

Número: _____ Nome: Pinto da Costa

Modelo dimensional bastante semelhante à época normal do mesmo ano, não justifica a sua realização. AKA, vejam a resolução do modelo dimensional do outro exame. Obrigado por virem ao meu TEDTalk

Grupo I - Modelação Dimensional

(6,5 valores)

Uma determinada empresa possui um sistema informático de registo das despesas de deslocação efectuadas pelos seus colaboradores. Os colaboradores da empresa são caracterizados por: número/identificador; nome; morada; código postal; localidade; categoria profissional; e, data de admissão na empresa. Os colaboradores da empresa pertencem ainda a departamentos, sendo estes caracterizados por um código e um descritivo (*e.g.*: departamento comercial).

As despesas de deslocação encontram-se associadas a centros de custos (*e.g.*: um centro de custo pode corresponder a um projecto em curso). Um centro de custo é caracterizado por: código; descrição; data de abertura; data de encerramento; e colaborador responsável pelo centro de custo.

Sempre que ocorre uma deslocação (nacional ou internacional) de um colaborador são registados os seguintes dados: local de destino; morada; código postal; localidade; país; data de ida; data de regresso; objectivo/finalidade da deslocação; valor total orçamentado; colaborador envolvido na deslocação; e centro de custo que suporta/financia a deslocação.

As despesas efectuadas são classificadas em função do seu tipo (*e.g.*: combustível; portagens; viagem de avião; táxi). Um tipo de despesa é caracterizado por um código e pelo respectivo descritivo.

Para cada uma das despesas que ocorreram no âmbito de uma deslocação é registado: o valor (na moeda original em que ocorreu), a respectiva moeda, o dia e hora em que ocorreu, e o tipo de despesa a que se refere. Note-se que numa deslocação ao estrangeiro podem estar envolvidas despesas em moedas diferentes (*e.g.*: Euros e Libras).

1. Seguindo a metodologia de *Kimball*, desenvolva o processo de análise dimensional, a fim de definir e criar o esquema conceptual para um *data mart* que permita realizar análises de

dados variadas às despesas de deslocação dos colaboradores, de acordo com a realidade descrita. Apresente todos os factos, dimensões, granularidade e todos os aspectos relevantes para o projecto de *data mart*.

2. Admita que se efectuam sistematicamente análises, por colaborador, ao número total de deslocações efectuadas e ao valor total despendido (em Euros) em cada mês. Acrescente ao modelo dimensional anterior o que for necessário (*i.e.*, tabela de factos e/ou dimensão) para aumentar a performance destas análises/consultas. **Criar a DimMonth**

Grupo II – Escolha Múltipla

(1 valor cada questão correcta/-0,5 cada questão errada)

Nas questões seguintes assinale apenas uma só alternativa correspondendo à que considera correcta.

1. Um facto/medida do tipo *saldo* existente numa tabela de factos de um *data mart* bancário:
 - 🍏 Pode ser adicionado ao longo de qualquer dimensão existente.
 - 🍏 Pode ser adicionado apenas ao longo da dimensão data e/ou tempo.
 - 🍏 **Não pode ser adicionado ao longo de qualquer dimensão.**
 - 🍏 Pode ser adicionado ao longo de todas as dimensões, à excepção da dimensão data e/ou tempo.
2. A existência de chaves de substituição em armazéns de dados justifica-se por:
 - 🍏 Proteger o sistema operacional de alterações que ocorram no armazém de dados.
 - 🍏 Associar uma semântica às chaves do sistema operacional, o que é útil aquando da realização de consultas.
 - 🍏 **Razões de performance.**
 - 🍏 Todos os motivos anteriores.
3. Considere uma situação de chegada de um facto que deveria ter sido carregado há muito tempo atrás e em que as dimensões possuem implementado um mecanismo de SCD de tipo 2. Para proceder ao seu carregamento no armazém de dados é necessário:
 - 🍏 Proceder exactamente da mesma forma como são carregados todos os outros factos.
 - 🍏 **Determinar as chaves de substituição dos registos das dimensões que à data do facto em questão se encontravam em vigor.**
 - 🍏 Determinar as chaves de substituição dos registos das dimensões cuja data de expiração (*expired_date*) seja inferior/menor que a data do facto em questão.
 - 🍏 Armazená-lo numa tabela de factos especial que contenha essas situações anormais.

4. Uma possível estratégia de optimização em armazéns de dados envolve a criação de agregações. As agregações podem ser colocadas em novas tabelas (estratégia 1) ou na própria tabela de factos atómica (estratégia 2).
- 🍏 Na estratégia 1 as tabelas de agregações encontram-se visíveis para os utilizadores.
 - 🍏 Na estratégia 2 é necessário colocar um atributo adicional em todas as dimensões.
 - 🍏 Na estratégia 2 não é necessário qualquer cuidado adicional quando se realizam análises de dados.
 - 🍏 Número de registos inseridos em cada uma das estratégias não é o mesmo.
5. No *On-Line Analytical Processing* (OLAP) há diversas operações de análise de dados que os utilizadores podem realizar, nomeadamente:
- 🍏 Operação de *Roll-up* que consiste em agregar os dados e efectuar análises com um maior nível de detalhe.
 - 🍏 Operação de *Dice* que consiste em extrair um sub-cubo do cubo original.
 - 🍏 Operação de *Slice* que consiste em seleccionar um subconjunto de dados restringindo a análise a duas ou mais dimensões.
 - 🍏 Todas as afirmações que constam dos pontos anteriores são válidas.

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. O atributo *IsCurrent*, cuja finalidade é indicar se um dado registo de uma dimensão é o corrente/actual, é suficiente para implementar um mecanismo de SCD (*slowly changing dimension*) de Tipo 2.

Falso - O *effectiveDate* e *ExpiredDate* também são necessários para esse efeito.

2. Para minimizar o espaço ocupado pelo armazém de dados, os atributos relacionados com o aspecto *tempo* (e.g.: hora, minuto, segundo, período do dia) devem ser armazenados na dimensão *Data*.

Falso - Estes atributos dizem respeito à dimensão *Time*, pois caso contrário, ao serem agrupados na *DlmDate* a sua quantidade de registos iria aumentar exponencialmente

3. Num armazém de dados que é abastecido com base em fontes diferentes (*e.g.*: bases de dados diferentes), não se justifica armazenar a proveniência/origem de cada registo individual.

Falso - É importante possuir um identificador para serem associados às respetivas origens da informação, no caso de problemas de qualidade de dados.

Grupo IV – Questão de Desenvolvimento

(2,5 valores)

Um aspecto essencial num armazém de dados é assegurar a integridade referencial entre a(s) tabela(s) de factos e as dimensões. Explique as diferentes técnicas/formas que podem ser utilizadas para garantir essa integridade referencial.

Para este efeito, existem 3 técnicas:

"Check before loading" - Verificação à priori (adição de factos ou eliminação de registos das dimensões) dos dados. Melhor abordagem.

"Check while loading" - Engine da DB já faz por default (enforcement). No entanto, pode ser um processo lento.

"Check after loading" - Periódicamente, realiza verificações de integridade, procurando chaves estrangeiras inválidas.
