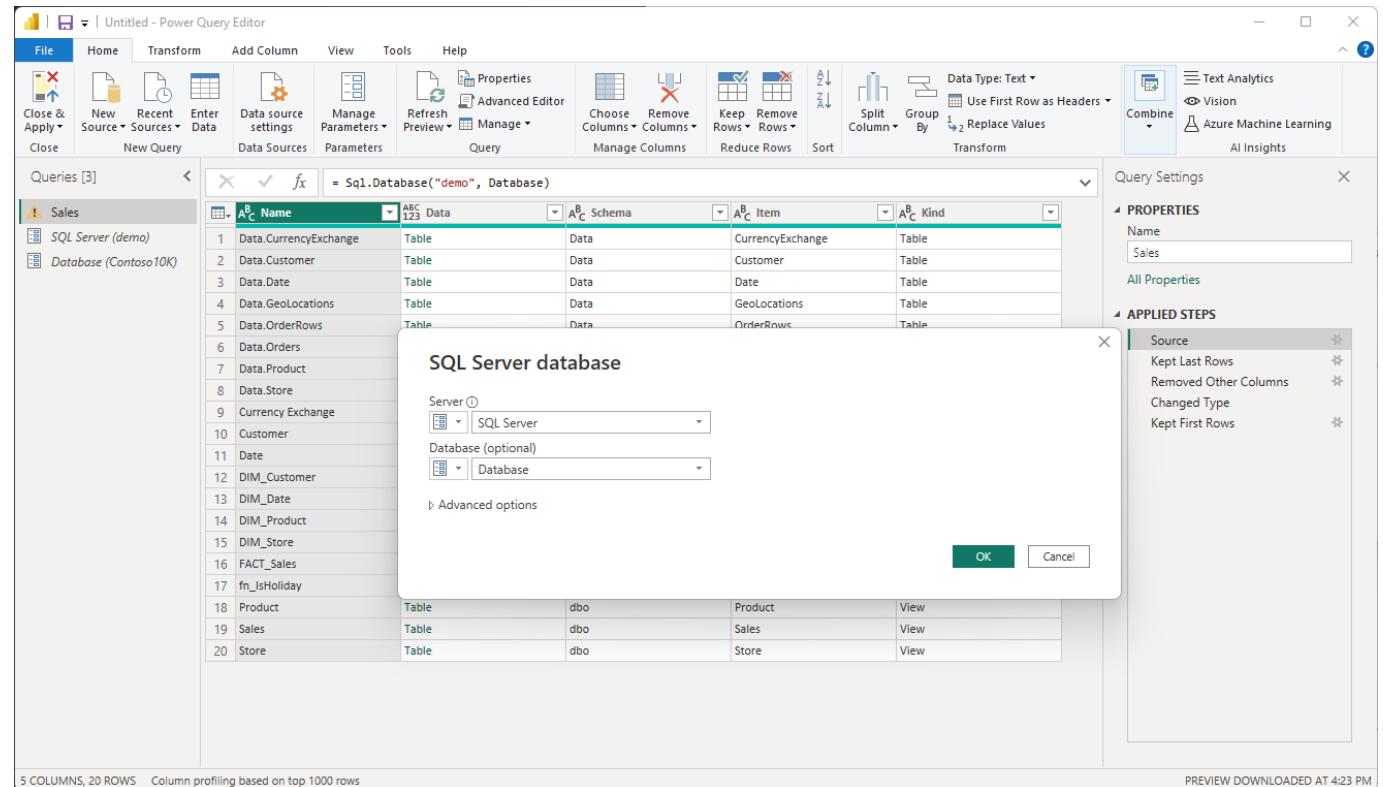
- This is the query sent when the model is refreshed
- This is not the query sent to the tabular engine when you interact with the visuals
- Put off breaking the query folding till the later steps

Parameters



Creating parameters

Using Parameters



Using a parameter for the Server and Database

Goals of Power Query Steps

- Get smallest dataset possible early in query steps
 - Remove Columns
 - Filter Rows
- Unique Values
- Preserve Query Folding as long as possible

Data Model

Turn off Auto Date Tables

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
- Published dataset settings
- Query reduction
- Report settings

Background Data

- Always allow data previews to download in the background
- Allow data previews to download in the background according to each file's setting
- Never allow data previews to download in the background

Parallel loading of tables

When you load data into Power BI (via import or DirectQuery), each data table is backed by a Power Query query. These queries are evaluated simultaneously instead of one-by-one, which can speed up the process. In certain situations, you might want to adjust the default number of simultaneous query evaluations and memory used. [Learn more](#)

Maximum number of simultaneous evaluations:

Maximum memory used per simultaneous evaluation (MB):

Time intelligence

Auto date/time for new files [Learn more](#)

Data Cache Management Options ⓘ

Currently used: 655 KB

Clear Cache

Maximum allowed (MB):

Restore Defaults

Q&A Cache Options ⓘ

Currently used: 1.43 MB

Clear Cache

OK **Cancel**

Hidden Date Tables

- >  Customer
- >  Date
- >  DateTableTemplate_a987aaf9-4324-4fff-9e6b-91eb917a6145
 -  Date
 - >  Date Hierarchy
 - 1²₃ Day
 - A^B_C Month
 - 1²₃ MonthNo
 - A^B_C Quarter
 - 1²₃ QuarterNo
 - 1²₃ Year
- >  LocalDateTable_329979ec-f948-448d-8d40-da32309c689b
- >  LocalDateTable_565d20a7-37da-41eb-acea-3d1681afac62
- >  LocalDateTable_6082b466-aa3a-416d-92f5-dee233105c41
- >  LocalDateTable_729c8b24-1385-4545-897f-e90fb9602a1d
- >  LocalDateTable_86220e95-4654-4888-819a-8af2ddbe1da
- >  LocalDateTable_8ce283be-c847-412c-907b-87781a0d19e1
- >  Product
- >  Sales
- >  Store

Relationships

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
Published dataset settings
Query reduction
Report settings

Type Detection

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 ⓘ [Learn more](#)

Background Data

Allow data previews to download in the background

Parallel loading of tables ⓘ

Maximum number of concurrent jobs [Learn more](#)

Default

One (disable parallel loading)

Custom

Q&A

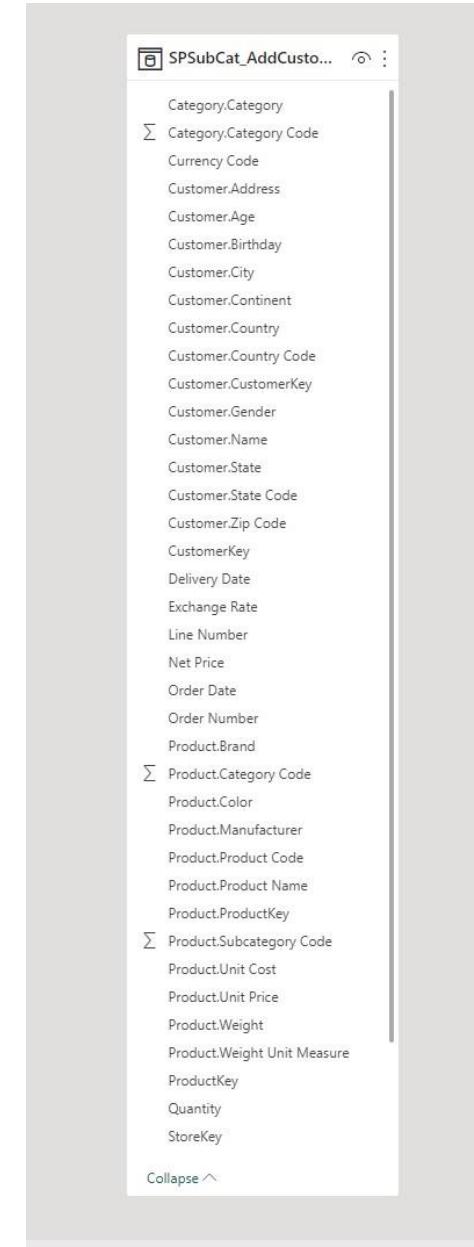
Turn on Q&A to ask natural language questions about your data ⓘ [Learn more](#)

Share your synonyms with everyone in your org

OK

Cancel

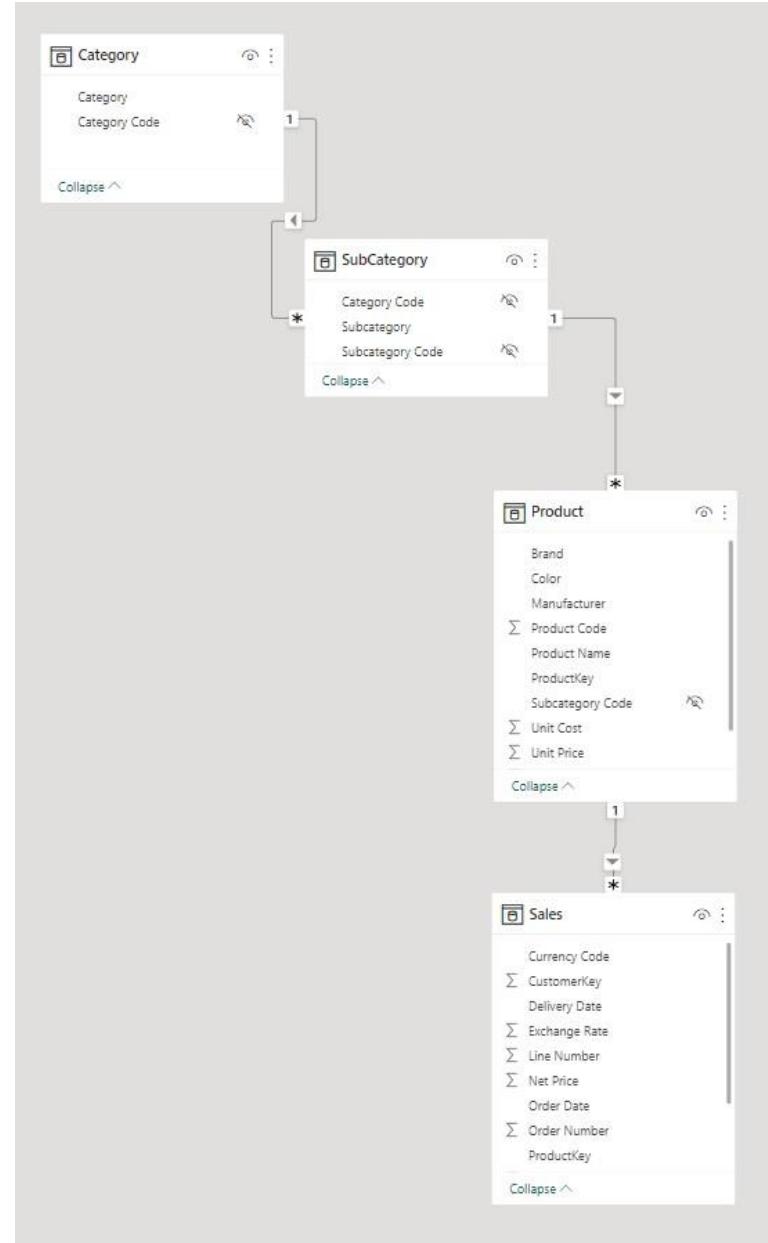
One big wide table (or tall)



Wide Table

- Difficult to find the column you are looking for
- Refresh can take longer
- DAX is more complicated
- Performance is better with Star Schema

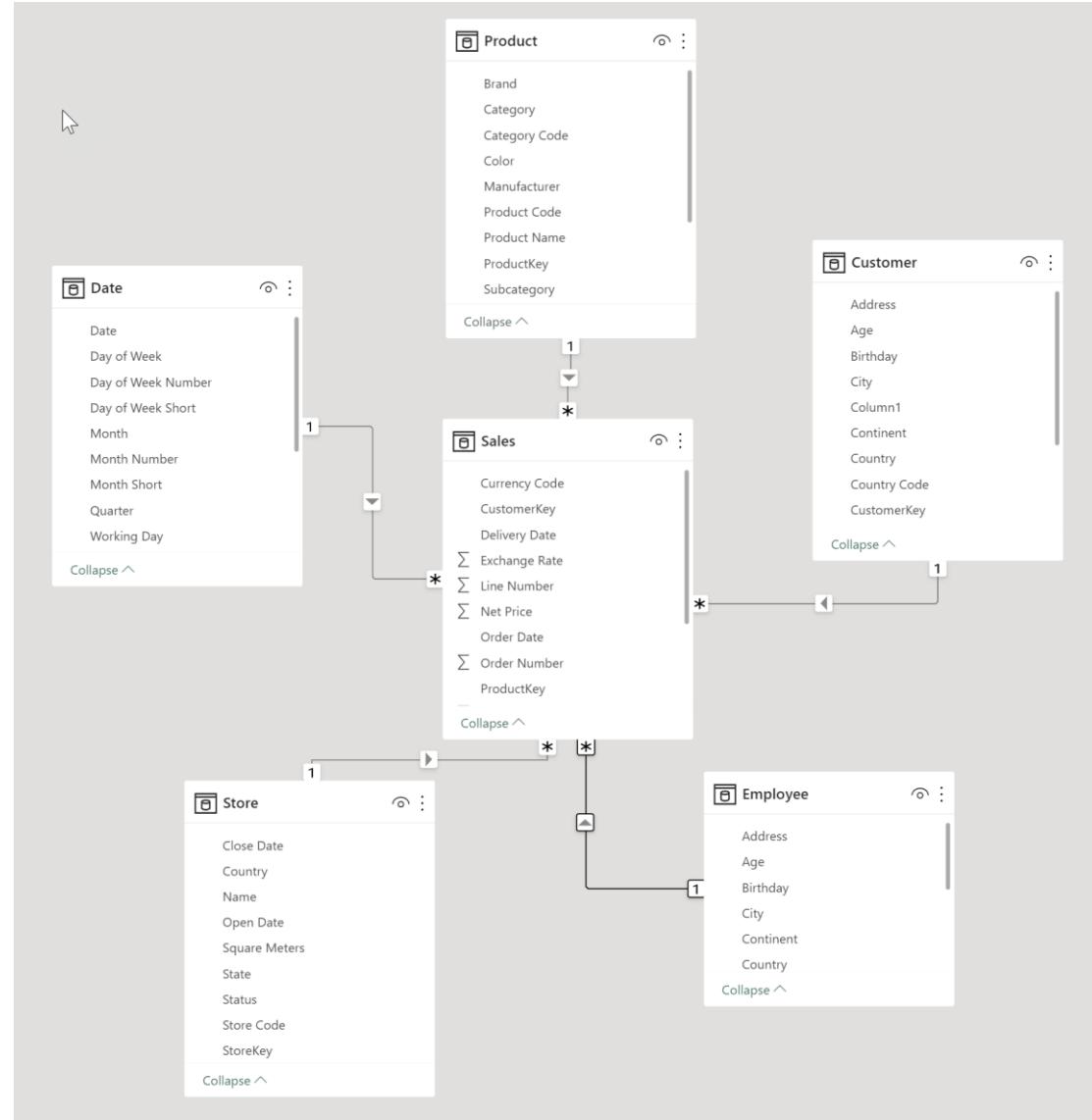
What is Snowflake Schema



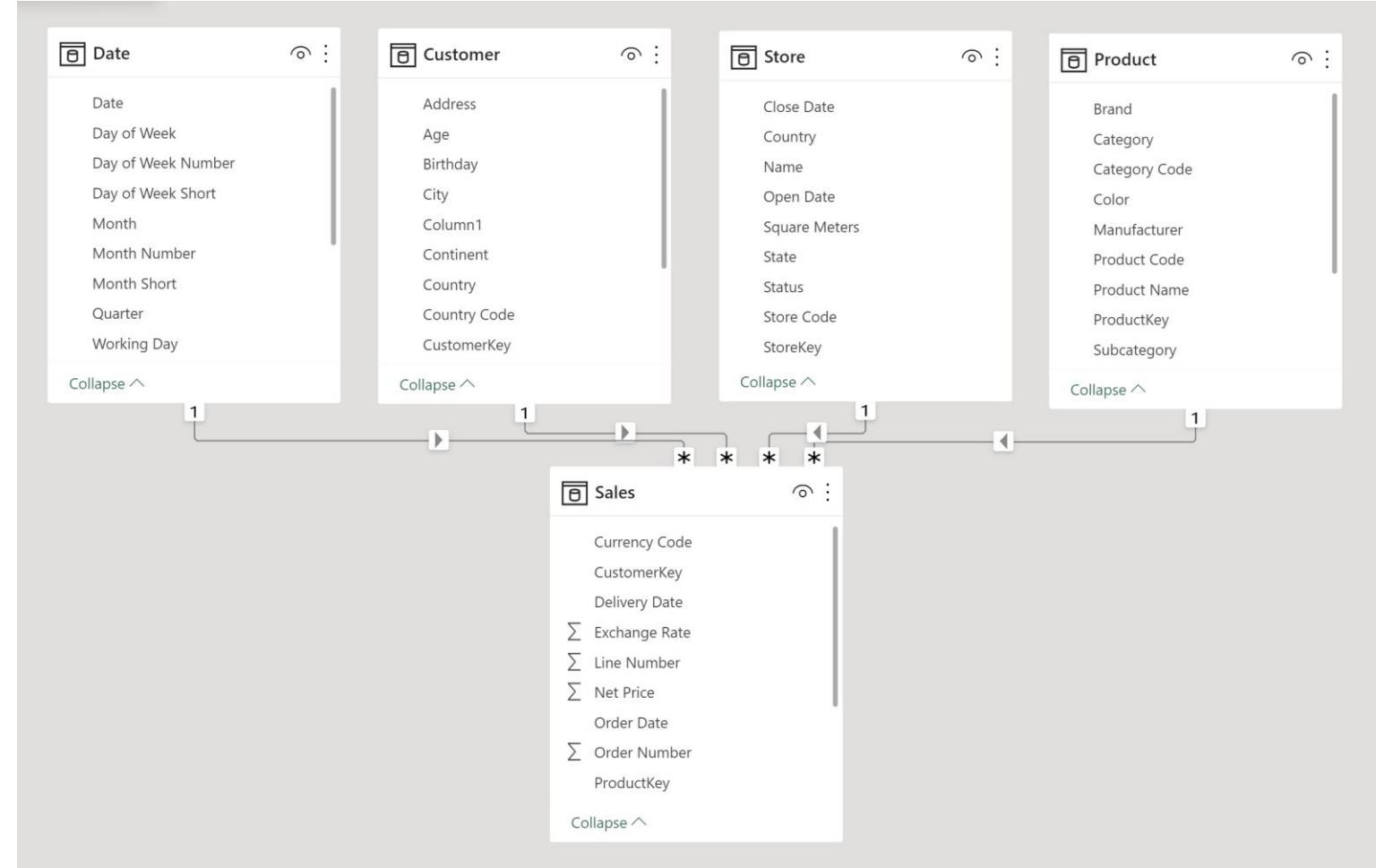
Other examples of Snowflake Schema

- General Ledger
 - Header and Detail
 - Combine into one table
- Employee table with sub type of employee and type of employee.

What is a Star Schema



Collie Method



Snowflake versus Star Schema

Snowflake

- Harder to understand as a developer
- Expanded table issue with DAX

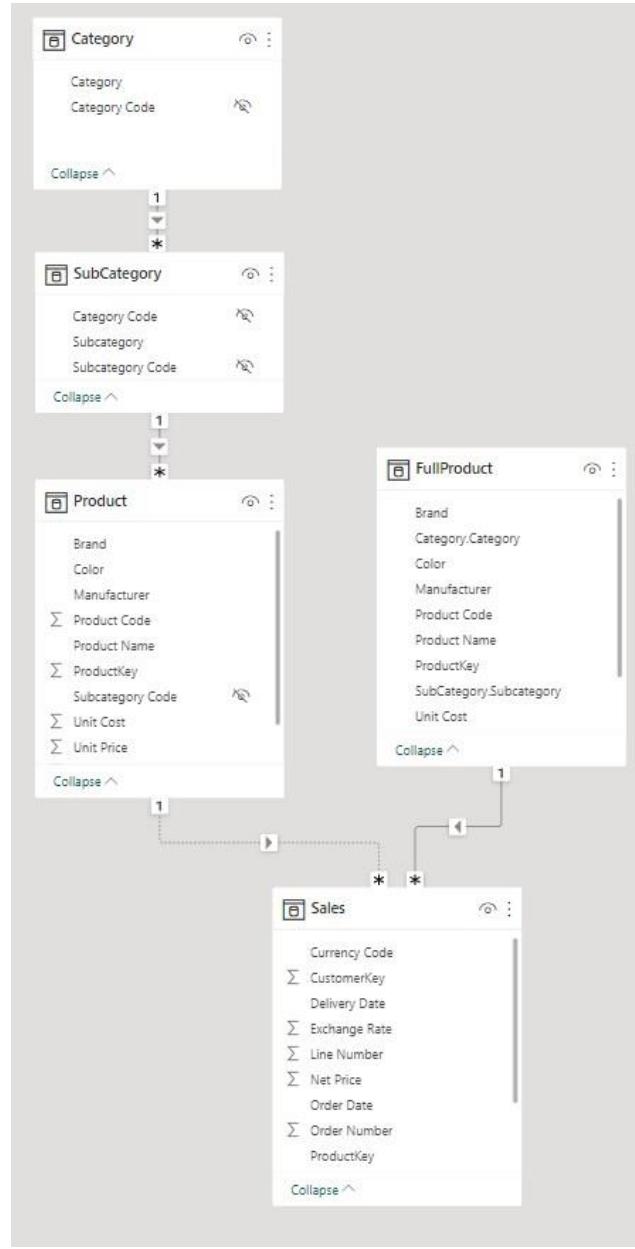
Star Schema

- Simpler
- Overall better performance

Expanded Tables

- When you apply a table filter it really is the expanded table.
`CALCULATE (`
 `COUNTROWS (Subcategory),`
 `Product`
`)`
- It is the expanded table for Product. So only those subcategories that have a product
 - Can't reference base table

Snowflake to (Super) Star Schema



Hiding Columns

- Hide Key Columns
- Build Measures on top of value columns and hide base column
- Implicit Measures versus Explicit Measures

Implicit Measures

- Currently if you create Calculation Groups it disables implicit measures
- Discourage Implicit Measures
 - True/False
- Build explicit Measures

Calculated Columns

- Create all calculated columns in source if possible
- Optimal sorting algorithm only on imported columns

Tabular Editor

- With C# scripts you can automate creating or changing aspects of the model
 - For example, create a measure for every column that has the word amount

Tabular Editor – Best Practice Analyzer

Manage Best Practice Rules

Current model

Rule collections:

- (Effective rules)
 - Rules within the current model
 - Rules for the local user
 - Rules on the local machine

Add... Remove ▲ ▼

Rules in collection:

Rule name	Scope	
[Error Prevention] Data columns must have a source c...	Data Columns	New rule...
[Error Prevention] Expression-reliant objects must have...	Measures,Calculated Columns,C...	Clone rule
[Error Prevention] Relationship columns should be of t...	Relationships	Edit rule...
Formatting		
[Formatting] Add data category for columns	Columns	Delete rule
[Formatting] Do not summarize numeric columns	Columns	
[Formatting] First letter of objects must be capitalized	Tables,Measures,Hierarchies,Cal...	
[Formatting] Format flag columns as Yes/No value strin...	Columns	
[Formatting] Hide fact table columns	Columns	
[Formatting] Hide foreign keys	Columns	
[Formatting] Mark primary keys	Columns	
[Formatting] Month (as a string) must be sorted	Columns	
[Formatting] Objects should not start or end with a space	Model,Tables,Measures,Hierarch...	
[Formatting] Percentages should be formatted with tho...	Measures	
[Formatting] Provide format string for "Date" columns	Columns	
[Formatting] Provide format string for "Month" columns	Columns	
[Formatting] Provide format string for measures	Measures	
[Formatting] Relationship columns should be of integer ...	Columns	
[Formatting] Whole numbers should be formatted with t...	Measures	
Maintenance		
[Maintenance] Calculation groups with no calculation it...	Calculation Groups	Move to...
[Maintenance] Ensure tables have relationships	Tables Calculated Tables	

OK Cancel

Bravo: Date Table

Bravo does
not support
AutoDate



Unsupported

Models with auto date/time option enabled are not supported.

[Disabling auto date-time in Power BI \(video\)](#)

Time Intelligence

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
- Published dataset settings
- Query reduction
- Report settings

Background Data

- Always allow data previews to download in the background
- Allow data previews to download in the background according to each file's setting
- Never allow data previews to download in the background

Parallel loading of tables

When you load data into Power BI (via import or DirectQuery), each data table is backed by a Power Query query. These queries are evaluated simultaneously instead of one-by-one, which can speed up the process. In certain situations, you might want to adjust the default number of simultaneous query evaluations and memory used. [Learn more](#)

Maximum number of simultaneous evaluations

Maximum memory used per simultaneous evaluation (MB)

Time intelligence

Auto date/time for new files [Learn more](#)

Data Cache Management Options ⓘ

Currently used: 655 KB

Clear Cache

Maximum allowed (MB):

Restore Defaults

Q&A Cache Options ⓘ

Currently used: 1.43 MB

Clear Cache

OK **Cancel**

Type of Calendar

Contoso10K > Manage Dates

Model Check



This model is compatible with the Manage Dates feature.

You can create new date tables without worrying about breaking measures or reports.

Calendar Interval Dates Holidays Time Intelligence

Choose a calendar template to apply to this model. Bravo will create the required tables or update them while keeping the existing relationships intact.

Template

Select **Monthly** for calendar based on different number of months. Set **Weekly** for 445-454-544-ISO calendars. Use **Custom** for flexible calendars of variable length. [Manage Templates](#)

Monthly - Fiscal ▾

First Month of the Year

For Weekly ISO, set *January*.

March ▾

Interval

Calendar Interval Dates Holidays Time Intelligence

Select a date interval for your model.

Date Interval



Choose how to determine the date interval of the model. Leave *First Year* and/or *Last Year* empty to use the automatic scan.

First Year

2018

Last Year

2024

Select Measures for Time Intelligence

StarSchema_Date > Manage Dates

Model Check

This model is compatible with the Manage Dates feature.

You can create new date tables without worrying about breaking measures or reports.

Calendar Interval Dates Holidays Time Intelligence

Create the most common Time Intelligence DAX functions available in your model.

Time Intelligence Functions

Target Measures Choose the measure used to generate the Time Intelligence functions.

Choose Measures... ▾

Search

TABLE \ COLUMN

Sales

The screenshot shows the 'Manage Dates' page in Power BI. On the left, a 'Model Check' box indicates compatibility with the feature, noting that new date tables can be created without breaking existing measures or reports. The main area is titled 'Time Intelligence Functions' and includes a 'Target Measures' section where 'Sales' is selected. A red box highlights this selection. At the top, the 'Time Intelligence' tab is selected in the navigation bar. The overall interface is clean and modern, typical of Microsoft's Power BI tool.

Time Intelligence Measures

←  StarSchema_Date > Manage Dates > Preview

Dates Time Intelligence

Search 

↳ To-date growth

- ↳ Sales Amount
 - MOMTD % Sales Amount
 - MOMTD Sales Amount
 - PMTD Sales Amount
 - PQTD Sales Amount
 - PYTD Sales Amount
 - QOQTD % Sales Amount
 - QOQTD Sales Amount
 - YOYTD % Sales Amount
 - YOYTD Sales Amount

A

A

A

A

A

A

A

A

A

A

A

A

A

A

A

A

↳ To-date total

- ↳ Sales Amount
 - MTD Sales Amount
 - QTD Sales Amount
 - YTD Sales Amount

A

A

A

A

Bravo: Optimize Model

Find unused columns, size of columns, and unique elements

Bravo Optimize Model



Your dataset is **1.48 MB** large and contains **66** columns, **56** of which are not referenced within the model.

Search Column

VPAX

COLUMN	TABLE	CARDINALITY	SIZE	WEIG...
Smaller columns...		10.44 K	1.08 MB	73%
Product Name	Product	2.52 K	145.70 KB	10%
ProductKey	Product	2.52 K	97.86 KB	6%
Product Code	Product	2.52 K	84.35 KB	6%
Date	Date	1.10 K	41.05 KB	3%
Birthday	Customer	511	40.34 KB	3%
		19.60 K	1.48 MB	100%

Unreferenced columns can generally be removed from the model to optimize performance. Before removing them, make sure you are not using these columns in any reports, which Bravo cannot determine.

The diagram illustrates the relationships between the columns identified in the table. It shows five boxes: 'Product Name' (yellow), 'ProductKey' (grey), 'Product Code' (yellow), 'Date' (grey), and 'Birthday' (yellow). 'Product Name' and 'ProductKey' are grouped together. 'Product Code' is separate. 'Date' and 'Birthday' are also grouped together. This visualizes how some columns are referenced by multiple tables (like Product Name and ProductKey being part of the Product fact table), while others are unique to specific dimensions (like Product Code, Date, and Birthday).

Visualize Data

Conclusion

- Shape data as much as possible in Power Query
- Import it into a Star Schema
- Use external tools to help optimize model

Resources

- [Guide to Query Folding](#)
- [Improve Model Performance](#)
- [Bravo for Power BI](#)
- [Tabular Editor 2](#)



SQLSATURDAY

OCT, 7th 2023

Orlando, FL

Scan the QR code on
the room poster to
fill out session and
event evaluations



Thank you

Jason Romans

thedaxshepherd@gmail.com

<https://thedaxshepherd.net/> 

<https://www.linkedin.com/in/jason-r-sql-jar>