

# Einführung in Datenbanken

---

## Übung 15: Lösung der Aufgaben zu Normalformen (BCNF)

Prof. Dr. Stefan Brass

PD Dr. Alexander Hinneburg

Martin-Luther-Universität Halle-Wittenberg

Wintersemester 2020/21

<http://www.informatik.uni-halle.de/~brass/db20/>

# Inhalt

1 Präsenzaufgabe 14

2 Hausaufgabe 14



# Lösung zu Präsenzaufgabe 14 (1)

- Die funktionale Abhängigkeit  $A \longrightarrow B$  kann bei dieser Tabelle gar nicht verletzt werden:

R		
A	B	C
1	2	3
4	x	y

- Man würde zwei Zeilen mit dem gleichen Wert für  $A$ , aber verschiedenem Wert für  $B$  benötigen.
- Mit den vorgegebenen Werten hat man aber schon verschiedene Werte für  $A$ .

# Lösung zu Präsenzaufgabe 14 (2)

- Daher kann man sich auf die verlangte Verletzung von  $B \rightarrow C$  konzentrieren:

R		
A	B	C
1	2	3
4	x	y

- x muss zwingend den Wert 2 haben.
- y kann beliebige gewählt werden, nur nicht 3.

R		
A	B	C
1	2	3
4	2	5

# Lösung zu Präsenzaufgabe 14 (3)

- Bonusaufgabe: Geben Sie eine dritte Zeile ( $a, b, c$ ) an, so dass  $A \rightarrow B$  weiter gilt, aber  $A \rightarrow C$  verletzt ist.

R		
A	B	C
1	2	3
4	2	5
4	2	6

- Man braucht jetzt einen doppelten Wert für  $A$ , im Beispiel wurde  $4$  gewählt ( $1$  würde natürlich auch gehen).
- Damit  $A \rightarrow B$  weiter gilt, muss  $B$  den gleichen Wert wie in der anderen Zeile mit dem gleichen  $A$ -Wert haben.
- Der Wert für  $C$  muss dagegen verschieden sein.



# Hausaufgabe 14 (1)

- Gegeben sei folgende Tabelle mit einigen Daten einer Schule:

SCHULE				
SCHUELER	KLASSE	FACH	LEHRER	RAUM
Karin Koch	7a	Deutsch	Herr Goetz	305
Karin Koch	7a	Englisch	Frau Rowling	305
Karin Koch	7a	Mathematik	Frau Noether	305
Frank Fischer	7a	Deutsch	Herr Goetz	305
Frank Fischer	7a	Englisch	Frau Rowling	305
Frank Fischer	7a	Mathematik	Frau Noether	305
Ronald Richter	10b	Mathematik	Frau Noether	210
Ronald Richter	10b	Englisch	Herr Pratchett	210
Ronald Richter	10b	Pyrotechnik	Herr Shimizu	210
Ronald Richter	10b	Religion	Herr Lewis	210

Die letzte Spalte heißt eigentlich „Klassenraum“. Das ist wichtig, weil die funktionale Abhängigkeit „Klasse bestimmt Klassenraum“ vorausgesetzt wird.



# Hausaufgabe 14 (2)

- Es seien folgende funktionale Abhängigkeiten gegeben:
  - SCHUELER  $\longrightarrow$  KLASSE
  - KLASSE, FACH  $\longrightarrow$  LEHRER
  - KLASSE  $\longrightarrow$  KLASSENRAUM
  - LEHRER, KLASSENRAUM  $\rightarrow$  LEHRER
  - SCHUELER, FACH  $\rightarrow$  KLASSENRAUM

- Z.B. bedeutet die erste funktionale Abhängigkeit (FA)

SCHUELER  $\longrightarrow$  KLASSE:

„Wenn zwei Zeilen den gleichen Wert in der Spalte SCHUELER haben, dann müssen sie auch den gleichen Wert in der Spalte KLASSE haben.“

Jeder Schüler gehört zu einer eindeutig bestimmten Klasse, diese ist nicht vom Unterrichtsfach abhängig.

# Hausaufgabe 14a: Weitere FA (1)

- Geben Sie eine weitere funktionale Abhängigkeit an,
  - die nicht von den explizit gegebenen Abhängigkeiten impliziert wird,
  - aber im Beispielzustand der Tabelle gilt.

Es ist nicht verlangt, dass die funktionale Abhängigkeit im realen Leben realistisch ist.

- Mögliche Lösung:

KLASSENRAUM  $\longrightarrow$  KLASSE

Dies wurde bei den funktionalen Abhängigkeiten vergessen. Es macht sogar im realen Leben Sinn: Ein Raum kann nur für eine Klasse Klassenraum sein (Forts. siehe nächste Folie).

# Hausaufgabe 14a: Weitere FA (2)

- Um zu prüfen, dass

$$\text{KLASSENRAUM} \longrightarrow \text{KLASSE}$$

im Beispiel-Zustand gilt, muss man nachschauen, dass in Gruppen von Zeilen mit gleichem **KLASSENRAUM** auch überall die gleiche **KLASSE** steht.

Man könnte eine SQL-Anfrage dafür schreiben.

- Zum Beweis, dass diese FA nicht von den gegebenen FAen impliziert wird, kann man die Attributhülle

$$\{\text{KLASSENRAUM}\}^+ = \{\text{KLASSENRAUM}\}$$

berechnen (bezüglich der gegebenen FAen).

- Da sie nicht **KLASSE** enthält, wird **KLASSENRAUM**  $\longrightarrow$  **KLASSE** nicht impliziert.

# Hausaufgabe 14a: Weitere FA (3)

- Es gibt noch weitere Lösungen, z.B.

LEHRER  $\longrightarrow$  FACH

- Im Beispiel-Zustand unterrichtet jeder Lehrer nur ein Fach.
- Dies wäre in der Realität aber eher nicht so.

Das kann natürlich von der Schule abhängen. Man muss im Rahmen des Datenbank-Entwurfs welche funktionalen Abhängigkeiten vorausgesetzt werden können.

- Man könnte diese beiden funktionalen Abhängigkeiten auch weiter abschwächen, z.B. wäre auch eine Lösung:

LEHRER, KLASSENRAUM  $\rightarrow$  FACH

# Hausaufgabe 14b: Attributhülle (1)

- Berechnen Sie die Attributhülle  $\{\text{SCHUELER}\}^+$ .
- Lösung: Man startet mit dem gegebenen Attribut:

$$\{\text{SCHUELER}\}^+ = \{\text{SCHUELER}, \dots\}.$$

- Aufgrund der funktionalen Abhängigkeit

$$\text{SCHUELER} \longrightarrow \text{KLASSE}$$

kann man KLASSE hinzufügen:

$$\{\text{SCHUELER}\}^+ = \{\text{SCHUELER}, \text{KLASSE}, \dots\}.$$

Wenn die linke Seite einer funktionalen Abhängigkeit bereits in der Attributhülle ist, kann man auch die rechte Seite hinzufügen.

- Fortsetzung siehe nächste Folie.

# Hausaufgabe 14b: Attributhülle (2)

- Da man nun KLASSE hat, kann man die funktionale Abhängigkeit

KLASSE  $\longrightarrow$  KLASSENRAUM

anwenden:

$\{\text{SCHUELER}\}^+ = \{\text{SCHUELER}, \text{KLASSE}, \text{KLASSENRAUM}\}.$

- Weitere funktionale Abhängigkeiten können nicht angewendet werden:
  - KLASSE, FACH  $\longrightarrow$  LEHRER
  - LEHRER, KLASSENRAUM  $\rightarrow$  LEHRER
  - SCHUELER, FACH  $\rightarrow$  KLASSENRAUM



# Hausaufgabe 14d: Schlüssel (1)

- Geben Sie alle minimalen Schlüssel der Tabelle an. Begründen Sie Ihre Aussage, d.h.
  - dass es jeweils ein Schlüssel ist,
  - dass dieser bzw. diese Schlüssel minimal sind, und
  - dass es keine weiteren Schlüssel gibt.
- Lösung: Wir müssen  $\subseteq$ -minimale Mengen  $\mathcal{A}$  von Attributen finden, so dass  $\mathcal{A}^+$  die Menge aller Attribute der Tabelle ist.

Ein Schlüssel bestimmt alle Attribute der Tabelle funktional.

Die Attributhülle  $\mathcal{A}^+$  ist die Menge alle von  $\mathcal{A}$  funktional bestimmten Attribute.



# Hausaufgabe 14d: Schlüssel (2)

- Die gegebenen funktionale Abhängigkeiten sind:
  - SCHUELER  $\longrightarrow$  KLASSE
  - KLASSE, FACH  $\longrightarrow$  LEHRER
  - KLASSE  $\longrightarrow$  KLASSENRAUM
  - LEHRER, KLASSENRAUM  $\rightarrow$  LEHRER
  - SCHUELER, FACH  $\rightarrow$  KLASSENRAUM
- Die Attribute SCHUELER und FACH kommen auf keiner rechten Seite vor.
- Wenn sie nicht von Anfang an in  $\mathcal{A}$  enthalten sind, können sie nicht in  $\mathcal{A}^+$  enthalten sein.

# Hausaufgabe 14d: Schlüssel (3)

- Die beiden Attribute **SCHUELER** und **FACH** müssen also in jedem Schlüssel enthalten sein.
- Wenn man nun

$$\{\text{SCHUELER, FACH}\}^+$$

berechnet, stellt man fest, dass es alle Attribute sind:

- **SCHUELER**  $\longrightarrow$  **KLASSE**
  - **KLASSE, FACH**  $\longrightarrow$  **LEHRER**
  - **KLASSE**  $\longrightarrow$  **KLASSENRAUM**
- Damit haben wir Glück: Es gibt nur einen minimalen Schlüssel, nämlich diese beiden Attribute.

# Hausaufgabe 14e: BCNF prüfen (1)

- Ist die Tabelle bezüglich der fünf gegebenen funktionalen Abhängigkeiten in BCNF? Begründen Sie Ihre Aussage.
- Lösung: Nein, die drei ersten funktionalen Abhängigkeiten verletzen BCNF, da die linke Seite kein Schlüssel (nicht notwendig minimal) ist, und sie nicht trivial sind:
  - SCHUELER  $\rightarrow$  KLASSE
  - KLASSE, FACH  $\rightarrow$  LEHRER
  - KLASSE  $\rightarrow$  KLASSENRAUM
  - LEHRER, KLASSENRAUM  $\rightarrow$  LEHRER
  - SCHUELER, FACH  $\rightarrow$  KLASSENRAUM

# Hausaufgabe 14e: BCNF prüfen (2)

- Die vierte funktionale Abhängigkeit verletzt BCNF nicht, da sie trivial ist:

LEHRER, KLASSENRAUM  $\rightarrow$  LEHRER

Die rechte Seite (LEHRER) kommt auch links vor.

- Die fünfte funktionale Abhängigkeit verletzt BCNF nicht, da links ein vollständiger Schlüssel steht:

SCHUELER, FACH  $\rightarrow$  KLASSENRAUM

Es könnten links auch noch weitere Attribute stehen, das wäre kein Problem. Man kann die BCNF-Bedingung auch ohne Vorab-Berechnung der Schlüssel testen, indem man die Attributhülle der linken Seiten berechnet. Dies muss jeweils die Menge aller Attribute sein (es sei denn, die FA wäre trivial, dann kann man sie sowieso streichen).

# Hausaufgabe 14e: BCNF herstellen (1)

- Zweiter Teil der Aufgabe: Falls die Tabelle nicht in BCNF sein sollte, überführen Sie die Tabelle durch verlustlose Aufspaltung in mehrere Tabellen, die in BCNF sind.

Alternativ dürfen Sie auch den 3NF Synthesealgorithmus verwenden.

- Verletzt  $\alpha \rightarrow \beta \in \mathcal{F}$  die BCNF, erzeugt man

- $R_1$  mit den Attributen  $\alpha^+$ .

Dies enthält natürlich die Attribute  $\alpha \cup \beta$  (wie in der oben gezeigten Dekomposition), könnte aber noch weitere Attribute enthalten (aber nur, wenn noch mehr FAen BCNF verletzen).

- $R_2$  mit den Attributen  $(\mathcal{A} \setminus \alpha^+) \cup (\alpha \setminus \beta)$ .

# Hausaufgabe 14e: BCNF herstellen (2)

- Wenn nur eine funktionale Abhängigkeit  $\alpha \rightarrow \beta$  BCNF verletzt, erzeugt man einfach eine Relation mit den Attributen  $\alpha \cup \beta$  (wobei  $\alpha$  Schlüssel ist), und entfernt die rechte Seite  $\beta$  aus der ursprünglichen Relation.

Voraussetzung dabei ist, dass  $\alpha \cap \beta = \emptyset$ . Ansonsten entfernt man den „trivialen Anteil“  $\alpha \cap \beta$  einfach von der rechten Seite.

- Im Beispiel gibt es leider mehrere Verletzungen.

Z.B. wäre das Ergebnis schlecht, wenn man zuerst die Verletzung durch **SCHUELER**  $\rightarrow$  **KLASSE** angeht: Dann würde man das Attribut **KLASSE** aus der ursprünglichen Tabelle entfernen (nach obigem Algorithmus immerhin zusammen mit **KLASSENRAUM**). Anschließend könnte man die funktionale Abhängigkeit **KLASSE**, **FACH**  $\rightarrow$  **LEHRER** nicht mehr ausdrücken. Man bekommt durch sukzessive Aufspaltung immer Tabellen in BCNF, und die Aufspaltung ist verlustlos, nicht immer das gefühlte beste Ergebnis.

# Hausaufgabe 14e: BCNF herstellen (3)

- Man sollte bei mehreren Verletzungen die FAen möglichst so anordnen, dass Attribute auf der rechten Seite (die entfernt werden) nicht in einer späteren FA links vorkommen.
- Starten wir also mit:

KLASSE, FACH  $\rightarrow$  LEHRER

Man könnte genauso gut mit  $\text{KLASSE} \rightarrow \text{KLASSENRAUM}$  starten, dagegen wäre  $\text{SCHUELER} \rightarrow \text{KLASSE}$  ungünstig.

- Man legt eine neue Tabelle mit den Attributen der FA an:  
LEHRER\_EINTEILUNG(KLASSE, FACH, LEHRER)
- Die rechte Seite wird aus der ursprünglichen Tabelle entfernt:  
SCHULE1(SCHUELER, KLASSE, FACH, KLASSENRAUM)

Dabei bilden KLASSE und FACH einen Fremdschlüssel zu LEHRER\_EINTEILUNG.

# Hausaufgabe 14e: BCNF herstellen (4)

- Nun verletzt **SCHULE1** noch immer BCNF, z.B. wegen  
**KLASSE  $\rightarrow$  KLASSENRAUM**.

Auch wegen **SCHUELER  $\rightarrow$  KLASSE**, aber das heben wir uns noch auf.

- Also legt man eine neue Tabelle mit den Attributen der FA an:  
**KLASSEN(KLASSE, KLASSENRAUM)**.

Die linke Seite der FA wird immer Schlüssel. Sie bestimmt nach der Konstruktion ja die übrigen Attribute.

- Wieder entfernen wir die rechte Seite der FA aus der ursprünglichen Tabelle:

**SCHULE2(SCHUELER, KLASSE  $\rightarrow$  KLASSEN, FACH)**



# Hausaufgabe 14e: BCNF herstellen (5)

- Auch **SCHULE2** verletzt noch BCNF wegen  
**SCHUELER**  $\rightarrow$  **KLASSE**,
- Also machen wir auch dazu eine Tabelle:  
**KLASSEN\_EINTEILUNG**(**SCHUELER**, **KLASSE** $\rightarrow$ **KLASSEN**).
- Die rechte Seite wird aus der ursprünglichen Tabelle entfernt, übrig bleibt:

**LERNT**(**SCHUELER** $\rightarrow$ **KLASSEN\_EINTEILUNG**, **FACH**)

Leider wird dabei der ursprüngliche Fremdschlüssel zerstört: Bei **SCHULE1** wurde festgelegt, dass **KLASSE** und **FACH** zusammen einen Fremdschlüssel bilden, der auf **LEHRER\_EINTEILUNG** verweist. Jetzt werden die beiden Attribute aber auseinander gerissen.

# Hausaufgabe 14e: BCNF herstellen (6)

- Das Endergebnis ist also:
  - $\text{LEHRER\_EINTEILUNG}(\underline{\text{KLASSE}}, \underline{\text{FACH}}, \text{LEHRER})$
  - $\text{KLASSEN}(\underline{\text{KLASSE}}, \text{KLASSENRAUM})$ .
  - $\text{KLASSEN\_EINTEILUNG}(\underline{\text{SCHUELER}}, \text{KLASSE} \rightarrow \text{KLASSEN})$ .
  - $\text{LERNT}(\underline{\text{SCHUELER}} \rightarrow \text{KLASSEN\_EINTEILUNG}, \underline{\text{FACH}})$

Da es für diese Tabelle keine FAen mehr gibt, sind beide Attribute zusammen der Schlüssel.

- Im Prinzip hat man aus jeder der interessanten FAen eine Tabelle gemacht, und  $\text{LERNT}$  ist der Rest.

Der 3NF-Syntheselgorithmus macht im Prinzip genau das. Im Skript ist bei diesem Algorithmus erklärt, wie man die gegebenen FAen aufbereiten muss, damit das funktioniert. Im Beispiel würden bei der Berechnung der „kanonischen Überdeckung“ die vierte und fünfte FA eliminiert.

# Hausaufgabe 14e: Exkurs zu 4NF (1)

## Nicht prüfungsrelevant:

- Das Beispiel ist leider komplizierter als beabsichtigt.
- Intuitiv ist die Tabelle

LERNT(SCHUELER, FACH)

für diese Anwendung falsch.

- Die Fächer hängen normalerweise nur von der Klasse ab, und nicht vom einzelnen Schüler. Dann würde man eher diese Tabellen wollen:
  - KLASSEN(KLASSE, KLASSENRAUM)
  - FAECHER(KLASSE→KLASSEN, FACH, LEHRER)
  - KLASSEN\_EINTEILUNG(SCHUELER, KLASSE→KLASSEN)

# Hausaufgabe 14e: Exkurs zu 4NF (2)

- Die gegebenen funktionalen Abhängigkeiten sagen aber nicht aus, dass die Fächer nur von der Klasse abhängen.

Das kann man beweisen, indem man einen DB-Zustand macht, der alle FAen erfüllt, aber zwei Schüler in der gleichen Klasse enthält, die verschiedene Fächer haben.

- Es kommt ja auch vor, dass Klassen bei bestimmten Fächern aufgespalten werden:
  - Ein Teil der Schüler hat Latein, der andere Teil Französisch.
  - Oder evangelische Religion, katholische Religion, Ethik.
- Wenn die Fächer tatsächlich vom Schüler abhängen und nicht von der Klasse, war der obige Entwurf doch richtig.

# Hausaufgabe 14e: Exkurs zu 4NF (3)

- Will man ausdrücken, dass zwei Schüler in der gleichen Klasse immer die gleiche Menge von Fächern haben, braucht man dazu „mehrwertige Abhängigkeiten“:
  - **KLASSE**  $\longrightarrow$  **SCHUELER**
  - **KLASSE**  $\longrightarrow$  **FACH, LEHRER**
- Diese Art von Integritätsbedingungen bedeutet: „Wenn zwei Zeilen in der linken Seite übereinstimmen, kann man die Werte der rechten Seite vertauschen, und bekommt zwei Zeilen, die auch in der Tabelle enthalten sein müssen.“

Z.B. kann man für zwei Zeilen mit der gleichen **KLASSE** die **SCHUELER** vertauschen: Die so entstandenen „neuen“ Zeilen müssen bereits in der Tabelle enthalten sein.

# Hausaufgabe 14e: Exkurs zu 4NF (4)

- Normalisierung mit mehrwertigen Abhängigkeiten führt zu 4NF.
- Die Reihenfolge der Normalformen ist: 2NF, 3NF, BCNF, 4NF, 5NF.
  - Normalformen später in der Liste implizieren Normalformen früher in der Liste.
  - Z.B. ist eine Tabelle in BCNF automatisch in 3NF.
  - Von diesen Normalformen basieren 2NF, 3NF und BCNF auf funktionalen Abhängigkeiten.

Wenn man mehr will, braucht man allgemeinere Integritätsbedingungen.

- Dies alles wird ausführlich in der Vorlesung „Datenbank-Entwurf“ im Master behandelt.