

Normalisierung einer DB

1.-3. Normalform (ausreichend guter Entwurf)

(es gibt auch die BCNF sowie die 4. und 5. Normalform)

Ziel: klare, logische Struktur + Vermeidung von Anomalien/ Inkonsistenzen, Redundanzen
(d.h. durch die Normalisierung soll die **Integrität (s.a.u.)** der Daten sichergestellt werden)

Beachte: „Ein *exakter Modellentwurf (ERM/RDM)* macht eine Normalisierung eigentlich überflüssig.“

- 1. Normalform:** alle Attributwerte sind **atomar** (nicht aus mehreren Elementen [durch Komma getrennt] zusammengesetzt) – bei Bedarf Auslagern in zusätzlichen Zeilen oder Einfügen zusätzlicher Spalten oder Erstellen neuer Tabellen (und Anpassen von Schlüsseln/ Beziehungen)
- 2. Normalform:** 1. Normalform + jedes Nicht-Schlüssel-Attribut ist **(voll) funktional abhängig** vom Gesamtschlüssel (und nicht bereits von Teilen davon) [– bei einspaltigem Primärschlüssel automatisch erfüllt] – ggf. Aufsplittern in mehrere Tabellen (und Anpassen von Schlüsseln/ Beziehungen)
- 3. Normalform:** 2. Normalform + **keine (direkten) funktionalen Abhängigkeiten** zwischen Nicht-Schlüssel-Attributen – ggf. weitere Tabellen (und auf Verbleib von Fremdschlüsseln achten)

Nach der 3. Normalform folgt oft wieder **Denormalisierung** zwecks Laufzeitverbesserung (durch Normalisierung entstehen oft mehr Tabellen, Abfragen werden teils komplizierter/ langsamer und der Gesamtüberblick geht ggf. in Teilen verloren).

Beispiel: CD-Sammlung (*Primärschlüssel stets kursiv*)

CD_Lied				
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}

Im Attribut **Album** werden Albumtitel und Interpret gespeichert, im Attribut **Titelliste** eine Menge von Titeln – man kann so z.B. nicht nach einzelnen Liedern suchen - **1. Normalform:**

CD_Lied						
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

Der **Albumtitel**, der **Interpret** und das **Erscheinungsjahr** sind zwar von der **CD_ID** als solche abhängig, nicht aber vom **Track** - **2. 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

In der Relation **CD** ist z.B. **Albumtitel** direkt vom Primärschlüssel **CD_ID** abhängig, das **Gründungsjahr** hingegen hängt bereits direkt vom Interpreten ab - **3. Normalform**:

CD				Künstler		
CD_ID	Albumtitel	Interpret_ID	Erscheinungsjahr	Interpret_ID	Interpret	Gründungsjahr
4711	Not That Kind	311	2000	311	Anastacia	1999
4712	Wish You Were Here	312	1975	312	Pink Floyd	1965
4713	Freak of Nature	311	2001			

Lied		
CD_ID	Track	Titel
4711	1	Not That Kind
4711	2	I'm Outta Love
4711	3	Cowboys & Kisses
4712	1	Shine On You Crazy Diamond
4713	1	Paid my Dues

Integrität von Daten

Die Integrität/ Konsistenz von Daten in relationalen Datenbanken wird durch das Aufstellen von sogenannten **Integritätsbedingungen** definiert. Man unterscheidet dabei zusätzlich verschiedene Arten der Integritätsbestimmungen:

- Bereichsintegrität
 - Der Wert jedes Attributes muss in einem bestimmten Wertebereich liegen.
- Entitätsintegrität:
 - Der Primärschlüssel jedes Objekts muss eindeutig sein. Der Feldinhalt darf dabei auf keinen Fall leer sein.
- Referentielle Integrität:
 - Der Inhalt eines Fremdschlüsselfeldes muss entweder leer (*null*) sein oder ein Objekt mit einem solchen (Primär-)Schlüssel muss existieren.
- Logische Konsistenz:
 - Der Benutzer kann auch zusätzliche Integritätsbestimmungen definieren (z. B. bei einer Stammbaumdatenbank: Die Kinder müssen nach den Eltern geboren worden sein). Solche Bedingungen können in der Regel vom Datenbanksystem nicht kontrolliert werden und müssen deshalb vom Benutzer selbst erfüllt werden.