

Administração de Banco de Dados

Centro Federal de Educação Tecnológica Celso Suckow da Fonseca – CEFET-RJ

Lista de exercícios de processamento de consultas

1. Quais são as técnicas para se realizar consultas pontuais em uma relação? Estas técnicas diferem na presença de índices primários e secundários?
2. Quais são as técnicas para se realizar consultas por faixa de valor em uma relação? Estas técnicas diferem na presença de índices primários e secundários?
3. Considere o esquema Pessoa(CPF, Nome, Sexo, Idade, Telefone, Cidade, Estado) para o cadastro de uma grande empresa de venda de produtos via internet. Assuma a presença dos seguintes índices: (i) primário <Nome, CPF>, (ii) secundário <CPF>, (iii) secundário <Sexo>, (iv) secundário <Estado, Cidade>.
 - a. Avalie a cada um dos índices criados, indicando se você apoiaria a sua criação. Indique se você criaria algum outro índice.
 - b. Apresente a consulta equivalente em álgebra relacional e indique as estratégias mais adequadas, no que tange a seleção, para as seguintes consultas:
 - i. `Select Nome, CPF from Pessoa where Nome like 'João%'`
 - ii. `Select Avg(Idade) from Pessoa where Sexo = 'F'`
 - iii. `Select Nome from Pessoa where CPF = 01234567890`
4. Quais são as diferentes formas de se realizar ordenação? Descreva os algoritmos e apresente a complexidade aproximada?
5. No processamento de junções, explique os seguintes algoritmos:
 - a. Nested-loop join
 - b. Merge join
 - c. Hash join
6. Explique a complexidade do algoritmo Nested-loop para duas relações R e S no melhor caso $O(n+m)$ e no pior caso $O(n*m)$, onde n e m são respectivamente o número de blocos de R e S.
7. Explique a complexidade do algoritmo Merge-Join para duas relações R e S é aproximadamente no melhor caso $O(n+m) + \delta$, onde n e m são respectivamente o número de blocos de R e S. O que representa δ ?
8. Explique a complexidade do algoritmo Hash-Join para duas relações R e S é aproximadamente no melhor caso $O(n+m) + \epsilon$, onde n e m são respectivamente o número de blocos de R e S. O que representa ϵ ?
9. Por que não é desejável forçar os usuários a escolher explicitamente uma estratégia de processamento de consultas? Existem casos nos quais pode ser desejável ao usuário saber os custos das candidatas a estratégias de processamento de consulta? Explique a sua resposta.

10. Considere a seguinte consulta SQL para um banco de dados de uma instituição financeira:

```
select T.branch-name
from branch T, branch S
where T.assets = S.assets
and S.branch-city = "Brooklyn"
```

Escreva uma eficiente expressão de álgebra relacional que seja equivalente a esta consulta. Justifique.

11. Quando os SGBD costumam usar índices de árvore (Bitmap e Árvore B+) e os índices de Hash? Justifique.
12. Seja as relações $R1(\underline{A}, B, C)$ e $R2(\underline{C}, D, E)$ tem as seguintes propriedades: $R1$ têm 20.000 tuplas, $R2$ têm 45,000 tuplas, 25 tuplas de $r1$ cabem em um bloco, e 30 tuplas de $r2$ cabem em um bloco. Estime o número de acessos a blocos para as seguintes estratégias: $R1 \bowtie R2$ para Nested Loop, Merge-Join e Hash-Join.
13. Considere as duas consultas equivalentes: $\sigma_{\text{saldo} < 2500}(\pi_{\text{saldo}}(\text{conta}))$ e $\pi_{\text{saldo}}(\sigma_{\text{saldo} < 2500}(\text{conta}))$, indique as formas de execução possíveis para estas consultas. Qual das duas seria a mais adequada?