

Número: _____ Nome: António Costa

Ramo: TVI

Grupo I - Modelação Dimensional

(6,5 valores)

Uma determinada empresa possui um sistema que monitoriza o acesso e permanência dos seus colaboradores nas várias salas da empresa. Paralelamente, o sistema é também utilizado para fazer o registo diário de presenças. Para efectuar a monitorização, todas as portas da empresa possuem sensores de passagem que são capazes de identificar/detectar os cartões de identificação que os colaboradores possuem. Sempre que é detectada uma passagem, é registado o instante de tempo e a porta onde esta ocorreu, bem como o colaborador que a originou. O sistema permite, assim, saber quanto tempo (em segundos) um colaborador esteve no interior de uma determinada sala, bem como os instantes de tempo de entrada e de saída (resultante dos dados produzidos pelos sensores das portas).

Cada porta encontra-se identificada por um código, sabendo-se se dão acesso ao exterior ou são internas. Neste caso, é conhecida a sala a que dão acesso. Acerca das portas, sabe-se ainda se são portas de segurança (impedem a passagem do fogo em caso de incêndio) ou não.

As salas encontram-se identificadas por um código, possuem uma determinada função (*e.g.*, sala de reuniões), área e capacidade (em número de colaboradores). Cada sala localiza-se num edifício, que é conhecido pela sua designação. O acesso a uma sala pode ser possível através de uma ou mais portas (*e.g.*, existentes em salas diferentes).

Cada colaborador possui um número de identificação, nome, data de nascimento e categoria profissional. Cada colaborador pertence a um único departamento, sendo este caracterizado por um código, um descritivo (*e.g.*, departamento comercial) e o número de colaboradores que dele fazem parte. Cada colaborador encontra-se associado a um tipo de horário, sendo este identificado por um código, descritivo, total de horas diário, hora de entrada e hora de saída.

1. Seguindo a metodologia *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 aos movimentos 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, e com carácter mensal ao: número de dias de trabalho; número de horas de trabalho; e, número de horas extra. O que acrescentaria ao modelo dimensional anterior (*i.e.*, tabela de factos e/ou dimensão) para aumentar a performance destas análises/consultas. Criaria-se uma dimensão com esse nível de date, criando a DimMonth

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.

1. Entre as arquitecturas de armazéns de dados *BUS* de *Ralph Kimball* e *Corporate Information Factory* de *Bill Inmon* existem diversas semelhanças. Qual das seguintes afirmações não corresponde a uma semelhança existente:
 - ☐ Existência de processos de gestão/manipulação dos metadados.
 - ☐ Existência de processos de extracção, transformação e carregamento dos dados.
 - ☐ Maior potencialidade/flexibilidade do armazém de dados advém do dados serem armazenados ao nível mais elementar.
 - ☒ Exceptuando a *data staging area* todos os outros componentes seguem o modelo dimensional no armazenamento dos dados.
2. Comparativamente a uma dimensão, uma tabela de factos é caracterizada por:
 - ☐ Possuir uma estrutura normalizada, muitos atributos e poucos registos.
 - ☐ Possuir uma estrutura desnormalizada, muitos atributos e poucos registos.
 - ☒ Possuir uma estrutura normalizada, poucos atributos e muitos registos.
 - ☐ Possuir uma estrutura desnormalizada, poucos atributos e muitos registos.
3. A *margem_bruta* ($(\text{valor_vendas} - \text{custo_vendas})/\text{valor_vendas}$) obtida a partir de uma tabela de factos de um *data mart* de vendas constitui uma medida:
 - ☐ Aditiva.
 - ☐ Semi-aditiva.
 - ☒ Não aditiva.
 - ☐ Não pode ser classificada em qualquer uma das alternativas anteriores.

4. Índices do tipo *Bitmap*:

- ☒ Especialmente úteis quando combinados em *queries* que envolvem múltiplos atributos.
- ☐ Podem ser usados como índice da chave primária de uma tabela de dimensão.
- ☐ Podem ser criados sobre múltiplos atributos, formando um índice composto.
- ☐ Possuem todas as características que constam das afirmações anteriores.

5. Comparativamente ao MOLAP (*Multidimensional OLAP*), o ROLAP (*Relational OLAP*) possui a seguinte vantagem:

- ☐ Melhor performance, devido à utilização de indexação específica e optimização ao nível do armazenamento.
- ☐ Menor espaço em disco, resultante da utilização de técnicas de compressão.
- ☒ Maior escalabilidade, em virtude de suportar a manipulação de volumes de dados muito elevados.
- ☐ Cálculo automático de agregações a partir da granularidade existente.

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. A existência de atributos do tipo *EffectiveDate* e *ExpiredDate* ou *IsCurrent* é dispensável na implementação de um mecanismo de SCD (*Slowly Changing Dimension*) – Tipo 2.

Falso - Uma SCD de tipo 2 implica a adição de um novo registo e a atualização da coluna, *Expireddate*, no anterior (antigo). Estes atributos são importantes, pois permitem identificar o registo atual em uso e os registos com valores passados e os seus intervalos temporais respetivos.

2. Uma dimensão *Junk* é utilizada em armazéns de dados para armazenar as combinações possíveis de um conjunto de atributos demográficos, o que permite a sua remoção de uma dimensão.

Falso - Para esse efeitos, são usadas minidimensões e não dimensões junk.

-
-
3. No particionamento horizontal (estratégia de optimização em armazéns de dados), os dados apenas podem ser particionados/separados em intervalos de valores.

FALSO- Para além de Range partitioning, também existe list partitioning

Grupo IV – Questão de Desenvolvimento

(2,5 valores)

Em determinadas situações há necessidade de proceder à correcção de factos já anteriormente carregados na tabela de factos. Explique as formas que podem ser utilizadas para proceder a essa correcção.

Para realizarmos a correcção de factos, possuímos 3 maneiras de o fazer:

1- Negar o facto e recarregá-lo. Isto implica criar um duplicado, negar todas as suas medidas, assim elas "cancelam-se".

Isto realiza-se por motivos de auditoria (razão principal) e medir/capturar entradas defeituosas é importante na perspetiva de negócio (razões analíticas)

2- Atualizar o facto, através de uma simples query SQL (Update)

3- Apagar o facto e recarregá-lo. Através de uma eliminação física (o registo é, efetivamente, apagado) ou através de uma eliminação lógica, o registo fica marcado como "Deleted".
