

Ottimizzazione Fisica (Parte 3)

Lorenzo Vaccarecci

7 Maggio 2024

```
1  SELECT *
2  FROM Noleggi N, Video V, Film F
3  WHERE N.colloc = V.colloc AND
4  V.titolo = F.titolo AND
5  V.regista = F.regista
```

Il risultato del join tra Noleggi e Video deve essere memorizzato nel buffer se ci sta, altrimenti su disco.

1 Pipeline

Si può applicare solo se l'operatore fisico esegue un ciclo sulla relazione outer. Non tutti gli algoritmi possono essere eseguiti in pipeline:

- Nested loop semplice
- Nested loop a blocchi
- Indexed nested loop

In questo caso, il pipeline viene applicato alla relazione outer mentre la relazione inner, utilizzata ad ogni ciclo per i confronti, deve essere materializzata

_____ = ammette elaborazione in pipeline (nel join su relazione outer)

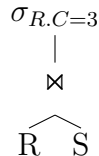
Operatore fisico			
Operatore logico	Iterazione	Uso di indici	Partizionamento
ordinamento		X	merge sort a due fasi
σ	X	X	
π^D	X		
π (con eliminazione duplicati)		X	Ordinamento (uso di funzione hash, non presentato)
\bowtie	nested loop semplice e a blocchi	index join	merge join hash join

2 Stima del costo di un piano di esecuzione

Per ogni piano, la stima viene fatta a partire dalle foglie (**bottom-up**). Per stimare il costo di un piano di esecuzione vengono considerati i **costi di accesso** e per ogni nodo dell'albero il **costo dell'operazione** e la **dimensione del risultato**. Il tutto poi sommato.

Le statistiche non sono sempre aggiornate, infatti sono aggiornate periodicamente. Molti DBMS prevedono un comando per richiedere l'aggiornamento delle statistiche (**ANALYZE** in PostgreSQL).

2.1 Esempio di stima del costo: Selezione



Il numero di tuple restituito dipende da quante tuple di R soddisfano la condizione (P).

Fattore di selettività: $F(P)$, se è < 1 è molto selettivo (poche tuple che soddisfano P) e viceversa.

Quante sono le tuple restituite da $\sigma_P(R)$?

$$F(P) \cdot T(R)$$

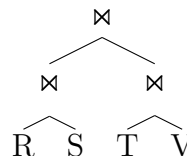
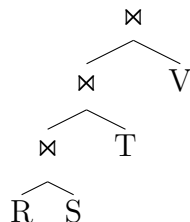
Come si stima il fattore di selettività?

$F(P) = \frac{\text{Casi favorevoli}}{\text{Casi possibili}} = \frac{1}{V(C,R)}$, a numeratore c'è il numero di valori della condizione. (*Più formule nelle slide p.63*)

3 Euristica per join

3.1 Piano Left Deep

$$R \bowtie S \bowtie T \bowtie V$$



L'albero di sinistra è Left Deep, quello di destra no.

3.2 Progettazione Fisica

Dato un carico di lavoro, l'attività di progettazione fisico si preoccupa di progettare uno schema fisico, in termini di insieme di indici creati, che permetta di rendere il più possibile efficiente l'esecuzione delle operazioni contenute nel workload. Una scelta impropria degli indici può portare a:

- inutilizzo degli indici
- scansioni sequenziali per uno o pochi record
- join multipli con tempi di esecuzione elevatissimi
- spreco di spazio disco

3.2.1 Approccio generale

Si considerano uno alla volta le interrogazioni, partendo da quelle usate più frequentemente. Per ogni interrogazione Q :

- si ipotizza un potenziale piano di esecuzione ottimale
- si individuano gli indici che permettono al sistema di prendere in considerazione il piano individuato
- si valuta empiricamente l'efficacia degli indici creati
- se gli indici non sono efficaci, cercare di comprendere la motivazione ed eventualmente rimuoverli dal livello fisico

Poi si valuta se gli indici identificati al passo i (per la query Q_i) sono compatibili con gli indici identificati fino al passo $i-1$ (per la query Q_{i-1}), ed eventualmente si apportano correttivi, **solo un indice per relazione può essere clusterizzato**.