

# Basi di dati, primo modulo

6 febbraio 2003

Tempo a disposizione: un'ora e trenta minuti. Libri chiusi.

## Domanda 1 (25%)

Considerare i seguenti schedule:

1.  $r_1(y)r_2(z)r_2(y)w_2(y)w_2(z)r_1(z)$
2.  $r_2(x)w_2(x)r_1(x)w_1(x)$
3.  $r_1(x)r_2(z)w_2(z)w_1(x)r_2(x)w_2(x)$
4.  $r_2(x)w_2(x)r_3(x)w_3(x)r_1(y)w_2(y)$
5.  $w_1(z)r_2(z)r_3(z)w_3(z)w_3(y)$

Specificare, con una breve giustificazione, a quali delle seguenti classi ciascuno di essi appartiene: S (seriale), VSR (view-serializzabile), CSR (conflict-serializzabile), 2PL (generabile da uno scheduler basato sul lock a due fasi) e TS (generabile da uno scheduler che utilizzi il metodo dei timestamp; si assuma che l'ordinamento degli identificatori delle transazioni corrisponda a quello dei timestamp).

## Domanda 2 (25%) Considerare la seguente interrogazione in SQL:

```
SELECT DISTINCT C, L
FROM R1, R2, R3
WHERE R1.C = R2.D AND R2.F=R3.G AND R1.B=R3.L AND R3.H>10
```

Mostrare un possibile piano di esecuzione (in termini di operatori dell'algebra relazionale e loro realizzazioni; prestare attenzione anche alla DISTINCT, in quanto le realizzazioni degli operatori non producono necessariamente insiemi, ma liste di ennuple), giustificando brevemente le scelte più significative, con riferimento alle seguenti informazioni sulla base di dati:

- la relazione  $R_1(\underline{ABC})$  ha 100.000 ennuple, una struttura heap e un indice secondario su  $C$ ;
- la relazione  $R_2(\underline{DEF})$  ha 30.000 ennuple, una struttura heap e un indice secondario sulla chiave  $D$ ;
- la relazione  $R_3(\underline{GHL})$  ha 10.000 ennuple, una struttura heap e un indice secondario sulla chiave  $G$ .

## Domanda 3 (35%)

Si consideri una relazione IMPIEGATO(Matricola,Cognome,Nome,DataNascita) con un numero di ennuple pari a  $N$  abbastanza stabile nel tempo e una dimensione di ciascuna ennupla (a lunghezza fissa) pari a  $L$  byte, di cui  $K$  per la chiave.

Supporre di avere a disposizione un DBMS che permetta strutture fisiche disordinate (heap) e hash e che preveda la possibilità di definire indici secondari e operi su un sistema operativo che utilizza blocchi di dimensione  $B$  e con puntatori ai blocchi di  $P$  caratteri.

Siano le seguenti le operazioni principali:

1. ricerca esatta sul numero di matricola con frequenza  $f_1$
2. ricerca sul cognome (anche su sottostringa iniziale) con frequenza  $f_2$ ; mediamente una richiesta restituisce 4 record.

Individuare alcune (almeno una) possibili organizzazione fisica per tale relazione e calcolare (approssimativamente) il numero di accessi a memoria secondaria (nell'unità di tempo) supponendo  $N = 5.000.000$ ,  $L = 125$ ,  $K = 5$ ,  $B = 1.000$ ,  $P = 4$ ,  $f_1 = 100$ ,  $f_2 = 1.000$ .

## Domanda 4 (15%)

Illustrare brevemente la tecnica chiamata "predicate lock" e spiegare perché è importante e come può essere realizzata.