

Basi di dati, primo modulo

9 novembre 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 D e con puntatori ai blocchi di P caratteri.

Indicare quale possa essere l'organizzazione fisica preferita nel caso in cui le operazioni principali siano le seguenti:

1. ricerca sul cognome (o una sua sottostringa iniziale, abbastanza selettiva, in media una sottostringa identifica $S = 10$ ennuple) con frequenza f_1
2. ricerca sul numero di matricola, 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

assumendo $N = 10.000.000$, $L = 100$, $K = 5$, $D = 500$, $P = 4$, $f_1 = 100$, $f_2 = 2.000$, $f_3 = 1$

Domanda 2 (25%)

Descrivere il processo di esecuzione e ottimizzazione delle interrogazioni nelle basi di dati relazionali.

Domanda 3 (20%)

Illustrare il protocollo del commit a due fasi utilizzando un esempio della vita quotidiana, quale l'organizzazione di una riunione di lavoro per la quale è prevista la partecipazione di tre persone insieme alla disponibilità di una sala riunioni.

Domanda 4 (25%)

Indicare quali potrebbero essere le differenze principali fra la base di dati utilizzata da una azienda per la gestione quotidiana del magazzino dei materiali di consumo (carta, cancelleria e simili) e quella (basata comunque sui dati della precedente) utilizzata per pianificare gli acquisti del magazzino stesso e per valutare nel tempo l'andamento dei consumi.