

Número: _____ Nome: Carlitos

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 caracterizada 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*.

2. Tenha em conta o requisito de que as análises multidimensionais de dados devem poder ser realizadas em qualquer moeda seleccionada pelo utilizador, independentemente da moeda original em que tenha ocorrido o movimento no terminal ATM. Para além da DimCurrency, seria necessária uma nova tabela de factos, FactCurrencyRate. Assim seria possível trabalhar com qualquer moeda

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. Uma tabela de factos inclui:
 - ☐ Uma chave primária formada sempre pelo conjunto de chaves estrangeiras e um conjunto de medidas (factos) cuja análise é relevante para o negócio.
 - ☐ Uma chave primária formada pelo conjunto ou subconjunto de chaves estrangeiras e sempre por um conjunto de medidas (factos) cuja análise é relevante para o negócio.
 - ☐ Um conjunto de medidas (factos) cuja análise é relevante para o negócio e, pelo menos, uma dimensão degenerada (*degenerate dimension*).
 - ☒ Uma chave primária formada pelo conjunto ou subconjunto de chaves estrangeiras e, eventualmente, um conjunto de medidas (factos) cuja análise é relevante para o negócio.
2. Na implementação de um mecanismo de SCD (*Slowly Changing Dimension*) – Tipo 2 numa dimensão:
 - ☐ É obrigatório acrescentar-se simultaneamente os atributos *EffectiveDate* (ou *StartDate*), *ExpiredDate* (ou *EndDate*) e *IsCurrent*.
 - ☐ Não é obrigatório acrescentar qualquer atributo adicional.
 - ☒ Pelo menos, há que acrescentar os atributos *EffectiveDate* (ou *StartDate*) e *ExpiredDate* (ou *EndDate*).
 - ☐ É suficiente acrescentar o atributo *IsCurrent* para efetuar uma correta manipulação do histórico.
3. Entre as arquiteturas de armazéns de dados *Bus* de Ralph Kimball e *Corporate Information Factory* (CIF) de Bill Inmon existem algumas diferenças. Uma das seguintes afirmações corresponde a uma dessas diferenças:
 - ☒ Na arquitetura *Bus* os *data mart* seguem o modelo dimensional enquanto na arquitetura CIF os *data mart* seguem o modelo relacional.
 - ☐ Na arquitetura *Bus* defende-se que o maior poder/flexibilidade dos dados se encontra no nível atómico enquanto na arquitetura CIF defende-se que o maior poder/flexibilidade se encontra no nível mais agregado.
 - ☐ Na arquitetura *Bus* os *data mart* agregados não são o mesmo que os *data mart* na arquitetura CIF.
 - ☐ Nenhuma das anteriores afirmações representa corretamente uma diferença entre ambas as arquiteturas.
4. Na indexação de uma tabela de factos:
 - ☐ O índice correspondente à chave primária pode ser do tipo *B-Tree* ou *Bitmap*.
 - ☒ Não se justifica a criação de índices isoladamente para cada um dos atributos que faz parte da chave primária.
 - ☐ O atributo *DateKey* deve encontrar-se logo na primeira posição do índice referente à chave primária.
 - ☐ Nunca se justifica a criação de índices nos atributos que são factos/medidas.

5. No *On-Line Analytical Processing* (OLAP) há diversos tipos de operações de análise de dados que os utilizadores podem realizar, nomeadamente:
- ☐ Operação de *Drill-down* que consiste em efetuar análises de dados a um menor nível de detalhe/granularidade.
 - ☐ Operação de *Roll-up* que consiste em efetuar análises de dados a um maior nível de detalhe/granularidade.
 - ☐ Operação de *Dice* que consiste em selecionar um sub-cubo a partir do cubo de dados original.
 - ☐ Operação de *Slice* que consiste em visualizar os mesmos dados, mas de uma perspetiva diferente.

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. Quando num armazém de dados se integram dados com origem em diferentes fontes é fundamental o armazenamento da proveniência de cada registo.

Verdade - Sim, os dados devem possuir um identificador da fonte de dados respetivo, que será visível pelos end-users, que poderão refletir se as dimensões dizem respeito aos seus dados e respetivos sistemas operacionais

2. As medidas semi-aditivas (e.g., quantidade em stock; saldo bancário) podem ser agregadas por qualquer dimensão que faça parte de um armazém de dados.

Falso - Medidas semi-aditivas só podem ser agregadas por algumas dimensões e não todas

3. Admitindo que há um conjunto de 10 atributos diferentes do tipo booleano (i.e., *flags*) numa tabela que serve de fonte para uma tabela de factos, tal implica que se criem 10 dimensões distintas (uma dimensão para cada atributo) no modelo dimensional.

Falso - Estes atributos devem ser colocados numa dimensão junk

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 verificada/assegurada.

Check before loading: verificar antes de adicionar factos, de remover registos das dimensões -> melhor abordagem

Check during loading: A engine faz isto por default, é giro mas é lento

Check after loading: Assume-se que não há integridade referencial na base de dados, necessita de checks periódicos -> lento