

Tecnologie delle Basi di Dati M

Appello del 22/7/2011

Esercizio 1 (2 punti)

Data la relazione con schema:

```
Esami(matricola, codice, data, voto)
```

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 (data). 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 = 320K
- Numero di chiavi (data) = 5K
- Dimensione matricola = 12 byte
- Dimensione codice = 8 byte
- Dimensione data = 4 byte
- Dimensione voto = 1 byte
- Dimensione RID = 4 byte
- Dimensione PID = 2 byte
- Percentuale di riempimento foglie = 65%

Esercizio 2 (5 punti)

Date le relazioni con schema:

```
Esami(matricola, codice, data, voto)
```

```
Studenti(matricola, nome, cognome, data, luogo, cdl)
```

si ottimizzi l'esecuzione della seguente interrogazione SQL:

```
SELECT S.nome, S.cognome, E.voto
FROM Studenti S, Esami E
WHERE S.matricola=E.matricola
      AND E.data >= 1-1-2011
      AND S.cdl = 0937 OR S.cdl = 0234
```

tenendo conto che dai cataloghi della base di dati risulta:

- Numero di tuple Esami = 320K
- Numero di pagine Esami = 200
- Numero di tuple Studenti = 20K
- Numero di pagine Studenti = 2K
- Indice unclustered (TID ordinate) su E.data: numero foglie = 50, numero anni = 10
- Indice clustered su (E.matricola, E.codice): numero foglie = 20, numero corsi = 100
- Indice unclustered su S.matricola: numero foglie = 200
- Indice clustered su cdl: numero foglie = 5, numero corsi di laurea = 20

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 = 200:

R	$\Phi(R, P)$
50	44.33749
100	78.84591
150	105.7043
200	126.6084
250	142.8784
300	155.5416
350	165.3974
400	173.0684
450	179.0388
500	183.6856

R	$\Phi(R, P)$
550	187.3023
600	190.1172
650	192.3081
700	194.0133
750	195.3405
800	196.3734
850	197.1774
900	197.8031
950	198.2902
1000	198.6692

R	$\Phi(R, P)$
1050	198.9642
1100	199.1938
1150	199.3726
1200	199.5117
1250	199.6199
1300	199.7042
1350	199.7698
1400	199.8208
1450	199.8605
1500	199.8914

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)

Supponendo di dovere ottimizzare un'interrogazione su due relazioni in cui il predicato di join sia del tipo $A.attr < B.attr$, si indichi come dovrebbero essere modificati i vari algoritmi di join conosciuti, precisando quali di essi non risultino applicabili/efficienti.

Soluzione Esercizio 1

Dimensionamento relazione:

Dimensione di ogni tupla = $12 + 8 + 4 + 1 = 25B$

Numero di tuple per pagina = $(4096 - 96)/25 = 4000/25 = 160$

Numero di pagine della relazione = $NT/160 = 320000/160 = 2000$

Dimensionamento indice (data):

Numero di chiavi = 5K, mediamente ci sono $320K/5K = 64$ tuple per ogni valore di chiave.

Dimensione di ogni record (foglia) = $4 + 64 \times 4 = 260B$

Dimensione "reale" foglia = $(4096 - 96) \times 0.65 = 2600B$

Numero di record per foglia = $2600/260 = 10$

Numero di foglie = $5K/10 = 500$

Dimensione di ogni record (nodo interno) = $4 + 2 = 6B$

Numero nodi livello 1 = $500 \times 6/4000 = 1$

Il B⁺-tree corrispondente si compone quindi di 2 livelli per un totale di 1 nodo interno (radice) e 500 foglie.

Soluzione Esercizio 2

Selettività dei predicati:

Predicato $E.data \geq 1-1-2011 = 1/10 = 0.1$

Predicato $S.cd1 = 0937 = 1/20 = 0.05$ per ogni valore di cd1

Predicato di join = $1/20K$ (chiave esterna)

Accesso a S:

Costo scan sequenziale = **2000**

Costo indice su cd1: $2 \times (NL \times 0.05 + NP \times 0.05) = 2 \times (5 \times 0.05 + 2000 \times 0.05) = 2 \times (1 + 100) = 202$

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

Numero tuple residue = $2 \times NT \times 0.05 = 2000$

Accesso a E:

Costo scan sequenziale = **200**

Costo indice su data: $NL \times 0.1 + \Phi(NT \times 0.1, NP) = 50 \times 0.1 + \Phi(320K \times 0.1, 200) = 5 + \Phi(32000, 200) = 5 + 200 = 205$

Costo indice su (E.matricola, E.codice): $NL/20K + NP/20K = 1 + 1 = 2$

Numero tuple residue = $NT \times 0.1 = 32000$

Costi di join:

S esterna: costo = costo indice su cd1 + $2000 \times$ costo indice matricola = $202 + 2000 \times 2 = 4202$

E esterna: costo = costo sequenziale + $32000 \times$ costo indice su matricola = $200 + 32000 \times 2 = 64200$

Il numero di risultati dell'interrogazione è $320K \times 0.1 \times 0.1 = 3200$