

Uma aplicação de Data Warehouse para apoiar negócios

André Vinicius Gouvêa Monteiro

Marcos Paulo Oliveira Pinto

Rosa Maria E. Moreira da Costa

*Universidade do Estado do Rio de Janeiro - UERJ
IME - Dept de Informática e Ciência da Computação
Rua São Francisco Xavier 524 - 6º andar - Bl. B, CEP 20550-013
Rio de Janeiro - RJ - Brasil*

RESUMO

Este trabalho descreve o processo de construção de uma ferramenta que explora a tecnologia de Data Warehouse. Os dados foram coletados em uma empresa, que é referência na área de manutenção de equipamentos, e padronizados para facilitar a consulta. Este tipo de aplicação pode apoiar a tomada de decisões em empresas de negócios.

ABSTRACT

This work describes the construction process of a tool, which uses the Data Warehouse technology. These data had been collected from a company that is well known in the equipments support area, and had been standardized to facilitate the consultation. This kind of application can support the decisions tasks in business enterprises.

1. Introdução

A partir da década de 1990, com a explosão no uso dos computadores pessoais e o crescimento da Internet, aliados à crescente facilidade de acesso a dados, as organizações começaram a perceber que fatores como diferencial competitivo e ascensão no mercado são obtidos mais facilmente por aqueles que aplicam estratégias de consulta aos dados operacionais.

Tendo em vista a constante necessidade de se analisar grandes volumes de dados em ambiente de produção, desenvolvedores sentiram a necessidade de novas estruturas que suportassem, de forma otimizada, os conceitos de multidimensionalidade de dados e navegabilidade hierárquica facilitada. Neste sentido, o conceito de *Business Intelligence* vem oferecer meios de organizar processos de extração de conhecimento das bases de dados existentes nas organizações, explorando principalmente, o potencial da tecnologia de Data Warehouse.

O processo de construção de um Data Warehouse vem sendo considerado como a melhor abordagem para a transformação das grandes quantidades de dados existentes nas organizações em informação útil e confiável, que atenda ao processo de tomada de decisão.

O objetivo deste trabalho é descrever o processo de construção de um ambiente que explora a tecnologia de Data Warehouse considerando a facilidade de acesso a dados distribuídos, inclusive pela Internet, compondo assim um Sistema de Apoio à Decisão. As informações contidas no Data Warehouse foram coletadas de bases de dados pertencentes a uma empresa que é referência do setor de comercialização de equipamentos e produtos para manutenção. Estes foram padronizados de forma planejada, para que seus usuários obtivessem uma consulta direcionada, composta de fases bem simplificadas.

Para atingir este objetivo, o trabalho está organizado em seções. Esta Introdução; a seção 2 que descreve as principais características envolvidas em *Business Intelligence*; a seção 3 que aborda os elementos que compõem a interface de apresentação dos dados gerados no ambiente de *Business Intelligence*; os aspectos envolvidos na modelagem do Data Warehouse, para apoiar a gerência de uma empresa de comercialização de equipamentos e produtos para manutenção são descritos na seção 4; finalmente, a seção 5 tece as considerações finais e aponta possíveis extensões do trabalho.

2. Business Intelligence (Inteligência Organizacional)

Na economia globalizada a informação adquire um papel muito importante, já que em geral, o mundo ficou mais competitivo, as empresas ficaram mais ágeis e os clientes mais exigentes. Há alguns anos, afirmamos que estávamos entrando na era da informação. Hoje, percebemos que esta previsão vem se confirmando, pois o rápido acesso à informação se tornou imprescindível no contexto empresarial. Utilizar a informação dos clientes ou de processos internos gera vantagens competitivas no atual mercado globalizado. Antecipar possíveis necessidades dos clientes permite oferecer algum tipo de serviço ou produto individualizado, estreitando assim os relacionamentos e criando possíveis vínculos de fidelidade.

Neste contexto, *Business Intelligence* (BI) pode ser definido como um conjunto integrado de conceitos e metodologias, que permite o processo de análise de dados acumulados em uma empresa, a fim de

extrair uma certa inteligência ou conhecimento, apoiando a tomada de decisão com a informação correta e dentro de um prazo adequado [1]. Alguns dos conceitos e tecnologias que compõem esta arquitetura são exibidos na Figura 2.1.

Os produtos de BI existem desde a década de 70 e tinham como característica o uso intenso e a exaustiva programação, com altos custos de implantação. Com o surgimento dos bancos de dados relacionais, dos computadores pessoais e das interfaces gráficas como Windows, OS2, etc, aliados ao aumento da complexidade dos negócios, começaram a surgir os primeiros produtos realmente direcionados aos analistas de negócios.

Em geral, os sistemas de BI oferecem possibilidades de:

- Extrair e integrar dados de múltiplas fontes;
- Fazer uso de experiências prévias;
- Analisar dados contextualizados;
- Trabalhar com hipóteses;
- Procurar relações de causa e efeito;

- Transformar os registros obtidos em informação útil para o conhecimento empresarial.

Considera-se que existem 2 tipos de aplicações para o negócio:

- Aplicações que sustentam o negócio: são as aplicações críticas, ou On-Line Transaction Processing - OLTP, que trazem retorno de investimento baseado no melhor desempenho do negócio;
- Aplicações que analisam o negócio: são as aplicações de apoio à decisão – Decision Support System (DSS). Nelas, o retorno do investimento é baseado no melhor conhecimento do negócio e no desenvolvimento de estratégias refinadas.

Segundo Kimball [2], o poder de soluções de *Business Intelligence* pode proporcionar a organização uma substancial vantagem competitiva em relação à seus competidores. Um pequeno investimento numa solução adequada de BI pode produzir retornos significativos.

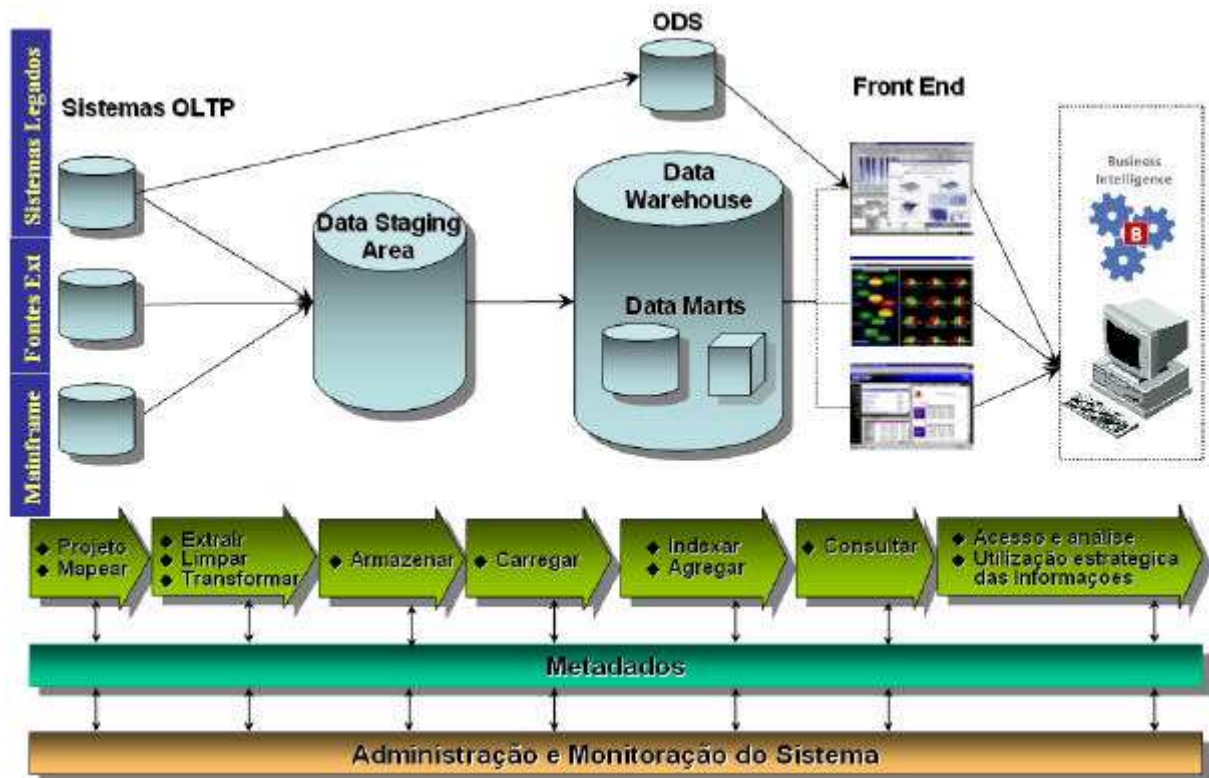


Figura 2.1. Arquitetura para *Business Intelligence* [1].

A seguir são apresentados os principais conceitos que compõem a arquitetura de *Business Intelligence*, dentre os quais destacam-se o OLTP, o ETL e Data Staging Área, além do Data Warehouse [3].

2.2 Sistemas OLTP

Os On-Line Transactions Processing – OLTP (Processamento de Transação On-Line) são sistemas operacionais de registro que capturam as transações

diárias da empresa. Estes são responsáveis pela persistência dos dados que os aplicativos da empresa manipulam. Podemos citar como exemplo: dados do *mainframe*, arquivos, banco de dados e sistemas legados. É destes sistemas que os dados serão retirados a fim de serem transformados em informação para a empresa.

Em geral, os sistemas de origem são projetados com o propósito de estarem disponíveis 100% do tempo

e serem muito rápidos. Se tivéssemos um sistema on-line de vendas e a empresa resolvesse começar a executar consultas imensas e demoradas, o sistema ficaria muito lento e provavelmente, não conseguiria registrar a venda por estar ocupado processando a consulta.

2.3 ETL e Data Staging Area

O processo denominado “Extraction, Transformation and Loading” - ETL (Extração, Transformação e Carga) consiste na primeira etapa do processo de obtenção de dados dos Sistemas OLTP para o ambiente de Data Warehouse.

O processo de extração envolve a leitura e compreensão de dados de origem além de cópia dos dados necessários ao Data Warehouse na Data Staging Area. A Data Staging Area (Área de Representação de Dados) realiza a ligação entre os Sistemas OLTP e o Data Warehouse. No Data Warehouse, dados operacionais brutos são transformados em um formato de *Warehouse* prontos para serem consultados e utilizados pelo usuário.

Enquanto os dados chegam na Data Staging Area, ocorrem muitas transformações, como filtragem dos dados (correções de erros de digitação, solução de conflitos de domínio, tratamento de elementos ausentes ou a divisão em formatos padrão), combinação de dados de várias origens, cancelamento de dados duplicados e atribuição de Chaves de Warehouse. Essas transformações são anteriores ao carregamento dos dados nos Data Marts. A Data Staging Area é dominada pelas atividades simples de classificação e processamento seqüencial. Em muitos casos, ela não se baseia na tecnologia relacional, podendo ser formada por um sistema de arquivos simples.

2.4 Data Warehouse

O Data Warehouse é um repositório de dados provenientes dos dados operacionais (dos Sistemas OLTP), onde se cria um ambiente homogêneo e padronizado, com finalidade de propiciar análises de negócio concentradas em um só local. Com isto, desenvolve-se um ambiente de observação de desempenho, de comparação e análise de dados a fim de extrair informações relevantes ao negócio. Um exemplo seria a comparação e análise da quantidade de novos pedidos de algum produto em relação aos pedidos da semana anterior, analisar por que novos clientes se associaram e por que deixaram de ser clientes da empresa.

Para que o Data Warehouse seja considerado bem sucedido, a comunidade de negócio deve aceitá-lo e ter confiança nas informações geradas a partir dele. E para isto, o Data Warehouse precisa possuir alguns requisitos, dentre os quais destacam-se:

- Deve permitir fácil acesso à informação, com conteúdo compreensível e dados intuitivos e óbvios para o usuário da área de negócios,

além de retornar resultados às consultas no menor intervalo de tempo de espera possível;

- Deve apresentar as informações da empresa de modo consistente, para isso devemos ter dados realmente confiáveis. Portanto, devemos dar uma atenção muito especial à etapa de ETL.
- Deve ser adaptável e flexível a mudanças, às necessidades dos usuários, às condições comerciais vigentes, aos dados e à tecnologia;
- Deve ser um ambiente seguro, pois muitas das informações são confidenciais;
- Deve conter os dados específicos para o domínio contemplado.

Os Data Warehouse não precisam ser construídos de uma vez, pois a complexidade seria elevada, podendo em vez disso, abordar processos de negócios de maneira crescente e assim, serem concebidos aos poucos.

Neste caso, torna-se importante introduzir o conceito de Data Mart, que é um subconjunto lógico e físico da área de apresentação do Data Warehouse. Logo, um Data Warehouse é composto por um conjunto de Data Marts integrados. Originalmente, os Data Marts eram definidos como subconjuntos altamente agregados de dados, normalmente escolhidos para responderem a uma questão de negócio específica. Essa definição não funcionou porque levou a Data Marts isolados, que não eram flexíveis e não podiam ser combinados entre si. Atualmente, o Data Mart é definido como um conjunto flexível de dados, de preferência baseado em dados mais atômicos quanto possível e apresentados em um modelo dimensional, que é mais resistente a consultas *ad hoc* de usuários.

Os Data Mart's podem ser vinculados usando-se técnicas de interligação quando suas dimensões estão em conformidade. Em sua forma mais simples, um Data Mart representa dados de um único processo de negócio.

Em geral, os Data Mart's podem ser construídos através de Bancos de Dados Relacionais ou Bancos de Dados OLAP – Online Analytical Processing (Cubos). As diferenças principais entre ambos residem na forma de armazenagem dos dados e na linguagem de consulta, que no caso dos cubos, não é a SQL e sim, uma linguagem especial para lidar com estruturas multidimensionais.

Ambos os tipos de repositórios apresentam vantagens e desvantagens. Bancos de dados relacionais apresentam flexibilidade e potencial para consultas *ad hoc*, um de seus pontos fortes. São sabidamente mais flexíveis quando são usados com uma estrutura de dados normalizada. Por sua vez, Bancos de Dados OLAP permitem manipular diretamente objetos multidimensionais. Uma típica consulta OLAP requer a integração de diversas relações através de operações de junção. Seu desempenho é melhor para consultas baseadas em chaves do que consultas baseadas em conteúdo.

Para atender aos requisitos deste tipo de transações, fornecedores de SGBD's relacionais têm adicionado funcionalidades a seus produtos, que incluem extensões às estruturas de armazenamento e aos operadores relacionais e esquemas de indexação especializados. Estas técnicas podem melhorar o desempenho para recuperações por conteúdo através da pré-junção de tabelas usando índices, ou pelo uso de listas de índices totalmente invertidas.

2.5 ODS

O Operational Data Store - ODS (Depósito de Dados Operacionais) é implementado para produzir relatórios operacionais, principalmente, quando nem os sistemas legados nem os sistemas de processamento de transações on-line (OLTP) mais modernos os fornecem. Esses relatórios caracterizam-se por um conjunto limitado de consultas fixas que podem estar interligados em uma aplicação de relatório. Os relatórios atendem às exigências de tomada de decisões mais estratégicas da empresa. As agregações de melhoria de desempenho, as seqüências de acontecimentos históricos importantes e a ampla atribuição descritiva são categoricamente excluídas do ODS.

O ODS suporta interações em tempo real, principalmente aplicações de negociação de serviços entre clientes-empresa, onde é necessário acessar os históricos dos relacionamentos.

As informações obtidas no processo de BI precisam ser apresentadas de maneira objetiva e clara. A seguir, são apresentados os principais conceitos explorados para otimizar a apresentação das informações geradas no Data Warehouse.

3. Camada de apresentação de dados – Front-End

A camada de apresentação de dados, também chamada de Front-end, é a interface que apresenta os dados disponíveis ao usuário e é normalmente composta, no ambiente de *Business Intelligence*, por ferramentas de criação de relatórios padronizados, de consultas *ad hoc*, que são as Ferramentas OLAP e de Data Mining.

Cada uma destas tecnologias desempenha um papel fundamental nesta camada. As ferramentas de criação de relatórios padronizados possibilitam a criação de relatórios pré-definidos e que geralmente, são os mais comumente utilizados pelos usuários. As ferramentas de consulta OLAP possibilitam a criação de relatórios personalizados. Já as ferramentas de Data Mining, que são mais complexas, são utilizadas para a descoberta de padrões e relações entre os dados, informações estas, que geralmente não se consegue obter com as outras ferramentas.

O Data Mining é adequado para analisar grupos de dados que são difíceis de analisar usando apenas a função OLAP, visto que esses grupos são grandes demais para serem 'navegados', ou explorados manualmente, ou ainda porque contêm dados muito densos ou não-intuitivos para serem compreendidos.

Assim, a diferença básica entre OLAP e Data Mining está na maneira como a exploração dos dados é realizada. Na análise OLAP, a exploração é feita através da verificação, isto é, o analista conhece a questão, elabora uma hipótese e utiliza a ferramenta para refutá-la ou confirmá-la. Com o Data Mining, a questão é total ou parcialmente desconhecida e a ferramenta é utilizada para a busca de conhecimento.

4. Modelando um Data Warehouse

O sucesso do desenvolvimento de um Data Warehouse depende da escolha correta da estratégia a ser adotada, de forma que seja adequada às características e necessidades específicas do ambiente onde será implementado. Existe uma variedade de abordagens para o desenvolvimento de Data Warehouses, devendo-se fazer uma escolha fundamentada em pelo menos três dimensões: escopo do Data Warehouse (departamental, empresarial, etc), grau de redundância de dados e tipo de usuário alvo.

O escopo de um Data Warehouse pode ser tão amplo quanto aquele que inclui todo o conjunto de informações de uma empresa ou tão restrito quanto um Data Warehouse pessoal de um único gerente. Quanto maior o escopo, mais valor o Data Warehouse tem para a empresa e mais cara e trabalhosa será a sua criação e manutenção. Por isso, muitas empresas tendem a começar com um ambiente departamental e só após obter um retorno de seus usuários, expandir para outras áreas.

Os padrões de uso de um Data Warehouse também constituem um fator importante na escolha de alternativas para o ambiente. Relatórios e consultas pré-estruturadas podem satisfazer o usuário final e geram pouca demanda sobre o SGBD e sobre o ambiente servidor. Análises complexas, por sua vez, típicas de ambientes de suporte à decisão, exigem um padrão de desempenho mais elevado. Ambientes dinâmicos, com necessidade de constantes mudanças são mais bem atendidos por uma arquitetura simples e de fácil alteração, ao invés de uma estrutura mais complexa, que precise de reconstrução a cada mudança. A freqüência da necessidade de atualização também é determinante: grandes volumes de dados, que são atualizados em intervalos regulares, favorecem uma arquitetura centralizada [4].

4.1 Data Warehouse para a área de comercialização de equipamentos e produtos para manutenção

A seguir, são descritos os passos necessários para o projeto e implementação do Data Warehouse proposto. A metodologia utilizada para a modelagem é o processo de criação dimensional em quatro etapas proposto por Ralph Kimball [2].

4.1.1 O processo de extração, transferência e carga dos dados - ETL entre os Sistemas OLTP e a Data Staging Area

Esta etapa iniciou-se através da análise detalhada do modelo de dados dos Sistemas OLTP, a fim de seleccionar um subconjunto que possibilitasse originar o modelo de dados da Data Staging Area, contendo somente dados realmente necessários para a criação do

Data Mart de Vendas. O modelo de dados da Data Staging Area foi gerado a partir de um subconjunto do modelo de dados do Sistema OLTP (Figura 4.1).

Todo o armazenamento dos dados foi feito no Microsoft® SQL Server™ 2000 e a ferramenta utilizada para realizar toda automatização da etapa de ETL foi O DTS (Data Transformation Services), que compõe a suíte do Microsoft® SQL Server™ 2000.

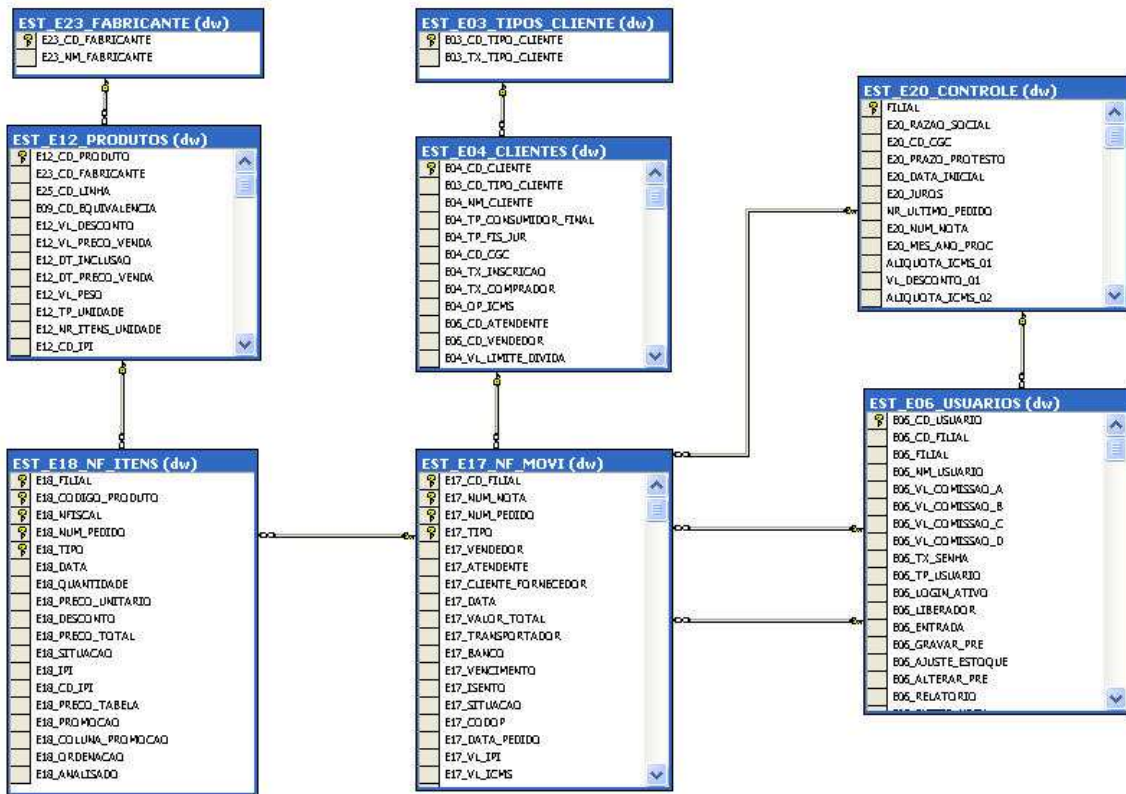


Figura 4.1 - Modelo de dados da Data Staging Area e de parte do Modelo dos Sistemas OLTP.

4.1.2 Definição do grão da Tabela de Fatos

Com a análise do Modelo de Dados da Data Staging Área foi definida a granularidade (grão) da tabela de fatos. Um registro na tabela de fatos representa a quantidade vendida de um produto e seu valor total. As

tabelas de Dimensão foram definidas com base na granularidade da tabela de fatos, e são elas: Dimensão Vendedor, Dimensão Cliente, Dimensão Tempo, Dimensão Filial, Dimensão Fabricante (Figura 4.2).

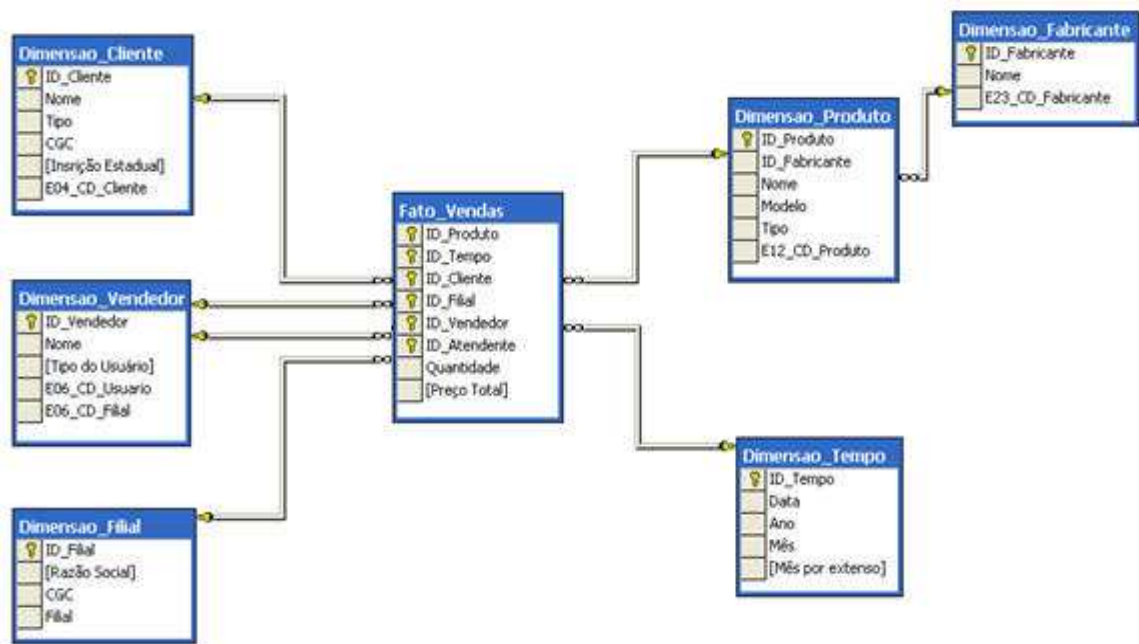


Figura 4.2 - Modelo Dimensional

4.1.3 ETL da Data Staging Area para o Data Warehouse

A extração, transformação e carregamento dos dados (ETL) a partir da Data Staging Area para o Modelo Dimensional foram efetuadas a partir de inúmeras transformações, como filtragem dos dados (correções de erros de digitação, solução de conflitos de domínio, tratamento de elementos ausentes, registros inválidos), combinação de dados de várias origens, cancelamento de dados duplicados e atribuição de chaves de Warehouse.

4.1.4 Os Data Mart's

Neste estágio o Data Warehouse já possui um Data Mart de Vendas completamente funcional, construído com base em tecnologia de Banco de Dados Relacional e que já pode ser acessado por ferramentas de consultas e de geração de relatórios.

A fim de prover uma alternativa aos Data Mart's baseados em tecnologia de Banco de Dados Relacional foi construído um outro Data Mart, também

chamado Vendas, mas só que baseado em tecnologia de Banco de Dados OLAP (Cubo).

Consultas OLAP são muito mais rápidas e permitem maior interatividade dos usuários, se comparado às consultas Relacionais. Cubos OLAP armazenam vários níveis de dados sumarizados em estruturas de dados altamente otimizadas para consultas de usuários. Podem realizar cálculos complexos, que normalmente não são possíveis em Bancos de Dados Relacionais por causa de suas limitações analíticas.

OLAP tem se mostrado um modelo mais intuitivo para o usuário, pois provê uma visão multidimensional dos dados, estendendo-as além das análises de duas dimensões padrão. Os usuários podem facilmente detalhar ou generalizar através dos vários níveis de dados sumarizados, trocar as linhas pelas colunas, mudando assim a orientação da visão dos dados.

4.1.4.1 Criação do Cubo Vendas

A criação do Cubo Vendas foi realizada utilizando como fonte o Data Mart Vendas, criado anteriormente, Figura 4.3.

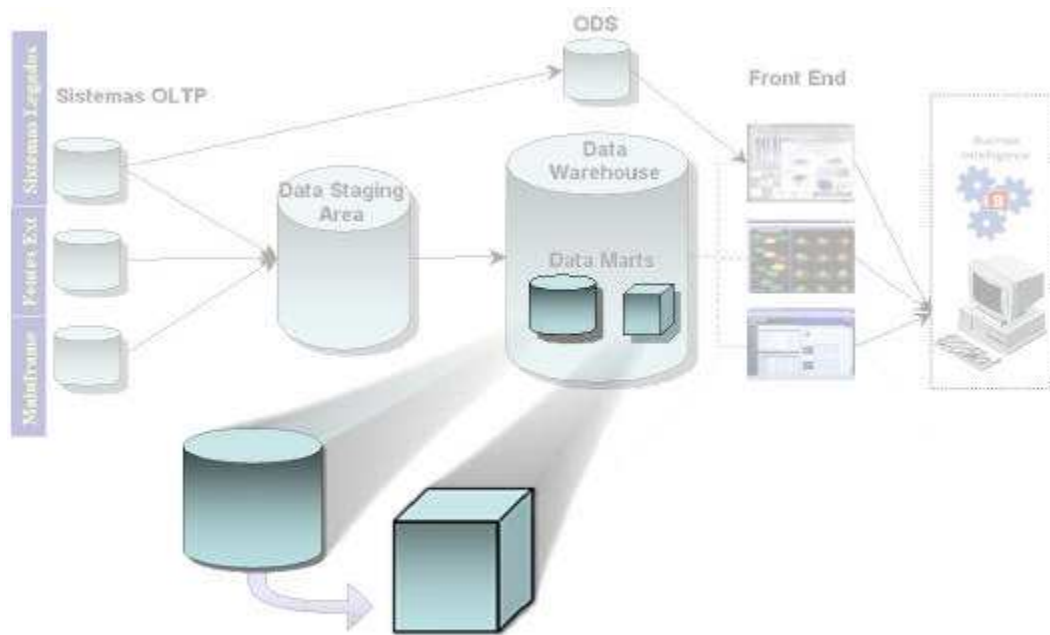


Figura 4.3 Criação do Cubo Vendas

As dimensões do Data Mart Vendas (Relacional) foram preservadas no cubo com exceção da Dimensão_Vendedor, que foi dividida em duas outras (Vendedor e Atendente) com intuito de permitir um filtro por vendedor e atendente quando se estiver acessando o cubo.

A criação do cubo se iniciou pelas Dimensões, em seguida seus níveis e por último com a definição da Tabela de Fatos. No canto esquerdo da Figura 4.4 estão listados os Fatos e as Dimensões do cubo com seus respectivos níveis. À direita está o esquema do cubo, com as tabelas que dão origem às Dimensões e Fatos.

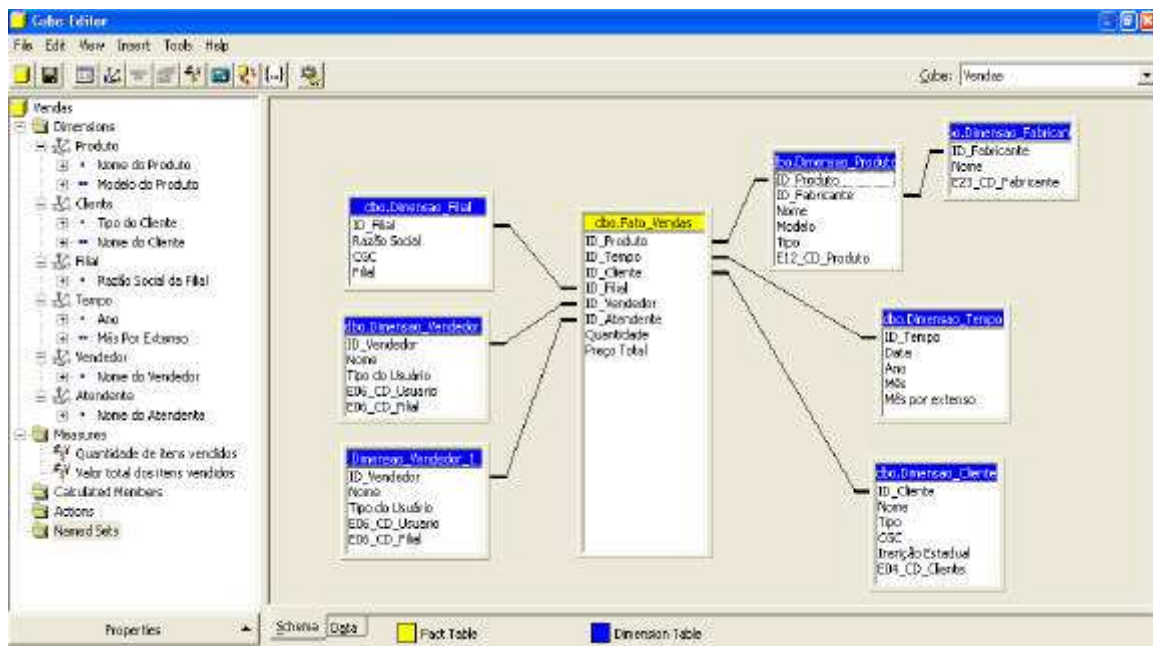


Figura 4.4 Cubo Vendas

4.1.4.2 Armazenamento de dados no Analysis Services

Após a construção do cubo é necessário processá-lo para que possa se conectar aos dados de origem e finalmente ser consultado. Mas antes disso, é importante escolher a forma como os dados serão armazenados.

O Microsoft® SQL Server™ 2000 Analysis Services suporta três formas de armazenamento:

- OLAP Multidimensional (MOLAP): Dados detalhados e agregações são armazenados em formato de cubo do Analysis Server;
- OLAP Relacional (ROLAP): Os dados detalhados e agregações são armazenados no Banco de Dados Relacional. Os únicos dados armazenados no Analysis Server são as hierarquias dimensionais;
- OLAP Híbrido (HOLAP): Os dados detalhados são armazenados no Banco de Dados e as agregações são armazenadas em formato multidimensional.

No Cubo Vendas, foi utilizada a forma de armazenamento MOLAP, devido aos seguintes fatores:

- Armazenamento em disco não era problema;

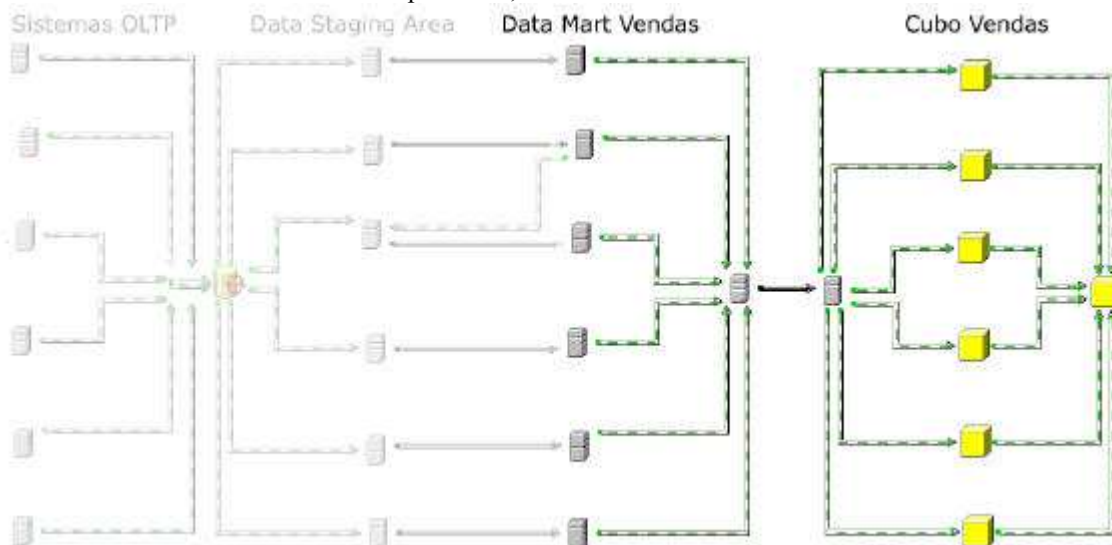


Figura 4.5 – Automatização de geração do Cubo Vendas a partir do Data Mart Vendas

4.1.6 Front-End – Camada de acesso ao Cubo Vendas

Com a conclusão de todo o processo, desde a carga na Data Staging Area até o processamento do Cubo Vendas, já é possível acessá-lo por meio de ferramentas. Neste sentido, foram utilizados o Excel e a Ferramenta XX como interface de acesso ao cubo. Esta ferramenta, que está em fase final de implementação, é uma aplicação que permite aos usuários visualizar dados de cubos do Analysis Services através de um

- Buscava-se a melhor performance em termos de rapidez nas consultas para usuários;
- Os Bancos de Dados de origem deveriam somente ser acessados no período de 01:00 às 04:00h, já que o movimento era reduzido neste horário, permitindo assim que o ambiente de produção não ficasse sobrecarregado.

4.1.5 Automatização de geração do Cubo Vendas

A automatização representa uma redução dos esforços para se manter um Data Warehouse, apesar da necessidade constante de manutenção e desenvolvimento.

Neste caso, foi elaborado um processo de automatização da geração do cubo, com a utilização do DTS (Data Transformation Services), que compõe a suíte do Microsoft® SQL Server™ 2000.

A parte de ETL já havia sido toda automatizada anteriormente e foi incluído mais este passo no processo de geração automática do ambiente. Além do ETL para a Data Staging Area e para o Data Mart Vendas, foi inserido o processo de geração automática do Cubo Vendas, como mostra a Figura 4.5.

navegador, não requerendo para isso componente(s) do Analysis Services instalado(s) na máquina do cliente.

A seguir são apresentados os resultados de algumas consultas, exibindo a interface implementada. As consultas submetidas ao Cubo Vendas através destas duas aplicações podem ser visualizadas, nas Figuras 4.6 e 4.7 demonstrando o acesso através do Excel e as duas seguintes, as Figuras 4.8 e 4.9, demonstram o acesso através da Ferramenta XX.

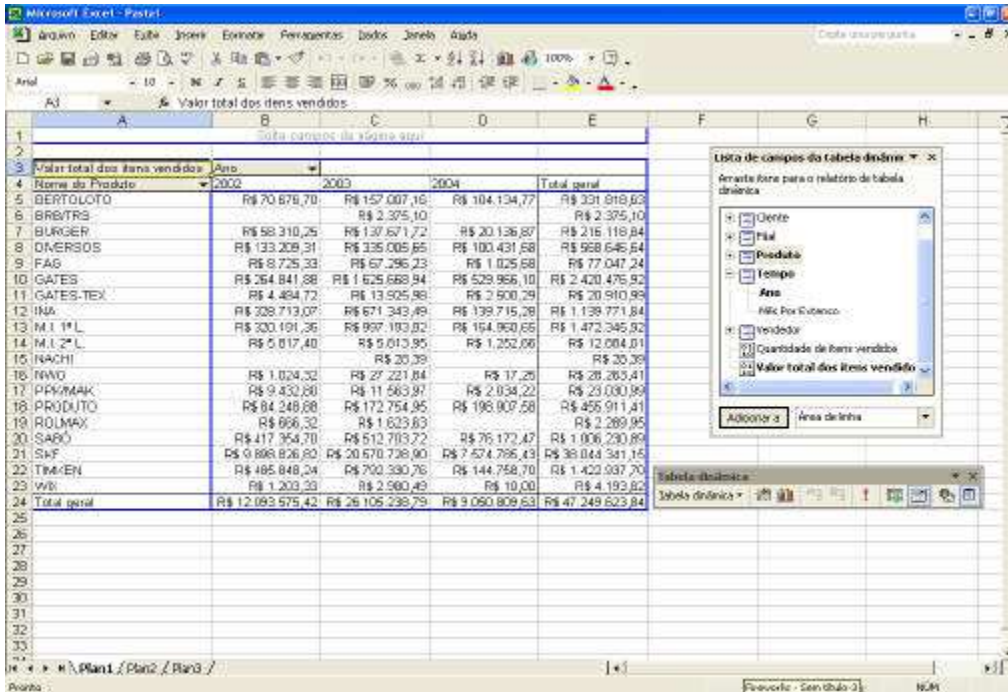


Figura 4.6 – Consulta Valor total das vendas por ano e produto

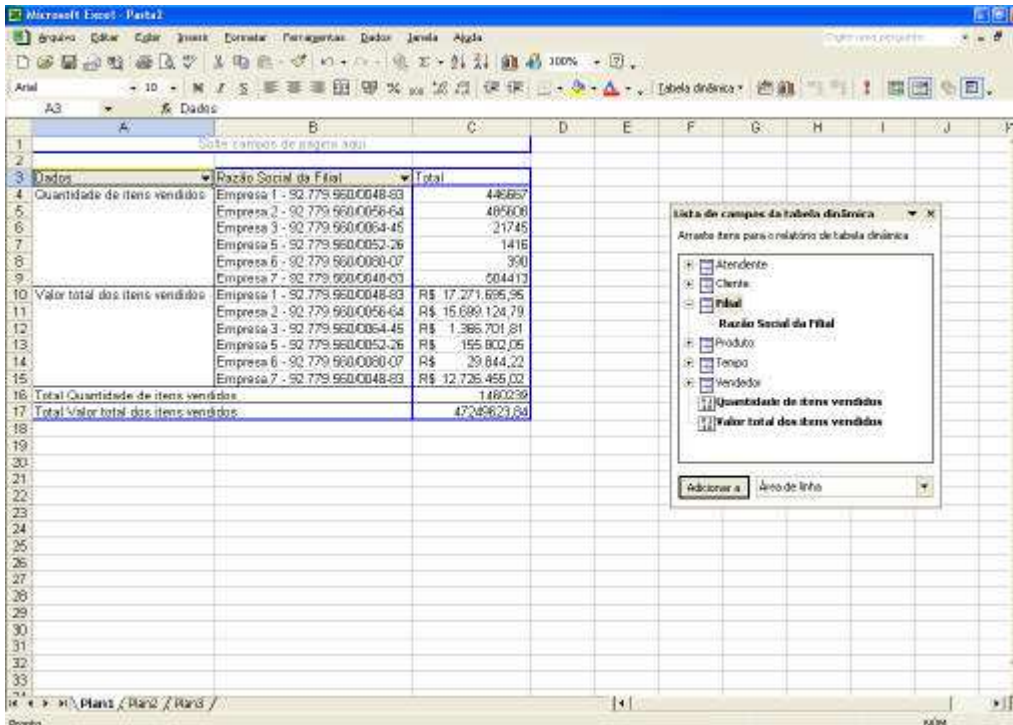


Figura 4.7 – Consulta Valor total das vendas e Quantidade de itens vendidos por filial

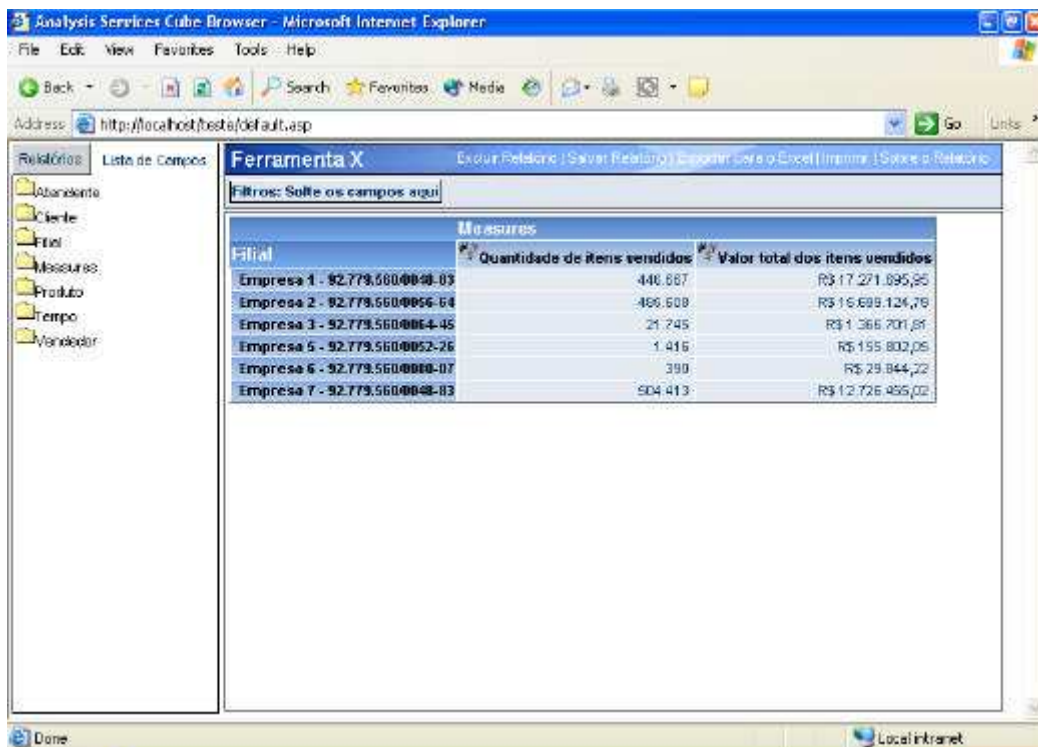


Figura 4.8 – Consulta Valor total das vendas e Quantidade de itens vendidos por filial

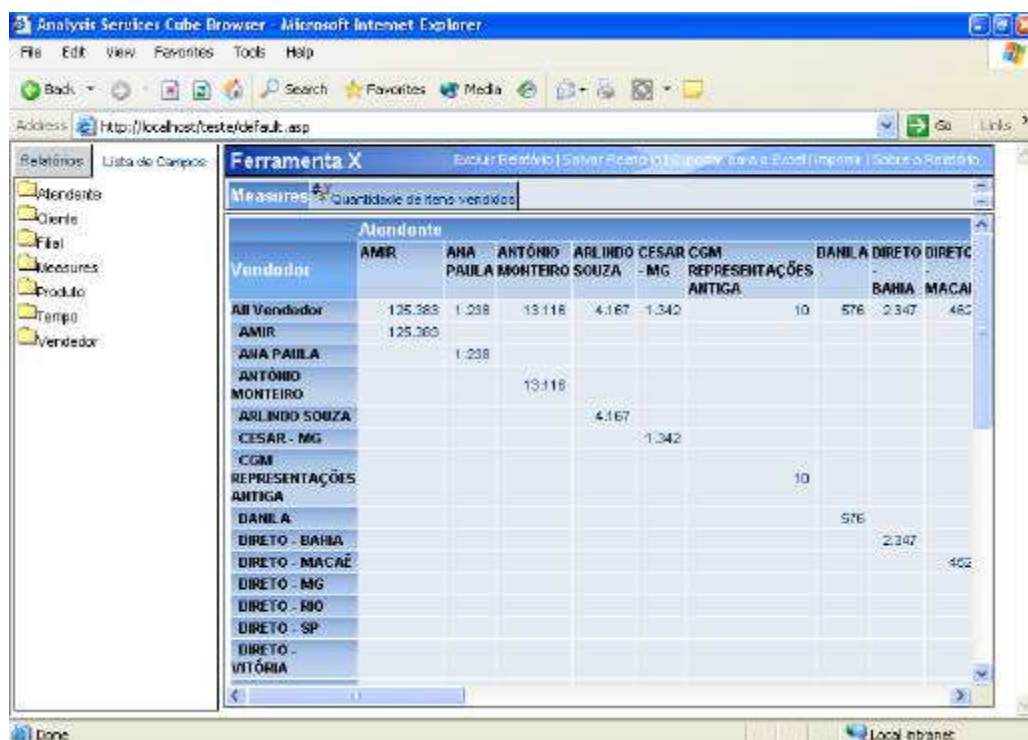


Figura 4.9 – Consulta Quantidade de itens vendidos por vendedor e por atendente

5. Conclusões

Atualmente, as organizações devem possuir meios de reunir informações relevantes sobre seus clientes e

processos para que possam sobreviver em um mercado globalizado, com níveis crescentes de concorrência.

A exploração de Data Warehouses se mostra de extrema importância ao suporte de *Business*

Intelligence, independente de assunto ou área em questão. Pode-se dizer que *Business Intelligence* é o futuro para as organizações que desejam além de informação, conhecimento.

Entretanto, na construção dos Data Warehouse, alguns aspectos devem ser considerados, para minimizar as dificuldades inerentes à construção destas aplicações. Neste caso, confirmando as previsões, a parte de ETL demandou cerca de 70% do esforço de desenvolvimento. As consultas SQL se tornaram complexas devido ao aumento de verificações necessárias a fim de garantir a integridade dos dados que compuseram o Data Staging Área. Estas verificações poderiam ser menores, caso os sistemas OLTP fossem criados baseados em técnicas de modelagem de banco de dados relacionais, através de estruturas normalizadas.

Visando ampliar as possibilidades de consulta, alguns processos de Data Mining devem integrar os Data Marts, apoiando a descoberta de relações entre as informações, que enriqueceria as análises dos negócios.

Espera-se que este trabalho sirva de exemplo para apoiar o desenvolvimento de outros ambientes de Suporte à Decisão apoiados na tecnologia de Data Warehouse

Referências

- [1] THONSEM, Erik. OLAP: Construindo sistemas de informações multidimensionais. Tradução da 2ª edição original. Rio de Janeiro: Campus, 2002.
- [2] KIMBALL, Ralph; ROSS, Margy. The Data Warehouse Toolkit. Tradução da 2ª edição original. Rio de Janeiro: Campus, 2002.
- [3] SERRA, Laércio. A essência do *Business Intelligence*. São Paulo: Berkeley Brasil, 2002.
- [4] WILBERT, Marta; MACEDO, Mauricio; PALMA, Samir; Inteligência na modelagem de dados; Monografia de final de curso de graduação - IME/UERJ, 2003.