

# Tecnologia delle basi di dati

## Basi di dati, primo modulo

20 settembre 2004

Tempo a disposizione: due ore. Libri chiusi..

**Nota:** è richiesta una “bella copia” comprensibile e ordinata.

### Domanda 1

Illustrare, brevemente, ma in modo ordinato, le differenze nelle tecniche di implementazione fra i livelli di isolamento **SERIALIZABLE** e **REPEATABLE READ**, facendo anche riferimento alle diverse strutture fisiche, primarie e secondarie, che possono essere coinvolte. Spiegare, quindi, perché le conseguenti differenze di prestazioni possono essere in alcuni casi enormi e in altri relativamente piccole.

### Domanda 2

Si considerino un sistema con blocchi di dimensione  $B = 1000$  byte e puntatori ai blocchi di  $P = 2$  byte e una relazione  $R(\underline{A}, B, C, D, E)$  di cardinalità pari circa a  $N = 1.000.000$ , con ennuple di  $L = 50$  byte e campo chiave  $A$  di  $K = 5$  byte. Valutare i pro e i contro (in termini di numero di accessi a memoria secondaria e trascurando le problematiche relative alla concorrenza) relativamente alla presenza di un indice secondario sulla chiave  $A$  e di un altro, pure secondario, su  $B$ , in presenza del seguente carico applicativo:

1. inserimento di una nuova ennupla (con verifica del soddisfacimento del vincolo di chiave), con frequenza  $f_1 = 100$
2. ricerca di una ennupla sulla base del valore della chiave  $A$  con frequenza  $f_2 = 100$
3. ricerca di ennuple sulla base del valore di  $B$  con frequenza  $f_3 = 500$

### Domanda 3

Una catena di fast-food gestisce le attività utilizzando, presso ciascun locale, una base di dati con le seguenti relazioni:

- Prodotti(CodiceProdotto, Descrizione, Prezzo, Categoria)
- Categorie(CodiceCategoria, Descrizione)
- Vendite(NumeroScontrino, Ora, NomeVenditore)
- DettaglioVendite(NumeroScontrino, CodiceProdotto, Quantità)

Si noti che

- La base di dati è comune ai vari negozi per quanto riguarda codici, descrizioni e categorie dei prodotti, mentre differisce per i prezzi dei prodotti e ovviamente per tutto ciò che riguarda le vendite.
- 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 relativa alle vendite dei prodotti, avente come misure le quantità vendute e gli incassi, che permetta di effettuare analisi di tipo spaziale (vendita nei singoli negozi e nelle aree geografiche, quali città, province e regioni), di tipo temporale (giorni, fasce orarie della giornata, giorni della settimana, mesi, ...) e sui prodotti (singoli e per categoria). Allo scopo:

1. specificare un possibile dettaglio del data mart;
2. individuare quali dati siano necessari nel data mart e non presenti nelle basi di dati operative;
3. specificare i principali processi necessari per l'alimentazione del data mart, con le relative scadenze e le operazioni (ad esempio in termini di istruzioni SQL) da svolgere allo scopo.

### Domanda 4

Considerare le seguenti richieste di lettura e scrittura ricevute da un gestore del controllo di concorrenza basato su timestamp (assumendo che si tratti delle prime richieste ricevute dopo l'avvio del sistema):

$$r_1(y), w_1(y), r_2(x), r_8(x), r_5(x), r_4(v), w_4(v), r_3(v), w_7(x), r_6(x), w_9(x), w_8(x)$$

Indicare quali vengono accettate e quali rifiutate e, di conseguenza, quali transazioni vengono uccise.