

OTTIMIZZAZIONI:

NESTED LOOP JOIN: $NP(R) + NR(R) \cdot NP(S)$

Ma Se Aggiungiamo un Indice B+Tree su S il Costo Cambia e diventa: $NP(R) + NR \cdot (\text{Profondità Indice} + \text{Selettività})$

Attributo A Nella
Relazione S

1 Pagina di Indice
Per Ogni livello

$$\frac{NR(R)}{VAL(A, S)}$$

Quante Righe Per Ogni
Valore di A

BLOCK NESTED LOOP JOIN:

Sfrutto i Dati che ho già in Memoria e Scaricare Pagina di S Solo Se ho finito tutti i Controlli

Ciascuna pagina di S (Tab. Interna) viene Letta **Per Ogni Pagina di R invece che Per Ogni Riga di R**

Loop

$$\begin{aligned} &\forall t_r \in R \{ \\ &\quad \forall t_s \in S \{ \quad \text{caricata interamente} \\ &\quad \quad \text{SE } (t_s, t_r) \in PJ \{ \\ &\quad \quad \quad \text{aggiungi} \\ &\quad \quad \} \\ &\quad \} \\ &\} \end{aligned}$$

BLOCK NESTED

$$\begin{aligned} &\forall \text{PAGINA } P_r \text{ di } R \{ \\ &\quad \forall \text{PAGINA } P_s \text{ di } S \{ \\ &\quad \quad \forall \text{RIGA } t_r \in R \{ \\ &\quad \quad \quad \forall \text{RIGA } t_s \in S \{ \\ &\quad \quad \quad \quad \text{SE } (t_s, t_r) \in PJ \{ \\ &\quad \quad \quad \quad \quad \text{aggiungi} \\ &\quad \quad \quad \} \\ &\quad \quad \} \\ &\quad \} \end{aligned}$$

Ottimizzo il Caricamento delle Pagine di S visto che Vengono Caricate \forall PAGINA di R e NON più per Ogni RIGA.

IL Costo DIVENTA $\Rightarrow NP(R) + NP(R) \cdot NP(S)$

MERGE-SCAN JOIN:

Si può Applicare Quando:

1. Entrambe Tab hanno le TUPLE **Ordinate** rispetto agli **Attributi di JOIN**

2. JOIN è un **uguaglianza Stretta** Tra Attributi di JOIN (EQUI)

R

22	a
32	s
45	h
87	b

S

22	z	8
22	s	7
35	h	4
45	s	9
45	c	3
87	h	5
87	g	6

$R.A = S.A$

Condizioni Soddisfatte

Sfutto l'ORDINAMENTO per Evitare CONFRONTI INUTILI ed IL Costo diventa: $NP(R) + NP(S)$

R

22	a
32	s
45	h
87	b

S

22	z	8
22	s	7
35	h	4
45	s	9
45	c	3
87	h	5
87	g	6

Da Qui in poi
so che A è
diverso

Se ho un Indice Su ENTRAMBE, devo fare una Stima per Eccesso e il costo è $\Rightarrow NR(R) + NR(S)$ se INDICE Sta Nel BUFFER
Posso NOTARE che Indice Qui non aiuta di molto, ma Potrebbe peggiorare se non Sta Nel BUFFER

HASH JOIN:

Solo con EQUI-JOIN, ma non richiede presenza di INDICI Ne che le Tabelle Siano ORDINATE rispetto all' Attributo di JOIN.

Con Relazioni Molto Grandi È VANTAGGIOSO.

Si Basa Sull' Avere a disposizione una fz. di HASH h che possiamo Applicare Agli Attributi di JOIN.

Se Parto da 2 Tuple: $t_R \in R$ e $t_S \in S$ e

$$t_{R.J} = t_{S.J} \Leftrightarrow h(t_{R.J}) = h(t_{S.J})$$

Ma può essere che ci Siano dei falsi Positivi (COLLISIONI), che Sono dei CASI Limitati e Potrei usarli come una SCREMATURA

se $t_{R.J} \neq t_{S.J} \Rightarrow$ sicuramente $h(t_{R.J}) \neq h(t_{S.J})$

IO EN Comune \Rightarrow Partizionare R ed S in Base ai Valori di h

Ricerca delle TUPLE che Soddifano PJ avviene Solo Tra le Partizioni con lo Stesso Valore di h

COSTO \Rightarrow ① Costo rice HASHMAP $\Rightarrow NP(R) + NP(S)$

② Ricerca \Rightarrow caso pessimo $= NP(R) \cdot NP(S)$

caso Medio dipende da #Buffer ma
ANCHE da distribuzione dei Valori

OTIMIZZATORE INTERROGAZIONI:

Genera **Tutti i possibili piani di Esecuzione** per poi valutare con formule **Approssimate** (MODELLO DI COSTO) il Costo di Ogni Alternativa.

Si Sceglie l'ALBERO di Costo minimo

Operatori/Algoritmi Alternativi
Uso o Meno di INDICI
Determino l'ORDINE

Stima Cardinalità RELazione (#Tuple)

Stima Dimensione Ogni TUPLA

Stima Dimensione Attributi

Stima #Valori distinti \forall attributo in Tab.

Stima Valore min & max attributo

STIMA COSTO INTERROGAZIONE:

INGREDIENTE (Codice, Nome, Calorie)

Composizione (Ricetta, **Ingrediente**, Quantità)

RICETTA (Codice, Nome, Regione, Tempo-Preparazione)

❧ Trovare Gli Ingredienti Usati in Ricette della Regione Veneto Riportando il codice, Nome e le Calorie dell'Ingrediente

```
SELECT R.codice, I.Nome, I.calorie
FROM Ricetta R JOIN Composizione C ON R.codice = C.ricetta
      JOIN Ingrediente I ON C.ingrediente = I.codice
WHERE R.regione = 'Veneto'
```

❧ Calcolare il costo (#Accessi Mem. Secondaria) con queste ipotesi:

- ① Selezione su RICETTA viene fatta con una Scansione Seq.
- ② Ordine Esecuzione del JOIN (RMC) ❧ I
- ③ JOIN fatto con NESTED LOOP JOIN
- ④ Risultato del primo JOIN intermedio Rimane Nel BUFFER

❧ Ci Sono le Seguenti Statistiche SULLE TABELLE:

	NP	NR	VAL(REGIONE)	VAL(ricetta)
RICETTA	12	260	20	
INGREDIENTE	40	1200		
COMPOSIZIONE	200	13000		260

Ci Sono 3 Operazioni

1 Selezione

R.REGIONE = 'veneto'

Scan Seq

COSTO = NP(R) = 12

Già State Selezionate

2 JOIN

R X C

*

NP(R) + NR(R) · NP(S)

12 + $\frac{NR(rietta)}{VAL(REGI, RIETTA)} \cdot 200$

12 + $\frac{260}{1,20} \cdot 200 = 2612$

Selezione + Join

* Se UNISCO Selezione con Scansione per il JOIN Vado a Cambiare il Valore 'NR(R)' visto che NON Saranno TUTTE ma Sono Solo Quelle 'VENETO'

Risultato del Primo Join

R' X S

NP(R') + NR(R') · NP(S)

Mantenuto nel BUFFER

COSTO = 0

$\frac{NR(C)}{VAL(R,C)} \cdot (\#Ric. Veneto)$

so 13000
1,200
13
260
1,20

COSTO ⇒ 650

Costo Complessivo \Rightarrow Somma JOIN

$$\Rightarrow 26000 + 2612 = 28612$$

Così cambia se è disponibile un INDICE B+Tree su Attributo Codice di INGREDIENTE (con PROFONDITÀ 2)

Dove lo posso usare Questo Indice? Nel Secondo JOIN, il Primo Rimane Invariato

da $[NP(R) + NR(R) \cdot NP(S)]$ a $[NP(R) + NR(R) \cdot (Prof. Indice + SElettività_{(codice, Indice)})]$

$$\text{Costo DIVENTA} \Rightarrow 0 + 650 \cdot \left(2 + \frac{NR(\text{ingrediente})}{val(cod, Ingrid)} \right)$$

$$\Rightarrow 1950$$

\rightarrow Codice è chiave Primaria ed Allora è uguale a $NR(\text{ingrediente})$

$$\text{Costo TOTALE} \Rightarrow 2612 + 1950 = 4562$$

ESEMPIO:

PAZIENTE (codice_ssn, Nome, Cognome,)

VISITA (codice_ssn, discode, Data, Ora-I, Ora-F, ..)

MEDICO (codisc, Cognome, Nome, Specialità)

① Se c'è Select Sequenziale è COSTO LETTURA + SCRITTURA

Si Azzerà se Viene Mantenuto nel Buffer

② NR Vanno Stimati (Quelle Sopravissute alla SELECT)

③ Se è PRIMARY KEY allora sono Pari a NR