

Corso di Basi di dati — Prova scritta — 15 giugno 2000

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

Domanda 1 (30%)

Si consideri una relazione IMPIEGATO(Matricola,Cognome,Nome,DataNascita) con un numero di ennuple pari a N molto variabile nel tempo e una dimensione di ciascuna ennupla (a lunghezza fissa) pari a L byte, di cui K per la chiave.

Supporre di avere a disposizione un DBMS che permetta strutture fisiche disordinate (heap), ordinate (con indice primario sparso) e hash e che preveda la possibilità di definire indici secondari e operi su un sistema operativo che utilizza blocchi di dimensione B e con puntatori ai blocchi di P caratteri.

Calcolare (approssimativamente) il numero di accessi a memoria secondaria (nell'unità di tempo) supponendo per la suddetta relazione varie organizzazioni (quelle che a prima vista si ritengono preferibili) nel caso in cui le operazioni principali siano le seguenti:

1. ricerca sul numero di matricola, con frequenza f_1
2. ricerca sul cognome (o una sua sottostringa iniziale, abbastanza selettiva, in media una sottostringa identifica $S = 4$ ennuple) con frequenza f_2
3. ricerca sulla base di un intervallo della data di nascita (poco selettivo), con frequenza f_3 molto minore di f_1 e f_2

In particolare, indicare la soluzione preferita (assumendo $N = 10.000.000$, $L = 100$, $K = 5$, $B = 1000$ e $P = 4$) nel caso

$$(a) f_1 = 100, f_2 = 1.000, f_3 = 1$$

Domanda 2 (30%)

Mostrare in dettaglio come le tecniche per il controllo di concorrenza basate su locking e su timestamp riescono a impedire il presentarsi della seguente anomalia (in particolare, mostrare lo schedule insoddisfacente e come ciascuna delle due tecniche modifica tale schedule).

1. Un signore ha dieci milioni sul proprio conto corrente e firma due assegni da tre milioni ciascuno. I due beneficiari si presentano quasi contemporaneamente a due impiegati diversi della stessa banca, ciascuno dei quali verifica che i soldi sono disponibili (ci sono dieci milioni) e, pagato l'assegno, registra il nuovo saldo di sette milioni.

Domanda 3 (20%)

Descrivere i principali algoritmi per l'esecuzione dell'operatore di join, discutendone anche la complessità (distinguendo casi diversi a seconda delle strutture fisiche definite sulle relazioni).

Domanda 4 (20%)

Si considerino due transazioni identiche che verificano che non esista un impiegato con una certa matricola e poi lo inseriscono. Spiegare:

1. perché una forma semplice di lock non risolve il problema;
2. quale tecnica permette invece di risolverlo.