

Número: \_\_\_\_\_ Nome: João Figueiredo

Ramo: \_\_\_\_\_

## Grupo I - Modelação Dimensional

(6,5 valores)

O sistema informático de uma empresa que comercializa máquinas de construção regista as despesas de deslocação/representação efectuadas pelos seus vendedores a potenciais clientes. Os vendedores da empresa são caracterizados por: número/identificador; nome; morada; código postal; telemóvel; data de admissão na empresa; e volume de vendas no último ano. A cada código postal corresponde a respectiva localidade.

Os vendedores encontram-se afectos às diferentes secções de venda existentes na empresa (*e.g.*; máquinas de terraplanagem; gruas). Uma secção é caracterizada por: código; descrição; data de criação; data de extinção (no caso desta ter sido encerrada); número de vendedores afectos e, vendedor-chefe que assume a responsabilidade pela secção.

Sempre que ocorre uma deslocação (nacional ou internacional) de um vendedor a um potencial cliente são registados os seguintes dados: identificador do potencial cliente (gerado automaticamente); nome; morada; código postal; país; data da deslocação; número de dias despendidos; valor total orçamentado inicialmente; e vendedor ou vendedores envolvidos na deslocação.

As despesas efectuadas são classificadas de acordo com o seu tipo (*e.g.*; refeições; viagem de avião; táxi). Um tipo de despesa é caracterizado por um código e pelo respectivo descritivo.

Para cada despesa individual ocorrida no âmbito de uma deslocação é registado: o valor (na moeda original em que ocorreu), o vendedor que a originou, a respectiva moeda, o dia e hora em que ocorreu, e o tipo de despesa a que se refere. Numa deslocação a um potencial cliente estrangeiro podem estar envolvidas despesas em moedas diferentes.

Seguindo a metodologia *Kimball*, desenvolva o processo de análise dimensional, a fim de definir e criar o esquema conceptual para um *data mart* que permita realizar análises de dados variadas, apresentando todos os factos, dimensões, granularidade e todos os aspectos relevantes para o projecto de *data mart*. Na elaboração do *data mart* tenha ainda em atenção o seguinte requisito:

- As análises multi-dimensionais de dados devem poder ser realizadas em qualquer moeda seleccionada pelo utilizador, independentemente da moeda em que tenha ocorrido cada despesa.

## Grupo II - Múltipla Escolha

(1 valor cada questão correcta/-0,5 cada questão errada)

Nas questões seguintes assinale apenas uma só alternativa correspondendo à que considera correcta.

1. A área de manipulação de dados (*data staging area*), existente nas arquitecturas de *Ralph Kimball* e *Bill Inmon*:
  - ☐ Não permite o acesso por parte dos utilizadores finais seja por que motivo for.
  - ☐ Não constitui suporte de armazenamento para a elaboração de relatórios.
  - ☐ Apenas suporta operações de leitura/escrita por parte dos processos de ETL.
  - ☒ Possui todas as características que constam das afirmações anteriores.
2. Um facto/medida do tipo *unidades\_em\_stock* existente numa tabela de factos de um *data mart* de inventário:
  - ☐ Pode ser adicionado ao longo de qualquer dimensão existente.
  - ☒ Pode ser adicionado apenas ao longo da dimensão data e/ou tempo.
  - ☐ Não pode ser adicionado ao longo de qualquer dimensão.
  - ☐ Pode ser adicionado ao longo de todas as dimensões, à excepção da dimensão data e/ou tempo.
3. A nível de performance, a melhor opção para garantir a integridade referencial de um facto que é carregado na tabela de factos consiste em:
  - ☒ Efectuar uma verificação da existência das chaves de substituição ainda antes de se proceder ao carregamento do facto em questão.
  - ☐ Efectuar a verificação da existência das chaves de substituição no momento em que o facto está a ser carregado, tirando partido do mecanismo da base de dados que assegura a manutenção da integridade referencial.
  - ☐ Efectuar a verificação da existência das chaves de substituição só após o carregamento dos factos e apenas periodicamente.
  - ☐ Utilizar qualquer uma das anteriores alternativas, uma vez que são iguais a nível de performance.

4. Uma possível estratégia de optimização em armazéns de dados envolve a criação de partições. Qual das seguintes afirmações não é correcta?
- ☒ Em armazéns de dados, o atributo data deve ser o eleito no particionamento horizontal.
  - ☐ No particionamento horizontal cada ficheiro criado possui exactamente a mesma estrutura/esquema.
  - ☐ O particionamento horizontal é mais usado do que o particionamento vertical.
  - ☐ Acessos a dados em múltiplas partições são sempre mais rápidos do que aceder a dados que se encontram numa só tabela.
5. O *Multidimensional On-Line Analytical Processing* (MOLAP) constitui a forma tradicional/clássica de OLAP. Uma das seguintes características não é válida neste tipo de OLAP:
- ☐ Dados encontram-se armazenados numa matriz multidimensional optimizada.
  - ☐ Obriga ao processamento prévio dos dados num hipercubo.
  - ☐ Permite manipular volumes de dados muito elevados.
  - ☒ Menor espaço necessário em disco para o armazenamento do hipercubo, em virtude da utilização de técnicas de compressão.

**Grupo III – Verdadeiros ou Falsos com Justificação**  
**(2 valores cada questão)**

Indique se as seguintes afirmações são verdadeiras ou falsas, apresentando a respectiva justificação.



1. O nível de granularidade adoptado numa tabela de factos é unicamente relevante segundo as perspectivas técnica e de projecto.

Falso - Além dessas 2 perspetivas, também é necessário a perspetiva de negócio, quando determinamos o nível de granularidade

---

---

---

---

2. Uma mini-dimensão é utilizada em armazéns de dados para armazenar conjuntos de atributos do tipo textual e do tipo *flag* (booleano), o que permite a sua remoção da tabela de factos.

Falso - O processo descrito diz respeito a uma junk dimensions e não a uma mini-dimensão. Uma minidimensão serve para este efeito: utilizada em armazéns de dados para armazenar as combinações

possíveis de um conjunto de atributos demográficos, o que permite a sua remoção de uma dimensão

---

---

3. No armazenamento de agregações em armazéns de dados (estratégia de optimização) podem ser utilizadas duas técnicas diferentes.

Verdadeiro - Existem 2 técnicas para armazenar os agregados:

---

1 - Os factos agregados e dimensões agregadas são armazenados em novas tabelas separados dos dados base atómicos.

---

---

2 - Os factos agregados são armazenados na tabela de factos atómica e os atributos de nível são armazenados nas dimensões

---

#### **Grupo IV – Questão de Desenvolvimento** **(2,5 valores)**

No registo das alterações que podem ocorrer aos atributos de uma dimensão, existem três técnicas disponíveis. Caracterize cada uma dessas técnicas e exemplifique o seu funcionamento num cenário de actualização ao valor de um atributo.

Existem 3 técnicas (mais até :) ), mas serão referenciados os tipos 1, 2 e 3.

---

Tipo 1: Dada uma alteração, num atributo de tipo 1 é reescrito, no próprio registo.

---

---

Tipo 2: Dada uma alteração, num atributo de tipo 2, é criado um novo registo dimensional e adicionado no registo anterior (onde se encontra o valor passado) a data de expiração.

---

---

Tipo 3 : Dada uma alteração, num atributo de tipo 3, é criada uma nova coluna na dimensão, com a informação previamente disponível no registo e o atributo tipo 3 é alterado com a informação mais recente

---

---

---

---

---

---

---

---