

Universidade Federal de Santa Catarina

Programa de Pós-graduação em

Engenharia de Produção

**A IMPLEMENTAÇÃO E UTILIZAÇÃO DE *DATA*
WAREHOUSE EM INSTITUIÇÕES PÚBLICAS NO BRASIL :
UM ESTUDO DESCRITIVO DAS IMPLICAÇÕES
ENVOLVIDAS**

Dissertação de Mestrado

Marcus Mosquéra Bomfim

Florianópolis

2001

**A IMPLEMENTAÇÃO E UTILIZAÇÃO DE *DATA*
WAREHOUSE EM INSTITUIÇÕES PÚBLICAS NO BRASIL :
UM ESTUDO DESCRITIVO DAS IMPLICAÇÕES
ENVOLVIDAS**

Universidade Federal de Santa Catarina

Programa de Pós-graduação em

Engenharia de Produção

**A IMPLEMENTAÇÃO E UTILIZAÇÃO DE *DATA*
WAREHOUSE EM INSTITUIÇÕES PÚBLICAS NO BRASIL :
UM ESTUDO DESCRITIVO DAS IMPLICAÇÕES
ENVOLVIDAS**

Marcus Mosquéra Bomfim

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina como requisito parcial para obtenção do título de Mestre em Engenharia de Produção

Florianópolis

2001

Marcus Mosquéra Bomfim

**A IMPLEMENTAÇÃO E UTILIZAÇÃO DE *DATA WAREHOUSE* EM INSTITUIÇÕES PÚBLICAS NO BRASIL :
UM ESTUDO DESCRITIVO DAS IMPLICAÇÕES
ENVOLVIDAS**

Esta dissertação foi julgada e aprovada para a obtenção do título de **Mestre em Engenharia de Produção** no **Programa de Pós-Graduação em Engenharia de Produção** da Universidade Federal de Santa Catarina

Florianópolis, 15 de outubro de 2001

Prof. Ricardo de Miranda Barcia, Ph.D.

Coordenador do Curso

BANCA EXAMINADORA

Prof. Alejandro Martins Rodrigues, Dr.
Orientador

Prof. Nilson Ribeiro Modro, Msc.
Tutor de Orientação

Prof. Aran Bey Tcholakian Morales, Dr.

Prof. José Leomar Todesco, Dr.

Dedico este trabalho

Aos meus pais, meus filhos e meu irmão.

Agradecimentos

Agradeço primeiramente a Deus que me presenteou com o dom da vida, da liberdade, inteligência e a coragem de lutar para esta realização.

A meus pais Arnaldo e Marília, meu irmão Francisco Henrique, por este sonho que sonhamos juntos, pelas noites que passamos acordados, pelos dias tensos, pelos momentos de perdão, pelos gestos de carinho, por tudo o que sei e simplesmente, por eu existir, eu amo vocês!

Aos meus filhos, Marcus Henrique, Nathália e Paulo Henrique, em reconhecimento aos muitos momentos roubados do nosso convívio, durante a realização deste trabalho.

A todos que dispensaram seus momentos preciosos para me esclarecer com suas experiências na carreira por mim abraçada.

Aos integrantes das Instituições Públicas pesquisadas, pela cordialidade e atenção dispensadas.

Ao professor Dr. José Leomar Todesco pelos conhecimentos transmitidos na disciplina “Data Warehouse”, que foram importantes para a realização deste trabalho.

Aos meus orientadores, pelos conhecimentos e direcionamentos que me levaram à conclusão desta jornada.

Aos que amo por fazerem seus o meu sonho, pelo apoio incondicional, pelo incentivo, pela alegria de minha conquista.

“Vivemos em uma era de enormes mudanças institucionais, talvez sem precedentes desde o início da Era Industrial. É difícil encontrar instituições – governamentais, educacionais ou empresariais –, em qualquer lugar, que estejam correspondendo às expectativas da sociedade.”

Peter M. Senge

SUMÁRIO

1	INTRODUÇÃO.....	1
1.1	Formulação do Problema	2
1.2	Justificativa / Relevância da Pesquisa	3
1.3	Objetivos	6
1.3.1	Geral	6
1.3.2	Específicos	6
1.4	Estrutura do Trabalho.....	7
2	INSTITUIÇÕES PÚBLICAS NO BRASIL.....	9
2.1	Considerações Iniciais.....	9
2.2	Serviço Público	9
2.2.1	Serviço Privativo do Estado	10
2.2.2	Serviço de Utilidade Pública	10
2.2.3	Prestação de Serviço Mista.....	10
2.3	Administração Pública.....	11
2.4	Organização da Administração Pública Brasileira	11
2.4.1	Administração Direta ou Centralizada	12
2.4.2	Administração Indireta ou Descentralizada	12
2.4.3	Instituições Públicas Autárquicas	13
2.4.4	Instituições Públicas Parestatais	13
2.4.5	Empresas Públicas	14
2.4.6	Sociedade de Economia Mista.....	14
2.4.7	Fundações Públicas	15
2.4.8	Serviços Sociais Autônomos	16
2.5	Considerações Finais.....	16
3	AMBIENTE DE DATA WAREHOUSE.....	18
3.1	Considerações Iniciais.....	18
3.2	O que é Data Warehouse?	18
3.3	Características Importantes do Data Warehouse.....	20
3.3.1	Orientado a assunto	20
3.3.2	Integrado.....	21
3.3.3	Variação com o tempo.....	23
3.3.4	Não-volátil.....	24
3.4	A Estrutura do Data Warehouse	25
3.4.1	Dados Atuais	25
3.4.2	Dados Detalhados Antigos	26
3.4.3	Dados Ligeiramente Resumidos	26
3.4.4	Dados Altamente Resumidos	27
3.4.5	Metadados	27
3.5	Arquitetura de Data Warehouse.....	28
3.5.1	Modelo de Arquitetura Genérica de Data Warehouse	29
3.5.2	Operational Data Store - ODS	30
3.5.3	Data Mart – DM	32
3.5.4	Modelo de Arquitetura de DW por Ralph Kimball	33
3.6	Modelagem dos dados	35
3.6.1	Modelo Dimensional	36
3.6.2	Gerenciamento do volume de dados.....	38
3.6.3	Granularidade e Particionamento	38
3.7	Estratégias arquitetônicas para a implementação de Data Warehouse	42
3.7.1	Estratégia “Top Down”	42
3.7.2	Estratégia “Bottom-up”	42

3.7.3	Estratégia “Intermediária”	43
3.8	Metodologias de Desenvolvimento	44
3.8.1	Metodologia de desenvolvimento baseada em dados – W.H.Inmon	44
3.8.2	Metodologia dos nove pontos de decisão – Ralph Kimball.....	47
3.9	Fatores Críticos de Sucesso no Desenvolvimento do Ambiente de Data Warehouse	49
3.10	Aplicações Comerciais de Tecnologias de Data Warehouse	53
3.11	Considerações Finais.....	54
4	METODOLOGIA DA PESQUISA	56
4.1	Considerações Iniciais.....	56
4.2	Classificação da Pesquisa.....	56
4.3	Universo e Amostra.....	57
4.4	Limites da Pesquisa.....	58
4.5	Levantamento de Dados.....	58
4.6	Variáveis da Pesquisa	60
4.7	Análise e Interpretação dos Dados.....	60
4.8	Etapas da Pesquisa	61
4.9	Desenvolvimento da pesquisa	62
4.10	Considerações Finais.....	62
5	RESULTADOS DA PESQUISA.....	63
5.1	Considerações Iniciais.....	63
5.2	Análise Referencial.....	63
5.3	Análise Descritiva.....	67
5.3.1	Serviço Federal de Processamento de Dados - SERPRO	67
5.3.2	Banco do Brasil	73
5.3.3	Instituto Brasileiro de Geografia Estatística - IBGE	79
5.3.4	Banco Central do Brasil - BC.....	84
5.3.5	Empresa Brasileira de Pesquisa Agropecuária - Embrapa.....	88
5.3.6	Agência Nacional de Energia Elétrica - ANEEL	93
5.3.7	Tribunal Superior Eleitoral - TSE	98
5.4	Considerações Finais.....	101
6	CONCLUSÕES E TRABALHOS FUTUROS	102
6.1	Conclusões.....	102
6.2	Trabalhos Futuros.....	104
	BIBLIOGRAFIA	106
	GLOSSÁRIO	111
	ANEXO I - INSTRUMENTOS DE COLETA DE DADOS.....	115

LISTA DE QUADROS

Quadro 1 - Alguns exemplos de autarquias.....	13
Quadro 2 - Alguns exemplos de empresas públicas.....	14
Quadro 3 - Alguns exemplos de Sociedades de Economia Mista.....	15
Quadro 4 - Alguns exemplos de fundações públicas.....	15
Quadro 5 - Alguns exemplos de Serviços Sociais Autônomos.....	16
Quadro 6 - Algumas funcionalidades necessárias ao ambiente DW na passagem dos dados do ambiente operacional para o ambiente DW.....	22
Quadro 7 - Descrição das camadas do modelo de arquitetura proposto por Orr.....	29
Quadro 8 - Descrição dos componentes do modelo de arquitetura definido por Kimball.....	34
Quadro 9 – Casos recentes de aplicações comerciais de tecnologias de DW.....	54
Quadro 10 - Conjunto de variáveis propostas para analisar a implementação de tecnologias de DW, em uma Instituição Pública.....	60
Quadro 11 - Etapas do processo de análise e interpretação dos dados desta pesquisa.....	61
Quadro 12 – Resultados da Análise Referencial.....	64
Quadro 13 – Resultados da Análise Referencial (continuação).....	65
Quadro 14 – Resultados da Análise Referencial (continuação).....	66
Quadro 15 - Avaliação dos Protótipos de DW Desenvolvidos pelas Empresas Fornecedoras.....	90
Quadro 16 - Plano de Trabalho para o desenvolvimento do Data Mart do CNE.....	101

LISTA DE FIGURAS

Figura 1 - DW - Dados para informação	19
Figura 2 - Definição do Data Warehouse	20
Figura 3 - Orientação por assunto do DW	21
Figura 4 - A questão da integração	22
Figura 5 - A questão da variação em relação ao tempo	23
Figura 6 - A questão da não-volatilidade.....	25
Figura 7 - A estrutura de dados dentro do DW	26
Figura 8 - Componentes de uma arquitetura que devem ser considerados antes da implementação do DW ..	29
Figura 9 - Modelo de arquitetura proposto por Orr	30
Figura 10 - A arquitetura onde o ODS é uma parte integral.....	31
Figura 11 - Visão de múltiplos Data Marts integrados	33
Figura 12 - Modelo de arquitetura proposto por Kimball.....	35
Figura 13 - Modelo dimensional de um negócio: visualização dos dados como um cubo	36
Figura 14 - Um modelo dimensional de transação de seguro com uma tabela de fatos e sete tabelas de dimensão	38
Figura 15 - O balanceamento no gerenciamento da questão da granularidade.....	39
Figura 16 - A importância do particionamento dos dados no projeto de dados do DW	41
Figura 17 – Etapas da Metodologia 1 de Inmon.....	45
Figura 18 – Etapas da Metodologia 2 de Inmon.....	46
Figura 19 – Etapas da Metodologia 3 de Inmon.....	47
Figura 20 - Visão geral da arquitetura ERM	105

RESUMO

BOMFIM, Marcus Mosquéra. **A implementação e utilização de *Data Warehouse* em Instituições Públicas no Brasil : um estudo descritivo das implicações envolvidas.** 2001. 119 p. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis.

A demanda por informações estratégicas para o apoio à tomada de decisão, nessa sociedade do 3^o. milênio, tem sido um dos grandes motivos para investimentos maciços em soluções informatizadas dentro de instituições, públicas ou privadas. Contudo, apesar dos investimentos serem intensos, ao final da maioria dos projetos, implementados nessas instituições, as necessidades do tomador de decisão não são satisfeitas, principalmente devido às constantes mudanças dos cenários no processo decisório. Nas instituições públicas, esse problema se acirra ainda mais, pois nelas é comum a predominância da decisão política em detrimento de soluções técnicas. Assim, observa-se nessas instituições que vários fatores acabam por dificultar a implementação de tecnologias que permitam facilitar ao nível gerencial tomar decisões baseadas em dados, fatos e informações. Para se decidir é necessário ter-se informações que consubstanciem a decisão. Hoje tem aparecido uma série de ferramentas que auxiliam o tomador de decisões, o corpo gerencial da empresa. Dentre as várias ferramentas existentes hoje no mercado, talvez devido à flexibilidade e ao foco centrado num processo evolutivo contínuo, o que permite o desenvolvimento de aplicações voltadas à tomada de decisão, a que se destaca é a que permite a utilização de uma metodologia de desenvolvimento baseada em *Data Warehousing* o qual viabiliza o desenvolvimento de sistemas de informações permitindo o acompanhamento das mudanças de necessidades dos tomadores de decisões. Contudo, se por um lado essas ferramentas parecem ser um poderoso aliado de gestores, por outro são várias as dificuldades encontradas para sua implementação e uso dentro de empresas, em particular as públicas. Assim, esse trabalho tem sua ênfase na identificação e análise das dificuldades sentidas por instituições do setor público brasileiro na implementação e utilização de *Data Warehouse*.

ABSTRACT

The demand of strategic information to support the decisions, in this third millennium, is going one of greater reasons to massive investments in informative solutions within institutions, public or private. Although many investments, most projects implemented in this institutions isn't satisfactory to decision making, principally concern to constant changes in the scenes of decisive process. In Public Institutions this adds the difficulties, inasmuch as it is common the politic decision in detriment of technical solutions. So we apperceive in this Institutions that various factors difficult the implementation of technologies which permit decision making, founded on facts, data and informations. To decide it is necessary to have information to allow the decision. Today has appearing a series of tools that assist the decision makers, the managers of the companies. Among the several existent tools today in the market, maybe due to the flexibility and to the focus centered in a continuous evolutionary process, what allows the development of applications returned to make decision, the one that it highlight is the one that allows the use of a development methodology based in *Data Warehousing* which makes possible the development of systems of information allowing the attendance of the changes of needs of the decision makers. However, if on one side those tools seem to be a powerful ally of managers, for other side they are several the difficulties found for its implementation and use inside of companies, in matter the Public Institutions. This way, that work has its emphasis in the identification and analyze of the difficulties felt by institutions of the Brazilian public sector in the implementation and use of *Data Warehouse*.

1 INTRODUÇÃO

Na medida em que os recursos públicos estão cada vez mais escassos, as Instituições Públicas devem procurar utilizá-los com maior eficácia e eficiência, para o melhor atendimento às demandas que a sociedade exige. Para tanto, os gestores dessas Instituições precisam de métricas que representem o que acontece nas suas unidades organizacionais conhecendo quando, onde e como os problemas ocorrem, identificando alternativas de soluções, através de análises e simulações, em várias perspectivas, melhorando as suas tomadas de decisões ((GIACOMONI, 2000), (KOHAMA, 2000), (FURLAN, 1997)).

Geralmente, as grandes Instituições Públicas incorrem em processos de entropia, onde as suas unidades procuram atingir as suas metas sem integração entre as mesmas. É preciso melhorar a comunicação entre as partes, identificando o desempenho de cada uma na Instituição em termos de eficácia e eficiência.

Os gestores da Instituição Pública precisam melhorar sua visão sistêmica do todo organizacional, através de informações sintéticas e representativas que, possibilitem gerenciar a atuação de suas unidades organizacionais, bem como, a análise de probabilidades em diversas perspectivas multidimensionais ((CHIAVENATTO,1993), (BIO,1989), (FURLAN,1997)).

Para se gerenciar bem é necessário que se tenha acesso aos dados transformados em informações, de forma tempestiva. O bom administrador é, principalmente, aquele que tem a sua disposição o maior número possível de informações sobre o seu negócio. Estar bem informado, hoje, é uma questão de vida ou morte para a maioria das empresas. Contudo, as instituições públicas, em particular as brasileiras, não podem se eximir de serem austeras, eficazes e eficientes em suas funções. Essa postura auxilia, no mínimo, na redução de custos e conseqüente redução do déficit público.

Ora, as tecnologias da Informação podem ajudar esse administrador nessa árdua tarefa. Tecnologias como *Data Warehouse*, *Data Mart* e *Datamining*, podem contribuir para uma boa administração e gerencia da instituição pública.

Esse trabalho, portanto, procurou analisar, à luz dessas tecnologias da informação, as dificuldades encontradas na implementação e utilização de *Data Warehouse* por parte de instituições públicas no Brasil.

1.1 Formulação do Problema

As unidades organizacionais de uma Instituição Pública, como nas grandes empresas, geralmente possuem sistemas de informações que suprem as suas necessidades de forma transacional e local, onde o produto desses sistemas é a informação operacional, reducionista e mecanicista, óbvia para o usuário, ocultando muitas vezes os resultados positivos e negativos. Os dados históricos e externos são mal persistidos senão, perdidos. Os gerentes dessa instituição dispõem de grandes custos para coletar, limpar, agregar, desagregar e analisar estes dados em diferentes perspectivas multidimensionais .

Algumas Instituições Públicas estão implementando e utilizando tecnologias de *Data Warehouse* – DW para a agilização e aprimoramento do processo decisório. DW é um “ambiente” para um grande banco de dados analítico da organização, onde os dados transacionais são extraídos de diversas fontes, limpos, armazenados, integrados, gerenciados e disponibilizados em interfaces amigáveis e flexíveis, permitindo aos gerentes a análise em diversas perspectivas multidimensionais, bem como o acesso em massa e com baixo tempo de resposta.

O sucesso na implementação de DW exige mudanças de paradigmas, comprometimento dos níveis tático-estratégico, ou seja, suporte político, bem como orçamento para aquisição de ferramentas e capacitação técnica ((FURLAN,1997), (INMON, 1997b), (KIMBALL, 1998b), (McGEE, 1997)).

Diante deste cenário, essa pesquisa pretende elucidar as seguintes perguntas:

Qual a importância da utilização de tecnologias de *Data Warehouse* para as Instituições Públicas Brasileiras?

Quais as dificuldades relacionadas à implementação de tecnologias de *Data Warehouse*, em Instituições Públicas no Brasil?

Quais os problemas enfrentados pelas Instituições Públicas Brasileiras que utilizam as tecnologias de Data Warehouse no seu dia-a-dia?

1.2 Justificativa / Relevância da Pesquisa

O atendimento das demandas crescentes da sociedade, pressiona o nível estratégico e os gerentes das Instituições Públicas a tomarem decisões em processos decisórios cada vez mais complexos, para tanto, necessitam de informações estruturadas, obtidas através dos sistemas de informações executivas (SIE) e dos sistemas de informações gerenciais (SIG), e principalmente, de informações tempestivas e não estruturadas, obtidas através dos sistemas de apoio à decisão (SAD). A tecnologia de DW se apresenta como a infra-estrutura de base de dados para os sistemas executivos, gerenciais e principalmente para os sistemas de suporte à decisão – DSS ((CHIAVENATTO,1993), (ANSOFF, 1990a), (FURLAN,1997), (INMON, 1997b)).

No nível operacional das grandes Instituições Públicas, como em todas as organizações empresariais, atuam diversos Sistemas de Informações de Processamento de Transações (SPT), também chamados de Processamento Transacional On-Line (*On-Line Transaction Processing* – OLTP). Os sistemas de processamento de transações não atendem como ferramentas de apoio à decisão.

Na maioria das Instituições Públicas as informações de suporte à decisão são derivadas de sistemas de informações não integrados, tais como: sistemas corporativos, gerenciais, estatísticos, e em muitos casos, através de planilhas eletrônicas, ferramentas estatísticas e outras. Geralmente, esses sistemas utilizam tecnologias de informação diferentes, sendo os dados coletados, preparados, armazenados por diferentes indivíduos, em lugares, épocas e formatos diferentes. Esses dados são voláteis ou mal persistidos, deixando de ser fontes históricas. O acesso restrito e a documentação desatualizada, ocasionam a retenção das informações pelas pessoas envolvidas diretamente com o sistema, o que dificulta a análise tempestiva (eventual e a tempo) para uma tomada de decisão imediata.

As Instituições Públicas também precisam trabalhar com as informações externas à organização, bem como, integrar as mesmas com as suas informações operacionais, gerenciais e executivas.

A integração de diversas fontes internas, bem como, das fontes externas, vai depender de compatibilizar a tecnologia, modelagem e formatação de dados. Geralmente, essas fontes possuem séries temporais diferentes e unidades de medidas diferentes.

Em muitas organizações empresariais, as tecnologias de DW são utilizadas como agente integrador de diversas fontes de informações. O resultado consiste em ter informações disponibilizadas de forma agregada, desagregada, resumida, sintetizada ou primitiva, com acesso rápido e fácil, escalabilidade e segurança. As tecnologias de DW implementadas nessas organizações, formam uma infra-estrutura de base de dados apropriada para os Sistemas de Suporte à Decisão – DSS, como os OLAP's - Processamento Analítico *On-line* - OLAP. Os OLAP's são interfaces e aplicações que disponibilizam informações com abordagem dimensional para o suporte à decisão. Essa infra-estrutura de base de dados também serve como uma fonte organizada para a utilização de tecnologias de *Datamining*. Essas tecnologias, através de inteligência aplicada, procuram o conhecimento escondido na base de dados bem como, padrões repetitivos e informação potencial, não trivial, onde ferramentas buscam padrões, buscam comportamento, sazonalidade, etc., reconhecendo esses padrões e buscando projeções futuras de acontecimentos.

As Instituições Públicas existem para atender determinadas necessidades da sociedade, e no atendimento dessas necessidades muitos padrões, comportamentos e sazonalidades poderiam estar sendo identificados, possibilitando com isso a aplicação dos recursos com maior efetividade ou seja, com maior eficácia e eficiência ((FURLAN,1997), (KIMBALL, 1998b), (INMON, 1997a), (STAIR, 1998), (LAUDON, 1999)).

Os clientes de uma Instituição Pública são os cidadãos, e estes a consideram um bem da sociedade. A sociedade necessita conhecer o papel, os produtos, a eficácia e eficiência da Instituição Pública. Aliando as tecnologias de *Internet* com as tecnologias

de DW, a Instituição Pública vai estar mais próxima do cidadão bem como, os seus executivos e gerentes que através da intranet da Instituição, poderão realizar análises eventuais e elaborar informações não-estruturadas de qualquer localidade que tenha acesso à Internet. Enfatizando este contexto, Kimball (1999), ao anunciar que a Internet está também mudando os DW's, afirma que:

“A grande novidade não é a tecnologia e sim o impacto da Internet. Esse baque esta sendo sentido de duas formas. Primeiro, todos os serviços de *Data Warehousing* estão sendo disponibilizados na Web. Os usuários esperam poder acessar o repositório de dados do trabalho, de casa e de diferentes locais remotamente. Em segundo lugar, os cliques nos sites da companhia estão se tornando uma importante fonte de dados. Nós chamamos isso de fluxo de cliques, que registra todos os gestos realizados de cada cliente. Esse fluxo vai estar muito relacionado ao faturamento de empresas de comercio eletrônico. Esse registro é talvez a maior fonte de dados potencial de qualquer *Data Warehouse*.”

Por outro lado, a atividade fim da maioria das Instituições Públicas não é a informática. Essas Instituições atendem à sociedade nas áreas de: educação, saúde, pesquisa, segurança, infra-estrutura e etc. Na maioria dessas Instituições a importância da informática está ligada ao apoio operacional. Geralmente, no organograma funcional dessas Instituições, a informática está subordinada a uma diretoria ou departamento de administração, onde o corpo técnico de informática é voltado para o desenvolvimento e manutenção de sistemas baseados em aplicações (requisitos). O ciclo de vida de desenvolvimento de sistema (System Development Life Cycle - SDLC), tradicionalmente utilizado no desenvolvimento de sistemas que são baseados em requisitos não é válido para o ambiente de DW. Para a implementação com sucesso de DW é necessário que a instituição tenha um corpo técnico capacitado e dedicado a essa tecnologia ((INMON, 1997a), (KIMBALL, 1998b), (McGEE, 1997)).

Grandes Instituições Públicas implementaram ou estão tentando implementar DW. Algumas Instituições Públicas possuem recursos orçamentários garantidos para a implementação, todavia o sucesso da implementação não está garantido (McGEE, 1997). A maioria das Instituições Públicas não dispõe de recursos orçamentários garantidos para investir na implementação de DW. Nesses casos, procuram alternativas de implementação próximas de sua realidade ((FURLAN,1997), (GIACOMONI, 2000), (KIMBALL, 1998a)).

Grandes Instituições Públicas podem possuir várias unidades de negócios diferentes. Uma das estratégias para atender separadamente as unidades de negócios são os *Data Mart*'s. Um *Data Mart* pode ser um subconjunto lógico de um completo DW. Em muitos casos, os executivos e gerentes desconhecem a importância da disponibilidade das informações analíticas representativas do negócio da Instituição, bem como, de ferramentas de DSS. O desenvolvimento de um protótipo de *Data Mart* pode convencer os dirigentes sobre a importância da tecnologia ((FURLAN,1997), (KIMBALL, 1998a)).

Portanto, tecnologias como essas podem se tornar essenciais um futuro próximo não só para instituições públicas bem como também para as instituições privadas. Tudo isso, e ainda o interesse do autor em trabalhar com *Data Warehouse* servem para justificar e mostrar a relevância do tema.

1.3 Objetivos

1.3.1 Geral

Analisar a implementação e utilização de tecnologias de *Data Warehouse* em Instituições Públicas no Brasil, identificando as dificuldades e problemas envolvidos.

1.3.2 Específicos

Identificar as dificuldades, com base na experiência pessoal de coletar informações, encontradas na implantação e utilização de *Data Warehouse* em instituições públicas brasileiras;

Avaliar os pontos críticos envolvidos na implementação e utilização de *Data Warehouse* em instituições públicas no Brasil;

Apresentar as necessidades e mostrar a importância de se ter disponível os dados de cunho gerenciais e científicos de forma integrada e acessível para análise sob diferentes perspectivas em apoio à tomada de decisões, para as instituições públicas;

Identificar os problemas específicos inerentes à implementação de DW em instituições públicas, no Brasil, tais como: coleta, preparação e armazenamento dos dados de várias fontes que possuam diferentes formatos de bases de dados e dados históricos, com séries temporais diferentes;

Analisar como o DW vem sendo implementado instituições públicas;

1.4 Estrutura do Trabalho

Esse trabalho analisa a implementação e utilização de tecnologias de *Data Warehouse* em Instituições Públicas no Brasil. Assim, os capítulos foram divididos de acordo com os temas pertinentes à realização do trabalho, em de seis capítulos distribuídos segundo a seqüência apresentada a seguir.

No primeiro capítulo é apresentada a introdução, sendo composta da formulação e relevância do problema, objetivos, justificativa e a organização do estudo.

No segundo capítulo são apresentados alguns conceitos inerentes às Instituições Públicas Brasileiras com o objetivo de dar ao leitor maiores subsídios para que o mesmo possa entender melhor o processo de tomada de decisão e liberação de verbas públicas, e, conseqüentemente as dificuldades enfrentadas pelas Instituições Públicas no processo de *Data Warehousing*.

No terceiro capítulo são apresentados alguns conceitos sobre o Ambiente de *Data Warehouse*, tais como: Características importantes; Estrutura; Arquitetura e infraestrutura; Estratégias de implementação; e Metodologias de desenvolvimento.

No quarto capítulo é apresentada a metodologia da pesquisa, onde são detalhados itens da pesquisa, tais como: Classificação da pesquisa; Universo e a amostra; Limites da pesquisa; Levantamento de dados; Variáveis da pesquisa; Análise e Interpretação dos Dados; Etapas da pesquisa; e Desenvolvimento da Pesquisa.

No quinto capítulo são apresentados os resultados das análises descritiva e referencial.

No sexto capítulo são apresentadas às conclusões e as recomendações para trabalhos futuros.

2 INSTITUIÇÕES PÚBLICAS NO BRASIL

2.1 Considerações Iniciais

Para que se possa entender melhor o processo de *Data Warehousing* nas Instituições Públicas no Brasil, é necessário conhecer as características das mesmas.

O conceito explícito de Instituição Pública no Brasil não foi encontrado na literatura pesquisada. Na tentativa de esclarecer esse conceito recorreu-se aos conhecimentos de serviço público, direito administrativo e administração pública. Através de revisão de literatura estes conhecimentos são descritos neste capítulo.

2.2 Serviço Público

Os serviços públicos são o conjunto de atividades e bens que são prestados pelo Estado ou delegados por concessão ou permissão sob condições impostas e fixadas por ele, à coletividade, visando proteger e proporcionar o maior grau possível de bem-estar social, ou “da prosperidade pública”. O Estado é a instituição prestadora de serviços que podem abranger atividades que competem exclusivamente ao Estado, e atividades exercidas por delegações do mesmo, atendendo a interesses coletivos ou ao bem estar geral (MEIRELLES, 1984).

Os serviços públicos podem ser classificados como próprios ou impróprios. Os serviços públicos próprios são aqueles que o Estado os executa diretamente por meio de seus agentes ou indiretamente por regime das concessões. Os serviços públicos impróprios são os que, embora atendendo às necessidades coletivas, como o os serviços diretos, não são executados pelo Estado, nem direta, nem indiretamente, mas devem receber autorização, permissão ou concessão do órgão estatal, a quem incumbe a tarefa de regulamenta-los ((CRETELLA,1995), (AMARAL,1996)).

2.2.1 Serviço Privativo do Estado

Na Constituição Federal¹, está especificado quais os serviços que exigem a centralização e competem exclusivamente ao *Estado*, que são os seguintes: os que dizem respeito às relações diplomáticas e consulares; os que se referem à defesa e segurança do território nacional; os concernentes a emissão de moeda e os de controle e fiscalização de instituições de crédito e de seguros; os de manutenção do serviço postal e do Correio Aéreo Nacional; os relativos ao estabelecimento e execução de planos nacionais de educação e saúde, bem como planos regionais de desenvolvimento; os que se relacionam ao poder de polícia e segurança pública; os que garantem a distribuição da justiça e outros que exigem medidas compulsórias em relação aos indivíduos.

2.2.2 Serviço de Utilidade Pública

Serviços de utilidade pública são os serviços prestados por delegação e regulamentação do Estado. Esses serviços são de interesse coletivo, portanto, o prestador não pode obter vantagens ou lucros em detrimento da coletividade, ainda quando dentro das possibilidades legais. Ao usuário é garantido não só o direito de obtenção e gozo, mas também a sua regular e permanente prestação. São exemplos típicos os serviços prestados a consumidores domiciliares, como eletricidade, telefone, água encanada, gás etc (KOHAMA, 2000).

2.2.3 Prestação de Serviço Mista

Segundo Kohama (2000), prestação de serviço mista é aquela que por dever é prestada pelo Estado e, entretanto, pode ser executada também por pessoa física ou jurídica de caráter privativo, independentemente de delegação para tanto.

A Constituição Federal do Brasil, especifica a prestação de serviço mista em relação à educação nos artigos 205 e 209, onde textualmente se vê:

Art. 205. A educação, direito de todos e dever do Estado e da família, será promovida e incentivada com a colaboração da sociedade, visando ao pleno

¹ *Constituição Federal*, artigo 21, incisos I, III, VIII, IX, X, XIII; e artigos 142 e 144

desenvolvimento da pessoa, seu preparo para o exercício da cidadania e sua qualificação para o trabalho.

Art. 209. O ensino é livre à iniciativa privada, atendidas as seguintes condições:

- I - Cumprimento das normas gerais de educação nacional;
- II - Autorização e avaliação de qualidade pelo Poder Público.

2.3 Administração Pública

Para Meirelles (1984), a Administração Pública é todo o aparelhamento do Estado, preordenado à realização de seus serviços, visando a satisfação das necessidades coletivas, em forma direta, contínua e permanente, e com sujeição ao ordenamento jurídico vigente.

Segundo Cretella (1995), Administração Pública não é só governo, poder executivo, a complexa máquina administrativa, o pessoal que a movimenta (conceito formal), como também a atividade desenvolvida (conceito material) por esse indispensável aparelhamento que possibilita o Estado o preenchimento de seus fins, através de atos concretos e executórios, para a consecução direta, ininterrupta e imediata dos interesses públicos.

2.4 Organização da Administração Pública Brasileira

A Constituição Federal do Brasil consagra os Poderes do Estado, inerentes ao Estado de Direito, através do seguinte texto: “São Poderes da União, independentes e harmônicos entre si, o Legislativo, o Executivo e o Judiciário.”

O Poder Legislativo tem a função normativa, isto é, aquela relativa a elaboração de leis. O Poder Executivo tem a função administrativa, ou seja, a execução da lei em ato individual e concreto. O Poder Judiciário tem a função judicial, ou a que se refere à aplicação aos litigantes. Um Poder investido na sua função, não poderá exercer a função de outro Poder, salvo exceções previstas em lei (KOHAMA, 2000).

Celso Antônio Bandeira de Mello (1979), por sua vez, afirma que de acordo com os termos do Decreto-lei 200, de 25/02/1967, a administração direta é realizada pelo conjunto de órgãos integrados nos Ministérios ou diretamente subordinados à Presidência da República. Administração Indireta é a desenvolvida por: autarquias; empresas públicas; sociedades de economia mista e fundações públicas.

A Administração Pública compreende os órgãos da Administração Direta ou Centralizada e os da Administração Indireta ou Descentralizada (Cretella, 1995).

2.4.1 Administração Direta ou Centralizada

Segundo Cretella (1995), classifica-se a administração em direta ou centralizada, quando o *serviço público* é gerido pelo próprio pessoal do *Estado* (União, Estados, Municípios), o “centro”, a pessoa política, a Chefia do Poder Executivo máxima e seus auxiliares, constituída dos serviços integrados na estrutura administrativa da Presidência da República e dos Ministérios, no âmbito federal, e do Gabinete do Governador e Secretarias de Estado, no âmbito estadual, e, na administração municipal, deve seguir estrutura semelhante.

2.4.2 Administração Indireta ou Descentralizada

Segundo Cretella (1995), a reforma administrativa², denominou a administração indireta como sendo as autarquias, as empresas públicas e sociedades de economia mista. As fundações públicas são uma espécie do gênero “autarquia”.

Kohama (2000), por sua vez, afirma que a administração indireta é composta por instituições com personalidade jurídica de direito público, também chamadas de *autarquias*, e as de direito privado, também chamadas de *paraestatais*.

² Decreto-lei nº 200, de 25 de fevereiro de 1967, dispõe sobre a organização Administração Federal, estabelece diretrizes para a Reforma Administrativa.

2.4.3 Instituições Públicas Autárquicas

Segundo Mello, C. (1993), as autarquias poderiam ser definidas como “pessoas jurídicas de direito público de capacidade exclusivamente administrativa”. O Decreto-lei 200 as caracteriza, nos seguintes termos: “o serviço autônomo, criado por lei, com personalidade jurídica, patrimônio e receitas próprios, para executar atividades típicas da Administração Pública, que requeiram, para seu melhor funcionamento, gestão administrativa e financeira descentralizada”.

Como definição, o enunciado normativo não vale nada. Sequer permite ao intérprete identificar quando a figura legalmente instaurada tem ou não natureza autárquica, pois deixou de fazer menção ao único traço que interessaria referir: a personalidade de direito público (MELLO, C, 1993).

O Quadro 1 mostra alguns exemplos de autarquias.

Quadro 1 - Alguns exemplos de autarquias.

Especialização	Autarquia
Autarquia Previdenciária	- Instituto de Assistência Médica e Previdência Social - Instituto de Previdência do Estado de São Paulo
Autarquias Profissionais	- Conselho Federal de Contabilidade - Conselho Federal de Engenharia - Ordem dos Advogados do Brasil
Autarquias Industriais	- Depto. Nacional de Estradas de Rodagem - Depto. Nacional de Águas e Energia Elétrica - Depto. Estadual de Estradas de Rodagem - Depto. Estadual de Águas e Energia Elétrica
Autarquias Especiais	- Banco Central do Brasil - Comissão Nacional de Energia Nuclear - Universidade de São Paulo – USP - Universidade de Campinas – UNICAMP
Autarquias Educacionais	- Faculdades Isoladas

Fonte: KOHAMA (2000)

2.4.4 Instituições Públicas Paraestatais

Kohama (2000, p. 41), conceitua instituições paraestatais como “pessoas jurídicas de direito privado, cuja criação é autorizada por lei, com patrimônio público ou misto, para a realização de atividades, obras ou serviços de interesse coletivo, sob normas e controle do Estado”. Essas instituições podem executar atividades econômicas com fins lucrativos, como podem perseguir fins não lucrativos, para tanto deverá ser escolhida a estrutura e a organização adequadas a cada finalidade. As instituições

parestatais da administração indireta ou descentralizada, podem ser constituídas das seguintes formas:

- Empresas Públicas;
- Sociedades de Economia Mista; e
- Fundações.

2.4.5 Empresas Públicas

Para Mello C. (1993), Empresa Pública Federal é a pessoa jurídica criada por lei como instrumento de ação do *Estado*, com personalidade de direito privado, mas submetida a certas regras especiais decorrentes de ser coadjuvante da ação governamental, são constituídas sob quaisquer das formas admitidas em direito e cujo capital seja formado unicamente por recursos de pessoas de direito público interno ou de pessoas de suas administrações indiretas, com predominância acionária residente na esfera federal.

O Quadro 2 mostra alguns exemplos de empresas públicas.

Quadro 2 - Alguns exemplos de empresas públicas.

Esfera Administrativa	Empresas Públicas
A nível da União	- Rede Ferroviária Federal - Casa da Moeda do Brasil - Caixa Econômica Federal - Empresa Brasileira de Correios e Telégrafos
A nível do Estado	- Caixa Econômica Estadual - Imprensa Oficial do Estado - Companhia de Metropolitano de São Paulo
A nível do Município	- Empresa Municipal de Urbanismo – EMURB

Fonte: KOHAMA (2000)

2.4.6 Sociedade de Economia Mista

Para Mello, C. (1993), deve-se entender Sociedade de Economia Mista como a pessoa jurídica criada por lei, como um instrumento de ação do *Estado*, dotada de personalidade de direito privado, mas submetida a certas regras especiais decorrentes desta sua natureza auxiliar da atuação governamental, constituída sob a forma de sociedade anônima cujas ações com direito a voto pertençam em sua maioria à União ou

instituição de sua administração indireta. O Quadro 3 mostra alguns exemplos de Sociedades de Economia Mista.

Quadro 3 - Alguns exemplos de Sociedades de Economia Mista.

Esfera Administrativa	Sociedades de Economia Mista
Ao nível da União	- Petrobrás – Petróleo Brasileiro S.A. - Banco do Brasil S.A. - Telesp – Telecomunicações de São Paulo S.A.
Ao nível do Estado	-Eletropaulo – Eletricidade de São Paulo S.A. -Banespa – Banco do Estado de São Paulo S.A.
Ao nível do Município	-Prodram-SP – Cia. de Processamento de Dados do Município de São Paulo

Fonte: KOHAMA (2000).

2.4.7 Fundações Públicas

O Decreto-lei 200, modificado pela Lei 7.596, de 10/04/87, define a fundação pública como: “a entidade dotada de personalidade jurídica de direito privado, sem fins lucrativos, criada em virtude de autorização legislativa, para o desenvolvimento de atividades que não exijam execução por órgãos ou entidades de direito público, com autonomia administrativa, patrimônio gerido pelos respectivos órgãos de direção, e funcionamento custeado pela União e de outras fontes”.

Para MELLO, C. (1993), é absolutamente incorreta a afirmação normativa de que as fundações públicas são pessoas de direito privado. Na verdade, são pessoas de direito público, consoante, aliás, universal entendimento, que só no Brasil foi concedido. Se lhe atribuiu a titularidade de poderes públicos (e não meramente o exercício deles) e disciplinou-a de maneira a que suas relações sejam regidas pelo direito público, a pessoa será de direito público, ainda que se lhe atribua outra qualificação. Entretanto, foram batizadas de pessoas de direito privado, apenas para se evadirem destes controles moralizadores ou, então, para permitir que seus agentes acumulassem cargos e empregos, o que lhes seria vedado se fossem reconhecidas como pessoas do direito. O Quadro 4 mostra alguns exemplos de fundações públicas.

Quadro 4 - Alguns exemplos de fundações públicas.

Esfera Administrativa	Fundações Públicas
Ao nível da União	- Funai – Fundação Nacional do Índio (Assistência Social) - Fundação Universidade de Brasília (Ensino)
Ao nível do Estado	- Fundação Padre Anchieta (Educação) - Fundação de Amparo a Pesquisa do Estado de São Paulo

Fonte: KOHAMA (2000).

2.4.8 Serviços Sociais Autônomos

Para Kohama (2000), os Serviços Sociais Autônomos são aqueles autorizados por lei, com personalidade de direito privado, com patrimônio próprio e administração particular, com a finalidade específica de assistência ou ensino a certas categorias sociais ou determinadas categorias profissionais, sem fins lucrativos. Embora entidades paraestatais, oficializadas pelo Estado, os Serviços Sociais Autônomos não fazem parte integrante da Administração Indireta ou Descentralizada, mas trabalham ao lado do Estado, e em virtude do interesse coletivo dos serviços que prestam, o poder público as autoriza e as ampara, através de dotações orçamentárias ou contribuições parafiscais, para a sua manutenção. Como entidades paraestatais que recebem dinheiro público, sujeitam-se a prestação de contas ao órgão estatal a que estejam vinculadas e por ele são supervisionadas. O Quadro 5 mostra alguns exemplos de Serviços Sociais Autônomos.

Quadro 5 - Alguns exemplos de Serviços Sociais Autônomos.

Sigla	Serviços Social Autônomo
SESI	- Serviço Social da Indústria
SESC	- Serviço Social do Comércio
SENAI	- Serviço Nacional de Aprendizagem Industrial
SENAC	- Serviço Nacional de Aprendizagem Comercial
SEBRAE	- Serviço Brasileiro de Apoio às Micro e Pequenas Empresas
	- Associação das Pioneiras Sociais

Fonte: KOHAMA (2000).

2.5 Considerações Finais

Na busca da conceituação de Instituição Pública no Brasil, neste capítulo foram abordados os conhecimentos sobre Serviço Público, Administração Pública e Organização da Administração Pública Brasileira. Foram classificados e caracterizados os tipos de Serviço Público bem como, as Instituições Públicas no Brasil.

Ao contrário, a Administração Pública hoje é conhecida por sua morosidade e pela má aplicação de verbas públicas. Normalmente as informações necessárias à tomada de decisão estão espalhadas por vários sistemas ou simplesmente inexistem em

algumas instituições, não há dados suficientes para uma tomada de decisão eficiente, o que leva muitas vezes a decisões de cunho político.

A sociedade de modo geral, não mais aceita a má administração das Instituições Públicas, exigindo respostas rápidas e eficientes. Caberia aos administradores dessas instituições procurar maneiras mais eficientes de obter as informações estratégicas de cada Instituição, para com isso poderem tomar decisões mais acertadas.

Verifica-se então que os administradores das Instituições Públicas estão diante de um dilema, por um lado eles precisam ampliar e modernizar os serviços para melhor atender à sociedade e por outro lado possuem um orçamento limitado. Para resolver esse problema, esses administradores necessitam de informações de apoio à decisão inclusive para decidir como melhor alocar os recursos disponíveis. As tecnologias de *Data Warehouse* despontam como alicerce aos sistemas de apoio à decisão. No capítulo seguinte são abordados os conhecimentos sobre *Data Warehouse*.

3 AMBIENTE DE DATA WAREHOUSE

3.1 Considerações Iniciais

Segundo Adelman & Lebaron (1997), o Ambiente de Apoio à Decisão é constituído pelos Sistemas de Apoio à Decisão e Sistemas de Informações Estratégicas. Estes sistemas basicamente promovem apoio à decisão, utilizando ferramentas e métodos de análise sobre os dados coletados nas empresas. São desenvolvidos com intuito de dar suporte aos gerentes e pessoas responsáveis pela tomada de decisão, permitindo assim uma visualização de vários aspectos do problema a ser analisado.

A multiplicação dos Sistemas de Apoio à Decisão resultou em dificuldades no controle e gerenciamento de dados. Nesse contexto surge o ambiente de *Data Warehouse*. Essa tecnologia é uma combinação de conceitos destinados ao suporte da decisão, com o propósito de habilitar o usuário final a tomar decisões melhores e mais rápidas.

O ambiente de *Data Warehouse* passou a funcionar como fonte de dados para os Sistemas de Apoio à Decisão. As preocupações sobre onde buscar as informações e quando carregá-las ficam sobre responsabilidade do DW, permitindo aos Sistemas de Apoio à Decisão se preocuparem apenas com as questões sobre o processamento de consultas.

Este capítulo aborda os conhecimentos sobre DW que envolve os conceitos, tais como: características, arquitetura e infra-estrutura, estratégias de implementação, metodologias de desenvolvimento e fatores críticos de sucesso no desenvolvimento e manutenção do ambiente de *Data Warehouse*.

3.2 O que é Data Warehouse?

Data Warehouse - DW é um ambiente estruturado projetado formando um consistente depósito de dados não-volátil, com dados integrandos de múltiplas fontes de aplicação, incluindo dados históricos e dados externos, com longo horizonte temporal,

disponíveis e sumarizados para análises rápidas, auxiliando os analistas, gerentes e executivos da organização nos processos decisórios e no contexto dos negócios da empresa ((CW, 2000), (FEW, 2001), (INMON, 2001), (SAMANT, 2000), (SADANANDAN, 2001)).

Segundo Singh (2001 p. 12), DW é uma tecnologia de gestão e análise de dados, constituindo “um ambiente de *suporte a decisão* que alavanca dados armazenados em diferentes fontes e os organiza e entrega aos tomadores de decisões da empresa, independente de plataforma que utilizam ou de seu nível de qualificação técnica”.

O propósito do DW não é suportar as operações ou transações da organização, mas centralizar os dados extraídos de aplicações operacionais e bancos de dados. Em geral, um DW requer a consolidação de outros recursos de dados além dos armazenados em banco de dados relacionais, incluindo informações provenientes de planilhas eletrônicas, documentos textuais, etc. O DW não é construído para suportar o processo funcional ou operacional da empresa, ou seja, não é o fim, mas é o meio para facilitar o uso da informação, como mostra a Figura 1 ((INMON, 1993), (KIMBALL, 1998b), (SINGH, 2001)).

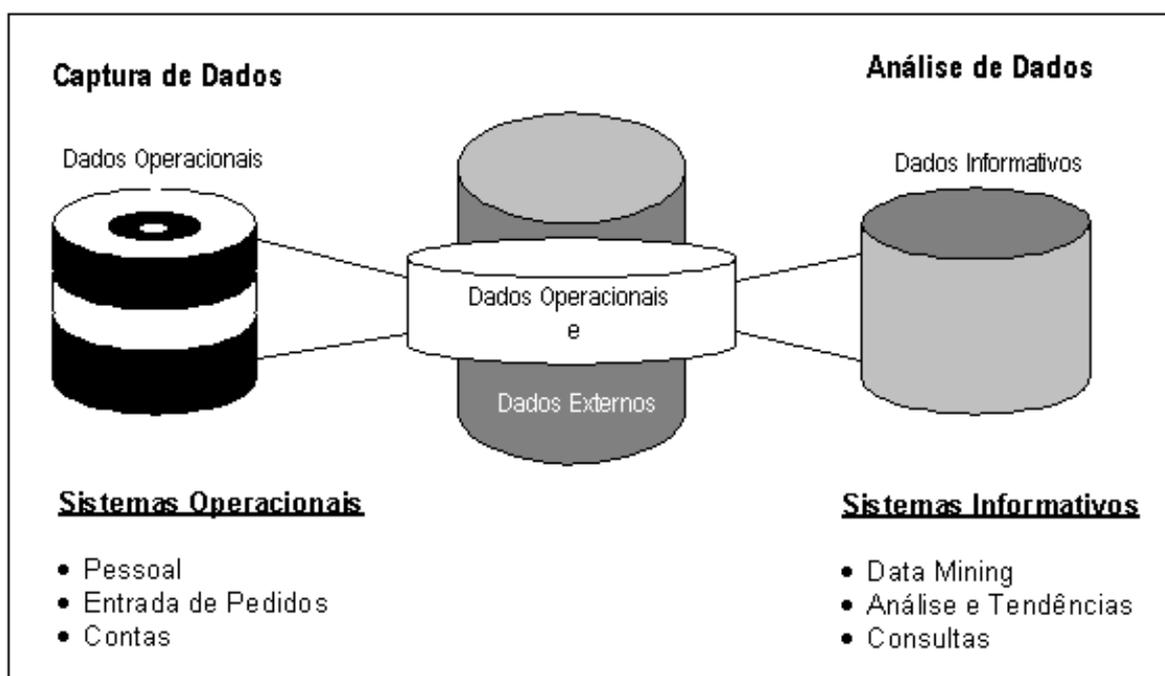


Figura 1 - DW - Dados para informação.
Fonte: (SINGH, 2001).

3.3 Características Importantes do *Data Warehouse*

Inmon (1993), define o *Data Warehouse* como “um conjunto de dados com as seguintes características: orientados por assunto ou negócio, integrados, variáveis com o tempo e não-voláteis, e que fornecem suporte ao processo de tomada de decisão do negócio”, como visto na Figura 2. A seguir serão detalhadas essas características.

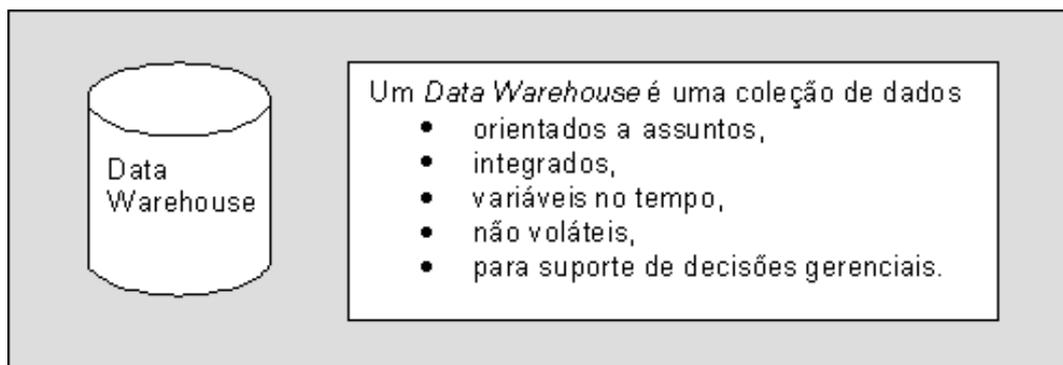


Figura 2 - Definição do *Data Warehouse*.
Fonte: (INMON, 1997a).

3.3.1 Orientado a assunto

Os dados operacionais ou funcionais são organizados em torno das aplicações da empresa, executam e registram as transações rotineiras necessárias para conduzir o negócio. O DW baseia-se nos principais assuntos ou negócios da organização que tenham sido definidos no modelo de dados ((INMON, 1997a), (LAUDON, 1999)).

Os dados orientados a aplicações possuem detalhes que satisfazem os requisitos imediatistas do processamento funcional. Esses detalhes podem ser irrelevantes ao analista de SAD. O DW não inclui dados que não serão usados para processamento SAD (INMON, 1997b).

Inmon (1997a), exemplifica o caso de uma companhia de seguros, onde os dados são organizados em torno de aplicações (SPT) que podem ser automóvel, saúde, vida e perdas. O DW da companhia é baseado em assuntos ou negócios da empresa que podem ser cliente, apólice, prêmio indenização, como mostra a Figura 3.

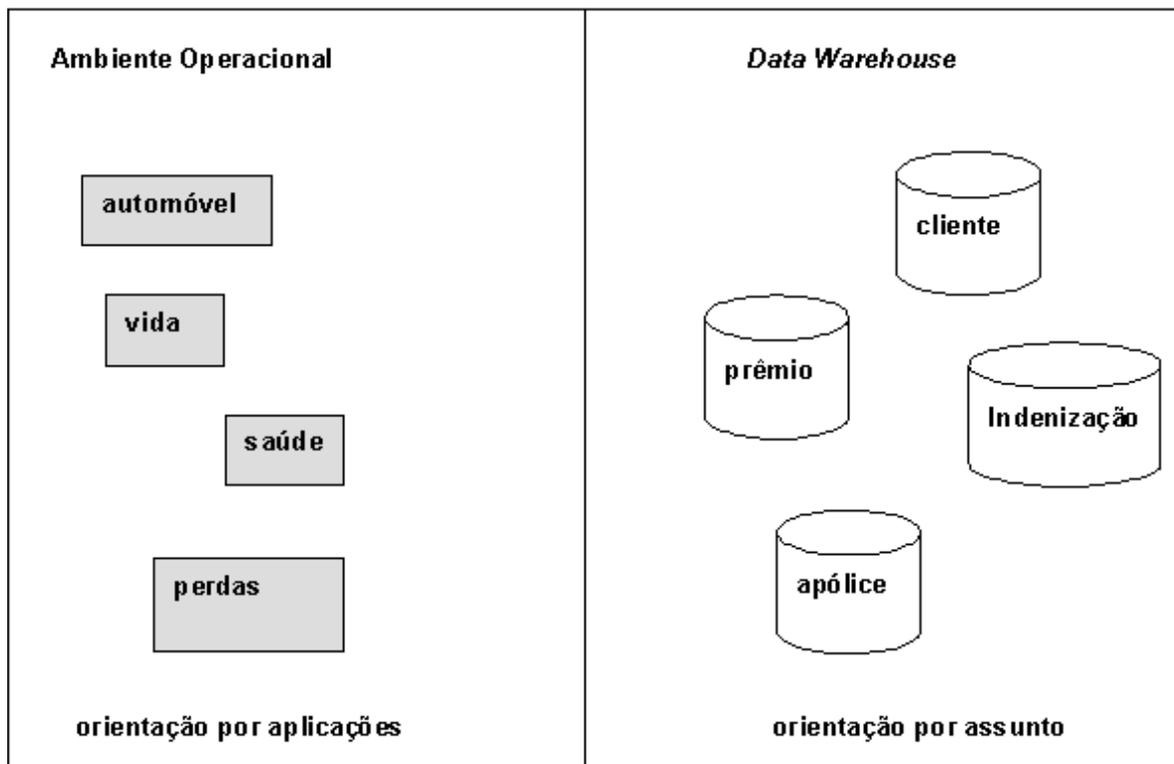


Figura 3 - Orientação por assunto do DW.

Fonte: (INMON, 1997a).

3.3.2 Integrado

A característica mais importante do ambiente de DW, é que por dentro, o DW é integrado. A integração de dados ocorre quando os dados são passados do ambiente operacional baseado em aplicações para o ambiente de DW (INMON, 1997a).

As aplicações não apresentam coerência, surgem ao longo dos anos de sob formas diferentes. A integração dos dados de diversas aplicações é feita de diferentes formas na consistência de nomes, na consistência de variáveis de medidas, na consistência da codificação das estruturas, na consistência dos atributos físicos dos dados e assim por diante. Outro aspecto importante é que a codificação para o DW deve ser feita de forma independente da aplicação de origem ((INMON, 1997a), (INMON, 1997b)). A Figura 4, mostra a questão da integração.

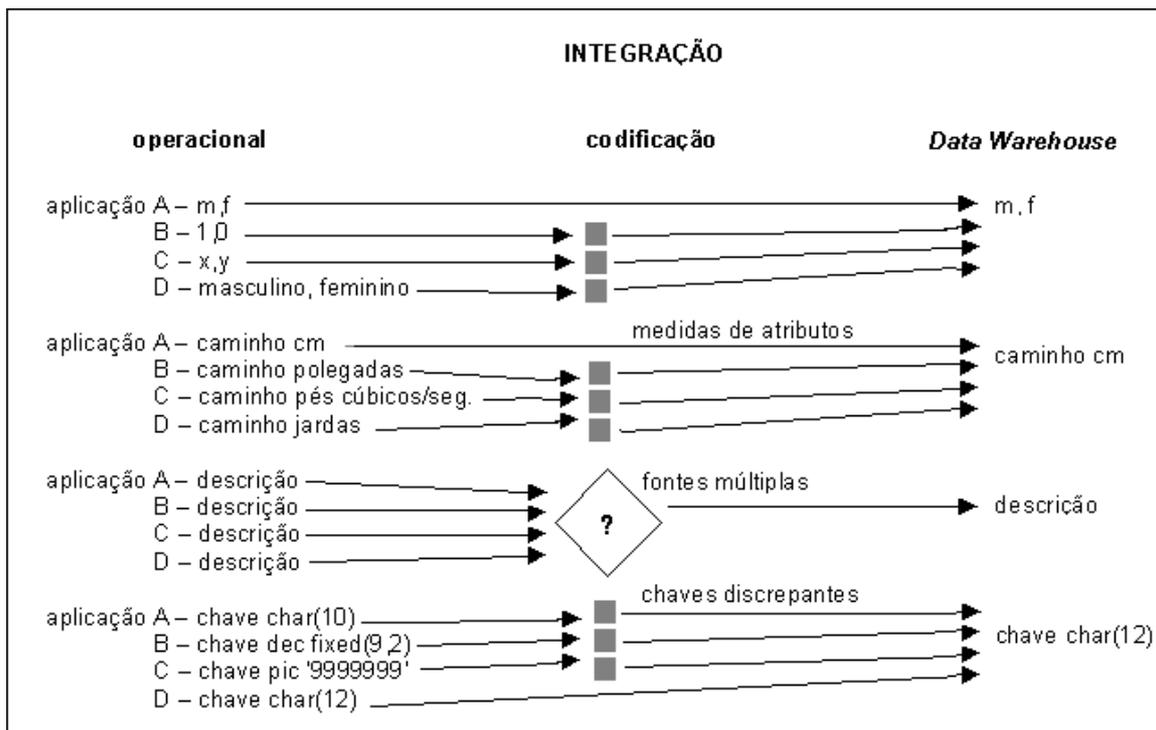


Figura 4 - A questão da integração - Fonte: (INMON, 1997a).

Os processos de extração, transformação e carga (ETC), executam a grande e complexa tarefa da passagem dos dados do ambiente operacional herdado para o ambiente DW. O Quadro 6 apresenta algumas funcionalidades necessárias aos processos de ETC (INMON, 1997a). O Quadro 8 na descrição do componente “*Back Room*” do modelo de arquitetura definido por Kimball, descreve os processos de ETC (KIMBALL, 1998a).

Quadro 6 - Algumas funcionalidades necessárias ao ambiente DW na passagem dos dados do ambiente operacional para o ambiente DW

Funcionalidades necessárias
• A extração de dados do ambiente operacional para o ambiente de DW demanda mudança de tecnologia, por exemplo, SGBD de DW.
• A seleção e validação dos dados do ambiente operacional.
• As chaves de entrada operacionais geralmente precisam ser reestruturadas antes de serem gravadas. Em casos simples, um elemento de tempo é acrescentado à estrutura da chave.
• Os dados são reformatados .
• Trabalhar com várias fontes de dados. A lógica deve ser esclarecida para que a fonte de dados apropriada contribua com seus dados segundo o conjunto correto de condições.
• Intercalar vários arquivos de entrada.
• Trabalhar com vários resultados de diferentes níveis de resumos produzidos pelo mesmo programa de carga.
• A eficiência na escolha dos dados de entrada para extração.
• Frequentemente é necessário resumir dados.
• A alteração de nomes de elementos de dados durante a passagem do ambiente operacional para o ambiente de <i>Data Warehouse</i> deve ser registrada.
• Os registros de entrada com formatos não-padronizados.

<ul style="list-style-type: none"> • A conversão precisa ser feita.
<ul style="list-style-type: none"> • Talvez o pior de tudo: Os relacionamentos entre os dados, que embutidos nos antigos programas herdados precisam ser compreendidos. Geralmente esses relacionamentos são misteriosos, difíceis de alterar e não-documentados.
<ul style="list-style-type: none"> • Processamento com volumes massivos de entrada de dados. Implementação de cargas em paralelo ou leitura em paralelo.
<ul style="list-style-type: none"> • O projeto do DW deve adequar-se ao modelo de dados corporativos. A entrada para o DW ajusta-se às especificações de uma aplicação que foi escrita há muito tempo.

Fonte: (INMON, 1997a).

3.3.3 Variação com o tempo

Segundo Singh (2001, p.14) “variação com o tempo significa que os dados estão associados a um ponto no tempo(ou seja, semestre, ano fiscal e período de pagamento)”.

Para Inmon (1997b), todos os dados no DW são precisos em algum instante no tempo. No ambiente operacional, os dados estão corretos como no momento do acesso. Em virtude dos dados no DW serem corretos como em algum momento no tempo, “não exatamente agora”, é dito que estes dados “variam com o tempo” (INMON, 1997b). A Figura 5 apresenta a variância no tempo dos dados do DW.

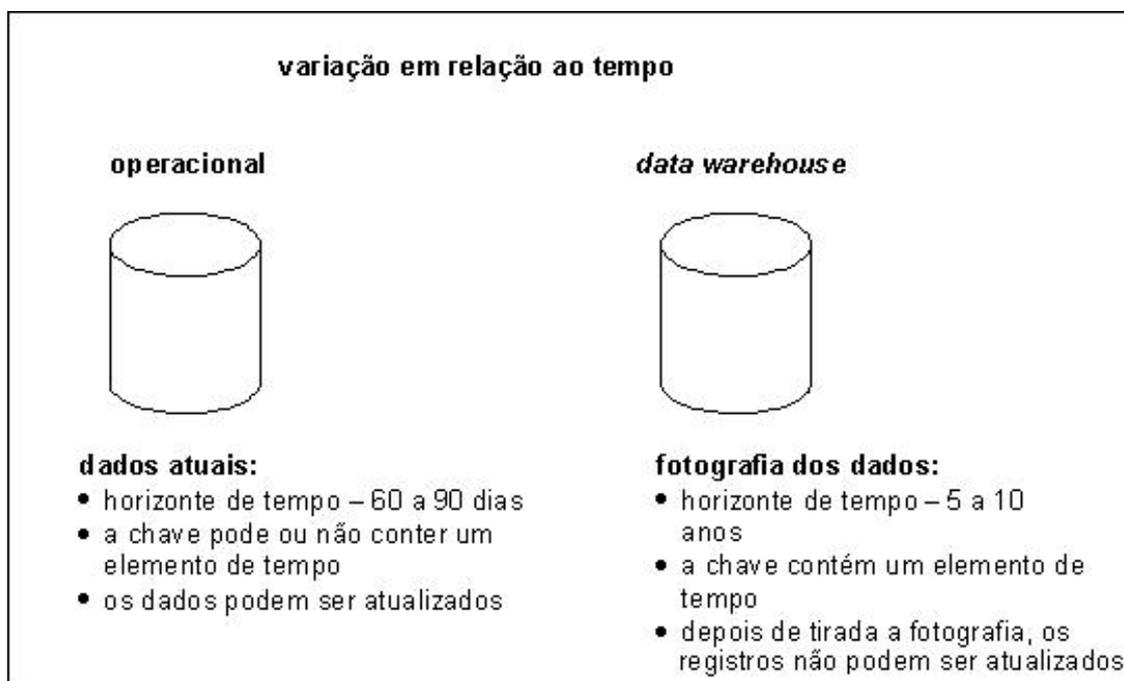


Figura 5 - A questão da variação em relação ao tempo.

Fonte: (INMON, 1997b).

Segundo Kimball (1998b, p. 6)

“o melhor sistema OLTP é um status instantâneo dos negócios de uma organização, atualizado constantemente à medida que as transações são concretizadas(...) Por exemplo, à medida que os pedidos são atendidos, uma relação de pedidos a atender muda. Os relacionamentos de fornecedores de peças pode estar mudando enquanto se observa o banco de dados.”

Esse tipo de banco de dados OLTP é chamado de banco de dados instantâneo (*twinkling database*).

“Essa natureza instantânea dos bancos de dados OLTP é o primeiro tipo de inconsistência temporal que evitamos em *Data Warehouse*. O segundo tipo de inconsistência temporal em um banco de dados OLTP é a falta de suporte explícito para representar corretamente o histórico do passado.”

No DW o armazenamento é uma seqüência de tempo explícita, onde são transferidos instantâneos estáticos (*static snapshots*) do OLTP para o DW como uma seqüência de camada de dados, muito semelhante às formações geológicas. O DW é uma seqüência de tempos, e como os geólogos, pode-se escavar as camadas para saber como eram os negócios do passado. Essa técnica de DW é chamada de mudança gradual nos eixos temporais (*slowly changing dimensions*), é considerada a principal técnica para representar o passado, onde as duas inconsistências temporais nos bancos de dados OLTP são resolvidas. A primeira, o DW não se modifica durante o dia quando os usuários estão fazendo as suas consultas. A segunda, armazenando as informações cuidadosamente em cada instantâneo no DW, pode-se representar todos os pontos de tempo anteriores corretamente (KIMBALL, 1998b).

3.3.4 Não-volátil

No ambiente DW, não-volátil significa que os dados não são alterados depois de inseridos. Os bancos de dados operacionais/OLTP mudam todas as vezes que uma transação é processada. No ambiente DW, alterações instantâneas não são permitidas ((SINGH, 2001), (TANLER, 1998), (KIMBALL, 1998b)).

Inmon (1997a), afirma que no ambiente de DW os dados são inseridos e acessados em grandes quantidades, mas a atualização dos dados geralmente não ocorre no ambiente de DW. No ambiente operacional, os dados são regularmente acessados um registro por vez e podem ser atualizados. A Figura 6 mostra a questão não-volatilidade.

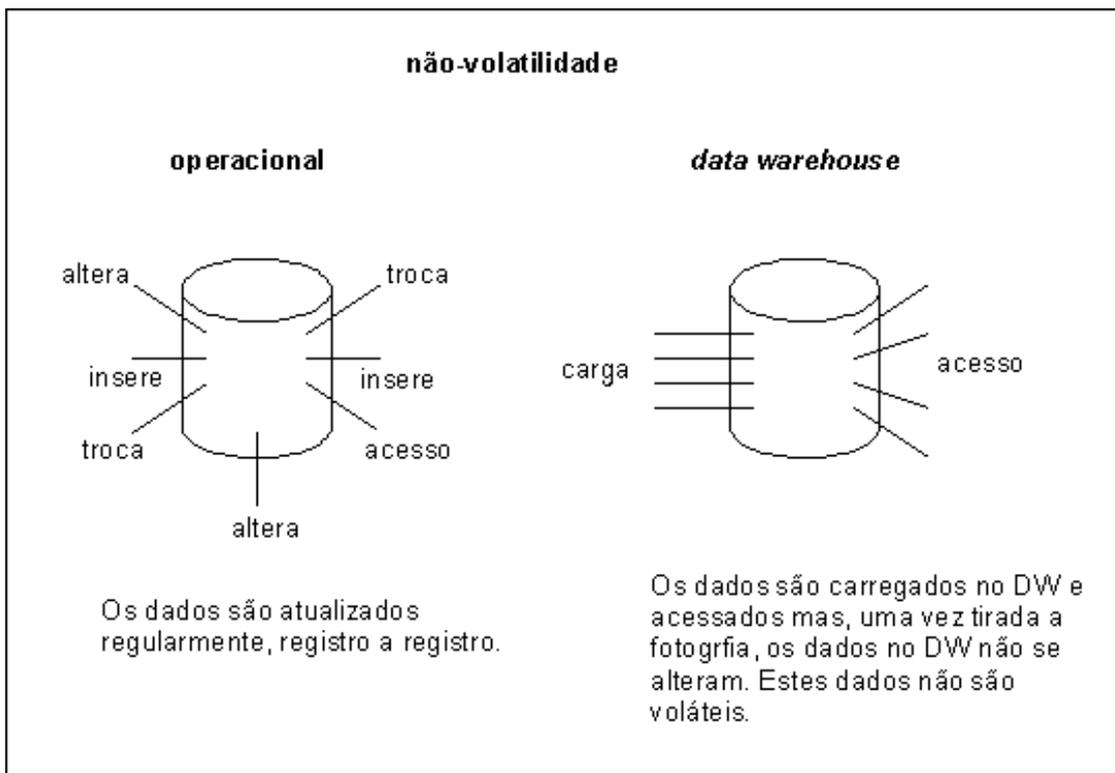


Figura 6 - A questão da não-volatilidade.
Fonte: (INMON, 1997a).

3.4 A Estrutura do *Data Warehouse*

O DW possui uma estrutura distinta. Há diferentes níveis de sintetização e detalhe que descrevem o DW, e, adicionalmente, diferentes níveis de idade dos dados (INMON, 1997a). A Figura 7 mostra a estrutura geral e os componentes do DW:

3.4.1 Dados Atuais

Os dados detalhados atuais são os que exigem mais atenção, por diversas razões (INMON, 1997b):

- Estes dados refletem os acontecimentos mais recentes, sempre de grande interesse;
- São volumosos devido serem armazenados no nível mais baixo de granularidade;
- Geralmente são armazenados em disco, um meio de acesso rápido mas de gerenciamento difícil e caro.

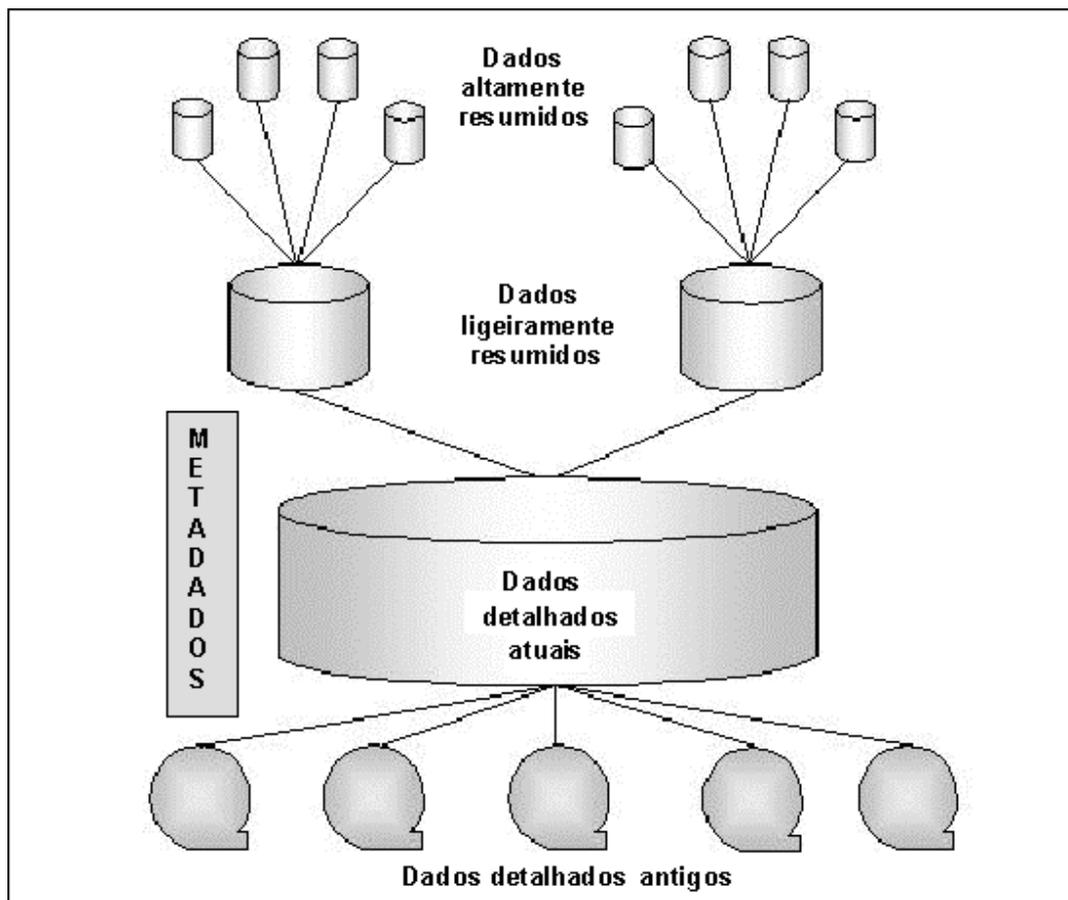


Figura 7 - A estrutura de dados dentro do DW.
Fonte: (INMON, 1997b).

3.4.2 Dados Detalhados Antigos

Os dados detalhados antigos são guardados em meios de armazenamento de massa, sendo acessados sem muita frequência e armazenados em um nível de detalhe consistente com o detalhe dos dados atuais. Geralmente estes dados são armazenados em algum meio alternativo que não seja em disco, em função do grande volume de dados e do acesso não frequente (INMON, 1997b).

3.4.3 Dados Ligeiramente Resumidos

Os dados neste nível são resumidos dos detalhes dos dados atuais. Na construção deste nível o arquiteto de dados geralmente enfrenta os seguintes problemas (INMON, 1997b):

- Qual o espaço de tempo que o resumo será feito?

- Qual o conteúdo e atributos dos dados ligeiramente resumidos?

3.4.4 Dados Altamente Resumidos

Os dados altamente resumidos são compactos e de fácil acesso. Às vezes estes dados são encontrados dentro das paredes da tecnologia que cerca o ambiente de DW e em outros casos fora. Independente da localização, os dados altamente resumidos fazem parte do DW (INMON, 1997b).

Singh (2001), afirma que nem todos os resumos realizados são armazenados no DW. Haverá muitas ocasiões em que a análise será feita em um tipo ou outro de resumo será produzido. Somente os resumos utilizados com frequência serão armazenados no DW.

3.4.5 Metadados

Os metadados são dados sobre os dados do DW, constituindo assim o o componente mais importante da estrutura do DW. Os metadados não contêm dados retirados diretamente do ambiente operacional, portanto se situam em uma dimensão diferente dos outros dados do DW. Os metadados contêm, pelo menos, as seguintes informações ((INMON, 1997b), (INMON, 1997a)):

- A estrutura dos dados;
- Os Algoritmos usados para a sintetização;
- A fonte de dados que alimenta o DW;
- O mapeamento desde o ambiente operacional até o DW;
- O modelo de dados;
- O relacionamento entre o modelo de dados e o DW;
- O histórico de extrações.

É necessário entender quais os dados que estão disponíveis no DW e onde estão localizados, para acessar com eficiência os dados do DW. Os metadados constituem um catálogo dos dados do DW e os indicadores/ponteiros para esses dados. Os metadados também podem conter as seguintes informações (SINGH, 2001):

- Estatística de uso de dados;
- Medições de desempenho;

- Diferentes elementos por atributo.

As informações contidas nos metadados têm um papel muito importante no DW e são utilizadas de diversas maneiras (INMON, 1997b):

- No auxílio dos analistas de SAD na localização dos dados do DW;
- Como guia para o mapeamento dos dados em sua transformação desde o ambiente operacional até o DW;
- Como guia para os algoritmos utilizados para sintetização entre os dados detalhados atuais e os dados altamente resumidos; etc.

3.5 Arquitetura de *Data Warehouse*

Segundo Singh (2001), arquitetura é um conjunto de regras que regulamentam uma estrutura para um produto ou projeto de um sistema. Os fundamentos da arquitetura de DW são muito mais importantes do que as ferramentas específicas para implementação da mesma. O principal fundamento de qualquer DW é ter flexibilidade suficiente para que a organização de acordo com as suas necessidades, possa modificar e analisar os seus dados. A arquitetura de DW deve ser custo-eficiente, adaptável e de fácil implementação (SINGH, 2001). A Figura 8 ilustra vários componentes de uma arquitetura de dados que devem ser considerados antes de se implementar qualquer DW.

Para Inmom, o design de arquitetura do DW deve ser voltado para satisfazer a necessidade de informação da organização não se restringido por nenhuma tecnologia. Uma arquitetura física de cliente/servidor possibilita separar a aquisição de dados, os requerimentos de processamento, bem como o acesso a dados e as funções de processamento de manipulação de dados (INMON, 1999).

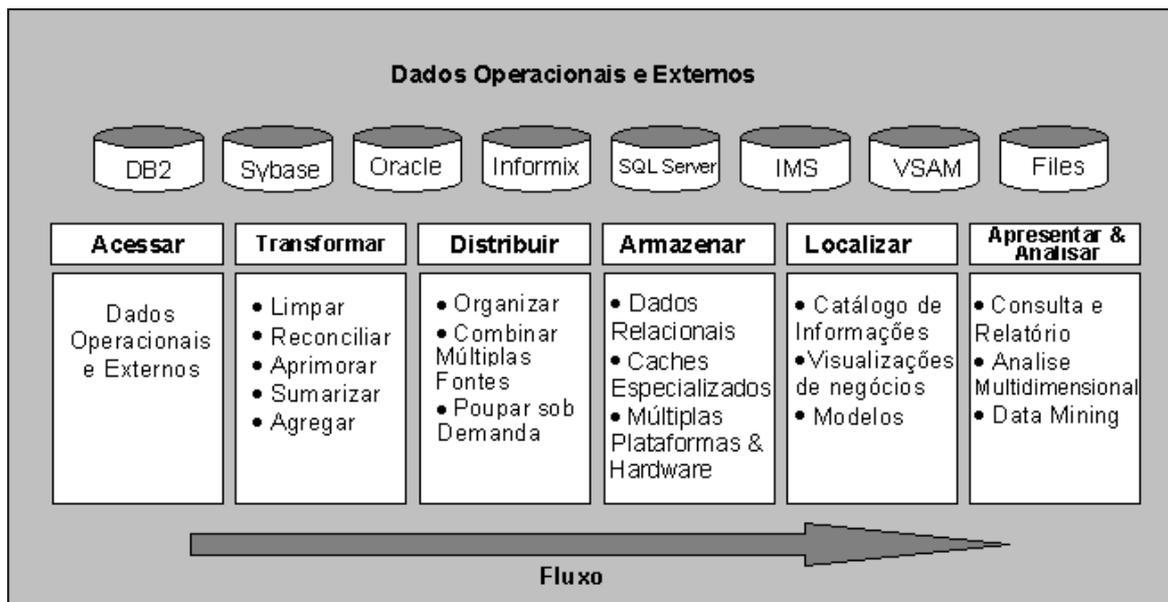


Figura 8 - Componentes de uma arquitetura que devem ser considerados antes da implementação do DW.

Fonte: (SINGH, 2001).

Kimball, destaca a importância do plano de arquitetura na construção do projeto de DW como ferramenta de comunicação, planejamento e flexibilidade, que facilita o aprendizado e aumenta a produtividade (KIMBALL, 1998a).

3.5.1 Modelo de Arquitetura Genérica de Data Warehouse

Orr (1996), propôs um modelo de arquitetura genérica de DW, com o objetivo de representar a funcionalidade de um DW, permitindo que diferentes abordagens de arquiteturas, onde um único componente de software pode atender várias camadas propostas. O Quadro 7 descreve cada camada do modelo de arquitetura proposto por Orr e ilustrado na Figura 9.

Quadro 7 - Descrição das camadas do modelo de arquitetura proposto por Orr.

Camada	Descrição da camada
Dados Operacionais / Dados Externos	Constitui-se dos dados dos SPT/OLTP e informações provenientes de fontes externas.
Acesso à Informação Information Access	Constitui-se de ferramentas (SAD, SIE, Data Mining), utilizadas para obtenção de informações executivas e informações para apoiar a tomada de decisões, que possibilitam a manipulação, análise e apresentação dos dados.
Acesso ao Dados - Data Access	Esta camada funciona com uma interface entre as ferramentas de acesso à informação e os bancos de dados operacionais. Permite que todos os dados da organização seja acessíveis, o que se chama "acesso universal de dados".
Dicionário de Dados (Metadados) -	Metadados são os dados sobre os dados da organização, sendo fundamentais para possibilitar o acesso universal dos dados. É necessário

Data Directory	que exista uma grande variedade de metadados no ambiente de DW para se ter um DW totalmente funcional.
Gerenciamento de Processo - Process Management	Gerencia várias tarefas que devem ser cumpridas para construir e manter as informações do DW e do dicionário de dados (Metadados).
Transporte - Application Messaging	Gerencia o transporte de informações pelo ambiente de rede. Pode ser usada para isolar aplicações, operacionais e informativas, a partir do formato exato em cada ponta. Também é usada na coleta de mensagens e transações e entregá-las em locais e tempos determinados.
Gerenciamento de Replicação - Data Staging	Consiste dos processos necessários para selecionar, editar, sumarizar, combinar e carregar dados operacionais e/ ou externos no DW.
Data Warehouse (físico)	É o local onde estão fisicamente os dados utilizados para obter informações. Em muitas situações os dados podem estar fisicamente armazenados no ambiente transacional e externo, mas são representados e acessados no DW como visões lógica ou virtual dos dados.

Fonte: ORR (1996)

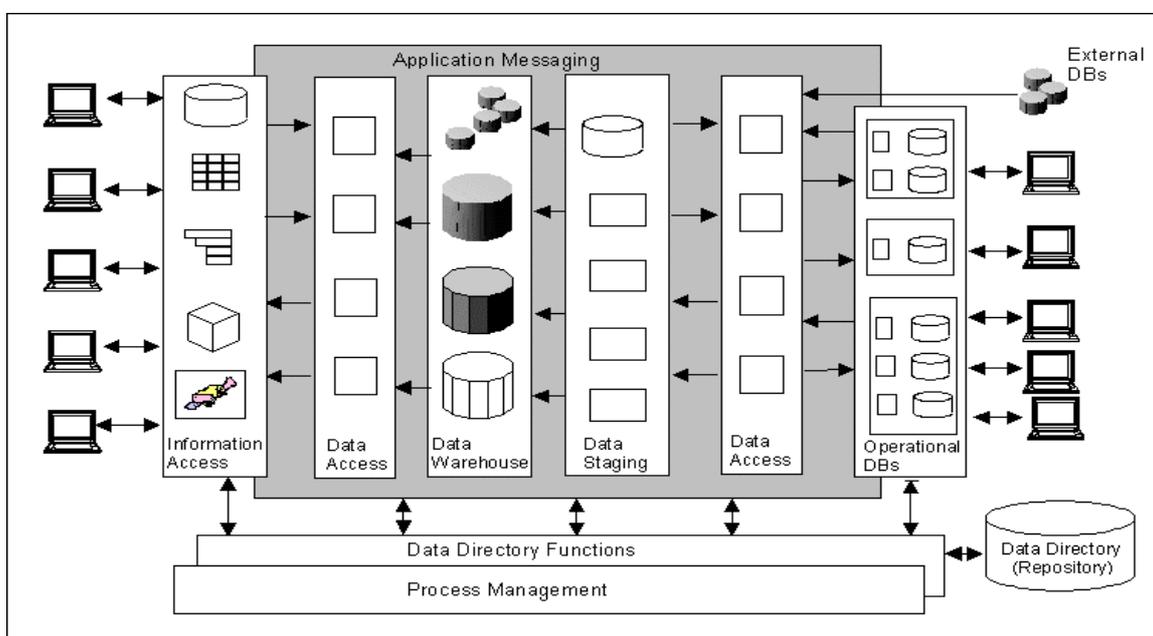


Figura 9 - Modelo de arquitetura proposto por Orr.

Fonte: (ORR, 1996).

3.5.2 Operational Data Store - ODS

É comum o ambiente operacional antigo não ser integrado. Com o passar do tempo as grandes corporações construíram grandes e desordenadas coleções de aplicações operacionais, carinhosamente chamadas de “sistemas legados” ou ambiente operacional “teia de aranha”. Essas aplicações foram sendo construídas uma a uma, separadamente das outras aplicações, com recursos tecnológicos diferentes. Por muitas razões, a tradução dos modelos de dados e processos dos sistemas legados em novos sistemas integrados, é uma transição muito difícil. Apesar de ser importante e

estratégico para o ambiente SAD, o DW não pode ajudar a corporação que esteja com grandes problemas com a falta de integração no ambiente operacional (INMON, 1997b).

Dentro desse contexto intransigente e de aplicações não integradas, aparece à noção de armazenamento de dados operacionais (*Operational Data Store* - ODS). O ODS se aplica ao processamento informacional dentro da arena de operações do dia-a-dia, e deve ser construído de modo separado do DW, conforme ilustra a Figura 10. O DW é usado para análises de longo prazo ou determinação de tendências, enquanto o ODS atende a um público de escritório ou tomadores de decisão corporativo no dia-a-dia, ou seja, o ODS serve para decisões “instantâneas” (INMON, 1997b).

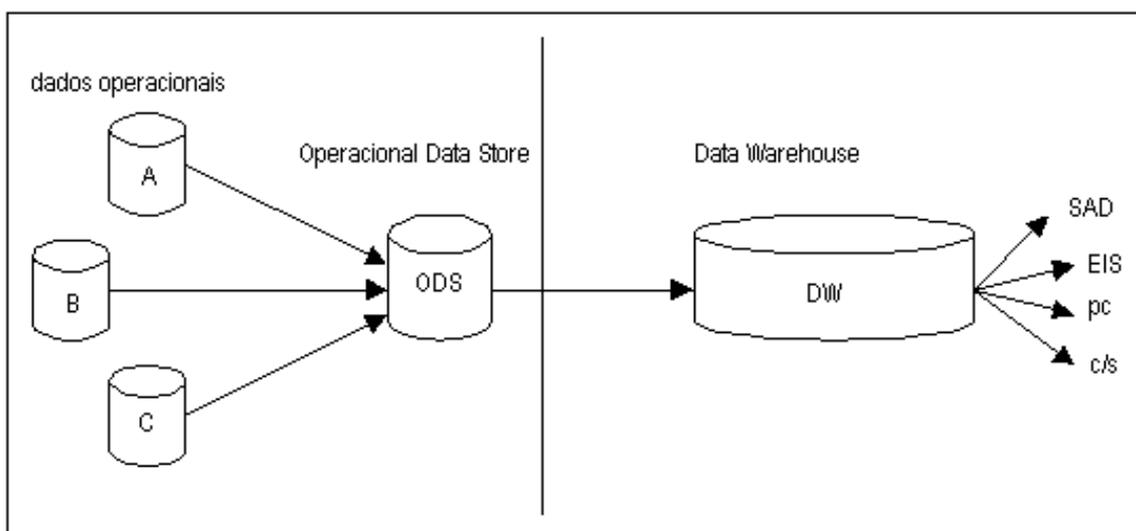


Figura 10 - A arquitetura onde o ODS é uma parte integral.
Fonte (INMON, 1997b).

Kimball (1998a), por sua vez, afirma que muitos administradores têm ambivalência ou confusão de sentido sobre ODS ser uma parte do ambiente de DW. Esses administradores muitas vezes se perguntam, “deveríamos construir um ODS?”, e/ou “nós temos certeza sobre o que é um ODS?”.

O ODS serve como ponto de integração dos sistemas operacionais, e em muitos casos o ODS é implementado para melhor suportar o relacionamento um-para-um (one-to-one), e sofre constantes acessos e alterações, deste modo, pode-se interpretar o ODS como um valioso componente do ambiente operacional. Todavia, o ODS é um depósito de dados atuais detalhados para suporte à decisão, principalmente quando se

necessita da análise no menor grão possível, e para a mineração de dados (*Data Mining*) identificando correlações e padrões, deste modo pode-se também entender o ODS como o nível atômico mais baixo do DW, ou seja, parte de um DW (KIMBALL, 1998a).

Kimball (1998a), afirma ainda, que a tecnologia de hardware e software para suportar DW está amadurecida, é possível armazenar mais dados, extrair e limpar os dados legados rapidamente, modelar para compreensão e elaboração de *queries* (indagações) mais sofisticadas e com alto desempenho. Ao contrário da suposição original do ODS, no DW pode-se armazenar as transações individuais de um grande negócio em uma série histórica temporal. A definição de ODS fica restringida somente ao suporte para integração dos dados transacionais dos sistemas operacionais.

3.5.3 Data Mart – DM

Originalmente um *Data Mart* era definido como um subconjunto altamente agregado de dados do DW, desempenhando um papel de um DW departamental ou direcionado a uma área específica do negócio. A organização pode construir uma série de *Data Marts* e ao longo do tempo vincula-los através de um DW lógico, com a vantagem de terem o retorno rápido, envolvendo o usuário final que consegue avaliar os benefícios extraídos do investimento ((SINGH, 2001), (MACHADO, 2000), (KIMBALL, 2000)). Segundo Kimball (2000), a *definição* original não funcionava porque os *Data Marts* não podiam ser combinados entre si, inflexíveis, tinham uma estrutura *stovepipe* (“chaminé de fogão”). O *Data Mart* agora é definido como um conjunto flexível de dados, freqüentemente com base no dado mais atômico (granular), sendo possível os dados serem extraídos de uma fonte original, e modelados de forma dimensional que é mais flexível para as consultas inesperadas (não estruturadas) dos usuários. Os *Data Marts* podem ser integrados utilizando técnicas de *drill across*³ através das (*conformed dimensions*) dimensões conformadas ou adaptadas⁴ (KIMBALL, 2000). Um DW é o conjunto de *Data Marts* integrados (KIMBALL, 1998a). A Figura 11 apresenta uma visão dessa abordagem.

³ Solicitar dados de uma ou mais tabelas de fatos de uma seqüência de valores em um único relatório (KIMBALL, 1998).

⁴ Dimensões são adaptadas quando são exatamente idênticas (incluindo as chaves) (KIMBALL, 2000).

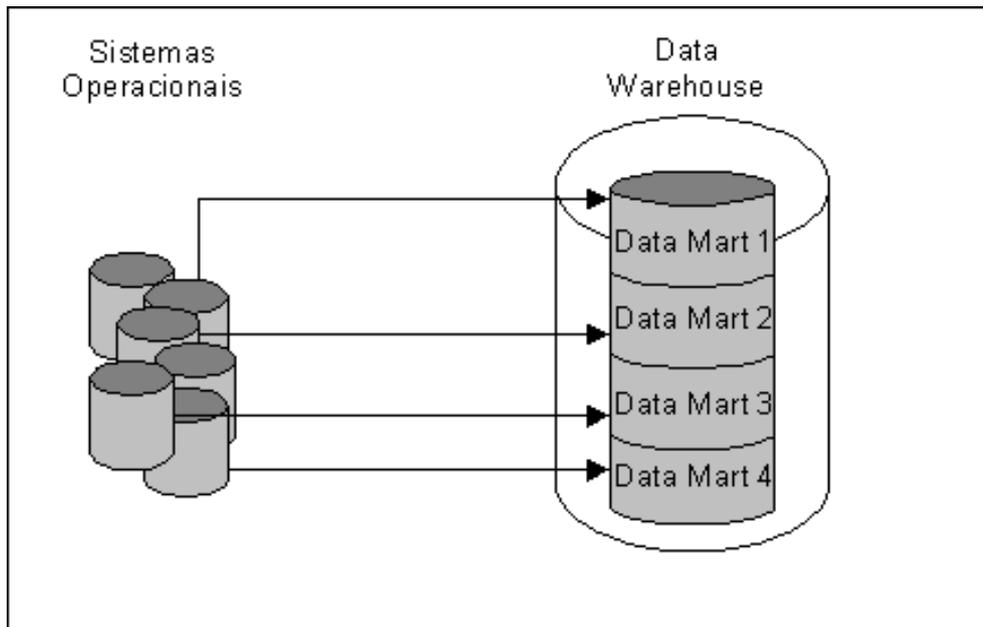


Figura 11 - Visão de múltiplos *Data Marts* integrados.
Fonte: (MACHADO, 2000).

Os *Data Marts* podem ser integrados ou interconectados através do compartilhamento de dados, provendo um aumento na capacidade e qualidade de visão corporativa dos dados e informações, porém essa integração é uma opção que irá aumentar consideravelmente a administração do DW (MACHADO, 2000).

3.5.4 Modelo de Arquitetura de DW por Ralph Kimball

Kimball (1998a), definiu um modelo de arquitetura de DW para a modelagem técnica do futuro DW. Essa modelagem deve representar os requisitos do negócio, bem como, possibilitar estimar custos e prazos de execução. O modelo de arquitetura é dividido em dois grandes componentes, o *Back Room* e o *Front Room*. O Quadro 8 descreve esses componentes ilustrados na Figura 12.

Quadro 8 - Descrição dos componentes do modelo de arquitetura definido por Kimball.

Componente	Descrição do Componente
Back Room (<i>Extração, Transformação e Carga – ETC</i>)	Neste componente concentra todas as ferramentas e processos necessários para aquisição dos dados dos sistemas fontes (sistemas operacionais ou transacionais, e a fontes externas), ou seja, a extração dos dados nos sistemas fontes, com as transformações adequadas, no tempo adequado e o armazenando na Área Intermediária de dados (<i>Data Staging Area</i>), para posterior carga no DW físico e a atualização do metadado (<i>Metadata Catalog</i>) que contém todos os significados (documentação) desses dados desde a origem, as suas transformações, até o armazenamento, sendo crucial para interação do usuário com os dados armazenados.
Front Room	Neste componente concentra todas as ferramentas e processos necessários para interação do usuário final com o DW, principalmente os aplicativos OLAP (On-line Analytical Processing) que irão possibilitar o acesso fácil, eficiente à informação, bem como, o processamento analítico da mesma. Também ferramentas de busca e apresentação dos dados (<i>warehouse browsing</i>), acesso e segurança, gerenciamento de consultas, relatórios padronizados e o monitoramento de atividades.

Fonte: Kimball (1998a)

Uma característica marcante no modelo de arquitetura definido por Kimball, é ser um modelo de *Data Marts* integrados, apesar dos *Data Marts* serem implementados separadamente, eles são interconectados, provendo uma visão corporativa dos dados. A Figura 13 ilustra a arquitetura definida por Kimball, onde podemos ver ao centro o DW físico com os *Data Marts* dimensionais dos dados atômicos e os *Data Marts* dimensionais dos dados agregados. Os *Data Marts* dimensionais são interligados através do *Data Warehouse Bus* onde estão as tabelas de dimensões e fatos conformados, que estabelecem a relação entre componentes de um *Data Mart* com os de outro *Data Mart*, isso ocorre tanto no nível atômico como no nível agregado ((MACHADO, 2000), (KIMBALL, 1998a)).

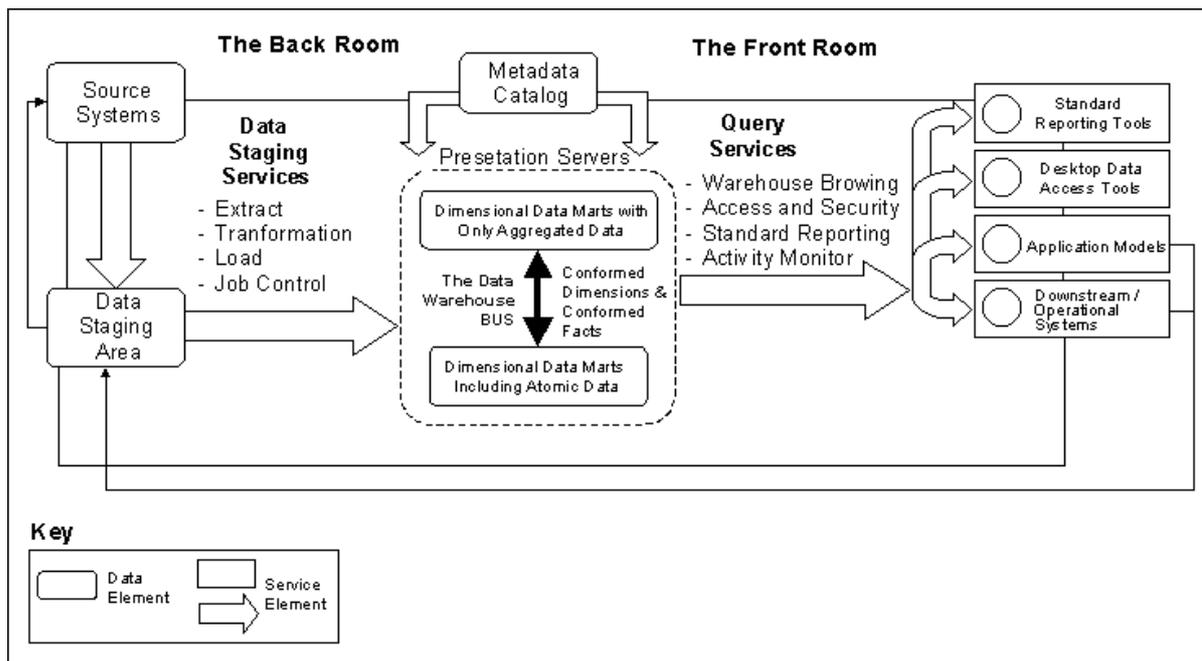


Figura 12 - Modelo de arquitetura proposto por Kimball.
Fonte: Adaptado de Kimball (1998a).

3.6 Modelagem dos dados

A compreensão dos dados é um dos maiores problemas no desenvolvimento do DW. A construção do modelo de dados é fundamental para o desenvolvimento, ajudando a compreender as regras de negócio e a organização dos dados para o melhor tempo de resposta (KIMBALL, 1998b).

Segundo Singh (2001), o modelo de dados do DW representa os requisitos de informações integradas, os requisitos de análise e o suporte à decisão de toda a organização.

A modelagem de dados para DW é completamente diferente da modelagem utilizada nos sistemas operacionais (OLTP) (MACHADO, 2000). Nos ambientes operacionais, o modelo utilizado é o Modelo Entidade/Relacionamento (MER) que utiliza uma técnica que busca remover qualquer redundância de dados, qualquer transação que atualize os dados será efetuada em apenas um ponto do banco de dados (KIMBALL, 1998b). No ambiente de DW, é utilizado o modelo dimensional que utiliza uma técnica que suporta o ambiente para análise multidimensional dos dados (MACHADO, 2000).

Os diagramas de entidade/relacionamento são difíceis de visualizar e memorizar, são simétricos, isto significa que a aparência do diagrama pode ser modificada arbitrariamente, mas sem mudar o significado lógico do diagrama. As tabelas parecem iguais, são representadas da mesma forma ficando difícil identificar qual delas é a maior ou a mais importante, e quais tabelas contêm valores numéricos do negócio (KIMBALL, 1998b).

Kimball (1998b, p. 142), afirma que “modelos de entidade/relacionamento são um desastre para consultas porque não podem ser entendidos pelo usuário nem navegados de forma útil pelo software de SGBD. Os modelos entidade/relacionamento não podem ser utilizados como base para *Data Warehouse*”.

3.6.1 Modelo Dimensional

O modelo dimensional é geralmente visualizado através do desenho de um cubo. Um cubo pode conter três ou mais dimensões que participam de um fato (assunto de negócios), conforme ilustra a Figura 13, cada ponto interno contém as medidas (o fato) relativas à combinação das dimensões (produto, mercado e tempo). A modelagem dimensional é mais simples, mais expressiva e mais fácil de entender que a modelagem (MER) ((MACHADO, 2000), (KIMBALL, 1998b), (FEW, 2001)).

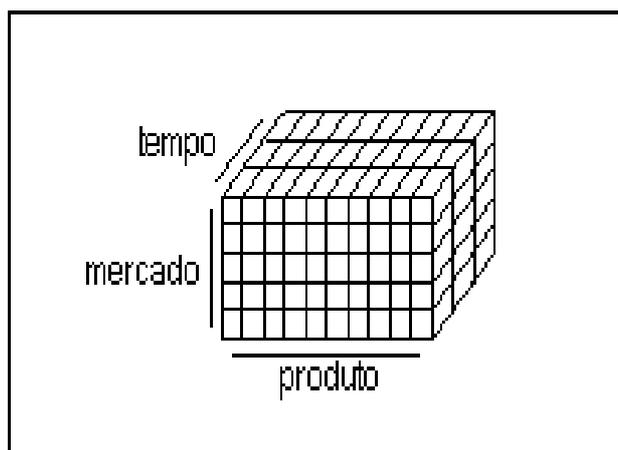


Figura 13 - Modelo dimensional de um negócio: visualização dos dados como um cubo.
Fonte: Adaptado de Kimball (1998b).

Os modelos dimensionais são construídos em torno de fatos numéricos de medida em um instante específico de tempo ou de medida acumulada sobre um período

de tempo. Esses fatos são as medições numéricas do negócio, que podem ser a venda de um produto no varejo, o preço de uma ação em um determinado momento de tempo ou a alteração do salário por promoção. Os fatos medidos são coletados dos sistemas de computador de produção e colocados em tabelas denominadas, apropriadamente, tabelas de fatos (KIMBALL, 2000).

Os elementos que participam de um fato (medida do negócio) são as dimensões (MACHADO, 2000). As dimensões são descrições textuais das dimensões do negócio no momento que o fato é gerado, em síntese, são os atributos de um fato ((KIMBALL, 1998),(KIMBALL, 2000)). Como exemplo, podemos considerar – Se o fato medido é a quantidade de cobertura de um automóvel incluída em uma nova apólice sendo criada, então talvez conheçamos as seguintes dimensões deste fato (KIMBALL, 2000):

- “A data dessa transação de seguro;
- O cliente;
- O agente de seguro;
- O item coberto (o carro);
- O tipo de cobertura (por exemplo, cobertura abrangente de responsabilidade);
- O tipo de transação (por exemplo, criar cobertura);
- O formato da apólice (por exemplo, um tipo específico de apólice contendo o conjunto conhecido de parágrafos jurídicos).”

As descrições textuais de cada dimensão são armazenadas em tabelas separadas de dimensão com uma chave primária. Na tabela de fato é colocada uma chave estrangeira que se associa a uma chave primária correspondente em uma tabela de dimensão, deste modo as dimensões ficam associadas ao fato. Na Figura 14 podemos verificar que esse modelo é parecido com uma estrela, onde a tabela de fatos está no centro e as tabelas de dimensão estão em torno dela. Esse modelo de dimensional é conhecido como modelo estrela (*star schema*) (KIMBALL, 2000).

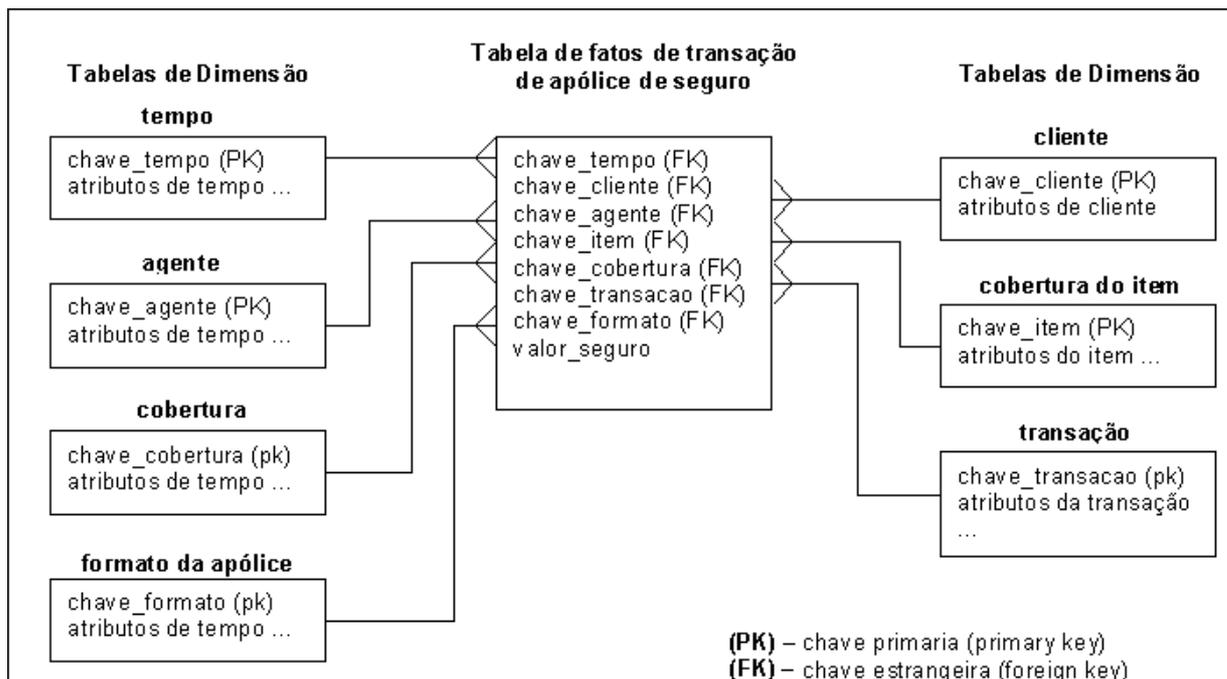


Figura 14 - Um modelo dimensional de transação de seguro com uma tabela de fatos e sete tabelas de dimensão.

Fonte: Adaptado de KIMBALL (2000).

3.6.2 Gerenciamento do volume de dados

Segundo Inmon (2001, p. 183) “uma das verdadeiras questões do design que o projetista enfrenta é gerenciar volumes maciços de dados”. Inmon também demonstra como os projetistas gerenciam grandes volumes de dados:

- “Por meio de técnicas de amostragem;
- Subdividindo os dados em tabelas separadas, onde cada tabela pode ser gerenciada individualmente;
- Operando abaixo do nível mais inferior de granulosidade. Isso geralmente é uma opção pobre e é feita quando não há outra maneira de realizar a tarefa;
- Quebrando o processo analítico em uma série de pequenos passos seqüenciais e executando cada um dos passos por vez e de forma independente.”

3.6.3 Granularidade e Particionamento

Nesta seção são descritos dois aspectos importantes no projeto de um DW. Segundo Inmon, nenhum outro aspecto de projeto terá importância se a granularidade

não for conduzida de forma apropriada e se o particionamento não for cuidadosamente projetado e implementado (INMON, 1997a).

3.6.3.1 Granularidade

Para Inmon, a granularidade é a principal questão de projeto, pois afeta profundamente o volume dados no DW e o tipo de consulta a ser oferecida. O nível de detalhe de uma consulta vai determinar o balanceamento do nível de granularidade contido nas unidades de dados no DW, e conseqüentemente o volume de dados. Quanto mais detalhe, mais baixo o nível de granularidade. Quanto menos detalhe, mais alto o nível de granularidade

A maioria das organizações encontra grandes dificuldades no balanceamento da granularidade, sendo que a melhor solução consiste em alguma estrutura de vários níveis de granularidade. O balanceamento no gerenciamento da questão da granularidade deve ser feito cuidadosamente no início do projeto e construção do DW, com mostra a Figura 15 (INMON, 1997a).

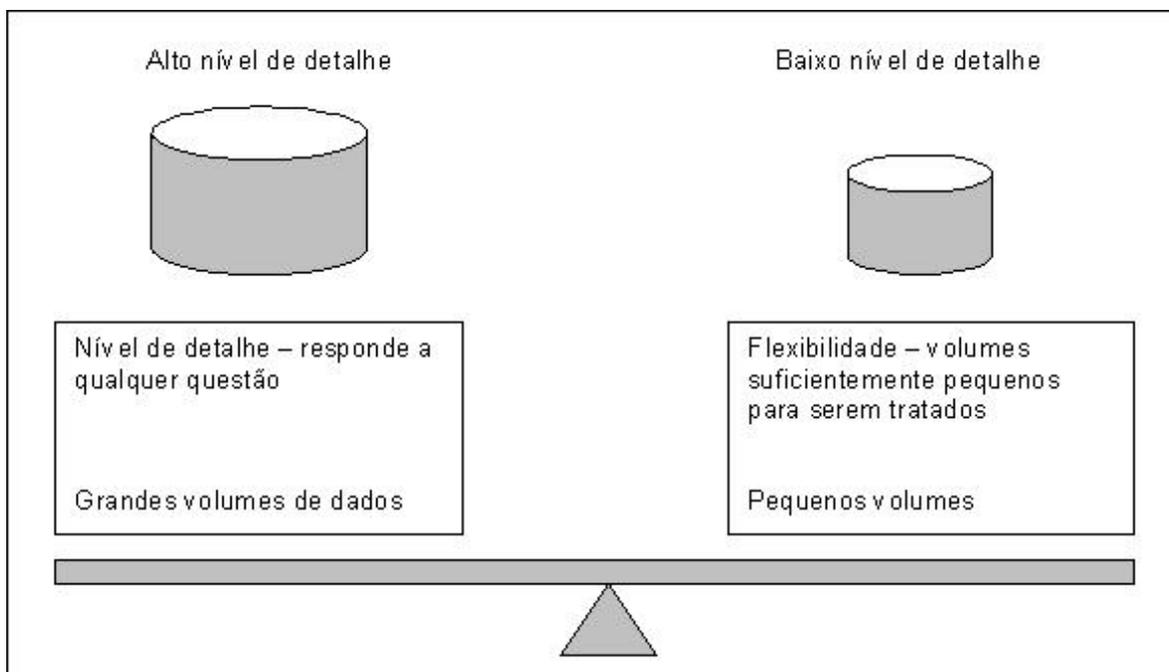


Figura 15 - O balanceamento no gerenciamento da questão da granularidade.
Fonte (INMON, 1997a).

Segundo Kimball, na maioria das vezes é preciso expressar os dados no menor nível de granularidade, para que as consultas possam aprofundar-se com precisão no

DW. A integração de duas fontes normalmente significa expressá-las no menor nível de granularidade comum a ambas as fontes. Por exemplo uma fonte baseada em meses e outra em semanas, a alternativa mais eficiente é integrar as fontes no nível diário (KIMBALL, 1998b).

Singh, afirma que sob muitos aspectos utilizar o maior nível de detalhamento representa segurança, mesmo com a disponibilidade de outros níveis sumarizados. Quanto maior for o nível de sumarização, maior a eficiência e a velocidade de obtenção dos dados, ou seja, os analistas de SAD manipulam mais os dados sumarizados (SINGH, 2001).

A maioria das organizações precisa dispor de dois níveis de granularidade no DW. Precisa do nível baixo de granularidade em função da segurança e da precisão obtida. Precisa também do nível mais alto de granularidade em função da velocidade e eficiência na obtenção dos dados ((INMON, 1997a), (KIMBALL, 1998b), (SINGH, 2001)).

3.6.3.2 Particionamento

Para Inmom, depois da granularidade outra questão importante do projeto dos dados é o particionamento de dados. O particionamento de dados é a separação do dados de detalhe corrente em unidades físicas separadas que podem ser tratadas de forma independente, por serem menores permitem maior flexibilidade no gerenciamento dos dados (INMON, 1997a). A Figura 16 mostra a importância do particionamento dos dados no projeto de dados do DW.

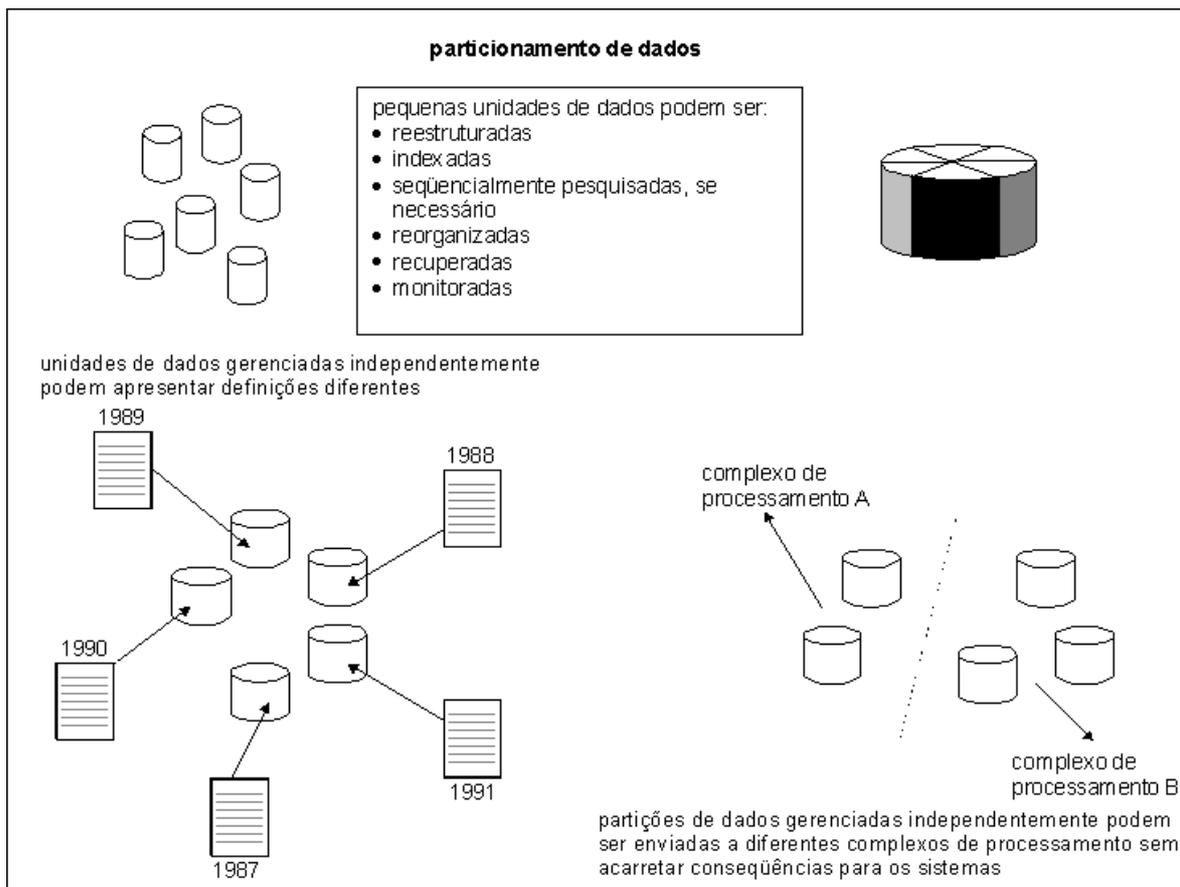


Figura 16 - A importância do particionamento dos dados no projeto de dados do DW.

Fonte: (INMON, 1997a).

Segundo Singh, o particionamento pode ser feito no nível do SGBD e/ou no nível da aplicação. O particionamento no nível SGBD, devido o conhecimento das partições, o gerenciamento e a infra-estrutura ficam a cargo do SGBD. No nível de aplicação o particionamento é feito pelo código da aplicação controlado pelo programador, o que possibilita maior flexibilidade de gerenciamento de dados (SINGH, 2001).

Inmon, chama o nível do SGBD de nível de sistema. Para ele esse nível pressupõe uma definição de dados única, o que dificulta para o DW que trabalha com horizontes temporais de até 10 anos, a definição dos dados, normalmente, é alterada com o passar do tempo. No nível de aplicação, o SGBD e o sistema operacional (SPT/OLTP) não tomam conhecimento do particionamento, sendo possível existir uma definição de dados diferente para cada ano (INMON, 1997a).

3.7 Estratégias arquitetônicas para a implementação de *Data Warehouse*

No planejamento da estratégia arquitetônica para a implementação de *DW*, encontram-se três estratégias básicas: “top-down”, “bottom-up” e intermediária. A escolha da estratégia é fator importante na seleção da tecnologia apropriada para o desenvolvimento deste ambiente. (MELLO, 1997), (HACKNEY, 2001).

3.7.1 Estratégia “Top Down”

Nessa estratégia o processo se inicia com a criação de um *Data Warehouse* e depois parte-se para a segmentação, ou seja, divide-se o *DW* em áreas menores gerando assim pequenos bancos orientados por assunto departamentalizados, ou seja, os *Data Marts*.

O *DW* integra todos os dados num formato comum e num ambiente comum de software. Todos os recursos de uma organização e todos os dados necessários para o suporte à decisão são consolidados na construção do *DW*.

Vantagens: Todos os *Data Marts* originados a partir do *DW*, utilizam a mesma arquitetura e dos dados deste *DW*, permitindo uma fácil manutenção; o *DW* concentraria todos os negócios da empresa, sendo possível a partir dele extrair níveis menores de informações; esse tipo de arquitetura garantiria a existência de um único conjunto de aplicações para extração, limpeza e integração dos dados, além de processo centralizados de manutenção e monitoração.

Desvantagens: Longo tempo de implementação, alta taxa de risco; a demora do projeto e a falta de retorno podem frustrar as expectativas dos usuários, fazendo com que o projeto venha a ser abandonado.

3.7.2 Estratégia “Bottom-up”

Nessa estratégia são desenvolvidos *Data Marts* independentes que podem se transformar em *Data Warehouses* ou não.

Na estratégia “*Bottom-up*” a situação é inversa à estratégia “*Top-Down*”. Muitas organizações preferem primeiro criar um banco de dados para somente uma área ou departamento. Com isso os custos são bem inferiores de um projeto de *DW* completo. A partir da visualização dos primeiros resultados partem para outra área assim sucessivamente até resultar num *DW*. A arquitetura proposta por Kimball (1998a), de *Data Marts* dimensionais que são interligados através do *Data Warehouse Bus* onde estão as tabelas de dimensões e fatos conformados, permite a conexão dos *Data Marts* formando um *DW* corporativo.

Vantagens: Permite um rápido desenvolvimento, conseqüentemente apresentando resultados rápidos e um elevado nível de confiança; a elaboração de *Data Marts* incrementais, permite que os principais negócios sejam enfocados inicialmente, sem que haja gastos no desenvolvimento de áreas não essenciais da organização.

Desvantagens: Requer um maior controle do negócio como um todo; os desenvolvedores têm um maior trabalho para extrair e combinar as fontes individuais; tem-se que coordenar esforços e recursos de múltiplas equipes; pode gerar *Data Marts* desvinculados do todo.

3.7.3 Estratégia “Intermediária”

Segundo Hackney (1998), esta estratégia tem a finalidade de integrar a estratégia “*top-down*” com a “*bottom-up*”. Nesta abordagem efetua-se a modelagem de dados do *Data Warehouse*, sendo o passo seguinte a implementação de partes desse modelo. Essas partes são escolhidas por área de interesse e constituem os *Data Marts*. Cada *Data Mart* gerado a partir do modelo de dados do *Data Warehouse* é integrado no modelo físico do *DW*. A principal vantagem desta estratégia é a garantia de consistência dos dados. Essa garantia é obtida em virtude do modelo de dados para o *DW* ser único possibilitando realizar o mapeamento e o controle dos dados.

3.8 Metodologias de Desenvolvimento

Para Chaudhuri (1997), para se construir um *Data Warehouse* é preciso um processo complexo e longo, o qual exige extensa modelagem do negócio da empresa e isso pode exigir muito tempo para que o mesmo seja bem sucedido. Por outro lado, para Mello (1997), esse processo abrange a integração e transformação das fontes de dados, ou seja, a modelagem e construção do DW, a conseqüente construção dos *Data Marts* e o subseqüente desenvolvimento das aplicações OLAP visando atender aos usuários finais.

Serão apresentadas a seguir as metodologias propostas por Kimball e Inmon. Ambas enfatizam a importância da modelagem de dados, apresentando o desenvolvimento do ambiente com uma arquitetura “top-down”.

3.8.1 Metodologia de desenvolvimento baseada em dados – W.H.Inmon

O *Data Warehouse* é desenvolvido com objetivo de atender às necessidades da organização quanto à suporte à decisão.

Para Inmon, o importante é assegurar que o usuário final entende que o *Data Warehouse* não é um banco de dados corporativo. Inmon utiliza o modelo corporativo da empresa como o início da modelagem e isso mostra uma visão genérica dos dados do negócio, a facilidade de crescimento e o desenvolvimento em paralelo de diferentes necessidades do negócio. A metodologia de Inmon é apresentada em 3 fases: METHOD1, METHOD2 e METHOD3.

METHOD1 – Fase direcionada para análise dos sistemas e processamento operacional. É a fase operativa, subdividida nas seguintes fases conforme a Figura 17.

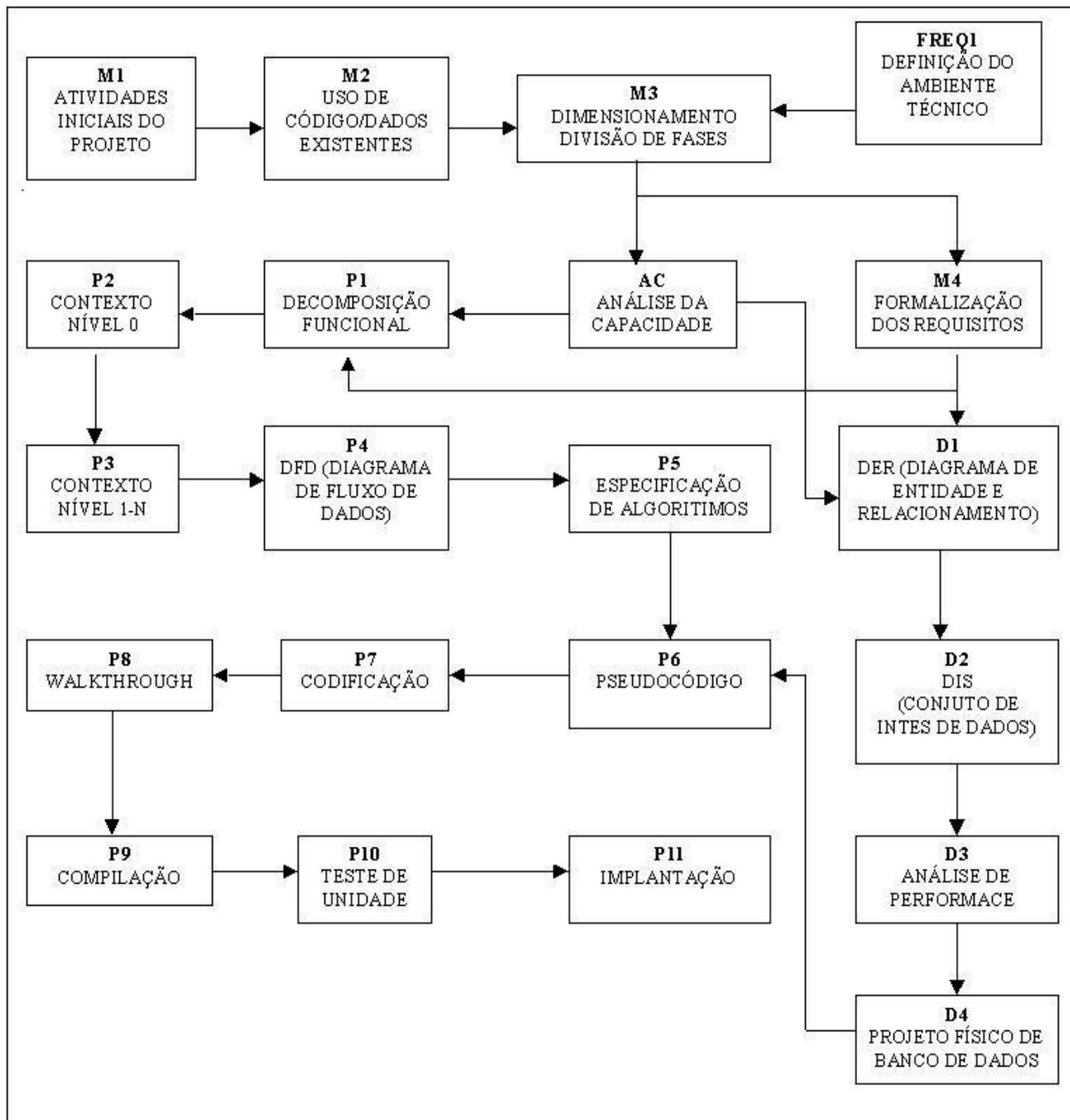
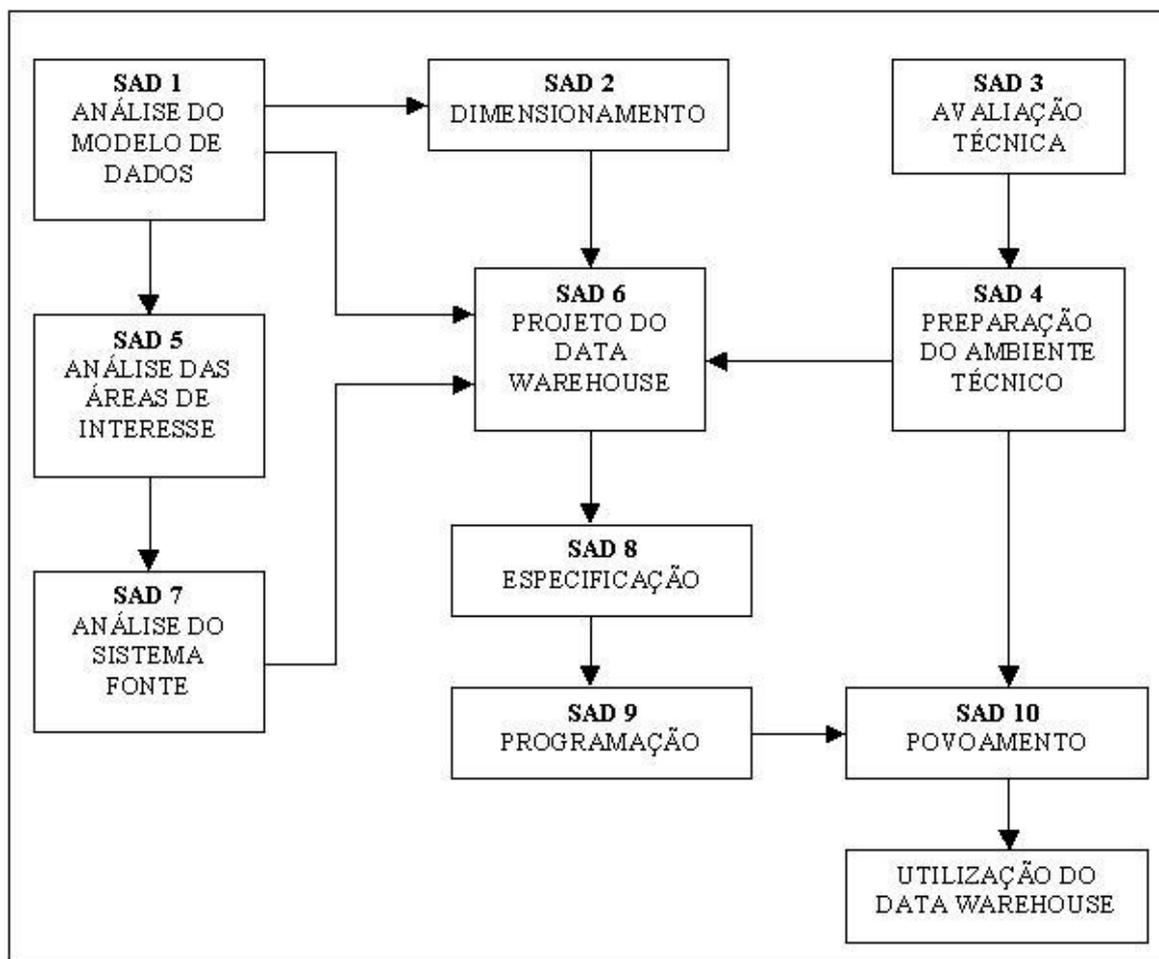


Figura 17 – Etapas da Metodologia 1 de Inmon.
Fonte: SOARES (1998).

METOD2 – Esta fase é voltada para processamento SAD orientado ao *Data Warehouse*. Esta fase se concentra em um modelo de dados que permita organizar a redundância dos dados. E a parte responsável pelo desenvolvimento do DW. Representa a fase de construção. A Figura 18 apresenta as etapas da metodologia 2.



**Figura 18 – Etapas da Metodologia 2 de Inmon.
Fonte: SOARES (1998).**

METOD3 – Esta fase da metodologia descreve a utilização do *Data Warehouse*. Representa a fase de utilização interativa do DW. Esta fase representa a utilização dos dados do DW para fins de análise. Ela apresenta um desenvolvimento diferente das fases anteriores, pois os processo de desenvolvimento começa pelos dados do DW, os requisitos não são conhecidos no início do processo e o processamento é feito de modo interativo e heurístico. A Figura 19 apresenta as fases dessa metodologia:

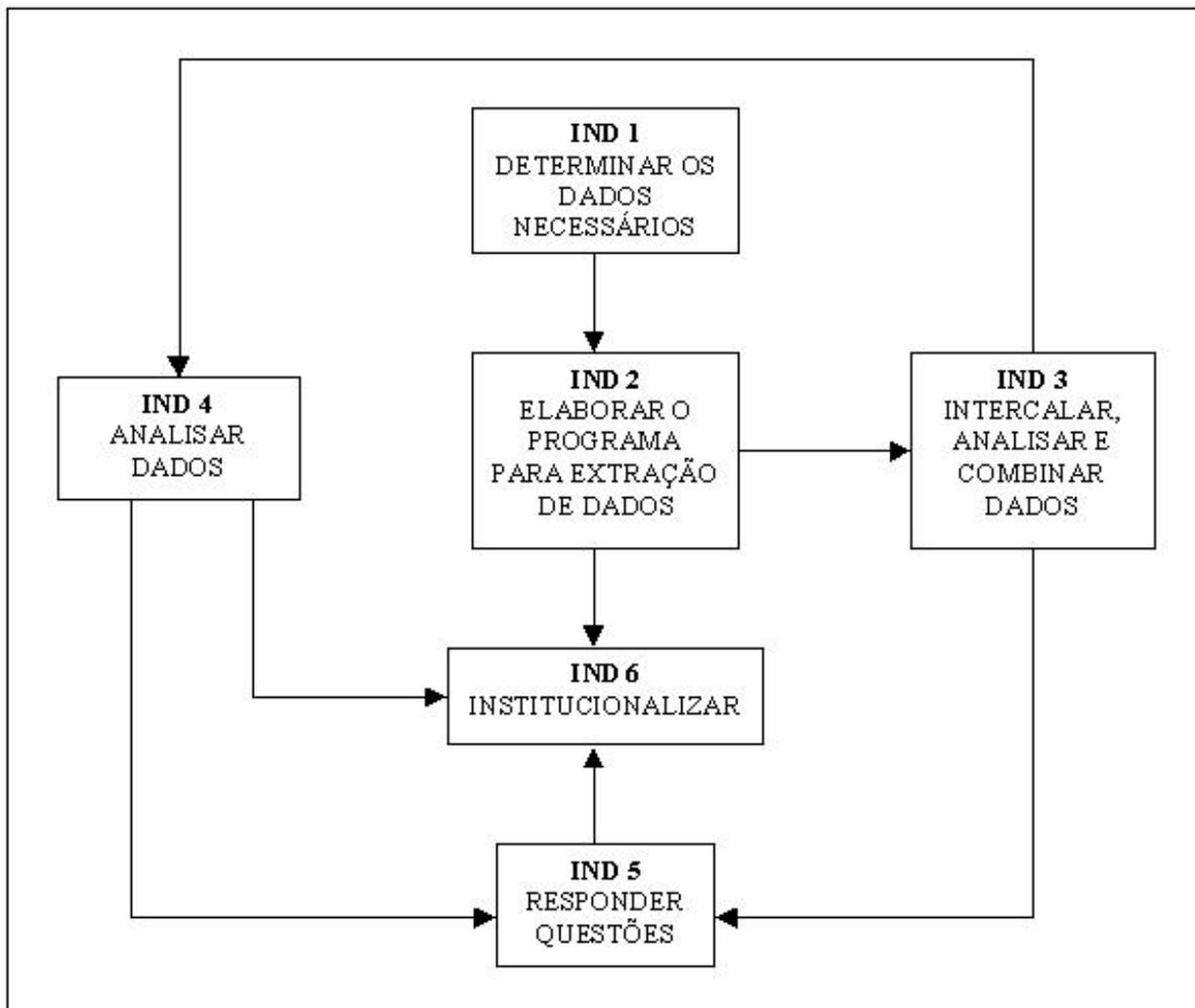


Figura 19 – Etapas da Metodologia 3 de Inmon.
Fonte: SOARES (1998).

3.8.2 Metodologia dos nove pontos de decisão – Ralph Kimball

A visão Kimballiana trabalha com o modelo dimensional, mostrado pelo esquema estrela, e é usado para o desenvolvimento do *Data Warehouse*. Uma empresa de grande porte apresenta vários esquemas. Esses esquemas devem ter tabelas de dimensões que possam ser compartilhadas entre esses esquemas (KIMBALL,1996).

O DW é conceituado pela utilização de estruturas multidimensionais, seja ela pela implementação de um esquema estrela em um Sistema Gerencial de Banco de Dados Relacional (SGBDR) ou através de um Sistema Gerencial de Banco de Dados Multidimensional (SGBDM). Esta modelagem deve seguir os seguintes propósitos:

- Deixar que o usuário final avalie uma determinada situação sobre diversas formas, de acordo com a sua necessidade;
- Permitir análises complexas e uma melhor visualização da informação;
- Disponibilizar a informação em uma estrutura que facilite o trabalho das ferramentas que a manipularão.

Segundo Kimball (1997), o desenvolvimento do ambiente deve seguir algumas etapas mostradas a seguir:

- Deve ser selecionado um processo de negócio para modelar. Esse processo deve ser uma operação importante na empresa e deve ser suportado por algum tipo de sistema de onde seja possível coletar dados;
- Deve ser selecionado a granularidade do negócio. O item mais detalhado (grão) é o nível atômico que representará esse processo na tabela de fatos. Transações individuais, instantâneos individuais diários e instantâneos individuais mensais são considerados grãos típicos;
- Deve ser selecionado as dimensões que serão aplicadas a cada registro da tabela de fato. Para cada dimensão escolhida, deve se descrever todos os atributos de dimensão que preenchem cada tabela dimensional;
- Deve ser selecionado os atos importantes que irão popular cada registro da tabela de fatos.

De acordo com Kimball (1996), para tomar o processo detalhado, são estabelecidos nove pontos de decisão que consistem em:

1. Deve-se identificar a tabela de fatos. Neste ponto são identificados os principais processos da empresa onde os dados serão coletados;
2. É estabelecido o detalhamento (granularidade) de cada tabela de fatos. Neste ponto é determinado o nível de detalhe de interesse para a análise;

3. Deve-se descobrir as dimensões de cada tabela de fatos. A partir da granularidade desejada são obtidas as dimensões primárias. Outras dimensões adicionais poderão surgir ao longo do desenvolvimento. Estas dimensões não são necessárias para a definição da granularidade da tabela de fatos;
4. Obter os fatos, incluindo fatos pré-calculados. Neste ponto serão estabelecidos os fatos mensuráveis;
5. Obter os atributos da dimensão com descrições completas e terminologia apropriada;
6. Rastrear dimensões de modificação lenta. Representa o rastreamento de dimensões que sofram mudanças graduais ao longo dos eixos temporais;
7. Analisar agregados, dimensões heterogêneas, minidimensões, modos de consultas e outras decisões de armazenamento físico;
8. Selecionar a amplitude de tempo do histórico do banco de dados; e
9. Estabelecer os intervalos com que os dados serão extraídos e carregados no *DW*.

3.9 Fatores Críticos de Sucesso no Desenvolvimento do Ambiente de *Data Warehouse*

Nesta seção são apresentados alguns fatores críticos que são determinantes para o desenvolvimento de um Ambiente de *DW*, com sucesso ((FEW, 2001), (BARBIERI, 2001, p.59-62), (HACKNEY, 2001) (PERKINS, 2000), (GRIFFIN, 2000)).

Definição do escopo do projeto: É fundamental a definição da abrangência e dos produtos a serem entregues. Esses produtos são definidos pelos usuários que, muitas vezes, não conseguem definir, claramente, os problemas de negócios e como resolvê-los. Devem ser observados cuidados especiais para evitar, a todo custo, o gigantismo de propostas ou projetos particularizados demais. O projeto deve ser desenvolvido por

partes, com recursos e produtos bem definidos e prazos razoáveis sendo, plenamente, acordado com os usuários.

Administração de expectativas: Como parte da definição do escopo do projeto, deveriam ser definidas expectativas claras, realistas e condizentes com o escopo do projeto, recursos e prazos possíveis. Os usuários devem ser bem informados do escopo do projeto e, sempre atualizados. Deve-se evitar mudar o escopo do projeto livremente, quando isso acontecer, às expectativas devem ser repensadas e acordadas com os usuários. Surpresas acontecem em um projeto de DW mas, com a administração das expectativas elas vão ser evitadas e no final, os usuários ficarão satisfeitos porque será fornecido o que foi acordado.

Patrocinador forte: No projeto deve haver a presença forte do usuário, com poder para direcionar o foco do projeto, definindo pendências, dirimindo dúvidas na esfera negocial, bem como, garantindo os recursos necessários. Esse usuário deve ser participante do projeto, como o grande patrocinador, com todos os ônus e bônus dessa condição.

Envolvimento dos usuários: É importante a conscientização dos usuários sobre as condicionantes, para o sucesso do projeto, bem como, as possibilidades vislumbradas, aumentando o compromisso para diminuir as resistências ao projeto. Deve-se manter os usuários aderidos. Todos os potenciais usuários do DW, inclusive os executivos devem estar, efetivamente, envolvidos no projeto, desenvolvimento e gerenciamento. Eles são os donos dos dados operacionais e conhecedores do negócio.

Pessoal qualificado no projeto: O projeto de DW é desenvolvido em várias etapas. Para tanto, é necessária uma equipe tecnicamente qualificada para a consecução das etapas, de forma coesa e integrada com o projeto. As restrições de pessoal qualificado aumentam o tempo de desenvolvimento, o custo das correções e frustram as expectativas, tornando o projeto mais caro. É importante fixar uma política de treinamento que, quando necessária, permita a aquisição de ajuda externa de consultores. Um consultor, às vezes, pode ensinar por um dia, o que se levaria um mês ou mais, para aprender por conta própria. O contato com os consultores, desde o início

do desenvolvimento, favorece a qualificação do pessoal interno e torna o aprendizado mais auto-suficiente.

Estratégia de implementação arquitetônica: A estratégia de implementação arquitetônica vai determinar o plano arquitetônico da estrutura do ambiente do DW, bem como, toda a infra-estrutura tecnológica a ser empregada no projeto. As estratégias básicas são a “bottom-up”, “top-down” e a intermediária. A estratégia deve ser adotada, levando em consideração os recursos TI existentes e os possíveis, bem como, o nível de conscientização das tecnologias de DW na organização. A estratégia adotada vai ser um dos fatores mais críticos de sucesso do ambiente de DW na organização.

Extração, transformação e carga dos dados (ETC): Normalmente de 60% a 80% de um projeto de DW, é composto dos processos de ETC. Os dados operacionais devem ser limpos, editados e transformados antes de serem carregados no DW. É fundamental conhecer e definir os dados necessários, com o mapeamento criterioso das fontes dos dados, em função da qualidade, natureza, periodicidade de atualização, dos sistemas mantenedores e perspectivas de duração das fontes. Nem todos os dados operacionais disponíveis são, necessariamente, úteis para o DW. As ferramentas próprias para os processos de ETC ainda são primitivas ou difíceis de usar, geralmente, os processos de ETC são desenvolvidos e mantidos pela própria equipe do projeto. Quando os dados fontes mudam na sua estrutura ou periodicidade de atualização, os processos de ETC devem ser adequados a essas mudanças.

Modelagem dos Dados: É determinante para se evitar futuros problemas de desempenhos do DW. O modelo dimensional é o mais adequado para um DW, é mais fácil para os usuários finais entender e navegar, enquanto que o modelo de entidade-relacionamento (ER) é ideal para os OLTP e ERP. É importante o trabalho de engenharia direcionado para construir um modelo de dados, com dados úteis, agregações e granularidades corretas.

Metadados: São essenciais em todos os níveis do ambiente do DW, sendo o centro nervoso. Além de conterem os dados sobre os dados do ambiente do DW, como o nome sugere, são usados para gerenciar e controlar o mesmo, bem como, ser o dicionário dos dados, com toda documentação sobre o dado, desde as origens até o

armazenamento, isso inclui a documentação sobre os processos de ETC, modelo de dados, estrutura e significado do dado. Para os usuários, os metadados são o catálogo sobre os dados contidos no ambiente do DW, facilitando a localização e o uso desses dados em as suas análises. É difícil encontrar um repositório de metadados que envolva todo ambiente do DW, o que seria ideal, pelo menos, as ferramentas OLAP possuem meios mínimos de documentação dos dados que possibilitem a localização, o acesso e as análises, por parte dos usuários.

Arquitetura e Infra-estrutura Tecnológica: É importante analisar, cuidadosamente, a plataforma tecnológica, onde residirão os SGBDs relacionais ou dimensionais, os softwares operacionais básicos, capacidade de processamento e armazenamento, escalabilidade das máquinas, middlewares, softwares de ETC, softwares de metadados, softwares de acesso aos dados do ambiente de DW (ferramentas dos usuários finais), etc. Deve-se levar em consideração a maturidade geral dos produtos, que vão fazer parte do ambiente do DW.

Ferramentas dos usuários finais: As ferramentas dos usuários são a vitrine do ambiente do DW. Para a maioria dos usuários finais, a maior parte da complexidade do ambiente do DW é transparente, para eles essas ferramentas definem todo o ambiente do DW. É importante a escolha de ferramentas maduras no mercado, que possam ser integradas ao ambiente do DW, com a interação fácil para localização e acesso aos dados do ambiente do DW, bem como, a fácil execução de consultas, relatórios e as análises on-line.

Acompanhamento: A utilização pelos usuários deve ser acompanhada e analisada. É muito importante o *feedback* dos usuários envolvidos e do patrocinador, para identificar e corrigir as possíveis falhas do projeto. À medida que, os produtos resultantes do ambiente do DW forem aparecendo, é importante também fazer marketing, através de palestras de demonstração para as gerencias e anúncios na Intranet, para o público alvo. Os usuários e patrocinadores devem saber que, até aquele momento, a relação custo/benefício do projeto foi favorável.

3.10 Aplicações Comerciais de Tecnologias de *Data Warehouse*

Para Kimball (2000), na última década, as tecnologias de DW tiveram o papel principal de publicar os ativos das empresas que estão na forma de dados para os seus analistas internos e gerência. Com a explosão do uso da *Web* para impulsionar os negócios das empresas, cresceu a audiência para dados de DW sobre o relacionamento com os clientes, parceiros e concorrentes. “A revolução da *Web* impulsionou o *Data Warehouse* para o centro do palco porque, em muitas situações, o *Data Warehouse* deve ser o motor que controla ou analisa a experiência na *Web*”.

Segundo Barbieri (2001), estamos entrando na era do *Business Intelligence* – BI, cujo objetivo é desenvolver percepções, entendimentos e conhecimento, melhorando o processo de tomada de decisão, através de habilidades de se estruturar, acessar e explorar informações, normalmente guardadas no ambiente de DW. Os pensadores do amanhã já imaginam a próxima era de transformação: a gestão do conhecimento (*Knowledge Management System* – KMS). Enquanto BI transforma dados em informação, produzindo cubos e relatórios, etc., a KMS realiza as combinações, compilações, pontos de discussão, transformando as informações em conhecimento. Novos conceitos surgem e estão envolvidos com as tecnologias de DW, destacando o *Customer Relationships Management* – CRM, que são tecnologias e soluções que visam a aproximação do cliente, através da busca do conhecimento analítico de seus anseios, necessidades e tendências, bem como, uma estrutura de serviços, capaz de satisfazê-lo antes que o corrente o faça.

Com a explosão de novos conceitos como ERP, CRM e BI, as tecnologias de DW estão conquistando espaço cada vez maior, nas empresas nacionais. Segundo Camargo (2001), “a prioridade de investimentos de TI para os próximos doze meses serão soluções de CRM, DW e BI”. O Quadro 9 apresenta alguns casos recentes de aplicações comerciais de tecnologias de DW.

Quadro 9 – Casos recentes de aplicações comerciais de tecnologias de DW.

Instituição	Atuação	Aplicações	Softwares
Lojas Marisa	Rede de varejo	Carência de informações gerenciais. Este sistema envolve pedidos, notas, estoques e outras tarefas.	sistema operacional: Windows NT; SGBD: Microsoft SQL Server 7.0; Ferramenta de BI: MicroStrategy 7.0.
Companhia Energética de Goiás- CELG	Distribuição de energia elétrica	Gestão de pessoal.	sistema operacional: Windows NT; SGBD: Microsoft SQL Server 7.0; Ferramenta de BI: MS SQL Server Olap.
Robert Bosch América do Sul	Indústria - setores: automotivo; bens consumo; bens produção.	Área comercial, expandindo para as áreas de logística, qualidade, compras e finanças	sistema operacional: Windows NT; ERP: R/3 SAP versão 4.6C; SGBD Oracle 8i; ferramenta de BI: Business Information Warehouse da SAP.
Grupo VR	Serviços corporativos:	A idéia é obter informações de todos os canais: Internet, portal, ERP, CRM (call center).	sistema operacional: HP-UX; SGBD: Sybase 11.92 ERP: Peoplesoft versão 7.70 GA; CRM: Vantive 8.2.1; ferramenta de BI: Business Objects ou da Cognos.
MD Papéis	Indústria de papel	Área comercial, sendo que a fábrica já iniciou o processo na parte industrial. Em seguida parte para recursos humanos e, por último, econômico e financeiro.	sistema operacional: Windows NT; SGBD: Microsoft SQL Server 7.0; Ferramenta de BI: - MS SQL Server Olap; - Data mining sob a plataforma DSS (SAD) da Accrue Pilot Software.
Roche do Brasil	Indústria farmacêutica	Receber todas as informações críticas e gerenciais para facilitar a toma de decisão. O objetivo é integrar ao data warehouse às soluções de ERP, CRM, call center e do Web site.	sistema operacional: HP9000-UNIX e Windows NT; ERP: R/3 SAP (UNIX/Oracle); SGBD Oracle 8i e Microsoft SQL Server 7.0; ferramenta de BI: Business Information Warehouse da SAP.
Visanet	Processamento de cartão de crédito	Inicialmente aplicações nas Áreas financeira e comercial. A consolidação do sistema em todas as áreas deve acontecer até o final do ano 2001.	sistema operacional: HP Openview (HP); e Windows NT; ERP: R/3 SAP (UNIX/Oracle); SGBD Oracle 8i e Microsoft SQL Server 7.0; ferramenta de BI: MicroStrategy 7.
Nestlé Brasil Ltda.	Indústria de alimentos	desenvolveu-se um aplicativo (Java) que apresenta aos usuários do sistema os principais indicadores de desempenho da Nestlé. Entre os índices medidos estão custos, serviços ao cliente, performance de vendas, produção, serviços de fornecedores e posição de estoque.	ERP: R/3 SAP; ferramentas de BI: Genexus Corporate, BI Explorer e Java
Santista Têxtil S/A	Indústria têxtil	Atualmente, a empresa está focada na implantação do ERP e em projetos de otimização da intranet e do data warehouse.	sistema operacional: Windows NT; SGBD: Microsoft SQL Server 7.0; Ferramenta de BI: MicroStrategy 7.0.
Sabesp	Saneamento básico	Nas áreas de faturamento, arrecadação e ordens de reforma da Sabesp. Além disso, a Sabesp pretende suportar e preservar o ambiente de dados para as mudanças previstas em todo seu parque legado de TI. "Estamos construindo o alicerce de nossa nova casa".	sistema operacional: Unix; SGBD: Teradata 2.3; Ferramentas de BI: Business Objects 5 e WebIntelligence 2.5, ambos da Business Object.

Fonte: Computerworld (2001)

3.11 Considerações Finais

Muitas empresas, principalmente grandes empresas, já chegaram à conclusão que a tomada de decisões baseada em fatos e de forma tempestiva é o novo diferencial

deste século. Porém como tomar decisões se as informações encontram-se espalhadas por vários lugares da organização? O *Data Warehouse* vem respondendo a esta pergunta de forma satisfatória.

O *Data Warehouse* pode ser implementado de várias formas, exige uma mudança de filosofia de tratamento de dados, na maioria das vezes é um processo demorado, dependendo da estratégia de abordagem adotada (top-down, bottom-up ou intermediária), podendo requerer inclusive mudança de tecnologias/parque tecnológico. Porém suas vantagens no processo decisório das empresas, como ferramenta auxiliar no processo de “garimpagem” e seleção de dados tem se mostrado eficaz.

O processo *Data Warehousing* exige pessoal capacitado. Deve-se tentar evitar que o seu tempo de implementação se estenda muito, evitando assim desperdício de dinheiro e garantindo continuidade do processo.

O processo de *Data Warehousing* pode auxiliar os tomadores de decisões das empresas públicas, principalmente devido ao seu tamanho e dificuldades orçamentárias, na obtenção de informações de forma tempestiva e conseqüentemente no menor desperdício de dinheiro público e na melhoria dos serviços oferecidos à população.

O objetivo deste trabalho é analisar a implementação e utilização de tecnologias referenciadas neste capítulo, em Instituições Públicas no Brasil. Este trabalho foi executado na forma de pesquisa qualitativa, descritiva e exploratória. No capítulo seguinte é apresentada a metodologia desta pesquisa.

4 METODOLOGIA DA PESQUISA

4.1 Considerações Iniciais

Para Silva & Menezes (2001, p. 9)

“A pesquisa é um trabalho em processo não totalmente controlável ou previsível. Adotar uma metodologia significa escolher um caminho, um percurso global do espírito.”

Este capítulo apresenta a metodologia que determinou a estrutura, os procedimentos e o desenvolvimento do presente trabalho de pesquisa.

4.2 Classificação da Pesquisa

A presente pesquisa caracteriza-se quanto à forma de abordagem do problema como uma pesquisa qualitativa que, segundo Yin (1984), tem a sua essência no uso da observação detalhada do mundo natural, feita pelo pesquisador, e ressaltando que esta observação é necessariamente, baseada em modelo teórico. Para Liebscher (1988) esta abordagem justifica-se quando o fenômeno em estudo é complexo, de natureza social e não tende à quantificação, que nesta pesquisa é a implementação e uso de tecnologias de *Data Warehouse*, em Instituições Públicas no Brasil.

Bogdan e Biklen (apud Triviños, 1987) indicam as principais características da pesquisa qualitativa que servirão de base para este trabalho. São elas:

A pesquisa qualitativa tem como fonte direta dos dados o ambiente natural e o pesquisador como instrumento-chave;

Os dados coletados, em sua maioria, são descritivos;

Os pesquisadores qualitativos preocupam-se com o processo e não simplesmente com os resultados e o produto;

A análise dos dados tende a ser um processo indutivo;

O “significado” é a preocupação essencial da abordagem qualitativa.

Quanto aos seus objetivos, a presente pesquisa caracteriza-se como exploratória descritiva. Segundo Antonio Carlos Gil (apud Silva & Menezes, 2001), a pesquisa exploratória busca tornar o problema mais explícito, enquanto que a pesquisa descritiva objetiva descrever características de uma determinada população, fenômeno ou o estabelecimento de relações entre variáveis. Para Rudio (1986), o pesquisador deve procurar conhecer e interpretar a realidade, sem nela interferir para modificá-la.

Quanto aos procedimentos técnicos, a presente caracteriza-se como uma pesquisa bibliográfica e estudo descritivo e análise documental. Segundo Antonio Carlos Gil (apud Silva & Menezes, 2001), a pesquisa bibliográfica é elaborada a partir de material já publicado em livros, artigos, revistas, jornais e Internet. Segundo Triviños (1987, p.110) “Os estudos descritivos exigem do pesquisador uma série de informações sobre o que se deseja pesquisar(...) O estudo descritivo pretende descrever ‘com exatidão’ os fatos e fenômenos de determinada realidade.”

4.3 Universo e Amostra

O universo da pesquisa é composto por todas instituições públicas brasileiras sob a administração direta ou indireta do Estado.

A pesquisa baseou-se numa amostra não-probabilística intencional. Segundo Marconi & Lakatos (1988), em uma amostra não-probabilística intencional, o pesquisador está interessado na opinião de determinados elementos do universo que são considerados relevantes para escolha, mas não representativos do mesmo. Pressupõe que estes elementos, através de ações ou intenções, têm a propriedade de influenciar a opinião dos demais.

A aplicação de soluções de *Data Warehouse* pelas organizações no Brasil é recente. Durante a pesquisa poucas instituições públicas do Brasil tinham ou estavam implementando *Data Warehouse*, justificando assim a amostra ser intencional.

As Instituições Públicas foram selecionadas considerando-se a relevância, experiência, tamanho e abrangência foram as seguintes:

- Serviço Federal de Processamento de Dados – SERPRO;
- Instituto Brasileiro de Geografia Estatística – IBGE;
- Banco do Brasil S.A;
- Banco Central do Brasil – BC;
- Tribunal Superior Eleitoral;
- Agência Nacional de Energia Elétrica – ANEEL;
- Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA.

4.4 Limites da Pesquisa

Os limites desta pesquisa não são as soluções de implementação e utilização de tecnologias de *Data Warehouse* para as Instituições Públicas no Brasil, mas a pretensa colaboração de fatos que possam auxiliar nas tomadas de decisão, sobre a implementação e utilização dessas tecnologias por essas Instituições.

4.5 Levantamento de Dados

O levantamento dos dados da pesquisa foi efetuado com a utilização dos seguintes instrumentos de coleta de dados:

- Pesquisa bibliográfica;
- Entrevistas semi-estruturadas; e
- Análise documental.

Para Marconi & Lakatos (1988), a finalidade da pesquisa bibliográfica é colocar o pesquisador em contato direto com tudo aquilo que foi escrito sobre determinado assunto.

Segundo Triviños (1987), a entrevista semi-estruturada é um dos principais meios que o pesquisador dispõe para realizar a coleta de dados. Ela possibilita que o

entrevistado alcance a liberdade, dentro do foco principal colocado pelo entrevistador, seguindo a linha de seu pensamento e de suas experiências.

Nas instituições pesquisadas foram realizadas entrevistas semi-estruturadas para os seguintes grupos de entrevistados:

Grupo Usuários - GU: os dirigentes, gerentes e profissionais, que foram, são ou serão usuários finais de tecnologias DW;

Grupo Técnicos - GT: os técnicos responsáveis pelas atividades de Tecnologia da Informação (TI), que foram, são, ou serão envolvidos em atividades de projeto, implementação, manutenção e monitoramento de tecnologias de DW.

Com os entrevistados do primeiro grupo, GU, foram coletados dados que possibilitaram descrever e identificar a importância da utilização de tecnologias de DW pela instituição, focalizando:

- os negócios da Instituição;
- informações executivas, gerenciais e para suporte à tomada de decisão;
- ferramentas de análise multidimensional;
- expectativas, resultados e dificuldades na utilização.

Com os entrevistados do segundo grupo, GT, foram coletados dados que possibilitaram descrever e identificar as atividades de projeto, implementação, monitoramento e manutenção de tecnologias de DW na Instituição, focalizando:

- projeto;
- hardware e software;
- expectativas, resultados e dificuldades para implementação, monitoramento e manutenção.

4.6 Variáveis da Pesquisa

Segundo Triviños (1987), as variáveis são características observáveis de algo, que podem apresentar diversos valores. Sem as variáveis é impossível ter uma idéia clara dos rumos e conteúdos da pesquisa. As hipóteses estão constituídas por uma, duas ou mais variáveis.

A seguir, são apresentadas no Quadro 10, as variáveis que foram consideradas nesta pesquisa, relacionadas aos dados coletados, para analisar a implementação de tecnologias de DW, em uma Instituição Pública:

Quadro 10 - Conjunto de variáveis propostas para analisar a implementação de tecnologias de DW, em uma Instituição Pública.

Variável	Dados
<i>Ambiente da Instituição Pública</i>	missões; metas; atividades; negócios; estrutura organizacional; perfil dos entrevistados
<i>Maturidade</i>	histórico das atividades; liderança ⁵ ; governança ⁶ ; aplicações; nível e disseminação do conhecimento; nível de implementação; nível de utilização; parcerias; expectativas; resultados.
<i>Fontes dos Dados do DW</i>	fontes externas; sistemas operacionais (SPT/OLTP); hardware; ferramentas (software); banco de dados (SGBD); nível de integração; sistemas de gestão empresarial (ERP); gerenciamento de relacionamento com o cliente (CRM); monitoração de acessos na Web; repositório de dados operacionais (ODS);
<i>Ambiente do DW</i>	projeto; arquitetura; modelagem lógica e física dos dados; metodologia de desenvolvimento; hardware; ferramentas (software); banco de dados (SGBD); metadados; segurança; monitoramento; manutenção; alocação dos técnicos.
<i>Acesso aos Dados do DW</i>	sistemas de informações executivas (SIE); sistemas de apoio à decisão (SAD); ferramentas de consultas (SQL); ferramentas de processamento analítico on-line (OLAP) e de prospecção de dados (DATA MINING); Internet e intranet (WEB); tempo de resposta;
<i>Dificuldades para implementação de DW, na Instituição pública</i>	liderança; governança; orçamento; custos; cultura de serviço público; burocracia da administração pública; sistemas legados; resistência dos usuários; quantidade e qualificação dos técnicos.

4.7 Análise e Interpretação dos Dados

A análise de conteúdo foi o método utilizado para análise e interpretação dos dados desta pesquisa. Segundo Bardin (apud Triviños, 1987), a análise de conteúdo é um conjunto de técnicas de análise das comunicações entre os homens, com

⁵ Para o General H. Normam Schwarzkopf (apud Hartman & Sifonis, 2000), “A liderança é uma forte mistura da estratégia com a disposição. Na falta de uma delas, melhor que seja a estratégia”.

⁶ A governança define as estruturas, os papéis, as responsabilidades, políticas e padrões, metas e medições, e a gerência financeira (Hartman & Sifonis, 2000).

ênfase no conteúdo das mensagens, para se obter indicadores que permitam a inferência de conhecimento. Para Triviños (1987), a análise de conteúdo adquire a sua força e seu valor com o respaldo do referencial teórico.

A seguir, são apresentadas no Quadro 11, as etapas do processo de análise e interpretação dos dados através do método de análise de conteúdo:

Quadro 11 - Etapas do processo de análise e interpretação dos dados desta pesquisa.

Etapa	Descrição da Etapa
<i>Pré-análise</i>	transcrição para documentos digitados, dos conteúdos das fitas gravadas e das anotações efetuadas durante as entrevistas semi-estruturadas
<i>Descrição Analítica (Análise Descritiva)</i>	análise descritiva dos documentos digitados na etapa de pré-análise, codificando, classificando e categorizando as informações importantes de acordo com as variáveis previamente definidas na pesquisa e os referenciais teóricos. Os resultados da análise descritiva são, então, processados e apresentados separadamente por variável e por instituição pesquisada.
<i>Interpretação Referencial (Análise Referencial)</i>	análise comparativa, estabelecendo relações entre as instituições pesquisadas, com embasamento nos resultados da etapa de descrição analítica, sendo processada e apresentada separadamente por variáveis levantadas durante a etapa de descrição analítica.

Fonte: adaptado de Triviños (1987)

4.8 Etapas da Pesquisa

A presente pesquisa foi desenvolvida através das seguintes etapas:

- escolha do tema;
- formulação do problema;
- determinação de objetivos;
- revisão de literatura;
- coleta de dados:
 - definição das variáveis a serem analisadas;
 - elaboração dos roteiros de entrevistas semi-estruturadas;
 - realização das entrevistas semi-estruturadas nas instituições; públicas;
- tabulação de dados:

- análise de descritiva das entrevistas de acordo com as variáveis e o referencial teórico, bem como a análise referencial;
- apresentação dos resultados da análise descritiva e referencial, por instituição;
- conclusão .

4.9 Desenvolvimento da pesquisa

O desenvolvimento da pesquisa teve duração de 3 meses, que envolveu contatos iniciais com as Instituições Públicas, entrevistas, coleta de documentos, análise e tabulação das informações e validação das informações pelas Instituições pesquisadas.

Na etapa de contatos iniciais, pode-se destacar o importante papel da Secretaria de Administração Estratégica – SEA da Embrapa que funcionou como facilitadora, intermediando os contatos iniciais com as Instituições Públicas e fornecendo carta de apresentação para as mesmas.

As entrevistas foram feitas em datas esparsas, pois foi verificada a disponibilidade de tempo dos entrevistados. Alguns dados tiveram que ser coletados após as entrevistas, sendo necessários em alguns casos, contatos telefônicos com outros envolvidos no processo para complementação das informações.

Após a tabulação dos dados, os mesmos foram enviados para a validação pelos responsáveis, dependendo da disponibilidade de tempo dos mesmos, o que demandou um tempo maior no processo de pesquisa.

4.10 Considerações Finais

Este capítulo abordou a metodologia que determinou o planejamento e a implementação da presente pesquisa. No capítulo seguinte serão apresentados os resultados desta pesquisa.

5 RESULTADOS DA PESQUISA

5.1 Considerações Iniciais

Este capítulo apresenta os resultados desta pesquisa, de acordo com o método de análise de conteúdo, que foi utilizado para análise e interpretação dos dados levantados por esta pesquisa, conforme citado no capítulo anterior, neste capítulo são apresentados os resultados Análise Descritiva e da Análise Referencial. Veja o Quadro 11.

Os resultados da Análise Referencial são elaborados a partir dos resultados da Análise Descritiva e constituem a síntese da pesquisa. Nos resultados da Análise Descritiva, estão os detalhes de cada Instituição pesquisada, colocados de forma descritiva. Com o objetivo de tornar a leitura mais objetiva neste capítulo, os resultados da Análise Referencial são apresentados antes dos resultados da Análise Descritiva.

5.2 Análise Referencial

A seguir, nos Quadros 12, 13 e 14, são apresentados os resultados da análise referencial, separadamente por variáveis relevantes, que foram levantadas durante a análise descritiva.

Quadro 12 – Resultados da Análise Referencial

Instít. Pública	Missão	Assuntos/Negócios	Perfil dos Entrevistados	Metodologia de Desenvolvimento
Embrapa	Viabilizar soluções para o desenvolvimento sustentável do agronegócio brasileiro.	Modelo de Gestão Estratégico; Sistema de avaliação das unidades da Embrapa; Sistema Embrapa de Planejamento; Informações sócio-econômicas sobre pesquisa agropecuária; Gestão de Pessoal; e Sistema de Premiação.	Chefe da DTI; Coordenadora de Processos e Software; Administrador de Banco de Dados; Administrador de Dados; e Analistas de Negócios da SEA.	Metodologia de desenvolvimento de DW da Oracle; ou baseada na metodologia proposta por Kimball (1998a)
ANEEL	Proporcionar condições favoráveis para que o mercado de energia elétrica se desenvolva com equilíbrio entre os agentes em benefício da sociedade.	Indicadores de qualidade de energia elétrica das concessionárias; Informações contábeis das concessionárias; e Central de ouvidoria (<i>call center</i>).	Chefe de Gestão do Conhecimento; Administrador de Banco de Dados; e Administrador de Dados.	Desenvolvimento baseado na metodologia proposta por Kimball (1998a) do ciclo de desenvolvimento do DW
Serpro	Fornecer soluções, com base em tecnologias da informação, para êxito das decisões e operações da administração pública.	Gestão de Pessoal da União; Comércio Exterior; Patrimônio da União; Ações do Governo Federal; Receita da União; e Tesouro Nacional.	Chefe Departamento de Consultoria Tecnológica em Integração de Dados; Analista de Negócios; e Consultor de DW.	Desenvolvimento baseado na metodologia proposta por Kimball (1998a) do ciclo de desenvolvimento do DW
TSE/SSE	Prover informações de dados estatísticos acerca da distribuição e perfil do eleitorado nacional e resultado das eleições.	Cadastro Nacional de Eleitores – CNE.	Coordenador de Sistemas Administrativos; Chefe da Seção de Estatística Eleitoral – SEE.	Metodologia CTM/IS – GER, baseada no método JAD: - Planejamento e entendimento; - Levantamento dos requisitos; - Especificação do <i>Data Mart</i> ; Construção do <i>Data Mart</i> ; - Testes, treinamento, implantação e monitoramento.
Banco Central do Brasil	Ser o agente da sociedade brasileira na promoção da estabilidade do poder de compra da moeda brasileira.	Câmbio; Dívida Externa; Supervisão do mercado financeiro; e Comércio Exterior.	Coordenador de DW; e Consultor de DW.	Metodologia própria baseada no método JAD: - Definição do Escopo do <i>Data Mart</i> ; - Levantamento das necessidades; -Modelagem multidimensional; - Projeto físico, implementação e carga do <i>Data Mart</i> no DW; - Projeto e implementação de ETC; - Testes, avaliação, homologação, treinamento, implantação e monitoramento.
Banco do Brasil	É ser a solução em serviços e intermediações financeiras, atender às expectativas de clientes e acionistas, fortalecer o compromisso entre funcionários e a Empresa ser útil à sociedade.	Gestão de Pessoal; e Unidades Estratégicas de Negócios: UEN Comercial; UEN Governo; UEN Internacional; UEN Crédito; UEN Rural e Agroindustrial; e UEN Seguridade.	Analista de Suporte a Bancos de Dados; Administrador de Dados; Administrador de Banco de Dados; Analista de Negócios Área de Seguros; e Analista do OLTP de Seguros.	Metodologia própria baseada nos livros dos escritores de DW, tais como Kimball, Inmon, entre outros. A seguir as fases: - Levantamento das Necessidades através de uma matriz fato x dimensão; -Construção no DW corporativo (plataforma alta); - Implementação de ETC p/ o DW; - Construção do <i>Data Mart</i> (plataforma baixa); - Carga no <i>Data Mart</i> ; -Testes, implantação, treinamento e monitoramento.
IBGE	Retratar o Brasil com informações, necessárias ao conhecimento da realidade e ao exercício da cidadania	Série agropecuária; Pesquisa Nacional por Amostra de domicílios; Censo 1991; e Censo 2000.	Chefe do Departamento de Metodologia e Banco de Dados; e Chefe do Projeto Banco de Microdados Multidimensionais.	Metodologia formal de desenvolvimento de DW da ORACLE, baseada na UML.

Quadro 13 – Resultados da Análise Referencial (continuação)

Instit. Pública	Estratégia	Tecnologias das Fontes de Dados	Back Room (ETC)	Metadados	Granularidade e Particionamento
Embrapa	Bottom-up ou Intermediária	SGBD Ingres (plataforma baixa); Linguagens de acesso nativo ao SGBD Ingres; Ferramenta estatística SAS; Arquivos Paradox; Planilha Eletrônica MS-Excel; Arquivos Texto; e Editor de Texto MS-Word.	Não implementado; Probabilidade para uso de tecnologias Oracle	não implementado	não implementado
ANEEL	Bottom-up	Microsoft SQL Server 7.0 (plataforma baixa).	Microsoft Data Tranformation Services – DTS	não implementado	De acordo com os requisitos de análise e desempenho
Serpro	Intermediária	SGBD ADABAS (plataforma alta); Linguagens de programação NATURAL de acesso nativo ao SGBD ADABAS; e Ferramenta estatística SAS.	<ul style="list-style-type: none"> •Plataforma Alta <ul style="list-style-type: none"> ○ NATURAL; •Transmissão (Alta x Baixa) <ul style="list-style-type: none"> ○ Tempus Connectivity Solution – TCS; •Plataforma Baixa <ul style="list-style-type: none"> ○ SAS; OraclePL/SQL; ○ Database Procedures. 	<ul style="list-style-type: none"> •MS-Word; •HTML; •Microstrategy; •Oracle Discoverer. 	O nível de granularidade é o mais atômico possível o Tema Pessoal, está sendo usado particionamento físico por mês
TSE/SSE	Bottom-up	SGBD ORACLE 8i (plataforma baixa); Linguagens de programação de acesso nativo ao SGBD ORACLE; Planilha Eletrônica MS-Excel; Arquivos Texto; Editor de Texto MS-Word.	<ul style="list-style-type: none"> •Oracle Data Mart Builder; •Database procedures. 	não implementado	De acordo com os requisitos de análise e desempenho
Banco Central do Brasil	Intermediária	SGBD ADABAS (plataforma alta); Linguagem de programação NATURAL de acesso nativo ao SGBD ADABAS.	<ul style="list-style-type: none"> •Plataforma Alta <ul style="list-style-type: none"> ○ NATURAL; •Plataforma Baixa <ul style="list-style-type: none"> ○ Database procedures; 	<ul style="list-style-type: none"> •CASE System Architect; •Microstrategy. 	O nível de granularidade é mais atômico do que o desejado, mas ainda é a solução para atender os usuários. Estudos para implementação de um ODS.
Banco do Brasil	Top-down	SGBD ADABAS (plataforma alta); SGBD DB2 (plataforma alta); Linguagens de programação NATURAL e COBOL;	<ul style="list-style-type: none"> •Plataforma Alta (DW) <ul style="list-style-type: none"> ○ NATURAL; ○ COBOL; •Transmissão (Alta x Baixa) <ul style="list-style-type: none"> ○ Visual Warehouse; •Plataforma Baixa (<i>Data Marts</i>) <ul style="list-style-type: none"> ○ Visual Warehouse; 	•CASE Erwin;	No DW corporativo o nível de granularidade é o mais atômico possível. Nos Data Marts de acordo com os requisitos de análise e desempenho
IBGE	Intermediária	Arquivo em fita magnética (plataforma alta); Banco de Dados Estatístico – RAPID; e Ferramenta estatística SAS.	Ferramentas de carga do SGBD Oracle	Metamodelo desenvolvido em UML / Java / SGBD Oracle 8i.	No DW procura-se trabalhar com uma granularidade mais fina possível

Quadro 14 – Resultados da Análise Referencial (continuação)

Instit. Pública	Modelagem dos Dados	Ambiente Físico dos Dados	Front Room (Informação)	Dificuldades
Embrapa	Modelagem multidimensional (star schema)	<ul style="list-style-type: none"> •SGBD Oracle 8i; •Servidor Sun Enterprise 5000; •Sistema Operacional Sun Solaris 2.8. 	Oracle Discoverer	Quantidade reduzida de profissionais de TI; Pouca percepção das tecnologias de DW pelo estratégico da organização e pelos profissionais de TI; Recursos orçamentários reduzidos; Estrutura organizacional descentralizada; Sistemas operacionais redundantes e sem integração; e Burocracia do processo de licitação.
ANEEL	Modelagem multidimensional (star schema)	<ul style="list-style-type: none"> •SGBD MS SQL Server 7.0; •Servidor (Pentium) provisório; •Sistema Operacional Windows NT4.0 	Planilha eletrônica MS-Excel (tabelas e gráficos dinâmicos conectados ao MS SQLServerOlap)	Quantidade reduzida de profissionais de TI; Capacitação de tecnologias de DW para os profissionais de TI que na Instituição na maioria é terceirizada; Resistências às tecnologias de DW pelos profissionais TI envolvidos com os sistemas operacionais; Indisponibilidade dos usuários finais para montar as análises de dados.
Serpro	Modelagem multidimensional, star schema e snowflake	<ul style="list-style-type: none"> •SGBD Oracle 8i; •Serv. Digital Alpha Server 8400; •Sist.Oper. Unix 4.0B 	<ul style="list-style-type: none"> •Olap Microstrategy 6; •Oracle Discoverer. 	Falta de envolvimento/comprometimento de profissionais de TI com os perfis necessários; Falta de definição de um escopo para a primeira versão dos projetos; Falta de uma previsão orçamentária compatível com o projeto; e Burocracia do processo de licitação.
TSE/SSE	Modelagem multidimensional (star schema)	<ul style="list-style-type: none"> •SGBD Oracle 8i; •2 servidores Pentium III Xeon 500; •Sist.Oper. Windows NT 4.0 	Oracle Discoverer	Rotatividade de funcionários públicos; Implementação próxima ao período eleitoral; e Falta de disponibilidade dos profissionais de TI para capacitação das tecnologias de DW.
Banco Central do Brasil	Modelagem multidimensional (snowflake)	<ul style="list-style-type: none"> •SGBD DB2/UDB (IBM); •Servidor Sun Enterprise 4000; •Sist.Oper. Sun Solaris 2.6. 	Olap Microstrategy 7	Tecnologias de DW ainda não disseminadas em todas as áreas que, potencialmente, poderiam utiliza-las; Quantidade reduzida de profissionais de TI; e Recursos orçamentários reduzidos.
Banco do Brasil	Modelagem ExR no DW; e Modelagem Multidimensional (star schema) nos Data Marts.	<ul style="list-style-type: none"> •Plataforma Alta <ul style="list-style-type: none"> ○ SGBD DB2; ○ Mainframe IBM; ○ Sist.Oper. OS390 versão 7; •Plataforma Baixa <ul style="list-style-type: none"> ○ SGBD DB2/UDB; ○ Servidor IBM F50/RS ○ Sist.Oper. Unix AIX 4.2. 	<ul style="list-style-type: none"> •Business Objects; •DB2 OLAP Server; 	Sistemas operacionais com tecnologias diversas e pouca integração; Resistências às tecnologias de DW pelos profissionais TI envolvidos com os sistemas operacionais; Atendimento de demanda muito grande de informações estratégicas onde algumas área não foram atendidas no tempo esperado; e Reestruturação organizacional em curso.
IBGE	Modelagem multidimensional (star schema)	<ul style="list-style-type: none"> •SGBD Oracle 8i; •Serv. IBM SP2; •Sist.Oper. Unix AIX 	Interfaces em Java (Web) e SAS AF, com o metamodelo associado às mesmas.	Quantidade reduzida de profissionais de TI; Recursos orçamentários dependentes de momentos de financiamento atípicos como o Censo 2000; e Enormes volumes de dados que precisam ser carregados no DW;

5.3 Análise Descritiva

A seguir, são apresentados os resultados da Análise Descritiva das Instituições Públicas pesquisadas, separadamente por Instituição e as variáveis desta pesquisa. Esses resultados são a síntese das análises das entrevistas e dos documentos levantados em cada instituição. Deve-se considerar que os resultados desta análise, são referentes ao período de realização desta pesquisa.

5.3.1 Serviço Federal de Processamento de Dados - SERPRO

5.3.1.1 Ambiente da Instituição Pública

O Serviço Federal de Processamento de Dados – SERPRO, é uma empresa pública, vinculada ao Ministério da Fazenda, dotada de personalidade jurídica de direito privado, com patrimônio próprio, autonomia administrativa e financeira.

A missão do SERPRO é fornecer soluções, com base em tecnologias da informação, para o êxito das decisões e operações da administração pública federal, com inovação, qualidade e segurança, a preços competitivos.

A modelo de gestão do SERPRO é estruturado por processos. Essa estrutura é composta da diretoria onde se encontra o presidente, abaixo da diretoria tem as Unidades de Processos Corporativos – UC que são unidades estratégicas voltadas para inovação tecnológica. Abaixo das UC's estão as Unidades de Negócio e as Unidades de Gestão e Infra-estrutura.

As Unidades de Negócios, são as superintendências que atendem os grandes clientes do Serpro. As Unidades de Gestão e Infra-estrutura, são as superintendências de apoio tecnológico às unidades de negócio. Para esta pesquisa destacamos as seguintes superintendências:

- Planejamento, Orçamento e Gestão – SUNAR;
- Administração Tributária – SUNAT; e

- Consultoria Tecnológica – SUPCT.

A SUNAR e a SUNAT, são Unidades de Negócio. A SUPCT é uma Unidade de Gestão e Infra-estrutura.

5.3.1.2 Maturidade

Os sistemas operacionais vinham atendendo aos clientes do SERPRO de forma satisfatória, porém, esse atendimento era direcionado aos usuários operacionais. Os módulos gerenciais desses sistemas não atendiam as demandas de informações não estruturadas para a tomada de decisão, chamadas de Apurações Especiais (AESPs). Essas demandas eram constantes e na maioria das vezes, não era possível atender dentro da expectativa do cliente em relação ao prazo, em muitos casos, quando a demanda era atendida, a decisão já havia sido tomada com a informação que o cliente dispunha. Em muitos casos, após analisar o resultado, o cliente solicitava modificações no mesmo, gerando nova demanda a qual passava a concorrer com outras que já estavam na fila, enfim, não se conseguia atender bem os clientes tomadores de decisão.

Na urgência de uma solução para o problema, foram testadas algumas tecnologias que ofereciam SAD (Sistemas de Apoio à Decisão), porém, sem sucesso, gerando até um certo descrédito nos clientes, quando se falava em nova solução de suporte à decisão.

Em meados de 1995 o SERPRO tomou conhecimento da tecnologia de DW. Começou então, uma parceria de risco com alguns fornecedores de *softwares* de DW, para desenvolvimento de protótipos para algumas áreas de negócios atendidas pelo SERPRO, sendo uma das quais a área de Recursos Humanos e Folha de Pagamento dos Servidores Públicos Federais Cíveis do Poder Executivo. Essa área de negócio é atendida pela SUNAR que, juntamente com a SUPCT desenvolveu o protótipo do Tema⁷ Pessoal.

Concluído o protótipo do Tema Pessoal, foi realizada uma apresentação para os usuários do Ministério do Planejamento, Orçamento e Gestão – MP (na época Ministério da Administração Federal e Reforma do Estado - MARE), os quais demonstraram grande interesse na solução apresentada. Então, usando tecnologias de DW, foi desenvolvida a

⁷ Tema é a denominação da área de negócio

primeira versão do Tema Pessoal, que foi o primeiro tema a entrar em produção no dia 02 de janeiro de 1997.

Os maiores usuários do Tema Pessoal são as áreas que fornecem informações para apoiar os tomadores de decisão e a área de auditoria do MP.

Novos temas foram desenvolvidos no DW, sendo que, alguns foram implementados em produção e outros estão paralisados.

O SERPRO tem desenvolvido o DW de forma corporativa e incremental, com soluções para os clientes nas áreas de Pessoal, Comércio Exterior, Patrimônio de União, Ações do Governo Federal, Receita, Tesouro Nacional, etc. O fato é que esse DW passou a ser a solução proposta pelo SERPRO, para apoiar o processo decisório do Governo Federal.

A Instituição sempre envia representantes para seminários, congressos e treinamentos específicos de tecnologias de DW.

5.3.1.3 Fontes dos Dados do DW

As principais fontes de dados do DW são os sistemas OLTP que atendem aos clientes do SERPRO. A maioria desses sistemas foi desenvolvida em plataforma alta, ou seja, em computador de grande porte, utilizando o SGBD ADABAS. Esses sistemas possuem bases de dados relacionadas com o assunto de interesse e que estão disponíveis para os processos de extração, transformação e carga - ETC. Para esses processos, atualmente o SERPRO está utilizando soluções próprias, porém, no momento desta pesquisa estava avaliando alguns *softwares* prontos, com o objetivo de vir a utilizá-los também.

A integração de várias fontes é feita através do compartilhamento do significado dos dados. É uma característica da qual permite que informações de fontes distintas possam ser vinculadas entre si e apresentadas com significado correto, lado a lado, no mesmo relatório.

No caso do Tema Ações do Governo Federal (denominado ÍCONE), as principais fontes de dados externos são:

- Instituto Brasileiro de Geografia Estatística – IBGE;
- Caixa Econômica Federal – CEF;
- Banco do Brasil;
- Tribunal Superior Eleitoral – TSE;
- Banco Nacional de Desenvolvimento Econômico e Social - BNDES;
- Banco do Nordeste – BNB;
- Secretaria do Tesouro Nacional – STN; e
- Ministérios.

5.3.1.4 Ambiente do DW

A metodologia de desenvolvimento do DW no SERPRO é baseada na metodologia proposta por Kimball (1998a) do ciclo de desenvolvimento do DW.

Para a infra-estrutura de DW, foram adquiridos equipamentos com grandes áreas de disco para armazenar os grandes volumes de dados e processadores paralelos para proporcionar melhores tempos de resposta.

A infra-estrutura do DW é em plataforma baixa, ou seja, em arquitetura Cliente x Servidor.

Em função da grande maioria dos sistemas OLTP estarem em ambiente computacional de plataforma alta, persistidos em SGBD ADABAS, os processos de extração e parte dos processos de transformação, são executados utilizando soluções próprias do SERPRO, através de ferramenta de programação NATURAL, dependendo do projeto, portanto, a *Staging Area* para transformar e integrar os dados operacionais antes de serem carregadas para o DW, começa no próprio ambiente operacional. Os arquivos resultantes desses processos de extração e transformação, no ambiente operacional, são transportados para a plataforma baixa onde se encontra o DW, através da ferramenta de transmissão de arquivos *Tempus Connectivity Solution* – TCS. Após o processo de

transporte, ocorrem ainda processos de transformação através da ferramenta de estatística SAS, onde os dados sofrem as últimas transformações e são carregados no DW, através das ferramentas *Oracle PL/SQL* e *Oracle SQL*, ambas da ORACLE. No momento desta pesquisa, algumas ferramentas de ETC, estavam sendo testadas e avaliadas.

Em alguns Temas, os processos de transformação estão em plataforma baixa, no ambiente computacional do DW. Em outros, parte ou todos os processos de transformação estão em plataforma alta, no ambiente computacional dos sistemas OLTP.

Os dados do DW estão armazenados ao nível de granularidade mais atômico possível e de forma corporativa incremental, persistidos em um *SGBD Oracle 8i*, que está residente em um computador servidor *Digital Alpha Server 8400*, composto de 8 processadores paralelos de 625 Megahertz cada, 6 Gigabytes de memória principal (*Radom Aleatory Memory* – RAM), 750 Gigabytes de espaço. Neste servidor o sistema operacional é o *UNIX 4.0B*.

Os usuários acessam os dados do DW, através de microcomputadores PC com sistema operacional Windows versão 95 ou superior. Esses microcomputadores estão em rede local, através do sistema operacional Windows NT 4.0. Essas redes locais estão conectadas a intranet do SERPRO.

Para o Tema Pessoal, está sendo usado particionamento físico por mês. Para os demais, está sendo usado o próprio particionamento do ORACLE.

A modelagem dos dados é multidimensional, tanto na visão estrela (*star schema*), como na visão flocos de neve (*snowflake*). A visão estrela é mais utilizada para comunicação com os clientes, enquanto a visão flocos de neve é mais utilizada pela equipe de desenvolvimento e implementação.

Para definir e atualizar os metadados é utilizado, a ferramenta *Word* da *Microsoft*. Posteriormente esses metadados são disponibilizados para acesso via WEB, através de HTML. Também são utilizados recursos da ferramenta OLAP, que oferece campos destinados à descrição do atributo ou do fato.

5.3.1.5 Acesso aos Dados do DW

Para acessar os dados do DW e realizar suas análises, os usuários finais utilizam ferramentas OLAP, são elas, a *MicroStrategy Web e MicroStrategy Agent* da MICROSTRATEGY BUSINESS INTELLIGENCE e a *Oracle Discoverer* da ORACLE. Os profissionais de TI, responsáveis pelo desenvolvimento e manutenção, além dessas ferramentas, utilizam também as ferramentas de SQL da ORACLE.

Os resumos e agregações são disponibilizados para acesso pelos usuários finais no mesmo ambiente do DW corporativo, podendo o usuário acessar, até mesmo, os dados mais atômicos, obedecendo à abrangência do Tema.

5.3.1.6 Dificuldades para implementação de DW

Durante o processo de levantamento dos dados na instituição para essa pesquisa, foram identificadas as seguintes dificuldades de implementação:

Falta de envolvimento/comprometimento de profissionais, com os perfis necessários. Para o desenvolvimento de um projeto de DW, é imprescindível que a equipe designada para o trabalho, além de estar envolvida e comprometida com o resultado deste trabalho, reúna, também, o domínio completo dos seguintes pontos:

- Conhecimento de modelagem multidimensional;
- Conhecimento do negócio do cliente;
- Conhecimento das necessidades do cliente;
- Conhecimento das bases e sistemas de origem (sistemas legados);
- Falta delimitação de um escopo para a primeira versão do projeto. O cliente tende a querer todas os dados existentes nos sistemas legados (fontes), desde a primeira versão. Isso, geralmente aumenta a complexidade do modelo, gerando dificuldade na utilização pelos usuários, o que leva ao desânimo e abandono, deixando a impressão de que DW é muito complicado e difícil de ser usado;
- A Falta de uma previsão orçamentária compatível com o projeto;

- As colocações anteriores são inerentes a qualquer tipo de empresa (pública ou privada), porém, a Empresa Pública convive com mais um fator, que é o processo de concorrência pública pelo qual normalmente, é demorado, levando muitas vezes ao atraso no desenvolvimento do projeto, conseqüentemente, atrasos na disponibilização para os clientes.

5.3.2 Banco do Brasil

5.3.2.1 Ambiente da Instituição Pública

O Banco do Brasil é uma instituição financeira de economia mista com fins lucrativos.

A missão do Banco do Brasil é ser a solução em serviços e intermediações financeiras, atender às expectativas de clientes e acionistas, fortalecer o compromisso entre funcionários e a Empresa e ser útil à sociedade.

O Banco do Brasil é a maior instituição financeira da América Latina. Seu conglomerado constitui-se das seguintes empresas: BB Previdência, Cobra, BB Leasing, BB Cartões, BB Banco de Investimentos, BB Financeira, BB Corretora, BB DTVM, BB Europe, BB AG. Viena, BB Leasing Co., BAMB, Brasilsaúde, Brasilcap, Brasilprev, SBCE, Upsicard, Visanet, Tecban, Aliança do Brasil, Brasilseg Participações, Brasil Veículos, BB Turismo, BB Securities.

A estrutura organizacional interna do Banco do Brasil constitui-se da Assembléia Geral de Acionistas, do Conselho Fiscal, do Conselho de Administração, da Auditoria, da Unidade de Estratégia, Marketing e Comunicação, da Consultoria Jurídica, da Secretaria Executiva, da Presidência e de sete diretorias que constituem as seguintes Unidades: Distribuição, Varejo e Serviços, BB Franquia, Banco Postal, Comercial, Internacional, BB Internet, Infra-estrutura, Tecnologia, Governo, Rural e Agroindustrial, Crédito, Reestruturação de Ativos Operacionais, Gestão de Pessoas, Gestão de Recursos de Terceiros, Finanças, Gestão de Riscos, Mercado de Capitais, Controladoria, Controles Internos, Contadoria. A presidência do BB é indicada pelo Governo Federal. Os diretores também podem ser indicados pelo Governo.

Para a coordenação do suporte ao cliente em seus 7.498 pontos de atendimento (2.908 agências) em todo o Brasil e em 33 dependências no exterior, o Banco do Brasil conta ainda com suas Superintendências Estaduais e Regionais.

Destacamos para esta pesquisa as seguintes gerências da Unidade de Função Tecnologia, envolvidas no desenvolvimento dos *Data Marts* e do DW corporativo:

- Gerência do Sistema de Informações - Produtos – GESIP;
- Gerência do Sistema de Informações – Gestão – GESIG;
- Gerência de Tecnologia de Software e Hardware – TECHS;
- Gerência de Administração de Dados e Sistemas Especiais- DEPDA;
- Gerência de Processamento de Informações – DEPRO.

Destacamos também as seguintes Unidades Estratégicas de Negócios que tiveram *Data Marts* projetados:

- UEN Comercial;
- UEN Governo;
- UEN Internacional;
- UEN Crédito;
- UEN Rural e Agroindustrial;
- UEN Seguridade.

A Unidade de Função UF Gestão de Pessoas, também teve um *Data Mart* projetado.

5.3.2.2 Maturidade

No momento desta pesquisa, o Banco do Brasil estava passando por um processo de reestruturação organizacional de sua área de tecnologia, porém alguns *Data Marts* já estavam implantados para algumas áreas negociais, tais como a Unidade Estratégica de Negócio Seguridade.

O processo de DW começou no Banco do Brasil devido à necessidade da Controladoria do Banco em obter informações precisas para cada área comercial e de tentar padronizar essas informações. Havia, na época, uma deficiência de informações, devido à dificuldade de disponibilizar dados a partir dos sistemas transacionais. Um pouco antes do processo de DW, os usuários de cada área podiam construir suas próprias consultas ao SGBD DB2 através do SQL, só que o aumento dessas consultas passou a concorrer com o atendimento On-line das agências e houve a necessidade de fechar as consultas SQL ao usuário. Passaram a ser liberadas apenas consultas via SQL Batch executado em horário noturno, ainda assim com controle de consumo de recursos.

A Controladoria fez um levantamento, por dois anos, das informações necessárias para cada área, sendo gerado um documento com o levantamento dessas informações. Esse documento foi entregue à Administração de Dados por volta de março de 1999, para que a mesma providenciasse as informações pedidas.

Essa demanda gerou uma pesquisa com o objetivo de tentar obter uma tecnologia que atendesse tal necessidade. A solução encontrada foi a implantação de tecnologias de DW.

As necessidades de informações gerenciais foram mapeadas. Foram criadas planilhas de definição dessas necessidades de informação e elaborado um plano de arquitetura e infra-estrutura; enfim, toda estrutura foi construída e está funcionando. Porém, investimentos para implementação do DW em novas áreas estão suspensos até que o processo de reestruturação organizacional da área tecnológica termine e sejam definidas novas metas para o DW e para a área de sistemas.

5.3.2.3 Fontes dos Dados do DW

O Banco do Brasil possui um legado de sistemas transacionais que está funcionando e é estável. Portanto, a maior fonte dos dados de DW vem deste legado. As informações são retiradas dos SGBDs ADABAS e DB2 para carga do DW corporativo. A linguagem principal para extração dos dados e carga do DW é o COBOL e NATURAL. O COBOL mostrou um melhor desempenho para retirada das informações de carga do DW do que o NATURAL. Portanto, a maioria dos programas de extração está hoje em COBOL.

Não são utilizadas fontes de dados externas para alimentação do DW.

5.3.2.4 Ambiente do DW

O ambiente de DW foi construído de forma científica, baseado nos livros dos escritores de DW, tais com Kimball, Inmon, entre outros.

A estratégia de implementação adotada foi pela construção de um DW corporativo em plataforma alta (mainframe) e de *Data Marts* por negócio/assunto em plataforma baixa. O DW corporativo é carregado, semanalmente, a partir dos diversos sistemas transacionais, conforme foi citado na seção anterior desta pesquisa. Os *Data Marts* são derivados do DW corporativo. O DW corporativo é modelado no nível de granularidade mais atômico possível e os *Data Marts* de acordo com os requisitos de análise demandados pelos usuários finais, bem como, a preocupação com o desempenho.

O modelo de dados do DW corporativo foi construído de maneira incremental, no nível de granularidade mais atômico possível, levando à elaboração de um grande modelo de Entidade e Relacionamento - ER que, a princípio, contemplava a área comercial do Banco. Após algumas tentativas, chegou-se a um modelo que utiliza chaves artificiais para qualquer tabela, resguardando as chaves naturais. O histórico também é guardado no modelo criado.

A fim de melhor atender às necessidades de informações consolidadas para os gestores de cada negócio/assunto, os dados dos *Data Marts* são modelados através da modelagem multidimensional utilizando-se o modelo *Star Schema*, onde é feita a

consolidação dos dados por fatos e dimensões. O modelo *Star Schema* foi seguido de acordo com a literatura, não havendo deturpações. Não há relacionamentos entre fatos e também não há relacionamentos entre dimensões.

Os *Data Marts* são implementados de acordo com as necessidades de informações de cada área. A equipe de Administração de Dados estabeleceu sua própria metodologia de levantamento de informações junto a cada área de negócio, onde é feita a coleta dos requisitos necessários utilizando-se uma matriz de fato x dimensão.

No momento desta pesquisa, ainda não haviam sido implementados metadados para o ambiente de DW. Para os modelos de dados e o dicionário de dados, foi utilizada a ferramenta CASE Erwin.

A infra-estrutura do DW corporativo é em plataforma alta, onde os dados são armazenados em um *SGBD DB2* instalado em um mainframe IBM de última geração, com sistema Operacional OS390 versão 7 que foi comprada especificamente para o DW.

A infra-estrutura dos *Data Marts* é em plataforma baixa, onde os dados são armazenados em um *SGBD DB2* instalado em um computador servidor IBM F50 / RS , composto de 2 (dois) processadores paralelos de 332 Megahertz cada, 1 (um) Gigabyte de RAM, 6 (seis) discos de 4,5 Gigabytes cada. Neste servidor o sistema operacional é o *UNIX AIX 4.2*. No momento desta pesquisa os *Data Marts* estavam sendo migrados para um computador servidor IBM SP50 com 4 (quatro) processadores paralelos de 332 Megahertz cada.

A ferramenta *Visual Warehouse* é utilizada para transferir as informações do DW corporativo para os *Data Marts*. Outras ferramentas de automação de extração de dados dos sistemas transacionais estavam sendo testadas quando começou a reformulação do Banco, o que adiou um pouco o processo de aquisição de novas ferramentas.

5.3.2.5 Acesso aos Dados do DW

Para consulta aos *Data Marts* foi utilizado o *Business Objects*, encontrando-se em fase de testes o uso do *DB2 OLAP Server*.

O usuário final só acessa os dados dos *Data Marts*. Num primeiro momento foi construído um DW tático para que o usuário pudesse extrair seus dados através de SQL. A navegação nesse DW era muito difícil e muitas vezes o usuário tinha que conhecer detalhes técnicos do mesmo para a obtenção de uma melhor performance. Era também comum a necessidade de *join* de várias tabelas, o que tornava complexas as consultas. Esses e outros fatores relacionados levaram à desinstalação do DW tático.

O acesso aos dados dos *Data Marts* é feito através da ferramenta *Business Objects*, estando em teste o uso da ferramenta *IBM DB2 OLAP Server*.

5.3.2.6 Dificuldades para implementação de DW

O Banco do Brasil possui uma série de sistemas e equipamentos que fazem parte do seu legado. Os dados são armazenados em vários bancos de dados e utilizam várias linguagens. Fora isso, o Banco possui diversas áreas negociais que têm necessidades de informações variadas e distintas. O DW surgiu como facilitador para a obtenção de informações gerenciais a partir de estudos efetuados pelo escalão médio da Empresa, não tendo sido submetido à alta administração.

O conhecimento para a construção do DW foi obtido através dos livros e através de tentativas e erros da equipe. Não houve uma assessoria especializada para indicar quais os melhores caminhos a serem seguidos. Entretanto, houve troca de informações com outras empresas públicas que também estavam implementando o DW.

O pessoal responsável pelos sistemas transacionais ofereceu um certo grau de resistência no que se referia a atender solicitações de informações, porque estas informações não eram consideradas prioritárias, justamente pelo fato de não serem geradoras de receita direta para o Banco.

O DW no Banco do Brasil foi pouco utilizado, mesmo tendo sido construídos diversos *Data Marts*. Poucas informações foram disponibilizadas.

O tempo para implantação de um DW foi também o grande impeditivo à sua implantação total. Começou a haver uma demanda muito grande de informações estratégicas e por várias questões algumas áreas não foram atendidas no tempo esperado,

como no caso da Unidade Estratégica de Negócios Rurais. Levou-se muito tempo tentando obter as informações com precisão e qualidade, e quando as mesmas estavam quase disponíveis, iniciou-se a reestruturação organizacional da área tecnológica, o que ocasionou a parada no processo de construção de novos *Data Marts*.

Hoje, no processo de reestruturação organizacional em curso, houve o redirecionamento para a implantação apenas de *Data Marts* isolados, tendendo a acabar com o DW corporativo. Cada área teria apenas as informações necessárias para ela, sem haver integração entre os *Data Marts*.

5.3.3 Instituto Brasileiro de Geografia Estatística - IBGE

5.3.3.1 Ambiente da Instituição Pública

O Instituto Brasileiro de Geografia e Estatística – IBGE é uma instituição da administração pública federal, subordinado ao Ministério do Planejamento, Orçamento e Gestão, que se constitui no principal provedor de dados e informações do país, que atendem às necessidades dos mais diversos segmentos da sociedade civil, bem como, dos órgãos das esferas governamentais federal, estadual e municipal. Sua missão é “Retratar o Brasil com informações, necessárias ao conhecimento da sua realidade e ao exercício da cidadania”.

O IBGE oferece uma visão completa e atual do País, através do desempenho de suas principais funções: produção e análise de informações estatísticas, coordenação e consolidação das informações estatísticas, produção e análise de informações geográficas, coordenação e consolidação das informações geográficas, estruturação e implantação de um sistema de informações ambientais, documentação e disseminação de informações e, coordenação dos sistemas estatístico e cartográfico, nacionais.

O IBGE possui quatro diretorias e dois outros órgãos centrais: Para que suas atividades possam cobrir todo o território nacional, o IBGE possui a rede nacional de pesquisa e disseminação, assim discriminada: 8 Departamentos Regionais (BA/CE/DF/MG/PA/PE/PR/SP), 5 Divisões de Geociências (BA/CE/DF/PA/SC), 27 Divisões de Pesquisa (26 nas capitais dos estados e 1 no Distrito Federal), 27 Setores de

Documentação e Disseminação de Informações (26 nas capitais e 1 no Distrito Federal), 544 Agências de Coleta de dados nos principais municípios.

5.3.3.2 *Maturidade*

No momento desta pesquisa, o DW do IBGE estava em fase de implementação no DEBAD – Departamento de Banco de Dados. O DW no IBGE, seria uma evolução do banco de dados estatístico que era utilizado desde a década de 80, o RAPID. O mesmo era um banco de dados que atendia, plenamente, o IBGE porém, por ser um modelo proprietário em plataforma alta que não sofreu alterações de versão, desde o final da década de 80, tornou-se desatualizado em termos de tecnologia. O RAPID possuía uma modelagem muito próxima da modelagem multidimensional, trabalhava com o conceito de matrizes transpostas, e já implementava todo um nível de compressão de dados, para dados ausentes e dados escassos, sabendo trabalhar muito bem uma matriz, espaço e informação, o que facilitou o processo de migração dos dados do RAPID para o DW.

O projeto de DW começou a ser desenvolvido em 1999, com o desenvolvimento de um projeto de avaliação para incorporação da tecnologia de modelos multidimensionais, como evolução do banco de dados estatístico. O RAPID foi utilizado, como paradigma para comparação e incorporação de tecnologia. O projeto foi bastante feliz, o que levou a diretoria a autorizar o prosseguimento para a transformação do projeto de prospecção num projeto de remodelagem da base de dados e construir uma base nos padrões multidimensionais de DW.

A ORACLE Consultoria foi contratada para o desenvolvimento do DW, eles utilizaram a metodologia formal para o desenvolvimento, toda ela baseada em UML.

O DW estava sendo implementado de forma incremental. A primeira pesquisa a ser armazenada, foi uma série agropecuária com cerca de 3 milhões de registros e em torno de 12 variáveis. Outra importante pesquisa armazenada, foi a Pesquisa Nacional por Amostra de Domicílios, possuía em torno de 300 (trezentas) variáveis e vinte e seis milhões de registros. O Censo de 1991, estava sendo carregado exigindo do IBGE grande atenção e esforço na otimização de acesso devido ao seu grande volume, cerca de 182 milhões de registros. O Censo 1991 possuía em torno de (146) cento e quarenta seis milhões de

registros de pessoas e (36) trinta e seis milhões de domicílios. No questionário básico, foram 12 variáveis de pessoas e vinte e poucas variáveis de domicílio e a amostra estava começando a ser carregada. A amostra continha, (180) cento e oitenta variáveis sobre pessoas e sessenta e poucas sobre domicílios, com um volume em torno de 15% dos registros do formulário básico. O modelo já estava definido faltando fazer a extração dos dados das bases antigas, para carregar nas bases novas.

A estratégia utilizada foi carregar o censo de 91 para, em caso de pane, verificar o que precisaria ser feito em termos técnicos. O censo 2000 seria carregado em meados de setembro.

O processo de implementação do DW no IBGE só estaria, completamente, implementado no segundo semestre de 2001, com o início oficial do serviço de difusão de microdados na Internet.

5.3.3.3 Fontes dos Dados do DW

A principal fonte de dados do DW é o Departamento de Pesquisas Estatísticas. O principal ponto de acumulação do acervo está em arquivos planos, armazenados em fitas magnéticas, em ambiente de plataforma alta. A administração dos dados é feita, através de um robô que auxilia a organizar e a disponibilizar todas informações requeridas, controlando o acesso automático às fitas magnéticas.

5.3.3.4 Ambiente do DW

O DW, está residente em plataforma baixa, utilizando dois computadores servidores IBM SP2 e um IBM RISC AIX. A base é controlada no nó de um servidor IBM SP2, com 3 processadores. A máquina principal possui 6 gigabytes e o IBGE já estuda aumentá-la para armazenamento dos dados. No momento desta pesquisa, estava sendo instalado um novo equipamento com 2 Terabytes, um equipamento da classe storage, um IBM-Shark, sendo que, 1 Terabyte será reservado, para o DW.

O IBGE fez a opção por trabalhar com ambiente aberto e compatível, com a quantidade de variáveis que estavam trabalhando, utilizando uma modelagem ROLAP sobre um banco ORACLE 8i.

A ferramenta SAS, também é utilizada para extração e análise dos dados.

Em termos de informação, o DW procura trabalhar com uma granularidade mais fina possível com o objetivo de permitir a maior liberdade possível, na área de análise de informações.

A modelagem multidimensional, para o meio estatístico tem necessariamente, no mínimo três tipos de componentes: o componente espacial, componente temporal e o componente substantivo, além das outras dimensões que se associam a essa informação. No componente espacial, a tendência é obter o menor seguimento espacial, relacionado à informação, no caso do censo demográfico o setor censitário, agrupamento de cerca de 250 domicílios.

5.3.3.5 Acesso aos Dados do DW

Há uma interface Java implementada através de uma em *Applet*, rodando no *browser*, e uma *servlet* rodando no IBM SP2 para fazer o acesso ao banco ORACLE. Na realidade são dois aplicativos Java, um applet controlando o browser do usuário e uma servlet recebendo as solicitações de busca, lendo o metamodelo, estruturando as consultas, agregando a informação, devolvendo para o browser.

O SAS possui uma interface gráfica construída, através do SAS AF. Este permite estruturar aplicações, facilitando desenhar a tela e fazer uma programa, em baixo de seu *Tool Kit* (conjunto de ferramentas). A camada de acesso do SAS AF, lê o metamodelo também utilizado pela Interface Java, permitindo que o usuário selecione as informações, através de listas de informações e no momento que o usuário determine a execução da consulta, então ele traduz o desejo do usuário numa query SQL, a qual se preocupará com o nome que está na base, o join das tabelas, o predicado de seleção. O metamodelo que está associado a esta interface, apresenta para o usuário as hipóteses de manuseio do banco que ele dispõe.

A camada de metainformação, envolve todo projeto de DW, dispensando o usuário saber como a informação foi armazenada. A modelagem física, que está sendo empregada é, totalmente, transparente para o usuário, ele vê a metainformação e através dela é que manuseia a base. Como estas tabelas estão organizadas, se vai ter que expandir ou não a amostra, se vai ter que fazer join, formação, qual a chave de *join* que terá que usar, isso não é problema do usuário. A camada de metainformação obtida, é uma camada de metainformação ativa, porque manuseia dados, através dela. Tanto o acesso Java quanto o SAS, estão customizados dessa forma.

Existe também uma implementação feita com *Oracle Discoverer*, mas ela não tem a importância das implementações em Java e SAS, porque depende de um metamodelo próprio, que não é o metamodelo do IBGE, e apesar de ser uma interface OLAP não é uma interface, perfeitamente, integrada ao projeto de DW.

5.3.3.6 *Dificuldades para implementação de DW*

Numa instituição pública, a implantação de novas tecnologias, sempre, demanda tempo e também que esteja prevista no orçamento. O DW do IBGE foi viabilizado com os recursos alocados ao censo Demográfico, onde está destacado como uma das suas principais novidades tecnológicas. O processo de desenvolvimento do DW é um processo demorado e que no IBGE já se trabalha há dois anos e meio. Para manter um projeto de longo prazo no governo brasileiro, é necessário de um momento atípico de financiamento, e assim, foi o Censo 2000. Graças a esse regime excepcional de orçamento, o DEBAD tem conseguido manter uma equipe de consultoria, trabalhando dentro da empresa, que vem permitindo o desenvolvimento do projeto.

O maior desafio para a implantação do DW no IBGE, foi o volume de dados que precisam ser carregados no mesmo.

A última fase do projeto, só deverá terminar em meados de 2002, quando então deverá se criar, rotinas automatizadas de carga e limpeza de dados no DW.

5.3.4 Banco Central do Brasil - BC

5.3.4.1 Ambiente da Instituição Pública

O Banco Central do Brasil - BC, é uma Autarquia Federal integrante do Sistema Financeiro Nacional – SFN, foi criado em 31.12.64, com a promulgação da Lei nº 4.595, para ser o agente da sociedade brasileira na promoção da estabilidade do poder de compra da moeda brasileira, através dos seguintes objetivos:

- zelar pela adequada liquidez da economia;
- manter as reservas internacionais do País em nível adequado;
- estimular a formação de poupança em níveis adequados às necessidades de investimento do País; e
- zelar pela estabilidade e promover o permanente aperfeiçoamento do Sistema Financeiro Nacional.

A estrutura organizacional do BC é composta pela presidência e uma diretoria colegiada composta por 6 (seis) diretores. Destacamos para esta pesquisa a Diretoria de Administração – DIRAD, a Diretoria de Fiscalização – DIFIS e a Diretoria de Assuntos Internacionais – DIREX.

Na DIRAD, destacamos o Departamento de Informática – DEINF, na DIFIS destacamos o Departamento de Supervisão Indireta – DESIN, na DIREX, destacamos o Departamento de Capitais Estrangeiros e Câmbio – DECEC.

O DEINF é responsável pela gestão de TI. O DESIN é responsável pelas áreas de supervisão indireta do mercado financeiro. O DECEC é responsável pelas áreas de câmbio, dívida externa, e comércio exterior.

5.3.4.2 Maturidade

O Banco Central, no exercício de suas atribuições, realiza frequentemente atividades de monitoração e controle sobre o sistema financeiro, política econômica,

câmbio, reservas internacionais, dívida externa e comércio exterior. Essas atividades são gerenciadas pelos analistas do BC, que precisam de informações com correção e tempestividade. Diante disso, houve a grande necessidade de implantação de um ambiente de DW no banco, para que as informações estivessem orientadas ao negócio, servindo como subsídio para a tomada de decisão.

No intuito de atender as demandas citadas, em 1998 foram iniciados estudos preliminares para a implementação do ambiente de DW. Estes estudos tiveram o apoio de uma consultoria para a definição das áreas de negócio a serem consideradas para a implantação. A partir desses estudos, foi decidido atender inicialmente a área de câmbio.

No momento desta pesquisa, já haviam sido implantados 3 (três) projetos na área de câmbio, são eles: contratação de câmbio, eventos de câmbio e movimentação diária. No entanto, já estavam em andamento projetos para as áreas de dívida externa, supervisão do mercado financeiro e comércio exterior.

Os gestores dos projetos citados, são os mesmos dos sistemas transacionais correspondentes.

A Instituição sempre envia representantes para seminários, congressos e treinamentos específicos de tecnologias de DW. Todos os técnicos e usuários envolvidos possuem treinamento.

5.3.4.3 Fontes dos Dados do DW

Os projetos implantados correspondem aos sistemas transacionais da mesma área. Os *Data Marts* “contratação de câmbio”, “eventos de câmbio” e “movimentação diária”, armazenam dados derivados dos sistemas operacionais de contratação de câmbio, de eventos de câmbio e da movimentação diária do câmbio, respectivamente.

No momento desta pesquisa, basicamente, todos os dados operacionais estavam armazenados no SGBD ADABAS em plataforma alta. As únicas fontes de dados externas são algumas informações de pessoas jurídicas obtidas da base de dados da Receita Federal.

5.3.4.4 Ambiente do DW

A metodologia de desenvolvimento de um projeto de DW no BC possui as seguintes fases:

- definição do escopo do projeto do *Data Mart* a ser construído através de reuniões iniciais com os usuários. Nessas reuniões é utilizada a técnica formal de levantamento de dados, denominada *Joint Application Design - JAD*;
- levantamento das necessidades, através de reuniões utilizando a técnica JAD, onde são discutidas as necessidades de informações, além dos meios a serem utilizados para adquiri-las;
- modelagem multidimensional, a ser validada pelos usuários;
- projeto de ETC, onde, são feitas as especificações dos programas de carga, sendo estes repassados à área de desenvolvimento pertinente;
- em paralelo ao projeto ETC ocorre o projeto físico do *Data Mart* no DW corporativo com a construção do ambiente de banco de dados do DW no SGBD;
- em paralelo às duas últimas fases citadas ocorre também o treinamento dos usuários na ferramenta de acesso e análise dos dados;
- finalizado a implementação do projeto ETC, o mesmo é submetido ao processo de teste, onde são realizados os acertos necessários;
- após a carga dos dados, o *Data Mart* é avaliado e homologado pelos usuários finais;
- treinamento e implantação.

A infra-estrutura do DW é em plataforma baixa, ou seja, em arquitetura Cliente x Servidor. Os dados do DW estão armazenados em um *SGBD DB2/UDB* instalado em um computador servidor *Sun Enterprise 4000*, composto de 8 (oito) processadores paralelos de 250 Megahertz cada, 2 Gigabytes de memória principal (*Radom Aleatory Memory* –

RAM), 2 (dois) discos internos com 2,1 Gigabytes cada, e 4 (quatro) discos externos com 176,4 Gigabytes cada. Neste servidor o sistema operacional é o *SUN SOLARIS 2.6*.

O nível de granularidade é mais atômico do que o desejado pelos técnicos, porém ainda é a solução para atender determinadas necessidades dos usuários. No momento desta pesquisa, os técnicos discutiam sobre o desenvolvimento de um ambiente de ODS.

Os processos de extração e transformação dos dados operacionais são feitos ainda em plataforma alta no ambiente operacional, através de programas escritos em Natural. O processo de validação e carga dos dados no DW é feito, através de ferramentas do SGBD DB2 como *SQL Storage Procedures* e regras de integridade. Esses processos que compõem o processo de ETC, são agendados e ativados automaticamente.

A modelagem dos dados é a multidimensional na abordagem *snowflake*, que era a mais indicada para o uso da ferramenta OLAP, denominada *MicroStrategy*.

No momento desta pesquisa, nenhuma ferramenta específica para gerenciamento de metadados, era utilizada. O dicionário de dados foi construído em cima da ferramenta CASE *System Architect*. Também eram utilizados recursos da ferramenta OLAP, que oferece campos destinados à descrição do atributo ou do fato.

5.3.4.5 Acesso aos Dados do DW

Para acessar os dados do DW e realizar suas análises, os usuários finais utilizam a ferramenta OLAP *MicroStrategy 7* da MICROSTRATEGY BUSINESS INTELLIGENCE. Tal solução, permite acessos aos dados, tanto na interface cliente/servidor, quanto na WEB (Internet e Intranet). A dificuldade maior dos usuários é quanto ao entendimento de alguns conceitos característicos de uma ferramenta OLAP, como: drill, métrica, fato, lookup, transformação, aditividade, etc. O tempo de resposta é aceitável, porém os técnicos acreditam que pode ser melhorado, através de incrementos no banco de dados tais como: novos índices, particionamentos e reorganização física, em vista dos discos e controladoras disponíveis, etc.

Os resumos e agregações são disponibilizados para acesso pelos usuários finais, no mesmo ambiente do DW corporativo, podendo o usuário acessar até mesmo os dados mais atômicos, obedecendo à abrangência da área de negócio.

5.3.4.6 Dificuldades para implementação de DW

As tecnologias de DW ainda não estão disseminadas em todas as áreas que, potencialmente, poderiam utilizá-las. Assim, um dos grandes desafios dos técnicos é mostrar, “vender” a importância dessas tecnologias para diversas áreas do banco, pois a cultura de SAD ainda é uma abordagem relativamente nova, enquanto é muito forte a cultura de aplicações OLTP em plataforma alta. Outra grande dificuldade é a carência de pessoal técnico para diminuir o tempo de implementação dos projetos. Além disso, a instituição enfrenta as conhecidas restrições orçamentárias do setor público.

5.3.5 Empresa Brasileira de Pesquisa Agropecuária - Embrapa

5.3.5.1 Ambiente da Instituição Pública

A Empresa Brasileira de Pesquisa Agropecuária – Embrapa, é uma empresa pública, vinculada ao Ministério da Agricultura e do Abastecimento, dotada de personalidade jurídica de direito privado, com patrimônio próprio, autonomia administrativa e financeira.

A missão da Embrapa é viabilizar soluções para o desenvolvimento sustentável do agronegócio brasileiro por meio de geração, adaptação e transferência de conhecimentos e tecnologias, em benefício da sociedade.

A Embrapa atua por intermédio de 40 unidades descentralizadas e 15 unidades centrais. O nível estratégico da instituição é composto pelo conselho de administração, diretoria executiva e o diretor presidente. As unidades centrais assessoram o nível estratégico da Embrapa, entre elas destacamos para a esta pesquisa, a Secretaria de Administração Estratégica – SEA e o Departamento de Tecnologia de Informação – DTI (antigo DIN).

A SEA é a unidade central responsável pelos processos de planejamento e gestão da estratégia organizacional da Embrapa. A DTI é a unidade responsável em subsidiar a Diretoria Executiva na formalização da política de informação e informática, e manter os instrumentos destinados a sua consecução, bem como coordenar e executar a política estabelecida.

No momento desta pesquisa, a Embrapa estava passando por uma grande mudança na sua estrutura organizacional, sendo introduzido um modelo de gestão estratégica por processos, com base na tecnologia de *Balanced Scorecard* - BSC, que visa exatamente, especificar e mapear os processos organizacionais da empresa, bem como definir metas e os objetivos estratégicos.

5.3.5.2 Maturidade

No momento desta pesquisa, não havia implementação efetiva de tecnologias de DW na Embrapa. Em 1999 a SEA iniciou um projeto que contava com a participação da DTI, com o objetivo de identificar as necessidades prioritárias da Embrapa, bem como as tecnologias de DW mais adequadas para o atendimento dessas necessidades e nesse projeto foram desenvolvidos protótipos com alguns fornecedores de tecnologia DW. O Quadro 24 mostra o resultado da avaliação desses protótipos, todavia, a Embrapa estava fechando a negociação de um grande contrato com a ORACLE que envolve tecnologias de: SGBD, WEB, Java, Corba, portal, workflow, CRM e ferramentas para DW. Essa negociação paralisou o processo de aquisição de uma das tecnologias aprovadas, em função da possibilidade de utilização das ferramentas de DW da ORACLE.

A SEA, desempenhando o seu papel, vem trabalhando para conscientizar o nível estratégico sobre a importância do uso de ferramentas que possibilitem o acesso fácil e analítico das informações estratégicas. Além desses protótipos feitos por fornecedores de tecnologias de DW, foi desenvolvido um protótipo com a tecnologia de *Balanced Scorecard* da Oracle, com as informações do Modelo de Gestão Estratégico da Embrapa – MGE e do Sistema de Avaliação das Unidades – SAU.

Quadro 15 - Avaliação dos Protótipos de DW Desenvolvidos pelas Empresas Fornecedoras

Fornecedor	Ferramenta	Data da Apresentação	Avaliação
Oracle	Discovery	07-05-1999	AR
SAS	SAS/Warehouse	05-05-1999	AR
Business Intelligence Company - Choice	O3	06-05-1999	RE
Informix	Metacube	13-05-1999	AP
Sagent	Sagent Solution	12-08-1999	AP

AP = Aprovado

AR = Aprovado com restrições

RE = Reprovado.

As tecnologias de DW ainda não constituem prioridade para a DTI, que enfrenta outros desafios, considerados mais emergentes, como soluções Web para intranet/extranet, segurança, integração de sistemas corporativos e CRM. Os técnicos sabem que com a implementação de CRM, abre-se um caminho para a implementação de DW. Tecnologias de ERP não são descartadas, mas também não estão na lista de prioridades.

A Embrapa investe, constantemente, em treinamento de pós-graduação, através de um programa oficial, onde alguns técnicos de TI realizam trabalhos acadêmicos que possibilitem a aquisição de conhecimentos sobre DW. Os técnicos com a função de administração de banco de dados, eventualmente participam de congressos e seminários sobre bancos de dados, e atualizam seus conhecimentos sobre as tecnologias de DW.

5.3.5.3 Fontes dos Dados do DW

Apesar da Embrapa não ter implementações efetivas de DW, este estudo levantou algumas outras fontes de dados, além do MGE e do SAU, potenciais para povoar o DW da instituição.

A Embrapa é uma empresa de pesquisa, onde os objetivos e recursos de qualquer atividade, desde administração até a pesquisa agropecuária, são definidos através da figura de projeto. O Sistema Embrapa de Planejamento – SEP, visa fazer todo o gerenciamento

do planejamento até a aprovação do projeto, bem como a execução, através do acompanhamento até a fase de avaliação. O Sistema de Informação Gerencial da Embrapa – SIGER, é o sistema OLTP que automatiza o SEP, os dados são armazenados em arquivos Paradox isolados nas unidades da Embrapa, dificultando a integração já que a figura do projeto está presente em praticamente todos os outros sistemas de informação da instituição. O SIGER está sendo revisto para utilizar tecnologias que possibilitem, a melhor integração com os sistemas corporativos da instituição.

Todos os empregados são obrigados a realizar o planejamento de suas atividades a cada ano de acordo com os projetos. Durante o ano essas atividades são acompanhadas e avaliadas, em consequência disso, o empregado é avaliado para fins de promoção e premiação. O sistema OLTP que automatiza esse processo é o Sistema de Planejamento, Acompanhamento e Avaliação do Trabalho Individual – SAAD. O empregado também é avaliado no trabalho em equipe. Esse sistema tem os dados armazenados em cada unidade da Embrapa em SGBD Ingres. Os dados de todas as unidades são integrados na sede da instituição através de processos *in-batch* internos da aplicação SAAD.

O Sistema de Informações de Recursos Humanos – SIRH é o sistema OLTP que automatiza todo processo de recursos humanos da instituição. Esse OLTP possui informações sobre os empregados desde a origem da instituição.

Existem informações históricas das pesquisas que são isoladamente armazenadas em microcomputadores, através de ferramentas de estatística como o SAS, de planilhas eletrônicas como o MS-Excel e de editores de texto como o MS-Word.

A instituição pública denominada “Instituto Brasileiro de Geografia Estatística – IBGE” disponibiliza dados sócio-econômicos para sociedade. Na SEA parte desses dados referentes a pesquisa agropecuária, são tratados e organizados, formando a base de dados AGROTEC que é disponibilizada na Web/intranet da instituição através de consultas SQL. O objetivo é que essa base seja utilizada como fonte para pesquisas realizadas pelo corpo de pesquisadores da Embrapa.

5.3.5.4 Ambiente do DW

No momento desta pesquisa, o ambiente do DW na Embrapa ainda não estava definido. Estudos estavam sendo realizados para a definição da estratégia de implementação do DW, através de um plano de arquitetura e infra-estrutura, inicialmente com base em tecnologias ORACLE. Essa definição será validada e aperfeiçoada, através da implementação de um projeto piloto sobre os dados da AGROTEC.

5.3.5.5 Acesso aos Dados do DW

Semelhante a variável anterior, as formas e ferramentas de acesso aos dados ainda não estavam definidas, entretanto na implementação do projeto piloto, inicialmente será utilizado a ferramenta *Oracle Discoverer*.

5.3.5.6 Dificuldades para implementação de DW

As principais dificuldades relatadas sobre implementação de DW na Embrapa foram as seguintes:

- A quantidade de profissionais de TI na instituição é bastante reduzida. As contratações de profissionais são feitas somente através de concursos públicos. Poucos concursos foram autorizados nos últimos 10 anos. A instituição é obrigada a ter um limite fixo de empregados, e quando as vagas surgem, em muitos casos, a prioridade não é para TI, até porque TI não é a atividade fim da instituição;
- A instituição pode terceirizar as atividades de TI, mas é preciso ter os processos bem definidos, bem como a alocação dos escassos profissionais de TI no acompanhamento desses processos;
- Devido ao número reduzido de profissionais de TI, existe sobrecarga de trabalhos, principalmente naqueles profissionais mais capacitados. Esses profissionais sobrecarregados ficam sem tempo para a fundamental atualização tecnológica de seus conhecimentos;

- Apesar da Embrapa ter orçamento aprovado para os seus projetos ainda depende de liberação dos recursos financeiros, através do Ministério da Agricultura, que muitas vezes não é feita dentro dos prazos previstos;
- A estrutura organizacional da Embrapa é constituída de várias unidades descentralizadas dispersas no território nacional, soluções de TI nessas unidades, foram alcançadas sem a preocupação de integração com o todo da instituição, gerando uma cultura de utilização de várias tecnologias para as mesmas soluções. Hoje a instituição possui alguns sistemas corporativos com grandes dificuldades de integração. No início de 2001 foi criado um grupo permanente de trabalho dentro da instituição com membros dos gestores dos processos e profissionais de TI, a fim de encontrar soluções para integração das informações dentro da instituição. Esse grupo tem encontrado dificuldades na disponibilidade dos membros para a execução das atividades do grupo;
- A burocracia do processo de licitação, além da pouca agilidade, muitas vezes engessa a evolução da TI na Embrapa. As dificuldades com relação a recursos humanos de TI dentro da instituição são evidentes.

5.3.6 Agência Nacional de Energia Elétrica - ANEEL

5.3.6.1 Ambiente da Instituição Pública

A Agência Nacional de Energia Elétrica - ANEEL, é uma autarquia sob regime especial, vinculada ao Ministério de Minas e Energia, com personalidade jurídica de direito público e autonomia patrimonial, administrativa e financeira.

A missão da ANEEL é proporcionar condições favoráveis para que o mercado de energia elétrica se desenvolva com equilíbrio entre os agentes e em benefício da sociedade.

A estrutura organizacional da ANEEL não é hierárquica funcional, ela é de gestão de processos, onde cada área de atuação é um processo. A estrutura é composta por uma diretoria colegiada composta por 5 (cinco) diretores com a mesma função, 24 processos definidos. Dos processos definidos, destacamos para essa pesquisa o processo de

Planejamento e Gestão Administrativa e desse processo a Superintendência de Gestão Técnica da Informação – SGI.

A SGI tem por finalidade executar as atividades relacionadas ao processo de gestão da informação, no âmbito interno e externo da Autarquia.

5.3.6.2 *Maturidade*

A SGI trabalha, principalmente, com dados agregados estratégicos sobre a regulação e fiscalização do setor de energia elétrica. Esses dados são obtidos, principalmente das diversas concessionárias de serviços públicos de energia elétrica.

A idéia de DW surgiu com a necessidade de integrar e agregar os dados, bem como o metadado desses dados. Os assessores da diretoria e as áreas específicas como a fiscalização, precisavam ter esses dados integrados e acessíveis para análises e avaliações mais criteriosas.

No início de 1999, começaram os estudos sobre a utilização de tecnologias de DW. Através de um contrato com a TBA, que é uma empresa representante dos produtos da Microsoft em Brasília, começou uma fase experimental com o desenvolvimento de 2 (dois) protótipos, tendo como objetivo capacitar os técnicos e sensibilizar os clientes da SGI. Foram feitos cursos com a finalidade de se obter o objetivo dos protótipos.

Os protótipos desenvolvidos foram sobre os indicadores de qualidade de energia elétrica das concessionárias e sobre as informações contábeis das concessionárias. Esses protótipos foram baseados em informações reais, demandadas pelos clientes da SGI. Os resultados esperados foram atingidos.

Os principais usuários finais das informações são os assessores da diretoria e os superintendentes. Esses usuários não estão disponíveis para a análise dos dados, eles estão interessados na informação final. Torna-se necessário ter pessoal disponível para a implementação dessas análises. A percepção desses usuários é usar ferramentas simples de visualização, onde as visões já estão prontas.

A ANEEL está disponibilizando uma central de ouvidoria (*call center*) onde o cliente, em última instância, pode recorrer à Instituição. Se o cliente recorrer e não for

última instância, a ANEEL direciona o chamado para a devida empresa concessionária. As empresas concessionárias passam a ser avaliadas por parâmetros de atendimentos e reclamações que a ANEEL controla, através da central de ouvidoria. A expectativa é que isso seja possível com a utilização tecnologias de CRM e DW.

A Instituição sempre envia representantes para seminários, congressos e treinamentos específicos de tecnologias de DW.

Foi formado um grupo de técnicos da área de administração de bancos de dados, para a pesquisa e desenvolvimento das tecnologias de DW. Existe o interesse da SGI em ter inicialmente um técnico exclusivo para o atendimento dos usuários finais quanto à análise dos dados. A idéia é que a equipe aumente ou que se terceirize tanto o desenvolvimento quanto à análise dos dados.

5.3.6.3 Fontes dos Dados do DW

As principais fontes são as informações agregadas exigidas das empresas concessionárias e enviadas para a ANEEL, geralmente na periodicidade mensal. O grande problema é como centralizar e padronizar as informações de 64 empresas do setor elétrico com tecnologias de TI diferentes. Hoje a SGI define a padronização dos dados que devem ser enviados para a ANEEL, sempre de forma agregada. Foi criada uma *extranet* para o setor elétrico que passou a ser o caminho de tráfego dessas informações. A SGI tem planos para que essa *extranet* seja um portal de DW, um *Data Webhouse*, possibilitando uma visão integrada de todo mercado de energia elétrica.

Outra fonte dos dados é a central de ouvidoria, onde os dados irão alimentar um *Data Mart* a ser implementado. A SGI, numa atitude pró-ativa, condicionou essa implementação como pré-requisito do CRM que está sendo implantado para atender a central de ouvidoria.

Além dos dados das concessionárias é necessário buscar outros dados de todos os agentes que estão envolvidos em energia elétrica espalhados no Brasil. A principal fonte desses dados tem sido o Instituto de Brasileiro de Geografia Estatística - IBGE.

5.3.6.4 Ambiente do DW

Os protótipos foram desenvolvidos com base em infra-estrutura e ferramentas disponíveis no pacote do SGBD *Microsoft SQL Server 7.0*. Além do SGBD esse pacote contém a ferramenta *Microsoft Data Transformation Services (DTS)* para os processos de extração, de transformação e carga. O pacote também contém a ferramenta *Microsoft SQL OLAP*.

No momento desta pesquisa a SGI estava trabalhando no projeto de arquitetura e infra-estrutura, pesquisando tecnologias de DW que pudessem ser agregadas ao ambiente Microsoft.

O desenvolvimento dos protótipos foi em *Data Marts* isolados, baseado na metodologia proposta por Kimball (1998a) do ciclo de desenvolvimento do DW. Nesse desenvolvimento não foi implementado metadados para o DW.

Foi destinado para o DW um servidor que era de contingência, especificamente para DW, não foi comprado nada até hoje. A SGI está com as metas para esse ano de aquisição de software e hardware necessários para a implementação efetiva de DW. Foi orçado em torno de 500.000 reais para ser investido na infra-estrutura do DW.

5.3.6.5 Acesso aos Dados do DW

Acreditava-se que uso de uma ferramenta OLAP iria resolver o problema de acesso porém, o uso da mesma ocupava um tempo excessivo do analista, para importar os modelos para o universo OLAP, além da não adaptação dos usuários finais, que mesmo visualmente tinham que fazer análises que não estavam acostumados a fazer.

Foi proposta uma solução pela Microsoft que realmente funcionou, a utilização da planilha eletrônica Excel, através de recursos de tabelas e gráficos dinâmicos conectados a fontes de dados do *Microsoft SQL Server OLAP*. Essa solução aumentou a necessidade de profissionais que, partindo de um demanda, execute a tarefa de análise dos dados e disponibilize o resultado através da ferramenta Excel. No momento desta pesquisa essa tarefa estava sendo executada pelos administradores de dados.

5.3.6.6 *Dificuldades para implementação de DW*

A ANELL é uma agência executiva, e por isso já tem uma nova visão de instituição pública, que tem os problemas burocráticos por ser pública, porém com maior capacidade de auto-gestão, possibilitando a característica de ser muito rápida em investimentos e compras. Dentro da limitação orçamentária a Instituição pode realizar contratos com terceiros, como é caso do contrato especial com a Microsoft. Hoje o ambiente de software é quase todo Microsoft.

A SGI é uma superintendência ligada a tecnologia, mas com a preocupação da melhor utilização dos recursos, portanto se tem a preocupação em relação à implementação de tecnologias de DW, em função da necessidade de investimentos. Existem pequenas resistências por parte dos demais profissionais de TI, principalmente dos técnicos de desenvolvimento que consideram que as demandas de informações estratégicas mais agregadas poderiam ser atendidas com o desenvolvimento por eles sem a implementação de tecnologias de DW.

A maioria dos profissionais de TI na Instituição é terceirizada, e se tem um problema da terceirização que é a entrega do profissional já capacitado. Os profissionais terceirizados, envolvidos com uma nova tecnologia, terão que ser capacitados, além da dificuldade de encontrar profissionais capacitados no mercado, não existe a cultura de treinamento por parte da terceirizada.

O usuário final quer a informação já estruturada e disponibilizada de forma fácil e visual, não tem disponibilidade para montar a análise dos dados. Existe a necessidade de profissionais analistas de apoio à tomada de decisão, que conheçam o negócio, e que fiquem dedicados a montar essas análises. Hoje essas tarefas são executadas pela própria SGI, através dos administradores de banco de dados.

A quantidade de profissionais envolvidos com as tecnologias de DW ainda é pequena, além desses profissionais terem outras atribuições de administração de dados.

5.3.7 Tribunal Superior Eleitoral - TSE

5.3.7.1 Ambiente da Instituição Pública

O Tribunal Superior Eleitoral - TSE, é um órgão da Justiça Eleitoral, pertencente ao Poder Judiciário.

O nível estratégico do TSE é formado por uma Presidência, Ministros, Corregedoria Geral Eleitoral e a uma Diretoria Geral.

Destacamos para essa pesquisa a Seção de Estatística Eleitoral – SEE, a Seção de Alistamento Eleitoral – SAE e a Seção de Administração de Banco de Dados – SABD, ambas fazem parte da Secretaria de Informática.

A SEE tem como missão, prover informações (dados estatísticos) acerca da distribuição e perfil do eleitorado nacional e resultado de eleições, através da extração e tratamento dos dados do Cadastro Nacional de Eleitores - CNE e dos Bancos de Dados de Eleições. Essas informações são disponibilizadas através de bancos de dados específicos, sistemas, relatórios e internet para setores internos do TSE, TREs e público externo em geral, visando, entre outros objetivos, auxiliar a tomada de decisão e estudos no sentido de nortear ações de:

- No âmbito da Justiça Eleitoral:
- Revisões eleitorais;
- Correções de distorções;
- Definição e mensuração dos recursos computacionais;
- Outros.

No âmbito do público externo:

- Definições de estratégias políticas;
- Auxílio em estudos acadêmicos;

- Divulgação pela mídia e órgãos oficiais de pesquisa;
- Outros.

5.3.7.2 *Maturidade*

No momento desta pesquisa, havia um protótipo implantado. Esse protótipo, tem o objetivo de introduzir o conceito de DW na organização, através da implementação de um *Data Mart*, tendo como área de negócio (assunto) o Cadastro Nacional de Eleitores - CNE.

Os usuários finais desse protótipo, são o estratégico do TSE, os Secretários, Coordenadores, Assessores e Técnicos em geral. As expectativas desses usuários foram todas atingidas, e estão descritas a seguir:

- Consulta a dados históricos e suporte na análise da evolução do eleitorado;
- Suporte na estimativa de custos para com o eleitorado;
- Suporte na análise de expansão do uso de urnas eletrônicas;
- Suporte na verificação da necessidade de revisão do eleitorado; e
- Possibilidade de disponibilizar informações diretamente na Intranet e Internet.

Devido o protótipo ter sido desenvolvido em parceria com uma empresa terceirizada, até o momento desta pesquisa, a Instituição ainda não havia enviado funcionários para participações em seminários, congressos, ou treinamentos específicos, sobre a implementação e utilização de tecnologias de DW.

5.3.7.3 *Fontes dos Dados do DW*

A principal fonte dos dados para o protótipo é sistema OLTP Cadastro Nacional de Eleitores – CNE. Os dados desse sistema estão armazenados através de SGBD ORACLE.

A principais fontes externas do protótipo, são os dados obtidos pelo IBGE sobre a população brasileira, no nível de município.

5.3.7.4 Ambiente do DW

Para o projeto de *Data Mart* do CNE, foram adquiridos dois servidores, cada um com 2 (dois) processadores Pentium III Xeon 500, 512 megabytes de Memória RAM, 18.2 gigabytes de espaço em disco e 1 (um) megabyte de cachê. Foi também adquirido o pacote de software *Oracle Data Mart Suit* da ORACLE, para criação, montagem, extração e visualização das informações. Esse pacote contém as seguintes ferramentas:

- Oracle *Data Mart* Designer;
- Oracle *Data Mart* Builder;
- Oracle 8 Enterprise Server;
- Oracle Application Server;
- Oracle Discoverer; e
- Oracle Discoverer Viewer.

A metodologia utilizada para realização das etapas do projeto é a metodologia CTM/IS - GER, baseada no método JAD (*Joint Application Design*) da IBM Canadense. Sua dinâmica dá ênfase à participação direta de pessoas que, efetivamente operam o dia-a-dia das áreas e processos que estão em foco, mediante a realização de reuniões de trabalho. O Quadro 25, mostra o plano de trabalho elaborado para o desenvolvimento do *Data Mart* do CNE. No momento desta pesquisa o desenvolvimento se encontrava na última etapa, o *Data Mart* implantado e sendo monitorado no ambiente de produção.

Para efetuar a extração dos dados de origem (cadastro nacional de eleitores e outras fontes), o tratamento, a agregação e conseqüente carga destes dados no *Data Mart*, foram construídos planos de carga, com base nos metadados definidos com a utilização da ferramenta Oracle *Data Mart* Builder, bem como, a criação de funções no SGBD ORACLE para determinadas transformações que retornam resultados já prontos do OPTP, para serem utilizados pelos planos de carga, melhorando assim, o desempenho de execução dos mesmos.

A modelagem dos dados adotada, foi a multidimensional com modelos estrela.

Quadro 16 - Plano de Trabalho para o desenvolvimento do *Data Mart* do CNE.

Etapa	Fases
Etapa 1 – Planejamento e Entendimento dos Trabalhos.	Fase 1 – Planejamento inicial do projeto
Etapa 2 – Levantamento dos Requisitos Iniciais.	Fase 1 – Identificação da área de negócios; Fase 2 – Definição e análise dos questionamentos gerenciais; Fase 3 – Identificação e entendimento das origens dos dados.
Etapa 3 – Especificação do <i>Data Mart</i>.	Fase 1 – Configuração do ambiente de desenvolvimento; Fase 2 – Modelagem multidimensional de dados; Fase 3 – Definição de metadados e elaboração dos planos de carga.
Etapa 4 – Construção do <i>Data Mart</i>.	Fase 1 – Criação e otimização do banco de dados do DM; Fase 2 – Carga inicial e incremental de dados; Fase 3 – Criação do ambiente de usuários finais.
Etapa 5 – Teste, Implantação e Monitoramento do DM no Ambiente de Produção.	Fase 1 – Teste do DM pelos usuários finais; Fase 2 – Implantação do DM no ambiente de produção; Fase 3 – Monitoramento da utilização do DM.

5.3.7.5 Acesso aos Dados do DW

As ferramentas de acessos aos dados do *Data Mart* são o *Oracle Discoverer* e o *Oracle Discoverer Viewer* para acesso via Web, ambas da ORACLE. Essas ferramentas são consideradas de fácil manuseio técnico, porém, para os usuários finais é necessário treinamento. Em computadores com poucos recursos, o tempo de resposta nos acessos via Web, tem sido elevado.

5.3.7.6 Dificuldades para implementação de DW

As principais dificuldades relatadas foram as seguintes: Rotatividade de funcionários públicos, execução da implementação próximo ao período eleitoral e falta de disponibilidade para curso das ferramentas utilizadas.

5.4 Considerações Finais

Neste capítulo foram apresentados os resultados desta pesquisa, que foi executada conforme a metodologia descrita no capítulo anterior. Estes resultados fundamentam as conclusões e recomendações, que serão apresentadas no capítulo seguinte.

6 CONCLUSÕES E TRABALHOS FUTUROS

6.1 Conclusões

Esta pesquisa teve como objetivo a análise da implementação e utilização de tecnologias de *Data Warehouse* em Instituições Públicas no Brasil, identificando as dificuldades e problemas envolvidos. Este objetivo foi desafiador, considerando que, para análise e identificação das dificuldades fez-se necessário o estudo exploratório e descritivo nas Instituições, portanto, abrangendo diversos aspectos sobre implementação e utilização de tecnologias de *Data Warehouse*.

Ficou evidente a importância da implementação e utilização de tecnologias de DW em Instituições Públicas, independente da área de atuação. As Instituições pesquisadas atuam em diversas áreas, tais como: pesquisa agropecuária; mercado de energia elétrica; processamento de dados e tecnologias da informação; estatística eleitoral; controle da moeda brasileira; serviços e intermediações financeiras; e informações que retratam o Brasil.

Foram implementadas ou se pretende implementar, tecnologias de DW em diversos negócios distintos e específicos da área de atuação de cada instituição. A implementação e utilização dessas tecnologias nas Instituições pesquisadas, não foram direcionadas para negócios afins.

Em todas as instituições pesquisadas, a iniciativa da implementação de tecnologias de DW, não partiu dos dirigentes mas, dos profissionais envolvidos com os processos de atendimento de informações de apoio às tomadas de decisão. Diante da crescente demanda dessas informações e da ineficiência e até ineficácia dos sistemas transacionais, esses profissionais passaram a procurar alternativas para o atendimento, tornando-se líderes na implementação e utilização de tecnologias de DW.

Os altos custos de projeto e as limitações de recursos foram algumas das dificuldades, encontradas por estes profissionais junto aos dirigentes e os demais profissionais de TI.

Em todas as instituições pesquisadas, foram relatadas dificuldades oriundas de limitações de recursos orçamentários, de falta de pessoal qualificado, de treinamento e da burocracia do processo de licitação.

Na maioria das instituições pesquisadas, a atividade fim não é TI, dificultando bastante a alocação de recursos para a implementação de tecnologias de DW. A resistência à tecnologia de DW pelos profissionais de TI, envolvidos com os sistemas transacionais, bem como, a pouca integração e a diversidade tecnológica desses sistemas, são também dificuldades comuns, na maioria dessas instituições.

As dificuldades levaram esses líderes a implementar soluções adequadas aos recursos disponíveis ou tangíveis. A seguir, são apresentadas de forma genérica as conclusões obtidas sobre essas soluções.

Com relação à estratégia arquitetônica de implementação, as instituições que adotaram as estratégias “top-down” e “intermediária”, foram principalmente, devido à quantidade reduzida de profissionais qualificados, bem como, a integração e o monitoramento dos dados em um ambiente de DW centralizado. As instituições que adotaram a estratégia “bottom-up”, foram principalmente, devido à necessidade de convencimento dos benefícios do uso da tecnologia, através de implementação de projetos pilotos.

Quanto à metodologia de desenvolvimento, a maioria das instituições preferiu adotar metodologias adaptadas, em alguns casos da metodologia proposta por Kimball (1998a) do ciclo de desenvolvimento do DW, em outros casos do método JAD.

A maioria das instituições, ainda não possui uma estrutura de metadados que envolva todo o ambiente de DW. As instituições que implementaram metadados, fizeram através de soluções próprias com recursos de HTML, ferramentas OLAP, ferramentas CASE e no caso específico da instituição IBGE foi desenvolvido um metamodelo.

A maioria das instituições, possui as fontes de dados transacionais em plataforma alta e o ambiente de DW em plataforma baixa. No caso específico da instituição Banco do Brasil, o DW corporativo é em plataforma alta sendo que, os *Data Marts* derivados do DW corporativo, estão em plataforma baixa em um mesmo ambiente. Essas instituições utilizam soluções próprias de ETC com recursos de ferramentas nativas de SGBD,

linguagens de programação, e ferramentas específicas para migração de dados entre plataformas alta e baixa. As demais instituições utilizam soluções de ETC disponíveis no pacote do SGBD adotado.

A modelagem de dados multidimensional, foi adotada por todas as instituições. No caso específico do Banco do Brasil o DW corporativo é modelado, através do modelo de Entidade x Relacionamento.

Em relação ao ambiente de DW em plataforma baixa, a maioria das instituições utiliza sistema operacional baseado no padrão UNIX. As demais instituições utilizam o sistema operacional Windows NT. Nesses ambientes os dados são armazenados em SGDBs relacionais.

A maioria das instituições, adotou ferramentas OLAP para o acesso e análises dos dados do DW e a conseqüente informação de apoio à tomada de decisão. A instituição IBGE desenvolveu interfaces em Java e SAS AF, associadas a um metamodelo próprio.

É importante salientar, que esta pesquisa também demonstrou que os processos de *Data Warehousing*, nas Instituições Públicas pesquisadas, estão em constante evolução que, apesar das dificuldades e problemas relatados, a evolução e a disseminação das tecnologias de DW, tornam esses processos cada vez mais efetivos.

6.2 Trabalhos Futuros

Em continuidade ao trabalho desta pesquisa, recomenda-se o desenvolvimento de pesquisa quantitativa com o objetivo de identificar áreas potenciais de implementação e utilização de tecnologias de DW no serviço público, assim como os benefícios e requisitos necessários. Essa pesquisa recomendada deverá trabalhar com os vários aspectos relatados nos resultados da presente pesquisa.

Outra recomendação, é o estudo aprofundado sobre a implementação de tecnologias de *Enterprise Relationship Management* – ERM, em Instituições Públicas. Esse estudo além rever e atualizar os conceitos sobre tecnologias de DW que foram referenciados na presente pesquisa, deverá também referenciar os conceitos de outras tecnologias que,

juntamente com as tecnologias de DW, compõem o conceito de ERM, veja a Figura 20 com a visão geral da arquitetura ERM.

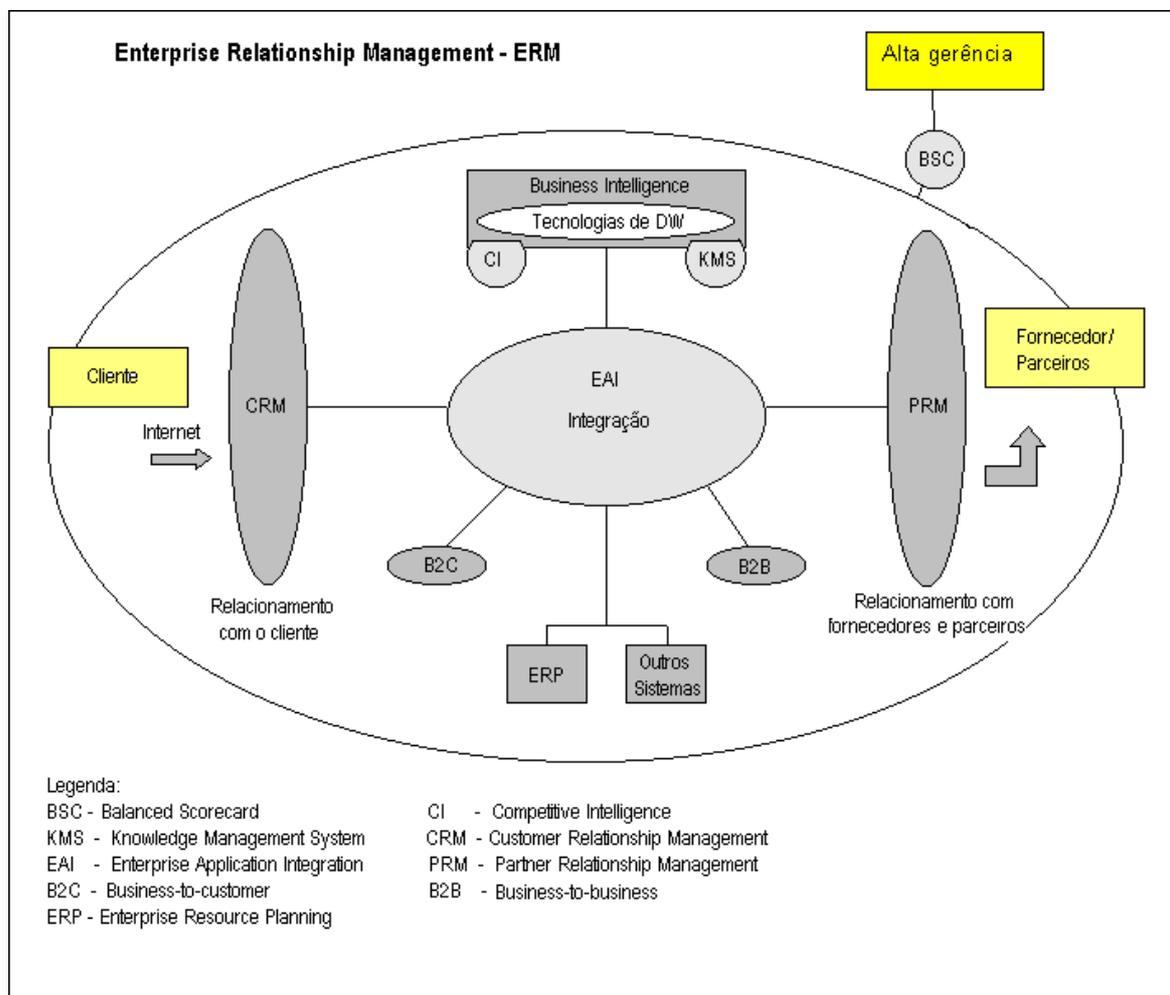


Figura 20 - Visão geral da arquitetura ERM.
Fonte: Adaptado de BARBIERI (2001).

BIBLIOGRAFIA

1. ADELMAN, S.; LEBARON, M. *Meta Data Standards*. Review Magazine. Dezembro, 1997.
2. AMARAL, Antônio Carlos Cintra. **Concessão de Serviço Público**. São Paulo: Malheiros Editores, 1996.
3. ANSOFF, Igor. **A Nova Estratégia Empresarial**. São Paulo : Atlas, 1990b.
4. ANSOFF, Igor. **Administração Estratégica**. São Paulo : Atlas, 1990a.
5. BARBIERI, Carlos, **BI – Business Intelligence: Modelagem e Tecnologia**. Rio de Janeiro: Axcel Books do Brasil, 2001.
6. BARQUINI, Ramon, *Planning and Designing the Warehouse*, New Jersey, Prentice-Hall, 1996, p. 311.
7. BELL, Judith. *Doing Your Research Project: A Guide for the First-time Researchers in Education and Social Science*. 2ªed. Milton Keynes, England: Open University Press, 1989. 145p.
8. BIO, Sérgio Rodrigues. **Sistemas de Informação, um Enfoque Gerencial**. São Paulo: Atlas, 1989.
9. CAMARGO, Luiz. **Para que serve mesmo? Estudo aponta panorama da tecnologia da informação nas grandes corporações do país**, 2001. Disponível em: <<http://www.informationweek.com.br>> . Acesso em 30 ago. 2001.
10. COMPUTERWORLD, Revista de Tecnologia e Negócios do Mundo Corporativo. *Cases*. Disponível em: <<http://www.computerworld.com.br/cases/default.asp>>. Acesso em 17 set. 2001.
11. CRETELLA, José J. **Curso de Direito Administrativo**. Rio de Janeiro: Forence, 1995.
12. CW, NRC IT. *Data Warehouse*, 2000. Disponível em:<http://paper.eainet.com/~dev/nrcitcw/ttt_data_warehouses.html> . Acesso em 30 ago. 2001.
13. DEVLIN, B. *Data Warehouse: from arquitetura to implementation*. Addison Wesley, 1997
14. FEW, Stephen C. *Archiving Data Warehouse Success: How to Avoid a Data Warehouse Meltdown*, 2001. Disponível em:

- <http://www.brio.com.au/library/white_papers/wp_meltdown.html> . Acesso em 30 ago. 2001.
15. FURLAN, J. D. Sistemas de Informações Estratégicas. 1997. **Anais Eletrônicos**. Disponível em: <<http://www.abep.sp.gov.br/Anais2511.html#Sistemas>> Acesso em 15 mar. 2001
 16. GIACOMONI, James, **Orçamento Público**. 9ª ed. São Paulo: Atlas, 2000
 17. GRIFFIN, Jane. *Fundamental Pitfalls to Avoid in your Data Warehouse*, 2000. Disponível em: <<http://www.datawarehouse.com/iknowledge/articles/article.cfm?ContentID=218>> . Acesso em 30 ago. 2001.
 18. GUPTA, C. R. **A introduction to Data Warehousing**, 1997. Disponível em: <<http://system-services.com/dwintro.htm>> Acesso em: 20 jan. 2001
 19. HACKNEY, Douglas. *Achitectures and Approaches for Successful Data Warehouses*, 2001. Disponível em: <http://itreports.computerworld.com/data/detail?id=993739718_147&type=RES&x=1421808361> > . Acesso em 30 ago. 2001.
 20. HACKNEY, Douglas. *Data Warehouse Delivery: Who are You? Part I. DM Review Magazine*, v. 8, n. 2, 1998.
 21. HARJINDER, G., RAO P. C. **The Official Guide to Data Warehousing**. The Corporation, 1996
 22. HARRISON, T. H., **Intranet Data Warehouse: Ferramentas e Técnicas para a Utilização do Data Warehouse na Intranet**. São Paulo, Berkeley Brasil, 1998
 23. IBM, Information Warehouse: An introduction, GC26-4876-00, IBM Corporation, 1991
 24. INMON, Bill. *Information Management: Charting the Course: What a Data Warehouse Is not*, 2001. Disponível em: <<http://www.dmreview.com/portal.cfm?NavID=91&EdID=3606&PortalID=25&Topic=55>> . Acesso em 30 ago. 2001.
 25. INMON, W. H. **Building the Data Warehouse**. New York: John Wiley & Sons, Inc., 1993
 26. INMON, W. H. **Como Construir o Data Warehouse**. Rio de Janeiro: Campus, , 1997

27. INMON, W. H. et al. **Gerenciando Data Warehouse**, São Paulo: Makron Books, 1999
28. INMON, W. H., HACKATHORN, Richard D. **Como usar o Data Warehouse**, Rio de Janeiro: IBPI, 1997
29. KIMBALL, R. **Data Warehouse Toolkit: Técnicas para Construção de Data Warehouse Dimensionais**. São Paulo: Makron Books, 1998b.
30. KIMBALL, R. **The Data Warehouse Lifecycle: Expert Methods for Designing, Developing and Deploying Data Warehouses**. New York John Wiley & Sons Inc., 1998a
31. KIMBALL, R., MERZ, R., **Data Webhouse: Construindo o Data Warehouse para Web**. Rio de Janeiro: Campus, 2000
32. KIMBALL, R., **Novos segredos dos armazéns de dados**, Computerworld, n° 307, 11 out. 1999, Entrevista concedida a Lucas Freitas
33. KOHAMA, HEILIO, **Contabilidade Pública – Teoria e Prática**. São Paulo: Atlas, 2000
34. LAUDON, Kenneth C., et al. **Sistemas de Informação**. Rio de Janeiro: LTC, 1999
35. LIEBSCHER, P. **Quantily with quality? Teaching quantitative and qualitative methods in a LIS Master's program**. Library Trends, v.46, n.4, Spring, 1998. p. 668-680
36. MACHADO, Felipe Nery Rodrigues. **Projeto de Data Warehouse – Uma visão Multidimensional**. São Paulo: Érica, 2000
37. MCGEE, Kimberlee R. **The Bottlenecks of Implementing a Successful Data Warehouse**. Dissertação (Master of Science in Interdisciplinary Studies). The University of Texas at Arlington, 1997,
38. MEIRELLES, Hely Lopes. **Direito Administrativo Brasileiro**, São Paulo: Revista dos Tribunais, 1984
39. MELLO, Celso Antonio Bandeira de. **Curso de Direito Administrativo**. São Paulo: Malheiros Editores, 1993
40. MELLO, Celso Antonio Bandeira de. **Prestação de Serviços Públicos e Administração Indireta**. São Paulo: Revista dos Tribunais, 1979
41. MELO, Rubens Nascimento. **Data Warehousing (Tutorial)**. In: XIII SBBB – Simpósio Brasileiro de Banco de Dados, Salvador, 1997.

42. OEA. Programa de Administração para o Desenvolvimento. Avaliação Orçamentária , **Revista ABOP**, maio/agosto, 1975
43. OLIVEIRA, Djalma de Pinho Rebouças de. **Sistemas de Informações Gerenciais**. São Paulo: Atlas, 1996
44. ONEIL, B. **Oracle Data Warehousing**. Indianapolis: Sams Publishing, 1997
45. ORFALI, Robert et al. **The Essential Client/Server Survival Guide**. New York: John Wiley & Sons, Inc., 1996
46. ORR, Ken. **Data Warehousing Tecnology**. The Ken Orr Institute, 1996
47. **PDRAE** - Plano Diretor da Reforma do Aparelho do Estado. Ministério da Administração e Reforma do Estado, 1995:19
48. PERKINS, A. **Critical Sucess Factors for Data Warehouse Engeneering**, 2000. Disponível em: <<http://www.visible.com/about/whitep.html>>. Acesso em 12 fev. 2001.
49. RUDIO, F.V. **Introdução ao projeto de pesquisa científica**. Petrópolis: Vozes, 128 p., 1986.
50. SADANANDAN, P. **Data Warehouse as a Great Knowledge Enabler**, 2001. Disponível em: <<http://www.intelesoftech.com/whitepaper.html>>. Acesso em 30 ago. 2001.
51. SAMANT, M. **Data Warehouse Concept Paper**, 2000. Disponível em: <http://itreports.computerworld.com/data/detail?id=991673572_267&type=RES&x=1959076736> . Acesso em 30 ago. 2001.
52. SARNEY, José. **Palavras do Presidente**. Presidência da República, 1985, v.2. p.83.
53. SENGE, Peter M. **A quinta disciplina: Arte e prática da organização que aprende**. 5ª ed. São Paulo: Best Seller, 1999.
54. SILVA, E. L., MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. 2.a ed. rev. Laboratório de Ensino a Distância da UFSC, Florianópolis, 2001.
55. SILVA, Fernando A. Rezende da. **Finanças Públicas**. São Paulo: Atlas, 1979
56. SINGH, Harry S. **Data Warehouse: Conceitos, Tecnologias, Implementação e Gerenciamento**. São Paulo: Makron Books, 2001.

57. SOARES, Vânia de Jesus Araújo. **Modelagem Incremental no ambiente de *Data Warehouse***. Tese de Doutorado apresentada à Universidade Federal do Rio de Janeiro. Rio de Janeiro : UFRJ, 1998.
58. SOUZA, José Pedro Galvão de. **Iniciação à Teoria do Estado**. São Paulo: Revista dos Tribunais, 1976
59. STAIR, Ralph M. **Princípios de Sistemas de Informação: Uma Abordagem Gerencial**. Rio de Janeiro: LTC, 1998.
60. TANLER, Rick. **Intranet *Data Warehouse***. Rio de Janeiro: Inforbook, 1998
61. TRIVIÑOS, A.N.S. **Introdução à Pesquisa em Ciências Sociais: a pesquisa qualitativa em educação**. São Paulo: Atlas, 1987.
62. WHITE, Gary Leon. **Data Integration With *Data Warehousing* and *Data Mining in Database Environments***. Department of Computer Science, University of Nevada, Las Vegas, 1996, Dissertação (Master of Science in Computer Science).
63. YIN, R.K. ***Case study research. Design and methods***. Beverly Hills, Sage, CA (USA), 1984.

GLOSSÁRIO

Data Webhouse – É a instanciação de Web do *Data Warehouse*.

B2B – (*Business-to-Business*) - É o termo que descreve o relacionamento entre empresas.

É comum para utilizá-lo para referir-se ao comércio ou a colaboração entre empresas, principalmente através de extranets ou da Internet.

B2C – (*Business-to-Consumer*) - Termo que descreve o relacionamento entre as empresas e seus consumidores. Normalmente faz referência ao atendimento direto ao cliente, através da Internet.

BSC – (*Balanced Scorecard*) - é uma técnica que permite a tradução da visão de futuro da organização em objetivos estratégicos correlacionados entre si. Tais objetivos são abrangidos por quatro dimensões: financeira, mercadológica, interna (processos), aprendizado e crescimento. A partir de indicadores e ações relacionados com objetivos assegura-se o alinhamento de iniciativas com a estratégia da organização em um processo de feedback e aprendizado contínuo.

Business Intelligence – conjunto de processos e ferramentas que possibilitam habilidades para estruturar, acessar e explorar informações, normalmente guardadas em DW e *Data Marts*, melhorando o processo de tomada de decisão.

CASE (*Computer-aided Software Engineering*) — engenharia de software auxiliada por computador.

Chave estrangeira — um atributo que não é uma chave primária em um sistema relacional, mas cujos valores são os da chave primária de uma outra relação.

Chave primária — um atributo que contém valores que identifica de modo único o registro no qual a chave existe.

CI – (*Competitive Intelligence*) - Disciplina capaz de integrar o planejamento estratégico, *marketing* e informação, objetivando o monitoramento constante do ambiente externo.

CRM – (*Customer Relationship Management*) (Gerência do Relacionamento com Clientes). CRM é o modelo de negócios centrado no cliente com a idéia de estabelecer relacionamento com os clientes de forma individual e depois usar as informações coletadas para tratar clientes diferentes de maneira diferente.

Drill-Down – é uma técnica que permite que o usuário desça ao nível de detalhes ao analisar um problema.

EAI – *Enterprise Application Integration* - Camada do ERM, responsável pela integração das camadas de CRM e PRM, assim como a integração entre os ERP e o mundo dos legados. Na camada EAI se posicionam tecnologias genéricas, que tem o objetivo de juntar as pontas dessas outras camadas, fazendo com que o processo básico da empresa não seja interrompido por hiatos tecnológicos.

ERP (*Enterprise Resource Planning*) - ERP é o termo que descreve uma série de atividades de gestão empresarial suportadas por aplicações de IT. Estas são compostas de muitos módulos, incluindo planejamento de produto, compras, estoque, relacionamento com fornecedores, atendimento ao cliente e acompanhamento de pedidos. Em seu uso corrente, o termo ERP engloba também módulos para as áreas financeira e de recursos humanos. Normalmente um ERP utiliza ou é integrado a um banco de dados, e a implantação de um sistema de ERP envolve uma profunda análise do negócio da empresa, treinamento de funcionários e modificações ou criação de procedimentos.

HTML (*Hypertext Mark-up Language*) - É uma linguagem padronizada de definição de layout de documentos e links de hipertexto, independente de plataforma. HTML é um subconjunto do SGML (*Standard Generalized Markup Language*) e é usada para

conectar documentos na *World Wide Web*.

Instantâneo (snapshot) — uma descarga de um banco de dados ou o arquivamento de dados fora de um banco de dados representando um determinado momento.

JAD (Joint Application Design) — uma união de pessoas —geralmente usuários finais — que criam e refinam requisitos de sistemas de aplicação.

KMS (Knowledge Management System) - foi desenvolvido na Carnegie Mellon University. É um sistema importante na evolução da hipermídia, sendo projetado para gerenciar grandes redes de hipertexto através de redes locais. Com estas características, esta ferramenta tem sido bastante utilizada para trabalho cooperativo, publicações eletrônicas, gerenciamento de projetos, mensagens eletrônicas e manuais técnicos.

PRM (Partner Relationship Management) - Metodologia e ferramentas que uma corporação emprega para auxiliar na gestão do relacionamento de interesses mútuos com outras empresas, como co-propriedade de marcas, produtos e patentes, parceria OEM, revenda ou distribuição e co-patrocínio. PRM tem muito em comum com o CRM e, como este, pode utilizar tecnologia de informação para manter e acompanhar esse tipo de relacionamento.

Roll Up – permite que o usuário reduza o escopo da análise. Ele sobe o nível de detalhe, ou seja, incrementa o nível de agregação.

SDLC (System Development Life Cycle) — o ciclo de vida do desenvolvimento de sistemas operacionais clássicos que tipicamente inclui o levantamento de necessidades, análise, projeto, programação, teste, integração e implementação.

SGBD (sistema de gerenciamento de bancos de dados) — um sistema de software baseado em computador e utilizado para conservar e gerenciar dados.

SGBD Multidimensional – é um gerenciador de banco de dados que utiliza uma estrutura

multidimensional para armazenamento de suas informações.

Slice-Dice – seleção e projeção. Tecnologia que permite ao usuário analisar a informação de diferentes prismas, atribuindo a consulta diferentes ângulos de análise.

Tecnologia da Informação (IT - Information Technology) - É o termo que engloba toda tecnologia utilizada para criar, armazenar, trocar e usar informação em seus diversos formatos (dados corporativos, áudio, imagens, vídeo, apresentações multimídia e outros meios, incluindo os que não foram criados ainda). É um termo conveniente para incluir a tecnologia de computadores e telecomunicações na mesma palavra. Essa convergência está conduzindo a "revolução da informação".

ANEXO I - INSTRUMENTOS DE COLETA DE DADOS

A seguir são apresentados os roteiros das entrevistas semi-estruturadas, utilizados como instrumento de coleta de dados. Cada roteiro serviu para direcionar o foco da entrevista, de acordo com o grupo do entrevistado. As perguntas foram abertas e o entrevistado teve a liberdade de discorrer a resposta, o que permitiu em alguns casos que, uma resposta servisse para responder outras perguntas que no momento ainda não haviam sido colocadas para o entrevistado. Houve também a preocupação do entrevistador, em não induzir a resposta.

INSTRUMENTO DE PESQUISA

ROTEIRO DE ENTREVISTA SEMI-ESTRUTURADA

Grupo Usuários - GU: os dirigentes, gerentes e profissionais, que foram, são ou serão usuários finais de tecnologias DW

1. Qual o seu cargo e função?
2. Em que departamento o Sr(a). está lotado(a)?
3. Quais as missões e metas da instituição?
4. Explique a estrutura organizacional da instituição, dando ênfase ao nível estratégico e as áreas envolvidas com as tecnologias de informação.
5. Qual a importância das informações estratégicas para a instituição? Existe algum órgão, departamento ou staff responsável pela gestão de informações estratégicas?
6. Explique os motivos e o histórico de utilização de tecnologias de *Data Warehouse*, na instituição. Quais os líderes e gestores da sua utilização?
7. Em que negócios/atividades da instituição se tem implementado tecnologias de *Data Warehouse*?
8. Quais são os usuários finais de tecnologias de *Data Warehouse*, na instituição?
dados importantes: analistas de SAD; analista de SIE; gerentes; executivos.
9. A instituição envia representantes para seminários, congressos, ou treinamentos específicos sobre a implementação e utilização de tecnologias de *Data Warehouse*?
10. Quais as expectativas dos usuários finais sobre a utilização de tecnologias de *Data Warehouse*, na instituição?
11. Quais as principais fontes de dados operacionais utilizados para carregar (alimentar) o *Data Warehouse*?
dados importantes: ERP; CRM; OLTPs.
12. Quais as principais fontes de dados externos utilizados para carregar (alimentar) o *Data Warehouse*?
13. Quais as ferramentas/aplicativos utilizadas para acessar os dados do *Data Warehouse*? Quais os usuários, e que dificuldades encontram nas mesmas?

dados importantes: SIE; SAD; OLAP; SQL; DATAMINING; a interface gráfica; tempo de resposta; intranet; Internet.

14. A instituição utiliza tecnologias de *Data Warehouse* em parceria (conexão) com outras instituições? Caso positivo, por favor, explique como?
15. Quais as dificuldades na implementação e utilização de tecnologias de *Data Warehouse*, inerentes ao fato da instituição ser pública?

dados importantes: cultura; burocracia do serviço público; liderança; governança; poucos investimentos; problemas orçamentários; treinamento; falta de pessoal técnico.

INSTRUMENTO DE PESQUISA

ROTEIRO DE ENTREVISTA SEMI-ESTRUTURADA

Grupo Técnicos - GT: os técnicos responsáveis pelas atividades de Tecnologia da Informação (TI), que foram, são, ou serão envolvidos em atividades de projeto, implementação, manutenção e monitoramento de tecnologias de DW.

1. Qual o seu cargo e função?
2. Em que departamento o Sr(a). está lotado(a)?
3. Explique os motivos e o histórico de utilização de tecnologias de *Data Warehouse*, na instituição. Quais os líderes e gestores da sua utilização?
4. Em que negócios/atividades da instituição se tem implementado tecnologias de *Data Warehouse*?
5. Quais são áreas da instituição e os técnicos envolvidos em atividades de projeto, implementação, manutenção e monitoramento de tecnologias de *Data Warehouse*?
6. A instituição envia representantes para seminários, congressos, ou treinamentos específicos sobre a implementação e utilização de tecnologias de *Data Warehouse*?
7. Quais as expectativas dos profissionais de tecnologia de informação (TI), sobre a utilização de tecnologias de *Data Warehouse*, na instituição?
8. Quais as principais fontes de dados operacionais utilizados para carregar (alimentar) o *Data Warehouse*? Por favor, explique o tipo de hardware e software utilizados, bem como, a integração dessas fontes.
dados importantes: ERP; CRM; OLTPs; SGBDs.
9. Quais as principais fontes de dados externos utilizados para carregar (alimentar) o *Data Warehouse*?
10. Foram realizados projetos de arquitetura e infra-estrutura? Caso positivo, por favor, explique os projetos.
11. Foi adotada alguma metodologia de desenvolvimento? Caso positivo, por favor, explique a metodologia.
12. Quais os investimentos de hardware e software que foram utilizados na implementação do *Data Warehouse*?

13. Explique o processo de extração, transformação e transporte dos dados das fontes operacionais e externas para serem armazenados no *Data Warehouse*. Por favor, identifique os softwares e hardwares, utilizados no processo.

dados importantes: integração dos dados operacionais/externos em um ODS ou no DW; SGBDs.

14. Explique como os dados estão armazenados no *Data Warehouse*. Por favor, identifique os softwares e hardwares utilizados.

dados importantes: SGBDs; modelagem lógica e física; granularidade e particionamento; *Data Marts*.

15. Explique como é atualizado e acessado o dicionário de dados (metadado) do *Data Warehouse*?

16. Quais as ferramentas/aplicativos utilizadas para acessar os dados do *Data Warehouse*? Que dificuldades os técnicos encontram nas mesmas?

dados importantes: SIE; SAD; OLAP; SQL; DATAMINING; a interface gráfica; tempo de resposta; intranet; internet.

17. Quais as dificuldades na implementação e utilização de tecnologias de *Data Warehouse*, inerentes ao fato da instituição ser pública?

dados importantes: cultura; burocracia; liderança; governança; poucos investimentos; problemas orçamentários; treinamento; falta de pessoal técnico