

Datenbankmanagement

Theorie

Prof. Dr. Gregor Hülksen

1	Einführung und Überblick
2	Modellierung
3	Normalisierung
3.1	Grundbegriffe Normalisierung
3.2	Normalformen (1-3)
3.3	Beispiel Normalisierung
4	Relationale Algebra
5	Lookup etc. in der Praxis
6	SQL – Data Definition Language
7	SQL – Data Manipulation Language
8	SQL – Trigger
9	SQL – Funktionen / Prozeduren
10	SQL – Datenschutz
11	Transaktionen

Definition: Normalisierung

Unter Normalisieren wird das Aufteilen der Daten in Relationen verstanden, wobei diese am Ende den Normalisierungsregeln entsprechen

Motivation

- Vermeidung von ungewollten Anomalien (Einfüge-, Änderungs- und Löschanomalien)
- Vermeidung von überflüssigen Informationen (Redundanzen)
- Zwang zum systematischen Datenbankentwurf (Bessere Übersichtlichkeit)

Vorgehensweise

- Analyse der Datenstruktur
- Schrittweise Überführung in die n-te Normalform
- Verlustfreie und Abhängigkeitsstreuende Zerlegung der Relationen
- Eliminierung von Redundanzen
- Eliminierung von Anomalien
- Bereinigung von Inkonsistenzen

Anomalien

Einfügeanomalie: Liegt vor, wenn die Eingabe einer Information zu einem Teilsachverhalt nicht angelegt werden kann

Änderungsanomalie: Liegt vor, wenn nicht alle (redundanten) Vorkommen einer Entität zugleich geändert werden.

Löschanomalie: Liegt vor, wenn durch das Löschen eines Datensatzes mehr Informationen als erwünscht verloren gehen.

ID	Name	Abteilung_ID	Abteilung
1	Brandt	1	Verkauf
2	Meier	2	Marketing
3	Kohl	2	Marketing
4	Schröder	3	Werkstatt

Verlustfreie Zerlegung / Definition

Man spricht von einer verlustfreien Zerlegung, wenn sie sich durch Join – Operationen rückgängig machen lässt, ohne das dabei zusätzliche Datensätze entstehen.

Motivation:

- Eine verlustfreie Zerlegung sichert die Wiederherstellbarkeit einer ursprünglichen Relation
- Mit Hilfe dieser verlustfreien Zerlegung werden auch die genannten Redundanzen der Daten beseitigt.

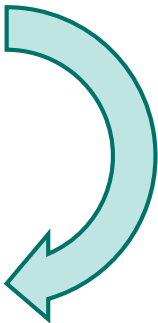
Verlustfreie Zerlegung 1 - Beispiel: Nicht verlustfreie Zerlegung

Student	LV	SGebiet
123321	STAT	BAU
123321	DBS	ARC
125674	STAT	ARC

Student	LV	LV	SGebiet
123321	STAT	STAT	BAU
123321	DBS	DBS	ARC
125674	STAT	STAT	ARC

Student	LV	SGebiet
123321	STAT	BAU
123321	STAT	ARC
123321	DBS	ARC
125674	STAT	ARC
125674	STAT	BAU

Ausgangsrelation



Zerlegung



Zusammenführung

Datensätze nicht
in der
Ausgangsrelation

Verlustfreie Zerlegung 2 - Beispiel: Verlustfreie Zerlegung

Vater	Mutter	Kind
Thomas	Stefanie	Frank
Thomas	Elke	Max
Markus	Petra	Tom

Vater	Kind	Mutter	Kind
Thomas	Frank	Stefanie	Frank
Thomas	Max	Elke	Max
Markus	Tom	Petra	Tom

Vater	Mutter	Kind
Thomas	Stefanie	Frank
Thomas	Elke	Max
Markus	Petra	Tom

Ausgangsrelation



Zerlegung

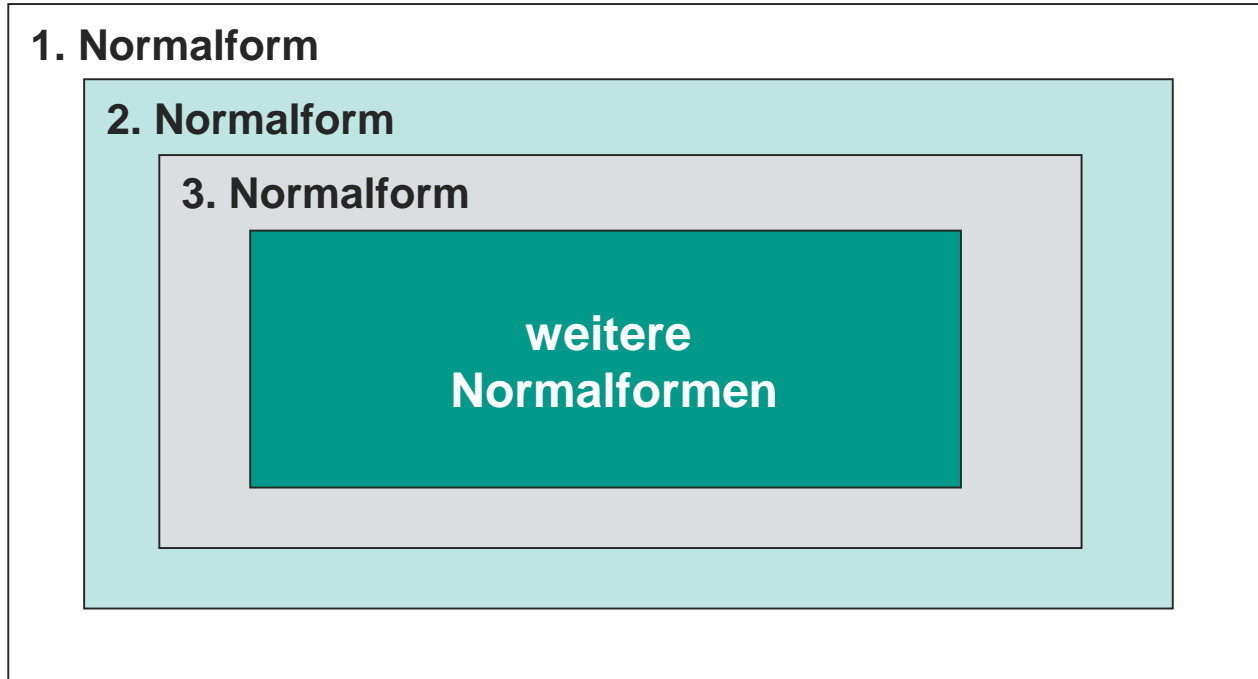


Zusammenführung

Ergebnis
entspricht der
Ausgangsrelation

3.2 Normalformen (1-3)

Normalformen



1. Normalform (1NF)

Eine Relation ist in der 1. Normalform, wenn der Wert eines jeden Attributes atomar ist

Atomare Werte bedeuten:

- keine Wertauflistungen
z.B. Management, Verkauf, Entwicklung
- keine zusammengesetzte Werte
z.B. Domagkstrasse, 48301 Münster

1. Normalform (1NF)

Beispiel:

Eine nicht normalisierte Relation

ID	Name	Adresse	Abteilung
1	Müller, Manfred	Dortmunder Str. 12, 48150 Münster	1,3,6

zusammengesetzte Werte

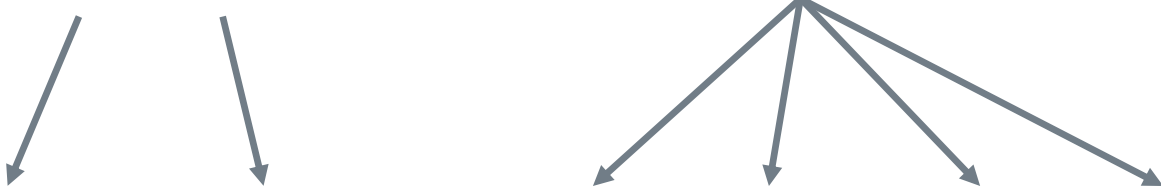
Werteauflählungen

1. Normalform (1NF)

Lösung: bei zusammengesetzten Werten

ID	Name		Adresse				Abteilung
1	Müller, Manfred		Dortmunder Str. 12, 48150 Münster				1,3,6


ID	Name	Vorname	Strasse	HausNr	PLZ	Ort	Abteilung
1	Müller	Manfred	Dortmunder Str.	12	48150	Münster	1,3,6



1. Normalform (1NF)


Lösung: bei zusammengesetzten Werten

ID	Name	Adresse	Abteilung
1	Müller, Manfred	Dortmunder Str. 12, 48150 Münster	1,3,6



ID	Name	Vorname	Strasse	HausNr	PLZ	Ort	Abteilung
1	Müller	Manfred	Dortmunder Str.	12	48150	Münster	1,3,6

(1NF)



ID	Name	Vorname	Strasse	HausNr	PLZ	Ort
1	Müller	Manfred	Dortmunder Str.	12	48150	Münster

ID	Abteilung
1	1
1	3
1	6

Funktionale Abhängigkeiten

Definition:

- X und Y seien zwei Teilmengen aus Attributen einer Relation R .
- Y heißt funktional abhängig von X , wenn folgendes gilt:

Für alle Tupel r, s aus R gilt:

aus $\text{Proj}X(s) = \text{Proj}X(r)$ folgt stets: $\text{Proj}Y(s) = \text{Proj}Y(r)$

- Die funktionale Abhängigkeit wird mit $X \rightarrow Y$ abgekürzt geschrieben, gelesen wird es als X bestimmt Y .

$\text{Proj}X(s)$ bezeichnet die Projektion von s auf X

Funktionale Abhängigkeiten

Beispiel:

ISBN	Titel	Autor	Verlag
0-1236-56	DBS	Müller	Mayer
0-1236-56	DBS	Franz	Mayer
0-1236-56	DBS	Werner	Mayer
0-1236-22	Das Märchenbuch	Grimm	Der Hörverlag
0-1236-22	Das Märchenbuch	Mahler	Der Hörverlag

FA: ISBN → Titel, Verlag

Keine FA: ISBN → Autor

Schlüssel

Eindeutiger Schlüssel

Eine Attributkombination L wird ein eindeutiger Schlüssel der Relation $R1(A_1, \dots, A_n)$ genannt, wenn

1. $L \rightarrow (A_1, \dots, A_n)$
2. (A_1, \dots, A_n) von keiner Teilmenge von L funktional abhängig ist

KundenNr	Name	Vorname	Geschlecht	Ort	Telefon
1	Müller	Andreas	M	Münster	0251-558863
2	Meier	Andreas	M	Münster	0251-489654
3	Müller	Karl	M	Köln	0221-563214
4	Wal	Silke	W	München	030-6698745

↑
Eindeutiger
Schlüssel

↑
Eindeutiger
Schlüssel

↑
Eindeutiger
Schlüssel

Schlüsselbegriffe

AuftragNr	Typ	KundenNr	AngNr	Bezahlte	Lieferdatum
1	Auftrag	1	2	01.02.2005	19.02.2005
2	Auftrag	2	22	01.03.2006	20.02.2006
3	Angebot	3	32		
4	Auftrag	4	4	08.06.2007	05.06.2007

↑
Primärschlüssel

↑
Zweitschlüssel

↖ ↗
Fremdschlüssel

2. Normalform (2NF)

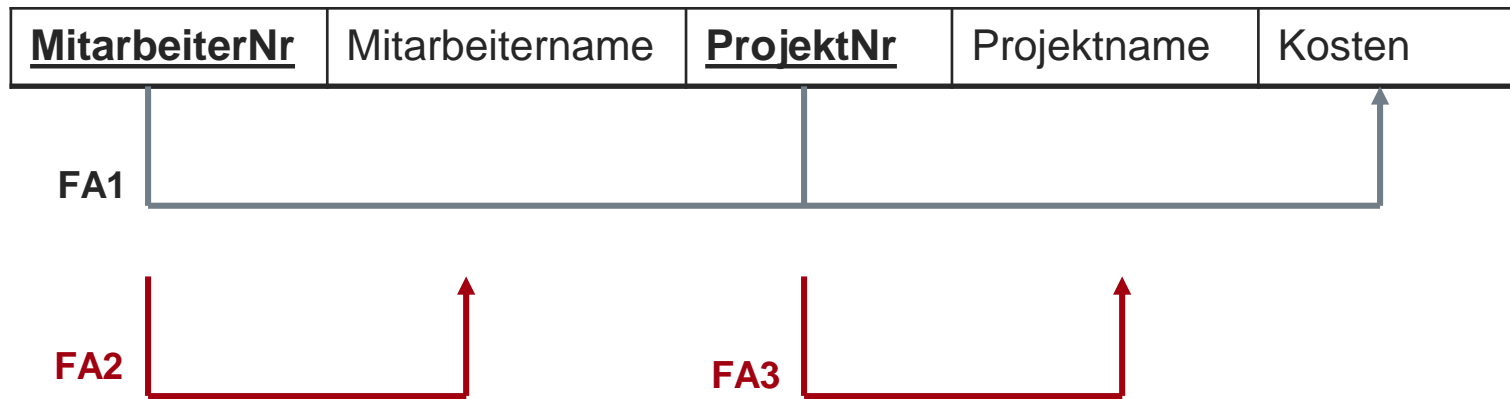
Eine Relation ist in der 2. Normalform, wenn sie sich in der ersten NF befindet und jedes Nichtschlüsselattribut von seinem gesamten Primärschlüssel voll funktional abhängig (FA) ist.

Anmerkung

- Falls kein zusammengesetzter Primärschlüssel existiert, befindet sich die Relation in der 2NF (wenn die Relation sich in der 1NF befindet)
- Das Ziel ist es, dass Nicht-Schlüssel-Attribute nur vom ganzen Schlüssel abhängen und nicht bereits schon von einem Schlüsselteil

2. Normalform (2NF)

Beispiel: Relation Projekt

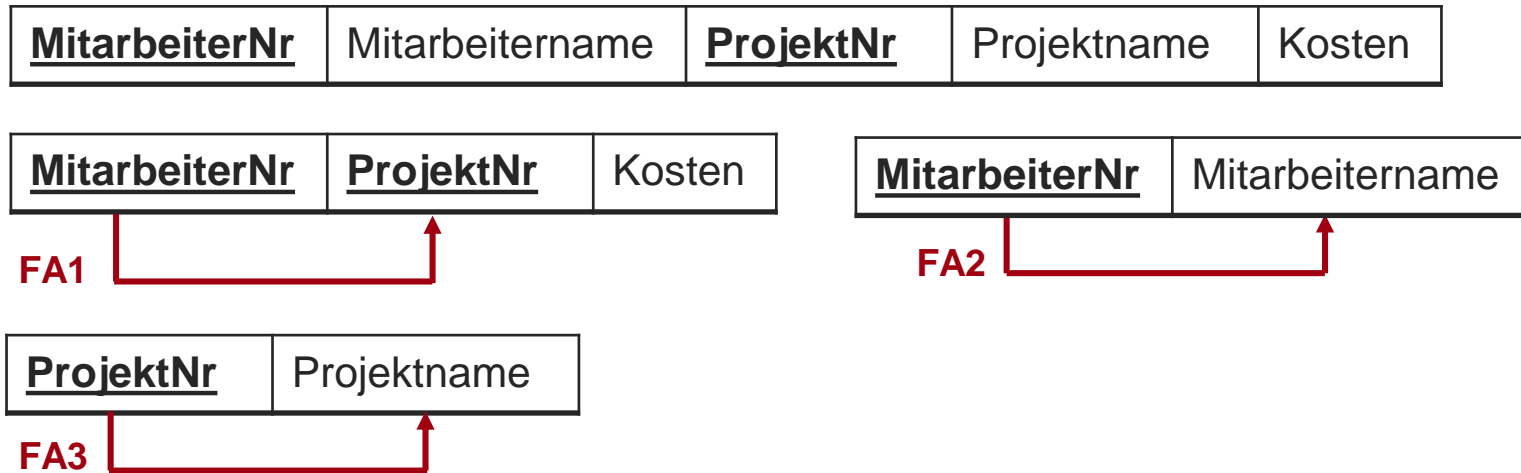


- Mitarbeitername ist abhängig von TS MitarbeiterNr (FA2)
- Projektname ist nur abhängig von TS ProjektNr (FA3)

Die gegebene Relation ist NICHT in der 2. Normalform

2. Normalform (2NF)

Lösung: Relation Projekt



Die von Schlüsselteilen funktional abhängigen Attribute werden zusammen mit dem Schlüssel jeweils in eine neue Relation überführt

Die gegebene Relation ist jetzt in der 2. Normalform

Übung: 2. Normalform (2NF)

Überprüfen Sie folgende Relationen, ob die 2NF vorliegt und überführen Sie die Relation gegebenenfalls in die 2 NF.

1. Lieferantenpreise (**WarenNr**, **LieferantenNr**, Artikelname, Preis)
Hinweis: Lieferant liefert Artikel zu bestimmten Preisen
2. Angestellte (**AngNr**, Name, Vorname, AbtNr, AbtName)
Hinweis: Mitarbeiter sind einer best. Abteilung zugeordnet
3. Bestellungen (**BestellNr**, **ArtikelNr**, KundenNr, Sachbearbeiter, Datum, Menge)
Hinweis: Kunden geben Artikelbestellungen auf

Lösung

1. Lieferantenpreise (ANr, LieferantenNr, Artikelname, Preis)

→ Lösung: Keine 2NF da Artikelname nur von ANr abhängt.
(ANr → Artikelname)

→ Zerlegung in: Lieferantenpreise (ANr, LieferantenNr, Preis)

Artikel (ANr, Artikelname)

2. Angestellte (AngNr, Name, Vorname, AbtNr, AbtName)

Lösung: 2NF liegt vor weil es keinen zusammengesetzten Schlüssel gibt

3. Bestellungen (BestellNr, ArtikelNr, KundenNr, Name, Datum, Menge)

Lösung: Keine 2NF da KundenNr, Datum nur von BestellNr abhängt.
(BestellNr → KundenNr, Name, Datum)

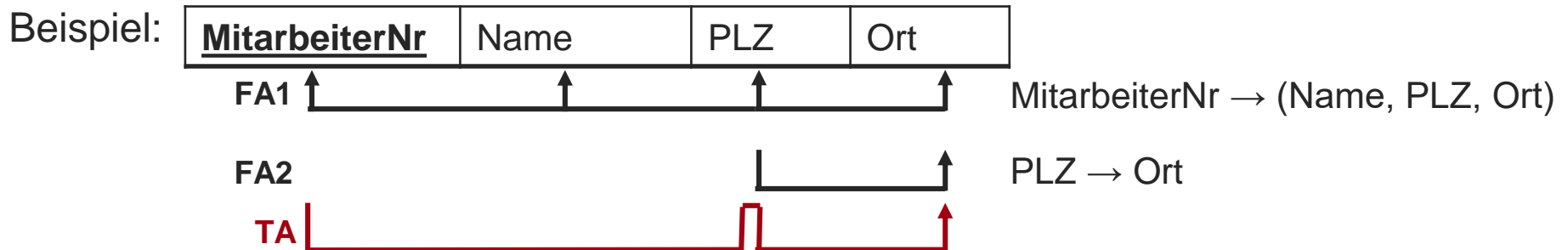
Zerlegung in: Bestellungen (BestellNr, KundenNr, Name, Datum)
Positionen (BestellNr, ArtikelNr, Menge)

Transitive Abhängigkeit

Ein Attribut C ist transitiv vom Primärschlüssel A abhängig, wenn ein Nichtschlüssel-Attribut B existiert, von dem C funktional abhängt, wobei gleichzeitig B von A funktional in Abhängigkeit stehen muss.

Schreibweise: $A \rightarrow B$ (B ist funktional abhängig von A) und
 $B \rightarrow C$ (C ist funktional abhängig von B)
 $\{A \rightarrow B, B \rightarrow C\} \Rightarrow A \rightarrow C$

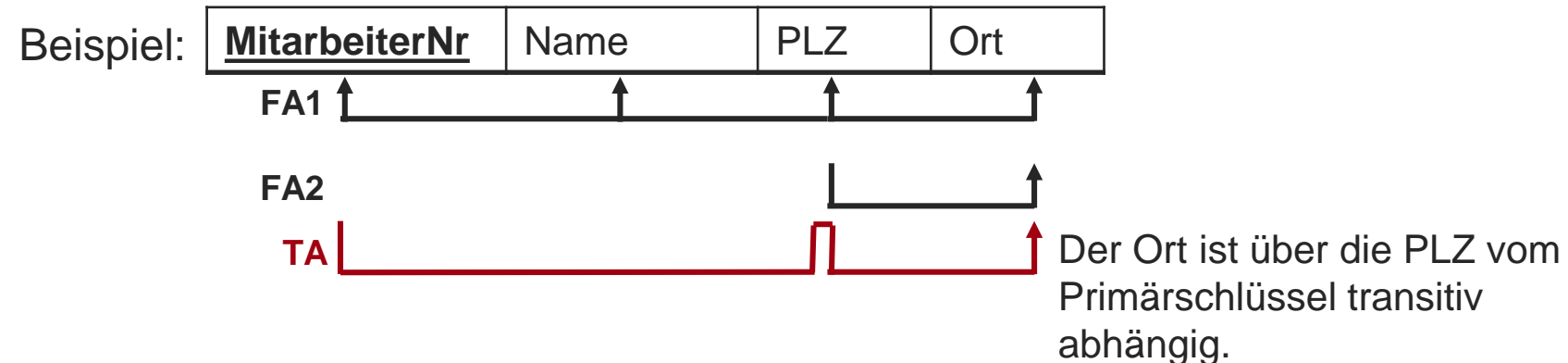
Hinweis: Ein Relationstyp kann nur in diesem Falle transitive Abhängigkeiten aufweisen, wenn neben dem Primärschlüssel mindestens zwei Nichtschlüsselattribute existieren, wovon eines von dem anderen funktional abhängt.



3. Normalform (3NF)

Eine Relation ist in der 3. Normalform, wenn die Restriktionen der 2. Normalform erfüllt sind und kein Nichtschlüsselattribut transitiv von dem Primärschlüssel abhängt.

Hinweis: Ziel der 3NF ist es, dass sich Nonprimattribute immer unmittelbar von dem Primärschlüssel ableiten lassen. Ein Ableitungsumweg über ein weiteres Nonprimattribut ist nach der Regel nicht zulässig



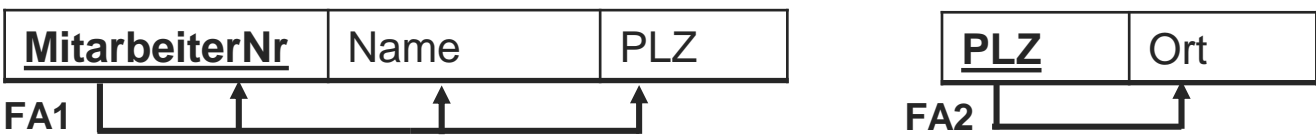
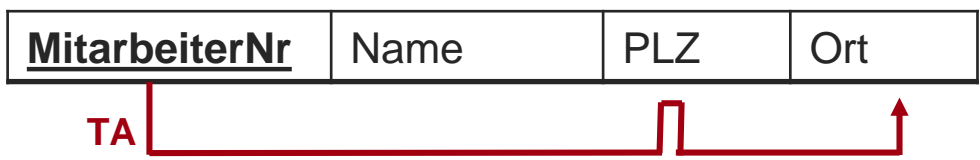
Die gegebene Relation befindet sich in der 2. NF, nicht aber in der 3. NF

3. Normalform (3NF)

Normalisierung:

Das Transitiv abhängige Attribut (Ort) wird mit dem determinierendem Attribut PLZ in eine weitere Tabelle überführt

Lösung:



Die gegebene Relation befindet sich jetzt in der 3. NF

Übung 3. Normalform (3NF)

Überprüfen Sie für die folgenden Relationen, ob die 3NF vorliegt und überführen Sie diese gegebenenfalls in die 3NF

1. Aufträge (**AuftragsNr**, KundenNr, Name, Datum)
Hinweis: Kunden initiieren Aufträge
2. Versicherte (**PersonenNr**, **VersicherungsNr**, Vertragsdatum)
Hinweis: Versicherte haben eindeutige Personen, bzw.
Versicherungsnummern, werden zu einem bestimmten Datum
abgeschlossen für einen Versicherten abgeschlossen

Exkurs: 4 Normalform (4NF)

Eine Relation ist in der 4 NF, wenn sie in Boyce-Codd Normalform ist und für jede mehrwertige Abhängigkeit einer Attributmenge Y von einer Attributmenge X gilt:

- Die mehrwertige Abhängigkeit ist trivial oder
- X ist ein Schlüsselkandidat der Relation

<u>MitarbeiterNr</u>	<u>Telefonnummer</u>
----------------------	----------------------

<u>MitarbeiterNr</u>	<u>Telefonnummer</u>	<u>Autokennzeichen</u>
----------------------	----------------------	------------------------

MitarbeiterNr	Telefonnummer	Autokennzeichen
1	0251-12345	MS-A-1
1	0251-123456	MS-A-2

MitarbeiterNr	Telefonnummer	Autokennzeichen
1	0251-12345	NULL
1	0251-123456	NULL
1	NULL	MS-A-1
1	NULL	MS-A-2

MitarbeiterNr	Telefonnummer	Autokennzeichen
1	0251-12345	MS-A-1
1	0251-123456	MS-A-1
1	0251-12345	MS-A-2
1	0251-123456	MS-A-2

Exkurs: 5. Normalform (5NF)

Eine Relation R ist in der 5NF (oder Project-Join-Normalform (PJNF)), wenn sie in der 4NF ist und für jede Join-Abhängigkeit (R1, R2, ..., Rn) gilt:

- Die Join-Abhängigkeit ist trivial oder
- Jedes Ri aus (R1, R2, ..., Rn) ist Schlüsselkandidat der Relation

Studiengang	Modul	Name
BBA	Controlling	Schmidt
BWI	IT-Basics	Schmidt
BBA	Controlling	Meier
BBA	Ökonomie	Müller



R1	
Studiengang	Modul
BBA	Controlling
BWI	IT-Basics
BBA	Ökonomie

R2	
Studiengang	Name
BBA	Schmidt
BWI	Schmidt
BBA	Meier
BBA	Müller

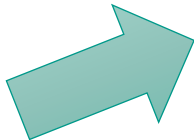


Join von R1 und R2		
Studiengang	Modul	Name
BBA	Controlling	Schmidt
BBA	Controlling	Meier
BBA	Controlling	Müller
BWI	IT-Basics	Schmidt
BBA	Ökonomie	Schmidt
BBA	Ökonomie	Meier
BBA	Ökonomie	Müller

Exkurs: 5. Normalform (5NF)

R3	
Modul	Name
Controlling	Schmidt
IT-Basics	Schmidt
Controlling	Meier
Ökonomie	Müller

Join von R1 und R2		
Studiengang	Modul	Name
BBA	Controlling	Schmidt
BBA	Controlling	Meier
BBA	Controlling	Müller
BWI	IT-Basics	Schmidt
BBA	Ökonomie	Schmidt
BBA	Ökonomie	Meier
BBA	Ökonomie	Müller



Join von R3 mit (Join von R1 und R2)		
Studiengang	Modul	Name
BBA	Controlling	Schmidt
BWI	IT-Basics	Schmidt
BBA	Controlling	Meier
BBA	Ökonomie	Müller



Studiengang	Modul

Studiengang	Name

Modul	Name

Zusammenfassung

Vorgehen bei Normalisierung

1. Prüfen, ob alle Attribute atomar sind → 1NF
2. Alle funktionalen Abhängigkeiten feststellen
3. Schlüssel und Nichtschlüssel-Attribute identifizieren
4. Prüfen, ob zusammengesetzte Schlüssel existieren → 2NF
5. Alle Abhängigkeiten zwischen Nichtschlüsselattributen entfernen → 3NF

3.3 Beispiel Normalisierung

Erster Entwurf für die Datenbank eines Obsthändlers

AuftragNr	Datum	Kunde	ArtikelNr	Bezeichnung	Menge	Bedienung
1	10.02.2012	1, Müller, Münster	1	Melone	1 Stück	2, Schmidt
1	10.02.2012	1, Müller, Münster	2	Bananen	2 Kilo	1, Maier
2	10.02.2012	2, Mayer, Hamburg	3	Paprika, Grün	0,5 Kilo	3, Kraus
2	10.02.2012	2, Meier, Hamburg	2	Bananen	2 Kilo	2, Schmidt
2	10.02.2012	2, Meier, Hamburg	23	Boskop	1 Kisten	1, Maier
3	11.02.2012	3, Franz, Roxel	3	Paprika, Rot	0,5 Kilo	4, Kaltmeier
3	11.02.2012	3, Franz, Roxel	102	Ananas	1 Stück	2, Schmidt
4	13.02.2012	1, Müller, Münster	23	Boskop	3 Kisten	1, Maier

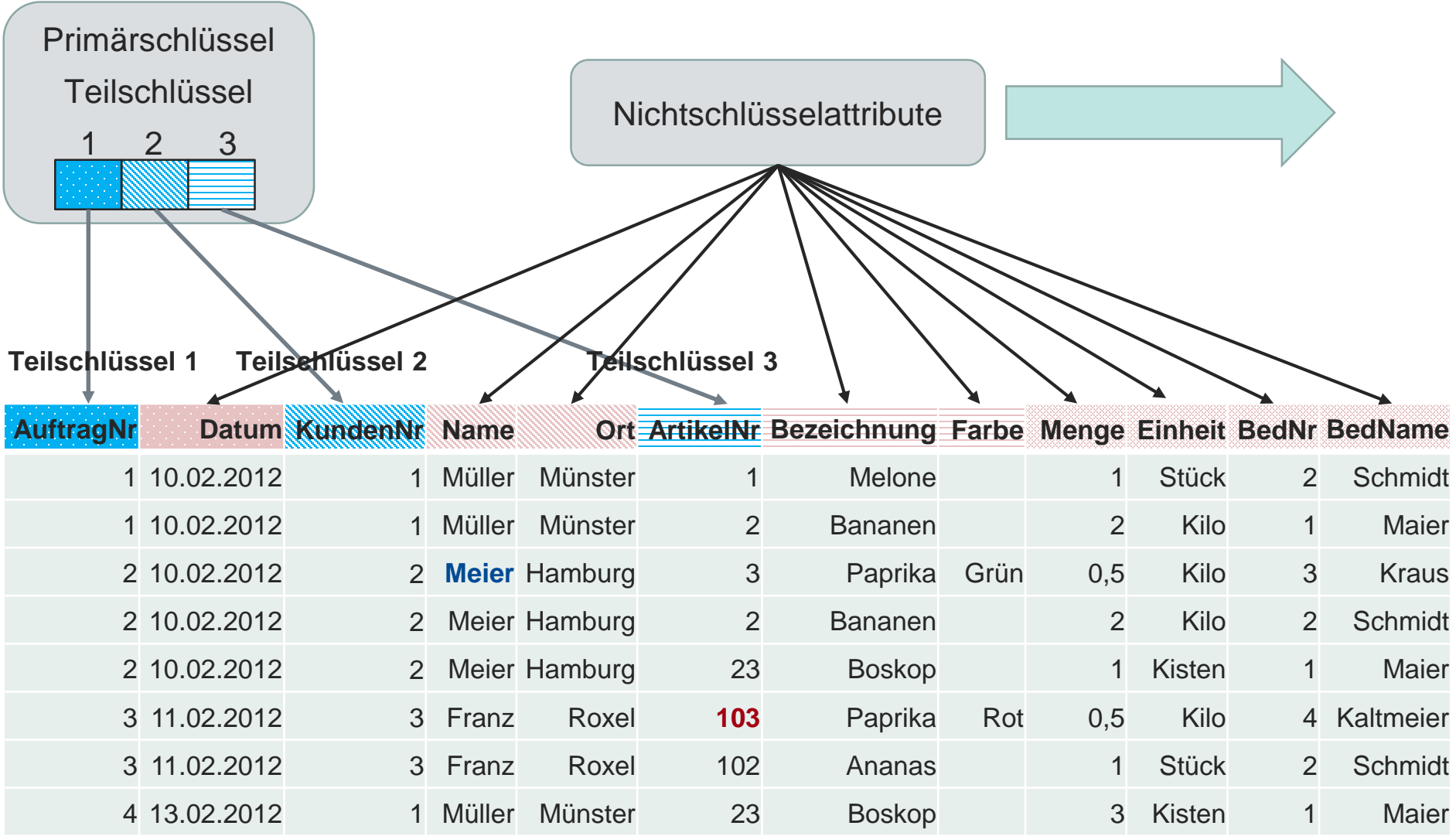
- Viele gleiche Einträge in den Spalten sorgt dafür, das die Datenbank schnell sehr groß wird
- Tippfehler beim Namen macht das auffinden des Kunden schwerer
- Die Spalte Kunde enthält zusammengesetzte Werte
- Artikelnummer 3 enthält 2 unterschiedliche Bezeichnungen
- Die Spalte Menge enthält die Stückzahl sowie die dazugehörige Einheit der Ware.

Maßnahmen zum Erreichen der ersten Normalform

- Die Spalte Kunde wird in KundeNr, Name und Ort aufgeteilt
- Die Spalte Bezeichnung wird in Bezeichnung und Farbe aufgeteilt
- Die Spalte Menge wird in Menge und Einheit aufgeteilt
- Tippfehler werden behoben (Meyer → Meier)
- Für die fehlerhafte ArtikelNr für Paprika (gleiche ArtikelNr für Rote und Grüne) wurde eine weitere ArtikelNr eingefügt (103)
- Bedienung wird in Bedienung Nr. und Bedienung Name aufgeteilt

AuftragNr	Datum	KundenNr	Name	Ort	ArtikelNr	Bezeichnung	Farbe	Menge	Einheit	BedNr	BedName
1	10.02.2012	1	Müller	Münster	1	Melone		1	Stück	2	Schmidt
1	10.02.2012	1	Müller	Münster	2	Bananen		2	Kilo	1	Maier
2	10.02.2012	2	Meier	Hamburg	3	Paprika	Grün	0,5	Kilo	3	Kraus
2	10.02.2012	2	Meier	Hamburg	2	Bananen		2	Kilo	2	Schmidt
2	10.02.2012	2	Meier	Hamburg	23	Boskop		1	Kisten	1	Maier
3	11.02.2012	3	Franz	Roxel	103	Paprika	Rot	0,5	Kilo	4	Kaltmeier
3	11.02.2012	3	Franz	Roxel	102	Ananas		1	Stück	2	Schmidt
4	13.02.2012	1	Müller	Münster	23	Boskop		3	Kisten	1	Maier

Sämtliche Anomalien sind weiterhin vorhanden!



Ergebnis: zweite Normalform

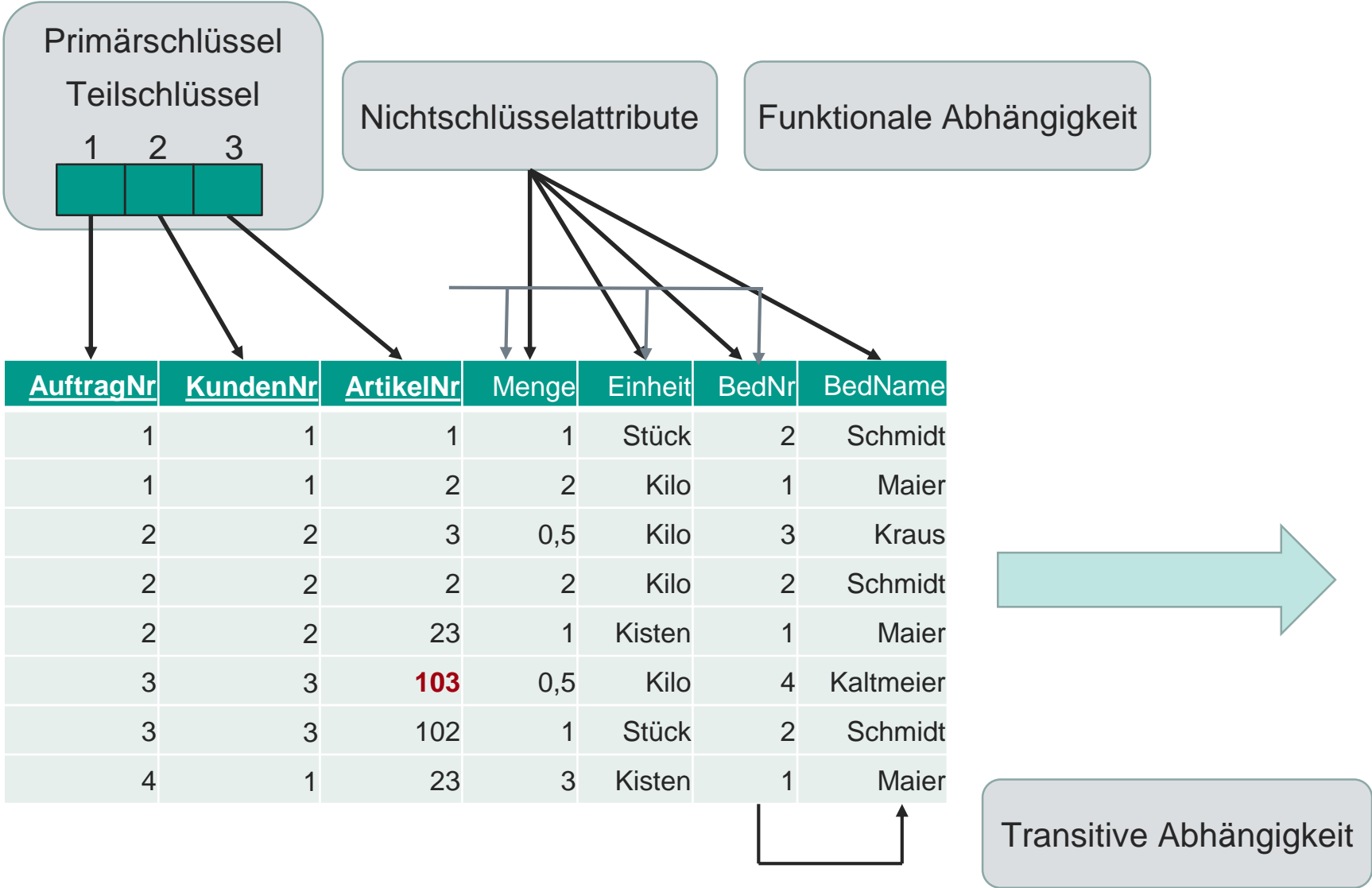
AuftragNr	Datum
1	10.02.2012
2	10.02.2012
3	11.02.2012
4	13.02.2012

KundenNr	Name	Ort
1	Müller	Münster
2	Meier	Hamburg
3	Franz	Roxel

ArtikelNr	Bezeichnung	Farbe
1	Melone	
2	Bananen	
3	Paprika	Grün
23	Boskop	
102	Ananas	
103	Paprika	Rot

AuftragNr	KundenNr	ArtikelNr	Menge	Einheit	BedNr	BedName
1	1	1	1	Stück	2	Schmidt
1	1	2	2	Kilo	1	Maier
2	2	3	0,5	Kilo	3	Kraus
2	2	2	2	Kilo	2	Schmidt
2	2	23	1	Kisten	1	Maier
3	3	103	0,5	Kilo	4	Kaltmeier
3	3	102	1	Stück	2	Schmidt
4	1	23	3	Kisten	1	Maier





Ergebnis: dritte Normalform

<u>AuftragNr</u>	Datum
1	10.02.2012
2	10.02.2012
3	11.02.2012
4	13.02.2012

<u>KundenNr</u>	Name	Ort
1	Müller	Münster
2	Meier	Hamburg
3	Franz	Roxel

<u>ArtikelNr</u>	Bezeichnung	Farbe
1	Melone	
2	Bananen	
3	Paprika	Grün
23	Boskop	
102	Ananas	
103	Paprika	Rot

<u>AuftragNr</u>	<u>KundenNr</u>	<u>ArtikelNr</u>	Menge	Einheit	BedNr
1	1	1	1	Stück	2
1	1	2	2	Kilo	1
2	2	3	0,5	Kilo	3
2	2	2	2	Kilo	2
2	2	23	1	Kisten	1
3	3	103	0,5	Kilo	4
3	3	102	1	Stück	2
4	1	23	3	Kisten	1

<u>BedNr</u>	BedName
1	Maier
2	Schmidt
3	Kraus
4	Kaltmeier

Normalisierung**Übung**

Gegeben ist folgenden Relation:

Bootsname	Segelfläche	Besatzung	Name	Start	Ziel	Länge
Skipper	50	4	KielCup	Lübeck	Kiel	200
Skipper	50	3	Ostseepokal	Rostock	Bornholm	180
Ariane	35	3	KielCup	Lübeck	Kiel	200
Ariane	35	4	Ostseepokal	Rostock	Bornholm	180
Ariane	35	2	Spreepokal	Lübeck	Kiel	200

- In welcher Normalform befindet sich die Relation?
- Überführen Sie die Relation ggf. in die 3NF

Physikalische Datenbankentwicklung

- Spezifisch für ein Datenbanksystem (MySQL, MS SQL, Oracle...)
- Optimierungsmöglichkeiten für Datenzugriffe (z. B. durch Index-Definitionen) einstellen
- Formulieren der Scripte / Kommandos zum Einrichten und Konfigurieren der Datenbank (in der Syntax des DBMS)
- Festlegungen zur Datensicherung