

Número: _____ Nome: _____

Grupo I - Modelação Dimensional (6,5 valores)

O Serviço Nacional de Saúde (SNS) possui um sistema informático que centraliza e armazena dados sobre as consultas médicas que realiza a nível nacional. Uma consulta médica decorre numa dada instituição de saúde (centro de saúde ou hospital) que integra a rede do SNS, sendo realizada por um médico a um utente. Assim, sempre que ocorre uma consulta médica é registada: a instituição onde decorreu; os intervenientes (i.e., o médico e o utente/paciente); a respetiva especialidade; data; hora de início; e, hora de término da consulta.

Uma instituição de saúde do SNS é caracterizada por: identificador único; nome (e.g., Centro de Saúde de Paranhos); morada; distrito; concelho; código postal; data de criação/abertura; e, contacto telefónico. Numa instituição de saúde trabalham/colaboram diversos médicos de várias especialidades, podendo inclusive um médico trabalhar em mais do que uma instituição de saúde. Cada médico é caracterizado por: número de registo na ordem dos médicos (identificador único); data de registo na ordem dos médicos; nome; género; data de nascimento; morada; código postal; e, contacto telefónico. Em certos casos, um médico pode ter habilitação para exercer mais do que uma especialidade (e.g., medicina interna e cirurgia geral).

As consultas encontram-se organizadas por especialidades médicas. Cada especialidade é caracterizada por: identificador único de registo na ordem dos médicos; designação da especialidade (e.g., pediatria; cardiologia); data de criação da especialidade (pela ordem dos médicos); e, número total de médicos (a nível nacional) que exercem a especialidade. As consultas visam fazer o acompanhamento médico dos utentes do SNS. Relativamente aos utentes, além de um identificador único (número de utente), estes são também caracterizados por: nome; data de nascimento; género; morada; código postal; e, contacto telefónico. Para além destes dados genéricos, um utente é também caracterizado, pela (última) altura conhecida, pelo (último) peso conhecido e pelo tipo de sangue. Estes dois últimos dados são especialmente importantes numa situação de urgência médica.

De notar que em todas as situações anteriores, o código postal encontra-se sempre associado uma localidade.

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 dados sobre as consultas médicas realizadas no SNS, de acordo com a realidade descrita. Apresente todos os factos, dimensões, granularidade e aspetos relevantes para o projeto de *data mart*.

A título de exemplo considere-se alguns exemplos de análises de dados que se pretendem realizar:

- N.º de consultas realizadas por um determinado utente, num determinado intervalo temporal.
- N.º de consultas realizadas por um determinado médico, num determinado intervalo temporal.
- N.º de consultas realizadas por especialidade médica, num determinado intervalo temporal.
- N.º de consultas de uma determinada especialidade médica, realizadas numa dada instituição, num determinado intervalo temporal.

Metodologia de Kimball:

Área de Negócio: Consultas médicas (realizadas no âmbito do SNS)

Granularidade: Consultas realizadas por utente, por médico, por especialidade, por instituição, por data e por tempo.

Dimensões: Utente; Especialidade; Médico; Instituição; Data; Hora

Factos: DuraçãoConsulta

Modelo Dimensional:

DimUtente	DimEspecialidade	DimMédico
UtenteKey (PK) NrUtente Nome DataNascimento Genero Morada Código Postal Localidade Contacto TipoSangue Altura Peso DataInicio DataFim Atual	EspecialidadeKey (PK) EspecialidadeID Descricao DataCriaçãoEspecialidade Nr Médicos	MédicoKey (PK) NrRegistoOrdem DataInscriçãoOrdem Nome Genero DataNascimento Morada CódigoPostal Localidade Contacto DataInicio DataFim Atual

DimInstituição	DimData	DimHora
InstituiçãoKey (PK) InstituiçãoID Nome Morada Distrito Concelho Código Postal Localidade Contacto	DataKey (PK) Fulldate Ano Semestre Trimestre Mês NomeMês Semana NrDiaAno NrDiaMes NrDiaSemana DiaSemana FimDeSemana ...	HoraKey (PK) Fulltime Hora Minuto Segundo Periodo HoraAlmoço HoraJantar ...

FactConsultas
UtenteKey (PK)
EspecialidadeKey
MédicoKey
InstituiçãoKey
DataKey (PK)
HoraInicioKey (PK)
HoraFimKey
DuraçãoConsulta

Em alternativa a PK pode ser formada por: UtenteKey; EspecialidadeKey; MedicoKey; DataKey

Grupo II - Múltipla Escolha

(1 valor cada resposta correta/-0,5 cada resposta errada)

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

Grupo III – Verdadeiros ou Falsos com Justificação

(2 valores cada resposta correcta)

Indique se as seguintes afirmações são verdadeiras ou falsas, apresentando a respetiva justificação.

1. No processo de carregamento de dados, recorrendo ao componente *Slowly Changing Dimension (Integration Services Project)*, a *business key* corresponde necessariamente a um único atributo.

Falso. A *business key* corresponde à chave primária da tabela em questão no sistema operacional. Se a chave primária do sistema operacional for um só atributo, a *business key* será formada apenas por esse atributo. No entanto, se a chave primária do sistema operacional for composta (i.e., formada por mais do que um atributo), a *business key* também será constituída por esses atributos.

2. Numa tabela de factos existe sempre somente um único índice, i.e., o que está inerente à sua chave primária.

Falso. Em **ROLAP (relational OLAP)** justifica-se também a existência de índices individuais para cada um dos atributos que formam essa chave primária, de modo a permitir ao motor de base de dados o melhor desempenho aquando da realização de análises/consultas de dados. Para além disso, pode também justificar-se a criação de índices sobre as próprias medidas da tabela de factos caso estas sejam utilizadas frequentemente em operações que envolvam filtragem de registos (e.g., n.º total de vendas em que o valor da venda foi superior a 200€).

3. No *SQL Server* é possível definir-se um índice do tipo *Bitmap* para atributos de dimensões que apresentem baixa cardinalidade (i.e., reduzido n.º de valores distintos).

Falso. O *SQL Server* não suporta índices do tipo *Bitmap*.

Grupo IV – Questão de Desenvolvimento

(2,5 valores)

Em função das características dos sistemas fonte podem ser utilizadas duas abordagens para a extração dos dados, nomeadamente: extração estática e extração incremental. Explique cada um destes tipos de extração e o respetivo cenário/contexto (do sistema fonte) em que são utilizados.

A extração estática consiste em extrair sempre todos os dados do sistema operacional (ou sistemas operacionais) de cada vez que é necessário carregar dados para o armazém de dados, o que normalmente acontece diariamente. Este tipo de extração é utilizado na primeira vez que se carrega o armazém de dados e necessariamente de todas as vezes seguintes sempre que o sistema operacional não possua forma de identificar os novos registos e os alterados desde o último carregamento (e.g., tal como aconteceu no caso prático usado nas aulas PL).

A extração incremental consiste em extrair do sistema operacional (ou sistemas operacionais) apenas os dados/registos que foram inseridos ou alterados desde o último carregamento. Neste tipo de extração não é necessário estar sistematicamente a extrair todos os dados do sistema operacional, mas apenas aquilo que verdadeiramente é relevante. Para que seja possível de implementar, entre outras alternativas, é necessário que o sistema operacional disponibilize metadados acerca da data de inserção e data de última atualização de cada registo.