

## 2. Esercizi tipologia B

B. Data una query SQL/algebra relazionale e alcune informazioni relative allo schema fisico

- individuare i cammini di accesso a una relazione
- individuare e rappresentare graficamente un piano fisico che soddisfa determinate caratteristiche (se esiste)
- determinare se un certo piano fisico ammette elaborazione in pipelining
- calcolare una stima della cardinalità del risultato di una query di selezione
- determinare quali indici creare per favorire la scelta di un piano di esecuzione fisico ritenuto ottimale
- discutere l'impatto della clusterizzazione in un certo piano fisico

### ESERCIZIO 1

Si considerino le relazioni con schema:

PIZZA (CodP, Nome, Costo)  
 CLIENTE (TelC, NomeC, Via, NCiv, NInt)  
 ORDINE (TelC<sup>CLIENTE</sup>, Data, CodP<sup>PIZZA</sup>, Qta)

si assuma inoltre che i cataloghi di sistema includano questo contenuto:

#### Relazioni

R	T(R)	B(R)
PIZZA	30	5
CLIENTE	500	100
ORDINE	10000	2500

#### Cluster

NomeC	Tipo	Su
C1	Index	PIZZA.Costo
C2	Hash	CLIENTE.TelC
C3	Hash	ORDINE.CodP

#### Indici

Chiave di ricerca	Tipo	H(I)	L(I)	CARD(I)
PIZZA.Costo	B+	2	5	20
PIZZA.Nome	B+	2	5	30
ORDINE.CodP	Hash			
ORDINE.Data	B+	2	30	300
ORDINE.TelC	B+	2	50	500
CLIENTE.TelC	Hash			

Solo due  
indici per PIZZA

#### Attributi

A	R	V(A,R)	Min(A,R)	Max(A,R)
CodP	PIZZA	30		
Nome	PIZZA	30		
Costo	PIZZA	20	4	14
TelC	CLIENTE	500		
TelC	ORDINE	500		
Data	ORDINE	300	'01-Jan-2008'	'09-Jan-2009'
CodP	ORDINE	30		

Qta	ORDINE	15	1	20
-----	--------	----	---	----

Considerare l'interrogazione

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');

- Determinare tutti i possibili cammini di accesso per le relazioni PIZZA, ORDINE e CLIENTE (considerare a questo proposito sia le condizioni di selezione sia le condizioni di join)

PIZZA

- Scansione sequenziale → valida SEMPRE
- I<sub>PIZZA</sub>(Costo), Costo < 5

Controlla se esiste un indice CodP per PIZZA in quanto CodP è l'attributo ANCHOR  
 si join con ORDINE. Quando l'attributo\* ma non c'è non esiste

## ORDINE

- JOINING SEQUENZIALE
- $I_{ORDINE}(DATA)$ ,  $'4-DEC-2008' \leq DATA \leq '6-DEC-2008'$
- $I_{ORDINE}(COPP)$ ,  $COPP = V$  CON  $V$  ATTRIBUTO DI JOIN
- $I_{ORDINE}(TELE)$ ,  $TELE = V$  CON  $V$  ATTRIBUTO DI JOIN

## CLIENTE

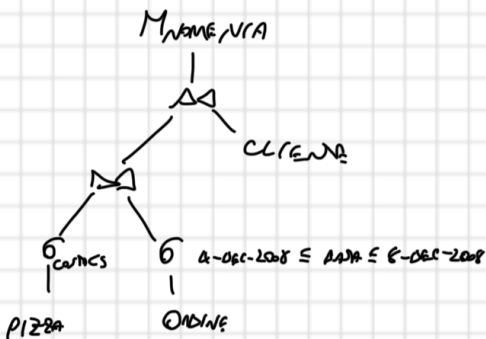
- JOINING SEQUENZIALE
- $I_{CLIENTE}(TELE)$ ,  $TELE = V$  CON  $V$  ATTRIBUTO DI JOIN

2. Presentare un esempio di piano di esecuzione fisico per l'interrogazione, supponendo di utilizzare un cammino di accesso sequenziale per tutte le relazioni, partendo dal piano di esecuzione logico in cui le selezioni vengono eseguite prima del join e la proiezione dopo.

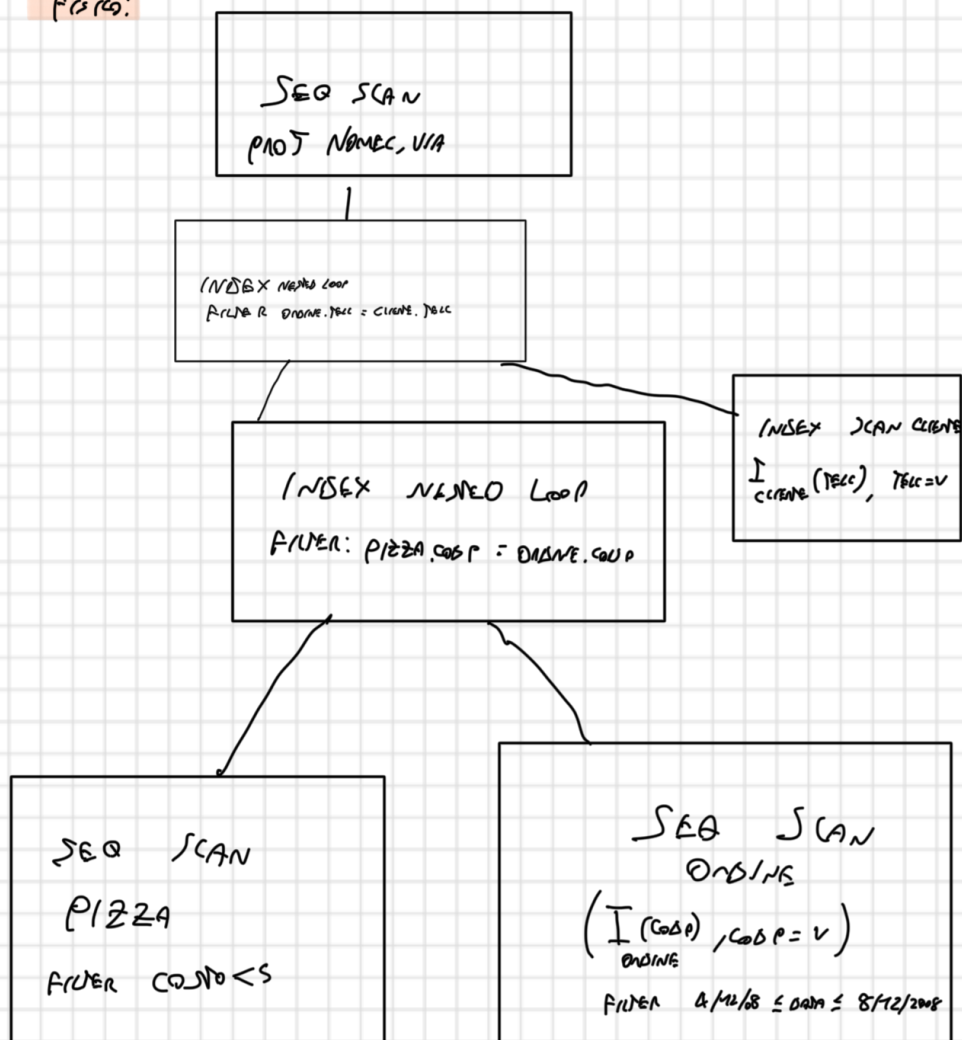
PRIMA FACCIAMO IL LOGICO, L'ESPRESSIONE IN ALGEBRA RELAZIONALE E':

$\pi_{NAME, VIA} (\sigma_{COSTO < 1 \wedge 4-DEC-2008 \leq DATA \leq 6-DEC-2008} (PIZZA \bowtie ORDINE \bowtie CLIENTE))$

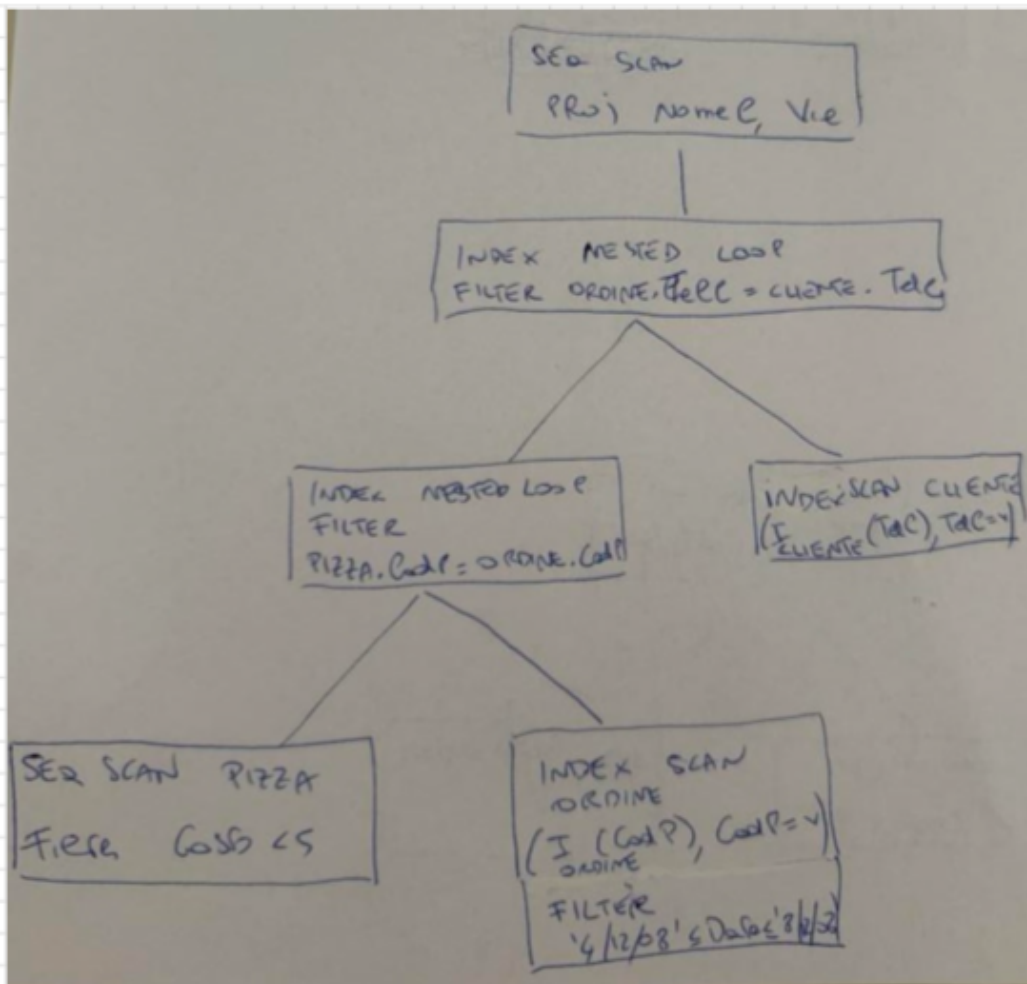
LOGICO:



FISICO:



3. Presentare, se esiste, un esempio di piano di esecuzione fisico per l'interrogazione, supponendo di utilizzare l'operatore di join index nested loop per entrambi i join.



4. Presentare, se esiste, un esempio di piano di esecuzione fisico per l'interrogazione, supponendo di utilizzare l'operatore di merge join per entrambi i join senza aggiungere ulteriori nodi rispetto al piano logico.

IL MERGE JOIN FUNZIONA SOLO SE ENTRAMBE LE TABELLE SONO ORDINATE RISPETTO ALL'ATTRIBUTO DI JOIN. L'INDICE  $I_{pizza}(CodP)$  NON ESISTE E L'INDICE  $I_{ordine}(Tdc)$  NON È CLUSIZZATO. IL PIANO NON ESISTE.

5. Determinare se i piani presentati ai punti (3) e (4) ammettono elaborazione in pipelining.

PIPELINE = FARE QUALCOSA IN PARALLELO. PER RISPONDERE A QUESTA DOMANDA SI DEVE GUARDARE QUESTA TABELLA:

SINTESI ELABORAZIONE DEGLI OPERATORI RELAZIONALI TRAMITE OPERATORI FISICI			
= ammette elaborazione in <u>pipeline</u> (nel join su relazione outer)			
Operatore fisico			
Operatore logico	Iterazione	Uso di indici	Partizionamento
ordinamento		x	merge sort a due fasi
$\sigma$	x	x	
$\pi^D$	x		
$\pi$ (con eliminazione duplicati)		x	Ordinamento (uso di funzione hash, non presentato)
$\bowtie$	nested loop semplice o a blocchi	index join	merge join hash join

PER CAPIRE SE AMMETTONO OPERAZIONI IN PIPELINE BISOGNA GUARDARE LA COLONNA ITERAZIONI E LEGGERE SE C'E' SCRITTA LA X.

IL PIANO AL PUNTO 3 AMMETTE ELABORAZIONI IN PIPELINE (LA SCANLINE JED, L'INDEX NESTED LOOP E LA PROIEZIONE SENZA LIMITAZIONE DI DUPLICATI AMMETTONO ELABORAZIONE IN PIPELINE)

## 6. I piani presentati ai punti (3) e (4) sono left deep?

IL PIANO 3 SONO LEFT DEEP PERCHE' LA RELAZIONE INTER DI OGNI OPERAZIONE DI JOIN (A BASTA PER CONVERGENZA NEL PIANO) E' SEMPRE UNA RELAZIONE DI BASE.

### ESERCIZIO 2

Si considerino le relazioni e lo schema fisico presentato nell'esercizio 1.  
Considerare l'interrogazione

```
SELECT TelC  
FROM ORDINE  
WHERE Qta = 5 AND CodP = 2
```

1. Calcolare una stima del numero di tuple restituite dall'interrogazione.

BISOGNA USARE QUESTA TABELLA:

#### STIMA DELLA DIMENSIONE DEL RISULTATO DELLA SELEZIONE

- $F(A=v) = 1/V(A,R)$
- $F(A \text{ IN } (v1, v2, \dots, vN)) = N * F(A=v)$
- $F(A > v) = (Max(A,R) - v) / (Max(A,R) - Min(A,R))$
- $F(A < v) = (v - Min(A,R)) / (Max(A,R) - Min(A,R))$
- $F(A1 = A2) = 1 / MAX(V(A1,R), V(A2,R))$
- $F(C1 \text{ AND } C2) = F(C1) * F(C2)$
- $F(C1 \text{ OR } C2) = F(C1) + F(C2) - F(C1) * F(C2)$
- $F(\text{NOT } C) = 1 - F(C)$

PER PRIMA COSA CALCOLIAMO IL FATTORE DI SELEZIONABILITA' DELLA CONDIZIONE:

$$F(Qta=5) = 1/V(qta, ORDINE) = 1/15$$

$$F(CodP=2) = 1/V(codp, ORDINE) = 1/30$$

$$F(Qta=5 \text{ AND } CodP=2) = F(Qta=5) * F(CodP=2) = 1/15 * 1/30 = 1/450$$

PER OTTENERE IL NUMERO STIMATO DI TUPLE RESTITuite, MOLTIPLICHIAMO IL NUMERO DI TUPLE DELLA RELAZIONE ORDINE PER IL VALORE DI SELEZIONABILITA':

$$T(ORDINE) * F(Qta=5 \text{ AND } CodP=2) = 10'000/450 = 22,22 \text{ OVVERO CIRCA 23 TUPLE}$$

2. Determinare i fattori booleani.

I fattori booleani sono  $Q_{ta} = 5$  e  $CodP = 2$  (sono due congiunti nella condizione

3. Supponendo di non avere ancora creato alcun indice e di poterne creare uno solo, su quale attributo converrebbe crearlo? Discutere l'impatto della clusterizzazione su questo indice. [la parte in blu è un chiarimento rispetto al testo discusso a lezione]

$CodP = 2$  PENSA' /TA IL FANTOMI DI SEIENRITA' PIU' BASSO