

Vorlesung Datenbanken im SomSem 2019

— Probeklausur —

Name, Vorname: _____

Matrikelnummer: _____

Studiengang: _____

Aufgabe	Max. Punkte	Punkte
1	22	
2	5	
3	24	
4	12	
5	8	
6	8	
Σ	79	

Hinweise:

- Bearbeitungsdauer: 90 Minuten
- Sie können die Folienkopien, das Textbuch sowie Ihre Unterlagen zur Vorlesung verwenden. Notebooks, PDAs und andere elektronische Geräte dürfen nicht verwendet werden.
- Schalten Sie für die Dauer der Klausur Ihr Mobiltelefon aus.
- Die Klausur besteht aus 6 Aufgaben auf 5 Seiten. Bitte überprüfen Sie die Vollständigkeit.
- **(Gilt nicht für die Probeklausur!)** Bitte benutzen Sie den zu jeder Aufgabe vorgegebenen Platz; dieser sollte zur Lösung der Aufgabe ausreichen. Wenn Sie noch die Rückseite verwenden, markieren Sie bei der Aufgabe klar, dass es eine Fortsetzung auf der Rückseite gibt.
- Tauschen Sie auf keinen Fall irgendwelche Notizen etc. mit den Nachbarn aus.
- Fragen Sie, wenn Ihnen eine Aufgabe nicht klar ist!
- Benutzen Sie keinen Bleistift.

Aufgabe 1: ER Modellierung

(11+2+9=22 Punkte)

Nach mehreren Jahren der Papierverwaltung haben sich die Redakteure des Bammentaler Kuriers entschieden, eine Datenbank zur Verwaltung von Artikeln und Photos zu verwenden. Nachfolgend sind die Anforderungen angegeben, die sich aus der Anforderungsanalyse ergeben haben.

- Ein Photo hat eine Nummer (ID), ein Datum sowie eine Höhe und eine Breite.
- Bei einem Photo wird unterschieden, ob das Photo von einem Photographen des Bammentaler Kuriers gemacht wurde oder ob es von einer Agentur gekauft wurde. Im ersten Fall wird nur der Name des Photographen gespeichert, im zweiten Fall wird der Name der Agentur sowie der Kaufpreis des Photos gespeichert.
- Informationen zu Personen, die auf einem Photo zu sehen sind, werden separat gespeichert. Zu einer Person wird der Name, das Geburtsdatum sowie der Bereich ("Sport", "Politik", "Film" etc.) gespeichert.
- Zu jedem Photo wird gespeichert, welche Personen auf dem Photo zu sehen sind.
- Ein Artikel hat einen Titel, eine Liste (keine Menge!) von Autoren sowie ein Datum, an dem der Artikel im Bammentaler Kurier erschienen ist. Informationen zu Autoren werden separat von den Artikeln verwaltet. Für jeden Artikel wird auch noch verwaltet, welche Personen in dem Artikel erwähnt werden und welche Photos zusammen mit dem Artikel gezeigt werden.

Führen Sie die nachfolgenden Entwurfstätigkeiten aus, basierend auf den oben angegebenen Anforderungen. Verwenden Sie nur Notationen aus der Vorlesung.

- a. Geben Sie ein ER-Schema an, das die obigen Informationen entsprechend beschreibt. Spezifizieren Sie geeignete Primärschlüsselattribute sowie Kardinalitäten mithilfe der [min,max]-Notation zu den Beziehungstypen.
- b. Geben Sie umgangssprachlich zwei wichtige und notwendige Integritätsbedingungen an, die Sie nicht in Ihrem ER-Schema bzw. im ER-Modell ausdrücken können.
- c. Übersetzen Sie Ihr ER-Schema in eine Menge von Relationenschemata. Geben Sie alle Primär- und Fremdschlüssel zu den Relationen an.

Beispieldatenbank:

Die Vereinigung der Speditionen in Baden-Württemberg (VSBW) hat ein Datenbanksystem zur Verwaltung der Mitglieder, Zweigstellen und Transportrouten gekauft. Das dazugehörige Datenbankschema besteht aus den folgenden vier Relationen.

```
STADT(plz, name, bundesland)
SPEDITION(sname, plz, zweigst, lkws)
ZWEIGSTELLE(sname, plz, lkws)
ROUTE(sname, start-plz, end-plz, km)
```

In der Relation STADT werden Informationen über Städte verwaltet, inklusive Postleitzahl, Name der Stadt und Bundesland.

Die Relation SPEDITION verwaltet Informationen zu den einzelnen Speditionen. Eine Spedition hat einen Namen (sname), eine Anzahl von Zweigstellen (zweigst) und eine Gesamtzahl von LKWs (lkws). Zu jeder Spedition wird auch die Postleitzahl des Hauptsitzes der Spedition abgespeichert.

Ein Spedition kann mehrere Zweigstellen haben, welche auch immer den Hauptsitz beinhaltet. Für jede Zweigstelle wird der Name der Spedition, die Postleitzahl der Zweigstelle sowie die Anzahl der LKWs bei dieser Zweigstelle gespeichert.

Schließlich werden noch in der Relation ROUTE Informationen zu den Strecken gespeichert, die von den verschiedenen Speditionen gefahren werden. Hierzu werden neben der Spedition (sname) die Postleitzahl des Start- und Zielortes sowie die Entfernung zwischen den beiden Orten verwaltet.

Aufgabe 2: Integritätsbedingungen

(3+2=5 Punkte)

- Welche Attribute in den obigen Relationen sind Fremdschlüsselattribute und welche Relationen referenzieren diese Attribute? Sie können hierzu folgende Notation verwenden, um zu beschreiben, dass ein Fremdschlüssel der Relation 1 auf den Primärschlüssel einer Relation 2 verweist: $\langle \text{Relation1} \rangle . \langle \text{Attribut(e)} \rangle \longrightarrow \langle \text{Relation2} \rangle$.
- Gegeben sei folgende Integritätsbedingung: *Die Gesamtzahl der LKWs einer Spedition muss mindestens gleich der Summe der LKWs an den Zweigstellen der Spedition sein.* Geben Sie alle Operationen (insert, update, delete auf den obigen Relationen) an, die diese Bedingung verletzen können.

Aufgabe 3: SQL-Anfragen

(2+4+3+4+5+3+3=24 Punkte)

Formulieren Sie folgende Anfragen an die Beispieldatenbank in SQL:

- Geben Sie die Postleitzahl der Zweigstellen an, die mindestens 50% aller LKWs einer Spedition haben.
- Welche Speditionen haben ihre Zweigstellen nur in einer Stadt (aber eventuell unter verschiedenen Postleitzahlen)? Formulieren Sie diese Anfrage ohne Aggregationen oder Gruppierung. Auszugeben sind Name der Spedition und Name der Stadt.
- Wie viele Zweigstellen hat die Spedition mit den meisten LKWs? Auszugeben sind der Name der Spedition und die Anzahl der Zweigstellen.

- d. Welche Speditionen haben ihren Hauptsitz in “Karlsruhe“ und bedienen mehr als 50 verschiedene Routen?
- e. Gibt es Paare von Routen, die die gleichen Start- und Endpostleitzahlen aber verschiedene Entfernungen haben? Auszugeben sind hierzu komplette Tupel der Relation ROUTE.
- f. Geben Sie für jede Route mit mehr als 700 km die Anzahl der Speditionen an, die diese Route bedienen. Gesucht sind nur Routen, die von mehr als 5 Speditionen bedient werden. Auszugeben sind neben der Anzahl der Speditionen die Start- und Endpostleitzahl.
- g. Die Spedition “HotWheels” hat zwei neue Zweigstellen in den Städten mit den Postleitzahlen “69120“ und “76646” aufgemacht. Insgesamt hat die Spedition 20 neue LKWs, die zu gleichen Teilen auf die beiden neuen Zweigstellen verteilt werden. Geben Sie alle Update-Operationen auf entsprechende Relationen an, um diese Änderungen durchzuführen.

Aufgabe 4: Anfragebearbeitung und -optimierung

(1+2+3+6=12 Punkte)

- a. Was wird durch die folgende SQL-Anfrage bestimmt?

```
select S.sname, T.name, R.km
from SPEDITION S, ROUTE R, STADT T
where S.sname = r.sname and
      R.end-plz = 69120 and R.start-plz = T.plz
```

- b. Übersetzen Sie diese Anfrage **direkt** in einen (kanonischen) Ausdruck der Relationenalgebra. Verwenden Sie nur die Grundoperationen der Algebra (beachten Sie hierbei, dass z.B. der Join nur eine abgeleitete Operation ist).
- c. Führen Sie eine algebraische Optimierung für den unter b) bestimmten Ausdruck unter Verwendung der Transformationsregeln durch. Es reicht aus, wenn Sie hierzu nur das Endergebnis (Anfragebaum) angeben.
- d. Angenommen für die Primärschlüssel in den obigen Relationen existiert jeweils ein geclusterter Index. Es sei weiterhin angenommen, dass ein nicht geclusterter Index für das Attribut end-plz existiert.

Geben Sie einen effizienten Ausführungsplan für die von Ihnen unter c) beschriebene Anfrage an. Welche Algorithmen verwenden Sie für die Joins? Begründen Sie Ihre Wahl. Begründen Sie auch, warum der von Ihnen erstellte Ausführungsplan am effizientesten ist.

Aufgabe 5: Relationale Algebra

(4+2+2=8 Punkte)

Formulieren Sie folgende Anfragen als Ausdrücke der relationalen Algebra. Wenn möglich, wenden Sie die Transformationsregeln an, um eine effiziente Anfrage zu erhalten.

- a. Welche Spedition (sname) aus Mannheim hat die meisten LKWs?
- b. Geben Sie alle Routen an, die von Speditionen bedient werden, die eine Zweigstelle in Sindelfingen haben.

Formulieren Sie die folgende Anfrage umgangssprachlich:

c. $\pi_{\text{sname}}(\sigma_{\text{lkws} > 100 \wedge \text{blkws} > 90}(\text{SPEDITION} \bowtie (\beta_{\text{blkws} \leftarrow \text{lkws}}(\pi_{\text{sname, lkws}}(\text{ZWEIGSTELLE}))))))$

Aufgabe 6: Transaktionsverwaltung

(5+3=8 Punkte)

a. Gegeben seien die beiden folgenden Schedules S_1 und S_2 :

S_1 :

T_1	T_2	T_3
read(X)	read(Z)	read(X)
read(Z)	read(Y)	read(Y)
write(X)	write(Z)	write(Y)
	write(Y)	

S_2 :

T_1	T_2	T_3
read(X)	read(Z)	
read(Z)		read(X)
write(X)		read(Y)
	read(Y)	write(Y)
	write(Z)	
	write(Y)	

Welcher der beiden Schedules ist serialisierbar? Zeichnen Sie jeweils den zugehörigen Konfliktgraphen und geben Sie gegebenenfalls jeweils einen seriellen Schedule an (hier reicht die Reihenfolge der Transaktionen).

b. Angenommen sei die folgende Transaktion T . Zu Beginn der Transaktion hat das Objekt A den Wert 100.

Zeit	Operation
1	read(A,a)
3	$a := a + 100$
5	write(A,a)
7	read(B,b)
9	write(B,b)
11	commit

- Angenommen sei eine weitere Transaktion T' , die das Objekt A zum Zeitpunkt 4 lesen will. Welchen Wert für A sollte T' zu lesen bekommen? Woher nimmt das DBMS den entsprechenden Wert für A ?
- Angenommen, das System stürzt zum Zeitpunkt 9 ab. Welchen Wert hat A nach einem Neustart des Systems? Begründen Sie Ihre Antwort.