

## Handout

### Themenfeld: Datenbanken und SQL

### Abschnitt: 04.01. Normalisierung

Autor: Thomas Krause

Stand: 14.11.2022 12:02:00

#### Inhalt

1	Einführung .....	2
2	Grundlagen.....	3
2.1	Was ist Normalisierung? .....	3
2.2	Normalform 1 .....	4
2.3	Normalform 2 .....	5
2.4	Normalform 3 .....	6
2.5	Welche Normalformen gibt es noch? .....	7
3	Weitere Quellen zum Thema .....	7



# 1 Einführung

## ZIELE:

- TN (Teilnehmer) erkennen die Bedeutung der Normalisierung in relationalen Datenbanken
- TN können NF1...3 grundlegend mit eigenen Worten beschreiben
- TN sind informiert, welche es noch gibt und welche wichtig sind für die Praxis

## INHALTE:

- Was ist Normalisierung?
  - Ziel?
  - wann?
  - wie?
- Normalformen im Detail NF1 ... NF3
  - Regeln
  - praktische Umsetzung
- welche sind wichtig, welche gibt es noch?



## 2 Grundlagen

### 2.1 Was ist Normalisierung?

Arbeitsschritte der Datenbank-Modellierung:

1. Szenario/ Pflichtenheft/ Lastenheft
2. Informationsanalyse/ Anforderungsanalyse
3. konzeptioneller Entwurf → ERM
4. logischer Entwurf → Relationenmodell (Tabellen)
5. Normalisierung → **Überprüfung** auf Einhaltung der Normalformen 1 ... 3
6. Implementierung in einem realen Datenbanksystem

#### Was ist Normalisierung?

- die sogenannte **Normalisierung** definiert Regeln und Maßnahmen für die (weitgehende) Sicherstellung einer
  - a. **redundanzarmen** Speicherung → jeder Fakt ist nur einmal in der Datenbank gespeichert
  - b. **Integritäts-/ Konsistenzsicherung** → es sind keine widersprüchliche Daten gespeichert
  - c. höchstmögliche **Flexibilität** (beim Erfassen, Ändern und Löschen von Daten)
  - d. **kurzen** Zugriffszeit
- NF 1 – 3 → sind für die Praxis wichtig

#### Welche Arten von N. gibt es?

es existieren die Normalformen 1, 2, 3, 4 und 5 sowie die Boyce-Codd-Normalform (BCNF)

#### Wie und wann führt man N. durch?

die Herstellung bzw. Überprüfung der Einhaltung der NF 1 ... 3 erfolgt i.d.R. parallel zur Erstellung des ERM und des Relationenschemas



## 2.2 Normalform 1

- Sind alle Einzelinformationen in eigenen Feldern gespeichert?
- **Eine Relation = Tabelle befindet sich in 1. Normalform (1NF), wenn alle Attributwerte elementar/ atomar (= unteilbar) sind.**
- Was bedeutet atomar: Attribut ist keine Menge, Aufzählungstyp, ... nicht mehrere Informationen, Aufzählungen o.ä. in einem Feld speichern
- **Bedingung für NF1 nicht erfüllt? → Lösungsmöglichkeiten?**
  - alle Informationen vereinzeln: Spalten = Datenfelder hinzufügen, Zeilen = Datensätze hinzufügen bzw. duplizieren
  - mehrere verschiedene Informationen auf mehrere Spalten = Datenfelder verteilen
  - mehrere gleichartige Information (z.B. Baustellen-Nr, Baustellen-Name) müssen auf mehrere Datensätze verteilt werden → Datensätze duplizieren
- für die Tabelle einen gültigen Primärschlüssel definieren, bei Bedarf muss ein zusammengesetzter Primärschlüssel gebildet werden (keinen künstlichen Primärschlüssel einführen)
- **WICHTIG: Jede Tabelle, die normalisiert wurde, muss dann auch über einen gültigen Primärschlüssel verfügen!**

Ausgangssituation (ein sogenanntes "naives Schema"):

Baustellen	Ableitung	Maschinen-berechtigung	Mitarbeiter	Wohnort	Bank-Verbindung
B021 „MIDL“ 12 Stunden Magdeburg, B112 „Kaufstadt“ 23 Stunden Hannover	12, Ausbau	J	Stein, Sven, Mitarbeiter-Nr.: M010	04838 Eilenburg	Konto: IBAN: DE43200701342478223697 BIC: LESBDELE287 Spar Bank 04229 Leipzig Kastanienallee 113 – 117 Tel.: 0341 – 998877 666
B253 „GaleriaX“ 37 Stunden Göttingen	9, Hochbau	N	Örtel, Olaf, Mitarbeiter-Nr.: M009	04105 Leipzig	Konto: IBAN: DE431003331122998247336 BIC: HASJDEHS881 City-Bank 06110 Halle/ Saale Handel-Allee 12 Tel.: 0345 – 586731 335
B056 „Brutto“ 21 Stunden Berlin, B112 „Kaufstadt“ 24 Stunden Hannover, B253 „GaleriaX“ 34 Stunden Göttingen	10, Haustechnik	J	Hahn, Harry, Mitarbeiter-Nr.: M021	04509 Delitzsch	Konto: IBAN: DE43200701342478221148 BIC: LESBDELE287 Spar Bank 04229 Leipzig Kastanienallee 113 – 117 Tel.: 0341 – 998877 668
B056 „Brutto“ 8 Stunden Berlin, B253 „GaleriaX“ 24 Stunden Göttingen	9, Hochbau	N	Holzer, Horst, Mitarbeiter-Nr.: M024	04119 Leipzig	Konto: IBAN: DE90128577701887277649 BIC: FRHEKDEM112 Geld-Bank 60438 Frankfurt am Main Main-Kai 201 – 203 Tel.: 069 – 887799 332

siehe Datei: **H\_04.01.01.Normalisierung\_hochbau.xlsx**

Zusammenfassung der Regel:

NF1: Eine Relation befindet sich in 1. Normalform (NF1), wenn alle Attributwerte elementar/ atomar (= unteilbar) sind. Sind alle Einzelinformationen in eigenen Felder abgespeichert? Sind alle Informationen in den Feldern atomar, also nicht weiter aufteilbar?



## 2.3 Normalform 2

- Eine Tabelle befindet sich in der zweiten Normalform, wenn sie sich  
(1) in NF 1 befindet und  
(2) jedes Attribut, das nicht Primärschlüsselattribut ist (wird Nichtschlüsselattribut genannt), vom **gesamten** Primärschlüssel (und nicht nur von Teilen des Primärschlüssels (bei zusammengesetzten Primärschlüsseln)) voll funktional abhängig ist.
- Hinweis: Ausgangstabellen, die sich in NF1 befinden und **keinen** zusammengesetzten Primärschlüssel haben, müssen **nicht** auf NF2 geprüft werden  
→ diese Tabellen können die NF2 nicht verletzen
- **Bedingung für NF2 nicht erfüllt? → Lösungsmöglichkeit?**
  - zunächst Ausgangstabellen auf Verletzungen der NF2 prüfen
  - Verletzungen der NF2 werden dadurch gelöst, dass die Ausgangstabelle in mehrere Tabellen zerlegt wird, in der die zusammengehörigen Informationen gespeichert werden können
- anschließend für alle Tabellen jeweils gültige Primärschlüssel definieren → es können auch zusammengesetzte Schlüssel sein
- **WICHTIG: Jede Tabelle, die normalisiert wurde, muss dann auch über einen gültigen Primärschlüssel verfügen!**

### Zusammenfassung der Regel:

NF2: Relation in NF1? Jedes Attribut ist vom **gesamten** Primärschlüssel und nicht nur von Teilen des Primärschlüssels abhängig.



## 2.4 Normalform 3

- Eine Tabelle befindet sich in der dritten Normalform (3NF), wenn sie sich
  - (1) in 2NF befindet und
  - (2) keine transitiven Abhängigkeiten aufweist (weitere Erläuterung unten bei Anforderungen)

Der Begriff '**transitiv**' kann mit '**indirekt**' übersetzt werden. Beispiel:  
 Eine transitive = indirekte Abhängigkeit liegt vor, wenn ein Attribut a von einem Attribut b und das Attribut b von einem Attribut c abhängig ist. Damit ist Attribut a indirekt, also transitiv, von c abhängig.
- Anforderungen:
  - jedes Attribut, das nicht zum Primärschlüssel gehört, muß **vollständig** vom Primärschlüssel abhängig sein  
 (Attribute, die nicht zum Primärschlüssel gehören, werden auch **Nichtschlüsselattribute** genannt.)
  - Alle **Nichtschlüsselattribute** müssen also untereinander voneinander unabhängig sein.
  - Es dürfen also auch keine berechneten Felder verwendet werden (z.B. Geburtsdatum, Alter → Alter berechnet sich aus Geburtsdatum).
  - Felder, die in keiner direkten Abhängigkeit vom Primärschlüssel stehen, müssen in einer eigenen Tabelle gespeichert werden.
  - **Bedingung für NF3 nicht erfüllt? → Lösungsmöglichkeit? →**  
 vorhandene Tabellen teilen und neue Tabellen einfügen
- anschließend für alle Tabellen einen gültigen Primärschlüssel definieren
- **HINWEIS** zu PLZ und Wohnort: Sind zwei unterschiedliche, voneinander unabhängige Informationen: Wohnort ist Ort, PLZ ist Kennzeichen für Zustellgebiet und Zustellgebiet ist nicht gleich Wohnort

### Zusammenfassung:

NF3: Relation in NF2? Jedes Attribut ist **vollständig** vom Primärschlüssel abhängig. Alle Attribute, die nicht Primärschlüssel sind, dürfen nicht voneinander abhängig sein. So dürfen auch keine berechneten Felder verwendet werden.



## 2.5 Welche Normalformen gibt es noch?

... und sonst noch (nur informativ)

- **Boyce-Codd-Normalform (BCNF):** Eine Relation ist in BCNF, wenn sie die Voraussetzungen der NF 3 erfüllt, und jede Determinante (Attributmenge, von der andere Attribute funktional abhängen) ein Schlüsselkandidat ist (oder die Abhängigkeit ist trivial). Ein Schlüsselkandidat ist eine Menge von Attributen, von der alle Attribute der Relation voll funktional abhängig sind.
- **NF 4:** Eine Datenbank ist dann in der 4. Normalform, wenn sie nur noch triviale mehrwertige Abhängigkeiten enthält oder die nicht-trivialen mehrwertigen Abhängigkeiten von Superschlüsseln ausgehen. Einfach gesagt: Es darf in einer Relation nicht mehrere, voneinander unabhängige,  $1:n$ - oder  $m:n$ -Beziehungen zu einem Schlüsselwert geben. ...
- **NF 5:** Eine Relation ist in 5NF, wenn sie sich nicht aus einfacheren Relationen (solche, die weniger Attribute enthalten) durch Verbundoperationen rekonstruieren lässt.

## 3 Weitere Quellen zum Thema

<https://www.youtube.com/watch?v=Hn6qyRkbPqI>

