

Klausur (Informatik)

Einführung in die Datenbanksysteme
 Datenbanken für die Bioinformatik
 Heinz Schweppe, Joos-Hendrik Böse

1	2	3	4	Σ
30	18	25	17	90

Name: _____

Matrikelnummer: _____

Studiengang: _____

Zu beachten:

- Namen und Matrikelnummer eintragen.
- Es sind keine Unterlagen erlaubt !
- Bitte sämtliche Lösungen auf die Blätter (auch Rückseite) der Klausur schreiben. Eigene Blätter sind nicht erlaubt.
- Täuschungsversuche jeglicher Art führen zum Nichtbestehen der Klausur (0 Punkte).
- Die Bearbeitungszeit für die Klausur beträgt 90 Minuten.

1. Aufgabe (Modellierung) 30 Punkte

Betrachten Sie folgende Beschreibung :

Ein Autohersteller, der nur einen einzigen Wagentyp herstellt (z.B. Golf), bietet seinen Kunden die individuelle Konfiguration ihres Neuwagens an. Eine Konfiguration besteht aus der Wahl einer Basisvariante und zusätzlichen Extras. Basisvarianten, z.B. „Golf Highline“, beschreiben den Motor und das Getriebe sowie die bereits serienmäßig enthaltenen Extras und haben natürlich auch einen Preis. Zu einer Basisvariante kann der Kunde nun weitere Extras hinzufügen, wie z.B. „Nebelscheinwerfer“ oder „Parkassistent“. Zu jeder Kundenkonfiguration wird das Erstellungsdatum benötigt und zu jedem Kunden Name und Email.

Ein Extra verfügt über einen Namen, eine Beschreibung sowie einen Preis. Es ist bei der Konfiguration zu beachten, dass bestimmte Extras nicht kompatibel mit jeder Basisvariante sind, z.B. kann das Extra „Tempomat“ nicht in die Basisvariante „Golf Comfort“ eingebaut werden.

Weiterhin verwaltet der Autohersteller Konfigurationen von Sondermodellen, die nicht von einem Kunden definiert worden sind, sondern in größeren Mengen für Händler vorproduziert werden. Ein Sondermodell besteht ebenfalls aus einer Basisvariante und zusätzlichen Extras. Z.B. basiert das Sondermodell „Golf Family“ auf der Basisvariante „Golf Comfort“ und verfügt zusätzlich zu den serienmäßigen Extras noch über die Extras „CD-Radio“ und „Kindersitz“. Alle Konfigurationen (Kundenkonfigurationen und Sondermodelle) haben eine eindeutige Nummer zur einfachen Identifizierung.

- Entwerfen Sie ein ER Modell für den oben beschriebenen Sachverhalt. Identifizieren Sie alle notwendigen Entitäten und Beziehungen. Tragen Sie die Kardinalitäten in der Ihnen aus der Veranstaltung bekannten Weise ein. Kennzeichnen Sie die Schlüssel der Entitäten. Fügen Sie keine künstlichen Schlüssel ein.
- Geben Sie das Relationenschema zum Entwurf an, fassen Sie die Relationen so weit wie möglich zusammen.

Benutzen Sie für Ihren Entwurf die letzte Seite, welche auch abgetrennt werden kann.
 Achten Sie darauf, dass die Seite mit Ihrer Matrikelnummer und Ihrem Namen beschriftet ist!

2. Aufgabe (Anfragen) 18 Punkte

Gegeben ist das folgende relationale Schema einer Katastrophendatenbank:

Ort(OrtID, Name, Land)

KatastrophenEreignis(OrtID, KatTypID, Datum, Beschreibung, Opfer)

KatastrophenArt(KatTypID, KatName)

Formulieren Sie die folgenden Anfragen in SQL:

1. Name und Land der Orte, in denen nie eine Katastrophe aufgetreten ist.
2. Der häufigste Katastrophentyp (Ausgegeben werden soll das Attribut KatName).
3. Alle Flutkatastrophen (KatName = „Flutkatastrophe“) im Jahr 2006 aufsteigend nach Datum geordnet.
4. Orte, an denen noch nie eine Atomkatastrophe aufgetreten ist, dafür aber eine Hungerkatastrophe.

Formulieren Sie folgende Anfragen in relationaler Algebra

5. Der Name des Katastrophentyps, der sich am 11.9.2001 in New York abgespielt hat.

6. Die Orte (Name), die von Godzilla oder King Kong (KatName) heimgesucht wurden.

3. Aufgabe 25 Punkte

Anmerkung: Bei den Multiple Choice Fragen der Klausur sind zum Teil auch mehrere Antworten möglich. Für jede richtig angekreuzte Aussage gibt es zwei Punkte, für falsch angekreuzte Aussagen einen Punkt Abzug.

- Wann kann es mehr als einen Primärindex geben ?
 - nie
 - wenn es mehr als einen Schlüsselkandidat gibt
 - immer

- Welcher wesentliche Unterschied besteht zwischen Primär- und Sekundärindex ? (ein Satz)

- Wie viel physische Schreiboperationen erfordert eine logische Schreiboperationen (Schreiben eines Datums) bei Raid-level 5 ?
 - einen Zugriff
 - zwei Zugriffe
 - drei Zugriffe

- Sequenzen (im Sinne von Oracle SEQUENCE) lassen sich mit Hilfe herkömmlicher Relationen emulieren. Dazu erstellt man eine Relation S, welche nur die aktuelle Sequenz-nummer hält. Erstellen Sie ein PL/SQL Programm (Pseudocode), welches die Erzeugung und der Lesen einer neuen Sequenznummer der Folge durchführt. Beachten Sie, dass die Erzeugung eines neuen Elements transaktionsgeschützt stattfinden muss.

- Entwerfen Sie eine Relation, die in 2. Normalform, aber nicht in 3. Normalform ist. Geben Sie die funktionalen Abhängigkeiten an.

- Betrachten Sie die folgende Instanz eines relationalen Schemas.

A	B	C
1	2	3
2	2	2
1	3	2
4	2	3

Enthält das Schema die funktionalen Abhängigkeiten $A \rightarrow B$ oder $BC \rightarrow A$?

- Ja
 Nein

- Eine SQL Anfrage besteht aus maximal 6 Klauseln (weitgehend unabhängigen Bestandteilen des Befehls), welche sind das ?
- Welches Problem kann bei der wechselseitigen referentiellen Integrität zwischen zwei Tabellen A und B auftreten und wie kann dieses gelöst werden ?

- Betrachten Sie das Schema $R(x)$ und die beiden Anfragen Q1 und Q2.

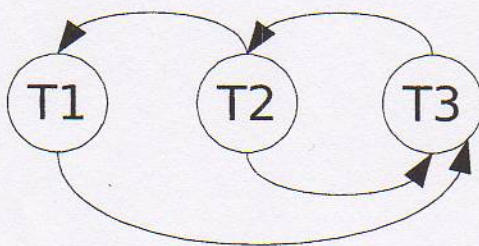
Q1: `SELECT x FROM R rr
WHERE NOT EXISTS (SELECT x FROM R WHERE x > rr.x);`

Q2: `SELECT MAX(x) FROM R;`

Kreuzen Sie die zutreffenden Aussagen an:

- Die Antwortmengen sind immer gleich
 - Die Antwortmenge von Q1 ist immer in der Antwort von Q2 enthalten.
 - Die Antwortmenge von Q2 ist immer in der Antwort von Q1 enthalten.
 - Die Antwortmengen sind immer verschieden
- In welcher Situation kann die *Thomas write rule* beim Timestamp Ordering Verfahren angewendet werden ?

- Betrachten Sie folgenden Konfliktgraphen:



Welche der folgenden Schedules entsprechen dem Konfliktgraphen ?

- $r1(y), w1(y), r3(y), w3(y), c3, r2(x), r1(x), w1(x), c1, r2(y), c2$
- $r3(c), w1(b), w3(c), r3(b), r1(d), r3(a), r2(b), c1, w3(b), w2(a), c3, c2$
- $r2(b), w1(b), w3(c), r3(b), r1(d), r3(a), r2(e), c1, w3(b), w2(a), c3, c2$

4. Aufgabe (Physisches Design) 17 Punkte

Betrachten Sie eine Festplatte mit einer Blockgröße von $b=512$ Bytes. Ein Blockpointer ist $pb = 6$ Bytes groß und ein Recordpointer $pr = 7$ Bytes. Die Relation $R(a,b,c,d)$ ist als Heap gespeichert und enthält 30.000 Records, wobei $c = 10$ Bytes groß ist. Es soll nun ein Index über das Attribut c mit Hilfe eines B+ Baumes (row-id Variante) erstellt werden. Gehen Sie im folgenden immer von 100% Füllgrad von inneren Knoten und Blattknoten aus. Berechnen Sie :

1. Den Verzweigungsgrad k des B+Baumes
2. Die Anzahl der Blätter des Baumes
3. Die Anzahl der Ebenen (ohne Blätter) in dem entstehenden B+ Baum
4. Die Gesamtzahl an Blocks, die vom B+ Baum benötigt werden.
5. Wie viele Blockzugriffe werden für eine Punktabfrage benötigt ?
6. Nehmen Sie an, der Wertebereich von Attribut c ist $[0;30.000]$ und jeder mögliche Wert in diesem Bereich kommt auch als Attributwert vor. Wie viele Blockzugriffe werden benötigt, um mit Hilfe einer Bereichsabfrage alle Records für $10.000 < c < 15.000$ zu lesen ?