



CrITÉrios de Correção

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

Período de Realização: decorre 27-01-2025 deste 15:00 com 1:30 horas de duração e tolerância de 0:15.

Data de Limite de Entrega: até 16:45 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 e-fólio é de 120 pontos = 12 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

Cotação: 2 2 2 2 2 2

Tópico 1 – Consultas e Transações

1. (2 valores) Relativo ao tema das transações, considere o seguinte sequenciamento $r1(X)$, $r2(X)$, $r2(Y)$, $r1(Y)$, $r2(X)$, $w2(X)$, $w1(X)$, $w2(Y)$. O referido sequenciamento é serializável a conflitos? Justifique a resposta.

Resposta:

Serialização a conflitos: um escalonamento é seriável por conflitos se o seu grafo de precedência não contiver ciclos. Na construção do grafo de precedências as arestas são montadas a partir das observações das transações que participa, da escala sendo duas transações T_i e T_j haverá uma aresta: $T_i \rightarrow T_j$ se forem observadas as seguintes condições:

- T_i executa *write*(q) antes de T_j executar *read*(q)
- T_i executa *read*(q) antes de T_j executar *write*(q)
- T_i executa *write*(q) antes de T_j executar *write*(q)

Sequenciamento $r1(X)$, $r2(X)$, $r2(Y)$, $r1(Y)$, $r2(X)$, $w2(X)$, $w1(X)$, $w2(Y)$

Divisão por recurso/item:	As precedências são as seguintes:
Item X: R1, R2, W2, W1	R1->W2, R2->W1, W2->W1
Item Y: R2, R1, W2	R1->W2

Conclusão: Existe um ciclo, $T1 \rightarrow T2 \rightarrow T1$, pelo que o escalonamento não é seriável a conflitos.

Critérios de correção:

- 1,0 valor, divisão por recurso/item e precedências
- 1,0 valor, conclusão e justificação
- erros, omissões, redundâncias: -20% a -100%

2. (2 valores) Considere o sequenciamento da pergunta 1. O referido sequenciamento é serializável a vistas? Justifique a resposta.

Resposta:

Serialização por vistas: Diz-se que “ T_i lê x de T_j ”, se $W_j(x)$ for a última operação de escrita antes de $R_i(x)$. Dois escalamentos são equivalentes a vistas, se a relação “lê de” em ambas for a mesma.

Dois escalamentos S1 e S2 são equivalentes se:

- Todos os elementos A têm o mesmo valor inicial
Se S1: $r_i(A)$ lê o valor inicial, então S2: $r_i(A)$ também lê o mesmo valor inicial
- Um elemento A tem a mesma vista nos dois escalamentos
Se S1: $w_j(A) \Rightarrow r_i(A)$ então em S2: $w_j(A) \Rightarrow r_i(A)$; têm a mesma vista
- Todos os elementos A têm a mesma vista final
Se S1: T_i termina $w_i(A)$, então S2: T_i também termina com $w_i(A)$; os valores finais são os mesmos.

Sequenciamento $r1(X)$, $r2(X)$, $r2(Y)$, $r1(Y)$, $r2(X)$, $w2(X)$, $w1(X)$, $w2(Y)$

S1 (non-serial)			S2 (serial)		
	T1	T2		T1	T2
	$r(X)$			$r(X)$	
		$r(X)$		$r(Y)$	
		$r(Y)$		$w(X)$	
	$r(Y)$				$r(X)$
		$r(X)$			$r(Y)$
		$w(X)$			$r(X)$
	$w(X)$				$w(X)$
		$w(Y)$			$w(Y)$

	X	Y		X	Y
initial read	T1	T2		T1	T2
update read					
final write	T1	T2		T2	T2

S1 não é equivalente a S2(serial schedule), portanto não é seriável a vistas.

Conclusão: a sequência não é serializável a vistas, pois:

- S1 e S2 não são equivalentes
- As escritas finais de X não são consistentes.

Critérios de correção:

- 1,0 valor, tabelas de recursos versus operações RW
- 1,0 valor, conclusão e justificação
- erros, omissões, redundâncias: -20% a -100%

3. (2 valores) Considere o sequenciamento da pergunta 1 e 2. O referido sequenciamento é serializável pelo 2PL com 'locks' exclusivos e partilhados? Justifique a resposta.

Resposta:

O protocolo 2-PL (*Two-Phase Locking Protocol*), requer que cada transação recorra a bloqueios, garantindo assim a seriabilidade, para aceder aos recursos em duas fases:

- 1- Fase de Crescimento: Uma transação pode obter bloqueios, mas não pode libertar nenhum;
- 2- Fase de Encolhimento: Uma transação pode libertar bloqueios, mas não consegue obter nenhum bloqueio novo.

Este protocolo, apesar de garantir a serialização não garante a recuperação a falhas nem ao 'deadlocks'.

Em 2-PL usamos a seguinte matriz de compatibilidade de 'locks':

State of the lock	Lock request type →	
	Shared	Exclusive
Shared	Yes	No
Exclusive	No	No

T1	T2	T1	T2	justificação
r(X)		x-lock(X), read(X)		T1 faz x-lock(X)
	r(X)		read(X)	x-lock(X) não é possível, T2 waiting
	r(Y)		read(Y)	T2 waiting
r(Y)		s-lock(Y), read(Y)		s-lock(Y)
	r(X)		read(X)	T2 waiting
	w(X)		write(X)	T2 waiting
w(X)		write(X), unlock		T1 unlock
	w(Y)		x-lock(X), read(X)	T2 prossegue
			s-lock(Y), read(Y)	
			read(X)	
			write(X), unlock	

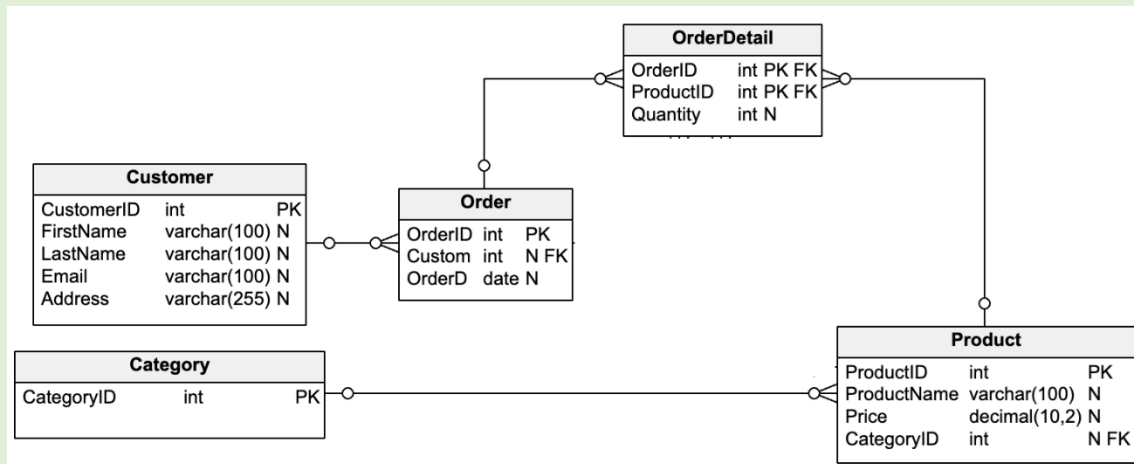
Conclusão: a sequência é serializável pelo 2PL

Critérios de correção:

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

Tópico 2 – Data Warehousing, Mining e Information Retrieval

4. (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):

- 1FD – constituída por uma poli-árvore, com a replicação das tabelas intermédias e de lookup que forem necessárias para evitar caminhos múltiplos;
- 2FD – constituída por várias árvores separadas, com a replicação das tabelas intermédias e de lookup que forem necessárias para evitar caminho múltiplos; esta FD é equivalente ao esquema em estrela ou ao esquema 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 usando a seguinte tabela e referindo os caminhos múltiplos

Antes 1 FD	
1FD	
2FD	
3FD	

Resposta:

Antes 1 FD	N.A.
1FD	N.A.
2FD	Não tem caminhos múltiplos, tem a estrutura de árvore: Customer -< Order -< Order_detail >- Product >- Category
3FD	N.A.

CrITÉrios de correção:

- 2,0 valores, representação da 2FD
- erros, omissões, redundâncias: -20% a -100%

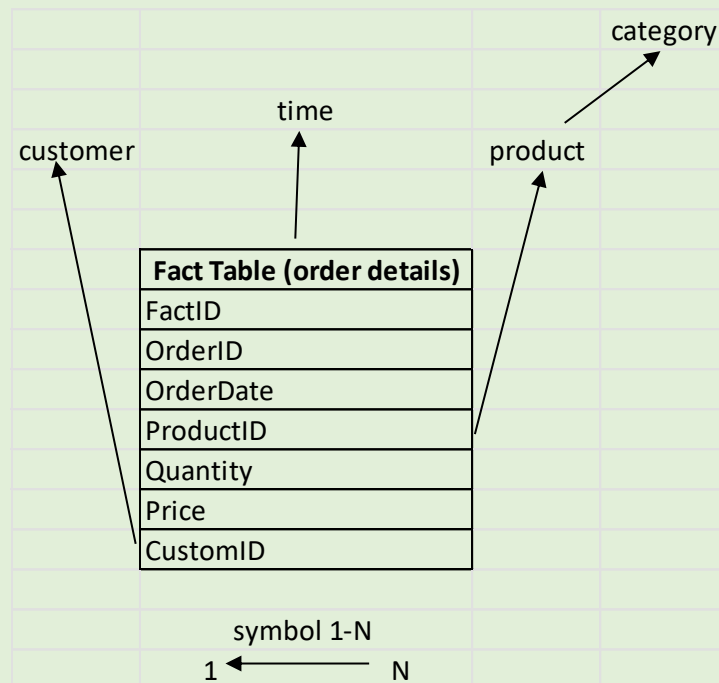
5. (2 valores) Considere a bases de dados da pergunta 4, desenhe um Data Mart, em estrela, com uma tabela de factos e com pelo menos três dimensões.

Considere os seguintes tipos de atributos da tabela de eventos:

- 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 o tipo de atributos que apresenta. Justifique a resposta.

Resposta:



Os atributos da tabela de Factos são Aditivos, dado que podemos encontrar um valor para cada FactID, usando o seguinte produto: $\text{Quantity} \times \text{Price}$.

Critérios de correção:

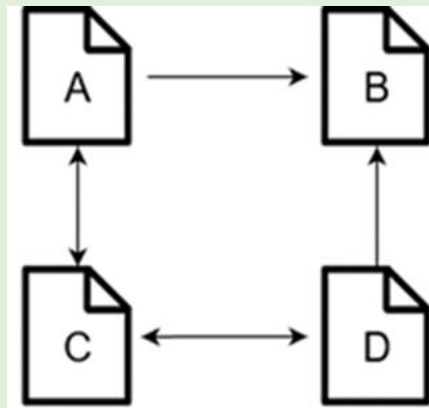
- 1,0 valor, tabela de factos com 3 dimensões, deve incluir o Tempo
- 1,0 valor, tipo de atributos Aditivos
- erros, omissões, redundâncias: -20% a -100%

6. (2 valor) Information Retrieval

Considere o algoritmo original de PageRank descrito por Lawrence Page and Sergey Brin em 1995 é dado por:

$$P[j] = \delta + (1 - \delta) * \sum_{i=1}^N (T[i,j] * P[i])$$

com $\delta = 0,5$. Encontre a ordenação das páginas para o seguinte conjunto. Justifique a resposta.



Resposta:

T[i,j]					
	A	B	C	D	soma
A	0	1/2	1/2	0	1
B	0	0	0	0	0
C	1/2	0	0	1/2	1
D	0	1/2	1/2	0	1

$$PR(A) = 0,5 + (1 - 0,5) * (1/2 * C)$$

$$PR(B) = 0,5 + (1 - 0,5) * (1/2 * A + 1/2 * D)$$

$$PR(C) = 0,5 + (1 - 0,5) * (1/2 * A + 1/2 * D)$$

$$PR(D) = 0,5 + (1 - 0,5) * (1/2 * C)$$

$$P[j] = \delta + (1 - \delta) * \sum_{i=1}^N (T[i, j] * P[i])$$

iteração	A	B	C	D
0	1	1	1	1
1	0,75	1	1	0,75
2	0,75	0,875	0,875	0,75
3	0,71875	0,875	0,875	0,71875
4	0,71875	0,859375	0,859375	0,71875
5	0,714844	0,859375	0,859375	0,714844
6	0,714844	0,857422	0,857422	0,714844
7	0,714355	0,857422	0,857422	0,714355
8	0,714355	0,857178	0,857178	0,714355
9	0,714294	0,857178	0,857178	0,714294
10	0,714294	0,857147	0,857147	0,714294
11	0,714287	0,857147	0,857147	0,714287
12	0,714287	0,857143	0,857143	0,714287
13	0,714286	0,857143	0,857143	0,714286
14	0,714286	0,857143	0,857143	0,714286
PageRank	2	1	1	2

Conclusão: $(B = C) > (A = D)$

Critérios de correção:

- 1,0 valor, tabela $T(i,j)$ e fórmulas
- 1,0 valor, resultado depois de algumas iterações
- erros, omissões, redundâncias ou apresentação desadequada: -20% a -100%

FIM