

Übung zur Vorlesung „Datenbanksysteme für Hörer anderer Fachrichtungen“ Richard Kuntschke (richard.kuntschke@in.tum.de)

Lösungen zu Blatt 1

Aufgabe 1

Kontrollierte redundante Datenhaltung kann aus Performancegesichtspunkten sinnvoll sein. Als Beispiel betrachte man die Personenverwaltung einer Universität. So sind etwa Verwaltungsangestellte, Professoren, wissenschaftliche Mitarbeiter und Studenten Angehörige einer Universität, deren personenbezogene Daten in einer Datenbank abgelegt werden. Anfragen, die an diese Datenbasis gestellt werden, können sich auf alle Personen beziehen (z.B. Einladung zu einer Fakultätsfeier), auf bestimmte Personengruppen (z.B. Erstellen der Lohnabrechnung für alle Angestellten) oder auch selektiv nur auf Teilgruppen (z.B. Einladung zu einer Informationsveranstaltung für alle Studenten im 4. Semester). Die einzelnen Personengruppen sind aber unterschiedlich groß. So ist davon auszugehen, dass es deutlich mehr Studenten als wissenschaftliche Mitarbeiter und mehr wissenschaftliche Mitarbeiter als Professoren gibt. Häufig auftretende Anfragen, die sich nur auf Professoren beziehen, sind dann ineffizient, wenn alle Personendaten zentral, also in einem Datensatz, abgespeichert sind. In diesem Fall wäre das Überprüfen aller Universitätsangehörigen und das Auswählen der Professoren notwendig.

Speichert man in einem Datensatz all die Information, die sich auf alle Personen beziehen (Name, Anschrift, etc.) und in spezialisierten Datensätzen die Informationen für Studenten (Name, Anschrift, Semester, Studienrichtung), Professoren (Name, Anschrift, Rang, Raum) usw., so kann man beide Typen von Anfragen effizient ausführen. Es ist dann möglich, allgemeine Anfragen, die auf alle Universitätsangehörigen zutreffen, wie auch selektive Anfragen effizient auszuführen. Die redundante Datenhaltung, in diesem Fall also das doppelte Ablegen allgemeiner personenbezogener Daten, wie Name und Anschrift, muss vom Datenbanksystem kontrolliert werden. Um einen konsistenten, d.h. stimmigen Dateninhalt zu gewährleisten, müssen Änderungen des Datenbestands (Einfügen, Löschen oder Modifikation einzelner Datensätze) sowohl in den spezialisierten Datensätzen, wie auch in dem globalen Datensatz der Personendaten durchgeführt werden. Diese Strukturierung von Personengruppen spiegelt sich im Datenbankentwurf auch in Form einer Generalisierung wieder.

Ein weiterer Anwendungsfall redundanter Datenhaltung bietet sich in verteilten Informationssystemen. Anfragen werden hier verteilt in einem Netzwerk abgearbeitet. Höhere Performanz kann dadurch erreicht werden, häufig angefragte Daten, die zudem relativ statisch sind, d.h. selten modifiziert werden, an den Knoten im Netzwerk zu replizieren, an denen die Anfrageauswertung erfolgt. Je nach Anwendungsfall muss dann natürlich die Konsistenz, d.h. die Gültigkeit oder auch die "Frische" der replizierten Daten durch das verteilte Datenbanksystem überwacht werden.

Aufgabe 2

Datenhaltung

Im Rahmen einer Universitätsverwaltung müssen unter anderem folgende Daten abgelegt und verwaltet werden:

- Daten über zentrale Verwaltungseinheiten, Fakultäten und Lehrstühle
- Gebäudedaten, z.B. Raumgröße und Zuordnung zu Lehreinrichtungen
- Informationen über Veranstaltungen (Vorlesungen, Praktika oder auch Gastvorträge), wie Titel und Beschreibung der Veranstaltung, Ort, Uhrzeit und Dozent
- Für bestimmte Lehrveranstaltungen (etwa Praktika) muss auch die Teilnehmerliste verwaltet werden
- Prüfungsdaten, die sich aus Prüfungstermin, Prüfer, Prüfling, Fach oder Fächerkombination und Prüfungsergebnis zusammensetzen

Benutzergruppen

Universitätsangehörige lassen sich unter anderem in folgende Benutzergruppen einteilen

- Studenten
- Professoren
- Wissenschaftliche Mitarbeiter
- Verwaltungspersonal

Anwendungsprogramme

- Studentenverwaltung: Dies schließt unter anderem das Eintragen und die Pflege personenbezogener Daten mit ein. Spezielle Applikationen sind auch für die Verwaltung von Prüfungsergebnissen (z.B. Erstellen von Zeugnissen) notwendig.
- Verwaltung der (Lehr-)Veranstaltungen. Dazu zählt die Zuteilung von Räumen und die Publikation von Lehrveranstaltungen.
- Personalverwaltung, etwa die Pflege der Personendaten, Abrechnung von Aufwendungen (z.B. Reisekosten) und Lohnabrechnung.

Verwaltung ohne Datenbanksystem

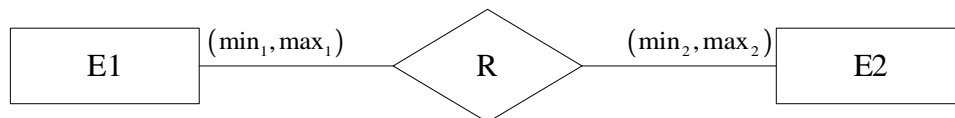
Ohne den Einsatz eines Datenbanksystems werden die entsprechenden Daten in getrennten Dokumenten in elektronischer oder auch schriftlicher Form verwaltet. Dies hat Nachteile in Bezug auf

- Ineffizienz bei Zugriffen: Die Selektion bestimmter Datensätze ist dann sehr aufwendig. Selbst wenn der Zugriff über Applikationen erfolgt, d.h. keine manuelle Suche, so entartet dies dann zu einer Überprüfung einer größeren Datenmenge. Sucht man beispielsweise die Studenten, die sich aktuell im 4. Semester befinden, so müssen dazu evtl. alle Studenten betrachtet und davon die, die im 4. Semester sind, ausgewählt werden. Datenbanken unterstützen hier optimierte Zugriffsmethoden.

- Änderung des Datenbestands: Nach Semesterende muss unter anderem die Semesterzahl der Studenten aktualisiert werden. Soll dies nicht manuell erfolgen, bedarf es dazu spezialisierter, d.h. eigens entwickelter Applikationen.
- Wahrung der Datenkonsistenz: Ohne effiziente Zugriffsmöglichkeiten auf die Daten (z.B. die gezielte Selektion eines Studenten) ist es ggf. sehr aufwendig zu überprüfen, dass Daten nicht mehrfach eingetragen werden, bzw. wenn Daten redundant verwaltet werden, dass alle Datensätze auf demselben Stand gehalten werden.
- Unterschiedliche Datenformate: Es besteht die Gefahr, dass die unterschiedlichen Informationseinheiten (etwa personenbezogene Daten und Veranstaltungsdaten) in unterschiedlichen Formaten abgelegt werden. Applikationen, die Zugriff auf beide Informationstypen benötigen, sind dann oft schwer zu realisieren und auch schwer wartbar.

Aufgabe 3

In folgender Abbildung sind (min, max)-Beziehungen schematisch dargestellt:



Funktionalitätsangaben werden durch (min, max)-Notation wie folgt ausgedrückt:

$1 : 1$:	$(\min_1, \max_1) = (0, 1)$	$(\min_2, \max_2) = (0, 1)$
$1 : N$:	$(\min_1, \max_1) = (0, *)$	$(\min_2, \max_2) = (0, 1)$
$N : 1$:	$(\min_1, \max_1) = (0, 1)$	$(\min_2, \max_2) = (0, *)$
$N : M$:	$(\min_1, \max_1) = (0, *)$	$(\min_2, \max_2) = (0, *)$