e-Fólio A

U.C. 21103

Sistemas de Gestão de Bases de Dados

2023-2024

Resolução e Critérios de Correção

INSTRUÇÕES

- O e-fólio é constituído por 3 perguntas com cotação de 1,0 valor cada. A cotação global é de 3 valores.
- O e-fólio deve ser entregue num único ficheiro PDF, não zipado, com fundo branco, com perguntas numeradas e sem necessidade de rodar o texto para o ler. Cada pergunta com uma ou mais páginas, deve ser iniciada numa nova página. Penalização de 10% a 100%.
- Não são aceites e-fólios manuscritos, i.e., tem penalização de 100%.
- O nome do ficheiro deve seguir a normal "eFolioA" + <nome estudante>.
- Durante a realização do e-fólio, os estudantes devem concentrar-se na resolução do seu trabalho individual, não sendo permitida a colocação de perguntas ao professor ou entre colegas.
- A interpretação das perguntas também faz parte da sua resolução, se encontrar alguma ambiguidade deve indicar claramente como foi resolvida.
- A legibilidade, a objetividade e a clareza nas respostas serão valorizadas, pelo que, a falta destas qualidades será penalizada.

Vetor Cotações:

1a	1b	2a	2b	3a	3b pergunta
4	6	4	6	5	5 décimas

Critérios de correção gerais: todas as respostas devem ser justificadas, incluir imagens e exemplos com vista a clarificar os argumentos expostos.

1) (1 valor) Considerando otimização de consultas em SQL, examine a seguinte tabela vendas com os campos (idVenda, dataVenda, valor, idProduto, idCliente).

a) Escreva uma instrução SQL que crie um índice que possa melhorar a performance de consultas que frequentemente filtram por idProduto e ordenam por dataVenda.

b) Explique como um plano de execução pode ser usado para otimizar consultas e como o usaria para analisar a eficiência do índice criado na parte (a).

Resposta:

a)
CREATE INDEX idx_vendas_produto_data ON vendas (idProduto, dataVenda);

b)
Um plano de execução é uma representação do caminho que o sistema de gestão de bases de dados (SGBD) escolhe para executar uma consulta. Mostra a ordem das operações, a forma como os índices serão utilizados, as junções entre tabelas, e as operações de agregação ou filtragem. Analisar o plano de execução de uma consulta pode revelar operações caras, tais como varreduras de tabela inteira ou junções ineficientes, e pode ajudar a identificar onde a adição de índices pode melhorar a performance.

Para analisar a eficiência do índice criado, um utilizador poderia executar a consulta com o comando EXPLAIN (ou EXPLAIN ANALYZE em alguns SGBD) para observar se o índice idx_vendas_produto_data está a ser utilizado efetivamente para filtrar por idProduto e ordenar por dataVenda. Se o índice for eficaz, o plano de execução deve mostrar que está a ser usado uma pesquisa indexada ao invés de uma varredura completa da tabela

Critérios de correção
alínea a) 4 décimas, criação adequada do índice
alínea b) 6 décimas, explicar plano de execução e a sua aplicação a este exemplo
- erros, omissões, redundâncias ou indentação desadequada: -20% a -100%

2) (1 valor) Considere as seguintes tabelas para uma biblioteca:

Estudantes (estudanteId -> nome, curso, dataNascimento)
Obras (obraId -> título, anoPublicação, editora)
Empréstimos (empréstimoId -> estudanteId, obraId, dataEmprestimo)

E a seguinte consulta SQL:

```
SELECT E.nome
FROM Estudantes E, Obras O, Empréstimos Em
WHERE E.estudanteId = Em.estudanteId
        AND O.obraId = Em.obraId
        AND O.editora = 'O'Reilly'
        AND YEAR(E.dataNascimento) BETWEEN 1990 AND 2000
```

Assumindo:

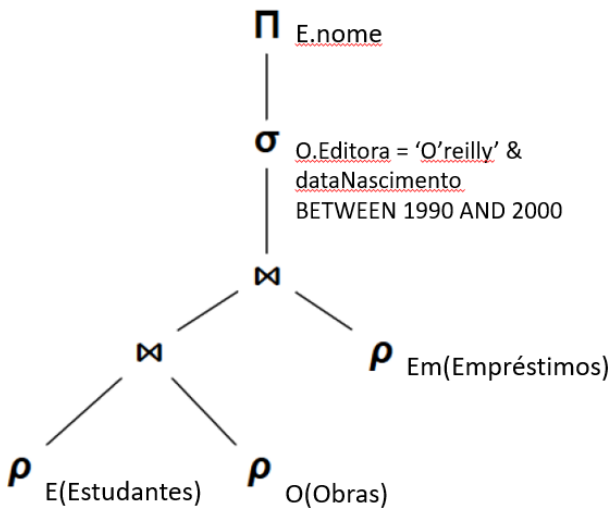
- Existem 15.000 registos de estudantes armazenados em 1.500 páginas.
- Existem 100.000 registos de obras armazenados em 10.000 páginas.
- Existem 500.000 registos de empréstimos armazenados em 25.000 páginas.
- Existem 1.000 editoras diferentes.
- As datas de nascimento dos estudantes variam de 1980 a 2010

a) Desenhe a árvore do plano de consulta, assumindo que não há índices e os dados não estão ordenados por nenhum atributo. Justifique e exemplifique a sua resposta?

b) Calcule o custo deste plano de consulta (i) e a cardinalidade do resultado(ii). Justifique e exemplifique a resposta.

Resposta:

a)



b)

i) Para a junção de blocos em ciclo ('block nested-loop join'), o número de blocos transferidos é $b_E+b_E\times b_{Em}$ e o número de blocos procurados é $2\times b_E$.

Blocos transferidos = $b_E+b_E\times b_{Em}$ = 1.500 + 1.500 x 25.000 = 37.501.500

Blocos procurados= $2\times b_E$ = 2 x 1.500 = 3.000

A junção seguinte na árvore é calculada da mesma forma.

ii) A estimativa da cardinalidade do resultado seria:

editoras: 1 editora 'O'Reilly' em 1.000 editoras, 1/1.000

datas de nascimento: 1990 a 2000 em 1980 a 2010, correspondendo a 11 anos de um intervalo de 31 anos, então 11/31

número de empréstimos: 500.000

cardinalidade = 500.000× 1/1.000 × 11/31 = aproximadamente 177 linhas ou tuplos.

Critérios de correção
alínea a) 4 décimas
alínea b) 6 décimas
- erros, omissões, redundâncias: -20% a -100%

3) (1 valor) No contexto do controle de concorrência em bases de dados, considere as seguintes transações simultâneas e suas operações sobre a base de dados:

Transação T1: R1(z), W1(z)
Transação T2: R2(y), W2(y)
Transação T3: R3(x), W3(x)
Transação T4: R4(z), R4(x), W4(x),

Onde R representa uma operação de leitura e W uma operação de escrita, com as letras x,y,z representando diferentes itens de dados na base de dados.

a) O que entende por sequenciamento serializável e não serializável? Justifique e exemplifique a resposta.

b) Dada a seguinte ordem de operações:
R1(z), R2(y), W1(z), R3(x), W2(y), R4(z), W3(x), R4(x), W4(x), determine se este sequenciamento é serializável a conflitos e/ou a vistas. Justifique sua resposta com base nas definições e no sequenciamento dado.

Resposta:

a)

sequenciamento serializável: Um sequenciamento é serializável quando as transações são executadas uma após a outra, sem qualquer sobreposição nas suas operações. Neste tipo de sequenciamento, todas as operações de uma transação são completadas antes de começar a execução de outra transação.

sequenciamento não serializável: Um sequenciamento é não serializável quando as operações de diferentes transações são intercaladas.

b)

O sequenciamento é serializável a vistas e conflitos, pois mantém a ordem de dependência de leitura/escrita, e é possível formar um sequenciamento serializável sem alterar o resultado final das transações.

Critérios de correção
alínea a) 5 décimas
alínea b): 5 décimas
- erros, omissões, redundâncias ou formatação desadequada: -20% a -100%