

UMA ANÁLISE DA NOVA GERAÇÃO DE SISTEMAS DE APOIO À DECISÃO

Carlos Alberto Ferreira Bispo

Dissertação apresentada à Escola de Engenharia
de São Carlos da Universidade de São Paulo,
como parte dos requisitos para a obtenção do
título de Mestre em Engenharia da Produção

ORIENTADOR: Prof. Dr. Edson W. Cazarini

SÃO CARLOS

1998

**À minha querida
esposa Regina e minhas queridas
filhas Ana Paula e Amanda pelo
apoio dado para que eu chegasse
até aqui.**

AGRADECIMENTOS

Aos meus pais, que faleceram durante o transcorrer do curso.

Ao Professor Edson Walmir Cazarini, pela sua amizade, seu apoio e sua orientação.

Ao Professor Renato Vairo Belhot, pela sua amizade, seu apoio e seu auxílio prestado.

Aos meus amigos Ricardo Reis Tavares, Vanderlei Neves, Lauro André Lançone, Marcos César Rodrigues Pinto e Welinton Ferreira dos Santos, pelo grande apoio, incentivo e amizade.

À Universidade de São Paulo, pela oportunidade de poder participar do Programa de Pós-Graduação em Engenharia da Produção.

A todos os colegas, professores e funcionários do Departamento de Engenharia da Produção da EESC/USP, que, direta ou indiretamente, deram sua colaboração.

A todos os amigos e colegas de trabalho da Academia da Força Aérea, que, direta ou indiretamente, deram sua contribuição.

SUMÁRIO

LISTA DE FIGURAS	v
LISTA DE TABELAS.....	vii
LISTA DE ABREVIATURAS E SIGLAS	viii
RESUMO	xi
ABSTRACT	xiii
CAPÍTULO 1 - INTRODUÇÃO.....	1
1.1 - OBJETIVOS.....	2
1.2 - ESCOPO.....	3
1.3 - A PESQUISA BIBLIOGRÁFICA	3
1.4 - ESTRUTURA	4
CAPÍTULO 2 - REVISÃO BIBLIOGRÁFICA	5
2.1 - A EVOLUÇÃO DO PROCESSO DECISÓRIO	5
2.1.1 - O processo decisório na Antigüidade	5
2.1.2 - A evolução da decisão no século XX	6
2.2 - CONCEITOS BÁSICOS SOBRE DECISÃO	7
2.2.1 - Os fatores de influência do processo decisório.....	7
2.2.2 - O processo decisório nas organizações.....	9
2.3 - A EVOLUÇÃO DOS SISTEMAS DE APOIO À DECISÃO	10
2.3.1 - O surgimento do suporte computacional à decisão.....	10
2.3.2 - Os Sistemas de Apoio à Decisão nas décadas de 60 e 70	11
2.3.3 - Os Sistemas de Apoio à Decisão na década de 80.....	12
2.3.4 - Os Sistemas de Apoio à Decisão na década de 90.....	13
2.4 - AS LINHAS DE PESQUISAS DOS SISTEMAS DE APOIO À DECISÃO.....	16

CAPÍTULO 3 - DATA WAREHOUSE.....	19
3.1 - CONCEITOS BÁSICOS.....	19
3.2 - AS PRINCIPAIS FERRAMENTAS UTILIZADAS EM UM DATA WAREHOUSE	21
3.3 - A MODELAGEM DOS DADOS	23
3.4 - A QUALIDADE DOS DADOS	27
3.5 - SEGURANÇA DE DADOS.....	31
3.6 - ALGUMAS ANÁLISES NECESSÁRIAS AOS NEGÓCIOS	33
3.7 - A ELABORAÇÃO DE UM PROJETO DE DATA WAREHOUSE	36
3.8 - A ESCOLHA DO TIPO DE DADO A SER ARMAZENADO NO DATA WAREHOUSE	45
3.9 - METADADOS.....	47
3.10 - DATA MART	49
CAPÍTULO 4 - OLAP - ON-LINE ANALYTICAL PROCESSING	51
4.1 - CONCEITOS BÁSICOS SOBRE OLAP.....	51
4.2 - ROLAP - RELATIONAL ON-LINE ANALYTICAL PROCESSING	59
4.3 - MOLAP - MULTIDIMENSIONAL ON-LINE ANALYTICAL PROCESSING	60
4.4 - ROLAP OU MOLAP, QUAL A MELHOR TECNOLOGIA OLAP?.....	72
4.5 - HOLAP - HYBRID OLAP.....	76
4.6 - WOLAP OU WEB OLAP.....	76
4.7 - GERENCIAMENTO DA FERRAMENTA OLAP.....	77
CAPÍTULO 5 - DATA MINING	79
5.1 - CONCEITOS DE DATA MINING.....	79
5.2 - TIPOS DE MODELOS CONSTRUÍDOS PELO DATA MINING	85
5.3 - AS TÉCNICAS UTILIZADAS PELO DATA MINING	88
5.4 - PLANEJAMENTO E GERENCIAMENTO DO DATA MINING.....	92
5.5 - APLICAÇÕES	94
CAPÍTULO 6 - EXEMPLOS DE EMPRESAS QUE IMPLANTARAM AS FERRAMENTAS E OS RESULTADOS OBTIDOS.....	95
CAPÍTULO 7 - TENDÊNCIAS	102
CAPÍTULO 8 - CONCLUSÕES.....	106
Anexo A - Um exemplo de um aplicativo de data mining: SuperQuery.....	109
Anexo B - Um exemplo de um produto OLAP: Powerplay 6.0	117
REFERÊNCIAS BIBLIOGRÁFICAS	123

BIBLIOGRAFIA COMPLEMENTAR 137**Apêndice 1- Alguns Produtos de Data Warehouse, OLAP e Data Mining..... 1**

1 - PRODUTOS PARA DATA WAREHOUSE.....	1
1.1 - Produtos para a extração e transformação de dados	1
1.2 - Produtos para limpeza e refinamento de dados.....	2
1.3 - Produtos para replicação ou transferência de dados	2
1.4 - Produtos para armazenamento de dados	2
1.5 - Produtos para gerenciamento e administração.....	4
1.6 - Produtos para geração de Consultas e Relatórios	5
1.7 - Produtos para metadados	6
2 - PRODUTOS OLAP.....	6
3 - PRODUTOS PARA DATA MINING.....	9

GLOSSÁRIO

LISTA DE FIGURAS

Figura 1 - Diagrama de um Modelo Entidade-Relacionamento	24
Figura 2 - Diagrama de um Modelo Dimensional	25
Figura 3 - Diagrama de um Modelo Dimensional	26
Figura 4 - Gráfico demonstrando a satisfação de um grupo de usuários com a implantação do data warehouse.....	43
Figura 5 - Modelo relacional (tabelas relacionadas).....	61
Figura 6 - Modelo multidimensional (duas dimensões).....	62
Figura 7 - Modelo relacional.....	63
Figura 8 - Modelo multidimensional (tridimensional).....	64
Figura 9 - Modelo relacional.....	64
Figura 10 - Modelo multidimensional (bidimensional)	65
Figura 11 - Visão 1: modelo por cor por concessionária	66
Figura 12 - Visão 2: cor por modelo por concessionária	66
Figura 13 - Visão 3: concessionária por cor por modelo	66
Figura 14 - Visão 4: cor por concessionária por modelo	67
Figura 15 - Visão 5: concessionária por modelo por cor	67
Figura 16 - Visão 6: modelo por concessionária por cor	67
Figura 17 - Seleção de dados (data dicing)	69
Figura 18 - Operações “roll-up” e “drill-down”	70
Figura 19 - Diagrama de uma Rede Neural Artificial.....	90
Figura 20 - Diagrama de uma Árvore de Decisão	91
Figura 21: A Janela Principal da SuperQuery.....	110
Figura 22: Janela onde é feita a seleção das colunas dependentes	111

Figura 23: Ativando ou Desativando Filtros Automáticos.	111
Figura 24: A janela de Ajustes de Descoberta de Fatos.....	113
Figura 25: O produto informando as regras durante o processamento	115
Figura 26: Uma visão da tabela de fatos	116
Figura 27 - Tela de um relatório financeiro gerado pelo PowerPlay 6.0.....	119
Figura 28 - Tela exibindo um item em destaque e possibilitando uma análise mais detalhado no mesmo	119
Figura 29 - Tela exibindo relatório contendo análise personalizada e pré-definida	120
Figura 30 - Tela contendo um dos diversos tipos de relatórios possíveis.....	120
Figura 31 - Tela contendo uma das análises disponíveis	121
Figura 32 - Tela contendo um dos tipos disponíveis de visualização e navegação dos dados.....	121

LISTA DE TABELAS

Tabela 1 - Custo de armazenamento por Megabyte.....	14
Tabela 2 - Dados fictícios que passarão pelo processo de limpeza de dados	29
Tabela 3 - Dados fictícios após a elementarização	29
Tabela 4 - Dados fictícios após a padronização.....	30
Tabela 5 - Tabela comparativa entre data mart e data warehouse	50
Tabela 6 - Comparação entre os ambientes de OLTP, data warehouse e OLAP.....	58
Tabela 7 - Exemplo de tabela contendo dados.....	83

LISTA DE ABREVIATURAS E SIGLAS

API	- Application Program Interface
CASE	- Computer Aided Software Engineering
CD-ROM	- Compact Disk - Read Only Memory
CEP	- Código de Endereçamento Postal
CGI	- Computer Graphics Interface
CIO	- Chief Information Officer
COBOL	- Common Business-oriented Language
DBA	- Data Base Administrator
DBMS	- Data Base Management System
DLL	- Dynamic Link Library
DOLAP	- Desktop On-Line Analytical Processing
DSS	- Decision Support Systems
EIS	- Executive Information Systems
ER	- Entidade-Relacionamento (Modelo)
ERP	- Enterprise Resource Planning
EUA	- Estados Unidos da América
FBI	- Federal Bureau of Investigation
GB ou Gb	- Gigabytes
GDSS	- Group Decision Support Systems
GUI	- Graphical User Interface
HOLAP	- Hybrid On-Line Analytical Processing
HTML	- HyperText Markup Language
HTTP	- HyperText Transport Protocol
IBM	- International Business Machines Corporation

IS	- Information Systems
IT	- Information Technology
KDD	- Knowledge Discovery in Databases
LAN	- Local Area Network
LHS	- Left Hand Side
MB ou Mb	- Megabytes
MCDA	- MultiCriterial Decision Analysis
MD	- Modelo Dimensional
MDB	- Multidimensional DataBase
MDDB	- MultiDimensional DataBase
MDM	- MultiDimensional Model
MER	- Modelo Entidade-Relacionamento
MIPS	- Millions of Instructions Per Second
MIS	- Management Information Systems
MOLAP	- Multidimensional On-Line Analytical Processing
MPP	- Massive Parallel Processing
MRP	- Materials Requirement Planning
MTBF	- Mean Time Between Failures
NBA	- National Basketball Association
ODBC	- Open Data Base Connectivity
ODS	- Operational Data Store
OLAP	- On-Line Analytical Processing
OLE	- Object Linking and Embedding
OLTP	- On-Line Transaction Processing
OOP	- Object-Oriented Programming
PC	- Personal Computer
RAD	- Rapid Application Development
RAM	- Random Access Memory
RDBMS	- Relational Data Base Management Systems
RHS	- Right Hand Side
RISC	- Reduced Instruction Set Computing
ROLAP	- Relational On-Line Analytical Processing

RSI	- Retorno Sobre o Investimento
SAD	- Sistemas de Apoio à Decisão
SGDB	- Sistemas de Gerenciamento de Banco de Dados
SI	- Sistemas de Informação
SMP	- Symmetrical Multi-Processing
SQL	- Structured Query Language
TB ou Tb	- Terabytes
TCP/IP	- Transmission Control Protocol/Internet Protocol
TI	- Tecnologia da Informação
WAN	- World Area Network
WWW	- World Web Wide

RESUMO

BISPO, C. A. F. (1998). *Uma análise da nova geração de sistemas de apoio à decisão*. São Carlos, 1998. 160 p. Dissertação (Mestrado) - Escola de Engenharia de São Carlos, Universidade de São Paulo.

Nesta dissertação são feitas três abordagens. Na primeira apresentam-se os componentes necessários para que se possa compreender melhor o cenário atual onde se encontram aqueles que são os responsáveis pelo processo decisório nas empresas. São abordados as evoluções do processo decisório e do suporte ao mesmo, suas etapas e os seus fatores de influência. A segunda abordagem é relativa às três ferramentas que constituem a nova geração de Sistemas de Apoio à Decisão. A primeira ferramenta é o data warehouse, um banco de dados específico para propósitos gerenciais e que é independente dos bancos de dados operacionais. A segunda ferramenta é o On-Line Analytical Processing (OLAP) e é utilizada para se realizar análises sofisticadas, que permitem aos seus usuários compreenderem melhor os negócios que são realizados na empresa. A última ferramenta é o data mining que permite que se faça uma análise nos dados armazenados, durante anos, para que se descubram relacionamentos ocultos entres os dados, revelando perfis de compras e de clientes; desta forma, as informações obtidas podem se tornar estratégias de negócios. Com a abordagem destas três novas ferramentas, deseja-se analisar o que existe de mais avançado, atualmente, para dar um melhor suporte ao processo decisório, sem entrar nos detalhes estritamente técnicos destas tecnologias. A terceira abordagem é constituída de exemplos de empresas que implementaram

estas ferramentas e os resultados obtidos, assim como pelas tendências destas ferramentas para os próximos anos.

Palavras-chave: Sistemas de Apoio à Decisão; Data Warehouse; On-Line Analytical Processing; Data Mining.

ABSTRACT

BISPO, C. A. F. (1998). *An analysis of the new generation of decision support systems*. São Carlos, 1998. 160 p. Dissertação (Mestrado) - Escola de Engenharia de São Carlos, Universidade de São Paulo.

In this dissertation we will deal with three approaches. On the first we present the necessary elements to make one understand better the current scenery where the responsible persons for the decision process of companies meet. The evolution of the decision process and its support, phases and influence factors. The second approach is related to the three tools that constitute the new generation of Decision Support Systems. The first tool is the data warehouse, a specific database for the managerial purposes that is independent from the operational databases. The second tool is the On-Line Analytical Processing (OLAP) used in carrying out sophisticated analyses allowing its users a better understanding of the business accomplished in the company. The last tool is the data mining that allows for an analysis of the data stored along the years so that one is able to find out the correct relationship among the collect's data, revealing business and clients profiles. In such way all the information gathered in the process can be converted into business strategy. With the approach of these three new tools we intend to analyze the most advanced techniques available nowadays to give a better decision support without getting into strictly technical details of these technologies. The third approach is made up of examples of companies that implemented such tools and the attained results, as well , the trends for these tools in the coming years.

Keywords: Decision Support System; Data Warehouse; On-Line Analytical Processing; Data Mining.

Capítulo 1 - Introdução

Esta dissertação visa dar uma contribuição no sentido de analisar algumas das novas ferramentas computacionais que podem auxiliar os processos de tomada de decisões empresariais.

Quanto mais conhecimento a humanidade adquire, mais complexos vão se tornando os sistemas, as empresas e a sociedade. Os novos e modernos conceitos administrativos e os avanços tecnológicos que não param de evoluir, estão tornando o ambiente das empresas cada vez mais complexo. Os administradores de empresa passam a maior parte de seu tempo levantando e analisando informações que os levarão às decisões necessárias para o gerenciamento da empresa e dos negócios.

Vários fatores estão contribuindo para mudanças nas empresas, no modo de se tomar decisões estratégicas em seus negócios (DATAMIND TECHNOLOGY CENTER, 1998). A crescente competição e mercados mais dinâmicos estão conduzindo as empresas a tomarem decisões mais rapidamente. Estruturas organizacionais mais niveladas estão trazendo mais funcionários para o processo decisório. O desejo dos acionistas, em aumentar os lucros, está desafiando as empresas a encontrarem soluções mais criativas para cultivarem seus negócios. As empresas tiveram que se direcionar para um aumento da eficiência, marketing mais agressivo e uma estrutura organizacional mais dinâmica. O enfoque primário está em alcançar e manter mais efetivamente os seus clientes e minimizar os custos para se obter uma maior competitividade.

No processo decisório das empresas existem muitas variáveis envolvidas e ainda existem interesses econômicos e políticos, alguns explícitos e outros ocultos. É necessário proporcionar um bom suporte a este processo para se obter o sucesso nos negócios. Os administradores têm que tomar as suas decisões com base na análise de todas as alternativas possíveis e encontrar a alternativa mais viável. Torna-se imprescindível a utilização do suporte computacional para o levantamento e as análises nas informações necessárias. O método de escolha da melhor alternativa deve ser um método reconhecido e confiável. As decisões empresariais devem ser tomadas com base em dados e/ou fatos comprovados e deve-se ter o máximo de certeza dos resultados a serem alcançados com a adoção da alternativa que for escolhida como sendo a mais viável.

As dificuldades relativas ao processo decisório não se constituem um problema recente. No próximo capítulo, será visto que sempre existiu a necessidade de basear-se em “algo” que servia de referência, para que se tomassem as decisões que eram consideradas sábias. Serão vistos, também, alguns dos antigos fatores que influenciavam as decisões antigamente e os que as influenciam atualmente.

1.1 - Objetivos

O objetivo principal desta dissertação é analisar as três ferramentas que constituem a nova geração de Sistemas de Apoio à Decisão (o data warehouse, o OLAP e o data mining), quanto ao suporte proporcionado pelas mesmas aos processos decisórios praticados nas empresas.

Para alcançar-se este objetivo é necessário obter-se os seguintes objetivos secundários:

1. mostrar as origens e o contexto histórico que levou à criação das ferramentas;
2. analisar as ferramentas quanto:
 - ⇒ à forma como elas auxiliam o processo decisório;
 - ⇒ às tecnologias envolvidas;
 - ⇒ à elaboração dos projetos de implantação;

- ⇒ ao gerenciamento das ferramentas;
- ⇒ aos benefícios, vantagens e desvantagens de sua utilização;
- ⇒ exemplos de empresas que já as estão utilizando e os resultados obtidos;
- ⇒ as tendências para os próximos anos;

1.2 - Escopo

O tema abordado por esta dissertação é complexo e multidisciplinar, envolve diversas tecnologias que fazem parte do ambiente informatizado ao qual pertencem as ferramentas. Envolve, também, diversos conceitos relativos a estas tecnologias e, ainda, os diversos conceitos da Administração de Empresas e dos Negócios.

De acordo com os objetivos propostos, não é necessário um estudo profundo nas tecnologias envolvidas ou nos conceitos da Administração. Serão citadas as tecnologias, assim como alguns de seus conceitos até um nível em que se possa compreender a importância destas tecnologias ou destes conceitos para as ferramentas abordadas.

1.3 - A pesquisa bibliográfica

Foi realizado o levantamento da literatura disponível e descobriu-se que a maior parte desta literatura é constituída por artigos. Por ser um campo de pesquisa recente, existem poucos livros publicados sobre as ferramentas. Apenas neste ano (1998) é que estão sendo lançados diversos livros nos EUA e alguns no Brasil. Boa parte dos artigos foram publicados em jornais e revistas especializadas e estão disponíveis na Internet. Isto auxiliou bastante o acesso a estas publicações.

Na fase posterior, foram selecionados os tópicos que abordavam os assuntos de interesse para esta dissertação. Parte das publicações abordavam aspectos estritamente técnicos das ferramentas, que extrapolam o escopo deste trabalho.

Após a fase de seleção dos assuntos de interesse, passou-se a analisá-los, a confirmar ou confrontar conceitos de diversos autores e a escolher os assuntos dos autores mais reconhecidos ou cuja linha de raciocínio fosse mais coerente com este trabalho. O último passo foi utilizar os conceitos selecionados, para redigir esta dissertação.

1.4 - Estrutura

A dissertação está dividida em oito capítulos. No capítulo 1, estão a introdução ao tema, os objetivos, o escopo, a pesquisa bibliográfica e a estrutura da dissertação.

No capítulo 2, é feita uma revisão bibliográfica, onde são abordados a evolução, as etapas e os fatores de influência do processo decisório, os conceitos básicos, a evolução e as linhas de pesquisas dos Sistemas de Apoio à Decisão.

Nos capítulos 3, 4 e 5, aborda-se, respectivamente, cada uma das três ferramentas, quanto aos aspectos propostos nos objetivos desta dissertação. No capítulo 6, são apresentados alguns exemplos de empresas que implantaram as ferramentas e os resultados obtidos. No capítulo 7, citam-se algumas tendências das ferramentas para os próximos anos. No capítulo 8, são apresentadas as conclusões e são feitas algumas propostas para futuros trabalhos.

Nos Anexos A e B, apresenta-se, respectivamente, um exemplo de produto para data mining e para OLAP. No Apêndice 1, é apresentada uma relação de produtos para data warehouse, OLAP e data mining. No glossário, apresenta-se uma relação de endereços eletrônicos onde podem ser acessados os glossários on-line.

Capítulo 2 - Revisão Bibliográfica

Para que se possa compreender melhor as necessidades das novas ferramentas, no suporte ao processo decisório, são abordados neste capítulo alguns tópicos que são considerados importantes. São apresentados a evolução e os fatores de influência do processo decisório, os conceitos básicos, a evolução e as linhas de pesquisas dos Sistemas de Apoio à Decisão.

2.1 - A evolução do processo decisório

A evolução do processo decisório é um tema extenso e bastante abrangente. Baseando-se em BISPO & CAZARINI (1998c), será apresentado um resumo dos pontos mais importantes desta evolução.

2.1.1 - O processo decisório na Antigüidade

Desde o início da civilização, o homem sempre procurou algo que o auxiliasse no seu processo decisório. Ele buscava nas divindades esse auxílio para que pudesse tomar decisões consideradas sábias. As diversas culturas antigas buscavam em pessoas com “poderes místicos” o auxílio para as suas decisões. Considerava-se que essas pessoas podiam ter contato direto com as divindades e que as orientações que esses místicos forneciam eram divinas. Assim, as decisões tomadas eram consideradas sábias e livres de erros. Quando a decisão tomada não proporcionava o resultado esperado, atribuíam-se os erros às divindades que não estavam felizes. Desta forma, as decisões tomadas nesta época eram fortemente influenciadas pelas entidades divinas e seus representantes.

Mais tarde surgiram divindades mais populares, como Cristo, Maomé e Buda. Surgiram, também, líderes religiosos, nas comunidades, que passaram a exercer influência direta nas decisões das pessoas, por meio dos preceitos religiosos de suas divindades. E, a exemplo do passado, quando o resultado das decisões proporciona o sucesso, justificava-se como sendo fruto do amor pela divindade. Porém, quando o resultado era um fracasso, afirmava-se que era falta de fé dos decisores. Assim, as entidades divinas continuaram a ter grande influência nas decisões.

Atualmente, a religião é apenas um dos fatores de influência do processo decisório; existem outros fatores que, em conjunto, exercem influências neste processo. Esses outros fatores sempre existiram, porém, só mais recentemente eles ganharam maior importância. Esses fatores serão vistos em outro tópico.

2.1.2 - A evolução da decisão no século XX

Os critérios de tomada de decisão, no começo do século, eram centrados no executivo principal, em geral o proprietário, que detinha as prerrogativas de escolher o que julgasse melhor para a empresa e para os trabalhadores (PEREIRA & FONSECA, 1997). Partia-se do pressuposto que os trabalhadores eram seres passivos e despreparados, sem capacitação e treinamento para atuarem nas decisões; eram avaliados quantitativamente, pelo que rendiam, e descartados, quando não fossem mais produtivos. Como o ambiente era estável e as informações eram restritas, supunha-se que os decisores detivessem conhecimento bastante amplo de todas as alternativas e de suas conseqüências. Acreditava-se que os decisores escolhiam sabiamente a melhor alternativa para cada problema surgido.

A década de 60 foi precursora de grandes mudanças, em todas as áreas, e ficou conhecida como “a década que mudou o mundo”. Na teoria administrativa, surgiu o movimento conhecido como Escola de Relações Humanas, oriundo da contribuição da Psicologia Social. Este movimento deu um grande passo ao reconhecer o trabalhador como um ser humano capaz de pensar, de decidir e de ser

motivado (PEREIRA & FONSECA, 1997). Os trabalhadores deixaram de ser simples “mão-de-obra”, ou seja, aqueles que utilizam apenas as mãos para efetuarem seus trabalhos e passaram a ser chamados de recursos humanos, quer dizer, aqueles que trabalham de corpo e alma, que se esforçam, quando motivados, e possuem habilidades que podem ser desenvolvidas e aproveitadas em benefício da empresa. Porém, a contribuição da Psicologia Social nas empresas foi a captação e a manipulação das pessoas para manter as organizações funcionando produtivamente e aumentando cada vez mais a lucratividade da empresa.

2.2 - Conceitos básicos sobre decisão

Segundo PEREIRA & FOSECA (1997) a palavra *decisão* é formada pelo prefixo *de* (prefixo latino aqui com o significado de parar, extrair, interromper) que se antepõe à palavra *caedere* (que significa cindir, cortar). Tomada ao pé da letra, a palavra *decisão* significa “parar de cortar” ou “deixar fluir”. Uma decisão precisa ser tomada sempre que estamos diante de um problema que possui mais que uma alternativa para a sua solução.

2.2.1 - Os fatores de influência do processo decisório

Toda decisão tem conseqüências e envolve riscos. O processo decisório é sempre estressante para os decisores e, geralmente, após as tomadas de decisão importantes acontece um relaxamento geral nos organismos dos mesmos (PEREIRA & FONSECA, 1997). Segundo os autores, em algumas pessoas, o stress provocado pelo processo decisório é marcado por claras manifestações psicossomáticas, tais como dores de cabeça, perturbações gástricas, manifestações de euforia, depressão ou outros sintomas físicos.

Existem alguns fatores que exercem influência direta no processo decisório. Antes de existir o problema, já existem estes fatores e, quando o problema surge, o mesmo é inserido em um cenário onde estes fatores estão embutidos. Normalmente, estes fatores não fazem parte das variáveis que compõem o problema, ou mesmo das

alternativas propostas para a solução do mesmo, porém, estes fatores estão sempre presentes e exercem influência na solução do problema.

Alguns destes fatores exercem influência individual ou coletiva. São eles:

- ⇒ a inteligência e a cultura;
- ⇒ o nível social;
- ⇒ o sexo;
- ⇒ a religião;
- ⇒ os costumes e as crenças;
- ⇒ a ética moral e a ética profissional;
- ⇒ a saúde física e a mental;
- ⇒ a influência familiar;
- ⇒ o fator emocional, na hora exata da tomada da decisão.

Nas empresas existem os fatores de influência que lhes são inerentes:

- ⇒ a necessidade de produtos ou serviços com qualidade;
- ⇒ a necessidade de atendimento rápido e personalizado;
- ⇒ a necessidade de preços competitivos;
- ⇒ as condições impostas por clientes e/ou fornecedores;
- ⇒ as exigências dos governos e da sociedade;
- ⇒ a concorrência;
- ⇒ as notícias da mídia;
- ⇒ a cultura organizacional;
- ⇒ as tecnologias utilizadas;
- ⇒ o melhor emprego dos recursos existentes;
- ⇒ as normas existentes na empresa;
- ⇒ a legislação em vigor;
- ⇒ o meio ambiente.

Como pode ser notado, tomar decisões no contexto atual é algo complexo. Administrar esta complexidade é o desafio de todo dirigente moderno. Boas ferramentas para dar suporte ao moderno processo decisório são imprescindíveis.

2.2.2 - O processo decisório nas organizações

A função dos administradores da empresa é extremamente decisória (PEREIRA & FONSECA, 1997). Os processos administrativos são basicamente processos decisórios. As decisões gerenciais afetam diretamente a sobrevivência da empresa e a vida das pessoas que giram em torno dela, sejam elas empregados, acionistas, fornecedores ou clientes. Por isso, o decisor é exposto a inúmeros fatores de influência (já vistos anteriormente). Todas as pessoas afetadas cobram dele um resultado satisfatório. Muitas vezes, clientes, acionistas e empregados solicitam soluções diferentes e até antagônicas para solucionar um mesmo problema. O cliente solicita o quarteto: qualidade, preço, entrega e atendimento. Os acionistas exigem garantia de lucro e o retorno sobre o investimento, combinados com o alcance das metas e com a imagem positiva da empresa e de seus produtos. E os empregados pressionam por melhores condições no trabalho, incluindo melhores salários, garantia de emprego, assistência médica e odontológica, e outras reivindicações mais específicas.

É necessário estabelecer prioridades, quando existem objetivos e posições antagônicas, carências ou disputas de recursos e de informações; transformar os objetivos organizacionais em objetivos coletivos, por meio da participação e da visão do futuro que deve ser compartilhado com todos os membros da organização. E é necessário, também, buscar a satisfação dos clientes ou usuários, sem se descuidar dos interesses dos acionistas e das necessidades dos empregados. As decisões são atos de poder, no momento em que alocam recursos, definem estratégias, conduzem o destino de organizações e de pessoas, os gerentes assumem uma dimensão política muito semelhante à de um governo (PEREIRA & FONSECA, 1997). No cotidiano, a viabilização desse processo, que, na maioria das vezes, é cheio de conflitos de interesses, exige objetivos compartilhados, liderança, comunicação efetiva e habilidade de negociação constante.

A informação é algo que alguém deseja obter e está disposto a pagar por ela; não é tangível e nem mensurável, mas é um produto valioso no mundo contemporâneo, porque proporciona poder (GATES, 1997). A busca de informações é atualmente o alvo da maior parte dos governos e das empresas. É através da informação que é possível ter um melhor suporte ao processo decisório. É função das ferramentas, que vão dar o suporte a este processo, levantar as informações necessárias de uma forma rápida, confiável e exibi-las de uma forma facilmente compreensível.

2.3 - A evolução dos Sistemas de Apoio à Decisão

SPRAGUE & WATSON (1991) afirmam que, no início da década de 70, várias empresas e vários grupos de pesquisas começaram a pesquisar e desenvolver Sistemas de Apoio à Decisão, que passaram a ser caracterizados como sistemas computacionais interativos que auxiliavam no processo decisório de problemas considerados não estruturados. Porém, segundo os autores, na década seguinte, vários pesquisadores e desenvolvedores de sistemas, ampliaram a definição destes sistemas de modo a incluir qualquer sistema capaz de dar alguma contribuição ao processo decisório, desde que possuam as seguintes características:

- ⇒ serem voltados para problemas menos estruturados e menos especificados com os quais os gerentes deparam;
- ⇒ combinem o uso de modelos ou técnicas analíticas a funções tradicionais de acesso e recuperação de informações;
- ⇒ concentrem-se especificamente em recursos que facilitem seu uso para pessoal não especializado em computação e
- ⇒ enfatizem a flexibilidade e a adaptabilidade de acomodar mudanças no ambiente e na abordagem ao processo decisório.

2.3.1 - O surgimento do suporte computacional à decisão

Segundo POWER (1997), o conceito de suporte computacional à decisão surgiu com a evolução de duas áreas de pesquisa: os estudos teóricos sobre o

Processo de Tomada de Decisão Organizacional, feitos no Carnegie Institute of Technology, durante as décadas de 50 e 60, e os trabalhos realizados com Sistemas Computacionais Interativos no Massachusetts Institute of Technology, nos anos sessenta.

2.3.2 - Os Sistemas de Apoio à Decisão nas décadas de 60 e 70

Segundo PEARSON & SHIM (1995), COSTA (1997) e FISHER (1998), os primeiros Sistemas de Apoio à Decisão (SAD) surgiram nos anos 60 e 70, para dar suporte aos gerentes na solução de problemas gerenciais não estruturados. Estes SADs pioneiros eram muitos caros, de uso muito específico e difíceis de se operar; talvez por isso não tenha havido maiores investimentos neste setor. Nesta época, para as empresas, o mercado era estável, havia menos concorrência e existia a fidelidade dos clientes. O grande problema, naquele período, era que as bases de dados dos sistemas computacionais não possuíam a arquitetura necessária para a realização de pesquisas típicas de Sistemas de Apoio à Decisão e, devido à falta de dados históricos, existia muita dificuldade para a criação de relatórios e para se realizar as análises necessárias ao gerenciamento dos negócios. A ênfase dos SADs não estava no processo decisório, mas no suporte computacional para o desenvolvimento rápido das aplicações.

Neste período, os sistemas computacionais para o suporte à decisão eram desenvolvidos na própria empresa ou por terceiros. Esses sistemas eram desenvolvidos inicialmente para auxiliar a resolver problemas gerenciais específicos na empresa, depois, eram aperfeiçoados para englobar outros problemas gerenciais. Porém, nenhum destes sistemas conseguiu dar um bom suporte ao dinâmico processo de tomada de decisão, fornecendo as informações necessárias e no momento certo. Estes sistemas se tornavam imensos, complexos e não conseguiam acompanhar as mudanças das necessidades gerenciais. Tornavam-se difíceis de usar e de se fazer a sua manutenção. Além disso, requeriam um orçamento significativamente grande.

Um problema comum nas empresas ocorria quando os gerentes ou os analistas necessitavam realizar consultas *ad hoc* ou confeccionar relatórios que envolviam muitos dados. Era necessária a confecção dos mesmos, através alguma linguagem de programação, geralmente o COBOL; havia a necessidade de que o pedido entrasse na fila de espera para desenvolvimento de novos produtos e, quando chegava a vez de ser confeccionado, demorava-se ainda um bom tempo para sua realização.

2.3.3 - Os Sistemas de Apoio à Decisão na década de 80

Somente nos anos 80, quando surgiram os primeiros Sistemas de Gerenciamento de Banco de Dados (SGDB), é que se tornou possível um melhor acesso aos dados disponíveis, à sua formatação e à construção de consultas e relatórios de uma forma mais prática, rápida e barata. Porém, sempre que era necessária uma análise mais sofisticada nos dados, esta análise era realizada de forma “artesanal”, ou seja, os dados eram colecionados, depois formatados, conforme necessário, e só então, eram feitas as análises necessárias, muitas vezes fora de qualquer sistema computacional. Com a chegada ao mercado de novos softwares, como as planilhas eletrônicas e os programas de visualização gráfica dos dados, estas atividades se tornaram menos complexas. Mas, ainda faltava se desenvolver uma ferramenta que realmente auxiliasse os gerentes no processo de tomada de decisão, que facilitasse a visão e a análise dos negócios e que lhes fornecesse as informações necessárias de modo fácil, rápido e confiável.

Mesmo com os avanços obtidos, ainda existia o grande problema de que a modelagem dos dados era baseada na estrutura dos processos ao invés de na estrutura dos negócios; isso facilitava bastante o gerenciamento das atividades operacionais e dificultava o gerenciamento das atividades gerenciais. Era difícil se elaborar um relatório que contivesse uma análise, cruzando informações de sistemas diferentes, o que era constantemente necessário aos gerentes. Quando era necessário se obter este tipo de relatório, o mesmo continuava a ser confeccionado artesanalmente, o que se tornava lento, dispendioso e atrasava o processo decisório. Muitas vezes o setor de

desenvolvimento de sistemas se propunha a desenvolver sistemas que resolvessem tais problemas, porém o desenvolvimento destes sistemas levavam meses e até anos e, muitas vezes, quando ficavam prontos, já não atendiam mais às necessidades dos gerentes, pois mudanças na forma de se gerenciar os negócios tiveram que ser postas em prática devido às necessidades do mercado.

Neste mesmo período, começaram a surgir os primeiros sistemas, especialmente desenvolvidos para os gerentes, os chamados Sistemas de Informação para Executivos (Executive Information Systems - EIS). Porém, o gerenciamento das empresas e dos negócios estavam evoluindo mais rapidamente que estes sistemas.

2.3.4 - Os Sistemas de Apoio à Decisão na década de 90

Segundo THE (1998) e WELDON (1998), surgiram no início desta década, as ferramentas CASE e as Linguagens de Quarta Geração. Estas ferramentas prometiam resolver os problemas dos usuários finais que precisavam de informações rápidas e não tinham tempo a perder com o desenvolvimento de sistemas específicos, para atender as suas necessidades. Através das Linguagens de Quarta Geração, foi possível confeccionar relatórios e realizar consultas de uma maneira bem mais rápida e prática. E, através das ferramentas CASE, foi possível desenvolver sistemas de uma maneira mais rápida e mais simples. Porém, ambas as ferramentas não eram versáteis o suficiente para conseguir atender a todas as necessidades gerenciais. Eram fortemente baseadas na manipulação e navegação dos dados, mas eram pobres em análises sofisticadas. As planilhas eletrônicas tentaram amenizar a falta de ferramentas para a análise dos dados e ainda permitiam a geração de gráficos com extrema facilidade. Contudo, ainda não conseguiam atender a todas as necessidades dos gerentes que gastavam mais tempo manipulando dados e construindo modelos do que realizando as análises necessárias às suas atividades.

Algumas ferramentas, que surgiram mais recentemente, também fizeram suas promessas, a reusabilidade demonstrou ser um valioso recurso da OOP -

Programação Orientada ao Objeto - e da RAD - Desenvolvimento Rápido de Aplicação, todavia ainda falta flexibilidade a estas ferramentas (WELDON, 1998).

Com o passar do tempo, as empresas foram crescendo, os negócios foram aumentando, o que aumentou o volume de dados armazenados. Houve a necessidade de se aumentar o número de gerentes ou de se dividir as tarefas de gerenciamento em níveis gerenciais. Isso fez com que a necessidade de análise dos dados crescesse e ainda criou a necessidade de respostas mais rápidas, confiáveis e que melhor se adaptassem às necessidades do gerenciamento da empresa e dos negócios.

Um dos fatores que contribuíram para a evolução dos bancos de dados foi a queda crescente nos custos com armazenagem (RADEN & PETERSON, 1998). Um artigo no jornal COMPUTERWORLD (1998c) demonstra a evolução do custo de armazenamento por Megabyte.

Ano	Custo por Megabyte (US\$)
1988	11,540
1992	3,000
1995	0,330
1998	0,044
2001	0,006

Tabela 1 - Custo de armazenamento por Megabyte

Foram elaborados novos métodos de gestão empresarial, como o Gerenciamento pela Qualidade Total (CAMPOS, 1992) e a Reengenharia (HAMMER, 1994). Surgiu, também, a globalização devido à velocidade com que as informações chegam a qualquer parte do planeta e, somando-se com as exigências cada vez maiores dos clientes, fez com que a concorrência aumentasse espantosamente e em nível mundial. Combinando tudo isto, tem-se um cenário onde existe intensa e constante pressão sobre os gerentes, para conhecer melhor seus negócios e seus clientes. Em conseqüência, houve um aumento das necessidades gerenciais, novos tipos de consultas e de análises nos dados passaram a ser

necessárias e, portanto, novos sistemas que dessem um melhor suporte ao processo decisório precisavam ser elaborados.

Segundo FISHER (1998), quando as necessidades de mercado e progresso tecnológico convergem, eles realizam as principais mudanças na prática dos negócios. Nos últimos anos, a evolução das Tecnologias da Informação tornou possível, para muitas empresas, enfrentar o ambiente competitivo a que estão sujeitas. Em alguns casos, o efeito foi bastante dramático e resultou na transformação de modelos empresariais e até mesmo na criação de novos tipos de empreendimentos.

Nesta década, foram desenvolvidos diversos sistemas para dar suporte à solução dos problemas das empresas. Entre as novas ferramentas de Tecnologia da Informação, está a ferramenta para gestão integrada da empresa, o ERP (Enterprise Resource Planning) e a nova geração de Sistemas de Apoio à Decisão: o data warehouse, o OLAP e o data mining (BISPO & CAZARINI, 1998b). A ferramenta ERP é utilizada para o gerenciamento no ambiente operacional da empresa, enquanto as três novas ferramentas estão sendo utilizadas no ambiente gerencial. Estas ferramentas revelam-se muito úteis no gerenciamento moderno dos negócios e estão auxiliando os gerentes a enfrentar os desafios dos novos tempos.

Com o surgimento das ferramentas data warehouse e OLAP, as consultas e os relatórios passaram a ser confeccionados pelos próprios usuários, sem um conhecimento profundo de tecnologias computacionais. A confecção dos mesmos é rápida, barata, confiável e facilmente adaptável aos modelos dos negócios. O uso destas ferramentas faz com que os gerentes gastem um tempo bem menor manipulando dados e construindo os modelos necessários para suas análises, permitindo-lhes usar melhor seu tempo para as análises necessárias e a geração de soluções para seus problemas.

Outro avanço ocorrido nesta década é com relação às inovações tecnológicas. Até alguns anos atrás, existia uma defasagem de meses ou de anos de atraso até que

estas inovações chegassem no Brasil. Atualmente, quando surgem estas inovações, seu lançamento é simultâneo em diversos países, inclusive aqui.

2.4 - As linhas de pesquisas dos Sistemas de Apoio à Decisão

EOM (1996) fez uma análise de cocitação entre as literaturas publicadas (livros e artigos), entre os anos de 1971 a 1993, a respeito dos Sistemas de Apoio à Decisão. Uma das atividades desta análise foi a identificação das principais linhas de pesquisas deste campo, das quais segue abaixo um resumo sobre o que as mesmas representam.

- 1. Sistemas de Apoio à Decisão em Grupo (GDSS).** Pesquisa como os avançados recursos atuais de software e de hardware, incluindo comunicação à distância, podem auxiliar a realizar análises e elaborar estratégias coletivamente, mesmo que os participantes destas atividades estejam em diversos lugares, distantes geograficamente.
- 2. Fundamentos sobre DSS.** São as pesquisas sobre definições, conceitos, arquiteturas, taxonomias, desenvolvimento e evolução dos Sistemas de Apoio à Decisão.
- 3. Interface com o usuário.** Enfoca principalmente o estudo da evolução do modo pelo qual os dados são introduzidos nos sistemas e como são visualizados os resultados fornecidos. Esta linha de pesquisa é uma das que proporcionou uma melhor aceitação dos sistemas computacionais no ambiente das empresas.
- 4. Modelos de gerenciamento.** Estuda a estrutura, a representação, o processamento, a integração, a aplicação e a interpretação dos modelos lógicos. Através da criação de um modelo abstrato, que representa algo no mundo real, é possível a elaboração de sistemas computacionais que auxiliarão os usuários a gerenciar melhor aquilo que foi modelado.

5. **Análise Multicritério de Apoio à Decisão (MCDA).** Trabalha com a análise de problemas semi-estruturados ou não estruturados, que envolvem múltiplos critérios de avaliação, múltiplos objetivos ou ambos.
6. **Psicologia Cognitiva / Ciência Cognitiva.** São campos interdisciplinares que estudam os diversos componentes da capacidade de aprendizado humano e como eles interferem nos processos decisórios. O foco principal é como os adultos absorvem o conhecimento adquirido, a variação da aprendizagem entre diferentes indivíduos e entre os grupos diferentes.
7. **Inteligência Artificial.** Estuda como esta tecnologia pode auxiliar no desenvolvimento de sistemas computacionais baseados no conhecimento e como estes sistemas podem auxiliar os processos decisórios.
8. **Ciência Organizacional.** Estuda a estrutura organizacional das empresas, para que a sua compreensão possa auxiliar na criação dos modelos abstratos que levarão ao desenvolvimento dos sistemas computacionais.
9. **Ciência dos Sistemas.** Estuda o estado de complexidade dos sistemas, seus componentes, seus inter-relacionamentos e os relacionamentos com outros sistemas dentro das empresas. Através da melhor compreensão destes sistemas, é possível a construção de sistemas computacionais que auxiliem a gerenciar melhor o funcionamento dos mesmos.
10. **Sistemas Multicritério de Apoio à Decisão / Sistemas de Apoio à Negociação.** Ambos os sistemas são construídos com base nos modelos da Análise Multicritério de Apoio à Decisão. Os referidos modelos já atingiram um estágio de desenvolvimento muito avançado, porém, somente recentemente é que o estágio de evolução da Tecnologia da Informação permitiu o desenvolvimento de sistemas com base naqueles modelos.

11. **Implementação.** Este último campo de pesquisa é o estágio final de todos os outros e onde os mesmos se inter-relacionam. Portanto, este campo estuda o relacionamento e o nível de harmonização entre os campos anteriores.

O fato de surgir uma nova geração de Sistemas de Apoio à Decisão não descarta e não substitui os sistemas antigos e tradicionais. Muitas vezes, os novos sistemas trabalham em conjunto com os antigos, para a solução dos problemas, para o gerenciamento dos negócios e para a elaboração de novas estratégias. Por exemplo, as informações obtidas através do OLAP ou do data mining podem alimentar um Sistema Multicritério de Apoio à Decisão ou qualquer outro sistema que trabalhe na linha de pesquisa operacional ou otimização. Os vários sistemas, que foram desenvolvidos para realizar um tarefa específica, continuarão a ter o seu lugar garantido e até mesmo continuarão a ser confeccionados e aperfeiçoados, principalmente os sistemas utilizados para otimização, como a maximização do uso dos recursos disponíveis (recursos humanos, matéria-prima, layout de máquinas, colheitadeiras, etc.).

Após uma melhor compreensão da evolução e dos fatores de influência do processo decisório, dos conceitos básicos, da evolução e das linhas de pesquisas dos Sistemas de Apoio à Decisão, é possível uma melhor assimilação em relação ao que será abordado sobre as três novas ferramentas de suporte à decisão. O próximo capítulo abordará a primeira destas ferramentas: o data warehouse.

Capítulo 3 - Data Warehouse

O capítulo anterior apresentou uma revisão bibliográfica, proporcionando uma melhor compreensão sobre o processo decisório, seu suporte e sobre os Sistemas de Apoio à Decisão.

3.1 - Conceitos básicos

Como se trata de uma nova ferramenta que, embora já esteja em uso em diversas empresas, está em fase de aperfeiçoamento, o data warehouse ainda não possui uma definição universalmente aceita. Nele estão envolvidos conceitos tecnológicos da área de banco de dados, redes de computadores e processamento distribuído, como também diversos conceitos de Administração de Empresas. Alguns desenvolvedores de sistemas auxiliam esta confusão. Uma definição simples poderia ser: “um data warehouse é um grande banco de dados, elaborado com a finalidade de dar suporte ao processo decisório, onde os dados que o povoarão são obtidos através dos bancos de dados dos aplicativos operacionais da empresa.”

Willian H. Inmon, um dos precursores da ferramenta, a define da seguinte forma: “*Um data warehouse é um conjunto de dados baseado em assuntos, integrado, não volátil, e variável em relação ao tempo, de apoio às decisões gerenciais.*” (INMON, 1997, p. 33). Segundo o autor, a ferramenta é:

⇒ orientada aos principais assuntos ou negócios da empresa como clientes, vendas, produtos, apólices, tratamentos, seguros, viagens, etc., enquanto os sistemas de informações tradicionais são orientados a processos como estoques, entradas e saídas de materiais, compras e vendas, faturamento, contabilidade, etc.;

- ⇒ integrada, ou seja, trabalha de forma a padronizar os termos e as estruturas técnicas que são utilizados nos sistemas de informações tradicionais, por exemplo, nestes diversos sistemas o sexo pode ser armazenados como: “m” ou “f”, “O” ou “1”, “x” ou “y”, “macho” ou “fêmea”, “homem” ou “mulher”, “dama” ou “cavalheiro” e outras formas; no data warehouse apenas uma destas formas poderá aparecer padronizando esta referência;
- ⇒ não volátil, ou seja, no ambiente operacional, os dados sofrem as alterações necessárias como: incluir, alterar ou excluir dados; porém, no data warehouse os dados permitem apenas duas atividades: a sua carga para o banco de dados e as consultas; os dados nunca podem ser alterados;
- ⇒ variável no tempo, ou seja, a estrutura dos dados do data warehouse sempre contém algum elemento de tempo, enquanto nos sistemas de informações tradicionais isso não ocorre obrigatoriamente; nestes sistemas, o horizonte de tempo é normalmente de 2 a 3 meses, enquanto no data warehouse este horizonte é de 5 a 10 anos.

O data warehouse, que também é chamado de armazém de dados, tem como objetivo básico satisfazer as necessidades dos usuários (normalmente executivos, gerentes e analistas de negócios) quanto ao armazenamento dos dados que servirão para se realizar as consultas e análises necessárias para o gerenciamento dos negócios.

Um data warehouse é um sistema complexo que integra muitos componentes: diversos tipos de software e de hardware, redes de computadores, sistemas de comunicações de dados, servidores, mainframes e sistemas de administração de banco de dados, como também muitas pessoas de diferentes unidades organizacionais, com objetivos diferentes (LAMBERT, 1997). As principais ferramentas que compõem o data warehouse serão abordadas no próximo tópico. Porém, para os usuários, não importam os conceitos ou as definições da ferramenta, e sim se ela consegue proporcionar o suporte necessário aos seus processos decisórios.

O conceito de data warehouse baseia-se na aplicação de antigas idéias que somente agora puderam ser viabilizadas pela conjunção de diferentes tecnologias (GONÇALVES, 1997).

Um dos benefícios, proporcionados pelo data warehouse, é a diminuição do tempo que os gerentes levam para obter as informações necessárias aos seus processos decisórios, com a eliminação de tarefas operacionais, como pesquisa e identificação dos dados necessários. A sua grande vantagem é permitir a tomada de decisão baseada em fatos (TAURION, 1998a). Esta ferramenta deve tornar-se imprescindível para os gerentes conseguirem administrar seus negócios nos próximos anos. Como o ambiente de negócios está se tornando cada vez mais dinâmico, é extremamente necessário que as regras de negócios sejam incorporados às aplicações, que as estruturas dos sistemas computacionais se ajustem aos negócios e que o tempo de resposta destes sistemas seja cada vez menor.

Ao reunir informações dispersas por diversos bancos de dados e plataformas distintas, o data warehouse permite que sejam feitas análises bastante eficazes, transformando dados esparsos em informações estratégicas, antes inacessíveis ou subaproveitadas (TAURION, 1997).

O data warehousing, um termo bastante empregado pelos usuários, é o processamento dos dados dos bancos de dados fontes até o armazenamento final no data warehouse, assim como o processamento dos dados durante as consultas e análises.

3.2 - As principais ferramentas utilizadas em um data warehouse

São apresentadas as principais ferramentas que são utilizadas em um data warehouse. Os produtos, para estas ferramentas, são apresentados no Apêndice 1.

1. *Ferramenta para armazenamento*: são os bancos de dados, considerados o coração do data warehouse e parte imprescindível do projeto.

2. *Ferramenta para a extração de dados*: busca, nas bases de dados operacionais, os dados que vão ser armazenados no data warehouse.
3. *Ferramenta para a transformação de dados*: ajusta os dados para o formato do data warehouse. Este formato auxilia as futuras pesquisas.
4. *Ferramenta para o refinamento ou limpeza de dados*: faz os ajustes necessários nos dados, fazendo correções, desmembramento e fusões de dados, quando necessário, visando melhorá-los para facilitar as futuras pesquisas.
5. *Repositórios de metadados*: são bancos de dados onde são armazenados os metadados. Este assunto será abordado no tópico 3-10.
6. *Ferramenta para transferência de dados e replicação*: pode ser considerada um subconjunto da ferramenta de extração. Não faz nenhum tipo de processamento ou transformação, apenas transfere um dado de um lugar “A” para “B”. Geralmente, é utilizada para facilitar e dar uma resposta mais rápida às consultas ou análises, movendo os dados para um lugar apropriado e fazendo o que for necessário para agilizar o serviço solicitado.
7. *Ferramenta para gerenciamento e administração*: o gerenciamento é o monitoramento dos bancos de dados, por exemplo, quanto ao desempenho, integridade e segurança de dados; enquanto a administração é o monitoramento do suporte ao sistema, tais como, os recursos humanos e os esquemas manutenção preventiva e corretiva dos equipamentos.
8. *Ferramentas para gerenciamento de consultas*: fazem consultas e/ou geram relatórios, extraindo os dados do data warehouse, resumindo-os e apresentando-os em um formato apropriado.

9. *Ferramentas para gerenciamento de relatórios*: são semelhantes às ferramentas do item anterior, porém elas estão voltadas à geração de relatórios mais complexos contendo, por exemplo, dados sintéticos e analíticos em conjunto, gráficos, e outros tipos de visualização dos dados.

3.3 - A modelagem dos dados

Segundo WELDON (1998), um modelo é uma representação abstrata de um objeto real ou de um ambiente. A modelagem de dados é a prática de elaborar um banco de dados usando modelos de dados já consagrados. O processo funciona assim: primeiro, desenvolve-se um modelo conceitual de alto-nível do processo empresarial ou da atividade que se deseja modelar. Depois, usa-se este modelo para derivar um modelo lógico, no qual os dados são abordados com mais detalhes. Finalmente, a partir do modelo lógico elabora-se o modelo físico que provê todos os detalhes da implementação do banco de dados. De uma perspectiva lógica, faz-se a modelagem de dados para se aumentar a compreensão dos problemas empresariais e identificar os componentes básicos nos quais a solução será construída.

O modelo Entidade-Relacionamento (ER ou MER) divide os dados em diversas tabelas, que se relacionam entre si, formando um complexo diagrama, como mostrado na Figura 1. Este tipo de diagrama é muito difícil de se interpretar e analisar, inclusive pelos próprios projetistas, e não retrata a realidade dos negócios. Sua estrutura é importante para a eficiência e o desempenho no ambiente operacional, onde aplicativos específicos trabalham com esta estrutura e, normalmente, não são necessárias consultas que extrapolem o âmbito desses aplicativos. Quando consultas extras são necessárias, fazem-se atualizações nos referidos aplicativos, para que incorporem as novas consultas, porém, isto pode demandar até meses de trabalho e nem sempre o usuário dispõe de todo este tempo.

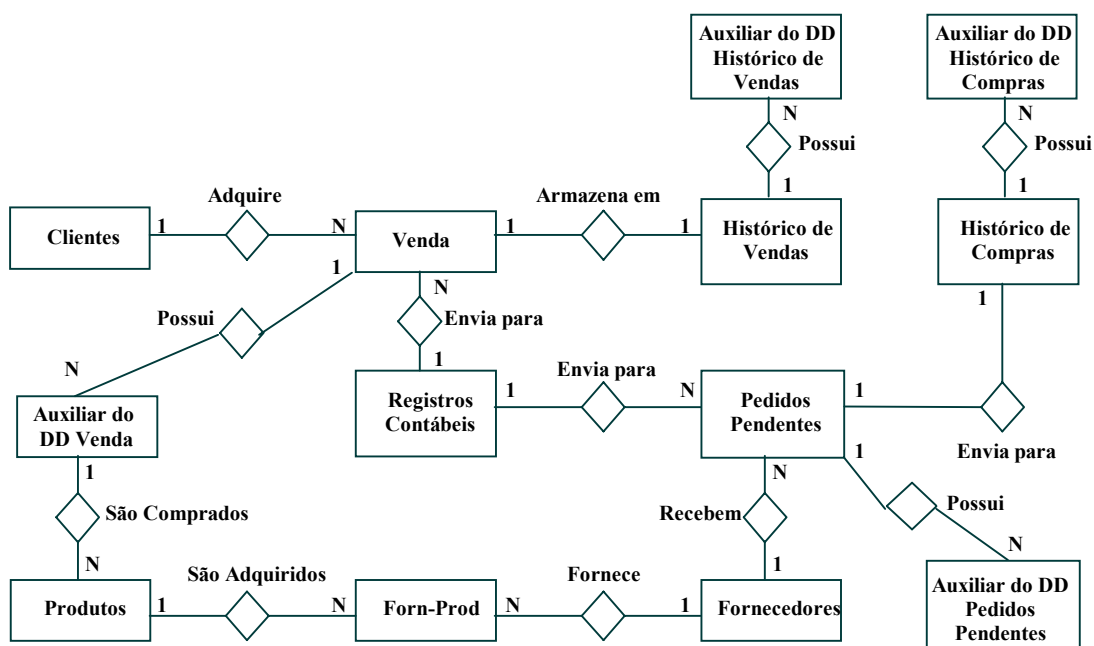


Figura 1 - Diagrama de um Modelo Entidade-Relacionamento

Para KIMBALL (1997b, 1998a e 1998b), o modelo Entidade-Relacionamento não é o mais adequado para se analisar os dados no ambiente gerencial, o modelo dimensional (MD) é o mais apropriado para este ambiente. Segundo o autor, este modelo também é conhecido por “star join scheme” ou simplesmente esquema estrela. O modelo é mais fácil para se consultar e analisar os dados, produz um banco de dados com menos tabelas e menos índices, apresenta os dados em um padrão, possui uma estrutura mais intuitiva e permite o acesso aos dados com alto desempenho (Figuras 2 e 3).

O modelo dimensional é assimétrico, ou seja, possui uma grande tabela, que é a principal, está localizada no centro do diagrama e possui outras tabelas secundárias ao seu redor, que são menores e que se relacionam com a tabela principal. A tabela central é chamada de tabela de fatos e as demais são chamadas tabelas de dimensão. A tabela de fatos armazena as medidas numéricas do negócio, por exemplo: unidades vendidas em cada transação, unidades produzidas em cada lote, etc. Esta tabela é composta por uma chave composta. As tabelas de dimensão armazenam as

descrições textuais das dimensões do negócio, por exemplo, a dimensão produto possui os dados sobre a marca, a categoria, o tipo de embalagem, o tamanho do produto, etc. As dimensões cliente e fornecedor possuem os seguintes dados: nome, endereço completo e outros dados específicos que auxiliam as análises do negócio. Cada tabela de dimensão tem uma única chave primária, que corresponde aos componentes da chave composta da tabela de fatos. O data warehouse possui várias tabelas de fatos, cada uma representando um assunto ou negócio diferente dentro da organização.

A elaboração de um modelo dimensional é um processo “top-down”: primeiro identificam-se os processos empresariais que representem assuntos ou negócios da empresa e servem como fontes das tabelas de fatos, depois povoam-se estas tabelas (KIMBALL, 1998b). Descreve-se cada assunto ou negócio por tantas dimensões quantas sejam necessárias. É importante que as tabelas de dimensão permaneçam em um único nível, sem serem normalizadas, senão o modelo torna-se mais difícil de ser compreendido e utilizado pelo usuário (KIMBALL, 1998i).

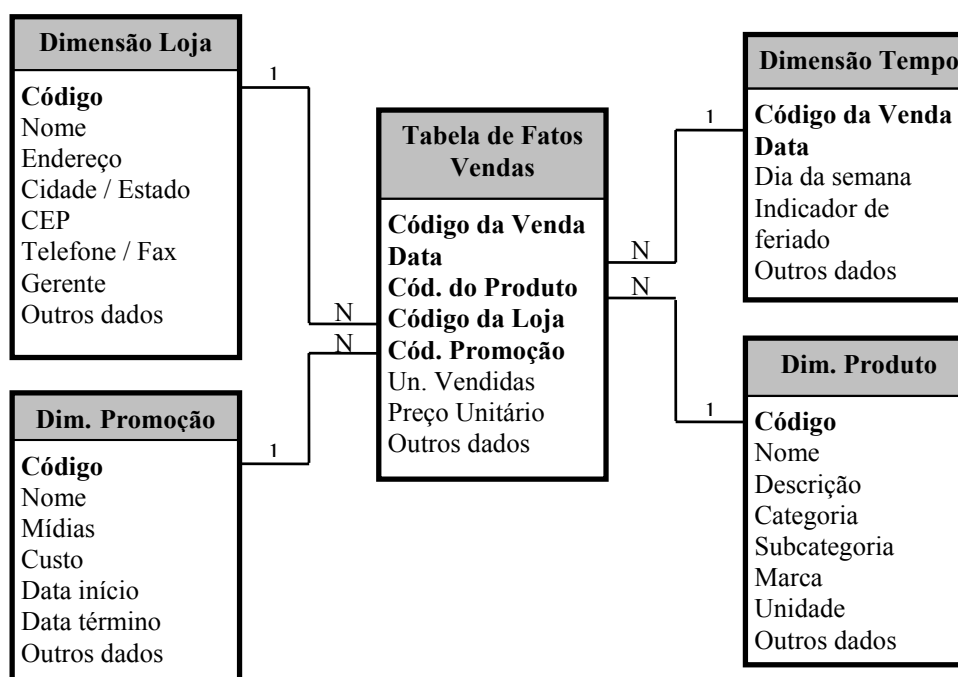


Figura 2 - Diagrama de um Modelo Dimensional

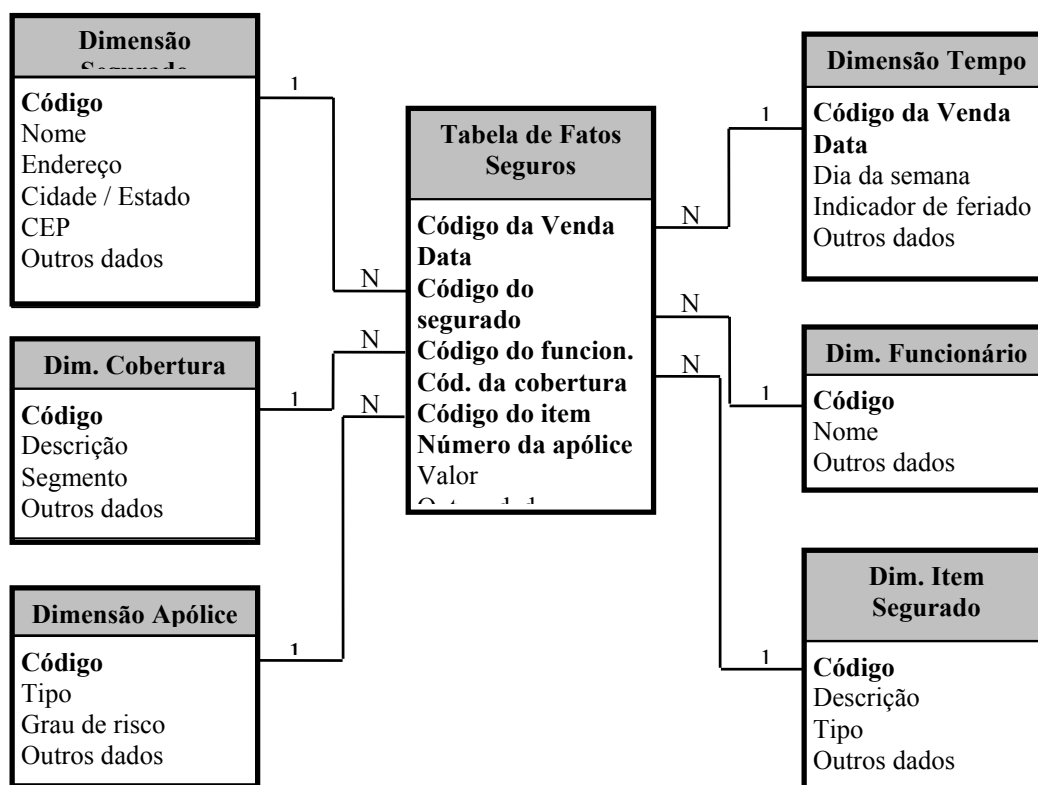


Figura 3 - Diagrama de um Modelo Dimensional

O modelo dimensional usa fatos, dimensões, hierarquias e esparsividade; é muito mais simples e elegante se comparado ao MER, e expressa o modo natural de os usuários raciocinarem (RADEN, 1997b). Este modelo pode ser implementado em um banco de dados relacional, em um multidimensional ou até mesmo em um banco de dados orientado a objeto. Para construir o modelo dimensional, são feitas seis perguntas básicas:

1. Qual negócio será modelado?
2. Quais são as medidas ou os fatos?
3. Qual o nível de granularidade?
4. Quais são as dimensões?
5. Quais são os atributos das dimensões?
6. Os atributos são estáveis ou variáveis com o passar do tempo, e qual a cardinalidade?

3.4 - A qualidade dos dados

A qualidade dos dados é o estado de perfeição, validade, consistência e precisão que os dados apresentam durante a sua utilização (HUFFORD, 1998). Como os dados armazenados servirão para dar suporte aos gerentes nos processos decisórios, estes dados devem ser totalmente confiáveis. Diversas são as situações onde a má qualidade dos dados podem comprometer sua análise e levar a uma solução que não seria necessariamente a mais adequada para resolver determinado problema.

Dois exemplos apresentados por KIMBALL (1998c) ilustram melhor o problema da má qualidade dos dados. Primeiro exemplo: uma pesquisa para detectar os clientes que adquiriram, diversas vezes, o mesmo produto. Porém, não é possível uma realização eficaz da pesquisa, devido ao fato de que, em cada venda realizada, os dados, tanto do cliente como do produto, foram armazenados de formas diferentes. Segundo exemplo: quando é necessário se fazer contatos contínuos com os clientes. A empresa pode enviar múltiplas correspondências absolutamente iguais ou dar múltiplos telefonemas ao mesmo cliente, devido à duplicidade de seus dados no banco de dados. Ou, ainda, quando é oferecido um produto a um cliente que já o adquiriu, recentemente, da própria empresa. Pode até haver dados errados, sobre os clientes, armazenados no banco de dados, e, quando se for fazer os contatos necessários, pode-se passar por situações constrangedoras. Isso destrói a credibilidade de qualquer empresa.

Resolver o problema de qualidade de dados não é uma tarefa fácil. ENGLISH (1998) e HUFFORD (1998) sugerem que sejam usados os mesmos princípios desenvolvidos e utilizados nos programas de Gerenciamento pela Qualidade Total, desenvolvidos pelos diversos mestres da área como ISHIKAWA (1984), FEIGENBAUN (1991), GARVIN (1992), CLAUSING (1993), JURAN & GRZYNA (1993), SUZAKI (1993) e GOESTCHI & DAVIS (1994). O tema Gerenciamento pela Qualidade Total foi abordado em BISPO et al. (1995 e 1997). Considerando que os dados são um produto de um processo empresarial, pode-se aplicar os mesmos

princípios para se melhorar sua qualidade. Não serão abordados os detalhes desse gerenciamento, por extrapolar o escopo desta dissertação.

Como o data warehouse não produz dados, ou seja, os dados armazenados nele são oriundos de outras fontes de dados, para melhorar sua qualidade, há a necessidade de se melhorar os processos empresariais que produzem esses dados.

LAMBERT (1997) propõe a identificação de pessoas que serão as responsáveis pela qualidade dos dados nos processos departamentais onde os dados se originam e que alimentarão o data warehouse. Propõe, ainda, o estabelecimento de diretrizes para administrar os recursos de dados incorporados. Estas diretrizes incluem padrões para se definir os elementos, a modelagem e as políticas de compartilhamento dos dados com os diferentes grupos dentro e fora da organização. Segundo o autor, é muito importante distinguir administração de dados de administração de banco de dados. Administradores de dados são orientados aos negócios, enfocam o significado e uso dos dados. Administradores de banco de dados (DBA) são orientados à tecnologia e se preocupam com a confiabilidade, integridade e desempenho de aplicações de banco de dados. Um DBA corrige erros técnicos do banco de dados, já o administrador de dados lida com a qualidade dos dados.

ENGLISH (1998), HUFFORD (1998) e KIMBALL (1998c) recomendam que para melhorar a qualidade dos dados é necessário se realizar um processo chamado limpeza dos dados. Este processo extrai os dados dos bancos de dados fontes, transforma-os até deixá-los em um estado de boa qualidade e carrega-os no armazém de dados. Segundo os autores, os esforços consideráveis e o custo para se limpar os dados não adicionam valor aos mesmos, porém aumentam a sua usabilidade e confiabilidade. A melhoria da qualidade dos dados busca minimizar custos e eliminar problemas com a perda de oportunidades causadas por dados ruins. Alcançar a qualidade significa satisfazer constantemente as expectativas dos clientes. As exigências dos clientes podem ser especificações escritas formais ou expectativas mentais informais. Em um data warehouse, os clientes são os usuários, suas

necessidades são a geração de informações e as especificações são: usabilidade, confiabilidade, interface amigável e alto desempenho na recuperação dos dados.

KIMBALL (1998c) recomenda desmembrar a limpeza dos dados em seis passos: elementarização, padronização, verificação da consistência, emparelhamento, verificação doméstica e documentação. Para exemplificar, são apresentados alguns dados fictícios (Tabela 2), que representam um único registro de um banco de dados.

Nome	Eng. José M. Silva
Endereço	R. Sto. Antonio, 201 - Vila Brás
Cidade/Estado	São José - MG
CEP	13630-000
Telefone	019-565-7000

Tabela 2 - Dados fictícios que passarão pelo processo de limpeza de dados

O primeiro passo para realizar a limpeza dos dados é a **elementarização**, ou seja, divisão dos dados até atingirem a forma mais elementar possível (Tabela 3). Este passo alterará não somente a estrutura de dados deste registro, mas de todo o banco de dados.

Nome	José M. Silva
Profissão	Eng.
Endereço	R. Sto. Antônio, 201
Complemento	
Bairro	Vila Brás
Cidade	São José
Estado	MG
CEP	13630-000
DDD	019
Telefone	565-7000

Tabela 3 - Dados fictícios após a elementarização

O segundo passo é a **padronização** dos elementos. No campo endereço, após confirmação, muda-se o “R.” para “Rua” e “Sto.” para “Santo”. No campo profissão, após uma verificação, muda-se “Eng.” para “Engenheiro de Produção” e no campo

nome, muda-se o “M.” para “Maria da”. Após as mudanças, os dados são mostrados na Tabela 4.

Nome	José Maria da Silva
Profissão	Engenheiro de Produção
Endereço	Rua Santo Antônio, 201
Complemento	
Bairro	Vila Brás
Cidade	São José
Estado	MG
CEP	13630-000
DDD	019
Telefone	565-7000

Tabela 4 - Dados fictícios após a padronização

O terceiro passo é a **verificação de consistência** dos elementos padronizados, ou seja, verificar se há algum erro no seu conteúdo. Há um engano no endereço, o campo cidade apresenta o dado “São José” e o campo CEP apresenta “13630-000”. A verificação de consistência constata que o referido CEP não pertence àquela cidade, significando que um dos dois campos não está correto e há a necessidade de uma checagem extra nestes dados.

O **emparelhamento** consiste em verificar se existem outros registros com os dados de José Maria Silva. Se houver, é necessário assegurar-se de que todos os dados são idênticos. A confirmação significa que está havendo duplicidade de dados e é necessário eliminá-la. Se os dados não forem totalmente idênticos, é necessário uma checagem minuciosa neles.

A **verificação doméstica (householding)** consiste em procurar se existem, em outros registros, pessoas que possuam exatamente o mesmo endereço. Se houver, pode-se verificar se estas pessoas possuem o mesmo sobrenome, o que caracterizaria que pertencem a uma mesma família. Se houver como confirmar esta informação, a mesma deve passar a integrar os dados do banco de dados, pois pode ser útil futuramente.

O sexto passo consiste em **documentar** os resultados dos passos anteriores e armazená-los no banco de dados para metadados.

3.5 - Segurança de dados

KIMBALL (1998e) faz um alerta para os perigos da falta de segurança de dados em um data warehouse. O autor lembra que a meta de todo data warehouse é tornar os dados corporativos disponíveis, facilitar-lhes o acesso e torná-los compreensíveis. A maioria das corporações investiram em centenas ou milhares de computadores, para transformar esta meta em uma realidade. Muitas redes de computadores são tão grandes e complicadas que se torna difícil controlá-las. Por outro lado, existem os hackers, os crackers e os spooks, que vêm violando sistemas computacionais em rede no mundo inteiro. O autor sugere que, além de um projeto de segurança para o ambiente de rede, também deve haver um projeto específico para o data warehouse. Neste projeto devem constar:

- ⇒ **regras de segurança:** quais usuários podem ter acesso e/ou podem fazer a manutenção de quais tipos de dados;
- ⇒ **tecnologia de segurança:** como o usuário é identificado, como sua autorização é liberada pelo servidor, qual o caminho de acesso aos dados quanto à proteção contra invasões, os esquemas de interrupção dos sistemas de segurança e o controle do dispositivos de acesso remotos à rede.
- ⇒ **administração da segurança:** quem controla as regras de segurança, como são administradas as senhas e como os esquemas de segurança são comunicados aos técnicos e usuários.

FONTES (1998) afirma que no projeto do data warehouse, devem estar previstas as normas, as políticas, os termos de compromisso e as auditorias que serão adotados com relação à segurança de dados. No projeto, devem constar as normas a serem seguidas e os termos de compromisso que serão assumidos por todos aqueles que tiverem algum tipo de contato com os dados do sistema. O principal documento gerado pela política de segurança de dados deve ser assinado pelo presidente da empresa, para demonstrar a todos que a informação é um recurso importante e que

existem responsabilidades no seu uso. Devem estar previstas, também, auditorias esporádicas que se farão no sistema para se verificar se as regras adotadas estão sendo cumpridas.

A política de segurança de dados deve definir como será o esquema de segurança de acesso aos dados pelos usuários e pelo pessoal de suporte técnico. Deverá definir os níveis hierárquicos de acesso aos dados e como controlar cuidadosamente do esquema de distribuição, além da troca periódica de senhas de acesso. Outra definição que deverá estar bem clara, na política de segurança de dados, deve ser o esquema de comunicação interna e externa da empresa. As comunicações devem ser bem monitoradas; deve-se ter a garantia de que os dados estão fluindo com integridade e os esquemas de segurança não estão sendo violados. Deve haver a proteção contra os “assaltos eletrônicos”, realizados pelos hackers e crackers. Estas ações podem ser feitas a distância e dificilmente deixam rastro.

Outro fator que pode influenciar a segurança de dados é o ambiente físico onde os equipamentos serão instalados. Deve-se ter o cuidado de respeitar as especificações técnicas dos fornecedores dos equipamentos. Não é nada agradável se fazer um alto investimento em tecnologia e, na hora de sua utilização, ocorrerem surpresas desagradáveis, devido a problemas com a sua instalação física.

A revista LANTIMES (1998) mostra os resultados de uma pesquisa feita em 1997, nos EUA, pelo Computer Security Institute em conjunto com o FBI, com 500 corporações americanas entre instituições financeiras, órgãos do governo e universidades. A pesquisa relaciona os principais tipos de violações eletrônicas e seus percentuais:

- ⇒ 72% - prejuízos com a quebra da segurança da rede;
- ⇒ 44% - acesso não autorizado de funcionários;
- ⇒ 42% - perdas contabilizadas;
- ⇒ 25% - ataques por negligência com serviços;
- ⇒ 24% - invasão por meio de sistemas externos;
- ⇒ 18% - roubo de informações confidenciais;

- ⇒ 15% - fraudes financeiras;
- ⇒ 14% - sabotagem de dados ou redes.

A pesquisa anterior mostra ainda que, em 1995, o Departamento de Defesa dos Estados Unidos sofreu 250.000 tentativas de ataques de hackers, dentre os quais 64% conseguiram violar os esquemas de segurança e somente 4% foram identificados.

Outro problema, relativo à segurança de dados e que não deve ser esquecido, são os vírus de computador. Com um único exemplo, é possível demonstrar os prejuízos que isso pode acarretar. Uma das maiores instituições financeiras brasileiras teve que paralisar, momentaneamente, as suas atividades no dia 7 de julho de 1998, devido ao ataque do vírus Knight 7, que não foi detectado pelos seus sistemas de proteção antivírus, pois era uma nova versão do referido vírus a qual os antivírus ainda desconheciam (LANTIMES, 1998). Houve prejuízos financeiros, devido à paralisação e ainda houve prejuízo para a imagem da empresa por causa do fato ocorrido.

3.6 - Algumas análises necessárias aos negócios

A globalização trouxe uma árdua concorrência a nível mundial. As empresas, até mesmo para poder sobreviver nos novos tempos, estão tendo que adotar posições mais agressivas no gerenciamento dos negócios. Hoje, já não se pode mais afirmar que os clientes são fiéis e que nunca deixarão de consumir produtos ou serviços de determinada empresa. E, para enfrentar essa nova realidade, as empresas estão adotando programas de reestruturação organizacional da própria empresa e dos negócios, sempre tendo como foco central os clientes. Programas, como o de Gerenciamento pela Qualidade Total (CAMPOS, 1992) e Reengenharia dos Negócios (HAMMER, 1994), fazem justamente a reestruturação da empresa para que a mesma se remodele completamente, para poder atender melhor as novas necessidades dos clientes.

Para que os objetivos das empresas sejam cumpridos, elas têm que passar a conhecer melhor os seus clientes e seus hábitos. Uma das formas de se obter este resultado é através de uma análise nos dados históricos dos negócios realizados com os clientes. Outra forma é através do preenchimento de questionários pelos clientes, em que eles descreverão seus hábitos particulares e de negócios.

Segundo ARNUM (1998), os produtos e os serviços fornecidos pelas empresas embutem a sua imagem, ou seja, bons produtos ou serviços levam os clientes a acreditar que são fornecidos por empresas com qualidade. A recíproca também é verdadeira. O melhor modo de se projetar uma imagem favorável da empresa é passar aos clientes a imagem de que se está mantendo sempre em evolução, atendendo as necessidades deles, realizando dinamicamente as mudanças que sejam necessárias, em relação às preferências deles.

FREEDMAN et al. (1998) afirmam que, da forma como o mercado está competitivo, até mesmo uma vantagem de uma fração de porcentagem (valor inferior a 1%) pode representar milhares ou até milhões de dólares em negócios. As empresas que querem prosperar devem ter a habilidade de reter bons clientes a um custo mínimo. Para se atingir esta meta, as ferramentas analíticas de alto desempenho e as ferramentas de predição estão se tornando imprescindíveis.

Os programas de reestruturação citados servem para direcionar os negócios para os clientes, utilizando diversos conceitos administrativos, mas, independentemente do conceito adotado, o grande ponto em comum e indiscutível que existe entre eles é o enfoque na estratégia de se voltar os negócios para o atendimento das necessidades dos clientes. É necessário conhecê-los melhor e, para fazê-lo, é necessário que se obtenha a maior quantidade possível de seus dados.

Alguns dados que devem ser levantados e armazenados, para futuras análises são: os fatos e os dados do mercado (dados históricos e atuais), as tendências projetadas por especialistas da área e os fatos e dados dos concorrentes. Não é necessário questionar os concorrentes ou espioná-los para obter seus dados. Alguns

dados são obrigatoriamente divulgados, tais como o balanço anual e a verificação de vendas, que são divulgados por auditores independentes. Uma pesquisa junto aos clientes também pode fornecer informações importantes, como a opinião deles sobre os produtos dos concorrentes. Cruzando as informações do mercado com as dos concorrentes e com as da empresa, monta-se um cenário da situação histórica e atual da empresa.

Em BISPO & CAZARINI (1998d) foi abordado algumas das análises que costumam ser feitas na maior parte dos conceitos modernos de administração de empresas que, juntamente com as análise feitas, para se direcionar os negócios a atender as necessidades dos clientes, ajudam a formar as estratégias de gerenciamento dos negócios são análise do/da:

- ⇒ ambiente sócio-econômico e do público alvo dos produtos ou serviços;
- ⇒ evolução histórica das características dos produtos ou serviços até se chegar ao estado atual e verificação de tendências futuras, tanto da empresa como dos concorrentes;
- ⇒ evolução histórica, situação atual e tendências da demanda dos produtos ou serviços da empresa e dos concorrentes;
- ⇒ evolução histórica do desenvolvimento tecnológico da empresa e dos concorrentes;
- ⇒ situação financeira histórica e atual da empresa e dos concorrentes;
- ⇒ evolução histórica, situação atual e tendências quanto à participação do mercado da empresa e dos concorrentes;
- ⇒ evolução histórica, situação atual e tendências no volume de produção, das vendas de produtos ou dos serviços prestados, da empresa e dos concorrentes;
- ⇒ evolução histórica, situação atual e tendências em investimentos na produção, nas vendas de produtos ou serviços, da empresa e dos concorrentes;
- ⇒ evolução histórica, situação atual e tendências do retorno dos investimentos em produção ou em serviços, da empresa e dos concorrentes;
- ⇒ evolução histórica, situação atual e tendências do retorno dos investimentos em tecnologia da informação da empresa e dos concorrentes;

- ⇒ evolução histórica, situação atual e tendências da estrutura organizacional da empresa e dos concorrentes;
- ⇒ evolução histórica, situação atual e tendências do crescimento da empresa da empresa e dos concorrentes.

Apesar de todas as análises realizadas acima, deve-se observar que o ambiente do gerenciamento dos negócios é bastante dinâmico, diante das diversas estratégias que são adotadas pela concorrência, para tentarem se expandir. É necessário traçar-se o cenário atual real com o máximo de informações possíveis e fazer-se simulações de cenários futuros, que podem vir à tona e se preparar antecipadamente para os eventuais imprevistos. Deve-se também traçar um cenário futuro que servirá de base aos objetivos da empresa e dos negócios.

O data warehouse deve favorecer estes tipos de análise através do armazenamento dos dados citados e do rápido acesso aos mesmos quando necessário.

3.7 - A elaboração de um projeto de data warehouse

A elaboração de um projeto de data warehouse que sirva para auxiliar no gerenciamento dos negócios, não é uma tarefa fácil, devido ao fato de envolver diversos conceitos e diversas tecnologias, já citadas anteriormente, que deverão ser integradas para que trabalhem em harmonia. E mesmo com toda a evolução tecnológica atual, existem muitas limitações tecnológicas que devem ser respeitadas e adequadas, durante a fase da elaboração do projeto (BISPO & CAZARINI, 1998e).

Uma empresa pode ter um excelente Retorno Sobre o Investimento (RSI) com um data warehouse pequeno e poucos usuários, e outra pode ter um retorno menos positivo, embora tenha terabytes de dados e centenas de usuários. O tamanho, por si, não é fator determinante do sucesso, o seu uso como ferramenta de suporte à decisão é que é o principal fator (TAURION, 1998a).

Um data warehouse não pode, nem deve ser um projeto exclusivo da área de sistemas (TAURION, 1997). A equipe que vai elaborar o projeto do data warehouse deve ser composta por pessoas tanto da área de negócios, como da área tecnológica. O pessoal da área de negócios deve monitorar o projeto para que o mesmo esteja sempre adequado para atender as necessidades dos negócios. O pessoal da área tecnológica deve dar o suporte necessário em todas as fases do projeto e da implementação do sistema, sempre com a preocupação básica de fazer com que o sistema esteja o mais adequado possível, para dar suporte às necessidades da empresa.

A empresa não deve esquecer seu lado cultural: se não for afeita a riscos, não deve usar excessivamente as tecnologias de ponta, mas buscar conciliar estas tecnologias com outras mais familiares (TAURION, 1998a). Segundo o autor, como as tecnologias que envolvem um data warehouse ainda são imaturas, sempre é bom deixar um orçamento prevendo uma quase inevitável substituição de alguma ferramenta por outra.

As necessidades de informações por parte dos usuários variam com o passar do tempo, adequando-se às necessidades dos negócios. Estas necessidades são impostas por diversos eventos e fatores de influência internos e externos, que ocorrem e que afetam a empresa. A elaboração do projeto do data warehouse deve prever esta dinamicidade das necessidades de informações. As ferramentas de pesquisa e de análise devem ser versáteis para que possam acompanhar estas mudanças.

Segundo TAURION (1998b), uma maneira simples de saber se um data warehouse será útil a uma empresa é responder ao pequeno check-list abaixo. Quanto maior o número de respostas "sim", maior é o seu potencial de uso.

1. A empresa baseia-se em informações para a tomada de decisões?
2. O segmento de negócios da empresa é caracterizada por uma forte concorrência e mudanças rápidas?
3. A base de clientes é grande e diversificada?

4. Os dados estão armazenados em diversos locais?
5. Os dados estão duplicados e espalhados por diversos sistemas?
6. Os dados estão em formatos e especificações diferentes?
7. A sua empresa está distribuindo o processo decisório, buscando maior agilidade e rapidez?

Com base em ALVES (1997), AMARAL Jr. (1997), BARBIERI (1997), FURLAN (1997), INMON (1997), MCELREATH (1997), BISPO & CAZARINI (1998a), INMON & HACKATHORN (1998b), KIMBALL (1998a) e TAURION (1998b), são apresentadas as principais etapas que fazem parte de um projeto lógico de data warehouse. Estas etapas não estão em uma ordem seqüencial, algumas delas podem ser realizadas paralelamente a outras ou a ordem pode ser alterada, de acordo com a necessidade; isto será definido pela equipe que está elaborando o projeto. São elas:

- ⇒ identificar os objetivos da organização, sob a óptica de seus executivos, por exemplo: “a empresa pretende crescer dentro de seu segmento de negócio?”;
- ⇒ identificar os processos de negócio diretamente relacionados a esses objetivos, com base nas prioridades emergentes da empresa sempre associadas com as prioridades reais dos negócios da empresa, por exemplo: Clientes, Finanças, Vendas, Produção, etc.;
- ⇒ definir as informações que são necessárias para dar suporte aos processos decisórios e onde essas informações serão obtidas;
- ⇒ modelar os dados que o data warehouse deverá conter;
- ⇒ determinar a granularidade e as agregações dos dados;
- ⇒ definir e detalhar as tabelas de fatos;
- ⇒ definir e detalhar as dimensões;
- ⇒ criar os metadados;
- ⇒ definir a frequência de atualização do data warehouse com dados dos aplicativos operacionais;
- ⇒ definir o tempo em que os dados se manterão armazenados;
- ⇒ definir as especificações técnicas e as alternativas tecnológicas para a implementação física;

- ⇒ escolher cuidadosamente o fornecedor dos produtos, pois trata-se de um investimento grande e de longo alcance e decisões errôneas podem ser fatais à organização;
- ⇒ criar o banco de dados físico do data warehouse;
- ⇒ povoar o data warehouse a partir das várias fontes;
- ⇒ gerar os aplicativos necessários para os usuários ou, de alguma outra forma, fornecer-lhes as ferramentas de consulta que lhes permitam ter acesso às informações armazenadas;
- ⇒ dar aos usuários e técnicos o conhecimento e o treinamento necessários para a utilização e manutenção da ferramenta;
- ⇒ prever nos orçamentos os gastos que se farão necessários, com a evolução tecnológica, pois haverá um aumento gradativo no volume de dados o que acarretará um aumento no número de consultas e análises; um suporte computacional cada vez melhor será necessário.

Para KONDRATIUK (1998), atualmente para dimensionar o problema do hardware para suportar o volume de dados do data warehouse, pode-se ter, com menos de 5 GB de dados, uma plataforma PC; entre 5 e 100 GB, exige-se um equipamento RISC de grande porte ou um mainframe; de 100 a 300 GB, são necessários servidores SMP ou grandes mainframes; acima de 300 GB, os servidores paralelos específicos para data warehouse são a solução.

Segundo VILELA (1997), o investimento em pessoal envolvendo a alta administração, usuários e profissionais de tecnologia não é um entrave, mas o meio de viabilização do projeto. O comprometimento dessas pessoas é fundamental.

Entre os fatores que podem gerar problemas na elaboração do projeto estão: custo de implementação (consultoria, custos internos, etc.), custo de operação e manutenção, mudanças estruturais na empresa e rejeição por parte dos funcionários (SEIDLER, 1995). O autor apresenta um conjunto de regras básicas que devem ser observadas na elaboração do projeto:

1. definir bem o escopo da implementação;

2. escolher o pacote com consciência;
3. analisar profundamente os processos alvos da empresa;
4. buscar o comprometimento da alta gerência;
5. envolver profundamente os usuários;
6. usar uma metodologia reconhecida e aprovada;
7. comunicar a todos muito bem o alvo e as conseqüências da implementação;
8. escolher muito bem a equipe de implementação;
9. motivar os envolvidos em todos os níveis e
10. estabelecer metas claras.

KIMBALL (1998h) cita as principais funções que um gerente de data warehouse deve exercer:

- ⇒ colecionar os dados introduzidos vindos de diversas fontes;
- ⇒ assegurar a qualidade destes dados introduzidos, fazendo as correções necessárias, removendo erros, eliminando dados nulos e combinando as múltiplas fontes;
- ⇒ tornar o data warehouse confiável;
- ⇒ ser nomeado na organização para servir como um comunicador claro;
- ⇒ ser dirigido pelas contínuas mudanças da necessidades da empresa e dos negócios e possibilitar o crescimento do volume de informações;
- ⇒ ser dirigido pelas rápidas mudanças tecnológicas em sua área, especialmente a Internet;
- ⇒ estar atento ao significado do data warehouse para a empresa e proporcionar as facilidades necessárias para o processo decisório.

No projeto do data warehouse, deve estar definida qual é a freqüência de atualização do data warehouse com dados dos aplicativos operacionais, bem como o tempo em que os dados se manterão armazenados (MCELREATH, 1997). Segundo o autor, dados muito antigos poderão produzir uma resposta que não é necessariamente condizente com a realidade e induzir a uma decisão que não seria a mais acertada. O armazenamento somente com os dados mais recentes também pode levar a resultados errôneos semelhantes. Portanto, esta última definição é muito importante e deverá ser

amplamente discutida com os analistas de negócios e com os futuros usuários do sistema.

Segundo HACKATHORN (1998b), a popularidade do data warehousing, infelizmente obscureceu alguns de seus objetivos básicos. Existem negligências que estão levando a obtenção de resultados errôneos, através da execução de procedimentos errados. A ferramenta oferece uma grande variedade de modelos que tentam descrever os negócios, e há o problema da compreensão correta das informações geradas. Há, ainda, o desafio de manter a consistência e a credibilidade dos dados armazenados e dos dados gerados nas consultas e relatórios. Pode acontecer, também, que dois usuários interpretem os mesmos dados de maneiras diferentes, sendo que, na verdade, apesar de usarem os mesmos dados, estão realizando análises diferentes ou observando pontos-de-vista diferentes. Como o ambiente é extremamente dinâmico, uma variação mínima na consulta pode levar a resultados completamente diferentes, que podem ser confundidos pelo usuário “distraído”. Os dados podem, também, ser interpretados por pessoas com pouca capacidade de análise dos negócios, ou maus administradores e estes, para superar suas dificuldades, podem dar um jeito de fazer a culpa recair sobre a ferramenta. O autor sugere que, já no projeto, para assegurar uma imagem consistente do data warehouse, estejam previstas formas de se checar freqüentemente a validade e a consistência dos dados armazenados e das consultas e relatórios gerados:

- ⇒ **cheque de uniformidade** - assegura que os valores dos dados estão dentro de limites preestabelecidos, com base em parâmetros pré-definidos ou determinados estatisticamente;
- ⇒ **cheque de versão** - verifica as mudanças, através das especificações dos metadados; um exemplo de erro comum é uso do ano civil, sendo que houve mudança para ano fiscal;
- ⇒ **cheque de perfeição** - determina se as agregações dos dados estão completas e corretas, elas são extremamente úteis, porém, podem camuflar erros ocorridos no nível dos dados com maior granularidade;
- ⇒ **cheque de conformidade** - faz a correlação formal entre os dados com as suas respectivas fontes;

⇒ **cheque de genealogia** - realiza uma completa auditoria dos dados desde as suas respectivas fontes.

A análise de viabilidade para um projeto de data warehouse deve ser realizada com um enfoque estratégico e deve ser encarado como um instrumento de vantagem competitiva (VILELA, 1997). Apesar de ter um retorno difícil de ser mensurado em números, os ganhos sob o ponto-de-vista estratégico podem ser percebidos rapidamente.

Um estudo realizado pela The Conference Board, em conjunto com a Price Waterhouse, com 50 corporações americanas com faturamento anual superior a US\$ 1 bilhão (COMPUTERWORLD, 1998b), apontou que 62% das corporações respondeu que está parcialmente satisfeita com os resultados obtidos com a implementação de um data warehouse, ou seja, em alguns setores, foi alcançado o resultado almejado e em outros isso não foi possível. Para 7% das corporações, os resultados superaram as expectativas, 19% se disse totalmente satisfeita com os resultados alcançados e 12% afirmou que não alcançou os resultados esperados (Figura 4). O estudo relata ainda que 18% das corporações realizou cortes em seus orçamentos com o data warehouse e cerca de 50% dos projetos em andamento estão atrasados em seus cronogramas.



Figura 4 - Gráfico demonstrando a satisfação de um grupo de usuários com a implantação do data warehouse

Uma análise dos resultados apresentados reforça a tese da dificuldade de se elaborar um projeto eficaz. As empresas que se dizem satisfeitas e mais que satisfeitas alcançam apenas 26% do total. As empresas que obtiveram resultados parciais alcançam 62%; isto se deve ao fato de que a implementação nos diversos setores da empresa produz resultados diferentes: alguns setores conseguem obter um bom retorno financeiro do investimento realizado e outros não. Isto deve fazer parte do projeto, através de estudos de viabilidade da implementação da ferramenta nos diversos setores da empresa. Apenas 12% das empresas não obteve o resultado esperado, o que pode significar um projeto mal elaborado.

Segundo STREHLO (1998), o término da implementação de um projeto de data warehouse não significa que o trabalho tenha terminado. É necessário fazê-lo trazer reais benefícios aos negócios e justificar o investimento realizado. O data warehouse irá crescer continuamente em tamanho e continuará a gerar custos que têm que ser revertidos em benefícios à empresa. Um data warehouse não é algo que

se constrói e depois só se faz a manutenção: é um processo contínuo e evolutivo. Não importa o quanto seja bom, se somente algumas pessoas o utilizam, pode-se ter um baixo retorno de investimento. Porém, envolvendo os usuários potenciais nas diversas fases do projeto, existirão maiores chances de se ter sucesso com a sua implementação.

A manutenção do data warehouse pode causar perda de renda, produtividade, e até mesmo de clientes (RADEN & PETERSON, 1998). Segundo os autores, as causas mais freqüentes de problemas na ferramenta são fracassos de dispositivos e de controladores, tempos de carga e de backup longos, enquanto os aplicativos operacionais estão executando outras atividades rentáveis.

Segundo a GENTIA SOFTWARE (1998), com a evolução tecnológica, houve um aumento significativo no tamanho dos sistemas de informação, o que provocou o aumento das possibilidades de falhas no hardware e no software. Isto ocorre porque mais componentes físicos e lógicos estão envolvidos, é o chamado tempo médio entre falhas (MTBF). Assim, as falhas devem ser esperadas e é essencial que elas não prejudiquem o processamento ou o armazenamento dos dados.

KIMBALL (1998d) sugere o uso da compressão de dados para o armazenamento dos dados do data warehouse. O autor afirma que o data warehouse pode ser constituído de gigabytes ou terabytes de dados estáticos que não sofrem alterações, apenas atualizações (inclusão de novos registros) e consultas. Dado que existe um custo para o armazenamento dos dados, através da compressão dos dados, este custo pode ser minimizado.

Com o data warehouse, a empresa pode mudar de aspecto (SZLEJF, 1997). Com a implantação da ferramenta, estão surgindo novos profissionais que estão trazendo novos conceitos, novas análises, novas visões de medições dos negócios e do mercado e produzindo novos resultados.

O data warehouse é uma tecnologia frágil e cara, se as instalações não são adequadas às exigências empresariais, o sistema entrará em desuso; se o retorno de investimento não aparecer, o sistema pode ser considerado como um luxo; se os executivos delegam seu uso, o lucro cultural está perdido (HORROCKS, 1997). Porém, segundo o autor, se a ferramenta foi bem planejada, cuidadosamente implementada e é dirigida para a chamada “inteligência dos negócios”, ela pode auxiliar os líderes empresariais nos seus desafios.

3.8 - A escolha do tipo de dado a ser armazenado no data warehouse

LAMBERT (1997) divide os tipos de dados a serem armazenados no data warehouse em primitivos e derivados. Alguns autores usam a denominação de dados atômicos e dados sumarizados. O referido autor afirma que dado primitivo é aquele que descreve um objeto individual ou um evento, normalmente, não pode ser calculado, mas deve ser medido e colecionado. Já o dado derivado é aquele que descreve uma coleção de objetos ou eventos, ele é calculado com base nos dados primitivos ou em outros dados derivados.

A escolha de dados atômicos para o armazenamento em um data warehouse traz diversos benefícios, porém possui algumas desvantagens. O maior benefício está na possibilidade de uma pesquisa em uma base de dados mais rica, proporcionando uma análise mais aprofundada nos dados, o que permite a verificação do histórico e, através de ferramentas apropriadas, verificar-se tendências, sazonalidades, fazer previsões, realizar auditorias e se elaborar cenários. Desta forma, o usuário do data warehouse tem à sua disposição os dados que permitirão amplas formas de pesquisas, proporcionando amplos campos de visão para solucionar seus problemas. A principal desvantagem é que o armazenamento de dados atômicos ocupará um espaço muito maior nos dispositivos de armazenamento, assim como serão necessários equipamentos mais modernos e robustos, para que não haja baixa performance nas consultas e análises dos dados. Isto significa que um orçamento maior que deve ser destinado ao projeto. Como toda empresa necessita que os

investimentos realizados tenham um retorno financeiro adequado, isto pode ser um empecilho.

A escolha de dados sumarizados para o armazenamento em um data warehouse também traz benefícios e desvantagens. O maior benefício é que os dados já estão sumarizados, ou seja, já estão resumidos e armazenados em um formato no qual são mais consultados. Ocupam menos espaço nos dispositivos de armazenamento e a performance das consultas e das análises dos dados é melhor. Os equipamentos necessários poderão ser de uma robustez inferior e os custos são menores. A desvantagem é que o armazenamento dos dados sumarizados limita bastante a capacidade de pesquisa e de análise, por parte daqueles que se vêem diante de problemas complexos e que necessitam de informações mais detalhadas, para se buscar melhores soluções para seus problemas. Um problema enfrentado, por quem usa este tipo de armazenamento, é o modo pelo qual os dados são sumarizados. A escolha deve ser um consenso entre os diversos tipos de usuários do sistema. Os negócios são dinâmicos e surgem diversos tipos de problema a todo instante. A função do data warehouse é dar suporte ao levantamento das informações que serão necessárias aos processos decisórios. Portanto, esta escolha deve ser muito bem feita.

A maioria das empresas estão optado pelas duas formas de armazenamento simultaneamente. Desta forma, somam-se as vantagens e reduzem-se as desvantagens de ambas.

A granularidade é o nível de detalhamento ou de sumarização dos dados armazenados. Por exemplo, os dados do setor de vendas podem ser armazenados com a granularidade a nível de transação individual ou sumarizados a nível diário, semanal, quinzenal, mensal, trimestral, semestral ou anual.

ADELMAN (1998) alerta para o perigo da explosão do volume de dados armazenados no data warehouse. Segundo o autor, se não houver um controle sobre quais dados serão armazenados, corre-se o risco de se armazenar um grande volume

de dados inúteis apresentando um alto custo e baixo desempenho. Com base nos conceitos do autor, recomendam-se alguns controles para evitar este problema:

- ⇒ fazer todos os usuários e a equipe de suporte reconhecer que o tamanho pode ter um impacto negativo no custo, desempenho, disponibilidade, risco e na administração do data warehouse;
- ⇒ implementar uma metodologia de justificação de custo;
- ⇒ educar os usuários sobre a importância da participação deles no processo de se determinar quais dados são necessários;
- ⇒ avaliar junto aos usuários a necessidade de se manterem armazenados dados atômicos e dados sumarizados juntos; esta necessidade pode mudar com o passar do tempo;
- ⇒ permitir o compartilhamento dos dados, em lugar de cada usuário ter a própria cópia dos mesmos dados;
- ⇒ usar os metadados para auxiliar a minimizar o armazenamento de dados redundantes;
- ⇒ monitorar a frequência de uso dos dados;
- ⇒ analisar e determinar até quando os dados históricos devem permanecer on-line;
- ⇒ analisar e determinar se todos os dados atômicos devem ser mantidos on-line;
- ⇒ determinar as exigências de tempo de resposta e se os dados históricos podem ser restabelecidos por processamentos menos caro.

3.9 - Metadados

Os metadados são usados para controlar o uso, a segurança, a integridade e o custo do data warehouse (MCELREATH, 1997). Ajudam os administradores de banco de dados (DBAs) a povoarem e gerenciarem o data warehouse, e os usuários a saber a origem dos dados. Não existem muitas ferramentas disponíveis, no mercado, para gerenciá-los e muitas empresas costumam desenvolver as suas próprias ferramentas.

KIMBALL (1998g) apresenta vários exemplos de metadados:

- ⇒ dados sobre os sistemas de informação que servem de fontes de dados;

- ⇒ descrições dos negócios envolvidos em cada fonte de dados;
- ⇒ estruturas dos bancos de dados fontes;
- ⇒ descrições das propriedades de cada fonte de dados;
- ⇒ frequência de atualização das fontes de dados;
- ⇒ frequência de atualização do data warehouse;
- ⇒ fidelidade dos dados, por exemplo: se todos os dados de janeiro foram recebidos durante o mês de janeiro;
- ⇒ relações e restrições de integridade referencial;
- ⇒ purgação e períodos de retenção;
- ⇒ regras de replicação e de distribuição;
- ⇒ métodos e regras de agregação.
- ⇒ limitações legais do uso de cada fonte de dados;
- ⇒ horários de trabalhos dos sistemas onde se encontram as fontes de dados;
- ⇒ métodos e permissões de acesso, privilégios e senhas das fontes de dados;
- ⇒ dados sobre o conjunto de ferramentas para extração automática;
- ⇒ extratos ou resultados de cada processo de extração, incluindo o tempo necessário, conteúdos e nível de perfeição da operação;
- ⇒ agenda de transmissão de dados e os resultados destas operações;
- ⇒ definições da tabelas de fatos e das dimensões;
- ⇒ especificações do trabalho de união dos dados de várias fontes, incluir/excluir campos e atributos;
- ⇒ especificações da limpeza dos dados;
- ⇒ especificações da transformação dos dados;
- ⇒ especificações das agregações;
- ⇒ especificações para a auditoria de dados;
- ⇒ versões dos diversos softwares envolvidos em todas as fases;
- ⇒ estatísticas de uso dos dados;
- ⇒ regras de segurança para a extração, limpeza e transformação dos dados;
- ⇒ regras de segurança para a transmissão dos dados;
- ⇒ regras de segurança para o armazenamento dos dados;
- ⇒ descrição das diversas tabelas utilizadas para o armazenamento;
- ⇒ descrição dos diversos índices utilizados;

- ⇒ especificações dos dispositivos de armazenamento;
- ⇒ definições dos processos de visualização dos dados;
- ⇒ definições dos tipos de consultas e tipos de relatórios possíveis;
- ⇒ especificações da documentação e do treinamento para os técnicos e usuários;
- ⇒ estatísticas do uso da ferramenta.

Não existe um padrão de utilização dos metadados. Cada empresa os utilizará da forma que julgar necessária, para exercer o controle efetivo sobre os dados armazenados no data warehouse.

3.10 - Data Mart

Da mesma forma que o data warehouse, o data mart ainda não possui uma definição universalmente aceita e também está em fase de aperfeiçoamento. Os data marts são subconjuntos de dados, dentro de um data warehouse, projetados para dar suporte a negócios de unidades organizacionais específicas (NIMER, 1998). Segundo o autor, os data marts são muito interessantes para resolver certos problemas, mas não são necessariamente substitutos de um projeto de data warehouse. Um data mart não deve ser um pequeno data warehouse, com a finalidade de ser rápido ou possuir dados ainda não ajustados para o data warehouse (KIMBALL, 1997a).

Os projetos de data marts se justificam em poucos casos, basicamente naqueles onde a alta gerência ainda não está convencida quanto à viabilidade e vantagens que a tecnologia do data warehouse pode prover às corporações. Neste caso, os data marts são viáveis, por apresentarem resultados mais rápidos, demoram entre 4 e 12 meses para serem implementados e, em consequência, começam a dar resultados mais rápidos. Os data warehouses têm prazos que variam entre 1 a 5 anos para implementação completa.

STRANGE (1998) apresenta uma tabela comparativa entre data mart e data warehouse:

Propriedades	Data Warehouse	Data Mart
Âmbito	aplicações neutras	aplicações específicas
	centralizado e compartilhado	departamento ou área de usuário
Assuntos	múltiplas áreas	área única
Fontes de dados	muitas	poucas
Tempo de implementação	<ul style="list-style-type: none"> • 9 a 18 meses para 1º estágio (duas ou três áreas de assuntos); • múltiplos estágios de implementação; 	4 a 12 meses
Características	flexível	restritivo
	nível estratégico	nível tático

Tabela 5 - Tabela comparativa entre data mart e data warehouse

Segundo DEPOMPA (1998), muitas empresas construíram seus data marts especializados para atender as necessidades imediatas de consultas e análises dos dados, sem criar primeiro uma infra-estrutura física e lógica para o data warehouse. Na hora de se criar o data warehouse, os problemas aparecerão.

O data warehouse é uma ferramenta que veio suprir os anseios dos gerentes e analistas de negócios que, durante as últimas décadas, não puderam utilizar os recursos tecnológicos existentes, para lhes proporcionar um suporte eficiente aos seus processos decisórios. Por meio desta ferramenta, atualmente, estes usuários podem obter ganhos quantitativos e qualitativos no suporte ao processo decisório. Eles perdem menos tempo com o acesso e a análise dos dados e podem dedicar mais tempo elaborando estratégias para seus negócios, baseando-se em informações e fatos analisados e comprovados. Mas isto não é tudo: o data warehouse sozinho não pode realizar todas as consultas e análises que os seus usuários necessitam. No próximo capítulo, apresenta-se a segunda ferramenta da nova geração de Sistemas de Apoio à Decisão, o OLAP, a qual proporciona análises mais sofisticadas nos dados.

Capítulo 4 - OLAP - On-Line Analytical Processing

O capítulo 3 apresentou uma solução para o problema do armazenamento e do acesso aos dados para propósitos gerenciais. Esta ferramenta já permite às empresas gerenciarem melhor os seus dados corporativos, dando-lhes grande auxílio em seus processos decisórios e no gerenciamento dos negócios. Porém, consultas e análises mais sofisticadas e cálculos mais complexos estão sendo cada vez mais necessários. Normalmente, as decisões são tomadas baseando-se em comparações e em tendências; é necessário realizar-se análises em diversas perspectivas (dimensões) do negócio com o passar do tempo. Também é necessário construir-se modelos de negócios para se realizar planejamentos, com base nos dados históricos disponíveis, ou seja, simular cenários a fim de se preparar para as incógnitas do futuro. Estes tipos de consultas, análises e cálculos não estão disponíveis nas ferramentas que acompanham o pacote de produtos para o data warehouse.

Para suprir estas necessidades das empresas, surgiu a segunda ferramenta da nova geração de Sistema de Apoio à Decisão, o OLAP - On-Line Analytical Processing, que será abordado neste capítulo.

4.1 - Conceitos básicos sobre OLAP

A ferramenta OLAP é constituída de um conjunto de tecnologias especialmente projetadas para dar suporte ao processo decisório através de consultas, análises e cálculos mais sofisticados nos dados corporativos, estejam armazenados

em um data warehouse ou não, por parte dos seus usuários, geralmente analistas, gerentes e executivos.

Esta ferramenta tornou-se a sucessora dos Sistemas de Informações para Executivos (EIS - Executive Information Systems) e ainda possui um âmbito mais amplo que a ferramenta superada (SIMON, 1998). O OLAP não somente dá suporte ao processo decisório, como também às estratégias dos negócios.

O OLAP permite aos seus usuários ganharem perspicácia nas consultas e análises dos dados, através de um acesso consistente, interativo e rápido em uma grande variedade de possíveis visões dos dados (FORSMAN, 1998). Esta ferramenta transforma dados crus em informações que são facilmente compreendidas pelos usuários e refletem a real dimensionalidade dos negócios da empresa. Segundo o autor, a ferramenta aumenta a produtividade dos usuários, a sua flexibilidade permite-lhes tornar-se mais auto-suficientes. Permite-lhes, também, modelar problemas que seriam impossíveis de serem solucionados usando outras ferramentas menos flexíveis.

A ferramenta OLAP permite, aos seus usuários, ter acesso aos dados que descrevem os negócios da empresa, permitindo-lhes uma melhoria na compreensão, gerenciamento e planejamento destes negócios. Permite, ainda, analisar as múltiplas dimensões dos dados usados nas empresas, em qualquer combinação e em qualquer ângulo, além de identificar tendências e descobrir o que está conduzindo os negócios.

No OLAP as respostas não são automáticas. Trata-se de um processo interativo, onde o usuário formula hipóteses, faz consultas, recebe informações, verifica um dado específico em profundidade e faz comparações (CARVALHO, 1997b). Ajuda os usuários a sintetizarem as informações sobre a empresa, através de comparações, visões personalizadas, análises estatísticas, previsões e simulações (FIGUEIREDO, 1998).

A maioria das ferramentas OLAP são implementadas para ambientes multiusuário e arquitetura cliente/servidor, o que proporciona respostas rápidas e consistentes às consultas iterativas executadas pelos usuários, independentemente da complexidade da consulta (FIGUEIREDO, 1998).

A ferramenta pode ser usada em diversas funções organizacionais:

- ⇒ Departamentos de Finanças - para planejar orçamentos e realizar análises financeiras;
- ⇒ Departamento de Vendas - para fazer análises e estimativas de vendas;
- ⇒ Departamento de Marketing - para realizar pesquisas e análises de mercado, estimativas, análises de clientes e segmentação de mercado;
- ⇒ Manufatura - para realizar o planejamento, análises da produção e análises de falhas ou defeitos.

Segundo FIGUEIREDO (1998), MOREIRA FILHO (1998), PENDSE (1998a) e TYO (1998), a principal característica das ferramentas OLAP é permitir uma visão conceitual multidimensional dos dados de uma empresa. Esta visão é muito mais útil para os usuários que a visão tradicional baseada em tabelas (modelo entidade-relacionamento), utilizada nos sistemas de processamento de transação (OLTP). A visão multidimensional dos dados é um conceito que pode parecer algo completamente abstrato e irreal; porém, é mais natural, mais fácil e intuitiva, permitindo a visão dos negócios da empresa em diferentes perspectivas e, assim, transformando os usuários em exploradores de informações.

As ferramentas OLAP permitem aos usuários analisar os dados em dimensões múltiplas, como região, produto, tempo e vendedor. Cada dimensão também pode conter hierarquias, por exemplo, a dimensão tempo pode conter as hierarquias ano, trimestre, mês, semana ou dia. A dimensão região pode ter as hierarquias continente, país, estado, cidade ou bairro. Os dados, nestas dimensões, são agregados, ou seja,

são resumidos, mas pode-se navegar livremente de uma hierarquia para outra, até chegar-se na máxima granularidade dos dados, o nível das transações realizadas.

Com base nos dados fornecidos por RADEN (1997a e 1998a), FIGUEIREDO (1998) e FORSMAN (1998), sobre especificações técnicas de alguns produtos OLAP, são apresentadas as principais características da ferramenta:

- ⇒ permitir a visão multidimensional dos dados;
- ⇒ possuir inteligência de tempo (saber reconhecer, por exemplo, os dias da semana, os dias úteis e não úteis e os feriados);
- ⇒ realizar cálculos complexos ;
- ⇒ criar agregações e consolidações;
- ⇒ fazer previsões, análises de tendências, comparações e análises estatísticas avançadas;
- ⇒ construir cenários a partir de suposições e fórmulas aplicadas pelos analistas aos dados históricos disponíveis;
- ⇒ fazer cálculos e manipular dados através de diferentes dimensões;
- ⇒ navegar com extrema facilidade entre os diversos níveis de agregação dos dados;
- ⇒ analisar os dados, através de qualquer combinação possível entre os mesmos, possibilitando as mais variadas visões possíveis do negócio;
- ⇒ formar, trabalhar e analisar subgrupos dos dados com as mesmas possibilidade do conjunto inteiro;
- ⇒ analisar os dados de acordo com as regras do negócio como: crescimento, porcentagem, mudanças ou alterações e análise de mercado;
- ⇒ permitir alterações da moeda corrente sempre que necessário;
- ⇒ permitir aos usuários uma grande variedade de opções de relatórios que lhes permita traduzir melhor suas necessidades de informações, incluindo documentos em HTML;
- ⇒ trabalhar com múltiplas fontes de dados como data warehouse, data marts, bancos de dados operacionais e outras fontes externas de dados;

- ⇒ realizar análises avançadas como: análise de série temporais, análise de estoques, regressão linear e não linear, probabilidades, desvio padrão, correlação e análise de variância, análise de Monte Carlo e análises do tipo “O que ... se ...”;
- ⇒ dar suporte a ambientes de Workgroup e ambientes distribuídos;
- ⇒ possuir várias funções matemáticas e estatísticas que habilitam o desenvolvimento de cenários, comparações entre conjuntos e subconjuntos de dados;
- ⇒ ter interface com diversas outras ferramentas comuns e de grande utilização pelos usuários, como: planilhas eletrônicas, ferramentas geradoras de consultas e relatórios, Web Browsers, ferramentas de análises estatísticas, ferramentas de visualização gráfica.

No Anexo B é apresentado o exemplo de um produto OLAP e suas propriedades. No Apêndice 1 são citados vários produtos OLAP acompanhados de um pequeno resumo de suas características.

O termo OLAP foi definido por E.F. Codd, que também criou, em 1993, um conjunto de 12 regras que servem para avaliar a ferramenta. Apesar desse ter sido um trabalho encomendado pela empresa Arbor, para diferenciar seu banco de dados multidimensional Essbase, este termo foi adotado pela comunidade da Tecnologia da Informação, as regras logo se espalharam pelo mercado e hoje servem como referência para desenvolvedores e usuários de produtos OLAP.

Segundo CODD et al. (1998), as doze regras para avaliar os produtos OLAP são:

1. **Visão conceitual multidimensional** - a visão de um usuário, no universo dos negócios, é multidimensional por natureza. Esta visão do usuário facilita tanto as análises como também os cálculos inter e intra dimensionais, através de um modelo analítico mais intuitivo. Os usuários podem manipular os modelos multidimensionais de dados mais facilmente e intuitivamente.

2. **Transparência** - o OLAP deve poder interagir facilmente com os produtos *front-end* habituais dos usuários, por exemplo: planilhas eletrônicas, processadores de textos ou aplicativos DSS. Se a ferramenta está em uma arquitetura cliente/servidor, então este fato deve ser bem transparente ao usuário. Deve ter uma verdadeira arquitetura de sistemas abertos e deve permitir embutir uma ferramenta analítica em qualquer lugar onde o usuário deseje, sem provocar impactos na sua funcionalidade.
3. **Acessibilidade** - a ferramenta OLAP tem que traçar seu próprio esquema lógico para tratar com os dados heterogêneos armazenados e executar qualquer conversão necessária a fim de apresentar ao usuário uma única, coerente e consistente visão dos dados.
4. **Desempenho consistente de fornecimento de informações** - mesmo com o aumento do número de dimensões ou do tamanho do banco de dados, o usuário não deve perceber uma degradação significativa no desempenho do fornecimento de informações.
5. **Arquitetura cliente/servidor** - atualmente a maioria dos dados são armazenados em um mainframe ou em um servidor de rede e estes dados são acessados através de computadores pessoais. Portanto, é necessário que a ferramenta seja capaz de operar em um ambiente cliente/servidor.
6. **Dimensionalidade genérica** - a estrutura básica dos dados e o formato dos relatórios não devem ser influenciados por qualquer dimensão de dados. Devemos poder executar manipulações ou cálculos intra e inter dimensões.
7. **Manipulação dinâmica da matriz esparsa** - para qualquer determinada matriz esparsa de dados, existe um e somente um esquema físico ótimo. Este esquema provê máxima eficiência e operacionalidade. A ferramenta deve poder ajustar seu

esquema físico, para o máximo desempenho, baseando-se na densidade dos dados armazenados.

8. **Suporte multiusuário** - muitas vezes, vários usuários necessitam trabalhar simultaneamente com o mesmo modelo analítico ou criar modelos diferentes a partir dos mesmos dados. A ferramenta têm que prover esse acesso simultâneo, sem prejuízo à integridade e segurança dos dados.
9. **Operações irrestritas com dimensões cruzadas** - quaisquer duas (ou mais) células, apesar de pertencerem a dimensão diferentes dentro de um determinado modelo analítico, devem poder ser usadas para realizar quaisquer cálculos. Esses cálculos não podem ser restritos a uma dimensão de cada vez. Qualquer conjunto de células de dados deve poder ser acessado para cálculos a qualquer momento.
10. **Manipulação intuitiva dos dados** - todo o processo de criação de modelos, manipulação de dados e realização de cálculos deve acontecer da forma mais intuitiva possível, por exemplo, através de um simples click do mouse em um determinado botão do sistema que é, intuitivamente, o mais indicado para realizar aquela tarefa.
11. **Relatórios flexíveis** - a análise e a apresentação dos dados tornam-se mais simples quando linhas, colunas e células, que vão ser comparadas visualmente, são organizados por algum agrupamento lógico que aconteça naturalmente nos negócios. Relatórios devem ser capazes de apresentar os dados de forma sintetizada ou informações que são o resultado de cálculos de um modelo criado, de acordo com qualquer visão possível. Isto significa que linhas, colunas ou títulos de página devem, cada um, ser capaz de conter e exibir até N dimensões cada, onde N é o número de dimensões no modelo analítico criado.
12. **Dimensões e níveis de agregação ilimitados** - qualquer ferramenta OLAP deve poder acomodar pelo menos quinze e até vinte dimensões de dados, dentro de um

modelo analítico comum. Além disso, cada uma das dimensões deve permitir um número ilimitado de níveis de agregação definidos pelos usuários.

A HYPERION (1998) apresenta uma tabela comparando algumas características das ferramentas OLTP, data warehouse e OLAP.

Sistema	OLTP	Data Warehouse	OLAP
Propósito	operacional	armazenamento e acesso aos dados histórico detalhados	analítico
Tipo de acesso	leitura e escrita	somente leitura	leitura e escrita
Modo de acesso	atômico	consultas e relatórios	iterativo, comparativo e investigativo
escopo	aplicações específicas	dados corporativos	análise dos dados
nível de detalhe	transação	dados limpos e sumarizados	sumarizados e calculados
estrutura dos dados	normalizados	desnormalizados	dimensional e hierárquicos
implementação	vários meses ou anos	vários anos	semanas

Tabela 6 - Comparação entre os ambientes de OLTP, data warehouse e OLAP

PENDSE (1998b) afirma que as ferramentas de análises multidimensionais existem há mais tempo que os banco de dados relacionais, desde a década de 70. Porém, somente na década de 90 atribuiu-se a elas o termo OLAP.

Existem 4 tipos de estruturas OLAP:

- ⇒ Relational OLAP (ROLAP);
- ⇒ Multidimensional OLAP (MOLAP);
- ⇒ Hybrid OLAP (HOLAP) e o
- ⇒ WOLAP ou Web OLAP.

Nos próximos tópicos, serão abordadas cada uma das quatro estruturas possíveis à ferramenta OLAP.

4.2 - ROLAP - Relational On-Line Analytical Processing

O ROLAP é uma simulação da tecnologia OLAP, feita em banco de dados relacionais, que, por utilizar a estrutura relacional, possui a grande vantagem de não ter restrições no volume de armazenamento de dados (CARVALHO, 1997b). Esta estrutura utiliza o modelo dimensional (item 3.4) para gerenciar os dados.

A principal vantagem de se adotar uma ferramenta ROLAP é a utilização de uma tecnologia estabelecida, de arquitetura aberta e padronizada como é a relacional, beneficiando-se da diversidade de plataformas, escalabilidade e paralelismo de hardware (FIGUEIREDO, 1998). Segundo a autora, quanto às limitações, cita-se o pobre conjunto de funções para análises dimensionais e a inadequação do esquema estrela (*star scheme*), para se realizar a manipulação dos dados proporcionando a visão multidimensional dos dados.

Segundo a GENTIA SOFTWARE (1998), as ferramentas ROLAP podem realizar o processamento dos dados para efetuar as consultas, análises ou cálculos no modelo dimensional de duas formas, dependendo da ferramenta e do suporte de hardware, gerar múltiplos e complexos passos em linguagem SQL:

1. no próprio servidor OLAP ou
2. carregando os dados necessários em outro equipamento, que pode ser outro servidor ou os equipamentos dos clientes no ambiente cliente/servidor.

Isto ocorre porque a linguagem SQL não pode executar as atividades do modelo dimensional diretamente (GENTIA SOFTWARE, 1998). Diferentes fornecedores de ferramentas ROLAP usam técnicas diferentes para superar esta dificuldade e alguns conseguiram alcançar um nível surpreendente de funções de

atividades do modelo dimensional com a linguagem SQL, mas às custas de ter que realizar o processamento com consideráveis múltiplos passos e utilizando diversas tabelas temporárias.

4.3 - MOLAP - Multidimensional On-Line Analytical Processing

As necessidades de consulta no ambiente dos negócios, requerem, freqüentemente, visões dos negócios sob várias perspectivas (KENAN TECHNOLOGIES, 1995). Raramente um modelo de negócio é representado por menos de três dimensões (MOREIRA FILHO, 1998). Por exemplo, em uma análise de vendas é necessário cruzarem-se informações da dimensão produto com a dimensão tempo e ainda com a dimensão região, obtendo assim, em cada dado, a interseção entre as três dimensões, ou seja, as vendas por produto por mês e por região.

A modelagem multidimensional é a técnica utilizada para se ter uma visão multidimensional dos dados, com que os dados são modelados em uma estrutura multidimensional conhecida por cubo. As dimensões do cubo representam os componentes dos negócios da empresa. A célula resultante da interseção das dimensões é chamada de medida e geralmente representa dados numéricos como “unidades vendidas”, “lucros” ou “custos.” No exemplo anterior, a medida é o total de vendas e as dimensões são o produto, o tempo (mês) e região.

Através da visão multidimensional, pode-se comparar qualquer parte do negócio com qualquer outra parte e definir novas análises conforme necessário, sem se ter que projetar um novo banco de dados para cada análise a ser realizada (GENTIA SOFTWARE, 1998). Um número imprevisível de tipos de análises podem ser necessários, em um curto e intenso período de trabalho.

Segundo FIGUEIREDO (1998) e PENDSE (1998b), o MOLAP é uma classe de sistemas que permite a execução de análises bastante sofisticadas, usando bancos

de dados multidimensionais (MDB ou MDDB). Na ferramenta MOLAP, os dados são mantidos em estruturas de dados do tipo “array” de maneira a prover um ótimo desempenho no acesso a qualquer dado. A forma de acesso e de agregação dos dados fazem com que esta ferramenta tenha um excelente desempenho. Além de ser rápida, outra grande vantagem é o rico e complexo conjunto de funções de análises que oferece.

Nos dois exemplos abaixo, pode-se notar a diferença entre dados armazenados em tabelas (modelo relacional) e os mesmos dados armazenados em uma estrutura multidimensional. As Figuras 5 a 8 apresentam uma estrutura de negócio de uma concessionária de veículos. A Figura 5 mostra uma tabela que contém três tipos de carros, onde cada um deles está relacionado com outra tabela, contendo as suas vendas por cor. A Figura 6 mostra os mesmos dados, em uma estrutura multidimensional (bidimensional), demonstrando sua maior simplicidade, facilidade de compreensão e manipulação dos dados.

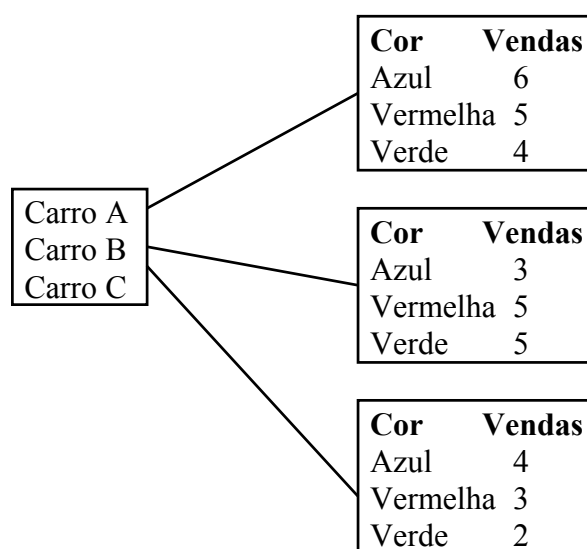


Figura 5 - Modelo relacional (tabelas relacionadas)

Modelo	Carro A	Carro B	Carro C
	6	3	4
	5	5	3
	4	5	2
	Azul	Vermelho	Verde



Figura 6 - Modelo multidimensional (duas dimensões)

As Figuras 7 e 8 exibem a mesma estrutura de negócios, porém, de uma forma um pouco mais complexa; agora temos a visão da rede de concessionárias. A Figura 7 apresenta um tabela contendo os dados das concessionárias, cada concessionária está relacionada com uma tabela, contendo os tipos de carros, e, por sua vez, cada tipo de carro está relacionado com outra tabela, contendo suas vendas por cor. No total, são necessárias diversas tabelas para representar este negócio. A Figura 8 apresenta os mesmos dados em uma simulação visual do modelo multidimensional (tridimensional). Esta figura torna mais nítida a simplicidade, facilidade de compreensão e de manipulação dos dados para este tipo de análise do negócio.

A manipulação dos dados, incluindo as consultas, é muito mais rápida e intuitiva no modelo multidimensional em comparação ao modelo relacional. Nos exemplos acima, enquanto uma busca no modelo relacional exige a navegação entre diversas tabelas, no modelo multidimensional isto não é necessário, o que o torna mais eficiente e com melhor desempenho. Algumas das vantagens dos bancos de dados multidimensionais sobre os relacionais são:

- ⇒ maior facilidade de navegação e apresentação dos dados;
- ⇒ maior facilidade na manutenção dos dados;
- ⇒ maior desempenho.

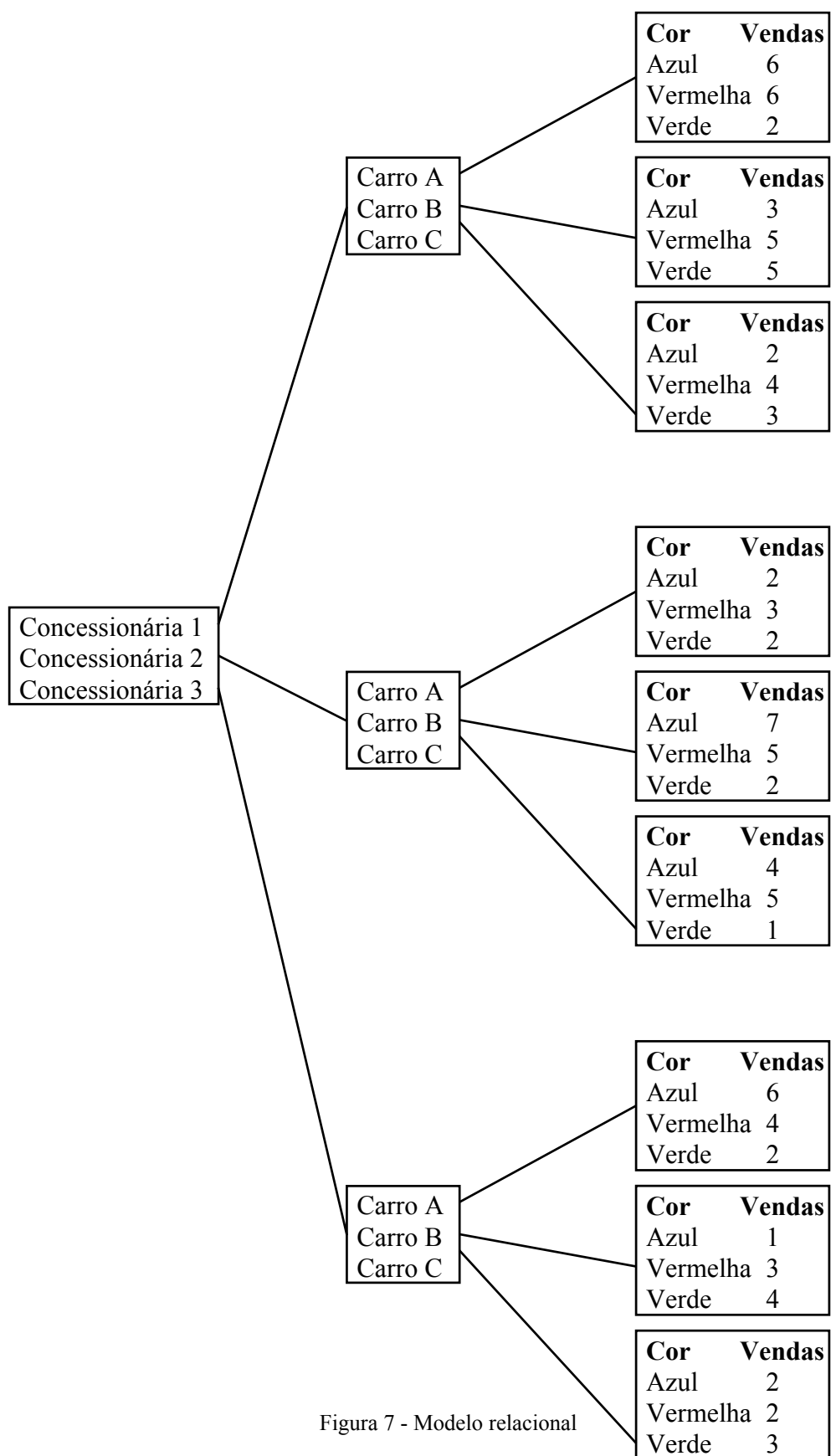


Figura 7 - Modelo relacional

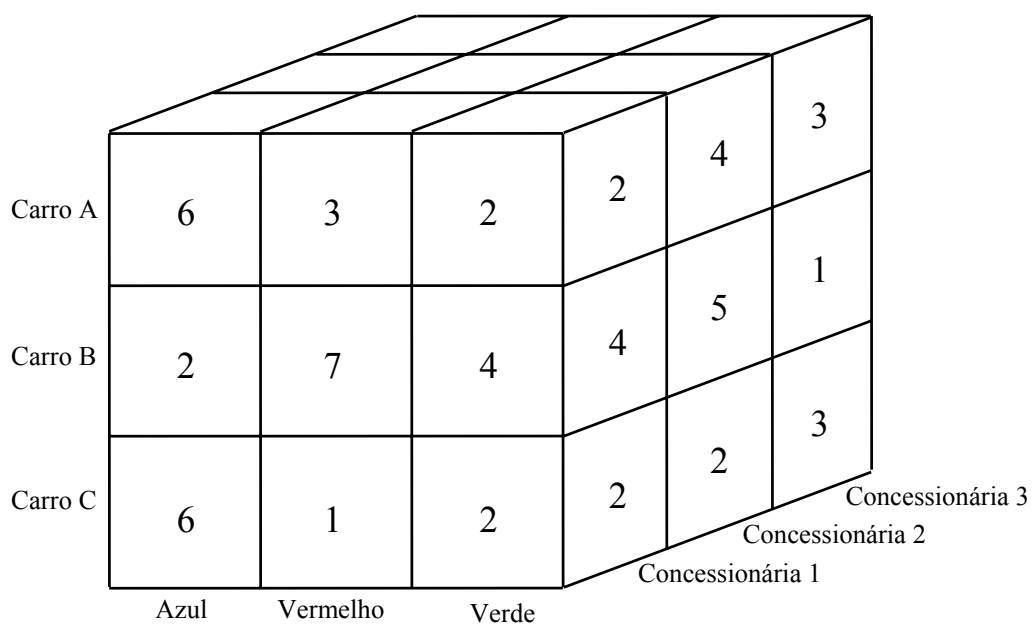


Figura 8 - Modelo multidimensional (tridimensional)

O modelo multidimensional não é apropriado para todos os tipos de situações, como por exemplo:

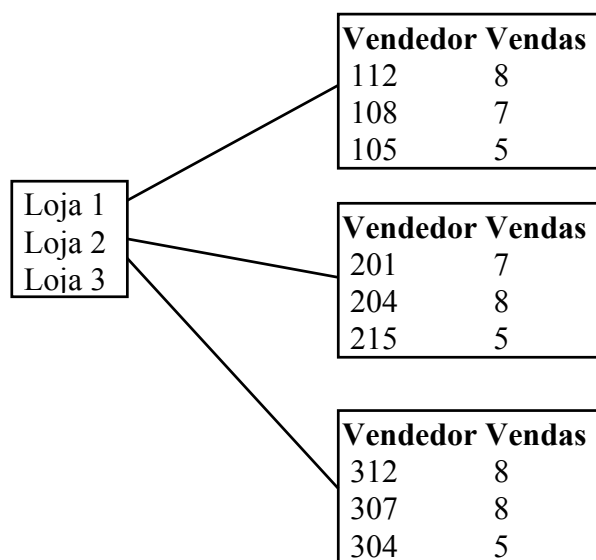


Figura 9 - Modelo relacional

Vendedor	Loja 1	Loja 2	Loja 3
112	8		
108	7		
105	5		
201		7	
204		8	
215		5	
312			8
307			8
304			5

Figura 10 - Modelo multidimensional (bidimensional)

Os dados no exemplo acima não deveriam ser representados pelo modelo multidimensional, porque não há inter-relacionamento entre eles, o que ocasionou a esparsividade dos dados, ou seja, a maioria das células não foram preenchidas e os dados se tornaram esparsos. Isto traz uma regra fundamental para a utilização do modelo multidimensional: quanto maior o inter-relacionamento entre os dados, melhor será a utilização destes dados pelo modelo.

O modelo multidimensional permite a visão dos dados de diversas formas, com extrema facilidade, o que o torna extremamente flexível e faz com que seja de grande utilidade para os processos de consultas e análises, em consequência, torna-se uma poderosa ferramenta no suporte ao processo decisório. Retornando ao exemplo de rede de concessionárias de vendas de carros, as Figuras 11 a 16 apresentam uma simulação de como visualizar, de diversas formas, os mesmos dados.

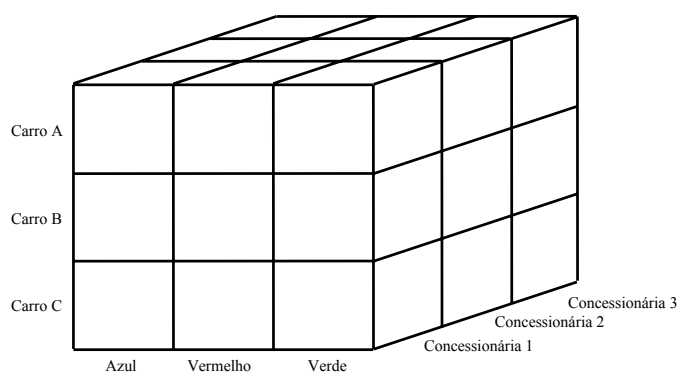


Figura 11 - Visão 1: **modelo por cor por concessionária**

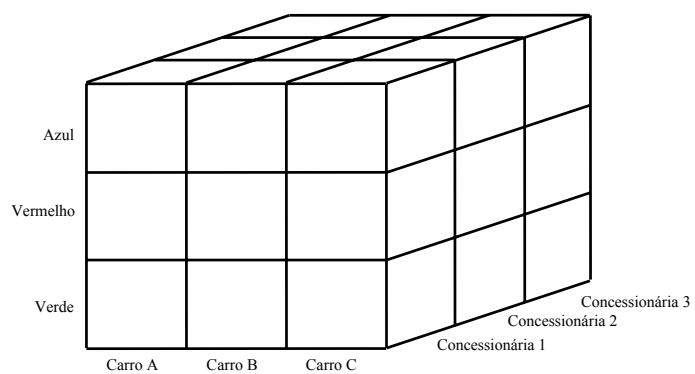


Figura 12 - Visão 2: **cor por modelo por concessionária**

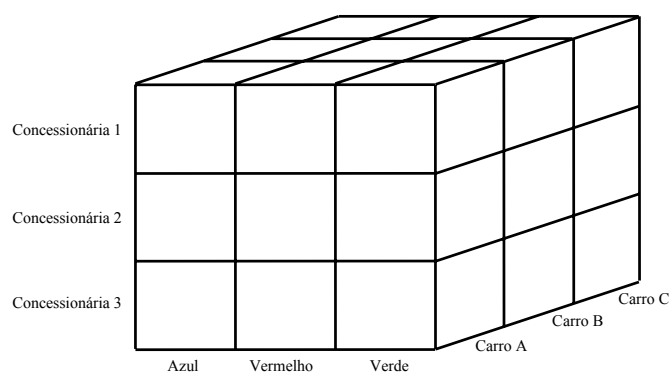


Figura 13 - Visão 3: **concessionária por cor por modelo**

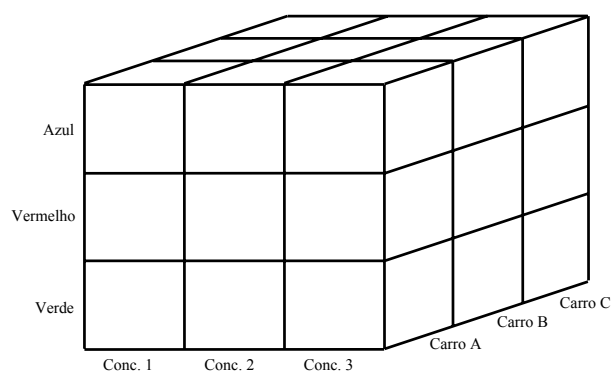


Figura 14 - Visão 4: **cor** por **concessionária** por **modelo**

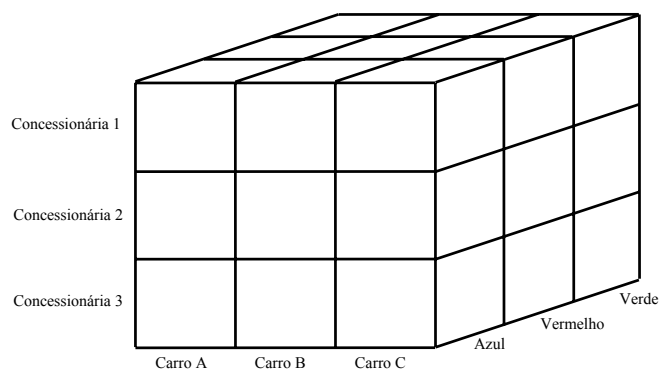


Figura 15 - Visão 5: **concessionária** por **modelo** por **cor**

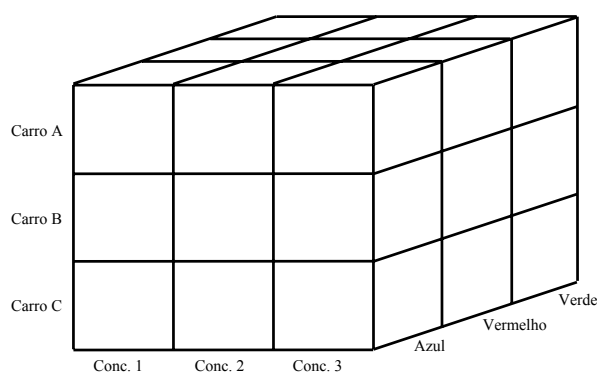
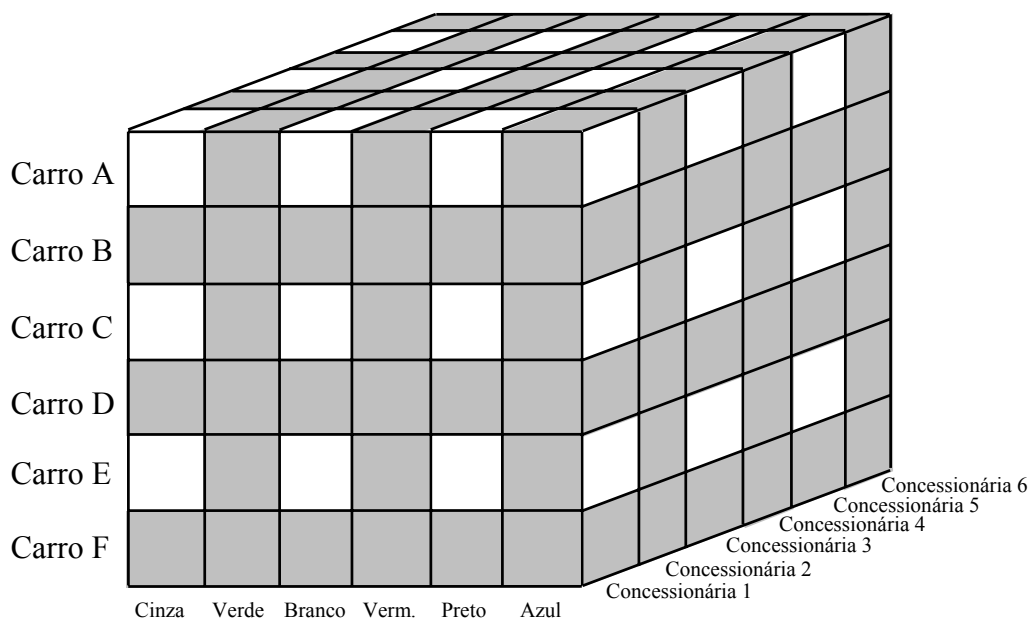


Figura 16 - Visão 6: **modelo** por **concessionária** por **cor**

Segundo a KENAN TECHNOLOGIES (1995), qualquer das seis visões exibidas acima pode ser obtida rápida e facilmente através de uma operação chamada “data slicing”. Esta operação é bastante simples para os bancos de dados

multidimensionais, não havendo a necessidade de reestruturação dos dados. A quantidade de visões possíveis aumenta exponencialmente com o número de dimensões, ou seja, esta quantidade (n) é obtida calculando-se o fatorial do número de dimensões existentes (d), assim $n = d!$. Isto proporciona ao usuário diversas formas de visualizar seus negócios. Um banco de dados multidimensional permite ao usuário navegar rapidamente e escolher a visão dos dados que melhor lhe satisfaça.

O exemplo das concessionárias de vendas de carros será ampliado para trabalhar com 6 lojas, 6 tipos de carros e 6 cores, gerando uma matriz tridimensional de $6 \times 6 \times 6$. Se o usuário quiser determinar o volume de vendas, para os modelos carro B, D e F, nas cores verde, vermelho e azul, nas concessionárias 2, 4 e 6, através de uma operação chamada “data dicing”, ele seleciona os dados necessários ao longo de cada dimensão. Isto está ilustrado na Figura 17. É criado um banco de dados temporário, que possui somente os dados solicitados e as mesmas propriedades do banco de dados de origem.



↓ data dicing

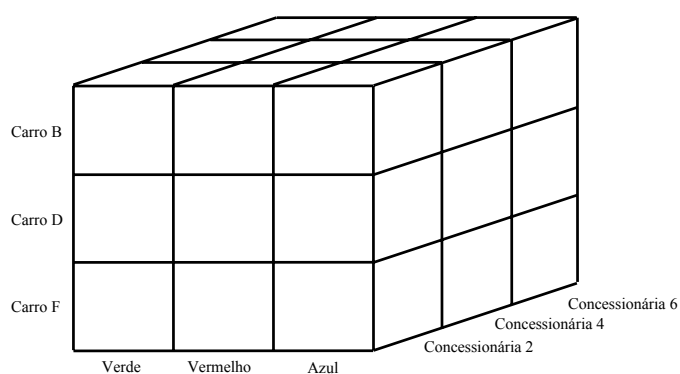
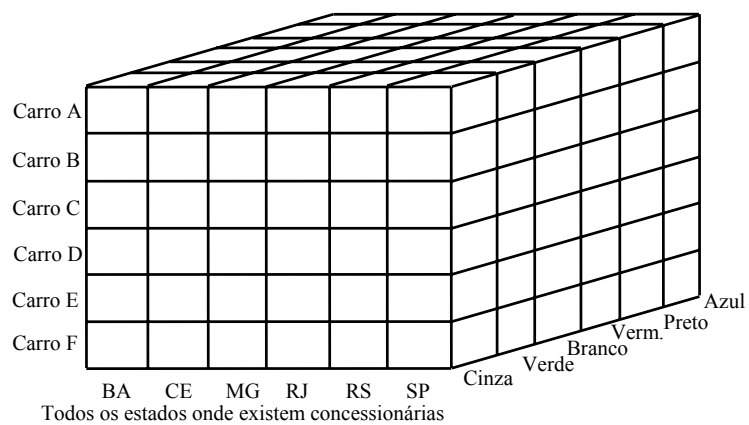
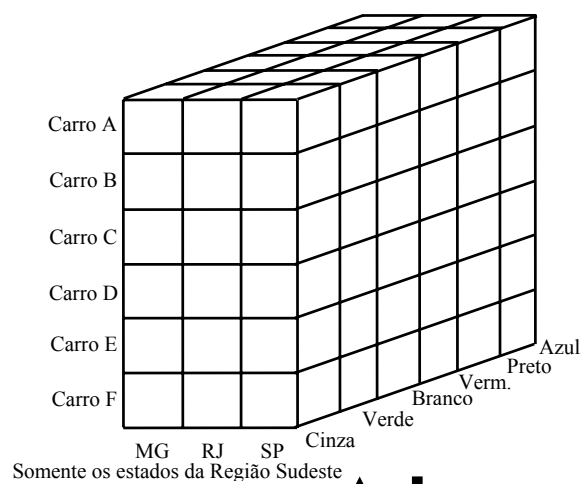


Figura 17 - Seleção de dados (data dicing)

Outro tipo de consulta ou análise necessária é a hierárquica; no exemplo, usando a dimensão região, seria verificar os resumos das vendas por região do Estado, depois por Estado, e finalmente por região do país. É possível se criar dimensões separadas para cada hierarquia, porém, uma solução melhor é definir os níveis de agregação relacionados com cada dimensão. Cada hierarquia é uma agregação do nível hierárquico imediatamente anterior, a navegação para um nível hierárquico acima é chamada de “roll-up” e a navegação para um nível hierárquico abaixo é chamado de “drill-down” (KENAN TECHNOLOGIES, 1995).



drill-down ↑ ↓ roll-up



drill-down ↑ ↓ roll-up

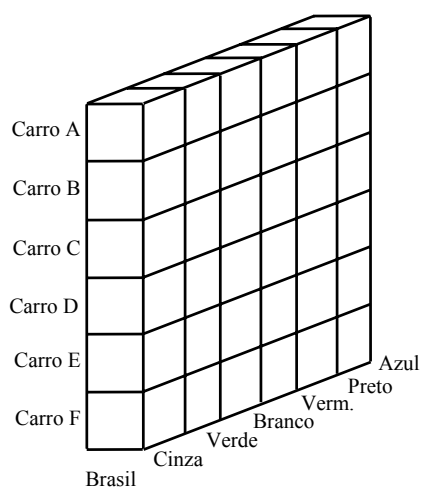


Figura 18 - Operações “roll-up” e “drill-down”

As consultas, nos bancos de dados multidimensionais, podem ser realizados diretamente nos mesmos ou pode-se gerar relatórios de diversos tipos e formas, com a mesma facilidade com que se navega pelo banco de dados (KENAN TECHNOLOGIES, 1995). Pode-se, ainda, utilizar-se das interfaces com outros sistemas, como planilhas eletrônicas, outros bancos de dados, programas de geração de relatórios e DSS específicos. Permite também que diversos usuários façam consultas simultâneas, cada um definindo sua própria visão dos dados, sem a interferência de um usuário com o outro. Possibilita, ainda, a utilização de múltiplos usuários, sem apresentar queda de desempenho, pois trabalham em ambiente cliente/servidor, o que permite a distribuição e processamento de dados entre as máquinas do cliente e do servidor.

O armazenamento em um banco de dados multidimensional pode ter entre metade até um quinto do espaço utilizado para armazenar os mesmos dados, em um banco de dados relacional (PENDSE, 1998c). Isto acontece porque não são necessários índices e a estrutura multidimensional é mais enxuta. Porém, no MOLAP, como existe a possibilidade dos dados esparsos, pode ocorrer a chamada “explosão” de armazenamento do banco de dados, ou seja, um imenso banco de dados multidimensional, contendo poucos dados armazenados.

Segundo RADEN (1997a), FIGUEIREDO (1998) e GENTIA SOFTWARE (1998), existem duas arquiteturas para os sistemas MOLAP: a hipercubo e a de multicubos. Apesar de se utilizar o termo cubo, que dá a impressão de trabalhar com apenas três dimensões, qualquer uma das duas arquiteturas pode trabalhar com diversas dimensões.

Na arquitetura hipercubo, há um único cubo onde os dados são armazenados. Por exemplo: um cubo, cujas medidas são “vendas”, pode possuir as dimensões “produto”, “região”, e “mês”, proporcionando, através de cada célula, uma visão das vendas por produtos, por região por mês.

A vantagem desta arquitetura é o rápido tempo de resposta, independentemente do número de dimensões envolvidas na consulta. A desvantagem é de uma maior necessidade de espaço, para armazenamento em disco e a maior possibilidade de ocorrência dos dados esparsos.

Na arquitetura multicubos, os dados são armazenadas em vários cubos, sendo que, em cada cubo, são agrupadas as dimensões que mais se relacionam. Por exemplo, nos dados das “vendas” podemos colocar em um cubo as dimensões “produto”, “região” e “mês”, em outro, as dimensões “produto”, “custo” e “departamento”, e em um outro, as dimensões “produto”, “faturamento” e “ano”.

As vantagens desta arquitetura são a menor utilização de espaço de armazenamento em disco, por diminuir o problema dos dados esparsos, e o melhor desempenho em consultas em um único cubo. A desvantagem ocorre, quando é necessário realizar-se uma consulta em mais de um cubo. Neste caso, é exigido bastante tanto do software, como do hardware e há uma queda no desempenho proporcionando respostas um pouco mais demoradas, dependendo da consulta realizada.

Segundo RADEN (1997a e 1998a) e FIGUEIREDO (1998) existem algumas limitações e desvantagens nos sistemas MOLAP. Os bancos de dados multidimensionais são sistemas proprietários que não seguem padrões, ou seja, cada desenvolvedor cria a sua própria estrutura para o banco de dados e as suas próprias ferramentas de suporte ao mesmo. Nos MDBs, um dos recursos para garantir um excelente desempenho é manter os “arrays” na memória, isso acaba limitando sua capacidade a alguns gigabytes de dados.

4.4 - ROLAP ou MOLAP, qual a melhor tecnologia OLAP?

Existe um grande debate a este respeito. Cada uma possui vantagens e desvantagens, como foi demonstrado nos tópicos anteriores. Antes de se optar por

qualquer delas, é importante se compreender muito bem os negócios e saber-se qual é o suporte necessário para gerenciá-lo. A partir daí, começa-se o processo de escolha da ferramenta que seja a solução mais adequada.

Segundo RADEN (1997a e 1998a), todos os bancos de dados relacionais possuem o mesmo modelo relacional, têm as mesmas características fixas e podem ser acessados por uma linguagem comum, o SQL. A exata situação oposta existe nos MDBs, armazenam os dados em “arrays” e cada produto é substancialmente diferente de qualquer outro. Os MDBs não têm nenhum método de acesso padrão, como a linguagem SQL.

A GENTIA SOFTWARE (1998) fez algumas comparações entre os dois tipos de tecnologia OLAP que podem auxiliar na escolha do tipo que seja o mais adequado para atender as necessidades dos negócios:

- ⇒ **O ROLAP pode fornecer mais dados que o MOLAP?** As ferramentas ROLAP podem gerenciar bancos de dados maiores que a maioria das ferramentas MOLAP, mas o desempenho e a manutenção limitam o ROLAP. Porém, frequentemente são usados subconjuntos de grandes volumes de dados, para consultas e análises, isto é vantajoso para os dois tipos de OLAP.
- ⇒ **Ferramentas ROLAP podem fazer uma melhor análise em tempo real dos dados operacionais?** Geralmente não é desejável executar uma análise multidimensional complexa, diretamente em um banco de dados operacional. Se algumas consultas precisam ser realizadas com as informações detalhadas mais recentes de uma aplicação OLTP, isto é possível tanto ao ROLAP, como ao MOLAP e os tempos de respostas de ambos são semelhantes.
- ⇒ **O ROLAP é melhor que o MOLAP em aplicações onde os dados mudam rapidamente?** O processo de otimização, atualização do dicionário de dados, reindexação e atualização das agregações é semelhante no ROLAP e no MOLAP.

- ⇒ **Dados armazenados no ROLAP estão mais disponíveis por estarem em tabelas?** Os dados armazenados no ROLAP podem ser acessados por uma maior variedade de ferramentas. Porém, para adquirir um desempenho aceitável e um armazenamento de dados eficiente, os dados são armazenados, freqüentemente, em numerosas tabelas relacionais com formatos não óbvios, o que prejudica a compreensão da estrutura dos dados e a sua manipulação.
- ⇒ **Os dados são carregados mais rapidamente em um ROLAP?** O carregamento dos dados inclui a leitura dos registros de dados, a validação, o armazenamento, a indexação, pré-cálculos e revisões, tudo isso enquanto se mantém o banco de dados on-line e consistente. É muito improvável que este processo global seja mais rápido ou mais barato em um ROLAP que em um MOLAP bem projetado.
- ⇒ **O MOLAP proporciona cálculos mais rápidos e melhor desempenho de recuperação dos dados que o ROLAP?** O MOLAP é projetado e desenvolvido especificamente para análises multidimensionais. O ROLAP necessita de mais processamento, para executar as mesmas tarefas que o MOLAP. Porém, uma solução ROLAP bem projetada e otimizada (incluindo o hardware) pode, na prática, ter um desempenho semelhante ao de um bom MOLAP.
- ⇒ **Indexação avançada permite às ferramentas ROLAP simularem os bancos de dados multidimensionais?** As ferramentas ROLAP, que usam indexação avançada, ocupam menos espaço e são mais ágeis, mas, utilizando somente este recurso, ainda não podem alcançar o mesmo desempenho das ferramentas MOLAP.
- ⇒ **As ferramentas ROLAP obtêm mais vantagens por meio do processamento paralelo?** Isto é verdade para partes da aplicação que são processadas, usando SQL no RDBMS. Deixa de ser verdade, se há um servidor de aplicações separado, que não pode explorar as tecnologias de processamento paralelo (SMP ou MPP),

ou se os dados têm que ser movidos para a máquina do cliente a fim de realizar o processamento. Os desenvolvedores de ferramentas MOLAP também estão explorando esta tecnologia; em breve este recurso será incorporado à ferramenta.

- ⇒ **As ferramentas ROLAP podem executar cálculos multidimensionais completos?** A linguagem SQL tem uma funcionalidade limitada para realizar cálculos entre as dimensões. Porém, as ferramentas ROLAP superaram esta dificuldade gerando complexos múltiplos passos SQL. Assim, podem fornecer uma maior quantidade de funções de cálculos multidimensionais, mas ainda é menor que a quantidade que um MOLAP pode fornecer.
- ⇒ **As ferramentas ROLAP integram-se melhor com dados não numéricos?** A maioria das ferramentas MOLAP podem fazer esta tarefa tão bem como qualquer ferramenta ROLAP.
- ⇒ **Um MOLAP é necessário para análises simples?** Normalmente não são. Análises simples são relativamente fáceis de implementar, usando ferramentas ROLAP ou as ferramentas geradoras de relatórios que acompanham os bancos de dados relacionais, ou até mesmo as planilhas eletrônicas. Algumas análises básicas, como resumos de vendas, de produção e de qualidade não justificam o uso de uma ferramenta MOLAP. Como existem estas ferramentas mais baratas e populares, o uso destas sempre deveria ser considerado, antes de um realizar um alto investimento em uma solução OLAP (seja um MOLAP ou um ROLAP).
- ⇒ **Aplicações ROLAP estão disponíveis em mais plataformas de servidores?** Alguns produtos ROLAP estão disponíveis em mais plataformas que a maioria dos produtos MOLAP.
- ⇒ **Aplicações ROLAP oferecem um custo mais baixo que as MOLAP?** Pode parecer óbvio que o MOLAP deva valer mais, porque é necessário um banco de dados exclusivo, com taxa de licença e treinamentos adicionais. Porém, o pacote

de ferramentas ROLAP também é caro e também requer um bom treinamento. Devido ao fato de o MOLAP requerer menos espaço de armazenamento em disco, menos processamento, menos esforços de implementação e menos refinação, o custo global normalmente é mais baixo, mesmo que a taxa de licença do banco de dados seja mais alta e sejam necessários treinamentos adicionais.

4.5 - HOLAP - Hybrid OLAP

Atendendo as necessidades dos clientes e frente à concorrência dos desenvolvedores de soluções, os desenvolvedores de produtos OLAP estão tentando maximizar suas vantagens e minimizar suas desvantagens. Para realizar isto, muitos desenvolvedores passaram a utilizar um sistema híbrido que utiliza os dois tipos de OLAP, ou seja, o ROLAP e o MOLAP. A este novo tipo de OLAP deu-se o nome de HOLAP. Desta forma, os produtos ROLAP estão incorporando banco de dados multidimensionais, para poder oferecer aos seus clientes as vantagens das duas tecnologias. Os produtos MOLAP também estão incorporando aos seus produtos os bancos de dados relacionais, devido à sua facilidade de armazenamento e maior número de interfaces com outros produtos.

As ferramentas HOLAP são inteligentes e selecionam automaticamente a tecnologia mais adequada, de acordo com a atividade que será executada, proporcionando-lhe o máximo desempenho.

4.6 - WOLAP ou Web OLAP

Já existe a migração da tecnologia OLAP para o ambiente da Internet, a nova versão da ferramenta está sendo chamada de WebOLAP ou WOLAP (GERBER, 1998b).

Para CARICKHOFF (1997), as facilidades são: a possibilidade de plataformas independentes para dar suporte a usuários distantes, aplicações de groupware, facilidade de aprendizado e de manutenção. As dificuldades são: as limitações dos

recursos da Internet, as interfaces e as funcionalidades, quando comparados com o ambiente cliente/servidor. Segundo o autor, novas tecnologias para Internet, como Java e ActiveX estão tentando solucionar estas dificuldades. Durante a fase de transição, o uso misto das arquiteturas cliente/servidor e Internet são recomendados, até que as tecnologias envolvidas na Internet estejam totalmente aperfeiçoadas e confiáveis, e a migração para a mesma seja inevitável.

4.7 - Gerenciamento da ferramenta OLAP

Segundo a GENTIA SOFTWARE (1998), existem alguns pontos, nas ferramentas OLAP, que precisam ser gerenciados da melhor maneira possível:

- ⇒ **Customização** - deve ser possível proporcionar aos usuários um acesso simultâneo aos mesmos dados e que se possa usá-los e visualizá-los de maneiras diferentes. As interfaces customizadas devem ser colocadas no repositório de metadados de modo a facilitar a sua localização.
- ⇒ **Processamento distribuído** - os usuários de ferramentas OLAP não precisam ter acesso aos dados somente enquanto conectados ao servidor. É essencial que se possa carregar subconjuntos de dados para uso off-line, incluindo os notebooks.
- ⇒ **Facilidade de administração** - muitos negócios modernos estão sendo orientados por “workgroup”, para realizar projetos e tarefas. Deve ser fácil prover e manter os grupos de usuários.
- ⇒ **Portabilidade de plataforma** - as ferramentas OLAP não devem ser restringidas por qualquer recurso técnico, relativo ao suporte de sistemas operacionais diferentes ou interfaces de usuário. Aplicações construídas, em qualquer plataforma, devem poder ser executadas imediatamente em qualquer outra plataforma, sem perda de funcionalidade e realizando as adaptações necessárias.

No próximo capítulo será apresentada a terceira ferramenta da nova geração de Sistemas de Apoio à Decisão, o data mining.

Capítulo 5 - Data Mining

O capítulo 3 apresentou uma solução para o problema do armazenamento dos dados para propósitos gerenciais. O capítulo 4 apresentou uma solução para as necessidades de consultas sofisticadas e análises mais complexas. Estas duas ferramentas proporcionam às empresas um grande auxílio em seus processos decisórios, assim como no gerenciamento dos seus negócios.

Porém, uma das necessidades das empresas modernas é direcionar os seus negócios para os clientes. Isto é preconizado pelos programas de modernização das empresas como a Gestão pela Qualidade Total, proposta por autores como ISHIKAWA (1984), FEIGENBAUN (1991), GARVIN (1992), CLAUSING (1993), JURAN & GRYNA (1993), SUZAKI (1993) e GOESTCHI & DAVIS (1994) e a Reengenharia proposta por HAMMER (1994). As empresas já possuem, em seus bancos de dados, os históricos contendo as transações de seus clientes. Faz-se necessário ter uma ferramenta que estude esses dados e proporcione a oportunidade de se conhecer melhor os clientes, para que se possa direcionar os negócios para eles. O data mining foi desenvolvido com esta finalidade. Por meio desta ferramenta, é possível se realizar análises em um grande volume de dados armazenados, no qual pode-se descobrir, por exemplo, tendências e padrões de consumo. Essas novas informações podem ser utilizadas para se direcionar os negócios para os clientes.

5.1 - Conceitos de data mining

Segundo KIMBALL (1998f), as raízes do data mining estão nos anos 60, quando era chamado de análise estatística e que consistia em rotinas estatísticas

clássicas como correlação, regressão e outras. Nos anos 80, à análise estatística clássica foi acrescentado um conjunto maior de técnicas, como a Lógica Nebulosa (Fuzzy Logic), redes neurais, árvores de decisão e outras técnicas de Inteligência Artificial. Somente nos anos 90, foi possível aplicar-se as poderosas técnicas de Inteligência Artificial para auxiliarem na solução dos problemas das empresas.

MOXON (1998) afirma que o data mining é um conjunto de técnicas automáticas, usadas para explorar exaustivamente e descobrir relacionamentos complexos em um grande conjunto de dados. O conjunto de dados, na maioria das vezes, está armazenado em um banco de dados em forma de tabelas. Porém, segundo o autor, as técnicas também podem ser aplicadas a outras representações de dados, como banco de dados multidimensionais, textos e ambiente de multimídia. HODEL (1998) complementa afirmando que o data mining é o processo de extrair informação significativa e previamente desconhecida, pertencente a um grande banco de dados, e usar aquela informação para tomar decisões importantes na empresa.

O termo “data mining” (também conhecido como “mineração de dados” ou ainda “garimpagem de dados”) descreve uma variedade de ferramentas que processam dados e geram estratégias que aumentam a utilidade dos dados armazenados em bancos de dados corporativos (DATASAGE, 1998). Segundo a publicação, atualmente, enquanto não há uma definição universalmente aceita, o termo é amplamente usado sempre que um processo tenta descobrir informações escondidas em um banco de dados.

SULAIMAN & SOUZA (1997 e 1998) afirmam que a busca automatizada de conhecimento, em bancos de dados, é chamada de Knowledge Discovery in Databases (KDD - Prospecção de Conhecimento em Bancos de Dados). Segundo os autores, a expressão garimpagem de dados é comumente usada por estatísticos, analistas de sistemas e pela comunidade de Tecnologia da Informação, enquanto KDD é mais usada pelos pesquisadores de inteligência artificial.

Segundo EDELSTEIN (1997a) e DATAMIND TECHNOLOGY CENTER (1998), a ferramenta lê automaticamente os dados armazenados e descobre relacionamentos importantes, gerando informações que podem não ser intuitivas ou óbvias e permitem aos seus usuários tomarem melhores decisões. Estas informações são apresentadas em um modelo que pode ser visualizado em um gráfico, uma planilha eletrônica ou um arquivo de texto. O data mining pode ser usado com uma boa variedade de fontes de dados, incluindo os bancos de dados dos aplicativos operacionais e os Sistemas de Apoio à Decisão específicos. Porém, a ferramenta é muito mais efetiva quando combinada com um data warehouse, onde os dados já estão limpos e consistentes. O uso de ambas ferramentas habilita descobertas mais abrangentes e precisas.

Para FREEDMAN et al. (1998), encontrar novos clientes é uma tarefa difícil e cara. Se, em contrapartida, forem usados os recursos do data mining para se conhecer melhor os hábitos dos clientes atuais, pode tornar-se mais fácil a tarefa de descobrir-se quais deles são usuários potenciais para a aquisição de um novo produto e, ainda, elaborar-se um perfil de potenciais compradores deste novo produto entre os não clientes da empresa. As informações levantadas, com o uso desta ferramenta, podem revelar não só os bons clientes, como também quais consumidores estão propensos a deixar de ser clientes da empresa, isto antes mesmo de o fato acontecer, possibilitando que os contatos necessários possam serem feitos antes da perda dos referidos clientes.

Para MENCONI (1998), com base nos dados a respeito dos hábitos de consumo dos clientes, seus hobbies e nos dados sobre suas transações comerciais e financeiras, é possível se traçar associações que revelem grandes filões de mercado.

Para NIMER & SPANDRI (1998), os produtos de data mining estão melhorando em termos de qualidade, mas ainda têm sérios problemas de performance, velocidade de processamento e grau de sofisticação.

Existe uma hierarquia de complexidade nos dados armazenados (SULAIMAN & SOUZA, 1997). Simples dados que se encontram armazenados e que não transmitem por si só qualquer informação, na escala hierárquica pertencem ao primeiro nível, são os **dados**. Por exemplo, alguns números armazenados, estes por si só não podem ter qualquer significado. Porém, se for atribuído um significado a estes números, colocando-os dentro de um contexto, estes dados passam para o segundo nível na hierarquia, o das **informações** ou dos **fatos**. No exemplo, se afirmar-se que aqueles números representam valores que são investimentos dos clientes, esses dados passam a ter um contexto e se transformam em informações. A missão dos Sistemas de Informações é justamente transformar os dados em informações, através da geração de consultas, relatórios e outros tipos de acessos aos dados.

Se, através dos dados armazenados, conseguir-se elaborar uma norma ou uma regra, esses dados assumem o terceiro nível na hierarquia, a **norma** ou **regra**. No exemplo anterior, se forem deduzidos que alguns dos valores acima são considerados bons investimentos, aí obteve-se uma regra. Poderia ser descoberto que um grande volume de pequenos investidores traz melhores resultados que poucos investidores com grandes valores depositados. A regra deduzida seria: os investidores que realizam depósitos com a média de um valor X são mais rentáveis que outros que realizam depósitos com a média Y. A identificação dos clientes que são considerados bons investidores constitui o quarto e último nível da hierarquia, o **conhecimento**. Somente através da tecnologia do data mining é possível se chegar aos terceiro e quarto níveis hierárquicos da complexidade dos dados.

Em data mining, fala-se freqüentemente em encontrar **padrões**, **regras** e **fatos** nos dados armazenados (AZMY, 1998). E o que são, respectivamente, padrões, regras e fatos? Em uma tabela contendo dados, como no exemplo da Tabela 7, um **padrão** é definido como um conjunto de linhas que compartilham os mesmos valores com duas ou mais colunas.

Item	Tipo	Cor
1	carro	vermelho
2	carro	vermelho
3	carro	vermelho
4	carro	azul
5	moto	azul

Tabela 7 - Exemplo de tabela contendo dados

Na Tabela 7, temos 3 itens (1, 2 e 3) que compartilham os mesmos valores, em duas colunas (Tipo e Cor) e pode-se afirmar: “*A maioria das carros são da cor vermelha*”. O fator de confiança é 75% e é calculado dividindo-se o número de itens que possuem os dados Tipo = carro e Cor = vermelho (3 itens) pelo número de itens que têm Tipo = carro e todas as Cores (4 itens).

Assim, da mesma forma que se procuram os padrões, também podem-se procurar as **exceções**. Por exemplo, o item 4 da Tabela 7 representa uma exceção. Este item possui o único carro na cor azul, enquanto todos os outros carros são da cor vermelha. Esta exceção ficaria mais interessante, se houvessem mil carros na cor vermelha e somente um na cor azul. A descoberta das exceções também é importante para as empresas e para os negócios, pois através dela é possível se descobrir erros, fraudes ou mesmo exceções que possam ser interessantes para alguns tipos de negócios, como no ramo de modas ou qualquer ramo de negócios que trabalhe com inovações.

Para AZMY (1998), um **fato** é representado por um padrão com fator de confiança superior a 50%. Na Tabela 7 temos um fato: 75% dos carros são da cor vermelha. A partir deste fato pode ser deduzida uma **regra**: “*Se item = carro então cor = vermelho.*”

A estrutura básica da descoberta de conhecimento, a partir dos dados armazenados, possui a seqüência descrita acima. Primeiro, o data mining procura os

identificar os padrões, depois transforma-os em fatos e então deduzem-se regras que levarão à descoberta de novos conhecimentos.

O data mining, quando está procurando padrões, faz uma análise do nível de agrupamento de cada coluna, ou seja, faz uma busca, para formar conjuntos de itens contendo os mesmos dados, dentro da mesma coluna. A ferramenta elimina, do processo de busca, as colunas que não formam agrupamentos significativos. O próximo passo é fazer os relacionamentos entre as colunas em busca de se encontrar os padrões. É possível se realizar a configuração necessária, determinando o nível de relacionamento entre as colunas, por exemplo, podemos configurar o sistema para descobrir padrões que ocorram em um nível superior a 5%; neste caso os que representem um valor inferior ao determinado serão desprezados. Um bom sistema de mineração de dados deve permitir que se eliminem do processo de descoberta os relacionamentos óbvios, como relacionar as cidades onde residem os clientes com os seus respectivos Estados e CEPs. O Anexo A contém algumas telas de um aplicativo de data mining com suas configurações e aplicações.

A tecnologia utilizada no data mining pode fazer uma procura em grandes quantidades de dados armazenados e informar somente padrões e relacionamentos que podem ser interessantes para os negócios. AZMY (1998) afirma que existem outras ferramentas, que também permitem a descoberta de conhecimento em um banco de dados, como as Análises Estatísticas, as ferramentas OLAP, a Visualização de Dados e alguns algoritmos de Inteligência Artificial, porém, ao contrário destas ferramentas, a mineração de dados não exige que um usuário faça perguntas específicas, o processo de descoberta do conhecimento é realizado automaticamente. O autor afirma, ainda, que a vantagem do data mining é que encontrar padrões interessantes é mais fácil e possui um custo menor, se comparado com as outras ferramentas.

Porém, para MENCONI (1998), o data mining está longe de ser a resposta para todos os males do mundo corporativo; é uma solução cara e exige informações muito bem organizadas para funcionar bem, caso contrário, pode até agravar a

situação de uma empresa em dificuldades. SMALL (1998) complementa, afirmando que os resultados obtidos, através do data mining, melhoram os negócios em uma organização já próspera, proporcionando as adaptações necessárias para se adequar às informações que foram descobertas. A ferramenta não proporciona necessariamente mudanças revolucionárias que poderiam até vir a recuperar empresas à beira da falência. Segundo os autores, o data mining é uma poderosa ferramenta para empresas com uma visão a longo prazo.

A facilidade de uso é um fator crítico para a aceitação de qualquer ferramenta de Tecnologia da Informação pelos profissionais de negócios. O data mining oferece algumas funções muito sofisticadas, porém a tecnologia utilizada está embutida no software, e os usuários, ao utilizá-lo, não precisam ser “experts” em técnicas estatísticas avançadas ou em Inteligência Artificial. Pelo contrário, os usuários podem trabalhar com formatos simples e familiares, como telas e relatórios pré-modelados, ou ainda exportar os resultados para planilhas eletrônicas, processadores de textos ou para outras ferramentas de apoio à decisão.

SMALL (1998) afirma que nenhuma técnica de análise pode substituir a experiência e o conhecimento nos negócios e seus mercados, pelo contrário, a mineração de dados torna este conhecimento e esta experiência mais importantes que nunca. Através do data mining, passa-se a conhecer melhor os negócios e os clientes e podem-se criar melhores estratégias para gerenciá-los.

5.2 - Tipos de modelos construídos pelo data mining

Quando se for adquirir um produto de data mining, a primeira coisa a se avaliar é saber quais modelos ele constrói e se esses modelos podem dar o suporte necessário ao processo decisório e ao gerenciamento dos negócios. Analisando os conceitos apresentados por ADRIAANS & ZANTINGE (1996), FAYYAD et al. (1996), EDELSTEIN (1997a e 1997b), SIMOUDIS (1997), GERBER (1998a), KIMBALL (1998f), MOXON (1998) e TKACH (1998), levantou-se os principais

modelos que os sistemas de data mining podem construir, para realizar suas descobertas.

- ⇒ **Modelo de agrupamento.** Segmenta um conjunto de dados em grupos diferentes, cujos itens são semelhantes. Usando o agrupamento, a ferramenta de mineração de dados descobre grupos diferentes dentro dos dados. Isto pode ser aplicado a problemas tão diversos, quanto detectar defeitos de fabricação ou identificar grupos com afinidades, para adquirirem novos cartões de crédito;

- ⇒ **Modelo de classificação.** Reconhece o padrão que descreve o agrupamento ao qual um determinado item pertence. Faz isto examinando os itens existentes que já foram agrupados e deduzindo regras daquele padrão. A classificação pode ajudar a descobrir, por exemplo, as características de clientes potenciais, fiéis ou problemáticos. Também pode ajudar a determinar que tipo de promoção é mais adequada para determinada categoria de clientes, de forma que seja feito, somente, o investimento necessário para aquela categoria de clientes;

- ⇒ **Modelo de descoberta de associação.** Descobre regras a partir dos padrões. Por exemplo, o modelo poderia descobrir que, quando qualquer cliente compra um item **X**, em **Z%** das vezes, ele compra também o item **Y**. Uma aplicação típica é a análise de vendas no varejo (também conhecida como análise de cestas de mercado), onde o modelo é usado para identificar as afinidades entre produtos adquiridos. Outras aplicações são: identificar os relacionamentos entre procedimentos médicos e os sintomas dos pacientes; analisar as operações e aplicações financeiras dos clientes, para identificar os seus relacionamentos; e analisar o resultado de promoções de vendas conjuntas;

- ⇒ **Modelo de descoberta de seqüência.** É semelhante ao modelo anterior, exceto que o relacionamento entre os itens são seqüenciais no tempo. Por exemplo, se um evento **X** ocorre, em **Z%** das vezes, então ocorre um evento **Y** dentro de **W** dias. Com este modelo, pode-se tirar proveito do tempo decorrido entre os eventos ocorrido e a ocorrer, que compõem a seqüência. Por exemplo, se um conjunto de

uma mesa com quatro cadeiras é comprado, em 40% das vezes, um fogão será comprado dentro de um mês e, em 20% das vezes, um refrigerador será comprado dentro de dois meses. Pode-se, então, providenciar para que não faltem esses produtos e um contato, devidamente planejado, com o cliente poderia ser feito;

- ⇒ **Modelo de regressão.** Usa uma série de valores existentes para prever outros valores. Valores como previsão de vendas, custos de armazenagem e taxas de falhas, que são difíceis de prever por outras técnicas, podem ser calculados por este modelo;
- ⇒ **Modelo de predição.** A predição é a habilidade para elaborar cenários diferentes, para antecipar certos resultados. Por exemplo, um gerente de banco poderia usar o modelo para prever se é provável que um cliente possa oferecer algum tipo de risco em uma concessão de crédito. Com base naquela predição, o gerente pode decidir aceitar ou rejeitar o crédito daquele cliente. Este modelo é apoiado pelos modelos de agrupamento, de descoberta de associações, de descoberta de seqüências, pela técnica de visualização e por detecção de divergências, usando técnicas estatísticas.

As atividades desenvolvidas pelo data mining, para realizar as descobertas e predições, combinam diversos modelos para produzir os resultados desejados, sem que o usuário precise selecionar quais serão os modelos necessários para realizar a atividade escolhida.

Segundo SWAMI (1998), o modelo de descoberta de associações tem muitas aplicações e é o modelo mais utilizado. Por exemplo, nas vendas a varejo, em que cada compra realizada consiste em um conjunto de itens adquiridos por um cliente, em uma única visita a um ponto de vendas, poderia ocorrer a seguinte de regra de associação: “*80% das pessoas que compram os produtos P e Q compram também o produto R.*” O valor 80% é o fator de confiança. Os produtos *P* e *Q* constituem o conjunto de itens, no lado esquerdo da regra (LHS - Left Hand Side), e o produto *R* é o item do lado direito da regra (RHS - Right Hand Side). Algumas aplicações para as

regras geradas pelo modelo de descoberta de associação são: o uso da regra que tem um produto

- ⇒ **X** no RHS para analisar quais produtos impulsionam a venda do produto **X**;
- ⇒ **Y** no LHS para analisar quais produtos sofreriam queda nas vendas, caso o produto **Y** deixasse de ser oferecido por qualquer motivo;
- ⇒ **X** no LHS e um produto **Y** no RHS para analisar quais produtos deveriam ser vendidos em conjunto com o produto **X**, para impulsionar as vendas do produto **Y**.

5.3 - As técnicas utilizadas pelo data mining

SMALL (1998) afirma que todas as técnicas utilizadas atualmente na mineração de dados são extensões naturais e generalizações de métodos analíticos já conhecidos há décadas. As redes neurais foram desenvolvidas nos anos quarenta, os modelos de classificação e de regressão foram usados nos anos sessenta. O que é novo na mineração de dados é que, somente recentemente, passou a ser possível se aplicar estas técnicas para se auxiliar no processo decisório e para encontrar novas estratégias para os negócios, graças ao aumento da disponibilidade de dados históricos e à queda nos custos do armazenamento e do processamento de dados. O recente interesse pela mineração de dados ocorreu, principalmente, devido às melhores interfaces com o usuário.

O data mining utiliza diversas técnicas, que detectam padrões automaticamente, adicionando inteligência à análise dos dados e tornando-a independente do usuário (CAMPOS, 1997b). Baseado em ADRIAANS & ZANTINGE (1996), FAYYAD et al. (1996), EDELSTEIN (1997a), BRAND & GERRITSEN (1998), FREEDMAN et al. (1998), KIMBALL (1998f), MOXON (1998) e PILOT SOFTWARE (1998), descreve-se a seguir as principais técnicas usadas no data mining.

1. **Lógica Nebulosa (Fuzzy Logic)**. Esta técnica é utilizada para capturar informações vagas, em geral descritas na sua forma natural, e convertê-las para

um formato numérico, para facilitar as suas análises. Esta técnica trabalha com a teoria dos conjuntos nebulosos (Fuzzy Set) que tem mostrado ser muito apropriada para se trabalhar com vários tipos de dados e informações, muitas vezes superando os resultados obtidos pelas técnicas estatísticas e probabilísticas tradicionais. Por exemplo, na frase: “Se o tempo de um investimento é longo e o sistema financeiro não tem sido muito estável, então a taxa de risco do investimento é muito alta”. Nesta frase, os termos “longo”, “não muito estável” e “muito alta” são informações em sua forma natural e são termos muito vagos para serem analisados. A representação numérica destas informações vagas pode ser feita através dos conjuntos nebulosos e as análises realizadas através da lógica nebulosa.

- 2. Redes neurais artificiais.** Mais conhecidas como redes neurais, trabalham de forma semelhante às redes neurais biológicas. Em sua estrutura, possui uma coleção de nós conectados (Figura 19). Nesta coleção, existe uma camada de nós de entrada, outra camada de nós de saída, podendo ter ainda uma ou várias camadas de nós intermediários. A camada de nós de saída e a(s) camada(s) de nós intermediário(s) (se houver) realizam o processamento dos dados da entrada proporcionando um resultado na saída. Uma rede neural artificial é capaz de “aprender” a executar suas tarefas por meio de um treinamento apropriado, conhecido como treinamento supervisionado. Em cada passo do treinamento, o conjunto de dados de saída é comparado com um resultado já conhecido e, se for diferente, uma correção é calculada e processada automaticamente nos nós que realizam o processamento na rede. Esses passos serão repetidos até que uma condição de parada seja atingida, ou um número de correções previamente estipulado seja alcançado. As redes neurais artificiais trabalham somente com números e os demais tipos de dados precisam ser convertidos em números para a sua utilização. Os processamentos gerados por esta técnica são considerados opacos, ou seja, os resultados são obtidos sem se interpretar claramente como a técnica chegou aos mesmos. Essa falta de explicações inibe a confiança, a aceitação e a aplicação dos resultados. Alguns produtos de redes neurais artificiais possuem algoritmos que podem traduzir o modelo utilizado pela técnica, em um

conjunto de regras que ajudam a compreender como se chegou aos resultados apresentados. Dependendo da aplicação, a técnica pode consumir um tempo muito longo de aprendizagem, porém, depois de realizada a aprendizagem, ela pode produzir resultados mais rápidos que outras técnicas. As redes neurais artificiais são muito utilizadas em atividades que envolvem classificação dos dados.

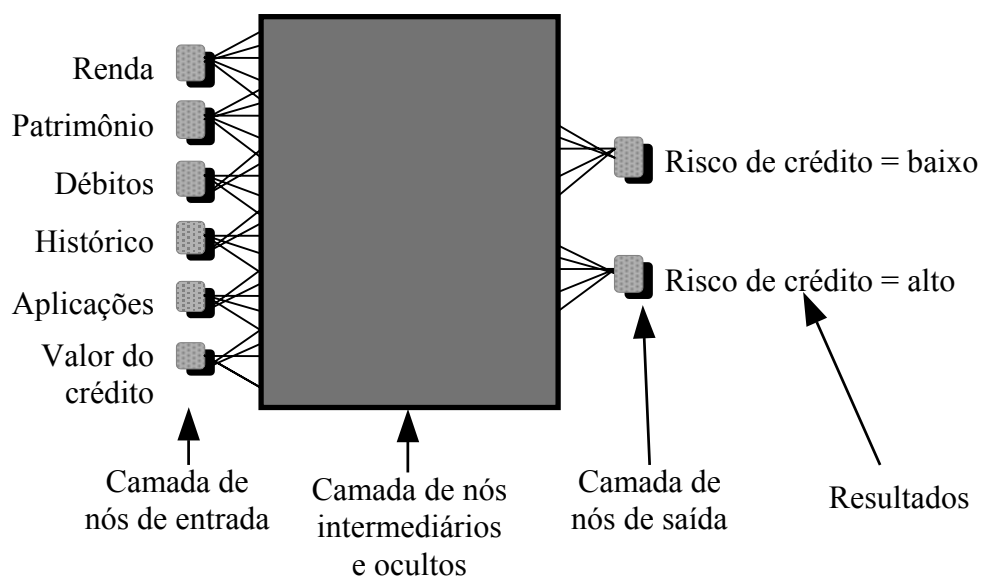


Figura 19 - Diagrama de uma Rede Neural Artificial

3. **Árvores de decisão.** Possui este nome porque a sua estrutura se assemelha a uma árvore. A sua estrutura é muito fácil de entender e de assimilar. Dividem os dados em subgrupos, com base nos valores das variáveis. O resultado é uma hierarquia de declarações tipo “Se ... então ...” que são utilizadas, principalmente, para classificar dados. Houve uma onda de interesse em produtos baseados em árvores de decisão, principalmente porque eles são mais fáceis de se compreender o seu funcionamento e como são obtidos os resultados. Porém, esta técnica, apesar de ser mais versátil que as redes neurais artificiais, também é restrita a alguns tipos de análises.

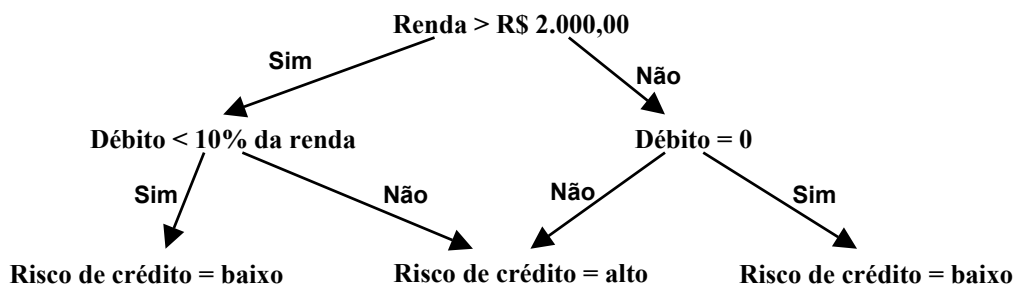


Figura 20 - Diagrama de uma Árvore de Decisão

4. **Algoritmos genéticos.** É uma técnica de otimização que usa processos como combinação genética, mutação e seleção natural, baseando-se nos conceitos de evolução.
5. **Regras de indução.** Esta técnica cria conjuntos hierárquicos ou não hierárquicos de condições (regras) que conduzem a um certo resultado. Pode analisar, entre diversas alternativas, qual delas pode apresentar o melhor resultado. Esta técnica, também, é muito utilizada na construção do modelo de predição.
6. **Nearest neighbor.** É uma técnica usada no modelo de classificação. Tenta buscar os k vizinhos mais próximos. Em outras técnicas preditivas, os dados são processados para se criar o modelo da técnica. Ao invés, na técnica do Nearest Neighbor o treinamento dos dados constituem a própria técnica. Quando um novo dado é fornecido à técnica, todos os dados são usados para se encontrar um subconjunto de dados que são muito semelhante ao novo dado, ou seja, a técnica usa os próprios dados para identificar os casos de semelhantes. Esta técnica possui algumas desvantagens, como os resultados são fornecidos sem a descrição de como se chegou a eles, além das implicações de se ter que comparar um novo dado com todos os demais que já estão armazenados. Em alguns softwares de data mining esta técnica é utilizada em conjunto com outras técnicas, para se obter um modelo de classificação mais eficaz.
7. **Visualização.** Alguns autores a consideram como modelo e outros como uma técnica embutida nos aplicativos de mineração de dados. Dada suas características, prefere-se considerá-la uma técnica. Apesar de não automatizar nenhuma atividade, ela possibilita ao usuário obter uma compreensão melhor e

mais intuitiva dos dados e dos resultados apresentados. Foi esta técnica que impulsionou o uso das ferramentas de mineração de dados.

Outras boas referências bibliográficas sobre as técnicas utilizadas pelo data mining são: WU (1995), BIGUS (1996), BERRY & LINOFF (1997), WEISS & INDURKHYA (1997), AMAND & BUECHNER (1998), FAYYAD & GEIGER (1998) e WESTPHAL & BLAXTON (1998).

Normalmente, os produtos para mineração de dados combinam as diversas técnicas, para se construir um produto mais preciso e mais rápido (EDELSTEIN, 1997a).

5.4 - Planejamento e gerenciamento do data mining

MENCONI (1998) menciona que um bom projeto de implantação de data mining implica a compra de software, a contratação de serviços de consultoria e, muitas vezes, uma atualização dos equipamentos. Pequenas e médias empresas podem gastar de 50.000 a 500.000 dólares na implantação do projeto. Em grandes conglomerados, esse valor pode saltar para 2 milhões de dólares.

Segundo EDELSTEIN (1997b), algumas ferramentas de mineração de dados exigem que o formato dos dados, do banco de dados fonte, seja do mesmo tipo dos dados de seus arquivos internos, outras já podem trabalhar diretamente com o banco de dados fonte e a ferramenta se encarregará de fazer as adaptações necessárias.

SMALL (1998) afirma que, embora as ferramentas para mineração de dados possam explorar os dados e descobrir seus relacionamentos ocultos, é preciso ser bem específico na sua utilização para que se possa obter bons resultados. Simplesmente entregar um banco de dados, como fonte de dados, para uma ferramenta de data mining e esperar, por exemplo, que a ferramenta encontre perfis de clientes que melhorem a eficiência de uma campanha de mala direta, não é um uso tão eficaz da ferramenta. Faz-se necessário ser mais específico nas metas: para

melhorar o retorno financeiro de uma mala direta, a meta poderia ser encontrar os clientes que compraram os artigos mais caros anteriormente. Por outro lado, para se aumentar o número de respostas da mala direta, a meta poderia ser encontrar os clientes que responderam às correspondências anteriores. Ter-se uma boa ferramenta e não saber como usá-la eficientemente pode ser até perigoso para a empresa.

Quanto maior o volume de dados, que possa ser processado pelo data mining, maior poderá ser a confiança nos resultados obtidos (SWAMI, 1998); porém, uma maior capacidade de processamento será necessária e custos mais elevados também terão que ser considerados. SMALL (1998) afirma que uma maior quantidade de dados a respeito dos itens são úteis somente quando eles podem contribuir, gerando mais informações durante processo de descoberta de padrões. Por exemplo: uma empresa administradora de cartões de crédito pode ter muitas informações sobre o uso do seus cartões de crédito, pelos seus clientes, e ter também informações sobre quais são os outros cartões de créditos que os seus clientes possuem, mas nenhuma informação sobre o uso dos outros cartões de créditos dos clientes (os cartões das empresas concorrentes). Neste exemplo, os dados armazenados não proporcionaram novas informações ou novos conhecimentos e podem ser considerados dados sem qualidade. Adicionar muitos dados, mas que forneçam poucas informações, diminui a eficácia e o desempenho da mineração de dados.

Para SMALL (1998), muitos dos modelos usados na mineração de dados foram desenvolvidos para analisar grandes volumes de dados. Mas um banco de dados de tamanho moderado ou pequeno também pode revelar valiosas informações. Às vezes, apesar de termos um volumoso banco de dados, não é muito prático trabalharmos com todos os seus dados, é mais produtivo trabalharmos com amostras que podem representar perfeitamente todo o conjunto de dados.

5.5 - Aplicações

Para o DATAMIND TECHNOLOGY CENTER (1998), três áreas-chave em que se espera que o data mining represente um papel importante nos negócios são:

1. **Marketing.** Algumas empresas gastam milhões de dólares para localizarem os clientes potenciais, por meio de campanhas de marketing, como anúncios e correspondências enviadas aos clientes. As empresas podem aperfeiçoar suas campanhas de marketing para obterem um maior retorno de investimento, usando o data mining para definir, entre os seus clientes, quais são o mais prováveis a adquirir um novo produto ou serviço. Por exemplo, uma campanha de mala direta pode ser enviada somente a 10.000 clientes qualificados, ao invés de se enviar a um total de 100.000 clientes, sem uma análise prévia.
2. **Atendimento ao consumidor.** Frequentemente, o único fator que diferencia os fornecedores é a qualidade no atendimento ao consumidor. Entender melhor as exigências dos clientes e ajudar as empresas a desenvolverem um serviço mais atraente é uma das funções do data mining.
3. **Descoberta de fraude.** As empresas perdem uma parte significativa da rentabilidade, através de práticas fraudulentas. As habilidades automatizadas de reconhecimento de padrões e exceções do data mining provêm uma alternativa poderosa para a descoberta de fraudes.

O próximo capítulo será dedicado a mostrar alguns casos práticos de empresas que implementaram as três tecnologias e os resultados alcançados pelas mesmas.

Capítulo 6 - Exemplos de empresas que implantaram as ferramentas e os resultados obtidos

Nos capítulos 3, 4 e 5 foram apresentadas as três ferramentas da nova geração de Sistemas de Apoio à Decisão. Explicou-se o que são estas tecnologias, como funcionam, as tecnologias envolvidas, como cada uma destas ferramentas pode dar a sua contribuição ao processo decisório das empresas e, também, como elas estão sendo utilizadas para transformar os dados armazenados em informações ou novos conhecimentos que podem servir para se criarem novas estratégias para os negócios.

Neste capítulo, serão apresentados alguns exemplos de empresas que implantaram as ferramentas e os resultados obtidos, coletados na literatura disponível e, através da análise destes exemplos, será possível obter-se uma visão da utilidade e das aplicações das três ferramentas.

O exemplo de mineração de dados mais divulgado é o de uma cadeia americana de lojas, o Wall-Mart. Ao procurar eventuais relações entre o volume de vendas e os dias da semana, o software de mineração de dados apontou que, às sextas-feiras, as vendas de cervejas cresciam na mesma proporção que as de fraldas. Em uma análise mais detalhada, usando a mesma ferramenta, revelou-se que, ao comprar fraldas para seus bebês, os pais aproveitavam para abastecer o estoque de cerveja para o final de semana. A providência tomada, para confirmar ou refutar a hipótese, foi a de colocar-se os dois produtos lado a lado. A hipótese foi confirmada com um crescimento nas vendas dos dois produtos.

No Brasil, MENCONI (1998) relata o exemplo das Lojas Brasileiras. A empresa utilizou ferramentas para mineração e análise dos dados, para poder conhecer melhor o seu negócio. Através destas ferramentas, a empresa conseguiu reduzir a quantidade de produtos expostos em suas prateleiras de 51.000 para apenas 14.000. Passou a oferecer aos clientes de cada região do país apenas os produtos mais consumidos e típicos daquela região. Desta forma, foi possível uma economia com compras desnecessárias, um melhor posicionamento dos produtos nas prateleiras, economia com estoques, pessoal, custos operacionais e ainda um atendimento mais personalizado aos clientes de cada região do país.

Outro exemplo relatado por MENCONI (1998) é do Bank of America. Este banco usou uma ferramenta de mineração de dados para descobrir novos negócios no mercado financeiro. Para concretizar um desses novos negócios, o banco enviou cartas, oferecendo linhas de crédito especiais para os bons clientes cujos filhos estivessem com a idade entre 18 e 21 anos e, portanto, precisassem de dinheiro para ajudar os filhos a comprar o próprio carro, uma casa ou arcar com os custos da faculdade. Segundo o autor, o resultado obtido em três anos foi que o banco lucrou 30 milhões de dólares.

A CAIXA ECONÔMICA FEDERAL (1998) está iniciando um projeto de implantação de um data warehouse. A empresa definiu, como prioridade, a construção de bases de dados estratégicas que possam servir de apoio aos seus executivos nos processos decisórios. Nesse processo, estão sendo consideradas as dificuldades hoje existentes na recuperação de dados históricos, a falta de padronização, a integração dos diversos sistemas e a existência de informações dispersas geograficamente. A implantação facilitará a execução de vários projetos na empresa, uma vez que permitirá o acesso rápido às informações. Dentre eles está o projeto de segmentação comportamental dos clientes, que dará início ao processo de marketing de relacionamento, possibilitando que a empresa conheça e acompanhe o cliente e considere o seu valor para a empresa e suas necessidades por produtos e serviços. Os resultados deste trabalho vão subsidiar o planejamento estratégico, a distribuição e comercialização de produtos, assim como as campanhas de marketing.

A IBM (1998) apresenta um exemplo em que apenas um data warehouse e as ferramentas de consultas amigáveis foram o suficiente para atender as necessidades da empresa Aetna U.S. Healthcare. Esta empresa administradora de planos de saúde possui 14 milhões de clientes em 50 estados dos EUA. Ela administra hospitais, médicos, fabricantes de medicamentos e produtos hospitalares, entre outros. Seu grande problema era prover seus administradores com informações rápidas, precisas e de um modo prático, como se faz necessário em um ambiente hospitalar. A implantação de um data warehouse, atualizado constantemente, conseguiu resolver o problema de informações gerenciais, sem prejudicar as informações operacionais.

A IBM (1998) relata outro exemplo de empresa do ramo de planos de saúde, a Blue Cross & Blue Shield. Esta empresa também enfrentava problemas para administrar um grande volume de dados e para fazer as análises avançadas necessárias, principalmente com relação aos custos dos tratamentos em relação aos valores pagos pelos clientes. A solução foi implantar um data warehouse e uma ferramenta OLAP. A partir daí, os profissionais do setor financeiro da empresa puderam produzir melhores relatórios analíticos, simular cenários, analisar melhor os custos e as mensalidades. Desta forma, foi possível se conhecer melhor os negócios da empresa e obter um melhor controle sob o mesmo. Somente após a implantação das ferramentas, foi possível um crescimento planejado e controlado pela empresa.

Segundo a IBM (1998), na CIGNA, outra empresa do ramo de planos de saúde, o maior problema são as fraudes. A empresa estima que bilhões de dólares são perdidos anualmente somente em fraudes. Estas fraudes ocorrem em falsificação de documentos e práticas de super-faturamento, praticadas por médicos, clínicas e hospitais. A descoberta, prevenção e recuperação, mesmo de uma pequena porcentagem destas irregularidades, pode significar uma economia de milhões de dólares e a cobrança de um preço mais justo nas mensalidades dos associados. A descoberta de fraudes é uma tarefa que está ficando cada mais difícil, os fraudadores estão desenvolvendo técnicas cada vez mais sofisticadas e variadas para cometerem as suas irregularidades. Como a quantidade de dados é muito grande, a análise e a

checagem da validade de cada dado sobre os tratamentos é praticamente impossível. Por meio de uma ferramenta de data mining, foi possível se levantar o perfil dos diversos tipos de tratamento, do atendimento e dos conveniados. Através do levantamento dos dados daqueles que realizaram tratamentos que fugiram aos perfis especificados, foi possível uma investigação mais direcionada e mais eficaz.

A Empire Blue Cross and Blue Shields, também do setor de planos de saúde, estava obtendo crescentes prejuízos com fraudes, segundo a IBM (1998). A empresa possui 4,5 milhões de associados e administra cerca de 50 milhões transações por ano, entre consultas e tratamentos. A situação chegou ao ponto de ser necessário se criar um setor específico para se detectar e tomar as providências para tentar eliminar as fraudes. Após a implantação de um data mining e um longo processo de investigação, diversos fraudadores foram identificados, processados, presos e tiveram que restituir o dinheiro comprovadamente desviado dos cofres da empresa.

As empresas do setor de telecomunicações estão necessitando usar o ambiente de data warehouse, OLAP e data mining para adquirirem novos clientes, reterem os atuais clientes, direcionarem melhor suas campanhas de marketing e desenvolverem serviços mais personalizados (IBM, 1998). Somente desta forma está sendo possível enfrentar a acirrada concorrência neste setor, nos EUA e Canadá. A combinação de acesso rápido e confiável aos dados (data warehouse), análises sofisticadas dos dados (OLAP) e descoberta de perfis e características dos clientes (data mining) ajudam os gerentes a obter um cenário bem claro dos seus negócios, quem são seus clientes, o que eles querem, seus hábitos e costumes, suas necessidades e o que fazer para satisfazê-los. Desta forma, estas empresas estão se tornando mais eficientes e deixando seus clientes mais satisfeitos.

Bons resultados também foram obtidos pela empresa farmacêutica Merck-Medco (MCCARTHY, 1997). A empresa utilizou um software de data mining para descobrir os vínculos existentes entre as enfermidades e os tratamentos realizados pelos seus clientes. Assim, conseguiu definir quais remédios são mais efetivos para

cada paciente. Os resultados alcançados foram tratamentos mais efetivos e com um menor custo para os pacientes (economia de 15% em média).

O setor bancário também é um dos grandes usuários das novas ferramentas. A IBM (1998) relata o caso do AIB Bank, um banco irlandês com um patrimônio de 30 bilhões de dólares. O banco tinha dificuldades para descobrir os hábitos e para prever o comportamento dos clientes em determinadas situações. Para resolver este problema, o banco resolveu adotar uma ferramenta de data mining. Assim, foi possível analisar os seus bilhões de registros de transações armazenados e segmentar seus clientes em classes. Ao analisar o comportamento dos clientes dessas classes, foi possível conhecê-los melhor e construir-se um relacionamento mais duradouro, através da criação de produtos e serviços mais personalizados e de acordo com cada categoria.

A IBM (1998) relata o exemplo da empresa CIGNA, do ramo de seguros. Após a implantação de um data warehouse, a empresa conseguiu administrar melhor os seus negócios. Anteriormente, os executivos analisavam os negócios realizados através de relatórios padronizados que continham os dados das transações, porém, sempre que se desejava fazer outros tipos de análises, que dependessem de outro tipo de levantamento de dados, era uma tarefa complexa e desestimulante. Com a implementação do data warehouse, o levantamento de dados para as pesquisas e análises se tornou extremamente fácil e versátil. Desta forma, foi possível administrar melhor os riscos deste tipo de negócio, fazer análises mais sofisticadas nos custos e na rentabilidade e planejar melhor o crescimento da empresa.

Outro exemplo relatado pela IBM (1998) é o da IMPCO, uma das maiores empresa de marketing do mundo. A empresa mudou completamente a forma com que praticava o seu negócio, investindo em equipamentos e programas para mineração de dados, o que lhe deu um novo impulso nos negócios. Até mudou o nome do negócio de “marketing direto” para “marketing estratégico”. Após a implantação do data mining, a empresa passou a solicitar aos clientes que lhe fornecessem não só os dados básicos dos clientes, mas também as suas transações,

para que se pudesse fazer uma análise dos hábitos para conhecer melhor os clientes alvos das campanhas de marketing. Através desta prática foi possível realizar melhores campanhas de marketing, direcionadas a uma faixa específica de clientes alvo, que proporcionam um melhor retorno e a um custo bem inferior.

Não são somente as empresas do setor privado que estão obtendo bons resultados com as novas ferramentas, o setor público também está trabalhando com as mesmas e chegando aos mesmos resultados que o setor privado. Os governos, em todos os níveis, e os seus órgãos estão enfrentado um problema semelhante ao enfrentado pelas empresas do setor privado, ou seja, melhorar a qualidade de seus produtos ou serviços, aumentar a arrecadação, diminuir o déficit orçamentário, reduzir as práticas fraudulentas e abusivas, melhorar a administração dos seus recursos disponíveis, seguir fielmente seus programas de governo e alcançar as metas planejadas.

Muitos órgãos do governo, como bancos e prestadores de serviços de telecomunicações, são concorrentes diretos de empresas do setor privado, portanto, para poder competir no mesmo nível, estes órgãos estão adotando as mesmas soluções que os concorrentes do setor privado. Existem alguns problemas que são exclusivos dos órgãos de governos, como a arrecadação de impostos. Neste caso, o grande problema é a sonegação. Segundo a IBM (1998), para solucionar este problema diversos governos federais (inclusive no Brasil), estaduais (principalmente nos estados dos EUA) e de algumas cidades americanas estão adotando o data mining para o auxílio de detecção de sonegação de impostos, através da segmentação dos contribuintes em classes distintas, a análise de seus bens e a evolução do patrimônio, construindo, desta forma, um perfil das diversas classes de contribuintes. Isto facilita muito o processo de investigação de sonegação.

Até mesmo nos esportes já está sendo utilizada a ferramenta de mineração de dados, para auxiliar a melhorá-los. Segundo BHANDARI et al. (1997) e IBM (1998), um software de data mining chamado Advanced Scout (AS) foi desenvolvido especialmente para a NBA (National Basketball Association) e foi distribuído para as

suas equipes, para que estas possam utilizá-lo e possam tirar o melhor proveito do software. Cada equipe utiliza o produto para analisar os jogos dos adversários, tentar compreender suas jogadas e se preparar melhor para enfrentá-los. As partidas são minuciosamente descritas, jogada a jogada, não somente os lances ocorridos, mas também as alternativas para cada jogada. Estas descrições são convertidas em dados numéricos e depois são submetidas os software AS, para que o mesmo descubra os perfis dos jogadores, das jogadas e da equipe adversária. Desta forma, a NBA pretende melhorar o nível das equipes e também do próprio campeonato, tornando-o mais interessante e mais disputado, obtendo um público maior nos estádios, além de, evidentemente, um retorno financeiro maior. Todos saem ganhando, o público assiste a melhores jogos, as equipes e a NBA aumentam o retorno financeiro.

Com os exemplos apresentados neste capítulo, foi possível ter-se uma compreensão melhor das finalidades, possibilidades, aplicações e resultados que podem ser obtidos com o uso das novas ferramentas.

No próximo capítulo, serão mostradas as tendências que as três ferramentas terão nos próximos anos.

Capítulo 7 - Tendências

Nos capítulos 3, 4 e 5, foram apresentadas as três ferramentas que constituem a nova geração de Sistemas de Apoio à Decisão. No capítulo anterior, foram apresentados alguns exemplos de empresas que implantaram as ferramentas e os resultados obtidos, para que fosse possível ter uma compreensão melhor das finalidades, possibilidades, aplicações e resultados que se pode obter com o uso das novas ferramentas.

As três ferramentas fazem parte também das novas Tecnologias da Informação. Estas tecnologias estão evoluindo constantemente. Primeiro, para se tentar atender cada vez mais as necessidades dos seus usuários. Segundo, porque existem diversos desenvolvedores destas ferramentas que são concorrentes entre si e uma das formas de se obter um crescimento das vendas destes produtos é proporcionar o seu aperfeiçoamento. Terceiro, porque, como já se afirmou anteriormente, as três ferramentas ainda estão em fase de evolução e ainda há muito a ser feito. Neste capítulo, serão apresentadas as tendências que terão as novas ferramentas nos próximos anos.

Segundo CARVALHO (1997a), uma das tendências para o data warehouse e o data mining, até então exclusivos do ambiente cliente/servidor, é a migração também para o ambiente da Internet, com a denominação de Netwarehouse e Netmining, respectivamente. Foi apresentado, no capítulo 4, o conceito de Web

OLAP ou WOLAP, onde foi visto que a ferramenta OLAP já está atuando no ambiente da Internet, através das Intranets ou das Extranets.

GERBER (1998a) e PENDSE (1998b) vão mais longe em suas previsões. Para os autores, devido à migração das ferramentas para o ambiente da Internet, estão sendo feitos grandes investimentos para se realizar as adaptações necessárias. Com estas adaptações, estão diminuindo as fronteiras entre o data warehouse, o OLAP e o data mining. Alguns produtos para OLAP e data warehouse estão incorporando algoritmos de mineração de dados. Algumas ferramentas para data warehouse estão incorporando as ferramentas OLAP. Alguns produtos para mineração de dados estão incorporando recursos para manipulação de dados antes exclusivos dos bancos de dados.

O acesso a um data warehouse, através da Internet, é uma solução aberta e permite o uso de qualquer browser da Web e o acesso em qualquer lugar onde haja conexão com a Internet. Este tipo de acesso pode ser um recurso cujo custo-benefício pode alcançar economias notáveis. Porém, o assunto segurança sempre está presente quando se trata de usar a Internet, para se obter acesso aos dados da empresa (RADEN, 1998b). Segundo o autor, existem vários estudos para se tentar solucionar este problema. A criptografia dos dados é uma das soluções propostas e que está sendo adotada no ambiente das transações via Internet. A outra solução, que também pode ser utilizada em conjunto com a criptografia, é o uso de programas de segurança conhecidos por “firewall”. Esta solução diminuiu drasticamente o problema de acesso e uso indevido dos dados. Outros problemas para o uso da Internet são: interface mais pobre que a do ambiente cliente/servidor, compatibilidade entre os servidores Web e os servidores cliente/servidor e a velocidade de transmissão dos dados também inferior ao ambiente cliente/servidor.

Até pouco tempo atrás, a união do ambiente do data warehouse com a Internet era tida como uma hipótese remota para as corporações. Agora, esta hipótese começa a se tornar realidade, trazendo como benefício imediato o acesso aos dados por um maior número de usuários. Alguns fatores, como menores custos

operacionais, manutenção mais fácil em relação ao ambiente cliente/servidor; interface de acesso mais amigável ao data warehouse e o uso de ferramentas OLAP, estão levando muitas empresas a montarem data warehouse no ambiente da Internet. Com o crescimento de algumas empresas que possuem filiais em diversos países e a necessidade de gerenciar seus dados no ambiente cliente/servidor, a única alternativa seria através de uma WAN, cujos custos de implantação, de operação e de manutenção são muito altos. A Internet está sendo uma solução natural às WANs.

Outra tendência natural, para as três novas ferramentas, é a incorporação das mesmas, gradativamente, pelos pacotes de gestão, os ERP. Para gerenciarem as empresas e os negócios, os gerentes e analistas precisam tanto de ferramentas para o gerenciamento do ambiente operacional da empresa (por exemplo, os pacotes integrados de gestão empresarial) como de ferramentas para o gerenciamento do ambiente gerencial (por exemplo, as três ferramentas apresentadas neste trabalho). Para se tomar diversos tipos de decisão, faz-se necessário, primeiro, uma consulta no ambiente gerencial (por exemplo, para se verificar tendências) e logo após outra consulta no ambiente operacional, para se verificar a exata situação atual, somente então pode-se chegar a uma conclusão que levará a uma decisão. Esta divisão entre os dois ambientes, que é um tanto incômoda, não poderá durar muito tempo e, em um futuro não muito distante, deverão se integrar e formar uma única ferramenta.

Para confirmar a tendência levantada acima, MOAD (1997) relata que diversas empresas desenvolvedoras de soluções para o ambiente ERP, estão começando a integrar ferramentas para data warehouse e OLAP com os seus produtos. O autor cita algumas empresas que estão iniciando esta integração: Baan Co., Computers Associates International, Oracle Corp., PeopleSoft Inc., SAP AG e SSA. Cada uma destas empresas possui seus próprios planos para realizar esta integração.

Neste capítulo, foram apresentadas algumas tendências que poderão ocorrer com as três novas ferramentas nos próximos anos. Algumas já estão em andamento, como a migração para o ambiente da Internet, a união das ferramentas e a integração

das mesmas com o ambiente do ERP. A última tendência, a união do ambiente operacional com o ambiente gerencial, somente o tempo poderá confirmá-la ou refutá-la.

No próximo capítulo serão apresentadas as conclusões deste trabalho.

Capítulo 8 - Conclusões

O sucesso da implantação de um data warehouse depende de um projeto muito bem elaborado, onde sejam bem delineados todos os itens que o compõem. Cuidados especiais devem ser tomados com os seguintes itens: qualidade, segurança e credibilidade dos dados. Após realizada a implantação, devem ser acrescentados, também, cuidados com o controle sobre os custos operacionais, com as manutenções preventiva e corretiva do ambiente computacional e também com a evolução deste ambiente, para que não se alcance o saturamento do sistema provocado pelo crescimento no volume dos dados armazenados.

Um dos pontos fundamentais, nas fases de elaboração e de implantação do projeto e também na fase operacional do data warehouse, são os recursos humanos. Todo o suporte computacional elaborado e implantado servirá para dar suporte aos seus usuários para que tomem decisões importantes para a empresa e para os negócios. Não adianta realizar um alto investimento apenas no ambiente computacional e não se investir nas pessoas que farão este ambiente ganhar vida. Boas ferramentas e bons equipamentos nas mãos de pessoas despreparadas e/ou desmotivadas têm um efeito tão prejudicial como um alto investimento em um ambiente computacional ruim. Esta segunda hipótese (investimento em um ambiente computacional ruim) ainda pode ser superada com o emprego de recursos humanos muito bem preparados e altamente motivados e estes podem reverter uma tendência de fracasso. Porém, caso ocorra a primeira hipótese, as chances de se obter um fracasso são muito maiores. Portanto, o investimento nos recursos humanos é tão ou até mais importante que o investimento realizado na aquisição de novas tecnologias.

Outro fator de extrema importância para o sucesso do empreendimento é a criação e a manutenção de um clima de cooperação e de harmonia entre as pessoas dos diversos setores que compõem o ambiente do data warehouse. A soma dos esforços individuais levará a melhores resultados comparando-se com os resultados que seriam obtidos com esforços individuais isolados.

As ferramentas OLAP e data mining podem ser implantadas isoladamente e independentemente da implantação do data warehouse. Porém, sem o data warehouse, o uso destas ferramentas terá que ser feito diretamente sobre os diversos aplicativos operacionais espalhados pela empresa, é justamente para evitar este tipo de problema e suas conseqüências que foi criado o data warehouse.

Na prática, as implantações das três ferramentas ocorrem na seqüência em que são apresentadas neste trabalho. Primeiro cria-se o data warehouse para que haja a disponibilidade dos dados para uso exclusivo dos propósitos gerenciais. Depois, a própria disponibilidade dos dados habilita e facilita as consultas, cálculos e análises mais sofisticadas, os quais são sempre necessários aos negócios, fazendo com que, naturalmente, se recorra às ferramentas OLAP. E, finalmente, após se explorar os dados de diversas maneiras com a segunda ferramenta, o próximo passo, também naturalmente, é o de se tentar extrair o conhecimento oculto entre os dados, os quais podem gerar novas estratégias para os negócios. Esta última atividade é realizada com a terceira ferramenta, o data mining. Dificilmente, se esta ordem for alterada, pode-se obter melhores resultados que os apresentados seguindo a ordem considerada natural de implantação das ferramentas.

Para a confecção desta dissertação, buscou-se uma diversificada literatura para poder analisar o tema, de forma a não se ficar preso aos conceitos de poucos autores ou literaturas. Pelo contrário, pode-se confrontar ou confirmar os conceitos de autores distintos, possibilitando a seleção dos conceitos considerados mais adequados para este trabalho. Tem-se a convicção de que a forma com que foi abordado este complexo tema foi suficiente para que fossem alcançados os objetivos

propostos. Quem desejar se aprofundar um pouco mais tecnicamente nos conhecimentos relativos às ferramentas apresentadas poderá consultar as literaturas apresentadas nas referências bibliográficas e na bibliografia complementar. Para que se pudesse mostrar o que existe de mais atual a respeito destas ferramentas deu-se prioridade às literaturas mais recentes.

Para futuros trabalhos, seguem-se as seguintes sugestões:

- ⇒ um estudo da viabilidade de cada uma das três ferramentas em cada área da empresa, como vendas, marketing, produção, finanças, patrimônio, planejamento e outros setores;
- ⇒ um estudo da viabilidade de cada uma das três ferramentas para os diversos tipos de negócios. Por exemplo: comércio, indústria, consultorias, educação, saúde, transporte, alimentício, bancos, comunicações e outros setores;
- ⇒ uma análise do custo-benefício em empresas que já implantaram as ferramentas, confirmando ou negando a eficácia delas para o suporte ao gerenciamento dos negócios;
- ⇒ um estudo da viabilidade da implantação destas ferramentas em empresas de pequeno e médio porte;
- ⇒ um estudo sobre quais os níveis gerenciais em que estas ferramentas prestam o melhor suporte, se no nível estratégico, tático ou operacional.

Anexo A - Um exemplo de um aplicativo de data mining: SuperQuery

Este exemplo foi extraído do artigo de AZMY (1998). Segundo o autor, o SuperQuery contém muitas facilidades para preparar, unir, dividir, classificar e resumir tabelas de dados. Além disso, o SuperQuery tem outros recursos gráficos e estatísticos que ajudam a analisar e explorar dados.

O autor apresenta as seguintes afirmações, para servir de exemplo, a fim de explicar o funcionamento do produto:

- Todas linhas que têm **Produto = Jaqueta**, também têm **Lucro = Alto**.
- A maioria das linhas que têm **Produto = Chapéu**, também têm **Lucro = Negativo**.
- Somente 1% das linhas que têm **Produto = Chapéu**, têm **Lucro = Alto**.

Para o autor, as duas primeiras declarações são **fatos** e simplesmente dizem que as jaquetas são lucrativas e chapéus não são. A última declaração é uma **exceção** e mostra que há poucas linhas de chapéus que são muito lucrativas. Estas linhas são interessantes, porque eles podem dar uma pista sobre quando as linhas de chapéu são lucrativas.

O SuperQuery informa o fator de confiança e o nível de linhas que dão apoio a cada fato ou exceção. A segunda declaração pode ter um fator de confiança de 80% e pode ter um nível de apoio de 1000 linhas.

O SuperQuery armazena todas os fatos descobertos em uma tabela chamada Fact Table. Esta tabela é tratada como qualquer outra tabela de dados. Isto significa que o usuário pode usar todas as ferramentas de análise do próprio SuperQuery, para explorar e entender os fatos encontrados. Torna-se mais importante quando o SuperQuery descobre vários fatos. A visão da Fact Table tem uma característica adicional que é o botão Drill-Down. Este botão permite ao usuário inspecionar algum fato e ver as linhas que lhe dão apoio. Esta última atividade exibe o conteúdo das linhas que dão apoio ao fato atual ou à exceção e torna disponível ao usuário a utilização de gráficos e estatísticas sobre um determinado fato.

A interface do usuário do SuperQuery possui “défauls” para todos os controles. O usuário necessita somente prover um nome para a tabela de saída e pressionar o botão “Run”.

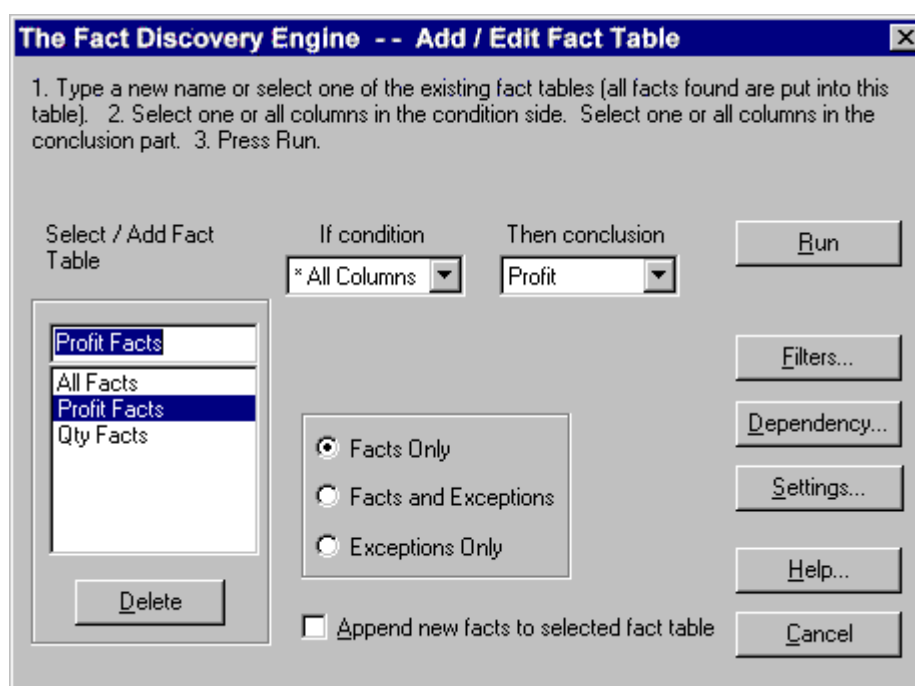


Figura 21: A Janela Principal da SuperQuery.

É importante evitar perder tempo descobrindo fatos já conhecidos. O objetivo do SuperQuery é informar somente os fatos interessantes; o produto permite ao usuário entrar em grupos de colunas que já se sabe que são colunas dependentes e o seu relacionamento não é útil. A Figura 22 mostra a janela que permite esta seleção.

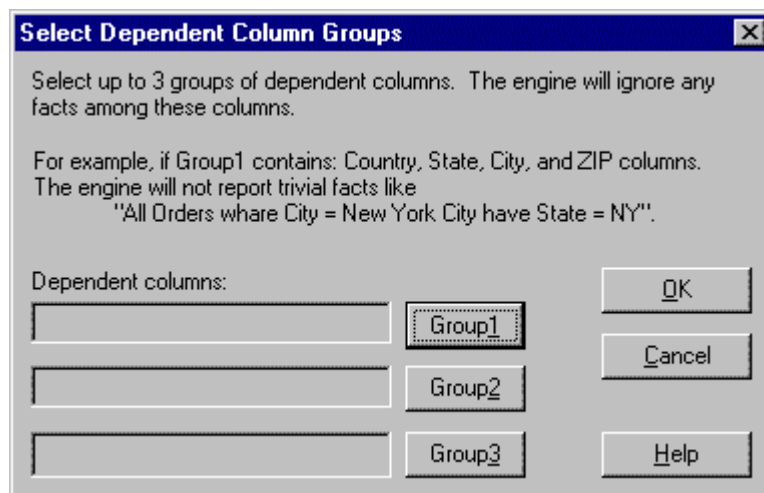


Figura 22: Janela onde é feita a seleção das colunas dependentes

É importante, também, não informar nenhum fato que possa ser uma consequência de outros fatos mais gerais. Há várias outras dependências que o SuperQuery pode descobrir automaticamente e evitá-las, como, por exemplo: fatos que contêm valores em branco, fatos entre uma coluna calculada e suas colunas fontes ou fatos que obtêm como resultado valores óbvios. A Figura 23 mostra todos os padrões que o produto ignora automaticamente, permitindo ao usuário ativar ou desativar os filtros que julgar necessário.

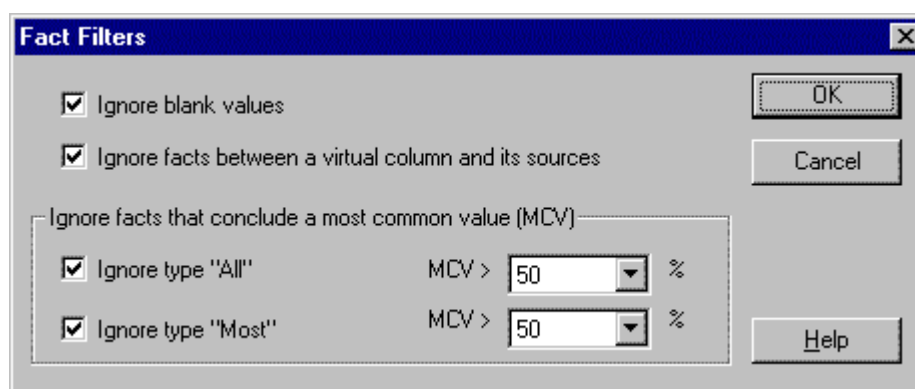


Figura 23: Ativando ou Desativando Filtros Automáticos.

⇒ **Ignore blank values**: quando informando os fatos, não informa qualquer fato que tenha um valor em branco em sua condição ou partes de conclusão.

- ⇒ **Ignore facts between a virtual column and its sources:** não informa os fatos que são o resultado de dependências existentes entre uma coluna virtual (calculada) e suas colunas fontes.
- ⇒ **Ignore facts that conclude a most common value (MCV):** suponha-se que uma coluna Lucro possa ter 3 valores: baixo, médio ou alto, e médio é o valor mais comum. Qualquer fato que conclua este valor é normalmente desinteressante. Seja o exemplo, um fato como: “*A maioria dos Produto = Jaqueta tem Lucro = Médio*” pode não ser interessante, desde que se saiba que a maioria dos produtos tem lucro médio. Por outro lado, um fato como: “*A maioria dos Produto = Jaqueta tem Lucro = Alto*” é interessante, desde que *Lucro = Alto* não seja esperado. Este filtro pode ser fixado para fatos dos tipos “All” e “Most”. A porcentagem com que um valor é considerado mais comum, também pode ser especificado.

O SuperQuery proporciona ao usuário o controle total na qualidade dos fatos encontrados. Isto é feito permitindo-se ao usuário configurar as exigências das linhas que dão apoio aos fatos (Figura 24). Observe-se o fato: “*A maioria dos produtos onde Produto = Jaqueta tem Lucro = Alto*” é apoiado por 100 linhas, isto significa isso há 100 linhas na tabela, onde o “*Produto = Jaqueta*” e ao mesmo tempo, “*Lucro = Alto*”. O usuário pode aumentar ou pode diminuir a exigência do nível de linhas que dão apoio a um fato. Aumentando o nível de linhas que dão apoio, reduz o número de fatos descobertos, porém é provável que esses fatos sejam interessantes e/ou importantes. Por outro lado, às vezes, é necessário se encontrarem mais padrões ou fatos. Neste caso, é diminuído o nível de apoio, e o processo de descoberta de fatos ocupará mais tempo para processar. Cada slider a ser configurado tem dois números: a porcentagem do lado esquerdo é o fator de confiança, o número ao lado direito é o valor do próprio slider.

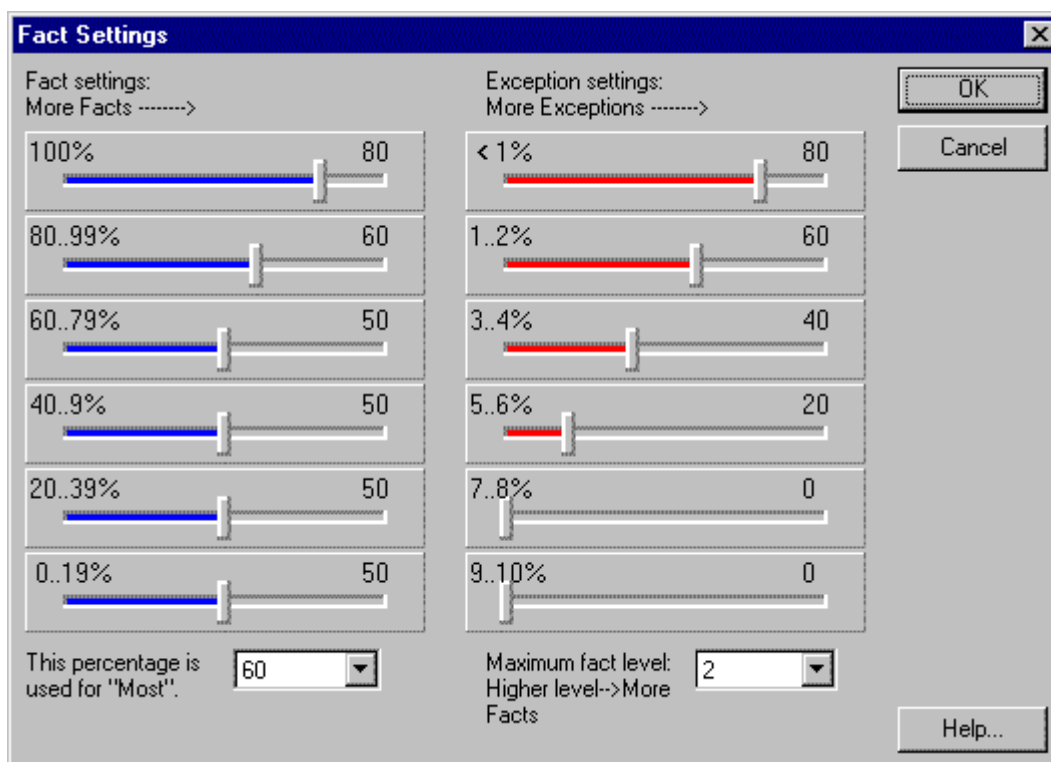


Figura 24: A janela de Ajustes de Descoberta de Fatos

Os sliders azuis controlam o número de Fatos encontrados em diferentes fatores de confiança. O primeiro slider é para fatos de tipo “All”, ou seja, um fator de confiança de 100%. O outro slider é para fatos de tipo a “Most” e tem um fator de confiança no alcance 80% a 99%. Os outros sliders operam com fatores de confiança menores. Aumentando o slider, aumentam o número de fatos fornecidos, diminuindo o nível de linhas de apoio aos fatos naqueles fatores de confiança. Por outro lado, diminuindo o slider, diminui o número de fatos fornecidos, exigindo que cada fato seja apoiado por mais linhas. Isso significa que o produto poderia ser configurado em uma grande variedade de modos. Por exemplo, poderia encontrar somente fatos com 100% de confiança, ou fatos com confiança no alcance 80% a 99%, ou qualquer combinação.

Os sliders vermelhos são as Exceções. Cada slider controla o número de exceções encontrado. Aqui se usam níveis de agrupamento, em vez de níveis linhas que dão apoio ao fato. Para uma exceção ser informada em uma coluna, esta coluna tem que ter poucos valores dominantes ou agrupamentos. Quando aumentamos o

slider, o número de exceções solicitada aumenta e o nível de agrupamento exigido é diminuído.

O SuperQuery também permite ao usuário selecionar um nível mais alto de um fato para se fazer a busca, ou seja, um fato como: “*A maioria dos produtos onde Produto = Jaqueta tem Lucro = Alto*” é considerado um fato com nível 1, porque tem apenas 1 condição. Por outro lado, um fato como “*A maioria dos produtos onde Produto = Jaqueta e Cor = Preto tem Lucro = Alto*” é um fato com nível 2, porque tem 2 condições. Normalmente, usa-se, no máximo, o nível 3, onde aparecem 3 condições.

O autor relata que o SuperQuery tem três formatos, para apresentar seus fatos descobertos e/ou as exceções:

- ⇒ Declarações IF / Then
- ⇒ Declarações All / Most
- ⇒ Tabela de fatos

Enquanto o produto está realizando as descobertas, são informados os padrões ou fatos, através de regras do tipo declarações “Se ... Então ...”, como mostrado na Figura 25.

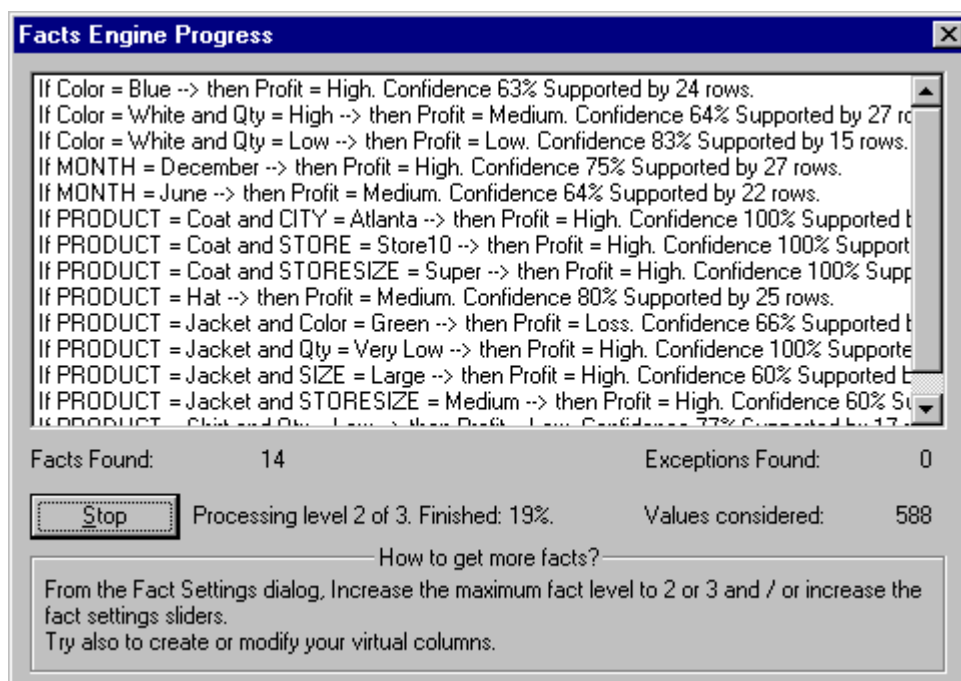


Figura 25: O produto informando as regras durante o processamento

O formato da exibição da tabela de fatos é mais satisfatório para manipular um grande número de fatos. Considerando que são representados fatos em uma tabela de dados normal, é possível usar vários filtros e até mesmo visualização gráfica. A partir da tabela de fato, pode-se:

- ⇒ filtrar fatos sobre uma coluna específica,
- ⇒ focar fatos do tipo “All” ou “Most”,
- ⇒ focar fatos com qualquer fator de confiança específico,
- ⇒ focar fatos que são apoiados por um número específico de linhas,
- ⇒ olhar os fatos em um formato gráfico,
- ⇒ usar o SuperQuery, na tabela de fatos, para ver se há qualquer padrão entre os próprios fatos.

A Figura 26 mostra uma visão da tabela de fatos e alguns comentários feitos pelo autor.

The Fact Table View

Drill down button creates a view with all rows supporting the current fact.

Type	IfColumnn	IfValue	ThenColumnn	ThenValue	Percent	No. Row	Level
Most	PRODUCT	Hat	Profit	Medium	80	25	1
Most	Color	Blue	Profit	High	63	24	1
Most	MONTH	June	Profit	Medium	64	22	1
Most	MONTH	December	Profit	High	75	27	1
Most	PRODUCT&Color	Jacket&Green	Profit	Loss	66	10	2
Most	PRODUCT&SIZE	Jacket&Large	Profit	High	60	9	2
Most	PRODUCT&STORESIZE	Jacket&Medium	Profit	High	60	9	2
All	PRODUCT&Qty	Jacket&Very Low	Profit	High	100	4	2
Most	PRODUCT&Qty	Shin&Low	Profit	Low	77	17	2

Most Sales where PRODUCT = Jacket and SIZE = Large → have Profit = High.
Percentage 60% Supported by 9 rows.

Confidence

Fact is supported by 9 rows.

Show other pane

This text changes as you move in the fact table to explain each row

Figura 26: Uma visão da tabela de fatos

Para finalizar, o autor cita outras funções do SuperQuery:

- ⇒ Wizards para ajudar os usuários a realizarem as tarefas sem esforços,
- ⇒ Consulta visual para extrair ou fazer amostras dos dados das linhas,
- ⇒ Colunas virtuais para preparar e classificar os dados das linhas,
- ⇒ Gráficos automáticos e estatísticas para visualizar e explorar os dados,
- ⇒ Tabela sumária para se agrupar linhas em classes,
- ⇒ Organizador para as várias visões de análise.

Anexo B - Um exemplo de um produto OLAP: Powerplay 6.0

Este exemplo foi extraído do site do produto na Internet (COGNOS, 1998). Segundo o desenvolvedor do produto, o PowerPlay 6.0 é uma ferramenta para análise multidimensional robusta e geração de consultas e relatórios sobre os negócios da empresa. Permite aos usuários analisar múltiplas dimensões dos dados em qualquer combinação, identificar tendências, descobrir como os negócios estão sendo conduzidos e gerar os relatórios necessários. Devido aos seus recursos e seu desempenho, permite aos decisores dedicarem mais tempo analisando os dados e buscando novas estratégias para os negócios.

O produto trabalha com um banco de dados multidimensional (MOLAP) chamado PowerCube, que gera as estruturas de dados chamadas cubos, as quais representam os modelos de negócios que os usuários podem utilizar para identificar tendências, realizar medições e administrar o desempenho dos negócios.

Entre as principais características e recursos do PowerPlay 6.0 estão:

- ⇒ navegação drill-down ou roll-up, que permite respostas rápidas para perguntas críticas sobre os negócios, movendo-se facilmente de um nível de agregação dos dados a outro mais resumido (roll-up) ou mais detalhado (drill-down);
- ⇒ seleção de dados, que permite que se selecione somente os dados necessários e se construa outro banco de dados temporário para as consultas e análises necessárias (slice and dice);
- ⇒ visualização dos dados em qualquer ângulo dos negócios que seja necessário;

- ⇒ visualização dos dados de diversas formas, incluindo gráficos, para a melhor compreensão dos fatores chaves que estão conduzindo os negócios (Figura 27);
- ⇒ possui diversas funções que executam os cálculos mais comuns nas análises dos negócios (Figura 27);
- ⇒ enfoque nos produtos que mais requerem atenção, destacando-os ou proporcionando análises mais personalizadas pré-definidas (Figura 28);
- ⇒ identificação e destaque dos melhores e dos piores itens em termos de desempenho;
- ⇒ análise dos dados de acordo com as regras do negócio, já pré-estabelecidas, como evolução, porcentagens, médias e variações das vendas;
- ⇒ reúne os requisitos essenciais dos negócios, que permitirão a confecção de consultas e relatórios, onde serão utilizados dados somente de algumas categorias de itens, porém, em múltiplos níveis de agregação;
- ⇒ permite a configuração e formatação dos dados de acordo com as necessidades;
- ⇒ possui diversas funções automáticas que representam atividades analíticas dos negócios;
- ⇒ permite construir e distribuir relatórios dinâmicos que satisfazem as necessidades dos usuários através do uso de hierarquias aninhadas, onde somente algumas categorias de itens são ampliadas aos níveis de maiores detalhes;
- ⇒ cria relatórios que podem conter dados de múltiplas dimensões;
- ⇒ escolha de uma grande variedade de opções de formatos de relatórios para proporcionar ao usuário utilizar aquele que seja mais adequado às necessidades;
- ⇒ navegação até o maior nível de detalhe possível;
- ⇒ realiza as conversões necessárias;
- ⇒ compara resultados;
- ⇒ publica relatórios como documentos HTML;
- ⇒ constrói os cubos rapidamente;
- ⇒ gera macros automaticamente;
- ⇒ utiliza as dimensões de tempo de forma avançada;
- ⇒ permite a personalização dos menus;
- ⇒ possui help on-line.

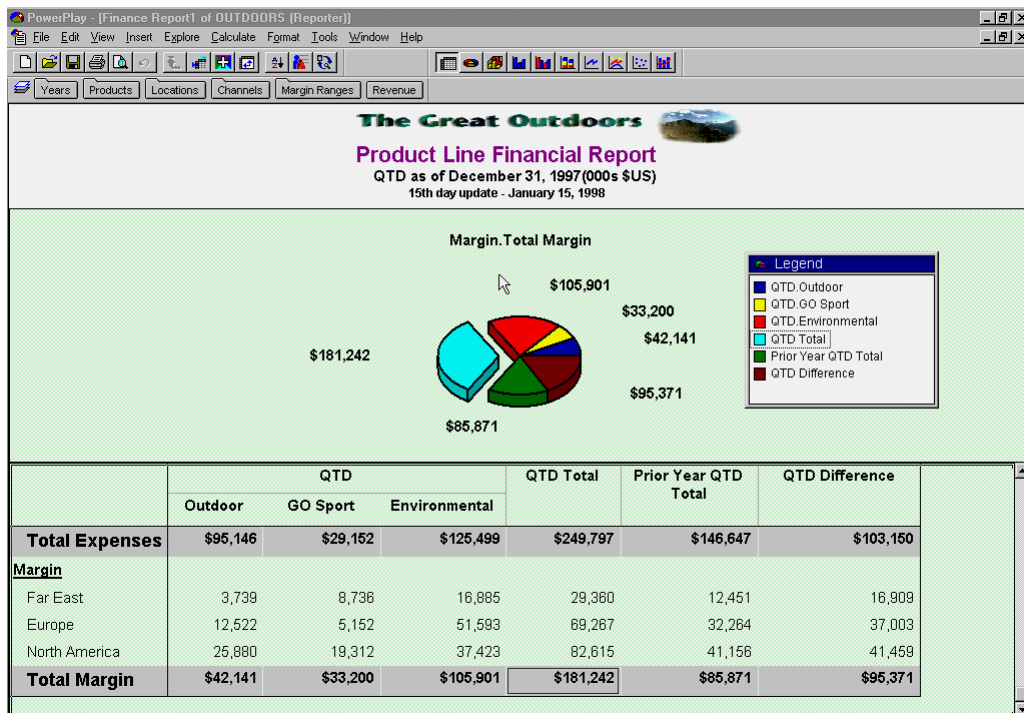


Figura 27 - Tela de um relatório financeiro gerado pelo PowerPlay 6.0

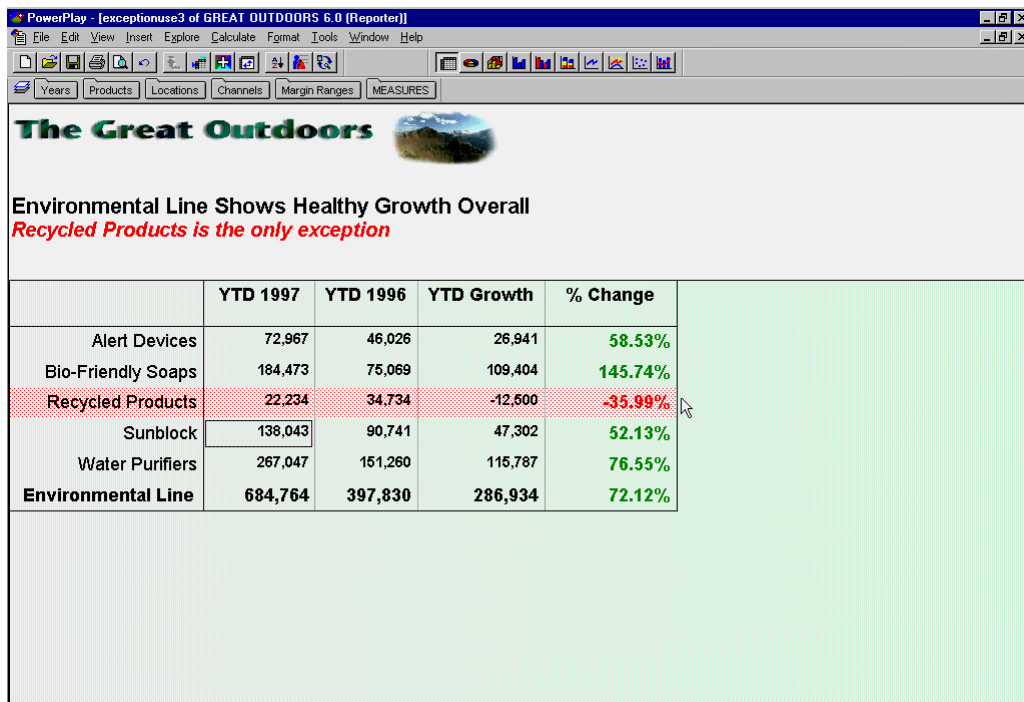


Figura 28 - Tela exibindo um item em destaque e possibilitando uma análise mais detalhado no mesmo

PowerPlay - [nestuse of Great Outdoors 6.0 (Reporter)]

File Edit View Insert Explore Calculate Format Tools Window Help

Years Products Locations Channels Margin Ranges Revenue

The Great Outdoors

Outdoor Products and Environmental Line
Margin Ranges by Channel (US\$)

			1996	1997
Outdoor Products	Channels	Margin Ranges	451,162	429,291
		Under 20%	0	3,750
		20% - 65%	451,162	425,541
		Over 65%	0	0
	Independent		27,940	61,953
	Sports Chain		81,262	99,622
	Camping Chain		151,757	143,508
	Mass Marketer		85,102	90,442
	GO Outlet		105,101	33,766
	Environmental Line	Channels	Margin Ranges	397,830
Under 20%			45,243	80,270
20% - 65%			0	0
Over 65%			352,587	604,494
Independent		39,239	57,055	
Sports Chain		92,781	169,995	
Camping Chain		116,682	209,580	
Mass Marketer		84,958	172,298	
GO Outlet		64,170	75,836	

Outdoor Products.Channels.20% - 65%

Figura 29 - Tela exibindo relatório contendo análise personalizada e pré-definida

PowerPlay - [Finance Report4 of FINANCE (Reporter)]

File Edit View Insert Explore Calculate Format Tools Window Help

Years Products Geography Income Stmt. Lines % of Plan MEASURES

The Great Outdoors

Monthly Operating Statement by Product
December FY98
(US \$000s)

	Environmental Line			Outdoor Products			GO Sport Line			Total Actual	Total Plan	Total % Plan
	Actual	Plan	% Plan	Actual	Plan	% Plan	Actual	Plan	% Plan			
QTD												
Revenue	198,902.53	181,270.42	109.73%	112,122.17	142,860.36	78.48%	48,904	50,904	96.07%	359,929	375,035	95.97%
COGS												
Materials	-22,550.33	-19,945.71	113.06%	-51,232.33	-72,396.30	70.77%	-13,112	-13,766	95.25%	-86,894	-106,108	81.89%
Mfg/Pkg	-35,220.04	-33,876.64	103.97%	-29,727.15	-30,430.64	97.69%	-10,615	-10,024	105.90%	-75,562	-74,331	101.66%
Shipping	-9,603.70	-9,657.67	99.44%	-14,186.45	-14,555.68	97.46%	-5,426	-5,112	106.14%	-29,216	-29,325	99.63%
SG & A												
Salary	-4,631.39	-4,631.39	100.00%	-5,029.96	-5,029.96	100.00%	-1,999	-1,999	100.00%	-11,660	-11,660	100.00%
Commission	-13,923.21	-12,688.91	109.73%	-3,363.68	-4,285.80	78.48%	-2,445	-2,545	96.07%	-19,732	-19,520	101.09%
Marketing	-7,670.49	-6,532.00	117.43%	-8,559.61	-7,895.74	111.23%	-199	-244	81.43%	-16,429	-14,472	113.52%
Bad Debt	-6,298.94	-1,781.18	353.64%	-2,044.50	-1,450.00	141.00%	-812	-812	100.00%	-9,156	-4,043	226.44%
Other Selling	-618.33	-971.24	63.66%	-9,599.45	-8,952.15	107.23%	-2,235	-2,026	110.33%	-12,453	-11,949	104.21%
General	-444.71	-535.64	83.02%	-1,357.28	-1,401.72	96.83%	-121	-209	57.67%	-1,923	-2,147	89.56%
Administration	-6,416.54	-6,509.92	98.57%	-6,974.50	-7,076.00	98.57%	-2,771	-2,811	98.57%	-16,162	-16,397	98.57%
Other	-3,978.09	-3,625.38	109.73%	-2,242.48	-2,857.19	78.49%	-978	-1,018	96.07%	-7,199	-7,501	95.97%
Taxes	-5,967.09	-5,438.10	109.73%	-3,363.68	-4,285.80	78.48%	-1,467	-1,527	96.07%	-10,798	-11,251	95.97%
Total QTD	81,579.67	75,076.64	108.66%	-25,558.90	-17,556.62	145.58%	6,725	8,812	76.32%	62,746	66,332	94.59%
Prior QTD	52,364.55	50,365.13	103.97%	-32,811.26	-26,692.54	122.92%	-2,917	-1,947	149.87%	16,636	21,726	76.57%

Figura 30 - Tela contendo um dos diversos tipos de relatórios possíveis

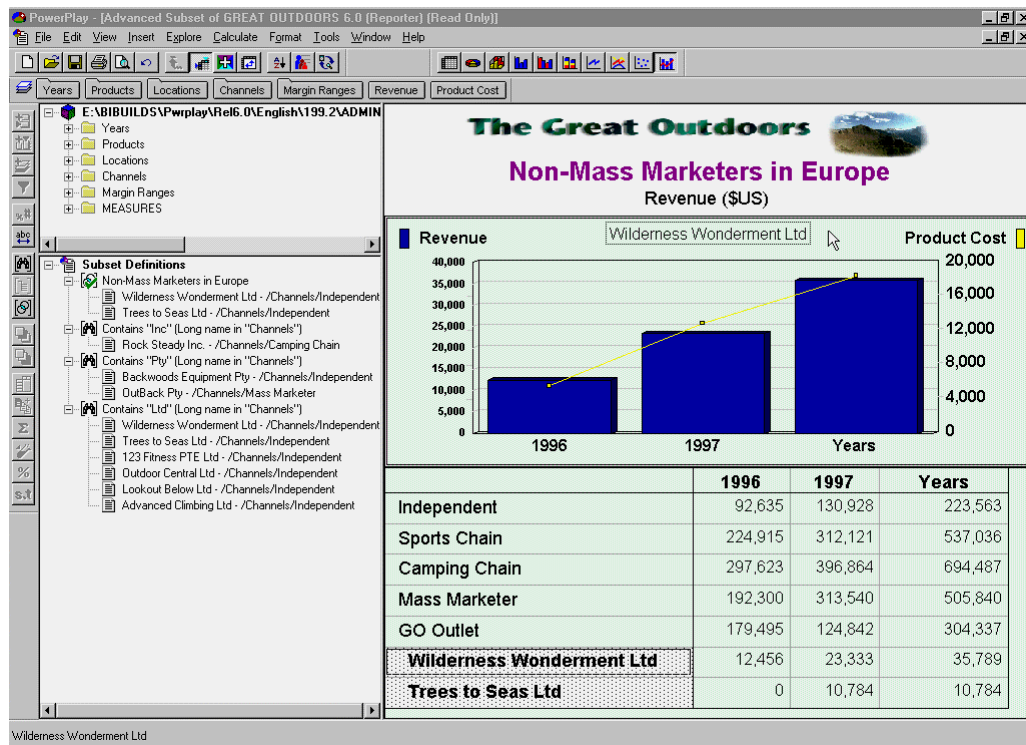


Figura 31 - Tela contendo uma das análises disponíveis

YEARS DIVISIONS PRODUCTS PLATFORMS

REVENUE TYPES CHANNELS

Years, Divisions, Products, Platforms, Revenue Types, Channels

Revenue	US	Europe	Campac	Latin America	Australia/N.Z.	Unknown	DIVISIONS
1992	58796.444	46369.407	11848.564	585.339	8184.622	10370.574	136154.950
1993	59259.956	48984.263	15930.926	759.310	7539.504	2170.491	134644.450
1994	63096.952	51985.734	17626.352	1201.784	8648.645	23.300	142582.767
1995	69976.691	63266.897	18480.606	1242.653	8335.460	0.000	161302.307
1996	30400.745	23915.228	7242.911	389.431	5512.615	0.000	67460.930
YEARS	281530.788	234521.529	71129.359	4178.517	38220.846	12564.365	642145.404

COCOS

Figura 32 - Tela contendo um dos tipos disponíveis de visualização e navegação dos dados

O produto permite o gerenciamento de até 50 milhões de linhas de dados e até 500.000 categorias de itens, proporcionando-lhes acesso fácil e rápido e permitindo-lhes distribuição e atualização:

- ⇒ permite a construção do PowerCube em servidores Unix e Windows NT, possibilitando a construção de cubos de qualquer tamanho;
- ⇒ distribui o PowerCube para clientes Windows, servidores, LANs e Web;
- ⇒ permite que se povoe o PowerCube com dados do data warehouse, de data marts, de bancos de dados relacionais e de planilhas eletrônicas;
- ⇒ permite definir classes de usuários e, em cada classe, a definição dos níveis de acesso dos usuários, permitindo-lhes ter acesso mediante as restrições necessárias;
- ⇒ acesso aos dados por meio de conexão direta na rede, dial-up, LAN ou Internet.

Algumas especificações técnicas do PowerPlay 6.0:

- ⇒ hardware mínimo: PC 486, 16 Mb RAM e 36 Mb de espaço em disco;
- ⇒ plataforma Windows: Windows 95, 98 e NT;
- ⇒ plataforma UNIX: HP, UX, AIX, Sol Solaris e DEC UNIX;
- ⇒ bancos de dados suportados: Informix, Oracle, Sybase SQL Server, Microsoft SQL Server, IBM DB2 Server e IBM DB2 Universal Database;
- ⇒ interface com outros produtos OLAP: Arbor Essbase, Oracle Express, IBM DB2 OLAP, Microsoft OLAP Server.

REFERÊNCIAS BIBLIOGRÁFICAS

ADELMAN, S. (1998). *The data warehouse database explosion*. Data Management Review, 1996.

[http:// http://www.data-warehouse.com/resource/articles/adelman.htm](http://www.data-warehouse.com/resource/articles/adelman.htm) (13 jan.).

ADRIAANS, P.; ZANTINGE, D. (1996). *Data mining*. s.l., Addison-Wesley.

ALVES, A. (1997). Prospecção. *Developers' Magazine*, n. 6, p. 40-41, fev.

AMARAL Jr., A. (1997). Desmitificando definitivamente o data warehouse. *Developers' Magazine*, n. 6, p. 14-17, fev.

ANAND, S.; BUECHNER, A. (1998). *Decision support using data mining*. s.l., Financial Times Pitman Publishing.

ARNUM, E. (1998). *Billing for telecommunications services*. EMMS, Silicon Graphics Computer Systems. <http://www.sgi.com/Technology/billing.html> (05 ago).

AZMY, A. (1998). *SuperQuery: data mining for everyone*. Azmy Thinkware. <http://www.azmy.com/wp1.htm> (22 jan.).

BARBIERI, C. (1997). *Novos conceitos, técnicas novas*. Computerworld, Guideline: Data Warehouse, p. 7, fev. Suplemento especial.

- BERRY, M.; LINOFF, G. (1997). *Data mining techniques for marketing, sales and customer support*. S.l., John Wiley & Sons.
- BHANDARI, I.; COLET, E; PARKER, J.; PINES, Z.; PRATAP, R. (1997). Advandec Scout: data mining and knowledge discovery in NBA data. *Data Mining and Knowledge Discovery*, n. 1, p. 121-125.
- BIGUS, J. (1996). *Data mining with neural networks*. s.l., McGraw-Hill.
- BISPO, C. A. F.; ULIAN, L. G., LEME, E. R.. (1995) *Programa de qualidade total*. Ribeirão Preto. /Monografia de Formatura (graduação) - Universidade de Ribeirão Preto/
- BISPO, C.A.F.; ULIAN, L.G.; LEME, E.R.; CAZARINI, E.W. (1997). Implantação de um programa de Qualidade Total. In: ENCONTRO REGIONAL DA ANGRAD, Ribeirão Preto, 1997. *Anais*. Ribeirão Preto, p. 15-22.
- BISPO, C.A.F.; CAZARINI, E.W. (1998a). Conceitos básicos e a elaboração de um projeto lógico de um data warehouse. (CD-ROM) In: ENCONTRO NACIONAL DE ENGENHARIA DA PRODUÇÃO, 18., / CONGRESSO INTERNACIONAL DE ENGENHARIA INDUSTRIAL, 4., Niterói, 1998. *Anais*. Niterói, TEP- UFF, artigo 221.doc.
- BISPO, C.A.F.; CAZARINI, E.W. (1998b). A nova geração de Sistemas de Apoio à Decisão. (CD-ROM) In: ENCONTRO NACIONAL DE ENGENHARIA DA PRODUÇÃO, 18., / CONGRESSO INTERNACIONAL DE ENGENHARIA INDUSTRIAL, 4., Niterói, 1998. *Anais*. Niterói, TEP- UFF, artigo 222.doc.
- BISPO, C.A.F.; CAZARINI, E.W. (1998c). A evolução do processo decisório. (CD-ROM) In: ENCONTRO NACIONAL DE ENGENHARIA DA PRODUÇÃO, 18., / CONGRESSO INTERNACIONAL DE ENGENHARIA INDUSTRIAL, 4., Niterói, 1998. *Anais*. Niterói, TEP- UFF, artigo 94.doc.

- BISPO, C.A.F.; CAZARINI, E.W. (1998d). A contribuição do data warehouse e suas ferramentas para o gerenciamento dos negócios. (CD-ROM) In: SIMPÓSIO DE ENGENHARIA DA PRODUÇÃO, 5., Bauru, 1998. *Anais*. Bauru, UNESP, Faculdade de Engenharia de Bauru, Departamento de Engenharia da Produção, Trabalho04.doc.
- BISPO, C.A.F.; CAZARINI, E.W. (1998e). Uma visão corporativa no projeto de um data warehouse. (CD-ROM) In: SIMPÓSIO DE ENGENHARIA DA PRODUÇÃO, 5., Bauru, 1998. *Anais*. Bauru, UNESP, Faculdade de Engenharia de Bauru, Departamento de Engenharia da Produção, Trabalho78.doc.
- BRAND, E.; GERRITSEN, R. (1998). *Data mining and knowledge discovery*. DBMS Magazine, Jul. 1998. Data Mining Solutions Supplement. <http://www.dbmsmag.com/9807m01.html> (27 ago.).
- CAIXA ECONÔMICA FEDERAL. (1998). *Informação: maior ativo das empresas*. Jornal da Caixa, p 1-4, jul. /Encarte/
- CAMPOS, A. (1997a). Estudo inglês analisa ferramentas OLAP. *Computerworld*, 12 maio, p. 11. Guideline.
- CAMPOS, A. (1997b). Data mining fica mais fácil e barato. *Computerworld*, 19 maio, p. 8.
- CAMPOS, V. F. (1992). *TQC: controle da qualidade total (no estilo japonês)*. Belo Horizonte, Fundação Christiano Ottoni / Escola de Engenharia da UFMG.
- CARICKHOFF, R. (1997). *A new face for OLAP*. Internet Systems, Jan. 1997. <http://www.dbmsmag.com/9701i08.html> (16 dez.).
- CARVALHO, J. (1997a). Data mining em intranets. *Computerworld*, 4 ago., p. 7.

- CARVALHO, J. (1997b). *OLAP sem segredos*. Computerworld, nov. 1997. <http://www.idg.com.br/computerworld/computerworld/236/236indep01.htm> (28 dez.).
- CIOS, K.; PEDRYCZ, W.; SWINIARSKI, R. (1998). *Data mining methods for knowledge discovery*. s.l., Kluwer.
- CLAUSING, D. (1993). *Total Quality development*. s.l. Asme Press.
- CODD, E. F.; CODD, S. B.; SALLEY, C. T. (1998). *Providing OLAP (On-Line Analytical Processing) to users-analysts: An IT mandate*. Arbor Software. http://www.arborsoft.com/essbase/wht_ppr/coddc0.html (28 jan.).
- COGNOS (1998). *PowerPlay 6.0*. <http://www.cognos.com/powerplay/info.html> (09 out.).
- COMPUTERWORLD. (1997a). *Data warehouse: a tecnologia que revoluciona os negócios*. Rio de Janeiro, p. 8-10, fev. Guideline.
- COMPUTERWORLD. (1997b). *Data mining ou DSS*. Jun., 1997. <http://www.idg.com.br/computerworld/computerworld/213/guideline/213guide04.htm> (28 dez.).
- COMPUTERWORLD. (1998a). *Data Warehouse*. Rio de Janeiro, p. 4-7, fev. Guideline.
- COMPUTERWORLD. (1998b). *Suspensão dos projetos de data warehouse*. Rio de Janeiro, p. 8, jul.
- COMPUTERWORLD. (1998c). *Evolução dos bytes*. Rio de Janeiro, p. 22-23, set.

- COSTA, P. W. A. (1997). Como surgiram os data warehouses? *Computerworld*, 03 nov., p. 16.
- DATAMIND TECHNOLOGY CENTER. (1998). *Data mining for competitive advantage*. http://www.datamindcorp/paper_advantage.html (16 jan.).
- DATASAGE. (1998). *CirrusNet: an automatic pattern recognition solution for data mining*. http://www.cirrusrec.com/Papers/DM_WP/DM_WP.html (22 jan.).
- DEPOMPA, B. (1998). *Companies that start with data marts find it difficult to move to a warehouse*. Information Week, 11 Nov. 1996.
<http://techweb.cmp.com/iw/605/05olwa1.htm> (14 jan.).
- EDELSTEIN, H. (1997a). *Technology how to: mining data warehouses*. Information Week, 8 Jan., 1996. <http://techweb.cmp.com/iw/561/61oldat.htm> (30 dez.).
- EDELSTEIN, H. (1997b). *Mining for gold*. Information Week, 21 Apr. <http://techweb.cmp.com/iw/627/27olmin.htm> (30 dez.).
- ENGLISH, L. P. (1998). *Help for data quality*. Information Week, 7 Oct. 1996. <http://techweb.cmp.com/iw/600/00oldat.htm> (06 jan.).
- EOM, S. B. (1996). Mapping the intellectual struture of research in Decision Support Systems through autor cocitacion analysis (1971-1993). *Decision Support Systems*, Elsevier, n. 16, p. 315-338.
- FAYYAD, U.; SHAPIRO, G. P.; SMYTH, P.; UTHURUSAMY, R. (1996). *Advances in knowledge discovery and data mining*. s.l., AAAI / MIT Press.
- FAYYAD, U.; GEIGER, D. (1998). Density-based indexing for nearest-neighbor queries. *Microsoft Research Technical Report*, MR-TR-98-58, Sept.

- FEIGENBAUN, A. (1991). *Total quality control*. New York, MacGraw Hill.
- FIGUEIREDO, A. M. C. M. (1998) Molap x Rolap: embate de tecnologias para data warehouse, *Developers' Magazine*, n. 18, p. 24-25, fev.
- FISHER, L. M. (1998). *Along the infobahn: data warehouses*. Strategy & Business, Third Quarter, 1996.
<http://www.strategy-business.com/technology/96308/page1.html> (02 jan.).
- FONTES, E. (1998). Protegendo a informação: fator crítico para o negócio. *Developers' Magazine*, n. 18, p. 32-33, fev.
- FORSMAN, S. (1998). *OLAP Council white paper*. OLAP Council.
<http://www.olapcouncil.org/research/whtpapco.html> (01 fev.).
- FREEDMAN, R.; KLEIN, R.; LEDERMAN, J. (1998). *Leveraging your hidden assets to improve ROI: a case study in the credit card business*. Capital Markets, 1995. http://www.think.com/html/products/darwin/r_intro.htm (17 jan.).
- FURLAN, J. D. (1997). *Modelagem de negócio*. São Paulo, Makron Books.
- GATES, B. (1995). *A estrada do futuro*. São Paulo, Companhia das Letras.
- GARVIN, D. A. (1992). *Gerenciando a qualidade*. São Paulo, Quality Mark.
- GENTIA SOFTWARE. (1998). *OLAP for enterprise*.
<http://www.gentia.com/products/gseolap.exe> (29 jan.).
- GERBER, C. (1998a). *Dissecting data mining*. Datamation.
<http://www.datamation.com/PlugIn/workbench/datamine/stories/dissect.htm> (15 jan.).

GERBER, C. (1998b). *Excavate your data*. Datamation.

<http://www.datamation.com/PlugIn/workbench/datamine/stories/excav.htm> (16 jan.).

GOESTCHI, D.L.; DAVIS, S. (1994). *Introduction to quality total: quality, productivity and competitiveness*. New York, Macmillan.

GONÇALVES, J. L. M.. (1997). Data warehouse é necessariamente um megaprojeto? *Developers' Magazine*, n. 6, p. 12-13, fev.

HACKATHORN, R. (1998a). *Data warehousing energizes your enterprise*. Datamation, 1 Feb. 1995.

<http://www.datamation.com/plugin/issues/1995/feb1/02asw100.html> (03 jan.).

HACKATHORN, R. (1998b). *Data warehousing's credibility crisis*. Byte, Aug., 1997. <http://www.byte.com/art/9708/sec4/art1.htm> (19 ago.).

HAMMER, M. (1994). *Reengenharia: revolucionando a empresa em função dos clientes, da concorrência e das grandes mudanças da gerência*. Rio de Janeiro, Campus.

HODEL, A. (1998). *Data mining: a new weapon for competitive advantage*. Software Quaterly, IBM, 1995.

<http://www.software.ibm.com/sq/issues/vol24/data.htm> (16 jan.).

HORROCKS, C. (1997). *Data warehouses: an executive information perspective*. Perspectives, Oct. 1995. http://www.csc.com/about/tech_dw_arch.html (29 dez.)

HUFFORD, D. (1998). *Data warehouse quality*. Data Management Review, Feb./Mar. 1996. <http://www.data-warehouse.com/resource/articles/huffor7.htm> / <http://www.data-warehouse.com/resource/articles/huffor8.htm> (03 jan.).

- HYPERION. (1998). *The role of the OLAP server in a data warehousing solution*. <http://www.hyperion.com/whitepapers.cfm> (10 out.).
- IBM. (1998). *Business Intelligence success stories from IBM*. Business Intelligence - Customer Experiences. <http://direct.boulder.ibm.com/bi/info/customer.htm> (06 ago.).
- INMON, W. H. (1997). *Como construir o data warehouse*. Rio de Janeiro, Editora Campus.
- INMON, W. H.; HACKARTHORN, R. D. (1997). *Como usar o data warehouse*. Rio de Janeiro, Infobook / IBPI Press.
- ISHIKAWA, K. (1984). *Total quality control - estratégia e administração da qualidade*. São Paulo, IM&C Internacional.
- JURAN, J. M.; GRZYNA, F. (1993). *Quality analysis and planning*. New York, MacGraw Hill.
- KENAN TECHNOLOGIES (1995). *An introduction to multidimensional database technology*. s.l., White Paper by Kenan Systems Corporation.
- KIMBALL, R. (1997a). *Dangerous preconceptions*. DBMS Magazine, Aug. 1996. <http://www.dbmsmag.com/9608d05.html> (26 dez.).
- KIMBALL, R. (1997b). *A dimensional modeling manifesto*. DBMS Magazine, Aug. 1997. <http://www.dbmsmag.com/9708d15.html> (16 dez.).
- KIMBALL, R. (1998a). *Data warehouse toolkit*. São Paulo, Makron Books.
- KIMBALL, R. (1998b). *Is ER modeling hazardous to DSS?* DBMS Magazine, Oct. 1995. <http://www.dbmsmag.com/9510d05.html> (27 ago.).

- KIMBALL, R. (1998c). *Dealing with dirty data*. DBMS Magazine, Sept. 1996.
<http://www.dbmsmag.com/9609d14.html> (05 jan.).
- KIMBALL, R. (1998d). *It's time for data compression*. DBMS Magazine, Oct. 1996.
<http://www.dbmsmag.com/9610d05.html> (27 ago.).
- KIMBALL, R. (1998e). *Hackers, crackers and spooks*. DBMS Magazine, Apr. 1997.
<http://www.dbmsmag.com/9704d05.html> (27 ago.).
- KIMBALL, R. (1998f). *Digging into data mining*. DBMS Magazine, Oct. 1997.
<http://www.dbmsmag.com/9710d05.html> (27 ago.).
- KIMBALL, R. (1998g). *Meta meta data data*. DBMS Magazine, Mar. 1998.
<http://www.dbmsmag.com/9803d05.html> (27 ago.).
- KIMBALL, R. (1998h). *Professional boundaries*. DBMS Magazine, Jul. 1998.
<http://www.dbmsmag.com/9807d05.html> (27 ago.).
- KIMBALL, R. (1998i). *Help for dimensional modeling*. DBMS Magazine, Aug. 1998. <http://www.dbmsmag.com/9808d05.html> (27 ago.).
- KONDRATIUK, E. R. (1998). Data warehouse: detalhes que fazem a diferença. *Developers' Magazine*, n. 18, p. 22, fev.
- LAMBERT, B. (1997). *Data warehousing fundamentals: what you need to know to succeed*. Data Management Review, Mar. 1996.
<http://www.data-warehouse.com/resource/articles/lamber8.htm> (29 dez.)
- LANTIMES. (1998). *O perigo ronda as empresas*. São Paulo, n. 3, p. 4-6. Edição especial.

MANZONI Jr., R. (1997). O segredo da produtividade está no uso da informação. *Computerword*, 28 a 30 abr., p. 10-11.

MCCARTHY, V. (1997). *Strike it rich*. Datamation.
<http://www.datamation.com/workbench/datamine/stories/rich.htm> (18 dez.)

MCELREATH, J. (1997). *Data warehouses: an architectural perspective*. Perspectives, Oct. 1995. http://www.csc.com./about/tech_dw_arch.html (29 dez.).

MENCONI, D. (1998). A mineração de informações. *Info Exame*, n. 144, p. 92-93, mar.

MINTZBERG, H.; RAISINGHANI, D.; THEORET, A. (1976). The structure of unstructured decision process. *Administrative Science Quarterly*, v. 21, n. 2, p. 246-275.

MOAD, J. (1997). *Buried info treasure*. PC Week, 13 Oct. 1997.
<http://www.zdnet.com/pcweek/news/1013/13erp.html> (22 dez.).

MOREIRA FILHO, T. R. (1998). On-line analytical processing server (servidor OLAP). *Developers' Magazine*, n.18, p. 28-29, fev.

MOXON, B. (1998). *Defining data mining*. DBMS, Data Warehouse Supplement, Aug. 1996. <http://www.dbmsmag.com/9608d53.html> (05 ago.).

NIMER, F. (1998). Analisando o retorno sobre o investimento de data warehouse. *Developers' Magazine*, n. 18, p. 16-17, fev.

NIMER, F.; SPANDRI, L. C. (1998). Obtendo vantagem competitiva com o uso de data mining. *Developers' Magazine*, n. 18, p. 30-31, fev.

- NUTT, P. C. (1993). The identification of solutions ideas during organizational decision making. *Management Sciences*, v. 39, n. 9, p. 1071-1089.
- PENDSE, N. (1998a). *Virtual OLAP: getting real*. Datamation, Sept. 1997.
<http://www.datamation.com/plugin/workbench/olap/stories/virt.html> (31 jan.).
- PENDSE, N. (1998b). *OLAP omnipresent*. Byte, Feb. 1998.
<http://www.byte.com/art/9802/sec6/art7.htm> (19 ago.).
- PENDSE, N. (1998c). *Database explosion*. The OLAP Report, 1998.
<http://www.olapreport.com/DatabaseExplosion.htm> (19 ago.).
- PEREIRA, M. J. L. B.; FONSECA, J. G. M. (1997). *Faces da decisão: as mudanças de paradigmas e o poder da decisão*. São Paulo, Makron Books.
- PEARSON, J. M.; SHIM, J. P. (1995). An empirical investigation into DSS structures and environments. *Decision Support Systems*, n. 13, p. 141-158.
- PILOT SOFTWARE. (1998). *An introduction to data mining*.
<http://ww.pilotsw.com/dmpaper/dmindex.htm> (16 jan.).
- POWER, D. (1997). *A brief history of Decision Support Systems*.
<http://power.cba.uni.edu/isworld/dsshistory.html> (19 dez.).
- RADEN, N. (1997a). *Data, data everywhere*. Information Week, 30 Oct. 1995.
http://members.aol.com/nraden/iw_mct01.htm (26 dez.).
- RADEN, N. (1997b). *Modeling the data warehouse*. Information Week, Jan. 1996.
http://member.aol.com/nraden/iw0196_1.htm (30 dez.).
- RADEN, N. (1998a). *Maximizing your warehouse*. Information Week, Mar. 1996.
<http://techweb.cmp.com/iw/571/71olwar.htm> (08 jan.).

- RADEN, N. (1998b). *Warehouses and the Web*. Information Week, May 1996. <http://techweb.cmp.com/iw/579/79olweb.htm> (15 jan.).
- RADEN, N.; PETERSON, M. (1998). *Who's minding the cache?* Information Week, Jun. 1997. <http://techweb.cmp.com/iw/637/37oldat.htm> (15 jan.).
- SABHERWAL, R.; GROVER, V. (1989). Computer support for strategic decision making process: review and analysis. *Decision Sciences*, v. 20, n. 1, p. 54-76.
- SAHIN, K. (1997). *Multidimensional database technology and data warehousing*. Database Journal, Dec. 1995. http://www.kenan.com/acumate/byln_mdw.htm (31 dez.).
- SEIDLER, C. (1995). Dez mandamentos. *Computerworld*, 10 jul., p. 18-23.
- SIMON, A. (1998). *Better clients, better decisions*. Byte, Jan. 1997. <http://www.byte.com/art/9701/sec7/art3.htm> (19 ago.).
- SIMOUDIS, E. (1997). *Data mining: a technology comes of age*. Software Quaterly, IBM, vol. 24, 1995. <http://www.software.ibm.com/sq/issues/vol24/datatech.htm> (29 dez.).
- SMALL, R. D. (1998). *Debunking data mining myths*. Information Week, 20 Jan. 1997. <http://techweb.cmp.com/iw/614/14oldat.htm> (02 jan.).
- SPRAGUE, R. H.; WATSON, H. J. (1991). *Sistema de Apoio à Decisão: colocando a teoria em prática*. Rio de Janeiro, Campus.
- STRANGE, K. (1998). *Can data marts grow?* CIO Magazine, 1 Jul. 1997. http://www.cio.com/archive/070197_gartner_content.html (05 jan.).

- STREHLO, K. (1998). *Data warehousing: avoid planned obsolescence*. Datamation, 15 Jan. 1996.
<http://www.datamation.com/plugin/issues/1996/jan15/01bsw200.html> (03 jan.).
- SULAIMAN, A.; SOUZA, J. M. (1997). Prospecção de conhecimento em banco de dados. *Developers' Magazine*, n. 6, p. 38-39, fev.
- SULAIMAN, A.; SOUZA, J. M. (1998). Representação do conhecimento em data warehouse. *Developers' Magazine*, n. 20, p. 42-43, abr.
- SUZAKI, K. (1993). *The new shop floor management*. New York, The Free Press.
- SWAMI, A. (1998). *Data mining with Silicon Graphics technology*. Silicon Graphics, 1995. <http://www.sgi.com/Technology/data-mining.html> (16 jan.).
- SZLEJF, M. (1997). *Computerworld technology breakfast: data warehouse*. Computerworld, set. 1997.
<http://www.idg.com.br/computerworld/computerworld/228/debate.htm> (28 dez.).
- TAURION, C. (1997). Data warehouse: estado de arte e estado de prática. *Developers' Magazine*, n. 6, p. 10-11, fev.
- TAURION, C. (1998a). Data warehouse: vale a pena gastar milhões investindo em um? *Developers' Magazine*, n. 18, p. 10-11, fev.
- TAURION, C. (1998b). O data warehouse será útil para a sua organização?. *Developers' Magazine*, n. 18, p. 26-27, fev.
- THE, L. (1998). *OLAP answers tough business questions*. Datamation, May 1995.
<http://www.datamation.com/plugin/workbench/olap/stories/05aev.html> (31 jan.).

- TKACH, D. S. (1998). *Information mining with the IBM Intelligent Miner family*. IBM, n. 1, Feb. 1998.
<http://www.software.ibm.com/data/iminer/foridata/download/whitefam3.html> (18 jul.).
- TYO, J. (1998). *Viewing data your way*. Information Week, 8 Jul. 1996.
<http://techweb.cmp.com/iw/587/87olpro.htm> (15 jan.).
- UNITED STATES DEPARTMENT OF THE INTERIOR. (1998). *Decision process guidebook*. <http://www.usbr.gov/Decision-Process/execsum.htm> (16 out.).
- URIS, A.(1989). *O livro de mesa do executivo*. São Paulo, Pioneira.
- VILELA, E. (1997). *Computerworld technology breakfast: prévia data warehouse*. Computerworld, 1997. Suplemento Especial.
<http://www.computerworld.com.br/polemica/data.htm> (28 dez.).
- WELDON, J. L. (1998). *A career in data modeling*. Byte, Jun. 1997.
<http://www.byte.com/art/9706/sec7/art3.htm> (19 ago.).
- WEISS, S. M.; INDURKHYA, N. (1997). *Predictive data mining: a practical guide*. s.l., Morgan Kaufmann.
- WESTPHAL, C.; BLAXTON, T. (1998). *Data mining solutions*. s.l., John Wiley.
- WU, X. (1995). *Knowledges acquisition from databases*. s. l., Ablex.

BIBLIOGRAFIA COMPLEMENTAR

- ABERDEEN GROUP (1998). *Data warehouse query tool: evolving to relational OLAP*. Market Viewpoint, vol. 8, n. 8, 7 Jul. 1995.
http://www.strategy.com/dwf/aber_.htm (28 jan.).
- ATRE, S. (1998). *Selecting end-user decision support tool: five key issues*. Atre Associates, 1996. <http://www.iqsc.com/products/analyst/5keytxt.htm> (28 jan.).
- BALLINGER, C.; OLSON, M. (1998). *High noon for TPC-D*. Database Programming & Design, Sept. 1997. <http://www.dbpd.com/9709tpcd.htm> (06 jan.).
- BAUM, D. (1998). *Warehouse mania*. Lantimes, 20 Nov. 1995.
<http://www.wcmh.com/lantimes/95nov/511b067a.html> (15 jan.).
- BAUM, D. (1998). *Planning and implementing a data warehouse*. Byte, Jun. 1997.
<http://www.byte.com/art/9706/sec20/art1.htm> (19 ago.).
- BAUMANN, P. (1998). *Sorting terabytes: multidimensional database systems provide greatly improved performance with large data volumes*. Byte, Jan. 1998.
<http://www.byte.com/art/9801/sec17/art3.htm> (19 ago.).
- BOHN, K. (1997). *Converting data for warehouses*. DBMS Magazine, Jun. 1997.
<http://www.dbmsmag.com/9706d15.html> (26 dez.).

BROOKS, P.; WOLLENWEBER, C. (1998). *Reporting against large databases*. DBMS Magazine, Aug. 1996. <http://www.dbmsmag.com/9608d55.html> (11 jan.).

CAMPOS, A. (1997). *OLAP: segmento em plena transformação*. Computerworld, abr. 1997. <http://www.idg.com.br/computerworld/206/coluna/206col01.htm> (28 dez.).

COMPUTERWORLD. (1997). *A construção passo a passo*. Rio de Janeiro, p. 6, fev. Suplemento Especial - Data Warehouse.

COMPUTERWORLD. (1998). *Integrando o pool de conhecimento*. Rio de Janeiro, jan. Suplemento Especial.

DARLING, C. (1998). *Dig deep to strike data access gold*. Datamation, 1 Nov. 1995. <http://www.datamation.com/plugin/issues/1995/nov1/11aev100.html> (05 jan.).

DARLING, C. (1998). *Think outside the OLAP box*. Datamation, Apr. 1996. <http://www.datamation.com/PluIn/workbench/olap/stories/04beval1.htm> (31 jan.).

DARLING, C. (1998). *Manage your reporting environment*. Datamation, 1 May 1996. <http://www.datamation.com/plugin/issues/1996/may1/05aevall1.html> (07 jan.).

DARLING, C. (1998). *How to integrate your data warehouse*. Datamation, 15 May 1996. <http://www.datamation.com/plugin/issues/1996/may15/05beval1.html> (30 dez.).

DARLING, C. (1998). *Datamining for the masses*. Datamation. <http://www.datamation.com/workbench/datamine/stories/masses.htm> (15 jan.).

- DARLING, C. (1998). *Ease implementation woes with packaged datamarts*. Datamation, Mar. 1997. <http://www.datamation.com/plugin/workbench/dmarts/stories/ease.htm> (05 jan.).
- DATAMIND TECHNOLOGY CENTER. (1998). *A model for effective customer-oriented market plans: an Arbor Software & DataMind white paper*. http://www.datamindcorp/paper_comp.html (16 jan.).
- DATAMIND TECHNOLOGY CENTER. (1998). *Agent network technology*. http://www.datamindcorp/paper_agentnetwork.html (16 jan.).
- DEMAREST, M. (1998). *Improving data legibility in Decision Support Systems*. DBMS Magazine, May 1994. <http://vista.hevanet.com/demarest/marc/schema.htm> (02 jan.).
- DEMAREST, M. (1998). *Building the data mart*. DBMS Magazine, Jul. 1994. <http://vista.hevanet.com/demarest/marc/marts.htm> (02 jan.).
- DEMAREST, M. (1998). *A data warehousing evaluating model*. Oracle Technical Journal, Oct. 1995. <http://vista.hevanet.com/demarest/marc/oracle7.htm> (02 jan.).
- GLYMOUR, C.; MADIGAN, D.; PREGIBON, D.; SMYTH, P. (1997). *Statistical themes and lessons for data mining*. *Data Mining and Knowledge Discovery*, n. 1, p. 11-28.
- GREENFIELD, L. (1998). *Don't let data warehousing gotchas getcha*. Datamation, Mar. 1 1996. <http://www.datamation.com/plugin/issues/1996/mar1/03asoft1.html> (05 jan.).
- HACKNEY, D. (1997). *Incremental data mart enterprise architecture*. Data Management Review, May 1997. http://www.entergroupltd.com/columns/5-97-1_enterprise_architecture.htm (31 dez.).

- HILL, B. (1998). *Sizing parallel servers today*. Database Programming & Design, Sept. 1997. <http://www.emergent.com/epress/earticle/psize.html> (12 jan.).
- HOFLAND, P.; USTLER, J. (1998). *Data mining at your desk*. Byte, Jul. 1997. <http://www.byte.com/art/9707/sec17/art1.htm> (19 ago.).
- INMON, W. H. (1998). *The data warehouse budget*. Data Management Review. <http://www.data-warehouse.com/resource/articles/inmon1.htm> (13 jan.).
- INMON, W. H. (1998). *Does your datamart vendor care about your architecture?* Datamation, Mar. 1997. <http://www.datamation.com/PlugIn/workbench/dmarts/stories/care.htm> (26 ago.).
- INMON, W. H. (1998). *What is a data warehouse?* Prism, v. 1, n. 1, 1997. http://www.cait.wustl.edu/cait/papers/prism/vol1_no1 (22 jul.).
- KIMBALL, R. (1997). *Mastering data extraction*. DBMS Magazine, Jun. 1996. <http://www.dbmsmag.com/9606d05.html> (30 dez.).
- KIMBALL, R. (1998). *Factless fact tables*. DBMS Magazine, Sept. 1996. <http://www.dbmsmag.com/9609d05.html> (27 ago.).
- KIMBALL, R. (1998). *Casual (not casual) dimensions*. DBMS Magazine, Nov. 1996. <http://www.dbmsmag.com/9611d05.html> (27 ago.).
- KIMBALL, R. (1997). *Features for query tools*. DBMS Magazine, Feb. 1997. <http://www.dbmsmag.com/9702d05.html> (31 dez.).
- KIMBALL, R. (1998). *Relocating the ODS*. DBMS Magazine, Dec. 1997. <http://www.dbmsmag.com/9712d05.html> (27 ago.).

- KIMBALL, R. (1998). *Bringing up supermarts*. DBMS Magazine, Jan. 1998.
<http://www.dbmsmag.com/9801d14.html> (27 ago.).
- LAMBERT, B. (1997). *Break old habits To define data warehousing requirements*.
Data Management Review, Dec. 1995.
<http://www.data-warehouse.com/resource/articles/lamber11.htm> (26 dez.)
- LEWISON, L. (1997). *Data mining: intelligent technology gets down to business*. PC
AI, Nov./Dec. 1993. http://axp16.iie.org.mx/Monitor/v01n03/ar_sbc1.htm (22
dez.).
- MANNI, L. C.; DORSA, L. F. A. (1998). Data warehouse: gerenciando a qualidade
dos dados. *Developers' Magazine*, n. 18, p. 20, fev.
- MEREDITH, M. E.; KHADER, A. (1998). *Divide and aggregate: designing large
warehouses*. Database Programming and Design, Jun. 1996.
<http://www.dbpd.com/khader.htm> (05 jan.).
- MUNDY, J. (1997). *From the ground up: building a data warehouse*. Sybase Server
Magazine, Fall, 1995.
<http://www.leland.stanford.edu/group/da/sybasepub.html> (31 dez.).
- PALMA, S. (1998). Os componentes funcionais de um data warehouse. *Developers
Magazine*, n. 18, p. 18-19, fev.
- PARSAYE, K. (1997). *Surveying decision support: new realms of analysis*. Database
Programming & Design, Apr. 1996. <http://www.dbpd.com/parsaye.htm> (27 dez.).
- PAUL, L. G. (1998). *Anatomy of a failure*. CIO Magazine, Nov. 1997.
http://www.cio.com/archive/enterprise/111597_data_content.html (05 jan.).

- RADEN, N. (1998). *Push back in push technology*. Internet Systems Supplement, Nov. 1997. <http://www.dbmsmag.com/9711i15.html> (21 ago.).
- RUDIN, K. (1998). *The habits of successful MPP sites*. DBMS, Oct. 1995. <http://www.emergent.com/epress/earticle/mpp.html> (14 jan.).
- RUDIN, K. (1998). *Practical advice on implementing parallel processing systems*. Database Programming & Design, Dec. 1995. <http://www.emergent.com/epress/earticle/dppd.html> (14 jan.).
- SANTOS, J. (1997). Bons negócios com data warehouse. *Computerworld*, 11 ago., p. 6-7.
- SAS INSTITUTE (1998). *OLAP tools and techniques within the SAS system*. SAS Institute White Paper. <http://www.sas.com/software/olap/whitepaper/technology.htm> (28 jan.).
- SAYLOR, M. J.; ACHARYA, M. G.; MOIN, A. (1997). *Data warehouse application administration: five questions IS managers should ask vendors*. Data Management Review, Jun. 1997. http://www.strategy.com/DWF/DMR_5questions.html (22 dez.).
- SCHWARZKOPF, A. B. (1998) The virtual data warehouse for small business. In: AMERICAS CONFERENCE ON INFORMATION SYSTEMS, 3., Indiana, 1997. *Proceedings*. Indianapolis, Association for Information Systems, 1997. <http://hsb.baylor.edu/ramsower/ais.ac.97/papers/schwarz.htm> (26 jan.).
- TIEZZI, G. (1997). O planejamento estratégico da informação. *Developers' Magazine*, n. 6, p. 24-25, fev.
- TYO, J. (1998). *Slicing data on the desktop*. Information Week, 15Jul. 1996. <http://techweb.cmp.com/iw/588/88olola.htm> (15 jan.).

WHITE, C. J. (1998). *The IBM business intelligence software solution*. IBM, 1998.
<http://www.software.ibm.com/data/pubs/papers/bisolution/index.html> (18 jul.).

WILLIAMS, J. (1997). *Tools for traveling data*. DBMS, Jun. 1997.
<http://www.dbmsmag.com/9706d16.html> (31 dez.).

Apêndice 1- Alguns Produtos de Data Warehouse, OLAP e Data Mining

Com base nos dados apresentados CAMPOS (1997a), CARVALHO (1997b), EDELSTEIN (1997) e COMPUTERWORLD (1997a e 1998a), são apresentados os principais produtos para o data warehouse, o OLAP e o data mining. No data warehouse, são apresentados produtos para as diversas ferramentas que foram referenciadas no tópico 3.3. Os autores afirmam, em seus artigos, que os dados apresentados foram fornecidos pelas empresas que desenvolvem os respectivos produtos. O nome que se encontra entre parêntesis, ao lado do nome do produto, é o nome da empresa que o desenvolveu ou de seu fornecedor.

1 - Produtos para data warehouse

1.1 - Produtos para a extração e transformação de dados

Info Pumb (Platinum Technology)

Faz a extração, refinamento e distribuição dos dados. Realiza a replicação de dados bidirecional e em ambientes de data warehouse, data mart e cliente/servidor.

Prism Warehouse Executive (Prism Solutions)

Produto para extração, transformação, integração e conversão de arquivos e dados das bases operacionais para o data warehouse.

Visual Warehouse (IBM)

Permite extrair e transformar uma grande variedade de dados usando SQL e armazená-los em bancos de dados.

1.2 - Produtos para limpeza e refinamento de dados

Data Provider (OSC)

Gerenciador de qualidade de dados, que permite extração e limpeza dos dados.

Info Refiner (Platinum Technology)

Automatiza a extração, refinação e movimentação de grandes volumes de dados legados e estruturas seqüenciais para o ambiente cliente/servidor.

Prism Quality Manager (Prism Solutions)

Proporciona a habilidade de auditar, monitorar, melhorar e certificar a qualidade dos dados em pontos-chave do ciclo de vida do data warehouse.

1.3 - Produtos para replicação ou transferência de dados

Data Bridge (Attachmate)

Servidor de replicação para bases de dados relacionais em servidores Windows NT ou Unix. Automatiza o processo nos bancos de dados Oracle, Red Brick, Informix e outros. Permite filtragem dos dados e atualizações.

InfoTransport (Platinum Technology)

Proporciona alta velocidade para replicação de dados, quando utilizado em conjunto com o produto InfoRefiner do mesmo fabricante.

1.4 - Produtos para armazenamento de dados

CA-Ingres (Computer Associates)

Banco de dados relacional dedicado a sistemas abertos, com suporte a TCP/IP para comunicação aberta.

DB/2 (IBM)

Solução desenhada para um gerenciamento de banco de dados para suporte à decisão.

Enterprise Connect (Sybase)

Família de produtos que simplifica o acesso a múltiplas fontes de dados em sistemas heterogêneos. Conectividade, replicação, mapeamento de objetos e software de messaging unificam as fontes de informações distribuídas em ambiente único de modo que os formatos e a localização dos dados sejam transparentes para o usuário do data warehouse.

Online Dynamic Server (Informix Software)

Gerenciador de banco de dados relacional, baseado na arquitetura escalável. Apresenta recursos que fornecem alta disponibilidade de informações e alta performance na execução de consultas complexas sobre grandes volumes de dados.

Oracle Universal Server (Oracle)

Um conjunto de produtos que viabiliza o gerenciamento dos dados. O gerenciador de banco de dados oferece uma série de componentes que podem ser escolhidos pelo usuário de acordo com os seus aplicativos.

Power Designer (Sybase)

Solução de modelagem de dados integrada que possibilita a criação de modelos baseados em conceitos comuns e geração de aplicações.

SAS System (SAS Institute)

Kit de produtos escaláveis desenvolvido para gerenciamento e manutenção de sistemas corporativos de informações.

SQL Server (Sybase)

Banco de dados para data warehouses.

1.5 - Produtos para gerenciamento e administração**Pinecone Content Tracker (Pinecone Systems)**

Monitora o conteúdo do data warehouse a partir de métricas pré-definidas e de consultas.

Pinecone Cost Tracker (Pinecone Systems)

Aloca custos fixos e variáveis aos usuários, a partir da monitoração do acesso e utilização do ambiente.

Pinecone Refreshment Tracker (Pinecone Systems)

Produto destinado aos administradores, é capaz de monitorar o grau de utilização do data warehouse a partir de rotinas manuais ou automatizadas.

Pinecone Usage Tracker (Pinecone Systems)

Monitora o acesso e uso do data warehouse, permitindo a identificação dos dados não utilizados e dos gargalos do sistema.

Prism Schedule Manager (Prism Solutions)

Gerenciador de Schedules, que cria um mapa das rotinas de produção e o transporte de arquivos de fontes operacionais até o data warehouse ou o data mart.

SourcePoint (Software AG)

Gerencia o processo de extração, transformação e carga dos dados no ambiente do data warehouse. Permite processamento paralelo.

1.6 - Produtos para geração de Consultas e Relatórios

Brio Query Enterprise Server (Brio Technology)

Servidor de consultas, permitindo agendamento e acesso, via cliente/servidor ou Web, a data warehouses e data marts.

Crystal Info (Seagate)

Permite elaborar consultas e relatórios, possibilita agendamento para processamento automático.

Crystal Reports (Seagate)

Permite elaborar consultas e relatórios através de uma galeria de modelos.

Forest & Trees (Platinum Technology)

Possibilita a criação de relatórios, consultas e análises gerenciais das informações.

IQ/Objects (IQ Software)

Gerador inteligente de relatórios com capacidades gráficas e publicação na Internet.

Impromptu (Cognos)

Possibilita a construção de consultas interativas com a visão do BD orientado a negócios.

InfoMaker (Sybase)

Gerador de relatórios.

SAS Access (SAS Institute)

Possui mecanismos de acesso aos principais bancos de dados, sejam eles relacionais ou não.

1.7 - Produtos para metadados

Prism WareHouse Directory (Prism Solutions)

Produto que permite a criação e manutenção de metadados técnicos, operacionais e de negócio, com integração, armazenamento e acesso pelos usuários e administradores.

Repository (Platinum Technology)

Conjunto de produtos para a administração de metadados corporativos, definindo o ambiente de migração, mapeamento de origens e alvos dos dados.

2 - Produtos OLAP

Acumen / Acumate (Kenan Systems)

Desenvolvida para grandes projetos de carga e manipulação de dados. Permite a construção de aplicações complexas, mas requer programadores capacitados. Possibilita análise multidimensional aleatória.

Brio Query Explorer (Brio Techonology)

Produto destinado aos administradores do ambiente do data warehouse, possibilita a geração de relatórios, consultas, gráficos e análises multidimensionais.

Brio Query Navigator (Brio Techonology)

Produto destinado aos usuários do ambiente data warehouse, permite o acesso a relatórios, consultas, gráficos e análises multidimensionais OLAP.

Business Objects (Business Objects)

Sistema integrado para geração de consultas, relatórios, atividades OLAP e para data mining. Possui geração automática de SQL, cria metadados, integra-se a sistemas ERP e possui recursos para a Internet.

CFI (Costumer Focus Inc.)

Família de produtos integrados voltada ao mercado financeiro. Facilita as análises de lucros, de riscos, a performance e gerencia as campanhas de marketing.

DSS Agent (IQ Software)

Destinado à análise multidimensional com acesso direto aos bancos de dados relacionais.

DSS Designer (Pilot Software)

Utilizada para desenvolvimento de interfaces amigáveis, modelagem e construção de bancos de dados multidimensionais.

DSS Analysis Server (Pilot Software)

Ferramenta destinada ao acesso e compartilhamento de data marts, permite a extração de dados de banco de dados relacionais.

Esperant (Speedware)

Destinado à geração de consultas e relatórios OLAP, permitindo o acesso a múltiplas bases de dados, criação de visualizações personalizadas dos dados e relatórios flexíveis.

Essbase (Arbor)

Servidor OLAP que permite simultaneamente armazenar, partilhar, analisar e atualizar os dados, utilizando um número ilimitado de dimensões.

Holos (Seagate)

Produto para o desenvolvimento de aplicações OLAP em ambiente cliente/servidor, fornecendo flexibilidade e escalabilidade.

IQ/Vision (IQ Software)

Produto para análise multidimensional, que permite ao usuário verificar tendências e diferentes formas de visualização da informação.

MetaCube ROLAP Option (Informix)

Possui os principais recursos OLAP, como transparência de dados agregados e amostragem estatísticas de dados, realiza comparações complexas e cálculos avançados.

Media MR (Speedware)

Produto HOLAP, que associa as características ROLAP e MOLAP, permitindo carregar dados de tabelas relacionais automaticamente para o cubo multidimensional.

Oracle Discoverer (Oracle)

Produto ROLAP para extração de dados e consultas *ad hoc*, orientada a negócios.

Oracle Express Server e Oracle Express Personal (Oracle)

Ambos são produtos MOLAP indicados para data marts de áreas específicas. Permitem análises de informações sob diversas perspectivas.

Oracle Express Analyser (Oracle)

Permite ao usuário realizar consultas e análises.

PowerPlay (Cognos)

Produto para análise multidimensional interativa. Suporta grandes volumes de dados e permite a navegação direta e diversas formas de apresentação dos dados.

Relational Access Manager e Relational Access Administrator (Oracle)

Ambos produtos fazem a comunicação entre os bancos de dados relacionais e multidimensionais, permitindo que a informação venha de qualquer tipo de fonte.

VentoMap (Vento Software)

Desenvolvido para empresas de telecomunicações e de varejo, possibilita a análise dos indicadores de performance e realiza projeções e análises de tendências.

3 - Produtos para data mining

4Thought (Cognos)

Foi projetado para construir modelos de regressão e série temporais, embora também possa ser usado para classificação. Utiliza redes neurais para construir estes modelos. O desenvolvimento e as capacidades da análise de modelos do produto são extensas, possui muitos recursos disponíveis para se trabalhar em análises de séries temporais. Os resultados podem ser exportados para vários formatos, incluindo o Microsoft Excel e o Lotus 1-2-3.

BusinessMiner (Business Objects)

Possibilita que usuários não técnicos encontrem relacionamentos previamente não detectados entre seus dados corporativos.

Darwin (Thinking Machines)

Conjunto de produtos orientado para classificação e regressão. O StarTree constrói uma árvore de decisão, o StarNet cria modelos que usam redes neurais e algoritmo genético, o StarMatch produz modelos que usam o algoritmo nearest neighbor e o StarData é usado para preparação e análise de dados. Os algoritmos possuem o recurso de processamento paralelo para gerenciar grandes quantidades de dados.

DataBase Mining Marksman (HNC Software)

Projetado para trabalhar com bancos de dados de aplicações comerciais. É vendido como uma combinação de hardware e software. O componente de hardware é um PC standard, com uma placa aceleradora que contém 16 processadores paralelos e permite ao produto, de um forma rápida e automática, construir muitas redes neurais com arquiteturas diferentes para se selecionar a melhor. O produto descobre relacionamentos entre atributos, computando fortes relacionamentos entre

todos os pares de campos. Isto é útil para explorar dados e identificar colunas altamente correlacionadas.

DataCruncher (DataMind)

Projetado para prever quais clientes podem vir a deixar de consumir os produtos da empresa. Foi desenvolvido especialmente para a indústria de telecomunicações. É uma ferramenta cliente/servidor que usa uma técnica proprietária de construção de modelos chamada Agent Network Technology. Constrói modelos de classificação usando árvore de decisão. Não suporta paralelismo.

Data Mining Solution (SAS Institute)

É um módulo de sistema da SAS para data mining. O produto provê uma interface com um conjunto extenso de opções para construir os modelos. Inclui as aplicações da Rede Neural SAS e da Árvore de Decisão SAS. Antes de gerar o modelo, pode-se explorar os dados, através da ferramenta SAS Insight Visualization.

Decision Series (NeoVista Solutions)

É um conjunto de produtos para data mining que provêm diferentes modelos e tecnologias. O DecisionNet utiliza a técnica de redes neurais que é usada para classificação e regressão. O DecisionAR pode ser usado para associação e descoberta de sucessões. O DecisionCL é usado para realizar agrupamentos. O DecisionAccess provê a função de preparação de dados.

//Discovery (HyperParallel)

O //Discovery (pronuncia-se Parallel Discovery) é um conjunto de produtos de data mining para classificação, regressão, agrupamento, associação e sequenciamento. Provê uma interface de comando de linha, projetada para ser usada através de uma equipe da HyperParallel, treinada para construir as aplicações para os clientes. Por conseguinte, a empresa vende seu conjunto de produtos empacotado com seus serviços.

DSS Discovery Server (Pilot Software)

Produto voltado à realização de análises preditivas e data mining.

Intelligent Miner (IBM)

É um conjunto de produtos para data mining, para realizar classificação, associação e descoberta de sucessões, série temporais, agrupamento e regressão. A maioria dos algoritmos foram desenvolvidos para suportar o processamento paralelo. Trabalha em conjunto com o DB2, um banco de dados desenvolvido também pela IBM, porém, o produto suporta outras fontes de dados.

KnowledgeSeeker (Angoss Software)

Produto desktop ou cliente/servidor que usa árvores de decisão para construir modelos preditivos, através de uma interface agradável. Os recursos interativos deixam os usuários explorarem os dados, dividindo-os em nodos selecionados na árvore de decisão ou forçando uma divisão particular que poderia ser interessante. Os usuários também podem desdobrar o modelo, exportando as regras descobertas para um editor de textos ou para um gerenciador de consultas SQL.

MineSet (Silicon Gráficos)

Conjunto de produtos para data mining que combina os modelos de classificação e de associação com a técnica de visualização. Nele estão incluídos, o Gerador de Regra de Associação, a Árvore de Decisão para se realizar a classificação e um utilitário para a determinação do grau de importância das colunas.

Pattern Recognition Workbench (Unica Technologies)

Conjunto de produtos para construir modelos de classificação, agrupamento, série temporais e modelos de regressão. Além de construir os modelos com redes neurais, provê algoritmos para regressão logística e regressão linear. O produto provê uma interface no estilo de planilha eletrônica. Os dados devem ser importados para uma ou mais planilhas eletrônicas; logo após, são preparados para se realizar a mineração de dados com um extenso conjunto de funções do produto, que gerará modelos alternativos automaticamente e procurará a melhor solução. Também provê

uma variedade de recursos de visualização para monitorar a construção do modelo e interpretar os resultados.

Scenário (Cognos)

Produto para data mining que permite identificar relacionamentos ocultos nos dados.

GLOSSÁRIO

Devido ao fato de o assunto abordado nesta dissertação abranger diversos conceitos e tecnologias, tanto da área da Administração como da área de Informática, julgou-se inviável criar-se um glossário contendo a descrição dos termos que abrangem o referido assunto. Na área tecnológica existe, também, a volatilidade dos termos, ou seja, com os rápidos avanços desta área, novos termos surgem constantemente, outros tornam-se obsoletos rapidamente e alguns termos normalmente mudam de significado adaptando-se às novas tecnologias.

Levando-se em consideração que existem diversos glossários eletrônicos, publicados por conceituadas entidades, que são atualizados freqüentemente e estão disponíveis na Internet, resolveu-se citá-los aqui, para que se possa, por meio do acesso aos mesmos, ter-se sempre disponível os principais termos empregados neste assunto e com os seus significados sempre atualizados. Seguem-se a relação de endereços eletrônicos:

- ⇒ Acronym Glossary - <http://observer.gsfc.nasa.gov/sec2/glossary.html>
- ⇒ DSS Glossary of Terms - <http://dss.cba.uni.edu/glossary/dssglossary.html>
- ⇒ DSS Glossary List - <http://www.adpc.purdue.edu/DSS/glossary.htm>
- ⇒ SIPS Technical Glossary - <http://www.state.nc.us/SIPS/glossary/glossmp.htm>
- ⇒ Student Computing Guide 98 - <http://www.qut.edu.au/scg/glossary.html>
- ⇒ Intranet Glossary - <http://www.pb.uillinois.edu/Intranet/html/glossary.htm>
- ⇒ University of Oregon Data Warehouse Glossary -
<http://oregon.uoregon.edu/~sisdwh/glossary.html>

- ⇒ OLAP Council definitions - <http://www.olapcouncil.org/research/glossaryly.htm>
- ⇒ The OLAP Report Glossary - <http://www.olapreport.com/glossary.htm>
- ⇒ Glossary of Terms - <http://www.sba.udayton.edu/hoffer/mis385/glossary.htm>
- ⇒ Dimensional Systems Glossary of OLAP - <http://www.dimsys.com/glossary.html>
- ⇒ Data Warehouse Glossary - IBM - <http://publib.boulder.ibm.com:80/cgi-bin/bookmgr/BOOKS/SG242238/GLOSSARY>
- ⇒ Industry Jargon Glossary - <http://www.d2k.com/gloss.htm>
- ⇒ Pilot Software Glossary of Data Mining -
http://wwwpilotsw.com/r_and_t/whtpaper/datamine/dmglos.htm
- ⇒ Two Crows Data Mining Glossary - <http://twocrows.com/glossary.htm>
- ⇒ Datamation OLAP Definitions -
<http://www.datamation.com/PlugIn/workbench/olap/stories/04bevgl.htm>
- ⇒ Datamation Data Warehouse Glossary -
<http://www.datamation.com/PlugIn/workbench/dwhouse/dwgloss.htm>
- ⇒ Oracle OLAP Glossary - <http://www.oracle.com/products/olap/html/glossary.html>
- ⇒ Data Warehouse Survival Kit -
http://www.data-warehouse.com/survival/k_glossary.htm
- ⇒ BusinessObjects Glossary - <http://tech.busobj.com/services/glossary.htm>
- ⇒ Datasage Glossary -
<http://www.ardentsoftware.com/datasage/partner.glossary.htm>
- ⇒ MicroStrategy Data Warehouse Glossary -
http://www.strategy.com/dw_forum/glossary.htm
- ⇒ sas.com interactive Glossary -
<http://www.sas.com/software/sas.com/current/glossary/dwdef.html>
- ⇒ SAS Institute Glossary: Data Mining -
http://www.sas.com/new/newsroom/crm/gloss_dm.html
- ⇒ Executive Information Systems Glossary - <http://wwwbcs.co.za/eis/glossary.htm>
- ⇒ Glossary of Data Busswords - <http://www.acxiom.com/glossary.htm>
- ⇒ Glossary of DataSet Terms - <http://www.ds-dataset.com/Glossary.html>
- ⇒ AIXOMA Glossary - http://www.axioma.co.at/research/gl_content.html
- ⇒ P-Six Logic Glossary - <http://www.psixlogic.com/glossary.htm>

- ⇒ CODA Glossary - <http://www.coda-financials.com/glossary/glossarya.htm>
- ⇒ Glossary of OLAP - <http://www.applix.com/tm1/OLAP/olapglos.htm>
- ⇒ Glossary - <http://www.kivaproductions.com/glossary.html>
- ⇒ Glossary - <http://www.survey.com/glossary.html>
- ⇒ Glossary - <http://nulook.uneb.edu/about/glossary.html>
- ⇒ Glossary - <http://www.sesh.com/Guide/Glossary.html>
- ⇒ Steve Elkins OLAP Glossary - <http://altaplana.com/olap/glossary.addendum.txt>
- ⇒ PC Glossary - <http://www.funpc.com/glossary.htm>