

Aufgabenblatt 1

Physische Speicherstrukturen

- Abgabetermin: **Dienstag, 05.11.13 (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.
- Abgabesystem unter
<http://www.dcl.hpi.uni-potsdam.de/submit>
 - ausschließlich pdf-Dateien
 - eine Datei pro Aufgabe namens Aufgabe-<aufgabenNr>.pdf
 - jedes Blatt beschriftet mit Namen

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 durchschnittliche Latenzzeit für einen Block? **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 und ein Hauptspeicher von 100 MiB.

- a) Wie lange dauert das Lesen und Schreiben von Tupeln insgesamt, wenn 10 Millionen Tupel per Two Phase Multiway Merge Sort sortiert werden müssen? Ein Block umfasst 100 Tupel. Zur Optimierung der Zugriffsgeschwindigkeit sind die Daten zuvor optimal über die Zylinder verteilt worden. *HINWEIS: Nutze die Ergebnisse aus Aufgabe 1!* **9 P**
- b) Was ist die Mindestgröße des Hauptspeichers (in Anzahl von Blöcken), um die 10 Millionen Tupel aus Aufgabe a) per Two Phase Multiway Merge Sort zu sortieren? Begründen Sie ihre Antwort! **3 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 definierten 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**