

Basi di dati II, primo modulo — Tecnologia delle basi di dati

24 settembre 2010 — Compito A

Rispondere alle prime tre domande su questo foglio e alla quarta sul foglio separato.

Tempo a disposizione: un'ora e trenta minuti.

Cognome _____ Nome _____ Matricola _____ Ordin. _____

Domanda 1 (20%) Si consideri una relazione $R(\text{CodiceCliente}, \text{Cognome}, \text{Nome}, \text{Regione})$. Spiegare perché, per la ricerca di tutti i clienti di una certa regione, un indice secondario sull'attributo **Regione** potrebbe in alcuni casi risultare conveniente e in altri no (mostrare anche esempi che illustrino le due situazioni da un punto di vista quantitativo).

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 (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 4 (40%) Una catena di negozi gestisce le attività utilizzando, in ciascun negozio, una base di dati con le seguenti relazioni (su cui sono definiti gli ovvi vincoli di integrità referenziale):

- Prodotti(CodiceProdotto, Descrizione, Prezzo, Categoria, Marca)
- Marca(Codice, Nome)
- Categorie(Codice, Descrizione, MacroCategoria)
- MacroCategorie(Codice, Descrizione)
- Vendite(NumeroScontrino, Ora)
- DettaglioVendite(NumeroScontrino, CodiceProdotto, Quantità)

Si noti che

- Le informazioni relative alle vendite vengono mantenute solo nel corso della giornata.
- Il prezzo di un prodotto può variare da un giorno all'altro.

Utilizzando tali informazioni, la catena vuole realizzare un data mart relativo alle vendite dei prodotti, avente come misure le quantità vendute e gli incassi, che permetta di effettuare analisi di tipo temporale (incluse, oltre ai giorni, anche le fasce orarie della giornata, ad esempio 9-10, 10-11 e così via, oppure mattina e pomeriggio) e su prodotti (singoli o per categoria e/o per marca) e sui negozi. Allo scopo, specificare un possibile dettaglio del data mart; in particolare

1. **specificare esplicitamente la grana scelta**, supponendo che la quantità delle vendite sia tale che si è deciso di non utilizzare il massimo livello di dettaglio, ma solo quello strettamente indispensabile (in altri termini, la grana non deve essere il singolo dettaglio di vendita, ma una opportuna aggregazione);
2. mostrare gli schemi delle tabelle (tabella dei fatti e tabelle delle dimensioni) indicare anche le dimensioni i cui dati non provengono dalla base di dati sopra mostrata

(Rispondere sul foglio separato)

Cognome _____ Nome _____ Matricola _____ Ordin. _____

A large, empty rectangular box with a thin black border, occupying most of the page below the header. It is intended for a drawing or a detailed answer.

Basi di dati II, primo modulo — Tecnologia delle basi di dati

24 settembre 2010 — Compito B

Rispondere alle prime tre domande su questo foglio e alla quarta sul foglio separato.

Tempo a disposizione: un'ora e trenta minuti.

Cognome _____ Nome _____ Matricola _____ Ordin. _____

Domanda 1 (20%) Si consideri una relazione $R(\text{CodiceCliente}, \text{Cognome}, \text{Nome}, \text{Città})$. Spiegare perché, per la ricerca di tutti i clienti di una certa città, un indice secondario sull'attributo Città potrebbe in alcuni casi risultare conveniente e in altri no (mostrare anche esempi che illustrino le due situazioni da un punto di vista quantitativo).

Domanda 2 (20%) Si consideri una relazione $R(\underline{A} \ B \ C \ D \ E)$, in cui gli attributi hanno tutti la stessa dimensione d (ad esempio, ma è irrilevante, 4Byte), molto più piccola della dimensione del blocco pari a P . Si supponga che la relazione sia molto grande (L 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 (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 i della base di dati, essi la trasformano nello stesso modo (e quindi $S_1(i) = S_2(i)$, se con $S(i)$ indichiamo l'istanza della base di dati ottenuta applicando a i 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 4 (40%) Una catena di negozi gestisce le attività utilizzando, in ciascun negozio, una base di dati con le seguenti relazioni (su cui sono definiti gli ovvi vincoli di integrità referenziale):

- Prodotti(CodiceProdotto, Descrizione, Prezzo, Categoria, Produttore)
- Produttore(Codice, Nome)
- Categorie(Codice, Descrizione, MacroCategoria)
- MacroCategorie(Codice, Descrizione)
- Vendite(NumeroScontrino, Ora)
- DettaglioVendite(NumeroScontrino, CodiceProdotto, Quantità)

Si noti che

- Le informazioni relative alle vendite vengono mantenute solo nel corso della giornata.
- Il prezzo di un prodotto può variare da un giorno all'altro.

Utilizzando tali informazioni, la catena vuole realizzare un data mart relativo alle vendite dei prodotti, avente come misure le quantità vendute e gli incassi, che permetta di effettuare analisi di tipo temporale (incluse, oltre ai giorni, anche le fasce orarie della giornata, ad esempio 9-10, 10-11 e così via, oppure mattina e pomeriggio) e su prodotti (singoli o per categoria e/o per produttore) e sui negozi. Allo scopo, specificare un possibile dettaglio del data mart; in particolare

1. **specificare esplicitamente la grana scelta**, supponendo che la quantità delle vendite sia tale che si è deciso di non utilizzare il massimo livello di dettaglio, ma solo quello strettamente indispensabile (in altri termini, la grana non deve essere il singolo dettaglio di vendita, ma una opportuna aggregazione);
2. mostrare gli schemi delle tabelle (tabella dei fatti e tabelle delle dimensioni) indicare anche le dimensioni i cui dati non provengono dalla base di dati sopra mostrata

(Rispondere sul foglio separato)

Cognome _____ Nome _____ Matricola _____ Ordin. _____

A large, empty rectangular box with a thin black border, occupying most of the page below the header. It is intended for a drawing or a detailed answer.