

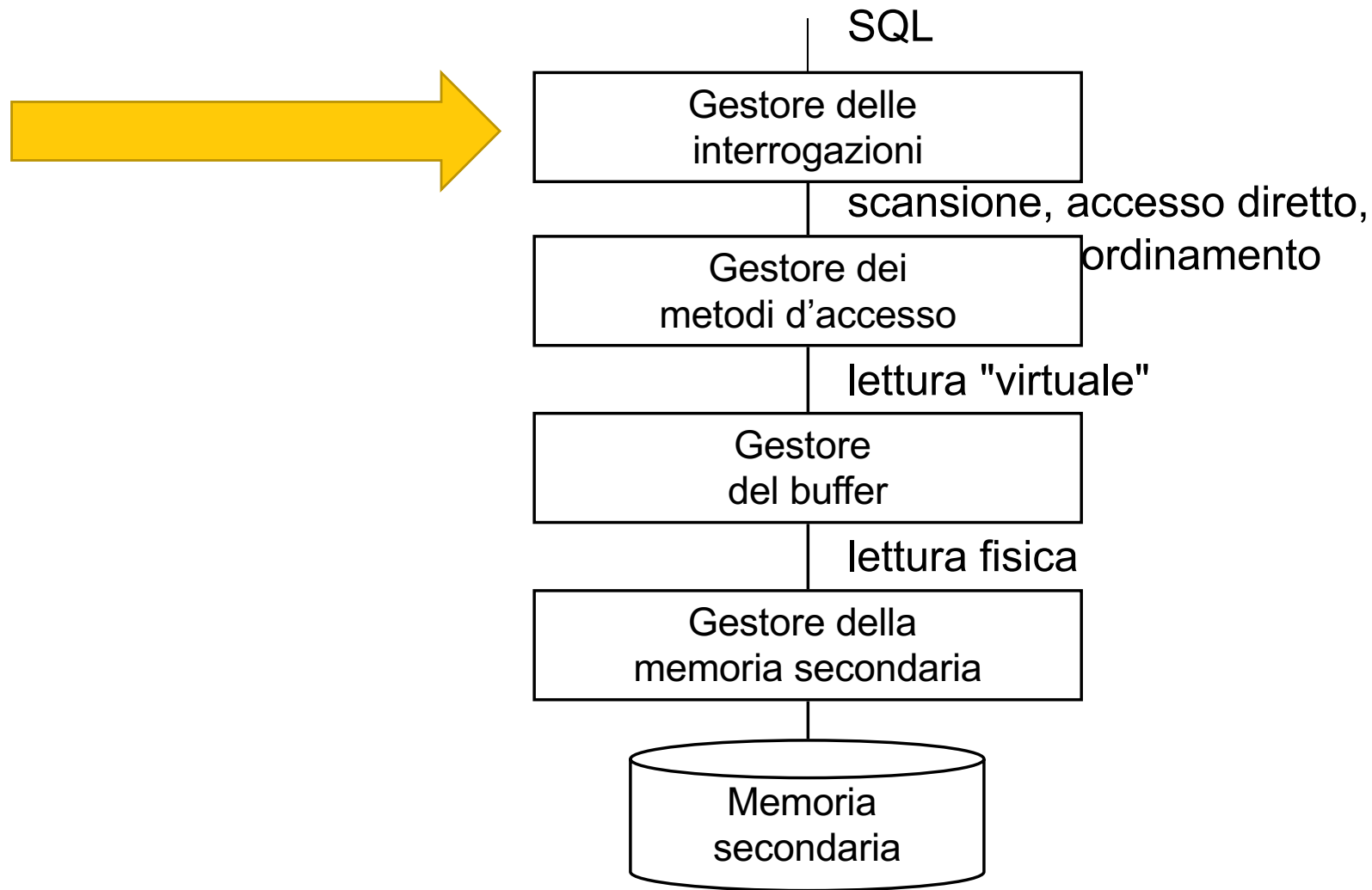
Basi dati II

Ottimizzazioni query

Cap. 7 Sistemi di Gestione dati- Catania et al.

Cap. 16-17 Sistemi di Basi di Dati,
Fondamenti e complementi R. Elmasri et al. Pearson edizioni.
(lucidi tratti dall'ultimo libro)

Gestore degli accessi e delle interrogazioni

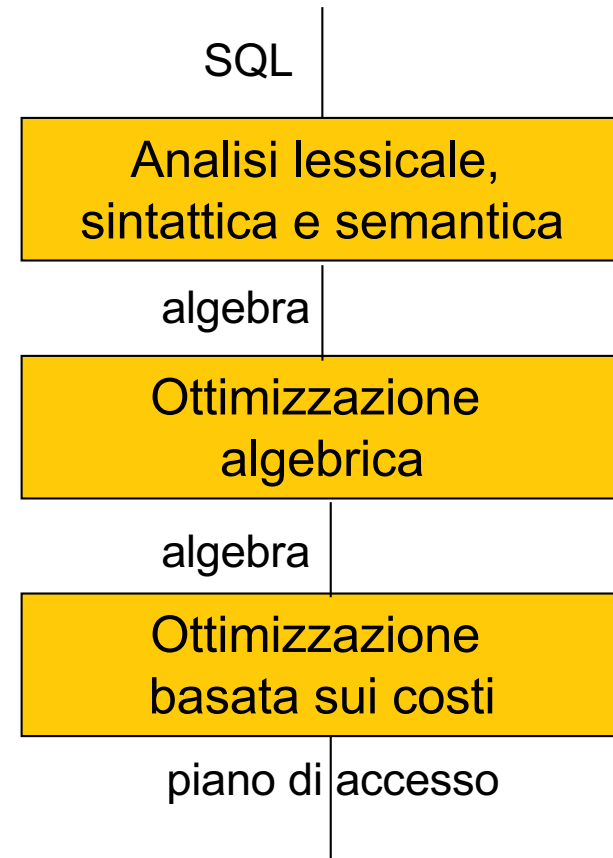


Esecuzione e ottimizzazione delle interrogazioni

- **Query processor** (**Ottimizzatore**):
un modulo del DBMS che decide come eseguire le query
 - Le query sono espresse ad alto livello (SQL)
 - l'ottimizzatore sceglie la strategia per eseguire la query SQL (di solito fra diverse alternative)

Esecuzione e ottimizzazione delle interrogazioni

- **Query processor** (**Ottimizzatore**): un modulo del DBMS che decide come eseguire le query
 - Le query sono espresse ad alto livello (SQL)
 - l'ottimizzatore sceglie la strategia per eseguire la query SQL (di solito fra diverse alternative)



Algebra relazionale

- Insieme di operatori
 - su relazioni
 - che producono relazioni
 - e possono essere composti

Ripasso veloce sulle relazioni

Relazione

- Siano $D_1, D_2, \dots, D_k \in D$ domini

Una relazione su D_1, D_2, \dots, D_k è un sottoinsieme finito del prodotto cartesiano $D_1 \times D_2 \times \dots \times D_k$

- Una relazione, sottoinsieme del prodotto cartesiano di k domini, ha **grado** k
- Ogni tupla di una relazione di grado k ha k componenti, una per ogni dominio su cui è definita la relazione cui la tupla appartiene
- La **cardinalità** di una relazione è il numero di tuple appartenenti alla relazione

Relazioni – notazione per nome

- Viene associato un nome, detto **nome di attributo**, ad ogni componente delle tuple in una relazione

La coppia (nome di attributo, dominio) è detta **attributo**

- L'uso degli attributi permette di
 - denotare le componenti di ogni tupla per nome piuttosto che per posizione
 - fornire maggiori informazioni semantiche sulle proprietà che ogni componente delle tuple in una relazione modella

Relazione - schema

- R nome di relazione
- $\{A_1, A_2, \dots, A_n\}$ un insieme di nomi di attributi
- $\text{dom} : \{A_1, A_2, \dots, A_n\} \rightarrow D$ una funzione totale che associa ad ogni nome di attributo in $\{A_1, A_2, \dots, A_n\}$ il corrispondente dominio

La coppia $(R(A_1, A_2, \dots, A_n), \text{dom})$ è uno **schema di relazione**

- U_R denota l'insieme dei nomi di attributi di R
 - $U_R = \{A_1, A_2, \dots, A_n\}$

Esempio relazioni

Film

titolo	regista	anno	genere	valutaz
underground	emir kusturica	1995	drammatico	3.20
edward mani di forbice	tim burton	1990	fantastico	3.60
nightmare before christmas	tim burton	1993	animazione	4.00
ed wood	tim burton	1994	drammatico	4.00
mars attacks	tim burton	1996	fantascienza	3.00
il mistero di sleepy hollow	tim burton	1999	horror	3.50
big fish	tim burton	2003	fantastico	3.10
la sposa cadavere	tim burton	2005	animazione	3.50
la fabbrica di cioccolato	tim burton	2005	fantastico	4.00
io non ho paura	gabriele salvatores	2003	drammatico	3.50
nirvana	gabriele salvatores	1997	fantascienza	3.00
mediterraneo	gabriele salvatores	1991	commedia	3.80
pulp fiction	quentin tarantino	1994	thriller	3.50
le iene	quentin tarantino	1992	thriller	4.00

Video

colloc	titolo	regista	tipo
1111	underground	emir kusturica	v
1112	underground	emir kusturica	d
1113	big fish	tim burton	v
1114	big fish	tim burton	d
1115	edward mani di forbice	tim burton	d
1116	nightmare before christmas	tim burton	v
1117	nightmare before christmas	tim burton	d
1118	ed wood	tim burton	d
1119	mars attacks	tim burton	d
1120	il mistero di sleepy hollow	tim burton	d
1121	la sposa cadavere	tim burton	d
1122	la fabbrica di cioccolato	tim burton	d
1123	la fabbrica di cioccolato	tim burton	d
1124	io non ho paura	gabriele salvatores	d
1125	nirvana	gabriele salvatores	d
1126	mediterraneo	gabriele salvatores	d
1127	pulp fiction	quentin tarantino	v
1128	pulp fiction	quentin tarantino	d
1129	le iene	quentin tarantino	d

Noleggio

colloc	dataNol	codCli	dataRest
1111	01-Mar-2006	6635	02-Mar-2006
1115	01-Mar-2006	6635	02-Mar-2006
1117	02-Mar-2006	6635	06-Mar-2006
1118	02-Mar-2006	6635	06-Mar-2006
1111	04-Mar-2006	6642	05-Mar-2006
1119	08-Mar-2006	6635	10-Mar-2006
1120	08-Mar-2006	6635	10-Mar-2006
1116	08-Mar-2006	6642	09-Mar-2006
1118	10-Mar-2006	6642	11-Mar-2006
1121	15-Mar-2006	6635	18-Mar-2006
1122	15-Mar-2006	6635	18-Mar-2006
1113	15-Mar-2006	6635	18-Mar-2006
1129	15-Mar-2006	6635	20-Mar-2006
1119	15-Mar-2006	6642	16-Mar-2006
1126	15-Mar-2006	6610	16-Mar-2006
1112	16-Mar-2006	6610	18-Mar-2006
1114	16-Mar-2006	6610	17-Mar-2006
1128	18-Mar-2006	6642	20-Mar-2006
1124	20-Mar-2006	6610	21-Mar-2006
1115	20-Mar-2006	6610	21-Mar-2006
1124	21-Mar-2006	6642	22-Mar-2006
1116	21-Mar-2006	6610	?
1117	21-Mar-2006	6610	?
1127	22-Mar-2006	6635	?
1125	22-Mar-2006	6635	?

Cliente

codCli	nome	cognome	telefono	dataN	residenza
6610	anna	rossi	01055664433	05-Ott-1979	via scribanti 16 16131 genova
6635	paola	bianchi	0104647992	12-Apr-1976	via dodecaneso 35 16146 genova
6642	marco	verdi	3336745383	16-Ott-1972	via lagustena 35 16131 genova

Algebra Relazionale

- Esistono cinque operazioni di base:
 - Proiezione
 - Selezione
 - Prodotto cartesiano
 - Unione
 - Differenza
- Queste operazioni definiscono completamente l'algebra 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 aggiuntive, che possono essere espresse in termini delle cinque operazioni di base
 - Non aumentano il potere espressivo dell'algebra ma sono comode a fini pratici
 - Es: Operazione fondamentale: join

Algebra Relazionale

- Esistono cinque operazioni di base:
 - Proiezione
 - Selezione
 - Prodotto cartesiano
 - Unione
 - Differenza
- Operazioni derivate:
 - Intersezione
 - join

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

- La proiezione genera un insieme T di tuple con m attributi (grado m)
- 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
- La cardinalità di T può essere diversa dalla cardinalità di R
 - Eliminazione duplicati

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

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

Proiezione - esempio

titolo	anno
underground	1995
edward mani di forbice	1990
nightmare before christmas	1993
ed wood	1994
mars attacks	1996
il mistero di sleepy hollow	1999
big fish	2003
la sposa cadavere	2005
la fabbrica di cioccolato	2005
io non ho paura	2003
nirvana	1997
mediterraneo	1991
pulp fiction	1994
le iene	1992

$\Pi_{\text{titolo,anno}}(\text{Film})$

genere
drammatico
fantastico
animazione
fantascienza
horror
commedia
thriller

$\Pi_{\text{genere}}(\text{Film})$

Algebra Relazionale

- Esistono cinque operazioni di base:
 - Proiezione
 - Selezione
 - Prodotto cartesiano
 - Unione
 - Differenza
- Operazioni derivate:
 - Intersezione
 - join

Selezione

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

$$\sigma_F (R)$$

genera una relazione che contiene tutte le tuple di R che
verificano 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
- La cardinalità del risultato è minore o uguale alla cardinalità della relazione di partenza

Selezione - predicati

Un predicato F su una relazione R ha una delle seguenti forme:

- predicato semplice
 - $A \text{ op } v$
 - $A \text{ op } A'$dove:
 - A e A' sono attributi di R
 - op è un operatore relazionale di confronto $>, <, >=, <=, =, \neq$
 - v è una costante compatibile con il dominio di A
- combinazione booleana di predicati semplici, ottenuta con i connettivi

\wedge (AND), \vee (OR), \neg (NOT)

Selezione – esempi di predicati

- $\text{codCli}=6635$ predicato semplice
- $\text{dataNol}=\text{dataRest}$ predicato semplice
- $\text{codCli}=6635 \wedge \text{dataNol}=\text{dataRest}$
 combinazione booleana

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

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

A	B	C
a	b	c
c	b	d

$$\sigma_{B=b \vee A=c}(R)$$

A	B	C

$$\sigma_{B=b \wedge A=c}(R)$$

Selezione - esempio

	colloc	dataNol	codCli	dataRest
$\sigma_{\text{codCli}=6635}(\text{Noleggio})$	1111	01-Mar-2006	6635	02-Mar-2006
	1115	01-Mar-2006	6635	02-Mar-2006
	1117	02-Mar-2006	6635	06-Mar-2006
	1118	02-Mar-2006	6635	06-Mar-2006
	1119	08-Mar-2006	6635	10-Mar-2006
	1120	08-Mar-2006	6635	10-Mar-2006
	1121	15-Mar-2006	6635	18-Mar-2006
	1122	15-Mar-2006	6635	18-Mar-2006
	1113	15-Mar-2006	6635	18-Mar-2006
	1129	15-Mar-2006	6635	20-Mar-2006
	1127	22-Mar-2006	6635	?
	1125	22-Mar-2006	6635	?

$\sigma_{\text{dataRest}=\text{dataNol}}(\text{Noleggio})$

colloc	dataNol	codCli	dataRest
--------	---------	--------	----------

Algebra Relazionale

- Esistono cinque operazioni di base:
 - Proiezione
 - Selezione
 - Prodotto cartesiano
 - Unione
 - Differenza
- Operazioni derivate:
 - Intersezione
 - join

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

Prodotto cartesiano

- Si applica solo se R e S hanno schemi disgiunti
- La cardinalità del risultato è il prodotto delle cardinalità degli argomenti
- Si applica solo se R e S hanno schemi disgiunti:
 - Se le due relazioni hanno attributi con lo stesso nome, è necessario **ridenominare** gli attributi in una delle due relazioni

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

R X 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

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
- La relazione ottenuta ha lo stesso grado della relazione R ed ha lo stesso contenuto

Ridenominazione - esempio

- La ridenominazione:

$$\rho_{\text{colloc,dataNol,codCli,dataRest} \leftarrow \text{video,dataIn,cliente,dataFine}} (Noleggio)$$

cambia lo schema di Noleggio da
Noleggio(colloc,dataNol,codCli,dataRest)
a
Noleggio(video,dataIn,cliente,dataFine)

Algebra Relazionale

- Esistono cinque operazioni di base:
 - Proiezione
 - Selezione
 - Prodotto cartesiano
 - **Unione**
 - Differenza
- Operazioni derivate:
 - Intersezione
 - join

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 schema
- La relazione risultato ha lo stesso schema delle relazioni argomento
- Le tuple duplicate vengono eliminate
- Il grado della relazione risultato è uguale al grado delle relazioni operandi

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

Unione - esempio

anno	genere
1990	fantastico
1993	animazione
1994	drammatico
1996	fantascienza
1999	horror
2003	fantastico
2005	animazione
2005	fantastico

R

anno	genere
2003	drammatico
1997	fantascienza
1991	commedia

S

anno	genere
1990	fantastico
1993	animazione
1994	drammatico
1996	fantascienza
1999	horror
2003	fantastico
2005	animazione
2005	fantastico
2003	drammatico
1997	fantascienza
1991	commedia

$R \cup S$

Algebra Relazionale

- Esistono cinque operazioni di base:
 - Proiezione
 - Selezione
 - Prodotto cartesiano
 - Unione
 - Differenza
- Operazioni derivate:
 - Intersezione
 - join

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 schema
- La relazione risultato ha lo stesso schema delle relazioni argomento
- Il grado della relazione risultato è uguale al grado delle relazioni operandi

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

Differenza - esempio

anno	genere
1990	fantastico
1993	animazione
1994	drammatico
1996	fantascienza
1999	horror
2003	fantastico
2005	animazione
2005	fantastico

R

anno	genere
2003	drammatico
1997	fantascienza
1991	commedia

S

anno	genere
1990	fantastico
1993	animazione
1994	drammatico
1996	fantascienza
1999	horror
2003	fantastico
2005	animazione
2005	fantastico

$R - S$

Algebra Relazionale

- Esistono cinque operazioni di base:
 - Proiezione
 - Selezione
 - Prodotto cartesiano
 - Unione
 - Differenza
- Operazioni derivate:
 - Intersezione
 - join

Operazioni derivate - 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 relazioni può essere fatta solo se hanno lo stesso schema
- La relazione risultato ha lo stesso schema delle relazioni argomento
- Il grado della relazione risultato è uguale al grado delle relazioni operandi

Operazioni derivate - intersezione

- L'intersezione è un'operazione derivata
- Può infatti essere definita come segue

$$R \cap S = R - (R - S)$$

Operazioni derivate – 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

A	B	C
d	a	f

$R \cap S$

A	B	C
a	b	c
c	b	d

$R - S$

A	B	C
d	a	f

$R - (R - S)$

Operazioni derivate – intersezione - esempio

anno	genere
1990	fantastico
1993	animazione
1994	drammatico
1996	fantascienza
1999	horror
2003	fantastico
2005	animazione
2005	fantastico

R

anno	genere
2003	drammatico
1997	fantascienza
1991	commedia

S

anno	genere
------	--------

$R \cap S$

Operazioni derivate – intersezione - esempio

genere
fantastico
animazione
drammatico
fantascienza
horror

$$GR = \Pi_{\text{genere}}(R)$$

genere
drammatico
fantascienza
commedia

$$GS = \Pi_{\text{genere}}(S)$$

genere
fantastico
animazione
drammatico
fantascienza
horror
commedia

$$GR \cup GS$$

genere
fantastico
animazione
horror

$$GR - GS$$

genere
drammatico
fantascienza

$$GR \cap GS$$

Algebra Relazionale

- Esistono cinque operazioni di base:
 - Proiezione
 - Selezione
 - Prodotto cartesiano
 - Unione
 - Differenza
- Operazioni derivate:
 - Intersezione
 - join

Operazioni derivate - join

Il join (detto anche theta-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**

Operazioni derivate - join

- Come per il prodotto cartesiano, gli schemi delle due relazioni argomento devono essere disgiunti e lo schema della relazione risultato è dato dalla loro unione
- Il grado della relazione risultato è uguale alla somma dei gradi delle relazioni operandi

Operazioni derivate – 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

$R \bowtie_{B < D} S$

Esempio relazioni

Film

titolo	regista	anno	genere	valutaz
underground	emir kusturica	1995	drammatico	3.20
edward mani di forbice	tim burton	1990	fantastico	3.60
nightmare before christmas	tim burton	1993	animazione	4.00
ed wood	tim burton	1994	drammatico	4.00
mars attacks	tim burton	1996	fantascienza	3.00
il mistero di sleepy hollow	tim burton	1999	horror	3.50
big fish	tim burton	2003	fantastico	3.10
la sposa cadavere	tim burton	2005	animazione	3.50
la fabbrica di cioccolato	tim burton	2005	fantastico	4.00
io non ho paura	gabriele salvatores	2003	drammatico	3.50
nirvana	gabriele salvatores	1997	fantascienza	3.00
mediterraneo	gabriele salvatores	1991	commedia	3.80
pulp fiction	quentin tarantino	1994	thriller	3.50
le iene	quentin tarantino	1992	thriller	4.00

Video

colloc	titolo	regista	tipo	Noleggio	colloc	dataNol	codCli	dataRest
1111	underground	emir kusturica	v		1111	01-Mar-2006	6635	02-Mar-2006
1112	underground	emir kusturica	d		1115	01-Mar-2006	6635	02-Mar-2006
1113	big fish	tim burton	v		1117	02-Mar-2006	6635	06-Mar-2006
1114	big fish	tim burton	d		1118	02-Mar-2006	6635	06-Mar-2006
1115	edward mani di forbice	tim burton	d		1111	04-Mar-2006	6642	05-Mar-2006
1116	nightmare before christmas	tim burton	v		1119	08-Mar-2006	6635	10-Mar-2006
1117	nightmare before christmas	tim burton	d		1120	08-Mar-2006	6635	10-Mar-2006
1118	ed wood	tim burton	d		1116	08-Mar-2006	6642	09-Mar-2006
1119	mars attacks	tim burton	d		1118	10-Mar-2006	6642	11-Mar-2006
1120	il mistero di sleepy hollow	tim burton	d		1121	15-Mar-2006	6635	18-Mar-2006
1121	la sposa cadavere	tim burton	d		1122	15-Mar-2006	6635	18-Mar-2006
1122	la fabbrica di cioccolato	tim burton	d		1113	15-Mar-2006	6635	18-Mar-2006
1123	la fabbrica di cioccolato	tim burton	d		1129	15-Mar-2006	6635	20-Mar-2006
1124	io non ho paura	gabriele salvatores	d		1119	15-Mar-2006	6642	16-Mar-2006
1125	nirvana	gabriele salvatores	d		1126	15-Mar-2006	6610	16-Mar-2006
1126	mediterraneo	gabriele salvatores	d		1112	16-Mar-2006	6610	18-Mar-2006
1127	pulp fiction	quentin tarantino	v		1114	16-Mar-2006	6610	17-Mar-2006
1128	pulp fiction	quentin tarantino	d		1128	18-Mar-2006	6642	20-Mar-2006
1129	le iene	quentin tarantino	d		1124	20-Mar-2006	6610	21-Mar-2006
					1115	20-Mar-2006	6610	21-Mar-2006
					1124	21-Mar-2006	6642	22-Mar-2006
					1116	21-Mar-2006	6610	?
					1117	21-Mar-2006	6610	?
					1127	22-Mar-2006	6635	?
							5	?
							2	?
							2	?

Cliente

codCli	nome	cognome	telefono	dataN	residenza
6610	anna	rossi	01055664433	05-Ott-1979	via scribanti 16 16131 genova
6635	paola	bianchi	0104647992	12-Apr-1976	via dodecaneso 35 16146 genova
6642	marco	verdi	3336745383	16-Ott-1972	via lagustena 35 16131 genova

Operazioni derivate – join - esempio

- Determinare il titolo ed il regista dei film noleggiati il 15 Marzo 2006 dal cliente di codice 6635

Operazioni derivate – join - esempio

- Determinare il titolo ed il regista dei film noleggiati il 15 Marzo 2006 dal cliente di codice 6635

$$\Pi_{\text{titolo, regista}} \left(\sigma_{\text{codCli}=6635 \wedge \text{dataNol}='15\text{--Mar--}2006' (\text{Noleggio}) \bowtie_{\text{colloc}=c} \rho_{\text{colloc} \leftarrow c}(\text{Video}) \right)$$

titolo	regista
la sposa cadavere	tim burton
la fabbrica di cioccolato	tim burton
big fish	tim burton
le iene	quentin tarantino