

Número: _____ Nome: _____

Grupo I - Modelação Dimensional
(6,5 valores)

A Sociedade Interbancária de Serviços (SIBS) possui um sistema operacional que regista todos os movimentos que ocorrem nacionalmente e internacionalmente na rede de terminais ATM (caixas multibanco e terminais de pagamento existentes nos estabelecimentos comerciais), realizados pelos clientes dos bancos portugueses. Sempre que ocorre um movimento é registado: o cartão multibanco que o originou; o instante de tempo local (data e hora com granularidade até ao segundo); o instante de tempo universal UTC (data e hora com granularidade até ao segundo); o terminal ATM onde ocorreu; o tipo de movimento produzido; a moeda local utilizada; e, o respetivo valor monetário.

Cada terminal ATM é caracterizado por: código do terminal (único para cada terminal); morada; código postal; localidade; localização detalhada (uma vez que o terminal ATM pode estar numa loja de um shopping); país; e, a indicação se esta se encontra ativa ou não (o terminal ATM pode ter sido removido). Para que se possa efetuar um movimento ATM é necessário um cartão multibanco. Cada cartão multibanco é caracterizado por: número (único para cada cartão); tipo de cartão (débito ou crédito); número de conta bancária a que se encontra associado; data de início de validade; e, data de fim de validade. Nos terminais ATM podem ocorrer diferentes tipos de movimento (e.g., levantamento de numerário; depósito de numerário; pagamento de compras; pagamento de serviços). Cada tipo de movimento é caracterizado por: código (diferente para cada tipo de movimento); descritivo; e, tipo de movimento contabilístico que origina (débito ou crédito).

Cada cartão multibanco encontra-se associado a um cliente da rede ATM. Sobre cada um destes clientes é conhecido: número de cliente (único para cada cliente); nome; morada; código postal; localidade; país; e contacto telefónico. Cada cartão multibanco encontra-se também associado a um banco (que o emitiu). Sobre cada banco conhece-se: código internacional de banco (diferente para cada banco); nome; morada da sede; código postal, localidade, país, capital social; data de abertura; e, eventualmente, data de encerramento.

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 na rede de terminais ATM, 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*.

Metodologia de Kimball:

- **Área de Negócio:** Movimentos em Terminais ATM
- **Granularidade:** Movimentos por terminal, por cartão, por banco, por cliente, por tipo de movimento, por moeda, por data e por tempo.
- **Dimensões:** TerminalATM; Cartão; TipoMovimento; Cliente; Banco; Data; Tempo; Moeda
- **Factos:** Montante-local; Montante-std

Modelo Dimensional:

| DimTerminalATM | DimCartao | DimTipoMovimento |
|--|--|---|
| TerminalATMKey (PK) CódigoTerminalATM Morada Localização CódigoPostal Localidade País Ativa | CartaoKey (PK) NumeroCartao ClienteKey TipoCartao BancoKey NrConta DataInicioValidadeKey DataFimValidadeKey | TipoMovimentoKey (PK) CodigoTipoMovimento TipoMovimento Descritivo |

| DimCliente | DimBanco | DimData |
|--|--|--|
| ClienteKey (PK) NrCliente Nome Morada CódigoPostal Localidade País ContactoTelefónico DataInicio DataFim Atual | BancoKey (PK) CodigoBanco Nome Morada CódigoPostal Localidade País Capital Social DataAberturaKey DataEncerramentoKey DataInicio DataFim Atual | DataKey (PK) DataCompleta Ano Semestre Quadrimestre Trimestre Mes Semana DiaSemana DiaAno Dia ... |

| DimTempo | DimMoeda | FactMovimentosATM |
|---|--|---|
| TempoKey (PK) TempoCompleto Hora Minuto Segundo Período HoraAlmoço HoraJantar ... | MoedaKey (PK) Descritivo Simbolo | CartaoKey (PK) UTCDDataKey (PK) UTCTempoKey (PK) LocalDataKey LocalTempoKey TerminalATMKey ClienteKey BancoKey MoedaKey Montante_Local Montante_Std |

2. Tenha em conta o requisito de que as análises multidimensionais de dados devem poder ser realizadas em qualquer moeda selecionada pelo utilizador, independentemente da moeda original em que tenha ocorrido o movimento no terminal ATM.

| FactCambio |
|---|
| DataKey (PK) MoedaOrigemKey (PK) MoedaDestinoKey (PK) FatorConversão |

Grupo II - Múltipla Escolha

(1 valor cada resposta correta/-0,5 cada resposta errada)

Grupo III – Verdadeiros ou Falsos com Justificação

(2 valores cada resposta correta)

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

1. Abstraindo de todos os outros fatores que possam condicionar a escolha do nível de granularidade de um armazém de dados, a melhor opção consiste em adotar o nível de granularidade mais elementar.

Verdadeiro, uma vez que do nível de granularidade mais elementar/detalhado (*e.g.*, ao nível da linha da fatura) é sempre possível agregar os dados para níveis de granularidade menos elementares/detalhados (*e.g.*, ao nível da fatura). O contrário é que não é verdade, o que impossibilita análises de dados futuras que se queiram fazer a esse nível de granularidade mais elementar/detalhado.

2. A existência de atributos como *Data Criação* e *Data Última Alteração* dos registos, nas tabelas do(s) sistema(s) fonte permite otimizar o processo de extração de dados para a *Staging Area*.

Verdadeiro, a existência da *Data Criação* e *Data Última Alteração* permitem otimizar a extração dos dados para a *Staging Area*, de modo que esta possa ser feita de forma incremental. Em vez de se extrair diariamente sempre a totalidade dos dados do(s) sistema(s) fonte, pode extrair-se apenas os registos que são novos (com base na *Data Criação*) e os que, já tendo sido extraídos anteriormente, sofreram alguma alteração (com base na *Data Última Alteração*).

3. Na tabela de factos apenas podem existir atributos cujo tipo de dados seja numérico.

Falso, na tabela de factos existe pelo menos um atributo para manter a ligação com o sistema fonte, denominado de *degenerated dimension* (*e.g.*, *TransactionID*). O seu tipo de dados tem de ser exatamente o mesmo que o utilizado no sistema de onde provém. Se o tipo de dados for um *char* ou *varchar* na tabela de factos tem de se usar esse tipo. No exemplo prático usado nas aulas PL acontecia isso, em que o atributo *TransactionID* existente na tabela de factos possuía com tipo de dados *char(11)*.

Grupo IV – Questão de Desenvolvimento

(2,5 valores)

Um aspeto fundamental num armazém de dados é a manutenção da integridade referencial entre as tabelas de factos e as dimensões. Apresente e explique as diferentes formas como a integridade referencial pode ser assegurada/mantida.

A verificação da integridade referencial entre a tabela de factos e as respetivas dimensões pode ser assegurada das seguintes formas:

- **Antes dos factos/registos serem carregados na tabela de factos ou de serem eliminados registos das dimensões. Esta constitui a melhor aproximação, tendo sido inclusivamente a que foi seguida no exemplo prático das aulas PL e também a que foi seguida nos trabalhos práticos.**
- **Fazer a verificação durante o carregamento dos factos/registos para a tabela de factos. A responsabilidade de fazer essa verificação é passada para o motor da base de dados que verifica se as restrições chave estrangeira – chave primária, definidas entre a tabela de factos e as dimensões, são cumpridas. Um erro/exceção é gerado quando tal não acontece.**
- **Fazer a verificação após o carregamento dos factos/registos na tabela de factos. Nesta situação, as restrições chave estrangeira – chave primária não se encontram definidas, o que permite o carregamento de qualquer facto/registo. Para assegurar a integridade referencial é necessário realizar verificações periódicas da existência de chaves estrangeiras inválidas na tabela de factos. Basicamente, procura-se identificar as chaves estrangeiras existentes na tabela de factos que não existam nas respetivas dimensões.**