

Basi di dati, primo modulo — 15 luglio 2003

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

Domanda 1 (30%) Si consideri la seguente base di dati relazionale, relativa alle vendite presso una catena di supermercati:

- Vendite(NumeroScontrino, CodNegozio, CFCliente, Data)
- ElementiVendita(NumeroScontrino, NumeroLinea, CodArticolo, Quantità)
- Articoli(CodArticolo, Descrizione, CodMarca, Prezzo)
- Clienti(CFCliente, Cognome, Nome, Età, CodCategoria, CittàResidenza)
- Negozi(CodNegozio, Nome, Indirizzo, Città)
- Città(Nome, Provincia)
- Province(Sigla, Regione)
- Marche(CodMarca, Nome, CodNazione)
- Nazioni(CodNazione, Nazione)
- Categorie(CodCategoria, Descrizione)

Costruire, sulla base di esso, un data mart (basato su uno schema a stella) che permetta di analizzare le vendite rispetto a tutte le dimensioni per le quali sono disponibili dati.

Domanda 2 (30%) Siano r_1 ed r_2 due relazioni che occupano rispettivamente $B_1 = 20.000$ e $B_2 = 100$ blocchi.

1. Calcolare il numero di accessi a memoria secondaria necessario per eseguire un join su di esse con il metodo *nested-loop*, supponendo di avere a disposizione un buffer di dimensione pari a 60 blocchi. In particolare, valutare quale delle due relazioni è più conveniente usare nel ciclo esterno e come debbono essere usate le pagine del buffer.
2. Valutare se e quanto la presenza di indici o strutture hash può rendere più efficiente il join discusso al punto precedente.

Domanda 3 (20%) Considerare la seguente interrogazione in SQL:

```
SELECT A, D, H
FROM R, S, T
WHERE E = B AND C = G
```

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

- $R(A, C, E)$ frammentata orizzontalmente:
 - $R_a = \sigma_{C > 100}(R)$
 - $R_b = \sigma_{C \leq 100}(R)$
- $S(B, D, F)$ frammentata verticalmente:
 - $S_a = \pi_{B, D}(S)$
 - $S_b = \pi_{B, F}(S)$
- $T(G, H)$ frammentata orizzontalmente:
 - $T_a = \sigma_{G > 200}(T)$
 - $T_b = \sigma_{G \leq 200}(T)$

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

Domanda 4 (20%) Sia data una relazione Vendite(Libro, Libreria, Data, Quantità) su cui si eseguono interrogazioni volte a calcolare il totale delle vendite in un certo periodo. Indicare (con un breve commento, non più di tre righe) quale livello di isolamento (READ UNCOMMITTED, READ COMMITTED, REPEATABLE READ o SERIALIZABLE) si potrebbe scegliere in ciascuno dei seguenti casi

1. l'interrogazione è eseguita con lo scopo di conoscere indicativamente il numero di copie vendute di ciascun libro, mentre vengono effettuati alcuni inserimenti, relativi ad un insieme di piccole librerie (cioè con poche vendite);
2. l'interrogazione è eseguita, con lo scopo di stilare la classifica delle vendite, mentre vengono inseriti i dati di alcune librerie;
3. l'interrogazione è eseguita, con lo scopo di stilare la classifica delle vendite, mentre vengono corretti i dati di alcune librerie, ma senza aggiungerne di nuovi;
4. l'interrogazione è eseguita in un tempo morto (cioè in assenza di aggiornamenti).