

Datenbanksysteme II
SS 2009

Übungsblatt 6: Hardwaregrundlagen

Besprechung: 22.06.2008 und 24.06.2008

Aufgabe 6-1 *Sequentieller Scan vs. Index*

Es werde ein komplexes Selektionsprädikat auf einer Tabelle aus 1.000.000 Tupeln ausgewertet. Ein Tupel belege hierbei 100 Bytes. Für die Operation stehen 10 MByte Datenbankpuffer zur Verfügung. Die Auswertung des Prädikats benötige 10 μ s CPU-Zeit. Die Daten des Plattenlaufwerkes seien wie folgt gegeben:

- $t_{seek} = 4$ ms
- $t_{lat} = 2$ ms
- Transferrate = 50 MByte/s.

(a) Zunächst sei die Tabelle in einer Datei organisiert, die sequentiell gelesen wird.

- (i) Wie viel Zeit benötigt das Einlesen der Datei? Fällt die Latenzzeit ins Gewicht?
- (ii) Wie viel CPU-Zeit benötigt die Selektionsanfrage?

(b) Nun sei die Tabelle in einem Index (B^+ -Baum) organisiert. Eine Seite habe eine Größe von 4 KByte. Die Speicherauslastung betrage 70%.

- (i) Wie viele Datenseiten werden benötigt, wenn die Datenseiten untereinander vorwärts und rückwärts verkettet sind und ein Zeiger 4 Byte benötigt?
- (ii) Wie viele Directory-Seiten werden benötigt, wenn man davon ausgeht, dass der Schlüssel 20 Byte und der Zeiger auf die Sohnseite 4 Byte benötigt?
- (iii) Der Index habe auf dem Selektionsprädikat optimale Selektivität (*best case*). Wie teuer ist die Auswertung (CPU und I/O)?
- (iv) Wie teuer ist die Auswertung (CPU und I/O) bei schlechter Selektivität (*worst case*)? Ist die Selektion CPU- oder I/O-bound?