

Datenbanksysteme (2h)	schriftliche Einzelprüfung	05.05.2008	1
-----------------------	----------------------------	------------	---

Aufgabe 1 [Relationale Abfragen: 30 Punkte]

Gegeben ist folgendes Relationenschema:

benutzer (userid, name, gruppe, team, email)

dokument (titel, beschreibung, groesse, endung, erstellt_am, geaendert_am)

berechtigung (userid, titel, typ)

benutzer.userid \diamond *berechtigung*, *dokument.titel* \diamond *berechtigung*

dokument.endung IN {'pdf', 'ppt', 'doc', 'odt', 'odp'}

berechtigung.typ IN {'upload', 'overload', 'download'}

(Hinweis: Die Größe der Dokumente wird in KiB angegeben)

Formulieren Sie die folgende Abfrage in **Relationenalgebra**:

- a. (8 Punkte) Ermitteln Sie Name und Gruppe aller Benutzer, die für zumindest zwei Dokumente größer gleich 1024 KiB eine 'download' Berechtigung aufweisen.

Formulieren Sie die folgende Abfrage im **Tupelkalkül**:

- b. (7 Punkte) Geben Sie Titel, Beschreibung und Erstellungsdatum aller Präsentationsdokumente ('pdf', 'ppt', 'odp') an die größer als 2048 KiB sind.

Formulieren Sie die folgende Abfrage in **SQL99**:

- c. (8 Punkte) Ermitteln Sie Paare von Benutzern (d.h. jeweils 2 Benutzer), die der selben Gruppe, jedoch einem unterschiedlichen Team zugeordnet sind, wobei weiters darauf zu achten ist, dass zumindest ein Dokument existiert, für welches **beide** Benutzer eine 'download' Berechtigung aufweisen. Geben Sie jeweils beide Namen eines solchen Benutzerpaares aus.

Formulieren Sie die folgende Abfrage in **SQL99**:

- d. (7 Punkte) Ermitteln Sie Name, Gruppe und Team aller Benutzer, die zumindest fünf **beliebige** Dokumente downloaden dürfen. Sortieren Sie die Datensätze zuerst nach Gruppe (aufsteigend), sollte die Gruppe gleich sein, dann nach Name (absteigend).

Aufgabe 2 [Query Optimierung: 30 Punkte]

Gegeben ist die folgende Abfrage in relationaler Algebra auf dem Relationenschema $R_1(F, A, N, T), R_2(V, I, E, R)$:

$$\pi_{N,I,E}(\pi_{F,N,V,I,E,R}(\sigma_{(A>13 \wedge T=8 \wedge I<82)}(R_1 \bowtie_{(A=I)} R_2)))$$

- a. (4 Punkte) Stellen Sie den Ausdruck grafisch dar.
- b. (12 Punkte) Führen Sie eine heuristische Optimierung mit Hilfe der in der Vorlesung angegebenen Äquivalenzumformungsregeln durch, mit dem Ziel, den Abarbeitungsaufwand zu minimieren. Stellen Sie den umgeformten Ausdruck grafisch dar und geben Sie die Reihenfolge der verwendeten Äquivalenzumformungsregeln an. Begründen Sie Ihre Vorgangsweise.
- c. (14 Punkte) Nehmen Sie an, dass die Relation R_1 15.000 und die Relation R_2 50.000 Datensätze enthält, wobei die Blockgröße für beide Relationen 5 ist. Für den Join wird das Nested-Loop Verfahren verwendet (Memorygröße 1 Block) und die Selektivität der Selektionen ist jeweils 1/10 (Annahme der Unabhängigkeit). Geben Sie die Kosten (Anzahl von Blockzugriffen) für
- den ursprünglichen Ausdruck und
 - den umgeformten Ausdruck an.

Nehmen Sie an, dass die Abarbeitung der Ausdrücke Pipelining nützt.

Aufgabe 3 [Formaler Datenbankentwurf: 15 Punkte]

Gegeben ist folgendes Relationenschema:

$$RS_1 = (\{X, A, V, I, E, R\}, \{XI \rightarrow E, I \rightarrow XVR, XA \rightarrow VIE\})$$

- a. (5 Punkte) Geben Sie für RS_1 die minimale Überdeckung der funktionalen Abhängigkeiten an.

Gegeben ist folgendes Relationenschema:

Datenbanksysteme (2h)	schriftliche Einzelprüfung	05.05.2008		2
-----------------------	----------------------------	------------	--	---

$$RS_2 = (\{F, E, N, D, R, I, C, H\}, \{ED \rightarrow FRI, I \rightarrow FN, EI \rightarrow DR\})$$

- b. (10 Punkte) Bestimmen Sie für RS_2 alle Schlüsselkandidaten, sowie *prime* und *nicht-prime* Attribute.

Aufgabe 4 [Speicher und Indexstrukturen: 15 Punkte]

- a. (5 Punkte) Was versteht man unter variabel langen Records?
- b. (10 Punkte) Welche Methoden zu deren Repräsentierung haben Sie in der Vorlesung kennengelernt? Geben Sie jeweils geeignete Beispiele sowie Vor- und Nachteile der beschriebenen Methoden an.

Aufgabe 5 [Serialisierbarkeit: 10 Punkte]

- a. (5 Punkte) Beschreiben Sie die Grundkonzepte des Two-Phase Locking Protokolls und illustrieren Sie die Varianten strict (streng), conservative (konservativ) und rigorous (streng+konservativ) anhand eines 2-dimensionalen Diagramms, wobei auf der x-Achse die Zeit und auf der y-Achse die Anzahl der Sperren angegeben wird.
- b. (5 Punkte) Verhindert das Two-Phase Locking Protokoll das Auftreten von Deadlocks? Wenn ja, beschreiben Sie, warum das so ist. Andernfalls geben Sie ein Beispiel für einen Schedule mit Two-Phase Locking, bei dem ein Deadlock auftritt.