

### Übungsblatt 5

Unterlagen zur Vorlesung: „www.dvs.informatik.uni-kl.de/courses/DBSREAL/“

#### Aufgabe 1: Direkte und Indirekte Satzadressierung

12

a) Vergleichen Sie den Speicherplatzbedarf für folgende Adressierungsprinzipien:

1. logische Byte-Adresse
2. TID-Konzept
3. DBK-Konzept
4. DBK/PPP-Konzept.

Die Ausprägungen eines zu adressierenden Satztyps seien jeweils auf ein Segment der Größe  $2^{32}$  Byte beschränkt. Es können mehrere Satztypen in einem Segment gespeichert sein. Welche Faktoren werden durch die einzelnen Adressierungskonzepte bei gegebener Pointerlänge begrenzt:

- $N_{REC}$  : Anzahl der Sätze,
- $S_l$  : logische Satzlänge,
- $L_s$  : Seitengröße,
- $m_k$  : Anzahl der Seiten (im Segment)?

Wie hoch ist der gesamte Speicherplatzbedarf für die Adressierung eines Satzes, wenn n verschiedene Pointer auf den Satz zeigen? Die Anzahl der Überläufer bei TID-Konzept liege bei 10%.

b) Schätzen sie den Zugriffsaufwand bei den verschiedenen Adressierungskonzepten ab. Wieviele physische E/A-Zugriffe sind bei einer Folge wahlfreier Zugriffe über einen Indexbaum mit 3 Stufen im Mittel zu erwarten, bis der gesuchte Datensatz gefunden ist,

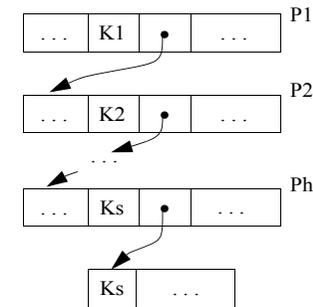
1. beim TID-Konzept
2. beim DBK-Konzept
3. beim DBK/PPP-Konzept?

Der DB-Puffer besitze 40 Seitenrahmen zu 4 K Byte; die Ersetzungsstrategie sei LRU. Es seien  $N_{REC}=10^5$  Datensätze über  $10^4$  Seiten verteilt. Der Indexbaum habe neben der Wurzel 9 Seiten auf der 2. Indexstufe und 1000 Seiten auf der 3. Indexstufe. Die Zuordnungstabelle bestehe aus Seiten der Größe 4 KB und besitze PP-Einträge der Länge 4 Byte. Die Anzahl der Überläufer sowie die Anzahl der falschen PPPs sei 10%.

Vergleichen sie die Anzahl der Zugriffe, wenn über die interne logische Adresse (TID, DB-Key) direkt zugegriffen wird.

c) Ermitteln Sie die Anzahl der physischen Zugriffe bei der fortlaufenden Verarbeitung aller Datensätze über eine Indexstruktur, wenn mit

1. dem TID-Konzept



2. dem DBK-Konzept
3. dem DBK/PPP-Konzept

adressiert wird. Welche Ergebnisse erhalten Sie unter den Annahmen von b) im *best case* (Clusterbildung) und im *worst case*?

#### Aufgabe 2: Pointer-Swizzling bei Bäumen

190

Gegeben ist ein vollständig belegter B\*-Baum der Klasse  $T(k, k^*, h)$ , der in seinen Blattseiten auf Sätze verweist, die keine weiteren Verweise enthalten.

Die B\*-Baum-Seiten besitzen folgendes Format:

Innerer Knoten	$Z_0$	$R_1$	$Z_1$	$R_2$	$Z_2$	...	$R_p$	$Z_p$	freier Platz
----------------	-------	-------	-------	-------	-------	-----	-------	-------	--------------

$R_i$  = Referenzschlüssel (Wegweiser),  $k \leq p \leq 2k$

$Z_i$  = Verweis auf Sohnseite

Blattknoten	P	$S_1$	$D_1$	$S_2$	$D_2$	...	$S_j$	$D_j$	freier Platz	N
-------------	---	-------	-------	-------	-------	-----	-------	-------	--------------	---

P = PRIOR-Zeiger, N = NEXT-Zeiger,  $k^* \leq j \leq 2k^*$

$S_i$  = Satzschlüssel

$D_i$  = Verweis auf Satz (referenzierte Speicherung)

Eine innere Seite (inkl. der Wurzel) habe also bei voller Belegung  $2k+1$  Verweise auf Seiten und eine Blattseite auf der Stufe h habe  $2k^*$  Verweise auf Sätze.

Anfangs seien keine Seiten des B\*-Baumes und keine der referenzierten Sätze im Hauptspeicher. Im folgenden soll Copy-Swizzling analysiert werden.

Skizzieren Sie für die verschiedenen Swizzling-Varianten die einzelnen Aktionen, die aus einem Suchvorgang für den Satz mit Schlüssel  $K_s$  resultieren, wenn dabei die Wurzelseite  $P_1$ , auf Baumebene 2 die Seite  $P_2$ , ... und auf Baumebene h die Seite  $P_h$  aufgesucht werden. Der Suchpfad ergebe sich durch die Schlüssel (Wegweiser)  $K_1$  in  $P_1$ ,  $K_2$  in  $P_2$ , ...,  $K_s$  in  $P_h$ :

Wie viele Seiten, Deskriptoren und Sätze sind jeweils einzulagern und aufzusuchen, wenn als Verfahren

