

Número: _____ Nome: João e Gaspar

Grupo I - Modelação Dimensional (6,5 valores)

Uma Instituição de Ensino Superior de grande dimensão possui um sistema de informação que, entre outros, armazena dados relativos à realização de exames. Um exame pode ser referente à época normal, época de recurso/melhoria, ou época especial, de um dado ano lectivo. Cada ano lectivo é caracterizado por: um código numérico (identificador único), o descritivo (*e.g.*, 2013/14), as datas de início e de fim do 1º e 2º semestre, as datas de início e de fim das épocas de exame do 1º e 2º semestre, e o número de dias de aulas anual (varia em função dos feriados).

Um exame ocorre num determinado dia e hora (de início), possui uma duração e, eventualmente, uma tolerância adicional. Um exame diz respeito a uma determinada disciplina, de um dado curso, sendo a disciplina da responsabilidade de um determinado docente (regente). Cada disciplina é caracterizada por: um código numérico (identificador único), a sigla, e o nome. A mesma disciplina pode ser comum a cursos diferentes (*e.g.*, licenciatura e mestrado), em anos curriculares diferentes (*e.g.*, 3º ano da licenciatura e 1º ano de mestrado), e possuir um número de créditos diferente. Cada curso é caracterizado por: um código numérico (identificador único), o descritivo do curso (*e.g.*, mestrado em engenharia informática), a data em que foi aprovado, a data da última remodelação, o ano lectivo de arranque, e o código do departamento a que pertence. Cada docente é caracterizado por: um código numérico (identificador único), a sigla, o nome, a área de formação (*e.g.*, engenharia informática), o grau académico (*e.g.*, doutoramento), e o código do departamento a que pertence. A cada código do departamento (*e.g.*, DEI) encontra-se sempre associado um descritivo (*e.g.*, departamento de engenharia informática).

Em cada exame de uma disciplina, em cada curso, o sistema de informação armazena dados sobre o número de alunos inscritos, o número de alunos que efetivamente estiveram presentes e, entre estes, o número de alunos que obtiveram aprovação. No caso das disciplinas comuns a cursos diferentes, normalmente o exame é realizado no mesmo dia e hora, mas nem sempre isto acontece.

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 multidimensionais de dados variadas aos exames realizados, 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 pretendem efetuar análises de dados às aprovações/reprovações por sala em que decorrem os exames de uma determinada disciplina, de um dado curso (*e.g.*, em função: do edifício e/ou piso a que a sala pertence; da lotação da sala; da sala possuir ou não climatização). O que acrescentava ao modelo dimensional elaborado na alínea anterior para suportar estas análises ?

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. Um dos motivos para que sistemas operacionais e sistemas de armazéns de dados se encontrem em ambientes distintos/separados resulta da necessidade de:
 - ☐ Grande volume de dados históricos dos sistemas operacionais.
 - ☒ Consultas complexas, envolvendo grande volume de dados, que vulgarmente ocorrem nos sistemas de armazéns de dados.
 - ☐ Consolidação de dados que é necessário realizar nos sistemas operacionais.
 - ☐ Todos os motivos apresentados nas alíneas anteriores.
2. Na arquitetura *CIF* de *Bill Inmon* existe um componente designado de *Operational Data Store* (ODS) que possui a seguinte característica:
 - ☐ Os dados encontram-se armazenados segundo o formato dimensional.
 - ☐ Quase todos os dados encontram-se agregados, existindo apenas alguns no nível mais básico/elementar.
 - ☒ Constitui um repositório de integração de dados em que os dados são atualizáveis.
 - ☐ Além dos dados atuais, também armazena dados históricos.
3. Frequentemente há inúmeros atributos textuais e do tipo booleano (flags) cuja inclusão na tabela de factos parece justificar-se. Nestas situações, a melhor opção a nível de modelo dimensional consiste em:
 - ☐ Colocar todos esses atributos na tabela de factos, mantendo os respectivos tipos de dados.
 - ☐ Colocar cada atributo textual e cada atributo do tipo booleano em dimensões separadas (uma dimensão para cada atributo).
 - ☒ Efetuar combinações dos diversos valores desses atributos e colocá-los numa só dimensão ou mesmo em mais do que uma dimensão, caso as combinações resultem num número muito elevado de registos.
 - ☐ Eliminar todos esses atributos, uma vez que não podem constar de uma tabela de factos.
4. Na extração de dados a partir dos sistemas fonte/operacionais é vulgar utilizar-se:
 - ☐ Extração incremental para carregar o armazém de dados inicialmente e extração estática para os carregamentos seguintes.
 - ☒ Extração estática para carregar o armazém de dados inicialmente e extrações incrementais para os carregamentos seguintes.
 - ☐ Extração estática ou extração incremental de forma indiferenciada no carregamento inicial e nos carregamentos seguintes do armazém de dados.
 - ☐ Outras estratégias de carregamentos mais adequadas que não as anteriores.
5. O *Híbrido On-Line Analytical Processing* (HOLAP) constitui a abordagem que a generalidade das ferramentas comerciais OLAP atualmente utiliza. Seguindo esta abordagem:
 - ☐ Os dados agregados são armazenadas em estruturas de dados relacionais.

- ☐ Os dados no nível de granularidade mais básico/elementar são armazenados em estruturas de dados multidimensionais.
- ☒ Os dados agregados são armazenados em estruturas de dados na memória primária (RAM).
- ☐ Apenas existem dados no nível de granularidade mais básico/elementar, sendo estes armazenados na memória primária (RAM).

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. Abstraindo de todos os outros factores que possam condicionar a escolha do nível de granularidade, a melhor opção consiste em adoptar o nível de granularidade mais elementar.

Verdadeiro - Pois o utilizador poderá fazer todo o tipo de consultas que desejar

2. Considerando que um atributo vai sofrer inúmeras alterações ao longo do tempo, um mecanismo de controlo de alterações de tipo 3 (SCD de tipo 3) é o mais adequado.

Falso - O tipo 3 apenas permite observar a última alteração. Numa situação de várias alterações, não é possível ver o seu histórico.

3. A operação de *drill-down* suportada pelas ferramentas OLAP permite efetuar análises de dados a um menor nível de granularidade/detalhe.

Falso - É o contrário. Numa abordagem drill-down, poderemos fazer consultas com um nível de detalhe maior

Grupo IV – Questão de Desenvolvimento (2,5 valores)

Uma estratégia de optimização que pode ser adoptada em armazéns de dados consiste na criação de particionamentos/partições (normalmente, na tabela de factos). Apresente as vantagens e desvantagens da criação de particionamentos, e explique os diferentes tipos de particionamentos existentes.

As partições, quando realizadas, melhorarão a performance (vantagem), no entanto, a complexidade de gestão de tabelas aumenta.

Existem 2 técnicas para realizar partições: horizontal e vertical.

Na partição horizontal, são criadas partições consoante as linhas/registos das tabelas através de 1 de 2 condições: por intervalo de valores ou por lista de valores.

Na Partição vertical, os valores são separados por coluna e colocados em registos diferentes. É menos usado, o seu esquema difere de partição para partição.
