

# DATABASE



## DHBW Stuttgart

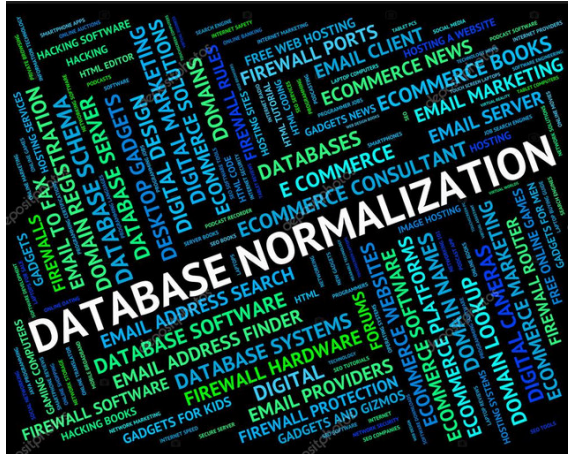
### Datenbanken I

### Kapitel 4 – Normalisierung

Modul: T3INF2004

## **Nutzungshinweis:**

**Diese Unterlagen dürfen ausschließlich von Mitgliedern  
(das sind Studierende, Bedienstete)  
der Dualen Hochschule Baden-Württemberg Stuttgart eingesetzt werden.  
Eine Weitergabe an andere Personen oder Institutionen ist untersagt.**



- Das Verfahren der Normalisierung, wurde von Codd (1972) vorgeschlagen
- War ursprünglich ein Entwurfsverfahren, wie die ER-Methode
- Normalisierung wird nur als Analyseverfahren verwendet
- „Gute“ Schemata sollen erzeugt werden
- Mängel im Datenbankentwurf erkenn- und korrigierbar

## Projekt\_Projektleiter

<u>Proj-Nr</u>	Bezeichnung	Beginn	Pers-Nr	Vorname	Nachname
4711	Fahrzeugversuchssystem für Firma WMB	15.3.2015	2	Rita	Schulze
3050	Erweiterung Personal-Datenbank Firma Kleinert	13.5.2018	3	Werner	Maier
2020	Schnittstellen zwischen Produktion und Verkauf erstellen	1.2.2018	2	Rita	Schulze
1234	Erweiterung interne Datenbank für unser Softwarehaus	1.4.2017	1	Hans	Müller
3091	Elektronische Erfassung der Prüfstandsdaten für Firma WMB	1.9.2018	3	Werner	Maier

## Einfügeanomalie

- Neuer PL „Schmidt“ (ohne Projekt) soll eingefügt werden, was passiert?

1010	NULL	NULL	5	Schmidt	Karl
------	------	------	---	---------	------

- Was passiert, wenn nun ein neues Projekt mit Herrn Schmidt angelegt wird?

## Projekt\_Projektleiter

<u>Proj-Nr</u>	Bezeichnung	Beginn	Pers-Nr	Vorname	Nachname
4711	Fahrzeugversuchssystem für Firma WMB	15.3.2015	2	Rita	Schulze
3050	Erweiterung Personal-Datenbank Firma Kleinert	13.5.2018	3	Werner	Maier
2020	Schnittstellen zwischen Produktion und Verkauf erstellen	1.2.2018	2	Rita	Schulze
1234	Erweiterung interne Datenbank für unser Softwarehaus	1.4.2017	1	Hans	Müller
3091	Elektronische Erfassung der Prüfstandsdaten für Firma WMB	1.9.2018	3	Werner	Maier

## Löschanomalie

- Projekt 1234 wird vom Kunden zurückgezogen und gelöscht, was passiert?

Der Projektleiter Müller .....

1. Alle Entitäten werden Relationen
2. Alle Beziehungen werden Relationen
3. Bei 1:N und N:1 Beziehungs-Relationen können deren Attribute mit der N-Relation zusammengezogen werden.  
Der Primärschlüssel (der 1-Relation) wird dann zum Fremdschlüssel
4. Bei 1:1 Relationen erfolgt die Zusammenfassung so, dass möglichst wenig NULL-Werte entstehen
5. Aus M:N Beziehungen werden eigenständige (Beziehungs-) Relationen erstellt

1. Alle Entitäten werden Relationen
2. Alle Beziehungen werden Relationen
3. Bei 1:N und N:1 Beziehungs-Relationen können deren Attribute mit der N-Relation zusammengezogen werden.  
Der Primärschlüssel (der 1-Relation) wird dann zum Fremdschlüssel
4. Bei 1:1 Relationen erfolgt die Zusammenfassung so, dass möglichst wenig NULL-Werte entstehen
5. Aus M:N Beziehungen werden eigenständige (Beziehungs-) Relationen erstellt

## Projekt\_Projektleiter

<u>Proj-Nr</u>	Bezeichnung	Beginn	Pers-Nr	Vorname	Nachname
4711	Fahrzeugversuchssystem für Firma WMB	15.3.2015	2	Rita	Schulze
3050	Erweiterung Personal-Datenbank Firma Kleinert	13.5.2018	3	Werner	Maier
2020	Schnittstellen zwischen Produktion und Verkauf erstellen	1.2.2018	2	Rita	Schulze
1234	Erweiterung interne Datenbank für unser Softwarehaus	1.4.2017	1	Hans	Müller
3091	Elektronische Erfassung der Prüfstandsdaten für Firma WMB	1.9.2018	3	Werner	Maier

## Änderungsanomalie

➤ Für Herrn Maier, der die Firma verlässt, kommt Herr Kunze, was passiert?

Je mehr Redundanzen, desto größer ist die Wahrscheinlichkeit für

.....

Schlecht designte Relationen führen zu ..... und dies sollte vermieden werden



# Die Projektion ( $\Pi$ )

- Mit der Projektion ( $\Pi$ ) erhalten wir die Attribute einer Relation.

$\Pi_{\text{Vorname, Nachname}}(\text{Projekt\_Projektleiter})$

Vorname	Nachname
Hans	Müller
Rita	Schulze
Werner	Maier

Formal:  
..... [  
 $R' \subseteq R$ ]

## Definiton:

Funktionale Abhängigkeit (FA; englisch functional dependency, FD) sind Konzepte der relationalen Entwurfstheorie und bilden die Grundlage für die Normalisierung von Relationenschemata. (Wikipedia)

## Formal:

Es seien zwei Attributmengen (Teilmengen)  $X$  und  $Y$  gegeben mit  $X, Y \subseteq [R]$   
 $X \longrightarrow Y$  heißt funktional abhängig (FA) wenn gilt:  
es existiert eine Funktion

.....

für  $x \in \Pi_X(R)$ ,  $y \in \Pi_Y(R)$  für alle möglichen Instanzen von  $R$ .

Die Werte von ..... die Werte von ...

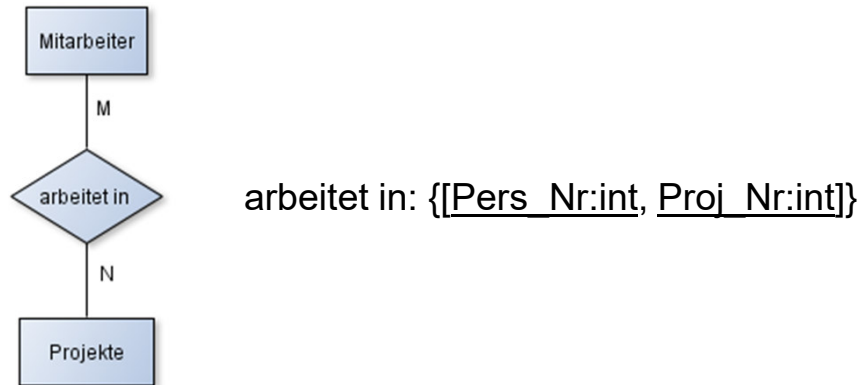
## Triviale FA

Ein Attribut ist immer funktional abhängig:

- von sich selbst
- und von jeder Obermenge von sich selbst

Dies bezeichnet man als triviale FA oder trivial.

Triviale funktionale Abhängigkeiten ..... immer erfüllt.



$\{\text{Pers\_Nr}, \text{Proj\_Nr}\} \rightarrow \{\text{Pers\_Nr}, \text{Proj\_Nr}\}$  .....

$\{\text{Pers\_Nr}\} \rightarrow \{\text{Proj\_Nr}\}$

$\{\text{Proj\_Nr}\} \rightarrow \{\text{Pers\_Nr}\}$

# Beispiel für funktionale Abhängigkeit

<u>Pers-Nr</u>	Vorname	Nachname	Geburtsdatum
1	Hans	Müller	13.5.1989
2	Rita	Schulze	2.2.1974
3	Werner	Maier	1.7.1990
4	Karin	Schwarz	05.12.1988

**Welche funktionale Abhängigkeiten können bestimmt werden?**

{Pers-Nr}      {Vorname, Nachname, Geburtsdatum} .....

{Nachname}    {Vorname}

{Geburtsdatum}    {Nachname, Vorname, Pers-Nr}

Die funktionalen Abhängigkeiten müssen für alle ..... (R)  
gelten und nicht nur für einige ..... Instanzen

- Funktionale Abhängigkeiten können mit Hilfe der Interferenzregeln noch erweitert werden

Mitarbeiter: [{PersNr,Name,GebDatum,AusweisNr,TelNr,Raum,Straße,PLZ,Ort}]

Welche funktionale Abhängigkeiten bestehen?

$\{PersNr\} \rightarrow \{Name, GebDatum, AusweisNr, TelNr, Raum, Straße, PLZ, Ort\}$

$\{Straße, Ort\} \rightarrow \{.....\}$

$\{TelNr\} \{Vorname, Nachname, Raum\}$

$\{AusweisNr\} \rightarrow \{.....\}$

- Die Menge aller zu einem Schema  $[R]$  definierten funktionalen Abhängigkeiten bezeichnen wir als **F**.
- Diese Menge aller möglichen herleitbaren funktionalen Abhängigkeiten einer Relation nennt man **.....**
- Diese wird dann als **.....** der Menge **F** bezeichnet

- Die Hülle ( $F^+$ ) kann durch die Anwendung der Interferenzregeln von Armstrong hergeleitet werden

(1)	$Y \in X$	$\Rightarrow$	$X \rightarrow Y$	Reflexivität
(2)	$X \rightarrow Y$	$\Rightarrow$	$XZ \rightarrow YZ$	Verstärkung
(3)	$X \rightarrow Y, Y \rightarrow Z$	$\Rightarrow$	$X \rightarrow Z$	Transitivität
(4)	$X \rightarrow YZ$	$\Rightarrow$	$X \rightarrow Y, X \rightarrow Z$	Zerlegung (Dekomposition)
(5)	$X \rightarrow Y, X \rightarrow Z$	$\Rightarrow$	$X \rightarrow YZ$	Vereinigung
(6)	$X \rightarrow Y, WY \rightarrow Z$	$\Rightarrow$	$WX \rightarrow Z$	Pseudotransitivität

## Formal:

### Superschlüssel:

Jede Teilmenge  $S_i \subseteq [R]$  für die gilt:

$S_i \rightarrow [R]$

$[R] \rightarrow [R]$

➤ Dieser muss nicht ..... sein.

➤ Es **kann** gelten:.....

Schlüsselkandidaten  $\subseteq$  .....

Es kann mehrere Schlüsselkandidaten ( $k_j$ ) geben.

.....

oder:

.....

### Primärschlüssel

Ist ein (willkürlich) ausgewählter Schlüsselkandidat.



Ein Attribut wird Primattribut oder „prim“ genannt, wenn es in irgendeinem Kandidatenschlüssel von  $[R]$  vorkommt.

Formal:

Alle Attribute in  $\bigcup_j K_j$  heißen prim

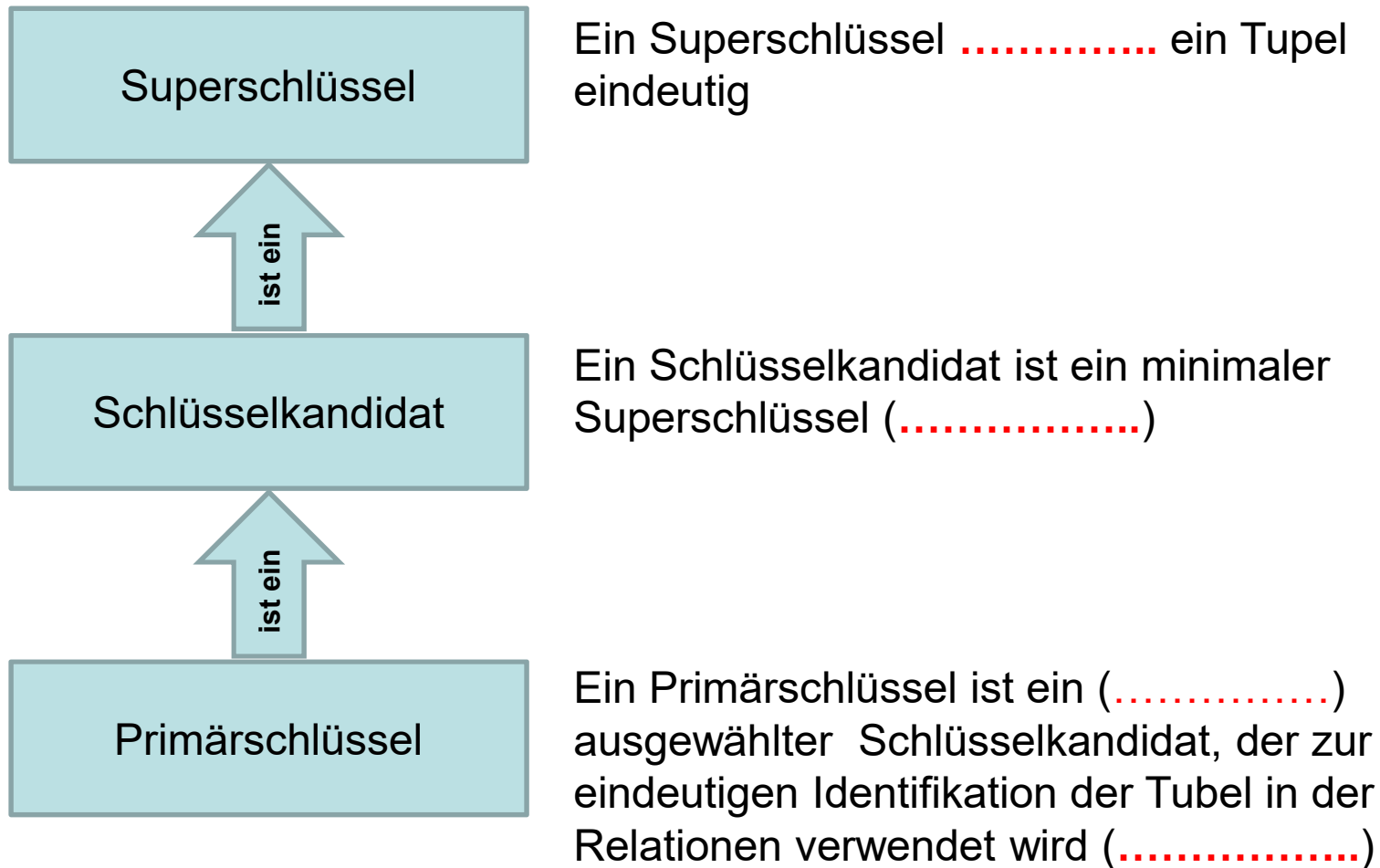
KFZ_Kennz	FIN	Marke	Typ	Leistung
ES-NT 2030	WVWZZZ1JZ3W386752	Volkswagen	Golf	85
BN-VR 22	WBABE7325VA36703	BMW	3-Reihe	115
HH-TT 1234	WDD1690071J236589	Mercedes	A-Klasse	180
B-XY 234	W0L0JBF19X1123456	Opel	Vectra B	115

Welche funktionalen Abhängigkeiten gibt es?

$\{\text{FIN, KFZ-Kennz}\} \rightarrow \{\text{Marke, Typ, Leistung}\}$  .....

$\{\text{FIN}\} \rightarrow \{\text{KFZ-Kennz, Marke, Typ, Leistung}\}$  .....

$\{\text{KFZ-Kennz}\} \rightarrow \{\text{.....}\}$  .....



- Prüfen, ob eine Relation bestimmten Anforderungen genügt.
- Falls nicht, muss transformiert werden, bis die Normalform erfüllt ist.
- Normalformen bauen aufeinander auf.

# 1. Normalform

**Def:**

Eine Relation erfüllt die erste Normalform wenn jede Entität für jedes Attribut der Relation nur einen Datenwert besitzt.

<u>Pers-Nr</u>	Vorname	Nachname	Abschlüsse	Abt_Nr	Abt_Bez	Abt_Leiter
1	Hans	Müller	Dipl. Informatiker (FH) 2000	3	Service	Lehmann
2	Rita	Schulze	Großhandelskauffrau 2000, Dipl. Wirtschaftsinformatik 2004	3	Service	Lehmann
3	Werner	Maier	Fachinformatiker 2000, Kaufmann 2002, Bachelor of Science Informatik 2006	2	Programmieren 1	Langer
4	Karin	Schwarz	Kauffrau für Bürokommunikation 2001, Dipl. Wirtschaftsinformatik 2005	1	Programmieren 2	Hausmann

**Lösung:** Alle mehrwertigen Attribute in eine neue .....

## 2. Normalform

**Def:**

Eine Relation erfüllt die zweite Normalform (2NF), wenn diese sich in der ersten Normalform befindet und alle Nichtschlüsselattribute nur durch den ..... festgelegt werden.

- Alle Nichtschlüsselattribute müssen ..... vom gesamten Schlüssel abhängig sein.

<u>Pers-Nr</u>	Vorname	Nachname	<u>Projekt-Nr</u>	Bezeichnung	Geleistete_Std
1	Hans	Müller	4711	Fahrzeugversuchssystem für Firma WMB	125
2	Rita	Schulze	3050	Erweiterung Personal-Datenbank Firma Kleinert	206
2	Rita	Schulze	2020	Schnittstellen zwischen Produktion und Verkauf erstellen	110
3	Werner	Maier	1234	Erweiterung interne Datenbank für unser Softwarehaus	154
3	Werner	Maier	2020	Schnittstellen zwischen Produktion und Verkauf erstellen	144
4	Karin	Schwarz	1234	Erweiterung interne Datenbank für unser Softwarehaus	154

## 2. Normalform

- Nichtschlüsselattribute sind?  
*Vorname, Nachname, Bezeichnung, Geleistete\_Std*  
....dürfen nur vom gesamten Primärschlüssel abhängig sein  
(vollfunktional abhängig)
- Welches Nichtschlüsselattribut erfüllt diese Bedingung?  
*Geleistete\_Std*

**To Do:** Zwei neue Relationen bilden

**Hinweis:** Alle Relationen, bei denen der Primärschlüssel nur aus einem Attribut besteht, sind automatisch in der 2.NF.



## 2. Normalform

- Nichtschlüsselattribute sind?  
....., *Nachname*, *Bezeichnung*, .....  
....dürfen nur vom gesamten Primärschlüssel abhängig sein  
(vollfunktional abhängig)
- Welches Nichtschlüsselattribut erfüllt diese Bedingung?  
.....

**To Do:**..... *Relationen bilden*

**Hinweis:** Alle Relationen, bei denen der Primärschlüssel nur aus einem Attribut besteht, sind automatisch in der 2.NF.



# 3. Normalform

## Def:

Eine Relation ist in der 3.Normalform (3NF), wenn diese die 2. Normalform erfüllt und .....zwischen einem Nichtschlüsselattribut und einem Schlüsselkandidaten besteht.

## Formal

- Ein Relationsschema [R] ist in 3NF, wenn für jede für [R] geltende funktionale Abhängigkeit der Form

$$X \rightarrow \alpha, \quad X \subseteq [R] \text{ und } \alpha \in [R]$$

mindestens eine der drei Bedingungen gilt:

$\alpha \in X$ , ( dann ist die FA ..... )

- $\alpha$  ist in einem ..... von [R] enthalten  $\Rightarrow \alpha$  ist „prim“
- X ist ein ..... von [R]

# 3. Normalform

<u>Proj-Nr</u>	Bezeichnung	Beginn	Pers-Nr	Vorname	Nachname
4711	Fahrzeugversuchssystem für Firma WMB	15.3.2015	2	Rita	Schulze
3050	Erweiterung Personal-Datenbank Firma Kleinert	13.5.2018	3	Werner	Maier
2020	Schnittstellen zwischen Produktion und Verkauf erstellen	1.2.2018	2	Rita	Schulze
1234	Erweiterung interne Datenbank für unser Softwarehaus	1.4.2017	1	Hans	Müller
3091	Elektronische Erfassung der Prüfstandsdaten für Firma WMB	1.9.2018	3	Werner	Maier

1. Prüfen, ob eine Relationen in 1NF und 2NF ist
2. .... Aufstellen
3. Schlüsselkandidaten definieren
4. Auf Definition ..... prüfen
5. Wenn nicht in 3NF, dann zerlegen

# 3. Normalform

Formale Prüfung auf 3NF:

Schlüsselkandidat = ..... (Primärschlüssel)

FA's aufstellen:

$Projekt\_Nr \rightarrow \{Bezeichnung, ....., \}$

$..... \rightarrow \{Vorname, Nachname\}$

$Projekt\_Nr \rightarrow \{Pers\_Nr, Vorname, Nachname\}$

1. Zerlege die FA's in einzelne Abhängigkeiten (Dekomposition).

2. Prüfe ob gilt:

- X ist ein Superschlüssel von [R]

**oder**

- $\alpha$  ist prim

➡ die Relation ist in 3NF



**Def:**

Eine Relation ist in Boyce-Codd Normalform, wenn jede Determinante ein ..... ist.

**Erklärung**

Eine Determinante ist eine Attributmenge (linke Seite), von der ein anderes Attribut vollständig funktional abhängig ist.

**Formal**

Ein Relationsschema  $[R]$  ist in Boyce-Codd NF, wenn für jede für  $[R]$  geltende funktionale Abhängigkeit der Form

.....  $X \sqsubseteq [R]$  und  $\alpha \in [R]$

mindestens eine der beiden Bedingungen gilt:

- $\alpha \in X$ ,
- $X$  ist ein ..... von  $[R]$
- $\alpha$  ist in einem Schlüsselkandidaten von  $[R]$  enthalten  $\Rightarrow \alpha$  ist „prim“

# Boyce-Codd Normalform

Projekt-Nr	Pers-Nr	PL-Name	Projekt-Std
4711	1	Müller	125
3050	2	Schulze	206
2020	2	Schulze	110
1234	3	Maier	154
2020	3	Maier	144
1234	4	Schwarz	154

$\{Pers - Nr, \dots\dots\dots\} \rightarrow \{Projekt - Std\}$

$\{Pers - Nr\} \rightarrow \{\dots\dots\dots\}$

$\{\dots\dots\dots\} \rightarrow \{Pers - Nr\}$

➤ 3NF ist  $\dots\dots\dots$ , *Pers-Nr* und *PL-Name* sind „ $\dots\dots\dots$ “ (Teil eines Kandidaten-schlüssels)

➤ *Pers-Nr* und *PL-Name* sind aber kein Superschlüssel  
=> Relation nicht in  $\dots\dots\dots$

➤ To Do: FA auslagern welche die BCNF verletzt

*R1*: {  $\dots\dots\dots$ ,  $\dots\dots\dots$  }

*R2* : { *Pers-Nr*, *Projekt-Nr*,  $\dots\dots\dots$  }

# Von 3NF nach BCNF und zurück

- Eine Relation  $R$  ist in BCNF, wenn diese keine überlappenden Schlüsselkandidaten hat.
- Umgekehrt gilt: wenn  $R$  in 3NF aber nicht in BCNF  
    ➔  $R$  hat zwei sich überlappende Schlüsselkandidaten

.....

Gegeben ist:

- eine Relation  $R$  mit den Attributen  $a, b, c, d$
- die FA's

$$\{a, b\} \rightarrow \{d\}$$

$$\{b\} \rightarrow \{c\}$$

$$\{c\} \rightarrow \{b\}$$

**Aufgabe:** Prüfen Sie in welcher NF sich die Relation befindet.

Die Relation ist in 3NF, aber nicht in ....., da  $\{c\}$  und  $\{d\}$  keine ..... sind.

[R] zerlegt in [R1] und [R2]

dann gilt:  $[R] = [R1] \cup [R2]$

...man spricht dann von einer .....**Zerlegung**

## Ausprägung der Zerlegung

$R1 = \dots (R)$

$R2 = \dots (R)$

## Verbundtreue oder verlustlose Zerlegung

Wenn gilt: ..... = R (für alle Ausprägungen von R)

## Formale Prüfung auf verlustlos

$[R1] \cap [R2] \rightarrow [R1] \in F^+$

oder

$[R1] \cap [R2] \rightarrow [R2] \in F^+$

...existiert für einer der beiden Teilschemata eine .....,  
so ist die Verbundtreue erfüllt.

# Verlustlose Zerlegung

<u>Proj-Nr</u>	Bezeichnung	Beginn	Pers-Nr	Vorname	Nachname
4711	Fahrzeugversuchssystem für Firma WMB	15.3.2015	2	Rita	Schulze
3050	Erweiterung Personal-Datenbank Firma Kleinert	13.5.2018	3	Werner	Maier
2020	Schnittstellen zwischen Produktion und Verkauf erstellen	1.2.2018	2	Rita	Schulze
1234	Erweiterung interne Datenbank für unser Softwarehaus	1.4.2017	1	Hans	Müller
3091	Elektronische Erfassung der Prüfstandsdaten für Firma WMB	1.9.2018	3	Werner	Maier

..... → {*Vorname, Nachname*}

..... → {*Bezeichnung, Beginn,*}



# Verlustlose Zerlegung

## Zerlegung

### R1 (Projekte)

<u>Proj-Nr</u>	Beginn	Pers-Nr	Vorname	Nachname
4711	15.3.2015	2	Rita	Schulze
3050	13.5.2018	3	Werner	Maier
2020	1.2.2018	2	Rita	Schulze
1234	1.4.2017	1	Hans	Müller
3091	1.9.2018	3	Werner	Maier

### R2 (Projektname)

<u>Nachname</u>	<u>Bezeichnung</u>
Schulze	Fahrzeugversuchssystem für Firma WMB
Maier	Erweiterung Personal-Datenbank Firma Kleinert
Schulze	Schnittstellen zwischen Produktion und Verkauf erstellen
Müller	Erweiterung interne Datenbank für unser Softwarehaus
Maier	Elektronische Erfassung der Prüfstandsdaten für Firma WMB

Der Schnitt beider Relationen ist der .....  
Prüfen wir nun, ob diese Attribute in den Kriterien entsprechen.

$[R1] \cap [R2] \rightarrow [R1] \in F^+$

$[R1] \cap [R2] \rightarrow \{.....\}$

$\{Nachname\} \rightarrow \{Pers - Nr, Beginn, Vorname, Nachname, \dots\}$

$[R1] \cap [R2] \rightarrow [R2] \in F^+$

$\{Nachname\} \rightarrow \{Bezeichnung\}$

Die Zerlegung ist nicht .....



## Zerlegung

### R3 (Projekte)

<u>Proj-Nr</u>	Beginn	Bezeichnung	Pers-Nr
4711	15.3.2015	Fahrzeugversuchssystem für Firma WMB	2
3050	13.5.2018	Erweiterung Personal-Datenbank Firma Kleinert	3
2020	1.2.2018	Schnittstellen zwischen Produktion und Verkauf erstellen	2
1234	1.4.2017	Erweiterung interne Datenbank für unser Softwarehaus	1
3091	1.9.2018	Elektronische Erfassung der Prüfstandsdaten für Firma WMB	3

### R4 (Personen)

<u>Pers-Nr</u>	Vorname	Nachname
2	Rita	Schulze
3	Werner	Maier
1	Hans	Müller

$[R3] \cap [R4] \rightarrow [R3] \in F^+$

$[R3] \cap [R4] \rightarrow \{Pers - Nr\}$

$\{Pers - Nr\} \rightarrow \{Proj - Nr, Beginn, Bezeichnung, Pers - Nr\}$

$[R3] \cap [R4] \rightarrow [R4] \in F^+$

$\{Pers - Nr\} \rightarrow \{Vorname, Nachname\} \dots\dots$

Die Zerlegung ist verlustlos



- Relationsschemata [R] sollte abhängigkeitsbewahrend zerlegt werden.
- Gilt bis 3NF und nicht mehr .....
- **Prüfung:**
  - Nach der Zerlegung eines Relationsschemas [R] in verschiedene Teilschemata [R<sub>i</sub>], kann jede FA in mindestens einer der [R<sub>i</sub>] dargestellt werden.

**Formal:**

$$\left( \left( \Pi_{[R1]} (F) \right) \cup \dots \cup \left( \Pi_{[Rn]} (F) \right) \right)^+ = F^+$$

# Abhängigkeitsbewahrende Zerlegung

## Zerlegung

### R3 (Projekte)

<u>Proj-Nr</u>	Beginn	Bezeichnung	Pers-Nr
4711	15.3.2015	Fahrzeugversuchssystem für Firma WMB	2
3050	13.5.2018	Erweiterung Personal-Datenbank Firma Kleinert	3
2020	1.2.2018	Schnittstellen zwischen Produktion und Verkauf erstellen	2
1234	1.4.2017	Erweiterung interne Datenbank für unser Softwarehaus	1
3091	1.9.2018	Elektronische Erfassung der Prüfstandsdaten für Firma WMB	3

### R4 (Personen)

<u>Pers-Nr</u>	Vorname	Nachname
2	Rita	Schulze
3	Werner	Maier
1	Hans	Müller

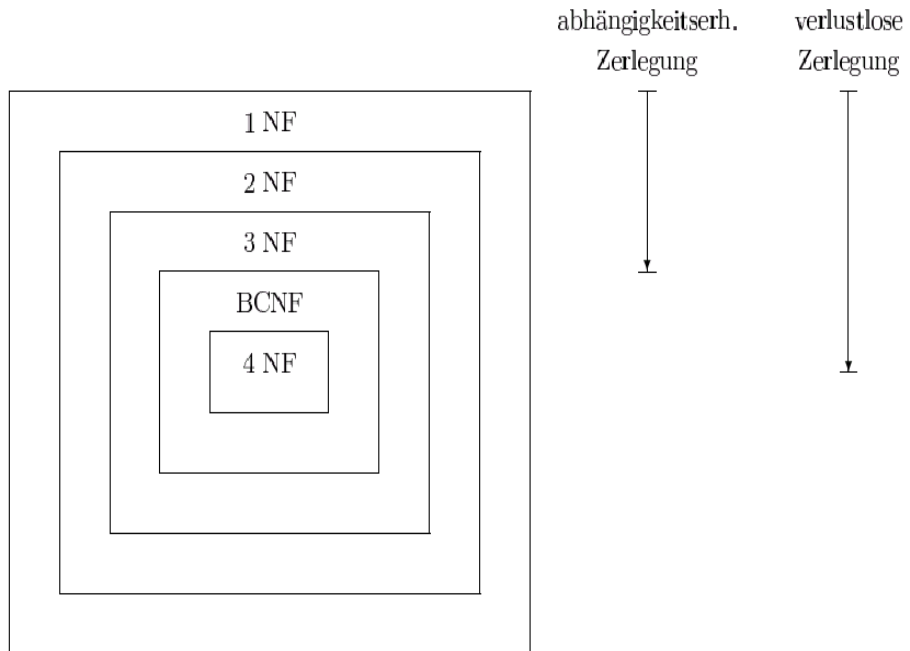
$Pers - Nr \rightarrow \{ \dots, \dots \}$   
 $\dots \rightarrow \{Bezeichnung, Beginn\}$

Prüfen, ob die FA's auf den Teilrelationen abgebildet werden können



Die dargestellte Zerlegung ist .....

# Darstellung der Normalformen



## Ende Kapitel 4

