

Universidade Federal do Ceará
Campus Quixadá
Prof.^a Livia Almada
Construção de Sistemas de Gerência de Banco de Dados
Lista - Processamento de Consultas

- 1) Explique o processo envolvido no processamento de uma consulta dada de entrada. Apresente um exemplo para facilitar a compreensão.
- 2) Como é feita a estimativa do custo de cada consulta pelo SGBD?
- 3) O que é o catálogo e quais informações importantes do processamento de consultas ele armazena?
- 4) Considere uma consulta de seleção simples da forma **R.atributo op valor**. Quais são os caminhos de acesso possíveis em cada um desses casos:
 - a. Não há nenhum índice e o arquivo não é ordenado;
 - b. Não há nenhum índice, mas o arquivo é ordenado.
- 5) Se um índice de árvore B+ corresponde à condição de seleção, como o índice clusterizado afeta o custo? Discuta isso em termos de seletividade da condição.
- 6) Considere uma relação R(a,b,c,d,e) contendo 5.000.000 registros, onde cada página de dados da relação contém 10 registros. R é organizada como um arquivo ordenado com índices secundários. Suponha que R.a seja uma chave candidata para R, com valores residindo no intervalo de 0 a 4.999.999, e que R esteja armazenada na ordem de R.a. Para cada uma das consultas, indique qual das estratégias (ou combinação delas) provavelmente é mais barata. Explique.
 - Acessar diretamente o arquivo ordenado R.
 - Usar índice de árvore B+ clusterizado no atributo R.a.
 - Usar um índice de hash no atributo R.a.
 - Usar índice de árvore B+ clusterizado nos atributos (R.a, R.b).
 - Usar índice de hash nos atributos (R.a, R.b).
 - Usar índice de árvore B+ não clusterizado no atributo R.a.
 - a. `SELECT * FROM R WHERE a<50.000 AND b<50.000`
 - b. `SELECT * FROM R WHERE a=50.000 AND b<50.000`
 - c. `SELECT * FROM R WHERE a>50.000 AND b=50.000`
 - d. `SELECT * FROM R WHERE a=50.000 AND b=50.010`
 - e. `SELECT * FROM R WHERE a!=50.000 AND b=50.000`
 - f. `SELECT * FROM R WHERE a<50.000 OR b=50.000`
- 7) Considere a junção R join S on R.a=S.b. A métrica de custo é o número de E/S de página, a não ser que seja denotado de outra forma, e o custo da gravação do resultado deve ser ignorado.

A relação R contém 10.000 tuplas e tem 10 tuplas por página.
A relação S contém 2.000 tuplas e também tem 10 tuplas por página.
O atributo b da relação S é a chave primária de S.
As duas relações são arquivos de heap simples.
Nenhuma das relações tem qualquer índice.

- a. Qual é o custo da junção de R e S utilizando uma junção de loops aninhados simples?
- b. Qual é o custo da junção de R e S usando uma junção de loops aninhados orientada a página?

8) Considere a seguinte consulta SQL:

```
SELECT T.nome_agência  
FROM agência T, agência S  
WHERE T.fundos > S.fundos and S.cidade_agência =  
"Maringá"
```

Considere como se cada tupla de Agência possui 40 bytes, são 1000 páginas ao total e cada página contém 100 tuplas. Escreva um plano para a consulta apresentada acima. Calcule o custo do seu plano escolhido em termos de I/O.

- 9) Você optaria por outro plano se soubesse que existe um índice sobre o atributo cidade_agência? Justifique sua resposta.