

Número: \_\_\_\_\_ Nome: Dragon ball gt

### Grupo I - Modelação Dimensional (6,5 valores)

A Liga Portuguesa de Futebol possui um sistema informático que armazena variados dados sobre as diversas competições futebolísticas (amadoras e profissionais) em Portugal que decorrem anualmente, no âmbito de cada época desportiva. Em particular, sobre cada competição é armazenado: um identificador único, a descrição (e.g.: Liga Portuguesa; Taça da Liga), o n.º de jornadas ou eliminatórias, a indicação se é uma competição profissional ou amadora, a data em que a competição foi fundada e, se for o caso, a data em que a competição foi extinta. Cada época é caracterizada por um identificador único; por um descritivo (e.g., época 2019/20); pelo n.º de competições profissionais; pelo n.º de competições amadora; e, pelas datas oficiais de início e fim de época futebolística.

Cada competição futebolística (em cada época) encontra-se organizada em jornadas ou eliminatórias. Cada jornada ou eliminatória é caracterizada por um código único e por um descritivo (e.g.: Jornada 8; meias-finais). A disputar as várias competições encontram-se os clubes. Sobre o clube é armazenado: o seu número de registo na Liga, o nome pelo qual é conhecido, a morada, o código postal, o número de identificação de pessoa coletiva e a data de fundação. Cada código postal encontra-se associado uma localidade.

O sistema operacional da Liga armazena dados sobre os jogos que ocorrem entre clubes, no âmbito de cada jornada ou eliminatória, de cada competição (profissional ou amadora), em cada época futebolística. Em particular, sobre cada jogo realizado é registado: o clube visitado; o clube visitante; a data e hora em que ocorreu o jogo; o n.º de golos marcados pelo (clube) visitado; o n.º de golos marcados pelo (clube) visitante; o n.º de faltas cometidas pelo visitado; o n.º de faltas cometidas pelo visitante; o n.º de cartões amarelos mostrados ao visitado; o n.º de cartões amarelos mostrados ao visitante; o n.º de cartões vermelhos mostrados ao visitado; o n.º de cartões vermelhos mostrados ao visitante; e, o tempo efetivo de jogo (i.e., em que o jogo não esteve parado).

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 dados sobre os jogos de futebol, 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 análises de dados agrupadas por época, competição e clube aos golos marcados, golos sofridos, n.º de faltas cometidas, n.º de faltas sofridas, n.º de cartões amarelos mostrados e n.º de cartões vermelhos mostrados. Explique o que acrescentaria ou alteraria a nível do esquema do modelo dimensional elaborado na alínea anterior, de modo a suportar a realização deste tipo de análises de dados da forma mais eficiente/otimizada.

## 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. No processo de carregamento de uma dimensão com elevado volume de dados, recorrendo ao componente *Slowly Changing Dimension (Integration Services Project)*, deve definir-se um índice único (i.e., não admite valores duplicados) para o atributo que constitui a *business key* (chave primária no sistema operacional).

Verdadeiro. No processo de carregamento de uma dimensão com um elevado volume de dados usando o componente Slowly Changing Dimension (SCD) em um projeto do Integration Services (SSIS), é comum definir um índice único para o atributo que constitui a chave de negócios (business key), que normalmente corresponde à chave primária no sistema operacional.

2. Numa tabela de factos nunca pode existir um atributo cujo tipo de dados seja textual.

Verdadeiro - Têm de ser numéricos

3. Na definição do índice da chave primária da tabela de factos, a posição que o atributo *DateKey* ocupa é irrelevante a nível da performance das consultas/análises de dados.

Falso - É relevante, a posição de DateKey deve ser a primeira

## 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. Dependendo da situação, há três estratégias diferentes que podem ser adotadas para solucionar o problema. Explique e exemplifique a nível do modelo dimensional cada uma dessas três estratégias.

3 formas distintas -

adicionar colunas na tabela de factos,

Adicionar uma moeda de uniformização e uma local

possuir uma nova tabela de factos que faça o câmbio de moedas diferentes