

Basi di dati, primo modulo

15 novembre 2001

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

Domanda 1 (35%)

Si consideri una relazione IMPIEGATO(Matricola,Cognome,Nome,DataNascita) con un numero di ennuple pari a N abbastanza stabile 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) 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.

Siano le seguenti le operazioni principali:

1. ricerca sul numero di matricola con frequenza f_1
2. ricerca esatta sul cognome con frequenza f_2

Individuare alcune (almeno una) possibili organizzazione fisica per tale relazione e calcolare (approssimativamente) il numero di accessi a memoria secondaria (nell'unità di tempo) supponendo $N = 5.000.000$, $L = 125$, $K = 5$, $B = 1.000$, $P = 4$, $f_1 = 100$, $f_2 = 1.000$.

Domanda 2 (35%)

Descrivere sinteticamente i principali algoritmi per l'esecuzione dell'operatore di join, indicando anche il numero di accessi a memoria secondaria necessari per la loro esecuzione, in funzione del numero di blocchi dei due file, con riferimento alle varie alternative per le rispettive strutture fisiche.

Domanda 3 (15%)

Può succedere che, dopo aver chiesto ad un agente di viaggio informazioni sulla disponibilità di posti su un volo e avendo ottenuto una risposta affermativa, una successiva richiesta di prenotazione dia esito negativo. La ragione di tale comportamento risiede nel fatto che le due operazioni vengono svolte in transazioni separate. Spiegare perché non conviene riunirle in un'unica transazione.

Domanda 4 (15%)

Indicare quali fra le seguenti affermazioni sono vere per i data warehouse:

1. Le attività supportate sono quelle quotidiane (ad esempio la vendita e le attività di sportello)
2. Gli utenti sono prevalentemente di livello più alto nella gerarchia aziendale rispetto a quelli che utilizzano i sistemi OLTP
3. Le proprietà "acide" non sono rilevanti
4. I dati sono, istante per istante, una sintesi dei dati operativi
5. Le strutture fisiche cercano di conciliare le esigenze degli aggiornamenti e quelle delle interrogazioni
6. Le operazioni sono complesse e non predefinite