E-fólio Global | Instruções para a realização do E-fólio

SISTEMAS DE GESTÃO DE BASES DE DADOS | 21103 | ÉPOCA NORMAL

Período de Realização: decorre 09-02-2022 deste 15:00 com 2:30 horas de duração

Data de Limite de Entrega: até 17:30 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 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!

Tópico 1 – Transações e Consultas

1. (2 valores) Relativo ao tema das Transações, considere o seguinte sequenciamento r1(x), r2(x), w2(x), r3(y), w4(x), r5(x), w5(z), w1(z).
a) O que entende por sequenciamento serializável a conflitos? O que entende por sequenciamento serializável a vistas? Justifique e exemplifique a resposta.
b) O referido sequenciamento é serializável a conflitos? O referido sequenciamento é serializável a vistas? Justifique e exemplifique a resposta.

2. (2 valores) Relativamente à concorrência em bases de dados:
a) Defina o protocolo 2-PL relativamente à seriabilidade e à recuperação e ao 'deadlock'. Justifique a resposta.
b) Considere o protocolo 2-PL e acrescente os operadores X-lock(_), S-lock(_) e Unlock(_). Existe a possibilidade de 'deadlock'? Complete a tabela e justifique a resposta.

T1	T2	Justificação
Read (A) Read (B) . . Write (B) Write (A) . .	. . Read (B) Read (A) . . Write (A) Write (B)	

3. (2 valores) Considere a seguinte sequência de 'logs' das transações T1, T2 e T3.

Log	Redo Phase	Undo Phase
Start T1 Read T1, D Checkpoint Write T1, D, 20, 25 Start T2 Read T2, B Checkpoint Commit T1 Start, T3 Write T2, B, 12, 18 Commit T2 Read T3, D Write T3, D, 25, 15 System crash ←-----		

a) Represente as transações e os checkpoints na linha do tempo (diagrama temporal).
b) Acrescente os registos na recuperação. Complete a tabela e justifique a resposta.

Tópico 2 – Data Warehousing, Mining e Information Retrieval

4. (4 valores) Considere a seguinte base de dados que vai servir de fonte de dados a um "Data Warehouse".

a) Encontre a 1FD (1ª forma desnormalizada) e a 2FD (2ª forma desnormalizada).
1FD – obter uma poli-árvore, replicando as tabelas que forem necessárias;
2FD – obter árvores separadas, replicando as tabelas que forem necessárias.
Justifique a resposta.

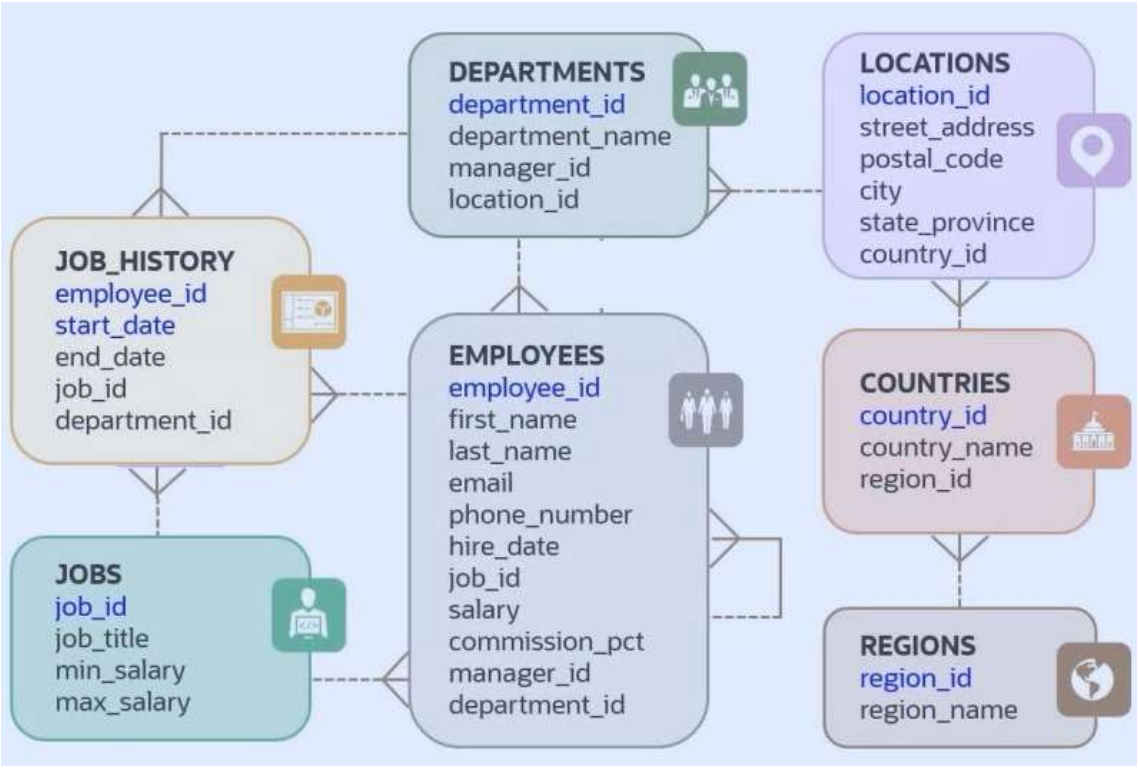
b) Verifique se a tríade (Departements, Employees, Job_History) é ou não redundante. Justifique a resposta.

c) Considerando que:

- Aditivos: são atributos que podem ser agregados (somados) por todas as dimensões, ex: valor da venda (usar Sum() sempre)
- Semi-aditivos: são atributos que podem ser agregados (somados) por algumas as dimensões, ex: quantidade (usar Sum() em condições particulares)
- Não-aditivos: são atributos que não podem ser agregados (somados), ex: preço unitário (usar Average() por exemplo)
- Sem factos: só existem identificadores (usar a função Count() dos identificadores).

Para as tabelas de factos encontradas defina os atributos aditivos, semi-aditivos, não-aditivos e sem factos. Justifique a resposta.

d) De seguida, defina um DW, em estrela ou constelação, com pelo menos três dimensões para cada tabela factos.



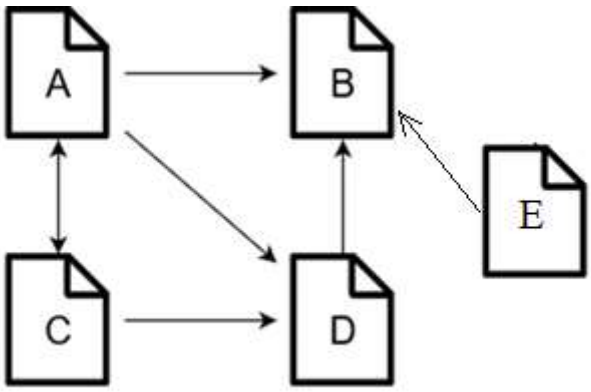
5. (2 valor) Information Retrieval

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

$$P[j] = \delta + (1 - \delta) * \sum_{i=1}^{N} (T[i,j] * P[i])$$

com $\delta = 0{,}5$.

Encontre a ordenação das páginas para o seguinte conjunto:



FIM