

UFPB - UNIVERSIDADE FEDERAL DA PARAÍBA

CCT - CENTRO DE CIÊNCIAS E TECNOLOGIA

DSC - DEPARTAMENTO DE SISTEMAS E COMPUTAÇÃO

COPIN - COORDENAÇÃO DE PÓS-GRADUAÇÃO EM INFORMÁTICA

MESTRADO EM INFORMÁTICA

DISSERTAÇÃO DE MESTRADO

AMBIENTE COMPUTACIONAL PARA GERÊNCIA ESTRATÉGICA -  
DATA WAREHOUSING: UMA SURVEY

AÍDRE DA CUNHA GUEDES DE SOUZA  
(MESTRANDO)

PROF. DR. ULRICH SCHIEL  
(ORIENTADOR)

SAPIENTIA AEDIFICAT

CAMPINA GRANDE  
ABRIL / 1999

**AÍDRE DA CUNHA GUEDES DE SOUZA**

**AMBIENTE COMPUTACIONAL PARA GERÊNCIA ESTRATÉGICA -  
DATA WAREHOUSING: UMA SURVEY**

**Dissertação apresentada à  
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM  
INFORMÁTICA do DEPARTAMENTO DE  
SISTEMAS E COMPUTAÇÃO do CENTRO DE  
CIÊNCIAS E TECNOLOGIA da UNIVERSIDADE  
FEDERAL DA PARAÍBA, como requisito  
parcial para obtenção do grau de  
MESTRE EM INFORMÁTICA.**

**Área de Concentração: SISTEMAS DE INFORMAÇÃO E BANCO DE DADOS**

**Orientador: PROF. DR. ULRICH SCHIEL**

**Apoio: IBM DO BRASIL E CNPQ**

**CAMPINA GRANDE  
ABRIL / 1999**



S729a Souza, Aídre da Cunha Guedes de.  
Ambiente computacional para gerência estratégica : data  
warehousing : uma survey / Aídre da Cunha Guedes de Souza.  
- Campina Grande, 1999.  
158 f.

Dissertação (Mestrado em Informática) - Universidade  
Federal da Paraíba, Centro de Ciências e Tecnologia, 1999.  
"Orientação : Prof. Dr. Ulrich Schiel".  
Referências.


1. Banco de Dados - Sistemas de Informação. 2. Ambiente  
Computacional - Gerência Estratégica. 3. Data Warehousing.  
4. Dissertação - Informática. I. Schiel, Ulrich. II.  
Universidade Federal da Paraíba - Campina Grande (PB). III.  
Título

CDU 004.65(043)

**AMBIENTE COMPUTACIONAL PARA GERÊNCIA ESTRATÉGICA –  
DATA WAREHOUSING: UMA SURVEY**

**AÍDRE DA CUNHA GUEDES DE SOUZA**

**DISSERTAÇÃO APROVADA EM 19.04.1999**



**PROF. ULRICH SCHIEL, Dr.  
Orientador**



**PROF. MARCUS COSTA SAMPAIO, Dr.  
Examinador**



**PROF. RUBENS NASCIMENTO MELO, Dr.  
Examinador**

**CAMPINA GRANDE – PB**

*Dedico esse trabalho a:*

*meus pais, por razões óbvias;*

*Ana Paula Mesquita, pela ajuda preciosa, amiga e afetuosa nos inúmeras momentos em que teve a oportunidade de colaborar e assim fez;*

*José Carlos Neves, pelo apoio amigo, esclarecimentos, incentivos e recomendação no meu ingresso no mestrado;*

*Vicente Moreira, sempre solidário e amigo, agindo com a postura de um verdadeiro Educador, com dicas, orientações e muito estímulo, antes e 'remoto', durante o curso;*

*Ulrich Schiel e Marcus Sampaio, na certeza de ter plantado uma semente boa.*

## **Agradecimentos**

A todos os funcionários da UFPb ligados ao mestrado em Informática, muito obrigado!

Ao amigoso povo Campinense que nos acolhe enquanto estudantes, sempre de braços abertos provendo-nos a infra-estrutura necessária para triunfamos no nosso trabalho, meus maiores agradecimentos.

A IBM do Brasil (bolsa de dois anos) e ao CNPq (bolsa de seis meses), pelos suportes financeiros.

Aos meus colegas do antigo Banco Econômico, por todo companheirismo e solidariedade, uma saudosa e grande gratidão.

A todos que contribuíram de forma direta ou indireta para a realização deste trabalho, não citando nomes para não cometer injustiças, o meu mais sincero muito obrigado!

## Resumo

SOUZA, Aídre da Cunha Guedes – *Ambiente Computacional para Gerência Estratégica - Data Warehousing: Uma Survey*, Dissertação (Mestrado em Informática), 158 p., Campina Grande, Universidade Federal da Paraíba, abril 1999.

As abordagens dos sistemas de informação gerenciais voltados para a análise e suporte à decisão (EIS - Executive Information System e DSS - Decision Support System) baseadas em Repositório de Dados do tipo Armazém de Dados, ou seja, Data Warehouse (DW), têm crescido de forma extraordinária. Contudo, muitos dos projetos que utilizam essa tecnologia não têm tido o êxito esperado. Alguns sequer conseguiram ser concluídos de forma a atender ao que se destinam: disponibilizar informações corporativas consolidadas e consistentes para a gerência estratégica. Por ser um trabalho que envolve grandes custos de desenvolvimento, implantação e manutenção, o fracasso em um projeto destes resulta em prejuízos fabulosos.

Este trabalho apresenta uma visão geral e detalhada do ambiente computacional voltado à gerência estratégica, *Data Warehousing*, de forma genérica, enfocando o que acreditamos ser o maior ponto crítico do sucesso deste tipo de projeto: a real compreensão da sua funcionalidade enquanto tecnologia, e as suas interações e características fundamentais de funcionamento, tais como os tipos de processamentos e a temporalidade do/no seu repositório de dados, o *Data Warehouse*. Abordam-se as óticas acadêmica e comercial, pois, em se tratando desta tecnologia ter seu habitat no mercado, suas muitas informações práticas nos foram úteis na compreensão da atualidade deste tema. O trabalho também refere-se ao seu surgimento, assim como à sua interação com outras tecnologias afins; ainda ressalta muitos pontos críticos de entendimento que encontramos no decorrer de projetos deste gênero.

## Abstract

SOUZA, Aídre da Cunha Guedes – *Ambiente Computacional para Gerência Estratégica - Data Warehousing: Uma Survey*, Dissertação (Mestrado em Informática), 158 p., Campina Grande, Universidade Federal da Paraíba, abril 1999.

The use of management information systems for decision support (EIS - Executive Information System e DSS - Decision Support System) based on data repository of kind Data Warehouse (DW), has been largely increased. Although, many of the projects that use this technology hasn't reached the expected success. Some of them not even could be finished in order to attend its requirements: to provide consolidated and consistent corporate informations for strategic management. As a work that involves large expenses in development, implementation and maintenance, the fail of such a project results in great damage.

This work shows a general and detailed view of the computational environment for strategic management, *Data Warehousing*, in a generic way, focusing what we believe is the major critical point of success of this kind of project: the real comprehension of its functionality as technology, and its interactions and fundamental features of working, like the kind of processing and the temporality of/in the data repository, the *Data Warehouse*. Academic and commercial approaches are done, since, talking about a technology that has its habitat on the market, much of the lots of practical information has been useful to the comprehension of the news on this subject. The work also relates the emergence of this new technology, as well as to its interactions with other related technologies; still emphasizing many critical points of its understanding that we found during the development of this project.



## Sumário

<b>1. INTRODUÇÃO .....</b>	<b>13</b>
1.1. OBJETIVO .....	13
1.2. MOTIVAÇÃO DO ESTUDO .....	14
1.3. RELEVÂNCIA DO TEMA .....	16
1.4. ESTRUTURA .....	18
<b>2. SISTEMAS DE INFORMAÇÃO.....</b>	<b>20</b>
2.1. IMPORTÂNCIA, CONTEXTO E EVOLUÇÃO .....	20
2.2. SEPARAÇÃO FUNCIONAL.....	25
2.2.1. <i>Sistemas de Automação</i> .....	26
2.2.1.1. Legacy .....	28
2.2.1.2. On Line Transaction Processing.....	30
2.2.2. <i>Sistemas Gerenciais</i> .....	31
2.2.2.1. Management Information Systems (MIS).....	32
2.2.2.2. Executive Information System (EIS).....	33
2.2.2.3. Decision Support System (DSS).....	36
2.2.2.4. Comparativo entre MIS, EIS e DSS.....	41
2.3. RESTRIÇÕES AMBIENTAIS PARA OS SISTEMAS GERENCIAIS .....	44
2.3.1. <i>A Problemática para a Efetivação dos EIS e DSS</i> .....	48
2.3.2. <i>A Solução Encontrada com o Uso do Data Warehouse</i> .....	52
2.4. RESUMO .....	54
<b>3. ON-LINE ANALYTICAL PROCESSING (OLAP) .....</b>	<b>57</b>
3.1. HISTÓRICO .....	58
3.1.1. <i>Origem</i> .....	58
3.1.2. <i>As Regras para OLAP</i> .....	60
3.1.3. <i>Movimentação Comercial</i> .....	62
3.2. DEFINIÇÃO E FINALIDADE.....	62
3.3. OPERAÇÕES.....	66
3.4. RELATÓRIOS.....	70
3.5. REPOSITÓRIO OLAP .....	72
3.6. MODELAGENS PARA REPOSITÓRIOS OLAP .....	75
3.6.1. <i>OLAP Multidimensional Relacional</i> .....	75
3.6.2. <i>OLAP Multidimensional Pura</i> .....	82
3.6.3. <i>Comparativo ROLAP X MOLAP</i> .....	84
3.7. ARQUITETURAS OLAP.....	86
3.7.1. <i>Cliente Gordo</i> .....	87
3.7.2. <i>Cliente Magro</i> .....	88
3.8. RESUMO .....	89
<b>4. DATA WAREHOUSING (DWING) .....</b>	<b>91</b>
4.1. CONTEXTO, IMPORTÂNCIA E ENFOQUES .....	91

4.2.	DEFINIÇÃO E ARQUITETURA BÁSICA .....	94
4.3.	DATA WAREHOUSE (DW) .....	97
4.3.1.	<i>Os Dados</i> .....	98
4.3.2.	<i>O Tempo dos Dados</i> .....	99
4.3.3.	<i>A Modelagem do DW</i> .....	102
4.3.4.	<i>A Funcionalidade do Armazenamento</i> .....	106
4.3.5.	<i>O Tempo de Armazenamento</i> .....	108
4.3.6.	<i>Os Metadados</i> .....	109
4.4.	OPERATIONAL DATA STORE (ODS).....	113
4.5.	ARQUITETURAS DE DWING .....	119
4.5.1.	<i>Camada Única (One-tier architecture)</i> .....	121
4.5.1.1.	<i>Camada Única Autônoma (Stand-alone One-tier Architecture)</i> .....	122
4.5.1.2.	<i>DW Virtual</i> .....	122
4.5.2.	<i>Duas Camadas (Two-tier Architecture)</i> .....	123
4.5.2.1.	<i>DW Padrão</i> .....	124
4.5.2.2.	<i>DW Dimensional</i> .....	124
4.5.2.3.	<i>DW Federativo</i> .....	127
4.5.3.	<i>Três Camadas (Three-tier Architecture)</i> .....	128
4.6.	INTERAÇÃO COM OUTRAS TECNOLOGIAS EMERGENTES .....	132
4.6.1.	<i>DW e OLAP</i> .....	132
4.6.2.	<i>DW e OLSP</i> .....	133
4.6.3.	<i>DW INTERNET, INTRANET E EXTRANET</i> .....	133
4.7.	PROBLEMAS EM ABERTO.....	136
4.8.	O ESTADO DA ARTE.....	137
4.8.1.	<i>Manutenção On-line / Atualização Automática</i> .....	137
4.8.2.	<i>Retroalimentação</i> .....	138
4.8.3.	<i>Reatividade e Tecnologia PUSH</i> .....	139
4.9.	RESUMO .....	140
<b>5.</b>	<b>CONCLUSÃO .....</b>	<b>142</b>
5.1.	DIFICULDADES ENCONTRADAS.....	144
5.2.	CONTRIBUIÇÃO.....	145
5.3.	SUGESTÕES PARA TRABALHOS FUTUROS .....	146
	<b>GLOSSÁRIO .....</b>	<b>147</b>
	<b>ÍNDICE REMISSIVO .....</b>	<b>157</b>

## Lista de Figuras

Figura 1.1 - Relações Intra e Interdepartamentais e Interempresariais.....	15
Figura 1.2 - Ambiente Data Warehousing.....	17
Figura 2.1 - Evolução dos Sistemas de Informação.....	24
Figura 2.2 - Separação Funcional dos SI.....	25
Figura 2.3 - Pirâmide Hierárquica com os respectivos SI.....	28
Figura 2.4 - Análise Multidimensional de Dados [WHITE, 1997].....	34
Figura 2.5 - Como Funcionam as Associações [KOURY, 1996].....	37
Figura 2.6 - Comparativo entre os SG.....	42
Figura 2.7 - Arquitetura de Desenvolvimento Espontâneo (A Teia de Aranha).....	44
Figura 2.8 - Diferenças entre Processamentos Operacionais e de Suporte à Decisão [BRACKETT, 1996].....	46
Figura 2.9 - Projeto de Topologia da Rede do Antigo Banco Econômico.....	48
Figura 2.10 - Ambientes de Repositórios de Dados do Antigo Banco Econômico.....	49
Figura 2.11 - A Localização do Data Warehouse.....	52
Figura 2.12 - Ambiente de Geração Espontânea Integrado com DW.....	53
Figura 2.13 - Tipos de SI e seus Tipos de Processamento.....	56
Figura 3.1 - Análise Multidimensional: O Cubo de Dados.....	59
Figura 3.2 - Exemplo de Hierarquia de Canal de Vendas [INMON, 1997].....	67
Figura 3.3 - Exemplo de Operação de <i>Drill-across</i> .....	68
Figura 3.4 - Exemplo de Análise Seletivo-Restritiva.....	69
Figura 3.5 - Exemplo de um Relatório Gerado por um SGBDR [KIMBALL, 1996].....	69
Figura 3.6 - Exemplo do Operação de <i>Pivot</i> do Relatório da Figura 3.5.....	70
Figura 3.7 - O Relatório Ideal de Data Warehouse [KIMBALL, 1998].....	71
Figura 3.8 - Serviços OLAP: Componentes Básicos.....	72
Figura 3.9 - Diferenças entre BD Operacional e Repositório OLAP.....	73
Figura 3.10 - Descrição de um <i>Star Schema</i> .....	76
Figura 3.11 - Exemplo de <i>Star Schema</i> para Vendas em Supermercado.....	78
Figura 3.12 - <i>Star Schema</i> com Mais de Um Fato.....	80
Figura 3.13 - Modelagens ROLAP.....	81
Figura 3.14 - Armazenamento Multidimensional sob a Ótica Comercial.....	83
Figura 3.15 - Comparação de Estilos de OLAP [BERSON & SMITH, 1997].....	84
Figura 3.16 - Modelagem ROLAP versus MOLAP.....	85
Figura 3.17 - Extremos da Arquitetura de Cliente OLAP.....	86
Figura 3.18 - Arquitetura de OLAP – <i>Cliente Gordo</i> .....	87
Figura 3.19 - Arquitetura de OLAP – <i>Cliente Magro</i> .....	88
Figura 4.1 - Diagrama Básico de Data Warehousing.....	96
Figura 4.2 - Tabela de Tempos Sexagesimais com Data Gregoriana e Tempo Absoluto (ABSTIME).....	101
Figura 4.3 - Acompanhamento Temporal dos Saldos de um Cliente Hipotético.....	103
Figura 4.4 - Saldo de Contas Bancárias de Dois Clientes em um Período.....	104
Figura 4.5 - Tabela Normalizada não Temporal.....	105

Figura 4.6 - Modelagem Normalizada Temporal .....	105
Figura 4.7 - Exemplo do Registro com Diferentes Tamanhos do Campo NOME DO CLIENTE entre Dois Sistemas Bancários .....	111
Figura 4.8 - Posição de um ODS no Diagrama Básico de DWing.....	114
Figura 4.9 - Utilização de Vários ODS no Diagrama Básico de DWing.....	114
Figura 4.10 - Diagrama Básico de DWing Visto sob a Ótica de Camadas.....	120
Figura 4.11 - Arquitetura de Camada Autônoma Única .....	122
Figura 4.12 - Arquitetura de Camada Única – DW Virtual .....	123
Figura 4.13 - Arquitetura de Duas Camadas – DW Padrão .....	124
Figura 4.14 - Arquitetura de Duas Camadas – DW Dimensional .....	125
Figura 4.15 - Arquitetura de Duas Camadas – DDW com OLSP Externo.....	126
Figura 4.16 - Arquitetura de Duas Camadas – DW Federativo .....	127
Figura 4.17 - Arquitetura de Três Camadas .....	128
Figura 4.18 - Arquitetura CIF [INMON ET AL, 1996] .....	128

## Lista de Siglas

- 3FN – Terceira Forma Normal
- AM – Análise Multidimensional
- API – “Application Program Interface”
- BD – Banco de Dados
- BDM – Banco de Dados Multidimensional
- BDR – Banco de Dados Relacional
- BDRT – Banco de Dados Relacional Temporall
- BDT – Banco de Dados Temporal
- CIF – “Corporate Information Factory”
- DDW – “Dimensional Data Warehouse”
- DSS – “Decision Support System”
- DW – “Data Warehouse”
- DW-cd – “Data Warehouse Current Datail”
- DWing – “Data Warehousing”
- DW-s – “Data Warehouse Summary”
- ECA – Evento, Condição, Ação
- EDP – “Eletrônic Data Processing”
- EIS – “Executive Information System”
- IT – “Infomation Tecnology”
- MDOLAP – “Multidimensional On Line Analytical Processing”
- MIS – “Management Information Systems”
- MOLAP – “Multidimensional On Line Analytical Processing”
- NC – “Network Computer”
- OLAP – “On-line Analytical Processing”
- OLSP – “On-line Synthetical Processing”
- SGBD – Sistema Gerenciador de Bando de Dados
- SGBDM – Sistema Gerenciador de Bando de Dados Multidimensionais
- SGBDR – Sistema Gerenciador de Bando de Dados Relaional

## 1. INTRODUÇÃO

Neste capítulo abordamos os aspectos informativos a nível estrutural dessa dissertação. Veremos o seu objetivo, a motivação que nos levou a pesquisar tal assunto, a relevância do tema escolhido e, faremos ao final, uma sucinta descrição da disposição dos assuntos nos capítulos desta monografia.

### 1.1. Objetivo

A dissertação objetiva ser uma referência acadêmica detalhada e didática sobre o tema *Data Warehousing* (DWing) – ambiente computacional para consultas destinadas ao âmbito gerencial a nível estratégico, baseado em um repositório<sup>1</sup> de dados histórico chamado de *Data Warehouse* (DW). Para melhor compreendermos este ambiente computacional, abordamos os Sistemas<sup>2</sup> de Informação (SI)<sup>3</sup> que de forma integrante participam do DWing, pois muitos ainda confundem as finalidades e fronteiras de cada um deles, isto é, onde começam e terminam as funções dos SI e começam as do DW. Serão detalhados os aspectos temporais do DW, ainda pouco referenciados (e até esquecidos, apesar da sua importância absoluta) nestes projetos, até o momento. Ao final, faremos a correlação entre o DW e outras tecnologias emergentes no mercado. Desta forma, teremos maior segurança no entendimento do processo de DWing no que diz respeito à real

---

<sup>1</sup> *Repositório de dados*, ou simplesmente *repositório*, é qualquer forma de armazenamento (mídia) no qual seja possível o registro de dados e recuperação dos mesmos de forma confiável.

<sup>2</sup> “Sistema é um conjunto de partes coordenadas, que concorrem para a realização de um conjunto de objetivos, segundo um plano. Qualquer sistema pode ser encarado como um subsistema de um maior, sendo isso denominado Hierarquia de Sistemas” [CAUTELA ET AL, s.d.].

<sup>3</sup> “SI é qualquer sistema usado para prover informações (incluindo o seu processamento), qualquer que seja sua utilização” [CAUTELA ET AL, s.d.].

funcionalidade do DW, seus aspectos temporais, a sua interação com as diversas fontes de dados operacionais e ferramentas administrativas de produtividade.

## **1.2. Motivação do Estudo**

Conforme [WIDOM, 1995], DWing tem sido um assunto que tem crescido bastante comercialmente, no entanto, no meio acadêmico – em particular pela comunidade de pesquisadores da área de Bancos de Dados (BD) –, ele tem tido atenção limitada. Materiais comerciais sobre DWing não faltam, especialmente impressos de divulgação publicitária de produtos. Contudo, ao tentarmos obter informações mais detalhadas e independentes, ou seja, documentos didáticos e abertos – livre de fornecedores, com seus produtos e metodologias específicas –, sentimos dificuldade de encontrá-los.

Por se tratar de um assunto apropriado para grandes organizações comerciais do tipo bancos, redes de supermercados, etc., o tema oferece, para aqueles que não obtiveram uma prévia experiência no segmento corporativo, relativa dificuldade na perfeita compreensão dos conceitos envolvidos. Isso se deve ao aspecto da departamentalização<sup>4</sup> organizacional e seus reflexos de independência funcional enquanto unidade administrativa e, ao mesmo tempo, da sua interdependência funcional enquanto corporação, não esquecendo também as possíveis relações interempresariais cada vez mais presentes. Na composição da imagem geral da instituição para a administração estratégica da organização identificamos então relações intra e interdepartamentais, conforme veremos a seguir.

Como podemos observar na Figura 1.1, a Empresa A, um supermercado por exemplo, tem vários departamentos: A1, A2, ..., An, tais como: compras, vendas, marketing, etc.. Supondo-se que seus diversos departamentos sejam micro-informatizados, existe uma chance grande de que haverá entre alguns deles um relacionamento não só intradepartamental, como também, interdepartamental. Por exemplo, o departamento de compras precisa saber do de vendas, informações sobre o estoque, para poder efetuar as

---

<sup>4</sup>Divisão dos órgãos que compõem a estrutura empresarial feita por um esquema que assegura homogeneidade e equilíbrio [CHIAVENATO, 1979].

compras rotineiras. Por sua vez, o departamento de marketing precisa saber do departamento de vendas, quais os produtos que existem no estoque, assim como a sua quantidade para posicionar-se com propagandas e promoções, e assim por diante. Isto seriam exemplos de relações interdepartamentais.

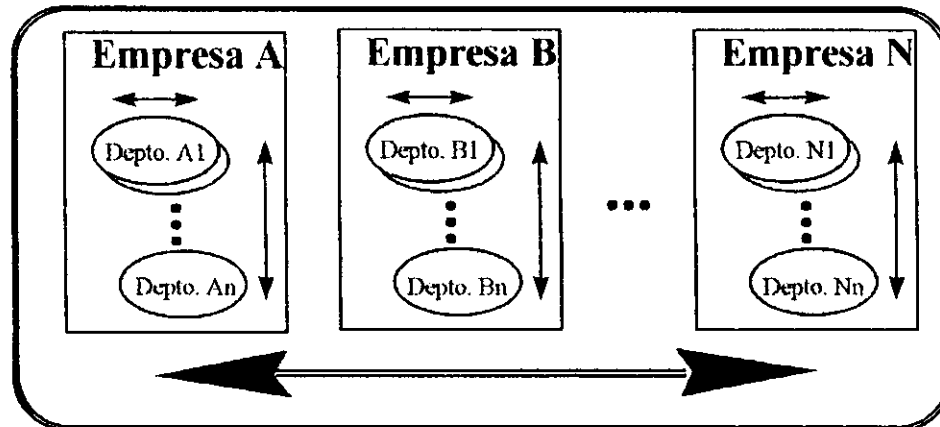


Figura 1.1 - Relações Intra e Interdepartamentais e Interempresariais.

Usando ainda o nosso exemplo do supermercado, é fácil perceber também a existência de relações interempresariais, tais como com fornecedores, promotores, etc., na nossa figura simbolizadas pelas Empresas B, ..., N. Com a economia globalizada, as relações interempresariais estão cada vez mais presentes e intensas no meio comercial. Para o corpo gerencial de uma instituição, seja qual for o ramo a que pertença, devido a esse novo e crescente tipo de economia, faz-se cada vez mais essencial a visão geral da mesma – mais conhecida como *visão corporativa* –, objetivando perfeita integração sistêmica, para melhor se atuar no campo estratégico, por parte daqueles responsáveis pelo destino da corporação.

O autor da presente dissertação já participou de vários projetos corporativos, trazendo para esse documento relatos e opiniões de forma a enriquecer este trabalho com a sua própria experiência. Sendo assim, a nossa vivência de quase dez anos na área de processamento de dados, numa grande instituição financeira com mais de 160 anos de



existência, o Banco Econômico S/A<sup>5</sup> (pivô de um conglomerado de outras empresas, formando um grande grupo financeiro do segmento comercial Brasileiro, o Grupo Econômico), nos trouxe a devida visão corporativa necessária a elucidar o entendimento e a importância de projetos de DWing para instituições com tamanho porte e nos motivar para nossos estudos nesta direção.

### 1.3. Relevância do Tema

Muito tem-se investido ultimamente, na informática, como diferencial competitivo. “Atualmente 20% das empresas de médio e grande porte tem um Data Warehouse (DW) em funcionamento e esse número tenderá a 80% dentro dos próximos 5 anos”, afirmam os consultores David Wells e Philip Carnelley em pesquisa encomendada pela empresa inglesa Ovum [CAMPOS, 1996].

Apesar do contínuo e garantido crescimento dos DW, o êxito esperado neste tipo de empreitada não tem sido o esperado. Na mídia não chegam os casos de fracasso, mas sabemos que eles tem ocorrido e muitos dos projetos efetivados foram suspensos pelas mais diversas causas. A mais presente é a falta da perfeita compreensão e domínio desta tecnologia.

A Figura 1.2 nos dá uma idéia da complexidade de um ambiente de DWing. Nela vemos diversidade de hardwares, sob a forma de inúmeros micros ligados ou não em redes locais, computadores de grande porte (*mainframes*), diversas mídias, etc., e, não tão visível, imaginamos a multiplicidade de softwares existentes para operacionalizar o diversificado ambiente, que retrata a realidade de muitas das grandes corporações empresariais, supermercados, bancos, etc.. Nota-se também que, no lado direito desta

---

<sup>5</sup>O Banco Econômico S/A fundado em 1834, foi dividido em 1996 em duas partes, após ficar sob a intervenção do Banco Central (BC) por nove meses. Uma parte (tida como boa) foi vendida ao Banco Excel (tornando-se Banco Excel Econômico), e a parte (dita ruim) ficou com o BC, a fim de resgatar os créditos considerados podres e empréstimos concedidos e tidos como de difícil liquidação. Atualmente, mais precisamente nesse ano (1998) o Banco Excel Econômico foi comprado pelo espanhol Banco Bilbao Vizcaya.

figura, estão algumas pessoas que representam o grupo seletivo de profissionais que trabalham com ferramentas voltadas às determinações estratégicas, também conhecidas comercialmente como “decision makers”<sup>6</sup>. Eles interagem com outro grupo de profissionais chamados de *knowledge workers*, que são os especialistas em conhecimento em termos genéricos. Juntos beneficiam-se com o DWing. Ao centro da figura está o DW, como sendo o pivô informativo do DWing. O DW foi idealizado como uma solução prática, motivada pelas necessidades prementes do moderno mercado competitivo, conforme veremos ao longo deste documento.

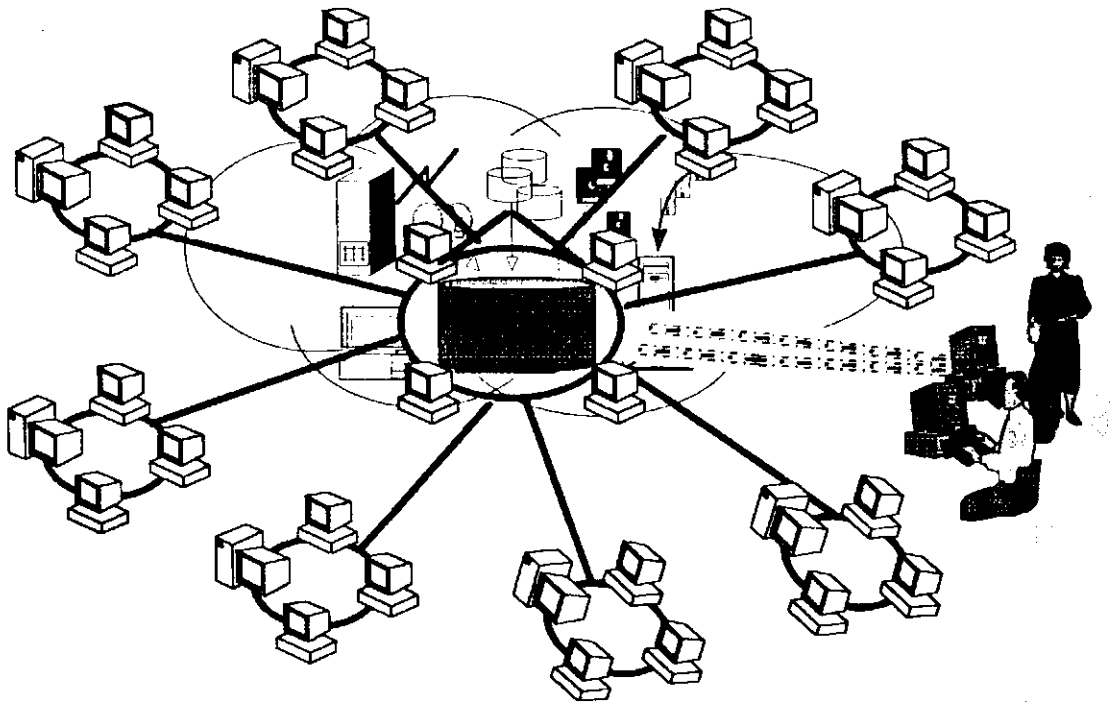


Figura 1.2 - Ambiente Data Warehousing.

A característica singular desse trabalho está na forma com que ele foi elaborado, isto é, baseando-se em necessidades atuais e prementes do mercado e contando com as opiniões dos experientes consultores no assunto, registradas em rica e diversificada bibliografia consultada. Afinal, a solução DW foi idealizada com uma solução prática,

<sup>6</sup>Há quem traduza naturalmente esta expressão inglesa para “tomadores de decisão”, porque é comum utilizar-se a expressão “ele tomou uma decisão”. Contudo, segundo o dicionário Aurélio, o verbete tomador, está ligado a operações de crédito. Sendo assim, na necessidade de tradução, sugerimos a expressão “determinadores”, que, tendo sentido apropriado, é correta e também sugestiva.

motivada pelas necessidades essencialmente comerciais, conforme veremos ao longo deste documento.

Muitos produtos já estão disponíveis no mercado. Muitas soluções imediatistas estão sendo adotadas, em função da atual competitividade empresarial. Mas, o sucesso da solução DWing a longo prazo está substancialmente atrelado ao perfeito equacionamento da dimensão do projeto e ao conhecimento da razão da existência das várias arquiteturas com suas vantagens, limitações, temporalidade envolvida, etc., enfim dominar ao máximo possível essa tecnologia emergente em consonância com as muitas outras já existentes no mercado. Neste ponto nosso documento oferece uma singular contribuição.

#### **1.4. Estrutura**

A disposição dos assuntos nessa dissertação foi dividida em quatro partes e distribuída por seis capítulos da seguinte forma:

- ☐ parte I (capítulo 1): apresenta e traça as linhas gerais do trabalho, esclarecendo os seus objetivos, motivação para o estudo, situa-o no contexto da sua importância para a academia e para o mercado e descreve-o estruturalmente;
  - ☐ parte II (capítulos 2 e 3): aborda diversos tipos de SI, seus momentos tecnológicos, objetivos, tipos de processamento, classifica-os, compara-os, evoca os problemas de implementação de alguns desses sistemas que evidenciaram a necessidade do surgimento da solução baseada no DW, e descreve detalhadamente o processamento analítico (OLAP - "On-line Analytical Processing"), tecnologia emergente e em franca expansão;
  - ☐ parte III (capítulo 4): aborda DWing, define e qualifica-o de forma completa, descrevendo-o através de uma arquitetura básica, exibindo também outras arquiteturas e relatando tecnologias afins utilizadas nesses projetos, interações mais fortes com redes
-

do tipo INTERNET (a grande rede mundial), conceitua e analisa vantagens e desvantagens no uso de *Data Marts* (subconjunto de um DW qualquer), comenta os problemas em aberto, cita o estado da arte e examina os aspectos temporais propondo enfoques mais realistas na representação desses aspectos na área de BD e, especialmente, no âmbito dos BD Temporais (BDT);

□ parte IV (capítulo 5): relata a conclusão dos nossos estudos materializados sob a forma dessa dissertação de mestrado, relatando as dificuldades encontradas, ressaltando a contribuição e sugerindo trabalhos futuros.

## 2. SISTEMAS DE INFORMAÇÃO

Neste capítulo abordamos os diferentes tipos de Sistemas de Informação (SI) existentes hoje e os objetivos a que se destinam. Citamos a sua importância para o cidadão moderno e posicionamo-los no contexto atual. Descrevemos a sua evolução tecnológica, classificando-os, segundo os seus tipos de processamentos em dois básicos e alguns sub-tipos, de acordo com a finalidade a que se destinam: *automação* e *gerenciamento* de uma organização. Ao final, eles são comparados, terminando com uma visão da problemática das restrições dos ambientes de processamento dessas organizações, que justificam a necessidade do surgimento da solução baseada no DW.

### 2.1. *Importância, Contexto e Evolução*

Cada vez mais, notamos a crescente importância da informação para o homem. De acordo com [KORFIAGE, 1997], o dia-a-dia do homem depende de informações, sendo elas das mais diversas em fontes e tipos, desde atravessar a rua (tendo que se informar através das cores dos semáforos) até obter diagnósticos de complexos problemas médicos. Algumas destas informações, como as dos semáforos, podem ser obtidas de forma bastante simples. Contudo, outras tantas dependem da busca e compilação de múltiplas fontes. A informática, ao mesmo tempo que oferece subsídios para facilitar o acesso e mesmo a manipulação das informações, aumenta o interesse na sua utilização e, conseqüentemente, as suas necessidades de uso, conforme a contemporânea teoria econômica baseada na dicotomia dos recursos escassos e necessidades ilimitadas [ROSSETTI, 1984].

Ao longo do tempo, através da evolução dos SI (cuja definição foi vista na seção 1.1) podemos constatar mudanças nas diversas formas de atuação e abrangência da informática (tecnologia que lida com informação através do uso de equipamentos e

procedimentos com o processamento eletrônico dos dados). Portanto, o que era apenas uma forma “inteligente” de se automatizar um processo baseado em informações, tornou-se uma maneira, não apenas de controle, mas, inclusive, um meio de *qualitativamente* avaliar, planejar, projetar e fundamentar decisões (e mesmo até explicitá-las!) sobre aspectos os mais estratégicos possíveis para as organizações. Isto é, sob o enfoque comercial e segundo [KELLY, 1996], os gerentes não se baseiam no que está *bom* ou *mau* na empresa, mas sim nas boas ou más informações que possuem no momento. Elas serão o subsídio fundamental para a orientação daqueles responsáveis na administração executiva pelas decisões, os “decision makers” (em português, determinadores e, de forma mais adaptada à realidade comercial, executivos), a fim de se posicionarem orientados nas ditas *boas* ou *más* informações oriundas dos dados dessa organização, manipulados de forma a traduzir o seu contexto comercial, não só o momentâneo, mas também (e de forma ideal) o projetado. Assim requer o cada vez mais competitivo mundo dos negócios, que, com grande interesse nas organizações, comercialmente denomina-se de Tecnologia da Informação (IT – do inglês “Information Technology”) esse campo de estudo.

O grande desafio a que os SI estão cada vez mais sujeitos é, em última instância, o de tornar disponíveis *informações corretas, precisas, atualizadas e concisas, na apresentação adequada e no momento ideal*. Essa é a forma que traduz o diferencial competitivo esperado dos SI, pelas organizações, conforme requer o complexo e impaciente mundo capitalista globalizado. Mas, para isso teremos que entender o longo, porém paradoxalmente rápido, caminho que percorremos para chegarmos ao atual nível de informatização.

Ainda de acordo com [KELLY, 1996], passamos por três eras distintas: Automação, Controle e Produtividade. A Era da Automação é caracterizada pelos sistemas de processamento eletrônico de dados, os chamados EDP (do inglês “Electronic Data Processing”), onde o computador é responsável por executar tarefas simples, repetitivas, tais como cálculos e comparações entre cadeias de caracteres, valendo-se apenas da sua rapidez e precisão. Já na Era do Controle, identificada pelos sistemas de informações gerenciais (MIS – “Management Information Systems”), verifica-se a geração

diversificada de relatórios básicos gerenciais do tipo totalizações de campos<sup>7</sup> e grupos destes.

Nesta época, na Era do Controle, os sistemas – devido exatamente à grande ênfase no próprio controle empresarial – produziam reduzida informação realmente gerencial, apesar de terem a denominação que assim sugeria. O modelo de evolução dos SI tinha como paradigma a substituição sistêmica – ou seja, sistemas eram sempre substituídos por outros com outras facilidades, mas que estavam sempre tornando disponível as funções básicas operacionais, de forma a “forçar” a substituição completa do sistema anterior pelo novo –, o que comprometia sobremaneira e ainda distorcia a validade deste modelo como evolutivo [KELLY, 1996].

Acrescentemos a isso duas questões que influenciaram bastante esse processo evolutivo para sairmos da Era do Controle e irmos efetivamente para sistemas gerenciais. A primeira é o processamento *on-line*, ou seja, interativo, extremamente moderno e atraente para a época. A popularização dos sistemas *on-line* fez com que o uso do processamento *batch*, isto é, em lote, antes soberano e absoluto, só fosse utilizado quando estritamente necessário – como no caso do sistema bancário de compensação de cheques –, o que contribuiu para a mudança, em larga escala, da forma de interação do usuário com a máquina. Portanto, o perfil do usuário de informática mudou, passando, ele agora, a viver bem mais próximo dela.

A segunda questão, um passo evolutivo dado no tipo de processamento *on-line*, reafirmando-o enquanto tipo de interação homem-máquina, foi a chegada dos Sistemas Gerenciadores de Bancos de Dados (SGBD) baseados no conceito de transação para aplicações tradicionais. Conforme [LIMA & TOLEDO, 1997], estas aplicações são caracterizadas por possuírem curta duração e não necessitarem de cooperação<sup>8</sup>, aspectos

---

<sup>7</sup>Estas “totalizações” são conhecidas na prática por “quebras” (ex. quebra por código do produto, quebra por tipo de empresa, etc.).

<sup>8</sup>Ultimamente surgiram novos modelos de transações mais apropriados às novas necessidades evolutivas do processamento transacional, onde as propriedades de Atomicidade e Isolamento são revistas [SAMPAIO, 1995].

que permitem o gerenciamento de transações baseado no conceito de serializabilidade [BERNSTEIN ET AL., 1987]. Este tipo de gerenciamento agrupa um conjunto de operações em unidades com propriedades conhecidas como ACID (Atomicidade, Consistência, Isolamento e Durabilidade) [LIMA & TOLEDO, 1997].

Por um lado a chegada do processamento transacional baseado na teoria da serializabilidade auxiliou mais ainda o desenvolvimento das eras da automação e controle, pois se presta perfeitamente à operacionalização e o controle de negócios, onde os dados precisam ser constantemente atualizados. Por outro, necessitou-se abstrair da idéia de processamento transacional para se ir mais adiante e continuar o processo evolutivo dos SI em direção ao âmbito gerencial estratégico, cuja ênfase está na consulta dos dados.

A evolução dos SI estratégicos se deu contando com a influência das modernas interfaces gráficas, que modificaram bastante a aparência e funcionamento desses sistemas, tornando-os mais amigáveis, ou seja, mais agradáveis de se interagir. Nesta época, o enfoque na interação Homem-máquina era grande e muitas idéias, para sistemas gerenciais, antigas, porém futuristas para época, foram resgatadas.

Surgiram alguns sistemas especialmente voltados para o corpo executivo empresarial (gerentes, gestores, etc.). Conhecidos como EIS (do inglês "Executive Information Systems"), esses sistemas são caracterizados pela possibilidade de visualização e interação com os dados diferentemente combinados entre si, isto é *agregados*, e sob vários níveis de detalhe, ou seja, *granularidades*. A forma de processamento passou de transacional para analítica. Eles serão comentados na seção 2.2.2.2 e amplamente explorados no capítulo 3, devido à sua crescente popularidade nos projetos de DWing.

Com a disseminação dos microcomputadores, surgiram novos softwares. Estes micros passaram a conviver com o que já havia no mundo do processamento de dados sem que houvesse uma troca dos mesmos, ou seja, as outras plataformas existentes, tais como *mainframes* e computadores de médio porte. Desta forma quebrou-se o paradigma da substituição sistêmica. Atingimos assim a era atual, ou seja, a Era da Produtividade dos



“decision makers” (simplicadamente a Era da Produtividade), identificada por possibilitar maior liberdade a esse perfil novo de usuário no seu trabalho de busca por informações específicas, tendo o seu apogeu nos sistemas de apoio (ou suporte) à decisão (DSS, do inglês “Decision Support Systems”). Estes últimos sistemas têm pretensões de capacitar o usuário na sua atividade de decisão, de forma a mostrar, guiar e até descobrir informações não intuitivas que naturalmente não seriam inteligíveis, tais como padrões de comportamento de clientes segundo as mais diversas situações, etc.. Apesar de estar ainda em estágio rudimentar, limitando-se mais ao nível conceitual, selecionamos maiores informações esclarecedoras sobre eles na seção 2.2.2.3.

Veremos, ainda nesse capítulo, na próxima seção, maiores detalhes sobre cada um desses sistemas citados. A ilustração da Figura 2.1 resume o que foi dito sobre os SI e a sua

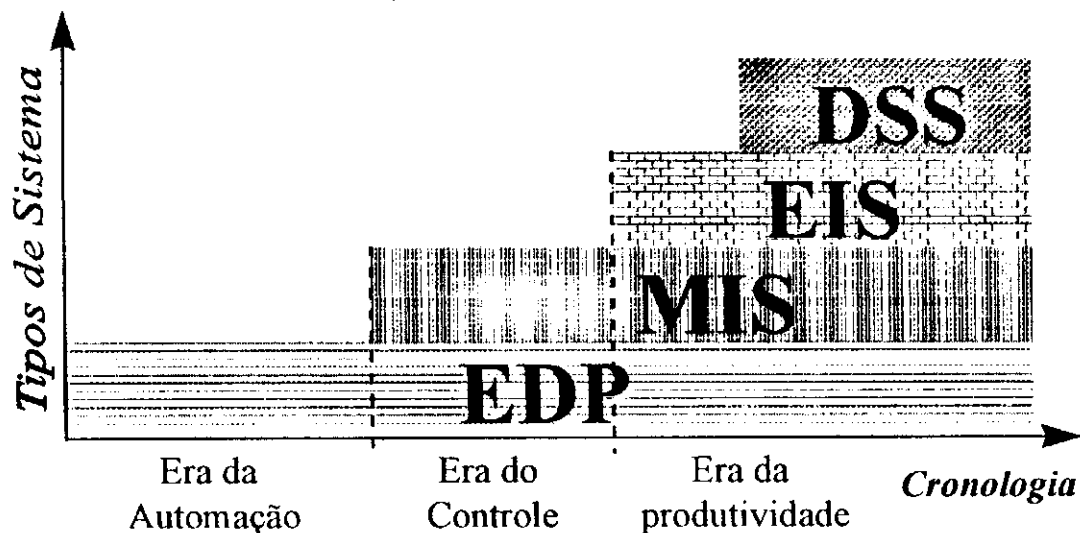


Figura 2.1 - Evolução dos Sistemas de Informação

cronologia. Nota-se na prática que, quando as empresas optam por enveredar pela informática, em todos os seus níveis de informatização, as eras sucedem-se nesta mesma ordem de cronologia. Elas passam de uma era evolutiva para outra sucessivamente.

Vale ressaltar que alguns autores, como por exemplo [INMON, 1996], consideram que o DSS é uma atualização do MIS, ou seja, o MIS moderno. Essa posição se deve à questão prática do uso dos dados, ou seja, ao objetivo a que o processamento dos dados se

destina, no caso, uso gerencial. Essa questão será abordada na seção 2.2.2.3 em detalhes. Contudo, adiantamos que, devido aos aspectos tecnológicos das eras da automação e controle serem completamente diferentes da Era da Produtividade – segundo características filosóficas já citadas, tais como o paradigma da substituição sistêmica e a forma de utilização dos bancos de dados, com aplicações tradicionais –, concordamos com [KELLY, 1996], que o DSS não é e nem será a evolução natural do MIS, pois, nem no que se refere à forma de processamento (transacional tradicional), nem pelas informações por eles produzidas, possuem grande semelhança.

## 2.2. Separação Funcional

Conforme dito anteriormente, a questão prática do uso dos dados, ou seja, o objetivo a que o processamento dos dados se destina, nos leva a um outro tipo de classificação baseado na separação funcional, não mais cronológica, mas fundamentada na sua abordagem prática. De acordo com [BRACKETT, 1996], as organizações podem usar os dados para operar seus negócios e para evolui-los.

Dessa forma e seguindo esse enfoque prático, constatamos os seguintes tipos sistemas de informação, segundo a sua finalidade: Sistemas de Automação (SA), ou seja, para operar o negócio, e Sistemas Gerenciais (SG), isto é, para evolui-los.

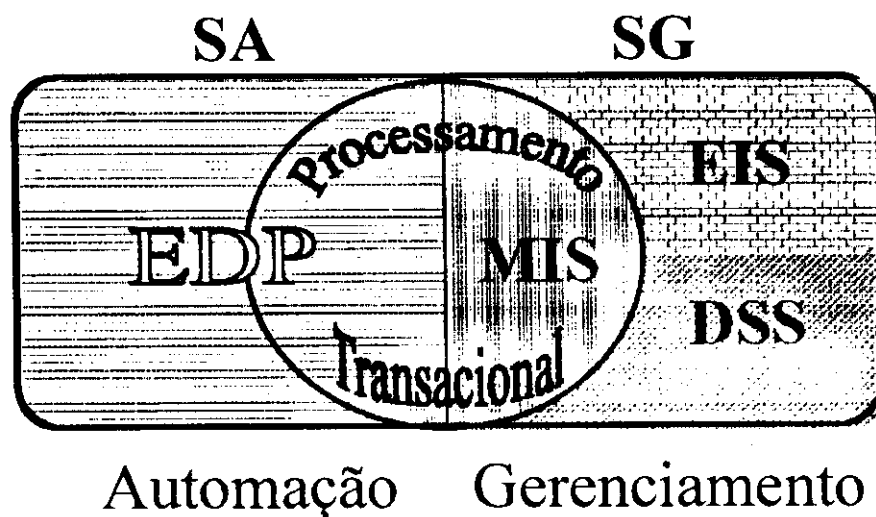


Figura 2.2 - Separação Funcional dos SI

Adotamos a notação de SA e SG, apesar de existirem outras terminologias atuais, tais como *sistemas operacionais* para os SA (que serão vistas mais adiante nas suas devidas seções), que, por serem equivocadamente ambíguas no nosso idioma, poderiam comprometer, momentaneamente, o perfeito entendimento do real sentido de cada uma delas.

A separação funcional dos SI (Figura 2.2) pode ser integrada com o que foi visto em relação aos vários tipos de sistemas na seção anterior, de forma agrupada conforme o tipo de objetivo. Do lado esquerdo estão os EDP, sistemas originalmente voltados à automação de procedimentos. Já do lado direito estão os três tipos de sistemas direcionados para o gerenciamento organizacional. É interessante notar o círculo que contorna os dois tipos de sistemas: SA e SG, ou seja automação e gerenciamento, indicando processamento transacional compreendendo praticamente a totalidade dos MIS.

Vale ressaltar que, enquanto os dados de entrada para os SA são oriundos do mundo real, os dados de entrada dos SG são provenientes dos dados processados e armazenados nos SA. Desta forma, os SG são alimentados pelos SA. Este ponto será mais amplamente discutido na seção 2.3, onde abordaremos as dificuldades envolvidas neste processo. Veremos a seguir mais detalhes sobre cada um desses dois tipos de sistema, SA e SG, e os seus tipos peculiares de processamento.

### 2.2.1. Sistemas de Automação

Principalmente no meio dos estudiosos e praticantes do DWing, os Sistemas de Automação (SA) são conhecidos como *sistemas operacionais*. [HAMMERGREN, 1996] afirma que esses sistemas respaldam os negócios do dia-a-dia, respondendo a eventos e transações. São os EDP, quer seja na sua forma *batch*, quer seja na sua forma *on-line*. A última forma, *on-line*, utiliza normalmente processamento transacional para aplicações tradicionais. Contudo, isso não impede que possam ser utilizados outros conceitos, tais como processamento transacional para aplicações cooperativas [BARGHOUTI & KAISER, 1991; ELMAGARMID, 1994 In: LIMA & TOLEDO, 1997], apesar de termos visto, na nossa

pesquisa, uma associação direta entre *sistemas operacionais* e processamento transacional para aplicações tradicionais por parte dos autores. Acreditamos que isso seja devido ao estado ainda pouco avançado e difundido, a nível comercial, das aplicações cooperativas, contrastando com o tema DWing, solenemente explorado comercialmente, o qual veremos com maiores detalhes no capítulo 4.

Vale salientar que a denominação de *sistemas operacionais* (em inglês “**operating systems**”) – expressão consagrada mundialmente pela informática para designar a camada de programas e rotinas de controle de hardware, como por exemplo UNIX, VM<sup>9</sup>, MVS<sup>9</sup>, OS/2<sup>9</sup>, MS-DOS<sup>10</sup>, WINDOWS<sup>10</sup>, etc. – refere-se, no estudo de DWing, aos sistemas administrativos informatizados voltados à operacionalidade empresarial (em inglês “**operational systems**”), ou seja, são os também conhecidos como *sistemas aplicativos*, tais como folha de pagamento, controle de estoque, etc., necessários à operacionalidade cotidiana da empresa dita informatizada.

Essa idéia dúbia da expressão *sistemas operacionais*, encontrada na nossa pesquisa, constitui um grave problema de ambigüidade gerada na tradução do inglês para o português, conforme citado acima, das expressões “operating system” e “operational system”. Constatamos isso não só ao consultarmos bibliografia traduzida existente, como [INMON, 1997], mas também ao verificamos na linguagem falada. Acreditamos que essa terminologia derivou da forte influência do grupo dos Administradores de Empresas, sempre envolvidos nestes tipos de projetos, ficando registrado ao longo do (pouco) tempo, desta forma, na bibliografia.

A Figura 2.3 nos mostra a “pirâmide administrativa” preenchida pelos SI correspondentes a cada nível hierárquico organizacional. Os níveis são responsáveis por atividades diferenciadas. De cima para baixo temos: *nível estratégico* (direção), o *nível tático* (controle) e o *nível operacional* (execução). Este último, provavelmente, foi que deu origem à expressão *sistemas operacionais*, aos EDP. Durante essa leitura, nos referiremos

---

<sup>9</sup>Marca registrada da IBM.

<sup>10</sup>Marca registrada da MICROSOFT.



## Níveis Organizacionais

Figura 2.3 - Pirâmide Hierárquica com os respectivos SI

aos *sistemas operacionais* (na tradução agora de “operational systems”) como SA, para melhor nos distanciarmos deste problema de ambigüidade. Não esqueçamos também que a forte influência do mundo prático, assim como a pouca atenção dada até então a esse importante assunto (DWing), pelo meio acadêmico – em particular, segundo [WIDOM, 1995], pela comunidade de BD – contribuíram de forma agravante no sentido de deixar perdurar, equivocadamente, tal expressão ambígua no nosso idioma.

De forma geral, podemos definir resumidamente os SA como sistemas que, a partir dos dados de entrada oriundos do mundo real, os processam eletronicamente de forma a gerar resultados como saída que contenham informações que atendam ao nível operacional de uma organização.

### 2.2.1.1. Legacy

“Legacy” do inglês, traduzido para o português como legado, herdado, é a denominação dada aos sistemas ativos existentes nas organizações e, conseqüentemente, os dados deles decorrentes. Até aqui podemos garantir a harmonia de pensamentos entre diversos os autores [INMON, 1996; BURETTA, 1997; BISCHOFF & ALEXANDER, 1997; BRACKETT, 1996; GILL & RAO, 1996] em relação a essa terminologia. Contudo, como

veremos a seguir, a forma de processamento (*on-line* ou *batch*), assim como a plataforma (sob a microinformática, ou não), faz com que as suas opiniões diverjam.

Para alguns autores tais como [INMON, 1996] e [BURETTA, 1997], independentemente da plataforma, dos antigos *mainframes* (porém bastante atuais nas grandes corporações), dos ultrapassados computadores de médio porte, ou mesmo da moderna microinformática, sistema *legacy* é qualquer sistema que manipule dados operacionais organizacionais. Resumidamente, a denominação sistema *legacy*, para eles, objetiva identificar os SA, independentemente do tipo de processamento (*on-line* ou *batch*), tratando-se de uma apologia àqueles sistemas encontrados na organização no momento em que deseja-se enveredar pela Era da Produtividade.

Para [BISCHIOFF & ALEXANDER, 1997], *legacy* refere-se à produção *antiga*, quer seja de sistemas, aplicações ou dados. [BRACKETT, 1996] acrescenta, fazendo uma explanação mais prática do assunto, que uma característica marcante dos sistemas *legacy* é a sua difícil manutenção, possuindo geralmente pouca documentação, limitada forma de manipulação dos dados e atendendo a um objetivo específico predeterminado. Pelas descrições citadas, percebemos que tudo nos leva aos velhos sistemas *batch*. Contudo, [GILL & RAO, 1996] são os que abordam esse assunto tecnicamente de forma mais clara e taxativa, dizendo que os dados *legacy* são em *batch* ou arquivados, aplicações operacionais que não precisem de nenhum tipo de *controle de concorrência*, isto é, não utilizam necessariamente nem processamento transacional, nem de BD, dessa maneira eliminando qualquer forma *on-line* de referência. Concordamos com a definição, a mais clara, desses dois últimos autores, em relação a associação dos sistemas *legacy* com os antigos sistemas *batch*, por achar que, em se tratando de tantos neologismos criados neste tema (DWing), temos um termo especialmente criado para descrever aplicações não *batch*, portanto *on-line*, e que exigem controle de concorrência, que é *On-line Transaction Processing* (OLTP), que veremos a seguir. Antes porém, gostaríamos de finalizar essa seção citando um exemplo típico de aplicação *legacy*: a aplicação bancária de atualização de contas correntes, em particular a rotina de compensação de cheques, cuja atualização é em *batch* e em um tempo predeterminado, independentemente da tecnologia empregada nesta operação.

### 2.2.1.2. On Line Transaction Processing

A importância da questão tempo de resposta [INMON, 1996; KIMBALL, 1996; HAMMERGREN, 1996] e, principalmente, a capacidade dos SGBD relacionais (SGBDR) [DATE, 1991] de armazenar consistentemente dados caracteriza bem o tipo de processamento conhecido como OLTP (“On Line Transaction Processing” [KIMBALL, 1996; HAMMERGREN, 1996; YAZDANI & WONG, 1998] ou “Online Transaction Processor” [INMON, 1996; INMON ET AL., 1996]).

[GILL & RAO, 1996] especificam OLTP como “um termo usado para definir qualquer sistema de software que lide com transações que utilizam dados entre a fonte do dado e o banco de dados no momento em que a transação ocorreu”, ou seja, sistemas de software que, se utilizando do processamento transacional para aplicações tradicionais, espelhem no BD imediatamente o dado no seu estado momentâneo. Para [KIMBALL, 1996], “OLTP é a descrição original de todas as atividades e sistemas associados à entrada de dados confiável em um banco de dados” e acrescenta, “é mais freqüentemente utilizada como referência aos BD Relacionais (BDR), embora OLTP possa ser de forma genérica usada para descrever qualquer ambiente de processamento transacional”.

Segundo Yazdani e Wong [Yazdan98], em OLTP as transações são curtas e em grande volume, oferecendo controle de concorrência em atualização/inclusão/eliminação (de registros) no processamento *on-line*, além de muita freqüência de rotinas e consultas.

Outra característica marcante referente à modelagem dos sistemas OLTP, é que estejam na terceira forma normal (3FN [DATE, 1991]) [INMON ET AL., 1996; BURETTA, 1997]. [BURETTA, 1997] acrescenta que as transações de atualização devem ser relativamente curtas, devido à existência de vários usuários compartilhando a mesma base de dados.

[BRACKETT, 1996] afirma que OLTP é apropriado para operações comerciais cotidianas e também é uma outra forma de se referir ao processamento operacional. Vale ressaltar que essa última frase causa uma dificuldade de entendimento para os leitores do

assunto. Ou seja, da mesma forma que para alguns autores os sistemas *legacy* englobam os OLTP, ocorre aqui o oposto: OLTP é tratado como a totalidade dos SA. Não vemos isso de forma aceitável e, para sermos coerentes com o exposto anteriormente sobre *legacy*, não concordamos em estender o conceito de OLTP para todos os SA, salvo aqueles *on-line*, conforme o nome sugere: sistemas de processamento transacional *on-line*.

Para finalizar o assunto SA, [TANLER, 1998] cita, como clássico exemplo de OLTP, o sistema de reservas de passagens aéreas. Veremos agora outros tipos de sistemas, não mais voltados à automação, mas sim ao gerenciamento.

### 2.2.2. Sistemas Gerenciais

Conforme vimos anteriormente, na seção 2.2 deste capítulo, nós chamamos os Sistemas Gerenciais de SG. São sistemas que atuam com a finalidade de prover ao usuário auxílio nas tarefas administrativas de controle e planejamento, ou seja, são ferramentas gerenciais de apoio ao corpo administrativo tático-estratégico de uma organização com a finalidade de evoluir os negócios.

Inicialmente, é importante lembrarmos que, apesar de estarem agrupadas sob o título de SG, estaremos abordando sistemas que surgiram em épocas tecnológicas diferentes: a Era do Controle – como representante os MIS – e a Era da Produtividade – dos sistemas EIS e DSS. Alguns autores agrupam os sistemas da Era da Produtividade e se referem a eles como EIS/DSS, ou ainda, somente DSS [BRACKETT, 1996; INMON ET AL., 1996; BISCHOFF & ALEXANDER, 1997]. Não concordamos com essa simplificação, pois tais sistemas – como veremos a seguir – possuem características completamente diferentes.

Outro ponto relevante e que causa alguma confusão, para aqueles que passam a se interessar pelo assunto SG, é a idéia de alguns autores [INMON, 1996] conceberem DSS como um novo nome para o MIS, partindo do pressuposto evolutivo, isto é, o MIS evoluiu e agora é chamado de DSS. Esse ponto foi ligeiramente comentado na seção 2.1 (onde já nos posicionamos em relação a essa comparação) e será mais detalhadamente abordado



adiante na seção 2.2.2.3. Veremos a seguir cada um desses sistemas e suas diferenças sob as óticas tanto técnicas, quanto administrativas. Começaremos, cronologicamente, pelos MIS.

### 2.2.2.1. Management Information Systems (MIS)

Para [CAUTELA ET AL., s.d.], *Management Information Systems (MIS)* “é todo e qualquer sistema que produza posições atualizadas a nível corporativo, resultado da integração de vários grupos de sistemas de informação, utilizando recursos de consolidação e interligação de entidades dentro de uma organização”. Ele continua: “é o SI que engloba todos os componentes da organização e todos os seus níveis de decisão”.

Como vemos, a pretensão dos MIS são (eram) bastante arrojadas de englobar “todos os componentes da organização e todos os seus níveis de decisão”. Justamente, o contexto de mudança de enfoque no uso dos dados a que ele se dispunha, com o objetivo de fornecimento do ferramental necessário à administração empresarial, o fizeram revolucionar à época, com a quebra do paradigma de que a informática estava restrita ao processamento eletrônico de dados (ou seja o EDP). Nesse momento encontramos a “porta de saída” da Era da Automação. Contudo, não foi possível atingir plenamente as prerrogativas a que o MIS almejava devido, basicamente, por dois pontos de cruciais, que impediram o sucesso da sua implementação.

O primeiro ponto é a questão de “produzir posições atualizadas a nível corporativo”, algo que de forma consolidada e confiável, nas grandes corporações, não é tão trivial de ser obtido. Isso ficará mais evidente na seção 2.3, sobre “Restrições Ambientais para os G”. O outro ponto foi (é), mais uma vez, o tipo de processamento utilizado, isto é, o transacional clássico, que impossibilitava maiores análises por parte do usuário, devido à sua falta de usabilidade dos dados – que sugeria um outro tipo de interação e enfoque dos mesmos –, agravado pelas limitadas interfaces existentes na época. Até os BD mais modernos da época, os BDR, eram eficazes e eficientes no armazenamento dos dados, mas não no fornecimento destes, em igual proporção qualitativa. Veremos

outras variações de tipos de processamento para SG ainda nas seções 2.2.2.2 e 2.2.2.3 que atendem melhor às prerrogativas com tal amplitude gerencial qual o MIS originalmente pleiteava.

Como ilustração do que foi realizado efetivamente em termos práticos de MIS, temos as diversas informações tiradas (ou mesmo por eles fornecidas<sup>11</sup>, conforme o paradigma da substituição sistêmica) dos sistemas EDP, através de somatórios e totalizações em percentuais consolidados por agrupamento funcional, tais como setores, departamentos, etc., ou seja, são relatórios gerenciais predefinidos e disponíveis a nível de sistema, que traduzem controle operacional de forma setorial, local e departamental, porém, não *a nível corporativo*, conforme sugeria seu conceito inicial. Por exemplo, um relatório bancário das contas correntes que “amanheceram” negativas, dentro e fora do limite de endividamento de cada cliente, por gerente de conta e de cada agência. Justamente por causa deste paradigma, as funções pertinentes aos MIS terminavam por serem incluídas em muitos dos EDP, uma vez que o tipo de processamento transacional era compatível com ambos, o que só veio a mudar com a chegada dos EIS, que veremos a seguir.

#### **2.2.2.2. Executive Information System (EIS)**

Com a evolução das interfaces gráficas e o reconhecimento do valor dos dados organizacionais, por parte dos administradores modernos, como o grande patrimônio do futuro numa sociedade informatizada que se organiza como sociedade pós-industrial – girando em torno de valores informacionais, cognitivos e selecionáveis, conforme [MASUDA, s.d.] –, surgiram diversos sistemas rotulados como EIS (do inglês, “Executive Information System”), de forma a contemplar os novos anseios e necessidades.

Esses novos sistemas comportam um ferramental, voltado ao gerenciamento estratégico, que provê análises de fatos comerciais registrados na forma de dados

---

<sup>11</sup> Alguns EDP acabaram por incluir as funções de controle dos MIS.

operacionais, sob várias óticas de atributos, simultaneamente, isto é, *Análise Multidimensional* (AM). Ou seja, AM pode ser vista como a técnica de visualização simultânea, por diversos ângulos administrativos, do agrupamento criterioso dos descritores de um fato comercial. Por exemplo, imaginemos a comercialização de um produto qualquer, que naturalmente envolve os fatos compras e vendas, registrados na organização na forma de dados operacionais. Identifiquemos então algumas possíveis dimensões tais como: mercado (composta por alguns atributos tais como: local de negociações de compras e vendas), produto (com: descrição do produto, marca, categoria), tempo (com: época de negociações de compras e vendas), transporte (com: descrição do

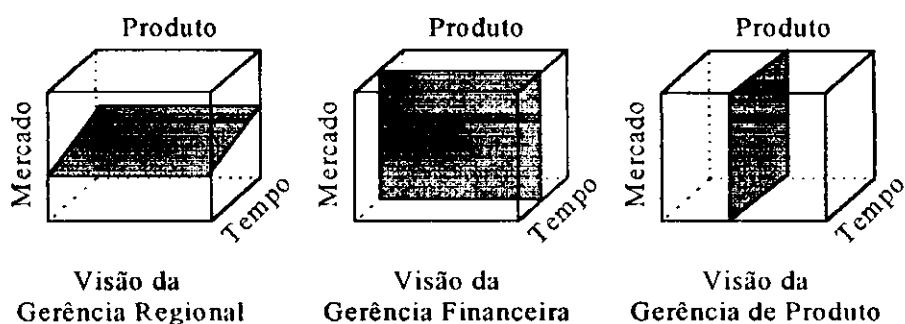


Figura 2.4 - Análise Multidimensional de Dados [WHITE, 1997]

roteiro, empresa transportadora), etc. Na Figura 2.4 são apresentadas três dimensões: mercado, produto e tempo, sob óticas gerenciais distintas: gerências regional, financeira e de produto, obtemos um panorama de uma dimensão qualquer em função das duas outras. Esta análise é particularmente interessante às gerências, de forma a auxiliar o trabalho pertinente a cada uma delas, individualmente.

Detalhemos mais a Figura 2.4. Ela é composta por três cubos. No primeiro cubo, direcionado à gerência regional na visualização do potencial de consumo de um mercado, temos um panorama dos produtos ao longo do tempo. Já no cubo do meio, importante para a gerência financeira na elucidação da rentabilidade comercial, temos um panorama dos produtos e mercados em função do tempo. E, por último, temos no terceiro cubo um panorama dos mercados ao longo do tempo, que é fundamental para a gerência de produto

na decisão da continuidade de comercialização de um produto qualquer, em um determinado mercado.

Os EIS são caracterizados pelo enfoque à importância do uso e manipulação dos dados operacionais e seus atributos<sup>12</sup>, na representação, compreensão e entendimento de fatos comerciais inerentes aos negócios. Nos EIS, os atributos dos dados operacionais são separados em dois tipos e agregados segundo um critério de natureza comum, relevante e concernente estrategicamente ao negócio como um todo. Atributos quantitativos ligados diretamente a um fato comercial qualquer (por exemplo compra, venda, etc.) são chamados de atributos do *fato*, por exemplo, para o *fato* venda, valor da venda, quantidade vendida, etc.. Atributos qualitativos ligados ao contexto do fato, isto é, tempo, geografia, etc., são chamados de atributos de *dimensão* -- onde *dimensão* é designação dada ao agrupamento criterioso de atributos qualitativos de um fato comercial e rotulados segundo um enfoque administrativo relevante -- por exemplo, ainda utilizando como *fato* venda, no caso da *dimensão* tempo, instante da venda, estação climática, etc., para a *dimensão* geografia, cidade da venda, estado, região, etc..

Afim de possibilitar a execução destes tipos de consultas, utiliza-se, não mais a forma clássica de processamento transacional, devido à mesma não oferecer a flexibilidade necessária no trato dos dados para consultas deste gênero, e sim a forma de processamento analítico denominado de OLAP (do inglês, "On-line Analytical Processing")<sup>13</sup> [KIMBALL, 1996; TANLER & DROST, 1996; GILL & RAO, 1996; BRACKETT, 1996, BURETTA, 1997; Yazdan98; WHITE, 1997; MENNINGER, 1997]. Por outro lado, constatamos que em [INMON, 1996], assim como na sua versão portuguesa [INMON, 1997], encontramos a sigla OLAP como "Online Analog Process", o que causa uma ligeira confusão para os leitores. Porém, essa divergência de significados da palavra OLAP não afeta o seu entendimento, havendo um consenso geral, sobre o que se espera da sigla em termos técnicos. Acreditamos tratar-se de um equívoco neste exemplar, pois, posteriormente, [INMON ET AL.,

---

<sup>12</sup>Atributos são características qualitativas e quantitativas de um dado.

<sup>13</sup>Para a maioria dos autores este é o significado da sigla OLAP, diferenciando apenas a palavra "analytic", usada por uns no lugar de "analytical", tendo ambos os adjetivos o mesmo significado em inglês.

1997] definem esse termo de acordo como os demais autores, ou seja, não mais como “Online Analog Process”, mas sim como “On-line Analytical Processing”.

Dedicamos um capítulo inteiramente a esse tema, o capítulo 3, devido à sua importância contextual na realidade atual do DWing (muitos confundem OLAP com DWing, conforme veremos na seção 4.6.1, do capítulo 4) e por ele possuir um rico e polêmico histórico. Como exemplo de EIS, temos nas modernas interfaces gráficas as chamadas ferramentas OLAP, tais como planilhas eletrônicas, que disponibilizam, entre muitas facilidades de manipulação, visualizações mutidimensionais dos dados.

### **2.2.2.3. Decision Support System (DSS)**

Ainda bastante confundidos com os outros sistemas gerenciais e cada vez mais populares, devido à constante divulgação da sigla pela mídia especializada em propaganda, os DSS, sistemas de apoio (ou suporte) à decisão (do inglês “Decision Support System”), estão em voga no âmbito empresarial. Apesar de tanta divulgação, na prática, os seus resultados ainda são bastante tímidos comercialmente e existe pouca evidência concreta de sistemas DSS “puros”.

Dizemos DSS “puros”, pois, conforme [KELLY, 1996], os DSS estão intrinsecamente ligados à heurística. Ele lembra que heurística vem do grego e tem o significado de inventar; descobrir. No DSS a heurística é fundamental. O DSS é indicado quando precisamos explorar os dados e se possui e um ambiente desestruturado (de difícil entendimento), uma pobre definição ou mesmo a ausência de um problema especificamente definido.

Carentes de uma formalização, ou pelo menos de uma tentativa de efetivá-la, como se deu com EIS/OLAP (vide seção 3.1, do capítulo 3) e ainda com poucos resultados aparentes a nível de mercado, o mais comum são os fornecedores recorrerem aos conceitos de EIS e mesmo MIS de forma a “encorpá-lo” e torná-lo mais consistente, e conseqüentemente, atraentes comercialmente. Daí a razão de termos utilizado a expressão

DSS “puro”, pois mistura-se fluentemente EIS e DSS, “empacotando-os” ao final como DSS, ou mesmo dizendo-se que o DSS é o MIS moderno, aproveitando-se de um conceito que já possui um histórico. Veremos mais adiante, ainda nesta seção, motivos suficientes para destacar o DSS dos demais. Constatamos também que uma outra forma de alguns autores referirem-se ao DSS é sob o título de *Data Mining*, ou seja, traduzido literalmente do inglês, “mineração de dados”, que não deixa de estar congruente com o que entendemos e chamamos aqui de DSS “puro”.

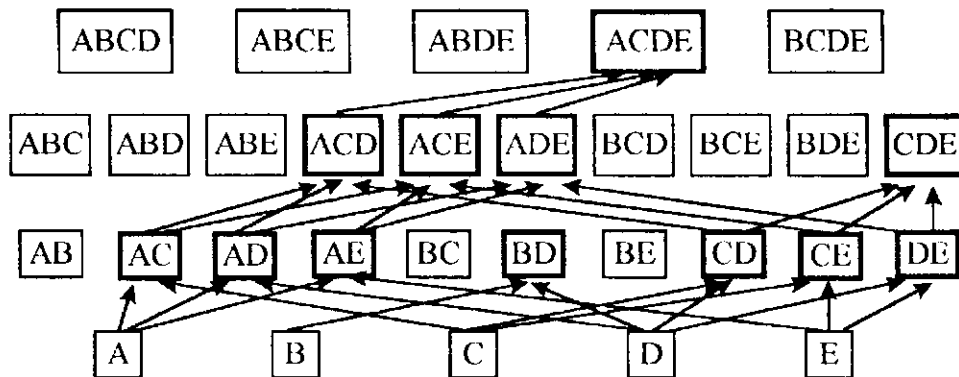


Figura 2.5 - Como Funcionam as Associações [KOURY, 1996]

A Figura 2.5 nos mostra o funcionamento de um tipo de técnica de *Data Mining* chamada de *associação*, que é a identificação de eventos que ocorrem juntos a partir de coleções lógicas [KOURY, 1996]. Nela estão “todas as combinações possíveis entre os cinco itens A, B, C, D e E. As setas apontam as combinações significativas que ocorrem com relativa frequência, como mostra o exemplo [...]. Usuário de *Data Mining*, por associação, se defronta tipicamente com centenas ou milhares de itens, resultando num número imenso de combinações. O desafio consiste em achar combinações significativas entre os dados sem ter de examinar um número excessivo de combinações falsas” [KOURY, 1996]. De um modo geral, *associações* são relacionamentos significativos entre dados armazenados em grandes BD. Esta técnica foi desenvolvida especialmente para resolver problemas de mercado de varejo, conforme esta mesma fonte. Seu princípio está na análise dos registros originados dos pontos de venda (tipos caixas de supermercados, lojas, etc.), a fim de otimizar a disposição da arrumação dos itens (nas prateleiras das seções)

pela identificação de quais deles tem algum tipo de associação com outros produtos na sua venda. Dessa forma visa-se a facilitação, para os clientes, da busca pelo consumo dos itens.

[GUROVITZ, 1997] lembra do clássico exemplo das fraldas e cervejas, que é uma típica solução de *Data Mining* por *associação*, ocorrido numa das maiores redes de varejo dos Estados Unidos. Lá descobriu-se uma estranha relação entre o consumo de fraldas descartáveis e cervejas. Assim sendo, puseram os produtos lado a lado. O resultado foi que as vendas de ambas dispararam. A explicação está no fato dos compradores serem, em geral, homens e aproveitarem para comprar cervejas. Ele ainda cita outros casos como o da rede varejista que descobriu que a venda de colírios aumentava na véspera de feriados. Nesse último caso a explicação do “porque os colírios eram mais vendidos” ainda não está clara. Ele ainda cita outros casos de sucesso no uso de *Data Mining*.

Rakesh Agrawal, pesquisador de algoritmos *Data Mining*, afirma que “existem ferramentas estatísticas disponíveis para lidar com parte dos dados, mas o problema é o enorme volume”, acrescentando, “pesquisar relacionamentos entre itens de varejo se torna um problema sério quando envolve 10 mil itens, como tipicamente é o negócio de varejo” [KOURY, 1996].

O *Data Mining* possui dois diferentes ramos, conforme afirmam [GILL & RAO, 1996]. Um voltado para análises estatísticas e outro voltado para a descoberta do conhecimento (do inglês “Knowledge Discovery”). *Data Mining* tem sido amplamente estudado pela academia. Inúmeros e complexos algoritmos têm sido elaborados sobre esse assunto. Alguns rápidos e voltados para grandes BD, usando para isso regras de associação [AGRAWAL & SRIKANT, 1994]. Porém, para ver a efetividade destes algoritmos na prática, os especialistas em conhecimento tiveram que se conscientizar de que se precisa de *dados corporativos consolidados, ou seja*, dados da organização atualizados, sem ambiguidades, nem redundâncias, para se obter algum resultado dentro do esperado. [KELLY, 1996] ressalta essa questão da concordância de vários autores quanto a questão imperativa do uso de dados corporativos (vide seção 2.3) neste tipo de processamento, enquanto que salienta sobre as diversas discordâncias em relação ao nível de granularidade dos dados requerido nesses sistemas.

É importante notar que nem todas as razões pertencentes a uma análise de reconhecimento de padrões ou mesmo tendência de comportamento são claramente inteligíveis. Justamente aí é que está a maior importância e contribuição dada pelos DSS. Por isto, inclusive, que se denomina de a Era da Produtividade dos *decision makers* a junção dos DSS com os EIS/OLAP.

Conforme dito na seção 2.1, sobre MIS, há quem faça uma correlação temporal entre os MIS e DSS, achando que um é uma evolução natural do outro. Para melhor abordarmos esse tema, inicialmente vamos definir decisão. Herbert Simon, apud [KELLY, 1996], no seu trabalho “The New Science of Management Decision”, foi quem primeiro identificou a distinção entre *decisões* ‘programadas’ e ‘não-programadas’, acrescentando que, para Simon, “‘*decisões*’ são não-programadas, não-estruturadas e conseqüentes”.

[KELLY, 1996] ainda cita o trabalho de Gorry e Scott-Morton, “A Framework for Management Information Systems”, baseado no trabalho de Robert Anthony, sobre o ato de decidir de formas estruturadas e não-estruturadas, que rejeita a idéia da época de um só sistema para abranger todas as necessidades gerenciais. Conforme vimos na seção 2.2.2.1, a pretensão daquilo que esperavam do MIS era deveras ambiciosa, fazendo-se uma adequação conceitual modular mais realista, o que resultou na prática nos sistemas MIS, EIS e DSS.

Gorry e Scott-Morton, conforme [KELLY, 1996], se basearam no trabalho “Planning and Control Systems” da “Harvard Business Scholl” de Robert Anthony que descreve três categorias de atividade gerencial, sendo elas: controle operacional, controle gerencial e planejamento estratégico, isto é, a execução de tarefas, supervisão daqueles engajados em executar as tarefas e decisão sobre os objetivos da organização e o desenvolvimento de recursos na busca desses objetivos, respectivamente. A Figura 2.3, da seção 2.2.1, no capítulo 2, exhibe graficamente essa distinção entre as categorias de atividade gerencial, ao mesmo tempo que situa os SI por nos vistos.



De forma inequívoca, [KELLY, 1996] relata: “O ato de decidir estruturadamente foi consignado por Glory e Scott-Morton para o mundo do MIS, o qual produz relatórios predefinidos em intervalos predefinidos. Decisões desestruturadas foram localizadas dentro do arcabouço de Glory e Scott-Morton em um novo ambiente de SI o qual foi chamado por eles de DSS. Eles também notaram que a habilidade requerida neste novo ambiente seria claramente diferente daquele requerido no mundo MIS”.

Outro aspecto um tanto quanto confuso na literatura é a relação feita entre EIS e DSS. Brackett, por exemplo, junta EIS (, OLAP) e DSS e chama de *suporte à decisão*, porém, ao final ele define “processamento de suporte a decisão como processamento analítico” [BRACKETT, 1996]. Chega-se até a afirmar que o “EIS tem um importante papel na figura maior do DSS” [INMON ET AL., 1996]. Isso porque para alguns autores tais como [INMON ET AL., 1996], os DSS são sinônimos de SG.

Há também quem envolva o conceito de OLAP na definição de DSS [BISCHIOFF & ALEXANDER, 1997]. Não acreditamos que devemos associar OLAP ao DSS, pois as características do DSS estão mais para outro tipo de processamento que, de forma diametralmente oposta ao analítico, está mais para sintético. Conforme o Dicionário Aurélio, sintético vem do grego *synthetikós*, o 'que compõe ou reúne'; 'hábil em compor' e tem como um dos seus significados de '*relativo a síntese*'. Já síntese, também do grego *synthesis*, 'composição', pelo latim *synthese*, significa entre outros sentidos '*operação mental que procede do simples para o complexo*'; por extensão *resumo*. Ora, o DSS opera perfeitamente nesse sentido de caminhar no sentido do simples, os dados, para o complexo, as informações estratégicas, com o objetivo de achar um padrão, uma tendência, etc., atuando a sua “operação mental” no “cérebro eletrônico” do computador e chegando a um resumo deduzido. Dessa forma podemos denominar, coerentemente com as outras siglas em inglês, o DSS como o tipo de processamento OLSP (“On-line Synthetical Processing”). Pode parecer a princípio mais um neologismo (entre os muitos que há) neste tema, mas é de fundamental importância essa distinção entre OLTP e OLAP, da mesma forma que tornou-se essencial distinguir OLAP de OLSP.

Encontramos na nossa pesquisa bibliográfica dois autores que reconheceram também a necessidade de se destacar o tipo de processamento dos DSS criando um denominação nova. [BERSON & SMITH, 1997] ao referirem-se ao processamento dos DSS, utilizam a denominação de OLCP (“On-line Complex Processing”). Eles reconhecem este como um processamento complexo, pois está baseado no conhecimento da totalidade dos dados organizacionais corporativos consolidados e consistentes, de forma a aplicar *complexos* algoritmos de *Data Mining*, para reconhecer padrões ou mesmo tendências de comportamento dos dados. Ainda conforme [BERSON & SMITH, 1997], os SGBD OLCP devem não apenas processar grandes quantidades de dados, mas sim ter a “habilidade de entendê-los”.

Apesar de valorizar a iniciativa da distinção e destaque ao processamento sintético como OLCP, consideramos não muito feliz a utilização do termo “complexo”, pois trata-se de um termo deveras impreciso, pois o conceito de complexo varia em função do referencial, já a terminologia “sintético” traduz com mais precisão a idéia.

#### 2.2.2.4. Comparativo entre MIS, EIS e DSS

A Figura 2.6 nos exhibe um resumo comparativo em forma de tabela, de maneira a elucidar resumidamente os SI do tipo SG, tirando qualquer dúvida que ainda possa pairar sobre os esses sistemas. Nela vemos os seguintes parâmetros diferenciais:

- **âmbito organizacional** – é o escopo de abrangência que o sistema pode/deve atingir;
- **interação** – é a forma com que o usuário interage com o sistema;
- **foco da análise** – é a expectativa do usuário na interação com o sistema, ou seja, o que podemos esperar de forma genérica desses sistemas, resumidamente expressa aqui por uma pergunta;

	MIS	EIS	DSS
Âmbito Organizacional	Departamental, isto é local, setorial	Corporativo, ou seja, global	Corporativo, ou seja, global
Interação	Consultas limitadas e previsíveis	Consultas aleatórias baseadas no conceito de dimensão	Consultas baseadas em metas
Foco da Análise	O que está acontecendo nos negócios atualmente?	O que está acontecendo no negócio a nível corporativo ultimamente ?	Quais as evidências, prognósticos e providências ?
Fator de Confiança	Derivado dos dados	Derivado do usuário	Derivado dos dados
Tipo de Processamento	OLTP	OLAP	OLSP
Granularidade	Primitivos, a nível de detalhe	Dados resumidos e agregados, podendo também estar dados primitivos	Ainda sem consenso, podendo ser dados primitivos, agregados e/ou resumidos.
Atualização dos dados de entrada	Dinâmica	Estática, programada	Estática, programada
Volatilidade no armazenamento	Corrente	Estática, geração de histórico	Estática, geração de histórico
Abrangência temporal	Atualidade	Variante no tempo	Variante no tempo
Repositório	Organizado com o objetivo de oferecer respostas imediatas a solicitações de armazenamento e recuperação dos dados	Organizado com vista à facilitação das consultas do usuário na forma de navegação nos dados na ótica de dimensões	Organizado com a finalidade de ser voltado para acessos automáticos orientados pela ferramenta de <i>Mining</i> .
Técnica de Análise	Consultas ad-hoc	Seleção transversal dimensional	Descobrimto automático
Controle do processo	Predefinido pelo usuário	Controlado na interação com o usuário	Controlado pelas evidências do <i>Mining</i>
Estado Tecnológico	Maduro	Maduro	Maduro em análises estatísticas e emergindo na descoberta do conhecimento
Contribuição	Controle operacional	Acompanhamento gerencial	Planejamento estratégico

Figura 2.6 - Comparativo entre os SG

- **fator de confiança** – é a responsabilidade no uso dos dados, ou seja, ao interagir com o sistema, evidencia-se uma maior ou menor participação do usuário no resultado do seu uso. Traduz a parcela de confiabilidade do sistema analisado-se como uma ferramenta;
- **tipos de processamento** – é a forma com que os dados são manipulados pelos sistemas;
- **granularidade** – e a forma com que os dados que entram no sistema, segundo a sua importância a nível de consolidação: detalhadamente, agregadamente, ou ainda, resumidamente;
- **atualização dos dados de entrada** – é a forma com que os sistemas necessitam encontrar os dados de entrada;
- **volatilidade no armazenamento** – indica o estado de conservação dos dados de entrada para o sistema;
- **abrangência temporal** – é a delimitação do escopo do tempo dos sistemas;
- **técnica de análise** – é a técnica<sup>14</sup> básica empregada na consulta;
- **controle do processo** – é a forma de domínio do andamento do processo;
- **estado tecnológico** – é o estado de maturidade da tecnologia empregada;
- **contribuição** – é a área de maior atuação e influência organizacional do sistema.

Ainda nesta figura, podemos notar através dos parâmetros *atualização dos dados de entrada*, *volatilidade no armazenamento*, *abrangência temporal* e, principalmente, *tipo de processamento*, que MIS difere bastante dos demais. Esse é um demonstrativo da distinção não só em relação à manipulação e ao armazenamento dos dados – que no caso dos sistemas EIS e DSS precisam de dados históricos, dados acumulados, ou seja, armazenamento temporal –, mas também ao tipo de processamento empregado. Conforme já mencionado anteriormente na seção 2.2.2.1, que justifica a limitada abrangência dos MIS a nível efetivo e aponta os dois pontos básicos. Em outras palavras, percebemos mais claramente no quadro que, além das questões ‘tipo de processamento’ (diferente do

---

<sup>14</sup>A técnica “seleção transversal dimensional”, também conhecida como “slice and dice”, será detalhadamente apresentada na seção 3.3 do capítulo 3.3. sobre Operações.

transacional utilizado) e de ‘produzir posições atualizadas a nível corporativo’ (não ser tão trivial), temos também a questão do armazenamento temporal – conforme indicam os parâmetros *volatilidade no armazenamento e abrangência temporal* –, que requerem uma mudança conceitual ambiental não observada antes, conforme veremos na seção seguinte.

### 2.3. Restrições Ambientais para os Sistemas Gerenciais

Nesta seção veremos, além das diferenças marcantes entre as características dos ambientes dos SA e EIS/DSS, a principal questão que fez com que inicialmente o MIS e posteriormente o EIS não obtivessem o resultado esperado, por razões distintas, porém pelo mesmo motivo, isto é, a dificuldade da obtenção dos *dados corporativos consolidados*.

Certamente uma das causas foi (e continua sendo) o fenômeno da descentralização departamental que as organizações sofrem, apoiadas infra-estruturalmente pela microinformática. Em especial, em relação aos sistemas produzidos neste ambiente da microinformática, [INMON, 1996] define como “arquitetura de desenvolvimento espontâneo”, o fenômeno de geração indiscriminada e setorial dos SA (também conhecida como “teia de aranha”), em uma organização. A Figura 2.7 nos mostra objetos vazados e

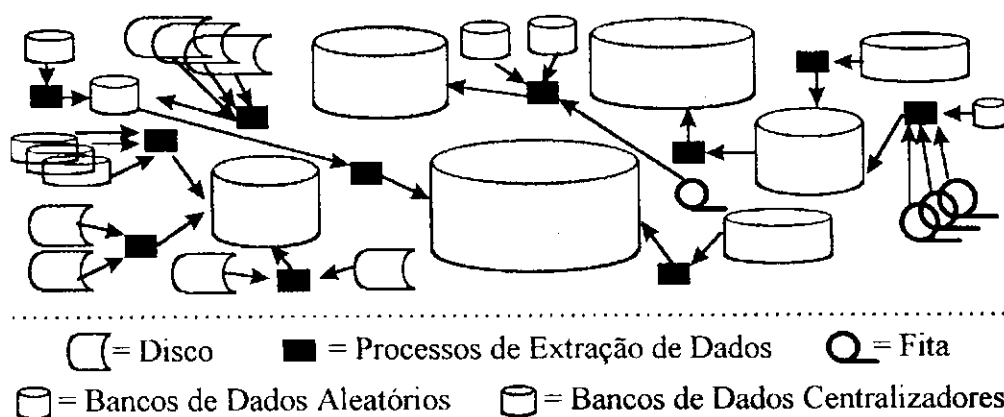


Figura 2.7 - Arquitetura de Desenvolvimento Espontâneo (A Teia de Aranha)

coloridos com cinza representando os vários repositórios gerados em uma corporação e ligados por setas que orientam o fluxo de dados de/para os quadrados que indicam processos de *extração* de dados (melhor detalhado na seção seguinte com um caso real). A cor cinza foi usada para destacar os repositórios obtidos a partir de outros previamente existentes, através do uso de técnicas de replicação<sup>15</sup> de dados, cada vez mais presentes nas grandes corporações.

Como podemos observar nesta figura hipotética, acompanhando o fluxo de dados mostrado pelas setas, chegamos a quatro bases distintas destacadas na figura pela cor cinza. A partir delas teremos que fazer uma consolidação, checando a consistência dos dados, a fim de obter uma posição globalizada. Algo que naturalmente requer consumo de recursos (físicos e de pessoal), muitas horas de máquina e nem sempre agrada, ao final, aos seus usuários.

Conforme já estamos cientes (vide na tabela da Figura 2.6 da seção anterior, item volatilidade no armazenamento), existe mais uma forma especial de armazenamento dos dados para os sistemas EIS e DSS que não atinge o MIS (pelo menos como ele acabou sendo implementado, voltado exclusivamente para o controle), que é, além dos dados corporativos estarem consolidados, eles precisam guardar o seu histórico, ou seja, ser temporais. A temporalidade dos dados será mais detalhada na seção 4.3.2 do capítulo 4, onde levantaremos também a questão do tempo de armazenamento, (seção 4.3.5), devido às questões temporais terem grande importância neste estudo e apesar da pouca atenção que tem sido dada a este ponto pela (quase) totalidade dos autores consultados.

Antes, porém, é interessante notarmos as várias diferenças entre os sistemas EIS/OLAP e DSS/OLSP e os SA atentando para a observação de Kimball quanto à importância da separação física destes ambientes. Segundo as diferenças existentes pelos critérios de [KIMBALL, 1996]: tipos de usuário, conteúdo e estrutura dos dados, hardware e software, administração e gerenciamento dos sistemas, e os seus ritmos diários de processamento, ele reconhece a necessidade de dois ambientes de processamento distintos.

---

<sup>15</sup>A terminologia "replicação", usada na literatura para identificar a manutenção automática de réplicas de tabelas relacionais, consistentemente, em um ambiente distribuído, tem sido usada, inclusive, para caracterizar a propagação assíncrona de dados de um sistema para outro [BURETTA, 1997].

Na Figura 2.8, [BRACKETT, 1996] nos mostra, sob forma de quadro comparativo, as diferenças entre a abordagem operacional, ou seja os SA, e a voltada para o suporte à decisão, mais precisamente os EIS e DSS, segundo dez parâmetros importantes. As diferentes características encontradas nelas, refletem as particularidades de cada tipo de ambiente. Os parâmetros utilizados foram:

	<b>Operacional</b>	<b>Suporte a Decisão</b>
<b>Volatilidade</b>	Dinâmica	Estática
<b>Fluência</b>	Corrente	Histórico
<b>Instância</b>	Tempo simples	Variante no tempo
<b>Granularidade</b>	Primitiva e detalhada	Derivada e resumido
<b>Atualizações</b>	Contínuas, aleatórias	Planejadas, periódicas
<b>Estrutura</b>	Estática	Dinâmica
<b>Normalização</b>	Altamente normalizada	Moderadamente Normalizada
<b>Índices</b>	Mínimo de índices	Máximo de índices
<b>Flexibilidade</b>	Baixa, repetitiva	Alta, heurística
<b>Desempenho</b>	Alto, resposta rápida	Baixa, resposta longa

Figura 2.8 - Diferenças entre Processamentos Operacionais e de Suporte à Decisão [BRACKETT, 1996]

- **volatilidade** - indica se os dados serão ou não atualizados durante a execução de uma consulta;
- **fluência** – indica a gama de tempo utilizada no armazenamento dos dados;
- **instância** – indica a influência do tempo ligada a uma dada consulta;
- **granularidade** - exprime como os dados serão armazenados para consultas;
- **atualizações** - exprime o critério de atualização;
- **estrutura** - exprime a forma interna de organização dos atributos;
- **normalização** - indica, para o caso de BDR, o critério de normalização;
- **índices** - expressa a importância da organização para acesso aos dados;
- **flexibilidade** - expressa a qualidade do processamento;
- **desempenho** - é o tempo de resposta esperado no processamento.

Apesar de diferentes características dos SG e SA, a entrada para o processamento dos SG é sempre originário de pelo menos um SA. Já que os SG são alimentados pelos SA, é até um tanto intuitivo se imaginar que a convivência desses dois tipos de sistemas em um só ambiente não seria realmente o mais razoável. Porém, muitos fornecedores ignoraram esse fato no intuito de comercializar os seus produtos, o que causou grande impacto operacionalmente nas instituições que partiram para implantar ambos tipos de sistemas em um só ambiente. Enquanto os SA precisam estar sempre *eficientemente atualizados* (livres de outros tipos de processamentos que consumam máquina) de forma a permitir o fluxo normal da empresa, os SG, e mais precisamente os EIS e DSS, necessitam de:

- histórico de dados;
- consistência temporal, isto é, estar coerentemente atualizado de forma corporativa, não necessariamente *on-line*, conforme as regras do negócio. Por exemplo, uma análise – que geralmente requer outras seguintes – deve capturar dados de forma não forçosamente a mostrar a posição mais atual, mas sim obrigatoriamente uma posição consistente e coerente, que reflita a realidade corporativa consolidada;
- **armazenar seus dados em repositórios com características especiais, isto é, voltada para recuperação dos dados de forma otimizada conforme o tipo de processamento desejado (analítico ou sintético).** Este aspecto merece um destaque por se tratar de um ponto nevrálgico em projetos de DWing, sendo ele um possível fator de sucesso (ou de fracasso, caso seja ignorado), e
- processar os dados de forma analítica (gerando dados agregados e resumidos para o caso dos EIS) ou forma sintética (através de algoritmos genéticos, redes neurais, etc.).

O impacto da não obtenção dos *dados corporativos consolidados* para o MIS foi o de não se conseguir atingir o *nível corporativo* conforme os seus objetivos conceituais (abordados na seção 2.2.2.1). Já para o EIS o cenário foi diferente, pois não se imaginava que o contexto de obtenção dos *dados corporativos consolidados* diariamente demandasse um alto grau de dificuldade. Assim, estamos diante da problemática do processo de extração dos dados corporativos em grandes organizações.



O autor deste documento, no exercício da função de analista de sistemas de um grande banco cuja alta gerência optou por instalar EIS, vivenciou essa questão e mostra o panorama de uma organização que, mesmo tendo considerado a questão das diferenças de ambientes de processamentos operacional e de suporte à decisão, sentiu os reflexos notando impactos generalizados por quase todo o parque computacional. Vale ressaltar que nesta época nem se falava nos atuais sistemas DSS (vistos na seção 2.2.2.3).

### 2.3.1. A Problemática para a Efetivação dos EIS e DSS

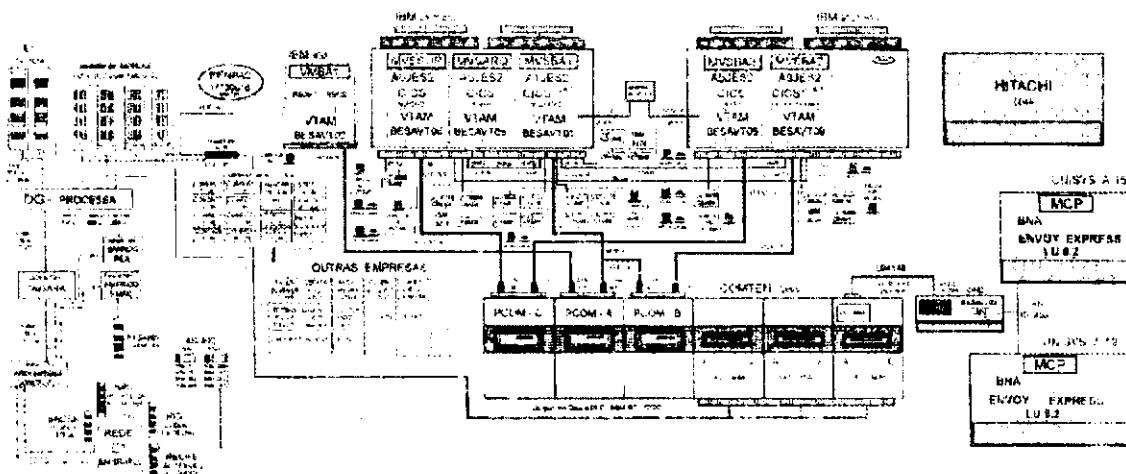


Figura 2.9 - Projeto de Topologia da Rede do Antigo Banco Econômico<sup>16</sup>

Para elucidar o problema que norteia as grandes corporações, nada melhor que exemplificarmos com um fato ocorrido no antigo Banco Econômico S/A<sup>17</sup> ao adquirir um EIS. Possuindo não só uma arquitetura de desenvolvimento espontâneo, isto é, uma estrutura descentralizada com centenas de micros (distribuídos por todo o Brasil), mas

<sup>16</sup>Esboço gentilmente cedido pelo colega bancário Carlo Muller.

<sup>17</sup>Em 1995, o Banco Econômico S/A possuía 279 agências totalmente informatizadas, 2.667 caixas e terminais de clientes, 128 postos de serviços, 114 terminais eletrônicos, 720 mil clientes no Brasil e exterior, 10.200 funcionários, 74 mil acionistas e era o 7º maior banco privado do país [TONIELLO ed., 1995].

também cinco *mainframes* de dois fornecedores distintos, essa organização detinha uma grande redundância de dados nos seus SA, conforme sugerido Figura 2.9. Essa redundância se dava, principalmente, devido à necessidade da criação de vários repositórios, a fim de transferir os dados inter-plataformas (entre os *mainframes*, Figura 2.10), utilizando-se para isso várias técnicas de replicação. [BURETTA, 1997] aborda algumas destas técnicas empregadas em detalhes. Contudo, como temos visto, para que o EIS funcione plenamente é imperativo que tenhamos os dados organizacionais consolidados corporativamente e de forma consistente. A Figura 2.10 nos mostra outro aspecto importante e que dificultava essa organização a nível corporativo, ou seja, a pluralidade e diversidade tecnológica dos repositórios utilizados.

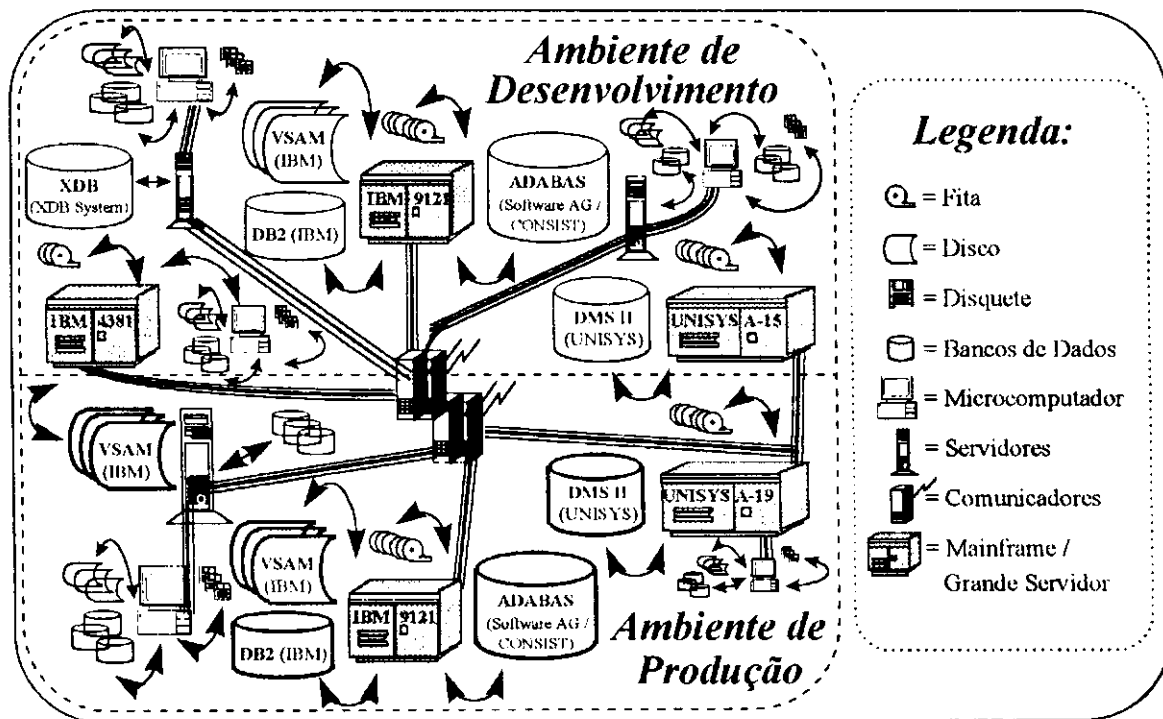


Figura 2.10 - Ambientes de Repositórios de Dados do Antigo Banco Econômico

Para alimentar o EIS de dados, inicialmente a empresa utilizou-se de um *mainframe*, a máquina dedicada aos sistemas de produção (MVSBA1, na Figura 2.9). As rotinas de *extração* de dados, voltadas à alimentação da base de dados do sistema EIS, consumiam muito tempo de processamento e atrasava as outras aplicações de produção. Então, imediatamente, transferiram o processamento dessas rotinas do EIS para a máquina

da área de desenvolvimento de sistemas (MVSBA2, nesta mesma Figura 2.9). Neste momento, o próprio autor deste documento sentiu as conseqüências dessa mudança. Agora ele próprio sendo vítima do impacto dessas rotinas, que, por serem voltadas à cúpula da organização, possuíam alta prioridade de execução em máquina. Porém, mesmo assim, levava quase toda a tarde rodando, em detrimento de outras atividades que utilizavam o parque computacional. Além disso, outras conseqüências existiram, tais como: muitos contratos de consultoria externa tiveram seus prazos prolongados, implantações de produtos adiadas, etc., o que gerou um custo adicional extra para o banco, que não foi nada agradável aos seus dirigentes.

Todo o entrave foi causado pelo processamento das rotinas de extração de dados, responsável por levar dados corporativos consolidados e consistentes, para a base de dados do EIS para então ele poder “produzir posições atualizadas a nível corporativo”. Eram basicamente dois programas em execução que consumiam o dia inteiro rodando. Eles eram fundamentais para que o EIS funcionasse de forma atualizada diariamente. Contudo, a demora no processamento diário de atualização da base de dados do EIS refletiram na simpatia deste tipo de sistema, na organização. Isso gerou um descontentamento amplo, tanto do corpo técnico – muitos tendo seus trabalhos afetados em prazos –, quanto da própria alta cúpula – então usuária do sistema –, que não se conformavam com tanta demora na atualização da base dos dados e toda a repercussão negativa causada pela sua tão recente implantação.

Relacionamos agora o ocorrido nesta organização em 1992, que não trata-se de um caso singular, mas um dos muitos ocorridos para aqueles que aderiram ao EIS naquela época, com algumas observações feitas por [INMON, 1996] referentes às diferentes necessidades entre o ambiente operacional e o informacional / analítico:

- **os dados são fisicamente diferentes** - razão pela qual tivemos forçosamente que executar sempre as duas rotinas “pesadas” que preparavam os dados corporativamente para o EIS, de forma consolidada e consistente;
- **as tecnologias são fundamentalmente diferentes** - exatamente por essa razão não foi possível se utilizar um dos ambiente já existentes. Foi gerado um novo

ambiente exclusivo para executar o EIS, apesar da extração e consolidação de dados ser sempre gerada forçosamente em um dos *mainframes*. Neste ponto o banco agiu da melhor forma possível e aconselhável, separando fisicamente o ambiente onde rodava o EIS dos operacionais existentes;

- **a comunidade de usuários é diferente** – o EIS tem como comunidade usuária a alta gerência, porém a sua alimentação dependia da utilização do ambiente operacional. Esse ponto deveria ter sido considerado, pois muitos tiveram seus trabalhos afetados pelas rotinas de extração de dados para o EIS, sem que tivessem a consciência do que realmente aquilo significava. Alguns mesmo ficaram inconformados com a lentidão dos sistemas, gerando um desestímulo geral que, por fim, foi traduzido em baixa produtividade circunstancial;
- **o processamento de dados é diferente** - aqueles que conviviam com processamento operacional não compreendiam outro tipo de processamento e terminavam querendo viabilizar o impossível, ou seja, gerar aplicativos analíticos em ambiente transacional. O que mais uma vez resultava numa baixa produtividade.

Não é necessário dizer que esse sistema acabou tornando-se um incômodo generalizado para a corporação, pois o corpo técnico não se conformava com a sua existência, questionando-a. E a cúpula administrativa, que recebia mais queixas do seu corpo técnico do que recomendações deles no seu uso, utilizava-o timidamente. Muitos dos dirigentes eram senhores de idade, que, apesar de conhecerem muito do sistema bancário, não tinham grande simpatia e familiarização no uso do computador como ferramenta cotidiana, reservando-o ao estritamente essencial. Para eles, o que muitos viam na tela eram informações que eles próprios intuía pela sua própria experiência no ramo (alguns com mais de vinte anos de organização) ou mesmo pelas deduções serem muito óbvias na prática do negócio bancário. Com experiências como esta, o EIS ficou marcado negativamente em muitas corporações. Porém a idéia, não só não foi abandonada, mas hoje é reconhecidamente valorizada para as organizações como fundamental na busca e impressão do seu diferencial competitivo na atual economia dita globalizada. E para este autor (que em muito teve seu trabalho e da equipe de consultores afetado) serviu de experiência e motivação, na qualidade de técnico, na direção da busca de um novo

conhecimento, de uma nova cultura de vanguarda de uso dos dados, que até então desconhecia.

### 2.3.2. A Solução Encontrada com o Uso do *Data Warehouse*

Hoje sabemos que, para se pensar em implantar um EIS/DSS, é necessário se ter em mente todo um projeto de geração de um repositório corporativo temporal especialmente voltado para esse fim. A esse repositório específico damos o nome de DW e ao projeto de construção DWing. Ele deverá contemplar vários aspectos, conforme veremos no capítulo 4, e principalmente prever a geração e manutenção dos dados corporativos em sua base, acumulativamente e de forma a disponibilizá-los sempre que necessários a tais SG sem comprometer o andamento dos demais sistemas de produção da organização (os SA).

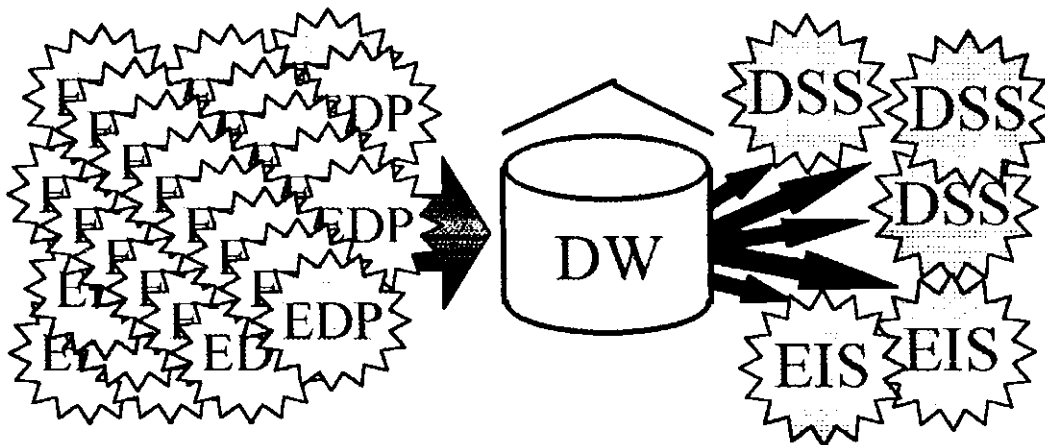


Figura 2.11 - A Localização do Data Warehouse

A Figura 2.11 nos dá uma idéia do ponto onde se situa o DW, de onde as informações fluem (das estrelas simbolizando os vários EDP da organização para o DW de forma consistente, consolidada e acumulativa), para daí estarem disponíveis para os sistemas EIS e DSS (também simbolizados por outros tipos de estrelas). Mas, nem sempre isso foi tão claro, conforme vimos na seção anterior sobre a nossa própria experiência enquanto analista de sistemas de um grande banco. Não se tratou de um caso singular, muitas foram as tentativas de várias corporações que, com projetos de EIS teoricamente

viáveis e exequíveis, demonstraram-se ao final impraticáveis. Muito foi gasto, não só em valores, mas também sob diversas formas de utilização de recursos (físicos e de pessoal), além das inúmeras tentativas de implantação, sem o sucesso esperado.

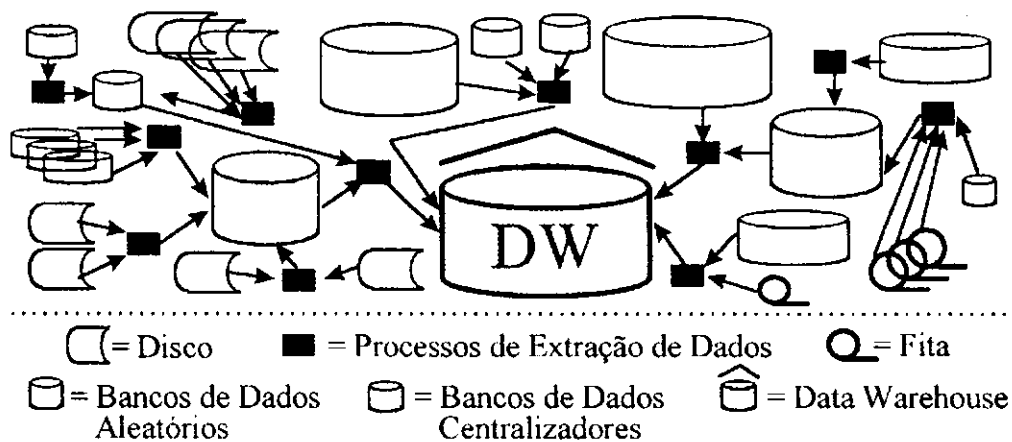


Figura 2.12 - Ambiente de Geração Espontânea Integrado com DW

O DW vem como solução para o problema da integração corporativa. Conforme a Figura 2.12, todas as mídias, através dos processos de extração de dados, convergem para esse repositório central, que é o DW, representado aqui na cor cinza e diferenciado dos bancos de dados centralizados por ter na parte superior um sinal em forma de acento circunflexo. De uma certa forma, é o resgate do conceito original de BD centralizado, onde todos os dados da corporação residiriam em um só repositório, com finalidade bem definida de armazenamento dos dados organizacionais e cercado de rotinas para torná-lo eficiente, consistente e confiável.

Contudo, apesar do DW ser uma solução na viabilização da produção de informações estratégicas a nível corporativo, a construção e manutenção desse repositório centralizado não é trivial. Uma das razões é oriunda da necessidade de armazenamento histórico antes tido como uma função secundária no processamento operacional e agora fundamental processamento estratégico.

Muitos associam o DW ao processamento analítico de forma a confundir as tecnologias, porém veremos mais adiante, na seção 3.7 do próximo capítulo especialmente destinado ao OLAP, que essa associação, preferencialmente, não deve ser feita.

Inmon, reconhecido como o “o pai do DW”<sup>18</sup>, é taxativo quando diz que “o DW proporciona o alicerce de dados de que o analista de EIS precisa para atender eficientemente ao processamento EIS”. Essa afirmação pode ser estendida ao DSS, pois a importância do DW é viabilizar a centralização dos dados da organização de forma consolidada e consistente, resolvendo, assim, o problema da “teia de aranha” já visto na Figura 2.7, da seção 2.3. O DW será amplamente visto na seção 4.3 e todo o processo a que ele está ligado será exaustivamente visto no capítulo 4 dedicado ao DWing.

#### 2.4. *Resumo*

Vimos a divisão dos SI em três eras distintas: automação, controle e produtividade dos “decision makers”. Essas eras existiram tanto no mundo do desenvolvimento da informática como ciência, quanto nas empresas ao optarem por enveredar pela informática em todos os seus níveis administrativos, sucedendo-se nesta mesma ordem de cronologia, isto é, elas são sucessivas no contexto evolutivo da informatização empresarial também. Mas, as eras seguintes não eliminaram, e sim incorporaram as anteriores.

Distinguimos então os SI em dois grupos. Os SA, que são sistemas voltados à produtividade operacional, e os SG, aqueles voltados ao gerenciamento empresarial. Os SA estão divididos em dois tipos: *legacy* e OLTP. Vimos também que não há uma concordância generalizada entre os autores a respeito da fronteira e mesmo do reconhecimento do que seja *legacy*, que para nós são os sistemas *batch*, e OLTP, que são os sistemas *on-line* transacionais.

---

<sup>18</sup>[INMON, 1997].

Os SG são divididos em MIS, EIS e DSS. O MIS foi o primeiro sistema conceitualmente voltado ao gerenciamento. Possuía pretensões grandes, contudo a tecnologia empregada na época acabou inviabilizando-o parcialmente. OLTP, então ideal para os SA, não era o tipo de processamento apropriado para o MIS, conforme a sua definição conceitual de um sistema gerencial e amplo o suficiente para “englobar todos os componentes da organização e todos os níveis de decisão”. Portanto, o MIS só pode contemplar o aspecto do controle gerencial, já que o estágio evolutivo da informática na época assim permitia.

Como uma forma alternativa do processamento vigente nas organizações, o OLTP, surgiu o OLAP, de forma a possibilitar maior interação com o usuário, aproveitando-se das modernas interfaces gráficas, e armazenando os dados em repositórios independentes e especialmente projetados para possibilitar consultas a várias óticas de atributos de dados operacionais sobre fatos comerciais, chamada Análise Multidimensional. Assim nascia o EIS, tal qual o MIS quanto às suas pretensões como SG, porém mais maduro e aproveitando-se da sua experiência, o EIS veio com pretensões de atingir o topo da pirâmide administrativa (vide a Figura 2.3, na seção 2.2.1).

Foi uma grande revolução a possibilidade, agora real, do usuário interagir com o computador na busca do conhecimento organizacional de forma a atingir o nível estratégico corporativo. Contudo, assim como os MIS, os EIS também não fizeram o sucesso imediato que se esperava. Isso porque para eles, mudando o processamento, não mais transacional e sim analítico, tudo estaria resolvido. Porém, foi a questão da descentralização generalizada que as empresas sofreram, a nível departamental com a adoção da microinformática (formando a conhecida “teia de aranha”, vide a Figura 2.7, na seção 2.3), que realmente influenciou negativamente os resultados das tentativas de implantação neste período. No entanto, muitas empresas chegaram a implantar EIS e tiraram muitas conclusões dessas experiências.

Subseqüentemente ao OLAP, o processamento analítico, surgiu o OLSP, este agora o processamento sintético, baseado na busca e identificação de informações extraídas dos dados corporativos, viabilizando o DSS/*Data Mining* base da busca do diferencial



comercial competitivo na economia globalizada. Desta forma, os sistemas OLTP, incontestáveis e imbatíveis tecnologicamente para o processamento voltado à produção organizacional, passaram a dividir o seu espaço no processamento de dados com os OLAP e OLSP, ou seja, os EDP (e MIS) continuaram existindo como antes, porém convivendo harmoniosamente com os novos sistemas EIS e DSS.

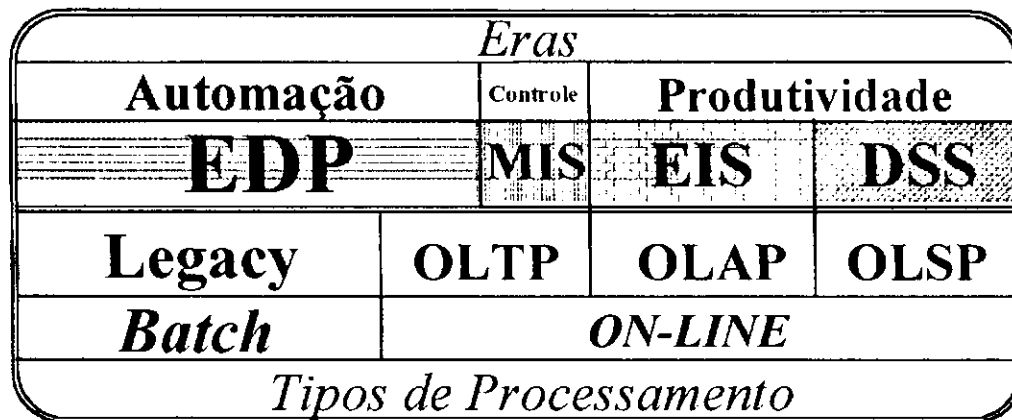


Figura 2.13 - Tipos de SI e seus Tipos de Processamento

A prática mostrou que a efetivação do EIS em grandes organizações (valendo também para os DSS), necessita da freqüente obtenção dos dados corporativos diários, de forma a integrar os seus repositórios de OLAP (e de OLSP), para viabilizar o(s) seu(s) funcionamento(s). A tarefa de extração diária dos dados departamentais consolidados corporativamente não é tão trivial (vide relato na seção 2.3.1), e os sistemas OLTP são em sua quase totalidade voláteis, ou seja, as informações são guardadas de forma que as novas sobrepõem as velhas, não guardando assim o histórico dos dados. Por isso, algo de novo, intermediário e de suma importância teve de ser criado para viabilizar o funcionamento dos sistemas EIS e DSS. A solução prática foi criar um repositório centralizado que mantivesse de forma atualizada a imagem corporativa organizacional. A esse repositório dá-se o nome de DW. Evidenciamos que ele é o intermediário entre os SA e os EIS / DSS distribuídos na organização. A Figura 2.13 resume o que nós vimos neste capítulo sobre os SI. O próximo capítulo abordará OLAP, pois, conforme dito antes, a sua importância contextual na realidade atual do DWing é grande.

### 3. On-line Analytical Processing (OLAP)

Neste capítulo abordamos mais detalhadamente o *processamento analítico* visto resumidamente na seção 2.2.2.2 do capítulo 2. Conhecido como OLAP, sigla do inglês “On-line Analytical Processing”, trata-se de um tipo de processamento de dados que, baseado em dados operacionais e seus atributos, provê ao usuário, a visualização dos fatos comerciais por diversos ângulos administrativos simultaneamente. É através do OLAP que se viabiliza a Análise Multidimensional (AM) dos sistemas EIS.

No OLAP, os atributos dos dados operacionais são separados em dois tipos segundo a sua natureza, isto é, atributos quantitativos, conhecidos como atributos de *fato*, e atributos qualitativos. Esses últimos, agregados segundo um critério de natureza comum, relevante e concernente estrategicamente aos negócios, são denominados de *dimensões*. Portanto, os atributos dos dados operacionais podem ser divididos em:

- ◆ atributos de *fato* – atributos quantitativos ligados diretamente a um fato comercial qualquer (compra, venda, etc.), por exemplo, para o *fato* venda, valor da venda, quantidade vendida, etc.;
- ◆ atributos de *dimensão* - atributos qualitativos ligados ao contexto do fato, isto é, tempo, geografia, etc., por exemplo, ainda utilizando como *fato* venda, no caso da *dimensão* tempo, instante da venda, estação climática, etc., para a *dimensão* geografia, cidade da venda, estado, região, etc..

A importância do OLAP tem crescido no âmbito comercial a cada dia, principalmente pela crescente competitividade gerada pela concorrência de mercado e, agravada em última instância pela globalização. Essa tecnologia, vem sendo cada vez mais necessária à sobrevivência e desenvolvimento das empresas, de forma a evidenciá-las positivamente como organizações modernas, perante os seus clientes atuais e futuros em

potencial. Veremos maiores detalhes a seguir, citando um breve histórico, que mostra a influência comercial desde o seu surgimento, a etimologia do termo OLAP, sua definição e finalidade, descrevendo as suas operações particulares, modelagens e, ao final, serão mostradas as arquiteturas existentes.

### **3.1. Histórico**

Inicialmente veremos a origem do estudo e a etimologia da sua sigla. Em seguida as diversas regras em que se baseiam, a polêmica gerada pelas condições com que foram originalmente definidas e, por fim, a influência do mercado da informática sobre esse ponto. É interessante salientar que, a rigor, o DW não fazia parte do estudo de OLAP, mas, em virtude dos problemas ambientais que restringem a sua perfeita implantação – conforme visto na seção 2.3 – acabou-se impondo o DW como elemento essencial em um projeto de OLAP.

#### **3.1.1. Origem**

Conforme [GILL & RAO, 1996], a partir da forte influência e sucesso causados pelas doze regras desenvolvidas por E. F. Codd, em 1985, para BDR, oito anos mais tarde, ele próprio, junto com S. B. Codd e C. T. Salley desenvolveram outras doze regras, agora, para o desenvolvimento de ferramentas para processamento analítico [CODD, 1993]. São os criadores da expressão abreviada OLAP, com o significado oriundo do artigo acima referido.

A repercussão deste segundo trabalho de Codd, referente a OLAP, não obteve o sucesso esperado, em comparação com o primeiro, sobre BDR. Segundo [KIMBALL, 1996], as doze regras são muito vagas para serem usadas como diretrizes de avaliação de um sistema dessa natureza, acrescentando que “a filosofia OLAP necessita de critérios mais específicos”. Isso foi devido ao trabalho ter sido concluído a partir de um estudo de caso

particular [KIMBALL, 1996; GILL & RAO, 1996]. Mas, nem por isso esse tema caiu no esquecimento, muito pelo contrário, ele tem crescido progressiva e amplamente.

Assim evidenciam-se, os conceitos de **dimensão** (de dados) como *um qualificador conceitual que provê um significado ou métrica, tal como geografia, tempo, etc.* e de **AM**, como sendo a *análise simultânea de múltiplas dimensões de dados* [GILL & RAO, 1996]. “A dimensão de dados é também o mais alto nível no caminho da consolidação dos dados” [GILL & RAO, 1996]. Na Figura 3.1 vemos um cubo sobre um eixo tridimensional ortogonal. Quando se deseja representar modernamente AM, geralmente utiliza-se esta figura geometria. Ela é usada apesar da AM poder ser n-dimensional. Mas, por razões das limitações de representação espacial no plano, o *cubo* é a figura geométrica que oferece maior expressividade.

Ainda na Figura 3.1, nos eixos “x”, “y” e “z” ficam as *dimensões*, referentes a fatos comerciais, escolhidas e relevantes ao corpo gerencial e concernentes aos negócios, podendo ser, por exemplo, tempo, produto e geografia, respectivamente. Um *cubo de dados* pode ser manipulado de forma a espelhar os mais diversos ângulos organizacionais, segundo diferentes enfoques administrativos, voltados a distintas gerências, conforme vimos na seção 2.2.2.2 do capítulo 2.

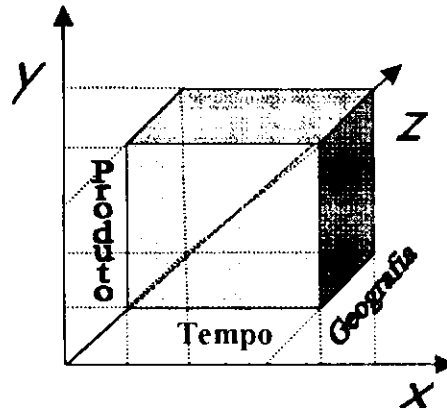


Figura 3.1 - Análise Multidimensional:  
O Cubo de Dados

Conforme [UHROWCZIC, 1997], “a AM não é um novo conceito ou tecnologia”, acrescentando que “há duas décadas atrás, alguns SGBD operacionais tinham sido usados como repositório de dados para dados agregados multidimensionais (tipicamente dimensões produto, geografia e tempo). Do mesmo modo, repositórios de dados especializados tanto como ferramentas para usuários finais para AM entraram no mercado logo depois”. [WHITE, 1997] afirma que a AM foi originada nos anos 70 por fornecedores que já dispunham de ferramentas de tempo compartilhado (“time sharing”) com esta

finalidade, tal com a APL da IBM e, posteriormente, pacotes do tipo planilhas. Com a complexidade das organizações e a necessidade de uma visão corporativa, estes instrumentos originais de análise multidimensional ficaram insuficientes por si só, requerendo outras tecnologias afins, como a de DW, de forma a integrar e fornecer consolidada e historicamente os dados para uso analítico organizacional.

Apesar dos BDR terem um cunho de suporte à operacionalização dos negócios, com o armazenamento confiável dos dados, [KIMBALL, 1996] esclarece que o objetivo dos BDR não é só obter dados, mas sim, fornecê-los também. Ele acrescenta que, para se bem realizar essa tarefa, referindo-se à AM, deve-se ignorar ou mesmo até empregar incorretamente algumas das facilidades principais dos BDR. Veremos melhor essa questão na seção 3.6, sobre “Modelagens para Repositórios OLAP”.

### 3.1.2. As Regras para OLAP

Relacionamos as primeiras e originais regras para OLAP definidas por Codd et al. apud [GILL & RAO, 1996] e comentamo-las a seguir:

1. **visão multidimensional conceitual** - porque a natural visão comercial do usuário empresarial requer AM, isto é, para as gerências é interessante ter diferentes e diversificados ângulos no negócio, segundo seus enfoques administrativos;
2. **transparência** - independência física com transparência lógica da ferramenta analítica para o usuário de negócios;
3. **acessibilidade** - acesso aos dados, pela ferramenta, de forma integrada e consistente, independentemente de onde e como os dados estão fisicamente;
4. **desempenho consistente nos relatórios** - o desempenho do sistema não deve ser degradado em função do aumento do número de dimensões envolvidas numa análise;
5. **arquitetura cliente/servidor** - facilidade de integração de novos usuários clientes, com uso de um servidor capaz de mapear e consolidar incompatíveis bancos de dados de forma **transparente** (conforme regra 2) criando um esquema conceitual único;

6. **dimensionalidade genérica** - padronização das dimensões quanto às suas estruturas e capacidades operacionais;
7. **capacidade de manipular dinamicamente matriz esparsa<sup>19</sup>** - otimização no armazenamento e trato dos dados, prevendo a ausência de valores em atributos de dados;
8. **suporte multiusuário** - permitir o acesso de vários usuários concorrentemente aos dados de forma garantida, quanto a integridade e segurança em operações de recuperação e atualização dos dados;
9. **operações cruzadas entre dimensões de forma irrestrita** - possibilidade de efetuar cálculos utilizando dados de diversas dimensões irrestritamente;
10. **manipulação intuitiva dos dados** - facilidade de navegar nos dados de forma direta, seguindo a própria intuição e vontade, de forma simples e direta, os quais devem atender plenamente às necessidades do negócio;
11. **flexibilidade nos relatórios** - possibilitar manipulações através de análises, sínteses, verificando os dados da forma que mais lhe convier: resumindo-os ou detalhando-os tranquilamente;
12. **níveis de agregações e dimensões ilimitadas** - o sistema deve acomodar pelo menos quinze dimensões dentro do modelo analítico, sendo que cada dimensão deve permitir o uso indeterminado de definições de agregações e níveis de resumos de uma linha de análise.

Além dessas doze regras, segundo Pense & Creeth, apud [WHITE, 1997], seis outras foram adicionadas em 1995 [PENSE & CREETH, 1995]. Contudo, a bibliografia por nós consultada não faz maiores referências a elas. Interessante notar também que as regras acima citadas e comentadas, assim como as seis outras que foram adicionadas posteriormente, tornaram-se apenas uma referência básica do assunto, pois, como veremos a seguir, houve inclusive uma influente segmentação comercial a esse respeito.

---

<sup>19</sup> A ocorrência de matriz esparsa, na AM, se dá pela intersecção vazia entre dimensões de um fato, por exemplo, as vendas (o fato) de janeiro (dimensão tempo) de guarda-chuvas (dimensão produto) na cidade do Salvador, estado da Bahia (dimensão geografia) praticamente inexistem, mas, já no mês de julho e sob as mesmas condições, muitos guarda-chuvas são vendidos.

### 3.1.3. Movimentação Comercial

Paralelamente aos 12 critérios de OLAP propostos por Codd et al. formou-se um consórcio de fornecedores auto-intitulando-se Fornecedores OLAP (“OLAP Vendors”) [KIMBALL, 1996]. A partir dessa movimentação dos fornecedores passamos a ter uma definição de OLAP orientada objetivamente, por eles, ao mundo dos negócios conforme veremos a logo a seguir, na seção 3.2.

De acordo com [UIROWCZIC, 1997], o interesse pela AM deve-se às emergentes tecnologia e ferramentas das estações trabalho (“workstations”) – que foram melhoradas<sup>20</sup> – e à disseminação e simpatia das planilhas para a grande maioria dos analistas de negócios. As ferramentas de usuário final variam das planilhas, cujo o uso comercial cresce a cada dia, até sistemas integrados de análise de dados. Pela sua constatação, os analistas de negócios usam mais planilhas, enquanto os executivos preferem as interfaces gráficas<sup>21</sup>.

## 3.2. Definição e Finalidade

Veremos agora o ponto de vista de alguns autores sob os seus entendimentos e o que eles consideram como finalidade do OLAP. Contudo, não deixamos de fora a sua ótica da movimentação comercial, ao contrário, iniciaremos esse tópico justamente com ela, por abordar de forma objetiva e bastante prática esse estudo. Uma definição sucinta de OLAP dada pelo grupo dos Fornecedores OLAP citada por [WHITE, 1997; INMON, 1997] é:

*“On-line Analytical Processing (OLAP) é uma categoria de tecnologia de software que capacita analistas, gerentes e executivos a obter a introspecção nos dados através de acesso rápido, consistente e interativo para uma larga variedade de possibilidades*

---

<sup>20</sup> Ultimamente a tecnologia na computação tem avançado nos aspectos velocidade no tempo de resposta, capacidade de armazenamento e usabilidade, tornando os computadores cada vez mais rápidos e fáceis de serem usados.

<sup>21</sup> Acreditamos que o autor quando menciona “interfaces gráficas”, esteja referindo-se a EIS/OLAP.

de visões da informação que fora transformada dos dados brutos para refletir a real dimensionalidade da empresa tal qual é entendida pelo usuário”.

[WHITE, 1997] cita uma definição mais simples (da revista DBMS de abril de 1995): “um nome fantasia para análise multidimensional”. [BRACKETT, 1996] refere-se a OLAP também como MDA (“Multiple Dimensional Analysis”), ou seja, AM. “MDA é a habilidade de manipular dados que foram agregados em várias categorias ou ‘dimensões’ ” [INMON, 1997].

[UHIROWCZIC, 1997] refere-se a AM, ou OLAP, como uma análise de agregações de dados do ponto de vista do negócio. [MENNINGER, 1997] define OLAP como a análise interativa dos dados que foram gravados ou dados futuros projetados. Neste ponto, vemos que [MENNINGER, 1997], em relação a [UHIROWCZIC, 1997], considera na AM inclusive o futuro, através do que ele chamou “dados futuros projetados”. Contudo, para [INMON, 1997], “mais sucintamente, OLAP é um conjunto de funcionalidades que tenta facilitar AM”, ou seja, é o tipo de processamento que viabiliza modernamente a AM que, como vimos, não é novidade.

O propósito da AM, ainda segundo o grupo dos Fornecedores OLAP, relatado por [INMON, 1997], é o de “(ajudar) o usuário (a) sintetizar as informações empresariais através de visões comparativas e personalizadas, tão bem quanto através da análise de dados históricos e projetados”. Portanto, precisamos descer em mais detalhes do que seja OLAP, visto que, essas definições se confundem. Para isso veremos o entendimento sobre o assunto, segundo os autores [GILL & RAO, 1996]. Para eles, OLAP é uma tecnologia de análise de dados que contempla os seguintes pontos:

- apresenta visão lógica multidimensional dos dados, independentemente de como eles estão armazenados;
  - é interativa nas consultas e análise dos dados. Essa interação se dá normalmente com múltiplos passos, envolvendo aprofundamento em sucessivos níveis menores de detalhes de dados ou níveis maiores de resumos e agregações;
-



- oferece capacidade de modelagem analítica, incluindo máquina de cálculos para taxas de derivação, etc., envolvendo medidas ou dados numéricos através de qualquer das dimensões;
- possibilita a criação de resumos, agregações<sup>22</sup>, hierarquias, e interroga todos os níveis de agregação e resumos em cada interseção dimensional;
- suporta modelos funcionais para previsões, tendências e análise estatísticas;
- recupera e exibe os dados em duas ou três dimensões em tabulação encadeada, tabelas e gráficos, com a fácil rotação dos eixos;
- responde às consultas rapidamente de forma que o processo de análise não seja interrompido e as informações não fiquem antiquadas;
- possui uma máquina multidimensional de armazenamento de dados, utilizando matrizes<sup>23</sup>, sendo elas a representação lógica das dimensões do negócio.

Os autores acrescentam que a tecnologia OLAP pode ser usada em várias áreas funcionais de negócios, tais como produto, vendas, análise de rentabilidade de marketing, etc.

[TANLER & DROST, 1996] listam quinze itens que consideram pré-requisitos mínimos, ditos “chaves” para uma lista de conferência na seleção de uma ferramenta para análise multidimensional. Em confronto com os oito pontos acima definidos, temos os seguintes itens adicionais:

- ◆ filtros - para uso de forma a restringir consultas, estes podem ser criados, salvos e compartilhados a partir de operações de arrasto de mouse e serem combinados entre as dimensões;
- ◆ capacidade, por software, de trabalho a nível de grupos (*groupware*) - compartilhamento de decisão por grupos de trabalho, ou mesmo a nível empresarial, de análises e seus resultados de forma segura e imediatamente atualizada, ou seja,

---

<sup>22</sup>Os autores [GILL & RAO, 1996] fazem uma ressalva que *agregações* algumas vezes são livremente referidas como *consolidações*.

<sup>23</sup> Neste trabalho utilizamos a terminologia *matriz* como tradução do inglês “array” (Camarão, P. - *Glossario de Informática*, Rio de Janeiro: LTC, 1994).

disponibilizar informação ao vivo, (*live information*) para a corporação. Isso feito sempre de forma segura para a corporação;

- ◆ poder criar, salvar e compartilhar cálculos simples e complexos, e obter suas respostas sempre dinamicamente, sem precisar ter que fazer rotinas (*stored procedures*) especialmente para cada novo tipo de consulta;
- ◆ conceito de tempo flexível - suporte a múltiplos períodos de tempo descontínuos, assim como seus derivados, tais como cálculos e variáveis do tipo “mais recente”, etc. Isso sempre sem armazená-los;
- ◆ metadados<sup>24</sup> - traduzindo redução no tempo das requisições de desenvolvimento e suporte ao usuário, além de administrar transparentemente as regras do negócio, facilita a navegação nos dados, pelo provimento de uma interface baseada em dados sobre os dados das dimensões;
- ◆ sistema aberto - independentemente do tipo de padrão do sistema ser local ou distribuído, deve existir acesso direto aos dados através dos SGBDR de forma padronizada e direta, sem necessitar de extrair ou armazenar este dados em ambiente proprietário multidimensional. Qualquer desenvolvimento feito em qualquer micro computador do tipo IBM PC (PC) ou equivalente deve poder efetuar AM;
- ◆ geração de SQL (linguagem padrão de acesso aos SGBDR ) automática - isso de forma eficiente, em tempo de execução e de acordo com as solicitações dos usuário e aplicação;
- ◆ esquema de suporte à matriz esparsa – por ser esse uma aspecto inerente à AM (pois interseções entre dimensões eventualmente são inexistentes), é relevante se ter em mente esse ponto, de forma a se evitar o registro de dados inexistentes, o que pode causar superlotação do repositório;
- ◆ leitura e gravação no repositório – a atividade gerencial de suporte à decisão necessita de fazer previsões, orçamentos, etc., o que sugere que o BD aceite além de leitura, também gravação.

Percebemos uma preocupação dos autores com alguns aspectos extras em relação aos anteriores [GILL & RAO, 1996]. Esses aspectos são: a filosofia *groupware* introduzida

---

<sup>24</sup>Metadados serão detalhadamente abordados sob a ótica do DW na seção 4.3.6 do capítulo 4.

oportunamente neste contexto, aspectos de implementação, tais como a geração automática de SQL e uso de sistema aberto. Contudo, vale uma ressalva ao ponto: leitura e gravação no repositório. Concordamos com os autores quanto às necessidades apontadas referente às atividades gerenciais de suporte à decisão, mas questionamos a vinculação delas à ferramenta de AM, como parte integrante dela.

Conforme acompanhamos os diversos pontos de vista sobre o que seria a definição e finalidade do OLAP, vemos, sob os seus diversos ângulos, que, segundo os valores e enfoques de cada um, temos diferentes abordagens. Contudo, elas não se contradizem, ao contrário, se somam, completando-as umas às outras. Uns abstraem-se mais que outros. Alguns ficam mais vinculados aos aspectos históricos e comerciais desse estudo, outros enfocam-no sob o prisma da arquitetura aberta de sistemas. Abordaremos a seguir a realidade operacional dessa tecnologia, mostrando as operações caracterizadas como Operações OLAP.

### 3.3. Operações

Além da facilidade em obter dados consolidados por dimensão, conforme [MATTISON, 1996], são esperadas as operações de *drill down / up / across* (análise prospectiva) de qualquer dimensão, *pivot* (redirecionar linha/coluna de) um relatório e *slice and dice* (trocar os critérios de restrição de qualquer uma das dimensões), mantendo a integridade dos valores pré-computados e seus subtotais envolvidos. Detalhando melhor:

- *Análise Prospectiva* – esta se subdivide em três modalidades ou operadores:
  - *drill-down* - análise prospectiva, caminhando para maior detalhamento de dados numa dimensão, ou seja, buscando particularidades;
  - *drill-up* - também conhecido como *roll-up* – análise prospectiva, caminhando para menor detalhamento de dados numa dimensão, ou seja, buscando uma visão macro;
  - *drill-across* - também conhecido como *cross-dimension* – análise prospectiva, caminhando entre dimensões, ou seja, a partir de uma dimensão mudar o enfoque analítico para outra dimensão;

- *Análise Seletiva-Restritiva* – subdividindo-se em dois modalidades ou operadores:
  - *slice* - (cortar, em português) seleção transversal dimensional, ou seja, análise dimensional em função da seleção e fixação uma dimensão;
  - *dice* - (fatiar, em português) estabelecer critérios restritivos na seleção transversal dimensional (*slice*), ou seja, restringir o foco analítico do que fora selecionado;
- *pivot* ou *pivoting* - redirecionamento (o rotacionamento) no sentido de linha/coluna em um relatório.

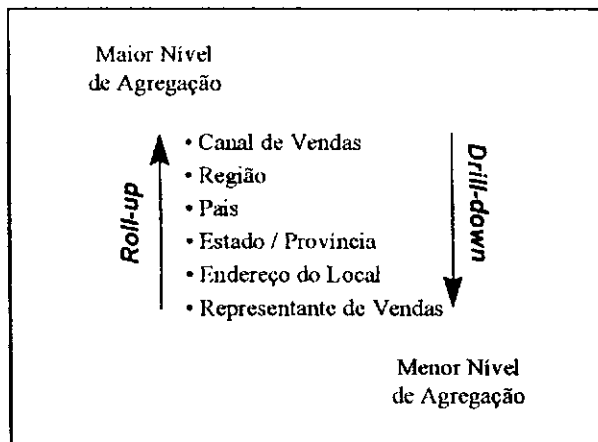


Figura 3.2 - Exemplo de Hierarquia de Canal de Vendas [INMON, 1997]

Na Figura 3.2 [INMON, 1997], vemos uma abordagem gráfica das operações de *drill-down* e de *drill-up*. Essa última, também conhecida como *roll-up* [GILL & RAO, 1996; INMON, 1997] – como mostra nesta figura – indica a saída de uma menor granularidade de dados (nível de detalhe), em busca de uma maior. No caso, saímos do representante de venda – que é a menor unidade

de granularidade nesta dimensão, neste caso hipotético, – e caminhamos em direção à maior granularidade, ou seja, a nível de canal de vendas. A essa operação de agregação denominamos *drill-up*, ou ainda *roll-up*. Já a operação de *drill-down* se caracteriza pelo caminho oposto (a busca do nível de maior detalhe), ou seja, a saída de uma maior agregação em direção a uma menor agregação. A granularidade em questão será diminuída no “drilling-down”. É conveniente atentar para o fato de que, em função da conveniência da análise em questão, poderemos opcionalmente estacionar num nível intermediário qualquer, durante a execução de uma dessas duas operações: *drill-up/down*.

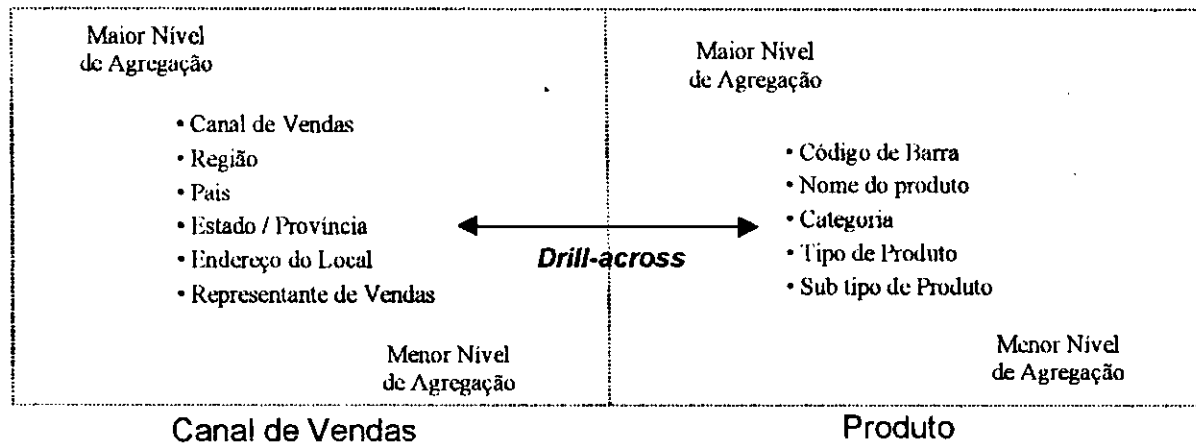


Figura 3.3 - Exemplo de Operação de *Drill-across*

Já na Figura 3.3, vemos um exemplo complementar das operações de análise prospectiva vistas na Figura 3.2. Enquanto as operações de *drill-up/down* são efetuadas sobre uma dimensão, *drill-across* opera sobre mais de uma dimensão, traduzindo exatamente a necessidade do usuário de mudar sua dimensão em foco. Nesta figura temos as dimensões Canal de Vendas e Produto. Poderíamos supor que, a partir do canal de vendas (atributo Estado/província da dimensão Canal de Vendas), desejássemos verificar os tipos de produtos (atributo Tipos de Produto, dimensão Produto) envolvidos comercialmente. Para isso, usando a operação de *drill-across*, partindo da dimensão Canal de Vendas em direção à dimensão Produto, poderemos continuar a análise prospectiva detalhando mais ou menos os dados, segundo diferentes níveis de agregação, conforme vimos, com os operadores de *drill-up/down*. Sempre que se deseja fazer uma análise prospectiva envolvendo uma mudança de dimensão, usa-se a operação de *drill-across*.

Na análise seletivo-restritiva, as operações de *slice & dice* são responsáveis pela seleção (*slice*) do foco analítico de interesse administrativo, ao mesmo tempo que pode restringir (*dice*) o âmbito angular da análise. Vimos anteriormente, na Figura 2.4, seção 2.2.2.2 do capítulo 2, um cubo visto sob três diferentes enfoques administrativos: gerências regional, financeira e de produto. Vemos na Figura 3.4 um cubo sob o enfoque da gerência regional. Para esta gerência, na identificação do potencial de consumo dos mercados, é interessante a *seleção* de cada Mercado, individualmente, para visualização dos produtos

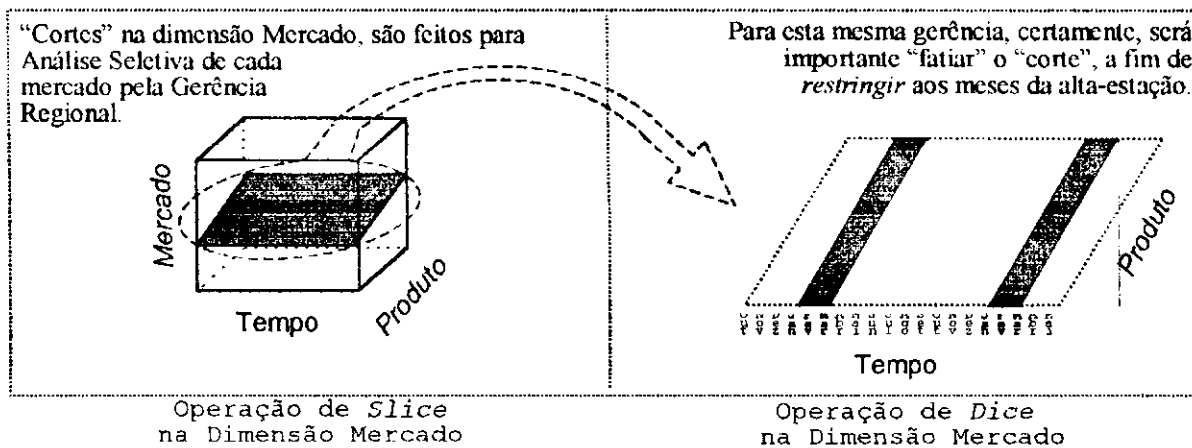


Figura 3.4 - Exemplo de Análise Seleto-Restritiva

ao longo do tempo. Porém, há períodos de tempo em que o comportamento do mercado sofre substanciais alterações. Por exemplo na época da alta-estação, compreendida pelos meses de Janeiro, Fevereiro e Março, nota-se, normalmente, um aumento generalizado de consumo de produtos. Um gerente experiente busca logo o confronto entre os produtos nesse período de tempo, na identificação do comportamento do mercado. A essa operação de *restrição*, na análise da *seleção* feita anteriormente, dá-se o nome de *dice*.

Produto	Região	Período	
		Este mês	Mês Anterior
Framis	Central	110	12
Framis	Leste	179	-3
Framis	Oeste	55	5
Widget	Central	66	2
Widget	Leste	102	4
Widget	Oeste	39	-9

Figura 3.5 - Exemplo de um Relatório Gerado por um SGBDR [KIMBALL, 1996]

Para finalizar as operações de OLAP, *pivot* ou *pivoting* (rotacionamento, redirecionamento), é definida como sendo a reorganização de um relatório qualquer elaborado sob forma de tabela, pela inversão arbitrária da posição dos cabeçalhos das linhas e colunas. Segundo [KIMBALL, 1996], esse

Período	Produto	Região		
		Central	Leste	Oeste
Este mês	Framis	110	179	55
Este mês	Widget	66	102	39
Mês anterior	Framis	12	-3	5
Mês anterior	Widget	2	4	-9

Figura 3.6 - Exemplo<sup>26</sup> do Operação de *Pivot* do Relatório da Figura 3.5

*Produto, Região e Período* e classificado por *Produto* e *Região*. Segundo [KIMBALL, 1998], em função do “SQL retornar os dados das consultas de forma muito peculiar e definida”, “todos os itens de uma determinada coluna possuem a mesma definição e os dados subjacentes são agregados unicamente por combinações de cabeçalhos de linha”. A Figura 3.6 traduz a operação de *pivot* na tabela da Figura 3.5, que fica agora classificada por *Período* e *Produto*.

### 3.4. Relatórios

A Figura 3.7 nos mostra um relatório multidimensional, no qual identificamos as dimensões produto (representada pelos produtos denominados Framis e Widget), região (selecionadas as áreas central, leste e oeste) e tempo (mês anterior, mês atual, ano anterior e ano atual). Os dados são consolidados, resumidos e totalizados por dimensão na sua forma absoluta e relativa, ou seja, expressos também sob forma de porcentagem. Neste relatório identificamos que alguns dados são impressos especialmente, de forma a destacá-los de acordo com a sua representatividade atual, ou seja, identificados em função da expressividade dos seus valores relativos momentâneos para o negócio.

<sup>25</sup> Provavelmente a inspiração de Kimball tenha sido no método *pivot* [RAO, 1952], para estimativa de determinantes, que, entre outros propósitos, é usada para se encontrar matriz inversa.

<sup>26</sup> O relatório extraído de [KIMBALL, 1996] foi alterado aqui, pois o original contém uma imprecisão no cabeçalho.

termo, foi criado para designar esta operação nos SGBDR no início da década de 1980 por ele<sup>25</sup>.

A Figura 3.5 nos mostra um exemplo de uma relatório do tipo tabela contendo os atributos

Produto	Região	Vendas no mês	Crescimento nas vendas X mês anterior	Vendas como % da categoria	Mudança nas vendas como % da categoria X mês anterior	Mudança nas vendas como % da categoria X ano anterior
Framis	Central	110	12%	31%	3%	7%
Framis	Leste	179	<b>(3%)</b>	28%	<b>(1%)</b>	3%
Framis	Oeste	55	5%	44%	1%	5%
<i>Total Framis</i>		344	6%	33%	1%	5%
Widget	Central	66	2%	18%	2%	10%
Widget	Leste	102	4%	12%	5%	13%
Widget	Oeste	39	<b>(9%)</b>	9%	<b>(1%)</b>	8%
<i>Total Widget</i>		207	1%	13%	4%	11%
<i>Total Geral</i>		551	4%	20%	2%	8%

Figura 3.7 - O Relatório Ideal de Data Warehouse [KIMBALL, 1998]

Kimball considera um relatório deste tipo como sendo um tipo de ideal de relatório de DW<sup>27</sup>. Vale ressaltar que, no momento em que ele destaca (com parênteses e em negrito) alguns dados relevantes ao negócio neste relatório, independentemente da razão, ele está extrapolando o conceito de EIS/OLAP, que é baseado puramente na AM, e parte para DSS/OLSP. Os dados destacados no relatório da Figura 3.7 tentam transmitir informações adicionais à sua análise. Para identificá-los, ele provavelmente utilizou-se de alguma forma de *Mining*, mesmo ainda que primitiva. Portanto, esse relatório, excluindo-se os destaques dados aos valores, é um modelo de relatório OLAP.

Apesar da idéia de OLAP poder ser operacionalizada através de relatórios, acreditamos que o uso das interfaces gráficas seja a forma ideal de interação no processamento analítico. Devido ao seu caráter de abstração, o uso dos conceitos de OO (Orientação a Objetos), provavelmente será fluentemente empregada no OLAP. Com os

<sup>27</sup> O referido autor não separa as diferentes funções inerentes aos repositórios de OLAP e DW. Para ele o repositório OLAP tem funções DW, constituindo-se um único repositório. Abordaremos detalhadamente essa questão, entre outras, sobre DW, no capítulo 4.



conceitos de OO aplicados ao OLAP nas interfaces gráficas, a “navegação intuitiva nas dimensões” (10ª regra OLAP) se efetua mais facilmente e, talvez, de forma mais dinâmica. Isso porque, um objeto poder ser visto como um exemplar ou uma instância de um conceito, construído talvez pela organização de outros objetos (por exemplo a dimensão Produto, composta por Código de Barra, Nome do Produto, Categoria, Tipo do Produto, etc.), mas sempre caracterizado por implementar um aspecto funcional coerente. Outros aspectos são: classificação, que hierarquizada facilita as operações *drill-up/down*, de análise prospectivas, nos atributos, vistos como objetos, e herança, na “dimensionalidade genérica” (6ª regra OLAP), ações comuns a ambas as classes de objetos poderão ser realizadas de formas diferentes segundo as características de cada dimensão.

A OO poderá facilitar ainda mais a elucidação das idéias intuitivas dos *knowledge workers* a respeito dos dados e flexibilizar sobremaneira o trabalho deles. Contudo, conforme [MENNINGER, 1997], muitas das ferramentas existentes hoje no mercado ainda são projetadas para sistemas orientados a registro e oferecem pouco ou nenhum suporte para AM. Porém, uma outra alternativa é um ambiente de aplicação especificamente projetado para construção de aplicações OLAP. Veremos na seção 3.7 alguns tipos de arquiteturas, mas antes, precisamos nos inteirar das modelagens existentes e utilizadas nas suas construções.

### 3.5. Repositório OLAP

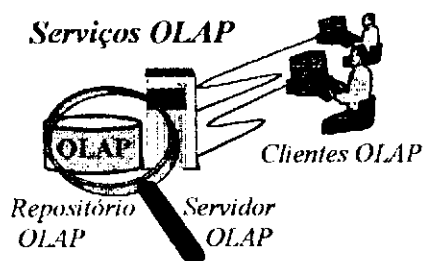


Figura 3.8 - Serviços OLAP:  
Componentes Básicos

A Figura 3.8 nos mostra os componentes básicos necessários ao OLAP. Nela vemos clientes OLAP (serão detalhados na seção 3.7, sobre Arquiteturas OLAP) um servidor OLAP, responsável pela realização das já vistas operações OLAP, e um repositório OLAP destacado pela lupa. O repositório OLAP difere de um BD operacional qualquer principalmente quanto à sua função. Enquanto os BD operacionais são organizados com o objetivo de oferecer respostas imediatas a

solicitações de armazenamento e recuperação dos dados, os Repositórios OLAP são organizados com vista à facilitação das consultas do usuário na forma de navegação nos dados na ótica de dimensões.

	<b>BD Operacional</b>	<b>Repositório OLAP</b>
<b>Prioridades do repositório</b>	1º plano a atualização 2º plano consultas	1º plano consultas 2º plano atualização
<b>Atualizações dos dados</b>	Contínuas, aleatórias	Planejadas, periódicas
<b>Granularidade dos dados</b>	Primitiva e detalhada	Dados resumidos e agregados, podendo também estar dados primitivos
<b>Instância dos dados</b>	Tempo simples	Variante no tempo
<b>Estrutura dos dados</b>	Nomalizada	Multidimensional
<b>Enfoque das consultas</b>	Simple e previsíveis, predefinidas pelo usuário	Sofisticadas e aleatórias, controladas pela interação com o usuário
<b>Flexibilidade das consultas</b>	Baixa, repetitiva	Alta, heurística
<b>Enfoque da modelagem</b>	Armazenamento consistente	Recuperação eficiente
<b>Controle de concorrência</b>	Efetivo	Inexistente
<b>Fluência no armazenamento</b>	Corrente	Histórico
<b>Volatilidade do armazenamento</b>	Dinâmica	Estática

Figura 3.9 - Diferenças entre BD Operacional e Repositório OLAP

Na Figura 3.9. vemos um quadro comparativo das diferenças entre BD operacional e Repositório OLAP segundo nove parâmetros importantes. As diferentes características encontradas nelas, refletem as particularidades de cada tipo de repositório. Os parâmetros utilizados foram:

- **prioridade do repositório** – esclarece o foco prioritário do repositório, ou seja, nos BD operacionais é básico que os dados estejam atualizados para que se efetue uma consulta qualquer, enquanto que para o outro o mais importante é que os dados estejam disponíveis para consultas especiais;
- **atualizações dos dados** - exprime o critério de atualizações dos dados;
- **granularidade dos dados** - exprime como os dados serão armazenados para consultas a nível de detalhes;
- **instância dos dados** – indica a influência do tempo ligada a uma dada consulta;

- **estrutura dos dados** - exprime a forma interna de organização dos atributos. Por exemplo, um dos gerentes da organização deseja analisar o *cube* de uma forma diferente da usual e para isso precisa de fazer uma agregação e ou inclusão/exclusão de um determinado atributo;
- **enfoque das consultas** - indica a qualidade de uma consulta, quanto à complexidade, a previsibilidade e maneira de interação com o usuário ;
- **flexibilidade das consultas** - expressa a qualidade de uma consulta, quanto ao seu grau de elaboração. Por exemplo, uso de múltiplas dimensões para atender aspectos heurísticos da base de dados;
- **enfoque da modelagem** - indica o objetivo da modelagem dos dados no repositório;
- **controle de concorrência** - indica a existência de um aspecto fundamental nos BD operacionais, contudo, inexistente nos Repositórios OLAP, que, por serem voltados à consultas multidimensionais, dispensam tal controle;
- **fluência no armazenamento** - indica a gama de tempo utilizada no armazenamento dos dados;
- **volatilidade no armazenamento** - indica se os dados serão ou não atualizados durante a execução de uma consulta.

Os BD operacionais têm como missão refletir, o mais rápido possível, a realidade do contexto a que está atrelado para, em algum momento, ser consultado. Já os repositórios OLAP não. Nos parâmetros **prioridade do repositório**, **enfoque das consultas**, **enfoque da modelagem**, e **controle de concorrência** residem os pontos principais de inadequação entre a abordagem do BD operacional como Repositório OLAP. Apesar de estarem sendo empregados na prática, conforme podemos constatar, devido ao que atribuímos pela “a inércia tecnológica”, advertimos: utilizar um BD operacional como Repositório OLAP não é o mais indicado, devido aos mecanismos que garantem a sua consistência e integridade acarretarem aumento do tempo de resposta das consultas, impactando no desempenho dos serviços OLAP.

Outra forma de alguns referirem-se ao Repositório OLAP é por BD Multidimensional (BDM) ou até por DW!<sup>28</sup> Originalmente, os BDM têm a sua modelagem proprietária, baseada em índices e matrizes multidimensionais, especialmente desenhada para atingir os objetivos de um Repositório OLAP. Contudo, com o progresso da pesquisa, foi possível criar uma forma simulada de modelar os BDM em BDR. Veremos as modelagens existentes dos Repositórios OLAP na seção seguinte.

### 3.6. Modelagens para Repositórios OLAP

Nesta seção veremos os dois tipos mais comuns de modelagens de dados para implementação lógica de Repositórios OLAP. São elas, a modelagem de simulação multidimensional para os DBR, conhecida por ROLAP (*Relational OLAP*), e a modelagem efetivamente multidimensional, chamada de MOLAP (*Multidimensional OLAP*). Estas duas formas apresentam substanciais diferenças que são discutidas, ao final, como vantagens e desvantagens, e mostradas a partir de uma análise comparativa entre essas duas abordagens, da qual concluiremos alguns pontos interessantes. Inicialmente, discutimos a abordagem relacional.

#### 3.6.1. OLAP Multidimensional Relacional

Ao vermos a expressão ROLAP - “Relational OLAP”, inicialmente, é natural a pressuposição lógica de uma base de dados normalizada, conforme recomenda-se para os BDR, contudo veremos que, na verdade, nesta modelagem os dados internamente não estarão rigorosamente desta maneira. Para atingirmos o nível de processamento analítico utilizando um BDR temos que “relaxar um pouco” a idéia de normalização [DATE, 1991] para todas as tabelas e elaborar uma nova estrutura que possibilitasse a implantação a nível

---

<sup>28</sup> Trata-se do DDW (*Dimensional Data Warehouse*), um modelo de DW que assume as funções de Repositório OLAP. Veremos maiores detalhes dele na seção 4.5.2.2.

lógico do conceito de multidimensão. Desta forma consegue-se simular um BDM utilizando-se linhas e colunas de um BDR.

Pioneiramente, Ralph Kimball, conforme [INMON ET AL., 1996], estruturou uma modelagem multidimensional para BDR baseada na aglutinação da maioria dos dados relacionados intrinsecamente a um fato comercial em uma determinada tabela, chamada de *fact table* (em português, tabela de fato), sendo esta cercada por outras tabelas menores de

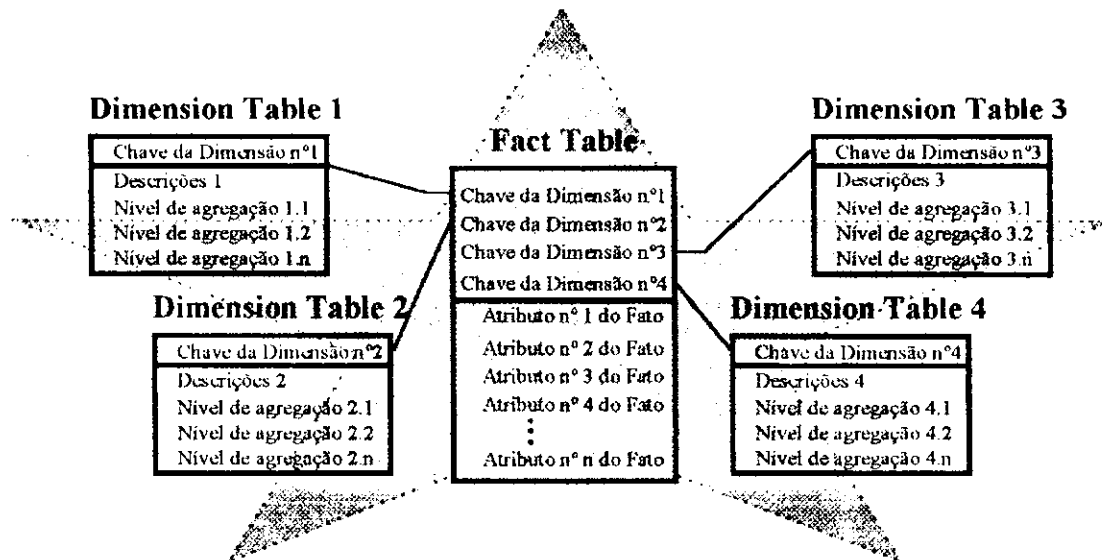


Figura 3.10 - Descrição de um *Star Schema*

apoio denominadas de *dimension tables* (em português, tabelas de dimensões). A Figura 3.10 nos mostra a disposição básica desta estrutura, denominada *Star (Join) Schema* – Esquema (de junção) Estrela. Podemos constatar que a denominação “estrela” é devido à semelhança com um polígono da disposição da tabela central (*fact table*) rodeada por outras secundárias (*dimension tables*). No centro do *star schema* está sempre uma (ou mais, como veremos a seguir) tabela de fato, a qual representa quantidades de dados que podem ser agregados sem perder seus significados e que precisam ser descritos por mais que duas (tabelas de) dimensões, além do tempo.

Dissecando as tabelas da Figura 3.10, vemos na tabela primária, a *fact table*, chaves de dimensões (casualmente 4 e numeradas de 1 a 4), que descrevem apontadores (chaves)

para as dimensões, e atributos relacionados com os fatos. Conforme vimos anteriormente, estes atributos são sempre quantitativos. São valores aditivos. Por exemplo, preço, quantidade, etc. Já nas *dimension tables*, vemos o respectivo apontador da específica dimensão, descrições<sup>29</sup> de atributos concernentes a essa dimensão em nível de detalhe e os seus níveis de agregação envolvidos. Por exemplo, suponhamos uma dimensão espacial geográfica qualquer. Podemos ter como descrição o número do local, o logradouro e agregações tais como bairro, cidade, estado, país, etc. Observando sempre que tratam-se de atributos qualitativos.

De acordo com [INMON, 1997], o *star schema* também é chamado tanto por *tall table* (em português, grande tabela) como por *modelo dimensional do BDR*. Acreditamos que a denominação de *tall table* derive do tamanho da *fact table*, que é a tabela central da estrutura, em relação ao seu tamanho consideravelmente maior que as demais. A Figura 3.11 exemplifica um *star schema* referente a vendas em supermercado. Nela encontramos no centro, como *fact table*, o *fato vendas*, com seus atributos quantitativos, *valor de vendas*, *itens vendidos* e *quantidades de compradores*. Por efeito de simplificação, colocamos apenas quatro dimensões: tempo, loja, ponto de venda (PDV) e produto, com os vários atributos qualitativos do *fato vendas* e agregações seletas por cada dimensão. Dessa forma, poderemos levantar facilmente informações sobre *a venda de um produto em um supermercado*, segundo as facilidades fornecidas através da prática das operações especiais já vistas na seção 3.3. tais como visualizar, por exemplo:

- ◆ produto mais vendido durante a alta-estação, por estados;
- ◆ loja mais dinâmica nas vendas dos produtos, durante a alta-estação;
- ◆ funcionários mais eficientes nas vendas, por loja, durante a alta-estação em função de cada produto;
- ◆ etc..

---

<sup>29</sup> [INMON, 1996] adverte que a descrições só devem ser usadas, se e somente se, forem constantemente utilizadas nas consultas.

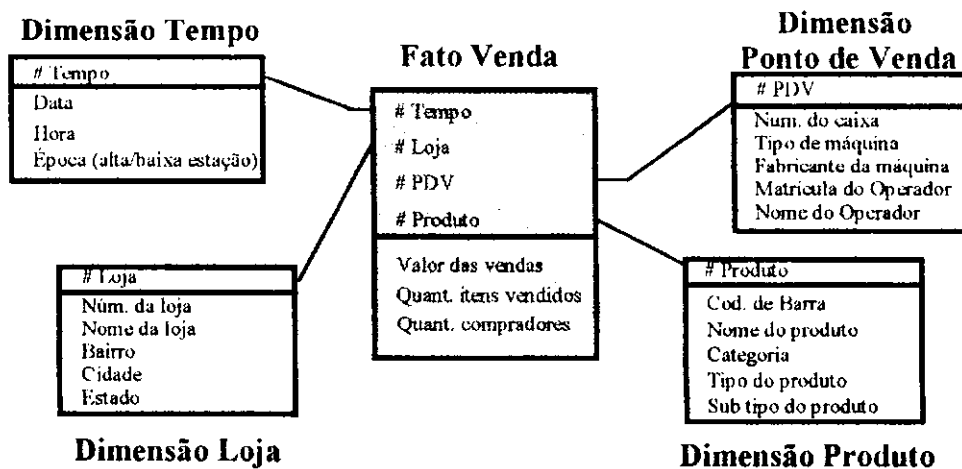


Figura 3.11 - Exemplo de *Star Schema* para Vendas em Supermercado

Utilizaremos ainda o exemplo da Figura 3.11 para elucidar algumas observações feitas por [INMON, 1997] sobre a *fact table* e as *dimension tables*. Na *fact table* encontramos:

- a **quantificação dos dados descritos pelas *dimension tables*** – no nosso exemplo as dimensões são *tempo*, *loja*, *PDV* e *produto* e na *fact table* estão quantificados os dados que apontam para elas por estes critérios de agregação;
- as **chaves formadas pela combinação das chaves das dimensões além de um atributo de tempo (a data, por exemplo) quando não há efetivamente a dimensão tempo** – isto é, uma atributo quantitativo de tempo com semântica de data (ou a própria data), tem que compor as chaves externas das dimensões da tabela de fato (a fim de efetuar-se a referência cruzada com a chave primária de uma dimensão em foco), a não ser que exista a dimensão tempo explícita, como é o caso do nosso exemplo, onde, na composição das chaves externas participa um ponteiro para a dimensão tempo;
- os **valores dos fatos devem ser aditivos<sup>30</sup>**, isto é, agregações do nível atômico ou quantidades de níveis atômicos. Não deve haver nem percentagens nem

<sup>30</sup>[KIMBALL, 1996] divide os fatos em *aditivos*, *semi-aditivos* e *não-aditivos*. A possibilidade de adição tem aqui o sentido de agregação, ou seja, um atributo do tipo aditivo é um atributo capaz de ser agregado a outros. São exemplos destes tipos de fatos, respectivamente, *região* (agregação de estados), *saldos*

**taxas** – no nosso exemplo *valor das vendas, quantidade de itens vendidos e quantidade de compradores*.

Já nas *dimension tables*, encontramos:

- **descrições dos dados organizados na *fact table*** – de um modo geral, considerado a dimensão *loja*, temos: número da loja, nome da loja, bairro, cidade e estado em que a loja se encontra;
- **a chave obrigatoriamente tendo valor único** – para nós, ainda sobre a dimensão *loja*, o campo número da loja (# loja) é uma chave primária, conforme o requisito;
- **a chave com o mais detalhado nível de agregação possível** – na *dimensão loja* temos o número da loja, que, além de ser único, identifica singularmente esta dimensão;
- **vários e manuseáveis níveis de agregação** – no nosso caso, indo do menor para o maior nível de agregação da *dimensão loja*, temos os campos *bairro, cidade e estado* da loja;
- **a possibilidade de dimensões que não tem hierarquia, apenas combinações de campos** – neste caso, no nosso exemplo a dimensão *ponto de venda* melhor se presta para elucidar esta observação. Nela encontramos campos como *número do caixa, tipo de máquina, fabricante da máquina, matrícula do operador e nome do operador*, muitos independentes entre si, do ponto de vista hierárquico, porém agregados nesta dimensão e sendo descritores (conforme visto na Figura 3.10).

---

(agregação parcial de lançamentos, segundo a sua natureza débito ou crédito) e *percentagem* (não há agregação, pois é uma medida relativa, não podendo ser agregada).



Conforme [INMON, 1997], o desenho lógico no ROLAP é uma combinação de dados propriamente normalizados e um ou mais *star schema*. Logo, é um erro imaginar-se que ROLAP exige apenas uma *fact table* central rodeada de várias *dimension tables*. Isso pode ser visto no exemplo da Figura 3.12, a qual contém duas tabelas de fatos: *Vendas*, que

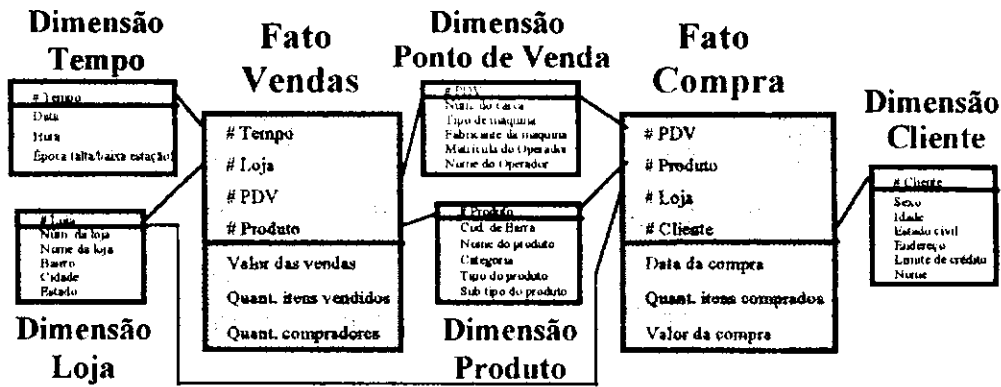


Figura 3.12 - *Star Schema* com Mais de Um Fato

se referem aos produtos vendidos pelos PDV nas lojas, e *Compra*, referente às compras de um determinado cliente, de um produto em um PDV de uma loja.

[KIMBALL, 1996] afirma que “a *fact table* armazena medições numéricas do negócio” e que “as *dimension tables* armazenam as descrições textuais das dimensões do negócio”. Ele acrescenta que a *fact table* é quem praticamente define o tamanho do repositório ROLAP (que ele denomina de *Dimensional Data Warehouse*), sendo elas geometricamente maiores que as *dimension tables*. Ele ainda adverte que tentar normalizar essa estrutura (as *dimension tables*, já que as *fact tables* são originalmente normalizadas) para se ganhar espaço em meio magnético (disco), além de ser “uma perda de tempo”, compromete a capacidade de pesquisa. Ainda de acordo com ele, “o espaço livre em disco obtido por meio da normalização de *dimension tables* representa normalmente menos que um por cento do espaço total necessário para a estrutura como um todo”.

A estrutura totalmente normalizada do *star schema* é conhecida como *snowflake schema*<sup>31</sup> (em português *Esquema focos de Neve*, como sugere o fato de sairmos

<sup>31</sup>Na versão portuguesa [KIMBALL, 1998], foi feita a adaptação para estrutura desmembrada.

normalizando tabelas de dimensão e criando outras tantas, como “flocos de neve caem em uma tempestade de neve”). Este tipo de estrutura é preferivelmente executada por desenvolvedores experientes acostumados a normalizar tudo em prol de um desempenho eficiente na manutenção dos BDR. Se por um lado melhoramos a eficiência da manutenção dos BDR com a maior normalização, dificultamos o seu uso na recuperação dos dados a nível dimensional.

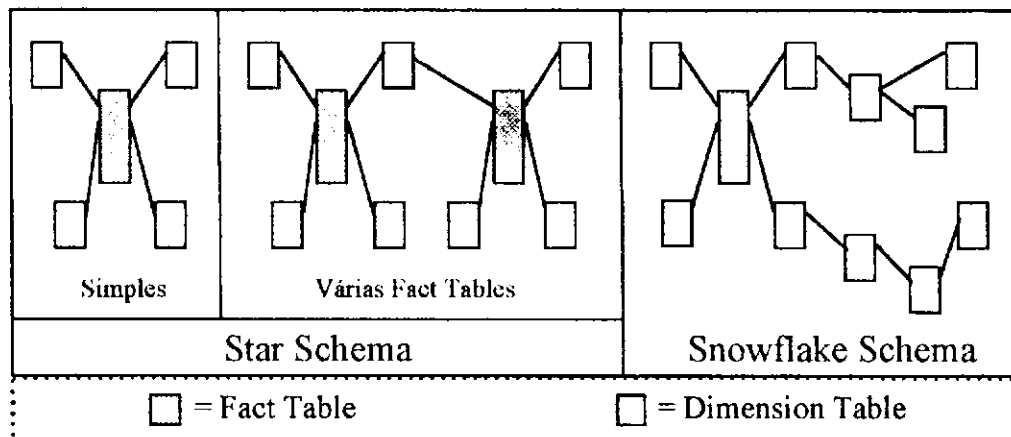


Figura 3.13 - Modelagens ROLAP

A Figura 3.13 condensa as modelagens vistas, ou seja, *star schema* simples, *star schema* com várias tabelas de fatos e *snowflake schema*. É interessante lembrar que todas essas modelagens têm como objetivo facilitar a recuperação dos dados, de forma a intuir (graficamente) ao usuário informações a respeito da organização que ora se mapeia com uma das estruturas escolhida e empregada. A vantagem do alto desempenho e facilidade armazenamento com integridade dos BDR normalizados não são traduzidas na mesma proporção no momento de recuperação dos dados a fim de sugerir uma visualização organizacional efetiva, ou seja, eles não se prestam à recuperação nesta mesma proporção de facilidade de se prestam no armazenamento de dados. Para acelerar o desempenho na recuperação, é costume *materializar-se*<sup>32</sup> as possíveis e prováveis visões dos usuários, especialmente nos dados agregados e resumos mais comuns. Mas, a popularidade do

<sup>32</sup>Do inglês “materialized view”, a terminologia “materializar” vem sendo cada vez mais usada pela academia para descrever o ato de gravar em meio magnético, geralmente disco, as visões mais referenciadas em consultas aos BD do tipo Repositório OLAP.

ROLAP tem crescido muito devido, basicamente, à possibilidade do uso da linguagem SQL padrão nas consultas. Já existem até extensões dela especialmente definidas para o processamento analítico que facilitam sobremaneira a atividade do programador / usuário.

Há também quem reconheça que uma boa solução pode estar na associação dos modelos *star schema* e o *snowflake schema*, selecionando apenas as maiores *dimension tables* para serem normalizadas, pois estas tabelas possuem geralmente dados das colunas altamente desnormalizados, isto é, duplicados [GILL & RAO, 1996].

### 3.6.2. OLAP Multidimensional Pura

A outra forma de modelarmos estruturalmente um repositório de dados voltado para consultas OLAP é conhecida como *multidimensional*, *multidimensional pura*, MOLAP - "Multidimensional OLAP", MDOLAP - "Multi Dimensional OLAP" ou ainda MDDB - "Multidimensional Database" (em português, BDM, para essa última sigla). Conforme [WHITE, 1997], "os dados nos Sistemas Gerenciadores de BDM (SGBDM) são armazenados em matrizes multidimensionais, onde cada célula na matriz é formada pela interseção de todas as dimensões".

Sem um padrão definido tal qual os BDR, temos notícia de dois tipos de estruturas de BDM. Conforme [GILL & RAO, 1996], os dados são logicamente armazenados em matrizes em um dos tipos a seguir: *multicube* ou *hypercube* (respectivamente em português, *multicubo* e *hipercubo*).

[WHITE, 1997] coloca que os dados são logicamente "quebrados" em *cubos de dados* em uma estrutura chamada de *multicube* com a finalidade de reduzir o efeito dos espaços vazios que ocorre na prática nas matrizes (esparsas) de dados para MOLAP. Contudo, ele explica ainda que alguns produtos não permitem esse tipo de armazenamento e requer que todos os dados sejam obrigatoriamente armazenados em um só cubo: *hypercube*. Para compensar, são utilizadas técnicas de compressão física dos dados e esquemas especiais de indexação, de forma a reduzir o impacto das matrizes esparsas e

também atingir um acesso rápido às células dos cubos de dados. A Figura 3.14 contém um exemplo bancário regional de contas correntes referentes a período mensal limitado.

Vale ressaltar que a ocorrência de matrizes esparsas se dá, inevitavelmente, devido à ausência de valores que atendam ao critério de interseção entre dimensões, por exemplo, em um sistema de vendas nem todo produto pode ser vendido em todos os departamentos ou durante todo um período de tempo, como esclarece [WHITE, 1997]. Portanto, quanto maior o número de dimensões, maior a probabilidade de existirem células vazias.

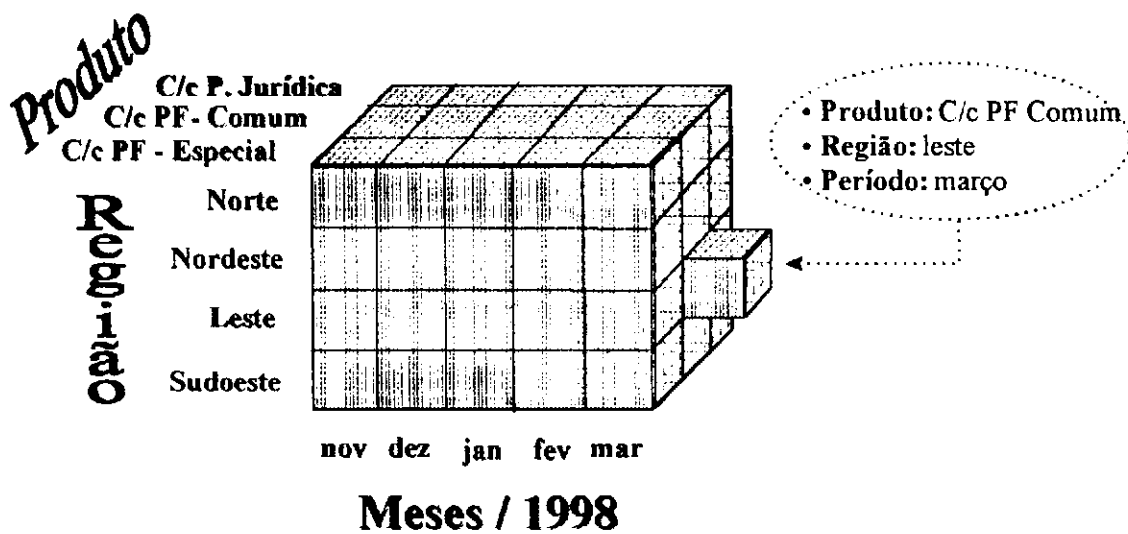


Figura 3.14 - Armazenamento Multidimensional sob a Ótica Comercial

A vantagem dessa modelagem de armazenamento dos dados em matrizes multidimensionais é que as consultas se dão operacionalmente exclusivamente a nível de índices. A atualização de valores das células não causa impacto nos índices, característica que permite facilmente implementarem-se aplicações de leitura-gravação e, portanto, facilita tanto a atualização dos dados, quanto a velocidade das operações OLAP. Contudo, ao alterarmos a estrutura dos dados, teremos que gerar novos índices, tratando-se de algo que requer tempo.

### 3.6.3. Comparativo ROLAP X MOLAP

Iniciaremos esse comparativo analisando a Figura 3.15 de [BERSON & SMITH, 1997], que nos exhibe a relação entre a complexidade e desempenho de uma aplicação OLAP, segundo as modelagens relacional e multidimensional. No gráfico, a área dos círculos indicam o tamanho da base de dados. Assim, para melhorarmos o desempenho das consultas e a sua complexidade devemos optar por MOLAP, mas, se quisermos ter um grande volume de dados necessariamente devemos escolher ROLAP.

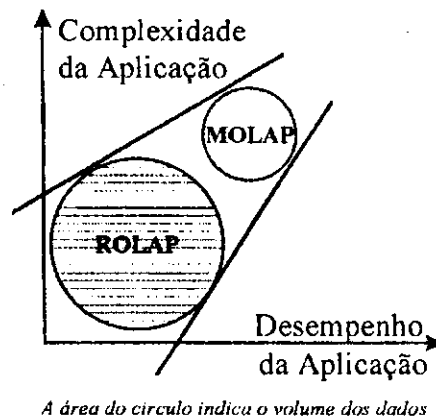


Figura 3.15 - Comparação de Estilos de OLAP [BERSON & SMITH, 1997]

Seguindo o comparativo, o MOLAP é proprietário de um SGBD e pode não possibilitar o uso de ferramentas de sistemas abertos. Já o ROLAP, por ter BDR como base, podemos empregar ferramentas e linguagens padronizadas, tais como SQL, pois os padrões existentes são apropriados a essa tecnologia. As interfaces existentes padronizadas podem não ser compatíveis ao MOLAP, ficando o usuário dependente de toda a arquitetura proposta pelo fornecedor. Contudo, vale ressaltar que muitos esforços têm sido feito no sentido de padronizar também o MOLAP.

Conforme vimos na seção 3.1.3, existe um grupo especialmente preocupado com o destino do OLAP a nível comercial. Segundo [WHITE, 1997], já há um estudo de padronização de API ("Application Program Interface", em português: interface para programas aplicativos), para os produtos OLAP. Ele cita o exemplo da necessidade que a empresa Microsoft teve e respondeu padronizando o acesso ODBC/SQL aos BDR. Provavelmente a falta de uma total transparência do armazenamento e da fisiologia do MOLAP nos SGBDM está ligada a interesses comerciais, devido justamente ao seu aspecto proprietário. Mas, com o uso da padronização as fronteiras da conectividade serão

ultrapassadas. Isso é fundamental para a ampla adesão comercial por parte dos usuários aos produtos MOLAP. [WHITE, 1997] nos alerta que os BDM são bons para manipular dados resumidos, mas nem sempre são tão bons no processamento analítico de grandes volumes de dados em altos níveis de detalhe, porém isso também tem tido a atenção devida pelos fornecedores de BDM.

<b>Característica</b>	<b>ROLAP</b>	<b>MOLAP</b>
Quanto ao volume de dados	Admite grandes volumes	Possui limite
Universo de consulta	Limitada, bastando efetuar-se a operação de junção necessária	Limitada aos caminhos aos índices pré-definidos no modelo de dados
Flexibilidade na alteração estrutural dos dados	Possui flexibilidade, bastando alterar a tabela desejada	Inflexível, pois obriga a total reindexação do repositório
Otimização de consulta	Possível, porém pouco provável, pois os DBR geralmente não aceitam otimização exclusiva para consulta	Previsto, principalmente em função do prévio conhecimento do padrão de utilização dos dados, na geração do caminho ótimo
Desempenho nas Operações OLAP	Lento em comparação aos BDM, que manipulam apenas índices	Eficiente devido ao uso efetivo dos índices das matrizes multidimensionais
Desempenho geral	Boa	Ótima
Carga de dados com alteração na estrutura dos mesmos	Normal	Lenta, devido à necessidade de indexação geral
Quanto à tecnologia	Conhecida e madura	Nova e em amadurecimento
Padronização	Existente	Ainda não há
Desempenho como ferramenta analítica	Razoável, pois, com já dissemos, os BDR foram idealizados para guardar dados	Ideal, pois ela é pré-concebida com essa função

Figura 3.16 - Modelagem ROLAP versus MOLAP

Um ponto relevante no ROLAP é a questão da facilidade de geração de algum dado resumido que se deseje e não se possui no BDR. Porém, no BDM não é possível de ser feito em tempo de execução, pois devem previamente ser gerados os índices e carregada na

matriz todos os dados que serão processados sempre e somente em tempo de carga. Já o tempo de resposta do ROLAP, em relação ao MOLAP, é mais lento para qualquer consulta. Na tentativa de minimizar isso é que desnormalizamos, conforme visto na seção 3.6.1., mesmo assim, o processamento dos SGBDM são mais rápidos com seus algoritmos baseados nos índices.

Para sintetizar a nossa análise, baseado-nos na abordagem de [INMON, 1996], fizemos um quadro comparativo das facilidades multidimensionais para alicerces de organização relacional e a organização multidimensional (por ele denominada de “*em cubo*”), apresentado na Figura 3.16, que, conforme esse autor, constatamos que não se trata de algo claro, ou mesmo intuitivo, a escolha de um ou outro modelo tecnológico. [GRAY & WATSON, 1998] afirmam que, até o momento, não é possível se precisar qual a forma de modelagem que dominará no futuro entre ROLAP e MOLAP.

### 3.7. Arquiteturas OLAP

Encontramos alguns tipos de arquiteturas empregadas no OLAP. Independentemente da questão vista na seção anterior sobre ROLAP e MOLAP, a idéia de arquitetura aqui, é a separação dos serviços OLAP, ou seja, onde ocorrerá o processamento analítico a nível do cliente.

Na Figura 3.17 visualizamos o *DW* no lado esquerdo, contendo dados históricos

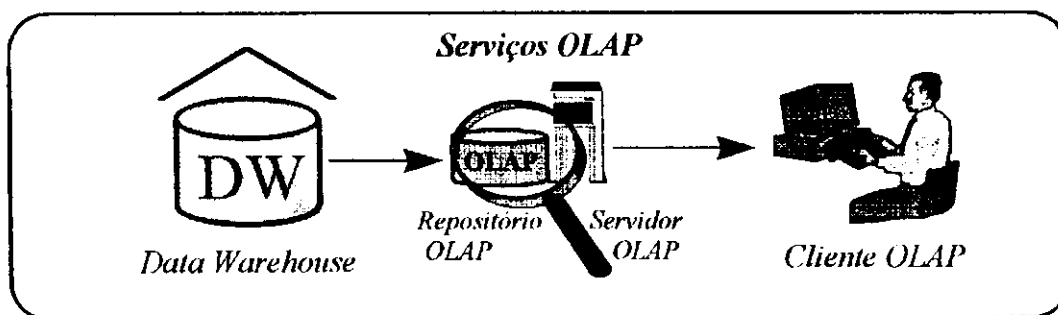


Figura 3.17 - Extremos da Arquitetura de Cliente OLAP

consolidados corporativamente, e um usuário no lado direito, com a sua ferramenta de produtividade para solicitar e visualizar os resultados das suas consultas OLAP. No meio há uma lupa intermediando o usuário do DW, simbolizando o fornecimento da visão OLAP, dada através do servidor OLAP, baseando-se, para isso, no repositório OLAP. Mas, onde a visão multidimensional realmente ocorre? A resposta a essa pergunta se desdobra de acordo com o extremo onde coloquemos a facilidade de serviços OLAP, conforme veremos nas duas formas a seguir.

Antes porém, vale a pena ressaltar que, os *Serviços OLAP* a que nos referimos, são o processamento da consulta, a recuperação (velocidade de pesquisa) e a apresentação dos dados (usabilidade). Para que isso ocorra, é necessária a prévia existência de um repositório de dados modelado conforme visto na seção 3.5 (repositório OLAP) e rotinas que processem consultas multidimensionais. O papel do DW é o de, a partir do armazenamento dos dados corporativos confiáveis e atualizados, alimentar o repositório OLAP. DW será amplamente discutido na seção 4.3 do capítulo 4.

### 3.7.1. Cliente Gordo

Cliente gordo (*Fat Client*) é quando o serviço OLAP (e eventualmente o armazenamento) reside(m) na máquina do próprio cliente. A máquina do cliente vai buscar os dados desejados no repositório OLAP e traz para si para manipulação

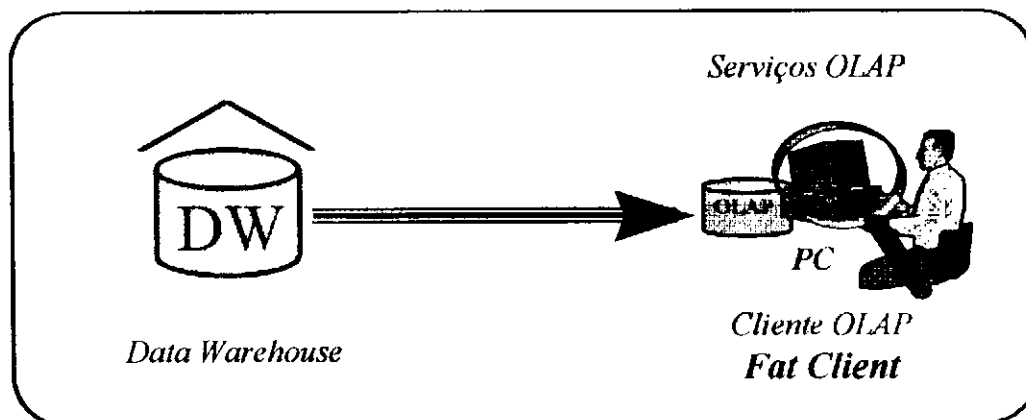


Figura 3.18 - Arquitetura de OLAP – *Cliente Gordo*



multidimensional. A Figura 3.18 nos mostra essa situação. A terminologia “gordo” vem do fato de nela, na máquina do cliente, estar o serviço de OLAP. O aspecto escalabilidade, ou seja, a capacidade de crescimento do equipamento, e usabilidade, a possibilidade de vários usuários simultâneos, constituem pontos fracos nesta arquitetura, pois, nela há o limite de um, e apenas um, cliente por máquina para Serviços OLAP, o que é bastante restritivo a nível corporativo. Porém, são pontos fortes nessa abordagem tanto a segurança, quanto a facilidade de manutenção, pois, o impacto de qualquer manutenção nessa ferramenta é minimizado pela sua singularidade. Contudo, uma vez ela fora de ordem, não há processamento analítico.

### 3.7.2. Cliente Magro

Já no Cliente magro (*Thin Client*), o serviço OLAP reside em um servidor distante do cliente. Ele pode utilizar um computador conhecido como NC - *network computer* (em português, computador de rede), que não possui grande capacidade de armazenamento em discos rígidos para programas e dados, e não são tão caros, tanto na compra, quanto na sua manutenção [GRAY & WATSON, 1998]. Portanto eles requerem uma infra-estrutura operacional que é dada por um Servidor OLAP, disponibilizando todos os tipos de serviços OLAP para todo o grupo de NC via um Servidor de Comunicação NC.

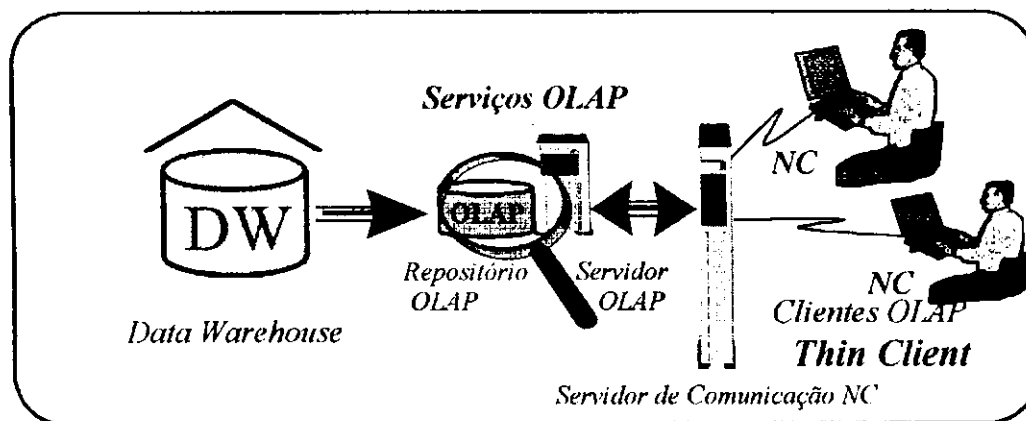


Figura 3.19 - Arquitetura de OLAP – *Cliente Magro*

A Figura 3.19 nos mostra dois servidores, o de OLAP e o de comunicação com os NC. A máquina do cliente encaminha uma solicitação OLAP e recebe a resposta dessa consulta a partir dos dados previamente organizados no repositório OLAP, processada pelo servidor OLAP e entregue a ao servidor de comunicação NC para o transporte até o NC solicitante. Esta arquitetura é mais flexível e tem sido utilizada cada vez mais. Contudo, no lugar dos NC tem-se utilizado PC. Acreditamos que isso seja devido ao estado imaturo dos NC, ainda sem um padrão definido e carentes de uma definição funcional mais consistente. Vários fornecedores têm investido neste tipo de máquina, mas ainda não houve um consenso sobre são pontos fortes o seu modo de operação como há nos PC. São pontos fortes, a escalabilidade e a usabilidade. São pontos fracos, a complexidade na manutenção e questão da segurança.

### **3.8. *Resumo***

Vimos, neste capítulo sobre OLAP, a história do seu surgimento, com a exposição da formulação das 12 regras do que seria o processamento analítico descrito a partir de um trabalho prático e toda a polêmica gerada a este respeito. Devido a toda controvérsia deste contexto na formalização do OLAP, foram expostas as opiniões de diversos autores especialistas no assunto. Ao vermos os diversos pontos de vistas sobre OLAP, constatamos a equivalência deste conceito com o de AM já existente, que na época em que surgiu não contava com os recursos gráficos das modernas interfaces de usuário e provavelmente por isso não obteve êxito comercial.

Agora a AM aparece de cara nova com várias operações especiais para manipular os dados visualmente a fim de possibilitar ao usuário análises por diversos ângulos empresariais através de diferentes óticas dos atributos dos fatos comerciais. Exemplificamos um modelo simples de relatório, onde fazemos ressalvas e distinções sobre as fronteiras do EIS/OLAP e outras tecnologias. Mostramos os tipos de modelagens mais usadas nos repositórios OLAP, assim como as arquiteturas viáveis, ressaltando que, independentemente do tipo de modelagem utilizada, é importante termos em mente que o

armazenamento dos dados é uma função do EIS/OLAP, para que não venhamos a confundi-la com outros tipos de tecnologia, no caso o DW, que veremos a seguir.

## **4. DATA WAREHOUSING (DWING)**

Neste capítulo abordamos o Data Warehouse (DW), repositório histórico de dados corporativos, e todo o seu cenário existencial, o Data Warehousing (DWing). Veremos a sua definição, importância, uma arquitetura básica para seu entendimento, outros tipos de arquiteturas existentes (classificadas por camadas), interações com outras tecnologias, tais como INTERNET e *Data Mart* e, ao final, faremos um breve relato tanto dos problemas ainda em aberto, quanto do estado da arte dessa tecnologia. É dada ênfase ao aspecto temporal do DW, propondo enfoques mais adequados à realidade dos fatos da representação desses aspectos na área de BD e, especialmente, no âmbito dos BDT, ainda pouco estudados na bibliografia sobre DW.

### **4.1. Contexto, Importância e Enfoques**

Como vimos no capítulo 2, mais precisamente na seção 2.3.2, idealizou-se um tipo particular de depósito de dados especialmente concebido de forma a viabilizar a utilização dos dados corporativos pelos sistemas EIS e DSS, o armazém de dados (DW, em inglês, Data Warehouse). Conforme dissemos no capítulo anterior, constatamos que, na verdade, trata-se de algo intermediário entre os SA e os EIS e DSS. É uma tecnologia de armazenamento de dados que permite fazer com que o dados operacionais, uma vez selecionados e armazenados de forma criteriosa e adequada, sejam aproveitados por aqueles SG de forma eficiente e que produzam resultados confiáveis, enquanto ferramentas gerenciais. A esse processo ambiental do DW dá-se o nome de DWing (do inglês e ainda sem tradução, "Data Warehousing").

Devido a fatores oriundos de uma economia globalizada que aumenta e diversifica ainda mais a concorrência de mercado, são requeridos mecanismos que identifiquem e

destaquem as instituições, ou seja, imprimam no mercado os seus diferenciais competitivos. Seguramente, os passos básicos nesta direção são: o autoconhecimento como organização, o mapeamento do comportamento dos fornecedores e a sua identificação com a clientela. Para isso precisamos viabilizar a integração sistêmica, de forma a objetivar a visão corporativa da instituição, para que o corpo gerencial pertencente ao nível estratégico organizacional (vide Figura 2.3 da seção 2.2.1 no capítulo 2 ) possa decidir melhor com base nas informações da organização de forma a destacá-la em meio ao mercado competitivo.

Integração sistêmica a nível organizacional não é algo fácil. A tarefa de selecionar os dados de forma a, em seguida, centralizá-los no DW, requer inicialmente um levantamento minucioso dos vários repositórios de dados operacionais da organização (ou seja compreender a “teia de aranha” da organização, vide a idéia dada pela Figura 2.7, da seção 2.3 no capítulo 2), a fim de se poder efetivamente gerar um DW (orientar o fluxo de dados departamentais em direção a um único recipiente, o DW, vide Figura 2.12, da seção 2.3.2 no capítulo 2). Isso não é trivial, e tem seu grau de complexidade diretamente proporcional ao porte institucional da organização que deseja trabalhar com sistemas EIS e DSS.

Surge então a idéia do DWing, como uma idéia modularizada – conforme veremos na próxima seção – de forma a prover, como subsidio tecnológico, a base necessária para atender a esse novo estado de economia competitivamente globalizada. Porém, conforme já afirmado e exemplificado com um estudo de caso na seção 2.3.1 do capítulo 2, constatamos que muitas tentativas no âmbito comercial nessa direção não obtiveram o êxito esperado.

Desta forma, iniciamos um estudo voltado para grandes e complexas corporações, objetivando atingir o maior grau de complexidade possível no assunto. Sendo assim, não poderíamos deixar de abordar, ao longo do documento, a instituição bancária – pivô comercial da economia –, por ser o sistema financeiro brasileiro um dos mais complexos, modernos e ágeis da atualidade. Por estas razões os bancos brasileiros estão sempre na vanguarda tecnológica.

As instituições financeiras, como um todo, passaram por algumas adaptações às novas necessidades de mercado, mas tendem a sofrer uma mudança estrutural no seu comportamento, em um futuro próximo, para a sua própria sobrevivência. Isso se deve à essa transformação sócio-político-econômica que está ocorrendo a nível mundial. Novos fatores agravantes da competitividade tais como a globalização, valor temporal [MASUDA, s.d.] e outros, estão cada vez mais presentes.

Consideremos o setor bancário. Muitos acreditam que, nesse novo cenário, a sobrevivência dos bancos dependa da diversificação nas vendas, extrapolando inclusive a área financeira, entrando em ramos de comércio de produtos não financeiros. Isso já é percebido na prática por alguns bancos que expandiram seus negócios, fazendo uso da sua própria estrutura informatizada para a venda de bilhetes, fornecimento de cupons [SOUCIE, 1996], acesso a cinemas com cartão magnético [LEANDRO, 1996], além da venda de seguros de vida e de veículos [TONIELLO ed., 1996].

Estes requisitos de mercado exigem novas tecnologias visando encontrar o público alvo ideal, para direcionar estas novas frentes de comércio. É aí então que se necessita da tecnologia de DWing como forma de integração corporativa, viabilizando também essas novas frentes de negócio.

Não devemos esquecer que o surgimento do DWing não se deu na academia e sim decorrente do “livre jogo das forças de mercado, segundo o qual os preços resultam da concorrência empresarial e da preferência dos consumidores” [ROSSETTI, 1984]. Conforme [GILL & RAO, 1996], “O dia-a-dia do usuário do DW é com o uso comercial”. Desta forma, a academia contribui hoje no incremento da solução encontrada pelas empresas – o DWing – a um problema tecnológico que eles mesmos criaram no seu ambiente organizacional – a sua multiplicidade de bases de dados devido à descentralização tecnológica departamental.

Um ponto de vista defendido<sup>33</sup> por alguns fornecedores é que o DWing possibilita uma maior “democratização na empresa”. Democratizar, ou seja, popularizar, o nível

---

<sup>33</sup>1º Seminário Internacional de Data Warehouse - 11 e 12 de março de 1997 - São Paulo/SP - Brasil.

estratégico de uma organização é, indiretamente, permitir e facilitar a circulação indiscriminada dos dados corporativos com o acesso livre a qualquer que seja o funcionário. Caso um funcionário não esteja suficientemente preparado, de forma consciente, da importância desses dados, as conseqüências do seu possível vazamento e uso pela concorrência poderá causar um grande impacto, pois eles são suficientes na conclusão das “boas ou más informações” (conforme visto seção 2.1) por parte da concorrência, que certamente agirá de forma a utilizá-las em seu benefício e contra os seus concorrentes.

Atribuir a característica de democratização no DWing, é uma distorção no seu enfoque funcional, enquanto tecnologia. Para isso existem algumas ferramentas de integração corporativa, tal como correio eletrônico (*e-mail*), plataforma para trabalho em grupo (*workgroups*), etc., que efetivamente atuam de forma a facilitar o fluxo comunicativo empresarial, proporcionando uma maior integração. Dessa forma, ao se tentar passar a idéia de se estar atacando e solucionando (pelo menos parcialmente) um dos maiores problemas das grandes corporações, a comunicação, está se distorcendo e confundindo a finalidade dessa tecnologia.

Para concluir, acreditamos que atribuir ao DWing mais funcionalidades do que a de ser um simples – e por isso eficiente, conciso – e daí preciso e não redundante – repositório de dados organizacionais, histórico, consolidado corporativamente para consultas por sistemas que atuem no nível estratégico empresarial, é “puro romantismo”, para não dizer, marketing enganoso.

#### **4.2. Definição e Arquitetura Básica**

O processo de construção de um repositório histórico corporativo atualizado e confiável – também chamado por muitos de “jornada” (devido à sua contínua e interminável tarefa de torná-lo atualizado de modo a refletir o mais preciso e rápido

---

possível a estrutura organizacional operacional integralmente) – é conhecido como DWing.

DWing é também definido como o processo de reunião de dados históricos em um *repositório*<sup>34</sup> [GILL & RAO, 1996]. Ainda de acordo estes autores, o DWing possui diversas fases para sua efetivação e elas devem conter basicamente os seguintes componentes com as funções que se seguem:

- executar a extração dos dados de sistemas de produção operacionais;
- “limpar e refinar” estes dados;
- adicionar o tempo e procedência dos dados extraídos;
- estruturar e armazenar os dados para possibilitar a gama de ferramentas de análises e os requisitos das análises de informação.

Como se pode observar, os aspectos temporal e espacial foram agrupados e isolados em uma única fase do roteiro desses autores, evidenciando um destaque a estes pontos. Como o horário (o tempo) varia em função das coordenadas geográficas (espaço), e a geografia é essencial para situarmos qualquer fenômeno multidimensional analítico ou sintético no DWing, precisaremos ter especificado de forma absoluta a geografia (ou então possuamos funções de conversão geográfico-temporal<sup>35</sup>) para situarmos o tempo de forma genérica no DW, já que tais projetos podem ter amplitude global. Por estas razões, dedicamos a seção 4.3.2 a esse assunto.

Qualquer arquitetura em DWing gira em torno do DW. A rigor, existem diversos tipos de arquiteturas para estes ambientes, conforme veremos na seção 4.5, que se diferem em quantidade, finalidade e interoperabilidades, ou seja, estes são sempre os elementos variantes nas diversas arquiteturas de DW. Porém, de uma maneira genérica, podemos dizer que um ambiente típico de DWing é constituído por um conjunto de procedimentos

---

<sup>34</sup> Para [GILL & RAO, 1996] *repositório* é o compêndio de todas definições relacionadas com tecnologia e especificações para itens que são importantes para as aplicações de software e bancos de dados correntemente desenvolvidas.

<sup>35</sup> A diferença temporal pode variar em até 12hs, para os casos de termos locais antípodas, com por exemplo, Brasil e Japão.



de extração, integração e transformação dos dados operacionais, um repositório principal chamado de DW e um ou mais sistemas de apoio à decisão com funções analíticas e/ou sintéticas gerenciais (os EIS e DSS) É possível uma perfeita compreensão do todo e da sua dinâmica a partir da arquitetura básica retratada na figura abaixo:

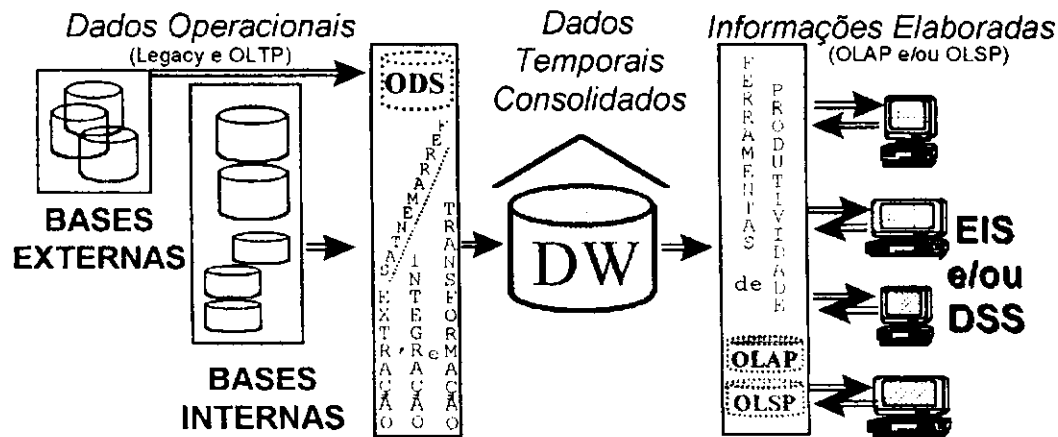


Figura 4.1- Diagrama Básico de Data Warehousing

De acordo com a Figura 4.1, caminhando da esquerda para direita, temos:

- *bases internas* – são fontes de dados operacionais da organização oriundas dos SA;
- *bases externas* – são fontes de dados de outras corporações, vindas via teleprocessamento (canal de voz/dados, internet, etc.), fitas, etc.;
- *ferramentas de extração, integração e transformação* – objetivam criar, a partir dos dados operacionais (*bases internas* e *externas*), a carga do *DW*, ou seja, alimentar este repositório de dados, através da replicação de dados, sem causar impacto no ambiente operacional de modo seletivo consolidado sob a ótica corporativa, e consistente e atualizada sob a ótica temporal;
- *DW* – banco de dados especialmente desenvolvido para suportar grandes volumes de dados derivados, armazenados historicamente, de forma a prover a centralização dos

dados corporativos para utilização a nível estratégico pela alta cúpula gerencial da organização;

- **ferramentas de produtividade** – objetivam, a partir do *DW*, a captação seletiva dos dados para eventualmente carregar repositórios específicos OLAP e OLSP dos sistemas EIS e DSS respectivamente (de preferência sem causar impacto no seu ambiente de consulta), formatando o seu conteúdo, executar processamentos tanto analítico, isto é, análises multidimensionais, quanto sintético, ou seja, *Data Mining*, de modo a permitir consultas diversas e imprevisíveis por parte do usuário dos EIS ou DSS, de forma que ele possa interagir com esses sistemas.

Vale ressaltar, contudo, que a prática do DWing tem evidenciado, para grandes e complexas instalações, que as **ferramentas de extração, integração e transformação** necessitam de um ou mais repositórios voláteis para viabilizarem as suas finalidades entre dois ou mais sistemas (ou mídias) que possuam redundância e/ou dependência dos dados na alimentação do DW. Esse tipo de repositório é chamado de **ODS** (do inglês *Operational Data Store*), com as características de um depósito de dados não temporal (não necessariamente um BD, podendo ser um simples arquivo seqüencial indexado), basicamente com a finalidade de integração corporativa de várias fontes de dados operacionais não integrados para, através da conciliação dos dados entre os sistemas e objetivando uma maior qualidade nos dados, atualizar o DW. Veremos ODS com maiores detalhes na seção 4.4, antes, porém, vamos ao âmago do DWing, o DW.

### 4.3. Data Warehouse (DW)

Conforme vimos na seção anterior, um DW é construído a partir de dados operacionais *legacy* e/ou OLTP, gerados pelos diversos sistemas da organização nos seus processamentos cotidianos. Esses dados são selecionados de forma consistente e consolidados corporativamente para armazenamento histórico, através de programas especiais extratores de dados. Este repositório é definido por Inmon da seguinte forma [INMON, 1996; GILL & RAO, 1996; BRACKETT, 1996]:

“o DW é uma coleção de dados integrada, orientada a assuntos, não volátil e temporal com a finalidade de respaldar decisões gerenciais”.

Ainda segundo [INMON, 1996], o DW é concebido para suportar funções de DSS, onde cada unidade de dado é relevante em algum momento no tempo. Vale lembrar que o que ele chama de “funções de DSS” equivale ao que chamamos de sistemas que atuam no nível estratégico empresarial, ou seja, os EIS/DSS. Para que se tenha um DW é fundamental levarmos em consideração alguns pontos. São eles: os dados, o tempo dos dados, a modelagem, a funcionalidade do armazenamento e o tempo do armazenamento. Estes elementos inter-relacionam-se e são responsáveis pelo bom funcionamento do DW.

#### 4.3.1. Os Dados

Com o atual desenvolvimento das várias tecnologias de armazenamento, conectividade<sup>36</sup> e a facilidade da elaboração de novos sistemas, temos que estar atentos à questão da duplicidade de um dado em relação a outros já existentes em outro sistema da organização. É o fenômeno dos “ambientes de desenvolvimento espontâneo”, por nós abordados na seção 2.3, visualizado na Figura 2.7, do capítulo 2, cada vez mais comuns nas grandes corporações.

Quando se possui várias fontes de dados, ou seja, lida-se com aplicativos de diferentes tecnologias operacionais, agrava-se a necessidade de se fazer uma verificação mais detalhada quanto às questões da origem, formatação e atualização dos dados nas suas diferentes fontes. Estes problemas são muito comuns em instituições que convivem com a informatização e as suas pluralidades tecnológicas, fruto da competição na busca de maior eficiência, eficácia e menor custo operacional (na prática perseguidos na respectiva ordem inversa).

---

<sup>36</sup>Conectividade é a capacidade de intercomunicação entre tecnologias, por exemplo uma determinada linguagem de um BD de um computador consultado os dados de outro BD em outra máquina.

O maior problema da duplicidade é a questão da falta de consistência que ela pode ocasionar na produção de informações. Caso estes dados refiram-se à mesma semântica, porém possuam valores diferentes, será necessária uma prévia análise com determinação de qual dado é o mais adequado para ser armazenado no DW. Portanto, a qualidade das informações contidas neste repositório e, por conseguinte, utilizadas posteriormente pelas ferramentas baseadas em DW, será função desta seleção prévia dos dados. Caso contrário, a confiabilidade no DW poderá ser comprometida.

Assim, fica evidenciado um pré-requisito básico no DWing, que se reflete diretamente no DW, ou seja, a criteriosa auditoria em busca da eliminação do problema de duplicidade dos dados, com a eleição do melhor dado. Segundo [INMON, 1996], o *melhor dado* é aquele que é mais recente, mais preciso, mais completo, mais perto da fonte externa e mais compatível estruturalmente. Perseguindo o *melhor dado* deparamos com um problema que muitas vezes acontece nas organizações, que é a desatualização dos dados. Porém esse não é um problema do DW e sim no DW. As conseqüências da não observância desse aspecto podem também afetar o sucesso do DWing. Por isso devemos organizar previamente os dados, de forma a selecionar apenas aqueles com características de *melhor dado*, para alimentar o DW, mantendo-o sempre atualizado e com dados de qualidade. É interessante observar que o melhor dado de hoje, pode não ser (e certamente não será) o *melhor dado* de amanhã. Isto é a influência do tempo que veremos a seguir.

#### 4.3.2. O Tempo dos Dados

Uma vez obtido o *melhor dado*, precisamos registrá-lo no DW de forma a construir a sua história, isto é, situá-lo no tempo identificando o período em que o mesmo permanece válido. Como já dissemos antes, na seção 4.2, a geografia influi na contagem do tempo, por isso, para trabalharmos com dados com diferentes coordenadas geográficas, precisamos equalizar o local, relativizando o tempo, ou então equalizarmos o local a partir de uma outra dimensão (, para termos, nos dados, uma relação temporal efetiva.

Por exemplo, imaginemos um banco com duas filiais no estado do Pará no Brasil<sup>37</sup>. O Pará possui dois fusos horários de uma hora entre eles. Dois cidadãos poderão, simultaneamente e em diferentes agências, fazer depósitos na conta de um cliente qualquer do banco neste estado e ter nos seus comprovantes de depósitos diferentes registros do mesmo instante.

Na alimentação do DW os fusos horários precisam ser considerados e processados de forma a refletir homogeneamente o tempo dos dados. A realidade da equivalência temporal baseada na equiparação geográfica pode nos levar a valores temporais, que antes eram diferentes, a valores relativos iguais e vice-versa. Dessa maneira poderemos realmente nos situar nos fatos, de forma fiel e refletindo a realidade de um panorama qualquer (como o do exemplo da movimentação bancária), de maneira a registrar a sua história. A partir desses dados temporalmente relativizados e equalizados globalmente será possível extrair informações estratégicas seguramente. Este procedimento é fundamental para permitir aos sistemas EIS e DSS os seus processamentos de forma confiável.

Resolvida a questão geográfica, fundamentalmente o registro do tempo se faz através de um atributo de tempo associando-o a atributos e/ou cada registro (tupla) variantes no tempo. Por exemplo, João, enquanto cliente do banco durante 3 anos, teve uma motocicleta durante dois anos e, posteriormente, comprou um automóvel. Esse atributo de tempo (normalmente uma data) pode ser convertido em número (inteiro ou de ponto flutuante). Um exemplo disso é a utilização do tempo absoluto do computador. Como exemplo prático temos o formato chamado de **ABSTIME** encontrado em alguns produtos tais como o CICS da IBM obtido através do comando **ASKTIME**, que informa o tempo do relógio do computador. Este formato de registro de tempo é definido como um número inteiro compactado e sinalizado de 15 posições<sup>38</sup>. Nele o tempo corrente é contado a partir da zero hora de primeiro de janeiro de 1900 e com granularidade em milésimos de segundo.

---

<sup>37</sup>O Brasil possui quatro fusos horários (hora legal de Brasília, +1h, +2h e -1h). Na alta estação é adotado o *horário de verão* para Brasília e algumas regiões do país. Nesta época, a diferença de fuso neste país chega a três horas em parte do Amazonas a oeste e no estado do Acre.

<sup>38</sup>Na linguagem COBOL é codificado como PIC S9(15) COMP-3.

Desta forma, operamos mais facilmente com o tempo e com uma precisão de milésimos de segundo. Usando-se este formato, a aritmética básica das operações de adição e subtração são suficientes na manipulação destes atributos, quando convertidos. Vemos na tabela da Figura 4.2, nas colunas sombreada na cor cinza, os valores referentes ao tempo normalmente utilizado (sistema sexagesimal e calendário gregoriano) na granularidade de milésimos de segundo. Nas colunas brancas temos a equivalência em tempo absoluto e na mesma granularidade.

<b>Tempo Sexagesimal</b>	<b>ABSTIME</b>	<b>(Tempo Sexagesimal em Milésimos de Segundo e) Data Gregoriana</b>	<b>ABSTIME</b>
um segundo	1.000	(00:00:30.154h) 20 de julho de 1900	17.280.030.154
um minuto	60.000	(08:50:00.000h) 01 de março de 1964	2.024.929.200.000
uma hora	3.600.000	(15:59:20.573h) 06 de setembro de 1998	3.114.118.220.573
um dia	86.400.000	(00:00:00.010h) 01 de janeiro de 2000	3.155.760.000.010

Obs.: O ABSTIME é definido com marco zero em primeiro de janeiro de 1900 e com a precisão em milésimos de segundo.

Figura 4.2 - Tabela de Tempos Sexagesimais com Data Gregoriana e Tempo Absoluto (ABSTIME)

Esse atributo de tempo deve ser associado aos outros atributos (aqueles que necessariamente variam em função do tempo) de forma a refletir a questão cronológica dos dados, ou seja, identificar claramente o seu período de validade. [KIMBALL, 1996] aborda esse aspecto temporal distinguindo o que ele chama de “dimensões de modificação lenta” para os atributos que variam no tempo de forma “lenta” e que são utilizados para elaboração de dimensões. Ele exemplifica com o atributo pessoal do estado civil de um cidadão e explicita três soluções, utilizando duas, que efetivamente guardam o histórico, o uso de um atributo de tempo em cada uma delas para resolver a situação de mudança desse campo que ele classifica como de “modificação lenta”.

Deste modo ele cria a modelagem dos outros atributos historicamente, partindo do pressuposto no caso da certeza da “lentidão” desses atributos. Contudo, com as mudanças imprevisíveis do mundo moderno, tendo em vista a revolução de hábitos, costumes,

práticas, cultura, economia, etc., hoje vivenciados, preferimos não partir desse pressuposto que consideramos um tanto incerto.

O uso de um simples *timestamp*, ou seja, um selo de tempo (podendo até ser do tipo ABSTIME ou outro qualquer) associado aos dados, mediante o contínuo armazenamento destes dados no DW, a fim de construir um histórico dos dados, não oferece muita flexibilidade ao DW na sua tarefa de tornar seus dados disponíveis eficientemente para as *ferramentas de produtividade* alimentarem os seus repositórios dos sistemas do tipo EIS/DSS. Isso porque um simples *timestamp* não reflete perfeitamente a validade do dado em foco, apenas a época do seu surgimento (ou seja seria como termos, uma certidão, ou data de nascimento do dado) no repositório impondo às *ferramentas de produtividade*, na melhor das hipóteses, ter que fazer outros acessos a outros dados do DW, a fim de obter efetivamente sua duração efetiva e construir (quando possível, pois na prática às vezes as informações são incompletas) o histórico de um dado, além de trazer alguns outros inconvenientes tais como a superpopulação dos repositórios com implicações maiores, caracterizando também um tipo de redundância. Esse é um problema que pode ser evitado com uma modelagem adequada, conforme veremos a seguir.

#### 4.3.3. A Modelagem do DW

Pelo caráter histórico do DW, ou seja, o armazenamento de informações no tempo, a modelagem temporal assume grande importância para a precisão dos sistemas com enfoques estratégicos. Uma modelagem temporal eficiente deve, dentre outras coisas, otimizar a acumulação de dados repetidos durante um período de tempo, já que eles são inevitáveis e fundamentais de serem registrados. Isto é, deve contemplar especialmente a inserção de dados repetidos em um período de forma a evitar a superpopulação do repositório.

Neste ponto do nosso trabalho, abstraímos totalmente da bibliografia especializada em DW, que, quando aborda essa questão de modelagem, associa precipitadamente funções OLAP aos DW e nos baseamos também no que observamos em campo –

problemas e empecilhos que resultaram em impactos práticos (tal qual a inflação nas bases de dados) na implementação de alguns DW – para sugerir uma modelagem fundamentada na modelagem temporal específica. Assim sendo, numa modelagem temporal são considerados os tempos, as regras e as hipóteses que se seguem [SNODGRASS, 1987]:

- atributos de tempo TS (do inglês “time-start”, tempo inicial) e TE (também do inglês “time-end”, tempo final);
- a informação é considerada válida em todos os tempos entre o TS e o TE;
- os atributos de tempo estão divididos em dois tipos: de validade e de transação.<sup>39</sup>

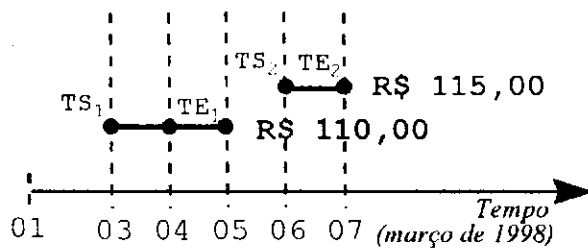


Figura 4.3 - Acompanhamento Temporal dos Saldos de um Cliente Hipotético

Vimos, na seção anterior, os atributos de tempo associados aos dados e suas formas de representação. Aqui, para traduzir a questão da duração efetiva de um dado levantada na seção anterior, [SNODGRASS, 1987] especifica dois instantes, o TS e TE, delimitando a

efetividade do dado nestes extremos e em todos os pontos nesse intervalo. A Figura 4.3 contém um gráfico do saldo bancário de um cliente qualquer. Nele encontramos informações sobre os saldos diários do período de três a sete de março de 1998. Lá vemos TS<sub>1</sub> e TE<sub>1</sub> equivalendo aos dias 03 e 05 de março de 1998 respectivamente, quando o cliente tem saldo de R\$ 110,00. Conforme o que foi afirmado e de acordo com essa figura, no dia 04 de março de 1998 esse cliente também possuía saldo de R\$ 110,00. No dia 06 de março de 1998 houve mudança no valor do saldo, quando nesse instante associamos outro par de valores, são eles TS<sub>2</sub> e TE<sub>2</sub>, 06 e 07 de março de 1998, respectivamente. Exatamente pelo caráter progressivo do tempo, ele possui uma ordenação natural, o que resulta numa classificação implícita, ou seja, o TS é sempre menor que o TE. Por essa razão, delimitamos sempre os períodos de tempo em que os valores se mantiveram constantes e

<sup>39</sup>O tempo de transação é o tempo de processamento, enquanto o tempo de validade é um tempo lógico atribuído. Por exemplo, um saque em um caixa eletrônico (“cash dispenser”) fora do horário de expediente bancário, tem como tempo de transação a hora efetiva do saque e como tempo de validade o próximo dia útil após o dia do saque.



atribuímos a eles um TS e um TE, no nosso caso da Figura 4.3 os saldos de R\$ 110,00 e R\$ 115,00, respectivamente os pares ordenados  $TS_1, TE_1$  e  $TS_2, TE_2$ .

Para melhor elucidar a questão das vantagens de uma modelagem temporal, tomemos ainda o exemplo de um sistema bancário qualquer de contas correntes. Suponhamos que se queira acumular o valor do saldo diário de clientes para uma futura análise qualquer. Nossos clientes serão **Ana Paula** e **Adriana** (observe que este último reflete o perfil exposto na Figura 4.3). Inicialmente, para efeito de simplificação, usaremos apenas um atributo de tempo, que no caso está expresso na forma de data gregoriana para ficar mais sugestivo, mas poderia estar em qualquer outra forma de expressão do tempo (como por exemplo o ABSTIME visto na seção anterior). Essa data reflete o saldo dos

Data	Cliente	
	Adriana	Ana Paula
03/03/98	R\$ 110,00	R\$ 105,00
04/03/98	R\$ 110,00	R\$ 115,00
05/03/98	R\$ 110,00	R\$ 120,00
06/03/98	R\$ 115,00	R\$ 125,00
07/03/98	R\$ 115,00	R\$ 130,00

clientes ao final do dia, conforme convenção bancária. Na Figura 4.4 temos uma relação dos saldos das nossas duas clientes durante o período semanal de três a sete de março de 1998.

Figura 4.4 - Saldo de Contas Bancárias de Dois Clientes em um Período.

Podemos notar, na planilha (Figura 4.4), o saldo de **Adriana** não varia durante os períodos de três a cinco e de seis a sete de março, enquanto que o de **Ana Paula** tem variação diária. Veremos a seguir duas modelagens relacionais: uma sem considerar a faixa temporal e outra baseando-se nela e usando os atributos de tempos TS e TE.

	Cliente	Data	Saldo
1	Adriana	03/03/98	R\$ 110,00
2	Adriana	04/03/98	R\$ 110,00
3	Adriana	05/03/98	R\$ 110,00
4	Adriana	06/03/98	R\$ 115,00
5	Adriana	07/03/98	R\$ 115,00
6	Ana Paula	03/03/98	R\$ 105,00
7	Ana Paula	04/03/98	R\$ 115,00
8	Ana Paula	05/03/98	R\$ 120,00
9	Ana Paula	06/03/98	R\$ 125,00
10	Ana Paula	07/03/98	R\$ 130,00

Figura 4.5 - Tabela Normalizada não Temporal

A Figura 4.5 (ao lado) nos mostra uma tabela normalizada que objetiva nos informar os saldos diários dos clientes. Podem ser observadas nela redundâncias, como por exemplo as linhas 1, 2, e 3. Todas elas têm a mesma informação, apenas havendo uma variação na data. Ou seja, no que concerne à cliente **Adriana**, poderíamos obter uma modelagem que suprimisse os dados repetidos. Este é o caso da tabela da Figura 4.6, que é projetada segundo a modelagem relacional temporal, abordagem utilizada em BDT.

Observamos que o número de linhas está reduzido em relação à tabela da Figura 4.5. Isso se deve à introdução da coluna **Data final** que, causa uma redução na quantidade de linhas da tabela.

	Cliente	Saldo	Data inicial	Data final
1	Ana Paula	R\$ 105,00	03/03/98	03/03/98
2	Adriana	R\$ 110,00	03/03/98	05/03/98
3	Ana Paula	R\$ 115,00	04/03/98	04/03/98
4	Ana Paula	R\$ 120,00	05/03/98	05/03/98
5	Ana Paula	R\$ 125,00	06/03/98	06/03/98
6	Adriana	R\$ 115,00	06/03/98	07/03/98
7	Ana Paula	R\$ 130,00	07/03/98	07/03/98

Figura 4.6 - Modelagem Normalizada Temporal

Como acabamos de constatar, tratando-se de redundância de origem temporal, para contornarmos tal situação é mais apropriado o emprego da modelagem baseada nos dois tempos (TS e TE), que delimitam a validade do dado. Essa melhoria pode ser incrementada

no DW com a adoção deste enfoque no seu projeto, porém, isto não tem sido praticado ou mesmo considerado nas bibliografias especializadas atuais.

A definição do DW como um BDT deverá facilitar a tarefa de centralizar os dados dispersos em uma grande corporação qualquer na tentativa de, a partir dessa associação tecnológica conceitual-funcional, poder tornar disponível todo e qualquer dado concernente à organização aos sistemas do tipo EIS/DSS, de forma a atender com informações os anseios e necessidades da sua cúpula estratégica corporativa.

A dinâmica de funcionamento de um BDT com o registro de dois tempos (TS e TE), por utilizar-se do tipo de processamento OLTP, comporta rápidas atualizações de forma confiável, ou seja, permite reter e tornar disponíveis no seu repositório os dados tão agilmente quanto possível e desejável. Vimos que o DW é um típico BDT, agora nos certificaremos, na seção seguinte, da sua funcionalidade, isto é, o armazenamento específico direcionado ao suprimento de dados para os sistemas EIS e DSS.

#### 4.3.4. A Funcionalidade do Armazenamento

Conforme [GILL & RAO, 1996]: “A arquitetura de um DW deve separar claramente as fontes de dados (quais dados estão disponíveis) da forma e formato nos quais os dados estão apresentados e estruturados para o usuário final”. Neste ponto, há uma clara separação entre a função do DW e a função das ferramentas de produtividade (especificadas na seção 4.2.). Este ponto é crucial para o entendimento desta tecnologia, simplificando o seu entendimento e dando a ele a sua dimensão exata, ou seja, simplesmente um repositório de dados temporais consolidados de forma corporativa.

É comum encontrarmos autores atribuindo funções de repositórios OLAP/OLSP ao DW. Muitos o fazem de forma implícita, já outros, a exemplo de [KIMBALL, 1996], utilizam-se de uma nova terminologia: o DDW (“Dimensional Data Warehouse”, em português DW Dimensional), um tipo de arquitetura que será abordada mais detalhadamente na seção 4.5.2.2. Essa abordagem de repositório multifunção não nos é

muito simpática, pois, conforme visto no capítulo 2, onde abordamos EIS/OLAP (seção 2.2.2.2) e DSS/OLSP (seção 2.2.2.3), eles possuem características bastante distintas. OLAP foi ainda mais detalhado no capítulo 3.

Ao se optar por incluir funções OLAP no DW, estamos transformando o repositório que deveria refletir a realidade dinâmica do ambiente operacional historicamente, em um depósito de dados estático e restrito, quanto à sua atualização, devido às características singulares inerentes a esse tipo de processamento. O mesmo ocorre ao DW se o imaginarmos com funções OLSP. Poderemos ter uma situação mais crítica ainda, onde precisaríamos atualizar o DW multifuncional, o DDW, com os dados mais recentes para necessidades de OLAP em que algum usuário estivesse fazendo uma consulta OLSP, e vice-versa. Aliás, as modelagens OLAP estão bem maduras, o que não acontece ainda com as OLSP. Associar DW a OLAP/OLSP é diminuir o poderio dessas tecnologias, enquanto ferramentas de processamento de informações, individualmente.

Dessa forma, diferenciamos as tecnologias para evitar o risco de tentarmos agrupar tudo e ao final não conseguirmos atingir o sucesso esperado. Ou mesmo termos que suspender todo o projeto, conforme já ocorrido em muitos casos sem que se constatasse de verdade o porque, devido à impossibilidade de atualização do DW em virtude do tempo por ele gasto para a inserção consolidada de novos dados aos já existentes nele.

Vemos o DW com o processamento OLTP, uma vez que ele deverá estar sempre atualizado corporativamente, lado a lado com os SA e de forma eficiente e eficaz. Como a ele não cabe a tarefa de fornecer informações, e sim de ser apenas um grande “armazém de dados”, o processamento transacional é bem adequado. Dessa forma garantimos a segurança e confiabilidade, com o emprego de uma tecnologia madura e apropriada a grandes volumes de dados. Utiliza-se então o BDR Temporal (BDRT), conforme veremos a seguir.

#### 4.3.5. O Tempo de Armazenamento

Abordamos, nas seções 4.3.2 e 4.3.3, a questão do *armazenamento do tempo nos dados*. Agora veremos o tempo de armazenamento dos dados, ou seja o *armazenamento dos dados no tempo*. Sabemos que a existência da memória do tempo no DW é inerente à sua função: o armazenamento temporal dos dados consolidados corporativamente. “Sistemas DW estão tornando-se populares no gerenciamento de dados variantes no tempo a fim de analisar tendências comerciais e fazer projeções” [BRACKETT, 1996]. Neste momento do nosso trabalho, queremos destacar esse outro ponto importante, que trata do tempo de permanência dos dados temporizados no DW. “Um horizonte de 60 a 90 dias é comum para sistemas operacionais, contudo, em se tratando de DW, o normal é de cinco a dez anos” [INMON, 1996].

A estimativa acima nos leva a refletir sobre essa importante, porém difícil questão a ser tratada, que foge do escopo da ciência da Computação, indo ao encontro da ciência da Administração. A questão de quanto tempo é preciso para manter os dados disponíveis no DW de forma que eles sejam efetivamente úteis às alçadas gerenciais estratégicas da organização, não é tão fácil de ser respondida. Por exemplo, no ambiente de uma instituição financeira do tipo banco comercial encontramos necessidades de armazenar o tempo explicitamente, em diferentes contextos, conforme requer este ramo de negócio, sempre obedecendo a normas legais ditadas pelo sistema financeiro do país. Neste mesmo ramo, porém mudando de uma economia de um país em desenvolvimento, como o Brasil, para um país de economia estável como os Estados Unidos da América, podemos constatar uma grande diferença.

No primeiro, o mercado é dinâmico, existem muitos índices econômicos e financeiros (tais como IPC—Índice de Preços ao Consumidor, etc.) que refletem a sua dependência mundial e induz as empresas a uma administração miope, ou seja, direcionada ao curto prazo, sendo esse prazo ainda mais encurtado em épocas de alta inflação (como ocorreu no final da década de 80 e início de 90), quando os padrões de consumo das classes dos cidadãos sofrem mudanças ainda mais radicais. Já numa economia estável, não são identificadas grandes e significativas mudanças.

Como sabemos, a história é cíclica. O que nos leva a refletir que os dados passados e ditos ultrapassados podem também ser aproveitados e analisados de maneira a extrair informações que venham a trazer melhor preparo internamente às empresas para enfrentar mudanças futuras, através do conhecimento do seu comportamento, enquanto empresa, nos vários períodos cíclicos e/ou críticos. O grau de confiança passado por qualquer sistema EIS/DSS está diretamente ligado à questão temporal, em relação ao período de armazenamento dos dados, ou seja, a precisão no processamento desses sistemas, em termos corporativos históricos, não será maior que o seu escopo temporal de armazenamento de um período histórico qualquer focado. Contudo, o período ideal de armazenamento dos dados em uma determinada empresa de modo a torná-los disponíveis no DW é uma questão multidisciplinar dos âmbitos das ciências da Computação e da Administração, conforme dito antes.

Com a abordagem temporal compreendida de maneira adequada, teremos então um DW mais enxuto e com melhor desempenho. Porém vale ressaltar que tanto o armazenamento, quanto as operações temporais exigem de um SGBD funcionalidades temporais (indexação temporal, linguagem temporal, entre outras). Infelizmente, comercialmente SGBD temporais ainda estão na infância. Isso constitui um dos pontos nevrálgicos da prática do DW, a falta não apenas de um BD preparado para lidar com tempo de maneira adequada, mais também do preparo do corpo técnico para manipular com esses aspectos eficientemente (tais como lógica temporal, informações incompletas). Vamos sair um pouco do âmago do armazenamento do DW para abordar, não mais *o como* armazenar mais *o quando e de onde* carregá-lo e atualizá-lo, ou seja, veremos agora os importantes metadados, responsáveis pelo correto armazenamento dos dados.

#### 4.3.6. Os Metadados

Como já foi dito acima, os dados são armazenados mediante a orientação dada pelos metadados. “Metadados são dados sobre dados” [INMON, 1996; BERSON & SMITH, 1997, GRAY & WATSON, 1998]. [INMON, 1996] ainda acrescenta que, apesar deles não

serem coisa nova, a produtividade na utilização do DW está ligada aos metadados, ressaltando que tipicamente eles mantêm as seguintes informações:

- a **estrutura dos dados** e as suas **rotinas de acesso**, segundo as diferentes visões técnicas do programador e do analista da área estratégica;
- a **origem** e a **transformação** sofrida pelos dados que ingressaram no DW;
- o **modelo de dados** e o seu **relacionamento** com o DW e
- o **histórico da extração dos dados**.

As **estrutura dos dados** e **rotinas de acesso** são vistas de duas formas diferentes pelos corpos técnicos e analistas, pois, para os primeiros – que constróem e mantêm o DW –, as informações que lhes são importantes relacionam-se com aspectos do bom armazenamento e a boa disponibilidade dos dados, enquanto, para os analistas – aqueles que tem o DW como a grande fonte de dados para as suas ferramentas de produtividade – o interesse está em saber quais, quando, como e quão recente os dados estão no DW. Desta forma são realmente necessárias duas visões destes importantes aspectos. Já a **origem** dos dados – indica a fonte do SA – e a **transformação** – possível conversão ou filtragens, etc. – sofrida por estes ao ingressarem no DW, que servem à gerência do mapeamento dos dados operacionais no DW, assim como o **modelo de dados** e o seu **relacionamento** com o DW, são basicamente informações técnicas e fundamentais nos momentos de adequação a novos dados correlatos e nas eventuais (as vezes um pouco frequentes) manutenções corretivas. Já o **histórico da extração dos dados** é o controle de acesso dos dados no DW. Ele também se faz essencial na busca da precisão da atualização, que será refletida nos sistemas EIS/DSS, através da repercussão na validade das informações derivadas dos dados.

Para esclarecer melhor esses pontos de forma a torná-los mais palpáveis, imaginemos uma informação cadastral de um determinado cliente de um banco qualquer. Descrevendo um pouco o perfil deste cliente, ele é possuidor de duas contas, uma tipo conta corrente e outra tipo conta poupança. Antes de ser correntista<sup>40</sup>, o mesmo era

---

<sup>40</sup>Correntista é o nome utilizado nos bancos para referenciar o cliente que possui conta corrente.

poupador<sup>41</sup>. Nesta instituição bancária hipotética o sistema de contas correntes era mais antigo que o sistema de poupança, pois a instituição comprou um sistema mais moderno ultimamente. Por essa razão, o sistema de contas correntes possui o tamanho de campo para o nome do cliente menor que os sistemas mais novos, conforme exigia os sistemas mais antigos com o objetivo de economia de memória e aumento de desempenho de alguns sistemas. No contas correntes o nome tem 20 posições já e na poupança tem 30 posições. A Figura 4.7 nos mostra um exemplo do registro de um cliente de nome *Alexandre Peixoto Guedes Neto* inicialmente no sistema de contas correntes e posteriormente (quatro anos mais tarde) no sistema de poupança.

0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 2	<i>Sistema de Contas Correntes</i> (max. de 20 caracteres)
1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0	
ALEXANDRE P. G. NETO	
<hr/>	
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 3	<i>Sistema de Contas Poupança</i> (max. de 30 caracteres)
1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0	
ALEXANDRE PEIXOTO GUEDES NETO	
Cadastramento do nome do cliente <b>ALEXANDRE PEIXOTO GUEDES NETO</b>	

Figura 4.7 - Exemplo do Registro com Diferentes Tamanhos do Campo  
NOME DO CLIENTE entre Dois Sistemas Bancários

Suponhamos que queiramos registrar essa informação sobre o cliente no nosso DW na época em que o dito cujo só possuía conta corrente. No nosso DW hipotético foi definido um tamanho de campo para o nome do cliente com 40 posições alfanuméricas. Inicialmente as informações do metadado NOME DO CLIENTE no DW eram derivadas do campo NOME DO CLIENTE do contas correntes. Posteriormente elas serão atualizadas com as referentes ao NOME DO CLIENTE do conta poupança, pois, em termos

<sup>41</sup>Poupador é o nome utilizado nos bancos para referenciar o cliente que possui conta poupança.



semânticos, o NOME DO CLIENTE no sistema de poupança é mais rico na sua quantidade de posições para representá-lo.

Assim, todas as informações devem ser registradas, pois há necessidade de haver uma homogeneidade consistente dos campos ao longo do tempo. Neste ponto sugerimos o emprego dos dois tempos (TS e TE) aplicados aos metadados<sup>42</sup> para melhor podermos responder aos pré-requisitos fundamentais do DW: integridade e consistência. Infelizmente isso não está claro nas bibliografias consultadas. Acreditamos que os metadados que estão sendo concebidos teoricamente e criados na prática não estão preparados para contemplar mudanças simples como essa que exemplificamos na busca do *melhor dado*, pois não é mencionado nada sobre a temporalidade dos metadados, que acabamos de abordar. Caso isso não seja considerado, teremos DW limitados e atravancados pela própria concepção do metadado. Por mais essa razão, o uso da temporalidade no metadado, que defendemos enfaticamente o uso da modelagem temporal que utiliza dois tempos que delimitam a validade dos dados (TS e TE) no DWing.

[GRAY & WATSON, 1998] alertam que na prática criar os metadados é uma das tarefas mais difíceis no DWing. Evidente que se trata de uma tarefa delicada, uma vez que qualquer falha na sua definição implicará diretamente sobre os dados do DW e finalizará afetando todos os sistemas que se baseiem no DW. Porém, a dificuldade na criação dos metadados que esses autores falam é agravada na medida em que se atribui ao DW funções de OLAP, como eles sugerem que seja. O exemplo disso está na subdivisão que é feita nos metadados, de forma a atender a funções EIS. Neste caso, as necessidades dos metadados OLAP são radicalmente diferente das necessidades do DW. Metadados puramente voltados para o DW são caracterizados exclusivamente por comportar informações técnicas, conforme já discriminadas.

Os metadados DW diferem dos metadados OLAP por razões intrínsecas funcionais<sup>43</sup>. Os metadados DW são substancialmente técnicos, de forma a atender a seu

---

<sup>42</sup>Corresponde ao versionamento de esquemas nos BDT.

<sup>43</sup>A diferença das funções puramente DW e OLAP, já foi por nós discutida tanto neste capítulo, na seção 4.3.4, quanto na seção 3.7 do capítulo 3 sobre OLAP.

compromisso como DW, ou seja, temático, voltado à gerência estratégica no suprimento de dados para ferramentas de produtividade, provendo os mais diversos dados organizacionais temporizados. Já os metadados OLAP<sup>44</sup> devem conter informações baseadas no negócio específico a que se destina, de forma a atender as diretivas dos usuários analistas estratégicos. Sairemos agora do centro do DWing, o DW, de dados, metadados e suas particularidades no DWing, para abordar outro componente do DWing que vem sendo muito referido na prática, o ODS, conforme dito anteriormente na seção 4.2.

#### **4.4. Operational Data Store (ODS)**

Na seção 4.2 introduzimos o Operational Data Store (ODS), caracterizando-o como um repositório volátil, não temporal, basicamente com a finalidade de integração corporativa de várias fontes de dados operacionais não integrados e não sendo necessariamente um BD. Contudo, apesar de volátil, não temporal e não necessitar de um BD, na prática é constatada a necessidade dele possuir um controle temporal simples, baseado em um selo de tempo (timestamp), podendo ser ABSTIME (vide Figura 4.2), de forma a identificar a movimentação dos dados no(s) sistema(s). Desta forma, ele fica caracterizado como um arquivo de LOG de um banco de dados qualquer, que registra as transações ocorridas nas bases de dados. A diferença é que, no caso do ODS serão registrados, após consolidação inter-sistêmica (para o caso de mais de uma mídia), os *melhores dados*, obtidos na movimentação gerada e consolidadas dos SA.

---

<sup>44</sup>Metadados OLAP são basicamente formados por três componentes: diretório técnico, diretório de negócios e navegador de informações. Este último é responsável pelo encaminhamento das operações OLAP, vide seção 3.3 no capítulo 3.

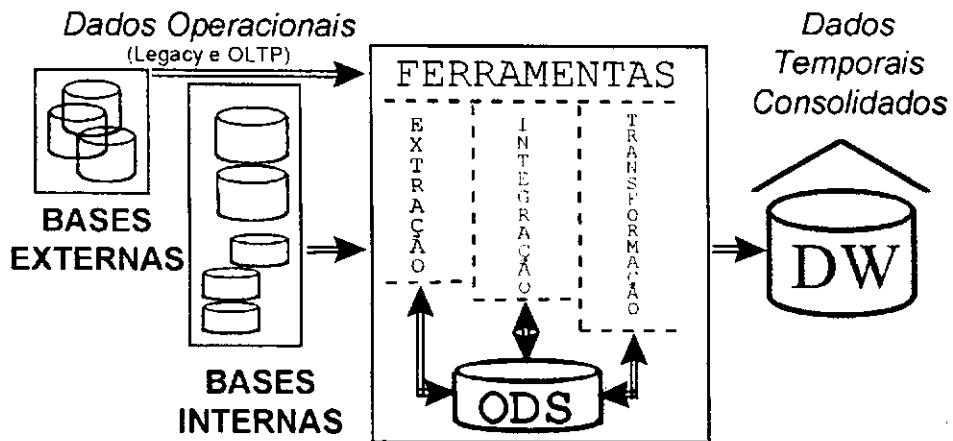


Figura 4.8 - Posição de um ODS no Diagrama Básico de DWing

A Figura 4.8 nos mostra a posição relativa do ODS, com base no diagrama mostrado anteriormente na Figura 4.1. A prática do DWing tem evidenciado, para grandes e complexas instalações, que as *ferramentas de extração, integração e transformação* necessitam de um ou mais repositórios voláteis para viabilizarem as suas finalidades entre dois ou mais sistemas (ou mídias) que possuam redundância e/ou dependência dos dados na alimentação do DW, assim nos mostra a Figura 4.9.

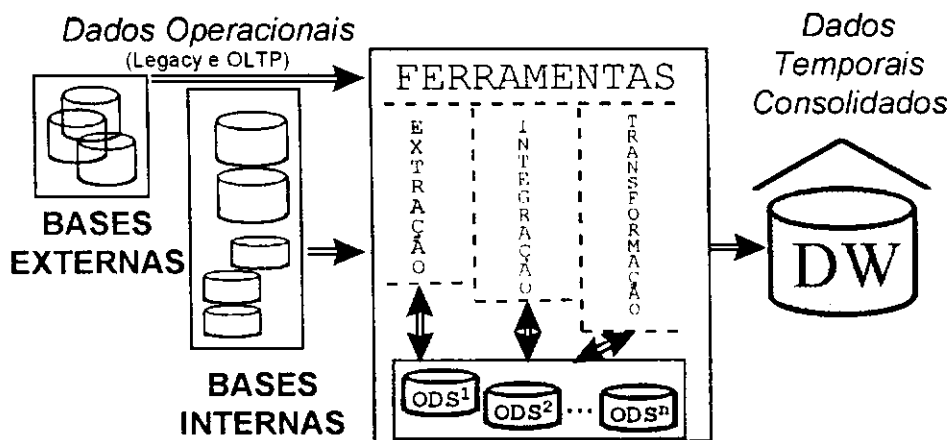


Figura 4.9 - Utilização de Vários ODS no Diagrama Básico de DWing

O ODS está sendo cada vez mais usado tanto como um meio de facilitação da preparação do DW a partir das diversas, heterogêneas e dispersas fontes de dados em uma organização, quanto como um meio de recepção de dados de clientes e parceiros comerciais a fim de permanecerem no mercado competitivo globalizado. Neste particular a INTERNET está sendo bastante empregada, na comunicação entre empresas, por ser um meio barato e ágil. Abordaremos mais adiante essa tecnologia emergente na seção 4.6.3. Através de listas de discussão<sup>45</sup> na INTERNET, a questão do ODS tem sido constantemente referenciada e muitas são as idéias expostas, desde a do usuário acessar diretamente o ODS, o que não consideramos uma proposta válida – pois o ODS (como o próprio nome já diz) não foi elaborado para ser consultado por usuários estrategistas – até a sua própria eliminação da arquitetura, em ambientes apenas com sistemas OLTP (sem *legacy*). Na prática, as grandes corporações possuem diferentes formatos e meios magnéticos (fitas, discos removíveis, cartuchos, disquetes, etc.), que justifica o uso do ODS.

Algumas questões derivadas do estudo do ODS no DW ainda trazem a tona alguns pontos relevantes, que nos sugere aprofundar-nos um pouco mais. Uma delas, por exemplo, é, justamente pelo ODS ter uma semelhança funcional com o arquivo de LOG de um BD no que diz respeito a mapear as alterações ocorridas em uma BD qualquer, os arquivos de LOG serem encarados como DW.

Sobre essa questão, tomamos, para clarear, a afirmação de [SULAIMAN & SOUZA, 1998], nos seus estudos sobre *Data Mining*, em que dizem “Os LOG dos BD são integrados. Uma vez que o LOG é a história das transações, a informação obtida é semanticamente relacionada à aplicação”. Com isso, intuitivamente pode-se pensar no uso indiscriminado do LOG de BD como um DW. Rigorosamente, apropriando-se do conceito teórico de DW cunhado por Inmon pode-se perceber facilmente essa diferença.

---

<sup>45</sup>Este autor é assinante das listas de DWing DWLIST ([dwlist@datawarehousing.com](mailto:dwlist@datawarehousing.com)), oriunda do “The Data Warehousing Group Institute”, vide <http://www.datawarehousing.com>, de onde observou várias discussões sobre ODS, inclusive participando ativamente, e DWLA ([ordonez@pe.net](mailto:ordonez@pe.net)), oriunda do Grupo Latino de DWing, vide <http://www.pe.net/~ordonez/dwla>.

Na prática, seria como possuir um único banco de dados em toda corporação. Mesmo assim não seria viável, pois o número de registros gerados no LOG teria que sofrer um pré processamento de seleção, uma vez que um campo qualquer de um BD pode sofrer várias modificações durante um dia, a exemplo do LOG de transações de um sistema de contas correntes em um banco comercial, que, ao final, geraria um “DW” altamente inflacionado de registros. O LOG de um sistema de contas correntes é acrescido de registros de movimentações de débitos e créditos, que na realidade não tem tanto valor estratégico significativo quanto em relação ao saldo da conta de um cliente, que é obtido no final do dia, diariamente, e não é normalmente gravado nele, de onde podemos obter informações mais relevantes sobre o próprio cliente. Além disso, sabemos que na prática as empresas possuem vários sistemas corporativos e com bases de dados dispersas, que geram LOG de transações independentemente.

Esses autores ainda pontuam, que “Os LOG são não voláteis. Eles têm que ser preservados para prevenir panes, mas também para serem auditados”. Esse aspecto temporal dos LOG é algo frágil quando aplicado ao estudo do DWing, pois, quando falamos em volatilidade em DW, estamos nos referindo desde meses a décadas. A depender do tipo de *Data Mining* que se venha a fazer, o uso do LOG pode, e segundo os autores, deve ser satisfatório. Concordamos e acreditamos ser uma forma interessante de se aproveitar uma estrutura de dados pertinente e integrante de qualquer BD para se realizar limitados e possivelmente úteis processamentos.

Vale destacar, contudo, que a solução ideal e de âmbito genérico a nível corporativo, e não apenas, tão somente para o *Data Mining*, é o uso de um DW. Por isso, ressaltamos que a sobreposição dos conceitos de LOG com DW, a rigor, não é correta. Seria como dizer que um conjunto de arquivos seqüenciais indexados referente a um sistema que possua internamente listas invertidas para o processamento, possa ser visto como um BD, devido equivalência na sua função “endógena” de armazenamento e recuperação dos dados.

Outra afirmativa a ser considerada é a de, possuindo-se um DW, se fazer acessos diretos aos arquivos dos SA para se obter dados mais recentes e utilizá-los nas *ferramentas*

*de produtividade*, uma idéia derivada da proposta de acesso direto aos ODS no ambiente de DWing. Não consideramos um proposta viável, pois os dados nos SA estão sendo constantemente modificados, o que não é algo interessante para as ferramentas de análise e síntese, obter uma posição não consolidada de um dado qualquer para processamento.

Vale ressaltar que [INMON ET AL, 1996], (talvez na primeira tentativa de definição formal do ODS) atribuem a ele – além das características acima citadas de um repositório volátil, não temporal, basicamente com a finalidade de integração corporativa de várias fontes de dados operacionais não integrados – as características de orientação por assunto e com alto grau de granularidade, ou seja, os dados com grande nível de detalhes. Concordamos com a questão do alto grau de granularidade dos dados, visto que, quanto mais completo e detalhado os dados possam ser corporativamente consolidados para alimentar o DW, mais preciso será esse repositório, e, por efeito cascata, teremos informações mais apuradas. Já a questão da orientação por assunto, acreditamos que não reflete a razão intrínseca da sua existência, ou seja, a viabilização técnica do DW para grandes e complexas corporações. “O ODS supre as necessidades do ambiente operacional, enquanto o DW supre as necessidades da comunidade da informação” [INMON ET AL, 1996]. Porém, conforme os referidos autores, eles enxergam esse repositório com valor tanto para o corpo técnico, quanto para os especialistas do negócio, razão pela qual ele o concebe desta forma, orientado a assuntos.

Os autores [INMON ET AL, 1996] afirmam ainda que o ODS é opcional, porém destacando que em função do tamanho da organização, da necessidade imediata de informações e da organização do seu parque informacional, existe ou justifica a sua adoção em um projeto de DWing. Para eles, a interseção dos fatores: organização grande, com necessidade de obtenção de informações tão breve quanto possível, assim que os dados sejam produzidos e a existência de parque computacional desorganizado, indicam o uso do ODS.

Consideramos que afirmar forçosamente que apenas a interseção desses três fatores justifica o emprego do ODS é demais exigente em termos de argumentação para a sua indicação numa arquitetura de DWing. Tanto o aspecto *tamanho avantajado da*

*corporação*, quanto a *desorganização do parque computacional operacional* por si só, já sugerem e aconselham na prática o ODS. Aliás, como não podemos prever nem o seu crescimento, nem a velocidade em que o mesmo possa ocorrer numa organização, acreditamos que um projeto de DWing, já deva iniciar com pelo menos um ODS para facilitar a integração potencial dos SA de forma a não afetar os sistemas do nível estratégico, no processo de escalabilidade sistêmica empresarial. Quanto à necessidade urgente do uso dos dados operacionais de forma a gerar informações estratégicas, acreditamos que esta seja um tendência natural, pois é sabido que qualquer usuário de informática tende a “apegar-se” e “depende” progressivamente dessa ferramenta, seja ela qual for e para qualquer fim a que se destine. Logo, acreditamos que temos que tratar esse fator como um “axioma” e já nos posicionarmos contando com essa necessidade, que se não aparecer imediatamente, a curto ou médio prazo, certamente, manifestar-se-á.

São definidas três classes de ODS, por [INMON ET AL, 1996], em função do sincronismo como a atualização do ambiente operacional. As classes I e II tem atualização síncrona, sendo uma *on-line* (com de dois a 3 segundos de diferença em relação à transação original) e a outra feita periodicamente (de meia em meia hora ou de uma em uma hora), respectivamente. Na classe III a atualização é assíncrona, ficando em função normalmente apropriada a períodos diários, ou seja de 24 em 24 horas. Notamos aqui, claramente, uma tentativa de contemplar todas as possibilidades de arquiteturas de DWing, de tal forma a facilitar oportunamente a alimentação do DW, isto é, DW com funções OLAP precisam forçosamente definir o tempo, de forma cronometrada, a sua atualização, enquanto banco de dados.

Neste ponto do nosso trabalho encontramos a necessidade de sintetizar as diversas arquiteturas baseadas em DW, tanto teóricas quanto práticas, independentemente do uso do ODS, para podermos visualizar melhor a questão da implementação de um ambiente de DWing. Para viabilizar essa reunião em grupos semelhantes utilizamos um critério classificatório comum no estudo arquitetural cliente-servidor, ou seja, o conceito de camadas arquiteturais, conforme veremos a seguir.

#### 4.5. Arquiteturas de DWing

No nosso trabalho de pesquisa, notamos que para muitos a diferença entre arquitetura e infra-estrutura não é tão clara. Conforme Hildebrand (1995) apud [GRAY & WATSON, 1998]: a arquitetura define as funções e responsabilidades de cada componente de uma aplicação, enquanto a infra-estrutura especifica o hardware e software para implementar esses componentes. Isto se deve, principalmente, devido à grande competitividade existente entre diversos fornecedores, tanto de software, quanto de hardware, que agora se fazem unidos em prol de chegarem mais rapidamente a um resultado satisfatório, para a comercialização de uma solução corporativa de DWing eficiente e eficaz, de forma a contemplar as necessidades exclusivas e particulares dos EIS e DSS.

Muito do que foi elucubrado e implementado até o momento sobre arquiteturas com DW e suas principais variações, compilamos e apresentamos a seguir. Contudo, ressaltando que, apesar dos diversos modelos existentes e como já dissemos anteriormente, é possível, uma perfeita compreensão do funcionamento, funções e responsabilidades de cada um dos componentes arquiteturais e suas dinâmicas, simplesmente a partir do diagrama básico já visto na seção 4.2, onde apresentamos uma arquitetura básica com a finalidade de elucidação, retratada na Figura 4.1 (e aqui classificada como DW padrão, conforme veremos na seção 4.5.2.1).

Para classificarmos as arquiteturas de DWing utilizamos a abordagem nativa do DWing, ou seja, o ambiente cliente/servidor. “O termo cliente/servidor originalmente foi aplicado à arquitetura de software que descrevia o processamento entre dois programas: uma aplicação e um serviço de suporte” [BERSON & SMITH, 1997]. “O DW é geralmente implementado em um ambiente cliente/servidor” [GRAY & WATSON, 1998]. Conforme [BERSON & SMITH, 1997], “o modelo computacional cliente/servidor cobre uma larga extensão de funções, serviços e outros aspectos do ambiente distribuído”. Desta forma, classificando as possíveis arquiteturas segundo o critério de camadas usado no estudo



cliente-servidor, pois contempla de forma adequada os aspectos técnicos mencionados ao longo deste capítulo e requeridos no DWing, chegamos a três tipos fundamentais de arquiteturas.

Não que essa seja a única forma de classificarmos os ambientes arquiteturais de DWing, pois vários critérios para analisá-los vem sendo utilizados, por alguns autores, tais como a abrangência do DW em relação ao âmbito corporativo, a paridade DW versus setores empresariais e outros tantos. Como a questão do tamanho do porte corporativo ser algo difícil de se mensurar, podemos incorrer em erros do tipo uma arquitetura ser vista e enquadrada de duas formas diferentes, em função do ponto de vista do analista que avaliou o tamanho da empresa para definir o tipo de arquitetura DW. O que ocorre é que autores diferentes classificam distintamente arquiteturas segundo diferentes (e até divergentes) pontos de vista, quando envolvemos o porte do negócio como referencial de classificação.

Veremos agora o que cada camada significa em cada um dos contextos. A primeira camada é sempre a interface com o usuário, isto é, trata-se do local onde o usuário visualiza o resultado das suas consultas e requisições ao ambiente OLAP e/ou OLSP. Nesta camada podemos ter clientes gordos (*fat clients*) ou magros (*thin clients*), conforme abordado no capítulo 3, seções 3.7.1 e 3.7.2, respectivamente.

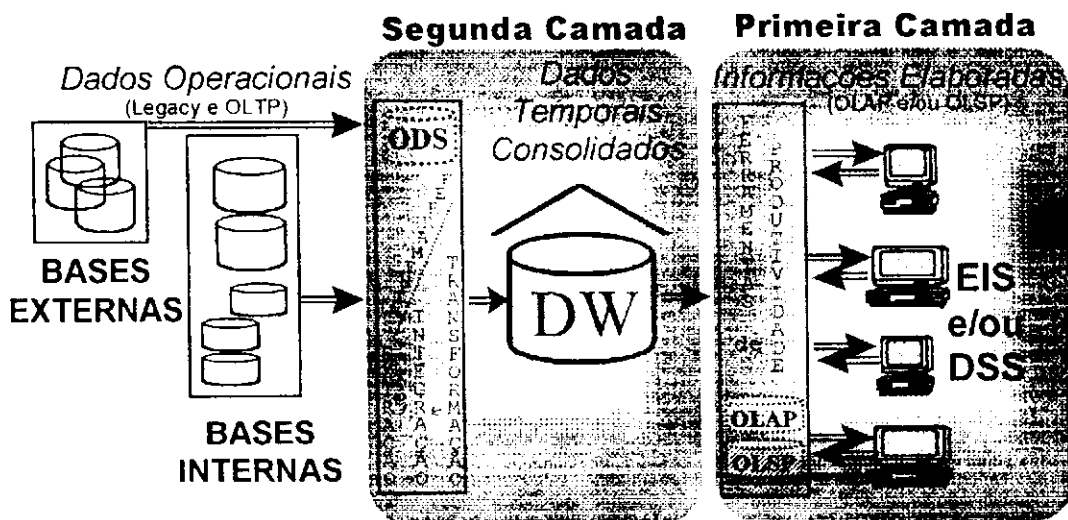


Figura 4.10 - Diagrama Básico de DWing Visto sob a Ótica de Camadas

Conforme a Figura 4.10, que nos mostra o diagrama básico da seção 4.2, agora dividido em camadas, onde separamos a primeira da segunda, onde observamos também que, como os dados operacionais constituem a entrada da nossa arquitetura, pois é algo previamente existente (podendo variar desde uma fita, até um complexo ambiente computacional integrado) não se atribui a ela o status de camada.

A primeira camada do DWing representa imutavelmente as interfaces e ferramentas de produtividade voltadas para o usuário. A partir daí veremos as outras possíveis arquiteturas e o que cada uma delas representa em função do número de camadas. Nesse caso específico (o da nossa arquitetura básica), a segunda camada é a camada do servidor de dados corporativos consolidados temporais, ou seja, a camada do DW. Iniciaremos a descrição dos modelos arquiteturais pela arquitetura de uma camada (do inglês “one-tier architecture”).

#### 4.5.1. Camada Única (*One-tier architecture*)

Na arquitetura de camada única não existe o repositório de DW, apenas a camada cliente, ou seja, há um compartilhamento do ambiente dos SA, de forma a acomodar também os softwares estratégicos. Podemos encontrar (se é que é possível!) duas versões dessa tipo de arquitetura. São elas: o “stand-alone one-tier architecture” e o DW virtual. Vale salientar que, conforme já discutimos no capítulo 2, precisamente nas seções 2.3 e 2.3.1, onde analisamos a questão detalhadamente e citamos um estudo de caso, respectivamente, essa arquitetura é impraticável para grandes corporações. Contudo, como há referências bibliográficas, vamos a elas.

#### 4.5.1.1. Camada Única Autônoma (*Stand-alone One-tier Architecture*)

Nesse tipo de arquitetura de camada única autônoma (do inglês, *Stand-alone One-tier Architecture*), os ambientes para EIS e DSS são implementados na mesma máquina em que os SA. Conforme [GRAY & WATSON, 1998], todas as funções são implementadas em uma única máquina física. É quando existe um limitado volume de dados e o número de usuários é pequeno. A vantagem é o baixo custo de implementação do ambiente especial para SG, porém, logo se sentirá as conseqüências dessa economia: a queda de desempenho da máquina operacional, isto é, se for realmente possível adaptá-la a conviver com dois ambientes computacionais completamente diferentes, conforme discutimos esse mérito no capítulo 2, seção 2.3.

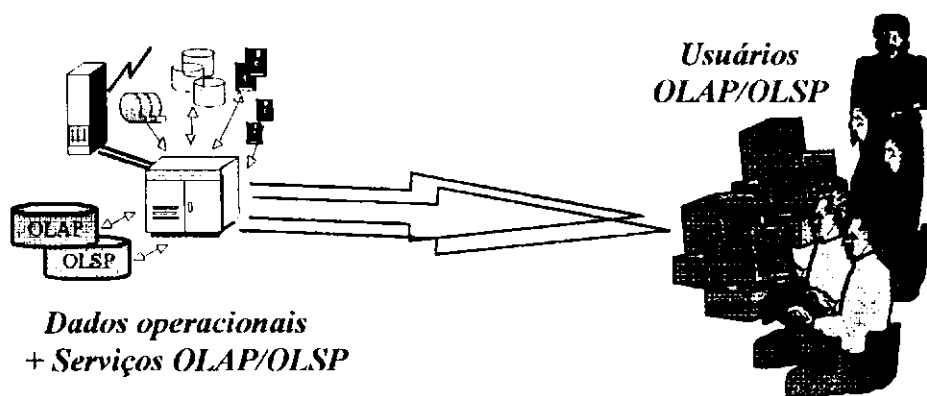


Figura 4.11 - Arquitetura de Camada Autônoma Única

#### 4.5.1.2. DW Virtual

Nesta arquitetura, conforme a Figura 2.12, não existe fisicamente o repositório onde residem os dados que serão a matéria prima para a consulta das ferramentas de produtividade. O DW é virtual [CAMPOS, 1996], ou seja, os dados necessários aos recipientes de armazenamento das ferramentas de produtividade OLAP/OLSP são carregados a partir de uma camada de middleware [MANZONI JR, 1996]. Não há o armazenamento temporal consolidado corporativamente, pois não existe DW físico, e sim

uma visão emulada de DW. Trata-se de uma arquitetura fundamentalmente teórica, pois, para ser implementada necessita de uma camada de middleware extremamente poderosa<sup>46</sup>, além do pré-requisito fundamental dos dados operacionais serem e/ou estarem necessariamente armazenados em bases de dados temporais, o que não reflete hoje a realidade das organizações independentemente dos seus portes.

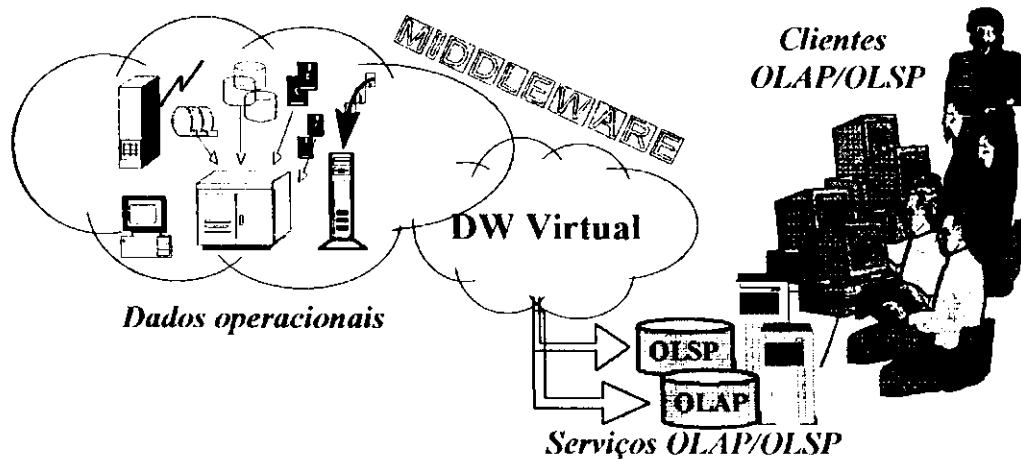


Figura 4.12 - Arquitetura de Camada Única – DW Virtual

#### 4.5.2. Duas Camadas (*Two-tier Architecture*)

Tratando-se da forma mais popular de arquitetura, atualmente a arquitetura de duas camadas pode ser vista e/ou abordada segundo três diferentes enfoques. Veremos inicialmente a arquitetura que chamamos de padrão. Aquele na qual está englobada o diagrama básico de DWing (Figura 4.1), o qual viemos até o momento nos utilizando.

<sup>46</sup>Apesar de não ter o propósito de ser utilizada com DWing, a Universidade de Stanford tem evoluído no estudo de integração de informações entre várias fontes de dados denominadas por eles de “mediator”, que pode ser visto em <http://www-db.stanford.edu> [WIEDERHOLD, 1997].

#### 4.5.2.1. DW Padrão

Essa arquitetura reflete a imagem genérica que temos de DW, estando cada componente devidamente modularizado e contemplando a possibilidade de escalabilidade corporativa tanto da área operacional, quanto da área estratégica empresarial, isto é, a possibilidade de crescimento organizacional tanto dos SA, quanto das ferramentas de EIS e DSS, de forma independente e que retrata a agilidade requerida pelas empresas da atualidade, como mostra a Figura 4.13.

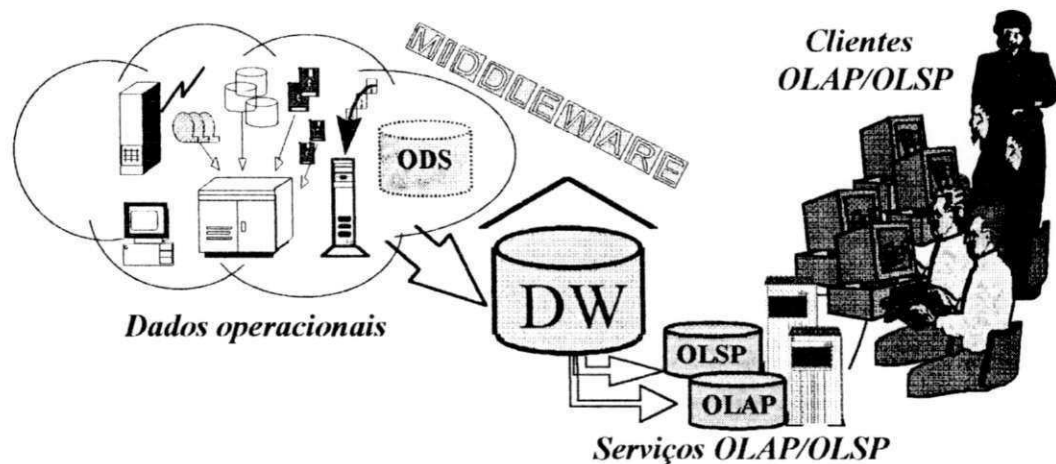


Figura 4.13 - Arquitetura de Duas Camadas – DW Padrão

#### 4.5.2.2. DW Dimensional

A arquitetura DW Dimensional (DDW, do inglês, *Dimensional Data Warehouse*) objetiva trazer para o repositório DW funções OLAP e OLSP, ou seja, é a idéia de termos em um único repositório, não só a missão de armazenar os dados – função legítima do DW – mas também reorganizá-los de forma a viabilizar tanto a análise multidimensional, quanto o descobrimento do conhecimento: *Data Mining*. Conforme dito na seção 4.3.4 sobre armazenamento, que é a principal funcionalidade do DW, essa abordagem de repositório multifunção não nos é muito simpática, pois, conforme visto no capítulo 2,

onde abordamos EIS/OLAP (seção 2.2.2.2) e DSS/OLSP (seção 2.2.2.3), eles possuem características distintas.

Devemos advertir ao leitor que, ao se optar por incluir funções OLAP e/ou OLSP no DW, estamos transformando o repositório que deveria refletir a realidade dinâmica do ambiente operacional historicamente, em um depósito de dados estático e restrito, quanto à sua atualização, devido às características singulares inerentes a esses tipos de processamentos. O que resulta em um repositório menos ágil, limitado pelas consultas dos usuários e conseqüentemente, menos atualizado. O mesmo ocorre na resposta das ferramentas de produtividade que, tendo um repositório único multifunção, resulta numa gama mais superficial de resultados enquanto sistemas EIS e/ou DSS, isto é, as respostas desses sistemas tendem a ter uma amplitude de complexidade menor, uma vez que não existe a dedicação exclusiva do repositório DDW para um processamento OLAP e/ou OLSP.

Existem ainda outros agravantes dessa reunião de diferentes funções no DDW, como, por exemplo, a necessidade de atualizar a função de armazenamento (DW) e estarmos processando uma análise multidimensional (função OLAP, ou então um processamento de *Mining* - OLSP -, indiferentemente) e algum outro usuário esteja inconformado precisando urgentemente fazer um outro processamento qualquer. Outro ponto fundamental a ser discutido é a se a modelagem sugerida pelo autor deste modelo arquitetural – [KIMBALL, 1996] – o *Star Schema*, é apropriado para processamentos OLSP

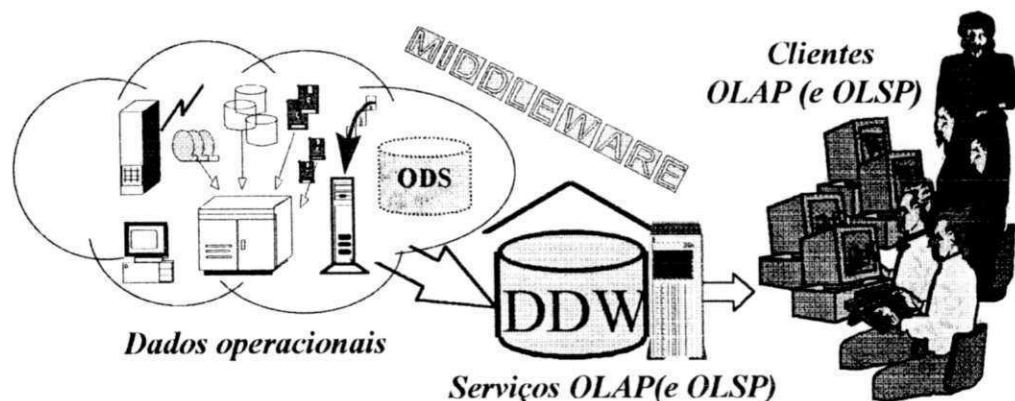


Figura 4.14 - Arquitetura de Duas Camadas – DW Dimensional

complexos. Acreditamos que não seja, pois a preocupação da modelagem citada é exclusivamente com a navegação nos dados, enquanto que o OLSP requer a manipulação matemática dos dados.

É interessante avaliar essas questões antes de se iniciar um projeto de DW, pois, apesar dessa arquitetura não possuir grandes dificuldades para se adaptar à tecnologia de armazenamento mais utilizada da atualidade, os DBR, ela impõe algumas limitações como as já expostas. Acreditamos que, para aqueles que implementam essa arquitetura, como mostrada na Figura 4.14, utilizem-na apenas para consultas OLAP variadas, porém limitando-se a mínimas informações de teor DSS (vide relatório do autor na Figura 3.7, na seção 3.4, do capítulo 3). Caso venham futuramente a optar por fazer processamentos OLSP complexos, adquirirão uma ferramenta DSS que incorpore um repositório OLSP (e conseqüentemente outro servidor, agora específico para o OLSP) de forma a atender às

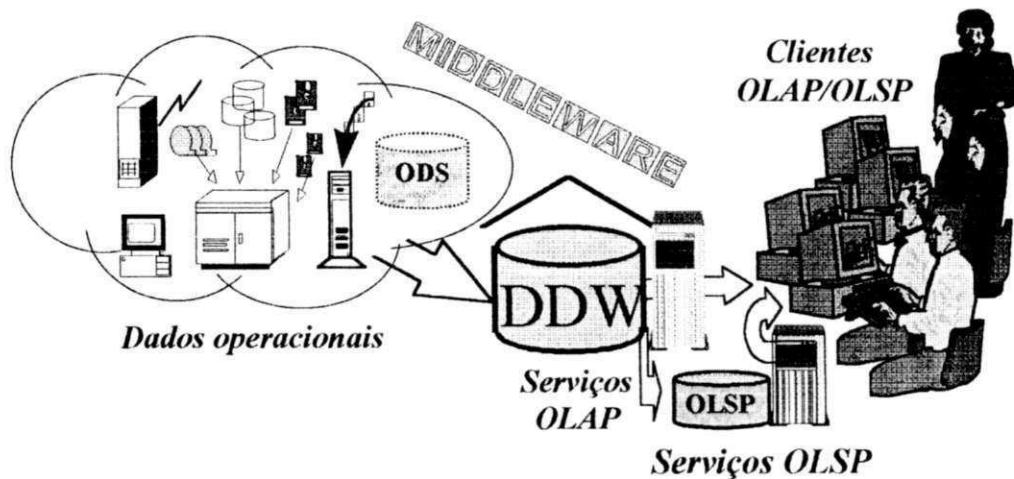


Figura 4.15 - Arquitetura de Duas Camadas – DDW com OLSP Externo

novas solicitações, conforme a Figura 4.15. Porém, a lentidão da atualização do repositório DDW, devido à sua modelagem voltada ao OLAP, gera impactos não apenas no DDW, mas também do novo repositório OLSP, que será atualizado via um repositório OLAP, repercutindo então no desempenho efetivo do projeto.

### 4.5.2.3. DW Federativo

O DW federativo [CAMPOS, 1996] constitui uma solução distribuída de repositórios do tipo *Data Mart* (DM), que são subconjuntos do DW, em que a visão do DW é dada através da “emulação de uma única visão lógica dos vários DM da empresa”. A grande vantagem desta arquitetura é a sua implementação gradual, prevista desde a sua concepção. Contudo, ficamos bastante dependentes do middleware envolvido para viabilizar o DW.

Vale salientar que um DW incompleto algumas vezes é confundido com um DM. Contudo, é interessante atentar para o fato que o DM deriva sempre do DW. Como conceitualmente o DM é um subconjunto do DW, é relevante que os DM independentes devem ter no seu projeto a particularidade do aspecto conceitual do modelo DW Federativo, conforme a Figura 4.16, bastante claro, pois, poderemos incorrer em erros tais

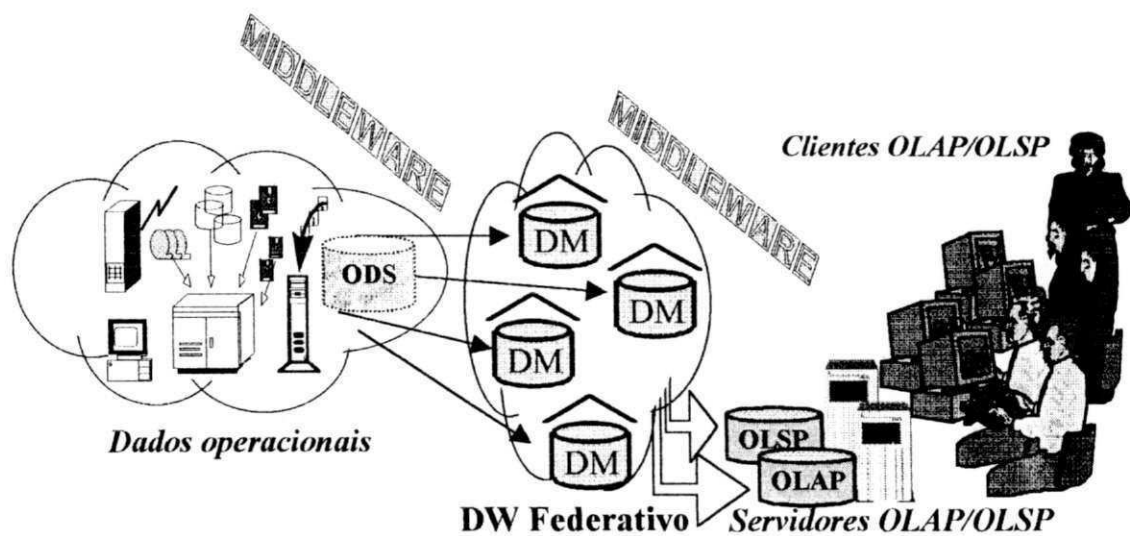


Figura 4.16 - Arquitetura de Duas Camadas – DW Federativo

como o de tratar qualquer um dos DM como um DW Padrão e acabar por comprometer conceitualmente o projeto em tempo de implementação e uso, resultando em provável geração de redundância de dados nos diversos repositórios DM, que finalizará com a possível inconsistência dos dados e a não viabilização da visão emulada e consistente do DW.



## 4.5.3. Três Camadas (Three-tier Architecture)

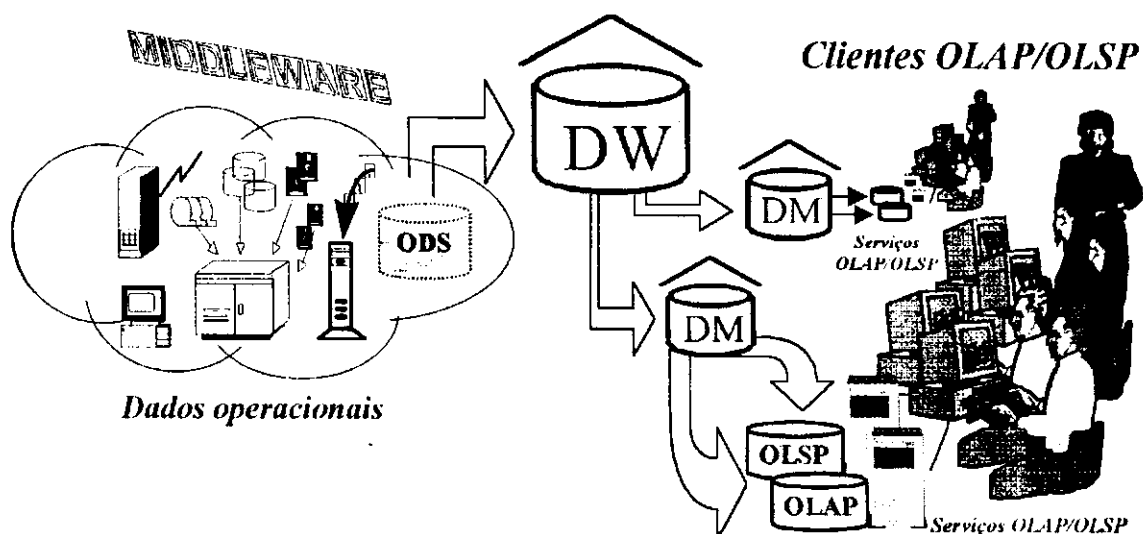


Figura 4.17 - Arquitetura de Três Camadas

Na arquitetura de três camadas, Figura 4.17, há um hardware intermediário, que constitui uma camada a mais, entre a camada cliente e a camada do DW vista na de duas camadas. Essa nova camada compõe-se de um ou mais repositórios do tipo DM. Essa nova camada atua principalmente de forma a influenciar positivamente todo o ambiente, através

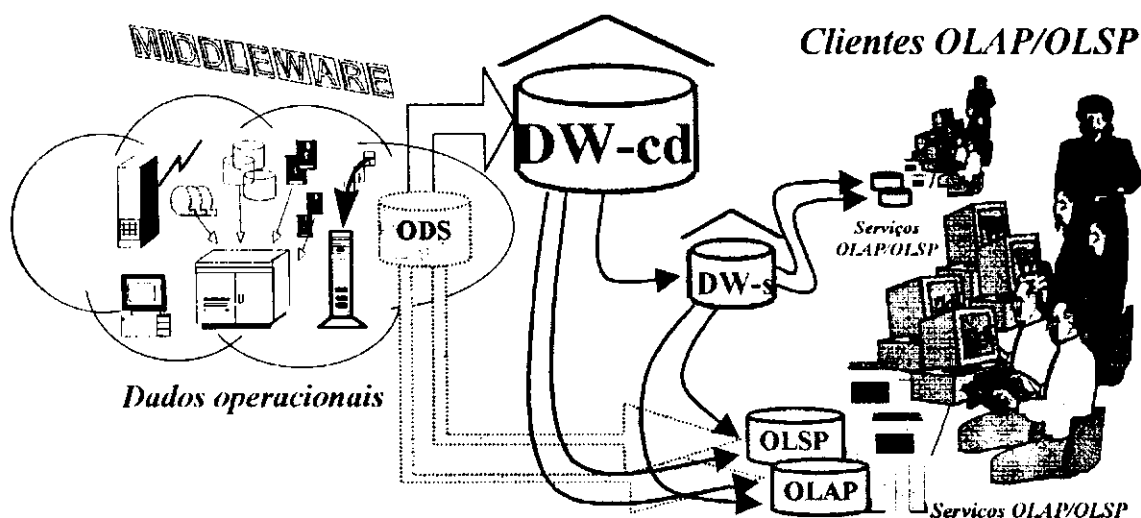


Figura 4.18 - Arquitetura CIF [INMON ET AL., 1996]

de tornar independentes a obtenção dos dados do DW pelos ambientes EIS e DSS, de forma que os processamentos OLAP e OLSP possam ser feitos em várias áreas da organização de maneira auto-suficiente departamentalmente.

Vale a pena salientar a existência de um modelo particular de arquitetura de três camadas chamada de CIF (Fábrica de Informações Corporativas, do inglês “Corporate Information Factory”) de [INMON ET AL., 1996], que difere conceitualmente dessa independência de DM departamentais. Nela existem um DW e um DM, chamados de DW-cd e DW-s, respectivamente. A Figura 4.18 exhibe-nos uma idéia global desta arquitetura. No detalhamento dos componentes da arquitetura CIF, sob a ótica temporal de armazenamento, temos:

- **ODS** – depósito de dados volátil, visando a carga futura dos seus dados para o DW. Os dados nele contidos têm até 24 horas;
- **DW-cd**<sup>47</sup> – depósito de dados temporal contendo informações detalhadas e com período de 24 horas a dois anos;
- **DW-s**<sup>48</sup> – depósito de dados temporal com a função de armazenar dados agregados consolidados, também conhecido como dados sumarizados<sup>49</sup>, ou seja, contém informações do DW (DW-cd) mais antigas e resumidas com mais de dois anos.

Como podemos observar nesse modelo, trata-se de uma arquitetura de três camadas onde a primeira camada pode ser clientes magros e/ou gordos, a segunda camada é a do DM – que nesse caso específico refere-se a um DM especializado (DW-s), ou seja, contém dados resumidos e consolidados temporalmente, que foram obtidos a partir do DW (DW-cd) – enquanto a terceira camada é a camada normal do DW. Aliás, essa arquitetura não se encaixa na de duas camadas, como alguns poderiam achar, pois, para se obter as informações para o DW-s, é necessário um processamento em cima dos dados obtidos no DW-cd. Sendo assim, é feito um pré-processamento no DW-cd para obter-se o DW-s, isto

<sup>47</sup>cd: *current detail*, detalhe corrente em inglês.

<sup>48</sup>s: *summary*, resumo em inglês.

<sup>49</sup>Sumarizado é um anglicismo que vem sendo cada vez mais usado nos meios do DWing.

é, são dados oriundos do DW-cd, que é genuinamente o DW e o DW-s um DM especializado.

Três observações devem ser feitas sobre a arquitetura CIF. Primeiramente, neste modelo a existência do ODS é opcional. Em segundo lugar, o tempo de “24 horas” a que se refere a CIF é o do período do dia, isto é, de 0 a 24 horas de relógio. Em terceiro lugar e não tendo sido mencionado antes, um outro componente desta arquitetura é o “BULK STORAGE”, que é um local onde informações antigas e pouco referenciadas são guardadas para consultas em meios menos caros de armazenamento eletrônico. Por se tratar de algo importante porém passivo nesta arquitetura, conforme [INMON ET AL., 1996], registramos aqui apenas a sua existência e possibilidade de uso.

Notamos também que existe uma preocupação singular na definição e organização dos repositórios sob a ótica temporal do armazenamento e manutenção dos dados. A definição de quais informações temporais serão migradas para o DW-cd e para o DW-s, e a que nível de granularidade a ser usado, serão de fundamental importância à sua construção.

Para clarear essa questão, imaginemos novamente um banco comercial. Para um gerente de contas correntes é importante saber o movimento diário dos correntistas, para verificar se eles ficaram negativos e, em caso afirmativo, em quanto. Porém, não menos relevante é conhecer o saldo médio<sup>50</sup> destes clientes para ajudar no critério de concessão de crédito. Já para o gerente de sucursal a atenção não está voltada nem para clientes individualmente, nem para períodos inferiores a vários meses. Para eles o que conta é o desempenho das agências a nível regional em determinados períodos não inferiores a trimestres. Aí temos casos de diferentes tipos de informações oriundas e consolidadas das agências a partir do movimento das contas dos clientes enfocadas em diversas granularidades de tempo.

Desta forma, notamos que como são necessárias informações operacionais resumidas para o controle gerencial, elas também sejam importantes a nível estratégico

---

<sup>50</sup>Geralmente o saldo médio é trimestral.

para as grandes empresas. Acreditamos que, a principal justificativa da necessidade de informações resumidas e temporais tenha sido o fundamento deste modelo, que consideramos muito interessante e aplicável de forma eficiente e eficaz.

Já que no exemplo citado anteriormente usamos a realidade bancária, devemos ressaltar que, nas grandes corporações tipo as do segmento bancário, confiamos que o ODS assuma um papel fundamental na integração corporativa. Sugerimos então, a utilização de mais de um ODS, justificado(s) pela possível e provável existência de diferentes plataformas operacionais e também pelos grandes volumes de transações geradas dos sistemas bancários. Acrescentando que, em relação ao tempo de armazenamento, a vida útil das suas informações no ODS também deve ser abstraída, em alguns momentos, do tempo “físico” de “24 horas” especificado por Inmon, ficando em função da frequência de atualização dos sistemas bancários.

Como atualmente a agilidade e a capacidade das organizações em adequar-se eficiente e eficazmente às necessidades de mudanças do mercado é algo imperativo para a própria sobrevivência das mesmas numa sociedade globalizada e altamente competitiva, acreditamos que a arquitetura de três camadas predominará e é, sem dúvida, aquela que melhor adapta-se a necessidades tais como escalabilidade, requerendo apenas a adaptação deste ambiente computacional aos novos requisitos do mercado e às características de novas ferramentas para SG de forma mais simples e mais rápida que a de duas camadas. Outros pontos vantajosos são: a maior independência de fornecedores de produtos voltados para o SG, a possibilidade de mais facilitada manutenção pelo seu carácter modular, a escalabilidade garantida e a agilidade em gerar departamentalmente um novo ambiente de SG, tendência natural observada nas empresas na administração descentralizada. Porém, certamente que esse modelo despense um maior custo, principalmente na sua implantação, sendo voltado para grandes organizações.

---

#### 4.6. Interação com Outras Tecnologias Emergentes

“Outro aspecto extremamente importante do DW é a capacidade de receber dados de e passar dados para uma ampla variedade de tecnologias. A tecnologia que suporta o DW praticamente não terá valor caso existam restrições severas ao recebimento e à transmissão de dados pelo DW” [INMON, 1996].

Com a idealização do DW imaginou-se prontamente as suas possíveis interações com outras tecnologias também emergentes, de forma a torná-las mais poderosas. Certamente a INTERNET consiste num grande veículo de disseminação de informações. Contudo, é importante e necessário atentar-se para a questão da crucial importância estratégica da natureza dos dados oriundos do DW, assim como das informações que deles possam advir. Inicialmente veremos o casamento do DW com OLAP.

##### 4.6.1. DW e OLAP

Em muitos momentos desse trabalho abordamos o processamento OLAP. Inicialmente no capítulo 2, seção 2.2.2.2, onde apresentamos o conceito de EIS e seu tipo de processamento, o OLAP, e fizemos uma ressalva da mistura conceitual, que alguns fazem sobre funções OLAP com as do DW. Posteriormente, dedicamos um capítulo integralmente a ele, o capítulo 3, no intuito de trazer à luz a questão do EIS/OLAP e suas funções particulares, necessidades e peculiaridades, enquanto um moderno SG. Contudo, não esgotamos o assunto apenas nessas referências, pois, como vimos posteriormente, na seção 4.5.2.2, há uma arquitetura de DW que prevê o armazenamento especialmente voltado para o OLAP. É interessante termos em mente a importância dessas tecnologias, e estarmos atentos das limitações que advirão ao ambiente de SG da organização ao optarmos pela junção em um único repositório dessas duas tecnologias complementares.

Pensando sob a ótica da convivência harmoniosa entre essas tecnologias, conforme [INMON, 1996], observa-se um relacionamento interessante e complementar entre o DW e

os SGBD multidimensionais. O DW contém dados no nível baixo, ou seja, no nível de detalhe, de forma a atender perfeitamente aos usuários de SGBD multidimensionais mais exigentes. Porém ele adverte que, na maioria dos casos, é um erro pensar em se utilizar SGBD multidimensionais para DW, reconhecendo ainda, que os SGBD multidimensionais são ótimos para aquilo a que se propõem<sup>51</sup>, a análise multidimensional, que não é a função de um DW. Ele ainda acrescenta que é um erro pensar que as tecnologias são mutuamente exclusivas. “SGBD multidimensional (OLAP) é um tecnologia e o DW é uma infraestrutura arquitetônica. Há um relacionamento complementar e simbiótico entre ambos” [INMON, 1996].

#### 4.6.2. DW e OLSP

Assim como o OLAP, o OLSP tem suas características particulares dos modernos SG. O DW relaciona-se também de forma complementar com essa tecnologia. Sendo que as duas maiores questões do OLSP são, sem dúvida, a necessidade dos dados serem essencialmente corporativos e a da incerteza geral que paira sobre a granularidade ideal que devemos utilizar no seu processamento, conforme relatado na seção 2.2.2.3 do capítulo 2. Desta forma, vemos que o DW contempla esses aspectos, fornecendo não só dados corporativos consolidados, mas também, devido ao seu alto grau de detalhamento dos dados, na granularidade que ele queira dispor em seu repositório particular OLSP.

#### 4.6.3. DW INTERNET, INTRANET E EXTRANET

A Internet constitui hoje talvez a mais popular das modernas tecnologias computacionais de que se tem conhecimento. É comum vermos hoje tanto cidadãos quanto empresas divulgarem seus endereços eletrônicos (e-mail) e das suas páginas virtuais da Internet (home-page). Frequente é também constatada a utilização de uma malha de

---

<sup>51</sup>Para Inmon os SGBD multidimensionais são perfeitos para serem usados com *Data Marts*.

comunicação intra-empresarial baseada na Internet internamente, ou seja, numa Intranet (do inglês “intra-Internet”), pois, deste modo as informações circulam mais rápido e facilmente entre os funcionários de uma mesma organização. A Intranet pode ser vista como a versão corporativa da Internet, servindo como um propulsor empresarial para a globalização [SERMOUD ed., 1996]. Algumas empresas estendem as suas Intranets interligando-as às de outras, tais como parceiros comerciais, fornecedores, distribuidores e afiliados, dando origem às chamadas Extranets (contração do inglês “External Intranet”, Intranet externa).

Toda essa movimentação em prol da facilitação de comunicação não poderia ficar à margem do estudo do DW. Contudo, um poderoso meio de comunicação como esse não está livre de problemas de segurança. “A segurança é um dos maiores obstáculos para a implementação de uma solução de DW para Intranet, em particular se ela estiver ‘linkada’ com a Internet” [TANLER, 1998]. O uso da Intranet em casamento com o DW é aconselhável, desde que seja de forma segura, pois pode facilitar desde o transporte dos dados tanto para as *ferramentas de extração, integração e transformação*, quanto às *ferramentas de produtividade*.

Com a tendência da fixação do trabalhador em sua residência (muitos já aderiram a essa linha, conforme o conceito SOHO - “Small Office Home Office”), vemos também a vantagem, com o DW disponível via Internet, de podermos ter a participação da alta cúpula gerencial, independente do local geográfico, de posse das informações de nível estratégico que deseje obter. Naturalmente que a Extranet deve estar suficientemente segura para não deixar vazar informações, pois, com a globalização, um alto funcionário pode precisar ao mesmo tempo estar em um local e ter que tomar uma decisão interna baseada em informações estratégicas para a empresa estando distante, em outro país por exemplo. E para isso ele, através da Internet, Intranet e Extranet poderá atender prontamente a esse requisito momentâneo.

Vale salientar que não só a alta cúpula poderá beneficiar-se com o DW. Como dissemos, parceiros comerciais, fornecedores, distribuidores e afiliados poderão compartilhar alguns dados provenientes do DW, ou seja, provavelmente será gerado um

DM especialmente voltado para consultas externas. Neste ponto ressaltamos que, também aqui, equivocadamente, há alguns que tencionam não apenas consultar o DW na Internet, mas alterá-lo. “Historicamente, os DW forneciam aos usuários um acesso somente de leitura aos recursos de informações, mas à medida que os DW se tornaram a base para aplicações do tipo previsão de vendas e criação de orçamentos, muitos DW dão privilégios de acesso de inserção e atualização limitados para alguns usuários. Entretanto o fornecimento de privilégios de escrita no banco de dados, mesmo que limitado, aumenta de maneira significativa as preocupações com relação à segurança porque os usuários podem corromper o conteúdo do DW - acidental ou intencionalmente” [TANLER, 1998].

A resposta para esse fato não deve ser a de desvirtuar a dinâmica funcional do DW, expondo-o à adulteração, e sim equacionar adequadamente o problema: existe um procedimento operacional externo que está sendo baseado em dados estratégicos internos. Como toda decisão estratégica deve, preferencialmente, ser baseada nos dados operacionais, há também uma tendência natural de se enxergar de forma miope a retroalimentação do DW (tanto por parte do corpo operacional, quanto pelo nível administrativo das organizações), devido aos fenômenos de parceria, alto custo de manutenção de estoque, terceirização, etc. sofridos pela maioria das grandes empresas nos dias de hoje e por parte daqueles ligados à empresa via Extranet.

A solução para esse problema está na criação de um sistema operacional que funcione sob a plataforma Internet e que tenha um DM especialmente consolidado para esse fim atuando como uma tabela de consulta, e que torne acessível e de modo transparente para o usuário um repositório para as inserções e alterações que sejam necessárias, prevendo tanto uma crítica, quanto fazendo a consistência do que ele venha a modificar, de forma que, ao final, se obtenha uma visão dessas alterações de maneira destacada, autenticada e com um LOG de transação, pois trata-se de uma operação de um SA. Falamos aqui de uma aplicação SA baseada em DM, o que é uma tendência atual.

Alguns ainda vêem a Internet como um DW [PU, 1997], baseando-se, dentre outras coisas, no aspecto de que na Internet há muitas fontes de dados potencialmente disponíveis e nela há preservação de qualidade de informação através do processo de manufatura de



informação. Faltam dois pontos na Internet que a impedem que ela possa ser considerada a rigor um DW. Primeiro ela não armazena dados de forma temporal e segundo os dados nela contidos não estão nem criticados, quanto mais consolidados de forma alguma.

#### **4.7. Problemas em Aberto**

Muito se tem a pesquisar em prol da melhoria de vários aspectos no DWing. Antigas metas do Processamento de Dados continuam (e sempre continuarão), tais como o armazenamento ótimo, velocidade de processamento, etc.. Em particular, a questão da obtenção do melhor dado para o DW ainda é algo muito subjetivo e mesmo indefinido. Certamente é uma função do negócio, mas há necessidade urgente de criar uma metodologia de obtenção desta qualidade de dado, que povoará o DW, fundamental na elaboração do DWing. Já há grandes referências, algumas com mais de 500 paginas tais como [BRACKETT, 1996], que abordam essencialmente o problema dos dados, contudo ainda é obscura a via-crúcis para aqueles que se deparam com a opção da escolha efetiva e eficaz do melhor dado para compor um DW.

Novos alvos para a Ciência da Computação passam a ser perseguidos, a exemplo da obtenção dos dados de forma mais “natural”. Terminais comerciais de venda (PDV- Ponto de Venda ou POS - “Point of Sale”, em inglês), que substituíram as antigas caixas registradoras, acumulam toda a relação das compras efetuadas por um cliente qualquer na granularidade de item e com detalhes como o peso da mercadoria, etc. Telefones celulares mapeam na companhia telefônica o trânsito de um cliente, mesmo que ele saia da sua área de mobilidade (“home”)<sup>52</sup>. Inevitavelmente serão dados para DW e transformar-se-ão em informações estratégicas. Não podemos deixar de sinalizar que, num futuro não tão distante, dados que venham a alimentar um DW, nada impede hoje de serem os dados de uma população, na granularidade do código genético dos cidadãos.

---

<sup>52</sup>Em telefonia celular, área de mobilidade é o local onde o cliente obteve a sua habilitação no serviço de telefonia celular. É discriminada, pois é onde o cliente paga a menor tarifa no uso do serviço.

Estudos multidisciplinares poderão e deverão ser feitos, no sentido de melhor lidar com muitos dos aspectos sutis e éticos dos dados e informações decorrentes deles, tanto nas suas obtenções, quanto nos seus usos. Os seus maus usos de forma estratégica pelas organizações (aqui ampliamos o sentido da palavra também para abranger as instituições governamentais) e as suas conseqüências - tanto boas, como adversas, não apenas ao ser humano na sua unidade, mas também a influência na sua organização social, precisam ser identificadas, acompanhadas e analisadas. Sendo assim, muitos problemas encontram-se em aberto quanto ao DWing e, como vemos, extrapolam o domínio tanto da Ciência da Computação quanto da ciência mais a ela afim, a Administração.

#### **4.8. O Estado da Arte**

O avanço da tecnologia da computação no que diz respeito à informática tem avançado no sentido de se atender prontamente não somente às solicitações atuais, mas de se adiantar ante aos acontecimentos. Seleccionamos três tópicos para expor, porém não constituem necessariamente o esgotamento do assunto intitulado como estado da arte do DWing. Veremos inicialmente um projeto acadêmico de vanguarda, o da Universidade de Stanford.

##### **4.8.1. Manutenção *On-line* / Atualização Automática**

Denominado de WHIPS (“Warehouse Information Project at Stanford”), o projeto de DWing da Universidade de Stanford [HAMMER ET AL, 1995] caracteriza-se pela sua arquitetura baseada em um repositório com dupla função. Conforme a nossa classificação anterior da seção 4.5, trata-se de um repositório de duas camadas do tipo DDW (vide Figura 4.14 na seção 4.5.2.2), diferenciando-se e destacando-se pelos aspectos de atualização do repositório DW de forma *on-line* e, de preferência, em tempo real. Dentre muitos aspectos, destacamos a particularidade da atualização *on-line*, ou seja, o processo de atualização do DW se dá sem que haja uma pausa na sua utilização por parte dos

usuários (o que acarreta uma complexidade enorme na geração, controle e gestão de várias versões de dados internas ao repositório) e o seu caráter de dinâmico, pois uma estrutura chamada de “monitor” detecta as alterações ocorridas nas fontes de informação e as propagam num formato genérico para uma outra estrutura chamada de “integrator”. É um projeto interessante que cria e trata vários problemas de grande complexidade (vide coleção de artigos na INTERNET em <http://www-db.stanford.edu/warehousing/publications.html>).

#### 4.8.2. Retroalimentação

Outro ponto que vem emergindo como uma possível tendência futura é a elaboração de novos projetos de DWing, contemplando a retroalimentação dos sistemas a partir do DW. Seria uma nova aplicação para DW especialmente definida para atuar não apenas na alta cúpula do meio estratégico empresarial, mas também contribuir para a área tática da organização. Algo que funcionaria como um oráculo corporativo onde algumas informações fossem apresentadas para o funcionário de modo que visualmente ele pudesse obter um panorama temporal e corporativo, a fim de posicionar-se frente às informações interativamente.

Como exemplo poderíamos ter um gerente de um banco de porte mundial, efetuando uma consulta DSS em um DW com a meta de descobrir e estimular clientes que sejam potenciais consumidores. Isso poderia ocorrer a partir dos diversos dados cadastrais dos clientes, seus produtos, nas agências em todo o país. O sistema de apoio à decisão, de posse dos dados consolidados temporalmente no DW, efetuando uma análise de segmentação de clientes, identifica uma classe emergente de clientes possíveis consumidores potenciais. Então, o gerente nacional de contas comanda alterações privilegiadas, para esse novo grupo de consumidores em potencial, de acordo com a política definida pela alta cúpula, tais como aumento de limite de endividamento no cheque, taxa diferenciada em aplicações, etc. Vemos que as mudanças em questão seriam feitas para o grupo seletivo a partir do DW, no qual estaria guardado (a nível de metadado) o endereço do parâmetro e o SA destino para alteração desejada, ou seja, para o sistema de

contas correntes seria comandada a alteração do campo do limite de endividamento, para o sistema de aplicações o valor da taxa de remuneração de empréstimo e assim por diante. Desta forma teríamos uma retroalimentação no DW, pois as alterações nos sistemas origens baseadas no DW, ao final seriam refletidas e repassadas para o DW, não gerando qualquer inconsistência e viabilizando também o trabalho da área tática empresarial.

#### 4.8.3. Reatividade e Tecnologia PUSH

Conforme [MELO ET AL, 1997], a evolução contínua, cada vez mais presente nas mais diversas áreas da vida moderna, das tecnologias de hardware e software, do ambiente e da abrangência das aplicações de BD, impulsiona alguns tópicos avançados no estudo dos BD, tais como BD Ativos, acesso via INTERNET e o próprio DW. Ao unirmo-los encontramos uma grande tendência no DWing, que é a elaboração de sistemas baseados em DW reativos, envolvendo a tecnologia PUSH, na Internet. Esclareceremos muito superficialmente cada um desses tópicos, de forma à mostrar que na sua interseção poderemos ter uma ferramenta bastante interessante e útil.

Um sistema reativo (tipo DSS) baseia-se num repositório do tipo BD Ativo. Ainda citando [MELO ET AL, 1997], os DB Ativos possuem as características especiais de “ter a capacidade de detectar a ocorrência de eventos, monitorar condições específicas sobre o estado do BD e executar algumas ações independente de qualquer solicitação externa ao sistema”. Desta forma e tendo como referencial o negócio a que o DWing se presta, são elaboradas regras de evento-condição-ação (ECA) especificadas por um modelo de conhecimento pertinente ao negócio. A utilização da reatividade em DWing já vem sendo bastante discutida<sup>53</sup> ultimamente. Contudo, a dinâmica da reatividade no DWing não teria tão grande valia, caso não houvesse uma comunicação efetiva e de forma rápida do ocorrido a parte interessada no acompanhamento do fato, no caso um ou mais funcionários.

---

<sup>53</sup>1º Congresso Internacional de *Data Mart* - 23 e 24 de setembro de 1997 - São Paulo/SP - Brasil

Nesse ponto é que entra o uso da tecnologia “push” (do inglês empurar), que, segundo [TAURION, 1997], é uma tecnologia que propõe levar informações diretamente ao usuário, passivamente, porém gera um grande inconveniente na rede em uso, devido ao alto tráfego nos servidores e demanda na rede. Ele ainda alerta que a tecnologia “push” baseada no acesso periódico em background a determinados servidores, ainda é bastante limitada, pois possui problemas técnicos de rede (demanda bastante largura de banda). Na prática não tem obtido êxito no fornecimento de informações em tempo real (quando aplicado na Internet) e sobrecarrega os servidores alvo. Trata-se de uma arquitetura cliente-servidor onde, através de um processo chamado de “unicast”, cada cliente é informado automaticamente de determinado assunto.

Deste modo, a tecnologia push mantém atualizados os clientes que, devidamente credenciados, são informados dos aspectos concernentes às situações dos dados empresariais, baseada na reatividade do BD, que por sua vez segue as suas regras ECA no seu modelo de conhecimento, pertinente ao negócio e definido através do processamento do DSS.

#### **4.9. Resumo**

A reunião de forma integrada e consistente dos dados oriundos dos processamentos transacionais dos sistemas operacionais, a fim de efetuarmos consultas OLAP e OLSP, requer um repositório com características especiais. Esse repositório singular denomina-se de Data Warehouse (DW). Então o DW nada mais é do que uma fonte de dados, especialmente tratada e preparada para consultas de informações estratégicas, ou seja, para os tipos emergentes de processamento: analítico e sintético.

Vimos a definição do DW, um BDT que tem como função atender às necessidades de conhecimento dos dados operacionais vistos organizacionalmente pelos EIS e DSS. O DW possui basicamente duas finalidades: *integrar os sistemas voltados à automação de forma a viabilizar a visão sistêmica da organização*, ou seja, dar a noção geral da

amplitude e complexidade dos negócios da organização, e *fornecer estes dados especialmente selecionados para os sistemas voltados ao gerenciamento organizacional.*

Foi enfocada também a importância do processo pertinente à criação e manutenção do DW, o DWing, e ressaltados os pontos relevantes no DW, ou seja, os dados, o tempo dos dados, a modelagem, a sua função crucial de armazenamento e traduzindo a importância dos metadados. Vimos a importância singular da temporalidade inerente ao DW, que imprimem no processo de desenvolvimento de DWing importantes detalhes no armazenamento e manipulação de dados históricos. Desta forma concluímos que, ao modelarmos eficientemente os dados de maneira a construir o seu histórico, eliminamos qualquer redundância de origem histórica e geramos um DW ágil e sempre pronto a prover todo tipo de dado que venha a ser solicitado, de forma consistente, temporal e segura.

Sob o enfoque prático, conceituamos e expusemos o ODS no contexto da sua importância efetiva nos projetos de DWing. Elaboramos uma classificação das arquiteturas baseadas em DW a partir do enfoque cliente-servidor. Mostramos a interação do DW com outras tecnologias emergentes, tais como OLAP, OLSP e INTERNET. Abordamos pontos que constituem problemas em aberto no DWing, especialmente a necessidade de elaboração de uma metodologia para obtenção do melhor dado que pertencerá ao DW, assim como os aspectos éticos concernentes ao uso efetivo do DW. Cada vez mais utilizado no ramo dos negócios para viabilizar a prática do suporte à decisão, de forma a imprimir no mercado o esperado diferencial competitivo empresarial, sistemas de DWing têm crescido muito. Deste modo e por último, mostramos um panorama das tendências de aplicações desta tecnologia.

Com isso, acreditamos que atentando para esses pontos adequadamente, estamos contribuindo para um DWing mais simples, do ponto de vista teórico – pela abordagem direta e prática da realidade das grandes corporações –, buscando mais eficiência e eficácia que serão sentidos pela gerência e por toda a organização, com a sua evolução, graças à utilização apropriada aquiteturalmente, independentemente do modelo adotado, de uma ferramenta administrativa poderosa com esta, que é o DW.

## 5. Conclusão

Conhecer DWing não é uma tarefa simples. Muitos agravantes de complexidade tais como: inúmeros neologismos, redefinições de palavras consagradas, vasta literatura comercial e pouca literatura acadêmica, contribuem para torná-lo de difícil domínio. Na nossa pesquisa tentamos elucidar o maior número de pontos possíveis. Muitos desses pontos carecem de maiores pesquisas científicas de forma a contemplar as necessidades do mercado, com soluções mais científicas e menos empíricas.

Entender os SI (EDP, MIS, EIS e DSS), assim como eras tecnológicas da informática (automação, controle e produtividade), são de fundamental importância para quem deseja enveredar pelo DWing. Isso pode ser constatado e concluído através da verificação da quantidade de projetos que fracassam na prática. Distinguir as tecnologias *legacy*, OLTP, DW, OLAP e OLSP, é básico para compreendermos o DWing de forma correta, porém isso ainda não é tão evidente para muitos. O conhecimento dos neologismos e redefinições utilizadas na prática são fundamentais na busca de uma equalização da linguagem empregada, visto que projetos com este grau de complexidade requerem o envolvimento de muitos fornecedores.

Ressaltamos também a importância da necessidade específica de armazenamento dos dados históricos por diversas razões (legais, concernentes ao negócio e operacionais) e as vantagens em bem guardar a memória do tempo no contexto de um repositório que se credita a função de ser um DW. Sugerimos para isso a conhecida abordagem dos BDT somada à modularização do projeto de DWing, limitando o DW a sua função crucial, o armazenamento corporativo consolidado e histórico dos dados, de forma a interagir com as demais tecnologias já existentes (OLAP, OLSP, INTERNET, etc.) e outras que venham a surgir isoladamente ou pela associação delas, para que não haja grandes transtornos de adequação à evolução tecnológica natural. Deve ser evitado o enfoque distorcido das

finalidades do DW, tanto como repositório OLAP, quanto como ferramenta de integração, comunicação e democratização das informações em uma corporação.

Notamos na nossa pesquisa que as abordagens temporais registradas no estudo do DWing pouco têm evidenciado o aspecto da multiplicidade semântica do tempo. Procuramos abordar esse ponto, mesmo que superficialmente, dando destaque à sua importância, ou seja, enfocando de forma bem diferente dos ambientes de DWing usuais, onde o tempo é tratado de uma maneira bastante primitiva. As referências feitas às questões temporais limitam-se a situar um evento no tempo e encarando-a no máximo como uma dimensão para o OLAP.

Oportuno e cabível foi a criação da terminologia que evidenciasse o tipo de processamento utilizado nas ferramentas do tipo DSS, que atuam como “*Data Mining*” na descoberta de informações estratégicas contidas nos dados. Devido às suas características singulares, requer uma denominação particular a nível de processamento por nós denominada de OLSP. Visto que o OLSP assume um papel equivalente ao OLAP, enquanto ferramenta de produtividade, levantamos a questão das diferentes características dos processamentos OLAP e OLSP, o que sugere repositórios individuais com características também distintas entre si e, obviamente, independentemente do DW. Desta forma, distanciamos e distinguimos mais claramente o DW do OLAP, fazendo aflorar os conceitos básicos do DW como tecnologia de integração corporativa.

Constatar que, na prática, o modelo de Kimball, hoje, é o mais popular. Ele compila análise multidimensional e *Data Mining*, a partir de um repositório único e temporal e com características de DW. O resultado disso está no relatório por ele proposto, fruto de processamentos tanto OLAP, quanto OLSP (esse de forma bastante elementar), oriundos de um único repositório com armazenamento relacional e modelado com o “star schema” de sua autoria e por ele denominado sutilmente de DDW. Esquema esse que contempla perfeitamente o processamento OLAP. Contudo, já não podemos afirmar o mesmo quanto ao OLSP, pois não se tem notícia do uso dessa modelagem neste tipo de processamento. Acreditamos que, com o desenvolvimento natural dos OLSP esse modelo tenha que sofrer uma revisão para atender a consultas OLSP complexas.



### 5.1. Dificuldades Encontradas

Listaremos a seguir os pontos que consideramos difíceis na confecção do presente trabalho:

- Ausência de bibliografia científica especializada que atendesse às nossas necessidades de pesquisa. Quando encontramos, limitava-se à alguma variação da arquitetura DDW, não contemplando de forma satisfatória as demais;
- Ausência de um consenso sobre a terminologia correntemente empregada no DWing, expressões como *legacy* e MIS têm diferentes interpretações;
- Limitado material no nosso idioma. Quando encontramos algo, continha algumas traduções literais que dificultavam, em muitos casos, o entendimento do sentido original das expressões empregadas, a exemplo da expressão “sistema operacional”;
- Ausência de bibliografia especializada produzida no nosso país, adequada às realidades nacionais que trouxessem as observações acima constatadas e outras, tais como a complexidade do sistema financeiro nosso e a riqueza de informações que podem ser obtidas dele;
- Em campo, os casos de sucesso são evidenciados, porém os fracassos, por razões estratégicas, são omitidos, dificultando a elucidação dos casos de real sucesso no DWing;
- Conceitos antigos revistos sob novas óticas, tais como multidimensionalidade, são envolvidos por aspectos comerciais, obscurecendo a técnica e mistificando o tema;
- Quebra de paradigmas consagrados, como, por exemplo, a normalização nos BDR e modo de operação (read/append only) dos BD, no uso em repositórios OLAP;
- Dificuldade de integração entre o que estava acontecendo em campo e as teorias existentes a respeito dos fenômenos envolvidos.

## 5.2. Contribuição

O presente documento aborda distintamente as tecnologias DW e OLAP, introduzindo o conceito de OLSP para os sistemas DSS/*Data Mining*, na tentativa de consolidar a forma singular de processamento sintético voltado ao nível estratégico empresarial. Traduz de forma clara e concisa todas essas tecnologias emergentes e suas interações. Aproximou um estudo acadêmico bastante maduro, o BDT, a uma necessidade tecnológica premente do mercado, o DW. Consagra a importância dos aspectos temporais no mundo comercial gerencial, a nível estratégico. Ressalta os vários tipos de abordagens temporais dos dados e nos dados, as suas utilizações práticas e importância no DWing.

Apesar do caráter multidisciplinar do estudo do DWing, aprofundamo-nos na área da Tecnologia da Informação de forma a verificar o emprego desta tecnologia sob a ótica prática. Quando se fez necessário, recorremos aos conceitos primitivos da informática, quando não, estivemos abertos a aceitar, de maneira positiva, o enfoque comercial que é dado nas novidades.

Desenvolvemos um estudo bastante abrangente sobre o tema DWing – o primeiro na nossa instituição – não só baseado na pesquisa bibliográfica, mas também na experiência pessoal do autor, com quase 10 anos de atividades em uma grande corporação tendo atuado no desenvolvimento e manutenção de vários sistemas, inclusive participando em diversos projetos de integrações sistêmicas.

Outro ponto é a questão do referencial básico de nosso estudo ser o de uma grande corporação, ou seja, idealmente uma empresa cujo produto é oriundo do processamento dos dados, tais como bancos. Bancos são grandes usuários de informática e lá encontramos uma complexidade tecnológica que requer soluções simples de funcionamento, porém sofisticadas na sua elaboração.

Não podemos deixar de destacar a elucidação feita a respeito dos tipos de processamento existentes, suas aplicações e características, assim como destacamos as diferentes necessidades de armazenar distintamente os dados para diferentes usos

estratégicos, isto é, para o uso analítico ou sintético. Destrinchamos as arquiteturas de DW existentes. Abordamos interações tecnológicas emergentes. Classificamos de forma evolutiva os sistemas de informação, situando-os nos contextos da informática, sempre de forma simples e didática, baseando-nos na aplicabilidade da tecnologia e sob a ótica da contribuição prática. Além disso, acreditamos que, por ser o pioneiro trabalho de tese sobre esse tema na nossa instituição, a UFPb, ele servirá de base para projetos na área de DW do departamento.

### **5.3. Sugestões para Trabalhos Futuros**

Listaremos algumas sugestões relacionadas com o tema:

- Estudo de um módulo temporal que relacione as dimensões tempo e espaço equalizando-as de forma a contemplar complexos processamentos que exijam diferentes referenciais tanto para múltiplas dimensões, quanto para uma única dimensão que possua dados variantes nos âmbitos espaciais e temporais;
- Monitor temporal para acompanhamento de consultas e classificação dos dados no DW de forma a evidenciar outros estudos relevantes na busca do *melhores dados*, tais como a distinção de dados que precisam ser obtidos com maior precisão, daqueles que não existem tal aspecto de rigorismo;
- Tratamento de dados incompletos e mesmo errados registradas nas bases operacionais e necessários no DW para uso em ambientes de EIS e DSS, de forma a identificar erros e falhas sistemáticos, tais como preenchimentos de campos, digitação de caracteres, etc.;
- Especificação formal de Repositório OLSP, que contemple as suas necessidades e
- Elaboração de uma metodologia de análise de custos que contemple o tempo de payback de um projeto de DWing, isto é, o tempo de retorno do investimento.

**ABSTIME** – Tipo de formatação de campo de tempo, com a precisão em milésimos de segundos, utilizado com parâmetro para o comando **ASKTIME** do **CICS**.

**AM** – Acrônimo para **Análise Multidimensional**.

**Análise Multidimensional** – Técnica que consiste na visualização simultânea, por diversos ângulos administrativos, do agrupamento criterioso dos descritores quantitativos e qualitativos de um fato comercial.

**ASKTIME** – Comando **CICS** que retorna o tempo imediato do relógio do computador.

**Banco de Dados Multidimensional** – Repositório de dados concebido especialmente para uso no **Processamento Analítico**. É um **Repositório OLAP** de modelagem proprietária, cuja a ênfase está na organização interna dos dados, feita sob forma de matrizes n-dimensionais, com o objetivo de obter velocidade na recuperação dos dados consolidados sob enfoque multidimensional.

**Banco de Dados Relacional** – Repositório de dados concebido especialmente para o **Processamento Operacional**. Caracteriza-se pela disposição dos dados sob a forma de tabelas, de modo a facilitar as suas manutenções, segundo a álgebra relacional, visando oferecer respostas imediatas e consistentes a solicitações de armazenamento e recuperação dos dados.

**Banco de Dados Relacional Temporal** – É o **BDR**, modelado para atuar como **BDT**.

**Banco de Dados Temporal** – É um repositório de dados no qual há tanto o registro, quanto a possibilidade de tratamento do tempo que é associado aos seus dados armazenados.

**BDM** – Acrônimo para **Banco de Dados Multidimensional**.

**BDR** – Acrônimo para **Banco de Dados Relacional**.

**BDRT** – Acrônimo para **Banco de Dados Relacional Temporal**.

**BDT** – Acrônimo para **Banco de Dados Temporal**.

**CICS** – Acrônimo para “**Customer Information Control System**”.

**CIF** – Acrônimo para “**Corporate Information Factory**”.

**Corporate Information Factory** – Tipo de arquitetura idealizada por William Inmon para grandes corporações.

**Customer Information Control System** – Produto IBM para teleprocessamento e gerenciamento de transações.

**Data Mart** – Qualquer subconjunto de um **Data Warehouse**.

**Data Mining** – Técnica que objetiva identificar informações não intuitivas que naturalmente não seriam inteligíveis.

**Data Warehouse** – Banco de Dados Temporal especialmente voltado para gerência estratégica.

**Data Warehouse Current Detail** – Componente da arquitetura **CIF**, trata-se de um **Data Warehouse** composto por dados bastante detalhados.

**Data Warehouse Summary** – Componente da arquitetura **CIF**, trata-se de um **Data Warehouse** composto por dados resultantes de agregações e/ou consolidações de dados.

**Data Warehousing** – Ato ou efeito da criação de um ambiente computacional especialmente voltado para gerência estratégica.

**DDW** – Acrônimo para “**Dimensional Data Warehouse**”.

**Decision Support System** – Tipo de sistema para informações gerenciais, voltado à gerência estratégica, que, a partir de uma meta do usuário, sintetiza informações para uso estratégico empresarial em futuras decisões a serem tomadas.

**Dimensional Data Warehouse** – Tipo de **Data Warehouse** baseado em **BDR**, com funções de **Repositório OLAP**. Para isso, ele utiliza alguma modelagem especial, que não a relacional completamente normalizada.

**Dimensões** – São agregações dos atributos qualitativos dos dados operacionais segundo um critério de natureza comum, relevante e concernente estrategicamente aos negócios.

**DSS** – Acrônimo para “**Decision Support System**”.

**DW** – Acrônimo para “**Data Warehouse**”.

**DW-cd** – Acrônimo para “**Data Warehouse Current Detail**”.

**DWing** – Acrônimo para “**Data Warehousing**”.

**DW-s** – Acrônimo para “**Data Warehouse Summary**”.

**ECA** – Acrônimo para Regra Evento, Condição, Ação

**EDP** – Acrônimo para “**Electronic Data Processing**”.

**EIS** – Acrônimo para “**Executive Information System**”.

**Electronic Data Processing** – Tipo de sistema de informação, voltado ao âmbito operacional organizacional, responsável pela automação de processos.

**Executive Information System** – Tipo de sistema de informação gerencial voltado à gerência estratégica, no qual cabe ao usuário, através de análise pessoal, inferir informações estratégicas empresariais, para se posicionar frente a futuras decisões a serem tomadas.

**Management Information Systems** – Tipo de sistema de informação gerencial voltado à gerência tática no acompanhamento dos negócios.

**MDOLAP** – Acrônimo para “**Multidimensional On Line Analytical Processing**”.

**MIS** – Acrônimo para “**Management Information Systems**”.

**MOLAP** – Acrônimo para “**Multidimensional On Line Analytical Processing**”.

**Multidimensional On Line Analytical Processing** – Modelagem de dados utilizada nos **BDM**, baseada no uso de matrizes n-dimensionais.

**NC** – Acrônimo para “**Network Computer**”.

**Network Computer** – Computador especialmente desenvolvido para utilização em ambiente de rede.

**OLAP** – Acrônimo para “**On-line Analytical Processing**”.

**OLSP** – Acrônimo para “**On-line Synthetical Processing**”.

**OLTP** – Acrônimo para “**On-line Transaction Processing**”.

**On-line Analytical Processing** – Tipo de processamento de dados voltado ao gerenciamento estratégico dos negócios. Baseia-se nos dados operacionais e seus atributos, provendo ao usuário a possibilidade de visualização dos fatos comerciais por diversos ângulos administrativos simultaneamente. É através do **OLAP** que se viabiliza a **Análise Multidimensional**.

**On-line Synthetical Processing** – Tipo de processamento de dados voltado ao gerenciamento estratégico dos negócios. Baseia-se nos dados operacionais e seus atributos, provendo ao usuário a identificação de padrões, tendências e até a descoberta de informações não intuitivas, que naturalmente não seriam inteligíveis, dos fatos comerciais por diversos ângulos administrativos. É através do **OLSP** que se viabiliza o **Data Mining**.

**On-line Transaction Processing** – Tipo de processamento de dados voltado à operacionalização dos negócios. Tem como meta a prontidão na resposta das solicitações de armazenamento e recuperação dos dados.

**Processamento Analítico** – ver **OLAP**.

**Processamento Operacional** – ver **EDP**.

**Processamento Sintético** – ver **OLSP**.

**Repositório OLAP** – Repositório de dados concebido especialmente para uso no Processamento Analítico. Os Repositórios OLAP são organizados com vista a facilitar as consultas do usuário na forma de navegação nos dados na ótica de dimensões.

**SGBD** – Acrônimo para **Sistema Gerenciador de Bando de Dados**.

**SGBDM** – Acrônimo para **Sistema Gerenciador de Bando de Dados Multidimensionais**.

**SGBDR** – Acrônimo para **Sistema Gerenciador de Bando de Dados Relacional**.

**Sistema Gerenciador de Bando de Dados** – Conjuntos de rotinas que garantem a consistência no armazenamento e recuperação dos dados, segundo critérios otimizados de acesso aos mesmos.

**Sistema Gerenciador de Bando de Dados Multidimensionais** – Conjuntos de rotinas que garantem a consistência no armazenamento e recuperação dos dados, segundo critérios otimizados de acesso aos dados, consolidados e organizados sob a ótica multidimensional simultânea.

**Sistema Gerenciador de Bando de Dados Relacional** – Conjuntos de rotinas que garantem a consistência no armazenamento e recuperação dos dados, segundo critérios otimizados de acesso aos mesmos, que são organizados sob a forma de tabelas e tendo como fundamento teórico a Álgebra Relacional.

**Snowflake Schema** – Tipo de modelagem de dados para uso nos **BDR** na simulação de **BDM**, visando a sua atuação como **Repositório OLAP**. Constitui-se de uma tabela central, chamada de tabela de fato, rodeada por outras tabelas normalizadas, chamadas de dimensões, que dão origem a novas tabelas, em efeito cascata.

**Star Schema** – Tipo de modelagem de dados para uso nos **BDR** na simulação de **BDM**, visando a sua atuação como **Repositório OLAP**. Constitui-se de uma tabela central, chamada de tabela de fato, rodeada por outras tabelas não normalizadas, chamadas de dimensões.

## Referências Bibliográficas

- [AGRAWAL & SRIKANT, 1994] AGRAWAL, R. & SRIKANT, R. – *Fast Algorithms for Mining Association Rules in Large Databases*, VLDB 94 - Santiago/Chile, Sept. 1994.
- [BARGHOUTI & KAISER, 1991] BARGHOUTI, N. & KAISER, G. – *Concurrency Control in Advanced Database Applications*, ACM Computing Surveys, mar. 1991.
- [BERNSTEIN ET AL, 1987] BERNSTEIN, P. A. HADZILACOS, V. & GOODMAN, N. – *Concurrency Control and Recovery in Database Systems*, Massachusetts: Addison-Wesley, 1987.
- [BERSON & SMITH, 1997] BERSON, A. & SMITH, S. J. – *Data Warehousing, Data Mining and OLAP*, New York: McGraw-Hill, 1997.
- [BISCHOFF & ALEXANDER, 1997] BISCHOFF, J. & ALEXANDER, T. – *Data Warehouse: Practical Advice from the Experts*, New Jersey: Prentice-Hall, Inc., 1997.
- [BRACKETT, 1996] BRACKETT, M. H. – *The Data Warehouse Challenge: Taming data chaos*, New York: John Wiley & Sons, 1996.
- [BURETTA, 1997] BURETTA, M. – *Data Replication - Tools and Techniques for Managing Distributed Information*, New York: John Wiley & Sons, 1997.
- [CAMPOS, 1996] CAMPOS, A. – Datawarehouses: Dividir para Conquistar. In *Computerworld*, 3 (181), IDG - Computerworld do Brasil Serviços e Publicações, 23 Set. 1996.
- [CAUTELA ET AL, s.d.] CAUTELA, A. L. POLLONI, E. G. F. LADESSA, J. F. TUMP S. & BERNARDO, V. – *MIS - A Estratégia da Informação*, 2ª. ed, São Paulo: Thema, s.d..
- [CHIAVENATO, 1979] CHIAVENATO, I. – *Teoria Geral da Administração*, ed revista, vl 1-2, São Paulo: McGraw-Hill, 1979.



- [CODD, 1993] CODD, E. F. – *Providing OLAP (On-line Analytical Processing) to User Analysts*, E.F. Cood and Associates, s.l., 1993.
- [DATE, 1991] DATE, C. – *Introdução a Sistemas de Bancos de Dados*, Rio de Janeiro: Campus, 4ª ed., 1991.
- [ELMAGARMID, 1994] ELMAGARMID, A. – *Database Transaction Models for Advanced Applications*, 1994.
- [GUROVITZ, 1997] GUROVITZ, H. – O Que Cerveja Tem a Ver com Fraldas. In *EXAME*, São Paulo: Abril, 09 abr. 1997.
- [GILL & RAO, 1996] GILL, S. H. & RAO, P. C. – *The Official Client/Server Computing Guide to Data Warehousing*, Indianapolis: Que Corporation, 1996.
- [GRAY & WATSON, 1998] GRAY, P. & WATSON, H. – *Decision Support in The Data Warehouse*, New Jersey: Prentice-Hall, Inc., 1998.
- [HAMMER ET AL, 1995] HAMMER, J., GARCIA-MOLINA, H., WILDOW, W., LABIO, W. & ZHUGE, Y. – The Standard Data Warehousing Project, In: *IEEE Data Engineering Bulletin, Special Issue on Materialized Views and Data Warehousing*, 18 (2), June 1995.
- [HAMMERGREN, 1996] HAMMERGREN, T. – *Data Warehousing - Building the Corporate Knowledgebase*, Boston: International Thomson Computer Press, 1996.
- [INMON, 1996] INMON, W. H. – *Building the Data Warehouse*, New York: John Wiley & Sons, 2nd ed., 1996.
- [INMON ET AL, 1996] INMON, W. H. IMHOFF, C. & BATTAS, G. – *Building the Operational Data Store*, New York: John Wiley & Sons, 1996.

- [INMON, 1997] INMON, W. H. – *Como Construir o Data Warehouse*, Rio de Janeiro: Campos, 1997.
- [INMON ET AL, 1997] INMON, W. H. WELCH, J. D. & GLASSEY, K. L. – *Managing the Data Warehouse*, New York: John Wiley & Sons, 1997.
- [KELLY, 1996] KELLY, S. – *Data Warehousing - The Route to Mass Customization Updated & Expanded*, Chichester: John Wiley & Sons, 1996.
- [KIMBALL, 1996] KIMBALL, R. – *The Data Warehouse Toolkit*, New York: John Wiley & Sons, Inc., 1996.
- [KIMBALL, 1998] KIMBALL, R. – *The Data Warehouse Toolkit*, São Paulo: Makron Books do Brasil, 1998.
- [KORFHAGE, 1997] KORFHAGE, R. R. – *Information Storage and Retrieval*, New York: John Wiley & Sons Inc., 1997.
- [KOURY, 1996] KOURY, L. M. ed. – Garimpendo Dados, (Artigo Traduzido da Revista Research, 2ª ed., 1996). In: *Informação IBM - 14 (76)*, Estado da Arte, out. 1996.
- [LEANDRO, 1996] LEANDRO, P. ed. – Filas Nunca Mais. In: *Perfil – Revista do Banco Excel Econômico*, Belo Horizonte: Artes Gráficas Formato, 1 (5), Out. 1996.
- [LIMA & TOLEDO, 1997] LIMA, G. & TOLEDO, M. – Um Modelo de Transações Cooperativas Integrado a um Modelo de Versões, In: *Anais do XII SBBB (Simpósio Brasileiro de Banco de Dados)*, Fortaleza, 13-15 out. 1997.
- [MANZONI JR, 1996] MANZONI JR, R. – Middleware Viabiliza Mundo Cliente/Servidor. In: *Computerworld*, 3 (173), IDG - Computerworld do Brasil Serviços e Publicações, 29 jul. 1996.

- [MASUDA, s.d.] MASUDA, Y. – *A Sociedade da Informação*, Rio de Janeiro: Rio, s.d..
- [MATTISON, 1996] MATTISON, R. – *Data Warehousing – Strategies Technologies and Techniques*, New York: The McGraw-Hill Companies, 1996.
- [MELO ET AL, 1997] MELO, R., Silva, S. & TANAKA, A. – *Banco de Dados em Aplicações Cliente-servidor*, Rio de Janeiro: Infobook, 1997.
- [MENNINGER, 1997] MENNINGER, D. – Building Object-Oriented OLAP Applications: Not Just Any Object-Oriented Tool Will Do. Apud: BARQUIN, R. C. & EDELSTEIN, H. A. *Building, Using, and Managing the Data Warehouse*, New Jersey: Prentice Hall PTR, 1997.
- [PENSE & CREETH, 1995] PENSE, N. & CREETH, R. – *The OLAP Report, Volumes 1 and 2*, Business Intelligence, s.l., 1995.
- [PU, 1997] PU, C. – Tutorial “The Internet, WWW and Databases” In: *Anais do XII SBBID (Simpósio Brasileiro de Banco de Dados)*, Fortaleza, 13-15 out. 1997.
- [ROSSETTI, 1984] ROSSETTI, J. P. – *Introdução à Economia*, São Paulo: Atlas - 10ª edição, 1984.
- [RAO, 1952] RAO, R. – *Advanced Statistical Methods in Biometric Research*, New York: Wiley, 1952.
- [SAMPAIO, 1995] SAMPAIO, M. C. – *Transactions Cooperatives: Une Approche Dirigée par les Données – Thèse de Doctorat*, Université de Montpellier II, 1995.
- [SERMOUD ed., 1996] SERMOUD, G. ed. – Guideline - Internet & Intranet - Informação ao Alcance de Todos. In: *Computerworld*, 3 (186), IDG - Computerworld do Brasil Serviços e Publicações, 28 Out. 1996.

- [SNODGRASS, 1987] SNODGRASS, R. – *The Temporal Query Language TQUEL*, ACM Transactions on Database Systems, July 1987.
- [SOUCIE, 1996] SOUCIE, R. – O Fim do Bancos? In: *Computerworld*, 3 (185), IDG - Computerworld do Brasil Serviços e Publicações, 21 Out. 1996.
- [SULAIMAN & SOUZA, 1998] SULAIMAN, A. & SOUZA, J. – A Decision Support System that Reverse Engineers Abstract Database Transactions - The conceptual Model. In: *Data Mining*, Southampton: WIT Press/Computational Mechanics Publications, 1998.
- [TANLER & DROST, 1996] TANLER, R. & DROST, K. – Multidimensional Analysis of Warehoused Data. In: MATTISON, R. *Data Warehousing - Strategies Technologies and Techniques*, New York: The McGraw-Hill Companies, 1996.
- [TANLER, 1998] TANLER, R. – *Intranet Data Warehouse*, Rio de Janeiro: Infobook, 1998.
- [TAURION, 1997] TAURION, C. – Qual o Valor da Tecnologia “push”? In *Computerworld*, 3 (225), IDG - Computerworld do Brasil Serviços e Publicações, 08 Set. 1997.
- [TONIELLO ed., 1995] TONIELLO, S. ed. – *BIS - A Revista do Banco Econômico*, ano XX, (495), Belo Horizonte: Artes Gráficas Formato, Ago. 1995.
- [TONIELLO ed., 1996] TONIELLO, S. ed. – Uma Fábrica de Negócios. In: *RH Integração – Revista do Banco Excel Econômico*, ano 1, (2), Salvador: Imprensa Rocha, Out. 1996.
- [UHIROWCZIC, 1997] UHIROWCZIC, P. – Trends in Data Warehousing. In: BISCHOFF, J. & ALEXANDER, T. *Data Warehouse: Practical Advice from the Experts*, New York: Prentice-Hall, Inc., 1997.
- [WIDOM, 1995] WIDOM, J. – Research Problems in Data Warehousing. In: *Proc. of 4<sup>th</sup> Int'l Conference on Information and Knowledge Management (CIKM)*, Nov. 1995.

- [WHITE, 1997] WHITE, C. – Multidimensional OLAP versus Relational OLAP. In: BISCHOFF, J. & ALEXANDER, T. *Data Warehouse: Practical Advice from the Experts*, New York: Prentice-Hall, Inc., 1997.
- [WIEDERHOLD, 1997] WIEDERHOLD, G. – “Integration of Information from Heterogeneous Sources” In: *Anais do XII SBBD (Simpósio Brasileiro de Banco de Dados)*, Fortaleza, 13-15 out. 1997.
- [YAZDANI & WONG, 1998] YAZDANI, S. & WONG, S. S. – *The Data Warehousing with Oracle - An Administrator's Handbook*, New Jersey: Prentice Hall PTR, 1998.

- ABSTIME, 100, 101, 102, 104, 113  
 AM, 34, 57, 59, 60, 61, 62, 63, 65, 66, 71, 72, 89  
 Análise Multidimensional. *Consulte* AM  
 API, 84  
*Application Program Interface. Consulte* API  
 ASKTIME, 100
- Banco de Dados Multidimensional. *Consulte*  
 BDM  
 Banco de Dados Relacional. *Consulte* BDR  
 Banco de Dados Relacional Temporal. *Consulte*  
 BDRT  
 Banco de Dados Temporal. *Consulte* BDT  
 BDM, 75, 76, 82, 85  
 BDR, 30, 32, 46, 58, 60, 75, 76, 77, 81, 82, 84,  
 85, 144  
 BDRT, 107  
 BDT, 19, 91, 105, 106, 112, 140, 142, 145
- CIF, 129, 130  
*Corporate Information Factory. Consulte* CIF
- Data Mining*, 37, 38, 41, 42, 55, 71, 97, 115, 116,  
 124, 125, 143, 145, 151, 155  
*Data Warehouse Current Detail. Consulte* DW-  
 cd  
*Data Warehouse Summary. Consulte* DW-s  
 DBMS. *Consulte* SGBD  
 DDW, 75, 106, 107, 124, 125, 126, 137, 143, 144  
*Decision Support System. Consulte* DSS  
 Descoberta do Conhecimento. *Consulte*  
 Knowledge Discovery  
*Dimensional Data Warehouse. Consulte* DDW  
 DSS, 24, 25, 31, 36, 37, 39, 40, 41, 42, 43, 44, 45,  
 46, 47, 48, 52, 54, 55, 56, 71, 91, 92, 96, 97,  
 98, 100, 102, 106, 107, 109, 110, 119, 122,  
 124, 125, 126, 129, 138, 139, 140, 142, 143,  
 145, 146  
 DW-cd, 129, 130  
 DW-s, 129, 130
- ECA, 139, 140  
 EDP, 21, 26, 27, 32, 33, 52, 56, 142  
 EIS, 23, 31, 33, 35, 36, 37, 39, 40, 41, 42, 43, 44,  
 45, 46, 47, 48, 49, 50, 51, 52, 54, 55, 56, 57,  
 62, 71, 89, 90, 91, 92, 96, 97, 98, 100, 102,  
 106, 107, 109, 110, 112, 119, 122, 124, 125,  
 129, 132, 140, 142, 146  
*Eletronic Data Processing. Consulte* EDP  
 Esquema Estrela. *Consulte* Star Schema  
 Esquema Flocos de Neve. *Consulte* Snowflake  
 Schema  
*Executive Information System. Consulte* EIS
- Forma Normal, 30
- Garimpagem de Dados. *Consulte* Data Mining
- IBM, 27, 60, 65, 100  
*Information Technology. Consulte* IT  
 IT, 21
- Knowledge Discovery, 38  
*knowledge workers*, 17, 72
- Legacy*, 28, 29, 31, 54, 97, 115, 142, 144
- Mainframe*, 16, 23, 29, 49, 51  
*Management Information Systems. Consulte* MIS  
 MDOLAP. *Consulte* MOLAP  
*Middleware*, 122, 123, 127  
 Mineração de Dados. *Consulte* Data Mining  
*Mining. Consulte* Data Mining  
 MIS, 21, 24, 25, 26, 31, 32, 33, 36, 37, 39, 40, 41,  
 42, 43, 44, 45, 47, 55, 56, 142, 144  
 MOLAP, 75, 82, 84, 85, 86  
*Multidimensional Database. Consulte* BDM  
*Multidimensional On Line Analytical Processing.*  
*Consulte* MOLAP
- NC, 88, 89  
*Network Computer. Consulte* NC  
 Normalização, 46, 75, 80, 81, 144
- ODBC, 84  
 OLAP, 18, 35, 36, 39, 40, 42, 45, 54, 55, 56, 57,  
 58, 60, 62, 63, 64, 66, 69, 71, 72, 74, 75, 82,  
 84, 85, 86, 87, 88, 89, 90, 107, 132  
 OLSP, 40, 41, 42, 55, 56, 71, 125, 145  
 OLTP, 29, 30, 31, 40, 42, 54, 55, 56, 97, 106,  
 107, 115, 142  
*On-line Analytical Processing. Consulte* OLAP  
*On-line Synthetical Processing. Consulte* OLSP
- Processamento Analítico. *Consulte* OLAP  
 Processamento Sintético. *Consulte* OLSP
- RDBMS. *Consulte* SGBDR  
 Regra ECA (Evento, Condição, Ação). *Consulte*  
 ECA  
*Relational Data Base Management System.*  
*Consulte* SGBDR  
 Repositório OLAP, 71, 72, 73, 74, 75, 81, 87, 89,  
 106, 126, 143, 144  
 Repositório OLSP, 126, 146  
 Repositório ROLAP, 80  
 ROLAP, 75, 80, 82, 84, 85, 86
- SGBDM, 82, 84, 86  
 SGBDR, 30, 65, 69, 70  
 Sistema Gerenciador de Banco de Dados.  
*Consulte* SGBD

Sistema Gerenciador de Banco de Dados  
Multidimensionais. *Consulte* SGBDM  
Sistema Gerenciador de Banco de Dados  
Relacional. *Consulte* SGBDR  
Sistemas Legados. *Consulte* Legacy  
*Snowflake Schema*, 80, 81, 82  
SQL, 65, 66, 70, 82, 84

*Star Join Schema*. *Consulte* Star Schema  
*Star Schema*, 76, 77, 80, 81, 82, 143

Tecnologia da Informação. *Consulte* IT  
Terceira Forma Normal (3FN). *Consulte* Forma  
Normal