

BASI DI DATI

LINGUAGGI PER BASI DI DATI

Prof. Fabio A. Schreiber
Dipartimento di Elettronica e Informazione
Politecnico di Milano

tratto da: Atzeni, Ceri, Paraboschi, Torlone - Basi di Dati - McGraw-Hill



CLASSIFICAZIONE DEI LINGUAGGI

- **LINGUAGGI FORMALI**
 - ALGEBRA RELAZIONALE
 - CALCOLO RELAZIONALE
 - DELLE TUPLE
 - DEI DOMINII
 - PROGRAMMAZIONE LOGICA
- **LINGUAGGI PROGRAMMATIVI**
 - SQL (Structured Query Language)
 - QBE (Query By Example)
 -

CLASSIFICAZIONE DEI LINGUAGGI

- **LINGUAGGI PROCEDURALI**

- SERVONO PER PRESCRIVERE IL PROCESSO ALGORITMICO MEDIANTE IL QUALE VIENE SODDISFATTA LA RICHIESTA (cosa e come)

- **LINGUAGGI NON PROCEDURALI**

- SERVONO PER DESCRIVERE LA FORMULAZIONE DEL RISULTATO (cosa)
- UN INTERPRETE SI INCARICA DI TROVARE IL MODO MIGLIORE PER RAGGIUNGERLO (come)

© Fabio A. Schreiber

Linguaggi per DB 2

CLASSIFICAZIONE DEI LINGUAGGI

- **LINGUAGGI DI DEFINIZIONE DEI DATI (DDL)**

- PER CREARE GLI SCHEMI DEI DATI E DEFINIRE LE LORO PROPRIETA'

- **LINGUAGGI DI MANIPOLAZIONE DEI DATI (DML)**

- PER AGGIORNARE (CREARE) LE ISTANZE DEI DATI
- PER L'INTERROGAZIONE DEI DATI (p.e. scrivere programmi applicativi)

© Fabio A. Schreiber

Linguaggi per DB 3

LINGUAGGI DI MANIPOLAZIONE DEI DATI

L'OPERAZIONE FONDAMENTALE PER MANIPOLARE UNA BASE DI DATI E' L'INTERROGAZIONE

- CON L'INTERROGAZIONE SI ESEGUE
L'ACCESSO AI DATI DESIDERATI
- AVENDO AVUTO ACCESSO AI DATI
DESIDERATI E' ANCHE POSSIBILE
 - MODIFICARLI (UPDATE)
 - CANCELLARLI (DELETE)
 - INSERIRNE DI NUOVI (INSERT)
 - AGGIUNGERNE (APPEND)

© Fabio A. Schreiber

Linguaggi per DB 4

ALGEBRA RELAZIONALE

E' UN LINGUAGGIO

- FORMALE
- PROCEDURALE
- PER FORMULARE INTERROGAZIONI

BASATO SU

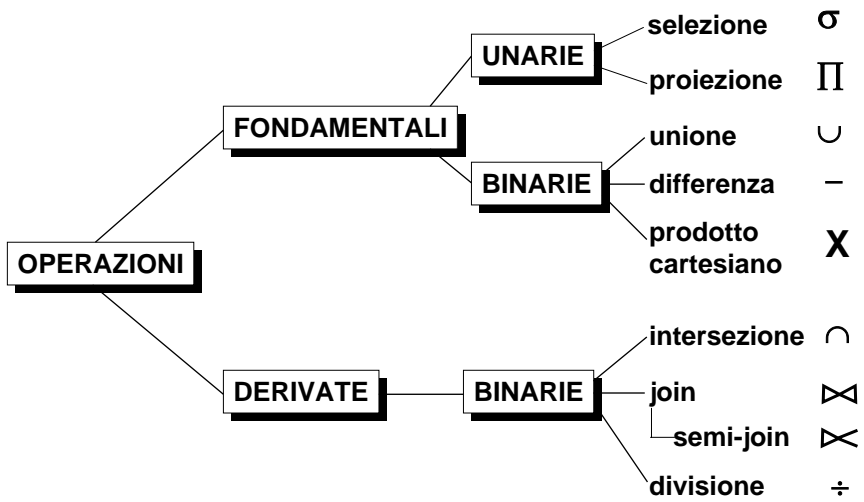
- 5 OPERAZIONI FONDAMENTALI

DEFINITO DA E. CODD NEL 1970

© Fabio A. Schreiber

Linguaggi per DB 5

OPERAZIONI DELL'ALGEBRA RELAZIONALE



© Fabio A. Schreiber

Linguaggi per DB 6

ESEMPIO

studente

MATR.	NOME	CITTA'	C-DIP
123	Carlo	Bologna	Inf
415	Paola	Torino	Inf
702	Antonio	Roma	Log

corso

C-CORSO	TITOLO	DOCENTE
1	matematica	Barozzi
2	informatica	Meo

esame

MATR.	C-CORSO	DATA	VOTO
123	1	7/9/1997	30
123	2	8/1/1998	28
702	2	7/9/1997	20

© Fabio A. Schreiber

Linguaggi per DB 7

ESEMPIO DI SELEZIONE

$\sigma_{\text{nome} = \text{"Paola"}} \text{ studente}$

E' UNA TABELLA CON

- **SCHEMA**
LO STESSO DI studente (STESSO GRADO)
- **ISTANZA**
LE TUPLE DI studente CHE SODDISFANO IL PREDICATO DI SELEZIONE (CARDINALITA' \leq)

MATR.	NOME	CITTA'	C-DIP
415	Paola	Torino	Inf

© Fabio A. Schreiber

Linguaggi per DB 8

SINTASSI DELLA SELEZIONE

$\sigma_P r$

σ	OPERATORE
P	PREDICATO
r	OPERANDO

P ESPRESSIONE BOOLEANA DI PREDICATI SEMPLICI

OPERAZIONI BOOLEANE

- AND ($P_1 \wedge P_2$)
- OR ($P_1 \vee P_2$)
- NOT ($\neg P_1$)

PREDICATI SEMPLICI

- TRUE, FALSE
- termine comparatore termine

COMPARATORE

=, \neq , <, \leq , >, \geq

TERMINE

- COSTANTE, ATTRIBUTO
- ESPRESSIONE ARITMETICA DI COSTANTI E ATTRIBUTI

© Fabio A. Schreiber

Linguaggi per DB 9

ESEMPIO DI SELEZIONE

σ (CITTÀ = "Torino") \vee ((CITTÀ = "Roma") \wedge \neg (C-DIP="Log")) studente

studente

MATR.	NOME	CITTA'	C-DIP
123	Carlo	Bologna	Inf
415	Paola	Torino	Inf
702	Antonio	Roma	Log

© Fabio A. Schreiber

Linguaggi per DB 10

ESEMPIO DI PROIEZIONE

$\Pi_{\text{NOME, C-DIP}}$ studente

E' UNA TABELLA CON

- SCHEMA
SOLO GLI ATTRIBUTI NOME E C-DIP (GRADO \leq)
- ISTANZA
LA RESTRIZIONE DELLE TUPLE DI studente SUGLI
ATTRIBUTI NOME E C-DIP (CARDINALITA' \leq)

NOME	C-DIP
Carlo	Inf
Paola	Inf
Antonio	Log

© Fabio A. Schreiber

Linguaggi per DB 11

SINTASSI DELLA PROIEZIONE

$$\Pi_{\text{ATTR1}, \dots, \text{ATTRn}} r$$

- FORMALMENTE LA PROIEZIONE ELIMINA I DUPLICATI

$$\Pi_{\text{C-DIP}} \text{studente}$$

C-DIP
Inf
Log

- NEI SISTEMI IN USO L'ELIMINAZIONE DEI DUPLICATI VA RICHIESTA ESPLICITAMENTE

© Fabio A. Schreiber

Linguaggi per DB 12

OPERAZIONI NON ALGEBRICHE

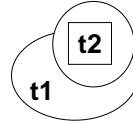
- **ASSEGNAZIONE**
SERVE A DARE UN NOME AL RISULTATO DI UN'OPERAZIONE ALGEBRICA
 - $\text{torinesi} = \sigma_{(\text{CITTÀ} = \text{"Torino"})} \text{studenti}$
 - $\text{informatici} = \sigma_{(\text{C-DIP} = \text{"inf"})} \text{studenti}$
 - **RIDENOMINAZIONE**
PERMETTE DI MODIFICARE I NOMI DEGLI ATTRIBUTI
 - $\rho_{\text{genitore} \leftarrow \text{padre}} \text{paternità}$
 - $\rho_{\text{genitore} \leftarrow \text{madre}} \text{maternità}$
- IN QUESTO MODO E' POSSIBILE RENDERE COMPATIBILI DOMINII ORIGINARIAMENTE DIVERSI

© Fabio A. Schreiber

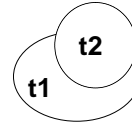
Linguaggi per DB 13

UNIONE E DIFFERENZA

- $\text{tabella1} \cup \text{tabella2}$



- $\text{tabella1} - \text{tabella2}$



**tabella1 E tabella2 DEVONO ESSERE
COMPATIBILI**

- STESSO GRADO
- DOMINI ORDINATAMENTE DELLO STESSO TIPO

© Fabio A. Schreiber

Linguaggi per DB 14

ESEMPIO DI UNIONE

informatici \cup torinesi

E' UNA TABELLA CON

- **SCHEMA**
LO STESSO DI **informatici**
- **ISTANZA**
L'UNIONE DELLE TUPLE DI **informatici** E **torinesi**

MATR.	NOME	CITTA'	C-DIP
123	Carlo	Bologna	Inf
415	Paola	Torino	Inf

© Fabio A. Schreiber

Linguaggi per DB 15

ESEMPIO DI DIFFERENZA

informatici – torinesi

E' UNA TABELLA CON

- **SCHEMA**
LO STESSO DI informatici
- **ISTANZA**
LA DIFFERENZA DELLE TUPLE DI informatici E torinesi

MATR.	NOME	CITTA'	C-DIP
123	Carlo	Bologna	Inf

© Fabio A. Schreiber

Linguaggi per DB 16

PRODOTTO CARTESIANO

$r \times s$

E' UNA TABELLA CON

- **SCHEMA**
TUTTI GLI ATTRIBUTI DI r E DI s
(GRADO ($r \times s$) = GRADO (r) + GRADO (s))
- **ISTANZA**
TUTTE LE POSSIBILI COPPIE DI TUPLE DI r E DI s
(CARD ($r \times s$) = CARD (r) * CARD (s))

© Fabio A. Schreiber

Linguaggi per DB 17

ESPRESSIONI ALGEBRICHE

- **CONCATENAZIONE DI PIU' OPERAZIONI ALGEBRICHE**
- **ESPRIMONO INTERROGAZIONI IN MODO FORMALE**
- **CONSENTONO DI ESTRARRE INFORMAZIONI DALLA BASE DI DATI**

© Fabio A. Schreiber

Linguaggi per DB 18

UN ALTRO ESEMPIO DI BASE DI DATI

- **correntista (NOME, VIA, CITTA')**
 - **cliente (CORRENTISTA, IMPIEGATO)**
 - **filiale (NOME, PATRIMONIO, CITTA')**
 - **depositi (FILIALE, NUMERO, CORRENTISTA, SALDO)**
 - **prestiti (FILIALE, NUMERO, CORRENTISTA, QUANTITA')**
-
- ```
graph TD; C[cliente] ---|CORRENTISTA| D[depositi]; C ---|CORRENTISTA| P[prestiti]; F[filiale] ---|FILIALE| D; F ---|FILIALE| P;
```

© Fabio A. Schreiber

Linguaggi per DB 19

## ESEMPIO DI USO DI ESPRESSIONI ALGEBRICHE

cliente

| CORRENTISTA                   | IMPIEGATO                    |
|-------------------------------|------------------------------|
| Rossi<br>Brambilla<br>Ferrari | Ghezzi<br>Ferrari<br>Ferrari |

correntista

| NOME                                             | VIA                                      | CITTA'                     |
|--------------------------------------------------|------------------------------------------|----------------------------|
| Rossi<br>Bianchi<br>Neri<br>Brambilla<br>Ferrari | Dante<br>Verdi<br>Po<br>Orefici<br>Verdi | MI<br>RM<br>TO<br>MI<br>TO |

**INTERROGAZIONE**  
**FORNIRE I NOMI E LE CITTA' DI**  
**RESIDENZA DEI CLIENTI**  
**DELL'IMPIEGATO "Ferrari"**

| $r_4$ | CLIENTE.<br>CORRENT. | CORRENT.<br>CITTA' |
|-------|----------------------|--------------------|
|       | Brambilla<br>Ferrari | MI<br>TO           |

© Fabio A. Schreiber

$r_1 = \text{cliente X correntista}$

| CLIENTE.<br>CORRENT.                                                          | CLIENTE.<br>IMPIEGATO                                                     | CORRENT.<br>NOME                                                         | CORRENT.<br>VIA                                     | CORRENT.<br>CITTA'                     |
|-------------------------------------------------------------------------------|---------------------------------------------------------------------------|--------------------------------------------------------------------------|-----------------------------------------------------|----------------------------------------|
| Rossi<br>Rossi<br>Rossi<br>Rossi                                              | Ghezzi<br>Ghezzi<br>Ghezzi<br>Ghezzi                                      | Rossi<br>Bianchi<br>Neri<br>Brambilla<br>Ferrari                         | Dante<br>Verdi<br>Po<br>Orefici<br>Verdi            | MI<br>RM<br>TO<br>MI<br>TO             |
| Brambilla<br>Brambilla<br>Brambilla                                           | Ferrari<br>Ferrari<br>Ferrari                                             | Rossi<br>Bianchi<br>Neri                                                 | Dante<br>Verdi<br>Po                                | MI<br>RM<br>TO                         |
| Brambilla<br>Brambilla<br>Ferrari<br>Ferrari<br>Ferrari<br>Ferrari<br>Ferrari | Ferrari<br>Ferrari<br>Ferrari<br>Ferrari<br>Ferrari<br>Ferrari<br>Ferrari | Brambilla<br>Orefici<br>Ferrari<br>Rossi<br>Bianchi<br>Neri<br>Brambilla | Orefici<br>Verdi<br>Dante<br>Verdi<br>Po<br>Orefici | MI<br>MI<br>TO<br>MI<br>RM<br>TO<br>MI |
| Ferrari                                                                       | Ferrari                                                                   | Ferrari                                                                  | Verdi                                               | TO                                     |

$r_1 = \text{cliente X correntista}$

$r_2 = \sigma_{\text{CLIENTE.IMPIEGATO}="Ferrari"} r_1$

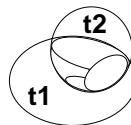
$r_3 = \sigma_{\text{CLIENTE.CORRENT}=\text{CORRENT.NOME}} r_2$

$r_4 = \pi_{\text{CLIENTE.CORRENT, CORRENT.CITTA'}} r_3$

Linguaggi per DB 20

## INTERSEZIONE

tabella1  $\cap$  tabella2



- tabella1 E tabella2 DEVONO ESSERE COMPATIBILI
- E' DERIVABILE TRAMITE LA SEGUENTE FORMULA

$$r \cap s = r - (r - s)$$

© Fabio A. Schreiber

Linguaggi per DB 21

## ESEMPIO DI INTERSEZIONE

**informatici**  $\cap$  **torinesi**

**E' UNA TABELLA CON**

- **SCHEMA**  
LO STESSO DI **informatici**
- **ISTANZA**  
L'INTERSEZIONE DELLE TUPLE DI **informatici** E **torinesi**

| MATR. | NOME  | CITTA' | C-DIP |
|-------|-------|--------|-------|
| 415   | Paola | Torino | Inf   |

© Fabio A. Schreiber

Linguaggi per DB 22

## ESEMPIO DI JOIN

**studente**  $\bowtie$  **esame**  
**studente.MATR=esame.MATR**

**E' UNA TABELLA CON**

- **SCHEMA**  
LA CONCATENAZIONE DEGLI SCHEMI DI **studente** E **esame**
- **ISTANZA**  
LE TUPLE OTTENUTE CONCATENANDO QUELLE TUPLE DI **studente** E **esame** CHE SODDISFANO IL PREDICATO

| MATR. | NOME    | CITTA'  | C-DIP | MATR. | C-CORSO | DATA     | VOTO |
|-------|---------|---------|-------|-------|---------|----------|------|
| 123   | Carlo   | Bologna | Inf   | 123   | 1       | 7/9/1997 | 30   |
| 123   | Carlo   | Bologna | Inf   | 123   | 2       | 8/1/1998 | 28   |
| 702   | Antonio | Roma    | Log   | 702   | 2       | 7/9/1997 | 20   |

© Fabio A. Schreiber

Linguaggi per DB 23

## SINTASSI DEL JOIN

$$r_1 \bowtie_P r_2$$

- E' EQUIVALENTE ALLA SEGUENTE ESPRESSIONE

$$\sigma_P r_1 \times r_2$$

- SINTASSI DEL PREDICATO DI JOIN
  - ESPRESSIONE CONGIUNTIVA DI PREDICATI SEMPLICI
    - ATTR<sub>1</sub> COMP ATTR<sub>2</sub>
      - ATTR<sub>1</sub> ∈ r<sub>1</sub>
      - ATTR<sub>2</sub> ∈ r<sub>2</sub>
      - COMP =, ≠, <, ≤, >, ≥
  - ATTRIBUTI OMONIMI SONO RESI NON AMBIGUI USANDO LA NOTAZIONE PUNTATA
    - r<sub>i</sub>. ATTR<sub>j</sub>

© Fabio A. Schreiber

Linguaggi per DB 24

## EQUI-JOIN E JOIN NATURALE

- EQUI-JOIN
  - IL PREDICATO AMMETTE SOLO CONFRONTI DI UGUAGLIANZA
- JOIN NATURALE
  - EQUI-JOIN DI TUTTI GLI ATTRIBUTI OMONIMI
    - SI OMETTE IL PREDICATO
    - SI ELIMINA LA COLONNA RIPETUTA

studente  $\bowtie$  esame

| MATR. | NOME    | CITTA'  | C-DIP | C-CORSO | DATA     | VOTO |
|-------|---------|---------|-------|---------|----------|------|
| 123   | Carlo   | Bologna | Inf   | 1       | 7/9/1997 | 30   |
| 123   | Carlo   | Bologna | Inf   | 2       | 8/1/1998 | 28   |
| 702   | Antonio | Roma    | Log   | 2       | 7/9/1997 | 20   |

© Fabio A. Schreiber

Linguaggi per DB 25

## SINTASSI DEL JOIN NATURALE

$$S \bowtie R = \Pi_{R \cup S} (r_{r.A1=s.A1 \wedge \dots \wedge r.An=s.An} S)$$

- $\{A1, \dots, An\} = R \cap S$
- $R \cap S = \Phi \Rightarrow \bowtie \equiv X$
- $\bowtie$  E' ASSOCIATIVA

© Fabio A. Schreiber

Linguaggi per DB 26

## JOIN NATURALE DI TRE TABELLE

studente  $\bowtie$  esame  $\bowtie$  corso

| MATR. | NOME    | CITTA'  | C-DIP | C-CORSO | DATA     | VOTO | TITOLO      | DOCENTE |
|-------|---------|---------|-------|---------|----------|------|-------------|---------|
| 123   | Carlo   | Bologna | Inf   | 1       | 7/9/1997 | 30   | matematica  | Barozzi |
| 123   | Carlo   | Bologna | Inf   | 2       | 8/1/1998 | 28   | informatica | Meo     |
| 702   | Antonio | Roma    | Log   | 2       | 7/9/1997 | 20   | informatica | Meo     |

© Fabio A. Schreiber

Linguaggi per DB 27

## SEMI-JOIN

**studente**  $\bowtie$  **esame**  
studente.MATR=esame.MATR

- E' EQUIVALENTE ALLA SEGUENTE ESPRESSIONE

$\Pi_{\text{studente.*}} (\text{studente} \bowtie \text{esame})$

- PRODUCE UN SOTTOINSIEME DI **studente**
  - solo gli studenti che hanno dato almeno un esame

studente

| MATR. | NOME    | CITTA'  | C-DIP |
|-------|---------|---------|-------|
| 123   | Carlo   | Bologna | Inf   |
| 702   | Antonio | Roma    | Log   |

© Fabio A. Schreiber

Linguaggi per DB 28

## SEMI-JOIN NATURALE

**studente**  $\bowtie$  **esame**

- E' EQUIVALENTE ALLA SEGUENTE ESPRESSIONE

$\Pi_{\text{studente.*}} (\text{studente} \bowtie \text{esame})$

- IN QUESTO ESEMPIO IL RISULTATO E' UGUALE AL CASO PRECEDENTE

studente

| MATR. | NOME    | CITTA'  | C-DIP |
|-------|---------|---------|-------|
| 123   | Carlo   | Bologna | Inf   |
| 702   | Antonio | Roma    | Log   |

© Fabio A. Schreiber

Linguaggi per DB 29

## DIVISIONE

$$r \div s$$

- E' EQUIVALENTE ALLA SEGUENTE ESPRESSIONE

$$\Pi_{R-S} r - \Pi_{R-S} ((\Pi_{R-S} r \times s) - r) \quad \text{CON } S \subseteq R$$

- E' UTILE PER ESPRIMERE INTERROGAZIONI CHE CONTENGONO IL QUANTIFICATORE UNIVERSALE  $\forall$

|     |   |   |   |   |   |     |   |   |   |
|-----|---|---|---|---|---|-----|---|---|---|
| $r$ |   |   |   |   |   | $s$ |   |   |   |
| X   | Y | W | Z | ÷ | W | Z   | = | X | Y |
| a   | b | c | d |   | c | d   |   | a | b |
| a   | b | e | f |   | e | f   |   | e | d |
| b   | c | e | f |   |   |     |   |   |   |
| e   | d | c | d |   |   |     |   |   |   |
| e   | d | e | f |   |   |     |   |   |   |
| a   | b | d | e |   |   |     |   |   |   |

© Fabio A. Schreiber

Linguaggi per DB 30

## ESEMPIO DI USO DELLA DIVISIONE

**TROVARE I NOMI DEI CORRENTISTI CHE HANNO UN DEPOSITO IN TUTTE LE FILIALI DI Milano**

- TUTTE LE FILIALI DI Milano

$$r_1 = \Pi_{\text{NOME}} (\sigma_{\text{CITTA}' = \text{"Milano"} \text{ filiale}})$$

- I CLIENTI DI CIASCUNA FILIALE

$$r_2 = \Pi_{\text{FILIALE, CORRENTISTA}} \text{ depositi}$$

- I CLIENTI DI  $r_2$  CHE SONO APPAIATI CON TUTTE LE FILIALI DI  $r_1$

$$r_3 = r_2 \div r_1$$

© Fabio A. Schreiber

Linguaggi per DB 31



## ESEMPIO DI ESPRESSIONE COMPLESSA

ESTRARRE IL NOME DEGLI STUDENTI CHE HANNO  
SOSTENUTO GLI ESAMI DI informatica E matematica  
LO STESSO GIORNO

$\Pi_{\text{NOME}} \text{ studente} \bowtie (\text{esame} \bowtie \sigma_{\text{TITOLO}=\text{"informatica"}} \text{ corso}) \bowtie$

$\bowtie \rho_{\text{MATR}=\text{MATRMAT} \wedge \text{DATA}=\text{DATAMAT}} \rho_{\text{MATRMAT},\text{DATAMAT}} \leftarrow \text{MATR},\text{DATA}$

$\Pi_{\text{MATR},\text{DATA}} (\text{esame} \bowtie \sigma_{\text{TITOLO}=\text{"matematica"}} \text{ corso})$

## ESEMPIO DI ESPRESSIONE COMPLESSA

ESTRARRE L'ULTIMO ESAME SOSTENUTO DA  
CIASCUNO STUDENTE

.....

$\text{esame} - \text{esame} \bowtie \rho_{\text{MATR}=\text{MATRNEW} \wedge \text{DATA} < \text{DATANEW}}$

$\bowtie \rho_{\text{MATRNEW},\text{DATANEW}} \leftarrow \text{MATR},\text{DATA} \Pi_{\text{MATR},\text{DATA}} \text{esame}$

## RAPPRESENTAZIONE DELLE ESPRESSIONI TRAMITE ALBERI

OGNI ESPRESSIONE DELL'ALGEBRA RELAZIONALE PUO' ESSERE RAPPRESENTATA CON UN ALBERO CHE ESPRIME L'ORDINE DI VALUTAZIONE DEGLI OPERATORI

- LA RADICE E' POSTA IN ALTO
- AD OGNI OPERATORE CORRISPONDE UN NODO INTERMEDIO
  - OPERATORI UNARI HANNO UN RAMO IN INGRESSO E UNO IN USCITA
  - OPERATORI BINARI HANNO DUE RAMI IN INGRESSO E UNO IN USCITA
- GLI OPERANDI (RELAZIONI) CORRISPONDONO ALLE FOGLIE

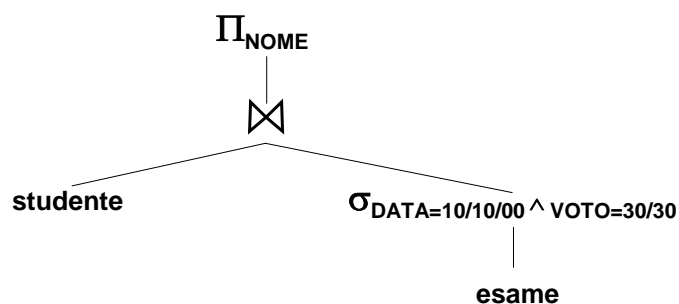
© Fabio A. Schreiber

Linguaggi per DB 34

## ESEMPIO DI ALBERO

NOME DEGLI STUDENTI CHE HANNO SOSTENUTO UN ESAME IL 10/10/00 CON VOTO PARI A 30/30

$\Pi_{\text{NOME}}$  studente  $\bowtie$   $\sigma_{\text{DATA}=10/10/00 \wedge \text{VOTO}=30/30}$  esame

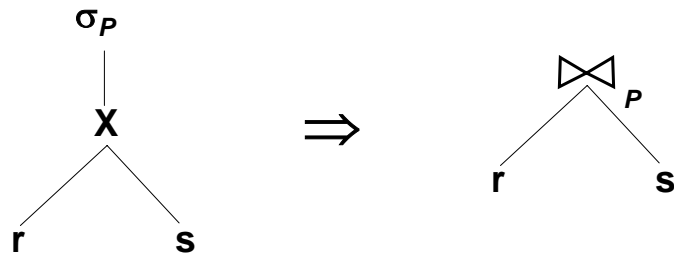


© Fabio A. Schreiber

Linguaggi per DB 35

## EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

### 1 - ELIMINAZIONE DEI PRODOTTI CARTESIANI



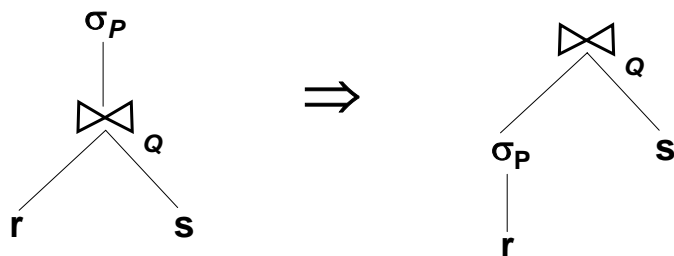
VALE SE  $P$  E' UNA CONGIUNZIONE DI PREDICATI  
ATTR *comp* ATTR

© Fabio A. Schreiber

Linguaggi per DB 36

## EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

### 2 - PUSH DELLA SELEZIONE RISPETTO AL JOIN



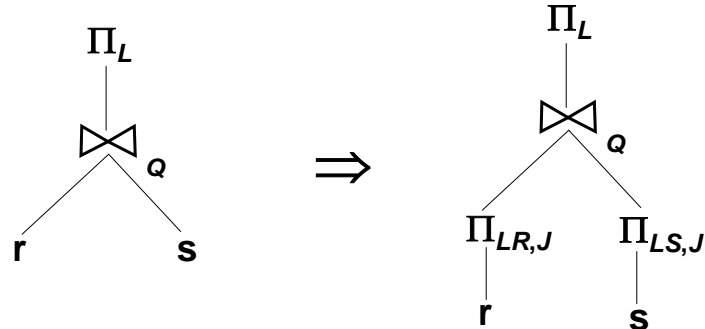
VALE SE  $P$  E' UN PREDICATO CHE SI APPLICA AI  
SOLI ATTRIBUTI DI  $r$

© Fabio A. Schreiber

Linguaggi per DB 37

## EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

### 3 - PUSH DELLA PROIEZIONE RISPETTO AL JOIN



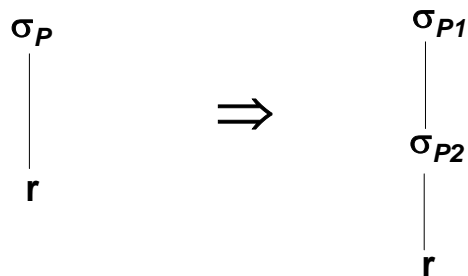
- $J$  SONO GLI ATTRIBUTI NECESSARI A VALUTARE  $Q$
- $LR = L - \text{SCHEMA}(s) - J$
- $LS = L - \text{SCHEMA}(r) - J$

© Fabio A. Schreiber

Linguaggi per DB 38

## EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

### 4 - IDEMPOTENZA DELLA SELEZIONE



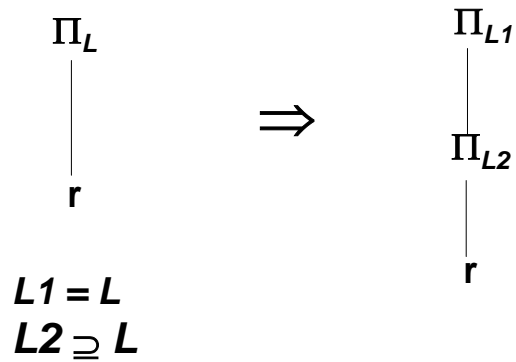
$$P = P1 \wedge P2$$

© Fabio A. Schreiber

Linguaggi per DB 39

## EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

### 5 - IDEMPOTENZA DELLA PROIEZIONE

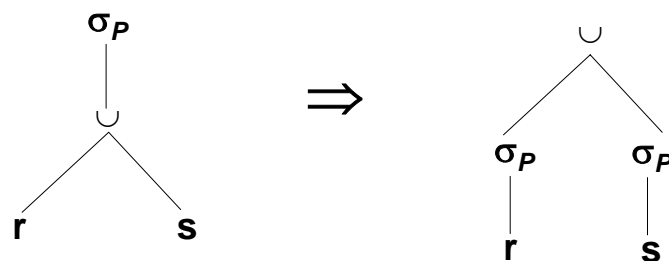


© Fabio A. Schreiber

Linguaggi per DB 40

## EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

### 6 - PUSH DELLA SELEZIONE RISPETTO ALL'UNIONE

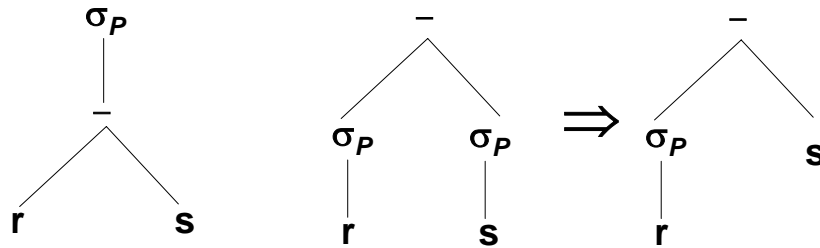


© Fabio A. Schreiber

Linguaggi per DB 41

## EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

### 7 - PUSH DELLA SELEZIONE RISPETTO ALLA DIFFERENZA

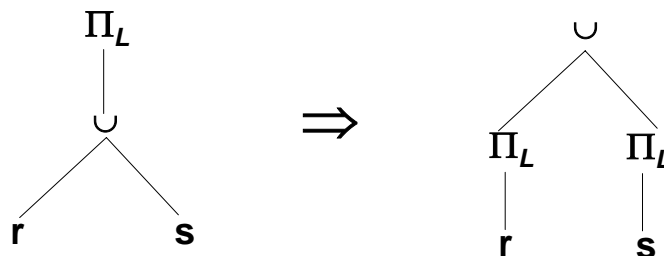


© Fabio A. Schreiber

Linguaggi per DB 42

## EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

### 8 - PUSH DELLA PROIEZIONE RISPETTO ALL'UNIONE



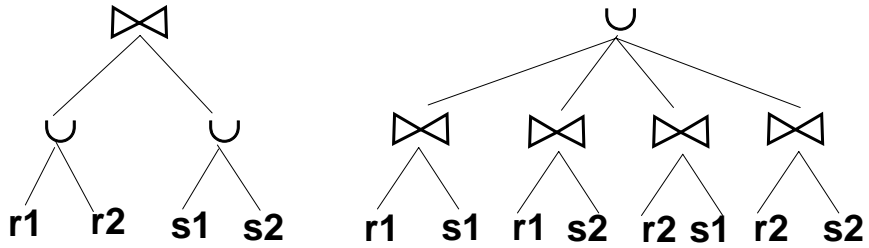
**ATTENZIONE: IL PUSH DELLA PROIEZIONE RISPETTO A DIFFERENZA E INTERSEZIONE NON VALE**

© Fabio A. Schreiber

Linguaggi per DB 43

## EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

### 9 - COMMUTAZIONE DI JOIN E UNIONE



© Fabio A. Schreiber

Linguaggi per DB 44

## FORMULE UTILI

$$r \cup r = r$$

$$r \cap r = r$$

$$r - r = \Phi$$

$$r \cap \sigma_P r = \sigma_P r$$

$$r \cup \sigma_P r = r$$

$$r - \sigma_P r = \sigma_{\neg P} r$$

$$\sigma_{P_1} r \cap \sigma_{P_2} r = \sigma_{P_1 \wedge P_2} r$$

$$\sigma_{P_1} r \cup \sigma_{P_2} r = \sigma_{P_1 \vee P_2} r$$

$$\sigma_{P_1} r - \sigma_{P_2} r = \sigma_{P_1 \wedge \neg P_2} r$$

$$\sigma_P \Phi = \Phi$$

$$\Pi_L \Phi = \Phi$$

$$r \cap \Phi = \Phi$$

$$r \cup \Phi = r$$

$$r - \Phi = r$$

$$\Phi - r = \Phi$$

$$r \cap \Phi = \Phi$$

$$r \cup \Phi = r$$

© Fabio A. Schreiber

Linguaggi per DB 45

## **OTTIMIZZAZIONE ALGEBRICA**

- **TRA TUTTE LE RAPPRESENTAZIONI EQUIVALENTI CONVIENE SCEGLIERE QUELLA MENO COSTOSA DA ESEGUIRE**
  - **MINIMIZZARE LA DIMENSIONE DEI RISULTATI INTERMEDI**
    - **UTILIZZARE, DOVE POSSIBILE, LE TRASFORMAZIONI DI PUSH (2,3,6,7,8)**
    - **USARE LE TRASFORMAZIONI DI IDEMPOTENZA (4, 5) PER GENERARE NUOVE SELEZIONI E PROIEZIONI**
  - **CERCARE DI COMBINARE I PRODOTTI CARTESIANI CON LE SELEZIONI I CUI PREDICATI APPARTENGONO AD ENTRAMBI GLI OPERANDI (DIVENTANO JOIN)**

© Fabio A. Schreiber

Linguaggi per DB 46

## **DIMENSIONI DEI RISULTATI INTERMEDI**

**VALUTARE IN GENERALE LE DIMENSIONI DEI RISULTATI INTERMEDI E' OPERAZIONE COMPLESSA**

**SI USANO METODI APPROSSIMATI FACENDO RICORSO AD OPERAZIONI STATISTICHE**

**SI SUPPONE UNIFORME LA DISTRIBUZIONE DEI VALORI NEI DOMINI E NON CORRELATI I DATI RELATIVI AI PREDICATI NELLE INTERROGAZIONI**

© Fabio A. Schreiber

Linguaggi per DB 47



## DIMENSIONI DEI RISULTATI INTERMEDI

### PARAMETRI QUANTITATIVI

- $CARD(r_i)$  NUMERO DI TUPLE DI  $r_i$
- $SIZE(t_i)$  DIMENSIONE (in byte) DELLA TUPLA DI  $r_i$
- $SIZE(A_j)$  DIMENSIONE (in byte) DELL'ATTRIBUTO  $A_j$  DI  $r_i$
- $VAL(A_j)$  NUMERO DI VALORI DISTINTI DELL'ATTRIBUTO  $A_j$  DI  $r_i$
- $S_i = 1/VAL(A_j)$  LA SELETTIVITA' DEL DOMINIO SUL QUALE E' DEFINITO L'ATTRIBUTO  $A_j$

© Fabio A. Schreiber

Linguaggi per DB 48

## DIMENSIONI DEI RISULTATI INTERMEDI

### OPERAZIONE DI SELEZIONE

$$CARD(\text{risultato}) = S_p * CARD(\text{operando})$$

$$S_p = \text{SELETTIVITA' DEL PREDICATO}$$

- $P = A \wedge B$                        $S_p = S_A * S_B$
- $P = A \vee B$                        $S_p = S_A + S_B - S_A * S_B$
- $P = \neg A$                           $S_p = 1 - S_A$

$$SIZE(\text{risultato}) = SIZE(\text{operando})$$

$$VAL(A_i) = 1 \quad i \equiv P$$

$$VAL(A_i) = col(CARD(\text{operando}), VAL(A_i), CARD(\text{risultato})) \quad i \notin P$$

© Fabio A. Schreiber

Linguaggi per DB 49

## ESEMPIO DI OTTIMIZZAZIONE

### DATO LO SCHEMA

- r(A,B,C)
- s(C,D,E)
- t(D,E,F,G)

### SI OTTIMIZZI L'ESPRESSIONE

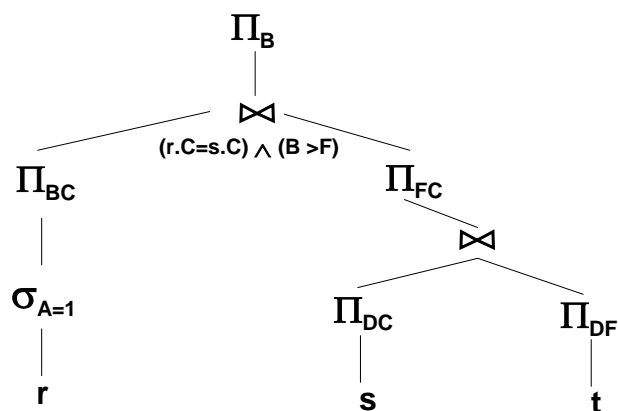
$$\Pi_B \sigma_{(r.C=s.C) \wedge (s.D=t.D) \wedge (r.A=1) \wedge (r.B>t.F)} r \times s \times t$$

© Fabio A. Schreiber

Linguaggi per DB 50

## ESEMPIO DI OTTIMIZZAZIONE

$$\Pi_B \sigma_{(r.C=s.C) \wedge (s.D=t.D) \wedge (r.A=1) \wedge (r.B>t.F)} r \times s \times t$$



© Fabio A. Schreiber

Linguaggi per DB 51

## MINIMA ESPRESSIONE CONTENITORE (CONTAIN VIEW)

DATE  $n$  INTERROGAZIONI SU UNA BASE DI DATI  
CONSENTE DI ESTRARNE I RISULTATI MEDIANTE  
SELEZIONI E PROIEZIONI SU DI ESSA

**DATO LO SCHEMA**

- $r(A,B,C)$
- $s(B,C,D,E,F)$
- $t(D, G)$

**SI RICAVI LA CONTAIN VIEW PER LE INTERROGAZIONI**

**1)  $\Pi_{r,C,G} \sigma_{(A=1) \wedge (r.B=s.B) \wedge (s.D=t.D) \wedge (E=1) \wedge (G>3)} r \times s \times t$**

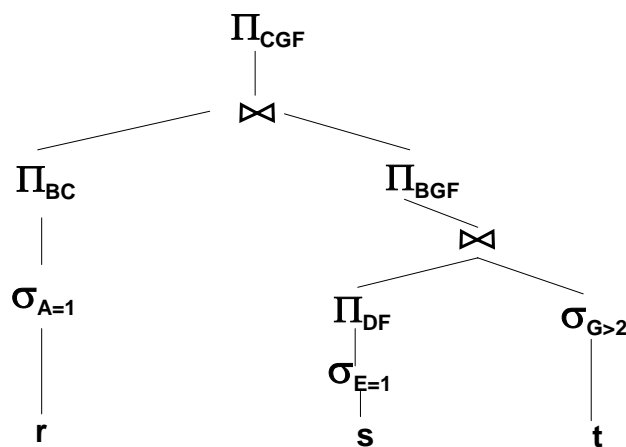
**2)  $\Pi_{r,C,F} \sigma_{(r.C>G) \wedge (A=1) \wedge (E=1) \wedge (F>3) \wedge (G>2) \wedge (s.D=t.D) \wedge (r.B=s.B)}$   
 **$r \times s \times t$****

© Fabio A. Schreiber

Linguaggi per DB 52

## ESTRAZIONE DELLA CONTAIN VIEW

RICERCA DELLA MASSIMA SOTTOESPRESSIONE COMUNE (MSE)



© Fabio A. Schreiber

Linguaggi per DB 53

## CALCOLO DELLE INTERROGAZIONI MEDIANTE LA MSE

1)

$\Pi_{CG}$   
|  
 $\sigma_{G>3}$   
|  
MSE

2)

$\Pi_{CF}$   
|  
 $\sigma_{G>G\wedge F>3}$   
|  
MSE

© Fabio A. Schreiber

Linguaggi per DB 54

## CALCOLO RELAZIONALE

- **E' UNA FAMIGLIA DI LINGUAGGI**
  - FORMALI
  - DICHIARATIVI (NON PROCEDURALI)
  - PER FORMULARE INTERROGAZIONI
- **DUE TIPI PRINCIPALI**
  - CALCOLO RELAZIONALE DELLE TUPLE (TRC)
  - CALCOLO RELAZIONALE DEI DOMINI (DRC)
- **HA UN POTERE ESPRESSIVO PARI A QUELLO DELL'ALGEBRA RELAZIONALE**

© Fabio A. Schreiber

Linguaggi per DB 55

## DEFINIZIONE FORMALE DEL TRC

### FORMA STANDARD

- $\{t \mid p(t)\}$  (TUTTE LE TUPLE  $t$  PER LE QUALI E' VERA  $p(t)$ )
- $p(t)$  E' UNA FORMULA COSTRUITA TRAMITE ATOMI

### ATOMI

- $t \in r$
- $t_1[A_1] \text{ comp } t_2[A_2]$
- $t[A] \text{ comp } k$

### COMPOSIZIONE DI FORMULE BEN FORMATE (WFF)

- UN ATOMO E' UNA WFF
- SE  $p$  E' UNA WFF, LO SONO ANCHE  $\neg p$  E  $(p)$
- SE  $p_1$  E  $p_2$  SONO WFF, LO SONO ANCHE  $p_1 \wedge p_2$ ,  $p_1 \vee p_2$ ,  $p_1 \Rightarrow p_2$
- SE  $p$  E' UNA WFF IN CUI  $s$  E' UNA VARIABILE, LO SONO ANCHE  $\exists s \in r (p(s))$  E  $\forall s \in r (p(s))$

© Fabio A. Schreiber

Linguaggi per DB 56

## PROPRIETA' DEL TRC

- $p_1 \wedge p_2 \equiv \neg(\neg p_1 \vee \neg p_2)$  (DE MORGAN)
- $\forall t \in r (p(t)) \equiv \neg \exists t \in r (\neg p(t))$
- $p_1 \Rightarrow p_2 \equiv \neg p_1 \vee p_2$
- **FORMULE UNSAFE**
  - $\{t \mid t \notin r\}$  FORNISCE UN RISULTATO INFINITO
- **SI RESTRINGE LA FORMULAZIONE DEL CALCOLO DELLE FORMULE AI DOMINII ATTIVI**
  - LE TUPLE CHE SODDISFANO UNA FORMULA POSSONO ESSERE COMPOSTE SOLAMENTE DA VALORI CHE COMPAGNO
    - ESPLICITAMENTE NELLA FORMULA
    - IN TUPLE DI RELAZIONI MENZIONATE NELLA FORMULA

© Fabio A. Schreiber

Linguaggi per DB 57

## ESEMPIO DI TRC

**NOME DEGLI STUDENTI CHE HANNO PRESO 30 IN  
“matematica”**

$$\{ t \mid \exists t1 \in \text{studente}, \exists t2 \in \text{esame}, \exists t3 \in \text{corso} \\ ( t[\text{NOME}] = t1[\text{NOME}] \wedge \\ t1[\text{MATR}] = t2[\text{MATR}] \wedge \\ t2[\text{C-CORSO}] = t3[\text{C-CORSO}] \wedge \\ t2[\text{VOTO}] = 30 \wedge \\ t3[\text{TITOLO}] = \text{“matematica”}) \}$$

© Fabio A. Schreiber

Linguaggi per DB 58

## ESEMPIO DI TRC

**MATRICOLA DEGLI STUDENTI CHE HANNO  
SOSTENUTO “matematica” MA NON “basi di dati”**

$$\{ t \mid \exists t1 \in \text{esame}, \exists t2 \in \text{corso} \\ ( t[\text{MATR}] = t1[\text{MATR}] \wedge \\ t1[\text{C-CORSO}] = t2[\text{C-CORSO}] \wedge \\ t2[\text{TITOLO}] = \text{“matematica”}) \wedge \\ \neg(\exists t3 \in \text{esame}, \exists t4 \in \text{corso} \\ (t[\text{MATR}] = t3[\text{MATR}] \wedge \\ t3[\text{C-CORSO}] = t4[\text{C-CORSO}] \wedge \\ t4[\text{TITOLO}] = \text{“basi di dati”})) \}$$

© Fabio A. Schreiber

Linguaggi per DB 59

## EQUIVALENZA TRA ALGEBRA E TRC

E' SUFFICIENTE MOSTRARE CHE SI POSSONO REALIZZARE I 5 OPERATORI FONDAMENTALI

- SELEZIONE

$$\sigma_{A=1} \Leftrightarrow \{ t \mid \exists t1 \in r ( t1[A] = 1 \quad t=t1 ) \}$$

- PROIEZIONE

$$\Pi_{AC} r \Leftrightarrow \{ t \mid \exists t1 \in r ( t[A,C] = t1[A,C] ) \}$$

- PRODOTTO CARTESIANO

$$r(A,B,C) \times s(D,E,F) \Leftrightarrow \{ t \mid \exists t1 \in r, \exists t2 \in s \\ ( t[A,B,C] = t1[A,B,C] \wedge t[D,E,F] = t2[D,E,F] ) \}$$

© Fabio A. Schreiber

Linguaggi per DB 60

## EQUIVALENZA TRA ALGEBRA E TRC

- UNIONE

$$r \cup s \Leftrightarrow \{ t \mid \exists t1 \in r \exists t2 \in s ( t=t1 \vee t=t2 ) \}$$

- DIFFERENZA

$$r - s \Leftrightarrow \{ t \mid \exists t1 \in r \quad t=t1 \wedge \neg(\exists t2 \in s, t=t2) \}$$

- ESEMPIO DI JOIN

$$r(A,C) \bowtie_{A=B} s(B,D) \Leftrightarrow \{ t \mid \exists t1 \in r, \exists t2 \in s \\ ( t[A,C] = t1[A,C] \wedge t[B,D] = t2[B,D] \wedge t[A]=t[B] ) \}$$

© Fabio A. Schreiber

Linguaggi per DB 61

$\sigma$   $\Pi$   $\cup$   $-$   $\times$   $\cap$   $\bowtie$   $\bowtie$   $\div$