

Número: \_\_\_\_\_ Nome: Chumbado

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

Uma dada companhia de seguros possui um sistema operacional que armazena dados sobre os processos de sinistro resultantes das participações dos seus segurados. Em particular, sobre cada sinistro é armazenada a apólice que o cobre, a data, a hora e o local em que ocorreu. Caso exista um terceiro envolvido no sinistro (e.g., no caso de um sinistro automóvel com outro automóvel), são também armazenados dados sobre esse terceiro. Uma vez que o sinistro pode não ser comunicado no mesmo dia em que ocorreu, a data de participação à companhia de seguros é, igualmente, armazenada. Quando o processo de sinistro é dado por concluído pela companhia de seguros, são armazenados os seguintes dados: data de conclusão; valor pago ao segurado como indemnização (que no limite pode ser zero); valor pago ao terceiro (que pode até não existir) como indemnização (que no limite pode ser zero); e, o valor total pago decorrente do sinistro.

Cada apólice é caracterizada por: um n.º de apólice (sempre diferente em cada apólice); um ramo; pela data em que foi constituída; eventualmente, pela data de anulação (por iniciativa do segurado ou pela própria companhia); e, naturalmente, pelo segurado que a subscreveu. Cada ramo é caracterizado por um identificador único e por um descritivo (e.g., ramo automóvel; ramo multirriscos habitação; ramo acidentes pessoais). A cada apólice encontra-se associado um conjunto de coberturas que descrevem aquilo a que o segurado, de facto, se encontra seguro. Uma cobertura é caracterizada por um identificador único, um descritivo (e.g., furto ou roubo; fenómenos da natureza; atos de vandalismo) e pelo respetivo capital seguro (i.e., valor monetário até ao qual o segurado poderá ser indemnizado em caso de sinistro).

Sobre o segurado são armazenados os seguintes dados: identificador único; nome; morada; código postal; contato telefónico; n.º de contribuinte; e, n.º de cartão de cidadão. Quando existe um terceiro envolvido no sinistro, os dados armazenados sobre este são exatamente os mesmos que os do segurado. Como já referido, sempre que ocorre um sinistro, é necessário registar o local da sua ocorrência. Sobre cada local é armazenado: um identificador único; a morada; o código postal; e, eventualmente, alguma observação (e.g., Km da estrada em que ocorreu um sinistro automóvel). De notar que a um código postal se encontra sempre associada uma dada localidade.

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 processos de sinistro, de acordo com a realidade descrita. Apresente todos os factos, dimensões, granularidade e aspetos relevantes para o projeto de *data mart*.

## 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. A nível de desempenho/performance a nível de consultas/análises de dados, um modelo dimensional no esquema em estrela é superior a um modelo dimensional no esquema em floco de neve.

Verdade - No snowflake quando repartimos tabelas, comprometemos performance

2. Aquando do carregamento da tabela de factos deve previamente proceder-se à eliminação das restrições chave primária – chave estrangeira (*primary key – foreign key constraints*), sendo estas recriadas após esse carregamento.

Verdadeiro - para podermos adicionar novos dados, as restrições têm de ser eliminadas

3. Num processo de carregamento de uma dimensão, recorrendo ao componente de *Slowly Changing Dimension* de um *Integration Services Project* do *Visual Studio*, nunca se justifica a criação de um índice em qualquer dos seus atributos.

Falso - Justifica-se a criação de índice na businessKey

## Grupo IV – Questão de Desenvolvimento

(2,5 valores)

Os índices aumentam consideravelmente a performance da realização de consultas/inquéritos (*queries*), mas penalizam a performance de inserções e atualizações. Num cenário de carregamento de uma tabela de factos (em que há inserções e atualizações de registos), descreva a estratégia adequada de manipulação dos índices, de modo que o carregamento seja feito da forma mais eficiente possível.

Passo a passo, o que fazer:

1- separar inserts de updates

2- Eliminar indexes não necessários para dar suport aos updates

3- Realizar updates

4- Eliminar os outros indexes

5- Fazer inserts

6- Voltar a criar os índices