Übung zur Vorlesung *Grundlagen: Datenbanken* im WS17/18

Harald Lang, Linnea Passing (gdb@in.tum.de)
http://www-db.in.tum.de/teaching/ws1718/grundlagen/

Blatt Nr. 09

Tool zum Üben der Normalformen: `http://normalizer.db.in.tum.de/`

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 wie die Relation $\mathcal{R} = \{A, B, C, D\}$.

- 1.NF:

  FDs:

  - $AB \to C$
  - $B \to D$

  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:

  - $AB \to C$
  - $C \to D$

  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:

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

  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:

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

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:

  - $AB \to CD$
  - $BC \to AD$

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

Hausaufgabe 2

Bestimmen Sie alle Kandidatenschlüssel der Relation R. Wenden Sie den Dekompositionsalgorithmus an, um die Relation R in die BCNF zu zerlegen und unterstreichen Sie die Schlüssel der Teilrelationen des Endergebnisses.

$$R = \{A, B, C, D, E, F\}$$

FDs:

1. $B \to DA$
2. $DEF \to B$
3. $C \to EA$

Prüfen Sie als erstes FD 1) ob Sie für die Zerlegung geeignet ist und - falls dies der Fall ist - verwenden Sie diese im ersten Zerlegungsschritt. Für diese Aufgabe ist zu bedenken, dass die oben angegebenen FDs eine Charakterisierung der insgesamt geltenden FDs sind. Die Menge der geltenden FDs ist größer. Wieso? Wie muss dies beim Dekompositionsalgorithmus genutzt werden?

Lösung:

- Dekompositionsalgorithmus:
  - Starte mit $Z := \{R\}$.
  - R in BCNF? - Nein, $B \to DA$ verletzt die BCNF.
    - Zerlegung anhand FD $B \to DA$, da $\{B\}$ kein Superschlüssel:
      $R_1 = \{A, B, D\}$ mit den FDs $F_1 = \{B \to DA\}$,
      $R_2 = \{B, C, E, F\}$ mit den FDs $F_2 = \{C \to E\}$, FD (2) geht verloren und FD (3) geht "teilweise" verloren: wenn $C \to AE$ gilt, dann gilt auch $C \to A$ und $C \to E$ (Dekompositionsregel), aber lediglich $C \to E$ bleibt erhalten.
      $Z := \{R_1, R_2\}$
  - R_1 in BCNF? - Ja.
  - R_2 in BCNF? - Nein, $C \to E$ verletzt die BCNF.

    * Zerlegung anhand FD $C \to E$, da $\{C\}$ kein Superschlüssel:
      $R_{2.1} = \{C, E\}$ mit den FDs $F_{2.1} = \{C \to E\}$,
      $R_{2.2} = \{B, C, F\}$ mit ausschließlich trivialen FDs.
      $Z := \{R_1, R_{2.1}, R_{2.2}\}$
  - $R_{2.1}$ in BCNF? - Ja.
  - $R_{2.2}$ in BCNF? - Ja.
- Ergebnis:

$$\begin{aligned} R_1 &= \{A, \underline{B}, D\} \\ R_{2.1} &= \{\underline{C}, E\} \\ R_{2.2} &= \{\underline{B}, \underline{C}, \underline{F}\} \end{aligned}$$

Im Allgemeinen ist eine gegebene FD-Menge weder minimal noch vollständig. Die angegebenen FDs enthalten also möglicherweise Redundanzen einerseits und andererseits werden triviale Abhängigkeiten i.d.R. nicht explizit mit angegeben. Bei der Ausführung des Dekompositionsalgorithmus müssen jedoch alle *geltenden* FDs betrachtet werden, die sich mit Hilfe der Axiome von Armstrong herleiten lassen (F^+). So gilt im obigen Beispiel in R_2 die FD $C \to E$, obwohl diese nicht explizit angegeben war.

Hausaufgabe 3

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: TicketID$\to$Festival, nicht jedoch TicketID$\to$Band, da i. A. mehrere Bands auftreten.

Nach Anwendung des Synthesealgorithmus erhält man:

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

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

Eine sinnvollere Zerlegung in

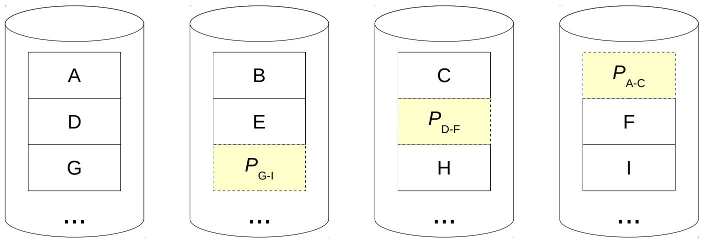
$$\{\underline{\text{TicketID}}, \text{Festival}\} \text{ und } \{\underline{\text{Festival, Band}}\}$$

erhält man nur dann, wenn auch die MVD Festival$\to\!\!\to$Band berücksichtigt wird.

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

Hausaufgabe 4

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.



a) Um welches RAID-Level handelt es sich?

b) Wieviele Festplatten können ausfallen, ohne dass mit Datenverlust zu rechnen ist? Geben Sie eine allgemeine Lösung für einen Verbund bestehend aus n Festplatten an.

c) 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 = _ _ _ _	$P_{\text{A-C}}$ = 1110
D = 0101	E = 1100	$P_{\text{D-F}}$ = _ _ _ _	F = 1100
G = 0011	$P_{\text{G-I}}$ = 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**	$P_{\text{A-C}}$ = 1110
D = 0101	E = 1100	$P_{\text{D-F}}$ = **0101**	F = 1100
G = 0011	$P_{\text{G-I}}$ = 1110	H = **1110**	I = 0011

Der Lehrstuhl für Datenbanksysteme wünscht frohe Weihnachten und schöne Feiertage!