

Sono date le seguenti query:

1. SELECT idFattura, COUNT(numRiga)
FROM DETTAGLIOFATTURA
WHERE idFattura > 10000
AND idProdotto BETWEEN 300 AND 500
GROUP BY idFattura
2. SELECT idProdotto, COUNT(idFattura)
FROM DETTAGLIOFATTURA
WHERE prezzoUnitario > 50
AND colore IN ('blu', 'verde')
GROUP BY idProdotto

sulla relazione:

DETTAGLIOFATTURA (idFattura, numRiga, idProdotto, colore, taglia, quantità, prezzoUnitario)

sulla quale sono costruiti due indici, uno clustered su idFattura ($h_{idFattura}=3$, $NL_{idFattura} = 1529$) e uno unclustered su idProdotto ($h_{idProdotto} = 3$, $NL_{idProdotto} = 1421$).

Si determini il miglior piano di accesso per la risoluzione della query tenendo conto dei seguenti dati:

NT = 250000, NP = 11500

$NK_{idFattura} = 20000$, $NK_{idProdotto} = 1000$, $NK_{taglia} = 5$, $NK_{colore} = 15$

$prezzoUnitario \in [10, 100]$

len(idFattura) = 4 byte, len(numRiga) = 4 byte, len(idProdotto) = 4 byte, len(colore) = 4 byte

len(taglia) = 4 byte, len(quantità) = 4 byte, len(prezzoUnitario) = 8 byte, len(TID) = 4 byte

D = 1 KB, u = 0.69

Si ipotizzi di utilizzare l'algoritmo Sort-Merge a Z=3 vie per l'eventuale ordinamento del risultato.

1) Query 1

$$f(p1) = \frac{\max(idFattura) - 10000}{NK_{idFattura}} = \frac{20000 - 10000}{20000} = \frac{1}{2}$$

$$f(p2) = \frac{500 - 300}{NK_{idProdotto}} = \frac{200}{1000} = \frac{1}{5}$$

$$ET = NT \times f(p1) \times f(p2) = 250000 \times \frac{1}{2} \times \frac{1}{5} = 25000$$

$$NP_R = \left\lceil \frac{ET \times (\text{len}(idFattura) + \text{len}(numRiga))}{D} \right\rceil = \left\lceil \frac{25000 \times (4 + 4)}{1024} \right\rceil = 196$$

$$C_{sort} = 2 \times NP_R \times \lceil \log_Z NP_R \rceil = 1960$$

- Accesso sequenziale

Il costo della scansione sequenziale è $C_{seq} = NP = \mathbf{11500}$

Se fosse possibile effettuare una ricerca dicotomica il costo sarebbe:

$$C_{seq} = \lfloor \log_2 NP \rfloor + \frac{NP}{2} = 13 + \frac{11500}{2} = \mathbf{5763}$$

- Accesso con indice su idFattura

$$\begin{aligned} C_{idFattura} &= (h_{idFattura} - 1) + \lceil f(p1) \times NL_{idFattura} \rceil + \lceil f(p1) \times NP \rceil \\ &= 2 + \left\lceil \frac{1}{2} \times 1529 \right\rceil + \left\lceil \frac{1}{2} \times 11500 \right\rceil = 2 + 765 + 5750 = \mathbf{6517} \end{aligned}$$

- Accesso con indice su idProdotto

$$\begin{aligned} C_{idProdotto} &= (h_{idProdotto} - 1) + \lceil f(p2) \times NL_{idProdotto} \rceil + EK_{idProdotto} \\ &\quad \times \phi\left(\frac{NT}{NK_{idProdotto}}, NP\right) + C_{sort} \\ &= (3 - 1) + \left\lceil \frac{1}{5} \times NL_{idProdotto} \right\rceil + f(p2) \times NK_{idProdotto} \\ &\quad \times \phi\left(\frac{250000}{1000}, 11500\right) + 1960 = 2 + 285 + 200 \times 247.3 + 1960 = \mathbf{51707} \end{aligned}$$

La ricerca dicotomica risulta il piano d'accesso più conveniente. Nel caso in cui ciò non sia possibile, il piano d'accesso migliore è quello basato sull'indice su *idFattura*.

2) Query 2

$$f(p1) = \frac{100-50}{100-10} = \frac{5}{9} \quad f(p2) = \frac{2}{NK_{colore}} = \frac{2}{15}$$

$$ET = NT \times f(p1) \times f(p2) = 250000 \times \frac{5}{9} \times \frac{2}{15} = 18519$$

$$NP_R = \left\lceil \frac{18519 \times (4 + 4)}{1024} \right\rceil = 145$$

$$C_{sort} = 1450$$

- Accesso sequenziale

$$C_{seq} = NP + C_{sort} = 11500 + 1450 = \mathbf{12950}$$

La ricerca dicotomica in questo caso non è possibile, in quanto non esistono predicati di selezione su *idFattura*.

- Accesso con indice su idFattura

Questo piano d'accesso non è certamente conveniente in quanto non sono presenti predicati su *idFattura* e la query non prevede un ordinamento su tale attributo. Il costo risulterà sicuramente superiore a quello della scansione sequenziale.

- Accesso con indice su idProdotto

Anche questo piano d'accesso può essere escluso a priori a causa dell'assenza di predicati su idProdotto. La presenza dell'indice permetterebbe di risparmiare il costo d'ordinamento (per il group by), ma il costo di reperimento delle pagine (per tutte le chiavi di idProdotto) tramite indice unclustered risulterebbe sicuramente troppo elevato.