

Klausur Datenbanken
Wintersemester 2006/2007
Prof. Dr. Wolfgang May
12. Februar 2007, 11-13 Uhr
Bearbeitungszeit: 90 Minuten

Vorname:

Nachname:

Matrikelnummer:

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 (bzw. CLZ: beim Prüfungsamt dort zu erfragen).

	Max. Punkte	Erreichte Punkte
Aufgabe 1 (ER-Modell)	15	
Aufgabe 2 (SQL und Relationale Algebra)	35	
Aufgabe 3 (Speicherung und Auswertung)	20	
Aufgabe 4 (Transaktionen)	20	
Summe	90	

Note:

Aufgabe 1 (ER-Modell [15 Punkte])

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

- Es gibt Tierarten. Jede *Tierart* hat einen Namen (z.B. “Löwe”, “Elch”, “Panda”, “Beutelteufel”, ...). Jede Tierart gehört zu einer *Gattung*, und erreicht ein bestimmtes *Gewicht*: Löwen sind Raubkatzen und bis zu 250kg schwer, Elche sind Huftiere und bis zu 800kg schwer, und Beutelteufel sind Beuteltiere und bis zu 10kg schwer. Tierarten haben ausserdem einen bevorzugten *Lebensraum*. Löwen leben in der Savanne, Elche in der Tundra, Beutelteufel leben im Urwald.
- Es gibt *Länder*. Länder haben einen Namen.
- Es gibt *Kontinente*. Kontinente haben einen Namen, Länder liegen auf Kontinenten.
- Der weltweite, freilebende *Bestand* der verschiedenen Tierarten verteilt sich auf viele Länder, oft auch auf bestimmte Gebiete in diesen Ländern. Z.B. leben 300000 Elche in Schweden, 200000 Elche in Norwegen, aber nur noch ca. 50000 Beutelteufel in Australien und ca. 1600 Pandas in China, genauer gesagt in der zu China gehörenden Gegend “Sichuan”.
- Ausserdem gibt es *Zoos*, die insbesondere auch bei der Erhaltung bedrohter Tierarten, wie z.B. dem Panda und dem Beutelteufel eine wichtige Rolle spielen. Zoos liegen in Städten, die wiederum in Ländern liegen; z.B. in Frankfurt in Deutschland. In jedem Zoo werden mindestens 3 Tierarten gehalten.
- In den Zoos gibt es ebenfalls (kleine) Bestände vieler Tierarten. Z.B. leben im Frankfurter Zoo 10 Elche, sowie 2 Pandas (und viele andere).

Name:

MatNr.:

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

Verwenden Sie für Aufgabe 2 die Datenbasis, die auf dem letzten Blatt der Klausur angegeben ist.

1. Geben Sie die **CREATE TABLE**-Statements zur Generierung der beiden Tabellen **Zoo** und **ZooBestand** so komplett wie möglich an (10P).
2. Geben Sie ein **SQL-Statement** an, das die Menge der Namen aller Raubkatzenarten angibt, die in deutschen Zoos gehalten werden. (3P)
3. Geben Sie einen **Ausdruck (oder Baum) der relationalen Algebra** an, der (2) beantwortet. (3P)
4. Geben Sie ein **SQL-Statement** an, das das folgende tut:
“Die Namen aller Tierarten, von denen es weltweit höchstens 50000 freilebende Exemplare gibt”. (4 P)
5. Geben Sie ein **SQL-Statement** an, das das folgende tut:
“Alle Namen von Tierarten, die in Namibia frei leben, und nicht in deutschen Zoos vertreten sind”. (5P)
6. Geben Sie einen **Algebra-Ausdruck oder -Baum** an, der die Namen aller Länder ergibt, in denen es keine freilebenden Raubkatzen gibt (5P)
7. Geben Sie eine **SQL-Anfrage oder einen Algebra-Ausdruck/Baum** an (d.h. entscheiden Sie selber, welche Anfragesprache Sie bevorzugen), der folgende Anfrage beantwortet: “Namen aller Tierarten, von denen es in *jedem* europäischen Zoo mindestens ein Exemplar gibt.” (5P).

Aufgabe 3 (Speicherung und Auswertung [20 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 Wertebereich von $R.A$ ist gleich dem Wertebereich von $S.E$, d.h. $\pi[A](R) = \pi[E](S)$.
- Die Tupel von R sind in beliebiger Reihenfolge gespeichert, es existieren keine Indexe auf R .
- Die Tupel von S sind so gespeichert, dass Tupel mit demselben E -Wert benachbart gespeichert sind. Eine weitergehende Ordnung existiert nicht.
- Für das Attribut $S.E$ ist ein B*-Baumindex angelegt, dessen Einträge jeweils auf das erste Tupel mit dem entsprechenden Wert von $S.E$ zeigen.

Daten:

- R enthält 2.000.000 Tupel [Hinweis: diese Schreibweise bedeutet in Deutschland "2 Millionen"].
- S enthält 4.000.000 Tupel.
- Die Tabellen sind wie üblich seitenweise abgelegt. Eine Seite (4KB) von R enthält 100 Tupel, eine Seite von S enthält 60 Tupel.
- Zu jedem Wert a von A gibt es genau 2 Tupel $\mu \in R$, so dass $\mu[A] = a$ ist.
- Zu jedem Wert e von E gibt es durchschnittlich 4 Tupel $\mu \in S$, so dass $\mu[E] = e$ ist.
- Für den B*-Baum über $S.E$ gilt folgendes: Jedes Blatt enthält durchschnittlich 100 Einträge. Jeder innere Knoten enthält durchschnittlich 200 Einträge.
- Im Hauptspeicher können 1000 Seiten gleichzeitig gehalten werden.

Fragen (jeweils mit kurzer Begründung zu beantworten):

1. Wieviele Werte enthält $\pi[A](R)$? (2P)
2. Wieviele Blätter enthält der B*-Baum über $S.E$? (2P)
3. Wie viele Ebenen hat der Baum?
Wieviele Knoten enthält der Baum insgesamt? (5P)
4. Die Anfrage $\pi[A, B, F](R \bowtie_{R.A=S.E} S)$ soll berechnet und am Bildschirm ausgegeben werden.
Wieviele Ergebnistupel bekommt man? (1P)
5. Beschreiben Sie, wie die Auswertung (unter Verwendung des Indexes) am besten

vorgeht.

Wieviele Hintergrundspeicherzugriffe auf die Datenbank werden benötigt?

(beschreiben Sie am besten zuerst grob, wie Sie vorgehen, und analysieren Sie dann die Zugriffe) (insgesamt 10P)

Name:

MatNr.:

Aufgabe 4 (Transaktionen [20 Punkte])

Gegeben ist der folgende Schedule:

$R_2B R_1A W_1C R_1B R_3C W_2B R_3B W_1A W_3B R_2A W_2A$

- a) Geben Sie den Dependency-Graphen von S an (berücksichtigen Sie dabei auch die Transaktionen T_0 und T_∞ wie in der Vorlesung). (4P)
- b) Geben Sie den Konfliktgraphen von S an. (4P)
- c) Ist S serialisierbar (mit Begründung) (2P)
- d) Geben Sie für jedes der folgenden Scheduling-Verfahren an, ob der Schedule nach dem Verfahren möglich ist (mit Begründung; 10P):
 - Überwachung des C-Graphen
 - Zeitstempel
 - Optimistisches Scheduling
 - 2-Phasen-Locking

Name:

MatNr.:

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

Die folgende Datenbasis wird in Aufgabe 2 verwendet.

Diese Datenbasis enthält die Daten über die weltweit freilebenden sowie in Zoos gehaltenen Bestände von verschiedenen Tierarten.

Tierart			
<u>Name</u>	<u>Gattung</u>	<u>Gewicht</u>	<u>Lebensraum</u>
Elch	Huftier	800	Tundra
Warzenschwein	Huftier	120	Savanne
:	Huftier	:	:
Rotes Känguruh	Beuteltier	80	Savanne
Beutelteufel	Beuteltier	10	Urwald
:	Beuteltier	:	:
Panda	Bär	150	Wald
:	:	:	:

Land	
<u>Name</u>	<u>Kontinent</u>
Deutschland	Europa
Schweden	Europa
China	Asien
Australien	Australien
Namibia	Afrika
:	:

Bestand			
<u>Tierart</u>	<u>Land</u>	<u>Gegend</u>	<u>Anzahl</u>
Elch	Schweden		300000
Elch	Norwegen		200000
Elch	:		:
Warzenschwein	Namibia		1000000
Warzenschwein	:		:
Rotes Känguru	Australien		8000000
Beutelteufel	Australien	Tasmanien	50000
Panda	China	Sichuan	1600
:	:	:	:

Zoo	
<u>Ort</u>	<u>Land</u>
Frankfurt	Deutschland
Berlin	Deutschland
Paris	Frankreich
Sydney	Australien
Peking	China
:	:

ZooBestand		
<u>Ort</u>	<u>Tierart</u>	<u>Anzahl</u>
Frankfurt	Elch	5
Berlin	Warzenschwein	7
Berlin	Panda	2
Peking	Panda	10
Paris	Warzenschwein	10
Sydney	Beutelteufel	30
:	:	:

Dabei wird angenommen, dass der Name einer Stadt weltweit eindeutig ist