Database Normal Forms

Joseph Hallett

February 1, 2023

University of BRISTOL

What’s all this about?

Database theory!

- So far we’ve discussed how to doodle database designs…
- We’ve discussed how to create tables in SQL

This time:

How do we design tables that are easy to use?

Lets start with our records database again...

We could store our data as follows:

Artist	Albums
The Beatles	Yellow Submarine, White Album, Rubber Soul
Milk Can	Make It Sweet
Dresden Dolls	Yes Virginia, No Virginia, The Dresden Dolls

Please, no.

This is a *terrible* idea

- Yes we have one big table which seems neater
- But its much harder to do anything actually with

For example:

- How many albums does each artist have?
- Change all of *Prince*'s albums after 1993 to being by a *Love Symbol*
- How many artists have an album with the same name?

Normal forms

Normal forms prevent this sort of insanity

- Using them requires discipline, and rememebering rules...
- But is worth it for your sanity in the short to medium term

First Normal Form

Each column shall contain *one* (and only one) value

Each row says describes *multiple* albums per artist…

Artist	Albums
The Beatles	Yellow Submarine, White Album, Rubber Soul
Milk Can	Make It Sweet
Dresden Dolls	Yes Virginia, No Virginia, The Dresden Dolls

First Normal Form

Lets fix that...

Artist	Album
The Beatles	Yellow Submarine
The Beatles	White Album
The Beatles	Rubber Soul
Milk Can	Make It Sweet
Dresden Dolls	Yes Virginia
Dresden Dolls	No Virginia
Dresden Dolls	The Dresden Dolls

Lets add some more data to our table

Artist	Album	Year	Prime Minister
The Beatles	Yellow Submarine	1969	Harold Wilson
The Beatles	White Album	1968	Harold Wilson
The Beatles	Rubber Soul	1965	Harold Wilson
Milk Can	Make It Sweet	1999	Tony Blair
Dresden Dolls	Yes Virginia	2006	Tony Blair
Dresden Dolls	No Virginia	2008	Gordon Brown
Dresden Dolls	The Dresden Dolls	2003	Tony Blair

Second Normal Form

Every non-key attributue is fully dependent on the key

In this case the key is *Artist, Album*

- And arguably *year* too if you’re gonna pull a Taylor Swift and rerelease all your albums…

Is *Prime Minister* dependent on the key?

- No. Put it in a different table.

Now it looks like

Artist	Album	Year
The Beatles	Yellow Submarine	1969
The Beatles	White Album	1968
The Beatles	Rubber Soul	1965
Milk Can	Make It Sweet	1999
Dresden Dolls	Yes Virginia	2006
Dresden Dolls	No Virginia	2008
Dresden Dolls	The Dresden Dolls	2003

Year	Prime Minister
1969	Harold Wilson
1968	Harold Wilson
1965	Harold Wilson
1999	Tony Blair
2006	Tony Blair
2008	Gordon Brown
2003	Tony Blair

Third Normal Form

> *Every non-key attribute must provide a fact about the key, the whole key and nothing but the key; so help me Codd.*

Lets add some extra information to our table of Prime Ministers...

Year	Prime Minister	Birthday
1969	Harold Wilson	1916-03-11
1968	Harold Wilson	1916-03-11
1965	Harold Wilson	1916-03-11
1999	Tony Blair	1953-05-06
2003	Tony Blair	1953-05-06
2006	Tony Blair	1953-05-06
2008	Gordon Brown	1951-02-20

Our key is (*Year*, *Prime Minister*); Birthday depends on *Prime Minister*.

- So every non-key depends on the key...
- So 2NF

But not 3NF as Birthday *doesn't* tell you a fact about the *whole key*... just the Prime Minister.

So split it up!

Year	Prime Minister
1969	Harold Wilson
1968	Harold Wilson
1965	Harold Wilson
1999	Tony Blair
2003	Tony Blair
2006	Tony Blair
2008	Gordon Brown

Prime Minister	Birthday
Harold Wilson	1916-03-11
Tony Blair	1953-05-06
Gordon Brown	1951-02-20

Why is this better?

- Now if we need to alter the birthday of a PM (or any other fact about that key)...
- ...then we only need to alter it in one place.

Other normal forms…

Boyce-Codd Normal Form

a set of one or more attributes (columns) that uniquely identify a row in a table.
Each table can have one or more candidate keys,
but only one of these candidate keys can be selected as the primary key for the table.

A slightly stronger form of 3NF…

- Sometimes called 3.5th Normal Form

Every possible *candidate key* for a table is also in 3NF.

- Split a 3NF table into tables with single candidate keys to get 3.5NF.

4th Normal Form

If multiple attributes in a table depend on the same key,

- Then those attributes should be dependant too
- Otherwise split them into separate tables…

5th Normal Form

It’s in 4th normal form and you can’t split it into more separate tables.

This is all getting a bit mathsy...

You can look up formal definitions for each of the normal forms

- (and you should)

But so long as you keep things *as separate as possible*, you'll usually hit at least 3NF by accident.

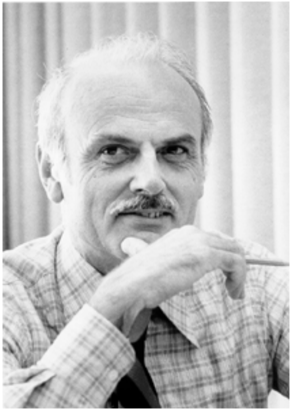
- ...and practically speaking your probably good then
- Getting it to 5NF *does* make things more flexible in the long run...
- But a 3.5NF database is often *good enough*.

Ultimately design is subjective (somewhat).

- ...but mathematical proof of flexibility is good right?

In conclusion

Every non-key attribute must provide a fact about the key, the whole key and nothing but the key; so help me Codd.



Ted Codd