

# Selezione con valori nulli

## Impiegati

Matricola	Cognome	Filiale	Stipendio
7309	Rossi	Roma	32
5998	Neri	Milano	45
9553	Bruni	Milano	NULL

**SEL** `Stipendio > 40` (Impiegati)

- la condizione atomica è vera solo per valori non nulli
- vediamo su Relax (relazione ImpiegatiTer)

<http://dbis-uibk.github.io/relax/calc/gist/1a9dc6cd0f3478388fc177dfc9b5a314>

# Un risultato non desiderabile

$$\text{SEL}_{\text{Stipendio}>40}(\text{Persone}) \cup \text{SEL}_{\text{Stipendio}\leq 40}(\text{Persone}) \neq \text{Persone}$$

- Vediamo su Relax.
- Perché?
  - Perché le selezioni vengono valutate separatamente!
- Ma anche

$$\text{SEL}_{\text{Stipendio}>40 \vee \text{Stipendio}\leq 40}(\text{Persone}) \neq \text{Persone}$$

- Perché?
  - Perché anche le condizioni atomiche vengono valutate separatamente!

# Selezione con valori nulli: soluzione

`SEL Stipendio > 40 (Impiegati)`

- la condizione atomica è vera solo per valori non nulli
- per riferirsi ai valori nulli esistono forme apposite di condizioni:

`IS NULL`

`IS NOT NULL`

- si potrebbe usare (ma non serve) una "logica a tre valori" (vero, falso, sconosciuto)

# Nota bene

- Relax usa

... = null

- Quindi:

$$\text{SEL}_{\text{Stipendio} > 40} (\text{Persone}) \cup \text{SEL}_{\text{Stipendio} \leq 40} (\text{Persone}) \cup \text{SEL}_{\text{Stipendio IS NULL}} (\text{Persone})$$
$$=$$
$$\text{SEL}_{\text{Stipendio} > 40 \vee \text{Stipendio} \leq 40 \vee \text{Stipendio IS NULL}} (\text{Persone})$$
$$=$$
$$\text{Persone}$$

## Impiegati

Matricola	Cognome	Filiale	Stipendio
5998	Neri	Milano	45
9553	Bruni	Milano	NULL

**SEL** (Stipendio > 40) OR (Stipendio IS NULL) (Impiegati)

# Selezione e proiezione

- Combinando selezione e proiezione, possiamo estrarre interessanti informazioni da una relazione

- matricola e cognome degli impiegati che guadagnano più di 50

Matricola	Cognome
7309	Rossi
5998	Neri
5698	Neri

**PROJ<sub>Matricola,Cognome</sub> (SEL<sub>Stipendio > 50</sub> (Impiegati))**

- Combinando selezione e proiezione, possiamo **estrarre** informazioni da **una** relazione
- Invece non possiamo
  - calcolare informazioni derivate
  - correlare informazioni presenti in relazioni diverse, né informazioni in ennuple diverse di una stessa relazione
- Vediamo altri operatori che permettono queste attività

# Aggregazione

- Il contenuto delle basi di dati viene spesso aggregato:
  - Il voto massimo nell'esame di basi di dati
  - Il voto medio per ciascun esame
  - Il numero di CFU conseguiti da ciascuno studente

# Aggregazione, sintassi e semantica

## • Sintassi

**AGG** *AttributiRaggruppamento; Funzione (Attributo), ... (Operando)*

- *Funzione*: count, sum, avg, max, min
- Semantica
  - Il risultato contiene la proiezione sugli attributi di raggruppamento e il valore delle funzioni in corrispondenza al sottoinsieme raggruppato
- simbolo più compatto ...  $\gamma$  (da GROUP BY)

# Versione semplice

- Senza attributo di raggruppamento
- Calcolo sull'intera relazione

*AGG* *Funzione (Attributo) (Operando)*

o meglio

*AGG* *Funzione (Attributo) → Nome (Operando)*

esami	Matricola	Voto	Codice
	3456	30	04
	3456	26	02
	9283	27	01
	6554	26	01
	6554	26	05

- Il numero totale di esami

**AGG** count(\*) → NumeroEsami (esami)

NumeroEsami
5

esami	Matricola	Voto	Codice
	3456	30	04
	3456	26	02
	9283	27	01
	6554	26	01
	6554	26	05

- Il voto medio complessivo

**AGG**  $\text{avg}(\text{Voto}) \rightarrow \text{MediaGenerale}(\text{esami})$

MediaGenerale
27

esami	Matricola	Voto	Codice
	3456	30	04
	3456	26	02
	9283	27	01
	6554	26	01
	6554	26	05

- Il numero di esami e il voto medio complessivo

NumeroEsami	MediaGenerale
5	27

**AGG**  $\text{count}(\ast) \rightarrow \text{NumeroEsami}, \text{avg}(\text{Voto}) \rightarrow \text{MediaGenerale} (\text{esami})$

$\gamma \text{ count}(\ast) \rightarrow \text{NumeroEsami}, \text{avg}(\text{Voto}) \rightarrow \text{Media} (\text{Esami})$

# Versione completa

- Con attributi di raggruppamento
- Calcolo sulle partizioni

*AGG* *AttributiRaggruppamento; Funzione (Attributo), ... (Operando)*

o meglio

*AGG* *AttributiRaggruppamento; Funzione (Attributo) → Nome, ... (Operando)*

esami	Matricola	Voto	Codice
	3456	30	04
	3456	26	02
	9283	27	01
	6554	26	01
	6554	26	05

- Il voto medio per ciascuno studente

**AGG** Matricola; avg(Voto) → Media (esami)

Matricola	Media
3456	28
9283	27
6554	26

esami	Matricola	Voto	Codice
	3456	30	04
	3456	26	02
	9283	27	01
	6554	26	01
	6554	26	05

- Numero esami e voto medio per ciascuno studente

**AGG** Matricola; avg(Voto) → Media, count(\*) → NumEsami (esami)

Matricola	Media	NumEsami
3456	28	2
9283	27	1
6554	26	2

# Valori nulli

- vengono trattati correttamente (cioè ignorati nelle medie, nelle somme e nei conteggi)

# count

- $\text{count}(\ast)$  conta le ennuple
- $\text{count}(A)$  conta le ennuple che hanno il valore di  $A$ 
  - non i valori distinti di  $A$
  - quindi in assenza di valori il risultato è lo stesso, in presenza di valori nulli per  $A$ 
    - $\text{count}(A)$  è il numero di ennuple che non hanno il valore nullo

**esami**

Matricola	Voto	Codice
3456	30	04
3456	NULL	02
9283	27	01
6554	26	01
6554	26	05

- Numero esami e voto medio per ciascuno studente

**AGG** `Matricola; avg(Voto) → Media, count(*) → NumEsami (esami)`

Matricola	Media	NumEsami
3456	30	2
9283	27	1
6554	26	2

- Numero voti e voto medio per ciascuno studente

**AGG** `Matricola; avg(Voto) → Media, count(Voto) → NumVoti (esami)`

Matricola	Media	NumVoti
3456	30	1
9283	27	1
6554	26	2

# Join

- il join è l'operatore più interessante dell'algebra relazionale
- permette di correlare dati in relazioni diverse

# Il solito esempio

(con nomi di attributi modificati in "esami")

studenti	<u>Matricola</u>	Cognome	Nome	Data di nascita
	6554	Rossi	Mario	05/12/1978
	8765	Neri	Paolo	03/11/1976
	9283	Verdi	Luisa	12/11/1979
	3456	Rossi	Maria	01/02/1978

esami	<u>Matricola</u>	Voto	<u>Codice</u>
	3456	30	04
	3456	24	02
	9283	28	01
	6554	26	01

corsi	<u>Codice</u>	Titolo	Docente
	01	Analisi	Mario
	02	Chimica	Bruni
	04	Chimica	Verdi

studenti

Matricola	Cognome	Nome	Data di nascita
3456	Rossi	Maria	01/02/1978

esami

Matricola	Voto	Codice
3456	30	04
3456	24	02

corsi

Codice	Titolo	Docente
02	Chimica	Bruni
04	Chimica	Verdi

(studenti JOIN esami) JOIN corsi

Matricola	Cognome	Nome	Data di nascita	Voto	Codice	Titolo	Docente
3456	Rossi	Maria	01/02/1978	30	04	Chimica	Bruni
3456	Rossi	Maria	01/02/1978	24	02	Chimica	Verdi

# Join naturale

- operatore binario (generalizzabile)
- produce un risultato
  - sull'unione degli attributi degli operandi
  - con ennuple costruite ciascuna a partire da una ennupla di ognuno degli operandi

# Join, sintassi e semantica

- alla lavagna

JOIN ~~∞~~

JOIN NATURALE

$R_1(X_1)$     $R_2(X_2)$

$$R_1 \bowtie R_2 = \left\{ t \text{ su } X_1 X_2 \mid \begin{array}{l} \text{esistono } t_1 \in R_1 \text{ e } t_2 \in R_2 \\ t[X_1] = t_1 \text{ e } t[X_2] = t_2 \end{array} \right\}$$

# Esempi

- alla lavagna
  - join completo, non completo, vuoto,  $m \times n$

$R_1$

IMPIEGATO	REPARTO
Rossi	A
Neri	B
Bianchi	B

$R_2$

REPARTO	CAPO
A	Mozzi
B	Bruni

$R_1 \bowtie R_2$

IMPIEGATO	REPARTO	CAPO
Rossi	A	Mozzi
Neri	B	Bruni
Bianchi	B	Bruni

$R_1$

IMPIEGATO	REPARTO
Rossi	A
Neri	B
Bianchi	B

$R_2$

REPARTO	CAPO
B	Morzi
C	Bruni

$R_1 \bowtie R_2$

IMPIEGATO	REPARTO	CAPO
Neri	B	Morzi
Bianchi	B	Morzi

$R_1$

IMPIEGATO	REPARTO
Rossi	A
Neri	B
Bianchi	B

$R_2$

REPARTO	CAPO
C	Mori
D	Bruni

$R_1 \bowtie R_2$

IMPIEGATO    REPARTO    CAPO

$R_1$

IMPIEGATO	REPARTO
Rossi	A
Neri	A
Bianchi	A

$R_2$

REPARTO	CAPD
A	Mozzi
A	Bruni

$R_1 \bowtie R_2$

IMPIEGATO REPARTO CAPD

---

⋮

6 tuple!

$R_1$	A	B	C
1	1	1	1
2	1	2	2
3	2	2	2

$R_2$	B	C	D
1	1	1	1
2	1	2	3
3	2	2	4

$R_1 \bowtie R_2$

A	B	C	D
1	1	1	1
3	2	2	4

# Cardinalità del join

- Il join di  $R_1$  e  $R_2$  contiene un numero di ennuple  
...
  - alla lavagna (chiavi, vincoli integrità referenziale)

$R_1$       $|R_1|$

$R_2$       $|R_2|$

$$0 \leq |R_1 \bowtie R_2| \leq |R_1| \cdot |R_2|$$

$R_1$

<u>A</u>	B
1	0
2	4
3	4

$R_2$

<u>B</u>	C
1	x
2	y
3	z
4	z
...	

$$0 \leq |R_1 \bowtie R_2| \leq |R_1|$$

$R_1$

<u>A</u>	B
1	1
2	4
3	4
	4

$R_2$

<u>B</u>	C
1	x
2	y
3	z
4	w
...	

VINCOLO DI INTEGRITA' REF.

FRA  $R_1(B)$  E LA CHIAVE B DI  $R_2$

$$|R_1 \bowtie R_2| = |R_1|$$

# Cardinalità del join

- Il join di  $R_1$  e  $R_2$  contiene un numero di ennuple
  - 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 ennuple è
  - compreso fra zero e  $|R_1|$
- se il join coinvolge una chiave di  $R_2$  e un vincolo di integrità referenziale verso di essa, allora il numero di ennuple è
  - pari a  $|R_1|$

## Cardinalità del join, 2

- $R_1(A,B)$  ,  $R_2(B,C)$
- in generale

$$0 \leq |R_1 \text{ JOIN } R_2| \leq |R_1| \times |R_2|$$

- se  $B$  è chiave in  $R_2$

$$0 \leq |R_1 \text{ JOIN } R_2| \leq |R_1|$$

- se  $B$  è chiave in  $R_2$  ed esiste vincolo di integrità referenziale fra  $B$  (in  $R_1$ ) e  $R_2$ :

$$|R_1 \text{ JOIN } R_2| = |R_1|$$