

Basi di dati, primo modulo

29 gennaio 2002

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

Domanda 1 (30%)

Come noto, 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 allocati (ad esempio ordinati) secondo i valori di opportuni campi dell'una e dell'altra relazione. Ad esempio, date due relazioni

- *Cliente*(PartitaIVA, RagioneSociale, Indirizzo)
- *Fatture*(Numero, Data, PartitaIVACliente, Importo)

questa tecnica (con riferimento agli attributi *PartitaIVA* di *Cliente* e *PartitaIVACliente* di *Fatture*) permetterebbe una memorizzazione contigua di ciascun cliente con le relative fatture.

Si supponga che le operazioni di interesse siano

1. stampa dell'elenco di tutti i clienti
2. stampa delle fatture di un cliente
3. stampa di tutti i clienti, ciascuno con le relative fatture

Spiegare, giustificando anche in termini quantitativi (con opportune ipotesi per la dimensione dei record e dei blocchi), quali operazioni potrebbero trarre vantaggio dal clustering e quali svantaggio (o né vantaggio né svantaggio). Ricordare che sulle relazioni possono essere definiti indici primari e secondari sia in presenza sia in assenza di clustering.

Domanda 2 (25%)

Discutere il ruolo che gli indici e le strutture hash possono avere nell'esecuzione dei join nelle basi di dati relazionali.

Domanda 3 (15%)

Indicare quali fra le seguenti affermazioni sono vere per i data warehouse:

1. Le proprietà “acide” sono rilevanti
2. I dati sono, istante per istante, una sintesi dei dati operativi validi in quell'istante
3. Le strutture fisiche cercano di conciliare le esigenze degli aggiornamenti e quelle delle interrogazioni
4. Le operazioni sono complesse e non predefinite
5. I dati hanno un contenuto storico che va al di là delle attività correnti
6. Gli utenti sono prevalentemente di livello più alto nella gerarchia aziendale rispetto a quelli che utilizzano i sistemi OLTP

Domanda 4 (30%)

Considerare i seguenti schedule:

1. $r_1(x)w_1(x)r_2(x)w_2(x)r_0(y)w_1(y)$
2. $r_1(x)w_1(x)r_2(x)w_2(x)r_3(y)w_1(y)$
3. $r_1(x)r_1(y)r_2(z)r_2(y)w_2(y)w_2(z)r_1(z)$
4. $w_0(x)r_2(x)r_1(x)w_2(x)w_2(z)$

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) and TS (generabile da uno scheduler che utilizzi il metodo dei timestamp; si assuma che gli identificatori delle transazioni corrispondano ai timestamp).