

”

Critérios de Correção

SISTEMAS DE GESTÃO DE BASES DE DADOS | 21103 | ÉPOCA RECURSO

Período de Realização: decorre 24-07-2025 desde 10:00 com 2:00 horas de duração e tolerância de 0:15.

Data de Limite de Entrega: decorre até 12:15 de Portugal Continental

Temática / Tema / Conteúdos: Sistemas de gestão de bases de dados

Objetivos: Reconhecer formas de armazenamento de dados e formas de otimização de consultas; reconhecer o sistema transacional e formas de recuperação de dados; reconhecer ambientes de Data Warehouse, Data Mining e Information Retrieval.

Trabalho a desenvolver: ambiente WISEflow, tipo FLOWlock

Critérios de avaliação e cotação: A cotação deste exame é de 200 pontos = 20 valores, pode encontrar as cotações parciais junto de cada pergunta. A interpretação das perguntas também faz parte da sua resolução, se encontrar alguma ambiguidade deve indicar claramente como foi resolvida. Critérios de avaliação gerais: (i) para a dificuldade de leitura (linhas cruzadas, letras com fontes desadequadas) a penalização é de 20% a 100%; (ii) para erros e omissões a penalização é de 20% a 100%.

Normas a respeitar: Deve redigir o seu E-fólio na Folha de Resolução disponibilizada na turma e preencher todos os dados do cabeçalho. Podem ser incluídas imagens e digitalizações de conteúdos produzido manualmente pelo estudante. Todas as páginas do documento devem ser numeradas. O documento A4 deve ser redigido em Times New Roman, tamanho de letra 12. O espaçamento entre linhas deve corresponder a 1,0 ou 1,5 linhas. Nomeie o ficheiro com o seu nome da plataforma. Finalmente deve gerar um PDF do documento. Deve carregar o referido ficheiro para a plataforma no dispositivo até à data e hora limite de entrega. Evite a entrega próximo da hora limite para se precaver contra eventuais problemas. O ficheiro a enviar não deve exceder 8 MB.

Votos de bom trabalho!

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

Questão: 1 2 3 4 5; 6 7 8 9 10

Cotação: 2 2 2 2 ; 2 2 2 2 2

Tópico 1 – Consultas e Transações

1. (2 valores) Otimização de consultas de um SGDB. Para a seguinte consulta desenhe duas árvores com os planos de execução de eficiência diferente, um otimizado e outro não otimizado.

```
select E.nome, U.nome, U.ects  
from estudante E, inscrito I, universidade U  
where E.id = I.estudante_id  
and I.universidade_id = U.id  
and U.nome = 'Universidade Aberta'
```

Critérios de correção:

- 1,0 valor, plano otimizado com Seleção nas folhas da árvore (antes das junções)
- 1,0 valor, plano não-otimizado com Seleção depois das junções
- erros, omissões, redundâncias: -20% a -100%

2. (2 valores) Concorrência no SGBD. Considere o seguinte sequenciamento r1(A), r2(B), r2(A), r1(B), w1(A), w2(B). O referido sequenciamento é serializável pelo 2PL com ‘locks’ exclusivos e partilhados? Justifique a resposta usando a seguinte tabela.

T1	T2	justificação

Critérios de correção:

- 1,0 valor, tabela com transações e justificações
- 1,0 valor, resposta final e justificação
- erros, omissões, redundâncias: -20% a -100%

3. (2 valores) Recuperação do SGBD. Considere o seguinte Log e preencha as colunas da tabela seguinte nas fases de Redo e Undo.

Log	Redo-Phase	Undo-Phase
Start T1		
Read T1		
Commit T1		
Checkpoint		
Start T2		
Read T2		
Start T3		
Checkpoint		
Write T2		
Commit T2		
Read T3		
Write T3		
System crash <-----		

Critérios de correção:

- 1,0 valor, tabela com coluna de redo-phase
- 1,0 valor, tabela com coluna de undo-phase
- erros, omissões, redundâncias: -20% a -100%

4. (2 valores) Qual a importância da otimização de consultas nas linguagens de programação declarativas como o SQL? Justifique a resposta e exemplifique com palavras suas.

Critérios de correção:

- 1,0 valor, resposta e justificação
- 1,0 valor, exemplos
- erros, omissões, redundâncias: -20% a -100%

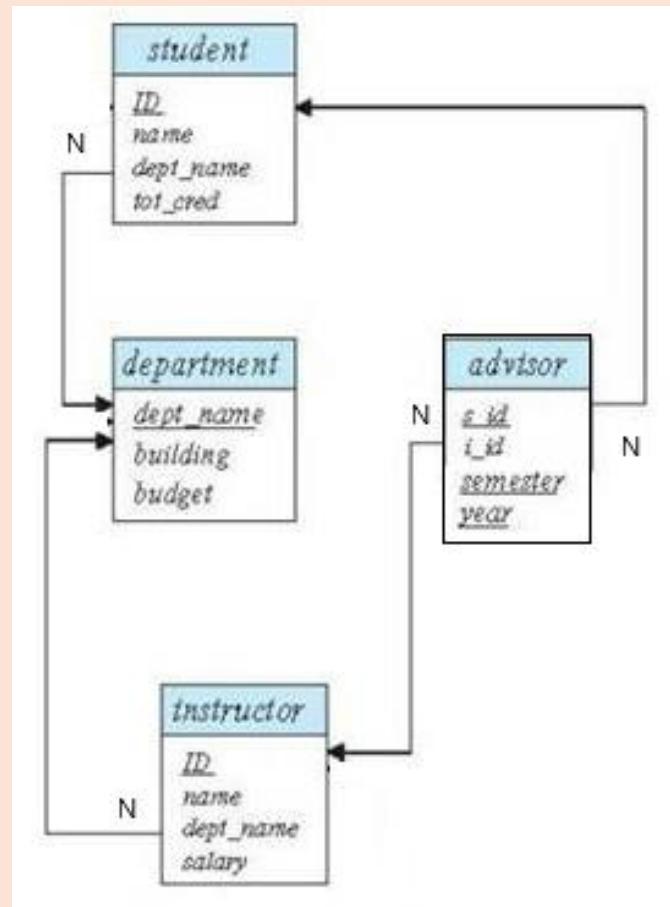
5. (2 valores) Qual é a diferença entre sistemas de bases de dados relacionais e NoSQL em relação à consistência dos dados? Justifique e exemplifique as vantagens e desvantagens com palavras suas.

Critérios de correção:

- 1,0 valor, definição, vantagens e desvantagens
- 1,0 valor, exemplos
- erros, omissões, redundâncias: -20% a -100%

Tópico 2 – Data Warehousing, Mining e Information Retrieval

6. (2 valores) Considere a seguinte base de dados que vai servir de fonte de dados a um “Data Warehouse”.



Considere ainda, as seguintes formas desnormalizadas (FDs):

- Antes da 1FD: constituída por várias poli-árvores interligadas;
- 1FD: constituída por uma poli-árvores separadas;
- 2FD: constituída por várias árvores separadas, depois da replicação das tabelas intermédias e de lookup para evitar caminho múltiplos; esta FD é equivalente ao esquema em estrela/floco-de-neve;
- 3FD: o processo de desnormalização termina com a junção de todas as tabelas da árvore com vista a uma rápida leitura dos dados.

Em que forma de desnormalização se encontra a bases de dados? Justifique a resposta e identifique os caminhos múltiplos no caso de existirem.

Critérios de correção:

- 1,0 valores, representação da 1FD
- 1,0 valores, identificação dos caminhos múltiplos
- erros, omissões, redundâncias: -20% a -100%

7. (2 valores) Considere a bases de dados da pergunta 4, desenhe um Data Mart, em estrela/floco-de-neve, com uma tabela de factos e com pelo menos três dimensões.

Considere a seguinte hierarquia de tipos de atributos de uma tabela:

- Aditivos: são atributos que podem ser agregados (somados) por todas as dimensões, ex: valor da venda (usar Sum() sempre)
- Semi-aditivos: são atributos que podem ser agregados (somados) por algumas as dimensões, ex: quantidade (usar Sum() em condições particulares)
- Não-aditivos: são atributos que não podem ser agregados (somados), ex: preço unitário (usar Average() por exemplo)
- Sem factos: só existem identificadores (usar a função Count() dos identificadores).

Para a tabela de factos defina a maior hierarquia do tipo de atributos que apresenta. Justifique a resposta.

Critérios de correção:

- 1,0 valor, tabela de factos com 3 dimensões, deve incluir o Tempo
- 1,0 valor, tipo de atributos: Sem Factos
- erros, omissões, redundâncias: -20% a -100%

8. (2 valor) Considere os dados das duas perguntas anteriores, e crie uma consulta OLAP (usando uma tabela de duas entradas) que evidencie o esforço (em número de estudantes) de cada professor (instructor) ao longo dos anos. Exemplifique a tabela de duas entradas e a consulta em SQL.

Critérios de correção:

- 1,0 valor, tabela pivot com 2 entradas
- 1,0 valor, código SQL
- erros, omissões, redundâncias: -20% a -100%

9. (2 valor) Information Retrieval

Explique o conceito de TF-IDF no contexto da Recuperação de Informação. Para que serve esta métrica e como ela contribui para a ordenação dos documentos? Justifique a resposta com um exemplo prático em palavras suas.

Critérios de correção:

- 1,0 valor, definição e utilidade
- 1,0 valor, exemplos
- erros, omissões, redundâncias ou apresentação desadequada: -20% a -100%

10. (2 valores) Data Mining

O que é *clusterização* em Data Mining? Descreva um algoritmo usado para essa tarefa e justifique sua aplicação com exemplos práticos.

Critérios de correção:

- 1,0 valor, definição
- 1,0 valor, exemplos
- erros, omissões, redundâncias ou apresentação desadequada: -20% a -100%

FIM