

Resolução e Critérios de Correção

U.C. 21103

Sistemas de Gestão de Bases de Dados

26 de fevereiro de 2016

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 das provas presenciais.
- 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 p-fólio 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

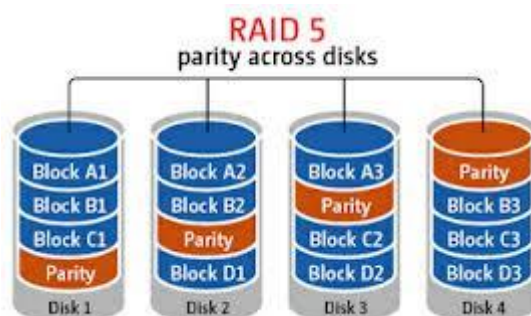
Grupo A – Sistemas de Bases de Dados

1. (2,5 valores) Uma falha de energia ocorreu quando um bloco de disco está a ser escrito, resultando num bloco parcialmente escrito. Suponha que blocos parcialmente escritos podem ser detetados. A propriedade de atomicidade dos blocos é dada quando o bloco do disco é totalmente escrito (ou seja, não são permitidas gravações parciais). Sugira um procedimento para conseguir o efeito de atomicidade dos blocos com um disco RAID nível 5. O procedimento deve envolver o trabalho no caso de recuperação da falha. Exemplifique o procedimento convenientemente.

(Resposta: 1 página)

Para o RAID 5 teremos a seguinte classificação segundo os critérios de desempenho e redundância:

	desempenho	redundância	
RAID	"striping"	"mirroring"	paridade
5	nível bloco	---	distribuída



No RAID 5 para garantir a atomicidade a escrita do bloco tem os seguintes passos:

- i. escrita da informação no 1º bloco
- ii. escrita da informação do bloco de paridade

No momento da recuperação, consistindo cada conjunto de bloco de ordem n de cada um dos discos é considerado. Se nenhum dos blocos do conjunto ter sido parcialmente escrito, e o conteúdo do bloco de paridade são consistentes com o conteúdo dos blocos de informação, em seguida, mais nenhuma ação devem ser tomadas.

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. Para 4 discos, com os dados D e a paridade P, teremos:

Drive 1	Drive 2	Drive 3	Drive 4
D0	D1	D2	P0
D3	D4	P1	D5
D6	P2	D7	D8
P3	D9	D10	D11

$$\begin{aligned}
 P0 &= D0 \text{ XOR } D1 \text{ XOR } D2 \\
 P1 &= D3 \text{ XOR } D4 \text{ XOR } D5 \\
 P2 &= D6 \text{ XOR } D7 \text{ XOR } D8 \\
 P3 &= D9 \text{ XOR } D10 \text{ XOR } D11
 \end{aligned}$$

No caso de falha da Drive 2, por exemplo, a informação pode ser reconstruída através do operador XOR.

A Drive 4 é utilizada para paridade	No caso de falha da Drive 2, esta pode ser reconstruída
0110 1101 Drive 1 XOR 1101 0100 Drive 2 XOR 0000 0000 Drive 3 1011 1001 Drive 4 Paridade P0	0110 1101 Drive 1 XOR 0000 0000 Drive 3 XOR 1011 1001 Drive 4 1101 0100 Drive 2 Recuperada

Visto que os blocos parcialmente escritos podem ser detetados, no caso de escrita na Drive 2 e falha de atualização na Paridade D0, teremos:

Escrita na Drive 2, dados D1	Na fase de recuperação, a paridade é atualizada
0110 1101 Drive 1 XOR 1111 1110 Drive 2 XOR 0000 0000 Drive 3 1011 1001 Drive 4 Paridade P0	0110 1101 Drive 1 XOR 1111 1110 Drive 2 XOR 0000 0000 Drive 3 1001 0011 Drive 4 Recuperada

Critério de correção:

- (1,0) descrição do procedimento
- (1,5) descrição da recuperação

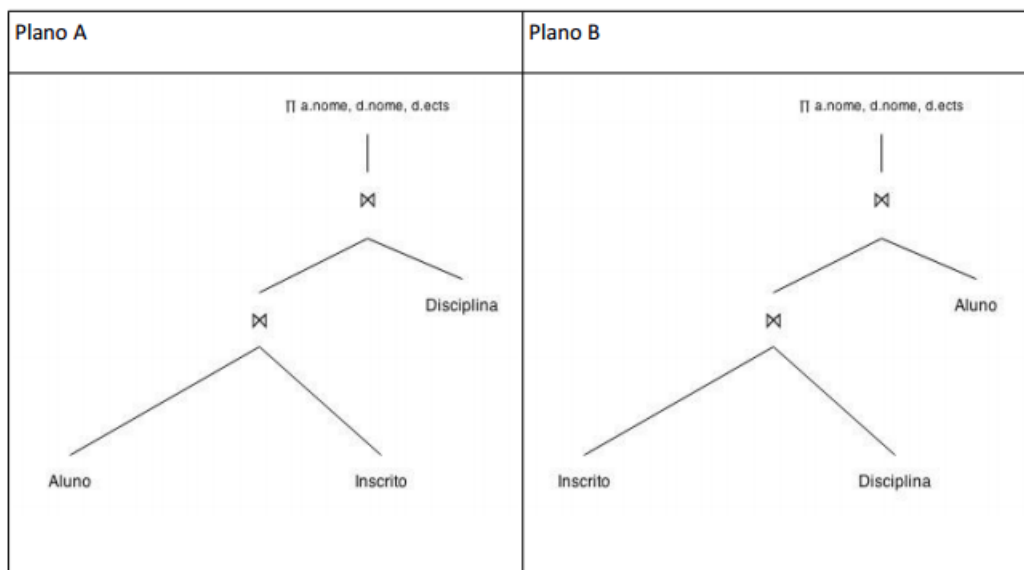
2. (2,5 valores) Explique porque é que o SGBD tem de criar planos de execução. De seguida escreva a seguinte consulta em álgebra relacional e desenhe pelo menos dois planos de execução, utilizando um estrutura em árvore.

```
select a.nome, d.nome, d.ects
from aluno a, inscrito i, disciplina d
where a.id= i.aluno_id
and i.disc_id = d.id
```

(Resposta: 1 página)

i) Visto que o SQL é um linguagem declarativa, o SGBD terá de encontrar a sequência de execução ou Plano de Execução.

ii) Os dois planos seguintes utilizam a estrutura “left-deep”:



Critério de correção:

- (1,0) parte i)
- (1,5) parte ii)
- penalização de 1 a 2 valores para árvores incorretas

3. (2,5 valores) Dado o seguinte sequenciamento/escalonamento (“schedule”) que envolve as duas transações:

T1: lê o dado x, soma 100 e escreve o dado x, i.e. R1(x), W1(x+100)

T2: lê o dado x, subtrai 50 e escreve o dado x, i.e. R2(x), W2(x-50)

Quantos e quais sequenciamentos possíveis? Para o valor inicial de $x = 10$, quais os respectivos valores finais para cada sequenciamento?

(Resposta: 1 página)

i) Existem 6 sequenciamentos possíveis.

S1	S2	S3	S4	S5	S6
R1(x) W1(x+100) R2(x) W2(x-50)	R1(x) R2(x) W1(x+100) W2(x-50)	R1(x) R2(x) W2(x-50) W1(x+100)	R2(x) W2(x-50) R1(x) W1(x+100)	R2(x) R1(x) W2(x-50) W1(x+100)	R2(x) R1(x) W1(x+100) W2(x-50)

ii) Para o valor inicial de $x=10$

S1	S2	S3	S4	S5	S6
$x=60$	$x=-40$	$x=110$	$x=60$	$x=110$	$x=-40$

Critério de correção:

- (1,5) parte i)

- (1,0) parte ii)

4. (2,5 valores) Explique e exemplifique o objetivo do mecanismo de “checkpoint”. Como é que a frequência do “checkpoint” afeta o desempenho do sistema quando não ocorre a falha? Qual a influência da frequência no tempo que leva para se recuperar de uma falha do sistema? Quantas vezes deve o “checkpoint” ser chamado?

(Resposta: 1 página)

O “checkpoint” (ou ponto de sincronização) é utilizado para reduzir o tempo de recuperação depois de uma falha do SGBD, quando da leitura dos “logs” (registro) do sistema. Se não existissem “checkpoints”, teria de ser feita a leitura de todo o “log” e a operação de “undo” e “redo” das respectivas transações. Se o “checkpoint” for realizado, as operações de escrita das transações passaram para os discos rígidos, sendo uma grande parte do “log” ignorado na fase de recuperação da falha.

Visto que os “checkpoints” degradam o desempenho do SGBD, a sua frequência deve ser reduzida se não for crítica a recuperação rápida. Se for preciso recuperação rápida, a frequência de “checkpoints” deve aumentar. Assim, para as variáveis frequência, desempenho do sistema e tempo de recuperação teremos:

- Sistema fiável (falhas são raras) tem maior intervalo entre checkpoints, com menos degradação do desempenho e uma recuperação mais lenta.
- Sistema menos fiável (falhas são mais frequentes) tem menor intervalo entre checkpoints, com mais degradação do desempenho e uma recuperação mais rápida.

Critério de correção:

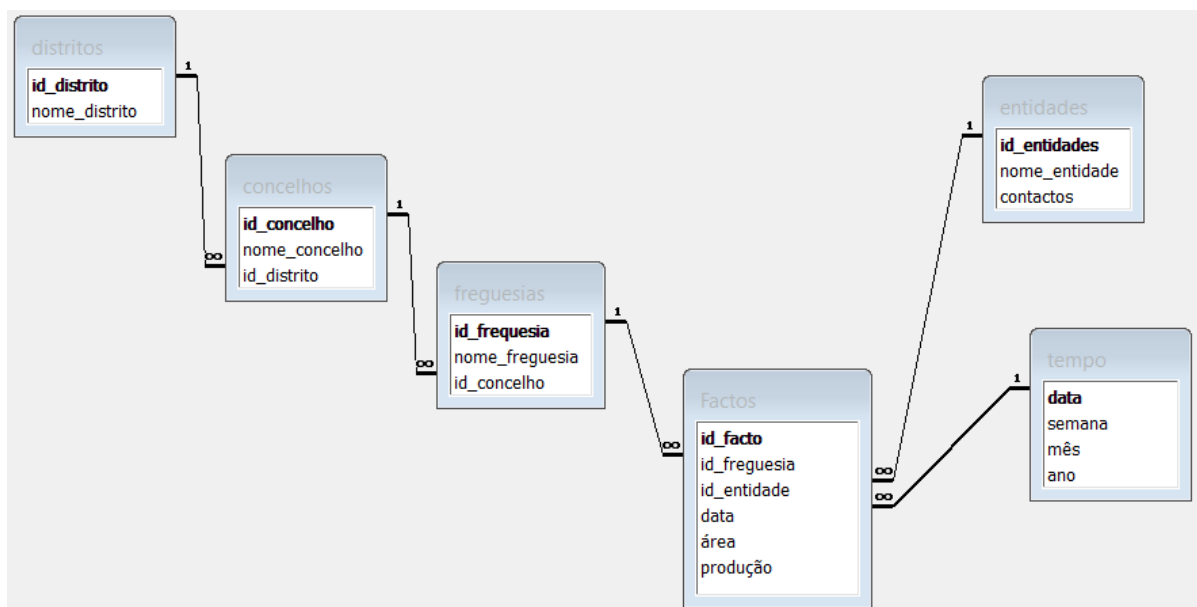
- (1,0) definição “checkpoint”
- (1,5) desempenho “checkpoint”

Grupo B – Prática em “Data Warehousing”

5. (2 valores) Um departamento do Estado pretende registar numa base de dados informação relativa a diversas explorações agrícolas. Sobre cada exploração agrícola pretende guardar a sua localização (freguesia, concelho e distrito), o nome da entidade responsável e a área ocupada por cada uma das suas principais produções. Cada freguesia, concelho e distrito possuem um código identificativo e um nome, sobre cada concelho deve ser registado o distrito a que pretende e sobre cada freguesia deve ser registado o concelho a que pertence. Existem explorações agrícolas de grande dimensão que ocupam mais de uma freguesia. Para cada entidade responsável deve estar registado com nome e contactos.

Pretendemos desenhar um “Data Warehouse” do seguinte sistema. Defina as tabelas de factos em primeiro lugar. De seguida, defina três dimensões para cada tabela de factos.

(Resposta: 1 página)



CrITÉrios de correção:

- criar DW com 1 tabelas factos com 3 dimensões
- penalização de 5 décimas na falta da dimensão tempo
- penalização para esquema mal desenhado
- penalização para atributos desadequados na tabela factos
- penalização para dimensões desadequadas
- penalização para ligações mal estabelecidas

FIM