

**Resolução e Critérios de Correção**

**U.C. 21103**

**Sistemas de Gestão de Bases de Dados**

**31 de janeiro de 2019**

**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  
C: 2.5 2.5 2.5 2.5 2.0

## Grupo A – Sistemas de Bases de Dados

1. (2,5 valores) Relativamente ao armazenamento e estrutura de ficheiros, existem várias formas de organização dos registos em ficheiros. Quais as formas que conhece? Exemplifique.

(Resposta: 1 página)

Resposta:

Formas de organização de registos em ficheiros:

- *Heap* (ou pilha),
- sequenciais,
- *hashing* e
- agrupados (*clustered*).

Ficheiros ‘heap’: os registos podem ser armazenados em qualquer parte do ficheiro, desde que exista espaço para tal. Os registos não são ordenados.

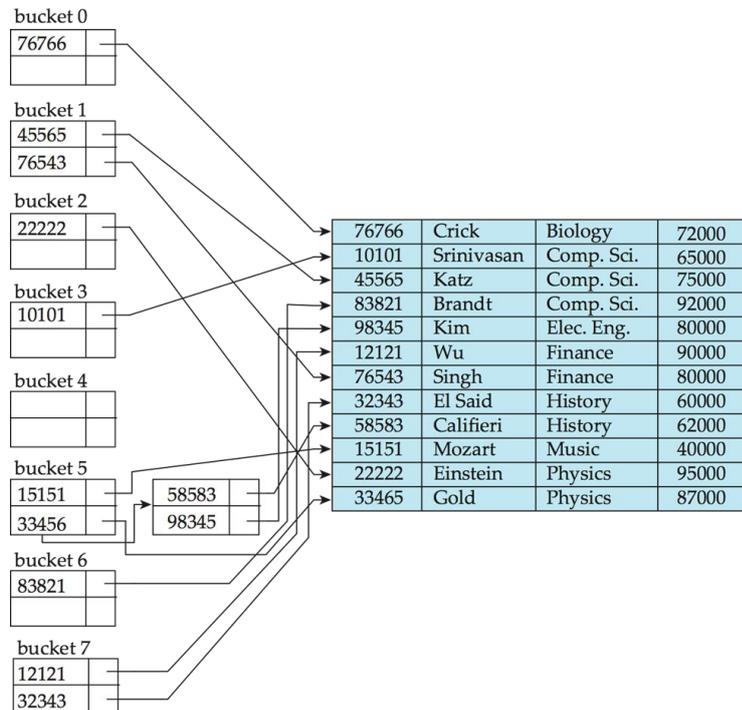


Ficheiros sequenciais: os registos são armazenados em ordem sequencial da chave de busca.

10101	Srinivasan	Comp. Sci.	65000	
12121	Wu	Finance	90000	
15151	Mozart	Music	40000	
22222	Einstein	Physics	95000	
32343	El Said	History	60000	
33456	Gold	Physics	87000	
45565	Katz	Comp. Sci.	75000	
58583	Califieri	History	62000	
76543	Singh	Finance	80000	
76766	Crick	Biology	72000	
83821	Brandt	Comp. Sci.	92000	
98345	Kim	Elec. Eng.	80000	
32222	Verdi	Music	48000	

The diagram shows a table of records with keys and values. Arrows indicate the sequential order of records based on their keys, showing that the records are not stored in the order they appear in the table.

Ficheiros ‘hash’: os registos são armazenados em “buckets” indexados por uma função ‘hash’.



Ficheiros agrupados: registos de várias tabelas diferentes podem ser armazenados no mesmo ficheiro; pode ser utilizado, quando uma junção é efetuada com frequência.

	<i>dept_name</i>	<i>building</i>	<i>budget</i>
<i>department</i>	Comp. Sci.	Taylor	100000
	Physics	Watson	70000

	<i>ID</i>	<i>name</i>	<i>dept_name</i>	<i>salary</i>
<i>instructor</i>	10101	Srinivasan	Comp. Sci.	65000
	33456	Gold	Physics	87000
	45565	Katz	Comp. Sci.	75000
	83821	Brandt	Comp. Sci.	92000

	<i>dept_name</i>	<i>building</i>	<i>budget</i>
multitable clustering of <i>department</i> and <i>instructor</i>	Comp. Sci.	Taylor	100000
	45564	Katz	75000
	10101	Srinivasan	65000
	83821	Brandt	92000
	Physics	Watson	70000

Os SGBD DB2 e PostgreSQL permitem apenas ficheiros ‘heap’. Oracle usa ficheiros ‘heap’ por omissão.

Nota: figuras retiradas de <http://www.db-book.com/>

Critério de correção:

- 2.5 valores para as 4 formas de organização
- erros, omissões ou redundâncias: -20% a -100%

2. (2,5 valores) Na otimização de consultas de um SGDB que informação precisa para melhorar a seguinte consulta? Desenhe um plano de execução e estime valores aleatórios para a informação que pretende:

```
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.districto = 'lisboa'
```

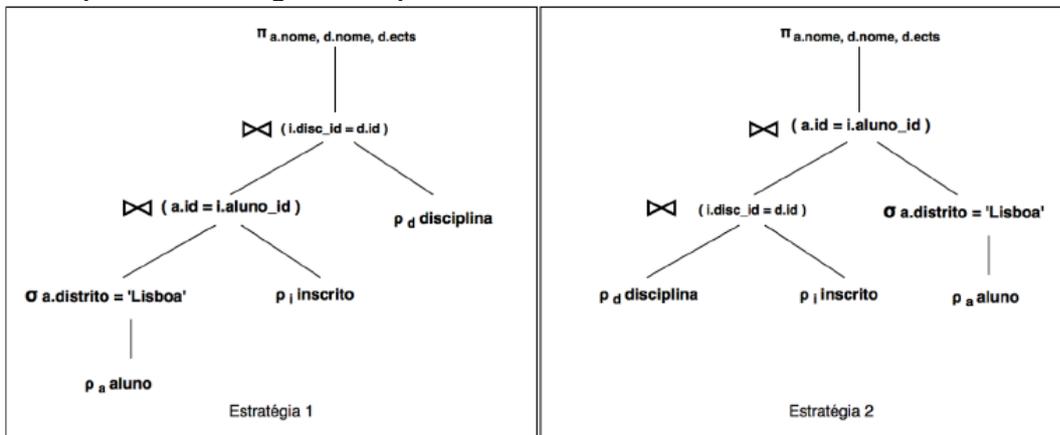
**(Resposta: 1 página)**

Resposta:

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';

Sendo possíveis os seguintes 2 planos:



Considerando os seguinte valores para Nr e V(A,r):

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

A dimensão estimada na junção:  $\min ( Nr*Ns/V(A,r), Nr*Ns/V(A,s) )$

		N	V(id-aluno)			N	V(id-disc)
s	aluno_Lisboa	400	400	s	disciplina	150	150
r	inscrito	10.000	1.000	r	inscrito	10.000	150
j1	$Nr*Ns/V(A,s)$	10.000		j1	$Nr*Ns/V(A,s)$	10.000	
j2	$Nr*Ns/V(A,r)$	4.000		j2	$Nr*Ns/V(A,r)$	10.000	
	Custo = $\min( j1,j2)$	4.000			Custo = $\min( j1,j2)$	10.000	
		N	V(id-disc)			N	V(id-aluno)
s	aluno_Lisboa  ><  inscrito	2.000	120	s	disciplina  ><  inscrito	10.000	1.000
r	disciplina	150	150	r	aluno_Lisboa	400	400
j1	$Nr*Ns/V(A,s)$	2.500		j1	$Nr*Ns/V(A,s)$	4.000	
j2	$Nr*Ns/V(A,r)$	2.000		j2	$Nr*Ns/V(A,r)$	10.000	
	dimensão = $\min( j1,j2)$	2.000			dimensão = $\min( j1,j2)$	4.000	
	Dimensão Plano 1	6.000			Dimensão Plano 2	14.000	

O Plano 1 apresenta uma menor dimensão estimada nas junções.

Critério de correção:

- 1,5 valores, informação necessária para otimizar a consulta
- 1,0 valores, cálculos de um plano execução
- erros, omissões ou redundâncias : -20% a -100%

3. (2,5 valores) Relativamente ao controlo de concorrência, defina o protocolo 2-PL. Quais as vantagens e desvantagens deste protocolo?

**(Resposta: 1 página)**

Resposta:

O protocolo 2PL (2-phase locking) é um algoritmo de bloqueio utilizado para o controlo de concorrência entre transações. Os algoritmos de bloqueio são os mais utilizados nos SGBD e, entre eles, o 2PL é o mais aplicado.

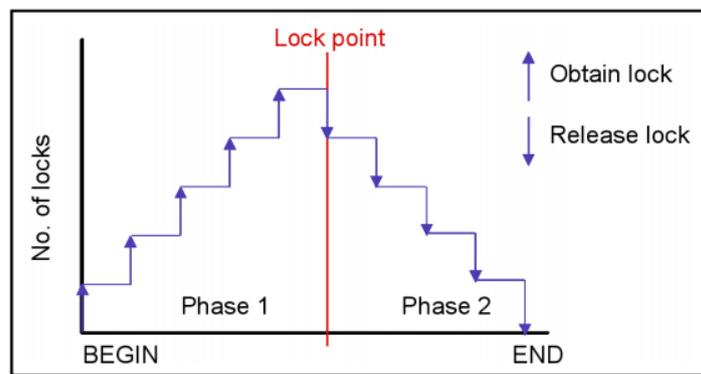
O protocolo 2PL utiliza dois tipos de bloqueios (locks):

- Bloqueio partilhado (S-lock) (utilizado nas operações de leitura (Read)): o item de dados pode ser partilhado por várias transações.
- Bloqueio exclusivo (X-lock) (utilizado nas operações de escrita (Write)): o item de dados não pode ser partilhado por várias transações.

O protocolo 2PL exige que todas as transações solicitem todos os bloqueios que necessitem antes de libertar qualquer um dos bloqueios que detenha. Desta forma a gestão de bloqueios é realizada em duas fases:

- Fase de crescimento ou expansão: a transação apenas pode adquirir bloqueios.
- Fase de encolhimento ou contenção: a transação apenas pode libertar bloqueios.

O ponto de mudança de fases é designado de ponto de bloqueio (lock point).



• Vantagens:

- produz escalonamentos serializáveis
- de fácil implementação

• Desvantagens:

- os pedidos de 'locks' são problemáticos para o 'lock manager'
- limita a concorrência, colocando em transações em fila de espera
- não evita a ocorrência de 'deadlock'
- não evita recuperações em cascata ('cascading rollback')

**Critério de correção:**

- (1,5) definições protocolo 2-PL
- (1,0) vantagens e desvantagens
- erros, omissões ou redundância: -20% a -100%

4. (2,5 valores) Relativamente ao controlo de concorrência, considere o protocolo 2-PL e explique detalhadamente a execução das seguintes transações: W1(x), R2(y), R1(x), Commit\_1, R2(x), W2(y), Commit\_2.

**(Resposta: 1 página)**

Resposta:

T1	T2	Gestão de 'Locks'
X-Lock(X)		Exclusive lock(T1, X) permite o acesso exclusivo da transação T1 ao item X
W(X)		
	S-Lock(Y)	Shared lock (T2, Y) permite o acesso partilhado da transação T2 ao item Y
	R(Y)	
R(X)		
UNLOCK(X)		Liberta lock(x)
Commit		
	S-Lock(X)	Shared lock (T2, X) permite o acesso partilhado da transação T2 ao item X
	R(X)	
	X-Lock(Y)	Exclusive lock(T2, Y) permite o acesso exclusivo da transação T2 ao item Y, fase de crescimento
	W(Y)	
	UNLOCK(X)	Liberta lock(X), fase de decrescimento
	UNLOCK(Y)	Liberta lock(Y)
	Commit	

Critérios de correção:

- (1,5) tabela com transações e locks
- (1,0) considerar as fases crescimento e decrescimento
- erros, omissões ou redundâncias: -20% a -100%

## Grupo B – Prática em “Data Warehousing”

5. (2 valores) O principal negócio de uma pizzaria é servir pizzas, para além das bebidas e sobremesas.



- Cada pizza tem um nome, um tipo de massa (massa fina, massa alta, etc.) e um preço. O preço da pizza é diferente se a pizza for pequena, média, ou grande. Certas pizzas só estão disponíveis em alguns tamanhos e no futuro poderá haver outros tamanhos também (p.ex. familiar). Cada pizza é feita com um conjunto de ingredientes. Cada ingrediente tem um nome e preço.
- Para além das pizzas, a pizzaria também serve bebidas e sobremesas. Estas são identificadas por um nome. As bebidas têm um nome e preço, enquanto as sobremesas têm um nome, descrição e preço. Além disso, existem vários tipos de bebidas: refrigerantes, cerveja e vinho. Os refrigerantes têm um preço diferente se forem em lata ou garrafa. A cerveja também tem um preço para lata, garrafa e copo. O vinho tem preço para garrafa e 1/2 garrafa.
- Finalmente, existem menus pré-definidos que incluem uma pizza, uma bebida, e uma sobremesa específicas. Estes menus são conhecidos por um número (Menu 1, Menu 2, etc.) e têm um preço. O cliente também pode encomendar um menu diferente, composto por quaisquer alimentos; nesse caso o preço do menu é calculado com base no preço dos produtos escolhidos.
- Os produtos são encomendados por clientes com número de contribuinte, nome, morada e telefone. Quando o cliente não apresenta o número de contribuinte este fica registado a zeros.
- O gerente da pizzaria pretende ter um registo exato de todas as vendas do seu negócio. Pretende saber que produtos/menus vendeu e a quem vendeu, por dia e por hora.

Pretendemos desenhar um “Data Warehouse” do seguinte sistema. Defina as tabelas de factos em primeiro lugar. De seguida, defina três dimensões para cada tabela de factos.

**(Resposta: 1 página)**

Resposta parcial:

1 tabela de factos (id, data, hora, nr\_contribuinte, nome\_produto, quantidade) e 3 dimensões: clientes, produtos, tempo

Critérios de correção:

- criar DW com 1 tabelas factos com 3 dimensões
- penalização de 5 décimas na falta da dimensão tempo
- penalização para esquema mal desenhado
- penalização para atributos desadequados na tabela factos
- penalização para dimensões desadequadas
- penalização para ligações mal estabelecidas
- erros, omissões ou redundância: -20% a -100%

**FIM**