

Resolução e Critérios de Correção

U.C. 21103

Sistemas de Gestão de Bases de Dados

18 de julho de 2012

INSTRUÇÕES

- O tempo de duração da prova de p-fólio é de 90 minutos.
- O estudante deverá responder à prova na folha de ponto e preencher o cabeçalho e todos os espaços reservados à sua identificação, com letra legível.
- Visto que o enunciado da prova não é utilizado para resposta, poderá ficar na posse do mesmo.
- Verifique no momento da entrega das folhas de ponto se todas as páginas estão rubricadas pelo vigilante. Caso necessite de mais do que uma folha de ponto, deverá numerá-las no canto superior direito.
- Em hipótese alguma serão aceites folhas de ponto dobradas ou danificadas.
- Exclui-se, para efeitos de classificação, toda e qualquer resposta apresentada em folhas de rascunho.
- Os telemóveis deverão ser desligados durante toda a prova e os objectos pessoais deixados em local próprio da sala de exame.
- O enunciado da prova é constituído por **2** páginas e termina com a palavra **FIM**. Verifique o seu exemplar do enunciado e, caso encontre alguma anomalia, dirija-se ao professor vigilante nos primeiros 15 minutos da mesma, pois qualquer reclamação sobre defeitos de formatação e/ou de impressão que dificultem a leitura não será aceite depois deste período.
- Utilize unicamente tinta azul ou preta.
- O exame é sem consulta. A interpretação das perguntas também faz parte da sua resolução, se encontrar alguma ambiguidade deve indicar claramente como foi resolvida.

A informação da avaliação do estudante está contida no **vetor das cotações**:

Questão: 1 2 3 4 5

C: 25 25 25 25 20 décimas

Grupo A – Sistemas de Bases de Dados

1. (2,5 valores) Na área de armazenamento de dados, o que entende por RAID (“redundant arrays of independent data”) nível 5? Qual o nº mínimo de discos para o RAID 5? Porquê?

(Resposta: 3/4 página, 1^a página)

(i) 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:

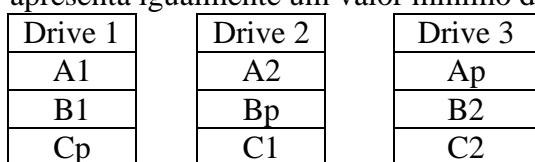
RAID	desempenho	redundância	
	"striping"	"mirroring"	paridade
0	nível bloco		
1		existe espelho	
2	nível bit		dedicada
3	nível byte		dedicada
4	nível bloco		dedicada
5	nível bloco		distribuída
6	nível bloco		duplamente distribuída
híbridos			
0+1	nível bloco	existe espelho	
5+1	nível bloco	existe espelho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

(ii) Número mínimo de discos: 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.

A Drive 3 é utilizada para paridade	A Drive 2, no caso de falha, pode ser reconstruída
01101101 Drive 1 XOR 11010100 Drive 2 10111001 Drive 3 Paridade	10111001 Drive 3 XOR 01101101 Drive 1 11010100 Drive 2 Recuperada

O nº mínimo de discos é portanto três, disposto da forma que se segue. A versão comercial apresenta igualmente um valor mínimo de três discos.



2. (2,5 valores) Considerando a base de dados dos Bancos, para a seguinte consulta SQL, encontre um plano eficiente.

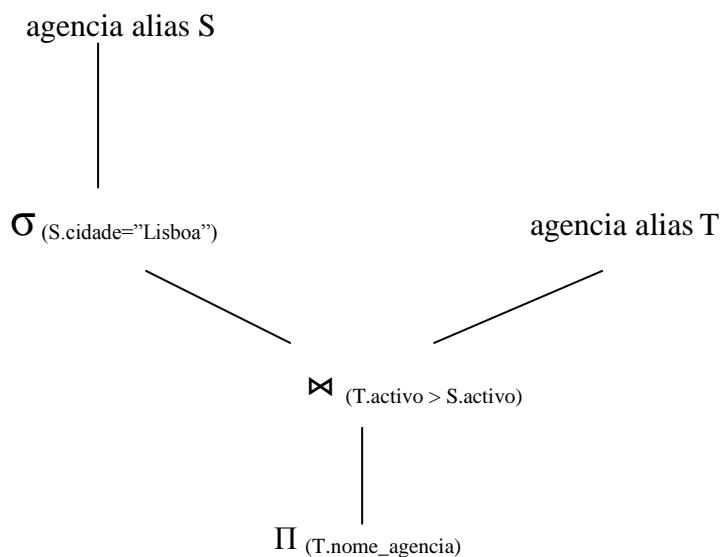
```
Select T.nome_agência
From agência T, agência S
Where T.activos > S.activos and S.cidade_agência = "Lisboa"
```

Desenhe uma árvore da expressão em álgebra relacional e escreva a expressão em álgebra relacional, equivalente à consulta. Justifique a resposta.

(Resposta: 1 página, 2ª página)

Resposta:

A forma mais eficiente é seleccionar em $S.\text{cidade} = \text{"Lisboa"}$;
 Em seguida realizar a junção para $T.\text{activo} > S.\text{activo}$;
 e por fim realizar uma projecção com $T.\text{nome_agencia}$



A expressão resulta:

$$\Pi_{(T.\text{nome_agencia})} [\sigma_{(S.\text{cidade} = \text{"Lisboa"})} (\text{agencia alias S}) \bowtie_{(T.\text{activo} > S.\text{activo})} (\text{agencia alias T})]$$

3. (2,5 valores) Dadas as transacções T1 e T2 onde foi aplicado o protocolo 2PL, verifique a possibilidade de "deadlock".

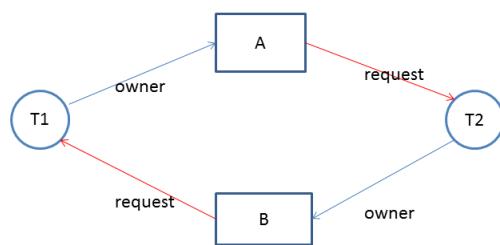
T1: lock-S(A); read(A); lock-X(B); read(B); if A=0 then B:=B+1; write(B); unlock(A); unlock(B);

T2: lock-S(B); read(B); lock-X(A); read(A); if B=0 then A:=A+1; write(A); unlock(B); unlock(A);

(Resposta: 1/2 página, 3^a página)

Resposta:

A execução destas transações pode resultar um “deadlock”.



Considere o seguinte exemplo:

T1	T2
lock-S(A)	
read(A)	lock-S(B)
lock-X (B)	read(B)
	lock-X (A)

4. (2,5 valores) Quais são as acções na recuperação de uma base de dados que utiliza o esquema de baseados em "logs" com modificação diferida dos dados, no caso de "roll back" de uma transacção?

(Resposta: 1/2 página, 3ª página)

Resposta:

Nos esquemas baseados em "logs" todos os "updates" são registados no "log", existindo duas variantes:

- no esquema de modificação diferida, durante a execução da transação, todas as operações de "write" são diferidas até que a transação atingir quase o "commit", nessa altura é utilizada a informação dos "logs";

- no esquema de modificação imediata, o sistema aplica todos os "updates" diretamente na base de dados; se existir um "crash" o sistema usa a informação do "log" para encontrar um estado estável.

Cada transação T precisa de ser "undone" se o "log" contém o registo (T, start) mas não tem o registo (T, commit); e cada transação T precisa de ser "redone" se o "log" contém o registo (T, start) e o registo (T, commit).

Grupo B – Prática em “Data Warehousing”

- 5.** (2 valores) Pretendemos desenhar um “Data Warehouse” do seguinte sistema. Defina a tabela de factos em primeiro lugar. De seguida, defina três dimensões para o “Data Warehouse” e apresente a tabela de factos associada às três dimensões.

"Um cinéfilo, detentor de uma coleção significativa de filmes, pretende uma aplicação que lhe permita armazenar e consultar os seus filmes. Pretende que os filmes possam ser consultados por título, país de origem, ano de realização, realizador e pelos actores intervenientes (cada filme tem um único realizador e vários actores). Para além da informação referida, é também necessário saber a duração de cada filme e, caso exista, o endereço na internet dos realizadores e actores. Os filmes são classificados por tipo: comédia, terror, ficção científica, etc. "

(Resposta: 1 página, 4^a página)

Critérios de correção:

- a tabela de Factos deve refletir os intervenientes nos filmes, os atores e os realizadores;
- as tabelas de Dimensões possíveis: pessoas, filmes, tempo;

FIM