

Tecnologie delle Basi di Dati M

Appello del 1/2/2011

Esercizio 1 (3 punti)

Date le relazioni con schema:

Corsi(codice, nome, corsodilaurea, docente)
CorsiDiLaurea(codice, presidente, sede)
Docenti(matricola, nome, dipartimento)

si supponga di dover ottimizzare la seguente interrogazione SQL:

```
SELECT C.nome, D.nome
FROM Corsi C, CorsiDiLaurea L, Docenti D
WHERE C.corsodilaurea=L.codice
      AND C.docente=D.matricola
      AND D.matricola=L.presidente
      AND L.sede="Bologna"
      AND D.dipartimento="DEIS"
```

Si indichino i piani di accesso che verrebbero generati da un ottimizzatore che utilizzi unicamente piani *left-deep*. In ogni piano si indichi anche quale sia l'attributo coinvolto nell'operazione di join. (Suggerimento: piani di accesso che presentino lo stesso ordine di accesso alle relazioni ma attributi di join diversi sono da considerarsi come piani diversi).

Esercizio 2 (4 punti)

Data la relazione con schema:

Dipartimenti(nome, sede, budget, direttore)

si ottimizzi l'esecuzione della seguente interrogazione SQL:

```
SELECT *
FROM Dipartimenti
WHERE sede IN ('Milano', 'Bologna', 'Roma', 'Firenze', 'Napoli')
      AND budget BETWEEN 1000 AND 2000
      AND direttore LIKE 'M%'
```

tenendo conto che dai cataloghi della base di dati risulta:

- Numero di tuple Dipartimenti = 1K
- Numero di pagine Dipartimenti = 50
- Numero di chiavi per l'attributo sede = 20
- Indice clustered su sede: numero foglie = 10, numero chiavi = 20
- Indice unclustered (TID ordinate) su budget: numero foglie = 25, valore minimo = 0, valore massimo = 10000
- Indice unclustered (TID ordinate) su direttore: numero foglie = 40, i valori iniziano con una lettera dell'alfabeto italiano (non per H).

Si stimi infine il numero di risultati dell'interrogazione.

Suggerimento: per la formula di Cardenas si utilizzino i seguenti valori, validi per $P = 50$:

R	$\Phi(R, P)$
5	4.80396
10	9.14636
15	13.07154
20	16.6196
25	19.82676
30	22.72578
35	25.34627
40	27.71498
45	29.85611
50	31.79152

R	$\Phi(R, P)$
55	33.54097
60	35.12234
65	36.55178
70	37.84387
75	39.01182
80	40.06756
85	41.02186
90	41.88447
95	42.6642
100	43.36902

R	$\Phi(R, P)$
105	44.00612
110	44.58201
115	45.10256
120	45.57311
125	45.99844
130	46.38291
135	46.73043
140	47.04457
145	47.32853
150	47.5852

Esercizio 3 (5 punti)

Si descriva l'algoritmo di inserimento di oggetti in un R-tree, stimando i costi attesi, compreso il caso di split di nodi dell'albero.

Esercizio 4 (3 punti)

Discutere quale potrebbe essere l'impatto dell'utilizzo di politiche "force" e "no steal" all'interno del buffer manager, considerando sistemi aventi differenti frequenze di aggiornamento dei dati e di accessi concorrenti.

Soluzione Esercizio 1

I piani che vengono generati sono i seguenti:

1. $(C \triangleright \triangleleft_{(C.docente=D.matricola)D}) \triangleright \triangleleft_{(C.corsodilaurea=L.codice)L}$
2. $(C \triangleright \triangleleft_{(C.docente=D.matricola)D}) \triangleright \triangleleft_{(D.matricola=L.presidente)L}$
3. $(D \triangleright \triangleleft_{(C.docente=D.matricola)C}) \triangleright \triangleleft_{(C.corsodilaurea=L.codice)L}$
4. $(D \triangleright \triangleleft_{(C.docente=D.matricola)C}) \triangleright \triangleleft_{(D.matricola=L.presidente)L}$
5. $(C \triangleright \triangleleft_{(C.corsodilaurea=L.codice)L}) \triangleright \triangleleft_{(C.dipartimento=D.codice)D}$
6. $(C \triangleright \triangleleft_{(C.corsodilaurea=L.codice)L}) \triangleright \triangleleft_{(L.presidente=D.matricola)D}$
7. $(L \triangleright \triangleleft_{(C.corsodilaurea=L.codice)C}) \triangleright \triangleleft_{(C.dipartimento=D.codice)D}$
8. $(L \triangleright \triangleleft_{(C.corsodilaurea=L.codice)C}) \triangleright \triangleleft_{(L.presidente=D.matricola)D}$
9. $(L \triangleright \triangleleft_{(L.presidente=D.matricola)D}) \triangleright \triangleleft_{(C.docente=D.matricola)C}$
10. $(L \triangleright \triangleleft_{(L.presidente=D.matricola)D}) \triangleright \triangleleft_{(C.corsodilaurea=L.codice)C}$
11. $(D \triangleright \triangleleft_{(L.presidente=D.matricola)L}) \triangleright \triangleleft_{(C.docente=D.matricola)C}$
12. $(D \triangleright \triangleleft_{(L.presidente=D.matricola)L}) \triangleright \triangleleft_{(C.corsodilaurea=L.codice)C}$

Soluzione Esercizio 2

Selettività dei predicati:

Predicato su sede = $1/20 = 0.05$ per ogni valore di sede

Predicato su budget = $(2000 - 1000)/(10000 - 0) = 0.1$

Predicato su direttore = $1/20 = 0.05$

Accesso a Dipartimenti:

Costo scan sequenziale = **50**

Costo indice su sede: $5 \times (NL \times 0.05 + NP \times 0.05) = 5 \times (10 \times 0.05 + 50 \times 0.05) = 5 \times (1 + 3) = \mathbf{20}$

Costo indice su budget: $NL \times 0.1 + \Phi(NT \times 0.1, NP) = 25 \times 0.1 + \Phi(1K \times 0.1, 50) = 3 + \Phi(100, 50) = 3 + 44 = \mathbf{47}$

Costo indice su direttore: $NL \times 0.05 + \Phi(NT \times 0.05, NP) = 40 \times 0.05 + \Phi(1K \times 0.05, 50) = 2 + \Phi(50, 50) = 2 + 32 = \mathbf{34}$

Il numero di risultati dell'interrogazione è $1000 \times 0.25 \times 0.1 \times 0.05 = \mathbf{1.25}$