



## **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: 00000efolioGlobal. 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.

## 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?

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

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

2.b) Considere que os valores  $\text{V}(\text{A}, x)$  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:

Select a.nome

From aluno a, inscrito i, disciplina d

Where a.id = i.aluno\_id

And i.disc\_id = d.id

And a.districto = 'Lisboa'

Tabela	N tuplos
disciplina	50
inscrito	1000
aluno	300
aluno Lisboa	50

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.

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

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

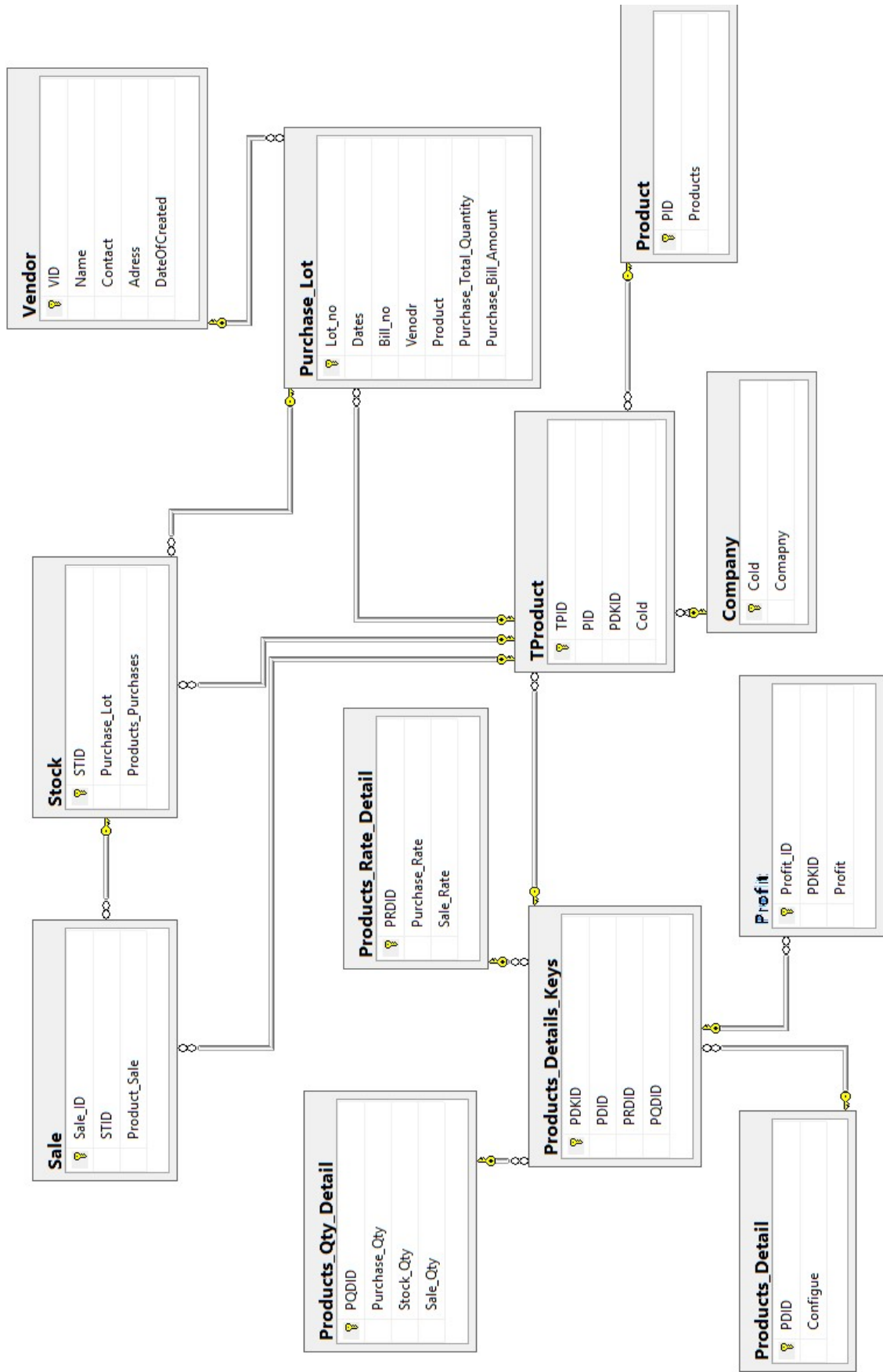
Represente as transações na linha do tempo e acrescente os registos na recuperação.

## 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.

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



FIM