Danilo de Melo Santos

Fabiane Aragão Dourado

Marcleiton Vilarouca Teixeira

APLICAÇÃO DA TECNOLOGIA DE DATA MART NO AMBIENTE COMERCIAL DA CONCESSIONÁRIA DE VEÍCULOS DFMAuto

Trabalho de Conclusão de Curso apresentado à Pró-Reitoria de Graduação no curso de Bacharelado em Sistemas de Informação da Universidade Católica de Brasília, como requisito para obtenção do título de Bacharel.

Orientador: Vilson Carlos Hartmann

Brasília
2006

Agradecimentos
Por Danilo de Melo Santos;
Chega ao fim uma longa jornada que exigiu muita dedicação, esforço e determinação na busca pelo conhecimento e aprimoramento, mas essa caminhada não poderia ter sido concluída sozinho, durante todo o tempo algumas pessoas estiveram apoiando e sempre dizendo: "O caminho é longo, mas tudo terá recompensa", o que me fazia levantar a cabeça e seguir em frente, mesmo que por vezes a descrença e o cansaço tenham tomado conta.
Agradeço agora às pessoas que estiveram me apoiando e confiando no meu potencial, ajudando a ver realmente a recompensa no final desta batalha e hoje sou grato a eles por terem sempre incitando a gana de lutador.
À minha mãe que, às vezes sem perceber, com pequenas ações me animou e me escutou logo no começo do dia com um bom papo à mesa do café da manhã. Por todas as noites que ela deixou minha comida preparada, minha cama arrumada com carinho por que eu chegaria tarde e cansado. Por continuamente reconhecer minha luta e insistir que o melhor caminho era o que eu estava tomando. Obrigado Mãe pelo amor e cuidado.
Ao meu pai que por maior que fosse a dificuldade sempre me ajudou. Também esteve presente nos meus momentos de fraqueza e soube dizer e fazer a coisa certa para que eu não deixasse tudo escapar entre dedos. Por ter me ensinado valores que levarei por toda a vida e mostrando que ninguém seria melhor que eu se eu não deixasse, que meu futuro estava apenas nos meus atos e no que eu estava construindo. Obrigado Pai, pelo homem que sou e pelo que aprendi com você.
À minha irmã que mesmo sem saber me ajudava com suas besteiras e palhaçadas. Sempre que acordada quando chegava da faculdade procurando um diálogo aparecendo no meu quarto dizendo qualquer besteira, me fazia esquecer dos problemas e dar boas risadas. Nayara, obrigado pelos bons momentos e risadas sem motivo. Amo você.
À minha Família, muito obrigado. Amo todos com muito carinho.
Aos amigos, Maria Fernanda (Fê obrigado pelo apoio, você também tem sua parcela nesse trabalho), Priscila (aprendi muito com você durante essa e outras jornadas), Izabela (obrigado por alguns puxões de orelha), Cris (muitas vezes cheguei na UCB com a cabeça melhor por sua causa), Badu (sem comentários, você é minha irmã). Obrigado a todas vocês.
Escolhi as melhores pessoas para fazer esse projeto. Marcleiton as risadas foram todas muito boas, os momentos de ralação, xingamento, besteirol, tudo sempre ajudou e muitas vezes conseguimos ultrapassar as dificuldades que a Universidade nos mostrava juntos. Valeu amigo, espero que essa amizade seja além destes cinco anos. Fabi, com suas tiradas clássicas e puxões de orelha às vezes nada discretos, mas sempre mostrava que a responsabilidade estava escapando e deixando a brincadeira tomar conta. As broncas foram muitas, mas a admiração por você e o que pude aprender com essa convivência é muito importante. Com vocês aprendi a ter mais determinação e obstinação.
Por fim, mas não menos importante, minha Neguinha. Obrigado por ter tido paciência e compreensão, por ter ensinado paciência e calma nos momentos de maior pressão, por ter me dado paz e serenidade quando o mundo estava desabando a minha volta. Você foi e, ainda é, um alicerce muito importante na minha vida. O que aprendi com você talvez nunca tenha nem ao menos percebido. Obrigado pelo amor, carinho, cuidado e atenção. Desde que apareceu na minha vida você se tornou essencial. Amo você.

Por Marcleiton Vilarouca Teixeira
Meus sinceros agradecimentos vão a todos que de alguma maneira cooperaram com esta minha conquista, não quero cometer a injustiça de deixar ninguém de fora.
Em especial, agradeço a minha Mãe, pessoa que sempre deu força, auxílio em tudo que necessitei, ficou acordada até mais tarde, acordou todos os dias cedo e fazia um café da manhã sempre com muito carinho, sempre se mostrou disposta a me ajudar, sempre torceu pelo meu sucesso, sem dúvidas a pessoa mais importante nesta caminhada.
Um grande e sincero afago a toda minha família, forte e vencedora, muito obrigado.
Agradeço aos amigos: Rodrigo Galdino dos Reis, um dos camaradas com quem consegui cursar algumas disciplinas, demos boas risadas em momentos únicos, deve treinar um pouco mais para jogar boliche comigo; Danilo de Melo Santos, parceiro de projeto final, de ônibus, de revezamento de nossos veículos, e de várias disciplinas, um irmão, desejo que esta amizade rompa a barreira de fim de graduação; Fabiane Aragão Dourado, parceira que intitulamos de Gerente neste projeto, falou o que podia nos momentos certos, muito ímpar essa mulher; João Barbosa França, uma pessoa que não tem nenhum pesar dos pesares, sempre com idéias novas; Alexandre Barbosa dos Santos, quando lembro deste camarada vem a mente esta frase: "Um irmão pode não ser um amigo, mas um amigo será sempre um irmão." Benjamin Franklin. Amigos por quem tenho muito apreço.
Agradeço a minha parceira, Miriane, que sempre me apoiou, deu conselhos de grande valia, estava ao meu lado sempre, torcendo e me auxiliando. Esta caminhada seria muito mais árdua sem ela. Ela com certeza também contribuiu neste trabalho e na minha vida. Amo Você.
Um muito obrigado ao Senhor Deus, que colocou essas pessoas na minha estrada, e me proporcionou tudo de bom, sempre me mostrando o lado bom da vida.

Por Fabiane Aragão Dourado

Chegou ao fim mais uma etapa de muitas que virão, claro, não consegui finalizar sozinha. Certa de que foi fruto de muita força de vontade e dedicação. Através dessa conquista consegui colher muitos frutos bons.

Agradeço em primeiro lugar a Deus, porque mesmo não merecendo, deu me a oportunidade de iniciar este curso e forças para concluí - lo. Tenho certeza que sem Ele não estaria agora escrevendo estas palavras de agradecimentos.

À minha mãe e ao meu pai (in memoriam), por não pouparem esforços para me educar e oferecer as melhores condições para que eu estudasse. A minha mãe, por todos dias acordar bem cedo, preparar um belo café da manhã e me proporcionar todo conforto que muitas vezes nem mereci. Por ela que sempre me ligava ou dormia tão tarde esperando para saber como estava e o que queria comer antes de dormir. Agradeço por todo o amor que tem por mim. A minha mãe e ao meu pai dedico todo conhecimento que adquiri, sem eles não seria nada. Amo muito minha família

Agradeço ao meus irmãos, ao Fábio por muitas vezes me ensinar o que ele mais sabe (graças a ele passei em EDA), por me incentivar sempre estudar e crescer profissionalmente, dando grande exemplo nesse sentido. Ao Fabrício, por falar que tudo é muito fácil, muito simples, sempre levando a vida numa boa, e com suas baboseiras me alegravam muitas vezes.

Ao meu Amor, namorado e amigo, Eduardo, que sempre me incentivou e incentiva muito a atingir minhas metas, me dar força em todos os momentos, me compreende, me alegra e tem muita paciência comigo. Uma pessoa que fez muita diferença na minha vida, transformou-a. Amo você

Aos meus amigos de curso, Marcleitom e Danilo, atrasados, mas muito dedicados. Foi muito bom concluir esse curso em conjunto, vocês foram mais que 10.

Enfim aos meus parentes e amigos que sempre me divertem.

TERMO DE APROVAÇÃO

Trabalho de Conclusão de Curso defendido e aprovado como requisito requisitos fundamental para obtenção do Título de Bacharel em Sistemas da Informação, em 24 de julho de 2006 pela banca examinadora constituída por:

_____

_____

_____

Brasília
UCB

RESUMO

Empresas de diversos ramos utilizam a tecnologia da informação para prover uma maior facilidade na realização de suas atividades de tomada de decisões. O ambiente de *Data Mart*, aliado a tecnologias de ETL e ferramentas OLAP permite ao tomador de decisões elaborar consultas analíticas a fim de gerar informações com alto grau de valor agregado, contudo, essa tecnologia por si só, não representa vantagem competitiva, devendo ser encarada apenas como um facilitador deste processo. Para a tecnologia de *Data Mart* existem três possíveis saídas: ferramentas OLAP, *Data Mining* ou relatórios estáticos. No presente projeto será desenvolvido um *Data Mart* que, aliado a uma ferramenta OLAP, será aplicado no setor de vendas da Concessionária de Veículos DFMAuto, que, motivada pela dificuldade em realizar consultas que efetivamente gerassem conhecimento para a organização, solicitou a concepção de um *Data Mart* que viabilizasse a construção dessas consultas de forma ágil e eficaz para a tomada de decisões. O estudo da viabilidade de atendimento da solicitação da DFMAuto demonstra diversas particularidades do ambiente de *Data Mart*, como por exemplo: aplicabilidade, estado da arte, capacidade de gerar informações, componentes e arquitetura. Para análise da viabilidade e validade do esforço para implantação desse ambiente, realizou-se a elaboração de relatórios sobre questionamentos de cunho gerencial feitos pela empresa DFMAuto. Desta forma o trabalho demonstra como a tecnologia de *Data Mart* permite ao gerente de projeto galgar suas decisões sobre uma base informacional consistente ao invés de seguir apenas sua experiência.

PALAVRAS-CHAVE: *Data Mart, On Line Analytical Processing.*

ABSTRACT

Companies of many types use information technology to provide a better facility in the accomplishment of their activities of taking decisions. The environment of Data Mart, allied to the ETL technologies and OLAP tools allows the decision taker to elaborate analytical consultations in order to generate information with a high degree of added value, however, this technology by itself does not represent competitive advantage and it must be faced only as a facilitator of this process. For the technology of Data Mart there are three possible exits: OLAP tools, Data Mining or static reports. In the present project it will be developed a Data Mart that allied to an OLAP tool, will be applied in the sales sector of the DFMAuto Vehicles Concessionaire, that motivated by the difficulty to execute seaches that effectively generated knowledge for its organization, requested the conception of a Data Mart that made possible the construction of these searches in a fast and efficient way to take decisions. The study of the viability to attend for the request of the DFMAuto demonstrates many particularitities of the Data Mart environment, for example: applicability, art state, capacity to generate information, components and architecture. For analysing the viability and validity of the effort for the implantation of this environment, reports were elaborated on questionings about managemental aspects made by the DFMAuto Company. In this way, the work demonstrates how the technology of Data Mart allows the project manager to take his decisions on a consistent informacional base instead of following only his experience.

KEY WORDS: *Data Mart*, *On Line Analytical Processing.*

SUMÁRIO

CAPÍTULO 1 ..... **14**

1.1 MOTIVAÇÃO **14**
1.2 BREVE HISTÓRICO **14**
1.3 PROBLEMAS DIAGNOSTICADOS **15**
1.4 DEPARTAMENTOS BENEFICIADOS **16**
1.5 OBJETIVOS DO TRABALHO **16**
1.5.1 OBJETIVO GERAL *16*
1.5.2 OBJETIVOS ESPECÍFICOS *17*
1.6 RECURSOS UTILIZADSOS **17**
1.7 RESULTADOS ESPERADOS **18**
1.8 ÁREAS AFETADAS PELA IMPLEMENTAÇÃO DO PROTÓTIPO **18**
1.9 CONSIDERAÇÕES SOBRE O CAPÍTULO **19**

CAPÍTULO 2 – BUSINESS INTELLIGENCE **20**

2.1 CONCEITUAÇÃO **20**
2.2 DADOS OPERACIONAIS x DADOS INFORMACIONAIS **21**
2.3 CARACTERÍSTICAS DO B.I **22**
2.4 COMPONENTES DE UM AMBIENTE DE B.I **23**
2.5 VENDAS E VALOR DAS INFORMAÇÕES X SEGURANÇA E PRIVACIDADE **24**
2.6 CONSIDERAÇÕES SOBRE O CAPÍTULO **25**

CAPÍTULO 3 – DATA WARE HOUSE E DATA MART **27**

3.1 *DATA WAREHOUSE:* CONCEITOS E FUNDAMENTAÇÃO **28**
3.2 VANTAGEM COMPETITIVA **29**
3.3 ARQUITETURA *DATA WAREHOUSE* **30**
3.4 OPÇÕES DE IMPLEMENTAÇÃO DO DW **34**
3.5 *DATA MART* **35**
3.6 PROCESSO DE EXTRAÇÃO, TRANSFORMAÇÃO E CARGA-ETL **38**
3.7 ODS – *OPERATION DATA STORE* **40**
3.7.1 ODS x DATA WAREHOUSE *41*
3.8 METADADOS **42**
3.8.1 CONCEITOS *43*
3.8.2 METADADOS E DATA WAREHOUSE/DATA MART *44*
3.8.3 O QUE GERA METADADOS? *45*
3.8.4 REPOSITÓRIO DE METADADOS *46*
3.9 CONSIDERAÇÕES SOBRE O CAPÍTULO **48**

CAPÍTULO 4 – MODELAGEM MULTIDIMENSIONAL **49**

4.1 MODELO DE DADOS **49**
4.1.1 MODELO RELACIONAL *50*
4.1.2 MODELAGEM DIMENSIONAL *51*
4.2 PASSOS PARA CONSTRUÇÃO DE UM MODELO MULTIDIMENSIONAL **52**
4.2.1 DEFINIÇÃO DA ÁREA DE NEGÓCIO *53*
4.2.2 DEFINIÇÃO DA GRANULARIDADE *53*
4.3 DEFINIÇÃO DAS TABELAS **54**
4.3.1 TABELA FATO *54*
4.3.2 DIMENSÃO *55*
4.4 NORMALIZAÇÃO DAS TABELAS **57**
4.4.1 ESQUEMA ESTRELA *57*
4.4.2 ESQUEMA FLOCO DE NEVE *59*
4.5 ABORDAGEM DE BILL INMON E RALPH KIMBALL **59**
4.6 CONSIDERAÇÕES SOBRE O CAPÍTULO **60**

CAPÍTULO 5 - *ON LINE ANALYTICAL PROCESSING* - **61**

OLAP **61**

5.1 FUNÇÕES DO OLAP **61**
5.2 VARIAÇÕES DO OLAP **64**
5.3 OLAP x OLTP **65**
5.4 CONSIDERAÇÕES SOBRE O CAPÍTULO **67**

CAPÍTULO 6 – *DATA MINING* **68**

6.1 CONCEITO DE *DATA MINING* **68**
6.1.1 IMPORTÂNCIA DO DATA MINING EM UM DATA WAREHOUSE *68*
6.1.2 PROCESSO DO DATA MINING *69*
6.1.3 TÉCNICAS DE DATA MINING *70*
6.2 CONSIDERAÇÕES SOBRE O CAPÍTULO **71**

CAPÍTULO 7 – ESTUDO DE CASO **72**

7 PROBLEMA ESTUDADO **72**

7.1 BASES DE DADOS ATUAIS **72**
7.2 PROTÓTIPO **73**
7.3 PROCLARITY **74**
7.4 DELIMITAÇÃO DO ESCOPO **74**
7.4.1 ALGUNS QUESTIONAMENTOS *75*
7.5 DEFINIÇÃO DAS DIMENSÕES **76**
7.5.1 DIMENSÃO CARROS *76*
7.5.2 DIMENSÃO CLIENTES *76*
7.5.3 DIMENSÃO PROMOÇÕES *77*
7.5.4 DIMENSÃO VENDEDORES *77*
7.5.5 DIMENSÃO TEMPO *77*
7.6 TABELA FATO: VENDAS **77**
7.7 MODELO DO NEGÓCIO – DATA MART VENDAS – ESQUEMA ESTRELA **79**
7.8 METADADOS – DATA MART DE VENDAS **80**
7.9 DEFINIÇÃO DA GRANULARIDADE **84**
7.10 PROCESSO DE CRIAÇÃO DO CUBO **84**
7.11 APLICAÇÃO – ORGANIZAÇÃO DO CUBO **85**
7.12 APLICAÇÃO – CONSULTAS **88**
7.13 SEGURANÇA DA INFORMAÇÃO **99**
7.14 TREINAMENTO **100**
7.15 TRABALHOS FUTUROS **100**

CONCLUSÃO **102**

SIGLAS E ABREVIAÇÕES **104**

REFERÊNCIAS BIBLIOGRÁFICAS **105**

LISTA DE FIGURAS

Figura 2.1 – Características do Business Intelligence 23
Figura 2.2 – Componentes do BI 24
Figura 3.1 – Modelo Arquitetural do Data Warehouse 31
Figura 3.1 – Representação do Data Mart independente 36
Figura 3.2 – Representação do Data Mart dependente 37
Figura 3.3 – Modelo de Data Mart - Top-down 37
Figura 3.4 – Modelo do Processo de ETL 39
Figura 3.5 – Diferença entre ODS e Data Warehouse 41
Figura 3.6 – Visão simplificada de uma arquitetura de Data Warehouse 44
Figura 3.7 – Repositório de Metadado 47
Tabela 4.2 - Requisitos de processamento transacional e analítico. 52
Figura 4.1 - Hierarquia explícita de Produto. 56
Figura 4.2 – Cubo de dados com 3 dimensões 57
Figura 4.3 – Representação da estrutura do Esquema Estrela. 58
Figura 5.1 – Representação do Drill up e Drill down 62
Figura 5.2 – Representação do processo de Pivot 63
Figura 5.3 – Processo de Slice And Dice 63
Figura 6.1 - Processo do Data Mining 69
Figura 7.1 – Bases de Dados Relacional da DFMAuto 72
Figura 7.2 – Modelagem do Data Mart de Vendas 79
Figura 7.3 – Modelo Estrela do Data Marta da DFMAuto 85
Figura 7.4 – Criação das Dimensões 86
Figura 7.5 – Criação do Cubo de Vendas da DFMAuto 87
Figura 7.6 – Cubo organizado conforme o Esquema Estrela 88
Figura 7.7 – Diagrama de Árvore de Decomposição 89
Figura 7.8 – Gráfico em pizza 90
Figura 7.9 – Gráfico em barras 91
Figura 7.10 - Tabela de consulta 92
Figura 7.11 – Tabela de Consulta para questionamento nº 3 93
Figura 7.12 – Tabela de Consulta para questionamento nº 4 94
Figura 7.13 – Árvore de decisão sobre a quantidade de vendas por período 95
Figura 7.14 – Tabela de consulta para o questionamento nº 6 96
Figura 7.15 – Tabela de consulta para o questionamento nº 7 97
Figura 7.16 – Gráfico em barras para responder ao questionamento nº 10 98
Figura 7.17 – Gráfico em barras para responder ao questionamento nº 9 99

LISTA DE TABELAS

Tabela 2.1 - Comparação entre dados de natureza operacional e informacional *20*
Tabela 3.1 – Comparação entre ODS, Data Warehouse, Data Mart *41*
Tabela 4.1 - Comparação entre modelo Relacional – E/R e modelo Dimensional *49*
Tabela 5.1 – Comparativo MOLAP x ROLAP *64*
Tabela 5.2 – Comparativo entre bancos operacionais e Data Warehouse *65*

INTRODUÇÃO

A evolução histórica dos valores pelos quais a sociedade em geral mede sua capacidade e competitividade mostra que da era industrial aos tempos atuais a informação assume um papel importante neste cenário.

Saber fazer uso da informação é o grande diferencial das organizações. A Tecnologia da Informação é vista como uma ferramenta estratégica central na busca pela vantagem competitiva. Atualmente, observamos que, na era do conhecimento, uma empresa que detém informação qualificada por si só já estabelece um diferencial competitivo.

A obtenção da informação não é grande problema se pensarmos que há diversos meios para se obter a informação, o problema é analisar a qualidade da informação, o tratamento que deve ser dado à informação para que tenha relevância em um processo de negócio.

A garantia da informação qualificada é um custo, logo, a obtenção da informação vem com um custo agregado, e este não deve ser maior que o retorno que esta informação pode dar à organização. Segundo Karl Albrecht, em um de seus artigos[1], que enfatiza a qualidade da informação, chamada por ele de "A Terceira Revolução da Qualidade", sendo a primeira revolução da qualidade do produto, e a segunda do serviço: "a redução de custos relativos à informação poderia apresentar uma grande oportunidade de aumento no retorno do investimento de várias empresas".

Com base no entendimento da revolução da qualidade surge a necessidade de se livrar e não mais gerar informações sem valor, as informações são armazenadas e não analisadas, muitas vezes são informações que não possuem relevância, mas que mantém um valor alto para ser mantida. Um exemplo clássico é a representação das informações. É possível demonstrar um relatório de dez páginas com apenas um gráfico, em uma página. Aumenta a praticidade e facilidade de se manusear.

Com o surgimento, na década de 90, do conceito para gerenciar as informações dentro de uma organização: o *Data Warehouse*(DW) e *Data Mart*(DM). As raízes da construção de DWs e DMs estão no aprimoramento das tecnologias de bancos de dados, que cada vez mais são capazes de manipular grandes volumes de dados. Muitas organizações

[1] Fonte: http://www.gepeq.dep.ufscar.br/base_conhec_det.php?id=7

construíram DWs ou DMs como solução para integrar dados de diversos bancos de dados operacionais e suportar a tomada de decisão com informações qualificadas (INMON, 1996).

Este trabalho pretende aplicar a tecnologia de DM para demonstrar a importância da análise da informação, através da implantação de ferramentas de extração e formatação dos elementos consolidados em banco de dados, gerando com isto um diferencial competitivo, uma vez que se usa a informação como apoio a tomada de decisão.

CAPÍTULO 1

1.1 MOTIVAÇÃO

Os bancos de dados são de vital importância para as organizações, embora os dados contidos nos mesmos sejam de difícil análise. Tudo isso porque, geralmente, grandes empresas detêm um grande volume de dados que estão armazenados em vários sistemas diferentes, não permitindo estruturar informações que permitam tomar decisões embasadas num histórico dos dados consistente.

Com o crescente aumento da competitividade e dinamismo dos negócios, há a necessidade das organizações tomarem decisões rápidas e precisas para alcançar um diferencial competitivo.

Como forma de solucionar esses problemas surgiu no mercado o conceito de *Data Warehouse*(DW) e *Data Marts*(DM) orientados às áreas de negócios da empresa, buscando dar aos gerentes acesso rápido e fácil a informações qualificadas, e a partir desses elementos a camada estratégica obter meios de fazer prospecção do seu negócio, gerando tendências e solucionando problemas.

Essa realidade de carência de informações qualificadas é constatada a partir dos problemas levantados na Empresa DFMAuto, logo há necessidade de implantação dessa tecnologia, com o objetivo de trazer vantagem competitiva e melhorias para o processo de negócio da empresa.

1.2 BREVE HISTÓRICO

Num contexto fictício, a empresa DFMAuto LTDA., atuante nas áreas de compra e venda de carros novos e usados, locação de veículos, venda de peças e acessórios automotivos e prestação de serviços de manutenção, trabalhava com diversas marcas de automóveis, o que dificultava o crescimento almejado pela diretoria da empresa. Isto diminuía o poder de negociação com fornecedores e compradores na obtenção de condições mais vantajosas nas transações comerciais. Não sendo uma forte concorrente no segmento de mercado em que atua, a diretoria optou por trabalhar unicamente com carros novos da marca General Motors.

A empresa utiliza, para os seus processos de tomada de decisão, vários relatórios estáticos, sendo que a cada nova informação, outros relatórios são gerados. Esses relatórios, muitas vezes, mostram-se confusos e gastam um tempo de análise considerável dos gestores.

Para fazer as projeções de compra e venda, a DFMAuto utiliza recursos do software Excel e Word, presente no pacote Microsoft Office. E para armazenamento dos dados é usado um banco de dados transacional. Mas com o grande crescimento observado nesse ramo de atuação, percebeu-se que essas alternativas já não atendiam mais a demanda de informações. Para tal, os diretores da empresa buscam uma forma de agilizar e aperfeiçoar os processos gerenciais já existentes, concluindo, assim, que alguma técnica deve ser utilizada para que seja possível uma melhor organização nos processos da empresa.

1.3 PROBLEMAS DIAGNOSTICADOS

Após uma análise do contexto vigente na DFMAuto, os principais problemas levantados foram:

- A emissão de relatórios para tomada de decisões, é complexa e ineficiente, produzindo uma grande quantidade de dados que, por vezes, não são necessários para o fim desejado;
- As informações contidas nas bases de dados do sistema não são utilizadas para fornecer aos gestores indicadores que auxiliem na tomada de decisão;
- Não há flexibilidade na geração dos relatórios;
- Relatórios gerenciais incompletos, aumentado o risco de erros na tomada de decisão no nível estratégico;

Outros pontos verificados, mas que não estão no escopo do protótipo que será apresentado:

- Os sistemas não oferecem bom desempenho de consulta;

- Os sistemas não incluem os recursos necessários para processar facilmente as questões relativas ao negócio e não oferecem suporte adequado para o armazenamento de grande volume de dados;
- Falta de pessoal qualificado;

1.4 DEPARTAMENTOS BENEFICIADOS

Como a implantação do *Data Mart* tem o intuito de proporcionar um ambiente com informações consistentes, compartilhadas e integradas, para prover às áreas gerenciais de compra e vendas da empresa. Os departamentos beneficiados como sistema serão:

- Gerência de vendas;
- Gerência de compras;
- Gerência de finanças;
- Diretorias.

1.5 OBJETIVOS DO TRABALHO

1.5.1 OBJETIVO GERAL

Compreender a utilização da tecnologia de *Data Mart* e *Business Intelligence*(BI) no processo de negócio de uma organização, bem como toda a arquitetura e passos que envolvem essas tecnologias, voltados para um Ambiente Empresarial de venda de veículos.

No contexto vigente e tendo conhecimento de que a empresa atua numa esfera altamente competitiva, as informações assumem um papel fundamental para o sucesso. Em face da enorme quantidade de informações despejadas nos sistemas existentes na empresa, surge a necessidade da aplicação da tecnologia de *Data Mart* para a principal área da empresa; área de vendas, e o BI que trabalhará os dados referentes a vendas de veículos, gerando relatórios de fácil análise para os processos de tomada de decisão, a fim de colocar a

empresa frente a outras do mesmo ramo e facilitar a geração de relatórios e análise das informações.

1.5.2 OBJETIVOS ESPECÍFICOS

Para o bom desenvolvimento deste trabalho e demonstração das técnicas utilizadas pretende-se:

- Disponibilizar de forma rápida e eficiente, utilizando-se da tecnologia de *Data Mart* as informações existentes nas bases de dados de forma que se tornem úteis aos gestores da DFMAuto em sua tomada de decisão;

- Estudo bibliográfico da tecnologia de *Data Warehouse, Data Mart* e *Business Intelligence*;

- Estudo de caso com aplicação da tecnologia de *Data Mart* e ferramenta OLAP (*On Line Analitcal Processing*);

- Gerar inteligência para a área de venda, como principal geradora de recursos da empresa.

1.6 RECURSOS UTILIZADSOS

Para o desenvolvimento do protótipo serão utilizados:

Software	Descrição	Valor
Micro computador	Pentium 4 Processador: 2.4 Ghz Memória: 40GB de HD, 512MB de RAM	R$ 3.000,00
Microsoft SQL Server 2000 Enterprise 1 Processor – Unlimited Clients	Sistema Gerenciador de Bando de Dados, versão 2000, utilizado para a criação da base de dados da Concessionário DFMAuto.	R$9.260,27
Analysis Services	Ferramenta OLAP Componente do Microsoft SQL Server, utilizada para	Incluso no pacote SQL Server

	geração do cubo de dados.	
DB Designer 4.0	Um software de design que integra modelagem, criação e manutenção de banco de dados, utilizado para a criação do modelo estrela da DFMAuto.	Software Livre
ProClarity Analytics Platform, versão 5.2	Ferramenta para orientar os tomadores de decisão a agirem em tempo real, utilizada para gerar relatórios, gráficos e planilhas com os dados da DFMAuto.	???

Fonte:*http://www.precomania.com/search_techspecs_full.php/masterid=32187*

1.7 RESULTADOS ESPERADOS

Este trabalho pretende demonstrar a viabilidade da implementação de um *Data Mart* e técnicas de BI na área de vendas de concessionárias de veículos. Trazendo, com isso, benefícios para empresa, como: flexibilidade de relatórios, a praticidade de apresentação de tabelas e gráficos, a possibilidade de gerar vários relatórios cruzando dados diversos em minutos, ganhando uma visão muito mais ampla com rapidez. Logo, de posse desses recursos, o gestor poderá identificar os fatores críticos de sucesso na área de vendas e planejar projetos futuros.

O presente trabalho não aborda a interferência da tecnologia implementada nos demais departamentos da empresa, restringindo-se ao departamento de vendas.

1.8 ÁREAS AFETADAS PELA IMPLEMENTAÇÃO DO PROTÓTIPO

Como este trabalho é voltado para a área de vendas de uma concessionária de veículos, as áreas afetadas são as de compra e venda de veículos e áreas financeiras, afetando, conseqüentemente, as funções gerenciais relativas a estas áreas, como gerente de compra,

gerente de vendas e gerente financeiro, dentre outras funções da camada estratégica da empresa, ou seja, com a otimização dos relatórios pressupõe uma melhoria em todos os setores envolvidos.

Alterando, portanto, alguns processos organizacionais da empresa, em virtude da tecnologia adotada.

1.9 CONSIDERAÇÕES SOBRE O CAPÍTULO

Uma breve análise sobre o contexto informacional da empresa DFMAuto demonstra a necessidade de implantação de um ambiente que permita uma melhor análise sobre a área de vendas da empresa. Setor com maior atividade.

Levando-se em conta então essa condição de área crítica da empresa e os problemas descritos no item 1.3 a adoção de um *Data Mart* associado a uma ferramenta OLAP, mostrou-se como escolha ideal para liquidar tais questões e permitir ao gerentes e diretores de vendas a obtenção de informações que possam gerar vantagem competitiva.

O estudo destas tecnologias como objetivo fim deste trabalho permitirá uma melhor compreensão sobre cada uma, além de possibilitar sua adequação fora do ambiente acadêmico, partindo para uma atividade prática dos ensinamentos e conceitos aprendidos durante este projeto.

CAPÍTULO 2 – BUSINESS INTELLIGENCE

A partir da década de 80, a informatização possibilitou grandes saltos operacionais nos negócios das empresas. Grandes sistemas foram construídos, possibilitando transações rápidas e eficientes, e gerando grandes volumes de dados. Isso resultou na dificuldade de extração de informação, ou seja, informações valiosas ficam armazenadas dentro da própria empresa, porém seu acesso é muito difícil. Há uma grande dificuldade de consolidação dessas informações e buscando resolver essas questões será demonstrado nesse capítulo, um estudo sobre *Business Intelligence* (BI).

2.1 CONCEITUAÇÃO

O conceito de *Business Intelligence* (BI) surgiu para resolver problemas como geração de grandes volumes de dados que impedem a extração de informações valiosas. E consiste de uma vasta categoria de tecnologias e programas aplicativos utilizados para extrair, armazenar, analisar e transformar grandes volumes de dados, produzindo um ambiente de conhecimento, onde há produção sistemática de informação gerencial, veloz e consistente, capaz de auxiliar empresas a tomarem as melhores decisões nos negócios, baseado em fatos reais, com profundidade suficiente para descobrir as causas de uma tendência ou de um problema (KAKINOHAMA, 2001).

A essência do BI está relacionada às formas alternativas de tratamento das informações.

O conceito de BI é um dos principais conceitos disponíveis atualmente no que diz respeito a gestão empresarial. Conhecimento do negócio na era da competição global e das comunicações on-line passou a ser chamado de *Business Intelligence* (BI) ou Inteligência de Negócios (IN), que é a utilização de uma série de ferramentas para coletar, analisar e extrair Informações, que serão utilizadas no auxílio ao processo de gestão e tomadas de decisão (SOUZA, 2003).

O ponto importante nessa tecnologia é a empresa poder direcionar todo seu capital intelectual, para a sua devida função, que é pensar. Os gerentes e diretores, poderão ter

as informações rapidamente, e também terão mais tempo para melhorarem todos seus processos e analisarem mais os seus dados, que passarão a ser valiosas informações.

O BI envolve sistemas de *Front End*, como SAD (sistemas de apoio a decisão), EIS (*executive information system*) e ferramentas de consulta analítica OLAP (*on-line analytical processing*); e sistemas de *Back-End*, como *data warehouse*, *data mart* e mineração de dados (*data mining*).

O *Business Intelligence* também pode ser entendido como um guarda chuva conceitual que envolve a Inteligência Competitiva (CI) e a Gerência de Conhecimento (KMS), pesquisa e análise de mercado, relacionados a nova era da Economia da Informação, dedicada a captura de dados, informações e conhecimentos que permitem as organizações competirem com maior eficiência no contexto atual, é um conjunto de ferramentas utilizado para manipular uma massa de dados operacional em busca de informações essenciais para o negócio (SOUZA, 2003).

2.2 DADOS OPERACIONAIS x DADOS INFORMACIONAIS

Dados operacionais e informacionais possuem objetivos diferentes e estão relacionados a sistemas diferentes, sistemas transacionais e sistemas de informações executivas. Ambos são fontes para solidificar os conceitos de BI.

Características	Dados Operacionais	Dados Informacionais
Conteúdo	Valores Correntes	Valores Sumariados, Calculados, Integrados de várias fontes.
Organização dos Dados	Por aplicação/sistema de informação	Por assunto/ Negócio
Natureza dos Dados	Dinâmica	Estática até a atualização dos dados.
Formato das Estruturas	Relacional, próprio para	Dimensional, simplificado,

	computação transacional.	próprio para atividades analíticas.
Atualização dos Dados	Atualização campo a campo	Acesso, sem atualização.
Uso	Altamente estruturado e processamento repetitivo.	Desestruturado, com processamento analítico/heurístico.
Tempo de resposta	Otimizado para 2 a 3 segundos.	Análises mais complexas, com tempos de respostas maiores.

Tabela 2.1 - Comparação entre dados de natureza operacional e informacional

Fonte: (BARBIERI, 2001)

2.3 CARACTERÍSTICAS DO B.I

De acordo com (SAKANAKA, 2001) as principais características do *Business Intelligence* são:

- Capacidade de extrair e integrar dados de múltiplas fontes;
- A transformação dos dados em informações úteis para o contexto da organização;
- Contextualização dos dados;
- Análise de tendências, possibilitando antecipar mudanças ocorridas no mercado;
- Descoberta de novos competidores;
- Desenvolvimento de estratégias e ações competitivas;

Figura 2.1 – Características do Business Intelligence

Fonte: (SAKANAKA, 2001)

2.4 COMPONENTES DE UM AMBIENTE DE B.I

O processo de desenvolvimento de BI envolve estruturas especiais de armazenamento como *Data Warehouses* (*DW*), *Data Marts* (*DM*) e ODS (*Operation Data Store*), com o objetivo de montar uma base de recursos informacionais, capaz de sustentar a camada de inteligência da empresa e possível de ser aplicada aos seus negócios, como elementos diferenciais e competitivos (BARBIERI, 2001).

Na figura 2.2, tem se além dos depósitos de dados, uma camada que é fundamental, ETC – Extração, Transformação e Carga dos dados, que é responsável por toda a preparação e carga dos dados nestes depósitos, *Data Warehouse*, *Data Mart* e ODS.

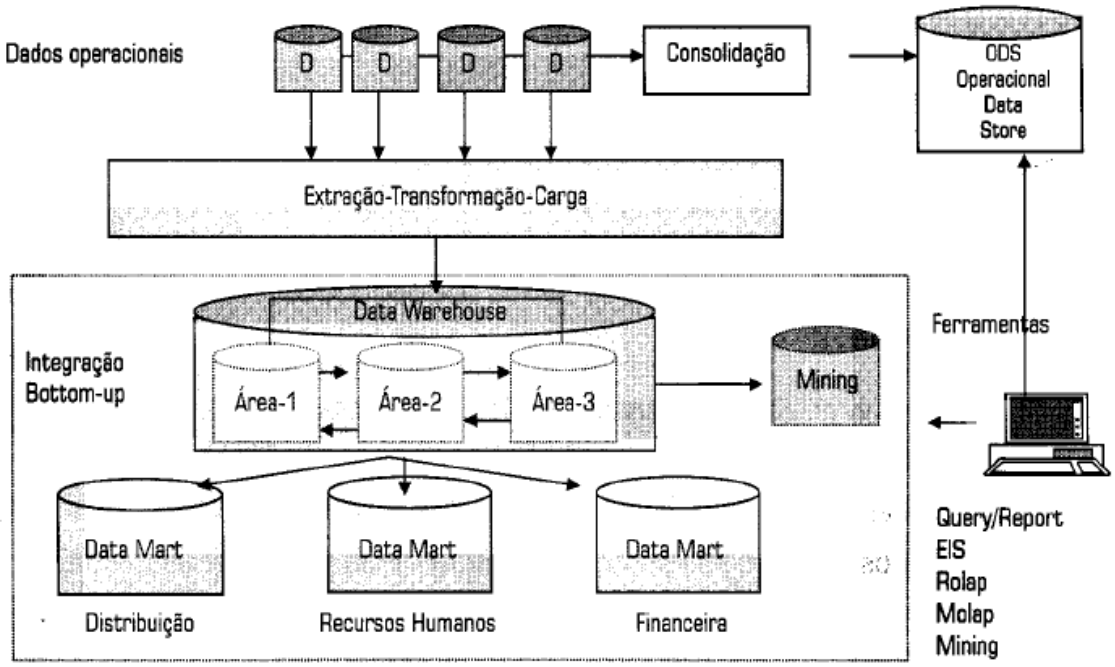


Figura 2.2 – Componentes do BI

Fonte: (BARBIERI, 2001)

2.5 VENDAS E VALOR DAS INFORMAÇÕES X SEGURANÇA E PRIVACIDADE

Segundo (BARBIERI, 2005):

A evolução das estratégias de BI - *Business Intelligence* certamente passará pela fase de comercialização de informações, insumo que valerá ouro nos próximos anos de competições acirradas. No Brasil, este aspecto ainda está pouco desenvolvido, motivado por restrições culturais, plenamente justificado, de possíveis usos indevidos das informações.

A primeira etapa se inicia através da cessão de uso de informações negociais para clientes especiais e posteriormente desembarca na fase de venda, onde cubos de informações poderão ser comercializados no mercado, mantidos, claro, os aspectos críticos de segurança e privacidade de informações.

A idéia é tentar atribuir valores que representem a maior utilização dos dados disponíveis num *mart* ou *warehouse*, tanto para usuários internos, quanto para externos. O usuário externo à empresa que também faz uso da informação pelas propostas de fidelização e aqueles que pagam pelo seu uso, numa relação, agora, de comércio de informação.

O valor da informação armazenado em um *Data Warehouse*/*Data Mart* será gradativamente ampliado na medida em que a democratização do seu uso se intensifique, inclusive para além das fronteiras das organizações.

Aspectos de segurança e privacidade são pontos chaves nessa abordagem, mas não deverão ser impeditivos.

Na medida em que as técnicas de *Business Intelligence* progridem e gradativamente passam a ser aplicadas nas empresa, alguns efeitos colaterais surgem, de mãos dadas com a melhoria da qualidade da informação que passou a ser produzida. O principal cuidado que advém é exatamente os aspectos relacionados com a privacidade e segurança dos dados. Enquanto no mundo transacional, os dados eram somente acessíveis aos administradores de bancos de dados, com a criação dos *Data Warehouse* e *Marts*, as coisas mudam.

Nas novas estruturas de informações dimensionais, os dados, agora transformados em peças valiosas de informações decisórias e comparativas, estão praticamente prontos e acessíveis, sem grandes trabalhos de busca detalhada ou elaboração. Neste momento, as empresas deverão repensar a suas políticas de acesso às informações. Os aspectos de segurança que embasam os sistemas transacionais deverão ser revistos e alguns pontos reforçados.

2.6 CONSIDERAÇÕES SOBRE O CAPÍTULO

Atualmente há a necessidade de integrar todos os dados de um departamento, ou todos os departamentos de uma empresa, qualquer que seja seu porte e segmento em que atua. Pois com a crescente competitividade entre elas implica na necessidade de aplicar conceitos e metodologias de BI .

Aplicando metodologias de BI, a empresa é capaz de transformar dados transacionais em informações estratégicas, definir e acompanhar indicadores de gerenciamento, reduzir custos operacionais, realizar projeções de crescimento, alinhar planos e metas e obter ganhos de produtividade, os quais são objetivos primordiais e necessários num mundo empresarial extremamente competitivo.

Uma das principais tendências atualmente é o aumento do número de projetos de consolidação das ferramentas de BI, voltados ao desenvolvimento de uma estrutura que atenda a organização de maneira abrangente.

CAPÍTULO 3 – DATA WARE HOUSE e DATA MART

No início da década de noventa a informação passou a ter um caráter crítico para o sucesso da empresa, as quais, em sua maioria, estavam despreparadas para trabalhar com este novo cenário.

Buscando agregar valor aos dados distribuídos dentro da empresa o *Data Warehouse* é concebido para permitir a valoração destes dados, a partir de ferramentas de extração de dados e análise analítica, possibilitando uma estruturação clara de informações antes escondidas em montanhas de dados.

Este capítulo trata da tecnologia de *Data Warehousing* que é, por fim, um repositório de dados, tendo como primeiro objetivo a integração efetiva de bases de dados operacionais para possibilitar o uso estratégico dos dados.

De acordo com o IDC (*International Data Corporation*), os pontos mais importantes considerados para a construção de um *Data Warehouse* são:

- A habilidade em focalizar processos empresariais e efetuar uma análise financeira completa destes processos, assim como permitir as organizações tomarem decisões baseadas na compreensão do sistema como um todo;

- A habilidade em racionalizar e automatizar o processo de construção de um repositório de informações integradas para uma empresa ao invés de desenvolver inúmeros sistemas de apoio à decisão e sua infra-estrutura correspondente;

- O preço de *hardware* e *software*, bem como os custos de armazenamentos relacionados ao desenvolvimento, organização, e manutenção de grandes repositórios de dados estão continuamente diminuindo;

- Os benefícios do *Data Warehouse* são estendidos à tomada de decisões estratégicas; e

- A habilidade em entender e administrar as perspectivas de uma organização, que pode economizar incontáveis horas de trabalho manual das organizações e pode ajudar a evitar cometer enganos dispendiosos.

3.1 *DATA WAREHOUSE:* CONCEITOS E FUNDAMENTAÇÃO

Conceitualmente, *Data warehouse* (*DW*) é um repositório de dados, onde ocorre a integração dos dados localizados em bancos transacionais e demais bases de dados da empresa, donde podem ser extraídas informações valiosas para geração de relatórios, análises e tomadas de decisões.

Data Warehouse é um ambiente onde os dados estão guardados de forma **uniforme, permitindo apenas a leitura e integrados**. Quando se integra dados de diferentes fontes, deve-se primeiro preocupar com a formatação desses dados. Por exemplo, o sexo pode estar guardado numa base de dados com 0 (zero) para feminino e 1 (um) para masculino, enquanto em outro banco, essa mesma informação é representada por 'M' e 'F', portanto, deve-se preocupar com a padronização dos dados, antes de realizar a integração.

Diferentemente de bancos transacionais, onde é representada a situação atual dos dados, o *Data Warehouse* armazenará dados históricos e de forma íntegra. Já que os dados estão corretamente formatados e devidamente armazenados para possibilitar a geração de uma informação valiosa, por que estes deveriam ser alterados ou excluídos?

A integração dos dados no *Data Warehouse* ocorre quando é realizada a carga do mesmo. Os dados são extraídos e transformados das bases de dados transacionais, para serem então “carregados” no *Data Warehouse*. A partir deste momento, os dados da empresa estarão integrados em apenas um local (no *Data Warehouse*), permitindo um acesso concorrente facilitado pelos diversos setores da empresa.

Os dados mantidos no *Data Warehouse* são, portanto, **não voláteis,** pois são guardados historicamente e permanentemente, não permitindo ainda, nenhuma outra tarefa a não ser a leitura; **e uniformes,** pois estão corretamente formatados, independentemente de sua origem, organizados de forma **integrada,** para garantir a facilidade de acesso.

3.2 VANTAGEM COMPETITIVA

As grandes decisões são tomadas por gerentes das empresas e baseiam-se primordialmente nas informações levantadas a partir de dados coletados e na experiência e capacitação profissional de cada qual. Mas o quão seguro é manter essa atividade concentrada apenas no tomador de decisões?

É inegável que por mais que essa decisão seja apoiada por uma ferramenta como o *Data Warehouse*, a experiência do tomador de decisões será crucial, mas ferramentas de auxílio à tomada de decisão, certamente facilitarão esse trabalho, além de desonerar o trabalho de diversos setores da empresa e permitir a continuidade nos serviços caso haja troca de funcionários.

A implantação dessas ferramentas são encaradas como vantagem competitiva pois poderá fornecer informações seguras sobre as transações da empresa e de forma rápida e eficiente, além de automatizar uma tarefa mentalmente cansativa, que seria a análise de uma montanha de dados, condensando-as até atingir o grau de viabilidade para avaliação.

Harry S. Singh, em seu livro *Data Warehouse*, Conceitos, Tecnologia, Implementação e Gerenciamento, refere-se à vantagem competitiva oferecida pelo *Data Warehouse* da seguinte forma: "Ao arquivar dados em ponto de armazenamento central, o *data warehouse* oferece uma representação integradas das múltiplas fontes de informações da empresa. Ele garante a consistência das normas de gerenciamento e das convenções aplicadas aos dados. Portanto, o *data warehouse* reflete as necessidade da corporação, não simplesmente as individuais."

A integração dos dados, possibilitando o acesso à informações importantes de forma centralizada e concorrente deve também ser destacado. A integração é tida, pois, os dados espalhados pela empresa são concentrados de forma não-volátil, uniforme e variável com o tempo num único repositório de dados, o que elimina a possibilidade de informações dúbias e facilita o acesso independentemente do qual setor da empresa tente acessá-los (SINGH, 2001).

A grande massa de dados armazenados ao longo de anos nas empresas dificultam enormemente a extração de informações valiosas para tomada de decisão e alcance de uma vantagem competitiva. Pensando nisso, *Data Warehouse* veio para integrar os dados espalhados pela empresa, de forma a mantê-los permanentemente, de forma íntegra e

consistente, realizando a tarefa cansativa de analisar e processar dados. Esse fator, certamente, permitirá entre outros aspectos algumas vantagens como:

- Sistemas e usuários apresentaram uma evolução constante;
- Melhoria na contribuição de cada funcionário na produtividade e na receita;
- Vantagem competitiva desempenhando os principais processos do negócio de forma melhor que os competidores e eliminando o maio número possível de práticas consumidoras de recursos;
- Aplicação da ciência à informação;
- Aprimoramento da seleção de marcados-alvo; e
- Determinação da eficácia de programas de *marketing*, permitindo eliminar programas fracos e aprimorar os fortes (SINGH, 2001).

3.3 ARQUITETURA *DATA WAREHOUSE*

Como já amplamente discutido *Data Warehouse* tem por função primeira permitir ao tomador de decisões, galgar seus atos administrativos sobre uma base de informações sólidas e consistentes, devidamente processadas e formatadas para melhor visualização e análise do gerente, configurando dessa forma, a informação em vantagem competitiva.

A arquitetura do *Data Warehouse*, segundo Harry S. Singh, está distribuída conforme a figura a seguir:

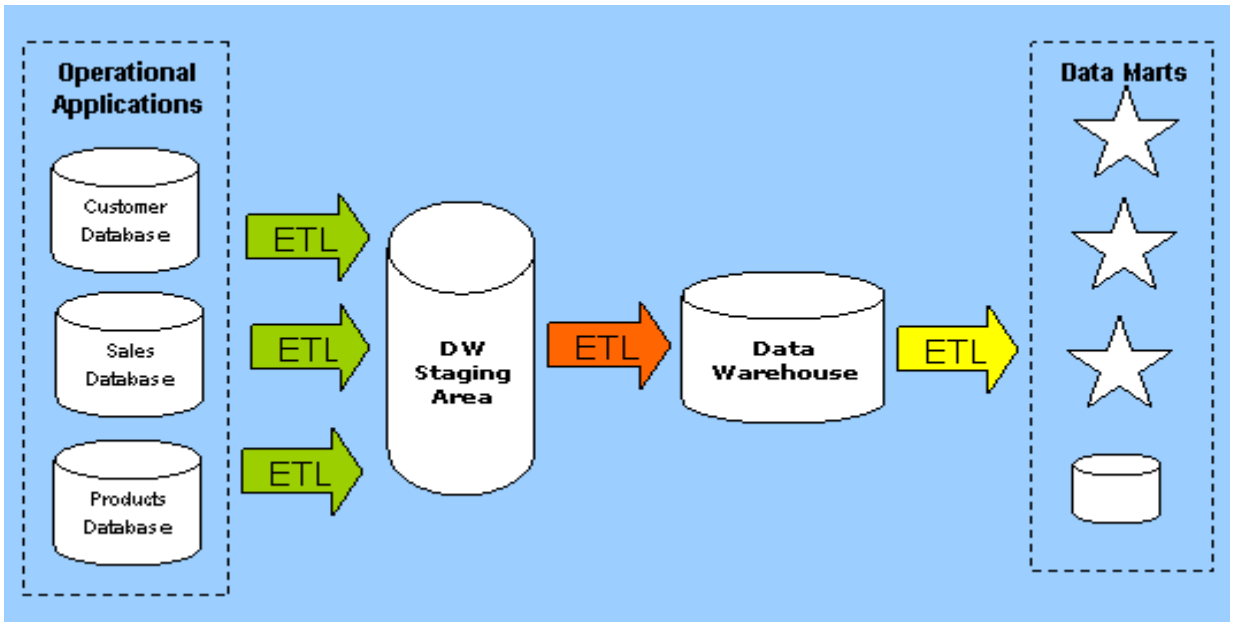


Figura 3.1 – Modelo Arquitetural do Data Warehouse

Fonte: (Guide to Data Warehousing)

Antes de se implantar um ambiente de *Data Warehouse* deve ser avaliada a natureza ocupacional da empresa e a real necessidade que esta tem em utilizar um ambiente tão robusto para processamento de dados operacionais.

Alguns estudiosos afirmam categoricamente que o investimento para se implantar um *Data Warehouse* não é justificável para um volume de dados pequeno (menos de 100.000 registros), e apenas citando Harry S. Singh "A arquitetura do *warehouse* deve ser custo-eficiente, adaptável e de fácil implementação. Uma tecnologia eficiente e confiável deve ser empregada – tecnologia que não só suporta a infra-estrutura técnica atual da organização mas que também seja flexível para um crescimento futuro."

Outro fator que vale ser cuidadosamente analisado é a forma que esta informação será distribuída pela empresa. Afinal, seria infrutífero investir na implantação de um *Data Warehouse* onde as informações será centralizadas, sendo que a empresa não possui uma boa disseminação de informações aos gerentes.

Como pode ser visto, suas principais partes são:

Camada dados operacionais/dados externos: Os dados das empresas estão espalhados em diversos sistemas de áreas distintas, representando por vezes, montanhas de dados aparentemente sem qualquer ligação.

Os dados operacionais são, portanto, os dados mantidos nos sistemas OLTP (Processamento de Transação *On-Line*), o que significa dizer que estes dados estão armazenados de forma volátil, detalhados, não redundantes e refletem apenas o estado atual das transações (HARRY, 2001).

Sob esta óptica os dados não são úteis ao *Data Warehouse*, no entanto, estes serão submetidos a um processo de extração, transformação e carga-ETL para que enfim, possam ser utilizados na geração de informações. Mais detalhes sobre o processo de ETL serão discutidos posteriormente.

Camada de acesso a informação: A camada de acesso à informação é onde o usuário final lidará com o ambiente de *Data Warehouse*. Esta camada permite a manipulação de dados para produção de relatórios, gráficos, planilhas, enfim, instrumentos para que o usuário final possa realizar a análise dos dados.

Nessa camada a informação é acessada através de ferramentas OLAP (Processamento Analítico *On-Line*). "A funcionalidade OLAP caracteriza-se pela análise multidimensional de dados consolidados da empresa suportando atividades analíticas e de navegação do usuário final que incluem cálculos e modelagens aplicadas ao longo de dimensões até hierárquicas e/ou ao longo de membros de análise de tendência em um período seqüencial de tempo" (SINGH, 2001).

Camada acesso a dados: A camada de acesso a dados será responsável pela comunicação entre a *camada de acesso à informação* e a *camada operacional*.

Atualmente, a linguagem que mais se destaca para este fim, é a SQL (Linguagem de Consulta Estrutura). Assim, a camada de acesso aos dados é responsável pela interface entre as ferramentas que acessam a informação e as bases de dados operacionais (SINGH, 2001).

Camada diretório de dados (Metadado): Metadados é o mapeamento (conteúdo e localização) dos dados contidos no *Data Warehouse*, indicando fatores como: sua

origem, seu destino e sua formação. Na arquitetura *Data Warehouse* o metadado tem ainda a função de definir os dados sob a perspectiva do tomador de decisões, estabelecendo dessa forma, um elo de ligação entre os dados e a ferramenta de suporte à decisão (SINGH, 2001).

Camada gerenciamento de processo: A tarefa de manter atualizado um banco de dados transaciona é demasiada dispendiosa, cansativa e onerosa, quando se leva em conta uma quantidade significativa de dados armazenados. Tal tarefa é por vezes distribuída pela empresa, tendo cada setor a função de manter (alterar, incluir e excluir) os próprios dados.

Na arquitetura *Data Warehouse* a camada de gerenciamento fica responsável por agendar as tarefas necessárias à construção e manutenção dos dados e metadados.

"A camada Gerenciamento de Processo funciona como um organizador dos vários processos que devem ocorrer para manter o *data warehouse* atualizado" (SINGH, 2001).

Camada *application messaging*: Esta camada é responsável pela coleta e envio de mensagens, contendo informações sobre transações (comerciais), entre as aplicações da empresa, podendo ainda, ser utilizado para isolar aplicações (deixando-as sem resposta, por exemplo). Os "mensageiros" ainda podem ser programados para levar determinada informação em determinada hora e dia, previamente definidos (HARRY S. SINGH-2001).

Camada *Data Warehouse* (física): "O *data warehouse* é o local onde estão os dados utilizados para as informações (...) Um *data warehouse* físico pode armazenar muitas cópias de dados operacionais e externos em um formato fácil de acessar e altamente flexível" (SINGH, 2001).

Camada *data staging*: Como já pode ser observado em alguns trechos deste projeto o *Data Warehouse* contém dados replicados para auxiliar em determinadas análises. Esta replicação, no entanto, não é feita ao mero a caso. Esta camada é conhecida como o gerenciador de replicação ou de cópia de dados, mas inclui também, "os processos necessários para selecionar, editar, sumarizar, combinar e carregar dados de bancos de dados operacionais e / ou externos no *data warehouse*" (SINGH, 2001).

3.4 OPÇÕES DE IMPLEMENTAÇÃO DO DW

Em desenvolvimento de sistemas distribuídos é comum o particionamento das aplicações em camadas, para facilitar o desenvolvimento e manutenção do *software*. Em desenvolvimento este conceito é conhecido como MVC (Modelo Visão Controle) onde a codificação, as regras de negócios e a camada de apresentação são distintamente separadas.

Quando se trata da implementação de *Data warehouse*, essa divisão é conhecida como arquitetura de um, dois, três ou quatro níveis, utilizando-se, similarmente, da separação entre cada bloco do *Data Warehouse*.

Com base em Harry S. Singh, os modelos de implementação podem ser descritas da seguinte forma:

Arquitetura de um nível: Consiste em uma arquitetura monolítica para o *Data Warehouse*, onde o próprio *Data Warehouse*, é o motor de inferência do DSS (*Decision Support System*) e o cliente DSS residindo sobre a mesma plataforma de hardware.

Arquitetura de dois níveis: Consiste em uma arquitetura que inclui dois componentes: um cliente e um servidor. O cliente tem como principal objetivo esconder a complexidade dos sistemas existentes. Esta abordagem é atrativa porque ela utiliza sistemas existentes legados como servidores de base de dados e requer o mínimo de investimentos adicionais em *hardware* e *software*. Todavia, esta arquitetura não é escalonável e não pode suportar um grande número de usuários finais em linha sem modificações adicionais de software.

Arquitetura de três níveis: Consiste em uma arquitetura que define três distintos níveis interface gráfica do usuário ou apresentação, lógica do negócio ou aplicação, e dados e coloca cada um destes níveis sobre seu próprio processador. Esta abordagem não sobrecarrega a estação de trabalho cliente, permitindo o uso de "clientes magros" (ou seja, clientes com pouca capacidade de memória e poder de processamento). Obtém-se claramente uma melhora em termos de escalabilidade, porém em oposição tem-se um aumento no custo e na complexidade da solução.

Arquitetura de quatro níveis: que consiste em uma arquitetura que inclui a replicação de dados, sob a condição que esta transferência possa ser executada sem poluir a

base local ou *warehouse* central. Usualmente, o próprio usuário final assume a responsabilidade pela recuperação e manutenção destes dados sobre seu *desktop*.

3.5 *DATA MART*

Muitas empresas procuram soluções oferecidas em *Data Warehouse* (*DW*). Ao verificarem que o desenvolvimento de um *Data Warehouse* tem um custo muito elevado e exige muito tempo para obter resultados, resolvem desistir ou optam por construir um *Data Mart* (*DM*), tanto pelo custo menor quanto pelo tempo menor se comparado a um *Data Warehouse*. Um outro motivo das empresas optarem por um *Data Mart* são as diversas modificações e avanços tecnológicos.

Assim como um *Data Warehouse*, o *Data Mart* é uma tecnologia que é desenvolvida com o intuito de se criar um repositório de dados que irão gerar informações para auxiliar o processo de tomada de decisão gerencial. Basicamente, um *Data Mart* se diferencia de um *Data Warehouse* pelo seu tamanho, complexidade de carga, tempo gasto para se obter a tecnologia implementada e principalmente o escopo do problema que está em questão para ser resolvido. O volume de dados de um *Data Mart* é menor, sua utilidade é mais restrita, pois apenas um setor ou alguns setores, grupo ou departamento irão usufruir de seus benefícios, não servindo para a empresa como um todo, como acontece em um *Data Warehouse* onde todos os setores podem ser beneficiados. Por exemplo, o setor de vendas de uma empresa pode construir um *Data Mart* com dados referentes às operações realizadas, tais como, o volume de vendas por período, sendo que o setor de Recursos Humanos não necessita saber qual foi o volume de vendas do ano 2005 para auxiliar em suas atividades, ou seja, dificilmente irá usufruir de informações contidas neste *Data Mart de Vendas*.

Os dados contidos em um *Data Mart* passam por uma seleção onde são corrigidos e adequados de acordo com uma parametrização anteriormente definida na modelagem.

É bastante comum verificar a denominação de um *Data Mart* como um *Data Warehouse* departamental, justamente pela sua destinação a setores específicos dentro de uma empresa (INMON-2001).

Uma empresa pode construir um *Data Mart* para cada setor especifico da empresa, consolidá-los e depois pode criar vínculos entre eles por meio de um *Data Warehouse* lógico que envolva toda a empresa (SINGH, 2001).

Existem duas grandes opiniões divergentes entre dois precursores de *Data Warehouse*s, *Ralph Kimball* que defende que os *Data Mart*s não devem ser departamentais, mas sim abrangendo "assuntos" que podem ser utilizados por diversos setores da empresa que façam uso das informações nele contido, e a partir destes *Data Mart*s devidamente inter-relacionados se chegaria a um *Data Warehouse*; já *Bill Inmon* enfatiza que os sistemas da empresa, como um todo, serve para alimentar o *Data Warehouse* e este deverá servir como fonte de dados para os *Data Mart*s da empresa, ele defende essa tese dizendo que ao criar primeiramente sistemas departamentais, *Data Mart*s, poderia fatalmente fugir do objetivo geral da empresa atendendo apenas as necessidades de departamentos.

Um *Data Mart* pode ser classificado como independente ou dependente. Ele é considerado independente, conforme figura 3.1, quando um determinado departamento da empresa implementa o *Data Mart* sem necessariamente contar com recursos da área de tecnologia de informação da empresa, isso se dá porque o próprio departamento é responsável pelo *Data Mart* e este departamento pode alimentar os dados do *Data Mart* diretamente, ou seja, a responsabilidade é exclusiva do departamento.

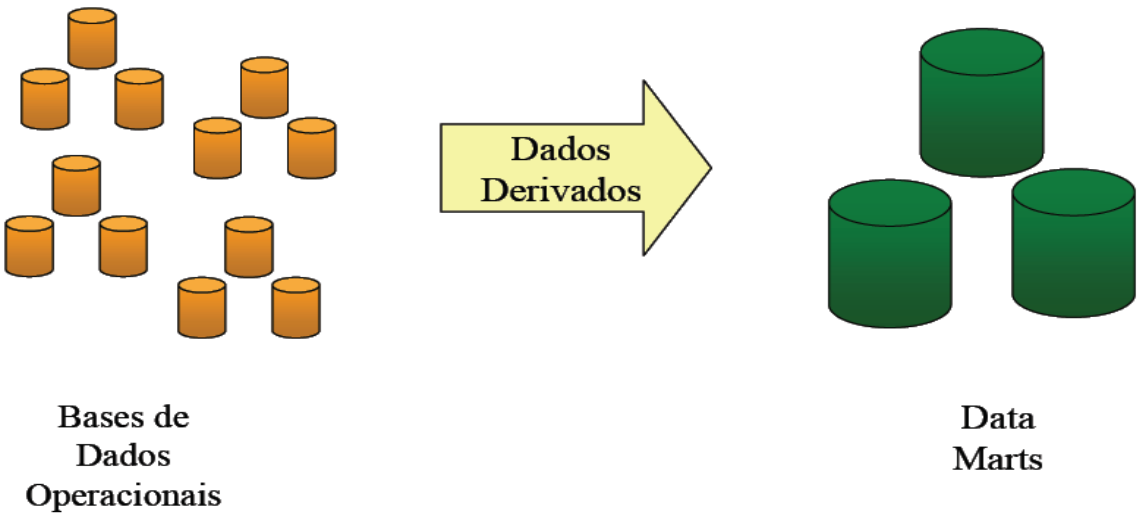


Figura 3.1 – Representação do Data Mart independente

Fonte: PUC – Rio de Janeiro, Laboratório de Tecnologia em Banco de Dados

Em um *Data Mart* dependente os dados são coletados das bases de dados administradas pelo setor de tecnologia de informação. Os dados de um *Data Mart* também podem ser adquiridos do banco de dados de um *Data Warehouse*, esses dados são ajustados às intenções do departamento ao qual será direcionado, conforme figura 3.2 (SING, 2001).

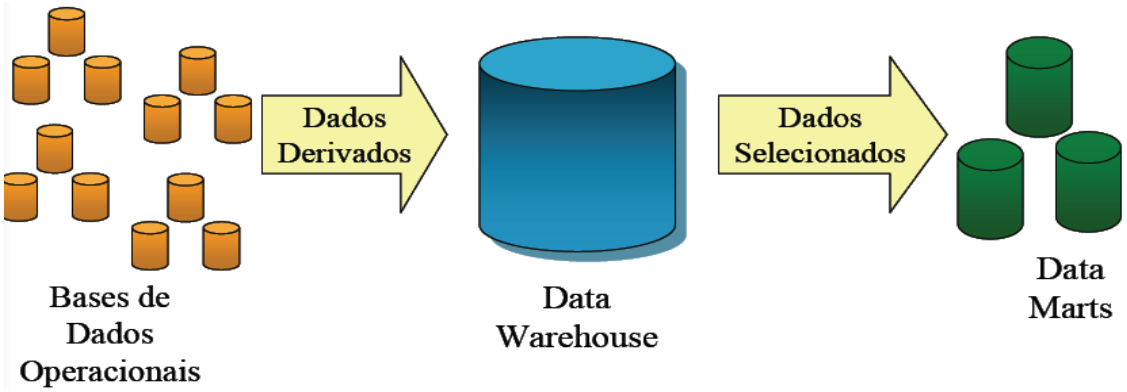


Figura 3.2 – Representação do Data Mart dependente

Fonte: PUC – Rio de Janeiro, Laboratório de Tecnologia em Banco de Dados

Com as idéias citadas anteriormente temos duas maneiras de se obter um *Data Mart*:

Top-down – ocorre quando uma organização obtém primeiramente um *Data Warehouse* e posteriormente fragmenta o *Data Warehouse* em assuntos de interesses de determinadas áreas e dando origem a bancos de dados menores que são os *Data Mart*.

Figura 3.3 – Modelo de Data Mart - Top-down

Fonte: PUC – Rio de Janeiro, Laboratório de Tecnologia em Banco de Dados

Botton-up – ocorre quando se opta em obter um *Data Mart* ou vários *Data Mart*s para cada área da empresa e posteriormente interligá-los e obter um *Data Warehouse*.

Em resumo, um *Data Mart* é feito para atender um setor, departamento ou grupo específico de uma determinada empresa, enquanto o *Data Warehouse* é feito para atender a toda empresa, sendo ambos feitos para auxiliar na decisão gerencial. A empresa pode ter um *Data Mart* também para verificar seus efeitos antes de decidir-se por um *Data Warehouse*.

Optar por uma dessas soluções é analisar o problema da corporação, e não simplesmente decidir-se por um *Data Mart* por ele ser mais barato ou mais rápido de ser implementado, ele pode não servir para o que a organização deseja.

3.6 PROCESSO DE EXTRAÇÃO, TRANSFORMAÇÃO E CARGA-ETL

O processo de estruturação dos dados no *Data Warehouse* é garantido através de um processo de ETL (Extração, Transformação e Carga) dos dados obtidos em bancos transacionais, arquivos de texto, arquivos DBF, arquivos do Access. O processo de definição das fontes de origem dos dados é o primeiro passo do ETL.

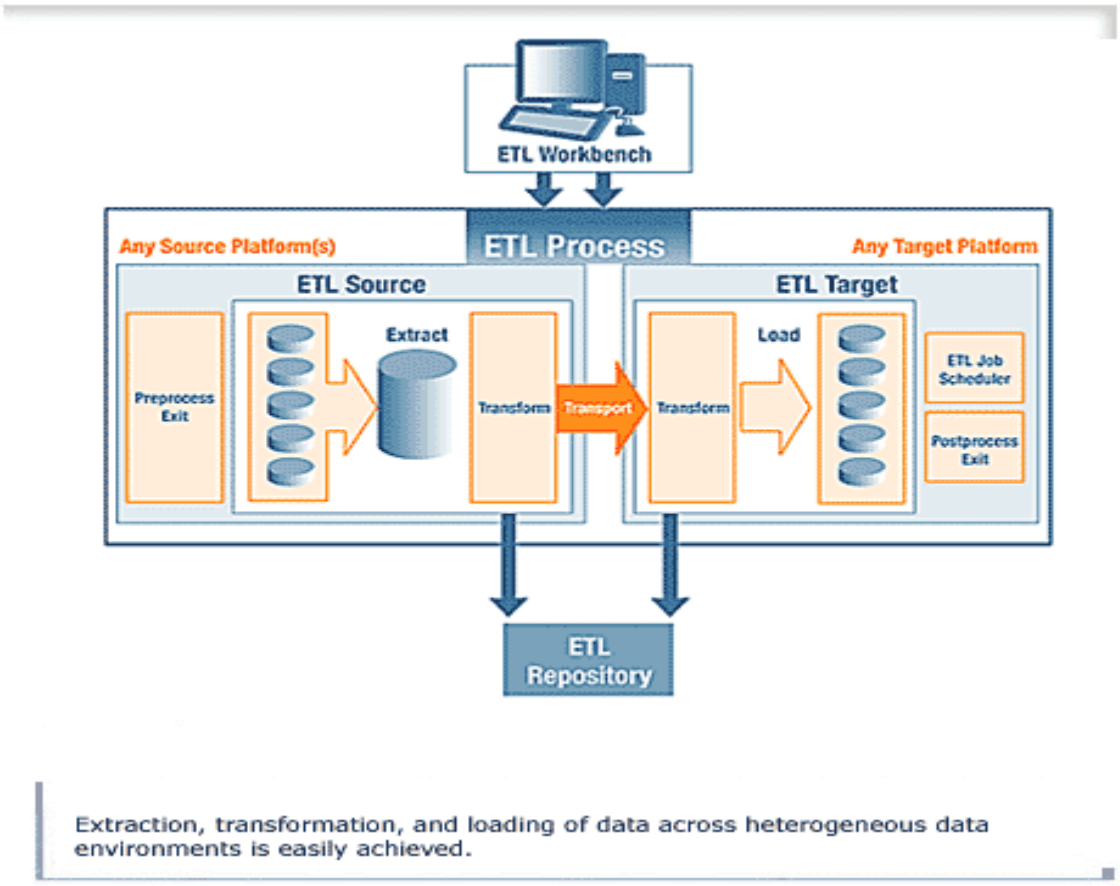


Figura 3.4 – Modelo do Processo de ETL

Fonte: Information Builders

Mas na extração desses dados não é incomum encontrar divergência de informações até mesmo sobre o mesmo atributo. Quando se negocia com diversas fontes de dados de origens distintas é comum encontrar-se, por exemplo, o atributo “sexo” sendo armazenado sobre a forma de “M” ou “F” para homem e mulher, respectivamente, ou “1” e “0”. Essas divergências impedem que sejam produzidas informações precisas, pois, ao se buscar o número de clientes separados por sexo certamente o gerente obterá um relatório irreal.

Para sanar essa questão o segundo passo do ETL é a transformação dos dados, onde as diferenças existentes para os mesmos atributos são tratadas até atingirem um formato consistente e íntegro.

Mas ocorre ainda a existência de campos não preenchidos, fazendo com o que os bancos de dados fiquem incompletos. A primeira solução para essa questão é a definição de valores padrões para esses campos, mas, essa metodologia não deve ser tomada como única “saída” para este problema. Um problema clássico citado na literatura sobre o assunto é

o vendedor de cartões de crédito, que pouco preocupado com a integridade e fidedignidade das informações cadastradas no sistema informam o CPF nº 999.999.999-99, o que, quando solicitado pelo gerente um relatório sobre os maiores clientes, produziria uma informação no mínimo curiosa.

Esses dados, onde o ELT não tem autonomia para utilizar um valor padrão, é mantido separado para preenchimento antes de prosseguir com o próximo passo: carga dos dados.

A carga é finalmente a formação do *Data Warehouse*, onde os dados estão devidamente formatados e de acordo com um padrão pré-estabelecido junto ao usuário. Como já demonstrado anteriormente, dentro da arquitetura do *Data Warehouse* existe uma camada responsável unicamente pela sumarização, formatação e padronização destes dados, além da definição da temporização destas cargas e outras atividades, já citadas quando da explanação da arquitetura *Data Warehouse*.

3.7 ODS – *OPERATION DATA STORE*

Operation Data Store-ODS é um conjunto dos dados integrados destinados ao *Data Warehouse*, funcionando como um repositório intermediário dos dados, podendo, também, ser utilizados nas consultas. A vantagem dessa abordagem é a otimização do processo de extração e carga dos dados no *Data Warehouse*.

A construção de um ODS é facultativa, entretanto, colabora para diminuir os esforços de construção de um *Data Warehouse*. Todo o esforço de integração entre os sistemas transacionais da empresa seriam depositados no ODS e a carga do *Warehouse* seria simplificada de maneira incomensurável (DWBRASIL, 2003).

Enquanto *Data Warehouse* e *Data Mart* se referem a estruturas dimensionais de dados, remodeladas com o objetivo de prover análises diferenciais, o conceito de ODS, por sua vez, está relacionado com o armazenamento e tratamento de dados operacionais, de forma também consolidada, porém sem características dimensionais. O conceito de ODS nasceu como uma solução intermediária entre os muitos arquivos e dados espalhados pela empresa, que careciam de uniformização e a proposta final de *Data Warehouse* e *Data Mart*. O ODS, além de ser a metade do caminho entre o legado e o *Data Warehouse*, também oferece

informações importantes do ponto de vista decisório, devido a sua característica de consolidação e integração de várias fontes de dados (BARBIERI, 2001).

Pelo fato do ODS conter os dados do ambiente operacional, já integrados, pode ser utilizado para agilizar o processo de consolidação, proporcionando um melhor desempenho na fase de atualização dos dados no *Data Warehouse*.

3.7.1 ODS x *DATA WAREHOUSE*

Se considerarmos o ODS proposto por Inmon como sendo uma necessidade de integração dos dados do ambiente operativo, podemos compará-lo com um *Data Warehouse* da seguinte forma (COSTA, ANCIAES, 2001):

- Tanto o *Data Warehouse* quanto o ODS são orientados por assuntos;
- O ODS e o *Data Warehouse* são integrados da mesma forma, porém,
- ODS é volátil, enquanto o *Data Warehouse* não;
- O ODS contém os dados mais atualizados, enquanto o *Data Warehouse* contém dados históricos e atualizados;
- Ao contrário do *Data Warehouse*, o ODS só contém dados detalhados, e não possui dados agregados e dados sumarizados.

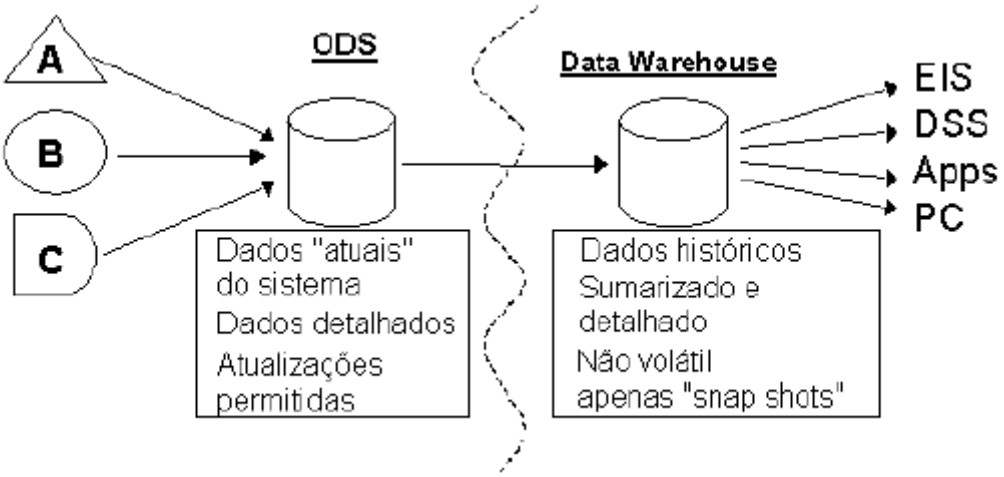


Figura 3.5 – Diferença entre ODS e Data Warehouse

Fonte: (COSTA, ANCIAES, 2001)

Ralph Kimball redefiniu o ODS como sendo uma estrutura de armazenamento de dados detalhados, não-volátil, orientada a assuntos e que suporta os sistemas do ambiente operacional com dados integrados. De qualquer forma, o uso de dados detalhados para tomada de decisões deveria se basear no nível de granularidade mais baixo suportado pelo Data Warehouse (KIMBALL *et al*, 1998).

A tabela 3.1 detalha as diferenças entre *ODS, Data Warehouse e Data Mart.*

Critério	ODS	Data Warehouse	Data Mart
Função	Suporte operacional a nível de departamento / área	Suporte à tomada de decisão corporativa, departamental ou de determinada área do negócio	Suporte à tomada de decisão departamental ou área de negócio
Fonte dos Dados	Sistemas operativos e fontes externas	Sistemas operativos e fontes externas	Sistemas operativos e fontes externas ou dados provenientes de um data warehouse
Esquemas	Normalizados	Dimensional ou relacional dependendo do uso	Dimensional para aumentar a performance
Escopo	Limitado	Amplo	Limitado
Tempo	Dados atuais e sincronizados com os sistemas operativos	Dados históricos, para permitir a realização de análises e planejamento estratégico	Dados históricos, para permitir a realização de análises e planejamento estratégico
Volatilidade/ Frequência de Atualização	Alta; os dados freqüentemente estão sendo atualizados para refletir as mudanças no ambiente operativo	Baixa/Média; os dados são estáticos e as mudanças dependem da freqüência de atualização (em geral, inserções)	Baixa; os dados são estáticos e as mudanças só ocorrem quando atualizações nos dados históricos são necessárias
Agregação dos Dados	Nível baixo; em geral, dados detalhados	Médio; alguns dados sumarizados são requeridos para aumentar performance e padronizar as análises. Níveis detalhados de dados transacionais históricos	Médio; alguns dados sumarizados são requeridos para aumentar performance e padronizar as análises. Somente contêm dados detalhados necessários para análises específicas

Tabela 3.1 – Comparação entre ODS, Data Warehouse, Data Mart

Fonte: (COSTA, ANCIAES, 2001)

3.8 METADADOS

O volume de dados utilizado para construção de um *Data Warehouse* facilmente atinge a casa de Giba ou Terabytes. A compreensão do funcionamento desses dados é de fundamental importância para implantação de um projeto de *Data Mart/Data Warehouse.*

Na estrutura desses ambientes repositórios de dados existe o conceito de metadados, utilizados para oferecem a semântica dos dados existentes na estrutura do *Data Mart/Data Warehouse* construído.

3.8.1 CONCEITOS

Metadados são dados que descrevem dados. É uma descrição concisa a respeito dos dados, estes podem ser documentos, coleção de documentos, gráficos, tabelas, imagens, vídeos, entre tantos outros. De uma forma mais completa, pode-se dizer que o metadado é a "descrição do dado, do ambiente onde ele reside, como ele é manipulado e para onde é distribuído" (COME, 1999).

Metadado é uma abstração do dado, ou seja, é o dado de alto nível que descreve o dado de baixo nível sendo o instrumento que transforma o dado sem valor em conhecimento. Por exemplo, é o metadado, na forma de definição de campo, que informa que uma dada cadeia de *bits* é um endereço de cliente, parte de uma imagem fotográfica ou parte do código de um programa de computador (COME, 1999).

Os metadados têm um papel importante na gestão de dados, pois a partir deles as informações são processadas, atualizadas e consultadas. As informações de como os dados foram criados/derivados, ambiente em que reside e/ou residiu, alterações feitas, entre outras são obtidas de metadados. Os metadados fornecem os recursos necessários para entender os dados através do tempo (SALETE, 2004).

É fato que quando não se documenta os dados, com o decorrer do tempo, fica sujeito à superposição de esforços de coleta e manutenção dos mesmos e vulneráveis aos problemas de inconsistências, assim sendo, são necessárias estratégias para gestão de metadados visto que não é trivial decidir quais metadados devem ser coletados e mantidos. Uma arquitetura de metadados deve ser flexível para permitir acréscimo ou decréscimo de metadados à medida que novas necessidades apareçam (SALETE, 2004).

3.8.2 METADADOS E DATA WAREHOUSE/DATA MART

Desde que surgiram os bancos de dados se fala sobre a importância de documentar os sistemas e os próprios bancos. Com o surgimento do conceito de *Data Warehouse*, essa importância aumentou muito. Segundo **(**IMNON, 1996), é por meio dos metadados que a utilização mais produtiva do *Data Warehouse* é alcançada. Os metadados mantêm informações sobre “o que está”, e “e onde" no ambiente de *Data Warehouse*.

Os metadados consistem em informações sobre dados que compõem o *Data Warehouse*. Eles são extremamente importantes dentro do ambiente de *Data Warehouse*, pois representam uma visão integrada das bases de dados que fazem parte deste ambiente. Eles são utilizados para construir, manter, gerenciar e utilizar o Data Warehouse (COSTA, ANCIAES, 2001).

No *Data Warehouse* documentar tudo é vital para a sobrevivência do projeto, pois o *Data Warehouse* pode ser um projeto bastante grande e se não houver uma documentação eficiente pode torna-se um caos e significar o fracasso do projeto.

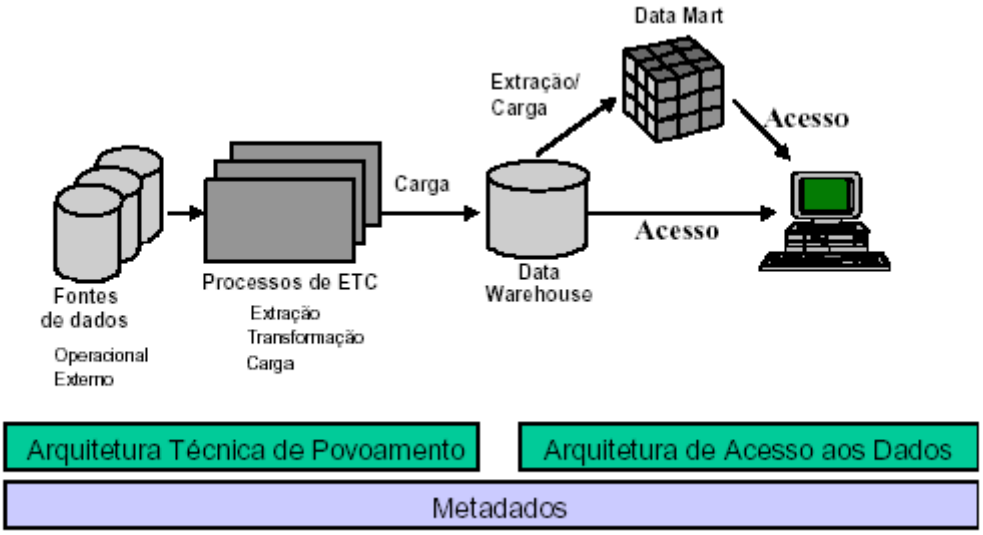


Figura 3.6 – Visão simplificada de uma arquitetura de Data Warehouse

Fonte: (COSTA, ANCIAES, 2001)

De acordo com a abordagem de Imnom, a figura 3.6 é uma visão simplificada de uma arquitetura de data warehouse, a qual abrange a arquitetura de dados, arquitetura técnica, administração e gerenciamento do *Data Warehouse* e, também, os metadados.

Não é difícil chegar a conclusão que os metadados são importantes em um *Data Warehouse*. Mas existem algumas perguntas básicas que também devem ser respondidas pelos metadados (COME, 1999) como:

- Que tabelas, atributos e chaves o *Data Warehouse* contém?
- Qual é a origem de cada conjunto de dados?
- Que transformação lógica foi usada na carga do dado?
- Como o metadado tem mudado ao longo do tempo?
- Quais *aliases* existem e como eles se relacionam?
- Quais são as referências-cruzadas entre termos técnicos e de negócios?
- Com qual freqüência os dados são carregados?
- Qual o volume de dados existente?

A complexidade dos metadados no *Data Warehouse* aumenta muito. Num sistema OLTP gera-se documentos somente sobre o levantamento dos dados, banco de dados e o sistema que alimenta o mesmo. No *Data Warehouse* além do banco gera-se uma documentação muito maior, que além de falar sobre o levantamento de dados e o banco, temos ainda o levantamento dos relatórios a serem gerados, de onde vêm os dados para alimentar o *Data Warehouse*, processos de extração, tratamento e rotinas de carga dos dados. Ainda podem gerar metadados as regras de negócio da empresa e todas as mudanças que elas podem ter sofrido, e também a freqüência de acesso aos dados (DW!, 2002).

3.8.3 O QUE GERA METADADOS?

Todas as fases de um projeto de *Data Warehouse/Data Mart*, desde o levantamento de requisitos até a visualização da informação geram metadados. Os metadados, segundo (IMNON, 1996) mantêm informações sobre:

- A estrutura de Dados segundo a visão do programador (visão tecnicista);
- A estrutura de Dados segundo a visão do analista de SAD (visão de negócios);
- A fonte de dados que alimenta o *Data Warehouse*;
- A transformação sofrida pelos dados no momento de sua migração para o *Data Warehouse*;
- O modelo de dados;
- Relacionamento entre o modelo de dados e o *Data Warehouse*;
- Histórico das extrações de dados;
- Informações referentes a relatórios gerenciais;
- Informações referentes aos processos de Carga.

3.8.4 REPOSITÓRIO DE METADADOS

Existem dois tipos de metadados que podem estar no repositório de metadados: técnicos e de negócios, conforme figura 3.7 (COME, 1999).

Metadados técnicos fornecem aos desenvolvedores e usuários técnicos de sistemas de suporte à decisão a confiança de que os dados estão corretos. Além disso, metadados técnicos são críticos para a manutenção e o crescimento contínuos de um *Data Warehouse*.

Eles abrangem todos os aspectos técnicos e funcionais do *Data Warehouse*, constituindo uma visão abstrata do mesmo e definindo todos os elementos e a forma como interagem entre si (COSTA, ANCIAES, 2001).

Metadados de negócios são a ligação entre o *Data Warehouse* e os usuários de negócios. Esses dados fornecem uma espécie de mapa aos usuários para que eles possam acessar os dados, tanto no *Data Warehouse* como nos *Data Marts*. Os metadados de negócios mostram em termos de negócios, que relatórios, *queries* e dados estão no *Data Warehouse*,

localização dos dados, confiabilidade dos dados, contexto dos dados, regras de transformação que foram aplicadas e quais as origens desses dados (COME, 1999).

O repositório de metadados ajuda tornar as informações em *DataWarehouse* e *Data Mart* mais visíveis e acessíveis aos usuários. Um repositório pode viabilizar ou não um *Data Warehouse*.

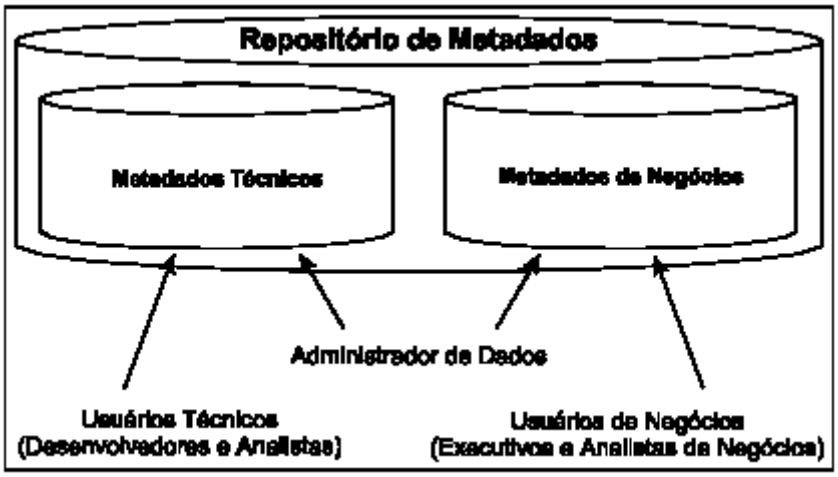


Figura 3.7 – Repositório de Metadado
Fonte: (COME, 1999)

Problemas que grandes corporações sofrem com a ausência de um ambiente de metadados (DWBRASIL, 2003):

- Os metadados são inacessíveis ao usuário de negócio;
- Os metadados estão armazenados em ambientes distintos;
- Usuários de Negócio são dependentes da área de Tecnologia;
- Dificuldade de mapeamento de dados entre o Data Warehouse e os Sistemas Operacionais;
- Sub-utilização do Ambiente de Informações Gerenciais;
- Alterações das Estruturas e Definições de Dados não são acompanhadas;
- Impossibilidade de se mensurar a utilização do ambiente;

- Visão não integrada da empresa;

3.9 CONSIDERAÇÕES SOBRE O CAPÍTULO

Independentemente da arquitetura para implementação adotada, *Data Warehouse* é uma tecnologia capaz de condensar informações e oferecer a estrutura necessária para uma modelagem multidimensional que aliada a uma ferramenta OLAP, gerará informações sobre um fato importante sob óticas diferentes.

Neste capítulo foram explorados os aspectos que fazem dessa tecnologia um diferencial na hora de agregar valor aos dados disseminados pela organização, organizando-os de forma devidamente padronizada para garantir a integridade da informação apresentada.

As características de integração, não volatilidade, capacidade histórica e uniformidade de dados são os grandes diferenciais relativamente a bancos de dados transacionais, o que dentre outras características, justifica o investimento nesse ambiente, uma vez que as vantagens frente aos competidores poderão ser mais facilmente alcançada o que permite ao tomador de decisões galgar suas ações em dados mais concretos que apenas sua intuição e experiência.

As estruturas integrantes do ambiente de *Data Warehouse/Data Mart* podem ser citadas como, importantes e essenciais, os metadados, que são de suma importância para o sucesso de um *Data Warehouse*. Ao se iniciar qualquer projeto deve sempre se preocupar com os mesmos, pois são eles que servirão como orientação para se guiar pelo emaranhado de tabelas, relatórios e dados quando estiverem perdidos.

Ainda, conforme tratado no presente capítulo, ETL, responsável pela extração, transformação e carga dos dados, garantindo dessa maneira uma formatação ideal dos dados absorvidos dos bancos de dados legados da empresa.

E por fim o ODS que permite a visualização de informações antes da carga no *Data Warehouse.* É importante salientar que a carga do repositório de dados dificilmente ocorrerá diariamente, no entanto, o gestor da informação necessita de informações que favoreça sua tomada de decisão. ODS permite então a construção dessa estrutura previamente a carga do *Data Warehouse.*

CAPÍTULO 4 – MODELAGEM MULTIDIMENSIONAL

As abordagens tradicionais de dados desenvolvidos ao longo dos últimos (vinte ou trinta) anos genericamente conhecida como modelagem de dados, foi um instrumento de extrema utilidade na formatação de estruturas capazes de serem implantadas e entendidas pelos gerenciadores de bancos de dados. Essa abordagem sempre primou pela representação de estruturas que melhor se ajustassem às características transacionais dos processamentos de dados (BARBIERI, 2004).

Sistemas que empregam essa abordagem constituem a camada básica e essencial de processos operacionais das empresas, mas com o aumento do volume de dados e busca de informações decisórias de forma mais ágil, os modelos de dados seguem uma outra arquitetura, o modelo dimensional, para atender processos analíticos.

4.1 MODELO DE DADOS

O primeiro passo para elaboração de um modelo de dados, concentra-se na observação dos fatos relevantes que ocorrem na realidade de um processo ou de uma empresa, com a finalidade de construir um sistema que possa automatizar as necessidades de informação dos mesmos.

Ao coletar e selecionar os fatos relevantes, o analista deve identificar os elementos geradores de informação, as leis que regem esta realidade, bem como, as operações que incidem sobre os elementos básicos (dados). Para registrar as necessidades é necessário fazer uso de um modelo, ou seja, algo que nos mostre como as informações estão relacionadas (fatos) (HOKAMA, 2004).

Com o desenvolvimento de outras necessidades, alavancadas por aspectos de competitividade e busca de diferenciais de negócios e conseqüente tomada de decisão, esses modelos se mostraram inadequados, pois, eram incapazes de atender tais demandas. (BARBIERI, 2001).

Como será visto a seguir, os modelos de dados podem ser classificados segundo a arquitetura que utilizam. O modelo relacional surgiu para atender os sistemas

transacionais (OLTP), já o modelo dimensional surgiu para atender os sistemas analíticos (OLAP).

Modelo Dimensional	Modelo Relacional E/R
Padrão de estrutura mais fácil e intuitiva	Modelo mais complexo
Anterior ao MER, anos 60.	Ênfase nos Bancos de Dados Relacionais, anos 70.
Tabelas Fato e tabelas Dimensão	Tabelas que representam Dados e Relacionamentos
Tabelas Fato são o núcleo – normalizadas.	Todas as tabelas são comumente normalizadas
Tabelas Dimensão são os pontos de entrada	As tabelas são indistintamente acessadas e de filtro inicial
Tabelas Dimensão opcionalmente normalizada	Todas as tabelas são comumente normalizadas
Modelo mais facilmente juntado.	Maior dificuldade de agregação pelo número maior de tabelas
Leitura mais fácil do modelo por usuários não especializados	Maior dificuldade de leitura pelo usuário não especializada

Tabela 4.1 - Comparação entre modelo Relacional – E/R e modelo Dimensional
Fonte: (HOKAMA, 2004)

4.1.1 MODELO RELACIONAL

A abordagem relacional está baseada no princípio de que as informações em uma base de dados podem ser consideradas como relações matemáticas e que estão representadas de maneira uniforme, através do uso de tabelas bidimensionais. Este princípio coloca os dados dirigidos para estruturas mais simples de armazenar dados, que são as tabelas, onde a visão do usuário é privilegiada (HOKAMA, 2004).

Esse modelo surgiu para atender sistemas transacionais que possuem operações atômicas e pré-definidas, geralmente, com um grande número de usuários simultâneos realizando operações repetidamente.

O modelo relacional deve estar preparado para ter um tempo de resposta pequeno, uma vez que, as transações são curtas e operam sobre poucos dados, visando a eliminação de redundância deixando os dados em um único lugar.

A base de dados de um modelo relacional reflete imediatamente as operações realizadas, com atualizações que são visualizadas em tempo real, não existindo um processo de carga de dados. A inserção de um registro ou a atualização de um registro já existente é imediatamente vista pelos usuários do sistema e geralmente não existe a necessidade de consultas a dados históricos (HOKAMA, 2004).

4.1.2 MODELAGEM DIMENSIONAL

A modelagem dimensional 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 de medição. Os modelos dimensionais são compreensíveis, previsíveis, ampliáveis e resistentes ao ataque específico de grupos de usuários de negócio, por se manter fiel à simplicidade, ter uma perspectiva voltada para as necessidades analíticas da empresa, e especialmente ao seu formato simétrico, em que todas as dimensões normalmente são iguais pontos de entrada na tabela de fatos (KIMBALL, 2002).

A modelagem multidimensional é a técnica de projeto mais freqüentemente utilizada para a construção de *Data Mart.* O objetivo é buscar um padrão de apresentação de dados que seja facilmente visualizado pelo usuário final e que possua um bom desempenho de consulta.

O modelo dimensional surgiu para atender sistemas de processamento analítico, com consultas para planejamento tático e estratégico da empresa. Atualmente a utilização desses sistemas pelo nível operacional das empresas vem crescendo, auxiliando o processo de tomada de decisões diárias. Normalmente, ele atende um pequeno número de usuários que realizam consultas planejadas (relatórios pré-definidos) e *ad-hoc* **(**HOKAMA, 2004).

O tempo de resposta é maior, devido a consultas longas que operam sobre grande volume de dados. Para melhor desempenho nas consultas, há redundância planejada dos dados, compensando os gastos com armazenamento e atualização das informações. O resultado é uma estrutura simples, com modelos que refletem o processo de análise de negócios. Os usuários somente realizam consultas na base de dados, podendo extrair e formatar seus próprios relatórios, não dependendo da equipe de tecnologia para isso.

Processamento Transacional:	**Processamento Analítico:**
Dados normalizados	Dados consistentes
Atualização em tempo real	Desempenho compatível com o volume de dados (ou fórmulas, etc)
Controle de concorrência	Clara representação do modelo de negócio (camada semântica)
Dados correntes	Ferramentas especiais para os usuários finais
Respostas imediatas	

Tabela 4.2 - Requisitos de processamento transacional e analítico.
Fonte: (HOKAMA, 2004)

O produto final da modelagem Dimensional á a produção de um modelo conceitual dimensional, formada por tabelas Fato e tabelas Dimensão.

4.2 PASSOS PARA CONSTRUÇÃO DE UM MODELO MULTIDIMENSIONAL

Há alguns passos para o bom desenvolvimento de um modelo multidimensional.

4.2.1 DEFINIÇÃO DA ÁREA DE NEGÓCIO

Uma área de negócio é devidamente escolhida, baseada nas prioridades emergentes da empresa. Parte-se, então, para definir aqueles que serão os processos-alvo do projeto de *Data Warehouse*/*Data Mart*. Os processos que compõem o modelo deverão ser analisados por uma abordagem que permita seu pleno entendimento (BARBIERI, 2001).

4.2.2 DEFINIÇÃO DA GRANULARIDADE

A questão mais importante de projeto que o desenvolvedor do *data mart/data warehouse* precisa enfrentar, refere-se à definição da granularidade, ou seja, o nível de detalhe ou de resumo dos dados existentes no *data warehouse/data mart*..

Quando a granularidade de um *data warehouse* é apropriadamente estabelecida, os demais aspectos de projeto e implementação fluem tranqüilamente; quando ela não é estabelecida, todos os outros aspectos se complicam (INMON, 1997).

Essa análise é de fundamental importância, pois define de forma combinatória, os níveis dimensionais que serão usados para o armazenamento dos dados. Os fatores que definem a escolha da granularidade estão relacionados com o volume de dados a ser mantido e com o processamento necessário para produzi-lo (BARBIERI, 2001).

Com um nível de granularidade mais alto o volume de dados é bem menor, como a quantidade de registro é geralmente bem grande, a força de processamento necessária para acessar os dados é um fator importante. O problema é que à medida que o nível de granularidade aumenta, o número de consultas que podem ser atendidas diminui, sendo que numa granularidade mínima as consultas mais detalhadas podem ser respondidas.

Pode-se ter, por exemplo, uma análise dos fatos em relação somente a ano e mês, como pode ter em relação ano, mês, dia e hora, o que diminuiria a granularidade, e aumentaria o volume de dados.

4.3 DEFINIÇÃO DAS TABELAS

Segue a definição das tabelas Fato, que representam o processo do negócio, e das tabelas Dimensões, que representam as entidades do negócio.

4.3.1 TABELA FATO

A tabela de fatos é a principal tabela de um modelo dimensional, onde as medições numéricas de interesse da empresa estão armazenadas (KIMBALL, 2002). A palavra "fato" representa uma medida dos processos que estamos modelando, como quantidades, valores e indicadores. A tabela de fatos registra os fatos que serão analisados. É composta por uma chave primária (formada por uma combinação única de valores de chaves de dimensão) e pelas métricas de interesse para o negócio.

As dimensões indicam a forma como as medidas serão vistas, os seja, são os aspectos pelos quais se pretende observar as métricas. A intersecção das chaves de dimensão define a granularidade da tabela de fatos (HOKAMA, 2004).

A tabela de fatos deve ser sempre preenchida com as medidas referentes ao fato. Não se deve preencher uma linha da tabela fato com zeros para representar que nada aconteceu (por exemplo, que não houve vendas de um produto em determinada data), pois isso faria com que a tabela de fatos crescesse demais.

Outro ponto importante é que a tabela de fatos deve representar uma unidade do processo do negócio, não devendo-se misturar assuntos diferentes numa mesma tabela de fatos.

Os atributos mais comuns em uma tabela de fatos são valores numéricos, métricas.

4.3.1.1 MODELAGEM DAS TABELAS FATOS

Uma técnica para modelar a tabela de fatos é responder as seguintes perguntas (HOKAMA, 2004):

- Que processo estamos modelando?

- O que usamos para medir este processo?
- Quais os indicadores críticos de desempenho desse processo?

Um processo é uma atividade de negócio natural executada na empresa. Ouvir os usuários é o meio mais eficiente de selecionar o processo de negócio. É importante lembrar que não estamos nos referindo a um departamento ou função de negócio de uma empresa quando falamos de processos de negócio. Se forem estabelecidos modelos dimensionais para cada departamento, inevitavelmente os dados serão duplicados com diferentes rótulos e terminologias (KIMBALL, 2002).

Os valores necessários para se medir o processo podem ser encontrados prontos no sistema origem, ou então podem ser calculados baseados em medidas do sistema origem. Os fatos típicos são valores numéricos aditivos (HOKAMA, 2004).

Os indicadores críticos de desempenho são os indicadores em que a empresa se baseia para fazer suas análises (HOKAMA, 2004).

4.3.2 DIMENSÃO

As tabelas dimensão representam entidades de negócios e constituem as estruturas de entrada que servem para armazenar informações como tempo, geografia, produto, cliente, etc. As tabelas dimensão tem relação 1:N com a tabela fato, e possuem um número significativo menor de linhas do que as tabelas fato. Devem ser entendidas como as tabelas que realizam os filtros de valores aplicados na manipulação dos fatos e por onde as consultas entram no ambiente *Data Warehouse*/*Data Mart* (BARBIERI, 2001).

A qualidade do banco de dados é proporcional à qualidade dos atributos de dimensões, portanto devem ser dedicados tempo e atenção a sua descrição, ao seu preenchimento e a garantia da qualidade dos valores em uma coluna de atributos (KIMBALL, 2002).

Cada dimensão é definida com uma única chave primária. Essa chave é a base da integridade referencial no relacionamento com a tabela de fatos. Uma tabela de dimensão é composta também de atributos, os melhores atributos são textuais e distintos, e devem

consistir de palavras reais, evitando-se o uso de códigos e abreviações. Esses atributos descrevem as linhas na tabela de dimensão.

Por exemplo, em uma tabela de dimensão Alimentos, o campo “perecível” deve ser preenchido com valores como “É perecível” ou “Não é perecível” ao invés de usar simplesmente “S” e “N” (HOKAMA, 2004). Em um relatório com a listagem de milhares de produtos, um valor descritivo tem muito mais utilidade do que códigos.

4.3.2.1 HIERARQUIA DE DIMENSÕES

As hierarquias são a base para a navegação entre os diferentes níveis de detalhe em uma estrutura multidimensional, descrevem a lógica dos relacionamentos entre os dados (HOKAMA, 2004).

Algumas estruturas hierárquicas são facilmente identificadas, como por exemplo, uma estrutura de tempo representada por horas, dias, semanas, meses, trimestres e anos ou uma estrutura geográfica representada por cidades, municípios, estados, regiões e países (KIMBALL, 1996).

A figura 4.1 representa a hierarquia para uma dimensão **Produto**. Essa hierarquia é constituída pelas dimensões **Tipo** e **Categoria**.

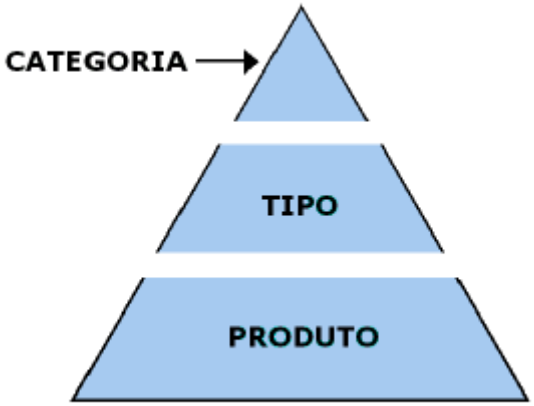


Figura 4.1 - Hierarquia explícita de Produto.
Fonte: (HOKAMA, 2004)

4.3.2.2 *DRILL-DOWN* E *ROLL-UP*

Uma característica bastante interessante nas análises dimensionais é a possibilidade de detalhamento e de sintetização das informações, que chamamos respectivamente de *drilldown* e *roll-up*. Veja detalhes no capítulo 5, referente à OLAP.

4.4 NORMALIZAÇÃO DAS TABELAS

As tabelas Dimensão, diferentemente das tabelas Fato, estão mais sujeitas ao processo de desnormalização. Existem duas correntes diferentes com relação aos aspectos de normalização das tabelas Dimensão (BARBIERI, 2004), há aqueles que adotam o Esquema Estrela e os que adotam o Esquema Flocos de Neve.

A representação dos dados em um *Data Warehouse* é estruturada como em "cubo de dados", ilustrado pela Figura 4.2. Esta estrutura é chamada modelo dimensional e implementa-se a partir de estruturas conhecidas como esquema estrela *(star)* e esquema floco de neve*(snowflake).*

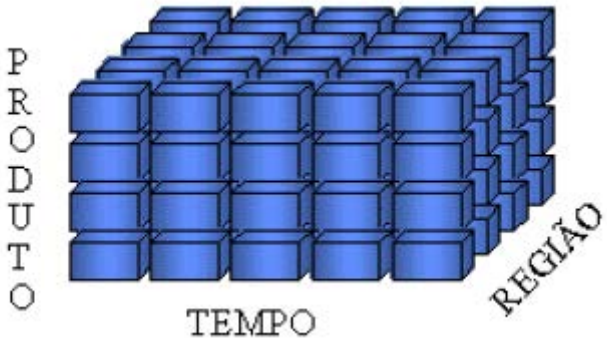


Figura 4.2 – Cubo de dados com 3 dimensões
Fonte: (HOKAMA, 2004)

4.4.1 ESQUEMA ESTRELA

O Esquema estrela, ilustrado na figura 4.3, é um dos esquemas dimensionais mais comuns. Consiste numa tabela de fatos e uma tabela para cada dimensão. A simplicidade

do esquema faz com que ele seja amplamente utilizado no mercado e, também, facilita o entendimento por parte dos usuários.

É uma estrutura simples, com poucas tabelas e ligações bem definidas, assemelha-se ao modelo de negócio. Permite a criação de um banco de dados que facilita a execução de consultas complexas, podendo ser realizadas de modo eficiente e intuitivo pelo usuário (HOKAMA, 2004).

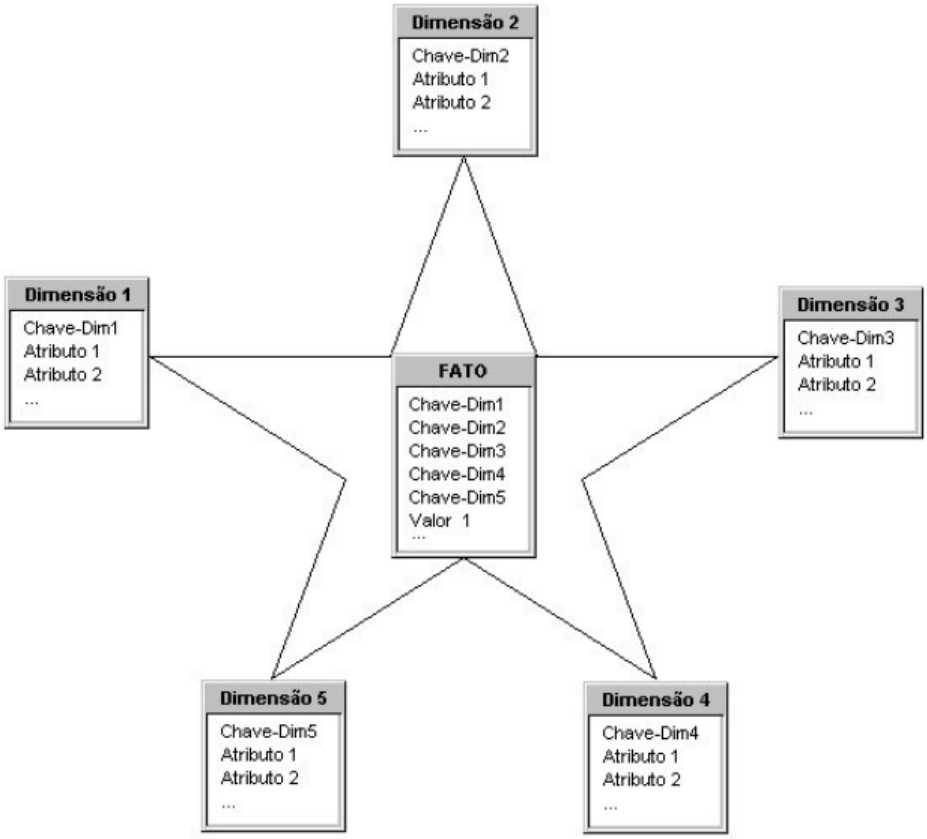


Figura 4.3 – Representação da estrutura do Esquema Estrela.

Fonte: (HOKAMA, 2004)

O nome "estrela" está associado à disposição das tabelas no modelo, que consiste de uma tabela central, a tabela de fatos, que se relaciona com diversas outras tabelas, as tabelas de dimensão.

4.4.2 ESQUEMA FLOCO DE NEVE

O esquema floco de neve é uma variação do esquema estrela, no qual todas as tabelas de dimensão são normalizadas na terceira forma normal (3FN), ou seja, são retirados das tabelas os campos que são funcionalmente dependentes de outros campos que não são chaves. Recomenda-se utilizar o esquema floco de neve apenas quando a linha de dimensão ficar muito longa e começa a ser relevante do ponto de vista de armazenamento ([HOKAMA, 2004).

Se as tabelas de dimensão forem normalizadas podem surgir alguns problemas. A complexidade do modelo de dados aumenta e, conseqüentemente, diminui a compreensão desse modelo por parte dos usuários. A performance das consultas realizadas pelos usuários tende a diminuir devido o aumento do número de junções, e essa sim pode ser crucial para o sucesso do ambiente visto que os usuários esperam retornos cada vez mais rápidos.

A decisão de optar pelo esquema estrela ou pelo floco de neve deve ser tomada levando-se em consideração o volume de dados, o SGBD e as ferramentas utilizadas. Outra possibilidade é mesclar o esquema estrela com o esquema floco de neve, normalizando apenas algumas dimensões (HOKAMA, 2004).

Dessa forma, pode-se economizar com espaço de armazenamento e melhorar a performance em alguns tipos de consultas, do que se fosse utilizado o esquema floco de neve puro.

4.5 ABORDAGEM DE BILL INMON E RALPH KIMBALL

A abordagem de Inmon se concentrou inicialmente no estilo mais tradicional de construção de banco de dados, onde se buscava uma forte interação entre todos os dados da empresa, que habitavam áreas funcionais diferentes. A Prism Solutions, empresa de Inmon ofereceu uma metodologia onde a ênfase era em cima de *Data Warehouse* propriamente dito, de onde outros depósitos de dados secundários (*Data Marts*) são originados e construídos. O grande diferencial de Inmon, em relação aos Bancos de Dados tradicionais, estava no oferecimento de um conjunto de ferramentas fundamentais para se estabelecer um projeto de *Data Warehouse*, principalmente se a sua abrangência for ampla (BARBIERI, 2004).

A abordagem de Kimball veio com um estilo mais simples e incremental. Aponta para projetos de *Data Marts* separados, que deverão ser integrados na medida da sua evolução. A essência da abordagem está na etapa de projeto dos *Data Mart*, centrada na modelagem dimensional, com o conceito de esquema estrela (*Star Schema*). Tem como desvantagem a possibilidade de se produzir diversos *Data Mart*, sem uma perfeita coesão entre eles, além de uma provável duplicação de esforços na fase de extração, preparação e carga dos dados (BARBIERI, 2004).

4.6 CONSIDERAÇÕES SOBRE O CAPÍTULO

A modelagem de dados passou por algumas modificações ao longo dos anos. Neste capítulo foram tratados alguns destes métodos, porém foi dada ênfase na modelagem de dados multidimensionais, que se inicia na observação dos fatos relevantes da empresa, em seguida definindo-se o grau de detalhamento que se deseja obter destes fatos considerando as diversas perspectivas que este fato pode apresentar.

As vantagens e desvantagens, comparativamente, entre um modelo relacional e um modelo multidimensional ficam claras quando suas características são sumarizadas na Tabela 4.1 e no Quadro 4.1.

O ponto crítico da construção desse modelo é a análise do negócio. É preciso ter claramente definidas as rotinas e processos da empresa e as métricas que influenciam em cada uma delas. Essa visão permite a construção das tabelas fatos, que são um retrato da ação e um meio para se obter informações mensuráveis.

Seguidamente a concretização dessa informação, as dimensões oferecerão a visão do fato escolhido de acordo com o prisma da dimensão. Portanto é importante compreender também os aspectos que influenciam naquele fato.

CAPÍTULO 5 - *ON LINE ANALYTICAL PROCESSING* - OLAP

O ambiente de *Data Warehousing* é concebido para prover um repositório de informações de forma concisa e sumarizada, permitindo uma recuperação de informações de forma otimizada para produção de relatório e meios de análise da informação para os gerente e tomadores de decisão.

Esses dados são extraídos através de *front-ends*. Dentre estes, pode-se citar ferramentas tais como o OLAP (*On-Line Analytical Processing*), e suas variantes, que aplicada sobre um ambiente de *Data Warehouse* e *Data Mart*, permitem analisar de forma multidimensional os dados, possibilitando a extração de informações valiosas (*DW* BRASIL, 2006).

OLAP é implementado usando-se uma arquitetura cliente/servidor para prover respostas rápidas às consultas, criando um cubo ou hipercubo, tendo em cada dimensão uma visão diferenciada de análise dos dados, o que permite ao usuário utilizar-se deste ferramental para ter uma visão da mesma informação sobre prismas diferentes (*DW* BRASIL, 2006).

A aplicação OLAP soluciona o problema de síntese, análise e consolidação de dados, pois é o processamento analítico online dos dados. Tem capacidade de visualização das informações a partir de muitas perspectivas diferentes, enquanto mantém uma estrutura de dados adequada e eficiente. A visualização é realizada em dados agregados, e não em dados operacionais porque a aplicação OLAP tem por finalidade apoiar os usuários finais a tomar decisões estratégicas (HOKAMA, 2004).

5.1 FUNÇÕES DO OLAP

Ferramentas OLAP permitem avaliar as informações sobre diversos pontos de vista e diferentes níveis de granularidade.

Dados sobre granularidade, cubo, tabelas fato e tabelas dimensão foram esclarecidos no capítulo 4, sobre modelagem multidimensional.

As análises oferecidas pelo OLAP são as seguintes:

- **Drill Across**: Está relacionado com o fato de poder pular de um esquema para outro, desde que ambos tenham algumas dimensões em conformidade, ou seja, as mesmas dimensões estão compartilhadas. Permite o tratamento de informações, que embora, correlacionadas, estão em estruturas separadas, porém unidas por algumas dimensões coerentes (BARBIERI, 2004).

- **Drill up/down**: Estão relacionados com o nível de granularidade dos dados armazenados. Drill down relaciona-se com o fato de sair de um nível mais alto da hierarquia e buscar informações mais detalhadas, ou seja, em níveis menores, na mesma dimensão. Dril up, também chamado, roll up, é o inverso, permite que o usuário reduza o escopo da análise.

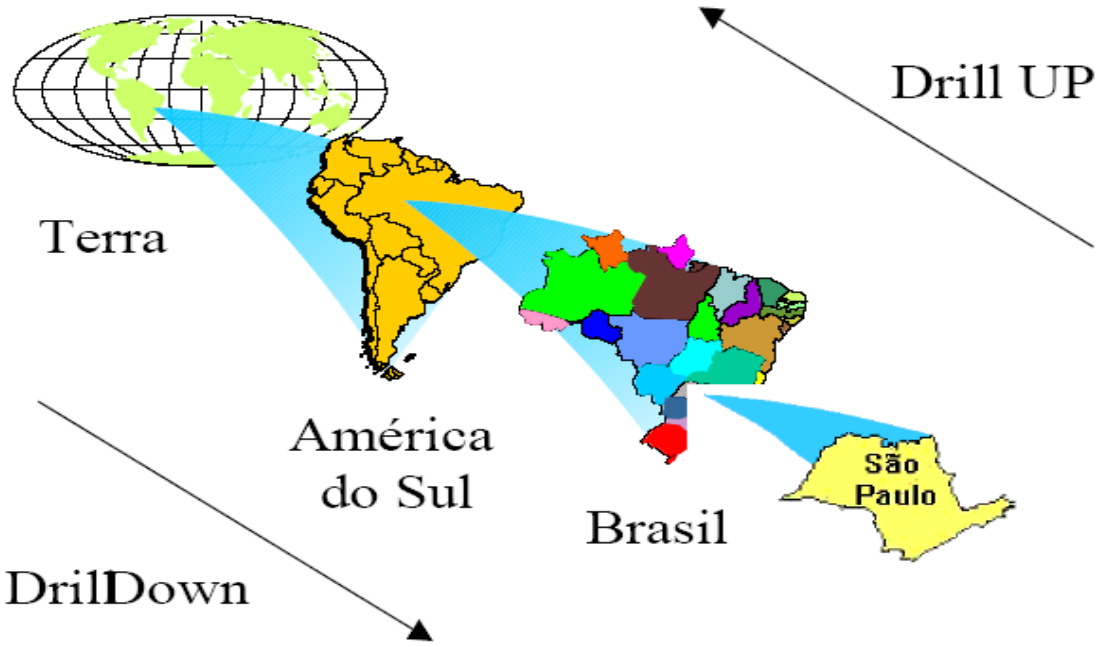


Figura 5.1 – Representação do Drill up e Drill down

Fonte: PUC – Rio de Janeiro, Laboratório de Tecnologia em Banco de Dados

- **Drill-Through:** Está relacionado com o fato de desejar uma informação num nível de detalhe menor do que aquele colocado na tabela Fato. Permite uma busca de informação além do nível de granularidade existente na estrutura dimensional (BARBIERI, 2004). É uma espécie de drill-down, só que busca o detalhe em uma outra estrutura.

- **Pivot**: As informações são organizadas em cubos para permitirem uma visão multidimensional do negócio, sendo que cada aresta é uma informação distinta, conforme pode ser observado da figura a seguir. A função *Slice e Dice* permite girar o foco da análise, "transformando" as colunas do cubo em linhas e as linhas em colunas (*DW* BRASIL-2006).

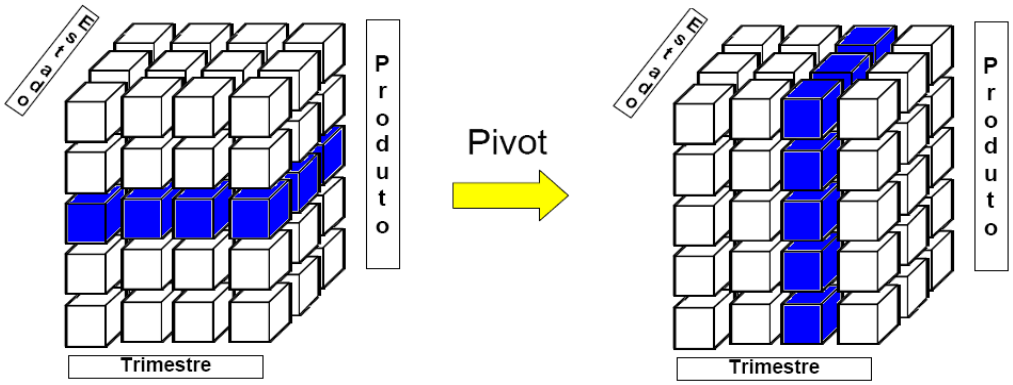


Figura 5.2 – Representação do processo de Pivot

Fonte: PUC – Rio de Janeiro, Laboratório de Tecnologia em Banco de Dados

- **Slice and Dice**: Permite analisar um valor simples em lugar de algo genérico.

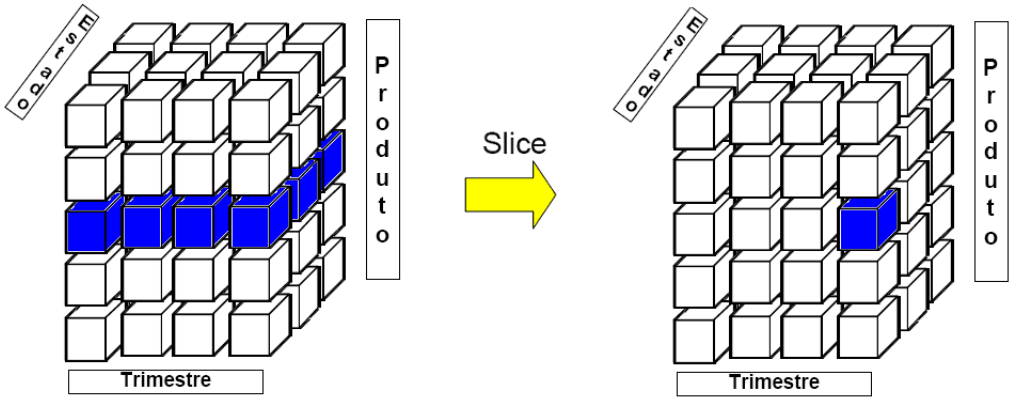


Figura 5.3 – Processo de Slice And Dice

Fonte: PUC – Rio de Janeiro, Laboratório de Tecnologia em Banco de Dados

- **Ranking**: Agrupamento da informação, geralmente numéricas, por ordem de valores, podendo partir dos maiores para menores ou inversamente.

- **Sorts**: Ordenação da informação de forma ascendente ou descendente.

5.2 VARIAÇÕES DO OLAP

Algumas variações são apresentadas nessa arquitetura, contanto, o que diferencia uma da outra, é a mudança apresentada na forma física de armazenamento das informações, mantendo a camada analítica semelhante entre elas. São as seguintes variações:

- **ROLAP** (OLAP Relacional): A consulta é enviada ao servidor de banco de dados relacional e processada no mesmo, mantendo o cubo no servidor. Essa arquitetura permite analisar enormes volumes de dados, em contra partida há sérios problemas de performance no servidor, em virtude da quantidade de usuários acessando-o simultaneamente.

- **MOLAP** (OLAP Multidimensional): Aplicada sobre Bancos de Dados Multidimensionais (MDB). O acesso aos dados ocorre diretamente no banco, ou seja, o usuário trabalha, monta e manipula os dados do cubo diretamente no servidor. Isso traz grandes benefícios aos usuários no que diz respeito à performance, mas tem problemas com escalibilidade além de ter um custo alto para aquisição.

- **HOLAP** (OLAP Híbrido): Também chamado de processamento híbrido. É uma mistura de tecnologias onde há uma combinação entre ROLAP e MOLAP. Combina a alta performance do MOLAP com a escalibilidade do ROLAP.

- **DOLAP** (Desktop OLAP): Nessa arquitetura há pouco tráfego na rede, visto que todo o processamento OLAP acontece na máquina cliente, e há maior agilidade de análise, além do servidor de banco de dados não ficar sobrecarregado, sem incorrer em problemas de escalabilidade.

CARACTERÍSTICA	ROLAP	MOLAP
Esquema	Utiliza Esquema Estrela Dimensões podem ser acrescentadas dinamicamente.	Utiliza Cubos de Dados Dimensões adicionais requerem que o cubo seja recriado.
Tamanho do Banco de Dados	Médio para grande.	Pequeno para médio
Arquitetura	Cliente/Servidor. Baseado em padrões abertos	Cliente/Servidor. Prioritária
Acesso	Suporte a consultas *ad hoc.* Dimensões limitadas	Limitado a dimensões predefinidas
Recursos	Altos	Muito Altos.
Flexibilidade	Alta	Baixa
Escalabilidade	Alta	Baixa
Performance	Bom para pequenos *datasets*, razoável para médios para grandes *datasets*.	Ótimo para pequenos *datasets*, razoável para médios para grandes *datasets*.

Tabela 5.1 – Comparativo MOLAP x ROLAP

Fonte: PUC – Rio de Janeiro, Laboratório de Tecnologia em Banco de Dados

5.3 OLAP x OLTP

Os sistemas de informações podem ser divididos em duas grandes partes. A primeira trata-se de sistemas utilizados dentro da empresa de forma operacional e transacional. São sistemas que procuram informatizar rotinas enfadonhas e/ou repetitivas, visando diminuir o esforço do funcionário para alcançar uma maior produtividade e um

melhor controle dos processos administrativos. Nesse tipo de aplicação são utilizados OLTP (*On-Line Transacional Processing*), podendo-se citar: sistemas de pagamento, controle de estoque, emissão de ordens de compra e etc.

Os outros se utilizam de OLAP (On-Line Analytical Processing) são sistemas voltados para análise do negócio, como Sistemas de Apoio à Decisão (SAD), *Enterprise Resource Planning (ERP) e Customer Relationship Management* (CRM), dentre outros.

CARACTERÍSTICA	**BD OPERACIONAL**	**DATA WAREHOUSE**
Objetivo	Operacional	Informativo
Processamento	OLTP	OLAP
Operação	Transações Simples	Consultas Complexas
Número de usuários	Milhares	Centenas
Usuário	Operadores, Projetistas de Sistema e Administradores de Sistema.	Analistas, Executivos e Usuários do conhecimento
Condições dos Dados	Dados Operacionais	Dados Analíticos
Estrutura dos Dados	Otimizados para transações	Otimizados para pesquisas complexas
Volume	MB e GB	GB e TB
Histórico	60 a 90 dias	5 a 10 anos
Granularidade	Detalhados	Detalhados e Agregados
Acesso a registro	Dezenas	Milhares
Atualização	Contínua (tempo real)	Periódica
Modelagem	Entidade-Relacionamento	Dimensional
Integridade	Transação	A cada atualização

Tabela 5.2 – Comparativo entre bancos operacionais e Data Warehouse

Fonte: PUC – Rio de Janeiro, Laboratório de Tecnologia em Banco de Dados

5.4 CONSIDERAÇÕES SOBRE O CAPÍTULO

Há alguns anos as bases de dados das empresas eram meramente transacionais, onde para a produção e análise das informações eram utilizadas ferramentas de processamento transacional em tempo real (OLTP). Com a nova modelagem dos dados e a criação de ambientes como *Data Warehouse* e *Data Mart*, este modo de avaliação do dado se mostrou ineficiente. Passou-se então, a existirem três saídas para esses novos meios de armazenamento de dados que são: Ferramentas que possibilitem uma análise analítica dos dados (OLAP), mineração de dados (*Data Mining*) e relatórios estáticos.

Este capítulo tratou da primeira saída oferecida. Essa ferramenta realiza a construção de cubos de dados, onde podem ser utilizadas várias funções do OLAP para avaliar a informação armazenada nesse cubo.

Ações que podem, por exemplo, permitir ao gerente subir e descer na estrutura de um cubo multidimensional e obter informações com alto ou baixo grau de especificidade.

Ferramentas OLAP apresentam ainda algumas variantes que permitem analisar não apenas cubos multidimensionais, mas também bancos relacionais. Essas variantes mostram-se importantes tendo em vista que numa organização estão presentes diversas bases de dados sobre formatos distintos.

CAPÍTULO 6 – *DATA MINING*

O armazenamento de informações nas empresas é algo fácil de ser feito, é diário e em grandes quantidades, abrangendo os diversos setores da empresa, porém o tratamento da relevância dessas informações é uma tarefa difícil. O intuito do *Data Mining* é fazer com que estes dados tenham relevância, ou seja, se transformem em informação importante na visão de quem realmente entende do ramo em que a empresa atua.

6.1 CONCEITO DE *DATA MINING*

"*Data mining* é o processo de extração de informações desconhecidas, porém significativas" (SINGH, 2001). Também conhecida como mineração de dados, o *Data Mining* explora dados geralmente ligados a negócios, pois serão esses dados que darão origem a informações de relevância que podem ser consideradas na tomada de decisão, sempre visando à vantagem competitiva. Esta informação é algo implícito que está previamente, na sua maioria, desconhecida.

O *Data Mining* não se trata de uma pesquisa ou um processamento de consultas dedutivas, mas sim, uma sistemática para a extração de informações de dados armazenados. Então, *Data Mining* é um processo de descoberta de informações, tais como: padrões, tendências, mudanças, anomalias, dependências, classificações, evoluções e associações; tudo isso é feito explorando bancos de dados.

6.1.1 IMPORTÂNCIA DO *DATA MINING* EM UM *DATA WAREHOUSE*

O *Data Mining* pode aumentar o valor de um *Data Warehouse*, pois no *Data Warehouse* deve-se saber o que perguntar, só assim, o usuário poderá identificar padrões, enquanto o *Data Mining* descobre os padrões implícitos nos dados, esses padrões muitas vezes não são identificados pelos usuários de um *Data Warehouse* (SINGH, 2001).

A utilização de um *Data Mining* em um *Data Warehouse*, é bastante viável, ao invés de se implementar um *Data Mining* em uma base onde ocorrem várias alterações durante o dia, exigindo um equipamento de rede e *hardware* bastante robusto para suportar o

paralelismo, no caso de um *Data Warehouse* as alterações não são tão constantes como nas bases de uso diário por diversos sistemas da empresa.

6.1.2 PROCESSO DO *DATA MINING*

A extração de dados feita no *Data Mining* ocorre geralmente aplicando atributos de seleção para relacionar um conjunto de dados que possuam relevância, esses dados agregados geram números estatísticos e resumos que podem ser úteis para uma tomada de decisão gerencial. Na figura a seguir temos o processo do *Data Mining*.

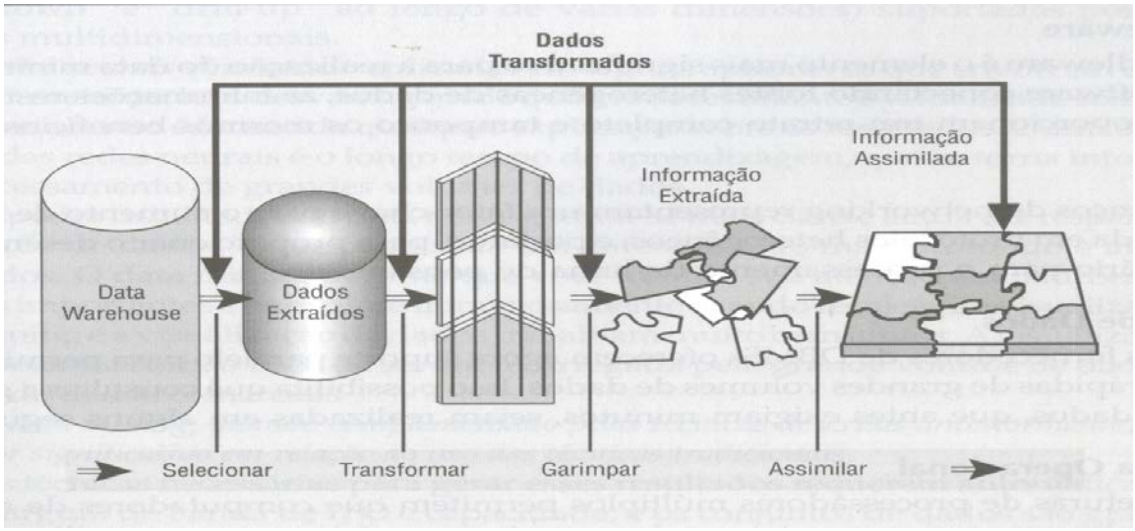


Figura 6.1 - Processo do Data Mining
Fonte: (SINGH, 2001)

A extração de informações ocorre em etapas, são elas:

Limpeza dos dados – os dados, nos seus devidos repositórios, geralmente não estão em um formato adequado para a garimpagem, por isso, nesta etapa os dados são limpos, onde os campos incompletos ou que não estejam de acordo, ou seja, não estão corretos, são corrigidos. Na figura 6.1 mostra uma extração em um *Data Warehouse*, o que facilita muito, pois em um *Data Warehouse* é pressuposto que os dados já estejam limpos e sem erros.

Extração dos dados – os dados são extraídos e colocados em uma nova base de dados de acordo com a parametrização previamente definida.

Transformação dos dados – os dados são transformados em informações, em resultados calculáveis, de modo que sejam facilmente manipulados.

Garimpagem – é a parte que demonstra as informações extraídas para verificação, valorização e análise, significa supor associações.

Assimilação – e o entendimento por parte do analista da informação fornecida pelo *Data Mining*

6.1.3 TÉCNICAS DE *DATA MINING*

O *Data Mining* apresenta um conjunto de técnicas e ferramentas, são estas técnicas e ferramentas que trabalham com a extração de dados. Segue algumas técnicas:

Redes neurais - Trata-se de uma técnica de difícil compreensão, onde se procura construir uma representação interna de padrões encontrados nos dados. Não é muito indicada quando ocorre a existência de muitos dados, pois as redes neurais constitui-se de entradas de dados exigindo uma grande circulação de dados podendo prejudicar a rede.

Introdução de Regras - Nesta técnica a interpretação de uma regra é geralmente feita pelo usuário, pois a este usuário é apresentada uma lista que indica tendências em um grupo de dados.

Visualização - Nesta técnica o programa de *Data Mining* efetua apenas a manipulação estatística. Como exemplo podemos verificar uma ferramenta que gere gráficos, são nestes gráficos que o usuário interpreta um espaço amostral de dados.

Davis King (KING, 1999) define que o método científico de um *Data Mining* se resume em 5 etapas, que são elas: definir o problema, gerar hipóteses/modelos, coletar dados, conduzir experimentos que gerem dados, testar modelos em confrontação com os dados e, por último, utilizar os resultados para gerar novas hipóteses, ou seja, os frutos gerados com o uso dos resultados da mineração de dados dá origem a novos dados a serem garimpados para que sejam verificados à ocorrência de novos padrões. Este processo ocorre, na sua maioria, com a intervenção da análise humana.

A grande importância do *Data Mining* é preparar dados para que sejam identificadas informações importantes. O resultado de um *Data Mining* é analisado com o intuito de se obter um novo conhecimento e definir a importância dessas informações e que elas sejam úteis para a tomada de decisão.

6.2 CONSIDERAÇÕES SOBRE O CAPÍTULO

Como citado na conclusão do capítulo anterior, *Data Mining* é uma das saídas possíveis de um *Data Mart/Data Warehouse*.

A mineração de dados é eficiente, juntamente com o Data Mart, para procurar por anomalias ou associações na base de dados. Essa mineração oferecerá um aumento na capacidade do *Data Mart*, pois, enquanto é preciso saber o que solicitar de informação o *Data Mining* aplicará uma sistemática para extração de informações que por ventura tenham passadas despercebidas.

Como já citado neste capítulo, a grande importância do *Data Mining* é preparar dados para que sejam identificadas informações importantes. O resultado de um *Data Mining* é analisado com o intuito de se obter um novo conhecimento e definir a importância dessas informações e que elas sejam úteis para a tomada de decisão.

CAPÍTULO 7 – ESTUDO DE CASO

7 PROBLEMA ESTUDADO

Neste capítulo serão demonstrados os resultados obtidos pela implantação do ambiente de *Data Mart* na empresa DFMAuto, bem como a verificação da validade dos conceitos acima apresentados.

Serão apresentadas ainda informações sobre o ferramental utilizado para construção desse ambiente.

7.1 BASES DE DADOS ATUAIS

A empresa DFMAuto apresenta em sua estrutura informacional a seguinte disposição:

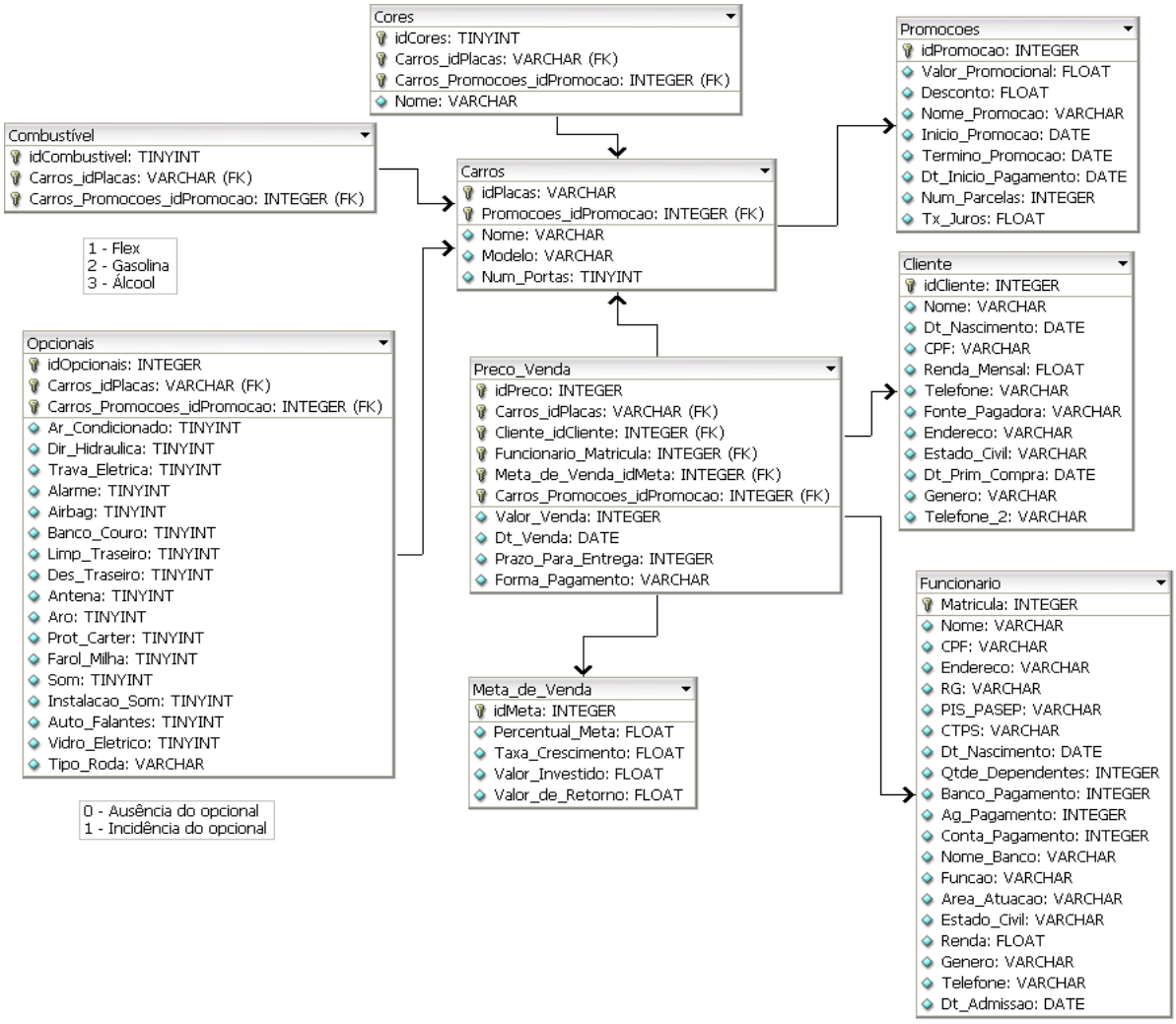


Figura 7.1 – Bases de Dados Relacional da DFMAuto

7.2 PROTÓTIPO

Conforme descrito no capítulo 1, a empresa DFMAuto enfrenta problemas para geração de informações que subsidiem a tomada de decisões, além de possuir um processo oneroso quando necessita realizar tal atividade, a emissão de relatórios para tomada de decisões é ineficiente, produzindo uma grande quantidade de dados, que devido a sua má formatação e apresentação, não são adequados ao fim desejado. Os relatórios são apresentados, muitas vezes, na forma de planilhas estáticas, de forma digital ou impressa, sem opção de outra forma para apresentação;

A arquitetura proposta por Ralph Kimball aborda um estilo simples e incremental, aborda a implementação de um *Data Mart* que permitirá à DFMAuto focar-se sobre a ação mais em evidência, as vendas, gerando informações valiosas aos gerentes e tomadores de decisão. Logo o *Data Mart*, sendo voltado somente para uma determinada área da empresa e apresentar custo e prazo menores, se mostrou como tecnologia adequada para suprir tal necessidade.

Como pode ser percebido nos pontos demonstrados no item 1.3, e avaliando-se a amplitude da questão, a implantação de um *Data Warehouse* seria demasiado complexo e dispendioso, já que a preocupação maior é na área de vendas.

O *Data Mart* será fisicamente estruturado adotando-se o Esquema Estrela, um dos esquemas dimensionais mais simples, o que faz com que seja amplamente utilizado no mercado. A escolha deste modelo garante uma maior simplicidade de implantação e entendimento por parte do usuário (gerentes e diretores). Este segundo ponto é de suma importância para o desenrolar do projeto, uma vez que para concepção de qualquer sistema, o maior ponto crítico para seu sucesso é o apoio da alta administração.

Serão utilizadas neste protótipo técnicas de BI para geração dos relatórios que darão base às tomadas de decisões. A essência do BI são as formas alternativas de tratamento da informação.

O BI envolve sistemas de *Back-End* e *Front-End*. Será implementado, como citado acima, o *Data Mart* como sistema *Back-End*, e utilizado uma ferramenta *Olap*, *Microsoft Analysis Services*, juntamente com um software, chamado ProClarity Analytics Platform, como *Front End.*

O Sistema Gerenciador de Banco de Dados a ser utilizado para armazenamento de dados será o SQL Server, versão 2000.

A Ferramenta *Proclarity* aplicada a um ambiente de *Data Mart* permite analisar de forma multidimensional os dados, possibilitando a extração de informações.

7.3 PROCLARITY

O *ProClarity Analytics Platform* é um ambiente de *software* que pode ser personalizado e otimizado para orientar os tomadores de decisão a agirem em tempo real.

Baseado em componentes para a construção de aplicações de BI, o *ProClarity Analytics Platform* utiliza o *Microsoft Analysis Services*, como servidor de processamento analítico on-line (OLAP). Sua arquitetura consiste em três camadas (apresentação, *business logic* e dados) e agrega mais de 30 componentes, que podem ser totalmente personalizados de acordo com a demanda de cada negócio.

Sua solução inclui um *Analytic Server*, camada intermediária que funciona como uma espécie de hub, centralizando e filtrando os dados de diversas fontes por meio do ambiente *Microsoft Analysis Services*. Através desse repositório de informações, os usuários do sistema acessam os relatórios e interagem com os dados necessários para otimizar a tomada de decisão. Essas informações, que podem ser extraídas pela Internet, Extranet, Intranet e até mesmo pelos aplicativos do *Microsoft Office*, geram conhecimento estratégico às empresas, o que assegura agilidade e confiabilidade no processo.

7.4 DELIMITAÇÃO DO ESCOPO

O primeiro passo para a construção de um ambiente de *Data Mart* é a análise do ambiente corporativo a fim de avaliar os elementos que deverão ser estruturados. (IMBALL-2001)

A área em maior evidência da empresa (com maior volume de operações) é o setor de Vendas. Segundo Kimball, a melhor opção para se implantar um *Data Mart* é a dissociação da empresa em diversos setores, gerando uma departamentalização funcional, logo, após a análise efetuada, o contexto para estudo será o setor de Vendas da DFMAuto.

Essa segregação permitirá avaliar de forma mais eficiente os fatos de maior relevância em cada setor, podendo, o *Data Mart*, posteriormente progredir para um *Data Warehouse*. Tal separação, no entanto, não significa dizer que as informações produzidas pelo setor de Vendas e armazenadas no *Data Mart* não estarão acessíveis, até certo nível, a outros setores.

7.4.1 ALGUNS QUESTIONAMENTOS

Durante as fases iniciais do projeto, que engloba a análise dos problemas, foram descobertas várias necessidades por parte dos usuários que nos permite tomar conhecimento sobre o domínio do problema.

Percebeu-se que diferentes usuários precisavam de uma maneira simples, rápida e fácil de enxergar os dados para que eles possam tomar decisões estratégicas a respeito de algum produto ou mesmo sobre ações de marketing. Alguns questionamentos que podem ser considerados relevantes para a tomada de decisão na empresa DFMAuto:

1. Qual o faturamento do período *x*?
2. Qual automóvel é mais vendido?
3. Quais modelos mais vendidos por vendedor?
4. Qual a faixa etária do cliente em relação ao carro que comprou?
5. Qual a quantidade de vendas por período?
6. Qual a quantidade de vendas por vendedor?
7. Quais os modelos mais vendidos de acordo com a cidade do cliente?
8. Qual modelo de carro mais vendido conforme o gênero do cliente?
9. Qual a quantidade de carros vendida em cada promoção?
10. Qual os modelos de carros mais vendidos conforme o gênero do vendedor?

7.5 DEFINIÇÃO DAS DIMENSÕES

Como mencionado anteriormente, o foco deste projeto será o setor de Vendas.

O setor de Vendas engloba a Divisão e a Gerência de Vendas, responsáveis especificamente, pela venda direta aos usuários e pela coordenação e adoção de ações relativas às vendas.

Considerando-se então as atividades neste contexto, as dimensões definidas para construção do ambiente de *Data Mart* são as seguintes:

- Carros;
- Clientes;
- Tempo;
- Promoções; e
- Vendedores.

O protótipo conterá então as cinco dimensões supra citadas que podem ser descritas da seguinte forma:

7.5.1 DIMENSÃO CARROS

Contém informações cadastrais dos carros que foram vendidos na empresa, assim como seus acessórios, opcionais, valor e forma de pagamento. Através dessa dimensão é possível gerar relatórios com resultados, como a forma de pagamento mais utilizada pelos clientes quando se compra carros com valores maiores de R$ 50.000,00.

7.5.2 DIMENSÃO CLIENTES

Contém informações sobre os clientes que efetuaram compras na empresa, incluindo faixa etária, renda e sexo. Com esses dados é possível gerar filtro relacionando a faixa etária do cliente e o carro que o mesmo comprou.

7.5.3 DIMENSÃO PROMOÇÕES

Contém informações sobre todas as promoções promovidas pela equipe de marketing da empresa. Através dessa dimensão pode-se filtrar em qual promoção há a maior número de venda de carros.

7.5.4 DIMENSÃO VENDEDORES

Contém informações cadastrais dos vendedores da empresa. Através dessa dimensão pode-se filtrar quais vendedores tem melhor rendimento em termos de número de venda na empresa.

7.5.5 DIMENSÃO TEMPO

Essa dimensão permite que se faça filtro por mês, bimestre, trimestre, semestre ou ano que esteja cadastrado na dimensão.

7.6 TABELA FATO: VENDAS

Todas as dimensões citadas acima permitirão analisar sob diferentes perspectivas um único fato: Vendas.

Portanto a tabela fatos do esquema apresentado representará as vendas de carros efetuados. A combinação desta tabela fato com as dimensões componentes do *Data Mart* permitirão aos gerente e diretores responderem questões de cunho gerencial, garantida a geração de informações estruturadas baseadas em fatos.

Será possível, por exemplo, determinar qual o tipo de carro mais vendido nos últimos quatro anos para pessoas entre 30 e 40 anos e que coincidentemente efetuaram suas compras em julho.

Por mais esdrúxula que a solicitação pareça, talvez essa informação permita ao gerente oferecer carros mais espaçosos concluindo que pessoas acima de 30 anos possuem filhos e que no mês de julho estão de férias e quem sabe saindo de viagem.

A possibilidade de geração de informações valiosas, adotando-se apenas as dimensões acima citadas é enorme, contudo, a preocupação maior quando da projeção deste

esquema foi em atender as necessidades informacionais da DFMAuto e a supressão dos problemas levantados, conforme citado no item 1.3.

Em suma, o grande diferencial que será oferecido pela solução adotada é a possibilidade de avaliar o mesmo fato sob diversos ângulos e considerando-se ainda, um período histórico. Tal fato no ramo da empresa em questão deverá permitir um ganho durante a análise das informações.

A tabela fato será composta, segundo a definição de Kimball, apenas por chaves estrangeiras (chaves primárias oriundas das dimensões) e métricas (valores que poderão ser calculados). Portanto se apresentam as seguintes chaves estrangeiras: identificadores das dimensões Carros, Tempo, Promoções, Vendedores e Clientes. Com referência às métricas: Quantidade de Vendas e soma do valor das vendas. Esta separação traz dois pontos importantes a serem considerados.

O primeiro deles é que toda consulta a ser realizada no *Data Mart* partirá da tabela fatos. Considerando o exemplo citado acima, o pedido de informação seria repassado diretamente à tabela fato, onde as chaves estrangeiras (organizadas cronologicamente) permitiriam a localização da informação solicitada em uma linha temporal.

Esta organização dos dados está em conformidade com o exposto no capítulo 4 referente a modelagem multidimensional, organização das tabelas fatos e dimensões e a organização do esquema estrela.

7.7 MODELO DO NEGÓCIO – DATA MART VENDAS – ESQUEMA ESTRELA

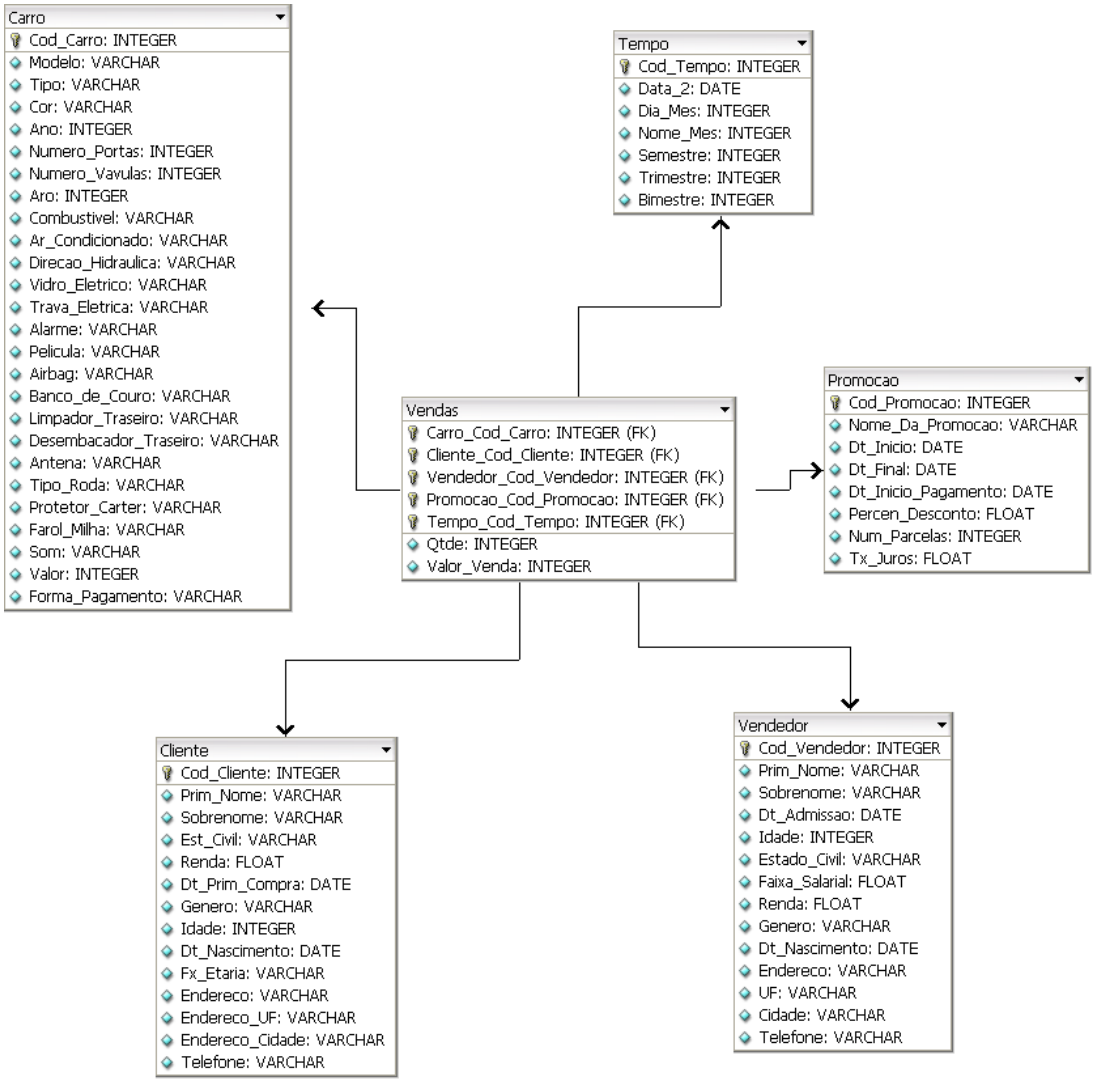


Figura 7.2 – Modelagem do Data Mart de Vendas

7.8 METADADOS – DATA MART DE VENDAS

Atributos - Tabela Fato	
Nome	***Descrição***
Cod_Carro	Identificador dos carros vendidos;
Cod_Cliente	Identificador dos clientes que já compraram na empresa;
Cod_Promocao	Identificador das promoções realizadas na empresa
Cod_Vendedor	Identificador dos vendedores da empresa
Cod_Tempo	Identificador da temporalidade
Qtde	Calcula a quantidade de vendas
Valor_Venda	Calcula o valor da venda

Atributos – Dimensão Carros	
Nome	***Descrição***
Cod_carro	Código numérico que identifica de maneira única um determinado carro;
Modelo	Nome que caracteriza o nome que recebe o modelo de determinado automóvel;
Tipo	Nome que caracteriza qual o tipo de veiculo, significando sua principal destinação, passeio ou automóvel utilitário;
Cor	nome que indica a cor a qual o automóvel foi pintado;
Ano	Indica o ano de fabricação e montagem do veículo;
Numero_Portas	Número que indica a quantidade de rodas que o automóvel possui;
Numero_Valvulas	Número que identifica quantas válvulas o automóvel possui em seu motor;
Aro	número que identifica o tamanho da roda que o automóvel possui;
Combustível	Nome que identifica qual(quais) é(são) o(s) combustível(is) que o automóvel é apto a funcionar;
Ar_Condicionado	Campo que indica, através de uma string, se o automóvel “possui” ou “não possui” ar condicionado;
Direção_Hidraulica	Campo que indica, através de uma string, se o automóvel “possui” ou “não possui” direção hidráulica;
Vidro_Eletrico	Campo que indica, através de uma string, se o automóvel “possui” ou “não possui” vidro elétrico;
Trava_Eletrica	Campo que indica, através de uma string, se o automóvel “possui” ou “não possui” trava elétrica;
Alarme	Campo que indica, através de uma string,

	se o automóvel “possui” ou “não possui” alarme;
Pelicula	Campo que indica, através de uma string, se o automóvel “possui” ou “não possui” película;
Airbag	Campo que indica, através de uma string, se o automóvel “possui” ou “não possui” airbag;
Banco_de_Couro	Campo que indica, através de uma string, se o automóvel “possui” ou “não possui” acentos revestidos em couro;
Limpador_Traseiro	Campo que indica, através de uma string, se o automóvel “possui” ou “não possui” limpador de vidro traseiro;
Desembaçador	Campo que indica, através de uma string, se o automóvel “possui” ou “não possui” desembaçador traseiro;
Antena	Campo que indica, através de uma string, se o automóvel “possui” ou “não possui” dispositivo que auxilia na recepção de ondas de rádio;
Tipo_Roda	Nome que indica o modelo de roda, podendo ser esportiva ou em aço;
Protetor_Carter	Campo que indica, através de uma string, se o automóvel “possui” ou “não possui” revestimento de proteção a possíveis impactos na parte inferior do motor;
Farol_Milha	Campo que indica, através de uma string, se o automóvel “possui” ou “não possui” farol extra acoplado ao pára-choque;
Som	Campo que indica, através de uma string, se o automóvel “possui” ou “não possui” aparelho de som;
Valor	Número que indica o valor pago na compra efetuada por um cliente de um automóvel, indica valor em moeda em circulação no Brasil, emitido ou cunhado pelo governo para ser utilizado como meio de pagamento, o Real;
Forma_Pagamento	Informará a modalidade escolhida pelo cliente para pagar o veículo

Atributos – Dimensão Clientes	
Nome	***Descrição***
Cod_Cliente	Código numérico que identifica de maneira única o cliente;
Prim_Nome	Nome que identifica o primeiro nome do cliente, que é uma pessoa física ou jurídica que adquiriu um automóvel junto a DFMAuto;
Sobrenome	Nome que identifica o segundo nome do cliente, que é uma pessoa física ou jurídica que adquiriu um automóvel junto a DFMAuto;
Est_Civil	Nome que indica a situação conjugal do cliente;
Renda	Valor da renda para pessoas físicas e valor do faturamento para pessoas jurídicas;
Dt_Prim_Compra	Data em que foi efetuada a primeira compra pelo cliente;
Gênero	Indica o sexo do cliente;
Dt_Nascimento	Data de nascimento para pessoas físicas e data de registro na junta comercial para pessoas jurídicas;
Idade	Idade calcula à partir da data de nascimento para pessoas físicas;
Fx_Etaria	Intervalo numérico em que o cliente está em relação a sua idade;
Endereço	Endereço da residência do contratante ou do estabelecimento quando se tratar de uma pessoa jurídica;
UF	Nome que indica a unidade federativa onde o cliente reside;
Cidade	Nome que indica a cidade onde o cliente reside
Telefone	Número que indica o telefone de contato do cliente;

Atributos – Dimensão Promoções	
Nome	***Descrição***
Cod_Promocao	Código numérico que identifica de maneira única a promoção;
Data_Inicio	Corresponde à data em que se iniciou a promoção;
Data_Final	Corresponde à data em que a promoção se encerrou;
Data_Inicio_Pagamento	Corresponde a data estipulada para se iniciar ou quitar um automóvel de acordo com a dada promoção;
Percent_Desc	Percentual de desconto concedido para a venda;

Num_Parcelas	Número máximo admitido de parcelas para pagamento;
Tx_Juros	Percentual estipulado de juros para as compras parceladas na promoção;

Atributos – Dimensão Vendedores	
Nome	***Descrição***
Cod_Vendedor	Código numérico que identifica de maneira única o vendedor;
Primeiro_Nome_Vendedor	Nome que identifica o primeiro nome do vendedor cliente;
Nome_Vendedor	Nome que identifica o nome completo do vendedor cliente;
Estado_Civil	Nome que indica a situação conjugal do vendedor;
Renda	Valor da renda média do vendedor
Gênero	Indica o sexo do vendedor;
Idade	Indica a idade do vendedor;
Data_Nascimento	Data de nascimento do vendedor;
Endereço	Endereço da residência do vendedor
UF	Nome que indica a unidade federativa do vendedor
Cidade	Nome que indica a cidade do vendedor
Telefone	Número que indica o telefone do vendedor;
Data_Admissao	Data em que o vendedor foi admitido

Atributos – Dimensão Tempo	
Nome	***Descrição***
Cod_Tempo	Código numérico que identifica de maneira única o período;
Data	Corresponde à data em que efetivamente foram realizadas vendas de automóveis.
Nome_Mes	Nome relativo ao período do mês;
Semestre	Refere-se ao primeiro semestre, que são os primeiros seis meses do ano, e o segundo semestre que são os seis últimos meses do ano;
Trimestre	Corresponde ao conjunto de três meses, sendo o primeiro trimestre formado por: janeiro, fevereiro e março; segundo trimestre: abril, maior e junho; quarto trimestre: julho, agosto e setembro; quarto trimestre: outubro, novembro e dezembro;
Bimestre	Corresponde a dois meses do ano em cada bimestre, sendo o primeiro bimestre

	formado por: janeiro e fevereiro; segundo bimestre: março e abril; terceiro bimestre: maio e junho; quarto bimestre: julho e agosto; quinto bimestre: setembro e outubro; sexto bimestre: novembro e dezembro;
Ano	Número de quatro dígitos que representa o ano;
Dia_Mês	Corresponde ao dia do mês.

7.9 DEFINIÇÃO DA GRANULARIDADE

Uma das questões mais importante do protótipo é definir a granularidade, o nível de detalhe que chegarão as consultas .

Para a elaboração dos relatórios será um adotado um nível de granularidade não muito baixo. Poderâo ser realizadas consultas a nível de mês, bimestre, trimestre, semestre e ano.

Poderia ser utilizado um nível de granularidade menor, a nível de Dia, mas o número de dados no *Data Mart* aumentaria consideravelmente o desempenho das consultas cairia.

7.10 PROCESSO DE CRIAÇÃO DO CUBO

Para atender às necessidades dos usuários, foi definido que para a visualização dos relatórios de forma dinâmica e flexível, os dados serão estudados a partir da formação de um cubo de dados, com cinco dimensões e 3 medidas. Este será construído através do *Analysis Manager*, funcionalidade do *Analysis Service*, ferramenta OLAP, da *Microsoft*.

De posse desse ferramental será possível, por exemplo, evoluir dentro da estrutura do cubo. Caso solicitado um dado mais genérico, o gerente poderá descer ou subir nessa estrutura (*drill up e drill down*), conforme a granularidade determinada no projeto de implementação do *Data Mart.*

7.11 APLICAÇÃO – ORGANIZAÇÃO DO CUBO

Utilizando o ferramental descrito nos itens anteriores foi possível construir o modelo estrela utilizando-se os atributos especificados. A construção do modelo estrela foi realizada utilizando-se o *SQL Server 2000*, organizando-se as dimensões e a fato, conforme figura a seguir:

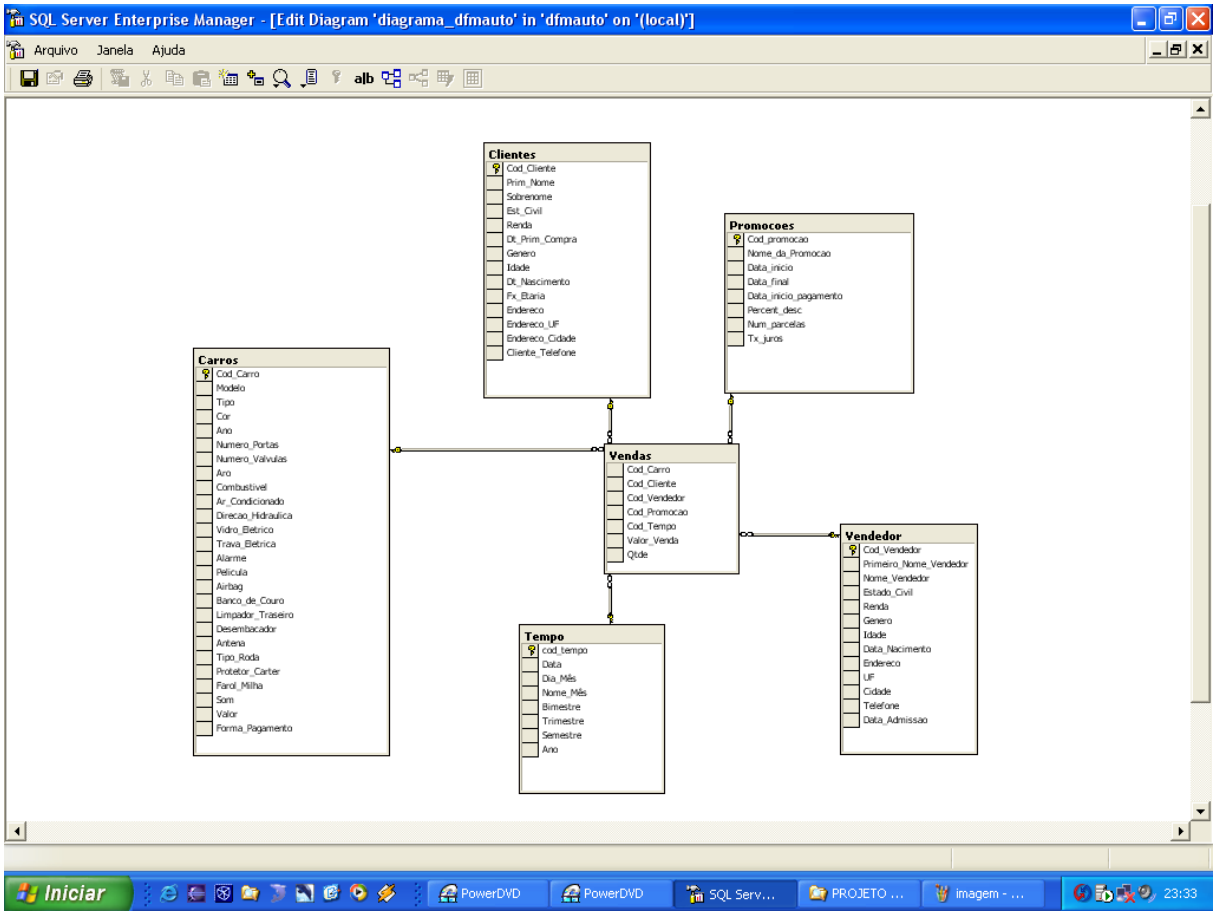


Figura 7.3 – Modelo Estrela do Data Marta da DFMAuto

Esta estrutura permitirá a realização das consultas que respondam aos questionamentos pontuados no item 7.4.1, além de possibilitar a geração de outros relatórios que o gerente por ventura necessite.

Utilizando-se o *Analysis Manager* do *SQL Server 2000* o cubo é organizado, realizando-se a criação das dimensões e da tabela fato, conforme o modelo estrela descreve na figura 7.3. A figura a seguir demonstra a criação de umas das dimensões que serão utilizadas:

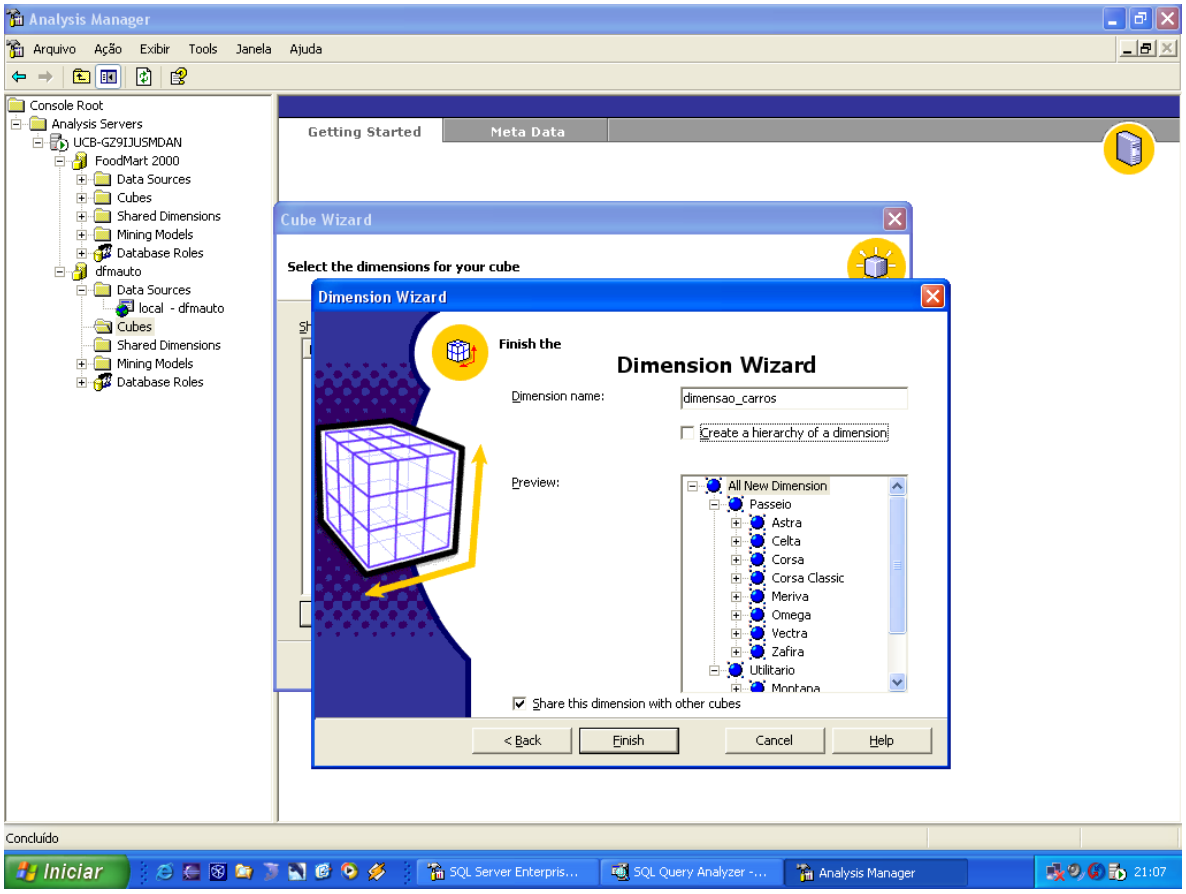


Figura 7.4 – Criação das Dimensões

Na criação da dimensão é possível determinar o nível da informação que a dimensão mostrará. Cada dimensão foi organizada de modo a permitir uma visualização das informações da forma mais genérica para a mais específica, descendo-se dentro da estrutura da dimensão, como pode ser observado, por exemplo, na figura 7.4. Os carros foram organizados primeiramente em Passeio e Utilitário, em seguida por Tipo, até chegar ao menor grau de especificação.

Após a organização das dimensões prossegue-se para a composição do cubo, conforme demonstrado na figura 7.5:

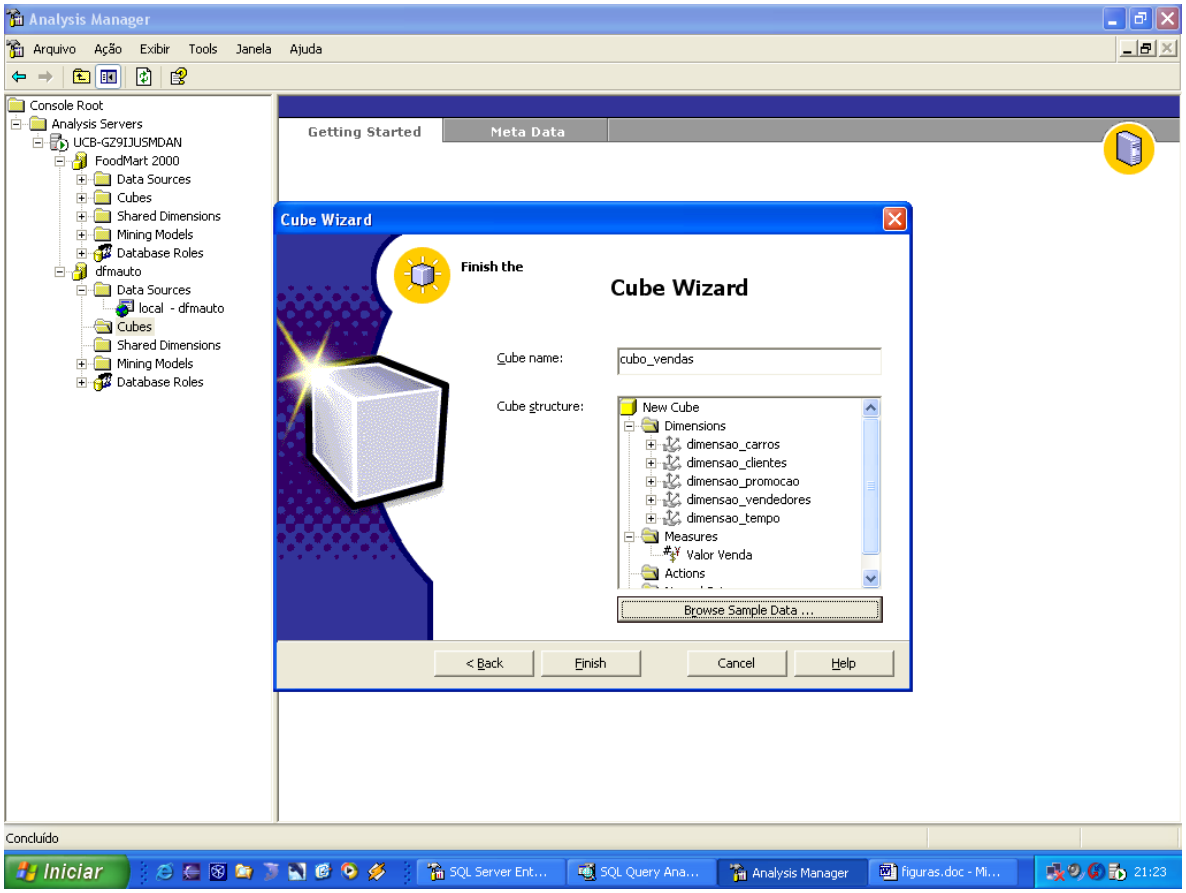


Figura 7.5 – Criação do Cubo de Vendas da DFMAuto

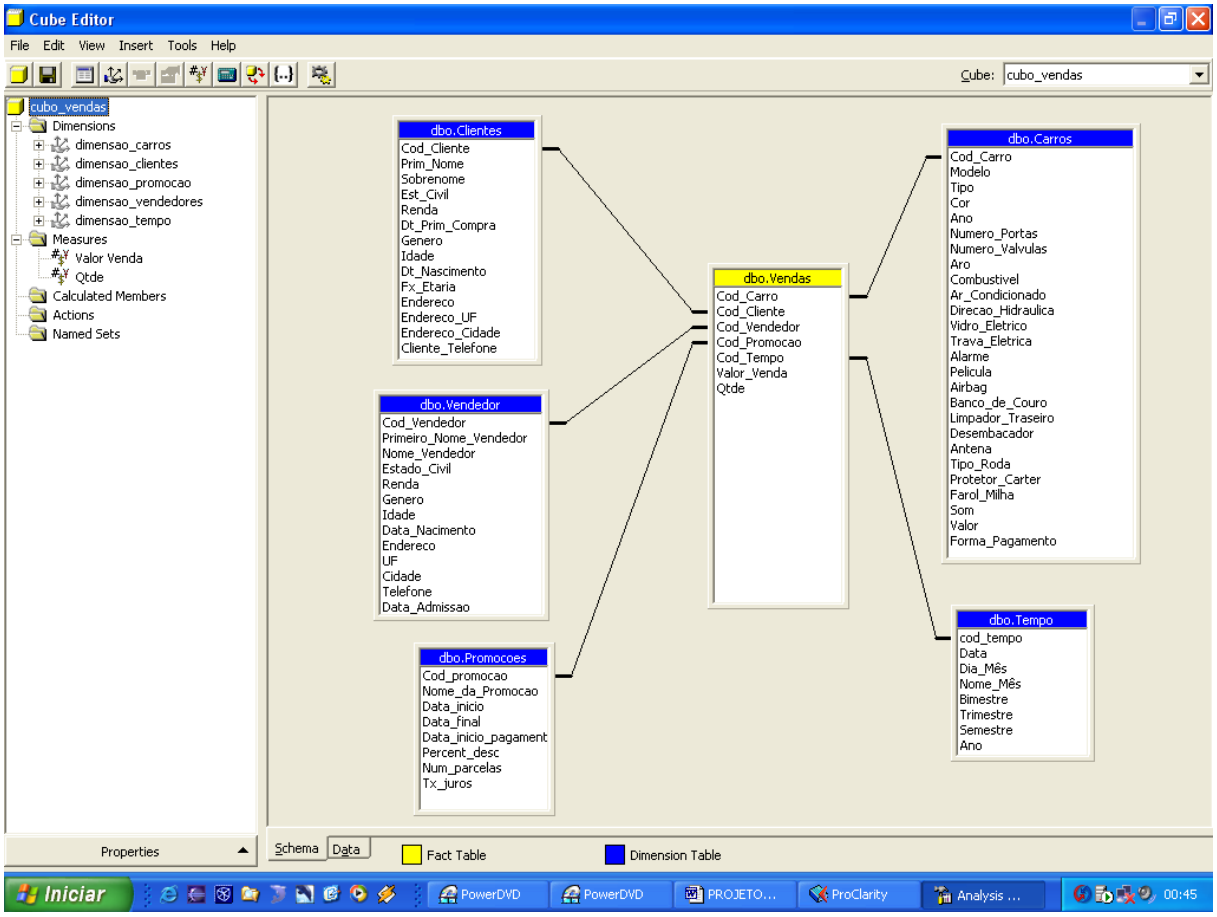


Figura 7.6 – Cubo organizado conforme o Esquema Estrela

7.12 APLICAÇÃO – CONSULTAS

Após o processamento do cubo utilizando-se a ferramenta *Analysis Manager*, torna-se possível responder aos questionamentos feitos pela gerência da DFMAuto.

Estas consultas são realizadas através da ferramenta *Proclarity Professional 5.2* que oferece três formas de visualização para as consultas: árvore de decisão, tabela e gráficos. As consultas foram realizadas para responder aos questionamentos da gerência, conforme item 7.1, mas também, foram demonstradas outras consultas, para exibir a funcionalidade e facilidade que esse ambiente de *Data Mart* passa a oferecer à DFMAuto. A seguir, são demonstradas as consultas realizadas:

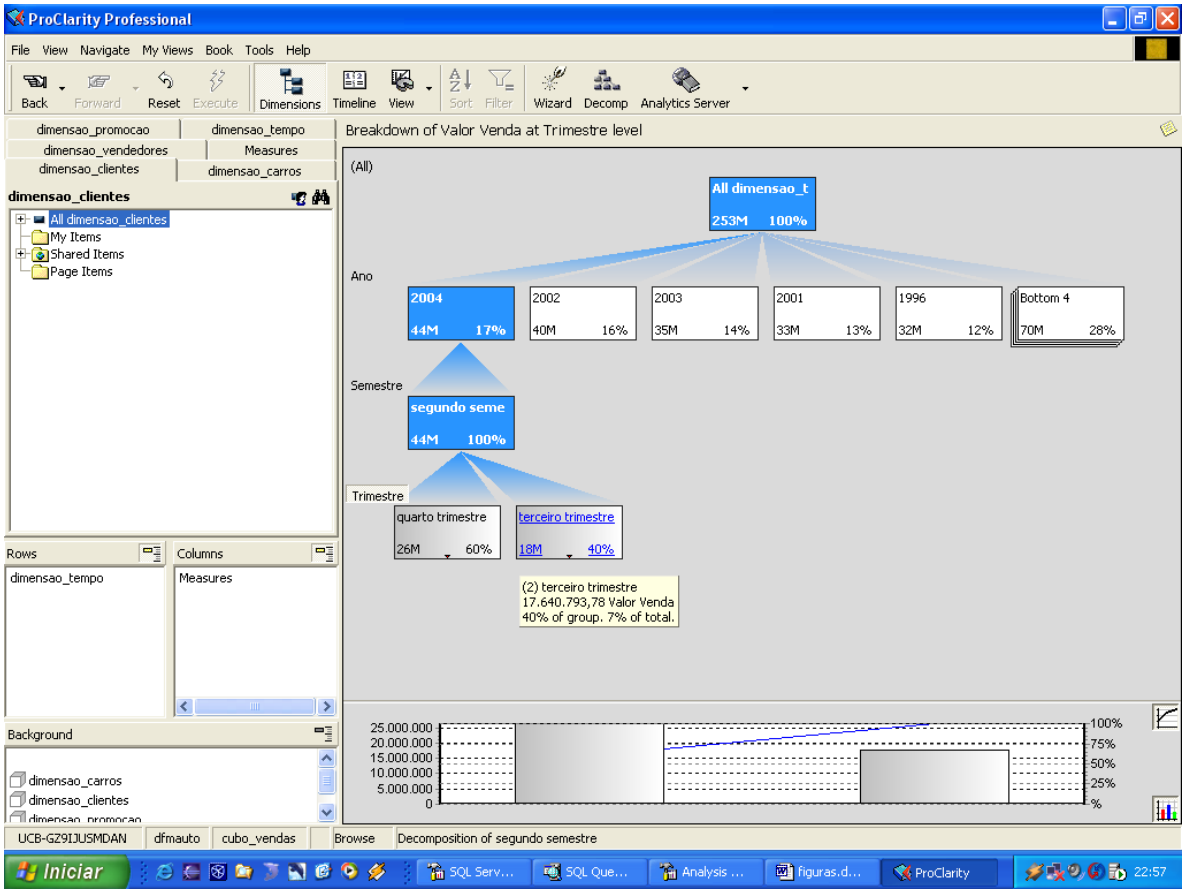


Figura 7.7 – Diagrama de Árvore de Decomposição

A figura 7.7 demonstra uma consulta realizada utilizando-se a árvore de decomposição para demonstrar o valor de vendas em determinado período. Com esta estrutura é possível realizar o processo de *Drill-Down* e *Drill-Up* para maior ou menor granularidade da informação. Nesta consulta em específico foi demonstrado o valor total de vendas no terceiro trimestre de 2004, mas como já citado, o gerente facilmente pode navegar dentro da estrutura de tempo e visualizar outros períodos.

A seguir, é utilizado um gráfico em pizza para demonstrar uma nova consulta:

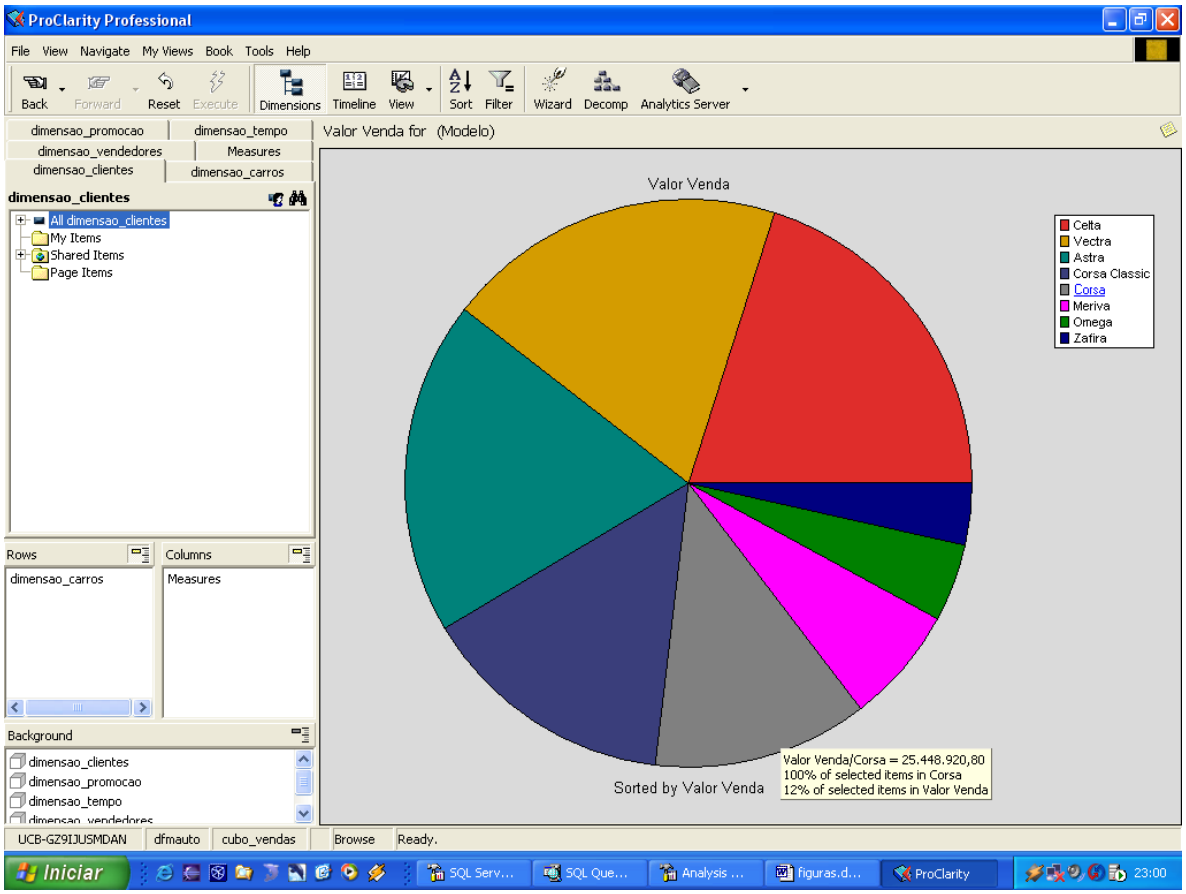


Figura 7.8 – Gráfico em pizza

A figura 7.8 demonstra a consulta realizada utilizando-se o gráfico em pizza para responder ao questionamento de qual o carro mais vendido e qual a representatividade do valor obtido com a venda deste modelo de carro no valor total arrecadado com as vendas.

É utilizado, posteriormente, o gráfico em barras conforme a figura 7.9:

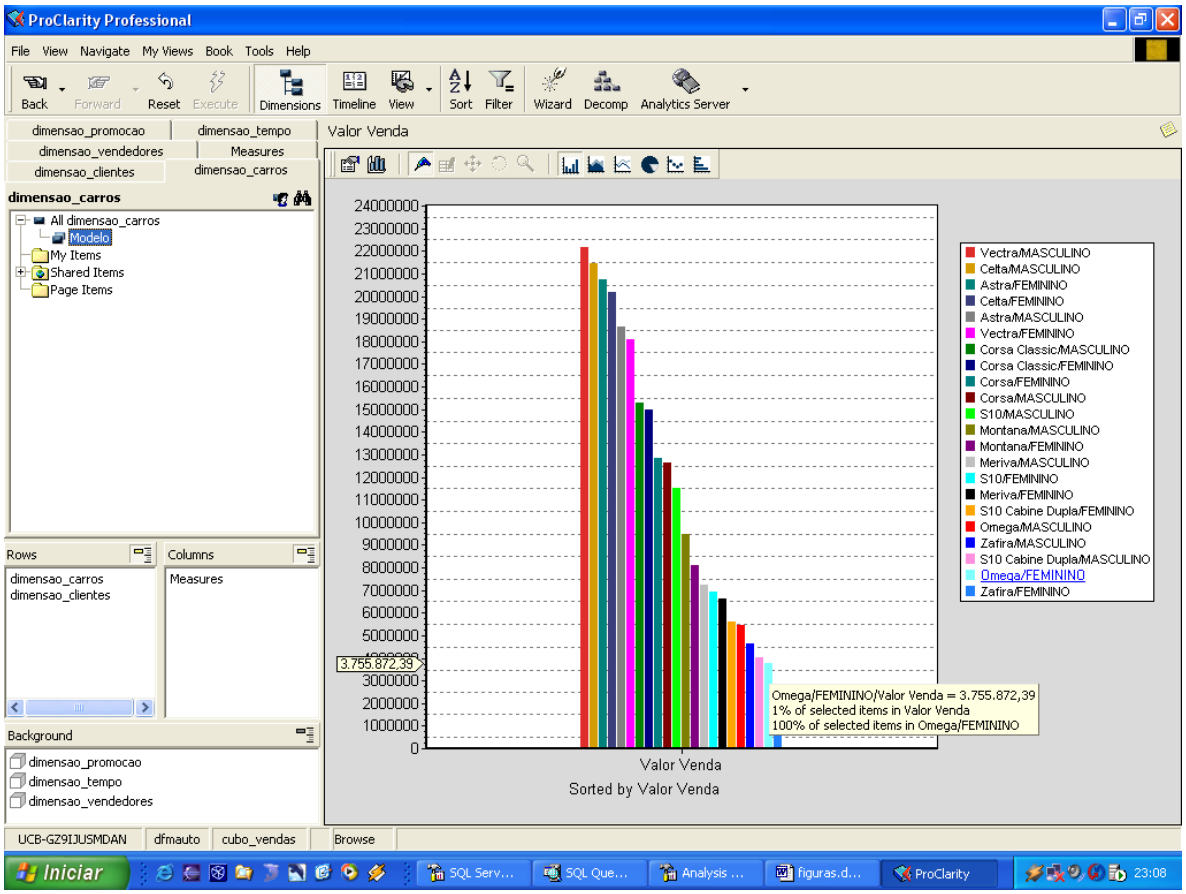
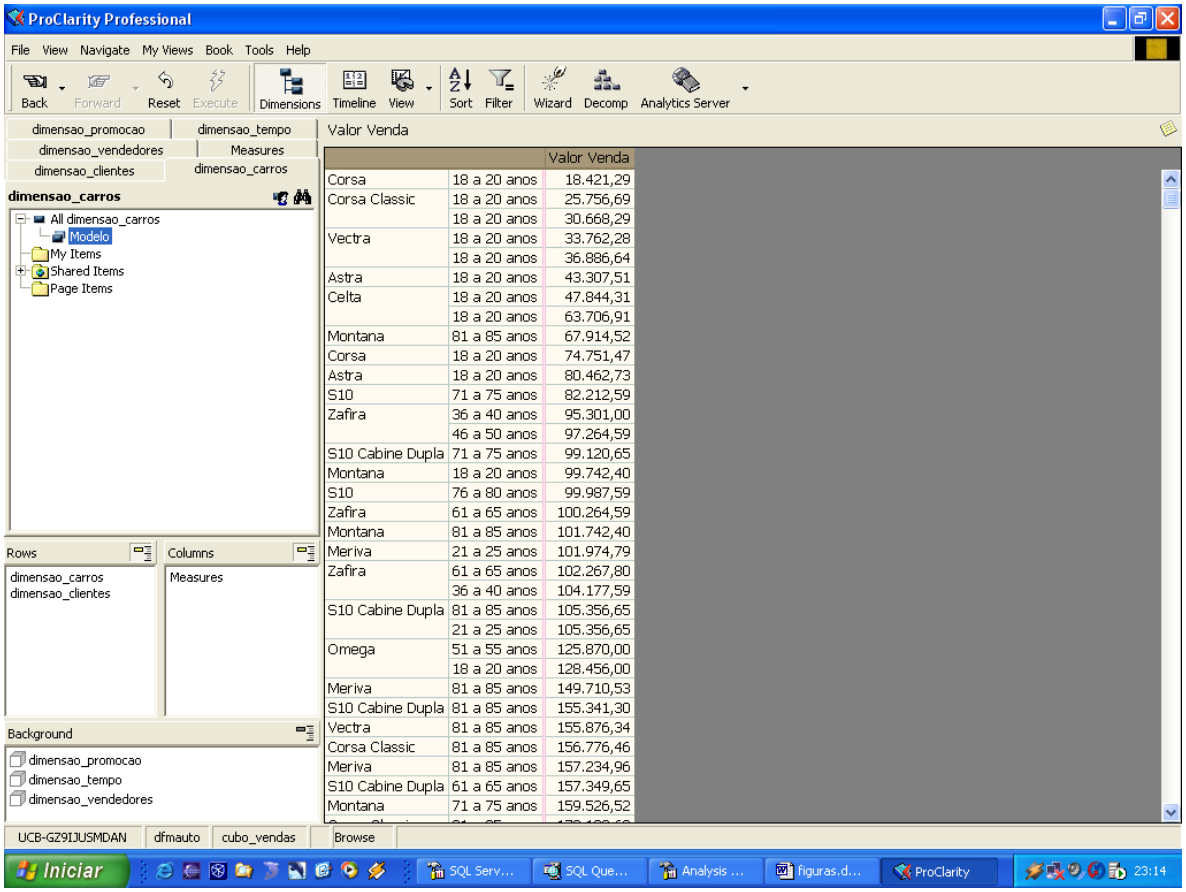


Figura 7.9 – Gráfico em barras

Nesta representação do gráfico com barras, foi demonstrado qual o carro mais vendido de acordo com o gênero (masculino ou feminino) do cliente. O ProCalrity permite ainda ao gerente, visualizar os valores que cada barra representa apenas colocando o ponteiro do mouse sobre o item desejado. Esta funcionalidade está disponível em todos as consultas gráficas.

O último tipo de consulta permitida pelo *ProClarity* (em formato de tabela) é demonstrado na figura a seguir:



		Valor Venda
Corsa	18 a 20 anos	18.421,29
Corsa Classic	18 a 20 anos	25.756,69
	18 a 20 anos	30.668,29
Vectra	18 a 20 anos	33.762,28
	18 a 20 anos	36.886,64
Astra	18 a 20 anos	43.307,51
Celta	18 a 20 anos	47.844,31
	18 a 20 anos	63.706,91
Montana	81 a 85 anos	67.914,52
Corsa	18 a 20 anos	74.751,47
Astra	18 a 20 anos	80.462,73
S10	71 a 75 anos	82.212,59
Zafira	36 a 40 anos	95.301,00
	46 a 50 anos	97.264,59
S10 Cabine Dupla	71 a 75 anos	99.120,65
Montana	18 a 20 anos	99.742,40
S10	76 a 80 anos	99.987,59
Zafira	61 a 65 anos	100.264,59
Montana	81 a 85 anos	101.742,40
Meriva	21 a 25 anos	101.974,79
Zafira	61 a 65 anos	102.267,80
	36 a 40 anos	104.177,59
S10 Cabine Dupla	81 a 85 anos	105.356,65
	21 a 25 anos	105.356,65
Omega	51 a 55 anos	125.870,00
	18 a 20 anos	128.456,00
Meriva	81 a 85 anos	149.710,53
S10 Cabine Dupla	81 a 85 anos	155.341,30
Vectra	81 a 85 anos	155.876,34
Corsa Classic	81 a 85 anos	156.776,46
Meriva	81 a 85 anos	157.234,96
S10 Cabine Dupla	61 a 65 anos	157.349,65
Montana	71 a 75 anos	159.526,52

Figura 7.10 - Tabela de consulta

A consulta realizada mostra qual tipo de carro mais comprado de acordo com a faixa etária do cliente, além de demonstrar quanto, em reais, cada uma representa.

As consultas ilustradas nas figuras de 7.7 a 7.10 demonstram a eficácia do ambiente construído para utilização da DFMAuto. Além destes questionamentos respondidos é possível organizar, facilmente, novas consultas considerando-se outras variáveis, cruzando-se dados de qualquer das dimensões.

Todas as consultas levam em consideração a métrica adotada (Valor das vendas). A métrica foi assim definida por ser o item de maior importância dentro do departamento escolhido.

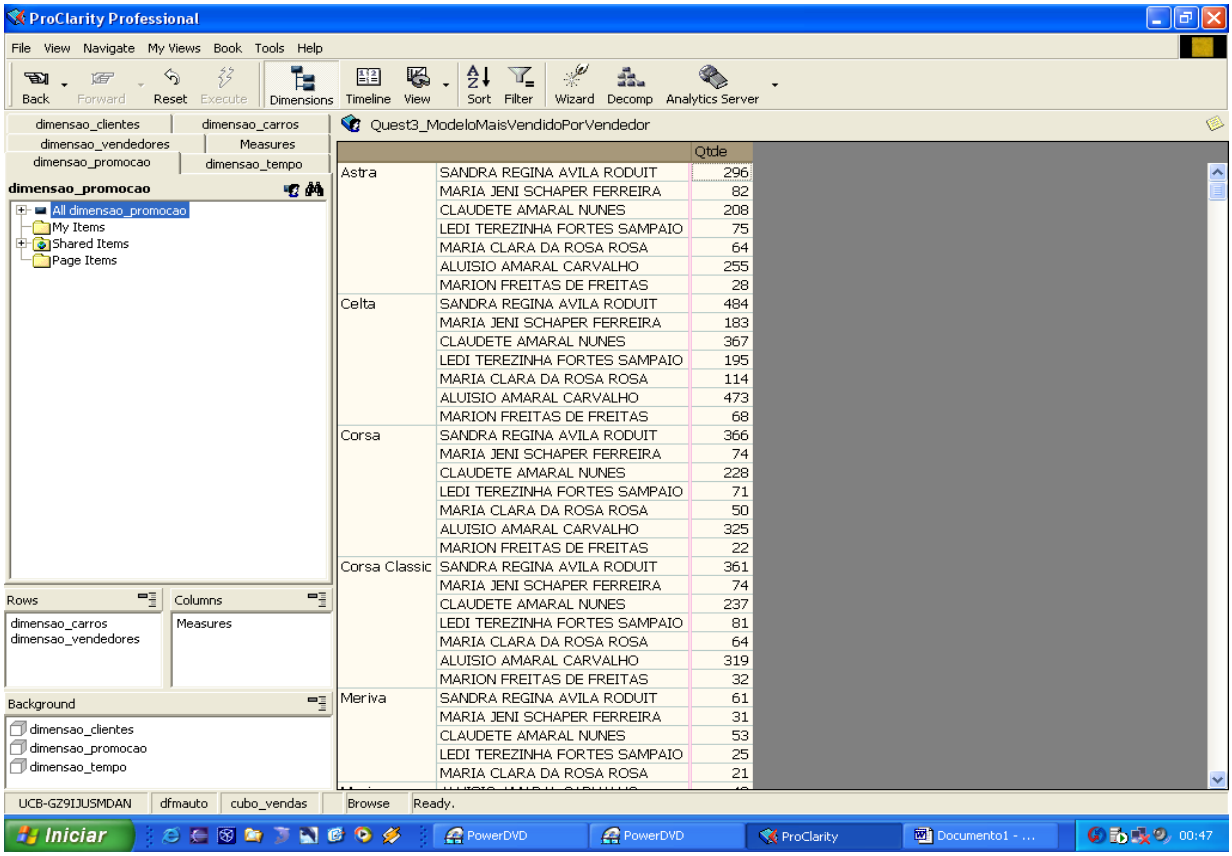


Figura 7.11 – Tabela de Consulta para questionamento nº 3

Para responder o questionamento sobre qual modelo de carro mais vendido por vendedor, foi utilizando uma tabela. A métrica “quantidade” permite realizar essa consulta separando por modelo e vendedores a quantidade que cada um vendeu.

Esse tipo de relatório permite ao Gerente, por exemplo, avaliar qual vendedor está mais apto para vender determinado modelo de carro, podendo então direcionar o funcionário para um tipo específico de veículo.

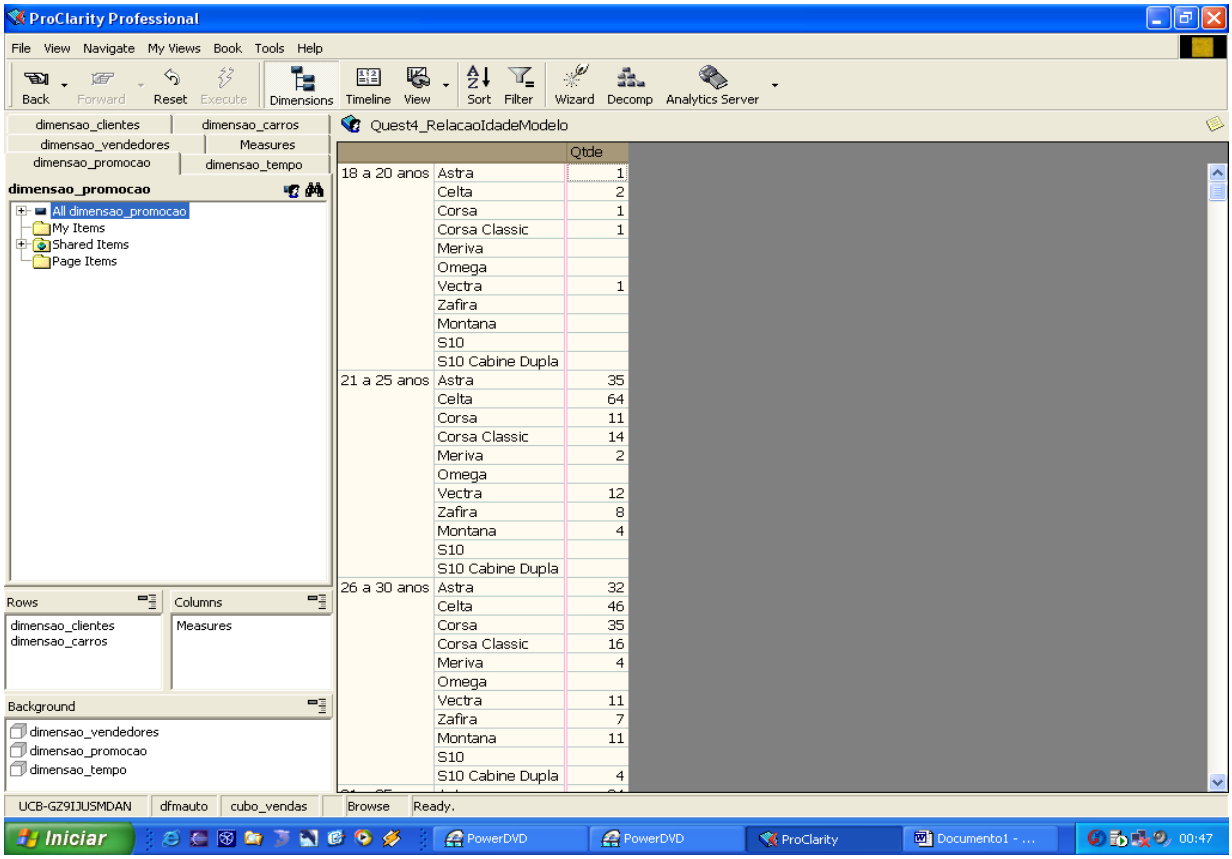


Figura 7.12 – Tabela de Consulta para questionamento nº 4

Este relatório responde ao questionamento sobre a relação existente entre a faixa etária do cliente e o carro que foi comprado.

Dessa forma é possível iniciar uma categorização do perfil dos clientes. Analisando o relatório é possível verificar que o modelo “Astra” é mais vendido para os clientes com faixa etária entre 21 e 25 anos. De posse dessa informação pode-se realizar uma promoção voltada para esta faixa etária de cliente, por exemplo.

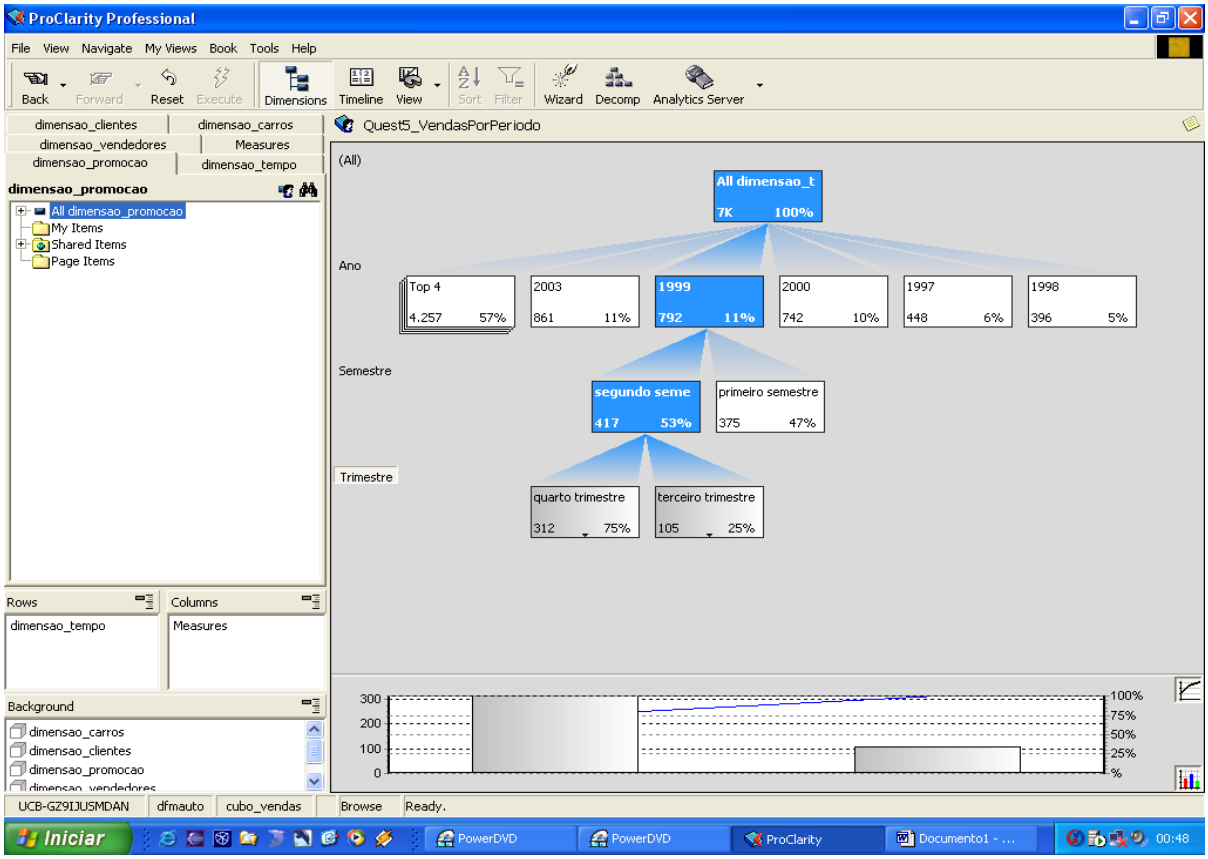


Figura 7.13 – Árvore de decisão sobre a quantidade de vendas por período

Uma das funções deste ambiente é proporcionar ao gerente uma forma de mais claramente os processos organizacionais.

Este diagrama permite verificar a quantidade de itens vendidos por período, além de possibilitar ao gerente “descer” ou “subir” na estrutura de tempo. Este relatório pode verificar a época do ano em que ocorrem mais vendas e direcionar uma promoção para um período que ocorra uma menor procura de veículos, por exemplo.

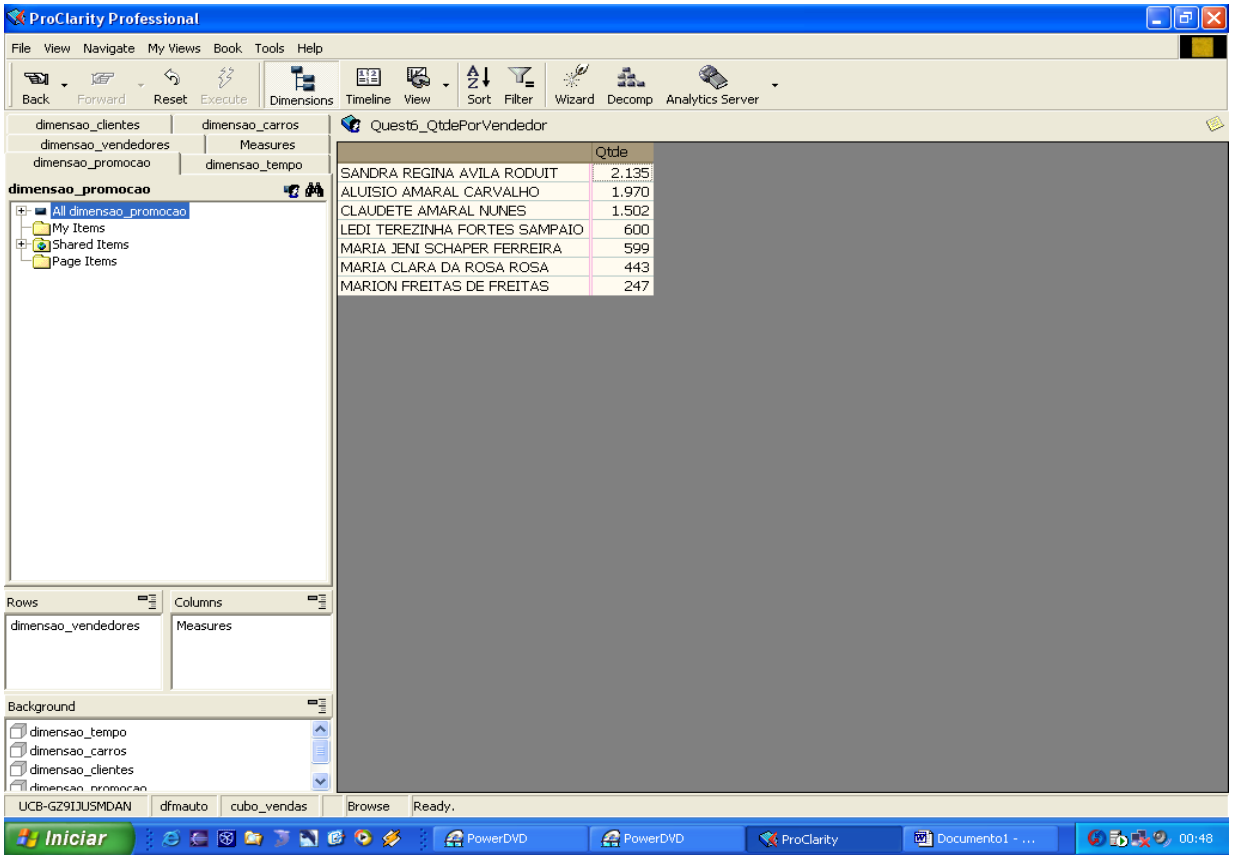


Figura 7.14 – Tabela de consulta para o questionamento nº 6

Para responder ao questionamento sobre qual vendedor realizou mais vendas foi utilizado um relatório em formato de tabela. Este relatório, contudo, pode ser facilmente transformado em um gráfico, caso o gerente ache mais fácil essa forma de análise.

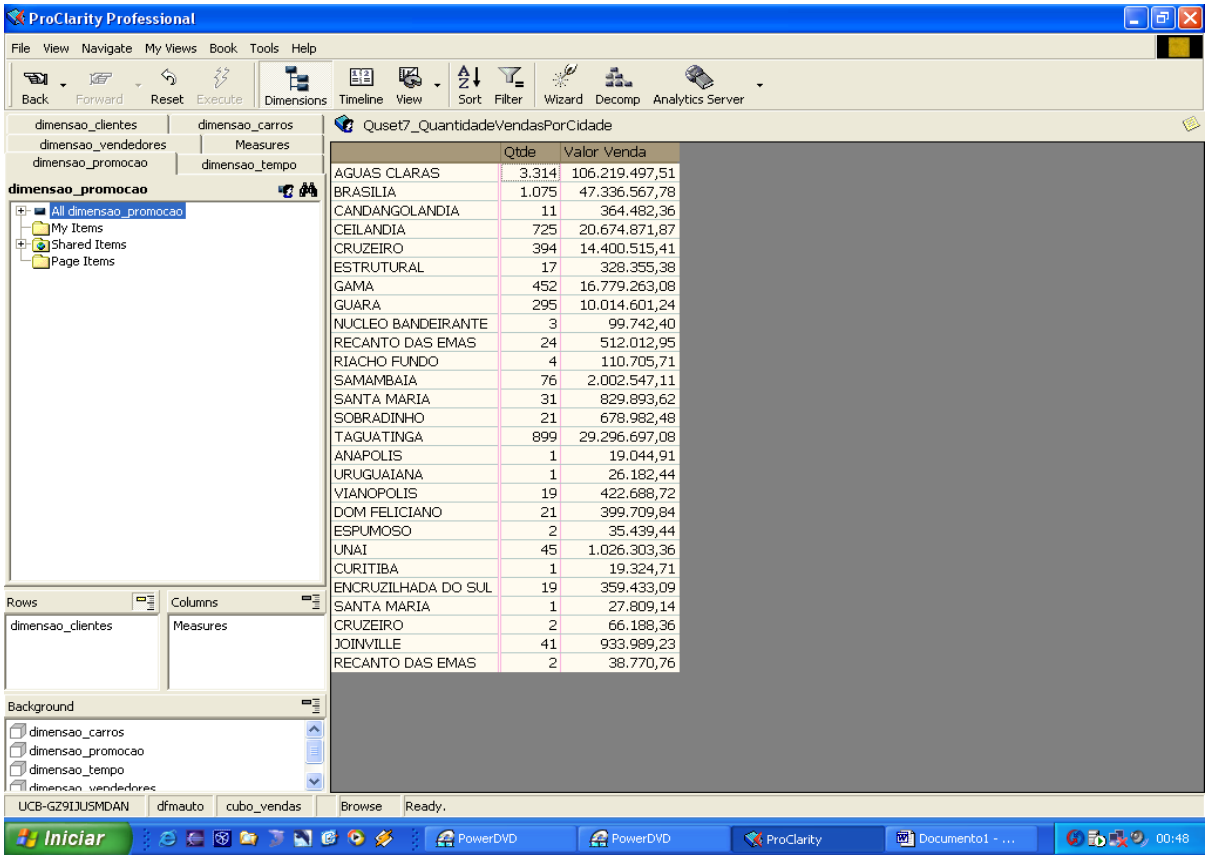


Figura 7.15 – Tabela de consulta para o questionamento nº 7

Uma consideração importante a ser feita na hora da elaboração de uma promoção é, certamente, para qual público ela será direcionada e, portanto, torna-se necessário saber para qual região seus produtos são mais vendidos.

O relatório acima demonstra a quantidade de vendas por cidade, conforme solicitado no item nº 7 do tópico 7.4.1.

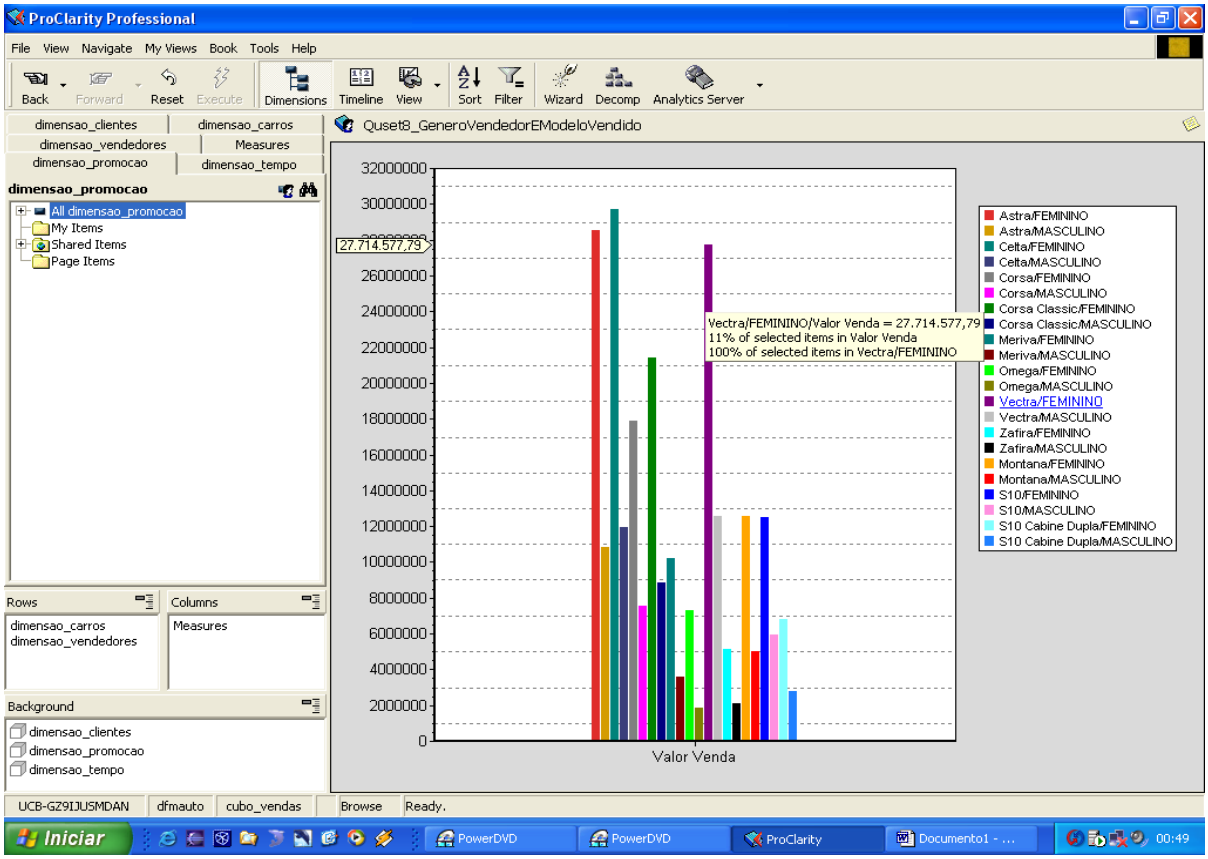


Figura 7.16 – Gráfico em barras para responder ao questionamento nº 10

Este relatório demonstra qual modelo de carro mais vendido conforme o gênero de vendedor. Utilizando-se de outros relatórios, já demonstrados neste capítulo, o gerente poderá então traçar um perfil ideal de vendedor para atender determinado tipo de cliente e para vender determinado tipo de carro.

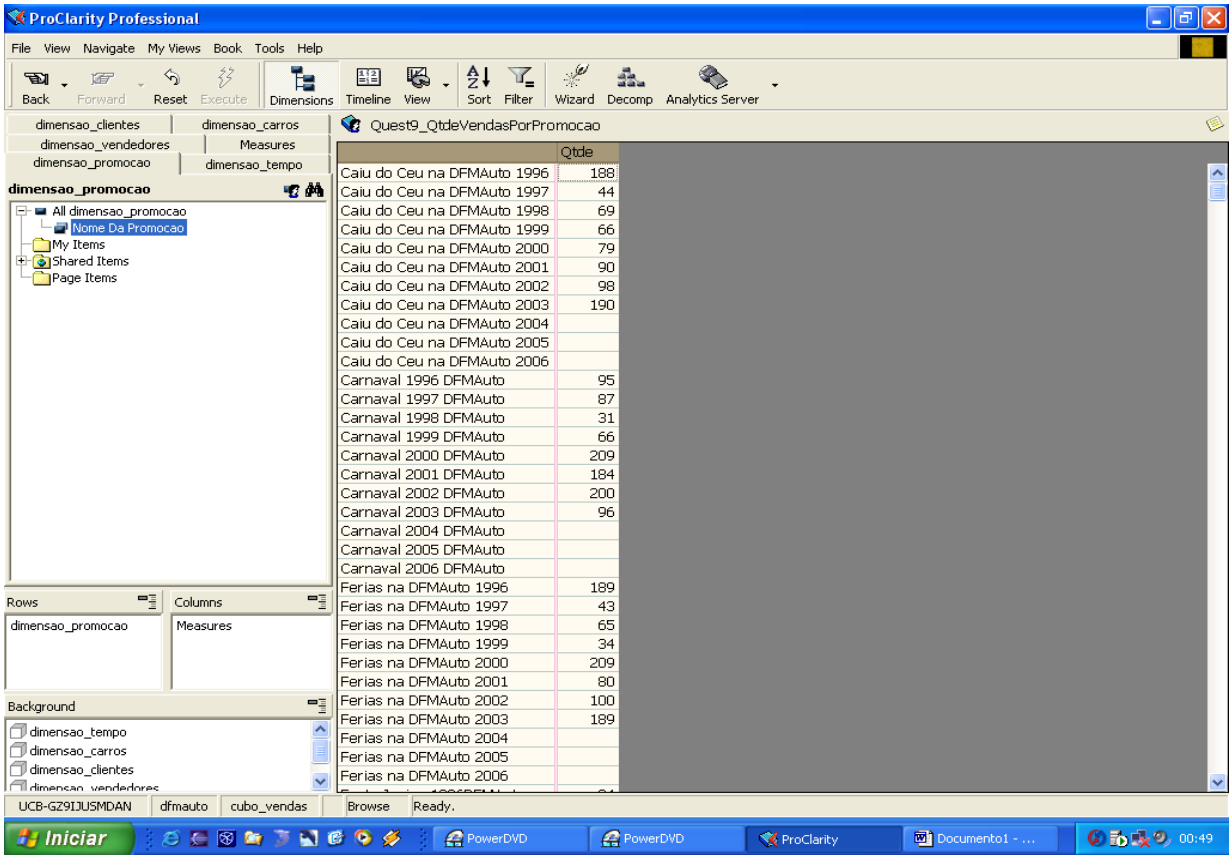


Figura 7.17 – Gráfico em barras para responder ao questionamento nº 9

É possível verificar, a partir desse relatório, a evolução das vendas ocorridas em cada promoção. Este tipo de informação permite ao gerente verificar a eficácia da promoção, avaliando a quantidade de produtos vendidos e determinando se a promoção foi ou não eficaz.

Estes relatórios trarão ao gerente a facilidade de cruzar dados importantes para seu negócio, pois o que antes eram apenas dados guardados em bases de dados transacionais, foram devidamente estruturadas para serem utilizadas em um ambiente de *Data Mart*, possibilitando uma visão analítica desses dados, gerando então informação.

7.13 SEGURANÇA DA INFORMAÇÃO

A empresa deve possuir uma política de proteção à informação, seja contra a perda, adulteração, espionagem e roubo, garantindo a confiabilidade, autenticidade e disponibilidade dos dados.

A política de segurança deve conter um esquema de importância e responsabilidade de documentos gerados, tanto por parte de quem gera os documentos quanto por parte daqueles que tem poder de analisar os documentos. Esses documentos não podem ser disponibilizados a todos na organização, como por exemplo, um relatório gerado com dados de clientes e faturamento da empresa em períodos; essas informações trazem um diferencial competitivo, mas em poder dos concorrentes podem se tornar uma grande arma contra a empresa a qual pertence à informação.

As máquinas, que contenham ferramentas de análise do cubo produzido, serão destinadas ao Gerente de Vendas e aos Diretores da empresa, que terão que disseminar a informação e analisar a informação dos resultados oferecidos pelo DM.

7.14 TREINAMENTO

Como se trata de uma nova tecnologia sendo implementada na empresa, os usuários devem receber treinamento adequando em relação a ferramenta que será implantada na empresa. Há previsão de um treinamento, não por parte de quem está desenvolvendo o DM, e sim em cursos oferecidos por escolas de informática que explicam como utilizar as ferramentas. Em relação ao desenvolvedor da solução haverá explicações e reconhecimento aos gerentes para que possam utilizar a ferramenta de forma adequada e possam extrair os relatórios desejados.

7.15 TRABALHOS FUTUROS

Data Mart/Data Warehouse têm como saída três opções: **relatórios estáticos**, onde o gerente pode pré-definir informações que constarão nas consultas e quando necessário, apenas atualizá-los, uma **ferramenta OLAP**, onde as consultas podem ser estruturadas de acordo com a necessidade do usuário, realizando o cruzamento entre dados que gerarão a informação requerida pelo gerente e por fim ***Data Mining,*** que poderá extrair informações valiosas, a partir do *Data Mart* ou *Data Warehouse*, que o gerente não tenha percebido.

No presente trabalho foi utilizado como saída do DM uma ferramenta OLAP, que proporcionará maior flexibilidade de consultas.

Como trabalho futuro propomos a integração de um *Data Mining* ao ambiente já construído que permitirá a geração de informações que podem ser valiosas ao gerente, mas que, porventura, tenham passadas despercebidas durante a manipulação dos dados pelo usuário.

Esta integração potencializará a capacidade de gerar informações valiosas ao tomador de decisão.

CONCLUSÃO

O início deste projeto se deu com o estudo das tecnologias envolvidas na concepção de um ambiente de repositório de dados. Este estudo possibilitou a correta escolha para implantação do citado ambiente. A escolha do *Data Mart* em detrimento ao *Data Warehouse* se mostrou crucial para o sucesso deste projeto, pois, sendo o *Data Mart* focado em um determinado departamento da empresa permitiu uma análise profunda sobre o processo em evidência dentro do Departamento de Vendas.

Após a opção da construção do *Data Mart* tornou-se evidente a necessidade de escolha da funcionalidade do *Data Mart*, ou seja, para qual função este se prestaria. Para tal, a necessidade de determinação do processo e das questões inerentes a este processo foram delimitadas, deixando claro o escopo do projeto. Sem esta delimitação do escopo, o projeto seria inviável, pois não seria possível nem se quer, determinar as dimensões, nem tampouco definir qual o fato a ser considerado.

Outra decisão importante a ser tomada nessa etapa é a granularidade do *Data Mart.* Foi definido um grau relativamente alto para este item, o que permite ao gerente sair de um período grande como, por exemplo, ano e atingir uma data específica.

As informações extraídas da fato vendas, serão verificadas a partir das dimensões. As dimensões permitem avaliar este fato sob óticas diferentes, o que permite ao gerente entender melhor seu negócio e tomar suas decisões baseadas num conjunto de informações consistentes.

O estudo de caso então se volta para a implantação deste ambiente na empresa DFMAuto. O ferramental adotado mostrou-se de difícil configuração, no entanto, a facilidade de uso denota que usuário final, que não participaria do processo de configuração, teria uma interação simplificada com as ferramentas.

O resultado esperado pela implantação desse ambiente foi alcançado com a resolução dos questionamentos levantados pela gerência da empresa, além de permitir a fácil construção de outras consultas.

O propósito principal para este estudo e para a implantação do ambiente *Data Mart* é galgar a gerência e o tomador de decisões de informações sólidas, facilitando a tomada de decisão e tirando a responsabilidade unicamente da experiência do profissional.

O ambiente não deve ser encarado como catalisador de vantagem competitiva e sim, como meio para alcançar esta vantagem, pois, este fator depende do profissional que fará uso destas informações.

SIGLAS E ABREVIAÇÕES

ETL – Processo de Extração, Transformação e Carga
DW – *Data Warehouse*
DM – *Data Mart*
BI – *Business Intelligence*
OLAP – *On Line Analitcal Processing*
MS-SQL – *Microsoft Structured Query Language*
DB Designer – *Data Base Designer*
Ghz – Giga Hertz
HD – Hard Disck
RAM – *Randomic Acces Memorie*
IN – Inteligência de Negócios
SAD – Sistemas de Apoio a Decisão
EIS – *Executive Information System*
CI – Inteligência Competitiva
KMS – *Knowledge Management System*
ODS – *Operation Data Store*
ETC – Extração, Transformação e Carga
EIS – *Executive Information System*
IDC – *International Data Corporation*
DSS – *Decision Support System*
Apps – *Adjustable Pressure Phase Separator*
3FN – Terceira Forma Normal
SGBD – Sistema Gerenciador de Banco de Dados
ROLAP – OLAP Relacional
MOLAP – OLAP Multidimensional
MDB – Bancos de Dados Multidimensionais
HOLAP – OLAP Híbrido
DOLAP – Desktop OLAP
OLTP – *On-Line Transacional Processing*
ERP – *Enterprise Resource Planning*
CRM – *Customer Relationship Management*
TB – Tera *Bytes*
GB – Giga *Bytes*
MB – Mega *Bytes*
UF – Unidade Federativa

REFERÊNCIAS BIBLIOGRÁFICAS

INMON, W. H. **Data Warehousing – Como transformar informações em oportunidades de negócios**, São Paulo: Berkeley, 2001.

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

PRESSMAN, Roger S. **Engenharia de Software**, Rio de Janeiro: McGraw-Hill, 2002.

SANTOS, Rafael. **Introdução à programação orientada a objetos usando Java**, Rio de Janeiro: Elsevier, 2003.

SINGH. Harry S. **Data Warehouse, Conceitos, Tecnologias, Implementação e Gerenciamento,** Rio de Janeiro: Makron Books, 2000.

ALCANTARA, Alexandre. Business Intelligence – Produzindo Resultados. **DW Brasil**. Disponível em:

<http://www.dwbrasil.com.br/html/artbi_20030602.html>. Acesso em: 09 mar. 2006.

BARBIERI, Carlos. **BI-Business intelligence: BI-Business intelligence** . Axcel Books: 2001. 424 p. BARBIERI, Carlos. **Business Intelligence**. Disponível em

<http://www.wgsystems.com.br/artigos>. Acesso em 26 mar. 2006.

CIELO, Ivã. **ETL – Extração, Transformação e Carga de Dados**. Disponível em

<http://www.datawarehouse.inf.br/artigos/etl.asp>. Acesso em: 09 mar. 2006.

COME, Gilberto de**. Os metadados no ambiente de data warehouse**. Disponível em:

<http://www.ead.fea.usp.br/Semead/4semead/Artigos/Mqi/Come.pdf>. Acesso em: 16 abr. 2006.

COSTA, ANCIAES. **Aspectos de Criação e Carga de um Ambiente de Data Warehouse**. Disponível em:

<http://dataware.nce.ufrj.br:8080/dataware/publicacoes/dataware/fisico/projetosFinais/datawarehousing/COSTA-ANCIAES-2001.pdf>. Acesso em: 14 abr. 2006.

COSTA, André Fernandes e ANCIAES, Felipe Curvello. **Aspectos De Criação E Carga De Um Ambiente De Data Warehouse**. Disponível em:

<http://dataware.nce.ufrj.br:8080/dataware/publicacoes/dataware/fisico/projetosFinais/datawarehousing/COSTA-ANCIAES-2001.pdf>. Acesso em 14 abr. 2006.

DW Architectura. **Guide to Data Warehousing.** Disponível em:

<http://data-warehouses.net/architecture>. Acesso em 05 jun. 2006

DW BRASIL. **Metadados**. Disponível em:

<http://www.dwbrasil.com.br/html/metadata.html>. Acesso em 12 abr. 2006

DW!. **Data Warehouse**. Disponível em:

<http://www.datawarehouse.inf.br/>. Acesso em 13 abr. 2006

FUJIWARA, Daniel. Bancos MOLAP. **DW Brasil**. Disponível em:

<http://www.dwbrasil.com.br/html/arto1_molap.html>. Acesso em: 14 mar. 2006.

HOKAMA, Daniel Del Bianco. **A modelagem de Dados no Ambiente Data WareHouse**. Disponível em:

<http://meusite.mackenzie.com.br/rogerio/tgi/2004ModelagemDW.pdf>. Acesso em 22 mar. 2006.

INFORMATION BUILDERS, **ETL Manager.** Disponível em:

<http://www.informationbuilders.fr>. Acesso em 05 jun. 2006

KAKINOHAMA, Regis Kichei. e-BI: **Uma Metodologia Para Gerenciamento de Projetos de Business Intelligence** . Disponível em:

<http://www.wgsystems.com.br/artigos/ebi-cits.pdf>. Acesso em: 05 mar. 2006.

KIMBALL, Ralph. **Especificações de Dimensões**. Disponível em:

<http://genesis.nce.ufrj.br/dataware/DataWarehouse/trabs9.htm>. Acesso em: 17 mar. 2006.

OLAP. **DW Brasil**. Disponível em:

<http://www.dwbrasil.com.br/html/olap.html>. Acesso em: 14 mar. 2006.

PENDSE, Nigel. What is OLAP? **The OLAP Report**. Disponível em:

<http://www.olapreport.com/fasmi.htm>. Acesso em: 14 mar. 2006.

OLAP architectures. **The OLAP Report**. Disponível em:

<http://www.olapreport.com/Architectures.htm>. Acesso em: 14 mar. 2006.

SAKANAKA, Paulo. e-BI: **Uma Metodologia Para Gerenciamento de Projetos de Business Intelligence**. Disponível em:

<http://www.wgsystems.com.br/artigos/regis1.htm>. Acesso em 22 mar. 2006.

SALETE, Maria. **Introdução a Metadados**. Disponível em:

<http://www.sqlmagazine.com.br/Colunistas/MariaSalete/02_Metadados.asp>. Acesso em: 12 abr. 2006.

SOUZA, Michael. **Business Intelligence**. Disponível em:

<http://www.imasters.com.br/artigo/1409/bi/business_intelligence>. Acesso em: 22 mar. 2006.

WG SYSTEMS – TECNOLOGIA PARA INTELIGÊNCIA EMPRESARIAL. **ETL (Extração, Transformação e Carga dos Dados)**. Disponível em:

<http://www.wgsystems.com.br/bi/etl.htm>. Acesso em: 09 mar. 2006.