

# Hauptdiplomklausur Datenbanksysteme I Wintersemester 2003/2004

Name: .....

Vorname: .....

Matrikelnummer: .....

Studienfach: .....

Wichtige Hinweise:

1. Prüfen Sie Ihr Klausurexemplar auf Vollständigkeit (14 Seiten).
2. Es sind keine Hilfsmittel zugelassen.
3. Die Klausur dauert 100 Minuten.
4. Jede Aufgabe ist auf dem zugehörigen Aufgabenblatt (und ggf. auf separaten Lösungsblättern) zu bearbeiten.
5. Vermerken Sie Ihren Namen und Ihre Matrikelnummer auf jedem Aufgaben- (bzw. Lösungs-)blatt. Blätter ohne Angabe von Namen und Matrikelnummer werden nicht bewertet.
6. Das Deckblatt sowie alle Aufgabenblätter (evtl. Lösungsblätter) sind abzugeben.

	maximale Anzahl Punkte	erreichte Anzahl Punkte
Aufgabe 1	15	
Aufgabe 2	15	
Aufgabe 3	7	
Aufgabe 4	24	
Aufgabe 5	4	
Aufgabe 6	8	
Aufgabe 7	13	
Aufgabe 8	14	
	100	

1. (a) (3 Punkte)

Was steht in der internen Datensatztafel einer Seite (Slots einer Slotted Page) im Hintergrundspeicher beim TID-Verfahren?

(b) (3 Punkte)

Erläutern Sie kurz den Unterschied zwischen den Strategien *force* und  $\neg$  *force* beim Einbringen von Änderungen einer Transaktion.

(c) (3 Punkte)

Geben Sie einen äquivalenten Ausdruck zu  $(R \bowtie_{R.A=S.B} S) \bowtie_{S.C=T.D} T$  nur mit Kreuzprodukten und Selektionen an.

(d) (3 Punkte)

Erläutern Sie kurz, was sich hinter dem WAL-Prinzip verbirgt.

(e) (3 Punkte)

Was ist ein „Dirty Read“?

2. (a) (5 Punkte)

Gegeben sei ein Relationenschema  $\mathcal{R}(A, B, C, D, E, F)$  mit folgenden funktionalen Abhängigkeiten  $\mathcal{F}_{\mathcal{R}} = \{B \rightarrow C, C \rightarrow DE, E \rightarrow F, F \rightarrow A\}$ . Geben Sie alle Schlüssel des Schemas an. In welcher höchsten Normalform befindet sich  $\mathcal{R}$ ?

(b) (5 Punkte)

Gegeben sei ein Relationenschema  $\mathcal{S}(A, B, C, D, E, F)$  mit folgenden funktionalen Abhängigkeiten  $\mathcal{F}_{\mathcal{S}} = \{E \rightarrow ACF, F \rightarrow BDE\}$ . Geben Sie alle Schlüssel des Schemas an. In welcher höchsten Normalform befindet sich  $\mathcal{S}$ ?

(c) (5 Punkte)

Das Relationenschema  $\mathcal{T}(A, B, C, D, E, F)$  mit folgenden funktionalen Abhängigkeiten  $\mathcal{F}_{\mathcal{T}} = \{AB \rightarrow CDE\}$  befindet sich momentan in 1NF. Fügen Sie eine funktionale Abhängigkeit in  $\mathcal{F}_{\mathcal{T}}$  ein, so dass  $A$  alleiniger Schlüssel und  $\mathcal{T}$  in 4NF ist.

3. (a) (3 Punkte)

Zwischen den Entitytypen  $A$  und  $B$  existiert die funktionale Beziehung  $A \rightarrow B$ . Stellen Sie diesen Sachverhalt in einem ER-Diagramm dar.

(b) (4 Punkte)

Zwischen den Entitytypen  $A$ ,  $B$  und  $C$  existieren die funktionalen Beziehungen  $A, B \rightarrow C$  und  $A, C \rightarrow B$ . Stellen Sie diesen Sachverhalt in einem ER-Diagramm dar.

4. Der Produktionsprozess einer Fabrik wird mit Hilfe eines relationalen Datenbanksystems verwaltet. Das Schema der Datenbank umfasst dabei drei Relationen:

- Fließband(FID, Ort)
- Produkt(PID, Name, QCheck)
- hergestellt(FID, PID, Datum, Uhrzeit)

Die Relation *Fließband* gibt an, welches Fließband an welchem Ort steht (z.B. Halle 1). In der Relation *Produkt* wird vermerkt, welches Produkt welchen Namen trägt und welche Qualitätsstufe das Produkt beim Qualitätscheck (QCheck) erreicht hat. In *hergestellt* wird festgehalten, an welchem Tag und zu welcher Uhrzeit ein Produkt von welchem Band gelaufen ist.

Formulieren Sie folgende Anfragen in SQL:

(a) (5 Punkte)

Geben Sie die Namen aller Produkte aus, die am 04.02.2004 von den Bändern in Halle 2 gelaufen sind.

(b) (4 Punkte)

Welche Stückzahl (d.h. wieviele Einheiten) hat das Band mit der FID=25 im Oktober 2003 produziert?

Fließband(FID, Ort)  
Produkt(PID, Name, QCheck)  
hergestellt(FID, PID, Datum, Uhrzeit)

---

(c) (6 Punkte)

Gibt es Fließbänder, die am 21.01.2004 zwischen 10:00 und 12:00 Uhr stillstanden (d.h. nichts produziert haben)?

Fließband(FID, Ort)  
Produkt(PID, Name, QCheck)  
hergestellt(FID, PID, Datum, Uhrzeit)

---

(d) (9 Punkte)

Welches Band liefert die qualitativ hochwertigsten Produkte (d.h. Anzahl der produzierten Einheiten mit QCheck=1 geteilt durch die Gesamtanzahl produzierter Einheiten für dieses Band ist maximal)?



5. Gegeben seien die Relationen  $R$ ,  $S$  und  $T$  mit folgendem Inhalt:

$R$			
$A$	$B$		
x	1		
x	2		
x	3		
y	1		
y	3		
y	4		
z	1		
z	2		
z	4		

$S$	$T$
$B$	$B$
1	1
2	4

(a) (2 Punkte)

Was ist das Ergebnis von  $R \div S$ ?

(b) (2 Punkte)

Was ist das Ergebnis von  $R \div T$ ?

6. Die Produktionsprozessdatenbank aus der vorletzten Aufgabe wird folgendermaßen erweitert. Jedes Produkt wird nun einer Charge zugeteilt:

- Produkt(PID, Name, QCheck, PCGID)

Außerdem soll in einer Tabelle *Charge* die Anzahl der Produkte die zu einer Charge gehören materialisiert werden (Chargen können dabei nur aus gleichen Produkten bestehen, PName steht für den Produktnamen):

- Charge(CGID, PName, Anzahl)

Da es Änderungen bei den Produktnamen geben kann, wird folgender Trigger definiert, um die Änderungen automatisch zu propagieren (bei Änderung eines Produktnamens werden alle Produktnamen dieser Charge automatisch angepasst):

```
create trigger PName
after update on Produkt
for each row
begin
    update Charge
    set    PName = new.Name
    where CGID = new.PCGID;
end
```

Um Änderungen an der Chargennummer zu propagieren, wird dieser Trigger definiert:

```
create trigger CGID
after update on Charge
for each row
begin
    update Produkt
    set    PCGID = new.CGID
    where PCGID = old.CGID;
end
```

(a) (3 Punkte)

Zu welchem Problem können diese Triggerdefinitionen führen?

(b) (5 Punkte)

Wie können Sie die automatische Propagierung der Änderungen sinnvoll durchführen?

7. (a) (4 Punkte)

Welche der folgenden Eigenschaften (seriell, serialisierbar, rücksetzbar, ACA, strikt) besitzt diese Historie?

$r_1[x] w_1[y] r_2[y] w_1[z] w_2[z] c_1 w_2[x] c_2$

(b) (4 Punkte)

Welche der folgenden Eigenschaften (seriell, serialisierbar, rücksetzbar, ACA, strikt) besitzt diese Historie?

$r_1[x] r_2[z] w_2[x] w_2[z] c_2 w_1[z] c_1$

(c) (5 Punkte)

In einem Datenbanksystem wird striktes 2PL gefahren. Es wird festgestellt, dass die Transaktionen lange warten müssen. Bei einer näheren Betrachtung wird festgestellt, dass die meisten Transaktionen folgendes Aussehen haben:

$T_i : w_i[x] w_i[a_i] w_i[b_i] w_i[c_i] c_i$

Was können Sie tun, um die Nebenläufigkeit zu erhöhen?

8. (a) (6 Punkte)

Auswertungspläne, die nur Reihenfolgen von Joins berücksichtigen, werden *Left-Deep Trees* genannt (d.h. nur auf der linken Seite eines Joinoperators darf eine Zwischenergebnisrelation stehen, siehe auch Abbildung 1).

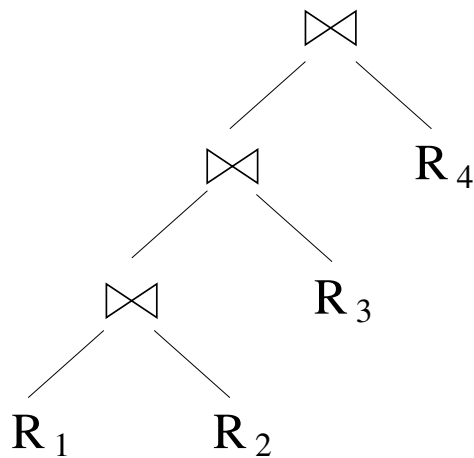


Abbildung 1: Left-Deep Tree

Wieviele verschiedene Left-Deep-Tree-Pläne gibt es, um die vier Relationen  $R_1, R_2, R_3$  und  $R_4$  zu joinen? (Gehen Sie davon aus, daß bei dem verwendeten Joinalgorithmus die Kosten für  $R_i \bowtie R_j$  gleich der Kosten für  $R_j \bowtie R_i$  sind und deswegen diese beiden Pläne als gleich betrachtet werden. Außerdem sind Kreuzprodukte erlaubt.)

(b) (8 Punkte)

Die Relation  $R$  enthält 6000 Tupel. Auf dieser Relation sollen nacheinander die Selektionen  $\sigma_1, \sigma_2, \sigma_3$  und  $\sigma_4$  ausgeführt werden. Da der Selektionsoperator kommutativ und assoziativ ist, spielt die Reihenfolge für das Ergebnis keine Rolle. Die Kosten (in Millisekunden) können für verschiedene Reihenfolgen aber durchaus verschieden sein. Tabelle 1 gibt die Selektivitäten und Auswertungskosten pro Tupel für jede Selektion an. Geben Sie die günstigste Auswertungsreihenfolge samt Kosten für die Gesamtauswertung an.

Selektion	Selektivität	Kosten pro Tupel
$\sigma_1$	$\frac{1}{2}$	5 ms
$\sigma_2$	$\frac{1}{20}$	1 ms
$\sigma_3$	$\frac{1}{15}$	2 ms
$\sigma_4$	$\frac{1}{10}$	5 ms

Tabelle 1: Selektivität und Kosten für Selektionen