

Número: _____ Nome: Tiago e João**Grupo I - Modelação Dimensional**
(6,5 valores)

Uma determinada Instituição Bancária possui um sistema operacional que regista os movimentos, dos seus clientes, nas contas de depósito à ordem. Os movimentos que ocorrem numa conta de depósito à ordem resultam de diversos eventos, como: depósito de numerário ao balcão; depósito de cheque ao balcão; depósito de numerário em caixa ATM (multibanco); depósito de cheque em caixa ATM; levantamento ao balcão; levantamento em caixa ATM; pagamento de cheque; pagamento de compra em caixa ATM; e, débito direto por prestador de serviços (e.g., eletricidade; água; telecomunicações). Os diversos eventos resumem-se a dois tipos de movimento nas contas de depósito à ordem: débito e crédito. Os movimentos a débito e crédito representam, respetivamente, a saída e entrada de valores monetários na Instituição Bancária.

As contas de depósito à ordem são caracterizadas por: um número (único para cada conta); um titular da conta (o cliente propriamente dito); um conjunto de até três pessoas com autorização para movimentação da conta; um NIB; a data de abertura; e, eventualmente, a data de encerramento da conta. As contas de depósito à ordem são ainda caracterizadas pelo tipo de conta (e.g., conta ordenado; conta jovem 18-30; conta estudante ensino superior). Para além deste descritivo, cada tipo de conta é ainda caracterizada por: um código (identificador único); uma data de lançamento; e, eventualmente, uma data de encerramento, caso esse tipo de conta já não se encontre em comercialização. Cada cliente possui uma ou mais contas de depósito à ordem (de tipos diferentes).

Cada cliente da Instituição Bancária é caracterizado por: o número de cliente (identificador único); o número de identificação fiscal; o nome; a morada; o código postal; o país; o contato telefónico; o e-mail; a data em que se tornou cliente; e, a data em que realizou o último movimento. A cada código postal, de cada país, corresponde a respetiva localidade. Sempre que ocorre um movimento, além deste, o sistema operacional da Instituição Bancária regista: o cliente que a originou; a conta de depósito à ordem envolvida; o instante de tempo em que ocorreu; e, o respetivo valor.

1. Seguindo a metodologia *Kimball*, desenvolva o processo de análise dimensional, a fim de definir e criar o modelo dimensional para um *data mart* que permita realizar análises multidimensionais de dados variadas aos movimentos dos clientes nas contas de depósito à ordem, de acordo com a realidade que acabou de ser descrita. Apresente todos os factos, dimensões, granularidade e todos os aspectos relevantes para o projecto de *data mart*.

2. Admita que mensalmente são realizadas as seguintes análises de dados:

- Dada uma conta de depósito à ordem, saber-se qual o montante total movimentado a débito e crédito.
- Dado um cliente, saber-se qual o montante total movimentado a débito e crédito nas várias contas de depósito à ordem que este possa possuir.

O que acrescentaria ou alteraria a nível do modelo dimensional para otimizar a realização deste tipo de análises?

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. Na constituição da chave primária de uma tabela de factos:

- ☐ Devem constar todos os atributos que são chaves estrangeiras para as dimensões.
- ☐ O atributo que representa a chave estrangeira para a dimensão Data deve ser o que surge em primeiro lugar.
- ☒ Deve constar unicamente o subconjunto de atributos, entre os que são chaves estrangeiras para as dimensões, que permite identificar de uma forma única cada facto/registo.
- ☐ Há que ter em atenção outros aspetos que não os referidos nas alíneas anteriores.

2. O nível de granularidade/detalhe adotado numa tabela de factos é importante:

- ☐ Porque não permite que se realizem análises abaixo desse nível de granularidade.
- ☐ Porque é dos principais fatores que influenciam o espaço ocupado pelo armazém de dados.
- ☐ Por ter implicação no volume de trabalho necessário para implementar o armazém de dados.
- ☒ Por todos os motivos apresentados nas alíneas anteriores.

3. Frequentemente os sistemas operacionais contêm atributos do tipo *flag* e outros atributos que possuem um conjunto reduzido de valores discretos. A melhor forma de armazenar estes valores num modelo dimensional passa por:

- ☐ Criar uma mini-dimensão contendo as diferentes combinações dos valores dos vários atributos.
- ☒ Criar uma dimensão *Junk* contendo as diferentes combinações dos valores dos vários atributos.
- ☐ Criar uma dimensão para cada um dos atributos.
- ☐ Colocar esses atributos na tabela de factos, de modo a que possam ser feitas análises com estes.

4. Na extração de dados a partir dos sistemas fontes/operacionais pode:

- ☐ Extrair-se unicamente os dados novos e/ou alterados desde o último carregamento, naquilo que se designa de extração estática.
- ☒ Utilizar-se uma abordagem de “força bruta”, extraíndo unicamente dados novos e/ou alterados.
- ☐ Realizar-se uma extração incremental, recorrendo-se a um algoritmo de CRC (*Cyclic-Redundancy Checksum*).
- ☐ Nenhuma das anteriores afirmações está correcta.

5. Uma estratégia de otimização vulgarmente utilizada em armazéns de dados envolve a criação de agregações. Os valores agregados:

- ☐ Têm que ser armazenados em novas tabelas de factos.
- ☐ Têm que ser armazenados nas tabelas de factos já existentes.

- ☐ Podem ser armazenados nas tabelas de factos já existentes, obrigando a que sejam feitas alterações a estas, mas não às dimensões.
- ☐ Podem ser armazenados nas tabelas de factos já existentes, obrigando a que sejam feitas alterações a estas e também às dimensões.

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. Na modelação dimensional não é admissível a existência de atributos numéricos numa dimensão. Este tipo de atributos corresponde a medidas e, como tal, deve ser colocado numa tabela de factos.

Falso - É admissível colocar atributos numéricos nas dimensões. O facto de serem numéricos não significa que sejam medidas/factos do negócio.

2. Um índice do tipo *Bitmap* não pode ser utilizado como chave primária de uma dimensão.

Falso - O SQL server não suporta índices bitmap

3. A existência do atributo *IsCurrent* numa dimensão é suficiente para cobrir todas as situações que possam ocorrer no carregamento de dados num armazém de dados que implementa um mecanismo de SCD (*Slowly Changing Dimension*) de Tipo 2, não sendo necessários os atributos *EffectiveDate* e *ExpiredDate*.

Falso - O *effectiveDate* e o *ExpiredDate* são indispensáveis na implementação de tipo 2. Caso contrário, nem seria possível haver um registo histórico

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

As principais arquiteturas de referência na área dos armazéns de dados são a *Bus Architecture* e a *Corporate Information Factory*, respetivamente, da autoria de *Ralph Kimball* e *Bill Inmon*. Apresente três semelhanças e três diferenças entre estas arquiteturas.

Semelhanças: Staging area, gestão de metadados, aquisição de dados e entrega de dados são processos separados; Poder de Informação é maior quando os dados são mantidos ao nível mais atómico/elementar; existência de data marts dimensionais.

Diferenças: No BUS DW não existe um repositório físico como no CIF DW; No BUS DW é usado o modelo estrela enquanto que no CIF DW é o esquema relacional; No BUS todos os componentes são dimensionais exceto a staging area