

Klausur Datenbanken
Wintersemester 2007/2008
Prof. Dr. Wolfgang May
5. Februar 2008, 11-13 Uhr
Bearbeitungszeit: 90 Minuten

Vorname:

Nachname:

Matrikelnummer:

Studiengang:

Bei der Klausur sind **keine Hilfsmittel** (Skripten, Taschenrechner, etc.) erlaubt. Handies müssen ausgeschaltet sein. Papier wird gestellt. Benutzen Sie nur die **ausgeteilten**, zusammengehefteten **Blätter** für Ihre Antworten. Schreiben Sie mit blauem/schwarzem Kugelschreiber, Füller, etc.; Bleistift ist nicht erlaubt.

Auf dem letzten Blatt finden Sie eine Datenbasis, die in den Aufgaben 1 und 2 verwendet wird. Trennen Sie es ggf. zur Bearbeitung der Aufgaben ab.

Zum **Bestehen** der Klausur sind **45** Punkte hinreichend.

- meine Note soll mit Matrikelnummer so bald wie möglich auf der Vorlesungs-Webseite veröffentlicht werden.
- meine Note soll nicht veröffentlicht werden; ich erfahre sie dann aus Munopag/Wopag/FlexNever oder beim zuständigen Prüfungsamt.

	Max. Punkte	Erreichte Punkte
Aufgabe 1 (ER-Modell)	15	
Aufgabe 2 (SQL und Relationale Algebra)	35	
Aufgabe 3 (Anfrageoptimierung und Auswertung)	21	
Aufgabe 4 (Transaktionen)	19	
Summe	90	

Note:

Name:

MatNr.:

Aufgabe 1 (ER-Modell [15 Punkte])

Geben Sie ein ER-Modell für den folgenden Sachverhalt an (die für die SQL-/Algebra-Aufgabe angegebene Datenbasis ist ein Teil dieses Szenarios). Geben Sie Schlüssel und Kardinalitäten sinnvoll an.

- Es geht um eine Datenbank über Restaurants.
- Jedes Restaurant hat einen Namen und befindet sich in einer Stadt (Hinweis: für Städte wird der Name als Schlüssel angenommen). In verschiedenen Städten kann es Restaurants mit demselben Namen geben (“Ratskeller”, “Zum Goldenen Ochsen”, “La Trattoria”). Jedes Restaurant hat eine Adresse.
- Jedem Restaurant ist eine Preiskategorie zugeordnet (extrem teuer=5, billig=1).
- Für jedes Restaurant ist angegeben, welche Art von Gerichten dort angeboten wird (Italienisch, Fisch, ...). Ein Restaurant kann dabei mehrere Einträge haben (z.B. passen Indisch und Vegetarisch gut zusammen). “Local” bedeutet, dass dort die landesübliche Küche gepflegt wird.

Weiterhin gibt es Personen:

- Jede Person hat einen Namen und lebt in einer Stadt.
- Jede Person bevorzugt ein oder mehrere Koch- bzw. Küchenstile (z.B. Fisch, Portugiesisch und French Cuisine).

Name:

MatNr.:

Aufgabe 2 (SQL und Relationale Algebra [35 Punkte])

Verwenden Sie für diese Aufgabe die Datenbasis, die auf dem letzten Blatt der Klausur angegeben ist.

- a) Geben Sie eine SQL-Anfrage *und* einen Algebra-Ausdruck bzw. Baum an, der die Menge der Namen aller Städte ergibt, in denen man italienisch essen kann. (3P)
- b) Geben Sie eine SQL-Anfrage *und* einen Algebra-Ausdruck bzw. Baum an, der die Namen aller Personen ergibt, die kein italienisches Essen mögen. (6P)
- c) Geben Sie eine SQL-Anfrage *und* einen Algebra-Ausdruck bzw. Baum an, die für jede Person angibt, in welchen Restaurants ihrer Heimatstadt sie landesübliche Gerichte (Style="Local" bedeutet, dass ein Restaurant landesübliche Küche anbietet) angeboten bekommt [Antworten als Paare (Person, Restaurant)]. (6P)
- d) Geben Sie eine SQL-Anfrage *oder* einen Algebra-Ausdruck bzw. Baum an, die die Menge der Namen aller Restaurants ergibt, in denen Alice und Peter sich zum gemeinsamen Essen verabreden können, und beide etwas finden, das sie mögen. (6P)
- e) Geben Sie eine SQL-Anfrage an, die für jede Stadt und Person die Anzahl der Restaurants ergibt, in denen diese Person gerne essen gehen würde. Es sollen nur Städte ausgegeben werden, in denen es mindestens drei solche Restaurants gibt. (7P)
- f) Geben Sie eine SQL-Anfrage *oder* einen Algebra-Baum *oder* einen Algebra-Ausdruck an, der alle Paare (Person, Stadt) zurückgibt, so dass jeder der von dieser Person bevorzugten Stile in dieser Stadt durch mindestens ein Restaurant vertreten ist. (7P)

Name:

MatNr.:

Aufgabe 3 (Anfrageoptimierung und Auswertung [21 Punkte])

Gegeben sind zwei Tabellenschemata R und S :

R		
<u>A</u>	<u>B</u>	C

S	
<u>E</u>	<u>F</u>

Die Aufgabe behandelt das Join $R \bowtie_{R.A=S.E} S$.

Physikalisches Datenbankschema:

- Schlüssel von R ist (A, B) , Schlüssel von S ist (E, F) .
- Der Datentyp von $R.A$ und $S.E$ ist "Integer", d.h., ganze nichtnegative Zahlen.
- Der Datentyp von $R.C$ ist CHAR(20), d.h. Zeichenketten der festen Länge 20.
- Die Tupel von R sind in beliebiger Reihenfolge gespeichert, es existieren keine Indexe auf R .
- Die Tupel von S sind in beliebiger Reihenfolge gespeichert, es existieren keine Indexe auf S .

Daten:

- R enthält 2.000.000 Tupel.
- S enthält 4.000.000 Tupel.
- Die Tabellen sind wie üblich seitenweise abgelegt. Jede Seite (4KB) von R enthält 50 Tupel, jede Seite von S enthält 100 Tupel.
- Im Hauptspeicher können 1020 Seiten gleichzeitig gehalten werden.
- 491, 907 und 1019 sind Primzahlen.

Fragen (jeweils mit kurzer Begründung zu beantworten):

- Wieviele Speicherseiten umfassen R und S ? (2P)
- Wieviele Bytes umfasst (grob) ein R -Tupel? (1P)
- Die Anfrage $\pi[A, B, F](R \bowtie_{R.A=S.E} S)$ soll berechnet und am Bildschirm ausgegeben werden.
Geben Sie den Algebra-Baum der Anfrage an und optimieren Sie die Anfrage algebraisch. (2P)
- Beschreiben Sie, wie man diese Anfrage auswertet, wobei ein *Hash-Join* verwendet werden soll; Sie dürfen dazu bis zu 1000 Hauptspeicherseiten verwenden.
Wieviele Hintergrundspeicherzugriffe auf die Datenbank werden benötigt?
(beschreiben Sie am besten zuerst grob, wie Sie vorgehen, und analysieren Sie dann die Zugriffe) (insgesamt 12P)
- (Bearbeiten Sie diesen Teil erst, wenn Sie die Lösung für (d) ausgearbeitet haben!)
Annahme: man weiss, dass $\pi[A](R) \subsetneq \pi[E](S)$ ("echte Teilmenge") ist, d.h. für z.B. 60% der Werte von $S.E$ kein passender Wert in $R.A$ vorhanden ist. An welcher Stelle (und wie) kann man das obige Verfahren dann optimieren? (4P)

Name:

MatNr.:

Name:

MatNr.:

Aufgabe 4 (Transaktionen [19 Punkte])

Gegeben ist der folgende Schedule S :

$$R_2A R_3A R_3B W_3C R_1C R_1B W_1B R_2B W_2A W_1C$$

- Geben Sie den D-Graphen von S an (4P).
- Geben Sie den C-Graphen von S an (4P).
- Ist S serialisierbar (mit Begründung; 2P)
- Betrachten Sie den *leicht veränderten* Schedule S' , bei dem die beiden angegebenen Aktionen vertauscht sind:

$$R_2A R_3A R_3B W_3C R_1C R_1B \boxed{R_2B W_1B} W_2A W_1C$$

Ist S' äquivalent zu S ?

Ist S' serialisierbar?

(jeweils mit Begründung; 3+3P)

- Ist S bzgl. einem Scheduler, der nach dem Zeitstempelverfahren arbeitet, zulässig (3P)?

Name:

MatNr.:

Name:

MatNr.:

[Trennen Sie dieses Blatt am besten vor Beginn der Bearbeitung ab]

Die folgende Datenbasis wird in Aufgabe 2 verwendet.

Restaurant				Styles		
City	Name	Address	Category	City	Restaurant	Style
Paris	L'Arpege	...	5	Paris	L'Arpege	French
Paris	Buddha Bar	...	5	Paris	Buddha Bar	Indian
Paris	Le Train Bleu	...	4	Paris	Le Train Bleu	French
Paris	Niollaville	...	3	Paris	Le Train Bleu	Local
München	Il Grappolo	...	3	Paris	Niollaville	Chinese
München	Makassar	...	3	München	Il Grappolo	Italian
München	Shoya	...	3	München	Makassar	Indian
München	Augustiner	...	2	München	Makassar	Fish
Vienna	Rimini	...	4	München	Shoya	Japanese
Vienna	Churrascaria	...	3	München	Augustiner	Local
Lisbon	Sette Mares	...	3	Vienna	Rimini	Italian
Lisbon	Bica do Sapata	...	4	Vienna	Churrascaria	Brazilian
Lisbon	Estoril Mandarin	...	5	Lisbon	Sette Mares	Fish
Lisbon	Armazem de Cachaca	...	4	Lisbon	Bica do Sapata	Portuguese
Lisbon	Casanostra	...	3	Lisbon	Bica do Sapata	Local
Lisbon	Os Tibetanos	...	3	Lisbon	Estoril Mandarin	Chinese
Roma	Gino da Trastevere	...	5	Lisbon	Armazem de Cachaca	Brazilian
Roma	Agata e Romeo	...	4	Lisbon	Casanostra	Italian
Roma	Le Relais de Jardin	...	5	Lisbon	Os Tibetanos	Vegetarian
Roma	Margutta Vegetariano	...	4	Roma	Gino da Trastevere	Fish
:	:	:	:	Roma	Gino da Trastevere	Local
				Roma	Agata e Romeo	Fish
				Roma	Agata e Romeo	Italian
				Roma	Le Relais de Jardin	French
				Roma	Margutta Vegetariano	Vegetarian
				:	:	:

City	
Name	Country
Lisbon	P
Paris	F
Rome	I
:	:

Person	
Name	Place
Alice	Paris
Peter	Vienna
:	:

likes	
Name	Style
Alice	French
Alice	Vegetarian
Alice	Indian
Alice	Fish
Alice	Local
Peter	Italian
Peter	Portuguese
Peter	Brazilian
Peter	Local
:	:

* information taken from several Marco Polo guides from the last decade.