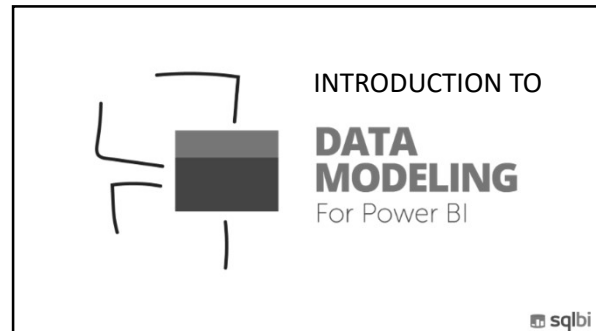
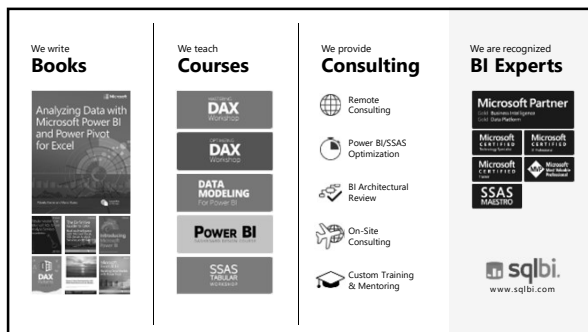




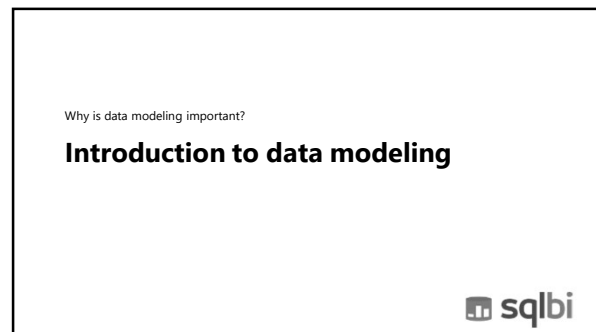
1



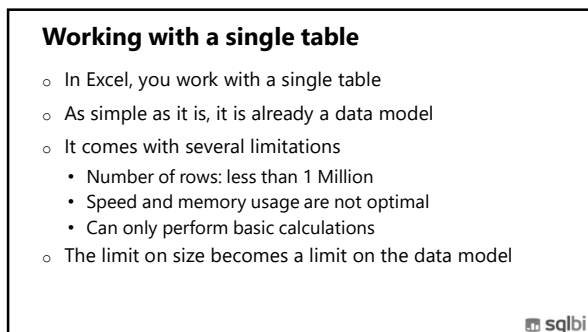
2



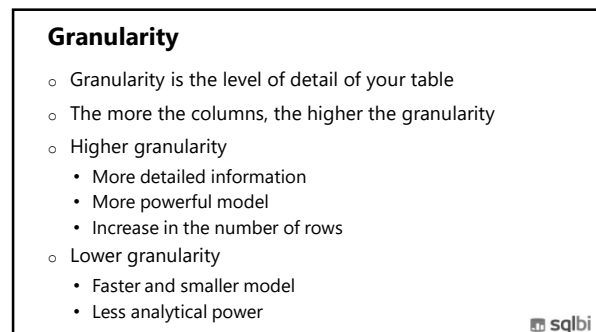
3



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	Category	Subcategory	Sales
Bikes	10,000	Bikes	Cross bikes	8,000
Helmets	5,000	Bikes	Mountain bikes	2,000
		Helmets	Colorful helmets	2,000
		Helmets	Lightweight helmets	3,000

sqlbi

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
Total	7,769,508.95

sqlbi

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]

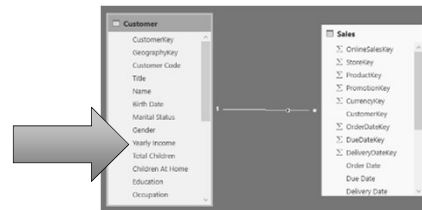
If YearlyIncome is a customer related information, then you need a separate Customer table to store it.

sqlbi

9

Customer is a business entity

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



sqlbi

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

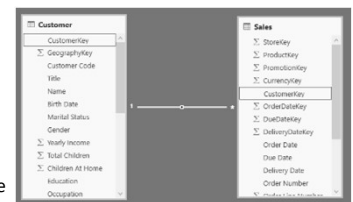
1 Entity = 1 Table

sqlbi

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



sqlbi

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	Brand Name	Product/Category Name
Adventure Works	Contoso	Coffee Machines
Contoso, Ltd	Contoso	Cell phones/Accessories
Adventure Works	Adventure Works	Televisions
Fabrikam, Inc.	Fabrikam	Cameras
Adventure Works	Contoso	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	Microscopy
Adventure Works	Contoso	Desktops
Contoso, Ltd	Contoso	Projectors & Screens
Contoso, Ltd	Contoso	Digital SLR Cameras
Adventure Works	Adventure Works	Desktops
Wide World Imports	Wide World Imports	Recording Pen
Wide World Imports	Wide World Imports	Recording Pen
Contoso, Ltd	Contoso	Microphones
A. Datum Corporation	A. Datum	Digital Cameras
Litware, Inc.	Litware	Washers & Dryers



16

Working with a single table

- All columns are denormalized

Year	Company	Manufacturer	Product Name	Product Category	Product Subcategory	Quantity	Value (\$)	Weight (kg)
2007-01-01	Adventure Works	Adventure Works	Coffee Machines	Home Appliances		35	14332.28	703.81
2007-01-01	Centros, Ltd.	Centros, Ltd.	Computers	Computers	Desktop Computers	2504	2104.43	14.43
2009-01-01	Adventure Works	Adventure Works	Televisions	TV and Video		124	5533.36	2818.6
2009-01-01	Fabrikam, Inc.	Fabrikam	Camcorders	Cameras and camcorders		282	10037.2	924.43
2009-01-01	Adventure Works	Adventure Works	Camcorders	Cameras and camcorders		14	440.43	4.43
2009-01-01	Centros, Ltd.	Centros, Ltd.	Cell Phones	Cell Phones		680	4320.2	9420.43
2009-01-01	Proseware, Inc.	Proseware, Inc.	Smartphones & Tablets	Smartphones & Tablets		7242.2	1420.43	1420.43
2009-01-01	Adventure Works	Adventure Works	Laptops	Computers		43	12573.2	9420.43
2009-01-01	The Home Company	The Home Company	Toast Toaster Bakes	Cell Phones		198	4803.67	2414.43
2009-01-01	Centros, Ltd.	Centros, Ltd.	Smartphones	Smartphones		213	213.43	213.43
2009-01-01	Fabrikam, Inc.	Fabrikam	Microphones	Home Appliances		43	4803.67	2414.43
2009-01-01	Adventure Works	Adventure Works	Computers	Computers		153	1733.2	1733.2
2009-01-01	Centros, Ltd.	Centros, Ltd.	Smartphones & Tablets	Smartphones & Tablets		43	1420.43	1420.43
2009-01-01	Centros, Ltd.	Centros, Ltd.	Digital Still Cameras	Cameras and camcorders		146	5337.2	5337.2
2009-01-01	Adventure Works	Adventure Works	Smartphones & Tablets	Smartphones & Tablets		146	5337.2	5337.2
2009-01-01	World Wide Importers	World Wide Importers	Recording Pen	Audio		43	9053.2	3627.2
2009-01-01	World Wide Importers	World Wide Importers	Recording Pen	Audio		2	1463.2	1463.2
2009-01-01	Centros, Ltd.	Centros, Ltd.	Home Appliances	Home Appliances		146	1463.2	1463.2
2009-01-01	A. Datum Corporation	A. Datum	Digital Cameras	Cameras and camcorders		345	7058.2	3227.2
2009-01-01	Uthmaniyah, Inc.	Uthmaniyah	Webcams & Webcams	Home Appliances		43	11263.2	11263.2

112603.8 56472.30
sqlbi

17

Normalized models (OLTP)

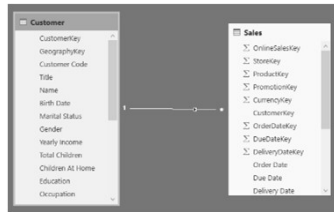
- This whole model represents a customer in an OLTP database
- There are a lot of different tables...
- Not a good model for queries



18

Denormalized model (BI model)

- Denormalization is welcome, to make the model easier



sqlbi

19

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

Introducing star schemas

sqlbi

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

sqlbi

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)

sqlbi

22

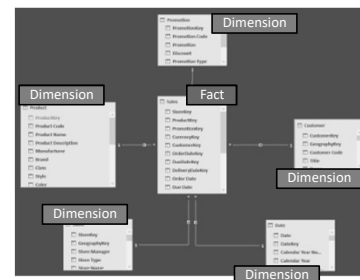
Placing tables in a diagram

- 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

sqlbi

23

Introducing star schemas



sqlbi

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

Data modeling scenarios



31

Common scenarios

- Header / detail tables
- Multiple fact tables
- Handling multiple dates
- Events with different durations



32

Let us see a first deviation from star schemas

Header / detail tables



33

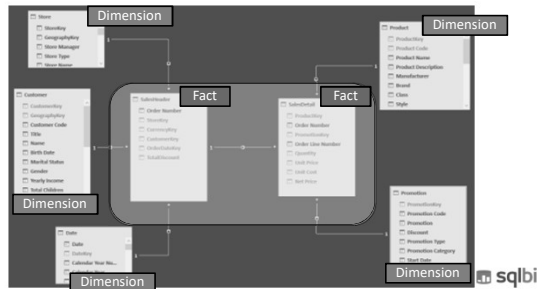
Introducing header/detail schemas

- Two fact tables, linked through a relationship
 - Invoices / lines of invoice
 - Orders / lines of the order
 - Teams / Individuals
- The model appears when you link fact tables
- Linking dimensions in hierarchies possible
 - Even if not a best practice
- Linking fact tables, increases the complexity and it is usually a very bad idea



34

Sales headers and sales details



35

Header/details issues

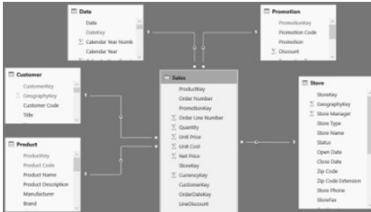
- Being two fact tables, both tables store information
- Aggregating the header produces incorrect results, if sliced by any dimension not linked to it

Continent	CY 2007	CY 2008	CY 2009	Total	Color	CY 2007	CY 2008	CY 2009	Total
Asia	\$56,282.74	\$52,457.75	\$42,562.23	\$151,302.72	Azure	\$158,506.66	\$120,458.29	\$116,667.34	\$395,632.29
Europe	\$52,153.03	\$25,881.80	\$32,700.45	\$110,735.27	Black	\$158,506.66	\$120,458.29	\$116,667.34	\$395,632.29
North America	\$50,070.90	\$42,118.73	\$41,404.66	\$133,594.29	Blue	\$158,506.66	\$120,458.29	\$116,667.34	\$395,632.29
Total	\$158,506.66	\$120,458.29	\$116,667.34	\$395,632.29	Brown	\$158,506.66	\$120,458.29	\$116,667.34	\$395,632.29
					Gold	\$158,506.66	\$120,458.29	\$116,667.34	\$395,632.29
					Green	\$158,506.66	\$120,458.29	\$116,667.34	\$395,632.29
					Grey	\$158,506.66	\$120,458.29	\$116,667.34	\$395,632.29
					Orange	\$158,506.66	\$120,458.29	\$116,667.34	\$395,632.29
					Pink	\$158,506.66	\$120,458.29	\$116,667.34	\$395,632.29
					Total	\$158,506.66	\$120,458.29	\$116,667.34	\$395,632.29

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

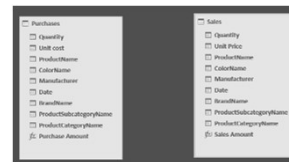
Using multiple fact tables

- 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



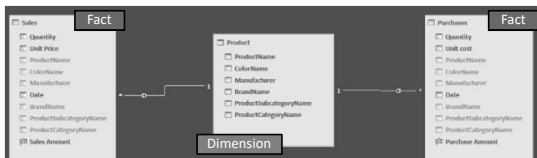
BrandName	Purchase Amount	Sales Amount
A. Datum	2,533,963.42	30,202,685.54
Adventure Works	6,048,167.59	30,202,685.54
Contoso	12,314,395.68	30,202,685.54
Fabrikam	10,003,071.13	30,202,685.54
Litware	6,377,548.93	30,202,685.54
Northwind Traders	1,713,836.80	30,202,685.54
Proseware	5,305,305.29	30,202,685.54
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
Total	54,339,597.80	30,202,685.54



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?

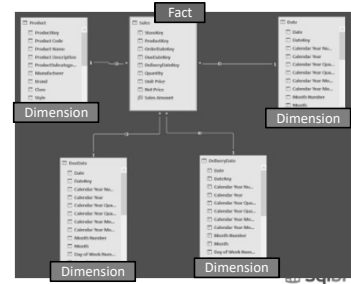
Handling multiple dates



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

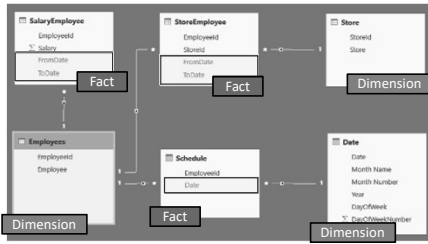
The scenario

- 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

Employees, with salary and stores

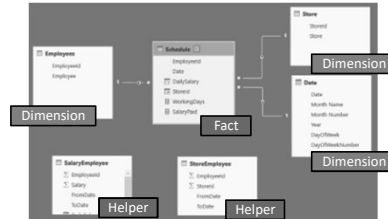


sqlbi

49

Precompute the values

- Using two calculated columns
- Remove the links with the bridge tables



sqlbi

50

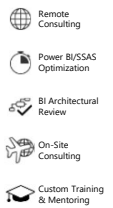
We write Books



We teach Courses



We provide Consulting



We are recognized BI Experts



sqlbi.
www.sqlbi.com

51

Thank you!



Check our articles, whitepapers and courses on
www.sqlbi.com

52