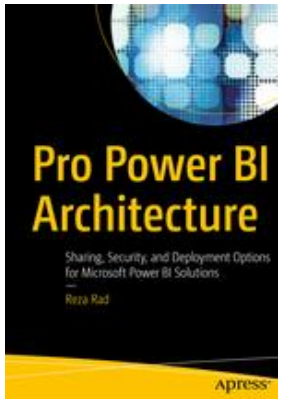
Basics of Power BI Modeling

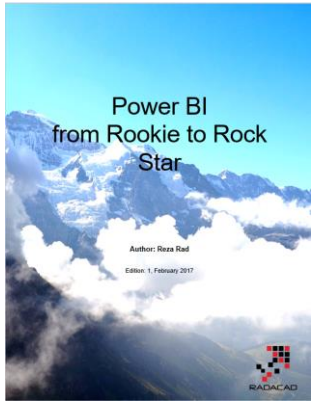
Reza Rad
Microsoft Data Platform MVP, Regional Director

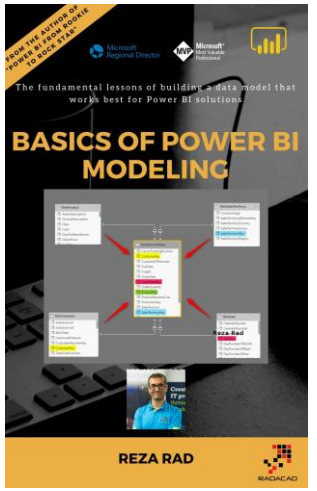
Reza Rad

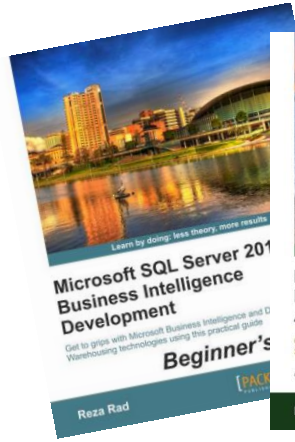
Consultant, Trainer
RADACAD

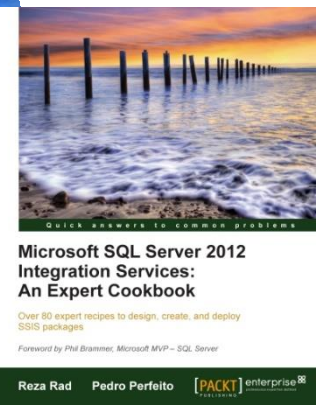
Consultant, Mentor, Trainer, Speaker
Blogger, (and videos on YouTube)
Microsoft Regional Director
Microsoft Data Platform MVP
Author of BI books
Author of Power BI from Rookie to Rock Star
Author of Pro Power BI Architecture
Author of Row-level security in Power BI
Author of Basics of Power BI Modeling











Microsoft Regional Director

Microsoft Most Valuable Professional

https://www.youtube.com/c/radacad

Classified as Microsoft Confidential

Agenda

Module 1: Relationships

Module 2: Dimensional Modeling

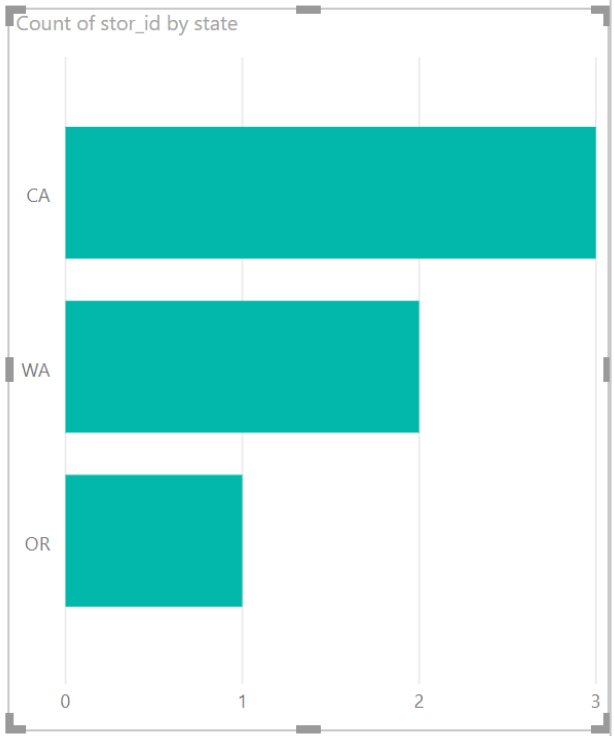
Module 3: Star schema in action

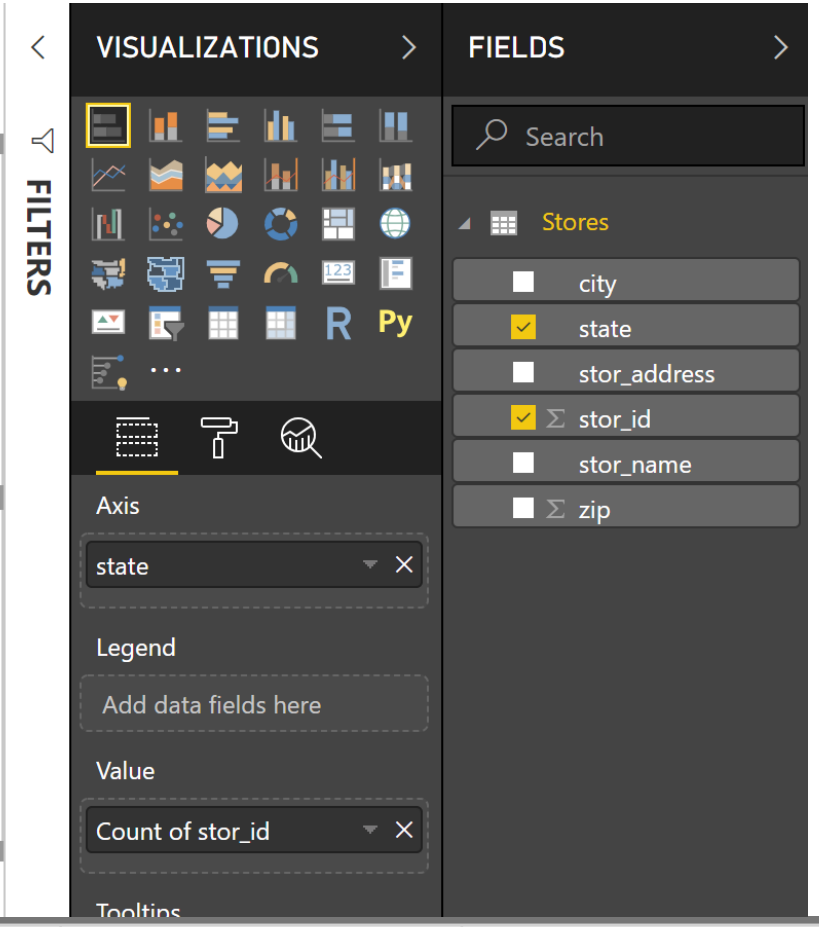
Module 4: calculations

Module 5: Better data model

What is the Relationship?

Filtering by state



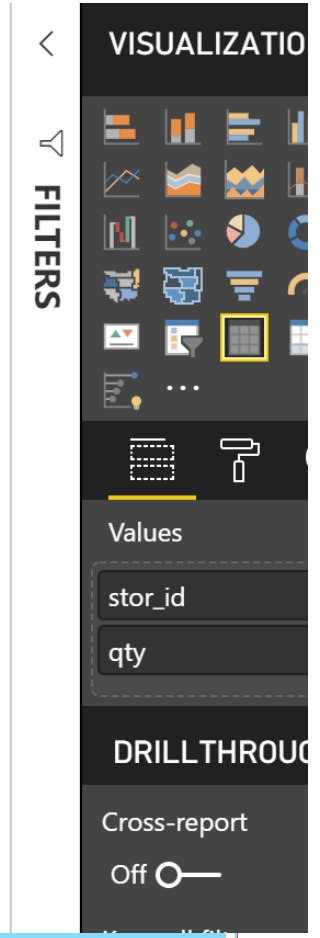


stor_id	stor_name	stor_address	city	state	zip
6380	Eric the Read Books	788 Catamaugus Ave.	Seattle	WA	98056
7066	Barnum's	567 Pasadena Ave.	Tustin	CA	92789
7067	News & Brews	577 First St.	Los Gatos	CA	96745
7131	Doc-U-Mat: Quality Laundry and Books	24-A Avogadro Way	Remulade	WA	98014
7896	Fricative Bookshop	89 Madison St.	Fremont	CA	90019
8042	Bookbeat	679 Carson St.	Portland	OR	89076

Filtering by stor_id

stor_id	ord_num	ord_date	qty		
6380	6871	14/09/1994 12:00:00 AM	5	Net	
6380	722a	13/09/1994 12:00:00 AM	3	Net	
7066	A2976	24/05/1993 12:00:00 AM	50	Net	
7066	QA7442.3	13/09/1994 12:00:00 AM	75	ON	
7067	D4482	14/09/1994 12:00:00 AM	10	Net	
7067	P2121	15/06/1992 12:00:00 AM	40	Net	
7067	P2121	15/06/1992 12:00:00 AM	20	Net	
7067	P2121	15/06/1992 12:00:00 AM	20	Net	
7131	N914008	14/09/1994 12:00:00 AM	20	Net	
7131	N914014	14/09/1994 12:00:00 AM	25	Net	
7131	P3087a	29/05/1993 12:00:00 AM	20	Net	
7131	P3087a	29/05/1993 12:00:00 AM	25	Net	
7131	P3087a	29/05/1993 12:00:00 AM	15	Net 60	PS3333
7131	P3087a	29/05/1993 12:00:00 AM	25	Net 60	PS7777
7896	QQ2299	28/10/1993 12:00:00 AM	15	Net 60	BU7832
7896	TQ456	12/12/1993 12:00:00 AM	10	Net 60	MC2222
7896	X999	21/02/1993 12:00:00 AM	35	ON invoice	BU2075
8042	423LL922	14/09/1994 12:00:00 AM	15	ON invoice	MC3021
8042	423LL930	14/09/1994 12:00:00 AM	10	ON invoice	BU1032

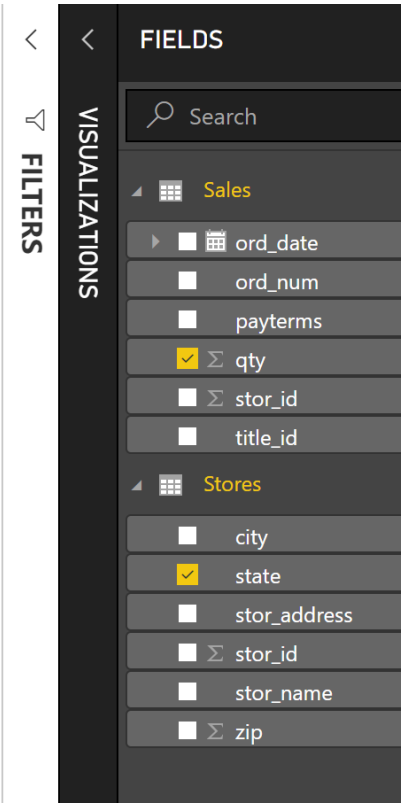
stor_id	qty
7131	130
7066	125
7067	90
8042	80
7896	60
6380	8
Total	**493**



No Filtering

stor_id	ord_num	ord_date	qty	payterms	title_id
6380	6871	14/09/1994 12:00:00 AM	5	Net 60	BU1032
6380	722a	13/09/1994 12:00:00 AM	3	Net 60	PS2091
7066	A2976	24/05/1993 12:00:00 AM	50	Net 30	PC8888
7066	QA7442.3	13/09/1994 12:00:00 AM	75	ON invoice	PS2091
7067	D4482	14/09/1994 12:00:00 AM	10	Net 60	PS2091
7067	P2121	15/06/1992 12:00:00 AM	40	Net 30	TC3218
7067	P2121	15/06/1992 12:00:00 AM	20	Net 30	TC4203
7067	P2121	15/06/1992 12:00:00 AM	20	Net 30	TC7777
7131	N914008	14/09/1994 12:00:00 AM	20	Net 30	PS2091
7131	N914014	14/09/1994 12:00:00 AM	25	Net 30	MC3021
7131	P3087a	29/05/1993 12:00:00 AM	20	Net 60	PS1372
7131	P3087a	29/05/1993 12:00:00 AM	25	Net 60	PS2106
7131	P3087a	29/05/1993 12:00:00 AM	15	Net 60	PS3333
7131	P3087a	29/05/1993 12:00:00 AM	25	Net 60	PS7777
7896	QQ2299	28/10/1993 12:00:00 AM	15	Net 60	BU7832
7896	TQ456	12/12/1993 12:00:00 AM	10	Net 60	MC2222
7896	X999	21/02/1993 12:00:00 AM	35	ON invoice	BU2075
8042	423LL922	14/09/1994 12:00:00 AM	15	ON invoice	MC3021
8042	423LL930	14/09/1994 12:00:00 AM	10	ON invoice	BU1032
8042	P723	11/03/1993 12:00:00 AM	25	Net 30	BU1111
8042	QA879.1	22/05/1993 12:00:00 AM	30	Net 30	PC1035

state	qty
CA	493
OR	493
WA	493
Total	**493**



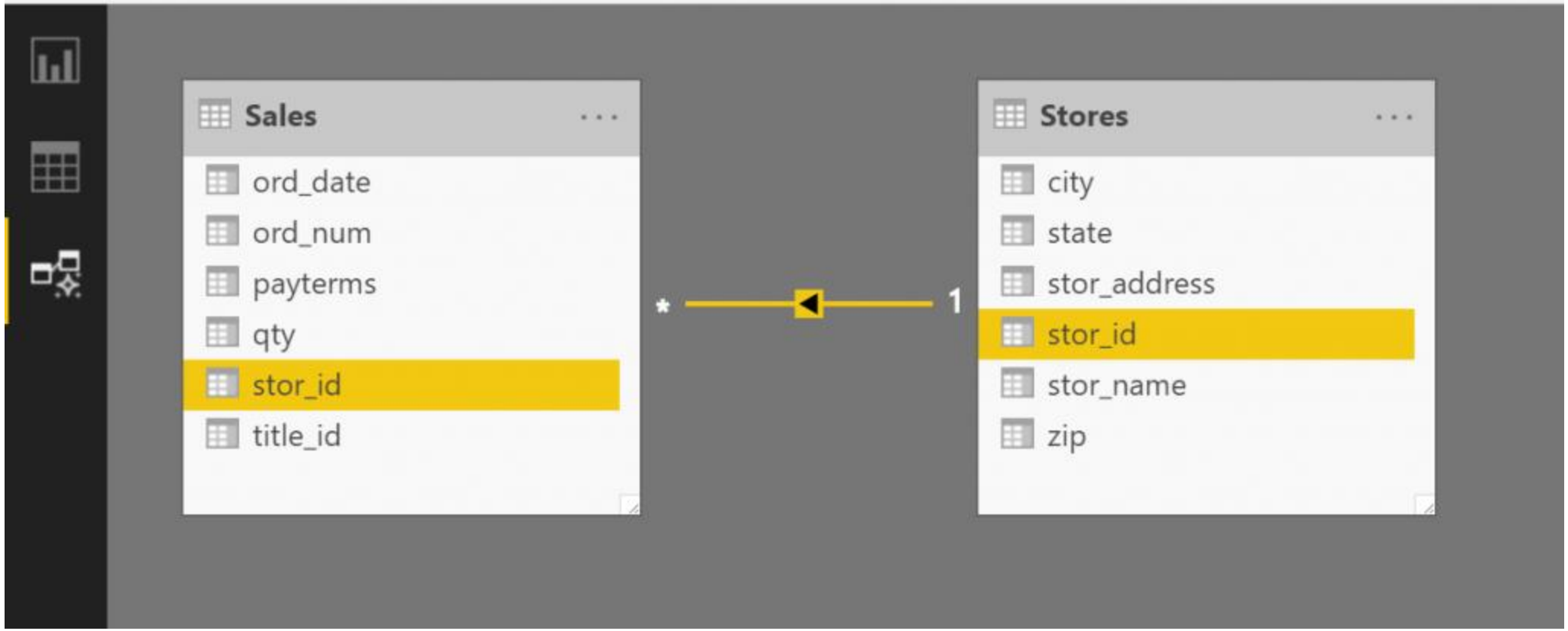
stor_id	stor_name	stor_address	city	state	zip
6380	Eric the Read Books	788 Catamaugus Ave.	Seattle	WA	98056
7066	Barnum's	567 Pasadena Ave.	Tustin	CA	92789
7067	News & Brews	577 First St.	Los Gatos	CA	96745
7131	Doc-U-Mat: Quality Laundry and Books	24-A Avogadro Way	Remulade	WA	98014
7896	Fricative Bookshop	89 Madison St.	Fremont	CA	90019
8042	Bookbeat	679 Carson St.	Portland	OR	89076

Lookup

stor_id	stor_name	stor_address	city	state	zip
6380	Eric the Read Books	788 Catamaugus Ave.	Seattle	WA	98056
7066	Barnum's	567 Pasadena Ave.	Tustin	CA	92789
7067	News & Brews	577 First St.	Los Gatos	CA	96745
7131	Doc-U-Mat: Quality Laundry and Books	24-A Avogadro Way	Remulade	WA	98014
7896	Fricative Bookshop	89 Madison St.	Fremont	CA	90019
8042	Bookbeat	679 Carson St.	Portland	OR	89076

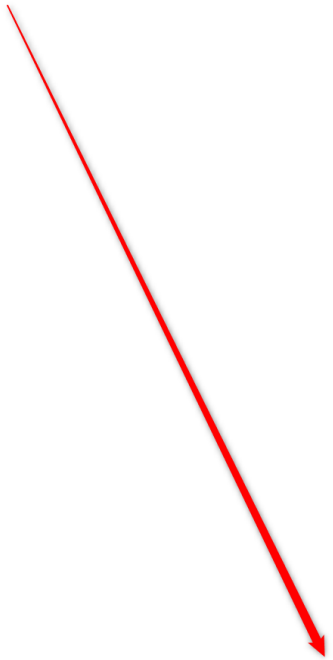
stor_id	ord_num	ord_date	qty	payterms	title_id
6380	6871	14/09/1994 12:00:00 AM	5	Net 60	BU1032
6380	722a	13/09/1994 12:00:00 AM	3	Net 60	PS2091
7066	A2976	24/05/1993 12:00:00 AM	50	Net 30	PC8888
7066	QA7442.3	13/09/1994 12:00:00 AM	75	ON invoice	PS2091
7067	D4482	14/09/1994 12:00:00 AM	10	Net 60	PS2091
7067	P2121	15/06/1992 12:00:00 AM	40	Net 30	TC3218
7067	P2121	15/06/1992 12:00:00 AM	20	Net 30	TC4203
7067	P2121	15/06/1992 12:00:00 AM	20	Net 30	TC7777
7131	N914008	14/09/1994 12:00:00 AM	20	Net 30	PS2091
7131	N914014	14/09/1994 12:00:00 AM	25	Net 30	MC3021
7131	P3087a	29/05/1993 12:00:00 AM	20	Net 60	PS1372
7131	P3087a	29/05/1993 12:00:00 AM	25	Net 60	PS2106
7131	P3087a	29/05/1993 12:00:00 AM	15	Net 60	PS3333
7131	P3087a	29/05/1993 12:00:00 AM	25	Net 60	PS7777
7896	QQ2299	28/10/1993 12:00:00 AM	15	Net 60	BU7832
7896	TQ456	12/12/1993 12:00:00 AM	10	Net 60	MC2222
7896	X999	21/02/1993 12:00:00 AM	35	ON invoice	BU2075
8042	423LL922	14/09/1994 12:00:00 AM	15	ON invoice	MC3021
8042	423LL930	14/09/1994 12:00:00 AM	10	ON invoice	BU1032
8042	P723	11/03/1993 12:00:00 AM	25	Net 30	BU1111
8042	QA879.1	22/05/1993 12:00:00 AM	30	Net 30	PC1035

Relationship means Filtering



Filtering one table by another

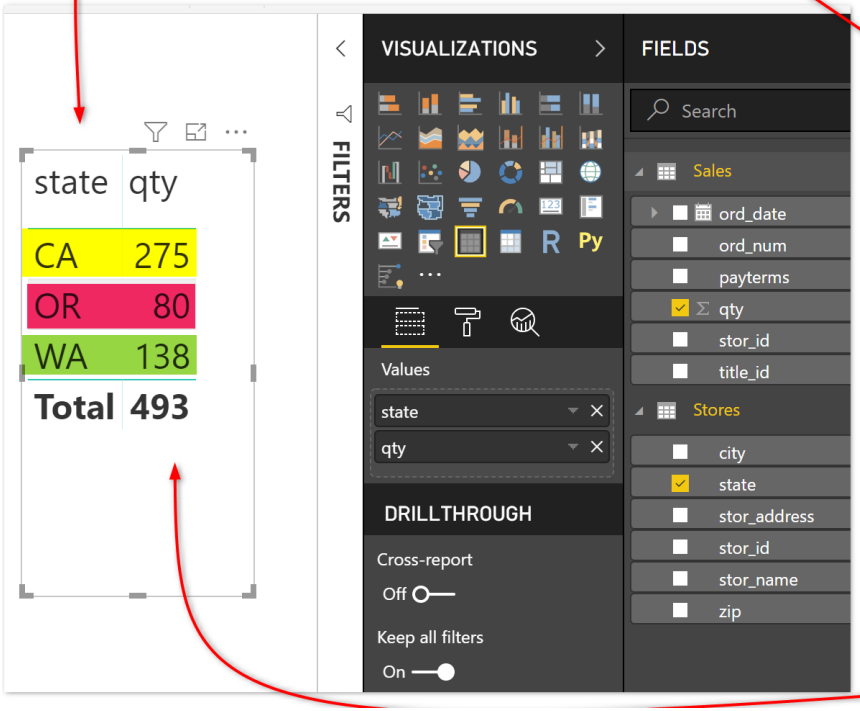
stor_id	stor_name	stor_address	city	state	zip
6380	Eric the Read Books	788 Catamaugus Ave.	Seattle	WA	98056
7066	Barnum's	567 Pasadena Ave.	Tustin	CA	92789
7067	News & Brews	577 First St.	Los Gatos	CA	96745
7131	Doc-U-Mat: Quality Laundry and Books	24-A Avogadro Way	Remulade	WA	98014
7896	Fricative Bookshop	89 Madison St.	Fremont	CA	90019
8042	Bookbeat	679 Carson St.	Portland	OR	89076



stor_id	ord_num	ord_date	qty	payterms	title_id
6380	6871	14/09/1994 12:00:00 AM	5	Net 60	BU1032
6380	722a	13/09/1994 12:00:00 AM	3	Net 60	PS2091
7066	A2976	24/05/1993 12:00:00 AM	50	Net 30	PC8888
7066	QA7442.3	13/09/1994 12:00:00 AM	75	ON invoice	PS2091
7067	D4482	14/09/1994 12:00:00 AM	10	Net 60	PS2091
7067	P2121	15/06/1992 12:00:00 AM	40	Net 30	TC3218
7067	P2121	15/06/1992 12:00:00 AM	20	Net 30	TC4203
7067	P2121	15/06/1992 12:00:00 AM	20	Net 30	TC7777
7131	N914008	14/09/1994 12:00:00 AM	20	Net 30	PS2091
7131	N914014	14/09/1994 12:00:00 AM	25	Net 30	MC3021
7131	P3087a	29/05/1993 12:00:00 AM	20	Net 60	PS1372
7131	P3087a	29/05/1993 12:00:00 AM	25	Net 60	PS2106
7131	P3087a	29/05/1993 12:00:00 AM	15	Net 60	PS3333
7131	P3087a	29/05/1993 12:00:00 AM	25	Net 60	PS7777
7896	QQ2299	28/10/1993 12:00:00 AM	15	Net 60	BU7832
7896	TQ456	12/12/1993 12:00:00 AM	10	Net 60	MC2222
7896	X999	21/02/1993 12:00:00 AM	35	ON invoice	BU2075
8042	423LL922	14/09/1994 12:00:00 AM	15	ON invoice	MC3021
8042	423LL930	14/09/1994 12:00:00 AM	10	ON invoice	BU1032
8042	P723	11/03/1993 12:00:00 AM	25	Net 30	BU1111
8042	QA879.1	22/05/1993 12:00:00 AM	30	Net 30	PC1035

Tables can filter each other through the relationship

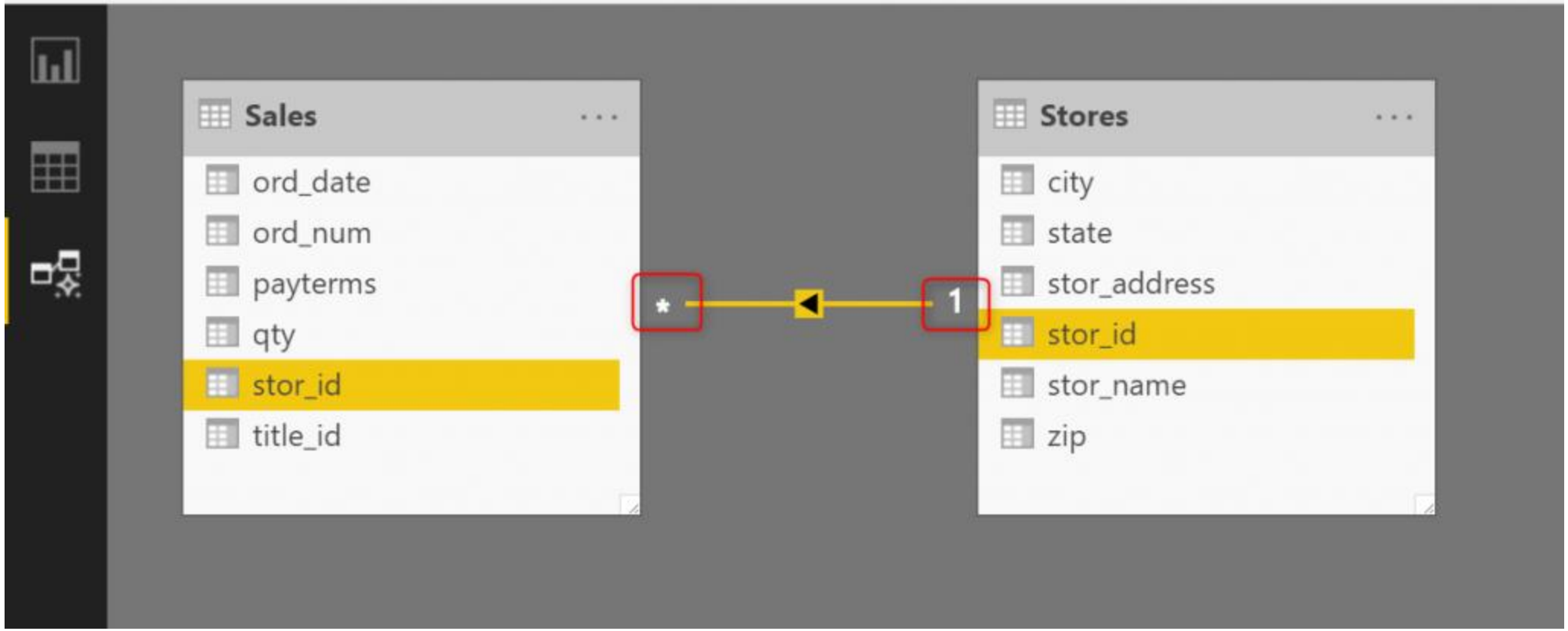
stor_id	stor_name	stor_address	city	state	zip
6380	Eric the Read Books	788 Catamaugus Ave.	Seattle	WA	98056
7066	Barnum's	567 Pasadena Ave.	Tustin	CA	92789
7067	News & Brews	577 First St.	Los Gatos	CA	96745
7131	Doc-U-Mat: Quality Laundry and Books	24-A Avogadro Way	Remulade	WA	98014
7896	Fricative Bookshop	89 Madison St.	Fremont	CA	90019
8042	Bookbeat	679 Carson St.	Portland	OR	89076



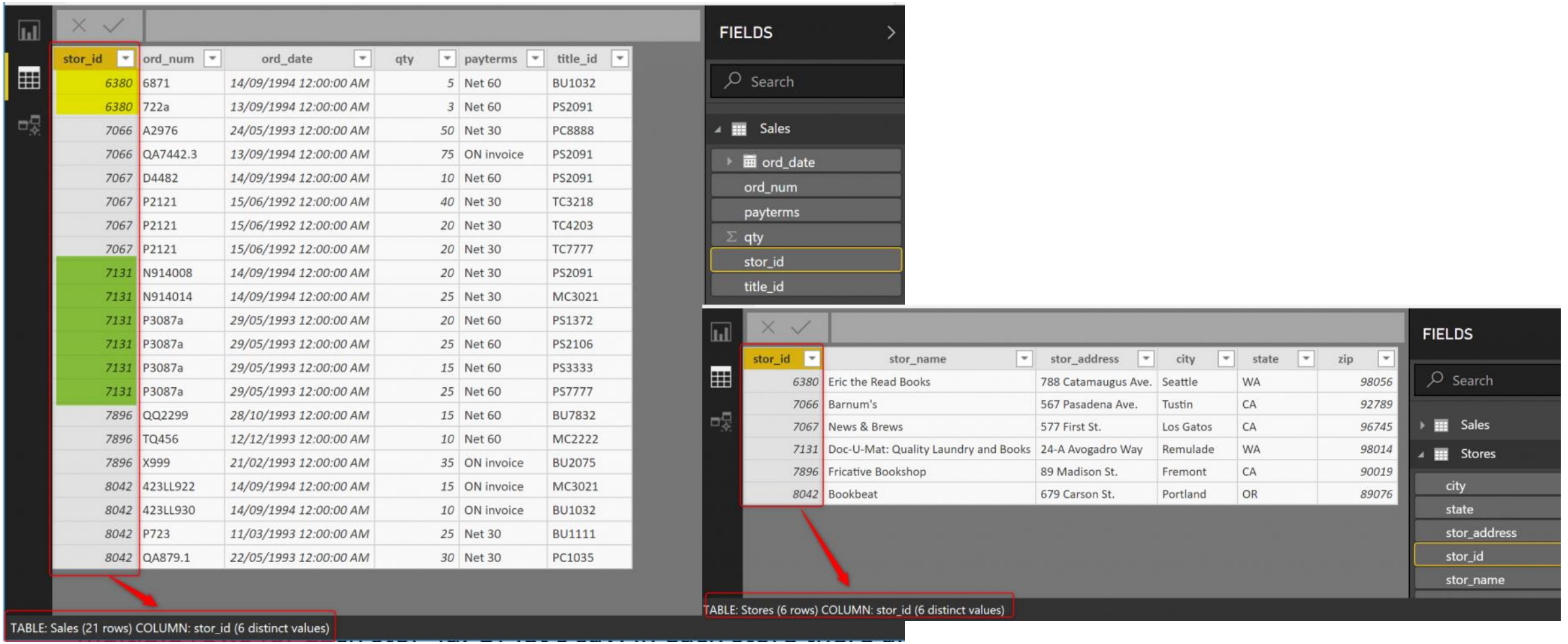
stor_id	ord_num	ord_date	qty	payterms	title_id
6380	6871	14/09/1994 12:00:00 AM	5	Net 60	BU1032
6380	722a	13/09/1994 12:00:00 AM	3	Net 60	PS2091
7066	A2976	24/05/1993 12:00:00 AM	50	Net 30	PC8888
7066	QA7442.3	13/09/1994 12:00:00 AM	75	ON invoice	PS2091
7067	D4482	14/09/1994 12:00:00 AM	10	Net 60	PS2091
7067	P2121	15/06/1992 12:00:00 AM	40	Net 30	TC3218
7067	P2121	15/06/1992 12:00:00 AM	20	Net 30	TC4203
7067	P2121	15/06/1992 12:00:00 AM	20	Net 30	TC7777
7131	N914008	14/09/1994 12:00:00 AM	20	Net 30	PS2091
7131	N914014	14/09/1994 12:00:00 AM	25	Net 30	MC3021
7131	P3087a	29/05/1993 12:00:00 AM	20	Net 60	PS1372
7131	P3087a	29/05/1993 12:00:00 AM	25	Net 60	PS2106
7131	P3087a	29/05/1993 12:00:00 AM	15	Net 60	PS3333
7131	P3087a	29/05/1993 12:00:00 AM	25	Net 60	PS7777
7896	QQ2299	28/10/1993 12:00:00 AM	15	Net 60	BU7832
7896	TQ456	12/12/1993 12:00:00 AM	10	Net 60	MC2222
7896	X999	21/02/1993 12:00:00 AM	35	ON invoice	BU2075
8042	423LL922	14/09/1994 12:00:00 AM	15	ON invoice	MC3021
8042	423LL930	14/09/1994 12:00:00 AM	10	ON invoice	BU1032

Cardinality of the Relationship?

Cardinality in the relationship



Unique values

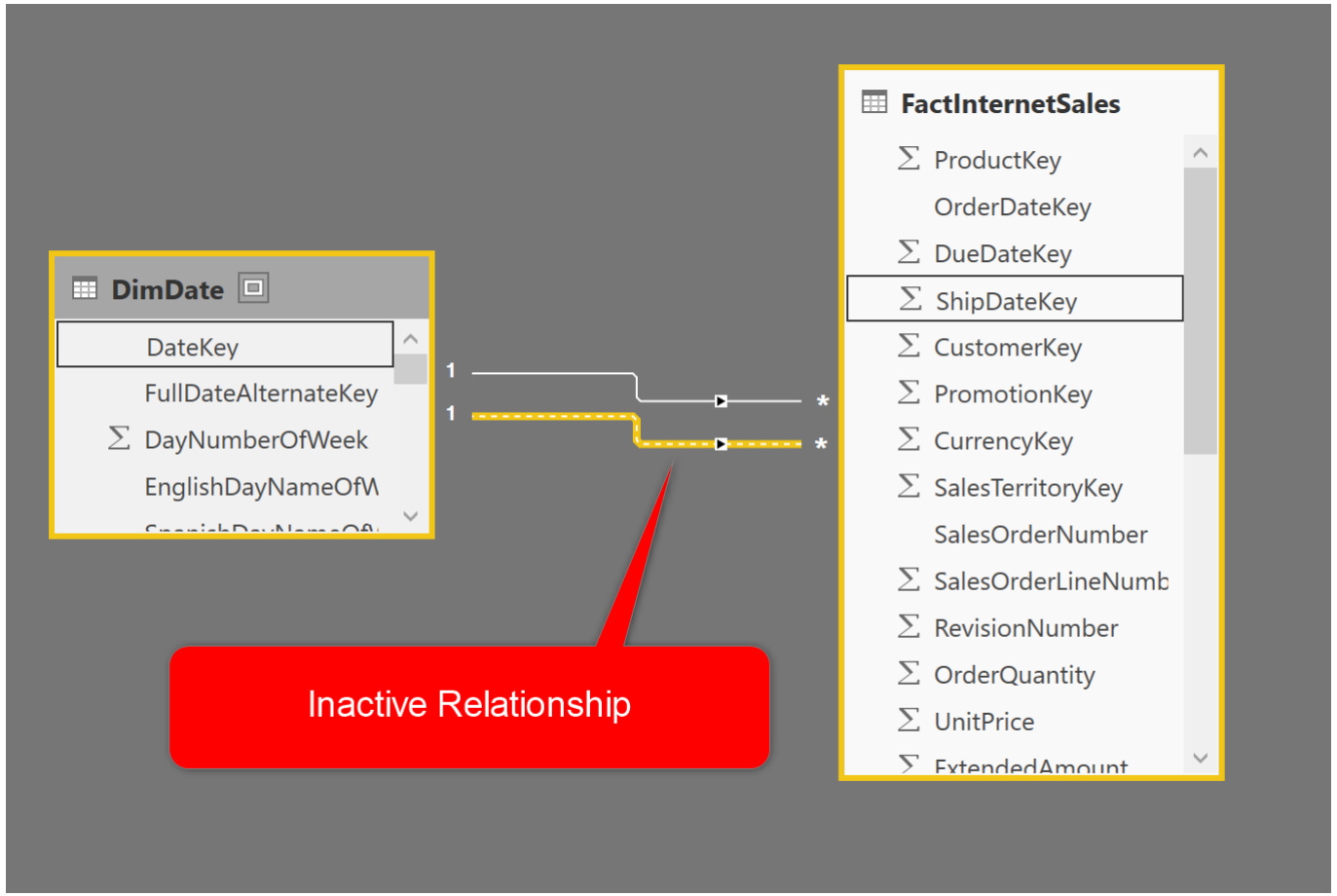


Cardinality

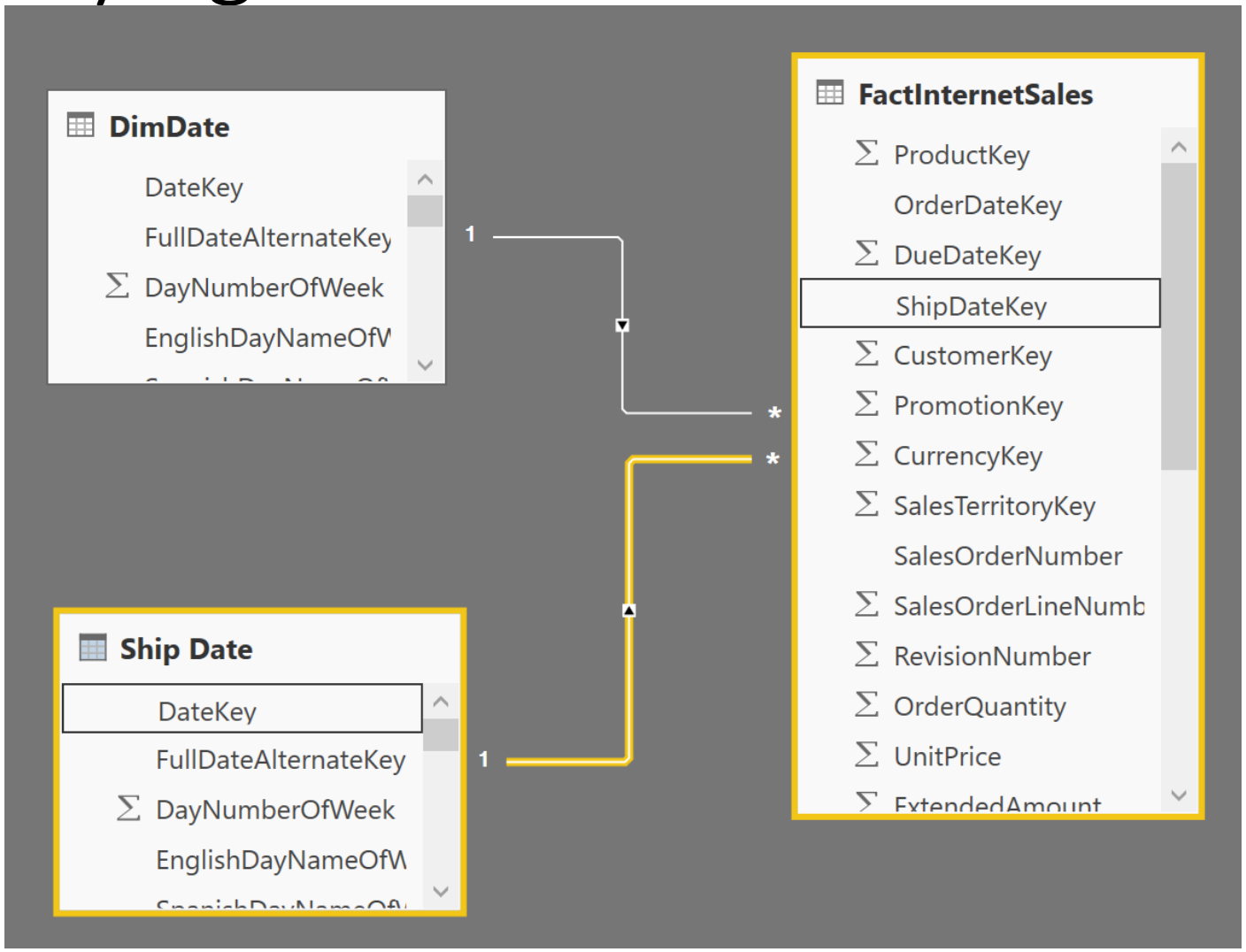
- One to Many
- Many to one
- One to One
- Many to Many

Active or Inactive Relationships?

If you have more than one relationship between two tables, only the first one is active

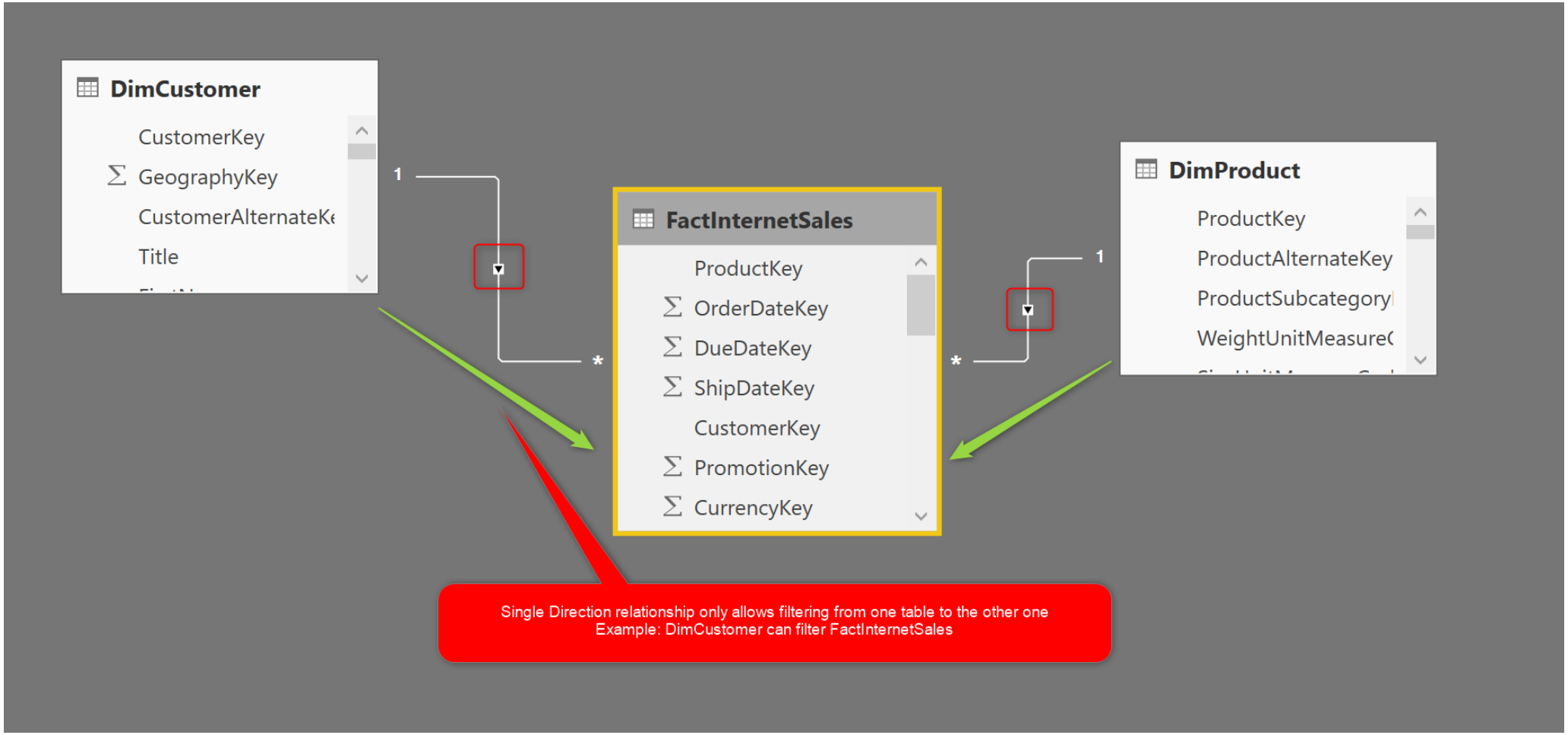


Role Playing Dimension

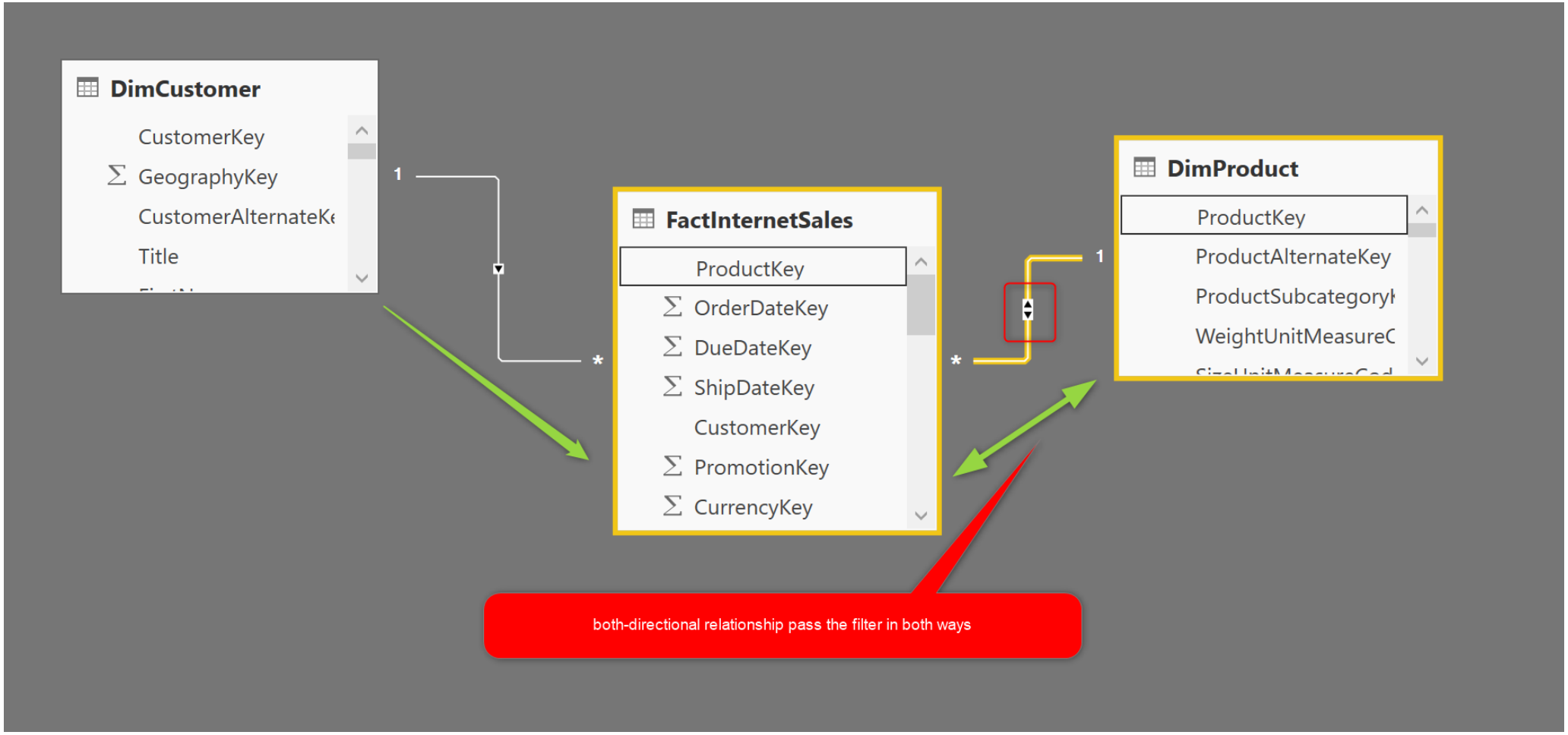


The Direction of Relationships

What is the meaning of the direction of the relationship



You can change the direction



Caution! Use these rarely and carefully

Many to Many relationship

Both-directional relationship

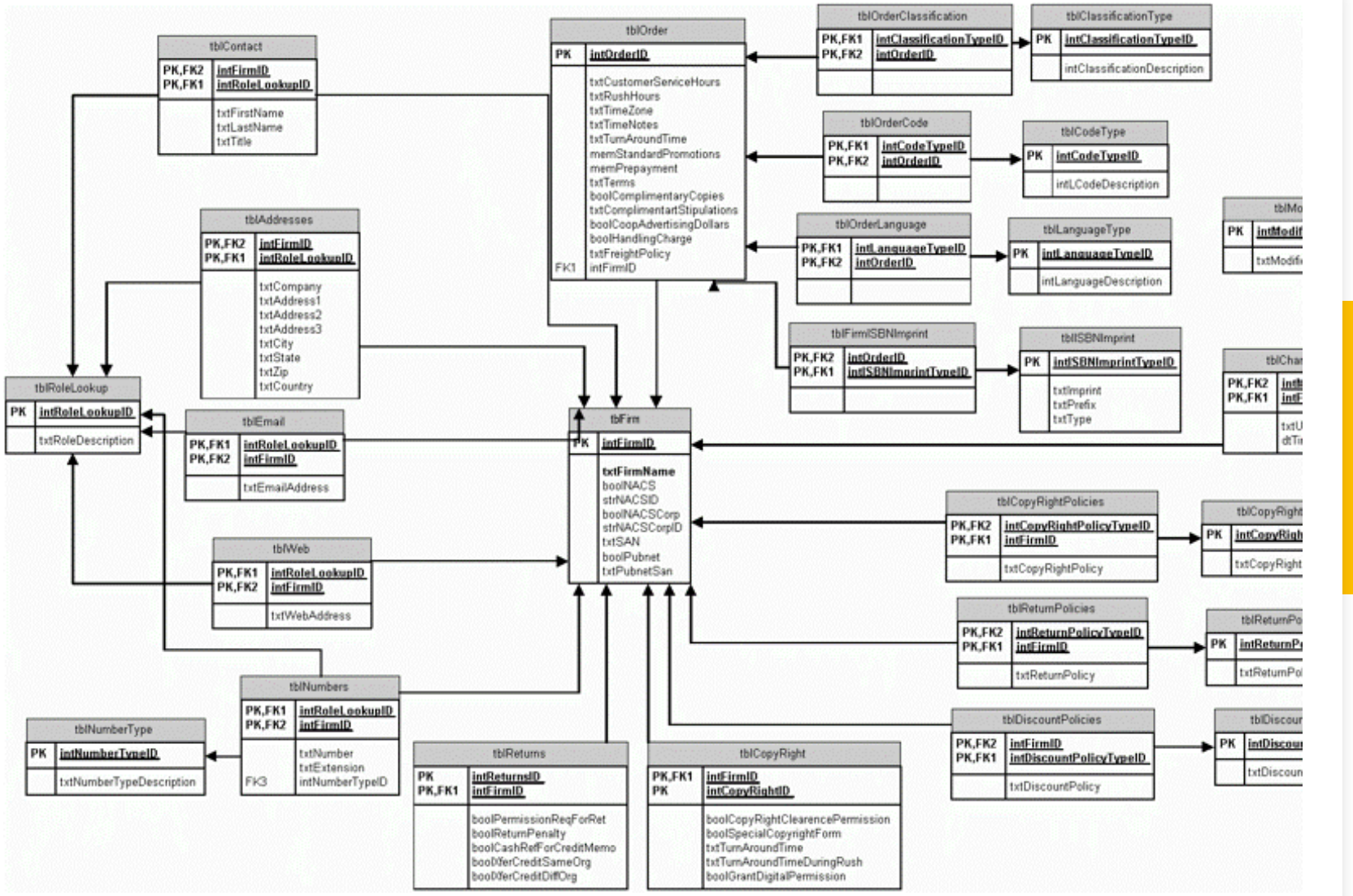
Inactive relationship without the UseRelationship measure

Relationship with multiple columns

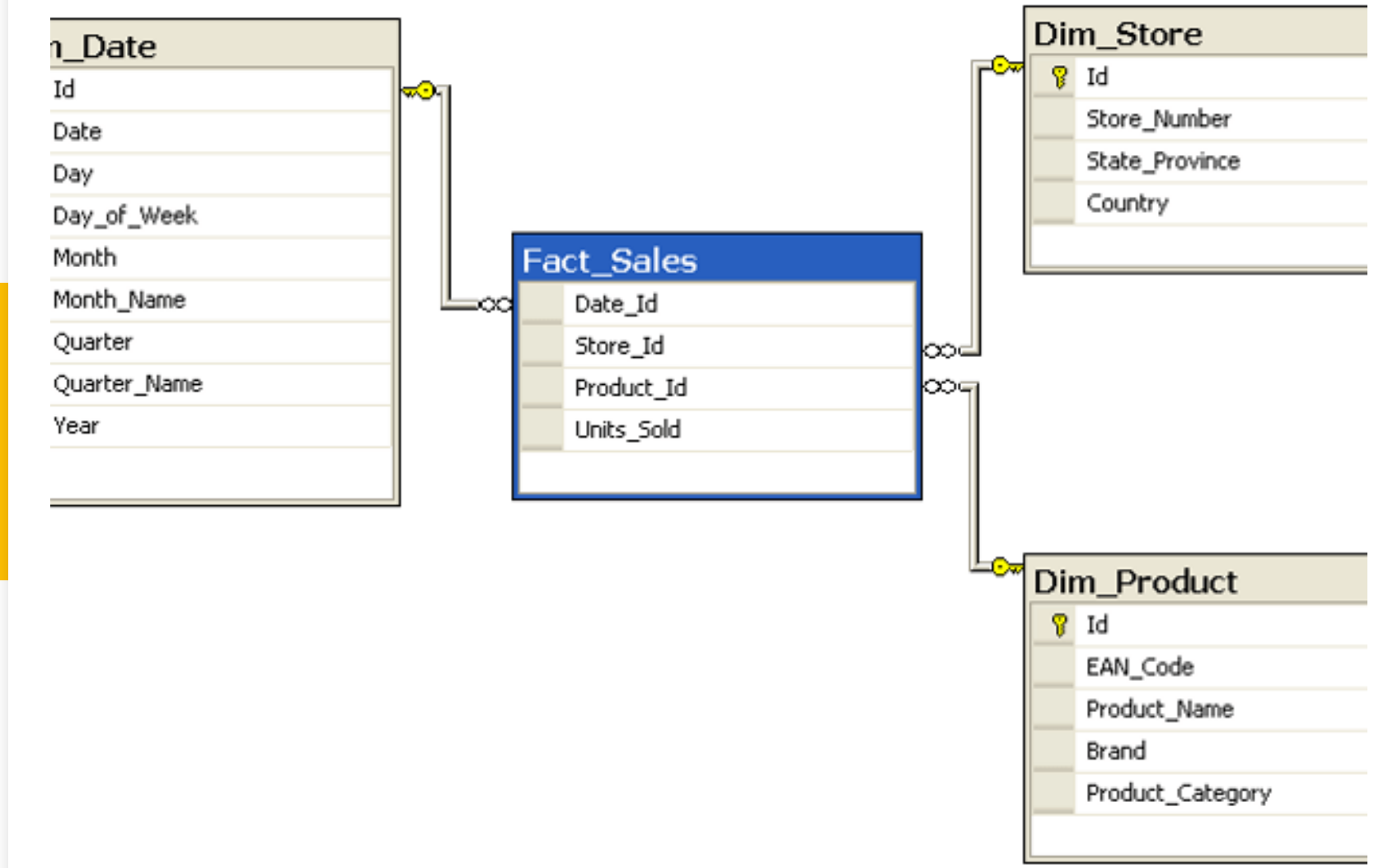
Module 2: Dimensional Modelling

Data preparation

The first and foremost important task in Power BI



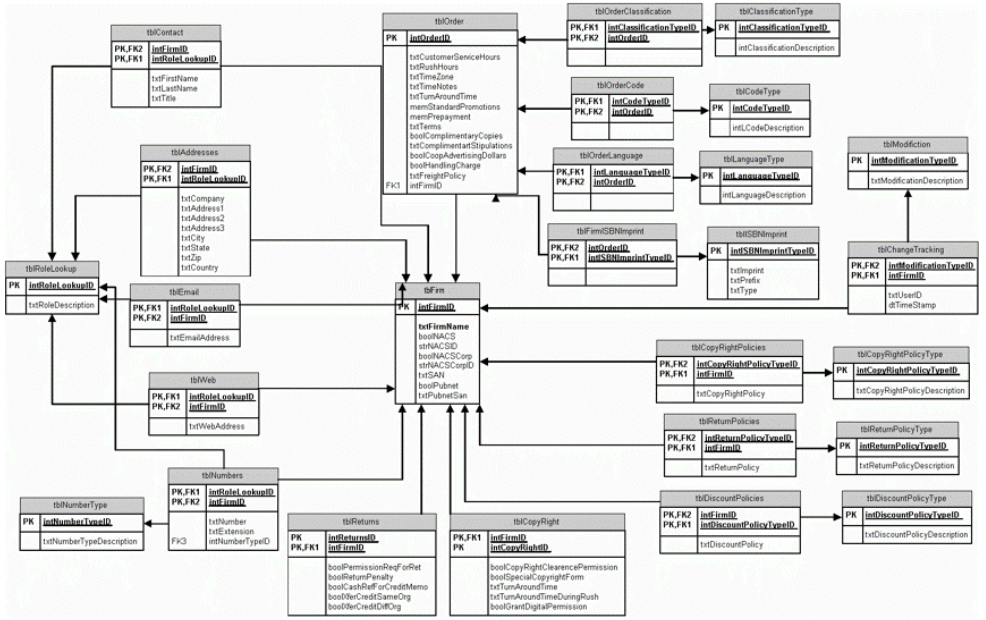
Operational database diagram



Reporting database diagram

You need a different layout

- You cannot use the table layouts as it is in the operational system or transactional database
- There should be a much simpler yet more effective layout for Power BI modelling.



There are two types of tables

You don’t have just one single table!

Not all tables are the same!

Some tables should be used as slicers

Some tables should be used as value

Tables that are used as slicer should filter those tables that should be used as value

We need to build a different Diagram

What is a Dimension Table and Why Say No to a Single Big Table

There are two types of tables

You don’t have just one single table!

Not all tables are the same!

Some tables should be used as slicers

Some tables should be used as value

Tables that are used as slicer should filter those tables that should be used as value

We need to build a different Diagram

Two types of fields

*Fields from a Dimension table are always used for **slicing and dicing** data of other tables (either through filters or slicers, or through axis or legend of the same visual), We call these fields, **Descriptive** fields.*

*Fields that are the **VALUE** part of a visualization (not for table, and a few other visuals), and sliced and diced by descriptive fields, often aggregated are **FACT** fields.*

Descriptive fields

Golden Rule: Fact and Descriptive Fields Not in the Same Table!

FirstName	LastName	EnglishEducation	MaritalStatus	Gender	BirthDate	YearlyIncome	SalesAmount	OrderQuantity
Jon	Yang	Bachelors	M	M	8/04/1966	90000	4892	4
Eugene	Huang	Bachelors	S	M	14/05/1965	60000	4903	2
Ruben	Torres	Bachelors	M	M	12/08/1965	60000	1075	6
Christy	Zhu	Bachelors	S	F	15/02/1968	70000	4891	3
Elizabeth	Johnson	Bachelors	S	F	8/08/1968	80000	1743	2
Julio	Ruiz	Bachelors	S	M	5/08/1965	70000	1718	10
Janet	Alvarez	Bachelors	S	F	6/12/1965	70000	1957	5
Marco	Mehta	Bachelors	M	M	9/05/1964	60000	3698	10
Rob	Verhoff	Bachelors	S	F	7/07/1964	60000	3508	9
Shannon	Carlson	Bachelors	S	M	1/04/1964	70000	4139	7
Jacquelyn	Suarez	Bachelors	S	F	6/02/1964	70000	3423	7
Curtis	Lu	Bachelors	M	M	4/11/1963	60000	4488	2
Lauren	Walker	Bachelors	M	F	18/01/1968	100000	3820	9
Ian	Jenkins	Bachelors	M	M	6/08/1968	100000	2581	1
Sydney	Bennett	Bachelors	S	F	9/05/1968	100000	2033	1
Chloe	Young	Partial College	S	F	27/02/1979	30000	3821	2
Wyatt	Hill	Partial College	M	M	28/04/1979	30000	3430	1
Shannon	Wang	High School	S	F	26/06/1944	20000	1324	2
Clarence	Rai	Partial College	S	M	9/10/1944	30000	1523	8
Luke	Lal	High School	S	M	7/03/1978	40000	1209	10
Jordan	King	High School	S	M	20/09/1978	40000	2808	10
Destiny	Wilson	Partial College	S	F	3/09/1978	40000	3732	2
Ethan	Zhang	Partial College	M	M	12/10/1978	40000	1765	7
Seth	Edwards	Partial College	M	M	11/10/1978	40000	4441	7
Russell	Xie	Partial College	M	M	17/09/1978	60000	4611	4
Alejandro	Beck	Partial High School	M	M	23/12/1945	10000	4195	6
Harold	Sai	Partial College	S	M	3/04/1946	30000	3392	5
Jessie	Zhao	Partial College	M	M	7/12/1946	30000	1284	1

Say No to a Single Table Including Everything!

- Maintenance of such a big table would be always a challenge. Every time you bring a new set of fields, you have to combine them into this table.
- The level of details stored in this table is dependent on the fields in it. If you bring new fields, you need to store more details! This is also a maintenance problem.
- Combining is not always possible! What if you also want to bring a budget into the same table? most of the time is not possible, because you may not have the budget data in the level of customer, you may have it only at the level of the month, and product. then you end up with multiple very large tables.

Maintenance Challenge

FirstName	LastName	EnglishEducation	MaritalStatus	Gender	BirthDate	YearlyIncome	SalesAmount	OrderQuantity	ProductName	ProductColor
Jon	Yang	Bachelors	M	M	8/04/1966	90000	4865	2	XYZ12	Red
Eugene	Huang	Bachelors	S	M	14/05/1965	60000	3400	10	ABC34	Green
Ruben	Torres	Bachelors	M	M	12/08/1965	60000	4116	7	XYZ12	Red
Christy	Zhu	Bachelors	S	F	15/02/1968	70000	1350	10	XYZ13	Red
Elizabeth	Johnson	Bachelors	S	F	8/08/1968	[illegible]	3302	2	ABC35	Green
Julio	Ruiz	Ba[illegible]					[illegible]9	7	XYZ13	Red
Janet	Alvarez	Ba[illegible]					[illegible]3	4	XYZ14	Red
Marco	Mehta	Ba[illegible]					[illegible]91	5	ABC36	Green
Rob	Verhoff	Bachelors	S	F	7/07/1964	60000	3859	4	XYZ14	Red
Shannon	Carlson	Bachelors	S	M	1/04/1964	70000	3175	9	XYZ15	Red
Jacquelyn	Suarez	Bachelors	S	F	6/02/1964	70000	3103	4	ABC37	Green
Curtis	Lu	Bachelors	M	M	4/11/1963	60000	1235	2	XYZ15	Red
Lauren	Walker	Bachelors	M	F	18/01/1968	100000	1689	5	XYZ16	Red
Ian	Jenkins	Bachelors	M	M	6/08/1968	100000	1479	7	ABC38	Green
Sydney	Bennett	Bachelors	S	F	9/05/1968	100000	4481	8	XYZ16	Red

High maintenance: Every time, you need to combine new fields into this table.

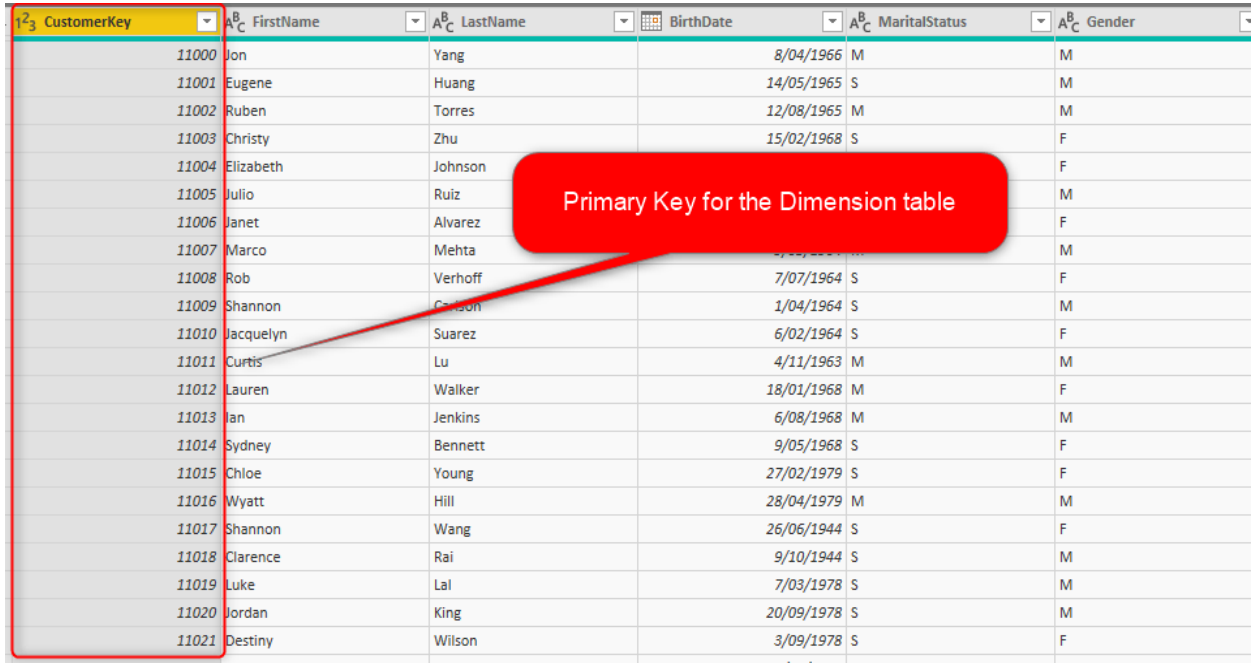
Combining is not always an option

- If the level of details in one or two fact fields is different, then you cannot store them in one table easily

BudgetAmount	ProductName	ProductColor
10000	XYZ12	Red
5000	XYZ14	Red
15000	XYZ16	Red
6000	ABC34	Green

Dimension Table

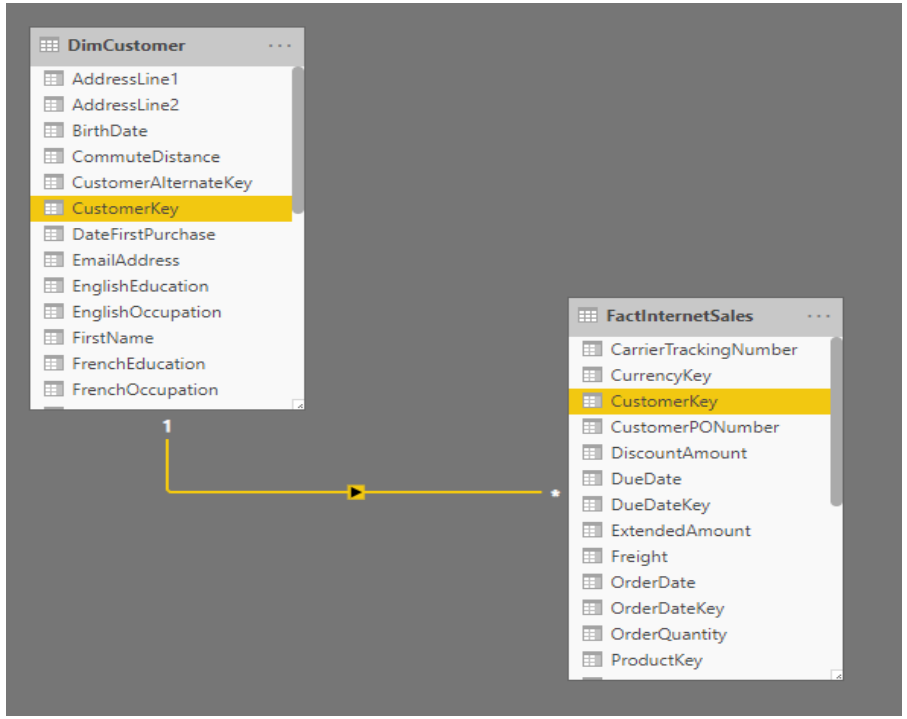
- *Dimension table is a table full of Descriptive Fields and zero Fact fields.*



CustomerKey	FirstName	LastName	BirthDate	MaritalStatus	Gender
11000	Jon	Yang	8/04/1966	M	M
11001	Eugene	Huang	14/05/1965	S	M
11002	Ruben	Torres	12/08/1965	M	M
11003	Christy	Zhu	15/02/1968	S	F
11004	Elizabeth	Johnson			F
11005	Julio	Ruiz			M
11006	Janet	Alvarez			F
11007	Marco	Mehta			M
11008	Rob	Verhoff	7/07/1964	S	F
11009	Shannon	Carlson	1/04/1964	S	M
11010	Jacquelyn	Suarez	6/02/1964	S	F
11011	Curtis	Lu	4/11/1963	M	M
11012	Lauren	Walker	18/01/1968	M	F
11013	Ian	Jenkins	6/08/1968	M	M
11014	Sydney	Bennett	9/05/1968	S	F
11015	Chloe	Young	27/02/1979	S	F
11016	Wyatt	Hill	28/04/1979	M	M
11017	Shannon	Wang	26/06/1944	S	F
11018	Clarence	Rai	9/10/1944	S	M
11019	Luke	Lal	7/03/1978	S	M
11020	Jordan	King	20/09/1978	S	M
11021	Destiny	Wilson	3/09/1978	S	F

Dimension table; Definition

- *Dimension table is full of Descriptive fields and zero Fact fields, and one or more Key fields. Key fields are row identifiers of the dimension table.*



When the Primary Key is Not a Primary Key!

Credit Card No	Firstname	Lastname	Exp Date
1234-1234-1234-1234	Reza	Rad	Mar-00
1111-1111-1111-1111	Leila	Etaati	Jul-99

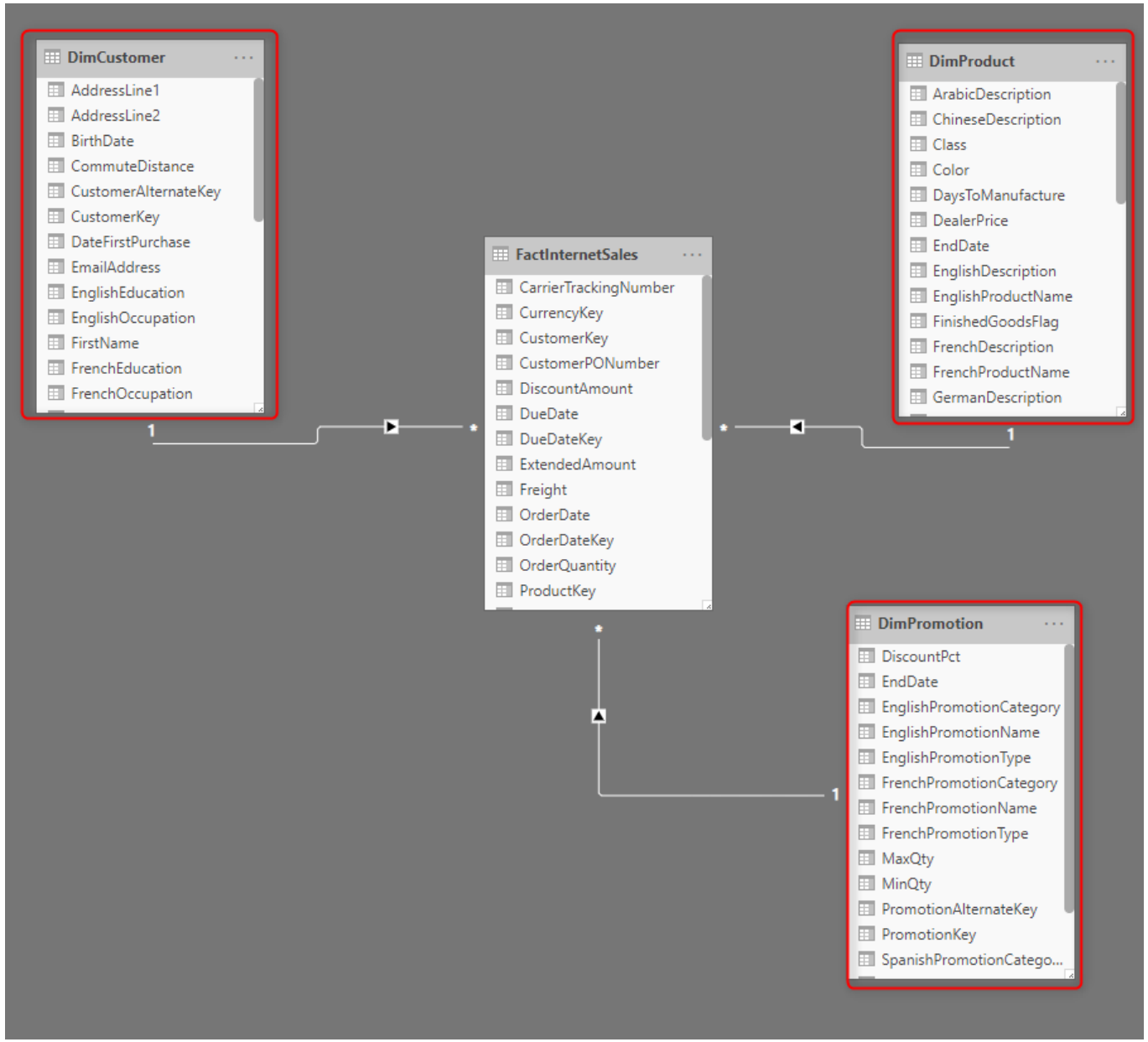
Surrogate Key	Credit Card No	Firstname	Lastname	Exp Date
1	1234-1234-1234-1234	Reza	Rad	Mar-00
2	1111-1111-1111-1111	Leila	Etaati	Jul-99
3	1234-1234-1234-1234	Reza	Rad	Mar-04
4	1234-1234-1234-1234	Reza	Rad	Mar-08

Credit Card No	Firstname	Lastname	Exp Date
1234-1234-1234-1234	Reza	Rad	Mar-00
1111-1111-1111-1111	Leila	Etaati	Jul-99
1234-1234-1234-1234	Reza	Rad	Mar-04
1234-1234-1234-1234	Reza	Rad	Mar-08

Dimension Tables are Wide and Shallow

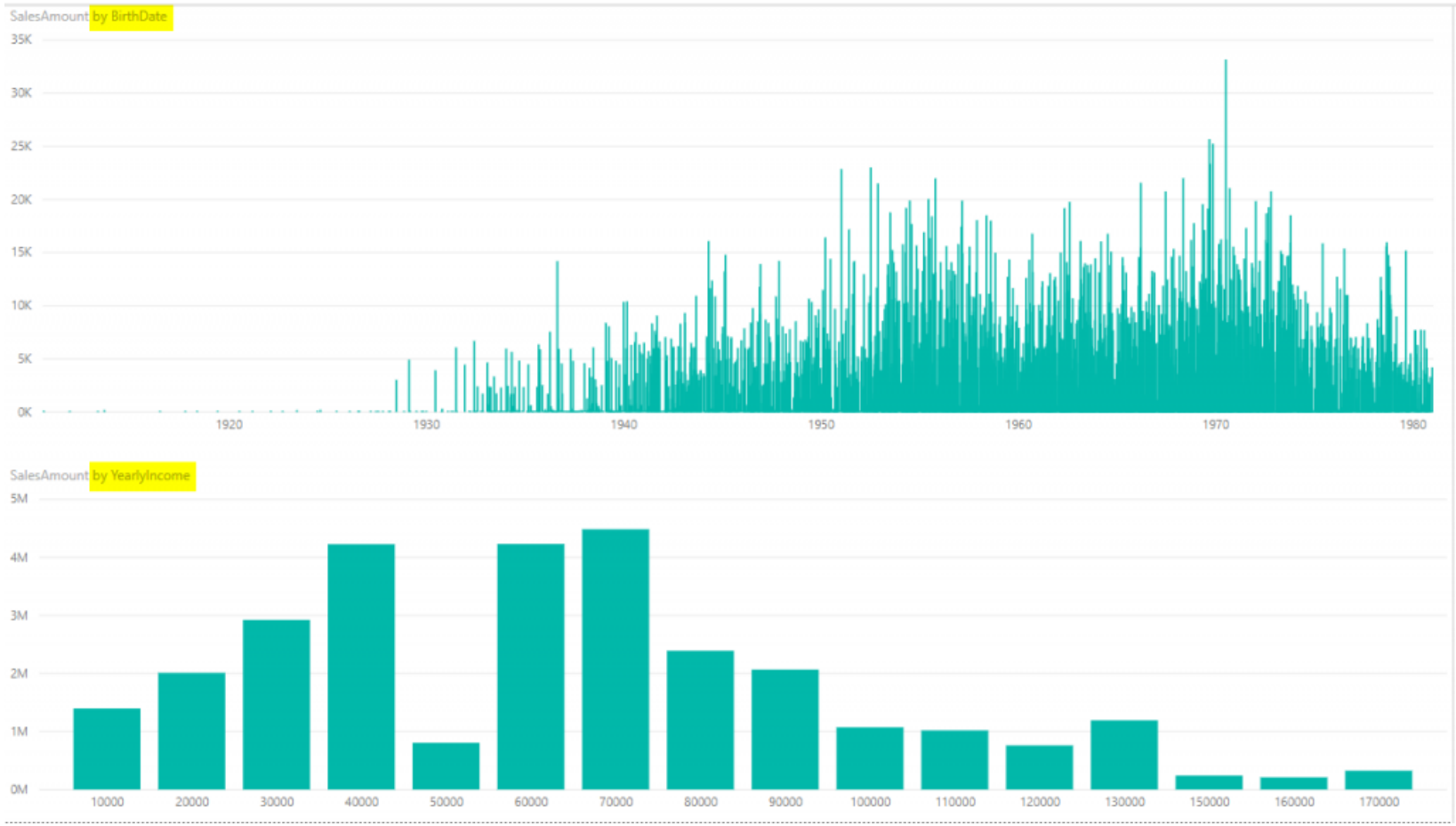
CustomerKey	GeographyKey	CustomerAlternateKey	Title	FirstName	MiddleName	LastName	NameStyle	BirthDate	MaritalStatus	Suffix	Gender	EmailAddress	YearlyIncome	TotalChildren	NumberChildrenAtHom
11471	207	AW00011471		Latasha		Suarez	False	Tuesday, 25 September 1973	S		F	latasha19@adventure-works.com	30000	0	
11602	135	AW00011602		Larry		Gill	False	Wednesday, 13 April 1977	S		M	larry16@adventure-works.com	30000	0	
11603	244	AW00011603		Geoffrey		Gonzalez	False	Sunday, 6 February 1977	S		M	geoffrey16@adventure-works.com	30000	0	
11604	275	AW00011604		Edgar		Sanchez	False	Friday, 3 June 1977	S		M	edgar21@adventure-works.com	30000	0	
11610	269	AW00011610		Blake		Collins	False	Wednesday, 23 April 1975	S		M	blake47@adventure-works.com	30000	0	
12515	189	AW00012515		Shelby		Bailey	False	Friday, 3 June 1977	S		F	shelby16@adventure-works.com	30000	0	
12517	133	AW00012517		Alexa		Watson	False	Thursday, 25 August 1977	S		F	alexa0@adventure-works.com	30000	0	
12518	161	AW00012518		Jacquelyn		Dominguez	False	Tuesday, 27 September 1977	S		F	jacquelyn13@adventure-works.com	30000	0	
12519	265	AW00012519		Casey		Gutierrez	False	Saturday, 17 December 1977	S		M	casey34@adventure-works.com	30000	0	
12524	211	AW00012524		Kate		Shan	False	Friday, 24 January 1975	S		F	kate8@adventure-works.com	30000	0	
12714	157	AW00012714		Colleen		Lu	False	Tuesday, 17 July 1973	S		F	colleen11@adventure-works.com	30000	0	
12716	185	AW00012716		Dale		Shen	False	Saturday, 16 March 1974	S		M	dale2@adventure-works.com	30000	0	
12718	193	AW00012718		Tammy		Sai	False	Thursday, 14 November 1974	S		F	tammy6@adventure-works.com	30000	0	
12728	131	AW00012728		Jeremiah		Stewart	False	Tuesday, 26 June 1979	S		M	jeremiah44@adventure-works.com	30000	0	
12871	233	AW00012871		Leah		Li	False	Wednesday, 6 October 1976	S		F	leah2@adventure-works.com	30000	0	
13671	173	AW00013671		Frank		Ramos	False	Thursday, 7 February 1974	S		M	frank25@adventure-works.com	30000	0	

Examples of Dimension Table



Slicer Tables: Dimensions

- They keep Descriptive Information
- Used for slicing and dicing
  - Slicers or filters
  - Axis or legend of charts
- Descriptive doesn't meant TEXT, it also means other types!
- Wide
- Shallow



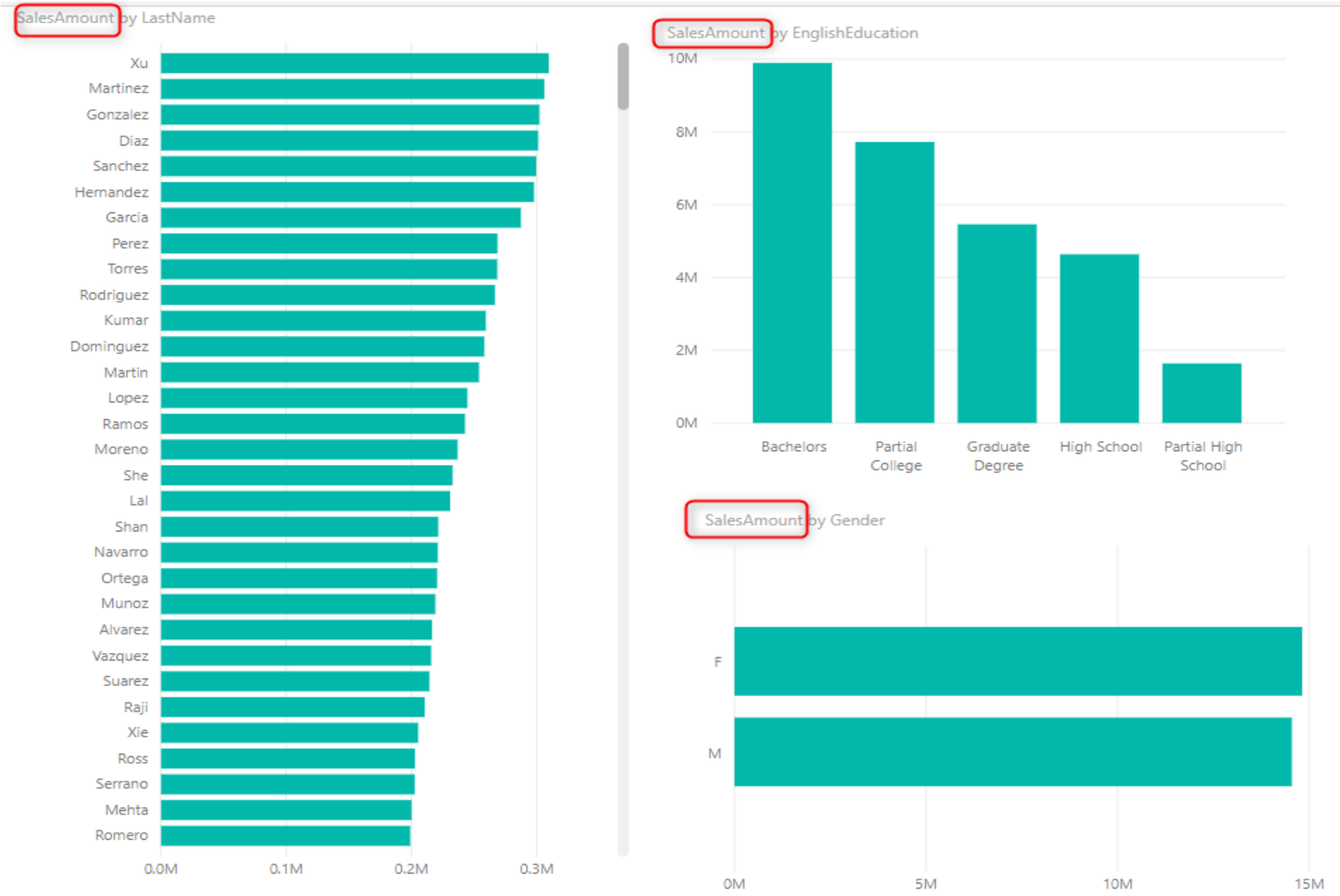
Fact tables

Two types of fields

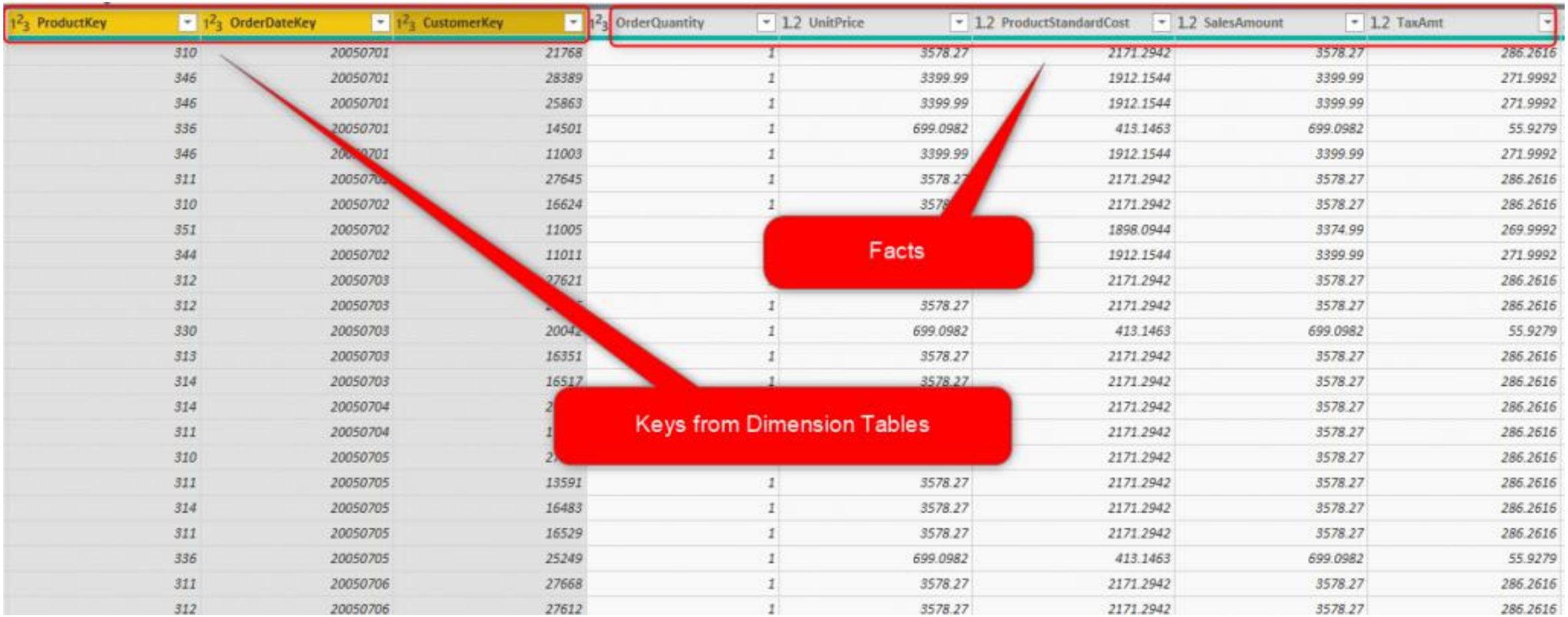
*Fields from a Dimension table are always used for **slicing and dicing** data of other tables (either through filters or slicers, or through axis or legend of the same visual), We call these fields, **Descriptive** fields.*

*Fields that are the **VALUE** part of a visualization (not for table, and a few other visuals), and sliced and diced by descriptive fields, often aggregated are **FACT** fields.*

Facts


SalesAmount by LastName
Xu
Martinez
Gonzalez
Diaz
Sanchez
Hernandez
Garcia
Perez
Torres
Rodriguez
Kumar
Dominguez
Martin
Lopez
Ramos
Moreno
She
Lal
Shan
Navarro
Ortega
Munoz
Alvarez
Vazquez
Suarez
Raji
Xie
Ross
Serrano
Mehta
Romero
0.0M
0.1M
0.2M
0.3M
SalesAmount by EnglishEducation
10M
8M
6M
4M
2M
0M
Bachelors
Partial College
Graduate Degree
High School
Partial High School
SalesAmount by Gender
F
M
0M
5M
10M
15M

Fact Tables



ProductKey	OrderDateKey	CustomerKey	OrderQuantity	UnitPrice	ProductStandardCost	SalesAmount	TaxAmt
310	20050701	21768	1	3578.27	2171.2942	3578.27	286.2616
346	20050701	28389	1	3399.99	1912.1544	3399.99	271.9992
346	20050701	25863	1	3399.99	1912.1544	3399.99	271.9992
336	20050701	14501	1	699.0982	413.1463	699.0982	55.9279
346	20050701	11003	1	3399.99	1912.1544	3399.99	271.9992
311	20050702	27645	1	3578.27	2171.2942	3578.27	286.2616
310	20050702	16624	1	3578	2171.2942	3578.27	286.2616
351	20050702	11005			1898.0944	3374.99	269.9992
344	20050702	11011			1912.1544	3399.99	271.9992
312	20050703	27621			2171.2942	3578.27	286.2616
312	20050703		1	3578.27	2171.2942	3578.27	286.2616
330	20050703	20042	1	699.0982	413.1463	699.0982	55.9279
313	20050703	16351	1	3578.27	2171.2942	3578.27	286.2616
314	20050703	16517	1	3578.27	2171.2942	3578.27	286.2616
314	20050704	2			2171.2942	3578.27	286.2616
311	20050704	1			2171.2942	3578.27	286.2616
310	20050705	2			2171.2942	3578.27	286.2616
311	20050705	13591	1	3578.27	2171.2942	3578.27	286.2616
314	20050705	16483	1	3578.27	2171.2942	3578.27	286.2616
311	20050705	16529	1	3578.27	2171.2942	3578.27	286.2616
336	20050705	25249	1	699.0982	413.1463	699.0982	55.9279
311	20050706	27668	1	3578.27	2171.2942	3578.27	286.2616
312	20050706	27612	1	3578.27	2171.2942	3578.27	286.2616

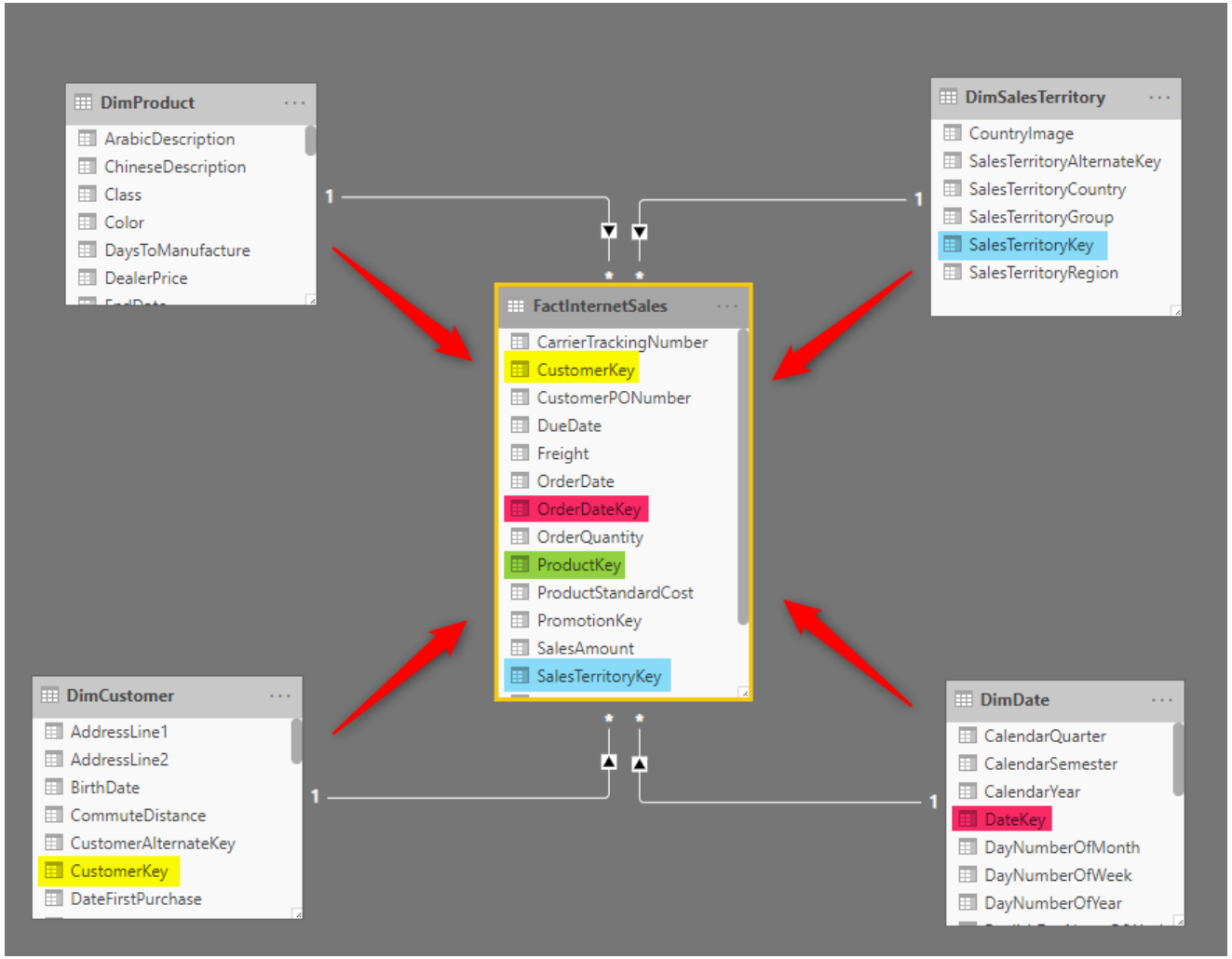
ProductKey	OrderDateKey	CustomerKey	OrderQuantity	UnitPrice	ProductStandardCost	SalesAmount	TaxAmt
310	20050701	21768	1	3578.27	2171.2942	3578.27	286.2616
346	20050701	28389	1	3399.99	1912.1544	3399.99	271.9992
346	20050701	25863	1	3399.99	1912.1544	3399.99	271.9992
336	20050701	14501	1	699.0982	413.1463	699.0982	55.9279
346	20050701	11003	1	3399.99	1912.1544	3399.99	271.9992
311	20050702	27645	1	3578.27	2171.2942	3578.27	286.2616
310	20050702	16624	1	3578.27	2171.2942	3578.27	286.2616
351	20050702	11005	1	3374.99	1898.0944	3374.99	269.9992
344	20050702	11011	1	3399.99	1912.1544	3399.99	271.9992
312	20050703	27621	1	3578.27	2171.2942	3578.27	286.2616

ProductKey	OrderDateKey	CustomerKey	PromotionKey	SalesTerritoryKey	OrderQuantity	UnitPrice	ProductStandardCost	SalesA
310	20050701	21768	1	6	1	3578.27	2171.2942	
346	20050701	28389	1	7	1	3399.99	1912.1544	
346	20050701	25863	1	1	1	3399.99	1912.1544	
336	20050701	14501	1	4	1	699.0982	413.1463	
346	20050701	11003	1	9	1	3399.99	1912.1544	
311	20050702	27645	1	4	1	3578.27	2171.2942	
310	20050702	16624	1	9	1	3578.27	2171.2942	
351	20050702	11005	1	9	1	3374.99	1898.0944	
344	20050702	11011	1	9	1	3399.99	1912.1544	
312	20050703	27621	1	4	1	3578.27	2171.2942	
312	20050703	27616	1	4	1	3578.27	2171.2942	
330	20050703	20042	1	10	1	699.0982	413.1463	

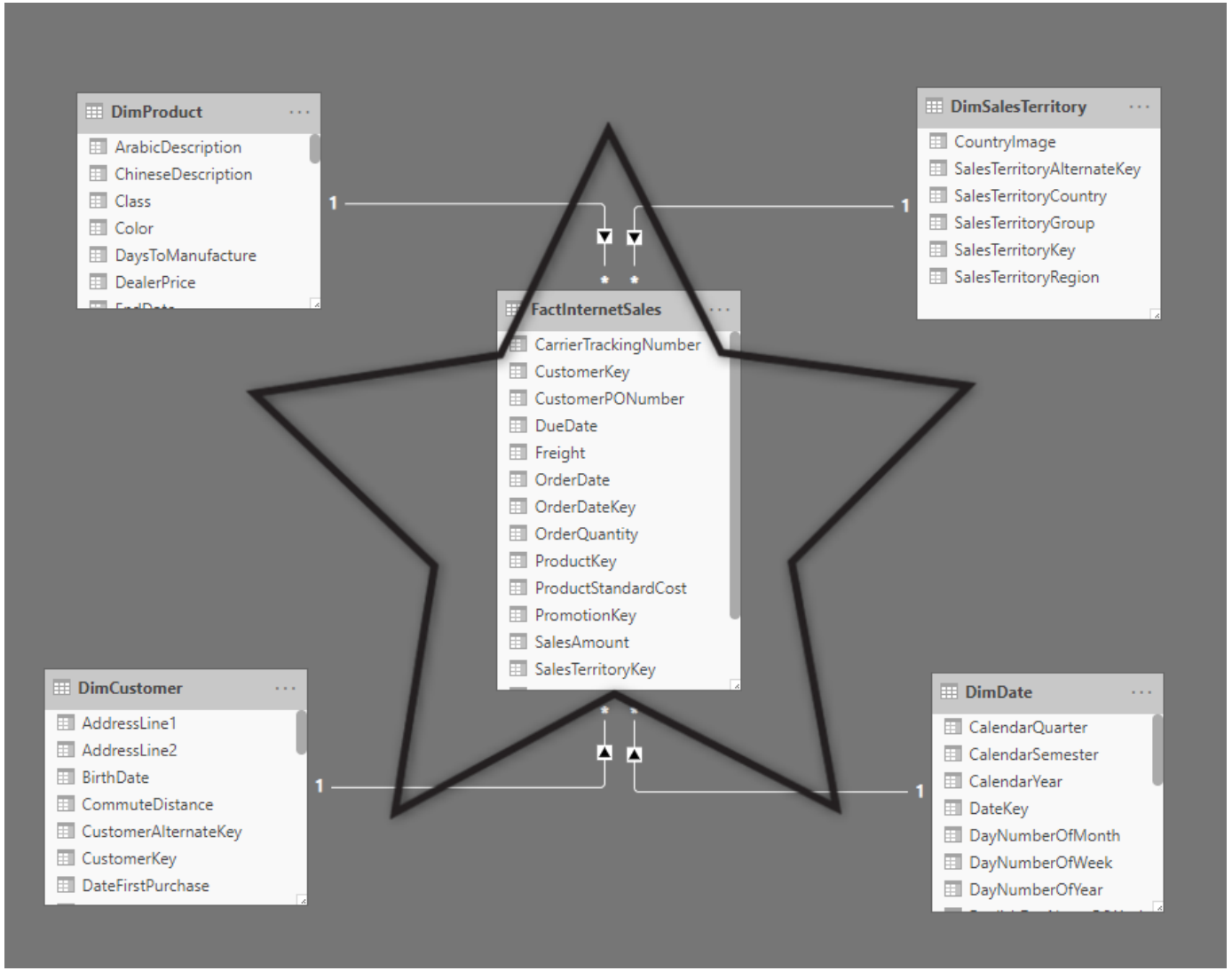
Grain

Relationship between Fact and Dimension tables

- One to many
- From dimension to fact
- Single directional
- Star Schema



Star Schema


DimProduct
ArabicDescription
ChineseDescription
Class
Color
DaysToManufacture
DealerPrice
DimSalesTerritory
CountryImage
SalesTerritoryAlternateKey
SalesTerritoryCountry
SalesTerritoryGroup
SalesTerritoryKey
SalesTerritoryRegion
FactInternetSales
CarrierTrackingNumber
CustomerKey
CustomerPONumber
DueDate
Freight
OrderDate
OrderDateKey
OrderQuantity
ProductKey
ProductStandardCost
PromotionKey
SalesAmount
SalesTerritoryKey
DimCustomer
AddressLine1
AddressLine2
BirthDate
CommuteDistance
CustomerAlternateKey
CustomerKey
DateFirstPurchase
DimDate
CalendarQuarter
CalendarSemester
CalendarYear
DateKey
DayNumberOfMonth
DayNumberOfWeek
DayNumberOfYear
1
1
1
1

Types of fact table

- Transactional
- Snapshot
- Accumulating snapshot
- Bridge

Transactional fact table

- *The grain of a transactional fact table is at the transaction level.*

ProductKey	OrderDateKey	CustomerKey	PromotionKey	SalesTerritoryKey	OrderQuantity	UnitPrice	ProductStandardCost	1.
310	20050701	21768	1	6	1	3578.27	2171.2942	
346	20050701	28389	1	7	1	3399.99	1912.1544	
346	20050701	25863	1	1	1	3399.99	1912.1544	
336	20050701	14501	1	4	1	699.0982	413.1463	
346	20050701	11003	1	9	1	3399.99	1912.1544	
311	20050702	27645	1	4	1	3578.27	2171.2942	
310	20050702	16624	1	9	1	3578.27	2171.2942	
351	20050702	11005	1	9	1	3374.99	1898.0944	
344	20050702	11011	1	9	1	3399.99	1912.1544	
312	20050703	27621	1	4	1	3578.27	2171.2942	
312	20050703	27616	1	4	1	3578.27	2171.2942	
330	20050703	20042	1	10	1	699.0982	413.1463	
313	20050703	16351	1	9	1	3578.27	2171.2942	
314	20050703	16517	1	9	1	3578.27	2171.2942	
314	20050704	27606	1	1	1	3578.27	2171.2942	
311	20050704	13513	1	8	1	3578.27	2171.2942	

Periodic snapshot fact table

- The periodic snapshot fact table is aggregated on the desired period.

ProductKey	PeriodKey	CustomerKey	OrderQuantity	SalesAmount
310	200507	21768	1	3578.27
346	200507	28389	1	3399.99
346	200507	25863	1	3399.99
336	200507	14501	1	699.0982
346	200507	11003	1	3399.99
311	200507	27645	1	3578.27
310	200507	16624	1	3578.27
351	200507	11005	1	3374.99
344	200507	11011	1	3399.99
312	200507	27621	1	3578.27
312	200507	27616	1	3578.27
330	200507	20042	1	699.0982
313	200507	16351	1	3578.27
314	200507	16517	1	3578.27
314	200507	27606	1	3578.27
311	200507	13513	1	3578.27
310	200507	27601	1	3578.27

Accumulating snapshot fact table

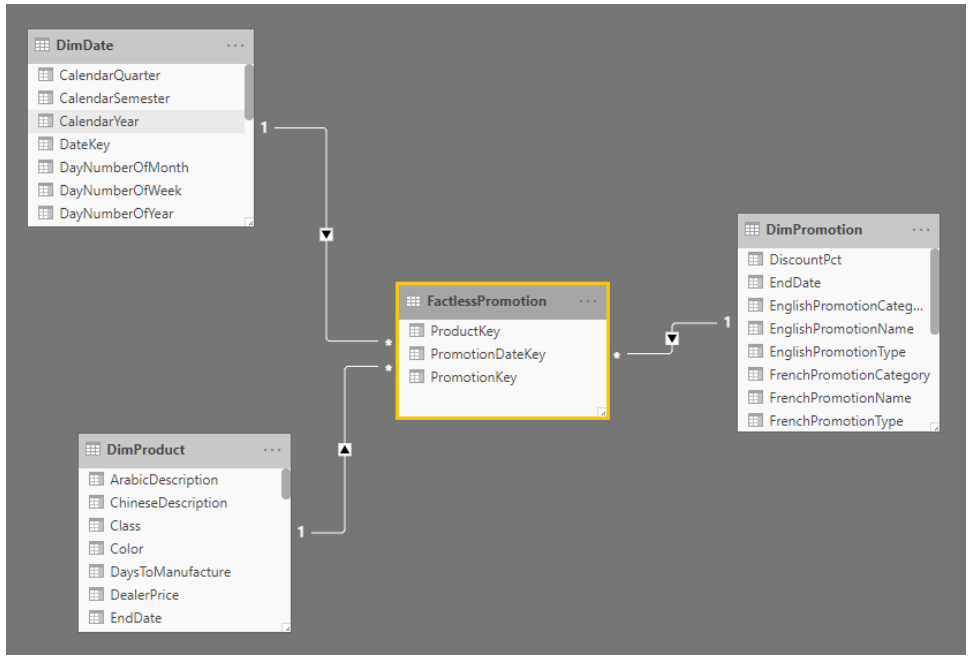
- *The accumulating-snapshot fact table is great for process-oriented analysis, such as workflow.*

WorkOrderKey	StartDate	CreatedBykey	ApprovalStepResultKey	Amount	ApprovedBy	ArchiveStatusKey	ArchiveDate
123	2/02/2019	23		2000			
124	3/03/2019	23	10	200	45		
125	1/10/2018	23	10	500	40	2	20/12/2018

Factless fact table: Bridge table

- *Factless Fact tables are great for analyzing NOT existence of the data.*

ProductKey	PromotionDateKey	PromotionKey
310	20050701	1
346	20050701	1
346	20050701	1
336	20050701	1
346	20050701	1
311	20050702	1
310	20050702	1
351	20050702	1
344	20050702	1
312	20050703	1
312	20050703	1
330	20050703	1
313	20050703	1
314	20050703	1



ProductKey	OrderDateKey	CustomerKey	OrderQuantity	UnitPrice	ProductStandardCost	SalesAmount	TaxAmt
310	20050701	21768	1	3578.27	2171.2942	3578.27	286.2616
346	20050701	28389	1	3399.99	1912.1544	3399.99	271.9992
346	20050701	25863	1	3399.99	1912.1544	3399.99	271.9992
336	20050701	14501	1	699.0982	413.1463	699.0982	55.9279
346	20050701	11003	1	3399.99	1912.1544	3399.99	271.9992
311	20050702	27645	1	3578.27	2171.2942	3578.27	286.2616
310	20050702	16624	1	3578.27	2171.2942	3578.27	286.2616
351	20050702	11005	1	3374.99	1898.0944	3374.99	269.9992
344	20050702	11011	1	3399.99	1912.1544	3399.99	271.9992
312	20050703	27621	1	3578.27	2171.2942	3578.27	286.2616
312	20050703	27616	1	3578.27	2171.2942	3578.27	286.2616
330	20050703	20042	1	699.0982	413.1463	699.0982	55.9279
313	20050703	16351	1	3578.27	2171.2942	3578.27	286.2616
314	20050703	16517	1	3578.27	2171.2942	3578.27	286.2616
314	20050704	27606	1	3578.27	2171.2942	3578.27	286.2616
311	20050704	13513	1	3578.27	2171.2942	3578.27	286.2616
310	20050705	27601	1	3578.27	2171.2942	3578.27	286.2616

Value Tables: Facts

- Explains happening of an action
- Keeps values (numeric, aggregable)
- Used as Values in charts and visuals
- GRAIN: Level of details in the table

Do You Need a Date Dimension?

What is a date dimension?

- Other names: calendar table, date table, etc.
- One row per date
- Not missing any day
- Columns for date attributes: Year, Qtr, Month, Week, IsHoliday, etc.

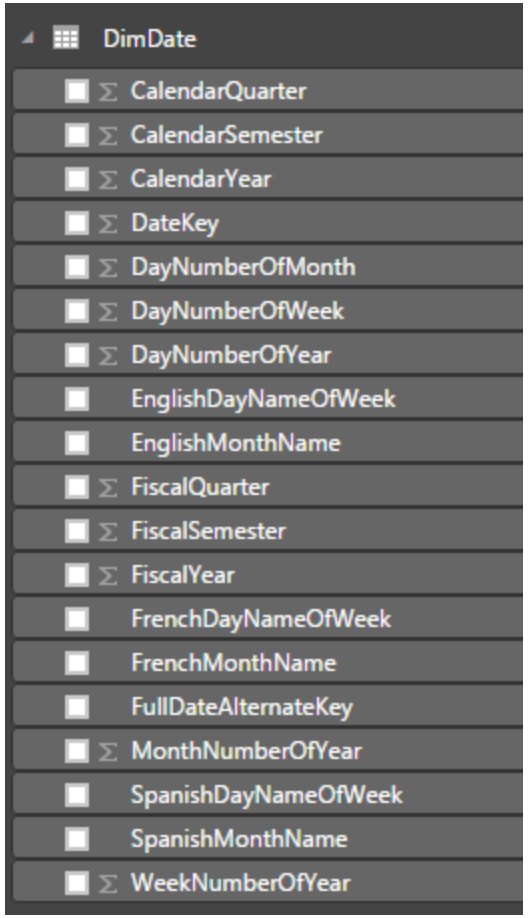
	Year	Month	Day	FullDateAl...	DateKey	DateFullName	Fiscal Year	Fiscal Quarter	Calendar Quar...	IsWeekDay	DayOfWeek	Month Name	Day of Week Name
1	2013	6	15	6/15/2013	20130615	15 June 2013	2013	4	2	0	6	June	Saturday
2	2013	6	16	6/16/2013	20130616	16 June 2013	2013	4	2	0	0	June	Sunday
3	2013	6	17	6/17/2013	20130617	17 June 2013	2013	4	2	1	1	June	Monday
4	2013	6	18	6/18/2013	20130618	18 June 2013	2013	4	2	1	2	June	Tuesday
5	2013	6	19	6/19/2013	20130619	19 June 2013	2013	4	2	1	3	June	Wednesday
6	2013	6	20	6/20/2013	20130620	20 June 2013	2013	4	2	1	4	June	Thursday
7	2013	6	21	6/21/2013	20130621	21 June 2013	2013	4	2	1	5	June	Friday
8	2013	6	22	6/22/2013	20130622	22 June 2013	2013	4	2	0	6	June	Saturday
9	2013	6	23	6/23/2013	20130623	23 June 2013	2013	4	2	0	0	June	Sunday
10	2013	6	24	6/24/2013	20130624	24 June 2013	2013	4	2	1	1	June	Monday
11	2013	6	25	6/25/2013	20130625	25 June 2013	2013	4	2	1	2	June	Tuesday
12	2013	6	26	6/26/2013	20130626	26 June 2013	2013	4	2	1	3	June	Wednesday
13	2013	6	27	6/27/2013	20130627	27 June 2013	2013	4	2	1	4	June	Thursday
14	2013	6	28	6/28/2013	20130628	28 June 2013	2013	4	2	1	5	June	Friday
15	2013	6	29	6/29/2013	20130629	29 June 2013	2013	4	2	0	6	June	Saturday
16	2013	6	30	6/30/2013	20130630	30 June 2013	2013	4	2	0	0	June	Sunday
17	2013	7	1	7/1/2013	20130701	01 July 2013	2014	1	3	1	1	July	Monday
18	2013	7	2	7/2/2013	20130702	02 July 2013	2014	1	3	1	2	July	Tuesday
19	2013	7	3	7/3/2013	20130703	03 July 2013	2014	1	3	1	3	July	Wednesday
20	2013	7	4	7/4/2013	20130704	04 July 2013	2014	1	3	1	4	July	Thursday
21	2013	7	5	7/5/2013	20130705	05 July 2013	2014	1	3	1	5	July	Friday
22	2013	7	6	7/6/2013	20130706	06 July 2013	2014	1	3	0	6	July	Saturday

Why date dimension?

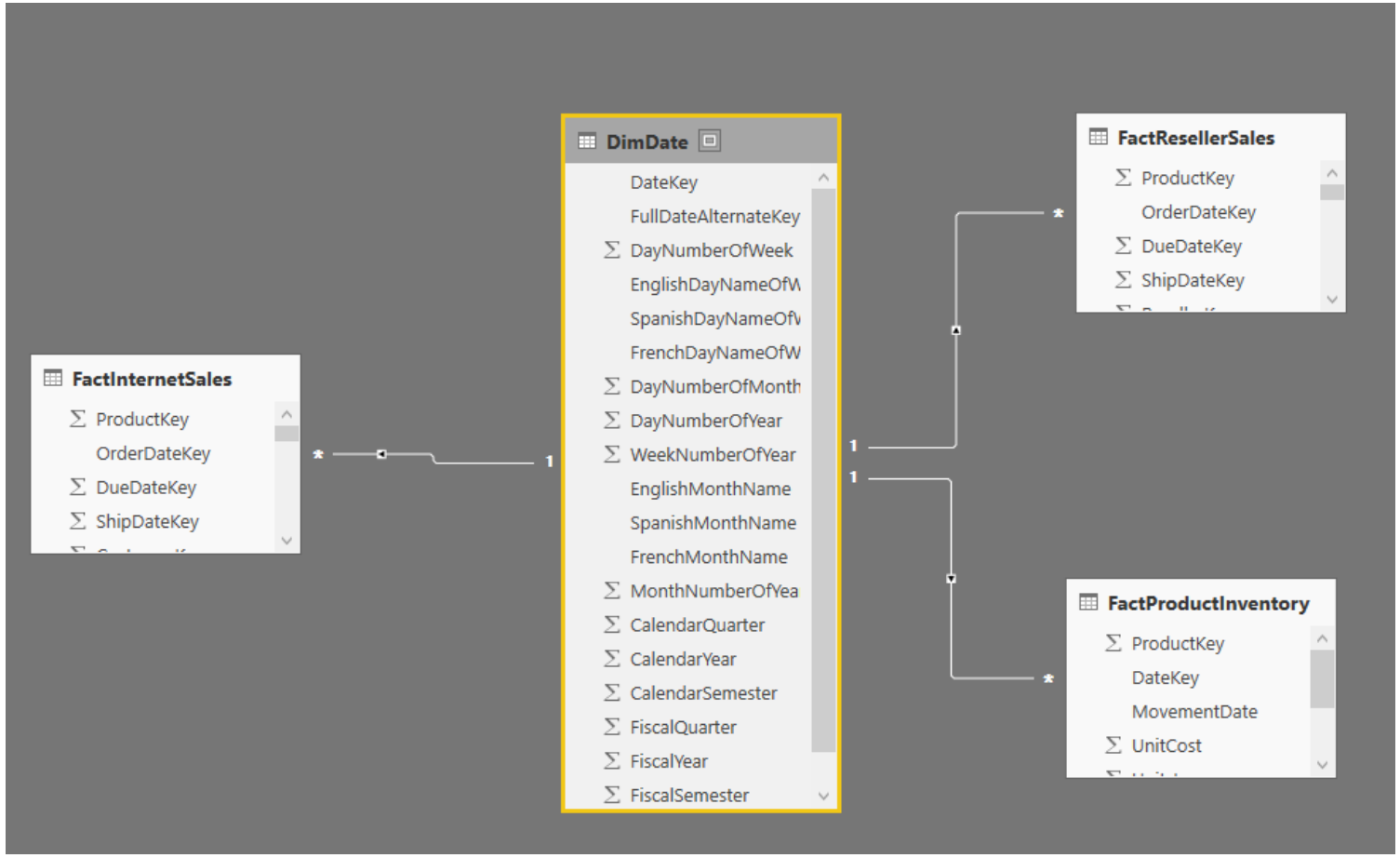
- It gives you the ability to slice and dice values by many date attributes such as week number, half-year, day of the year, and etc.
- It makes the analysis consistent.
- It gives you the ability to do analysis based on public holidays (Easter Monday, Good Friday, and etc.)
- There are some BI tools extended functions that need to work with a Date Dimension.

Powerful slicing and dicing

- Weekday analysis of revenue (by Monday, Tuesday, …., Sunday)
- Which week in the year generates the least revenue or most revenue?
- What about Fiscal? What is the revenue for the Fiscal year?
- And many other questions.



Consistency in analysis

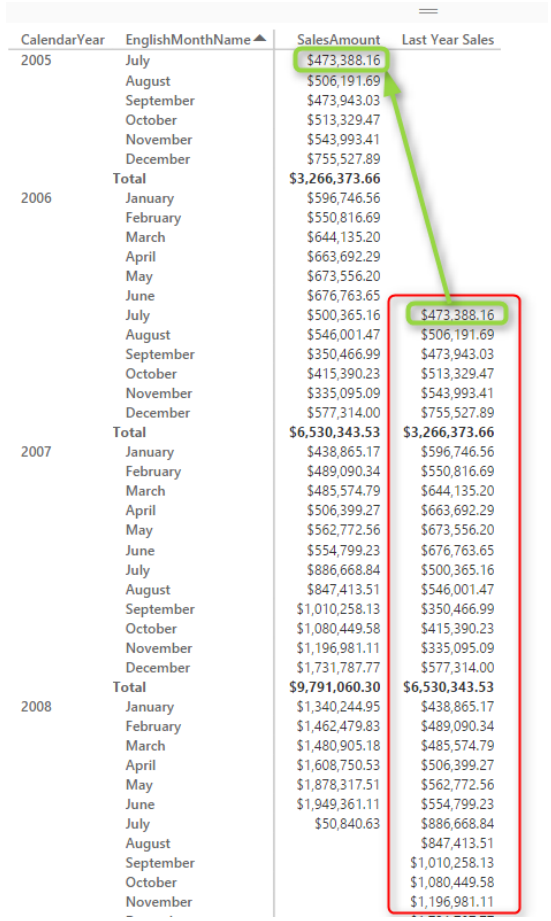


Public holidays insights

- Was the sales at Easter this year better than last year? (you don't want to calculate Easter in Power BI, of course)
- How is the revenue is public holidays compared with weekends?
- How were the sales for the store on opening day? (special event)
- and many other questions.

	Holiday	Date
47	Westland Anniversary	11/28/2016
48	Chatham Islands Anniversary	11/28/2016
49	Christmas Day	12/27/2016
50	Boxing Day	12/26/2016
51	New Year's Day	1/3/2017
52	Day after New Year's Day	1/2/2017
53	Wellington Anniversary	1/23/2017
54	Auckland Anniversary	1/30/2017
55	Nelson Anniversary	1/30/2017
56	Waitangi Day	2/6/2017
57	Taranaki Anniversary	3/13/2017
58	Otago Anniversary	3/20/2017
59	Daylight Saving ends	4/2/2017
60	Good Friday	4/14/2017
61	Easter Monday	4/17/2017
62	Easter Tuesday ?	4/18/2017

Extended Functions for Time Intelligence



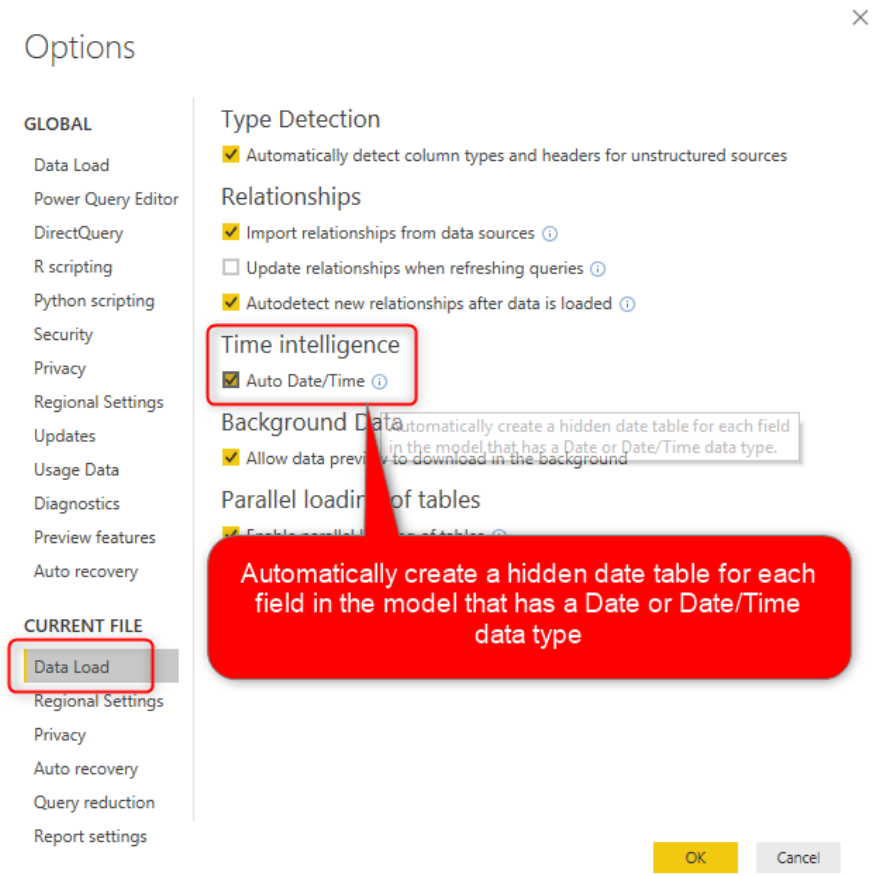
CalendarYear	EnglishMonthName	SalesAmount	Last Year Sales
2005	July	$473,388.16	
	August	$506,191.69	
	September	$473,943.03	
	October	$513,329.47	
	November	$543,993.41	
	December	$755,527.89	
	Total	**$3,266,373.66**	
2006	January	$596,746.56	
	February	$550,816.69	
	March	$644,135.20	
	April	$663,692.29	
	May	$673,556.20	
	June	$676,763.65	
	July	$500,365.16	$473,388.16
	August	$546,001.47	$506,191.69
	September	$350,466.99	$473,943.03
	October	$415,390.23	$513,329.47
	November	$335,095.09	$543,993.41
	December	$577,314.00	$755,527.89
	Total	**$6,530,343.53**	**$3,266,373.66**
2007	January	$438,865.17	$596,746.56
	February	$489,090.34	$550,816.69
	March	$485,574.79	$644,135.20
	April	$506,399.27	$663,692.29
	May	$562,772.56	$673,556.20
	June	$554,799.23	$676,763.65
	July	$886,668.84	$500,365.16
	August	$847,413.51	$546,001.47
	September	$1,010,258.13	$350,466.99
	October	$1,080,449.58	$415,390.23
	November	$1,196,981.11	$335,095.09
	December	$1,731,787.77	$577,314.00
	Total	**$9,791,060.30**	**$6,530,343.53**
2008	January	$1,340,244.95	$438,865.17
	February	$1,462,479.83	$489,090.34
	March	$1,480,905.18	$485,574.79
	April	$1,608,750.53	$506,399.27
	May	$1,878,317.51	$562,772.56
	June	$1,949,361.11	$554,799.23
	July	$50,840.63	$886,668.84
	August		$847,413.51
	September		$1,010,258.13
	October		$1,080,449.58
	November		$1,196,981.11

Ways to have a date dimension

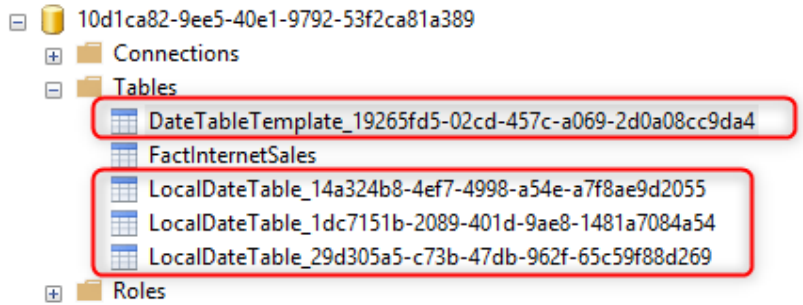
- Built-in
  - By default provided in Power BI
- Custom
  - Build using T-SQL (database script)
  - Build using Power Query
  - Build using DAX expression

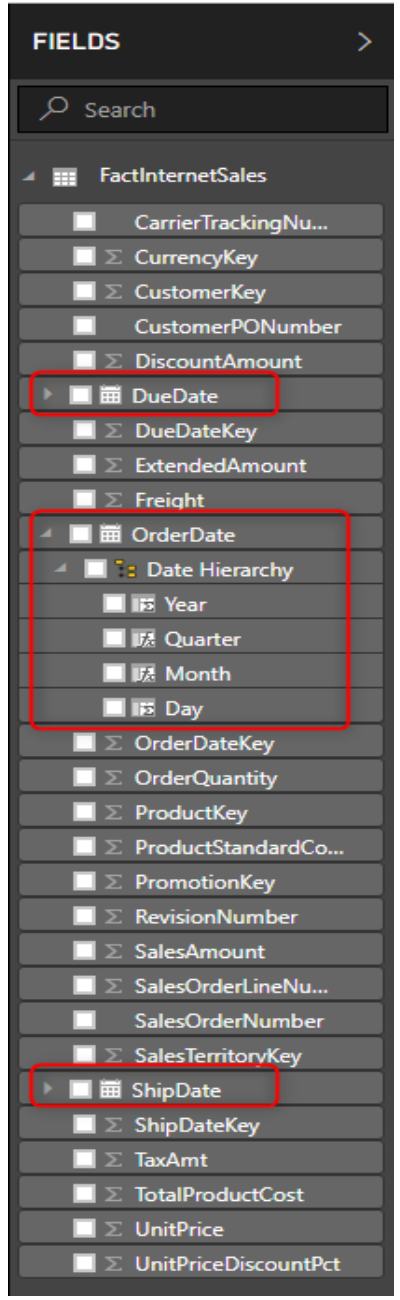
Power BI Date dimension; Default or Custom?

Default date table



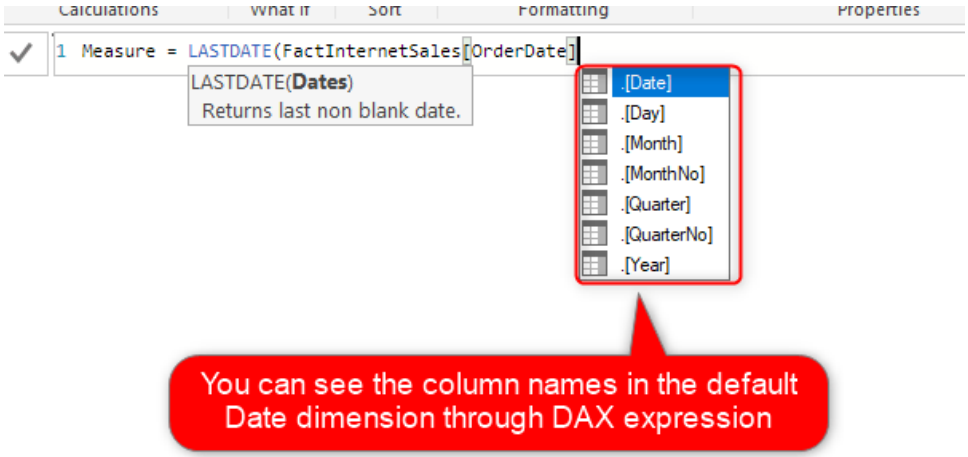
Every date field is a table





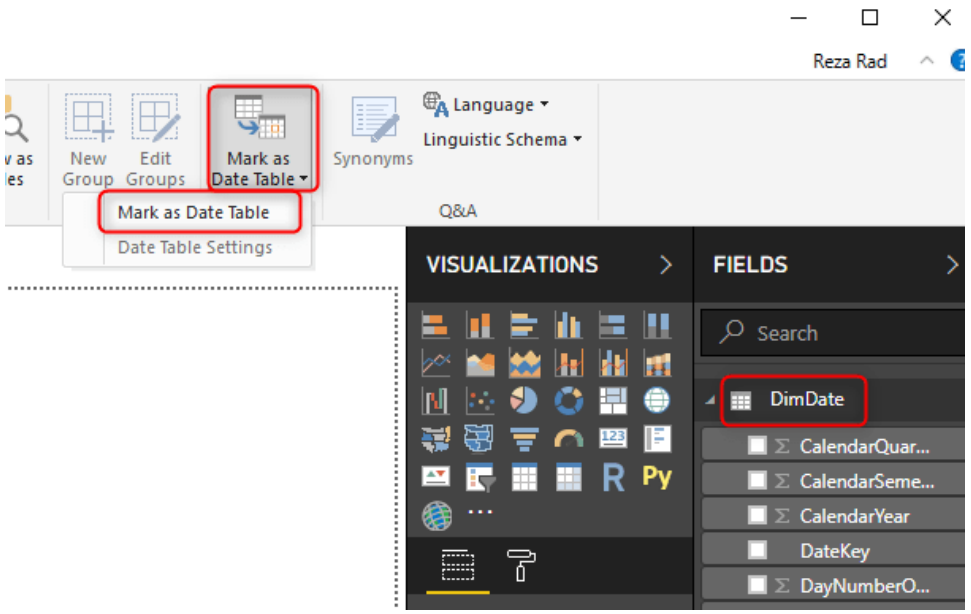
When you write DAX expressions

- *Sales YTD = TOTALYTD(*
- *SUM(FactInternetSales[SalesAmount]),*
- *FactInternetSales[OrderDate].[Date])*



Custom Date dimension

- Mark as date table



DAX expression

```
1  Sales YTD = TOTALYTD(
2      SUM(FactInternetSales[SalesAmount]),
3      DimDate[FullDateAlternateKey])
```

FullDateAlternateKey	SalesAmount	Sales YTD
Friday, 1 July 2005	14,477.34	14,477.34
Saturday, 2 July 2005	13,931.52	28,408.86
Sunday, 3 July 2005	15,012.18	43,421.04
Monday, 4 July 2005	7,156.54	50,577.58
Tuesday, 5 July 2005	15,012.18	65,589.75
Wednesday, 6 July 2005	14,313.08	79,902.83
Thursday, 7 July 2005	7,855.64	87,758.47
Friday, 8 July 2005	7,855.64	95,614.11
Saturday, 9 July 2005	20,909.78	116,523.89
Sunday, 10 July 2005	10,556.53	127,080.42
Monday, 11 July 2005	14,313.08	141,393.50
Tuesday, 12 July 2005	14,134.80	155,528.30
Wednesday, 13 July 2005	7,156.54	162,684.84
Thursday, 14 July 2005	25,047.89	187,732.73

Default Vs Custom?

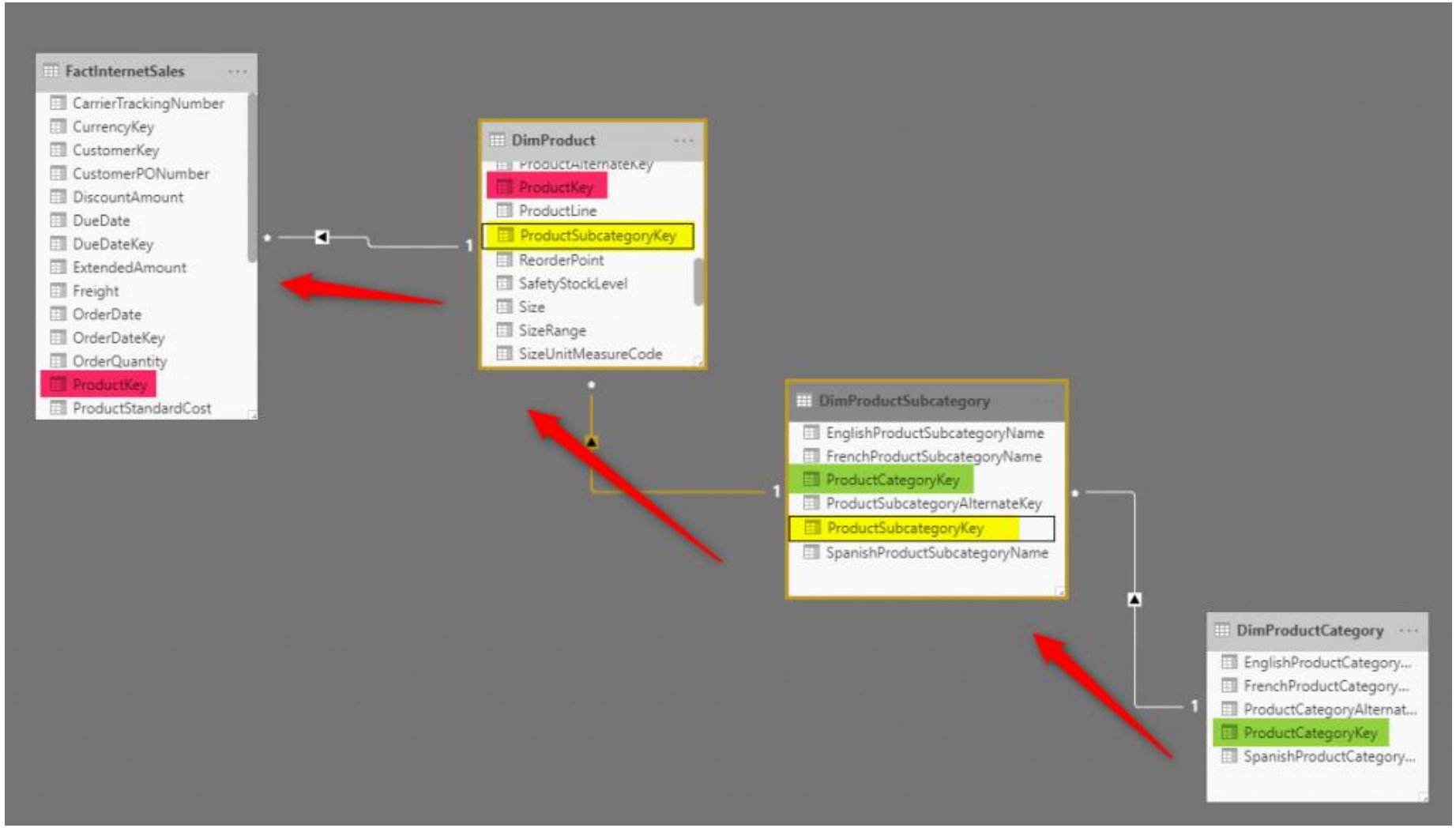
- Modeling can start easier with the default
  - No need to create a custom table
  - Every date field has a date table
  - Default hierarchies
- Advanced analysis is only possible with Custom
  - Public holidays, or specific working days analysis
  - A single date table to filter multiple tables
  - Better memory management
  - Better modeling

Script to build a Date dimension

Script to build a Time dimension

Star Schema and how to build it

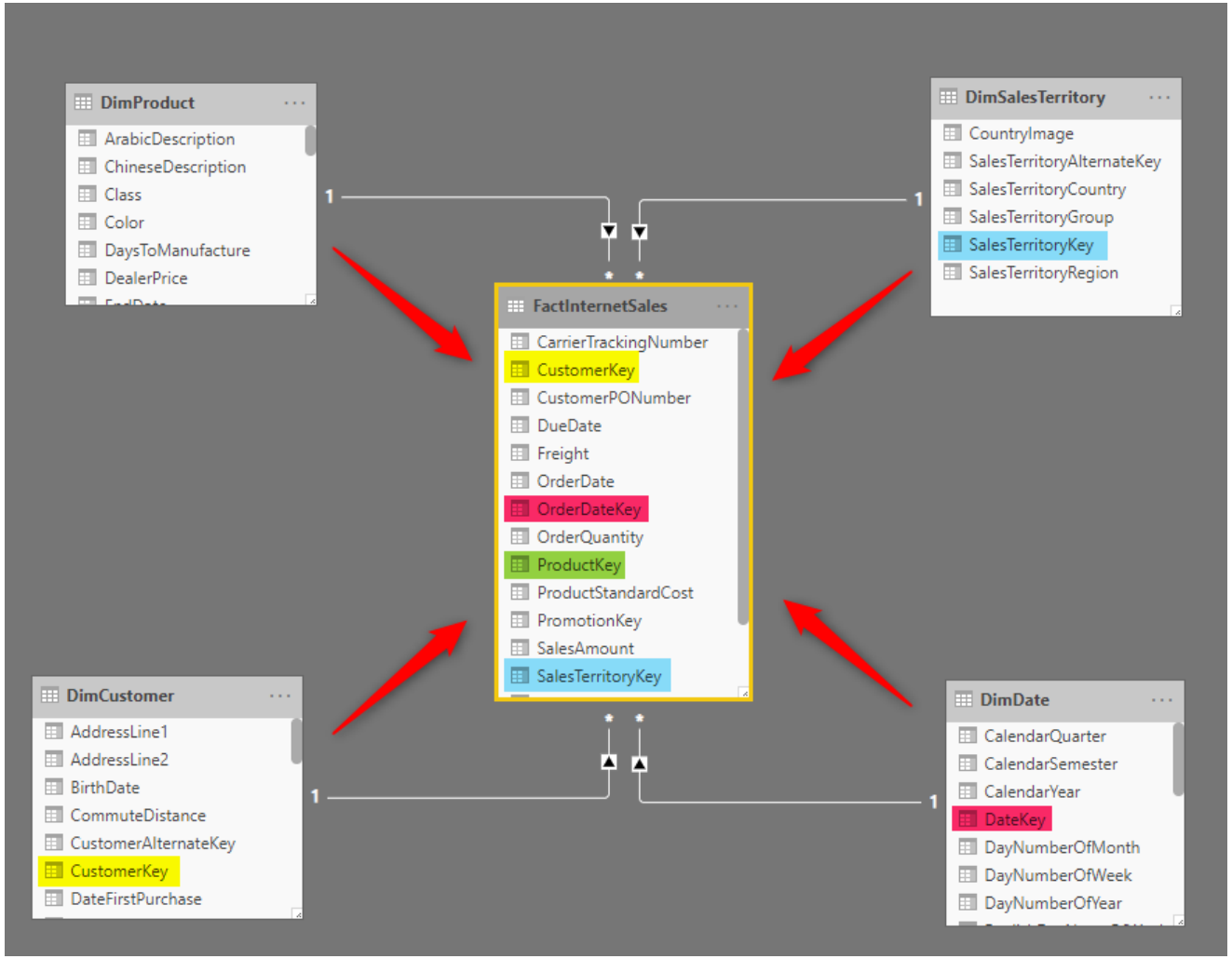
No dimension to dimension connection



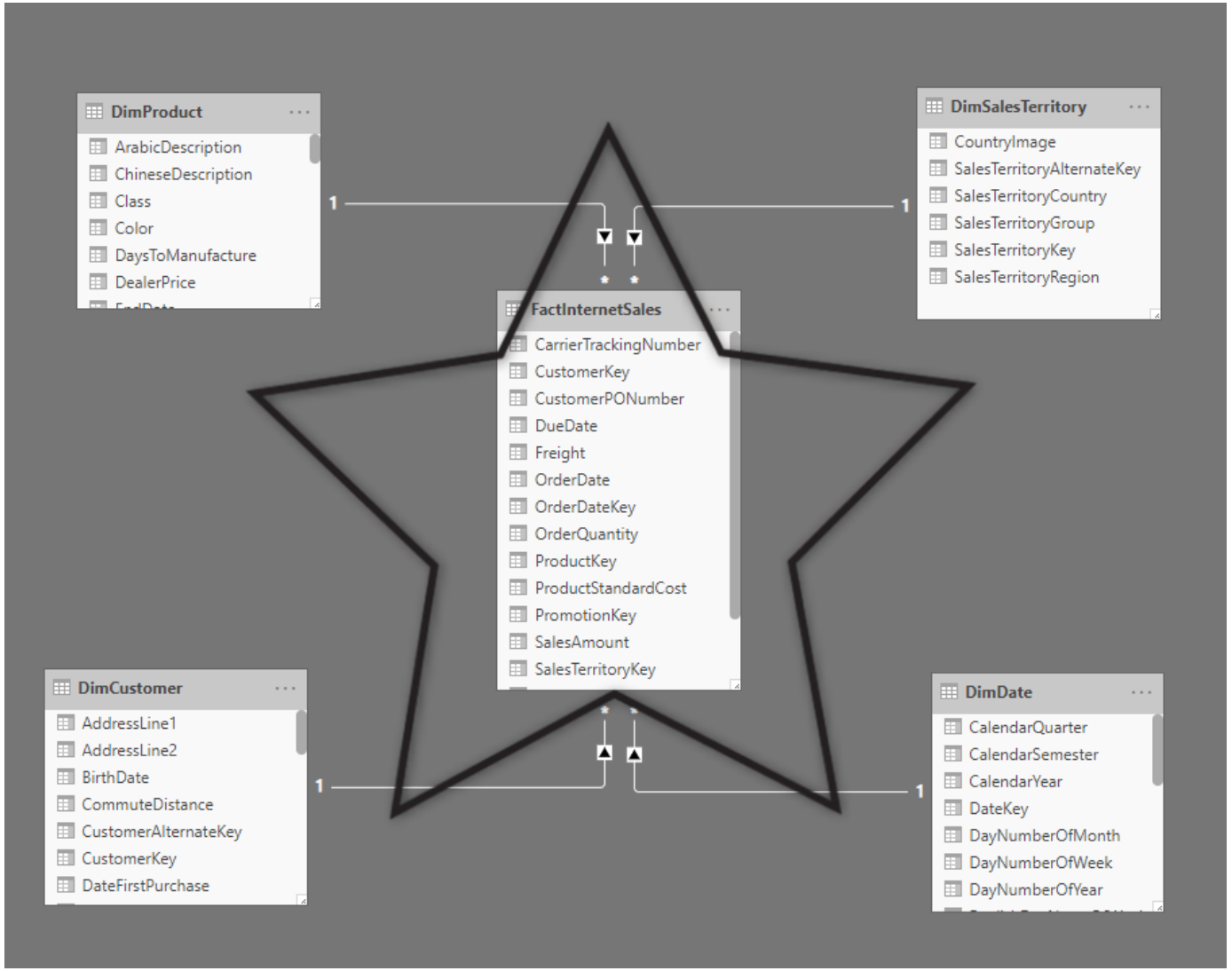
NO to Snowflake Schema

Relationship between Fact and Dimension tables

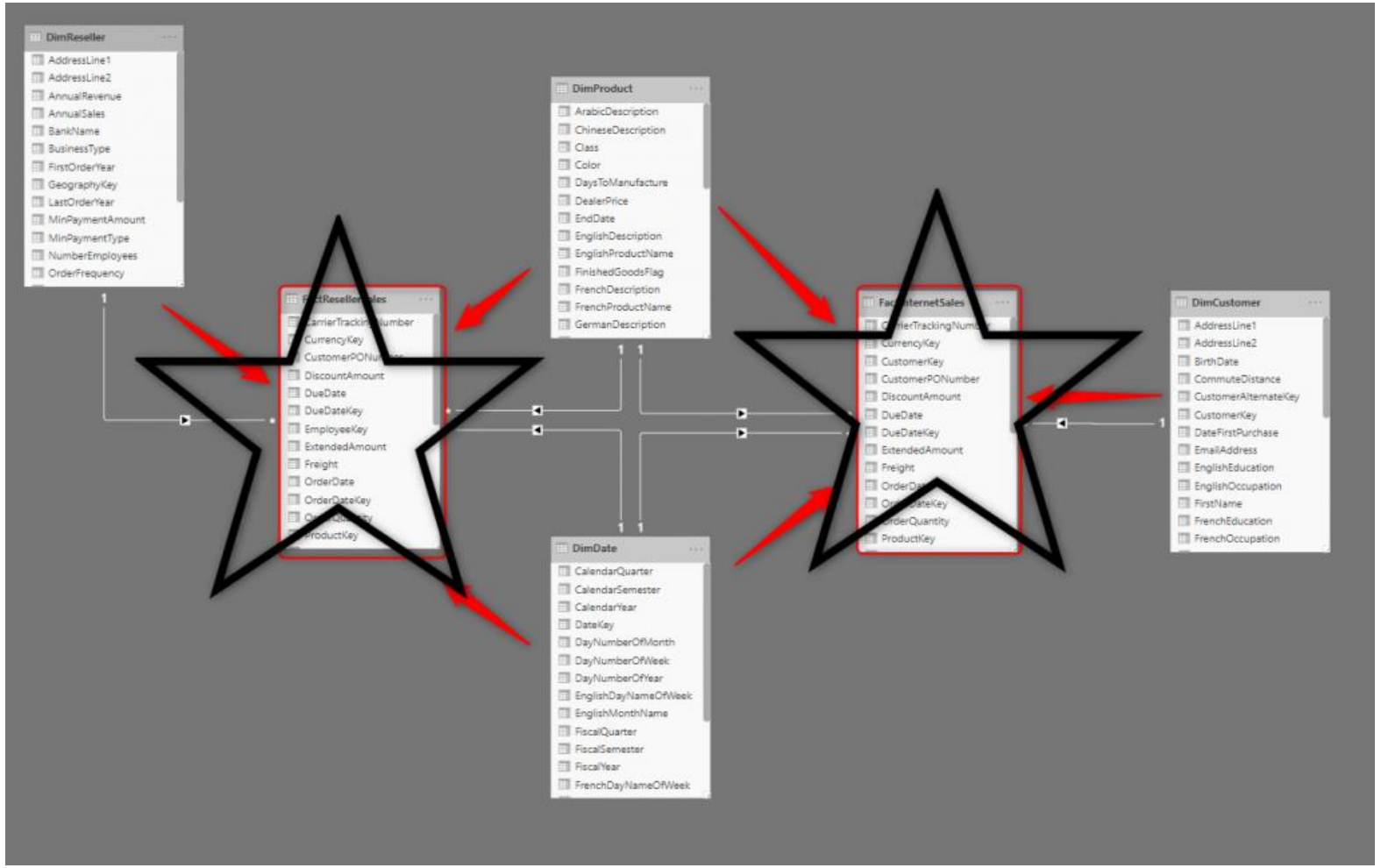
- One to many
- From dimension to fact
- Single directional
- Star Schema



Star Schema


DimProduct
ArabicDescription
ChineseDescription
Class
Color
DaysToManufacture
DealerPrice
DimSalesTerritory
CountryImage
SalesTerritoryAlternateKey
SalesTerritoryCountry
SalesTerritoryGroup
SalesTerritoryKey
SalesTerritoryRegion
FactInternetSales
CarrierTrackingNumber
CustomerKey
CustomerPONumber
DueDate
Freight
OrderDate
OrderDateKey
OrderQuantity
ProductKey
ProductStandardCost
PromotionKey
SalesAmount
SalesTerritoryKey
DimCustomer
AddressLine1
AddressLine2
BirthDate
CommuteDistance
CustomerAlternateKey
CustomerKey
DateFirstPurchase
DimDate
CalendarQuarter
CalendarSemester
CalendarYear
DateKey
DayNumberOfMonth
DayNumberOfWeek
DayNumberOfYear
1
1
1
1

Star Schema is not just one fact table!



Star Schema for a Data Model is like a Conductor for an Orchestra. The best analytics outcome would be possible with the star schema, like the best music outcome with a good conductor.

Module 3:
Star schema in action

Combining dimension tables using Power Query

Creating a shared dimension using Power Query

What fields to hide?

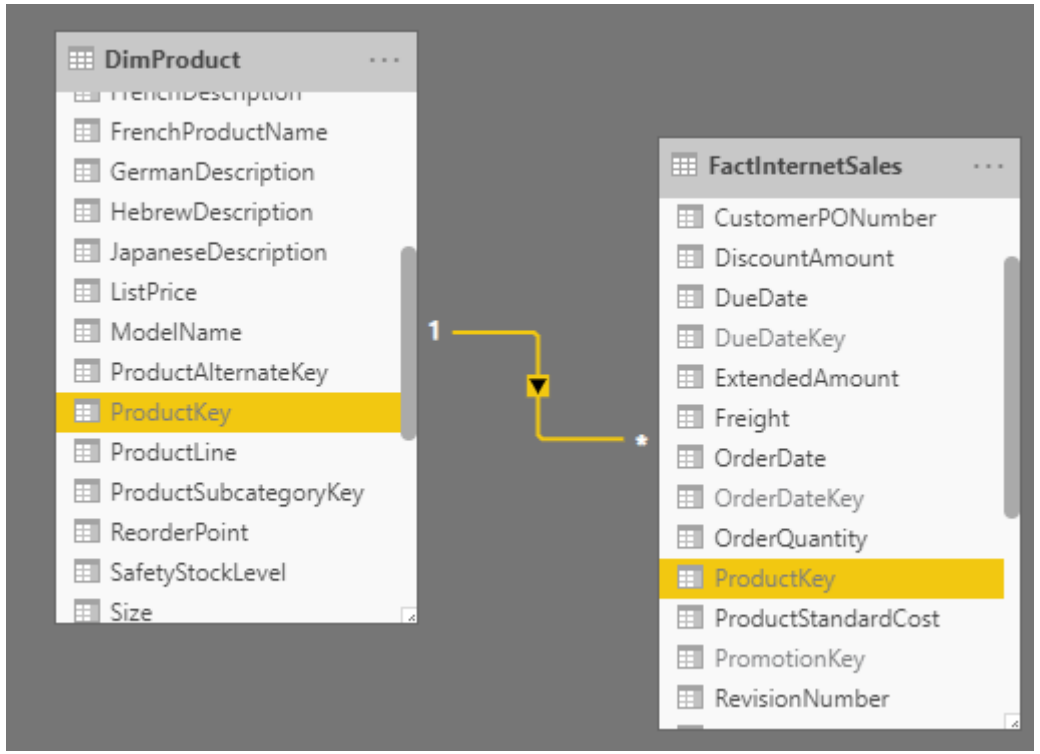
Consider REMOVE before hiding

- Unnecessary fields has to be REMOVED, otherwise they are consuming memory

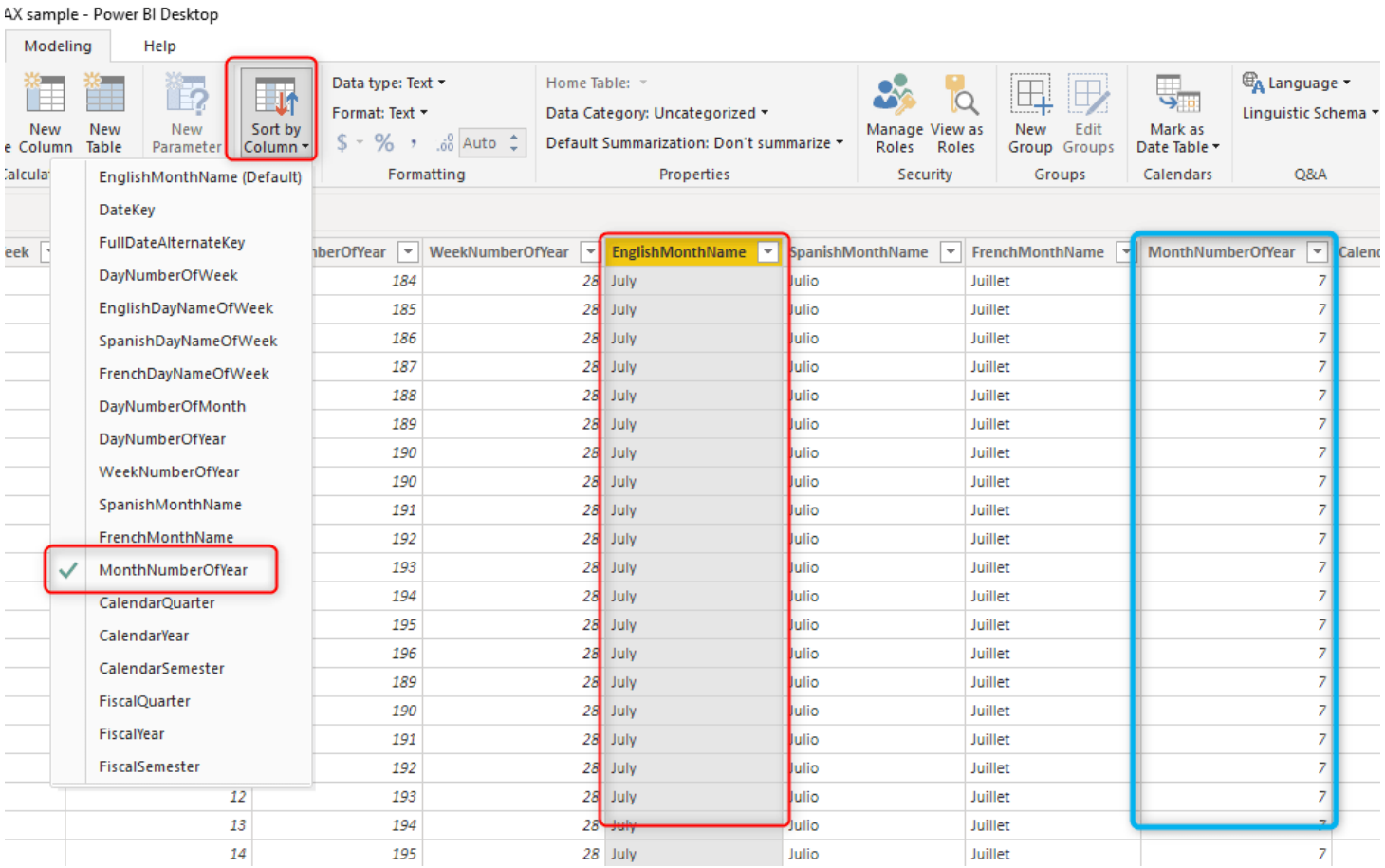
Hide these fields

- Fields used in a Relationship
- Fields used to Sort Other Fields
- Fields used in Hierarchy
- Fields used in DAX Calculations, but not in visuals directly

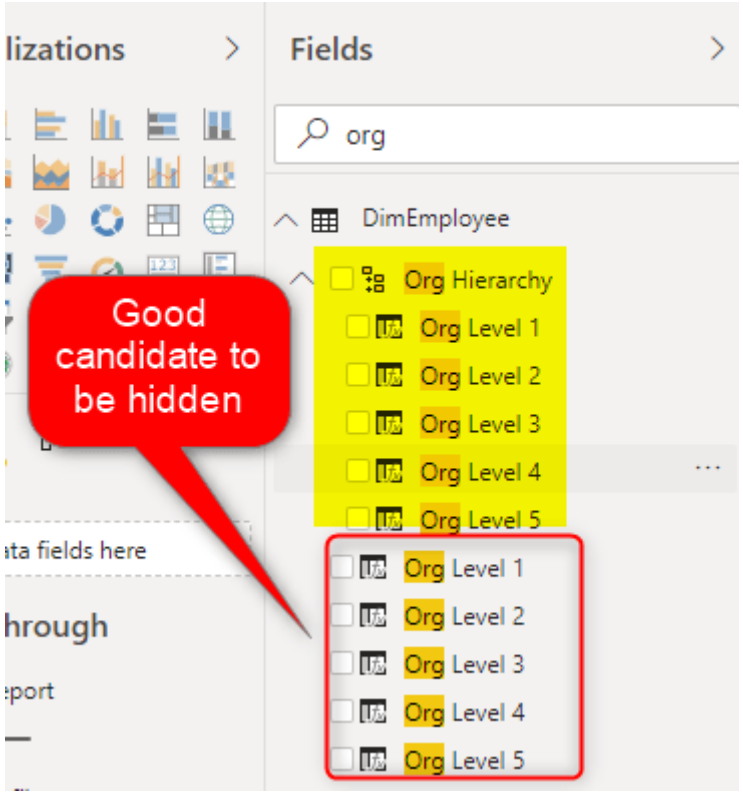
Fields used in a relationship



Fields used to sort other fields



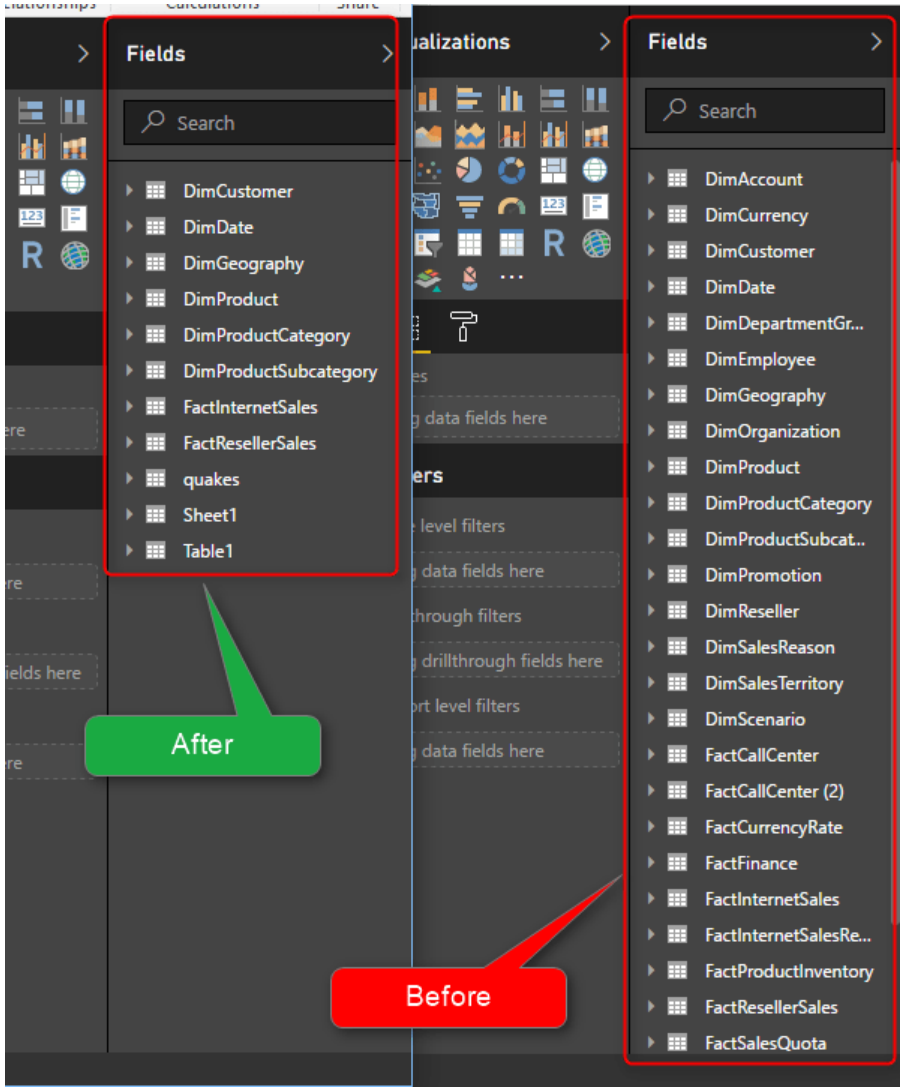
Fields used in a hierarchy



Fields used in calculations, but not visuals directly

Power BI Helper

- https://radacad.com/power-bi-helper
- Free tool to find out fields to be hidden



Build your first star schema in action

How to use time and date dimensions in a Power BI model?

Multiple fact tables: Budget Vs. Actual

Fact tables with different granularity

One dimension filters another dimension

Module 4: Calculations

Calculations in Power Query [M] or DAX?

Power Query: M

- Data transformation language

```
let
    FirstAndLastDayOfTheMonth = (date) =>
        let
            dated=Date.FromText(date),
            year=Date.Year(dated),
            month=Date.Month(dated),
            FirstDateText=Text.From(year)&"-"&Text.From(month)&"-01",
            FirstDate=Date.FromText(FirstDateText),
            daysInMonth=Date.DaysInMonth(dated),
            LastDateText=Text.From(year)&"-"&Text.From(month)&"-"&Text.From(daysInMonth),
            LastDate=Date.FromText(LastDateText),
            record=Record.AddField([],"First Date of Month",FirstDate),
            resultset=Record.AddField(record,"Last Date of Month",LastDate)
        in
            resultset
in
    FirstAndLastDayOfTheMonth("30/07/2015")
```

DAX

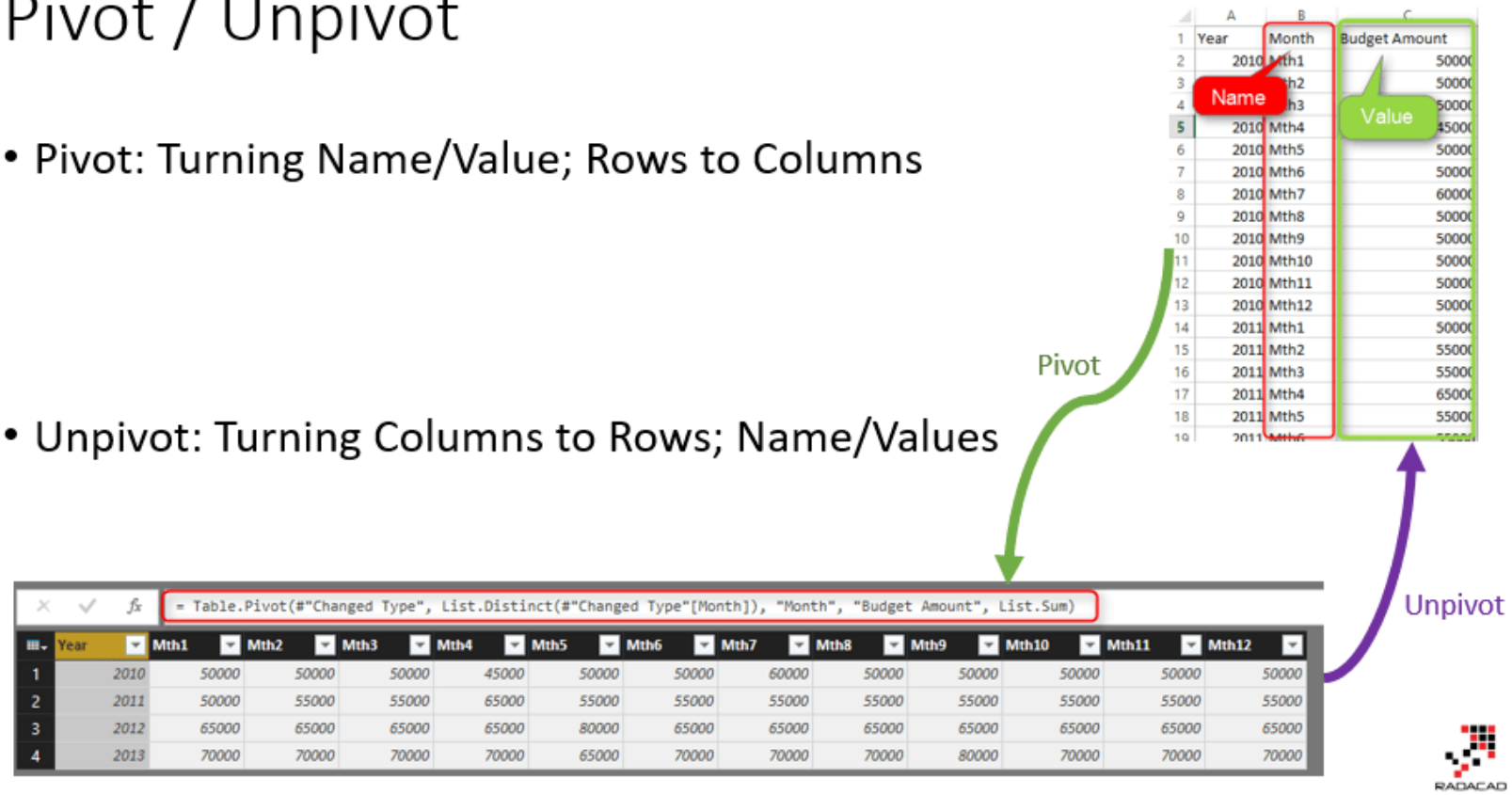
- Data Analysis eXpression language

```
Sales Rolling 12 Months =
CALCULATE(
    SUM(FactInternetSales[SalesAmount]),
    DATESBETWEEN
           (DimDate[FullDateAlternateKey],
               NEXTDAY(SAMEPERIODLASTYEAR(LASTDATE(DimDate[FullDateAlternateKey]))),
               LASTDATE(DimDate[FullDateAlternateKey])
            ),
    ALL(DimDate)
    )
```

Usages of Power Query

Pivot / Unpivot

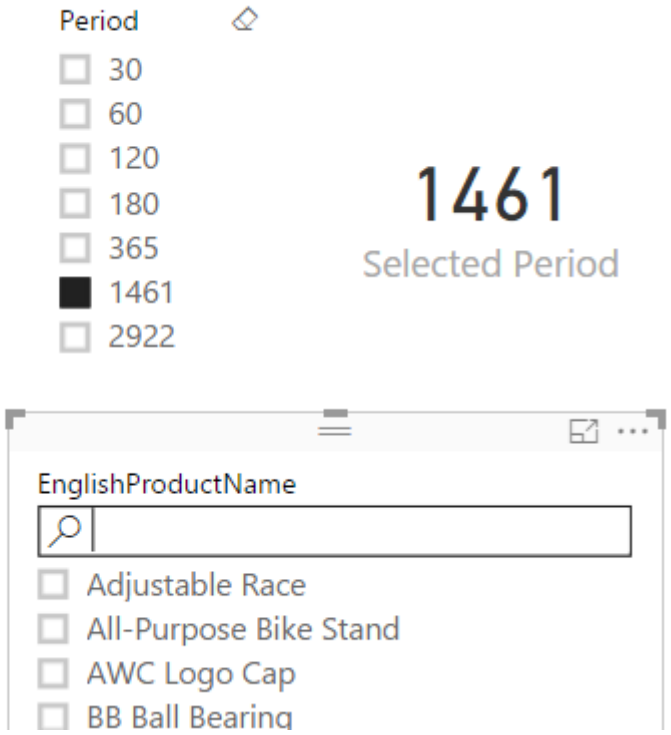
- Pivot: Turning Name/Value; Rows to Columns
- Unpivot: Turning Columns to Rows; Name/Values



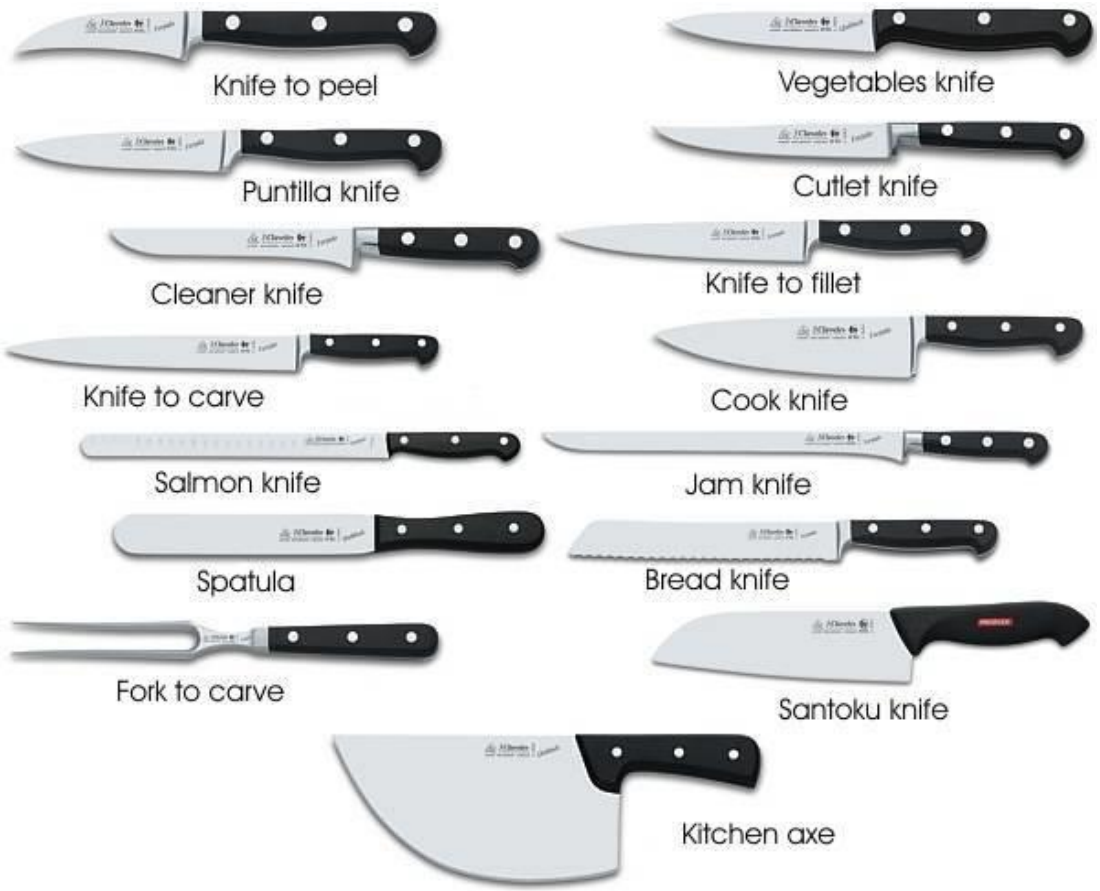
	Year	Mth1	Mth2	Mth3	Mth4	Mth5	Mth6	Mth7	Mth8	Mth9	Mth10	Mth11	Mth12
1	2010	50000	50000	50000	45000	50000	50000	60000	50000	50000	50000	50000	50000
2	2011	50000	55000	55000	65000	55000	55000	55000	55000	55000	55000	55000	55000
3	2012	65000	65000	65000	65000	80000	65000	65000	65000	65000	65000	65000	65000
4	2013	70000	70000	70000	70000	65000	70000	70000	70000	80000	70000	70000	70000

Usages of DAX

FullName	Total Revenue	Last Period Revenue	Lost Customers	New Customers
Aaron Adams	$118	$118	0	1
Aaron Alexander	$70	$70	0	1
Aaron Allen	$3,400		1	0
Aaron Baker	$1,751	$1,751	0	1
Aaron Bryant	$134	$134	0	1
Aaron Butler	$15	$15	0	1
Aaron Campbell	$1,155	$1,155	0	1
Aaron Carter	$40	$40	0	1
Aaron Chen	$40	$40	0	1
Aaron Coleman	$62	$62	0	1
Aaron Collins	$6,047	$2,469	0	0
Aaron Diaz	$6,030	$2,451	0	0
Aaron Edwards	$94	$94	0	1
Aaron Evans	$2,433	$2,433	0	1



Calculated column



reference: http://forkitchen.blogspot.co.nz/2008/10/what-are-different-types-of-kitchen.html

Calculated Table

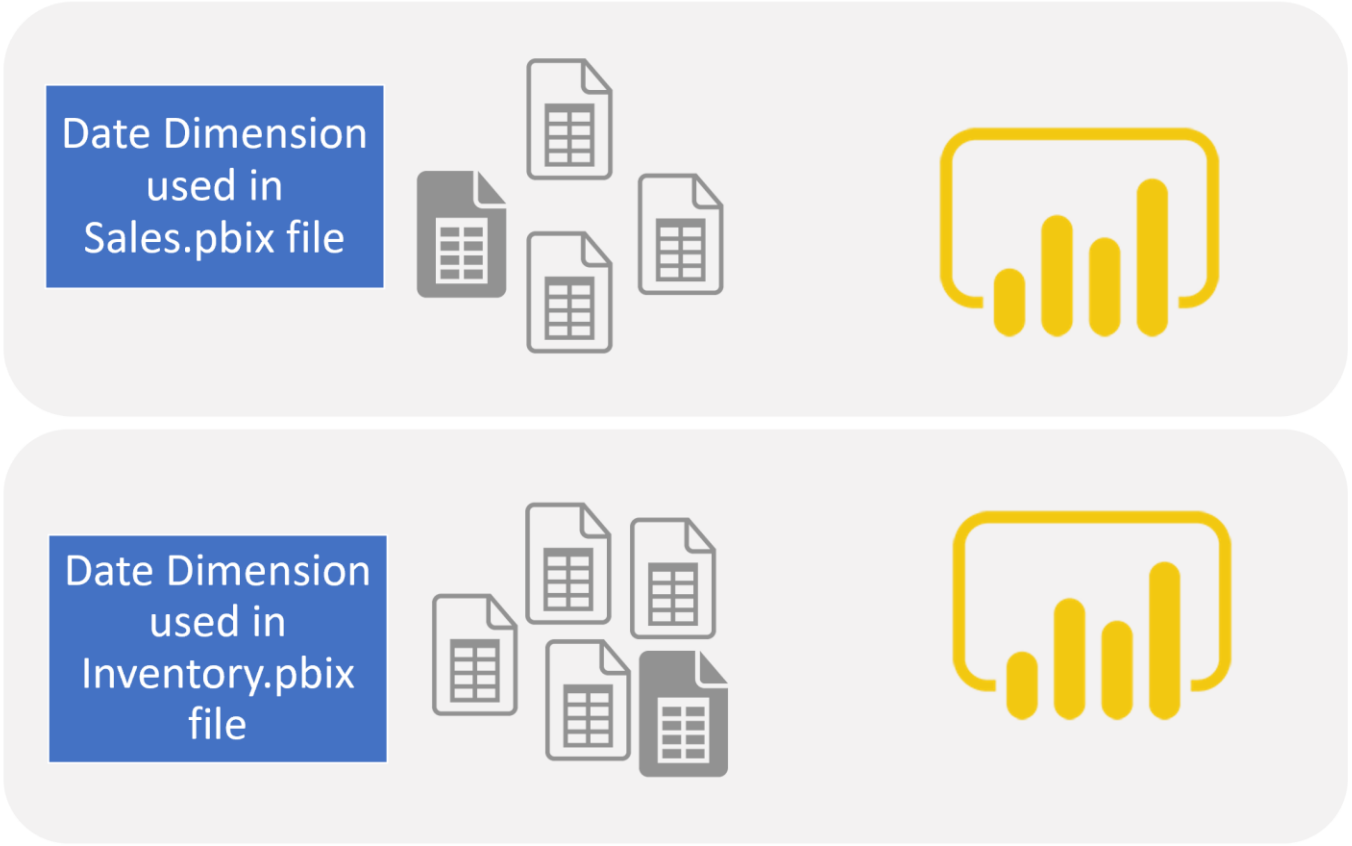
Role-Playing Dimensions

DAX Measure Vs. Calculated Column?

Module 5:
Better data model

Move shared tables to dataflows

Using One Power Query Table in Multiple Power BI Reports



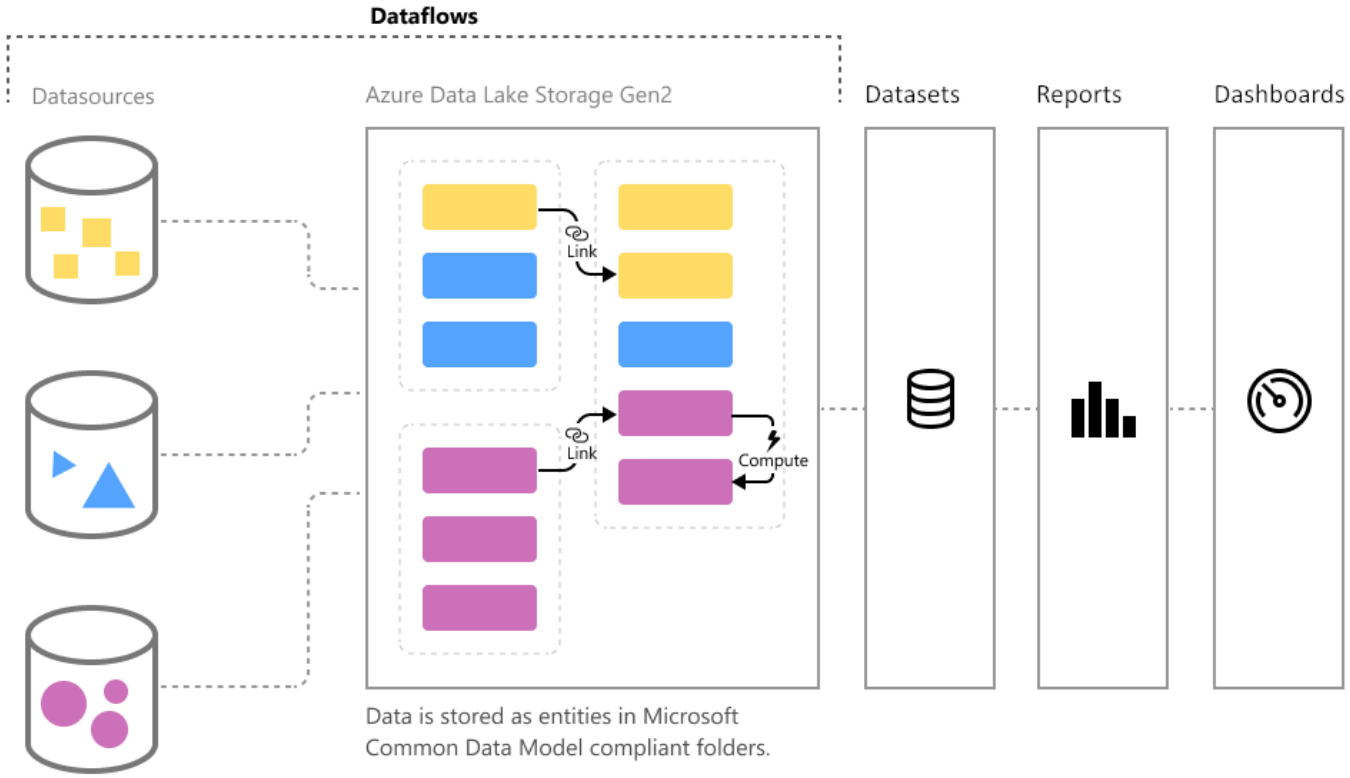
Date Dimension transformation executed multiple times, when only once is needed

The Solution

Dataflow for Transformation

What is Dataflow?

- Dataflow is a **Power Query** process that runs in the **cloud** **independently** from any Power BI reports.



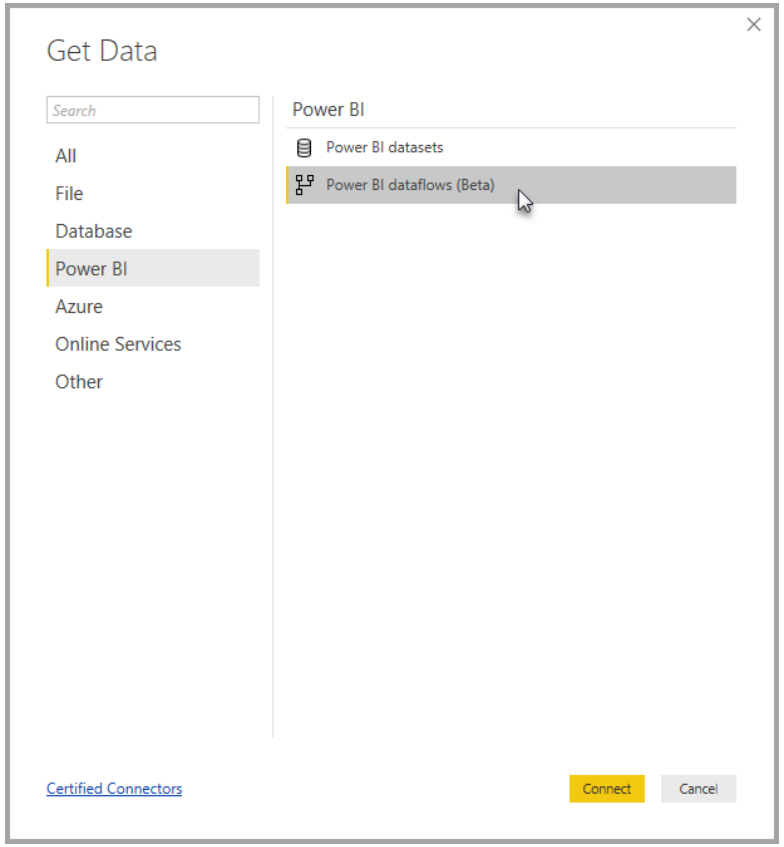
Where the Output Stored?

- Dataflow stores the data in the Azure Data Lake storage.

But I Don’t Have Azure Data Lake Subscription!

- Dataflow manages the Data Lake configurations internally. You won’t need anything except your Power BI accounts and subscriptions.

Power BI Can Do Get Data from Dataflow

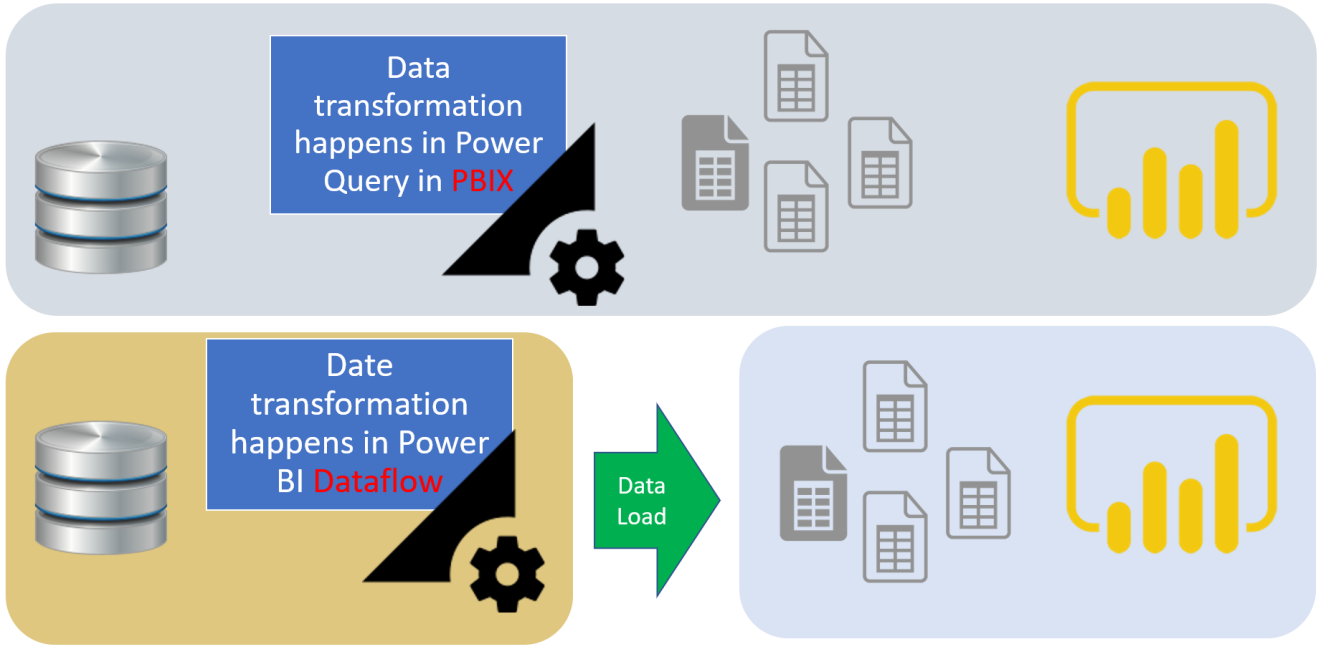


Using One Power Query Table in Multiple Power BI Reports

- Re-usable tables or queries across multiple Power BI files, are one of the best candidates for Dataflow.

Date Dimension transformation executed only once, and then used multiple times

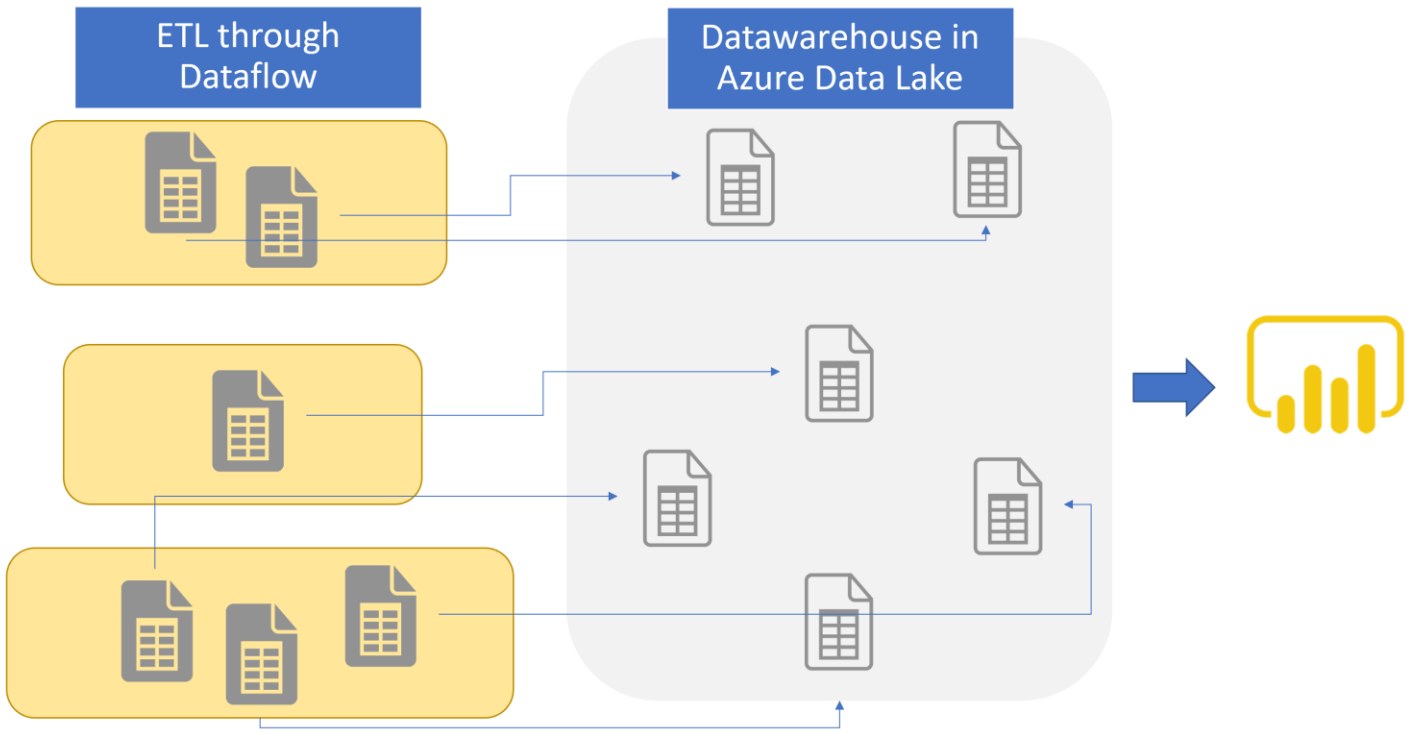
FASTER refresh time



Push heavy transformations to Power BI dataflow

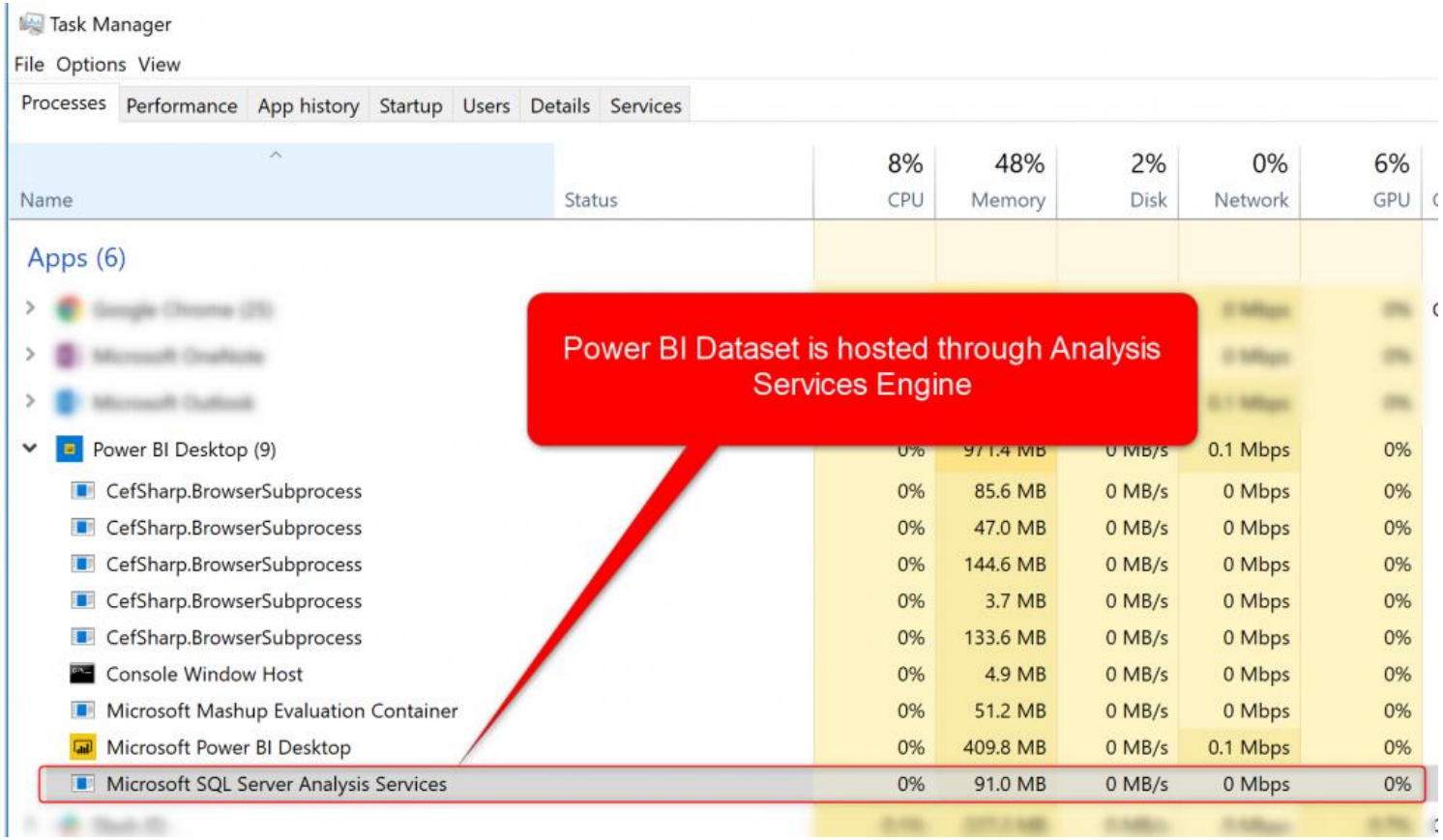
Centralized Data Warehouse

- Dataflow can be the ETL engine, that fuels the centralized data warehouse in Azure data lake storage.



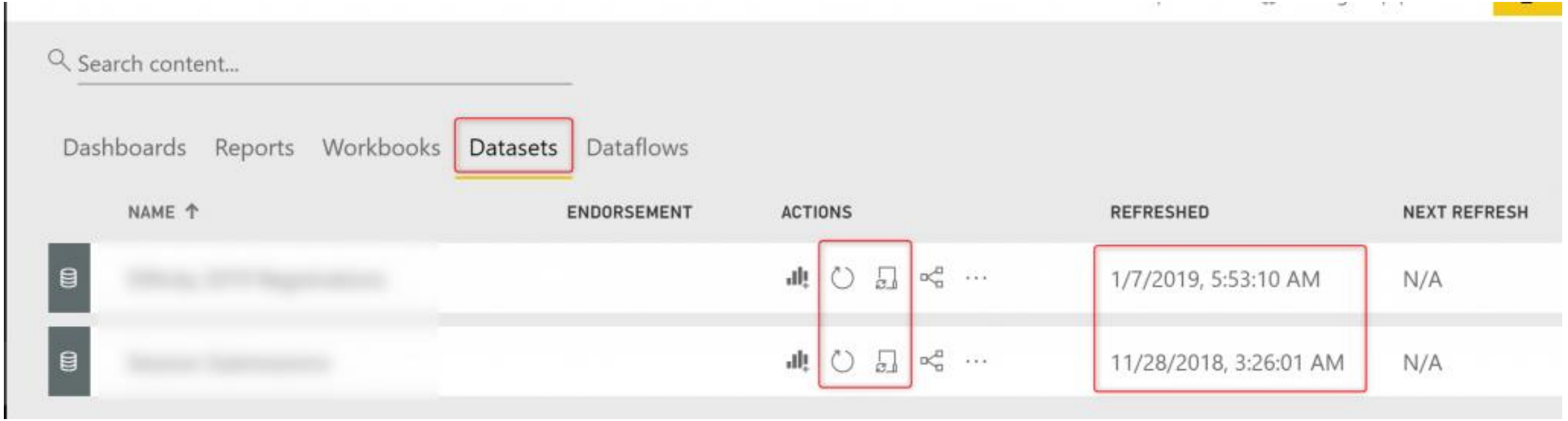
Dataflow Fuels the Azure Data Lake as the ETL fuels the Data Warehouse

Shared dataset



- The **report** is the visualization layer of your Power BI implementation
- The **dataset** includes the data, tables, relationships, calculations, and connection to the data source.

What is the Dataset?



Dataset in the Service

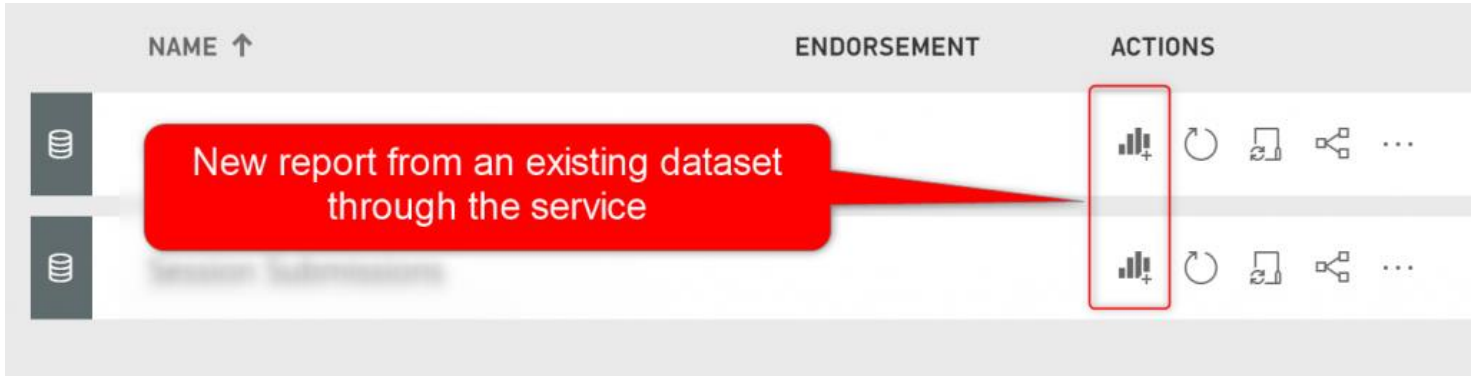
Multiple report visualizers using the same calculation (DAX)?

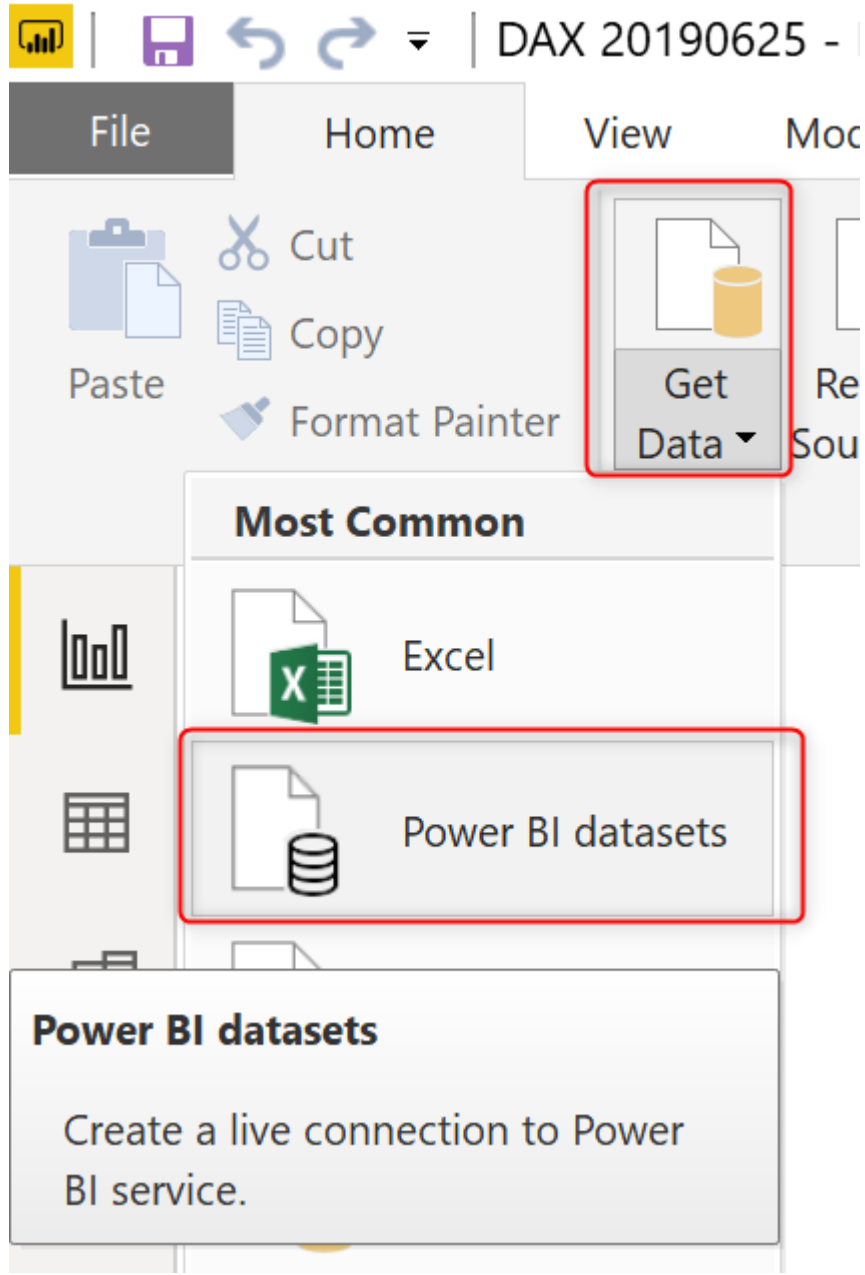
The Solution

Shared dataset

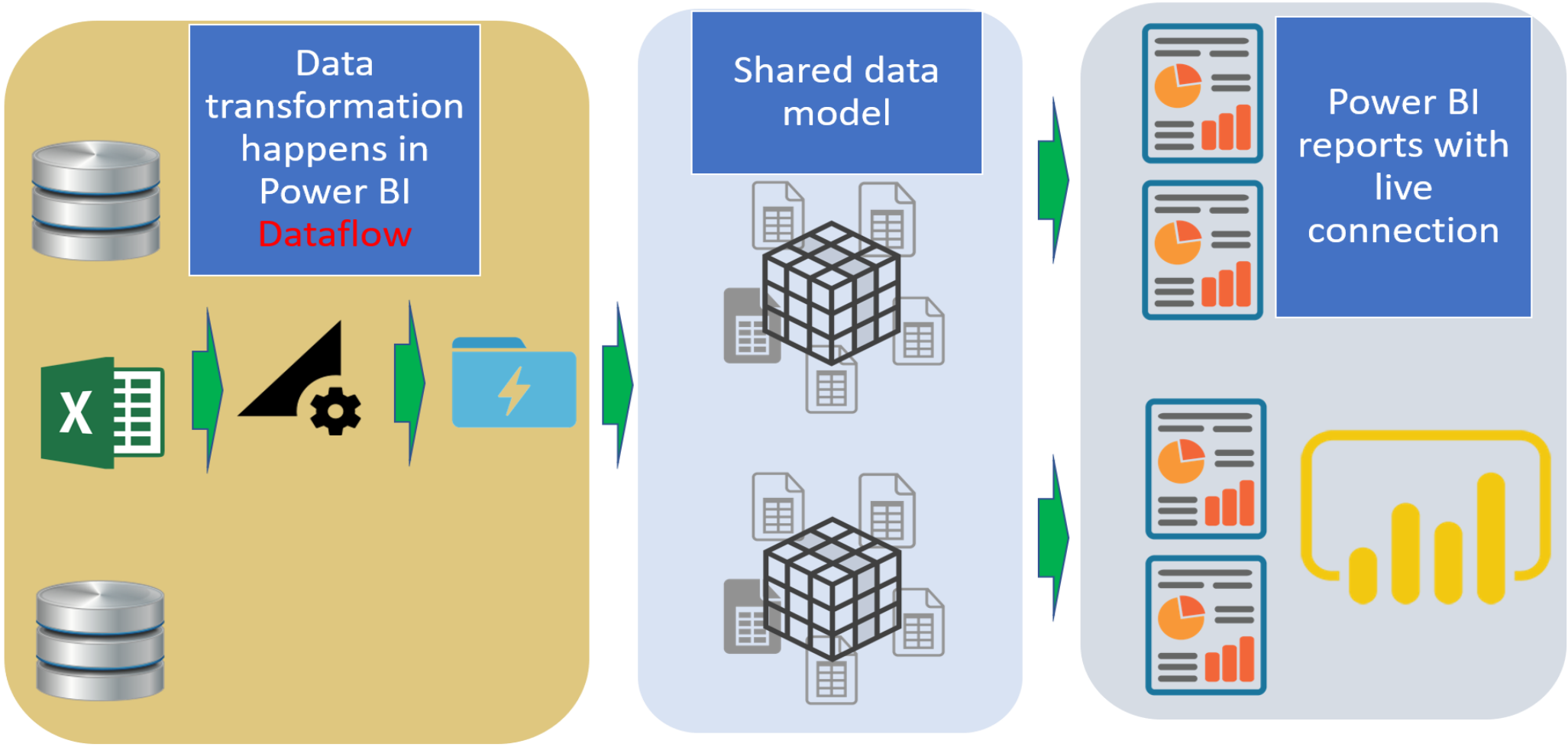
Shared Dataset

- Dataset used in multiple reports





Power BI Architecture for Multi-Developer Tenant Using Dataflows and Shared Datasets



Power BI Solution Architecture

Reduce the size of Power BI file

Be careful of dates