

Tecnologie delle Basi di Dati M

Appello del 21/6/2010

Esercizio 1 (2 punti)

Data la relazione con schema:

Studenti(matricola, datanascita, luogonascita, indirizzo)

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 composto (luogonascita, datanascita). Si supponga di avere pagine di dimensione 4KB, di cui 96 B riservati per il page header, e si considerino i seguenti valori:

- Numero di tuple = 160K
- Numero di chiavi (luogonascita) = 400
- Numero di chiavi (datanascita) = 8000
- Dimensione matricola = 4 byte
- Dimensione datanascita = 6 byte
- Dimensione luogonascita = 20 byte
- Dimensione indirizzo = 20 byte
- Dimensione RID = 4 byte
- Dimensione PID = 4 byte
- Percentuale di riempimento foglie = 75%

Esercizio 2 (4 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>60000
      AND R.luogo = "Bologna"
```

tenendo conto che dai cataloghi della base di dati risulta:

- Numero di tuple Personale = 100K
- Numero di pagine Personale = 1K
- Numero di responsabili = 100
- Indice unclustered (TID ordinate) su luogo: numero foglie = 400, numero chiavi = 100
- Indice unclustered (TID ordinate) su stipendio: numero foglie = 2500, valore minimo = 10000, valore massimo = 210000
- Indice clustered su matricola: numero foglie = 3600

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 = 1000:

R	$\Phi(R, P)$
100	95.20785
200	181.3512
300	259.293
400	329.8141
500	393.6211
600	451.3531
700	503.5886
800	550.8509
900	593.6134
1000	632.3046

R	$\Phi(R, P)$
1100	667.3121
1200	698.9866
1300	727.6454
1400	753.5757
1500	777.0372
1600	798.265
1700	817.4718
1800	834.8499
1900	850.5735
2000	864.8001

R	$\Phi(R, P)$
2100	877.6722
2200	889.3187
2300	899.8565
2400	909.3909
2500	918.0176
2600	925.823
2700	932.8852
2800	939.2751
2900	945.0566
3000	950.2876

Esercizio 3 (5 punti)

Si illustrino i principali problemi dovuti alla concorrenza di transazioni e si mostri come il protocollo Strict 2PL ne impedisca il verificarsi.

Facoltativo (1 punto): si indichino quali caratteristiche del protocollo o delle transazioni siano responsabili del non verificarsi di ciascun malfunzionamento.

Esercizio 4 (4 punti)

Indicare quale sia il ruolo svolto dalle statistiche su DB/indici nell'ottimizzazione delle interrogazioni. In particolare, si discuta sull'opportunità di aggiornare frequentemente tali statistiche e/o di riservare più/meno spazio ad alcune statistiche rispetto ad altre.

Soluzione Esercizio 1

Dimensionamento relazione:

Dimensione di ogni tupla = $4 + 6 + 20 + 20 = 50B$

Numero di tuple per pagina = $(4096 - 96)/50 = 4000/50 = 80$

Numero di pagine della relazione = $NT/800 = 160000/800 = 2000$

Dimensionamento indice (luogonascita, datanascita):

Numero di chiavi = $400 \times 8000 > 160K$ = Numero di tuple, quindi si può supporre non ci siano duplicati.

Dimensione di ogni record (foglia) = $4 + 6 + 20 = 30B$

Dimensione "reale" foglia = $(4096 - 96) \times 0.75 = 3000B$

Numero di foglie = $160K \times 30/3000 = 1600$

Dimensione di ogni record (nodo interno) = $4 + 6 + 20 = 30B$

Numero nodi livello 1 = $1600 \times 30/4000 = 12$

Numero nodi livello 2 = $12 \times 30/4000 = 1$

Il B⁺-tree corrispondente si compone quindi di 3 livelli per un totale di 13 nodi interni e 1600 foglie.

Soluzione Esercizio 2

Selettività dei predicati:

Predicato su stipendio = $(210000 - 60000)/(210000 - 10000) = 0.75$

Predicato su luogo = $1/100 = 0.01$

Predicato di join = $1/100K$ (chiave esterna)

Accesso a P:

Costo scan sequenziale = **1000**

Costo indice su stipendio: $NL \times 0.75 + \Phi(NT \times 0.25, NP) = 2500 \times 0.75 + \Phi(100K \times 0.75, 1K) = 1875 + 1000 = 2875$

Numero tuple residue = $NT \times 0.75 = 75000$

Accesso a R:

Costo scan sequenziale = **1000**

Costo indice su luogo: $NL \times 0.01 + \Phi(NT \times 0.01, NP) = 400 \times 0.01 + \Phi(100K \times 0.01, 1K) = 637$

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

Numero tuple residue = $NT \times 0.25 = 1000$

Costi di join:

P esterna: costo = costo sequenziale + $75000 \times$ costo indice matricola = $1000 + 75000 \times 2 = 151000$

R esterna: costo = costo indice luogo + $1000 \times$ costo sequenziale = $637 + 1000 \times 1000 = 1000637$

Il numero di risultati dell'interrogazione è $100K \times 0.01 \times 0.75 = 750$