

PRIMEIRA PROVA DE LABORATÓRIO DE BANCO DE DADOS

ACH2025 TURMA 2019 104 – TEÓRICA

Prof. José de J. Pérez-Alcázar

1. Para cada uma das perguntas seguintes marque as afirmações que são verdadeiras. Duas erradas anulam uma boa. 40%

- Sobre criação de índices:
 - a. Para agilizar as consultas devemos criar muitos índices sobre um arquivo ()
 - b. Uma busca indexada sempre é mais eficiente que uma busca sequencial ()
 - c. Os índices secundários são bons no caso de consultas de igualdade pelo campo índice e são poucos os valores ()
 - d. A criação de índices sobre uma tabela não cria “overhead” nas atualizações sobre essa tabela ()
 - e. Todas as anteriores são verdadeiras ()
- Os planos de execução serializáveis são:
 - a. Corretos, porque geram os mesmos resultados que um plano serial. ()
 - b. Equivalentes a um plano resultante da execução serial das suas transações. ()
 - c. Mais eficientes que os planos com leitura repetitiva. ()
 - d. $r1(X), r2(X), w1(X), r1(Y), w2(X), w1(Y)$ é conflito serializável ()
 - e. $r1(X), w1(X), r2(X), w2(X), r1(Y), w1(Y)$ é conflito serializável ()
- Dado um arquivo com um tamanho de bloco (B) igual a 1024 bytes, e um arquivo índice secundário com tamanho do campo de busca 9 bytes e pointer 6 bytes. Esse campo é chave primária. Suponhamos que existem 400000 registros no arquivo de dados. Então:
 - a. O número de blocos do arquivo índice será 5883 ()
 - b. No caso de ser criado um índice multinível este teria 4 níveis ()
 - c. O último nível (raiz) teria 2 blocos ()
 - d. O fan-out do índice seria de 68 ()
 - e. Seriam feitos 4 acessos a disco ()

- No caso da avaliação de uma junção:
 - a. O algoritmo de laços aninhados por blocos pode ser usado só para os casos de uma junção por igual. ()
 - b. Se as tabelas estiverem ordenadas uma boa maneira de fazer a junção é usar o algoritmo de junção por fusão (merge) ()
 - c. A junção hash é, talvez, a melhor maneira de implementar uma junção por igual, no caso das tabelas não estiverem ordenadas. ()
 - d. O algoritmo de junção por fusão (merge) é útil para qualquer tipo de junções ()
 - e. Nenhuma das anteriores é verdadeira ()

- Suponha (por simplicidade neste exercício) que somente uma tupla se encaixa em um bloco e a memória mantenha no máximo 3 frames de página (blocos). Suponha também que é usado o algoritmo de sort-merge para classificar as seguintes tuplas sobre o primeiro atributo: (kangaroo, 17), (wallaby, 21), (emu, 1), (wombat, 13), (platypus, 3), (lion, 8), (warthog, 4), (zebra, 11), (meerkat, 6), (hyena, 9), (hornbill, 2), (baboon, 12). Podemos afirmar, então:
 - a. Usando números de tuplas: t_1, \dots, t_{12} . As sequências ordenadas (runs) iniciais têm três blocos cada, uma delas é $\{t_{12}, t_{11}, t_{10}\}$ ()
 - b. Cada passagem mistura duas sequências, então na primeira passagem, uma das sequências (runs) será: $\{t_3, t_1, t_6, t_5, t_2, t_4\}$ ()
 - c. Outra sequência deve ser: $\{t_{12}, t_{11}, t_{10}, t_9, t_7, t_8\}$ ()
 - d. Na segunda passagem o algoritmo termina. ()
 - e. O número total de passagens é dado por $\lceil \log M(br/M) \rceil$. Onde M é o número máximo de blocos em memória (buffer) ()

2. Considere a seguinte consulta SQL para o banco de dados bancário estudado em aula: 30%

```
SELECT T.nome_agência
FROM agência T, agência S
WHERE T.ativos > S.ativos and S.cidade_agência = 'Brooklyn'
```

Escreva uma expressão em álgebra relacional que seja equivalente a essa consulta. Represente ela como uma árvore e olhe como poderiam ser aplicadas heurísticas (regras de equivalência) para fazer mais eficiente a consulta. Justifique a sua resposta.

3. Considere o esquema da relação Marinheiros: 30%

Marinheiros(mid: integer, mnome: string, avaliacao: integer, idade: real)

Assuma que cada tupla de Marinheiros é de tamanho 50 bytes, que uma página pode conter 80 tuplas, e que nós temos 500 páginas. Para cada uma das seguintes condições de seleção, estime o número de páginas recuperadas, dada a informação no catálogo descrita na pergunta.

- a. Assuma que nos temos uma árvore B+ T sobre o campo de busca (Marinheiros.mid), e assumo que a altura do T é 4. o número de páginas folhas é 50, o mínimo valor do campo no índice é 1 e o máximo é 100000.
 - i. $\sigma_{\text{Marinheiros.mid} < 50000}(\text{Marinheiros})$
 - ii. $\sigma_{\text{Marinheiros.mid} = 50000}(\text{Marinheiros})$
- b. Assuma que nos temos um índice hash T sobre o campo de busca (Marinheiros.mid), e assumo o mínimo valor do campo no índice é 1 e o máximo 100000.
 - i. $\sigma_{\text{Marinheiros.mid} < 50000}(\text{Marinheiros})$
 - ii. $\sigma_{\text{Marinheiros.mid} = 50000}(\text{Marinheiros})$