UNIVERSIDADE ABERTA
www.uab.pt

Ministério da Educação e Ciência

p-Fólio

Resolução e Critérios de Correção

U.C. 21103

Sistemas de Gestão de Bases de Dados

13 de julho de 2015

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.
- Visto que o enunciado da prova não é utilizado para resposta, poderá ficar na posse do mesmo.
- 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 **2** 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: 25 25 25 25 20 décimas

Grupo A – Sistemas de Bases de Dados

1. (2,5 valores) 1) Na área de armazenamento de dados, o que entende por RAID (“redundant arrays of independente data”).

(Resposta: 1 página)

Com vista a melhorar o desempenho nos acessos a disco e aumentar a segurança (através de redundância) existem três atributos que diferenciam a classificação dos discos RAID: "striping", "mirroring" e paridade.

- Por "stripping" entende-se a segmentação em faixas com vista a melhorar o desempenho com múltiplos acessos a disco.
- Por "mirroring" entende-se que existe uma duplicação, cópia ou espelho.
- Por paridade é uma operação para deteção e correção de erros.

Para os vários tipos de RAID teremos a seguinte classificação segundo os critérios de desempenho e redundância:

	desempenho	**redundância**	
RAID	**"striping"**	**"mirroring"**	**paridade**
0	nível bloco		
1		existe espelho	
2	nível bit		dedicada
3	nível byte		dedicada
4	nível bloco		dedicada
5	nível bloco		distribuída
6	nível bloco		duplamente distribuída
híbridos			
0+1	nível bloco	existe espelho	
5+1	nível bloco	existe espelho	distribuída

Para além dos RAID de 0 a 6 podem ser combinados híbridos como os RAID "0+1" e o "5+1".

O RAID 5 é o RAID mais utilizado na indústria e caracteriza-se por um "stripping" ao nível dos blocos. Relativamente à redundância de dados, não tem "mirroring" e tem paridade distribuída.

A paridade do RAID 5 utiliza a função XOR, se um disco falha, os dados dos outros dois podem ser combinados e reconstruída a informação em falta.

Critério de correção:
- (1,5) definir e explicar os 3 atributos RAID
- (1,0) associar aos critérios de desempenho e redundância

2. (2,5 valores) Dado o seguinte sequenciamento ("schedule") que envolve as transações T1 e T2, o sequenciamento é conflito-serializável ("conflict serializable")? Justifique detalhadamente a sua resposta.

T1	T2
read(A) write(A)	
	read(A) read(B)
read(B) write(B)	

(Resposta: 1 página)

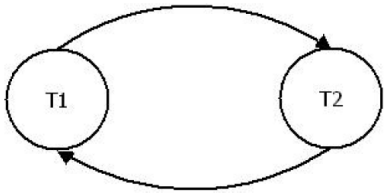
Na construção do grafo de precedências as arestas são montadas a partir das observações das transações que participa, da escala sendo duas transação Ti e Tj haverá uma aresta Ti -> Tj se forem observadas as seguintes condições:

1. Ti executa write(A) antes de Tj executar read(A)
2. Ti executa read(A) antes de Tj executar write(A)
3. Ti executa write(A) antes de Tj executar write(A)

No caso em questão:

write(A)	read(A)	T1-> T2
read(B)	write(B)	T2->T1

Graficamente obtemos um grafo cíclico:



Conclusão: o sequenciamento que envolve 2 transações **não é conflito-serializável**. Para a sequência (T1 -> T2) temos os conflito write(A) –> read(A) e para a sequência (T2 -> T1) encontramos o conflito read(B) –> write(B).

Critério de correção:
- (1,5) solução: não é conflito-serializável
- (1,0) justificação detalhada da resposta

3. (2,5 valores) O que entende por algoritmo ARIES, "Algorithms for Recovery and Isolation Exploiting Semantics", para recuperação de uma base de dados?

(Resposta: 1 página)

O "Algorithms for Recovery and Isolation Exploiting Semantics", ARIES, distingue-se pela utilização de um *Log* com "Log Sequence Number" (LSN), uma "Dirty Page Table" (DPT) e uma "Transaction Table" (TT).

- A "Dirty Page Table" (DPT) - contém uma entrada para cada página suja no *buffer* e o LSN correspondente à primeira atualização dessa página.

- "Transaction Table" (TT) - contém uma entrada para cada transação ativa, com informações sobre o ID de transação, o LSN do registo de *log* mais recente para a transação e o status da transação.

No ARIES distinguem-se três fases: análise, *redo* e *undo*.

- Na fase prévia de análise, são criadas as DPT e TT. O *log* é varrido desde o *checkpoint* até ao ponto de *crash*. Na DPT é registado o primeiro LSN para cada página e a TT o último LSN para cada transação. O ARIES também regista os *rollbacks* das transações através dos "Compensation Log Records" (CLR) i.e. que compensam as alterações de transações incompletas ou abortadas. Assim no *log* existem comandos de *write, commit, end, abort,* CLR (Compensation Log Record).

- O *Redo* inicia com o menor LSN da DPT. O *log* é varrido desde o menor LSN até ao ponto de *crash*, voltando a aplicar as alterações (*writes* ou CLR) do *log.*

- O *Undo* é iniciado no maior LSN da TT, o *log* é varrido do fim para o princípio. A instrução de CLR é acrescentada ao *log* para cada ação de desmontagem.

Os resultados na recuperação de cada transação da base de dados são de *Commit* ou *Rollback.*

Critérios de correção:
- (1,0) componentes ARIES
- (1,5) fases ARIES

4. (2,5 valores) Em "Information Retrieval" o que entende por "PageRank"? Qual a forma de calcular esta métrica?

(Resposta: 1 página)

O *PageRank* é o algoritmo que permite calcular o "valor" de uma página na Web. O valor da página não depende apenas da quantidade de *links* apontados para ela, mas do "valor" das páginas que apontam para ela.

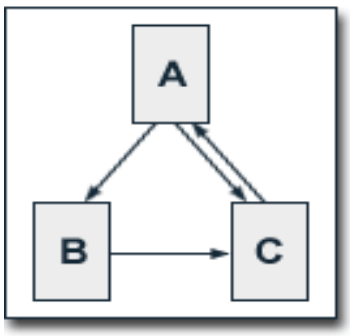
O algoritmo original de PageRank descrito por Lawrence Page and Sergey Brin em 1995 é dado por:

PR(A) = (1-d) + d (PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))

onde

- PR(A) é o PageRank da página A,
- PR(Ti) é o PageRank das páginas Ti que estão ligadas (apontam) para a página A,
- C(Ti) é o número de apontadores ("outbound links") na página Ti
- d é o fator de amortecimento que varia em 0 e 1.

Exemplo:



Seja d=0.5,
PR(A) = 0.5 + 0.5 (PR(C) /1)
PR(B) = 0.5 + 0.5 (PR(A) / 2)
PR(C) = 0.5 + 0.5 (PR(A) / 2 + PR(B) /1)

Resolvendo o sistema de 3 equações e 3 incógnitas obtemos os seguintes PR:
PR(A) = 14/13 = 1.07692308
PR(B) = 10/13 = 0.76923077
PR(C) = 15/13 = 1.15384615

Critério de correção:
- (1,00) definição
- (1,50) forma de cálculo

Grupo B – Prática em "Data Warehousing"

5. (2 valores) Pretendemos desenhar um "Data Warehouse" do seguinte sistema. Defina a tabela de factos em primeiro lugar. De seguida, defina três dimensões para o "Data Warehouse" e apresente a tabela de factos associada às três dimensões.

O aeroporto da Portela resolveu organizar a sua informação num sistema de bases de dados relativa à formação das equipas de manutenção das aeronaves.

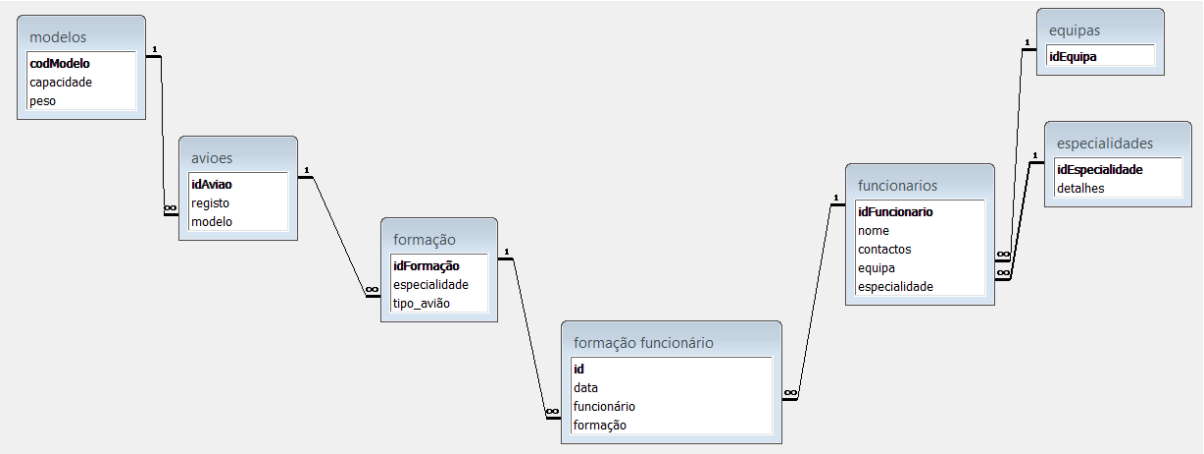
- Cada avião tem um número de registo e cada avião é de um modelo específico. O aeroporto pode acolher um certo número de modelos de aviões e cada modelo tem um código de modelo (ex. Airbus320, Boeing747), bem como uma capacidade e um peso.

- Cada equipa da manutenção tem um ou mais chefes, vários "aviónicos" (verificação de peças), vários mecânicos, vários técnicos de manutenção (combustível, etc). Cada funcionário da manutenção deve estar registado com nome e contactos.

- Cada funcionário tem formação e é avaliado regularmente na sua especialidade e para cada tipo de avião.

(Resposta: 1 página)

A base de dados correspondente aos requisitos definidos terá o seguinte aspeto, onde se distinguem o registo da formação dos funcionários.



Dado que pretendemos evitar "connection traps", teremos uma tabela de factos relativa à formação.

Critérios de correção:
- criar DW com 1 tabelas de factos com informação do funcionário, da formação (avião e especialidade) e da data da formação
- penalização até 50% para esquema mal desenhado
- penalização até 50% atributos desadequados na tabela factos
- penalização até 50% dimensões desadequadas
- penalização até 50% ligações mal estabelecidas

FIM