

Prof. Dr. Guido Moerkotte

Email: moer@pi3.informatik.uni-mannheim.de

Pit Fender

B6, 29, Raum C0.05

68131 Mannheim

Telefon: (0621) 181-2517

Email: pfender@pi3.informatik.uni-mannheim.de

Anfrageoptimierung

Herbst-/Wintersemester 07

11. Übungsblatt

5. Dezember 2007

Aufgabe 1

Gegeben eine Relation mit 6 Tupeln und 3 Seiten (2 Tupel pro Seiten). Berechnen Sie diesmal (im Gegensatz zum letzten Übungsblatt) mit Hilfe der Yao-Formel die durchschnittliche Anzahl gelesener Seiten für 2 gelesene Tupel.

Lösung

Es ergeben sich folgende Werte:

$$N = 6$$

$$m = 3$$

$$B = n = 2$$

$$k = 2$$

$$\bar{\mathcal{Y}}_n^{N,m}(k) = \bar{\mathcal{Y}}_2^{6,3}(2) = 1,8$$

Aufgabe 2

Aufgabe 2 a)

Gegeben eine Relation mit 1000 Tupeln auf 100 Seiten (10 Tupel pro Seite). Berechnen Sie den erwarteten Abstand zwischen der ersten und der letzten gelesenen Seite, wenn 50 verschiedene Tupel gelesen werden.

Lösung

- Gelesene Seiten (Yao): $\bar{\mathcal{Y}}_n^{N,m} \approx 40.27$ Seiten
- Abstand: $\bar{\mathcal{B}}_{\text{tot}} = \frac{Bb+b}{b+1} \approx 98.55$ Seiten (für Tabellen-Scan); alternativ $\bar{\mathcal{B}}_{1\text{-span}} = \frac{Bb-B+2b}{b+1} \approx 97,1$ Seiten (für Range-Scan in B-Baum)

Aufgabe 2 b)

Wie ändert sich die obige Abschätzung, wenn die 50 Tupel Duplikate enthalten können?

Lösung

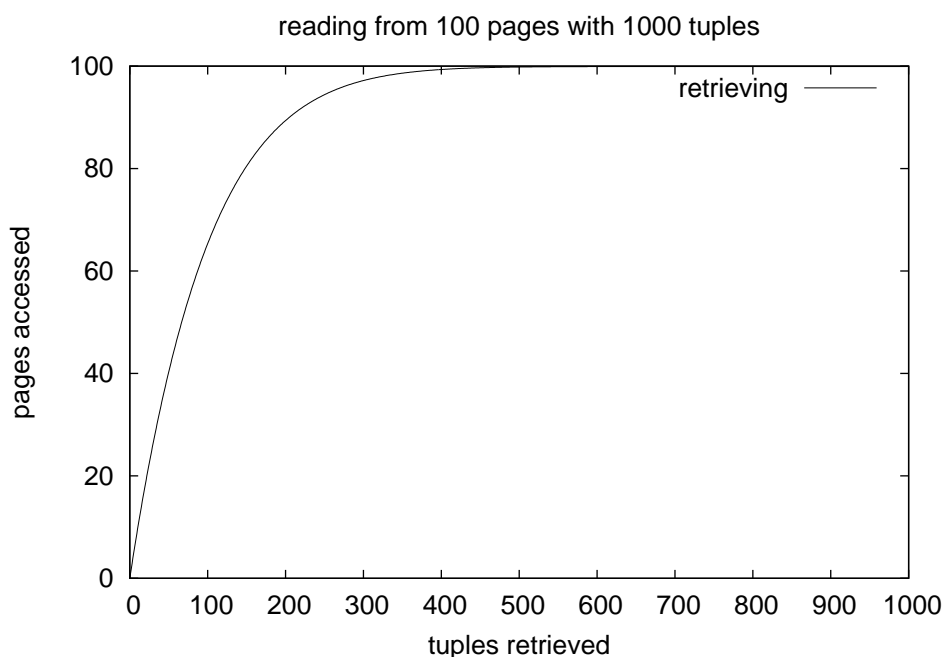
- Gelesene Seiten (Cheung): $\overline{Cheung}_n^{N,m} \approx 38,77$, alternativ mit $\mathcal{Y}_n^{N,m}(\mathcal{D}(N, k)) \approx 38,39$
- Abstand: $\overline{\mathcal{B}}_{\text{tot}} = \frac{Bb+b}{b+1} \approx 98,5$ Seiten; alternativ: $\overline{\mathcal{B}}_{1\text{-span}} = \frac{Bb-B+2b}{b+1} \approx 96,9$ Seiten

Aufgabe 3

Gegeben eine Relation mit 1000 Tupeln und 100 Seiten. Plotten Sie die Anzahl der erwarteten Seitenzugriffe für 1-1000 Tupel mit Hilfe der Yao-Formel.

Lösung

yao.java : plote Ergebnisse mit gnuplot



Aufgabe 4

Als zusätzlichen Parameter wird $C =$ Anzahl Seiten pro Zylinder in die Yao-Formel eingeführt. Berechnen Sie die durchschnittliche Anzahl von zugegriffenen Zylindern, wenn k Tuple gelesen werden.

Lösung

$$\frac{m}{C} \left[1 - \frac{\binom{N-n*C}{k}}{\binom{N}{k}} \right]$$

Aufgabe 5

Geben Sie zwei Anfragen an, für deren Kostenabschätzung die Yao-Formel verwendet werden kann.

Lösung

1. `select distinct *`
`from a a`
`where a.x=10`
2. `select distinct *`
`from a a,b b`
`where a.x=10`
`and a.y=b.y`

Aufgabe 6

Aufgabe 6 a)

Sei n die Anzahl der Seiten einer Relation. Die Zeit für einen sequentiellen Seitenzugriff sei $D_{seq} = 0,5 \frac{ms}{page}$, und für einen wahlfreien Seitenzugriff $D_{rand} = 5,0 \frac{ms}{page}$. Gegeben sei ein Prädikat mit Selektivität s . Angenommen, man kann mit einem wahlfreien Seitenzugriff auf die Tupel der Relation zugreifen, die das Prädikat erfüllen.

Berechnen Sie die Selektivität s , ab der die wahlfreien Seitenzugriffe schneller sind, als ein sequentieller Scan der Relation. Gehen Sie zur Vereinfachung davon aus, daß für den Zugriff auf die Daten im Index je nur ein wahlfreier Seitenzugriff nötig ist.

Lösung

$$\begin{aligned} n \cdot D_{seq} &> s \cdot n \cdot D_{rand} \\ s &< 0,1 \end{aligned}$$

D.h. für Selektivitäten kleiner als 10% lohnt in diesem Beispiel ein direkter wahlfreier Seitenzugriff.

Aufgabe 6 b)

Seien R und S zwei Relationen, die r bzw. s Seiten benötigen. Sei p ein Selektionsprädikat auf Relation S . Berechnen Sie in dem Plan $R \bowtie^{NL} (\sigma_p(S))$, ab welcher Selektivität $sel(p)$ es sinnvoll ist, die Tupel von $\sigma_p(S)$ zu materialisieren.

Die Zeit für einen sequentiellen schreibenden Seitenzugriff sei dabei D_{wseq} .

Lösung

$$\begin{aligned}
 (r \cdot D_{seq}) \cdot (s \cdot D_{seq}) &> (r \cdot D_{seq}) \cdot (sel(p) \cdot s \cdot D_{seq}) + (sel(p) \cdot s \cdot D_{wseq} + s \cdot D_{seq}) \\
 &= sel(p) \cdot (r \cdot s \cdot D_{seq}^2 + s \cdot D_{wseq}) + s \cdot D_{seq} \\
 sel(p) &> \frac{r \cdot D_{seq}^2 - D_{seq}}{r \cdot D_{seq}^2 + D_{wseq}}
 \end{aligned}$$

Für $r = 1$ und $D_{seq} = D_{wseq} = 5.0 \frac{ms}{page}$ ist $sel(p) = 2/3$. Für größere r nimmt auch $sel(p)$ zu. Temp lohnt also oft.