

Übung zur Vorlesung *Grundlagen: Datenbanken* im WS16/17

Harald Lang, Linnea Passing (gdb@in.tum.de)

<http://www-db.in.tum.de/teaching/ws1617/grundlagen/>

Blatt Nr. 10

Hausaufgabe 1

Geben Sie für jede der Normalformen 1NF, 2NF, 3NF, BCNF, 4NF jeweils eine Relation mit FDs an, so dass die Relation in der gewünschten Normalform ist (und in keiner höheren).

Lösung:

Für alle Normalformen betrachten wir die Relation $\mathcal{R} = \{A, B, C, D\}$.

- 1.NF:

FDs:

- $AB \rightarrow C$
- $B \rightarrow D$

Die Relation ist nicht mengenwertig, daher 1. NF. D ist lediglich von B abhängig, der Kandidatenschlüssel ist aber AB , weswegen D nicht voll funktional vom Kandidatenschlüssel abhängig ist, daher keine 2. NF.

- 2. NF:

FDs:

- $AB \rightarrow C$
- $C \rightarrow D$

Jedes Attribut der Relation ist voll funktional abhängig vom Kandidatenschlüssel AB , daher 2.NF. Das Attribut D ist transitiv und nicht direkt vom Kandidatenschlüssel abhängig, darum nicht 3. NF.

- 3. NF:

FDs:

- $AB \rightarrow CD$
- $BC \rightarrow AD$
- $D \rightarrow C$

Für alle FDs gilt entweder, dass sie trivial ist, dass die linke Seite Superschlüssel ist oder dass die rechte Seite in einem Kandidatenschlüssel enthalten ist, daher 3. NF. Bei der BCNF fällt die dritte erlaubte Art von FD weg, daher FDs müssen trivial sein oder ihre linke Seite Superschlüssel. Da die dritte FD des Beispiels dies verletzt ist die Relation nicht in BCNF und daher genau in 3. NF.

- BCNF:

FDs:

- $AB \rightarrow CD$
- $BC \rightarrow AD$
- $D \twoheadrightarrow C$

BCNF, da die BCNF verletzende FD aus dem Beispiel für 3. NF entfernt wurde. Nicht 4. NF weil eine nicht trivial MVD gilt, deren linke Seite nicht Superschlüssel ist.

• 4. NF

FDs:

- $AB \rightarrow CD$
- $BC \rightarrow AD$

Nach Entfernung der nicht trivialen MVD dann auch 4. NF.

Hausaufgabe 2

Überführen Sie verlustlos in die 4. NF:

Relation

$$\mathcal{R} = \{A, B, C, D\}$$

mit den FDs/MVDs

$$\begin{array}{l} AB \twoheadrightarrow C \\ BC \twoheadrightarrow D \\ BA \rightarrow CD \\ DA \rightarrow B \end{array}$$

Lösung:

Kandidatenschlüssel: $\kappa_1 = \{A, B\}$ und $\kappa_2 = \{A, D\}$

- Betrachte $AB \twoheadrightarrow C$ in \mathcal{R} :
 - MVD ist nicht trivial ($AB \cap C = \emptyset$ und $AB \cup C \subset \mathcal{R}$),
 - aber die linke Seite ist ein Superschlüssel
 - und stellt somit keine Verletzung der 4. NF dar.
- Betrachte $BC \twoheadrightarrow D$ in \mathcal{R} :
 - MVD ist nicht trivial ($BC \cap D = \emptyset$ und $BC \cup D \subset \mathcal{R}$),
 - und die linke Seite ist **kein** Superschlüssel.
 - Die 4. NF wird dadurch verletzt.
- Zerlege \mathcal{R} anhand $BC \twoheadrightarrow D$ in
 - $\mathcal{R}_1 = \{B, C\} \cup \{D\} = \{B, C, D\}$, und
 - $\mathcal{R}_2 = \mathcal{R} \setminus \{D\} = \{A, B, C\}$, mit den FDs/MVDs
 - $F_{\mathcal{R}_1} = \{\}$, $D_{\mathcal{R}_1} = \{BC \twoheadrightarrow D\}$, und
 - $F_{\mathcal{R}_2} = \{BA \rightarrow C\}$, $D_{\mathcal{R}_2} = \{AB \twoheadrightarrow C\}$.

Die FDs $BA \rightarrow D$ und $DA \rightarrow B$ gehen verloren.

- Betrachte $BC \twoheadrightarrow D$ in \mathcal{R}_1 :
 - MVD ist trivial, da $BC \cup D = \mathcal{R}_1$,
 - und stellt somit keine Verletzung der 4. NF dar.
 - \mathcal{R}_1 kann nicht weiter zerlegt werden.
- Betrachte $AB \twoheadrightarrow C$ in \mathcal{R}_2 :
 - MVD ist trivial, da $AB \cup C = \mathcal{R}_2$,
 - und stellt somit keine Verletzung der 4. NF dar.
 - \mathcal{R}_2 kann nicht weiter zerlegt werden.

Ergebnis: $\mathcal{R}_1 = \{\underline{B}, \underline{C}, D\}$, $\mathcal{R}_2 = \{\underline{A}, B, C\}$

Hausaufgabe 3

Gegeben sei die durch folgende SQL-Statements definierte Ausprägung einer Relation.

```
create table kinder_fahrraeder (  
  person varchar(100) not null,  
  kind_name varchar(100) not null,  
  kind_alter integer not null,  
  fahrrad_typ varchar(100) not null,  
  fahrrad_farbe varchar(100) not null  
);  
insert into kinder_fahrraeder values  
( 'Thomas', 'Markus', 10, 'Trekking-Fahrrad', 'schwarz' ),  
( 'Thomas', 'Markus', 10, 'Mountainbike', 'rot' ),  
( 'Thomas', 'Johanna', 5, 'Trekking-Fahrrad', 'schwarz' ),  
( 'Thomas', 'Johanna', 5, 'Mountainbike', 'rot' );
```

Es gelten die beiden komplementären MVDs

- 1) $person \twoheadrightarrow \{kind_name, kind_alter\}$ und
- 2) $person \twoheadrightarrow \{fahrrad_typ, fahrrad_farbe\}$

sowie die FD

- 3) $kind_name \rightarrow kind_alter$.

- (a) Laura, das dritte Kind von Thomas, wird geboren. Fügen Sie Laura per SQL-Insert-Statement hinzu und beachten Sie dabei die MVDs. Formulieren Sie Ihr Statement so, dass es auch ohne Kenntnis der Fahrräder von Thomas funktioniert (d.h. nicht `insert ... 'Mountainbike', 'rot'`);).
- (b) Allgemein gesprochen: In eine Relation $R = \{A, B, C\}$ mit den MVDs $A \twoheadrightarrow B$ und $A \twoheadrightarrow C$ soll für ein bestimmtes a in Spalte A ein neuer Wert b in Spalte B eingefügt werden. Wie viele Tupel müssen hinzugefügt werden, damit die MVDs weiterhin gelten?
- (c) Was passiert, wenn Thomas seine beiden Fahrräder verkauft?
- (d) Überführen Sie die Relation `kinder_fahrraeder` mit dem Dekompositionsalgorithmus in die 4. NF.

- (e) Schreiben Sie ein SQL-Statement um zu prüfen ob die MVDs der Relation `kinder_fahrraeder` erfüllt sind.

Lösung:

- (a) Zunächst werden alle Fahrräder von Thomas ermittelt, anschließend wird Laura kombiniert mit jedem Fahrrad in die Tabelle eingefügt.

```
with fahrraeder as (
  select distinct person, fahrrad_typ, fahrrad_farbe
  from kinder_fahrraeder
  where person='Thomas'
)
insert into kinder_fahrraeder (
  select person, 'Laura', '0', fahrrad_typ, fahrrad_farbe
  from fahrraeder
);
```

- (b) b muss kombiniert mit jedem Wert in C eingefügt werden, die Anzahl neuer Tupel ist daher:

```
select count(distinct C) from R where A=a;
```

- (c) Da die fahrradbezogenen Spalten nicht nullable sind, müssen sämtliche Tupel gelöscht werden, sodass auch die Informationen zu den Kindern verloren gehen. Durch Normalisierung kann diese Anomalie verhindert werden.

- (d) Zunächst wird anhand der MVDs zerlegt, wir erhalten die Relationen:

$$\mathcal{R}_1 = \{\underline{\text{person}}, \underline{\text{kind_name}}, \text{kind_alter}\} \text{ und}$$

$$\mathcal{R}_2 = \{\underline{\text{person}}, \underline{\text{fahrrad_typ}}, \underline{\text{fahrrad_farbe}}\}.$$

Anschließend wird R_1 anhand der FD weiter zerlegt:

$$\mathcal{R}_{1a} = \{\text{person}, \underline{\text{kind_name}}\} \text{ und}$$

$$\mathcal{R}_{1b} = \{\underline{\text{kind_name}}, \text{kind_alter}\} \text{ und}$$

$$\mathcal{R}_2 = \{\underline{\text{person}}, \underline{\text{fahrrad_typ}}, \underline{\text{fahrrad_farbe}}\}.$$

Alternativ kann auch mit der Zerlegung anhand der FD begonnen werden, die finale Zerlegung ist dieselbe.

- (e) Eine Möglichkeit ist es, das Kreuzprodukt aus allen Fahrrädern und Kindern pro Person zu bilden, und die tatsächliche Ausprägung abzuziehen. Ist das Ergebnis des SQL-Statements leer, so sind die MVDs erfüllt.

```
with kinder as (
  select distinct person, kind_name, kind_alter
  from kinder_fahrraeder
), fahrraeder as (
  select distinct person, fahrrad_typ, fahrrad_farbe
  from kinder_fahrraeder
)
select f.person, k.kind_name, k.kind_alter, f.fahrrad_typ, f.fahrrad_farbe
from fahrraeder f, kinder k
where f.person=k.person
except
select * from kinder_fahrraeder;
```

Hausaufgabe 4

Gegeben sei das folgende Schema:

$$\text{Lala} = \{\text{Festival, TicketID, Band}\}$$

Bestimmen Sie die geltenden Abhängigkeiten zwischen den Attributen, sodass sich mithilfe eines aus der Vorlesung bekannten Algorithmus eine semantisch sinnvolle Zerlegung ergibt. Geben Sie die Zerlegung und den verwendeten Algorithmus an.

Lösung:

Es gilt die FD: $\text{TicketID} \rightarrow \text{Festival}$, nicht jedoch $\text{TicketID} \rightarrow \text{Band}$, da i. A. mehrere Bands auftreten.

Nach Anwendung des Synthesealgorithmus erhält man:

$$\{\underline{\text{TicketID}}, \text{Festival}\} \text{ sowie } \{\underline{\text{TicketID}}, \text{Band}\} = \kappa$$

Die Relation $\{\underline{\text{TicketID}}, \text{Band}\}$ ist im Kontext von Festivals jedoch wenig hilfreich.

Eine sinnvollere Zerlegung in

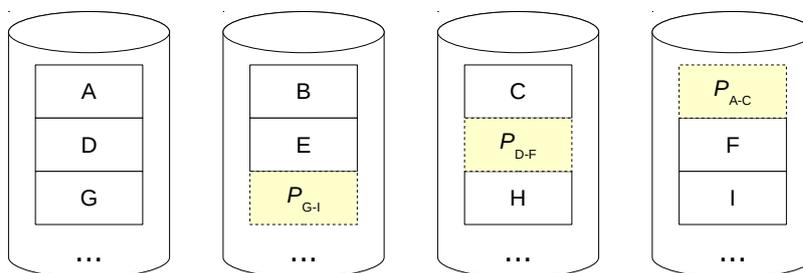
$$\{\underline{\text{TicketID}}, \text{Festival}\} \text{ und } \{\underline{\text{Festival}}, \text{Band}\}$$

erhält man nur dann, wenn auch die MVD $\text{Festival} \twoheadrightarrow \text{Band}$ berücksichtigt wird.

Die Zerlegung anhand des Dekompositionsalgorithmus für die 4.NF ist jedoch nicht eindeutig. Es muss als erstes die MVD $\text{Festival} \twoheadrightarrow \text{Band}$ betrachtet werden; andernfalls führt der Dekompositionsalgorithmus zum gleichen Ergebnis wie der Synthesealgorithmus.

Hausaufgabe 5

Die folgende Abbildung zeigt einen Festplattenverbund bestehend aus vier Laufwerken, auf welchen die Datenblöcke A bis I gespeichert sind. Die Blöcke P_i enthalten Paritätsinformationen.



- Um welches RAID-Level handelt es sich?
- Wieviele Festplatten können ausfallen, ohne dass mit Datenverlust zu rechnen ist? Geben Sie eine allgemeine Lösung für einen Verbund bestehend aus n Festplatten an.
- Kann die Ausfallsicherheit erhöht werden? Begründung?

- d) Welchen weiteren Vorteil bietet das gezeigte RAID-System neben der Ausfallsicherheit?
- e) Nach einem Festplattendefekt enthalten die Datenblöcke die folgenden Binärdaten. Rekonstruieren Sie die Datenblöcke der $Disk_2$ mithilfe der XOR-Verknüpfung.

$Disk_0$	$Disk_1$	$Disk_2$	$Disk_3$
A = 1111	B = 1001	C = - - - -	P_{A-C} = 1110
D = 0101	E = 1100	P_{D-F} = - - - -	F = 1100
G = 0011	P_{G-I} = 1110	H = - - - -	I = 0011

Lösung:

- a) 5
- b) 1, unabhängig von n .
- c) Ja, z.B. mit einem RAID-6 (Ausfall zweier Platten kann kompensiert werden) oder RAID-15 (das RAID-5 wird zusätzlich nochmal gespiegelt).
- d) Höherer Datendurchsatz.
- e) Die Rekonstruktion der Datenblöcke unterscheidet sich rechnerisch nicht von der Berechnung der Parität.

$Disk_0$	$Disk_1$	$Disk_2$	$Disk_3$
A = 1111	B = 1001	C = 1000	P_{A-C} = 1110
D = 0101	E = 1100	P_{D-F} = 0101	F = 1100
G = 0011	P_{G-I} = 1110	H = 1110	I = 0011