

Tecnologie delle Basi di Dati M

Appello del 25/2/2011

Esercizio 1 (2 punti)

Data la relazione con schema:

Studenti(matricola, nome, residenza, luogonascita, datanascita)

si effettui una stima del numero di pagine necessarie per memorizzare la relazione e del numero di livelli (e di nodi) di un B⁺-tree costruito sull'attributo (residenza). Si supponga di avere pagine di dimensione 2 KB, di cui 48 B riservati per il page header, e si considerino i seguenti valori:

- Numero di tuple = 10K
- Numero di chiavi (residenza) = 1K
- Dimensione matricola = 4 byte
- Dimensione nome = 32 byte
- Dimensione residenza = 20 byte
- Dimensione luogonascita = 20 byte
- Dimensione datanascita = 4 byte
- Dimensione RID = 4 byte
- Dimensione PID = 4 byte
- Percentuale di riempimento foglie = 90%

Esercizio 2 (5 punti)

Data la relazione con schema:

Personale(matricola, nome, data, luogo, stipendio, responsabile)

si ottimizzi l'esecuzione della seguente interrogazione SQL:

```
SELECT P.matricola, R.matricola
FROM Personale P, Personale R
WHERE P.responsabile=R.matricola
      AND P.stipendio BETWEEN 30000 AND 60000
      AND R.luogo in ('Milano', 'Bologna')
```

tenendo conto che dai cataloghi della base di dati risulta:

- Numero di tuple Personale = 60K
- Numero di pagine Personale = 2K
- Numero di responsabili = 100
- Indice unclustered (TID disordinate) su luogo: numero foglie = 400, numero chiavi = 100
- Indice unclustered (TID ordinate) su stipendio: numero foglie = 2400, valore minimo = 20000, valore massimo = 200000
- Indice clustered su matricola: numero foglie = 3000

Si disegni infine l'albero corrispondente al piano di accesso di costo minimo e stimi il numero di risultati dell'interrogazione.

Suggerimento: per la formula di Cardenas si utilizzino i seguenti valori, validi per P = 2000:

R	$\Phi(R, P)$
0	0
500	442.4958
1000	787.0904
1500	1055.444
2000	1264.425
2500	1427.17
3000	1553.907
3500	1652.604
4000	1729.465
4500	1789.32

R	$\Phi(R, P)$
5000	1835.933
5500	1872.232
6000	1900.501
6500	1922.515
7000	1939.658
7500	1953.009
8000	1963.405
8500	1971.502
9000	1977.807
9500	1982.717

R	$\Phi(R, P)$
10000	1986.541
10500	1989.519
11000	1991.838
11500	1993.644
12000	1995.05
12500	1996.145
13000	1996.998
13500	1997.662
14000	1998.179
14500	1998.582

Esercizio 3 (5 punti)

Si illustri il funzionamento dell'indice B⁺-tree in fase di inserimento di oggetti, mostrandone le prestazioni in termini di costo medio per oggetto inserito.

Esercizio 4 (3 punti)

Gli algoritmi presentati per la valutazione efficiente di join sono adatti ad essere utilizzati nel caso in cui la condizione di join sia una condizione di uguaglianza. Si discuta sulla possibile estensione degli algoritmi *sort-merge join* e *hash join* per la valutazione efficiente di outer join (destro, sinistro e full).

Soluzione Esercizio 1

Dimensionamento relazione:

Dimensione di ogni tupla = $4 + 32 + 20 + 20 + 4 = 80B$

Numero di tuple per pagina = $(2048 - 48)/80 = 4000/80 = 25$

Numero di pagine della relazione = $NT/25 = 10000/25 = 400$

Dimensionamento indice (residenza):

Numero di chiavi = 1K, quindi mediamente ci sono $100K/1K = 10$ tuple per ogni valore di chiave.

Dimensione di ogni record (foglia) = $20 + 10 \times 4 = 60B$

Dimensione "reale" foglia = $(2048 - 48) \times 0.90 = 1800B$

Numero di record per foglia = $1800/60 = 30$

Numero di foglie = $1000/30 = 34$

Dimensione di ogni record (nodo interno) = $20 + 4 = 24B$

Numero nodi livello 1 = $34 \times 24/2000 = 1$

Il B⁺-tree corrispondente si compone quindi di 2 livelli per un totale di 1 nodo interno (radice) e 34 foglie.

Soluzione Esercizio 2

Selettività dei predicati:

Predicato su stipendio = $(60000 - 30000)/(200000 - 20000) = 0.166$

Predicato su luogo = $1/100 = 0.01$ per ogni valore di luogo

Predicato di join = $1/60K$ (chiave esterna)

Accesso a P:

Costo scan sequenziale = **2000**

Costo indice su stipendio: $NL \times 0.166 + \Phi(NT \times 0.166, NP) = 2400 \times 0.166 +$

$\Phi(60K \times 0.166, 2K) = 400 + \Phi(10K, 2K) = 400 + 1986 = 2386$

Numero tuple residue = $NT \times 0.166 = 10000$

Accesso a R:

Costo scan sequenziale = **2000**

Costo indice su luogo: $2 \times (NL \times 0.01 + NT \times 0.01) = 2 \times (400 \times 0.01 + 60K \times 0.01)$

$= 2 \times (4 + 600) = 1208$

Costo indice su matricola: $1 + 1 = 2$

Numero tuple residue = $2 \times NT \times 0.01 = 1200$

Costi di join:

P esterna: costo = costo sequenziale + $10000 \times$ costo indice matricola

$= 2000 + 10000 \times 2 = 22000$

R esterna: costo = costo indice luogo + $1200 \times$ costo sequenziale = $1208 + 1200 \times 2000 =$

2401208

Il numero di risultati dell'interrogazione è $60K \times 0.02 \times 0.166 = 200$