

Il Modello Relazionale

1

Il modello relazionale

Il modello relazionale

- Proposto da E. F. Codd nel 1970 per favorire l'indipendenza dei dati e reso disponibile come modello logico in DBMS reali nel 1981
- si basa sul concetto matematico di **relazione**, questo fornisce al modello una base teorica che permette di dimostrare formalmente proprietà di dati e operazioni
- le relazioni hanno una rappresentazione naturale per mezzo di tabelle

2

Il modello relazionale

Il modello relazionale

- Il modello relazionale, sebbene non sia stato il modello usato nei primi DBMS, è divenuto lentamente il modello più importante al punto che è oggi comunemente usato in quasi tutti i DBMS disponibili a livello commerciale
- la ragione principale della popolarità di questo modello è che fornisce linguaggi semplici e di tipo dichiarativo, ma al tempo stesso potenti, con cui esprimere le operazioni per l'accesso e la manipolazione dei dati

3

Il modello relazionale

Relazione

- D_1, D_2, \dots, D_n (n insiemi anche non distinti)
- il prodotto cartesiano $D_1 \times D_2 \times \dots \times D_n$, è l'insieme di tutte le **tuple** ordinate (d_1, d_2, \dots, d_n) tali che $d_1 \in D_1, d_2 \in D_2, \dots, d_n \in D_n$
- una relazione su D_1, D_2, \dots, D_n è un sottoinsieme del prodotto cartesiano $D_1 \times D_2 \times \dots \times D_n$
- D_1, D_2, \dots, D_n sono i domini della relazione
- una relazione su n domini ha **grado** n
- il numero di tuple è la **cardinalità** della relazione
- nelle applicazioni reali, la cardinalità è sempre finita

4

Il modello relazionale

Dominio

- Un dominio è un insieme (anche infinito) di valori:
- Esempi:
 - l'insieme dei numeri interi
 - l'insieme delle stringhe di caratteri di lunghezza 20
 - l'insieme $\{0,1\}$

5

Il modello relazionale

Esempio

- $D_1 = \{a,b\}$
- $D_2 = \{x,y,z\}$
- prodotto cartesiano $D_1 \times D_2$

a	x
a	y
a	z
b	x
b	y
b	z

- una relazione $r \subseteq D_1 \times D_2$

a	x
a	z
b	y
b	z

6

Il modello relazionale

Relazione matematica, proprietà

- In base alle definizioni, una relazione matematica è un **insieme** di tuple **ordinate**:
 (d_1, d_2, \dots, d_n) tali che $d_1 \in D_1, d_2 \in D_2, \dots, d_n \in D_n$
- una relazione è un **insieme**, quindi:
 - non è definito alcun ordinamento fra le tuple
 - le tuple di una relazione sono distinte l'una dall'altra
- una tupla è al proprio interno **ordinata**: l' i -esimo valore di ciascuna proviene dall' i -esimo dominio; è cioè definito un ordinamento fra i domini

7

Il modello relazionale

Notazioni

- Sia r una relazione di grado k :
 - sia t una tupla di r
 - sia i un intero appartenente all'insieme $\{1, \dots, k\}$
 - $t[i]$ denota la i -esima componente di t

Esempio:

Sia $r = \{(0,a), (0,c), (1,b)\}$

Sia $t = (0,a)$ una tupla di r

$t[2] = a$

$t[1] = 0$

8

Il modello relazionale

Modello relazionale

- Una relazione può essere vista, alternativamente, come una **tabella**, in cui ogni riga è una tupla ed ogni colonna corrisponde ad una componente
- alle colonne sono associati dei nomi, detti **nomi di attributo**
la coppia (nome di attributo, dominio) è detta **attributo**
- l'insieme degli attributi di una relazione ne costituisce lo **schema**

9

Il modello relazionale

Modello relazionale

- Se una relazione ha nome R ed attributi di nomi rispettivamente A_1, A_2, \dots, A_k , lo schema è spesso indicato con
$$R(A_1, A_2, \dots, A_k)$$
- inoltre $UR = \{A_1, A_2, \dots, A_k\}$ viene usato per denotare l'insieme di tutti i nomi di attributo della relazione R

10

Il modello relazionale

Esempio

- Info_Città

Città	Regione	Popolazione
Roma	Lazio	3000000
Milano	Lombardia	1500000
Genova	Liguria	800000
Pisa	Toscana	150000

Schema: Info_Città(Città,Regione,Popolazione)

11

Il modello relazionale

Modello relazionale

- In questa definizione del modello relazionale, le componenti delle tuple sono denotate tramite i nomi di attributi (notazione per nome in contrasto con la notazione per posizione)
- dato uno schema di relazione $R(A_1, A_2, \dots, A_k)$, una tupla t su tale schema può essere rappresentata tramite la notazione $[A_1:v_1, A_2:v_2, \dots, A_k:v_k]$ dove v_i ($i=1, \dots, k$) è un valore appartenente al dominio di A_i (indicato con $\text{dom}(A_i)$)
- inoltre $t[A_i]$ denota il valore dell'attributo A_i della tupla t

12

Il modello relazionale

Esempio

$t = [\text{Città: Roma, Regione: Lazio, Popolazione: 3000000}]$

è una tupla definita sullo schema `Info_Città`

$t[\text{Città}] = \text{Roma}$

Il valore dell'attributo `Città` per la tupla t è `Roma`

13

Il modello relazionale

Valori nulli

- Non sempre sono disponibili informazioni sulle entità del dominio applicativo rappresentato nella base di dati:
 - alcune tuple possono non avere un valore per un qualche attributo
- si introduce un valore speciale (**valore nullo**) che denota la mancanza di valore (spesso denotato con '?')

14

Il modello relazionale

Il concetto di chiave

- Data una relazione, la **chiave** della relazione è un insieme di attributi che distingue tra loro le tuple della relazione
- più precisamente, un insieme X di attributi di una relazione R , è *chiave* di R se verifica entrambe le seguenti proprietà:
 - 1 qualsiasi sia lo stato di R , non esistono due tuple distinte di R che abbiano lo stesso valore per tutti gli attributi in X
 - 2 nessun sottoinsieme proprio di X verifica la proprietà (1)

15

Il modello relazionale

Esempio

Nell'esempio precedente:

$\text{chiave}(\text{Info_Città}) = (\text{Città})$

se non esistono città con lo stesso nome in regioni diverse

$\text{chiave}(\text{Info_Città}) = (\text{Città}, \text{Regione})$

se esistono città con lo stesso nome in regioni diverse

16

Il modello relazionale

Il concetto di chiave

- Una chiave non può avere valori nulli
- una relazione può avere più di un insieme X che verifica le proprietà viste
- in alcuni casi, può essere necessario scegliere una chiave se il sistema usato non supporta più chiavi
- in tal caso, il termine **chiavi candidate** viene usato per indicare le possibili chiavi
- il termine **chiave primaria** viene usato per indicare la chiave selezionata

17

Il modello relazionale

Chiavi candidate

- Un criterio nella scelta della chiave primaria consiste nello scegliere tra le chiavi candidate quella più frequentemente usata nelle interrogazioni
- un altro criterio è scegliere la chiave che contiene il minor numero di attributi

18

Il modello relazionale

Il concetto di chiave esterna

- Date due relazioni R ed R' tali che:
 - R abbia un insieme di attributi X
 - R' abbia come chiave un insieme Y di attributi
- Y è **chiave esterna** di R su R' se Y è un sottoinsieme di X
- in altre parole, se una relazione R ha tra i suoi attributi un insieme di attributi che costituisce la chiave di una relazione R', allora tale insieme di attributi è una chiave esterna di R su R'
- R' è detta **relazione riferita**

19

Il modello relazionale

Il concetto di chiave esterna

- Le chiavi esterne permettono di collegare tra loro tuple di relazioni diverse e costituiscono un meccanismo, detto per valore, per modellare le associazioni tra entità
- una tupla che deve riferire un'altra tupla include tra i suoi attributi uno o più attributi il cui valore è il valore della chiave della seconda tupla

20

Il modello relazionale

Esempio

- Definiamo due relazioni che contengono informazioni riguardanti i dipendenti di un'azienda ed i dipartimenti in cui l'azienda è organizzata
- Le relazioni sono definite come segue:
 - Impiegati (Imp#, Nome, Mansione, Data_A, Stipendio, Premio_P, Dip#)
chiave(Impiegati) = Imp#
chiave_esterna(Impiegati) = Dip#
(relazione riferita: Dipartimenti)
 - Dipartimenti(Dip#, Nome_Dip, Ufficio#, Divisione#, Dirigente)
chiave(Dipartimenti) = Dip#

21

Il modello relazionale

Impiegati

Imp#	Nome	Mansione	Data_A	Stipendio	Premio_P	Dip#
7369	Rossi	ingegnere	17-Dic-80	1600,00	500,00	20
7499	Andrei	tecnico	20-Feb-81	800,00	?	30
7521	Bianchi	tecnico	20-Feb-81	800,00	100,00	30
7566	Rosi	dirigente	02-Apr-81	2975,00	?	20
7654	Martini	segretaria	28-Set-81	800,00	?	30
7698	Blacchi	dirigente	01-Mag-81	2850,00	?	30
7782	Neri	ingegnere	01-Giu-81	2450,00	200,00	10
7788	Scotti	segretaria	09-Nov-81	800,00	?	20
7839	Dare	ingegnere	17-Nov-81	2600,00	300,00	10
7844	Turni	tecnico	08-Set-81	1500,00	?	30
7876	Adami	ingegnere	28-Set-81	1100,00	500,00	20
7900	Gianni	ingegnere	03-Dic-81	1950,00	?	30
7902	Fordi	segretaria	03-Dic-81	1000,00	?	20
7934	Milli	ingegnere	23-Gen-82	1300,00	150,00	10
7977	Verdi	dirigente	10-Dic-80	3000,00	?	10

22

Il modello relazionale

Dipartimenti

Dip#	Nome_Dip	Ufficio	Divisione	Dirigente
10	Edilizia Civile	1100	D1	7977
20	Ricerche	2200	D1	7566
30	Edilizia Stradale	5100	D2	7698

23

Il modello relazionale

Integrità referenziale

- L'integrità referenziale rappresenta un importante vincolo di integrità semantica
- se una tupla t riferisce come valori di una chiave esterna i valori v_1, \dots, v_n allora deve esistere nella relazione riferita una tupla t' con valori di chiave v_1, \dots, v_n
- le relazioni Impiegati e Dipartimenti verificano l'integrità referenziale
- si consideri la seguente tupla e si assuma che sia inserita nella relazione Impiegati:
[Imp#: 7899, Nome: Smith, Mansione: tecnico, Data_A:03-Dic-81, Stipendio:2000, Premio_P: 100, Dip#: 50]
- tale tupla viola l'integrità referenziale in quanto non esiste un dipartimento (nella relazione Dipartimenti) che abbia numero 50

24

Il modello relazionale

Integrità referenziale

- I linguaggi per basi di dati (SQL) permettono all'utente di specificare per quali relazioni e quali attributi è necessario mantenere l'integrità referenziale (e le azioni da eseguire in caso di violazione)

25

Il modello relazionale

Algebra Relazionale

26

Il modello relazionale

Operazioni nel Modello Relazionale

- Le operazioni sulle relazioni possono essere espresse in due formalismi di base:
 - **Algebra relazionale:** le interrogazioni (query) sono espresse applicando operatori specializzati alle relazioni
 - **Calcolo relazionale:** le interrogazioni (query) sono espresse per mezzo di formule logiche che devono essere verificate dalle tuple ottenute come risposta all'interrogazione
 - i due formalismi (sotto opportune ipotesi) sono equivalenti

27

Il modello relazionale

Algebra Relazionale

- Esistono cinque operazioni di base:
 - Unione
 - Differenza
 - Prodotto cartesiano
 - Proiezione
 - Selezione
- queste operazioni definiscono completamente l'algebra relazionale

28

Il modello relazionale

Algebra Relazionale

- Ogni operazione restituisce come risultato una relazione: è pertanto possibile applicare una operazione al risultato di un'altra operazione (proprietà di chiusura)
- esistono operazioni addizionali, che possono essere espresse in termini delle cinque operazioni di base

29

Il modello relazionale

Algebra Relazionale

- Tali operazioni non aggiungono potere espressivo all'insieme delle operazioni di base, ma sono utili come abbreviazioni; di queste la più importante è l'operazione di *join*
- rispetto alla notazione per nome del modello relazionale, può essere utile introdurre una ulteriore operazione di *renaming* che permette di modificare i nomi degli attributi

30

Il modello relazionale

Algebra Relazionale - Unione

- L'unione di due relazioni R, S, indicata con $R \cup S$ è l'insieme delle tuple in R, S o in entrambe
- l'unione di due relazioni può essere fatta solo se hanno lo stesso grado; inoltre il primo attributo di R deve avere dominio compatibile con il primo attributo di S, il secondo attributo di R deve avere dominio compatibile con il secondo attributo di S, e così via
- le tuple duplicate vengono eliminate
- il grado della relazione risultato è uguale al grado delle relazioni operandi

31

Il modello relazionale

Algebra Relazionale - Unione

Esempio:

A	B	C
a	b	c
d	a	f
c	b	d

R

D	E	F
b	g	a
d	a	f

S

$R \cup S$

A	B	C
a	b	c
d	a	f
c	b	d
b	g	a

32

Il modello relazionale

Algebra Relazionale - Differenza

- La differenza di due relazioni R ed S, indicata con $R-S$ è l'insieme delle tuple che sono in R ma non in S
- la differenza (come l'unione) può essere eseguita solo se le relazioni hanno lo stesso grado e gli attributi hanno domini compatibili
- il grado della relazione risultato è uguale al grado delle relazioni operandi

33

Il modello relazionale

Algebra Relazionale - Differenza

- Esempio:

A	B	C
a	b	c
d	a	f
c	b	d

R

D	E	F
b	g	a
d	a	f

S

R-S

A	B	C
a	b	c
c	b	d

34

Il modello relazionale

Algebra Relazionale – Prodotto Cartesiano

- Il prodotto cartesiano di due relazioni R ed S, di grado k_1, k_2 , indicato con:

$R \times S$

è una relazione di grado k_1+k_2 le cui tuple sono tutte le tuple che hanno:

- come prime k_1 componenti le tuple di R
- come seconde k_2 componenti le tuple di S

35

Il modello relazionale

Algebra Relazionale – Prodotto Cartesiano

- Nella relazione risultato i nomi dei primi k_1 attributi sono i nomi degli attributi della relazione R e i nomi degli ultimi k_2 attributi sono i nomi degli attributi della relazione S
- se le due relazioni hanno attributi con lo stesso nome, è necessario ridenominare gli attributi in una delle due relazioni

36

Il modello relazionale

Algebra Relazionale – Prodotto Cartesiano

- Esempio:

A	B	C
a	b	c
d	a	f
c	b	d

R

D	E	F
b	g	a
d	a	f

S

A	B	C	D	E	F
a	b	c	b	g	a
a	b	c	d	a	f
d	a	f	b	g	a
d	a	f	d	a	f
c	b	d	b	g	a
c	b	d	d	a	f

R X S

37

Il modello relazionale

Algebra Relazionale - Proiezione

- La proiezione di una relazione R su un insieme di attributi $A = \{A_1, \dots, A_m\}$, $A \subseteq U_R$, indicata con:

$$\Pi_{A_1, \dots, A_m}(R)$$

- è una relazione di grado m le cui tuple hanno come attributi solo quelli specificati in A
- pertanto la proiezione genera un insieme T di m-tuple tali che se $t = [A_1:v_1, \dots, A_m:v_m]$ è in T allora esiste una tupla t' in R tale che per ogni A_i in A, $t[A_i] = t'[A_i]$
- nella relazione risultato gli attributi hanno l'ordine specificato in A

38

Il modello relazionale

Algebra Relazionale - Proiezione

- Esempio:

A	B	C
a	b	c
d	a	f
c	b	d

R

B
b
a

$\Pi_B(R)$

A	C
a	c
d	f
c	d

$\Pi_{A,C}(R)$

B	A
b	a
a	d
b	c

$\Pi_{B,A}(R)$

39

Il modello relazionale

Algebra Relazionale - Predicati

- Un predicato F su una relazione R ha una delle seguenti forme:
 - predicato semplice
 - combinazione booleana di predicati semplici (tali combinazioni sono ottenute con i connettivi AND, OR e NOT)

40

Il modello relazionale

Algebra Relazionale - Predicati

- Un predicato semplice ha una delle seguenti forme:
 - $A \text{ op } \textit{costante}$
 - $A \text{ op } A'$

A e A' sono attributi di R , \textit{op} è un operatore relazionale di confronto $>, <, >=, <=, =$, ecc., *costante* è una costante compatibile con il dominio di A

41

Il modello relazionale

Algebra Relazionale – Esempi di predicati

- $B=b$ predicato semplice forma 1)
- $A=C$ predicato semplice forma 2)
- $B=b \text{ OR } A=C$ combinazione booleana
- $B=b \text{ AND } A=C$ combinazione booleana
- $\text{NOT } B=b$ combinazione booleana

42

Il modello relazionale

Algebra Relazionale - Selezione

- La selezione su una relazione R , dato un predicato F , indicata con:

$$\sigma_F(R)$$

è una relazione che contiene tutte e sole le tuple che verificano il predicato F

- il grado della relazione risultato è uguale al grado della relazione operando; i nomi degli attributi della relazione risultato sono gli stessi della relazione operando
- se nessuna tupla di R verifica il predicato F , allora il risultato è una relazione vuota

43

Il modello relazionale

Algebra Relazionale - Selezione

- Se k è il grado di R , la selezione genera un insieme T di k -tuple
- sia $t = A_1.v_1, \dots, A_k.v_k$ una k -tupla in T
- t è tale che:

$$F(A_1/t[A_1], \dots, A_k/t[A_k])$$

è vera, dove la notazione $A_i/t[A_i]$ indica la sostituzione in F del nome di attributo A_i (se tale nome compare in F) con il valore $t[A_i]$ dell'attributo di nome A_i in t

44

Il modello relazionale

Algebra Relazionale - Selezione

- Esempio:

A	B	C
a	b	c
d	a	f
c	b	d

R

A	B	C
a	b	c
c	b	d

$\sigma_{B=b}(R)$

A	B	C
d	a	f

$\sigma_{NOTB=b}(R)$

A	B	C
a	b	c
c	b	d

$\sigma_{B=b \text{ OR } A=C}(R)$

$\sigma_{B=b \text{ AND } A=C}(R)$

45

Il modello relazionale

Algebra Relazionale - Ridenominazione

- La ridenominazione di una relazione R rispetto ad una lista di coppie di nomi di attributi $(A_1, B_1), \dots, (A_m, B_m)$ tale che A_i è un nome di attributo di R, denotata con:

$$\rho_{A_1, \dots, A_m \leftarrow B_1, \dots, B_m}(R)$$

ridenomina l'attributo di nome A_i con il nome B_i

- la ridenominazione è corretta se il nuovo schema della relazione R ha attributi con nomi tutti distinti

46

Il modello relazionale

Algebra Relazionale - Ridenominazione

- Esempio:
La ridenominazione:

$$\rho_{A,B,C \leftarrow AA, BB, CC}(R)$$

cambia lo schema $R(A,B,C)$ in $R(AA, BB, CC)$

47

Il modello relazionale

Algebra Relazionale – Esempi

- Q1: selezionare i nomi degli impiegati che hanno uno stipendio maggiore di 2000

$$\Pi_{\text{Nome}}(\sigma_{\text{Stipendio} > 2000}(\text{Impiegati}))$$

Nome
Rosi
Blacchi
Neri
Dare
Verdi

48

Il modello relazionale

Algebra Relazionale - Esempi

- Q2: selezionare i nomi ed i numeri di dipartimento degli impiegati che hanno uno stipendio maggiore di 2000 e hanno mansione di ingegnere

$\Pi_{\text{Nome, Dip\#}} (\sigma_{\text{Stipendio} > 2000 \text{ AND } \text{Mansione} = \text{'ingegnere'}} (\text{Impiegati}))$

Nome	Dip#
Neri	10
Dare	10

49

Il modello relazionale

Algebra relazionale - Operazioni di base

- Sia $R = (A_1, \dots, A_k)$ uno schema di relazione
- indichiamo con $\mathfrak{R}(R)$ l'insieme di tutte le relazioni su tale schema

- $_ \cup _ : \mathfrak{R}(R) \times \mathfrak{R}(R) \rightarrow \mathfrak{R}(R)$
 $r1 \cup r2 = \{t \mid t \in r1 \vee t \in r2\}$

- $_ \setminus _ : \mathfrak{R}(R) \times \mathfrak{R}(R) \rightarrow \mathfrak{R}(R)$
 $r1 \setminus r2 = \{t \mid t \in r1, t \notin r2\}$

50

Il modello relazionale

Algebra relazionale - Operazioni di base

- $_ \times _ : \mathfrak{R}(R1) \times \mathfrak{R}(R2) \rightarrow \mathfrak{R}(R1 \cup R2)$
 con $R1 \cap R2 = \emptyset$
 $r1 \times r2 = \{t1 \cdot t2 \mid t1 \in r1, t2 \in r2\}$
- $\pi_{R'} _ : \mathfrak{R}(R) \rightarrow \mathfrak{R}(R')$ con $R' \subseteq R$
 $\pi_{R'}(r) = \{t[R'] \mid t \in r\}$
- $\sigma_{F_} : \mathfrak{R}(R) \rightarrow \mathfrak{R}(R)$
 $\sigma_F(r) = \{t \mid t \in r, F(t)\}$

51

Il modello relazionale

Algebra Relazionale - Join

- Il join di due relazioni R ed S sugli attributi A di R ed A' di S, indicato con
 $R \bowtie_{A\theta A'} S$
 è definito come $\sigma_{A\theta A'}(R \times S)$
- il join è quindi un prodotto cartesiano seguito da una selezione
- il predicato $A\theta A'$ è detto predicato di join

52

Il modello relazionale

Algebra Relazionale - Join

- Il grado della relazione risultato è uguale alla somma dei gradi delle relazioni operandi
- spesso il join è indicato con le seguenti notazioni: $R.A\theta S.A'$ oppure $R[A\theta A']S$
- il join prende il nome di equijoin quando l'operatore θ nel predicato di join è l'operatore di uguaglianza

53

Il modello relazionale

Algebra Relazionale - Join

- Esempio:

A	B	C
1	2	3
4	5	6
7	8	9

R

D	E
3	1
6	2

S

A	B	C	D	E
1	2	3	3	1

$R \bowtie_{A=E} S$

A	B	C	D	E
1	2	3	3	1
1	2	3	6	2
4	5	6	6	2

$R \bowtie_{B<D} S$

54

Il modello relazionale

Algebra Relazionale – Join naturale

- L'operazione di join naturale è una 'semplificazione' del join
- si consideri l'interrogazione "ritrovare tutti gli impiegati e gli uffici dove lavorano". Se usiamo il join, tale interrogazione è espressa come:

$\Pi_{Nome, Ufficio} (Impiegato \bowtie_{Impiegato.Dip\#=Dipartimento.Dip\#} Dipartimento)$

- notare che questo join impone l'uguaglianza degli attributi che appaiono in entrambe le relazioni

55

Il modello relazionale

Algebra Relazionale – Join naturale

- E' un tipo di join molto frequente
- l'operazione di join naturale indica un tipo di join basato sull'eguaglianza degli attributi comuni a due relazioni
- ha senso solo nella notazione con nome, a differenza delle altre operazioni

56

Il modello relazionale

Algebra Relazionale – Join naturale

- R, S relazioni, $\{A_1, \dots, A_k\} = U_R \cap U_S$ insieme degli attributi presenti sia nello schema di S che nello schema di R, $\{I_1, \dots, I_m\} = U_R \cup U_S$ insieme degli attributi nello schema di R o nello schema di S

- l'espressione che definisce il join naturale è

$$\Pi_{I_1, \dots, I_m} \left(\sigma_c \left(R \times \left(\rho_{A_1, \dots, A_k \leftarrow S.A_1, \dots, S.A_k} (S) \right) \right) \right)$$

57

Il modello relazionale

Algebra Relazionale – Join naturale

- Nella formula precedente c indica la formula

$$A_1 = S.A_1 \text{ AND } A_2 = S.A_2 \text{ AND } \dots \text{ AND } A_k = S.A_k$$

- il join naturale esegue pertanto un join uguagliando gli attributi con lo stesso nome delle due relazioni e poi elimina gli attributi duplicati

- il join naturale si indica con $R \triangleright \triangleleft S$

58

Il modello relazionale

Algebra Relazionale – Join naturale

- Esempio:

A	B	C
a	b	c
d	b	c
b	b	f
c	a	d

R

B	C	D
b	c	d
b	c	e
a	d	b

S

A	B	C	D
a	b	c	d
a	b	c	e
d	b	c	d
d	b	c	e
c	a	d	b

$R \bowtie S$

59

Il modello relazionale

Algebra relazionale - Divisione

- si supponga che gli impiegati della base di dati di esempio siano assegnati a dei corsi di aggiornamento; ogni impiegato in genere partecipa a più corsi e viceversa ogni corso è seguito da più impiegati; vengono inoltre rappresentate le informazioni relativamente ai corsi
- si supponga, pertanto, che siano definite le seguenti relazioni

Segue (Imp#, Corso#)

Corsi (Corso#, Argomento, Durata)

60

Il modello relazionale

Algebra relazionale - Divisione

- un possibile contenuto delle due relazioni è il seguente:

Segue		Corsi		
Imp#	Corso#	Corso#	Argomento	Durata
7369	10	20	CAD	5
7369	20	10	Basi di dati	3
7369	30	30	Basi di dati	2
7782	10	40	Sistemi oper.	4
7782	40			

61

Il modello relazionale

Algebra relazionale - Divisione

- consideriamo la seguente interrogazione: "trovare il numero di impiegato degli impiegati che seguono tutti i corsi il cui argomento è basi di dati"
- il numero di corso dei corsi il cui argomento è basi di dati è ottenuto come segue:

$$R1 = \Pi_{\text{Corso\#}} (\sigma_{\text{Argomento} = \text{'Basi di dati'}} (\text{Corsi}))$$

il risultato dell'espressione R1 è il seguente insieme di numeri di corso {10,30}

62

Il modello relazionale

Algebra relazionale - Divisione

- il risultato della interrogazione è dato, pertanto, da tutti quegli impiegati che appaiono nella relazione Segue con ognuno dei numeri di corso determinati da R1
- il risultato della nostra interrogazione è pertanto solo l'impiegato il cui numero è 7369
- l'operazione che permette di eseguire l'interrogazione precedente è l'operazione di divisione

63

Il modello relazionale

Algebra relazionale - Divisione

- Date due relazioni R ed S con insiemi di attributi U_R ed U_S , rispettivamente, e tali che $U_R \supset U_S$, l'operazione di divisione di R per S è denotata da

$$R \div S$$

ed è espressa come segue:

$$\Pi_{(U_R - U_S)}(R) - \Pi_{(U_R - U_S)}((\Pi_{(U_R - U_S)}(R) \times S) - R)$$

- l'espressione a destra del $-$ determina tutte le tuple di R che non sono associate ad almeno una tupla di S

64

Il modello relazionale

Algebra relazionale - Divisione

- l'interrogazione dell'esempio precedente è espressa come segue

$$\text{Segue} \div \Pi_{\text{Corso\#}} (\sigma_{\text{Argomento} = \text{'Basi di dati'}} (\text{Corsi}))$$

$$R = \text{Segue}$$

$$S = \Pi_{\text{Corso\#}} (\sigma_{\text{Argomento} = \text{'Basi di dati'}} (\text{Corsi}))$$

$$S = \{10, 30\}$$

$$U_R = \{\text{Imp\#}, \text{Corso\#}\}$$

$$U_S = \{\text{Corso\#}\}$$

65

Il modello relazionale

Algebra relazionale - Divisione

$$1) \Pi_{(U_R - U_S)} (R) = \Pi_{\text{Imp\#}} (R) = \begin{array}{c} \text{Imp\#} \\ \hline 7369 \\ 7782 \end{array}$$

$$2) \Pi_{(U_R - U_S)} (R) \times S = \Pi_{\text{Imp\#}} (R) \times S =$$

Imp#	X	Corso#	=	Imp#	Corso#
7369		10		7369	10
7782		30		7369	30
				7782	10
				7782	30

66

Il modello relazionale

Algebra relazionale - Divisione

$$3) (\Pi_{(UR - US)}(R) \times S) - R = (\Pi_{Imp\#}(R) \times S) - R =$$

Imp#	Corso#	-	Imp#	Corso#	=	Imp#	Corso#
7369	10		7369	10		7782	30
7369	30		7369	20			
7782	10		7369	30			
7782	30		7782	10			
			7782	40			

67

Il modello relazionale

Algebra relazionale - Divisione

$$4) \Pi_{(UR - US)}((\Pi_{(UR - US)}(R) \times S) - R) =$$

$$\Pi_{Imp\#}((\Pi_{Imp\#}(R) \times S) - R) = \frac{Imp\#}{7782}$$

$$5) \Pi_{(UR - US)}(R) - \text{Calcolato al passo (1)}$$

$$\Pi_{(UR - US)}((\Pi_{(UR - US)}(R) \times S) - R) \text{ Calcolato al passo (4)}$$

Imp#	-	Imp#	=	Imp#
7369		7782		7369
7782				

68

Il modello relazionale

Algebra relazionale - Operazioni derivate

- Sia $R = (A_1, \dots, A_k)$ uno schema di relazione
- indichiamo con $\mathfrak{R}(R)$ l'insieme di tutte le relazioni su tale schema
- $_ \cap _ : \mathfrak{R}(R) \times \mathfrak{R}(R) \rightarrow \mathfrak{R}(R)$
 $r1 \cap r2 = r1 \setminus (r1 \setminus r2) = \{t \mid t \in r1, t \in r2\}$
- $_ |x|_F _ : \mathfrak{R}(R1) \times \mathfrak{R}(R2) \rightarrow \mathfrak{R}(R1 \cup R2)$
 con $R1 \cap R2 = \emptyset$
 $r1 |x|_F r2 = \sigma_F(r1 \times r2) =$
 $\{t1 \cdot t2 \mid t1 \in r1, t2 \in r2, F(t1, t2)\}$

69

Il modello relazionale

Algebra relazionale - Operazioni derivate

- $_ |x| _ : \mathfrak{R}(R1) \times \mathfrak{R}(R2) \rightarrow \mathfrak{R}(R1 \cup R2)$
 $r1 |x| r2 = \{t \mid t[R1] \in r1, t[R2] \in r2\}$
 - se $R1 \cap R2 = \emptyset$ $r1 |x| r2 = r1 \times r2$
 - se $R1 = R2$ $r1 |x| r2 = r1 \cap r2$
- $_ \div _ : \mathfrak{R}(R1) \times \mathfrak{R}(R2) \rightarrow \mathfrak{R}(R1 \setminus R2)$ con $R1 \supset R2$
 $r1 \div r2 = \{t \mid \forall t2 \in r2 \exists t1 \in r1 \text{ t.c.}$
 $t1[R1 \setminus R2] = t, t1[R2] = t2\}$

70

Il modello relazionale

Calcolo Relazionale

71

Il modello relazionale

Algebra vs. calcolo

- L'algebra relazionale è un linguaggio procedurale: nello specificare un'espressione algebrica, dobbiamo indicare le operazioni necessarie per calcolare il risultato della query, insieme all'ordine in cui queste operazioni devono essere svolte
- nel calcolo relazionale viene data una descrizione formale del risultato, senza specificare come ottenerlo

72

Il modello relazionale

Calcolo relazionale - Varianti

- Due varianti del calcolo relazionale:
 - Tuple relational calculus (TRC)
 - le variabili rappresentano tuple
 - Domain relational calculus (DRC)
 - le variabili rappresentano valori di domini

73

Il modello relazionale

Calcolo relazionale

- In TRC una query è un'espressione della forma
 $\{t:U|P(t)\}$
ossia è definita come l'insieme di tutte le tuple definite su un insieme di attributi U tali che il predicato P è vero per t
- notazione: t.A indica il valore della tupla t per l'attributo A, $t \in R$ indica che t è nella relazione R
- esempio: determinare tutti gli impiegati il cui stipendio è maggiore di 2000
 $\{t: U_{\text{Impiegati}} | t \in \text{Impiegati} \wedge t.\text{Stipendio} > 2000\}$

74

Il modello relazionale

Calcolo relazionale - Esempi

- Trovare il nome degli impiegati il cui stipendio è maggiore di 2000
 $\{t:\{\text{Nome}\}|\exists s(s \in \text{Impiegati} \wedge s.\text{Stipendio} > 2000 \wedge s.\text{Nome} = t.\text{Nome})\}$
- t è una variabile che indica tuple appartenenti ad una relazione che ha come schema $\{\text{Nome}\}$
- la notazione $\exists t(Q(t))$ indica che esiste una tupla t tale che $Q(t)$ è vera

75

Il modello relazionale

Calcolo relazionale - Esempi

- Trovare i nomi e gli uffici degli impiegati che hanno uno stipendio maggiore di 2000
 $\{t:\{\text{Nome}, \text{Ufficio}\}|\exists s(s \in \text{Impiegati} \wedge s.\text{Stipendio} > 2000 \wedge s.\text{Nome} = t.\text{Nome} \wedge \exists u(u \in \text{Dipartimenti} \wedge s.\text{Dip\#} = u.\text{Dip\#} \wedge u.\text{Ufficio} = t.\text{Ufficio}))\}$

76

Il modello relazionale

Calcolo relazionale - Esempi

- Trovare i nomi degli impiegati che hanno uno stipendio maggiore di 2000 oppure lavorano in un dipartimento della divisione D1
 $\{t:\{\text{Nome}\}|\exists s(s \in \text{Impiegati} \wedge s.\text{Nome}=t.\text{Nome} \wedge s.\text{Stipendio}>2000 \vee \exists u(u \in \text{Dipartimenti} \wedge s.\text{Dip\#}=u.\text{Dip\#} \wedge u.\text{Divisione}=\text{"D1"}))\}$

77

Il modello relazionale

Calcolo relazionale - Sintassi

- *Atomi* – gli atomi sono:
 - $s \in R$ (R è un nome di relazione ed s è una variabile)
 - la tupla s appartiene alla relazione R
 - $s.A\theta u.A'$ (s ed u sono variabili, θ è un operatore relazionale di confronto, A ed A' sono nomi di attributi)
 - il valore di A nella tupla s è in relazione θ con il valore di A' nella tupla u
 - $s.A\theta a$ (s è una variabile, θ è un operatore relazionale di confronto, A è un nome di attributo, a è una costante)
 - il valore di A nella tupla s è in relazione θ con il valore a

78

Il modello relazionale

Calcolo relazionale - sintassi

- *Formule*

- ogni atomo è una formula, tutte le occorrenze delle variabili dell'atomo sono libere
- se ϕ_1 e ϕ_2 sono formule, allora $\phi_1 \wedge \phi_2$, $\phi_1 \vee \phi_2$, $\neg \phi_1$ sono formule, le occorrenze delle variabili sono libere o legate a seconda di come sono in ϕ_1 e ϕ_2
- se ϕ è una formula allora $\exists s(\phi)$, $\forall s(\phi)$ sono formule tutte le occorrenze di s in ϕ sono legate al quantificatore \exists (oppure \forall)
- se ϕ è una formula allora (ϕ) è una formula

79

Il modello relazionale

Calcolo relazionale - Sintassi

- *Variabili libere e legate*

- data una formula F ed una variabile x , x è libera in F se e solo se x non è quantificata
 - $\forall x$ (quantificazione universale)
 - $\exists x$ (quantificazione esistenziale)

80

Il modello relazionale

Calcolo relazionale - Esempi

- $\exists s(s \in \text{Impiegati} \wedge s.\text{Stipendio} > 2000)$
è una formula legale, tutte le occorrenze di s sono legate
- $\exists s(s \in \text{Impiegati} \wedge x.\text{Stipendio} > 2000 \wedge x.\text{Dip\#} = y.\text{Dip\#})$
è una formula legale, tutte le occorrenze di s sono legate mentre quelle di x e y sono libere

81

Il modello relazionale

Calcolo relazionale - Sintassi

- *Espressioni del calcolo*
 - Un'espressione (o query) del calcolo su tuple ha la forma
$$\{x:U|f(x)\}$$
dove U è un insieme di attributi, f è una formula legale del calcolo, x è libera in $f(x)$ ed è l'unica variabile libera in $f(x)$

82

Il modello relazionale

Calcolo relazionale - Esempi

- L'espressione
 $\{y:\{\text{Dip}\#\}|\exists x(x\in \text{Impiegati}\wedge x.\text{Stipendio}>2000\wedge x.\text{Dip}\#=y.\text{Dip}\#)\}$
 è un'espressione corretta di TRC che è soddisfatta da tutti i numeri dei dipartimenti che hanno almeno un impiegato che guadagna più di 2000
- L'espressione
 $\{y:U_{\text{Impiegati}}|\forall y(y\in \text{Impiegati}\wedge y.\text{Mansione}=\text{'ingegnere'})\}$
 non è un'espressione corretta di TRC, in quanto y non è libera

83

Il modello relazionale

Calcolo relazionale – Esprimere l'algebra con il TRC

- Unione: $R\cup S$ $\{t:U_R|t\in R\vee t\in S\}$
- Differenza: $R-S$ $\{t:U_R|t\in R\wedge\neg t\in S\}$
- Prodotto cartesiano:
 $R\times S$ siano $U_R=\{A_1,\dots,A_n\}$ e $U_S=\{A'_1,\dots,A'_m\}$ gli insiemi degli attributi di R ed S
 $\{t:U_R\cup U_S|\exists x\exists y(x\in R\wedge y\in S\wedge x.A_1=t.A_1\wedge\dots\wedge x.A_n=t.A_n\wedge y.A'_1=t.A'_1\wedge\dots\wedge y.A'_m=t.A'_m)\}$

84

Il modello relazionale

Calcolo relazionale – Esprimere l'algebra con il TRC

- Proiezione: $\Pi_{A_1, \dots, A_k}(R)$
 $\{t: \{A_1, \dots, A_k\} \mid \exists x (x \in R \wedge x.A_1 = t.A_1 \wedge \dots \wedge x.A_k = t.A_k)\}$
- Selezione: $\sigma_F(R)$
 $\{t: U_R \mid t \in R \wedge F'\}$
dove F' è la formula F con ogni attributo A sostituito da $t.A$

85

Il modello relazionale

Calcolo relazionale – Potere espressivo

- L'algebra e il calcolo relazionale hanno lo stesso potere espressivo?
- la semantica di un'interrogazione (in algebra o in calcolo) è una funzione che trasforma una base di dati relazionale (insieme di relazioni) in una nuova base di dati relazionale
- algebra e calcolo hanno lo stesso potere espressivo se per ogni interrogazione Q_1 in uno dei due formalismi esiste un'interrogazione Q_2 nell'altro la cui semantica è la stessa funzione

86

Il modello relazionale

Calcolo relazionale – Potere espressivo

- Non tutte le espressioni del calcolo possono essere tradotte in equivalenti espressioni dell'algebra
- esempio: l'espressione

$$\{t:U_R \mid \neg t \in R\}$$

- sebbene sintatticamente corretta, se almeno uno dei domini degli attributi di R è un insieme infinito, questa espressione è soddisfatta da un numero infinito di tuple
- il risultato non sarebbe una relazione!

87

Il modello relazionale

Calcolo relazionale – Potere espressivo

- Nozione di *formula indipendente dal dominio*
- una formula è indipendente dal dominio se la sua valutazione genera sempre lo stesso risultato anche supponendo di estendere la base di dati con nuove relazioni o nuove tuple, contenenti valori non presenti nella base di dati di partenza
- la nozione di indipendenza dal dominio è però indecidibile
- si introduce quindi una condizione sintattica (*safety*) sufficiente a garantire l'indipendenza dal dominio

88

Il modello relazionale

Calcolo relazionale – Potere espressivo

- L'espressione
 $\{t:U_R \mid \neg t \in R\}$
non è safe
- il calcolo relazionale safe e l'algebra relazionale hanno lo stesso potere espressivo
- la traduzione da un formalismo all'altro può essere effettuata in tempo polinomiale nella dimensione dell'espressione