

Número: _____ Nome: Sara

Grupo I - Modelação Dimensional (6,5 valores)

Uma determinada empresa detentora da concessão de autoestradas possui um sistema operacional que regista o instante de tempo da passagem de cada veículo, nos diversos pórticos das autoestradas, em função de um dispositivo eletrónico existente no interior do veículo. Os pórticos encontram-se nos ramais de entrada/saída das autoestradas, sendo caracterizados por: um identificador; a autoestrada em que se encontram; o respetivo Km; e, pela indicação se é um pórtico de entrada ou saída da autoestrada. O identificador do pórtico é único, não existindo em qualquer outra autoestrada um pórtico com igual identificador.

A empresa é concessionária de várias autoestradas. Cada autoestrada é caracterizada por: uma designação única (e.g., A1; A2; A3); localidade de início; localidade de fim; distância em Kms do início ao fim; n.º de locais de acesso (i.e., entradas/saídas) existentes; data de início da construção; e, data de fim da construção (do último troço), uma vez que uma autoestrada vai sendo progressivamente construída ao longo do tempo.

Para poder circular numa autoestrada, cada veículo tem que possuir obrigatoriamente um dispositivo eletrónico. Cada dispositivo é caracterizado por: um número de série (único), a data de fabrico, a marca, o modelo; a matrícula do veículo em que se encontra instalado/colocado; a respetiva classe de pagamento do veículo (i.e., motociclos pertencem à classe 1; ligeiros pertencem à classe 2; pesados pertencem à classe 3); e, o código de cliente/utilizador a quem o dispositivo pertence.

Os dispositivos são adquiridos à empresa pelos clientes/utilizadores da(s) autoestrada(s). Cada cliente é caracterizado por: código de cliente (identificador único); nome; morada; código postal; e, número de identificação fiscal. A cada código postal do cliente corresponde a respetiva localidade. Cada cliente pode possuir um ou mais identificadores, em função do número de veículos que possui a circular na(s) autoestrada(s) da concessionária.

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 às passagens dos veículos nos pórticos das autoestradas, 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 a empresa pretende efetuar as seguintes análises de dados:

- Matricula dos veículos que nunca passaram em determinado pórtico
- Pórticos em que nunca passou uma determinada matrícula
- Dias em que uma dada matricula nunca passou em determinado pórtico

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

Falso - É importante manter um registo da origem dos dados.

2. A fase de carregamento de uma dimensão envolve a eliminação e recriação das restrições chave primária – chave estrangeira existentes com a(s) tabela(s) de factos.

Falso - Este processo é apenas aplicado a tabelas de factos

3. Uma das vantagens do *Multidimensional On-Line Analytical Processing* (MOLAP) é que necessita de menos espaço de armazenamento do que o *Relational On-Line Analytical Processing* (ROLAP).

Falso - É ao contrário

Agregados -> sumário da informação (conjunto de medidas), guardadas na fact table com o propósito de apoiar queries

Vantagens --> performance, transparente aos users e partilhada pelos mesmos

Técnicas -> group by na staging por atributo, group by por operação no data mart, adição de dados atômicos
