

Tecnologie delle Basi di Dati M

Appello del 28/6/2012

Esercizio 1 (2 punti)

Data la relazione con schema:

Personale(codice, cognome, nome, datanascita, dipartimento)

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 (cognome). Si supponga di avere pagine di dimensione 8 KB, di cui 192 B riservati per il page header, e si considerino i seguenti valori:

- Numero di tuple = 62.5K
- Numero di chiavi (cognome) = 12.5K
- Dimensione codice = 4 byte
- Dimensione cognome = 32 byte
- Dimensione nome = 16 byte
- Dimensione datanascita = 10 byte
- Dimensione dipartimento = 2 byte
- Dimensione RID = 6 byte
- Dimensione PID = 4 byte
- Percentuale di riempimento foglie = 90%

Esercizio 2 (5 punti)

Date le relazioni con schema:

Studenti(matricola, nome, data, luogo)

Corsi(codice, nome, cdl)

Voti(matricola, codice, data, voto)

si ottimizzi l'esecuzione della seguente interrogazione SQL (si considerino solamente i piani che verrebbero generati da un ottimizzatore che utilizzi unicamente piani *left-deep*):

```
SELECT S.nome, C.nome, V.voto, V.data
FROM Studenti S, Voti V, Corsi C
WHERE S.matricola=V.matricola
      AND C.codice=V.codice
      AND S.data BETWEEN '1/1/1973' AND '31/12/76'
      AND C.cdl in ('IngInf', 'IngGest')
```

tenendo conto che dai cataloghi della base di dati risulta:

- Numero di tuple *Studenti* = 10K, numero di pagine *Studenti* = 1K
- Numero di tuple *Corsi* = 100, numero di pagine *Corsi* = 10
- Numero di tuple *Voti* = 100K, numero di pagine *Voti* = 5K
- Indice unclustered su *S.matricola*: numero foglie = 100
- Indice clustered su *S.data*: numero foglie = 40, valore minimo = 1/1/75, valore massimo = 31/12/94
- Indice unclustered (TID ordinate) su *C.cdl*: numero foglie = 10, numero chiavi = 10
- Indice unclustered su *C.codice*: numero foglie = 4
- Indice unclustered (TID ordinate) su *V.matricola*: numero foglie = 200
- Indice unclustered (TID ordinate) su *V.codice*: numero foglie = 100

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 rispettivamente per $P = 10$ e $P = 5000$:

R	$\Phi(R, 10)$
2	1.9
4	3.439
6	4.68559
8	5.695328
10	6.513216
12	7.175705
14	7.712321
16	8.14698
18	8.499054
20	8.784233

R	$\Phi(R, 5000)$
100	99.0164
200	196.072
300	291.2056
400	384.4552
500	475.8582
600	565.451
700	653.2697
800	739.3492
900	823.7241
1000	906.4281

R	$\Phi(R, 5000)$
1100	987.4943
1200	1066.955
1300	1144.842
1400	1221.187
1500	1296.02
1600	1369.371
1700	1441.269
1800	1511.744
1900	1580.823
2000	1648.534

Esercizio 3 (5 punti)

Illustrare le differenze tra i seguenti tipi di indici: denso/sparso, clustered/unclustered, primario/secondario. Si indichi quali combinazioni di tali indici non sono ammissibili.

Esercizio 4 (3 punti)

Descrivere la realizzazione della scrittura di una pagina, il commit e l'abort in un sistema che adotti una politica *noUndo-Redo*.

Soluzione Esercizio 1

Dimensionamento relazione:

Dimensione di ogni tupla = $4 + 32 + 16 + 10 + 2 = 64B$

Numero di tuple per pagina = $(8192 - 192)/64 = 8000/64 = 125$

Numero di pagine della relazione = $NT/125 = 62500/125 = 500$

Dimensionamento indice (cognome):

Numero di chiavi = 5K, mediamente ci sono $62.5K/12.5K = 5$ tuple per ogni valore di chiave.

Dimensione di ogni record (foglia/nodo interno) = $4 + 32 = 36B$

Dimensione "reale" foglia = $(8192 - 192) \times 0.90 = 7200B$

Numero di record per foglia = $7200/36 = 200$

Numero di foglie = $500/200 = 3$

Numero nodi livello 1 = $3 \times 36/8000 = 1$

Il B⁺-tree corrispondente si compone quindi di 2 livelli per un totale di 1 nodo interno (radice) e 3 foglie.

Soluzione Esercizio 2

Selettività dei predicati:

Predicato $S.data = (77-75)/(95-75) = 0.1$

Predicato su $cd1 = 1/10 = 0.1$ per ogni valore di $cd1$

Predicato di join su $matricola = 1/10K$ (chiave esterna)

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

Accesso a S:

Costo scan sequenziale = **1000**

Costo indice su $data: NL \times 0.1 + NP \times 0.1 = 40 \times 0.1 + 1000 \times 0.1 = 4 + 100 = 104$

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

Numero tuple residue = $NT \times 0.1 = 1000$

Accesso a C:

Costo scan sequenziale = **10**

Costo indice su $cd1: 2 \times (NL \times 0.1 + \Phi(NT \times 0.1, NP)) = 2 \times (10 \times 0.1 + \Phi(100 \times 0.1, 10)) =$

$2 \times (1 + \Phi(10, 10)) = 2 \times (1 + 7) = 16$

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

Numero tuple residue = $2 \times NT \times 0.1 = 20$

Accesso a V:

Costo scan sequenziale = **5000**

Costo indice su $matricola: NL/10K + \Phi(NT/10K, NP) = 200/10K + \Phi(100K/10K, 5K) =$

$1 + 10 = 11$

Costo indice su $codice: NL/100 + \Phi(NT/100, NP) = 100/100 + \Phi(100K/100, 5K) =$

$1 + 907 = 908$

Costi di join:

S-V-C: costo = costo indice su $S.data + 1000 \times$ costo indice $V.matricola + 1000 \times 10 \times$ costo indice $C.codice = 104 + 1000 \times 11 + 10000 \times 2 = 31104$

C-V-S: costo = costo indice su $cd1 + 20 \times$ costo indice $V.codice + 20 \times 1000 \times$ costo indice

$S.matricola = 10 + 20 \times 908 + 20000 \times 2 = 58170$

I piani con V esterna hanno tutti costi maggiori poiché non hanno selettività sul risultato.

Il numero di risultati dell'interrogazione è $100K \times 0.1 \times 0.2 = 2000$