

U.C. 21103

Sistemas de Gestão de Bases de Dados

26 de janeiro de 2017

-- INSTRUÇÕES --

- O tempo de duração da prova de exame é de 150 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.
- Verifique no momento da entrega da(s) folha(s) 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 objetos pessoais deixados em local próprio da sala de exame.
- O enunciado da prova é constituído por 3 páginas, 9 perguntas 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

Grupo A – Sistemas de Bases de Dados

1. (2,5 valores) Índices servem para acelerar o processamento de consultas, mas geralmente é uma má ideia para criar índices em todos os atributos, e cada combinação de atributos, que sejam potencialmente utilizáveis numa pesquisa. Explique porquê.

2. (2,5 valores) Considere as relações $r_1(A, B, C)$, $r_2(C, D, E)$, e $r_3(E, F)$, com chaves primárias A , C , e E , respetivamente. Suponha que r_1 tem 1000 tuplos, r_2 tem 1500 tuplos, e r_3 tem 750 tuplos. Estimar o tamanho de $r_1 \bowtie r_2 \bowtie r_3$, e dar uma estratégia eficiente para a executar (nota: o operador \bowtie é o *natural join*).

3. (2,5 valores) Considere as seguintes duas transações:

T₁: read(A)
read(B)
if A>B then B:=B+A;
write(B).

T₂: read(B)
read(A)
if B<0 then A=2*A;
write(A).

Adicione instruções de *lock* e *unlock* às transações de modo a ficarem no protocolo *two-phase lock*. É possível a execução destas duas tarefas resultarem num *deadlock*? Justifique.

4. (2,5 valores) Descreva os benefícios inconvenientes de uma arquitetura orientada à fonte (*source-driven*) para recolha de dados em um *data warehouse*, em comparação com uma arquitetura orientada ao destino (*destination-driven*).

Grupo B – Prática em “Data Warehousing”

5. (2 valores) Considere uma base de dados de um clube de atletismo que possui uma pista de treino circular. Existe um sistema automático para alimentar uma base de dados com informações sobre os treinos dos atletas. Sempre que um atleta utiliza a pista, os tempos são registados por volta, sendo estes valores agrupados num período de treino se as voltas forem consecutivas. O registo de treino tem a distância associada, devido a ter o número de voltas. Em todos os regístros é guardado também a data, bem como a temperatura e percentagem de humidade, que são fornecidos também automaticamente por um sistema complementar. Cada atleta apenas entra na pista com a sua identificação, que permite o sistema associar o atleta correto a cada registo.

Pretendemos desenhar um “*Data Warehouse*” deste sistema. Defina as tabelas de factos em primeiro lugar e as tabelas de dimensões, bem como as respetivas associações.

Grupo C – Sistemas de Bases de Dados (Exame)

6. (2 valores) É importante obter-se rapidamente a localização de um bloco no buffer, no caso de estar em memória, ou saber-se que não está em memória. Dado que os buffers de base de dados são muito grandes, que estrutura de dados (na memória) aconselha para esta tarefa?

7. (2 valores) Considere a seguinte consulta de SQL (os campos ID são chave das respetivas tabelas):

```
SELECT curso.nome_curso, local.nome_local, Count(*) As inscritos  
FROM curso, local, inscrito  
WHERE inscrito.local_id=local.local_id  
    AND inscrito.curso_id=curso.curso_id  
GROUP BY local.nome_local, curso.nome_curso  
HAVING curso.nome_curso="Informática";
```

Escreva uma expressão eficiente em álgebra relacional, equivalente à consulta. Justifique.

8. (2 valores) Implementações de sistemas de base de dados têm dado muito mais atenção para as propriedades ACID implementações dos sistemas de ficheiros dos sistemas operativos. Porquê?

9. (2 valores) Na avaliação de um classificador, são utilizadas as seguintes medidas de qualidade: *accuracy*; *recall*; *precision*; *specificity*. Defina cada uma e diga se é possível um sistema obter *recall* 100% e *precision* menor que 100%. Caso seja possível, esse sistema teria alguma utilidade?

FIM