

U.C. 21103

Sistemas de Gestão de Bases de Dados

2021-2022

Resolução e Critérios de Correção

INSTRUÇÕES

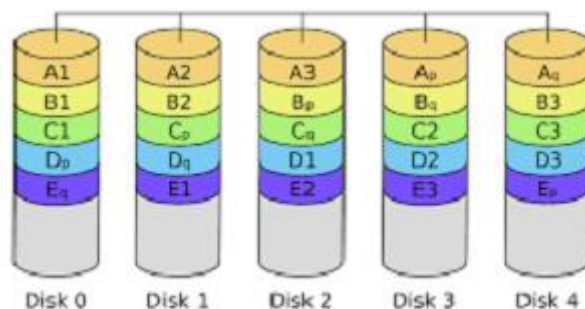
- O e-fólio é constituído por 3 alíneas com cotação de 1,0 valor cada. A cotação global é de 3 valores.
- O e-fólio deve ser entregue num único ficheiro PDF, não zipado, com fundo branco, com perguntas numeradas e sem necessidade de rodar o texto para o ler. Cada pergunta com uma ou mais páginas, deve ser iniciada numa nova página. Penalização de 1 a 3 valores.
- Não são aceites e-fólios manuscritos, i.e., tem penalização de 100%.
- O nome do ficheiro deve seguir a normal “eFolioA” + <nº estudante> + <nome estudante com o máximo de 3 palavras>
- 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
3	7	6	4	3	7	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) Relativamente ao tema do armazenamento.



a) Explique por palavras suas o que entende por RAID 0, RAID 5 e RAID 6. Justifique e exemplifique a resposta.

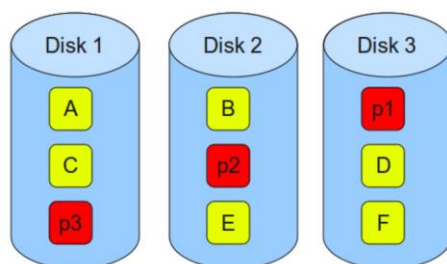
b) Considere 5 discos com 2TB cada. Calcule o espaço de armazenamento útil para RAID 0, RAID 5 e RAID 6. Encontre as expressões genéricas (fórmulas) para os três tipos de RAID.

Resposta:

a) Os RAID (redundant arrays of independent disks) distinguem-se da seguinte forma:

RAID 0 - Nonredundant stripping, os dados são lidos/escritos em múltiplos discos, como se fosse um único disco, melhorando assim o desempenho. É necessário no mínimo dois discos. O lado negativo é que não há tolerância a falhas (cópias ou paridade).

RAID 5 – Block interleaved distributed parity, é a configuração RAID mais comum nos servidores empresariais. A paridade está distribuída pelos vários discos. Permite a falha de um disco. É necessário no mínimo três discos.



RAID 6 – P+Q Redundancy scheme, é semelhante ao RAID 5, mas armazena informação extra, permitindo a falha de dois discos. A paridade e a recuperação de erros estão distribuídas pelos vários discos. É necessário no mínimo quatro discos.

b) Temos 5 discos de 2 TB cada. O espaço útil é o seguinte:

- RAID 0, sem redundância: 10 TB úteis + 0 TB proteção = 10TB
- RAID 5, paridade: 8 TB úteis + 2 TB proteção = 10 TB
- RAID 6, paridade e recuperação: 6 TB úteis + 4 TB proteção = 10 TB

Expressões genéricas para N discos:

- RAID 0: espaço útil = $1 - 0/N = 100\%$
- RAID 5: espaço útil = $1 - 1/N$, para 5 discos: $1 - 1/5 = 4/5 = 80\%$
- RAID 6: espaço útil = $1 - 2/N$, para 5 discos: $1 - 2/5 = 3/5 = 60\%$

Para a redundância genérica r com N discos, o espaço útil = $1 - r/N$

Critérios de correção

alínea a):

- 3 décimas, para definições RAID

alínea b):

- 3 décimas, para cálculo do espaço útil
- 4 décimas, expressões genéricas
- erros, omissões, redundâncias ou indentação desadequada: -20% a -100%

2) (1 valor) Relativamente ao tema dos índices (Hash-table e B⁺-tree) considere o esquema da base de dados da Figura 1.

a) Quando se deve usar índice suportados por Hash-tables e por B⁺-tree? Escolha uma tabela para aplicar um índice suportado por uma Hash-table. Escolha outra tabela para aplicar um índice suportado B⁺-tree. Justifique todas as respostas.

b) Escreva em SQL os comandos CREATE INDEX para as tabelas que escolheu.

Resposta:

a) Formas de usar os índices:

Hash-table: para múltiplas leituras e poucas alterações

B⁺-tree: usado em múltiplas alterações ('insert' e 'delete')

Quais as tabelas mais adequadas:

Hash-tables: Addresses (justificação: porque tem menos alterações)

B⁺-tree: Order_Products ou Deliveries_Products (porque tem mais alterações)

b) Usando a seguinte sintaxe SQL:

```
CREATE INDEX index_name ON table_name [USING method]
(column_name [ASC|DESC] [NULL{FIRST|LAST}])
```

```
CREATE INDEX Addresses_index ON Addresses USING btree
(address_id)
```

```
CREATE INDEX Order_Products_index ON Order_Products USING hash
(order_id, product_id)
```

Critérios de correção

alínea a):

- 3 décimas, formas de usar os índices
- 3 décimas, tabelas mais adequadas

alínea b):

- 4 décimas, SQL com método
- erros, omissões, redundâncias ou indentação desadequada: -20% a -100%

3) (1 valor) No tema do Processamento e Otimização de Consultas, considere o caso da operação de seleção em que, o tempo de procura (seek) $t_s=4$ milissegundos e o tempo de transferência de um bloco $t_T=0.1$ milissegundos. Considere ainda um ficheiro com 25000 linhas, em que cada bloco contem 500 linhas.

a) O que entende por tempo de procura (seek) t_s ? Calcule o custo com varrimento linear.

b) Calcule o custo da seleção com índice B⁺-tree, com máximo grau de 10, para uma comparação de igualdade com a chave primária.

Resposta:

a) Tempo de procura (seek) t_s , é o tempo de posicionamento do braço do disco, quando é necessário a leitura/escrita de dados num determinado setor.

t_s – tempo procura = 4 ms

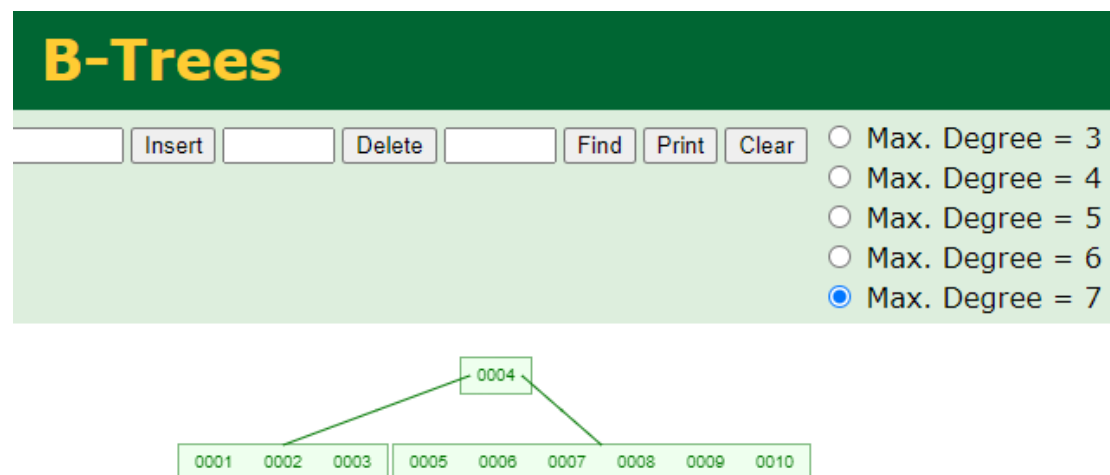
b – número blocos = $25000/500=50$ blocos

t_T – tempo transferência bloco =0.1 ms

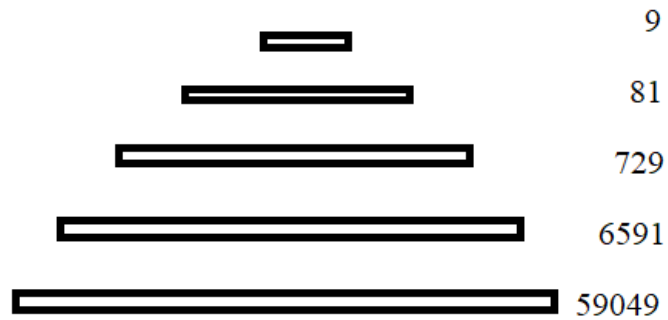
Custo com varrimento linear= $t_s + b \cdot t_T = 4 + 50 \cdot 0.1 = 9$ ms

(ver A1 página 543)

b) Numa B⁺-tree, com máximo grau é dado pelo comprimento máximo do nó. Exemplo de comprimentos máximo de 7. Neste exemplo a altura do índice h_i é de 1.



Para 25000 registos, numa B-tree equilibrada a altura dos índices h_i é de 4:



Utilizando logaritmos, $\text{LOG}_9(25000) = 4.6$

t_s – tempo procura = 4 ms

t_T – tempo transferência bloco = 0.1 ms

h_i – altura do índice = 4

Custo índice B⁺-tree = $(h_i + 1) \cdot (t_T + t_s) = (4+1) \cdot (0.1 + 4) = 20.5$ ms

(ver A2 página 543)

Critérios de correção

alínea a):

- 3 décimas, fórmula A1

alínea b):

- 3 décimas, altura índice

- 4 décimas, fórmula A2

- erros, omissões, redundâncias ou indentação desadequada: -20% a -100%

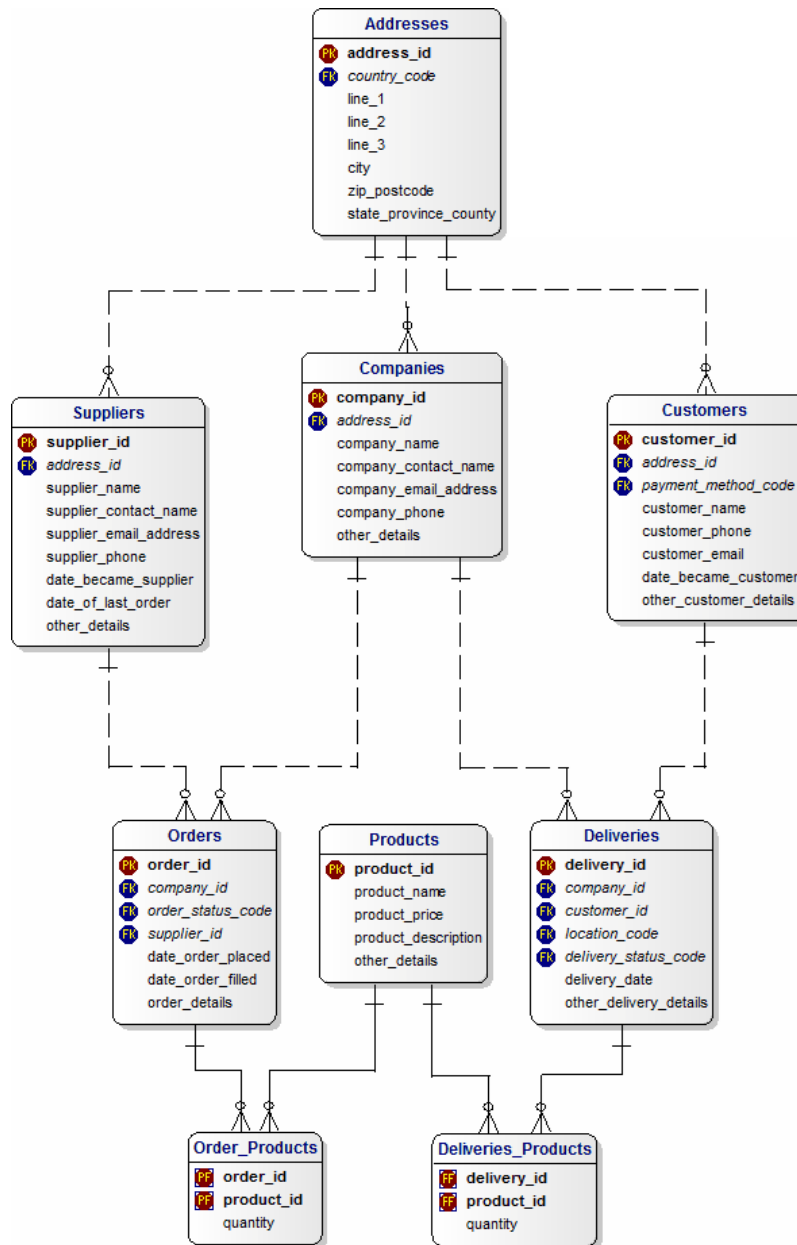


Figura 1 – Esquema de base de dados