

Tecnologie delle Basi di Dati M

Appello del 13/2/2012

Esercizio 1 (2 punti)

Data la relazione con schema:

Studenti(codice, cognome, nome, datanascita, cd1)

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 *sparso* costruito sull'attributo (datanascita). Si supponga di avere pagine di dimensione 4 KB, di cui 96 B riservati per il page header, e si considerino i seguenti valori:

- Numero di tuple = 30K
- Numero di chiavi (datanascita) = 2K
- Dimensione codice = 12 byte
- Dimensione cognome = 32 byte
- Dimensione nome = 16 byte
- Dimensione datanascita = 4 byte
- Dimensione cd1 = 2 byte
- Dimensione RID = 6 byte
- Dimensione PID = 4 byte
- Percentuale di riempimento foglie = 80%

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<20000
      AND R.luogo in ('Milano', 'Roma', 'Napoli')
```

tenendo conto che dai cataloghi della base di dati risulta:

- Numero di tuple Personale = 24K
- Numero di pagine Personale = 1.5K
- Numero di responsabili = 200
- Indice unclustered (TID ordinate) su luogo: numero foglie = 20, numero chiavi = 200
- Indice clustered su stipendio: numero foglie = 600, valore minimo = 10000, valore massimo = 40000
- Indice unclustered su matricola: numero foglie = 500

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 = 1500:

R	$\Phi(R, P)$
20	19.87384
40	39.48436
60	58.83507
80	77.92938
100	96.77072
120	115.3624
140	133.7078
160	151.8101
180	169.6726
200	187.2984

R	$\Phi(R, P)$
220	204.6907
240	221.8525
260	238.787
280	255.4971
300	271.9858
320	288.256
340	304.3107
360	320.1526
380	335.7847
400	351.2097

R	$\Phi(R, P)$
420	366.4302
440	381.4492
460	396.2691
480	410.8927
500	425.3225
520	439.5611
540	453.6111
560	467.475
580	481.1551
600	494.654

Esercizio 3 (5 punti)

Si confrontino i protocolli 2PL e Strict 2PL, specificando in particolare i motivi per cui il secondo è da preferirsi rispetto al primo.

Esercizio 4 (3 punti)

Si illustri quale sia l'effetto di incrementare/decrementare la dimensione della pagina dati su una (o più) struttura hash a scelta, discutendone in particolare l'impatto sui costi di inserimento e ricerca.

Soluzione Esercizio 1

Dimensionamento relazione:

Dimensione di ogni tupla = $12 + 32 + 16 + 4 + 2 = 66\text{B}$

Numero di tuple per pagina = $(4096 - 96)/66 = 4000/66 = 60$

Numero di pagine della relazione = $NT/60 = 30000/60 = 500$

Dimensionamento indice (datanascita):

Numero di chiavi = $2K$, mediamente ci sono $30K/2K = 15$ tuple per ogni valore di chiave.

Dimensione di ogni record (foglia/nodo interno) = $4 + 4 = 8\text{B}$

Dimensione "reale" foglia = $(4096 - 96) \times 0.80 = 3200\text{B}$

Numero di record per foglia = $3200/8 = 400$

Numero di foglie = $500/400 = 2$

Numero nodi livello 1 = $2 \times 8/4000 = 1$

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

Soluzione Esercizio 2

Selettività dei predicati:

Predicato $P.stipendio < 20000 = (20000 - 10000)/(40000 - 10000) = 0.33$

Predicato su luogo = $1/200 = 0.005$ per ogni valore di luogo

Predicato di join = $1/24K$ (chiave esterna)

Accesso a P:

Costo scan sequenziale = **1500**

Costo indice su stipendio: $NL \times 0.33 + NP \times 0.33 = 600 \times 0.33 + 1500 \times 0.33 = 200 + 500 = 700$

Numero tuple residue = $NT \times 0.33 = 8000$

Accesso a R:

Costo scan sequenziale = **1500**

Costo indice su luogo: $3 \times (NL \times 0.005 + \Phi(NT \times 0.005, NP)) = 3 \times (20 \times 0.005 + \Phi(24K \times 0.005, 1.5K)) = 3 \times (1 + \Phi(120, 1500)) = 3 \times (1 + 116) = 351$

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

Numero tuple residue = $3 \times NT \times 0.005 = 360$

Costi di join:

P esterna: costo = costo indice su stipendio + $8000 \times$ costo indice matricola = $700 + 8000 \times 2 = 16700$

R esterna: costo = costo indice luogo + $360 \times$ costo indice su stipendio = $351 + 360 \times 700 = 252351$

Il numero di risultati dell'interrogazione è $24K \times 0.015 \times 0.33 = 120$