

Basi di dati, primo modulo

Prova di autovalutazione — 5 maggio 2004

Note

- Gli homework sono facoltativi ma è estremamente importante svolgerli (e anche discuterli ad esempio sul forum), perché le domande proposte nei compiti di esame possono essere molto simili.
- Questo homework verrà discusso in aula nell'esercitazione prevista per il giorno 14 maggio (proseguibile se necessario il 17 o il 18). La discussione dovrà avvenire attraverso una presentazione delle soluzioni fatta da studenti (e non da parte del docente, che tutt'al più potrà commentarle).
- Studenti interessati a sviluppare in dettaglio (ma comunque entro il 14 maggio) la soluzione al primo quesito sono invitati a contattare il docente, che potrà prendere in considerazione la possibilità di considerare tale attività ai fini dell'esame.

Domanda 1

Sperimentare le strutture fisiche messe a disposizione da DB2 ed esaminare il processo di ottimizzazione. In particolare:

1. individuare (attraverso i manuali, l'help ed eventualmente documenti disponibili in rete) le strutture fisiche messe a disposizione da DB2
2. esaminare il comportamento del sistema rispetto ad alcune interrogazioni interessanti (ad esempio varianti di quelle discusse a lezione, sia quella "simbolica" sui lucidi principali sia quelle nei lucidi sul tuning), avendo definito opportune strutture fisiche e avendo costruito relazioni di esempio sufficientemente grandi; cercare anche di mostrare come al variare di parametri, strutture fisiche e dimensioni, possa variare il piano di accesso scelto dal sistema.

Domanda 2

Valutare il costo, in termini di numero di accessi a memoria secondaria (senza considerare gli eventuali benefici derivanti dalla gestione dei buffer), dell'esecuzione di una operazione di join con il metodo merge scan di due relazioni R_1 e R_2 , assumendo:

- blocchi di dimensione $B = 1000$
- puntatori ai blocchi dimensione $P = 2$
- record di dimensione rispettivamente $L_1 = 40$ e $L_2 = 50$
- attributo di join di dimensione $A = 4$
- cardinalità rispettivamente $N_1 = 200.000$ e $N_2 = 1.000.000$

in ciascuno seguenti casi (gli indici citati sono relativi all'attributo di join)

1. indice primario per entrambe le relazioni
2. indice primario per R_1 e secondario per R_2
3. indice secondario per entrambe le relazioni

Domanda 3

Alcuni DBMS permettono una tecnica di memorizzazione chiamata "co-clustering" o "clustering eterogeneo," in cui un file contiene record di due o più relazioni e tali record sono allocati (ad esempio ordinati) secondo i valori di opportuni campi dell'una e dell'altra relazione. Ad esempio, date due relazioni

- *Ordini*(CodiceOrdine, Cliente, Data, Totale)
- *LineeOrdine*(CodiceOrdine, Linea, Prodotto, Importo)

questa tecnica (con riferimento agli attributi *CodiceOrdine* delle due relazioni) permetterebbe una memorizzazione contigua di ciascun ordine con le rispettive "linee d'ordine," cioè dei prodotti ordinati (ciascun ordine fa riferimento a più prodotti, ognuno su una "linea").

Con riferimento all'esempio, indicare quali delle seguenti operazioni possono trarre vantaggio dall'uso di questa opportunità e quali ne possono essere penalizzate (spiegare la risposta possibilmente anche in termini quantitativi, attraverso l'uso di esempi, che prevedano ipotesi sulle dimensioni dei vari attributi; ipotizzare anche una valutazione di convenienza complessiva, rispetto a possibili frequenze delle tre operazioni):

1. stampa dei dettagli (cioè delle linee d'ordine) di tutti gli ordini (ordinati per codice)
2. stampa dei dettagli di un ordine
3. stampa delle informazioni sintetiche (codice, cliente, data, totale) di tutti gli ordini