

Join 2

Outer Join - Join Esterni

L'effetto collaterale, talvolta dannoso, del join è il **tralasciare informazioni rilevanti**.

L'**outer join** prevede che tutte le tuple diano un contributo al risultato, eventualmente **ponendo valori null**, laddove fosse necessario.

Left outer Join

Estende solo le tuple del primo operando

Right outer Join

Estende solo le tuple del secondo operando

Full outer Join

Estende solo le tuple di entrambi gli operandi

...Outer Join...

r_2

Film	Nome Regista	Cognome Regista
La strada	Federico	Fellini
2001 Space Odissey	Stanley	Kubrick

r_1

Film	Nome Interprete	Cognome Interprete
La strada	Giulietta	Masina
Taxi Driver	Robert	De Niro
La strada	Anthony	Quinn

...Outer Join...

$r_1 \bowtie_{\text{LEFT}} r_2$

Film	Nome Interprete	Cognome Interprete	Nome Regista	Cognome Regista
La strada	Giulietta	Masina	Federico	Fellini
Taxi Driver	Robert	De Niro	Null	Null
La strada	Anthony	Quinn	Federico	Fellini

$r_1 \bowtie_{\text{RIGHT}} r_2$

Film	Nome Interprete	Cognome Interprete	Nome Regista	Cognome Regista
La strada	Giulietta	Masina	Federico	Fellini
2001 Space Odissey	Null	Null	Stanley	Kubrick
La strada	Anthony	Quinn	Federico	Fellini

...Outer Join

$$r_1 \bowtie_{\text{FULL}} r_2$$

Film	Nome Interprete	Cognome Interprete	Nome Regista	Cognome Regista
La strada	Giulietta	Masina	Federico	Fellini
2001 Space Odyssey	Null	Null	Stanley	Kubrik
La strada	Anthony	Quinn	Federico	Fellini
Taxi Driver	Robert	De Niro	Null	Null

Warning #1: "Null" non significa che non ci sono interpreti per quel film, ma che io non possiedo l'informazione nel mio db riguardo gli interpreti!!!

Warning #2: Il join fa proliferare i valori "Null". E' proprio necessario?

Null

L'informazione è fatta di dati!

Film	Nome_reg	Cognome_reg	Nome_Inter	Cognome_Inter
Taxi Driver	Null	Null	Robert	De Niro
Null	Null	Null	Robert	De Niro
Null	Sergio	Leone	Robert	De Niro
Null	Null	Fellini	Null	Null

Join n-ario

Proprietà del **join naturale**:

- è **commutativo** $r_1 \bowtie r_2 = r_2 \bowtie r_1$
 - è **associativo** $r_1 \bowtie (r_2 \bowtie r_3) = (r_1 \bowtie r_2) \bowtie r_3$
- $\Rightarrow r_1 \bowtie r_2 \bowtie r_3 \bowtie \dots \bowtie r_n$ oppure $\bowtie_{i=1}^n r_i$

Siano X_1 ed X_2 gli insiemi su cui sono definiti gli operandi di join,

Se $X_1 = X_2 \Rightarrow r_1(X_1) \bowtie r_2(X_2) = r_1(X_1) \cap r_2(X_2)$

Se $X_1 \cap X_2 = \emptyset \Rightarrow$

- il risultato è ancora definito sull'unione di X_1 e X_2 , ma non essendoci attributi in comune, la **condizione per la partecipazione al join è sempre verificata**.

\Rightarrow il join diventa un **prodotto cartesiano**, che contiene tutte le tuple dei due operandi di partenza

Prodotto Cartesiano...

Il **prodotto cartesiano** è un operatore definito su relazioni che non hanno attributi in comune. La cardinalità del risultato è pari al prodotto delle cardinalità degli operandi.

Warning: l'uso del termine è improprio, dal momento che il prodotto cartesiano è un insieme di coppie ordinate, mentre in questo caso si ottiene un insieme di tuple

PUNTATE

Sequel	Puntata
Il commissario Montalbano	Puntata 1
Il commissario Montalbano	Puntata 2

CASTING

CF	Nome Interprete	Cognome Interprete
1234	Luca	Zingaretti
2345	Katharina	Bohm
3456	Cesare	Bocci



...Prodotto Cartesiano

EPISODI ⋈ CASTING

Sequel	Puntata	CF	Nome Interprete	Cognome Interprete
Il commissario Montalbano	Puntata 1	1234	Luca	Zingaretti
Il commissario Montalbano	Puntata 1	2345	Katharina	Bohm
Il commissario Montalbano	Puntata 1	3456	Cesare	Bocci
Il commissario Montalbano	Puntata 2	1234	Luca	Zingaretti
Il commissario Montalbano	Puntata 2	2345	Katharina	Bohm
Il commissario Montalbano	Puntata 2	3456	Cesare	Bocci



Theta-join...

Il theta join è un prodotto cartesiano seguito da una selezione.

- Sia F una **formula proposizionale** utilizzabile in una selezione,
- le relazioni r_1 ed r_2 non hanno attributi in comune,

$$r_1 \bowtie_F r_2 = \sigma_F (r_1 \bowtie r_2)$$

⊙ Puntata = Puntata 1 (*CASTING ⋈ EPISODI*)

Sequel	Puntata	CF	Nome Interprete	Cognome Interprete
Il commissario Montalbano	Puntata 1	1234	Luca	Zingaretti
Il commissario Montalbano	Puntata 1	2345	Katharina	Bohm
Il commissario Montalbano	Puntata 1	3456	Cesare	Bocci

...Theta-join

Un theta join in cui la condizione di selezione F sia una **congiunzione di atomi di uguaglianza**, con un attributo della prima relazione e uno della seconda, è chiamato **equi-join**.

L'equi join ed il theta join hanno una notevole rilevanza, in quanto la maggior parte dei sistemi di basi di dati li utilizzano per correlare le informazioni.

Interrogazioni...

- Dato uno **schema R di basi di dati**, un' **interrogazione** è una funzione che per ogni istanza r di R produce una relazione su un dato insieme di attributi X .
- Data un'espressione **E** di un linguaggio di interrogazione, l'applicazione di E ad una istanza r , rappresenta una interrogazione **$E(r)$** .
- Si consideri lo schema seguente:
 - FILM (Titolo, Cognome Regista, Nome Regista, Anno)
 - INTERPRETE (CF, Film, Cognome Interprete, Nome Interprete)
- Interrogazione: *trovare il titolo dei film, il cognome e nome del regista, il cui film è stato prodotto dopo il 1960.*

π Titolo, Cognome Regista, Nome Regista (σ Anno > 1960 (FILM))



...Interrogazioni...

FILM

Titolo	Nome Regista	Cognome Regista	Anno
La strada	Federico	Fellini	1954
2001 Space Odissey	Stanley	Kubrick	1968
Raging Bull	Martin	Scorsese	1980



Titolo	Nome Regista	Cognome Regista	Anno
2001 Space Odissey	Stanley	Kubrick	1968
Raging Bull	Martin	Scorsese	1980



...Interrogazioni

Trovare il cognome ed il nome degli interpreti dei film prodotti prima del 1960

π *Cognome Interprete, Nome Interprete* (INTERPRETI \bowtie Titolo=Film (σ Anno < 1960 (FILM)))

INTERPRETE

Film	Nome Interprete	Cognome Interprete	CF
La strada	Giulietta	Masina	1234
Taxi Driver	Robert	De Niro	2345
La strada	Anthony	Quinn	3456



Nome Interprete	Cognome Interprete
Giulietta	Masina
Anthony	Quinn



Equivalenza espressioni algebriche...

$$E_1 \equiv_R E_2 \Leftrightarrow E_1(r) = E_2(r) \forall r \in R$$

$$E_1 \equiv E_2 \Leftrightarrow E_1 \equiv_R E_2 \forall \text{ schema } R$$

L'equivalenza delle espressioni algebriche è fondamentale per calcolare il **costo** di un'interrogazione.

Bisogna ricondursi a trasformazioni di equivalenza che:

- riducono le dimensioni dei risultati intermedi
- preparano l'espressione all'applicazione di una trasformazione che riduce i risultati intermedi.



...Equivalenza espressioni algebriche...

Atomizzazioni di selezioni: Una selezione congiuntiva può essere sostituita da una cascata di selezioni atomiche:

$$\sigma_{F_1 \wedge F_2}(E) \equiv \sigma_{F_1}(\sigma_{F_2}(E))$$

Idempotenza delle proiezioni: una proiezione può essere trasformata in una cascata di proiezioni che eliminano gli attributi in fasi diverse.

$$\pi_X(E) \equiv \pi_X(\pi_{XY}(E))$$

Anticipazione della selezione rispetto al join

$$\sigma_{F_1}(E_1 \bowtie E_2) \equiv E_1 \bowtie \sigma_{F_1}(E_2)$$

...Equivalenza espressioni algebriche...

Anticipazione della proiezione rispetto al join

- Siano E_1 ed E_2 definite su X_1 ed X_2
 - $Y_2 \subseteq X_2$ e $Y_2 \supseteq (X_1 \cap X_2)$ [cioè $X_2 - Y_2$ non sono coinvolti nel join]
- $$\Rightarrow \pi_{X_1 Y_2} (E_1 \bowtie E_2) \equiv E_1 \bowtie \pi_{Y_2} (E_2)$$

Inglobamento di una selezione in un prodotto cartesiano a formare un join

$$\sigma_F (E_1 \bowtie E_2) \equiv E_1 \bowtie_F (E_2)$$

....Equivalenza espressioni algebriche...

Distributività della selezione rispetto all'unione:

$$\sigma_F (E_1 \cup E_2) \equiv \sigma_F (E_1) \cup \sigma_F (E_2)$$

Distributività della selezione rispetto alla differenza:

$$\sigma_F (E_1 - E_2) \equiv \sigma_F (E_1) - \sigma_F (E_2)$$

Distributività della proiezione rispetto all'unione:

$$\pi_X (E_1 \cup E_2) \equiv \pi_X (E_1) \cup \pi_X (E_2)$$

....Equivalenza espressioni algebriche

$$\sigma_{F_1 \vee F_2}(R) \equiv \sigma_{F_1}(R) \cup \sigma_{F_2}(R)$$

$$\sigma_{F_1 \wedge F_2}(R) \equiv \sigma_{F_1}(R) \cap \sigma_{F_2}(R) \equiv \sigma_{F_1}(R) \bowtie \sigma_{F_2}(R)$$

$$\sigma_{F_1 \wedge \neg F_2}(R) \equiv \sigma_{F_1}(R) - \sigma_{F_2}(R)$$

$$E \bowtie (E_1 \cup E_2) \equiv (E \bowtie E_1) \cup (E \bowtie E_2)$$

Algebra con valori Nulli...

Titolo	Nome Regista	Cognome Regista	Anno
La strada	Federico	Fellini	1954
2001 Space Odyssey	Stanley	Kubrick	1968
Raging Bull	Martin	Scorsese	NULL



Il valore Unknown

Not	
T	F
F	T
U	U

And	T	U	F
T	T	U	F
U	U	U	F
F	F	F	F

Or	T	U	F
T	T	T	T
U	T	U	U
F	T	U	F

Il paradosso: $\sigma_{\text{Anno} \geq 1960}(\text{FILM}) \cup \sigma_{\text{Anno} < 1960}(\text{FILM})$ non Restituisce la Relazione di partenza, perché la terza tupla non verifica né la prima, né la seconda condizione!!!!

Si introducono due nuove forme di condizione atomica:

- A IS NULL, vero su una tupla t se t[A] è Null
- A IS NOT NULL, vero se su una tupla t t[A] è specificato

ora $\sigma_{\text{Anno} \geq 1960 \vee \text{Anno} < 1960 \vee \text{Anno IS NULL}}(\text{FILM})$
include anche la terza tupla

Corrado Aaron Visaggio

21

Viste

Una vista è una relazione derivata da altre relazioni.

Esistono due tipi di relazioni derivate:

- **viste materializzate**: relazioni derivate effettivamente memorizzate nella base di dati.
 - Calcolo della vista complesso
 - necessità di riferimento frequente ed immediato
 - Ma non sono supportate!!!!
- **Relazioni virtuali**: relazioni definite per mezzo di funzioni, non memorizzate nella base di dati, ma utilizzabili come se lo fossero.
 - Evitare componenti non rilevanti.
 - Ridurre la complessità delle espressioni da gestire.
 - Introduzione di meccanismi specifici di riservatezza
 - Agevolare la ristrutturazione delle basi di dati.

Corrado Aaron Visaggio

22



Any question?

