

Normalisierung

Mehrstufiger Verbesserungsprozess mit dem Ziel der Gewinnung „gutmütiger“ Relationen.

- Formale Betrachtung von Datenmengen und ihrer semantischen Zusammenhänge (\rightarrow Codd)
- Analyse der Abhängigkeiten zwischen den semantischen Inhalten (Attributen) einer Relation
- Eliminierung unerwünschter Abhängigkeiten durch Aufspaltung der Relationen

4 Ziel der Normalisierung

- Vermeidung/Eliminierung von Redundanzen
- Vermeidung/Eliminierung von Anomalien
 - Änderungsanomalie
 - Einfügeanomalie
 - Löschanomalie

- Gefahr des Auftretens inkonsistenter Datenbankzustände
 - Fehlinformationen
 - Fehlentscheidungen
 - Akzeptanzverlust des Informationssystems
 - Kollaps des Informationssystems
- Hoher Pflegeaufwand zum Sicherstellen bzw. Wiederherstellen eines konsistenten Zustandes

4

Beispiel zu Mutationsanomalien

Ändern des Titel ‚Wirtschaftsinformatik‘ in ‚Grundlagen der Wirtschaftsinformatik‘

<u>Sign</u>	<u>Autor</u>	Titel	Jahr	SID	Sachgebiet	<u>Schlagwort</u>
QH1	Ferstl	Wirtschaftsinformatik	2013	GRD	Grundlagen	Informationssysteme
QH1	Sinz	Wirtschaftsinformatik	2013	GRD	Grundlagen	Informationssysteme
QH1	Ferstl	Wirtschaftsinformatik	2013	GRD	Grundlagen	Objektmodellierung
QH1	Sinz	Wirtschaftsinformatik	2013	GRD	Grundlagen	Objektmodellierung
QH2	Wirth	Modula-2	2004	PRG	Programmierung	Algorithmus
QH2	Wirth	Modula-2	2004	PRG	Programmierung	Modul
QH3	Wirth	Pascal	2007	PRG	Programmierung	Algorithmus

Ändern des Titel ‚Wirtschaftsinformatik‘ in ‚Grundlagen der Wirtschaftsinformatik‘

<u>Sign</u>	<u>Autor</u>	<u>Titel</u>	<u>Jahr</u>	<u>SID</u>	<u>Sachgebiet</u>	<u>Schlagwort</u>
QH1	Ferstl	Grundlagen der Wirtschaftsinformatik	2013	GRD	Grundlagen	Informationssysteme
QH1	Sinz	Grundlagen der Wirtschaftsinformatik	2013	GRD	Grundlagen	Informationssysteme
QH1	Ferstl	Grundlagen der Wirtschaftsinformatik	2013	GRD	Grundlagen	Objektmodellierung
QH1	Sinz	Wirtschaftsinformatik	2013	GRD	Grundlagen	Objektmodellierung
QH2	Wirth	Modula-2	2004	PRG	Programmierung	Algorithmus
QH2	Wirth	Modula-2	2004	PRG	Programmierung	Modul
QH3	Wirth	Pascal	2007	PRG	Programmierung	Algorithmus

Änderungsanomalie: Der Titel muss an mehreren Stellen geändert werden.

Problematisch: Kleine Fehler führen dazu, dass Änderungen vergessen werden.

Ziel: Wir wollen die Liste der Schlagworte um „Datenstruktur“ erweitern.

<u>Sign</u>	<u>Autor</u>	Titel	Jahr	SID	Sachgebiet	<u>Schlagwort</u>
QH1	Ferstl	Wirtschaftsinformatik	2013	GRD	Grundlagen	Informationssysteme
QH1	Sinz	Wirtschaftsinformatik	2013	GRD	Grundlagen	Informationssysteme
QH1	Ferstl	Wirtschaftsinformatik	2013	GRD	Grundlagen	Objektmodellierung
QH1	Sinz	Wirtschaftsinformatik	2013	GRD	Grundlagen	Objektmodellierung
QH2	Wirth	Modula-2	2004	PRG	Programmierung	Algorithmus
QH2	Wirth	Modula-2	2004	PRG	Programmierung	Modul
QH3	Wirth	Pascal	2007	PRG	Programmierung	Algorithmus

Ziel: Wir wollen die Liste der Schlagworte um „Datenstruktur“ erweitern.

<u>Sign</u>	<u>Autor</u>	Titel	Jahr	SID	Sachgebiet	<u>Schlagwort</u>
QH1	Ferstl	Wirtschaftsinformatik	2013	GRD	Grundlagen	Informationssysteme
QH1	Sinz	Wirtschaftsinformatik	2013	GRD	Grundlagen	Informationssysteme
QH1	Ferstl	Wirtschaftsinformatik	2013	GRD	Grundlagen	Objektmodellierung
QH1	Sinz	Wirtschaftsinformatik	2013	GRD	Grundlagen	Objektmodellierung
QH2	Wirth	Modula-2	2004	PRG	Programmierung	Algorithmus
QH2	Wirth	Modula-2	2004	PRG	Programmierung	Modul
QH3	Wirth	Pascal	2007	PRG	Programmierung	Algorithmus
QH3	Wirth	Pascal	2007	PRG	Programmierung	Datenstruktur

Einfügeanomalie: Möchte man ein neues Schlagwort hinzunehmen, so muss auch gleich dazu ein Buch hinzugefügt werden.

4

Beispiel zu Mutationsanomalien

Löschen des Buches aus dem Bestand.

<u>Sign</u>	<u>Autor</u>	Titel	Jahr	SID	Sachgebiet	<u>Schlagwort</u>
QH1	Ferstl	Grundlagen der Wirtschaftsinformatik	2013	GRD	Grundlagen	Informationssysteme
QH1	Sinz	Grundlagen der Wirtschaftsinformatik	2013	GRD	Grundlagen	Informationssysteme
QH1	Ferstl	Grundlagen der Wirtschaftsinformatik	2013	GRD	Grundlagen	Objektmodellierung
QH1	Sinz	Grundlagen der Wirtschaftsinformatik	2013	GRD	Grundlagen	Objektmodellierung
QH2	Wirth	Modula-2	2004	PRG	Programmierung	Algorithmus
QH2	Wirth	Modula-2	2004	PRG	Programmierung	Modul
QH3	Wirth	Pascal	2007	PRG	Programmierung	Algorithmus
QH3	Wirth	Pascal	2007	PRG	Programmierung	Datenstruktur

4

Beispiel zu Mutationsanomalien

Löschen des Buches aus dem Bestand.

<u>Sign</u>	<u>Autor</u>	Titel	Jahr	SID	Sachgebiet	<u>Schlagwort</u>
QH1	Ferstl	Grundlagen der Wirtschaftsinformatik	2013	GRD	Grundlagen	Informationssysteme
QH1	Sinz	Grundlagen der Wirtschaftsinformatik	2013	GRD	Grundlagen	Informationssysteme
QH1	Ferstl	Grundlagen der Wirtschaftsinformatik	2013	GRD	Grundlagen	Objektmodellierung
QH1	Sinz	Grundlagen der Wirtschaftsinformatik	2013	GRD	Grundlagen	Objektmodellierung
QH2	Wirth	Modula-2	2004	PRG	Programmierung	Algorithmus
QH2	Wirth	Modula-2	2004	PRG	Programmierung	Modul
QH3	Wirth	Pascal	2007	PRG	Programmierung	Algorithmus

Löschanomalie: Es gehen mehr Informationen verloren als beabsichtigt ist.
Das Schlagwort „Datenstruktur“ existiert nicht mehr.

4 Literaturempfehlung (Normalisierung)

Lackes, R., Siepermann, M.:
Wohlstrukturiertheit von Daten in betrieblichen
Informationssystemen. In : wisu – das wirtschafts-
studium, Zeitschrift für Ausbildung, Examen,
Berufseinstieg und Fortbildung, 32. Jahrgang, Heft 6,
Düsseldorf 2003, S. 787-794.

4 Schlüsselbegriffe

Identifikator

Ein Identifikator ist ein Attribut oder eine Attributkombination, welche jedes einzelne Element einer Relation eindeutig identifiziert.



Schlüsselkandidat

Ein Schlüsselkandidat ist ein Identifikator, dessen Attributkombination minimal ist, d.h. keine Attributteilmenge des Schlüsselkandidaten ist ebenfalls ein Identifikator.



4 Schlüsselbegriffe

Primärschlüssel

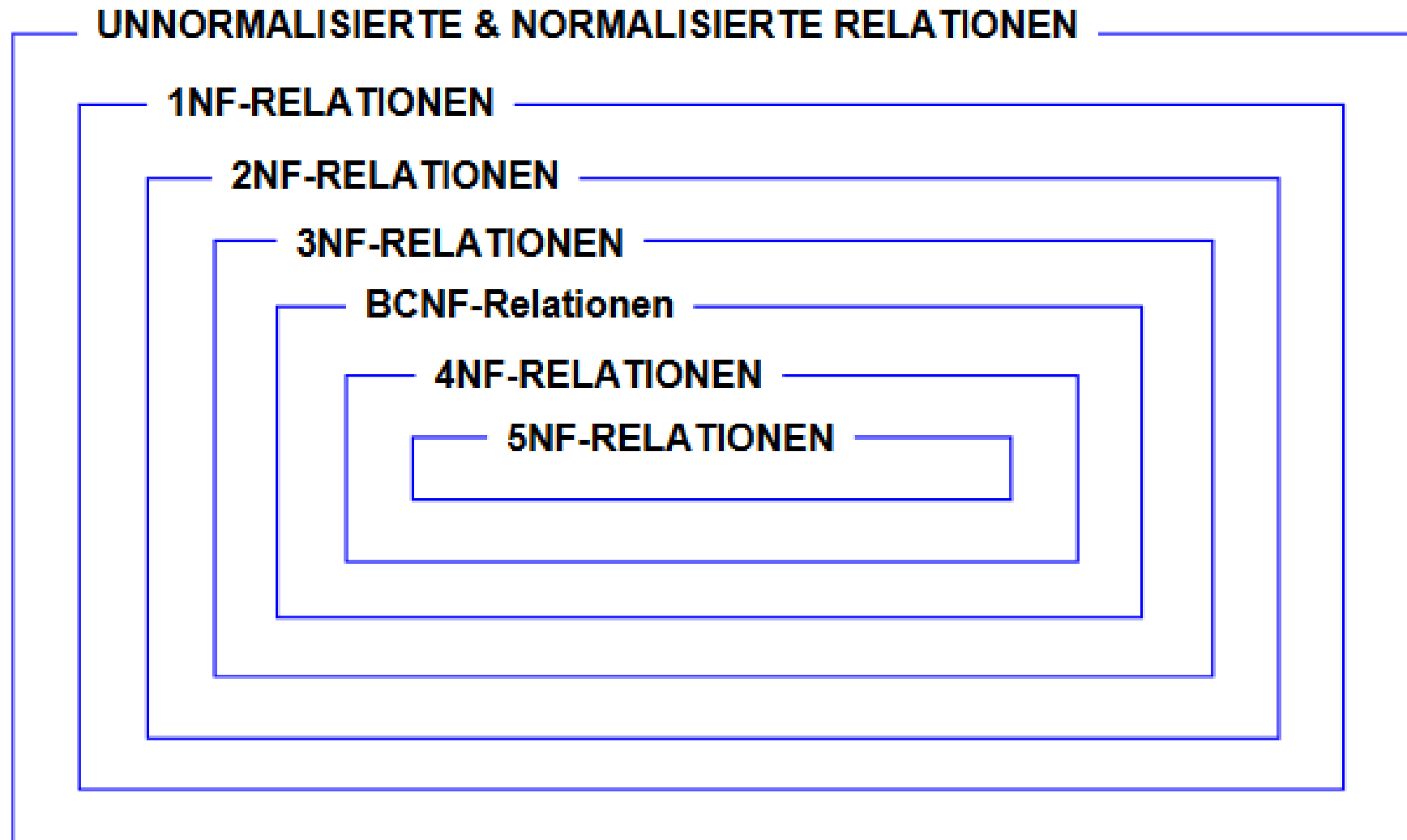


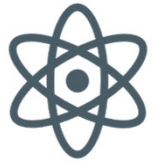
Der Primärschlüssel einer Relation ist der Schlüsselkandidat, der zur Identifikation der Elemente einer Relation explizit herausgestellt wird.

Fremdschlüssel



Unter einem Fremdschlüssel versteht man den geerbten Primärschlüssel einer anderen Relation.

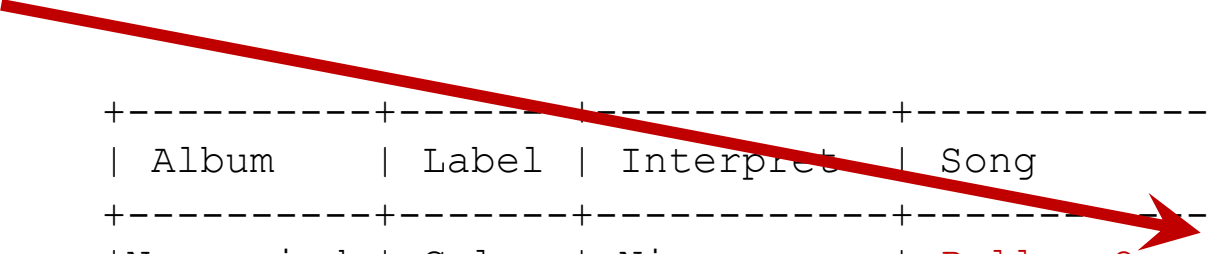




Eine Relation R befindet sich genau dann in erster Normalform (1NF), wenn alle Attributwerte atomar sind und die Relation frei von Wiederholungsgruppen ist.

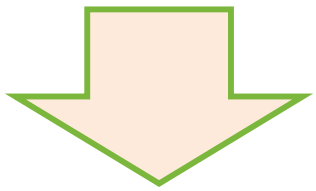
Überführung in die erste Normalform (1NF)

Ein Relationenschema R ist in 1NF, wenn jedes Attribut einen **atomaren Wertebereich** hat und die Relation frei von **Wiederholungsgruppen** ist.



Album	Label	Interpret	Song
Nevermind	Color	Nirvana	Polly, On a Plain
The Wall	King	Pink Floyd	Vera

Unnormalisierte Form,
d.h. nicht in 1NF



albumid	songid	Album	Label	Interpret	Song
1	1	Nevermind	Color	Nirvana	Polly
1	2	Nevermind	Color	Nirvana	On a Plain
2	1	The Wall	King	Pink Floyd	Vera

Atomarer Wertebereich: Nicht zusammengesetzt, z.B. Adresse nicht atomar
Keine Wiederholungsgruppen: Auch keine Attribute Song1, Song2, etc.



Eine Relation R befindet sich genau dann in zweiter Normalform (2NF), wenn R in erster Normalform vorliegt und alle Nichtschlüsselattribute **vollfunktional** vom Primärschlüssel **abhängig** sind.



Seien X und Y Attributkombinationen in R.

Funktionale Abhängigkeit

Y heißt funktional abhängig von X in R
(X bestimmt Y funktional, also $X \rightarrow Y$),
wenn es in der Relation R keine zwei Tupel geben darf,
die in ihrem Wert zu X, aber nicht in ihrem Wert zu Y
übereinstimmen.

Was ist wovon funktional abhängig?

Titel	Jahr	Dauer	FilmTyp	StudioName	StarName
Star Wars	1977	124	Farbe	Fox	Carrie Fisher
Star Wars	1977	124	Farbe	Fox	Mark Hamill
Star Wars	1977	124	Farbe	Fox	Harrison Ford
Good-Bye Lenin!	2003	121	Farbe	WDR	Daniel Brühl
Troja	2004	156	Farbe	Warner Bros	Brad Pitt
Troja	1956	121	SW	Warner Bros	Stanley Baker

Wir gehen davon aus, dass (1) in einem Jahr es nie 2 Filme mit gleichem Namen gibt, (2) ein Film-Name von einem Studio lizenrechtlich geschützt ist und (3) vor 1970 alles SW, danach alles in Farbe ist.



Seien X, Y, Z Attributkombinationen in R .

Vollfunktionale Abhängigkeit

Y heißt vollfunktional abhängig von X in R (X bestimmt Y voll funktional, also $X \Rightarrow Y$), wenn $X \rightarrow Y$ gilt und X minimal ist, also wenn keine Attributmenge $Z \subset X$ existiert, so dass $Z \rightarrow Y$. X wird Determinante genannt.

Ist dies eine volle oder eine partielle Abhängigkeit? Keine, weil ...

Titel	Jahr	Dauer	FilmTyp	StudioName	StarName
Star Wars	1977	124	Farbe	Fox	Carrie Fisher
Star Wars	1977	124	Farbe	Fox	Mark Hamill
Star Wars	1977	124	Farbe	Fox	Harrison Ford
Good-Bye Lenin!	2003	121	Farbe	WDR	Daniel Brühl
Troja	2004	156	Farbe	Warner Bros	Brad Pitt
Troja	1956	121	SW	Warner Bros	Stanley Baker

Wir gehen davon aus, dass (1) in einem Jahr es nie 2 Filme mit gleichem Namen gibt, (2) ein Film-Name von einem Studio lizenzrechtlich geschützt ist und (3) vor 1970 alles SW, danach alles in Farbe ist.



Seien X, Y, Z Attributkombinationen in R .

Vollfunktionale Abhängigkeit

Y heißt vollfunktional abhängig von X in R (X bestimmt Y voll funktional, also $X \Rightarrow Y$), wenn $X \rightarrow Y$ gilt und X minimal ist, also wenn keine Attributmenge $Z \subset X$ existiert, so dass $Z \rightarrow Y$. X wird Determinante genannt.

Ist dies eine volle oder eine partielle Abhängigkeit? Partielle.

Titel	Jahr	Dauer	FilmTyp	StudioName	StarName
Star Wars	1977	124	Farbe	Fox	Carrie Fisher
Star Wars	1977	124	Farbe	Fox	Mark Hamill
Star Wars	1977	124	Farbe	Fox	Harrison Ford
Good-Bye Lenin!	2003	121	Farbe	WDR	Daniel Brühl
Troja	2004	156	Farbe	Warner Bros	Brad Pitt
Troja	1956	121	SW	Warner Bros	Stanley Baker

Wir gehen davon aus, dass (1) in einem Jahr es nie 2 Filme mit gleichem Namen gibt, (2) ein Film-Name von einem Studio lizenrechtlich geschützt ist und (3) vor 1970 alles SW, danach alles in Farbe ist.

Überführung in die zweite Normalform (2NF)

Ein Relationenschema R befindet sich in 2NF, wenn es in 1NF ist und die Nichtschlüsselattribute von den Schlüsseln **vollständig abhängen**

albumid 	songid 	Album	Label	Interpret	Song
1	1	Nevermind	Color	Nirvana	Polly
1	2	Nevermind	Color	Nirvana	On a Plain
2	1	The Wall	King	Pink Floyd	Vera

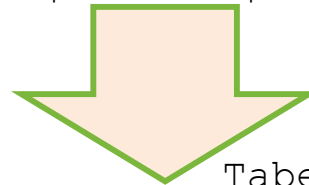


Tabelle 'songs'




id 	albumid	Song
1	1	Polly
2	1	On a Plain
3	2	Vera

Tabelle 'album'

id 	Album	Label	Interpret
1	Nevermind	Color	Nirvana
2	The Wall	King	Pink Floyd



Eine Relation R befindet sich genau dann in dritter Normalform (3NF), wenn R in zweiter Normalform vorliegt und kein Nichtschlüsselattribut **transitiv** vom Primärschlüssel abhängt.

4 Transitive Abhängigkeit

→ →

Seien X, Y, Z Attributkombinationen in R .

Transitive Abhängigkeit

Z heißt transitiv abhängig von X in R

(X bestimmt Z transitiv, also $X \rightarrow :: \rightarrow Z$),

wenn es ein $Y \neq X$ und $Y \neq Z$ gibt mit $X \rightarrow Y$ und $Y \rightarrow Z$.

Id	Titel	Jahr	Dauer	FilmTyp	StudioName	StarName
1	Star Wars	1977	124	Farbe	Fox	Carrie Fisher
2	Star Wars	1977	124	Farbe	Fox	Mark Hamill
3	Star Wars	1977	124	Farbe	Fox	Harrison Ford
4	Good-Bye Lenin!	2003	121	Farbe	WDR	Daniel Brühl
5	Troja	2004	156	Farbe	Warner Bros	Brad Pitt
6	Troja	1956	121	SW	Warner Bros	Stanley Baker

Wir gehen davon aus, dass (1) in einem Jahr es nie 2 Filme mit gleichem Namen gibt, (2) ein Film-Name von einem Studio lizenzrechtlich geschützt ist und (3) vor 1970 alles SW, danach alles in Farbe ist.

4 Überführung in die dritte Normalform (3NF)

Ein Relationenschema R befindet sich in 3NF, wenn es in 2NF ist und kein Nichtschlüsselattribut von einem Schlüsselkandidaten **transitiv abhängig** ist.

Tabelle 'songs'				Tabelle 'album'			
id	albumid	Song		id	Album	Label	Interpret
1	1	Polly		1	Nevermind	Color	Nirvana
2	1	On a Plain		2	The Wall	King	Pink Floyd
3	2	Vera		3	Meddle	King	Pink Floyd

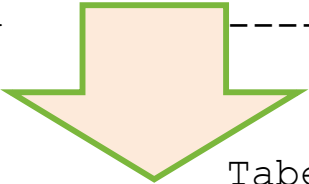


Tabelle 'interpreten'				Tabelle 'album'			
id	Interpret	Label		id	Album	interpretid	
1	Nirvana	Color		1	Nevermind	1	
2	Pink Floyd	King		2	The Wall	2	
				3	Meddle	2	

Tabelle 'songs' unverändert.

Realitätskonforme Modellierung: Erhaltene Entitätstypen machen „Sinn“.

4 Aufgabe I - Normalisierung

Gegeben ist folgende Tabelle (Ausschnitt), welche die Mengen angibt, die ein Kunde im letzten Jahr von einem bestimmten Artikel bestellt hat (unabhängig von der Anzahl der einzelnen Bestellungen):

Kunden-Nr	Kunden-Name	Artikel-Nr	Artikel-Bezeichnung	Menge
1	Maier	100	Monitor XY	3
2	Müller	234	Maus Quick	1
1	Maier	300	Drucker Printfix	6
4	Schmidt	234	Maus Quick	2

- a) Legen Sie einen Primärschlüssel (Primary Key, PK) fest.
- b) Welche Normalform liegt vor?
- c) Geben Sie alle Gründe an, warum höhere Normalformen nicht erreicht wurden.
- d) Überführen Sie die Tabelle in die 3. Normalform (Tabellen angeben).

Gegeben ist folgende Tabelle für die Ergebnisse eines Prüfungszeitraums:

Matrikel-Nr.	Name	Vorlesungs-Nr.	Vorlesungs-Titel	Note
123456	Maier	F005	WINF 1	2,3
234567	Müller	F005	WINF 1	3,7
123456	Maier	F010	WINF2	1,7
345678	Schmidt	F012	Statistik	4,0

- a) Legen Sie einen Primärschlüssel (Primary Key, PK) fest.
- b) Welches sind die Fremdschlüssel (Foreign Key, FK) dieser Relation?
- c) Welche Attribute sind funktional von Matrikel-Nr. abhängig?
- d) Welche Attribute sind vollfunktional von der Kombination aus Matrikel-Nr. u. Vorlesungs-Nr. abhängig?
- e) Welche Attribute sind transitiv von Matrikel-Nr. abhängig?
- f) In welcher Normalform befindet sich die Tabelle (Begründung)?

4 Fallbeispiel

Die Medizinische Anwendungstechnik GmbH (MAT) möchte ihre zentralen Daten in einer Datenbank verwalten. Diese soll folgende Sachverhalte abbilden:

Ein Mitarbeiter, bestehend aus einem Namen, Vornamen und einer MitarbeiterNr, besitzt verschiedene Qualifikationen und führt unterschiedliche Tätigkeiten aus. Er ist für maximal ein Produkt zuständig, für das er mehrere Kunden betreut, die jedoch bei anderen Mitarbeitern bzgl. weiterer Produkte betreut werden können.

Ein Produkt, für das mehrere Mitarbeiter zuständig sein können, wird an verschiedenen Standorten (Fabriken) auf unterschiedlichen Bändern produziert. Die verschiedenen Abteilungen sind an jeweils einem Standort angesiedelt. Das Produkt hat eine Kategorie und Produkt/Band haben eine Bezeichnung.

4 Fallbeispiel

Die Medizinische Anwendungstechnik GmbH (MAT) möchte ihre zentralen Daten in einer Datenbank verwalten. Diese soll folgende Sachverhalte abbilden:

Ein Mitarbeiter, bestehend aus einem Namen, Vornamen und einer MitarbeiterNr, besitzt verschiedene Qualifikationen und führt unterschiedliche Tätigkeiten aus. Er ist für maximal ein Produkt zuständig, für das er mehrere Kunden betreut, die jedoch bei anderen Mitarbeitern bzgl. weiterer Produkte betreut werden können.

Ein Produkt, für das mehrere Mitarbeiter zuständig sein können, wird an verschiedenen Standorten (Fabriken) auf unterschiedlichen Bändern produziert. Die verschiedenen Abteilungen sind an jeweils einem Standort angesiedelt. Das Produkt hat eine Kategorie und Produkt/Band haben eine Bezeichnung.

4 Fallbeispiel

Relation 1: Abteilungen mit deren Fabriken und Standorten

<u>ANr</u>	AName	FNr	FName	Ort
1	Controlling	1	Fab 1	Silschede
2	Einkauf	2	Fab 2	Heeren-Werve
3	Vertrieb	1	Fab 1	Silschede
4	F & E	3	Fab 3	Marsberg
5	Programmierung	4	Fab 4	Lobberich

Relation 2: Kunden mit deren Produkten

<u>KNr</u>	KName	KVorname	<u>PNr</u>	MNr
1	Pullach	Ingo	1	2
1	Pullach	Ingo	2	3
2	Tullberg	Christian	1	2
2	Tullberg	Christian	2	1

4 Fallbeispiel

Die Medizinische Anwendungstechnik GmbH (MAT) möchte ihre zentralen Daten in einer Datenbank verwalten. Diese soll folgende Sachverhalte abbilden:

Ein Mitarbeiter, bestehend aus einem Namen, Vornamen und einer MitarbeiterNr, besitzt verschiedene Qualifikationen und führt unterschiedliche Tätigkeiten aus. Er ist für maximal ein Produkt zuständig, für das er mehrere Kunden betreut, die jedoch bei anderen Mitarbeitern bzgl. weiterer Produkte betreut werden können.

Ein Produkt, für das mehrere Mitarbeiter zuständig sein können, wird an verschiedenen Standorten (Fabriken) auf unterschiedlichen Bändern produziert. Die verschiedenen Abteilungen sind an jeweils einem Standort angesiedelt. Das Produkt hat eine Kategorie und Produkt/Band haben eine Bezeichnung.

4 Fallbeispiel

Relation 3: Mitarbeiter

<u>MNr</u>	MName	MVorname	Tätigkeit	Qualifikation
1	Mocke	Agatha	Marketing Kommunikation	Direct Sales Master Rhetorik
2	Danndahl	Claus	PR Organisation	Coaching ZMR
3	Platten	Susanne	Programmierung Kommunikation	Java CORBA

4 Fallbeispiel

Die Medizinische Anwendungstechnik GmbH (MAT) möchte ihre zentralen Daten in einer Datenbank verwalten. Diese soll folgende Sachverhalte abbilden:

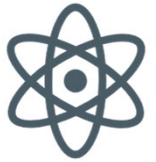
Ein Mitarbeiter, bestehend aus einem Namen, Vornamen und einer MitarbeiterNr, besitzt verschiedene Qualifikationen und führt unterschiedliche Tätigkeiten aus. Er ist für maximal ein Produkt zuständig, für das er mehrere Kunden betreut, die jedoch bei anderen Mitarbeitern bzgl. weiterer Produkte betreut werden können.

Ein Produkt, für das mehrere Mitarbeiter zuständig sein können, wird an verschiedenen Standorten (Fabriken) auf unterschiedlichen Bändern produziert. Die verschiedenen Abteilungen sind an jeweils einem Standort angesiedelt. Das Produkt hat eine Kategorie und Produkt/Band haben eine Bezeichnung.

4 Fallbeispiel

Relation 4: Produkte

<u>PNr</u>	PBez	PKat	<u>FNr</u>	<u>BNr</u>	BBez
1	Dialysegerät	2F	1	1	3c
1	Dialysegerät	2F	1	2	1b
1	Dialysegerät	2F	2	1	3c
1	Dialysegerät	2F	2	2	1b
2	Frequenzmesser	G6	1	1	3c
2	Frequenzmesser	G6	1	2	1b
3	Röntgengerät	K2	1	1	3c
3	Röntgengerät	K2	2	1	3c
3	Röntgengerät	K2	4	1	3c
3	Röntgengerät	K2	4	3	6f



Eine Relation R befindet sich genau dann in erster Normalform (1NF), wenn alle Attributwerte atomar sind und die Relation frei von Wiederholungsgruppen ist.

4 Beispiel zur 1. Normalform

Ausgangstabelle

Relation 3: Mitarbeiter

<u>MNr</u>	MName	MVorname	Tätigkeit	Qualifikation
1	Mocke	Agatha	Marketing Kommunikation	Direct Sales Master Rhetorik
2	Danndahl	Claus	PR Organisation	Coaching ZMR
3	Platten	Susanne	Programmierung Kommunikation	Java CORBA

4 Beispiel zur 1. Normalform

Ergebnis: Tätigkeiten und Qualifikationen sind atomar

Relation 3: Mitarbeiter

<u>MNr</u>	MName	MVorname	<u>Tätigkeit</u>	<u>Qualifikation</u>
1	Mocke	Agatha	Marketing	Direct Sales Master
1	Mocke	Agatha	Marketing	Rhetorik
1	Mocke	Agatha	Kommunikation	Direct Sales Master
1	Mocke	Agatha	Kommunikation	Rhetorik
2	Danndahl	Claus	PR	Coaching
2	Danndahl	Claus	PR	ZMR
2	Danndahl	Claus	Organisation	Coaching
2	Danndahl	Claus	Organisation	ZMR
3	Platten	Susanne	Programmierung	Java
3	Platten	Susanne	Programmierung	CORBA
3	Platten	Susanne	Kommunikation	Java
3	Platten	Susanne	Kommunikation	CORBA



Eine Relation R befindet sich genau dann in zweiter Normalform (2NF), wenn R in erster Normalform vorliegt und alle Nichtschlüsselattribute vollfunktional vom Primärschlüssel abhängig sind.

4 Beispiel zur 2. Normalform

Ausgangstabelle

Relation 2: Kunden

<u>KNr</u>	KName	KVorname	<u>PNr</u>	MNr
1	Pullach	Ingo	1	2
1	Pullach	Ingo	2	3
2	Tullberg	Christian	1	2
2	Tullberg	Christian	2	1

4 Beispiel zur 2. Normalform

Ergebnis

Relation 2a: Zuordnung von Kunden zu Produkten mit Mitarb.

<u>KNr</u>	<u>PNr</u>	MNr
1	1	2
1	2	3
2	1	2
2	2	1

Relation 2b: Kunden

<u>KNr</u>	KName	KVorname
1	Pullach	Ingo
2	Tullberg	Christian

4 Beispiel zur 2. Normalform

Ausgangstabelle

Relation 3: Mitarbeiter



<u>MNr</u>	MName	MVorname	<u>Tätigkeit</u>	<u>Qualifikation</u>
1	Mocke	Agatha	Marketing	Direct Sales Master
1	Mocke	Agatha	Marketing	Rhetorik
1	Mocke	Agatha	Kommunikation	Direct Sales Master
1	Mocke	Agatha	Kommunikation	Rhetorik
2	Danndahl	Claus	PR	Coaching
2	Danndahl	Claus	PR	ZMR
2	Danndahl	Claus	Organisation	Coaching
2	Danndahl	Claus	Organisation	ZMR
3	Platten	Susanne	Programmierung	Java
3	Platten	Susanne	Programmierung	CORBA
3	Platten	Susanne	Kommunikation	Java
3	Platten	Susanne	Kommunikation	CORBA

4 Beispiel zur 2. Normalform

Ergebnis

Relation 3a: Tätigkeit und Q.

<u>MNr</u>	<u>Tätigkeit</u>	<u>Qualifikation</u>
1	Marketing	Direct Sales Master
1	Marketing	Rhetorik
1	Kommunikation	Direct Sales Master
1	Kommunikation	Rhetorik
2	PR	Coaching
2	PR	ZMR
2	Organisation	Coaching
2	Organisation	ZMR
3	Programmierung	Java
3	Programmierung	CORBA
3	Kommunikation	Java
3	Kommunikation	CORBA


Relation 3b: Mitarbeiter

<u>MNr</u>	<u>MName</u>	<u>MVorname</u>
1	Mocke	Agatha
2	Danndahl	Claus
3	Platten	Susanne

4 Beispiel zur 2. Normalform

Ausgangstabelle

Relation 4: Produkte



<u>PNr</u>	PBez	PKat	<u>FNr</u>	<u>BNr</u>	BBez
1	Dialysegerät	2F	1	1	3c
1	Dialysegerät	2F	1	2	1b
1	Dialysegerät	2F	2	1	3c
1	Dialysegerät	2F	2	2	1b
2	Frequenzmesser	G6	1	1	3c
2	Frequenzmesser	G6	1	2	1b
3	Röntgengerät	K2	1	1	3c
3	Röntgengerät	K2	2	1	3c
3	Röntgengerät	K2	4	1	3c
3	Röntgengerät	K2	4	3	6f

4 Beispiel zur 2. Normalform

Ergebnis

Relation 4a:
Zuordnungen

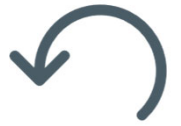
<u>PNr</u>	<u>FNr</u>	<u>BNr</u>
1	1	1
1	1	2
1	2	1
1	2	2
2	1	1
2	1	2
3	1	1
3	2	1
3	4	1
3	4	3

Relation 4b: Produkte

<u>PNr</u>	PBez	PKat
1	Dialysegerät	2F
2	Frequenzmesser	G6
3	Röntgengerät	K2

Relation 4c:
Bänderbez.

<u>BNr</u>	BBez
1	3c
2	1b
3	6f

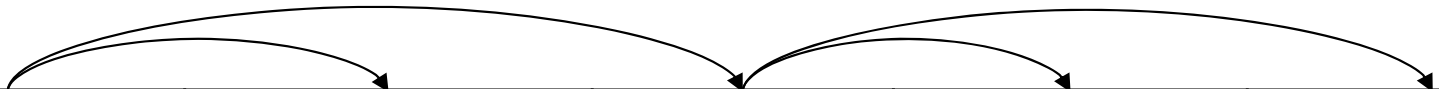


Eine Relation R befindet sich genau dann in dritter Normalform (3NF), wenn R in zweiter Normalform vorliegt und kein Nichtschlüsselattribut transitiv vom Primärschlüssel abhängt.

4 Beispiel zur 3. Normalform

Ausgangstabelle

Relation 1: Abteilungen mit deren Fabriken und Standorten



<u>ANr</u>	AName	FNr	FName	Ort
1	Controlling	1	Fab 1	Silschede
2	Einkauf	2	Fab 2	Heeren-Werve
3	Vertrieb	1	Fab 1	Silschede
4	F & E	3	Fab 3	Marsberg
5	Programmierung	4	Fab 4	Lobberich

4 Beispiel zur 3. Normalform

Ergebnis

Relation 1a: Abteilungen

<u>ANr</u>	AName	FNr
1	Controlling	1
2	Einkauf	2
3	Vertrieb	1
4	F & E	3
5	Programmierung	4


Relation 1b: Fabriken

<u>FNr</u>	FName	Ort
1	Fab 1	Silschede
2	Fab 2	Heeren-Werve
3	Fab 3	Marsberg
4	Fab 4	Lobberich

4 Definition der Boyce-Codd Normalform

Ein oder mehrere Attribute eines Relationenschemas fungieren als **Determinante**, wenn andere Attribute des Relationenschemas funktional von diesen abhängen.

Ein **Schlüsselkandidat** ist eine *minimale* Menge von Attributen, die die Tupel (Datensatz) einer Relation eindeutig identifiziert.



Eine Relation R befindet sich genau dann in Boyce-Codd Normalform (BCNF), wenn R in dritter Normalform vorliegt und jede Determinante in R auch Schlüsselkandidat ist.

4 Beispiel zur Boyce-Codd Normalform

Ausgangstabelle

Relation 2a: Zuordnung von Kunden zu Produkten mit Mitarb.

<u>KNr</u>	<u>PNr</u>	MNR
1	1	2
1	2	3
2	1	2
2	2	1

Attributkombinationen	Identifikator	Schlüssel-kandidat	Determinante	Abhängiges Attribut
(KNr, MNR, PNr)	X			
(KNr, PNr)	X	X	X	MNR
(KNr, MNR)	X	X	X	PNr
(MNR, PNr)				
(MNR)			X	PNr
(KNr)				
(PNr)				

4 Beispiel zur Boyce-Codd Normalform

Ergebnis

Relation $2a_1$:
Zuordnung Kunde zu Mitarb.

<u>KNr</u>	<u>MNr</u>
1	2
1	3
2	2
2	1

Relation $2a_2$:
Zuordnung Kunde zu Produkt

<u>MNr</u>	PNr
2	1
3	2
1	2

4 Mehrwertige Abhängigkeit

Seien X, Y, Z Attributkombinationen in R .

Mehrwertige Abhängigkeit

Eine mehrwertige Abhängigkeit von X in R liegt genau dann vor,

$$\text{also } X \twoheadrightarrow Y,$$

wenn die Werte von X eine Menge von Werten in Y bestimmen, unabhängig von den Werten der restlichen Attributen von R .

4 Beispiel zur 4. Normalform


Ausgangstabelle

Relation 3a: Tätigkeit und Qualifikation



<u>MNr</u>	<u>Tätigkeit</u>	<u>Qualifikation</u>
1	Marketing	Direct Sales Master
1	Marketing	Rhetorik
1	Kommunikation	Direct Sales Master
1	Kommunikation	Rhetorik
2	PR	Coaching
2	PR	ZMR
2	Organisation	Coaching
2	Organisation	ZMR
3	Programmierung	Java
3	Programmierung	CORBA
3	Kommunikation	Java
3	Kommunikation	CORBA

Eine mehrwertige Abhängigkeit $X \twoheadrightarrow Y$ ist **nicht trivial**, wenn außer X und Y noch weitere Attribute im zugehörigen Relationenschema vorkommen.



Eine Relation R befindet sich genau dann in vierter Normalform (4NF), wenn R in BCNF vorliegt und alle mehrwertigen Abhängigkeiten in R trivial sind.

4 Beispiel zur 4. Normalform

Ausgangstabelle

Relation 3a: Tätigkeit und Qualifikation



<u>MNr</u>	<u>Tätigkeit</u>	<u>Qualifikation</u>
1	Marketing	Direct Sales Master
1	Marketing	Rhetorik
1	Kommunikation	Direct Sales Master
1	Kommunikation	Rhetorik
2	PR	Coaching
2	PR	ZMR
2	Organisation	Coaching
2	Organisation	ZMR
3	Programmierung	Java
3	Programmierung	CORBA
3	Kommunikation	Java
3	Kommunikation	CORBA

Liegt 4. NF vor? Falls ja, warum? Falls nein, wie bringt man es in 4. NF?

4 Beispiel zur 4. Normalform

Ergebnis

Relation 3a₁: Tätigkeit

<u>MNr</u>	<u>Tätigkeit</u>
1	Marketing
1	Kommunikation
2	PR
2	Organisation
3	Programmierung
3	Kommunikation

Relation 3a₂: Qualifikation

<u>MNr</u>	<u>Qualifikation</u>
1	Direct Sales Master
1	Rhetorik
2	Coaching
2	ZMR
3	Java
3	CORBA


Seien X, Y, Z Attributkombinationen in R .

Verbundabhängigkeit

Eine Relation R genügt der Verbundabhängigkeit,

also $\bowtie(X, Y, \dots, Z)$,

genau dann, wenn R gleich dem Verbund aus seinen Projektionen X, Y, \dots, Z ist.



Eine Relation R befindet sich in fünfter Normalform (5NF) genau dann, wenn R in vierter Normalform vorliegt und jede Verbundabhängigkeit in R eine Folge der Schlüsselkandidaten in R ist.

4 Beispiel zur 5. Normalform

Ausgangstabelle

Relation 4a: Zuordnungen

<u>PNr</u>	<u>FNr</u>	<u>BNr</u>
1	1	1
1	1	2
1	2	1
1	2	2
2	1	1
2	1	2
3	1	1
3	2	1
3	4	1
3	4	3

4 Beispiel zur 5. Normalform

Ergebnis

Relation 4a₁

<u>PNr</u>	<u>FNr</u>
1	1
1	2
2	1
3	1
3	2
3	4

Relation 4a₂

<u>PNr</u>	<u>BNr</u>
1	1
1	2
2	1
2	2
3	1
3	3

Relation 4a₃

<u>FNr</u>	<u>BNr</u>
1	1
1	2
2	1
2	2
4	1
4	3

4 Sechs Strukturregeln

- **Ziel**
Analyse von Zusammenhängen *zwischen* Relationen,
nicht innerhalb von Relationen wie bei der Normalisierung.
- **Ergebnis**
Global normalisierte Datenbasis

Lokales Attribut

Ein lokales Attribut kommt nur in einer einzigen Tabelle vor und ist dort nicht Bestandteil des Primärschlüssels.

Globales Attribut

Ein globales Attribut ist in mindestens einer Tabelle Teil des Primärschlüssels und kann so über Schlüsselvererbung in mehreren Tabellen vorkommen.

1. Strukturregel

Jede Relation muss einen Identifikationsschlüssel haben.

2. Strukturregel

Eine Datenbasis muss aus Relationen in 3NF bestehen, welche ausschließlich globale und lokale Attribute enthalten.

4 Statischer und dynamischer Wertebereich

Statischer Wertebereich

Ein statischer Wertebereich umfasst eine feste Menge von Werten, die sich im Verlauf der Zeit nicht ändert. Er wird bei der Definition der Datenbasis festgelegt.

Beispiele für statische Wertebereiche:

Integer, Zeichenkette mit 20 Zeichen

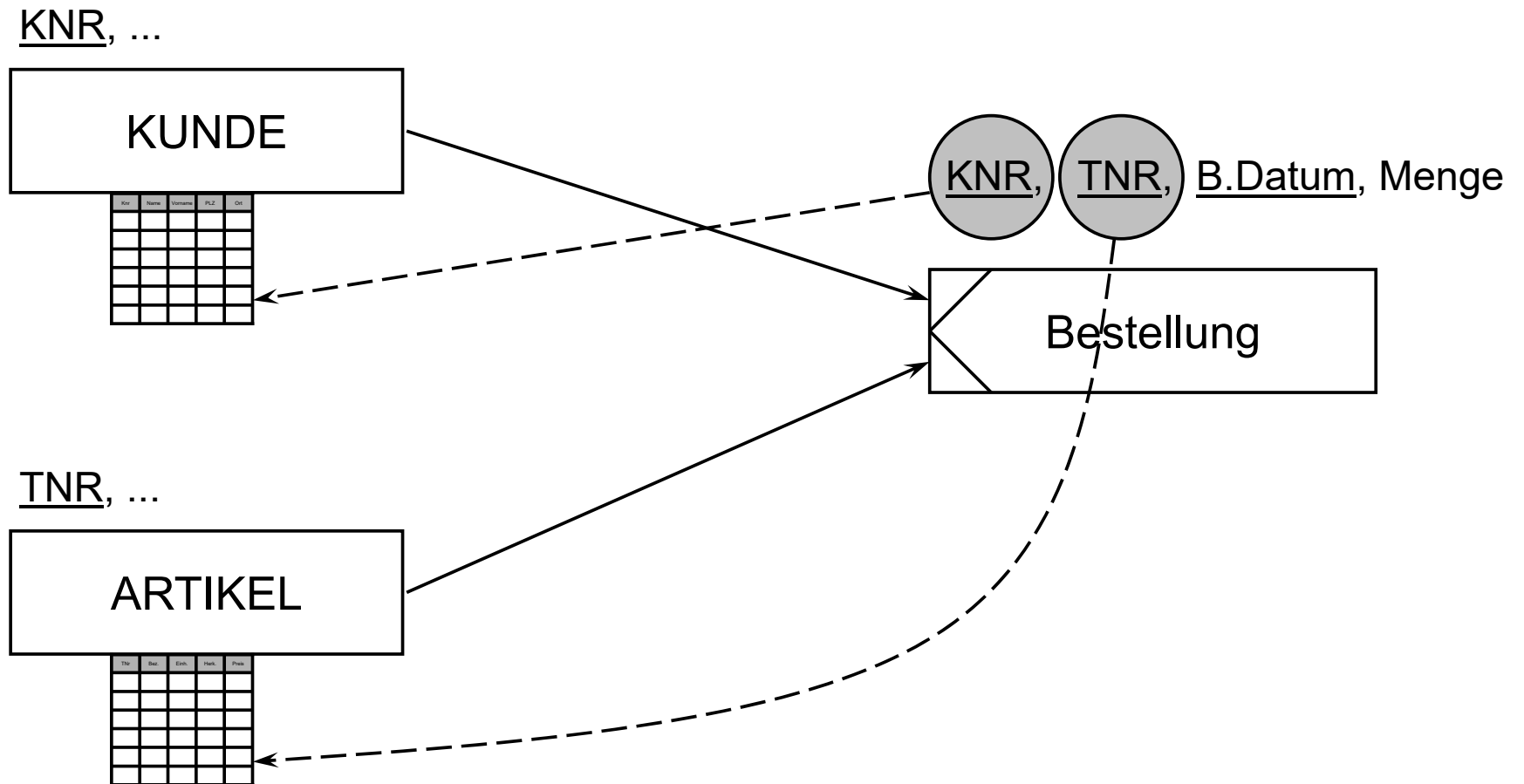
Dynamischer Wertebereich

Ein dynamischer Wertebereich besteht aus einer Menge von Werten, die in einem Schlüsselattribut einer fremden Relation als Attributwerte vorhandener Datensätze auftreten.

3. Strukturregel

Lokale Attribute haben statische Wertebereiche.

Globalen Attribute haben nur in einer Relation einen statischen Wertebereich. Dort sind sie ein ID-Schlüssel. In allen anderen Relationen darf das Attribut nur als Fremdschlüssel mit dynamischem Wertebereich auftreten (Bedingung der referentiellen Integrität).

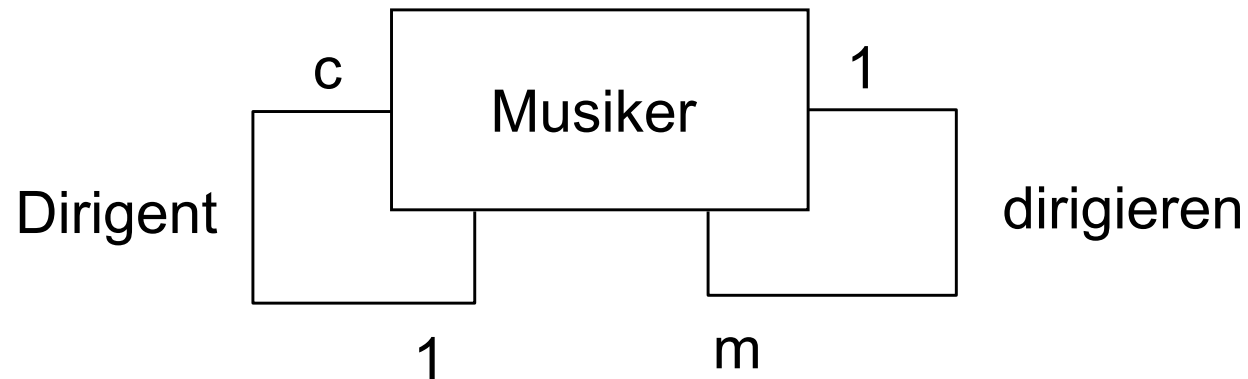


4. Strukturregel

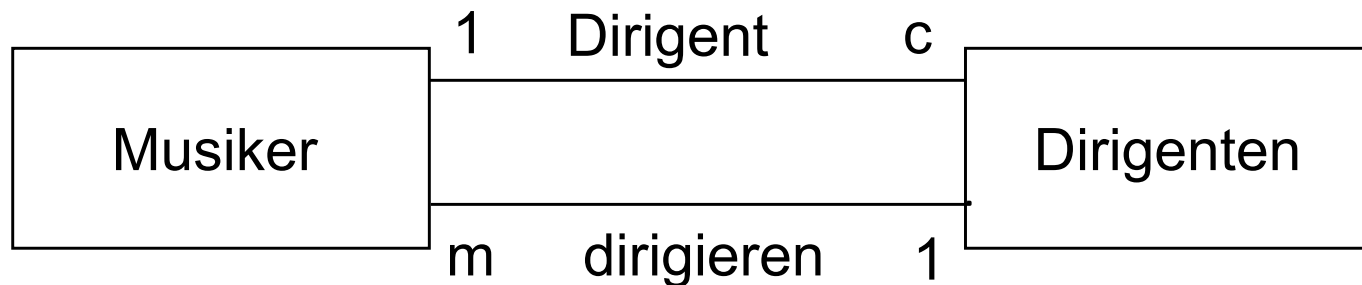
Rekursive Beziehungen zwischen Relationen sind unzulässig.

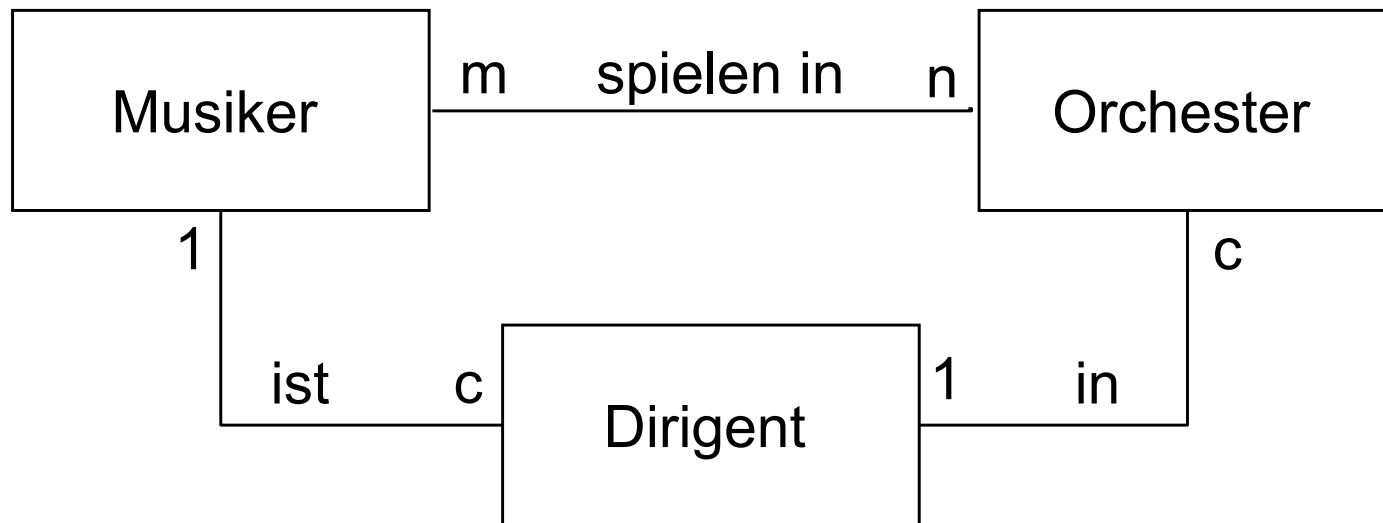
Dies bedeutet: In einer Relation R_1 darf ein globales Attribut nur mit einem Fremdschlüssel gebildet werden, dessen Ursprungsrelation R_2 unabhängig von R_1 definiert werden kann.

Falsch nach 4. Strukturregel



Falsch nach 4. Strukturregel

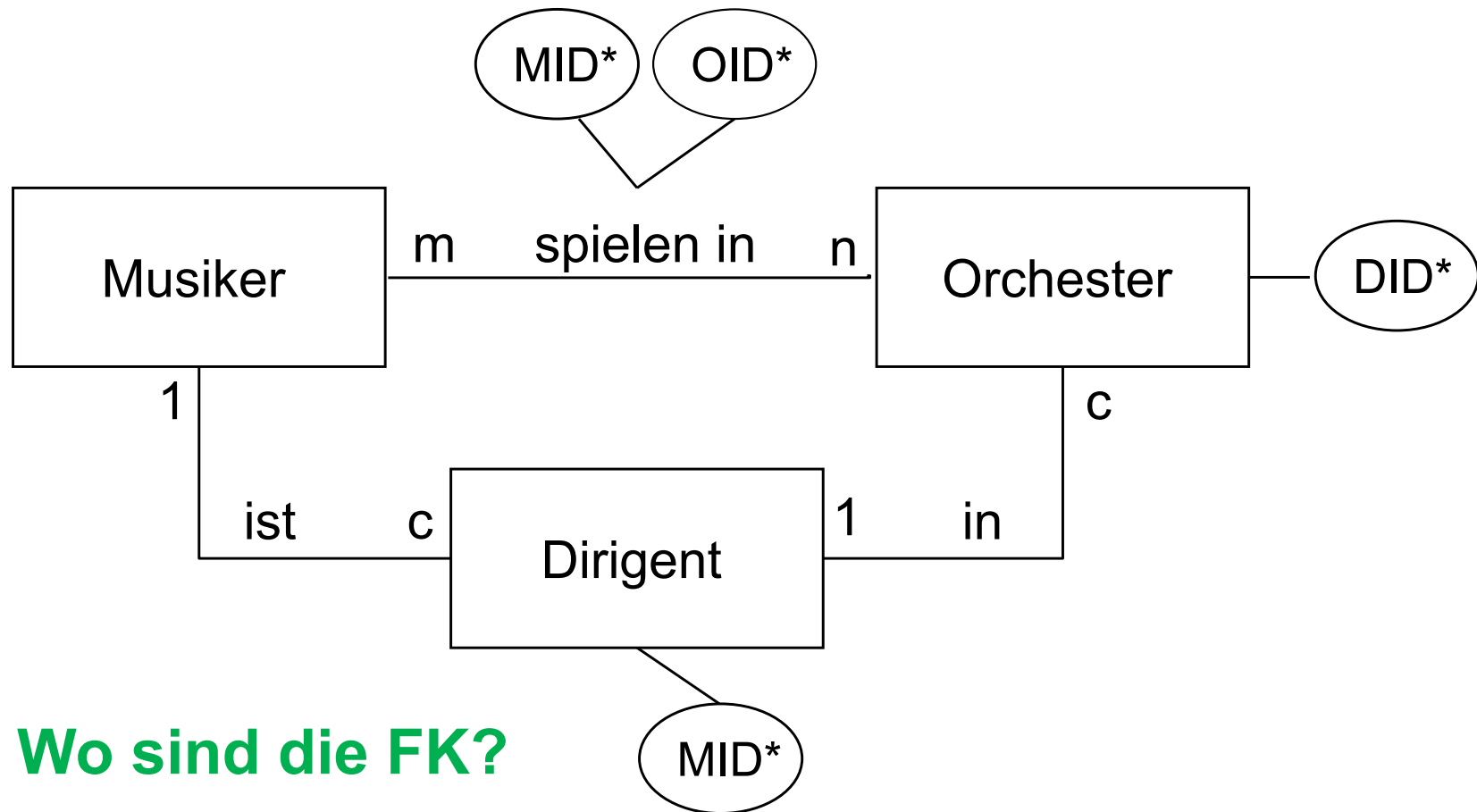




Wo sind die FK?

4

Beispiel zur 4. Strukturregel

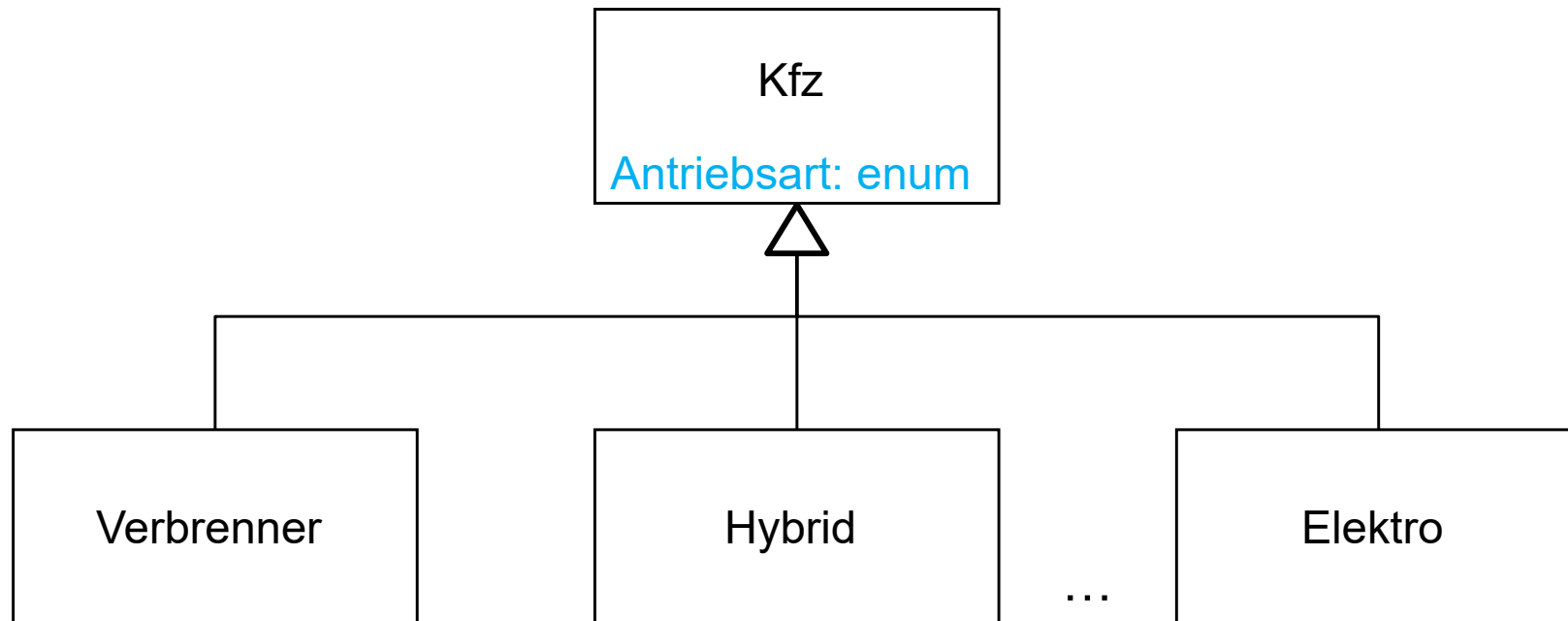


Wo sind die FK?

In welcher Reihenfolge werden die Tabellen gefüllt?

5. Strukturregel

Vorhandene Ober- und Untermengenbeziehungen zwischen Relationen sind präzise darzustellen. Die Zuordnung einer Entität zu einer disjunkten, spezialisierten Untermenge wird durch ein diskriminierendes Attribut in der generalisierten Relation ausgedrückt.



6. Strukturregel

Wenn in einer Tabelle globale Attribute als Fremdschlüssel eingeführt werden sollen, dann nur wenn dies eine größtmögliche Einschränkung (möglichst wenige Tupel in der Entitätsmenge) des zulässigen dynamischen Wertebereiches mit sich bringt.

Kunden (KNr, Vorname, Nachname, PLZ, Ort)

Man könnte bei obiger Relation auf die Idee kommen, die Nachnamen in einer eigenen Tabelle zu verwalten:

Kunden (KNr , Vorname, NNr, PLZ, Ort)

Nachnamen (NNr , Nachname)

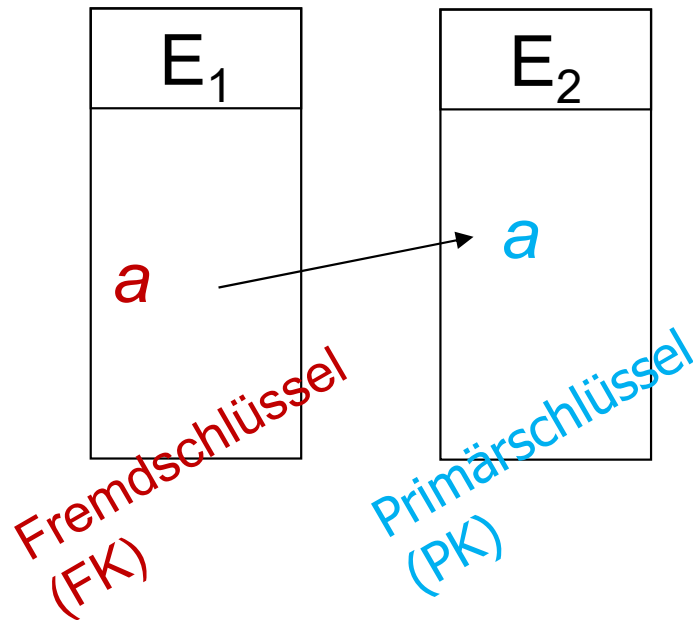
6. Strukturregel: Dieser FK NNr soll **nicht eingeführt** werden, da es gibt jedoch gibt derart viele Nachnamen, dass es keinen Sinn macht, diese in einer eigenen Tabelle zu verwalten. Es ist einfacher, den Nachnamen einfach einzugeben.

Ziel der 6. Strukturregel:

Unnötiges Aufteilen von Tabellen verhindern.

- Aufgrund der Forderung nach Normalisierung enthalten Tabellen Verweise auf Einträge anderer Tabellen.
- Wichtig ist: Diese Einträge sollten immer vorhanden sein.
- Die **referenzielle Integrität** (RI) definiert Regeln, die dieses Vorhandensein sicherstellen soll
- Nach den Regeln dürfen Datensätze (über ihre Fremdschlüssel) nur auf existierende Datensätze verweisen.
- Bei Änderungen von Werten, auf die verwiesen wird, wird geprüft, ob die Regeln noch intakt bleiben.

(Gute) Datenbanken unterstützen die RI.



Regeln beim Löschen/Update in E_2 :

RESTRICT: Nicht möglich, erst muss Referenz entfernt werden.

CASCADE: Updates werden auf E_1 übertragen
Löschung in $E_2 \rightarrow$ Löschung in E_1

SET NULL: Setze Attribut in E_1 auf NULL.

- Der Eintrag mit Schlüssel a muss in E_2 enthalten sein.
- Notlösung: Falls nötig und möglich kann die Spalte in E_1 den Wert NULL erhalten, um zu verdeutlichen, dass es keinen passenden Primärschlüssel gibt.
- RI kann durch SQL-Statements (DDL) garantiert werden!

4 Referentielle Integrität – Änderungsregeln

- **RESTRICT**

Zurückweisen der Änderungsoperation, d.h. solange abhängige Objekte bestehen, ist PK-Wert nicht löscherbar.

- **NO ACTION**

Durchführen der Operation und Verifizieren der Fremdschlüsselbeziehungen am Ende.

Z.B. kann ein Primärschlüssel geändert werden, wenn die Semantik der Anweisung und aller involvierten Trigger dafür sorgt, dass am Ende alles passt.

- **CASCADE**

Propagieren der Änderung, d.h. alle abhängigen Datensätze werden ebenfalls gelöscht

4 Referentielle Integrität – Änderungsregeln

- SET NULL

Verweise auf NULL setzen, d.h. die Werte der abhängigen Fremdschlüssel werden auf Null gesetzt

- SET DEFAULT

Verweise auf Defaultwert setzen, d.h. die Werte der abhängigen Fremdschlüssel werden auf Defaultwert gesetzt

Gegeben sei folgende Relation:

*Relation R(Matrikel-Nr, Name, Vorlesungs-Nr,
Vorlesungs-Titel, Studiengang-Kürzel, Studiengang-Name)*

Jede Vorlesung ist dabei eindeutig einem Studiengang zugeordnet. Eine Zuordnung der Studierenden zu einem Studiengang soll hier unberücksichtigt bleiben.

- a) Was verhindert, dass sich die Relation in der 2NF befindet?
- b) Überführen Sie die Relation in die 2NF. Benutzen Sie höchstens 3 Relationen.
- c) Was verhindert jetzt die 3NF?
- d) Bilden Sie nun die 3NF.

Gegeben ist folgende Tabelle, die einen Stundenplan für ein bestimmtes Semester wiedergibt. Dabei wird angenommen, dass in einem Raum zu einem Zeitpunkt immer nur ein Fach stattfindet.

Wochentag	Block	Raum-Nr.	Raumgröße	Fach-Nr.	Titel	Dozent
Donnerstag	2	A2	160	WK1203	DBS	Kammer
Freitag	3	B04	40	WK2102	BI	Ritz
...						

- Geben Sie die Tabelle in der 1NF an.
- Nennen Sie konkret alle Gründe (mit Attributnamen), warum sich die Tabelle nicht in 2NF befindet.
- Überführen Sie die Relation in die 2NF. Nennen Sie konkret alle Gründe (mit Attributnamen), warum sich die neuen Relationen noch nicht in 3NF befindet.
- Überführen Sie die Relationen in die 3NF.

4 Kontrollfragen

- Welche Gründe sprechen für den Einsatz des Relationenmodells?
- Was versteht man unter einem Schlüssel?
- Welche Zwecke erfüllen Identifikations- und Klassifikationsschlüssel?
- Erklären Sie die Bedeutung eines Primär-, Sekundär-(Index-) und Fremdschlüssels.
- Wie geht man bei der Ableitung des Relationenmodells aus dem ER-Modell vor?
- Erläutern Sie den Begriff der „Normalisierung“. Welche Vorteile ergeben sich durch eine normalisierte Speicherung und mit welchen neuen Herausforderungen ist man konfrontiert?
- Wann ist eine Relation in 1.Normalform (NF), wann in 2.NF und wann in 3.NF? Geben Sie jeweils Beispiele für diese Normalformen an.
- Was versteht man unter „Referenzieller Integrität“?
- Geben Sie mögliche Änderungsregeln für Fremdschlüsselverbindungen an.