

Übung zur Vorlesung *Grundlagen: Datenbanken im WS23/24*

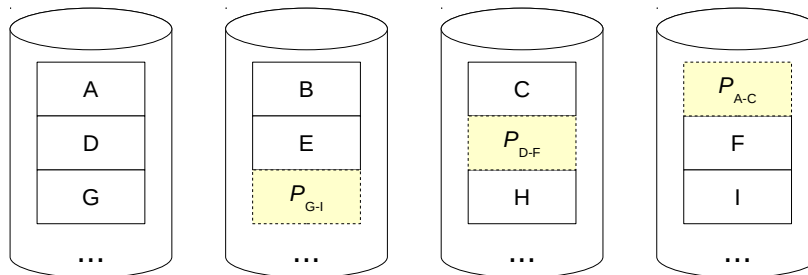
Christoph Anneser, Michael Jungmair, Stefan Lehner, Moritz Sichert, Lukas Vogel
 (gdb@in.tum.de)

<https://db.in.tum.de/teaching/ws2324/grundlagen/>

Blatt Nr. 09

Hausaufgabe 1

Die folgende Abbildung zeigt einen Festplattenverbund bestehend aus vier Laufwerken, auf welchen die Datenblöcke A bis I gespeichert sind. Die Blöcke P_i enthalten Paritätsinformationen.



- Um welches RAID-Level handelt es sich?
- Wieviele Festplatten können ausfallen, ohne dass mit Datenverlust zu rechnen ist? Geben Sie eine allgemeine Lösung für einen Verbund bestehend aus n Festplatten an.
- Kann die Ausfallsicherheit erhöht werden? Begründen Sie!
- Welchen weiteren Vorteil bietet das gezeigte RAID-System neben der Ausfallsicherheit?
- Nach einem Festplattendefekt enthalten die Datenblöcke die folgenden Binärdaten. Rekonstruieren Sie die Datenblöcke der $Disk_2$ mithilfe der XOR-Verknüpfung.

$Disk_0$	$Disk_1$	$Disk_2$	$Disk_3$
A = 1111	B = 1001	C = _ _ _ _	P_{A-C} = 1110
D = 0101	E = 1100	P_{D-F} = _ _ _ _	F = 1100
G = 0011	P_{G-I} = 1110	H = _ _ _ _	I = 0011

Hausaufgabe 2

Gegeben sei ein Array von 1.000.000.000 8-Byte-Integer-Werten und ein Programm, das alle Werte aufsummiert.

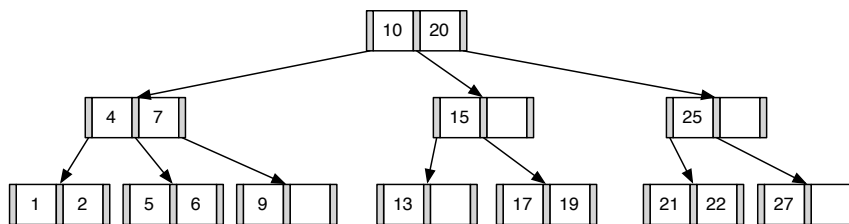
Das Programm wird auf einem System mit 16 GB Hauptspeicher und einer herkömmlichen Magnetfestplatte (Größe 1 TB), auf der alle Werte sequentiell gespeichert sind, ausgeführt.

Ein Random Access auf die Festplatte dauert 10 ms, beim sequentiellen Lesen hat sie einen Durchsatz von 160 MB/s. Das Summieren zweier Werte im Hauptspeicher dauert 1 ns.

(1 MB = 10^6 B und 1 TB = 10^{12} B)

- Gehen Sie davon aus, dass alle Werte bereits im Hauptspeicher liegen. Wie lange läuft das Programm?
- Nun liegen alle Werte ausschließlich auf der Festplatte. Wie lange läuft das Programm jetzt?
- Auf der Festplatte liegt jetzt zusätzlich nach jedem 100.000. Wert die Summe der 100.000 davorliegenden Werte. Wie lange läuft das Programm, wenn es nur diese Summen aufsummiert?

Hausaufgabe 3



- Fügen Sie die 3 in den gezeigten B-Baum ein. Zeichnen Sie das Endergebnis. Zeichnen Sie jeweils den kompletten Baum oder machen Sie **deutlich**, falls Teile des Baumes unverändert bleiben. Verwenden Sie den aus der Vorlesung bekannten Algorithmus.
- Entfernen Sie aus dem **ursprünglichen Baum** den Eintrag 20. Zeichnen Sie das Ergebnis der Operation. Sollte es mehrere richtige Lösungen geben, genügt es, wenn Sie hier eine angeben. Zeichnen Sie jeweils den kompletten Baum oder machen Sie **deutlich**, falls Teile des Baumes unverändert bleiben. Verwenden Sie den aus der Vorlesung bekannten Algorithmus.

Hausaufgabe 4

Bestimmen Sie k für einen B-Baum, der die folgenden Informationen aller Menschen auf der Erde (ca. 10 Milliarden) enthalten soll: Namen, Land, Stadt, PLZ, Straße und Hausnummer (insgesamt ca. 100 Byte). Dabei ist die Steuernummer eindeutig und 64 Bit lang und wird im B-Baum als Suchschlüssel verwendet. Gehen Sie bei der Berechnung davon aus, dass eine Speicherseite 16 KiB groß ist und ein Knoten des B-Baums möglichst genau auf diese Seite passen sollte.

(1 KiB = 2^{10} Byte)

Hausaufgabe 5

Geben Sie eine Permutation der Zahlen 1 bis 24 an, so dass beim Einfügen dieser Zahlenfolge in einen (anfängs leeren) B-Baum mit Grad $k = 2$ ein Baum minimaler Höhe entsteht. Skizzieren Sie den finalen Baum.

Bonusaufgabe 6

Implementieren Sie einen B+-Baum in C++. Es sollten mindestens die Funktionen *insert* und *lookup* unterstützt werden. Zur Vereinfachung können Sie annehmen, dass lediglich Schlüssel-Wert-Paare bestehend aus Integern eingefügt werden.

Diese Aufgabe ist für diejenigen gedacht, die sich über den Vorlesungsstoff hinaus mit dem Thema Datenbanken beschäftigen wollen. Sie wird nicht in der Übung besprochen und ist nicht klausurrelevant. Falls Sie Feedback wünschen, können Sie Ihre Lösung gerne an gdb@in.tum.de senden.

**Der Lehrstuhl für Datenbanksysteme wünscht
frohe Weihnachten und schöne Feiertage!**

