

Tecnologia delle basi di dati

Prova di autovalutazione — 23 maggio 2005

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 6 o 7 giugno. La discussione dovrà avvenire attraverso una presentazione delle soluzioni fatta da studenti (e non da parte del docente, che tutt'al più potrà commentarle).

Domanda 1

Molti DBMS hanno introdotto recentemente indici basati su una tecnica innovativa, detta “bitmap,” con le seguenti caratteristiche:

- l'indice, attributo A con K valori diversi, per una relazione con N ennuple, è costituito da K vettori (uno per ciascun valore dell'attributo A) di N bit ciascuno (un bit per ogni ennupla della relazione): l' i -esimo bit del vettore associato al valore a_j è 1 se il valore della i -esima ennupla sull'attributo A è a_j e 0 se il valore è diverso
- L'accesso ai vettori è organizzato per mezzo di un albero. In sostanza, rispetto ad un B⁺-tree, abbiamo una struttura di accesso simile, ma con foglie diverse: nel B⁺-tree, per ogni valore dell'attributo abbiamo una lista di indirizzi, mentre nell'indice bitmap abbiamo un vettore di bit (deve poi esistere una tabella che associ gli indirizzi delle ennuple ai numeri da 1 a N).

Sulla base degli elementi forniti:

1. valutare lo spazio necessario per le foglie di un indice bitmap, confrontandolo con quello necessario per le foglie di un B⁺-tree; in particolare, commentare come varia tale confronto in funzione del valore di K ;
2. approfondire il ragionamento al caso in cui vi siano più indici bitmap su attributi diversi;
3. spiegare come possono essere realizzate selezioni congiuntive su più attributi su ciascuno dei quali sia definito un indice bitmap;
4. spiegare per quali ragioni un indice bitmap può risultare vantaggioso nel contesto dei data warehouse.

Domanda 2

L'ufficio statistico dell'ateneo riceve spesso, dai presidi di facoltà e da altri docenti, richieste volte a conoscere:

- Il numero di immatricolati per la prima volta nel sistema universitario nazionale in un corso di laurea in un certo anno accademico (nel seguito si usa il termine *coorte* per fare riferimento a tale insieme di studenti);
- Per una specifica coorte e uno specifico anno accademico, il numero degli studenti che appartengono a ciascuna delle seguenti categorie:
 - iscritti a ciascuno degli anno di corso;
 - iscritti fuori corso;
 - laureati;
 - non più iscritti al corso di laurea, non laureati, ma iscritti ad altro corso dello stesso ateneo;
 - non più iscritti al corso di laurea, non laureati, e non iscritti ad altro corso dello stesso ateneo.
- Per una specifica coorte, con riferimento ad una certa data, con riferimento agli studenti ancora iscritti (in quella data), numero di studenti che hanno conseguito crediti in numero compreso in un certo intervallo (supponendo di interesse gli intervalli multipli di 10 e quelli multipli di 15).
- Per una specifica coorte, per uno specifico corso, il numero di studenti che hanno, ad una certa data, superato il relativo esame.
- Per una specifica coorte, per uno specifico corso, il voto medio riportato dagli studenti che hanno, ad una certa data, superato il relativo esame.

Progettare un data mart che permetta di rispondere alle esigenze sopra formulate, supponendo che le informazioni necessarie allo scopo siano nella base di dati delle segreterie studenti. In particolare,

1. mostrare i frammenti di schema (ER e relazionale) della base di dati delle segreterie che si suppone di utilizzare come sorgente dei dati;
2. mostrare gli schemi a stella dei data mart;
3. mostrare le trasformazioni necessarie per passare dalla sorgente ai data mart.