

Critérios de Correção

# sisteMas de gestão de Bases de DAdos | 21103 | Época Recurso

## Período de Realização: decorre 21-07-2020 deste 10:00 com 3 horas de duração

## Data de Limite de Entrega: decorre 21-07-2020 até 13:00 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: Resolução de um conjunto de exercícios.

## 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 seu E-fólio não deve ultrapassar 1 página por pergunta. 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 número de estudante, seguido da identificação do E-fólio, segundo o exemplo apresentado: 000000efolioGlobal. Finalmente deve gerar um PDF do documento. Deve carregar o referido ficheiro para a plataforma no dispositivo E-fólio Global 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! Luís Cavique.

Vetor das perguntas: 1 2 3 4 5.1 5.2

Vetor das cotações: 2 2 2 2 2 2 somando 12 valores

**Grupo A – Sistemas de Bases de Dados**

**1.** (2 valores) Os índices aceleram o processamento das consultas. Contudo, não é boa prática a criação de índices para todos os atributos que são potencialmente pesquisáveis. Exemplifique a razão desta boa prática. Quais os atributos que são indexados automaticamente?

**Resposta:**

- As razões da referida boa prática são:

1. Cada índice extra requer espaço de armazenamento adicional.

2. Todo índice requer tempo adicional da CPU e sobrecarga de E/S do disco durante inserções e exclusões.

3. Os índices das chaves não-primárias podem precisar ser alterados nas atualizações, embora um índice na chave primária não possa ser (isso ocorre porque as atualizações normalmente não modificam os atributos da chave primária).

4. Para consultas que envolvem condições em várias chaves de pesquisa, a eficiência pode ser razoável, mesmo que apenas algumas das chaves tenham índices neles. Portanto, o desempenho da base de dados é degradado adicionando índices quando já existem muitos índices.

- Os atributos que são indexados automaticamente, na maior parte das bases de dados, são as chaves primárias. Para cada tuplo de uma chave primária que é inserido na base de dados são verificadas as regras de não-duplicação através do índice. Sem indexação a duplicação de tuplos é permitida.

Critério de correção:

- 1 valor razão da boa prática

- 1 valor índices criados automaticamente

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

**2.** (2 valores) Na otimização de consultas de um SGDB, considere que a estimativa do custo de junção é dado por Min[Nr\*Ns/V(A,r), Nr\*Ns/V(A,s)], em que Nx representa o número de tuplos de tabela x e que V(A,x) representa o número distinto de elementos do atributo A na tabela x.

**2.a)** Explique por palavras suas a expressão dada.

**2.b)** Considere que os valores V(A, x) foram estimados em 1/4 do tuplos da relação x. Calcule e desenhe o melhor e o pior plano de execução para a seguinte consulta com os seguintes números de tuplos:

|  |  |
| --- | --- |
| **Tabela** | **N tuplos** |
| disciplina | 50 |
| inscrito | 1000 |
| aluno | 300 |
| aluno Lisboa | 50 |

Select a.nome

From aluno a, inscrito i, disciplina d

Where a.id = i.aluno\_id

And i.disc\_id = d.id

And a.distrito ='Lisboa'

**Resposta:**

2.a) Sabendo que:

- Nx representa o número de tuplos de tabela x

- V(A,x) representa o número distinto de elementos do atributo A na tabela x

O custo do produto cartesiano é dado por Nr\*Ns. Contudo, a junção r|><|s é uma operação mais eficiente, permitindo que se divida o produto pelo número de tuplos do atributo que junta as duas tabelas, neste caso, A. Assim teremos custo\_junção = Nr\*Ns/V(A,r). Como devemos verificar o mesmo processo para V(A,s), teremos: custo\_junção= Min[Nr\*Ns/V(A,r), Nr\*Ns/V(A,s)].

Exemplo para as tabelas r e s com Nr=5 e Ns=9, com o atributo de junção id:

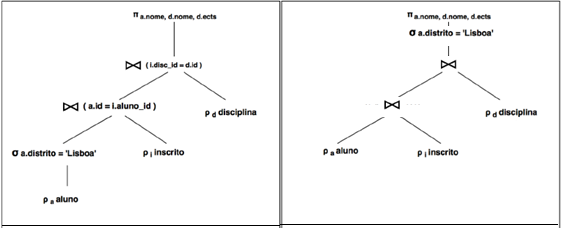


No exemplo o custo da junção r|><|s tem dimensão exatamente de 9 tuplos.

2.b) Iremos aplicar 1/4 do tuplos da relação *x* quando o número de tuplos for desconhecido:







Plano 1 com custo= 400 (um dos melhores) e plano 3 (pior) com custo =1200

Critério de correção:

- 2.a) 1,0 valor, explicação e exemplo da expressão

- 2.b) 1,0 valor, cálculo e escolha de dois planos

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

**3.** (2 valores) Considere a seguinte sequência: r2(x), r3(y), r2(x), r1(x), r1(y), r2(x), w2(y), w2(z), w1(z), w3(x). Desenhe o grafo de precedência e verifique se existe ciclicidade.

**Resposta:**

Para construir o grafo de precedências há que observar a sequência das operações para cada recurso/item. Assim, sendo duas transações Ti e Tj e um recurso X, será desenhada

uma aresta de Ti para Tj (Ti → Tj) num dos seguintes casos:

1. Se Ti executa Write(X) antes de Tj executar Read(X)

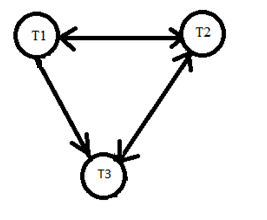
2. Se Ti executa Read(X) antes de Tj executar Write(X)

3. Se Ti executar Write(X) antes de Tj executar Write(X)

Dado: r2(x), r3(y), r2(x), r1(x), r1(y), r2(x), w2(y), w2(z), w1(z), w3(x).

|  |  |
| --- | --- |
| Divisão por recurso/item:  Item x: r2, r1, w3  Item y: r3, r1, w2  Item z: w2, w1 | As precedências são as seguintes:  r2 -> w3, r1 -> w3  r3 -> w2, r1 -> w2  w2 -> w1 |

Graficamente obtemos um grafo cíclico T1, T3, T2, T1:

****

Critério de correção:

- 0,5 valores, resposta de grafo cíclico

- 1,5 valores, explicação detalhada

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

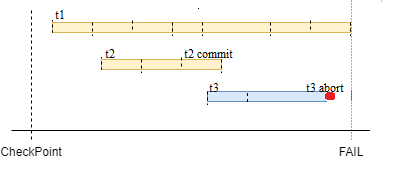
**4.** (2 valores)Relativamente à recuperação de sistemas, aplique o algoritmo de recuperação ao seguinte log:

|  |  |  |
| --- | --- | --- |
| LSN Log  00 check point  10 end check point  20 T1 writes P1  30 T2 writes P2 | 40 T3 writes P3  50 T2 commit  60 T3 writes P2  70 T2 end | 80 T1 writes P5  90 T3 abort  --- CRASH |

Represente as transações na linha do tempo e acrescente os registos na recuperação.

**Resposta:**

|  |  |  |
| --- | --- | --- |
| LSN 00 begin check point  LSN 10 end check point  LSN 20 update: T1 writes P1  LSN 30 update: T2 writes P2  LSN 40 update: T3 writes P3  LSN 50 T2 commit  LSN 60 update: T3 writes P2  LSN 70 T2 end  LSN 80 update: T1 writes P5  LSN 90 T3 abort  ------------------------------------------  LSN 100 CLR: Undo T1 LSN 80  LSN 110 CLR: Undo T3 LSN 60  LSN 120 CLR: Undo T3 LSN 40  LSN 130 CLR: Undo T1 LSN 20 | Redo-phase  Undo-list=T1,T2,T3  Redo T1 writes P1  Redo T2 writes P2  Redo T3 writes P3  Undo-list=T1,T3  Redo T3 writes P2  --  Redo T1 writes P2  FAIL | Fim Undo-phase  Undo-list=vazia  Undo T1  --  Undo T3  --  Undo T3  --  Undo T1  Início Undo-phase  Undo-list=T1,T3 |

****

Em conclusão para cada transação: T1 fez rollback (redo, undo); T2 fez commit (redo); T3 fez rollback (redo, undo)

Critério de correção:

- 1,0 valores, fases de Redo e Undo

- 1,0 valores, diagrama temporal

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

**Grupo B – Prática em “Data Warehousing”**

**5.** (4 valores)Considere a seguinte base de dados de compra e venda de produtos. Pretendemos desenhar um “Data Warehouse” do sistema da figura.

**5.a)** Defina as tabelas de auxiliares (‘lookup’), intermédias e de factos em primeiro lugar.

**Resposta:**



**5.b)** De seguida, defina pelo menos três dimensões para cada tabela de factos.

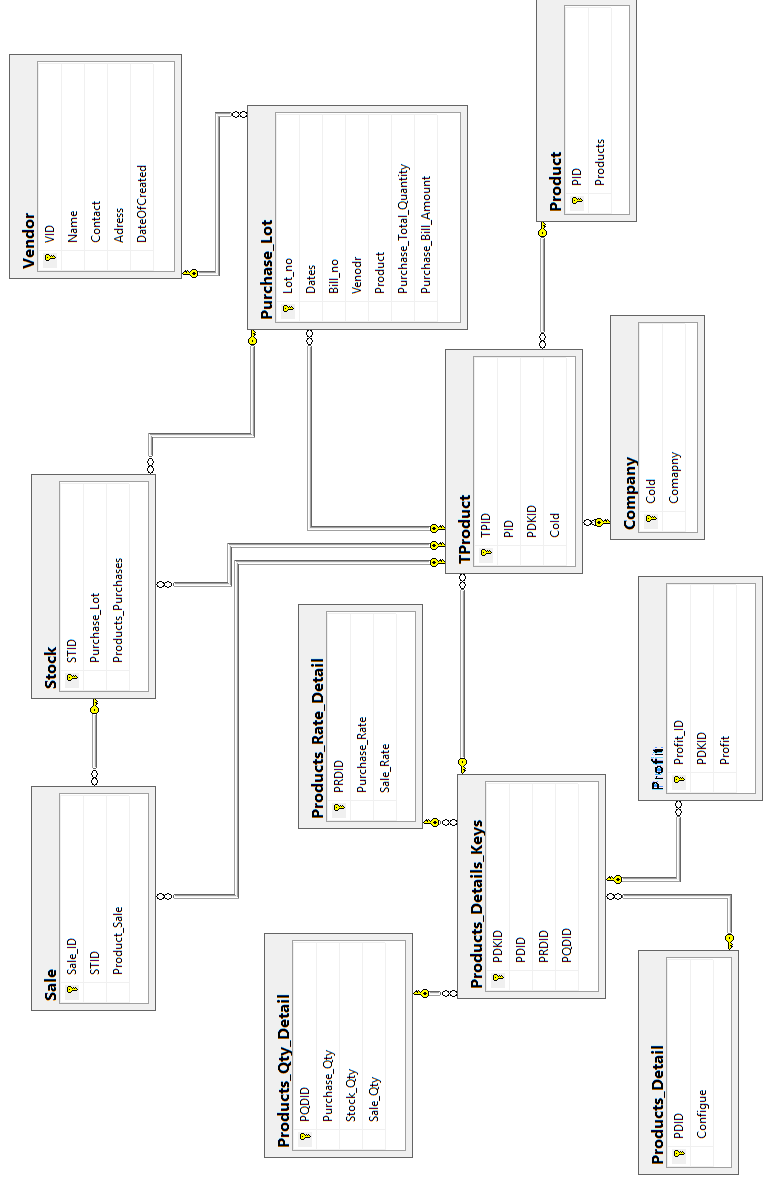


Critério de correção:

- a) 2,0 valores, 3 tipos tabelas

- b) 2,0 valores, tabelas de factos e dimensões

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



**FIM**