

Normalisierung von Datenbanken



Photo by Tobias Fischer on Unsplash

1. Normalform: „Atomare Attribute“

Nur atomare Attributwerte

- **Nicht** in 1. Normalform:

CD_ID	Album	Gründungsjahr	Erscheinungsjahr	Titelliste
4711	Anastacia - Not That Kind	1999	2000	{1. Not That Kind, 2. I'm Outta Love, 3. Cowboys & Kisses}
4712	Pink Floyd - Wish You Were Here	1965	1975	{1. Shine On You Crazy Diamond}
4713	Anastacia - Freak of Nature	1999	2001	{1. Paid my Dues}

Titelliste, Interpret und Album liegen nicht **atomar** vor, sondern in einem Attribut „versteckt“. Man kann Titel oder Interpreten hier nicht „direkt“ abfragen.

- Überführung in 1. Normalform

CD_ID	Albumtitel	Interpret	Gründungsjahr	Erscheinungsjahr	Track	Titel
4711	Not That Kind	Anastacia	1999	2000	1	Not That Kind
4711	Not That Kind	Anastacia	1999	2000	2	I'm Outta Love
4711	Not That Kind	Anastacia	1999	2000	3	Cowboys & Kisses
4712	Wish You Were Here	Pink Floyd	1965	1975	1	Shine On You Crazy Diamond
4713	Freak of Nature	Anastacia	1999	2001	1	Paid my Dues

Alle wesentlichen Eigenschaften der *modellierten Miniwelt* liegen in eigenen Attributen vor. Wie fein das sein muss hängt vom Modell ab (Datum? Tag/Monat/Jahr?)

2. Normalform: „Monothematische Entitäten“

1. Normalform + Nicht-Schlüsselattribute von keinem Teil eines Schlüssels abhängig
 (Einfacher aber ungenauer: „Für jeden Primärschlüssel sind alle Attributwerte eindeutig“)

- Nicht in 2. Normalform:**

CD_ID	Albumtitel	Interpret	Gründungsjahr	Erscheinungsjahr	Track	Titel
4711	Not That Kind	Anastacia	1999	2000	1	Not That Kind
4711	Not That Kind	Anastacia	1999	2000	2	I'm Outta Love
4711	Not That Kind	Anastacia	1999	2000	3	Cowboys & Kisses
4712	Wish You Were Here	Pink Floyd	1965	1975	1	Shine On You Crazy Diamond
4713	Freak of Nature	Anastacia	1999	2001	1	Paid my Dues

Redundante Speicherung der Attributwerte für Albumtitel, Interpret ... bedingt
 Notwendigkeit *mehrfacher Änderungen*. Gefährdung der Integrität.

- Überführung in 2. Normalform durch Aufteilung in mehrere Tabellen:

CD					Lied		
CD_ID	Albumtitel	Interpret	Gründungsjahr	Erscheinungsjahr	CD_ID	Track	Titel
4711	Not That Kind	Anastacia	1999	2000	4711	1	Not That Kind
4712	Wish You Were Here	Pink Floyd	1965	1975	4711	2	I'm Outta Love
4713	Freak of Nature	Anastacia	1999	2001	4711	3	Cowboys & Kisses
					4712	1	Shine On You Crazy Diamond
					4713	1	Paid my Dues

2. Normalform: „Monothematische Entitäten“

Die 2NF entsteht oft schon „automagisch“ beim DB-Design mit einem ER-Diagramm, wenn man beim Überführen in Relationen den Grundsatz

1 Entität ↔ 1 Tabelle

einhält und die Attribute hinreichend atomar wählt.

3. Normalform: „Keine transitiven Abhängigkeiten“

2. Normalform + kein Nichtschlüsselattribut hängt von einem Schlüsselattribut transitiv ab.

- **Nicht in 3. Normalform:**

CD					Lied		
CD_ID	Albumtitel	Interpret	Gründungsjahr	Erscheinungsjahr	CD_ID	Track	Titel
4711	Not That Kind	Anastacia	1999	2000	4711	1	Not That Kind
4712	Wish You Were Here	Pink Floyd	1965	1975	4711	2	I'm Outta Love
4713	Freak of Nature	Anastacia	1999	2001	4711	3	Cowboys & Kisses
					4712	1	Shine On You Crazy Diamond
					4713	1	Paid my Dues

CD_ID → Interpret → Gründungsjahr

Das Gründungsjahr hängt transitiv vom Schlüsselattribut CD_ID ab das erzeugt eine Redundanz!

- **3. Normalform durch weitere Tabelle**

CD				Künstler			Lied		
CD_ID	Albumtitel	Interpret_ID	Erscheinungsjahr	Interpret_ID	Interpret	Gründungsjahr	CD_ID	Track	Titel
4711	Not That Kind	311	2000	311	Anastacia	1999	4711	1	Not That Kind
4712	Wish You Were Here	312	1975	312	Pink Floyd	1965	4711	2	I'm Outta Love
4713	Freak of Nature	311	2001				4711	3	Cowboys & Kisses
							4712	1	Shine On You Crazy Diamond
							4713	1	Paid my Dues

Weitere Normalformen für die Schule wenig relevant

- Boyce-Codd-Normalform (BCNF)
- 4NF
- 5NF

[https://de.wikipedia.org/w/index.php?title=Normalisierung_\(Datenbank\)](https://de.wikipedia.org/w/index.php?title=Normalisierung_(Datenbank))

Praktisch wenig relevant, da:

- 3NF verhindert bei guter Modellierung Redundanzen und Anomalien und sorgt dadurch für Zielerreichung unter den Gesichtspunkten **Redundanz**, **Performance**, und **Flexibilität**

3NF kann gut „flexibel“ um neue Anforderungen des Modells erweitert werden, ohne die Konsistenz zu zerstören.

SQL-Abfragen werden durch weitere Normalisierung komplexer, weil noch mehr Tabellen abgefragt und verknüpft werden müssen.