

Query Optimizer

Thursday, September 25, 2025 5:33 PM

Diketahui relasi dalam skema basisdata relasional sebuah tempat persewaan buku sebagai berikut (yang digaris bawahi adalah primary key).

Pelanggan = {pid, nama, umur, alamat}
Buku = {bid, judul, penulis}
Peminjaman = {pid, bid, tanggal}
FK: Peminjaman (pid) → Pelanggan (pid);
Peminjaman (bid) → Buku (bid)

Asumsikan:

- Relasi Pelanggan terdiri atas 10000 tuple yang tersimpan dalam 1000 blok.
- Relasi Buku terdiri atas 50000 tuple yang tersimpan dalam 5000 blok.
- Relasi Peminjaman terdiri atas 300000 tuple yang tersimpan dalam 15000 blok.
- Ada 500 orang penulis buku berbeda.
- Umur pelanggan bervariasi antara 7 s.d. 24 tahun.
- Distribusi seragam (uniform) pada nilai atribut penulis dan umur.

Diaumsikan ukuran tuple hasil join antara relasi Buku dan Peminjaman adalah 2 kali ukuran tuple relasi Peminjaman.

Menggunakan skema relasi tersebut, tetapi:

- terdapat primary index dengan key penulis pada relasi Buku, dengan kedalaman index = 2.
- terdapat secondary index dengan key bid pada relasi Peminjaman, dengan kedalaman index = 3.

Diketahui query dalam aljabar relasional sebagai berikut.

$$\Pi_{\text{nama}} ((\sigma_{\text{umur} \geq 21}(\text{Pelanggan})) \bowtie ((\sigma_{\text{penulis} = \text{'Andrea Hirata'}}(\text{Buku}) \bowtie \text{Peminjaman}))$$

1. Tuliskan sebuah evaluation plan yang memanfaatkan kedua index yang disebutkan di atas untuk query tersebut dan estimasi cost untuk menjalankan plan tersebut. Estimasi pula jumlah tuple yang akan dihasilkan. Gunakan cara pada latihan query processing.

2. Tuliskan sebuah ekspresi aljabar relasional yang ekivalen dengan ekspresi tersebut (dengan menerapkan sejumlah equivalence rules) dan kemudian tuliskan sebuah evaluation plan berdasarkan ekspresi baru yang dihasilkan serta estimasi cost untuk plan tersebut. Evaluation plan yang dibuat dapat menggunakan/tidak menggunakan index yang tersedia.

index hanya bisa
& originalnya

- ① query awal: $\Pi_{\text{nama}} ((\sigma_{\text{umur} \geq 21}(\text{Pelanggan})) \bowtie ((\sigma_{\text{penulis} = \text{'Andrea Hirata'}}(\text{Buku}) \bowtie \text{Peminjaman}))$
- evaluation plan. heuristic optimization dengan early selection untuk mengurangi ukuran relasi sebelum operasi join
- (a) seleksi pada tabel Buku
- Karena ada primary index di nonkey,
→ cost pencarian via index = $2 + \left(\frac{50.000}{5.000} \right) = 2 + 10 = 12 \text{ block transfers}$
- estimasi ukuran hasil result: $\frac{50.000}{100} = 500$
- (b) join (a) dengan peminjaman
- menggunakan index nested-loop join untuk cari di secondary index pada Peminjaman
→ cost pencarian via index = kedalaman = 3
- dengan rata-rata peminjaman per buku: $\frac{n_{\text{peminjaman}}}{n_{\text{buku}}} = \frac{300.000}{50.000} = 6$
- cost block transfers: hasil (a) \times (cost index + $\frac{\text{rata-rata peminjaman per buku}}{10} + b(a)$)
 $= 100 \times (3 + 6) + 10 = 910 \text{ block transfers}$
- estimasi ukuran hasil result: $n(a) \times \text{rata-rata peminjaman per buku} = 100 \times 6 = 600 \text{ tuples}$
- (c) seleksi pada tabel Pelanggan
- linear scan → 1000 block transfers
- estimasi ukuran hasil result: $\frac{1}{18} \times n_{\text{pelanggan}} = \frac{1}{18} \times 10.000 = 222 \text{ tuples}$
- (d) join (b) dengan (c)
- menggunakan nested-loop join
- $n(c) \times b(b) + b(c) = 222 \times 60 + \frac{n(c)}{f(c)}$
 $= 222 \times 60 + \frac{222}{10} = 133.543 \text{ block transfers}$
- estimasi ukuran hasil result: $\frac{1}{18} \times n(c) = \frac{1}{18} \times 600 = 133 \text{ tuples}$

estimasi ukuran hasil result: $\frac{1}{18} \times N(c) = \frac{1}{18} \times 600 = 133$ tuples

(e) proyeksi nama

menggunakan pipelining sehingga

estimasi cost: 0 block transfer

estimasi ukuran hasil result = 133 tuples

SEHNGGA,

estimasi cost: 12 + 910 + 1000 + 133.543 = 135.465 block transfers

estimasi ukuran tuple yang dihasilkan = 133 tuples

② ekspresi aljabar relasional yang ekuivalen =

rule 6 (a) associative rule, sehingga Pelanggan \bowtie Peminjaman dulu, baru hasilnya \bowtie Buku

$T_{\text{nama}}((\sigma_{\text{penulis: 'Andrea Hirata'}}(\text{Buku}) \bowtie \sigma_{\text{umur} \geq 21}(\text{Pelanggan})) \bowtie \text{Peminjaman})$

evaluation plan. heuristic optimization dengan early selection untuk mengurangi ukuran relasi sebelum operasi join (tanpa index)

(a) seleksi pada tabel Buku

(linear scan) \rightarrow 5000 block transfers

estimasi hasil result = $\frac{5000}{500} = 100$ tuples

(b) Seleksi pada tabel Pelanggan

estimasi cost: $b_{\text{pelanggan}} = 1.000$ block transfers

estimasi hasil result: $\frac{1}{18} \times 1.000 = 2.222$ tuples

(c) join (a) dan (b)

nested-loop join dengan (a) sebagai outer-loop

$N(a) \times b(b) = 100 \times 223 + 10 = 22310$ block transfers

estimasi hasil result = $2.222 \times 100 = 222.200$ tuples

Y

Jumlah block:

$\frac{222.200}{10} = 22.220$ block

(d) join (c) dengan Peminjaman

nested-loop join dengan (c) sebagai outer-loop

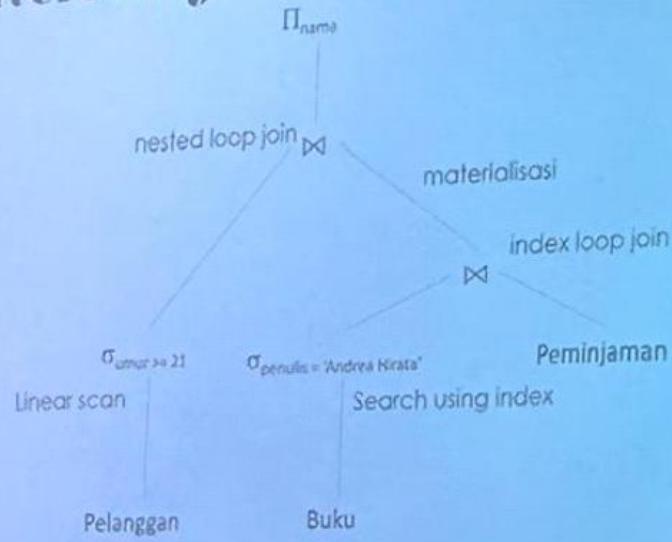
$N(c) \times b(\text{peminjaman}) + b(c) = 222.200 \times 15.000 + 44.440 = 3,33 \times 10^9$ block transfers

= 3.333.022.220 block transfers

Hasil akhir tetapi 133 tuples

dan estimasi cost = 3.333.050.530 block transfers

Latihan Alternatif Solusi



$\Pi_{name} ((\sigma_{umur \geq 21}(\text{Pelanggan})) \bowtie ((\sigma_{penulis = 'Andrea Hirata'}(\text{Buku})) \bowtie \text{Peminjaman}))$

Latihan Alternatif Solusi (lanjutan)

Operasi	Cost (#block transfer)	Banyaknya tuple
$\sigma_{penulis = 'Andrea Hirata'}(\text{Buku})$ Search using primary index Hasil operasi = A	$b_{Buku} = c + b_A = 2 + (100/10) = 12$	$n_A = 50000/500 = 100$
A \bowtie Peminjaman Asumsi: Index loop join Hasil operasi = B	$ \begin{aligned} &n_A * (c+1) + b_B \\ &= 100 * (3 + 6) + 60 \\ &= \text{150000} + 60 \\ &\quad \text{Hasil perlu dimaterialisasi karena akan} \\ &\quad \text{digunakan sbg inner relation utk operasi} \\ &\quad \text{berikutnya. } b_B \text{ adalah # blok yang ditulis ke} \\ &\quad \text{disk.} \\ &f_{Peminjaman} = 20 \\ &f_B = \frac{1}{2} f_{Peminjaman} = 10 \text{ (karena ukuran tuple} \\ &\text{hasil join 2 x ukuran tuple peminjaman)} \\ &b_B = n_B/f_B = 600/10 = 60 \end{aligned} $	$ \begin{aligned} &n_B < 300000 \\ &= 300000 / 500 \\ &= 600 \end{aligned} $ <p>bid adalah key pada Buku jadi # tuple hasil operasi < # tuple Peminjaman; 300000/500 karena A hanya melibatkan penulis = 'Andrea Hirata' dan ada 500 tuple berbeda</p>

Latihan Alternatif Solusi (lanjutan)

Operasi	Cost (#block transfer)	Banyaknya tuple
$\sigma_{umur \geq 21}(Pelanggan)$ Linear scan Hasil operasi = C	$b_{Pelanggan} = 1000$	$n_C = 4/18 * 10000$ = 2222 karena hanya melibatkan data Pelanggan umur ≥ 21
$C \bowtie B$ Asumsi: nested loop join Hasil operasi = D	$n_C * b_B = 2222 * 60$ = 133320	$n_D < 600 = 4/18 * 600 = 133$ pid adalah key pada Pelanggan di C: 4/18 karena hanya melibatkan data Pelanggan umur ≥ 21
$\Pi_{\text{nama}}(D)$	0 (karena pipeline)	$n_{\Pi_{\text{nama}}(D)} = n_D = 133$
Total	$= 12 + 960 + 1000 + 133320$ = 140432 $\underline{\underline{= 135292}}$	133

KE