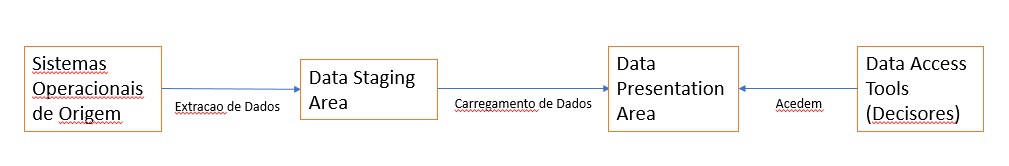
# Parte I — Sistemas de Apoio à Decisão e *Data Warehouses*

1. **2019.06** Desenhe um diagrama com as etapas do sistema ETL de um *data warehouse*, desde que os dados saem dos sistemas operacionais até que chegam aos decisores, e indique uma tarefa concreta de cada uma dessas etapas. Tenha o cuidado de incluir a *data presentation area* e *data staging area* no diagrama.



* Sistemas Operacionais – registam transações
* Dados são extraídos para a data staging área
* Na data staging área os dados são limpos, combinados e padronizados, são conformados a dimensões.
* Os dados depois de tratados são carregados na data presentation área
* A data presentation área contem os data marts que vão apresentar sumários de dados baseados num único processo de negocio
* As data access tools ajudam os decisores a aceder à data presentation área do data warehouse através de ferramentas de interrogação

1. **2019.04** Indique afirmações verdadeiras (V) ou falsas (F) sobre *data warehouses*.

F\_\_ Na abordagem *bottom-up*, os *data marts* são feitos a pedido e derivam do *data warehouse*.

F\_\_ A deteção de alterações nos dados é uma tarefa da etapa de transformação do sistema ETL.

V\_\_ O cálculo e armazenamento em disco de valores agregados são feitos na etapa de carregamento.

V\_\_ A vertente de aplicações de um projeto pode ser realizada autonomamente da vertente de tecnologias.

V\_\_ Facilitam o acesso a dados históricos, sendo esta uma vantagem face às bases de dados federadas.

1. **2018.06** Indique duas razões que levaram ao surgimento de sistemas OLAP, em complemento dos -tradicionais sistemas OLTP, tendo o cuidado de mencionar os respetivos públicos-alvo. Diga também por que razão continuam a existir sistemas OLTP nas organizações.

Os sistemas OLAP surgiram para facilitar a análise de dados e permitir a juncão de dados de varias fontes, sendo estes desenhados com os decisores da organização em mente, pois permitem gerar relatórios mais acessíveis. Os sistemas OLTP continuam a existir na organização pois permitem a entrada de dados na organização, que são depois extraídos e carregados num sistema OLAP.

1. **2018.06** Compare os modelos de dados dos sistemas OLTP e OLAP em função da quantidade e tamanho das tabelas, em termos relativos. Adicionalmente, dê um exemplo de operação de leitura e outro de escrita sobre os dados para cada um destes tipos de sistema. Por fim, indique se são as leituras ou as escritas, ou ambas, que dominam na utilização habitual de cada tipo de sistema.

OLAP- menos tabelas (esquema em estrela, uma tabela de factos ligada a tabelas de dimensão), tabelas com grande numero de colunas. As operações de escrita em sistemas OLAP são apenas realizadas quando os dados são carregados periodicamente ou quando existe necessidade de corrigir dados. As operações de leitura são relatórios gerados para decisores com os dados relevantes para a interrogação. As leituras são predominantes no sistema OLAP.

OLTP- mais tabelas (esquema floco de neve, tabelas ligadas entre si), tabelas com baixo numero de colunas. Operações de escrita são realizadas constantemente, sempre que existe uma nova transação na organização, sendo predominantes no sistema OLAP. As operações de leitura podem ser realizadas com interrogações diretamente sobre o sistema operativo, no entanto estas são de difícil execução e podem prejudicar o desempenho do sistema.

1. **2018.06** Indique afirmações verdadeiras (V) ou falsas (F) sobre bases de dados federadas.

F\_\_ Relatórios de apoio à decisão com dados históricos costumam ser viáveis.

V\_\_ O desempenho dos sistemas operacionais é afetado pelas interrogações analíticas.

V\_\_ O mediador permite abstrair o tipo dos vários sistemas operacionais envolvidos.

F\_\_ As respostas às perguntas analíticas têm dados mais antigos do que com *data warehouses*.

V\_\_ As interrogações analíticas são difíceis de construir pois não existe um modelo dimensional.

1. **2018.04** Indique afirmações verdadeiras (V) ou falsas (F) sobre dados abertos.

V\_\_ Podem ser combinados com outros dados abertos, e redistribuídos livremente.

F\_\_ São capturados de forma passiva por indivíduos, tal como os grandes dados *(big data)*.

F\_\_ Devem ser guardados em formato PDF, pois pode ser lido em quase todos os computadores.

F\_\_ Os enriquecedores colecionam e analisam dados, e cobram pelos seus resultados.

V\_\_ Oferecem fontes de dados externas adicionais a um *data warehouse*.

1. **2018.04** Descreva uma responsabilidade de cada etapa do sistema ETL de um data warehouse, e justifique a importância das mesmas para o apoio à decisão ou para o funcionamento diário da organização. Por exemplo: se essa responsabilidade não existisse, o decisor deixaria de poder fazer isto.

(E)extração – consiste na atualização periódica dos dados contidos no data warehouse através da extração de dados do sistema operacional (OLTP) para a data staging área. Sem esta operação, não seria possível manter os dados do data warehouse atualizados, o que prejudicaria os relatórios a que os decisores têm acesso.

(T)transformação – a etapa de transformação acontece na data staging área, e consiste na limpeza, combinação e conformação dos dados às dimensões, de forma a que os valores sejam consistentes com aqueles que existem já dentro do data warehouse. Sem esta fase, os relatórios gerados pelo sistema podem conter informação incorreta e o data warehouse tem mais vulnerabilidade a erros como duplicação de valores.

(L)Carregamento – a última etapa do sistema ETL é a etapa de carregamento, ou seja, a passagem dos dados transformados na data staging área para a data presentation área, onde os decisores podem ter acesso aos dados transformados. Sem esta etapa, não seria possível aos decisores acederem aos dados, pois as interrogações são sempre feitas sobre os dados que existem nesta área.

1. **2018.04** Preencha o texto em falta em cada afirmação sobre a construção de um *data warehouse*.
   * A atividade de escolha e instalação do produto está incluída na vertente de… tecnologia
   * Cria-se primeiro o *data warehouse*, do qual derivam os *data marts*, na abordagem… top down
   * O desenho e desenvolvimento do sistema ETL faz parte da vertente de… dados
   * Segundo Kimball, o primeiro *data mart* a ser construído é para o… processo de negócio com mais impacto e maior feasibilidade
   * As dimensões mais usadas na organização são mostradas na… bus matrix
2. **2017.04** Indique afirmações verdadeiras (V) ou falsas (F) sobre a gestão de dados mestre.

F\_\_ Só é útil quando a organização dispõe de um *data warehouse*.

V\_\_ Procura atingir o objetivo de haver uma versão única da verdade sobre os dados.

V\_\_ Facilita o carregamento de dados para as dimensões de um *data warehouse*.

F\_\_ A sua introdução numa organização não afeta o funcionamento dos sistemas operacionais.

V\_\_ Requer a atribuição de um dono e de uma autoridade responsável a todos os itens de dados.

1. **2016.06** Os *data warehouses* disponibilizam mais dados, melhores dados, e melhores ferramentas para o apoio à decisão. Indique qual o tipo de sistema informático que está a servir de base de comparação e apresente uma justificação para cada um dos três aspetos referidos.

Os data warehouse estão a ser comparados com os sistemas OLTP. Neste caso, o data warehouse guarda mais dados pois consegue guardar um histórico de dados, melhores dados devido à transformação de dados para serem relevantes a decisores e incorporação de dados de fontes externas na forma de dados abertos para melhor caracterizar os dados provenientes do sistema OLTP e também melhores ferramentas de apoio à decisão porque podem ser criadas vistas que contém os dados relevantes para cada processo de negocio e os relatórios podem ser gerados de acordo com as necessidades de cada decisor.

1. **2016.06** Um projeto típico de construção de um *data warehouse* contempla três vertentes principais logo a seguir ao planeamento e definição de requisitos de negócio, e antes da implantação e exploração na organização. Identifique essas vertentes, justifique se podem ser realizadas autonomamente umas das outras, e descreva o propósito de uma das atividades de cada vertente.

As três vertentes principais de um data warehouse a seguir ao planeamento e definição de requisitos de negócio são as tecnologias, dados e aplicação.

A vertente das tecnologias engloba o projeto de arquitetura (normalmente baseado em plataformas existentes) e a seleção e instalação de software.

Na vertente dos dados define-se a modelagem dimensional (processos, fatos, dimensões, medidas), o projeto físico (acesso ao disco rígido, distribuição e partição de dados, índices, compactação de dados) e o projeto e desenvolvimento de processos ETL (extração, transformação e carregamento de dados para o data warehouse).

Por último, na aplicação/utilização, surge o design das aplicações analíticas (painéis, relatórios dinâmicos) e o desenvolvimento de aplicações analíticas.

As três vertentes principais de um data warehouse, para uma definição clara e planeada do projeto, podem ser realizadas autonomamente entre si, contribuindo para um resultado final comum. Com o desenrolar do processo de negócio (crescimento e manutenção) podem existir alterações posteriores nas três vertentes.

1. **2016.04** Descreva o propósito da gestão de dados mestre numa organização, indique de que forma este serviço deve ser usado pelos sistemas operacionais, e justifique se facilita o desenvolvimento de *data warehouses*.

Os dados mestre estão acima do data warehouse e servem como validação para os dados que são depois introduzidos no data warehouse. Cada instância de um tipo de dados ou mesmo de cada dado, de cada linha da tabela, de dados mestre deve ser completamente conformado e validado. A gestão de dados mestre gere dados críticos comuns aos processos de negócios, focado na qualidade dos dados da organização. A relação dos dados mestre com sistemas operacionais ocorre na medida em que estes respondem a solicitações de dados com valores corretos e coerentes e aceitam novos pedidos de registo de dados para validação. Em relação ao desenvolvimento de data warehouses, a gestão de dados mestre, simplifica a limpeza e integração de dados de várias fontes e facilita o carregamento de dados em dimensões conformadas mestre. Desta forma é esperada uma redução de custos devido a dados inconsistentes e redundantes, maior agilidade para iniciar projetos que dependem de dados confiáveis e melhor conformidade com diretrizes e normas.

1. **2016.04** Indique frases verdadeiras ou falsas (V/F) sobre sistemas de informação para executivos.

F\_\_ Fornecem muitos relatórios predefinidos para cobrir o máximo de cenários de decisão.

V\_\_ Mostram tendências nos dados ao longo do tempo, abrangendo vários anos se necessário.

F\_\_ Disponibilizam e cruzam dados de múltiplas fontes, desde que apenas internas à organização.

V\_\_ Requerem pouco ou nenhum treino para poderem ser usados pelos decisores.

F\_\_ Representam os dados apenas em forma de gráficos, em vez de texto ou tabelas.

1. **2015.06** Justifique qual a ordem de desenvolvimento e operacionalização de sistemas numa cadeia de lojas de roupa: se primeiro um sistema OLAP e depois um OLTP, ou se o contrário. Descreva também duas operações típicas em cada tipo de sistema e o respetivo público-alvo.

Um sistema OLTP deve surgir primeiro que um sistema OLAP. Caso essa cadeia consiga fazer toda a análise com sistemas OLTP, ou seja, esteja mais orientada para transações, com poucos dados, sem preocupação com o histórico e atualizações continuas e tendo em consideração apenas fontes de dados internas, não é necessária a implementação de um sistema OLAP. Caso essa cadeia pretenda iniciar, por exemplo, uma expansão para outras cidades ou países ou criar um site na web, deve optar pelo desenvolvimento de um sistema OLAP, para fazer análises de padrões, dados agregados e relatórios analíticos, com histórico de fontes internas e externas. Sistemas OLTP dizem respeito a processos orientados para o mercado e sistemas OLAP a processos orientados para o cliente. Operações OLTP: INSERT e UPDATE; Operações OLAP: SELECT e GROUP BY

1. **2014.04** Indique afirmações verdadeiras (V) ou falsas (F) sobre sistemas de apoio à decisão.

?\_\_ As transações costumam ser curtas relativamente às dos sistemas operacionais.

V\_\_ As análises de dados tendem a ser exploratórias e a visar estudos de tendências.

F\_\_ A navegação nos dados é versátil, mas requer muito treino dos executivos.

F\_\_ Os dados disponibilizados para a tomada de decisão costumam ser apenas agregados.

F\_\_ Devido ao grande volume de dados o tempo de resposta do sistema costuma ser longo.

1. **2013.04** Justifique a existência de um sistema OLTP numa cadeia de lojas de roupa e descreva duas operações que nele possam ser executadas recorrentemente. Indique também uma razão para o desenvolvimento de um sistema OLAP na mesma organização, tendo o cuidado de mencionar os públicos-alvo dos dois tipos de sistema.

Um sistema OLTP existe numa cadeia de lojas de roupa como forma de registo de transações, para gerar os dados de entrada para um sistema OLAP. Exemplos de operações que podem ser executadas no sistema são o registo de uma transação (compra do produto a, cliente b, vendedor c, preço d) e uma interrogação (qual o preço do produto x). Alguns sistemas OLTP suportam interrogações mais complexas, no entanto isso pode diminuir o desempenho do sistema operacional. Caso essa cadeia consiga fazer toda a análise com sistemas OLTP, ou seja, esteja mais orientada para transações, com poucos dados, sem preocupação com o histórico e atualizações continuas e tendo em consideração apenas fontes de dados internas, não é necessária a implementação de um sistema OLAP. Quando uma empresa pretende fazer analises de tendências utilizando dados históricos, então torna-se necessária a implementação de um sistema OLAP, sendo o publico alvo deste os decisores.

1. **2013.04** Os dados mais recentes dos sistemas OLAP costumam ter um atraso temporal face aos dos sistemas OLTP. Apresente uma razão para esse desfasamento, e justifique se este é ou não aceitável para os decisores, relacionando com a frequência de atualização dos dados OLAP.

Os dados mais recentes dos sistemas OLAP costumam ter um atraso temporal face aos dos sistemas OLTP porque nos sistemas OLAP a atualização é planeada e em sistemas OLTP é atualização contínua. Este desfasamento de informação é aceitável para os decisores porque o objetivo dos sistemas OLAP é a possibilidade de gerar relatórios com dados históricos que permitem análise de padrões, pelo que uma atualização semanal, por exemplo, não vai causar grande impacto em muitas análises históricas.

**2012.04** Descreva as abordagens *top-down* e *bottom-up* de construção de *data warehouses*, indicando o foco inicial dos trabalhos e a forma de expansão para satisfazer necessidades futuras.

A abordagem Top-Down possui uma visão geral que vai do alto nível ao mais detalhado. Em primeiro lugar, detalha o plano completo do data warehouse e cria/transforma dimensões simplificadas em dimensões muito especializadas. Os data marts são construídos por solicitação, a partir de dados provenientes do principal data warehouse. Esta abordagem demonstra-se complexa, já que a visão geral possui grande abrangência, e só é possível começar a utilizar os data marts depois do data warehouse completo estar implementado. Top-Down é indicado apenas quando todo o processo é facilmente compreendido e simples de ser implementado.

Já a abordagem Bottom-Up vai do menor ao maior nível de detalhe. Constrói-se primeiro o data mart para o processo mais importante, sendo os outros data marts acrescentados à medida que existe necessidade deles, e o conjunto de todos os data marts em construção formam o data warehouse. A matriz de negócio funciona como uma linha de orientação para integração de dados e partilha de dimensões. A abordagem Bottom-Up é, tipicamente, a abordagem com melhores resultados, já que o projeto tem o desenvolvimento evolutivo e possui menores riscos.

1. **2012.04** Antes do surgimento, nos anos 70, dos *executive information systems* (EIS), as organizações tinham ao seu dispor *management information systems* (MIS). Indique o tipo de relatórios que eram gerados pelos MIS e apresente duas razões pelas quais os executivos sentiram necessidade dos EIS.

Os management information systems (MIS) são sistemas de informações usados para a tomadas de decisão e para a coordenação, controle, análise e visualização de informações numa organização. Esse estudo envolve, obviamente, pessoas, processos e tecnologia num contexto organizacional. Num ambiente corporativo, como tipicamente acontece, o objetivo final do uso de um sistema de informações é aumentar o valor e os lucros dos negócios. Os MIS são geradores periódicos de relatórios pré-definidos. A partir dos anos 70, passou a existir uma análise de tarefas específicas de decisão, usados por gestores, mas não por executivos. Os Executivos descontentes com a pouca flexibilidade dos relatórios e pouca relevância dos dados fornecidos pelos sistemas de informação e com o aparecimento dos executive information system (EIS) foram fazendo gradualmente a transição. O EIS é um tipo de sistema de suporte à gestão que facilita e dá suporte às informações de executivos superiores e às necessidades das tomadas de decisão, para além do fácil acesso a informações internas e externas relevantes para os objetivos organizacionais. O EIS enfatiza telas gráficas e interfaces fáceis de usar, perceber e analisar (mesmo pelos executivos), uma forte capacidade de gerar relatórios detalhados importantes nos processos de negócio. Através de EIS é possível comparar e destacar tendências, para delinear o desempenho e identificar oportunidades e problemas.

1. **2011.06** Descreva duas razões para o surgimento de sistemas OLAP em organizações que já dispunham de sistemas OLTP, justificando, em particular, que público-alvo mais sentiu a necessidade de mudança.

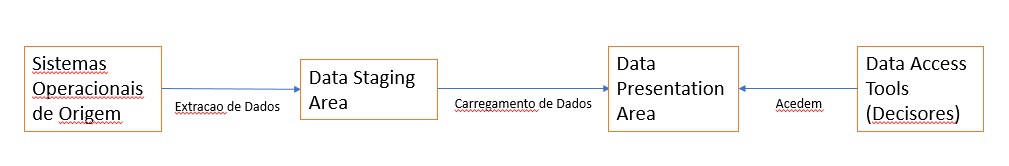
Os sistemas OLAP surgiram em organizações que já dispunham de sistemas OLTP devido à necessidade de analisar com maior profundidade os dados das organizações, com vista a identificar padrões de vendas e oportunidades de negocio, assim como para tornar mais acessíveis os relatórios com as medidas da organização. Os executivos/decisores foram o publico alvo que mais sentiu necessidade desta mudança.

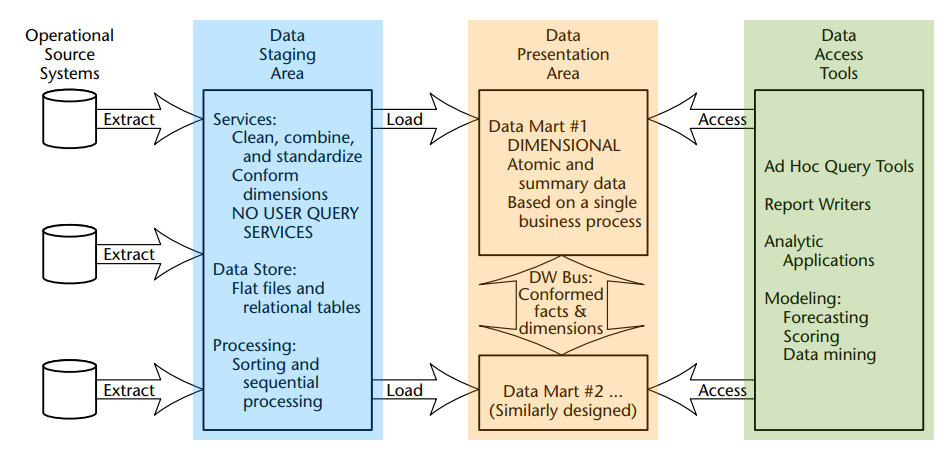
1. **2011.06** Compare os sistemas OLTP e OLAP segundo os critérios seguintes: a) número e origem de fontes de dados usadas nos relatórios; e b) frequência e momento de atualização dos dados.

|  |  |  |
| --- | --- | --- |
|  | OLTP | OLAP |
| a) | Número de fontes reduzido, sendo essas fontes de dados internas e sem histórico | Número de fontes de dados alto, sendo essas fontes de dados internas e externas, com histórico |
| b) | A atualização dos dados é contínua, com uma alta disponibilidade e processamento repetitivo | A atualização dos dados é, tipicamente, planeada, com uma disponibilidade relaxada e processamento exploratório |

1. **2011.06** Enquadre as etapas do processo de *extraction*, *transformation*, e *loading* (ETL) relativamente à *data staging area*, à *data presentation area*, e aos sistemas operacionais, e indique uma responsabilidade específica de cada etapa. Notas: pode fazer um esboço anotado se achar mais prático e não se limite a traduzir os termos para português.

O processo ETL costuma ser um processo complexo, moroso e transversal a toda a construção do data warehouse, pois envolve a extração, transformação e carregamento dos dados. Quando se extraem os dados, estes passam para uma data staging area, que se situa entre os sistemas operacionais e a data presentation data e onde são guardados os dados extraídos dos sistemas operacionais. Na data staging area os dados são limpos, combinados e standardizados. Após a passagem por essa “área de estágio” os dados são transformados e carregados para uma data presentation area, onde os dados prontos a usar são carregados, incluindo já índices e medidas agregadas pré-calculadas, para o data warehouse e posteriormente para os data marts necessários ao processo de negócio.





1. **2011.06** Justifique se interrogações analíticas realizadas sobre uma base de dados federada afetam ou não o desempenho das bases de dados operacionais da organização, bem como se é ou não simplificada a conjugação de dados históricos nessas análises.

As interrogações analíticas realizadas sobre uma base de dados federada colocam problemas aos sistemas operacionais, afetando assim o desempenho das bases de dados operacionais da organização. É dificil fazer otimização de interrogações numa base de dados federada bem como pode ser difícil obter informação histórica nessas análises.

1. **2011.04** Descreva o papel dos sistemas OLTP e OLAP no contexto de uma organização, identifique os públicos-alvo respetivos, e dê um exemplo de operação típica em cada tipo de sistema.

Os sistemas OLTP permitem a entrada de dados na organização, e consistem em processos rápidos e repetitivos que leem a menor quantidade de dados possível. Os sistemas OLTP são orientados para transações. Os sistemas OLAP recolhem dados de varias fontes e facilitam a análise dos dados. O seu público alvo são os decisores.

As operações típicas nos sistemas OLTP são por exemplo transferências bancarias, pagamentos, faturação. Nos sistemas OLAP as operações típicas são analise de produtos mais vendidos num fim de semana, historial de vendas por mês.

1. **2011.04** Compare os sistemas OLTP e OLAP segundo os três critérios seguintes: a) estrutura dos relatórios;

b) frequência de atualização dos dados; e c) abrangência temporal dos dados.

|  |  |  |
| --- | --- | --- |
|  | OLTP | OLAP |
| a) | Relatórios pré-definidos | Relatórios personalizados |
| b) | Constantemente atualizados | Atualizados periodicamente |
| c) | Dados atuais | Dados atuais e históricos |

1. **2011.04** Compare os *data warehouses* e as bases de dados federadas segundo os três critérios seguintes: a) necessidade de cópia dos dados dos sistemas operacionais; b) impacto das interrogações analíticas nos sistemas operacionais; e c) possibilidade de obtenção de dados históricos.

|  |  |  |
| --- | --- | --- |
|  | Data Warehouse | Base de Dados Federada |
| a) | Necessidade de copiar periodicamente a informação do sistema operacional para a data warehouse | Não há necessidade de copiar dados pois a base de dados federada é construída sobre o sistema operacional |
| b) | Impacto baixo | Performance dos OLTP afetada pelas interrogações na base de dados federada |
| c) | Dados históricos obtidos facilmente | Pode não ser possível obter dados históricos |

1. **2010.07** Os *data warehouses* consolidam dados de múltiplas fontes num só repositório. Em alternativa também podem ser usadas bases de dados federadas para responder a interrogações analíticas. Descreva uma desvantagem desta última opção face à primeira.

As desvantagens das bases de dados federadas são o facto das interrogações analíticas penalizarem a performance dos sistemas OLTP, a dificuldade da escrita de interrogações analíticas para os decisores e a dificuldade em obter dados históricos.

1. **2010.06** Explique qual o âmbito típico de um *data mart* dentro do contexto organizacional e apresente duas razões que levam estes últimos a serem usados em vez de os utilizadores acederem diretamente ao *data warehouse*.

O âmbito dos data marts é a possibilidade da criação de subconjuntos ou vistas da data warehouse. Estes últimos são utilizados pelos decisores em vez do acesso direto ao data warehouse pois são mais específicos para cada processo de negócio, o que facilita a análise dos dados, e também pelo facto de serem mais pequenos e mais eficientes.

1. **2009.06** No processo de ETL, justifique que tipo de técnica de mudança lenta é mais difícil de gerir, ilustrando os passos a seguir.

Tipo 3?

1. **2009.04** No contexto de um *data warehouse* defina sucintamente os seguintes conceitos: a) *data mart*; b) *data staging area*; e c) *drill across*.

Data mart – um data mart é um subconjunto ou vista sobre o data warehouse. São utilizados para processos de negócio específicos, de forma a tornair mais eficiente e mais direta a apresentação de dados para os decisores, pois apresentam menos informação que o data warehouse.

Data staging área – A área de data staging armazena dados extraídos de sistemas operacionais. Esta área não é acedida diretamente pelos decisores, os dados aqui armazenados têm de ser primeiro trabalhados para depois serem carregados na data presentation área.

Drill across - operação que permite a juncão de medidas de processos diferentes num mesmo relatório, possível apenas quando os atributos das tabelas de dimensão estão conformados e são partilhados pelos vários processos de negócio.

# Parte II — Relatórios Analíticos

1. **2019.04** Preencha o texto em falta em cada afirmação sobre relatórios analíticos.
   * Uma vantagem das extensões OLAP ao SQL é… simplificar interrogações
   * Ao ser arrastada uma medida para um relatório vazio é exibido… um número que corresponde à soma de todas as linhas que contêm essa medida
   * O arrastar de um atributo de uma dimensão para um relatório serve para… adicionar um termo de agrupamento dos dados (ex group by year)
   * Para obter dados mais detalhados deve ser usada a operação de… drill down
   * *Slicing* é um caso particularmente simples de… dicing
2. **2017.04** Indique afirmações verdadeiras (V) ou falsas (F) sobre relatórios analíticos.

F\_\_ A operação de *dicing* é um caso particularmente simples de *slicing*.

F\_\_ As medidas só podem ser colocadas num relatório após os atributos das dimensões.

F\_\_ Fazer *drill-down* geralmente revela menos valores do que os que estavam disponíveis.

V\_\_ Cada medida M colocada num relatório acrescenta, por omissão, SUM(M) ao SELECT em SQL.

F\_\_ Um atributo A posto nas linhas de um relatório vazio gera SELECT A … GROUP BY CUBE (A).

1. **2016.07** Indique afirmações verdadeiras (V) ou falsas (F) sobre relatórios analíticos em SQL.

F\_\_ Com SQL clássico, o número de SELECTs é igual ao número de dimensões no relatório.

V\_\_ A cláusula GROUP BY CUBE faz parte das extensões OLAP para SQL.

F\_\_ GROUP BY ROLLUP (A, B, C) inclui cálculos agregados em função só de C.

V\_\_ É incorreto usar SELECT A, B, SUM(C) FROM T GROUP BY ROLLUP (A, B, C).

F\_\_ Um só SELECT de SQL clássico pode gerar um relatório com totais em colunas e linhas.

1. **2016.06** No contexto da HyperVending, indique que totais e subtotais seriam calculados para a interrogação SELECT Ano, Sucursal, Cliente, SUM(Euros) FROM Vendas WHERE (Ano BETWEEN 2010 AND 2016) GROUP BY ROLLUP (Ano, Sucursal, Cliente).

Esta interrogação calcularia:

- Soma (Euros) de vendas por cliente por sucursal por cada ano entre 2010 e 2016;

- Soma (Euros) de vendas por sucursal por cada ano entre 2010 e 2016;

- Soma (Euros) de vendas por cada ano entre 2010 e 2016;

- Soma (Euros) de vendas

1. **2016.06** Indique afirmações verdadeiras (V) ou falsas (F) sobre relatórios analíticos.

F\_\_ São calculadas menos somas se for usado CUBE em vez de ROLLUP num GROUP BY.

V\_\_ As extensões OLAP ao SQL permitem otimizar o cálculo de valores agregados.

V\_\_ A operação de *slice* é um caso particularmente simples da operação *dice*.

F\_\_ *Roll-up* serve para procurar explicações detalhadas para os grandes totais.

V\_\_ Podem ser incluídos atributos de mais de duas dimensões num relatório.

1. **2015.06** Indique afirmações verdadeiras (V) ou falsas (F) sobre relatórios analíticos.

F\_\_ Só pode ser aplicada uma operação de *slice* em cada relatório.

V\_\_ As hierarquias de atributos de dimensões possibilitam o *roll-up* e o *drill-down*.

V\_\_ O *roll-up* serve para passar de dados mais detalhados para dados mais agregados.

F\_\_ As extensões OLAP ao SQL não simplificam a escrita de comandos que geram relatórios.

V\_\_ As medidas só podem ser colocadas num relatório após os atributos das dimensões.

1. **2014.04** Suponha um relatório com as vendas em euros por sucursal ao longo dos anos, inspirado no caso HyperVending. Identifique uma hierarquia de atributos numa das dimensões e descreva um cenário de *drill-down*, apresentando uma razão plausível para tal operação.

Uma hierarquia possível seria: ano – semestre – trimestre – mês. Aplicando um drill-down, ou seja, aumentando o nível de detalhe, poderia ser feita uma consulta a partir do 2º trimestre a qual o mês com mais vendas em cada sucursal. Razões plausíveis para essa consulta seria, por exemplo, ver o efeito de uma promoção em determinada sucursal num mês do 2º trimestre.

1. **2014.04** Indique afirmações verdadeiras (V) ou falsas (F) sobre relatórios analíticos.

V\_\_ Os atributos das dimensões podem ser usados para esmiuçar as medidas numéricas.

F\_\_ O *slicing* permite aplicar restrições por intervalo de valores numa dimensão.

F\_\_ A cláusula GROUP BY ROLLUP permite gerar totais para todas as combinações de atributos.

V\_\_ A operação de *dice* pode ser usada para restringir os valores que aparecem num relatório.

F\_\_ O *roll-up* serve para passar de dados mais agregados para dados mais detalhados.

1. **2013.04** Indique duas razões para a inclusão de extensões OLAP na linguagem SQL no contexto dos relatórios analíticos e descreva quais os totais e subtotais calculados na resposta à interrogação SELECT Cliente, Ano, SUM(Unidades) FROM Vendas GROUP BY ROLLUP (Cliente, Ano).

Sem a inclusão de extensões OLAP na linguagem SQL os relatórios analíticos ficariam muito extensos, com muitas interrogações e muita informação, o que o tornaria mais lento o desempenho do sistema e menos acessível a formulação das interrogações. As extensões OLAP na linguagem SQL permitem o cálculo de dados agregados para todas as dimensões e combinações, num formato mais simples de compreender.

Interrogação calcula os totais de unidades por cliente por ano, depois remove a linha mais à direita (ano) e calcula os totais por cliente, remove novamente a linha mais à direita e calcula o total final quando a lista de colunas está vazia.

1. **2012.06** Suponha um relatório dinâmico sobre audiências televisivas, envolvendo programas e espetadores. Identifique uma hierarquia de atributos numa das dimensões e descreva um cenário de realização da operação *roll-up* com uma medida numérica plausível.

Uma hierarquia de atributos numa das dimensões possível para este relatório seria a existência de uma hierarquia na dimensão data, por exemplo de hora, parte do dia (manha, tarde, noite), dia, o que tornaria possível ver o numero de espectadores total num dia, mas também qual a parte do dia onde existe maior numero de espectadores.

1. **2012.06** No contexto da HyperVending, mostre exemplos de linhas resultantes da interrogação SELECT País, Ano, SUM(Euros) FROM Vendas GROUP BY ROLLUP (País, Ano) e indique que linhas adicionais seriam mostradas caso tivesse sido usado GROUP BY CUBE (País, Ano).

Esta interrogação calcularia:

- Soma (euros) de vendas por pais por cada ano;

- Soma (euros) de vendas total por países;

- Soma (euros) de vendas total de todos os países.

Com a GROUP BY CUBE adiciona:

- Soma (euros) de vendas total por ano em todos os países

1. **2011.06** Suponha um relatório dinâmico da empresa HyperVending abrangendo sucursais e produtos. Identifique uma possível hierarquia de atributos e descreva um cenário de realização da operação *rollup*, para o qual deverá também identificar uma medida numérica plausível.

Uma possível hierarquia tendo em conta o cenário descrito poderia ser: Continente – País – Região – Cidade

Uma operação roll-up, ou seja, diminuir o nível de detalhe, poderia ser por exemplo, passar do total de vendas por cidade (Lisboa, Porto, Coimbra) para vendas por país (Portugal).

1. **2011.04** Considere a interrogação SELECT País, Ano, SUM(Euros) FROM Vendas GROUP BY CUBE (País, Ano) do caso HyperVending. Descreva o resultado que seria obtido.

Esta interrogação calcularia:

- Soma (euros) de vendas por pais por cada ano;

- Soma (euros) de vendas total por países;

- Soma (euros) de vendas total de todos os países;

- Soma (euros) de vendas total por ano em todos os países.

1. **2010.07** Suponha um relatório dinâmico que atualmente mostra as vendas em euros por sucursal ao longo dos anos, inspirado no caso HyperVending. Descreva o conteúdo do relatório se fizesse *drill-down* em 2010, bem como a sua assunção sobre a hierarquia de atributos subjacente.

Através do caso descrito, fazendo um drill-down nos dados de 2010, ou seja aumentando o detalhe, seria dado o acesso às vendas em euros por sucursal em cada semestre, trimestre e até mesmo mês ou dia se fosse necessário. Esta hierarquia permitiria ver a variação mais detalhada das vendas em cada sucursal, bem como a influência dos saldos ou datas importantes como o Natal, Dia do Pai ou Dia da Mãe.

1. **2008.04** Considere a seguinte tabela de folha de cálculo, criada para armazenar dados de audiências televisivas: Audiências, NomeDoEspetador, FaixaEtária, ClasseSocEconómica, DimAgregadoFamiliar, Data, HoraDeInício, HoraDeFim, MinutosVistos, NomeDoPrograma, TipoDePrograma, Canal. Nota: HoraDeInício e HoraDeFim referem-se respetivamente ao momento no tempo em que um determinado espetador sintonizou um canal e o deixou.

Usando as extensões OLAP do SQL, refira como faria para construir uma *pivot table* que descrevesse o número de espetadores que viram o programa semanal Batatoon distribuídos por faixa etária, nas semanas de junho de 2007.

SELECT Data, FaixaEtária, DimAgregadoFamiliar, NomeDoPrograma

FROM Audiências

WHERE (MONTH = 6) AND (YEAR = 2007) AND (NomeDoPrograma = “Batatoon”)

GROUP BY CUBE (FaixaEtária, DimAgregadoFamiliar, NomeDoPrograma).

# Parte III — Modelação Dimensional

1. **2019.06** Dê um exemplo de tabela de factos de tipo instantâneo periódico, com pelo menos uma medida, mencionando se cada facto abrange períodos variáveis ou fixos de tempo, se o carregamento de novos dados envolve só inserções, só atualizações, ou ambas as operações de escrita, e justifique se essa medida (ou uma de entre várias que tenha escolhido) é aditiva ou apenas semi-aditiva.

Uma tabela de facto do tipo instantâneo periódico pode ser por exemplo

Fatura mensal eletricidade

- Número fatura

- Mês

- Cliente

- Total euros

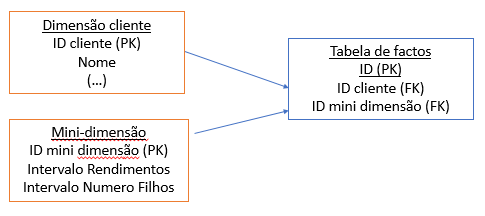
Cada facto abrange um período fixo de um mês, novos dados são carregados principalmente por inserções, a não ser que exista uma atualização devido a erros. A medida ‘e aditiva, pois pode calcular-se o total que foi gasto em energia ao longo das dimensões mês e cliente.

1. **2019.04** Explique de que forma as bifurcações e as minidimensões podem ajudar a controlar o crescimento de dimensões “monstras,” salientando o que têm em comum bem como as principais diferenças, e dando um exemplo adequado a cada uma dessas duas técnicas.

As bifurcações e as mini dimensões tem a vantagem de juntar numa tabela separada da dimensão alguns valores que se alteram ao longo do tempo, evitando um crescimento rápido da dimensão. As bifurcações são geralmente utilizadas para informação que muda de forma muito lenta, relacionada entre si e que muda em conjunto. Um exemplo de uma bifurcação será a informação demográfica da região onde um cliente habita. As bifurcações estão ligadas diretamente à dimensão através de uma chave estrangeira. As mini dimensões são utilizadas para guardar informação inserida em intervalos, que geralmente muda mais frequentemente do que uma bifurcação. A informação contida numa mini dimensão não necessita de estar relacionada entre si. Um exemplo de uma mini dimensão poderá ser a informação demográfica do cliente, por exemplo o seu intervalo de rendimentos, faixa etária ou numero de filhos. As mini dimensões geralmente estão ligadas à tabela de factos.

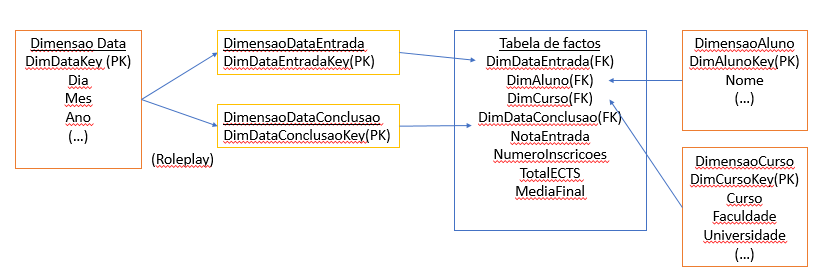
1. **2018.06** Complete cada afirmação sobre modelação dimensional.
   * A seguir à modelação das dimensões vem o passo de… identificação das medidas numéricas
   * Inventários semanais de armazéns, de natureza periódica, são aditivos ao longo da dimensão… armazém (localização)
   * A percentagem de espaço ocupado pelas tabelas de factos costuma ser aproximadamente…90%
   * Uma tabela de factos sem factos pode ser usada para… guardar factos sem medidas de negócio
   * A técnica de tipo 1 para mudanças lentas não permite… guardar histórico das mudanças
2. **2018.06** Considere a dimensão Cliente, com milhões de linhas e em rápido crescimento. Os decisores têm-se queixado da lentidão com que os relatórios são gerados, em particular os que incluem a evolução histórica dos atributos Rendimentos e NúmeroDeFilhos. Supondo que estes atributos são movidos para uma minidimensão, descreva as adaptações que seria necessário fazer aos dados dos atributos, desenhe um diagrama com o novo esquema em estrela, descreva um caso em que a minidimensão permitiria abrandar o crescimento da dimensão, e explique como poderia ser analisada a evolução histórica dos referidos atributos.

As adaptações necessárias aos atributos são a criação de intervalos para as medidas que tem uma maior evolução histórica como o atributo rendimentos e numero de filhos. A mini dimensão permite abrandar o crescimento da dimensão devido a utilização de intervalos em vez de valores concretos, o que diminui a quantidade de vezes que estes valores têm de ser atualizados. O histórico de mudanças é guardado através da atualização da chave da mini dimensão aquando da introdução de novos factos na tabela de factos.



1. **2018.06** Desenhe um esquema em estrela adequado ao registo de factos como o seguinte: no dia D1, o aluno A foi admitido no curso C da faculdade F da universidade U, com nota de entrada NE, tendo concluído o curso no dia D2, depois de se ter inscrito em L anos letivos, nos quais fez um total de T ECTS, com média final de MF valores. Indique também, para cada medida, se é ou não aditiva, justificando sucintamente, e proponha uma medida adicional que seja aditiva.

Para este esquema em estrela, as medidas são: nota de entrada (não aditiva, não faz sentido somar notas de entradas de alunos), numero de inscrições (pode ser aditiva, pode fazer sentido por exemplo investigar quantas inscrições para conclusão de curso), total ECTS (poderá ser aditiva em alguns casos, mas se o curso tiver um numero de ECTS certo para conclusão, pode não fazer sentido ser aditiva) e média final (não aditiva, não faz sentido somar médias). Um exemplo de uma medida adicional que seria aditiva poderia ser o total de propinas pago pelo aluno.



1. **2018.06** Indique afirmações verdadeiras (V) ou falsas (F) sobre tabelas de ponte.

V\_\_ Guardam caminhos entre todos os elementos de todos os níveis de uma hierarquia.

V\_\_ Permitem agregar medidas para cada elemento abaixo de um dado elemento.

V\_\_ São habitualmente usadas entre uma dimensão e uma tabela de factos.

F\_\_ Não podem ser usadas para representar hierarquias de profundidade variável.

F\_\_ A remoção de um elemento da hierarquia tem pouco impacto nas linhas da tabela.

1. **2018.06** Indique três critérios para a escolha de atributos a mover para uma bifurcação de uma dimensão “monstra”. Depois, para apenas um desses critérios, justifique se o mesmo também se aplica às minidimensões.

Os critérios para a escolha de atributos para uma bifurcação de uma dimensão monstra são:

- Atributos relacionados entre si

- Mudam juntos

- Mudam pouco frequentemente.

Para o caso das mini dimensões, não se aplica o critério de os atributos mudarem juntos, poe exemplo uma mini dimensão com atributos demográficos de um cliente pode conter informação sobre a faixa etária e a faixa de rendimentos, que podem mudar independentemente um do outro.

1. **2018.04** Indique afirmações verdadeiras (V) ou falsas (F) sobre modelação dimensional.

F\_\_ Relatórios já existentes devem ser a base de partida para a modelação.

V\_\_ As tabelas de factos sem factos servem para acompanhamento simples de eventos.

F\_\_ A técnica 3 para mudanças lentas é adequada para valores antigos e atuais de muitas colunas.

V\_\_ Uma dimensão conformada tem nomes e valores de atributos comuns em vários processos.

F\_\_ Chaves substitutas em dimensões impedem o registo do histórico de mudanças lentas.

1. **2018.04** Complete cada afirmação sobre modelação dimensional.
   * A decomposição de uma tabela de dimensão em tabelas mais pequenas designa-se… snowflaking
   * Saldos mensais de contas, de natureza periódica, são aditivos ao longo da dimensão… cliente
   * As dimensões que não têm uma tabela própria chamam-se… dimensões degeneradas
   * A mesma tabela de dimensão participar várias vezes num facto designa-se… roleplaying
   * A seguir à identificação dos processos de negócio a modelar, vem a etapa de… declarar o grão da tabela de factos
2. **2017.04** Indique afirmações verdadeiras (V) ou falsas (F) sobre a modelação dimensional.

F\_\_ A escolha do processo de negócio é feita depois de comparadas tabelas de factos alternativas.

F\_\_ Cada processo de negócio tem a sua própria matriz de exequibilidade/valor.

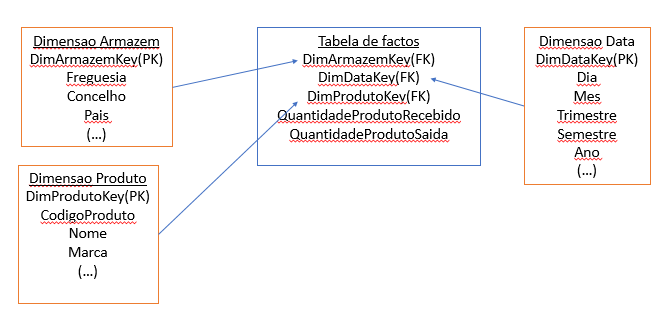
V\_\_ As medidas só devem ser identificadas depois de definidos os atributos das dimensões.

F\_\_ Os processos prioritários estão no quadrante superior direito da matriz de processos.

V\_\_ O grão determina os conceitos participantes em cada evento guardado na tabela de factos.

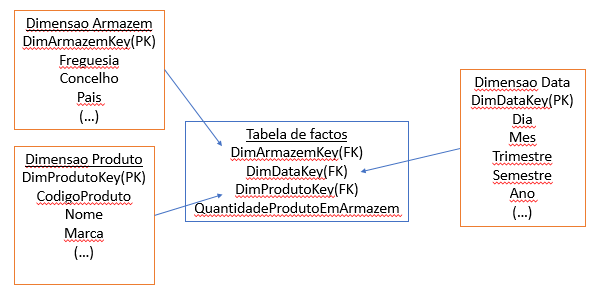
1. **2017.04** Considere os seguintes factos referentes a um só contexto: o armazém A recebeu U1 unidades do produto P no dia D1; e, do mesmo produto P saíram U2 unidades do mesmo armazém A no dia D2.
   1. Assumindo que se pretendem manter distintos os dois factos (e outros semelhantes), indique o tipo de tabela de factos e desenhe um esquema em estrela contendo as tabelas de factos e de dimensões, medidas numéricas, e vários atributos plausíveis, para além dos essenciais.

Para este caso poderia ser utilizado uma tabela de factos transacional, como no diagrama seguinte:



* 1. Pretende-se ter uma segunda tabela de factos, que regista a quantidade de produtos de cada tipo em cada armazém no final de cada dia. Indique o tipo desta tabela de factos, desenhe o esquema em estrela, mencione que dimensões podem ser reutilizadas (não precisa mostrar os seus atributos), e justifique se as medidas são aditivas.

Para este caso, a tabela de factos seria do tipo snapshot periódico. Todas as dimensões podem ser reutilizadas.



1. **2017.11** Considere uma dimensão onde está a ser usada a técnica de tipo 1 para registar mudanças lentas. Indique uma limitação dessa técnica, que pode ser ultrapassada com a adoção da técnica de tipo 2, e explique que alterações seria necessário fazer na estrutura e modo de utilização da tabela da dimensão.

A utilização da técnica de tipo 1 para registar mudanças lentas substitui o valor antigo pelo valor novo, não permitindo guardar um histórico das alterações. Para adotar uma técnica do tipo 2, serão adicionadas novas linhas à tabela da dimensão para cada mudança, cada uma destas linhas com a sua própria chave primaria, tendo, no entanto, de existir a chamada chave supernatural, que se mantem constante e que permite ver o histórico das mudanças.

1. **2017.04** Compare as bifurcações e as minidimensões em termos de: a) frequência de atualização dos valores dos atributos; e b) relação entre os atributos. Apresente também um exemplo adequado a uma destas técnicas para controlar o crescimento de dimensões “monstras”, justificando.

|  |  |  |
| --- | --- | --- |
|  | Bifurcacoes / Outriggers | Mini dimensões |
| a) | Atributos mudam lentamente | Atrubutos mudam com alguma frequência |
| b) | Atributos relacionados entre si, mudam juntamente | Atributos não necessitam de estar relacionados entre si |
| Exemplo | Informação demográfica da região onde um cliente vive (população, população idosa, população jovem) | Informação sobre a demografia de um cliente (faixa etária, intervalo de rendimentos) |

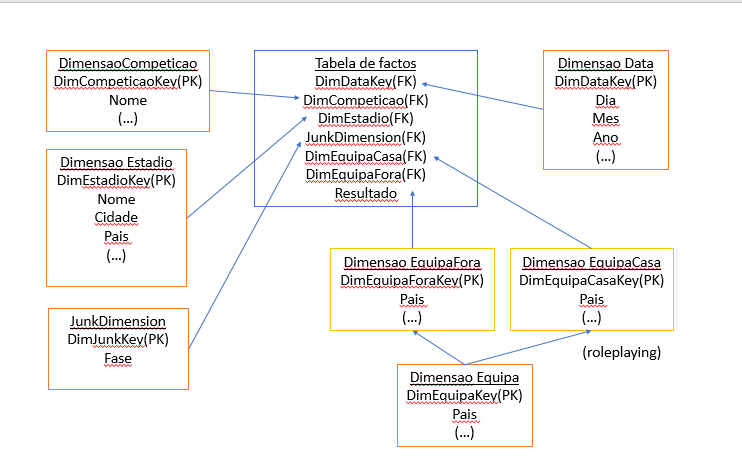
1. **2016.07** Considere este facto: no jogo da final do campeonato europeu de futebol, realizado no dia 10 de julho de 2016, no estádio Stade de France com 80000 lugares e lotação esgotada, que fica na cidade de Saint-Denis, a equipa de Portugal venceu a da França por 1 a 0.

Supondo que se pretende uma tabela de factos para guardar todos os jogos (ex. também quartos de final e meias finais) de campeonatos com equipas nacionais (ex. também os mundiais de futebol), indique: o tipo de tabela de factos, as medidas numéricas (caso existam), as dimensões, se há *role playing*, e alguns atributos representativos, incluindo chaves substitutas e estrangeiras. Nota: pode desenhar um diagrama de dados.

O tipo da tabela de factos é transação.

Medida numérica é o resultado

Dimensões: Data, Competição, Estádio, Equipa (tem roleplay), Junk



1. **2016.07** Indique afirmações verdadeiras (V) ou falsas (F) sobre esquemas em estrela.

F\_\_ Costumam ser usados em *data warehouses* e também em sistemas operacionais.

V\_\_ A tabela de factos está ligada a todas as tabelas de dimensões.

F\_\_ A dimensão com mais atributos está tipicamente no centro do esquema.

V\_\_ As medidas só são permitidas na tabela de factos.

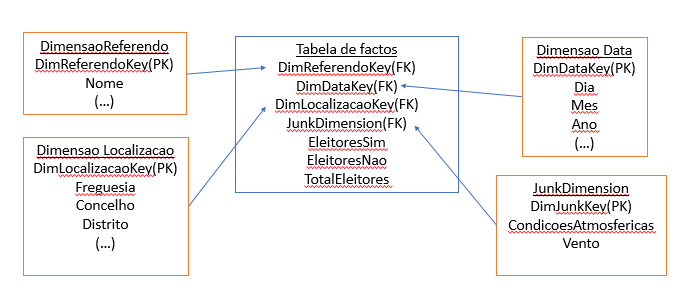
F\_\_ É comum duas ou mais dimensões estarem ligadas diretamente entre si.

1. **2016.06** Para o facto seguinte, indique o tipo de tabela de factos, as medidas numéricas (caso existam), as dimensões, se há *role playing*, e alguns atributos representativos, incluindo chaves substitutas e estrangeiras: no referendo R, do dia D, houve S eleitores a votar ‘sim’ e N a votar ‘não’, de um total de T eleitores da freguesia F, do concelho C, e distrito I, sendo também registadas as condições atmosféricas nessa freguesia e dia (ex. se choveu, nevou, ou fez sol, se houve vento forte ou fraco). Nota: pode desenhar um diagrama de dados.

Tipo de tabela: Transacao

Medidas numéricas: EleitoresSim, EleitoresNao, TotalEleitores

Dimensoes: Referendo, Localizacao, Data, Junk



1. **2016.06** Considere a dimensão Cliente, com milhões de linhas, na qual está a ser usada a técnica 2 para guardar o histórico de mudanças que vão ocorrendo. De entre os atributos da dimensão, existe um conjunto dedicado à demografia da região do cliente, o qual só será atualizado em 2020.
   1. Descreva como seria registada uma atualização do número de filhos de um cliente, incluindo o que seria guardado nos atributos demográficos, e cobrindo todos os atributos cujos valores seriam alterados.

Utilizando a técnica 2, seria introduzida uma nova linha na tabela da dimensão Cliente com o novo numero de filhos, com uma nova chave primaria, mas mantendo a chave supernatural da linha original e todos os outros atributos.

* 1. Explique como poderia ser controlado o crescimento da dimensão Cliente, particularmente no que diz respeito ao conjunto de atributos demográficos. Nota: pode fazer um diagrama, se for adequado.

Para melhor controlar o crescimento da dimensão Cliente pode ser criada uma mini dimensão para guardar os atributos demográficos, no entanto estes devem ser convertidos para intervalos de valores (ex. faixa etária, intervalo de rendimentos). O histórico de alterações destes valores é guardado na tabela de factos (tipo 4)

1. **2016.04** Indique afirmações verdadeiras (V) ou falsas (F) sobre tabelas de ponte.

V\_\_ Permitem navegar em hierarquias de profundidade variável.

F\_\_ Servem para ligar duas tabelas de factos e permitir relatórios transdepartamentais.

F\_\_ Cada linha da tabela inclui apontadores para os níveis ascendente e descendente.

V\_\_ Guardam caminhos entre todos os valores de todos os níveis de uma hierarquia.

F\_\_ Costumam ter menos linhas do que as tabelas de dimensões que lhes servem de base.

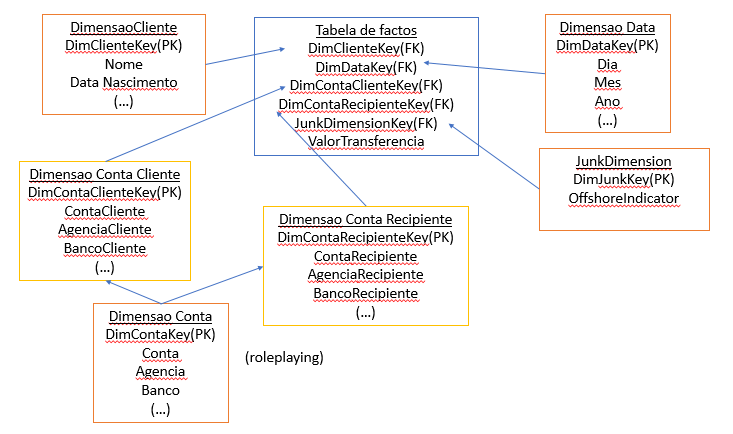
1. **2016.04** Para o facto seguinte, indique as medidas numéricas (se existirem), as dimensões, se há *role playing*, alguns atributos representativos, incluindo chaves substitutas e estrangeiras, e o tipo de tabela de factos: a pessoa P transferiu E euros na data D da conta C da agência A do banco B para a conta D da agência O, sinalizada como estando em local *offshore*, do banco F. Nota: pode desenhar um diagrama de dados.

Medidas numéricas - ValorTransferencia

Dimensões – Data, Cliente, Conta, Junk

Roleplaying - Conta

Tipo de tabela de factos – transação



1. **2015.06** Indique afirmações verdadeiras (V) ou falsas (F) sobre modelação dimensional.

V\_\_ As dimensões e as medidas costumam ser substantivos e os factos costumam ser verbos.

F\_\_ Numa tabela de factos apenas as medidas de negócio podem ser atributos numéricos.

V\_\_ Numa mesma tabela de dimensão podem coexistir várias hierarquias de atributos.

F\_\_ As tabelas de factos costumam ter muitas colunas e as de dimensões tendem a ser estreitas.

F\_\_ Uma minidimensão está ligada diretamente à dimensão "monstra" da qual derivou.

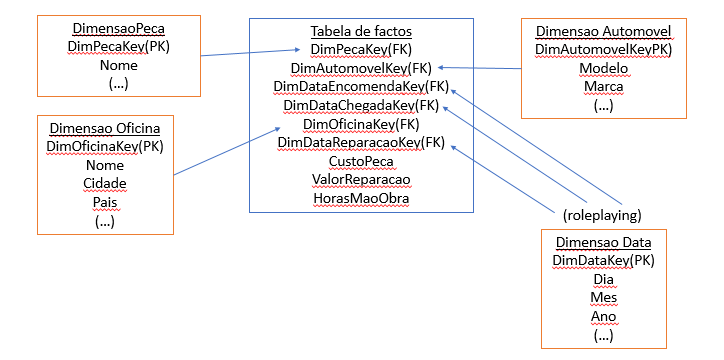
1. **2015.06** Para o facto seguinte, indique o tipo de tabela de factos, as medidas numéricas (caso existam), as dimensões, se há *role playing*, alguns atributos representativos, incluindo chaves substitutas e estrangeiras, e uma estimativa do número de linhas em cada tabela: a peça P, com um custo de E1 euros, do automóvel A, foi encomendada no dia D1 e chegou no dia D2 à oficina O, tendo sido aí feita uma reparação concluída no dia D3 no valor de E2 euros para H horas de mão-de-obra. Notas: pode desenhar um diagrama de dados, e considere que se trata de uma empresa de aluguer com centenas de automóveis de apenas três modelos.

Medidas numéricas – CustoPeca, ValorReparacao, HorasMaoObra

Dimensões – Data, Peca, Oficina, Automovel

Roleplaying – dimensão data

Tipo de tabela de factos – snapshot acumulativa



1. **2014.06** Compare as bifurcações e minidimensões em termos de: a) frequência de atualização dos atributos; b) relação entre os atributos; e c) modo de juntar os dados com os da dimensão “monstra” correspondente.

|  |  |  |
| --- | --- | --- |
|  | Bifurcacoes / Outriggers | Mini dimensões |
| a) | Atributos mudam lentamente | Atributos mudam com alguma frequência |
| b) | Atributos relacionados entre si, podem conter hierarquias | Atributos não necessitam de estar relacionados entre si |
| c) | Snowflake de uma dimensão monstra, transfere atributos que mudam pouco para outra tabela. Chave estrangeira para o outrigger na dimensão monstra. | Guarda as combinações possíveis de valores dos atributos na mini dimensão, dados são juntos com a dimensão monstra através de uma chave para a mini dimensão na tabela de factos (tipo 4) ou diretamente na tabela da dimensão monstra (tipo 5) |

1. **2014.06** Compare os três tipos de tabelas de factos em função do grão e dê um exemplo plausível de cada tipo.

Tabela de factos de transações: cada evento é armazenado na tabela de factos apenas uma vez, possui uma coluna de data indicando quando o evento ocorreu e uma coluna que identifica cada evento. O número de linhas é igual ao da tabela de origem. Uma linha numa tabela de factos da transação corresponde a um evento no espaço e no tempo

Um exemplo poderá ser quando existe uma tabela de vendas com pedidos de clientes. Grão: Uma linha por transação.

Tabela de factos instantâneos periódicos: uma linha numa tabela de factos de snapshot periódico resume muitos eventos de medição que ocorrem durante um período padrão, como um dia, uma semana ou um mês. Todo o sistema de origem é copiado na tabela de fatos regularmente, sendo que o mesmo evento pode ser armazenado várias vezes. Possui uma coluna de data do snapshot, indicando quando uma cópia da tabela de origem foi criada. O grão é o período, não a transação individual. Um exemplo deste tipo de tabela de factos poderá ser o saldo da conta bancária. No final de cada dia, os saldos de cada conta de cliente no banco são armazenados nesta tabela de saldo da conta. Grão: uma linha por saldo bancário por mês (período).

Tabela de factos instantâneos acumulativos: resume os eventos que ocorrem em etapas previsíveis entre o início e o final de um processo. Há uma chave estrangeira da dimensão data na tabela de fatos para cada evento crítico no processo. Além das datas de chaves estrangeiras associadas a cada etapa crítica do processo, as tabelas de fatos de snapshot acumulativo contêm chaves estrangeiras para outras dimensões e, opcionalmente, dimensões degeneradas. A atualização das linhas de fatos é exclusiva entre os três tipos de tabelas de fatos. Exemplo: Numa compra numa loja online, o cliente é de Lisboa e o artigo estava num armazém no Porto. Terá de existir a data do pedido, a data de saída do armazém do Porto, a data de chegada ao armazém de Lisboa, a data de saída do armazém de Lisboa e a data de entrega. Grão: uma atualização na tabela de fatos sempre que algo nesse processo de entrega tenha mudado (ciclo de vida).

1. **2014.06** Justifique se, numa tabela de factos de tipo transacional sobre encomendas, usaria ou não como medidas o preço unitário de cada produto e a percentagem de desconto aplicada ao preço base do produto para cada cliente que fez uma encomenda. Em caso negativo, indique que medidas usaria em alternativa.

A aplicação de percentagens de desconto face a preços base ou preços unitários é mais típica de tabelas de factos sem factos. A percentagem não pode ser considerada como medida na transação, logo não a usaria numa tabela de factos desse tipo. As medidas que utilizaria seriam o preço total, o desconto total e o numero de unidades do produto.

1. **2014.06** Explique a diferença entre medidas aditivas e semiaditivas, dando um exemplo de cada, e justifique qual dos dois tipos é preferível em *data warehouses*. Nota: tenha especialmente em conta o ponto de vista do decisor que está a compor relatórios.

As medidas aditivas permitem a soma de valores ao longo de todas as dimensões, sendo esta uma das operações mais úteis num data warehouse, enquanto que as medidas semi aditivas dizem respeito a factos de natureza periódica e onde as medidas são aditivas mas não em todas as dimensões. Normalmente as medidas que surgem na tabela de factos, do ponto de vista do decisor, são medidas somáveis, logo as medidas aditivas são mais utilizadas.

1. **2014.06** Apresente um exemplo de duas hierarquias para atributos muito relacionados entre si e pertencentes a uma mesma dimensão, e justifique se pode ou não ser útil um decisor que está a compor relatórios ter acesso a qualquer uma dessas hierarquias (ou, em geral, a hierarquias em condições semelhantes).

Tipicamente a dimensão data apresenta atributos muito relacionados entre si, por exemplo: Ano-Semestre-Trimestre-Mês-Dia e Ano fiscal-Semestre fiscal-Trimestre fiscal-Dia.

Para o decisor pode ser importante ter acesso a todas as hierarquias e até efetuar algumas operações ao cubo de dados em condições semelhantes, pois estas hierarquias dizem respeito, no fundo, ao mesmo (datas), o que se altera são os valores que compõem os atributos. Por exemplo, em 2020 as receitas fiscais baixaram porque no 2º trimestre o país entrou em estado de emergência devido à COVID-19.

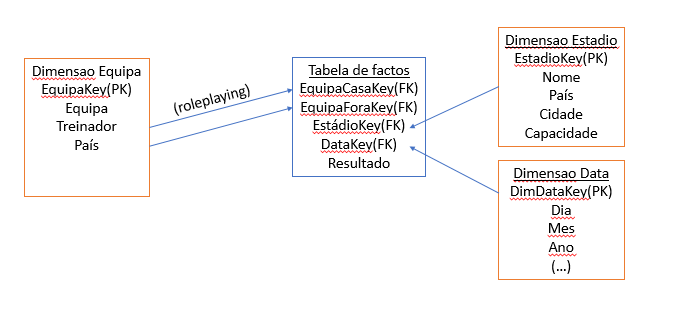
1. **2014.06** Para o facto seguinte, indique as medidas numéricas, caso existam, o tipo de tabela de factos, as dimensões, e alguns atributos representativos, incluindo chaves substitutas e estrangeiras: a equipa do país P1 treinada por T1 jogou com a equipa do país P2 treinada por T2 no dia D, no estádio E com capacidade para 55000 espetadores que fica na cidade C do país P3, tendo o resultado final sido 4 a 0.

Medidas numéricas – Resultado

Dimensões – Equipa, Data, Estádio

Roleplaying – Dim Data

Tipo de tabela de factos – transação



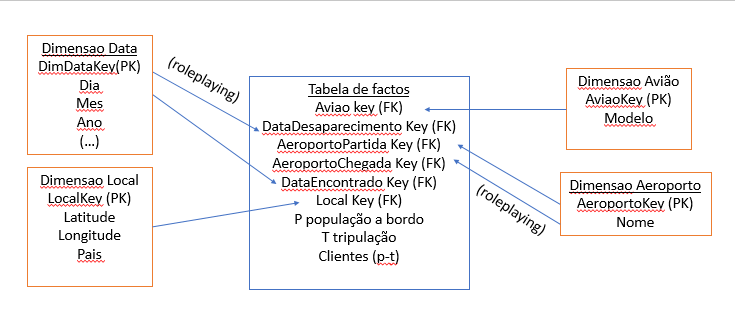
1. **2014.04** Para o facto seguinte, indique as medidas numéricas, caso existam, o tipo de tabela de factos, as dimensões, e alguns atributos representativos: o avião V com P pessoas a bordo, das quais T pertencentes à tripulação, partiu do aeroporto A1 e tinha destino previsto para o aeroporto A2, mas desapareceu dos radares no dia D1, tendo sido descoberto no local L, no dia D2. Nota: pode desenhar um diagrama de dados.

Medidas numéricas –população, tripulação, clientes

Dimensões – data, local, avião, aeroporto

Roleplaying – dimensões data e aeroporto

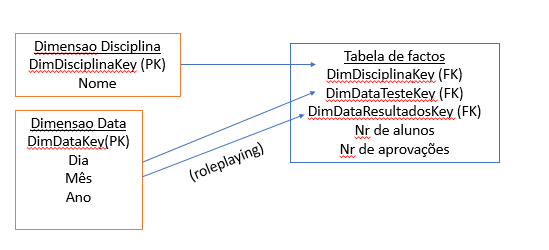
Tipo de tabela de factos – snapshot acumulativa



1. **2013.06** Considere a dimensão Cliente, com dados desde 2010, para já com apenas mil registos, na qual está a ser usada a técnica 2 para guardar mudanças lentas que vão ocorrendo. De entre os atributos da dimensão, existe um conjunto dedicado à demografia da região do cliente, o qual só será atualizado em 2020.
   1. Assuma que cada registo ocupa em média 2000 bytes dos quais 1000 bytes são para os valores dos atributos demográficos, que chaves primárias e estrangeiras ocupam 4 bytes, e ainda que existem 5 regiões diferentes. Calcule o espaço total ocupado pela dimensão e compare com o que seria necessário caso fosse utilizada uma bifurcação para controlar o crescimento da dimensão Cliente.
   2. Sabendo que as bifurcações, quando bem aplicadas, permitem poupanças significativas de espaço em disco, justifique se devem ou não ser usadas intensivamente no modelo de dados de um *data warehouse*.
2. **2012.04** Considerando os factos seguintes, indique: as medidas numéricas, caso existam; o tipo de cada tabela de factos; as dimensões e alguns atributos representativos; e se as dimensões podem participar em ambas as tabelas de factos. Nota: pode desenhar diagramas de dados.
   1. 48 alunos realizaram o teste de IPAI no dia 12 de abril de 2011, tendo as notas sido divulgadas em 3 de maio de 2011, e havendo obtido aprovação 37 alunos.

Medidas numéricas –numero de alunos, numero de aprovacoes

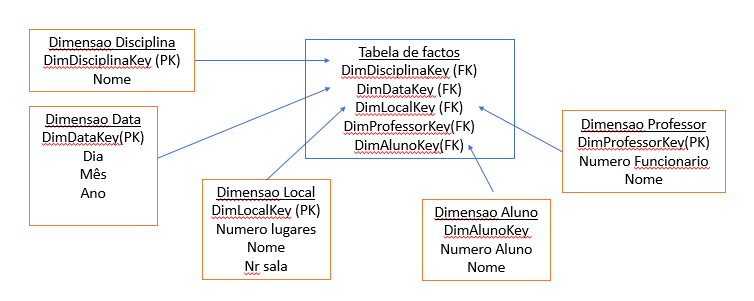
Dimensões – data, disciplina



* 1. O aluno Pedro, número 43210, fez o teste de IPAI de 17 de abril de 2012, ocorrido no anfiteatro 1.3.20, com 74 lugares, com vigilância do professor António, funcionário 1234.

Medidas numéricas –não existem – tabela de factos sem factos

Dimensões – data, local, disciplina, professor, aluno



As dimensões data, disciplina são partilhadas entre as duas tabelas de factos.

1. **2012.04** Apresente uma vantagem e uma desvantagem das técnicas de tipo 1 e 2 para registo de mudanças lentas numa dimensão, e descreva um cenário adequado a cada técnica.

|  |  |  |
| --- | --- | --- |
|  | Tipo 1 | Tipo 2 |
| Vantagens | Fácil e simples de perceber | Histórico permite perceber o impacto das decisões |
| Desvantagens | Não considera o histórico | Acelera o crescimento da tabela de dimensões |
| Cenarios | Mudanças de produto no departamento, apenas salva o departamento para onde foi mudado | Se o produto mudar de departamento, o tipo 2 adiciona linhas novas para guardar as alterações, sendo necessária uma chave supernatural |

1. **2011.06** Justifique se a matriz de processos *(bus matrix)* deve ser definida antes ou depois da modelação detalhada das dimensões, e indique o papel da matriz na conformação de dimensões.

Aquando da implementação da bus matrix, existe já a ideia das dimensões comuns mas não estão definidas as modelações detalhadas das mesmas, porque a bus matrix cruza dimensões dos dados com os processos de negócio. Já as dimensões são conformadas quando partilham atributos, desta forma a bus matrix facilita a visualização das dimensões conformadas.

1. **2011.06** Considerando os factos seguintes, indique os nomes e atributos das dimensões envolvidas bem como das medidas numéricas (caso existam), e mencione o tipo de tabela de factos.
   1. 113 pessoas candidataram-se em 2011 a cursos da FCUL através do programa Maiores de 23, tendo 12 candidatos obtido aprovação na prova realizada a 7 de Maio cujos resultados foram publicados a 31 de Maio.
   2. A fatura mensal de eletricidade do cliente C no mês M foi de E euros, dos quais T resultam do pagamento de uma taxa de audiovisual.
2. **2011.06** Considerando os factos seguintes, indique os nomes e eventuais atributos das dimensões envolvidas bem como das medidas numéricas (caso existam), e mencione também o tipo de tabela de factos.
   1. 17 partidos concorreram às eleições legislativas antecipadas de 5 de junho de 2011 em Portugal, dos quais 5 elegeram deputados.
   2. Nas legislativas de 5 de junho de 2011 o partido P elegeu o deputado N pelo distrito D.
3. **2011.06** Apresente um exemplo de *role playing* de dimensão e indique uma forma de o realizar usando conceitos de SQL, com o mínimo de recursos e de forma transparente para o utilizador.

O role-playing dá-se quando por exemplo numa dimensão data existem vários eventos com a mesma data, embora cada data tenha um significado diferente em cada dimensão/evento. Neste exemplo, as dimensões data são simuladas com vistas ou sinónimos de SQL.

1. **2011.06** Quanto mais fino o grão dos factos mais registos tendem a existir na tabela de factos. Por exemplo, basta um registo para guardar o total de uma fatura, mas são precisos vários para guardar os subtotais por produto. Nesta linha, justifique se a afirmação seguinte é ou não correta: “em geral, quanto mais fino o grão dos factos mais dimensões tendem a ser modeladas.”

A afirmação presente no enunciado é verdadeira, porque um grão fino possui maior nível de detalhe e logicamente mais dimensões a serem modeladas e mais atributos na chave primária.

1. **2011.06** Suponha uma tabela de factos de tipo transacional sobre vendas de produtos, com medidas PreçoUnitário e QuantidadeVendida. Justifique se ambas as medidas são aditivas e em caso negativo proponha uma alternativa em que tal aconteça sem haver perda de informação.
2. **2011.04** Considerando os dois factos seguintes, indique os atributos e nomes das dimensões envolvidas bem como das medidas numéricas (caso existam), e mencione também o tipo de tabela de factos em causa.
   1. 27 alunos inscreveram-se no teste de IPAI do dia 20 de Abril de 2010, tendo as notas sido publicadas em 22 de Abril de 2010, e tendo obtido aprovação 23 alunos.
   2. O aluno Pedro, número 43210, fez o teste de IPAI de 12 de Abril de 2011, ocorrido no anfiteatro 8.2.47, com 220 lugares, com vigilância do professor António, funcionário 1234.
3. **2010.07** Descreva um exemplo de tabela de factos de tipo instantâneo cumulativo, tendo o cuidado de mencionar se cada facto abrange períodos variáveis ou fixos de tempo e se o carregamento de novos dados envolve só inserções, só atualizações, ou se permite ambas as operações de escrita.

As tabelas de factos de tipo instantâneo cumulativo acompanham um processo recorrente, mas de duração variável, com etapas previsíveis, cumpridas em datas imprevisíveis. Dados novos através de inserções e atualizações de linhas.

1. **2010.07** Compare tabelas de factos de tipo instantâneo periódico e cumulativo segundo os dois critérios seguintes: a) períodos variáveis ou fixos representados em cada facto; e b) uso ou não de inserções/atualizações aquando do carregamento de dados.

a) Instantâneo periódico: Períodos fixos

Instantâneo cumulativo: Períodos variáveis

b) Instantâneo periódico: Inserções

Instantâneo cumulativo: Inserções e atualizações

1. **2010.06** Descreva os quatro passos recomendados para a modelação dimensional, exemplificando com o caso do projeto da disciplina. Nota: os passos referem dimensões, grão dos factos, processos de negócio, e medidas, não necessariamente por esta ordem.

1. Identificar processos a modelar: Matriz exequibilidade/valor salienta processos valiosos e com dados

2. Determinar grão da tabela de factos: Nível de detalhe máximo desejado para as análises

3. Modelar dimensões em tabelas: Dados que descrevem os processos da organização

4. Identificar medidas numéricas na tabela de factos: Medidas que servem para avaliar o negócio

1. **2009.07** Considere que foi feito um estudo sobre a aquisição de medicamentos em farmácias portuguesas. Para cada aquisição é registado o medicamento adquirido, o valor pago, o dia e a hora bem como informação sobre o cliente (nome, número de beneficiário, e residência). Para certos medicamentos passados com receita conhece-se ainda o médico envolvido e o valor da comparticipação. Os medicamentos têm determinados compostos ativos em quantidades diferentes dependendo da marca. Nota: a informação do cliente poderá não estar sempre disponível. Elabore o esquema para um *data warehouse* que permita recolher e analisar esta informação.
2. **2009.07** Bifurcações e minidimensões são estratégias diferentes de decomposição usadas para modelar tabelas de dimensão “monstras” (com muitas linhas e muitas colunas). Descreva em que consistem e ilustre as suas diferenças.

Nas bifurcações, os atributos devem mudar lentamente, os atributos estão relacionados entre si e é feito um snowflaking de uma dimensão “monstra” que transfere atributos que mudam pouco para outra tabela;

Nas mini-dimensões, os atributos devem mudar com frequência, não necessitando de estar relacionados entre si. Guardam as combinações possíveis de valores dos atributos através de tabela de factos com chave estrangeira adicional para mini-dimensão.

1. **2009.06** Elabore um esquema para um *data warehouse* que permita recolher e analisar a informação na seguinte situação, justificando as decisões tomadas.

Foi feito um estudo sobre a utilização de transportes públicos em várias cidades europeias. Para tal fezse o seguimento de um conjunto de passageiros, através dos registos de entrada e saída dos seus títulos de transporte para cada viagem encetada. Conhece-se a identidade dos passageiros seguidos bem como as suas caraterísticas socioeconómicas. Está ainda disponível informação sobre os veículos, as rotas seguidas, o número de paragens em cada viagem e os condutores.

1. **2009.04** No contexto de um *data warehouse* referente a infrações de trânsito, a GNR necessitou da criação de uma dimensão Condutor.
   1. Indique que atributos consideraria para a modelação dessa dimensão e justifique que tamanho teria ao fim de três anos sabendo que em média são autuados cerca de 10.000 novos condutores por mês (valores fictícios).
   2. Para o mesmo trabalho a GNR partiu de um conjunto de folhas Excel existentes em cada esquadra, que tinham a seguinte estrutura: Audiências, NomeDoCondutor, FaixaEtária, Veículo, EstadoDoVeículo, CartaDeCondução, ÁlcoolNoSangue, TipoDeInfracção, Hora, Data, Local, AgentesEmOperação, ValorDaMulta. Desenhe o diagrama de um *data warehouse* para este problema, expandindo, se necessário, os atributos das dimensões consideradas. Nota: pode referenciar a resposta da pergunta anterior.
2. **2008.09** Indique como faria para modelar uma hierarquia de profundidade fixa numa tabela de dimensão. Descreva também as diferenças caso a hierarquia fosse de profundidade variável. Nota: pode fazer uma ilustração se achar necessário.

Numa hierarquia de profundidade fixa, todos os níveis da hierarquia têm sempre valores. Estas hierarquias são de uso generalizado e fáceis de entender. As principais diferenças residem no facto de as hierarquias de profundidade variável possuirem alguns níveis que podem não estar preenchidos e servem para relações hierárquicas mais complexas.

1. **2008.09** Elabore um esquema para um *data warehouse* que permita recolher e analisar a informação na seguinte situação, justificando as decisões tomadas.

Uma empresa de sondagens faz entrevistas telefónicas recolhendo preferências de consumidores sobre refrigerantes e cervejas. Dentro das perguntas efetuadas, pergunta-se qual o grau de preferência por uma série de marcas, e até que ponto se recorda de ter visto propaganda a elas referida. A propaganda pode ser de rua, televisão, ou imprensa escrita. É recolhida também informação socioeconómica e geográfica sobre o cliente, sabendo-se ainda o momento em que a entrevista ocorreu e qual o entrevistador.

1. **2008.05** Indique três caraterísticas das tabelas de factos do tipo instantâneo periódico *(periodic snapshot)*, ilustrando com um exemplo.

Guardam desempenho acumulado em períodos fixos e regulares

Dados novos através de inserções

Algumas medidas podem valer zero

Ex: Saldo, comissões e juros em janeiro, fevereiro, março, …

1. **2008.04** Indique que estratégia utilizaria para modelar uma dimensão de clientes empresariais, que contém um conjunto de atributos qualitativos que muda com muita frequência. Para o problema concreto suponha que a dimensão tem 100.000 empresas distintas, com cerca de 120 atributos, dos quais 20, sofrem alterações todas as semanas.
2. **2008.04** No contexto de um *data warehouse* defina sucintamente os seguintes conceitos: a) grão; b) tabelas de factos sem factos; c) dimensões degeneradas.

a) O grão é o significado de uma linha na tabela de fatos e determina o nível máximo de detalhe;

b) As tabelas de factos sem factos guardam factos sem medidas de negócio e servem para simples acompanhamento de eventos;

c) As dimensões degeneradas dizem respeito a dimensões que servem apenas para agrupar factos. O agrupamento tem de ter significado forte no contexto. Não possui outros atributos próprios e não tem tabela de dimensão associada. As dimensões degeneradas podem ser descartadas, se não fizerem falta à análise ou substituídas por chaves estrangeiras para dimensões