

Datenbanken

Informatik, ICS und als Wahlfach

5. Normalformen/ Normalisierung

Prof. Dr. Markus Goldstein

SoSe 2022

5.1 Überblick Normalisierung

5.2 Erste Normalform

5.3 Zweite Normalform

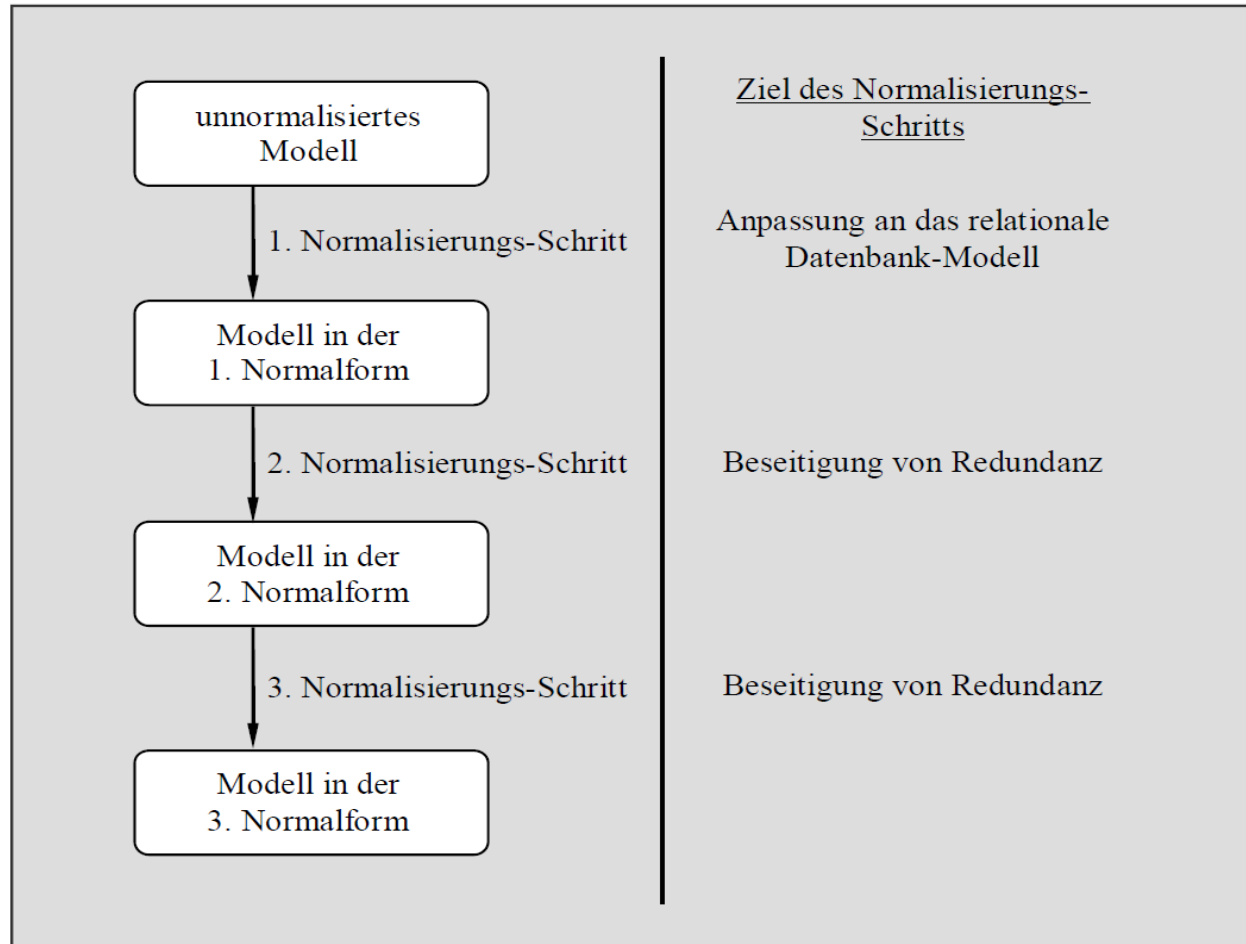
5.4 Dritte Normalform

5.5 Weitere Normalformen

- Erstellung eines konzeptionellen Schemas
- Überführung in ein relationales (logisches) Modell
- Manchmal mehrere Möglichkeiten zur Modellierung
- Wir haben schon Regeln für komplexe Eigenschaften kennengelernt → Kapitel 2.65
 - Subeigenschaft
 - Mehrere Ausstattungsmerkmale
 - Für ein „besseres Modell“
- Was ist ein **gutes Modell**?

- Gesucht: Objektive Qualitätskriterien
- Lösung: **Normalisierung**
 - Eigentlich für relationales Modell entwickelt
 - Aber auch für konzeptionelle Modelle anwendbar
- Überführung des Modells in **Normalformen**

- Erste, Zweite, und 3. Normalform (1NF, 2NF, 3NF)



5.1 Überblick Normalisierung

5.2 Erste Normalform

5.3 Zweite Normalform

5.4 Dritte Normalform

5.5 Weitere Normalformen

- Wir haben bereits komplexe Eigenschaften kennengelernt
 - Mehrere Werte gleichzeitig für ein Attribut einer Entität („Ausstattungsmerkmale“) → Problem der 1NF
 - [Subeigenschaften; später 3NF)]

- Weitere Varianten
 - Mehrwertiges Attribut
(z.B. Telefonnummer: CC, Vorwahl, Nummer)
 - Zusammengesetztes Attribut (z.B. Adresse)

Komplexe Eigenschaften - Beispiel

- Bestelldaten: 2 Artikel pro Kunde - mehrere Werte gleichzeitig
- Hier z.B. 4, 02.05.03, ... {L03,K201}, {Tee 125g, T-Shirt}
- Verletzt die erste Normalform

Bestell-		Kunden-			Artikel-			
Nr	Datum	Nr	Name	Adresse	Nr	Bezeichnung	Preis	Menge
4	02.05.03	121	Klein, Sara	Sara-Klein-Str. 1, Berlin	L03	Tee 125g	4.20	3
					K201	T-Shirt	5.00	3
5	03.05.03	12	Groß, Rolf	Siegfried-Groß-Alee 3, Berlin	K27	Schuhe	80.00	1
					K201	T-Shirt	5.00	7

Keine Relation!

Komplexe Eigenschaften - Beispiel

- Auflösung durch Wiederholung der Attribute

Bestell-		Kunden-			Artikel-			
Nr	Datum	Nr	Name	Adresse	Nr	Bezeichnung	Preis	Menge
4	02.05.03	121	Klein, Sara	Sara-Klein-Str. 1, Berlin	L03	Tee 125g	4.20	3
4	02.05.03	121	Klein, Sara	Sara-Klein-Str. 1, Berlin	K201	T-Shirt	5.00	3
5	03.05.03	12	Groß, Rolf	Siegfried-Groß-Alee 3, Berlin	K27	Schuhe	80.00	1
5	03.05.03	12	Groß, Rolf	Siegfried-Groß-Alee 3, Berlin	K201	T-Shirt	5.00	7

...

Zusammengesetzte Attribute

- Weiteres Problem: Zusammengesetzte Attribute
 - Auflösung durch Aufspaltung in eigene Attribute

Nachname	Vorname
Klein	Sara
Groß	Rolf

1. Normalform (1NF)

- Überführen der Unnormalisierten Form (UNF) in die 1NF
 - nur **atomare** Attributwerte
- Bei mehrwertigen Attribut
 - Einführung von eigenen Attributen (Spalten)
- Bei multiplen Eigenschaften
 - Einführung von „stark redundanten“ Tupeln (Zeilen)

1. Normalform (1NF)

- Weiteres Beispiel für **multiple Eigenschaften**
- **UNF**

<u>ArtNr</u>	Bezeichnung	LNr	Lieferant	...
2000	Server Xpert	L1	H. Schulze	...
3000	Trinitron 17" LCD	L1, L2	H. Schulze, Koch & Krug	...
5000	USB-Kabel	L1, L2	H. Schulze, Koch & Krug	...

- **1NF mit Erweiterung des Primärschlüssels**

<u>ArtNr</u>	Bezeichnung	<u>LNr</u>	Lieferant	...
2000	Server Xpert	L1	H. Schulze	...
3000	Trinitron 17" LCD	L1	H. Schulze	...
3000	Trinitron 17" LCD	L2	Koch & Krug	...
5000	USB-Kabel	L1	H. Schulze	...
5000	USB-Kabel	L2	Koch & Krug	...

1. Normalform (1NF) mit Redundanz

- 1NF nicht perfekt, da **Redundanzen** in der Tabelle

<u>ArtNr</u>	Bezeichnung	<u>LNr</u>	Lieferant	...
2000	Server Xpert	L1	H. Schulze	...
3000	Trinitron 17" LCD	L1	H. Schulze	...
3000	Trinitron 17" LCD	L2	Koch & Krug	...
5000	USB-Kabel	L1	H. Schulze	...
5000	USB-Kabel	L2	Koch & Krug	...

- Redundanzen aber unerwünscht, da
 - Speicherplatzverschwendung
(früher wichtig, heute nicht mehr essentiell)
 - Inkonsistenzen (Anomalien) möglich (**wichtiger Grund**)
Beispiel: Umbenennen von „USB-Kabel“
 - Vermeiden der Redundanzen in der 2NF und 3NF

- Änderungsanomalien in dieser Tabelle ...

<u>ArtNr</u>	Bezeichnung	<u>LNr</u>	Lieferant	...
2000	Server Xpert	L1	H. Schulze	...
3000	Trinitron 17" LCD	L1	H. Schulze	...
3000	Trinitron 17" LCD	L2	Koch & Krug	...
5000	USB-Kabel	L1	H. Schulze	...
5000	USB-Kabel	L2	Koch & Krug	...

- **Einfügeanomalie:** Neuer Lieferant (L3) nicht ohne mindestens einen Artikel einfügbar
- **Modifikationsanomalie:** Änderung der Artikelbezeichnung an mehreren Stellen (aka **Update-Anomalie**)
- **Löschanomalie:** Löschen von Artikeln führt ggf. zum Löschen der Lieferanten

5.1 Überblick Normalisierung

5.2 Erste Normalform

5.3 Zweite Normalform

5.4 Dritte Normalform

5.5 Weitere Normalformen

Definition **Funktionale Abhängigkeit** (engl. functional dependency):

Gegeben sind zwei Attribute A und B einer Relation R.

Umgangssprachlich:

„Wenn ich den Wert von A kenne, dann weiß ich auch eindeutig, welchen Wert B hat.“

Notation: ($A \rightarrow B$)

- B ist funktional abhängig von A
- A bestimmt/ impliziert B

Formal:

Seien $A, B \in \text{Attr}(R)$ (die Menge der Attribute von R). Das Attribut B ist funktional abhängig von A, wenn gilt:

Für alle Tupel $t_1, t_2 \in R$ gilt: $t_1[A] = t_2[A] \Rightarrow t_1[B] = t_2[B]$

($t_1[A]$ bedeutet: Der Wert des Attributs a im Tupel t_1)

Überprüfen auf funktionale Abhängigkeit

- Gibt es mindestens ein Paar von Tupeln, bei dem die Werte von A gleich sind, die von B aber unterschiedlich?
 - Ja: $a \rightarrow b$ gilt nicht
 - Nein: $a \rightarrow b$ gilt*
- * **Vorsicht:** nicht unbedingt basierend auf den aktuellen Daten. Betrachten sich auch mögliche zukünftige Daten!

Funktionale Abhängigkeit

Beispiel: Überprüfen auf funktionale Abhängigkeit

Kunden-Nr	Nachname	Vorname	Bestell-Nr
1	Klein	Sara	11
2	Groß	Rolf	22
1	Klein	Sara	33
3	Klein	Michael	44

- Es gilt nicht: Nachname \rightarrow Vorname (2 Vornamen für Klein)
- Es gilt: Vorname \rightarrow Nachname
(aber nur nach den vorliegenden Daten)
- Allgemein gilt aber Vorname \rightarrow Nachname **nicht**
(offensichtlich, da es Menschen mit gleichen Vornamen, aber unterschiedlichen Nachnamen geben kann)
- Offensichtlich: KundenNr \rightarrow Nachname gilt

Funktionale Abhängigkeit ist nicht auf einzelne Attribute beschränkt

Erweiterte Definition:

Für jede Wertekombination von A gibt es genau eine Wertekombination von B. Dann gilt funktionale Abhängigkeit

Beispiele:

- $KuNr \rightarrow \text{Vorname, Nachname}$
- $MatrNr, VLNr, \text{Prüfungsdatum} \rightarrow \text{Note}$

Triviale funktionale Abhängigkeit

- $A \rightarrow A$
- $(A, B) \rightarrow A$ bzw. $(A, B) \rightarrow B$

Funktionale Abhängigkeit

Beispiele erweiterte Definition

- (Nummer, gehört zu) → Fläche



- (ArtNr, LNR) → (Bezeichnung, Lieferant)

<u>ArtNr</u>	Bezeichnung	<u>LNR</u>	Lieferant
2000	Server Xpert	L1	H. Schulze
3000	Trinitron 17" LCD	L1	H. Schulze
3000	Trinitron 17" LCD	L2	Koch & Krug
5000	USB-Kabel	L1	H. Schulze
5000	USB-Kabel	L2	Koch & Krug

Artikel
<u>ArtNr</u>
Bezeichnung
<u>LNR</u>
Lieferant

Volle Funktionale Abhängigkeit

Volle funktionale Abhängigkeit ist:

- Gegeben: Kombination $A=(A_1, A_2, \dots, A_n)$
 A_i , $i=1, \dots, n$ sind Attribute oder Beziehungstyprichtungen

B voll funktional abhängig von A, wenn

- Werte von A_i , $i=1, \dots, n$ bekannt
- Wert von B daraus erschließbar
- **jedes** A_i , $i=1, \dots, n$ dafür notwendig
- Beispiel: $(A_1, A_2) \rightarrow B$ ist voll funktional abhängig, wenn weder $A_1 \rightarrow B$ noch $A_2 \rightarrow B$ schon „einzeln“ gilt.
- **Gegenteil: partielle funktionale Abhängigkeit**

Zweite Normalform (2NF)

Eine Relation ist in zweiter Normalform (2NF),

- wenn sie in 1NF ist, und
- wenn jedes nicht-Schlüsselattribut vom Schlüssel voll funktional abhängig ist.
D.h. es gibt kein Attribut, das von einem Teil des Schlüssels funktional abhängig ist.


Daraus folgt (Merkregel):

- Hat ein Entitätstyp einen nicht-zusammengesetzten Schlüssel (nur ein Attribut als PK), liegt die 2NF automatisch vor.

Zweite Normalform (2NF)

Beispiel

- Liegt hier die 2NF vor?



<u>Bestell Nr</u>	Bestell Datum	Kunden Nr	Kunden Name	Kunden Adresse	<u>Artikel Nr</u>	Artikel Bezeichnung	Artikel Preis	Menge
...

- Nein, da
 - (BestellNr, ArtikelNr) → ArtikelBezeichnung
keine volle funktionale Abhängigkeit (Bez. hängt nur von ArtNr ab), u.s.w.
- Überführung in die 2NF durch Auslagern der nicht voll funktional abhängigen Attribute in eigene Entitätstypen

Position:



<u>Bestell Nr</u>	<u>Artikel Nr</u>	Menge
...

Bestelldaten:



<u>Bestell Nr</u>	Bestell Datum	KundenNr	Kunden Name	Kunden Adresse
...

Artikel:



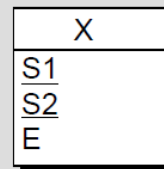
<u>Artikel Nr</u>	Artikel Bezeichnung	Artikel Preis
...

Zweite Normalform (2NF)

Allgemeine Regel für die 2NF:

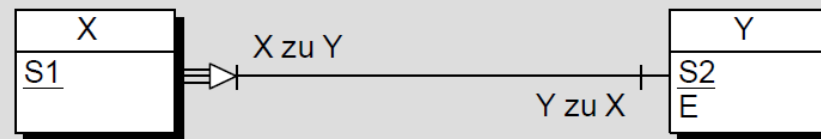
Vorgehen: Überprüfe alle Objekttypen mit einem zusammengesetzten Schlüssel daraufhin, ob eine beschreibende Eigenschaft E allein bereits von einem Teilschlüssel S2 des Gesamtschlüssels S1+S2 *funktional abhängig* ist!

Wenn ja, dann liegt folgende Situation vor:



Die Eigenschaft E ist vom Teilschlüssel S2 funktional abhängig.

Sie muss transformiert werden in:



Zweite Normalform (2NF)

Folge: Weniger Redundanz

Bestell-		Kunden-			Artikel-			
Nr	Datum	Nr	Name	Adresse	Nr	Bezeichnung	Preis	Menge
4	02.05.03	121	Klein	Sara-Klein-Str. 1	L03	Tee 125 g	4.20	3
4	02.05.03	121	Klein	Sara-Klein-Str. 1	K201	T-Shirt	5.00	3
5	03.05.03	12	Groß	Siegfried-Groß-Alee 3	K27	Schuhe	80.00	1
5	03.05.03	12	Groß	Siegfried-Groß-Alee 3	K201	T-Shirt	5.00	7
6	04.05.03	121	Klein	Sara-Klein-Str. 1	L04	Tee 250 g	7.90	1

Redundante Informationen

... wird zu (2NF):

Position:

Bestell-Nr	Artikel-Nr	Menge
4	L03	3
4	K201	3
5	K27	1
5	K201	7
6	L04	1

Bestelldaten:

Bestell-Nr	Bestell-Datum	Kunden-Nr	Kunden-Name	Kunden-Adresse
4	02.05.03	121	Klein	Sara-Klein-Str. 1
5	03.05.03	12	Groß	Siegfried-Groß-Alee 3
6	04.05.03	121	Klein	Sara-Klein-Str. 1

Artikel:

Artikel-Nr	Artikel-Bezeichnung	Artikel-Preis
L03	Tee 125 g	4.20
K27	Schuhe	80.00
K201	T-Shirt	5.00
L04	Tee 250 g	7.90

5.1 Überblick Normalisierung

5.2 Erste Normalform

5.3 Zweite Normalform

5.4 Dritte Normalform

5.5 Weitere Normalformen

Transitive Abhängigkeit

Definition Transitive Abhängigkeit

- Gegeben: Attribut A, B und C.
- C ist **transitiv abhängig** von A (über B: $A \rightarrow B$, $B \rightarrow C$), wenn
 - Wert von A ist bekannt
 - Daraus lässt sich Wert von B bestimmen
 - Aus dem Wert von B lässt sich C bestimmen

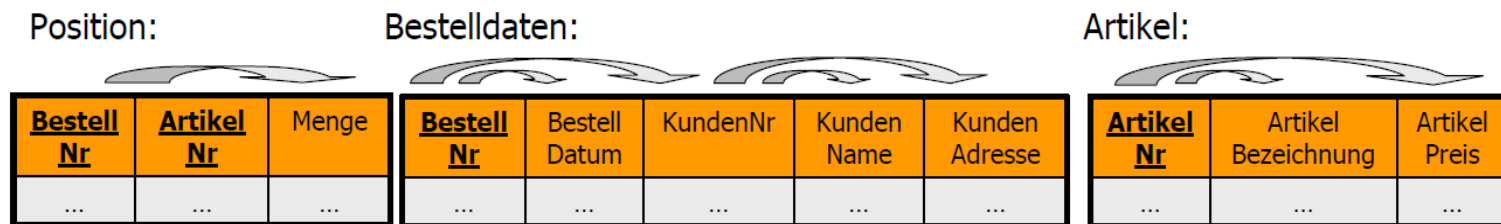
Erweiterte Definition:

- A kann auch eine *Menge* an Attributen $A=(A_1, A_2, \dots, A_n)$ sein
 A_i , $i=1, \dots, n$ Attribute oder Beziehungstyprichtungen
- B kann auch eine Menge an Attributen sein

Dritte Normalform (3NF)

Eine Relation ist in dritter Normalform (3NF),

- wenn sie in 2NF ist, und
- kein Nebenattribut (=nicht-Schlüsselattribut) transitiv funktional vom Schlüssel abhängig ist.
D.h. es gibt kein Nebenattribut, das von einem anderen Nebenattribut funktional abhängig ist.
- Beispiel – Bestelldaten verletzt 3NF



- Attribute *KundenName* und *KundenAdresse* sind nicht-Schlüsselattribute und jedes hängt funktional vom nicht-Schlüsselattribut *Kundennummer* ab

Dritte Normalform (3NF)

- Auflösung: Auslagern der Attribute, die die 3NF verletzen in neue Entitätstyp(en)
- Beispiel

Vorher:



<u>Bestell Nr</u>	Bestell Datum	KundenNr	Kunden Name	Kunden Adresse
...

Nachher:

Bestellung:



<u>Bestell Nr</u>	Bestell Datum	KundenNr
...

Kunde:



<u>Kunden Nr</u>	Kunden Name	Kunden Adresse
...

Dritte Normalform (3NF)

Folge: **Keine Redundanz mehr**

<u>Bestell-Nr</u>	Bestell-Datum	Kunden-Nr	Kunden-Name	Kunden-Adresse
4	02.05.03	121	Klein	Sara-Klein-Str. 1
5	03.05.03	12	Groß	Siegfried-Groß-Alee 3
6	04.05.03	121	<i>Klein</i>	<i>Sara-Klein-Str. 1</i>

Redundante Informationen

... wird zu (3NF):

<u>Bestell-Nr</u>	Bestell-Datum	Kunden-Nr
4	02.05.03	121
5	03.05.03	12
6	04.05.03	121

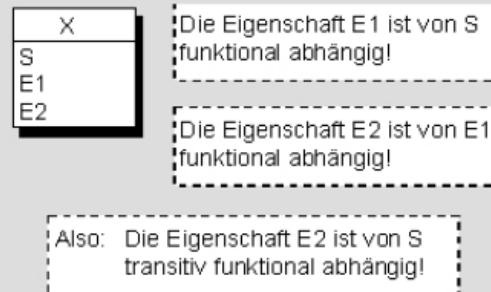
<u>Kunden-Nr</u>	Kunden-Name	Kunden-Adresse
121	Klein	Sara-Klein-Str. 1
12	Groß	Siegfried-Groß-Alee 3

Dritte Normalform (3NF)

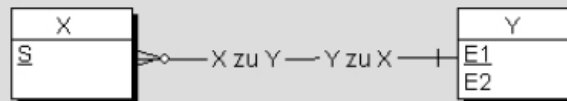
Allgemeine Regel für die 3NF:

Vorgehen: Überprüfe alle Objekttypen mit einem (elementaren oder zusammengesetzten) Schlüssel S und mindestens zwei beschreibenden Eigenschaften E1 und E2 daraufhin, ob E2 vom Schlüssel S transitiv funktional abhängig ist, ob also E2 zunächst von E1 und dann E1 von S funktional abhängen!

Wenn ja, dann liegt folgende Situation vor:



Diese muss transformiert werden in:



Dritte Normalform (3NF) ?

Zurück zum anderen Beispiel ...

<u>ArtNr</u>	Bezeichnung	<u>LNr</u>	Lieferant	...
2000	Server Xpert	L1	H. Schulze	...
3000	Trinitron 17" LCD	L1	H. Schulze	...
3000	Trinitron 17" LCD	L2	Koch & Krug	...
5000	USB-Kabel	L1	H. Schulze	...
5000	USB-Kabel	L2	Koch & Krug	...

- In welcher NF ist diese Tabelle?
- Was tun?

Dritte Normalform (3NF) ?

Zurück zum anderen Beispiel ...

Artikel

<u>ArtNr</u>	Bezeichnung	Kategorie
2000	Server Xpert	Server
3000	Trinitron 17" LCD	Monitor
5000	USB-Kabel	Zubehör
4000	USB-Maus	Zubehör

Lieferant

<u>LNr</u>	Lieferant	LTKrz	LTyp
L1	H. Schulze	HD	Händler
L2	Koch&Krug	HR	Hersteller
L3	Reisch	HR	Hersteller

Artikellieferant

<u>ArtNr</u>	<u>LNr</u>	Preis	Menge
2000	L1	9250,00	100
3000	L1	199,00	100
3000	L2	180,00	120
5000	L1	9,95	20
5000	L2	6,50	5

Und nun?

Dritte Normalform (3NF)

Zurück zum anderen Beispiel ... jetzt in 3NF

Artikel

<u>ArtNr</u>	Bezeichnung	KNr
2000	Server Xpert	1
3000	Trinitron 17" LCD	2
5000	USB-Kabel	3
4000	USB-Maus	3

Artikellieferant

<u>ArtNr</u>	<u>LNr</u>	Preis	Menge
2000	L1	9250,00	100
3000	L1	199,00	100
3000	L2	180,00	120
5000	L1	9,95	20
5000	L2	6,50	5

Kategorie

<u>KNr</u>	Kategorie
1	Server
2	Monitor
3	Zubehör

LieferantenTyp

<u>LTKrz</u>	LTyp
HD	Händler
HR	Hersteller

Lieferant

<u>LNr</u>	Lieferant	LTKrz
L1	H. Schulze	HD
L2	Koch&Krug	HR
L3	Reisch	HR

Zusammenfassung

Die 3NF ist zwangsweise nötig, um die Datenintegrität im relationalen Modell zu gewährleisten

- 1NF: Alle Attributwerte sind atomar
- 2NF: Nur bei zusammengesetzten Primärschlüsseln – nicht-Schlüsselattribute sind vom gesamten PK funktional abhängig
- 3NF: Kein nicht-Schlüsselattribut ist von einem anderen nicht-Schlüsselattribut funktional abhängig

5.1 Überblick Normalisierung

5.2 Erste Normalform

5.3 Zweite Normalform

5.4 Dritte Normalform

5.5 Weitere Normalformen

Weitere Normalformen (ohne Details)

Boyce-Codd Normalform (BCNF)

- Relation in 3NF und
Jeder Determinant ist ein Schlüsselkandidat. Ein Determinant ist eine Attributmenge, von der ein anderes Attribut vollständig funktional abhängig ist.
- Verletzung BCNF selten; setzt ebenso compound PK voraus
- „strenger“ als 3NF

4. Normalform

- In BCNF und
- Keine „mehrwertigen“ nicht-trivialen Abhängigkeiten im PK

5. Normalform

- In 4NF und gar keine „mehrwertigen“ Abhängigkeiten

Aufgaben

Bitte bearbeiten Sie jetzt die Aufgaben in Moodle zum Kapitel 5.

- **Teil A**