



Curso de Engenharia de Sistemas e Informática - 5º Ano

Ficha T. Prática n.º 1

Objectivo:

Estudo do paradigma multidimensional com introdução de uma extensão ao diagrama E/R para modelação multidimensional ME/R.

Introdução

OLAP surge como uma tendência em tecnologias de bases de dados, baseadas em vistas multidimensionais de dados.

Uma boa definição de OLAP pode tirar-se de [OLAP97]: “... OLAP (*On-Line Analytical Processing*) é uma categoria de tecnologias de software que permite aos analistas, gestores e executivos a análise de dados, através de um acesso interactivo, consistente e rápido, de uma grande variedade de possíveis vistas, a informação que foi transformada a partir de dados base, por forma a reflectir a dimensionalidade real da empresa da forma como é percebida pelo utilizador. As funcionalidade OLAP são caracterizadas por uma análise dinâmica multidimensional dos dados consolidados da empresa que suportam as actividades de análise e navegação incluindo cálculos e modelação aplicados entre dimensões, atravessando hierarquias e /ou entre membros, análise de tendências, sequências temporais, visualização de subconjuntos através de slicing, perfurar os dados para níveis mais profundos de consolidação, rotação de dados para novas comparações, etc...”.

As ferramentas OLAP são utilizadas muito frequentemente como ferramentas de *front end* em ambientes de *data warehousing* [SBH99]. Permitem uma análise interactiva de dados multidimensionais (MD). Desta forma a tecnologia de bases de dados multidimensionais tem-se tornado progressivamente mais importante e independentemente das arquitecturas que as implementam em termos de armazenamento, o paradigma multidimensional através do qual os dados são apresentados ao utilizador é central para toda esta temática. Vai-se assim, aqui, tratar deste aspecto fulcral e no final, introduzir-se-á uma técnica de modelação, sob a forma de uma extensão ao clássico diagrama E/R, decerto do conhecimento de todos.

Modelação Multidimensional

A modelação multidimensional de dados desempenha um papel chave durante a concepção de um data warehouse. O esquema do warehouse multidimensional oferece

uma visão integrada das fontes de dados operacionais e, assim, serve de núcleo ao data warehouse, devendo ser dada muita atenção ao desenvolvimento do seu esquema.

O processo de concepção do esquema é mostrado na figura 1.

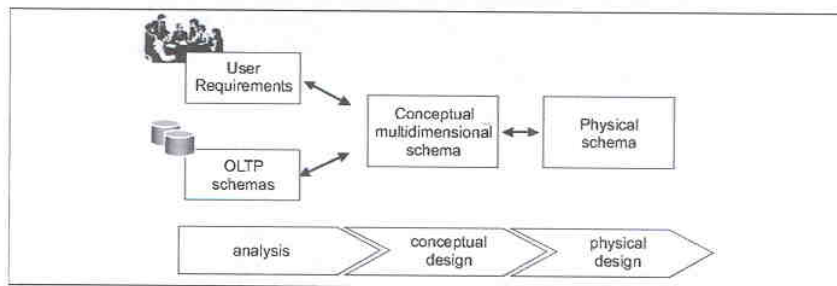


Fig. 1 Processo de concepção do esquema em ambientes de Data Warehousing.

A maioria dos projectos é influenciada pelos requisitos dos utilizadores e disponibilidade dos dados nos sistemas operacionais, sendo seguida uma abordagem evolucionária: começa-se com um protótipo que proporciona algumas funcionalidades e um dado conjunto de dados. Este protótipo é depois adaptado de acordo com as alterações e crescentes requisitos conhecidos através do feedback dos utilizadores. Uma das características do esquema concebido deverá ser a flexibilidade para assegurar a sua reutilização.

Paradigma Multidimensional

Para iniciarmos a abordagem a esta temática, comecemos com um exemplo.

Um construtor de automóveis pode pretender analisar as reparações de veículos para melhorar o seu produto, definir novas políticas de garantias e obter informação acerca da certificação de oficinas de manutenção. O construtor pretende analisar as reparações dos veículos (facto em análise) segundo um conjunto de perspectivas (dimensões). Esta metáfora é representada por um cubo (ou multicubo – para n dimensões), utilizada para representar a visão dos dados e é mostrada na figura 2.

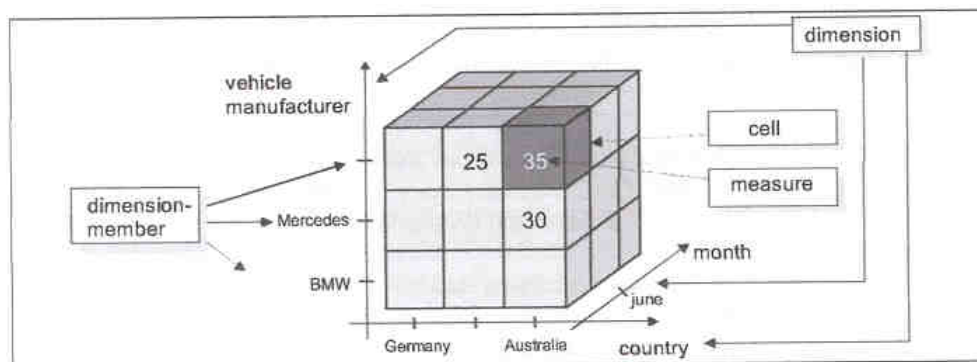


Fig. 2 – Visualização dum esquema multidimensional utilizando a metáfora do cubo.

No cubo temos o assunto em análise chamado facto (no exemplo, a reparação de veículos). As células do cubo de dados contêm as medidas (também chamadas de dados quantificáveis – pois que maioritariamente numéricos) que descrevem o facto (p. ex. custos e duração da reparação dos veículos).

Os eixos do cubo (chamados de dimensões ou dados qualificadores) representam diversas formas de analisar os dados (ex. veículo e data de reparação).

Esta visão dos dados é semelhante à noção de arrays. Mas há uma grande diferença: estes estão estruturados apenas segundo uma ordenação linear definida pelos índices. Em aplicações OLAP isto não é suficiente, pois que do ponto de vista do utilizador, os elementos de uma dimensão (denominados membros) não estão, em regra, ordenados linearmente (p. ex. garagens), excepto em algumas situações como é o caso da dimensão tempo.

A estruturação das dimensões é efectuada através de hierarquias contendo um conjunto distinto de membros (Ver fig. 3 – à esquerda). Níveis diferentes correspondem a distintos níveis de granularidade e formas de classificação. O nível A *rolls up* para o nível B se uma classificação dos elementos de A de acordo com os elementos de B for semanticamente significativa para a aplicação (ex. o nível “dia” *rolls up* para “mês”).

Um nível pode fazer *roll up* para qualquer número de níveis, permitindo assim existir múltiplas hierarquias numa dada dimensão. Isto será possível se existirem vários critérios de classificação para membros de uma dimensão (ex. garagens podem ser classificadas pela sua localização geográfica ou pelo seu tipo). Outra situação ocorre quando vários caminhos de *roll-up* são possíveis entre dois níveis – estamos em presença de caminhos alternativos (ver. fig. 3 – à direita).

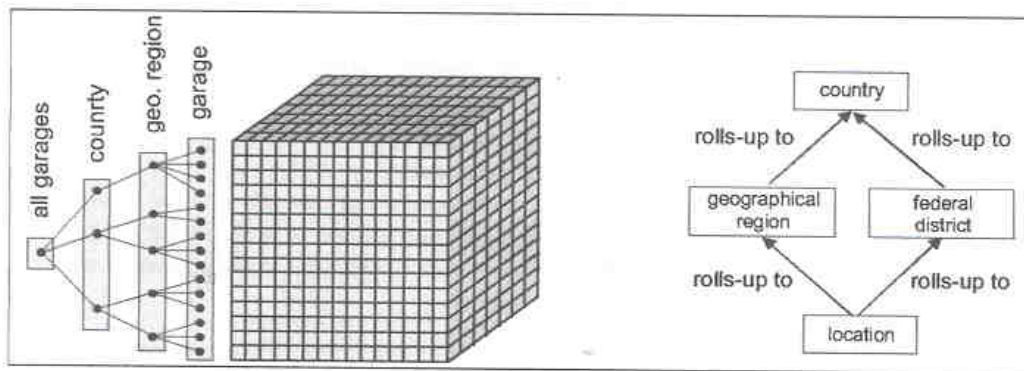


Fig. 3 - Estrutura da hierarquia de níveis de dimensões (à esquerda) e caminhos alternativos dentro de uma dimensão (à direita).

A estruturação das dimensões do ponto de vista do utilizador é completada com a utilização de atributos de nível de dimensão – descrevem os membros de um dado nível de dimensão, mas não definem hierarquias (ex. o nome e morada de um cliente ou nome de um gestor de uma região).

Um esquema complexo pode conter mais do que um cubo. Isso torna-se necessário se uma aplicação requer a análise de diferentes factos (ex. vendas de veículos e reparações). Cada cubo pode ter granularidades diferentes.

Modelação Multidimensional

No artigo [SBHD99], utilizado como base principal para fazer esta ficha, os autores propõem uma extensão ao modelo E/R que permita utilizá-lo para a modelação de dados multidimensionais, que denominaram de Modelo Multidimensional E/R (ME/R). Introduzem 3 novos elementos gráficos de notação apresentados na figura 4.

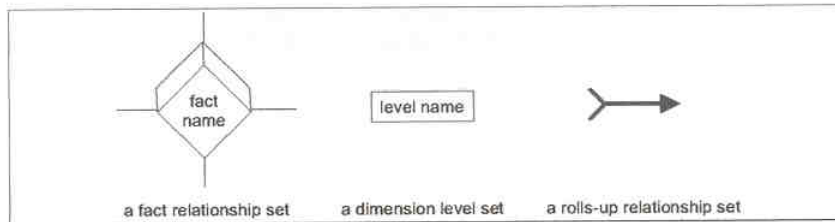


Fig. 4 – Notação gráfica de elementos dos diagramas ME/R.

Com este modelo poderem-se modelar os modelos multidimensionais de dados como se apresenta na fig. 5, incluindo também especializações como se apresenta na fig. 6 do artigo, ou ainda, representando cubos múltiplos, como é mostrado na fig. 7.

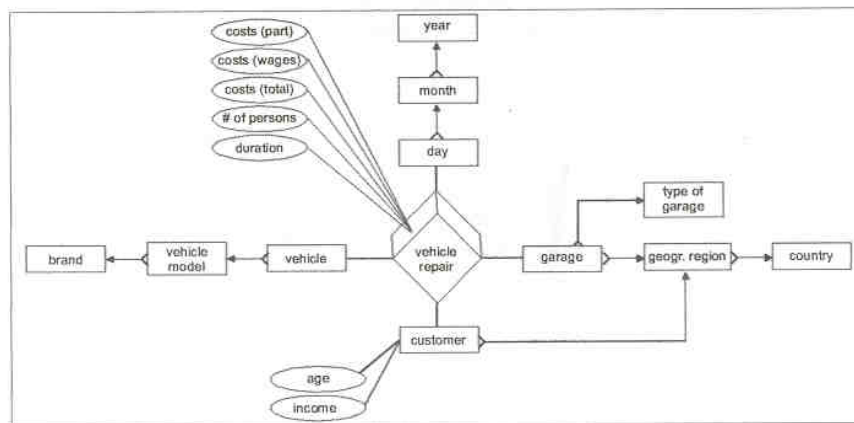


Fig. 5 – Um diagrama ME/R para análise de reparação de veículos.

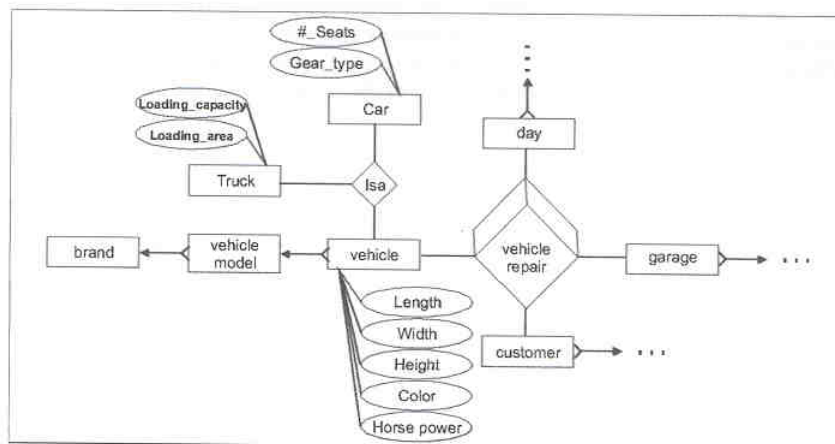


Fig. 6 – Combinação de notação ME/R com elementos construtivos clássicos em diagramas E/R.

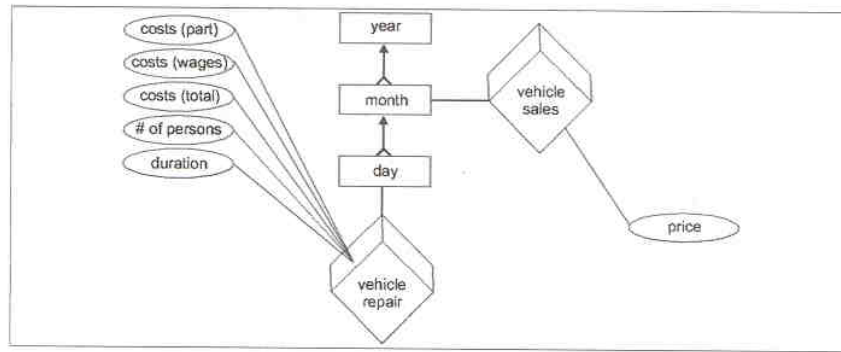


Fig. 7 – Cubos múltiplos partilhando uma dimensão em diversos níveis.

Referências e Fontes

- [OLAP97] OLAP Council. OLAP and OLAP Server Definitions. 1997
- [SBHD99] Carsten Sapia, Markus Blaschca, Gabriele Hofling, Barbara Dinter, “Extending the E/R Model for the Multidimensional paradigm”, in K.Kambayashi et. Al. “Advances in Database Technologies”, LNCS Vol. 1552, Springer-Verlag, 1999
- [SBH99] Carsten Sapia, Markus Blaschca, Gabriele Hofling “An Overview of Multidimensional Data Models”, FORWISS Technical Report FR-1999-001, 1999, disponível em <http://citeseer.ist.psu.edu/sapia99overview.html>.

Questões

1. No início é transcrito um excerto da definição dada pelo OLAP Council de OLAP. Caracterize e explique a definição, efectuando alguns comentários que lhe pareçam pertinentes, nomeadamente quanto à forma nativa de ver os dados pelo analista e à natureza do próprio modelo implícito.
2. Ao falar-se em modelação multidimensional diz-se que ela é central para o projecto dum Data Warehouse. Comente a afirmação.
3. Na fig. 1 mostra-se o processo de concepção do DW. Mostre em que medida a abordagem incremental está implícita ou como alteraria o esquema para o fazer.
4. O esquema multidimensional (mostrado na fig. 2) se parece igual à noção de arrays, é realmente algo diferente. Mostre em que medida.
5. O mapeamento do modelo multidimensional para o esquema físico pode realizar-se numa perspectiva relacional ou multidimensional. Surge assim o MOLAP e ROLAP. A priori e sem maiores conhecimentos, qual acha fornecerá um maior suporte às características base do paradigma multidimensional? Justifique.
6. Já no final, o autor introduz o modelo ME/R, uma extensão ao clássico diagrama E/R, mas aqui adaptado à modelação multidimensional. Pela análise das diversas figuras em que são mostrados modelos de DW, mostre alguns dos aspectos mais relevantes do modelo.
7. Nesse modelo são suportadas hierarquias múltiplas em dimensões? Mostre como indicando um exemplo retirado de um dos esquemas.