

1. Esercizi tipologia A

A. Data una query SQL/algebra relazionale:

- (a) se SQL, fornire la rappresentazione canonica in algebra relazionale
- (b) fornire una espressione algebrica equivalente a quella canonica o a quella data ma più/meno efficiente
- (c) fornire l'espressione algebrica ottimale equivalente a quella canonica o a quella data
- (d) individuare i fattori booleani nella condizione

ESERCIZIO 1

Si considerino le relazioni con schema:

PIZZA (CodP, Nome, Costo)

CLIENTE (TelC, NomeC, Via, NCiv, NInt)

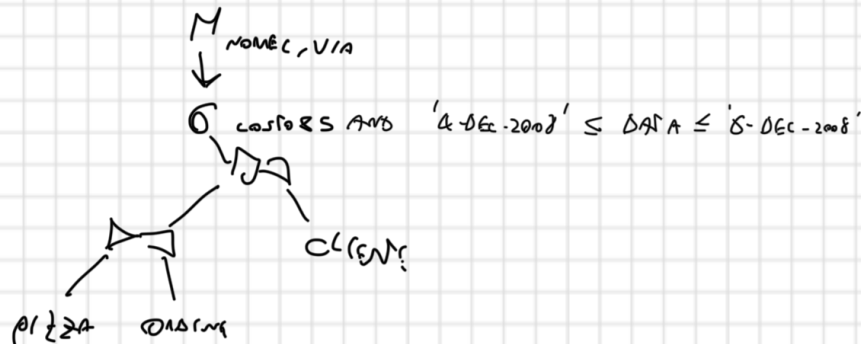
ORDINE (TelC^{CLIENTE}, Data, CodP^{PIZZA}, Qta)

e la query

```
SELECT NomeC, Via  
FROM PIZZA NATURAL JOIN ORDINE NATURAL JOIN CLIENTE  
WHERE Costo < 5 AND Data BETWEEN (DATE '4-Dec-2008' AND DATE '6-Dec-2008');
```

1. Presentare il piano logico corrispondente all'espressione algebrica canonica.

$\pi_{\text{NomeC}, \text{Via}} \left(\sigma_{\text{Costo} < 5 \text{ AND } '4\text{-Dec-2008'} \leq \text{Data} \leq '6\text{-Dec-2008'}} (\text{PIZZA} \bowtie \text{ORDINE} \bowtie \text{CLIENTE}) \right)$



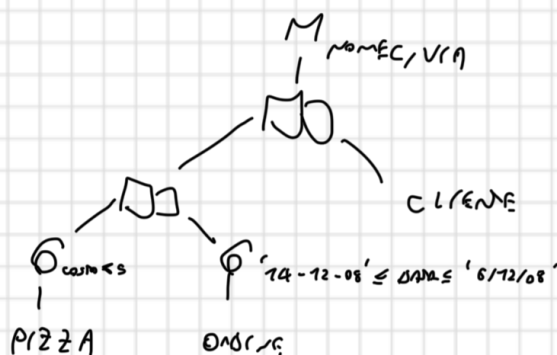
2. Individuare un piano logico equivalente al piano logico canonico ma più efficiente.

PER ESSERE EFFICIENTE ABBIAMO DUE MODI:

a) ANTICIPARE LA PROIEZIONE

b) ANTICIPARE LA SELEZIONE

ANTICIPIAMO LA SELEZIONE:



3) INDI VIGORARE I FATTORI BOOLEANI PER LA QUERY

I FATTORI BOOLEANI SONO CONDIZIONI ATOMICHE CHE SE SONO FALSE ALLORA L'ESPRESSIONE È FALSA TUTTA L'ESPRESSIONE.

IN QUESTO CASO SONO COSTO E DATA BETWEEN

4. Cosa cambierebbe relativamente ai fattori booleani se la query fosse:

```
SELECT NomeC, Via
FROM PIZZA NATURAL JOIN ORDINE NATURAL JOIN CLIENTE
WHERE Costo < 5 OR Data BETWEEN (DATE '4-Dec-2008' AND DATE '6-Dec-2008');
```

N.B. SUI FATTORI BOOLEANI

ESAME GIUGNO 2026 PTLA A

Si consideri il seguente schema relazionale:

```
IMPIEGATO(eid, didDIPARTIMENTO, sal, hobby)
DIPARTIMENTO(did, nomedip, piano, telefono)
BILANCIO(didDIPARTIMENTO, budget, uscite, entrate)
```

e la seguente interrogazione:

```
SELECT nomedip, budget
FROM IMPIEGATO NATURAL JOIN DIPARTIMENTO NATURAL JOIN BILANCIO
WHERE piano = 1 AND sal = 59000 AND hobby = "giardinaggio";
```

e le seguenti informazioni sullo schema fisico:

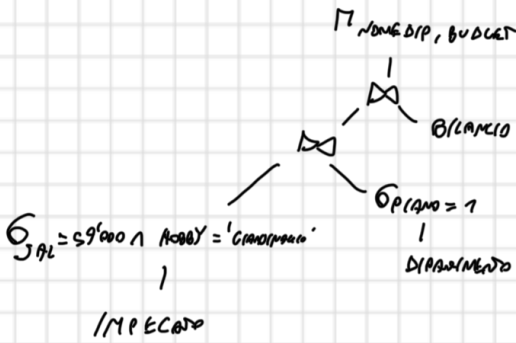
- indici ordinati (B+ alberi) su IMPIEGATO.eid, DIPARTIMENTO.piano, DIPARTIMENTO.did e BILANCIO.did; indice hash su IMPIEGATO.sal.
- i salari distinti degli impiegati sono 500, gli hobby diversi degli impiegati sono 200 e la compagnia occupa due piani dell'edificio;
- nella base di dati ci sono 50.000 impiegati e 5.000 dipartimenti (ognuno con le relative informazioni sul bilancio);

[Motivare tutte le risposte]

2. Presentare, se esiste, un piano logico di esecuzione più efficiente del piano canonico, evitando di inserire operazioni inutili.

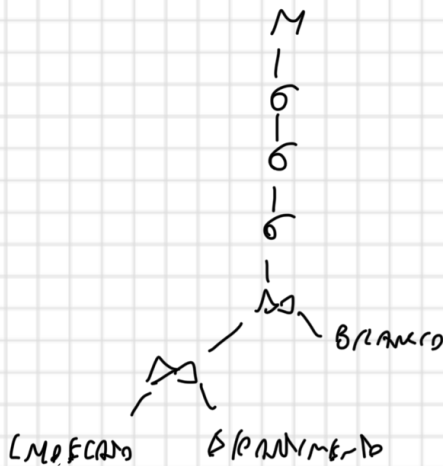
INOME DIP, BUDGET (6 piano=1 ∧ SAL=59000 ∧ HOBBY="GIARDINAGGIO") (IMPIEGATO AS DIPARTIMENTO AS BILANCIO)

PER ESSERE PIÙ EFFICIENTE XELLO DI ANTERIARE LA SELEZIONE



3. Presentare, se esiste, un piano logico di esecuzione meno efficiente del piano canonico, evitando di inserire operazioni inutili.

UN PIANO MENO ADIPICIENTE E', PER ESEMPIO, UN PIANO IN CUI SI SOSTITUISCE LA SEIBZOMF CON UNA SEQUENZA DI SEIBZOMI IN CASATA, UNA PER CIASCUNA COMBINAZIONE.



4. Supponendo che nel piano logico di esecuzione tutte le selezioni vengano eseguite prima del join mentre la proiezione venga eseguita dopo, stimare quante tuple verranno selezionate per la relazione IMPIEGATO.

PER RISPONDERE QUESTO ESERCIZIO DOBBIAMO GUARDARE LE SEGUENTI SUE:

STIMA DELLA DIMENSIONE DEL RISULTATO DELLA SELEZIONE

- Come si stima il fattore di selettività?
 - **Assunzione: uniformità di distribuzione dei valori di ogni attributo**, cioè si assume che ogni valore appaia con la stessa probabilità
 - Stima dipende dalla condizione di selezione
- ESEMPIO
- Condizione di selezione $A = v$
 - $F(A=v)$: probabilità che una tupla in input alla selezione soddisfi la condizione $A = v$
 - *Casi possibili*: tutti i possibili valori per A in R $V(A,R)$
 - *Casi favorevoli*: 1 (solo v)
 - Quindi $F(A=v) = 1 / V(A,R)$

$$\sigma_A = v$$

STIMA DELLA DIMENSIONE DEL RISULTATO DELLA SELEZIONE

- $F(A=v) = 1/V(A,R)$
- $F(A \text{ IN } (v_1, v_2, \dots, v_N)) = N * F(A=v)$
- $F(A > v) = (\text{Max}(A,R) - v) / (\text{Max}(A,R) - \text{Min}(A,R))$
- $F(A < v) = (v - \text{Min}(A,R)) / (\text{Max}(A,R) - \text{Min}(A,R))$
- $F(A_1 = A_2) = 1 / \text{MAX}(V(A_1,R), V(A_2,R))$
- $F(C_1 \text{ AND } C_2) = F(C_1) * F(C_2)$
- $F(C_1 \text{ OR } C_2) = F(C_1) + F(C_2) - F(C_1) * F(C_2)$
- $F(\text{NOT } C) = 1 - F(C)$

IMPRECATO NELLA SELEZIONE (ALBANO) HA UN AND. QUINDI USIAMO LA FORMULA:

$$F(C1 \text{ AND } C2) = F(C1) * F(C2) = F(SAL = 59'000 \text{ AND } HOBBY = 'GRANDINACCI') = F(SAL = 59'000) * F(HOBBY = 'GRANDINACCI') =$$

NOTIAMO CHE ADESSO ABBIAMO $F = \text{quanta}$ A QUESTO PUNTO USIAMO: $F(A=V) = 1/V(A,R)$ QUINDI:

$1/500 * 1/200$. PENSARE' SOD E 200? LA CONVERSIONE DI SELEZIONE SAL = 59'000 SI RIFERISCE

ALL'AMBITO SAL. DAL TESTO CHE SCIAMO CHE CI SONO 500 SALARI DISTINTI QUINDI $V(SAL, IMPRECATO) = 500$.

PER STIMARE IL NUMERO DI TUPLE, MOLTIPLICIAMO IL FATTORE DI SELEZIONABILITA' PER IL NUMERO DI TUPLE DELLA TABELLA IMPRECATO:

$$50'000 * 1/100000 = 1/2 \quad \text{QUINDI VERRA' RESTITUITA CIRCA UNA TUPLE.}$$

↑
DATO DA TESTO

5. Presentare un piano fisico di esecuzione che utilizzi un cammino di accesso indicizzato per almeno una relazione di base.

