

Número: _____ Nome: Arnaldo Armindo

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 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) e à gravação de programas anteriormente emitidos nesses canais (até aos últimos 7 dias). Cada canal é caracterizado por um número que identifica a posição do canal na grelha de canais (que pode mudar ao longo do tempo), por um código de identificação (que 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 programas que são emitidos pelos canais são caracterizados por um identificador único, pelo nome, pelo tipo (e.g., transmissão desportiva, documentário, filme, serviço noticioso), pela classificação etária a que se destina (e.g., todos, maiores de 12, maiores de 16) e pelas datas de emissão da primeira e última vez. Ao longo do tempo, o mesmo programa acaba por ser emitido diversas vezes.

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. O cliente subscreve um dos pacotes de televisão comercializados pela operadora (e.g., pacote básico de 30 canais; pacote 140 canais; pacote 200 canais; pacote 140 canais + Sport TV). Naturalmente, ao longo do tempo, o pacote de televisão subscrito pelo cliente pode sofrer alterações (e.g., passar do pacote de 140 canais para o pacote de 200 canais).

Sempre que é detetada interação/evento entre o cliente e a box, o sistema operacional da operadora de serviços de televisão regista: o cliente que a origina; o instante de tempo em que ocorre (data e hora); o serviço utilizado (i.e., emissão em direto ou programa anteriormente gravado); o canal respetivo; e, o programa que está a ser visualizado (resultante da emissão em direto ou da gravação acedida).

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, granularidade e todos os aspectos relevantes para o projeto de *data mart*.

2. Admita que se pretendem efetuar as seguintes análises de dados:

- Dado um programa (e.g., de elevada audiência) saber-se quais os clientes que nunca assistiram a esse programa.
- Dado um canal saber-se quais os clientes que nunca acederam a esse canal. Não sei :(
- Dado um cliente saber-se quais os tipos de programa a que este nunca acedeu.
- Dado um cliente saber-se quais os canais a que este nunca acedeu.

O que acrescentaria ou alteraria a nível do modelo dimensional para suportar análises deste tipo? Explique de que forma essas análises poderiam ser realizadas.

Grupo II - Múltipla Escolha
(1 valor cada questão correcta/-0,5 cada questão errada) :(

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. Uma tabela de factos armazena não só os eventos ocorridos mas também os que não ocorreram, de modo a possibilitar total flexibilidade na realização das análises de dados.

Falso - Não armazena eventos que não ocorreram, para não ocupar espaço

2. A fase de carregamento de dados num armazém de dados envolve que se eliminem e recriem os índices existentes.

Falso - Este procedimento só é aplicado às tabelas de factos e não ao armazém todo (?)

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

Identifique e explique as operações básicas suportadas pelas ferramentas OLAP (*On-Line Analytical Processing*).
Existem 6:

[illegible]
