

Número: _____ Nome: Diogo Matias

Grupo I - Modelação Dimensional (6,5 valores)

Uma determinada operadora de serviços de televisão possui um sistema operacional que regista todos os eventos gerados pelos clientes quando interagem com as *boxes* de televisão que lhes são disponibilizadas. A *box* permite o acesso aos serviços de televisão que a operadora disponibiliza, nomeadamente: à emissão em direto dos diversos canais (em sinal digital); a programas gravados (emitidos nos últimos 7 dias); ao videoclube; e, a jogos interativos. Cada canal é caracterizado por um número que identifica a posição do canal na grelha de canais (pode mudar ao longo do tempo), por um código de identificação (nunca altera), pelo nome (e.g.: RTP1; SIC; TVI), pelo género (e.g.: generalista; notícias; documentário; filmes; desporto), pela data de início de emissão na operadora e, eventualmente, pela data de fim de emissão (ao longo do tempo, alguns canais deixam de ser disponibilizados). Os conteúdos disponíveis são caracterizados por um identificador único, pelo nome, pelo tipo (i.e.: programa de televisão; filme do videoclube; jogo), pelo subtipo (e.g.: documentário; filme de animação; jogo de estratégia), pela classificação etária a que se destina (e.g.: todos; maiores de 6; maiores de 12), pela data em que o conteúdo foi disponibilizado pela primeira vez e, eventualmente, pela data em que o conteúdo deixou de estar disponível aos clientes.

Cada cliente da operadora de serviços de televisão é caracterizado por: código de cliente; nome; morada; código postal; número de identificação fiscal; data em que celebrou contrato com a operadora; e, eventualmente, data em que rescindiu contrato com a operadora. No caso de o cliente rescindir contrato e, posteriormente, voltar a celebrar novo contrato com a operadora, o código de cliente atribuído mantém-se inalterado. A cada código postal do cliente corresponde a respetiva localidade. Cada cliente tem acesso a uma *box* de televisão no âmbito do contrato que estabelece com a operadora de televisão. No entanto, mediante pagamento extra, pode contratar *boxes* adicionais. Cada *box* é caracterizada por: número de série; marca; modelo; data de início de instalação nos clientes; e, eventualmente, data de fim de instalação. Ao longo do tempo, as *boxes* tornam-se obsoletas, sendo substituídas por modelos mais recentes.

Sempre que é detetada interação/evento entre um cliente e uma *box*, o sistema operacional da operadora de serviços de televisão regista: a *box* em que ocorreu; o cliente respetivo; o instante de tempo (data e hora); o serviço utilizado (i.e.: emissão em direto; programa gravado; videoclube; jogo); e, o conteúdo que foi selecionado. No caso de se tratar de emissão em direto ou programa gravado é ainda registado o canal que passou a ser visualizado. O sistema operacional da operadora também regista o tempo decorrido entre um evento e o seguinte, representativo do tempo de visualização (em segundos) do conteúdo selecionado.

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 eventos originados pelos clientes nas suas interações com as *boxes* de televisão, de acordo com a realidade que acabou de ser descrita. Apresente todos os factos, dimensões, atributos e granularidade, assim como, todos os aspetos que considere relevantes para o projeto de *data mart* em questão.
2. Admita que se pretendem efetuar análises de dados com periodicidade trimestral ao total dos tempos de visualização, por conteúdo e por cliente (independentemente do n.º de *boxes* que este possa ter). Explique o que acrescentaria ou alteraria a nível do esquema do modelo dimensional elaborado na alínea anterior para suportar a realização deste tipo de análises da forma mais eficiente/otimizada. **DimTrimestre**

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.

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

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

1. A performance na realização das análises/consultas de dados é o único motivo pelo qual se adota o modelo dimensional em armazém de dados.

Falso - Existem outros 2 motivos:

Flexibilidade de mudança

Facilidade de entender por parte dos utilizadores

2. O padrão de operação/utilização de um sistema de armazém de dados é idêntico ao de um sistema operacional.

Falso - O sistema operacional serve para gerir recursos em tempo real da empresa.

O armazém de dados serve para guardar grandes quantidades de dados, possuir um histórico e permitir análises.

3. O frete existente ao nível da fatura pode ser armazenado, sem qualquer alteração, como medida/facto numa tabela de factos cuja granularidade seja ao nível da linha da fatura.

Falso - O frete deve ser dividido pelas linhas, pois numa tabela de factos não podem existir níveis de granularidade diferentes.

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

Na definição do conceito de armazém de dados são frequentemente utilizadas quatro palavras-chave. Identifique e explique o significado de cada uma dessas palavras-chave.

Subject-oriented, integrated, time variant and non-volatile

Subject-oriented -> focado na modelagem e análise de dados, criar sistemas simples e concisos

Integrado -> dados de várias fontes, limpeza e transformação de dados

Variável de tempo -> Horizonte temporal extenso, guarda o histórico ao longo do tempo, Classifica o período de tempo

Não Atualizável -> Armazenamento físico, não ocorre atualizações dos dados, não requer mecanismos de recuperação, apenas 2 operações: LOAD e ACCESS e Não pode ser atualizado pelos utilizadores finais.

(Não está nos slides :/)
