

Aufgabenblatt 1

Physische Speicherstrukturen

- Abgabetermin: **Mittwoch, 17.11.10 (23:59 Uhr)**
- Zur Prüfungszulassung muss ein Aufgabenblatt mit mind. 25% der Punkte bewertet werden und alle weiteren Aufgabenblätter mit mindestens 50% der Punkte.
- Die Aufgaben sollen in Zweiergruppen bearbeitet werden.
- Abgabe:
 - per E-Mail an `dbs2-2010@hpi.uni-potsdam.de` mit Subject
Abgabe DBS II: Aufgabenblatt <n> Namen
 - ausschließlich pdf-Dateien
 - eine Datei pro Aufgabe mit folgendem Dateinamen:
`blatt<aufgabenblattNr>aufgabe<aufgabenNr><Nachnamen>.pdf`
Bitte **keine Leerzeichen, Unterstriche, Umlaute, Sonderzeichen**, ... im Dateinamen!
 - **jedes Blatt beschriftet mit Namen und Matrikelnummern**
 - Wir korrigieren die Abgaben aufgabenweise. Das beschriebene Verfahren vereinfacht uns die Arbeit erheblich!

Aufgabe 1: Megatron 777

Betrachte die Festplatte Megatron 777 mit folgenden Merkmalen:

- Es gibt fünf 3,5" Platten, die jeweils zwei Plattenoberflächen besitzen.
 - 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 in GB? **1 P**
- b) Wieviele Blöcke liegen auf einem Zylinder? **1 P**
- c) Was ist die minimale, maximale und durchschnittliche Latenzzeit für einen Block? **2+2+2 P**
- d) Auf dieser Festplatte sei eine Datei gespeichert, die eine komplette Spur belegt. Wie lange dauert durchschnittlich das sequentielle Einlesen der gesamten Datei, wenn das Einlesen einer Spur nur an einer bestimmten Position auf der Spur beginnen darf. **2 P**

Aufgabe 2: Two Phase Multiway Merge Sort

Gegeben sei die Festplatte Megatron 777 aus Aufgabe 1 mit einem Hauptspeicher von 100 MB (= 100×2^{20} Bytes).

- a) 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. **9 P**
- b) Wie groß ist die minimale Blockanzahl, die für das Sortieren im Hauptspeicher mittels Two-Phase Multiway Merge-Sort benötigt wird? **3 P**

Aufgabe 3: 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 ein Block 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 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**