

SOAL

Diketahui relasi dalam skema basis data relasional sebuah tempat persewaan buku sebagai berikut (yang digaris bawah 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)

Berikut ini adalah statistik terkait ketiga relasi di atas.

- 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.

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

Diketahui *query* dalam aljabar relasional sebagai berikut.

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

Tuliskan sebuah *evaluation plan* untuk query tersebut dan estimasikan cost untuk menjalankan plan tersebut. Estimasi pula jumlah tuple yang akan dihasilkan.

JAWABAN

A. Evaluasi Query Seleksi

1. Seleksi pelanggan dengan umur ≥ 21 tahun

Relasi Pelanggan terdiri dari 10000 tuple dalam 1000 blok. Umur pelanggan antara 7 s.d 24 tahun (total 18 tahun), Karena distribusi umur adalah seragam, proporsi pelanggan yang berumur ≥ 21 tahun adalah

$$(24-21+1)/(24-7+1) = 4/18 = 0.222222$$

Estimasi hasil seleksi ini adalah

$$0.222222 \times 10000 = 2222.22 \text{ tuple}$$

Estimasi jumlah blok yang diperlukan untuk menyimpan hasil ini adalah

$$0.222222 \times 1000 = 222.222 \text{ blok}$$

Query ini harus melakukan linear search (harus scan semua 1000 blok), maka cost nya adalah 1000 blok transfers + 1 seek.

2. Seleksi buku dengan penulis "Andrea Hirata"

Relasi Buku terdiri dari 50000 tuple dalam 5000 blok. Karena ada 500 penulis yang tersebar secara seragam, proporsi buku oleh Andrea Hirata adalah

$$1/500 = 0.002$$

Estimasi jumlah buku yang ditulis oleh Andrea Hirata adalah

$$0.002 \times 50000 = 100 \text{ tuple}$$

Estimasi jumlah blok yang diperlukan adalah

$$0.002 \times 5000 = 10 \text{ blok}$$

Query ini harus melakukan linear search (harus scan semua 5000 blok), maka cost nya adalah 5000 blok transfers + 1 seek.

B. Evaluasi Join

1. Join antara Buku dan Peminjaman

Setiap join antara Buku dan Peminjaman menghasilkan tuple 2 kali lebih besar dari tuple Peminjaman, maka ukuran tuple baru adalah 2 kali ukuran tuple Peminjaman.

Estimasi jumlah blok yang dibutuhkan untuk menyimpan hasil ini bisa dihitung dengan asumsi join standar.

Relasi Peminjaman sebanyak 300000 tuple dalam 15000 blok. Karena buku Andrea Hirata ada di 100 tuple dalam 10 blok, maka

$br = 10$ blok (buku Andrea Hirata)

$bs = 15000$ blok (Peminjaman)

$cost = br * bs + br \text{ block transfers} + 2 * br \text{ seeks}$

$$= 10 * 15000 + 10 + 2 * 10 = 150010 \text{ block transfers} + 20 \text{ seeks}$$

2. Join antara Pelanggan dengan B1

Sama seperti asumsi B1, ukuran tuple hasil join Buku dan Peminjaman adalah 2 kali tuple Peminjaman. Terdapat 600000 tuple dalam 30000 blok (asumsi ukuran blok sama), maka

$br = 222$ blok (pelanggan umur ≥ 21)

$bs = 30000$ blok (buku join peminjaman)

$cost = br * bs + br \text{ block transfers} + 2 * br \text{ seeks}$

$$= 222 * 30000 + 222 + 2 * 222 = 6660222 \text{ block transfers} + 444 \text{ seeks}$$

Estimasi cost keseluruhan adalah

$$1000 + 5000 + 150010 + 6660222 = 6816232 \text{ block transfers dan}$$

$$1 + 1 + 20 + 444 = 466 \text{ seeks}$$

Estimasi tuple dari keseluruhan query sekitar 600000 tuple (best case), dimana setiap tuple pada Pelanggan bisa mencocokkan tuple hasil join Buku dan Peminjaman.