

Aufgabenblatt 1

Physische Speicherstrukturen und Datenrepräsentation

- Abgabetermin: **Mittwoch, 09.05.07**
- Das Aufgabenblatt gilt als bestanden, wenn mindestens 20 der 40 Punkte erreicht werden.
- Die Aufgaben sollen in Zweiergruppen bearbeitet werden.
- Bitte beschriftet *jedes* Blatt der Abgabe mit Namen und Matrikelnummern!
- Abgabe: Auf Papier im Fach „Datenbanksysteme II“ im Foyer oder per E-Mail als pdf oder doc an dbs2@hpi.uni-potsdam.de mit cc an brigitte.hobro@hpi.uni-potsdam.de

Aufgabe 1: Megatron 777

Betrachte die Festplatte Megatron 777 mit folgenden Merkmalen:

- Es gibt 5 Platten mit 10 Plattenoberflächen.
 - Jede Plattenoberfläche trägt 10.000 Spuren.
 - Jede Spur besteht durchschnittlich aus 1000 Sektoren der Größe 512 Byte.
 - 20 % jeder Spur werden von Lücken überdeckt.
 - Die Platten rotieren mit einer Geschwindigkeit von 10.000 U/min.
 - Die Bewegung des Lesekopfes über n Spuren benötigt $(1 + 0,001n)$ ms.
 - Ein Block umfasst 4 Sektoren.
- a) Was ist die Gesamtkapazität der Platte? **1 P**
- b) Wieviele Blöcke liegen auf einem Zylinder? **1 P**
- c) Was ist die minimale, maximale und durchschnittliche Transferzeit für einen Block? **2+2+4 P**
- d) Wie lange dauert es, 10 Millionen Tupel per Two Phase Multiway Merge Sort zu sortieren? Ein Block umfasst 100 Tupel. Zur Optimierung der Zugriffsgeschwindigkeit sind die Daten optimal über die Zylinder organisiert. **8 P**

Aufgabe 2: Elevator Algorithmus vs. First Come First Served

Gegeben sei eine Festplatte mit 8192 Zylindern. Die Seekzeit für n Zylinder ist $(1 + 0,002n)$ ms. Die durchschnittliche Zeit zur Rotation der Platte zum gesuchten Sektor (latency) beträgt 6,5 ms und die durchschnittliche Transferzeit für einen Sektor 0,5 ms. Der Plattenkopf ist initial auf Zylinder 4000 positioniert.

Betrachte die folgenden Anweisungen:

Eintreffen der Anweisung	0 ms	3 ms	11 ms	19 ms
Angefragter Zylinder	6500	2000	8000	3500

Wann werden die einzelnen Requests bearbeitet und wann sind sie abgeschlossen unter Verwendung

- a) des Elevator Algorithmus? **4 P**
- b) der First Come First Served Reihenfolge? **2 P**

Aufgabe 3: Datensätze fester vs. variabler Länge

Gegeben seien die folgenden Anforderungen für eine Relation einer Patientendatenbank: Es gibt 15 Attribute, die immer gefüllt werden (z. B. Name, Geburtsdatum) und weitere 25 optionale Attribute, die nicht für jeden Patienten bekannt oder relevant sind (z. B. aktuelle Blutzuckerwerte). Jedes der optionalen Attribute wird mit der Wahrscheinlichkeit p gefüllt.

Die Werte der erforderlichen Attribute werden in Feldern der festen Länge von 40 Byte gespeichert, die Werte der optionalen Attribute in Feldern der festen Länge von 10 Byte.

- a) Was ist die erwartete Größe eines Tupels (in Abhängigkeit von p) unter Verwendung von
- Datensätzen fester Länge? **1 P**
 - Datensätzen variabler Länge? Die optionalen Attribute sollen als getaggte Felder gespeichert werden. Das Tag jedes dieser Attribute umfasst 2 Byte und die Sequenz der Felder wird mit einem gesonderten Tag der Größe 2 Byte abgeschlossen. **1 P**
- b) Für welche Wahrscheinlichkeiten p sind Datensätze fester Länge zu bevorzugen? **2 P**

Aufgabe 4: Speichern von Relationen

Gegeben seien die folgenden beiden Relationen Hersteller und Produkte, wobei Produkte per Fremdschlüssel einem Hersteller zugeordnet sind. Alle Tupel werden als Datensätze fester Länge gespeichert.

Die Hersteller-Relation umfasst 25.000 Tupel. Jeder Datensatz ist 350 Byte lang. Die Produkte-Relation umfasst 75.000 Tupel. Jeder Produkte-Datensatz hat die Größe 300 Byte.

Die Datensätze sollen in Blöcken der Größe 4096 Byte gespeichert werden, wobei jeder Block 50 Byte für Header-Informationen reserviert. Es gibt keine Datensätze, die auf mehrere Blöcke aufgeteilt sind.

- a) Wieviele Blöcke werden benötigt, um beide Relationen zu speichern, wenn jede Relation am Stück (in getrennte Blockgruppen) gespeichert wird? **4 P**
- b) Wieviele Blöcke werden benötigt, um beide Relationen zu speichern, wenn jeder Hersteller und seine Produkte im selben Block gespeichert werden? Nimm an, dass jeder Hersteller 3 Produkte herstellt. **4 P**
- c) Der Workload auf der Datenbank umfasst zwei Typen von Anfragen:
- 1) Lesen aller Hersteller-Daten
 - 2) Join der Relationen Hersteller und Produkte

Welche der beiden obigen Speichervarianten (Aufgabenteil a und b) bevorzugst du für welche Anfrageart? Und warum? **4 P**