Data Modeling in SQL

2023-07-11

datacamp

About this case study

Data: Computer workstation sales, 1980–1996

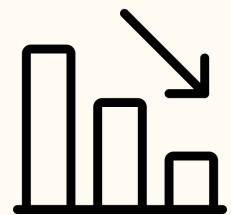http://five.dartmouth.edu/datasets

Primary Goal: Learn basics of dimensional data modeling with star ⭐ and snowflake ❄️ schema
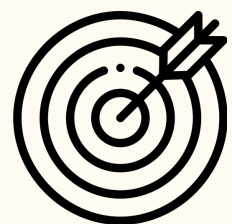
Secondary Goal: Build a basic data transformation pipeline for analytics

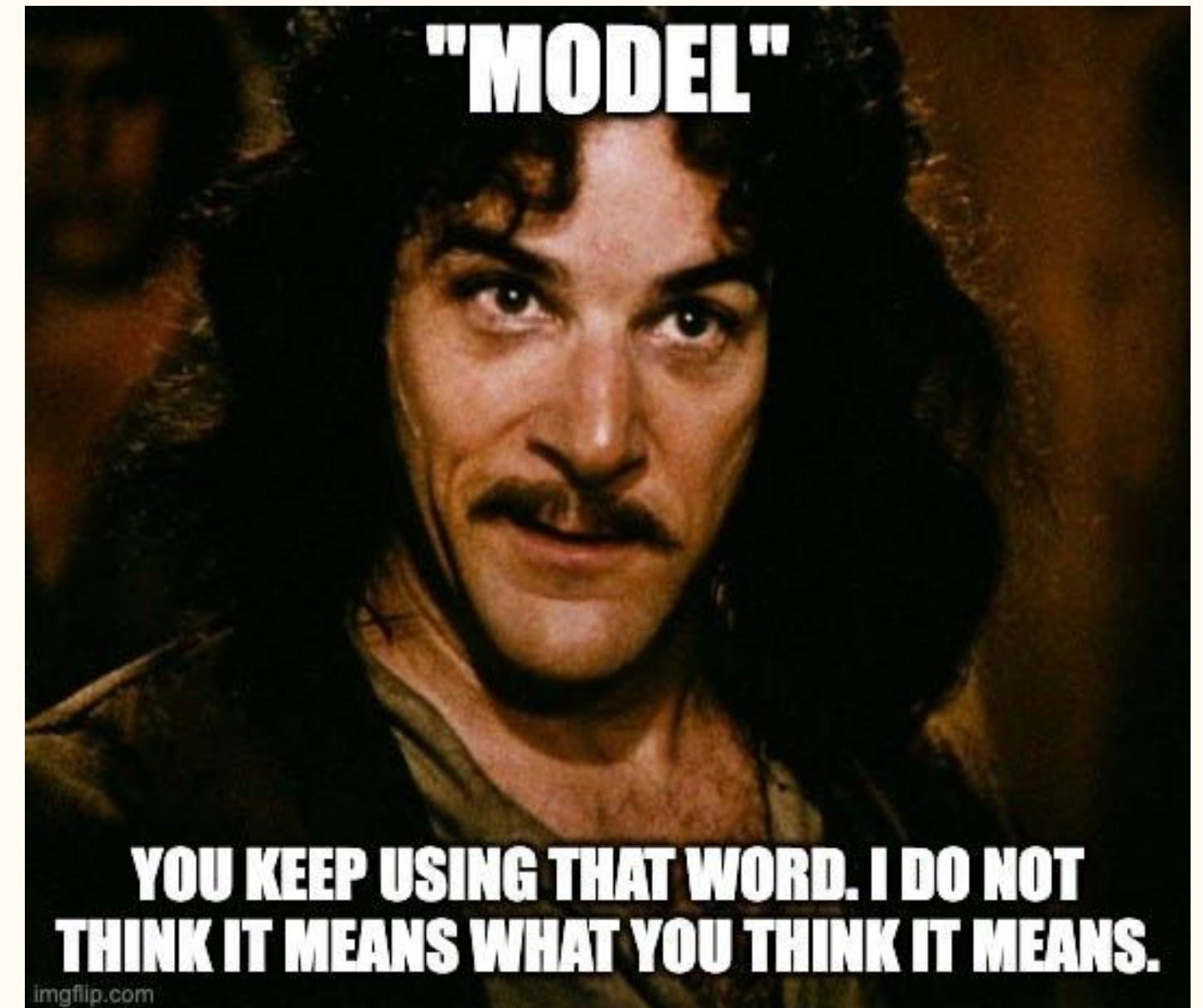**Why?** Poorly modeled data can lead to long and expensive 💸 queries

Soooo, what is data modeling?

Data model:

"a model is not simply the shape that data takes once uploaded to a database, but rather, the blueprint for business operations"

Gershkovich, Serge. Data Modeling with Snowflake: A practical guide to accelerating Snowflake development using universal data modeling techniques (p. 50). Packt Publishing. Kindle Edition.

Data modeling: the act of creating a model

"MODEL"

YOU KEEP USING THAT WORD. I DO NOT THINK IT MEANS WHAT YOU THINK IT MEANS.

imgflip.com

https://joereis.substack.com/p/joes-nerdy-rants-5

Before getting our hands dirty, some terminology

Primary key: A column (or set of columns) that uniquely identify a row

Foreign Key: A column that references another table's primary key

Fact/metric: The thing you are measuring (dollars, units, etc)

Dimension/Attribute: The different variables you want to slice and dice the facts by

Grain: The lowest level of detail captured

Steps of dimensional data modeling - Kimball

1. Define the business process
2. Declare the grain
3. Identify the dimensions
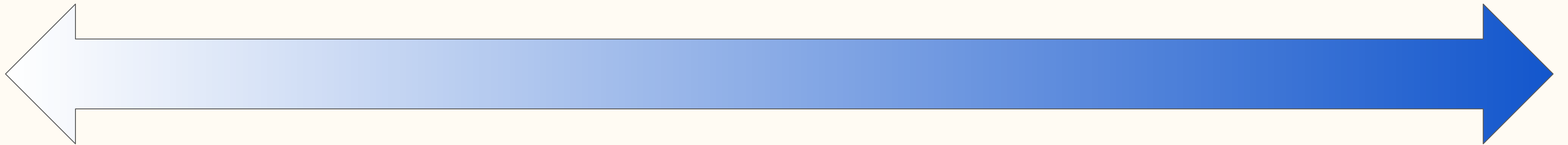4. Identify the facts

Let's get started!

Normalization

Normalization: Decreasing duplication in data

Less normalized

Less JOINs needed,
More data duplication,
Harder to modify data (UPDATE)

More normalized

More JOINs needed,
Less data duplication,
Easier to modify data (INSERT/DELETE)

How to remember types of normalization: The data is unique on the key (1NF), the whole key (2NF), and nothing but the key (3NF)