Basi di Dati

prof. Letizia Tanca

Linguaggi formali di interrogazione per il Modello Relazionale dei Dati

Linguaggi di interrogazione

Ricordiamo:

- Permettono di trovare un dato basandosi **sulle sue proprietà.**
- Permettono di trovare dati basandosi su **confronti tra i contenuti** di più tabelle.

CLASSIFICAZIONE DEI LINGUAGGI

- LINGUAGGI **FORMALI**
  - Algebra relazionale
  - Calcolo relazionale
    - Delle tuple
    - Dei domini
  - Datalog
- LINGUAGGI "**COMMERCIALI**"
  - SQL (Structured Query Language)
  - QUEL
  - QBE (Query By Example)

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 le istanze dei dati
  - per l'interrogazione dei dati

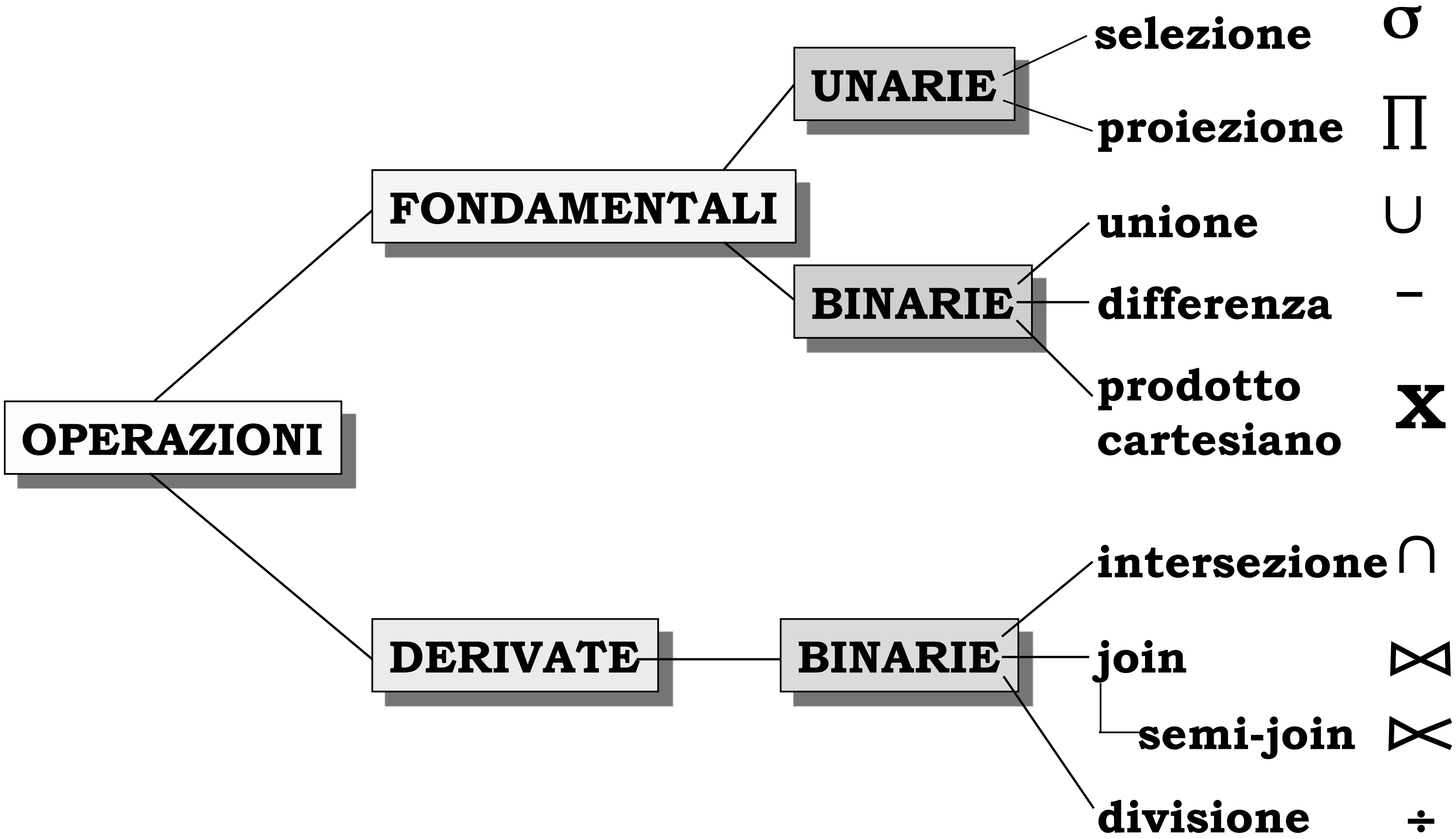
ALGEBRA RELAZIONALE

E' UN LINGUAGGIO
- formale
- funzionale
- per formulare interrogazioni

BASATO SU
- 5 operazioni fondamentali

DEFINITO DA E. CODD NEL 1970

OPERAZIONI DELL'ALGEBRA RELAZIONALE



ESEMPIO

studente

MATR.	NOME	CITTA'	CL
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

ESEMPIO DI SELEZIONE

$\sigma_{\texttt{nome = "Paola"}}$ **STUDENTE**

e' una tabella con

- **SCHEMA**

  lo stesso di STUDENTE (stesso grado)

- **ISTANZA**

  le tuple di STUDENTE che soddisfano il predicato di selezione (cardinalita' ≤)

MATR.	NOME	CITTA'	CL
415	Paola	Torino	Inf

SINTASSI DELLA SELEZIONE

σ_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

=, ≠, <, ≤, >, ≥

TERMINE

- **COSTANTE, ATTRIBUTO**
- **ESPRESSIONE ARITMETICA DI COSTANTI E ATTRIBUTI**

ESEMPIO DI SELEZIONE

$\sigma_{(\text{CITTÀ} = \text{“Torino”}) \vee ((\text{CITTÀ} = \text{“Roma”}) \wedge \neg (\text{CORSO}=\text{“Log”}))}$ **studente**

studente

MATR.	NOME	CITTA’	CL
~~123~~	~~Carlo~~	~~Bologna~~	~~Inf~~
415	Paola	Torino	Inf
~~702~~	~~Antonio~~	~~Roma~~	~~Log~~

Esempio di Proiezione

Π $_{\text{NOME, CL}}$ **studente**

e' una tabella con

- **SCHEMA**

  solo gli attributi NOME e CL (grado $\leq$)

- **ISTANZA**

  la restrizione delle tuple di studente sugli attributi NOME e CL (cardinalita' $\leq$)

NOME	CL
Carlo	Inf
Paola	Inf
Antonio	Log

SINTASSI DELLA PROIEZIONE

$$\Pi_{\text{ATTR1,,,,,ATTRn}}\ \mathbf{R}$$

- **formalmente la proiezione elimina i duplicati**
- **nei sistemi in uso l'eliminazione dei duplicati va richiesta esplicitamente**

$$\Pi_{\text{CL}}\ \mathbf{STUDENTE}$$

CL
Inf
Log

Esempio di schema di base di dati

PRESIDENTI (NOME-P, DATA-N, DATA-M, PARTITO, STATO, NOME-M)

CONGRESSI (# CONGRESSO, %S-REP, %C-REP, %S-DEM, %C-DEM)

AMMINISTRAZIONI (# AMMIN, DATA-IN, VICE-PRES, DATA-N-VP, NOME-P, DATA-N-P)

ELEZIONI (ANNO, VOTI-PRES, NOME-P, DATA-N-PRES, NOME-PERD,

DATA-N-PERD, VOTI-PERD)

STATI (STATO, POPOLAZ, # AMMIN.)

PRESID-CONGR (NOME-P, DATA-N, # CONGR)

Interrogazioni

- *Trovare l'anno di nascita del presidente J.F. Kennedy*
- *Trovare le persone che sono state presidenti OPPURE vicepresidenti in amministrazioni inaugurate dopo il 1880*
- *Trovare le persone che sono state presidenti E ANCHE vicepresidenti in qualche amministrazione inaugurata dopo il 1880*
- *Trovare le persone che sono state presidenti MA MAI vicepresidenti in amministrazioni inaugurate dopo il 1880*

OPERAZIONI NON ALGEBRICHE

- **ASSEGNAMENTO**

  Serve a dare un nome al risultato di un'operazione algebrica

  - **TORINESI = $\sigma_{(\text{CITTÀ} = \text{“Torino”})}$ STUDENTI**
  - **INFORMATICI = $\sigma_{(\text{CL} = \text{“inf”})}$ STUDENTI**

- **RIDENOMINAZIONE**

  Permette di modificare i nomi degli attributi

  - $\rho_{\text{genitore} \leftarrow\text{-- padre}}$ **PATERNITÀ**
  - $\rho_{\text{genitore} \leftarrow\text{-- madre}}$ **MATERNITA**

in questo modo e' possibile rendere compatibili domini originariamente diversi

Le operazioni INSIEMISTICHE: UNIONE, DIFFERENZA e INTERSEZIONE

- TABELLA1 ∪ TABELLA2
- TABELLA1 − TABELLA2
- TABELLA1 ∩ TABELLA2

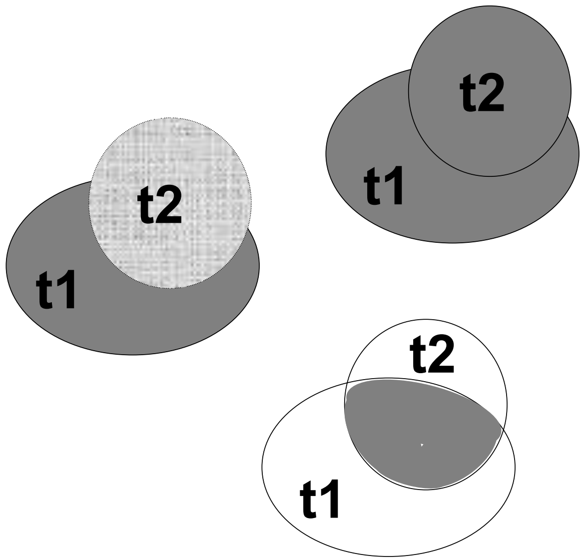


TABELLA1 e **TABELLA2** devono essere *compatibili:*

- stesso grado
- domini ordinatamente dello stesso tipo
- vale: **r ∩ s = r - (r-s)** perciò l'intersezione è un operatore **DERIVATO**

ESEMPIO DI UNIONE

INFORMATICI ∪ TORINESI

e' una tabella con

- **SCHEMA**
  lo stesso di INFORMATICI

- **ISTANZA**
  l'unione delle tuple di INFORMATICI e TORINESI

MATR.	NOME	CITTA'	CL
123	Carlo	Bologna	Inf
415	Paola	Torino	Inf

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'	CL
123	Carlo	Bologna	Inf

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'	CL
415	Paola	Torino	Inf

PRODOTTO CARTESIANO
(*quasi* un' operazione insiemistica)

R × S

e' una tabella con

- **SCHEMA**

  tutti gli attributi di R e di S

  (grado (R × S) = grado (R) + grado (S))

- **ISTANZA**

  tutte le possibili coppie di tuple di R e di S

  (card (R × S) = card (R) * card (S))

ESPRESSIONI ALGEBRICHE

- Concatenazione di piu' operazioni algebriche

- Esprimono interrogazioni in modo formale

- Consentono di estrarre informazioni dalla base di dati

UN ALTRO ESEMPIO DI BASE DI DATI

- **CORRENTISTA (NOMECORR, VIA, CITTA’)**
- **CLIENTE (CORRENTISTA, IMPIEGATO)**
- **FILIALE (NOMEFIL, PATRIMONIO, CITTA’)**
- **DEPOSITI (FILIALE, NUM-DEP, CORRENTISTA, SALDO)**
- **PRESTITI (FILIALE, NUM-PRES, CORRENTISTA, QUANTITA’)**

Esempio Di Uso Di Espressioni Algebriche

cliente

CORRENTISTA	IMPIEGATO
Rossi	Ghezzi
Brambilla	Ferrari
Ferrari	Ferrari

correntista

NOMECOR	VIA	CITTA'
Rossi	Dante	MI
Bianchi	Verdi	RM
Neri	Po	TO
Brambilla	Orefici	MI
Ferrari	Verdi	TO

r_1 = cliente X correntista

CLIENTE. CORRENT.	CLIENTE. IMPIEGATO	CORRENT. NOMECOR	CORRENT. VIA	CORRENT. CITTA'
Rossi	Ghezzi	Rossi	Dante	MI
Rossi	Ghezzi	Bianchi	Verdi	RM
Rossi	Ghezzi	Neri	Po	TO
Rossi	Ghezzi	Brambilla	Orefici	MI
Rossi	Ghezzi	Ferrari	Verdi	TO
Brambilla	Ferrari	Rossi	Dante	MI
Brambilla	Ferrari	Bianchi	Verdi	RM
Brambilla	Ferrari	Neri	Po	TO
Brambilla	Ferrari	Brambilla	Orefici	MI
Brambilla	Ferrari	Ferrari	Verdi	TO
Ferrari	Ferrari	Rossi	Dante	MI
Ferrari	Ferrari	Bianchi	Verdi	RM
Ferrari	Ferrari	Neri	Po	TO
Ferrari	Ferrari	Brambilla	Orefici	MI
Ferrari	Ferrari	Ferrari	Verdi	TO

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

r_4

CLIENTE. CORRENT.	CORRENT. CITTA'
Brambilla	MI
Ferrari	TO

r_1 = cliente X correntista

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

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

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

ESEMPIO DI JOIN

$$\text{STUDENTE} \bowtie_{\text{STUDENTE.MATR=ESAME.MATR}} \text{ESAME}$$

e’ una tabella con

- **SCHEMA**

  la concatenazione degli schemi di STUDENTE ed ESAME

- **ISTANZA**

  le tuple ottenute concatenando quelle tuple di STUDENTE ed ESAME che soddisfano il predicato

MATR.	NOME	CITTA’	CL	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

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 booleana di predicati semplici
    - $ATTR_1$ $COMP$ $ATTR_2$
      - $ATTR_1 \in R_1$
      - $ATTR_2 \in R_2$
      - ***COMP*** **=, ≠, <, ≤, >, ≥**
  - attributi omonimi sono resi non ambigui usando la notazione puntata: $R_I.ATTR_J$

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’	CL	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

SINTASSI DEL JOIN NATURALE

$$S \bowtie R = \Pi_{sch(R) \cup sch(S)} (R \bowtie_{R.A1=S.A1 \wedge , \ldots, \wedge R.AN=S.AN} S)$$

- $\{A1, \ldots, An\} = sch(R) \cap sch(S)$
- $sch(R) \cap sch(S) = \Phi \Rightarrow \bowtie \equiv \times$
- $\bowtie$ E' ASSOCIATIVO

JOIN NATURALE DI TRE TABELLE

studente $\bowtie$ esame $\bowtie$ corso

MATR.	NOME	CITTA'	CL	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

SEMI-JOIN

STUDENTE $\ltimes_{\text{STUDENTE.MATR=ESAME.MATR}}$ ESAME

- **e' equivalente alla seguente espressione**

$\Pi_{\text{STUDENTE.*}}$ **(STUDENTE** $\bowtie_{\text{STUDENTE.MATR=ESAME.MATR}}$ **ESAME)**

- produce un sottoinsieme di STUDENTE:

  solo gli studenti che hanno dato almeno un esame

studente

MATR.	NOME	CITTA'	CL
123	Carlo	Bologna	Inf
702	Antonio	Roma	Log

SEMI-JOIN NATURALE

STUDENTE $\ltimes$ ESAME

- E' EQUIVALENTE ALLA SEGUENTE ESPRESSIONE

$\Pi_{STUDENTE.*}$ ***(STUDENTE*** $\bowtie$ ***ESAME)***

- in questo esempio il risultato e' uguale al caso precedente

studente

MATR.	NOME	CITTA'	CL
123	Carlo	Bologna	Inf
702	Antonio	Roma	Log

DIVISIONE

$$R \div S$$

- e' equivalente alla seguente espressione

$$\Pi_{R\text{-}S}\, R - \Pi_{R\text{-}S}\, ((\Pi_{R\text{-}S}\, R \times S) - R)$$

con Schema(S) $\subseteq$ Schema(R)

- e' utile per esprimere interrogazioni che contengono il quantificatore universale $\forall$: *seleziona le tuple di R che contengono TUTTE le tuple di S*

r

X	Y	W	Z
a	b	c	d
a	b	e	f
b	c	e	f
e	d	c	d
e	d	e	f
a	b	d	e

$\div$

s

W	Z
c	d
e	f

=

X	Y
a	b
e	d

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

  $r_1 = \rho_{\textbf{filiale <-- nome}} \Pi_{\text{NOME}} (\sigma_{\text{CITTA' =“Milano”}} \text{filiale})$

- 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

  $r_3 = r_2 \div r_1$

PRECEDENZA DEGLI OPERATORI

- La precedenza funziona circa come in algebra elementare
- gli operatori unari (come il meno unario in algebra) hanno priorità massima
- il join e il prodotto cartesiano sono "come il prodotto"
- per il resto meglio mettere le parentesi

ESEMPIO DI ESPRESSIONE COMPLESSA

ESTRARRE I NOMI DEGLI STUDENTI CHE HANNO SOSTENUTO GLI ESAMI DI *informatica* E *matematica* LO STESSO GIORNO

Π_{NOME} (STUDENTE $\bowtie$ (ESAME $\ltimes$ $\sigma_{\text{TITOLO=“INFORMATICA”}}$ CORSO)

$_{\text{MATR=MATRMAT} \wedge \text{ DATA=DATAMAT}}$ $\rho_{\text{MATRMAT,DATAMAT <--- MATR,DATA}}$ $\Pi_{\text{MATR,DATA}}$

(ESAME $\ltimes$ $\sigma_{\text{TITOLO=“MATEMATICA”}}$ CORSO))

$\bowtie$

Esempio di schema di base di dati

PRESIDENTI (NOME-P, DATA-N, DATA-M, PARTITO, STATO, NOME-M)

CONGRESSI (# CONGRESSO, %S-REP, %C-REP, %S-DEM, %C-DEM)

AMMINISTRAZIONI (# AMMIN, DATA-IN, VICE-PRES, DATA-N-VP, NOME-P, DATA-N-P)

ELEZIONI (ANNO, VOTI-PRES, NOME-P, DATA-N-PRES, NOME-PERD, DATA-N-PERD, VOTI-PERD)

STATI (STATO, POPOLAZ, # AMMIN.)

PRESID-CONGR (NOME-P, DATA-N, # CONGR)

Interrogazioni

- ***Trovare l'anno di nascita del presidente J.F. Kennedy***
- ***Trovare gli anni in cui è stato eletto un presidente repubblicano proveniente dall'Illinois***
- ***Trovare i numeri di congressi presieduti dal presidente eletto nel 1955***
- ***Trovare i perdenti delle elezioni vinte da qualche presidente di nome Roosevelt***
- ***Trovare i nomi delle mogli dei presidenti provenienti dalla California eletti dopo il 1960***
- ***Trovare le persone che sono state presidenti OPPURE vicepresidenti in amministrazioni inaugurate dopo il 1880***
- ***Trovare le persone che sono state presidenti E ANCHE vicepresidenti in qualche amministrazione inaugurata dopo il 1880***
- ***Trovare le persone che sono state presidenti MA MAI vicepresidenti in amministrazioni inaugurate dopo il 1880***

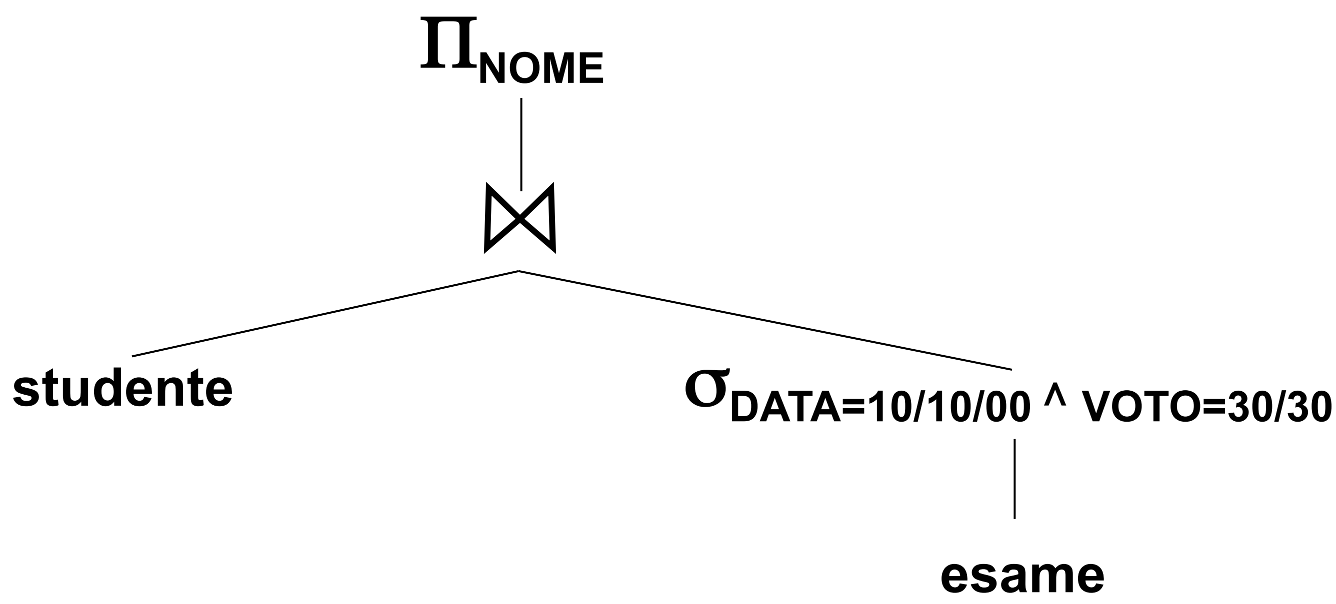
RAPPRESENTAZIONE DELLE ESPRESSIONI TRAMITE ALBERI

Ogni espressione dell'algebra relazionale può essere rappresentata mediante un albero sintattico che esprime l'ordine di valutazione degli operatori

ESEMPIO DI ALBERO

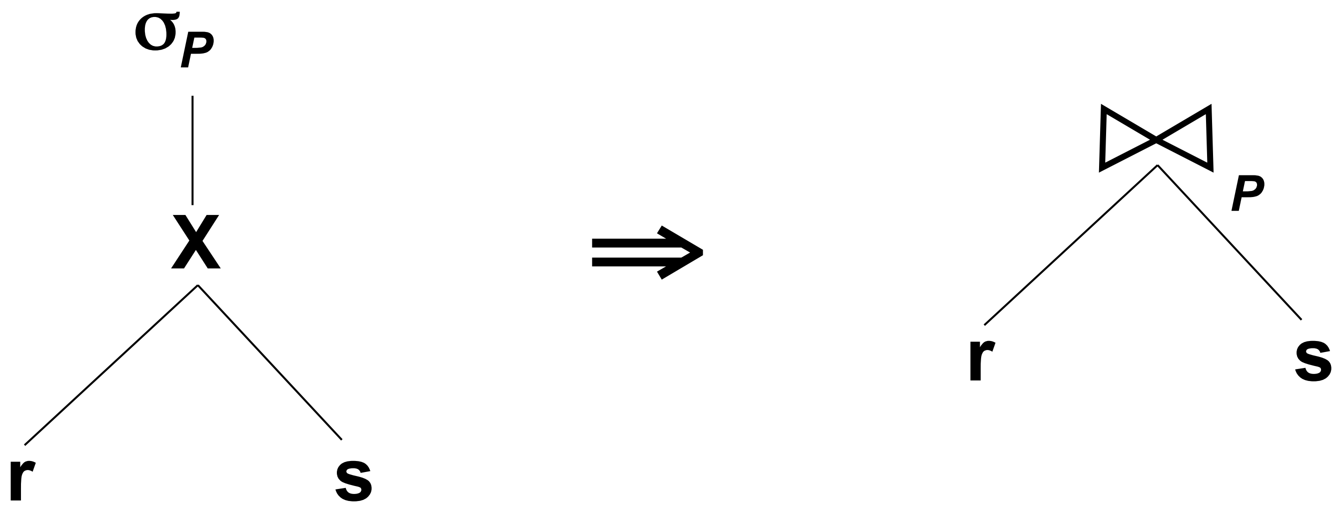
Nomi degli studenti che hanno sostenuto un esame il 10/10/00 con voto pari a 30/30

$$\Pi_{NOME}\ (\mathbf{STUDENTE} \bowtie \sigma_{DATA=10/10/00\ \wedge\ VOTO=30/30}\ \mathbf{ESAME})$$



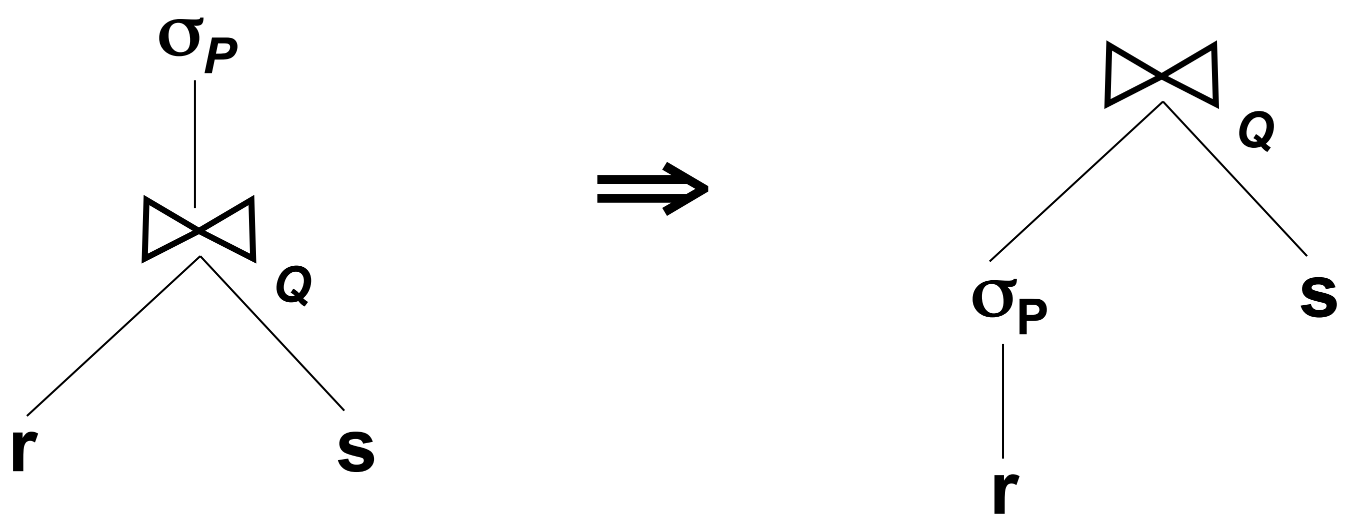
EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

1 - ELIMINAZIONE DEI PRODOTTI CARTESIANI



EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

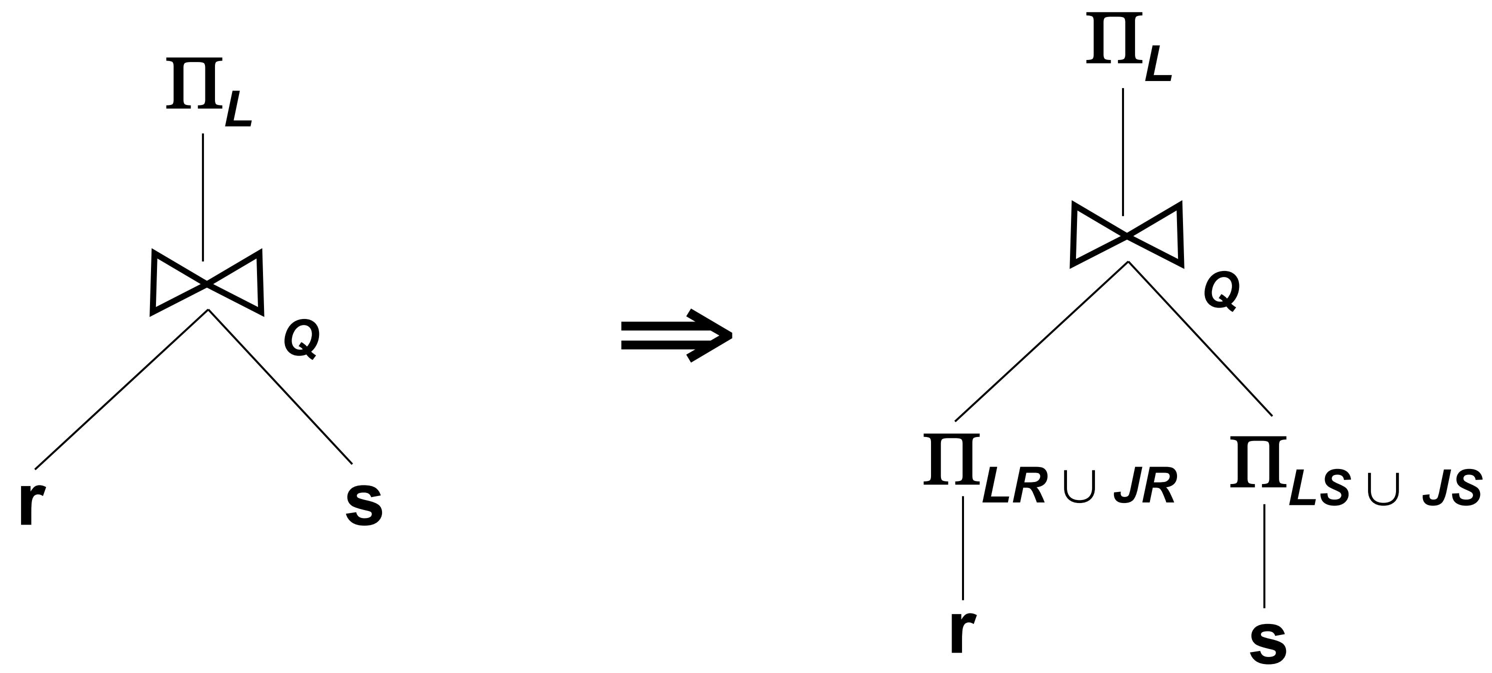
2 - PUSH DELLA SELEZIONE RISPETTO AL JOIN



VALE SE P SI APPLICA AI SOLI ATTRIBUTI DI R:
altrimenti possiamo suddividere P nei suoi *"congiunti"*

EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

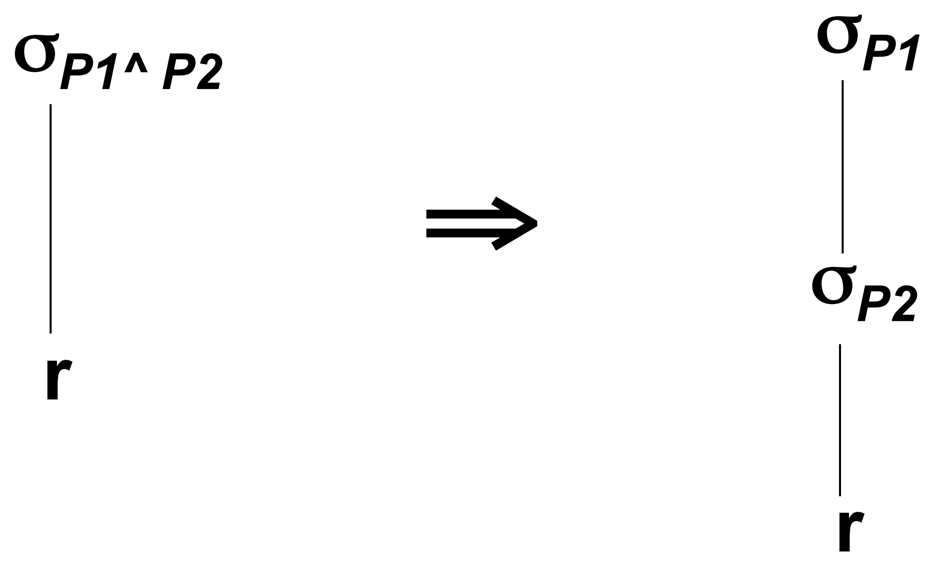
3 - PUSH DELLA PROIEZIONE RISPETTO AL JOIN



- ***J=JR U JS*** **SONO GLI ATTRIBUTI** ***"COINVOLTI NEL JOIN"***
- ***LR*** **=** ***L*** **- SCHEMA(s)**
- ***LS*** **=** ***L*** **- SCHEMA(r)**

EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

4 - ATOMIZZAZIONE DELLE SELEZIONI



EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

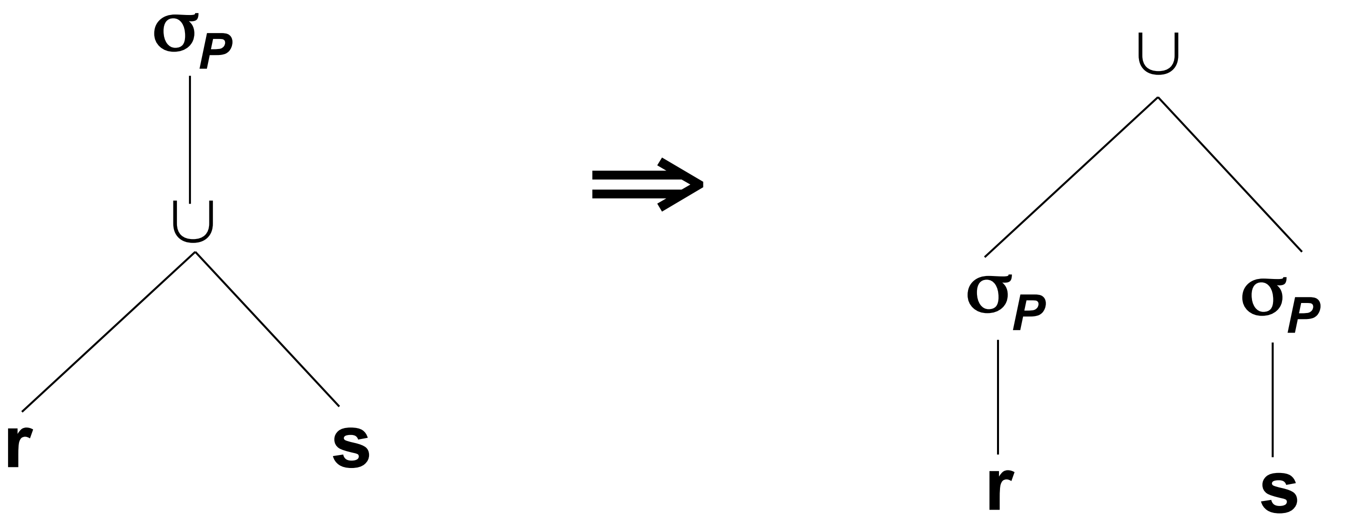
5 - IDEMPOTENZA DELLA PROIEZIONE

$$\begin{array}{c} \Pi_{L} \\ | \\ \mathbf{r} \end{array} \quad \Rightarrow \quad \begin{array}{c} \Pi_{L} \\ | \\ \Pi_{L2} \\ | \\ \mathbf{r} \end{array}$$

$$L \subseteq L2$$

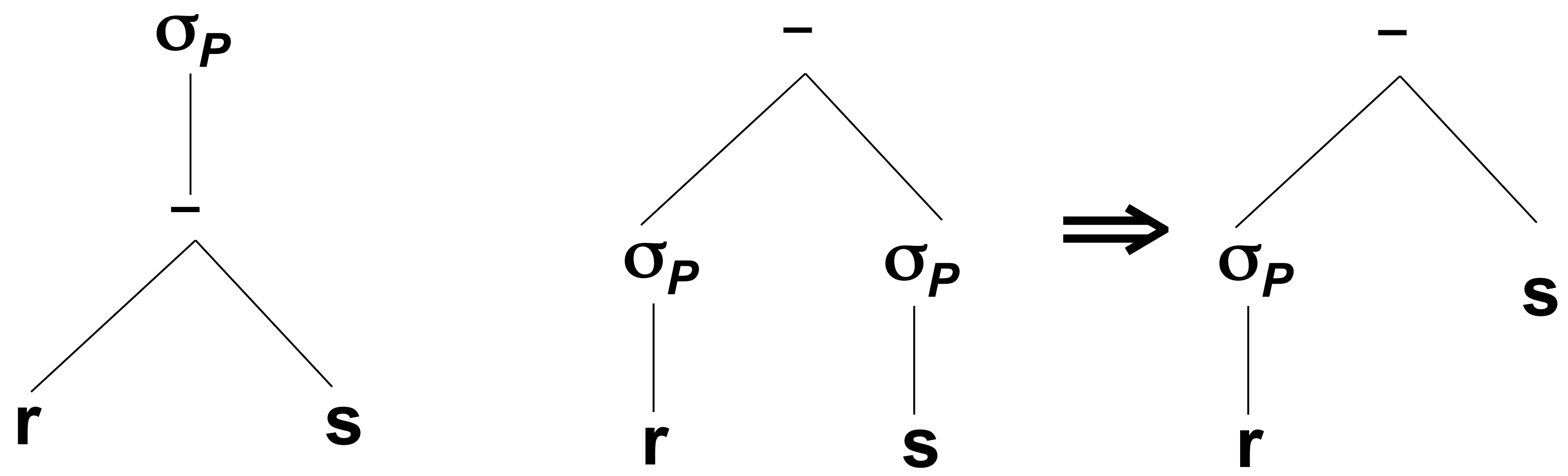
EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

6 - PUSH DELLA SELEZIONE RISPETTO ALL'UNIONE



EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

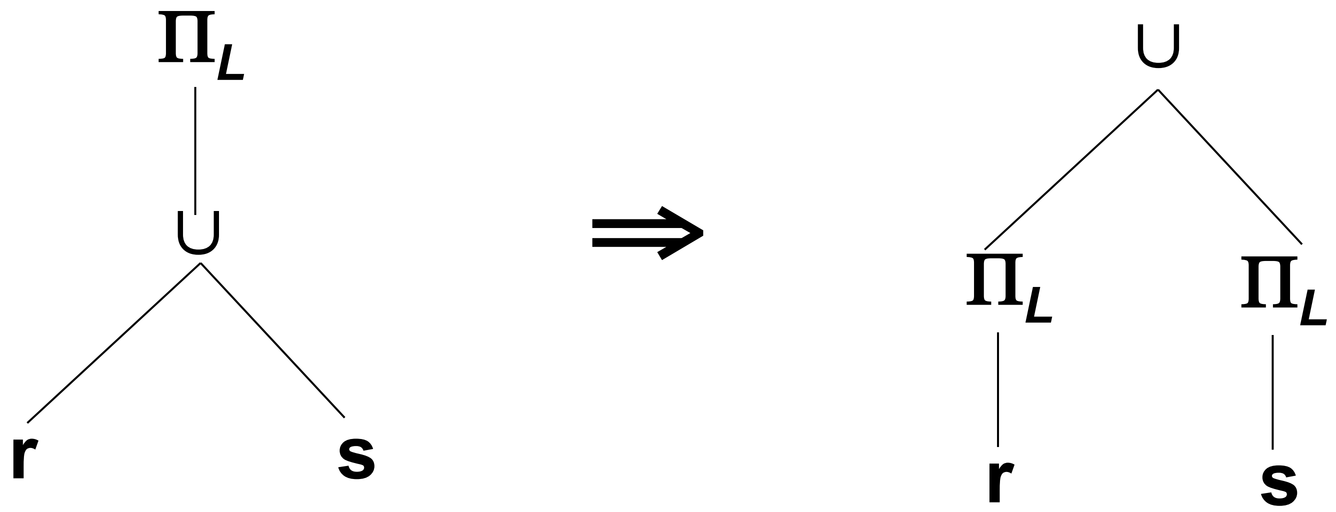
7 - PUSH DELLA SELEZIONE RISPETTO ALLA DIFFERENZA



Nota: la selezione commuta con tutte e tre le operazioni insiemistiche

EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

8 - PUSH DELLA PROIEZIONE RISPETTO ALL'UNIONE



Nota: la proiezione NON commuta con le altre operazioni insiemistiche

EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

La proiezione NON commuta con le altre operazioni insiemistiche

- **Differenza:** $R_1(A,B)$, $R_2(A,B)$

  $\Pi_A(R_1 - R_2)$ è diverso da $\Pi_A R_1 - \Pi_A R_2$

  (v. slide successiva)

- **Intersezione:** ovviamente non commuta *perché si deriva da unione e differenza*

La proiezione non commuta con la differenza

Clienti filiale A

CORRENTISTA	IMPIEGATO
Rossi	Ghezzi
Brambilla	Ferrari
Verdi	Ferrari

$\Pi_{correntista}$ (Clienti Filiale A – Clienti Filiale B)=
{(Rossi),(Verdi)}

Clienti filiale B

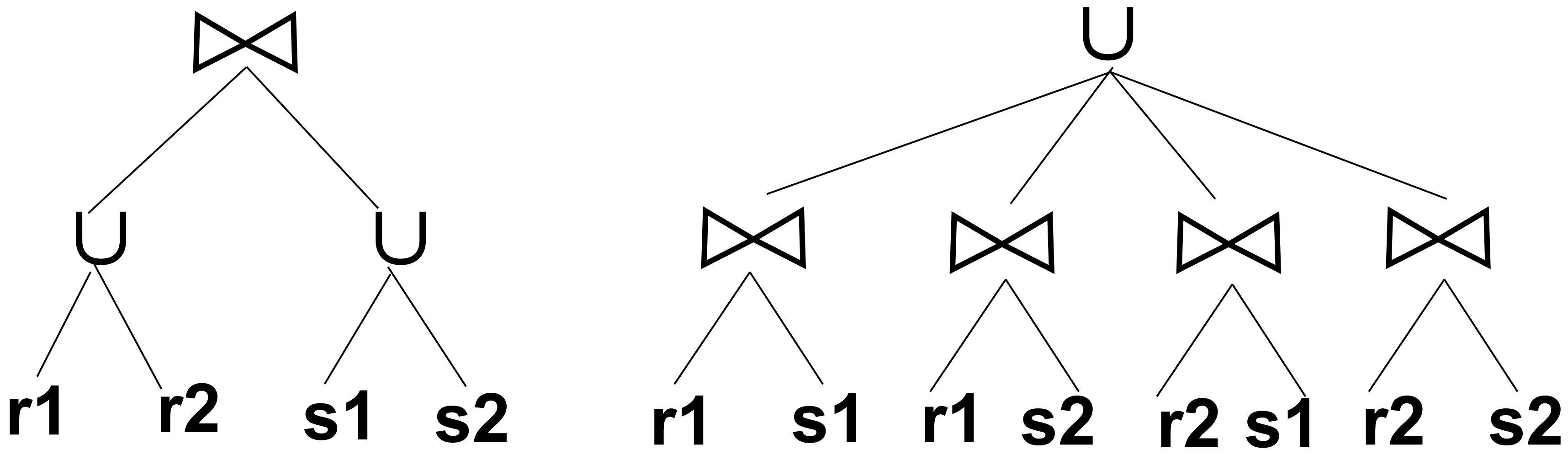
CORRENTISTA	IMPIEGATO
Rossi	Amato
Brambilla	Ferrari
Verdi	Lupo

$\Pi_{correntista}$ Clienti Filiale A –

$\Pi_{correntista}$ Clienti Filiale B = $\emptyset$

EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

9 – distributività di JOIN rispetto a UNIONE



FORMULE UTILI

$r \bowtie r = r$

$r \cup r = r$

$r - r = \emptyset$

$r \bowtie \sigma_P r = \sigma_P r$

$r \cup \sigma_P r = r$

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

$\sigma_{P1} r \bowtie \sigma_{P2} r = \sigma_{P1 \wedge P2} r$

$\sigma_{P1} r \cup \sigma_{P2} r = \sigma_{P1 \vee P2} r$

$\sigma_{P1} r - \sigma_{P2} r = \sigma_{P1 \wedge \neg P2} r$

$\sigma_P \emptyset = \emptyset$

$\Pi_L \emptyset = \emptyset$

$r \times \emptyset = \emptyset$

$r \cup \emptyset = r$

$r - \emptyset = r$

$\emptyset - r = \emptyset$

$r \bowtie \emptyset = \emptyset$

$r \cap \emptyset = \emptyset$

N.B. le formule che valgono per il join <u>valgono anche per l'intersezione</u>

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 atomizzazione ed 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)

ESEMPIO DI OTTIMIZZAZIONE

DATO LO SCHEMA

R(A,B,C)

S(C,D,E)

T(D,E,F,G)

SI OTTIMIZZI L' ESPRESSIONE

$$\Pi_{B}\, \sigma_{(R.C=S.C) \wedge (S.D=T.D) \wedge (R.A=1) \wedge (R.B>T.F)} (R \times S \times T)$$

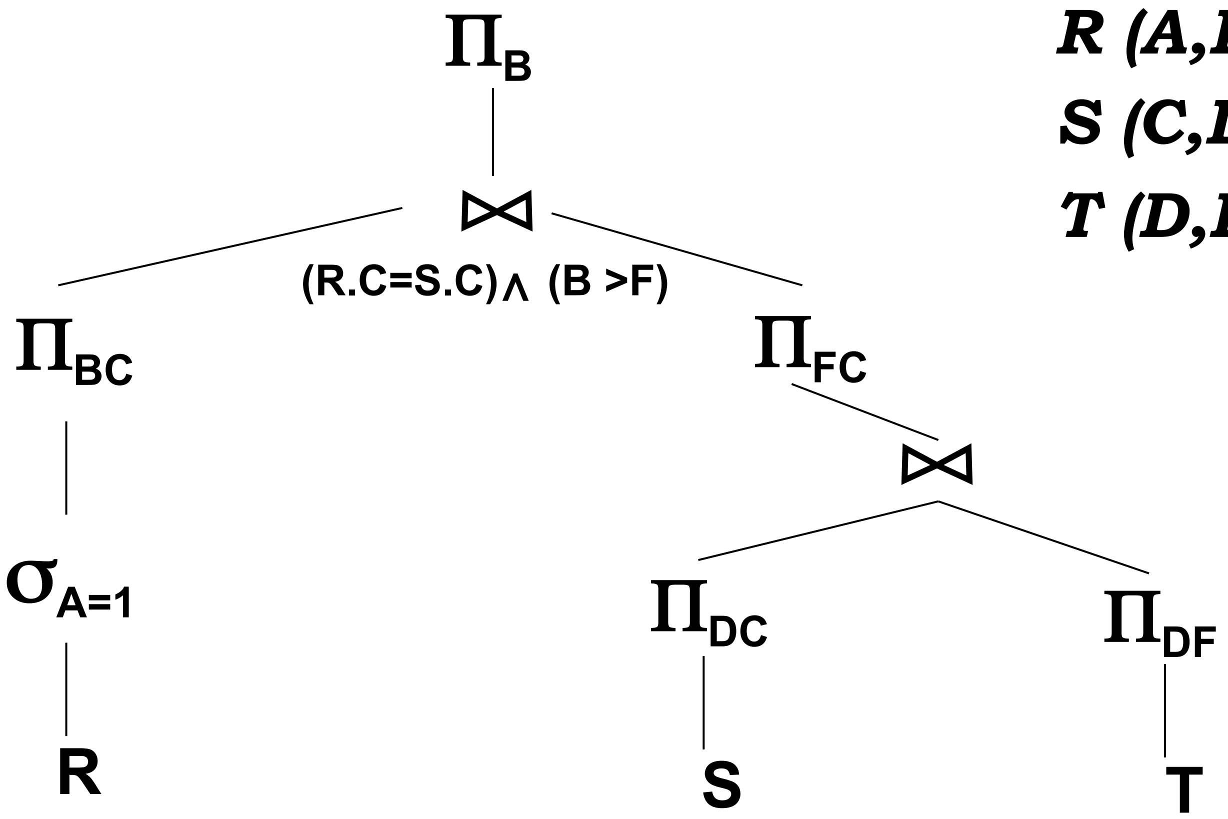
ESEMPIO DI OTTIMIZZAZIONE

$$\Pi_B \, \sigma_{(R.C=S.C) \wedge (S.D=T.D) \wedge (R.A=1) \wedge (R.B>T.F)} (R \times S \times T)$$

R (A,B,C)

S (C,D,E)

T (D,E,F,G)



ESERCIZI

- Disegnare gli alberi delle interrogazioni sui presidenti
- Ottimizzarli

Esempio di schema di base di dati

PRESIDENTI (NOME-P, DATA-N, DATA-M, PARTITO, STATO, NOME-M)

CONGRESSI (# CONGRESSO, %S-REP, %C-REP, %S-DEM, %C-DEM)

AMMINISTRAZIONI (# AMMIN, DATA-IN, VICE-PRES, DATA-N-VP, NOME-P, DATA-N-P)

ELEZIONI (ANNO, VOTI-PRES, NOME-P, DATA-N-PRES, NOME-PERD, DATA-N-PERD, VOTI-PERD)

STATI (STATO, POPOLAZ, # AMMIN.)

PRESID-CONGR (NOME-P, DATA-N, # CONGR)

Interrogazioni

- ***Trovare l'anno di nascita del presidente J.F. Kennedy***
- ***Trovare gli anni in cui è stato eletto un presidente repubblicano proveniente dall'Illinois***
- ***Trovare i numeri di congressi presieduti dal presidente eletto nel 1955***
- ***Trovare i perdenti delle elezioni vinte da qualche presidente di nome Roosevelt***
- ***Trovare i nomi delle mogli dei presidenti provenienti dalla California eletti dopo il 1960***
- ***Trovare le persone che sono state presidenti OPPURE vicepresidenti in amministrazioni inaugurate dopo il 1880***
- ***Trovare le persone che sono state presidenti E ANCHE vicepresidenti in qualche amministrazione inaugurata dopo il 1880***
- ***Trovare le persone che sono state presidenti MA MAI vicepresidenti in amministrazioni inaugurate dopo il 1880***

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 suppongono:

- uniforme la distribuzione dei valori nei domini
- non correlati i dati relativi ai predicati nelle interrogazioni

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 SE I VALORI SONO UNIFORM. DISTRIBUITI

DIMENSIONI DEI RISULTATI INTERMEDI

OPERAZIONE DI SELEZIONE $\sigma_P R$

CARD(risultato) = S_P * CARD(R)

S_P = SELETTIVITA' DEL PREDICATO

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

SIZE(risultato) = SIZE(R), cioè la tupla del risultato ha la stessa dimensione della tupla originaria

Inoltre, nel risultato sarà VAL(A_i) = 1 se P ≡ "A_i = v"

DIMENSIONI DEI RISULTATI INTERMEDI

OPERAZIONE DI PROIEZIONE: $\Pi_{A_{1,\ldots}A_N} R$

CARD$(\Pi_{A_{1,\ldots}A_N} R) <=$ VAL(A_1) *….* VAL(A_N)

SIZE$(\Pi_{A_{1,\ldots}A_N} R) =$ SIZE(A_1) + … + SIZE(A_N) , perché la tupla del risultato è formata dai soli attributi indicati nella proiezione