

# Corso di Basi di dati

2020/2021



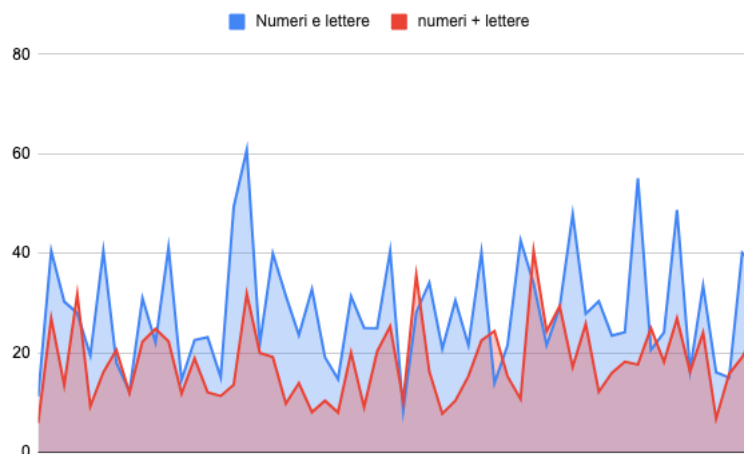
Paolo Buono  
paolo.buono@uniba.it



<http://ivu.di.uniba.it/people/buono.htm>

2

## numeri e lettere (57/108 risposte)



Basi di Dati + Laboratorio - Informatica Triennale - Corso A

193

## statistiche

- Tempo medio numeri (1-21): 10,22"
- Tempo medio lettere (a-z): 7,9"
- Somma media (numeri + lettere): 18,1"
- Media (numeri-lettere): 28,6"
- Switch overhead medio: 10,5"

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

194

## ALGEBRA RELAZIONALE

195

## Linguaggi per basi di dati

- operazioni sullo schema
  - DDL: data definition language
- operazioni sui dati
  - DML: data manipulation language
    - interrogazione ("query")
    - aggiornamento

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

196

## Linguaggi di interrogazione per basi di dati relazionali

- **Dichiarativi**
  - specificano le proprietà del risultato ("che cosa")
- **Procedurali**
  - specificano le modalità di generazione del risultato ("come")

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

197

## Linguaggi di interrogazione

- **Algebra relazionale**: procedurale
- **Calcolo relazionale**: dichiarativo (teorico)
- **SQL** (Structured Query Language): parzialmente dichiarativo (reale)
- **QBE** (Query by Example): dichiarativo (reale)

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

198

## Algebra relazionale

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

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

199

## Operatori dell'algebra relazionale

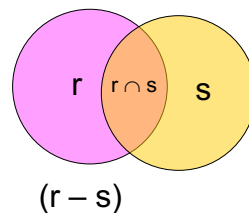
- unione, intersezione, differenza
- ridenominazione, selezione, proiezione
- join (join naturale, prodotto cartesiano, theta-join)

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

200

## Operatori insiemistici

- le relazioni sono insiemi
- i risultati debbono essere relazioni
- è possibile applicare **unione**, **intersezione**, **differenza** solo a relazioni definite sugli stessi attributi
- $r \cap s = r - (r - s)$



Basi di Dati + Laboratorio - Informatica Triennale - Corso A

201

## Unione

Laureati

Matricola	Nome	Età
7274	Rossi	42
7432	Neri	54
9824	Verdi	45

Quadri

Matricola	Nome	Età
9297	Neri	33
7432	Neri	54
9824	Verdi	45

Laureati  $\cup$  Quadri

Matricola	Nome	Età
7274	Rossi	42
7432	Neri	54
9824	Verdi	45
9297	Neri	33

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

202

## Intersezione

Laureati

Matricola	Nome	Età
7274	Rossi	42
7432	Neri	54
9824	Verdi	45

Quadri

Matricola	Nome	Età
9297	Neri	33
7432	Neri	54
9824	Verdi	45

Laureati  $\cap$  Quadri

Matricola	Nome	Età
7432	Neri	54
9824	Verdi	45

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

203

## Differenza

### Laureati

Matricola	Nome	Età
7274	Rossi	42
7432	Neri	54
9824	Verdi	45

### Quadri

Matricola	Nome	Età
9297	Neri	33
7432	Neri	54
9824	Verdi	45

### Laureati – Quadri

Matricola	Nome	Età
7274	Rossi	42

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

204

## Esempio unione

### Paternità

Padre	Figlio
Adamo	Abele
Adamo	Caino
Abramo	Isacco

### Maternità

Madre	Figlio
Eva	Abele
Eva	Set
Sara	Isacco

### Paternità $\cup$ Maternità

??

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

205

## Ridenominazione

- operatore monadico (con un argomento)
- "modifica lo schema" lasciando inalterata l'istanza dell'operando
- $\rho_{B_1 B_2 \dots B_k \leftarrow A_1 A_2 \dots A_k}(r) \forall t \in r \exists t' \text{ t.c.}$
- $\forall i = 1, \dots, k \quad t'[B_i] = t[A_i]$

$$\rho_{B_1 B_2 \dots B_k \leftarrow A_1 A_2 \dots A_k}$$

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

206

### Paternità

Padre	Figlio
Adamo	Abele
Adamo	Caino
Abramo	Isacco

$\rho_{\text{Genitore} \leftarrow \text{Padre}}(\text{Paternità})$

Genitore	Figlio
Adamo	Abele
Adamo	Caino
Abramo	Isacco

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

207



### Paternità

Padre	Figlio
Adamo	Abele
Adamo	Caino
Abramo	Isacco

### $\rho_{\text{Genitore} \leftarrow \text{Padre}}$ (Paternità)

Genitore	Figlio
Adamo	Abele
Adamo	Caino
Abramo	Isacco

### Maternità

Madre	Figlio
Eva	Abele
Eva	Set
Sara	Isacco

### $\rho_{\text{Genitore} \leftarrow \text{Madre}}$ (Maternità)

Genitore	Figlio
Eva	Abele
Eva	Set
Sara	Isacco

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

208

### $\rho_{\text{Genitore} \leftarrow \text{Padre}}$ (Paternità)

Genitore	Figlio
Adamo	Abele
Adamo	Caino
Abramo	Isacco

$\rho_{\text{Genitore} \leftarrow \text{Padre}}$  (Paternità)

$\rho_{\text{Genitore} \leftarrow \text{Madre}}$  (Maternità)

### $\rho_{\text{Genitore} \leftarrow \text{Madre}}$ (Maternità)

Genitore	Figlio
Eva	Abele
Eva	Set
Sara	Isacco

Genitore	Figlio
Adamo	Abele
Adamo	Caino
Abramo	Isacco
Eva	Abele
Eva	Set
Sara	Isacco

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

209

<b>Impiegati</b>	Cognome	Ufficio	Stipendio
	Rossi	Roma	55
	Neri	Milano	64
<b>Operai</b>	Cognome	Fabbrica	Salario
	Bruni	Monza	45
	Verdi	Latina	55
<p> <math>\rho</math> Sede, Retribuzione <math>\leftarrow</math> Ufficio, Stipendio (<b>Impiegati</b>)  <math>\cup</math>  <math>\rho</math> Sede, Retribuzione <math>\leftarrow</math> Fabbrica, Salario (<b>Operai</b>) </p>			
	Cognome	Sede	Retribuzione
	Rossi	Roma	55
	Neri	Milano	64
	Bruni	Monza	45
	Verdi	Latina	55

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

210

Selezione
<ul style="list-style-type: none"> <li>operatore monadico</li> <li>produce un risultato che <ul style="list-style-type: none"> <li>ha lo stesso schema dell'operando</li> <li>contiene un sottoinsieme delle ennuple dell'operando</li> <li>quelle che soddisfano una condizione</li> </ul> </li> </ul>

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

211

### Impiegati

Matricola	Cognome	Filiale	Stipendio
7309	Rossi	Roma	55
5998	Neri	Milano	64
9553	Milano	Milano	44
5698	Neri	Napoli	64

- impiegati che
  - guadagnano più di 50
  - guadagnano più di 50 e lavorano a Milano
  - hanno lo stesso nome della filiale presso cui lavorano

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

212

## Selezione, sintassi e semantica

- sintassi

$\sigma$  *Condizione* (*Operando*)

- *Condizione*: espressione booleana (come quelle dei vincoli di ennupla)
- semantica
  - il risultato contiene le ennuple dell'operando che soddisfano la condizione

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

213

- impiegati che guadagnano più di 50

### Impiegati

Matricola	Cognome	Filiale	Stipendio
7309	Rossi	Roma	55
5998	Neri	Milano	64
5698	Neri	Napoli	64

$\sigma_{\text{Stipendio} > 50}$  (Impiegati)

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

214

- impiegati che guadagnano più di 50  
e lavorano a Milano

### Impiegati

Matricola	Cognome	Filiale	Stipendio
5998	Neri	Milano	64

$\sigma_{\text{Stipendio} > 50 \text{ AND Filiale} = \text{'Milano'}}$  (Impiegati)

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

215

- impiegati che hanno lo stesso nome della filiale presso cui lavorano

### Impiegati

Matricola	Cognome	Filiale	Stipendio
9553	Milano	Milano	44

$\sigma_{\text{Cognome} = \text{Filiale}}(\text{Impiegati})$

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

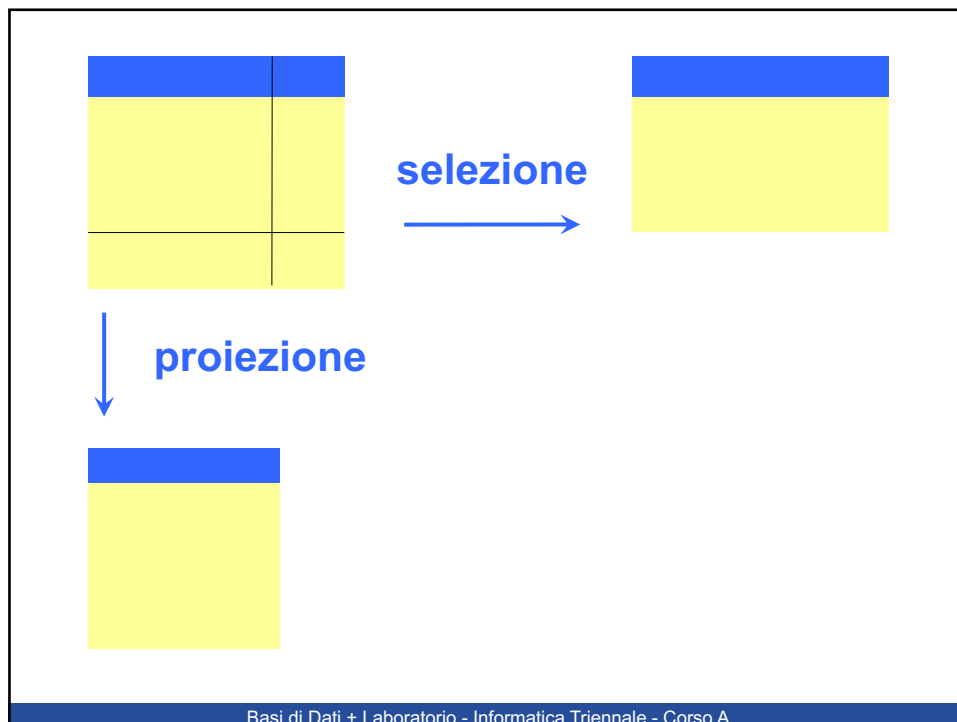
216

## Selezione e proiezione

- operatori "ortogonali"
- **selezione**:
  - decomposizione orizzontale
- **proiezione**:
  - decomposizione verticale

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

217



218

## Proiezione

- operatore monadico
- produce un risultato che
  - ha parte degli attributi dell'operando
  - contiene tutte le ennuple dell'operando

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

219

## Impiegati

Matricola	Cognome	Filiale	Stipendio
7309	Neri	Napoli	55
5998	Neri	Milano	64
9553	Rossi	Roma	44
5698	Rossi	Roma	64

- per tutti gli impiegati:
  - matricola e cognome
  - cognome e filiale

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

220

## Proiezione, sintassi e semantica

- sintassi

$\pi_{\text{ListaAttributi}}(\text{Operando})$

- semantica
  - il risultato contiene le ennuple ottenute da tutte le ennuple dell'operando ristrette agli attributi nella lista

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

221

- matricola e cognome di tutti gli impiegati

Matricola	Cognome
7309	Neri
5998	Neri
9553	Rossi
5698	Rossi

$\Pi_{\text{Matricola, Cognome}} (\text{Impiegati})$

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

222

- cognome e filiale di tutti gli impiegati

Cognome	Filiale
Neri	Napoli
Neri	Milano
Rossi	Roma

$\Pi_{\text{Cognome, Filiale}} (\text{Impiegati})$

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

223



## Proprietà selezione

- Cardinalità
  - $|\sigma_C(R)| \leq |(R)|$
- Grado
  - $\text{grado}(\sigma_C(R)) = \text{grado}(R)$
- Proprietà commutativa
  - $\sigma_{C1}(\sigma_{C2}(R)) = \sigma_{C2}(\sigma_{C1}(R))$
- Idempotenza
  - $\sigma_{C1}(\sigma_{C2}(R)) = \sigma_{C1 \wedge C2}(R)$

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

224

## Cardinalità della proiezione

- una proiezione contiene al più tante ennuple quante l'operando
  - può contenerne di meno
- se  $X$  è una superchiave di  $R$ , allora  $\pi_X(R)$  contiene esattamente tante ennuple quante  $R$
- $\pi_X(\pi_Y(R)) = \pi_X(R)$ 
  - se  $X \subseteq Y$

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

225

## $\sigma$ & $\pi$

- Combinando selezione e proiezione, si possiamo estrarre interessanti informazioni da una relazione
- Es.: 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))**

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

226

- Combinando selezione e proiezione, possiamo estrarre informazioni da **una** relazione
- spesso si vuole correlare informazioni presenti in relazioni diverse

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

227

## Join

- operatore più interessante dell'algebra relazionale
- permette di correlare dati in relazioni diverse

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

228

## Prove scritte in un concorso pubblico

- I compiti sono anonimi e ad ognuno è associata una busta chiusa con il nome del candidato
- Ciascun compito e la relativa busta vengono contrassegnati con uno stesso numero

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

229

1	25
2	13
3	27
4	28

1	Mario Rossi
2	Nicola Russo
3	Mario Bianchi
4	Remo Neri

Mario Rossi	25
Nicola Russo	13
Mario Bianchi	27
Remo Neri	28

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

230

Numero	Voto
1	25
2	13
3	27
4	28

Numero	Candidato
1	Mario Rossi
2	Nicola Russo
3	Mario Bianchi
4	Remo Neri

Numero	Candidato	Voto
1	Mario Rossi	25
2	Nicola Russo	13
3	Mario Bianchi	27
4	Remo Neri	28

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

231

## 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

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

232

## Join, sintassi e semantica

- $R_1(X_1), R_2(X_2)$
- $R_1 \text{ JOIN } R_2$  è una relazione su  $X_1X_2$   
$$\{ t \text{ su } X_1X_2 \mid \text{esistono } t_1 \in R_1 \text{ e } t_2 \in R_2 \\ \text{con } t[X_1] = t_1 \text{ e } t[X_2] = t_2 \}$$

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

233

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 ennupla contribuisce al risultato:
  - join **completo**

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

234

## 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

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

235

## Un join vuoto

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

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

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

236

## Un join completo, con $n \times m$ ennuple

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

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

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

237

## Risultato Impiegati JOIN Capi=?

Impiegati		Capi	
Impiegato	Reparto	Sede	Capo
Rossi	A	B	Mori
Neri	B	B	Bruni

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

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

238

## Join, sintassi e semantica

- $R_1(X_1), R_2(X_2)$
- $R_1 \text{ JOIN } R_2$  è una relazione su  $X_1X_2$ 

$$\{ t \text{ su } X_1X_2 \mid \text{esistono } t_1 \in R_1 \text{ e } t_2 \in R_2 \\ \text{con } t[X_1] = t_1 \text{ e } t[X_2] = t_2 \}$$

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

239



## 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, allora il numero di ennuple è pari a  $|R_1|$

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

240

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

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

241