

Basi di dati II, primo modulo — 17 febbraio 2011

Cognome _____ Nome _____ Matricola _____ Ordin. _____

Domanda 1 (20%) Si consideri una forma di equivalenza fra schedule denominata *final-state-equivalenza*, secondo la quale due schedule S_1 e S_2 sono equivalenti se, per ogni istanza d della base di dati, essi la trasformano nello stesso modo (e quindi $S_1(d) = S_2(d)$, se con $S(d)$ indichiamo l'istanza della base di dati ottenuta applicando a d le operazioni dello schedule S).

1. Formulare una definizione per questa proprietà, variante della definizione di view-equivalenza.
2. Spiegare, almeno intuitivamente, il rapporto che esiste fra final-state-equivalenza e view-equivalenza (sono equivalenti, incomparabili, oppure una implica l'altra?).

Domanda 2 (20%) Si consideri una relazione $R(\underline{A} B C D E)$, in cui gli attributi hanno tutti la stessa dimensione a (ad esempio, ma è irrilevante, 4Byte), molto più piccola della dimensione del blocco pari a P . Si supponga che la relazione sia molto grande (N ennuple) e che le operazioni più frequenti su di essa siano le seguenti:

o_1 SELECT * FROM R ORDER BY A, con frequenza f_1

o_2 SELECT A, B, C FROM R ORDER BY A, con frequenza $f_2 = 10 \times f_1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

(i) memorizzazione della relazione $R(\underline{A} B C D E)$ ordinata su A

costo unitario di o_1 : _____ costo unitario di o_2 : _____

costo complessivo: _____

(ii) memorizzazione delle proiezioni $R1(\underline{A} B C)$ e $R2(\underline{A} D E)$, entrambe ordinate su A

costo unitario di o_1 : _____ costo unitario di o_2 : _____

costo complessivo: _____

Supporre che il join possa essere eseguito con il metodo merge-join (e che il costo del join stesso sia trascurabile rispetto alle due scansioni).

Domanda 3 Considerare uno schema dimensionale relativo agli esami, che utilizzi, come tabella dei fatti e come una delle dimensioni, le relazioni come quelle qui schematizzate:

| <u>KStudente</u> | <u>KCorso</u> | <u>KData</u> | Voto | ... |
|------------------|---------------|--------------|------|-----|
| 301 | 201 | 405 | 25 | ... |
| 301 | 202 | 406 | 28 | ... |
| 302 | 201 | 407 | 30 | ... |
| 302 | 203 | 407 | 22 | ... |
| ... | ... | ... | ... | ... |

| <u>KCorso</u> | Titolo | Crediti | ... |
|---------------|-----------|---------|-----|
| 201 | Fisica I | 6 | ... |
| 202 | Chimica | 9 | ... |
| 203 | Geometria | 6 | ... |
| ... | ... | ... | ... |

Supporre che si presentino le seguenti esigenze di modifica:

- i corsi cambiano nome nel tempo: per esempio, il corso nella prima enupla potrebbe ad un certo punto cambiare nome da “Fisica I” in “Meccanica”; interessano selezioni e aggregazioni relative agli esami tanto con riferimento al nome del corso (al momento dell’esame) quanto alla sua identità (un codice che viene introdotto allo scopo, ma non sempre viene utilizzato, perché alcuni analisti preferiscono fare riferimento al nome corrente del corso); le modifiche sono rare, ma è possibile che ci siano corsi con vari cambiamenti di nome;
- per ogni corso, interessa rappresentare anche il docente, per supportare analisi sugli esami svolti da ciascun docente; i docenti cambiano nel tempo e passano da un corso all’altro (e possono anche tenere più corsi nello stesso momento, ma ogni corso ha, in un certo giorno, un solo docente); è disponibile l’informazione relativa ai docenti dei corsi nel tempo (per tutto il periodo, anche passato, di interesse).

Mostrare nuove versioni delle due tabelle che permettano di soddisfare le esigenze sopra citate (mostrare anche i dati, con riferimento a quelli presenti negli esempi sopra, aggiungendo nuovi dati ragionevoli, che permettano di comprendere le modifiche).