

Datenmanagement & -analyse

Datenmodellierung: Normalisierung relationaler Datenbanken

Prof. Dr. Christoph M. Flath

Lehrstuhl für WI & BA

Julius-Maximilians-Universität Würzburg

Sommersemester 2021

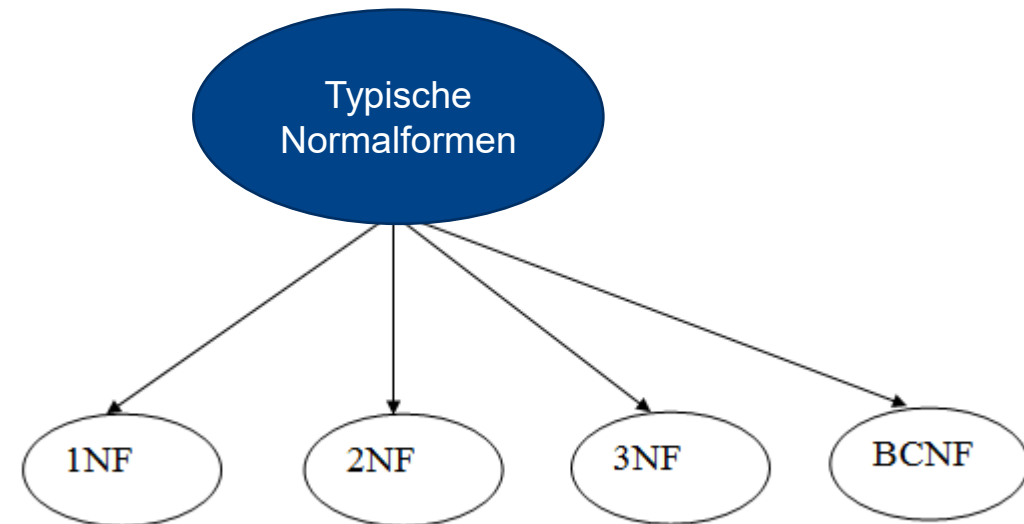


Relationale Datenbanken sind nicht per se systematisch

- Durch das Zusammenführen von Datenbeständen in ein Datenbanksystem sind nicht automatisch alle Nachteile der klassischen Datenspeicherung in Dateien aufgehoben!
- Im relationalen Modell müssen alle Tabellen **normalisiert**, d.h. anhand von Regeln transformiert werden
- Gründe für Normalisierung:
 - Vermeidung überflüssiger **Redundanzen**
 - Vermeidung von **Anomalien**
 - Sicherstellen der **Integrität**



- Normalisierung ist der Prozess der Organisation der Daten in der Datenbank
- Die Normalisierung wird verwendet, um die Redundanz einer Beziehung oder eines Satzes von Beziehungen zu minimieren
- Sie wird auch verwendet, um unerwünschte Merkmale wie Einfüge-, Aktualisierungs- und Lösungsanomalien zu beseitigen
- Die Normalisierung unterteilt die größere Tabelle in kleinere Tabellen und verknüpft sie mit Hilfe von Beziehungen
 - Schrittweise **Zerlegung** von Relationen in mehrere Relationen auf der Grundlage **funktionaler Abhängigkeiten**
- Verschiedene Normalformen legen dabei zunehmend strengere Kriterien an das Datenmodell an



- 1 Funktionale Abhängigkeit**
- 2 Erste Normalform**
- 3 Zweite Normalform**
- 4 Dritte Normalform**
- 5 Höhere Normalformen**

- Funktionale Abhängigkeit ist eine Beziehung, die zwischen zwei Attributen X und Y in einer besteht:
 $X \rightarrow Y$
- Die linke Seite wird als Determinante, die rechte Seite als abhängig bezeichnet.
- Beispiel: Eine Angestelltentabelle mit den Attributen An_Id, An_Name, An_Adresse
 - Hier kann das Attribut "An_Id" das Attribut "An_Name" der Angestelltentabelle eindeutig identifizieren, d.h. $An_Id \rightarrow An_Name$
 - Wir können sagen, dass An_Name funktional von An_Id abhängig ist.
- Funktionale Abhängigkeit besteht typischerweise zwischen dem Primärschlüssel und einem Nicht-Schlüssel-Attribut innerhalb einer Tabelle
- **Triviale funktionale Abhängigkeit**
 - $A \rightarrow B$ hat triviale funktionale Abhängigkeit, wenn B eine Teilmenge von A ist.
 - Bsp.: $\{An_id, An_Name\} \rightarrow An_Id$ ist eine triviale funktionale Abhängigkeit, da letzteres eine Teilmenge von $\{An_Id, An_Name\}$ ist
- **Nicht-triviale funktionale Abhängigkeit**
 - $A \rightarrow B$ hat eine nicht-triviale funktionale Abhängigkeit, wenn B keine Teilmenge von A ist.
 - Bsp.: $ID \rightarrow Name$, $Name \rightarrow Geburtstag$

- Die Axiome von Armstrong sind die grundlegende Inferenzregel
- Die Armstrong-Axiome werden verwendet, um auf funktionale Abhängigkeiten in einer relationalen Datenbank zu schließen.
- Die Inferenzregel kann auf eine Menge von funktionalen Abhängigkeiten angewendet werden, um weitere FA abzuleiten

1. Reflexivität

Wenn $X \supseteq Y$ dann $X \rightarrow Y$

$X = \{a, b, c, d, e\}, Y = \{a, b, c\}$

2. Erweiterungsregel

Wenn $X \rightarrow Y$ bestimmt, dann bestimmt $XZ \rightarrow YZ$ für jedes Z .

$X \rightarrow Y \rightarrow XZ \rightarrow YZ$

3. Transitivität

Wenn $X \rightarrow Y$ bestimmt und $Y \rightarrow Z$ bestimmt, dann muss $X \rightarrow Z$ auch bestimmen.

$ID \rightarrow Name, Name \rightarrow Geburtstag \rightarrow ID \rightarrow Geburtstag$

Axiome von Armstrong (2)

Um Herleitungen einfacher zu gestalten, können zusätzlich die folgenden (abgeleiteten) Regeln verwendet werden:

4. Vereinigungsregel

Wenn $X \rightarrow Y$ bestimmt und $X \rightarrow Z$ bestimmt, dann muss X auch Y und Z bestimmen

$$X \rightarrow Y \text{ und } X \rightarrow Z \Rightarrow X \rightarrow YZ$$

Beweis:

$$X \rightarrow Y \Rightarrow XX \rightarrow XY \Rightarrow X \rightarrow XY \text{ [Erweiterungsregel]}$$

$$XY \rightarrow Z \Rightarrow XY \rightarrow YZ \text{ [Erweiterungsregel]}$$

$$X \rightarrow YZ \text{ [Transitivität auf den beiden Ergebnissen zuvor]}$$

5. Dekompositionsregel

wenn $X \rightarrow Y$ und Z bestimmt, dann bestimmt $X \rightarrow Y$ und Z separat.

$$X \rightarrow YZ \Rightarrow X \rightarrow Y \text{ und } X \rightarrow Z$$

Beweis:

$$YZ \rightarrow Y \text{ (Reflexivität)}$$

$$X \rightarrow Y \text{ (Transitivität auf } X \rightarrow YZ \text{ und } YZ \rightarrow Y \text{)}$$

6. Pseudotransitivitätsregel

Wenn $X \rightarrow Y$ bestimmt und $YZ \rightarrow W$ bestimmt, dann bestimmt $XZ \rightarrow W$

$$X \rightarrow Y \text{ und } YZ \rightarrow W \Rightarrow XZ \rightarrow W$$

Beweis:

$$XZ \rightarrow YZ \text{ (Erweiterungsregel)}$$

$$XZ \rightarrow W \text{ (Transitivität)}$$

- 1 Funktionale Abhängigkeit
- 2 Erste Normalform
- 3 Zweite Normalform
- 4 Dritte Normalform
- 5 Höhere Normalformen

Erste Normalform (1NF)

- Eine Relation ist 1NF, wenn sie nur atomare Werte enthält
 - Ein Attribut einer Tabelle darf nicht mehrere Werte enthalten
 - Es darf nur ein einwertige Attribute geben

- Die erste Normalform lässt mehrwertige Attribute, zusammengesetzte Attribute und deren Kombinationen nicht zu.

AN_ID	AN_NAME	AN_TEL	AN_LAND
14	John	72728263, 90647382	BY
20	Harry	85747838	NRW
12	Sam	73903723, 85898303	BW

Erste Normalform (2)

AN_ID	AN_NAME	AN_TEL	AN_LAND
14	John	7272826385	BY
14	John	9064738238	BY
20	Harry	8574783832	NRW
12	Sam	7390372389	BW
12	Sam	8589830302	BW

- 1 Funktionale Abhängigkeit
- 2 Erste Normalform
- 3 **Zweite Normalform**
- 4 Dritte Normalform
- 5 Höhere Normalformen

Zweite Normalform (2NF)

- In der 2NF muss die Relation in der 1NF sein.
- In der zweiten Normalform sind alle Nicht-Schlüssel-Attribute funktional abhängig vom Primärschlüssel

PROF_ID	FACH	PROF_ALTER
25	BWL	30
25	WiInfo	30
47	BWL	35
83	BWL	38
83	VWL	38

Zweite Normalform (2NF)

- In der gegebenen Tabelle ist das Nicht-Primzahl-Attribut PROF_ALTER abhängig von PROF_ID, das eine richtige Teilmenge eines Kandidatenschlüssels ist
- Deshalb verstößt sie gegen die Regel für 2NF
- Um die gegebene Tabelle in 2NF zu konvertieren, zerlegen wir sie in zwei Tabellen

PROF_ID	PROF_ALTER
25	30
47	35
83	38

PROF_ID	FACH
25	BWL
25	WiInfo
47	BWL
83	BWL
83	VWL

- 1 Funktionale Abhängigkeit
- 2 Erste Normalform
- 3 Zweite Normalform
- 4 **Dritte Normalform**
- 5 Höhere Normalformen

Dritte Normalform (3NF)

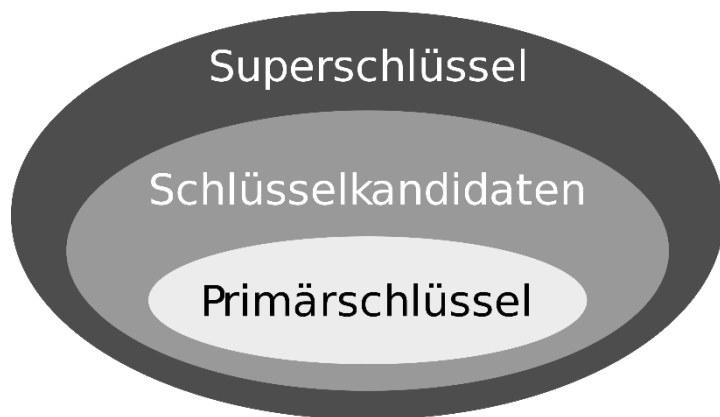
- Eine Relation ist in 3NF, wenn sie in 2NF ist und wenn keine Nichtschlüssel-Attribute **transitiv abhängig** vom Primärschlüssel der Relation sind
 - Es dürfen keine Nichtsschlüssel-Attribute indirekt über andere Nichtschlüssel-Attribute vom Primärschlüssel abhängen
 - Falls die Relation keine oder nur ein Nichtschlüssel-Attribut enthält, ist sie bereits in 3. Normalform
- 3NF wird verwendet, um die Datenduplikation zu reduzieren und Datenintegrität sicherzustellen

AN_ID	AN_NAME	AN_PLZ	AN_LAND	AN_Stadt
222	Harry	70180	BW	Stuttgart
333	Stephan	76133	BW	Karlsruhe
444	Lan	97070	BY	Würzburg
555	Katharine	80331	BY	München
666	John	10115	BE	Berlin

Schlüsselkandidat: {AN_ID}

Dritte Normalform (2)

- Hier sind AN_LAND & AN_STADT abhängig von AN_PLZ und AN_PLZ abhängig von AN_ID. Die Nicht-Prime-Attribute (AN_LAND, AN_STADT) sind transitiv abhängig vom Superschlüssel (AN_ID)
 - Das verstößt gegen die Regel der dritten Normalform
- Deshalb müssen wir die AN_STADT und AN_LAND in die neue Tabelle <ANGESTELLTE_ADRESSEN> verschieben, mit AN_PLZ als Primärschlüssel



AN_ID	AN_NAME	AN_PLZ
222	Harry	70180
333	Stephan	76133
444	Lan	97070
555	Katharine	80331
666	John	10115

AN_PLZ	AN_LAND	AN_Stadt
70180	BW	Stuttgart
76133	BW	Karlsruhe
97070	BY	Würzburg
80331	BY	München
10115	BE	Berlin

Weiteres Beispiel – Ausgangssituation: Nix

Klettersteigset Verleih									
HNR	Hersteller	ProdID	Produktname	SID	Schwierigkeit	TID1	Test1	TID2	Test2
P0002	Petzl	SAL	Salewa Ergo Zip	001	A	2	gut	1	sehr gut
				002	B	2	gut	1	sehr gut
				002	B	3	mittel	NULL	NULL
				003	C	1	sehr gut	NULL	NULL
		SE	Scorpio Eashook	001	A	1	sehr gut	NULL	NULL
				002	B	1	sehr gut	NULL	NULL
				003	C	2	gut	3	mittel
				003	C	1	sehr gut	2	gut
E0045	Edelrid	CC23	Cable Comfort	003	C	1	sehr gut	2	gut
				004	D	1	sehr gut	2	gut
				004	D	2	gut	3	mittel
		CV	Cable Vario	005	E	1	sehr gut	2	gut
				003	C	3	mittel	3	mittel
				001	A	5	mangelhaft	NULL	NULL
M0001	Mammut	TSB	Tec Step Bionic	001, 003	A, C	2	gut	3	mittel
				002	B	2	gut	3	mittel
				002	B	2	gut	3	mittel

1. Normalform

Klettersteigset Verleih							
HNR	Hersteller	ProdID	Produktname	SID	Schwierigkeit	TID	Test
P0002	Petzl	SAL	Salewa Ergo Zip	001	A	2	gut
P0002	Petzl	SAL	Salewa Ergo Zip	002	B	2	gut
P0002	Petzl	SAL	Salewa Ergo Zip	002	B	3	mittel
P0002	Petzl	SAL	Salewa Ergo Zip	003	C	1	sehr gut
P0002	Petzl	SE	Scorpio Eashook	001	A	1	sehr gut
P0002	Petzl	SE	Scorpio Eashook	002	B	1	sehr gut
P0002	Petzl	SE	Scorpio Eashook	003	C	2	gut
P0002	Petzl	SE	Scorpio Eashook	003	C	1	sehr gut
E0045	Edelrid	CC23	Cable Comfort	003	C	1	sehr gut
E0045	Edelrid	CC23	Cable Comfort	004	D	1	sehr gut
E0045	Edelrid	CC23	Cable Comfort	004	D	2	gut
E0045	Edelrid	CV	Cable Vario	005	E	1	sehr gut
E0045	Edelrid	CV	Cable Vario	003	C	3	mittel
E0045	Edelrid	CV	Cable Vario	001	A	5	mangelhaft
M0001	Mammut	TSB	Tec Step Bionic	001	A	2	gut
M0001	Mammut	TSB	Tec Step Bionic	003	C	2	gut
M0001	Mammut	TSB	Tec Step Bionic	002	B	2	gut

2. Normalform

Set-Test für Schwierigkeit		
#ProdID	#SID	#TID
SAL	001	2
SAL	002	2
SAL	002	3
SAL	003	1
SE	001	1
SE	002	1
SE	003	2
SE	003	1
CC23	003	1
CC23	004	1
CC23	004	2
CV	005	1
CV	003	3
CV	001	5
TSB	001	2
TSB	003	2
TSB	002	2

Produkte			
ProdID	Produktname	HNR	Hersteller
SAL	Salewa Ergo Zip	P0002	Petzl
SE	Scorpio Eashook	P0002	Petzl
CC23	Cable Comfort	E0045	Edelrid
CV	Cable Vario	E0045	Edelrid
TSB	Tec Step Bionic	M0001	Mammut

Test	
TID	Test
1	sehr gut
2	gut
3	mittel
5	mangelhaft

Schwierigkeit	
SID	Schwierigkeit
001	A
002	B
003	C
004	D
005	E

3. Normalform

Set-Test für Schwierigkeit		
#ProdID	#SID	#TID
SAL	001	2
SAL	002	2
SAL	002	3
SAL	003	1
SE	001	1
SE	002	1
SE	003	2
SE	003	1
CC23	003	1
CC23	004	1
CC23	004	2
CV	005	1
CV	003	3
CV	001	5
TSB	001	2
TSB	003	2
TSB	002	2

Produkte		
ProdID	Produktname	#HNR
SAL	Salewa Ergo Zip	P0002
SE	Scorpio Eashook	P0002
CC23	Cable Comfort	E0045
CV	Cable Vario	E0045
TSB	Tec Step Bionic	M0001

Hersteller	
HNR	Hersteller
P0002	Petzl
E0045	Edelrid
M0001	Mammut

Test	
TID	Test
1	sehr gut
2	gut
3	mittel
5	mangelhaft

Schwierigkeit	
SID	Schwierigkeit
001	A
002	B
003	C
004	D
005	E

- 1 Funktionale Abhängigkeit
- 2 Erste Normalform
- 3 Zweite Normalform
- 4 Dritte Normalform
- 5 **Höhere Normalformen**

Boyce Codd Normalform (BCNF)

- BCNF ist die erweiterte, strengere Version von 3NF
- Eine Tabelle ist in BCNF, wenn sie sich in 3NF befindet und für jede funktionale Abhängigkeit $X \rightarrow Y$, X der Superschlüssel der Tabelle ist
 - ein Nichtschlüssel-Attribut darf nicht einen Teil des zusammengesetzten Schlüssels bestimmen und so einen weiteren Schlüsselkandidaten bilden
 - Fakten sollen nur einmal gespeichert werden
 - Falls die Relation nur einen Schlüsselkandidaten hat, ist sie bereits in Boyce-Codd-Normalform

AN_ID	AN_STAAT	AN_ABT	ABT_TYP	AN_ABT_NR
264	Indien	Design	D394	283
264	Inden	QM	D394	300
364	UK	Vertrieb	D283	232
364	UK	Entwicklung	D283	549

$AN_ID \rightarrow AN_STAAT$

$AN_ABT \rightarrow \{ABT_TYP, AN_ABT_NR\}$

Schlüsselkandidat: $\{AN_ID, AN_ABT\}$

Die Tabelle ist nicht in BCNF, weil weder **AN_ABT** noch **AN_ID** allein Schlüssel sind.

Boyce Codd Normalform (2)

<u>AN_ID</u>	<u>AN_STAAT</u>
264	Indien
364	UK

<u>AN_ABT</u>	<u>ABT_TYP</u>	<u>AN_ABT_NR</u>
Design	D394	283
QM	D394	300
Vertrieb	D283	232
Entwicklung	D283	549

<u>AN_ID</u>	<u>AN_ABT</u>
264	Design
264	QM
364	Vertrieb
364	Entwicklung

Mapping-Tabelle

Nun wird die BCNF erfüllt, da der linke Teil der beiden funktionalen Abhängigkeiten ein Schlüssel ist.

Vierte Normalform (4NF)

- Eine Relation ist in 4NF, wenn sie in Boyce-Codd-Normalform ist und keine mehrwertige Abhängigkeit hat
- Wenn für eine Abhängigkeit $A \rightarrow B$ für einen einzelnen Wert von A mehrere Werte von B existieren, dann ist die Beziehung eine mehrwertige Abhängigkeit
- Die Beispieltabelle ist in 3NF, aber FACH und HOBBY sind zwei unabhängige Entitäten
 - Daher gibt es keine Beziehung zwischen FACH und HOBBY
 - In der Beispiel-Relation hat der Student mit STU_ID, 21 zwei Fächer, Informatik und Mathematik, und zwei Hobbys, Tanzen und Singen
 - Es besteht also eine mehrwertige Abhängigkeit von STU_ID, was zu unnötigen Wiederholungen von Daten führen kann

STU_ID	FACH	HOBBY
21	Informatik	Tanzen
21	Mathe	Singen
34	Chemie	Tanzen
74	Biologie	Cricket
59	Physik	Hockey

Vierte Normalform (2)

STU_ID	FACH
21	Informatik
21	Mathe
34	Chemie
74	Biologie
59	Physik

STU_ID	HOBBY
21	Tanzen
21	Singen
34	Tanzen
74	Cricket
59	Hockey

Weiteres Beispiel

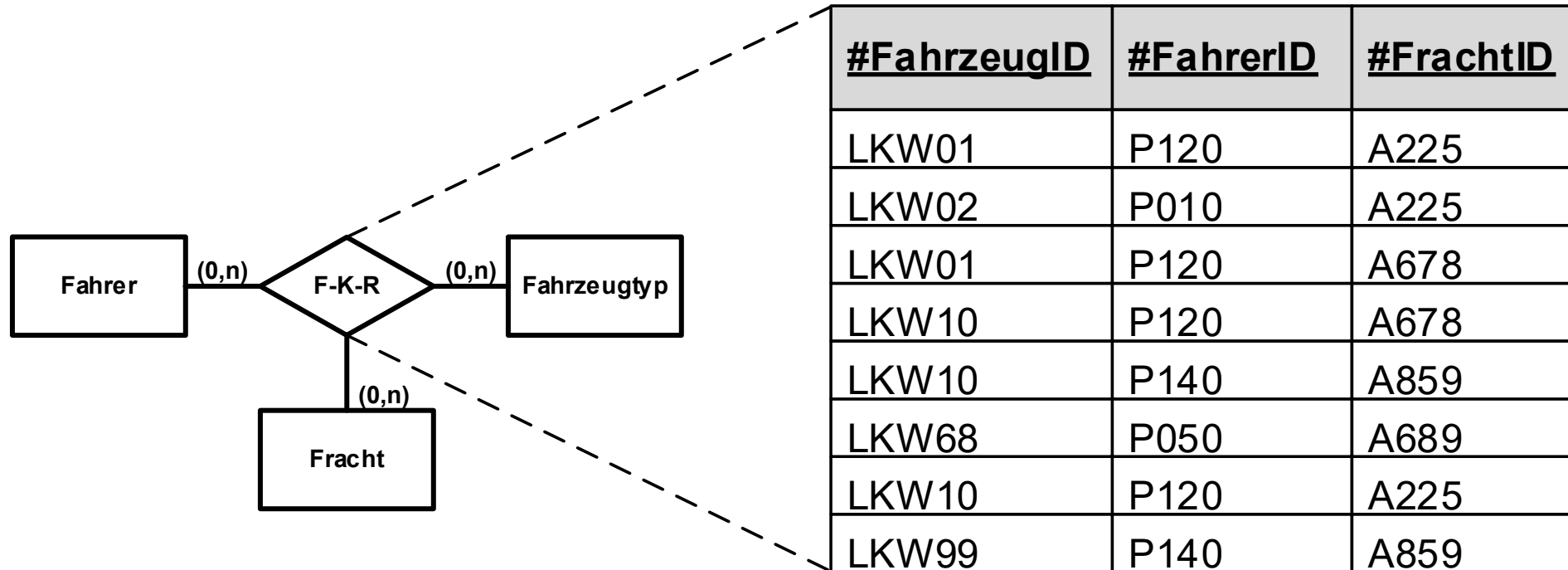
<u>#FahrzeugID</u>	<u>#FahrerID</u>	<u>#FrachtID</u>
LKW01	P120	A225
LKW02	P010	A225
LKW01	P120	A678
LKW10	P120	A678
LKW10	P140	A859
LKW68	P050	A689
LKW10	P120	A225
LKW99	P140	A859

<u>FrachtID</u>	Frachtgut	Gefahren- klasse
A225	Flüssigeis	unkritisch
A678	Brennstäbe	sehr kritisch
A859	Forscherbier	ultrakritisch
A689	Kunstschnee	kritisch

<u>FahrzeugID</u>	Fahrzeugbezeichnung	PS
LKW01	MB Actros	400
LKW02	MAN TGS	250
LKW10	Volvo FH16	380
LKW99	MB L-911	436
LKW68	MAN TGX	150

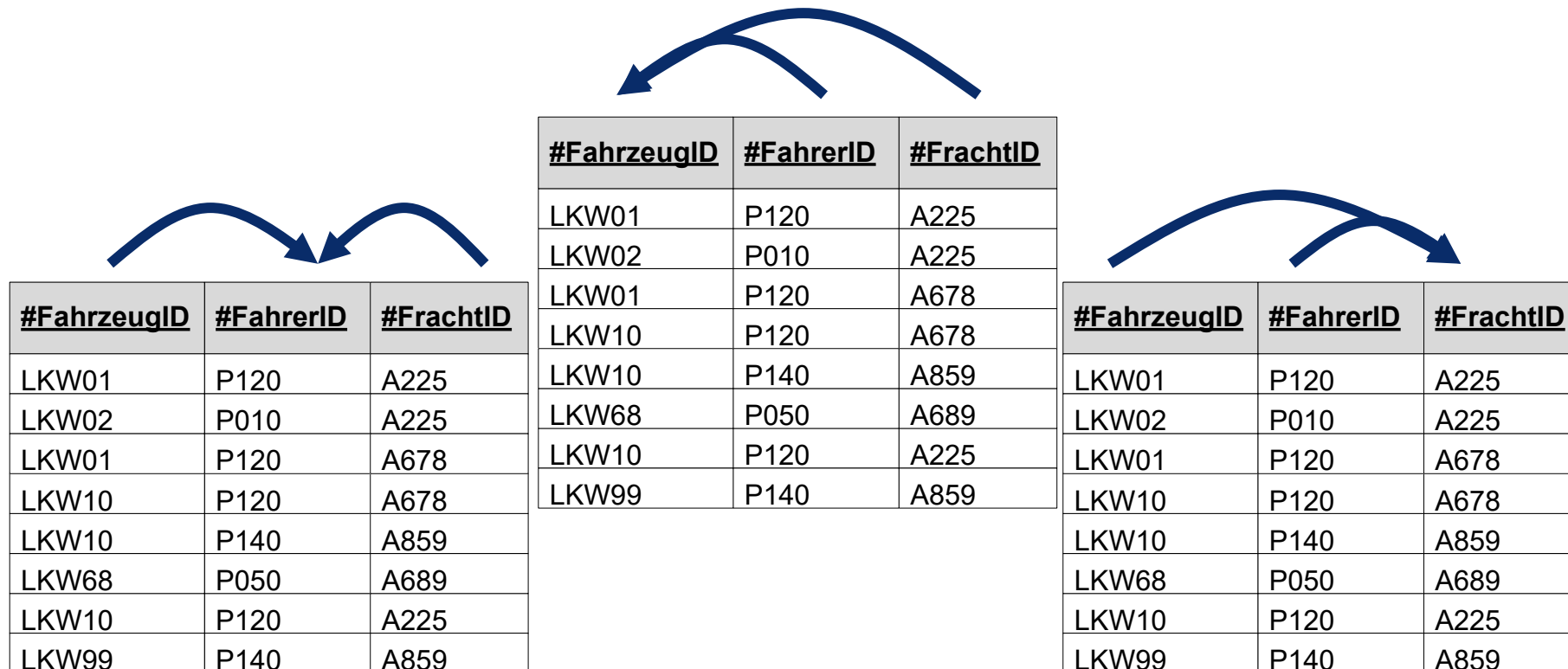
<u>FahrerID</u>	Fahrername	Alter
P010	Manfred Lebmann	40
P120	Wolfgang Taschenbier	38
P140	Karl-Heinz Böhmer	60
P050	Stanislaw Nowak	77

Weiteres Beispiel




Weiteres Beispiel

- Überprüfen auf mehrwertige Abhängigkeiten mit allen Attributen als jeweiliger Ausgangspunkt

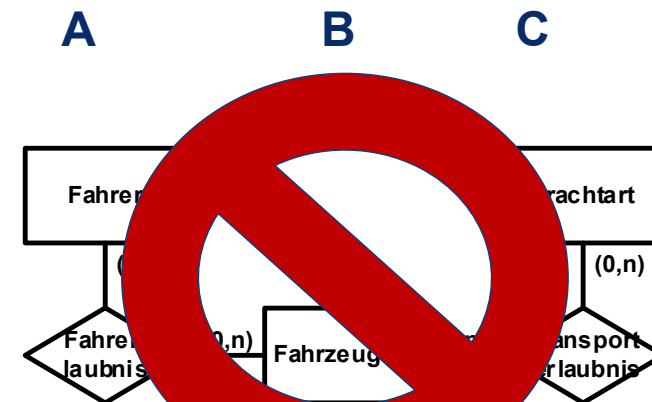


Weiteres Beispiel

A	B	C
<u>#FahrzeugID</u>	<u>#FahrerID</u>	<u>#FrachtID</u>
LKW01	P120	A225, A678
LKW02	P010	A225
LKW10	P120	A225, A678
	P140	A225
LKW68	P050	A689
LKW99	P140	A859



<u>#FahrzeugID</u>	<u>#FahrerID</u>	<u>#FrachtID</u>
LKW01	P120	A225
LKW02	P010	A225
LKW01	P120	A678
LKW10	P120	A678
LKW10	P140	A859
LKW68	P050	A689
LKW10	P120	A225
LKW99	P140	A859



Es darf sich zu keinem Wert von A für jede Kombination dieses Wertes mit einem Wert von B eine identische Menge an Werten von C ergeben

Keine mehrwertigen Abhängigkeiten!

Weiteres Beispiel

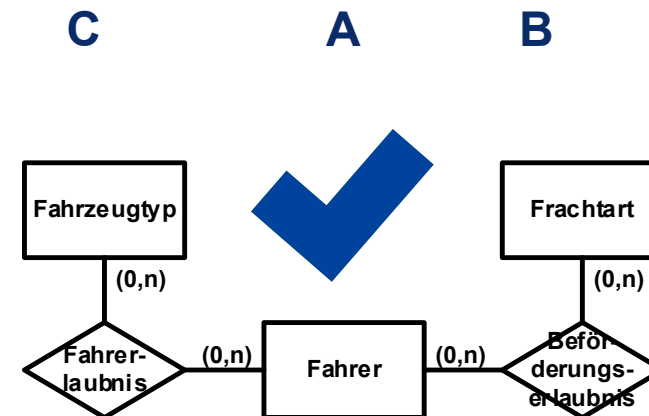
A	B	C
<u>#FahrerID</u>	<u>#FrachtID</u>	<u>#FahrzeugID</u>
P010	A225	LKW02
P050	A678	LKW68
P120	A225	LKW01, LKW10
	A678	LKW01, LKW10
P140	A859	LKW99



<u>#FahrzeugID</u>	<u>#FahrerID</u>	<u>#FrachtID</u>
LKW01	P120	A225
LKW02	P010	A225
LKW01	P120	A678
LKW10	P120	A678
LKW10	P140	A859
LKW68	P050	A689
LKW10	P120	A225
LKW99	P140	A859

Es darf sich zu keinem Wert von A für jede Kombination dieses Wertes mit einem Wert von B eine identische Menge an Werten von C ergeben

Potentielle mehrwertigen Abhängigkeiten vorhanden!

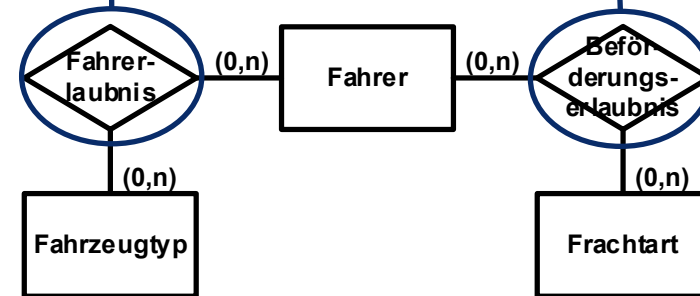
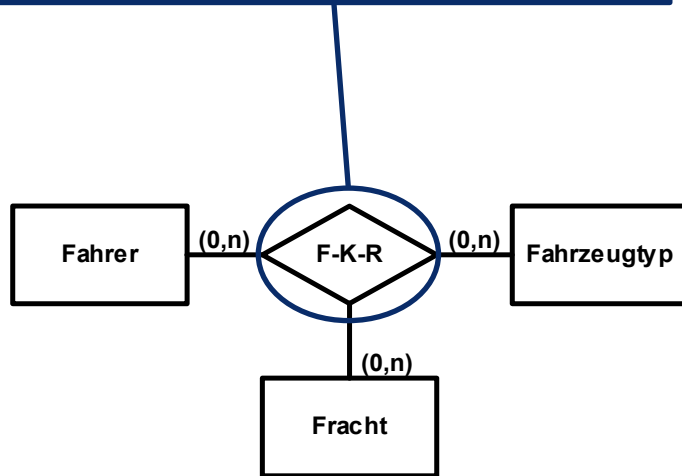


Weiteres Beispiel

#FahrzeugID	#FahrerID	#FrachtID
LKW01	P120	A225
LKW02	P010	A225
LKW01	P120	A678
LKW10	P120	A678
LKW10	P140	A859
LKW68	P050	A689
LKW10	P120	A225
LKW99	P140	A859


#FahrzeugID	#FahrerID
LKW01	P120
LKW02	P010
LKW10	P120
LKW10	P140
LKW68	P050
LKW99	P140

#FahrerID	#FrachtID
P010	A225
P050	A689
P120	A225
P120	A678
P140	A859




Weiteres Beispiel

<u>#FahrzeugID</u>	<u>#FahrerID</u>
LKW01	P120
LKW02	P010
LKW10	P120
LKW10	P140
LKW68	P050
LKW99	P140



<u>#FahrerID</u>	<u>#FrachtID</u>
P010	A225
P050	A689
P120	A225
P120	A678
P140	A859



<u>FrachtID</u>	Frachtgut	Gefahren- klasse
A225	Flüssigeis	unkritisch
A678	Brennstäbe	sehr kritisch
A859	Forscherbier	ultrakritisch
A689	Kunstschnee	kritisch



<u>FahrzeugID</u>	Fahrzeugbezeichnung	PS
LKW01	MB Actros	400
LKW02	MAN TGS	250
LKW10	Volvo FH16	380
LKW99	MB L-911	436
LKW68	MAN TGX	170



<u>FahrerID</u>	Fahrername	Alter
P010	Manfred Lebmann	40
P120	Wolfgang Taschenbier	38
P140	Karl-Heinz Böhmer	60
P050	Stanislaw Nowak	27



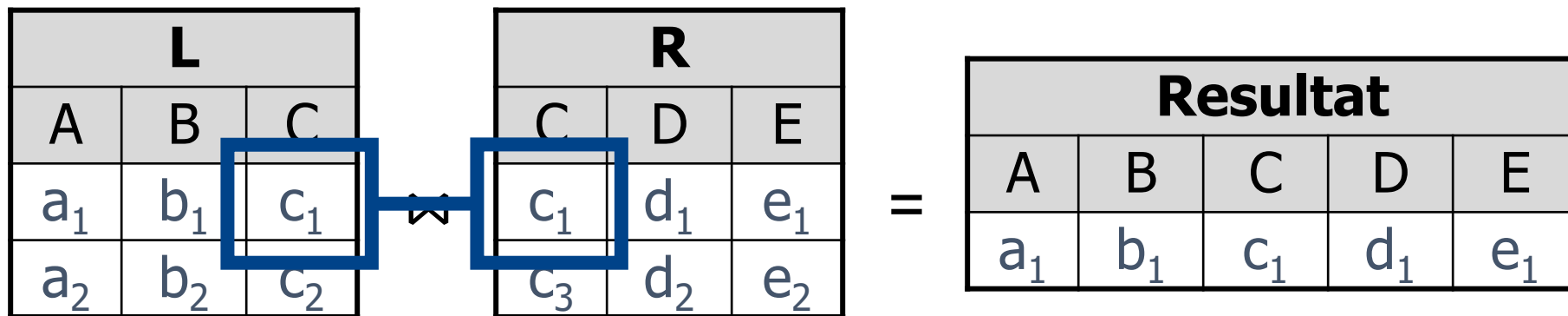
- Normalisierungen führten meist zu einer Zerlegung **einer** Relation in **zwei** andere
- Es kann sein, dass eine Relation **ohne Informationsverlust nicht in zwei, aber in drei oder mehr** Relationen zerlegt werden kann
- Diese Eigenschaft wird **n-Zerlegbarkeit** genannt
- n-Zerlegbarkeit bezieht sich auf Beziehungen, die über mehr als zwei Attribute greifen

Verlustfreier Join (Lossless Join)

- Zerlegungseigenschaft einer Relation
- Stellt sicher, dass bei der Wiedervereinigung von Einzelrelationen mittels **Natural-Join-Operation keine verfälschten Datensätze** gegenüber der Ursprungsrelation erzeugt werden
- **Natural-Join-Operation** verknüpft Datensätze zweier Relationen anhand gemeinsamer Attribute

Natural Join

- Natural Join (auch natürlicher Verbund)
 - **Kartesisches Produkt** zweier Tabellen
 - **Selektion** mit der Bedingung, dass der Inhalt bestimmter Spalten identisch sein muss
 - **Ausblendung** gleicher Spalten im Natural Join
- Eigenschaften
 - kommutativ ($L \bowtie R = R \bowtie L$)
 - assoziativ ($L \bowtie R) \bowtie T = L \bowtie (R \bowtie T)$)



5. Normalform (5NF)

- Eine Relation befindet sich in 5. Normalform wenn sie sich in 4. Normalform befindet und Informationen der Relation nicht durch einen **verlustfreien Join** wiederhergestellt werden können
 - sie darf sich **nicht verlustfrei in duale Einzelrelationen** zerlegen lassen
 - eine Relation mit weniger als 3 Attributen ist immer in 5. Normalform (wenn sie in 1. Normalform ist) und
 - eine Relation in 3. Normalform mit Nichtschlüssel-Attributen ist automatisch in 5. Normalform

FACH	STUDENT	SEMESTER
Informatik	Anshika	1. Semester
Informatik	John	1. Semester
Mathe	John	1. Semester
Mathe	Akash	2. Semester
Chemie	Praveen	1. Semester

5. Normalform (2)

- Im Beispiel belegt John sowohl Informatik als auch Mathe im ersten Semester, aber keine Mathe im zweiten Semester
 - Eine Kombination aus all diesen Feldern ist erforderlich, um gültige Daten zu erhalten.
- Angenommen, wir fügen ein neues Semester als Semester 3 hinzu, wissen aber nicht, welches Fach und wer dieses Fach belegen wird, also lassen wir Dozent und Fach als NULL stehen
 - Da aber alle drei Spalten zusammen als Primärschlüssel fungieren, können wir die anderen beiden Spalten nicht leer lassen.
- Um die Tabelle in 5NF umzuwandeln, können wir sie in drei Relationen P1, P2 und P3 zerlegen

FACH	STUDENT	SEMESTER
Informatik	Anshika	1. Semester
Informatik	John	1. Semester
Mathe	John	1. Semester
Mathe	Akash	2. Semester
Chemie	Praveen	1. Semester

5. Normalform (2)

FACH	STUDENT	SEMESTER
Informatik	Anshika	1. Semester
Informatik	John	1. Semester
Mathe	John	1. Semester
Mathe	Akash	2. Semester
Chemie	Praveen	1. Semester

SEMESTER	FACH
1. Semester	Informatik
1. Semester	Mathe
1. Semester	Chemie
2. Semester	Mathe

FACH	STUDENT
Informatik	Anshika
Informatik	John
Mathe	John
Mathe	Akash
Chemie	Praveen

SEMSTER	STUDENT
1. Semester	Anshika
1. Semester	John
1. Semester	John
2. Semester	Akash
1. Semester	Praveen

Weiteres Beispiel: Stand 4. Normalform

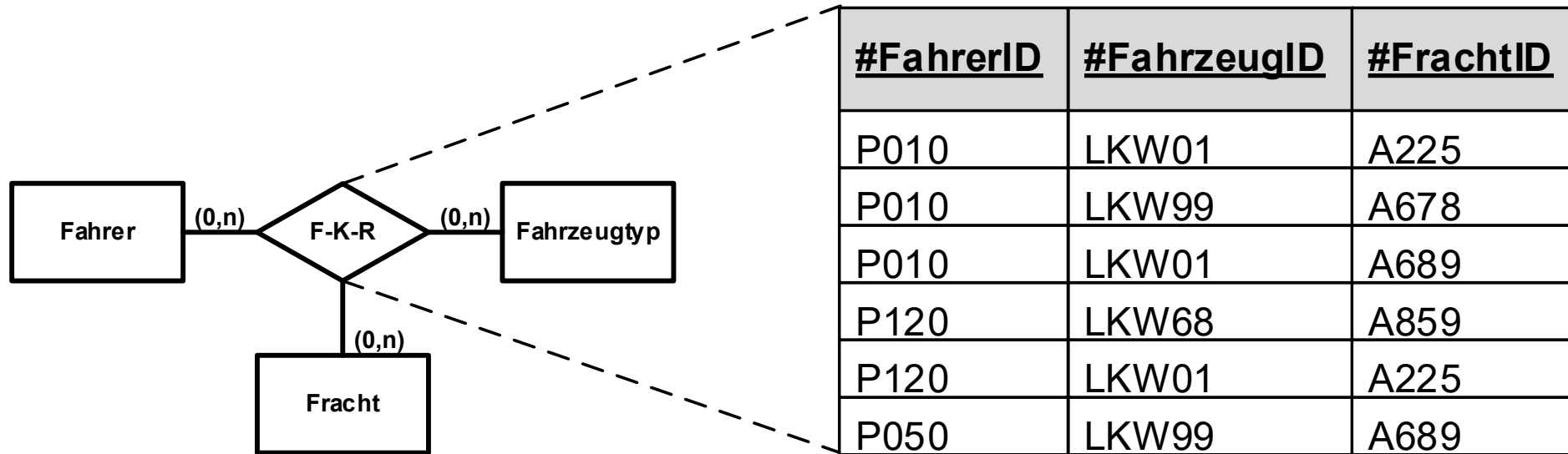
<u>#FahrerID</u>	<u>#FahrzeugID</u>	<u>#FrachtID</u>
P010	LKW01	A225
P010	LKW99	A678
P010	LKW01	A689
P120	LKW68	A859
P120	LKW01	A225
P050	LKW99	A689

<u>FrachtID</u>	Frachtgut	Gefahren- klasse
A225	Flüssigeis	unkritisch
A678	Brennstäbe	sehr kritisch
A859	Forscherbier	ultrakritisch
A689	Kunstschnee	kritisch

<u>FahrzeugID</u>	Fahrzeugbezeichnung	PS
LKW01	MB Actros	400
LKW02	MAN TGS	250
LKW10	Volvo FH16	380
LKW99	MB L-911	436
LKW68	MAN TGX	110

<u>FahrerID</u>	Fahrername	Alter
P010	Manfred Lebmann	40
P120	Wolfgang Taschenbier	38
P140	Karl-Heinz Böhmer	60
P050	Stanislaw Nowak	27

Weiteres Beispiel



Weiteres Beispiel

- Zerlegung der Relation in duale Einzelrelationen

<u>#FahrerID</u>	<u>#FahrzeugID</u>	<u>#FrachtID</u>
P010	LKW01	A225
P010	LKW99	A678
P010	LKW01	A689
P120	LKW68	A859
P120	LKW01	A225
P050	LKW99	A689

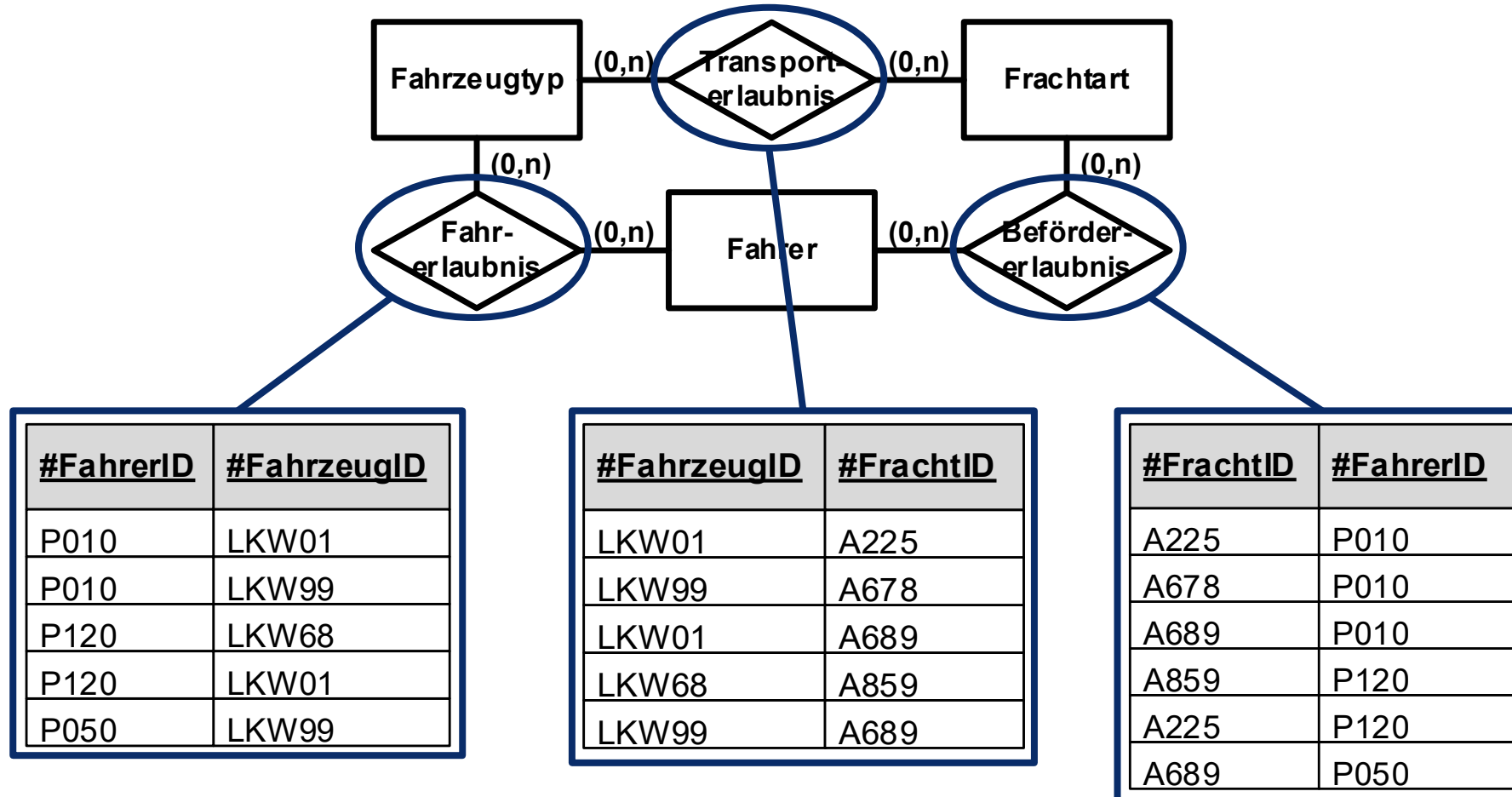
<u>#FahrerID</u>	<u>#FahrzeugID</u>
P010	LKW01
P010	LKW99
P120	LKW68
P120	LKW01
P050	LKW99

<u>#FahrzeugID</u>	<u>#FrachtID</u>
LKW01	A225
LKW99	A678
LKW01	A689
LKW68	A859
LKW99	A689

<u>#FrachtID</u>	<u>#FahrerID</u>
A225	P010
A678	P010
A689	P010
A859	P120
A225	P120
A689	P050

Weiteres Beispiel

- Zerlegung der Relation in duale Einzelrelationen

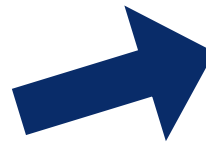


Weiteres Beispiel

- Verknüpfung zweier Einzelrelationen mittels Natural-Join-Operation

<u>#FahrerID</u>	<u>#FahrzeugID</u>
P010	LKW01
P010	LKW99
P120	LKW68
P120	LKW01
P050	LKW99

<u>#FahrzeugID</u>	<u>#FrachtID</u>
LKW01	A225
LKW99	A678
LKW01	A689
LKW68	A859
LKW99	A689



<u>#FahrerID</u>	<u>#FahrzeugID</u>	<u>#FrachtID</u>
P010	LKW01	A225
P010	LKW01	A689
P010	LKW99	A678
P010	LKW99	A689
P120	LKW68	A859
P120	LKW01	A225
P120	LKW01	A689
P050	LKW99	A678
P050	LKW99	A689

Weiteres Beispiel

- Verknüpfung der Ergebnisrelation mit weiterer Einzelrelation mittels Natural-Join-Operation

<u>#FrachtID</u>	<u>#FahrerID</u>
A225	P010
A678	P010
A689	P010
A859	P120
A225	P120
A689	P050



<u>#FahrerID</u>	<u>#FahrzeugID</u>	<u>#FrachtID</u>
P010	LKW01	A225
P010	LKW01	A689
P010	LKW99	A678
P010	LKW99	A689
P120	LKW68	A859
P120	LKW01	A225
P120	LKW01	A689
P050	LKW99	A678
P050	LKW99	A689

Weiteres Beispiel

- Abgleich von Ergebnisrelation mit Ursprungsrelation

<u>#FahrerID</u>	<u>#FahrzeugID</u>	<u>#FrachtID</u>
P010	LKW01	A225
P010	LKW01	A689
P010	LKW99	A678
P010	LKW99	A689
P120	LKW68	A859
P120	LKW01	A225
P050	LKW99	A689



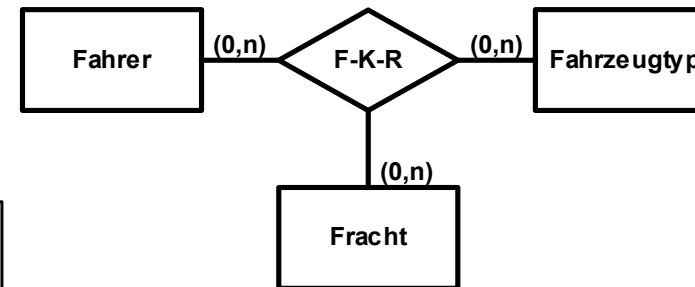
<u>#FahrerID</u>	<u>#FahrzeugID</u>	<u>#FrachtID</u>
P010	LKW01	A225
P010	LKW99	A678
P010	LKW01	A689
P120	LKW68	A859
P120	LKW01	A225
P050	LKW99	A689

- Lossless-Join-Eigenschaft liegt nicht vor
- Relation befindet sich bereits in 5. Normalform

Ergebnis in 5. Normalform

<u>#FahrerID</u>	<u>#FahrzeugID</u>	<u>#FrachtID</u>
P010	LKW01	A225
P010	LKW99	A678
P010	LKW01	A689
P120	LKW68	A859
P120	LKW01	A225
P050	LKW99	A689

<u>FrachtID</u>	Frachtgut	Gefahren- klasse
A225	Flüssigeis	unkritisch
A678	Brennstäbe	sehr kritisch
A859	Forscherbier	ultrakritisch
A689	Kunstschnee	kritisch

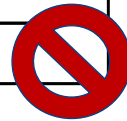


<u>FahrzeugID</u>	Fahrzeugbezeichnung	PS
LKW01	MB Actros	400
LKW02	MAN TGS	250
LKW10	Volvo FH16	380
LKW99	MB L-911	436
LKW68	MAN TGX	150

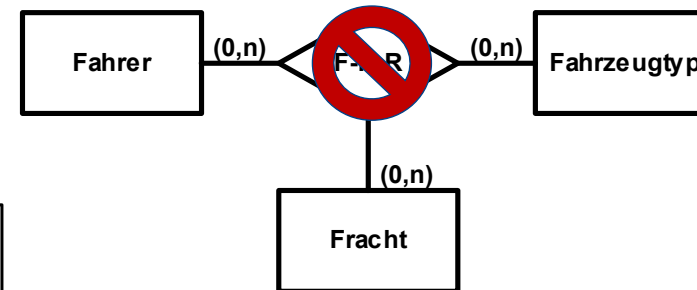
<u>FahrerID</u>	Fahrername	Alter
P010	Manfred Lebmann	40
P120	Wolfgang Taschenbier	38
P140	Karl-Heinz Böhmer	60
P050	Stanislaw Nowak	27

Was wäre wenn P010 mit LKW99 Fracht A689 fahren würde?

#FahrerID	#FahrzeugID	#FrachtID
P010	LKW01	A225
P010	LKW99	A678
P010	LKW01	A689
P010	LKW99	A689
P120	LKW68	A859
P120	LKW01	A225
P050	LKW99	A689



FrachtID	Frachtgut	Gefahren- klasse
A225	Flüssigeis	unkritisch
A678	Brennstäbe	sehr kritisch
A859	Forscherbier	ultrakritisch
A689	Kunstschnee	kritisch



FahrzeugID	Fahrzeugbezeichnung	PS
LKW01	MB Actros	400
LKW02	MAN TGS	250
LKW10	Volvo FH16	380
LKW99	MB L-911	436
LKW68	MAN TGX	150



FahrerID	Fahrername	Alter
P010	Manfred Lebmann	40
P120	Wolfgang Taschenbier	38
P140	Karl-Heinz Böhmer	60
P050	Stanislaw Nowak	27



Neues Ergebnis in 5. Normalform

#FahrzeugID	#FrachtID
LKW01	A225
LKW99	A678
LKW01	A689
LKW68	A859
LKW99	A689

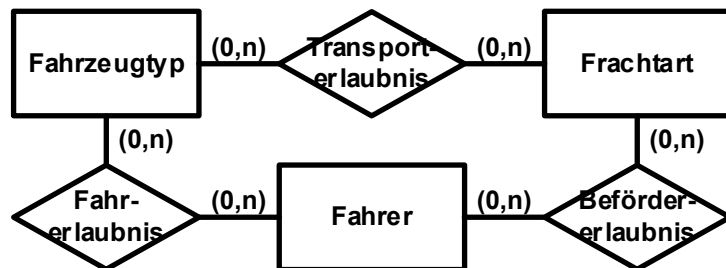
#FrachtID	#FahrerID
A225	P010
A678	P010
A689	P010
A859	P120
A225	P120
A689	P050

FrachtID	Frachtgut	Gefahrenklasse
A225	Flüssigeis	unkritisch
A678	Brennstäbe	sehr kritisch
A859	Forscherbier	ultrakritisch
A689	Kunstschnee	kritisch

FahrerID	Fahrername	Alter
P010	Manfred Lebmann	40
P120	Wolfgang Taschenbier	38
P140	Karl-Heinz Böhmer	60
P050	Stanislaw Nowak	

#FahrerID	#FahrzeugID
P010	LKW01
P010	LKW99
P120	LKW68
P120	LKW01
P050	LKW99

FahrzeugID	Fahrzeugbezeichnung	PS
LKW01	MB Actros	400
LKW02	MAN TGS	250
LKW10	Volvo FH16	380
LKW99	MB L-911	436
LKW68	MAN TGX	130



Und noch einmal!

- Es ist nicht immer sinnvoll, alle möglichen Normalisierungsschritte auch wirklich durchzuführen
- Vorteil
 - Vermeidung von Redundanz und Inkonsistenzen
- Nachteil
 - Komplexität
- Die Kenntnis der Normalisierungstheorie verbessert das Verständnis der attributbasierten Modellierung!