

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
2005.1



ANÁLISE, PROJETO E IMPLEMENTAÇÃO DE
UM DATA MART PARA AMBIENTES
DE BUSINESS INTELLIGENCE

TRABALHO DE GRADUAÇÃO
EM BANCO DE DADOS

Autor: Tiago Marafante Lins de Souza
Orientadora: Ana Carolina Salgado

(tmls@cin.ufpe.br)
(acs@cin.ufpe.br)

Agradecimentos

A Deus, inteligência suprema e causa primária de todas as coisas, por me abençoar e permitir que eu me formasse e, desta maneira, sentir que nessa existência subi um pequeno degrau na escalada da evolução intelectual.

A meu avô Jessé, nobre batalhador e grande empreendedor, o qual ajudou bastante meus estudos. Ele é simplesmente a pessoa que mais admiro na minha vida!

A minha amorosa mãe, meu pai e irmãos, pela compreensão, apoio e incentivo durante o decorrer de meus estudos.

A Alexandre Marafante, por ter sido e continuar sendo um excelente primo, amigo, sócio e mentor intelectual para a minha pessoa. Devo muito do que aprendi hoje a ele. Espero ter contribuído, estar contribuindo, e continuar contribuindo ainda mais para nosso sucesso intelectual, moral e profissional.

A meus amigos de faculdade e irmãos de consideração: Heitor Vital, Leandro Marques e Victor Wanderley, com os quais convivi em vários momentos durante a graduação, como estudos pra lista de exercícios, provas, projetos e até mesmo esta monografia, e hoje estão todos presentes em importante lugar em meu coração.

A minha namorada, pela companhia em momentos bons de lazer, namoro e divertimento e pela compreensão de minhas ausências até mesmo em fins de semana.

Aos professores do Centro de Informática, pelo empenho e qualidade no ato de ensinar, abrindo-nos as portas do amplo conhecimento em Ciência da Computação e aos demais funcionários por manter o CIn em funcionamento. A professora Carol, em especial, por ter me orientado com pleno profissionalismo e objetividade.

Resumo

Nas empresas que têm por principal finalidade a obtenção de lucro, a tomada de decisões com rapidez e qualidade representa o principal desafio experimentado pela pessoa responsável pela gestão do negócio. Para superar este desafio, existe a necessidade de tais empresas otimizarem seus processos de tomada de decisão. Essa necessidade pode ser justificada por tais otimizações favorecerem reduções de custo ou elevações de receita com aumento de lucro.

Como uma das respostas da Tecnologia da Informação para solucionar tais necessidades se baseia nos conceitos de *Business Intelligence* e *Data Marts*, este trabalho tem o objetivo de, via pesquisa, entender o conceito, as características e o estado da arte em *Business Intelligence*. Além disso, visa compreender as relações entre *BI* e *Data Marts* para, em seguida, aplicar tal conhecimento na construção de um *Data Mart* para *Business Intelligence* obtendo-se como resultado um ambiente de BI com conhecimento capaz de auxiliar a empresa a tomar decisões de negócios com mais garantia e sucesso.

O processo adotado para atingir o objetivo deste trabalho se divide em duas fases. Na primeira, será realizada uma pesquisa sobre o estado da arte em *Business Intelligence*. Na segunda, será feito um levantamento e análise de requisitos junto aos usuários, um projeto lógico (a modelagem) do *Data Mart*, a implementação de tal *Data Mart* num determinado sistema gerenciador de banco de dados e, por fim, serão pesquisadas, analisadas, preparadas e experimentadas ferramentas de BI que acessarão e analisarão as informações de negócio armazenadas no *Data Mart* implementado.

Índice

| | |
|--|----|
| Glossário | 7 |
| 1. Introdução | 8 |
| 1.1 Objetivos e Contexto | 9 |
| 1.2 Estrutura do Trabalho | 10 |
| 2. Principais Conceitos | 12 |
| 2.1 Definições | 12 |
| 2.2 O Problema | 17 |
| 3. Estado da Arte em <i>Business Intelligence</i> | 19 |
| 3.1 Histórico de BI | 19 |
| 3.2 Relações entre <i>Business Intelligence</i> e <i>Data Mart</i> | 20 |
| 3.3 O Estado da Arte | 21 |
| 4. Especificação do Projeto | 25 |
| 4.1 Principais Decisões de Projeto | 25 |
| 4.2 Usuários | 27 |
| 4.3 Requisitos e Escopo | 28 |
| 4.4 Arquitetura Tecnológica e Informacional do Ambiente | 30 |
| 4.5 Projeto lógico do <i>Data Mart Faturas</i> | 31 |
| 5. Implementação do <i>Data Mart</i> | 33 |
| 5.1 Principais Decisões Técnicas Sobre a Implementação | 33 |
| 5.2 Detalhamento do <i>Data Mart Faturas</i> | 35 |
| 5.3 Análise e Experimentação de Interfaces de Acesso e Análise de Dados no <i>Data Mart Faturas</i> | 41 |
| 6. Conclusões e Trabalhos Futuros | 43 |
| 7. Referências | 46 |
| Apêndice A – Código SQL do processo ETL do <i>Data Mart Faturas</i> | 50 |
| Apêndice B – Exemplo de Ferramentas de Acesso e Análise de Dados em <i>Data Warehouse</i> (ou <i>Data Mart</i>) | 52 |
| 8. Datas e Assinaturas | 54 |

Índice de Figuras

| | |
|--|----|
| Figura 1 – O processo de <i>Data Warehousing</i> . _____ | 15 |
| Figura 2 – Cubo multidimensional. _____ | 16 |
| Figura 3 – Um framework para <i>BI Value</i> [17]. _____ | 22 |
| Figura 4 – Abordagem <i>bottom-up</i> para construção de <i>Data Warehouse</i> . _____ | 26 |
| Figura 5 – Arquitetura do Ambiente. _____ | 30 |
| Figura 6 – Esquema estrela do <i>Data Mart Faturas</i> . _____ | 32 |
| Figura 7 – Etapa de Extração do processo ETL para o <i>Data Mart Faturas</i> . _____ | 38 |
| Figura 8 – <i>Screenshot</i> dos dados do <i>Data Mart Faturas</i> . _____ | 40 |
| Figura 9 – Exemplo de gráfico de uma das ferramenta de BI experimentadas. _____ | 42 |
| Figura 10 – Ferramenta OLAP <i>Cube Browser</i> experimentada no trabalho. _____ | 52 |
| Figura 11 – Aplicação de <i>Web-BI</i> construída no trabalho. _____ | 53 |

Índice de Tabelas

| | |
|--|----|
| Tabela 1 - Análise do Problema..... | 17 |
| Tabela 2 – Principais <i>softwares</i> de BI no mercado. | 24 |
| Tabela 3 – Diferenças entre sistemas transacionais e ambientes de suporte à decisão. | 27 |
| Tabela 4 – Principais requisitos do negócio. | 29 |
| Tabela 5 – Etapas de especificação do <i>Data Mart Faturas</i> | 31 |
| Tabela 6 – Detalhes das tabelas de dimensão do <i>Data Mart Faturas</i> | 35 |
| Tabela 7 – Detalhes da tabela de fatos do <i>Data Mart Faturas</i> | 37 |
| Tabela 8 – Análise de Interfaces de acesso e análise de dados no <i>Data Mart Faturas</i> | 41 |

Glossário

Benchmarking – Comparação (seletiva), referência, parâmetro comparativo, referencial de excelência.

Centro de competência (*competency center*) – Área essencial e estratégica da empresa.

DM – *Data Mart*.

DW – *Data Warehouse*.

IRPJ – Imposto de Renda das Pessoas Jurídicas.

Screenshot- Imagem capturada que serve de ilustração.

SGBD – Sistema gerenciador de banco de dados.

SQL – *Structured Query Language* (Linguagem de consulta estruturada). É a linguagem padrão para definir e manipular dados em banco de dados relacionais.

Transact-SQL - Dialeto de SQL para o SGBD *SQL Server* da *Microsoft*.

1. Introdução

No ambiente das empresas com fins lucrativos, com efeito, o principal desafio vivenciado pelo indivíduo que está responsável pela gestão do negócio é tomar decisões com rapidez e qualidade [2]. Com isto, no meio corporativo, existe a necessidade das empresas otimizarem seus processos de tomada de decisão. Dentre os motivos para justificar tal necessidade encontra-se o fato de tais otimizações viabilizarem aumento de lucro que, conforme *Peter Drucker* - respeitado autor da área de Administração de Empresas - em [13], representa um dos principais objetivos perseguidos por empresas privadas.

É oportuno frisar que tais decisões são baseadas em componentes racionais e emocionais cuja intensidade e direção dependem do fator informação [40]. Nesse contexto, profissionais de Tecnologia da Informação (TI), especificamente de áreas como Banco de Dados e Sistemas de Suporte a Decisão, trabalham com conceitos, técnicas e ferramentas que tanto atendem às necessidades citadas acima quanto organizam e valorizam o patrimônio de informações de negócios das empresas que implantam soluções tecnológicas desta natureza.

Dentre tais conceitos destaca-se, entre outros, os *Data Marts* e *Business Intelligence*. Um *Data Mart* [3] é um repositório de informações que dá suporte à decisão para um pequeno grupo de pessoas sobre um determinado assunto, por exemplo, sobre assuntos de negócio citados por *Ralph Kimball* em [22] como Vendas a varejo, Estoque, Compras, Gerenciamento de Pedidos, Gerência de relacionamento com clientes e Contabilidade. *Business Intelligence*, segundo a definição clássica de *Howard Dresner* [40] que é considerado o pai do termo, é “o processo com o intuito de explorar e analisar informações estruturadas e específicas de um domínio para enxergar tendências ou padrões e, através disto, produzir percepções e tirar conclusões. Domínios incluem clientes, fornecedores, produtos, serviços e concorrentes”.

Numa pesquisa internacional realizada em abril de 2004 pelo *Gartner Group*, ficou aparente que uma importante parte do globo aprovou o conceito de BI, uma vez que os números de tal pesquisa revelam que aproximadamente 10% das 2.000 (duas mil) maiores empresas do mundo têm alguma forma de centro de

competência em *Business Intelligence* [40]. Um outro fato recente e interessante, de nível nacional, reforça o alinhamento de *Business Intelligence* com o sucesso do negócio e pode ser observado na publicação dos lucros do primeiro semestre de 2005 do Banco Bradesco cujos números apontam um lucro recorde de R\$ 2,6 bilhões [5]. Abaixo seguem algumas palavras de João Penso, gerente *Business Intelligence* do Bradesco Seguros [6]: “Dotar a Bradesco Seguros de um poderoso instrumento para tomada de decisões gerenciais que a auxiliasse a manter sua liderança e lucratividade em um mercado extremamente competitivo, foi o grande desafio assumido pela Diretoria Geral de TI da Empresa. Assim, a partir do segundo semestre de 2001, teve início um arrojado e inovador projeto de *Data Warehouse* destinado a atender às necessidades de informação das diversas unidades de negócio da Seguradora. Com base nesta abrangente base de informações integradas, outros projetos de sucesso puderam ser implementados como o Painel de Controle baseado na metodologia de *Balanced Scorecard* e um moderno sistema de detecção de fraudes fundamentado em técnicas de *Gestão de Conhecimento e Redes Neurais*”

1.1 Objetivos e Contexto

De uma maneira geral, o objetivo deste trabalho é realizar uma pesquisa sobre o assunto *Business Intelligence* e, em seguida, implementar um *Data Mart* para ambientes de BI.

O objetivo específico deste trabalho é, via pesquisa, entender o conceito, as características e o estado da arte em *Business Intelligence* bem como compreender as relações entre BI e *Data Marts* para, em seguida, aplicar tal conhecimento na construção de um *Data Mart* para *Business Intelligence*, obtendo-se como resultado um ambiente de BI com conhecimento capaz de auxiliar a empresa a tomar decisões de negócios com mais garantia e sucesso. Sendo assim, a implementação deste projeto facilitaria bastante o trabalho de tomadores de decisão como empresários, gestores, analistas de negócio, funcionários e colaboradores da empresa, além de ser útil servindo como *case* para profissionais ou estudantes de Tecnologia da Informação do Centro de Informática da UFPE ou de outro lugar da Sociedade.

Conforme exposto no título do trabalho, para realizar o objetivo citado acima será necessário se fazer, além da pesquisa sobre BI, um levantamento e análise de requisitos junto aos usuários, um projeto lógico (a modelagem) do *Data Mart* e, por fim, a implementação de tal *Data Mart* num sistema gerenciador de banco de dados.

Este projeto se situa numa empresa real que apresenta tais problemas e necessita resolvê-los para alcançar melhores resultados de negócio. Esta empresa é cliente da Zênite Soluções em TI & comércio LTDA [44] que é onde o autor deste trabalho de graduação trabalha. Embora, no momento desta monografia, o projeto não se encontre formal e comercialmente acordado entre tais empresas, todo o desenvolvimento do projeto pôde ocorrer normalmente devido ao contato diário do autor com: os funcionários da empresa cliente, o sistema ERP (*Enterprise Resource Planning*) do cliente e, principalmente, seus bancos de dados. Este contato se dá de maneira profissional pelo fato do autor prestar serviço de consultoria, customização e implantação do sistema ERP do cliente.

Para construir o *Data Mart* deste trabalho, foram utilizados como base os princípios de modelagem dimensional de autoria de *Ralph Kimball* [22] e a experiência prática em banco de dados do autor.

1.2 Estrutura do Trabalho

A seguir, a estrutura do trabalho será descrita para introduzir o leitor no material contido nesta monografia. No capítulo 2 – Principais Conceitos – serão apresentadas definições relevantes ao tema abordado bem como será descrito o problema que a implementação almejada se propõe a resolver. No terceiro capítulo – Estado da Arte em *Business Intelligence* - será exibida uma breve pesquisa sobre o estado da arte em Business Intelligence onde serão descritas suas características bem como serão apresentadas algumas aplicações do conceito que estão sendo utilizadas juntamente com alguns fatos sobre o mercado de BI. O capítulo 4 – Especificação do Projeto – apresenta o projeto do que foi implementado no trabalho. Ele caracteriza a concepção e o planejamento do projeto, com a conseqüente definição do escopo do mesmo. Mostra as

principais decisões de projeto realizadas bem como descreve seus usuários e os requisitos levantados junto a tais usuários. Este capítulo se encerra com o projeto lógico, e o respectivo *star schema* (esquema estrela), do *Data Mart* criado neste trabalho. Considerações e explicações sobre como o projeto foi implementado são feitas no capítulo 5 – Implementação do *Data Mart* – onde são descritas as principais decisões técnicas sobre a implementação e onde o *Data Mart* construído é descrito em detalhes. Em seguida, continuando o capítulo, são apresentadas técnicas, ferramentas e scripts SQL (especificamente *Transact-SQL* que é um dialeto de SQL para o MS *SQL Server*) utilizados no processo ETL (do inglês *Extract, Transform and Load* – Extração, Limpeza, Transformação e Carga) para tal *Data Mart*. Finalizando o capítulo, são apresentadas quatro ferramentas de análise e acesso aos dados no *Data Mart* que foram pesquisadas, preparadas e experimentadas neste trabalho.

2. Principais Conceitos

Nesta seção serão apresentados os principais conceitos presentes neste trabalho, bem como suas respectivas definições e relações entre eles. Sempre que oportuno, os conceitos serão exemplificados ou ilustrados para que haja um bom entendimento dos termos, os quais serão utilizados ao longo do trabalho. Finalizando-se esta seção, o problema abordado será caracterizado e os principais desafios desta problemática serão apresentados.

2.1 Definições

Para um bom entendimento deste trabalho, os principais conceitos relevantes e necessários são:

- *Business Intelligence (BI)*
- *Data warehouse (DW)*
- *Data Mart (DM)*
- **Metadados**
- *Data warehousing*
- *Modelagem dimensional*
- **Tabela de fatos**
- **Tabela de dimensões**
- **ETL (Extração, Transformação, Limpeza e Carga)**
- *Data Staging Area*
- **OLAP (On-Line Analytical Processing)**
- **Cubo Multidimensional**
- **Sistemas ERP (Enterprise Resource Planning)**
- *Faturas*
- *DSS (Decision Support System)*

Antes de definir o conceito *Business Intelligence (BI)*, é esclarecedor dizer que tal termo trata-se de um “termo guarda-chuva” [32]. Vários autores se referem ao conceito como sendo um “guarda-chuva” em que estariam incluídos

outros conceitos como: *Data Warehouse*, *Data Mart*, *Data Warehousing*, *ETL*, *DSS*, *EIS (Executive Information System)*, *OLAP* e *data mining*.

Uma vez esclarecida tal peculiaridade, cabe salientar também que tal conceito apresenta-se definido de formas variadas. A explicação para tal fato é que diferentes pessoas e empresas criaram e adotaram diferentes definições do conceito. Partindo-se finalmente para uma definição, dentre as principais definições da área, destaca-se a definição clássica de *Howard Dresner*, considerado o pai do termo, que em [40] afirma que BI é “o processo com o intuito de explorar e analisar informações estruturadas e específicas de um domínio para enxergar tendências ou padrões e, através disto, produzir percepções e tirar conclusões. Domínios incluem clientes, fornecedores, produtos, serviços e concorrentes”. Cabe ressaltar que complementarmente à definição usando linguagem de negócios de *Dresner* [40], existe a definição em [9], mais formada por termos de Tecnologia da Informação, na qual *Business Intelligence* é uma categoria ampla de aplicativos e tecnologias para captar, armazenar, analisar e prover acesso aos dados corporativos de forma a auxiliar os tomadores de decisão a tomarem melhores decisões de negócio. Por exemplo, Aplicativos de BI [19] incluem *DSS*, consultas e relatórios, *OLAP*, análises estatísticas, *forecasting* e *data mining*.

Para se obter um bom ambiente de *Business Intelligence*, um dos componentes básicos, e potencialmente o mais necessário e importante, é um *Data Warehouse* (ou um *Data Mart*, para um ambiente de BI com menor abrangência). De acordo com *Bill Inmon* [20], um *Data Warehouse (DW)* “é uma coleção de dados orientada a assuntos, não volátil, integrada e variante com o tempo, modelada para atender necessidades de gerenciamento do processo de suporte à decisão.” Um importante aspecto do ambiente de *Data Warehouse* diz respeito aos metadados. Metadados são dados que fazem referência a outros dados ou, segundo *Inmon*, os metadados mantêm informações sobre "o que está e onde" no ambiente de DW.

Este grande armazém de dados serve para armazenar informações e propiciar a constituição de ambientes mais estruturados de suporte à decisão [19].

Por exemplo, típicos cenários de uso do DW por parte dos usuário são:

- Estudar o rumo que a empresa deve tomar (como se focar num produto ou serviço A ou B),

- Analisar ameaças iminentes (como fortes concorrentes já conhecidos ou possíveis novos concorrentes entrantes no mercado)
- Avaliar oportunidades escondidas (como padrões de venda de produtos no clássico exemplo das vendas de cerveja e fraudas no *Walmart* [18])
- Realizar comparações históricas (como vendas do mês de maio de 2005 comparada às vendas de maio de 2004)
- Determinar sazonalidades (como identificar que os produtos X, Y e Z vendem mais no Verão ou Inverno; no Carnaval, São João ou Natal)
- Acompanhar taxas de retenção e conquista de novos clientes
- Comparar eficiência da área de logística ao longo do tempo.

Após apontar para que servem e onde são utilizados os *Data Warehouses*, é imprescindível destacar um conceito central a este trabalho e que já foi citado na introdução –um repositório de informações que dá suporte à decisão para um pequeno grupo de pessoas sobre um assunto determinado - os chamados *Data Mart's* [3]. Kimball et al [20] definem um *Data Mart* como sendo “uma parte do *Data Warehouse* restrita a um único processo de negócio, ou a um grupo de processos de negócio relacionados entre si e voltados para um grupo de negócio particular.”

Neste trabalho, o processo de negócio contemplado pelo *Data Mart* é o *Faturamento*. É imperioso esclarecer que embora o processo de negócio modelado no *Data Mart* seja o *Faturamento*, tal área da empresa não é a dona do *Data Mart*, fato inclusive alertado e chamado de “mito” por Kimball em [22]. O assunto “*Faturas*”, por exemplo, interessará certamente à área Financeira, Contábil, de Marketing e de Vendas. Uma *Fatura* é um documento de controle operacional que contém uma relação de produtos (e/ou serviços), quantidade e preço [35]. Estas informações de faturas, bem como outras relacionadas a elas, serão trabalhadas e transformadas na seção 5.2.3 para finalmente serem armazenadas no *Data Mart* que será chamado de “*Data Mart Faturas*”.

Para se projetar um *Data Warehouse* ou um *Data Mart* é comum se utilizar a modelagem dimensional. Modelagem dimensional, segundo Kimball em [22], é uma “metodologia que permite modelar logicamente dados para melhorar o

desempenho de consultas e prover facilidade de utilização a partir de um conjunto de eventos básicos e de medição". As estruturas comuns utilizadas para modelar os dados são conhecidas como tabelas de fatos e dimensões. Uma tabela de fatos é usada para armazenar medidas numéricas, que são associadas a eventos de negócio. As tabelas de dimensões estão sempre acompanhadas das tabelas de fatos e servem como descrições textuais dos fatos. Na implementação, os atributos da tabela de dimensão são utilizados como fonte primária de restrição de consultas, agrupamentos de dados e rótulos de relatórios [22].

Outro conceito importante, e que as vezes levemente é confundido com o próprio conceito de DW, é o processo de *Data Warehousing* [24]. Tal conceito refere-se à seqüência de atividades realizadas para a criação e utilização de um DW. Este processo pode ser visto na Figura 1, conforme proposto por Machado [24]. O resultado do processo de *Data Warehousing* é um *Data Warehouse* ou um *Data Mart*, dependendo do objetivo do processo.

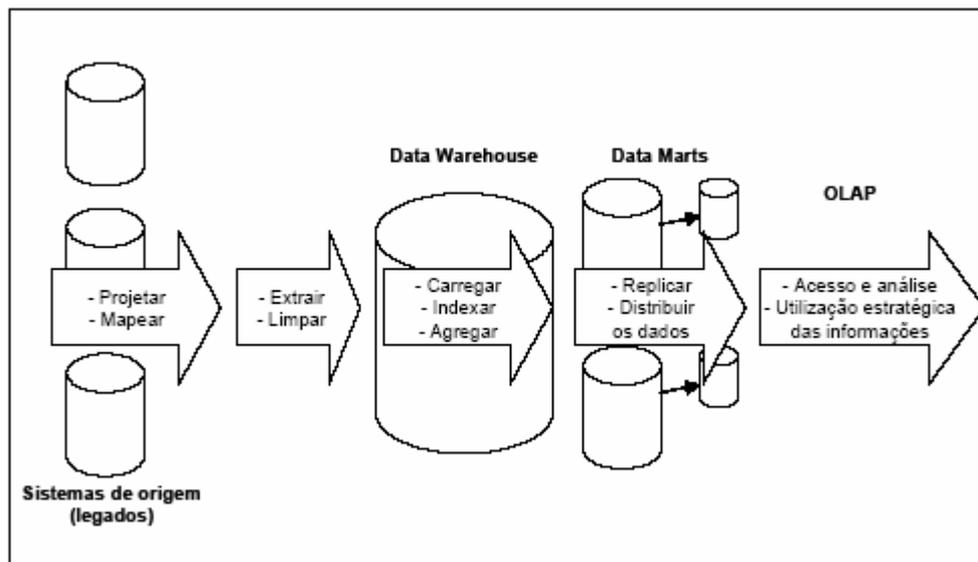


Figura 1 – O processo de *Data Warehousing*.

Uma das partes críticas do processo de *Data Warehousing* é a etapa de *ETL*. Estas etapas ocorrem numa área chamada de *Data Staging Area* a qual é uma área de processamento temporário onde ocorre a preparação dos dados que irão para o *Data Warehouse* (ou *Data Mart*). O conceito de *ETL* representa o processo que permite às instituições moverem dados de múltiplas fontes, reformatá-los, limpá-los e carregá-los em outro banco de dados (um *Data Mart* ou

Data Warehouse) para análise [14]. Esta análise é feita principalmente utilizando-se de processamento OLAP.

O termo *OLAP* significa *On-Line Analytical Processing* e foi definido por *E.F. Codd* em 1993. Por outro lado, em detrimento da definição de *Codd* e de acordo com a definição em [40], a qual é citada em mais de 120 sites de aproximadamente 30 países diferentes, tal conceito significa análise rápida de informações multidimensionais compartilhadas.

Um modelo para representar tais informações multidimensionais são os “cubos multidimensionais”. Um cubo é um conjunto que é comumente construído a partir de um subconjunto de um *Data Warehouse* e é organizado e sumarizado numa estrutura multidimensional definida por um conjunto de dimensões e fatos [36]. Uma ilustração de tal conceito é apresentada na Figura 2 [8] em que as dimensões são Linha de Produtos, Trimestre e Região e os fatos são as medidas numéricas do cubo:

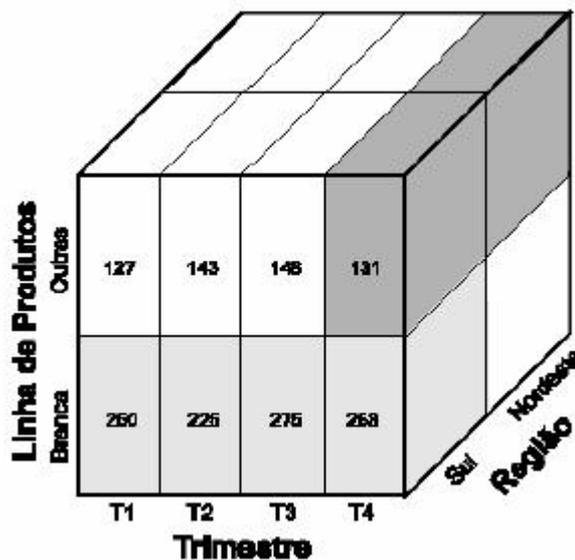


Figura 2 – Cubo multidimensional.

Tais cubos são os principais objetos relacionados a dados numa camada *OLAP*.

Neste projeto, dentre as fontes de dados citadas na definição de ETL, encontra-se principalmente: uma Base de dados de um sistema ERP.

Os Sistemas ERP - (do inglês *Enterprise Resource Planning*, Planejamento dos recursos de uma empresa) - são sistemas de gestão

empresarial que integram informações, departamentos, funções e processos de uma organização.

Um outro conceito importante e relacionado a este trabalho são os Sistemas de Suporte à decisão (sigla *DSS*, em inglês). Tais sistemas se caracterizam por fornecer informações sob demanda aos responsáveis por tomar decisões na empresa. Se uma empresa tem um ambiente de BI funcionando, este pode ser considerado um sistema de suporte a decisão.

2.2 O Problema

Após reuniões com os usuários para levantamento das necessidades deles e, posteriormente, uma análise do volume e tipo de informações presentes no banco de dados do sistema ERP do cliente, percebeu-se os seguintes problemas:

- Alto volume de informações apenas de nível operacional do negócio,
- Ausência de tecnologia de análise dimensional dos dados do negócio,
- Falta de ferramenta de acesso e inexistência de processo definido de exploração das informações do negócio.

Após o entendimento do problema, a abordagem realizada por este trabalho propôs as seguintes soluções, relacionadas aos seguintes conceitos exibidos na Tabela 1:

Tabela 1 - Análise do Problema.

| Problema identificado | Solução proposta | Conceito e Tecnologia |
|--|--|-----------------------|
| Alto volume de informações apenas de nível operacional do negócio | Construção de um repositório de informações integradas e de nível tático e estratégico | Data Mart |
| Ausência de tecnologia de análise dimensional dos dados do negócio | Utilização da Tecnologia de processamento analítico de dados | OLAP |
| Inexistência de processo definido de exploração das informações do negócio | Definição e implantação de um processo de Inteligência de Negócio para a empresa | Business Intelligence |

A implementação do principal artefato tecnológico a ser produzido neste trabalho, o *Data Mart*, não é uma tarefa fácil. Dentre os desafios específicos e presentes na implementação deste projeto, estão:

- Tamanho e complexidade da principal fonte de dados para o *Data Mart Faturas*. Esta fonte de dados é o banco de dados do sistema ERP do cliente do projeto cujo BD tem 1.041 tabelas.

Além dos desafios específicos deste projeto citados acima, de acordo com a literatura especializada da área, encontra-se também dificuldades como a tarefa de entender os processos e as regras do negócio da empresa, compreender os dados operacionais, manter a consistência e a credibilidade dos dados armazenados no DW (ou no *Data Mart*) e dos dados gerados nas consultas e relatórios. Interessante enfatizar que é possível que dois usuários interpretem os mesmos dados de maneira diferente, sendo que, na verdade, apesar de usarem os mesmos dados, estão realizando análises diferentes ou observando pontos-de-vista diferentes [19], [22]. Como o ambiente é dinâmico, uma variação mínima na consulta pode levar a resultados diferentes, que podem ser confundidos pelo usuário "distraído" o qual pode, eventualmente, querer colocar a culpa no *DM*.

Outros desafios relatados por estudiosos e profissionais da área em [38], [11][31] são:

- Sumarização e agregação de dados
- Qualidade dos dados: limpeza e refinamentos
- Gerência da expectativa do usuário
- Definição dos requisitos necessários para disponibilizar os dados do DW
- Dimensionamento de espaço em disco e processador necessário para o DW, já que eles são históricos e crescem indefinidamente consumindo rapidamente a capacidade de armazenamento e de processamento dos computadores atuais
- Integração de dados e metadados de várias fontes

Um reflexo de tais desafios pode ser notado na qualidade e tamanho da equipe de *Data Warehouse* descrita por Kimball em [22], onde se nota a necessidade de amplo conhecimento especialista dos profissionais envolvidos no desenvolvimento do DW.

3. Estado da Arte em *Business Intelligence*

Nesta seção serão apresentados o histórico da *Business Intelligence* (BI), as relações entre BI e *Data Marts* e o estado da arte na área.

3.1 Histórico de BI

Embora o termo BI seja relativamente novo, tendo surgido na década de 70, este conceito já era desenvolvido e aplicado de maneira empírica, milhares de anos atrás, por povos como Fenícios, Persas, Egípcios e outros orientais [19]. Tais povos cruzavam informações provenientes da natureza para tomar decisões que permitissem a melhoria de vida de suas comunidades. Por exemplo, os Egípcios, que viviam numa região quente e seca, mas são favorecidos com o maior rio do mundo em extensão, o Nilo, dependiam quase que totalmente das cheias periódicas, que adubavam as margens, para que os grãos pudessem ser semeados e grandes colheitas feitas. Há mais de 40 séculos, informações eram armazenadas sobre essas cheias e todos os episódios externos que aconteciam paralelamente, tais como estado da lua, temperatura, período da seca e de chuvas e o nível que o rio atingia. Com isso, era possível tomar decisões que resultassem na melhoria da qualidade de vida de suas comunidades.

A história de *Business Intelligence* que conhecemos hoje, começa na década de 70, quando alguns produtos de BI foram disponibilizados para os analistas de negócio. O grande problema era que esses produtos exigiam intensa e exaustiva programação, não disponibilizavam informação em tempo hábil nem de forma flexível, e além de tudo, tinham alto custo de implantação. Com o surgimento dos bancos de dados relacionais, dos PC's e das interfaces gráficas como o *Windows*, aliados ao aumento da complexidade dos negócios, começaram a surgir os primeiros produtos realmente direcionados aos analistas de negócios, que possibilitavam rapidez e uma maior flexibilidade de análise [2].

Em 1980 o termo foi registrado por *Howard Dresner* do *Gartner Group* [40] como mais um dos jargões administrativos que povoavam dicionários dos

executivos [19]. Com o passar do tempo, a expressão BI teve grande apoio na área de tecnologia, pois tal “inteligência de negócio” pode ser melhor explorada com a utilização de recursos de *software* e *hardware* adequados.

3.2 Relações entre *Business Intelligence* e *Data Mart*

Esta seção visa esclarecer as relações entre BI e *Data Marts* para que o leitor tenha total compreensão do significado da expressão “Um *Data Mart* para ambientes de BI”, a qual, desnecessário dizer, faz parte do título deste trabalho.

A primeira relação é que um *Data Mart* (ou um *Data Warehouse*) é um componente de um ambiente de BI. Importante citar a consequência deste fato: tal repositório de dados deve acompanhar as alternâncias do mundo exterior e as exigências dos analistas de negócio [8].

A segunda relação pode ser explicada analisando-se o significado dos termos da expressão “Inteligência de Negócio”. *Inteligência* significa “*Faculdade de entender, pensar, raciocinar e interpretar; entendimento, intelecto. 2. Compreensão, conhecimento profundo. 3. Pessoa de grande esfera intelectual. 5. Conluio, ajuste, combinação*” [35]. Utilizando-se do significado “*conhecimento profundo*”, conclui-se que BI visa dar conhecimento profundo sobre o negócio. Como para isto, segundo o pai do termo *Howard Dresner* em [40], é necessário conhecer clientes, fornecedores, concorrentes e produtos ou serviços da empresa, então, nesse contexto, aparecem os *Data Warehouses* ou *Data Marts* para armazenar tais informações e viabilizar análises que auxiliarão a empresa a ter um ambiente de BI presente.

Portanto, um “*Data Mart* para ambientes de BI” é um repositório que armazena importantes informações sobre clientes, fornecedores, concorrentes e produtos ou serviços da empresa de maneira a viabilizar um ambiente de conhecimento profundo sobre o negócio, um ambiente de BI.

3.3 O Estado da Arte

Esta seção apresenta, sem a pretensão de esgotar o assunto, o que está sendo utilizado na área de *Business Intelligence*.

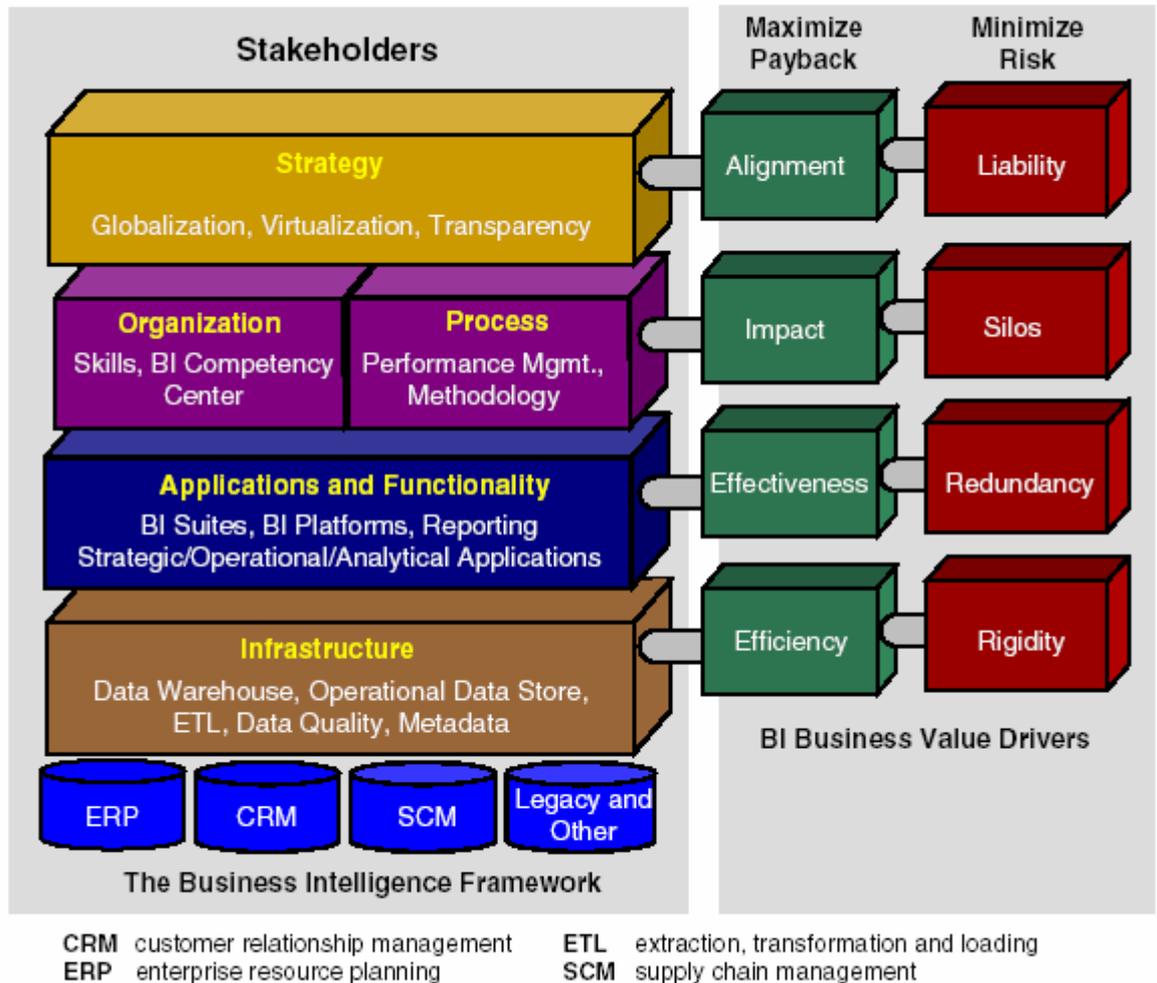
Do ponto de vista conceitual, destaca-se principalmente o conceito de *WBI* (*Web Business Intelligence* ou *Web-BI*). Os autores Tan et al. [35] e Henrichs e Lim [8] destacam o desenvolvimento e crescimento deste conceito o qual mescla *Data Warehousing* e sistemas de BI com a tecnologia da *Web*. As vantagens fundamentais do uso de sistemas de *WBI*, quando comparados com os sistemas tradicionais, incluem acesso universal ao *Web-BI* e capacidade de trabalho com sons e vídeos, além de textos, números e gráficos. A redução das barreiras tecnológicas e geográficas proporcionadas pelos sistemas de *Web-BI* beneficia bastante os negócios, pois tornam mais fácil e viável financeiramente se disponibilizar aos decisores informações relevantes para tomadas de decisão, mesmo que estes estejam geograficamente em localizações distantes, por exemplo numa viagem para outro estado ou país. Construir um *Web-BI* é uma tarefa complexa, pois deve considerar um sistema de alta disponibilidade, que esteja funcionando sete dias por semana, vinte e quatro horas por dia e deve manipular um grande número de solicitações concorrentes [28].

Dentre as quatro ferramentas de acesso e análise de dados para suporte à decisão analisadas, preparadas e experimentadas na seção 5.3 deste trabalho, uma delas implementa o conceito de *Web Business Intelligence*. Este recurso foi experimentado através da construção de uma aplicação de *Web-BI*, utilizando-se a tecnologia *OWC* (*Office Web Components*) [29]. Uma ilustração desta aplicação *web* pode ser vista no item “b)” do APÊNDICE B deste trabalho.

Outro conceito importante, mais voltado para empresas que investiram, estão investindo ou vão investir num projeto de BI, é o modelo “*The BI value framework*” [17], definido pelo *Gartner Group*. Este *framework* auxilia o gerenciamento de um projeto de BI, pois permite medir o sucesso de uma iniciativa de BI na empresa. Este sucesso, no modelo, é determinado pelo grau de *business value* (incluindo, por exemplo: eficiência, eficácia, impacto e grau de alinhamento aos objetivos estratégicos) que a iniciativa de BI proporciona para a empresa. O *framework* se baseia em métricas e princípios fundamentais que, se

seguidos, se propõem a maximizar o retorno do investimento no projeto e minimizar os riscos do mesmo. Uma ilustração do modelo pode ser vista na Figura 3.

The BI Value Framework



Source: Gartner Research (April 2004)

Figura 3 – Um framework para *BI Value* [17].

Conforme exibido na Figura 3, o framework é dividido em quatro níveis (*Strategy*, *Organization - Process*, *Applications and Functionality*, *Infrastructure*), cada um com medidas de *business value* ao longo das dimensões *Maximize Payback* e *Minimize Risk*. Quanto mais alto o nível, mais alto é o seu potencial para maximizar o retorno e minimizar riscos. Contudo, nos níveis mais altos, é mais difícil e complexo quantificar o *business value* proporcionado por tal nível.

Do ponto de vista de mercado, em termos de tipos de produtos, destacam-se os *Enterprise BI Suítes (EBIS)*, ferramentas de consulta e relatório, ferramentas de BI avançadas (ferramentas analíticas OLAP avançadas), e plataformas de BI para o desenvolvimento de aplicações de BI [3].

Continuando a visão mercadológica, alguns dos principais softwares de BI e softwares relacionados, juntamente com seus respectivos fornecedores, são descritos na Tabela 2. Excluindo-se a SAP, Microsiga, *Microsoft* e a Datasul, os outros fornecedores listados lideraram a categoria de BI da 4ª pesquisa de Marcas/Empresas da *Revista Info* do mês de abril de 2005.

Ainda no contexto do mercado de BI, segundo um relatório divulgado pela IDC, uma das principais tendências na área é o aumento do número de projetos de consolidação das ferramentas de BI, voltados ao desenvolvimento de uma estrutura que atenda à organização de maneira abrangente [8]. Neste aspecto, ressalta-se o crescimento das aplicações direcionadas ao monitoramento da performance empresarial (*Business Performance Management – BPM*) e gerenciamento de indicadores-chave de negócios (*Key Performance Indicator – KPI*), além da integração destas com as ferramentas tradicionais de BI, que vêm sendo implementadas nos últimos anos de maneira crescente no Brasil. Em tal relatório, a IDC mostra ainda que uma das principais áreas de crescimento em *Business Intelligence* será a aplicação das ferramentas de BI em dispositivos móveis.

**Análise, Projeto e Implementação de um *Data Mart* para
Ambientes de *Business Intelligence***

Tabela 2 – Principais *softwares* de BI no mercado.

| Softwares de BI * | Nome do fornecedor |
|---|---|
| <i>Cognos ReportNet, Cognos PowerPlay, Cognos NoticeCast, Cognos DecisionStream, Cognos Performance Applications</i> | COGNOS - http://www.cognos.com |
| Plataforma <i>Microstrategy (Application Integration, Closed-Loop BI, Corp. Performance Mgmt., Embedded Intelligence, Extranets, Portals, Security, Solutions for Excel, Solutions for SAP, Web Services)</i> | MICROSTRATEGY - www.microstrategy.com.br |
| <i>Crystal Reports, Crystal Reports Explorer, Live Office, Web Intelligence, OLAP Intelligence, BusinessObjects, Dashboard Manager, Performance Manager, Applications, BusinessObjects, Enterprise, Integration Kits, Analytic Engines, Data Integrator, Rapid Marts</i> | BUSINESS OBJECTS - www.businessobjects.com.br |
| <i>Business Intelligence 10g, Business Intelligence Beans 10g, Business Intelligence Solution, Business Intelligence Warehouse Builder 10g</i> | ORACLE - www.oracle.com |
| <i>SAS Enterprise BI Server (SAS Web Report Studio, SAS Add-In for Microsoft Office, SAS Information Delivery Portal, SAS Information Map Studio, SAS Integration Technologies)</i> | SAS - www.sas.com |
| <i>Hyperion BI Platform (Hyperion Essbase, Hyperion Metrics Builder, Hyperion Performance Suite, Hyperion Intelligence, Hyperion Analyzer, Hyperion Visual Explorer, Essbase Spreadsheet Services, Hyperion SQR, Hyperion Reports, Hyperion Application Builder, Essbase Administration Services, Hyperion Deployment Services, Essbase Integration Services, Hyperion Application Link, Hyperion Hub, Hyperion MDM Server)</i> | HYPERION - www.hyperion.com |
| <i>DB2 UDB DWE (DW Editions), DB2 OLAP Server, WebSphere Information Integrator, Query Management Facility, DB2 Alphablox, O DB2 Entity Analytics</i> | IBM - www.ibm.com |
| <i>SAP Business Intelligence, SAP Business Warehouse</i> | SAP - www.sap.com |
| SigaDW | Microsiga Software - www.microsiga.com.br |
| <i>MS SQL Server, MS Analysis Services, MS DTS, MS Data Analyzer, MS Excel e Office Web Components, Reporting Services, Business Scorecard Accelerator, Maestro (previsão [17] de lançamento para setembro/2005)</i> | Microsoft - www.microsoft.com.br |

* Inclui softwares não específicos de BI, mas relacionados.

4. Especificação do Projeto

Nesta seção serão apresentadas as fases de planejamento, análise e projeto lógico do *Data Mart*. Sendo assim, serão definidos o escopo e os elementos humanos e técnicos presentes no ambiente do projeto, bem como serão especificadas as características do *Data Mart* a ser construído. É importante ressaltar que, dada a inegável importância das pessoas diretamente afetadas por este projeto, como diretores, gerentes e analistas de negócio, também serão feitas considerações sobre os usuários envolvidos no projeto. Naturalmente, serão especificados também os requisitos que tais usuários desejam que o projeto atenda. Por fim, serão apresentados os resultados técnicos desta fase do trabalho como o importante e clássico *Star schema* (modelo estrela) o qual é o principal objetivo desta fase.

4.1 Principais Decisões de Projeto

As principais decisões de projeto, tomadas na fase de concepção do mesmo, foram:

- Construir inicialmente um *Data Mart* (utilizar abordagem *bottom-up* de *Kimball*)
- Contemplar primeiro o processo de negócio Faturamento
- Criar o *Data Mart* especificamente para ambientes de *Business Intelligence*

O que motivou a necessidade de tais decisões serem tomadas foram basicamente três questionamentos sobre o projeto:

a) Por quê construir inicialmente um *Data Mart* e não um *Data Warehouse*?

Porque construir um *Data Warehouse* é uma iniciativa cara, complexa e relativamente demorada. Em contrapartida, construir inicialmente um *Data Mart* permite a empresa experimentar a tecnologia e decidir se o investimento vale a pena, ou seja, diminui-se o risco do projeto optando-se por começar pelo *Data*

Mart. Tal abordagem está alinhada, inclusive, a alguns fornecedores de soluções de *Data Warehouse* conforme citado em [3]. Uma ilustração desta abordagem *bottom-up* é apresentada na Figura 4 [31]:

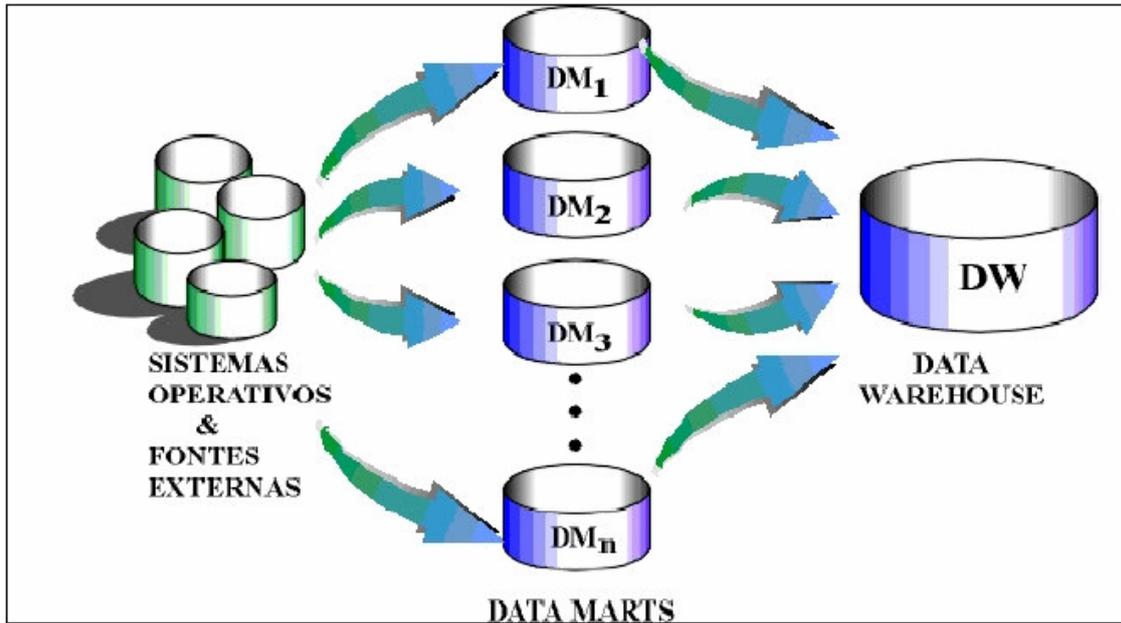


Figura 4 – Abordagem *bottom-up* para construção de *Data Warehouse* .

b) Qual a razão em se construir primeiro um *Data Mart* sobre Faturas?

Primeiramente, do ponto de vista conceitual, pelo fato dos dados produzidos pelo processo de Faturamento serem considerados “o banco de dados mais eficiente” uma vez que este BD combina os clientes, produtos e os componentes de lucratividade (custo e preço) da empresa. Este argumento está alinhado aos princípios de *Ralph Kimball* em [22].

Em segundo lugar, do ponto de vista do cliente, pelo fato de, neste projeto, o Faturamento representar um processo de negócio cujo módulo no Sistema ERP do cliente já se encontra implantado e em produção.

Portanto, contemplar inicialmente o assunto “Faturas” torna-se uma escolha eficaz e prioritária.

c) Por que um *Data Mart* para *Business Intelligence*?

Devido aos benefícios e vantagens de negócio propiciadas pela *Business Intelligence* como antecipação às mudanças de mercado, antecipação às ações dos concorrentes, conhecimento sobre o negócio (como a lucratividade), aprendizado pelo sucessos e falhas internas e dos concorrentes, visão clara sobre novos negócios, percepção sobre novos produtos ou serviços que podem trazer benefícios para o negócio [19].

4.2 Usuários

Diferentemente dos projetos de sistemas transacionais, em que o foco de trabalho é no sistema (metodologias de desenvolvimento, métricas de software, testes, análise de impactos e riscos e implantação), nos projetos de *Data Warehouse* ou *Data Mart* a atenção está totalmente voltada para os usuários e para o banco de dados que armazenará o *DW* ou o *DM* [19]. É preciso lembrar também que tais projetos consomem, geralmente, bastante tempo e recursos financeiros de maneira que para uma empresa aceitar tal investimento é primordial que o projeto tenha um “patrocinador”, ou seja, uma pessoa com influência na organização e que dê apoio e acredite no projeto.

Algumas diferenças que podem ser percebidas pelos usuários entre os sistemas transacionais e o ambiente de suporte a decisão do qual o *Data Mart* faz parte encontram-se na Tabela 3, adaptada de [19]:

Tabela 3 – Diferenças entre sistemas transacionais e ambientes de suporte à decisão.

| Característica | Sistema transacional | Ambiente de suporte à decisão |
|---------------------------|----------------------|---|
| Pergunta do usuário | O quê ? | Por quê? |
| Função do usuário | Operar a empresa | Dirigir a empresa |
| Registros acessados | Um por vez | Muitos ao mesmo tempo |
| Ênfase de processamento | Atualização | Consulta |
| Tempo de resposta exigido | Poucos Segundos | Minutos (ou Muitos segundos) |
| Utilização | Repetitiva, focada | Imprevisível, sem periodicidade definida |
| Perfil do usuário | Operacional | Tático e estratégico (Analítico, decisório) |

Dentre os objetivos dos usuários do *DW* ou do *DM*, os quais são profissionais especializados no negócio da empresa, entre outros, estão:

- Identificar ameaças e oportunidades de negócios.
- Expandir a participação da empresa em outros segmentos de mercado.
- Elevar as vendas e incrementar as receitas
- Reduzir custos
- Aumentar o lucro do negócio

Neste projeto percebeu-se a presença principalmente dos seguintes usuários de negócio: diretor de vendas, diretor de compras, gerente geral, gerente comercial (que no caso engloba o setor de faturamento), analistas de negócio, além do analista-desenvolvedor do *Data Mart*. Percebe-se que tais usuários buscam interpretar como os negócios têm sido administrados e conduzidos ao longo do tempo, além de também procurarem saber o que pode ser feito para se ter mais inteligência nos negócios.

4.3 Requisitos e Escopo

Após sucessivas reuniões com os usuários para levantamento de requisitos, tais requisitos foram registrados e analisados. Depois de analisados e refinados, por não apresentarem conflitos, inconsistências ou incompletudes, prosseguiu-se com um trabalho de estudo e entendimento dos dados necessários para suportar os requisitos de negócio contemplados pelo escopo deste projeto, ou seja, os dados relativos às transações de faturamento da empresa do cliente.

Nesta etapa, percebeu-se que o fato do sistema ERP do cliente já ser integrado amenizou a complexidade de localização e acesso a tais dados, fato este que, conforme será mostrado no Capítulo 5 deste trabalho, diminuirá a dificuldade do processo ETL para o *Data Mart Faturas*.

Após entendimento e constatação das disponibilidades dos dados, os principais requisitos de negócio que o *Data Mart Faturas* se propõe a atender foram agrupados na Tabela 4, juntamente com suas respectivas importâncias para um ambiente de *Business intelligence*.

Análise, Projeto e Implementação de um *Data Mart* para
Ambientes de *Business Intelligence*

Tabela 4 – Principais requisitos do negócio.

| Descrição do requisito | Importância no negócio |
|---|--|
| Informações sobre quantidade total de produtos faturados opcionalmente por Filial, por Produto, por Cliente, por Vendedor, por Data, por Armazém. | Auxilia decisões baseadas em volume de vendas. Por exemplo, um grande volume de vendas do produto P1 influenciará o decisor de compras a adquirir e manter um estoque adequado do produto. |
| Informações sobre valor total (quantidade x preço unitário) faturado opcionalmente por Filial, por Produto, por Cliente, por Vendedor, por Data, por Armazém. | Afeta, por exemplo, a classificação contábil da empresa que se refletirá, entre outros, na alíquota e valor de IRPJ a serem pagas ao governo. |
| Informações de preço médio praticado opcionalmente por Filial, por Produto, por Cliente, por Vendedor, por Data, por Armazém. | Crucial no gerenciamento de clientes e útil para se fazer <i>benchmarking</i> junto à concorrência. Por exemplo, se na empresa o preço médio do produto P2 é R\$ 60,00 e na concorrência o mesmo produto, nas mesmas condições, ficaria por R\$ 45,00, então providências devem ser tomadas imediatamente, como reduzir o preço trocando de fornecedor. |
| Informações de custo médio opcionalmente por Filial, por Produto, por Cliente, por Vendedor, por Data, por Armazém. | Influencia decisões, por exemplo, num contexto de logística, sobre “qual galpão deverá ser usado para armazenar os produtos?” Baseando-se no custo de armazenagem, que varia por armazém e compõe o custo dos produtos, pode-se reduzir os custos dos produtos escolhendo-se os armazéns com menores custos de armazenagem. |
| Informações de custos total (quantidade x custo) dos produtos opcionalmente por Filial, por Produto, por Cliente, por Vendedor, por Data, por Armazém. | Serve para se ter uma percepção do montante de capital necessário para se comprar tal produto. Por exemplo, se a empresa não estiver com suficiente disponibilidade financeira em caixa, tal produto não deve ser comprado no momento. Caso esta informação não fosse conhecida, poderia-se gerar uma inadimplência junto ao fornecedor do produto. |
| Cálculo de margem de lucro opcionalmente por Filial, por Produto, por Cliente, por Vendedor, por Data, por Armazém. | A margem de lucro é um indicador crítico da eficiência do negócio. Ao mesmo tempo em que uma boa margem de lucro é geralmente sinônimo de bons resultados e conseqüente satisfação do empreendedor do negócio, uma margem de lucro baixa ou negativa pode significar sua falência. Sendo assim, é importante que tal indicador seja fornecido pelo <i>Data Mart</i> e exaustivamente monitorado. |

Além dos requisitos definidos na Tabela 4, o *Data Mart Faturas* também deve atender requisitos não-funcionais como:

- Performance
- Clareza na modelagem dimensional
- Segurança
- Integridade e Consistência
- Extensibilidade

4.4 Arquitetura Tecnológica e Informacional do Ambiente

O ambiente tecnológico e informacional construído neste trabalho é composto por seis tipos de elementos, a saber:

- Fontes de dados operacionais
- Sistema ERP
- *Data Mart Faturas*
- Metadados
- *OLAP*
- Interface

Tais elementos formam a arquitetura do ambiente, cuja especificação é encontrada na Figura 5 (esta imagem é uma adaptação da figura em [17]).

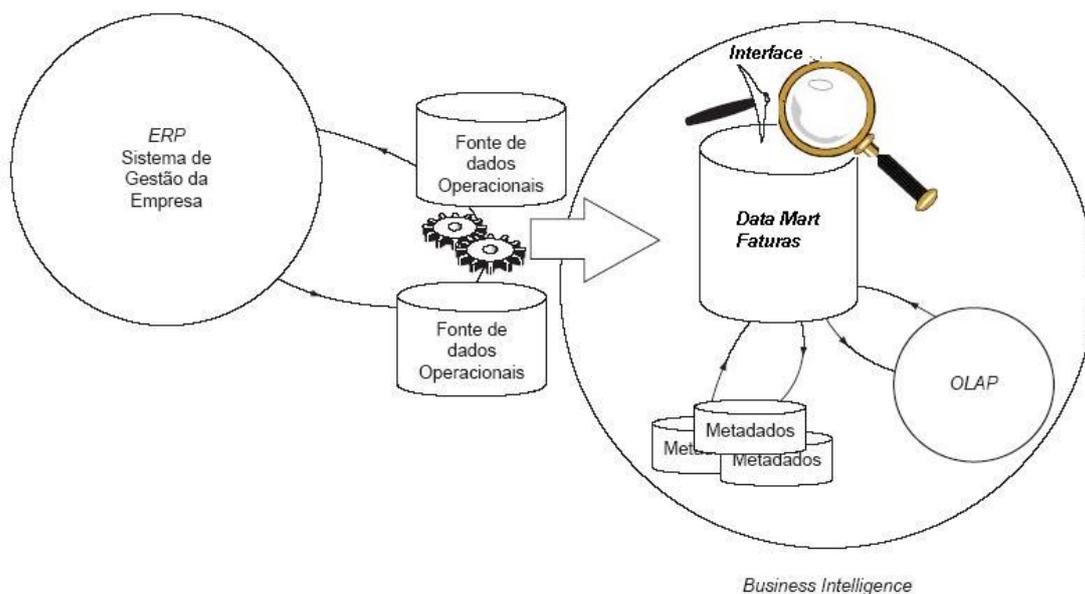


Figura 5 – Arquitetura do Ambiente.

É oportuno lembrar que serão enfatizadas as características do *Data Mart Faturas* por tal elemento representar o alvo principal deste trabalho.

4.5 Projeto lógico do *Data Mart Faturas*

O objetivo do *Data Mart Faturas* é atender aos requisitos definidos na Tabela 4 e, desta forma, complementar e dar suporte informacional ao ambiente de *Business Intelligence*. A especificação do *Data Mart* é resultado de quatro etapas, que levam em consideração sistematicamente uma ordem específica, e que são descritas por *Ralph Kimball* nas etapas de criação de um banco de dados dimensional [21], [22]. Tais etapas e as respectivas especificações geradas neste trabalho são descritas na Tabela 5:

Tabela 5 – Etapas de especificação do *Data Mart Faturas*.

| Etapa | Especificação Gerada |
|------------------------------------|--|
| 1. Selecione o processo de negócio | Processo de negócio selecionado: Faturamento |
| 2. Declare a granularidade | Cada linha da tabela de fatos representará UM ITEM (PRODUTO) DE UMA FATURA DE VENDA |
| 3. Escolha as dimensões | FILIAL CLIENTE ARMAZÉM VENDEDOR PRODUTO DATA FATURA |
| 4. Identifique os fatos | QUANTIDADE_FATURADA PRECO_MEDIO VALOR_FATURADO CUSTO_MEDIO CUSTO_FATURADO MARGEM_DE_LUCRO |

Após o fim da realização das quatro etapas acima, têm-se o projeto lógico do *Data Mart Faturas*, cujo esquema estrela pode ser visto na Figura 6:

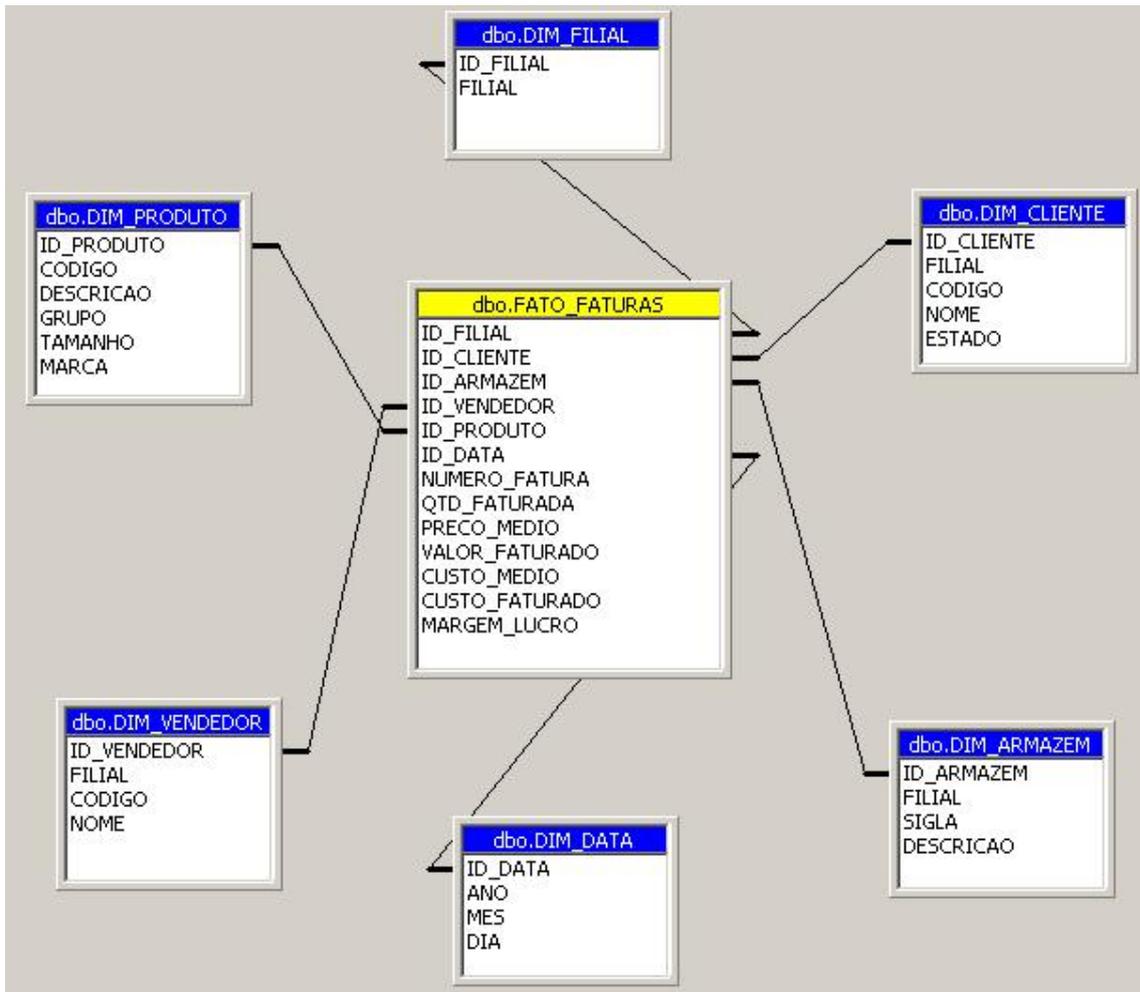


Figura 6 – Esquema estrela do *Data Mart Faturas*.

Neste momento, finaliza-se a especificação do ambiente do *Data Mart*. O principal artefato produzido até agora foi o Esquema Estrela do *Data Mart Faturas*, segundo a nomenclatura da modelagem dimensional por *Kimball* em [21], [22]. No capítulo 5 – Implementação – este esquema estrela será mapeado num modelo físico para concluir sua implementação. Também no próximo capítulo serão descritas interfaces de análise e acesso ao *Data Mart Faturas*, as quais foram pesquisadas, preparadas e experimentadas neste trabalho.

5. Implementação do *Data Mart*

O objetivo desta seção do trabalho foi realizar o projeto físico do *Data Mart Faturas*, para se obter as tabelas de fato e dimensões implementadas e carregadas. Serão relatadas as principais decisões técnicas que foram tomadas nesta fase do projeto como escolhas do esquema de modelagem dimensional, do SGBD, da alocação física do *Data Mart Faturas*, das ferramentas *ETL* e da camada *OLAP* a serem utilizados. Em seguida, serão detalhados alguns aspectos de implementação do *Data Mart Faturas* como detalhes das tabelas de dimensão e fatos e o processo *ETL* para tal *Data Mart*. Além disto, será exibido um *screenshot* do *Data Mart Faturas* após o carregamento dos dados. Finalizando a seção, será feita uma análise e escolha de uma interface de acesso e análise de dados a tal repositório de informações de negócio. Esta é uma etapa crítica e desafiadora do projeto pois, como amplamente conhecido pela comunidade de Tecnologia da Informação, freqüentemente é exatamente na implementação que grandes dificuldades emergem, de maneira que são exigidas soluções racionais e criativas do profissional ou equipe que está conduzindo o projeto.

5.1 Principais Decisões Técnicas Sobre a Implementação

5.1.1 Esquema de Modelagem Dimensional

O esquema escolhido foi o Esquema Estrela devido aos ganhos de performance propiciados por tal esquema. Esta decisão, inclusive, representa um esforço alinhado ao atendimento do requisito de performance levantado junto aos usuários. Caso se escolhesse, por exemplo, o esquema *Snowflake*, os benefícios atingidos seriam a redução de armazenamento necessário para o *Data Mart Faturas*, porém tal vantagem de redução de espaço em disco não é prioritária comparada com a vantagem de performance alcançada com o esquema estrela.

5.1.2 SGBD

Primeiramente é importante frisar que a escolha do SGBD é, sem dúvidas, uma das mais importantes questões técnicas inerentes a este projeto pois o

SGBD afeta diretamente questões como escalabilidade do *Data Mart*, desempenho e conseqüente tempo de respostas das consultas e comunicação com aplicações de *Business Intelligence*. Neste trabalho, o SGBD escolhido foi o *Microsoft SQL Server 2000* [37] por este SGBD atender bem às questões citadas acima e, principalmente, pelo fato do banco de dados do sistema ERP do cliente ser o *Microsoft SQL Server 2000*.

5.1.3 Alocação física do *Data Mart Faturas*

O *Data Mart Faturas* foi fisicamente alocado segundo as práticas recomendadas por especialistas da área, ou seja, num novo banco de dados, diferente do banco de dados operacional da empresa. Esta abordagem proporciona como benefício, principalmente, o ganho de performance. Como no momento em que esta monografia está sendo escrita o projeto encontra-se em fase de testes, o computador que armazenou o *Data Mart* foi uma estação de trabalho normal, sem ser um servidor. Quando o projeto for colocado em produção, o *Data Mart* sairá do ambiente de testes e ficará alocado num servidor apropriado, inclusive será recomendado que tal servidor seja exclusivo também por questões de performance.

5.1.4 Ferramenta ETL

Na etapa de extração decidiu-se por utilizar a ferramenta *DTS* [30] (*Data Transformation Services*) da *Microsoft* [26] pelo fato da mesma ser nativa ao *MS SQL Server 2000* além de prover boa funcionalidade como importação e exportação de dados entre BD distintos.

Já nas etapas de transformação e carga houve a escolha de implementá-las via comandos da linguagem SQL escritos manualmente pelo autor utilizando-se como ferramenta o *SQL Query Analyzer*, que é uma ferramenta nativa do *MS SQL Server 2000*. O principal motivo que influenciou esta decisão foi o sólido conhecimento e familiaridade do autor com a linguagem SQL.

O processo ETL para este projeto será detalhado e exemplificado posteriormente neste documento.

5.1.5 Camada OLAP

A camada OLAP escolhida foi o *Analysis Services* do *MS SQL Server 2000*. Esta decisão foi influenciada pelo SGBD escolhido ter sido o *SQL Server 2000* também da *Microsoft*. Esta decisão tem as vantagens de forte integração entre o *SQL Server 2000* e o *Analysis Services* além do fator custo adicional pela ferramenta ser reduzido a zero, uma vez que o *Analysis Services* é incluso na aquisição ou licenciamento do *SQL Server 2000*.

5.2 Detalhamento do *Data Mart Faturas*

Nesta seção serão apresentadas as tabelas de dimensão e de fatos, o processo ETL e um *Screenshot* dos dados carregados no *Data Mart Faturas* após a etapa de carga.

5.2.1 Detalhes das tabelas de dimensão

Alguns detalhes das 06 (seis) principais tabelas de dimensões como nome físico, atributos e quantidade de registros são apresentados na Tabela 6:

Tabela 6 – Detalhes das tabelas de dimensão do *Data Mart Faturas*.

| Nome Dimensão | Físico da | Atributos | Quantidade de registros |
|---------------|-----------|---|-------------------------|
| DIM_FILIAL | | <u>ID_FILIAL</u> , FILIAL | 6 |
| DIM_CLIENTE | | <u>ID_CLIENTE</u> , FILIAL, CODIGO, NOME, ESTADO | 5.824 |
| DIM_ARMAZEM | | <u>ID_ARMAZEM</u> , FILIAL, SIGLA, DESCRICAO | 81 |
| DIM_VENDEDOR | | <u>ID_VENDEDOR</u> , FILIAL, NOME | 61 |
| DIM_PRODUTO | | <u>ID_PRODUTO</u> , CODIGO, DESCRICAO, GRUPO, TAMANHO E MARCA | 1.285 |
| DIM_DATA | | <u>ID_DATA</u> , ANO, MÊS, DIA | 518 |

Também foi modelada a seguinte dimensão de degeneração [22]:

Tabela 5.2 – Tabela de dimensão de degeneração do *Data Mart Faturas*.

| Nome da dimensão | Atributo |
|------------------|---------------|
| DIM_FATURA | NUMERO_FATURA |

Tal dimensão de degeneração, conforme terminologia de *Kimball* e que é uma dimensão com apenas um atributo, será mapeada fisicamente como atributo da tabela de fatos, conforme será apresentado adiante no documento [22].

5.2.2 Detalhes das tabelas de fatos

Alguns detalhes da tabela de fatos do *Data Mart Faturas* como nome físico do fato, sua descrição e sua fórmula de cálculo em relação às tabelas da *Data Staging Area* são apresentados na Tabela 7:

Tabela 7 – Detalhes da tabela de fatos do *Data Mart Faturas*.

| Nome Físico do Fato | Descrição | Fórmula de Cálculo |
|---------------------|---------------------------------------|---|
| QTD_FATURADA | Quantidade faturada do produto | SUM(C6_QTDVEN) |
| PRECO_MEDIO | Preço médio do produto | AVG(C6_PRCVEN) |
| VALOR_FATURADO | Valor total do produto na fatura | SUM(C6_QTDVEN * C6_PRCVEN) |
| CUSTO_MEDIO | Custo médio do produto | AVG(C6_CUSTLOC*2.46) Obs: 2.46 é a cotação do dólar no dia da carga, 20.07.2005, que no caso era 2.46 em www.bb.com.br |
| CUSTO_FATURADO | Custo total do produto na fatura | SUM(C6_QTDVEN * C6_CUSTLOC*2.46) |
| MARGEM_LUCRO | Margem de lucro percentual do produto | 100 * ((AVG(C6_PRCVEN) - AVG(C6_CUSTLOC*2.46)) / AVG(C6_CUSTLOC*2.46)) |
| NUMERO_FATURA | Número da Fatura | C6_NUM (agrupado numa instrução GROUP BY) |

5.2.3 O Processo ETL para o *Data Mart Faturas*

A etapa de Extração foi implementada com a rotina de importação de dados da ferramenta *DTS*. Esta rotina extrai os dados do banco de dados do sistema ERP e os copia para a *Data Staging Area*, que no caso foi alocada fisicamente no mesmo banco de dados do *Data Mart Faturas*. Para se otimizar repetidas execuções desta rotina e obter maior rapidez na etapa de Extração, tal

rotina foi encapsulada num “*DTS Package*”, um recurso disponível na ferramenta. Uma ilustração de tal rotina é apresentada na Figura 7:

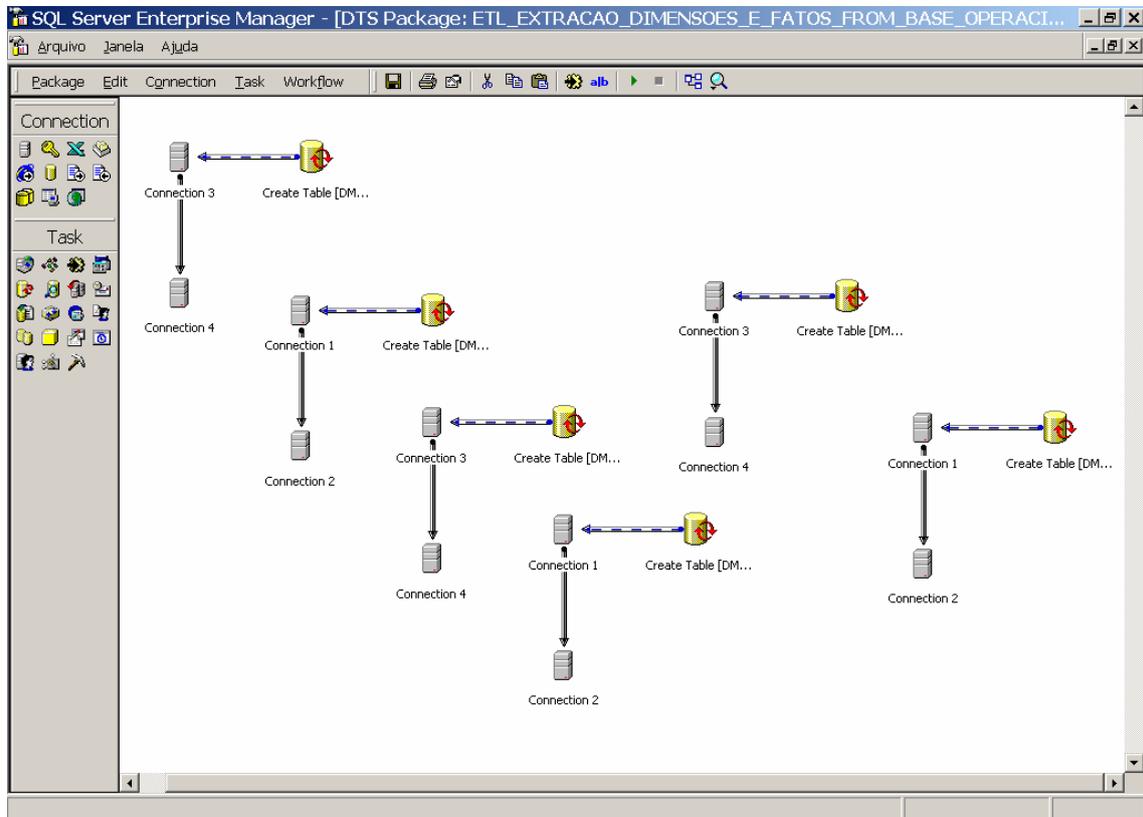


Figura 7 – Etapa de Extração do processo ETL para o *Data Mart Faturas*.

Na figura 7, os elementos *connections* são usados para se conectar ao BD do sistema ERP e ao BD do *Data Mart Faturas* e copiar as tabelas desejadas do ambiente operacional para a *Data Staging Area*. Após esta etapa de Extração, tais tabelas estão disponíveis para as etapas de Transformação (e Limpeza) e Carga.

Neste projeto, nas etapas de Transformação, Limpeza e Carga foram usadas técnicas de agregação (SUM e AVG), filtragem (WHERE) e inserção de dados (INSERT INTO E SELECT INTO). Ambas as técnicas foram implementadas via comandos *SQL*. Alguns exemplos desses comandos, em código *Transact-SQL*, serão apresentados a seguir:

- Etapa ETL de transformação, limpeza e criação da dimensão Produto em relação às tabelas da *Data Staging Area*:

```
SELECT
    IDENTITY(int, 1, 1) as ID_PRODUTO,
    B1_COD AS CODIGO,
    B1_DESC AS DESCRICAO, B1_GRPSM AS GRUPO,
    B1_TAMANHO AS TAMANHO, B1_MARCA AS MARCA
INTO DIM_PRODUTO
FROM SB1010
WHERE D_E_L_E_T_ = ''
```

- Etapa ETL de transformação, limpeza e criação da dimensão Data:

```
SELECT
    IDENTITY(int, 1, 1) AS ID_DATA,
    SUBSTRING(M2_DATA, 1, 4) AS ANO,
    SUBSTRING(M2_DATA, 5, 2) AS MES,
    SUBSTRING(M2_DATA, 7, 2) AS DIA
INTO DIM_DATA
FROM SM2010
WHERE
    D_E_L_E_T_ = '' AND
    M2_DATA BETWEEN '20040101' AND '20060101'
ORDER BY M2_DATA
```

Obs: Durante este Processo *ETL*, na etapa de carga do DM, foram encontrados custos zerados em três produtos. Este problema foi resolvido através de uma limpeza em tais dados. Para se efetivar esta solução foi realizada uma conversa com os usuários onde ficou decidido, por sugestão do autor-desenvolvedor, atualizar tais dados baseando-se no valor correto destes custos após consulta ao setor de Estoque e Custos da Empresa. Esta decisão foi implementada via o comando *UPDATE* da linguagem SQL.

Outros códigos *Transact-SQL* de transformação (e limpeza) e criação de outras tabelas do *Data Mart Faturas* estão disponíveis no APÊNDICE A deste documento.

5.2.4 Screenshot dos dados do *Data Mart Faturas*

Para se fazer uma experimentação do resultado alcançado com a finalização do processo *ETL* para o *Data Marta Faturas*, a Figura 8 apresenta um *screenshot* dos dados após serem carregados em tal *Data Mart*. A Figura 8 exhibe os 10 (dez) primeiros registros do *Data Mart Faturas* ordenados por preço médio de venda:

```
SELECT TOP 10
  ID_FILIAL, ID_ARMAZEM, ID_VENDEDOR,
  ID_PRODUTO, ID_DATA, PRECO_MEDIO
FROM FATO FATURAS
ORDER BY PRECO_MEDIO DESC
```

| | ID_FILIAL | ID_ARMAZEM | ID_VENDEDOR | ID_PRODUTO | ID_DATA | PRECO_MEDIO |
|----|-----------|------------|-------------|------------|---------|-------------|
| 1 | 1 | 13 | 35 | 651 | 432 | 2190.0 |
| 2 | 1 | 13 | 35 | 711 | 416 | 1900.0 |
| 3 | 4 | 34 | 12 | 1133 | 487 | 1260.0 |
| 4 | 6 | 40 | 21 | 1125 | 472 | 1250.0 |
| 5 | 6 | 40 | 21 | 1125 | 509 | 1250.0 |
| 6 | 6 | 40 | 18 | 1125 | 486 | 1200.0 |
| 7 | 6 | 40 | 18 | 790 | 485 | 1200.0 |
| 8 | 6 | 40 | 18 | 790 | 486 | 1200.0 |
| 9 | 6 | 40 | 18 | 1125 | 478 | 1200.0 |
| 10 | 6 | 40 | 18 | 1125 | 472 | 1200.0 |

Figura 8 – Screenshot dos dados do *Data Mart Faturas*.

5.3 Análise e Experimentação de Interfaces de Acesso e Análise de Dados no *Data Mart Faturas*

Para complementar o trabalho e validar o funcionamento do *Data Mart* implementado, foram pesquisadas, avaliadas, preparadas e experimentadas as seguintes ferramentas de acesso e análise de dados no *Data Mart Faturas*:

Tabela 8 – Análise de Interfaces de acesso e análise de dados no *Data Mart Faturas*.

| Ferramenta | Vantagens | Desvantagens |
|---|--|---|
| <i>Cube Browser</i> . [40] | Custo adicional zero pois é uma ferramenta nativa do programa <i>Analysis Manager</i> o qual vem no <i>Microsoft Analysis Services</i> do <i>SQL Server 2000</i> | - Relativa baixa usabilidade da ferramenta - Não é uma ferramenta WEB |
| <i>AnalyzeOnDemand 8.3.1.122</i> – fornecido pela empresa SOLONDE [1] | Alta Funcionalidade, como consultas OLAP, Query, Relatórios, Dash Boards e Gráficos. | - Custo de U\$ 1.995,00 ¹ |
| <i>OWC (Office Web Component)</i> - da Microsoft [29] | -Viabiliza acesso Web -Possibilidade de salvar cubos para análise off-line | -Média usabilidade -necessário baixar segurança no navegador para usá-lo |
| <i>Microsoft Excel</i> [28] | Custo adicional zero por ser nativa do <i>Microsoft Office</i> , alta facilidade de uso | Pouca escalabilidade (Limitação de tamanho dos dados a serem analisados) |

¹ – Recuperado de <https://www.solonde.com/opt/buy.php> on-line em 27 de julho de 2005.

Para ilustrar o uso de uma das interfaces acima, na Figura 9 encontra-se um gráfico com o total faturado Por Filial, Por Estado, obtido através da ferramenta *AnalyzeOnDemand* [1] consultando o *Data Mart Faturas*.

Análise, Projeto e Implementação de um *Data Mart* para Ambientes de *Business Intelligence*

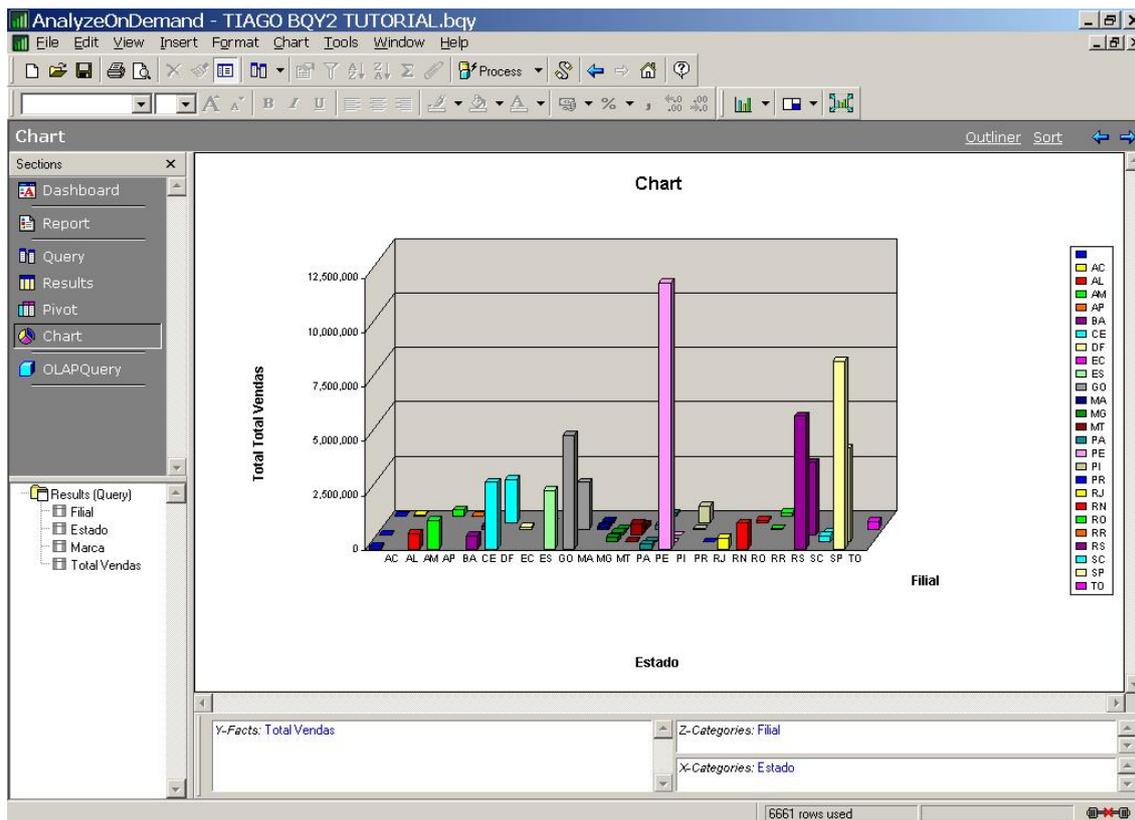


Figura 9 – Exemplo de gráfico de uma das ferramenta de BI experimentadas.

No APÊNDICE B desta monografia encontram-se *screenshots* das interfaces *Cube Browser* e *OWC (Office Web Component)*.

Neste ponto do trabalho o *Data Mart Faturas* está pronto para ser utilizado e amplamente explorado pela empresa para auxiliá-la a tomar decisões de negócio com mais certeza e garantia de sucesso, caracterizando, desta maneira, um interessante ambiente de *Business Intelligence*.

6. Conclusões e Trabalhos Futuros

Na concepção deste trabalho, o principal objetivo proposto inicialmente consistia na construção de um *Data Mart* para *Business Intelligence*. Naquele momento, não estava planejada uma pesquisa sobre o estado da arte em BI, nem pesquisa, análise, preparação e experimentação de quatro ferramentas de acesso e análise de dados em *Data Warehouse* (ou *Data Mart*). Em contrapartida, fazendo-se uma avaliação objetiva ao término do trabalho, percebe-se, entre os resultados positivos, que a pesquisa sobre o estado da arte em BI fez parte do trabalho e foi realizada de maneira efetiva, contribuindo para o esclarecimento dos conhecimentos específicos e necessários para a etapa seguinte, a de implementação. Além disso, o *Data Mart* foi projetado e implementado, atendendo aos requisitos levantados junto aos usuários e alcançando, satisfatoriamente, o principal objetivo inicialmente definido. Adicionalmente, foram pesquisadas, avaliadas, preparadas e experimentadas quatro ferramentas de acesso e análise aos dados do *Data Mart*, conforme descritas na seção 5.3 e apêndice B deste relatório. Finalizando-se os resultados positivos, constatou-se que a solução técnica obtida encontra-se alinhada ao estado da arte em *Business Intelligence*, conforme descrito na seção 3.3 deste trabalho. Tais resultados propiciam benefícios como auxílio a tomadas de decisões com mais garantia e sucesso, além da liberação dos sistemas transacionais para a realização exclusiva das transações operacionais de controle da empresa, obtendo-se ganho de performance, bem como viabilizam a diminuição da necessidade de profissionais de TI para construção de relatórios estáticos, reduzindo-se custos com tecnologia.

Por outro lado, fazendo-se uma crítica objetiva deste trabalho, dentre os resultados negativos percebidos está a predominância, e conseqüente dependência, de um único fornecedor, no caso a *Microsoft*, nas tecnologias utilizadas (SGBD, ferramenta ETL, camada OLAP). Além desse, percebe-se que não foram realizados testes e validações de maneira exaustiva aos dados carregados no *Data Mart*.

Dentre as principais dificuldades vivenciadas ao longo do trabalho está a presença de ruídos nos dados do ambiente operacional, como no exemplo de produtos com custos zerados, mostrado na seção 5.2.3 deste documento, o que,

do ponto de vista técnico, dificulta o processo *ETL* e, do ponto de vista do usuário, se não tratados, diminui a qualidade dos dados armazenados no *Data Mart* e visualizados no ambiente de BI. Além dessa, outra dificuldade é a manipulação de fatos não-aditivos (que não podem ser somados) pois, por exemplo ao se consultar os campos *preço médio, custo médio e margem de lucro* numa ferramenta OLAP, tais valores podem ser visualizados de maneira incorreta, dependendo da abordagem utilizada para calculá-los [22]. Por fim, está a dificuldade de se manter a integridade e consistência dos dados e uma boa performance das consultas ao *Data Mart*.

Como sugestões para trabalhos futuros seria importante, para melhorar os resultados alcançados, a realização de testes exaustivos sobre o ambiente construído, por exemplo, utilizando-se a abordagem de cruzamento dos dados fornecidos pelo *Data Mart* com dados fornecidos por relatórios obtidos no sistema ERP da empresa. Este trabalho propiciaria uma validação das funcionalidades implementadas e a constatação do nível de qualidade das informações no *Data Mart*, que é um fator crítico para o sucesso do projeto [19]. Além desse, outro importante trabalho de melhoria seria a construção de aplicações de mineração de dados (*data mining*) para exploração do *Data Mart* implementado e viabilizar, por exemplo, o entendimento de padrões e tendências no negócio. Outra iniciativa de trabalho futuro, de natureza empreendedora e que caracterizaria a utilização da solução obtida para resolver problemas de suporte à decisão em empresas, seria considerar o fato do projeto se encontrar especificado e implementado e concentrar esforços para formalizar o projeto, de maneira que o mesmo evoluísse para ser implantado num ambiente real de negócio de uma empresa cliente.

Como avaliação final do trabalho e de seus resultados, felizmente se estabeleceu o fato de que o projeto avançou barreiras, pois além de se realizar uma pesquisa sobre o estado da arte em *Business Intelligence* e levantar os requisitos, analisar, projetar e implementar um *Data Mart* para ambientes de BI, também foram pesquisadas, analisadas, preparadas e experimentadas quatro ferramentas de acesso e análise dos dados no *Data Mart*. Com isso, construiu-se um ambiente completo de utilização e exploração de informações reais de negócio para tornar a *Business Intelligence* presente na empresa e permitir que o ciclo de conhecimento profundo e aprendizado contínuo sobre o negócio seja alcançado. Sendo assim, o ambiente tecnológico e informacional está

Análise, Projeto e Implementação de um *Data Mart* para
Ambientes de *Business Intelligence*

implementado para, após ser validado, auxiliar decisões com mais rapidez, garantia e sucesso. O próximo passo depende do talento dos seres humanos que utilizarão esta excelente ferramenta, a qual permite-lhes transformar dados em informação, informação em conhecimento profundo, conhecimento profundo em decisões otimizadas e decisões otimizadas em bons resultados e sucesso nos negócios.

7. Referências

- [1]. *AnalyzeOnDemand 8.3.1.122*.
http://cms.solonde.com/cms/front_content.php?idcat=32, on-line em 27 de julho de 2005.
- [2]. Área de *Business Intelligence* do DwBrasil.
<http://www.dwbrasil.com.br/html/bi.html>, on-line em 04 de agosto de 2005.
- [3]. Área de Data Mart do DwBrasil. <http://www.dwbrasil.com.br/html/dmart.html>, on-line em 14/04/2005.
- [4]. Banco Bradesco S/A. <http://www.bradesco.com.br>, on-line em 09 de agosto de 2005.
- [5]. Bancos têm lucro recorde – edição do dia 08 de agosto de 2005 do Jornal da globo.
<http://jg.globo.com/JGlobo/0,19125,VTJ0-2742-20050808-106022,00.html>, on-line em 09 de agosto de 2005.
- [6]. "Bradesco Seguros - a informação a serviço dos negócios" Case João Penso, Gerente Business Intelligence, Bradesco Seguros.
<http://64.233.187.104/search?q=cache:ZtgFRpPi4EQJ:www.ideti.com.br/inteligenciaorganizacional/rj/palestras.htm+bradesco+%22business+intelligence%22&hl=pt-BR>, on-line em 09 de agosto de 2005.
- [7]. *Business Intelligence - Gartner Group - Datapro Summary*.
www2.ufp.pt/~lmbg/formacao/mestrados/bintelligence.PDF, on-line em 12 de agosto de 2005.
- [8]. Caraciolo, J. Modelagem Incremental De um Ambiente de Data Warehousing: Uma Abordagem da Arquitetura Bottom-up. Trabalho de Graduação. 2004. Centro de Informática-UFPE.
- [9]. Definições de Business Intelligence do site Conhecimento Empresarial.
http://www.conhecimentoempresarial.com.br/define_bi.htm, on-line em 11 de julho de 2005.

- [10]. Definição de *Business Intelligence* do site *whatis.com*, http://searchcrm.techtarget.com/sDefinition/0,290660,sid11_gci213571,00.html, on-line em 12 de julho de 2005.
- [11]. Dias, M. Mattos, M. Romão, W. Todesco, J. Pacheco, R. Pesquisas Futuras em Data Warehouse. <http://www.cin.ufpe.br/~crc/antigo/TBD2/PesquisaFuturasEmDatawarehouse.pdf>, on-line em 18 de julho de 2005.
- [12]. Diniz, D. Parente, D. Vieira, F. Course DW. http://www.owg.com.br/html/body_course_dw.html, on-line em 11 de agosto de 2005.
- [13]. Drucker, P. (1995). Introdução à administração. (3ª edição). São Paulo: Pioneira.
- [14]. *ETL – Computerworld February 02, 2004*, <http://www.computerworld.com/databasetopics/businessintelligence/datawarehouse/story/0,10801,89534,00.html>, on-line em 16 de julho de 2005.
- [15]. Heinrichs, J. H.; Lim, J. S. *Integrating web-based data mining tolls with business models for knowledge management. Decision Support Systems*, v. 35, n. 1, p. 103-112, 2003.
- [16]. IDC Brasil aponta crescimento no segmento de software de *Business Intelligence*. http://www.idcbrasil.com.br/brasil/telas/pagina.asp?id_area=3&n=56, on-line em 14 de agosto de 2005.
- [17]. Fortulan, M. Gonçalves, E. Uma Proposta de Aplicação de Business Intelligence no Chão-de-fábrica, www.scielo.br/pdf/gp/v12n1/a06v12n1.pdf, on-line em 19 de julho de 2005.
- [18]. Gurovitz, H. (1997). O Que Cerveja Tem a Ver Com Fraudas? Revista Informática Exame. Ano 12, nº 139. Outubro/1997.
- [19]. Hackathorn, R. (1998). *Data warehousing's credibility crisis*. Byte, Aug., 1997. www.byte.com/art/9708/sec4/art1.htm, on-line em 18 de julho de 2005.
- [20]. Immon, W., Building the Data Warehouse. Wiley Computer Publishing (2nd Edition), 1996.

- [21]. Kimball, R. et al. *The Data Warehouse Lifecycle Toolkit – Expert Methods for Designing, Developing, and Deploying Data Warehouses*. New York: John Wiley & Sons, Inc., 1998. 771p.
- [22]. Kimball, R. *The Data Warehouse Toolkit: Guia completo para modelagem dimensional*. Tradução de Ana Beatriz Tavares, Daniela Lacerda. - Rio de Janeiro: Campus, 2002.
- [23]. Leme, T. “Business Intelligence no Microsoft Excel”. Axcel Books, 2004.
- [24]. Machado, F. N. R. *Projeto de Data Warehouse: uma visão multidimensional*. Editora Érica, São Paulo, 2000.
- [25]. *Management Update: The Cornerstones of Business Intelligence Excellence. The BI Value Framework (2004)*.
www.gartner.com/2_events/crmawards/2004/images/management_upda.pdf, on-line em 14 de agosto de 2005.
- [26]. *Microsoft Corporation*. <http://www.microsoft.com>, on-line em 26 de julho de 2005.
- [27]. Microsoft Lança software para BI.
<http://www.itweb.com.br/noticias/artigo.asp?id=91717>, on-line em 14 de agosto de 2005.
- [28]. Microsoft Office OnLine: Excel. <http://office.microsoft.com/pt-br/FX010858001046.aspx>, on-line em 28 de julho de 2005.
- [29]. Microsoft Office Web Components (OWC).
<http://www.microsoft.com/downloads/details.aspx?FamilyID=982b0359-0a86-4fb2-a7ee-5f3a499515dd&displaylang=pt-br>, on-line em 28 de julho de 2005.
- [30]. *Microsoft SQL Server: Data Transformation Services (DTS)*.
<http://www.microsoft.com/sql/evaluation/features/datatran.asp>, on-line em 26 de julho de 2005.
- [31]. Mosquéra, M. *Dissertação de Mestrado - Universidade Federal de Santa Catarina. A Implementação e Utilização de Data Warehouse em Instituições públicas no Brasil: Um Estudo Descritivo Das Implicações Envolvidas*.
<http://teses.eps.ufsc.br/defesa/pdf/7377.pdf>, on-line em 18 de julho de 2005.

- [32]. Quarles, D. *Business Intelligence - The Umbrella Term*. Final report of a field exploration. Faculteit der Exacte Wetenschappen-vrije Universiteit Amsterdam. www.few.vu.nl/stagebureau/werkstuk/werkstukken/werkstuk-quarles.doc, on-line em 11 de julho de 2005.
- [33]. Santos, R. Ângelo F., *Fastcube: Uma Metodologia de Construção de Data Mart*. http://infodw.port5.com/aula_dm.pdf, on-line em 07 janeiro de 2005.
- [34]. Shim, J. P.; Warkentin, M.; Courtney, J.; Power, D. J.; Sharda, R.; Carlsson, C. Past, present, and future of decision support technology. *Decision Support System*, v. 33, n. 2, p. 111-126, 2002.
- [35]. Software DIC Michaelis - Moderno Dicionário da Língua Portuguesa.
- [36]. *SQL Server Books OnLine – Help do MS SQL Server 2000*.
- [37]. *SQL Server Home*. <http://www.microsoft.com/sql>, on-line em 25 de julho de 2005.
- [38]. Souza, M. iMasters - Business Intelligence, Melhorando nossos conceitos em Datawarehouse (3ª parte), <http://www.imasters.com.br/artigo.php?cn=1802&cc=59>, on-line em 18.07.2005.
- [39]. Tan, X.; Yen, D. C.; Fang, X. *Web warehousing: web technology meets data warehousing. Technology in Society*, v. 25, n. 1, p.131-148, 2003.
- [40]. *The Gartner Fellows: Howard Dresner's Biography*, http://www.gartner.com/research/fellows/asset_79427_1175.jsp, on-line em 16 de julho de 2005.
- [41]. *The OLAP Report: What is OLAP?*, <http://www.olapreport.com/fasmi.htm>, on-line em 16 de julho de 2005.
- [42]. Vieira, P. Introdução a *Business Intelligence*. Aula da disciplina Economia e Estratégia para Empreendedores do Cin UFPE.
- [43]. Viewing Cube Data - Cube Browser.
http://msdn.microsoft.com/library/default.asp?url=/library/en-us/olapdmad/agcubesview_428x.asp, on-line em 10 de agosto de 2005.
- [44]. Zênite Soluções em TI & comércio LTDA. www.zenite.inf.br, on-line em 18 de maio de 2005.

Apêndice A – Código *SQL* do processo *ETL* do *Data Mart Faturas*

Etapa ETL de transformação, limpeza e criação da dimensão Cliente:

```
SELECT
    IDENTITY(int, 1, 1) AS ID_CLIENTE, A1_FILIAL AS FILIAL,
    A1_COD AS CODIGO, A1_NOME AS NOME, A1_EST AS ESTADO
INTO DIM_CLIENTE
FROM SA1010
WHERE D_E_L_E_T_ = ''
```

Etapa ETL de transformação, limpeza e criação da dimensão Produto

```
SELECT
    IDENTITY(int, 1, 1) as ID_PRODUTO, B1_COD AS CODIGO,
    B1_DESC AS DESCRICAO, B1_GRPSM AS GRUPO,
    B1_TAMANHO AS TAMANHO, B1_MARCA AS MARCA
INTO DIM_PRODUTO
FROM SB1010
WHERE D_E_L_E_T_ = ''
```

Etapa ETL de transformação, limpeza e criação da dimensão Armazém:

```
SELECT
    IDENTITY(int, 1, 1) as ID_ARMAZEM, Y2_FILORI AS FILIAL,
    Y2_SIGLA AS SIGLA, Y2_DESC AS DESCRICAO
INTO DIM_ARMAZEM
FROM SY2010
WHERE D_E_L_E_T_ = ''
```

Etapa ETL de transformação, limpeza e criação da dimensão Vendedor:

```
SELECT
    IDENTITY(int, 1, 1) as ID_VENDEDOR, A3_FILIAL AS FILIAL,
    A3_COD AS CODIGO, A3_NOME AS NOME
INTO DIM_VENDEDOR
FROM SA3010
WHERE D_E_L_E_T_ = ''
```

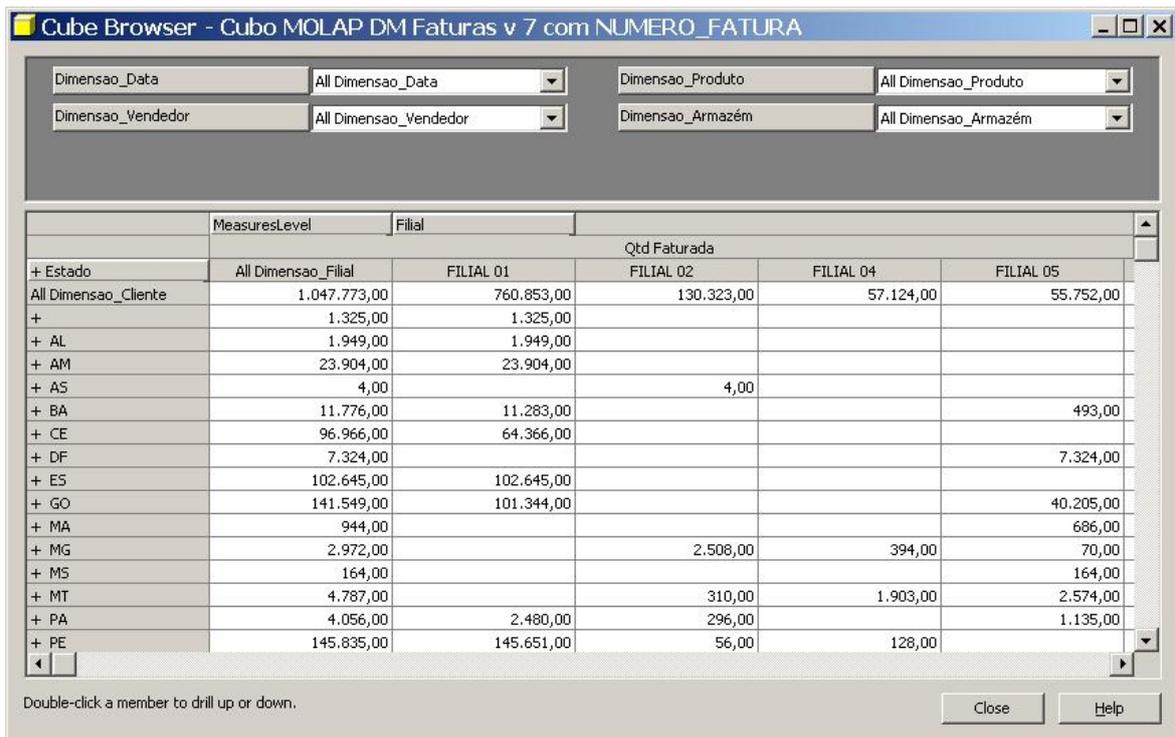
Análise, Projeto e Implementação de um *Data Mart* para
Ambientes de *Business Intelligence*

Etapa ETL de transformação e criação da dimensão Filial:

```
SELECT 1 AS ID_FILIAL, 'FILIAL 01' AS FILIAL INTO DIM_FILIAL
INSERT INTO DIM_FILIAL VALUES (2, 'FILIAL 02')
INSERT INTO DIM_FILIAL VALUES (3, 'FILIAL 03')
INSERT INTO DIM_FILIAL VALUES (4, 'FILIAL 04')
INSERT INTO DIM_FILIAL VALUES (7, 'FILIAL 05')
INSERT INTO DIM_FILIAL VALUES (6, 'FILIAL 06')
```

Apêndice B – Exemplo de Ferramentas de Acesso e Análise de Dados em *Data Warehouse* (ou *Data Mart*)

- a) Ferramenta Cube Browser: Ferramenta OLAP, nativa do MS *Analysis Services*, de acesso e análise aos dados do *Data Mart Faturas*.



The screenshot shows the 'Cube Browser' window with the title 'Cubo MOLAP DM Faturas v 7 com NUMERO_FATURA'. It features four dropdown menus for dimensions: Dimensao_Data, Dimensao_Vendedor, Dimensao_Produto, and Dimensao_Armazém, all set to 'All Dimensao_...'. Below is a pivot table with 'MeasuresLevel' and 'Filial' as row labels, and 'Qtde Faturada' as the column label. The table data is as follows:

| MeasuresLevel | Filial | Qtde Faturada | | | | |
|----------------------|---------------------|---------------|------------|-----------|-----------|--|
| + Estado | All Dimensao_Filial | FILIAL 01 | FILIAL 02 | FILIAL 04 | FILIAL 05 | |
| All Dimensao_Cliente | 1.047.773,00 | 760.853,00 | 130.323,00 | 57.124,00 | 55.752,00 | |
| + | 1.325,00 | 1.325,00 | | | | |
| + AL | 1.949,00 | 1.949,00 | | | | |
| + AM | 23.904,00 | 23.904,00 | | | | |
| + AS | 4,00 | | 4,00 | | | |
| + BA | 11.776,00 | 11.283,00 | | | 493,00 | |
| + CE | 96.966,00 | 64.366,00 | | | | |
| + DF | 7.324,00 | | | | 7.324,00 | |
| + ES | 102.645,00 | 102.645,00 | | | | |
| + GO | 141.549,00 | 101.344,00 | | | 40.205,00 | |
| + MA | 944,00 | | | | 686,00 | |
| + MG | 2.972,00 | | 2.508,00 | 394,00 | 70,00 | |
| + MS | 164,00 | | | | 164,00 | |
| + MT | 4.787,00 | | 310,00 | 1.903,00 | 2.574,00 | |
| + PA | 4.056,00 | 2.480,00 | 296,00 | | 1.135,00 | |
| + PE | 145.835,00 | 145.651,00 | 56,00 | 128,00 | | |

Figura 10 – Ferramenta OLAP *Cube Browser* experimentada no trabalho.

Na Figura 10, está ilustrada uma consulta à quantidade faturada de produtos por Filial por Estado da dimensão Cliente.

b) Aplicação de *WBI*: ferramenta de *Web-BI*, implementada utilizando-se a tecnologia *OWC* (*Office Web Component*).

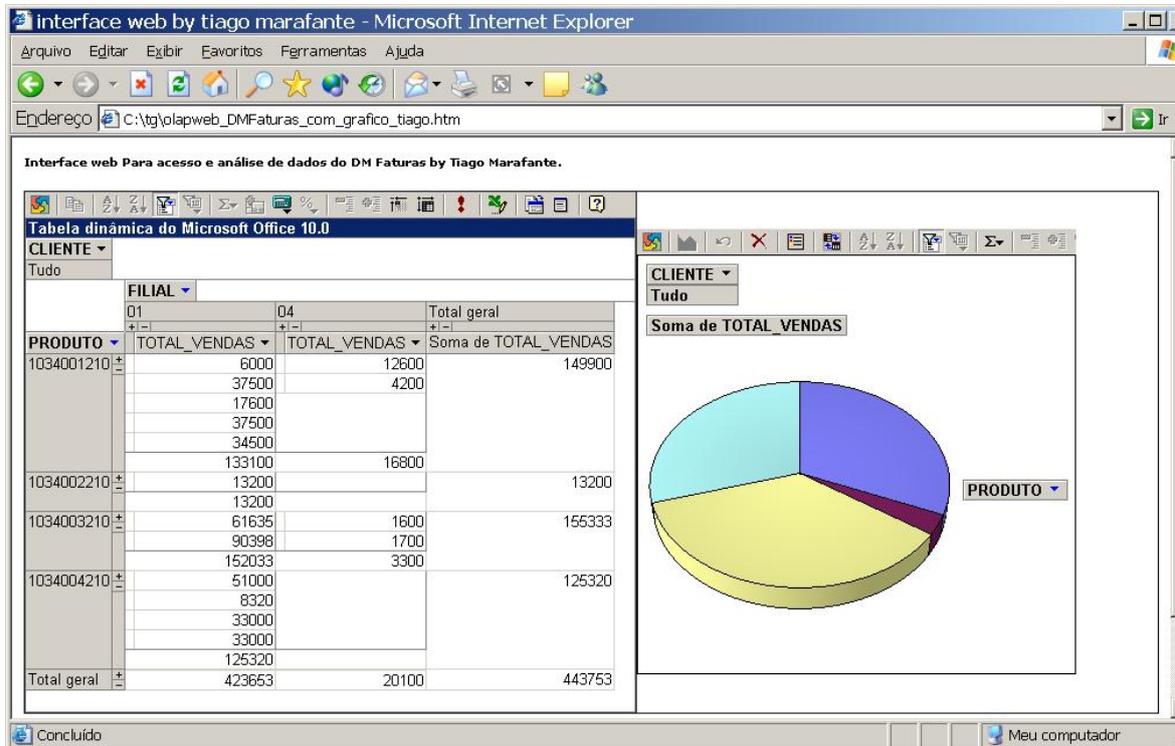


Figura 11 – Aplicação de *Web-BI* construída no trabalho.

Na Figura 11, percebe-se uma característica marcante dos ambientes de BI que é a presença de gráficos para melhorar a facilidade de uso e interpretação das informações fornecidas. No exemplo, o gráfico representa a distribuição do total de vendas faturadas de quatro produtos específicos, para todos os clientes de apenas duas filiais.

8. Datas e Assinaturas

Recife, agosto de 2005,

Tiago Marafante Lins de Souza
Aluno

Ana Carolina Salgado
Orientadora