

Business Intelligence: data warehousing na prática

Sergio Fred Andrade



Business Intelligence: data warehousing na prática

Business Intelligence: data warehousing na prática

Direcionado para pequenas empresas

Sergio Fred Andrade

[Capítulo 1 - Business Intelligence 5](#)

[Evolução histórica dos dados 10](#)

[Componentes e características de Data Warehouse 10](#)

[Arquitetura 12](#)

[Granularidade 13](#)

[Extração, Transformação e Carga de Dados \(ETL\) 15](#)

[Tabela de Fatos 17](#)

[Tabelas de Dimensão 18](#)

[Esquema Estrela 18](#)

[Técnicas de visualização da informação \(OLAP\) 21](#)

[Capítulo 3 - Base de dados de exemplo 24](#)

[Banco de dados de produção 24](#)

[O modelo conceitual através de ontologia de domínio 28](#)

[Metodologia para construção da ontologia de compras 30](#)

[Roteiro para Construção da Ontologia de Empresa Comercial 34](#)

[Da Ontologia para a Matriz de Barramento 44](#)

Business Intelligence: data warehousing na prática

[Capítulo 5 – Modelos lógico-multidimensional 52](#)

[Integração ontologia – modelo lógico-dimensional 52](#)

[Capítulo 6 - Arquitetura para data warehousing 55](#)

[Data mart de compras, vendas e logística e finanças 55](#)

[Diagrama de conversão de DM para DW 58](#)

[Data warehouse central 60](#)

[Considerações finais 61](#)

[Referências 62](#)

A quem é direcionado este Livro

Este conteúdo é direcionado para projetistas de software, programadores, analistas de sistemas, gestores empresariais, pesquisadores, professores e estudantes que buscam o entendimento sobre *data warehousing* e técnicas de visualização de informação através de OLAP para *Business Intelligence* (BI), do ponto de vista prático e que possam aplicar em seus projetos de software, acadêmico e de pesquisas.

É apontado, especificamente, para aqueles profissionais que buscam noções claras e objetivas de como fazer *data warehousing* na prática para alcançar os resultados certos que promovam a extração do conhecimento em bases de dados diversas e super povoadas, visando auxiliar no processo decisório organizacional e na análise de dados departamentais e empresariais.

As bases de dados consideradas aqui formam a origem para todo o trabalho de *business intelligence* e exemplos como compor os *data mart's* e o *data warehouse* central.

São três bancos de dados departamentais dos setores de gestão de contratos, compras e vendas e, logística de entregas de mercadorias, sem integração, de dados fictícios, empregados numa empresa de pequeno porte de material para construção civil.

Boa leitura e bom aproveitamento.

Capítulo 1 - Business Intelligence

As empresas são organismos vivos que produzem e registram dados e informações sobre diversos assuntos. Numa Operadora de Telefonia moderna, a cada ligação que opera produz inúmeros dados e essas informações são registradas num banco de dados. A cada ligação completada ou perdida os dados são inseridos num repositório. São exemplos de data, hora, números de origem e destino, titular da linha de origem e destino, local de origem e destino, tempo de duração se completou a ligação, e outros. Agora, vamos pensar quantas ligações são feitas numa determinada cidade, estado e país. Podemos ficar aqui, exemplificando diversas empresas, um supermercado, um posto de combustíveis, uma empresa de transporte coletivo, e outras.

Essas empresas registram informações de todas as suas atividades que acontecem nos setores. Ou seja, são dados departamentais sobre suas atividades rotineiras, como estoques, vendas, compras, manufatura, prestação de serviços, finanças, contabilidade, recursos humanos, marketing, segurança, patrimônio e outros.

Vamos imaginar que essas informações departamentais, são muitas vezes postas em relatórios que necessitam tempo de dedicação e leitura para interpretação e extração de conhecimento ao processo decisório.

Esses relatórios são uniformizados baseados em matriz bidimensional, contendo caracteres, de disposição pré-formatada com colunas fixas que recebem dados recuperados das tabelas do banco de dados para preencherem as linhas e exibirem informação sempre na mesma posição. Chegam para o gestor e nem sempre são lidos e vão se amontoando período após período. Geralmente escondem informação histórica crucial para gestão organizacional e não contribuem para novas descobertas nem situações inesperadas que podem estar ali camufladas.

Para resolver esse problema surgiu a *Business Intelligence*, uma concepção, arquitetura e modelagem sistêmica que organiza os dados, de diversos formatos e origens, como informações históricas para as atividades de análise, interpretação e extração de conhecimento objetivando o processo decisório. São diversas ferramentas, métodos e procedimentos para a prática do BI, em diversos tipos como *Data Warehouse (DW)*, *Data Mart (DM)*, ferramentas de *On-Line Analytical Process (OLAP)*, mineração de dados, entre outras.

Atualmente conhecidos como Sistemas de Apoio à Decisão (SAD) e são classificados para área de Ciência de Dados, originados de *Big Data* para servir à *analytical process*. Muitas vezes, é um processo oferecido às Organizações e Empresas de grande porte, em razão de elevado custo de projeto, desenvolvimento e aquisição.

Observe que estamos descrevendo situação de empresas de que contam com especialistas de várias áreas, estrutura organizacional planejada e atividades normatizadas e padronizadas. Ou, possa terceirizar essas tarefas facilmente.

Numa pequena empresa que também geram muitos dados, mas quase nunca emitem relatórios para tomada de decisão, em razão do elevado custo de pessoal especializado, sistemas gerenciais informatizados e impressão gráfica, essa situação fica mais difícil, ou seja, o uso de informações para o processo decisório é quase nulo ou inexistente.

Business Intelligence: data warehousing na prática

Porém, como pouco esforço e utilizando ferramentas de domínio público pode-se modelar e implementar aplicações para BI e aproveitar os benefícios dessa tecnologia para o processo de decisão empresarial.

O propósito deste livro está exatamente direcionado para os pequenos empreendimentos e suas necessidades de conhecimento dos dados que produzem e que não têm condições para adquirir um sistema de BI pelo elevado custo de obtenção. Os objetivos específicos desta obra incluem:

1. Demonstrar a técnica de *data warehousing*, cubos multidimensionais e técnicas de OLAP, sem auxílio de ferramentas proprietárias;
2. Proporcionar técnicas facilitadas para construção de cubos multidimensionais acessíveis aos micros empreendimentos;
3. Oferecer condições fáceis de entendimento para projetistas e programadores das técnicas de *data mart*, *data warehouse* e OLAP, para aplicação em website e aplicativo mobile;
4. Promover disponibilização de tecnologia em BI acessíveis e de menor custo possível;
5. Promover melhor entendimento e prática para estudantes de BI e SAD, nas disciplinas de cursos de graduação e pós-graduação.

Organização do conteúdo no livro

A estrutura deste livro está apresentada em 6 Capítulos descritos a seguir:

O primeiro uma noção sobre *BI* que tece uma rápida apresentação sobre a importância da utilização para entendimento das informações nos processos decisórios, as técnicas de BI e seus principais tipos.

No segundo Capítulo, é escrita a fundamentação para *data warehousing* e as principais técnicas para visualização de informação em procedimentos OLAP.

No terceiro Capítulo, o sobre os dados departamentais, bancos de dados e SGBD utilizados nos exemplos práticos do livro, para permitir melhor entendimento e possibilitar que o leitor refaça os procedimentos indicados.

O quarto Capítulo trata da modelagem conceitual para base multidimensional e aborda uma adaptação a matriz barramento e os elementos fundamentais para *data warehousing*.

No Capítulo quinto são apresentados o modelo lógico-multidimensional e a notação em álgebra multidimensional seguida para explicar os modelos, diagramas e relacionamentos utilizados fisicamente.

No último e sexto Capítulo, a implementação física dos *Data Mart's* e do *Data Warehouse* central, com a utilização da técnica estrutural de *bottom-up*.

Capítulo 2 - Fundamentação prática de data warehousing e OLAP

Business Intelligence: data warehousing na prática

A tecnologia de *data warehousing* suporta várias fontes e bases de dados diversificadas, que são periodicamente extraídas, adaptadas e uniformizadas e carregadas numa base estruturada multidimensional, preservando a evolução histórica das informações.

Essa base multidimensional deve ter dados limpos, estruturados e relacionados em todas as dimensões e mensurações para facilitar o compartilhamento através de cubos dimensionais ajustados para visualização das informações num processo decisório.

Importante dizer que um *data warehouse* é orientado por assunto. Numa empresa dizemos que é orientado por departamento, numa pesquisa é orientado por temática científica, na gestão pública é orientado por políticas públicas, numa escola pode ser orientado por matéria acadêmica, assim por diante. Observa-se que existem aí domínios e núcleos operacionais, ou seja, a atividade-fim ou operações da produção sempre orienta a formação de cubos dimensionais num ramo específico.

Na Figura 1 é ilustrado um exemplo onde os dados são originados de vários bancos de dados departamentais, ajustados e carregados num *data warehouse*, onde as informações são relacionadas e integradas para facilitar a visualização da informação e o processo de decisão.

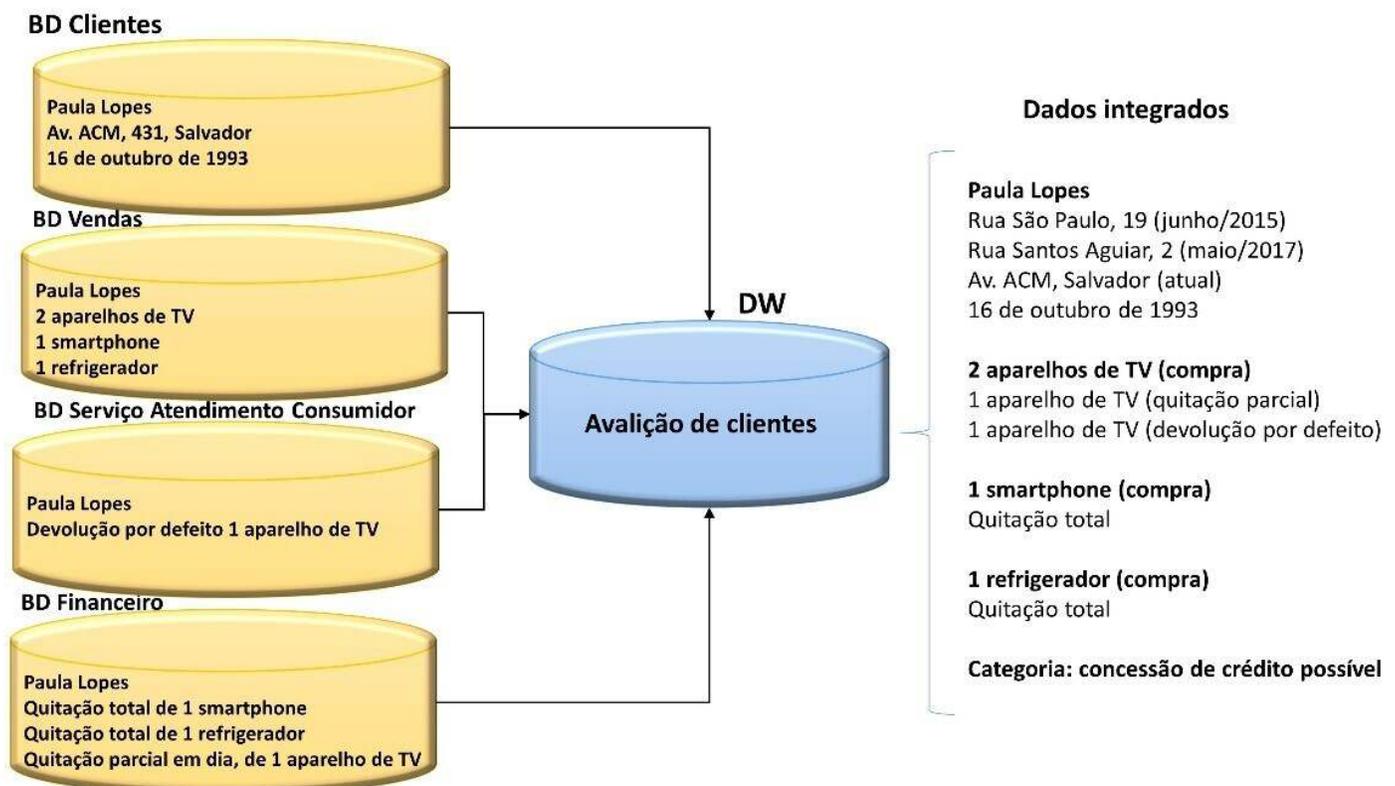


Figura 1 - Dados em bases de produção e no DW

Business Intelligence: data warehousing na prática

Observa-se que os banco de dados integrados são afins e complementares. Mas cada um isoladamente, não oferece a informação adequada para o processo de tomada de decisão. Nota-se também, a evolução histórica foi preservada, quando os vários endereços do cliente foram mantidos possibilitando melhores análises do contexto, inclusive um tratamento estatístico para avaliação conceitual de crédito.

As informações são retratadas geralmente por ferramentas de visualização de cubos e pivot dimensionais, OLAP, que ilustram consultas com gráficos e tabelas dinâmicas para o processo decisório.

A seguir são relatados os elementos importantes para Data Warehouse que foram adotados neste trabalho para nortear o mapeamento dos bancos de dados de origem, o processo ETL, a construção dos três Data Mart's influentes e o Data Warehouse central. Como também, os modelos conceituais e os modelos dimensionais trabalhados no exemplo prático.

Evolução histórica dos dados

Geralmente os sistemas de produção inserem, atualizam, removem e consultam dados num banco de dados. Esses sistemas são conhecidos por OLTP (*On Line Transaction Processing*), pois cumprem essas quatro transações básicas num Sistema Gerenciador de Banco de Dados (SGBD).

Os dados em OLTP's sempre são atualizados sobrescrevendo os registros anteriores. Por exemplo, em um *e-commerce* quando o comprador muda seu endereço os registros de rua, número de casa, bairro, cidade, estado da federação, casa alugada ou própria e outros afins, sobrescrevem os dados anteriores perdendo-se o valor histórico.

Esses valores, o anterior e o atual, são cruciais para uma loja de vendas. Imagine se o diretor de vendas deseja saber o potencial de pagamento do cliente, analisando sua capacidade econômica por tipo e categoria de moradia através da rua e bairro onde morava antes e mora agora, se era de aluguel e se agora é proprietário?

Essa situação no data warehouse é preservada porque existe evolução histórica registrada pelas cargas de dados periódicas, sempre mantendo-se os anteriores.

Componentes e características de Data Warehouse

A arquitetura de um *data warehouse* apresenta alguns elementos que o compõe e um conjunto básico de características que formam o modelo tecnológico.

Segundo Kimball e Ross (2002), os componentes do DW são necessários para composição do sistema e cada um atende a uma função específica. Os quatro componentes distintos são: sistemas operacionais de origem, *data staging area*, área de apresentação de dados e ferramentas de acesso a dados. A Figura 2 ilustra esses elementos integrados.

Business Intelligence: data warehousing na prática

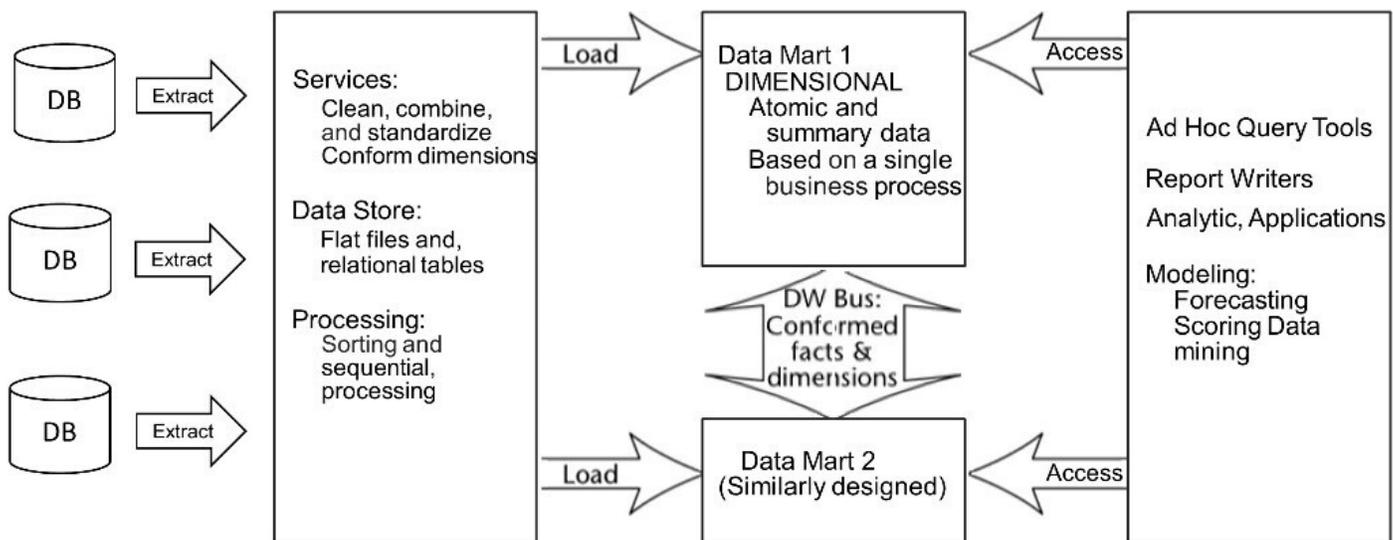


Figura 2 – Elementos básicos do *data warehouse*. Kimball e Ross (2002, p. 9).

Sistemas OLTP, de produção, são sistemas de captura dos dados das produções da organização. São considerados externos ao DW, servem para adição, edição, exclusão e consultas limitadas a registros, com pouca ou nenhuma evolução histórica. Geralmente cada sistema de origem é uma aplicação independente e trata os registros por vez, a cada acesso.

Data staging area é um conjunto de processos que normalmente é denominado de ETL (*Extraction, Transformation and Load*), ou seja, extração, transformação e carga, podendo ser também uma área de armazenamento temporária. Abrange todo mecanismo de extração de dados dos sistemas operacionais, a leitura e a sua compreensão. A transformação dos dados compreende o filtro, a limpeza, a preparação, correções de erros e integração dos dados para a próxima etapa que é a de carga no DW.

Apresentação de dados é o local onde as informações ficam armazenadas e organizadas em possibilidade de acesso, diretamente pelos usuários e por sistemas especialistas de consultas e relatórios para análise. Constitui no DW propriamente dito e, dependendo da arquitetura aplicada, pode ser formado por um ou vários *data mart's*, ou mesmo, podendo dele formar vários *data mart's*, que representa um domínio ou alguns processos do negócio.

As ferramentas de acesso a dados, último componente citado, formam os recursos necessários para os usuários acessar a base do DW através de *data mart*. Sendo considerado o principal ponto para as consultas como apoio às tomadas de decisões. São vários os recursos existentes nessas ferramentas que usam a forma dimensional específica e geralmente são baseados em cubos multidimensionais de dados, como exemplo de aplicações OLAP (*On-Line Analytical Processing*).

Business Intelligence: data warehousing na prática

Arquitetura

Data Mart (DM) é um subconjunto lógico e físico da área de apresentação do *data warehouse*. São DW's departamentais, pois têm as mesmas características de um *data warehouse* comum, e representam determinados processos de negócios. Tornam-se consideravelmente menos onerosos e sua construção é muito mais rápida em relação ao projeto do DW corporativo (BONIFATI, 2001).

Segundo Kimball e Ross (2002, p. 455), DM é um conjunto flexível baseado nos dados com alto nível de granularidade, extraído de fontes operacionais de informação e apresentado em um modelo dimensional. Um conjunto de DM representa processos de negócios e forma o barramento do DW corporativo. O que o torna mais funcional diante das consultas inesperadas dos usuários.

Já na visão de Inmon (2005), DM ou base multidimensional oferece flexibilidade aos usuários finais e são contidos do DW central. Deste DW, originam os dados e proporcionam a fonte robusta e útil para os data mart's.

Kimball e Inmon são considerados precursores da tecnologia de *data warehousing*, com grande contribuição para a produção do conhecimento nessa área. Mesmo assim, pode-se observar que são de orientações divergentes, embora os objetivos sejam os mesmos.

Essa divergência proporciona tipos de implementação de arquiteturas diferentes, com vantagens e desvantagens cada uma, a depender da situação, dos propósitos considerados para aplicação e dos problemas a serem resolvidos.

Os tipos de implementação de arquiteturas mais comuns são *top-down* – onde os DM's são criados a partir do DW central, orientação defendida de Inmon. E, *bottom-up* – onde os DM's são construídos representando áreas setoriais da organização e o seu conjunto forma o DW central, orientação defendida por Kimball. Estas implementações são descritas nos capítulos específicos seguintes sobre essas arquiteturas.

Granularidade

O conceito mais claro sobre granularidade menciona: “a granularidade diz respeito ao nível de detalhe ou de resumo contido nas unidades de dados existentes do *data warehouse*. Quanto mais detalhe, mais baixo o nível de granularidade. Quanto menos detalhe, mais alto o nível de granularidade.” (INMON, 2005).

Definir a granularidade de dados constitui num dos mais importantes estágios do projeto de DW, porque ela determina o volume de dados que são armazenados na base do *data warehouse* e influencia o tipo de consulta que pode ser atendida. Essa definição deve ser em níveis apropriados e em conformidade com as necessidades dos usuários.

Business Intelligence: data warehousing na prática

Para Kimball e Ross (2002), declarar a granularidade significa especificar com clareza o que uma linha do relacionamento das tabelas da base de dados do DW representa. A granularidade exprime o nível de detalhes associados às medidas dos fatos, como os dados descritivos e suas mensurações ou quantificações.

São exemplos de declarações de granularidade: um item de uma linha individual de um pedido de compras em certo dia ou de um item de uma determinada nota fiscal; um instantâneo diário dos níveis de estoques de cada material; ou, um instantâneo mensal de uma compra por fornecedor. Percebe-se que para cada nível de detalhe um horizonte de tempo é estabelecido.

Portanto, há uma relação inversa proporcional ao nível de granularidade e a possibilidade de utilização dos dados para consulta. Ou seja, definindo-se um nível muito baixo de granularidade é possível responder a quaisquer consultas, pois o nível de detalhe na base é maior. Situação que pode ser notada na Figura 3, onde a técnica *drill-down* representa essa orientação com o exemplo de um item de um pedido de vendas por dia.

Porém, com um alto nível de granularidade, o número de questões a que os dados podem satisfazer é limitado, pois o nível de detalhe na base é menor. Na Figura 3 observa-se essa situação, a técnica de visualização da informação apropriada é a *drill-up* que permite agregar dados tornando-se baixo nível de detalhamento, ideal para totalizações, sumarizações ou elevações em níveis hierárquicos.

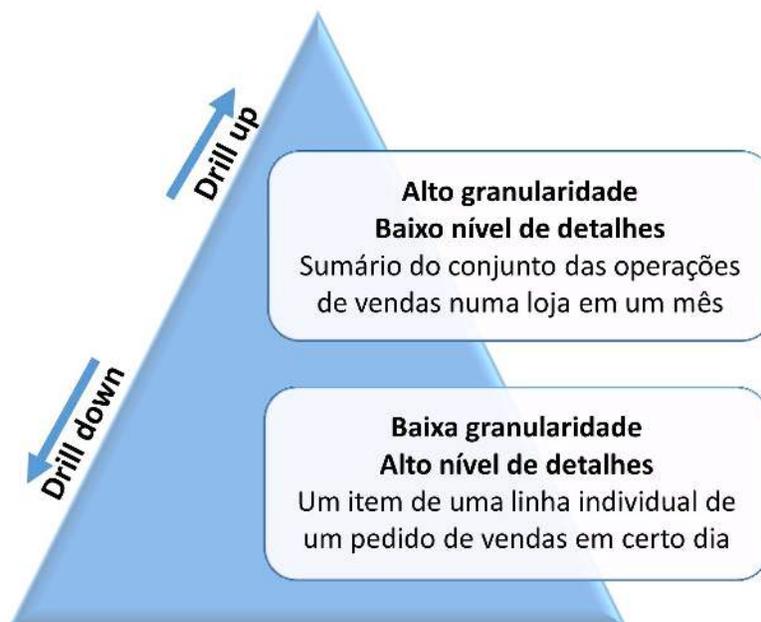


Figura 3 - Níveis de dados para granularidade

A escolha do nível de granularidade, ser mais ou menos detalhada, também influencia no volume de dados registrados na base do DW, como também, na eficiência e rapidez nas consultas. Com um nível de granularidade mais baixo o volume de dados torna-se significativamente grande, ainda mais considerando um horizonte de tempo entre 5 a 10 anos.

Business Intelligence: data warehousing na prática

São situações para se preocupar durante a fase de projeto do DW, motivo pelo qual, torna-se a granularidade fator importante e condicionante na definição do modelo de *data warehouse*.

Extração, Transformação e Carga de Dados (ETL)

O processo de *data staging area* requer etapas de Extração, Transformação e Carga (ETL) de dados para povoamento o *data mart* ou diretamente para o *data warehouse*, a depender da implementação utilizada.

Trata-se de um esforço extremamente necessário para obtenção de dados dos sistemas operacionais de transação e estima-se que se consuma um terço do tempo e orçamento para construção de DW. E que, em razão da alta curva de aprendizagem para ferramentas de ETL, muitas organizações preferem adotar seu próprio desenvolvimento, manutenção e uso dessas aplicações.

Para Kimball e Ross (2002), na prática, o processo de ETL inicia nos dados brutos dos sistemas operacionais de transação e os prepara para o modelo dimensional na área de apresentação de dados, que servirá para disponibilizar consultas para os usuários finais. As etapas principais envolvem:

- **extração de dados dimensionais de sistemas operacionais de origem** – pode-se tomar como fontes dos dados internos e externos, banco de dados corporativos, arquivos planos de texto, planilhas, arquivos semiestruturados ou não, de *intranet* ou *internet*;
- **limpeza de valores de atributo** – tomar ações que envolvam análises de nomes, endereços, valores descritos inconsistentes, decodificações alteradas ou ausentes, códigos sobrepostos, dados inválidos, datas sem padrão, unidades de medidas sem uniformização, dentre outras;
- **gerenciamento de atribuições de chaves substitutas** – atribuir e manter chaves primárias substitutas para tabelas dimensões, através de referência cruzada, no formato numérico e sequencial para garantir a integridade referencial e os valores históricos, mesmo quando houver alteração nos sistemas de origem;
- **transformar os dados de fatos** – essas transformações são cálculos numéricos, totalizações, conversão de tempo, equivalência de moedas, valores de unidades de medidas, normalização de fatos e tratamento de valores nulos;
- **construir tabelas de fatos de agregação** – tabelas de fatos normalmente são integradas conforme a granularidade definida, agregando-se valores de fatos de medição e uniformizando descrições textuais para as tabelas de dimensões;
- **fazer carga em massa de dados** – é o processo de povoamento propriamente dito após a etapa de limpeza, agregação e transformação. Deve-se manter históricos cíclicos e gerenciamento do volume de dados.

A Figura 4 ilustra um processo básico de ETL com as etapas necessárias para povoamento dos dados no DM e DW.

Business Intelligence: data warehousing na prática

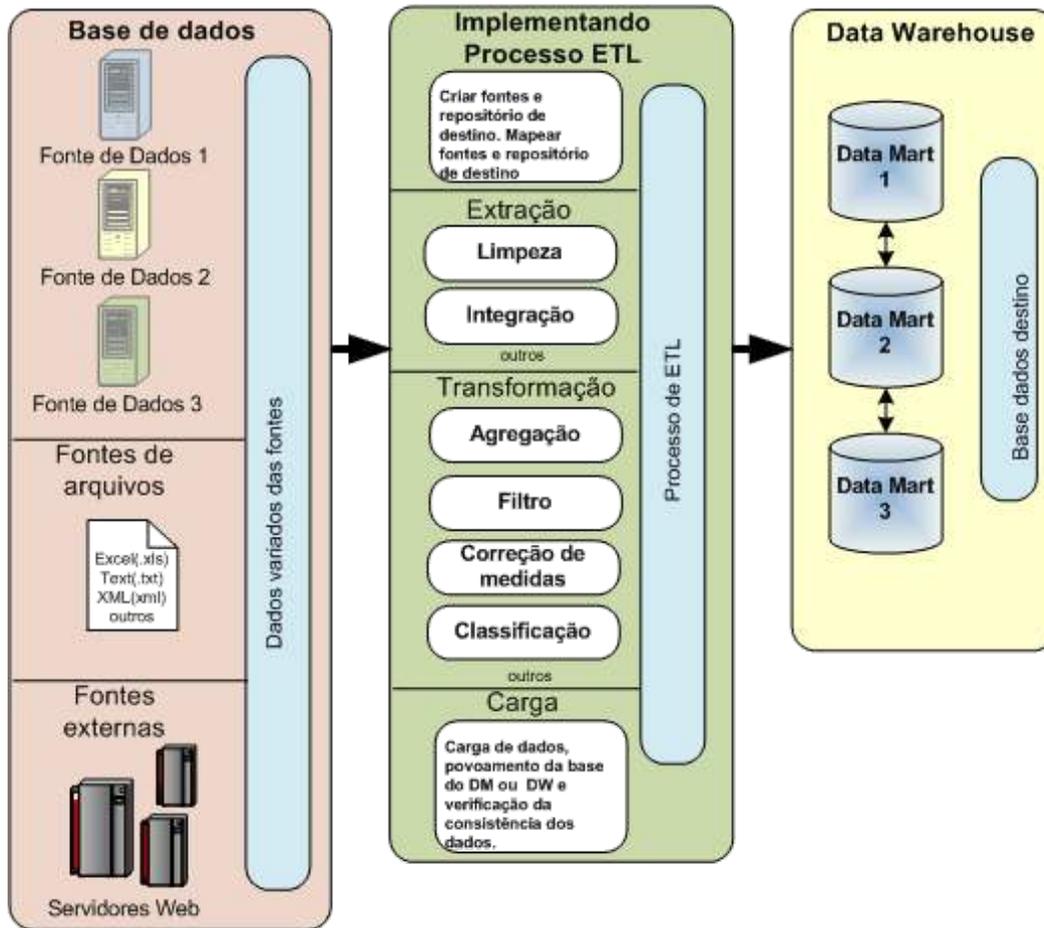


Figura 4 - Processo de ETL.

Em razão do propósito deste livro abordar o diretamente processo de *data warehousing* e técnicas OLAP, não serão detalhadas as etapas influentes no processo ETL.

Dimensões e fatos sumarizados

O DW utiliza a modelagem dimensional, que induz ao formato de um cubo, podendo apresentar três ou mais dimensões que participam de fatos sobre um assunto de negócio, o qual contém em cada ponto interno medidas relacionadas com as próprias dimensões. Os elementos básicos que formam essa modelagem no banco de dados do DW são as tabelas de fatos e dimensões (BONIFATI, 2001). A Figura 5 ilustra exemplo desse relacionamento.

Business Intelligence: data warehousing na prática

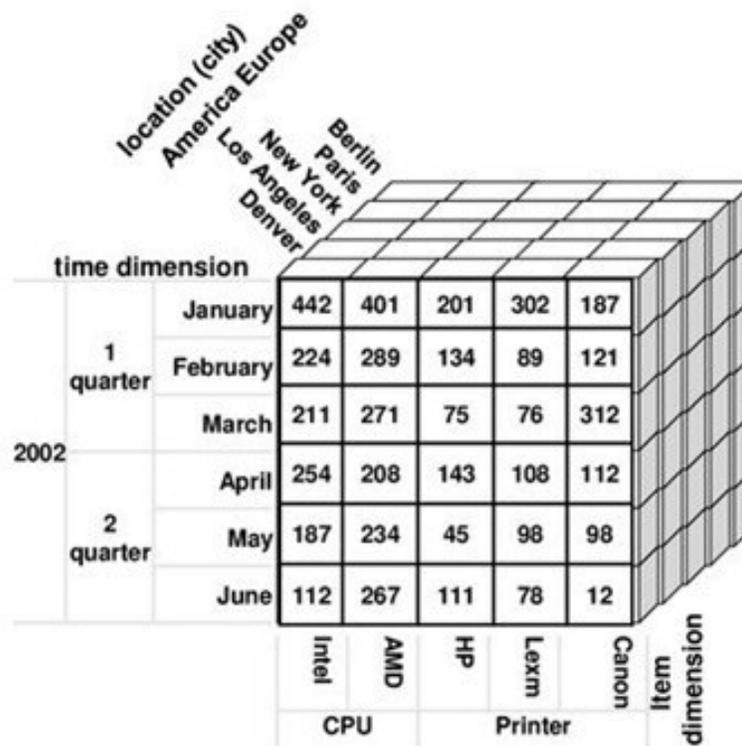


Figura 5 - Cubo de dados com dimensões e fatos.

Tabela de Fatos

Corresponde a principal tabela do modelo pois armazena as medidas numéricas e aditivas do negócio. Os dados correspondem a quantitativos, totais de valores por tempo ou outro parâmetro, números estatísticos e outras mensurações. Cada linha na tabela de fatos constitui numa medição e todas as linhas devem estar ajustadas na mesma granularidade.

A tabela de fatos relaciona-se às tabelas de dimensões através de chaves estrangeiras – *foreign key* - notação *FK*. Cada chave estrangeira deve corresponder a uma chave primária na tabela de dimensão referenciada. Quando todas as chaves estrangeiras se corresponderem com as respectivas chaves primárias de cada tabela dimensional, então, pode-se considerar que o conjunto de tabelas satisfaz ao modo de integridade referencial, podendo-se acessar as tabelas de fatos através das tabelas de dimensões.

Em geral, a chave primária da tabela de fatos é formada por um composto das chaves estrangeiras, sendo denominada de chave concatenada ou composta. No modelo dimensional as tabelas de fatos expressam a relação de um-para-muitos com as tabelas de dimensões.

Tabelas de Dimensão

São tabelas que armazenam textos descritivos das dimensões dos processos de negócios, que têm atributos representados por texto como, nome, endereço, especificação, marca, unidade de fornecimento entre outros.

Cada tabela de dimensão possui muitos atributos ou colunas que descrevem o sobre o assunto do domínio específico, ou seja, qualificam a entidade do processo de negócio.

Em cada tabela de dimensão deve existir uma chave primária identificadora – *primary key* – *PK*, do tipo *surrogate key* - *SK*, substituta, numérico auto incremental e sem significância com o domínio, que é referenciada pela chave estrangeira na tabela de fatos, formando a base da integridade referencial. Portanto, essas chaves correspondem ao ponto de acesso para as tabelas de fatos e são usadas como interface entre os usuários e o DW.

Normalmente, na arquitetura esquema estrela a qual é seguida pela metodologia deste livro, as tabelas de dimensão aceitam redundâncias e não possuem regras rígidas de normalização, para caracterizar a evolução histórica dos fatos registrados ao longo do tempo. Como também, propiciar melhor desempenho nas consultas e facilidade de entendimento aos usuários.

Esquema Estrela

A reunião da tabela de fatos e das tabelas de dimensões representa o esquema multidimensional do DM ou DW, que em outras palavras, é a tabela de fatos - formada por dados numéricos e de mensurações, associada às tabelas de dimensões que armazenam atributos descritivos dos processos de negócios.

Essa associação caracteriza uma estrutura semelhante ao formato de uma estrela, denominada de *star schema* ou esquema estrela, termo segundo Kimball e Ross (2002), que tem sua origem nos primórdios dos bancos de dados relacionais.

O esquema estrela apresenta uma estrutura física simples, poucas tabelas e relacionamentos, que promove aos usuários facilidade de entendimento e compreensão dos dados. Toda dimensão forma pontos simetricamente iguais de entrada para a tabela de fatos. A Figura 6 ilustra um exemplo de esquema estrela.

Business Intelligence: data warehousing na prática

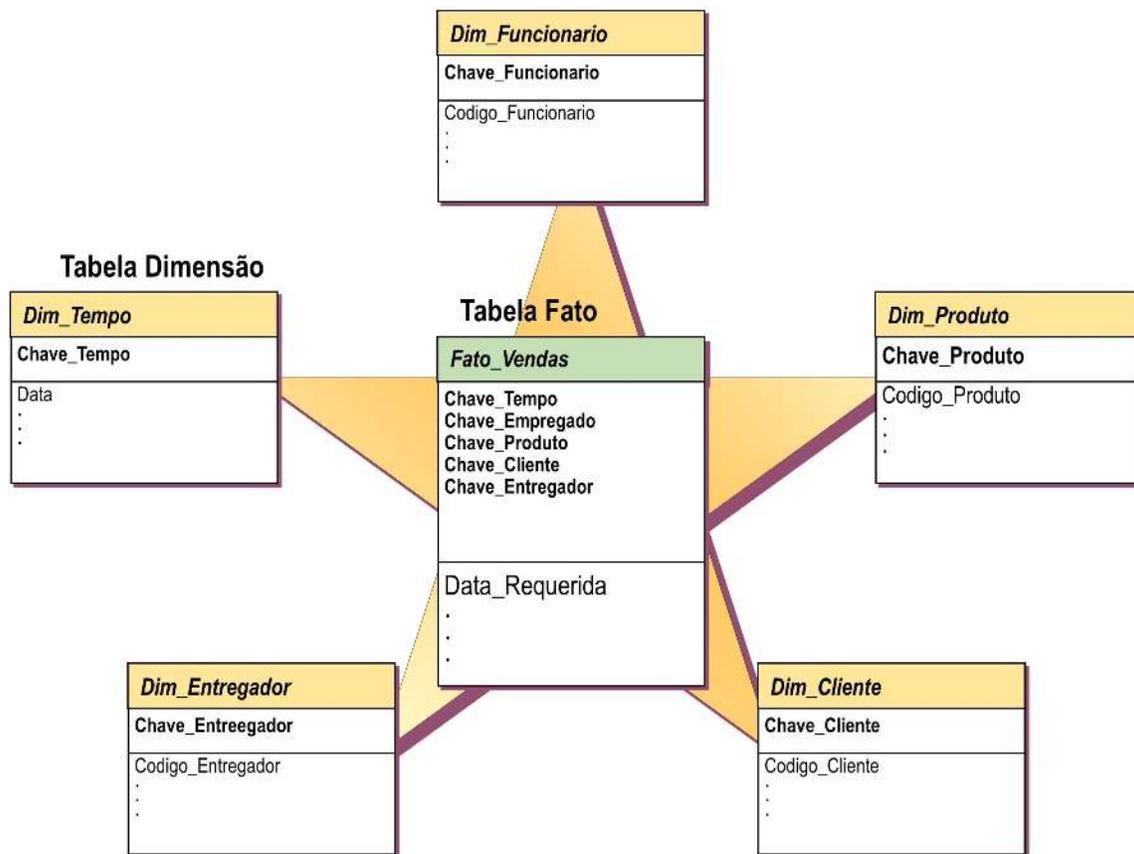


Figura 6 – Representação de esquema estrela tradicional.

Segundo Kimball e Ross (2002), o modelo dimensional com esquema estrela apresenta diversas vantagens para utilização no DW. Que são:

- arquitetura padrão e previsível;
- dimensões do modelo são equivalentes, ou seja, podem ser vistas como pontos simétricos para a tabela de fatos;
- modelo flexível, permite a inclusão de novos elementos de dados e mudanças que ocorram no projeto;
- facilidade na alteração das tabelas de fatos e dimensões;
- todas as aplicações que existiam antes da mudança efetuada, permanecem em funcionamento sem problemas.

Business Intelligence: data warehousing na prática

Um aspecto importante no modelo dimensional são as chaves primárias das tabelas de dimensões que são identificadoras e, denominadas de chaves substitutas ou *surrogate keys*. Devem ser um atributo único e definido pelo sistema como inteiros atribuídos sequencialmente. São utilizadas apenas para permitir a junção entre as tabelas de dimensões e a tabela de fatos. Portanto, não são usadas as chaves provenientes dos outros sistemas fontes, como exemplo do código de produto ou do código de cliente.

A razão da utilização de chaves substitutas é para manter o DW isolado das regras operacionais visando gerar, atualizar, excluir e reutilizar os códigos utilizados nos sistemas operacionais de transação. Desta forma, o DW deve manter as informações durante muito tempo e não pode ficar vulnerável a problemas de sobreposição de chaves no caso de consolidações de dados. Outro aspecto importante, diz respeito a alterações de atributos nos sistemas transacionais, as chaves substitutas serão as bases para se manter o histórico destas alterações no DW.

Existe também uma variação do modelo dimensional *star schema* chamado de *snowflake* ou floco de neve, que constitui numa extensão do modelo estrela. O modelo floco de neve corresponde a extensão da ponta de uma estrela, representado por alguma tabela dimensão que passa ter outra tabela relacionada ou, ser o centro de outra estrela.

Esse modelo decorre da desnormalização e redução de cardinalidade do modelo estrela, violando as tabelas originais ao longo de hierarquias existentes em seus atributos. Kimball e Ross (2002) aconselham o não uso dessa estrutura, devido ao impacto da complexidade deste tipo de modelo sobre o usuário final, enquanto que o ganho em termos de espaço seria pouco relevante.

Neste Livro, não é considerado o modelo *snowflake* em razão da utilização da arquitetura *bottom-up*, com construção de DM no primeiro estágio, e da aceitação da concepção de Kimball et al. (1998), que advoga um sistema menor, departamental, com evolução gradativa.

Cubos dimensionais

Os cubos dimensionais são estruturas abstratas que representam relações entre as dimensões ou descrições dos assuntos, direcionadas aos fatos em quantificações ou totalizações.

Exemplo de um cubo dimensional de vendas, teria as dimensões produto, cliente, local e data. E os fatos sobre as mensurações sobre essas vendas, quantidades vendidas, totais de vendas, totais de recebimentos à vista, totais de recebimentos a prazo, estatísticas de vendas, segmentação por clientes, por local, por período, e outros.

Os cubos dimensionais formam a base para aplicação das técnicas de visualização OLAP, que permitem navegar entre as dimensões e suas mensurações, ter informações em níveis hierárquicas, dados integrados ou expandidos, em rotação nas dimensões permitindo diferentes visões, com fatiamento de dimensões possibilitando restrições nas visões de dados e outros.

Técnicas de visualização da informação (OLAP)

OLAP é um processo ou tecnologia para gerar visualização de informações. É classificado como um componente para finalizar uma arquitetura de DW e ambos fazem parte dos elementos de um BI. Geralmente é apresentado como ferramenta para visualização de informações e têm as seguintes funções: apresentação dimensional de dados, hierarquização para expansão e contração dados, rotação multidimensional de dados, visualização de cubos de dados, visualização de dados em *pivot*, transposição e cruzamento de tabelas.

Os recursos de OLAP tratam da saída de dados originados, se possível, da base de um DM ou DW. Que podem ser armazenados de arquitetura ou tipologia diferente como:

ROLAP (OLAP Relacional) - os dados são armazenados em SGBD relacional;

MOLAP (OLAP Multidimensional) - os dados são armazenados na forma multidimensional;

HOLAP (OLAP Híbrido) - trata-se de um método de armazenamento híbrido entre ROLAP e MOLAP;

DOLAP (OLAP Desktop) - dados multidimensionais armazenados num servidor SGBD e cedido para uma estação desktop.

A técnicas da aplicação de OLAP ilustradas na Figura 7, geralmente são empregadas com as seguintes funções:

Drill-down - ocorre quando o usuário aumenta o nível de detalhe da informação, diminuindo a granularidade (A granularidade determina quais os tipos de consultas podem ser empregados no DW. Ela influencia diretamente na velocidade do acesso às informações e no volume de dados armazenados);

Drill-through - passagem de uma dimensão de dados numa consulta para outra, sem perder o foco das informações mensuradas. Por exemplo, pode-se visualizar as vendas numa região e passar para vendas por cliente, sem perder o propósito das informações;

Slice - possibilita determinar um alvo da análise numa específica região da informação com uma ou mais dimensões. Por exemplo pode-se ter o propósito de analisar uma filial da rede da loja, seus determinados produtos e em períodos de vendas específicos ou gerais;

Dice - requer restrições específicas para consultas como filial da loja, período de vendas sobre determinado produto;

Pivot - Movendo-se entre os níveis de dados, desde os conjuntos mais detalhados até conjuntos de dados mais resumidos. Por exemplo, um nível pode ser sobre o total de vendas em cada estado ao longo do tempo, enquanto o usuário pode se aprofundar para ver mais detalhes de cada produto;

Business Intelligence: data warehousing na prática

Roll-up - resumindo dados combinando atributos baseados em uma hierarquia. Por exemplo, dados sobre as vendas em cada estado podem ser acumulados em dados sobre áreas geográficas maiores.

Crosstab - trata-se do movimento de duas ou mais dimensões ou conjunto de dados em tabelas cruzadas, onde num ordenamento os dados em linha passam para colunas mantendo-se a coerência das informações.

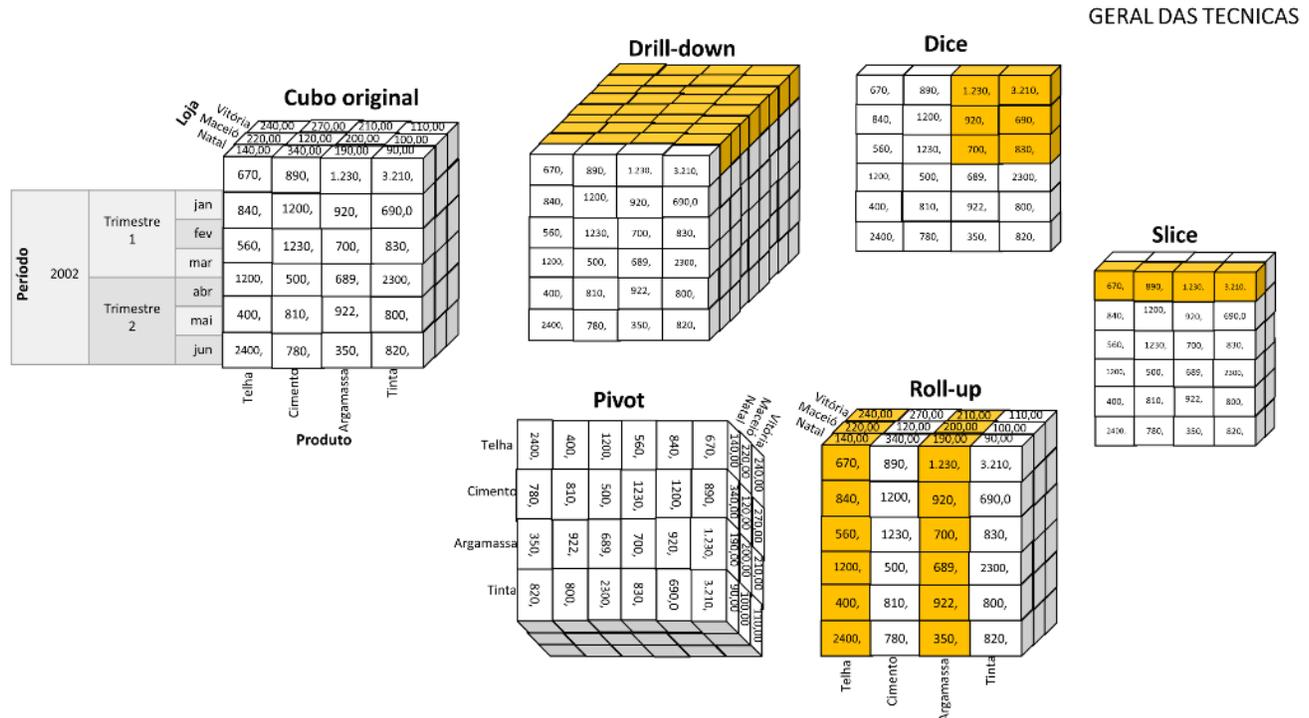


Figura 7 - Cubos com técnicas OLAP.

Capítulo 3 – Base de dados de exemplo

A partir deste Capítulo descreveremos as bases de dados fontes utilizadas para exemplificação da metodologia e resultados práticos da modelagem conceitual, lógico-dimensional e implementação física da base multidimensional, para formação dos *data mart's* e *data warehouse*. Para tanto, apresentaremos os bancos de dados (BD) empresariais de produção adotados como exemplo, e os modelos produzidos para os sistemas de BI.

Business Intelligence: data warehousing na prática

Banco de dados de produção

Com propósito de demonstrar o processo de modelagem estamos considerando três bancos de dados departamentais dos setores de gestão de contratos, compras e vendas e, logística de entregas de mercadorias, sem integração, de dados fictícios, empregados numa empresa tradicional de pequeno porte do comércio de material para construção civil.

Descreveremos cada banco de dados, em separado, para melhor compreensão do processo de negócio.

Banco de dados de gestão de contratos

Trata de um repositório de sistema para auxiliar na gestão dos contratos gerados nas vendas de mercadorias, para atendimento a pedidos de clientes e entrega por serviços por *delivery* através da frota da empresa. Toda venda gera um contrato e seus aditivos e um frete de entrega de mercadoria, tem um supervisor para acompanhamentos da logística até o endereço do cliente. Ver esquema lógico na Figura 8.

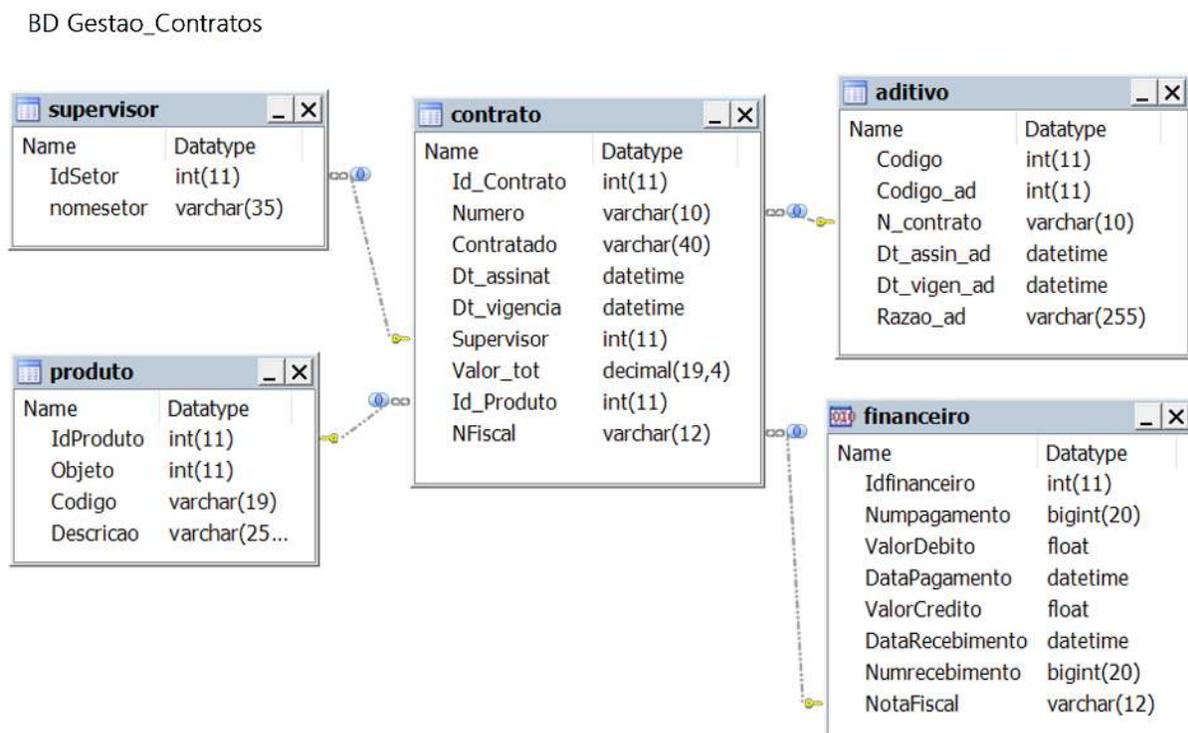


Figura 8 - Banco de dados gestão de contratos.

Banco de dados de compras e vendas

Business Intelligence: data warehousing na prática

Na Figura 9, é mostrado o esquema lógico do banco de dados do sistema de comercial que registra as compras e vendas de mercadorias e suas movimentações quantitativas e financeiras na gestão do estoque de mercadorias. Nas compras há o pagamento e o processo de recebimento de produtos dos fornecedores. Depois, o controle de estoque e disponibilização dos produtos para os setores de vendas de mercadorias com o devido recebimento financeiro.

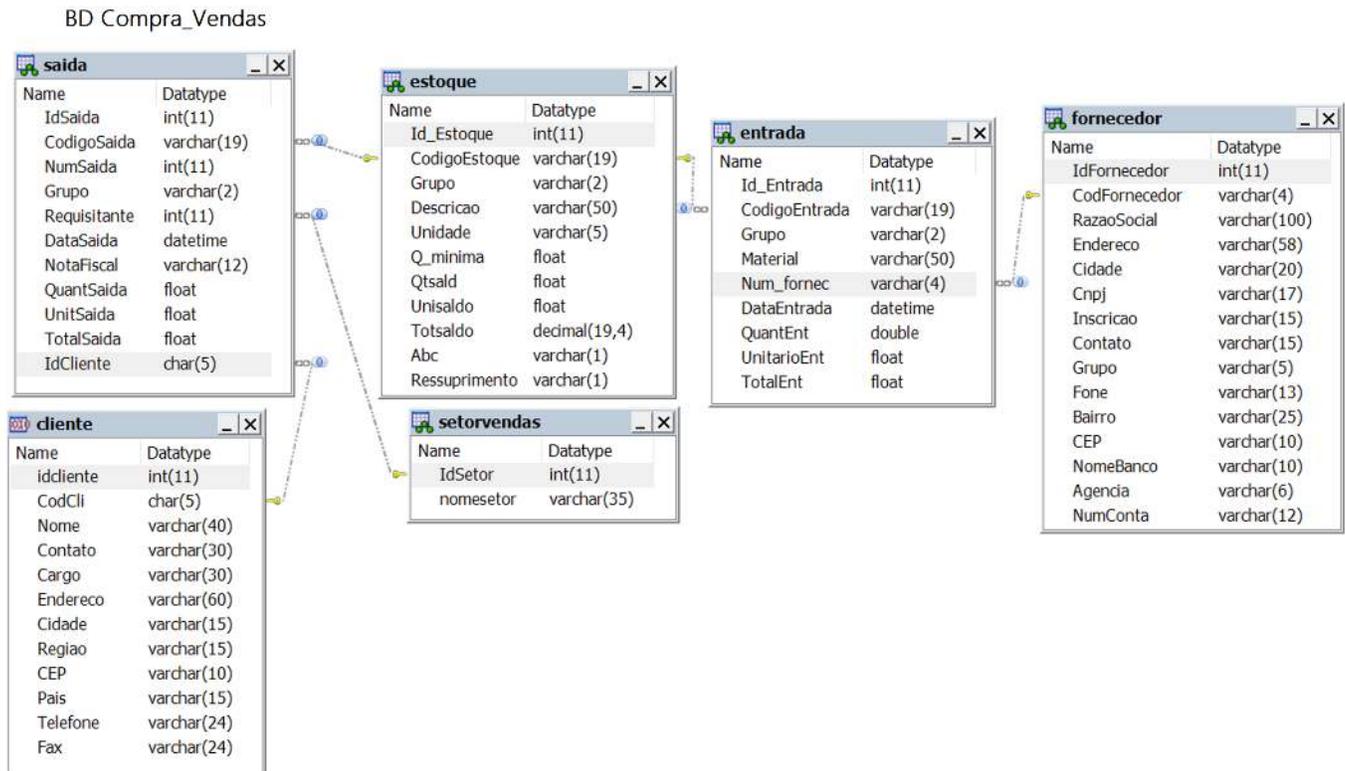


Figura 9 - Banco de dados de compras e vendas.

Banco de dados de logística de entregas

Em todo processo comercial as entregas são na modalidade *delivery* através da frota interna. O sistema de logística tem um banco de dados, ver na Figura 10, que registra a movimentação das entregas, as cargas, origem e destino, os abastecimentos dos veículos e o motorista responsável.

Business Intelligence: data warehousing na prática

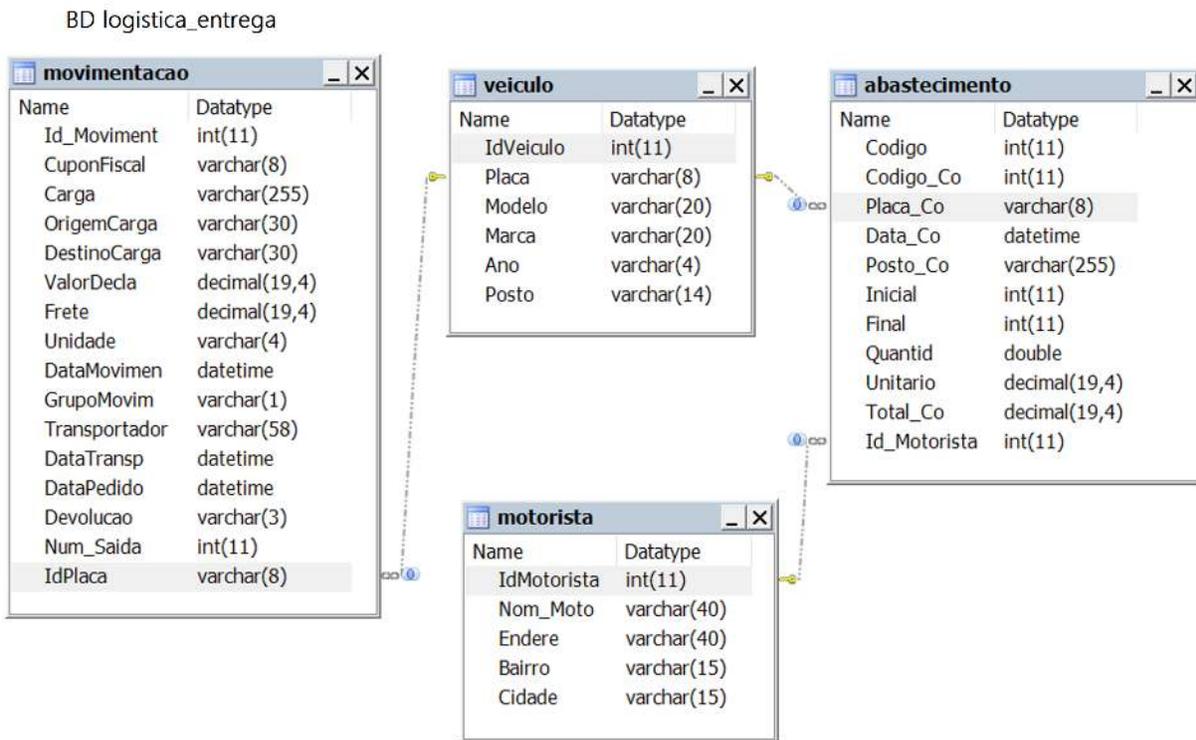


Figura 10 - Banco de dados de logística das entregas.

Diante das fontes de dados e da caracterização dos processos de negócio, decidiu-se adotar a arquitetura ROLAP, na construção de três *Data Mart* compras, vendas e logística e finanças, conforme ilustra a Figura 11.

A modelagem conceitual é apresentada a seguir em dois eixos – através da matriz de barramento e através de ontologia de domínio. São abordagens diferentes, mas que seguem o mesmo sentido e objetivos do processo de *data warehousing*.

A abordagem pela matriz de barramento é adaptada do modelo tradicional de Kimball et al. (2005), porém com incrementos direcionados para as operações de OLAP. A abordagem pela ontologia de domínio é uma inovação, e apresenta formas de automatizar a construção de consultas OLAP através de mecanismos de semânticas por reconhecimento de máquinas.

Business Intelligence: data warehousing na prática

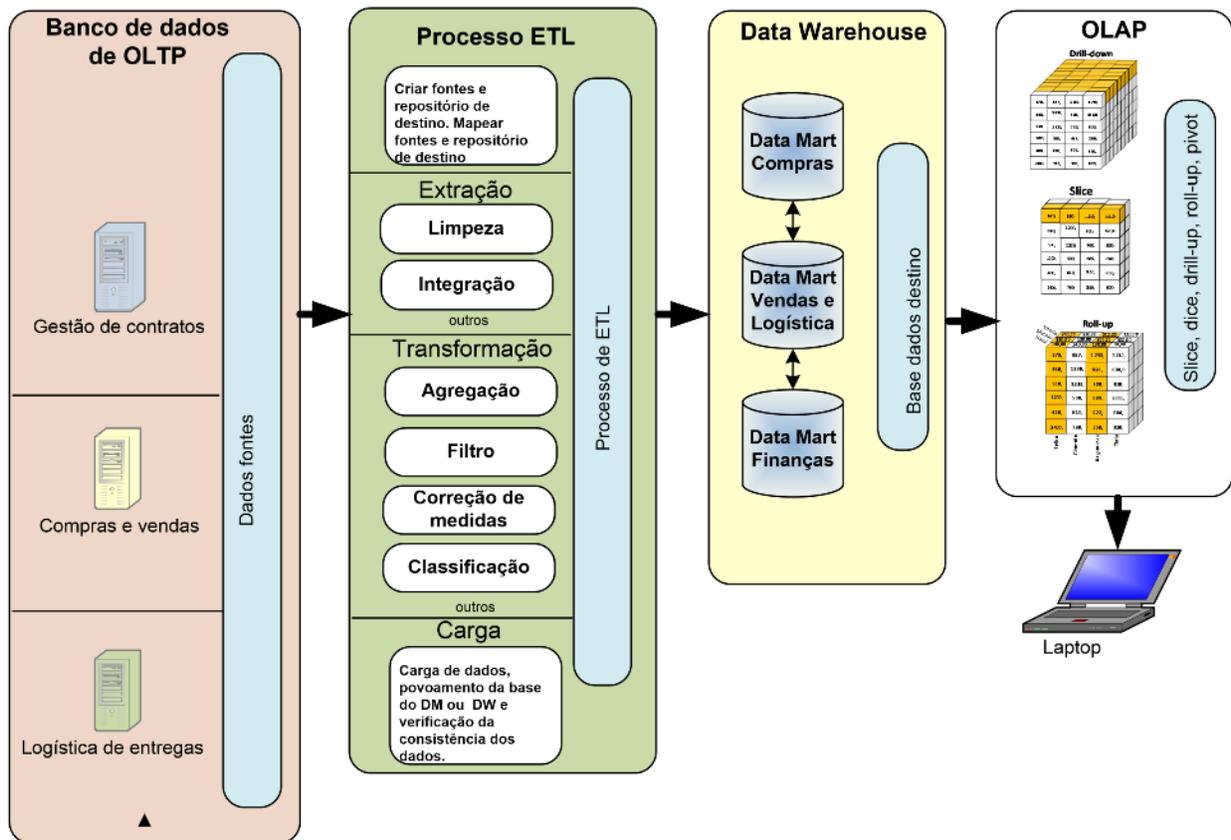


Figura 11 - Estrutura do modelo de *Data Warehouse*.

Capítulo 4 - Modelo Conceitual Multidimensional

O modelo conceitual através de ontologia de domínio

A presente abordagem para modelagem multidimensional visa prover condições para que o analista e o usuário possam entender mais facilmente o modelo conceitual utilizando a linguagem própria do domínio, o que possibilita resolver um problema sobre a dificuldade para construir cubos de dados e uso das técnicas mais apropriadas de OLAP.

Para tanto, foram considerados os três BD's departamentais de uma empresa comercial, apresentados anteriormente, com dados referentes as compras e vendas, gestão de contratos e logística, para mostrar o processo de modelagem.

Modelo Conceitual

O dicionário de dados dos BD's fontes, a descrição do domínio organizacional, as regras do negócio e as operações organizacionais formam a base de requisitos para a modelagem conceitual multidimensional. Esses conceitos são usados para os termos da taxonomia da ontologia empresa comercial, onde as classes, subclasses, propriedades e associações entre as classes, servem para estruturar os esquemas dos DM's.

Business Intelligence: data warehousing na prática

Foi utilizado o software Protégé 5.5, descritos em Rubin et al. (2007), para auxiliar no processo de produção da ontologia, geração da taxonomia e sua validação, por ser uma ferramenta de fácil acesso e ampla documentação. Porém, esse roteiro usando ontologia é opcional, podendo o modelo conceitual ser elaborado com a experiência e ajuda dos técnicos do ramo específico, neste caso, o de empresas comerciais.

As operações organizacionais indicadas na elicitação de requisitos são candidatas às classes da ontologia e representam os objetos do domínio. As características desses objetos são relacionadas às propriedades das classes da ontologia e são confrontadas com os atributos das tabelas dos BD's fontes.

Dos atributos selecionados é feita triagem conforme as propriedades das classes da ontologia e as operações organizacionais, podendo ser integrados, resumidos ou descartados. Desta forma, as propriedades e atributos são vinculados às dimensões e fatos no propósito de atender fielmente às consultas organizacionais que devem representar alternativas para tomada de decisão.

Após isso, é composta a matriz de barramento dinâmica demonstrada a seguir. Trata-se de uma estrutura onde os questionamentos para consultas organizacionais estão vinculados às operações organizacionais, em cada linha, e apontam para as dimensões e fatos que estão em colunas, formando uma intersecção de dependência. Essa estrutura dinâmica permite incluir, editar ou excluir processos de negócios ou consultas decisórias. Serve para todo ciclo de vida do DW e deve orientar a modelagem multidimensional, a arquitetura do DW, a geração dos cubos dimensionais dos DM's e uso das técnicas OLAP.

Conhecendo os questionamentos sobre os processos de negócio, pode-se chegar as dimensões e fatos. As entidades fatos constituem as principais tabelas do modelo multidimensional, pois armazenam dados numéricos com relacionamento direto para todas as dimensões descritivas, definindo a granularidade dos dados, a qual corresponde aos dados atômicos.

A dimensão "data", que representa temporalidade expandida ou contraída, como na ordem trimestral, mensal, quinzenal, diária ou, vice-versa, tem particularidade diferente das outras entidades, pois é formada por uma conjunção de valores inerentes às diversas dimensões, podendo representar as ocorrências históricas.

Metodologia para construção da ontologia de compras

Para exemplificar o modelo de dados conceitual para os Data Mart's e Data Warehouse foi utilizado a técnica da abstração da ontologia do domínio direcionado apenas para compras em empresas comerciais. Mas, exemplo ilustrado a seguir serve também para vendas, logística, finanças e outros processos organizacionais em empresas comerciais, guardado as devidas especificidades de cada subárea.

Ontologia de domínio serve para elaboração de terminologia única, com o propósito que o conhecimento produzido possa ser compartilhado e permita a sua reutilização. Gruber (1993) diz que uma ontologia deve ser considerada como especificação formal, explícita e de conceituação compartilhada. De certa forma, esse conceito é considerado na literatura como um consenso comum. O termo formal indica que ontologia pode ser processada por computadores.

Business Intelligence: data warehousing na prática

Sendo um domínio, uma área de interesse por assunto que pode ser interpretado formalmente, principalmente por sistemas de informação através de objetos, atributos e relacionamentos, que podem ser representados numa linguagem formal, onde há possibilidade de descrever a ontologia num software estruturado para reconhecimento de entidades e execução de métodos comuns.

Aplicações geralmente utilizadas de ontologias estão presentes em sistemas como gestão do conhecimento, reconhecimento de linguagem natural, modelagem de empresas, modelagem de sistemas de conhecimento, módulos de interoperabilidade entre sistemas, aplicações de comércio eletrônico, *Web* semântica, interligação de banco de dados entre outros. (GRUBER, 1993).

A abordagem Methontology, de Fernández-Lopez et al. (1997), foi utilizada para modelagem da ontologia, combinada com o roteiro de construção de ontologias de Noy et al. (2001), em razão da integração conceitual entre esses métodos.

Os passos adotados foram os seguintes: a) determinar o domínio e o escopo da ontologia; b) considerar a reutilização de ontologias existentes; c) enumerar termos importantes para a ontologia; d) definir as classes e a hierarquia entre as classes; e) indicar as propriedades das classes; f) definir das características das propriedades em tipo e tamanho; e, g) criar instâncias individuais das classes ou subclasses na hierarquia.

Em uma concepção ampla, a metodologia para construção da ontologia de Empresa Comercial adotada neste livro, pode ser resumida nas seguintes ações de desenvolvimento:

a) elaborar um planejamento das etapas e tarefas que serão executadas, organizando as ações, pré-definindo o tempo gasto para efetivação e os recursos que serão envolvidos, como exemplo: profissionais, software, metodologia, pesquisa sobre o domínio, entre outros;

b) conhecer o domínio em todos os aspectos, através de pesquisas e de especialistas ou peritos no assunto, utilizar de técnicas apropriadas de elicitação do conhecimento, listar as fontes de conhecimento, definir previamente a elaboração de cada processo e apresentar a metodologia e técnicas que se utilizará;

c) após a obtenção do conhecimento com profundidade, deve-se especificá-lo num modelo conceitual, onde se descreve os problemas e as soluções;

d) transformar o modelo conceitual num modelo formal, tornando compatível e empregando as técnicas de representação do conhecimento;

e) as ontologias devem ser reutilizáveis, portanto, é necessária tentativa de integração da ontologia que está em desenvolvimento com o maior número permitido de outras ontologias existentes;

f) é necessário tornar a ontologia acessível a computadores, sendo preciso sua implementação em linguagem formal;

Business Intelligence: data warehousing na prática

g) realizar validações, antes de tornar-se a ontologia acessível, avaliando sua consistência e eliminando a possibilidade de erros conceituais;

h) é necessária documentação da ontologia para que se possa permitir ser reutilizada e compartilhada; e,

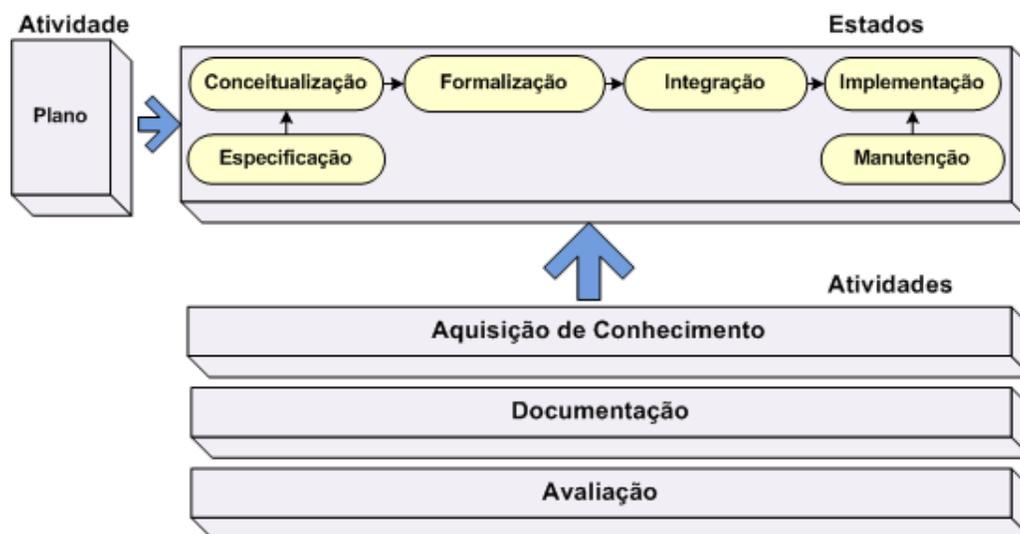
i) a manutenção da ontologia deve ser prevista, pois a qualquer tempo o conhecimento do domínio pode ser modificado, com inclusões ou alterações de conceitos e definições.

Em razão da integração conceitual e de definirem as etapas para execução de cada tarefa e técnica utilizada, os objetos de saída e avaliações, maior detalhamento na metodologia, recomendação de ciclo de vida, e, aplicação independente para estratégia de construção.

O guia do roteiro de Noy et al. (2001) utiliza o software Protégé e, entre outras linguagens, a *OWL (Web Ontology Language)* e *RDF/RDFS (Resource Description Framework/Schema)*, ambas recomendadas pela *W3C (W3C Consortium)*, para implementação e validação da sua ontologia exemplo. Essas são as ferramentas e linguagens utilizadas, neste trabalho, para construção da ontologia para compras para empresas comerciais.

A *Methontology* é uma metodologia com muitos recursos que permite indicar a seqüência e o nível de detalhamento que as atividades são elaboradas. É baseada no padrão do *IEEE (Institute of Electrical and Electronics Engineers)*, para o desenvolvimento de software possibilitando identificar quais as atividades deverão ser executadas no processo de construção de ontologias.

O ciclo de vida para a ontologia apresentada pela *Methontology* forma um conjunto de etapas que devem ser executadas durante o processo de construção. As etapas são: especificação, conceitualização, formalização, integração, implementação e manutenção. E estão ilustradas no ciclo de vida da Figura 12 e, descritos a seguir.



Business Intelligence: data warehousing na prática

Figura 12 – Ciclo de vida para construção de ontologia, Methontology. Fernández et al. (1997).

- **Especificação** - determina o propósito e o escopo da ontologia a ser elaborada, incluindo os objetivos e a utilização, assim como, o grau de formalidade, as características e granularidade de detalhe com concisão e consistência;
- **Conceitualização** - estrutura o domínio em um modelo de conhecimento, descreve os significados dos vocabulários num glossário de termos, que inclui classes, subclasses, propriedades e restrições;
- **Formalização** - representação em linguagem formal para reunir termos do glossário para reconhecimento através de computadores;
- **Integração** - constitui na reutilização de definições já desenvolvidas por outras ontologias, através de meta-ontologia e pesquisas em bibliotecas existentes;
- **Implementação** - utilização de ambiente de desenvolvimento para codificar em uma linguagem formal, através de analisador sintático e léxico, tradutores, editor, navegador e avaliadores;
- **Avaliação** - acompanhar num julgamento técnico o resultado desenvolvido da ontologia, através da verificação e validação, para garantir que a documentação corresponda ao sistema que se pretende representar.

Roteiro para Construção da Ontologia de Empresa Comercial

A construção da ontologia em questão decorre da necessidade de se definir e compartilhar entendimentos comuns sobre **o conhecimento do domínio de compras**, através de especialistas da área, da literatura pertinente e procedimentos padrões. Utilizando agentes de software com o propósito de estruturar modelagem dimensional para o *data mart* proposto.

Para a construção de ontologia existe um direcionamento de posições entre os pesquisadores da área, de que:

- não há um único roteiro ou guia para elaborar a modelagem do domínio, a solução mais viável depende da aplicação que se pretende alcançar;
- quase sempre o processo de construção necessariamente obedece a estágios interativos;
- conceitos numa ontologia respeitam, na maioria das vezes, às formas sintáticas de substantivos e verbos em sentenças que especificam o domínio.

O roteiro usado de Noy et al. (2001) serviu de base para a construção da ontologia em questão. É procedido pelos sete passos seguintes, que estão adaptados ao contexto da proposta deste trabalho.

a) Passo 1º - Determinar o domínio e o escopo da ontologia

O propósito da ontologia de compras para empresas comerciais é definir o domínio da área de conhecimento de compras nessas instituições, com vocabulário próprio e taxonomia, contemplando suas atividades básicas, os procedimentos padrões, as etapas nos processos de aquisição, estoques e logísticas, as modalidades de licitações, documentos padronizados, profissionais envolvidos e setores acadêmicos e administrativos como usuários do sistema.

Para aquisição do conhecimento específico, que é uma atividade independente, toma-se como fonte os especialistas da área do domínio, legislação pertinente, livros técnicos, manuais, dados armazenados e outras ontologias, se existir. As técnicas para obtenção do conhecimento são: entrevistas, análise de textos, diagnósticos, dentre outras.

O escopo incluirá um conjunto de significados a serem representados num glossário de termos e estrutura de hierarquia de classes, subclasses, *slots* (propriedades) e restrições, com características e granularidade. Utilizando uma linguagem formal para processamento por computadores.

No caso específico, a utilização dessa ontologia servirá para descrever de maneira formal as informações conceituais do domínio, de modo a permitir o reuso e compartilhamento para constituir a base dos modelos conceitual e dimensional de um *data mart*, na arquitetura *bottom-up*, visando a interoperabilidade entre sistemas, para apoio ao processo decisório comercial.

Business Intelligence: data warehousing na prática

As respostas às questões seguintes também ajudarão no desenvolvimento da ontologia:

- A ontologia deverá contemplar que domínio do conhecimento?
- Para qual finalidade será usada a ontologia?
- Quais tipos de questionamentos ou informações a ontologia deve responder?
- Quem usará e manterá a ontologia?

b) Passo 2º - Considerar a reutilização de ontologias existentes

É conveniente considerar e verificar a existência de outras ontologias nas diversas fontes de consultas, este procedimento permite possível reutilização de ontologias disponíveis. Pois pode ser requisito do sistema interagir e colaborar com outras aplicações que utilizam da representação de vocabulário idêntico, como também, aproveitar partes de uma representação já definida.

Há bibliotecas de ontologias reusáveis na *Web* que podem ser consultadas. Os exemplos são: *University Stanford - Protégé*, *DAML Ontology*, *UNSPSC (United Nations Standard Product and Service Codes)*, *RosettaNet*, *Universal Repository*, *Ontolingua Server*.

Para consecução da ontologia proposta, executou-se uma pesquisa nas bibliotecas de ontologias disponíveis na *web*, além do Instituto Brasileiro de Ciência e Tecnologia (IBICT), *Thesaurus*^[1] Brasileiro, do *Thesaurus* da UNESCO, do *Eurovoc Thesaurus* e do *Library and Archives Canada*. E, não foi encontrada nenhuma ontologia que contemplasse o domínio de empresa comercial, até o momento.

c) Passo 3º - Enumerar termos importantes para a ontologia

Ontologia utiliza uma representação formal do conhecimento. Para tanto, é necessário a identificação dos termos que compõem o domínio de interesse, para que sejam formalizados num segundo momento.

Neste estágio, é importante que seja construída uma lista de termos significativos para o entendimento do domínio e esses termos precisam ser especificados como conceitos ou propriedades. Algumas perguntas podem ser respondidas para facilitar este procedimento, são elas:

- Quais os termos são necessários para tratar do assunto?
- Quais propriedades existem nesses termos?
- Que propriedades têm as condições apresentadas pela ontologia?

Elabora-se uma lista ampla de termos significativos relacionados com a área de interesse do domínio e as relações entre os conceitos e suas outras propriedades. É obtida uma lista composta de possíveis termos da nova ontologia, através das diversas fontes.

Business Intelligence: data warehousing na prática

Embora ainda provisório, procura-se determinar a função de cada termo eleito, já sinalizando se o termo é uma classe, uma subclasse, propriedade ou relação. Esta lista deverá ser ordenada e servirá de base para a constituição da hierarquia da ontologia. O Quadro 1 mostra os termos iniciais escolhidos para a ontologia de empresa comercial.

A lista encontra-se ordenada por termos mais significativos, com possibilidade de constituir classes e subclasses e suas propriedades condicionadas, podendo representar atributos numa dependência de hierarquia.

Autorização Fornecimento	Estoque	Material/Serviço
Fornecedor	Entradas	Serviço
Material/Serviços	Recebimento de Material	Mercadoria Comprada
Coleta de Preços	Nota Fiscal	Produto Vendido
Fornecedor	Inspeção Material	Material Estoque
Preço	Saídas	Código Material
Descrição Material/Serviço	Distribuição de Material	Grupo Material
Quantidade Material/Serviço	Saldos	Classe Material
Prazo Entrega	Fornecedor	Família Material
Condições de Pagamento	Pessoa Jurídica	Unidade Física
Comprador	Pessoa Física	Compras
Nome	Finanças	Coleta de Preços
Matrícula	Recebimento	Vendas
Contrato	Modalidades	Material/Serviço
Fornecedor	Execução contrato	Quantidade vendida
Objeto	Fluxo de caixa	Unidade Física
Vigência	Recebimento	Cliente
Valor	Pagamento	Entregas
Obrigações Contratuais	Cobrança	Loja Vendas
Aditivo	Modal	Responsável
Logística	Acordos	Localização
Roteiro de entrega	Crédito	Entrega Delivery
Veículo de carga	Juros de mora	Rota Entrega

Quadro 1 – Exemplo de lista de termos para ontologia de empresa comercial.

Business Intelligence: data warehousing na prática

d) Passo 4º - Definir as classes e a hierarquia das classes

Há várias possibilidades de enfoques para construir uma hierarquia de classes.

No modelo *top-down* o processo tem início com a definição dos conceitos mais genéricos e depois segue para os mais específicos. No desenvolvimento *bottom-up*, toma-se o caminho contrário, começando com a atribuição das classes mais específicas e depois são agrupadas em classes com definições de conceitos mais genéricos. Existe também, uma associação dessas técnicas, onde são definidos os conceitos mais importantes, e então, realiza a generalização e especificação adequadamente, usando os dois modelos.

O modelo escolhido para a ontologia em construção foi o *top-down*, em razão da facilidade de identificação dos termos desse campo de conhecimento, que são muito utilizados na literatura técnica de administração comercial, nos procedimentos operacionais de compras, vendas e estoques, e em geral, são definidos do aspecto mais amplo para o particular.

A partir da lista elaborada no passo anterior, ilustrada na Quadro 8, procurou-se analisar os termos que representaram os objetos e têm existência independente uns dos outros. Esses termos têm a possibilidade de se tornar classes da ontologia e constituem umas das principais referências na hierarquia.

Em um estágio seguinte, foram estruturadas as classes numa taxonomia fazendo o seguinte questionamento:

Sendo a classe A uma superclasse da classe B, então cada instância de B pode conter atributos e métodos de A ?

Alguns termos da lista também foram descartados, quando não representavam a generalidade do conceito de objeto e hierarquia sobre outros objetos, ou representavam alguma propriedade de objetos, ou ainda, eram caracterizados como instâncias de algum objeto. Exemplo de: autorização de fornecimento de material/serviço, contrato de fornecimento, mapa de comparação de preços, fases de compras ou vendas.

Termos da lista que não foram utilizados na determinação de classes, foram atribuídos conceitos como atributos dessas classes e chamados de propriedades, relações ou *slots*, em razão da pouca ou nenhuma representatividade desses termos para definição dos objetos. Por exemplo: preço, prazo de entrega, condições de pagamento, responsável, unidade física, entre outros. Pois não foram considerados significantes para o domínio, porém, a depender da necessidade decisória podem voltar ao modelo. Lembra-se que a matriz de barramento é dinâmica e permite edições.

A Figura 13 ilustra o resultado da definição das classes da proposta de ontologia de compras nas universidades públicas do estado da Bahia.

Business Intelligence: data warehousing na prática

Classes de ontologia

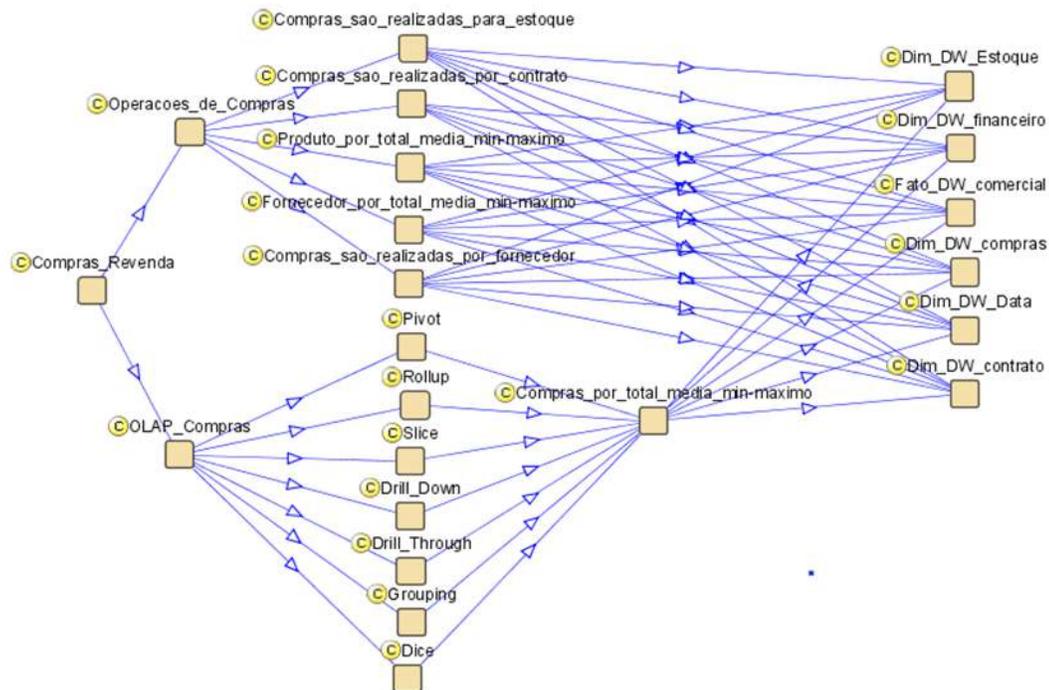


Figura 13 – Classes da ontologia de empresa comercial.

e) Passo 5º - Definir as propriedades das classes

As classes listadas não oferecem informação suficiente para responder aos questionamentos realizados no Passo 1º. Após a definição das classes e subclasses, deve-se descobrir a estrutura interna dos conceitos, os atributos e características que formam as propriedades e suas relações.

Em geral, há vários tipos de propriedades de objetos que podem constituir em propriedades numa ontologia, como atributos intrínsecos, extrínsecos e partes físicas e abstratas, considerando que o objeto é um todo estruturado.

Essas propriedades referem-se aos termos presentes na primeira lista criada anteriormente e que não foram definidas como classes ou subclasses. Assim, essas propriedades podem corresponder tanto aos atributos da classe como ao relacionamento existente entre esta e outras classes.

Todas as subclasses herdam propriedades e relação de sua classe. Por exemplo: as propriedades da classe fornecedor serão herdadas por todas as subclasses de fornecedor, que incluem pessoa jurídica e pessoa física. O Quadro 1 mostra as propriedades resultantes para as classes e subclasses, em ordem alfabética.

Business Intelligence: data warehousing na prática

Aditivo contratual	Email	FinanceiroRecimento	QuantidadeItem
Banco	EnderecoCliente	FinanceiroPagamento	Razao_Social
CaracterizacaoFinanceira	EnderecoFornecedor	InscricaoEstadual	RazaoAditivo
CaracterizacaoContratacao	EspecificacaoProduto	ItemNotaFiscal	Requisitante
Cidade	EspecificacaoItem	JulgamentoPropostas	ResponsavelSetor
Classe	EspecificacaoMaterial	JustificativaEstoque	SaldoAnterior
CNPJ	EspecificacaServico	MatriculaComprador	SaldoAtual
CodigoServico	Estado	NomeComprador	SetorVenda
CodigoMaterial	EstadoFederativo	NumeroEmpenho	SetorFiscalizador
ContratoVendas	FamiliaMaterial	NumeroEntrada	Telefones
ContatoFornecedor	Finalidade	NumeroRequisicao	UnidadeFisica
CPF/CNPJ	FiscalContrato	NumeroSaida	ValorCompraDireta
DataAquisição	FormalizacaoContrato	ObjetoContratual	ValorVendas
DataRequisicao	FornecedorColeta	ObrigacoesContratuais	ValorTotalContrato
DataEntrega	FornecedorContratado	OrcamentoSetor	ValorTotalMovimento
DataVenda	FornecedorCadastro	PrazoVigênciaAditivo	ValorUnitario
DataContrato	GrupoMaterial	QuantidadeVendida	VigenciaContrato

Quadro 1 - Propriedades (*slots*) das classes e subclasses.

f) Passo 6º - Definição das características das propriedades

As propriedades podem ter diferentes características que descrevem o tipo de valor, os valores admitidos, a cardinalidade e outras características:

▪ **Tipo de valor:** é necessário definir os tipos de valores das propriedades de acordo com seu comportamento, ou seja, o tipo de valor que poderá preencher uma determinada propriedade. Exemplo: a descrição de material é uma cadeia de caracteres, podendo aceitar alfanuméricos e outros símbolos, já o preenchimento de um valor total admite-se somente valores numéricos. Os tipos de valores mais comuns são:

- String: valores simples que representam um conjunto de caracteres;
- Número: descrevem valores numéricos, como *float* e *integer*;

Business Intelligence: data warehousing na prática

- Booleano: *flags* simples que podem assumir valores sim/não;
- Enumerado: especifica uma lista de valores admitidos para as propriedades;
- Instância: permite definição de relacionamento entre objetos. As propriedades com tipo de valor instância devem definir uma lista das classes admitidas e quais as instâncias podem se relacionar.
- **Cardinalidade:** determina a quantidade de elementos que uma propriedade pode ter. Os sistemas podem diferenciar somente entre cardinalidade única - que permite no máximo um valor e, cardinalidade múltipla - permite qualquer quantidade de valores. Podendo também ser representada na forma de cardinalidade mínima e máxima, ou seja, a quantidade mínima e máxima de valores permitidos. Uma cardinalidade “N” significa que uma propriedade deve ter pelo menos “N” valores. Por exemplo: cada fornecedor da coleta de preços tem uma cardinalidade mínima de 1 (um), ou seja, cada coleta de preços precisa ao menos um fornecedor.

Uma cardinalidade máxima “M” significa que uma propriedade pode ter até “M” valores. Algumas vezes pode ser útil fixar a máxima cardinalidade no valor 0 (zero). Esta definição indicará que uma propriedade não pode ter nenhum valor particular para uma subclasse específica.

- **Domínio e escopo:** ocorre quando a definição de uma propriedade é associada a várias classes. Estas classes agregadas são chamadas de domínio da propriedade, que se refere ao conjunto de indivíduos ao qual a propriedade está aplicada. Caso a propriedade seja do tipo que permita instâncias como valor, deverá fazer referência a uma outra classe. Assim, todos os valores desta propriedade devem ser instância das classes que estão sendo relacionadas. As classes admitidas para as propriedades do tipo instância são chamadas de escopo de uma propriedade. A Figura 14 mostra como exemplo a composição da propriedade “Especificacao”, com tipo de valor, cardinalidade, domínio e escopo, na tela apropriada do software *Protégé 3.2*, para essa definição.

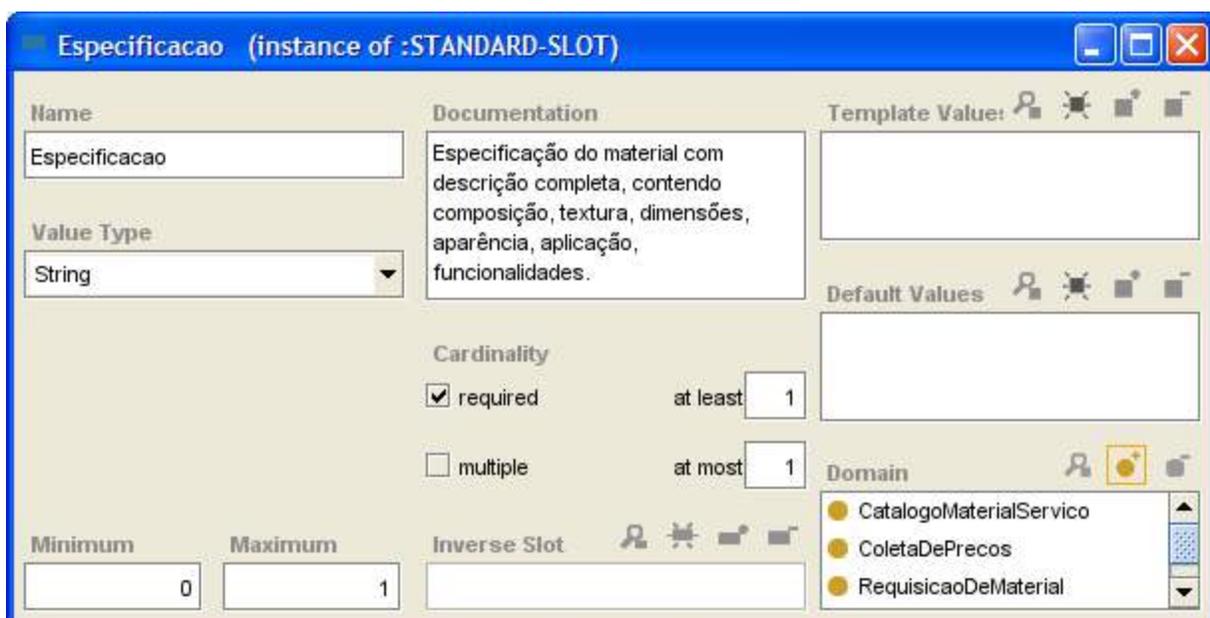


Figura 14 – Propriedade “Especificacao” com as devidas características.

Business Intelligence: data warehousing na prática

g) Passo 7º - Criar instâncias

O último passo na construção da ontologia consiste em criar instâncias individuais das classes na hierarquia, ou seja, vários indivíduos distintos para as classes.

Para definição de uma instância individual de uma classe, requer: eleger uma classe, criar uma instância da classe e preencher os valores das propriedades. Por exemplo: uma instância individual “Nota Fiscal de Saída nº. 10806” representa um tipo específico de vendas em varejo e pedido fidelizado. Esta instância define os seguintes valores: cliente cadastrado, contrato fidelizado de venda, entrega por delivery, fiscalização de contrato, atendimento pós venda, entre outros.

Da Ontologia para a Matriz de Barramento

O dicionário de dados dos BD's, a descrição do domínio organizacional, as regras do negócio e as operações organizacionais formam a base de requisitos para a modelagem conceitual multidimensional. A matriz de barramento é usada para estruturar essa modelagem.

Segundo Kimball et al. (2005), matriz de barramento é uma composição onde os processos do negócio, em cada linha, apontam para as dimensões que os incluem, e que, por sua vez, as dimensões são relacionadas aos processos do negócio que estão em colunas, formando uma intersecção de dependência.

Conhecendo processos de negócio e as dimensões, pode-se chegar aos questionamentos sobre seus dados numéricos, estatísticos, mensurações, volumes e totalizações, definindo-se as entidades fatos.

As entidades fatos constituem as principais tabelas do modelo multidimensional, pois armazenam dados numéricos com relacionamento direto para todas as dimensões, definindo a granularidade dos dados. A granularidade corresponde aos dados atômicos ou as informações que podem ser subdivididas ou integradas.

A dimensão “data”, que representa temporalidade expandida ou contraída, como na ordem semestral, trimestral, mensal, quinzenal, semanal, diária ou, vice-versa, tem particularidade diferente das outras entidades, pois é formada por uma conjunção de valores inerentes às diversas dimensões, podendo representar as ocorrências históricas.

Matriz de barramento dinâmica

O modelo conceitual aplicado neste livro usa uma adaptação da Matriz de Barramento de Kimball et al. (2005). Portanto, para responder aos questionamentos de cada processo organizacional e de negócio são necessárias informações provenientes das dimensões e fatos vinculados a ele, assim como, orientações ao processo decisório das técnicas OLAP que serão utilizadas.

Business Intelligence: data warehousing na prática

Os termos empresariais usados para composição da matriz de barramento, no Quadro 2, são os seguintes: consulta decisória – proposição com dimensões do domínio diante de um questionamento empresarial sobre a operacionalização administrativa ou de produção; operação organizacional – ação departamental na empresa sobre um assunto setorial que pode estruturar um *Data Mart* (DM); processo de negócio – integração entre operações departamentais para alcance dos objetivos empresariais e fonte para as consultas decisórias.

A matriz de barramento é considerada na presente metodologia como dinâmica, pois serve em todo ciclo de vida do DW, e deve ser aplicada na modelagem conceitual multidimensional, na arquitetura do DW, na geração dos cubos dimensionais dos DM's e na orientação das técnicas OLAP. É formada pelos seguintes elementos:

1. Processos de negócios e consultas OLAP: é um grupo de colunas que representa as atividades formadas por operações organizacionais, consultas decisórias, fontes dimensionais e técnicas OLAP.

1) Operações organizacionais: são operações organizacionais específicas ou competências departamentais, referentes às atividades meio e fim nas organizações empresariais, governamentais e outras. Ou seja, são operações diretas, inerentes aos departamentos como produção, comercialização, estoques, logística, recursos humanos e outras, cujas atividades são auxiliadas por setores que compõem a estrutura organizacional de uma instituição e têm suas atribuições definidas internamente.

As operações organizacionais não são consideradas departamentos de uma empresa, não são estruturas ou locais específicos para aglutinar operações afins, mas sim integração de ações na empresa que são complementares e necessárias, podendo estar em departamentos diferentes e, portanto, serem registrados em bases de dados diversos. Por exemplo, pode-se ter departamento de vendas, departamento de logística e departamento financeiro. São setores diferentes, mas suas atividades são complementares: após as vendas, realiza-se entrega do produto e recebe-se o pagamento devido. Isto pode ser representado por apenas um *Data Mart*, se for de conveniência do modelo conceitual-dimensional do DW.

Em geral, as operações organizacionais ou processos de negócios dão origens aos *Data Mart's*, que são orientados por assuntos integrados e não somente departamentais.

2) Consultas decisórias: são orientações relativas aos processos de negócio, que auxiliam nas consultas para tomadas de decisões através das dimensões, fatos e atributos específicos e relacionadas, descritos com aspectos qualitativos ou quantitativos. Servem para auxiliar o tomador de decisão nas formulações de perguntas que serão respondidas pelas técnicas OLAP em conformidade com os cubos dimensionais.

Exemplo: totais de vendas, por região, em determinado período, emitidas por determinado vendedor. Neste caso, são indicadas consultas que empreguem dimensões e fatos e visualizada por técnicas em “cubo” como *dice*, *crossstab*, *drill-through*, *roll-up*, *pivot* e *grouping*;

Business Intelligence: data warehousing na prática

3) Fontes: a matriz de barramento traz orientações para utilização das dimensões ou fatos. A depender da consulta requerida, poderão ser empregadas somente dimensões para respostas qualitativas ou, dimensões e fatos para respostas com mensurações e totalizações. Esta orientação ajuda na construção do cubo dimensional com a indicação de quais dimensões, fatos e atributos farão parte do cubo dimensional para emprego da adequada técnica OLAP;

4) OLAP: é a aplicação da técnica OLAP (*slice, dice, drill-down, drill-through, rollup, pivot, grouping e crosstab*) sugerida para obtenção da resposta requerida de acordo com o processo de negócio e a operação organizacional.

1. **Dimensões e fatos:** representam todas as dimensões e fato do modelo estrela. A escolha das dimensões deve refletir as operações organizacionais representadas nos bancos de dados fonte: gestão de contratos, compras e vendas, e, logística de entregas de mercadorias. Recomenda-se os seguintes passos:

1) **Identificar os processos de negócio.** Para que haja produção empresarial, há um fluxo de informações entre os processos de negócios que ocorrem nos departamentos, os quais estão representados pelos Bancos de Dados. No exemplo deste livro, inicia-se com um processo de **compra**, depois emissão do **contrato de compra**, o **pagamento** dos produtos comprados, e o recebimento em **estoque**, para disponibilização às **vendas**, emissão de contrato de venda, recebimento financeiro e entrega da mercadoria. Por fim, habilita-se um novo processo de compra para ressuprimento. Figura 15.

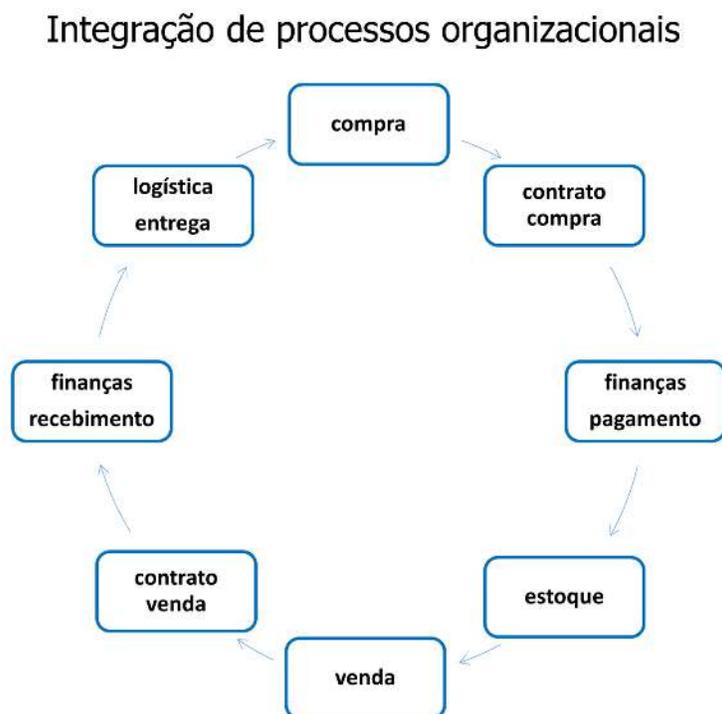


Figura 15 - Integração nos processos organizacionais representada dos BD's departamentais.

Business Intelligence: data warehousing na prática

2) **Identificar a integração organizacional existente entre os BD's de produção.** Como trata-se de uma única empresa os BD's departamentais representam dados integrados que formam um processo empresarial único e cíclico, conforme Figura 16.

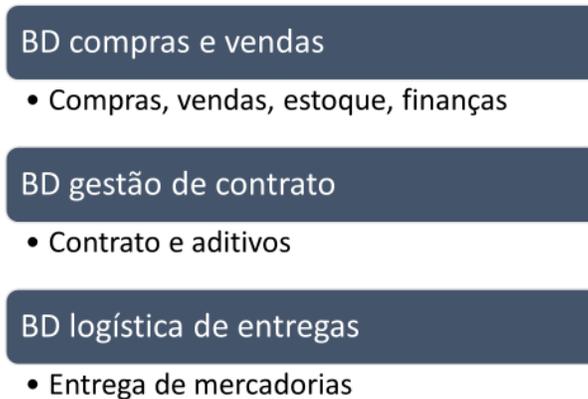


Figura 16 - Integração organizacional originada dos BD's departamentais.

As operações organizacionais da Matriz de Barramento Dinâmica, que estão indicadas a seguir, foram originadas pelos BD's de produção (em parênteses), que representam atividades compartilhadas departamentais:

Processo organizacional	Banco de dados de origem
Compras para revenda	BD compras e vendas (Fig. 6)
Vendas comerciais	BD compras e vendas (Fig. 6)
Produtos estocados	BD compras e vendas (Fig. 6)
Finanças	BD compras e vendas (Fig. 6)
Contratos de compras	BD gestão de contratos (Fig. 5)
Transporte <i>delivery</i>	BD logística de entregas (Fig. 7)

3) **Determinar as dimensões que devem ser originadas das operações organizacionais.** Neste modelo dimensional, por convenção, as tabelas são nomeadas usando-se o prefixo “Dim_dw_” para melhor identificação no esquema estrela e implementação física. Estas tabelas dimensionais são:

Dim_dw_contrato, Dim_dw_compras, Dim_dw_finanças, Dim_dw_estoque, Dim_dw_vendas, Dim_dw_logística, Dim_dw_contrato, Dim_dw_data.

A tabela de dimensão *Dim_dw_data*, é uma tabela específica que armazena a série de tempo do histórico de registros contidos nos sistemas OLTP. Seus atributos são especiais e servem também para auxiliar na granularidade dos DM's. Já que a granularidade definida corresponde a um item de um pedido e este está condicionado à data de aquisição.

Business Intelligence: data warehousing na prática

4) **Determinar a tabela fato:** constitui na principal tabela do modelo dimensional, pois representa a definição de granularidade do DM e tem como objetivo o armazenamento dos dados de medição, dados quantitativos, valorativos e aditivos. Todas as medições em uma tabela de fatos devem estar alinhadas na mesma granularidade. Por convenção deste livro, a tabela de fatos é nomeada pelo prefixo “**Fat_dw**”, para facilitar a *identificação no* esquema estrela. *Sua denominação é Fat_dw_Comercial.*

Business Intelligence: data warehousing na prática

Processos de negócios e consultas OLAP				Dimensões e fato									
Operações organizacionais	Consultas decisórias	Fonte	OLAP	Dim Compras	Dim Contrato	Dim Estoque	Dim Vendas	Dim Financeiro	Dim Transporte	Dim Data	Fat dw_Comercial		
Compras para revenda	Compras são realizadas por fornecedor por: região, cidade, bairro, produto, funcionário, período, contrato.	Busca em dimensões	Dice, Crosstab, Pivot, Grouping, Roll-up	X	X	X		X		X	X		
	Compras são realizadas por contrato por: região, cidade, produto, cliente, fornecedor, período, funcionário.			X	X			X		X	X		
	Compras são realizadas para estoque: por produto, funcionário, cliente, período, contrato.			X	X	X		X	X	X	X		
Contratos de compras	Contratos são emitidos por fornecedores por: região, produto, funcionário, cliente, período, aditivo.			X	X			X		X	X		
Vendas comerciais	Vendas são realizadas para cliente: por região, cidade, produto, funcionário, fornecedor, período, contrato.				Roll-up			X	X	X	X	X	X
	Vendas são realizadas por produto por: região, cliente, funcionário, fornecedor, período, contrato.						X	X	X	X	X	X	X
	Vendas são realizadas por Setor de Vendas por: região, cidade, cliente, vendedor, período, contrato.						X	X	X	X	X	X	X
Transporte delivery	Transporte por: origem, destino, produto, loja, cidade, veículo.							X	X		X	X	X
Compras para revenda	Compras por: total, média, min-máximo, contagem por período, região, fornecedor, clientes, produto, vendedor.			Busca em fatos e dimensões	Dice, Crosstab, Pivot, Grouping, Roll-up, Drill-down, Drill-Through	X	X	X		X		X	X
	Produto por: total, média, min-máximo, contagem por período, região, fornecedor, cliente, vendedor.					X	X	X		X		X	X
	Fornecedor por: total, média, min-máximo, contagem por período, região, produto, cliente, vendedor.	X	X			X		X		X	X		
Vendas comerciais	Vendas por: total, média, min-máximo, contagem por período, região, cidade, clientes, produto, vendedor, loja.							X	X	X	X	X	X
	Produto por: total, média, min-máximo, contagem, região, fornecedor, cliente, vendedor, loja.							X	X	X	X	X	X
	Cliente por: total, média, min-máximo, contagem por período, vendas, região, produto, cliente, vendedor, loja.							X	X	X	X	X	X
Transporte delivery	Transporte por: total, média, min-máximo, contagem por vendas, cliente, região, loja, produto, vendedor.							X	X		X	X	X
Produtos estocados	Estoque por: produto, compras, vendas, fornecedor, região, loja, período, quantidade estoque, total estoque, média, min-máximo, ressuprimento, classificação.					X	X	X	X	X	X	X	X
Finanças	Movimento Financeiro por: pagamento, recebimento, valores em estoque, período, vendas, produtos, lojas, região, cidade.					X	X	X	X	X		X	X

Quadro 2 – Matriz barramento adaptado de Kimball et al. (2005).

Capítulo 5 – Modelos lógico-multidimensional

Integração ontologia – modelo lógico-dimensional

Essa estrutura de modelo lógico é ilustrada como um esquema estrela e contém uma tabela de fatos, várias tabelas de dimensões descritivas e uma tabela dimensão específica para data – originadas da arquitetura da matriz de barramento.

Em seguida, aplicou-se a notação em álgebra relacional e a respectiva matriz de barramento para composição do modelo lógico multidimensional.

Nesse contexto, foram selecionadas as entidades e os atributos influentes ao processo decisório, que formaram as tabelas de dimensões e fatos para cada DM, que depois resultaram a estrutura do DW.

A justificativa para aplicação de ontologia, no processo de modelagem conceitual e multidimensional para DW, é a busca de técnicas para automatizar o processo de reconhecimento do modelo de DW, reconhecimento por máquinas do modelo multidimensional, promover conhecimento aos usuários dos procedimentos de consultas OLAP de forma mais facilitada e prover nos sistemas de informação condições para gerar automaticamente as consultas no DW e processar consultas OLAP de acordo com os questionamentos para o processo decisório.

Essa metodologia, busca dar condições para os usuários, leigos na tecnologia de BI ou DW, possa entender mais facilmente o processo de consultas multidimensionais e construir suas consultas OLAP segundo suas necessidades de informações. Utilizando sua linguagem organizacional ou empresarial.

Isso, busca resolver um problema que existe sobre a dificuldade pelo empresário ou usuário, sobre os procedimentos de BI ou DW, construir cubos de dados, consultas multidimensionais e uso das técnicas OLAP (drill-up, drill-down, slice, dice, roll-up, etc), quando eles geralmente, não tem conhecimento de modelagem relacional, elementos de banco de dados, tabelas, chaves, relacionamentos e outros fatores técnicos.

A resolução deste problema, passa pelo emprego de ontologia, sobre taxonomia de termos do domínio que os usuários estão acostumados a lidar na sua atividade. O uso da ontologia deve facilitar o entendimento do uso da tecnologia, da operação das aplicações e das consultas para responder mais rápido aos questionamentos empresariais.

Notação algébrica para o modelo lógico

Na modelagem lógica adotada foi implementado um esquema ROLAP seguindo a notação primária adaptada do modelo de álgebra de Thomas e Datta (2001). Foram incluídos na notação um cartesiano para junções com chaves substitutas em (R) e um elemento de agrupamento com agregação por função (G), tornando um conjunto sêxtuplo $DW = \{D, M, A, f, R, G\}$, onde os elementos integrados indicam as características de DW. A saber:

Business Intelligence: data warehousing na prática

- a), representam os atributos descritivos do domínio, onde cada d_i é o nome de uma dimensão extraída da relação;
- b), representam os dados numéricos e mensurações do domínio, onde cada f é o nome de um fato extraído da relação;
- c)}, cada f é uma propriedade extraída de uma dimensão ou de um fato da relação;
- d) A função f é mapeada como $f: D \rightarrow A$, i.e., para cada entidade dimensão existe um conjunto de atributos. Cada f é disjuncto para atributos, i.e., $i, j, i \neq j, f(d_i) \cap f(d_j) = \emptyset$;
- e) X , é um conjunto de relacionamentos um-para-muitos entre dimensões e fatos, na seguinte restrição algébrica: $|X| .pk = .fk$, através de chaves substitutas *primary key* (pk) e *foreign key* (fk);
- f) $G = G_1, G_2, \dots, G_m$ $g(l_1, l_2, \dots, l_r)$, é operação de agrupamento com agregação g por função f para soma, contagem, média, máximo ou mínimo sobre dados numéricos em fatos ou atributos de dimensões, a depender do cubo multidimensional empregado.

Assume-se que a visão de um cubo multidimensional é definida na expressão base *Project-Select-Join* (PSJ) da álgebra relacional. Assim, se $\{r_1, r_2, \dots, r_m\}$ são n relações do esquema de DW, V é definido como um cubo para consultas conforme exemplo para *roll-up* a seguir. (técnicas assemelhadas mudam somente a forma de agrupamento, agregação, união ou outros, e não são mostradas por falta de espaço.)

$$V = \pi(\sigma |X| .pk = .fk) (r_1, r_2, \dots, r_m)$$

Em que (π) é a projeção de um conjunto de nomes de atributos, (σ) é uma expressão condicional e $\{r_1, r_2, \dots, r_m\}$ são as relações base.

Como um exemplo ilustrativo, baseado na notação da álgebra multidimensional para modelagem lógica, imagina-se que os usuários precisam conhecer informações sobre compras e seus fornecedores, locais, períodos, quantitativos e valores. Ou seja, dados correspondentes aos fatos $M = \{quantidade_produtos, valor_total_compras\}$ e nas dimensões DATA, COMPRAS e FORNECEDOR.

Em outras palavras, eles estão interessados em questionar: "Quais os totais de compras para um produto em determinado fornecedor, por local, durante certo intervalo de tempo?" (questão que envolve as três dimensões).

A dimensão DATA contém as propriedades dia, mês e ano. A dimensão COMPRAS tem as propriedades fornecedor, grupo de produtos e unidade física. E a dimensão FORNECEDOR tem: razão social, endereço, cidade e estado. Na entidade fato estão os dados numéricos de quantidade de produtos e total de compras. Desta forma, obteve-se:

$A = \{dia, mês, ano, fornecedor, grupo_produto, razão_social, unidade_física, endereço, estado, cidade, quantidade_produtos, valor_compras\}$.

Business Intelligence: data warehousing na prática

Cada entidade é descrita por propriedades específicas e são disjuntas mutuamente. Respeitando a notação citada, o cubo multidimensional ficou assim definido:

$f(\text{DATA}) = \{\text{dia}, \text{mês}, \text{ano}\}$

$f(\text{COMPRAS}) = \{\text{fornecedor}, \text{grupo_produto}, \text{unidade física}\}$

$f(\text{FORNECEDOR}) = \{\text{razão social}, \text{endereço}, \text{estado}, \text{cidade}\}$

$f(\text{FATOS}) = \{\text{quantidade_produtos}, \text{valor_compras}\}$

É possível representar a consulta pela notação algébrica resumida e instrução de SQL seguinte:

$V = \pi(\sigma |X| =)$

and ... (r1, r2, ..., rm)

```
SELECT dia, mes, ano, fornecedor, grupo_produto, unidade_física, razão_social, endereço, estado, cidade,
quantidade_produtos, format(Sum(valor _compras), 2) as Total_Compras FROM dim_data, dim_compras,
dim_fornecedor, fato_dw_comercial WHERE id_data=iddimdata and id_compras=iddimcompras and
id_fornecedor=iddimfornecedor GROUP BY dia, mes, ano, fornecedor, grupo_produto, unidade_física, razão_social,
endereço, cidade, estado, quantidade_produtos WITH ROLLUP
```

Capítulo 6 - Arquitetura para data warehousing

Data mart de compras, vendas e logística e finanças

Os DM's foram estruturados na arquitetura ROLAP. A saber: a) **DM - compras**, b) **DM – vendas e logística**, e c) **DM - finanças**, conforme ilustra a Figura 4. A integração dos DM's foi realizada pela granularidade representada no modelo o qual corresponde a um item de produto, por sempre estar presente no pedido de compra, na nota fiscal de venda, na entrega por *delivery* e no recebimento financeiro.

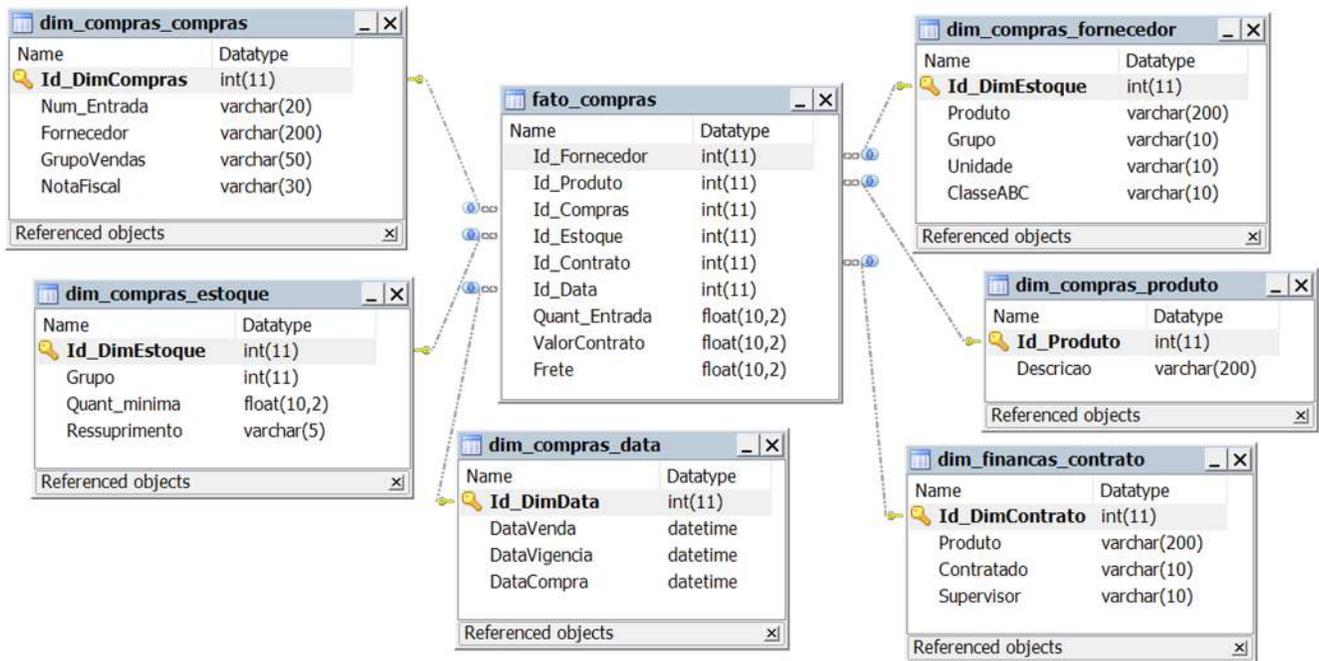
Um item de produto através de seu código permite associar os três BD's em questão, representando o ciclo das atividades empresariais, ou seja: início do processo de compra, emissão do contrato de compra, o pagamento dos produtos comprados e o recebimento em estoque. Depois, a disponibilização do estoque às vendas, que fazem entrega da mercadoria, com o devido recebimento financeiro. E, por fim, habilita-se um novo processo de compra para ressuprimento.

O BD DM_compras é um *data mart* relativo ao domínio das atividades de compras de mercadoria da empresa. Essas mercadorias são recebidas, verificadas, estocadas e colocadas para venda.

As compras são realizadas em um fornecedor, segundo um contrato firmado. As mercadorias são colocadas em estoque e será feito um pagamento ao fornecedor por essa aquisição. A tabela fato_compras do Data Mart de Compras registra movimento de quantidade de entrada em estoque, valores da entrada ou compra e frete pago.

Business Intelligence: data warehousing na prática

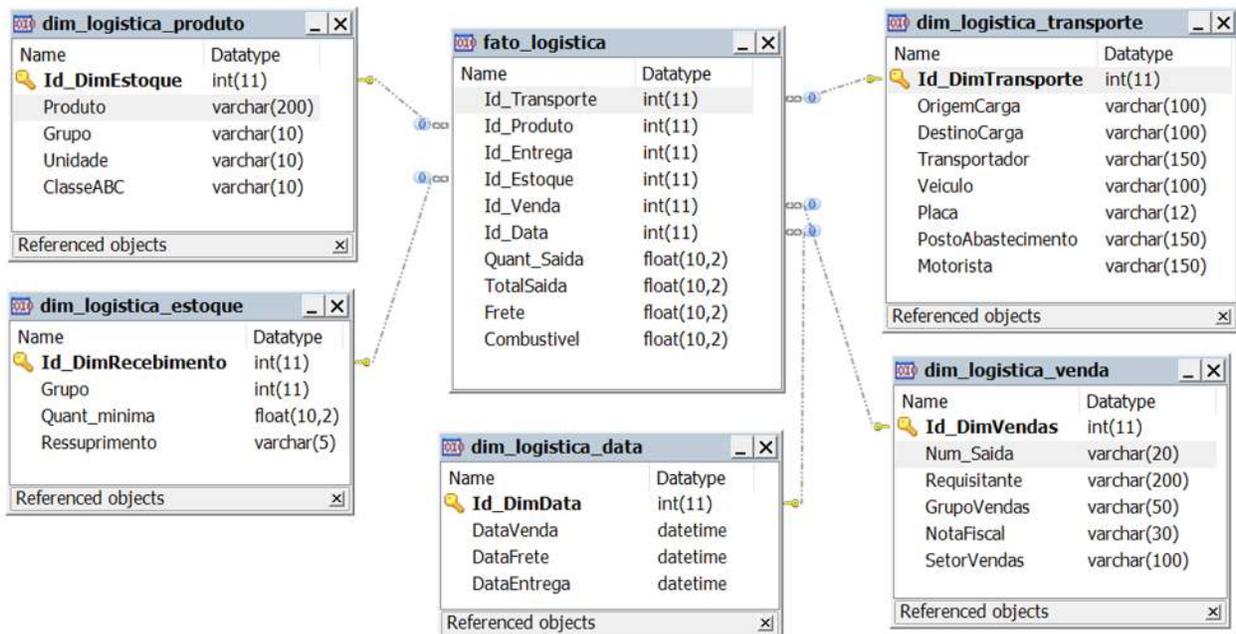
DM_compras



O DM_logistica representa o domínio do estoque com saldo anterior, entrada, saída para venda e saldo atual de cada mercadoria. O serviço de transporte de entregas após vendas. A tabela fato_logistica representa registro quantitativos de saída de vendas, total de vendas, saldo em estoque, quantidade mínima de segurança, ponto de ressuprimento, frete e combustível.

Business Intelligence: data warehousing na prática

DM_logistica



O DM_financeiras é um domínio importante que representa dados de finanças, contratos, recebimentos e pagamentos para toda atividade comercial. Esse domínio faz uma ligação entre as compras e as vendas com suporte de registro para pagamentos e recebimentos. A tabela fato_financeiras registra quantitativo financeiro dos estoques, compromissos contratuais, pagamentos de compras, recebimentos de vendas, modalidades de pagamentos/recebimentos e despesas como frete, combustíveis, pessoal e outras.

Business Intelligence: data warehousing na prática

DM_financas

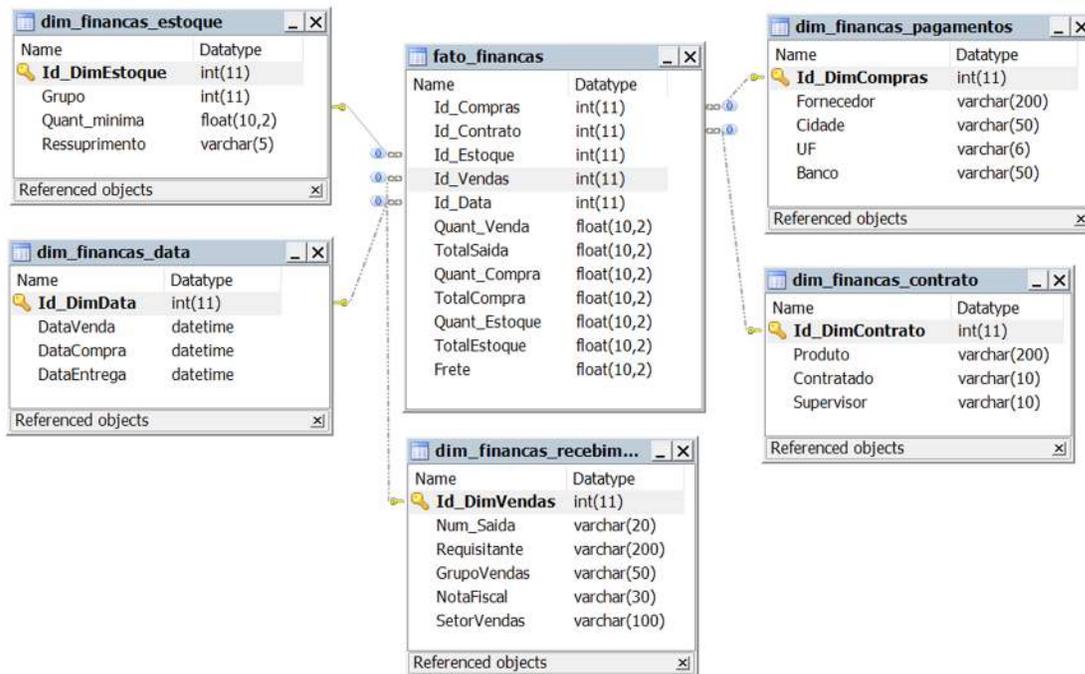


Diagrama de conversão de DM para DW

Após integração de todos os DM's apresentados modelou-se o esquema estrela do DW. Este modelo é completo e capaz de responder às questões gerais de todos os setores, fazendo as combinações necessárias entre o fato e as suas dimensões. O modelo está representado na Figura 18 e serve de suporte geral à empresa onde pode ser fonte maior de informações, pois engloba as áreas envolvidas.

Business Intelligence: data warehousing na prática

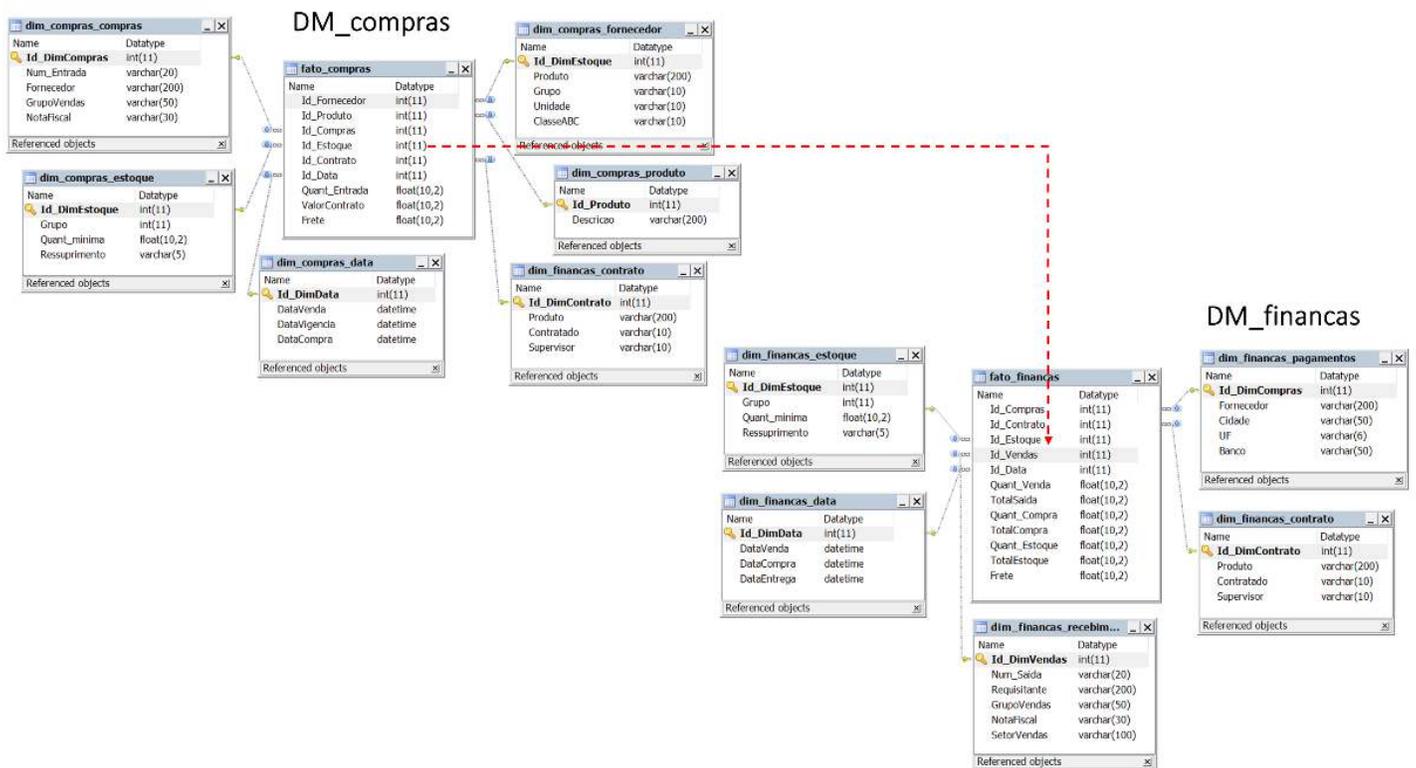


Figura 17. Integração entre DM compras e DM finanças.

Data warehouse central

O DW central após modelagem conceitual e multidimensional é mostrado na Figura 18. Trata-se de um DW no esquema estrela pronto para oferecer condições de consultas para tabelas, gráficos e dashboards para tratamento estatístico e analítico próprio para tomada de decisões.

As fontes originaram das bases de dados de compras, logística e vendas e finanças com aporte importante para decisões estratégicas empresariais de um DW central pronto para se extrair informações para as decisões gerenciais.

O DW está apropriado para reconhecimento em várias ferramentas para visualização de dados, como o exemplo do *MS PowerBi*. Pois esta como outras ferramentas oferecem diversos métodos para consultas de informações com detalhes importantes para *data analytics*.

Business Intelligence: data warehousing na prática

DW – Comercial

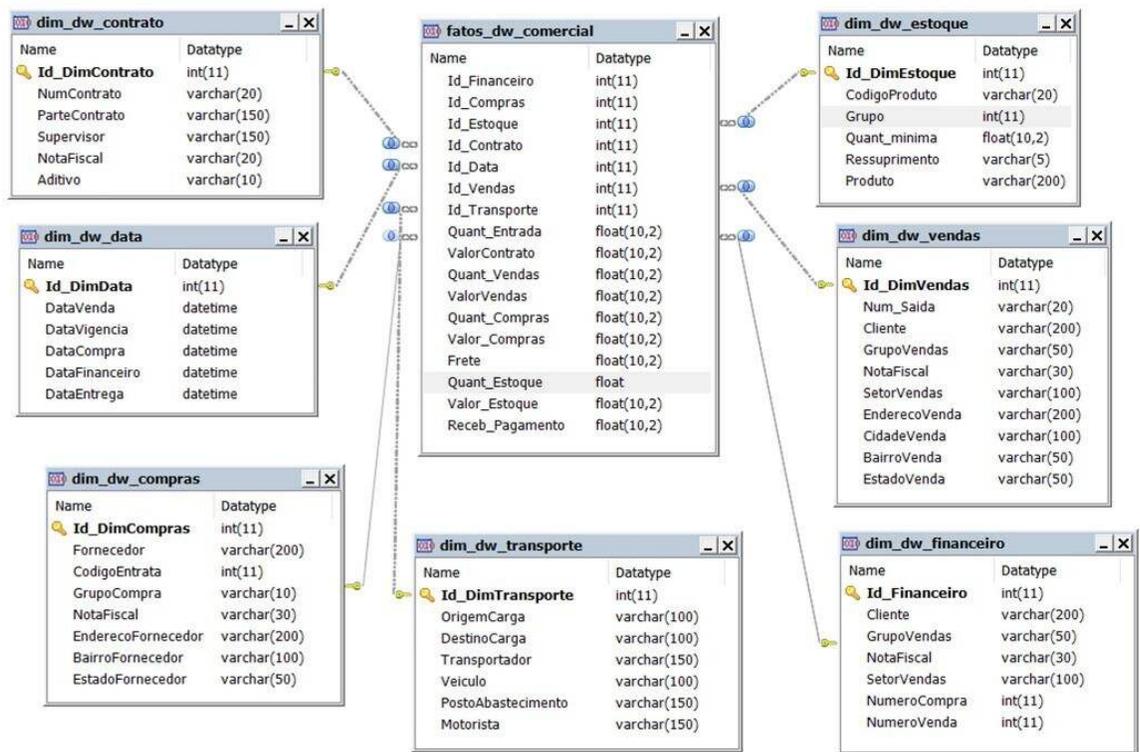


Figura 18. Data Warehouse estruturados para bottom-up.

Considerações finais

Com esse e-book espera-se auxiliar na formação do conhecimento para construção de BI e DW num domínio de empresas comerciais de pequeno porte, com ferramentas de fácil acesso e domínio público.

Através dos registros históricos pode-se fazer planejamento e acompanhamento com mais acurácia das operações comerciais com informações importantes para o processo decisório.

Na prática, trata-se da possibilidade de oferta de uma aplicação tecnológica para a gestão empresarial que pode servir para diversas empresas que enfrentam esse problema de acesso as ferramentas para *data analytics*.

Com este trabalho, espera-se contribuir com o ensino e aprendizagem da modelagem multidimensional em cursos tecnológicos e de bacharelado na área da tecnologia da informação, como também, para startups e empreendedores que projetam *business intelligence* para empresas comerciais.

Referências

BONIFATI, Angela et al. Warehousing workflow data: Challenges and opportunities. In: VLDB. 2001. p. 649-652.

Business Intelligence: data warehousing na prática

Datta, Anindya, and Helen Thomas. "The cube data model: a conceptual model and algebra for on-line analytical processing in data warehouses." *Decision Support Systems* 27.3 (1999): 289-301

Data Mining. A Knowledge Discovery Approach, Krzysztof J. Cios, Witold Pedrycz, Roman W. Swiniarski, Lukasz A. Kurgan. Ed. Springer, 2007.

FERNÁNDEZ-LÓPEZ, Mariano; GÓMEZ-PÉREZ, Asunción; JURISTO JUZGADO, Natalia. Methontology: from ontological art towards ontological engineering. 1997.

GRUBER, Tom R. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*. 1993, 5, p. 199-220. USA.

INMON, William H. Building the data warehouse. John Wiley & Sons, 2005.

KIMBALL, Ralph; ROSS, Margy. The Data Warehouse Toolkit - Guia Completo para Modelagem Dimensional. Tradução da 2ª ed. Rio de Janeiro: Campus, 2002.

MACHADO, F.N.R. Projeto de Data Warehouse: Uma Visão Multidimensional, São Paulo: Érica, 2000

NOY, Natalya F. et al. *Ontology development 101: A guide to creating your first ontology*. 2001.

RUBIN, Daniel L.; NOY, Natalya F.; MUSEN, Mark A. Protege: a tool for managing and using terminology in radiology applications. *Journal of digital imaging*, v. 20, n. Suppl 1, p. 34-46, 2007.

1. O *Thesaurus* é um instrumento que reúne termos escolhidos a partir de uma estrutura conceitual previamente estabelecida e destinados à indexação e à recuperação de documentos e informações num determinado campo do saber. ↑

Business Intelligence: data warehousing na prática

Sobre o autor Sergio Fred Andrade: PhD in Science, Master in Systems and Computing, specialist in Advanced Computing and Distance Education, Bachelor in Information Systems and Mathematics. He has experience in computing, analysis and information systems, database, data science, and software engineering.

