Kapitel 8

Formaler Datenbankentwurf

Inhalt

8.1 Zerlegung eines Relationenschemas

Eine **Zerlegung** (decomposition) eines Relationenschemas $R = (S, F)$ ist eine Menge von Relationenschemata $\{R_1 = (S_1, F_1), R_2 = (S_2, F_2), ..., R_n = (S_n, F_n)\}$, wobei:

(1) $S = S_1 \cup S_2 \cup ... \cup S_n$

(2) $F_i \equiv \{X \rightarrow Y \mid X \rightarrow Y \in F^+ , XY \subseteq S_i\}$, d.h. F_i^+ ist die Projektion der Abhängigkeitsmenge F^+ auf S_i ($F_i^+ = F^+ [S_i]$)

(3) $(F_1 \cup F_2 \cup ... \cup F_3)^+ \subseteq F^+$

Die Zerlegung heißt **abhängigkeitstreu** (dependency preserving), wenn jede Abhängigkeit F_i aus der ursprünglichen Tabelle erhalten bleibt

$(F_1 \cup F_2 \cup ... \cup F_n)^+ = F^+$

Die Zerlegung heißt **verbundtreu** (verlustfrei, lossless), wenn sich genau alle Tupel der ursprünglichen Relationen durch Join wiederherstellen lassen. Für die Relation r (S) gilt also:

$$r (S) = r(S_1) \bowtie r(S_2) \bowtie ... \bowtie r (S_n)$$

Beispiel: nicht abhängigkeitstreue Zerlegung

Sei R = ({A B C D E}, {A → BCD, CD → E, EC → B}) ein Relationenschema

Eine mögliche Zerlegung von R ist $\{R_1, R_2\}$ mit

R1 = ({A B C D}, {A → BCD, CD → B)

R2 = ({C D E}, {CD → E})

Die Abhängigkeit EC → B wird nicht durch die entweder auf R_1 oder R_2 anwendbaren FD aus F impliziert

Die Zerlegung ist also nicht abhängigkeitstreu

Beispiel: verbundtreue Zerlegung

$R = (\{A\ B\ C\}, \{A \rightarrow B, B \rightarrow C\})$

wird zerlegt in

$Q = (\{AB\}, \{A \rightarrow B\})$ und

$P = (\{BC\}, \{B \rightarrow C\})$

r

A	*B*	*C*
a1	b1	c1
a2	b2	c1

q

A	*B*
a1	b1
a2	b2

$\bowtie$ p

B	*C*
b1	c1
b2	c1

= r'

A	*B*	*C*
a1	b1	c1
a2	b2	c1

Beispiel: nicht verbundtreue Zerlegung

$R = (\{A\ B\ C\}, \{A \rightarrow B, C \rightarrow B\})$

$Q = (\{A\ B\}, \{A \rightarrow B\}), P = (\{B\ C\}, \{C \rightarrow B\})$

r

A	B	C
a1	b1	c1
a2	b1	c2

q

A	B
a1	b1
a2	b1

$\bowtie$ p

B	C
b1	c1
b1	c2

= r'

A	B	C
a1	b1	c1
a1	b1	c2
a2	b1	c1
a2	b1	c2

Theoreme für Zerlegung

Satz: Eine **Zerlegung** von

$R = (S, F)$ in $\{R_1 = (S_1, F_1), R_2 = (S_2, F_2), \ldots R_n = (S_n, F_n)\}$

ist **verbundtreu**, wenn bzgl.

$F' = F_1 \cup F_2 \cup \ldots \cup F_n$,

die Hülle eines der generierten Relationenschemata S_i^+ gleich S ist

Satz: Jedes Relationenschema kann **verlustfrei und abhängigkeitstreu** in ein 2NF und 3NF Schema **zerlegt** werden.

Zu einem Schema kann es mehrere verlustfreie und abhängigkeitstreue Zerlegungen geben. Es ist darauf zu achten, eine semantisch gute Zerlegung zu finden!

Verbundtreue Zerlegung in 3NF

gegeben: Relationenschema R = (S, F)

gesucht: verbundtreue Zerlegung

Algorithmus verbundtreue Zerlegung in 3NF

Suche FD $Y \rightarrow A$, die 3NF verletzt // d.h. $Y \neg$ Oberschlüssel, $A \neg$ prim)

zerlege in

S1 = S - A

S2 = YA

wende Algorithmus auf S1 und S2 an.

Beispiel

R = ({A B C D E} {AB → C, D → E, C → DE}

Schlüsselkandidaten: {AB}, C → DE verletzt 3NF

Zerlegung in: R_1 = ({A B C} {AB → C}) und R_2 = ({C D E} {C → DE, D → E})

3NF-Synthesealgorithmus

Gegeben: Relationsschema R = (S, F)

Gesucht: verbund- und abhängigkeitstreue Zerlegung

1) Finde eine kanonische Überdeckung F_C für F
   - 1a) Linksreduktion
   - 1b) Rechtsreduktion
   - 1c) Entfernen „sinnloser“ Abhängigkeiten
   - 1d) Zusammenfassen gleicher linker Seiten
2) Erstelle für jede funktionale Abhängigkeit $\alpha \rightarrow \beta \in F_C$ ein Relationenschema $R_\alpha = (S_\alpha, F_\alpha)$, mit $S_\alpha = \alpha \cup \beta$ und $F_\alpha = \{\alpha' \rightarrow \beta' \in F_C | \alpha' \cup \beta' \subseteq S_\alpha\}$
3) Wenn kein Relationenschema R_α einen Schlüsselkandidaten für R enthält, ergänzen mit R_0 = (K, ∅), wobei K ein Schlüsselkandidat für R = (S, F) ist.
4) Entferne Relationenschemata R_α, falls $S_\alpha \subseteq S_{\alpha'}$ für irgendein anderes $R_{\alpha'}$ der erzeugten Relationenschmata. (R_α ist in $R_{\alpha'}$ vollständig enthalten.)

3NF-Synthesealgorithmus Beispiel

R=(S,F)

S={ABCDEFG}

F={AC→F,A→BE,B→C,DA→G,D→AE}

1a) A→F, A→BE, B→C, D→G, D→AE

1b) A→F, A→BE, B→C, D→G, D→A

1c) keine Abhängigkeiten der Form X→Ø

1d) A→BEF, B→C, D→AG

2) R1 = ({<u>A</u>BEF},{A→BEF})
R2 = ({<u>B</u>C},{B→C})
R3 = ({<u>D</u>GA},{D→AG})

3) Schlüsselkandidat in ursprünglicher Relation war D, D ist in R3 enthalten, also nichts hinzuzufügen

4) Kein Relationenschema enthält ein anderes vollständig. Es wird daher kein Relationenschema entfernt.

8.3 BOYCE-CODD Normalform

Ein Relationenschema R = (S, F) ist in **Boyce-Codd Normalform** (BCNF), wenn für alle FD $X \rightarrow A$ gilt: $X \rightarrow S$ oder $A \in X$.

D.h. das Relationenschema ist in 3NF und es darf keine Abhängigkeiten zwischen Schlüsselkandidaten geben

Es gilt folgender Satz:

BCNF $\Rightarrow$ 3NF

Die BCNF ist strikter als 3NF für Relationen mit mehreren, zusammengesetzten, nichtdisjunkten Schlüsselkandidaten.

Beispiel BCNF

R = ({Matr#, LVLeiter, LV}, {Matr# LV → LVLeiter, LV-Leiter → LV})

(Matr#, LV) ist Schlüssel, also verletzt die Abhängigkeit LVLeiter → LV die BCNF (aber nicht 3NF).

Zerlegung in:

R1 = ({Matr#, LVLeiter}, { })

R2 = ({LVLeiter, LV}, {LVLeiter → LV})

Diese Zerlegung in BCNF ist verbundtreu, aber nicht abhängigkeitstreu.

Bei Zerlegung in BCNF ist es nicht immer möglich, eine verbund- und abhängigkeitstreue Dekomposition zu finden. In einem solchen Fall ist auf Verbundtreue zu achten, auf Abhängigkeitstreue muss verzichtet werden.

Dekompositionsalgorithmus
Verbundtreue Zerlegung in BCNF

Gegeben: Relationenschema R = (S, F)

1) F_C := kanonische Ueberdeckung von F
2) Suche eine "bösartige" Abhängigkeit $X \rightarrow Y$, wobei X keinen Schlüsselkandidaten enthält, und Y nicht in X vorkommt.

   Existiert keine solche Abhängigkeit, ist das Schema in BCNF.

   Gibt es eine solche Abhängigkeit, dann spalte R in 2 Schemata

   R1 = (S - Y, F1),
   R2 = (XY, F2)

   wobei F1 und F2 die Projektion von F^+ auf S-Y bzw. XY ist

   $F1^+ = F^+[S – Y]$
   $F2^+ = F^+[XY]$
3) Wende das Verfahren rekursiv auf R1 und R2 an.

Beispiel

Schema U = ({K, L, D, R, S, N}, F)
Die Attribute haben folgende Bedeutung: K: Kurs, L: Lektor, D: Dauer, R: Raum, S: Student, N: Note
F enthält folgende funktionale Abhängigkeiten K → L, DR → K, DL → R, KS → N, DS → R
Wie leicht zu sehen ist, ist DS der einzige Schlüsselkandidat.

Zerlegung: U = ({K, L, D, R, S, N}, {K → L, DR → K, DL → R, KS → N, DS → R})

bösartige Abhängigkeit: KS → N

U1 = ({K, L, D, R, S}, {K → L, DR → K, DL → R, DS → R})	Schlüssel: DS
U2 = ({K, S, N}, {KS → N}	Schlüssel: KS, ist in BCNF

Zerlegung U1: K → L

U1.1 = ({K, D, R, S}, {KD → R, DR → K, DS → R})	Schlüssel: DS
U1.2 = ({K, L}, {K → L})	Schlüssel: K, ist in BCNF

Zerlegung U1.1 KD → R

U1.1.1 = ({K, D, S}, {DS → K})	Schlüssel: DS
U1.1.2 = ({K, D, R}, {KD → R, DR → K})	Schlüsselkandidaten: KD, DR

beide Schemata sind in BCNF.

Die Zerlegung ist also: (U2, U1.2, U1.1.1, U1.1.2)

8.4 Mehrwertige Abhängigkeiten und 4. Normalform

Sei R ein Relationenschema, X, Y ⊆ R, Z = R - XY

Eine Relation r (R) erfüllt die **mehrwertige Abhängigkeit** (multivalued dependency, MVD), **X ↠ Y**, wenn für zwei beliebige Tupel t1 und t2 in r mit t1(X) = t2(X) zwei Tupel t3 und t4 existieren mit
t3(X) = t1(X) und t3(Z) = t1(Z) und t3(Y) = t2(Y)
t4(X) = t1(X) und t4(Z) = t2(Z) und t4(Y) = t1(Y)

Beispiel MVD, X ↠ Y

Aus

X	Y	Z
X	Y1	Z1
X	Y2	Z2

folgt

X	Y	Z
X	Y1	Z1
X	Y1	Z2
X	Y2	Z1
X	Y2	Z2

Intuitiv: Jeder X-Wert x bestimmt eine Menge von Y-Werten, die für jede x, z-Kombination immer vollständig auftreten müssen.

Mehrwertige Abhängigkeiten

Autos

Autotyp	Motor	Farbe
BMW 318i	Benzin	Schwarz
BMW 318i	Benzin	Blau
BMW 318i	Benzin	Rot
BMW 318i	Diesel	Schwarz
BMW 318i	Diesel	Blau
BMW 318i	Diesel	Rot

Motor und Farbe sind unabhängig!

Mehrwertige Abhängigkeiten treten immer paarweise auf:

$X \twoheadrightarrow Y, Z = (R - XY) \quad \Rightarrow \quad X \twoheadrightarrow Z$

Im Beispiel: Autotyp ↠ Motor, Autotyp ↠ Farbe

Schreibweise: $X \twoheadrightarrow Y \mid Z$

d.h. Autotyp ↠ Motor | Farbe

Eine mehrwertige Abhängigkeit $X \twoheadrightarrow Y$ ist **trivial**, wenn entweder $Y \subseteq X$ oder $XY = R$ ist

Beispiel MVD Anomalien

Mitarbeiter-Relation, Annahme: Fähigkeiten und Sprachkenntnisse sind unabhängig,

Angestellter ↠ Fähigkeit|Sprache

Mitarbeiter

MName	Fähigkeit	Sprache
Huber	Kochen	Deutsch
Huber	Kochen	Englisch
Huber	Kochen	Französisch
Huber	Steno	Deutsch
Huber	Steno	Englisch
Huber	Steno	Französisch

Anomalien

- Insertion Anomalie: Wenn eingefügt werden soll, dass "Huber" auch "Maschinschreiben" kann, so müssen in diesem Fall 3 Tupel eingefügt werden.
- Update Anomalie: Wenn wir für "Huber" die Sprache "Englisch" auf "Amerikanisch" ändern wollen, so müssen mehrere Tupel geändert werden.
- Deletion Anomalie: Wenn wir die Fähigkeit "Kochen" löschen wollen, müssen wir mehrere Tupel ändern. Wenn wir dann noch "Steno" löschen, geht auch verloren, welche Sprachen Huber beherrscht.

Lösung: Zerlegung in 2 Relationen

Fähigkeiten

MName	Fähigkeit
Huber	Kochen
Huber	Steno

Sprache

MName	Sprache
Huber	Deutsch
Huber	Englisch
Huber	Französisch

Anomalien nicht mehr vorhanden

Axiomensystem für mehrwertige und funktionale Abhängigkeiten

(M1) **Reflexivität** (reflexivity) $Y \subseteq X \quad \vdash X \rightarrow Y$

(M2) **Erweiterung** (augmentation) $X \rightarrow Y, Z \subseteq W \vdash XW \rightarrow YZ$

(M3) **Transitivität** (transitivity) $X \rightarrow Y, Y \rightarrow Z \vdash X \rightarrow Z$

(M4) **Komplementregel** (complementation)

$X \twoheadrightarrow Y \vdash X \twoheadrightarrow R\text{-}XY$

(M5) **Mehrwertige Erweiterung** (multivalued augmentation)

$X \twoheadrightarrow Y, Z \subseteq W \vdash XW \twoheadrightarrow YZ$

(M6) **Mehrwertige Transitivität** (multivalued transitivity)

$X \twoheadrightarrow Y, Y \twoheadrightarrow Z \vdash X \twoheadrightarrow Z\text{-}Y$

(M7) **Replikation** (replication)

$X \rightarrow Y \vdash X \twoheadrightarrow Y$

(M8) **Koaleszenz** (coalescence)

$X \twoheadrightarrow Y, Z \subseteq Y, \exists W \; W \cap Y = \emptyset, W \rightarrow Z \quad \vdash \quad X \rightarrow Z$

Die Regeln sind korrekt und vollständig (sound and complete)

Weitere Regeln

(M9) **Mehrwertige Vereinigung** (multivalued union)

$X \twoheadrightarrow Y,\ X \twoheadrightarrow Z \mid\!- X \twoheadrightarrow YZ$

(M10) **Mehrwertiger Schnitt** (multivalued intersection)

$X \twoheadrightarrow Y,\ X \twoheadrightarrow Z \mid\!- X \twoheadrightarrow Y \cap Z$

(M11) **Mehrwertige Differenz** (multivalued difference)

$X \twoheadrightarrow Y,\ X \twoheadrightarrow Z \mid\!- X \twoheadrightarrow Z\text{-}Y,\ X \twoheadrightarrow Y\text{-}Z$

4. Normalform

Sei FM eine Menge funktionaler und mehrwertiger Abhängigkeiten.

Die **Hülle** $FM^+ = \{X \twoheadrightarrow Y \mid FM \vdash X \twoheadrightarrow Y\}$

Sei FM eine Menge funktionaler und mehrwertiger Abhängigkeiten über dem Relationenschema R.

R ist in **4. Normalform (4NF)**, wenn für jede nichttriviale mehrwertige Abhängigkeit $X \twoheadrightarrow Y \in FM^+$ gilt, dass $X \rightarrow R \in FM^+$ (d.h. X ist Oberschlüssel)

4 NF: $X \twoheadrightarrow Y \Rightarrow Y \subseteq X$ oder $XY = R$ oder $X \rightarrow R$

Theoreme zu MVD

Eine Relation r über R ist **verlustfrei zerlegbar** in R1 = XY und R2 = XZ, genau dann, wenn gilt: $X \twoheadrightarrow Y \mid Z$

Es gilt: **4 NF ⇒ BCNF ⇒ 3 NF ⇒ 2 NF ⇒ 1 NF**

Normalisieren: durch Aufspalten (siehe Beispiel)