

**21103 - Sistemas de Gestão de Bases de Dados  
2013-2014  
e-fólio A**

**Resolução e Critérios de Correção**

PARA A RESOLUÇÃO DO E-FÓLIO, ACONSELHA-SE QUE LEIA ATENTAMENTE O SEGUINTE:

- 1) O e-fólio é constituído por 4 perguntas. A cotação global é de 2 valores.
- 2) 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. Penalização de 1 a 2 valores.
- 3) Não são aceites e-fólios manuscritos, i.e. tem penalização de 100%.
- 4) O nome do ficheiro deve seguir a normal “eFolioA” + <nº estudante> + <nome estudante com o máximo de 3 palavras>
- 5) 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.
- 6) A interpretação das perguntas também faz parte da sua resolução, se encontrar alguma ambiguidade deve indicar claramente como foi resolvida.
- 7) A legibilidade, a objectividade e a clareza nas respostas serão valorizadas, pelo que, a falta destas qualidades serão penalizadas.

Vetor Cotações  
1 2 3 4 pergunta  
5 5 5 5 décimas

1) Na área de armazenamento de dados, o que entende por RAID (“redundant arrays of independent data”). Desenvolva em 200 palavras.

Resposta:

Com vista a melhorar o desempenho nos acessos a disco e aumentar a segurança (através de redundância) existem três atributos que diferenciam a classificação dos discos RAID: "striping", "mirroring" e paridade.

Por "striping" entende-se a segmentação em faixas com vista a melhorar o desempenho com múltiplos acessos a disco.

Por "mirroring" entende-se que existe uma duplicação, cópia ou espelho.

Por paridade é uma operação para deteção e correção de erros.

Para os vários tipos de RAID teremos a seguinte classificação segundo os critérios de desempenho e redundância:

	<b>desempenho</b>	<b>redundância</b>	
<b>RAID</b>	<b>"striping"</b>	<b>"mirroring"</b>	<b>paridade</b>
0	nível bloco		
1		existe espelho	
2	nível bit		dedicada
3	nível byte		dedicada
4	nível bloco		dedicada
5	nível bloco		distribuída
6	nível bloco		duplamente distribuída
híbridos			
0+1	nível bloco	existe espelho	
5+1	nível bloco	existe espelho	distribuída

Para além dos RAID de 0 a 6 podem ser combinados híbridos como os RAID "0+1" e o "5+1".

O RAID 5 é o RAID mais utilizado na indústria e caracteriza-se por um "striping" ao nível dos blocos. Relativamente à redundância de dados, não tem "mirroring" e tem paridade distribuída.

A paridade do RAID 5 utiliza a função XOR, se um disco falha, os dados dos outros dois podem ser combinados e reconstruída a informação em falta.

Critério de correção:

- (0,3) definir RAID

- (0,2) definir os atributos

2) Considere uma “hash-table” estática com a função  $h(x) = x \bmod 8$  e a dimensão do “bucket” igual a 3, como na figura em baixo. Insira numa tabela vazia os seguintes valores: 17, 12, 35, 11, 19, 5, 29, 7, 23, 31, 27 e 46.

Key	Value	
125	Data	Bucket 1
240	Data	
500	Data	
227	Data	Bucket 2
440	Data	
...		
850	Data	Bucket N
900	Data	

Resposta:

Existe uma colisão no “Bucket” 3 que é ultrapassada pela criação de um apontador para um novo conjunto de “keys” de “overflow”.

Bucket	key	
0		
1	17	
2		
3	35, 11, 19	→ [ 27 ]
4	12	
5	5, 29	
6	46	
7	7, 23, 31	

Critério de correção:

- (0,3) aplicar (mod 8)
- (0,2) aplicar “overflow”

3) Qual o número de acessos a disco na operação de álgebra relacional de divisão.

Resposta:

Suponha  $r(A,B)$  e  $s(B)$ . No exemplo temos  $A$ =pilotos e  $B$ =aviões; pretende-se saber se existe algum piloto que voa todos os aviões referidos na relação  $s$ :

$r$	A	B	$\div$	$s$	B	=	$r \div s$	A
	Xavier	B52			A380			Zito
	Yam	A380			B52			
	Yam	B747			B747			
	Zito	A380						
	Zito	B52						
	Zito	B747						

a) O algoritmo segue os seguintes passos:

- Começamos por ordenar a relação  $s$  por  $B$ . De seguida ordenamos a relação  $r$  por  $(A,B)$ .
- Varrendo a relação  $r$  considere o atributo  $A$ . Em simultâneo, varrendo a relação  $s$  verifique se os atributos  $B$  são iguais na relação  $r$  e  $s$ . Se for o caso, o resultado da divisão é o atributo  $A$  e continua a pesquisa.

A relação  $r$  é varrida uma única vez. A relação  $s$  é varrida  $N$  vezes, sendo  $N$  o número de atributos distintos de  $A$ .

O número total de acessos a disco, depois das tabelas ordenadas, é de  $|r|+N*|s|$ , onde  $N$  corresponde ao número de atributos distintos de  $A$ .

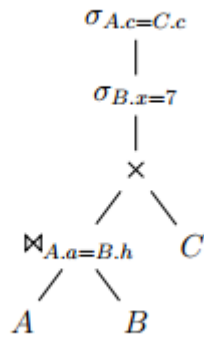
b) Foi ainda considerada correta a resposta obtidas através da implementação possível em SQL, que recorre ao produto cartesiano.

Sabendo que  $r \div s = \Pi_{R-S}(r) - \Pi_{R-S}((\Pi_{R-S}(r) \times s) - r)$ , sendo  $r$  e  $s$  quaisquer relações,  $R$  e  $S$  o conjunto de atributos daquelas relações  $r$  e  $s$ , e  $S \subseteq R$  (Maier, 1983).

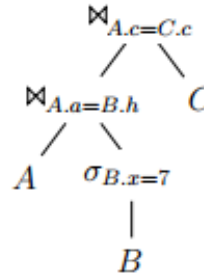
**Critério de correção:**

- (0,3) explicar a operação de divisão,
- (0,2) valor de  $|r|+N*|s|$ ,

4) Otimize a consulta e apresente a justificação. De seguida, esclareça qual o papel da otimização de consultas na determinação dos planos.



árvore original



Resposta: árvore otimizada

i) A seleção de  $B.x=7$  deve ser realizada na base da árvore. O produto cartesiano associado à seleção  $A.c=C.c$  é equivalente a uma junção com a condição  $A.c=C.c$ .

ii) o papel da otimização de consultas na determinação dos planos das consultas.

O plano de execução da consulta pretende ser o mais eficiente possível, com vista a consumir o menor número de recursos computacionais.

A otimização da consulta é realizada através de árvores equivalentes, regras de equivalência e heurísticas. As estimativas de custo de cada solução têm em conta as operações e as dimensões das tabelas.

Depois de encontrar uma solução ótima, ou quase-ótima, a consulta é implementada conforme o plano (ou estratégia) previamente definida.

Critério de correção:

- (0,3) árvore ótima: produto cartesiano e  $B.x=7$ ;
- (0,2) plano da consulta