Klausur zum Kurs 1665 Datenbanksysteme

Beachten Sie bitte die folgenden Hinweise:

- Die Bearbeitungszeit beträgt 3 Stunden.
- Es sind maximal 94 Punkte erreichbar.
- Als **Hilfsmittel** sind Schreibuntensilien und unbeschriebene Papierbögen zugelassen.
- Schreiben Sie mit Tinte oder Kugelschreiber.
- Füllen Sie das **Deckblatt** aus und heften Sie es zusammen mit der **Aufgabenstellung** vor ihre **Lösungsblätter**. Geben Sie alles zusammen nach Beendigung der Klausur ab.
- Schreiben Sie auf jedes Lösungsblatt oben rechts ihren Namen, ihre Matrikelnummer und ggf., zu welcher Aufgabe es gehört.
- Überprüfen Sie bitte die Vollständigkeit der Aufgabenstellungen: Die Klausur umfaßt das Deckblatt, eine Teilnahmebescheinigung, dieses Blatt und 8 Aufgaben.
- Die korrigierte Klausur erhalten Sie in 4 - 6 Wochen zusammen mit einer Musterlösung zurück.
- Sollten Sie eine Klausurteilnahmebescheinigung benötigen, füllen Sie bitte den entsprechenden Vordruck am Ende der Klausur aus und heften Sie ihn hinter das Deckblatt. Die Bescheinigung wird von der FernUniversität abgestempelt und Ihnen zusammen mit der korrigierten Klausur zugeschickt.

Wir wünschen Ihnen viel Erfolg!

Aufgabe 1: Datenbank, DBMS, DBS (7 Punkte)

Eine Datenbank ist eine integrierte Ansammlung von Daten, die allen Benutzern eines Anwendungsbereiches als gemeinsame Basis aktueller Information dient.

Erläutern Sie kurz die Schlüsselwörter *integriert* und *gemeinsame Basis* obiger Definition.

Definieren Sie kurz die folgenden Begriffe:

- Datenbankmanagementsystem (DBMS)
- Datenbanksystem (DBS)

Definieren Sie die folgenden Begriffe:

- Datenunabhängigkeit
- Integrität der Daten

Aufgabe 2: ER-Modellierung (16 Punkte)

Erstellen Sie ein ER-Diagramm entsprechend folgender Punkte:

- Eine Firma besteht aus mehreren Abteilungen (die Firma muss nicht modelliert werden). Jede Abteilung besitzt einen eindeutigen Abteilungsnamen und eine eindeutige Nummer; ferner besitzt jede einen Standort.

- Die Firma hat mehrere Angestellte, zu denen jeweils der Name, die eindeutige Sozialversicherungsnummer, die Adresse, das Gehalt und das Geburtsdatum abgelegt werden sollen.

- Ein Angestellter arbeitet für genau eine Abteilung. Für jeden Angestellten wird der direkte Vorgesetzte notiert.

- Jede Abteilung wird von einem bestimmten Angestellten der Firma geleitet. Jeder Angestellt leitet maximal nur eine Abteilung. Es soll jeweils das Datum, ab dem der Angestellte die Leitung übernommen hat, festgehalten werden.

- Eine Abteilung hat die alleinige Verantwortung für eine Reihe von Projekten, die jeweils durch ein Projektkürzel und eine Nummer identifiziert werden. Ferner hat ein Projekt ein Start- und Enddatum zur Kennzeichnung der Projektlaufzeit.

- Ein Angestellter kann an mehreren Projekten arbeiten, für die nicht notwendigerweise die Abteilung verantwortlich ist, der er zugeordnet ist. Es soll die Stundenzahl verfolgt werden, die ein Mitarbeiter an einem Projekt arbeitet.

Aufgabe 3: Formulierung von Queries (15 Punkte)

Eine Bibliothek habe zur Verwaltung Ihrer Buchausleihe folgende Relationenschemata definiert:

BUCH(B_NR, AUTOR, TITEL, VERLAG, ERSCHEINUNGSJAHR, STANDORT)
ENTLEIHER(E_NR, NAME, ANSCHRIFT)
BUCH_ENT(B_NR, E_NR, ENTLEIHDATUM)

Formulieren Sie die folgenden Anfragen. Geben Sie dabei nur die Anfragen, nicht aber die Ergebnisrelationen an:

a) Formulieren Sie in SQL: (3 Punkte)
Finde alle Autoren, deren Buch im gleichen Verlag erschienen ist wie das Buch 'Database Managementsystems' von R. Ramakrishnan.

b) Formulieren Sie in der Relationenalgebra: (3 Punkte)
Finde die Namen und Anschriften aller Entleiher, die nach dem 18.03.2005 ein Buch des Autos A. Silberschatz ausgeliehen haben.

c) Definieren Sie in SQL eine Sicht, die für jeden Standort die Anzahl der zugeordneten Bücher ausgibt. (2 Punkte)

d) Formulieren Sie in SQL: (3 Punkte)
Finde die Namen und Anschriften aller derjenigen Entleiher, die seit dem 15.02.2005 ein Buch aus dem 'W.H. Freeman & Co. Verlag' mehr ausgeliehen haben.

e) Formulieren Sie im Relationenkalkül: (4 Punkte)
Finde die Namen und Anschriften aller Entleiher, die ihre entliehenen Bücher an mehreren Tagen ausgeliehen haben.

Aufgabe 4: Normalformen (14 Punkte)

Gegeben sei das folgende relationale Datenbankschema mit den Relationen:

Einkaufs_Abt
(Abt_Name, Abt_Leiter, Produkt_Gruppe, EK_Lager, Lager_Leiter)

Bestellung
(Lieferanten_Bez, L_Anschrift, Waren_Bez, Produkt_Gruppe, Menge, Best_Datum)

Es gelten die folgenden funktionalen Abhängigkeiten:

Abt_Name	→ Abt_Leiter
Abt_Name	→ Produkt_Gruppe
Produkt_Gruppe	→ EK_Lager
EK_Lager	→ Lager_Leiter
Lieferanten_Bez	→ L_Anschrift
Lieferanten_Bez, Waren_Bez, Best_Datum	→ Menge
Waren_Bez	→ Produkt_Gruppe

a) Welchen Normalformen genügen die Relationen und welchen nicht (2NF, 3NF, BCNF)? Begründen Sie Ihre Antwort.
b) Welche Anomalien können beim Arbeiten mit der obigen Datenbank auftreten? Nennen sie jeweils ein kurzes Beispiel.
c) Erstellen Sie durch Zerlegung ein Relationenschema, dass in 3NF ist.

Aufgabe 5: Query-Optimierung (14 Punkte)

Im Skript werden einige Heuristiken eingeführt, mit denen die algebraische Optimierung von Anfragen durchgeführt werden kann. Gegeben sei das folgende Relationenschema für die Verwaltung von Ferienwohnungen:

FERIENWOHNUNG (**F_NR**, ORT, BESITZER, MIETPREIS)
MIETER (**M_NR**, NAME, ANSCHRIFT, BERUF)
BUCHUNG (**F_NR**, **M_NR**, VON, BIS, PERSONEN)

Die fettgedruckten Attribute geben die Schlüssel der jeweiligen Relationen an.

Gegeben sei der folgende Ausdruck:

MIETER [M_NR = M_NR] (BUCHUNG [F_NR, M_NR, PERSONEN]) [BERUF <> ‚Kaufmann'] [PERSONEN > 2] [F_NR = F_NR] FERIENWOHNUNG [BERUF = ‚Anwalt' v PERSONEN > 10] [BESITZER = ‚Klein'] [NAME = ‚Mayer' v NAME = ‚Müller'] [NAME, BERUF, PERSONEN, F_NR, BESITZER]

(a) Zeichnen Sie für den obigen Ausdruck den entsprechenden Operatorbaum (2 Punkte)

(b) Optimieren Sie den angegebenen Ausdruck mit Hilfe der Ihnen bekannten Regeln und geben Sie den optimierten Ausdruck (6 Punkte) und den zugehörigen Operatorbaum (2 Punkte) an.

(c) Beschreiben Sie nun zwei jetzt noch durchführbare Optimierungen auf der physischen Ebene. (4 Punkte)

Aufgabe 6: Serialisierbarkeit (12 Punkte)

Gegeben sei das Transaktionssystem T = (T1, T2, T3, T4). Weiterhin sei folgender Schedule gegeben:

T1	T2	T3	T4
R1(d)			
		R3(b)	
	R2(c)		
	R2(d)		
			R4(a)
		W3(b)	
			R4(b)
		R3(a)	
	W2(d)		
		R3(c)	
			W4(a)
R1(a)			
			R4(d)
			W4(b)
W1(a)			
		W3(c)	

(a) Erstellen Sie den Abhängigkeitsgraphen zum oben angegebenen Schedule. (5 Punkte)

(b) Warum ist der Schedule nicht serialisierbar? Begründen Sie Ihre Antwort bitte kurz! (1 Punkt)

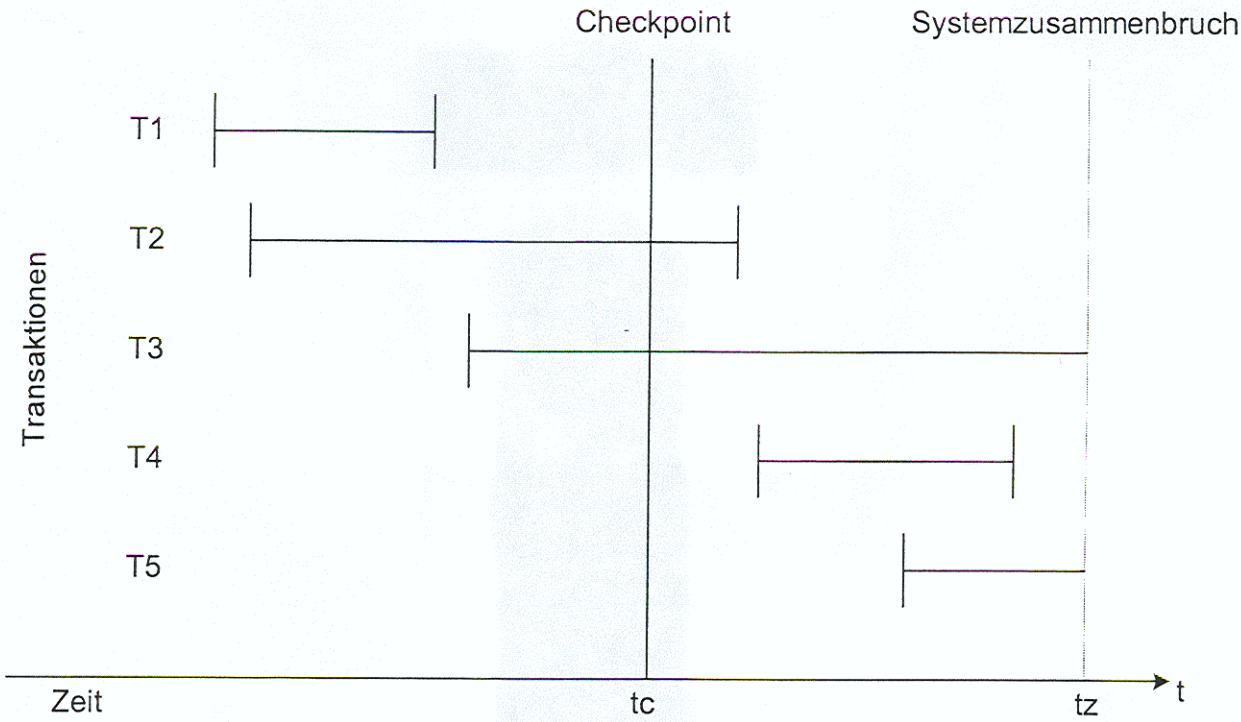
(c) Wieviele und welche Transaktionen müssen **mindestens** entfernt werden, damit das Transaktionssystem serialisierbar wird? (2 Punkte)

(d) Geben Sie eine serielle Ausführung an, zu dem der aus (c) resultierende (serialisierbare) Schedule äquivalent ist. (2 Punkte)

(e) Geben Sie ein Beispiel für ein Transaktionssystem, dass nicht zyklenfrei ist aber trotzdem serialisierbar. (2 Punkte)

Aufgabe 7: Recovery (10 Punkte)

Betrachten Sie die folgende Situation:



Zum Zeitpunkt tz ereignet sich ein Systemzusammenbruch. Der letzte Checkpoint wurde zum Zeitpunkt tc geschrieben.

(a) Wie sieht das Recovery für die obige Situation aus, wenn man davon ausgeht, dass das Schreiben eines Checkpoints das Herausschreiben aller veränderter Daten erzwingt? Was geschieht mit den einzelnen Transaktionen T1 bis T5? (5 Punkte)

(b) Was geschieht mit den einzelnen Transaktionen T1 bis T5, wenn das Herausschreiben der veränderten Daten nicht beim Schreiben des Checkpoints, sondern bei EOT geschieht? (5 Punkte)