

# Hauptdiplomklausur Datenbanksysteme I Sommersemester 2004

Name: .....  
Vorname: .....  
Matrikelnummer: .....  
Studienfach: .....

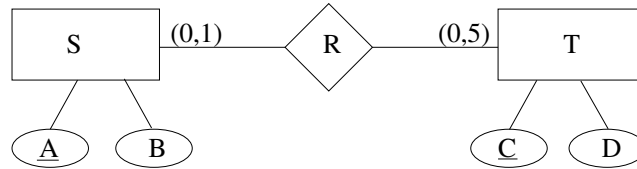
## Wichtige Hinweise:

1. Prüfen Sie Ihr Klausurexemplar auf Vollständigkeit (16 Seiten)
2. Es sind keine Hilfsmittel zugelassen.
3. Die Klausur dauert 100 Minuten.
4. Jede Aufgabe ist auf den zugehörigen Aufgabenblättern (und ggf. auf separaten Lösungsblättern) zu bearbeiten.
5. Vermerken Sie Ihren Namen und Matrikelnummer auf jedem Aufgaben- und Lösungsblatt. Blätter ohne Namens- und Matrikelangabe werden nicht bewertet.
6. Das Deckblatt sowie alle Aufgabenblätter sind abzugeben.

	maximale Anzahl Punkte	erreichte Anzahl Punkte
Aufgabe 1	9	
Aufgabe 2	23	
Aufgabe 3	22	
Aufgabe 4	15	
Aufgabe 5	11	
Aufgabe 6	12	
Aufgabe 7	8	
	100	

1. (a) (4,5 Punkte)

Für das folgende ER-Diagramm sind alternative Abbildungen auf ein relationales Schema gegeben. Nennen Sie für jede der Alternativen die Probleme dieser Abbildung.



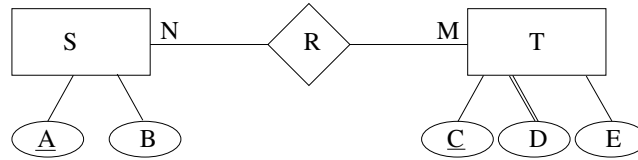
i. S: {[A, B, C]}  
T: {[C, D]}

ii. S: {[A, B]}  
T: {[C, D, A]}

iii. S: {[A, B]}  
T: {[C, D]}  
R: {[ $A_1, A_2, A_3, A_4, A_5, \underline{C}$ ]}  
in Relation R enthalten die Attribute  $A_i$  Werte von Attribut A aus Relation S.

(b) (4,5 Punkte)

Für das folgende ER-Diagramm sind alternative Abbildungen auf ein relationales Schema gegeben. Nennen Sie für jede der Alternativen die Probleme dieser Abbildung.



i. S: {[A, B, C]}  
T: {[C, D, E]}

ii. S: {[A, B]}  
T: {[C, D, E]}  
R: {[A, C]}

iii. S: {[A, B]}  
T: {[C, E]}  
R: {[A, C]}  
T': {[C, D]}

2. (a) (3 Punkte)

Welche Eigenschaften gelten für den Semi-Join?

	Ja	Nein
$ A \bowtie B  \leq  A \times B $		
$A \bowtie B = B \bowtie A$		
$A \bowtie B = \Pi_A(A \bowtie B)$		
$A \bowtie (B \bowtie C) = (A \bowtie B) \bowtie C$		
$\sigma_p(A \bowtie B) = \sigma_p(A) \bowtie B$		
$\sigma_p(A \bowtie B) = A \bowtie \sigma_p(B)$		

---

Gegeben sei folgende SQL-Anfrage:

```

SELECT DISTINCT x.a FROM R AS x
WHERE EXISTS ( SELECT * FROM S AS y
                WHERE y.c = 5
                AND x.a = y.b
                AND EXISTS ( SELECT * FROM T z
                            WHERE y.c = z.d ) )

```

---

(b) (4 Punkte)

Geben Sie einen äquivalenten Ausdruck in relationaler Algebra an.

(c) (3 Punkte)

Geben einen äquivalenten Ausdruck im Tupelkalkül an.

(d) Gegeben sei folgende SQL-Anfrage:

```
SELECT DISTINCT R.a
FROM R, S
WHERE R.a = S.b
      AND S.c = 5
```

i. (3 Punkte)

Überführen Sie die SQL-Anfrage in einem kanonischen Operatorbaum.

ii. (7 Punkte)

Optimieren Sie die SQL-Anfrage, so daß die Anzahl der gelesenen Tupel minimiert wird. Nehmen Sie für Joins eine Nested-Loop-Implementierung an. Für die Optimierung sind folgende Informationen gegeben:

- $|R| = 1.000$
- $|S| = 20.000$
- $sel(R.a) = 1/500$
- $sel(S.b) = 1/20$
- $sel(S.c) = 1/100$
- $sel(R \bowtie S) = 1/200$

Wieviele Tupel werden im optimierten Plan gelesen?

(e) (3 Punkte)

Welches oberstes Optimierungskriterium wird in verteilten Datenbanksystemen verwendet? Warum ist das ein sinnvolles Optimierungskriterium?

3. Gegeben sei folgendes relationales Schema für die Verwaltung einer Konferenzdatenbank.

Autor    {[Email, Name, Adresse, Institut]}

Referee  {[Email, Name]}

Paper    {[PaperID, Titel, POrt, PZeit, präsentiertVon]}

schreibt {[T\_Email, PaperID]}

bewertet {[R\_Email, PaperID, Bewertung]}

*Autoren* der Konferenz *schreiben* Beiträge (*Paper*) für die Konferenz. Einer der Autoren eines Papers präsentiert sein Paper (Fremdschlüssel *präsentiertVon*) auf der Konferenz. Die Präsentationen finden zu einer bestimmten Zeit an einem bestimmten Ort statt. Für Paper, die auf der Konferenz präsentiert werden, ist die Kombination von *POrt*, *PZeit* und *präsentiertVon* eindeutig. Alle anderen Paper (die abgelehnten Paper) haben NULL-Werte für den Ort, die Zeit und das Attribut *präsentiertVon*. Jedes eingereichte Paper wird durch eine feste Anzahl *Referees* *bewertet*. Je höher der Wert einer Bewertung ist, desto besser schätzt ein Referee das Paper ein.

Formulieren Sie die folgenden Anfragen in SQL.

- (a) (3 Punkte)

Geben Sie eine Liste der Titel aller Paper (auch diejenigen, die nicht präsentiert werden) zusammen mit allen Autoredaten an. Sortieren Sie die Liste absteigend nach PaperID.

Autor	{[ <u>Email</u> , Name, Adresse, Institut]}
Referee	{[ <u>Email</u> , Name]}
Paper	{[ <u>PaperID</u> , Titel, POrt, PZeit, präsentiertVon]}
schreibt	{[ <u>T_Email</u> , PaperID]}
bewertet	{[ <u>R_Email</u> , PaperID, Bewertung]}

---

(b) (6 Punkte)

Geben Sie den Titel sowie den Ort und die Zeit der Präsentation des Papers mit der besten Einzelbewertung an. Geben Sie auch den Namen des Autors an, der das Paper präsentiert, und die Email-Adresse des Referees an, der die Bewertung vergeben hat.

(c) (6 Punkte)

Geben Sie die Differenz zwischen der durchschnittlichen Bewertung der präsentierten Paper und den nicht präsentierten Paper an.



Autor    {[Email, Name, Adresse, Institut]}

Referee  {[Email, Name]}

Paper    {[PaperID, Titel, POrt, PZeit, präsentiertVon]}

schreibt {[T\_Email, PaperID]}

bewertet {[R\_Email, PaperID, Bewertung]}

---

(d) (7 Punkte)

Welches Institut hat die beste durchschnittliche Bewertung für alle seine Paper?

4. Gegeben sei ein Relationenschema  $\mathcal{R}(A, B, C, D, E, F)$  mit funktionalen Abhängigkeiten  $\mathcal{F}_{\mathcal{R}}$ . Lösen Sie in jeder Teilaufgabe folgende Teilprobleme:

- Geben Sie alle Kandidatenschlüssel von  $R$  an.
- In welcher höchsten Normalform ist  $R$ ?
- Zerlegen Sie  $R$  verlustlos in Teilrelationen, so dass die Teilrelationen in der nächst höheren Normalform sind. Verwenden Sie als Grundlage für die Zerlegung die gegebenen funktionalen Abhängigkeiten. Wenn  $R$  bereits in der höchsten bekannten Normalform ist, machen Sie deutlich, daß keine Zerlegung nötig ist.

(a) (5 Punkte)

$$\mathcal{F}_{\mathcal{R}} = \{D \rightarrow E, EF \rightarrow ABCD\}.$$

(b) (5 Punkte)

$$\mathcal{F}_{\mathcal{R}} = \{AB \rightarrow C, A \rightarrow D, DE \rightarrow F, E \rightarrow B\}.$$

- Geben Sie alle Kandidatenschlüssel von  $R$  an.
  - In welcher höchsten Normalform ist  $R$ ?
  - Zerlegen Sie  $R$  verlustlos in Teilrelationen, so dass die Teilrelationen in der nächst höheren Normalform sind. Verwenden Sie als Grundlage für die Zerlegung die gegebenen funktionalen Abhängigkeiten. Wenn  $R$  bereits in der höchsten bekannten Normalform ist, machen Sie deutlich, daß keine Zerlegung nötig ist.
- 

(c) (5 Punkte)

$$\mathcal{F}_R = \{A \rightarrow E, B \rightarrow ADF, C \rightarrow B, D \rightarrow C, D \twoheadrightarrow F\}.$$



(b) (9 Punkte)

Geben Sie für jede Historie an welche Eigenschaft aus den Tabellen sie besitzt und welche nicht.

i.  $r_1[x], r_2[x], w_1[x], w_2[y], r_1[y], c_2, a_1$

	Ja	Nein
serialisierbar		
rücksetzbar		
vermeidet kask. Rückts.		
strikt		

ii.  $w_1[x], w_2[y], w_3[x], c_3, r_2[x], c_2, w_1[y], c_1$

	Ja	Nein
serialisierbar		
rücksetzbar		
vermeidet kask. Rückts.		
strikt		

iii.  $r_1[x], w_2[x], w_1[y], c_1, r_3[y], c_3, w_4[y], c_2, a_4$

	Ja	Nein
serialisierbar		
rücksetzbar		
vermeidet kask. Rückts.		
strikt		

6. In einem Datenbanksystem werden die drei Transaktionen  $T_1, T_2$  und  $T_3$  verzahnt ausgeführt. Die (schreibenden) Zugriffe der Transaktionen auf die Datenelemente  $A, B, C, D$  und  $E$  werden in einem Log-File mitprotokolliert. In Schritt 12 wurde ein Sicherungspunkt ausgeführt bei dem alle Änderungen in die Datenbasis eingebracht wurden.

Schritt	$T_1$	$T_2$	$T_3$	Log-Datei
				[LSN, TA, Redo, Undo, PreviousLSN]
1.	BOT			[#1, $T_1$ , BOT, -, 0]
2.	$r(B, b_1)$			
3.	$b_1 = b_1 - 5$			
4.	$w(B, b_1)$			[#2, $T_1$ , $B = B - 5, B = B + 5$ , #1]
5.	$r(A, a_1)$			
6.		BOT		[#3, $T_2$ , BOT, -, 0]
7.		$r(B, b_2)$		
8.			BOT	[#4, $T_3$ , BOT, -, 0]
9.			$r(A, a_3)$	
10.			$a_3 = a_3 + 10$	
11.			$w(A, a_3)$	[#5, $T_3$ , $A = A + 10, A = A - 10$ , #4]
12.	Sicherungspunkt			[#6, $T_1, T_2, T_3$ , Sicherungspunkt, #1 ]
13.			Commit	[#7, $T_3$ , Commit, -, #5]
14.		$b_2 = b_2 + 12$		
15.		$w(B, b_2)$		[#8, $T_2$ , $B = B + 12, B = B - 12$ , #3]
16.	Abort			[#9, $T_1$ , Abort, -, #2]

(a) (3 Punkte)

Geben Sie eine Definition für WAL an. Trägt das Anwenden von WAL zur Dauerhaftigkeit (Durability) in DBMS bei? Begründen Sie Ihre Antwort kurz.

(b) i. (1 Punkt)

Welche Art von Sicherungspunkt wurde verwendet?

ii. (2 Punkte)

Wozu werden die Informationen im Logeintrag des Sicherungspunkts benötigt?

iii. (2 Punkte)

Welche Auswirkungen hat der Sicherungspunkt aus (b) für die Recovery?

(c) (4 Punkte)

Welche Operationen führt das Datenbanksystem aus, nachdem Transaktion  $T_1$  ein Abort ausführt?

## 7. (8 Punkte)

Gegeben sei ein verteiltes Datenbanksystem. Die Datensätze A, B, C und D sollen den Stationen  $S_1$ ,  $S_2$  und  $S_3$  des VDBMS zugeordnet werden. Auf den Stationen  $S_j$  werden die Transaktionen  $T_1$ ,  $T_2$  und  $T_3$  ausgeführt.

Die linke Tabelle gibt an, welche Daten wie oft von den drei Transaktionen verwendet werden. Die rechte Tabelle enthält, wie oft jede Transaktion auf welcher Station ausgeführt wird.

Daten	Transaktion		
	$T_1$	$T_2$	$T_3$
A	5		2
B		4	
C		2	3
D	1		2

Station	Transaktion		
	$T_1$	$T_2$	$T_3$
$S_1$	1	3	2
$S_2$	2		5
$S_3$		7	2

Ein Datensatz wird auf der Station allokiert, auf der die meisten lokalen Referenzen – Anfragen und Updates – erfolgen.

Wie sollen die Datensätze auf die Stationen verteilt werden?