

Procjene troškova

Selekcija

- bez indeksa, linearna pretraga: $B(r)$
- pomoću indeksa idx (B-stabla)
 - jedna n-torka (preko ključa relacije): $d(idx) + 1$
 - više n-torki
 - *clustered index*: $d(idx) +$ broj blokova u kojima se nalaze tražene n-torke
 - *non-clustered index*: vrijednosti u listovima indeksa jesu sortirane, ali tražene n-torke mogu biti znatno raspršene među blokovima s podacima. Ekstremni slučaj: za dohvat svake n-torke potrebna je jedna U/I operacija. Ako je broj n-torki koje se dohvaćaju velik, trošak može biti i veći nego trošak linearne pretrage pa se tada indeks ne koristi

External sort-merge

- ukupan broj potrebnih U/I operacija: $2 \cdot B(r) + 2 \cdot B(r) \lceil \log_{M-1} \lceil B(r) / M \rceil \rceil$
 - uz uključen trošak zapisivanja konačnog rezultata, inače za $B(r)$ manje!
 - ako se koristi *pipelining*, ne postoji trošak zapisivanja konačnog rezultata

Nested-loop join

- najlošiji slučaj: samo po jedan blok svake relacije stanu u glavnu memoriju
 - procjena troška: $N(r) \cdot B(s) + B(r)$ (nisu uključeni troškovi zapisivanja rezultata)
- najbolji slučaj: obje relacije ili samo manja relacija (manja relacija se tada koristi kao unutarnja) stanu u glavnu memoriju
 - procjena troška: $B(r) + B(s)$ (nisu uključeni troškovi zapisivanja rezultata)

$t_r \cdot t_s$: n-torka dobivena ulančavanjem n-torki t_r i t_s

$N(r)$, $N(s)$: broj n-torki u relacijama r i s

$B(r)$, $B(s)$: broj blokova u relacijama r i s

Block nested-loop join

- najlošiji slučaj: samo po jedan blok svake relacije stanu u glavnu memoriju
 - procjena troška: $B(r) \cdot B(s) + B(r)$ (nisu uključeni troškovi zapisivanja rezultata)
- najbolji slučaj: (jednako kao u *nested-loop join*)
 - procjena troška: $B(r) + B(s)$ (nisu uključeni troškovi zapisivanja rezultata)

Ako je za operaciju spajanja na raspolaganju glavna memorija veličine M blokova

- 1 blok za čitanje unutarnje relacije, 1 blok za izlazni međuspremnik
- vanjska petlja: umjesto po jedan blok, $B(r) / (M - 2)$ puta učitava po $M-2$ blokova
- procjena troška: $B(r) / (M - 2) \cdot B(s) + B(r)$
(nisu uključeni troškovi zapisivanja rezultata)

Indexed nested-loop join

- najlošiji slučaj: samo po jedan blok svake relacije stanu u glavnu memoriju
 - procjena troška: $N(r) \cdot c + B(r)$ (nisu uključeni troškovi zapisivanja rezultata)

c : broj U/I operacija za dohvat n -torke iz s pomoću indeksa (\sim dubina stabla)

Sort-merge join

- procjena troška: $B(r) + B(s)$ (nisu uključeni troškovi zapisivanja rezultata)

Hash join

- procjena troška: $3 \cdot (B(r) + B(s))$ (nisu uključeni troškovi zapisivanja rezultata)

Eliminacija duplikata / Projekcija

- procjena troška: $3 \cdot B(r)$ (nisu uključeni troškovi zapisivanja rezultata)

Agregacija i grupiranje

Ako je rezultat prevelik za raspoloživi prostor glavne memorije

- procjena troška(za *hash*): $3 * B(r)$ (nisu uključeni troškovi zapisivanja rezultata)

Ako se rezultat može pohraniti u raspoloživi prostor glavne memorije

- procjena troška: $B(r)$ (nisu uključeni troškovi zapisivanja rezultata)

Unija, presjek, razlika

- procjena troška:
 - ako su r i s sortirane
 - $B(r) + B(s)$
 - ako r i s nisu sortirane
 - trošak sortiranja (r) + trošak sortiranja (s) + $B(r) + B(s)$

Procjena broja n-torki

Jednostavna selekcija

$$\sigma_{A=v}(r)$$

- uz pretpostavku: jednolika razdioba vrijednosti atributa A
$$N(r) / V(A, r)$$
- ako $V(A, r)$ nije poznata: $N(r) / 10$

$$\sigma_{A \leq v}(r)$$

- ako je vrijednost v poznata u trenutku optimizacije
 - za $v < \min(A, r)$ 0
 - za $v \geq \max(A, r)$ $N(r)$
 - inače $N(r) \cdot (v - \min(A, r)) / (\max(A, r) - \min(A, r))$
- ako vrijednost v ili $\min(A, r)$, $\max(A, r)$ nisu poznate u trenutku optimizacije
$$N(r) / 3$$
 - u tom slučaju se isti izraz koristi i za nejednakosti oblika $<, \geq, >$

$$\sigma_{A \text{ LIKE izraz}}(r)$$

$$N(r) / 5$$

Složena selekcija - konjunktivna forma

$$\sigma_{F_1 \wedge F_2 \wedge \dots \wedge F_m}(r)$$

- procijenjeni broj n-torki u rezultatu je dakle
$$N(r) \cdot (N_{F_1}(r) \cdot N_{F_2}(r) \cdot \dots \cdot N_{F_m}(r)) / (N(r))^m$$
 - odnosno $N(r) \cdot f(F_1, r) \cdot f(F_2, r) \cdot \dots \cdot f(F_m, r)$

Složena selekcija - disjunktivna forma

$$\sigma_{F_1 \vee F_2 \vee \dots \vee F_m}(r)$$

- procijenjeni broj n-torki u rezultatu je dakle
$$N(r) \cdot (1 - (1 - N_{F_1}(r) / N(r)) \cdot (1 - N_{F_2}(r) / N(r)) \dots (1 - N_{F_m}(r) / N(r)))$$
- odnosno $N(r) (1 - (1 - f(F_1, r)) \cdot (1 - f(F_2, r)) \dots (1 - f(F_m, r)))$

Spajanje

$$r(R) \bowtie s(S)$$

- ako je $R \cap S = \emptyset \Rightarrow N(r) \cdot N(s)$
- ako je $R \cap S = \text{ključ u } R \Rightarrow N(s)$ (jer se svaka t_s može spojiti s najviše jednom t_r)
- ako je $R \cap S = \text{ključ u } S \Rightarrow N(r)$ (jer se svaka t_r može spojiti s najviše jednom t_s)
- $R \cap S = A$, A je atribut ili skup atributa koji nije ključ niti za R niti za S
- dakle svaka t_r će se spojiti sa $N(s) / V(A, s)$ n-torki t_s . Procijenjeni broj n-torki
$$N(r) \cdot N(s) / V(A, s)$$
- procjena za $s \bowtie r$: $N(s) \cdot N(r) / V(A, r)$
- očito, ako je $V(A, r) \neq V(A, s)$, procjene se razlikuju! U takvom slučaju, kao točnija procjena uzima se manja od dviju prikazanih, dakle
$$N(r) \cdot N(s) / \max(V(A, r), V(A, s))$$

Unija, presjek, razlika, projekcija, agregacija

$\sigma_{F_1}(r) \cup \sigma_{F_2}(r)$ unija podskupova iste relacije

- procjenjuje se jednako kao rezultat operacije $\sigma_{F_1 \vee F_2}(r)$

$\sigma_{F_1}(r) \cap \sigma_{F_2}(r)$ presjek podskupova iste relacije

- procjenjuje se jednako kao rezultat operacije $\sigma_{F_1 \wedge F_2}(r)$

$r \cup s \Rightarrow N(r) + N(s)$

$r \cap s \Rightarrow \min(N(r), N(s))$

$r \setminus s \Rightarrow N(r)$

$\pi_L(r) \Rightarrow V(L, r)$

$\rho_{\mathcal{AF}_1(B_1), \mathcal{AF}_2(B_2), \dots, \mathcal{AF}_n(B_n)}(r) \Rightarrow V(L, r)$