DATENMODELLIERUNG IV

NORMALISIERUNG

DATENBANKSYSTEME 3. JAHRGANG

ABTEILUNG INFORMATIONSTECHNIK

WAS BISHER GESCHAH ...

Bisher:

aus Beschreibung des Geschäftsfeldes

-> (E)ER-Diagramm

-> Relationenmodell

-> Skripte zur Tabellenerstellung

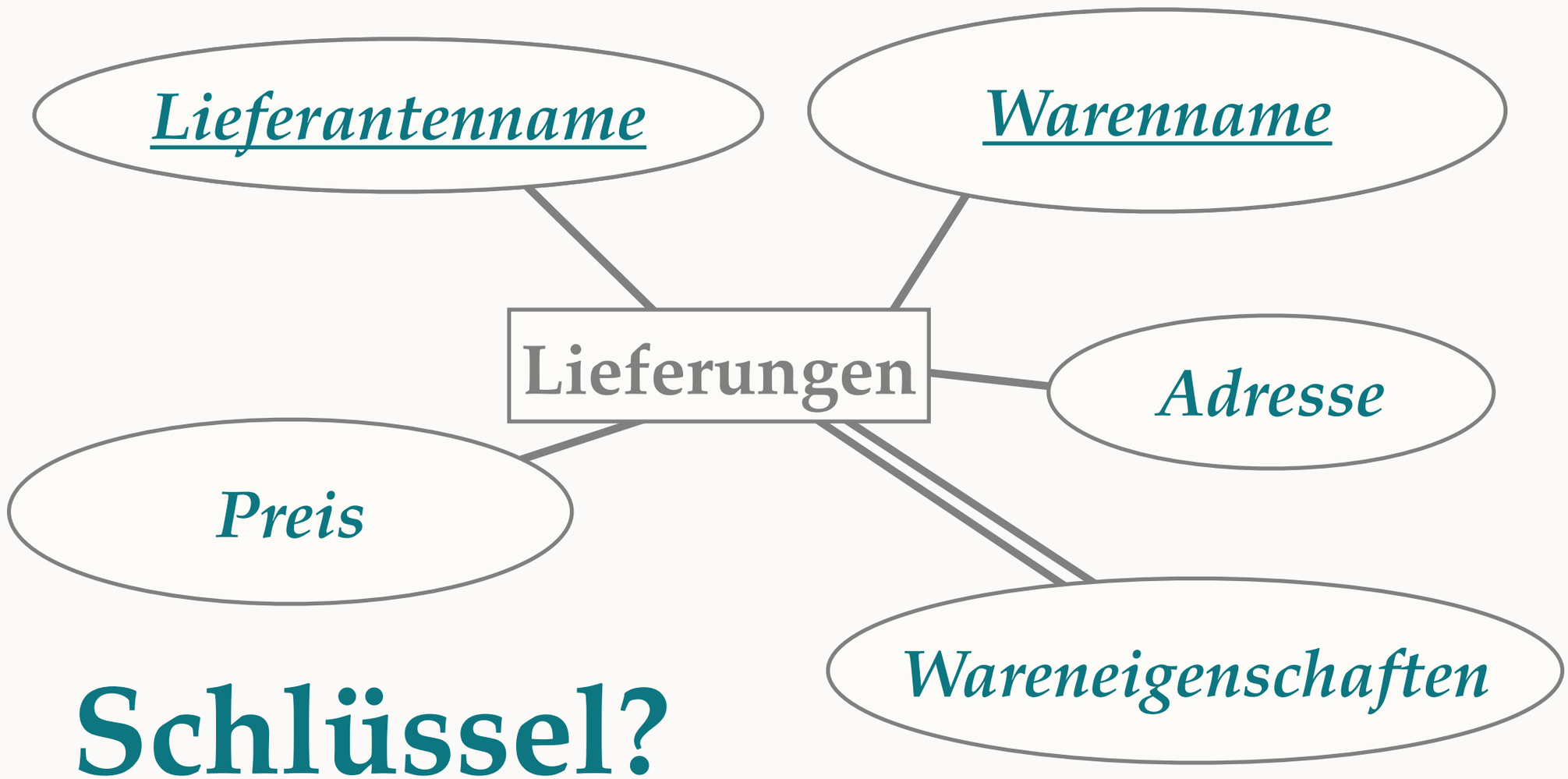
BEISPIEL

Beschreibung des Geschäftsfeldes:

Eine Firma bezieht ihre Waren über verschiedene Lieferanten. Jeder ***Lieferant*** liefert bestimmte ***Waren*** mit bestimmten ***Eigenschaften*** zu einem bestimmten ***Preis***. Für jeden Lieferanten soll die ***Adresse*** gespeichert werden.

BEISPIEL

EER-Diagramm:



Schlüssel?

BEISPIEL

Relationenmodell:

Lieferungen (<u>LName</u> [Text 50 Zeichen, nicht leer],
LAdresse [Text 100 Zeichen],
<u>WName</u> [Text 50 Zeichen, nicht leer],
Preis [positive ganze Zahl],
Eigenschaften [Text so groß wie möglich]
)

BEISPIEL

Tabellenerstellungs-Skript:

```
CREATE TABLE Lieferungen (
    LName VARCHAR(50) NOT NULL,
    LAdresse VARCHAR(100) NOT NULL,
    WName VARCHAR(50),
    Preis INT,
    Eigenschaften VARCHAR(255),
    PRIMARY KEY (LName, WName)
) Engine=InnoDB;
```

BEISPIEL

resultierende Tabelle Lieferungen:

LName	LAdresse	WName	Preis	Eigenschaften
Heinz Meier GmbH	Troststraße 25, 1100 Wien	Lattenrost 200x100	75	Härte 3, nicht lackiert, Lieferzeit 1 Tag
Frischwaren Kg	Dammgasse 15, 1120 Wien	Gänsefedern Klasse 1	22	von glücklichen Gänsen, superweich
Elektromax	Hirtplatz 27, 1120 Wien	Nachttischlampe „Hellfire“	78	mit Montagematerial, Lieferzeit 3 Tage
Elektromax	Hirtplatz 27, 1120 Wien	Nachttischlampe „Düsterwald“	64	ohne Montagematerial, Selbstabholung

Probleme: Redundanz Mehrwertigkeit

optimal?

PROBLEM: MEHRWERTIGKEIT

Bedeutung:

In einem Feld sind mehrere Werte gespeichert. Man sagt, das Feld ist ***nicht atomar.***

- ★ verhindert genauere Spezifikation und Einschränkung der Werte
  z.B: Feld Eigenschaften: „Lieferzeit: 1 Woche“
  -> keine Einschränkung auf Zahlen oder gültige Werte möglich

- ★ verhindert/erschwert Suchmöglichkeiten:
  z.B. „In welchen Bezirken kann morgen die Ware geholt werden“
  -> benötigt PLZ und Lieferzeit, die nur Teil eines Feldes sind.

LÖSUNG: ATOMARE DATEN

Ein Feld mit ***atomaren Daten*** darf nicht mehrere Datenwerte gleicher Art enthalten

Idee:

Preis	Eigenschaft1	Eigenschaft2	Eigenschaft3
75	Härte 3	Lieferzeit 1 Tag	nicht lackiert
12	von glücklichen Gänsen	superweich	NULL
78	mit Montagematerial	Lieferzeit 3 Tage	NULL
64	ohne Montagematerial	Selbstabholung	NULL

Schlechte Idee

Reihenfolge? Erweiterbarkeit? Null-Werte?

ATOMARE DATEN

Eine Datensatz mit ***atomaren Daten*** darf nicht mehrere Felder mit der gleichen Art von Daten (Feld-Gruppen) enthalten.

Aber ein Feld darf in ***verschiedene Arten*** von Information getrennt werden:

LName	LAdresse	LPLZ	LOrt
Heinz Meier GmbH	Troststraße 25	1100	Wien
Frischwaren Kg	Dammgasse 15	1120	Wien
Elektromax	Hirtplatz 27	1120	Wien
Elektromax	Hirtplatz 27	1120	Wien

1. NORMALFORM (1NF)

Definition:

Ein Datenbank-Schema ist dann in einer NF, wenn alle zugehörigen Tabellen in dieser NF sind.

Eine Tabelle ist dann in 1NF, wenn

1. jeder Datensatz einen eindeutigen Bezeichner (***Primärschlüssel***) hat **und**
2. jeder Datensatz ausschließlich ***atomare*** Werte enthält (atomare Felder, keine Feld-Gruppen)

WIE ATOMAR IST ATOMAR?

Ist diese Tabelle atomar?

SVNr	Name	Adresse
1234020499	Mitzi Meier	Birkenstraße 28, 1200 Wien
9999131204	Hansi Müller	Donauplatz 3, 3400 Klosterneuburg

für den Pizzalieferanten wird das genügen

Jetzt vielleicht?

für das Meldeamt wird diese Form wohl besser sein

SVNr	Vorname	Nachname	Straße	Haus nummer	Tür nummer	PLZ	Ort
1234020499	Mitzi	Meier	Birkenstraße	28	NULL	1200	Wien
9999131204	Hansi	Müller	Donauplatz	3	11	3400	Klosterneuburg

Was atomar ist, hängt vom Anwendungsfall ab!

ÜBUNGSBEISPIEL

Bringe diese Tabelle in 1NF:

Superhelden

Name (Initialen)	Fähigkeiten	Stadt	Erzfeind
Müllmann (MM)	reinigt gut, Angst vor Bleiche	Gotham	ID: 4, Stadt: Gotham
Börsenmann (BÖ)	macht Geld aus nichts, fürchtet Finanzamt	Vaduz	ID: 8, beheimatet in Berlin
Supertyp (SU)	kann fliegen, aber Angst vor Vögeln	Gotham	ID: 3, Stadt: Metropolis
Vogelfrau (VF)	kann fliegen, mag keine Flugzeuge	Metropolis	ID: 4, tätig in Gotham
Supertyp (SU)	hat Superkraft nur nicht gegen Gummi	Rom	ID: 5, Cannes

Name	Kraft	Schwäche	Stadt	Initialen	Erzfeind-ID	Erzfeind-Stadt
Müllmann	reinigt gut	Bleiche	Gotham	MM	4	Gotham
Börsenmann	macht Geld aus nichts	Finanzamt	Vaduz	BÖ	8	Berlin
Supertyp	kann fliegen	Vögel	Gotham	SU	3	Metropolis
Vogelfrau	kann fliegen	Flugzeuge	Metropolis	VF	4	Gotham
Supertyp	Superkraft	Gummi	Rom	SU	5	Cannes

ZURÜCK ZUM BEISPIEL

da gibt es noch ein anderes Problem:

LName	LAdresse	WName	Preis	Eigenschaften
Heinz Meier GmbH	Troststraße 25, 1100 Wien	Lattenrost 200x100	75	Härte 3, nicht lackiert, Schulterzone, Lieferzeit 1 Tag
Frischwaren Kg	Dammgasse 15, 1120 Wien	Gänsefedern Klasse 1	22	von glücklichen Gänsen, Lieferzeit 3 Wochen
Elektromax	Hirtplatz 27, 1120 Wien	Nachttischlampe „Hellfire“	78	mit Montagematerial, Lieferzeit 3 Tage
Elektromax	Hirtplatz 27, 1120 Wien	Nachttischlampe „Düsterwald“	64	ohne Montagematerial, Lieferzeit 1 Woche

Probleme: Redundanz

Mehrwertigkeit

PROBLEM: REDUNDANZ

Bedeutung:

Für jede Ware wird die Adresse des Lieferanten gespeichert -> die Adresse ist mehrfach vorhanden

- führt zu unnötiger Speicherverschwendung
- verursacht ***Änderungsanomalien***
  - UPDATE: Adressänderungen in mehr als 1 Datensatz notwendig.
  - INSERT: Einfügen eines Lieferanten erfordert Ware.
  - DELETE: Löschen der letzten Ware löscht eines Lieferanten löscht alle weiteren Infos wie Adresse des Lieferanten.

PROBLEM: REDUNDANZ

Was tun?

Um das Problem der Redundanzen (und andere) zu lösen, müssen wir uns die Beziehungen der Felder untereinander ansehen.

Wir kennen bereits Schlüssel als besondere Felder. Schlüssel machen Datensätze eindeutig. Man kann sagen, verschiedene Schlüssel bedeuten i.A. verschiedene Daten in den Nichtschlüssel-Feldern.

Die Nichtschlüssel-Felder sind also in gewisser Weise von den Schüsselfeldern ***abhängig***.

ABHÄNGIGKEITEN

Definition:

Wenn sich Daten eines Feldes ändern (könnten), weil die Daten eines anderes Feldes geändert wurden, ist das erste Feld vom zweiten ***funktional abhängig***.

Kurzsschreibweise: A → B

man sagt B ist von A funktional abhängig

z.B:

LName → LAdresse

... verschiedene Lieferanten haben in der Regel verschiedene Adressen

LName,Ware → Preis

... ändert sich der Lieferant oder die Ware ändert sich damit auch (ev.) der Preis

ABHÄNGIGKEITEN

Es gibt Unterschiede:

LName,Ware → Preis

Der Preis hängt vom gesamten Schlüssel ab. Ändert sich eines der beiden Schlüsselfelder, könnte sich auch der Preis ändern.

-> Preis ist ***voll funktional abhängig***

LName → LAdresse

Die Adresse hängt nur vom Lieferantennamen ab. Sie ist unabhängig von der eingetragenen Ware.

-> LAdresse ist ***partiell funktional abhängig***

2. NORMALFORM (2NF)

Definition:

Eine Tabelle ist dann in 2NF, wenn

1. sie in 1NF ist **und**

2. keine ***partiell funktionalen*** Abhängigkeiten besitzt.

Daraus resultierende Hilfsregeln:

1. Tabellen mit einspaltigem ***synthetischem*** Primärschlüssel sind automatisch in 2NF.

2. Tabellen, die ausschließlich Primärschlüsselfelder haben sind automatisch in 2NF (z.B. m:n-Hilfstabellen).

ÜBUNGSBEISPIEL

Suche Abhängigkeiten und ihre Art:

Superhelden

Name	Kraft	Schwäche	Stadt	Initialen	Erzfeind-ID	Erzfeind-Stadt
Müllmann	reinigt gut	Bleiche	Gotham	MM	4	Gotham
Börsenmann	macht Geld aus nichts	Finanzamt	Vaduz	BÖ	8	Berlin
Supertyp	kann fliegen	Vögel	Gotham	SU	3	Metropolis
Vogelfrau	kann fliegen	Flugzeuge	Metropolis	VF	4	Gotham
Supertyp	Superkraft	Gummi	Rom	SU	5	Cannes

Name → Initialen (partiell funktional abhängig)
Name, Kraft → Schwäche (voll funktional abhängig)
...

Name $\not\rightarrow$ Erzfeind-Stadt? oder Kraft $\not\rightarrow$ Erzfeind-Stadt?

Erzfeind-Stadt ist in der Tabelle zu den Schlüsselfeldern völlig ***unabhängig***.

ZURÜCK ZUM BEISPIEL

Wie erreichen wir 2NF?

(betrachten wir die Spalte Eigenschaften vorläufig als atomar)

LName	LAdresse	WName	Preis	Eigenschaften
Heinz Meier GmbH	Troststraße 25, 1100 Wien	Lattenrost 200x100	75	Härte 3, nicht lackiert, Schulterzone, Lieferzeit 1 Tag
Frischwaren Kg	Dammgasse 15, 1120 Wien	Gänsefedern Klasse 1	22	von glücklichen Gänsen, Lieferzeit 3 Wochen
Elektromax	Hirtplatz 27, 1120 Wien	Nachttischlampe „Hellfire“	78	mit Montagematerial, Lieferzeit 3 Tage
Elektromax	Hirtplatz 27, 1120 Wien	Nachttischlampe „Düsterwald“	64	ohne Montagematerial, Lieferzeit 1 Woche

Lösung: Zerlegung in mehrere Tabellen.

ZERLEGUNG ZU 2NF

Wie zerlegen?

1. Zuerst müssen wir für jedes Nichtschlüssel-Feld bestimmen, wie es von den Schlüsseln abhängig ist.

2. Alle partiell funktional abhängigen Felder, die von den gleichen Feldern abhängen, werden jeweils in eine neue Tabelle verschoben.
   Primärschlüssel der neuen Tabelle werden die Felder der jeweils linke Seite der funktionalen Abhängigkeit.
   Gleichzeitig werden diese ***Fremdschlüssel*** in der ursprünglichen Tabelle.

3. Alle voll funktional abhängigen Felder verbleiben in der Tabelle.

ZERLEGUNG ZU 2NF

Beispiel:

1. Abhängigkeit der Nichtschlüssel-Felder bestimmen

LName	LAdresse	LPLZ	LOrt	WName	Preis	Eigenschaften
Heinz Meier GmbH	Troststraße 25	1100	Wien	Lattenrost 200x100	75	Härte 3, nicht lackiert, Schulterzone, Lieferzeit 1 Tag

gemeinsame linke Seite

LName → LAdresse (partiell funktional abhängig)

LName → LPLZ (partiell funktional abhängig)

LName → LOrt (partiell funktional abhängig)

LName,WName → Preis (voll funktional abhängig)

LName,WName → Eigenschaften (voll funktional abhängig)

ZERLEGUNG ZU 2NF

Beispiel:

2. Zerlegung in mehrere Tabellen

LName	LAdresse	LPLZ	LOrt	WName	Preis	Eigenschaften
Heinz Meier GmbH	Troststraße 25	1100	Wien	Lattenrost 200x100	75	Härte 3, nicht lackiert, Schulterzone, Lieferzeit 1 Tag

Lieferanten

LName	LAdresse	LPLZ	LOrt
Heinz Meier GmbH	Troststraße 25	1100	Wien

Alle partiell funktional abhängigen Felder kommen in eine neue Tabelle, mit der gemeinsamen linken Seite (LName) als Primärschlüssel

LName wird Fremdschlüssel

alle voll funktional abhängigen Felder bleiben in der ursprünglichen Tabelle

Lieferungen

LName	WName	Preis	Eigenschaften
Heinz Meier GmbH	Lattenrost 200x100	75	Härte 3, nicht lackiert, Schulterzone, Lieferzeit 1 Tag

ÜBUNGSBEISPIEL

Bringe diese Tabelle in 2NF:

Autoverleih

Kunde	Datum	Uhrzeit	Fahrzeug	Dauer	Tarif	Type	Hersteller
Hr. Müller	28.02.08	08:00	W 234C	2	Werktag Frühaufsteher	Focus 2l	Ford
Fr. Meier	29.02.08	19:00	W 234C	24	Schalttag Abendtarif	Focus 2l	Ford
Hr. Hinter	29.02.08	10:00	W 9988B	6	Schalttag Geschäftszeit	Golf 5 GTI	VW
Fr. Vorder	23.03.08	10:00	W 9988B	4	Wochenend-tarif	Golf 5 GTI	VW

gemeinsame linke Seiten

1. Abhängigkeiten finden

Datum,Uhrzeit,Fahrzeug → Kunde (voll funktional abhängig)
Datum,Uhrzeit,Fahrzeug → Dauer (voll funktional abhängig)
Datum,Uhrzeit → Tarif (partiell funktional abhängig)
Fahrzeug → Type (partiell funktional abhängig)
Fahrzeug → Hersteller (partiell funktional abhängig)

ÜBUNGSBEISPIEL

2. Tabelle zerlegen - für jede linke Seite eine neue Tabelle

Tarife

Datum	Uhrzeit	Tarif
28.02.08	08:00	Werktag Frühaufsteher

Datum,Uhrzeit → Tarif

Fahrzeuge

Fahrzeug	Type	Hersteller
W 234C	Focus 2l	Ford

Fahrzeug → Type
Fahrzeug → Hersteller

3. voll funk. Felder in ursprünglicher Tabelle lassen - Fremdschlüssel anlegen

Autoverleih

Kunde	*Datum*	*Uhrzeit*	*Fahrzeug*	Dauer
Hr. Müller	28.02.08	08:00	W 234C	2

Fremdschlüssel nicht vergessen

Datum,Uhrzeit,Fahrzeug → Kunde
Datum,Uhrzeit,Fahrzeug → Dauer

UNABHÄNGIGKEIT

Wir erinnern uns an das Superhelden-Beispiel:

Name $\not\rightarrow$ Erzfeind-Stadt? oder Kraft $\not\rightarrow$ Erzfeind-Stadt?

Erzfeind-Stadt ist in der Tabelle zu den Schlüsselfeldern völlig ***unabhängig***.

Superhelden

Name	Kraft	Schwäche	Stadt	Erzfeind-ID	Erzfeind-Stadt
Müllmann	reinigt gut	Bleiche	Gotham	4	Gotham
Börsenmann	macht Geld aus nichts	Finanzamt	Vaduz	8	Berlin

Felder, die von keinem anderen Feldern abhängig sind nennt man:

(vollständig funktional) unabhängig

TRANSITIVE ABHÄNGIGKEIT

Aber stimmt das in diesem Fall?

Nein, denn

Erzfeind-ID → Erzfeind-Stadt

Daher: Erzfeind-Stadt ist von einem Nichtschlüssel-Feld funktional abhängig.

Eine Abhängigkeit eines Nichtschlüssel-Feldes von einem anderen Nichtschlüssel-Feld nennt man eine

transitive funktionale Abhängigkeit

3. NORMALFORM (3NF)

Definition:

Eine Tabelle ist dann in 3NF, wenn

1. sie in 2NF ist **und**

2. keine ***transitiven*** Abhängigkeiten besitzt.

3. NORMALFORM (3NF)

Wie erreichen wir 3NF?

Superhelden

Name	Kraft	Schwäche	Stadt	Erzfeind-ID	Erzfeind-Stadt
Müllmann	reinigt gut	Bleiche	Gotham	4	Gotham
Börsenmann	macht Geld aus nichts	Finanzamt	Vaduz	8	Berlin

Lösung: Zerlegung in mehrere Tabellen (wie bei 2NF)

Superhelden

Name	Kraft	Schwäche	Stadt	*Erzfeind-ID*
Müllmann	reinigt gut	Bleiche	Gotham	4
Börsenmann	macht Geld aus nichts	Finanzamt	Vaduz	8

Erzfeinde

Erzfeind-ID	Stadt
4	Gotham
8	Berlin

ÜBUNGSBEISPIEL

Bringe diese Tabelle in 3NF:

Pferderennen

Rennen#	Pferd	Gestüt	Nachname	Vorname	Straße	PLZ	Stadt	Bundesland	Renndatum
1	Hasibutz	Gaulhof	Superjockey	Lisa	Haferg. 3	5360	Ried	Salzburg	02.05.08

1. Tabelle auf 2NF kontrollieren: nur ein synthetischer Schlüssel → 2NF

2. transitive Abhängigkeiten ermitteln

Pferd → Gestüt

Nachname, Vorname → Straße, PLZ, Stadt, Bundesland

PLZ → Stadt, Bundesland

ÜBUNGSBEISPIEL

2. Tabelle zerlegen - für jede linke Seite eine neue Tabelle

Pferde

Pferd	Gestüt
Hasibutz	Gaulhof

Pferd → Gestüt

kein brauchbarer Schlüssel, daher synthetischen Ersatzschlüssel bilden

Jockeys

JockeyID	Nachname	Vorname	Straße	*PLZ*
1	Superjockey	Lisa	Haferg. 3	5360

Nachname, Vorname → Straße, PLZ, Stadt, Bundesland

Orte

PLZ	Stadt	Bundesland
5360	Ried	Salzburg

PLZ → Stadt, Bundesland

3. nicht transitiv abhängige Felder in Tabelle lassen - Fremdschlüssel anlegen

Pferderennen

Rennen#	*Pferd*	*JockeyID*	Datum
1	Hasibutz	1	02.05.08

SINN VON NORMALFORMEN

Durch die Normalisierung erreichen wir:

- ★ Daten sind einfach abzufragen, da immer nur ein Wert pro Feld vorhanden ist.
- ★ Nur wirklich zusammengehörige Daten stehen in einer Tabelle.

Aber:

Durch gute Modellierung bereits des ER-Diagramms können viele Normalisierungsschritte vorweg-genommen werden.

MEHR NORMALFORMEN

Ausblick:

Außer den drei vorgestellten Normalformen gibt es noch strengere wie die

- ⋆ Boyce-Codd-Normalform
- ⋆ 4. Normalform

Mit diesen werden wir uns aber nicht beschäftigen (→ Universität).