

Tecnologia delle basi di dati (ex Basi di dati, primo modulo) 25 settembre 2006

Tempo a disposizione: due ore. **Nota:** è richiesta una “bella copia” comprensibile e ordinata.

Domanda 1 (25%) Considerare le seguenti richieste ricevute da un gestore del controllo di concorrenza (assumendo che si tratti delle prime richieste ricevute dopo l’avvio del sistema e indicando con c_i il commit della transazione i , che permette il rilascio dei lock da essa acquisiti):

$r_3(x), r_2(x), r_4(y), w_2(x), c_2, r_6(y), r_1(x), c_1, w_3(x), c_3, w_4(y), c_4, w_7(x), c_7, w_6(y), c_6, r_5(x), c_5$

Indicare possibili effetti del controllo della concorrenza (indicare cioè quali operazioni vengono eseguite e in quale ordine) prodotti da controllori dei due tipi principali:

- basato su 2PL; in questo caso supporre che: (a) quando una transazione viene bloccata a causa della mancata concessione di un lock, le sue richieste “rinviate” arrivino poi una dopo l’altra, quando il lock viene concesso; (b) che lo stallo venga immediatamente rilevato e che venga risolto uccidendo la transazione che ha formulato l’ultima delle richieste che hanno causato lo stallo; (c) ogni transazione uccisa per risolvere lo stallo venga riavviata subito e sia in grado di richiedere immediatamente le azioni svolte in precedenza (dopo però le concessioni di lock rese possibili dalla sua uccisione);
- basato su timestamp; in questo caso supporre che (a) l’identificatore della transazione corrisponda, al solito, al timestamp (quindi t_i è più giovane di t_j se e solo se $i > j$); (b) ogni transazione uccisa per risolvere lo stallo venga riavviata subito (con un timestamp opportuno) e sia in grado di richiedere immediatamente le azioni svolte in precedenza.

Domanda 2 (30%) Una catena di negozi gestisce le attività utilizzando, in ciascun negozio, una base di dati con le seguenti relazioni:

- Prodotti(CodiceProdotto,Descrizione,Prezzo,Categoria)
- Categorie(Codice,Descrizione,MacroCategoria)
- MacroCategorie(Codice,Descrizione)
- Vendite(NumeroScontrino,Ora)
- DettaglioVendite(NumeroScontrino,CodiceProdotto,Quantità)

Si noti che

- Le informazioni relative alle vendite vengono mantenute solo nel corso della giornata.
- Il prezzo di un prodotto può variare da un giorno all’altro.

Utilizzando tali informazioni, la catena vuole realizzare un data mart relativo alle vendite dei prodotti, avente come misure le quantità vendute e gli incassi, che permetta di effettuare analisi di tipo temporale (incluse, oltre ai giorni, anche le fasce orarie della giornata, ad esempio 9-10, 10-11 e così via, oppure mattina e pomeriggio) e su prodotti (singoli e per categoria) e sui negozi. Allo scopo:

- specificare un possibile dettaglio del data mart; al riguardo, si supponga che la quantità delle vendite sia tale che si è deciso di non utilizzare il massimo livello di dettaglio, ma solo quello strettamente indispensabile (in altri termini, la grana non deve essere il singolo dettaglio di vendita, ma una opportuna aggregazione; **specificare esplicitamente la grana scelta**)
- specificare l’interrogazione SQL necessaria per produrre, giornalmente, le nuove ennuple da inserire nella tabella dei fatti (utilizzare eventualmente una o più viste per facilitare la scrittura dell’interrogazione)

Domanda 3 (25%) Si considerino un sistema con blocchi di dimensione $B = 1000$ byte e puntatori ai blocchi di $P = 2$ byte e una relazione $R(\underline{A}, B, C, D)$ di cardinalità pari circa a $N = 1.000.000$, con ennuple di $L = 50$ byte e campo chiave A di $L_A = 8$ byte. Valutare i pro e i contro (in termini di numero di accessi a memoria secondaria e trascurando le problematiche relative alla concorrenza e i benefici derivanti dai buffer e dalla contiguità) relativamente alla presenza di un indice secondario sulla chiave A e di un altro, pure secondario, su B (di $L_B = 2$ byte), in presenza del seguente carico applicativo:

- inserimento di una nuova ennupla (con verifica del soddisfacimento del vincolo di chiave), con frequenza $f_1 = 200$
- ricerca di una ennupla sulla base del valore della chiave A , con frequenza $f_2 = 100$
- ricerca di ennuple sulla base del valore di B , con frequenza $f_3 = 500$
- elenco di tutte le ennuple, ordinato secondo sulla base del valore di B , con frequenza $f_4 = 10$

Domanda 4 (20%) Illustrare brevemente le tecniche per la realizzazione dell’operatore di join, commentandone anche la complessità e l’utilizzo delle strutture fisiche.