

Join (naturale)

E' l'operatore più caratteristico dell'algebra relazionale, che evidenzia la proprietà del modello relazionale di essere basato su valori. Non ha un corrispettivo nella teoria degli insiemi

L'operatore di join (naturale) correla dati contenuti in relazioni diverse

*Nella situazione più comune, produce una **relazione definita sull'unione degli schemi degli operandi**, le cui tuple sono ottenute combinando le tuple degli operandi **che hanno valori uguali** su attributi comuni (cioè che hanno lo stesso nome nei due schemi)*

Join

Il join naturale $r_1 \bowtie r_2$ di $r_1(X_1)$ e $r_2(X_2)$ è una relazione definita su $X_1 \cup X_2$ (che si può scrivere $X_1 X_2$)

$$r_1 \bowtie r_2 = \{ t \text{ su } X_1 X_2 \mid t[X_1] \in r_1 \text{ e } t[X_2] \in r_2 \}$$

Il grado di tale relazione (numero attributi) è minore o uguale alla somma dei gradi delle due relazioni poiché gli attributi omonimi compaiono una sola volta

Per questo motivo la definizione implica, normalmente, che una tupla del risultato sia ottenuta concatenando ogni tupla della prima relazione con ogni tupla della seconda in cui gli attributi comuni hanno lo stesso valore

Join

Due casi limite:

- **join vuoto**, le due tabelle hanno alcuni attributi in comune ma nessuna tupla di un operando è combinabile con nessuna tupla dell'altro
- **prodotto cartesiano** degli operandi: ciascuna delle tuple di un operando è combinabile con tutte le tuple dell'altro. Può verificarsi se le due relazioni non hanno attributi in comune

Join

- **Se $X_1 \cap X_2$ è vuoto** il join naturale equivale al prodotto cartesiano fra le relazioni.

Dimostrazione

Le due relazioni non hanno attributi comuni => ogni tupla della prima relazione può «concatenarsi» con ogni tupla della seconda, soddisfacendo la definizione di join

- **Se le due relazioni hanno lo stesso schema ($X_1 = X_2$)** il join naturale equivale all'intersezione fra le relazioni.

Dimostrazione

Ogni tupla può «concatenarsi» solo con una tupla uguale => è presente nel risultato finale se e solo se è presente in entrambe le relazioni **NB La concatenazione genera la tupla stessa**

Join

Se ciascuna tupla di ciascun operando contribuisce ad almeno una tupla del risultato il join si dice ***completo***

Se per alcune tuple non è verificata la corrispondenza e quindi non contribuiscono al risultato, le tuple si dicono ***dangling***

Impiegato	Reparto
Rossi	A
Neri	B
Bianchi	B

Reparto	Capo
A	Mori
B	Bruni

Impiegato	Reparto	Capo
Rossi	A	Mori
Neri	B	Bruni
Bianchi	B	Bruni

- ogni tupla contribuisce al risultato: il join è **completo**

Un Join non completo

Impiegato	Reparto
Rossi	A
Neri	B
Bianchi	B

Reparto	Capo
B	Mori
C	Bruni

Impiegato	Reparto	Capo
Neri	B	Mori
Bianchi	B	Mori

Un Join vuoto

Impiegato	Reparto
Rossi	A
Neri	B
Bianchi	B

Reparto	Capo
D	Mori
C	Bruni

Impiegato	Reparto	Capo
-----------	---------	------

Join: proprietà

- $r_1 \diamond r_2$ contiene un numero di tuple compreso fra zero (*risultato = insieme vuoto*) e il prodotto di $|r_1|$ e $|r_2|$ (*risultato = prodotto cartesiano*)
- Se $r_1 \diamond r_2$ è **completo** allora contiene un numero di tuple pari almeno al massimo fra $|r_1|$ e $|r_2|$ (*nessuna tupla resta esclusa, quindi tutte le tuple della relazione di cardinalità maggiore partecipano sicuramente al join*)
- Se $X_1 \cap X_2$ contiene una chiave per r_2 , allora $r_1(X_1) \diamond r_2(X_2)$ contiene al più $|r_1|$ tuple (*ogni tupla di r_1 si associa al più a una tupla di r_2*).

Join: proprietà

- se il join coinvolge una chiave di r_2 e un vincolo di integrità referenziale da r_1 a r_2 , allora il numero di tuple è pari a $|r_1|$

(ogni tupla di r_1 si associa ad una e una sola tupla di r_2)

- $r_1 \diamond r_2 = r_2 \diamond r_1$ il join è *commutativo*

- $(r_1 \diamond r_2) \diamond r_3 = r_1 \diamond (r_2 \diamond r_3)$ il join è *associativo*

Quindi sequenze di join possono essere scritte senza parentesi

Join

Per correlare attributi con nome diverso (se cioè $X_1 \cap X_2$ è vuoto) è possibile fare il *theta-join*, definito come un prodotto cartesiano seguito da una selezione

$$r_1 \diamond_F r_2 = \sigma_F (r_1 \diamond r_2)$$

dove F è la condizione di selezione.

Se F è una condizione di uguaglianza fra un attributo della prima relazione e uno della seconda, allora siamo in presenza di un *equi-join*. E' il tipo di join più comune.

Sono importanti formalmente:

- il join naturale è basato sui *nomi* degli attributi
- equi-join e theta-join sono basati sui *valori*

Join

Il theta-join, espresso come prodotto cartesiano seguito da una selezione, è il tipo di join operativamente più generale. Infatti:

- Consente di associare (concatenare) tuple attraverso la corrispondenza fra attributi arbitrari e non necessariamente omonimi
- Consente di basare l'associazione su una arbitraria condizione e non solo su una condizione di uguaglianza

Impiegati

Impiegato	Reparto
Rossi	A
Neri	B
Bianchi	B

Reparti

Codice	Capo
A	Mori
B	Bruni

Impiegati JOIN_{Reparto=Codice} Reparti

Impiegato	Reparto	Codice	Capo
Rossi	A	A	Mori
Neri	B	B	Bruni
Bianchi	B	B	Bruni