

“

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

**Período de Realização:** decorre 27-01-2025 desde 15:00 com 1:30 horas de duração e tolerância de 0:15.

**Data de Limite de Entrega:** até 16:45 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:** ambiente WISEflow, tipo FLOWlock

**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 nome da plataforma. Finalmente deve gerar um PDF do documento. Deve carregar o referido ficheiro para a plataforma no dispositivo 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 – Consultas e Transações

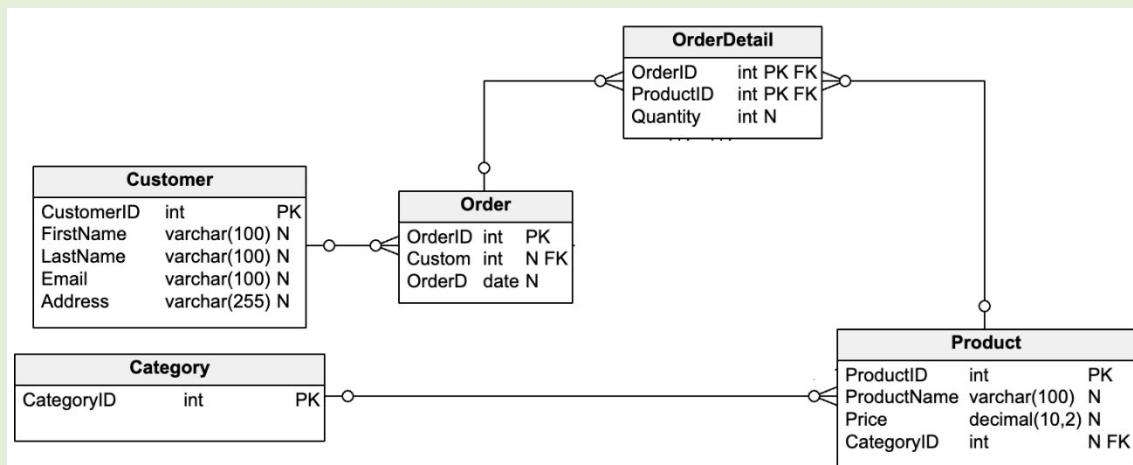
**1.** (2 valores) Relativo ao tema das transações, considere o seguinte sequenciamento r1(X), r2(X), r2(Y), r1(Y), r2(X), w2(X), w1(X), w2(Y). O referido sequenciamento é serializável a conflitos? Justifique a resposta.

**2.** (2 valores) Considere o sequenciamento da pergunta 1. O referido sequenciamento é serializável a vistas? Justifique a resposta.

**3.** (2 valores) Considere o sequenciamento da pergunta 1 e 2. O referido sequenciamento é serializável pelo 2PL com ‘locks’ exclusivos e partilhados? Justifique a resposta.

## Tópico 2 – Data Warehousing, Mining e Information Retrieval

**4.** (2 valores) Considere a seguinte base de dados que vai servir de fonte de dados a um “Data Warehouse”.



Considere ainda, as seguintes formas desnormalizadas (FDs):

- 1FD – constituída por uma poli-árvore, com a replicação das tabelas intermédias e de lookup que forem necessárias para evitar caminhos múltiplos;
- 2FD – constituída por várias árvores separadas, com a replicação das tabelas intermédias e de lookup que forem necessárias para evitar caminho múltiplos; esta FD é equivalente ao esquema em estrela ou ao esquema floco-de-neve.
- 3FD – o processo de desnormalização termina com a junção de todas as tabelas da árvore com vista a uma rápida leitura dos dados.

Em que forma de desnormalização se encontra a bases de dados? Justifique a respostas usando a seguinte tabela e referindo os caminhos múltiplos

Antes 1 FD	
1FD	
2FD	
3FD	

**5.** (2 valores) Considere a bases de dados da pergunta 4, desenhe um Data Mart, em estrela, com uma tabela de factos e com pelo menos três dimensões.

Considere os seguintes tipos de atributos da tabela de eventos:

- 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).

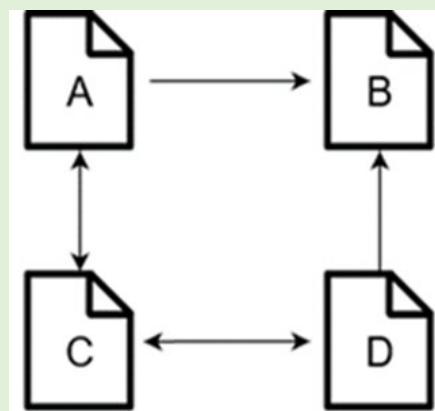
Para a tabela de factos defina o tipo de atributos que apresenta. Justifique a resposta.

**6.** (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. Justifique a resposta.



**FIM**