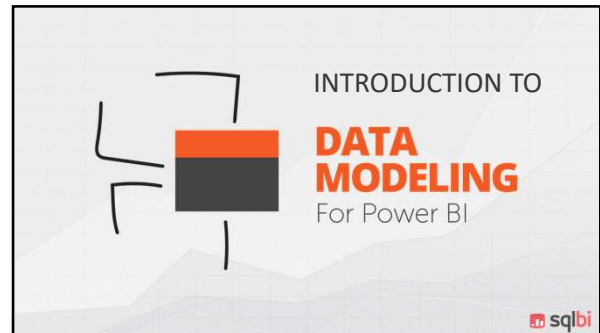
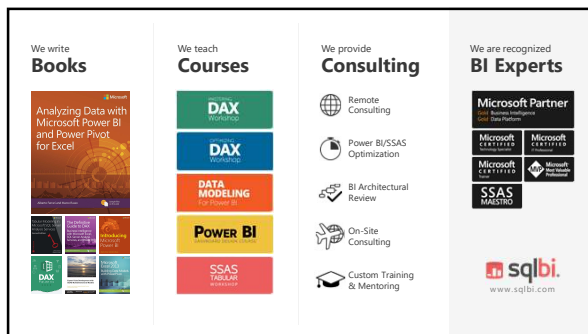




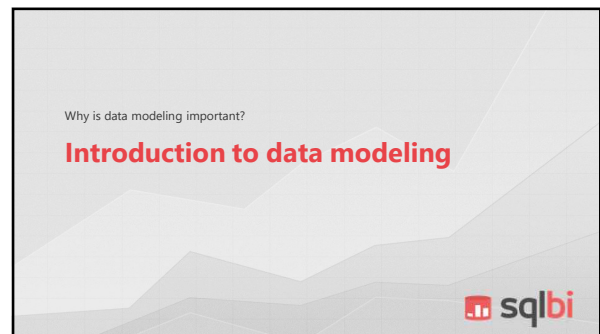
1



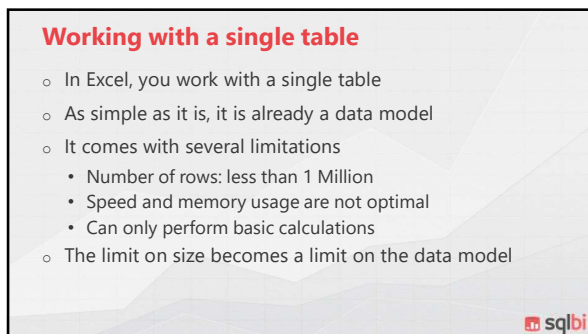
2



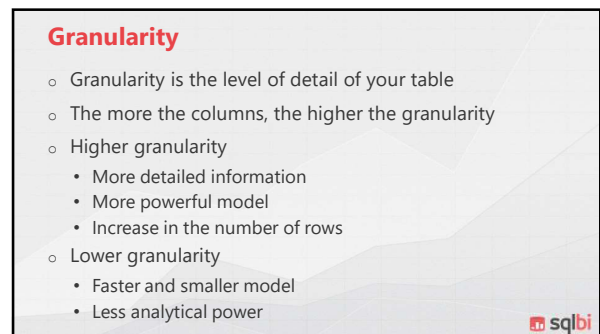
3



4



5



6

### Granularity and table size

- Increasing granularity increase the size of the model
  - More columns → More rows
- You quickly hit the limit of 1M rows of Excel...

Category	Sales
Bikes	10,000
Helmets	5,000

Category	Subcategory	Sales
Bikes	Cross bikes	8,000
Bikes	Mountain bikes	2,000
Helmets	Colorful helmets	2,000
Helmets	Lightweight helmets	3,000

7

### Scattered information

- Higher granularity is not always the best choice
- Too high is as bad as too low
- Example: yearly income repeated on every row

Education	AverageYearlyIncome
	10,000,000.00
Bachelors	63,702.67
Graduate Degree	67,481.24
High School	46,624.58
Partial College	53,519.76
Partial High School	36,584.57
<b>Total</b>	<b>7,769,508.95</b>

8

### Leveraging the data model

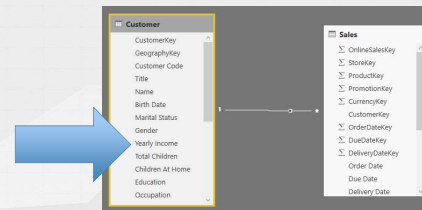
- Using the data model, you can load multiple tables
- Load Customers and Sales as separate tables
- Two tables need to be linked through a relationship
- Sales[CustomerKey] = Customer[CustomerKey]

If YearlyIncome is a customer related information, then you need a separate Customer table to store it.

9

### Customer is a business entity

- Being a business entity, it deserves a table by itself



10

### Business Entities

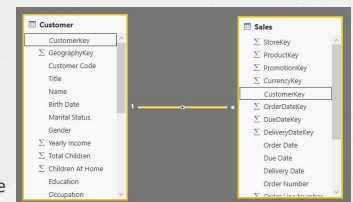
- Each business has different entities
  - Products, Customers, Resellers
  - Patients, Medications, Doctors
  - Claims, Customers
  - Teams, Workers, Buildings, Projects
  - Software, Features, Bugs, Customers
- Each business entity has unique characteristics

1 Entity = 1 Table

11

### Relationships

- Many tables need relationships
- Links two tables
- Has a direction
  - Customer: one side
  - Sales: many side
  - Many sales, one customer
- Best practice: same column name in both tables



12

## Granularity with multiple tables

- With multiple tables, granularity is a different topic
- Each table has its own granularity
  - Customers: at the customer level
  - Date: at the date level
  - Product: at the product level
- Sales has granularity defined by related tables
  - Customer, Date and Product level
  - If you have those three tables
- We will come back to granularity pretty often...



13

Adding and removing tables is the key skill of any data modeler

## Normalization and denormalization



14

## Normalization vs Denormalization

- Normalization is the process of organizing the columns (attributes) and tables of a database to reduce data redundancy and improve data integrity
- Denormalization is the opposite of normalization, that is increasing data redundancy, with the goal of improving the understanding of the model
- Let us see the concept with some examples



15

## Normalization of a table

Manufacturer	BrandName	ProductSubcategoryName
Adventure Works	Adventure Works	Coffee Machines
Contoso, Ltd	Contoso	Cell phones Accessories
Adventure Works	Adventure Works	Television/DVR
Fabrikam, Inc.	Fabrikam	Camcorders
Adventure Works	Adventure Works	Laptops
Contoso, Ltd	Contoso	Cell phones Accessories
Proseware, Inc.	Proseware	Projectors & Screens
Adventure Works	Adventure Works	Laptops
The Phone Company	The Phone Company	Touch Screen Phones
Contoso, Ltd	Contoso	Home & Office Phones
Fabrikam, Inc.	Fabrikam	Microwaves
Adventure Works	Adventure Works	Desktops
Contoso, Ltd	Contoso	Desktops & Screens
Contoso, Ltd	Contoso	Digital SLR Cameras
Adventure Works	Adventure Works	Desktops
Wide World Importers	Wide World Importers	Recording Pen
Wide World Importers	Wide World Importers	Recording Pen
Contoso, Ltd	Contoso	Microwaves
A. Datum Corporation	A. Datum	Digital Cameras
Litware, Inc.	Litware	Washers & Dryers

Brand Code	Brand Name
1	Adventure Works
2	Contoso
3	Fabrikam
4	Proseware
5	The Phone Company
...	...



16

## Working with a single table

- All columns are denormalized

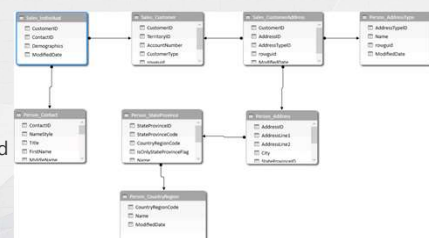
OrderDate	Manufacturer	BrandName	ProductSubcategoryName	ProductCategoryName	SalesQuantity	Subtotal	TotalCost
2007-08-11	Adventure Works	Adventure Works	Coffee Machines	Home Appliances	50	24152.298	7651.54
2008-10-22	Contoso, Ltd	Contoso	Cell phones Accessories	Cell phones	2040	22504.88	5245.94
2009-01-11	Adventure Works	Adventure Works	Television	TV and Video	154	51593.356	28346.4
2009-01-21	Fabrikam, Inc.	Fabrikam	Camcorders	Cameras and camcorders	252	162007.2	76700.43
2007-12-31	Adventure Works	Adventure Works	Laptops	Computers	29	14208.43	7944.32
2007-06-22	Contoso, Ltd	Contoso	Cell phones Accessories	Cell phones	680	6357.24	3420.44
2007-06-22	Proseware, Inc.	Proseware	Projectors & Screens	Computers	56	73417.6	30786.94
2007-08-23	Adventure Works	Adventure Works	Laptops	Computers	43	22672.2	9954.6
2009-01-30	The Phone Company	The Phone Company	Touch Screen Phones	Cell phones	198	48500.37	24344.56
2008-01-14	Contoso, Ltd	Contoso	Home & Office Phones	Cell phones	355	7251.536	3934.64
2007-09-30	Fabrikam, Inc.	Fabrikam	Microwaves	Home Appliances	44	4805.604	2824.24
2007-11-13	Adventure Works	Adventure Works	Desktops	Computers	153	47517.97	26254.02
2008-12-06	Contoso, Ltd	Contoso	Projectors & Screens	Computers	52	30796.4	6477.2
2007-11-14	Contoso, Ltd	Contoso	Digital SLR Cameras	Cameras and camcorders	146	53393.5	23876
2009-12-30	Adventure Works	Adventure Works	Desktops	Computers	32	13457.75	7952.97
2009-01-11	Wide World Importers	Wide World Importers	Recording Pen	Audio	42	7996.92	3607.26
2009-08-11	Wide World Importers	Wide World Importers	Recording Pen	Audio	9	1406.1	749.16
2009-05-28	Contoso, Ltd	Contoso	Microwaves	Home Appliances	78	9955.268	5129.27
2008-02-18	A. Datum Corporation	A. Datum	Digital Cameras	Cameras and camcorders	345	70989.33	32872.56
2007-08-15	Litware, Inc.	Litware	Washers & Dryers	Home Appliances	69	112603.8	56472.35



17

## Normalized models (OLTP)

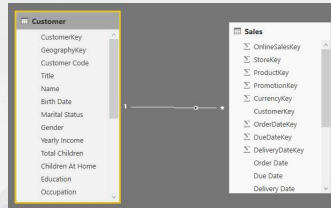
- This whole model represents a customer in an OLTP database
- There are a lot of different tables...
- Not a good model for queries



18

## Denormalized model (BI model)

- Denormalization is welcome, to make the model easier



19

Star schemas are the most popular way of modeling data in Business Intelligence

## Introducing star schemas



20

## Separation between facts and dimensions

- Different entities need different ways of handling
- Fact: *something that happened*
  - The sale of a product to a customer
  - A cash withdrawal on an ATM machine
  - The signature of an order
  - The prescription of a medical treatment
- Dimension: *something that describes a fact*
  - Attribute of a fact
  - The name of the customer, or of the patient
  - The date when the fact happened
  - The currency of the cash withdrawal



21

## What makes a dimension?

- One business entity = one table
- Attributes of an entity in the same table
- Customer is a business entity
  - Attributes: city, country, region, education, gender, age
- Usually Country is not an entity
  - It is an attribute of other dimensions
  - Country of customer, country of store
- Exception: demographic data
  - Measure: population (fact table)
  - Dimension: country (which is an entity in this model)



22

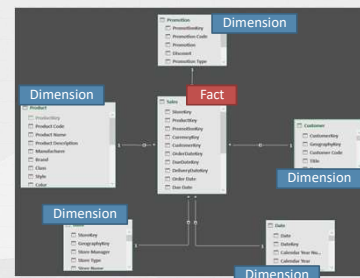
## Placing tables in a diagram

- Fact table
  - Stands, alone, in the center
- Dimensions
  - All around the fact table
  - Directly linked to it
- The figure that appears looks like a star
- Hence, the name: star schema



23

## Introducing star schemas



24

### Star schemas

- Very easy to understand at first glance
  - You slice by dimensions and aggregate facts
  - There is no ambiguity
  - One level of indirection makes it easy to see roles of tables
- Very fast
  - Modern engines are optimized for star schemas
- Drive a clean modeling path
  - Numbers go in the fact table
  - Strings go in the dimension
  - Everything else... we need to understand what it is



25

### If you don't have a star schema

- Most of the times, you are in trouble
- Any model change towards a star schema is a good step
- We will see several examples of this
- Your model is not different from all the other ones
  - As anybody else, you have a "special" model
  - With special requirements and special calculations
  - However, a star schema will fit it well!
- If you are unable to identify facts and dimensions
  - It is likely you still have to understand well the model



26

### Why data modeling is useful



27

### Why choosing a different model?

- If the model is not the right one
  - DAX code tends to be very complex
  - Formulas are hard to think at
  - Complexity turns into performance issues
- With the correct model
  - DAX code is simple, as it should be
  - Performance is great
- Building the right model requires experience



28

### Tasks of a data modeler

- Data modeling means
  - Knowing several patterns
  - Being able to match your model to a pattern
  - Apply the pattern
  - Adapt the small differences appearing in custom models
- You learn patterns with experience
- In this course, we present multiple patterns
- The goal is not learning them, but seeing them in action
  - Learning requires time, you will do it later



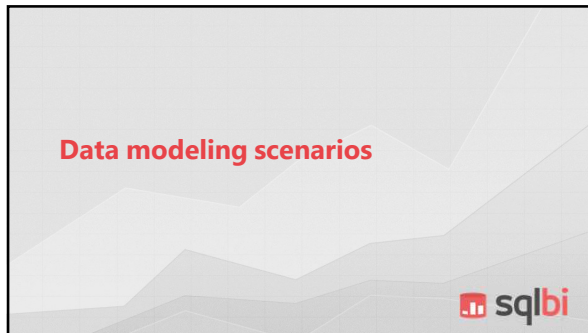
29

### Is your model a different one?

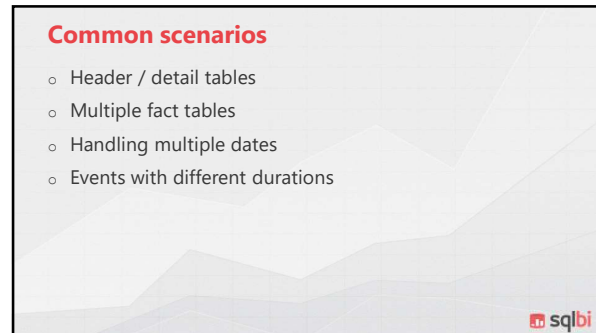
- At the beginning, you always feel your model is different than the standard ones
- 99.9% of the times, this is not the case
- Do not deviate from standard modeling, unless you really know what you are doing
- Business Intelligence was born in **1958**
- In **60 years**, we analyzed nearly any existing model
- And we found star schemas to be the best option



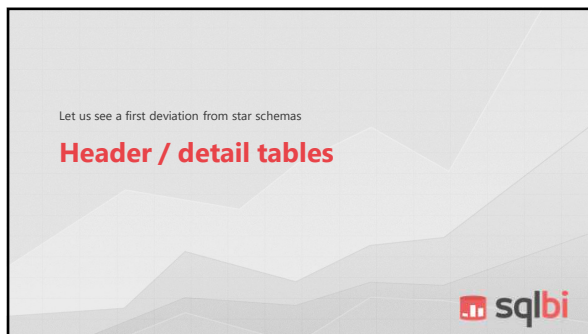
30



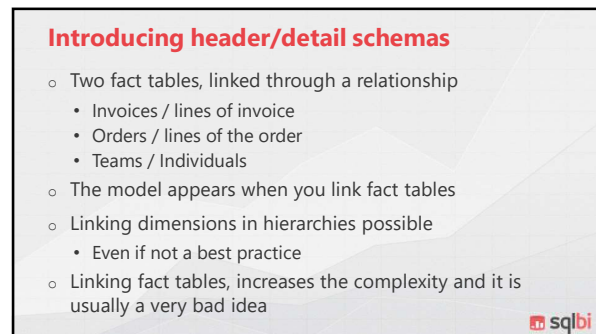
31



32



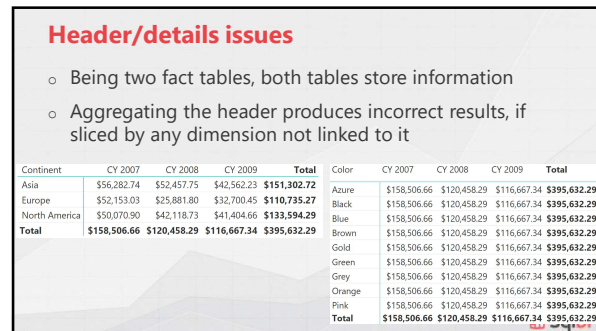
33



34



35



**Header/details issues**

- Being two fact tables, both tables store information
- Aggregating the header produces incorrect results, if sliced by any dimension not linked to it

Continent	CY 2007	CY 2008	CY 2009	Total
Asia	\$56,282.74	\$52,457.75	\$42,562.23	<b>\$151,302.72</b>
Europe	\$52,153.03	\$25,881.80	\$32,700.45	<b>\$110,735.27</b>
North America	\$50,070.90	\$42,118.73	\$41,404.66	<b>\$133,594.29</b>
<b>Total</b>	<b>\$158,506.66</b>	<b>\$120,458.29</b>	<b>\$116,667.34</b>	<b>\$395,632.29</b>

Color	CY 2007	CY 2008	CY 2009	Total
Azure	\$158,506.66	\$120,458.29	\$116,667.34	<b>\$395,632.29</b>
Black	\$158,506.66	\$120,458.29	\$116,667.34	<b>\$395,632.29</b>
Blue	\$158,506.66	\$120,458.29	\$116,667.34	<b>\$395,632.29</b>
Brown	\$158,506.66	\$120,458.29	\$116,667.34	<b>\$395,632.29</b>
Gold	\$158,506.66	\$120,458.29	\$116,667.34	<b>\$395,632.29</b>
Green	\$158,506.66	\$120,458.29	\$116,667.34	<b>\$395,632.29</b>
Grey	\$158,506.66	\$120,458.29	\$116,667.34	<b>\$395,632.29</b>
Orange	\$158,506.66	\$120,458.29	\$116,667.34	<b>\$395,632.29</b>
Pink	\$158,506.66	\$120,458.29	\$116,667.34	<b>\$395,632.29</b>
<b>Total</b>	<b>\$158,506.66</b>	<b>\$120,458.29</b>	<b>\$116,667.34</b>	<b>\$395,632.29</b>

36



## Back to a star schema

Once correctly denormalized, the model becomes a star schema again.



Once you identify the set of facts and dimensions, you no longer need header/detail tables.

Star schemas are always the best choice

37

Computing over multiple star schemas

## Multiple fact tables

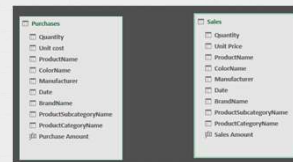
38

## Using multiple fact tables

- o Very common scenario
  - Sales and purchases
  - Orders and shipments
  - Sales and weather information
- o What we cover in this section
  - Build the correct set of dimensions
  - Use one fact table to filter the other one(s)

39

## Denormalized fact tables

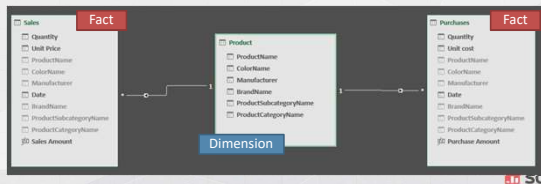


BrandName	Purchase Amount	Sales Amount
A. Datum	2,533,963.42	30,202,685.54
Adventure Works	6,048,167.59	30,202,685.54
Contoso	12,314,395.68	30,202,685.54
Fabrikam	10,003,071.13	30,202,685.54
Litware	6,377,548.93	30,202,685.54
Northwind Traders	1,713,836.80	30,202,685.54
Proseware	5,305,305.29	30,202,685.54
Southridge Video	2,199,989.35	30,202,685.54
Tailspin Toys	646,571.47	30,202,685.54
The Phone Company	3,045,608.33	30,202,685.54
Wide World Importers	4,151,139.81	30,202,685.54
<b>Total</b>	<b>54,339,597.80</b>	<b>30,202,685.54</b>

40

## Building a star schema

- o A proper star schema is nearly always the best choice
- o But how do we build the Product table?



41


## Options to build the new dimension

- o Use an SQL view, if feasible
- o Use M code in Power Query
  - Available in Excel and Power BI
- o Use DAX code and build a calculated table
  - Available in Power BI and SSAS 2016
- o You need a key for the new dimension
  - Easy in SQL
  - Harder in M or DAX, if primary key not already available

42

In a fact table you might have multiple dates, how should you handle them?

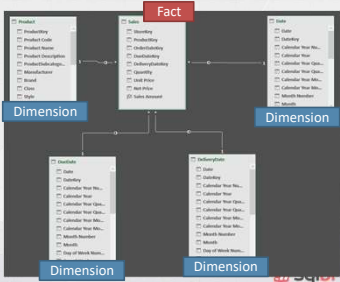

## Handling multiple dates



43

## Multiple date tables



- Multiple date tables
- Single fact table
- The model becomes more complicated
- Slicing multiple fact tables becomes troublesome
- Not a best practice

44

## Multiple relationships with date


- One table, multiple relationships
- Only one relationship can be active

45

Different events, different durations, different fact tables...


## Events with different durations



46

## Different durations


- This scenario happens when you have
  - Multiple fact tables
  - Each fact table contains some sort of event
  - The start date and the duration of different events is unrelated
- Example
  - Fact: hours worked by employees
  - Fact: the store where the employee is working
  - Fact: the salary of the employee, changing over time



47

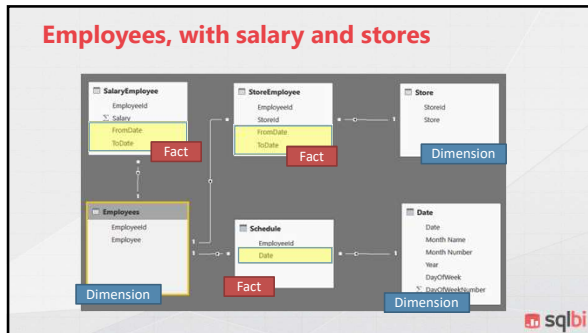
## The scenario

- SalaryEmployee
  - Salary of an employee
  - From date, to date
- StoreEmployee
  - Assignment of an employee to a given store
  - From date, to date
- Schedule
  - Working schedule of an employee
  - Daily granularity

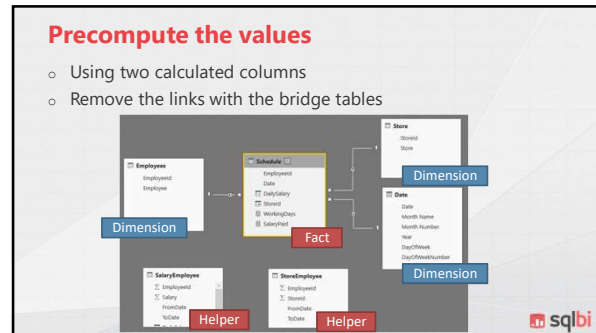


48






49



50


We write

**Books**



We teach

**Courses**




We provide

**Consulting**

- Remote Consulting
- Power BI/SSAS Optimization
- BI Architectural Review
- On-Site Consulting
- Custom Training & Mentoring

We are recognized

**BI Experts**



sqlbi.com

51

# Thank you!



Check our articles, whitepapers and courses on [www.sqlbi.com](http://www.sqlbi.com)

52