

Resolução e Critérios de Correção

U.C. 21103

Sistemas de Gestão de Bases de Dados

11 de fevereiro de 2020

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.
- 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 **3** 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

Cotação: 2 2 2 2 4

Grupo A – Sistemas de Bases de Dados

1. (2 valores) Relativamente ao armazenamento e estrutura de ficheiros, o que entende por dicionário de dados? Exemplifique.

(Resposta: 1 página)

O dicionário de dados contém os “dados acerca dos dados”, também conhecido por catálogo, ou recentemente referido como metadados. O dicionário de dados armazena:

- nomes das tabelas
- nomes, tipos e comprimento dos atributos
- regras de integridade
- informação dos índices
- tipos de armazenamento das tabelas e número de tuplos

Para exemplificar um Dicionário de Dados, de modo a garantir uma gestão eficaz, seria de implementar a seguinte regra para normalização dos nomes das aplicações, tabelas, atributos, índices e demais elementos do dicionário de dados:

- exemplo de aplicação, começa por G e tem 3 letras: Gcli - aplicação cliente
- exemplo de tabela, começa por T e tem 5 a 7 letras: Tclient - tabela clientes
- exemplo dos atributos, começa por e tem 5 a 7 letras: Anumero - número

Assim, GcliTclient.Anumero corresponde ao número da tabela cliente da aplicação cliente.

Critério de correção:

- 1 valor para definição
- 1 valor para exemplo
- erros, omissões ou redundâncias: -20% a -100%
- nota: muitos estudantes confundiram dicionário de dados com índices

2. (2 valores) Na otimização de consultas de um SGDB que informação precisa para melhorar a seguinte consulta? Estime valores aleatórios para a informação em falta e desenhe dois planos de execução.

Tabela r	Nr tuplos
disciplina	150
inscrito	10.000
aluno	1.000
aluno Lisboa	400

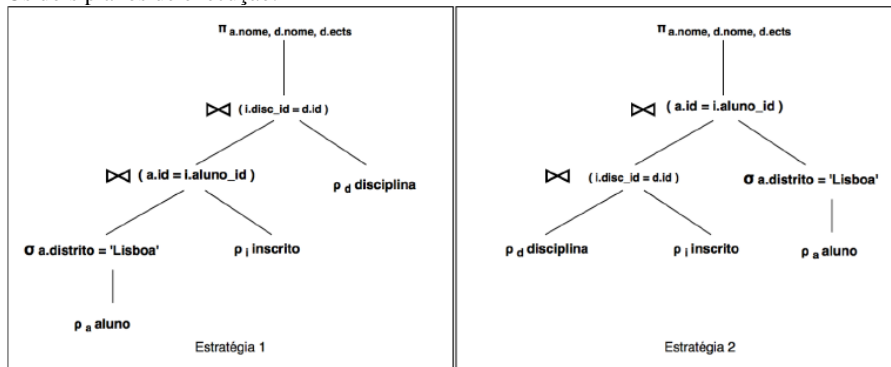
```
select a.nome, d.nome, d.ects
from aluno a, inscrito i, disciplina d
where a.id=i.aluno_id
and i.disc_id=d.id
and a.district='lisboa'
```

(Resposta: 1 página)

Informação necessária para otimizar a consulta:

- número de tuplos/linhas das tabelas Nr;
- número de valores distintos das colunas envolvidas nas junções V(A,r);
- a existência de histogramas com a distribuição dos dados;
- a existência de um índice em distrito e valores da seleção distrito='lisboa';

Os dois planos de execução:



Plano 1

	aluno_Lisboa >< inscrito	N	V(id-aluno)
s	aluno_Lisboa	400	200
r	inscrito	10.000	5.000
j1	Nr*Ns/V(A,s)	20.000	
j2	Nr*Ns/V(A,r)	800	
	Custo = min(j1,j2)	800	

Plano 2

	disciplina >< inscrito	N	V(id-disc)
s	disciplina	150	75
r	inscrito	10.000	5.000
j1	Nr*Ns/V(A,s)	20.000	
j2	Nr*Ns/V(A,r)	300	
	Custo = min(j1,j2)	300	

	[aluno_Lisboa >< inscrito] >< disciplina	N	V(id-disc)
s	aluno_Lisboa >< inscrito	800	400
r	disciplina	150	75
j1	Nr*Ns/V(A,s)	300	
j2	Nr*Ns/V(A,r)	1.600	
	Custo = min(j1,j2)	300	

	[disciplina >< inscrito] >< aluno_Lisboa	N	V(id-aluno)
s	disciplina >< inscrito	300	150
r	aluno_Lisboa	400	200
j1	Nr*Ns/V(A,s)	800	
j2	Nr*Ns/V(A,r)	600	
	Custo = min(j1,j2)	600	

Custo Plano 1	1.100
---------------	-------

Custo Plano 2	900
---------------	-----

Os valores $V(A, r)$ foram estimados em metade do tuplos da relação r .
Para determinar o custo foi aplicada a expressão: $\min [Nr*Ns/V(A,r); Nr*Ns/V(A,s)]$
Os planos de execução: melhor é o plano 2;

Critério de correção:

- 0,5 valores, informação necessária
- 1,5 valores, escolha e cálculo dos dois planos
- erros, omissões, redundâncias: -20% a -100%

3. (2 valores) Relativamente à gestão de transações, considere o seguinte:

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

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

Os valores iniciais são A=0 e B=0. Existe ainda um requisito adicional de consistência da base de dados: A=0 ou B=0. Mostre e exemplifique que qualquer execução com as duas transações preserva a consistência da base de dados.

(Resposta: 1 página)

Vamos considerar as transações T1 e T2 isoladamente.

T1	T2
r(A)	
r(B)	
if A=0: B++	
w(B)	
	r(B)
	r(A)
	if B=0: A++
	w(A)

Existem 2 execuções possíveis que preservam a regra de consistência (A=0 ou B=0) que só dependem da transação inicial:

- execução T1, * (seguida de outra qualquer transação T1 ou T2), A=0 e B<>0

Exemplo; T1, T1, T2, T2 obtemos A=0, B=2

- execução T2, * (seguida de outra qualquer transação T1 ou T2), A<>0 e B=0

Exemplo; T2, T1, T2, T2 obtemos A=3, B=0

Critério de correção:

- 1 valor, prova que qualquer execução das duas transações preserva a consistência

- 1 valor, explicação

- erros, omissões, redundâncias: -20% a -100%

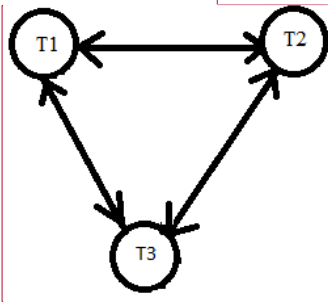
4. (2 valores) Considere a seguinte sequência: $r2(x)$, $w3(y)$, $w2(x)$, $w1(x)$, $r1(y)$, $r2(x)$, $w2(y)$, $w2(z)$, $w1(z)$, $w3(x)$. Desenhe o grafo de precedência e verifique se existe ciclicidade.

(Resposta: 1 página)

Dado: $r2(x)$, $w3(y)$, $w2(x)$, $w1(x)$, $r1(y)$, $r2(x)$, $w2(y)$, $w2(z)$, $w1(z)$, $w3(x)$.

Divisão por recurso/item:	As precedências são as seguintes:
Item x: $r2$, $w2$, $w1$, $r2$, $w3$	$w2 \rightarrow w1$, $w2 \rightarrow w3$, $w1 \rightarrow r2$, $w1 \rightarrow w3$, $r2 \rightarrow w3$
Item y: $w3$, $r1$, $w2$	$w3 \rightarrow r1$, $w3 \rightarrow w2$, $r1 \rightarrow w2$
Item z: $w2$, $w1$	$w2 \rightarrow w1$

Graficamente obtemos um grafo cíclico:



- Critério de correção:
- 0,5 valores, resposta de grafo cíclico
 - 1,5 valores, explicação detalhada
 - erros, omissões, redundâncias: -20% a -100%

Formatou: Cor do tipo de letra: Automática

Formatou: Cor do tipo de letra: Automática

Formatou: Cor do tipo de letra: Automática

Comentado [LMPSCS1]:

Grupo B – Prática em “Data Warehousing”

5. (4 valores) Considere a seguinte base de dados de aluguer de DVD. Pretendemos desenhar um “Data Warehouse” do seguinte sistema.

- a) Defina as tabelas de auxiliares (‘lookup’), intermédias e de factos em primeiro lugar.
- b) De seguida, defina pelo menos três dimensões para cada tabela de factos.

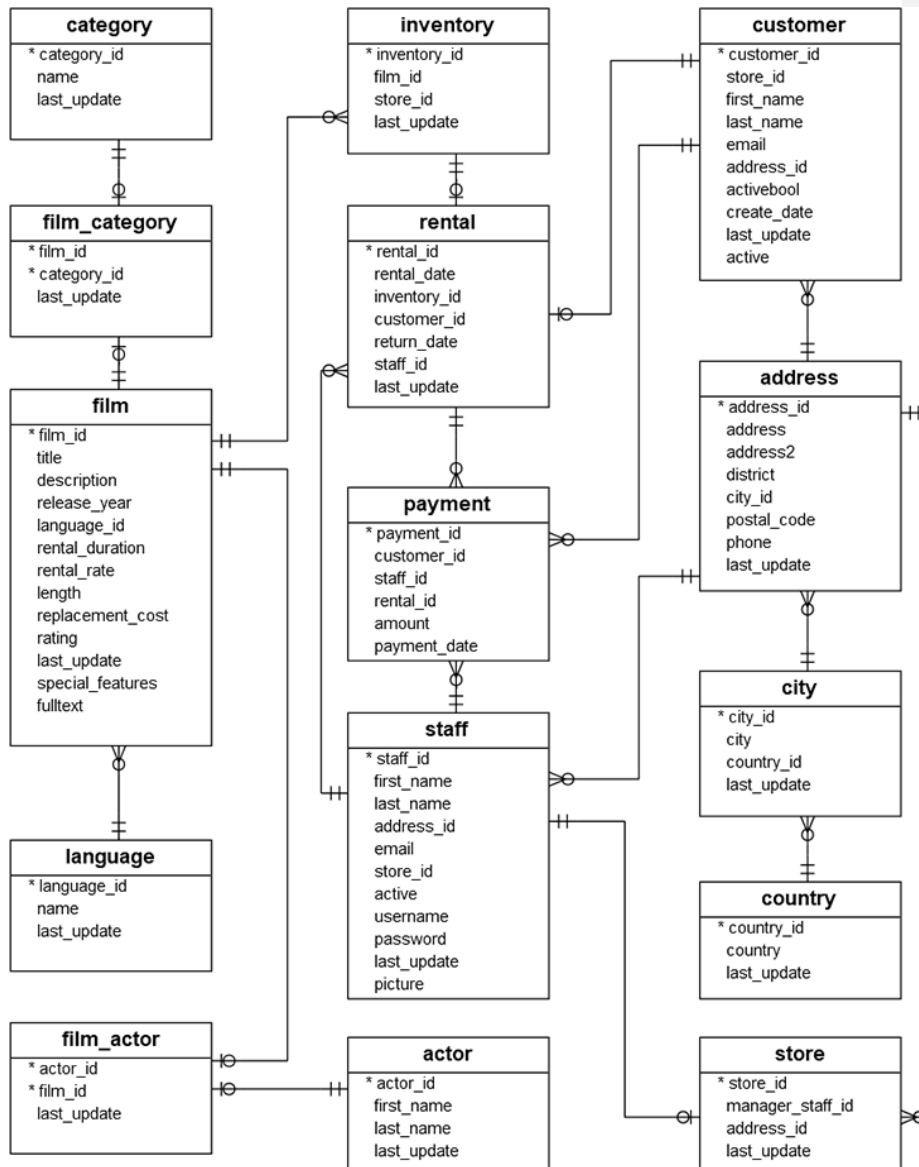
(Resposta: 1 página)

5.a)

table name	type of table
country	lookup
city	intermediate
address	intermediate
customer	intermediate
store	intermediate
staff	intermediate
rental	intermediate
payment	fact
actor	lookup
language	lookup
category	lookup
film	intermediate
inventory	intermediate
film_category	fact
film_actor	fact
film_text	fact

5.b)

table name	payment	film_category	film_actor	film_text
country	X			
city	X			
address	X			
customer	X			
store	X			
staff	X			
rental	X			
payment	additive			
actor			X	
language	X			
category		X		
film	X	X	X	X
inventory	X			X
film_category		without facts		
film_actor			without facts	
film_text				without facts
time	X	X	X	X



FIM