

U.C. 21103

Sistemas de Gestão de Bases de Dados

13 de fevereiro 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 **vector 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) Para sistemas de gestão de bases de dados qual a melhor política de substituição das memórias temporárias ("buffers"): LRU (*least recently used*) ou MRU (*most recently used*)? Justifique a sua resposta.

(Resposta: 3/4 página, 1ª página)

Resposta:

- As políticas de substituição das memórias temporárias (buffers) pretendem minimizar o acesso a disco.

- Nos sistemas operativos geralmente utilizam a política LRU (*least recently used*) de substituição de blocos, i.e., se um bloco tem de ser substituído vamos remover o que foi utilizado menos recentemente.

- Nas bases de dados a estratégia LRU é uma má estratégia. Por exemplo, na junção de duas tabelas (empréstimo \bowtie cliente) o algoritmo "nested loop join" contém um ciclo interno e um ciclo externo:

```
01 For each tuple in empréstimo
02   For each tuple in cliente
03     If empréstimo.id_cliente=cliente.id_cliente
04       Begin ..... end
05   End for
06 End for
```

- No caso dos blocos com informação do cliente, depois de tratados, podem ser imediatamente removidos. Assim, para as bases de dados, a substituição dos blocos mais recentemente utilizados é a estratégia mais adequada, i.e., MRU (*most recently used*).

Manual: As políticas de substituição dos buffers, página 462 e seguintes.

CrITÉrios de correcção:

- 1,0 valores, referir a MRU
- 1,5 valores, justificação

2. (2,5 valores) Mostre como deriva uma das seguintes expressões relacionais utilizando regras de equivalência para satisfazer a igualdade: $\sigma_{\theta_1 \wedge \theta_2} (E1 \bowtie_{\theta_3} E2) = \sigma_{\theta_1} (E1 \bowtie_{\theta_3} (\sigma_{\theta_2} (E2)))$ onde θ_2 envolve só atributos de E2.

(Resposta: 1 página, 2ª página)

Resposta:

Utilizando as seguintes regras de equivalência:

Regra A: $\sigma_{\theta_1 \wedge \theta_2} (E) = \sigma_{\theta_1} (\sigma_{\theta_2} (E))$ uma operação de selecção conjuntiva pode ser decomposta por uma sequência de selecções individuais;

Regra B: $\sigma_{\theta_1} (E1 \bowtie_{\theta} E2) = (\sigma_{\theta_1} (E1)) \bowtie_{\theta} E2$ a operação de selecção distribuí sobre uma teta-junção quando todos os atributos de selecção θ_1 envolvem só os atributos da expressão E1.

Podemos encontrar as equivalências com as seguintes justificações:

$\sigma_{\theta_1 \wedge \theta_2} (E1 \bowtie_{\theta_3} E2)$	dado
$\sigma_{\theta_1} (\sigma_{\theta_2} (E1 \bowtie_{\theta_3} E2))$	aplicando a regra A
$\sigma_{\theta_1} (E1 \bowtie_{\theta_3} (\sigma_{\theta_2} (E2)))$	aplicando a regra B

Manual: Regras de equivalência, página 571 e seguintes.

Crítérios de correcção:

- 1,0 valores, derivação das equivalências
- 1,5 valores, justificação das regras

3. (2,5 valores) Dadas as transacções T1 e T2, acrescente as instruções "lock" e "unlock" às transacções de forma a aplicar o protocolo 2PL.

T1: read(A); read(B); if A=0 then B:=B+1; write(B);

T2: read(B); read(A); if B=0 then A:=A+1; write(A);

(Resposta: 1/2 página, 3ª página)

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);
--	--

Manual: Protocolos baseados em "locks", página 635 e seguintes.

Critérios de correcção:

- 1,5 valores, aplicação apropriada dos lock-S e lock-X

- 1,0 valores, aplicação apropriada dos unlock

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 imediata dos dados, no caso de "roll back" de uma transacção?

(Resposta: 1/2 página, 3ª página)

Se a transacção contem os registos <start> e <commit> é necessário refazer a transacção com o procedimento redo().

Se a transacção contem os registos <start> mas não contem o registo <commit> é necessário desfazer a transacção com o procedimento undo().

Manual: "Log based recovery: immediate database modification", página 694 e seguintes.

Critérios de correcção:

- 1,0 valores, transacção com registos <start> e <commit>

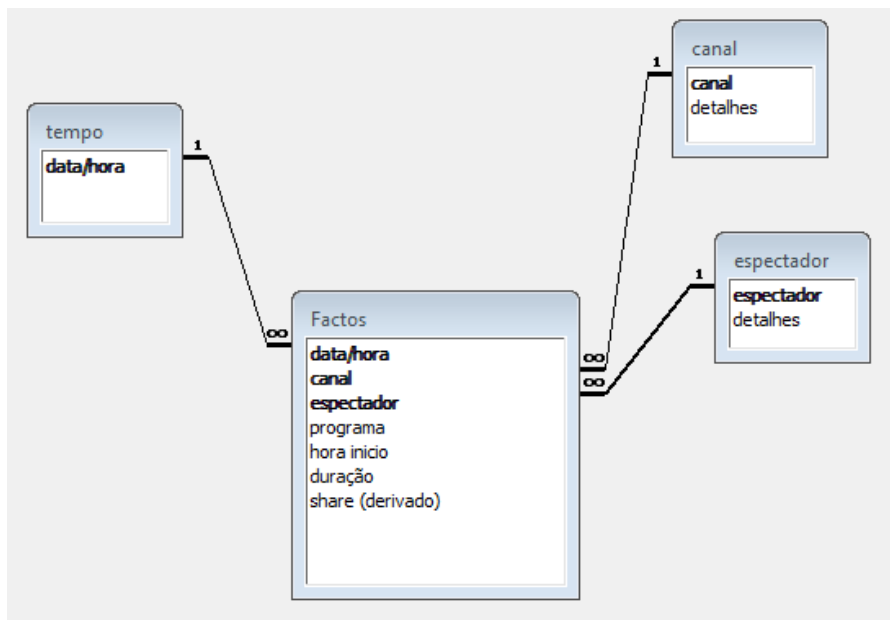
- 1,5 valores, transacção com registos <start> mas não contem o registo <commit>

Grupo B – Prática em “Data Warehousing”

5. (2 valores) Pretendemos desenhar um “Data Warehouse” para o 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.

"Uma entidade autónoma para cálculo e divulgação das audiências ("share") das estações de televisão pretende um novo sistema de informação que auxilie a armazenar a programação diária. Quem consultar a programação através de um "browser" deverá poder visualizar por cada canal, para cada dia, a sequência de programas, com indicação da hora de início, duração do programa e respectivo "share". Para calculo do "share" existe um ficheiro de grandes dimensões com a informação dos telespectadores seleccionadas, com registos a cada minuto, por forma a saber em cada minuto quantos espectadores estavam a visualizar um determinado canal."

(Resposta: 1 página, 4ª página)



Critérios de correcção:

- a tabela de Factos e as dimensões devem reflectir que para cada minuto existe um espectador a num determinado canal, com a chave concatenada <data/hora + canal + espectador> de forma a encontrar o "share" de cada canal a cada minuto;
- penalização de 0,5 valores por cada um dos elementos (data/hora, canal, espectador) em falta;
- penalização de 1,0 a 2,0 valores para dimensões descontextualizadas;

FIM