Il Modello Relazionale

1

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

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 <u>di tipo dichiarativo</u>, ma al tempo stesso potenti, con cui esprimere le operazioni per l' accesso e la manipolazione dei dati

Relazione

- D_1, D_2, …, D_n (n insiemi anche non distinti)
- il prodotto cartesiano $D_1 \times D_2 \times \ldots \times D_n$, è l'insieme di tutte le **tuple** ordinate (d_1, d_2, …, d_n) tali che $d_1 \in D_1$, $d_2 \in D_2$, …, $d_n \in D_n$
- una relazione su D_1, D_2, …, D_n è un sottoinsieme del prodotto cartesiano $D_1 \times D_2 \times \ldots \times D_n$
- D_1, D_2, …, 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

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}

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

Relazione matematica, proprietà

- In base alle definizione, una relazione matematica è un **insieme** di tuple **ordinate**:
  (d_1, d_2, …, d_n) tali che $d_1 \in D_1$, $d_2 \in D_2$, …, $d_n \in D_n$
- 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

Notazioni

- Sia r una relazione di grado k:
  - sia t una tupla di r
  - sia i un intero appartenente all' insieme{1,...,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

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**

Modello relazionale

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

Esempio

- Info_Città

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

Schema: Info_Città(Città,Regione,Popolazione)

Modello relazionale

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

Esempio

t=[Città: Roma, Regione: Lazio, Popolazione:3000000]

è una tupla definita sullo schema Info_Città

t[Città] = 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

chiave(Info_Città) = (Città)

se non esistono città con lo stesso nome in regioni diverse

chiave(Info_Città) = (Città,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**

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

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#

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

Dipartimenti

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

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 $v_1,\ldots\ldots,v_n$ allora deve esistere nella relazione riferita una tupla t' con valori di chiave $v_1,\ldots\ldots,v_n$
- 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

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

31

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 relazione 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

32

Algebra Relazionale - Unione

Esempio:

R

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

S

D	E	F
b	g	a
d	a	f

$R \cup S$

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

33

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

34

Algebra Relazionale - Differenza

- Esempio:

R

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

S

D	E	F
b	g	a
d	a	f

R-S

A	B	C
a	b	c
c	b	d

35

Algebra Relazionale – Prodotto Cartesiano

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

  R X 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

36

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

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

Algebra Relazionale - Proiezione

- La proiezione di una relazione R su un insieme di attributi A={A_1,…A_m}, A ⊆ U_R, indicata con:

$$\Pi_{A1,\dots,Am}(R)$$

- 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,…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

Algebra Relazionale - Proiezione

- Esempio:

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

R

A	C
a	c
d	f
c	d

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

B	A
b	a
a	d
b	c

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

B
b
a

$\Pi_B(R)$

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)

Algebra Relazionale - Predicati

- Un predicato semplice ha una delle seguenti forme:
  - A op *costante*
  - A op A'

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

Algebra Relazionale – Esempi di predicati

- B=b predicato semplice forma 1)
- A=C predicato semplice forma 2)
- B=b OR A=C combinazione booleana
- B=b AND A=C combinazione booleana
- NOT B=b combinazione booleana

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 $t=[A_1:v_1, \ldots A_k:v_k]$ una k-tupla in T
- t è tale che:
  $$F(A_1/t[A_1], \ldots, A_k/t[A_k])$$
  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

$\sigma_{NOT B=b}(R)$

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 $(A_1,B_1), \ldots, (A_m,B_m)$ tale che A_i è un nome di attributo di R, denotata con:
  $$\rho_{A_1,\ldots,A_m \leftarrow B_1,\ldots,B_m}(R)$$
  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

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

Nome	Dip#
Neri	10
Dare	10

Algebra relazionale - Operazioni di base

- Sia R = (A_1, ..., A_k) uno schema di relazione
- indichiamo con $\Re$(R) l' insieme di tutte le relazioni su tale schema
- $\_ \cup \_ : \Re(R) \times \Re(R) \to \Re(R)$
  $r1 \cup r2 = \{t \mid t \in r1 \vee t \in r2\}$
- $\_ \setminus \_ : \Re(R) \times \Re(R) \to \Re(R)$
  $r1 \setminus r2 = \{t \mid t \in r1, t \notin r2\}$

Algebra relazionale - Operazioni di base

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

Algebra Relazionale - Join

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

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.AθS.A' oppure R[AθA']S
- il join prende il nome di equijoin quando l'operatore θ nel predicato di join è l'operatore di uguaglianza

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

$R \rhd\lhd_{A=E} S$

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

$R \rhd\lhd_{B<D} S$

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:

$$\Pi_{Nome,Ufficio}\left(\text{Impiegato} \rhd\lhd_{\text{Impiegato.Dip\#=Dipartimento.Dip\#}} \text{Dipartimento}\right)$$

- 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, $\{A_1,\ldots,A_k\}=U_R\cap U_S$ insieme degli attributi presenti sia nello schema di S che nello schema di R, $\{I_1,\ldots,I_m\}= U_R\cup U_S$ insieme degli attributi nello schema di R o nello schema di S
- l'espressione che definisce il join naturale è

$$\Pi_{\text{I1,...,Im}}\left(\sigma_c\left(R\times\left(\rho_{\text{A1,...,Ak}\leftarrow\text{S.A1,...,S.Ak}}(S)\right)\right)\right)$$

57 Il modello relazionale

Algebra Relazionale – Join naturale

- Nella formula precedente *c* indica la formula

  A1=S.A1 AND A2 = S.A2 AND … AND Ak = S.Ak
- 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 $R \rhd\lhd S$

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

$R \rhd\lhd S$

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

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

Corsi

Corso#	Argomento	Durata
20	CAD	5
10	Basi di dati	3
30	Basi di dati	2
40	Sistemi oper.	4

61

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:

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

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

62

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

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:

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

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

64

Algebra relazionale - Divisione

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

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

R = Segue

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

S = {10, 30}

U_R = {Imp#, Corso#}

U_S = {Corso#}

65

Algebra relazionale - Divisione

1) $\Pi_{(UR - US)} (R) = \Pi_{Imp\#} (R) =$

Imp#
7369
7782

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

Imp#
7369
7782

X

Corso#
10
30

=

Imp#	Corso#
7369	10
7369	30
7782	10
7782	30

66

Algebra relazionale - Divisione

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

Imp#	Corso#
7369	10
7369	30
7782	10
7782	30

\-

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

=

Imp#	Corso#
7782	30

Algebra relazionale - Divisione

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

Imp#
7782

5) $\Pi_{(UR - US)}(R) -$ Calcolato al passo (1)
$\Pi_{(UR - US)}((\Pi_{(UR - US)}(R) \times S) - R)$ Calcolato al passo (4)

Imp#
7369
7782

\-

Imp#
7782

=

Imp#
7369

Algebra relazionale - Intersezione

- L’intersezione di due relazioni R, S, indicata con
  $R \cap S$
  è l’insieme delle tuple contenute in R e in S
- l’intersezione di due relazione 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
- il grado della relazione risultato è uguale al grado delle relazioni operandi

Algebra relazionale - Intersezione

- L’intersezione è un’operazione derivata
- può infatti essere definita come segue:
  $R \cap S = R - (R - S)$
- Cioè:
  - si determinano le tuple che stanno in R ma non in S (R-S)
  - si tolgono da R tutte le tuple che non stanno in S (R - (R - S))
  - le tuple risultanti stanno sia in R che in S

Algebra relazionale - Intersezione

- Esempio

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

R

D	E	F
b	g	a
d	a	f

S

$R \cap S$

A	B	C
d	a	f

Algebra relazionale - Intersezione

- R - S

A	B	C
a	b	c
c	b	d

- R - (R-S)

A	B	C
d	a	f

Algebra relazionale - Operazioni derivate

- Sia R = (A_1, ..., A_k) uno schema di relazione
- indichiamo con $\Re(R)$ l' insieme di tutte le relazioni su tale schema
- $\_ \cap \_ : \Re(R) \times \Re(R) \to \Re(R)$

  $r1 \cap r2 = r1 \setminus (r1 \setminus r2) = \{t \mid t \in r1, t \in r2\}$
- $\_ |x|_F \_ : \Re(R1) \times \Re(R2) \to \Re(R1 \cup R2)$

  con $R1 \cap R2 = \varnothing$

  $r1 |x|_F r2 = \sigma_F (r1 \times r2) =$

  $\{t1 \cdot t2 \mid t1 \in r1, t2 \in r2, F(t1,t2)\}$

73 Il modello relazionale

Algebra relazionale - Operazioni derivate

- $\_ |x| \_ : \Re(R1) \times \Re(R2) \to \Re(R1 \cup R2)$

  $r1 |x| r2 = \{t \mid t[R1] \in r1, t[R2] \in r2\}$

  - se $R1 \cap R2 = \varnothing$ $\quad r1 |x| r2 = r1 \times r2$
  - se $R1 = R2$ $\quad r1 |x| r2 = r1 \cap r2$
- $\_ \div \_ : \Re(R1) \times \Re(R2) \to \Re(R1 \setminus R2)$ con $R1 \supset R2$

  $r1 \div r2 = \{t \mid \forall t2 \in r2\ \exists t1 \in r1$ t.c.

  $t1[R1 \setminus R2] = t,\ t1[R2] = t2\}$

74 Il modello relazionale

Calcolo Relazionale

75 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

76 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
- vedremo solo TRC

77 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

  $\{t: U_{Impiegati} | t \in Impiegati \wedge t.Stipendio > 2000\}$

78 Il modello relazionale

Calcolo relazionale - Esempi

- Trovare il nome degli impiegati il cui stipendio è maggiore di 2000

  {t:{Nome}|∃s(s∈ Impiegati∧s.Stipendio>2000∧ s.Nome=t.Nome)}
- t è una variabile che indica tuple appartenenti ad una relazione che ha come schema {Nome}
- la notazione ∃t(Q(t)) indica che esiste una tupla t tale che Q(t) è vera

79 Il modello relazionale

Calcolo relazionale - Esempi

- Trovare i nomi e gli uffici degli impiegati che hanno uno stipendio maggiore di 2000

  {t:{Nome,Ufficio}|∃s(s∈ Impiegati∧
  s.Stipendio>2000∧s.Nome=t.Nome∧
  ∃u(u ∈ Dipartimenti ∧s.Dip#=u.Dip#∧
  u.Ufficio=t.Ufficio))}

80 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:{Nome}|∃s(s∈ Impiegati ∧ s.Nome=t.Nome ∧
  s.Stipendio>2000 ∨ ∃u(u ∈ Dipartimenti∧
  s.Dip#=u.Dip#∧u.Divisione="D1")))}

81 Il modello relazionale

Calcolo relazionale - Sintassi

- *Atomi* – gli atomi sono:
  - s∈ R (R è un nome di relazione ed s è una variabile)
    - la tupla s appartiene alla relazione R
  - s.Aθu.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 θa (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

82 Il modello relazionale

Calcolo relazionale - sintassi

- *Formule*
  - ogni atomo è una formula, tutte le occorrenze delle variabili dell'atomo sono libere
  - se φ1 e φ2 sono formule, allora φ1∧φ2, φ1∨φ2, ¬φ1 sono formule, le occorrenze delle variabili sono libere o legate a seconda di come sono in φ1 e φ2
  - se φ è una formula allora ∃s(φ), ∀s(φ) sono formule tutte le occorrenze di s in φ sono legate al quantificatore ∃ (oppure ∀)
  - se φ è una formula allora (φ) è una formula

83 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
    - ∀x (quantificazione universale)
    - ∃x (quantificazione esistenziale)

84 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 mente quelle di x e y sono libere

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)

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

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:
  RxS siano $U_R =\{A_1,\ldots,A_n\}$ e $U_S =\{A'_1,\ldots,A'_m\}$ gli insiemi degli attributi di R ed S
  $\{t: U_R \cup U_S |\exists x(\exists y(x \in R \wedge y \in S \wedge x.A_1=t.A_1 \wedge \ldots \wedge x.A_n=t.A_n \wedge y.A'_1=t.A'_1 \wedge \ldots \wedge y.A'_m=t.A'_m))\}$

Calcolo relazionale – Esprimere l'algebra con il TRC

- Proiezione: $\Pi_{A_1,\ldots A_k}(R)$
  $\{t:\{A_1,\ldots,A_k\}|\exists x(x \in R \wedge x.A_1=t.A_1 \wedge \ldots \wedge x.A_k=t.A_k\}$
- Selezione: $\sigma_F(R)$
  $\{t:U_R|t \in R \wedge F'\}$

dove F' è la formula F con ogni attributo A sostituito da t.A

Calcolo relazionale – Potere espressivo

- L'algebra relazionale e il calcolo relazionale hanno lo stesso potere espressivo?
- Cioè: tutto le operazioni esprimibili mediante espressioni dell'algebra relazionale possono essere espresse mediante espressioni del calcolo relazionale e viceversa?

Calcolo relazionale – 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 Q1 in uno dei due formalismi esiste un'interrogazione Q2 nell'altro la cui semantica è la stessa funzione

91 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!

92 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 i domini associati agli attributi presenti nella base di dati con nuovi valori, non presenti nella base di dati di partenza
- si introduce una condizione sintattica (*safety*) per garantire questa proprietà (non la vediamo)
- **Idea**: mi restringo ad interrogazioni il cui risultato dipende solo da valori contenuti nella base di dati di partenza

93 Il modello relazionale

Calcolo relazionale - Potere espressivo

- la nozione di indipendenza dal dominio è però indecidibile
- Si introduce quindi una condizione sintattica (*safety*) sufficiente a garantire l'indipendenza dal dominio
- non vediamo questa condizione

94 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

95 Il modello relazionale

Perché due linguaggi?

- Algebra relazionale
  - linguaggio procedurale
  - utile per il sistema
- Calcolo relazionale
  - linguaggio dichiarativo
  - utile per l'utente
- i linguaggi di interrogazione nei sistemi reali si basano sul calcolo
- l'algebra relazionale viene utilizzata come linguaggio interno
  - l'interrogazione utente viene compilata in un'espressione dell'algebra relazionale
  - il sistema utilizza le proprietà dell'algebra per stabilire quale espressione permette di risolvere l'interrogazione iniziale nel modo più efficiente

96 Il modello relazionale

97

Esempio

- Determinare i nomi degli impiegati che lavorano nella divisione D1
- Calcolo
  {t:{Nome}|∃s(s∈Impiegati ∧ s.Nome=t.Nome ∧
  ∃u(u ∈Dipartimenti ∧ s.Dip#=u.Dip#∧u.Divisione="D1")))}
- Algebra
  (a) $\Pi_{Nome}\left(\sigma_{Divisione=D1}\left(\text{Impiegati} \rhd\lhd \text{Dipartimenti}\right)\right)$

Il modello relazionale

98

Esempio (continua)

- Poiché l'attributo Divisione ∈ $U_{Dipartimenti}$, è possibile anticipare la selezione
  (b) $\Pi_{Nome}\left(\text{Impiegati} \rhd\lhd \sigma_{Divisione=D1}\left(\text{Dipartimenti}\right)\right)$
- Supponiamo che:
  - Impiegati contenga n_1 tuple
  - Dipartimenti contenga n_2 tuple (ragionevolmente $n_2 <= n_1$)
  - i dipartimenti della divisione D1 sono n_3 ($n_3 <= n_2$)
- Costi
  - (a): join:n_1n_2,selezione: n_1, proiezione: n_1, totale: $n_1n_2 +2n_1$
  - (b): selezione: n_2, join: n_1n_3, proiezione: n_1, totale: $n_1n_3 +n_1$+n2
    - Costo(a) > Costo(b)
  - il sistema sceglie l'espressione (b)

Il modello relazionale