Universidade de São Paulo
Faculdade de Economia, Administração e Contabilidade
Departamento de Administração

Contribuição ao Estudo da Implementação de Data Warehousing: Um Caso no Setor de Telecomunicações

Gilberto de Come

Orientador: Prof. Dr. Hiroo Takaoka

São Paulo
Novembro/2001

Reitor da Universidade de São Paulo
Prof. Dr. Jacques Marcovitch

Diretor da Faculdade de Economia, Administração e Contabilidade
Prof. Dr. Eliseu Martins

Chefe do Departamento de Administração
Prof. Dr. Claudio Felisoni de Angelo

Universidade de São Paulo
Faculdade de Economia, Administração e Contabilidade
Departamento de Administração

Contribuição ao Estudo da Implementação de Data Warehousing: Um Caso no Setor de Telecomunicações

Gilberto de Come

Orientador: Prof. Dr. Hiroo Takaoka

Dissertação apresentada ao Departamento de Administração da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo como requisito parcial para obtenção do título de Mestre em Administração de Empresas.

Área de Concentração:
Métodos Quantitativos e Informática

São Paulo
Novembro/2001

FICHA CATALOGRÁFICA

Come, Gilberto de
Contribuição ao estudo da implementação de Data Warehousing: um caso no setor de telecomunicações / Gilberto de Come. – São Paulo : FEA/USP, 2001.
133 p.

Dissertação - Mestrado
Bibliografia

1. Tecnologia da informação 2. Banco de dados 3. Sistemas de informação 4. Telecomunicação I. Faculdade de Economia , Administração e Contabilidade da USP.

CDD – 658.4038

À minha esposa **Fátima** e aos meus filhos **Paula** e **Eduardo**,
pela compreensão, paciência e incentivo durante os anos
dessa jornada. Vocês são o meu maior tesouro.

Agradecimentos

Uma dissertação de mestrado não é um trabalho isolado que pode ser realizado sem a colaboração de pessoas, mais especificamente de amigos. Nesse sentido eu gostaria de agradecer formalmente àqueles que me ajudaram nessa longa jornada.

Inicialmente os meus "chefes", nos tempos da CTBC, que apoiaram a realização desse curso. Ao Sr. **Paulo Eduardo Lopes Passos** e ao Sr. **Daniel Fernandes Pedão** agradeço a compreensão pelas ausências programadas e também pelas eventuais.

Agradeço ao Prof. Dr. **Hiroo Takaoka**, orientador desse trabalho, que participou desde a fase inicial de escolha do tema durante suas aulas de Análise de Informações. As orientações objetivas, bem como o apoio e a amizade, foram fundamentais para o desenvolvimento do trabalho.

Um trabalho acadêmico não envolve apenas estudos e pesquisas e nesse sentido eu gostaria de agradecer à **Licia Abe**, assistente do Prof. Hiroo e também à **Valéria**, do setor de pós-graduação, que ajudaram a resolver pendências administrativas e que sempre tiveram a paciência necessária para solucionar minhas dúvidas.

Para aqueles que participaram das entrevistas, vale um agradecimento especial ao Sr. **Júlio César Muraro** (Marketing) e à Sra. **Márcia Madaloso** (Sistemas) que forneceram informações importantes para o entendimento do projeto e sempre demonstraram muita disposição no atendimento de minhas solicitações. Não menos importantes, nessa fase, foram as participações do Sr. **Daniel F. Pedão** (Diretor de Sistemas), **Márcia Siqueira** e **Flávia** (Sistemas), **César Pascual** (Consultoria), **Audrey C. Mattos** (Marketing), **Alexandre Pinho** (Infra-estrutura de informática) e **Vírginio Mattos** (EDS).

Um agradecimento particularmente especial vai para meu irmão **Eduardo** que desde o início foi o grande incentivador e a quem devo o "ponta-pé" inicial dessa jornada. Seu apoio foi muito importante durante todo esse tempo.

Aos meus pais, **Roque** e **Ignez**, agradeço pela educação e pelos ensinamentos que serviram de base para a chegada até aqui. Agradeço também à minha irmã **Claudia** que sempre colaborou quando foi necessário.

Agradeço a **Deus** que tanto tem abençoado minha vida, meu lar, minha família.

Muito obrigado a todos.

Resumo

O desenvolvimento da tecnologia de informação, aliado à globalização e ao aumento da competitividade nos mais variados setores, está facilitando a integração entre o mercado produtor e consumidor. A cada negócio realizado, ou seja, a todo momento, uma grande quantidade de dados é gerada e armazenada, passando a ser um recurso importante da empresa.

O grande problema está em como tratar esse grande volume de dados, geralmente espalhados por diversos sistemas de informação, de forma a obter alguma vantagem competitiva através da informação certa no momento certo.

O *Data Warehouse* surgiu nos anos 90 como uma maneira flexível e eficiente de tratar esses grandes volumes de dados e obter informações que auxiliem no processo para tomada de decisão, mas sua implementação não é tão simples. Projetos dessa natureza demandam uma grande quantidade de tempo e dinheiro e portanto precisam de uma atenção especial.

O setor de telecomunicações tem passado por uma grandes transformações nos últimos tempos. Além das novas tecnologias, o aumento da competição entre as empresas está cada vez maior. No Brasil a grande virada ocorreu em 1998 quando as empresas do setor foram privatizadas e o mercado foi aberto para a concorrência. Nesse novo cenário as empresas precisam entender o mercado em que atuam e mais que isso, conhecer muito bem o cliente. Nesse sentido o *Data Warehouse* surge como uma interessante alternativa.

Este trabalho procura entender como foi, ou como está sendo feita a implementação do *Data Warehouse* em uma empresa do setor de telecomunicações que atua no Brasil, através da análise dos aspectos gerenciais, aspectos tecnológicos e aspectos humanos envolvidos no projeto.

O levantamento bibliográfico, começa pela apresentação do ambiente de estudo, ou seja, o setor de telecomunicações no Brasil, o *Data Warehouse* e como eles se relacionam. Em seguida são apresentados vários conceitos – gerenciais, tecnológicos e humanos – que serviram de base para o desenvolvimento e conclusão do trabalho.

A pesquisa em campo foi realizada através do método de Estudo de Caso por ser entendido como uma forma eficiente de responder aos objetivos propostos, pois possibilita a consideração dos mais variados aspectos relativos ao fato estudado, no caso, a implementação do *Data Warehouse* em uma empresa de telecomunicações.

Abstract

The information technology evolution, allied to globalization and the increasing in competitiveness in different segments, is facilitating the integration between producer and consumer market. In each accomplished business, in other words, at every moment, a great amount of data is generated and stored, becoming an important resource of the company.

The big problem is how to handle this great data volume, usually spread in several information systems, in order to get some competitive advantage through the right information in the right moment.

The Data Warehouse appeared in the nineties, as a flexible and efficient way to handle these data and to get information, which aids in the process of decision-making, but its implementation is not that simple. This kind of project demands a great amount of time and money and therefore they need a special attention.

The telecommunications (telecom) segment is going through great changes in the last times. Besides the new technologies, the increasing in competition among the companies is becoming larger. In Brazil an important milestone happened in 1998 when the telecom companies were privatized and the market was opened for the competition. In this new situation the companies need to understand the market and more than that, know the customer very well. To address this need, the Data Warehouse appears as an interesting alternative.

This dissertation tries to understand how it was, or how the Data Warehouse implementation is being made in a company of the Brazilian telecom segment, through the analysis of the managerial, technological and human aspects involved in the project.

The literature review begins presenting the study environment, in other words, the Brazilian telecom segment, the Data Warehouse and the relation between them. Then several concepts – managerial, technological and humans – are presented and used as foundation for the development and conclusion of this work.

The empirical research was conducted by the Study of Case method, which was considered as an efficient way to reach the proposed objectives, as it makes possible the analyses of various aspects of the studied fact, in this case, Data Warehouse implementation in a telecom company.

Índice

Lista de Figuras vi

Lista de Tabelas vi

Capítulo 1 – O Problema 1

1.1 Introdução 1
1.2 Justificativas e Importância do Tema 4
1.3 Problema de Pesquisa 5
1.4 Objetivos 5
1.4.1 Objetivo principal *6*
1.4.2 Objetivos secundários *6*
1.5 Definições Operacionais 6
1.6 Organização do Trabalho 7

Capítulo 2 – O Ambiente de Estudo 8

2.1 O Setor de Telecomunicações 8
2.2 Data Warehousing 14
2.2.1 Estrutura dos dados *16*
2.2.2 Processamento Operacional x Processamento Informacional *17*
2.2.3 ODS - Operational Data Store *18*
2.2.4 OLTP x OLAP *19*
2.2.5 Data Warehouse *x SIG x SAD x SIE* *20*
2.2.6 Fábrica de Informações Corporativas *22*
2.3 O Data Warehousing no Setor de Telecomunicações 23

Capítulo 3 – O Projeto 26

3.1 Projeto 26
3.1.1 Motivação *27*
3.1.2 Patrocinador *29*
3.2 Aspectos Gerenciais 30
3.2.1 Gerência do Tempo *31*
3.2.2 Gerência de Riscos *32*
3.2.3 Gerência dos Custos *33*
3.2.4 Gerência da Comunicação *36*
3.3 Aspectos Tecnológicos 37
3.3.1 Arquitetura *37*
3.3.2 Hardware *40*
3.3.3 Granularidade e Sumarização *42*
3.3.4 Extração, Transformação e Carga *42*
3.3.5 Metadados *43*
3.3.6 Segurança *45*
3.3.7 Metodologia *46*
3.3.8 Modelagem de Dados *50*
3.4 Aspectos Humanos 55
3.4.1 Usuário *57*
3.4.2 Equipe de Projeto *58*
3.4.2 Facilidade de Uso e Utilidade *60*
3.4.3 Treinamento *61*
3.4.4 Envolvimento e Participação *62*
3.4.5 Consultoria *63*

Capítulo 4 – Metodologia de Pesquisa 65

4.1 O Estudo de Caso 66
4.2 Modelo Conceitual 69

4.3 O Caso ... 70
4.4 Instrumentação ... 70
4.5 Respondentes ... 71
4.6 Desenvolvimento da pesquisa ... 71

Capítulo 5 – O Estudo do Caso ... 72

5.1 A Empresa ... 72
5.2 Projeto ... 76
5.2.1 Motivação ... 77
5.2.2 Patrocínio ... 79
5.2.3 Considerações Iniciais ... 80
5.3 Aspectos Gerenciais ... 81
5.3.1 Gerência do Tempo ... 81
5.3.2 Gerência dos Riscos ... 83
5.2.3 Gerência dos Custos ... 84
5.2.4 Gerência da Comunicação ... 85
5.4 Aspectos Tecnológicos ... 86
5.4.1 Arquitetura ... 86
5.4.2 Hardware ... 88
5.4.3 Granularidade e Sumariação ... 89
5.4.4 Extração, Transformação e Carga ... 90
5.4.5 Metadados ... 92
5.4.6 Segurança ... 92
5.4.7 Metodologia ... 93
5.4.8 Modelagem de Dados ... 94
5.5 Aspectos Humanos ... 95
5.5.1 Usuário ... 95
5.5.2 Equipe do Projeto ... 95
5.5.3 Facilidade de Uso e Utilidade ... 98
5.5.4 Treinamento ... 98
5.5.5 Envolvimento e Participação ... 99
5.5.6 Consultoria ... 100

Capítulo 6 – Sumário e Conclusões ... 101

6.1 Sumário ... 101
6.2 Projeto ... 103
6.3 Aspectos Gerenciais ... 104
6.4 Aspectos Tecnológicos ... 106
6.5 Aspectos Humanos ... 110
6.6 Considerações Finais ... 113
6.7 Propostas para novos estudos ... 114

Glossário ... 115

Anexo 1 – Roteiro de Entrevistas ... 119

Anexo 2 – Endereços na Internet ... 122

Anexo 3 – Exemplos do Modelo Dimensional ... 123

Anexo 4 – Autorização para publicação ... 125

Referências Bibliográficas ... 127

Lista de Figuras

Figura 1 - Um dos desafios de um *Data Warehouse*: integrar dados 2
Figura 2 - O ambiente de *Data Warehouse* 15
Figura 3 - Integração dos dados 15
Figura 4 - Estrutura dos dados no *Data Warehouse* 16
Figura 5 - Fábrica de Informações Corporativas 23
Figura 6 - Modelo de sucesso de projeto 27
Figura 7 - Investimentos em *Data Warehouse* 34
Figura 8 - Abordagens *Bottom-Up* e *Top-Down* 40
Figura 9 - O processo de Extração, Transformação e Carga 43
Figura 10 - Desenvolvendo o *Data Warehouse* 47
Figura 11 - SDLC Clássico x SDLC *Data Warehouse* 50
Figura 12 - Modelo Entidade-Relacionamento 51
Figura 13 - Modelo Dimensional 52
Figura 14 - A Tabela de Fatos 52
Figura 15 - Modelo "Floco de Neve" 53
Figura 16 - O "Cubo" 55
Figura 17 - Alinhando pessoas, processos e tecnologia 56
Figura 18 – Entrade de Dados e Saída de Informações 60
Figura 19 - Envolvimento do usuário no ciclo de vida de um SI 62
Figura 20 - Modelo Conceitual 70
Figura 21 - Principais áreas envolvidas 77
Figura 22 - Arquitetura do *Data Warehouse* 86
Figura 23 - Organograma básico da equipe 96

Lista de Tabelas

Tabela 1 – Acessos fixos telefônicos por mês (em milhões de linhas) 12
Tabela 2 – Acessos fixos telefônicos por ano (em milhões de linhas) 13
Tabela 3 – Evolução da densidade telefônica (em %) 13
Tabela 4 - Sistema Operacional x Sistema Informacional 18
Tabela 5 - Comparação entre SIG, SAD e SIE 22
Tabela 6 - A falta de elementos de patrocínio 30
Tabela 7 - *Data Mart* x *Data Warehouse* 38
Tabela 8 - *Bottom Up* e *Top Down* - Vantagens e Desvantagens 40
Tabela 9 - Metodologias/Fases para o *Data Warehouse* 48
Tabela 10 - Análise com uma dimensão 54
Tabela 11 - Análise com duas dimensões 54
Tabela 12 - Análise com três dimensões 54
Tabela 13 - Papéis no ciclo de vida do *Data Warehouse* 59
Tabela 14 - Metas da Anatel alcançadas 74
Tabela 15 - Fases da Metodologia 93
Tabela 16 - Papéis e responsabilidades no projeto 97

Capítulo 1 – O Problema

1.1 Introdução

O desenvolvimento da tecnologia de informação (TI) tem permitido que as empresas manipulem um grande volume de dados ao mesmo tempo que facilita a integração entre o mercado produto e consumidor. A cada operação, a cada negócio realizado, ou seja, a todo momento, uma quantidade enorme de dados sobre a empresa, o cliente, os produtos, o mercado, etc., são gerados e armazenados passando a ser um importante recurso corporativo. O problema é que esses dados, geralmente, encontram-se espalhados por diversos sistemas e exigem um grande esforço na tentativa de integrá-los para que possam ter alguma utilidade.

Segundo Torres (1995), no mundo atual dificilmente se pode competir, na grande maioria dos negócios, sem o uso da TI, em outras palavras, existe um relacionamento muito grande entre a TI e o comportamento estratégico de uma organização. Nesse sentido a TI não é apenas um elemento operacional ou integrante de produtos e serviços pois tem seu valor estratégico numa de suas funções mais tradicionais que seria o fornecimento de informações para a gestão. Dessa forma, muitas informações de naturezas estratégicas podem tornar a empresa mais competitiva na medida que melhoram a tomada de decisão (Torres, 1995). O uso estratégico da informação tem se tornado uma necessidade cada vez mais clara para as empresas. Quando a informação é bem utilizada ela pode agregar valor ao negócio, além de gerar outros benefícios para a empresa, como reduzir custos por exemplo.

Na tentativa de tratar o problema, as empresas têm implantado diversos Sistemas de Apoio à Decisão (SAD[1]) para conseguir informações através de seus dados. Basicamente o apoio à decisão consiste no processo de agrupar, estruturar, manipular, armazenar, acessar, apresentar e distribuir informações de negócios de maneira oportuna, ou seja, a informação certa no momento certo e na quantidade certa. Sprague (1991) coloca que um SAD concentra-se em recursos que simplifiquem seu uso para um pessoal não especializado em computação de forma a facilitar o processo de tomada de decisão.

Segundo Watson (1998), grandes varejistas, bancos e Telcos[2] começaram a construir grandes armazéns de dados em meados dos anos 80, mas somente no início dos anos 90 Bill Inmon criou o termo *Data Warehouse*. A partir de então, mais ferramentas se tornaram disponíveis e o *Data Warehouse* tornou-se um dos itens mais importante no mundo da computação corporativa.

Nos últimos anos o *Data Warehouse* vem oferecendo às organizações uma maneira flexível e eficiente de obter as informações que os gestores necessitam nos processos decisórios e se caracteriza como uma função de apoio à decisão.

[1] Tradução para a sigla DSS (*Decision Support System*)
[2] Telco, forma abreviada de "*telephone company*", é o termo normalmente usado para designar as empresas de telefonia.

Knowles (1996) utiliza uma lógica interessante para dizer como o *Data Warehouse* é importante para a empresa: "Poder faz dinheiro. Conhecimento é poder. *Data Warehouse* aumenta o conhecimento. Portanto, *Data Warehouse* faz dinheiro."

Gurovitz (1999) cita um estudo do IDC (International Data Corporation) que coloca o *Data Warehouse* como a melhor chance para a TI finalmente mostrar ao que veio, gerando ganhos de tempo e dinheiro com informações acessíveis aos executivos quando e como eles quiserem.

Segundo Ralph Kimball (1998), uma autoridade nesse assunto, *Data Warehouse* "é o lugar onde as pessoas podem acessar seus dados". Já Wang (1998) tem uma definição um pouco mais elaborada quando diz que *Data Warehouse* "é o processo pelo qual os dados relacionados de vários sistemas operacionais são fundidos para proporcionar uma única e integrada visão de informação de negócios que abrange todas as divisões da empresa".

Normalmente os sistemas de informação das empresas estão literalmente "distribuídos" o que dificulta o entendimento da corporação como um todo. A maioria dos sistemas é instalada com uma visão local e seu principal propósito é resolver um problema singular e isolado, tal como folha de pagamento, planejamento de manufatura, acompanhamento de clientes, etc. Embora não necessariamente incorreta, esta abordagem começa a apresentar problema quando uma visão cruzada é necessária para o entendimento da dinâmica da situação. A integração dessas informações dispersas é um dos desafios do *Data Warehouse* (Figura 1).

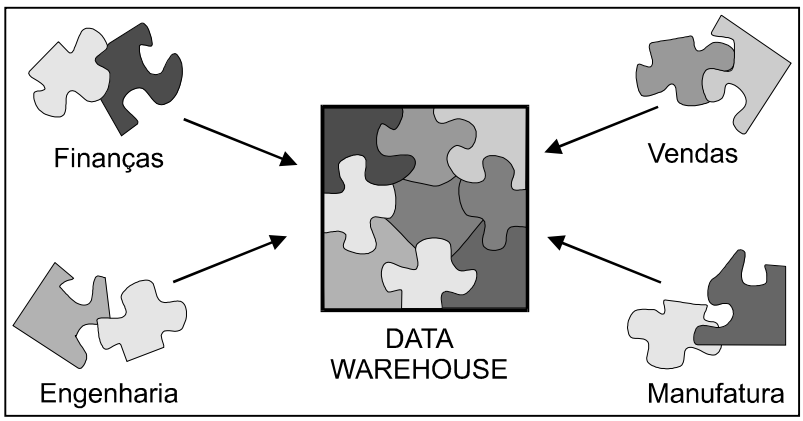


Figura 1 - Um dos desafios de um *Data Warehouse*: integrar dados
Fonte: elaborado pelo autor

A simples integração não é suficiente pois, após a coleta, os dados devem ser analisados para determinar sua significância. Falhas na implementação de sistemas e métodos para análise desses dados colocarão a empresa em desvantagem no mercado que está cada vez mais competitivo. Quanto mais os dados se tornam disponíveis, mais complexo se torna o processo de localizar e extrair informações realmente importantes.

Para Gagnon (1999), as questões sobre cliente satisfeito, maiores lucros e melhoria da participação no mercado são fundamentais para os negócios no competitivo mercado dos anos 90. Nesse caso é importante encontrar uma forma para o negócio poder tratar todas essas questões ao mesmo tempo. Sem dúvida a TI tem uma participação muito importante nesse contexto e o *Data Warehouse* pode ser uma resposta pois coleta os dados operacionais, dados históricos e dados externos; consolida-os em um banco de dados; gerencia-os e molda-os em um formato orientado a assuntos que é otimizado para o usuário final acessar e analisar.

Singh (1997) nos relata um breve fato que serve para ilustrar, com bom humor, a necessidade de informações de qualidade baseadas em dados confiáveis no ambiente organizacional:

"Ao final de uma palestra, o executivo encontra um grande guru da administração e pergunta: 'Oh guru, como eu posso garantir o sucesso para minha empresa?' O guru responde: 'Informação'. Então o executivo faz outra pergunta: 'Oh guru, e o quê eu mais devo temer?' O guru responde: 'Dados'."

Gurovitz (1999) procura mostrar a importância estratégica de um *Data Warehouse* a partir de três exemplos de utilização em setores que vêm fazendo grande uso de TI como componente estratégico para organização:

- No setor bancário temos o banco Itaú (pioneiro no uso de *Data Warehouse* no Brasil) que costumava enviar mais de 1 milhão de malas diretas para os correntistas mas cujo retorno não passava de 2%. A partir da análise dos dados armazenados, as cartas passaram a ser enviadas apenas para aqueles com maior chance de responder. Com isso a taxa de retorno subiu para 30% e a conta do correio foi reduzida para um quinto.

- Uma das maiores redes de varejo americana, o Wal-Mart, descobriu uma interessante relação entre a venda de fraldas e cervejas (este é um caso clássico no mundo do *Data Warehouse*), bem como entre a boneca Barbie e barras de chocolate. Com base nessas informações os produtos foram colocados lado a lado e as vendas aumentaram.

- A Sprint, um dos grandes do setor de telecomunicações americano, desenvolveu um método capaz de prever com 61% de certeza se um cliente trocaria de empresa no prazo de dois meses. Através de um marketing agressivo, conseguiu evitar a deserção de 120.000 clientes e uma redução de 35 milhões de dólares em faturamento.

Esses são apenas alguns exemplos de resultados obtidos a partir da boa utilização de um *Data Warehouse*. Mas isso não ocorre de forma rápida em uma empresa. Os projetos nesse sentido normalmente requerem altos investimentos e um certo tempo (às vezes anos) para acumular dados suficientes.

Os números com relação ao volume de dados armazenados também não são pequenos. O Wal-Mart (o maior *Data Warehouse* do mundo), em 1999, possuia cerca de 23 trilhões de caracteres, ou 60.000 anos de jornal diário. Isso equivale a 10 vezes toda a história humana, desde o surgimento da escrita (Gurovitz, 1999). Sem dúvida são números impressionantes que nem sequer poderiam ser imaginados até um certo tempo atrás.

Baker & Baker (1999) citam uma pesquisa do META Group Inc. onde mais de 90% das empresas da Fortune 2000 informaram ter um *Data Warehouse* ou estar em processo para desenvolver um. Com certeza a implementação vai variar de empresa para empresa pois depende de muitos fatores, mas uma coisa é certa, ela deverá ser realizada, caso contrário, a empresa tende a deixar de ser competitiva.

O propósito dessa pesquisa é analisar um caso de implementação de *Data Warehousing*[3] analisando suas principais características e fatores que foram considerados impactantes por alguns envolvidos no processo. Optou-se por estudar a situação em uma empresa de telecomunicações por três motivos principais:

- a crescente e forte concorrência existente nesse setor;
- o uso intensivo de tecnologia e
- a enorme quantidade de dados disponíveis.

1.2 Justificativas e Importância do Tema

A cada dia, o volume de dados disponíveis é maior e a necessidade de transformar esses dados em informação tem sido um grande desafio para as empresas. São cada vez maiores as pressões para prover informações de melhor qualidade para tomada de decisões, em formatos que sejam de fácil acesso e manipulação (Wang, 1998).

Como foi dito anteriormente, os dados nas empresas costumam estar distribuídos, dificultando muito o processo de transformá-los em informação consolidada que sirvam de base para a tomada de decisão. Segundo Singh (1997), as empresas normalmente não sofrem de falta de dados, mas de uma abundância e redundância de dados inconsistentes. O *Data Warehouse* vem sendo apontado como uma boa forma de por "ordem" nesse grande volume de dados ao mesmo tempo que pode transformá-los em informações úteis.

As empresas fazem grandes investimentos em TI na tentativa de gerenciar os negócios com mais eficiência. A necessidade de informação tem feito com que as empresas, venham investindo na construção de SADs que por sua vez estão adquirindo complexidade crescente e utilizando técnicas de tratamento de dados cada vez mais sofisticadas. As ferramentas para isso estão se tornando cada vez mais poderosas e o *Data Warehouse* faz parte desse conjunto que reflete o reconhecimento de que a exploração de dados é o caminho para vantagem competitiva. Como os investimentos são altos, os projetos não podem simplesmente ser "implantados" sem uma análise profunda de vários fatores, sejam técnicos ou administrativos, que podem garantir o sucesso do empreendimento trazendo o retorno esperado. Pinto & Slevin (1988) dizem que o processo de implementação de um projeto representa um desafio constante para os gerentes.

Herdlein (1996) coloca que construir um *Data Warehouse* provavelmente envolverá mais recursos humanos e de capital que qualquer outro projeto de TI que a empresa já possuiu. Nessas proporções, as chances de falhas não são pequenas. Muitos projetos de *Data Warehouse* param por causa de infra-estrutura técnica ineficiente, falta de suporte executivo, inexperiência das equipes de projeto e longo prazo para apresentação de resultados.

[3] Em uma tradução literal temos que "*Data Warehousing*" significa "Armazenamento de Dados". Neste trabalho vamos adotar o termo para designar a ampla variedade de soluções e produtos voltados para o gerenciamento, organização e exploração da base de dados corporativa (*Data Warehouse*). Normalmente os autores usam *Data Warehousing* e *Data Warehouse* sem distinção.

Portanto, o reconhecimento dos fatores impactantes é essencial para que o *Data Warehouse* seja projetado e construído corretamente de forma a propiciar uma significativa liberdade de acesso a dados, gerando amplos benefícios para qualquer organização. A necessidade da informação vem fazendo com que o *Data Warehouse* se torne um fator crítico de sucesso para a própria empresa. Por mais que uma empresa seja eficiente nas suas diversas áreas operacionais, se ela estiver vulnerável em fatores críticos de sucesso é muito provável que não consiga ter competitividade. Conforme Singh (1997), o *Data Warehouse*, não é simplesmente um produto ou processo, mas uma estratégia que reconhece a necessidade de consolidar os dados armazenados em sistemas de informações dedicados a ajudar os profissionais de negócios a tomarem decisões mais rápidas e efetivas.

Neste trabalho é adotado "fator impactante" ao invés de "fator crítico de sucesso" pois, como veremos adiante, existe controvérsia com relação ao término de um projeto de *Data Warehouse*. Alguns autores afirmam que é um processo constante, e sendo assim fica difícil determinar o ponto exato do término do projeto como um todo para definir se o mesmo teve sucesso ou não.

A pesquisa procura entender como o *Data Warehouse* foi ou está sendo implementado na empresa analisando as principais características apontadas na literatura relacionadas ao projeto, aspectos gerenciais, aspectos tecnológicos e aspectos humanos envolvidos no processo. A realização da pesquisa através de um estudo de caso permitiu conhecer também outros aspectos que foram considerados importantes para a implementação do *Data Warehouse*. Apesar do trabalho não estar focado diretamente no sucesso do projeto, essa visão é importante na exposição dos dados e conclusões já que, conforme Pinto & Slevin (1988a), muitos projetos de sistemas são percebidos como fracasso apesar de terem cumprido as especificações iniciais e outros são considerados sucessos, mesmos não tendo cumprido tais itens.

A grande maioria dos livros e artigos sobre *Data Warehouse* trata o assunto de forma técnica e os estudos acadêmicos ainda são poucos, talvez por ser um tema relativamente recente. Dessa forma, acredita-se que o resultado final desse trabalho terá sua validade para todos aqueles profissionais que estão de alguma forma envolvidos em um projeto de *Data Warehouse*.

1.3 Problema de Pesquisa

No sentido mais amplo, o problema é uma questão proposta para ser discutida e resolvida pelas regras da lógica e de outros meios que se dispõe. No caso desta pesquisa estamos interessados em descobrir como ocorre a implementação de *Data Warehousing* em uma empresa do setor de telecomunicações. Sendo assim a questão específica deste trabalho seria:

"Como foi, ou como está sendo feita a implementação de *Data Warehousing* na empresa?"

1.4 Objetivos

Neste trabalho os objetivos serão divididos em duas categorias: objetivo principal e objetivos secundários.

1.4.1 Objetivo principal

- Caracterizar o processo de implementação de *Data Warehousing* pelas perspectivas de aspectos gerenciais, aspectos tecnológicos e aspectos humanos.

1.4.2 Objetivos secundários

- Caracterizar, em linhas gerais, o processo de *Data Warehousing* e suas características no setor de telecomunicações.

- Identificar os aspectos que são apontados pela literatura como sendo impactantes na implementação de *Data Warehousing*.

- Identificar os aspectos que são apontados pelos respondentes (estudo de caso) como sendo impactantes na implementação de *Data Warehousing* e não foram destacados na revisão da literatura.

1.5 Definições Operacionais

Esse tipo de tratamento é essencial para o entendimento da metodologia e dos resultados da pesquisa. É preciso definir operacionalmente as variáveis e palavras visando a homogeneidade de interpretação e de medição no contexto metodológico do trabalho científico. Trata-se de estabelecer uma ponte entre os conceitos e as observações, atribuindo significado a um constructo ou variável, especificando as atividades ou operações necessárias para medí-lo ou manipulá-lo (Martins, 1990). Nesse trabalho vamos definir operacionalmente os principais termos que estão envolvidos na questão principal e nos objetivos:

Data Warehousing
Compreende a ampla variedade de soluções e produtos voltados para o gerenciamento, organização e exploração da base de dados corporativa, assim como sistema de apoio a decisão.

Implementação
Refere-se ao ato ou efeito de implementar, ou seja, dar execução a um plano, programa ou projeto.

Fatores Impactantes
São aqueles fatores que contribuem diretamente para o alcance dos objetivos.

Projeto
Um projeto é uma organização de pessoas dedicadas a um propósito ou objetivo específico (Pinto & Slevin, 1988). Essa definição atende bem aos propósitos desse trabalho pois outras características normalmente associadas a projetos como, início e fim definidos e orçamento limitado, algumas vezes, são temas polêmicos em se tratando de *Data Warehouse*. Vale ressaltar que o projeto, nesse caso, não deve ser confundido com o desenho (*design*) do *Data Warehouse.*

Aspectos Gerenciais
Os aspectos gerenciais estão voltados para o gerenciamento de um projeto e consistem em coordenar, integrar e monitorar todas as atividades necessárias para alcançar os objetivos do projeto.

Aspectos Tecnológicos
As características relacionadas ao ***software*** (programas) e ***hardware*** (equipamentos) que fazem parte do projeto de *Data Warehouse*.

Aspectos Humanos
De forma geral é o estudo científico do relacionamento entre os humanos e o seu ambiente de trabalho. O conhecimento adquirido deste estudo é usado para criar sistemas e ambientes de trabalho que ajudam a fazer pessoas mais produtivas e mais satisfeitas com seu trabalho (Beard & Peterson, 1988). Essa pesquisa procura mostrar de que forma os humanos (área de informática e usuários) estiveram envolvidos no projeto.

1.6 Organização do Trabalho

Além desse capítulo inicial, onde o problema foi apresentado e os objetivos da pesquisa foram definidos, este trabalho compreende os seguintes capítulos:

Capítulo 2 – O Ambiente de Estudo,
onde busca-se caracterizar o ambiente em que ocorre a pesquisa. Em primeiro lugar é feito um resumo histórico do setor de telecomunicações no Brasil, desde o início até os momentos atuais. Em seguida são mostrados os conceitos de *Data Warehouse* em linhas gerais, destacando algumas características que o diferencia de outros sistemas. Finalmente é feita uma apresentação geral de como estes dois ambientes se relacionam atualmente.

Capítulo 3 – O Projeto,
onde é feita toda a fundamentação teórica para o desenvolvimento da pesquisa. A partir da literatura existente, buscou-se os principais conceitos envolvidos com os objetivos do trabalho relativos a aspectos gerenciais, aspectos técnicos e aspectos humanos.

Capítulo 4 – Metodologia de Pesquisa,
onde é feita a justificativa para o método de pesquisa escolhido, o estudo de caso, assim como são citadas suas principais características. Também são apresentados o modelo conceitual e a forma de desenvolvimento da pesquisa.

Capítulo 5 – O Estudo de Caso,
que se refere a pesquisa propriamente dita. É a organização das informações que foram coletadas na entrevistas na mesma estrutura apresentada no Capítulo 3.

Capítulo 6 – Sumário e Conclusões,
onde é feito um sumário de todo o trabalho realizado e também são apresentadas as principais conclusões, baseadas no relacionamento entre a teoria (Capítulo 3) e a prática (Capítulo 5). Também são apresentadas algumas propostas para outros estudos.

Capítulo 2 – O Ambiente de Estudo

Como o trabalho vai analisar um caso de *Data Warehousing* no setor de telecomunicações, é importante colocar uma base teórica, em linhas gerais, tanto do setor de telecomunicações como do *Data Warehouse* e em seguida mostrar algumas características do setor de telecomunicações que indiquem a necessidade de um projeto dessa natureza. Características de projeto e seus aspectos gerenciais, aspectos tecnológicos e aspectos humanos também fazem parte dos objetivos do trabalho e serão analisados com mais detalhes no próximo capítulo.

Por se tratar de um assunto que se tornou popularmente conhecido a partir dos anos 90, a literatura existente é bastante recente, podendo-se até dizer que a grande parte surgiu nos últimos quatro anos. Ralph Kimball e Bill Inmon são dois autores considerados por muitos como os "pioneiros" do *Data Warehouse*, e isso significa que seus livros e artigos influenciaram, e ainda influenciam, muitas pessoas no processo de construir um *Data Warehouse*. Apesar de em muitos momentos seus livros serem bastante técnicos, permitem ao leitor uma boa base teórica a respeito do assunto.

Muitos autores não se limitam a escrever livros e costumam abastecer os periódicos (revistas técnicas) com suas visões sobre determinados pontos, normalmente de maneira curta e objetiva, mas que no conjunto podem ser de grande valia para este trabalho. Dentre os periódicos pode-se destacar: Computerworld, Data Management Review, DBMS Magazine e Datamation.

No início de uma pesquisa todo tipo de informação é válido no sentido de se formar uma visão abrangente do assunto, principalmente pelo fato de que a quantidade de trabalhos acadêmicos ainda é muito pequena nessa área. Na Internet podemos encontrar uma enorme quantidade de informações, que muitas vezes não são incluídas no trabalho por não estarem diretamente relacionadas ao tema mas que podem ser uma ponte para novas informações. Alguns *sites*[4] interessantes sobre o assunto são: "*Data Warehousing* Information Center", "*Data Warehousing* Knowledge Center", "The *Data Warehousing* Institute", além dos *sites* de Bill Inmon e Ralph Kimball.

2.1 O Setor de Telecomunicações

A história do setor de telecomunicações no Brasil não é recente sendo que os serviços de correio foram regulamentados em 1808 e a telegrafia elétrica iniciou seus serviços em 1852 no Rio de Janeiro. Durante 3 anos a rede telegráfica ficou restrita ao Rio de Janeiro e só em 1873 atingiu a região Nordeste, mais especificamente Recife e Maceió. Em 1874 um cabo submarino fez a ligação entre Rio, Bahia, Pernambuco e Pará. Ainda nesse mesmo ano ocorreu a ligação com Portugal (Brandão, 1996).

Graham Bell apresentou seu invento, o telefone, em uma exposição na Filadélfia em 1876 com a presença do então imperador do Brasil D. Pedro II. Apenas um ano depois foi instalado o primeiro telefone no país. Nesse período inicial os telefones eram instalados apenas para a

[4] Ver uma relação no Apêndice 2

comunicação entre repartições públicas, órgãos militares e corpo de bombeiros. O avanço do comércio no Rio de Janeiro levou o telefone para as lojas e armazéns no porto. No início do século muitas empresas estrangeiras instalaram escritórios no Brasil sendo que em 1907 o país já contava com 15.208 empresas, sendo 5.399 em São Paulo e 3.365 no Rio de Janeiro (Brandão, 1996). Em 1930 começa a exploração dos serviços de radiotelefonia internacional e em 1931 o governo estabelece um quadro regulatório[5] mínimo para os serviços telegráficos e telefônicos. É criada a Empresa de Correios e Telégrafos.

Segundo Siqueira (1993), a evolução do setor de telecomunicações no Brasil, até o início dos anos 90, poderia ser dividida em cinco etapas distintas de desenvolvimento:

- Estagnação (1946-1962)
- Reorganização (1962-1967)
- Decolagem (1967-1975)
- Turbulência (1975-1985)
- Crise (1985-1993)

A etapa da Estagnação se estendeu dos anos 40 até 1962, e nesse período não houve uma política para o setor. Mais de 900 companhias telefônicas e serviços municipais operavam no território nacional no final desse período com grandes dificuldades operacionais e de interligação. As concessões dos serviços de telecomunicações eram distribuídas pelos governos federal, estaduais e municipais o que gerava um crescimento desordenado, oneroso e sem compromisso de qualidade no setor (Telebrás, 2001). Em 1961 criou-se o CONTEL – Conselho Nacional de Telecomunicações.

Com a Lei 4.117 (CBT - Código Brasileiro de Telecomunicações) de 27 de agosto de 1962 teve início a etapa de Reorganização, que foi até 1967. No início desse período o Brasil tinha cerca de 1,2 milhões de telefones para uma população de mais de 75 milhões de habitantes. Até 1964 a densidade de telefones no Brasil era de pouco mais de 1 telefone para cada 100 habitantes.

A Embratel, operadora de serviços de longa distância, nasceu em 16 de setembro de 1965 com base nessa lei e tinha como principal missão interligar o território nacional e viabilizar a comunicação internacional automática (Paste, 2000). O Ministério das Comunicações foi criado em 1967 como o órgão do poder Executivo Federal encarregado da elaboração e do cumprimento das políticas públicas do setor de comunicações. Suas atividades abrangem três áreas fundamentais: radiodifusão, serviços postais e telecomunicações. O Contel e a Embratel foram logo vinculados a esse ministério.

[5] Pessini e Maciel (in Coutinho et al., 1995) definem "regulamentação" e "regulação" para mostrar as diferenças existentes entre os dois conceitos. "Regulamentação" é a ação exercida pelo Estado sobre um mercado em que o operador é único (monopólio) ou que os eventuais concorrentes não tenham condições de influenciar o mercado. O Estado tende a emitir normas e regulamentos e a intervir sobre a forma de se realizar os serviços. A "regulação", ao contrário, pressupõe a existência de mais de um operador (públicos ou não) que, em conjunto, influenciam o mercado, de modo que o Estado, como poder concedente, deverá ter uma de ação reguladora sobre a concorrência. Nesse caso, o Estado atua por meio de diretrizes e metas de caráter mais estratégico e, ao controlar os objetivos, permite maior liberdade quanto à forma de realizá-los.

Brandão (1996) diz que essas iniciativas criaram uma centralização administrativa fundamental, visto que os órgãos governamentais responsáveis pela coordenação das atividades de telecomunicações estavam pulverizados, o que dispersava as ações no setor. Com a aprovação do CBT o governo passou a monopolizar a concessão dos serviços e centralizou a operação do "Sistema de Telecomunicações". Em termos internacionais, foi um momento de transição e descontinuidade tecnológica em que os sistemas digitais começaram a substituir os analógicos. Nessa época, os serviços telefônicos estavam concentrados na região centro-leste do País, onde estavam mais de 60% dos terminais que eram explorados pela CTB (Companhia Telefônica Brasileira), empresa de capital canadense (Anatel, 2001).

A fase da Decolagem (1967-1975), ou ciclo expansivo como disse Brandão (1996), começou em 1967 e durou até a metade dos anos 70 quando a Telebrás deu início a um plano de dois milhões de linhas telefônicas. A Telebrás – Telecomunicações Brasileiras S.A. – nasceu em 1972, através da Lei nº 5.792, de 11 de junho, e era vinculada ao ministério das comunicações. Sua missão era de prestar serviços de telecomunicações com qualidade, diversidade e quantidade exigidas pelos mercado, contribuindo para o desenvolvimento político, econômico e social do País. Siqueira (1993) diz que o monopólio foi o primeiro modelo de organização das telecomunicações em todo o mundo e, até o início dessa fase, a maioria dos países adotava o regime do monopólio, seja estatal ou privado. Segundo Straubhaar (in Siqueira, 1993) na década de 70 o monopólio estatal tinha três características:

- Fixavam os preços de forma monopolística;
- Eram lentos no desenvolvimento, no lançamento e na implementação dos serviços (até mesmo na instalação de linhas telefônicas);
- Impunham restrições radicais até na permissão dos tipos de equipamentos que os clientes podiam conectar ao sistema.

Após sua criação, a Telebrás iniciou um processo de aquisição e absorção de empresas que prestavam serviços telefônicos no Brasil com a intenção de consolidá-las em empresas de âmbito estadual (Anatel, 2001). Em 1974, através do Decreto nº 74.379, a Telebrás foi designada "concessionária geral" para exploração dos serviços públicos de telecomunicações em todo o território nacional. Foi um período de inusitada expansão e extraordinárias transformações (Brandão, 1996; Coutinho et al., 1995). A TELESP (São Paulo) e a TELERJ (Rio de Janeiro) possuíam 524 mil e 451 mil telefones respectivamente, dos 1,5 milhões existentes no país nesse momento. Existiam cerca de 1.000 empresas, das quais 222 foram absorvidas (114 em 1973 e 108 em 1974), restando 778 empresas que representavam 5% do total de telefones instalados no país.

A fase da Turbulência foi a mais dinâmica e conturbada e se estendeu de 1975 até o fim do regime militar em 1985 (Siqueira, 1993). Apesar do crescimento contínuo (DDD, cabos submarinos, comunicações via satélite) começaram a surgir problemas com o monopólio. A prioridade dos investimentos foi reduzida e as tarifas locais eram corrigidas abaixo dos índices de inflação, o que representava um sério problema para as operadoras. O resultado foi o congestionamento dos serviços, a demanda represada, a ociosidade industrial e o retrocesso tecnológico. Mesmo assim consolidou-se o processo de incorporação das companhias telefônicas, ficando a Telebrás responsável pela operação de mais de 95% dos terminais telefônicos.

Durante a Crise, que foi de 1985 até meados dos anos 90, ocorreram problemas industriais e tecnológicos, elevação na demanda represada, operadoras estatais menos profissionais, politização das diretorias da Telebrás e suas subsidiárias e defasagem dos novos serviços (Siqueira, 1993). Apesar da carência de infra-estrutura de suporte ao desenvolvimento, inicia-se, a partir de 1985, a implantação de terminais digitais com os primeiros 6.120 sendo instalados em São Paulo. Em 1987 a taxa de digitalização do sistema era de apenas 2,9%. A partir dessa data foram estabelecidas diretrizes para que todas as centrais de comutação contratadas fossem digitais. A digitalização permitiu integrar diversos serviços numa mesma rede, viabilizando a utilização de um mesmo canal para voz, dados e imagens além de multiplicar a velocidade de transmissão e fidelidade dos sinais.

As grandes marcas da década de 80, segundo Brandão (1996) foram: atrasos na expansão da rede, troca e implantação de equipamentos, descontinuidades na extensão da digitalização e outros que degradaram e encareceram os serviços do sistema.

A Constituição de 5 de outubro de 1988, em seu artigo 21, inciso XI, ratificou a posição do governo sobre os serviços de telecomunicações no país:

"*Art. 21 – Compete à União: (...)*
XI – Explorar, diretamente ou mediante concessão a empresas sob controle de capital estatal, os serviços telefônicos, telegráficos, de transmissão de dados e demais serviços públicos de telecomunicações, assegurada a prestação de serviços de informações por entidades de direito privado através da rede pública de telecomunicações explorada pela União".

Brandão (1996) afirma que, apesar da recuperação dos investimentos na primeira metade dos anos 90, a demanda represada e a deterioração do tráfego, devido a congestionamento e baixa qualidade, ainda eram grandes. Somente 23% dos domicílios urbanos e cerca de 50% dos estabelecimentos de negócios possuíam telefone. Em 1993 a densidade era de 7 linhas telefônicas para cada 100 habitantes, 10^{o} lugar na América Latina e 42^{o} no mundo.

Apesar do Brasil continuar com o monopólio, Straubhaar (in Siqueira, 1993) diz que muitos outros países passaram a ingressar em processos conhecidos, dependendo das características, pelos nomes de *liberalização, desregulamentação e privatização*.

O papel das telecomunicações para a economia e a sociedade é um assunto onde não há dúvidas e isso reforça a necessidade de investimentos crescentes na expansão e na modernização do sistema brasileiro. As telecomunicações, como fator de integração regional e social, além de elemento de competitividade, precisam de adequação e atualização de forma contínua, frente às transformações organizacionais e tecnológicas que acontecem nos países industrializados (Brandão, 1996).

Durante os anos 90 a Telebrás esteve voltada para a retomada do crescimento e busca da qualidade na prestação dos serviços. Nessa época alcançou a marca de 10 milhões de telefones instalados e iniciou as instalações do Sistema Móvel Celular. Durante o período de 1976 a 1996, a população brasileira aumentou em aproximadamente 50% e o PIB cresceu quase 80% enquanto que a planta telefônica cresceu mais de 400% (Anatel, 2001). Nos últimos 20 anos, porém, o tráfego telefônico cresceu mais de 900% no serviço local e mais de 1.700% no serviço interurbano.

A partir de 1995, o governo do então presidente Fernando Henrique Cardoso, iniciou o processo para eliminar o monopólio estatal dos serviços de telecomunicações, o que foi feito através da Emenda Constitucional nº 8, para em seguida privatizar o sistema, o que acabou acontecendo em meados de 1998 pela LGT – Lei Geral de Telecomunicações, aprovada pelo Congresso Nacional de 16 de julho de 1997. Para facilitar o processo, o governo dividiu a Telebrás em três grandes operadoras locais, mantendo a Embratel como empresa de longa distância. As novas empresas eram a Tele Norte-Leste, a Tele Centro-Sul e a Telesp.

A LGT também criou a Anatel – Agência Nacional de Telecomunicações – para viabilizar as privatizações e depois exercer as atividades de regulamentar, outorgar e fiscalizar. A Anatel foi criada como uma autarquia especial, sendo administrativamente independente, financeiramente autônoma e não subordinada hierarquicamente a nenhum órgão do governo (Anatel, 2001). Dentro de suas principais atribuições destacam-se:

- Implementar a política nacional de telecomunicações;
- Propor o plano geral de metas para universalização dos serviços de telecomunicações;
- Atuar na defesa e proteção dos direitos do clientes (usuários dos serviços);
- Propor o Plano Geral de Outorgas;
- Estabelecer restrições, limites ou condições a grupos empresariais para obtenção e transferência de concessões, permissões e autorizações, de forma a garantir a competição e impedir a concentração econômica no mercado.
- Estabelecer a estrutura tarifária de cada modalidade de serviços prestados em regime público.

Em 1997, os telefones fixos no Brasil somavam pouco menos de 19 milhões, que representava uma densidade telefônica (quantidade de telefones para cada 100 habitantes) de 11,7%. Em 1999, o número subiu para 27,8 milhões de telefones com uma densidade de 17% e ao final de 2000 já eram cerca de 36 milhões de terminais instalados com densidade média de quase 21%. Os números praticamente dobraram com apenas um ano e meio de privatização do sistema. A Tabela 1 mostra a evolução dos telefones instalados após a privatização.

Jul/98	Jan/99	Jan/00	Fev/00	Mar/00	Abr/00	Mai/00	Jun/00	Jul/00	Ago/00	Set/00	Out/00
20,2	27,8	28,2	28,3	30,1	31,2	31,8	32,4	33,3	33,8	34,6	35,4

Tabela 1 – Acessos fixos telefônicos por mês (em milhões de linhas)
Fonte: Teletime (2001)

Até a privatização não se exigia das empresas compromissos relativos a expansão e universalização dos serviços (Paste, 2000). A metas eram fixadas de acordo com interesses e necessidades próprios. Não havia cobranças nem sansões por parte da Telebrás que era a "dona" do sistema. Antes do processo se efetivar, a Anatel assinou um contrato de concessão com todas as operadoras que apresentava metas de universalização e metas de qualidade que deveriam ser cumpridas a partir de 2000. Como incentivo ao cumprimento antecipados das metas, foi proposto o Plano Geral de Metas que liberaria novas outorgas para as empresas que anteciparem as metas do final de 2003 para o final de 2001. Além do cumprimentos das metas, as infrações cometidas pelas operadoras atualmente, pode gerar multas de até R$ 50 milhões.

A Anatel determinou 35 metas de qualidade para a telefonia local:

- 15 metas dizem respeito às ligações em si, divididas em grupos de três para a análise da qualidade em cada um dos períodos de maior movimento: matutino, vespertino e noturno. Essas metas envolvem, por exemplo, a obtenção de sinal de discar em até 3 segundos em 98% dos casos e completar 60% das chamadas originadas no serviço local.
- 4 metas dizem respeito ao atendimento de solicitação de reparos (ex: ter no máximo três solicitações de reparo por cem acessos instalados).
- 3 metas avaliam a taxa de atendimento às solicitações de mudança de endereços (ex: em três dias úteis em 95% dos casos para os clientes residenciais).
- 3 metas, uma para cada período do dia, determinam o atendimento ao cliente por telefone em até dez segundos em 98% dos casos.
- 2 metas avaliam o atendimento aos defeitos de telefones de uso público.
- 3 metas avaliam o atendimento ao cliente.
- 4 metas avaliam os erros na emissão das contas telefônicas.
- Finalmente a digitalização da rede local deve atingir no mínimo 75%.

Cumprindo as metas propostas pela Anatel, a empresa que opera na região Norte-Leste, por exemplo, poderá atuar também na região Centro-Sul ou mesmo em São Paulo. Isso vai aumentar muito a competição no setor e com certeza alterará o atual funcionamento do sistema. Atualmente o lucro das empresas é determinado pelo valor da tarifa, mas no novo cenário ele deverá ser resultado da escala da operação, ou seja, terá lucro a empresa que conseguir conquistar novos clientes, manter os atuais e fazer com que ambos utilizem intensivamente suas linhas (Colombini, 2001).

Até abril de 1997, era necessário dispor de R$ 1.117,63 para entrar em um Plano de Expansão de linhas telefônicas, mas nem sempre o compromisso firmado era cumprido corretamente pelas operadoras (Paste, 2000). Pelas novas regras, a habilitação de um telefone fixo passou a custar, na maioria dos estados, cerca de R$ 50,00.

Para 2000 havia a previsão para um total de 36 milhões de linhas instaladas em todo o Brasil sendo que São Paulo contribui com quase 12 milhões de linhas. A previsão para 2005 é que São Paulo chegue a 18,7 milhões de linhas. A Tabela 2 mostra a evolução das linhas instaladas no país de 1994 até a previsão para 2005.

1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
13,3	14,6	16,5	18,8	22,3	27,8	36,0	40,1	45,1	49,6	53,8	58,0

Tabela 2 – Acessos fixos telefônicos por ano (em milhões de linhas)
Fonte: Paste (2000)

Com relação a densidade telefônica média, havia a previsão de 20,9% para 2000. A densidade telefônica em São Paulo chegou a 32,2%, só perdendo para o Distrito Federal que apresentou 41,5%. A Tabela 3 mostra a evolução e a previsão da densidade telefônica no Brasil de 1994 até 2005.

1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
8,6	9,3	10,4	11,7	13,6	16,8	20,9	23,9	26,3	28,5	30,6	32,6

Tabela 3 – Evolução da densidade telefônica (em %)
Fonte: Paste (2000)

Segundo dados da Anatel, em setembro de 2000, os telefones fixos no Brasil estavam presente em 5.527 municípios. Isso quer dizer que a grande maioria das localidades está sendo atendida por, pelo menos, uma operadora de telefonia fixa.

2.2 Data Warehousing

A teoria de bancos de dados, bem como os bancos de dados já estão disponíveis há bastante tempo. No início os dados ficavam concentrados em um único banco de dados que atendia tanto ao processamento de transações (sistema operacional), que era predominante, e processamento analítico (sistema informacional).

Segundo Haisten (1999), a origem do *Data Warehouse* vem dos estudos do MIT (Massachusetts Institute of Technology) nos anos 70 que focavam o desenvolvimento de uma arquitetura técnica mais eficiente para sistemas de informação. Pela primeira vez foi feita uma distinção entre sistemas operacionais e aplicações analíticas e surgiu o princípio de separar esses dois tipos de processamento em projetos e armazéns de dados diferentes.

Para Ballard & Herreman (1998) e Teresko (1999), o conceito de *Data Warehousing* surgiu no início do anos 80 quando os sistemas gerenciadores de bancos de dados (SGBD) emergiram como produtos comerciais com facilidades para a computação de apoio à decisão. Teresko (1999) comenta que Bill Inmon, observou que estes repositórios de informação poderiam ser organizados em um bem corporativo que ele chamou de *Data Warehouse* e por causa disso Inmon é considerado o "pai do *Data Warehouse*". No início, o *Data Warehouse* consistia de instantâneos, ou subconjuntos dos dados operacionais que eram carregados em bancos de dados de apoio à decisão em períodos regulares que costumavam ser semanais ou mensais (Ballard & Herreman, 1998).

Haisten (1999) diz que em 1988, Barry Devlin e Paul Murphy[6], ambos da IBM, passaram a cuidar do problema da integração e introduziram o termo "armazém de informação" (*Information Warehouse*) e o definiram como "um ambiente estruturado suportando usuários finais no gerenciamento do negócio completo e suportando o departamento de TI no gerenciamento da qualidade dos dados". Nessa mesma época a IBM passou a tratar de um problema relacionado ao gerenciamento da informação que era a proliferação dos silos de informação. As empresas passaram a enfrentar a difícil tarefa de integrar dados de muitos sistemas separados (Haisten, 1999). O ambiente básico de um *Data Warehouse* está representado na Figura 2.

Para Berson (1997), os anos 90 poderiam ser lembrados como a época durante a qual as organizações começaram a reconhecer o uso estratégico dos dados como uma disciplina completamente diferente do uso operacional. Os sistemas operacionais tradicionalmente atendem requerimentos de missão crítica com processamentos *on-line* e *batch* (lotes). Por outro lado, o uso estratégico dos dados se caracteriza por pesquisas *on-line ad-hoc* e funções *batch* para apoio à decisão.

[6] Devlin, B. A. & Murphy, P. T., *An architecture for a business and information systems* – IBM Systems Journal, v.27, n.1, 1988

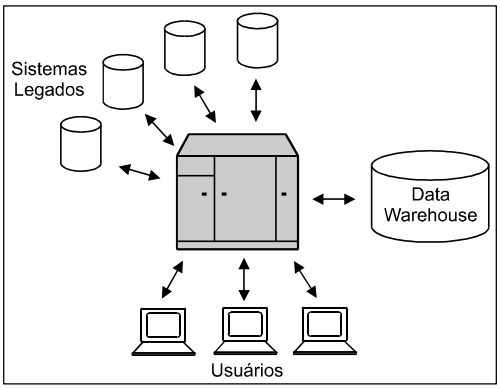


Figura 2 - O ambiente de *Data Warehouse*
Fonte: elaborado pelo autor

O *Data Warehouse* proporciona a integração dos dados na corporação, dessa forma não é preciso confrontar várias aplicações em busca de uma inteligência nos negócios, basta usar ferramentas analíticas apropriadas. Outra característica importante é sua natureza histórica, pois com informações históricas pode-se fazer uma estimativa do futuro. Por último o *Data Warehouse* oferece uma combinação de dados detalhados e sumariados (Teresko, 1999).

A definição clássica de *Data Warehouse* criada por Inmon (1997) é a seguinte :

"*Data Warehouse* é uma coleção de dados orientada por assuntos, integrada, variante no tempo, e não volátil, que tem por objetivo dar suporte aos processos de tomada de decisão".

Explicando melhor os principais termos envolvidos na definição temos :

- *Orientado por assuntos*: contém informações sobre temas específicos importantes para o negócio da empresa. Nos sistemas operacionais, de uma instituição financeira por exemplo, existem aplicações como empréstimo, poupança e cartão de crédito. O *Data Warehouse* é organizado em volta de assuntos principais tais como cliente, vendedor produto e atividade.

- *Integrado*: contém dados em estado uniforme, ou seja, existe uma consistência entre nomes, unidades das variáveis, etc. A integração, uma das mais importantes características do *Data Warehouse*, pode acontecer de várias formas: em consistentes convenções de nomes, consistentes medidas de variáveis, consistentes estruturas de dados, consistentes atributos físicos de dados, etc. A Figura 3 exemplifica essa integração:

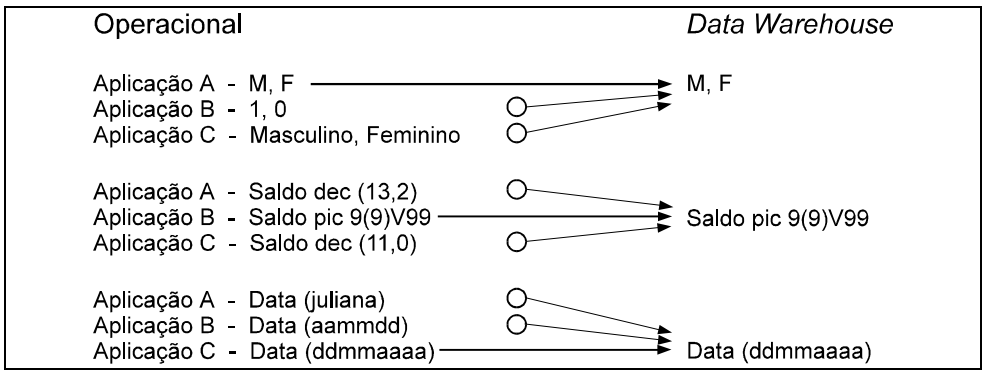


Figura 3 - Integração dos dados
Fonte: Inmon (1997); Inmon & Hackathorn (1994)

- *Variante no tempo*: contém dados não atualizáveis que se referem a algum momento específico. Os dados no *Data Warehouse* devem ser exatos em algum momento do tempo. Esta é uma característica básica diferente dos sistemas operacionais onde os dados devem ser exatos no momento do acesso (Inmon 1997; Inmon & Hackathorn, 1994).

- *Não volátil*: permite apenas a carga inicial dos dados e consultas a estes dados. Essa é outra característica que difere o *Data Warehouse* dos sistemas operacionais que permitem inclusões, exclusões e atualizações na base de registro-a-registro em tempo real (Inmon 1997; Inmon & Hackathorn, 1994).

Para Gray (1997) a definição criada por Inmon é bastante útil pois define o *Data Warehouse* em termos de atributos mensuráveis.

2.2.1 Estrutura dos dados

Já foi mencionado que os dados são a base para o *Data Warehouse* e por isso merecem uma atenção especial; mas o que seriam esses dados ? Isoladamente os dados não significam muita coisa pois o que interessa realmente é a informação. Dessa forma é necessário fazer uma distinção entre dados e informação. Conforme Tronchin (1998) *dados* são os componentes básicos a partir dos quais a informação é criada. *Informação* são dados inseridos em um contexto. *Contexto* é a situação que está sendo analisada. A partir da informação vem o *conhecimento*, que permite tomar decisões adequadas, trazendo vantagem competitiva.

Além de ser um repositório de dados "crus", o *Data Warehouse* é também um repositório de dados sumariados, ou agregados, em formato simplificado dos sistemas operacionais (Gray, 1997). Além disso existem diferentes níveis de idade dos dados (Singh, 1997). Basicamente, um *Data Warehouse*, é composto por cinco tipos de dados (Gray, 1997; Inmon & Hackathorn, 1994) que são esquematizados na Figura 4.

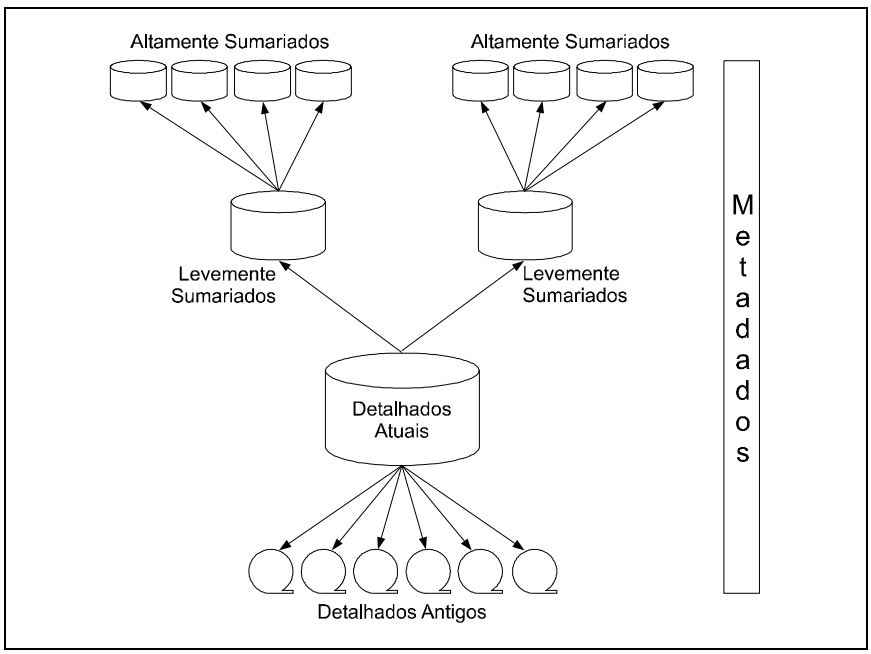


Figura 4 - Estrutura dos dados no *Data Warehouse*
Fonte: Inmon & Hackathorn (1994)

Os *dados atuais detalhados* refletem os acontecimentos mais recentes. Muitas vezes são uma simples replicação dos dados transacionais que foram "limpos" e carregados. Esses dados podem gerar um volume bastante grande apesar de nem todos os campos dos sistemas transacionais serem movidos para o *Data Warehouse*. Embora dito como "atual", este dado está atualizado somente até o momento em que é removido do sistema operacional.

Os *dados antigos detalhados* são os dados que já estão armazenados há um certo tempo. Não são acessados freqüentemente e são armazenados em um nível de detalhes consistente com os dados detalhados atuais. Pelo grande volume de dados e baixa quantidade de acessos, geralmente não é armazenado em disco.

Dados levemente sumariados: segundo Gray (1997), a experiência tem mostrado que a sumariação de dados em um formato que antecipe uma solicitação (por ser uma quantidade padrão), melhora o tempo de resposta e o uso do *Data Warehouse*. Esses dados são derivados de um baixo nível de detalhe encontrado no nível detalhado atual.

Os *dados altamente sumariados* incluem a habilidade para manter sumário de dados sobre longos períodos de tempo de forma que tendências possam ser estabelecidas. Armazenando esse tipo de dado, o tempo de resposta para essa informação também é melhorado. Dados altamente sumariados são compactos e facilmente acessáveis.

Os *metadados*, são basicamente definidos como "dados sobre os dados" e serão discutidos adiante em um tópico específico. Na Figura 4 fica fácil perceber a abrangência dos metadados na estrutura dos dados.

Gray (1997) diz que como qualquer produto de *software*, um *Data Warehouse* é de valor para a organização somente se é usado regularmente e o que se percebe é que quanto maior o nível de sumariação, mais o dado é usado; quanto mais sumariado é o dado, mais rápido é para recuperá-lo.

2.2.2 Processamento Operacional x Processamento Informacional

Voltando à questão da diferença entre operacional e informacional, Inmon (1997) coloca que existem várias razões para haver essa divisão:

- Os dados que atendem a necessidades operacionais são fisicamente diferentes dos dados que atendem a necessidades informacionais ou analíticas;
- A tecnologia de suporte ao processamento operacional é fundamentalmente diferente da tecnologia utilizada para prestar suporte a necessidades informacionais ou analíticas;
- A comunidade de usuários dos dados operacionais é diferente da que é atendida pelos dados informacionais ou analíticos;
- As características de processamento do ambiente operacional e do ambiente informacional são fundamentalmente diferentes.

O processamento informacional atende às necessidades dos gerentes no processo de tomada de decisão. Ele permite análises amplas com a finalidade de detectar tendências enquanto o processamento operacional está voltado para as necessidades do dia-a-dia. Se por um lado o

processamento operacional acessa alguns registros por vez o processamento informacional pode acessar milhares ou milhões de registros para fornecer um resultado. Segundo Berson (1997), o sistema informacional é capaz de responder a questões como "quais os três produtos que resultaram no maior número de chamados ao atendimento a clientes no último trimestre?".

A Tabela 4 apresenta um resumo das características de um sistema operacional e de um sistema informacional (*Data Warehouse*):

Característica	Sistema Operacional	Sistema Informacional
Tipo de dados	Detalhados	Detalhados e sumariados
Organização dos dados	Por aplicação	Por assunto
Estabilidade dos dados	Dinâmico	Estático
Qualidade dos dados	Na entrada	No processo (ETL[7])
Estrutura dos dados	Otimizados para transações	Otimizados para pesquisas complexas
Dados por transação	Poucos (dezenas)	Muitos (milhares)
Freqüência de acesso	Alta	Média para baixa
Volume de dados	Megabytes – Gigabytes	Gigabytes – Terabytes
Tipo de Informação	Atual e volátil	Histórica e não volátil
Operação	Atualização	Leitura e análise
Processamento	Dirigido à transação (OLTP[8])	Dirigido à análise (OLAP[9])
Uso	Operacional, repetitivo e estruturado	Informativo, analítico e não estruturado
Comunidade atendida	Funcional, com necessidades cotidianas. Decisões do dia-a-dia	Gerencial, com necessidades gerenciais. Decisões estratégicas. longo-prazo
Redundância	Não ocorre (normalizado)	Ocorre (desnormalizado)
Objetivo	Manutenção do negócio	Análise do negócio
Interação	Pré-definida	Pré-definida e ad hoc
Histórico	Baixo (até 3 meses)	Alto (até 10 anos)
Tempo de resposta	Até 2-3 segundos	De segundos a minutos
Atualização	Atualizado em tempo real	Atualizado periodicamente (*Batch*)
Disponibilidade	Alta	Atenuada

Tabela 4 - Sistema Operacional x Sistema Informacional
Fontes: Singh (1997); Berson (1997); Inmon (1997)

2.2.3 ODS - Operational Data Store

O ODS (*Operational Data Store*) não faz parte dos objetivos desse trabalho, mas será rapidamente explicado pois eventualmente é citado quando se fala em *Data Warehousing*. Segundo Inmon & Hackathorn (1994) o analista de SAD pode fazer o processamento informacional tanto do *Data Warehouse* como de sua variante o ODS, mas existem importantes diferenças entre os dois ambientes. O processamento informacional do *Data Warehouse* não é o mesmo do ODS.

O ODS é uma extensão do *Data Warehouse* no ambiente dos sistemas operacionais. Enquanto o *Data Warehouse* atende a comunidade gerencial, o processamento informacional do ODS se aplica ao cotidiano do campo operacional e está voltado para a comunidade de escritório que

[7] *Extraction, Transformation and Load* (Extração, Transformação e Carga)
[8] *On-Line Transaction Processing* (Processamento de Transações On-Line)
[9] *On-Line Analytical Processing* (Processamento Analítico On-Line)

toma decisões do dia-a-dia (questões detalhadas e imediatas) e operam em um nível mais alto que das aplicações operacionais. O ODS deve ser construído separado do *Data Warehouse* e não existem situações onde os dois devam ser combinados.

Enquanto o ODS possui dados atuais, dados "quase-atuais" (recentes), dados detalhados e atualizações, o *Data Warehouse* possui dados históricos, dados sumariados, dados detalhados e instantâneos não-voláteis. Normalmente o *Data Warehouse* possui muito mais dados que o ODS (Inmon & Hackathorn, 1994).

Fazendo uma definição do ODS, no mesmo estilo da definição clássica de *Data Warehouse*, Inmon (1997) diz que existem duas diferenças principais que estão sublinhadas para melhor identificação:

- Orientado a assuntos;
- Integrado;
- Não variante no tempo, ou seja, possui dados atuais ou quase-atuais para suporte à decisões operacionais detalhadas do dia-a-dia;
- Volátil, ou seja, pode ser atualizado

2.2.4 OLTP x OLAP

Na Tabela 4 vimos que os sistemas operacionais são voltados para aplicações (OLTP) enquanto os sistemas informacionais são voltados para análise (OLAP). Essas siglas muitas vezes são confundidas, até pela semelhança das letras, e inclusive o conceito de OLAP chega a ser confundido com *Data Warehouse*.

O termo OLAP foi introduzido por E. F. Codd, o pai dos bancos de dados relacionais, em um artigo da Computerworld no ano de 1993. Segundo Gray & Watson (1996), Codd chegou a conclusão que os bancos de dados relacionais para OLTP tinham chegado o máximo de suas capacidades em termos de visões de dados que eles podiam prover aos usuários.

OLAP é o acrônimo de "*On-Line Analytical Processing*" (Processamento Analítico On-Line), e ao contrário do OLTP, que é o acrônimo de "*On-Line Transaction Processing*" (Processamento de Transações *On-Line*) descreve uma classe de tecnologias que são projetadas para acessos *ad hoc* e análise de dados (Singh, 1997). Enquanto o OLTP praticamente não executa análises, o objetivo do OLAP é analisar relacionamentos complexos e buscar por padrões, tendências e condições de exceção.

Segundo Kimball (1998), um sistema OLTP pode processar milhares, ou até mesmo milhões, de transações por dia mas envolvem uma pequena quantidade de dados, um *Data Warehouse* normalmente processa poucas transações por dia mas cada transação poderá conter milhares ou até mesmo milhões de registros.

Samtani et al. (1998) fazem a separação entre *Data Warehouse* e OLAP quando dizem que o *Data Warehouse*, junto com OLAP, permite aos tomadores de decisão de negócios, abordar, analisar e entender os problemas de negócios usando uma certa criatividade. Theodoratos & Sellis (1999) reforçam essa distinção dizendo que o *Data Warehouse* armazena grandes

volumes de dados que podem ser usados pelas empresas para aplicações OLAP e sistemas de apoio à decisão, ou seja, o *Data Warehouse* é a base para o OLAP.

2.2.5 Data Warehouse *x SIG x SAD x SIE*

Se os termos OLAP, OLTP e *Data Warehouse* já podem gerar uma certa confusão, a confusão entre SIG, SAD, SIE e *Data Warehouse* tende a ser muito maior. Como já buscamos uma definição para *Data Warehouse*, vamos analisar um pouco melhor os conceitos de SIG, SAD e SIE e compará-los para melhor compreensão.

Sprague (1991), define SIG (Sistema de Informações Gerenciais, ou em inglês MIS - *Management Information System*) como um conjunto de sistemas e atividades necessárias ao gerenciamento, processamento e uso de informações como um recursos dentro da empresa.

Para Sprague (1991), a abordagem do SIG elevou o foco das atividades dos sistemas de informação, dando ênfase adicional na integração e planejamento das funções desses sistemas. Dessa forma o SIG contribuiu com um novo nível de informações que podiam atender algumas necessidades gerenciais mas ainda estava bastante voltada para fluxos de informações e arquivos de dados.

Os aplicativos podem ser classificados nas seguintes categorias, segundo Hogue (1991):
- Sistema de Processamento de Transações – SPT
- Sistema de Informações Gerenciais – SIG
- Sistema de Apoio à Decisão – SAD

Hogue (1991) concorda com Sprague quando diz que tanto SPT como SIG são direcionados a fluxos de informações estruturadas e acrescenta que servem de suporte para os processos nos níveis baixo e médio da estrutura organizacional da empresa. Uma diferença importante entre SAD e SPT/SIG é que no SAD a informação está disponível em diversos formatos e em SPT/SIG as informações são bastantes inflexíveis.

Segundo Sprague (1991), os conceitos de SAD foram desenvolvidos no início da década de 70 sob a denominação de "Sistemas de Decisões Gerenciais". Estes sistemas começaram a ser caracterizados como sistemas computacionais iterativos que ajudavam os responsáveis pela tomada de decisões a utilizar dados e modelos para resolver problemas não estruturados. Sprague (1991) diz que alguns autores, no início da década de 90, ampliaram a definição de SAD de modo a incluir qualquer sistema que pudesse contribuir para o processo decisório, o que permite que o termo possa ser atribuído a todos os tipos de processamento, menos o de transações. Os avanços da TI e seu uso no contexto de uma empresa levaram do SPT ao SAD, passando pelo SIG. Por essa perspectiva, o SAD complementa o SIG.

Mann e Watson[10] (in Hogue, 1991) dizem que SAD "é um sistema interativo que proporciona ao usuário acesso fácil a modelos decisórios e dados a fim de dar apoio a atividades de tomada de decisões semi-estruturadas ou não-estruturadas".

[10] Mann, R. & Watson, H. – *A contingency model for user involvement in SAD development* – MIS Quaterly, n.8, March/1984, p.27-38

O *Data Warehouse*, segundo Berson (1997), surgiu como um reconhecimento do valor e papel da informação e é o meio para o uso estratégico dos dados, sendo assim um *Data Warehouse* não é o mesmo que um SAD. Um *Data Warehouse* é uma plataforma que integra dados de qualidade melhorada para apoiar muitas aplicações e processos de apoio à decisão (SAD e SIE) dentro da empresa.

O termo SIE (Sistema de Informações Executivas, ou em inglês EIS - *Executive Information Systems*) foi criado no MIT no final da década de 1970 e seu conceito tem sido relacionado à área de SAD. Sistemas SAD e SIE são muito similares no sentido de fornecer informações para a tomada de decisão, entretanto, as aplicações SIE permitem maior flexibilidade para tratar os dados num estilo mais aceitável para tomadores de decisão seniores (Turban & Shaeffer, 1991).

Inmon (1997) diz que os SIE constituem uma das mais poderosas formas de computação pois permitem que o analista executivo localize problemas com precisão e detecte tendências importantes. O *Data Warehouse* atua no ambiente de SIE e torna sua tarefa mais fácil do que quando não há um alicerce de dados para operar, em outras palavras, o *Data Warehouse* proporciona o alicerce de dados de que o analista de SIE precisa para atender eficientemente ao processamento SIE.

Segundo Turban & Shaeffer (1991), conceitos como SIG, SAD e SIE são livres de contexto, ou seja, eles significam coisas diferentes para pessoas diferentes e por isso existem várias definições para esses termos. Sprague (1991) concorda com esse enfoque quando diz que as pessoas com diversas formações e provenientes de contextos diferentes têm visões diferentes do que sejam tais sistemas, dessa forma, um gerente e um cientista computacional raramente vêem as coisas da mesma maneira. A Tabela 5 faz um resumo das principais diferenças e semelhanças entre SIG, SAD e SIE.

Dimensão	SIG	SAD	SIE
Foco	Processamento de informações	Análise, apoio à decisão	Acesso ao *status*
Usuários típicos	Executivos de nível médio e baixo, às vezes sênior	Analistas, profissionais, gerentes	Executivos seniores
Objetivo	Eficiência	Eficácia	Conveniência
Aplicação	Controle de produção, projeção de vendas, análise financeira e gerenciamento de recursos humanos	Diversas áreas onde são tomadas decisões gerenciais	Análise de ambientes, avaliação de desempenho, identificação de problemas e oportunidades
Banco de dados	Da empresa	Especial	Especial
Apoio à decisão	Direto ou indireto, principalmente problemas estruturados, utilizando operações padrão, pesquisas e outros modelos	Apoio a atividades decisórias semi ou não-estruturadas, *ad hoc*	Indireto, principalmente para problemas de alto nível e não-estruturados
Tipo de informação	Relatórios sobre operações internas, fluxo estruturados, relatórios das operações internas	Informações de apoio para situações específicas	Novos itens, informações externas sobre clientes, concorrentes, e o ambiente, relatórios operações internas
Uso principal	Controle	Planejamento, organização e controle	Acompanhamento e controle

Capacidade de adaptação ao usuário	Normalmente nenhuma	Permite julgamento individual, recursos de simulação, algumas escolhas de estilos de diálogo	Adaptável ao estilo decisório do executivo, oferece várias opções de entrada
Gráficos	Desejável	Parte integrada de muitos SAD	Essencial
Facilidade conversacional	Desejável	Essencial se não forem usados intermediários	Essencial
Tratamento das informações	As informações são fornecidas a um grupo diversificado de usuários que as manipula	Informações fornecidas pelo SIE e/ou SIG são usadas como entrada para o SAD	Filtra e resume informações, acompanha dados e informações cruciais
Informações detalhadas	Inflexibilidade de relatórios, impossibilidade de obter detalhes com rapidez	Podem ser programadas no SAD	Acesso instantâneo aos detalhes de qualquer resumo
Banco de modelos	Modelos padrão disponíveis mas não gerenciáveis	Parte essencial do SAD	Pode ser acrescentado, normalmente não incluído ou limitado
Desenvolvimento	Fornecedores ou especialistas em sistemas de informação	Usuários, sozinhos ou com ajuda de especialistas em informática	Fornecedores ou especialistas em sistemas de informação
Hardware	*Mainframe*, micro ou distribuído	*Mainframe*, micro ou distribuído	Sistema distribuído
Natureza dos pacotes computacionais	Orientado por aplicativos recursos para geração de relatórios, modelos administrativos, contábeis, estatísticos e financeiros	Recursos para geração de modelos, simulação, aplicativos, e geradores de SAD	Interativa, acesso fácil a vários bancos de dados, acesso *on-line*, recursos sofisticados de DBMS, ligações complexas

Tabela 5 - Comparação entre SIG, SAD e SIE
Fonte: Turban & Shaeffer (1991)

Turban & Shaeffer (1991) dizem que os SIE têm sido freqüentemente associados a SAD e considerados por alguns como SAD para executivos mas a tabela anterior confirma que existem diferenças substanciais entre os dois sistemas. No sentido funcional, SIE e SAD são dois produtos diferentes, mas complementares.

2.2.6 Fábrica de Informações Corporativas

O conceito de Fábrica de Informações Corporativas - FIC (CIF – *Corporate Information Factory*) - segundo Imhoff (2000), define uma arquitetura lógica cujo propósito é entregar capacidade de inteligência e gerenciamento dos negócios dirigida por dados providos por operações de negócios. Essa fábrica provê uma arquitetura técnica estável e duradoura para qualquer tamanho de empresa desejando construir sistemas de apoio à decisão estratégicos e táticos. Basicamente a FIC consiste de produtores de dados e consumidores de informação.

Imhoff (2000) diz que o fluxo de dados/informação dentro da fábrica seria o seguinte: os produtores na FIC capturam e transformam os dados (Aquisição de Dados) dos sistemas operacionais provenientes de diversas fontes (Atividades e Dados Externos) e monta-os (Gerenciamento do Armazenamento) em um formato utilizável (*Data Warehouse* ou ODS) para serem utilizados pelos consumidores do negócio. Os consumidores da FIC conseguem a

informação produzida (Entrega dos Dados), manipulam-na (*Data Marts*) e a absorvem em seus próprios ambientes (Análises Estatísticas e Análises Analíticas).

A Figura 5 ilustra bem esse fluxo e resume todo o ambiente onde estão inseridos vários conceitos que foram discutidos até agora.

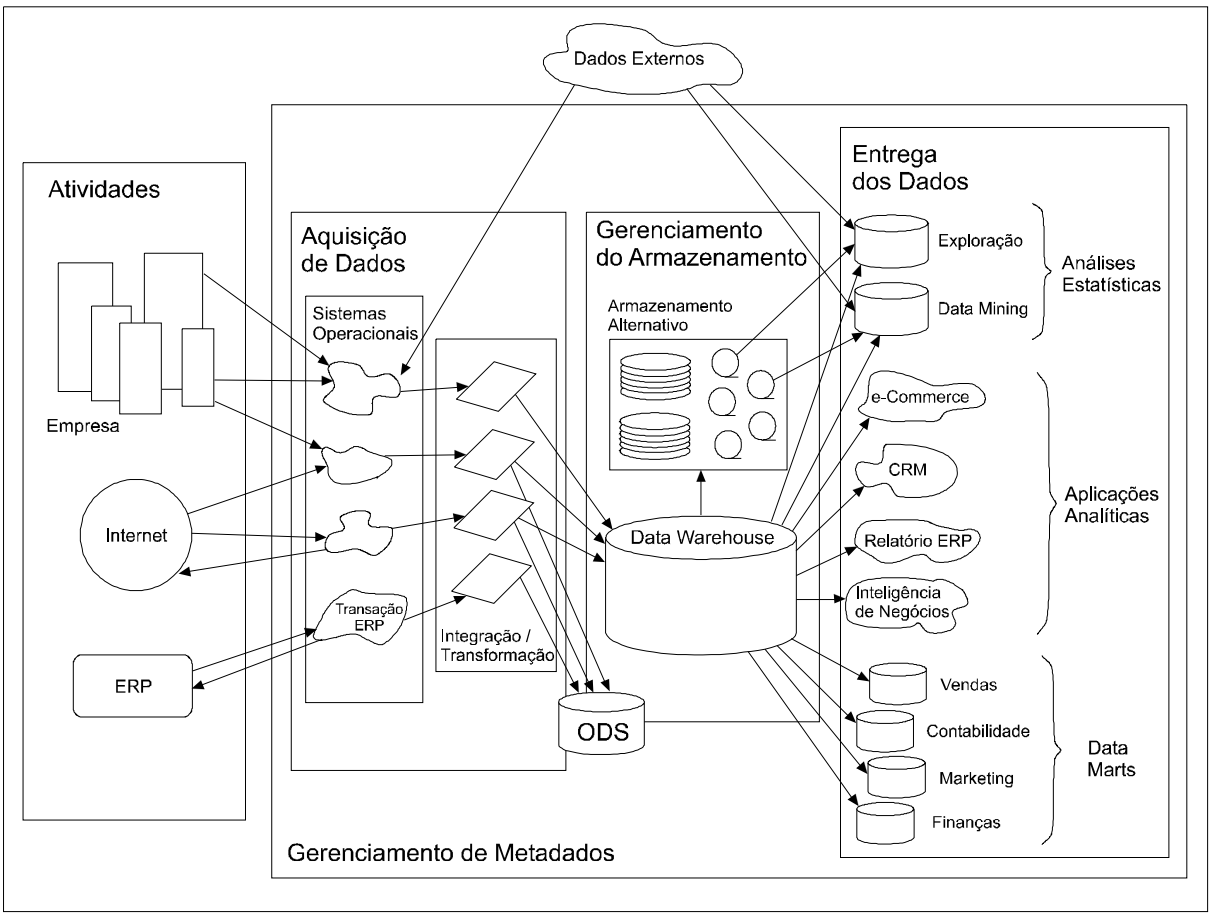


Figura 5 - Fábrica de Informações Corporativas
Fonte: http://www.billinmon.com

2.3 O Data Warehousing no Setor de Telecomunicações

A indústria de serviços de telecomunicações tem influenciado significativamente o crescimento de *Data Warehousing* nas duas últimas décadas (Peters, 2000). Segundo Schroeck (1999), *Data Warehousing* pode ajudar uma Telco a estimar o sucesso que terá um novo produto durante o seu ciclo de vida. Pode indicar como será a aceitação dessa nova oferta pelo cliente e também pode ajudar a criar cenários de preços e ganhos de receita.

De acordo com o META Group[11] (Peters, 2000), as Telcos lideram a indústria de *Data Warehousing* no tamanho médio dos *Data Warehouse*s, porcentagem de orçamento em TI gasto em *Warehousing* e o número de projetos envolvidos em *Data Warehousing*.

[11] *Data Warehousing* Trends and Opportunities, 1999

As Telcos possuem uma grande arma competitiva no mundo de hoje: dados digitalizados detalhados (Red Brick, 1998). Esses dados contém detalhes sobre clientes, chamadas telefônicas, contas, reclamações ou questões sobre produtos. Rufino (1998) acrescenta que, de maneira simples, um *Data Warehouse* para telecomunicações, consiste de um sistema de informação que armazena:

- dados do cliente, como nome, endereço, mais dados externos relacionados, como etnia, demografia, estilos de vida, etc.;
- histórico dos valores pagos (contas);
- detalhamento das chamadas: duração, origem e destino e
- produtos e serviços que o cliente possui ou possuiu.

Dessa forma, ao contrário de muitas empresas, não é problema para as Telcos conseguir dados para começar pois ela recebe diariamente milhões de registros de chamadas, etc. O grande desafio está em o que manter armazenado e por quanto tempo. O gerenciamento do grande fluxo de dados, através de muitos diferentes sistemas, também representa um desafio.

O Yankee Group[12] (Peters, 2000) identifica seis tendências que estão confrontando as Telcos atualmente: competição sem precedentes, rápida consolidação, explosão da capacidade de infra-estrutura, impacto da Internet, oportunidades de serviços da próxima geração e demanda do consumidor.

Face ao número grande de preocupações, novas pressões competitivas e oportunidades, as Telcos de hoje precisam explorar, de forma total e eficiente, cada parte da informação que possuem (Red Brick, 1998). Este é um pequeno argumento para reforçar que alguma forma de *Data Warehousing* é imperativa.

Taylor & Handen (1996) dizem que não é surpresa que as Telcos estão interessadas em construir *Data Warehouses* desde que essa tecnologia objetiva resolver um sério problema que vem afetando as Telcos: a inabilidade para extrair informações críticas para os negócios em bases de dados diferentes e conflitantes apesar da grande quantidade disponível de dados. O que é interessante sobre o fenômeno do *Data Warehouse* é que a indústria de telecomunicações tem se tornado a força pioneira nessa tecnologia.

Com o domínio das informações, que geram o conhecimento, é possível medir qual a probabilidade das pessoas mudarem de provedor de serviços, ou seja, qual a probabilidade de ocorrer *churn*[13] (Peters, 2000). Segundo Faltys (2000), nesse mercado de telecomunicações é muito importante ter condições de identificar fatores que levam a excessivas taxas de *churn* entre consumidores valiosos, como por exemplo o número de ligações não completadas ou planos de uso inapropriados.

[12] Rich, Rob, "The State of OSS: A Market in Transition" OSS2000. Orlando, January 31, 2000

[13] O termo *Churn* é derivado da palavra de origem inglesa que significa agitação, movimento. É usado para designar a falta de fidelidade dos clientes em determinado negócio (SPSS, 2000). Conforme Rufino (1999) a taxa de *Churn* é definida pela fórmula: clientes perdidos no mês / (clientes existentes no mês + mais clientes adicionados no mês)

As telecomunicações estão se movendo da era do marketing de massa para um marketing mais objetivo e dirigido, é o que diz Faltys (2000). Para ter sucesso nesse mercado altamente competitivo é preciso entender e tratar os clientes, e potenciais clientes, como indivíduos, ou seja, para conquistar uma vantagem competitiva, o serviço e o marketing devem ser individualizados. A situação mudou e os clientes podem escolher a companhia telefônica local conforme sua necessidade.

Este fato também vem se tornando realidade no Brasil após a abertura do mercado e a privatização das Telcos, que ocorreu em 1998 e deu início à competição nesse setor. A partir do momento que um consumidor pode escolher entre diferentes companhias telefônicas, o atendimento ao cliente (*customer care)* pode ser a chave para manter, ou conquistar, uma posição no mercado. Arnum (2000) diz que apesar das Telcos terem um grande banco de dados de clientes e saberem quem são esses clientes, elas não estão familiarizadas com eles. Muitos clientes sentem-se tratados como números. Para muitos deles o serviço de atendimento ao cliente é visto como um departamento de reclamações.

Os sistemas modernos de atendimento ao cliente devem resultar em atenção mais personalizada, menos tempo gasto para resolver problemas e aumento de vendas devido a promoção ativa de produtos e serviços (Arnum, 2000). Um sistema eficiente e eficaz como esse pode ajudar uma companhia a identificar quem são os clientes mais prováveis a querer um determinado serviço e ajudar um cliente a obter um conjunto de serviços otimizados. Como resultado consegue-se um maior nível de satisfação do consumidor e lealdade o que significa a retenção desse cliente e conseqüentemente uma posição competitiva mais forte. Com a opção de escolha, os clientes podem decidir mudar de companhia caso se sintam tratados como números. De qualquer forma, permanecer focalizado no cliente e seus comportamentos sem se preocupar com o que o cliente irá querer a seguir é um erro muito comum conforme Rufino (1998). Não basta conhecer cada cliente, ou mesmo vender para todo cliente, é preciso saber qual cliente comprará quais produtos, ou seja, identificar quem comprará o quê.

Conforme Rufino (1999), até pouco tempo atrás a cultura das Telcos era "conquistar o cliente a qualquer preço". A não ser pelas novas companhias que estão entrando no mercado e precisam conquistar um mínimo *market share* para continuar, a cultura atual é "reter o cliente" e "conquistar o bom cliente". Enquanto o custo para conseguir um novo cliente celular nos EUA é de US$ 400, o custo para manter um cliente é de aproximadamente US$ 60, ou seja, é importante manter o cliente e evitar o *churn*. *Data Warehousing* é uma grande ajuda nesse sentido.

Ao mesmo tempo, faz pouco sentido gastar recursos em clientes que não são rentáveis apenas para retê-los (Gomez, 1998). Isso põe uma nova ênfase no conhecimento sobre o cliente para os tomadores de decisão. Apesar de tudo, o conhecimento sobre o cliente e suas necessidades está entre as maiores vantagens que um provedor de serviços tem sobre a concorrência. Esse conhecimento permite responder a questões do tipo:

- Quem são meus clientes mais lucrativos?
- Que produtos eles estão comprando e em quais combinações?
- Quais são meus produtos mais lucrativos e por quê?
- Meus preços e serviços estão atraindo e retendo os clientes mais lucrativos?

Capítulo 3 – O Projeto

3.1 Projeto

Uma diferença entre os serviços continuados e os projetos é que os primeiros são contínuos e repetitivos enquanto que os projetos são temporários e únicos (PMI, 1996). Sendo assim, um projeto pode ser definido como um empreendimento temporário com o objetivo de criar um produto ou serviço único. Temporário pois deve ter um começo e fim bem definidos, e único pois o resultado, produto ou serviço, é alguma coisa diferente de outros produtos ou serviços semelhantes.

Para Pinto & Slevin (1988a) o projeto "é uma organização de pessoas dedicadas a um específico propósito ou objetivo, projetos geralmente envolvem tarefas grandes, caras, únicas ou de alto risco que têm que ser completadas em uma certa data, com uma certa quantidade de dinheiro, dentro de algum esperado nível de performance". Dessa forma, é importante que os projetos tenham objetivos bem definidos e recursos suficientes para executar as tarefas necessárias.

Avaliar o sucesso do projeto não faz parte do escopo desse trabalho, mas vale fazer alguns comentários sobre esse assunto, até para melhor entendimentos das características do caso que serão analisadas. Segundo Pinto & Slevin (1988a), muitos projetos terminam dentro do orçamento e cronograma e são considerados falhos, da mesma forma existem projetos que terminam atrasados e acima do orçamento e ainda assim são considerados como sucesso. Projetos "de sucesso" podem ser mal recebidos pelos usuários e serem pouco utilizados e outros que inicialmente foram percebidos como falhos podem ser tornar um sucesso com o passar do tempo.

Normalmente o sucesso de um projeto é avaliado por ele estar dentro, ou próximo, do orçamento e cronograma com um nível aceitável de performance. É fácil chegar a uma conclusão dessa forma pois são variáveis fáceis de quantificar e ficam dentro do domínio da organização do projeto (Pinto & Slevin, 1988a).

Baker et al. (1988) realizaram um estudo que constatou a importância de incluir a satisfação do cliente na avaliação de sucesso do projeto. Eles chegaram a conclusão de que o sucesso de um projeto é algo mais complexo que apenas estar dentro do custo e do cronograma. É muito importante que aqueles que estão ligados, ou são afetados, pelo projeto estejam satisfeitos. Segundo Pinto & Slevin (1988a) muitas organizações de projetos começaram a incluir a variável de satisfação do cliente para avaliação de sucesso do projeto. Por ser um conceito difícil de ser quantificado é freqüentemente medido por substitutos, como por exemplo, o número de reclamações. Os autores sugerem que o sucesso de um projeto tem dois componentes principais: as questões referentes ao projeto propriamente dito e as questões referentes ao cliente, ou usuário (Figura 6).

As áreas chave para o sucesso do projeto que estão associadas ao cliente seriam (Pinto & Slevin, 1988a):

- Uso – o projeto será/está sendo usado pelos seus usuários. Os usuários importantes, diretamente afetados pelo projeto, farão uso dele;

- Satisfação – o cliente está satisfeito com o processo pelo qual este projeto está sendo/foi completado;
- Efetividade – este projeto beneficiará diretamente os usuários, melhorando a eficiência ou efetividade do empregado.

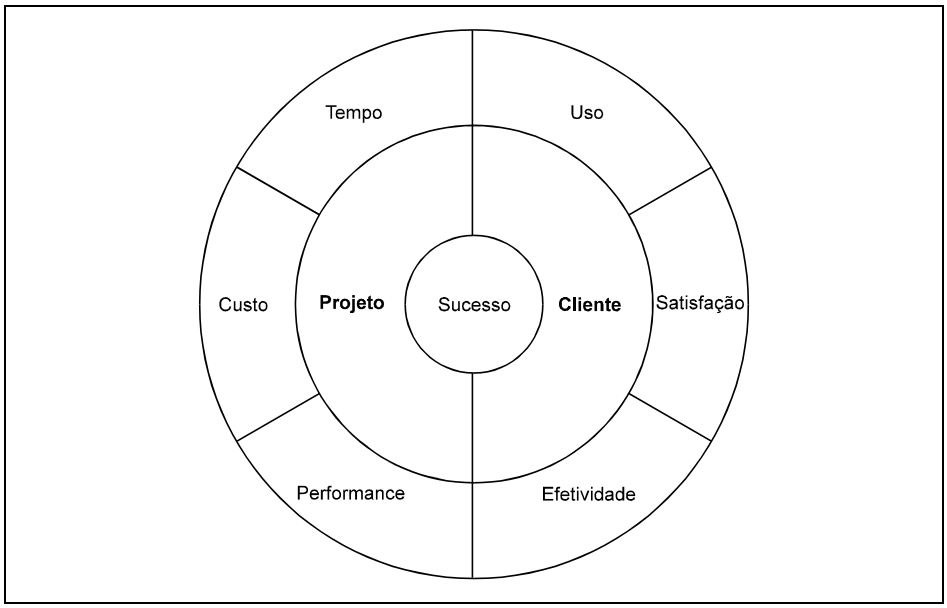


Figura 6 - Modelo de sucesso de projeto
Fonte: Pinto & Slevin (1988a)

Para Inmon (1997), é errado pensar que os enfoques para condução de projeto que funcionaram no passado serão úteis na construção do *Data Warehouse* pois os requisitos não podem ser conhecidos até que ele esteja parcialmente povoado e sendo usado, ou seja, espera-se até que o usuário diga "agora que vejo o que pode ser feito, posso dizer-lhe o que realmente seria útil". Por outro lado é errado pensar em não prever requisitos. Sendo assim a realidade estaria em algum ponto intermediário. O autor continua dizendo que de uma certa forma "projeto" não é uma descrição exata do que acontece durante a construção do *Data Warehouse*, uma vez que ele é construído de modo heurístico. O *Data Warehouse* deve ser primeiro povoado com alguns dados que serão examinados pelo analista de SAD e com base no seu *feedback* os dados são modificados e/ou outros dados são adicionados.

A implementação de um *Data Warehouse* em uma empresa, conforme Bachteal (1997), não termina na fase de implementação. Como o *Data Warehouse* cresce peça por peça e para garantir que futuras implementações tenham sucesso, é crítico aprender com os sucessos e falhas da primeira fase.

3.1.1 Motivação

Um novo projeto nasce a partir de alguma necessidade ou motivação. Essa definição pode ser a base para a determinação dos objetivos a serem alcançados. Segundo Kimball (1998) a necessidade de um *Data warehouse* poderia ser percebida pelas conversas dos gerentes de grandes companhias. Em geral as declarações dos gerentes seriam como estas:

- Nós temos montanhas de dados nessa companhia mas não podemos acessá-los;

- Nada enlouquece mais um gerente que ter duas pessoas apresentando um mesmo resultado de negócios com números diferentes;
- Nós queremos separar e analisar dados por diversas perspectivas;
- Mostre-me apenas o que é importante;
- Todos sabem que alguns dados não são muitos bons;

A falta de comprometimento no projeto pode não ser falha dos executivos, afirma Atre (1997c). Alguém tem que lhes mostrar o valor do projeto para o negócio, tal como *Data Warehouse*, o que pode não ser fácil de determinar e explicar.

Para Taylor & Handen (1996), estabelecer um *Data Warehouse* deve ser parte de cada plano de uma Telco. Para ter sucesso o projeto deve começar com um problema real e bem definido de negócio (Macdonald, 1998; Taylor & Handen, 1996). A companhia deve ter certeza de que ela faz uso apropriado da tecnologia disponível e ajusta sua metodologia de desenvolvimento para compensar os desafios técnicos e de negócios que um *Data Warehouse* apresentará.

Para Berson (1997) os problemas de negócios tratados por *Data Warehousing*, poderiam ser classificados como:

- Análise retrospectiva, que foca em assuntos de eventos passados e presentes. Por exemplo, uma organização pode decidir que requer uma análise em profundidade da performance da vendas da organização nos últimos dois anos por diferentes regiões geográficas, demográficas e tipos de produtos.
- Análise preditiva, que foca em prever certos eventos ou comportamentos baseados em informações históricas.

Segundo Singh (1997) o conceito de *Data Warehouse* vem da necessidade crescente de analisar rapidamente as informações de negócios que os tradicionais sistemas operacionais existentes não podem atender porque:

- Eles têm falta de dados históricos *on-line*;
- Os dados requeridos para análise residem em diferentes sistemas operacionais;
- A performance das pesquisas é extremamente pobre o que acaba impactando a performance dos sistemas operacionais;
- Os projetos de gerenciadores de bancos de dados operacionais são inadequados para o apoio à decisão

A idéia geral é dar sentido à enorme quantidade de dados, colecionados sobre os últimos anos na forma de inteligência de negócios para efetiva tomada de decisão (Chervinsk & Henderson, 1997). A necessidade de um *Data Warehouse* se torna latente quando a companhia passa a ter:

- Desejo de estar perto do consumidor;
- Produtos ou serviços baseado na informação;
- Um mercado altamente competitivo e em constante mudança;
- Dados armazenados em muitos sistemas diferentes;
- Muitos dados – possibilidade de "se afogar" neles;
- Necessidade de sistemas de apoio à decisão;
- A necessidade para pesquisas e relatórios específicos mais efetivos;

Para Man (1997) a motivação para construir um *Data warehouse* surge quando existem muitos sistemas diferentes e independentes juntamente com a necessidade de uma visão total do cliente. Para conseguir reter esses clientes será necessário acesso a mais e melhores informações.

A razão fundamental para construir um *Data Warehouse* é melhorar a qualidade da informação na organização, é o que diz Watson (1998). Ele realizou uma pesquisa em 1997 com 121 participantes de uma conferência sobre *Data Warehouse* patrocinada pelo *Data Warehousing* Institute. Quando foram solicitados a responder sobre as motivações de suas companhias para *Data Warehouse*, 38% dos respondentes indicaram que gostariam de ter melhor acesso a informação, 21% ter uma informação melhor e mais precisa, e 20% uma ter uma única fonte de dados.

De acordo com uma pesquisa do Technology Research Institute[14] (Red Brick, 1998), traçar o perfil do cliente é a prioridade número um em *Data Warehouse* para as Telcos. Isto é mais que apenas saber quem são os consumidores: é entender quais são os mais lucrativos, quais estão com mais probabilidade de *churn* e em quais é necessário trabalhar mais para reter.

Os principais assuntos para as Telcos atualmente são minimizar o *churn* e aumentar a retenção do cliente (Sequent, 2000). As empresas oferecem significantes incentivos para conquistar clientes, entretanto, custa muitas vezes mais ganhar clientes do que mantê-los pois cerca de 40% de todos os clientes de telecomunicações trocam de companhia anualmente.

Neste tópico verificamos a opinião de diversos autores sobre os fatores que motivam a construção de um *Data Warehouse* e que poderiam ser resumidos da seguinte forma – Um *Data Warehouse* se torna necessário quando:

- faltam dados históricos *on-line*;
- não existe flexibilidade para pesquisas nem sistemas de apoio à decisão;
- existe a necessidade de conhecer melhor o consumidor;
- a quantidade de dados é muito grande;
- os sistema não estão integrados;
- existe a necessidade de melhorar a qualidade das informações;
- é preciso minimizar o *churn* e aumentar a retençao de cliente e
- o mercado se torna altamente competitivo.

3.1.2 Patrocinador

A pesquisa de Watson (1998) mostrou que 48% dos *Data Warehouses* foram patrocinados pela TI e 41% tiveram um executivo sênior ou um gerente da área funcional como patrocinador do projeto, ou seja, não existe um definição clara com relação a quem deve ser o patrocinador do projeto. Esta variedade também é verdade quando se buscam os financiadores do projeto: 24% são financiados pela área de TI, 17% pela área funcional e 49% tanto pela

[14] *Data Warehousing and Decision Support Systems in Telecommunications*, Technology Research Institute, Executive Summary, October 1995, page 5

área de TI como pela área funcional. Para Watson (1998) a situação ideal é ter o patrocínio formado por TI e unidade de negócio.

Macdonald (1998) reforça a necessidade de um patrocinador quando diz que "cada implementação de *Data Warehouse* com sucesso necessita de um defensor forte". Normalmente um defensor de *Data Warehouse* tem que "carregar a bandeira" para fazê-lo um sucesso. Na maioria dos casos o defensor é um executivo sênior com grande entendimento do ambiente de TI.

Segundo Barquin et al. (1995), um projeto de *Data Warehousing* sem a correta corrente de patrocínio é como um automóvel com gasolina e óleo insuficientes, e um problema de ligação entre o volante e as rodas, ou seja, não vai alcançar seu objetivo. O autor diz que essa corrente inclui dois elementos chave acima do gerente do *Data Warehousing* (o líder do projeto). No nível mais alto estaria um patrocinador executivo com dinheiro para investir em uso efetivo da informação. O outro seria o "dirigente" do projeto que seria responsável por manter o projeto e o seu planejamento no caminho certo e estaria numa posição entre o patrocinador e o gerente do projeto. A Tabela 6 aponta alguns problemas que podem ocorrer com a corrente de patrocínio e como isso pode afetar o projeto.

Problema com a corrente de patrocínio	Preço pago pelo projeto
O patrocinador é um executivo de TI ao invés de um executivo de negócios fora da área de TI	O projeto é visto como um experimento tecnológico ao invés de um investimento estratégico no negócio
O patrocinador tem um orçamento limitado	Cada desafio técnico não esperado é uma crise. Como o dinheiro não é fácil de obter, o projeto passa a ser visto como problemático e devorador de orçamento
Não há o dirigente. O gerente do *Data Warehouse* se reporta diretamente ao patrocinador	Ninguém no projeto tem autoridade para amenizar a competição entre as definições de dados; o patrocinador se cansa dessa guerra e desiste de apoiar o projeto
O dirigente não conquistou o respeito dos colegas ao nível executivo	O conteúdo do *Data Warehouse* não é confiável, em parte porque ninguém pode assegurar a validade das definições usadas
O dirigente está animado ao invés de cético sobre tecnologia	O projeto é visto como um experimento técnico (ou brinquedo); as pessoas de negócios tendem a evitá-lo
O dirigente é indeciso ou se recusa atuar rapidamente	O projeto anda devagar, o apoio executivo se dissolve, o interesse desaparece, usuários buscam soluções alternativas

Tabela 6 - A falta de elementos de patrocínio
Fonte: Barquin et al. (1995)

Os principais pontos a destacar com relação ao patrocínio são:

- patrocínio formado para área de negócios e área de TI;
- patrocinador atuando como grande defensor do projeto e
- patrocinador com bom entendimento do ambiente de TI.

3.2 Aspectos Gerenciais

Por ser algo que demanda muito tempo e dinheiro, um projeto de *Data Warehousing* necessita dar atenção especial a vários fatores que podem comprometer o resultado final. Hall (1999) realizou uma pesquisa onde mais de um quarto dos respondentes disseram ter tido alguma

experiência em pelo menos um projeto falho, e somente 51% das companhias indicam que se sentem confiantes com a tecnologia de *Data Warehouse.* Bachteal (1997) afirma que o sucesso de um *Data Warehouse* vem e vai com a clareza de sua estrutura de projeto.

A gerência de projetos é a aplicação de conhecimentos, habilidades e técnicas para projetar atividades que visem atingir ou exceder as necessidades e expectativas das partes envolvidas, com relação ao projeto (PMI, 1996). Para que isso aconteça deve haver um equilíbrio entre

- Escopo, prazo, custo e qualidade;
- Diferentes necessidades e expectativas das partes envolvidas;
- Necessidades concretas e expectativas.

Segundo Adelman & Oates (1998), o gerenciamento de projetos de *Data Warehouse* trata com planos de projeto, entendimentos de escopo, recursos, cronogramas, controle de mudança, gerenciamento de risco, comunicação e ferramentas e metodologias para gerenciamento de projeto. Neste trabalho vamos analisar os itens em destaque considerando recursos como sendo recursos financeiros, ou seja, os aspectos que envolvem custos e orçamento.

Como em qualquer desenvolvimento de projeto, em um projeto de *Data Warehouse* deve haver um componente de gerenciamento do começo ao fim do projeto (Ballard & Herreman, 1998). Os autores ressaltam que esse é um componente de gerenciamento de projeto e não um componente de gerenciamento do *Data Warehouse*. A diferença é que o gerenciamento do projeto é finito no escopo e está preocupado com a construção do *Data Warehouse*, enquanto que o gerenciamento de um *Data Warehouse* é contínuo (como inventário por exemplo), e está preocupado com a execução dos processos de *Data Warehousing*.

Para Adelman & Oates (1998) o sucesso ou fracasso na implementação de um *Data Warehouse* está no bom gerenciamento do projeto. O fato de *Data Warehouses* serem diferentes dos sistemas operacionais não significa que não deve haver gerenciamento sobre cronograma, recursos, riscos, comunicação, etc.

3.2.1 Gerência do Tempo

Uma questão interessante em projetos de *Data Warehousing* se refere aos prazos. As características básicas de um projeto incluem a definição das datas de início e fim mas isso não é tão fácil em *Data Warehousing*. Bill Inmon acredita que a questão "quanto tempo se gasta para construir um *Data Warehouse*?" é na verdade duas questões. A primeira questão é "quanto tempo se gasta para construir a primeira iteração do *Data Warehouse*?" e a segunda é "quanto tempo se gasta para construir a última iteração do *Data Warehouse*?" (Baker & Baker, 1999).

A gerência do tempo de um projeto deve incluir todos os processos necessários para garantir que o projeto seja implantado no prazo previsto (PMI, 1996). Os principais processos envolvidos na gerência do tempo são:

- Definição das atividades - identificar as atividades específicas que devem ser realizadas para produzir os diversos subprodutos do projeto.

- Seqüenciamento das atividades - identificar e documentar as relações de dependência entre as atividades.
- Estimativa de duração das atividades - estimar a quantidade de períodos de trabalho que serão necessários para a implementação de cada atividade.
- Desenvolvimento do cronograma - analisar a seqüência e as durações das atividades, e os requisitos de recursos ara criar o cronograma do projeto.
- Controle do cronograma - controlar as mudanças no cronograma do projeto.

Segundo Adelman & Oates (1998), os cronogramas de um projeto de *Data Warehouse* são preparados antes do plano de projeto ter sido desenvolvido. É o plano de projeto que tem as durações, missões, tarefas predecessoras e é o único meio de determinar a duração do projeto Para o PMI (1996), desenvolver o cronograma significa determinar as datas de início e fim para as atividades do projeto. Só o planejamento e controle do cronograma permite estabelecer datas realísticas, monitorar o progresso do cronograma e atualizar as datas das atividades restantes para refletir as variações atuais previstas do inicio até o término do projeto, conforme Charoenngam & Kazi (1997).

Três meses pode ser considerado um tempo interessante para a liberação dos primeiros resultados. É um tempo suficiente para possibilitar o desenvolvimento técnico e um prazo aceitável para o usuário esperar sem passar a desacreditar no sistema (Baker & Baker, 1999).

Resumindo, pode-se dizer que em um projeto de *Data Warehouse* a gerência do tempo deve:

- garantir que o projeto seja implantado no prazo;
- preparar cronogramas antes do plano de projeto ter sido desenvolvido e
- garantir tempos de implementação que possibilitem o desenvolvimento técnico ao mesmo tempo que sejam aceitáveis para o usuário.

3.2.2 Gerência de Riscos

O gerenciamento do risco do projeto inclui o processo de identificar, analisar e responder aos riscos do projeto. Isso significa maximizar o resultado dos eventos positivos e minimizar as conseqüências dos eventos negativos (PMI, 1996). A identificação do risco consiste em determinar quais riscos provavelmente afetarão o projeto e documentar cada um deles. Essa análise deve ser conduzida durante todo o projeto considerando não só os riscos internos como os externos também. Os principais processos envolvidos na gerência de riscos são:

- Identificação dos riscos - determinar quais os riscos são mais prováveis de afetar o projeto e documentar as características de cada um.
- Quantificação dos riscos - avaliar os riscos e suas interações no sentido de avaliar possíveis conseqüências.
- Desenvolvimento das respostas aos risco - definir as melhorias necessárias para o aproveitamento de oportunidades e respostas às ameaças.
- Controle das respostas aos riscos - responder às mudanças nos riscos no decorrer do projeto.

A identificação dos riscos não é um evento pontual e deve ser realizado de forma regular ao longo do projeto (PMI, 1996). Esse processo deve abranger tanto os riscos internos quanto os externos. Os riscos internos são aqueles que a equipe do projeto pode controlar ou influenciar, tais como designação de pessoas e estimativas de custos. Os riscos externos são coisas além do controle ou influência da equipe, tais como mudanças no mercado ou ação governamental.

Conforme o PMI (1996), o risco, na sua forma literal, envolve somente a possibilidade de uma perda ou dano mas, no contexto do projeto, a identificação dos riscos diz respeito também às oportunidades e às ameaças.

Adelman & Oates (1998) dizem que cada projeto tem um grau de risco e o objetivo do gerente de projeto é reconhecer e identificar os riscos iminentes e adotar medidas para minimizá-los. A perda do patrocinador, por exemplo, é um risco sempre presente que pode ser minimizado com a identificação de pelo menos uma pessoa substituta que está interessada no projeto e poderia dispor de pessoas e orçamento bem como conduzir o gerenciamento para manter o projeto no caminho. Outro risco seria do usuário não usar o sistema. Esse problema pode ser minimizado ou eliminado tendo o usuário envolvido desde o começo em cada passo do projeto de implementação do *Data Warehouse*. Um bom projeto de banco de dados, monitoração constante e administradores de bancos de dados treinados podem minimizar o risco de baixa performance do sistema.

De forma geral, a gerência de riscos busca:

- identificar e documentar os riscos;
- desenvolver resposta aos riscos e
- ser realizada de forma regular durante o projeto.

3.2.3 Gerência dos Custos

A gerência do custo do projeto inclui os processos necessários para assegurar que o projeto seja concluído dentro do orçamento aprovado (PMI, 1996). Seus principais processos:

- Planejamento dos recursos - determinar quais recursos (pessoas, equipamentos, materiais) e que quantidade de cada recurso deve ser usada para executar as atividades do projeto.
- Estimativa dos custos - desenvolver uma estimativa dos custos dos recursos necessários à implementação das atividades do projeto.
- Orçamentação dos custos - alocar as estimativas de custos globais aos itens individuais de trabalho.
- Controle dos custos - controlar as mudanças no orçamento do projeto.

Charoenngam & Kazi (1997) dizem que completar os projetos dentro de seus orçamentos estabelecidos é um dos objetivos primários do gerenciamento. O sistema de planejamento e controle de custo sempre foi considerado como um fator crítico de sucesso, e Adelman & Oates (1998) acrescentam que é impossível desenvolver um orçamento preciso sem um bom plano como sua base.

Sem um orçamento adequado, o gerente de projeto não pode contratar pessoas, escolher os melhores produtos e ferramentas, adquirir o ***hardware*** correto e contratar consultores. Vários

fatores compõem o custo de um *Data Warehouse*: o tamanho do banco de dados; a complexidade e limpeza dos dados; o número de bancos de dados fonte e suas características; o número de usuários; a escolha das ferramentas; os requerimentos de rede; o treinamento, etc.

Adelman & Moss (2000) fazem uma observação importante quando dizem que a maioria das organizações focam no custo para implementar a aplicação inicial de *Data Warehouse* e dão pouca atenção ao seus custos de pós-implementação (manutenção e administração). Esses custos tendem a superar os custos iniciais com o passar do tempo, à medida que o *Data Warehouse* cresce em termos de tamanho, número de usuários e número de consultas/relatórios. Isso sem contar a introdução de novos *softwares* e novas versões para atender às novas necessidades bem como a atualização e crescimento dos elementos de *hardware* e rede.

De acordo com a Insigth (1998), o custo de construir um *Data Warehouse* para uma Telco pode estar entre US$ 40,000 até US$ 3 milhões. Esse valor depende muito do tamanho da Telco e do número de clientes envolvidos. A previsão global para os gastos em *Data Warehouse* pelos provedores de serviços de telecomunicações e projetos crescerá de US$ 284 milhões em 1998 para mais de US$ 2,18 bilhões em 2003 (Figura 7).

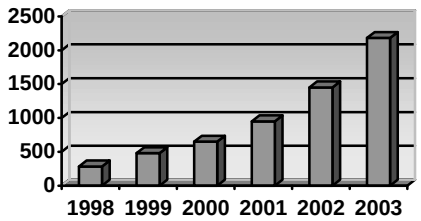


Figura 7 - Investimentos em *Data Warehouse*
Fonte: Insight (1998)

Não necessariamente a previsão e análise da perspectiva de desempenho financeiro do produto do projeto é feita no ambiente do projeto mas, quando isso ocorre, a gerência do custo do projeto inclui processos adicionais como, por exemplo, o retorno do investimento (ROI) (PMI, 1996).

Muitos artigos colocam que um ponto positivo do *Data Warehouse* se refere ao rápido retorno do investimento, no entanto esse cálculo nem sempre é uma tarefa fácil. Watson (1998) diz que calcular o ROI pode ser um desafio dada a incerteza dos benefícios, muitos deles intangíveis. Knowles (1996) e Power (1998) acrescentam que não é possível projetar um ROI pois o custo do *Data Warehouse* pode ser determinado mas os benefícios não, pois são difíceis de quantificar. Giurlani (1999) põe em dúvida a possibilidade de cálculo do ROI de qualquer tecnologia e reforça que em grande projetos, como *Data Warehouse*, a resposta pode ser não.

Power (1998) cita um estudo do IDC que, no início de 1996, mostrou um ROI médio de 401% em três anos. No entanto, diz que é preciso ter cuidado com esse tipo de pesquisa pois muitos projetos ainda não foram completados, sendo que os custos e a complexidade de um *Data Warehouse* corporativo podem virar de cabeça para baixo muitos dos esforços de um projeto piloto.

Quando decidindo por um projeto de *Data Warehouse*, os gerentes devem perguntar "quais são os resultados e benefícios esperados?" ao invés de "qual é o ROI antecipado?". Os gerentes não devem exigir um ROI positivo de um projeto de *Data Warehouse* proposto, mas eles devem exigir, e antecipar, resultados positivos.

Segundo Hackney (1999), tem havido um questionamento sobre usar ROI como uma medida para a comunidade teórica. O autor informa que, segundo Bill Inmon, "o ROI não é uma medida válida para medir o sucesso de um *Data Warehouse* pois você não pode medí-lo até que o *Data Warehouse* esteja completo e todos nós sabemos muito bem que um *Data Warehouse* nunca está 'completo' e a maioria dos negócios considera alguma forma de análise financeira como um viável e necessário critério de avaliação".

Hall (1999) realizou uma pesquisa onde somente 17% das companhias disseram que têm executado estudos de ROI em aplicações de *Data Warehouse*, e somente 13% estabeleceram métricas para apoiar o planejamento e desenvolvimento de aplicações.

Berson (1997) não vê problemas na determinação do ROI quando diz que o custo de um *Data Warehouse* pode parecer demasiadamente alto, mas os benefícios e o ROI superam o custo por uma margem muito significante. Ele cita a pesquisa do IDC em 1996 como um bom exemplo disso e coloca algumas razões para haver grandes retornos financeiros na implementação de um *Data Warehouse*:

- Habilidade para focar nos processos de negócios a executar uma análise financeira completa destes processos, permitindo às organizações tomar decisões baseadas no entendimento de todo o sistema ao invés de usar estimativas grosseiras baseadas em dados incompletos;
- Habilidade para racionalizar e automatizar o processo de construir um completo e integrado depósito de informações ao invés de desenvolver muitos sistemas individuais de apoio à decisão e a infra-estrutura correspondente;
- Os custos de *hardware*, *software*, e armazenamento relacionados ao desenvolvimento, preparação e manutenção de grandes depósitos de dados informacionais continua em declínio;
- Os benefícios do *Data Warehousing* podem ser facilmente entendidos para tomada de decisão estratégica que pode produzir grandes e tangíveis benefícios;
- A habilidade para simultaneamente entender e gerenciar as perspectivas macro e micro da organização pode economizar incontáveis horas de trabalho manual e pode ajudar a evitar caros erros que podem ser um resultado de suposições feitas sobre dados incompletos ou incorretos;

Os principais pontos apresentados nesse item foram:

- assegurar que o projeto seja concluído dentro do orçamento aprovado;
- vários fatores compõem o custo de um *Data Warehouse*;
- custos pós-implementação podem ser tão elevados quanto os custos iniciais e
- a determinação do ROI pode ser um grande desafio.

3.2.4 Gerência da Comunicação

O processo de comunicar envolve a troca de informações onde temos um emissor que é responsável por tornar a informação clara, coerente e completa permitindo que um receptor a receba corretamente. Esse receptor é responsável por garantir que a informação seja recebida de forma integral e entendida corretamente (PMI, 1996).

Na definição do PMI (1996), a gerência das comunicações do projeto inclui os processos requeridos para garantir a geração apropriada e oportuna, a coleta, a distribuição, o armazenamento e o controle básico das informações do projeto.

Adelman & Oates (1998) dizem que, provavelmente, o aspecto mais negligenciado em um projeto de *Data Warehouse* é manter todos informados. Todos inclui não somente a equipe do projeto mas também os patrocinadores e os usuários finais. Um plano de comunicação periódico, formal e informal, deve ser parte integral de cada projeto de *Data Warehouse*, principalmente quando for um grande projeto. Um plano de comunicação bem executado pode ajuda a manter rumores danosos e pressão ruim para um mínimo. É muito importante que a comunicação seja sempre honesta.

O planejamento das comunicações envolve determinar as informações e comunicações necessárias para os interessados: quem necessita de qual informação, quando necessitarão dela e como isso será fornecido para eles. A identificação das necessidades de informação e a determinação da forma de atender essas necessidade é um fator importante para o sucesso do projeto.

A participação do usuário no projeto de *Data Warehouse*, que será melhor discutida adiante, tem se mostrado um fator muito importante para o alcance dos objetivos, mas Arnold (1995) destaca que analistas e usuários, muitas vezes, encontram dificuldade em comunicar-se efetivamente e diz que a situação ideal é que eles trabalhem juntos, como uma verdadeira equipe. Todos os envolvidos no projeto devem estar preparados para enviar e receber informações na "linguagem" do projeto e devem entender como essa comunicação pode afetar o projeto como um todo.

Inmon & Hackathorn (1994) dizem que uma parte importante do projeto é fazer a comunicação aos níveis gerenciais. Os altos executivos precisam ser informados tanto dos sucessos como das falhas para decidirem onde a atenção deve ser concentrada. Quando essa comunicação não ocorre, as chances de sucesso no ambiente do *Data Warehouse* são limitadas. Não deve haver barreiras entre o andamento do projeto e a alta gerência.

Com relação à comunicação, pode-se resumir os principais aspectos como:

- manter todos informados é um fato bastante negligenciado em projetos de *Data Warehouse*;
- um plano de comunicação deve ser parte integral de cada projeto;
- todos devem estar preparados para a comunicação na linguagem do projeto e
- o trabalho em equipe facilita a comunicação.

3.3 Aspectos Tecnológicos

Segundo Watson (1998), os desafios na implementação de *Data Warehousing* incluem um complexo conjunto de componentes *hardware* e *software* com capacidades altamente especializadas, tais como multiprocessamento simétrico, bancos de dados multidimensionais e ferramentas para extração e limpeza dos dados. Vale lembrar que não é objetivo desse trabalho detalhar as características técnicas desses fatores e sim analisar sua importância para o projeto.

Para Inmon (1997), o *Data Warehouse* demanda um conjunto de características tecnológicas mais simples que seus predecessores por diversos fatores como, por exemplo, a inexistência de atualização *on-line*. No entanto, há um número considerável de aspectos tecnológicos relacionados ao *Data Warehouse.*

Para a IBM (1999), a experiência tem mostrado que em projetos onde a tecnologia tem sido percebida como uma falha, o problema normalmente não está propriamente na tecnologia, mas sim na forma que a tecnologia foi aplicada, ou seja, é aplicada para um problema de negócios errado, em escala errada, com insuficiente treinamento, com pouco ou nenhum pensamento para como o usuário necessita acessar os dados, etc.

3.3.1 Arquitetura

Singh (1997) diz que um *Data Warehouse* é uma arquitetura e não uma infra-estrutura mas o primeiro passo é definí-lo tanto em termos de arquitetura como de infra-estrutura. A arquitetura seria o conjunto de regras ou estruturas para o projeto completo de um produto ou sistema, enquanto infra-estrutura se refere a plataformas, bancos de dados, *gateways*, ferramentas e outros componentes requeridos para as funções da arquitetura. A tecnologia deve ser testada e confiável de modo a suportar não somente a infra-estrutura técnica atual mas com flexibilidade para um futuro crescimento.

Somente depois que a arquitetura do *Data Warehouse* foi definida a empresa deve começar a selecionar e implementar sua tecnologia de *Data Warehouse*. Caso contrário haverá grande probabilidade da tecnologia não suportar as necessidades da empresa, é o que diz Man (1997).

***Data Warehouse* x *Data Mart*[15]**

" … o *Data Warehouse* não é nada mais que a união de todos os *Data Mart*s ..." Ralph Kimball, December 29, 1997 (Inmon, 1999).

"Você pode pegar todos pequenos peixes do oceano e juntá-los mas ainda assim não terá uma baleia" Bill Inmon, January 8, 1998 (Inmon, 1999).

[15] Espécie de *Data Warehouse* departamental. Base de dados criada para propósitos de consultas e análises, destinada a atender segmentos específicos de usuários, como departamentos e organizações.

Inmon (1999) afirma que o assunto mais importante para a TI nos últimos anos é decidir se é melhor construir primeiro um *Data Warehouse* ou *Data Mart*. A escolha entre um único *Data Warehouse* corporativo e uma arquitetura consistindo de muitos *Data Marts* menores é um ponto de alguma controvérsia (Informix, 1998). A diferença não é necessariamente de custo, embora uma bem planejada solução corporativa possa requerer um grande investimento inicial, bem como evitar contínuas atualizações de *hardware* enquanto o *Data Warehouse* cresce.

Farias (1998) diz que, se por um lado o *Data Mart* agiliza o retorno do investimento e reduz os custos, pode também representar uma grande dor de cabeça para quem não planeja seu modelo, por confundir base de dados e qualidade das informações. Uma boa parte dos especialistas hoje defende a implementação de *Data Marts* como passo inicial, mas existe uma unanimidade em alertar o usuário que em momento algum ele pode esquecer o modelo corporativo, sob o risco de se amargar sérios prejuízos.

O problema para os usuários, segundo Farias (1998), começa com a proliferação de *Warehouses* departamentais, ou seja, os *Data Marts*. Com o bom retorno inicial as empresas tendem a criar outros, de forma desorganizada, sem integração.

A Tabela 7 traz um breve comparativo entre *Data Mart* e *Data Warehous*e a fim de deixar mais claros alguns aspectos que embasam a discussão desse tópico:

Data Mart	*Data Warehouse*
Departamental	Corporativo
Alto nível de granularidade	Baixo nível de granularidade
Pequena quantidade de dados históricos	Grande quantidade de dados históricos
Tecnologia otimizada para acesso e análise	Tecnologia otimizada para armazenamento e gerenciamento de grandes volumes de dados
Cada departamento tem uma estrutura diferente	Estrutura se adapta ao entendimento corporativo dos dados

Tabela 7 - *Data Mart* x *Data Warehouse*
Fonte: Inmon (1999)

Inmon (1999) diz que quando não se constrói um *Data Warehouse* corporativo as seguintes situações podem ocorrer :

- Redundância massiva de dados detalhados e históricos de um *Data Mart* para outro;
- Resultados inconsistentes e incompatíveis entre *Data Marts*;
- Uma interface intratável entre os *Data Mart*s e o ambiente de aplicações legadas, etc.

Em uma pesquisa conduzida pelo Atre Group, muitas companhias disseram que começaram com um *Data Mart* ao invés de *Data Warehouse* corporativo, mas menos de um quarto delas fizeram uma avaliação de riscos (Atre, 1997a). Aparentemente os usuários acreditam que *Data Marts* são um esforço de baixo risco.

Segundo Atre (1997a), o risco aumenta quando se constrói uma série de *Data Marts*, que podem ser muito diferentes. A incompatibilidade não acontece por questões de *hardware* e software, mas porque os *Data Mart*s foram criados com diferentes modelos de negócios e definições de dados. O departamento de finanças não necessariamente define o termo "cliente" da mesma forma que o departamento de marketing. Existe o perigo de que o sucesso rápido

possa conduzir a uma proliferação de *Data Marts* perpetuando problemas de "silos de informação" pela organização (Watson, 1998).

*Data Mart*s podem ser de duas formas: dependentes e independentes (Baker & Baker, 1999; Atre, 1997a). Um *Data Mart* dependente reúne suas informações específicas de assunto a partir de um *Data Warehouse* existente. Quando não existe um *Data Warehouse* corporativo e a informação específica é extraída diretamente de fontes de dados internas e/ou externas, o *Data Mart* é considerado independente. *Data Marts* dependentes são criados freqüentemente para prover mais rápido, ou mais eficiente, acesso ao repositório central de dados. Eles também oferecem uma linha de negócios mais autônoma. *Data Marts* independentes, por outro lado, rapidamente alavancam uma linha de negócios sem requerer a implementação de um *Data Warehouse* central.

Atre (1997a) verificou em sua pesquisa que a maioria dos respondentes construiu *Data Marts* únicos e independentes. Essa arquitetura parece ser mais fácil de desenvolver do que *Data Marts* dependentes que são alimentados por um *Data Warehouse* corporativo. Watson (1998) também compartilha essa visão quando diz que algumas companhias começam com um *Data Mart* independente pois é uma abordagem apropriada para um projeto voltado para aplicação que pode ser feito rapidamente, com um custo pequeno e alto retorno de investimento. Serve também como uma "prova-de-conceito" para *Data Warehouse.*

Dessa forma, ao invés de falar aos altos executivos sobre *Data Warehouse* corporativo, as pessoas freqüentemente acreditam que é mais fácil evitar as barreiras políticas e técnicas e construir *Data Marts* independentes. *Data Marts* custam menos inicialmente e podem precisar apenas de aprovação a nível de departamento (Atre, 1997c). Com a construção de um *Data Mart*, a esperança é que a companhia veja logo o seu benefício e torne-se dependente dele. Dessa forma, talvez, os executivos façam um compromisso financeiro para *Data Warehouse* e pagarão por capacidade corporativa.

Para Atre (1997c) a decisão de desenvolver um *Data Mart* ou um *Data Warehouse*, ou ainda ambos, não deve ser feita até que se faça uma avaliação das necessidades de apoio à decisão para o negócio como um todo. Mesmo que se decida por construir um *Data Mart* é preciso implementá-lo dentro de um estratégica corporativa.

As duas estratégias para implementação de *Data Warehousing* são conhecidas como "*Top Down*" e "*Bottom Up*". Na estratégia *Top Down* primeiro constrói-se o *Data Warehouse* corporativo e então faz-se a criação dos *Data Marts* dependentes a partir do *Data Warehouse* "pai". Na estratégia *Bottom Up* uma série de *Data Marts* individuais e incrementais são construídos em direção do *Data Warehouse* corporativo (Hackney, 1998; Berson, 1997). A Figura 8 mostra o desenvolvimento das duas abordagens.

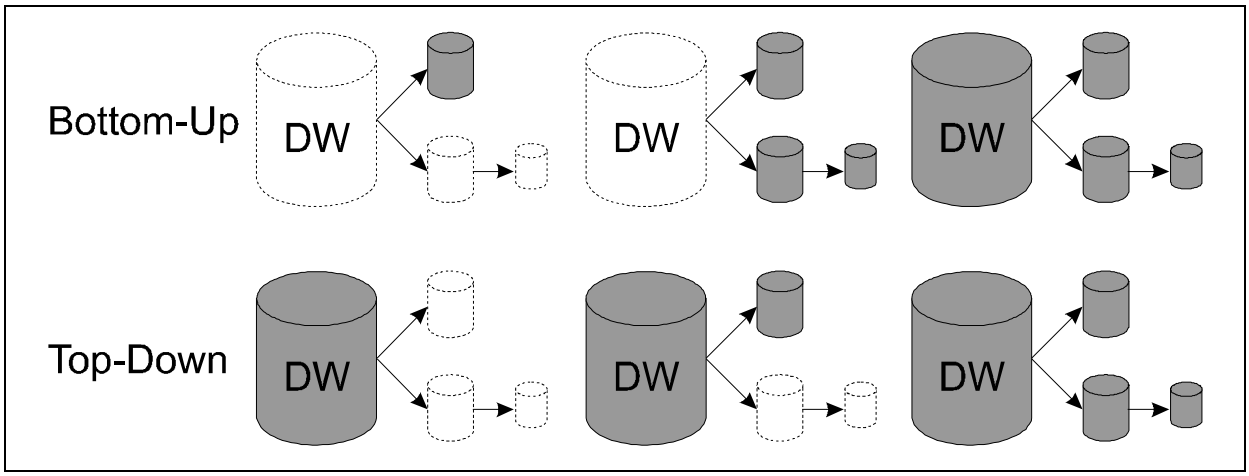


Figura 8 - Abordagens *Bottom-Up* e *Top-Down*
Fonte: Hacney (1998)

A seguir, a Tabela 8 faz uma rápida comparação entre as duas abordagens:

	Vantagens	Desvantagens
Bottom Up	• Implementação rápida • Rápido retorno do investimento • Equipe focada • Aplicações menores e dirigidas	• Desafio para obter visão corporativa • Várias equipes de gerenciamento de coordenação
Top Down	• Visão corporativa • Regras e controles centralizados • Melhores resultados a longo prazo	• Implementação muito longa • Custo elevado • Coordenação de muitos grupos diferentes na organização

Tabela 8 - *Bottom Up* e *Top Down* - Vantagens e Desvantagens
Fonte: Hackney (1998); IBM (1999)

3.3.2 Hardware

Inmon (1997) diz que uma outra importante diferença entre os ambientes operacional e de *Data Warehouse* se refere aos padrões de utilização de *hardware* que ocorrem em cada ambiente. No ambiente do *Data Warehouse* a utilização segue um padrão binário, ou seja, ou o *hardware* está sendo totalmente utilizado ou simplesmente não está e por isso qualquer cálculo envolvendo média de utilização não terá sentido. Esse fato reforça a necessidade de manter os ambientes separados pois é muito difícil tentar otimizar o equipamento para os dois tipos de processamento.

Planejamento de capacidade para *Data Warehouse* não é uma ciência exata (Man, 1997). De maneira geral, subestimar tem sido mais uma regra que uma exceção. Alguns especialistas aconselham a dobrar as estimativas iniciais dos requerimentos de *hardware* porque os usuários do *Data Warehouse* e a complexidade das pesquisas aumentam muito nos primeiros meses após a primeira implementação. Mesmo com capacidade inicial suficiente, é crítico escolher sistemas flexíveis para suportar o crescimento futuro, que é inevitável e difícil de quantificar. Macdonald (1998) diz que os assuntos de arquitetura e possibilidades em um sistema de *Data Warehouse* são enormes e a seleção da tecnologia correta é um fator crítico.

Lang (1997) relaciona algumas características que devem ser consideradas na escolha de uma plataforma tecnológica para *Data Warehousing*:

- *Velocidade.* É muito importante para a tomada de decisão em vários tipos de negócios. Esperar várias horas pelas informações e mais horas para executar uma análise é inaceitável num clima competitivo como o atual. Por causa da grande quantidade de dados mantida em um *Data Warehouse* o processamento rápido de uma pesquisa requer uma boa capacidade de processamento.

- *Escalabilidade*[16]. A quantidade de dados que é adicionada a maioria dos *Data Warehouses* cresce com grande velocidade. Esse fato cria a necessidade de grande poder de armazenamento e processamento para tratar os dados de forma eficaz. É necessário garantir que os sistemas cresçam a medida que as necessidades cresçam. Os servidores devem permitir um crescimento pela adição de capacidade processamento e memória interna. A arquitetura de armazenamento deve permitir aumentar a capacidade com baixo custo e sem degradar a performance.

- *Disponibilidade*. Grande parte dos negócios críticos possui serviços durante 24 horas e portanto é altamente desejável que os dados estejam disponíveis quando se precisar deles. Uma interrupção no sistema pode ser inaceitável nesse caso. O *Data Warehouse* deve estar seguro quanto a isso.

- *Acessibilidade*. O *Data Warehouse* pode prover uma visão operacional de um negócio no tempo e extrair seu conteúdo de muitas diferentes fontes. Eles necessitam integrar dados a partir de diferentes *hardwares*, sistemas operacionais e *softwares*.

Taylor & Handen (1996) acrescentam que para a seleção de um servidor é necessário verificar:

- A capacidade existente para ver se é possível aproveitar alguma coisa;
- Se a arquitetura é aberta e suporta uma variedade de softwares;
- Se o fornecedor provê acompanhamento e suporte adequados;
- Se o produto é reconhecido pelo mercado para evitar riscos.

No ambiente do *Data Warehouse* o tempo de resposta é bem diferente do ambiente operacional. Enquanto que no ambiente operacional o tempo de resposta é quase sempre um fator crítico, no ambiente do *Data Warehouse* o relacionamento não é tão direto e o tempo de resposta pode até ser medido em minutos e horas. De qualquer forma o tempo de resposta não deixa de ser um fator importante já que a atuação do usuário ocorre de forma iterativa (Inmon, 1997).

[16] Tradução livre para o termo em inglês "scalability" que se refere à capacidade de ajuste de configuração e tamanho.

3.3.3 Granularidade e Sumariação

Com relação às questões de armazenamento de dados é preciso considerar a granularidade que será adotada. Segundo Campos (1998), granularidade se refere ao nível de detalhe em que as unidades de dados são mantidas no *Data Warehouse*. Quanto maior o nível de detalhes, menor o nível de granularidade e quanto menor o nível de detalhes, maior o nível de granularidade. Esta é uma questão fundamental no projeto de um *Data Warehouse* porque afeta diretamente o volume de dados armazenados e ao mesmo tempo, o tipo de consulta que pode ser respondida, havendo portanto a necessidade de um acerto entre estes dois aspectos (Campos 1998; Inmon 1997; Ballard & Herreman, 1998). Inmon (1997) acrescenta que, normalmente, a empresa quer "fazer gol e defender ao mesmo tempo" porque existe uma grande demanda por eficiência no armazenamento e acesso aos dados bem como pela possibilidade de analisar os dados em maior detalhe.

Gardner (1998) chama a atenção para o fato de que enquanto sumariar dados destaca algumas condições, pode esconder outras. Ela funciona bem enquanto os negócios que foram usados para criar o sumário se mantém previsível e constante. Se essas condições mudam, os dados sumariados podem não atender às novas necessidades e, se os detalhes não estão disponíveis, a empresa pode não ter mais condições de usar esses dados.

Para Macdonald (1998), boa parte do sucesso de um *Data Warehouse* será baseada em dados históricos detalhados. No esforço de gerenciar grandes volumes de dados e a complexidade associados com navegação através de semanas, meses ou mesmo anos de dados detalhados de transação, o departamento de TI vai à direção de dados sumariados como estratégica preferida para implementação de suporte à decisão.

Inmon (1997) diz que quando um organização possui grandes quantidades de dados no *Data Warehouse*, faz sentido pensar em dois (ou mais) níveis de granularidade na parte detalhada do *Data Warehouse*. Ele coloca que essa necessidade é tamanha que poderia ser considerada padrão para grande parte das empresas.

3.3.4 Extração, Transformação e Carga

Outro fator tecnológico muito importante em um *Data Warehouse* se refere à extração, transformação e carga dos dados (ETL). Bill Inmon estimou que 80% do tempo gasto na construção de um *Data Warehouse* seria gasto em ETL (Gagnon, 1999). É um processo complexo que consome tempo mas também é a base para um *Data Warehouse* bom e funcional.

A qualidade na conversão de dados é importante para o *Data Warehouse* porque ele contém a informação que é chave para o processo de tomada de decisão. Para a empresa alcançar os objetivos do *Data Warehouse* ela deve entender quão crítico é o processo de conversão de dados. Para Hall (1999), garantir a integridade e limpeza dos dados é a tarefa mais exigente no processo de extração e transformação e Bohn (1997) acrescenta que a qualidade dos dados do *Data Warehouse* depende de se prestar muita atenção em cada atividade da conversão de dados.

A extração dos dados se faz a partir dos sistemas fonte, bancos de dados, arquivos ou ainda de fontes externas que podem ser dos mais variados formatos (Singh, 1997, Ballard & Herreman, 1998). Ferramentas de gerenciamento de dados são requeridas já na fase de extração e também serão necessárias nas fases seguintes de processo.

Um dos desafios na implementação de *Data Warehouse* é a transformação dos dados. Inconsistências de formatos podem existir em um único banco de dados e são praticamente certas quando vários bancos de dados são utilizados. Uma vez que as regras de transformação são estabelecidas, definições podem ser criadas e incluídas em rotinas de transformação. A transformação ocorre ao nível de registro ou ao nível de atributo e as definições, ou características da transformação, são armazenadas como metadados (Ballard & Herreman, 1998).

Finalmente, o processo de carga usa os dados criados no processo de transformação e transporta-os para o *Data Warehouse* de forma que possam ser utilizados. A Figura 9 procura esquematizar todo esse processo.

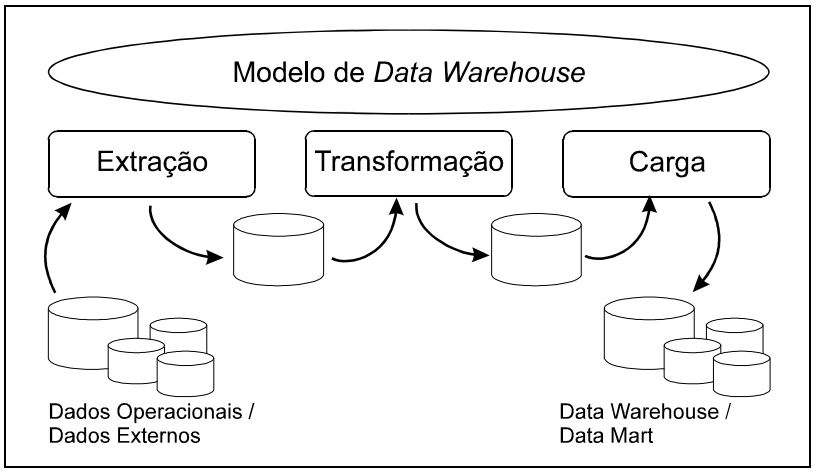


Figura 9 - O processo de Extração, Transformação e Carga
Fonte: Ballard & Herreman (1998)

Resumidamente, pode-se dizer que alguns fatores impactantes no processo de ETL são:

- qualidade na conversão dos dados;
- alto tempo gasto no processo e
- a necessidade de ferramentas de gerenciamento.

3.3.5 Metadados

Como já foi dito, os metadados são muito importantes no processo de *Data Warehousing*. De forma simples e direta, Singh (1997) diz que os metadados são o principal componente do *Data Warehouse*. A definição mais comum que se encontra na literatura sobre metadados é que eles são "dados sobre dados". De uma forma um pouco mais completa pode-se dizer que o metadado é a "descrição do dado, do ambiente onde ele reside, como ele é manipulado e para onde é distribuído". Uma outra forma, mais concisa, é definir metadado como "documentação" (Tronchin, 1998). É o metadado, na forma de definição de campo, que

informa que uma dada cadeia de bits é um endereço de cliente, parte de uma imagem fotográfica ou parte do código de um programa de computador.

O metadado no ambiente de *Data Warehouse* tem um papel bem diferente daquele que tem no ambiente operacional, segundo Inmon (1997). No ambiente operacional eles normalmente são deixados de lado e chegam ao nível de importância de uma documentação. No *Data Warehouse* os metadados são obrigatórios desde o início.

Os metadados devem isolar o usuário da complexidade de acessar informações distribuídas, enquanto facilita a atualização e sincronização de vários bancos de dados. Se isso não funcionar, os usuários voltarão a se encontrar com os problemas que o *Data Warehouse* pretendia resolver, ou seja, diferentes respostas para a mesma questão e a resultante falta de confiança na informação obtida.

Sem metadados, os dados não têm significado e localizar informações contidas em um *Data Warehouse* torna-se uma tarefa muito difícil. Os usuários de negócios serão como turistas deixados em uma nova cidade sem qualquer informação sobre essa cidade, e os administradores do *Data Warehouse* serão como os administradores da cidade que não têm idéia do tamanho dessa cidade e em que velocidade ela está crescendo. Um *Data Warehouse* sem metadados adequados é um armário-arquivo cheio de papéis, mas sem pastas ou etiquetas.

Existem algumas perguntas básicas que também devem ser respondidas pelos metadados:

- Que tabelas, atributos e chaves o *Data Warehouse* contém ?
- Qual é a origem de cada conjunto de dados ?
- Que transformação lógica foi usada na carga do dado ?
- Como o metadado tem mudado ao longo do tempo ?
- Quais *aliases*[17] existem e como eles se relacionam ?
- Quais são as referências-cruzadas entre termos técnicos e de negócios ?
- Com qual freqüência os dados são carregados ?
- Qual o volume de dados existente ?

Basicamente existem duas grandes fontes de metadados: a formal e a informal. Estas fontes compreendem os metadados técnicos e de negócios de uma organização.

As fontes de metadados formais são aquelas que já foram discutidas e documentadas. Normalmente estes metadados são armazenados em ferramentas ou documentos que são mantidos, distribuídos e reconhecidos por toda a organização. Estas fontes formais de metadados alimentam tantos os metadados técnicos como os de negócios.

Os metadados informais consistem do conhecimento corporativo, políticas e orientações que não estão em um formato padrão. Este é o tipo de informação que as pessoas "apenas sabem" e que faz parte do "conhecimento da empresa". Embora não seja formalmente documentado, este conhecimento é tão importante quanto as fontes formais de metadados.

[17] Nome alternativo que normalmente é mais fácil de memorizar.

De forma resumida, os metadados:

- são muito importantes no processo de *Data Warehouse*;
- são obrigatórios desde o início;
- devem isolar o usuário da complexidade de acessar informações e
- podem vir de fontes formais e informais.

3.3.6 Segurança

Por ser um ambiente que contém informações importantes sobre toda a empresa o *Data Warehouse* necessita de algum tipo de segurança e somente poucos usuários tem autorização para ver dados em qualquer lugar do *Data Warehouse* (Singh, 1997).

Apesar do grande esforço gasto na construção de um *Data Warehouse*, pouco se fala sobre segurança até o momento em que a implementação está concluída (Singh, 1997), mas em certas empresas a segurança é um assunto tão importante que deve ser tratado cedo, nas fases de modelagem do projeto.

Para Man (1997), a segurança do *Data Warehouse* deve incluir a segurança de acesso lógico e a segurança física dos dados mas Gagnon (1999) complementa dizendo que as considerações de segurança para *Data Warehouse* são diferentes daquelas usadas para os sistemas operacionais. O *Data Warehouse* é uma fonte de informações da empresa apenas para leitura, portanto os acessos para inclusão, atualização e exclusão são uma preocupação a menos para os administradores do ambiente. Por outro lado, existe a necessidade de proteger um bem valioso da organização contra acesso não autorizado e fazer o dado disponível para qualquer um dentro da empresa que possa fazer bom uso dele.

Segundo Singh (1997) os dados devem estar seguros, mas se controlados com muito rigor o valor do *Data Warehouse* nunca será totalmente percebido. Uma boa solução é permitir que todos na empresa tenham acesso às definições e variações, mas somente permitir acesso aos dados confidenciais com uma "necessidade de saber" comprovada.

Gagnon (1999) complementa essa idéia dizendo que para um *Data Warehouse* alcançar seus objetivos, os usuários têm que ter a habilidade para beneficiar-se dele, e portanto muitos usuários necessitarão de acesso aos dados, até mesmo mais do que tradicionalmente autorizado pela segurança no ambiente OLTP. De acordo com os especialistas, o "direito de saber" deve prevalecer sobre a "necessidade de saber", que poderia restringir o acesso ao *Data Warehouse* e minimizar sua eficiência. Se a organização não pode fazer uma mudança cultural, um *Data Warehouse* pode ser um gasto de dinheiro.

Dependendo do nível de segurança exigido o dados podem ser codificados (criptografados) e só poderão ser acessados por que tiver a chave de decodificação (Singh, 1997). Vários níveis de autorização de codificação podem ser concedidos aos usuários:

- Visão somente por usuários autorizados;
- Visão somente por usuários autorizados mas com proteção de cópia não autorizada. O usuário final ainda precisa da rotina de decodificação para vê-lo;
- Proteção de acesso indevido durante a transmissão sem a chave de decodificação.

Com relação a parte física também é preciso garantir segurança suficiente dos dados, através de ***backup***[18], armazenamento remoto, replicação, hardware tolerante a falhas, etc. para proteger os dados de perda devido a falha de energia, mau funcionamento do equipamento e sabotagem entre outros problemas que podem acontecer. O ***backup*** é importante mas ao mesmo tempo pode ser um problema para os administradores do *Data Warehouse* pois a quantidade de dados a serem salvos, geralmente, é muito grande o que pode consumir muito tempo se a solução não for otimizada.

Segundo os autores citados, a segurança no ambiente de *Data Warehouse*:

- deve ser discutida desde o início do projeto;
- é diferente daquela usada nos sistema operacionais e
- deve ser lógica (acesso aos dados) e física (***backup***, redundância, etc.).

3.3.7 Metodologia

Conforme Inmon (1997), as metodologias de desenvolvimento têm um apelo universal colocando o desenvolvedor no caminho racional, destacando o que precisa ser feito, em que ordem e em quanto tempo.

Embora a participação do usuário final seja importante, não significa que o projeto deva ser tratado exclusivamente pelo departamento do usuário final (SAS, 2000). Deve ser usada uma metodologia que se ajuste tanto à arquitetura de TI como à estratégia da empresa. Adelman & Oates (1998) acrescentam que um bom plano de projeto deve listar as tarefas críticas a serem executadas bem como os prazos de início e fim. Deve também identificar quem vai executar as tarefas, descrever aquilo que será criado e entregue e ainda identificar marcos para medir o progresso.

Inmon (1997) diz que muitas vezes as metodologias são decepcionantes e cita algumas razões para isso:

- As metodologias, normalmente, apresentam um fluxo plano e linear de atividades. Não há preocupação em rever uma ou mais atividades. No caso do *Data Warehouse* isso não faz muito sentido;
- As metodologias, de maneira geral, mostram como fazer alguma coisa e não o que realmente precisa ser feito, se atendo a muitos detalhes e casos especiais;
- As metodologias normalmente não fazem distinção entre processamento operacional e processamento informacional. Essa distinção é necessária para o sucesso da metodologia.

O *Data Warehouse* tem uma metodologia única embora, como no desenvolvimento tradicional, tenha quatro fases distintas: análise, projeto, construção e implementação (Taylor & Handen, 1996; Gardner, 1988; Ballard & Herreman, 1998). Embora tenha muitas similaridades com as metodologias tradicionais, a metodologia de *Data Warehouse* apresenta algumas importantes diferenças:

[18] Cópia dos dados que pode ser usada para restaurá-los em caso de problemas ou perda dos dados.

- Desenvolvimento iterativo, não linear;
- Ênfase no uso de protótipos alavancáveis;

Vários autores concordam com a característica de processo iterativo. Um *Data Warehouse* não pode ter uma implementação "*Big-Bang*", de uma só vez, mas deve ter projetos iterativos e incrementais (SAS, 2000). Segundo Taylor & Handen (1996), os negócios mudam muito rapidamente por isso a metodologia deve prover funcionalidade em passos iterativos. Inmon (1997) cita algumas razões que justificam a importância do desenvolvimento iterativo:

- O histórico de sucesso das aplicações sugere tal método;
- O usuário final não terá condições de expressar suas necessidades com clareza até que a primeira iteração seja feita;
- A gerência não se comprometerá por completo até que verdadeiros resultados tangíveis e claros sejam apresentados;
- Há necessidade de, rapidamente, obter resultados visíveis.

A Figura 10 tem como principal mensagem a natureza iterativa do desenvolvimento do *Data Warehouse*.

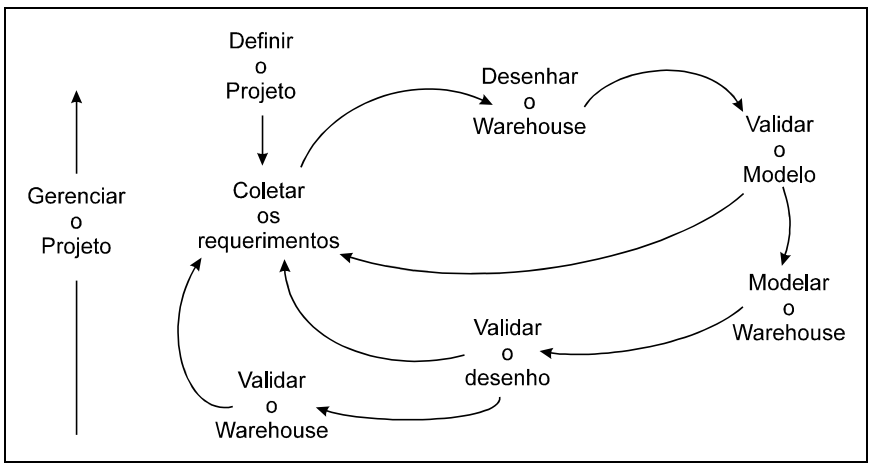


Figura 10 - Desenvolvendo o *Data Warehouse*
Fonte: Ballard & Herreman (1998)

Ballard & Herreman (1998) destacam essa característica iterativa quando dizem que todos os projetos têm algum grau de iteração mas os projetos de *Data Warehouse* levam isso ao máximo. Partes do *Data Warehouse* poderiam estar sendo entregues enquanto outras ainda estariam sendo desenvolvidas.

Taylor & Handen (1996) dizem que outra diferença na metodologia de *Data Warehousing* é o uso de protótipos alavancáveis. Eles são importantes para confirmar a interpretação dos requerimentos dos usuários. Diferente dos protótipos tradicionais, a característica de "alavancável" se refere ao fato do protótipo ir sendo revisado em cada fase do ciclo de vida do desenvolvimento até que seja colocado em produção.

A Tabela 9 mostra um resumo de metodologias propostas por alguns autores:

Autor	Fases Propostas
Chevinsky & Henderson	• Fazer plano de projeto; • Definir áreas para implementação e modelo de dados; • Identificar fontes de informação para o *Data Warehouse*; • Extrair e reformatar os dados; • Limpar e carregar os dados; • Prover ao usuário final acesso à informação e OLAP; • Distribuir os dados.
Adelman & Oates	Etapas específicas para o desenvolvimento do *Data Warehouse*: • Definir objetivos do *Data Warehouse*; • Definir bibliotecas de consulta; • Integrar metadados; • Definir estratégia de renovação (*refresh*); • Definir critério para seleção de produtos; • Preparar especificações de migração; • Desenvolver monitoramento das consultas.
Ralph Kimball	• Identificar qual processo de negócios vai ser modelado; • Definir a granularidade para determinar o nível de detalhes apresentado; • Definir as dimensões de cada tabela de fatos; • Identificar os fatos; • Analisar os atributos das dimensões; • Decidir sobre projeto físico; • Preparar dimensões para suportar evoluções; • Definir a duração do banco de dados (previsão do histórico); • Definir a freqüência com que os dados devem ser extraídos e carregados.
Alex Berson	• Juntar e analisar as necessidades de negócios. • Criar um modelo de dados e um projeto físico para o *Data Warehouse*. • Definir as fontes de dados. • Escolher a tecnologia e a plataforma de banco de dados. • Extrair os dados das fontes operacionais e executar o processo de ETL; • Escolher as ferramentas de acesso e informação do banco de dados. • Escolher o software de conectividade com o banco de dados. • Escolher o software de análise e apresentação dos dados. • Atualizar o *Data Warehouse*.
SAS Institute	Estas etapas são repetidas em ciclos: • Fase de justificação, onde se avalia o impacto do *Data Warehouse* para os negócios; • Fase de reunião dos requerimentos, onde a equipe de projeto junta as necessidades de informação dos departamentos dos usuários finais (o quê, em que formato, com que freqüência); • Fase de desenho/modelagem, onde o *Data Warehouse* lógico é desenhado, as fontes de dados são identificadas e as regras de transformação de negócios são definidas; • Fase de implementação, onde todos os programas e aplicações necessários são escritos; • Fase de revisão, onde ocorre a avaliação de todo o projeto, seus sucessos e suas implicações nos negócios da empresa.

Tabela 9 - Metodologias/Fases para o *Data Warehouse*
Fontes: Chevinsky & Henderson (1997); Adelman & Oates (1998); Kimball (1996); Berson (1997); SAS (2000)

Podemos observar que, com exceção do SAS Institute que propõe fases mais abrangentes, os autores focam em etapas que são bem específicas do *Data Warehouse* como por exemplo metadados, processo de ETL, dimensões, fatos, etc. Dimensões e fatos serão discutidos adiante no tópico sobre modelagem de dados.

Baseada nos Dados x Baseada no Usuário

Segundo Ballard & Herreman (1998), a reunião dos requerimentos para o desenvolvimento de um *Data Warehouse* pode assumir duas abordagens: baseada nos dados ou baseada no usuário.

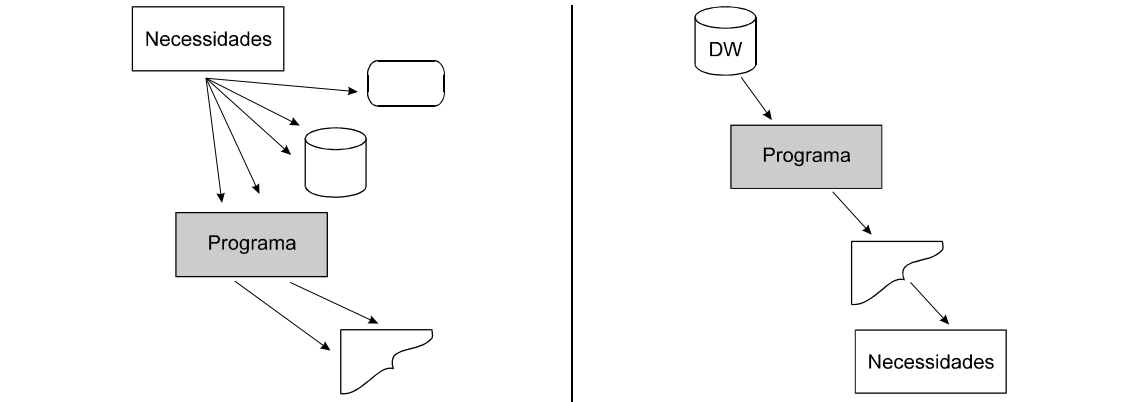
Quando a abordagem é voltada para os dados, a definição dos requerimentos é baseada no uso das fontes de dados que estão nos sistemas operacionais. Segundo Inmon (1997), existem pelo menos duas características que distinguem uma metodologia baseada nos dados:

- Ela constrói sobre código e dados que já foram criados em vez de fazer novos desenvolvimentos. Primeiro reconhece-se a redundância dos dados e depois, se os dados existem eles são aproveitados caso contrário são criados.
- Existe uma ênfase no armazenamento central de dados, ou seja, o próprio *Data Warehouse* como base para o processamento informacional.

Os benefícios desta abordagem estão no fato de poder fornecer todos os dados porque eles já foram limitados sobre o que estava disponível, ou seja, o usuário será provido com o que existe, e porque pode-se ganhar tempo nos estágios iniciais do projeto (Ballard & Herreman, 1998). Como desvantagens pode-se citar: o risco do entendimento incorreto dos requerimentos pelo menor envolvimento do usuário, o volume de dados que, se for muito grande, pode consumir muito tempo e a possibilidade dos requerimentos necessitarem de dados que não estão disponíveis no momento.

Quando a abordagem é baseada no usuário a definição dos requerimentos é feita com base nas funções executadas pelos usuários. Normalmente isso ocorre através de reuniões e entrevistas (Ballard & Herreman, 1998). A grande vantagem desta abordagem está em prover o que é necessário ao invés de prover o que está disponível. Por ter um escopo menor que a outra abordagem pode produzir um *Data Warehouse* em menor tempo. Uma aspecto negativo pode ser a expectativa do usuário que deve entender a possibilidade de alguns dados necessários não estarem disponíveis.

Segundo Inmon (1997), o ciclo de vida clássico – SDLC (*System Development Life Cycle*) – é baseado em requisitos, ou seja, primeiro é necessário entender as necessidades. O CLDS (sigla de SDLC ao contrário) é praticamente o inverso pois começa pelos dados. O CLDS é um ciclo de vida de desenvolvimento baseado em dados, ao passo que o SDLC é um ciclo de vida de desenvolvimento baseado em requisitos como mostrado na Figura 11.



SDLC Clássico	SDLC de *Data Warehouse*
• Levantamento de necessidades	• Implementar o *Data Warehouse*
• Análise	• Integrar os dados
• Projeto	• Procurar distorções
• Programação	• Programas para os dados
• Teste	• Projetar sistema SAD
• Integração	• Analisar os resultados
• Implementação	• Entender necessidades

Figura 11 - SDLC Clássico x SDLC *Data Warehouse*
Fonte: Inmon (1997)

3.3.8 Modelagem de Dados

Um modelo pode ser definido como uma abstração e reflexo do mundo real. Partindo desse ponto, um modelo de dados pode ser definido como uma representação gráfica dos dados dentro de uma área específica de interesse e tem o objetivo de transportar o significado dos dados, o relacionamento entre eles, os atributos dados e ainda gravar as definições precisas dos dados. (Singh, 1997).

Segundo Singh (1997), a modelagem de dados é uma técnica que permite estruturar dados em volta de conceitos de negócios de forma a prover a base para análises de dados sofisticadas. Ballard & Herreman (1998) dizem que, no início, a modelagem de dados para o ambiente de apoio à decisão não era de tanto interesse pois os sistemas se adequavam os modelos de dados dos sistemas operacionais pois não passavam de "instantâneos" de ambiente.

Para Kimball (1998) uma importante diferença entre o sistemas operacionais e o *Data Warehouse* está na organização dos dados, ou seja, no modelo de dados. Dessa forma os projetos de bancos de dados devem ser diferentes também pois no processamento de transações o objetivo é ter velocidade para acessar e atualizar um ou poucos registros, já no *Data Warehouse* o objetivo é acessar agregados, somas, médias, etc. (Barquin et al., 1995).

Modelo Entidade-Relacionamento

De forma simples, Kimball (1997) diz que o modelo Entidade-Relacionamento (Modelo ER ou apenas MER) é uma técnica de projeto lógico que busca remover a redundância nos dados,

o que é muito benéfico no processamento de transações. Um MER utiliza dois conceitos básicos: as *entidades* e os *relacionamentos* entre essas entidades. Se o modelo for mais detalhado ele também contém *atributos* que seriam as propriedades dessas entidades ou relacionamentos.

Um MER pode ter centenas de entidades lógicas e até milhares em sistemas de grande porte como os ERP (*Enterprise Resource Planning*). Na prática, cada uma dessas entidades se torna uma tabela física quando o banco de dados é implementado (Kimball, 1997). O MER não tem "regras de negócios", ele tem "regras de dados". A Figura 12 mostra o exemplo de um MER.

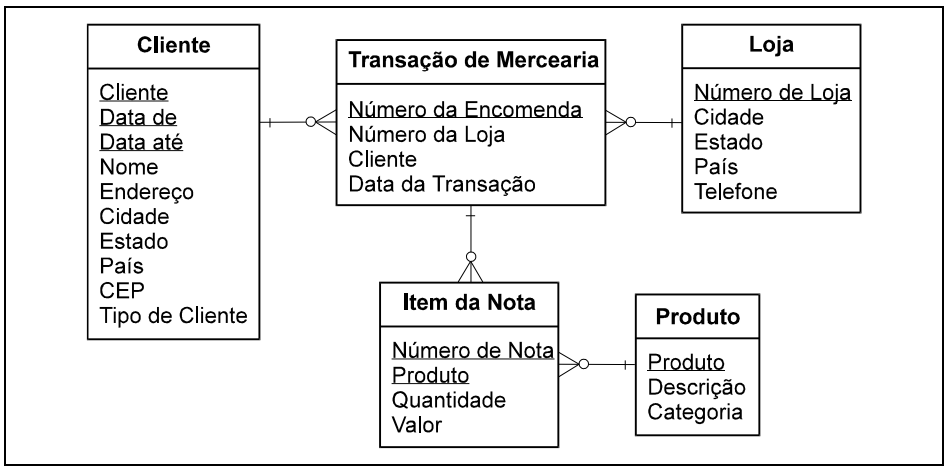


Figura 12 - Modelo Entidade-Relacionamento
Fonte: Tronchim (1998)

Modelo Dimensional/Multidimensional[19]

Segundo Kimball (1997), a modelagem dimensional "é uma técnica de projeto lógico que busca apresentar os dados em uma estrutura padrão e intuitiva que permite um acesso de alta performance". O autor acrescenta que esse é apenas um novo nome para uma técnica antiga usada para criar bancos de dados simples e compreensíveis. Ballard & Herreman (1998) dizem que a análise dimensional permite investigar um grande número de fatores envolvidos em um problema de negócios em diferentes níveis de detalhes. Três conceitos básicos estão envolvidos nessa modelagem: *fatos*, *dimensões* e *métricas* (ou atributos).

Um *fato* é uma coleção de itens de dados relacionados, consistindo de métricas e contexto de dados. De maneira geral cada fato representa um item ou transação de negócios. Os fatos são reunidos na *tabela de fatos*. Segundo Kimball (1997), as tabelas de fatos mais úteis contém fatos numéricos e aditivos que são importantes porque as aplicações de *Data Warehouse* normalmente recuperam muitos registros de uma única vez e uma tendência é somá-los para análise, o que não deixa de ser uma "compactação" que trará ganhos de performance.

A *dimensão* é uma coleção de membros ou unidades do mesmo tipo de visões (Ballard & Herreman, 1998). No diagrama, uma dimensão é usualmente representada por um eixo. Em

[19] Na literatura encontramos tanto o termo "dimensional" como "multidimensional", ou seja, não existe uma definição clara que indique uma diferença precisa entre os dois.

um modelo dimensional, cada ponto de dado em uma tabela de fato é associado com um e somente um membro de cada uma das múltiplas dimensões. As tabelas dimensionais armazenam as descrições textuais das dimensões do negócio. Kimball (1997) diz que cada tabela de dimensão tem uma chave primária que corresponde exatamente a um dos componentes da chave composta da tabela de fatos. A Figura 13 traz um exemplo simples de um modelo dimensional. Para efeito de comparação, esse exemplo utiliza o mesmo negócio apresentado na Figura 12 (Modelo ER).

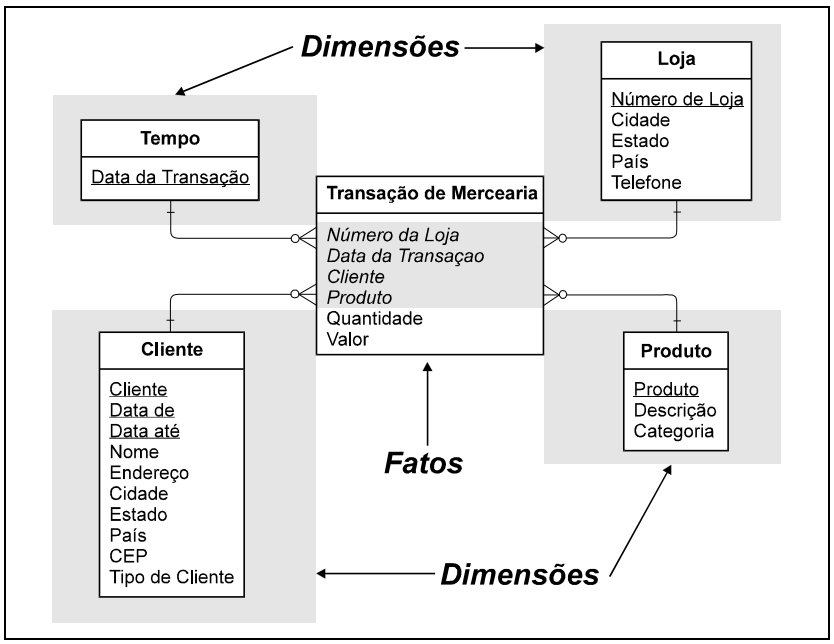


Figura 13 - Modelo Dimensional
Fonte: Tronchin (1998)

Para um melhor entendimento, a Figura 14 mostra o conteúdo de uma tabela de fatos.

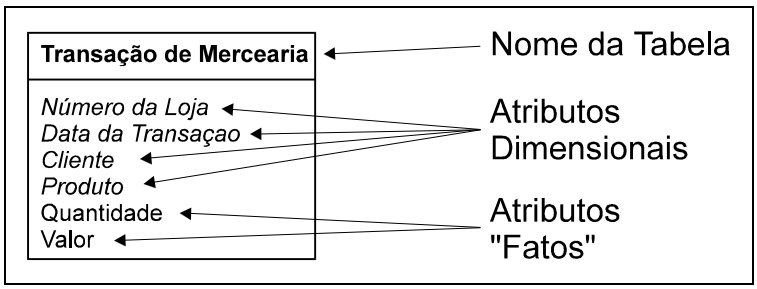


Figura 14 - A Tabela de Fatos
Fonte: Tronchin (1998)

A *métrica* pode ser definida como um atributo numérico de um fato, representando a performance ou comportamento do negócio para as dimensões. As métricas podem ser vendas em dinheiro, volumes de vendas, quantidade fornecida, etc.

Um outro nome para o modelo dimensional é *star schema* ou *star model* (modelo estrela), pois o diagrama se assemelha a uma estrela com uma tabela grande no centro cercada por tabelas auxiliares em um padrão radial (Kimball, 1998).

Singh (1997) diz que por ser muito simples, o modelo estrela apresenta algumas vantagens importantes:

- Permite que uma estrutura de dados complexa seja definida com um modelo de dados muito simples;
- Isto facilita a definição de relacionamentos hierárquicos dentro de cada dimensão, e simplifica a tarefa de criar junções através de múltiplas tabelas;
- Reduz o número de junções físicas e pesquisas que tem para processar. Isso aumenta bastante a performance.
- Pela simplificação da visão do modelo de dados, ele reduz as chances de usuários submeterem pesquisas incorretas que consomem recursos significantes e retornam informações imprecisas.
- Permite a expansão e desenvolvimento do *Data Warehouse* com baixa manutenção, ou seja, provê uma base flexível para o crescimento.

Um derivação do modelo estrela é o modelo "floco de neve" (***snowflake model***) que, segundo Ballard & Herreman (1998), é o resultado da decomposição hierárquica de uma ou mais dimensões. Tronchin (1998) simplifica a definição dizendo que o modelo "floco de neve" acrescenta estruturas hierárquicas ao modelo estrela como mostrado na Figura 15.

Quando um banco de dados pode ser visualizado como um "cubo" contendo três, quatro ou até cinco ou mais dimensões, as pessoas conseguem fatiar esse cubo e cortá-lo em cubos menores (processo conhecido como *slice and dice)* em qualquer de suas dimensões (Kimball, 1997; Tronchin, 1998). Essa metáfora de cubo transmite bem a idéia de múltiplas dimensões. Uma necessidade comum no mundo dos negócios, segundo Kimball (1998), seria: "Vendemos *produtos* em vários *mercados* e avaliamos nosso desempenho ao longo do *tempo*".

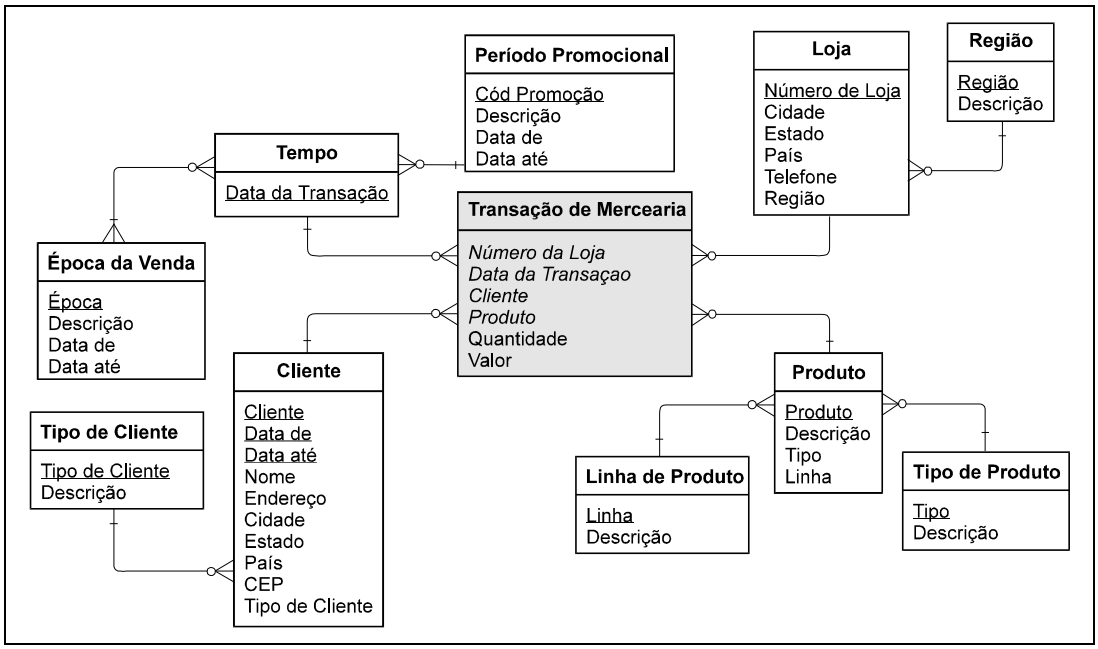


Figura 15 - Modelo "Floco de Neve"
Fonte: Tronchin (1998)

A seguir será apresentado um exemplo para ilustrar melhor esses conceitos. O dados da Tabela 10 mostram o resultado de uma análise do volume de vendas por região. Nesse caso existe uma única dimensão (Região) com a agregação de um fato relacionado (Vendas):

Região	Vendas
Sul	R$ 1.552 M
Nordeste	R$ 1.732 M

Tabela 10 - Análise com uma dimensão
Fonte: Tronchin (1998)

Pode-se aumentar o detalhamento da análise (processo conhecido como *drill-down*) adicionando-se a dimensão tempo (Trimestre), conforme mostrado na Tabela 11.

Região	Trimestre	Vendas
Sul	T1	R$ 377 M
	T2	R$ 368 M
	T3	R$ 423 M
	T4	R$ 384 M
Nordeste	T1	R$ 427 M
	T2	R$ 418 M
	T3	R$ 453 M
	T4	R$ 434 M

Tabela 11 - Análise com duas dimensões
Fonte: Tronchin (1998)

A Tabela 12 apresenta 3 dimensões (Região, Trimestre e Linha de Produtos) e um fato (Vendas). A cada nova dimensão acrescentada aumenta-se o nível de detalhe. Por exemplo, partindo de Região pode-se acrescentar Estado, Cidade, etc. e partindo de Trimestre pode-se acrescentar Mês e Dia, desde que o *Data Warehouse* mantenha esse nível de detalhe.

Região	Trimestre	Linha de Produtos	Vendas
Sul	T1	Linha Branca	R$ 250 M
		Outros	R$ 127 M
	T2	Linha Branca	R$ 225 M
		Outros	R$ 143 M
	T3	Linha Branca	R$ 275 M
		Outros	R$ 148 M
	T4	Linha Branca	R$ 253 M
		Outros	R$ 131 M
Nordeste	T1	Linha Branca	R$ 280 M
		Outros	R$ 147 M
	T2	Linha Branca	R$ 255 M
		Outros	R$ 163 M
	T3	Linha Branca	R$ 305 M
		Outros	R$ 148 M
	T4	Linha Branca	R$ 283 M
		Outros	R$ 151 M

Tabela 12 - Análise com três dimensões
Fonte: Tronchin (1998)

A Figura 16 mostra a representação dessas tabelas em formato de cubo. Se a necessidade for verificar as Vendas na Região Sul para a Linha de Produtos Branca em todo o ano o cubo é fatiado como mostrado pela hachura mais clara. Caso a necessidade seja analisar as Vendas nas Regiões Sul e Nordeste para a Linha de Produtos Outras e apenas no Quarto Trimestre tem-se o corte mostrado pela hachura mais escura.

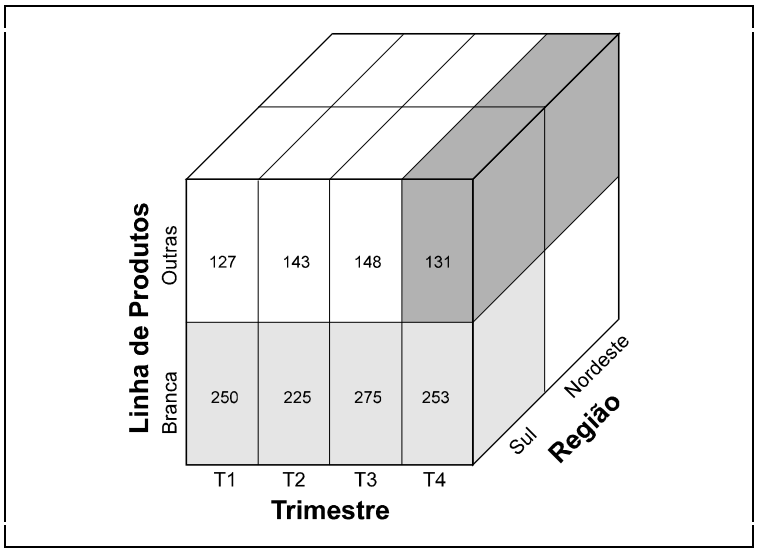


Figura 16 - O "Cubo"
Fonte: Elaborado pelo autor

Resumindo os conceitos aqui apresentados, pode-se dizer que o modelo E-R não difere da modelagem tradicional que é utilizada nas aplicações OLTP. No modelo dimensional as entidades de negócios são tabelas chamadas "dimensões" e as ocorrências do negócios, ou seja, os "fatos" são tabelas de métricas relacionadas às dimensões.

Os dados podem ser armazenados em um banco de dados relacional e mesmo assim podem ser vistos de forma multidimensional (Singh, 1997). Para isso utiliza-se uma ferramenta do tipo ROLAP (*Relational OLAP*). Dessa forma é possível manter a escalabilidade e performance dos bancos de dados relacionais sem perder a funcionalidade analítica.

3.4 Aspectos Humanos

Segundo Arnold (1995), as companhias geralmente implementam sistemas de informação com propósitos específicos tais como melhorar a produtividade da força de trabalho criando vantagem estratégica e competitiva, ou expandindo a efetividade dos processos de negócios, mas nem sempre esses objetivos são claramente percebidos. Martinsons & Chong (1999) dizem que a tecnologia é introduzida para estimular eficiência econômica mas acrescenta que se ela não é bem planejada para melhorar a eficiência dos recursos humanos, será um desastre.

A TI pode ajudar as pessoas a fazerem melhor o seu trabalho, mas somente se eles têm vontade para usar essa tecnologia e se eles tornam-se usuários efetivos. Muitas vezes o fracasso de projetos dessa natureza está associado à má utilização, sub-utilização ou abandono. Martinsons & Chong (1999) acrescentam que pouco menos de 10% das falhas podem ser atribuídas a problemas tecnológicos e, muito mais freqüentemente, resultados insatisfatórios ocorrem devido a fatores não técnicos.

Como mostrado na Figura 17, um sistema de informação efetivo consiste da aplicação certa de TI para ajudar as pessoas certas na execução do processo certo. A tecnologia tem conseguido avanços extraordinários e, para acompanhar esse ritmo, o elemento humano tem se tornado um fator determinante no sucesso de um sistema de informação. Uma consideração

inadequada e um fraco gerenciamento dos aspectos humanos podem atrapalhar o uso e efetividade de um SI (Martinsons & Chong, 1999).

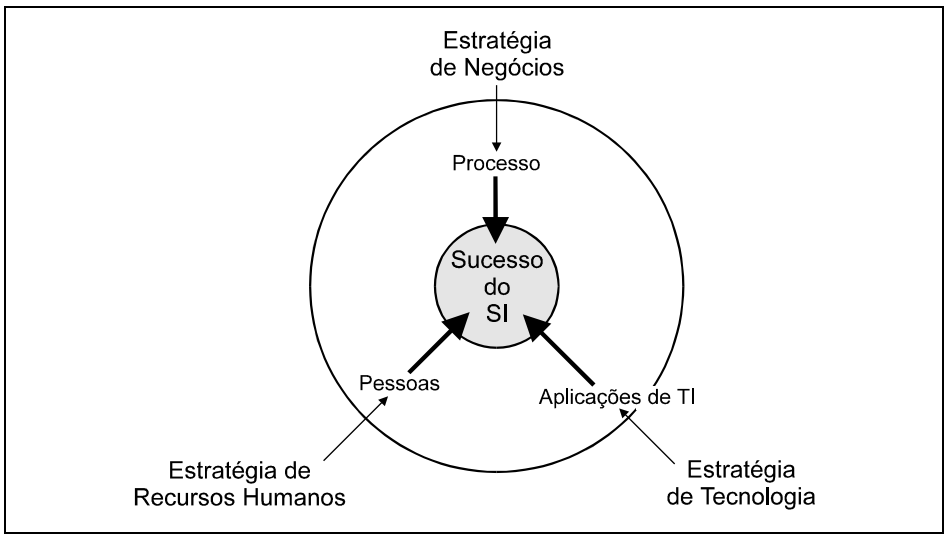


Figura 17 - Alinhando pessoas, processos e tecnologia
Fonte: Martinsons & Chong (1999)

Os sistemas de informação devem ser considerados não somente em termos de dinheiro, tecnologia, e processos de negócios, mas também em termos do seu efeito em atitudes e comportamentos (Martinsons & Chong, 1999). Sendo assim, o gerenciamento de recursos humanos é crítico quando se deseja perceber os potenciais benefícios de um investimento em TI. A melhoria nos negócios depende do trabalho mais inteligente e efetivo das pessoas.

Inmon & Hackathorn (1994) dizem que um aspecto essencial para o sucesso em um ambiente de *Data Warehouse* é o fator humano. Os melhores projetos e arquiteturas terão seu uso limitado caso não hajam pessoas disponíveis e preparadas para por os planos em ação. Para Singh (1997), os assuntos que envolvem pessoas podem levar o *Data Warehouse* a tornar-se o maior bem da empresa bem como o seu maior pesadelo.

Deck (1999) cita a opinião de algumas pessoas, diretamente envolvidas no ambiente de *Data Warehouse*, com relação à importância dos aspectos humanos:

- "Em projetos de *Data Warehouse*, ciências sociais podem ser tão importantes quanto ciência da computação" (Wayne Eckerson, vice-presidente de serviços de tecnologia do *Data Warehouse* Institute).
- "*Data Warehouses* não são simplesmente projetos porque ultimamente eles dependem muito dos usuários finais" (Jackie Sweeney, analista do International Data Corp).
- "Projetos de *Data Warehouse* falham mais por razões organizacionais do que por razões técnicas" (Herb Edelstein, analista do Two Crows Corp.).

Perkins (1999) resume a importância dos aspectos humanos quando diz que todos os potenciais usuários do *Data Warehouse*, mesmo os executivos, de cada unidade e nível da organização, devem ser ativamente envolvidos no projeto, desenvolvimento e gerenciamento do *Data Warehouse*. Os usuários do *Data Warehouse* terão a maior influência na aceitação do *Data Warehouse*, então é imperativo que suas necessidades sejam tratadas. Eles também são

os "donos" e "zeladores" dos dados operacionais e assim a melhor fonte para expertise no assunto.

3.4.1 Usuário

Inmon & Hackathorn (1994) dizem que, basicamente, o público atendido pelo ***Data Warehouse*** pode ser chamado de comunidade SAD, onde estariam as pessoas que tomam decisões de alto nível e de longo prazo. Também atende às comunidades de SIE, de análise de planilhas eletrônicas, de relatório ***ad hoc***, de relatórios periódicos, etc. Esse público variado é possível porque o ***Data Warehouse*** proporciona uma ampla variedade de usos para os seus dados que possuem duas características básicas: são integrados e são históricos. A característica da integração provê uma base para análises com abrangência corporativa e a característica histórica provê uma base para análises de tendências que permite ver a empresa em uma linha do tempo.

O usuário final pode ser definido como uma pessoa de negócios que pode, ou não, ter conhecimentos técnicos (Inmon, 1997; Inmon & Hackathorn,1994). Seu objetivo primário é fazer os negócios acontecerem da melhor maneira possível e ele utiliza o ***Data Warehouse*** como um meio para esse fim. Existem os usuários que fazem uso intensivo do ***Data Warehouse*** bem como aqueles que usam muito pouco. Pelo lado dos que fazem uso intensivo encontram-se os chamados ***power users***. São aqueles usuários que entendem de computação, estão motivados com o sistema e sabem bem o que estão procurando. Do outro lado está o usuário casual, que faz pouco uso do processamento informacional. Normalmente esse usuário entende de computação, mas não conhece todos os possíveis usos da informação para melhorar o processo de tomada de decisão.

Segundo Inmon (1997) é interessante verificar como uma analista de SAD entende a utilização do ***Data Warehouse***. Sua postura é de "me dê o que eu digo que quero e, então, eu poderei lhe dizer o que eu realmente quero", ou seja, o analista de SAD atua por meio de descobertas. Quando ele tem resultados em relatório, ou em uma tela, ele começa a explorar o potencial do sistema.

Basicamente, existem quatro classificações gerais de usuários em uma comunidade SAD: turistas, fazendeiros, exploradores e mineradores de dados (Inmon, 1998). Os turistas seriam aqueles que sabem como cobrir uma considerável quantidade de dados rapidamente, mas de maneira superficial. Os fazendeiros fazem a mesma atividade repetidamente, sabem o que querem antes de iniciarem a execução de uma pesquisa e operam de maneira previsível. Os exploradores não sabem o que querem e operam na intuição, criando grandes pesquisas (***queries***) em cima de dados históricos e detalhados. Finalmente temos os mineradores que, metodicamente, examinam grandes quantidades de dados em nível detalhado buscando a confirmação de hipóteses ou mesmo uma nova hipótese. Apesar de muitas semelhaças, exploradores e mineradores apresentam diferenças, por exemplo: os exploradores operam em uma mesma base de dados ao passo que os mineradores estão sempre mudando a base em que operam.

Resumindo o que foi discutido, o usuário de *Data Warehouse*:

- é uma pessoa que toma decisão de alto nível e longo prazo;
- é uma pessoa de negócios que pode, ou não, ter conhecimentos técnicos e
- atua por meio de descobertas;

3.4.2 Equipe de Projeto

Conforme o PMI (1996), as partes envolvidas são os indivíduos e organizações que estão diretamente envolvidos no projetos, ou ainda aqueles cujos interesses podem ser afetados, de forma positiva ou negativa, no decorrer do projeto ou mesmo após sua conclusão. De forma geral, os principais envolvidos em um projeto seriam :

- O Gerente do Projeto que é o indivíduo responsável pela gerência do projeto;
- O Cliente, que pode ser um indivíduo ou organização, que fará uso do produto do projeto. Podem existir múltiplas camadas de clientes;
- A Organização Executora que seria a empresa cujos funcionários estão mais diretamente envolvidos na execução do projeto;
- O Patrocinador, que pode ser um indivíduo ou grupo, dentro da organização executora, que provê os recursos financeiros, para o projeto.

Singh (1997) diz que os indivíduos que vão compor a equipe podem vir de diferentes disciplinas, tais como análise de negócios, modelagem de dados, administração de dados, desenvolvimento de aplicações e suporte aos usuários finais. De onde quer que os recursos venham é importante que a equipe esteja focada e motivada. Isso não é considerado uma tarefa fácil para as empresas mas o fato de trabalhar com uma tecnologia nova pode ser um ponto positivo nesse sentido. O conjunto de habilidades e responsabilidades necessárias para manter o *Data Warehouse* são variadas e complexas.

Treinamento adequado, planejamento, motivação e organização podem ajudar para que a equipe de *Data Warehousing* esteja posicionada para prover um alto nível de suporte e com isso melhorar o retorno sobre investimento. De qualquer forma esse suporte não será suficiente se os papéis, as responsabilidades e a estrutura organizacional não estiverem claramente definidos.

Independente do tamanho da estrutura, Singh (1997) diz que existem alguns grupos-chave que devem estar presentes para assegurar um suporte amplo e adequado. Cada um desse grupos tem um importante papel nas fases de desenvolvimento, implementação e manutenção do *Data Warehouse*. Como não é fácil definir exatamente onde termina o projeto de um *Data Warehouse* esse grupos continuam através da vida do *Data Warehouse*. De maneira geral esses grupos incluem:

- Patrocinadores executivos. Sem dúvida, um dos mais importantes papéis no gerenciamento de um *Data Warehouse* é a liderança do projeto. O patrocínio ao nível executivo é um tipo de liderança necessário para fazer o *Data Warehouse* uma prioridade dentro da organização. Segundo Barquin et al. (1995), sem um líder dinâmico, mesmo após um tempo da implementação, o *Data Warehouse* pode ter o seu sucesso ameaçado pois seu processo é uma jornada sem final definido.

- Analistas de negócios. É um papel crítico pois é ele quem inicia o projeto. Esse grupo é responsável pode identificar e definir objetivos e usuários para o *Data Warehouse* e garantir que os objetivos estratégicos dos negócios sejam alcançados.
- Suporte ao usuário final. Todo investimento feito em pessoas, tecnologia e dados pode ser perdido se os dados não forem entregues às pessoas certas que necessitam deles para dirigir, gerenciar e avaliar a organização.
- Suporte técnico. Esse grupo deve ter habilidade e experiência para definir os processos de coleta, transformação e carga dos dados bem como para elaborar os modelos de dados que serão a base para o *Data Warehouse*.

A Tabela 13, proposta por Kachur (2000), descreve uma lista de papéis e como seria a participação destes no ciclo de vida do *Data Warehouse*:

	Gerenciamento			Entrega								
Descrição do papel →	Comitê do Projeto	Patrocinador do Projeto	Gerente de Projeto	Analista de Negócios	Administrador de Metadados	Arquiteto de *Data Warehouse*	Administrador de banco de dados	Analista de Suporte à Decisão	Analista de sistemas de negócios	Arquiteto de infra-estrutura técnica	Administrador de qualidade de dados	Analista organizacional
Atividade do ciclo de vida ↓												
Planejamento e Avaliação	**SA**	**S**	**SV**	**P**	**P**	**P**	**P**	**P**	**P**	**P**	**P**	**P**
Piloto	**SA**	**S**	**SV**	**SV**	**P**	**P**	**P**	**P**	**P**	**P**	**P**	**P**
Desenho iterativo	**SA**	**S**	**SV**	**SV**	**P**	**SV**	**P**	**P**	**SV**	**P**	**P**	**P**
Teste e implementação	**SA**	**S**	**SV**		**SV**	**SV**	**P**	**P**	**SV**	**P**	**P**	**SV**
Manutenção e suporte	**SA**	**S**	**SV**		**S**	**SV**	**SV**	**SV**		**SV**	**SV**	**S**
Análise de evolução	**SA**	**S**	**SV**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**

Tabela 13 - Papéis no ciclo de vida do *Data Warehouse*
Fonte: Kachur (2000)
Legenda: (A)ceita – (P)roduz – (S)uporta/Apóia – (V)alida

A quantidade de pessoas que vão ser designadas para cada um destes papéis vai variar em função do escopo e tamanho do projeto. Uma pessoa também não está restrita a um único papel e sua participação vai depender de quão abrangentes são suas habilidades. Em grandes projetos pequenas equipe poderiam ser designadas para realizar cada papel (Kachur, 2000).

Segundo Barquin et al. (1995), um grande erro é colocar, para comandar o projeto de *Data Warehouse*, um gerente orientado à tecnologia ao invés de um gerente orientado ao negócio. Isso não quer dizer que pessoas técnicas não possam ser bons gerentes de projetos mas o *Data Warehouse* está voltado aos negócios e ter uma pessoa de negócios no comando pode ser um grande passo para o sucesso do projeto.

Dentro do conceito de Fábrica de Informações Corporativas, Imhoff (1999) diz que uma parte da equipe fica muito focada nos sistemas operacionais, na camada de transformação e integração e na obtenção dos dados. Uma outra parte ficaria "gravitando" em pesquisas, *Data Marts*, relatórios, etc. fornecendo informações para a comunidade de negócios. Os objetivos do primeiro grupo seriam: integração do dados, visibilidade (os dados devem estar em um formato conveniente e disponível para facilitar o acesso) e qualidade. O segundo grupo

deveria se preocupar basicamento com: capacidade (cada iteração desse grupo deve resolver um problema de negócio específico), clareza (a comunidade de negócios não pode usar essas capacidades sem primeiro entender seu propósito, como e quando usá-las) e segurança (garantia de que a informação não vai ser roubada, acessada indevidamente, etc.) A Figura 18 esquematiza esse ambiente e os objetivos dos dois grupos:

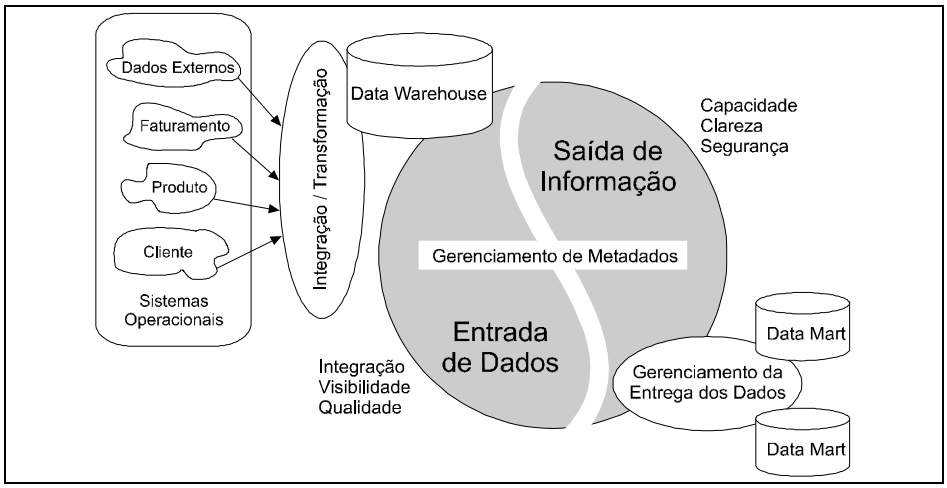


Figura 18 – Entrade de Dados e Saída de Informações
Font: Imhoff (1999)

3.4.2 Facilidade de Uso e Utilidade

O trabalho de Arnold (1995) destaca duas variáveis que são muito importantes na aceitação e uso de sistemas de informação: "facilidade de uso percebida" e "utilidade percebida". Facilidade de uso percebida é definida como "o grau para o qual uma pessoa acredita que usando um sistema particular deverá ser livre de esforço", e utilidade percebida é definida como "o grau para o qual uma pessoa acredita que, usando um determinado sistema, deverá aumentar sua performance no trabalho".

Portanto, mesmo que o projeto seja de qualidade, ele será de pouco valor para os usuários se não for percebido como útil em suas atividades diárias e também relativamente fácil de usar. Garantir que os usuários tenham essa percepção pode ser a ponte entre o novo sistema e o sucesso da implementação.

Quando o usuário está envolvido ele tende a usar o sistema e assim ele passa a utilizá-lo, sugere mudanças e vê resultados. Dessa forma ganha-se um aliado que passa a ser um vendedor do sistema e disseminador da tecnologia. Com isso o projeto tende ao sucesso.

Segundo Berson (1997), o uso do *Data Warehouse* inclui:

- Localização da informação correta.
- Apresentação da informação (relatórios, gráficos)
- Teste de hipóteses.
- Descoberta da informação.
- Compartilhamento da análise.

Com o uso efetivo, alguns benefícios tangíveis do *Data Warehouse* podem ser percebidos (Berson, 1997):

- Rotatividade do estoque melhorada.
- Custo de introdução de um produto é reduzido com melhoria na seleção do mercado alvo.
- Melhora na inteligência dos negócios é permitida pela melhora na qualidade e flexibilidade da análise do mercado disponível através estrutura de dados em vários níveis, que podem ir do detalhe até uma alta sumariação.

Se o *Data Warehouse* é usado, alguns benefícios intangíveis poderiam ser percebidos (Berson, 1997). Embora difíceis de quantificar, esses benefícios também devem ser considerados quando o *Data Warehouse* está sendo planejado:

- Melhoria da produtividade, por manter todos os dados requeridos em um único local.
- Redução do processamento, suporte e software de apoio redundantes em aplicações de apoio à decisão.
- Melhoria nas relações com os clientes através do melhor conhecimento das necessidades e tendências individuais.
- *Data Warehouse* pode prover percepções úteis nos processos de trabalho, resultando em um atalho no desenvolvimento de idéias para a reengenharia de processos.

3.4.3 Treinamento

Segundo Atre (1997b), freqüentemente o treinamento é pouco empregado e pouco apreciado. Usuários e pessoal técnico muitas vezes dizem que somente produtos requerem treinamento. Mas o treinamento freqüentemente falha onde não há fundamentação conceitual para o aprendizado.

Construir um *Data Warehouse* tecnicamente excelente e deixar os usuários perdidos nele sem nenhum treinamento é uma maneira de encontrar problemas (IBM, 1999). Programas de treinamento devem ser desenvolvidos não com objetivo na tecnologia por trás das telas mas, de preferência, com objetivo na forma de usar o sistema pelo ponto de vista do negócio.

Arnold (1995) diz que o jeito mais óbvio para tentar mudar a percepção das pessoas é através do treinamento e desenvolvimento, no entanto, se o treinamento pode influenciar ou não a percepção da utilidade e da facilidade de uso de maneira significante é uma questão pesquisável, mas existe uma boa razão para achar que sim. Nesse caso o treinamento e o desenvolvimento assumem uma importante dimensão e será necessário encontrar, com certeza através de rigorosa pesquisa, que métodos de treinamento mais influenciam no uso e aceitação de sistemas.

Martinsons & Chong (1999) chamam a atenção para a dificuldade existente em superar a resistência para uma mudança não desejada simplesmente através de treinamento ou persuasão forçada. Ao contrário, é necessário primeiro estabelecer as condições culturais e estruturais que conduziram à mudança. Uma vez que um ambiente favorável foi criado, treinamento pode ajudar a facilitar o processo de mudança.

Segundo Kimball (1998), a necessidade de treinamento, de maneira geral, concentra-se nos analistas que geram e modificam os relatórios pré-definidos e parametrizados e um dia de treinamento pode ser mais do que suficiente. Se for de um dia, o treinamento deve ser dividido em duas partes: uma parte que trate do conteúdo existente no *Data Warehouse* e outra sobre a utilização das ferramentas para acesso e apresentação dos dados.

3.4.4 Envolvimento e Participação

Segundo Martinsons & Chong (1999), as aplicações iniciais de TI demandavam muito conhecimento técnico e o envolvimento de pessoas estava limitado ao pessoal de informática. Os usuários identificavam suas necessidades no início do desenvolvimento e após um certo tempo recebiam um sistema terminado.

O problema pode ser maior quando os gerentes das áreas usuárias abdicam da sua responsabilidade sobre o sistema e deixam para os especialistas de TI, assumindo, de forma errônea, que o sucesso de um sistema de informação depende basicamente de critérios técnicos. Embora sejam bons tecnicamente, os especialistas de TI geralmente falham em considerar a importância dos aspectos humanos e a perspectiva do usuário (Martinsons & Chong, 1999).

Martinsons & Chong (1999) dizem que, como uma extensão à distinção entre os papéis ativo e passivo, os usuários podem assumir um dos seguintes três papéis durante o processo de adoção de um SI: proativo, de apoio e reativo (Figura 19).

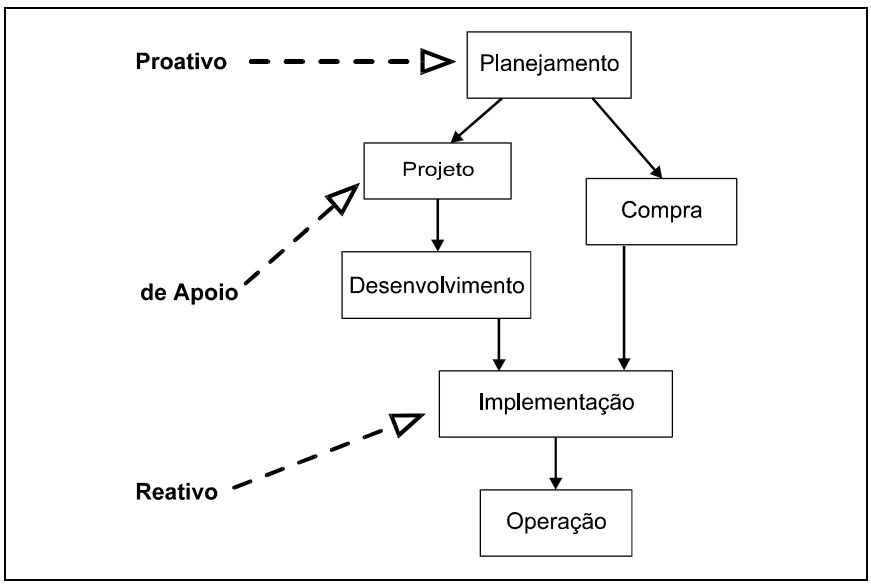


Figura 19 - Envolvimento do usuário no ciclo de vida de um SI
Fonte: Martinsons & Chong (1999)

No papel proativo, os usuários especialistas são envolvidos a partir do estágio de planejamento do SI. Eles ajudam no planejamento do novo sistema. Dificuldades potenciais podem então ser evitadas, ou pelo menos minimizadas, e os usuários podem ser preparados para a mudança esperada. A estrutura ideal de trabalho, tanto da perspectiva social como técnica, bem como obstáculos que atrapalham sua realização podem ser identificados.

No papel de apoio, os problemas dos usuários são antecipados e políticas para prevení-los ou tratá-los podem ser desenvolvidas. Usuários especialistas podem ajudar a projetar programas de treinamento e suporte com um foco comum em assuntos importantes.

No papel reativo, o envolvimento dos usuários especialistas é restrito a atual implementação do novo sistema de informações. Ao mesmo tempo, preocupações como performance negativa e insatisfação do empregado podem surgir.

Martinsons & Chong (1999) dizem que significantes melhorias nos níveis de satisfação do usuário foram relatados em casos onde os usuários tiveram um papel proativo ou de apoio ao invés de um reativo. Estes resultados não estabelecem uma causalidade entre um grande envolvimento do usuário e o sucesso do SI, entretanto eles indicam que um maior nível de envolvimento de usuários especialistas serão positivamente associados com o sucesso de um novo SI.

Para Berson (1997), grande parte das organizações tem um bom conhecimento sobre o desenvolvimento de sistemas operacionais mas, como já foi discutido nesse trabalho, requerimentos e ambientes associados com as aplicações de um *Data Warehouse* são diferentes. Sendo assim, a empresa necessita empregar práticas de desenvolvimento que são diferentes daquelas usadas para aplicações operacionais. Para compor o *Data Warehouse* os desenvolvedores precisam juntar os dados que estão nos sistemas operacionais bem como os dados externos e fazer as transformações necessárias, segundo os critérios dos negócios, antes de disponibilizá-los para os usuários. Como os usuários estão mais perto dos dados, sua participação é muito importante durante o desenvolvimento do projeto.

As restrições de tempo e recursos podem impedir o bom resultado de um sistema de informações, por isso é importante que os principais envolvidos com interesses e perspectivas conflitantes bem como diferentes níveis de poder político tenham a oportunidade de participar no processo de desenvolvimento e implementação do sistema (Martinsons & Chong, 1999).

3.4.5 Consultoria

Geralmente os projetos de *Data Warehouse* são complexos e ao mesmo tempo uma novidade para os técnicos da empresa que serão envolvidos. Baker & Baker (1999) dizem que existem muitas empresas de consultoria que conhecem *Data Warehouse* e podem ser contratadas para ajudar no projeto. Uma vez que o *Data Warehouse* é construído, é possível manter as ferramentas de projeto e gerenciamento sem manter os consultores ao lado.

Conforme Hall (1999), a maioria das organizações acha necessário usar consultores externos para aplicações específicas de *Data Warehouse*, e muitos os acharam úteis para conselhos em *Data Warehousing* também.

Adelman & Moss (2000) dizem que os gastos com consultores e contratados podem elevar muito o custo de um projeto. Esses consultores são usados para suprir a falta de experiência que possa existir na equipe de projeto e os contratados são usados para suprir a falta de pessoal habilitado. Basicamente seriam:

- Contratados de um produto específico que são alocados porque conhecem bem o produto e podem ajudar a instalar o produto ou até mesmo customizá-lo se for necessário. A instalação de uma ferramenta para extração, transformação e carga dos dados é um bom exemplo do tipo de serviço que pode ser prestado por esses contratados. Essas pessoas podem vir de um fornecedor específico ou serem independentes. Para controlar os custos é importante ter pessoas internas trabalhando ao lado dos contratados para absorver o conhecimento e entendimento do processo o que permitirá a continuidade do trabalho no futuro mantendo o que já foi realizado.

- Consultores gerais de *Data Warehouse* que podem estar atuando em uma área específica tal como modelagem de dados, performance, seleção de ferramentas, etc. Normalmente essas pessoas são envolvidas por um tempo menor que os contratados para um produto específico. Basicamente, o papel desses consultores é de trabalhar com o pessoal interno para completar uma tarefa, como por exemplo a seleção de uma ferramenta, e também transferir o conhecimento para que a tarefa possa ser executada no futuro sem a sua presença.

Adelman & Moss (2000) acrescentam que é normal que esses contratados e consultores sejam solicitados a participar de atividades que estão além do trabalho originalmente acordado e que podem contribuir para o sucesso do projeto.

Capítulo 4 – Metodologia de Pesquisa

A pesquisa é o procedimento racional e sistemático que tem como objetivo proporcionar respostas aos problemas que são propostos e assim contribuir com o conhecimento científico. A pesquisa é requerida quando não se dispõe de informação suficiente para responder ao problema, ou então quando a informação disponível se encontra em tal estado de desordem que não possa ser adequadamente relacionada ao problema (Gil, 1987). Segundo Yin (1989), a pesquisa deve ser conduzida por um projeto de pesquisa que visa conectar os dados empíricos às questões iniciais de estudo através de uma seqüência lógica o que permitirá chegar às conclusões. Neste trabalho estamos buscando informações para entender como aconteceu, ou está acontecendo, a implementação de *Data Warehouse* em uma empresa do setor de telecomunicações.

As pesquisas podem ser classificadas basicamente em três grandes grupos: descritivas, explicativas e exploratórias (Gil, 1987). A linha geral que conduzirá este trabalho será a da pesquisa descritiva.

A pesquisa descritiva tem como objetivo principal a descrição das características de determinada população ou fenômeno ou então, o estabelecimento de relações entre variáveis. São inúmeros os estudos que podem ser classificados sob este título e uma de suas características mais significativas está na utilização de técnicas padronizadas de coleta de dados, tais como o questionário e a observação sistemática (Gil, 1987). A pesquisa descritiva busca colher dados sem qualquer manipulação do contexto da pesquisa.

Segundo Rudio (1986), a pesquisa descritiva admite várias formas tais como estudos exploratórios, estudos descritivos, pesquisa de opinião, pesquisa histórica, estudos de caso, pesquisa documental, estudos causais comparativos, experimentos, histórias e *surveys*, entre outros.

A pesquisa nas ciências tem sido fortemente marcada ao longo dos anos, por estudos que valorizam a adoção de métodos quantitativos na descrição e explicação dos fenômenos de seu interesse. Hoje, no entanto, é possível identificar, com clareza uma outra forma de abordagem identificada como "qualitativa" (Godoy, 1995a).

Em linhas gerais, num estudo quantitativo o pesquisador preocupa-se com a medição objetiva e a quantificação dos resultados. Busca a precisão, evitando distorções na etapa de análise e interpretação dos dados, garantindo assim uma margem de segurança em relação às inferências obtidas. De maneira inversa, a pesquisa qualitativa não procura enumerar ou medir os eventos estudados, nem emprega instrumental estatístico na análise dos dados. Parte de questões ou focos de interesses amplos que vão se definindo à medida que o estudo se desenvolve. Envolve a obtenção de dados descritivos sobre pessoas, lugares e processo interativos, procurando compreender os fenômenos segundo a perspectiva dos sujeitos, ou seja, dos participantes da situação em estudo.

Winegardner (1999) coloca que o produto de um estudo qualitativo é ricamente descritivo, expresso em palavras e figuras ao invés de números. Descrições do contexto, os participantes, as atividades de interesse, bem como as próprias palavras dos participantes, normalmente são incluídas.

Embora haja muita diversidade entre os trabalhos denominados qualitativos, alguns aspectos essenciais identificam os estudos desse tipo:

- A pesquisa qualitativa tem o ambiente natural como fonte direta de dados e o pesquisador como instrumento fundamental.
- A pesquisa qualitativa é descritiva.
- Os pesquisadores qualitativos estão preocupados com o processo e não simplesmente com os resultados ou produto.
- significado que as pessoas dão às coisas e à sua vida são a preocupação essencial do investigador.
- Pesquisadores utilizam o enfoque indutivo na análise de seus dados.

Considerando-se o problema a ser pesquisado, será adotada a abordagem qualitativa pois o que se busca é o entendimento de um fenômeno, no caso, a implementação de *Data Warehouse.*

Partindo-se de questões amplas que vão se tornando claras no decorrer da investigação, o estudo qualitativo pode ser conduzido através de diferentes caminhos. Existem basicamente três tipos bastante conhecidos e utilizados de pesquisa qualitativa: a pesquisa documental, a etnografia e o estudo de caso (Godoy, 1995b). O estudo de caso será o método adotado nesse trabalho.

Para fundamentar teoricamente o trabalho, foi feita uma pesquisa bibliográfica. Esta pesquisa reuniu informações das obras (livros e artigos) de diversos autores que discutem de alguma forma o *Data Warehouse*, o setor de telecomunicações e o relacionamento entre ambos, seja para reunir argumentos que vão compor a solução do problema, seja para fornecer um maior embasamento teórico sobre o assunto. A principal vantagem da pesquisa bibliográfica reside no fato de permitir ao investigador a cobertura de uma gama de fenômenos muito mais ampla do que aquela que ele poderia pesquisar diretamente.

4.1 O Estudo de Caso

Para Godoy (1995b), o estudo de caso se caracteriza como um tipo de pesquisa cujo objetivo é uma unidade que se analisa profundamente. O propósito fundamental do estudo de caso, como tipo de pesquisa, é analisar intensivamente uma dada unidade social. Goode e Hatt (1979), dizem que o estudo de caso é um meio de organizar dados sociais preservando o caráter unitário do objeto social estudado e Bonoma (1985) resume o estudo de caso como a descrição de uma situação gerencial.

Bonoma (1985) coloca que, ao contrário das pesquisas quantitativas, o estudo de caso não visa a quantificação ou enumeração mas sim a descrição, a classificação e o desenvolvimento teórico, ou seja, o objetivo é a compreenssão. Selltiz et al. (1974) reforçam essa visão quando dizem que a compreensão de um fato ou situação pode ser feita através do conhecimento de exemplos selecionados, ou seja, os estudos de casos.

O estudo de caso tem se tornado a estratégia preferida quando os pesquisadores procuram responder às questões "como" e "por quê" certos fenômenos ocorrem, quando há pouca possibilidade de controle sobre os eventos estudados e quando o foco de interesse é sobre

fenômenos atuais, que só poderão ser analisados dentro de algum contexto da vida real. Estas três características se enquadram bem nessa pesquisa que procura saber *por quê* optou-se por *Data Warehouse* e *como* ocorreu, ou está ocorrendo a sua implementação em uma empresa do setor de telecomunicações. Nesta análise *não haverá manipulação de variáveis* e o caso está inserido em um *contexto da vida real*. Em se tratando de variáveis, podemos dizer que o estudo apresenta como variável dependente a *implementação do projeto* e como variável independente *o projeto.* Dentro do enfoque de projeto haverá um destaque para alguns *aspectos gerenciais, aspectos tecnológicos* e *aspectos humanos* que foram envolvidos. Poderíamos também citar variáveis intervenientes no processo como por exemplo a cultura, as políticas internas e a estrutura organizacional da empresa.

Yin (1989) coloca que o estudo de caso tem seu lugar no campo das pesquisas com, pelo menos, quatro tipos de aplicações possíveis. Uma que trata das ligações causais que ocorrem nas intervenções na vida real. Outra é a descrição do contexto da vida real na qual uma intervenção tenha ocorrido. Uma terceira refere-se à descrição, a partir de um caso ilustrativo (na qual se enquadra este estudo). Por último, o uso do estudo de caso para explorar situações que não têm um conjunto de resultados claros.

Uma pesquisa que usa o estudo de caso como fonte de informação pode utilizar um ou mais casos, Boyd & Stash (in Campomar, 1991) colocam que "o estudo de caso envolve a análise intensiva de um número relativamente pequeno de situações e, às vezes, o número de casos estudados reduz-se a um. É dada ênfase à completa descrição e ao entendimento do relacionamento dos fatores de cada situação, não importando os números envolvidos".

Segundo Yin (1989), existem cinco componentes em um estudo de caso. O primeiro componente se refere às questões de estudo que devem servir para a escolha da estratégia a ser utilizada. A intenção deste trabalho é descobrir por quê e como implementar o *Data Warehouse* em uma empresa do setor de telecomunicações. Como dito anteriormente, o estudo de caso é indicado para responder essas formas de perguntas.

O segundo componente trata das proposições do estudo que servem para manter o pesquisador no caminho correto. As proposições se referem àquilo que será estudado no escopo do trabalho. Segundo Yin (1989), essas proposições ajudam o pesquisador a buscar aquilo que ele está realmente interessado em responder. A proposição principal deste trabalho é caracterizar a implementação de *Data Warehouse*, com maior foco nos aspectos gerenciais, aspectos tecnológicos e aspectos humanos envolvidos.

A unidade de análise, que serve para delimitar a coleta e a análise de dados, é o terceiro componente de um estudo de caso. Uma unidade de análise poderia ser um indivíduo, um programa específico, um processo de implementação, uma mudança organizacional, etc. A unidade de análise desse estudo será o processo de implementação de *Data Warehousing* em uma empresa privada do setor de telecomunicações que atua no Brasil.

O quarto componente faz a ponte entre os dados encontrados e as proposições feitas anteriormente e, finalmente, o quinto componente é o critério para interpretar as descobertas do estudo que não pode ser definido precisamente por não conter dados para análises estatísticas. Esses dois componentes representam a análise do estudo de caso e servirá de base para a elaboração das conclusões.

A generalização de estudo caso para teoria, segundo Yin (1989), não deve se basear na generalização estatística, uma vez que os casos não são "unidades de amostragem" e não deveriam ser escolhidos por esta razão. Campomar (1991) coloca que as análises deverão ser feitas principalmente por analogias, contendo comparações com teorias, modelos e outros casos. As conclusões deverão ser específicas, com possíveis inferências (não estatísticas) e explicações permitindo que as generalizações sejam usadas como base para novas teorias e modelos.

Para Yin (1989), existem quatro critérios relevantes para a avaliação de um estudo de caso:

- Validade de Constructo, que diz respeito ao correto estabelecimento de medidas operacionais para os conceitos que estão sendo estudados. Isso é importante para minimizar a subjetividade na coleta dos dados. As táticas recomendadas para aumentar a validade de constructo são: utilizar múltiplas fontes de evidência na coleta dos dados, estabelecer uma cadeia de evidências e ter informantes-chave para examinar o relatório preliminar do estudo.

- Validade Interna, que compreende o entendimento das relações causais que mostram quando um evento conduz a outro. Yin (1989) diz que a validade interna está mais relacionada aos estudos explicativos e causais e que nos estudos de casos esse problema envolve o estabelecimento de inferências, que ocorre toda vez que um evento não pode ser diretamente observado.

- Validade Externa, que refere-se ao problema de saber se as conclusões do estudo de caso podem ser generalizadas. Esse assunto tem sido uma das maiores barreiras na realização dos estudos de casos. Yin (1989), argumenta que, ao contrário de outros tipos de pesquisa onde as conclusões de uma amostra podem ser generalizadas para um universo através de estatística, nos estudos de casos a generalização é analítica.

- Fidedignidade, que diz respeito a possibilidade de que o mesmo estudo possa ser repetido por um outro pesquisador e chegue aos mesmos resultados. Para que isso aconteça é recomendado que os procedimentos sejam devidamente documentados.

Para análise do caso, um dos métodos mais recomendados, segundo Yin (1989), é a comparação de padrões (***pattern-matching***) que consiste em comparar os padrões empíricos com os padrões previstos. Quanto maior for a coincidência entre os padrões maior será a validade interna do estudo. Uma outra alternativa é a elaboração de explicações (***explanation-building***) que seria um tipo especial de comparação de padrões.

De qualquer forma, não podemos esquecer que assim com outros tipos de pesquisa, o estudo de caso tem suas limitações. Yin (1989) relaciona as principais críticas feitas aos estudos de casos:

- Falta de rigor. Para Yin (1989), esse problema não é exclusivo do método de estudo de caso e vai depender do esforço do pesquisador a ocorrência, ou não, dessa situação.
- Generalização. No estudo de caso, uma das principais limitações se refere justamente às generalizações. Apesar de permitir uma verificação completa e profunda das variáveis escolhidas dentro da organização analisada, este tipo de pesquisa está limitado à situação

estudada. Yin (1989) diz que os estudos de casos são generalizáveis para proposições teóricas e não para populações ou universos.

- Tempo e relatórios. Essa crítica diz respeito ao longo tempo consumido pelos estudos e aos relatórios gerados que podem ser de leitura difícil de tediosa. Yin (1989) não aceita essa crítica dizendo que é possível realizar um estudo de caso que não seja longo e cujo relatório seja feito de forma apropriada.

Resumindo, temos que este trabalho será conduzido por uma abordagem qualitativa descritiva sob a forma de pesquisa bibliográfica e estudo de caso, pois, conforme o que foi exposto, é uma abordagem que possibilita a consideração dos mais variados aspectos relativos ao fato estudado.

4.2 Modelo Conceitual

A fim de esquematizar o escopo desta pesquisa, é apresentado um modelo conceitual (Figura 20) com suas principais características descritas a seguir:

- O quê se busca: caracterizar o processo de implementação de *Data Warehousing*. Serão analisados os seguintes aspectos:

  - Projeto: identificar as características gerais do projeto: a motivação para iniciar o projeto e o patrocinador.

  - Aspectos Gerenciais: identificar algumas características ligadas ao gerenciamento do projeto: gerenciamento do tempo (cronogramas), gerenciamento dos custos (orçamento), gerenciamento dos riscos e gerenciamento da comunicação.

  - Aspectos Tecnológicos: identificar os principais aspectos tecnológicos envolvidos no projeto: a arquitetura utilizada, os critérios para escolha do *hardware*, a granularidade empregada, o processo de extração, transformação e carga dos dados, a importância dos metadados, a segurança do ambiente, a metodologia utilizada e o modelo de dados adotado. Não faz parte dos objetivos deste trabalho o aprofundamento dos detalhes técnicos de cada item.

  - Aspectos Humanos: identificar os usuários, a equipe de projeto, a facilidade de uso e utilidade do *Data Warehouse*, o treinamento realizado, o envolvimento e participação do usuário e a necessidade de uma consultoria externa.

- O critério utilizado: Analisar o processo de implementação de *Data Warehousing* através de um estudo de caso.

- Condições que serão aceitas como dadas: Políticas internas, Estrutura organizacional, Fornecedores, Concorrência, etc.

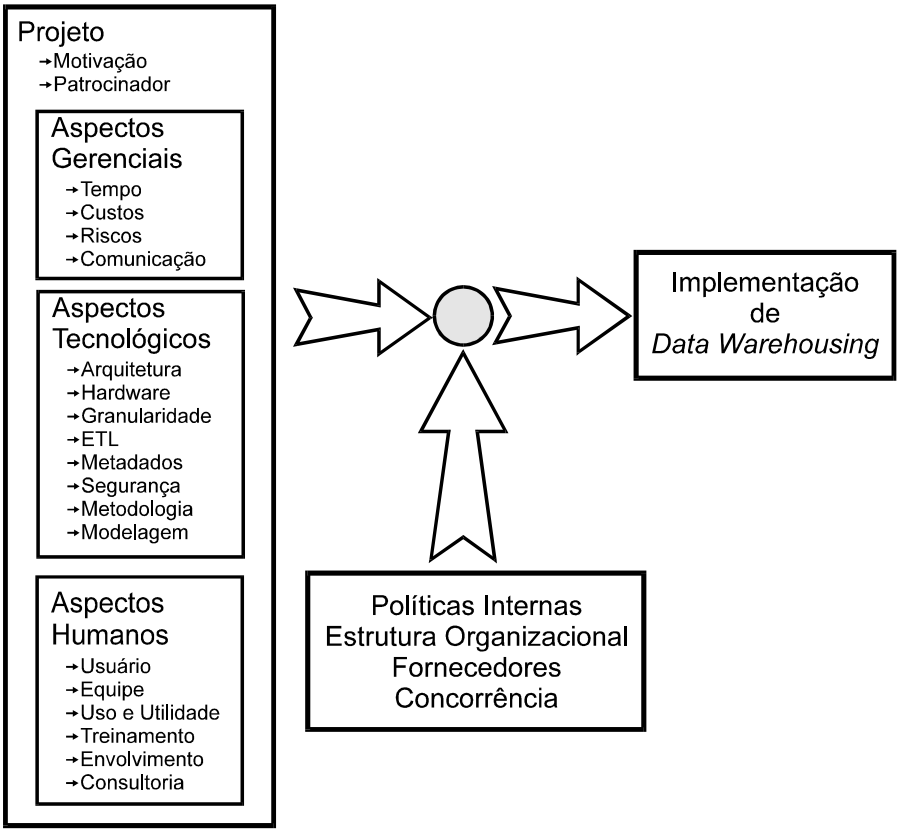


Figura 20 - Modelo Conceitual
Fonte: elaborado pelo autor

4.3 O Caso

O caso a ser analisado é da Telecomunicações de São Paulo S/A – Telesp, empresa de telefonia que atua no estado de São Paulo e passou a ser controlada pela Telefónica de España S.A. ("Telefônica"), a partir de 1998 com a privatização do setor de telecomunicações. Como citado anteriormente, a escolha do setor de telecomunicações se deve ao fato de ser um setor que se apóia fortemente na informática e a utiliza de forma estratégica na condução dos seus negócios. Vale ressaltar que a escolha também foi influenciada pela conveniência do pesquisador em ter contato e acesso facilitado a essa empresa.

4.4 Instrumentação

A entrevista é uma das mais importantes fontes de informação em um estudo de caso (Yin, 1989). Selltiz et al. (1974) afirmam que a entrevista oferece uma grande flexibilidade ao pesquisador na obtenção de informações e Martins (1990) complementa dizendo que a entrevista permite o relacionamento entre entrevistado e pesquisador, através de um diálogo orientado que busca informações e dados para a pesquisa.

Em uma entrevista, os respondentes chaves são convidados a comentar sobre certos eventos. Eles podem propor soluções ou prover percepções em eventos. Eles também podem validar

evidências obtidas de outras fontes. O pesquisador deve evitar tornar-se dependente de um único informante e buscar o mesmo dado de outras fontes para verificar sua autenticidade.

Existem várias formas de entrevista, como por exemplo a não-estruturada, estruturada e *survey*. Segundo Schrader (1971) na entrevista pessoal não-estruturada o questionário é substituído por um roteiro de entrevista que contém apenas algumas questões gerais que não prevêem alternativas de resposta. Também existe uma liberdade para o entrevistados que não fica preso a formulações prefixadas, podendo acrescentar perguntas adicionais e "perguntas de sondagem". Também não existe uma seqüência rígida para a ordem em que os assuntos são colocados, podendo inclusive ser invertida, conforme a necessidade. Para Yin (1989), a entrevista em estudos de casos podem ser:

- de natureza aberta-fechada onde o investigador solicita aos respondentes a apresentação dos fatos bem com as opiniões a eles relacionado;
- uma entrevista focada onde pode assumir uma natureza aberta-fechada ou conversacional, mas seguindo um roteiro estabelecido;
- uma entrevista do tipo *survey* onde as questões e respostas são mais estruturadas.

Segundo Yin (1989), um problema que pode ocorrer com as entrevistas é a influência dos observadores e entrevistadores que podem ser interpretadas de acordo com quem faz e relata a entrevista. Por outro lado, respondentes bem informados podem ser muito importantes para o resultado final. Esse estudo do caso será conduzido basicamente na forma de um roteiro de entrevista, a ser realizado de forma individual, com a intenção de obter o máximo de informação possível dos entrevistados.

4.5 Respondentes

A intenção inicial era de entrevistar dois profissionais da área de TI, o gerente de projeto e um analista e mais dois profissionais da área usuária. No final acabaram sendo entrevistadas nove pessoas sendo duas da área usuária, cinco da área de sistemas, uma da consultoria contratada e uma da empresa que tercerizou a área de produção de informática. O perfil dos respondentes foi definido de forma a tentar se obter uma visão por dois ângulos: desenvolvedor e usuário. O questionário foi único para todos os entrevistados na tentativa de se perceber o quanto cada profissional consultado conhecia do projeto como um todo.

4.6 Desenvolvimento da pesquisa

A pesquisa começou com um amplo estudo bibliográfico para aprofundar os conhecimentos sobre o tema e identificar os aspectos, relativos à questão principal, que já foram trabalhados pela literatura. Com base nesses aspectos foi elaborado o roteiro das entrevistas visando obter o máximo de informações possíveis. As entrevistas foram realizadas durante os meses de maio e junho de 2001, pelo próprio pesquisador, no sentido de esclarecer eventuais dúvidas e deixar o respondente a vontade para justificar qualquer ponto de vista que poderia ser importante para o resultado final e venha ratificar a validade interna da pesquisa. Outros detalhes para caracterização do caso e do projeto foram obtidos via pesquisa documental dentro da empresa. Os dados coletados são apresentados na forma de descrição e depois são confrontados com a literatura a fim de se chegar a uma conclusão. Vale lembrar que este tipo de pesquisa (estudo de caso) não permite generalização dos resultados.

Capítulo 5 – O Estudo do Caso

Durante a descrição do caso é adotado o nome "Telesp", que é o nome oficial da empresa, quando o assunto se refere à empresa residente no Brasil, mais especificamente em São Paulo, que está sendo estudada. Quando for citado o nome "Telefônica", trata-se do grupo Espanhol que tem atuação mundial. Isso é feito para evitar confusão com o nome fantasia da Telesp que é Telefônica. Qualquer referência a um momento atual ou recente se refere ao período de maio-junho de 2001, época em que foi feita a pesquisa (entrevistas).

5.1 A Empresa

No início dos anos 70, o Estado de São Paulo tinha 256 empresas de telefonia e a maior delas era a CTB – Companhia Telefônica Brasileira – subsidiária de uma multinacional. A Telesp surgiu em 12 de abril de 1973 e passou a encampar as empresas existentes no Estado, contratar e instalar terminais telefônicos e sistemas de transmissão.

A partir de então, a Telesp começou a prover serviços de telefonia pública no estado de São Paulo como uma operadora do sistema Telebrás. Em 1973, a CTBC – Companhia Telefônica da Borda do Campo, tornou-se uma subsidiária de Telesp. A CTBC começou sua atividades em 22 de março de 1954, como empresa privada, provendo serviços públicos de telefonia fixa para a região conhecida como Grande ABC que compreende os municípios de Santo André, São Bernardo do Campo, São Caetano do Sul, Diadema, Mauá, Ribeirão Pires e Rio Grande da Serra.

A Telesp Participações S.A. (TelespPar), foi constituída de acordo com o artigo 189 da Lei nº 9.472/97 – LGT e com base no Decreto nº 2.546 de 14 de abril de 1998, como parte do processo de cisão da Telebrás. Ela passou a ser a controladora da Telecomunicações de São Paulo S.A. – Telesp e da Companhia Telefônica da Borda do Campo – CTBC.

No leilão público realizado em 29 de julho de 1998, na Bolsa de Valores do Rio de Janeiro, as ações de controle da TelespPar foram adquiridas pela Tele Brasil Sul Participação S.A. – TBS, um consórcio que tinha participação majoritária da Telefônica Internacional S.A. – TISA com 69,04% das ações. Também compunham esse consórcio a Portugal Telecom S.A. (19,85%), Iberdrola Investimentos S.U.L (6,99%), Ptelecom Brasil S.A. (3,15%), Banco Bilbao Vizcaya (0,95%) e outros (0,02%). Desde que assumiu a operação da Telesp e CTBC, em agosto de 1998, a Telefônica iniciou um processo de expansão e ampliação do sistema de telefonia fixa do Estado de São Paulo. Nessa época a Telesp tinha 5,6 milhões de telefones instalados em todo o estado e a CTBC cerca de 600 mil. A área de cobertura da Telesp atende 625 municípios em São Paulo.

Em 30 de novembro de 1999 a TelespPar, devido a uma reestruturação societária, passou a se chamar Telecomunicações de São Paulo S.A. – Telesp. A empresa é a principal provedora de serviços públicos de telecomunicações de linha fixa no estado de São Paulo e seu contrato de concessão expira em 31 de dezembro de 2005, podendo ser renovado por um período de 20 anos. Em dezembro de 1999 a Telesp adquiriu participação acionária nas Centrais Telefônicas de Ribeirão Preto S.A. – CETERP.

Basicamente, os serviços de telefonia fixa oferecidos pela Telesp consistem em serviços locais, incluindo instalação, assinatura básica, serviço medido e telefones públicos; serviços intra-regionais de longa distancia; transmissão de dados; serviços de rede, incluindo interconexão e aluguel de facilidade e outros serviços.

A Telesp ainda não tem autorização para fornecer serviços inter-regionais de longa distância ou internacionais que abrangem as ligações entre um ponto na região e um ponto fora da região ou do país. Caso a Telesp consiga cumprir, até o final de 2001, as obrigações (metas) determinadas pela Anatel ela terá autorização para executar esses serviços a partir de 2002.

Quando iniciou suas operações em São Paulo, a Telefônica teve vários problemas. O foco da empresa estava voltado para atender a demanda reprimida e os problemas com as linhas já instaladas começaram a aumentar. Em determinado momento, os problemas chegaram a afetar simultaneamente mais de 10% dos aparelhos instalados na cidade, quando o percentual máximo aceitável é de 1% de linhas com defeito (Colombini, 2000). O Procon (Órgão de Proteção ao Consumidor) chegou a registrar 2.349 processos em Abril de 1999 mas esse número baixou para menos de 200 em novembro de 2.000. A empresa chegou a instalar 3 milhões de linhas em pouco mais de um ano e meio de atividades, graças a investimentos em infra-estrutura e tecnologia.

Após dois anos e meio de operação, a empresa contabilizou um investimento de R$ 25,2 bilhões. Esse valor representa o maior investimento já realizado por uma empresa estrangeira no Brasil nesse espaço de tempo (Colombini, 2000). Para 2001 já estão previstos mais R$ 4,5 bilhões. Foram instalados mais de 6 milhões de linhas o que praticamente dobrou o número inicial. A velocidade chegou a mais de 300 mil linhas por mês, o mesmo número que a antiga Telesp demorava cerca de um ano para alcançar.

No primeiro trimestre de 2001 a Telesp alcançou a marca de 11,2 milhões de linhas em serviço[20] o que representa um aumento de 27,9% nos últimos 12 meses. O total de linhas instaladas chegou a 13,5 milhões em todo o estado. A densidade telefônica chegou a 31,7 linhas para cada 100 habitantes o que significa um aumento de 6% no último ano. A Telesp espera, até o final do ano, poder atender as novas solicitações de linhas em menos de 15 dias. A quantidade de telefones públicos cresceu 25,3% e chegou a 258 mil terminais. Esse número já supera a meta da Anatel para dezembro de 2001 que é de 241 mil aparelhos.

A expansão da rede e a redução de preços fez com que os segmentos da sociedade que tinham pouco acesso aos serviços públicos básicos, como telefonia fixa, tivessem o maior crescimento desde a privatização dos serviços. A penetração de linhas telefônicas nos lares da classe C passou de 33%, em 1998, para 70% no final de 2000. Na classe D, o aumento foi bem maior, passando de 11% para 45% no mesmo período.

Outro resultado dos investimentos realizados é o índice de digitalização da rede que chegou a 94,7%. A digitalização é um fator importante para a expansão de novos serviços como as facilidades da linha inteligente (caixa postal, identificador de chamadas, transferência de ligações, etc.), e da conexão à Internet por banda larga (serviço Speedy). O Speedy conta com

[20] Vale destacar que o número de linhas efetivamente em serviço não é o mesmo (é menor) que o número de linhas instaladas. Uma linha instalada não necessariamente já está em serviço.

mais de 60 mil clientes e os serviços de linha inteligente possuem mais de 7 milhões de itens que foram vendidos a 3 milhões de clientes.

Todos esses fatores levaram a uma melhora financeira da empresa onde o balanço mostra que a receita operacional líquida subiu 25,3%, passando de R$ 1,61 bilhão para R$ 2,03 bilhões no primeiro trimestre de 2001. O resultado líquido foi de R$ 242 milhões.

A Telesp é uma das maiores empregadoras do Brasil, com cerca de 59 mil postos de trabalho, sendo 13.414 diretos e os demais nas prestadores de serviço. A produtividade também aumentou consideravelmente passando de 611 linhas por empregado em março de 2000 para 841 em março de 2001, ou seja, um aumento de 37,7%

A Telesp é a principal plataforma de operações da Telefônica na América Latina, atuando em um estado que movimenta 35,5% do PIB brasileiro. Todos esses números permitiram que a empresa cumprisse antecipadamente várias metas da Anatel até o final de 2000 como mostrado na Tabela 14:

Metas	Posição da Telesp		Metas para 31 de dezembro de		
	31/12/99	30/12/00	2000	2001	2002
Número de linhas instaladas (milhares)	8.648	11.186	8.320	9.727	n.d.
Número de aparelhos de telefones públicos em serviço (milhares)	197,6	223,4	216,0	241,0	n.d.
Serviço de telefones públicos em localidades sem acessos individuais (população com mais de)	Atingiu	Atingiu	1.000	600	600
Distância máxima entre telefones públicos dentro dos limites de localidades (em metros)	Atingiu	Atingiu	800	500	500
Tom de discar dentro de 3 segundos (% de casos)	99,8	99,9	98,0	98,0	98,0
Taxa de chamadas completadas local em períodos de picos (% de tentativas)	61,4	65,2	60,0	65,0	65,0
Taxa de chamadas completadas longa distância em períodos de picos (% de tentativas)	65,4	61,0	60,0	65,0	65,0
Solicitação de reparo por 100 acessos em serviço	2,7	2,8	3,0	2,5	2,5
Solicitação de reparo por 100 telefones públicos em serviço	11,2	10,3	15,0	12,0	12,0
Taxa de atendimento por telefone (auto-atendimento ou telefonista) em até 10 seg (%)	95,6	96,9	92,0	93,0	93,0
Reclamações de erros em contas por 1000 contas telefônicas	3,3	2,15	4,0	3,0	3,0
Nível mínimo de digitalização da rede (%)	86,2	93,8	75,0	85,0	85,0

Tabela 14 - Metas da Anatel alcançadas
Fonte: Internet Telefônica

A Telefônica no Brasil

As empresas de telecomunicações no Brasil faturaram 23,8 bilhões de dólares em 1999 e respondem por 55% do total movimentado pelas 200 maiores empresas de tecnologia, segundo a revista Info Exame (Info, 2000). A operação da Telefônica no Brasil tem sido a mais rentável entre as 11 que o grupo espanhol tem sob controle tanto na Europa com na América Latina. Em 2000, o faturamento foi de aproximadamente 9 bilhões de reais e o lucro chegou a 1,4 bilhão de reais. Isso significa que a cada 100 reais de receita, 15 se tornam lucro.

Passado o impacto inicial, ocasionado pelas inúmeras reclamações relativas aos serviços, a Telefônica passou a procurar outras área de atuação (Colombini, 2000). Sua primeira investida foi na área de internet com a compra do provedor Zaz, que passou a se chamar Terra. Atualmente o Terra é o segundo em seu segmento no mercado nacional. Depois entrou no ramo de *Call Center* adquirindo a A4 que pertencia ao Unibanco. Com o novo nome de Atento a empresa se tornou líder nesse mercado.

Em telefonia móvel a Telefônica é acionista majoritária a Celular CRT (Rio Grande do Sul) e da Tele Sudeste Celular que reúne a Telerj Celular (Rio de Janeiro) e a Telest Celular (Espírito Santo). Ainda participa, em conjunto com o Iberdrola, na gestão da Tele Leste Celular que reúne a Telebahia Celular (Bahia) e a Telergipe Celular (Sergipe). Junto com São Paulo, essas regiões, segundo o IBGE, concentram cerca de 75 milhões de habitantes e 61,3% do Produto Interno Bruto.

Além dessas empresas, o Grupo Telefônica possui outras operando no Brasil: Assist (atendimento domiciliar de telefonia), Emergia (rede de cabos submarinos para transmissão em alta velocidade), Telefônica B2B (comércio eletrônico), Fundação Telefônica (programas sociais), Telefônica Empresas (clientes corporativos), Telefônica Engenharia de Segurança (soluções de segurança para o mercado corporativo), TPI – Telefônica Publicidade e Informação (lista telefônicas), entre outras. As empresas administradas pela Telefônica no Brasil empregam direta e indiretamente cerca de 81.000 pessoas.

Além de conseguir o 1º lugar entre as 200 maiores empresas privadas de tecnologia no Brasil em 1999 e 2º entre as que mais investem em pesquisa (US$ 16.771,9), a Telefônica conseguiu outros números de destaque, dentro do setor de telecomunicações, conforme a revista Info Exame (Info, 2000):

- As maiores de Telecom - 1º lugar
- Excelência empresarial[21] - 1º lugar
- As maiores por venda - 1º lugar - US$ 4.318.107,7 mil
- As com maior lucro - 1º lugar - US$ 411.310,2 mil
- Os maiores empregadores - 1º lugar - 54.495
- Liderança de mercado - 1º lugar - 18,1%
- Endividamento geral - 3º lugar - 22,8%

[21] Indicador gerado pela soma de vários critérios utilizados pela revista Info Exame (Info, 2000).

A Telefônica no Mundo

A Telefônica está presente em mais de 40 países, reunindo cerca de 65 milhões de cliente e mais de 2 milhões de acionistas em um mercado potencial de 550 milhões de consumidores. É a nona maior empresa de telecomunicações do mundo, sendo a segunda maior na Europa, ficando atrás da empresa britânica Vodafone. O faturamento global no último ano chegou próximo de 27 bilhões de dólares.

O maior mercado nesse momento está na Espanha com mais de 18 milhões de linhas em serviço. Existem outros 16,5 milhões de linhas fixa em países como Argentina, Brasil, Chile, Peru, Porto Rico, El Salvador, Guatemala e Áustria. Algumas informações expressivas[22]:

- Tem aproximadamente 120 mil empregados no mundo inteiro;
- É a 5ª operadora do mundo em telefonia móvel, com 23 milhões de celulares;
- Tem mais de 41 milhões de clientes de telefonia fixa no mundo;
- É a 3ª provedora mundial de Internet e o 1ª em língua hispânica e portuguesa;
- É a 2ª provedora mundial de transmissão de dados;
- Teve uma valorização de 704% em Bolsa de Valores nos últimos 4 anos.

5.2 Projeto

O primeiro projeto de *Data Warehouse* na Telesp ocorreu em 1997 a partir de uma iniciativa do departamento de comunicação de dados. Eles contrataram a empresa Lucent para desenvolver uma solução para o gerenciamento de produtos de comunicação de dados, e o *Data Warehouse* fazia parte dessa solução. Era muito pequeno e específico, portanto poderia ser considerado um *Data Mart* de comunicação de dados. Esse *Data Mart* foi desenvolvido pela Informix durante três meses. Foi um projeto rápido pois os dados eram poucos e relativamente simples. A idéia do departamento de comunicação de dados era manter um sistema independente e portanto a tendência de crescimento era bem pequena. Esse *Data Mart* não teve sucesso pois os dados dos sistemas operacionais, não estavam totalmente consistentes e portanto as informações geradas não eram corretas em grande parte dos casos.

O grande projeto da Telesp teve início em 1998, após a privatização da Telebrás, onde a então Telesp foi adquirida pela Telefónica de España. Tudo começou em uma reunião com o novo vice-presidente de Planejamento Estratégico, recém chegado da Espanha, que já estava familiarizado com o conceito de *Data Warehouse* e também conhecia, e utilizava, o sistema que foi implantado para a telefonia fixa na matriz. O discurso inicial foi: "Nós precisamos ter um *Data Warehouse* para cuidar do marketing da empresa". A primeira área que iria se beneficiar do novo projeto seria a vice-presidência de Negócios para o Grande Público (VP-S). Nessa época, a VP-S administrava cerca de 4 milhões de linhas residenciais cujo faturamento correspondia a 40% do faturamento total da Telesp. A partir desse momento formou-se uma pequena equipe, composta por membros da Diretoria de Sistemas (RS) e da Superintendência de Informações de Marketing (S*M), que estava encarregada de dar início ao projeto. Para entender um pouco melhor o posicionamento dessas áreas dentro da empresa a Figura 21 apresenta um organograma resumido da organização.

[22] Informações disponíveis na Intranet da Telefônica.

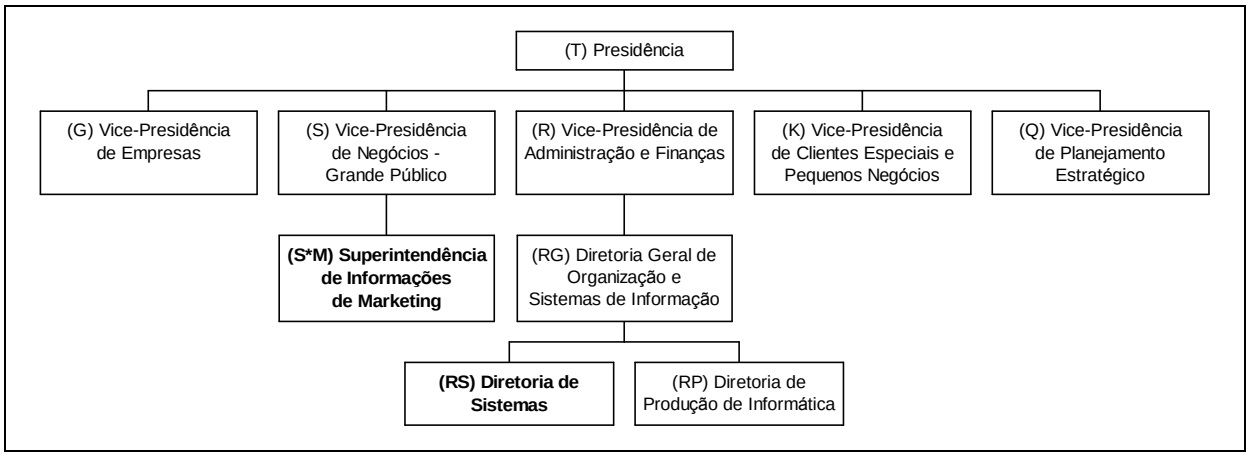


Figura 21 - Principais áreas envolvidas
Fonte: adaptado da Intranet da Telesp

O superintendente de Informações de Marketing diz que, basicamente, recebeu apenas o termo *Data Warehouse* para que fosse dado início ao projeto, ou melhor, aos estudos em cima desse assunto e isso ocorreu em agosto de 1998 com muitas pesquisas na literatura disponível e na Internet, além da participação em seminários nacionais e internacionais.

5.2.1 Motivação

O grande cenário do setor de telecomunicações brasileiro, nessa época, era a privatização do setor com a conseqüente liberação da concorrência e, para entrar nessa competição, a empresa precisaria estar preparada. A Telesp era uma empresa de engenharia, que conhecia muito bem os produtos, seus detalhes, os defeitos e suas causas, mas não conhecia bem o cliente. Era necessário mudar esse conceito e passar a ser uma empresa de negócios. Não havia referências sobre o cliente e a partir desse momento começou-se a pensar em segmentação para poder conhecer o cliente sob diferentes perspectivas.

Os sistemas transacionais, ou operacionais, não estavam preparados para suportar e preparar a empresa para a concorrência que já estava se delineando na época. Era senso comum que isso era um tema crítico pois o conceito de empresa voltada para a tecnologia tinha que mudar para a empresa que usa a tecnologia para fazer negócios. Estava ficando claro que o *Data Warehouse* poderia estar sendo usado para entender os cliente e gerar negócios.

O mercado de telecomunicações, de maneira geral, passou a demandar:

- Maior qualidade nos serviços oferecidos;
- Novos serviços;
- Entrega mais rápida;
- Tempo de vida menor dos produtos;
- Tratamento rápido e flexível para grandes volumes de dados;
- Redução de preços.

E para atender a essas necessidades a empresa precisaria:

- Conhecer melhor o cliente e suas necessidades;
- Aumentar a velocidade de oferta de novos produtos;
- Buscar meios mais eficientes para promover e distribuir novos produtos;
- Acompanhar e monitorar o cliente no pós venda;
- Evitar o *churn* dos clientes mais rentáveis;
- Disponibilizar informações corretas, de forma correta e no tempo certo.

Mas isso não é uma coisa que ocorre de forma simples, pois vários obstáculos precisariam ser superados:

- Os dados são muitos e as informações são poucas;
- O acesso aos dados não é tão simples;
- Não existe uma visão única da informação na organização (necessidade de integração);
- As informações históricas não atendem às necessidades;
- Consultas mais complexas são realizadas apenas por pessoas especializadas e consomem muito tempo;
- Os sistemas existentes não são flexíveis.

Sendo assim, pode-se dizer que o problema de negócios estava definido. Todos os envolvidos entendiam o problema da mesma maneira e colaboraram na especificação dos requisitos pois tinham muito interesse nas informações. Basicamente era preciso saber: O que aconteceu ? Por quê aconteceu ? O que vai acontecer ?

Os envolvidos no projeto tiveram conhecimento de várias aplicações de *Data Warehouse* que estavam sendo utilizadas no mercado de telecomunicações e poderiam contribuir para a transformação da empresa:

- Análise de Fraude:
    - Monitoração individualizada de clientes;
    - Detecção de desvios suspeitos chamada a chamada;
    - Segmentação de clientes segundo variáveis distintas como o valor da fatura;
- Marketing:
    - Melhor conhecimento de clientes e concorrentes;
    - Transição de um empresa orientada a produto para um empresa orientada ao cliente através de:
        - Identificação e reunião de dados sobre clientes atuais e futuros;
        - Análise do comportamento dos clientes para antecipar suas necessidades;
    - Segmentação do mercado em grupos homogêneos melhorando o rendimento das campanhas de marketing.
- GeoMarketing:
    - Análise do comportamento de canais de distribuição (geografia de pontos de venda, etc.)
    - Identificação de regiões adequadas para lançar determinados produtos;
    - Estabelecimento de políticas de preços locais;

- Gestão do *Churn*:
  - Análise do impacto dos planos de tarifa, próprios e dos concorrentes, na base de clientes;
  - Análise da rede para verificar sua qualidade com relação a qualidade de serviços;
- Vendas:
  - Previsões mais exatas, por região, áreas, endereço, etc.
  - Otimização da produção;
  - Avaliação e controle dos objetivos das vendas.
- Cuidados com o cliente:
  - Determinação de produtos e serviços que apresentam problemas;
  - Determinação dos perfis de clientes que apresentaram reclamações;
  - Minimização de perdas de clientes.
- Análise de tráfego:
  - Definição de padrões de tráfego para realizar a segmentação de clientes;
  - Análise automática de predições de tráfego;
  - Definição de melhores tarifas conforme o tipo de cliente.
- Análise da rede e serviços:
  - Analisar histórico de alarmes e incidências nos sistemas de gestão da rede;
  - Analisar a correlação entre alarmes e qualidade de serviço;
  - Analisar os dados históricos de qualidade de serviços.

Um dos grandes objetivos da empresa é "criar valor para o acionista mediante a satisfação e a fidelização dos clientes". Para isso ela precisava:

- Estar orientada para o cliente e não para a linha telefônica;
- Conquistar/reter os clientes rentáveis;
- Otimizar os investimentos em infra-estrutura;
- Acompanhar os programas de fidelizacao;
- Segmentar os clientes com base no tráfego.

O principal benefício esperado com o *Data Warehouse* era o conhecimento mais exato dos clientes, das redes que gere, do tráfego nessas redes, do uso que os clientes fazem dos serviços oferecidos, da eficiência dos canais de distribuição, do faturamento, etc. que permitiria uma melhora nos processos de tomada de decisão.

5.2.2 Patrocínio

Não foi necessário que uma área usuária, ou de sistemas, saísse atrás de um patrocinador pois isso já veio de forma *Top-Down*, ou seja, já tinha o apoio da alta direção da empresa. Como os espanhóis já conheciam o sistema da matriz, sabiam que ele também seria necessário aqui. O projeto tinha o apoio do próprio planejamento estratégico da empresa.

O patrocínio do *Data Warehouse* corporativo é realizado em conjunto, com uma parte sendo financiada pela VP-S, que paga pelo desenvolvimento da solução, e outra pela Diretoria de Sistemas (RS), que é encarregada de fazer toda a contratação de *hardware* e *software*. A VP-S foi o grande patrocinador inicial, não só pelos benefícios que estavam sendo esperados como também pela enorme quantidade de dados que precisavam ser tratados, o que não era possível nos sistemas existentes na época. A VP-G (Empresas) tem uma quantidade muito menor de

dados e consegue, talvez não da melhor forma, baixar em PC os dados dos sistemas transacionais que estão em *Mainframe* e fazer algum tipo de tratamento.

Para os *Data Marts* o negócio muda um pouco, sendo que cada unidade de negócio, no caso a VP-G e a VP-S pagam pelo desenvolvimento e pelos equipamentos mas a área de sistemas é que faz a contratação, coordena o desenvolvimento e a implementação. Como cada área de negócios tem sua ótica e sua prioridade é preciso ter um cronograma patrocinado por alguém que acredita no projeto e que, se precisar "brigar" com alguém, vai fazer. Em uma certa época houve também um *Data Mart* para a VP-K (Clientes Especiais) mas devido a reestruturações internas da organização seus dados foram distribuídos e ele deixou de existir.

No início de 2001, a Telesp fez um *Outsourcing* parcial da área de informática que envolveu toda a Diretoria de Produção (RP). Com essa mudança alguns itens, como por exemplo os discos magnéticos para armazenamento de dados, passaram a ser despesa ao invés de investimento e então as áreas tiveram que rever seus planos financeiros.

O grupo envolvido reconhece a importância do patrocínio em um projeto como esse e sabe também que se começarem a ocorrer conflitos o projeto tende ao insucesso. Vários milhões de reais já foram investidos no projeto da Telesp. Além de toda a infra-estrutura as ferramentas também são caras. Se o patrocinador não tiver esses conceitos bem entendidos ele não dará continuidade ao projeto.

Conforme a área usuária, já não existe mais a necessidade de "vender" o projeto, então a aprovação do orçamento é solicitada para todas as áreas que pedem alguma coisa. Tem acontecido o caso de áreas que cortam verba num primeiro momento mas depois percebem que aquilo é realmente necessário e voltam atrás.

5.2.3 Considerações Iniciais

Após um conhecimento geral do assunto, a Telesp chamou três empresas (Unisys, NCR e IBM) para apresentarem uma solução que atendesse aos interesses da empresa. O projeto estava começando em um ritmo alucinante e essas empresas deveriam estar apresentando propostas em 15 dias para um projeto que deveria ser concluído em 120 dias. A primeira fase do projeto incluía a carga de informações dos serviços comercializados, faturamento e tráfego das linhas individuais (Segmento Grande Público).

A Telefônica possui uma empresa no grupo que é especializada no desenvolvimento de soluções de TI – I+D[23] (Telefónica Investigación y Desarrollo) – e que já implantou *Data Warehouse* tanto na Telefônica fixa como na celular no território Espanhol. No segmento Celular foi implantada uma solução mais barata e mais aberta que no segmento fixo onde a solução era NCR (proprietária). Então a Telesp passou a considerar o I+D como mais um fornecedor de solução, não só por ser uma empresa do grupo mas também pelo conhecimento que já tinha do assunto, principalmente no setor de telecomunicações.

[23] Mais detalhes relativos às consultorias envolvidas no projeto serão vistos adiante.

Os representantes do S*M e do RS mais o diretor de Planejamento Estratégico foram para a Espanha com o objetivo de conhecer o projeto e como ele estava sendo usado. Na viagem de volta tiveram a companhia de uma equipe do I+D que veio para entender a necessidade e também preparar uma proposta de desenvolvimento.

A proposta apresentada pelo I+D incluía, além dos serviços de desenvolvimento, a solução de *hardware* e *software*. Essa solução estava baseada em plataforma SUN com banco de dados Oracle e as ferramentas do usuário que eram o Business Objects (relatórios), DSS Microstrategy (OLAP) e SAS (estatística). Essas ferramentas já eram de interesse da Telesp independente da empresa que ganhasse a concorrência. Era importante que a solução fosse a mais aberta possível permitindo total compatibilidade entre *hardware* e *software*, independente do fornecedor. Após uma análise comparativa de vários fatores, incluindo preço, o I+D foi considerado vencedor e portanto o desenvolvimento do projeto seria feito internamente no grupo Telefônica. Foi uma decisão conjunta entre a área usuária e a área de sistemas.

5.3 Aspectos Gerenciais

5.3.1 Gerência do Tempo

O questão dos prazos já era crítica desde o início dos trabalhos. As três primeiras empresas que foram chamadas para apresentar uma proposta tiveram apenas 15 dias para apresentar uma solução que, depois de aprovada, deveria ser implantada em 120 dias.

Normalmente, é feito um levantamento dos requisitos para determinar o tamanho daquilo que precisa ser feito e com base nisso pode-se montar um cronograma bem realista da situação. Isso não aconteceu pois havia uma definição da área usuária que os primeiros resultados deveriam estar acontecendo em quatro meses.

A equipe da Telesp sabia que poderia ter problemas caso o projeto demorasse muito tempo para apresentar resultados, poderia cair em descrédito por exemplo, e por isso resolveu trabalhar com prazos extremamente curtos. A primeira fase do projeto envolvia um conjunto de dados que, apesar do volume representativo, não seria difícil de ser obtido e facilitaria uma apresentação rápida de resultados.

O início real do projeto aconteceu somente em março de 1999 com a proposta do I+D de entregar os primeiros resultados em cinco meses, sendo que a carga do banco de dados e a emissão dos primeiros relatórios ocorreu em exatos cinco meses. O primeiro projeto, ou a primeira fase do projeto, ocorreu em nove meses, e a partir de então cada fase do projeto foi sendo planejada para um prazo total de doze meses.

Foram feitas várias reuniões com a consultoria (I+D) sobre aquilo que foi visto na Espanha e o que se esperava que fosse feito aqui, quais seriam as fases e quais seriam os primeiros módulos a serem implementados. De qualquer forma o I+D já tinha uma certa experiência nesse tipo de trabalho e procurou fazer um cronograma realista. Cada um dos módulos foi sendo finalizado dentro dos prazos determinados que sempre eram muito curtos conforme o padrão adotado, ou seja, nunca superior a quatro meses para gerar algum resultado.

A primeira entrega do projeto, que ocorreu em agosto de 1999, era composta basicamente por:

- *Data Warehouse* carregado com dados de clientes, serviços comercializados e faturamento;
- Geração de relatórios pré-definidos usando aplicações em Business Objects;
- Análises OLAP usando aplicações em DSS Microstrategy.

Na fase de especificação do projeto foi adotado, para primeiro corte, as linhas individuais porque cada cliente tem uma linha, um número telefônico, uma conta telefônica, etc. Também foi colocado o parque de linhas, que representa a planta instalada da companhia, e o movimento de altas e baixas, que é formado pelas ordens de serviços, para poder observar o comportamento dinâmico da planta, ou seja, quantas linhas foram instaladas e retiradas. Até então esse movimento era analisado por "fotos", o que poderia não refletir a realidade já que a entrada de uma linha e a retirada de outra poderia ocorrer em um intervalo muito curto e não ser registrado na "foto" o que poderia induzir a um comportamento estático da planta. Esse módulo deve representar cerca de 70% da planta da empresa e não havia outro sistema que permitisse o tratamento desejado dessa quantidade de dados. O projeto começou em cima de um critério que não poderia ser atendido nem pelos PCs, pelos altos volumes de dados, nem pelos sistemas transacionais, pela falta de flexibilidade nas análises.

No primeiro ano o foco foi nas linhas individuais e no segundo ano (2000) entraram os outros tipos de linhas: linhas tronco, DDR, 0800, TUP, 0900 (que existia na época). Entraram também as informações de dívida e arrecadação para fazer o perfil de clientes inadimplentes, etc.

Como o trabalho ia ser realizado, basicamente, por uma empresa contratada, o cronograma foi rigoroso e a equipe da Telesp acompanhava de perto as datas acordadas. O acompanhamento por parte dos usuários também era bastante grande, inclusive ao nível de vice-presidência que estavam sempre perguntando sobre a entrada de novos módulos. É verdade também que nem todos os cronogramas foram cumpridos com exatidão e qualidade ao mesmo tempo. Os cronogramas previam a data da primeira carga, mas isso não quer dizer que os dados já seriam válidos. Depois da carga existe um período de validação desses dados que é feito pela equipe do *Data Warehouse*. Esse período é considerado essencial para que seja dado o aceite do serviço entregue.

Conforme a área usuária, um grande problema era que mal se acabava a carga da primeira fase e já havia pedidos para a segunda fase que ainda nem estava contratada. Isso acabava gerando um conflito entre as áreas usuárias e de TI. Não havia um tempo hábil para a estabilização do sistema. Trabalhando dessa forma, não era rara a descoberta de informações inválidas que tinham sido geradas há cerca de dois ou três meses e que acabavam exigindo reprocessamento. O reprocessamento era inevitável nesses casos pois não faz sentido manter informações inválidas em um sistema dessa natureza, onde as informações são históricas e portanto não serão mais atualizadas, sem contar que essas informações são a base para a geração de relatórios, comparações, tendências, etc. que por sua vez serão usados para a tomada de decisão.

Os reprocessamentos não foram poucos e o usuário estava sempre cobrando por resultados. A equipe tinha consciência da necessidade de um período de estabilização para garantir a qualidade das informações. Mas não eram só esses reprocessamentos que afetavam o

cronograma de alguma forma. Segundo a área de sisteams, ocorreram vários problemas com o *hardware* na fase inicial onde, em determinado momento, o fornecedor chegou até a trocar o equipamento. Tudo isso junto acabava gerando um certo tumulto.

A fase inicial foi trabalhosa e problemática mas era um trabalho que precisava ser feito com muito cuidado pois se as informações não estivessem corretas o usuário poderia desistir de utilizar o sistema. A medida que o sistema ia sendo melhor entendido e utilizados pelos usuários, os problemas eram detectados mais rapidamente e conseqüentemente os prazos para correção dos erros eram menores.

A cada ano é feito um novo contrato com a consultoria, com término previsto, e o que vai ser contratado vai depender muito da necessidade do usuário. Apesar disso sempre aparecem pequenas novas necessidades que não estavam previstas e acabam sendo "encaixadas" no projeto.

A ferramenta utilizada para controle dos cronograma é o MS-Project da Microsoft.

5.3.2 Gerência dos Riscos

Como a equipe inicial fez várias pesquisas e participou de vários eventos sobre *Data Warehouse* sabia que 60%-70% dos projetos fracassam e que esse fracasso pode acontece por vários motivos. Basicamente, os principais riscos considerados foram: o patrocínio, a utilização e os prazos.

O patrocínio foi, talvez, o risco menos impactante pois o projeto começou com um "usuário de peso" que já tinha o conceito de *Data Warehouse* bem entendido e estava convencido de sua necessidade. Além de garantir o financiamento do projeto, o patrocínio forte contribui muito para que o usuário final seja incentivado a usar o sistema. É uma orientação que vem na forma *Top-Down* e por isso mesmo tem maior chance de dar certo. O marketing interno não precisa ser tão grande e agressivo para conseguir vender o sistema pois, de uma certa forma, já está vendido.

Na opinião do usuário, a utilização do sistema não está garantida apenas com isso, muito pelo contrário, se as informações não tiverem uma qualidade mínima o usuário pode deixar de usá-lo. E não fará isso por falta de vontade mas sim porque não há muito o que se fazer com uma informação inválida. Para evitar esse problema foi feito um trabalho muito forte com relação a validação dos dados que são carregados no *Data Warehouse*.

Era consenso da equipe que o projeto não poderia demorar muito para ser implantado e por isso os cronogramas foram bastante agressivos. Associado ao fato de que as coisas mudam rapidamente, o projeto foi dividido em vários módulos que deveriam estar sendo concluídos em prazos máximos de quatro meses. Esse período foi considerado suficiente para administrar a expectativa dos usuários até a liberação dos primeiros resultados.

Um outro fato que não foi desprezado foi a "novidade do assunto" em termos de Brasil. A equipe teve notícias de muitas empresas cujos projetos iniciais não deram certo, e a partir daí aprenderam com os erros e refizeram o *Data Warehouse*. A Telesp se considera com "sorte" já que o projeto inicial está dando certo sem alterações drásticas no modelo inicial. O I+D já

tinha experiência em *Data Warehouse* no setor de telecomunicações, e as contribuições que foram feitas pela equipe local fez com que o projeto ficasse melhor do que aquele que foi implementado na matriz.

Com os primeiros resultados positivos passou-se a procurar outras áreas na empresa para utilizar o *Data Warehouse*. Os riscos nesse momento já eram mínimos e o *Data Warehouse* era apresentado como uma solução que fornecia resultados concretos e tinha financiador definido (S*M e RS), ou seja, era só usar. Dessa forma era fácil vender alguma coisa, segundo a área usuária.

Em sua proposta inicial o I+D continha uma parte intitulada "Análise de Riscos" que incluia:

- Disponibilidade do cliente. Disponibilidade dos usuários finais dos diferentes módulos para uma correta e razoável definição de requisitos.
- Designição de responsável pelo *Data Warehouse* na área de sistemas.
- Arquitetura de *hardware/software* deve estar implantada antes do começo de cada atividade.
- Telesp deve fornecer infra de desenvolvimento para realização das atividades.

5.2.3 Gerência dos Custos

Em 1998, quando se iniciaram os estudos para a implantação de *Data Warehouse*, não havia orçamento definido já que o planejamento do ano corrente é feito no ano anterior e o *Data Warehouse* era um projeto que não havia sido previsto pois a nova empresa – Telefônica – tinha acabado de chegar ao Brasil. Para a elaboração de um primeiro orçamento, foi feita a média em cima das propostas que foram apresentadas pelas três empresas iniciais. Com a alteração do projeto para a inclusão do I+D na concorrência, o prazo para início real do projeto se estendeu o que permitiu um melhor planejamento, e então, em março de 1999, já havia um orçamento definido para o projeto.

A área de planejamento estratégico definiu que o orçamento do *Data Warehouse* deveria ficar restrito às áreas S*M e RS para evitar cortes que poderiam colocar o projeto em risco. Esse orçamento era muito centrado nos custos da consultoria e da própria área de sistemas que dizia qual era o valor estimado para novas aquisições de *hardware* e *software*.

A área de sistemas teve alguns problemas de corte de orçamento no primeiro ano. Houve uma redução em torno de 50% e o Real teve um significante desvalorização em relação ao dólar o que diminuiu mais ainda o dinheiro disponível. Isso implicou a compra de um *hardware* inicial com capacidade menor do que a prevista o que não seria suficiente para comportar o volume estimado da primeira fase. Durante o ano de 1999 aconteceram alguns remanejamentos de orçamento entre as área envolvidas para garantir, principalmente, a compra do *hardware*. As máquinas que iam sendo trocadas, em função das novas demandas e capacidade de processamento, eram alocadas para projetos menores, ou seja, não havia perda do investimento já realizado.

O planejamento anual é feito com uma estimativa de quanto se espera evoluir em *hardware*, *software*, discos para armazenamento das informações e consultoria, e tudo isso é colocado no

orçamento. Cada nova solicitação que surge durante o desenvolvimento do projeto é discutida e negociada com o usuário para a definição de prazos e investimentos.

No início, houve até a iniciativa de se fazer algum estudo baseado no ROI mas não foi necessário pois o conceito já havia sido provado na Espanha e ele não precisou ser vendido pelo ROI. A equipe tem notícias de que na Espanha o sistema funciona muito bem e o retorno do investimento ocorre num tempo bem curto (não foram coletados mais detalhes sobre esse assunto).

De qualquer forma a equipe entende que não é fácil fazer algum estudo nesse sentido já que os benefícios alcançados muitas vezes são intangíveis. Durante o primeiro ano muita coisa foi justificada com base nos relatórios que estavam sendo gerados. Os relatórios de tráfego mostravam valores que deveriam estar sendo cobrados mas não estavam. Ao longo do tempo esses valores foram sendo cobrados e o prejuízo diminuído. Isso já começou a mostrar uma forma de retorno do investimento.

Os primeiros relatórios também demonstraram que o usuário tinha uma outra fonte de informações, além do ambiente transacional, e então começaram a surgir relatórios de comportamento, desempenho, etc. que permitiriam desenvolver campanhas otimizadas para determinados segmentos de clientes com conseqüente melhoria dos resultados.

Apesar de saber do potencial do *Data Warehouse* a empresa necessitava de um valor para fazer o estudo de viabilidade do projeto. Em uma das vezes o próprio pessoal de sistemas fez uma pesquisa no *Data Warehouse* e verificou quanto havia de serviços de valor adicionado (relacionados as facilidades CPA[24]: identificação de chamadas, transferência de chamadas, etc.) no ano anterior e quanto existia hoje. Foi verificado que houve um crescimento em torno de 200% nos serviços durante esse período de um ano. Se for considerado um valor médio de R$ 3,00 para cada venda, chega-se a uma receita adicional em torno de R$ 14 milhões. O usuário afirma que não teria conseguido atingir esse número sem o uso do *Data Warehouse* que permitiu fazer o acompanhamento das faturas, o comportamento de tráfego, etc. e propôr o serviço certo para a pessoa certa na hora certa.

5.2.4 Gerência da Comunicação

Nesse ponto o projeto não teve maiores problemas pois havia um comprometimento de toda a equipe que era multidisciplinar e estava motivada para trabalhar com a nova tecnologia. Um representante da área de sistemas, um representante da consultoria e um representante do usuário formavam o comitê de acompanhamento do projeto.

Esse comitê se reúne quinzenalmente para verificar o andamento do projetos, discutir os principais problemas e analisar os próximos passos. Os problemas, em sua maioria, são resolvidos internamente pelo próprio comitê mas os problemas mais sérios podem chegar ao nível de diretoria. Cada um tem sua responsabilidade bem definida e a comunicação entre todos é boa pois o grupo está bem integrado.

[24] Central de Programa Armazenado. É a central telefônica digital que veio para substituir as antigas centrais analógicas.

No sentido de facilitar, não só a comunicação mas também a integração de todos os membros da equipe, eles foram todos alocados em um mesmo ambiente físico.

5.4 Aspectos Tecnológicos

5.4.1 Arquitetura

A consultoria I+D apresentou a proposta de uma plataforma que seguia os moldes do sistema da Telefónica Móviles (Celular da Espanha). A Telefônica Fixa da Espanha usava uma plataforma NCR que era proprietária e bem mais cara que a outra que foi proposta. Isso estava de acordo com as expectativas iniciais da equipe da Telesp que buscava a arquitetura mais aberta possível para evitar problemas entre o ***hardware*** e as soluções de ***software*** de mercado. A intenção também era de se evitar uma dependência grande com um único fornecedor.

A área de interconexão da empresa solicitou um sistema independente (DW-Itx) para armazenar apenas o tráfego que passava pelas centrais telefônicas e isso acabou se tornando um outro projeto. Não ficou no *Data Warehouse* corporativo, ou *Data Warehouse* de Marketing (DW-Mkt) como é conhecido internamente, pois volume de dados era bem considerável e poderia comprometer o sistema principal, além de ter um propósito bem específico com foco diferente das outras áreas usuárias (S, K e G).

Optou-se pela construção de um *Data Warehouse* corporativo desde o início pois a equipe tinha consciência de que, se começasse montando pequenos *Data Marts*, teria problemas no futuro. A integração, com certeza, seria muito difícil já que as áreas geralmente têm interesses diferentes o que iria levar cada *Data Mart* a ter suas características próprias. Uma das vantagens conhecidas dos *Data Marts* está na geração rápida dos primeiros resultados, o que ajuda a dar credibilidade ao projeto junto aos usuários. Para não ter problemas nesse sentido, os cronogramas do DW-Mkt foram montados com prazos bastante agressivos.

A VP-S e a VP-G foram as primeiras que quiseram ter uma certa "independência" através do *Data Marts*. O *Data Mart* que foi criado para a S é basicamente um espelho do *Data Warehouse* porém com menos histórico e periodicidade de carga menor, mas é um ambiente de uso exclusivo. A utilização do *Data Warehouse* corporativo traz informações mais "ricas" mas com os *Data Marts* as amostras podem ser trabalhadas com mais liberdade, em um ambiente independente, sem afetar o desempenho dos demais usuários. Os *Data Marts* são identificados como CRM, ou seja, temos o CRM-S e o CRM-G (no passado existiu também o CRM-K). A Figura 22 mostra um esquema da arquitetura utilizada.

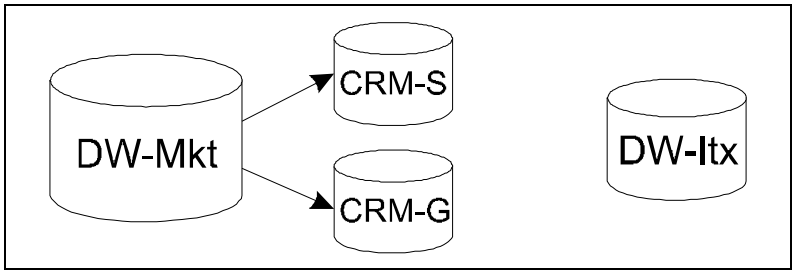


Figura 22 - Arquitetura do *Data Warehouse*
Fonte: elaborado pelo autor

Um exemplo citado para uso para um *Data Mart* é quando existe a necessidade de se montar uma conta telefônica para um Telefone de Uso Público (TUP), pois esse tipo de telefone não gera uma conta específica. Através das informações de tráfego faturado é possível conseguir todas as chamadas a cobrar que vêm de um TUP e aparecem em uma conta telefônica normal. De outro lugar obtém-se as informações do próprio TUP e dos interurbanos que ele fez. Também é possível saber quanto foi utilizado através de cartões telefônicos. A questão é que todas essas informações estão distribuídas no *Data Warehouse* com focos diferentes. Em um *Data Mart* essas informações são agrupadas em um único local (estrela) de fácil acesso, o que gera muita agilidade na geração de relatórios relativos aos TUPs.

A área de sistemas vem analisado com bastante critério as necessidades de um *Data Mart*. Não basta o usuário querer um ambiente independente para trabalhar, é preciso analisar a necessidade e ver se aquilo faz sentido estar em um *Data Mart*, como por exemplo, quando há a necessidade de se criar uma visão diferente para determinado negócio.

No início, o *Data Warehouse* tinha dois aplicativos principais desenvolvidos pelo I+D:

- Segmentação de clientes baseada em tráfego, onde, a partir da análise de uma amostra dos dados, é possível agrupá-los conforme suas características de consumo.
- Acumulação de tarifa, onde, a partir de um histórico de três meses de tráfego é possível analisar o impacto gerado no faturamento pela alteração do valor da tarifa em um determinado horário.

Outra aplicação interessante que existe é o simulador de pontos. Conforme o lugar do ponto de interconexão (ponto de conexão entre duas operadoras) a chamada é cobrada como interurbano ou não. A chamada paga a passagem pelo ponto de interconexão e dependendo do lugar desse ponto ela paga o interurbano (IU) também. Através do simulador de pontos é possível estudar o impacto de se retirar o ponto de interconexão de um determinado local e colocá-lo em outro: qual será a receita ? cobrar IU ou não ? etc.

Outras aplicações existentes são: faturamento, ordens de serviço, dívida e cobrança, parque de linhas, tráfego TUP, tráfego de interconexão, comunicação de dados, serviços de valor adicionado. Mais dois módulos estão sendo implementados: ordens de serviço pendentes e bilhetes de defeito. Para os bilhetes de defeito, o importante não é o defeito em si, mas sim a qualidade da engenharia da empresa. A intenção nesse caso e descobrir se o faturamento de um cliente caiu porque ele passou a usar a empresa concorrente ou caiu porque a rede estava indisponível para uso.

5.4.2 Hardware

O arquitetura inicial era composta de uma máquina E5500, da SUN, com seis processadores e *storage* A5000, com 750 GB úteis[25], também da SUN. A idéia inicial era de uma máquina modelo E6500 com 12 processadores mas isso não foi possível no primeiro momento por conta dos cortes que ocorreram no orçamento.

No segundo ano de funcionamento do *Data Warehouse*, a máquina passou a ser o modelo E10000, da SUN, com 32 processadores e o *storage* mudou para a EMC. No final do segundo ano já havia um total de 3 TB úteis de disco. O DW-Mkt tem cerca de 5.4 TB úteis, o DW-Itx 500 GB, o CRM-G 500 GB e o CRM-S tem cerca de 250 GB úteis. Existe uma expectativa de que até o final do ano o DW-Mkt possa chegar próximo aos 7 TB úteis. Com certeza é um crescimento que superou as previsões iniciais. A parte mais importante do *Data Warehouse*, o banco de dados, está protegida por discos em RAID-1 (segurança por espelhamento), enquanto que o restante, área de extração e tratamento de dados por exemplo, está em RAID-5 (segurança por paridade). O espaço em disco também foi um item importante na fase de análise de *hardware*.

Os equipamentos foram adquiridos para uso exclusivo do *Data Warehouse*. A máquina SUN E10000, além de comportar o DW-Mkt e o DW-Itx, está sendo usada para outros projetos menores, através da divisão do ambiente, mas a partir do momento que o *Data Warehouse* demandar mais recursos a prioridade será toda dele. Isso acontece hoje porque o sistema não está consumindo toda a capacidade da máquina e também é possível porque não é uma arquitetura dedicada exclusivamente ao *Data Warehouse* como é o caso da NCR. O sistema como um todo ainda precisa de muitos ajustes para que possa aproveitar melhor o *hardware* que hoje está sendo utilizado em cerca de 40% de sua capacidade. O CRM-G está instalado em uma máquina modelo E4500 e o CRM-S em uma máquina E5500, ambas da SUN.

Assuntos como velocidade, escalabilidade e disponibilidade sempre foram considerados importantes nesse projeto mas não foram totalmente contemplados no início por conta dos cortes de orçamento.

A janelta para os processo de ETL é suficiente para todas as cargas necessárias mas muitos relatórios ainda demoram para serem executados o que incomoda os usuários. Quando um usuário de TI está com uma consulta de baixa performance ele recorre a alguém de suporte (DBA) que por sua vez faz o ajuste fino nessa consulta para melhorar a performance. O usuário comum, quando necessário liga para a área de TI para conseguir ajuda ou acaba desistindo, o que não é bom pois pode acabar desestimulando o uso. De forma geral, as consultas são bem rápidas, principalmente quando se trabalha com os agregados, mas já ocorreram casos de consultas que chegaram a mais de 24 horas e o resultado foi inválido pois o critério utilizado considerou dados que ainda não estavam disponíveis.

Com a abertura do mercado esperada para 2002, existe uma expectativa de que os volumes, tanto de dados, aplicações e mesmo utilização, cresçam acima do esperado, que seria algo em

[25] É considerada área útil aquela realmente usada pelos dados. Quando os discos possuem segurança RAID-1 (espelhamento) uma área útil de 1 TB necessita de um total de 2 TB de área em disco (1 TB para dados + 1 TB para espelhamento desses dados).

torno de 25% ao ano, e por conta disso estão sendo analisadas alternativas para permitir esse crescimento sem prejuízo de performance.

Está sendo feito um ***benchmark*** com as empresas NCR e SUN. O ***hardware*** NCR utiliza processamento paralelo, é uma máquina dedicada para ***Data Warehouse*** e apresenta uma escalabilidade matemática, ou seja, se para atender 100 usuários será preciso utilizar 5 processadores, para atender 200 serão necessários 10 processadores. Além disso a solução NCR conta com o banco de dados Teradata que também é otimizado para o ***Data Warehouse***. O ***hardware*** da SUN conta o dobro de processadores (64) e máquinas trabalhando em ***cluster*** para garantir alta disponibilidade do sistema. O banco de dados utilizado na solução da SUN continua sendo o Oracle. Os itens mais importantes que estão sendo considerados nessa avaliação são: velocidade de carga, velocidade de consulta e escalabilidade.

5.4.3 Granularidade e Sumariação

Tudo que existe no ***Data Warehouse*** está ao nível de detalhes. Todas as informações geradas pela central telefônica e todos os itens de uma fatura, ou seja, tudo o que é cobrado por linha, por cliente, por produto, etc. Também fica armazenado todo o tráfego. O nível de detalhes, as agregações e a vida útil dessas agregações foram definidas pela equipe de projeto de acordo com as necessidades dos usuários.

O mesmo arquivo que é utilizado para fazer a carga do banco de dados (detalhes) é usado para a geração da agregações. O tempo de vida dessas agregações no ***Data Warehouse*** é maior que o de detalhes, por exemplo, enquanto o detalhe do tráfego fica armazenado por apenas três meses, em função do alto volume de dados, os seus agregados ficam retidos por treze meses. Existem solicitações para que esses agregados fiquem retidos por vinte meses, para permitir um estudo mais amplo das sazonalidades. Se o faturamento no último mês de janeiro, por exemplo, teve o tráfego muito reduzido é importante saber se foi um fato isolado ou isso sempre acontece nessa época. No início essa informação ficava retida por apenas doze meses o que não permitia esse tipo de estudo, então foi alterado para treze mas a necessidade já é maior. O prazo de armazenamento de alguns serviços detalhados é:

- 12 meses para serviços comercializados;
- 12 meses para faturamento detalhados;
- 15 meses para arrecadação;
- 12 meses para dívida;
- 3 meses para tráfego faturado;
- 3 meses para ***market share*** (tráfego) e
- 3 meses para painel de controle.

Apesar de existirem esses prazos definidos, o tempo de permanência não é tão rígido. Passado o prazo estipulado os dados vão para uma outra área em disco. Depois disso eles são transferidos para uma fita magnética. O tempo de permanência, tanto nessa nova área em disco como em fita vai depender do espaço disponível, ou seja, os dados só serão eliminados quando não houver mais espaço para guardar os dados novos.

Estão sendo criados mais agregados por linha telefônica, o que não havia até então. Sabia-se quanto estava sendo gerado de tráfego por localidade, por produto, etc. mas não por linha e surgiu a necessidade de acompanhar melhor esse comportamento. Qualquer alteração no *market share* da empresa, principalmente se for uma redução, requer mais informações e estudos e isso tem sido feito através do *Data Warehouse*.

Os agregados, muitas vezes, são criados pela própria área de sistemas em função dos novos relatórios que são solicitados pela área usuária. Outras vezes o próprio usuário solicita alguma inclusão. Existe um trabalho em andamento para verificar quais os agregados que estão realmente sendo utilizados e por quem. É um controle administrativo de utilização do *Data Warehouse*. Esse trabalho visa otimizar o ambiente, eliminando os agregados não utilizados, fundindo outros que poderiam ser apenas um e separando aqueles que são mais utilizados. Como resultado espera-se um redução do espaço de armazenamento e um aumento de velocidade nas consultas e cargas. No ano de 2000, foi adquirido o produto Syncsort, comercializado pela PHD, que reduz bastante o tempo para geração dos agregados.

5.4.4 Extração, Transformação e Carga

Diariamente são tratados milhões de registros que envolvem tráfego, faturamento e operações comercias. Nesse processo, existe uma etapa intermediária onde os dados são extraídos das tabelas dos sistemas em *mainframe*, tratados e indexados para então serem transferidos para o banco de dados Oracle no *hardware* da SUN.

Uma das principais bases de dados é o sistema CSO (Cadastro de Serviços em Operação – é o sistema de Atendimento ao Cliente – *Customer Care*), que roda em *mainframe*. Para carregar todas as informações, cerca de 15 milhões de registros, o processo de ETL demora algo em torno de 18 horas e é executado a cada final de semana, mas algumas informações como ordens de serviço completadas (cerca de 40.000), são carregadas diariamente. O tempo tende a diminuir bastante quando o Syncsort estiver totalmente implantado. Este produto é usado basicamente para fazer a classificação dos dados (*Sort*) e está sendo implantado em fases para cada assunto do *Data Warehouse*. Por exemplo, o tráfego, que era bastante crítico em termos de tempo nesse processo, teve um ganho médio bastante expressivo, em torno de 50%. Alguns procedimentos que antes demoravam quase quatro horas para serem executados agora estão sendo concluídos em cerca de 10 minutos.

As informações do faturamento são carregadas ao final de cada período de faturamento[26]. De qualquer forma existe uma preocupação em distribuir a carga dos dados para evitar acúmulo de dados em um único dia o que poderia gerar processamentos muito demorados. Para os integrantes do projeto, não faz sentido esperar um mês para carregar as informações do faturamento. O usuário precisa ter a informação disponível no menor tempo possível e se os dados estão disponíveis por quê não carregá-los? Não é necessário esperar um mês para verificar se houve ou não uma queda no faturamento, ou se houve algum erro no processamento desse faturamento. Dessa forma é possível fazer uma análise e ter a

[26] Em cada mês são processados vários períodos de faturamento que representam os ciclos de cobrança. Cada um desses ciclos gera contas telefônicas com dia de vencimento diferente. A quantidade de dados em cada ciclo varia bastante, passando de milhares de clientes nos menores para alguns milhões nos maiores.

expectativa de quanto vai ser o faturamento no mês seguinte nesse período, ou seja, faz-se um modelo preditivo que permite um trabalho proativo.

O usuário informa que por meio dessas análises é possível estimar, com bastante certeza, se o cliente vai ou não pagar a conta no próximo vencimento. O quanto esse cliente paga e não paga também é controlado pelo módulo de dívida e contas a receber. Quando um cliente não está dentro do padrão de pagamento, pode iniciar-se um estudo para propor alternativas que beneficiem ambos os lados. Por exemplo, se o vencimento da conta é no dia 5 e o cliente sempre paga no dia 10, pode ser feita a proposta de se mudar o ciclo de faturamento. Dessa forma o cliente deixa pagar multa e a empresa recebe o dinheiro na data certa.

Além do Syncsort, um outro produto usado no processo de ETL é o ETI-Extract, da Etica Software e comercializado pela IBM. Conforme a área de sistemas, este produto permite mapear os dados conforme a necessidade, por exemplo, o campo X do CSO deve passar por tais filtros e tais transformações e ser carregados no campo Y do Oracle. Os programas para o processamento, ou seja, os programa para extração e carga, podem ser gerados em Cobol, Natural (para acesso ao banco de dados Adabas) ou em linguagem C. Resumindo, a empresa utiliza o Syncsort para a classificação dos dados e ETI-Extract para a geração dos programas que farão o processamento dos dados. Basicamente, o ciclo de ETL poderia ser esquematizado da seguinte forma:

Para a equipe, um item considerado muito importante em todo o projeto é a qualidade dos dados, que vem através do processo de ETL. Existe um grupo de usuários que passa o dia validando os dados do *Data Warehouse*. Com isso, as informações estão cada vez mais próximas daquilo que é a expectativa de resultado. A precisão atual dos dados está em torno de 99%. A área usuária da Telesp diz que na Espanha, o usuário entende que uma qualidade em torno de 85% é suficiente para o *Data Warehouse*.

Houve uma necessidade muito forte para se buscar esse número pois as unidades de negócio foram criadas e os sistemas transacionais não conseguiam acompanhar o volume expressivo de projetos que começaram aparecer e também não existia uma fonte única e integrada de informações qualificadas. A área S*M começou a ser cobrada, por exemplo, pela causa de determinado faturamento não estar de acordo com uma conta emitida, e a partir de então o processo de ETL passou a ser refinado. Verifica-se o tamanho do arquivo de entrada e de saída, valores antes e depois, ou seja, busca-se pontos de controle para garantir que a informação será de boa qualidade. Em determinada ocasião, perdeu-se um registro entre vários milhões e, apesar de ser uma coisa muito pequena, buscou-se a causa para depois decidir se aquele registro era válido ou não. Se for uma situação nova, aquilo que era apenas um pode ser um milhão com o tempo, então é necessário saber o que aconteceu.

Após a validação do processo de ETL, ainda existe uma validação junto aos sistemas transacionais, que são a origem dos dados, para confirmar se aquilo que foi checado de maneira macro, continua valendo em uma análise mais detalhada. Um conta telefônica é localizada no *Data Warehouse* é montada com os itens que se encontram lá. No final o resultado deverá ser o mesmo. O objetivo não é substituir as funções existentes nos sistemas transacionais mas sim garantir que os valores que estão no *Data Warehouse* refletem a mesma

realidade. A medida que erros vão sendo encontrados, analisados e entendidos, o processo vai sendo melhorado. Um outro exemplo começando de maneira macro: os números do faturamento não conferem, então parte-se para uma análise dos itens que compõem esse faturamento (assinatura, serviço medido, interurbanos, tarifas de serviços, etc.). Verifica-se que houve uma discrepância no item assinatura e portanto é necessário saber o que significa esse item. A análise continua até o item mais atomizado (*drill-down*) para verificar se aquilo está de acordo com o sistema de faturamento da Telesp. Se realmente existe um erro ele pode estar tanto no *Data Warehouse* como no sistema transacional.

Além desses processos de qualidade, estão se buscando outros pois ficar comparando duas bases de dados não é tão fácil. A intenção é usar o Syncsort, durante o processo de ETL, para já fazer um batimento dos dados antes da carga.

5.4.5 Metadados

A equipe de *Data Warehouse* está ciente da importância dos metadados para o projeto mas eles foram implementados de uma forma muito simples. Eles são gerados pelas próprias ferramentas durante o processo de ETL.

Os metadados estão sendo usados basicamente para a administração do ambiente, com registros de controles de cargas e controles de usuários. Não estão representando exatamente o conceito de "dados sobre dados". O usuário não utiliza o *Data Warehouse* através dos metadados. Quando ele não conhece bem um dado pede auxílio para a equipe de suporte.

A equipe do projeto tem planos para incrementar a utilização dos metadados no próximo ano. Por exemplo, quando um usuário receber um relatório de faturamento ele terá informações para saber como está o processamento daquele faturamento, ou seja, saberá se aquelas informações são o resultado final ou parcial.

5.4.6 Segurança

Para a equipe a segurança física ainda é um ponto de falha no projeto. Não existe uma política de *backup* implantada até o momento. Os *backups* que ocorreram até hoje foram esporádicos e realizados em circunstâncias específicas, como por exemplo, antes de uma atualização da versão do Oracle. Atualmente está sendo feita a implementação de um sistema de *backup* corporativo que vai atender também ao *Data Warehouse*. O processo a entrar em operação prevê a execução de *backups* diários incrementais, ou seja, vai salvar somente aquilo que foi incluído no dia. Também está previsto um *backup* "*full*" quinzenal que vai salvar todo o *Data Warehouse*. Não existe plano de contingência definido e o *hardware* atual não contempla alta disponibilidade. Em caso de problemas que exijam a reconstrução dos dados as únicas alternativas disponíveis são: a última versão do *backup*, que pode estar muito desatualizado, ou o reprocessamento dos dados que nem sempre é possível.

A área de infraestrutura informa que não há problemas com o espaço físico onde está instalado o *hardware*. Ele fica em um prédio construído exlcusivamente para ser um *Datacenter* e portanto conta com itens como piso elevado, ar condicionado, *no-break*, acesso controlado, etc. Os discos que contém os dados, como já foi dito anteriormente, contam com a

proteção de espelhamento (RAID-1), e garantem um segurança relativa em caso de problemas físicos (quebra de um disco, por exemplo).

Com relação à segurança lógica das informações o controle já é maior. Por definições internas, uma área de negócios não está autorizada a acessar os dados de outra. Foi elaborada uma política de dados para o *Data Warehouse* dizendo quem é o dono da informação. Quem deseja acesso a um determinado dado, precisa pegar um formulário na Intranet da empresa, onde também está disponível a política, e dizer que informação será necessária e para qual finalidade. Existe também um termo de confidencialidade que deve ser enviado, junto com o outro formulário, para a equipe do projeto que faz as autorizações, se for o caso, e disponibiliza a ferramenta adequada (Business Objects, DSS Microstrategy ou SAS).

A área de sistemas informa que já ocorreram problemas com pessoas que conseguiram acessar a base de dados para gerar novas tabelas. Isso aconteceu porque o DSS Microstrategy cria tabelas temporárias durante a geração de um relatório e quem tinha acesso a isso também tinha acesso para trabalhar no banco de dados. Uma nova versão do produto corrigiu essa falha pois agora existe uma única chave interna para esse tipo de acesso que é independente da chave do usuário. De qualquer forma existe um processo intermediário para checar se o registro desejado está realmente liberado. Essas definições de segurança são feitas no Oracle e na ferramenta.

5.4.7 Metodologia

A metodologia usada no desenvolvimento do projeto para a Telesp foi a MDP (Metodologia do Projeto) que é própria do I+D. Essa metodologia é usada pela consultoria para o desenvolvimento de outros tipos sistema, portanto não tem nenhuma particularidade destacada que seja própria para o *Data Warehouse*, segundo a própria consultoria. As fases básicas dessa metodologia, bem como o percentual médio de tempo consumido por cada fase está representado na Tabela 15:

Fase do Projeto	% Tempo
• Definição do Produto	10 %
• Definição de Alto Nível	10 %
• Desenho Detalhado	20 %
• Desenvolvimento	50 %
• Passagem para a produção	10 %
• Suporte em Campo	-

Tabela 15 - Fases da Metodologia
Fonte: elaborado pelo autor

Para a fase de Suporte em Campo não foi considerado um consumo de tempo pois, depois de transferido para a produção, o sistema passa a ter um suporte contínuo.

Nas fases de definição (produto e alto nível) as principais atividades são:

- Criação do modelo conceitual corporativo;
- Identificação dos indicadores de negócio mais importantes;

- Definição das aplicações de Inteligência de Marketing;
- Determinação da arquitetura ótima para o *Data Warehouse*.

Na fase de desenho acontecerá:

- O desenho do modelo de dados (processos, dimensões e atributos, tabelas de fatos, granularidade, agregados, tratamento às dimensões que mudam com o tempo, duração histórica dos dados, etc.);
- O desenho das aplicações de Inteligência de Marketing (modelo de valor do cliente, análise e simulação de tarifas, segmentação baseada em padrões de tráfego);
- O desenho das demais aplicações (tarefas realizadas sobre as ferramentas de usuário final destinadas à criação de relatórios pré-definidos e à realização de análises dimensionais).

No desenvolvimento do *Data Warehouse* da Telesp não foi feito nenhum protótipo ou projeto piloto em função dos prazos agressivos que foram colocados. Após o desenvolvimento o projeto é diretamente implantado o que acaba gerando vários problemas, principalmente nos processo de ETL.

Os requerimentos para o desenvolvimento desse projeto foram baseados principalmente nas necessidades do usuário, que não eram poucas em função do novo cenário que estava se desenhando (concorrência). De qualquer forma, essas necessidades estão sendo atendidas apenas pelos dados que estão disponíveis nos sistemas transacionais

5.4.8 Modelagem de Dados

Apesar de utilizar banco de dados relacional (Oracle) a modelagem de dados adotada para este projeto foi a modelagem dimensional, ou *star schema*, que é o nome mais adotado pela equipe. O termo "estrela" é bem entendido por aqueles que estão participando do projeto. O fato de haver a necessidade de ver os dados de uma forma dimensional não quer dizer que eles precisam ser arquivados nesse mesmo modo. Uma visão dimensional da organização é muito mais relevante para a tomada de decisão que uma visão tradicional fornecida pelos sistemas transacionais. Essa combinação (relacional e dimensional) fornece o benefício da funcionalidade analítica enquanto mantém as vantagens dos bancos de dados relacionais, como arquivamento de dados ao nível de detalhes, performance e compatibilidade com várias plataformas de *hardware*. Em outras palavras, a opção pela modelagem dimensional está associada à performance desejada durante o acesso aos dados bem como na possibilidade de se analisar o negócio por diversas perspectivas e níveis de detalhes.

O modelo de dados dimensional é elaborado na fase de desenho do projeto quando são definidos: dimensões, métricas, fatos e relações. O resultado desse processo é o modelo dados do *Data Warehouse* para Telesp. A ferramenta utilizada para essa modelagem é o DW-Architect. Para acesso a esses dados é usada a ferramenta DSS Microstrategy que se enquadra na categoria da ferramentas ROLAP (*Relational* OLAP).

Exemplos do modelo dimensional são apresentados no Anexo 3.

5.5 Aspectos Humanos

5.5.1 Usuário

No início do projeto o *Data Warehouse* era uma grande novidade não só para a área de sistemas como para a área usuária. A equipe de usuários, na época a área S*M, sabia muito bem usar as ferramentas tradicionais (Access, Excel) para análise dos dados e tinha potencial de crescimento mas não assumiam muitos riscos. O mercado era muito restrito em se tratando de pessoas que conheciam bem, tanto o ambiente de *Data Warehouse* como o negócios de telecomunicações, mais especificamente de telefonia, então o gerente da área decidiu investir no seu próprio pessoal. Ter o usuário capacitado para usar o sistema foi considerado um fator fundamental para o sucesso do projeto. Essa equipe (cerca de 15 pessoas), foi promovida e passou a ser treinada para mudar a forma de atuação. A partir dos dados disponíveis eles deveriam extrair, ponderar, validar e questionar as informações em função dos resultados esperados. Para o usuário, esse é considerado o uso nobre do *Data Warehouse*. O desafio maior estava em associar o uso do sistema e das ferramentas ao conhecimento do negócio que já existia. O usuário conhece os números e o conceito a eles associados. Por exemplo, se na cidade de Bauru existem cerca de 50 mil telefones e a resposta mostra que existem 100 mil, fica fora da expectativa do usuário que esperava algo menor justamente por conhecer o negócio da empresa. O objetivo era que essas pessoas estivessem capacitadas num prazo de seis meses.

Um grande desafio foi eliminar a insegurança da equipe em relação ao novo ambiente. Até então, quando eles precisavam criar um indicador de negócios, saiam contatando várias pessoas para coletar as informações necessárias e isso deixaria de acontecer. Estava sendo dito para o usuário: "a partir de agora você não vai mais fazer isso pois digitar e coletar números é função da máquina". Foi necessária uma mudança de cultura.

Quando a qualidade dos dados chegou ao nível desejado, o convite para utilização do *Data Warehouse* passou a ser feito para as outras áreas da empresa. A VP-S é responsável por grande parte dos relatórios gerados a nível corporativo mas com a criação dos *Data Marts*, cada vez mais, as áreas estão trabalhando de forma independente, conforme suas necessidades.

5.5.2 Equipe do Projeto

A pequena equipe inicial, cuja missão era entender o conceito de *Data Warehouse* e verificar como ele poderia ser aplicado à empresa, foi formada por algumas pessoas das áreas S*M (Marketing) e RS (Sistemas). Após as definições e aprovações iniciais, incluindo orçamento, foi formada uma equipe com pessoas do marketing, de sistemas e do I+D, além da participação eventual dos fornecedores de *hardware* e *software*. Para completar a equipe técnica foi feito um contrato com o CPqD para alocação de pessoas especializadas. Basicamente o organograma original da equipe era o seguinte:

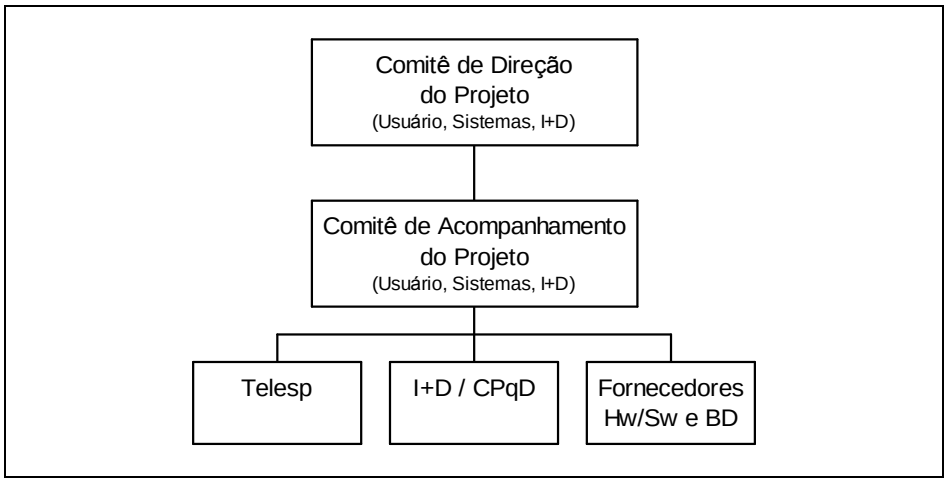


Figura 23 - Organograma básico da equipe
Fonte: documentação do projeto da Telesp

No Comitê de Direção do Projeto estão alocados os altos executivos (patrocinadores) das principais áreas envolvidas. O Comitê de Acompanhamento é encarregado da gestão do dia-a-dia do projeto, fazendo reuniões periódicas que permitem identificar e solucionar problemas. Essas reuniões contam com os membros desse comitê mais aqueles que sejam considerados necessários. As pessoas que iniciaram os estudos de *Data Warehouse* para a Telesp fazem parte desse comitê.

No grupo Telesp se enquadram os seguintes papéis e responsabilidades:

- Responsáveis pelas unidades de negócios para identificar os recursos de suporte ao projeto (analista de negócio, gerente de negócio, diretor de negócio), coordenar o suporte ao projeto;
- Responsáveis pelos sistemas operacionais para identificar os recursos de suporte ao projeto (analistas de dados, técnicos de suporte e infra-estrutura de rede), coordenar o suporte ao projeto.
- Desenvolvedores para o processo de ETL que devem conhecer as bases de dados e definir os processos para extração, transformação e carga dos dados.

No grupo I+D se enquadram os seguintes papéis e responsabilidades:

- Chefe de projeto que é responsável pela direção, coordenação e acompanhamento das atividades do projeto.
- Consultor de negócio para a determinação dos requisitos das aplicações.
- Consultor Técnico para definição dos requisitos de extração.
- Especialista em modelos de dados para definição do modelo lógico de dados.
- Especialista em ETI-Extract para treinar os desenvolvedores no uso da ferramenta.

Além desses, outros papéis, não menos importantes, fazem parte da equipe como por exemplo, especialistas nas ferramentas de acesso (BO, SAS, etc.), administrador de qualidade, etc. No grupo do CPqD se concentram basicamente as pessoas especializadas nos processo de ETL, mas também podem ser alocados outros perfis conforme as necessidades do projeto. O grupo de fornecedores participa, quando necessário, para alguma consultoria especializada ou treinamento em alguma ferramenta ou produto.

A equipe que participa do projeto do *Data Warehouse* pode ser considerada em dois grandes grupos: (1) a equipe de desenvolvimento dos módulos internos do *Data Warehouse*, ou seja, a definição das estrelas na modelagem dimensional, a configuração dos *softwares* de extração, etc. e (2) a equipe de negócios que faz a validação dos programas que a outra equipe estava desenvolvendo e distribui as informações para a empresa. Um exemplo de funcionamento desse esquema é mais ou menos o seguinte: a equipe de desenvolvimento define o processo de ETL para a carga faturamento das linhas individuais, durante cerca de quinze dias vários indicadores vão sendo criados e durante os outros quinze dias do mês a equipe de negócios faz a aferição desses indicadores, item a item, para garantir as informações antes de distribuí-las.

O processo de validação dos dados já não é tão intensivo pois as ferramentas de ETL estão calibradas, os bancos de dados foram "limpos", e os conceitos que ainda não estavam bem claros foram entendidos. O momento está muito mais voltado para o uso intensivo do *Data Warehouse*, através da mineração dos dados, modelos preditivos, etc. Isso não quer dizer que acabou o cuidado com a qualidade dos dados, muito pelo contrário, uma carga errada pode fazer com que o usuário perca a confiança na informação e isso pode por em risco a continuidade do projeto.

Desde o início a equipe está dedicada exclusivamente ao projeto. Só cuidam do *Data Warehouse*, do cronograma, do processo de ETL, da qualidade dos dados, etc. As consultorias contratadas mudam as pessoas conforme a necessidade e especialidade desejada para determinados casos mas a estrutura da equipe e grande parte das pessoas envolvidas, principalmente aquelas que são funcionárias da Telesp, permanecem as mesmas. As alterações na equipe podem ser definidas mais como agregações de pessoas do que como a troca das mesmas. Atualmente a equipe conta com cerca de 30 pessoas dedicadas exclusivamente ao projeto cujos papéis nas diversas fases do projeto poderiam ser resumidos como apresentado na Tabela 16:

	Gerenciamento		Negócios			Desenvolvimento					
Descrição do papel →	Comitê de Direção[27]	Comitê do Projeto	Analista de Negócios	Analista de Suporte à Decisão	Analista de Sistemas (Negócios)	Adm. de Banco de Dados	Arquiteto de Infra-estrutura	Especialista no processo de ETL	Especialista em BO, DSS, SAS	Adm. de Qualidade	Fornecedor de Hw / Sw
Fases do Projeto ↓											
Definição do Produto	SA	SV	P	P	P	P	P	P	P	P	S
Desenho Detalhado	SA	SV	SV	SV	P	P	P	P	P	P	S
Desenvolvimento	SA	SV	SV	SV	P	P	S	P	P	P	S
Passagem para a Produção	SA	SV	S	S	S	P	S	P	P	P	S
Suporte em Campo		SV	S	S	S	P		P	P		S

Tabela 16 - Papéis e responsabilidades no projeto
Fonte: elaborado pelo autor
Legenda da Tabela: (A)ceita - (P)roduz - (S)uporta/Apóia - (V)alida

[27] Nesse Comitê de Direção do Projeto está incluído o patrocinador

5.5.3 Facilidade de Uso e Utilidade

O *Data Warehouse* é fácil e ao mesmo tempo muito difícil de ser usado na opinião da área usuária. Ele é fácil de ser usado quando o foco está voltado para as ferramentas que são muito práticas. Sua utilização é muito difícil quando se fala no conceito de negócios. Em outras palavras, um analista de negócios do setor de telecomunicações pode facilmente entrar em um *Data Warehouse* do setor de varejo e encontrar os dados, mas certamente não saberá usar esses dados para gerar informações úteis já que não conhece esse ramo de negócios. Um "problema" que dificulta a utilização do *Data Warehouse* é que ele fornece uma flexibilidade tão grande que coisas que nunca foram vistas antes, relacionadas aos negócios, começam a aparecer.

Durante uma análise de comportamento do cliente é possível descobrir que um determinado cliente costuma faz vários interurbanos mas não executa chamadas locais. Esse comportamento não é normal e necessita de uma análise mais detalhada, através do processo de *drill-down*, até que se encontre a resposta para o problema. Segundo o usuário, é um processo "cirúrgico" que exige conhecimento do negócio.

A partir do momento que o ambiente está estabilizado e entendido fica fácil vender o sistema para outras áreas da empresa mostrando sua utilidade. Ele pode ser usado para validar vários processos que estão ocorrendo dentro da empresa, e também é muito mais flexível que os sistemas transacionais que estão voltados para o atendimento do dia-a-dia. Além disso reúne as informações relacionadas aos cliente e à empresa em um ambiente único e integrado. Se antes demorava-se um mês para conseguir determinadas informações, isso agora pode ser conseguido em poucos dias ou até mesmo horas. No início da utilização o *Data Warehouse* pode até ser considerado lento e complicado por ter muitas visões do negócio, o que exige mais atenção e conhecimento. Com o passar do tempo a produtividade tende a aumentar.

Quando a ferramenta é instalada para o usuário inicia-se um processo de monitoração da utilização. Quando um novo usuário tem dúvidas, ou sua consulta está demorando muito, a equipe do *Data Warehouse* captura a transação e verifica se o conceito que está sendo utilizado está correto. Atualmente a quantidade de consultas está em torno de 18 mil por mês, sendo que o principal usuário (S*M) executa entre 8 a 10 mil consultas por mês. O horário disponível para consultas, que são feitas através da rede interna, é das 08:00 h até as 18:00 h sendo que o restante do tempo é utilizado no processo de ETL. O período de maior utilização está no começo do mês em virtude da avaliação relativa ao mês anterior.

5.5.4 Treinamento

O pequeno grupo de pessoas que iniciou o projeto aprendeu os conceitos de *Data Warehouse* através da participação em vários seminários sobre o assunto. Mas a maior preocupação estava em capacitar o grupo de usuários que estariam sendo envolvidos pois os técnicos da consultoria, que iam realmente desenvolver o projeto, já estavam treinados.

Para os membros da equipe que eram funcionários da Telesp, foram realizados alguns treinamentos específicos que aos poucos foram sendo repassados para os demais pela própria convivência do dia-a-dia.

Um certo tempo antes a Telesp realizou um treinamento em grande escala para o sistema SAP e então decidiu-se usar essa mesma estrutura (espaço, PCs, etc.), que ainda estava montada, para oferecer vagas de treinamento em *Data Warehouse* para as unidades de negócios. Um argumento utilizado para incentivar os usuários a participarem foi de que os relatórios que estavam sendo emitidos até então pelo S*M poderiam estar sendo feitos por eles mesmos, de forma mais rápida, já que eles também eram conhecedores do negócio.

O usuário informa que o curso foi montado com o conceito de que não seria um simples curso de comandos mas sim um curso através do uso do *Data Warehouse* que já estava em funcionamento. Os comandos seriam usados para extrair informações reais. A intenção era que o usuário percebesse todo o poder das ferramentas em cima de dados que ele já conhecia. Para aplicar o treinamento foi utilizado o pessoal contratado junto ao CPqD que ajudou a desenvolver o projeto pois, além de terem o conhecimento de todo o ambiente e terem sido treinados no exterior, eles também tinham uma bagagem didática pelo tempo que foram instrutores do antigo sistema Telebrás. A intenção era que as pessoas saíssem desse treinamento prontas para fazer uso produtivo do *Data Warehouse*. Um dos objetivos desse treinamento era ter pessoas usando bem o *Data Warehouse* para ajudar a vender o produto. Quando alguém perguntasse "Por quê você usa ?" a resposta deveria ser "Porque a informação está lá e é confiável".

Após o treinamento, e antes de voltar à área de origem, o funcionário ainda ficava um tempo junto com a equipe de *Data Warehouse* fazendo uma espécie de "estágio". Ele estaria executando suas funções normais mas em contato direto com pessoas que poderiam ajudá-lo a esclarecer dúvidas e melhorar o entendimento dos conceitos e ferramentas. Esse tempo era variável e poderia chegar até cerca de um mês. Mais uma vez o foco não estava nos comandos da ferramenta mas em saber se os números que estavam surgindo refletiam aquilo que ele estava esperando. Esse contato mais direto foi importante até para o ajuste e criação de agregados.

Como as ferramentas, às vezes, não são suficientes para atender a todas as necessidades, foi criado um núcleo de analistas que se especializaram em SQL para acesso ao Oracle. O projeto foi baseado em *Star Schema* e o cruzamento entre diferentes dimensões pode não ser tão eficiente apenas com o uso das ferramentas e por isso necessita de algo mais direto, segundo a área de sistemas. Essas pessoas têm um profundo conhecimento do ambiente. No início a grande preocupação estava na busca de resultados, mas com o ambiente mais estabilizado e com a colaboração de uma consultoria da Oracle, a equipe está buscando melhorar a performance do sistema.

5.5.5 Envolvimento e Participação

Após a definição e contratação do projeto junto ao I+D houve uma reunião de *kick-off* onde foram convidados o presidente e os vice-presidentes para uma apresentação do *Data Warehouse* pelo diretor de informática que ressaltou as vantagens competitivas que isso traria e pediu a colaboração de todos no desenvolvimento referente àquilo que fosse responsabilidade de cada área.

A equipe reconhece que o projeto não teria sucesso sem um participação direta e efetiva do usuário que esteve envolvido em todas as fases de definição e desenvolvimento do projeto. Se

ele não estiver presente expondo suas necessidades, dificilmente o sistema será desenhado com "a cara dele". Na verdade os dados, até então operacionais, serão transformados em dados de negócios. A participação foi crescendo a medida que a qualidade das informações aumentava e o usuário adquiria confiança naquilo que estava sendo proposto. Até este momento, a participação do usuário tem sido bastante grande na definição e validação das novas etapas do projeto.

5.5.6 Consultoria

Duas consultorias foram contratadas para participar do desenvolvimento do projeto: Telefónica I+D e CPqD.

O I+D – Investigación y Desarrollo – foi criado pela Telefónica de España em março de 1988 com o objetivo de desenvolver atividades relacionadas à inovação e exploração tecnológica para o mercado de telecomunicações, principalmente para as operadoras do próprio grupo Telefônica.

O I+D foi contratado para atuar em todo o projeto, ou seja, levantamento de requisitos, desenvolvimento, implantação e suporte ao usuário. A experiência deles estava relacionada à implementação de ***Data Warehouse*** na própria Telefónica de España (telefonia fixa) e na Telefónica Móviles (telefonia celular). Projetos semelhantes estão sendo desenvolvidos em outras operadoras do grupo (Argentina, Chile, Peru e também na Telerj Celular).

O CPqD – Centro de Pesquisa de Desenvolvimento em Telecomunicações – nasceu em 1976, através de uma decisão do ex-sistema Telebrás, para suprir as dependências tecnológicas do setor. Em julho de 1998, com a privatização do sistema Telebrás, o CPqD tornou-se uma fundação de direito privado atuando como provedora de soluções tecnológicas para qualquer empresa interessada em seu trabalho.

O contrato com o CPqD tem a duração de um ano, com direito a renovação, e é feito na forma de horas/homem que são utilizadas da forma que a Telesp achar mais conveniente e portanto não ficam vinculadas às etapas do projeto. Sendo assim, elas podem ser utilizadas para atender às novas necessidades que vão surgindo durante o desenvolvimento do projeto e não estavam previstas.

As pessoas, tanto do I+D como do CPqD, podem ser trocadas, ou agregadas, conforme surgem novas necessidades da Telesp ou do próprio projeto, ou ainda, conforme a disponibilidade das mesmas.

Capítulo 6 – Sumário e Conclusões

6.1 Sumário

O desenvolvimento da TI tem proporcionado grandes transformações nos mais diversos mercados. À medida que a distância entre o produtor e o consumidor se tornam menores e a globalização aumenta, a quantidade de negócios realizados cresce bastante e conseqüentemente a quantidade de dados, que são gerados a todo momento. Um grande problema atual não é a falta de dados mas sim a abundância e redundância de dados inconsistentes e desintegrados. Nesse cenário o *Data Warehouse* surge como uma alternativa flexível e eficiente de obter informação através de grandes volumes de dados. Dessa forma é possível a geração de informações estratégicas que podem trazer uma vantagem competitiva para empresa.

O *Data Warehouse* tem se mostrado muito eficiente para o conhecimento do cliente e do mercado em que a empresa atua, mas os projetos dessa natureza costumam demandar altos investimentos e longos prazos de implementação. Por esses fatores, os primeiros a entrar nesse mundo foram os grandes varejistas, os bancos e as empresas de telecomunicações, todos atuando em mercados competitivos, altamente dependentes da tecnologia para conduzir os negócios e com poder para investimentos. De qualquer forma os projetos não são simples e requerem um certo cuidado com vários fatores, sejam técnicos ou administrativos, que podem garantir o sucesso do empreendimento e trazer o retorno esperado.

Para entender melhor esse ambiente esta pesquisa procurou entender "como foi, ou como está sendo feita a implementação de *Data Warehousing*" em uma empresa do setor de telecomunicações. A análise procurou identificar os principais aspectos gerenciais, tecnológicos e humanos envolvidos no projeto. O setor de telecomunicações foi escolhido por ser um setor que utiliza fortemente a TI na condução de seus negócios, e recentemente entrou em um novo cenário de competição. A pesquisa foi feita inicialmente através de uma revisão da literatura que serviu de base para a elaboração do questionário utilizado nas entrevistas que conduziram o estudo de caso. O método do estudo de caso foi escolhido pois permite a análise dos mais variados aspectos do caso estudado, o que facilitou a busca de respostas para as questões levantadas.

Para apresentar o ambiente de estudo é feito um breve histórico do setor de telecomunicações no Brasil bem como uma análise das principais características de um *Data Warehouse* e o relacionamento entre ambos.

O setor de telecomunicações, desde 1877 quando foi instalado o primeiro telefone no Brasil até os dias de hoje, tem passado por muitas transformações. No início existiam milhares de empresas atuando no território nacional e as políticas não eram claras até que em 1967 foi criado o Ministério da Telecomunicações para organizar o setor. Em 1972 foi criada a Telebrás com o objetivo de ser a concessionária geral para a exploração dos serviços públicos. Através da aquisição e absorção de empresas ela passou a administrar 95% dos telefones instalados no país. Depois de estagnações e crises, a retomada do crescimento teve início nos anos 90 e um grande marco ocorreu em 1998 com a privatização do setor. A partir de então teve início uma nova fase marcada pela livre competição, a princípio em nível regional e depois em nível nacional.

O termo *Data Warehouse* surgiu durante os anos 80 por intermédio de Bill Inmon, mas apenas nos anos 90 as empresas passaram a reconhecer o uso estratégico dos dados, ou seja, as informações para a tomada de decisão. A definição clássica diz que "*Data Warehouse* é uma coleção de dados orientada por assuntos, integrada, variante no tempo e não volátil, que tem por objetivo dar suporte aos processos de tomada de decisão". Nesse ambiente os dados são armazenados tanto no formato detalhado, para análises em profundidade, como no formato sumariado que permite maior agilidade nas consultas.

O surgimento do termo *Data Warehouse* aliado a outros novos termos de TI, por vezes, acabam gerando uma certa confusão com conceitos já existentes. O *Data Warehouse* está inserido no ambiente dos sistemas informacionais que é diferente dos sistemas operacionais. Enquanto os sistemas operacionais são voltados para as aplicações (OLTP – *On-Line Transaction Processing*) os sistemas informacionais são voltados para a análise (OLAP – *On-Line Analytical Processing*). Muitas vezes o termo OLAP chega a ser confundido com *Data Warehouse* mas na verdade o *Data Warehouse* é a base para as aplicações OLAP. A confusão continua quando se fala em SIG (Sistema de Informações Gerenciais), SAD (Sistema de Apoio à Decisão) e SIE (Sistema de Informações Executivas). Apesar de muitos pontos em comum, esses sistemas apresentam características próprias que permitem uma distinção.

O *Data Warehouse* tem se mostrado de grande valia para o setor de telecomunicações, que ao contrário de outros segmentos, não tem problema para ter dados dos cliente, pelo contrário, recebe diariamente milhões de registros de chamadas realizadas, etc. O desafio está em armazenar e gerenciar todo esse volume de dados de forma eficiente. Com o novo cenário de competição, o *Data Warehouse* pode ajudar a conhecer o cliente, identificar fatores que levam ao *churn*, reter o cliente, conquistar novos clientes, passar do marketing de massa para um marketing dirigido, identificar quem comprará o quê, etc.

O estudo de caso para análise dos aspectos gerenciais, tecnológicos e humanos envolvidos na implementação de um *Data Warehouse* foi realizado na Telesp (Telecomunicações de São Paulo S.A.) que se tornou uma empresa do grupo Espanhol Telefônica após a privatização ocorrida em 1998. Foram realizadas nove entrevistas individuais sendo duas na área usuária, cinco na área de sistemas, uma na consultoria e uma com a empresa que terceirizou a área de produção de informática. O questionário, apesar de ter algumas características específicas para cada grupo de entrevistados, foi realizado da mesma forma para todos. Isso permitiu analisar qual é o conhecimento de cada um com relação ao projeto como um todo e confrontar informações por diferentes perspectivas. Os principais envolvidos no projeto – gerente da área usuária, gerente de sistemas e gerente da consultoria – participaram desse levantamento o que aumenta a confiabilidade das informações apresentadas e a fidedignidade do método.

As análises e conclusões do caso serão apresentadas a seguir, seguindo a mesma estrutura de tópicos do levantamento bibliográfico e utilizará a comparação de padrões, ou seja, será feita a comparação entre os padrões empíricos com os padrões previstos na literatura.

6.2 Projeto

O projeto da Telesp teve início em 1998, logo após a privatização do setor, ou seja, é um projeto recente. Isso permitiu que a empresa analisasse alguns trabalhos já realizados e aprendesse bastante com os erros e acertos dos pioneiros, principalmente no Brasil. Como o processo de compras da Telesp não permite a escolha direta de um fornecedor, várias empresas foram chamadas para apresentar suas propostas com base em uma RFP (***Request for Proposal***) que foi elaborada. Esse processo permite uma concorrência entre os fornecedores que tende a reduzir o preço final da solução. De qualquer forma existe o risco de contratar um serviço incompleto caso a RFP não especifique detalhadamente aquilo que se deseja adquirir. A partir do momento que a área solicitante aprovar as propostas técnicas dos fornecedores, o critério para escolha do vencedor passa a ser o preço.

Motivação

Sem dúvida o grande motivo para implantação de um *Data Warehouse* na Telesp foi o novo cenário de concorrência que surgiu com a privatização do setor. Até então não havia concorrentes em função do monopólio e por isso a empresa não havia sentido falta de ter um conhecimento maior do cliente. Antes da privatização a empresa já começara a se organizar para o novo mercado mas um projeto dessa natureza não entrou nos planos, até porque não se sabia quem seriam os novos donos e qual a política que seria adotada.

Os novos donos chegaram e sabiam exatamente o que queriam. Queriam um *Data Warehouse*, principalmente para o marketing. Eles já conheciam o conceito e sabiam que ele seria necessário já que o foco da empresa estava mudando da tecnologia para o negócio. A situação estava como a descrita por Singh (1997): falta de dados históricos, dados espalhados por vários sistemas e sistemas não adequados para apoio à decisão. Resumindo, não havia informação de qualidade e, segundo Watson (1998), esse é um dos principais motivos que levam à construção de um *Data Warehouse*.

A mudança de foco para o negócio exigia o conhecimento do perfil do cliente, que de acordo com o Technology Research Institute (Red Brick, 1998) é a prioridade número um para as Telcos em se tratando de *Data Warehouse*. O conhecimento do cliente permitiria analisar seu comportamento e propor soluções que podem a ajudar a diminuir o ***churn*** que tem se tornado um dos principais assuntos no setor. Os motivos para a implantação de um *Data Warehouse* na Telesp não diferem daquilo que é apresentado na literatura a esse respeito com exceção de que um grande incentivo veio de um patrocinador forte que já tinha consciência do que era um *Data Warehouse* e sabia dos benefícios que ele poderia trazer.

Patrocinador

O projeto já nasceu com um patrocinador forte que era o vice-presidente de negócios para o grande público (clientes residenciais). Além disso o projeto também foi apoiado pelo planejamento estratégico da empresa. Esse patrocínio também contou com a área de sistemas que contribuiu para o financiamento do projeto. Essa situação, de patrocínio formado por área de negócios e TI, é considerada a ideal por Watson (1998) cuja pesquisa demonstrou que 49% das empresas adotam essa combinação. Barquin et. al. (1995) também acham importante que

o principal patrocinador seja um executivo da área de negócios para evitar que o projeto seja encarado como um experimento tecnológico ao invés de um investimento estratégico no negócio. De qualquer forma a área de sistemas também acredita bastante no *Data Warehouse*, tanto que está preparando um reestruturação do modelo para melhorar o foco no negócio e facilitar a vida do usuário.

6.3 Aspectos Gerenciais

O gerenciamento do projeto ocorreu, basicamente, por intermédio dos principais gerentes envolvidos, ou seja, da área usuária, da área de sistema e da consultoria contratada. Essas pessoas formavam o comitê de acompanhamento do projeto.

Gerência do Tempo

Bill Inmon disse que a questão sobre o tempo necessário para construção um *Data Warehouse* deve se dividir em duas: quanto tempo se gasta na primeira iteração e quanto tempo se gasta na última iteração. Em outras palavras, apesar de ser um projeto grande e demorado ele deve ser feito por partes. Baker & Baker (1999) dizem que a primeira fase pode ocorrer em poucos meses mas a última pode levar anos. Vários autores dizem que os primeiros resultados devem ocorrer num tempo relativamente curto para evitar que o sistema caia em descrédito.

A equipe da Telesp estava ciente dessa necessidade de resultados em curto prazo e decidiu que cada etapa do projeto teria um prazo máximo de quatro meses para gerar as primeiras informações. Os cronogramas são trabalhados de trás para frente, ou seja, é fixado o prazo máximo de 120 dias e a partir daí o tempo de cada tarefa é ajustado. Esse prazo está dentro da expectativa dos usuários. Para facilitar a primeira entrega foi selecionado um conjunto de dados, que apesar do volume expressivo, poderiam ser trabalhados com uma certa facilidade.

De qualquer forma nem sempre os prazos são exatos. A primeira entrega de resultados não significava resultados com qualidade o que exigia uma participação grande da equipe (sistemas e usuários) para validação das informações. Era uma situação já esperada e portanto também estava prevista no cronograma. O que não estava previsto eram os reprocessamentos e problemas de *hardware* que de uma certa forma acabavam comprometendo algumas datas.

Essa situação leva ao entendimento de que as boas informações demoravam mais que quatro meses para aparecer, mas com os primeiros resultados o usuário já passava a ter alguma coisa para trabalhar, o que era importante para controlar a expectativa.

O contrato com a consultoria é renovado a cada ano, o que vai de encontro ao planejamento e orçamento do *Data Warehouse* que é feito ano a ano.

Gerência dos Riscos

A equipe não fez um estudo detalhado de riscos nesse projeto com identificação, quantificação, desenvolvimento de respostas e controle das respostas, como recomendado

pelo PMI (1996), mas sabia que eles existiam e também quais eram os principais: patrocínio, utilização, e prazos.

Talvez o risco do patrocínio tenha sido o menor de todos pois o projeto já começou com o apoio de quem conhecia bem o conceito e já havia utilizado o produto. Adelman & Oates (1998) dizem que risco de se perder o patrocinador pode ser minimizado com a identificação de um substituto, mas esse não foi o caso. Com o *Data Warehouse* corporativo consolidado, os *Data Marts* passaram a ser financiados pelas próprias áreas de negócios interessadas. Isso significava mais gente envolvida apoiando o projeto, além da própria área de sistemas cuja atuação é bastante forte.

Adelman e Oates (1998) também citam o risco da não utilização que poderia estar sendo minimizado com a participação do usuário desde o começo, e isso realmente aconteceu na Telesp. Além disso houve uma preocupação muito grande com a qualidade das informações que foi considerada um fator de muita importância para garantir a utilização do *Data Warehouse*. O risco da demora na entrega de resultados foi gerenciado com prazos agressivos (até quatro meses) o que permitia controlar a expectativa do usuário. Não foi considerado, num primeiro momento, o risco de baixa performance do sistema. Esse assunto está sendo considerado agora, não tanto como um risco, mas sim como uma melhoria no sistema.

O I+D chegou a apresentar um documento com alguns riscos mas que poderia ser considerado mais como uma lista de premissas para início do proejto do que como uma análise de riscos pois, apesar da identificação, não foi realizado o processo de analisar e responder aos riscos do projeto. Por exemplo, ter a infra-estrutura para iniciar o projeto é mais uma condição necessária do que um risco que pode acontecer durante o desenvolvimento.

Gerência dos Custos

Adelman & Oates (1998) dizem que o orçamento é um recurso chave e isso foi praticamente garantido pela divisão do orçamento entre duas áreas: de negócios e de sistemas. Digo praticamente pois nunca se sabe quando algum fato externo vai afetar o planejamento financeiro da empresa. A confirmação de que esse processo de divisão pode funcionar bem veio logo no início quando a área de sistemas sofreu cortes no orçamento e foram necessários alguns remanejamento de verba para garantir o andamento do projeto.

No início a equipe teve problemas pois estava no terceiro trimestre do ano e não havia sido definido um orçamento para *Data Warehouse* no ano anterior. Como o prazo para início efetivo das contratações (consultoria, *hardware*, *software*, etc.) só ocorreu no ano seguinte, foi possível fazer um planejamento mais adequado dentro dos processos normais da empresa.

Com a terceirização da área de produção de informática em 2001 a Telesp teve que rever seu planejamento financeiro pois muitas coisas que eram investimento passaram a ser despesa, como por exemplo, a área de armazenamento em discos.

Não foram feitos estudos específicos de ROI pois o conceito já estava provado na Espanha e além disso a equipe reconhece a dificuldade de se realizar estudos dessa natureza já que grande parte dos benefícios são intangíveis. Essa dificuldade também é reconhecida por Knowles (1996), Power (1998) e Giurlani (1999). De qualquer forma a empresa exigiu um

estudo para verificar a viabilidade financeira das próximas fases do projeto e alguns casos foram demonstrados, como por exemplo, valores que não haviam sido cobrados em conta telefônica e passaram a ser cobrados ao longo do tempo diminuindo o prejuízo. Também foi colocado o expressivo aumento da venda de serviços agregados que, segundo a área de negócios, foi conseguido através do uso do *Data Warehouse*. Esses são alguns benefícios tangíveis que podem ser mensurados mas de qualquer forma existem muitos outros que não podem, como por exemplo a melhoria no processo de tomada de decisão. Vale ressaltar que esses estudos foram feitos depois que os resultados surgiram para mostrar que o *Data Warehouse* trazia benefícios, ou seja, não foi feita uma estimativa antes do projeto ser iniciado para quantificar os benefícios esperados das próximas fases. Os valores envolvidos no projeto não foram divulgados por se tratar de informação confidencial da empresa.

Gerência da Comunicação

Não havia um plano de comunicação formalmente definido que, segundo o PMI (1996), deve fazer parte da gerência de comunicação do projeto. De qualquer forma, reuniões eram marcadas quinzenalmente para discutir o andamento do projeto. Quando um problema não podia ser resolvido na própria reunião ele era escalado para um nível superior, ou seja, nem todas as informações eram compartilhadas com todos no mesmo momento.

Arnold (1995) diz que a situação ideal para facilitar a comunicação é ter todos trabalhando na mesma equipe. Isso de fato aconteceu pois a equipe da Telesp era formada por usuários, pessoal de sistemas e consultoria trabalhando juntos e todos trabalhavam em um mesmo espaço físico, o que facilitava bastante a comunicação interna.

6.4 Aspectos Tecnológicos

Arquitetura

A arquitetura inicial estava bem entendida e definida já no começo do projeto. A estrutura de *hardware* iria seguir aquilo que já havia sido implantando na Telefónica Móviles, ou seja, uma arquitetura aberta que permitisse compatibilidade com a grande maioria dos *softwares*.

Com relação à estrutura lógica do ambiente, não havia dúvidas que o *Data Warehouse* deveria ser corporativo. Apesar de ele nascer basicamente como um *Data Warehouse* de marketing, não se pensou em nenhum momento em começar por *Data Marts*. Aos poucos foram sendo agregados dados de faturamento, vendas, etc., o que ampliou bastante a funcionalidade do sistema. Segundo Inmon (1999), essa decisão, *Data Warehouse* ou *Data Mart*, tem sido um dos assuntos mais importante para a TI nos últimos anos e a decisão da Telesp foi por adotar a abordagem *Top-Down*. Essa abordagem *Top-Down* (*Data Warehouse* Corporativo **à** *Data Marts*) eliminou o risco apontado por Atre (1997a) de incompatibilidade entre modelos de negócios e definições de dados que cada *Data Mart* pode ter se forem construídos de forma independente.

Uma das desvantagens associadas aos *Data Warehouses* corporativos, que é o longo prazo de implantação, foi contornada com cronogramas agressivos para a entrega dos primeiros resultados. Num segundo momento foram criados *Data Marts* dependentes (esse conceito faz

parte da abordagem *Top-Down*) para atender às unidades de negócio de uma forma mais autônoma. Watson (1998) diz que muitas empresas começam por *Data Mart* para fazer uma "prova-de-conceito" mais isso não foi necessário pois na matriz (Espanha) já havia acontecido.

Hardware

Lang (1997) diz que velocidade, escalabilidade e disponibilidade são fatores muito importantes que devem ser considerados na escolha do ***hardware*** e a equipe da Telesp sabia disso desde o início mas, em função do corte orçamentário, a primeira máquina adquirida ficou abaixo da expectativa.

O ***hardware***, como um todo, está sub-utilizado (40% da capacidade) mas o usuário ainda reclama que algumas consultas são muito lentas. É muito provável que isso possa ser resolvido com um ajuste fino na configuração do equipamento, no banco de dados ou mesmo na aplicação.

De qualquer forma está sendo realizado um ***benchmark*** para verificar o quanto se pode obter de melhoria com uma atualização, ou mesmo troca, do ***hardware***. Os principais itens que estão sendo considerados são: velocidade (de carga e de consulta) e escalabilidade. O aumento acelerado no volume de dados e na utilização pode exigir uma melhora nesses itens para evitar que a janela de ETL seja insuficiente ou mesmo que as consultas fiquem lentas demais no futuro. Caso a análise dos resultados direcione para uma mudança de tecnologia (SUN-Oracle para NCR-Teradata), o sistema será migrado da forma que está, depois será feito um estudo para verificar a necessidade de redesenhar alguma coisa. Até o final dessa pesquisa o ***benchmark*** ainda não estava terminado.

Granularidade e Sumariação

O nível de granularidade no *Data Warehouse* da Telesp é bastante baixo, ou seja, o nível de detalhes é alto. Tudo é guardado ao nível de detalhes. Ao mesmo tempo que isso aumenta as possibilidades de análise, exige muita área física para armazenamento dos dados.

Um módulo que chama a atenção pelo volume expressivo de dados é o de bilhetagem total (CDR – *Call Data Record*) que contém dados detalhados de cada chamada telefônica realizada. O volume atualmente armazenado está por volta de 800 milhões de registros. Esse dado tem sido carregado no *Data Warehouse* em um formato já trabalhado visto que o volume diário que chega na empresa está em torno de 150 milhões de registros com tamanho de 80 ***bytes***. Este volume tende a crescer pois a Telesp está trabalhando de forma acelerada na instalação de telefones. Só para se ter uma idéia, o volume mensal de novos telefones significa um telefone instalado a cada 7 segundos.

Como era de se esperar, os agregados, ou dados sumariados, ficam retidos por um tempo maior que os dados detalhados. A medida que o usuário vai utilizando e entendendo melhor o *Data Warehouse* começam a surgir as necessidades de aumentar a retenção dos dados, principalmente dos sumariados, para aumentar a amplitude de tempo das análises o que permite verificar melhor as sazonalidades.

Um detalhe interessante é que muitos dados estão sendo retidos por mais tempo que o previsto em função do espaço disponível em disco. Se por um lado isso é bom, pode ser um problema quando o espaço for apenas o necessário e a disponibilidade dos dados for, aparentemente reduzida. De qualquer forma o usuário está ciente de qual é a retenção correta para cada tipo de dado.

Os agregados podem ser criados diretamente pela área de sistemas para atender certas necessidades detectadas, ou então solicitados pelo usuário. O crescimento não foi muito controlado e por isso está sendo feito um trabalho para a "limpeza" dos agregados o que acaba contribuindo para a performance do sistema.

Extração, Transformação e Carga

Vários autores enfatizam a importância desse processo e Inmon chegou a dizer que 80% do tempo gasto na construção de um *Data Warehouse* seria gasto em ETL. A quantidade de dados no processo de ETL da Telesp é bastante considerável e algumas cargas chegam a demorar em torno de 18 horas e por isso só podem ser executadas aos finais de semana. Esses números fizeram com que a equipe fosse atrás de alternativas para otimizar o processo e uma dessas alternativas foi adquirir uma nova ferramenta para classificação dos dados, que trouxe ganhos expressivos, mas ainda não está completamente instalada. A velocidade de carga é um dos itens importantes que estão sendo analisados no *benchmark.*

Bohn (1997) diz que a qualidade dos dados depende de se prestar muita atenção em cada atividade da conversão de dados, e isso vem sendo feito pela equipe da Telesp. A qualidade dos dados foi considerada um dos fatores críticos de sucesso do projeto. Sem ela provavelmente o usuário não estaria usando o sistema. Existe um grupo de pessoas dentro da equipe de *Data Warehouse* que passa o dia validando os dados. O objetivo é ter a qualidade muito próxima dos 100%.

Metadados

Apesar da literatura apontar os metadados como um dos principais, senão o principal, componente do *Data Warehouse*, eles não são amplamente divulgados no ambiente da Telesp e não foram considerados um fator crítico de sucesso para o projeto. É óbvio que existem os metadados, mas eles estão sendo mais utilizados para controles administrativos do ambiente do que para orientação dos usuários.

Apesar disso o usuário não tem tido problemas pois, quando não sabe alguma coisa, recorre a uma equipe especialista que fornece todas as orientações necessárias e supre a falta de conhecimento sobre os metadados. Quando o *Data Warehouse* estiver sendo amplamente utilizado dentro da empresa talvez a coisa já não seja tão simples assim e os metadados podem se tornar muito necessários.

Segurança

Podemos analisar a segurança por três perspectivas: ambiente onde estão os dados, os dados em si, e o acesso a esses dados.

Com relação ao ambiente onde estão os dados não existem maiores preocupações pois trata-se de um ambiente de *Datacenter*, com toda a infra-estrutura necessária para manter um sistema de informações da empresa, ou seja, controle de acesso ao ambiente, climatização, etc. De qualquer forma não existe um ambiente de contingência preparado para garantir a continuidade em caso de problemas mais sérios que tornem inviável a operação no local e o *hardware* ainda não é de alta disponibilidade (equipamento redundante, etc.). Os discos oferecem uma segurança relativa com a tecnologia RAID-1 (espelhamento de discos).

Os dados em si são uma preocupação atual. Apesar de um *Data Warehouse* não sofrer atualizações dos dados já existentes, eles precisam ser salvos (*backup*) de alguma maneira pois em caso de problemas físicos, como por exemplo um defeito nos discos, hajam condições de recuperação. Se um problema ocorrer em dados recentes, talvez reprocessamentos (ETL) resolvam a situação, mas isso dificilmente será possível com dados antigos, e mesmo que for, demandará tanto tempo para se voltar ao ponto atual que tornará o processo inviável. A equipe está trabalhando para que a implementação do processo de *backup* ocorra o mais rapidamente possível, enquanto isso são feitos *backups* eventuais de forma manual.

O acesso aos dados têm um controle bem maior. Já parte-se do princípio que uma área de negócios não pode ver os dados da outra. Isso visa que o usuário fique totalmente focado no seu negócio. De qualquer forma qualquer acesso aos dados deve obedecer à uma política que foi elaborada para esse fim.

Um ambiente dessa natureza realmente precisa de um controle de acesso bastante grande pois mantém informações estratégicas da empresa, mas deve ser um controle que não limite demais o acesso aos dados. Deve-se considerar a idéia proposta por Gagnon (1999), baseada na opinião de especialistas, de que o "direito de saber" deve prevalecer sobre a "necessidade de saber" para não minimizar a eficiência do *Data Warehouse*.

Metodologia

Vários autores dizem que apesar do *Data Warehouse* ter uma metodologia com fases – análise, projeto, construção e implementação – assim como ocorre no desenvolvimento tradicional de sistemas, ela apresenta importantes diferenças como por exemplo o desenvolvimento iterativo e os protótipos alavancáveis.

A metodologia utilizada na Telesp é a fornecida pela consultoria contratada. De maneira macro, é uma metodologia como muitas outras utilizadas no desenvolvimento de sistemas. Durante as entrevistas realizadas nesse estudo de caso não foi possível destacar nenhuma característica que possa ser considerada exclusiva para o desenvolvimento de um *Data Warehouse* mas talvez seja possível associar alguma coisa a partir de outras análises.

Por exemplo, a entrega dos primeiros resultados é feita já prevendo-se que será necessário um período para validação. Talvez seja possível dizer que essa primeira entrega seja um protótipo,

pois ele ainda não estará efetivamente em produção enquanto os dados não tiverem a qualidade desejada. A partir daí a equipe (sistemas e usuário) vai trabalhar para "corrigir" aquilo que não ficou bom, ou seja, a partir de um ponto de partida (protótipo) faz-se os acertos necessários para disponibilizar o sistema sem a necessidade de mudar o ambiente ("protótipo alavancável"). Esse período de validação (coletar requerimentos – validar – coletar requerimentos – validar – ...) mostra que também existe a iteratividade no processo.

A iteratividade também se mostra na estratégia de desenvolvimento que foi adotada pela consultoria:

- Definir, a partir da situação existente, os negócios que serão incorporados ao sistema;
- Definir o alcance das aplicações dessa fase;
- Definir a iteração do sistema cujo alcance foi especificados nos itens anteriores;
- Definir as linhas gerais de evolução do sistema (dados, negócios, ferramentas, etc.);
- Continuar o sistema com sucessivas iterações que seguirão o esquema anterior.

Esse esquema se enquadra naquilo que foi proposto por Ballard & Herrreman (1998) e está representado na Figura 10 (Capítulo 3 – item 3.3.7).

Não ficou claro se, para o desenho do *Data Warehouse*, foi adotada uma abordagem baseada nas necessidades do usuário ou nos dados disponíveis. Na verdade ocorreu um processo híbrido onde foram considerados tanto as necessidades como os dados existentes, ou seja, é feito aquilo que é necessário mas considerando os dados existentes.

Modelagem de Dados

O banco de dados do *Data Warehouse* é o Oracle, que é um banco de dados relacional, mas para poder usar as funcionalidade do OLAP foi utilizada uma modelagem dimensional. Dessa forma é possível obter visões do negócio por meio de várias dimensões. Para que haja a integração entre esses ambientes, relacional e dimensional, é utilizada uma ferramenta ROLAP.

O conceito de "estrela" (*star schema*) é bem entendido pela equipe e pelos usuários mais experientes. A visão "simples" dos dados que esse modelo fornece facilita o trabalho para definição de novas estrelas que se mostrem necessárias para atender a novas necessidades.

6.5 Aspectos Humanos

Usuário

Singh (1997) diz que poucas pessoas de negócios querem tornar-se especialistas em informática para poder usar os sistemas e por isso eles precisam de ferramentas que associem termos comuns de negócios com as definições dos dados. Os primeiros usuários do *Data Warehouse* conheciam muito bem o negócio mas até então tinham dificuldades para acessar as informações de forma integrada e faziam uso de ferramentas em PC, como Excel e Access, que não permitiam o tratamento de grandes volumes de dados. O gerente da área, que já fazia

parte da equipe de *Data Warehouse*, resolveu investir nesse pessoal pois acreditou no seu potencial. A intenção era ter todos atuando como analistas de SAD.

Para o *Data Warehouse* dar certo foi necessário mudar a forma de atuação das pessoas. Elas precisariam ser mais críticas, questionando as informações obtidas em função dos resultados esperados. Isso vai de encontro com o que disse Inmon (1997) sobre a postura de um analista de SAD: "me dê o que eu digo que quero e, então, eu poderei lhe dizer o que eu realmente quero". Isso significa explorar o sistema a partir de algum resultado. Os curtos prazos para a entrega dos primeiros resultados favorece esse tipo de atuação.

Nos usuários da Telesp é possível identificar os quatro perfís definidos por Inmon (1998): os turistas, que seriam aqueles usuários que conhecem muito bem o negócios e usam o *Data Warehouse* eventualmente para conseguir informações; os fazendeiros, que fornecem os relatórios gerenciais periódicos para toda a empresa; e finalmente temos aqueles usuários que conhecem muito bem o negócio e o sistema – *power users* – e podem ser os exploradores e/ou mineradores.

Equipe de Projeto

A equipe de projeto da Telesp está constituída com os elementos-chave que são considerados pelo PMI (1996) e por Singh (1997): gerente do projeto, cliente ou usuário, patrocinador, analistas de negócios, suporte ao usuário final. Além disso existe a participação, não tão direta dos fornecedores de *hardware* e *software*.

Essa equipe, que hoje é constituída por cerca de 30 pessoas, pode ser considerada em dois grandes grupos: o grupo de desenho e desenvolvimento, e o grupo de validação. Essa divisão não descaracteriza a equipe mas fornece uma estrutura que permite definir mais claramente os papéis e responsabilidades de cada membro. Esses dois grupos também se enquadram no conceito proposto por Imhoff (1999), onde o primeiro seria responsável pela entrada dos dados e o segundo faria a saída das informações para a empresa. No entanto, deve se considerar uma intersecção entre esse dois grupos para a busca da qualidade pois é um trabalho conjunto (geração e validação dos dados). A dedicação exclusiva ao projeto permitiu que os membros da equipe adquirissem um grande conhecimento técnico e do ambiente, além de uma facilitar a integração do grupo.

Facilidade de Uso e Utilidade

Arnold (1995) diz que, mesmo que o projeto seja muito bom, ele será de pouco valor para os usuários de não for percebido como útil e fácil de usar. Essa foi uma preocupação da equipe do projeto que considerou, por exemplo, a qualidade dos dados, como um risco ao sucesso do projeto. Mas por quê a qualidade ? Porque se os dados não tiverem qualidade, ou seja, não forem confiáveis, o usuário não vai acreditar no sistema e com isso tende a deixar de usá-lo. A participação do usuário desde o início do projeto é outro ponto importante para estimular o uso contínuo e crescente.

O *Data Warehouse* é considerado fácil de usar quando se fala das ferramentas que são relativamente simples e práticas mas considerado difícil de usar quando se fala no conceito de

negócios. E isso tem sua lógica, pois só é possível extrair informações úteis quando se conhece bem o ambiente de trabalho, ou seja, tanto os dados como o contexto em que os dados estão inseridos devem ser conhecidos. Quando o usuário não conhece bem os dados ele poderia estar recorrendo aos metadados que auxiliam muito nesse sentido mas isso não acontece. Ao invés disso o usuário recorre a uma equipe de suporte que conhece muito bem os dados. É um alternativa prática mas que pode sobrecarregar essa equipe caso o uso comece a crescer rapidamente.

Os benefícios que uma análise bem feita desses dados traz pode ser percebido em aplicações simples como por exemplo a emissão de uma mala direta. Num processo normal, um grupo de clientes seria selecionado aleatoriamente, talvez agrupados por região, para a divulgação de uma nova campanha. Ao invés de selecionar 100 mil nomes, e ter uma taxa de sucesso de 1%, o sistema permite fazer análises e filtros que limitem bastante essa quantidade de nomes mas cuja probabilidade de sucesso, em função do produto que está sendo oferecido, seja bem maior. Alguns benefícios diretos nesse caso são: a redução de custo com tarifas de postagem, menor tempo de campanha, aumento da receita em função da adesão por um maior número de pessoas.

Treinamento

A grande preocupação, com relação ao treinamento, estava voltada para o usuário final já que os técnicos das consultorias, que iriam desenvolver o projeto, já estavam treinados assim como os funcionários da Telesp que já faziam parte da equipe. Era um ambiente novo para os usuários e, segundo Martinsons & Chong (1999), se um ambiente favorável foi criado, o treinamento pode ajudar no processo de mudança.

Segundo a IBM (1999), os programas de treinamento devem ser desenvolvidos não com o objetivo da tecnologia por trás das telas mas, de preferência, com objetivo na forma de usar o sistema pelo ponto de vista do negócio e esse foi o enfoque dado aos treinamentos de *Data Warehouse* na Telesp.

Os treinamentos foram montados de forma a utilizar o ambiente real de trabalho, com isso o usuário já saia de lá com uma boa noção do poder das ferramentas em cima de dados, que a princípio, ela já conhecia. Depois do treinamento inicial, o usuário ainda permanecia um tempo junto à equipe do projeto fazendo um tipo de "estágio" para solidificar os conhecimentos adquiridos.

Um fato positivo para que os treinamentos dessem os resultados esperados foi a presença dos técnicos do CPqD. Eles tinham bastante conhecimento técnico, participaram do desenvolvimento do ambiente, tinham uma bagagem didática e aplicaram os treinamentos. Dessa forma praticamente todas as dúvidas poderiam ser resolvidas rapidamente.

Envolvimento e Participação

O usuário teve uma participação direta no projeto desde o início. A pequena equipe inicial já contava com o gerente da superintendência de marketing. Essa participação foi muito importante para o sucesso do projeto pois os usuários conhecem bem os dados, ou melhor, conhecem o negócio.

Segundo o modelo de Martinsons & Chong (1999), essa participação do usuário, que ocorre desde a fase de planejamento de um SI, é dita proativa. Isso permite evitar, ou pelo menos diminuir, dificuldades potenciais e além disso o usuário pode ser preparado para a mudança. Dificilmente o sistema será implantado de forma diferente daquilo que o usuário espera.

Consultoria

Conforme as considerações de Adelman & Moss (2000), um projeto pode contar com a participação de consultores e contratados, onde os consultores supririam a falta de experiência da equipe e os contratados supririam a falta de pessoal habilitado. Como a equipe interna não tinha experiência anterior no assunto, a participação das consultorias foi de extrema importância no desenvolvimento do projeto.

No projeto da Telesp encontram-se os dois casos: o I+D que atua no papel de consultoria e o CPqD no papel dos contratados. O I+D, que já tinha experiência em outros *Data Warehouses* no setor de telecomunicações, atua fortemente desde o levantamento dos requisitos até a implementação e suporte ao usuário. O pessoal do CPqD tem grande conhecimento técnico e é alocado principalmente para atuar no processo de ETL mas também foram muito importantes no treinamento do pessoal. Vale lembrar que o I+D, apesar de atuar com uma consultoria externa (contrato, etc.), faz parte do grupo Telefônica.

6.6 Considerações Finais

Apesar dos problemas iniciais do projeto, como o *hardware* restrito em função do corte de orçamento e a baixa qualidade inicial dos dados, o projeto nunca precisou ser refeito. O início do processo foi extremamente desgastante, tanto para a equipe de negócios como para a equipe de sistemas, principalmente após a emissão dos primeiros relatórios quando se percebeu que os números não estavam corretos. Iniciou-se, então, uma fase de questionamentos, ajustes, revisão de programas e conceitos para aprimorar a qualidade dos dados. Hoje o sistema funciona bem.

O *Data Warehouse* é um projeto estratégico para a Telesp e espera-se que em breve ele seja o único meio para o usuário ter acesso às informações. Os sistemas operacionais seriam destinados exclusivamente para operações do dia-a-dia. Para que isso aconteça, a área de sistemas está planejando uma reestruturação no modelo do *Data Warehouse* que pretende retirar alguns dados e acrescentar outros de forma a aumentar o foco no negócio. Essa reestruturação não está amarrada ao resultado do *benchmark*, mas caso decida-se por uma mudança de tecnologia os processos poderão acontecer em paralelo, cujo resultado final pode unir maior flexibilidade com maior performance. O *Data Warehouse* de Interconexão deve se

integrar ao *Data Warehouse* de Marketing formando uma única base de dados, ratificando o conceito de ambiente corporativo.

Também existem planos para fazer algum tipo de atualização *on-line* nos *Data Marts* para fazer uma gestão de eventos. Por exemplo, se um cabo telefônico foi rompido, o defeito é identificado por uma reclamação ou pelo gerenciamento da rede. Com essa informação atualizada é possível identificar todos os clientes afetados pelo defeito e fazer uma ação de marketing para antes mesmo que eles percebam o problema.

O *Data Warehouse* da Telesp é reconhecido dentro do grupo Telefônica como um projeto de sucesso. No passado as operadoras do grupo que atuam na América Latina tinham que ir até a Espanha para verificar o que estava sendo feito e hoje esse pessoal é encaminhado para conhecer o sistema implantado no Brasil. Se for relativo a telefonia celular ainda é na Espanha, mas com relação à telefonia fixa o modelo está no Brasil, ou melhor, na Telesp.

6.7 Propostas para novos estudos

Um assunto que chamou a atenção durante todas as entrevistas foi a ênfase dada para a necessidade de ter qualidade nos dados que estão *Data Warehouse*. Não há dúvida de que isso é importante, mas na Telesp espera-se que esse a qualidade seja de praticamente 100%. Também foi informado que na matriz do grupo um índice em torno de 85% já seria aceitável. Nesse sentido, seria interessante avaliar a questão da qualidade dos dados em outras empresas que fazem uso de um *Data Warehouse* com o foco voltado no processo da qualidade. Será que todas consideram a qualidade da mesma forma ?

O grande cenário que incentivou o *Data Warehouse* na Telesp foi a concorrência que passou a existir após a privatização do setor. Seria interessante pesquisar as outras empresas desse segmento, seja de telefonia fixa ou celular, para verificar o que elas estão fazendo com relação ao *Data Warehouse*. Quantos o consideram estratégico para conquistar e manter o cliente nesse novo mercado altamente competitivo ?

Glossário

Este glossário inclui alguns termos comuns usados em gerenciamento de projetos e *Data Warehousing* (PMI, 1996; Wang, 1998; Meta, 1998; Inmon, 1997; Kimball, 1998)

Acesso – a operação de busca, leitura ou gravação de dados em uma unidade de armazenamento.

Agregados – resumos pré-calculados ou pré-armazenados que são mantidos no *Data Warehouse* para melhorar o desempenho da consulta.

Ambiente Batch – Um modo de processamento predominantemente seqüencial. No modo batch as informações de entrada são reunidas e armazenadas para posterior processamento.

Aplicação – uma coleção de programas de software que permite ao computador realizar tarefas de alto nível como planilhas eletrônicas, gerenciamento financeiro, controle de estoque, folha de pagamento, orçamento, etc. Um conjunto de algoritmos e dados interligados para proporcionar suporte a uma necessidade organizacional.

Área estratégica de negócios – segmento atual ou potencial de mercado em que existe "espaço de mercado".

Backup – um arquivo que serve de base para a atividade de recuperação de um banco de dados. Consiste de um instantâneo de um banco de dados em algum momento anterior.

Banco de Dados – um conjunto de dados inter-relacionados armazenados segundo um esquema. Um banco de dados pode atender a uma ou várias aplicações. É também um mecanismo de armazenamento de dados que elimina redundância e conflito entre múltiplos arquivos de dados. Os dados são introduzidos uma vez e depois ficam disponíveis a todos os programas que deles necessitarem.

Cliente/Servidor – um relacionamento entre programas executando em máquinas separadas numa rede de computador. O servidor fornece serviços; o cliente consome serviços. Um servidor HTTP proporciona páginas Web acessíveis a muitos clientes simultaneamente.

Competência tecnológica – é a capacidade de obter e deter o conjunto de conhecimentos e instrumentos que se aplicam a uma área de atuação.

Computador – máquina eletrônica que realiza cálculos matemáticos ou lógicos de alta velocidade ou que monta, armazena, correlaciona, ou também processa e entrega informações derivadas de dados codificados de acordo com um programa predeterminado.

Controle – o processo de comparar a performance atual com a performance planejada, analisando as variações, avaliando as possíveis alternativas e tomando as ações corretivas apropriadas quando necessário.

Dado – a matéria-prima para os consumidores de informação criarem a informação.

Dado Agregado – o dado que resulta da aplicação de um processo para combinar elementos de dados. Dado que é acessado de forma coletiva ou sumariada.

Dado Atômico – elementos de dados que representam o mais baixo nível de detalhe.

Dados de Negócios – informação sobre pessoas, lugares, coisas, regras de negócios e eventos que é usada para operar os negócios.

Drill Down – um método de explorar dados detalhados que foram usados na criação de um nível sumariado de dados. Os níveis de *Drill Down* dependem da granularidade dos dados no *Data Warehouse.* Significa adicionar ou substituir um cabeçalho de linha em um relatório para aumentar o nível de detalhe das linhas do conjunto de respostas.

Drill Up – ao contrário do *Drill Down*, significa remover ou substituir um cabeçalho de linha em um relatório para reduzir o nível de detalhe das linhas do conjunto de respostas.

DSS (*Decision Support System*)– ver **SAD**.

Efetividade – é a relação entre os resultados alcançados e os objetivos propostos ao longo do tempo.

Eficácia – é a contribuição dos resultados obtidos para o alcance dos objetivos da empresa.

Eficiência – é a otimização dos recursos utilizados para obtenção dos resultados.

Escalabilidade (*Scalability*) – a habilidade de adaptação para suportar maiores ou menores volumes de dados e mais ou menos usuários. A habilidade para aumentar ou diminuir tamanho ou capacidade com o mínimo de impacto no custo do negócio.

Estratégia competitiva – voltada para analisar a forma de situar a empresa em um mercado competitivo, com destaque às formas de conseguir vantagens competitivas destinadas a assegurar sobrevivência.

Estrutura organizacional – é o delineamento interativo das responsabilidades, autoridades, comunicações e decisões dos indivíduos em cada unidade organizacional, da tarefa mais simples à abordagem mais ampla e importante, com suas funções e a relação de cada parte para com as demais e a empresa como um todo.

Fator Crítico de Sucesso – determinada atividade cujos resultados favoráveis são necessários para o alcance dos objetivos.

Gerenciamento de Projeto – a aplicação de conhecimento, habilidades, ferramentas e técnicas para atividades de projeto de forma a atender ou exceder as necessidades e expectativas dos interessados nesse projeto.

Globalização – é a tendência dos negócios a cruzarem os limites internacionais, tanto em marketing de produtos e serviços quanto em movimentação de matérias-primas e produtos acabados.

Granularidade – o nível de detalhe contido em uma unidade de dados. Quanto mais detalhe houver, mais baixo o nível de granularidade. Quanto menos detalhe houver, mais alto o nível de granularidade.

Heurística – o método de análise no qual o próximo passo é determinado pelos resultados do atual passo de análise. Usado no processamento de suporte à tomada de decisões.

Informação. Dado que foi processado de tal forma que pode melhorar o conhecimento da pessoa que o recebe. Dados que os seres humanos assimilam e interpretam para resolver um problema ou tomar uma decisão.

Integração de sistemas – o processo de implantar sistemas distintos de diferentes fornecedores a fim de que trabalhem juntos com um sistema de informação coeso para uma organização específica.

Meta – é o passo ou etapa perfeitamente quantificada e com prazo e responsável definidos para alcançar os desafios e objetivos da empresa.

Metadado – dados sobre dados. A descrição da estrutura, conteúdo, chaves, índices, etc. de dados. Qualquer dado mantido para suportar as operações ou o uso de um *Data Warehouse*.

Metodologia – um sistema de princípios, práticas e procedimentos aplicados a um ramo específico do conhecimento.

Modelos – são formas de representação gráfica da realidade. Provocam o debate de eventuais caminhos alternativos para o futuro.

Objetivo – é o alvo ou situação que se pretende alcançar.

OLAP – *On-Line Analytical Processing*. Processamento analítico on-line. Um termo que contrasta com OLTP. Um conjunto de princípios com definição imprecisa que fornece uma estrutura para suporte a decisão.

OLTP – *On-Line Transaction Processing*. Processamento de transações on-line. A descrição original de todas as atividades e sistemas associados à inserção de dados de forma confiável em um banco de dados. O termo pode ser usado para descrever um ambiente de processamento de transações.

Organização – é a capacidade de ordenação, estruturação e apresentação de um sistema, de um projeto, de um trabalho e dos recursos alocados.

PC – computador pessoal projetado para satisfazer aos requisitos pessoais de um único usuário. O primeiro PC comercialmente disponível foi o Altair em 1975. Em 1981, tanto a Apple Computers quanto a IBM tinham entrado neste novo mercado. Atualmente, PCs e software de PCs continuam a ser o segmento de mercado que mais depressa cresce na indústria de TI.

Planejamento – é a capacidade de diagnosticar e analisar situações atuais, de articular objetivos de forma integrada aos da empresa como um todo e de delinear estratégias – inclusive alternativas – para alcançar estes objetivos, bem como de políticas que servem de sustentação a este processo.

Processo – é um conjunto de atividades seqüenciais que apresentam relação lógica entre si, com a finalidade de atender e, preferencialmente, suplantar as necessidades e expectativas dos clientes externos e internos da empresa.

Programa – conjunto ordenado de instruções ao computador que resultam na conclusão de uma tarefa, tal como um processamento de fatura. Os programas podem ser escritos em qualquer linguagem de programação.

Projeto – é o trabalho a ser executado com responsabilidade de execução, resultados esperados com quantificação de benefícios e prazo para execução preestabelecidos, considerando os recursos humanos, financeiros, materiais e de equipamentos, bem como as áreas envolvidas necessárias a seu desenvolvimento. Um esforço temporário necessário para criar um produto ou serviço. Uma organização de pessoas dedicadas a um propósito ou objetivo específico.

Risco – uma ocorrência discreta que pode afetar o projeto para melhor ou pior.

SAD (Sistema de Apoio à Decisão) – o processo de reunião, estruturação, manipulação, armazenamento, acesso, apresentação e distribuição de informações de negócios em tempo oportuno.

SGBD – Sistema Gerenciador de Banco de Dados. Um sistema de software baseado em computador e utilizado para conservar e gerenciar dados.

Sistema – é um conjunto de partes interagentes e interdependentes que, em conjunto, formam um todo unitário com determinado objetivo e efetuam determinada função.

Slice & Dice – a descrição padrão da habilidade de acessar uma *Data Warehouse* por meio de qualquer de suas dimensões de forma igual.

Tecnologia de Informação (TI) – a força fundamental para reformular organizações aplicando investimentos em informática e comunicações, a fim de promover vantagem competitiva, serviço a clientes, e outros benefícios estratégicos. É um tripé tecnológico que congrega a tecnologia e a evolução dos computadores, telecomunicações e a automação de tarefas. Além dos aspectos puramente técnicos, sua concepção atual também busca a integração com os aspectos sociais da empresa.

Usuário – uma pessoa ou processo que emite comandos e mensagens para o sistema de informações.

Vantagem competitiva – vista pelo consumidor, representa aquilo que faz com que o produto ou serviço seja aceito entre alternativas de compras.

Anexo 1 – Roteiro de Entrevistas

Roteiro de Entrevistas

Esse roteiro servirá apenas como um referencial/guia para as entrevistas, sendo que a ordem e forma dos questionamentos bem como a inclusão e exclusão de itens serão feitos conforme a necessidade do entrevistador, respostas dadas e perfil dos respondentes.

- **Projeto**
  - Início
    - Quando começaram os estudos para *Data Warehouse* ?
    - Quem deu o passo inicial ?
  - Motivação
    - Quais eram os cenários e necessidades iniciais ?
    - O problema de negócios estava definido ?
    - Todos entendiam o problema da mesma maneira ?
    - Quais eram as metas e objetivos ?
    - Quem definiu os objetivos (TI, usuários, ambos) ?
  - Patrocínio
    - Como o projeto foi exposto para a alta direção ?
    - Quais as justificativas apresentadas ?
    - Quem (área / nível hierárquico) e quantos eram/são os patrocinadores ?
    - Quem financiou o projeto ?

- **Aspectos Gerenciais**
  - Tempo
    - Como foram definidos os cronogramas ?
    - Como estão sendo controlados ?
    - Estão sendo cumpridos ? Se não, quais as conseqüências ?
    - Existe um término previsto ?
  - Risco
    - Houve alguma análise de risco ? Qual ? O que foi considerado ?
    - Quais as medidas para eliminar/minimizar os riscos ?
    - O que fazer se ocorrer ?
  - Custos
    - Como foi definido o orçamento do projeto ?
    - Como está sendo controlado ?
    - Houve algum estudo de custo/benefício ?
    - Foi ou está sendo feito algum estudo dessa natureza ? De que forma ?
  - Comunicação
    - Havia um plano de comunicação bem definido ?
    - Era entendido por todos os envolvidos ?

- **Aspectos Tecnológicos**
  - Arquitetura
    - Como o *Data Warehouse* é definido/entendido pela empresa ?
    - Qual foi a arquitetura planejada ?
    - Que tipo de dados está no *Data Warehouse* ?
    - Que critério foi utilizado para a seleção dos dados ?
    - Qual o tamanho inicial e atual (GB/TB) do *Data Warehouse* ?
    - Em que velocidade o conteúdo está crescendo/modificando ?

- Quais as aplicações existentes ?

- Hardware
  - Como foram considerados e qual a importância dos itens: velocidade, escalabilidade e disponibilidade ?
  - Qual(is) o(s) produto(s) utilizado(s) ? Por quê ?
  - Os equipamentos são todos novos ? Adquiridos só para esse projeto ?
  - Foi considerada alguma taxa de crescimento ? Qual ? Para quanto tempo ?
  - O Hardware é otimizado para uso de *Data Warehouse ?*
- Granularidade
  - Qual o nível de granularidade adotado ?
  - Esse nível é constante o muda conforme a idade do dado ?
  - Quem definiu o conteúdo dos dados sumariados ? Qual o critério utilizado ?
- ETL
  - Qual o tempo gasto no processo ?
  - Qual o volume de dados na entrada e na saída do processo ?
  - Qual a periodicidade do processo ?
  - Como garantir qualidade, integridade e consistência dos dados no processo ?
  - Quais as principais características do processo de extração ?
- Metadados
  - Qual o conteúdo dos Metadados ?
  - Como foram definidas as regras de negócio ?
  - Quais são as fontes de dados ? (formais, informais, técnicos, negócios)
- Segurança
  - Como é feita a segurança física ? (*backup*, redundância)
  - Como é feita a segurança lógica ? (acesso aos dados, criptografia)
  - Existe algum esquema especial de contingência ?
- Metodologia
  - Foi utilizada alguma metodologia no projeto ? Qual ?
  - Houve algum protótipo / projeto piloto ? Por quê ?
  - O desenvolvimento foi baseado no usuário ou nos dados ? Por quê ?
- Modelagem de dados
  - Qual o modelo de dados utilizado ? Por quê ?
  - Qual(is) o(s) produto(s) utilizado(s) para SGBD? Por quê ?
  - Existe alguma limitação ou vantagem destacada ?

q **Aspectos Humanos**

- Usuário
  - Quem está utilizando o *Data Warehouse* ? Qual o perfil ?
  - Quantos são os usuários ? Qual a simultaneidade ?
- Equipe de Projeto
  - Qual a composição da equipe de projeto ? (quantidades, perfis)
  - Que critério foi utilizado para essa composição ?
  - A equipe ficou dedicada *full-time* para o projeto ?
  - Qual era o conhecimento/experiência inicial da equipe ?
  - Houve mudanças na equipe (estrutura/pessoas) ?
- Facilidade de Uso e Utilidade
  - O usuário percebe a utilidade do *Data Warehouse* ?
  - Como a utilização é medida e avaliada ?
  - Qual o período de utilização ? Existem picos ?
  - Como está sendo feito o acesso ? (Rede, Intranet, Internet)
  - Houve um aumento de produtividade após a implementação ?
  - Quais os resultados encontrados/percebidos ?

- O sistema é fácil de usar ? Por quê ?
- Os usuários conhecem bem o conteúdo do *Data Warehouse* ?
- Existe algum critério para determinar a satisfação do usuário ?

- Treinamento
  - Quem realizou treinamento e em quê?
  - Como o treinamento foi aplicado ?
  - Houve alguma replicação interna pelo pessoal já treinado ?
  - Em que fases do projeto ?
  - O treinamento foi técnico, de utilização ou ambos ?
- Envolvimento e Participação
  - Em que fase do projeto o usuário foi envolvido ?
  - Sua participação foi ativa ou passiva ?
- Consultoria
  - Houve a participação de consultores externos ? Quem ? Quantos ?
  - Qual a experiência dos consultores ?
  - Em qual(is) área(s) eles atuaram ?
  - Qual foi o tempo da consultoria ?
  - Houve a contratação temporária de pessoas ?

q **Outros**

- Aspectos importantes
  - Que fatores críticos de sucesso foram considerados ?
  - Que obstáculos foram considerados/identificados ?
  - Como está a situação atual ?
  - Quais os próximos passos ?

Anexo 2 – Endereços na Internet

- **BillInmon.com**
  http://www.billinmon.com
  Web Site oficial de Bill Inmon, um dos líderes em *Data Warehousing*. O site contém uma biblioteca com, artigos, *white papers*, bibliografia e um glossário de termos usados em *Data Warehousing*.

- **Ralph Kimball Associates**
  http://www.rkimball.com
  Ralph Kimball, outro pioneiro em *Data Warehouse* e proponente da abordagem dimensional para desenhar grandes *Data Warehouses.* O site contém vários artigos, revisão de livros e *links* para serviços de consultoria.

- **Stanford Data Warehousing Publications**
  http://www-db.stanford.edu/warehousing/publications.html
  O site oferece uma lista de publicações compiladas como parte do projeto de *Data Warehousing* da Universidade de Stanford. A lista é organizada por assuntos tais como, extração de dados, consistência do *Data Warehouse*, manutenção do *Data Warehouse*, etc.

- **The Data Warehousing Information Center**
  http://www.dwinfocenter.org/
  Os *links* apontam para publicações de *Data Warehousing* e apoio à decisão (artigos, *white papers*, livros, avaliações técnicas, periódicos) e outros serviços de interesse (consultores, treinamento, conferências).

- **Data Warehouse Report**
  http://datawarehouse.dci.com/
  O site contém vários artigos sobre *Data Warehouse* e *Data Mining*. O foco está na tecnologia, aplicações de negócios e tecnologias. É mantido pelo DCI.

- **DM Review**
  http://www.dmreview.com/
  O site da DM Review oferece o conteúdo da DM Review Magazine bem como artigos disponíveis somente na forma *on-line*, listas de recursos, fóruns de discussão, etc.

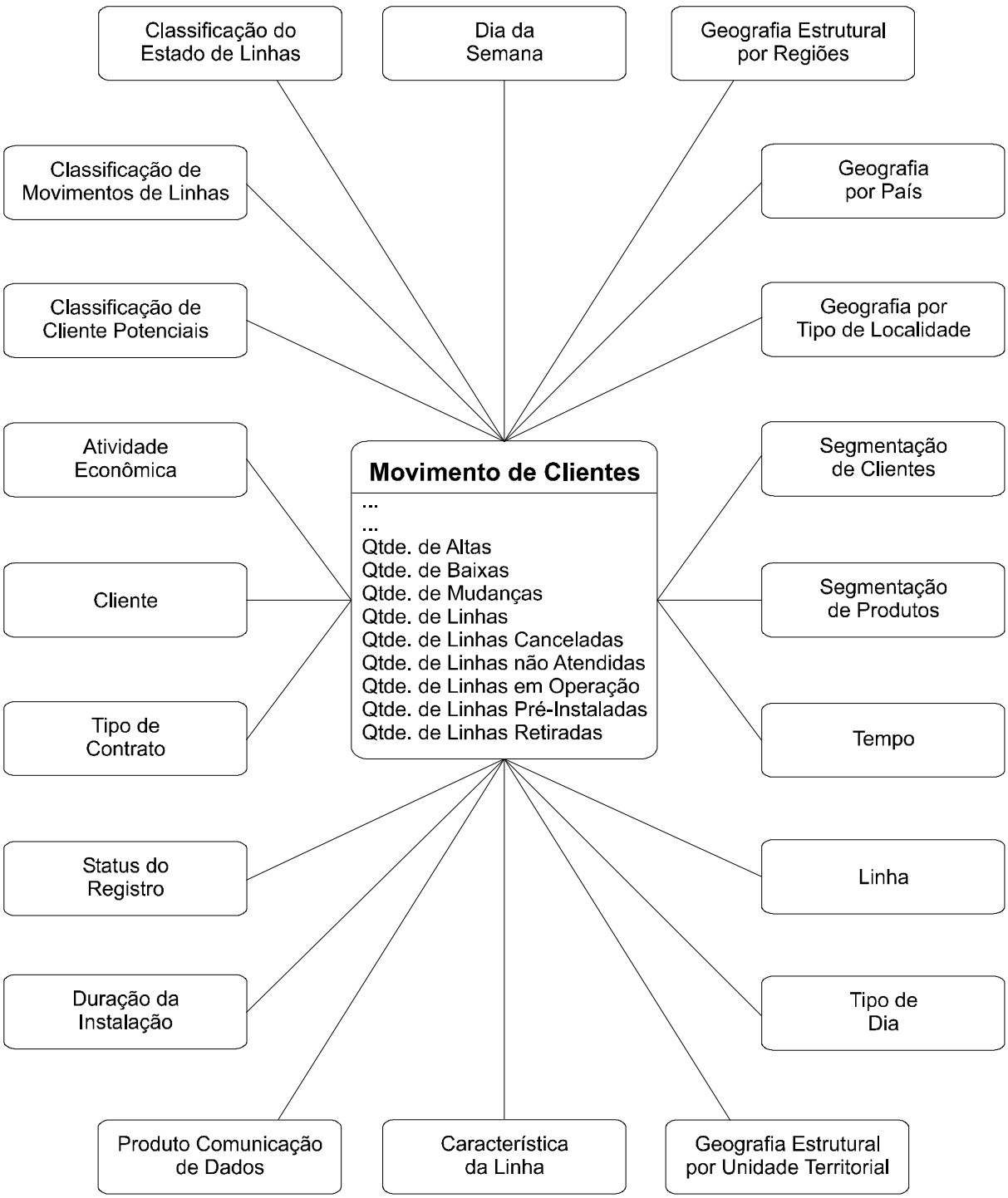
- **DataWarehousing.com**
  http://www.datawarehousing.com/
  Boa coleção de *links* para informações de *Data Warehouse* na Internet. Inclui companhias de *software* e consultoria, *papers* e artigos.

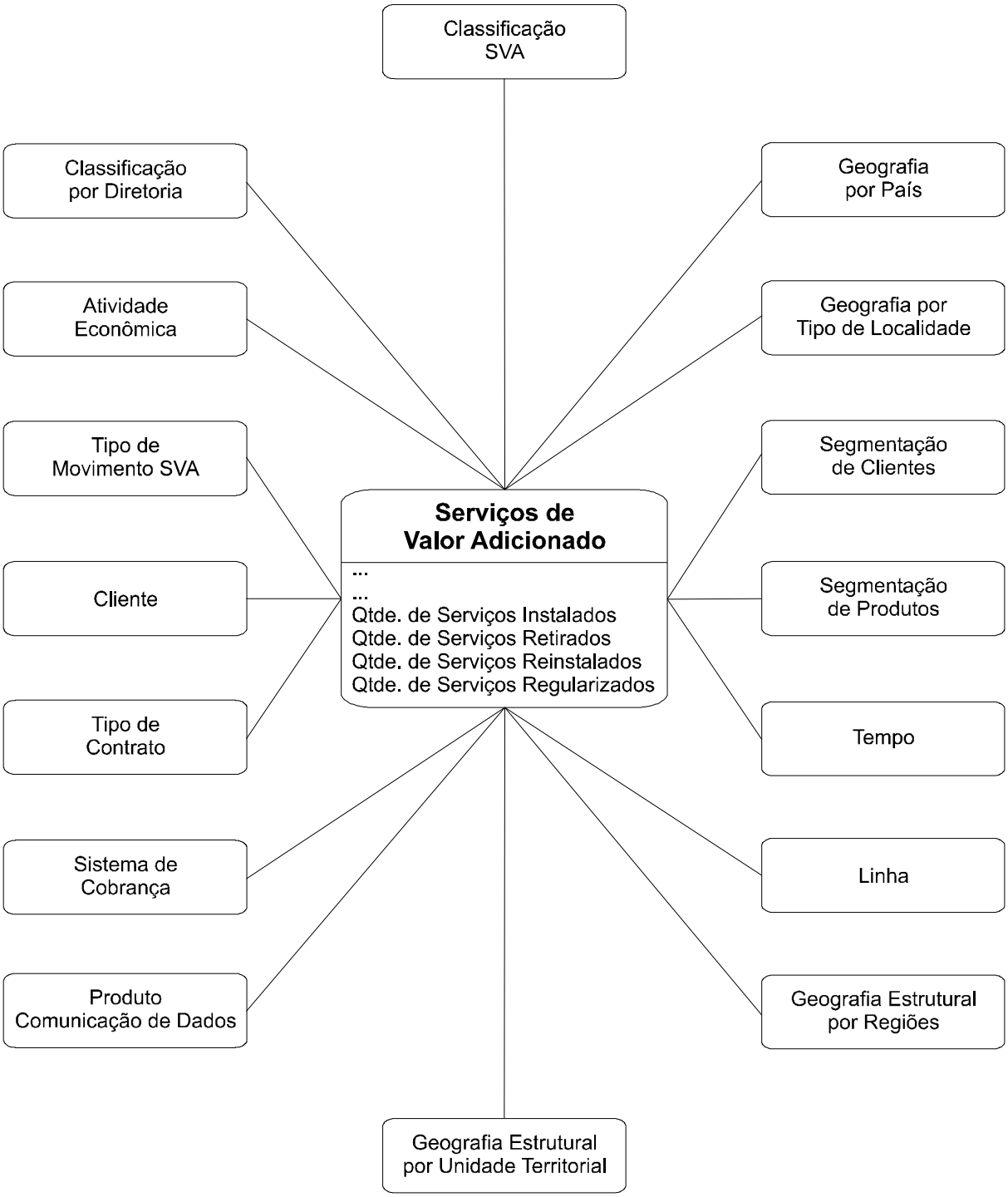
- **DataWarehouse.com**
  http://www.datawarehouse.com/
  Contém informações sobre os últimos desenvolvimentos em sistemas de informação e *Data Warehouse*. Fornece *links* para fóruns, notícias, soluções de fornecedores, etc.

- **Data Warehouse Institute**
  http://www.dw-institute.com/index.htm
  Associação para divulgação do conhecimento relacionado às tecnologias de *Data Warehouse*. Contém *links*, *white papers*, estudos de casos, informações de fornecedores, sites relacionados, etc.

- **Data Warehousing Knowledge Center**
  http://www.knowledgecenters.com
  Centro de informação para as tecnologias de *Data Warehouse*.

Anexo 3 – Exemplos do Modelo Dimensional




Classificação SVA
Classificação por Diretoria
Atividade Econômica
Tipo de Movimento SVA
Cliente
Tipo de Contrato
Sistema de Cobrança
Produto Comunicação de Dados
Geografia por País
Geografia por Tipo de Localidade
Segmentação de Clientes
Segmentação de Produtos
Tempo
Linha
Geografia Estrutural por Regiões
Geografia Estrutural por Unidade Territorial
Serviços de Valor Adicionado
...
...
Qtde. de Serviços Instalados
Qtde. de Serviços Retirados
Qtde. de Serviços Reinstalados
Qtde. de Serviços Regularizados

Anexo 4 – Autorização para publicação

Referências Bibliográficas

1. ADELMAN, Sid & MOSS, Larissa, ***Data Warehouse Cost*** – Addison Wesley Longman, May 2000 (extraído do livro *Data Warehouse* Project Management), consultada em 10/01/2001, disponível em http://www.ewsolutions.com/newsletter/17/1701.htm

2. ADELMAN, Sid & OATES, Joe, ***Data Warehouse Project Management*** – DM Review, May, 1998, consultada em 18/12/2000, disponível em http://www.dmreview.com/editorial/dmreview/print_action.cfm?EdID=1810

3. ANATEL, Agência Nacional de Telecomunicações – informações disponíveis em http://www.anatel.gov.br, 2001

4. ARNOLD, David E., ***The Role of HRD in the Successful Implementation of Information Systems*** – Human Resource Development Quarterly, v.7 n.3, Fall 1995, p.271-278

5. ARNUM, Eric, ***The Impact of Customer Care in Telecommunications*** – DM Review, consultada em 01/08/2000, disponível em http://www.dmreview.com/master.cfm?NavID=61&WhitePaperID=21

6. ATRE, Shaku, ***Learn the risks of marts*** – Computerworld, v.31, n.20, May 19 1997a, p.63-64

7. ATRE, Shaku, ***No panaceas*** – Computerworld, v.31, n.50, Dec 15 1997c, p.67-70

8. ATRE, Shaku, ***The trouble with training*** – Computerworld, v.31, n.42, Oct 20 1997b, p.73-76

9. BACHTEAL, Paul, ***Data Warehouses: Professional management key to successful implementation*** – Canadian Manager, v.22, n.2, Summer 1997 p.20-21

10. BAKER, B. N., MURPHY, D. C., FISHER, D. ***Factors Affecting Project Success*** – in Project Management Handbook, D. I. Cleland & W. R. King – New York: Van Nostrand Reinhold, 1988, p.902-919

11. BAKER, Sunny & BAKER, Kim, ***The best little warehouse in business*** – Journal of Business Strategy, v.20, n.2, Mar/Apr 1999, p.32-37

12. BALLARD, Chuck & HERREMAN, Dirk, ***Data Modeling Techniques for Data Warehousing*** – IBM, International Technical Support Organization, February, 1998.

13. BARQUIN, Ramon, PALLER, Allan & EDELSTEIN Herb, ***Ten Mistakes to Avoid for Data Warehousing Managers*** – The Data Warehouse Institute, Washington, D.C., 1995 (Appendix D in BERSON, Alex & SMITH, Stephen J. Smith, *Data Warehousing, Data Mining and OLAP* – USA, McGraw-Hill, 1997)

14. BEARD, Jon W. & PETERSON, Tim O., ***A Taxonomy for the Study of Human Factors in Management Information Systems***, In: CAREY, Jane M., *Human Management Information Systems* – New Jersey: Ablex Publishing Corporation, 1988, p.7-25

15. BERSON, Alex & SMITH, Stephen J. Smith, ***Data Warehousing, Data Mining and OLAP*** – USA, McGraw-Hill, 1997

16. BOHN, Kathy, ***Converting Data for Warehouses*** – DBMS, June 1997, consultada em 28/01/1999, disponível em http://www.dbmsmag.com/9706d15.html

17. BONOMA, Thomas V., ***Case research in marketing: opportunities, problems and a process*** – EUA: Journal of Marketing Research, May 1985, v.XXII, p.199-208

18. BRANDÃO, Carlos Antonio, ***Telecomunicações e Dinâmica Regional no Brasil*** – Tese de Doutorado, Instituto de Economia, Universidade Estadual de Campinas, Fevereiro, 1996

19. CAMPOMAR, M. Cortez, ***Do uso do "estudo de caso" em pesquisas para dissertações e teses em administração*** – Revista de Administração, São Paulo, FEA-USP, Julho-Setembro, 1991, v.26, n.3, p.95-97

20. CAMPOS, Maria Luiza, ROCHA FILHO, Arnaldo V. – ***Data Warehouse*** – Universidade Federal do Rio de Janeiro, consultada em 22/07/1998, disponível em http://tartaruga.nce.ufrj.br/dw/tutorial/inicio.htm

21. CHAROENNGAM, Chotchai & KAZI, Abdul Samad, ***Cost/Schedule Information System: A Human-Centered Approach*** – Cost Engineering, v.39, n.9, September 1997, p.29-35

22. CHERVINSK, Hannan & HENDERSON, Deborah, ***Building a cost-effective data warehouse*** – CMA Magazine, v.71, n.1, Feb 1997, p.19-22

23. COLOMBINI, Luís, ***Ímpeto de Matador*** – Revista Forbes Brasil, ano 2, n.13, 28 de março de 2001

24. COUTINHO, Luciano, CASSIOLATO, José Eduardo & SILVA, Ana Lúcia G. (Coord.), ***Telecomunicações, Globalização e Competitividade*** – Campinas, SP: Papirus, 1995

25. DECK, Stewart, ***Human Side Key to Data Warehousing*** – Computerworld, February 15 1999, v.33, n.7, p.14

26. FALTYS, John, ***Rules-Based Software for Telecommunications Targeted Marketing*** – DM Review, March 2000, consultada em 07/08/2000, disponível em http://www.dmreview.com/editorial/dmreview/print_action.cfm?EdID=2110

27. FARIAS, Andréa, ***A hora e a vez dos Data Marts*** – Computerworld, n.248, 06 de Abril de 1998, consultada em 08/10/1999, disponível em http://www.uol.com.br/computerworld/computerworld/248/soft_00.htm

28. GAGNON, Gabrielle, ***Data Warehousing: An Overview*** – PC Magazine, February 22 1999, consultada em 14/10/1999, disponível em http://www.zdnet.com/filters/printerfriendly/0,6061,388682-3,00.html

29. GARDNER, Stephen R., ***Building the Data Warehouse*** – Communications of ACM, September 1998, v.41, n.9, p.52-60

30. GIL, Antônio Carlos, ***Como elaborar projetos de pesquisa*** – São Paulo: Atlas, 1987

31. GIURLANI, Silvia, ***Um ROI cada vez mais difícil*** – Computerworld, n.309, 06/11/1999, consultada em 16/11/1999, disponível em http://www.uol.com.br/computerworld/technology/9911/9911roi.htm

32. GOODE, Willian J. & HATT, Paul, ***Métodos em pesquisa social*** – São Paulo: Companhia Editora Nacional, 1979

33. GODOY, Arilda Schmidt, ***Introdução à pesquisa qualitativa e suas possibilidades*** – São Paulo: Revista de Adm. de Empresas, Março/Abril 1995a – v.35 n.2 p. 57-63

34. GODOY, Arilda Schmidt, ***Pesquisa qualitativa: tipos fundamentais*** – São Paulo: Revista de Adm. de Empresas, Maio/Junho 1995b – v.35 n.3 p. 20-29

35. GOMEZ, John, ***Data Wareshousing for the Telecom Industry*** – DM Review, December 1998, consultada em 07/08/2000, disponível em http://www.dmreview.com/editorial/dmreview/print_action.cfm?EdID=260

36. GRAY, Paul, ***Data Warehouses, OLAP, Data Mining, and the New DSS*** – HEC Lausanne, 21-22 July 1997.

37. GRAY, Paul & WATSON, Hugh J., ***The New DSS: Data Warehouses, OLAP, MDD and KDD*** – University of Georgia, White Paper 1996, disponível em http://hsb.baylor.edu/ramsower/ais.ac.96/papers/graywats.htm, 15/01/2001

38. GUROVITZ, Helio, ***O que cerveja tem a ver com fraldas?*** – Informática Exame Online, consultada em 19/10/99, disponível em http://www2.uol.com.br/info/exame/armaz.html

39. HACKNEY, Douglas, ***How about 0% ROI?*** – *Data Mart* Review Magazine, January 1999, consultada em 12/09/2000, disponível em http://www.egltd.com/production/columns/1-99-5_zero_ROI.htm

40. HACKNEY, Douglas, ***Who Are You?*** – DM Review, February 1998, consultada em 14/10/1999, disponível em http://www.entergroupltd.com/colums/2-98-1_Who_Are_You.htm

41. HAISTEN, Michael, ***Real Time Data Warehouse: The Next Stage in Data Warehouse Evolution, part 1*** – DM Review, June 1999, consultada em 16/01/2001, disponível em http://www.dmreview.com

42. HALL, Curt, ***Data Warehousing for Business Intelligence*** – Report for IT & Software Professionals, March 1999, consultada 03/12/1999, disponível em http://www.cutter.com/itgroup/reports/dwissues.html

43. HERDLEIN, Stacey, ***Tips for Data Warehouse Project Managers to Build a Solid Foundation for a Data Warehouse*** – The Navigator, Fall 1996, consultada em 08/10/1999, disponível em http://www.idwa.org/fall96/pmtips.htm

44. HOGUE, Jack T., ***Uma estrutura para análise do envolvimento da gerência em sistemas de apoio à decisão,*** In: SPRAGUE, Ralph H. & WATSON, Hugh J. (Org.), Sistemas de apoio à decisão: colocando a teoria em prática – Rio de Janeiro: Campus, 1991, p. 57-75

45. IBM, ***A Pratical Guide to Getting Started with Data Warehousing*** – consultada em 23/08/1999, disponível em http://www.software.hosting.software.ibm.com/data/busn-intel/dwatgbroch

46. INFO Exame, ***As 200 de Info: as maiores empresas de Tecnologia do Brasil*** – Parte integrante da edição nº 173, Agosto/2000

47. INFORMIX Corporation, ***Data Warehousing for the Telecommunications Industry*** – White Paper 1998, consultada em 08/2000, disponível em http://www.dmreview.com

48. IMHOFF, Claudia, ***The Corporate Information Factory*** – DM Review, June 2000, consultada em 07/08/2000, disponível em http://www.dmreview.com

49. IMHOFF, Claudia, ***Are You an Inny or an Outty ?*** – DM Review, September 1999, consultada em 03/08/2001, disponível em http://www.dmreview.com/

50. INMON, Willian H., ***Data Mart ¹ Data Warehouse*** – DM Review, November 1999, consultada em 24/11/1999, disponível em http://www.dmreview.com

51. INMON, Willian H., ***Como construir o Data Warehouse*** - 2ª ed. – Rio de Janeiro: Campus, 1997

52. INMON, Willian H., ***The End-User Community*** – DM Review, December 1998, consultada em 03/08/2001, disponível em http://www.dmreview.com

53. INMON, Willian. H. & HACKATHORN, Richard D., ***Using the Data Warehouse*** – New York, John Wiley & Sons, 1994

54. INSIGHT Research Corporation, ***Data Warehouse in Telephone Networks*** – Perspective on the Telecom Marketplace, June 1998

55. KACHUR, Richard, ***The Data Warehouse Diary: Initiating a Data Warehouse Project, Part I*** – DM Review, December, 2000, consultada em 07/12/2000, disponível em http://www.dmreview.com/editiorial/dmdirect/

56. KIMBALL, Ralph, ***A Dimensional Modeling Manifesto***, DBMS, August 1997, consultada em 29/07/1998, disponível em http://www.dbmsmag.com/9708d15.htm

57. KIMBALL, Ralph, ***Data Warehouse Toolkit: Técnicas para construção de Data Warehouses Dimensionais*** – São Paulo: Makron Books, 1998

58. KNOWLES, Jim, ***Explore Data Warehousing*** – Datamation, December 1996, consultada em 29/01/1999, disponível em http://www.datamation.com/PlugIn/issues/1996/dec/12col40.html

59. LANG, Roger, ***5 Hardwares Considerations*** – Bank Marketing, v.29, n.4, February 1997, p.26

60. MacDONALD, Michael, ***Beat the odds*** – CMA Magazine, v.72, n.5, Jun 1998, p.16-18

61. MAN, Dick, ***The FAQs on Data Warehousing***, Bank Marketing, v.29, n.4, Apr 1997, p.38-42

62. MARTINS, Gilberto de Andrade, ***Manual para elaboração de monografias*** – São Paulo: Atlas, 1990

63. MARTINSONS, Maris G. & CHONG, Patrick K. C., ***The Influence of Human Factors and Specialist Involvement on Information Systems Success*** – Human Relations, v.52, n.1, January 1999, p.123-152

64. META Group & Software AG, ***Glossary*** – Alpha Box, 1998, consultada em 27/07/1998, disponível em http://www.data-warehouse.com/survival/k_glossary.htm

65. PASTE – Perspectivas para Ampliação e Modernização do Setor de Telecomunicações – Anatel, 2000

66. PERKINS, Alan, ***Critical Success Factors for Data Warehouse Engineering*** – Datawarehouse.com White Paper Library, consultada em 18/06/1999, disponível em http://www.datawarehouse.com/whitepaper/dw10.htm

67. PETERS, Richard, ***Data Warehousing in the Telecommunications Services Industry*** – DM Review, May 2000, consultada em 07/08/2000, disponível em http://www.dmreview.com/editorial/dmreview/print_action.cfm?EdID=2191

68. PINTO, Jeffrey K. & SLEVIN, Dennis P., ***Leadership, Motivation, and the Project Manager***, In: CLELAND, David, I., KING, Willian R., *Project management handbook* 2 ed. – New York: Von Nostrand Reinhold, 1988, p.739-770

69. PINTO, Jeffrey K. & SLEVIN, Dennis P., ***Project Success: Definitions and Measurement Techniques*** – Project Management Journal, v.XIX, n.1, February 1988a

70. PMI – Project Management Institute, ***A Guide to the Project Management Body of Knowledge*** – Newtown Square, PA 19073-3299 USA, 1996

71. POWER, D. J., ***Justifying a Data Warehouse Project*** – The On-Line Executive Journal for Data-Intensive Decision Support, February 3 and 10, 1998, v.2, n.5 and 6, consultada em 08/10/1999, disponível em http://dss.cba.uni.edu/papers/justifydw/indexp1.htm

72. RED BRICK Systems, Inc., ***Looking for New Answers: Data Warehousing in Telecommunications*** – 1998, consultada em 29/08/2000, disponível em http://www.redbrick.com/products/white/papers/telecom/telecom.html

73. RUDIO, Franz Vitor, ***Introdução ao projeto de pesquisa científica*** – Petrópolis: Vozes, 1986

74. RUFINO, Rogério, ***Churn*** – CTBC Telecom, Agosto de 1999, consultada em 29/08/2000, disponível em http://www.geocities.com/WallStreet/Exchange/8822/Noticias/1999/news2699.html

75. RUFINO, Rogério, ***Data Wareshousing*** – CTBC Telecom, Julho de 1998, consultada em 29/08/2000, disponível em http://www.geocities.com/WallStreet/Exchange/8822/Trends/trends26.html em

76. SAS Institute, ***Data Warehousing Methodology: A Management Summary*** – DM Review, December, 2000, consultada em 21/12/2000, disponível em http://www.dmreview.com/master.cfm?NavID=61&WhitePaperID=96

77. SAMTANI, Sunil; MOHANIA, Mukesh; KUMAR, Vijay; KAMBAYASHI, Yahiko, ***Recent Advances and Research Problems in Data Warehousing***, in: Kambayashi, Yahiko (Ed.), *Advances in Database Technologies* – ER '98 Workshops on *Data Warehousing*, Singapore, November, 1998, p. 81-92

78. SCHRADER, Achim, ***Introdução à Pesquisa Social Empírica: um Guia para o Planejamento, a Execução e Avaliação de Projetos de Pesquisas Não-Experimentais*** – Rio Grande do Sul: Editora Globo S.A., 1971

79. SCHROECK, Michael J., ***Insights from de Front Line: Understanding the Telecom Customer*** – DM Review, July 1999, consultada em 07/08/2000, disponível em http://www.dmreview.com/editorial/dmreview/print_action.cfm?EdID=1241

80. SELLTIZ, Claire et al., ***Métodos de pesquisa nas relações sociais*** – São Paulo: E.P.U. Editora Pedagógica e Universitária, 1974

81. SEQUENT Computer Systems, Inc., ***Data Wareshousing in the Telecommunications Industry: How it Works to Support the Business*** – DM Review, White Paper 2000, consultada em 01/08/2000, disponível em http://www.dmreview.com/master.cfm?NavID=61%WhitePaperID=90

82. SINGH, Harry, ***Data Warehousing: Concepts, Technologies, Implementations, and Management*** – Upper Saddle River, NJ: Prentice Hall, 1997

83. SIQUEIRA, Ethevaldo, ***Telecomunicações: Privatização ou Caos*** – São Paulo: TelePress Editora, 1993

84. SPRAGUE, Ralph H., ***Estrutura para o desenvolvimento de sistemas de apoio à decisão***, In: SPRAGUE, Ralph H. & WATSON, Hugh J. (Org.), Sistemas de apoio à decisão: colocando a teoria em prática – Rio de Janeiro: Campus, 1991, p. 9-42

85. SPSS, ***SPSS@work: Análise de Churn*** – 2000, consultada em 29/08/2000, disponível em http://www.spss.com.br/churn

86. TAYLOR, Jennifer & HANDEN, Lawrence, ***Managing the Store*** –Telephony, April 15, 1996, p.24-29

87. TELEBRÁS, Telecomunicações Brasileiras S.A. – informações disponíveis em http://www.telebras.com.br, 2001

88. TELETIME 2001 – Atlas Brasileiro de Telecomunicações.

89. TERESKO, John, ***Information Rich, Knowledge Poor?***, Industry Week, v.248, n.3, Feb 1999, p.19-24

90. THEODORATOS, Dimitri & SELLIS, Timos, ***Dynamic Data Warehouse Design***, in: Mohania, Mukesh (Ed.), *Data Warehousing and Knowledge Discovery* – First International Conference, DaWaK'99, Florence, Italy, August-September, 1999, p.1-10

91. TORRES, Norberto A., ***Competitividade empresarial com a tecnologia de informação*** – São Paulo: Makron Books, 1995

92. TRONCHIN, Valsoir, ***Análise, Modelagem e Implementação de Data Warehouses*** – São Paulo: Palestra da Fenasoft/98 em 20/07/98

93. TURBAN, Efraim & SHAEFFER, Donna M., ***Uma comparação entre Sistemas de Informação para Executivos, DSS e Sistemas de Informação Gerencial,*** In: SPRAGUE, Ralph H. & WATSON, Hugh J. (Org.), Sistemas de apoio à decisão: colocando a teoria em prática – Rio de Janeiro: Campus, 1991, p.345-361

94. WANG, Charles B., ***Techno Vision II*** – São Paulo: Makron Books, 1998

95. WATSON, Hugh J., ***Managerial Considerations*** – Communications of ACM, September 1998, v.41, n.9, p.32-37

96. WINEGARDNER, Karen E., ***The Case Study Method of Scholarly Research*** – The Graduate School of America, consultada em 19/10/1999, disponível em http://www.tgsa.edu/online/cybrary/case1.html

97. YIN, R. K., ***Case Study Research: Design and Methods*** – California: Sage Publications, 1989