

# Hauptdiplomklausur

## Datenbanksysteme I

### Sommersemester 2001

Name: .....

Vorname: .....

Matrikelnummer: .....

Studienfach: .....

Wichtige Hinweise:

1. Prüfen Sie Ihr Klausurexemplar auf Vollständigkeit (13 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. Lösungsblatt). Blätter ohne Namens- und Matrikelnummerangabe werden nicht bewertet.
6. Das Deckblatt sowie alle Aufgabenblätter (evtl. Lösungsblätter) sind abzugeben.

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

1. (je 1 Punkt)

Geben Sie für folgende Aussagen an, ob sie wahr oder falsch sind.

**Bei einer falschen Antwort werden Punkte abgezogen, die Gesamtpunktzahl kann jedoch nicht unter 0 Punkte sinken.**

	wahr	falsch
Das Generalisierungskonstrukt im E-R-Modell kann direkt ins relationale Modell übertragen werden, nicht jedoch ins objekt-orientierte Modell.	<input type="checkbox"/>	<input type="checkbox"/>
Die UndoNextLSN-Einträge werden nur in Compensation Log Records (CLRs) angelegt.	<input type="checkbox"/>	<input type="checkbox"/>
Folgende Äquivalenz ist allgemeingültig: $\Pi_{A_1, A_2, \dots, A_n}(R_1 \bowtie R_2) = \Pi_{A_1, A_2, \dots, A_n}(R_1) \bowtie \Pi_{A_1, A_2, \dots, A_n}(R_2)$	<input type="checkbox"/>	<input type="checkbox"/>
Sichten, in denen eine distinct-Klausel vorkommt, sind update-fähig	<input type="checkbox"/>	<input type="checkbox"/>

2. (a) (5 Punkte)

Gegeben sei ein Relationenschema  $\mathcal{R}(A, B, C, D, E, F, G)$  mit folgenden funktionalen Abhängigkeiten  $\mathcal{F}_{\mathcal{R}} = \{A \rightarrow B, B \rightarrow CD, C \rightarrow EF, D \rightarrow FG\}$ . In welcher höchsten Normalform befindet sich  $\mathcal{R}$ ?

(b) (5 Punkte)

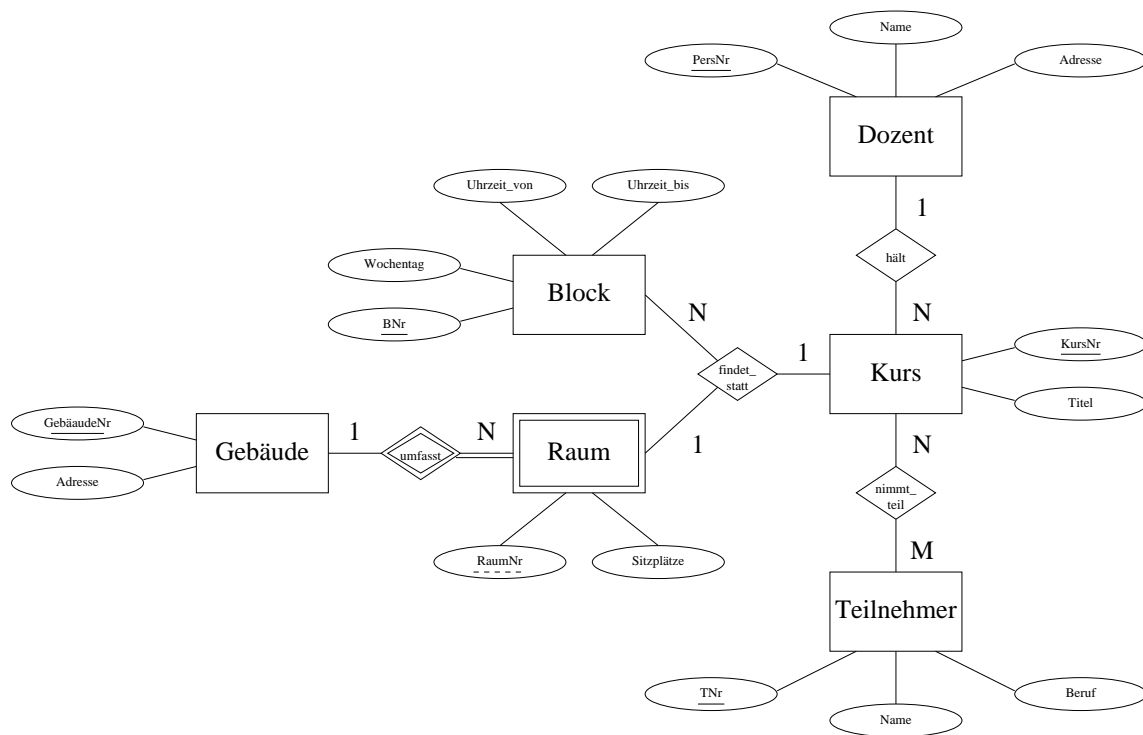
Gegeben sei ein Relationenschema  $\mathcal{S}(A, B, C, D, E, F)$  mit folgenden funktionalen Abhängigkeiten  $\mathcal{F}_{\mathcal{S}} = \{A \rightarrow F, B \rightarrow D, C \rightarrow B, CDF \rightarrow A, D \rightarrow E\}$ . In welcher höchsten Normalform befindet sich  $\mathcal{S}$ ?

(c) (5 Punkte)

Gegeben sei ein Relationenschema  $\mathcal{T}(A, B, C, D, E, F)$  mit folgenden funktionalen/mehrwertigen Abhängigkeiten  $\mathcal{F}_{\mathcal{T}} = \{A \rightarrow B, B \rightarrow C, CD \rightarrow A, D \rightarrow EF, E \rightarrow B, B \twoheadrightarrow C\}$ . In welcher höchsten Normalform befindet sich  $\mathcal{T}$ ?

3. (8 Punkte)

Für die Verwaltung des Kursangebots einer Fortbildungsakademie wurde folgendes konzeptuelles Datenbankschema entworfen:



Setzen Sie dieses E-R-Diagramm in ein relationales Schema um.

4. Ein Konferenzveranstalter hält seine Daten in einem relationalen Datenbanksystem mit folgendem Schema:

- Konferenz(Name, Ort, Anfangsdatum, Enddatum)
- Veranstaltung(VID, Titel, Datum, Uhrzeitvon, Uhrzeitbis, Konferenzname)
- Teilnehmer(PersNr, Name, Adresse, email)
- nimmt\_teil(PersNr, Konferenz, Status)
- besucht(PersNr, VID)

Die Relation *Konferenz* enthält allgemeine Daten zur Konferenz. Jede Konferenz besteht aus einer Reihe von Einzelveranstaltungen, die in der Relation *Veranstaltung* abgespeichert sind (gehen Sie davon aus, daß sich eine Einzelveranstaltung nicht über mehrere Tage hinziehen kann). *Teilnehmer* verwaltet die Daten der teilnehmenden Personen, während *nimmt\_teil* und *besucht* Auskunft darüber geben, welche Konferenzen bzw. Einzelveranstaltungen ein Teilnehmer besucht.

Formulieren Sie folgende Anfragen in SQL:

(a) (6 Punkte)

Geben Sie die Titel aller Veranstaltungen an, die am zweiten Tag der Konferenz "ICDE2001" nachmittags (Beginn nach 12:00 Uhr) stattfinden. (Sie dürfen dabei mit Datumsangaben ganz normal rechnen, also z.B. '30.09.2001' + 1 = '01.10.2001'.)

Konferenz(Name, Ort, Anfangsdatum, Enddatum)

Veranstaltung(VID, Titel, Datum, Uhrzeitvon, Uhrzeitbis, Konferenzname)

Teilnehmer(PersNr, Name, Adresse, email)

nimmt\_teil(PersNr, Konferenz, Status)

besucht(PersNr, VID)

---

(b) (7 Punkte)

Geben Sie die PersNr und Namen der Personen an, die die Veranstaltung 'Web DB' besucht haben, aber nicht die Veranstaltung 'DB and Internet'.

Konferenz(Name, Ort, Anfangsdatum, Enddatum)

Veranstaltung(VID, Titel, Datum, Uhrzeitvon, Uhrzeitbis, Konferenzname)

Teilnehmer(PersNr, Name, Adresse, email)

nimmt\_teil(PersNr, Konferenz, Status)

besucht(PersNr, VID)

---

(c) (9 Punkte)

Welche von den Konferenzen, die in New York stattfanden, hatte die größte Teilnehmerzahl?

5. Zwischen den Relationen  $R(\underline{\alpha}, \beta)$  und  $S(\underline{\gamma}, \delta)$  existiert eine Fremdschlüsselbeziehung:

```

create table R (
    α primary key
    β
    ...
)
create table S (
    γ primary key
    δ references R on α
    ...
)

```

Die Relationen  $R$  und  $S$  haben folgenden Inhalt:

$R$	
$\alpha$	$\beta$
$a_1$	$b_1$
$a_2$	$b_2$
$a_3$	$b_3$

$S$	
$\gamma$	$\delta$
$c_1$	$a_1$
$c_2$	$a_1$
$c_3$	$a_2$

Folgende Anfragen führen zu den folgenden Ergebnissen:

(a) (4 Punkte)

Anfrage

Ergebnis

delete from  $R$   
where  $\alpha = a_1$

$R$	
$\alpha$	$\beta$
$a_2$	$b_2$
$a_3$	$b_3$

$S$	
$\gamma$	$\delta$
$c_3$	$a_2$

Anfrage

Ergebnis

update  $R$   
set  $\alpha = a_5$   
where  $\alpha = a_2$

$R$	
$\alpha$	$\beta$
$a_1$	$b_1$
$a_5$	$b_2$
$a_3$	$b_3$

$S$	
$\gamma$	$\delta$
$c_1$	$a_1$
$c_2$	$a_1$
$c_3$	NULL

Welche weiteren Klauseln stehen in diesem Fall in der `create table S` Anweisung?



(b) (4 Punkte)

Anfrage

delete from  $R$   
where  $\alpha = a_1$

Ergebnis

$R$	
$\alpha$	$\beta$
$a_1$	$b_1$
$a_2$	$b_2$
$a_3$	$b_3$

$S$

$\gamma$	$\delta$
$c_1$	$a_1$
$c_2$	$a_1$
$c_3$	$a_2$

Anfrage

update  $R$   
set  $\alpha = a_5$   
where  $\alpha = a_2$

Ergebnis

$R$	
$\alpha$	$\beta$
$a_1$	$b_1$
$a_5$	$b_5$
$a_3$	$b_3$

$S$

$\gamma$	$\delta$
$c_1$	$a_1$
$c_2$	$a_1$
$c_3$	$a_5$

Welche weiteren Klauseln stehen in diesem Fall in der `create table S` Anweisung?

(c) (4 Punkte)

Was passiert beim Ausführen der Anweisung: *insert into S values (c<sub>4</sub>, a<sub>4</sub>);*?  
Ist das Resultat abhängig von den Klauseln in Teil (a) und (b)?

6. Für die Relation  $R$ , die auf 182 hintereinanderliegenden Blöcken auf der Platte gespeichert ist, wurde ein geballter  $B^+$ -Baum (clustered  $B^+$ -tree) auf dem Schlüsselattribut  $A$  mit der Höhe 2 angelegt (siehe Abbildung 1).

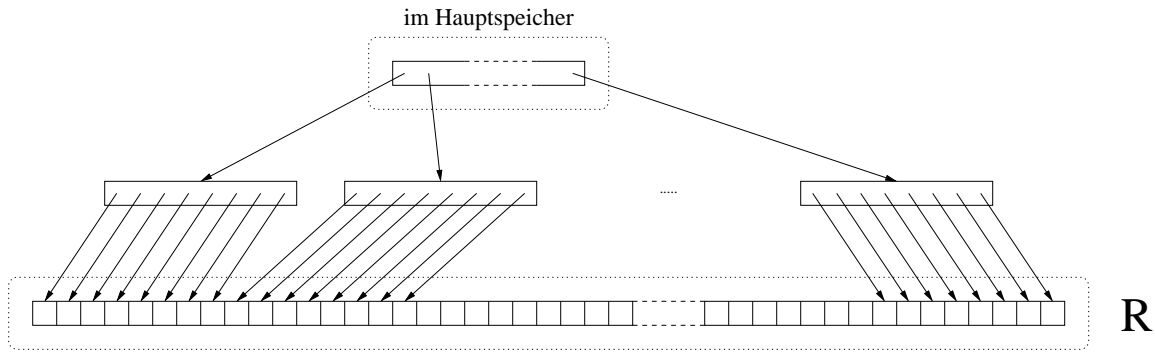


Abbildung 1: geclusterter  $B^+$ -Baum

Für einen Zugriff auf einen bestimmten Block auf einer Festplatte werden drei Arbeitsvorgänge benötigt. Zunächst muß der Schreib-/Lesekopf auf die entsprechende Spur plaziert werden. Die dazu benötigte Zeit wird als *Seek Time* ( $t_s$ ) bezeichnet. Dann wird gewartet, bis sich durch die Rotation der Platte der gesuchte Block am Kopf vorbeibewegt. Die zu erwartende Verzögerung wird *Latenzzeit* ( $t_l$ ) genannt. Im dritten Schritt wird der Block gelesen (*Lesezeit*,  $t_t$ ). Für diese Aufgabe gehen Sie von folgenden Daten für die Platte aus:

Seek Time ( $t_s$ )	12ms
Latenzzeit ( $t_l$ )	6ms
Lese- oder Transferzeit ( $t_t$ )	2ms

Bei einem sequentiellen Scan durch die Relation muß einmal gesucht werden, einmal die Latenzzeit abgewartet werden und 182 Blöcke in den Hauptspeicher transferiert werden.

(a) (6 Punkte)

Es werden Anfragen folgender Form an die Datenbank gestellt:

```
select *
from R
where A=23
or    A=19828
or    A=125
or    ...;
```

Ab wieviel verschiedenen gesuchten Werten für  $A$  ist ein sequentieller Scan genauso schnell wie die Zugriffe über den Index?

Gehen Sie davon aus, daß jede Seite des  $B^+$ -Baums in einem Block liegt und zum Holen dieses Blocks das Suchen, die Latenz und die eigentliche Übertragung abgewartet werden muß. Aus Effizienzgründen wird der Wurzelknoten des

B<sup>+</sup>-Baums im Hauptspeicher gehalten, muß also nicht von der Platte geholt werden. Gehen Sie weiterhin davon aus, daß jeder der gesuchten Werte für  $A$  auf einer anderen Blattseite liegt.

(b) (6 Punkte)

Nun werden folgende Anfragen an die Datenbank gestellt:

```
select *  
from R  
where A > 23;
```

Bei dieser Art der Anfrage (Range-Query) kann nach dem Finden des Einstiegs-  
punktes in die Relation sequentiell weitergescannt werden. Wieviele Blöcke der  
Relation  $R$  müssen bei einer Range-Query durchlaufen werden, so daß der se-  
quentielle Scan genauso schnell ist wie der Zugriff über den Index?

7. Im folgenden sind verschiedene Eigenschaften für Historien angegeben. Geben Sie jeweils eine Historie mit höchstens sieben Operationen (aus  $\{r_i, w_i, c_i, a_i\}$ ) an, die alle angegebenen Eigenschaften erfüllt.
- (a) (4 Punkte)  
serialisierbar, nicht rücksetzbar
- (b) (4 Punkte)  
serialisierbar, rücksetzbar, vermeidet kein kaskadierendes Rücksetzen (nicht ACA)
- (c) (5 Punkte)  
strikt, nicht serialisierbar
- (d) (5 Punkte)  
serialisierbar, vermeidet kaskadierendes Rücksetzen (ACA), das Recoveryprotokoll ARIES mit physischer Protokollierung (Before- und After-Images) funktioniert nicht (ARIES ist das Recoveryprotokoll, das im Buch beschrieben wird.)
- (e) (5 Punkte)  
serialisierbar, aber nicht von einem 2PL-Scheduler erzeugbar

8. In einem verteilten Datenbanksystem werden die Relationen  $R$  und  $S$  mit Hilfe des Semi-Joins  $R \bowtie S$  verknüpft. Die Relation  $R$  befindet sich dabei auf der Station  $S_1$ , die Relation  $S$  auf der Station  $S_2$ :

$S_1$		$S_2$	
$R$		$S$	
A	B	B	C
$a_1$	$b_2$	$b_1$	$c_1$
$a_2$	$b_2$	$b_3$	$c_2$
$a_3$	$b_2$	$b_4$	$c_2$
$a_4$	$b_3$	$b_5$	$c_2$
$a_5$	$b_5$	$b_6$	$c_3$
$a_6$	$b_5$		

(a) (2 Punkte)

Wie sieht die Teilrelation aus, die von Station  $S_1$  nach Station  $S_2$  geschickt wird?

(b) (2 Punkte)

Wie sieht die Teilrelation aus, die von Station  $S_2$  nach Station  $S_1$  zurückgeschickt wird?