

Esercizi sulla

Teoria Relazionale

F.M. Malvestuto

(Corsi di Laurea in INFORMATICA, A. A. 2002/2003)

Prova Scritta del 18 gennaio 2000

PARTE I - Basi di Dati relazionali

ESERCIZIO 1. Si consideri lo schema relazionale $\rho = [R, F]$, dove $R = abcdeg$ ed $F = \{(a, b), (ae, g), (bc, de)\}$. Sia $X = ac$.

1. Calcolare la chiusura di Armstrong di X rispetto a ρ .
2. Dare una derivazione della dipendenza funzionale (X, R) da F .

ESERCIZIO 2. Si consideri lo schema relazionale $\rho = [R, F]$, dove $R = \{matricola, prof, corso\}$ ed $F = \{(matricola, corso, prof), (prof, corso)\}$. Si provi che:

1. ρ non è in forma normale di Boyce-Codd.
2. Non esiste alcuna scomposizione di ρ che conservi le dipendenze.

PARTE II - Organizzazione fisica e gestione delle transazioni

ESERCIZIO 1. Sia T un B-albero in cui sia memorizzato un file "principale" (main) di 1000 record. Sappiamo che un blocco di memoria ha una capacità massima di 19 record, tanto per i blocchi contenenti record del main (che si considerano parte di T) quanto per quelli contenenti record dell'indice. Sappiamo inoltre che ogni blocco di T contiene esattamente 10 record.

- (a) Quanti livelli ha T ?
- (b) Che numero *minimo* di inserzioni è necessario effettuare in T perché esso cresca di un livello?
- (c) Che numero *massimo* di inserzioni è possibile effettuare in T prima che esso cresca di un livello?

Motivare le risposte. In particolare spiegare i calcoli eseguiti.

ESERCIZIO 2. Si considerino le seguenti transazioni:

T1: LOCK A; LOCK B; UNLOCK B; UNLOCK A
T2: LOCK B; LOCK C; UNLOCK C; UNLOCK B
T3: LOCK C; LOCK A; UNLOCK A; UNLOCK C

- (a) Esistono scheduling non seriali di T1, T2 e T3 ma serializzabili?
- (b) Esistono scheduling di T1, T2 e T3 non serializzabili?

Motivare le risposte.

Prova Scritta del 31 gennaio 2000

PARTE I - Basi di Dati relazionali

Svuole progettare una base di dati contenente i seguenti attributi.

a:	materia	attributo alfanumerico.
b:	codice	attributo numerico di tipo intero
c :	studente	alfanumerico,
d:	voto	attributo numerico di tipo intero

Dopo aver determinato le ovvie dipendenze funzionali esistenti tra gli attributi, si trovi uno schema della base di dati nella forma normale più alta che le dipendenze consentono.

Supponiamo di aver progettato la base di dati utilizzando due nomi di relazione r ed s i cui schemi siano rispettivamente $R = \{a, b\}$ ed $S = \{b, c, d\}$. Si vuole interrogare la base di dati per ottenere una relazione con schema $\{a, c, d\}$ che riporti gli esami che sono stati superati con il voto più basso in assoluto. Si formuli tale richiesta con un'espressione del calcolo relazionale e la si valuti sulla seguente base di dati attendendosi quanti più possibile ad uno schema algoritmico.

r		s		
a	b	b	c	d
Programmazione	1	1	Angelo	18
Algoritmi	2	1	Maria	25
Architetture	3	2	Angelo	30
Linguaggi	4	2	Carlo	25
Sistemi informativi	5	3	Valeria	28

PARTE II - Organizzazione fisica e gestione delle transazioni

Esercizio 1.

a) Si descriva la procedura di ricerca per una interrogazione a intervallo su k-d-tree.

b) Si supponga che per un dato un file $R(X,Y,Z)$ con X , Y e Z campi di tipo intero, sia stata adottata una organizzazione di tipo k-d-tree. A partire dal k-d-tree iniziale vuoto, si inseriscano nell'ordine i seguenti record: (5,5,5), (7,2,0), (4,3,3), (3,7,5), (3,7,1), (4,2,10), (2,7,3), (3,7,7), (4,1,5). Disegnare il k-d-tree risultante.

c) Illustrare la procedura di ricerca sul k-d-tree ottenuto al punto b) nel caso dell'interrogazione $[3 \leq X < 5]$ and $[Y \leq 4]$.

Esercizio 2.

a) Si dimostri che se il grafo di serializzabilità di uno scheduling S non ha cicli, allora S é serializzabile.

b) Si consideri lo scheduling S delle transazioni T_1, T_2, T_3, T_4 e T_5 dato da:

- | | |
|-----------------|-------------------|
| 1. T1: lock A | 11. T4: unlock C |
| 2. T1: lock B | 12. T2: unlock B |
| 3. T4: lock C | 13. T2: lock C |
| 4. T3: lock D | 14. T3: lock B |
| 5. T3: unlock D | 15. T5: unlock D |
| 6. T1: unlock B | 16. T4: unlock A |
| 7. T2: lock B | 17. T5: lock A |
| 8. T1: unlock A | 18. T2: unlock C |
| 9. T4: lock A | 19. T3: unlock B |
| 10. T5: lock D | 20. T5: unlock A. |

Si applichi l'algoritmo per il test di serializzabilità ad S.

Prova Scritta del 17 giugno 2000

PARTE I - Basi di Dati relazionali

1) Si consideri la relazione

corso	S	M	D
	Giacomo	Matematica	Gauss
	Giacomo	Fisica	Newton
	Carlo	Matematica	Leinitz
	Carlo	Fisica	Galileo

avente la seguente semantica: per ogni materia (**M**), ogni studente (**S**) di quella materia ha un unico docente (**D**), ogni docente (**D**) insegna una sola materia (**M**). Fornire il modello della relazione **corso** dettato da questa semantica. Provare che tale modello non è in forma normale di Boyce-Codd e discutere le anomalie che possono sorgere in fase d'aggiornamento. Scomporre (in maniera algoritmica) il modello in sottomodelli che siano in forma normale di Boyce-Codd ed analizzare il risultato in termini di conservazione dei dati e delle dipendenze.

2) Si considerino le due relazioni **regno**(**SOVRANO INIZIO FINE**) e **vita**(**NOME SESSO NASCITA MORTE**) che formano lo schema di una base di dati sulla dinastia scozzese degli Stuart. Dare un'espressione del calcolo relazionale che fornisca i nomi dei membri di sesso maschile della dinastia nati nel XVII secolo che non furono sovrani, e la si valuti sulla seguente base di dati:

regno	SOVRANO	INIZIO	FINE
	Giacomo I	1603	1625
	Carlo I	1625	1648
	Carlo II	1660	1685
	Giacomo II	1685	1688
	Maria II	1688	1694
	Anna	1702	1714

vita	NOME	SESSO	NASCITA	MORTE
	Giacomo I	M	1566	1625
	Elisabetta	F	1590	1662
	Carlo I	M	1600	1649
	Carlo II	M	1630	1685
	Maria	F	1631	1659
	Giacomo II	M	1633	1701
	Enrichetta A.	F	1640	1670
	Maria II	F	1662	1694
	Anna	F	1665	1714
	Giacomo E.	M	1686	1766

PARTE II - Organizzazione fisica e gestione delle transazioni

1. Descrivere in al più 70 parole la ricerca per interpolazione su un file con indice.

2. Considerare un file in cui ciascun record contenga un intero e su cui sia definito un indice a B-albero con blocchi di capacità $2e-1 = 2d-1 = 9$. Sia T l'albero ottenuto inserendo i primi 54 interi positivi nell'ordine naturale: 1, 2, ... 54.

- Descrivere cosa avviene in T a seguito della cancellazione del record contenente 46.

- Descrivere cosa avviene in T a seguito della cancellazione del record contenente 12.

3. Dimostrare che qualunque schedule di transazioni a due fasi è serializzabile.

Prova Scritta del 18 luglio 2000

PARTE I - Basi di Dati relazionali

1) Dato un modello $m = [R, F]$,

a) dare le definizioni di chiusura logica e di chiusura di Armstrong di F rispetto ad R , e calcolarle nel caso che $F = \emptyset$;

b) descrivere il metodo per costruire una relazione con schema R che soddisfi tutte e solo le dipendenze valide in m , ed applicarlo al caso in cui $R = \{\mathbf{a}, \mathbf{b}\}$ ed $F = \emptyset$.

Soluzione. La CHIUSURA LOGICA di F rispetto ad R è l'insieme delle dipendenze applicabili ad R che sono soddisfatte da tutte le relazioni presenti nella varietà del modello $m = [R, F]$, cioè da tutte le relazioni con schema R che soddisfano ognuna delle dipendenze in F . Nel caso in cui $F = \emptyset$, la varietà del modello m coincide con l'insieme di tutte le possibili relazioni con schema R . Ora, siccome ogni dipendenza banale applicabile ad R è sempre soddisfatta mentre, per ogni dipendenza non banale applicabile ad R , esiste almeno una relazione con schema R che non la soddisfa (dare un esempio), la chiusura logica di F rispetto ad R coincide con l'insieme delle dipendenze banali applicabili ad R .

La CHIUSURA DI ARMSTRONG di F rispetto ad R è l'insieme delle dipendenze derivabili da F applicando gli assiomi di Armstrong: A1-A6. (La definizione di derivabilità va richiamata). Nel caso in cui $F = \emptyset$, va provato che le uniche dipendenze derivabili da F con gli assiomi di Armstrong sono quelle che si ottengono con gli assiomi A1 ed A2 oppure con gli assiomi A1 ed A4 così che la chiusura di Armstrong di F rispetto ad R contiene tutte e solo le dipendenze banali applicabili ad R .

2) Si consideri il modello $m = [R, F]$ dove $R = \{\mathbf{Matricola\ di\ Studente}, \mathbf{Professore}, \mathbf{Corso}\}$ ed F contiene le due dipendenze

$\{\mathbf{Matricola\ di\ Studente}, \mathbf{Corso}\}$ determina $\mathbf{Professore}$

$\mathbf{Professore}$ determina \mathbf{Corso}

Assumendo che i domini dei tre attributi siano atomici, dire in quali forme normali si trova m motivando le risposte.

Soluzione. Dopo aver richiamato le definizioni di ciascuna delle forme normali, vanno determinate le chiavi di m usando l'algoritmo per il calcolo della chiusura di Armstrong di un insieme di attributi in m . Ve ne sono esattamente due: **$\{\mathbf{Matricola\ di\ Studente}, \mathbf{Corso}\}$** e **$\{\mathbf{Matricola\ di\ Studente}, \mathbf{Professore}\}$** . Ne segue che m non contiene attributi secondari e, quindi, è in 3.a forma normale (e, quindi in 2.a ed in 1.a). Inoltre va provato che dipendenza " **$\mathbf{Professore}$ determina \mathbf{Corso}** " viola la condizione di Boyce-Codd così che si può concludere che m non è in forma normale di Boyce-Codd.

Prova Scritta del 22 settembre 2000

PARTE I - Basi di Dati relazionali

1. Si consideri il modello $m = [R, F]$ dove

$R = \mathbf{abcdegh}$ ed $F = \{(\mathbf{ab}, \mathbf{e}), (\mathbf{ag}, \mathbf{d}), (\mathbf{be}, \mathbf{c}), (\mathbf{e}, \mathbf{g}), (\mathbf{cg}, \mathbf{h})\}$.

La dipendenza $(\mathbf{ab}, \mathbf{gh})$ è valida in M .

- Fornirne una derivazione utilizzando gli assiomi di Armstrong A1-A6
- Fornirne una derivazione nel sistema di calcolo basato sugli assiomi A1, A2 ed A6
- Fornirne una derivazione nonlineare nel sistema di calcolo basato sugli assiomi A1, A4 ed A7, e ricavare da questa una derivazione lineare
- Fornirne una derivazione nel sistema di calcolo grafico
- Dimostrarne la validità con l'algoritmo di Chang del calcolo proposizionale.

2. Si consideri la base di dati \mathcal{D} formata dalle due relazioni seguenti che riportano alcune informazioni sulla dinastia scozzese degli Stuart.

regno	SOVRANO	INIZIO	FINE	
	Giacomo I	1603	1625	
	Carlo I	1625	1648	
	Carlo II	1660	1685	
	Giacomo II	1685	1688	
	Maria II	1688	1694	
	Anna	1702	1714	
vita	NOME	SESSO	NASCITA	MORTE
	Giacomo I	M	1566	1625
	Elisabetta	F	1590	1662
	Carlo I	M	1600	1649
	Carlo II	M	1630	1685
	Maria	F	1631	1659
	Giacomo II	M	1633	1701
	Enrichetta A.	F	1640	1670
	Maria II	F	1662	1694
	Anna	F	1665	1714
	Giacomo E.	M	1686	1766

Si fornisca un'espressione del calcolo relazionale che fornisca il nome della prima regina della dinastia scozzese degli Stuart, e la si valuti sulla base di dati \mathcal{D} .

PARTE II - Organizzazione fisica e gestione delle transazioni

1. Qual è il massimo numero di blocchi che possono essere creati a seguito dell'inserzione di un record in un B-albero di profondità n ?

Qual è il numero massimo di blocchi che possono essere eliminati da un B-albero di profondità n a seguito di una cancellazione?

Motivare le risposte con due esempi.

2. Considerare due strategie alternative per la cancellazione di un record da un B-albero: nella prima le chiavi dei record (k, p) dell'indice vengono aggiornate in modo che k sia sempre uguale

alla più piccola chiave del file principale presente nel sottoalbero puntato da p . Nella seconda no. Mostrare un B-albero sul quale una cancellazione seguita da una inserzione producano nei due

casi una diversa distribuzione dei record del file principale.

3. Mostrare un insieme T di transazioni tali che: nessuna transazione in T segue il protocollo a due fasi, ma qualunque schedule di transazioni in T è seriale. T deve avere cardinalità maggiore di 1.

Prova Scritta del 15 gennaio 2001

PARTE I - Basi di Dati relazionali

1) Dare le definizioni di chiusura logica di un insieme di dipendenze nonché della sua chiusura di Armstrong.

Calcolare la chiusura di Armstrong di un insieme vuoto di dipendenze.

2) Si consideri un modello relazionale $[R, F]$ dove

$R = \{\text{Matricola di Studente, Professore, Corso}\}$

$F = \{(\{\text{Matricola di Studente, Corso}\}, \{\text{Professore}\}), (\{\text{Professore}\}, \{\text{Corso}\})\}$

In quali forme normali si trova il modello? (Provare la correttezza della risposta che si fornisce)

Soluzione

1) Dare le definizioni di chiusura logica di un insieme di dipendenze nonché della sua chiusura di Armstrong.

Sia R un insieme di attributi e sia F un insieme di dipendenze applicabili ad R .

La chiusura logica di F rispetto ad R è l'insieme delle dipendenze valide nel modello $[R, F]$, cioè l'insieme delle dipendenze applicabili ad R che sono soddisfatte da ogni relazione della varietà relazione (da definire) del modello $[R, F]$.

La chiusura di Armstrong di F rispetto ad R è l'insieme delle dipendenze applicabili ad R che sono derivabili da F con gli assiomi di Armstrong (da riportare).

Calcolare la chiusura di Armstrong di un insieme vuoto di dipendenze.

Sia R un insieme di attributi. La chiusura di Armstrong di F rispetto ad R risulta essere uguale all'insieme delle dipendenze banali applicabili ad R (Da provare che gli assiomi di Armstrong generano solo dipendenze banali).

2) Si consideri il modello $m = [R, F]$ dove $R = \{\text{Matricola di Studente, Professore, Corso}\}$ ed F contiene le due dipendenze:

$\{\text{Matricola di Studente, Corso}\}$ determina Professore
Professore determina Corso

Dire in quali forme normali si trova il modello m .

Il modello m ha due chiavi: $\{ \text{Matricola di Studente, Corso} \}$ e $\{ \text{Matricola di Studente, Professore} \}$. Dunque, i tre attributi in R sono tutti primari. Ne viene che, se i domini dei tre attributi sono atomici, allora m è sicuramente in terza forma normale (riportare la definizione) e, quindi in seconda e prima forma normale, ma non è in forma normale di Boyce-Codd (riportare la definizione) perché l'attributo (primario) Corso dipende parzialmente (e quindi transitivamente) dalla chiave $\{ \text{Matricola di Studente, Professore} \}$.

PARTE II - Organizzazione fisica e gestione delle transazioni

1. Chiamiamo "sicuro" un insieme S di transazioni tale che nessuna esecuzione concorrente di elementi di S può produrre una situazione di stallo (deadlock). Sia $A < B < \dots < Z$ un ordinamento su un insieme di risorse e sia $S1$ l'insieme di tutte le transazioni T che rispettano il protocollo che consiste nel richiedere i Lock rispettando l'ordinamento, cioè per le quali, se Lock X precede Lock Y in T , allora la risorsa X precede la risorsa Y nell'ordinamento. Sia $S2$ l'insieme ottenuto aggiungendo ad $S1$ la transazione $T1$:

Lock A
Lock C
Unlock C
Lock B
Unlock B
Unlock A

1.1 L'insieme $S2$ è sicuro? Motivare la risposta.

1.2 Se si è risposto negativamente in 1.1 fornire un esempio di stallo. Se si è risposto affermativamente fornire un nuovo protocollo tale che l'insieme di tutte le transazioni che lo rispettano sia sicuro ed includa $S2$.

2. Descrivere l'algoritmo di cancellazione di un record in un B-albero.

3. Considerare una base di dati in cui i record hanno come chiave un intero positivo. Le inserzioni nella base di dati si susseguono nel tempo rispettando sempre l'ordinamento naturale delle chiavi: 1, 2, ... n . Nessun intero positivo viene "saltato" e non sono previste cancellazioni di record. Quale (o quali) delle seguenti organizzazioni fisiche è (o sono) da preferirsi alle altre? B-albero; indice ISAM con ricerca binaria; indice ISAM con ricerca per interpolazione; insieme di bucket con funzione di accesso hash. Motivare la risposta.

Prova Scritta del 22 febbraio 2001

PARTE I - Basi di Dati relazionali

Si consideri lo schema relazionale formato dai quattro attributi **SOVRANO** **SESSO** **INIZIO** **FINE** **DURATA**. Si trovino le dipendenze e con queste si definisca un modello relazionale. Dire se il modello è in III forma normale. Se non lo è, trovare (algoritmicamente) una scomposizione del modello in III forma normale che conservi dati e dipendenze.

Si dia infine un'espressione del calcolo relazionale il cui valore sulla seguente relazione fornisca i nomi del re e della regina che hanno regnato più a lungo.

<i>regno</i>	SOVRANO	SESSO	INIZIO	FINE	DURATA
	Giacomo I	M	1603	1625	22
	Carlo I	M	1625	1648	23
	Carlo II	M	1660	1685	15
	Giacomo II	M	1685	1688	3
	Maria II	F	1688	1694	6
	Anna	F	1702	1714	12

PARTE II - Organizzazione fisica e gestione delle transazioni

1. Definire in dettaglio (incluso il formato dei blocchi e dei record) l'organizzazione fisica di una base di dati i cui record contengano parole di testo (una per ogni record). I record possono essere puntati e sono previste operazioni di *look-up*, inserzione e cancellazione. Mostrare lo stato della base di dati quando vi sia inserita una strofa della canzone *O' Sole mio*.

2. Considerare il seguente *schedule* di due transazioni T_1 e T_2 :

T_1 : Lock A

T_1 : Lock B

T_1 : Unlock B ()

T_2 : Lock B

T_2 : Unlock B ()

T_1 : Unlock A ()

T_1 : Lock C

T_1 : Unlock C ()

Chiamiamo rispettivamente a , b e c i valori iniziali delle risorse A , B e C . Esprimere in termini delle funzioni f , g , h ed k , associate rispettivamente alle azioni α , β e γ , i valori delle risorse A , B e C .

Prova Scritta del 5 giugno 2001

PARTE I - Basi di Dati relazionali

Esercizio 1. Il seguente algoritmo trova una chiave del modello $m = [R, F]$ dove $R = \{A_1, \dots, A_i\}$.

(1) $X := R$;

(2) Per $i = 1, \dots, n$

se la chiusura di Armstrong (rispetto ad F) di $X - \{A_i\}$
coincide con R allora $X := X - \{A_i\}$.

Dimostrare la correttezza dell'algoritmo.

Esercizio 2. Si consideri la base di dati $\{\text{studente}, \text{esame}\}$ dove gli schemi delle variabili relazionali **studente** ed **esame** sono rispettivamente $\{\text{NOME}, \text{MATRICOLA}, \text{NAZIONALITÀ}\}$ e $\{\text{MATERIA}, \text{MATRICOLA}, \text{VOTO}\}$. Si vuole sapere il nome e la nazionalità degli studenti che hanno preso 30 in tutti gli esami che hanno sostenuto. Si formuli la domanda nell'algebra relazionale o nel calcolo relazionale. La si valuti poi sullo stato della base di dati specificato dalle due relazioni:

studente	NOME	MATRICOLA	NAZIONALITÀ
	Al-Khuwarizmi	1	araba
	Gauss	2	tedesca
	Galois	3	francese
	Hilbert	4	tedesca

esame	MATERIA	MATRICOLA	VOTO
	Matematica	1	30
	Matematica	2	30
	Fisica	2	30
	Fisica	3	25
	Matematica	4	30
	Fisica	4	28

PARTE II - Organizzazione fisica e gestione delle transazioni

Esercizio 3. Si vuole memorizzare un file di record *non puntati* utilizzando un kd-tree. Ogni record contiene il nome e l'età di un bambino che frequenta la scuola materna del paese. Si assuma di utilizzare come chiavi sia il nome del bambino che l'età.

Si mostri lo stato dettagliato del kd-tree dopo l'inserimento dei seguenti record: Pia (anni 2), Camillo (anni 4), Rosina (anni 2), Piero (anni 2), Enzo (anni 4), Amelia (anni 4), Mariagrazia (anni 5), Angelo (anni 4), Mario (anni 4), Teresa (anni 5), Nadia (anni 4), Ugo (anni 2), Claudio (anni 3), Alessandro (anni 2).

Si descriva passo per passo la seguente operazione di ricerca nel kd-tree descritta informalmente: quali bambini hanno almeno 4 anni?

Esercizio 3. Sia $R_1 < R_2 < \dots < R_k$ un ordinamento totale delle risorse di un sistema concorrente. Si consideri il seguente protocollo P rispettato da tutte le transazioni del sistema: se una transazione effettua un lock su R_i e successivamente un lock su R_j , allora $i \leq j$.

Dimostrare o confutare formalmente la seguente congettura:

il protocollo P garantisce che tutti gli scheduling sono serializzabili.

Prova Scritta del 12 luglio 2001

PARTE I - Basi di Dati relazionali

Esercizio 1. Si consideri lo schema relazionale $S = \{\text{MATRICOLA}, \text{STUDENTE}, \text{MATERIA}, \text{DOCENTE}, \text{VOTO}\}$ e si assumano le dipendenze funzionali

MATRICOLA **STUDENTE**,
MATERIA **DOCENTE**,
MATRICOLA, **MATERIA** **VOTO**.

Supponendo che i domini degli attributi siano atomici, trovare una scomposizione dello schema S in III forma normale che conservi sia i dati che le dipendenze. (Riportare l'algoritmo di scomposizione ed applicarlo al caso in esame). Decidere se la scomposizione trovata è anche in forma normale di Boyce-Codd.

Dopo aver introdotto delle opportune variabili relazionali, una per ciascuno dei sottoschemi di S prodotti dalla scomposizione, formulare nell'algebra Relazionale e nel Calcolo Relazionale la seguente domanda: quali sono le coppie (**STUDENTE**, **DOCENTE**) tali che **STUDENTE** abbia superato con 30 almeno un esame sulla materia insegnata da **DOCENTE** ?

PARTE II - Organizzazione fisica e gestione delle transazioni

Esercizio 2. Si vuole memorizzare un file di record *non puntati* utilizzando un *B-tree*. Ogni record contiene un intero memorizzato in un campo che usiamo come chiave. Ogni blocco, che sia all'interno o che sia una foglia, può contenere al più 3 record. Si mostri lo stato dettagliato del B-tree dopo l'inserimento dei seguenti record:

149, 80, 122, 184, 78, 106, 47, 133, 64, 166, 85, 179.

Si mostri poi lo stato dettagliato del B-tree dopo l'inserimento di:

177, 176, 193, 48, 72, 115, 199, 93, 119, 107, 12, 1.

Infine, si mostri lo stato dettagliato del B-tree dopo la cancellazione di tutti i numeri compresi tra 170 e 179.

Si descriva passo passo la seguente operazione di ricerca nel B-tree descritta informalmente: quali record hanno un campo chiave maggiore od uguale a 100?

Esercizio 3. Si mostri che se S è uno *scheduling* qualsiasi di transazioni a "due fasi", allora S è serializzabile. Si dia per scontata la correttezza dell'algoritmo per il test di serializzabilità.

Prova Scritta del 17 settembre 2001

PARTE I - Basi di Dati relazionali

Esercizio 1. Si ricordi che un modello è in *forma canonica* se è in forma minimale, se è ridotto e se ogni dipendenza funzionale definita nel modello è semplice. Fornire un algoritmo che, dato un qualsiasi modello $m = [R, dom, F]$, permetta di costruire un modello in forma canonica che sia equivalente ad m . Si applichi tale algoritmo al modello $m = [R, dom, F]$ dove $R = \mathbf{abc}$ ed $F = \{\mathbf{ac} \rightarrow \mathbf{bc}, \mathbf{bc} \rightarrow \mathbf{a}, \mathbf{b} \rightarrow \mathbf{c}\}$ e si discuta l'unicità del risultato.

Esercizio 2. Si consideri la base di dati $\{\mathbf{studente}, \mathbf{tandem}\}$ dove gli schemi delle variabili relazionali **studente** e **tandem** sono rispettivamente $\{\mathbf{NOME}, \mathbf{MEDIA}\}$ e $\{\mathbf{COPPIA}, \mathbf{S1}, \mathbf{S2}\}$. Si vuole sapere per ogni coppia di studenti la media (dei voti) dello studente più bravo. Si formuli la domanda nel Calcolo Relazionale. La si valuti poi sullo stato della base di dati specificato dalle due relazioni:

studente	NOME	MEDIA
	Al-Khuwarizmi	25
	Gauss	26
	Galois	30
	Wiles	24
	Cantor	29

tandem	COPPIA	S1	S2
	alpha	Al-Khuwarizmi	Gauss
	beta	Galois	Hilbert
	gamma	Wiles	Cantor

Prova Scritta del 19 febbraio 2002

PARTE I - Basi di Dati relazionali

Si vuole progettare una base di dati contenente i seguenti attributi.

a:	materia	attributo alfanumerico
b:	docente	attributo alfanumerico
c :	studente	alfanumerico
d:	voto	attributo numerico di tipo intero

Dopo aver individuato le dipendenze funzionali esistenti tra gli attributi (si assuma tra l'altro che **a** determina **b**) si trovi uno schema della base di dati nella forma normale più alta che le dipendenze consentono.

Supponiamo di aver progettato la base di dati utilizzando due nomi di relazione **r** ed **s** i cui schemi siano rispettivamente **{a, b}** ed **{a, c, d}**. Si vuole interrogare la base di dati per ottenere una relazione con schema **{b, d}** che riporti il voto più alto dato da ciascun docente. Si formuli tale richiesta con un'espressione del calcolo relazionale e la si valuti sulla seguente base di dati attenendosi quanto più possibile ad uno schema algoritmico.

r		s		
a	b	a	c	d
Programmazione	Knuth	Algoritmi	Angelo	18
Algoritmi	Knuth	Algoritmi	Maria	25
Architetture	Tanenbaum	Linguaggi	Angelo	30
Linguaggi	Pascal	Basi di Dati	Carlo	25
Basi di Dati	Ullman	Basi di Dati	Valeria	30
		Programmazione	Angelo	27

SOLUZIONE

La domanda può essere formulata nel Calcolo Relazionale usando l'espressione

$$E = \{x(\mathbf{b}, \mathbf{d}) \mid f(x)\}$$

dove

$$f(x) = y(\mathbf{a}, \mathbf{b}) (r(y) \quad y(\mathbf{b}) = x(\mathbf{b}) \quad g(y) \quad h(y))$$

con

$$g(y) = z(\mathbf{a}, \mathbf{c}, \mathbf{d}) (s(z) \quad z(\mathbf{a}) = y(\mathbf{a}) \quad z(\mathbf{d}) = x(\mathbf{d}))$$

$$h(y) = u(\mathbf{a}, \mathbf{b}) (\neg r(u) \quad u(\mathbf{b}) \quad y(\mathbf{b}) \quad k(u))$$

dove

$$k(u) = v(\mathbf{a}, \mathbf{c}, \mathbf{d}) (\neg s(v) \quad v(\mathbf{a}) \quad u(\mathbf{a}) \quad v(\mathbf{d}) \quad x(\mathbf{d})).$$

Prova Scritta del 12 giugno 2002

PARTE I - Basi di Dati relazionali

1) Si consideri un modello relazionale con schema $R = \{A, B, C, D, E, F\}$ e dipendenze funzionali $\{A, C\} \rightarrow E, \{A, E\} \rightarrow F, \{B, C\} \rightarrow D, \{B, D\} \rightarrow A$.

1a) Trovare tutte le chiavi del modello.

1b) Applicare il Test di Normalità al modello. Se questo dà esito negativo, trovare una scomposizione conservativa in Terza Forma Normale. Applicare il Test di Normalità a ciascuna delle proiezioni così ottenute per sapere se queste sono anche in Forma Normale di Boyce-Codd.

SOLUZIONE

Per il teorema sull'unicità della chiave, la coppia $\{B, C\}$ è l'unica chiave del modello. Inoltre il modello non è normale e l'applicazione dell'algoritmo per la scomposizione conservativa in Terza Forma Normale dà $S = \{\{A, C, E\}, \{A, E, F\}, \{B, C, D\}, \{A, B, D\}\}$.

La proiezione del modello su $\{A, C, E\}$ ha per dipendenze funzionali

$\{A, C\} \rightarrow E$ e tutte le altre dipendenze funzionali valide nel modello che siano applicabili ad $\{A, C, E\}$.

Visto che, per ogni sottoinsieme nonvuoto X di $\{A, C, E\}$ con $X \neq \{A, C\}$, si ha $X^* = X$ l'unica dipendenza funzionale nonbanale valida nella proiezione del modello su $\{A, C, E\}$ è $\{A, C\} \rightarrow E$. Pertanto, la proiezione del modello su $\{A, C, E\}$ è un modello normale. Analogamente si prova che anche le proiezioni del modello su $\{A, E, F\}$, $\{B, C, D\}$ ed $\{A, B, D\}$ sono modelli normali.

2) Supponiamo di avere una base di dati con due sole variabili relazionali **corso** ed **esame** i cui schemi sono rispettivamente $\{A, B\}$ ed $\{A, C, D\}$ dove gli attributi sono così definiti:

A:	materia	attributo alfanumerico
B:	docente	attributo alfanumerico
C:	studente	alfanumerico
D:	voto	attributo numerico di tipo intero

Si vuole interrogare la base di dati per ottenere una relazione con schema $\{C\}$ che riporti gli studenti che hanno riportato il voto più alto dato da Knuth. Si formuli tale richiesta con un'espressione del calcolo relazionale e la si valuti sullo stato $\{r, s\}$ della base di dati dove r è il valore di **corso** dato dalla seguente relazione

A	B
Programmazione	Knuth
Algoritmi	Knuth
Architetture	Tanenbaum
Linguaggi	Pascal
Basi di Dati	Ullman

ed s è il valore di *esame* dato dalla seguente relazione

A	C	D
Algoritmi	Angelo	18
Algoritmi	Maria	27
Linguaggi	Angelo	30
Basi di Dati	Carlo	25
Basi di Dati	Valeria	30
Programmazione	Angelo	27

SOLUZIONE

La domanda può essere formulata nel Calcolo Relazionale usando l'espressione

$$E = \{x(\mathbf{C}) \mid f(x)\}$$

dove

$$f(x) = \exists y(\mathbf{A}, \mathbf{C}, \mathbf{D}) (esame(y) \wedge y(\mathbf{C}) = x(\mathbf{C}) \wedge g(y) \wedge h(y))$$

con

$$g(y) = \exists z(\mathbf{A}, \mathbf{B}) (corso(z) \wedge z(\mathbf{A}) = y(\mathbf{A}) \wedge z(\mathbf{B}) = \text{Knuth})$$

$$h(y) = \exists y'(\mathbf{A}, \mathbf{C}, \mathbf{D}) (\neg esame(y') \wedge y'(\mathbf{D}) = y(\mathbf{D}) \wedge k(y'))$$

dove

$$k(y') = \exists z'(\mathbf{A}, \mathbf{B}) (\neg corso(z') \wedge z'(\mathbf{A}) = y'(\mathbf{A}) \wedge z'(\mathbf{B}) = \text{Knuth}).$$

La valutazione dell'espressione E sullo stato $\{r, s\}$ della base di dati, richiede la sostituzione di x con ogni elemento c del dominio dell'attributo \mathbf{C} , e la successiva interpretazione della formula $f(x)|_{x=c}$. Ora, l'interpretazione di $f(x)|_{x=c}$ è pari a

$$t \wedge s \wedge I[t(\mathbf{C}) = c] \wedge I[g(y)|_{y=t}] \wedge I[h(y)|_{y=t}]$$

dove

$$I[g(y)|_{y=t}] = \bigvee_u \bigwedge_r I[u(\mathbf{A}) = t(\mathbf{A})] \wedge I[u(\mathbf{B}) = \text{Knuth}]$$

$$I[h(y)|_{y=t}] = \bigvee_v \bigwedge_s I[v(\mathbf{D}) = t(\mathbf{D})] \wedge I[k(y')|_{y'=v}]$$

con

$$I[k(y')|_{y'=v}] = \text{w} \text{ } r \text{ } I[w(\mathbf{A}) \text{ } v(\mathbf{A})] \text{ } I[w(\mathbf{B}) \text{ } \text{Knuth}].$$

Esplicitamente si ha che l'interpretazione di $f(x)|_{x=c}$ è vera se solo se $c \in \{\text{Angelo}, \text{Maria}\}$ e, dunque, la rispota alla domanda è la relazione

C
Angelo
Maria

Prova Scritta del 27 giugno 2002

PARTE I - Basi di Dati relazionali

ESERCIZIO 1. Si consideri il modello $M = [R, d, F]$ dove

$$R = \{a, b, c, d, e, f\}$$

$$F = \{a \rightarrow c, ad \rightarrow b, b \rightarrow d, c \rightarrow a, d \rightarrow e\}$$

Dare una derivazione regolare da F della dipendenza funzionale $ab \rightarrow ae$.

Dare un algoritmo per trovare una chiave di M che contenga l'attributo b . Sia questa X .

Posto $S = R - \{b\}$ e $Y = X - \{b\}$, decidere se Y è una chiave del modello $N = [S, d, G]$, dove G è l'insieme delle dipendenze funzionali valide in M applicabili ad S .

SOLUZIONE

Un esempio di derivazione regolare è

$$ab \rightarrow ab, b \rightarrow d, ab \rightarrow abd, d \rightarrow e, ab \rightarrow abde, ab \rightarrow ae$$

Un algoritmo per trovare una chiave X di M che contenga l'attributo b è

(1) $X := R$

(2) Per ogni $A \in R - \{b\}$, se $(X - \{A\})^+ = R$ allora $X := X - \{A\}$.

(3) Se $(X - \{b\})^+ = R$, allora concludere che non esiste nessuna chiave di M che contenga l'attributo b .

Se si esaminano gli attributi in R in ordine alfabetico, si trova $X = \{b, c, f\}$. Effettivamente, X è una chiave di M perché $(X - \{b\})^+ = R$.

Poniamo $S = \{a, c, d, e, f\}$ e $Y = \{c, f\}$. Per decidere se Y è una chiave del modello N , basta osservare che una dipendenza funzionale $Y \rightarrow A$

è applicabile ad S se e solo se $A \in S$

è valida in M se e solo se $A \in Y^+$

Pertanto, $Y \rightarrow A$ appartiene a G se e solo se $A \in S \cap Y^+$ e, quindi, la chiusura Y^* di Y rispetto a G è data da $Y^* = S \cap Y^+$. Dunque, Y è una chiave del modello N se e solo se

(1) $S \cap Y^+ = S$ o, equivalentemente, $S \subseteq Y^+$

(2) Per ogni $A \in Y, (Y - \{A\})^* \subseteq S$ o, equivalentemente, $S \subseteq (Y - \{A\})^+$.

Esplicitamente, $Y^+ = \{a, c, f\}$ così che già la condizione (1) non è soddisfatta. Pertanto, Y non è una chiave del modello N .

ESERCIZIO 2. Supponiamo di aver progettato una base di dati utilizzando due nomi di relazione *cattedra* ed *esame* i cui schemi siano rispettivamente {**a**, **b**} e {**a**, **c**, **d**}. Si vuole interrogare la base di dati per ottenere una relazione con schema {**b**, **c**} che contenga gli studenti (**c**) che abbiano superato tutti gli esami sostenuti con un voto non inferiore a 18 e contenga per ogni studente i docenti (**b**) delle materie oggetto degli esami da lui sostenuti. Si formuli tale richiesta con un'espressione dell'algebra relazionale e la si valuti sullo stato della base di dati

<i>cattedra</i>		<i>esame</i>		
a	b	a	c	d
Programmazione	Knuth	Algoritmi	Angelo	10
Algoritmi	Knuth	Algoritmi	Valeria	12
Architetture	Tanenbaum	Linguaggi	Angelo	30
Linguaggi	Pascal	Basi di Dati	Carlo	25
Basi di Dati	Ullman	Basi di Dati	Valeria	30
		Programmazione	Valeria	18
		Programmazione	Angelo	27

dettagliando i passi del processo di valutazione.

SOLUZIONE

Siano

$$\begin{array}{llll}
 E_1 = d_{<18}(esame) & E_2 = c(E_1) & E_3 = c(esame) & E_4 = E_3 - E_2 \\
 E_5 = a, c(esame) & E_6 = E_4 & E_5 & E_7 = cattedra & E_6 & E_8 = b, c(E_7)
 \end{array}$$

Allora E_8 è un'espressione dell'Algebra Relazionale per la domanda in esame. I valori delle otto espressioni sullo stato della base di dati sono rispettivamente

E_1	a	c	d
	Algoritmi	Angelo	10
	Algoritmi	Valeria	12
E_2		c	
		Angelo	
		Valeria	
E_3		c	
		Angelo	
		Valeria	
		Carlo	
E_4		c	
		Carlo	

<i>E</i> ₅	a Algoritmi Algoritmi Linguaggi Basi di Dati Basi di Dati Programmazione Programmazione	c Angelo Valeria Angelo Carlo Valeria Valeria Angelo	
<i>E</i> ₆	a Basi di Dati	c Carlo	
<i>E</i> ₇	a Basi di Dati	b Ullman	c Carlo
<i>E</i> ₈		b Ullman	c Carlo

PARTE II

ESERCIZIO 3. È dato un file di 250.000 record. I record sono a lunghezza fissa non puntati. Ogni record occupa 325 byte, di cui 75 per la chiave. Un puntatore a blocco occupa 5 byte. Un blocco di memoria contiene 2048 byte. Utilizziamo una organizzazione hash con una funzione che prende come argomento la chiave, e distribuisce uniformemente i record in 350 bucket.

- Quanti blocchi di memoria occupa la bucket directory?
- Quanti blocchi di memoria occupa ogni bucket?
- Qual è il numero medio di accessi a blocco richiesti per ricercare un record?

ESERCIZIO 4. Sia dato il seguente scheduling per l'insieme di transazioni T1, T2, T3

T1	T2	T3
	READ(X)	
		READ(Y)
	X = X - 20	
READ(X)		
X = X + 10		
		Y = Y - 5
WRITE(X)		
		WRITE(Y)
	WRITE(X)	
READ(Y)		
Y = Y + 20		
WRITE(Y)		

Il controllo della concorrenza è basato su timestamp, e le transazioni hanno i seguenti timestamp:

$$TS(T1) = 110, TS(T2) = 100, TS(T3) = 105.$$

Descrivere come vengono modificate durante lo schedule le variabili READ_TIMESTAMP e WRITE_TIMESTAMP associate agli item X ed Y, e come procedono le tre transazioni. Fornire inoltre i valori finali dei due item.

Prova Scritta del 12 luglio 2002

PARTE I - Basi di Dati relazionali

ESERCIZIO 1. Si consideri il modello $M = [R, d, F]$ dove

$$R = \{a, b, c, d\}$$

$$F = \{a \rightarrow bc, b \rightarrow bc, c \rightarrow bd, d \rightarrow ab, bc \rightarrow ad\}$$

e d è arbitrario.

1a) Dopo aver riportato l'Algoritmo di Semplificazione e l'Algoritmo di Riduzione, applicarli per ottenere

un modello minimale M' equivalente ad M ,

un modello ridotto M'' equivalente ad M ,

un modello N in forma canonica equivalente ad M .

1b) Dopo aver riportato l'Algoritmo di Scomposizione in Terza Forma Normale, applicarlo ad N . Sia S il ricoprimento di R così ottenuto.

1c) Dopo aver riportato il Test di Conservazione dei Dati ed il Test di Conservazione delle Dipendenze, applicarli per verificare che la scomposizione di N indotta da S conserva sia i dati che le dipendenze.

SOLUZIONE

1a) Una soluzione è data da

$M' = [R, d, F']$ dove $F' = \{a \rightarrow bc, b \rightarrow bc, c \rightarrow bd, d \rightarrow ab\}$ ($bc \rightarrow ad$ è ridondante).

$M'' = [R, d, F'']$ dove $F'' = \{a \rightarrow bc, b \rightarrow bc, c \rightarrow bd, d \rightarrow ab, c \rightarrow ad\}$: l'attributo b (o, alternativamente, l'attributo c) è eliminabile dal determinante della dipendenza funzionale $bc \rightarrow ad$.

$N = [R, d, G]$ dove $G = \{a \rightarrow b, b \rightarrow c, c \rightarrow d, d \rightarrow a\}$, cosicché ogni attributo forma una chiave.

1b) L'applicazione dell'Algoritmo di Scomposizione in Terza Forma Normale ad N genera il ricoprimento $S = \{ab, ad, bc, cd\}$ di R .

1c) Si verifica facilmente che il Test di Conservazione dei Dati dà esito positivo. Quanto al Test di Conservazione delle Dipendenze si osservi che, siccome ogni attributo forma una chiave di N , per ogni dipendenza $A \rightarrow B$ in G e per ogni $R' = \{A, B\}$ in S si ha che

$$(\{A\} \rightarrow R')^+ = \{A\}^+ = R$$

sicché $A \rightarrow B$ è sempre valida nel sottomodulo di N generato da S , il quale dunque risulta essere equivalente ad N .

ESERCIZIO 2. Sia r il nome di una relazione con schema $\{a, b, c\}$ dove a e b sono rispettivamente il nome ed il cognome di un padre, e c è il nome di un suo figlio (maschio). Si vuole conoscere il nome ed il cognome di tutti i nonni che non hanno nipoti che portino il loro nome. Si formuli tale richiesta con un'espressione E del Calcolo Relazionale e la si valuti sulla relazione r

a	b	c
Aldo	Rosso	Andrea
Aldo	Rosso	Mario
Franco	Rossi	Andrea
Mario	Bianchi	Filippo
Andrea	Rosso	Filippo
Giulio	Bianchi	Mario
Filippo	Bianchi	Giulio
Andrea	Rossi	Franco
Mario	Bianchi	Giulio

Ottenuto il valore $E(r)$ dell'espressione E , si esplicitino in dettaglio i passi dell'interpretazione per un'ennupla presente in $E(r)$ e per un'ennupla che non appartiene ad $E(r)$.

SOLUZIONE

La relazione richiesta deve contenere tutte le coppie (a, b) tali che

- (i) (a, b) sia un nonno
- (ii) per ogni figlio c di (a, b) si abbia che $c' \neq a$ per ogni figlio c' di c .

Introduciamo la variabile $x(a, b)$ per indicare la coppia (a, b) . Allora, la condizione (i) si esprime come

$$f(x) = \exists y(a, b, c) (r(y) \wedge y(a) = x(a) \wedge y(b) = x(b) \wedge g(x, y))$$

dove

$$g(x, y) = \exists z(a, b, c) (r(z) \wedge z(a) = y(c) \wedge z(b) = x(b))$$

La condizione (ii) si esprime come

$$h(x) = \forall u(a, b, c) (\neg r(u) \vee u(a) \neq x(a) \vee u(b) \neq x(b) \vee k(x, u))$$

dove

$$k(x, u) = \exists v(a, b, c) (\neg r(v) \wedge v(a) = u(c) \wedge v(b) = x(b) \wedge v(c) = x(a)).$$

Riassumendo abbiamo l'espressione

$$E = \{x(\mathbf{a}, \mathbf{b}) \mid f(x) = h(x)\}.$$

Il valore di E su r è la relazione

a	b
Aldo	Rosso
Filippo	Bianchi

Per $(a, b) = (\text{Aldo}, \text{Rosso})$, abbiamo $I[f(a, b) \rightarrow h(a, b)] = \text{VERO}$ perché

— $I[f(a, b)] = \text{VERO}$ quando le variabili y e z vengono sostituite rispettivamente con $(\text{Aldo}, \text{Rosso}, \text{Andrea})$ e $(\text{Andrea}, \text{Rosso}, \text{Filippo})$, e

— $I[h(a, b)] = \bigvee_{(a, b, c) \in r} I[(c, b, c') \wedge r] I[c' \neq a] = I[\text{Filippo} \neq \text{Aldo}] = \text{VERO}$.

Per $(a, b) = (\text{Andrea}, \text{Rosso})$, abbiamo $I[f(a, b) \rightarrow h(a, b)] = \text{FALSO}$ perché

— $I[f(a, b)] = \text{FALSO}$ perché Andrea Rosso non è un nonno in quanto suo figlio Filippo non risulta avere figli (cioè non esiste nessuna terna che può essere utilmente sostituita alla variabile z).

ESERCIZIO 3. Supponiamo di avere un file di 700.000 record. Ogni record occupa 180 byte, di cui 60 per il campo chiave. Ogni blocco contiene 2048 byte. Un puntatore a blocco occupa 5 byte. Non è richiesto l'allineamento dei campi ad indirizzi multipli di 4 e i record non sono puntati. Usiamo una organizzazione B-tree. Qual è il numero minimo di blocchi che dobbiamo utilizzare complessivamente per il file indice (totale dei blocchi di tutti i livelli) e per il file principale (livello foglia)?

ESERCIZIO 4. Sia dato il seguente schedule S per l'insieme di transazioni T1, T2, T3, T4, T5:

T1	T2	T3	T4	T5
LOCK A				
	LOCK B			
UNLOCK A				
	UNLOCK B			
	LOCK A			
		LOCK B		
	UNLOCK A			
				LOCK A
		UNLOCK B		
				UNLOCK A
		LOCK A		
			LOCK B	
		UNLOCK A		
			UNLOCK B	
LOCK B				
			LOCK A	
			UNLOCK A	
UNLOCK B				
				LOCK B
				UNLOCK B

Verificare se lo schedule è serializzabile, e in questo caso fornire lo schedule seriale equivalente (o gli schedule seriali equivalenti se ne esiste più di uno).

Soluzione degli Esercizi 1 e 2 della Prova Scritta di

Basi di Dati I

10 giugno 2003

ESERCIZIO 1. Si scriva un algoritmo che, facendo uso del solo Algoritmo di chiusura, decida se esista o meno un'unica chiave di un assegnato modello e, in caso affermativo, la trovi.

Si applichi l'algoritmo al modello $M = [R, F]$

$$R = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}\}$$

$$F = \{\mathbf{ab} \quad \mathbf{bc}, \mathbf{cd} \quad \mathbf{d}, \mathbf{c} \quad \mathbf{d}, \mathbf{cd} \quad \mathbf{de}\}$$

1)	$test := \text{Falso}; Z := \emptyset.$
2)	Per ogni f in F $Z' := dip(f) - det(f); Z :=$ $Z \cup Z'.$
3)	$Y := R - Z.$
4)	Calcolare la chiusura Y^+ di Y .
5)	Se $Y^+ = R$ allora $test := \text{Vero}.$

L'applicazione al modello M dà $test = \text{Vero}$ ed $Y = \{\mathbf{a}, \mathbf{b}\}.$

Si riporti il test di conservazione dei dati ed il test per la conservazione delle dipendenze. Li si applichi alla scomposizione di M indotta dal ricoprimento $S = \{\mathbf{ab}, \mathbf{cd}, \mathbf{ae}\}$ di R . Sulla base dei soli risultati ottenuti dai due test su S , elencare: quali scomposizioni di M conservano i dati, quali conservano le dipendenze, quali non conservano i dati e quali non conservano le dipendenze.

L'applicazione del test di conservazione dei dati così come l'applicazione del test di conservazione delle dipendenza danno esito negativo. Ne segue che ogni ricoprimento più fine di S induce una scomposizione di M che non conserva né i dati né le dipendenze.

ESERCIZIO 2. Sia r una variabile relazionale con schema $\{a, b, c\}$. Interpretare (attenendosi ad uno schema algoritmico) la formula

$$f = \exists y(a, b, c) (r(y) \wedge y(a) = \text{Aldo} \wedge y(b) = \text{Rosso} \wedge g(y))$$

dove

$$g(y) = \exists z(a, b, c) (r(z) \wedge z(a) = y(c) \wedge z(b) = y(b)),$$

sulla relazione

a	b	c
Aldo	Rosso	Andrea
Aldo	Rosso	Mario
Franco	Rossi	Andrea
Mario	Bianchi	Filippo
Andrea	Rosso	Filippo
Giulio	Bianchi	Mario
Filippo	Bianchi	Giulio
Andrea	Rossi	Franco
Mario	Bianchi	Giulio

Sia r questa relazione. Il seguente algoritmo fornisce il valore dell'interpretazione di f .

```

1)  valore := Falso.
2)  Per ogni terna  $(i, j, k)$    $r$  ripetere
      per ogni terna  $(i', j', k')$  ripetere
          se  $(i = \text{Aldo})$  e  $(j = \text{Rosso})$  e  $(i' = k)$  e  $(j' =$ 
       $j)$  allora  $\text{valore} := \text{Vero}$  ed Uscire.
```

L'applicazione dell'algoritmo dà valore Vero.

N.B. La formula f è interpretabile dal momento che non contiene variabili con occorrenze libere. Per lo stesso motivo, f non può essere utilizzata da sola per definire un'espressione del Calcolo Relazionale.

PARTE II

ESERCIZIO 3. Supponiamo di avere un file di 1.850.000 record. Ogni record occupa 370 byte, di cui 55 per il campo chiave. Ogni blocco contiene 2048 byte. Un puntatore a blocco occupa 5 byte. Non è richiesto l'allineamento dei campi ad indirizzi multipli di 4 e i record non sono puntati. Usiamo una organizzazione ISAM. Qual è il numero minimo di blocchi che dobbiamo utilizzare complessivamente per il file indice e per il file principale, supponendo di lasciare in ogni blocco il 20% (al massimo) di spazio libero?

ESERCIZIO 4. Sia dato il seguente schedule S per l'insieme di transazioni T1, T2, T3, T4, T5:

T1	T2	T3	T4	T5
LOCK A				
		LOCK B		
	LOCK C			
		UNLOCK B		
				LOCK B
				UNLOCK B
UNLOCK A				
LOCK B				
			LOCK A	
	UNLOCK C			
		LOCK C		
			UNLOCK A	
				LOCK A
		UNLOCK C		
				LOCK C
UNLOCK B				
				UNLOCK C
			LOCK C	
			UNLOCK C	
				UNLOCK A

Verificare se lo schedule è serializzabile, e in questo caso fornire lo schedule seriale equivalente (o gli schedule seriali equivalenti se ne esiste più di uno). In caso contrario indicare i cicli contenuti nel grafo.

Prova Scritta di **Basi di Dati I**

2 luglio 2003

PARTE I

ESERCIZIO 1. Si consideri il modello con schema $R = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}, \mathbf{f}, \mathbf{g}\}$ e dipendenze funzionali $F = \{\mathbf{a} \rightarrow \mathbf{d}, \mathbf{ab} \rightarrow \mathbf{de}, \mathbf{ce} \rightarrow \mathbf{f}, \mathbf{e} \rightarrow \mathbf{g}\}$. Trovare gli attributi primari e secondari del modello. Verificare che il modello non è in seconda forma normale esibendo un attributo secondario A ed una chiave X tale che A dipenda parzialmente da X . Applicare l'algoritmo di scomposizione in terza forma normale. Verificare che i modelli ottenuti dalla scomposizione sono tutti in forma normale di Boyce-Codd.

Soluzione. Per il teorema sull'unicità della chiave, il modello $M = [R, F]$ ha un'unica chiave che è data da $X = \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$. Pertanto, \mathbf{a}, \mathbf{b} e \mathbf{c} sono gli attributi primari di M , e $\mathbf{d}, \mathbf{e}, \mathbf{f}$ e \mathbf{g} sono gli attributi secondari di M . La dipendenza funzionale $\mathbf{a} \rightarrow \mathbf{d}$ in F dimostra poi che l'attributo secondario \mathbf{d} è parzialmente dipendente dalla chiave X ; pertanto, M non è in seconda forma normale. Per applicare l'algoritmo di scomposizione in terza forma normale, occorre prima fornire un modello equivalente ad M che sia in forma canonica. A questo scopo, si eliminano le dipendenze funzionali in F che non sono semplici: $F' = \{\mathbf{a} \rightarrow \mathbf{d}, \mathbf{ab} \rightarrow \mathbf{d}, \mathbf{ab} \rightarrow \mathbf{e}, \mathbf{ce} \rightarrow \mathbf{f}, \mathbf{e} \rightarrow \mathbf{g}\}$. Quindi, si eliminano le dipendenze funzionali ridondanti: $F'' = \{\mathbf{a} \rightarrow \mathbf{d}, \mathbf{ab} \rightarrow \mathbf{e}, \mathbf{ce} \rightarrow \mathbf{f}, \mathbf{e} \rightarrow \mathbf{g}\}$. Infine, ove possibile si riducono i determinanti (ma non è questo il caso). Dunque il modello $[R, F'']$ è equivalente ad M ed è in forma canonica. L'applicazione ad $[R, F'']$ dell'algoritmo di scomposizione in terza forma normale genera il ricoprimento $S = \{R_1, R_2, R_3, R_4, R_5\}$ con $R_1 = \{\mathbf{a}, \mathbf{d}\}$, $R_2 = \{\mathbf{a}, \mathbf{b}, \mathbf{e}\}$, $R_3 = \{\mathbf{c}, \mathbf{e}, \mathbf{f}\}$, $R_4 = \{\mathbf{e}, \mathbf{g}\}$ ed $R_5 = X = \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$. Per i modelli M_1, M_2, M_3, M_4, M_5 che si ottengono dalla scomposizione abbiamo le forme canoniche

$$\begin{array}{lll} [R_1, \{\mathbf{a} \rightarrow \mathbf{d}\}] & [R_2, \{\mathbf{ab} \rightarrow \mathbf{e}\}] & [R_3, \{\mathbf{ce} \rightarrow \mathbf{f}\}] \\ [R_4, \{\mathbf{e} \rightarrow \mathbf{g}\}] & [R_5, \emptyset] & \end{array}$$

e, dunque, essi sono tutti in forma normale di Boyce-Codd.

ESERCIZIO 2. Siano \mathbf{a}, \mathbf{b} e \mathbf{c} tre attributi binari e sia r una variabile relazionale con schema $R = \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$. Sia

$$E = \{x(R) \mid r(x) \rightarrow f\}$$

dove

$$f = \neg x(\mathbf{a}, \mathbf{b}) \vee (\neg y(\mathbf{a}) \wedge (x(\mathbf{a}) = y(\mathbf{a}) \wedge \neg x(\mathbf{b}) = 1)) .$$

Dimostrare nei dettagli che E è un'espressione del Calcolo Relazionale.

Se n sono le ennuple della relazione di nome r contenuta nello stato attuale D della base di dati, quante ennuple contiene la relazione che è il valore di E su D ?

Qual è un'espressione dell'Algebra Relazionale equivalente ad E ?

Soluzione. La formula $r(x) \wedge f$ è ben formata. La variabile x è

è parzialmente strutturata nella formula $x(b) = 1$ e vi ha schema $\{b\}$

è parzialmente strutturata nella formula $(x(a) = y(a) \wedge x(b) = 1)$ e vi ha schema $\{a, b\}$

è parzialmente strutturata nella formula $y(a) \wedge (x(a) = y(a) \wedge x(b) = 1)$ e vi ha schema $\{a, b\}$

ha solo occorrenze legate nella formula f

è ben strutturata nella formula $r(x) \wedge f$ con schema R .

La variabile y

è parzialmente strutturata nella formula $(x(a) = y(a) \wedge x(b) = 1)$ e vi ha schema $\{a\}$

ha solo occorrenze legate nelle tre formule

$$y(a) \wedge (x(a) = y(a) \wedge x(b) = 1) \quad f \quad r(x) \wedge f$$

Le variabili x ed y che compaiono nella formula f vi hanno solo occorrenze legate; pertanto, f è già interpretabile ed $I[f] = \text{FALSO}$ (esistendo coppie (a, b) con $b \neq 1$). Dunque, per ogni terna (a, b, c) si ha

$$I[r(x) \wedge (x(a) = y(a) \wedge x(b) = 1) \wedge f]_{x=(a,b,c)} = I[r(x) \wedge (x(a) = y(a) \wedge x(b) = 1)]_{x=(a,b,c)} \wedge I[f]_{x=(a,b,c)} = \text{FALSO} = \text{FALSO}$$

e quindi il valore E su ogni possibile stato della base di dati è una relazione vuota con schema $\{a, b, c\}$. Un'espressione E' dell'Algebra Relazionale equivalente ad E è

$$E' = r - r.$$