

# Hauptdiplomklausur Datenbanksysteme I Wintersemester 2005

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	14	
Aufgabe 2	18	
Aufgabe 3	22	
Aufgabe 4	15	
Aufgabe 5	18	
Aufgabe 6	13	
	100	

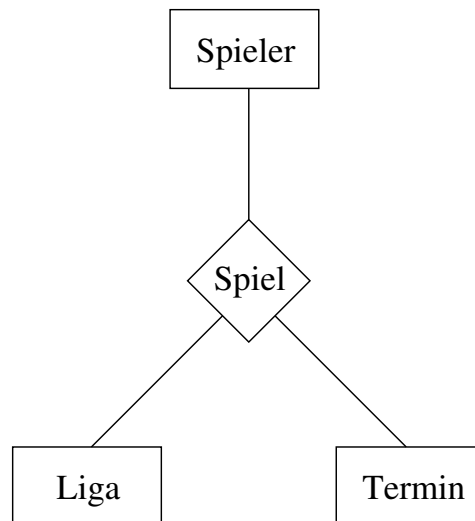
1. Erstellen Sie für folgende Szenarien je ein ER-Diagramm. Betrachten sie die Szenarien unabhängig voneinander.

(a) (6 Punkte)

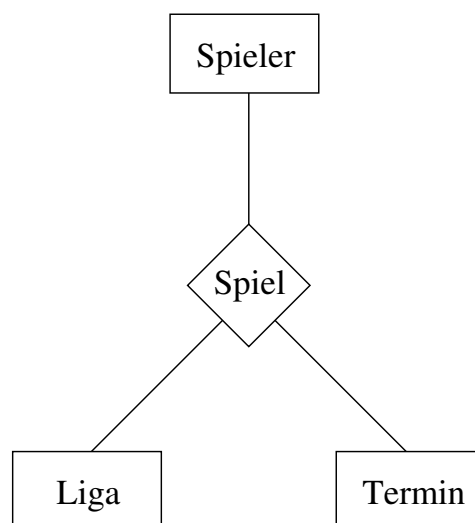
Gegeben sei eine 3-stellige Beziehung *Spiel*, die in der Modellierung für einen Tennisverein vorkommt. Die Beziehung besteht zwischen den Relationen *Spieler*, *Liga* und *Termin*. Dabei gelten folgende Annahmen:

- Ein Spieler spielt an einem Termin für genau eine Liga.
- An einem Spiel für eine Liga zu einem Termin sind zwei bis 4 Spieler beteiligt.
- Spieler spielen üblicherweise viele Spiele.
- Es gibt keine Spieler, Liga oder Termin ohne zugehöriges Spiel.

i. Tragen Sie in die folgende Abbildung die korrekten Funktionalitäten ein.



ii. Tragen Sie in die folgende Abbildung die korrekten Kardinalitäten ein.



(b) (5 Punkte)

In der Verwaltung der Vereinsmitglieder wird zwischen *Spielern*, *Schiedsrichtern* und Mitgliedern des *Vorstands* unterschieden. Für alle *Mitglieder* wird eine eindeutige *MitgliedsNr* vergeben. Außerdem wird für jedes Mitglied der *Name* und der *Mitgliedsbeitrag* verwaltet. Spezifisch für Spieler ist die *Position* innerhalb seiner Mannschaft (z.B. Position 1, 2, ... 6). Für Schiedsrichter wird festgelegt, in welchen *Ligen* sie tätig sein können, und für Mitglieder des *Vorstands* wird ein *Arbeitsbereich* angegeben.

Modellieren Sie den Sachverhalt in einem ER-Diagramm.

(c) (3 Punkte)

Für die Berechnung der Tabelle einer Liga werden Spielstände verwaltet. Jedes *Spiel* besteht aus mehreren *Sätzen*. Für jeden Satz wird die *Satznummer*, der *Punktstand* und die *Spielminuten* für diesen Satz gespeichert. Bei Spielen wird das *Datum* gespeichert, an dem das Spiel stattfand. Spiele werden über eine eindeutige *ID* identifiziert. Die Spiele und die Spielstände werden erst nach der Austragung eines Spiels in die Datenbank eingetragen.

Modellieren Sie den Sachverhalt in einem ER-Diagramm. Geben Sie Kardinalitäten bei der Beziehung zwischen Spielen und Sätzen an.

2. (a) (3 Punkte)

Welche der folgenden Designentscheidungen werden im physischen Datenbankentwurf getroffen.

	Ja	Nein
B-Baum als Index auf ein Attribut		
Materialisieren und Ballung einer Beziehung zwischen 2 Relationen		
Einrichten einer Sicht über 2 Relationen		
Überführen einer Relation in 3. Normalform		
Bestimmen des Schlüssels für eine Relation		
komprimierte Speicherung eines Attributs		

(b) (3 Punkte)

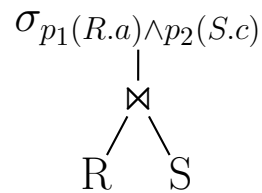
Gegeben sei folgende SQL-Anfrage:

```
SELECT R.a, S.d, T.e
FROM R, S, T
WHERE R.a = S.b
      AND T.e = S.d
      AND S.c = 3
```

Überführen Sie die SQL-Anfrage in einen kanonischen Operatorbaum.

(c) (12 Punkte)

Gegeben sei folgender Operatorbaum.



Dabei gelten folgende Parameter:

$$\begin{array}{llll}
 |R| = 10.000 & sel(R \bowtie S) = \frac{1}{10.000} & cost(R_1 \bowtie R_2) = |R_1| \cdot |R_2| \\
 |S| = 1.000 & sel(\sigma_{p_1(R.a)}) = \frac{1}{50} & cost(\sigma_{p_1(R.a)}(R_1)) = 100 \cdot |R_1| \\
 & sel(\sigma_{p_2(S.c)}) = \frac{1}{200} & cost(\sigma_{p_2(S.c)}(R_1)) = 1 \cdot |R_1|
 \end{array}$$

$R_1$  und  $R_2$  sind Teilausdrücke, die die notwendigen Variablenbindungen bereitstellen.

Finden Sie heraus, ob "Pushen" von Selektionen bei den gegebenen Daten sinnvoll ist, in dem Sie die Anordnung der Selektionen mit minimalen Kosten berechnen.

---

$$\begin{array}{llll} |R| = 10.000 & sel(R \bowtie S) = \frac{1}{10.000} & cost(R_1 \bowtie R_2) = |R_1| \cdot |R_2| \\ |S| = 1.000 & sel(\sigma_{p_1(R.a)}) = \frac{1}{50} & cost(\sigma_{p_1(R.a)}(R_1)) = 100 \cdot |R_1| \\ & sel(\sigma_{p_2(S.c)}) = \frac{1}{200} & cost(\sigma_{p_2(S.c)}(R_1)) = 1 \cdot |R_1| \end{array}$$

---

3. Gegeben sei folgendes relationales Schema für ein Weingut.

Rebsorte    {[Sorte, Name, Farbe]}  
 Wein        {[ID, Name, RSorte]}  
 Jahrgang    {[WeinID, Jahr, Preis, Qualität]}

In der Datenbank werden Weine verschiedener *Rebsorten* verwaltet. Die verschiedenen Weine werden in der Tabelle *Wein* verwaltet. Jeder Wein ist über den Fremdschlüssel *RSorte* genau einer Rebsorte zugeordnet. Für jeden *Jahrgang* wird für jeden Wein ein Preis und die Qualität vermerkt. Die Qualitäten “exzellent”, “sehr gut” und “gut” usw. werden auf die Ordinalzahlen 1, 2, 3 usw. abgebildet. Im folgenden sind Beispielausprägungen der Tabellen zu finden.

Rebsorte		
Sorte	Name	Farbe
1	Merlot	Rot
2	Riesling	Weiß
3	Spätburgunder	Rot

Wein		
ID	Name	RSorte
1	Marbuzet	1
2	Dr. Bassermann-Jordan	2
3	Herrgottsacker	2

Jahrgang			
WeinID	Jahr	Preis	Qualität
2	2003	20.00	1
2	2004	14.00	2
3	2001	50.00	1

---

Formulieren Sie die folgenden Anfragen in SQL.

(a) (4 Punkte)

Geben Sie Namen, Preis und Qualität aller Riesling-Weine des Jahrgangs 1996 an.



Rebsorte {[Sorte, Name, Farbe]}  
Wein {[ID, Name, RSorte]}  
Jahrgang {[WeinID, Jahr, Preis, Qualität]}

---

(b) (5 Punkte)

Geben Sie für jeden Rotwein (Farbe="Rot") die durchschnittliche Qualität und die Anzahl der Jahre an, die dabei berücksichtigt wurden.

(c) (6 Punkte)

Geben Sie alle Jahrgänge des Weins mit dem Namen "Marbuzet" an, die einen maximalen Preis über alle Jahrgänge dieses Weins erzielt haben und nicht exzellente Qualität (Qualität=1) hatten.

Rebsorte {[Sorte], Name, Farbe}  
Wein {[ID], Name, RSorte}  
Jahrgang {[WeinID], [Jahr], Preis, Qualität}

---

(d) (7 Punkte)

Finden Sie für alle Weine der Rebsorte “Merlot” diejenigen Weine heraus, die im Jahr 2004 einen geringeren Preis erzielt haben als den Durchschnittspreis dieses Weins über alle Jahre vor 2004.

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 – mindestens aber in der 3. Normalform. 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)

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

(b) (5 Punkte)

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

- 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 – mindestens aber in der 3. Normalform. 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)

$$\mathcal{F}_R = \{AB \rightarrow CDEF, B \twoheadrightarrow EF\}.$$

## 5. (a) (12 Punkte)

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

- i.
- $w_1[x], w_2[y], w_3[x], c_3, r_2[x], c_2, r_1[y], c_1$

	Ja	Nein
serialisierbar		
rücksetzbar		
vermeidet kask. Rückts.		
strikt		

- ii.
- $r_1[y], w_1[x], w_2[y], w_3[x], r_2[x], c_3, w_2[x], c_2, c_1$

	Ja	Nein
serialisierbar		
rücksetzbar		
vermeidet kask. Rückts.		
strikt		

- iii.
- $r_2[x], w_1[y], w_1[x], c_1, r_3[y], w_3[y], c_3, w_2[y], c_2$

	Ja	Nein
serialisierbar		
rücksetzbar		
vermeidet kask. Rückts.		
strikt		

- (b) Betrachten Sie folgende Historie  $H$ , in der die Transaktionen  $T_1$  und  $T_2$  auf die Datenelemente  $x$  und  $y$  zugreifen.

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

- i. (3 Punkte)

Kann die Historie  $H$  von einem Scheduler erzeugt worden sein, der das Zwei-Phasen-Sperrprotokoll befolgt? Begründen Sie Ihre Antwort.

- ii. (3 Punkte)

Kann der Scheduler die Operationen in  $H$  so umordnen, daß eine äquivalente Folge von Operationen entsteht, die das strenge 2PL befolgt? Begründen Sie Ihre Antwort.

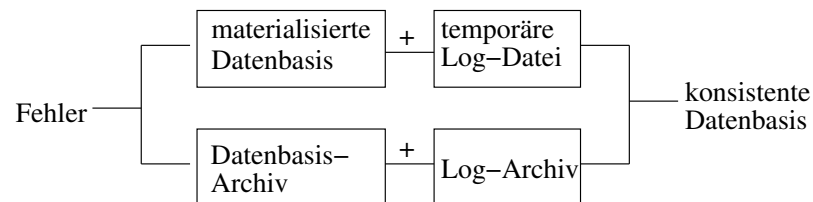
6. (a) (2 Punkte)

Definieren Sie, was Idempotenz der Recovery bedeutet.

(b) (2 Punkte)

Erläutern Sie, wie die Idempotenz in der Redo-Phase der Recovery gewährleistet wird.

(c) Nach einem Fehler muß die Recovery den letzten konsistenten Datenbankzustand wieder herstellen. In der folgenden Abbildung sind zwei Pfade angegeben, um dies zu erreichen.



i. (2 Punkte)

Unter welchen Umständen wird der obere Pfad durchlaufen?

ii. (2 Punkte)

Wann wird der untere Pfad durchlaufen?

In einem Datenbanksystem wird eine  $\neg force/\neg steal$  Strategie zum Einbringen von Änderungen in die Datenbasis verwendet.

(d) i. (2 Punkte)

Welche Operationen müssen unter diesen Voraussetzungen im Fehlerfall mit Verlust des Hauptspeicherinhalts ausgeführt werden? Begründen Sie kurz Ihre Antwort.

ii. (2 Punkte)

Welche Informationen müssen in den Log-Records der Log-Datei stehen, um eine korrekte Recovery zu gewährleisten?

iii. (1 Punkt)

Welche Informationen müssen in den CLR-Records der Log-Datei stehen, um eine korrekte Recovery zu gewährleisten?