

Basi di dati, primo modulo (vecchio ordinamento)

Prova parziale — 7 giugno 2002 — Compito A

Tempo a disposizione: un'ora e quarantacinque minuti. Libri aperti e appunti chiusi.

Domanda 1 (20%) I trigger sono citati esplicitamente da vari autori come “risposta” della tecnologia delle basi di dati alle esigenze che altri autori vedono soddisfatta dal paradigma orientato agli oggetti. Commentare brevemente.

Domanda 2 (25%) Considerare la seguente interrogazione in SQL:

```
SELECT A, D
FROM R, S
WHERE C > 150 AND E = B
```

definita con riferimento a due relazioni, definite e frammentate come segue (per essere poi distribuite):

- $R(A, C, E)$ frammentata orizzontalmente:
 - $R_a = \sigma_{C > 100}(R)$
 - $R_b = \sigma_{C \leq 100}(R)$
- $S(B, D, F)$ frammentata verticalmente:
 - $S_a = \pi_{B,D}(S)$
 - $S_b = \pi_{B,F}(S)$

Mostrare (ad esempio sotto forma di albero) l'espressione dell'algebra relazionale definita sui frammenti che realizza in modo efficiente tale interrogazione.

Domanda 3 (30%) Si consideri una base di dati relativa alle carriere degli studenti presso una università. In particolare vengono registrati i corsi (con codice, titolo, docente e facoltà), gli studenti (con matricola, anno di corso, nome, cognome e facoltà) e gli esami (ognuno relativo ad uno studente e un corso), con data, sessione e anno accademico (che sono almeno in parte indipendenti dalla data) e voto. In tale contesto, si vuole progettare un data mart per studiare l'andamento nel tempo degli esami, in funzione dei vari parametri disponibili (ma con il divieto di rappresentare i singoli esami, per ragioni di rispetto della privacy; non è sufficiente nascondere i dati anagrafici, è necessario prevedere forme di aggregazione). Mostrare

1. per la base di dati, lo schema concettuale;
2. per il datawarehouse:
 - (a) lo schema concettuale;
 - (b) i fatti, le misure e le dimensioni;
 - (c) lo schema dimensionale a stella.

Domanda 4 (25%) Il controllo di concorrenza basato su lock può essere considerato “pessimistico,” nel senso che porta a fermare transazioni che non è sicuro che vadano in conflitto. In alternativa, alcuni autori hanno proposto un controllo di concorrenza “ottimistico,” basato sulle seguenti ipotesi e principi:

- si assume che ciascuna transazione esegua prima tutte le proprie letture, poi svolga le proprie elaborazioni “privatamente” e infine chiedi l'autorizzazione ad andare in commit e quindi a scrivere
- (supponendo per semplicità che le letture siano seriali: prima fa tutte le proprie letture una transazione, poi le fa tutte una seconda, e così via) viene assegnato a ciascuna transazione un “timestamp” di lettura;
- il sistema ricorda, per ciascuna transazione andata in commit, quali sono i dati che ha letto e scritto e il grafo dei conflitti fra di esse (vuoto all'avvio e aggiornato come descritto al punto seguente)
- all'atto della richiesta di autorizzazione per il commit, si verifica quali siano le transazioni con cui la transazione è in conflitto lettura-scrittura e, se il grafo, con l'aggiunta di archi che descrivano tali conflitti (come nella verifica della CSR), risulta ciclico, la transazione viene rifiutata, altrimenti viene accettata e il grafo viene aggiornato.

Spiegare (intuitivamente):

1. perché il metodo garantisce la serializzabilità;
2. perché l'algoritmo risulta utilizzabile in pratica (con opportune tecniche per la gestione efficiente del grafo), mentre in generale la tecnica CSR non lo è;
3. a che punto possono essere eliminate dal grafo (al fine di evitare la crescita eccessiva del grafo stesso) le informazioni su una transazione andata in commit

Basi di dati, primo modulo (vecchio ordinamento)

Prova parziale – 7 giugno 2002 – Compito B

Tempo a disposizione: un'ora e quarantacinque minuti. Libri aperti e appunti chiusi.

Domanda 1 (20%) Alcuni autori sostengono che le basi di dati attive contribuiscono a completare la “migrazione storica” iniziata con il concetto di base di dati intesa come risorsa comune per più applicazioni. Commentare brevemente.

Domanda 2 (25%) Considerare la seguente interrogazione in SQL:

```
SELECT E, D
FROM T, R
WHERE A < 90 AND C = B
```

definita con riferimento a due relazioni, definite e frammentate come segue (per essere poi distribuite):

- $T(E, A, C)$ frammentata orizzontalmente:
 - $T_a = \sigma_{A > 100}(T)$
 - $T_b = \sigma_{A \leq 100}(T)$
- $R(B, D, F)$ frammentata verticalmente:
 - $R_a = \pi_{B,D}(R)$
 - $R_b = \pi_{B,F}(R)$

Mostrare (ad esempio sotto forma di albero) l'espressione dell'algebra relazionale definita sui frammenti che realizza in modo efficiente tale interrogazione.

Domanda 3 (30%) Si consideri una base di dati relativa alle carriere degli studenti presso una università. In particolare vengono registrati i corsi (con codice, titolo e facoltà), gli studenti (con matricola, anno di corso, nome, cognome e facoltà) e gli esami (ognuno relativo ad uno studente e un corso), con data (da cui dipendono funzionalmente sessione e anno accademico), docente esaminatore e voto. In tale contesto, si vuole progettare un data mart per studiare l'andamento nel tempo degli esami, in funzione dei vari parametri disponibili (ma con il divieto di rappresentare i singoli esami, per ragioni di rispetto della privacy; non è sufficiente nascondere i dati anagrafici, è necessario prevedere forme di aggregazione). Mostrare

1. per la base di dati, lo schema concettuale;
2. per il datawarehouse:
 - (a) lo schema concettuale;
 - (b) i fatti, le misure e le dimensioni;
 - (c) lo schema dimensionale a stella.

Domanda 4 (25%) Alcuni autori hanno proposto una variante del 2PL chiamata “locking altruistico” in cui una transazione, dopo aver ottenuto un lock su un oggetto e pur dovendo richiedere altri lock, può mettere a disposizione (“donare”) l'oggetto, permettendo ad altre transazioni (“usufruttuarie”) di utilizzarlo, sotto opportune condizioni (supponiamo per semplicità di illustrazione che i lock siano tutti esclusivi):

- ogni oggetto si può trovare (oltre che negli stati già noti, *libero*, *w-locked*) in uno dei seguenti stati: *donated-free*, *donated-w-locked* (con un solo usufruttuario)
- una transazione può donare un oggetto x dopo aver ottenuto il lock su di esso e prima di richiedere lock su altri oggetti, ma successivamente non può svolgere altre operazioni su x ; lock e unlock debbono seguire la regola a due fasi
- una transazione che dona un oggetto deve prima o poi rilasciarlo (“unlock”)
- una transazione usufruttuaria non può acquisire lock se non quelli acquisiti da un unico donatore e non può rilasciare i lock (né “subdonarli”) finché il donatore non li rilascia

Spiegare (intuitivamente):

1. perché il metodo garantisce la serializzabilità;
2. perché l'usufruttuario non può rilasciare i lock prima del donatore;
3. perché è utile che ogni usufruttuario abbia un solo donatore (in effetti, si possono definire varianti in cui ciò non è necessario, ma qui le ignoriamo per semplicità) e perché è utile che non acquisisca lock su oggetti liberi.

Basi di dati, primo modulo (vecchio ordinamento)

Prova parziale – 7 giugno 2002 – Compito C

Tempo a disposizione: un'ora e quarantacinque minuti. Libri aperti e appunti chiusi.

Domanda 1 (20%) Alcuni autori affermano che le funzionalità attive permettono di aumentare il grado di condivisione della risorsa “basi di dati” e quindi ne aumentano il valore. Commentare brevemente.

Domanda 2 (25%) Considerare la seguente interrogazione in SQL:

```
SELECT B, D
FROM S, T
WHERE E > 110 AND A = C
```

definita con riferimento a due relazioni, definite e frammentate come segue (per essere poi distribuite):

- $S(B, E, A)$ frammentata orizzontalmente:
 - $S_a = \sigma_{E > 100}(S)$
 - $S_b = \sigma_{E \leq 100}(S)$
- $T(C, D, F)$ frammentata verticalmente:
 - $T_a = \pi_{C,D}(T)$
 - $T_b = \pi_{C,F}(T)$

Mostrare (ad esempio sotto forma di albero) l'espressione dell'algebra relazionale definita sui frammenti che realizza in modo efficiente tale interrogazione.

Domanda 3 (30%) Si consideri una base di dati relativa alle carriere degli studenti presso una università. In particolare vengono registrati i corsi (con codice, titolo, docente e facoltà), gli studenti (con matricola, anno di corso, nome, cognome e facoltà) e gli esami (ognuno relativo ad uno studente e un corso), con data (da cui dipendono funzionalmente sessione e anno accademico) e voto. In tale contesto, si vuole progettare un data mart per studiare l'andamento nel tempo degli esami, in funzione dei vari parametri disponibili (ma con il divieto di rappresentare i singoli esami, per ragioni di rispetto della privacy; non è sufficiente nascondere i dati anagrafici, è necessario prevedere forme di aggregazione). Mostrare

1. per la base di dati, lo schema concettuale;
2. per il datawarehouse:
 - (a) lo schema concettuale;
 - (b) i fatti, le misure e le dimensioni;
 - (c) lo schema dimensionale a stella.

Domanda 4 (25%) Alcuni autori hanno proposto una variante del 2PL basata sulla gestione di più versioni di uno stesso oggetto, e in particolare, due, nella tecnica più semplice:

- una versione stabile, acceduta in lettura
- una versione provvisoria, prodotta da una transazione che vuole scrivere

Si possono avere tre tipi di lock (gestiti nel rispetto del protocollo a due fasi: dopo il primo rilascio nessuna nuova acquisizione):

- in lettura (compatibile con altri in lettura e con uno in scrittura)
- in scrittura (compatibile con lock in lettura ma non con altri in scrittura)
- in commit (incompatibile con tutti gli altri) utilizzato a fine transazione per andare in commit e far diventare stabili le versioni provvisorie degli oggetti scritti

Spiegare (intuitivamente):

1. perché il metodo garantisce la serializzabilità;
2. perché il metodo favorisce le prestazioni;
3. perché il metodo si coordina bene con quelli per la gestione dell'affidabilità.