

Tecnologie delle Basi di Dati M

Appello del 16/1/2013

Esercizio 1 (2 punti)

Data la relazione con schema:

Personale(codice, cognome, nome, 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 *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 = 100K
- Numero di chiavi (datanascita) = 20K
- Dimensione codice = 6 byte
- Dimensione cognome = 32 byte
- Dimensione nome = 32 byte
- Dimensione datanascita = 10 byte
- Dimensione RID = 8 byte
- Dimensione PID = 6 byte
- Percentuale di riempimento foglie = 80%

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.voto = '30L'
      AND S.luogo in ('Bologna', 'Ferrara', 'Modena')
```

tenendo conto che dai cataloghi della base di dati risulta:

- Numero di tuple Esami = 320K
- Numero di pagine Esami = 500
- Numero di tuple Studenti = 20K
- Numero di pagine Studenti = 1K
- Indice unclustered (TID ordinate) su E.voto: numero foglie = 160, numero voti = 16
- Indice clustered su (E.matricola, E.codice): numero foglie = 50, numero corsi = 100
- Indice clustered su S.matricola: numero foglie = 100
- Indice unclustered su luogo: numero foglie = 100, numero luoghi = 1K

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 = 500:

R	$\Phi(R, 500)$
50	47.63
100	90.72
150	129.70
200	164.97
250	196.89
300	225.76
350	251.88
400	275.52
450	296.90
500	316.24

R	$\Phi(R, 500)$
550	333.75
600	349.58
650	363.91
700	376.87
750	388.60
800	399.21
850	408.81
900	417.50
950	425.36
1000	432.47

R	$\Phi(R, 500)$
1050	438.90
1100	444.72
1150	449.99
1200	454.75
1250	459.06
1300	462.96
1350	466.49
1400	469.68
1450	472.57
1500	475.18

Esercizio 3 (5 punti)

Si illustri il funzionamento di un DBMS basato sul protocollo ARIES nella fase di restart seguita ad un malfunzionamento di tipo "system failure".

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 = $6 + 32 + 32 + 10 = 80\text{B}$

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

Numero di pagine della relazione = $NT/50 = 100000/50 = 2000$

Dimensionamento indice (datanascita):

Numero di chiavi = 20K, mediamente ci sono $100\text{K}/20\text{K} = 5$ tuple per ogni valore di chiave.

Dimensione di ogni record (foglia/nodo interno) = $6 + 10 = 16\text{B}$

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

Numero di record per foglia = $3200/16 = 200$

Numero di foglie = $2000/200 = 10$

Numero nodi livello 1 = $10 \times 16/4000 = 1$

Il B⁺-tree corrispondente si compone quindi di 2 livelli per un totale di 1 nodo interno (radice) e 10 foglie.

Soluzione Esercizio 2

Selettività dei predicati:

Predicato E.voto = '30L' = $1/16 = 0.0625$

Predicato S.luogo in ('Bologna', 'Ferrara', 'Modena') = $1/1000 = 0.001$ per ogni valore di luogo

Predicato di join = $1/20\text{K}$ (chiave esterna)

Accesso a S:

Costo scan sequenziale = **1000**

Costo indice su luogo: $3 \times (NL \times 0.001 + NT \times 0.001) = 3 \times (100 \times 0.001 + 20000 \times 0.001) = 3 \times (1 + 20) = 63$

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

Numero tuple residue = $3 \times NT \times 0.001 = 60$

Accesso a E:

Costo scan sequenziale = **500**

Costo indice su voto: $NL \times 0.0625 + \Phi(NT \times 0.0625, NP) = 160 \times 0.0625 + \Phi(320\text{K} \times 0.0625, 500) = 10 + \Phi(20000, 500) = 10 + 500 = 510$

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

Numero tuple residue = $NT \times 0.0625 = 20000$

Costi di join:

S esterna: costo = costo indice su luogo + $60 \times$ costo indice matricola = $63 + 60 \times 2 = 183$

E esterna: costo = costo sequenziale + $20000 \times$ costo indice su matricola = $500 + 20000 \times 2 = 40500$

Il numero di risultati dell'interrogazione è $320\text{K} \times 0.0625 \times 3 \times 0.001 = 60$