

## Übung zur Vorlesung *Grundlagen: Datenbanken* im WS19/20

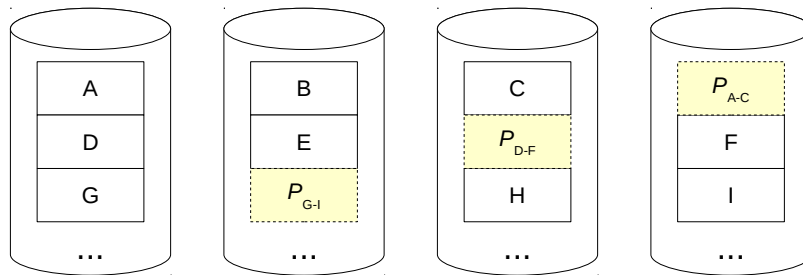
Christoph Anneser, Moritz Sichert, Lukas Vogel (gdb@in.tum.de)

<https://db.in.tum.de/teaching/ws1920/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? Begründung?
- 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)

- a) Gehen Sie davon aus, dass alle Werte bereits im Hauptspeicher liegen. Wie lange läuft das Programm?
- b) Nun liegen alle Werte ausschließlich auf der Festplatte. Wie lange läuft das Programm jetzt?
- c) 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

- a) Fügen Sie in einen anfänglich leeren  $B^+$ -Baum mit  $k = 3$  und  $k^* = 2$  die Zahlen eins bis fünfundzwanzig in aufsteigender Reihenfolge ein. In den Blattknoten werden TIDs verwendet. Was sind TIDs, wann lohnt sich ihre Verwendung, was ist die Alternative zu TIDs?
- b) Erläutern Sie die Vorgehensweise bei der Bearbeitung der folgenden Anfrage „Finde alle Datensätze mit einem Schlüsselwert zwischen 5 und 15.“

### Hausaufgabe 4

Fügen Sie nacheinander die folgenden Einträge in eine anfangs leere erweiterbare Hashtabelle, welche 2 Einträge pro Bucket aufnehmen kann, ein. Es soll effizient nach der **KundenNr** gesucht werden können.

KundenNr	Name
10	Müller
25	Meier
30	Schmidt
18	Krause
40	Schulz
45	Kaufmann

### Hausaufgabe 5

Gegeben sei eine erweiterbare Hashtabelle mit globaler Tiefe  $t$ . Wie viele Verweise zeigen vom Verzeichnis auf einen Behälter mit lokaler Tiefe  $t'$ ?