

Normalisierung

Foliensatz 3

DI(FH) Gerald Aistleitner, 2015/16

Normalisierung

- Sammlung von Verhaltensregeln, deren Einhaltung die Wahrscheinlichkeit von Redundanzen und damit die Wahrscheinlichkeit von Speicheranomalien reduziert
- Normalisierung betrachtet nur einzelne, isolierte Entitätsmengen und erkennt daher Redundanzen über mehrere Entitätsmengen nicht
→ nicht als vollständiges Regelwerk ansehen!
- Findet auf konzeptioneller Ebene statt
- Bei der Überleitung ins interne Modell wird tw. wieder denormalisiert (Performanceoptimierung)

Funktionale Abhängigkeit

- Darstellung: $\alpha \rightarrow \beta$
- α und β repräsentieren jeweils Mengen von Attributen (Teilmengen von R)
- Für alle Paare von Tupeln $r, t \in R$ mit $r.\alpha = t.\alpha$ muss auch gelten $r.\beta = t.\beta$
- Wenn 2 Tupel gleiche Werte für alle Attribute in α haben, dann müssen auch ihre β -Werte übereinstimmen.
- $\rightarrow \alpha$ -Werte bestimmen die β -Werte funktional bzw.
 $\rightarrow \beta$ -Werte sind funktional abhängig von α -Werten

Funktionale Abhängigkeit

- Beispiel: Prüfe auf funktionale Abhängigkeit

R				
	A	B	C	D
t	a_4	b_2	c_4	d_3
p	a_1	b_1	c_1	d_1
q	a_1	b_1	c_1	d_2
r	a_2	b_2	c_3	d_2
s	a_3	b_2	c_4	d_3

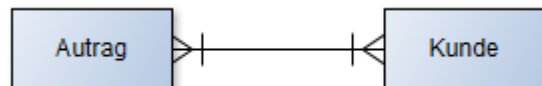
- $\{A\} \rightarrow \{B\}$?
- $\{A\} \rightarrow \{C\}$?
- $\{C,D\} \rightarrow \{B\}$?
- $\{B\} \rightarrow \{C\}$?

Ausgangsbeispiel

- Einfache Auftragsverwaltung basierend auf 2 Entitätsmengen

Auftrag				
AuftragsNr	ArtikelNr[1:10]	ArtikelBez	Betrag	Auftragsdatum
1	4, 5	HDMI-Kabel, TV	15, 550	2.1.2016
2	5, 9	TV, SAT-Receiver	550, 250	5.1.2016
3	4	HDMI-Kabel	15	7.1.2016

Kunde				
KundenNr	Name	PLZ	Ort	AuftragsNr[1:100]
1	Maier	4020	Linz	1, 2
2	Müller	5020	Salzburg	3



1. Normalform (1. NF)

- **Ein Relationstyp (Tabelle) ist in 1. Normal-form, wenn die Wertebereiche sämtlicher Attribute skalar sind.**

(ein Wertebereich ohne Wiederholungsgruppen heißt skalar oder atomar)

- Anmerkung:
Bei manchen DB-Systemen wird gezielt darauf verzichtet (Performancegründen)

2. Normalform (2. NF)

- ***Ein Relationstyp ist in 2. Normalform, wenn er in 1. Normalform ist und wenn jedes Nichtschlüsselattribut von jedem Schlüsselkandidaten voll funktional abhängig ist..***
- Kann nur verletzt sein, falls sich der Entitätsschlüssel aus mehreren Attributen zusammensetzt und die Entitätsmenge Nichtschlüsselattribute enthält.
- Lösung durch Auftrennen der Entitätsmenge

2. NF: Teilen einer Entitätsmenge

- N-Schlüsselattribute: $2^n - 1$ Kombinationen möglich
 - 1. Bestimmen von möglichen Schlüsselkombinationen
 - 2. Zuordnen von Nichtschlüsselattributen zu jener Schlüsselkombination, von welcher das Nichtschlüsselattribut abhängig ist
 - 3. Schlüsselkombinationen mit mind. 1 Nichtschlüsselattribut sind die ersten Resultatemengen.
 - 4. Schlüsselkombinationen ohne Nichtschlüsselattribute: prüfen ob weglassen ohne Informationsverlust möglich ist
 - 5. Benennung der Resultatemengen
- Kenntnis des Sachverhaltes notwendig !!

3. Normalform (3. NF)

- ***Ein Relationstyp ist in 3. Normalform, wenn er in 2. Normalform ist und kein Nichtschlüsselattribut transitiv von einem Schlüsselkandidat abhängt.***
- transitive Abhängigkeit:
Wenn Y von X funktional abhängig ist und Z von Y, so ist Z von X funktional abhängig.
- Transitiv abhängige Spalten werden in weitere Untertabellen ausgelagert, da sie nicht direkt vom Schlüsselkandidaten abhängen, sondern nur indirekt.

Boyce Codd Normalform (BCNF)

- ***Eine Relationstyp ist in BCNF, wenn jede Determinante vom Relationstyp ein Schlüsselkandidat ist.***
- Determinante:
Eine Menge von Attributen, von denen andere *voll funktional* abhängen.
- → Elimination von überlappenden Schlüsselkandidaten

4. Normalform (4. NF)

- **Ein Relationstyp befindet sich genau dann in der vierten Normalform, wenn er sich in der BCNF befindet und maximal eine nichttriviale mehrwertige Abhängigkeit enthält.**
- Mehrwertige Abhängigkeit: $X \twoheadrightarrow Y$
Wenn für 2 Attributmengen X und Y gilt, dass einem Wert von X eine Menge von Werten von Y unabhängig von den restlichen Werten zugeordnet wird, so besteht zw. X und Y eine mehrwertige Abhängigkeit.
- Eine mehrwertige Abhängigkeit einer Attributmenge Y von einer Attributmenge X ist trivial, wenn Y teil von X ist oder die Relation nur aus X und Y besteht.

5. Normalform (5. NF, Project-Join-Normalform)

- Ein Relation R ist in 5. Normalform, wenn sie in 4. NF ist und für jede Join-Abhängigkeit (R_1, R_2, \dots, R_n) gilt:
 - Die Join-Abhängigkeit ist trivial oder
 - Jedes R_i aus (R_1, R_2, \dots, R_n) ist Schlüsselkandidat
- Eine Relation R ist in 5. Normalform, wenn sie in 4. NF ist und keine mehrwertigen Abhängigkeiten enthält, die voneinander abhängig sind.

Vereinfacht:

- Ein Relationstyp ist in 5. Normalform, wenn er in der vierten Normalform ist und sich nicht ohne Informationsverlust in mehrere Tabellen aufspalten lässt.