

# Normalformen

---

## **Ziel:**

**Verbesserung der Qualität des DB-Entwurfs**

## **Vorgangsweise:**

- **Zerlegung der Relationen entsprechend den Normalformen-Regeln**
- **Normalformen sind ENTWURFSREGELN  
für den guten relationalen DB-Entwurf**

# Warum Normalformen

---

- **Qualität** eines relationalen Datenbankentwurfs bewerten
- **1. Redundanzfreiheit**
  - in mehreren Tabellen wird immer wieder ZUNAME, VORNAME, ADR, ... gehalten
- **2. Konsistenzbedingungen**
  - einhalten, die durch funktionale Abhängigkeiten gegeben sind
  - {PERSNR} --> PERSONEN-Tabelle
  - {PLZ} --> {BLAND, ORT, STRASSE}
- **3. Daten-Anomalien** vermeiden
- Normalformen garantieren obige Kriterien!!

# 'Schlechte' Relationenschemata

ProfVorl						
PersNr	Name	Rang	Raum	VorlNr	Titel	SWS
2125	Sokrates	C4	226	5041	Ethik	4
2125	Sokrates	C4	226	5049	Mäeutik	2
2125	Sokrates	C4	226	4052	Logik	4
...	...	...	...	...	...	...
2132	Popper	C3	52	5259	Der Wiener Kreis	2
2137	Kant	C4	7	4630	Die 3 Kritiken	4

Update-Anomalien

Einfügeanomalien

Löschanomalien

# Anomalien bei schlechten Relationenschemata

---

- Updateanomalien:
  - Wenn ein Professor einen anderen Raum bezieht, müssen alle Tupel geändert werden.
- Einfügeanomalie:
  - Was macht man mit Professoren, die keine Vorlesung halten?
- Löschanomalien:
  - Was passiert, wenn Kant seine einzige Vorlesung absagt ?
  - Wird er gelöscht?
- Lösung: Zerlegung der Relation in Teilrelationen

# Funktionale Abhängigkeiten

---

- Beim DB-Entwurf sind die funktionalen Abhängigkeiten (FD) zwischen Attributen bzw. Attributkombinationen sehr wichtig.
- FDs sind Integritätsbedingungen, die zu allen Zeiten in jedem DB-Zustand (=Ausprägung) eingehalten werden müssen.
- **Kenntnis und Beachtung v. FDs beim DB-Entwurf ist eine unabdingbare Voraussetzung für die Gewährleistung der Integrität einer Datenbank**
- FDs sind eine Verallgemeinerung des Schlüsselbegriffs

# Funktionale Abhängigkeiten

---

- Schema
  - $R = \{A, B, C, D\}$
- Ausprägung  $R$
- Seien  $\alpha \subseteq R$ ,  $\beta \subseteq R$
- $\alpha \rightarrow \beta$  genau dann wenn  $\forall r, s \in R$  mit  $r.\alpha = s.\alpha \Rightarrow r.\beta = s.\beta$

$R$			
$A$	$B$	$C$	$D$
a4	b2	c4	d3
a1	b1	c1	d1
a1	b1	c1	d2
a2	b2	c3	d2
a3	b2	c4	d3

$$\{A\} \rightarrow \{B\}$$

$$\{C, D\} \rightarrow \{B\}$$

$$\text{Nicht: } \{B\} \rightarrow \{C\}$$

Notationskonvention:

$$CD \rightarrow B$$

## Beispiel: Funktionale Abhängigkeiten

Stammbaum				
Kind	Vater	Mutter	Opa	Oma
Sofie	Alfons	Sabine	Lothar	Linde
Sofie	Alfons	Sabine	Hubert	Lisa
Niklas	Alfons	Sabine	Lothar	Linde
Niklas	Alfons	Sabine	Hubert	Lisa
...	...	...	Lothar	Martha
...	...	...	...	...

- Kind  $\rightarrow$  Vater,Mutter
- Frage: Welche 2 weiteren Fds gibt es
- Kind,Opa  $\rightarrow$  Oma
- Kind,Oma  $\rightarrow$  Opa

# Funktionale Abhängigkeiten

- Übliche Sprechweisen:
  - $\alpha \twoheadrightarrow \beta$  ,d.h. wenn  $\alpha$  bekannt ist, kennt man auch  $\beta$
  - $\alpha$  bestimmt  $\beta$  oder  $\beta$  hängt von  $\alpha$  ab
- Alle Tupel, die in  $\alpha$  den selben Wert aufweisen, müssen auch in  $\beta$  übereinstimmen.
- Überprüfe, ob die angegeb. FDs (a-d) beim Erstellen der Relation eingehalten wurden A B C D E F  
a)  $AB \rightarrow D$  b)  $C \twoheadrightarrow E$   
c)  $C \twoheadrightarrow F$  d)  $ABC \twoheadrightarrow E$ 
  - a e 3 g 5 p
  - a c 4 h 1 g
  - a b 3 f 5 p
  - a e 2 g 4 g
  - a k 3 f 7 p
  - a k 2 f 4 g



# Schlüssel

---

- $\alpha \subseteq \mathcal{R}$  ist ein *Superschlüssel* wenn gilt:

$$\alpha \rightarrow \mathcal{R}$$

- $\beta$  ist *voll funktional abhängig* von  $\alpha$  – in Zeichen  $\alpha \xrightarrow{\bullet} \beta$  – falls beide nachfolgenden Kriterien gelten:
  1.  $\alpha \rightarrow \beta$ , d.h.  $\beta$  ist funktional abhängig von  $\alpha$  und
  2.  $\alpha$  kann nicht mehr „verkleinert“ werden, d.h.

$$\forall A \in \alpha : \alpha - \{A\} \not\xrightarrow{\bullet} \beta$$

- $\alpha \subseteq \mathcal{R}$  ist ein *Kandidatenschlüssel* wenn gilt:

$$\alpha \xrightarrow{\bullet} \mathcal{R}$$

Städte			
Name	BLand	Vorwahl	EW
Frankfurt	Hessen	069	650000
Frankfurt	Brandenburg	0335	84000
München	Bayern	089	1200000
Passau	Bayern	0851	50000
...	...	...	...

Die Kandidatenschlüssel für die Relation *Städte* sind:

- {Name, BLand}
- {Name, Vorwahl}

Man beachte, daß zwei (kleinere) Städte dieselbe Vorwahl haben können.

# Transitive Abhängigkeit

- S Sei der Identifikationsschlüssel einer Relation R. B und C seien zwei weitere Attribute oder Attributskombinationen von R derart, daß die drei Attribute (Attributskombinationen) untereinander je distinkt sind. C ist **transitiv abhängig** von S falls jederzeit gilt:
  - $R.S \multimap R.B$       und    $R.B \multimap R.C$
  - $R.B \not\multimap R.S$
- PERSONAL (PersNr, Name, AbtNr, AbtBez)
  - $PersNr \rightarrow PERSONAL$  (insbes.:  $PersNr \rightarrow AbtNr$ )
  - $AbtNr \rightarrow AbtBez$
  - *AbtBez ist transitiv abhängig von PersNr*
- "Transitiv abhängig" ist also gleichbedeutend mit "abhängig auch über schlüsselfremde Umwege"

# 1. NF

---

- Eine Relation befindet sich in der 1. Normalform, wenn keines ihrer Attribute Attributwerte aufweist, die ihrerseits Mengen sind.

Personal:

<u>PID</u>	Person	AbtID	Abteilung	ProjID	Projekt	ProjZeit
1	Hans	1	Physik	11,12	A, B	60,40
2	Rolf	2	Chemie	13	C	100
3	Ursula	2	Chemie	11,12,13	A, B, C	20,50,30
4	Paul	1	Physik	11,13	A, C	80,20

Person\_Projekt:

<u>PID</u>	Person	AbtID	Abteilung	<u>ProjID</u>	Projekt	ProjZeit
1	Hans	1	Physik	11	A	60
1	Hans	1	Physik	12	B	40
2	Rolf	2	Chemie	13	C	100
3	Ursula	2	Chemie	11	A	20
3	Ursula	2	Chemie	12	B	50
3	Ursula	2	Chemie	13	C	30
4	Paul	1	Physik	11	A	80
4	Paul	1	Physik	13	C	20

## 2. NF (1)

---

Eine Relation  $\mathcal{R}$  mit zugehörigen FDs  $F$  ist in zweiter Normalform, falls jedes Nichtschlüssel-Attribut  $A \in \mathcal{R}$  voll funktional abhängig ist von jedem Kandidatenschlüssel der Relation.

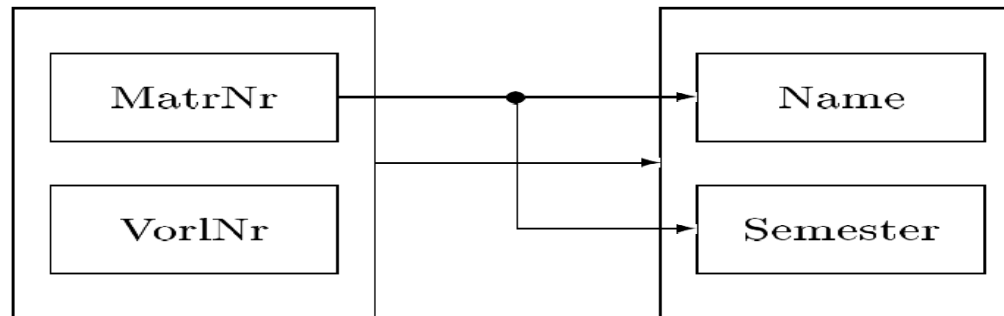
StudentenBelegung			
MatrNr	VorlNr	Name	Semester
26120	5001	Fichte	10
27550	5001	Schopenhauer	6
27550	4052	Schopenhauer	6
28106	5041	Carnap	3
28106	5052	Carnap	3
28106	5216	Carnap	3
28106	5259	Carnap	3
...	...	...	...

*Studentenbelegung* ist nicht 2 NF wegen

- $\{\text{MatrNr}\} \rightarrow \{\text{Name}\}$  und
- $\{\text{MatrNr}\} \rightarrow \{\text{Semester}\}$

## 2. NF (2)

---



- Einfügeanomalie: Was macht man mit Studenten, die keine Vorlesungen hören?
- Updateanomalien: Wenn z.B. „Carnap“ ins vierte Semester kommt, muß sichergestellt werden, daß alle vier Tupel geändert werden.
- Löschanomalien: Was passiert, wenn „Fichte“ ihre einzige Vorlesung absagt?

Zerlegung in:

- hören: {[MatrNr, VorlNr]} und
- Studenten: {[MatrNr, Name, Semester]}

Beide Relationen sind 2 NF (erfüllen sogar noch „höhere Gütekriterien“)

## 2. NF (3)

---

Projekt:

<u>ProjID</u>	Projekt
11	A
12	B
13	C

Personal:

<u>PID</u>	Person	AbtID	Abteilung
1	Hans	1	Physik
2	Rolf	2	Chemie
3	Ursula	2	Chemie
4	Paul	1	Physik

Projektzugehörigkeit:

<u>PID</u>	<u>ProjID</u>	ProjZeit
1	11	60
1	12	40
2	13	100
3	11	20
3	12	50
3	13	30
4	11	80
4	13	20

### 3. NF (1)

---

- Eine Relation befindet sich in der 3. Normalform, wenn sie sich in der 2. Normalform befindet und kein Attribut, das nicht zum Identifikationsschlüssel gehört, transitiv von diesem abhängt.
- Eine Relation befindet sich dann und nur dann in der 3. Normalform, wenn sie sich in der 2. Normalform befindet und kein NSA von einem anderen NSA funktional abhängig ist.
- Eine Relation befindet sich NICHT in der 3. Normalform, wenn ein NSA von einem anderen NSA funktional abhängig ist.
  - Bsp: {**PERSNR**, NAME, ABTNR, ABTNAME}
    - Weil ABTNR --> ABTNAME und beide sind NSA => nicht in 3.NF

## 3.NF (2)

---

Personal:

<u>PID</u>	Person	AbtID
1	Hans	1
2	Rolf	2
3	Ursula	2
4	Paul	1

Abteilung:

<u>AbtID</u>	Abteilung
1	Physik
2	Chemie

Projekt:

<u>ProjID</u>	Projekt
11	A
12	B
13	C

Projektzugehörigkeit:

<u>PID</u>	<u>ProjID</u>	ProjZeit
1	11	60
1	12	40
2	13	100
3	11	20
3	12	50
3	13	30
4	11	80
4	13	20



## Übung: CD\_Lieder

---

- [http://de.wikipedia.org/wiki/Normalisierung\\_\(Datenbank\)](http://de.wikipedia.org/wiki/Normalisierung_(Datenbank))

CD Lieder		
<i>CD_ID</i>	Album	Titelliste
4811	Anastacia - Not That Kind	{1. Not That Kind, 2. I'm Outta Love, 3. Cowboys & Kisses}
4713	Pink Floyd - Wish You Were Here	{1. Shine On You Crazy Diamond}

- **Verletzung der 1NF**
- Das Feld **Album** beinhaltet die Attributwertebereiche Interpret und Albumtitel.
- Das Feld **Titelliste** enthält eine Menge von Titeln.

## Übung: CD\_Lieder

---

- Frage: Ist CD\_Lieder in 2. NF?

CD_Lieder				
<i>CD_ID</i>	<i>Album</i>	<i>Interpret</i>	<i>Track</i>	<i>Titel</i>
4811	Not That Kind	Anastacia	1	Not That Kind
4811	Not That Kind	Anastacia	2	I'm Outta Love
4811	Not That Kind	Anastacia	3	Cowboys & Kisses
4712	Wish You Were Here	Pink Floyd	1	Shine On You Crazy Diamond

## Übung: CD\_Lieder

---

- Lösung: CD\_Lieder ist nun in 2.NF

CD			Lieder		
<i>CD_ID</i>	Album	Interpret	<i>CD_ID</i>	<i>Track</i>	Titel
4811	Not That Kind	Anastacia	4811	1	Not That Kind
4712	Wish You Were Here	Pink Floyd	4811	2	I'm Outta Love
			4811	3	Cowboys & Kisses
			4712	1	Shine On You Crazy Diamond

## Übung: CD\_Lieder

---

- Annahme: CD habe folgendes Aussehen:
- Frage: Ist CD in 3.NF?

<i><b>CD_ID</b></i>	<b>Album</b>	<b>Interpret</b>	CD
			<b>Gründungsjah r</b>
4811	Not That Kind	Anastacia	1999
4713	Freak of Nature	Anastacia	1999
4712	Wish You Were Here	Pink Floyd	1965

## Übung: CD\_Lieder

---

- Die Zerlegung entspricht nun der 3. NF, d.h.
  - Redundanzfrei, keine evtl. Insert/Update/Delete Anomalien

CD		CD_Künstler		Künstler	
<i>CD_ID</i>	Album	<i>CD_ID</i>	<i>I_ID</i>	<i>I_ID</i>	Interpret Gründungsja h
4811	Not That Kind	4811	2423	2423	Anastacia 1999
4713	Freak of Nature	4713	2423	3433	Pink Floyd 1965
4712	Wish You Were Here	4712	3433		

# Übung:

---

- 1. Normalisieren Sie die folgende , nicht normalisierte Tabelle bis zur dritten Normalform.

<u>TnNr</u>	<u>TnName</u>	<u>FirNr</u>	<u>FirmenName</u>	<u>KursTyp</u>	<u>KursTypName</u>	<u>Ort</u>	<u>Datum</u>
1962	Antkowiak	56	Helm AG	K1060	Systemmodelle	D	15.06.99
1958	Sieger	56	Helm AG	K1500	Forms	D	03.10.99
5324	Schuster	87	Schuster GbR	K1122	Reports	HH	06.09.99
8231	Hauser	102	Bauer KG	K1060	Systemmodelle	B	10.11.99
9243	Scherbaum	87	Schuster GbR	K1122	Reports	HH	06.09.99
2834	Adam	25	Eva Kosmetik GmbH	K1500	Forms	D	03.10.99
2936	Meister	142	Stahlbau Bronk	K1060	Systemmodelle	S	10.11.99
6352	Hinrichs	102	Bauer KG	K1070	PL/SQL	M	10.09.99
1962	Antkowiak	56	Helm AG	K1000	SQL	D	11.11.99
2932	Bond	7	British Goods Inc.	K3308	Admin Oracle8	S	13.12.99