

BASI DI DATI I

- Tutte le operazioni di Join (prima parte)
- Prima parte
- Seconda parte

OPERAZIONI DI JOIN (1° PARTE)



JOIN

- La sequenza di operazioni appena vista, riassumibile in $\sigma_{join-condition} (R \times S)$ è abbastanza frequente: per questo è stata creata un'operazione speciale, chiamata JOIN.
- La JOIN, denotata con \bowtie , è usata per combinare tuple relate in una sola tupla.
 - $\sigma_{join-condition} (R \times S)$ diventa $R \bowtie_{join-condition} S$



JOIN: ESEMPIO

- Supponiamo di voler trovare il nome del manager di ciascun dipartimento:
 - Combiniamo ogni tupla dipartimento con la tupla impiegato il cui SSN fa match col valore MGRSSN nella tupla dipartimento.
 - DEPT_MGR= department $\bowtie_{MGRSSN=SSN}$ EMPLOYEE
 - RESULT= $\pi_{DNAME, LNAME, FNAME}(\text{DEPT_MGR})$

DEPT_MGR	DNAME	DNUMBER	MGRSSN	• • •	FNAME	MINIT	LNAME	SSN	• • •
	Research	5	333445555	• • •	Franklin	T	Wong	333445555	• • •
	Administration	4	987654321	♦ ♦ ♦	Jennifer	S	Wallace	987654321	• • •
	Headquarters	1	888665555	♦ ♦ ♦	James	E	Borg	888665555	• • •



JOIN: ESEMPIO SU PRODOTTO CARTESIANO

- L'esempio precedente sul prodotto cartesiano può essere risolto rimpiazzando le operazioni:

- $\text{EMP_DEPENDENTS} = \text{EMPNAME} \times \text{DEPENDENTS}$
- $\text{ACTUAL_DEPENDENTS} = \sigma_{\text{SSN}=\text{ESSN}}(\text{EMP_DEPENDENTS})$

Con:

- $\text{ACTUAL_DEPENDENTS} = \text{EMPNAME} \bowtie_{\text{SSN}=\text{ESSN}} \text{DEPENDENTS}$



OPERAZIONI DI JOIN (2° PARTE)



NOZIONI SULLA JOIN

- Sintassi :

Date le relazioni $R(A_1, A_2, \dots, A_n)$ e $S(B_1, B_2, \dots, B_m)$

Scriveremo $R \bowtie_{\text{join_condition}} S$

- Il risultato della join ha $n+m$ attributi:

$Q(A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_m)$

- La condizione della join è della forma:

$\langle \text{condition} \rangle \text{ AND } \langle \text{condition} \rangle \text{ AND } \dots \text{ AND } \langle \text{condition} \rangle$
dove ogni condizione è della forma $A_i \theta B_j$,

con $A_i \in R, B_j \in S, \text{dom}(A_i) = \text{dom}(B_j)$ e $\theta \in \{=,,\neq,<,>,\leq,\geq\}$

- Tuple con attributi di join **null** non appaiono nel risultato.



TIPI DI JOIN

- Una join con θ generica è detta **THETA JOIN**
- Il join più comune ha come θ l'operatore di uguaglianza ed è detto **EQUIJOIN**:
 - Nei risultati dell'EQUIJOIN abbiamo sempre una coppia di valori di attributi identici.
- Poiché tale ripetizione è superflua, esiste una nuova operazione, chiamata **NATURAL JOIN** e denotata da $*$, che elimina il secondo attributo superfluo.
 - È in sostanza una EQUIJOIN con la rimozione degli attributi con valori uguali superflui



Impiegati

Impiegato	Reparto
Rossi	A
Neri	B
Bianchi	B

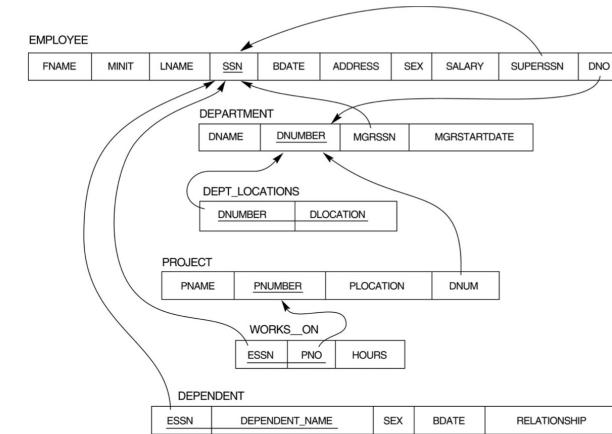
Reparti

Codice	Capo
A	Mori
B	Bruni

Impiegati $\triangleright\triangleleft_{\text{Reparto}=\text{Codice}}$ Reparti

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

NATURAL JOIN

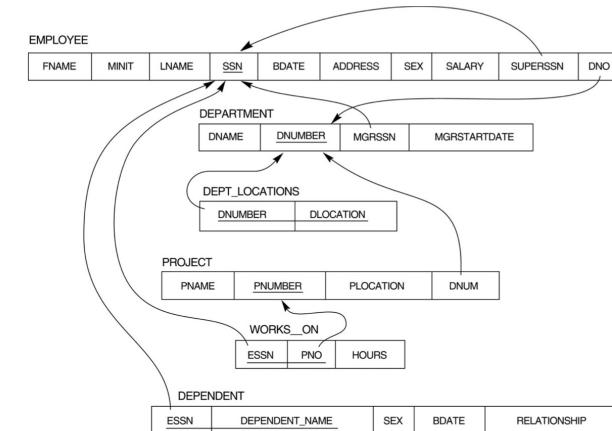


- La definizione standard di Natural join richiede che i due attributi di join (per ogni coppia) abbiano lo stesso nome. Se ciò non vale, occorre prima un'operazione di renaming.
 - $\text{DEPT} = \pi_{(\text{DNAME}, \text{DNUM}, \text{MGRSSN}, \text{MGRSTARTDATE})} (\text{DEPARTMENT})$
 - $\text{PROJ_DEPT} = \text{PROJECT} * \text{DEPT}$
 - L'attributo DNUM è detto “**attributo di join**”

PROJ_DEPT	PNAME	PNUMBER	PLOCATION	DNUM	DNAME	MGRSSN	MGRSTARTDATE
ProductX	1	Bellaire	5	Research	333445555	1988-05-22	
ProductY	2	Sugarland	5	Research	333445555	1988-05-22	
ProductZ	3	Houston	5	Research	333445555	1988-05-22	
Computerization	10	Stafford	4	Administration	987654321	1995-01-01	
Reorganization	20	Houston	1	Headquarters	888665555	1981-06-19	
Newbenefits	30	Stafford	4	Administration	987654321	1995-01-01	



NATURAL JOIN (2)



- Qualora gli attributi di join abbiano lo stesso nome, non è necessario il renaming.
- **Esempio:**
 - DEPT_LOCS = DEPARTMENT * DEPT_LOCATIONS

DEPT_LOCS	DNAME	DNUMBER	MGRSSN	MGRSTARTDATE	LOCATION
	Headquarters	1	888665555	1981-06-19	Houston
	Administration	4	987654321	1995-01-01	Stafford
	Research	5	333445555	1988-05-22	Bellaire
	Research	5	333445555	1988-05-22	Sugarland
	Research	5	333445555	1988-05-22	Houston

- In generale il Natural Join verifica l'uguaglianza di tutte le coppie di attributi.



NATURAL JOIN (3)

Definizione più generale:

- $Q = R^*(\langle \text{list}_1 \rangle)(\langle \text{list}_2 \rangle)S$
 - $\langle \text{list}_1 \rangle$ =lista di attributi da R
 - $\langle \text{list}_2 \rangle$ =lista di attributi da S
- Le liste sono usate per le uguaglianze tra coppie di attributi corrispondenti;
 - le condizioni sono combinate in AND.
 - Solo la lista corrispondente ad attributi della prima relazione, $\langle \text{list}_1 \rangle$ è mantenuta nel risultato.



NATURAL JOIN (4)

- Se nessuna combinazione di tuple soddisfa la condizione di join, la relazione risultante avrà zero tuple:
 - Date le relazioni R (con n_r tuple) ed S (con n_s tuple),
 - $n_q Q = R \bowtie_{\text{joincond}} S$ avrà da 0 a $n_r \cdot n_s$ tuple.
 - $n_q / (n_r \cdot n_s)$ è detta **join selectivity**.
- Se non esiste alcuna **<joincondition>**, tutte le tuple compariranno nella relazione risultante, ed il join diventa un prodotto cartesiano, detto **CROSS JOIN**.



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

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

- Ogni n -pla contribuisce al risultato:
 - join **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	Reparto	Capo
Rossi	A	D	Mori
Neri	B	C	Bruni
Bianchi	B		

Impiegato Reparto Capo



UN JOIN COMPLETO, CON N · M MAPPE

Impiegato	Reparto
Rossi	B
Neri	B

Reparto	Capo
B	Mori
B	Bruni

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



CARDINALITÀ DEL JOIN

- Il join di R_1 e R_2 contiene un numero di n -ple compreso fra zero e il prodotto di $|R_1|$ e $|R_2|$.
- Se il join coinvolge una chiave di R_2 , allora il numero di n -ple è compreso fra zero e $|R_1|$.
- Se il join coinvolge una chiave di R_2 e un vincolo di integrità referenziale, allora il numero di n -ple è pari a $|R_1|$.



JOIN, UNA DIFFICOLTÀ

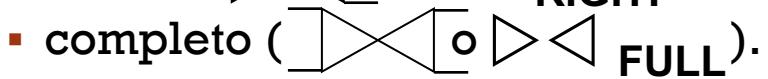
Impiegato	Reparto	Reparto	Capo
Rossi	A	B	Mori
Neri	B	C	Bruni
Bianchi	B		

Impiegato	Reparto	Capo
Neri	B	Mori
Bianchi	B	Mori

- Alcune n -ple non contribuiscono al risultato: vengono “tagliate fuori”.



JOIN ESTERNO

- Il join **esterno** estende, con valori nulli, le n -ple che verrebbero tagliate fuori da un join (**interno**).
- Esiste in tre versioni:
 - sinistro ( o  ), **LEFT**,
 - destro ( o  ), **RIGHT**,
 - completo ( o  ), **FULL**.



JOIN ESTERNO (2)

- **Sinistro**: mantiene tutte le n -ple del *primo operando*, estendendole con valori nulli, se necessario.
- **Destro**: ... *del secondo operando* ...
- **Completo**: ... *di entrambi gli operandi* ...



Impiegati

Impiegato	Reparto
Rossi	A
Neri	B
Bianchi	B

Reparti

Reparto	Capo
B	Mori
C	Bruni

Impiegati \bowtie_{LEFT} Reparti

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



Impiegati

Impiegato	Reparto
Rossi	A
Neri	B
Bianchi	B

Reparti

Reparto	Capo
B	Mori
C	Bruni

Impiegati \bowtie_{RIGHT} Reparti

Impiegato	Reparto	Capo
Neri	B	Mori
Bianchi	B	Mori
NULL	C	Bruni



Impiegati

Impiegato	Reparto
Rossi	A
Neri	B
Bianchi	B

Reparti

Reparto	Capo
B	Mori
C	Bruni

Impiegati \bowtie_{FULL} Reparti

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



L'OPERAZIONE DI DIVISIONE



L'OPERAZIONE DI DIVISIONE

- L'operazione di divisione è indicata con \div è un tipo particolare di interrogazione che talvolta si presenta nelle applicazioni di basi di dati.
- Un esempio è “*trova i nomi degli impiegati che lavorano a tutti i progetti su cui lavora ‘John Smith’*”.
- La divisione \div , di $r1$ per $r2$, con $r1$ su $R_1(X_1X_2)$ e $r2$ su $R_2(X_2)$, è (il più grande) insieme di tuple con schema X_1 tale che, facendo il prodotto cartesiano con $r2$, ciò che si ottiene è una relazione contenuta in $r1$.



ESEMPIO

Voli

Codice	Data
AZ427	21/07/2001
AZ427	23/07/2001
AZ427	24/07/2001
TW056	21/07/2001
TW056	24/07/2001
TW056	25/07/2001

Linee

Codice
AZ427
TW056

Voli ÷ Linee

Data
21/07/2001
24/07/2001

(Voli ÷ Linee) ▷◁ Linee

Codice	Data
AZ427	21/07/2001
AZ427	24/07/2001
TW056	21/07/2001
TW056	24/07/2001

La divisione trova le date con voli per tutte le linee



UN INSIEME COMPLETO DI OPERAZIONI

- Si può provare che $\{\sigma, \pi, \cup, -, \times\}$ è un **insieme completo**, cioè tutte le altre operazioni possono essere espresse come combinazioni di queste.

- **Esempi:**

- Intersezione:

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

- Join:

- $R \bowtie_{<\text{condition}>} S = \sigma_{<\text{condition}>} (R \times S)$

- Divisione:

- $R \div S = T1 \leftarrow \pi_{<\text{attribute of } R - \text{attribute of } S>} (R)$

- $T2 \leftarrow \pi_{<\text{attribute of } R - \text{attribute of } S>} ((S \times T1) - R)$

- $T \leftarrow T1 - T2$



ALGEBRA RELAZIONALE: LIMITI

- Ci sono però interrogazioni interessanti non esprimibili:
 - Calcolo di valori derivati: possiamo solo **estrarre** valori, non calcolarne di nuovi; calcoli di interesse:
 - a livello di n -pla o di singolo valore (conversioni somme, differenze, etc.);
 - su insiemi di n -ple (somme, medie, etc.).
 - Interrogazioni inerentemente **ricorsive**, come la **chiusura transitiva**.



FUNZIONI AGGREGATE E DI RAGGRUPPAMENTO

- *Non sono presenti nell'algebra relazionale base.*
- Operano su un insieme di dati per restituire come risultato una relazione con un solo valore.
- Sono funzioni applicate a collezioni di valori numerici:
 - SUM, AVERAGE, MAXIMUM, MINIMUM, COUNT.
- *Notazione:*

$\langle \text{attributi di raggruppamento} \rangle \mathcal{F}_{\langle \text{lista funzioni} \rangle}(R)$

- dove $\langle \text{attributi di raggruppamento} \rangle$ è una lista di attributi della relazione R e raggruppa le tuple presenti nella relazione sulla base dei loro valori,
- e $\langle \text{lista funzioni} \rangle$ è una lista di coppie $\langle \text{funzione} \rangle \langle \text{attributo} \rangle$:
 - $\langle \text{funzione} \rangle$ è una delle funzioni SUM,...
 - $\langle \text{attributo} \rangle$ è un attributo della relazione R.



ESEMPIO

- $T(N_D, N_IMP, STIP_MEDIO) \leftarrow_{N_D} \mathcal{F}_{COUNT\ SSN, AVERAGE\ STIPENDIO}(IMPIEGATO)$
- $R \leftarrow \sigma_{N_D, N_IMP, STIP_MEDIO}(T)$

R	N_D	N_IMP	STIP_MEDIO
5	4	33250	
4	3	31000	
1	1	55000	

- $S \leftarrow \mathcal{F}_{co}$

S	COUNT_SSN	AVERAGE_STIPENDIO
8	35125	



EQUIVALENZA DI ESPRESSIONI

- Due espressioni sono **equivalenti** se producono lo stesso risultato qualunque sia l'istanza attuale della base di dati.
- L'equivalenza è importante perché i DBMS cercano di eseguire espressioni equivalenti a quelle date, ma meno “costose”.
- Trasformazioni di equivalenza.





FINE

Per eventuali domande: (in ordine di preferenza personale)

- Ora.
- Chat di Teams
- Mail: silvio.barra@unina.it

