

AG Datenbanken und Informationssysteme

Wintersemester 2006 / 2007

Prof. Dr.-Ing. Dr. h. c. Theo Härder
Fachbereich Informatik
Technische Universität Kaiserslautern



<http://www.dvs.informatik.uni-kl.de>

4. Übungsblatt

Für die Übung am Donnerstag, 23 November 2006,
von 15:30 bis 17:00 Uhr in 13/222.

Aufgabe 1: 4NF am Beispiel Kredite

Gegeben sei folgende Relation:

KREDITE (KONTONR, LFD_KREDITNR, DATUM, NAME, ADRESSE)

In dieser Relation sollen folgende Abhängigkeiten gelten:

- Ein Kunde kann mehrere Konten haben.
- Ein Kunde kann mehrere Adressen haben.
- Ein Konto kann mehrere Besitzer haben.
- Auf einem Konto können mehrere Kredite aufgenommen werden.
- Ein Kredit auf einem Konto läuft seit einem festgelegten Datum

Folgendes Beispiel erfüllt diese Bedingungen:

KONTONR	LFD_KREDITNR	DATUM	NAME	ADRESSE
100	1	01.01.2002	Müller	KL
100	1	01.01.2002	Maier	MA
100	1	01.01.2002	Maier	SB
100	2	01.03.2002	Müller	KL
100	2	01.03.2002	Maier	MA
100	2	01.03.2002	Maier	SB
110	1	15.11.2001	Müller	KL
110	2	01.02.2002	Müller	KL
120	1	01.06.2002	Maier	MA
120	1	01.06.2002	Maier	SB

- Welche Abhängigkeiten bestehen in dieser Relation?
- Wie lässt sich die Relation in eine 4NF zerlegen?

Aufgabe 2: Synthese von Relationen in 3NF

Die Beziehungen der Miniwelt, einem anwendungsbezogenen Ausschnitt der realen Welt, sollen im Hinblick auf das Informationsbedürfnis der verschiedenen Anwendungen einer Datenbank unabhängig von ihrer physischen Realisierung im konzeptionellen Schema (logisches Datenmodell) beschrieben und dargestellt werden. Im Relationenmodell sind Relationen in 3NF zur Vermeidung bestimmter Anomalien beim Änderungsdienst als logische Datenstrukturen besonders gut geeignet.

Ein Weg zur Ableitung dieser logischen Datenstrukturen ist die Relationensynthese. Ausgangspunkt dafür ist die Beschreibung der Miniwelt durch Attribute und FAen, die entsprechend dem Informationsbedarf der geplanten Anwendungen zu erheben sind. Attribute und FAen dienen als Erhebungsprimitive für den IST-Zustand. Alle möglichen FAen, die gegebenenfalls aus der erhobenen Menge abgeleitet sind, bilden eine transitiv abgeschlossene Menge, die sich als gerichteter Graph darstellen lässt. Für den Entwurf von Relationen interessiert eine nichtredundante Überdeckung der FAen der transitiven Hülle, die eine redundanzfreie logische Datenstruktur garantiert. Aus dieser minimalen Menge von FAen lassen sich schrittweise Relationen in optimaler 3NF erzeugen.

Für folgendes Anwendungsbeispiel ist die Relationensynthese für Relationen in optimaler 3NF durchzuführen:

Es seien in einer Miniwelt folgende Attribute, die die Entities Student (STUD), Fachbereich (FB), Vorlesung (VL), Professor (PR), Forschungsgruppe (FG) und Fachrichtung (FR) beschreiben, bekannt: Name (NAME), Matrikelnummer (MATRNR), Geburtsdatum (GEB), Adresse (ADR), Semesteranzahl (SEMANZ), Fachbereichsnummer (FBNR), Fachbereichsname (FBNAME), Dekan (DEKAN), Fachbereichsadresse (FBADR), Professor (PROF), Vorlesung (VORL), Fachrichtung (FACHR), Forschungsgruppe (FG) und Stundenanzahl (STDANZ).

Folgende FAen seien erhoben worden:

NAME, GEB, ADR	→	MATRNR, SEMANZ
MATRNR	→	NAME, GEB, ADR, SEMANZ
MATRNR	→	FACHR, FBNR
FACHR	→	FBNAME
FG	→	FBNAME, FACHR
PROF	→	FACHR, FG, FBNR
FBNAME	→	DEKAN
DEKAN	→	FBNR
FBNR	→	FBADR
FBADR	→	FBNAME
VORL	→	FBNR, STDANZ

- Stellen Sie die Menge der FAen durch einen Graph dar.
- Der erhaltene FA-Graph hat Zyklen. Da jede FA-Beziehung eines der am Zyklus beteiligten Attribute für jedes Attribut des Zyklus gilt, kann der Zyklus separiert und durch ein Ersatzattribut (Ersatzknoten) im FA-Graph ersetzt werden. Zeichnen Sie den reduzierten FA-Graph nach Elimination der Zyklen
- Für jeden Zyklus kann isoliert eine minimale Überdeckung gefunden werden. Finden Sie für die Zyklen geeignete minimale Überdeckungen.
- Finden Sie eine minimale Überdeckung für den reduzierten FA-Graph und ersetzen sie die Ersatzattribute wieder durch eine geeignete minimale Überdeckung des entsprechenden Zyklus. Damit haben sie eine minimale Überdeckung für das Anwendungsbeispiel gefunden.
- Bilden Sie aus den FAen der gefundenen minimalen Überdeckung Relationen in 3NF und diskutieren Sie das Ergebnis.

Aufgabe 3: Entwurf eines Data Warehouse

Für den aus dem folgenden Anwendungsszenario hervorgehenden Datenbestand eines Unternehmens soll ein Data Warehouse zur Analyse der Verkaufszahlen erstellt werden.

Das Unternehmen erfasst alle Kunden mit einer eindeutigen ID, deren Namen und Geburtsdatum. Jeder Kunde wird einer Altersgruppe (z. B. Teenager) zugeteilt. Die angebotenen Produkte werden mit einer Produkt-ID, Namen und einem Preis gespeichert. Jedes Produkt wird zusätzlich mit einer Produktgruppe klassifiziert. Das Unternehmen besitzt zahlreiche Filialen. Für jede Filiale ist der Ort gespeichert, der wiederum einem der Gebiete Süd, Ost, West oder Nord zugeordnet ist. Der Verkauf eines Produkts an einen Kunden in einer Filiale wird zusammen mit dem Datum in der Unternehmensdatenbank abgelegt.

- a) Erstellen Sie für das beschriebene Anwendungsszenario ein E/R-Diagramm.
- b) Bilden Sie das E/R-Diagramm mit SQL auf ein relationales Datenbankschema ab.
- c) Erstellen Sie mit SQL das Stern-Schema für ein Data Warehouse, in das die Daten aus der Datenbank von b) geladen und mit dem die in d) aufgeführten Fragen beantwortet werden können.
- d) Formulieren Sie die folgenden Anfragen in SQL auf dem Datenbankschema des Data Warehouse:
 - (1) Wieviele Produkte wurden im ersten Quartal verkauft?
 - (2) Wieviele Produkte wurden davon in der fünften Kalenderwoche verkauft?
 - (3) Wieviele Produkte haben Kunden der Altersgruppe 'Teenager' im Gebiet 'Süd' gekauft?
 - (4) Wie hoch ist das Durchschnittsalter aller Kunden der Altersgruppe 'Rentner', die im Gebiet 'Nord' Produkte der Gruppe 'Mobiltelefon' außerhalb der Weihnachtszeit gekauft haben?

Aufgabe 4: SQL-Anfragen und Views am Beispiel „Personal-DB“

Gegeben sei die folgende Datenbank, die von der Finanzabteilung zur Berechnung der Löhne und Gehälter der Mitarbeiter (MA) der verschiedenen Abteilungen (ABT) genutzt wird.

```
MA (MANR, MANAME, MAVORNAME, ABTNR, FIRMENZUGEHORIGKEIT,
    KINDER, STEUERKLASSE, GEHALT, KRANKENKASSE, BEITRAGSSATZ)
ABT (ABTNR, ABTNAME, ABTLEITER, ABTORT)
```

ABTLEITER hat denselben Wertebereich wie MANR und ist Fremdschlüssel. Zur Erstellung verschiedener Statistiken sollen dynamische Sichten erzeugt werden, und zwar:

- a) Eine Sicht, die die Abteilungsnummer, den Abteilungsnamen, die Anzahl der Mitarbeiter der Abteilung, den Durchschnitt der Firmenzugehörigkeit und des Gehalts, das höchste Gehalt der Abteilung und die Differenz zwischen dem höchsten und niedrigsten Gehalt in der Abteilung umfasst.
- b) Eine Sicht, die, gestaffelt nach Krankenkasse und Kinderzahl, den durchschnittlichen Beitragssatz für Mitglieder von Abteilungen in Frankfurt, München oder Stuttgart beinhaltet.
- c) Eine Sicht, die Name, Vorname und Gehalt der Mitarbeiter enthält, die in Abteilungen arbeiten, deren Durchschnittsgehalt größer als 50.000 ist.
- d) Eine Sicht, die die Daten der Mitarbeiter in Steuerklasse 1 enthält, und eine weitere, die nur Mitarbeiter in Steuerklasse 1 mit mehr als 5 Jahren Firmenzugehörigkeit enthält.
- e) Formulieren Sie auf der ersten der beiden letzten Sichten die Anfrage nach den Daten aller Mitarbeiter, deren Abteilungsleiter 'Müller' heißt.
- f) Was passiert bei Änderungen auf Sichten, die Aggregatfunktionen beinhalten?