

Resolução e Critérios de Correção

U.C. 21103

Sistemas de Gestão de Bases de Dados

18 de julho de 2014

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

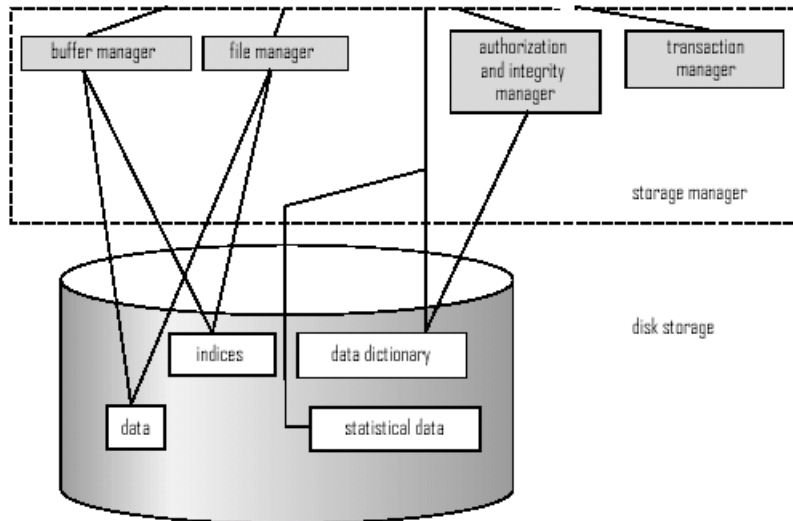
C: 25 25 25 25 20 décimas

Grupo A – Sistemas de Bases de Dados

1. (2,5 valores) O que é o Dicionário de Dados de um SGDB? Que informação contém? Qual a importância na organização?

(Resposta: 1 página)

O Dicionário de Dados é o conjunto metadados (dados acerca dos dados) do SGDB, que está geralmente associado ao nível do armazenamento em disco.



O Dicionário de Dados contém de forma normalizada os nomes dos diferentes elementos dos SGDB: atributos, tabelas, índices, vistas, utilizadores, programas, gatilhos, “procedures” e demais elementos. Por exemplo MRK-CLI-NOM, pode significar o atributo nome da tabela de clientes da aplicação/área de marketing, normalizado com 3 letras para a área, 3 letras para a tabela e 3 letras para o atributo.

Numa pequena base de dados não existe necessidade de normalização. Contudo, numa organização com várias áreas, utilizadores, tabelas e atributos é importante definir normas sintáticas para os nomes dos elementos, em que o Dicionário de Dados contém um vocabulário comum a todos os colaboradores.

Critérios de correção:

5 décimas: o que é

10 décimas: informação que contém

10 décimas: importância na organização

2. (2,5 valores) De quantas formas é possível realizar $r1 \bowtie r2 \bowtie r3$? Enumere as diferentes formas e explique como um SGBD deve escolher umas delas.
(Resposta: 1 página)

Existem $(2 \cdot (n-1))! / (n-1)!$ formas diferentes de juntar n tabelas. Para $n=3$ a enumeração das 12 formas possíveis é apresentada de seguida:

$r1 \bowtie (r2 \bowtie r3)$; $r2 \bowtie (r1 \bowtie r3)$; $r3 \bowtie (r1 \bowtie r2)$
 $r1 \bowtie (r3 \bowtie r2)$; $r2 \bowtie (r3 \bowtie r1)$; $r3 \bowtie (r2 \bowtie r1)$
 $(r2 \bowtie r3) \bowtie r1$; $(r1 \bowtie r3) \bowtie r2$; $(r1 \bowtie r2) \bowtie r3$
 $(r3 \bowtie r2) \bowtie r1$; $(r3 \bowtie r1) \bowtie r2$; $(r3 \bowtie r1) \bowtie r3$

Os SGBD para criar os planos de junção de tabelas não recorrem à enumeração completa. Os planos de junção são realizados por algoritmos eficientes de Programação Dinâmica ou através de abordagens heurísticas.

Critérios de correção:

10 décimas: enumeração formas diferentes

15 décimas: escolha realizada pelo SGBD

3. (2,5 valores) O que entende por ACID num sistema de base de dados transacional?
Exemplifique falhas das propriedades de atomicidade e consistência.
(Resposta: 1 página)

ACID é o acrónimo de quatro propriedades: Atomicidade, Consistência, Isolamento e Durabilidade.

(a) Atomicidade - a transação executa todas as ações ou pelo contrário nenhuma delas, isto é, não existem execuções parciais das transações. A ideia de atomicidade está associada à indivisibilidade; as transações são indivisíveis.

Para a transação T, que realiza uma transferência bancária, subtraindo 100 euros na conta A e adicionando 100 euros na conta B;

0001: T writes $A=A-100$

0002: T writes $B=B+100$

um exemplo de falha da atomicidade será a execução da primeira ação (T writes $A=A-100$) e a não execução da segunda ação (T writes $B=B+100$). Para este tipo de falha o SGDB deverá abortar a transação T.

(b) Consistência - a transação ao utilizar uma base de dados consistente deve deixá-la consistente no final da mesma transação. Sem o requisito da consistência, as quantidades de dinheiro podiam ser criadas ou destruídas numa transação. A responsabilidade da consistência recai sobre o programador da transação.

Para a transação T, que realiza uma transferência bancária, subtraindo 100 euros na conta A e adicionando 100 euros na conta B, um exemplo de falha de consistência seria

0001: T writes $A=A-100$

0002: T writes $B=B+150$

onde foi realizado um aumento de 50 euros sem qualquer explicação.

Critérios de correção:

5 décimas: o que é

10 décimas: atomicidade

10 décimas: consistência

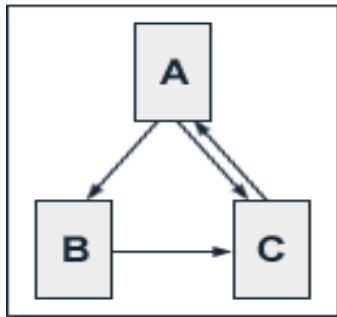
4. (2,5 valores) Em “Information Retrieval” o que entende por “PageRank”?
(Resposta: 1 página)

O algoritmo original de PageRank descrito por Lawrence Page and Sergey Brin em 1995 é dado por: $PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$

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:

$$PR(A) = 14/13 = 1.07692308$$

$$PR(B) = 10/13 = 0.76923077$$

$$PR(C) = 15/13 = 1.15384615$$

Fonte: <http://pr.efactory.de/e-pagerank-algorithm.shtml>

Critérios de Correção:

- explicação geral 15 décimas
- exemplo 10 décimas

Grupo B – Prática em “Data Warehousing”

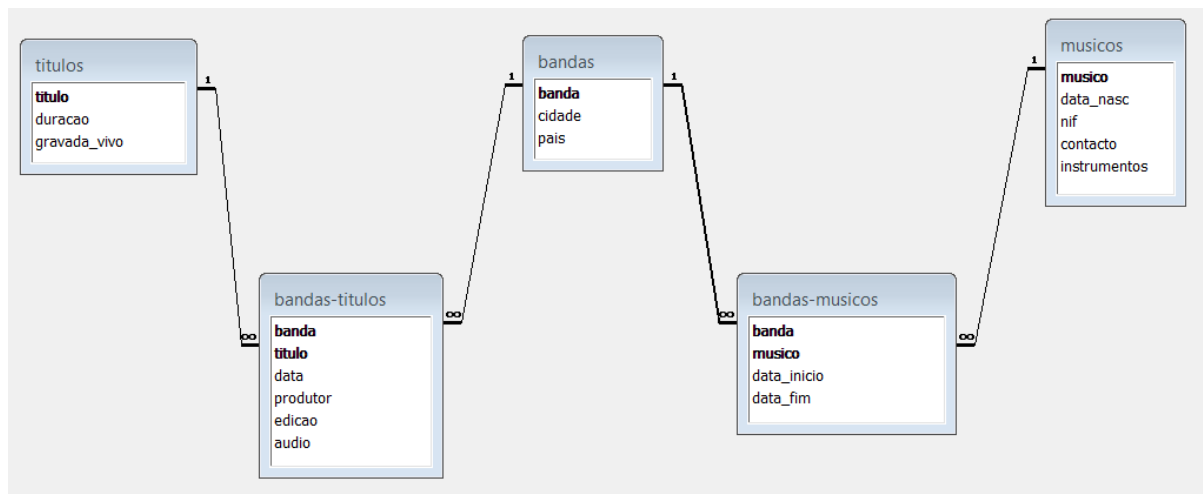
5. (2 valores) Pretendemos desenhar um “Data Warehouse” do seguinte sistema. Defina a tabela de factos em primeiro lugar. De seguida, defina três dimensões para o “Data Warehouse” e apresente a tabela de factos associada às três dimensões.

Considere que pretende criar uma base de dados para uma empresa discográfica, onde existem bandas musicais, com os atributos: nome e origem (cidade e país).

- Cada banda tem um ou mais músicos, com os atributos: nome, data de nascimento, número fiscal, contacto e instrumentos musicais que toca. Sabemos ainda que um músico pode tocar em várias bandas.
- As bandas criam músicas com os atributos: título, duração em segundo. Sabemos que podem existir músicas com o mesmo título tocadas por bandas diferentes. E que as músicas podem ser gravadas ao vivo ou em estúdio.
- As músicas são divulgadas ao público em edições áudio e vídeo. Uma edição pode ser um álbum, uma gravação isolada (single), ou um vídeo. É importante saber a data da edição e o nome do seu produtor.

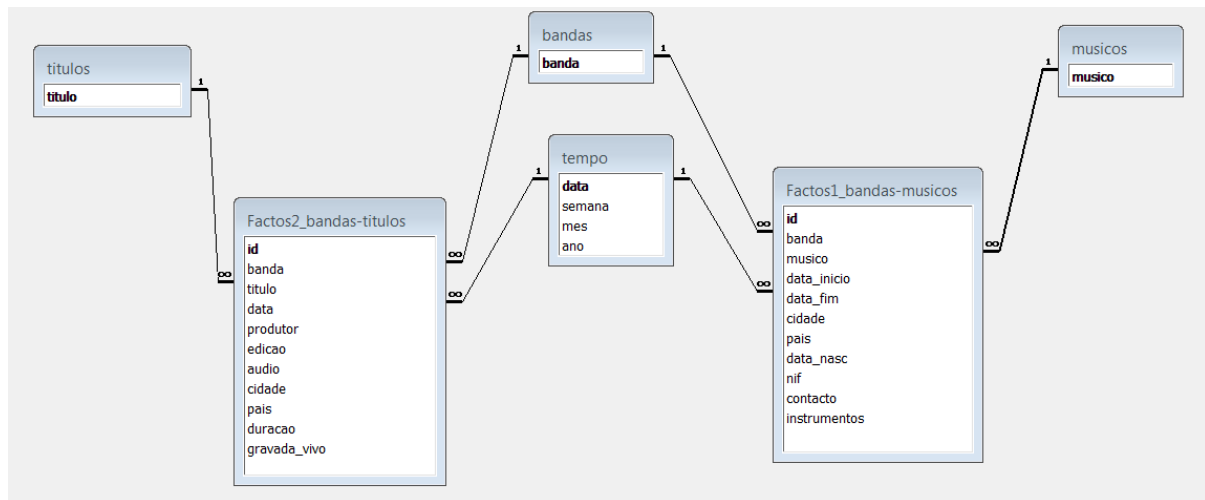
(Resposta: 1 página)

O esquema da base de dados original corresponde ao seguinte:

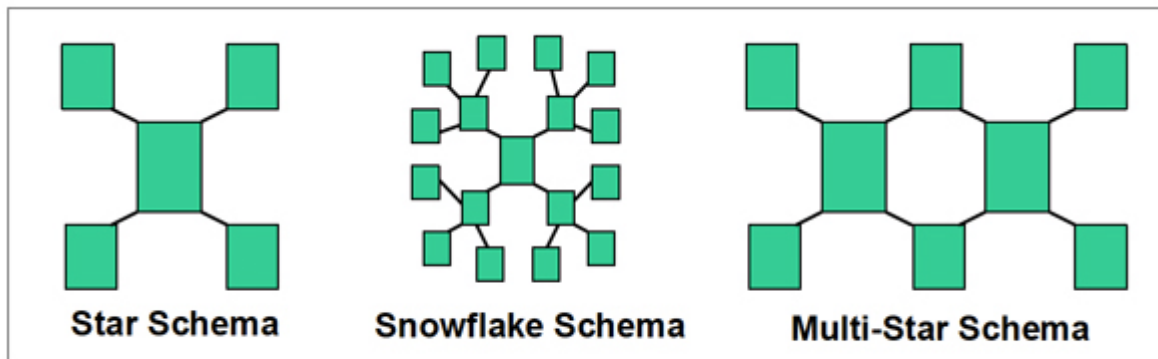


Para evitar junções N:N em (bandas-titulos |><| bandas |><| banda-musicos) deverão existir 2 tabelas de factos.

Assim obtemos o seguinte *Data Warehouse* com 2 tabela de factos associadas a três dimensões.



Os *Data Warehouses* (DW) com uma única tabela de factos são conhecidas por DW com esquema em estrela; com dimensões que têm mais de uma tabela são conhecidas por DW com esquema em floco-de-neve (snowflake); e quando têm mais do que uma tabela de factos são conhecidos como DW com esquema em constelações (ou multi-estrela);



Types of dimensional models

Critérios de Correção:

- 10 décimas: uma única tabela de factos
- 10 décimas: a segunda tabela de factos

FIM