

Basi di dati, primo modulo

14 novembre 2002

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

Domanda 1 (25%)

Considerare i seguenti schedule:

1. $r_1(y)r_4(z)r_4(y)w_4(y)w_4(z)r_1(z)r_2(z)r_3(z)w_3(z)$
2. $w_1(z)r_2(z)r_3(z)w_3(z)w_3(y)$
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. $r_4(z)r_4(y)w_4(y)w_4(z)r_2(x)w_2(x)r_1(x)w_1(x)$

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 (30%)

Alcuni DBMS permettono una tecnica di memorizzazione chiamata "co-clustering" o "clustering eterogeneo," in cui un file contiene record di due o più relazioni e tali record sono raggruppati (eventualmente, ma non necessariamente, ordinati) secondo i valori di opportuni campi dell'una e dell'altra relazione. Ad esempio, date due relazioni

- *Ordini*(CodiceOrdine, Cliente, Data, ImportoTotale)
- *LineeOrdine*(CodiceOrdine, Linea, Prodotto, Quantità, Importo)

questa tecnica (con riferimento agli attributi *CodiceOrdine* delle due relazioni) permetterebbe una memorizzazione contigua di ciascun ordine con le rispettive "linee d'ordine," cioè dei prodotti ordinati (ciascun ordine fa riferimento a più prodotti, ognuno su una "linea").

Con riferimento all'esempio, indicare quali delle seguenti operazioni possono trarre vantaggio dall'uso di questa opportunità e quali ne possono essere penalizzate (spiegare la risposta anche in termini quantitativi, individuando valori opportuni per i principali parametri di interesse; supporre che siano utilizzati indici su *CodiceOrdine*, in tutti i casi, due per la memorizzazione tradizionale e uno nel caso di utilizzo del cluster eterogeneo):

1. stampa dei dettagli (cioè delle linee d'ordine) di tutti gli ordini (ordinati per codice)
2. stampa dei dettagli di un ordine
3. stampa delle informazioni sintetiche (codice, cliente, data, totale) di tutti gli ordini

Domanda 3 (20%)

Discutere brevemente le ragioni che rendono non direttamente utilizzabili, in pratica, le tecniche di controllo di concorrenza basate sulla view-serializzabilità e la conflict-serializzabilità e quelle che invece rendono utilizzabili il 2PL e le tecniche basate su timestamp. Discutere anche la relazione che esiste fra le varie tecniche (quali sono più lasche e quali più restrittive).

Domanda 4 (25%)

Considerare la seguente interrogazione in SQL:

```
SELECT A, H
FROM R, S, T
WHERE C = D AND F = G AND H > 20
```

definita con riferimento a tre relazioni, definite e frammentate come segue (per essere poi distribuite):

- $R(A, B, C)$ frammentata orizzontalmente:
 - $R_a = \sigma_{C > 100}(R)$
 - $R_b = \sigma_{C \leq 100}(R)$
- $S(D, E, F)$ frammentata orizzontalmente:
 - $S_a = \sigma_{D > 120}(S)$
 - $S_b = \sigma_{D \leq 120}(S)$
- $T(G, H, L)$ frammentata verticalmente:
 - $T_a = \pi_{G, H}(T)$
 - $T_b = \pi_{H, L}(T)$

Mostrare (ad esempio sotto forma di albero) l'espressione dell'algebra relazionale definita sui frammenti che realizza in modo efficiente tale interrogazione.