

UNINDO O MODELO ENTIDADE / RELACIONAMENTO AO MODELO DIMENSIONAL COMO FORMA DE EXTRAIR INFORMAÇÕES GERENCIAIS DIRETAMENTE DA BASE DE DADOS OPERACIONAL*

João Alexandre Bonin de Mello**
Aran B. T. Morales***
José Leomar Todesco****

RESUMO: Este artigo propõe a utilização de um modelo híbrido, unindo o modelo entidade/relacionamento ao modelo dimensional, a fim de possibilitar que empresas de menor porte, que não dispõem de recursos para investimento em um *data warehouse* ou de *data-marts* separados do banco de dados transacional, possam utilizar-se de sistemas de informação executiva e de sistemas de apoio à decisão servindo-se do próprio banco de dados transacional. Primeiramente, o artigo classifica os sistemas de informação de acordo com sua finalidade no ambiente organizacional. Em seguida, discute o modelo entidade/relacionamento e sua importância para os sistemas transacionais e o modelo dimensional como suporte informacional. Finalmente, propõe um modelo híbrido, que é composto de uma tabela de fatos central e de várias tabelas dimensionais normalizadas e relacionadas entre si a fim de suportar a base de dados transacional.

PALAVRAS-CHAVE: *data Warehouse*; sistemas de informação; modelo entidade/relacionamento; modelo multidimensional.

UNITE THE ENTITY MODEL/RELATION TO DIMENSIONAL MODEL AS FORM TO EXTRACT MANAGERICAL INFORMATIONS DIRECTLY OF BASE OF OPERATIONAL DATA

ABSTRACT: This article propose the utilization of hybrid model, uniting the entity model/relation to dimensional model, with objective to make possible little enterprise that don't have resources to investment in *data warehouse* or *data-marts* separate of transitional database can use systems of information executive of systems of support to decision use the transitional database. First, this article classifies the systems of information with its objective in organizational environment. Second, discuss the entity model/relation and its importance to the

* Trabalho apresentado no Programa de Pós-graduação em Engenharia de Produção em Nível de Mestrado da Universidade Federal de Santa Catarina – UFSC.

** Professor do Curso de Sistema de Informação da UNIPAR – Toledo – Paraná, Mestre em Engenharia de Produção – UFSC – Universidade Federal de Santa Catarina. Rua Leon Deniz, 902, Toledo – PR. 85.900.000. Alexandre@unipar.br

** Dr. em Engenharia de Produção – UFSC – Universidade Federal de Santa Catarina. aran@eps.ufsc.br

*** Dr. em Engenharia de Produção – UFSC – Universidade Federal de Santa Catarina.

transitional systems and the dimensional model as information support. Finally offer a hybrid model that have a table of central facts and many dimensional tables normalized and related itself with objective of support the base of transitional data.

KEY-WORDS: data warehouse; systems of information; entity model/relation; multidimensional model.

. Introdução

A utilização de grandes bases de dados em sistemas de apoio à decisão e em sistemas de informação executiva ganhou força com a utilização das tecnologias de *data warehouse* e *data-mart*, sendo que a abordagem comercial mais utilizada no momento, neste tipo de sistema, é a modelagem dimensional. Este tipo de tecnologia possibilitou grande flexibilidade, agilidade e consistência às informações de nível estratégico e tático.

Entretanto, a utilização de *data warehouses* dimensionais pressupõe a utilização de dois ambientes separados, um para o processamento transacional, baseado no modelo entidade/relacionamento e outro para o ambiente de *data warehouse*, baseada no modelo dimensional. Este tipo de arquitetura pressupõe grandes investimentos em *hardware*, *software* e pessoal, a fim de manter os dois ambientes em pleno funcionamento, o que força uma grande quantidade de empresas, de menor porte, a abrir mão de informações estratégicas que poderiam estar contribuindo com uma significativa melhoria no processo de gestão das mesmas.

Este trabalho propõe um modelo híbrido de modelagem de dados, apoiado tanto no modelo dimensional quanto no modelo entidade/relacionamento. Este modelo propõe-se a possibilitar que empresas que não possuam recursos para investimento em dois ambientes, utilizem-se da base de dados do sistema transacional para retirar suas informações de apoio à decisão com a mesma flexibilidade do modelo dimensional e, ainda, com um bom nível de performance.

. Classificação dos Sistemas de Informação

Os sistemas de informação podem ser classificados de acordo com sua finalidade. De acordo com STAIR (1998), SPRAGUE e WATSON (1991), estes podem ser classificados em sistemas de processamento transacional, sistemas de informações gerenciais, sistemas de apoio à decisão e sistemas especialistas.

. Sistemas de processamento transacional

Os sistemas de informação transacional (SIT), atualmente chamados de OLTP (*On-line transaction processing*) foram os primeiros aplicativos de computador a serem desenvolvidos na maioria das organizações. Eles têm a tarefa de processar as funções básicas e rotineiras de uma organização, tais como processamento da folha de pagamento, faturamento etc. (STAIR, 1998).

. Sistemas de informação gerencial

Um sistema de informações gerenciais (SIG) tem o propósito de fornecer informações necessárias à medição da eficiência operacional da organização, dando ênfase às necessidades gerenciais através de informações resumidas obtidas a partir da filtragem e análise de dados altamente detalhados, obtidos a partir das bases de dados dos sistemas de processamento transacionais e de fontes externas, (STAIR, 1998; SPRAGHE E WATSON, 1991).

. Sistemas de apoio à decisão

Um sistema de apoio à decisão (SAD) tem por finalidade dar embasamento à tomada de decisões referentes a problemas específicos. Este tipo de sistema é geralmente utilizado quando a tomada de decisão torna-se difícil sem o apoio de um sistema que forneça informações específicas que possam dar o devido suporte à tomada de decisão.

Um SAD deve reconhecer que existem estilos gerenciais diferentes e tipos de decisões diferentes. Isso implica uma grande participação do usuário do SAD no seu desenvolvimento, (STAIR, 1998; SPRAGHE e WATSON, 1991).

. Sistemas de Informação Executiva

Os sistemas de informação executiva (*executive information systems* – EIS), oferecem um meio prático para que executivos acessem informações estratégicas da empresa de um modo agrupado e simplificado de acordo com suas necessidades, ajustadas ao seu estilo de trabalho (FURLAN, IVO e AMARAL, 1994).

O sucesso de um EIS depende não só da disponibilidade das informações necessárias ao processo decisório, como também da forma de extração e exploração destas informações utilizada no momento da tomada de decisão, uma vez que diferentes tomadores de decisão necessitam de diferentes informações e possuem estilos decisórios diferentes, (FREITAS e POZZEBOM, 2000).

STAIR (1998) resume algumas características importantes para um Sistema de Informações Executivas. Estas características são: facilidade de uso, manipulação de dados externos e internos, quantitativos e qualitativos, execução de sofisticadas análises de dados, oferecendo um alto grau de especialização, fornecendo flexibilidade e dando suporte a todos os aspectos da tomada de decisões (uma vez que decisões que utilizam fontes externas de informações podem ter um alto grau de incerteza), além de oferecer recursos abrangentes de comunicação, uma vez que executivos necessitam de comunicação instantânea com o ambiente interno e externo da organização.

FREITAS e POZZEBOM (2000), defendem uma mudança no foco dos sistemas de informações executivas para sistemas de informação empresarial, onde os tomadores de decisão de todos os níveis possam se beneficiar da oferta de informações. VOLOLINO, WATSON e ROBINSON *apud* FREITAS e POZZEBON (2000), definem EIS como uma tecnologia de informação para todos

os usuários finais da organização, sendo os executivos um subconjunto destes usuários, em que o suporte é ajustado de acordo com as necessidades de cada classe de usuários.

. Sistemas OLTP e o Modelo Entidade/Relacionamento

Sistemas OLTP processam uma enorme quantidade de transações diariamente. Consistência transacional e velocidade são requisitos imperativos neste tipo de sistema. Eles operam as atividades essenciais de uma organização (KIMBALL, 1988).

O modelo entidade/relacionamento aplica-se perfeitamente a este tipo de sistema, pois divide os dados em várias entidades distintas. Técnicas de normalização eliminam completamente a redundância de dados, reduzindo significativamente a probabilidade de consistência e aumentando a velocidade de processamento transacional, uma vez que as atualizações do banco de dados a cada transação resumem-se a uma pequena quantidade de registros por vez (KIMBALL, 1998, KROENKE, 1999).

O modelo da figura 1, ilustra uma simplificação do diagrama de entidades/relacionamentos de um sistema de vendas para uma loja, onde os produtos são divididos em seções.

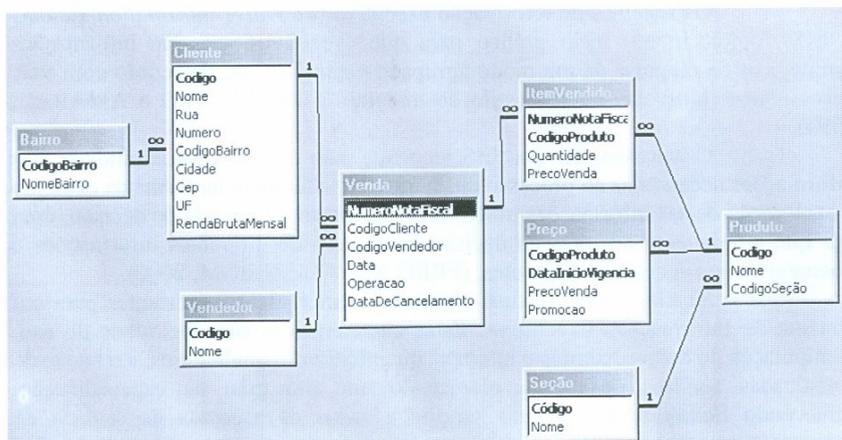


Figura 1 – Uma simplificação de um sistema de vendas implementado no modelo entidade/relacionamento

. Data Warehouse, Modelagem Multidimensional e Sistemas de Apoio à Decisão.

Até 1993, os fornecedores de RDBMS propunham um único sistema, baseado no modelo entidade-relacionamento para sistemas OLTP e SAD (KIMBALL, 1998). De fato, STAIR (1998), afirma que os sistemas informação

transacional são a base de dados para os sistemas de informação gerencial e sistemas de apoio à decisão.

Com o tempo, percebeu-se que o modelo entidade / relacionamento era ineficiente para sistemas de apoio à decisão. Esta ineficiência se dá devido à complexidade do modelo para o usuário, que tem dificuldades em entender o modelo e ao grande número de INNER JOINS necessários a extrair consultas mais complexas, o que torna o processamento extremamente lento (KIMBALL, 1998).

Outro problema é o que se chama de inconsistência temporal, que é o fato de as informações do OLTP sofrerem modificações ao longo do tempo, assim, duas consultas baseadas no mesmo parâmetro de entrada solicitadas em instantes diferentes poderão gerar resultados diferentes (KIMBALL, 1998).

Um data warehouse é o ponto central da arquitetura de processamento de informações para tomada de decisões, fornecendo o suporte informacional para os sistemas de apoio à decisão (SAD) e para os “*executive information systems*” (EIS). O data warehouse é construído sob uma perspectiva de armazenamento de informações de longo prazo. Sua base de dados é orientada por assunto, consolidada, integrada, variável com o tempo e não volátil (INMON e HACKCATHORN, 1997), fornecendo informações precisas, completas e apoiando a necessidade de informação do usuário a fim de lidar com aspectos estratégicos de longo prazo (HARRISON, 1998).

O data warehouse deve ser capaz de gerar rapidamente pequenos conjuntos de respostas baseados em uma infinidade de registros. Usuários de um data warehouse alteram constantemente o tipo de consulta efetuada, dependendo de suas necessidades. Estas consultas podem utilizar-se de centenas ou milhões de registros, causando impacto variável na performance do banco de dados (KIMBALL, 1998).

• O Modelo Dimensional

No modelo dimensional os dados são visualizados como um cubo de várias dimensões, de forma que os usuários possam cortar este cubo em quaisquer dimensões a fim de visualizar suas informações de forma simples e flexível. No modelo dimensional é possível obter respostas facilmente a perguntas como “Qual o volume de vendas do produto X, nos meses de dezembro, dos últimos cinco anos?” (HARRISON, 1998; KIMBALL, 1998). ”

O modelo dimensional, também chamado de esquema estrela, devido ao formato de seu diagrama, combina uma ou mais tabelas centrais em que os dados históricos do negócio (tabelas de fatos) são ligados a tabelas descritivas (tabelas dimensionais), tais como tempo, produto, localização geográfica etc.

As tabelas de fato são mantidas em linhas, cada uma contendo uma ou mais chaves estrangeiras correspondentes às dimensões do negócio. Cada chave estrangeira estabelece um relacionamento com uma tabela de dimensão.

As tabelas dimensionais descrevem as dimensões dos dados do negócio (períodos de tempo, produtos, mercados, contas etc.). As tabelas

dimensionais possuem diversos atributos descritivos e devem ser projetadas sob o ponto de vista do usuário do data warehouse.

É importante não utilizar as chaves candidatas determinadas pelas técnicas de normalização como chaves primárias da tabela de fatos e das tabelas de dimensão. O ideal é utilizar chaves artificiais, a fim de aumentar a flexibilidade e consistência, pois, desta forma, mudanças estruturais causam um impacto menor no modelo, além de aumentar a performance pela redução do tamanho dos índices, uma vez que chaves candidatas freqüentemente são compostas por mais de um campo.

As vantagens do modelo dimensional sobre o modelo entidade-relacionamento são a simplicidade, a flexibilidade e a velocidade para se extrair informações a partir de uma grande quantidade de dados primários, uma vez que o modelo é facilmente compreendido pelo usuário e as junções necessárias a uma consulta são reduzidas em relação ao modelo entidade-relacionamento.

A figura 2 transpõe o modelo entidade/relacionamento, ilustrado anteriormente, para o modelo dimensional.

. O modelo *snowflake*

O modelo *snowflake* emprega a normalização das tabelas dimensionais como forma de preservar a integridade referencial dos dados e reduzir a redundância e a desnormalização para aumentar o desempenho (HARRISON, 1988). A desvantagem do modelo *snowflake* em relação ao modelo estrela é o aumento da complexidade da estrutura, aumentando o número de *joins* e dificultando a compreensão do modelo por parte do usuário.

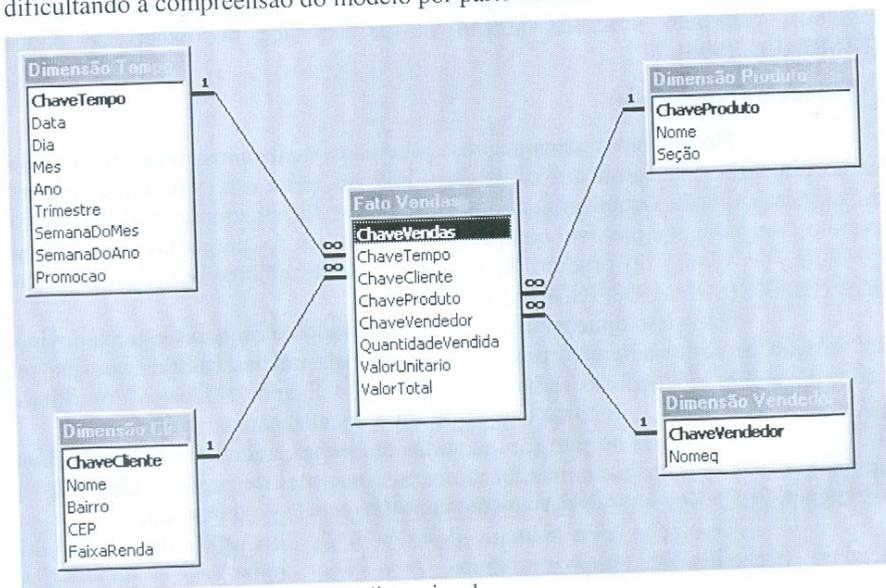


Figura 2 – Um exemplo de modelo dimensional

. Uma Proposta de Modelo Híbrido

A utilização do modelo entidade/relacionamento nos sistemas OLTP da organização e do modelo dimensional no *Data Warehouse* apresentam diversas vantagens, tais como:

- Possibilidade de se manter o banco de dados de OLTP normalizado e manter a performance alta, uma vez que este não tem a tarefa de recuperar grandes quantidades de dados de uma única vez.
 - Oferecer um ambiente de apoio à decisão flexível, simples de se utilizar e eficiente na performance, através da utilização de *Data Warehouses* dimensionais.
- Entretanto, este modelo apresenta algumas desvantagens, tais como:
- Necessidade de se aumentar os recursos humanos, de *hardware* e de *software* e, conseqüentemente, elevando os custos necessários para se manter as duas estruturas funcionando perfeitamente.
 - Necessidade de se utilizar um processo de transição (estagiamento) dos dados a fim de se converter as informações do OLTP para o *Data Warehouse*.
 - Necessidade de se investir em ferramentas específicas para estagiamento de dados.

Estas desvantagens podem retirar o *data warehouse* e seus benefícios do grupo de empresas que não dispõe de recursos financeiros necessários para tal.

A proposta é chegar a um modelo de dados híbrido que permita que se retirem informações estratégicas a partir da base de dados do OLTP, possibilitando, assim, que corporações que não dispõem de uma quantidade de recursos suficientes para investir em duas tecnologias possam se beneficiar da utilização do *data warehouses* para seus sistemas de apoio à decisão.

A idéia é manter as tabelas de fato com a menor unidade de informação possível, a fim de satisfazer o OLTP e o *Data Warehouse* e criar uma estrutura em estrela ao redor destas tabelas de fato. As tabelas dimensionais podem ser interligadas, de forma a preservar a integridade dimensional, a fim de suportar o OLTP.

A figura 3, representa o sistema exemplo utilizando-se de uma estrutura híbrida. No modelo proposto, foi acrescentada a dimensão Bairro, que deixou de ser um atributo de cliente. O motivo do acréscimo desta dimensão foi preservar a integridade referencial entre cliente e bairro e eliminar a redundância de dados. Pelo mesmo motivo foram acrescentadas uma dimensão Seção e um relacionamento entre seção e produto. A tabela preço também foi acrescentada ao modelo, apesar de não ser uma dimensão.

As entidades Venda e Item Vendido foram substituídas pela tabela Fato Vendas. É importante ressaltar que os registros desta tabela não deverão sofrer exclusões ou alterações, ao invés disso, o sistema deverá prover operações

de cancelamento e de correção, mediante a adição de novos registros. Desta forma, uma vez que um fato foi gravado, este não terá seus valores alterados, o que preserva a consistência das informações extraídas ao longo do tempo.

O modelo proposto, resume-se a incrementar o modelo dimensional com tabelas necessárias para suportar o OLTP e normalizar as tabelas de dimensão, a fim de se eliminar a redundância e facilitar a imposição da integridade referencial sem, contudo, criar um *snowflake*. Isso provoca um aumento da quantidade de dimensões, como ilustram os exemplos das dimensões Bairro e Seção, criadas para impor integridade referencial. É importante ressaltar que as tabelas de dimensão estão todas ligadas à tabela de fatos, para preservar as vantagens do modelo dimensional em estrela.

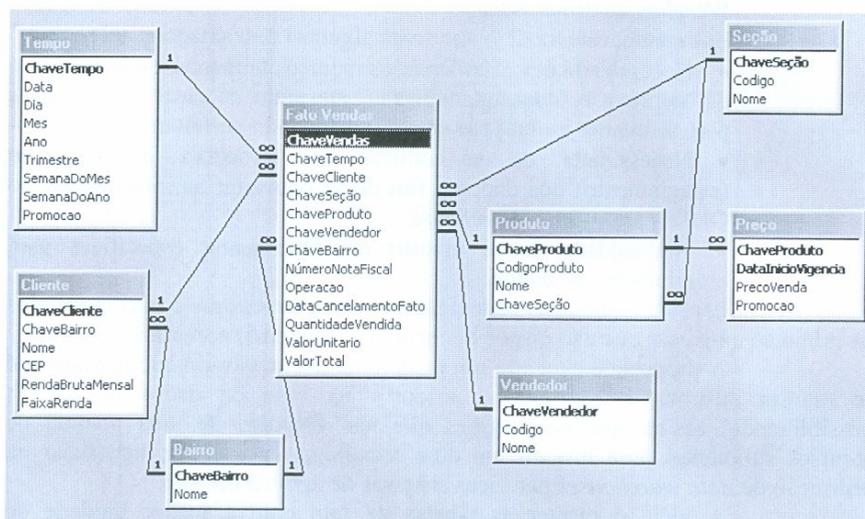


Figura 3 - Uma Proposta de Modelo Híbrido

As tabelas de fatos deverão ser mantidas com o menor grão possível, ou seja, elas deverão armazenar a menor unidade de informação possível. Caso o volume de informações armazenado dificulte a extração de informações, pode-se criar tabelas resumo para suprir esta deficiência.

. Conclusão

A utilização da internet como ferramenta de comércio se, por um lado, permite que a empresa amplie seu mercado para além de seus limites geográficos, por outro, eleva a concorrência ao âmbito global. É neste contexto que a utilização de sistemas de informação para apoio à decisão têm assumido um papel cada vez mais importante, independentemente, da área de negócios em que a empresa atua ou do porte da mesma. Para HARRISON (1998, p 3), “aumentar o

capital intelectual das empresas é uma necessidade competitiva (...). As organizações que usam com eficácia a tecnologia de informações adquirem conhecimento e velocidade para alcançar uma esmagadora superioridade nos mercados em que atuam”.

O modelo proposto não se destina a substituir *data warehouses* com *terabytes* de informações. Este modelo destina-se a suprir de informações aquelas empresas que não possuem recursos para investir nas tecnologias necessárias à construção de *data-marts* em separado.

Como se trata apenas de uma proposta ilustrada por um modelo simplificado, serão efetuados estudos mais aprofundados para validar o uso deste tipo de modelo no mundo real. Estes estudos começarão pela construção de protótipos de laboratório próximos da realidade das empresas. Estes protótipos serão exaustivamente testados, utilizando-se métricas apropriadas a fim de certificar-se da validade do modelo. Caso os estudos revelem a viabilidade do modelo, este deverá ser padronizado com um número significativo de regras que caracterizem o mesmo. Só então o modelo poderá ser implementado em sistemas reais.

. Referências Bibliográficas

FREITAS, Henrique e POZZEBOM, Marlei. **Características Desejáveis de um EIS – Enterprise Information System – Rumo à Proatividade**. PPGA – Universidade Federal do Rio Grande do Sul. Disponível em: <<http://read.adm.ufrgs.br/read05.artigo.eis.htm>> Acesso em: 10/11/2000.

FURLAN, José Davi; IVO, Ivonildo da Mota e AMARAL, Francisco Piedade. **Sistemas de Informação Executiva – EIS**. São Paulo: Makron Books, 1994.

HARRISON, Thomas. **Intranet Data Warehouse**. São Paulo: Berkeley, 1998.

INMON, W.H. e HACKARTHORN, Richard D. **Como usar o data warehouse**. Rio de Janeiro: IBPI Press, 1997.

KIMBALL, Ralph. **Data Warehouse Toolkit**. Makron Books: São Paulo, 1998a.

KIMBALL, Ralph. Et Al. **The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing Data Warehouses**. New York: Wiley Computer Publishing, 1998b.

KROENKE, David M. **Banco de Dados: Fundamentos, Projeto e Implementação – 6ª. Edição**. Rio de Janeiro: Livros Técnicos e Científicos, 1999.

SPRAGUE, Ralph H. Jr. e WATSON, Hugh J.. **Sistema de apoio à decisão: colocando a teoria em prática**. Rio de Janeiro: Campus, 1991.

STAIR, Ralph M.. **Princípios de sistemas de informação, uma abordagem gerencial**. Rio de Janeiro: Livros Técnicos e Científicos, 1998.

Data de Recebimento: 08/12/2002.

Data de Aceite: 28/05/2003.