

Número: _____ Nome: _____

Grupo I - Modelação Dimensional (6,5 valores)

Uma determinada empresa fornece serviços de alojamento de sítios *web* (i.e., *websites*) a diversos clientes empresariais. Cada sítio *web* alojado é caracterizado por: identificador único; nome; cliente proprietário do sítio *web*; data em que ficou acessível na *web*; e, se for caso disso, data em que deixou de estar on-line.

Um sítio *web* é composto por um conjunto de páginas *web*. Cada página é caracterizada por: identificador único; título; data em que a página foi colocada on-line; data em que deixou de estar on-line, se tal for o caso; e, nome físico do ficheiro que corresponde à página. Cada página é composta por diversos tipos de elementos (e.g.: cabeçalho; parágrafo; imagem; tabela; hiperligação; caixa de texto; botão de comando; etc.) que se podem repetir ao longo desta. Cada tipo de elemento é caracterizado por um identificador único e pelo respetivo descritivo.

Na sua interação com as páginas *web*, os utilizadores produzem diversos eventos (e.g.: clique; duplo clique; digitação de caracteres; *scroll* vertical, etc.). Cada evento diferente suscetível de ocorrer numa página é caracterizado pelo identificador único e pelo correspondente descritivo. Os eventos produzidos são registados, mas apenas para os utilizadores que se encontram registados num sítio *web*. Para efeitos de análise de dados, considerou-se que apenas se justifica fazê-lo para estes utilizadores e não para utilizadores anónimos (i.e., não registados). Sobre um utilizador que se regista num sítio *web* é armazenado: identificador único; nome de utilizador; *password*; nome; morada; código postal; e, país. A cada código postal corresponde a respetiva localidade.

Sempre que um utilizador produz um evento numa página de um determinado sítio *web*, é registada a data e hora em que tal aconteceu, bem como o tipo de elemento envolvido. Para efeito das análises de dados pretendidas, é suficiente saber-se o evento e o tipo de elemento envolvido (e.g., clique em hiperligação) e não o elemento concreto da página *web* em que tal ocorreu (e.g., uma página pode conter mais que uma hiperligação).

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 nas páginas dos sítios *web* alojados pela empresa, de acordo com a realidade descrita. Apresente todos os factos, dimensões, granularidade e aspetos relevantes para o projeto de *data mart*.

Metodologia de Kimball:

Área de negócio: Eventos produzidos em páginas web

Granularidade: Eventos por página de sítio *web* por utilizador por tipo de elemento por evento por utilizador por data e por hora

Dimensões: sítio web; página; utilizador; tipo de elemento; evento; data; hora

Factos: não tem

Modelo Dimensional:

DimSítioWeb

SítioKey (PK)
SiteID
Nome
ClienteProprietário
DataOnlineKey
DataOfflineKey

DimPáginaWeb

PáginaKey (PK)
PáginaID
Título
DataOnlineKey
DataOfflineKey
NomeFísico

DimTipoElemento

TipoElementoKey (PK)
TipoElementoID
Descritivo

DimTipoEvento

TipoEventoKey (PK)
TipoEventoID
Descritivo

DimUtilizador

UtilizadorKey (PK)
ClienteID
Login
Password
Nome
Morada
CódigoPostal
Localidade
País
DataEfetivo
DataExpiração
Atual

DimData

DataKey (PK)
DataCompleta
Ano
Mês
Dia
Semestre
Trimestre
SemanaDoAno
DiaDaSemana
...

DimHora

HoraKey (PK)
HoraCompleta
Hora
Minuto
Segundo
Período
HoraAlmoço
HoraJantar
...

FactEventos

DataKey (PK)
HoraKey (PK)
SítioKey
PáginaKey
TipoElementoKey
TipoEventoKey
UtilizadorKey (PK)

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. O espaço de armazenamento de um modelo dimensional no esquema em floco de neve é inferior ao espaço de armazenamento do mesmo modelo no esquema em estrela.

Verdadeiro. Embora o modelo em floco de neve seja, fundamentalmente, desnormalizado, possui um maior grau de normalização do que o modelo no esquema em estrela. Como o modelo em floco de neve é mais normalizado ocupa menos espaço de armazenamento do que o correspondente modelo no esquema em estrela.

2. Para aumentar o desempenho do carregamento de dados numa dimensão, justifica-se sempre a criação de um índice no atributo representativo da chave primária da tabela correspondente no sistema operacional (e.g.: *CardID*, *PaymentMethodID*).

Falso. Na maioria das situações justifica-se a criação de um índice no atributo representativo da chave primária da tabela correspondente no sistema operacional. No entanto, se a dimensão for muito pequena (por exemplo, *DimStore* ou *DimPaymentMethod* do exemplo das aulas PL), não há qualquer ganho que justifique a criação de um índice. Nestes casos, a dimensão acaba por ser armazenada em *cache*.

3. A existência de um atributo do tipo dimensão degenerada (em inglês: *degenerated dimension*) na tabela de factos nem sempre se justifica.

Falso. A existência de um atributo do tipo dimensão degenerada (e.g., *TransactionID*) justifica-se sempre na tabela de factos como forma de manter ligação ao sistema operacional. Esta ligação é especialmente importante na manutenção da rastreabilidade dos dados.

Grupo IV – Questão de Desenvolvimento

(2,5 valores)

Durante o processo de desenvolvimento de um armazém de dados surge, frequentemente, o problema dos valores monetários estarem em moedas diferentes (e.g., tal como aconteceu no trabalho prático). Dependendo da situação, há três abordagens/estratégias diferentes que podem ser adotadas para solucionar o problema. **Explique e exemplifique**, ao nível do modelo dimensional, cada uma dessas três abordagens/estratégias.

1ª abordagem/estratégia:

Repetir cada atributo monetário nas diferentes moedas em que se pretende efetuar análises de dados. Por exemplo, caso se pretenda efetuar análises de dados em euros, libras e dólares, cada atributo monetário (e.g., *PreçoUnitário*) da tabela de factos teria de estar em triplicado para cada uma das moedas (e.g., *PreçoUnitário_euros*; *PreçoUnitário_libras*; *PreçoUnitário_dolares*).

2ª abordagem/estratégia:

Adotar uma moeda de uniformização que permita a realização de análises globais de dados. Cada atributo monetário (e.g., *PreçoUnitário*) deve ser duplicado na moeda local (e.g., *PreçoUnitário_local*) e na moeda de uniformização (e.g., *PreçoUnitário_standard*). Aquando do carregamento dos dados, o valor na moeda local (seja qual for) tem de ser transformado nessa moeda de uniformização. Para que se saiba a que moeda corresponde os valores monetário armazenado nos atributos referentes à moeda local, é necessária a existência de uma dimensão moeda, com o seguinte esquema: *DimCurrency*(*CurrencyKey*; *CurrencyID*; *CurrencyName*; *CurrencySymbol*).

3ª abordagem/estratégia:

Seguir a mesma abordagem que a descrita na 2ª abordagem/estratégia, mas incluir também uma tabela de factos de câmbios que permitirá realizar análises de dados globais em qualquer moeda, para além daquela que tenha sido adotada como moeda de uniformização. Esta tabela possui o seguinte esquema: *FactCurrencyRate*(*Date*; *SourceCurrency*; *DestinationCurrency*; *ExchangeRate*).