

**UNIVERSIDADE FEDERAL DE SANTA CATARINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA  
COMPUTAÇÃO**

**Ivanise Volpato de Souza**

**Engenharia da Informação aplicada no desenvolvimento de *Data  
Warehouse***

Dissertação submetida à Universidade Federal de Santa Catarina como parte dos requisitos para a obtenção do grau de Mestre em Ciência da Computação.

Vitório Bruno Mazzola

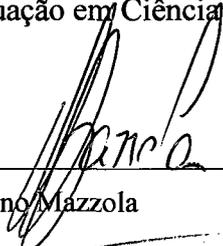
Florianópolis, Fevereiro de 2002.

As pessoas não perdem tempo na vida;  
o que se perde é a vida,  
ao perder-se tempo.”  
(Autor desconhecido)

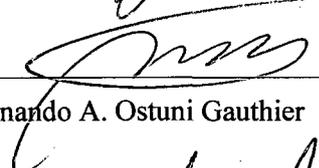
# Engenharia da informação aplicada no desenvolvimento de *Data Warehouse*

Ivanise Volpato de Souza

Esta Dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação Área de Concentração Sistemas de Computação e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.



Vitório Bruno Mazzola

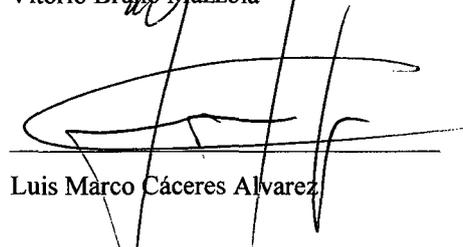


Fernando A. Ostuni Gauthier

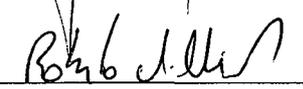
Banca Examinadora



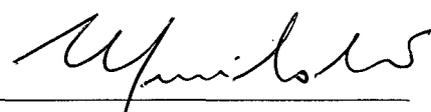
Vitório Bruno Mazzola



Luis Marco Cáceres Alvarez



Roberto Willrich



Murilo Silva de Camargo

Dedico este trabalho ao meu orientador e co-orientador,  
professores, amigos e a todos os meus familiares.

## **Agradecimentos**

A Deus Todo Poderoso, pela minha vida e por tudo que conquistei.

Aos professores integrantes da banca examinadora pela apreciação do presente trabalho.

Aos meus pais, meus irmãos e a toda minha família, pelo apoio e pelo incentivo. Em especial a minha tia Derce pela sua acolhida e ajuda durante todo o desenvolvimento deste trabalho.

Aos professores que, ao longo do curso, contribuíram de alguma forma para a realização deste trabalho.

À empresa Paradigma, em especial ao Fabio Vicentini, que ao longo deste trabalho deu toda a sua cordial atenção e ajuda.

Aos colegas da Pós-Graduação e aos amigos que durante esta caminhada fizeram-se presentes participando dos acontecimentos. Um especial agradecimento aos colegas Luiz, Sandro, Everton, Roberto, Christian, Leila, Adriana e Juliana.

Aos colegas integrantes do Laboratório de Gerência e Redes – LRG, em especial ao professor Carlos B. Westphall, por seu apoio e recursos disponibilizados.

Em especial, agradeço ao meu orientador, Dr. Vitorio Bruno Mazzola e meu co-orientador Luis Marco Cáceres Alvarez, pelo empenho e dedicação a este trabalho e às suas atividades acadêmicas.

## Sumário

<b>LISTA DE FIGURAS.....</b>	<b>VIII</b>
<b>LISTA DE TABELAS.....</b>	<b>IX</b>
<b>RESUMO.....</b>	<b>X</b>
<b>PALAVRAS-CHAVE.....</b>	<b>X</b>
<b>ABSTRACT .....</b>	<b>XI</b>
<b>1. INTRODUÇÃO .....</b>	<b>12</b>
1.1. VISÃO GERAL .....	12
1.2. OBJETIVO GERAL .....	13
1.3. OBJETIVOS ESPECÍFICOS .....	13
1.4. APRESENTAÇÃO DO TRABALHO .....	13
<b>2. ENGENHARIA DA INFORMAÇÃO .....</b>	<b>15</b>
2.1. ORIGEM .....	15
2.2. A PIRÂMIDE .....	16
2.2.1. <i>Nível 1 - Planejamento estratégico de informações</i> .....	17
2.2.2. <i>Nível 2 - Análise da área de negócio</i> .....	18
2.2.3. <i>Nível 3 - Projeto</i> .....	19
2.2.4. <i>Nível 4 - Construção</i> .....	20
2.2.5. <i>Nível 5 – Manutenção</i> .....	20
2.3. OS BLOCOS QUE COMPÕEM A ENGENHARIA DA INFORMAÇÃO .....	21
2.4. FERRAMENTAS CASE E ICASE.....	23
2.4.1. <i>A Enciclopédia</i> .....	24
2.5. O MODELO DE DADOS .....	26
2.6. CONCLUSÃO .....	26
<b>3. DATA WAREHOUSE .....</b>	<b>28</b>
3.1. CONCEITO DO <i>DATA WAREHOUSE</i> .....	28
3.2. CARACTERÍSTICAS DO <i>DATA WAREHOUSE</i> .....	30
3.2.1. <i>Orientado por Temas</i> .....	30
3.2.2. <i>Integrado</i> .....	31
3.2.3. <i>Variável no tempo</i> .....	31
3.2.4. <i>Não volátil</i> .....	31
3.3. MODELAGEM DOS DADOS .....	32
3.4. METADADOS .....	34
3.5. POVOAMENTO DE <i>DATA WAREHOUSE</i> .....	34
3.6. OBTENÇÃO DE INFORMAÇÕES DO <i>DATA WAREHOUSE</i> .....	35
3.7. CONCLUSÃO .....	35
<b>4. METODOLOGIAS DE DESENVOLVIMENTO DO DATA WAREHOUSE .....</b>	<b>37</b>
4.1. METODOLOGIA SEGUNDO RALPH KIMBALL.....	37
4.1.1. <i>Planejamento de projeto</i> .....	37

4.1.2. Definição de requisitos de negócio .....	38
4.1.3. Modelagem dimensional.....	38
4.1.4. Projeto físico .....	39
4.1.5. Projeto e desenvolvimento da classificação dos dados.....	39
4.1.6. Projeto da arquitetura técnica .....	39
4.1.7. Seleção e instalação de produtos .....	40
4.1.8. Especificação de aplicações do usuário final .....	40
4.1.9. Desenvolvimento de aplicação do usuário final .....	40
4.1.10. Desenvolvimento.....	40
4.1.11. Manutenção e crescimento .....	40
4.1.12. Gerenciamento de projeto .....	41
4.2. METODOLOGIA SEGUNDO INMON .....	41
4.2.1. Análise do modelo de dados .....	41
4.2.2. Dimensionamento .....	42
4.2.3. Avaliação técnica .....	42
4.2.4. Preparação do ambiente técnico.....	43
4.2.5. Análise das áreas de interesse.....	44
4.2.6. Projeto do Data Warehouse .....	44
4.2.7. Análise do sistema-fonte.....	45
4.2.8. Especificações .....	46
4.2.9. Programação.....	46
4.2.10. Povoamento .....	47
4.3. CONCLUSÃO .....	47
<b>5. APLICAÇÃO DA ENGENHARIA DA INFORMAÇÃO NO DESENVOLVIMENTO DO DATA WAREHOUSE.....</b>	<b>50</b>
5.1. OBJETIVO .....	50
5.2. BREVE HISTÓRICO SOBRE A ORGANIZAÇÃO .....	50
5.3. ESTABELECIMENTO DAS ETAPAS DA METODOLOGIA .....	51
5.3.1. Planejamento estratégico de informações.....	52
5.3.2. Análise das áreas de negócios da empresa .....	56
5.4. CONCLUSÃO .....	62
<b>6. MODELAGEM E IMPLEMENTAÇÃO DO MODELO DIMENSIONAL.....</b>	<b>63</b>
6.1. MODELAGEM DIMENSIONAL .....	63
6.2. COMPARAÇÃO ENTRE O MODELO RELACIONAL E O MODELO DIMENSIONAL .....	64
6.3. MODELAGEM DO MODELO DIMENSIONAL .....	65
6.3.1. Definição da área de negócio.....	65
6.3.2. Definição das tabelas de dimensão .....	66
6.3.3. Normalização das tabelas de dimensão .....	66
6.3.4. Relacionamentos de atributos das tabelas de dimensão .....	67
6.3.5. Definição dos atributos das tabelas de fato .....	67
6.4. IMPLEMENTAÇÃO DO MODELO DIMENSIONAL.....	68
6.4.1. Processamento Analítico On-line - OLAP .....	68
6.4.2. O Microsoft Analysis Manager .....	69
6.4.4. Origem de dados.....	70
6.4.5. Dimensões.....	71
6.4.6. Cubo .....	74

6.4.7. <i>Inclusão das dimensões no cubo</i> .....	77
6.5. VISUALIZAÇÃO DOS DADOS .....	78
6.6. CONCLUSÃO .....	81
<b>7. CONCLUSÃO E PERSPECTIVAS FUTURAS.....</b>	<b>82</b>
7.1. TRABALHOS FUTUROS .....	84
<b>8. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>85</b>
<b>ÍNDICE REMISSIVO.....</b>	<b>88</b>

## Lista de Figuras

FIG. 2.1 – PIRÂMIDE DA ENGENHARIA DA INFORMAÇÃO SEGUNDO (MARTIN, 1991). ...	16
FIG. 2.2 – PIRÂMIDE DA ENGENHARIA DA INFORMAÇÃO SEGUNDO (NETO, 1988).....	17
FIG. 2.3 – EXEMPLO DE DIAGRAMA DE FLUXO DE DADOS E DIAGRAMA DE AÇÃO CORRESPONDENTE. ....	19
FIG. 2.4 - BLOCOS DA ENGENHARIA DA INFORMAÇÃO. ....	22
FIG. 2.5 – REPRESENTAÇÃO GRÁFICA DA ENCICLOPÉDIA. ....	25
FIG. 3.1 – DADOS EXTRAÍDOS DE SISTEMAS OPERACIONAIS PARA O <i>DATA WAREHOUSE</i> . ..	29
FIG. 3.2 - DIAGRAMA DE ENTIDADE-RELACIONAMENTO. ....	32
FIG. 3.3 - DIAGRAMA DE MODELO DIMENSIONAL. ....	33
FIG. 4.1 – DIAGRAMA DO CICLO DE VIDA. ....	38
FIG. 4.2 – DESENVOLVIMENTO DO <i>DATA WAREHOUSE</i> . ....	43
FIG. 5.1 – RESUMO DA METODOLOGIA. ....	53
FIG. 5.2 – PLANEJAMENTO ESTRATÉGICO DE INFORMAÇÕES EM NÍVEL ESTRATÉGICO. ....	54
FIG. 5.3 – MODELO DE DADOS CORPORATIVO. ....	56
FIG. 5.4 – REPRESENTAÇÃO GRÁFICA DAS ENTIDADES <i>AGENTE</i> E <i>FORNECEDOR</i> . ....	58
FIG. 5.5 – ENTIDADE AGENTE FÍSICO E AGENTE JURÍDICO. ....	59
FIG. 5.6 – MODELO DE DADOS NORMALIZADO. ....	61
FIG. 6.1 – DEFINIÇÃO DAS TABELAS DIMENSÕES. ....	66
FIG. 6.2 – INTERFACE DO <i>MICROSOFT ANALYSIS MANAGER</i> . ....	69
FIG. 6.3 – CAIXA DE DIÁLOGO <i>NEW DATABASE</i> . ....	70
FIG. 6.4 – CAIXA DE DIÁLOGO <i>DATA LINK PROPERTIES</i> . ....	70
FIG. 6.5 – O <i>DIMENSION WIZARD</i> . ....	71
FIG. 6.6 – CAIXA DE DIÁLOGO <i>SELECT DIMENSION TABLE</i> . ....	72
FIG. 6.7 – CAIXA DE DIÁLOGO <i>DIMENSION WIZARD SELECT LEVELS</i> . ....	73
FIG. 6.8 – CAIXA DE DIÁLOGO <i>DIMENSION WIZARD FINISH</i> . ....	74
FIG. 6.9 – CAIXA DE DIÁLOGO <i>CUBE WIZARD SELECT FACT TABLE</i> . ....	75
FIG. 6.10 – CAIXA DE DIÁLOGO <i>CUBE WIZARD DEFINE MEASURES</i> . ....	76
FIG. 6.11 – CAIXA DE DIÁLOGO <i>CUBE WIZARD SELECT DIMENSIONS</i> . ....	76
FIG. 6.12 – CAIXA DE DIÁLOGO <i>CUBE WIZARD FINISH</i> . ....	77
FIG. 6.13 – MODELO DIMENSIONAL FINAL. ....	78
FIG. 6.14 – CAIXA DE FERRAMENTAS. ....	79
FIG. 6.15 – CAIXA DE DIÁLOGO <i>DATA LINK PROPERTIES</i> . ....	80
FIG. 6.16 – LISTA DE CAMPOS DA TABELA. ....	80

## Lista de Tabelas

TABELA 2.1 – EXEMPLO DE DICIONÁRIO DE DADOS.....	24
TABELA 5.1 – ATRIBUTOS DAS ENTIDADES IDENTIFICADAS. ....	57
TABELA 6.1 – MODELO DIMENSIONAL VERSUS MODELO RELACIONAL. ....	64

## Resumo

O presente trabalho propõe a utilização da Engenharia da Informação para a análise, incorporação de técnicas e procedimentos formais nas atividades de planejamento e de desenvolvimento de sistemas cuja informação estratégica é processada por intermédio dos procedimentos desenvolvidos pela tecnologia *Data Warehouse* cujo o conjunto de dados está baseado em informações integradas, não voláteis, variáveis no tempo, consistentes e de apoio às decisões estratégicas das empresas de *e-commerce*.

Apresentam-se duas abordagens que estabelecem as atividades necessárias para o desenvolvimento do *Data Warehouse* no planejamento estratégico das informações e análise da área de negócios. Para a implementação do *Data Warehouse* se utilizará a modelagem dimensional em formato de esquema estrela e o conceito de *ClickStream*, isso permite a construção de um protótipo, o qual valida os conceitos e os resultados esperados no uso da Engenharia da Informação e *Data Warehouse*.

### Palavras-chave

Engenharia da Informação, *Data Warehouse*, Modelagem Dimensional.

## **Abstract**

This work proposes the use of Information Engineering to the analysis and incorporation of formal techniques and procedures in the planning and system development activities, whose strategical information is processed through the procedures developed by the Data Warehouse technology, in which the data set is based on integrated, non-volatile, variable in the time, and consistent information that gives support to the strategical decision of the e-commerce companies.

Two approaches are presented that establishes the necessary activities for the Data Warehouse development in the information's strategical planning and the area's business analysis. For the Data Warehouse implementation a dimensional modeling in star scheme format and the ClickStream concept will be used, allowing a prototype construction, which validates the concepts and results expected using information engineering and Data Warehouse.

## **Key-words**

Information Engineering, Data Warehouse, Dimensional Modeling.

# 1. Introdução

## 1.1. Visão geral

A Engenharia da Informação (EI) com a utilização de técnicas similares às da Engenharia tradicional procura tratar as informações necessárias para o funcionamento do empreendimento de forma exata e com conceitos formais, a fim de desenvolver e manter sistemas de processamento de dados (NETO, 1988).

A velocidade do avanço tecnológico na área da EI e nas organizações marcou as últimas décadas deste século. As enormes mudanças ocorridas nesse período, no entanto, não se comparam com a verdadeira revolução que está em curso com a democratização e popularização do uso da informação.

Atualmente, torna-se necessário que as organizações transformem sua enorme quantidade de dados em uma grande vantagem competitiva. Esta é uma tarefa quase impossível sem o recurso dos grandes repositórios de dados que são capazes de realizar pesquisas personalizadas de acordo com o perfil do usuário.

No processo decisório das organizações existem vários fatores envolvidos. Isso exige que se disponibilize um bom suporte a este processo para se atingir o sucesso dos negócios. Os gerentes e diretores devem ter suas decisões tomadas com base na análise de todas as alternativas possíveis, encontrando assim, a mais viável.

É indispensável a utilização do suporte computacional para a aquisição e análise das informações necessárias. As decisões devem ser realizadas com base em dados e fatos comprovados, para que se possa ter o máximo de certeza dos resultados a serem alcançados com a escolha da melhor alternativa.

Em um ambiente onde a informação é um importante recurso estratégico diante da concorrência, a implementação de sistemas como o *Data Warehouse* é uma importante ajuda. Desse modo, o *Data Warehouse* define a informação como um banco de dados histórico, que busca organizar os dados corporativos, para dar subsídios de informações aos gerentes e diretores para tomada de decisões gerenciais e estratégicas.

## 1.2. Objetivo geral

O trabalho tem como principal objetivo propor a aplicação da Engenharia da Informação no processo da modelagem dimensional do *Data Warehouse* por intermédio do desenvolvimento de um modelo orientado a uma organização de comércio eletrônico, neste caso, a empresa Paradigma.

Pretende integrar-se a Engenharia da Informação à abordagem desenvolvida por Ralph Kimball, para o tratamento das necessidades de informação da organização cujo trabalho desenvolvido seguiu esta linha. Tentou-se eliminar um dos fatores críticos de sucesso no desenvolvimento de projetos de *Data Warehouse*. Objetivando uma melhor idéia dos resultados esperados do *Data Warehouse*.

## 1.3. Objetivos específicos

O trabalho tem como principais objetivos:

- estudar os principais conceitos relacionados à disciplina da Engenharia da Informação;
- estudar aspectos de arquitetura e mecanismos associados à operação de um *Data Warehouse*;
- analisar as alternativas em termos de metodologia para a concepção de *Data Warehouse*;
- verificar como integrar os conceitos da Engenharia da Informação no contexto de uma metodologia de concepção de *Data Warehouse*;
- estudar, analisar e aplicar os conceitos relacionados com a modelagem dimensional orientado a construção do *Data Warehouse*.

## 1.4. Apresentação do trabalho

Neste trabalho são apresentados no capítulo 2 conceitos de Engenharia da Informação, sua origem, as fases representadas pela pirâmide, os blocos que compõem a Engenharia da Informação e que formam um conjunto integrado de metodologias, as ferramentas computadorizadas *CASE* e *I-CASE*, e o modelo de dados.

O capítulo 3 descreve conceitos sobre *Data Warehouse*, suas características gerais, a modelagem dos dados, metadados, e os processos associados ao povoamento do *Data Warehouse*.

O capítulo 4 apresenta duas metodologias de desenvolvimento do *Data Warehouse*, as quais mostram os passos a seguir na manipulação da informação estratégica de uma organização de uma forma consistente e eficiente.

O capítulo 5 descreve a aplicação da Engenharia da Informação no desenvolvimento do Modelo Dimensional do *Data Warehouse* que fornece o acesso aos dados de modo que representam a informação real da organização que balizará a tomada de decisões.

O capítulo 6 apresenta a modelagem baseada no esquema estrela e a implementação do Modelo Dimensional no *Analysis Services*, modelo este obtido pela aplicação da Engenharia da Informação na organização.

Finalmente, o capítulo 7 apresenta as conclusões e perspectivas futuras e o capítulo 8 as referências bibliográficas.

## 2. Engenharia da Informação

*“Administrar bem um negócio é administrar seu futuro, e administrar o futuro é administrar informações”.* (FURLAN, 1994).

A Engenharia da Informação pode ser definida segundo (MARTIN, 1991) como “a aplicação de um conjunto interligado de técnicas formais de planejamento, análise e construção de sistemas de informação sobre uma organização como um todo ou em seus principais setores”, um exemplo do conceito é ilustrado na Figura 2.1.

De acordo com (NETO, 1988) “a Engenharia da Informação é um conjunto de técnicas e lógicas formais, aplicadas na pirâmide de dados, atividades, tecnologia e pessoas, que permite planejar, analisar, construir e manter sistemas de processamento de dados, de forma integrada e de interação mútua”, um exemplo do conceito é ilustrado na Figura 2.2.

Ao fazer referência às técnicas automatizadas, a Engenharia da Informação foi definida como “um conjunto de disciplinas automatizadas em nível de organização cuja finalidade é fornecer as informações certas às pessoas certas e na hora certa” (MARTIN, 1991).

Neste capítulo são apresentados conceitos sobre Engenharia da Informação. Na seção 2.1 é descrito sua origem; na seção 2.2 são descritas as fases, representadas pela pirâmide; na seção 2.3 são apresentados os blocos que compõem a Engenharia da Informação e formam um conjunto integrado de metodologias; na seção 2.4 são apresentadas as ferramentas computadorizadas *CASE* e *I-CASE*; a seção 2.5, descreve o modelo de dados; finalmente a seção 2.6 apresenta a conclusão .

### 2.1. Origem

De acordo com (NETO, 1988) a convivência com a necessidade de gerenciamento dos projetos de sistemas levou em 1981, os especialistas em informática, Clive Finkelstein e James Martin, a estabelecerem as bases de uma nova disciplina, denominada **Engenharia da Informação** que tem por objetivo incorporar técnicas e procedimentos formais às atividades de planejamento e desenvolvimento de sistemas.

A objetivo principal da Engenharia da Informação é aplicar o planejamento *top-down* (de cima para baixo), a modelagem de dados e de processos em uma organização como um todo, e não apenas em projetos isolados (NETO, 1988).

Para que isso fosse alcançado, a evolução das ferramentas integradas *CASE* (*Computer Aided Systems Engineering*) foi necessária para produzir um gerador de códigos para esta modelagem de dados (NETO, 1988).

## 2.2. A pirâmide

Segundo (MARTIN, 1991), a Engenharia da Informação possui quatro níveis e duas faces coesas, integradas, interativas e seqüenciais. Podemos representar esses níveis em uma pirâmide, como observado na Figura 2.1.

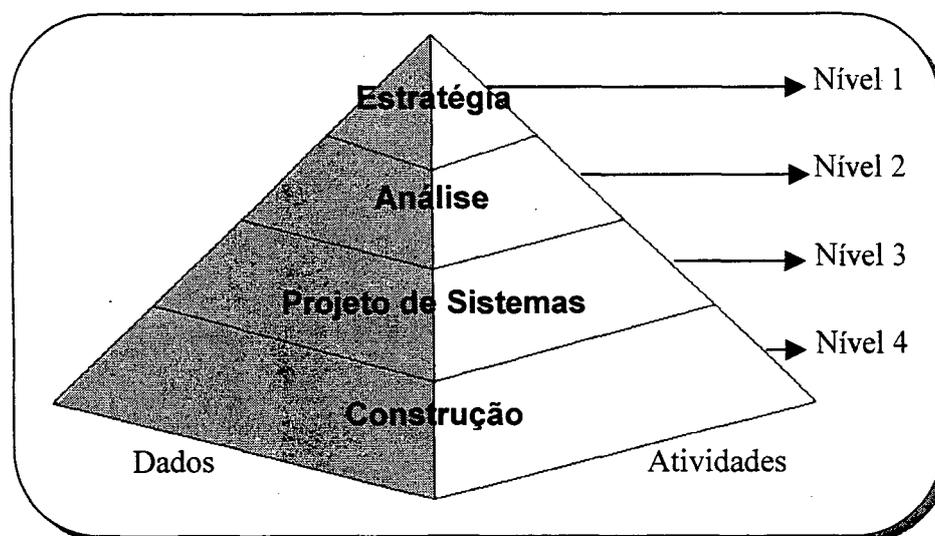


Fig. 2.1 – Pirâmide da Engenharia da Informação segundo (MARTIN, 1991).

Fonte: (MARTIN, 1991)

De acordo com (NETO, 1988), podemos representar a Engenharia da Informação como uma pirâmide de cinco níveis distintos e quatro faces compostas por: dados, atividades, tecnologia e pessoas. Cada uma dessas faces, dentro da abordagem *top-down*, apresenta cinco níveis integrados e progressivos para o desenvolvimento de sistemas de processamento de dados, abrangendo o Planejamento Estratégico, a Análise das Áreas de Negócios da Empresa, o Projeto, a Construção e a Manutenção do sistema, conforme ilustrado na Figura 2.2.

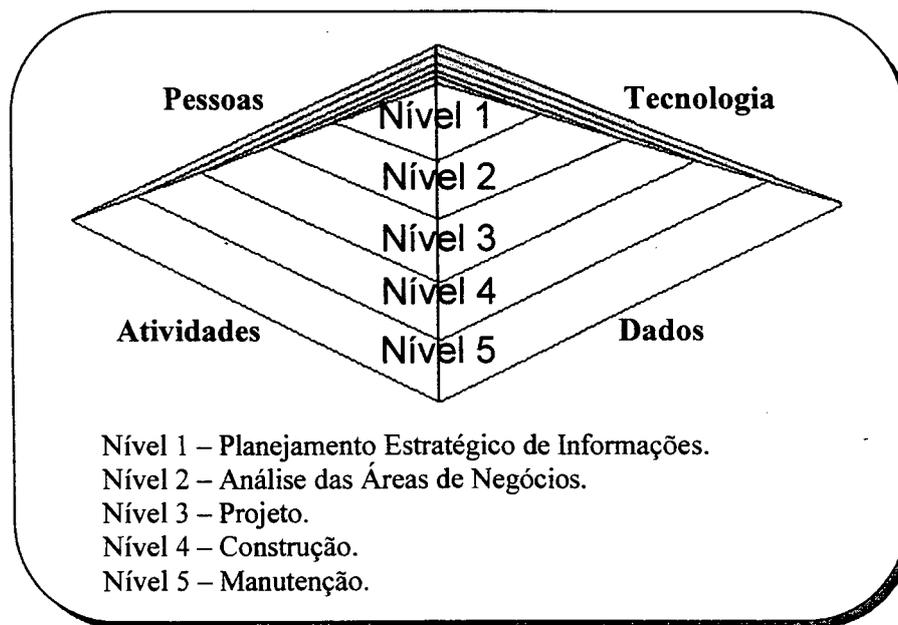


Fig. 2.2 – Pirâmide da Engenharia da Informação segundo (NETO, 1988).

Fonte: (NETO, 1988)

As duas concepções são muito semelhantes, o que diferencia a segunda da primeira citada é que a segunda incorporou mais um nível e duas faces à pirâmide. As duas metodologias possuem os mesmos conceitos e métodos, o que não interfere no presente trabalho, sendo possível utilizar a primeira ou a segunda concepção, ou então mesclar as duas.

Tais fases, observadas na Figura 2.1 e na Figura 2.2, são descritas nas sub-seções a seguir:

### 2.2.1. Nível 1 - Planejamento estratégico de informações

Segundo (MARTIN, 1991), a fase inicial da Engenharia da Informação corresponde ao **Planejamento Estratégico de Informações** onde são estabelecidos os propósitos básicos, para que se possa desenvolver sistemas computadorizados estáveis.

Nessa fase caso não se tenha uniformidade dos objetivos para se atingir os propósitos globais da organização, existirá um desencontro que se tornará desafinado e sem nenhuma interação, ocasionando uma independência negativa (NETO, 1988).

A partir de uma metodologia definida e aplicada no planejamento estratégico empresarial, esta adotará a missão, as metas, os desafios e os objetivos da organização, estabelecendo, em seguida, vínculos com os fatores críticos de sucesso que permitirão o alcance dos objetivos de longo, médio e curto prazo (NETO, 1988).

Dessa forma, o planejamento estratégico de informações, fornecerá informações integrados com os propósitos globais da organização, conhecendo-se detalhadamente a origem, o meio e o fim de todos os projetos desenvolvidos (NETO, 1988).

O planejamento estratégico de informações necessita do comprometimento da alta administração. A primeira preocupação refere-se ao uso estratégico da tecnologia, ter uma visão de como a informática poderá colocar a organização a frente de seus concorrentes. Esse processo resulta na identificação de problemas e soluções organizacionais e operacionais (MARTIN, 1991).

### **2.2.2. Nível 2 - Análise da área de negócio**

De acordo com (MARTIN, 1991) a segunda fase, **Análise da Área de Negócio**, é conduzida separadamente para cada área da empresa, onde é desenvolvido um modelo para os dados e para os processos necessários no funcionamento da organização.

Nesta fase a participação intensiva e o comprometimento dos usuários envolvidos na área da organização analisada é muito importante, de forma a estabelecer um modelo de dados e um modelo de processos detalhados, que sejam de acordo com a realidade, completos e de plena confiabilidade (NETO, 1988).

Segundo (MARTIN, 1991), nessa fase são definidas as classes de dados, decompostas em atributos, e os processos gerenciais, decompostos em atividades (que são todos os procedimentos necessários para executar determinado processo ou ação), gerando visões de contexto dos Diagramas de Fluxo de Dados (DFD) que decompostos até um maior nível de detalhamento, resultarão nos correspondentes Diagramas de Ação estruturado, como mostra a Figura 2.3.

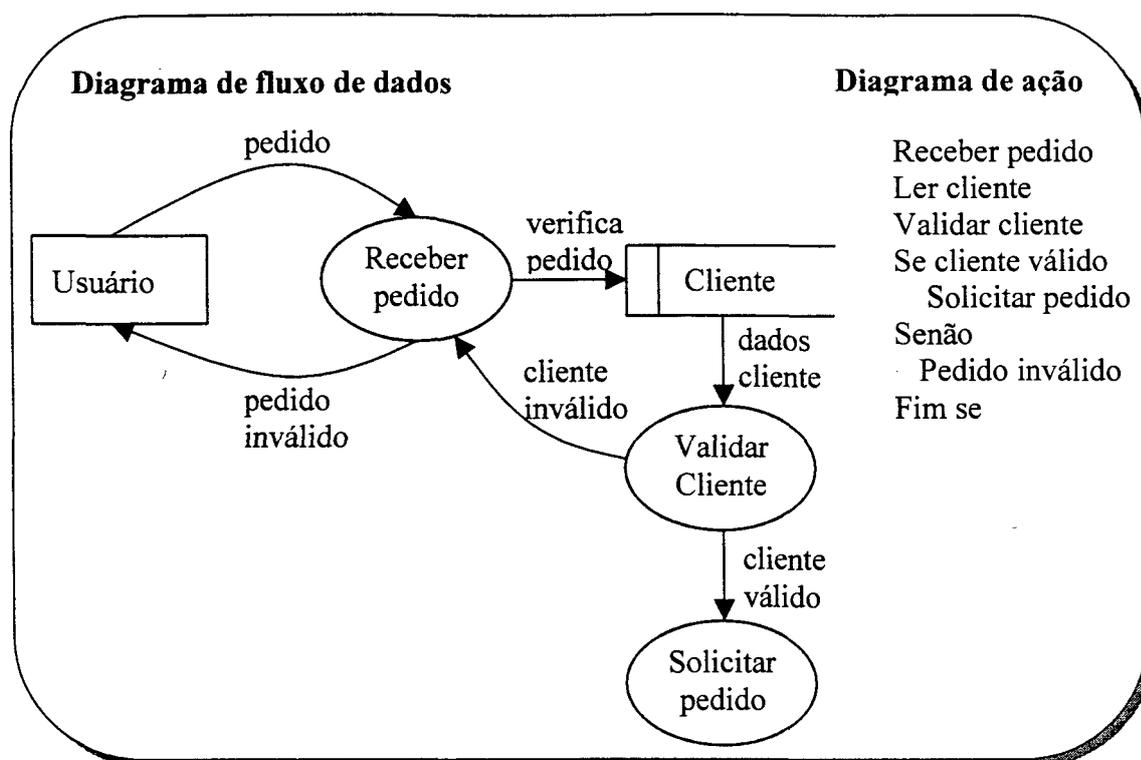


Fig. 2.3 – Exemplo de diagrama de fluxo de dados e diagrama de ação correspondente.

A análise da área de negócios não se destina a projetar sistemas. Ela simplesmente procura entender e modelar os processos e dados necessários para que uma determinada área da organização funcione (NETO, 1988).

### 2.2.3. Nível 3 - Projeto

De acordo com (MARTIN, 1991), analisada a área de negócio e identificados os processos gerenciais e as atividades críticas, inicia-se a fase de **Projeto**. Nesta fase se estabelece o submodelo de dados de cada atividade envolvida, migra-se os diagramas de fluxo de dados para os correspondentes diagramas de ação, como mostra o exemplo da Figura 2.3.

Igualmente a fase de Análise da Área de Negócio, a participação ativa dos usuários é fator primordial de sucesso no empreendimento, tanto na definição do submodelo de dados como na identificação detalhada dos procedimentos que compõem cada atividade e, especialmente, no teste dos protótipos criados, onde, em função de sua vivência com os sistemas atuais e da visão de suas necessidades, agilizará a finalização dos propósitos de projeto (NETO, 1988).

Essa fase se modifica drasticamente quando ferramentas de projeto automatizadas são utilizadas. Com essas ferramentas o projeto fica pronto em muito menos tempo, pois o projetista pode constantemente editar o projeto, acrescentando, eliminando e modificando blocos, ligações e aperfeiçoando os detalhes (MARTIN, 1991).

#### **2.2.4. Nível 4 - Construção**

Segundo (MARTIN, 1991) esgotadas todas as filtragens e refinamentos necessários na fase de Projeto, deve-se executar a **Construção** do sistema enfocado, com a utilização das ferramentas disponíveis, desde geradores de códigos até linguagens de quarta geração compatíveis com os equipamentos de processamento de dados da organização.

Essa fase caracteriza-se pelas particularidades de hardware e software, visto que cada organização apresenta configuração específica de equipamento e universo próprio de recursos de linguagens e de sistemas (NETO, 1988).

#### **2.2.5. Nível 5 – Manutenção**

De acordo com (NETO, 1988), desde que as quatro fases anteriores tenham sido executadas criteriosamente e metodologicamente, a manutenção se restringirá a ocorrências eventuais, visto que existe estabilidade no modelo de dados e dos processos vinculados, e somente mudanças políticas, econômicas e legislativas do ambiente externo poderão teoricamente resultar em alterações a serem modificadas no sistema.

No entanto, caso haja requerimentos não vinculados a essas mudanças, mas relacionados com alterações de procedimentos internos, é essencial analisar com os usuários envolvidos a real necessidade da adição de novos conceitos, e quais benefícios e problemas resultaria essa manutenção (NETO, 1988).

No caso de mudanças externas, é importante antes de qualquer modificação verificar se o sistema não previu a ocorrência da alteração. Apenas no caso de não se ter uma alternativa dentro do sistema é que se deve partir para a fase de manutenção (NETO, 1988).

## 2.3. Os Blocos que compõem a Engenharia da Informação

A Engenharia da Informação oferece um conjunto integrado de metodologias. Neste diagrama, cada bloco depende do bloco do nível inferior. Entretanto, os blocos podem ser montados de várias formas, como apresentado a seguir. A Figura 2.4, mostra os vários blocos que compõem a Engenharia da Informação.

Bloco 1, modelo da organização e planejamento estratégico de informações: está relacionado ao planejamento estratégico, onde são estabelecidos os objetivos da organização. Procura-se determinar quais informações são necessárias para que a organização atinja seus objetivos. É onde todos os outros blocos se apóiam (MARTIN, 1991).

Bloco 2, análise das entidades: um mapa contendo a visão geral dos dados necessários para o funcionamento da organização é criado. É realizada uma análise, de cima para baixo (*top-down*), dos tipos de dados que devem ser mantidos e como se relacionam entre si (MARTIN, 1991).

Bloco 3, modelos de dados detalhados: é desenvolvido o modelo de dados que cria o projeto lógico do banco de dados. Deve-se fazê-lo o mais consistente possível previamente a sua implementação (MARTIN, 1991).

Blocos 4 e 5, centro de informações e computação pessoal: são utilizadas linguagens e geradores de códigos por usuários finais (MARTIN, 1991).

Bloco 6, análise de processos: é efetuada a análise de uma determinada área de negócios. As funções da área em análise são decompostas em processos por meio de diagramas de decomposição. Diagramas de dependências ou diagramas de fluxos de dados podem ser desenhados, mostrando os inter-relacionamentos entre os processos, como ilustrado na Figura 2.4 (MARTIN, 1991).

Bloco 7, projeto de procedimentos: usa vários tipos de diagramas relacionados entre si por intermédio da enciclopédia (observe item 2.4.1). Um subconjunto do modelo de dados é extraído da enciclopédia, junto com informações que auxiliam a confecção do projeto (MARTIN, 1991).

Blocos 8 e 9, gerador de programas ou linguagem de quarta geração e protótipos: o projeto criado no bloco 7 pode ser implementado com um gerador de código ou com uma linguagem de quarta geração. Pode ser usado como uma alternativa para criar um

protótipo. Este protótipo pode ser modificado, sucessivamente, e pode tornar-se o código operacional (MARTIN, 1991).

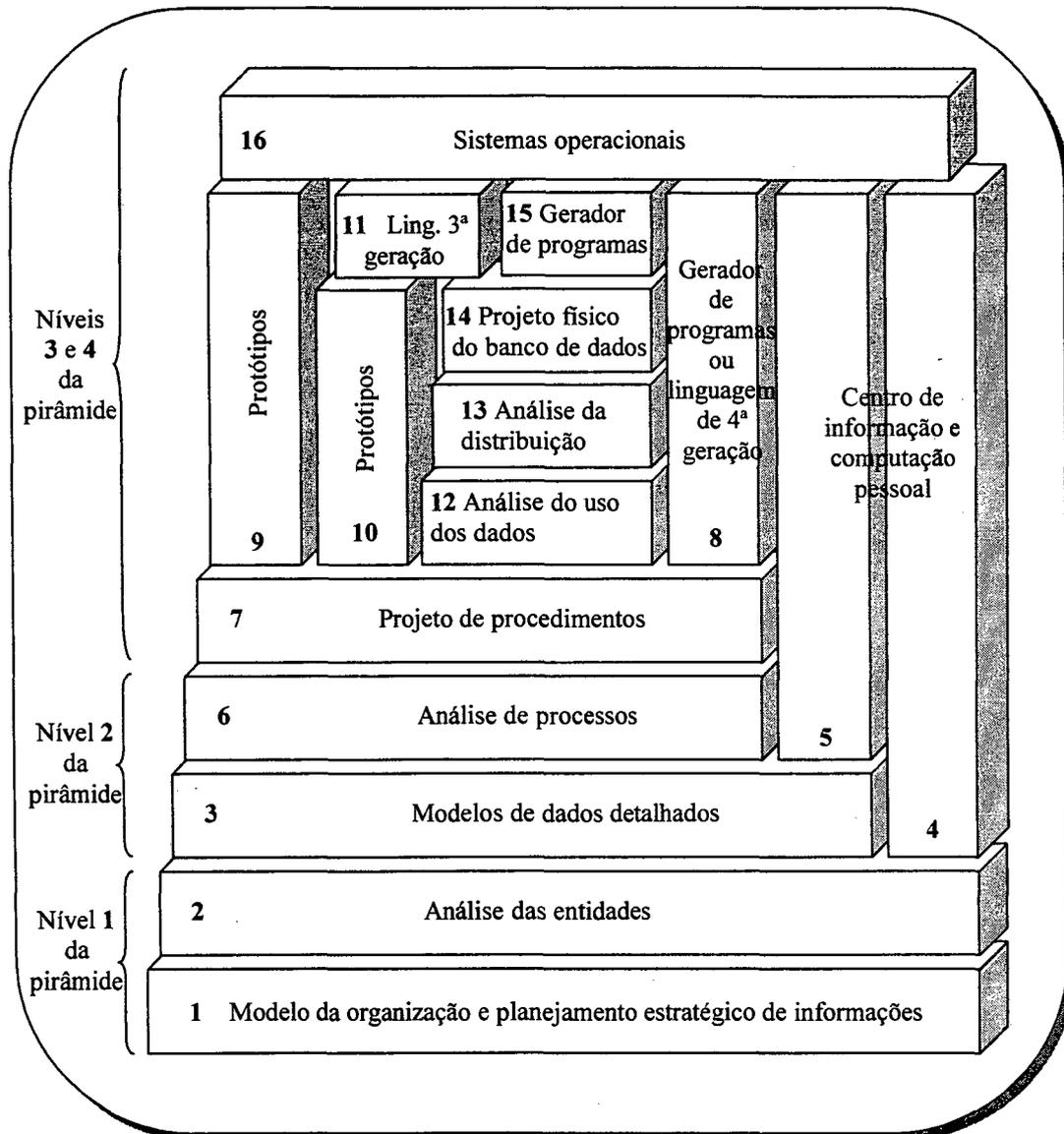


Fig. 2.4 - Blocos da Engenharia da Informação.

Fonte: (MARTIN, 1991)

Blocos 10 e 11, protótipos e linguagem de terceira geração: o protótipo pode ser convertido para uma linguagem de nível inferior, para que possa ser otimizado em termos de performance (MARTIN, 1991).

Bloco 12, análise do uso dos dados: é definido como os dados deverão ser utilizados, os caminhos de uso dentro do banco de dados, o volume de uso dos dados e a necessidade de tempo de resposta (MARTIN, 1991).

Bloco 13, análise da distribuição: existem diversas razões para a existência de sistemas distribuídos e diferentes formas de distribuição. Os modelos de dados podem ser divididos em bancos de dados e separados quando as bases de dados são utilizadas na implementação dos programas (MARTIN, 1991).

Bloco 14, projeto físico de banco de dados: está relacionado à conversão do modelo de dados e procedimentos no projeto físico do banco de dados (MARTIN, 1991).

Bloco 15, gerador de programas ou linguagem de quarta geração: a implementação pode ser feita por geradores de programas, linguagens de terceira e quarta geração. O ponto principal para atingirmos uma boa produtividade e uma alta qualidade é combinar as ferramentas de projeto automatizado da Engenharia da Informação com geradores de códigos (MARTIN, 1991).

Bloco 16, sistemas operacionais: preocupa-se com a manutenção do sistema que deverá ser realizada após uma modificação nos diagramas ou uma reconstrução de programas (MARTIN, 1991).

## **2.4. Ferramentas *CASE* e *ICASE***

As primeiras tentativas na área de Engenharia da Informação ocorreram antes que existissem as ferramentas computadorizadas de hoje. Apenas em meados de 1980, as ferramentas *CASE* tornaram-se populares e são usadas para descrever ferramentas poderosas para os analistas de sistemas, até então, feitas manualmente (MARTIN; 1991).

A grande vantagem da ferramenta *CASE* é que permite criar diagramas de grande complexidade com exatidão e rapidez, uma vez que manualmente, é quase impossível, e a probabilidade de ocorrer erros é muito grande (MARTIN, 1991).

Os diagramas são usados para representarem as informações que se baseiam no planejamento, uma visão geral do sistema, modelos e fluxos de dados, projetos detalhados e estruturas de programas. Devem ser completos o suficiente para servirem de base para a geração de programas e para a conversão automática de um tipo de

diagrama para outro. Também, devem ser claros e precisos, de maneira que quando o programador altere seu projeto, essa alteração não interfira no projeto de outros (MARTIN, 1991).

As ferramentas *I-CASE* (*CASE integrado*) geram programas executáveis automaticamente. Esse gerador de programas orientado a objetos permite uma grande integração das ferramentas de análise e projeto e oferece uma maior produtividade e qualidade em relação às ferramentas usadas separadamente (MARTIN, 1991).

As ferramentas *I-CASE* possuem grandes vantagens em relação às ferramentas *CASE* porque além de incorporarem as características da segunda, possuem um gerador de programas completamente integrado ao ambiente de projeto. Suportam as fases da vida do projeto de forma integrada e toda a documentação é gerada automaticamente (MARTIN, 1991).

#### 2.4.1. A Enciclopédia

A **enciclopédia** é um repositório computadorizado que armazena informações referentes ao planejamento, análise, projeto, construção e mais tarde, manutenção de sistemas (MARTIN, 1991). As ferramentas de Engenharia da Informação utilizam dois tipos de repositório:

- **Dicionário:** define a organização básica de um repositório. Contém uma lista de todos os arquivos, o número total de registros em cada arquivo, o nome, a descrição e o tipo de cada campo (MARTIN, 1991). Observe o exemplo ilustrado na Tabela 2.1.

Tabela 2.1 – Exemplo de dicionário de dados.

<i>Column_name</i>	<i>data_type</i>	<i>data_length</i>	<i>description</i>
Id	number	7	Primary key
first_name	varchar	25	First_name
last_name	varchar	25	Last_name
Salary	number	11	Salary
Department	varchar	25	Department

- **Enciclopédia:** durante o andamento das fases da Engenharia da Informação, os conhecimentos são acumulados e armazenados na enciclopédia (MARTIN, 1991). Observe a representação gráfica na Figura 2.5.

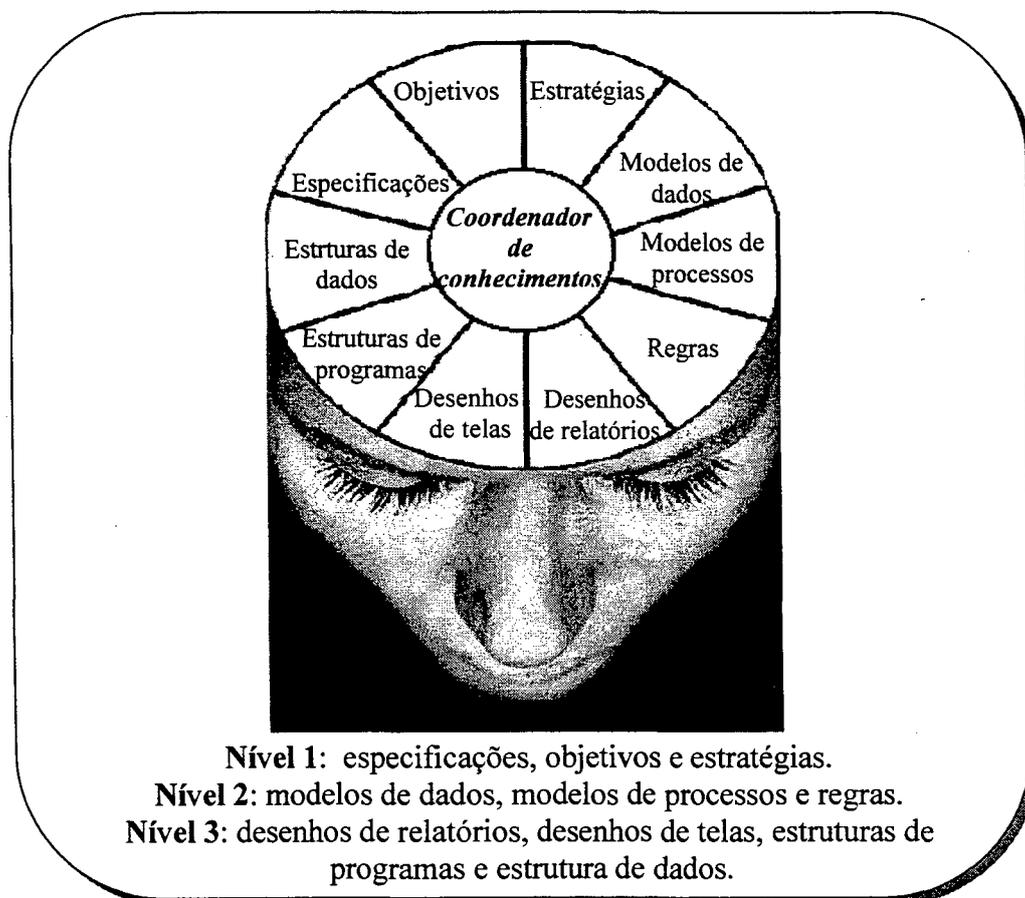


Fig. 2.5 – Representação gráfica da enciclopédia.

No nível 1 da pirâmide, as informações da enciclopédia estão relacionadas ao planejamento estratégico das informações, por exemplo, os fatores de sucesso críticos do negócio são armazenados e relacionados a outros aspectos do planejamento de sistemas. No nível 2 da pirâmide, os modelos de dados e os de processos são construídos dentro da enciclopédia. O nível 3 da pirâmide usa as informações contidas na enciclopédia, como apoio para a geração do projeto. Detalhes de tela, estruturas de programas e os bancos de dados são criados dentro da enciclopédia.

Durante o andamento das fases da Engenharia da Informação, os conhecimentos são acumulados e armazenados em uma enciclopédia. A enciclopédia é, portanto, uma

base de conhecimentos que não só armazena informações para o desenvolvimento como também auxilia a controlar a sua exatidão e validade (MARTIN, 1991).

## 2.5. O Modelo de dados

A Engenharia da Informação se preocupa com os dados, de forma que sejam o centro do processamento. Os tipos de dados usados em uma organização se bem projetados, não sofrem muitas mudanças (MARTIN, 1991).

O modelo de dados constitui o ponto principal sobre o qual é construído a maioria dos procedimentos computadorizados. É importante um projeto lógico, bem planejado das estruturas de dados, para que o banco de dados seja consistente e robusto. O processo de normalização permite que os dados tenham uma estrutura lógica correta representando as propriedades referentes a eles, podendo ser utilizados em vários tipos de aplicações (MARTIN, 1991).

Para construir o modelo de dados, a prioridade inicial é uma visão geral dos dados de toda a organização ou os principais setores dela. Após isso, é criado um modelo de dados normalizado, ou seja, são incluídos mais detalhes no modelo geral criado na fase anterior. Durante a fase de projeto, a estrutura dos dados adapta-se aos recursos de um banco de dados ou sistema de gerenciamento de banco de dados (MARTIN, 1991).

## 2.6. Conclusão

A Engenharia da Informação (EI) é uma metodologia desenvolvida para ser, segundo (MARTIN, 1991), “um conjunto de disciplinas automatizadas em nível de organização cuja finalidade é fornecer as informações certas às pessoas certas e na hora certa”.

O resultado esperado na implantação da Engenharia da Informação é obter sistemas de informação alinhados com os objetivos da organização, compatíveis entre si, sem redundâncias desnecessárias e de baixo custo de manutenção. Adicionalmente, por ser implantada com o suporte de software específico, a Engenharia da Informação gera um crescente repositório de conhecimentos sobre a organização, denominado por (MARTIN, 1991) de **enciclopédia**. Nela ficam registrados os seus modelos de dados, modelos de processos, projetos de sistemas e dicionários (SERAFIM FILHO, 1999).

As organizações necessitam de um enfoque metodológico no desenvolvimento de seus sistemas de informação e no uso da tecnologia da informação. As metodologias utilizadas, no entanto, precisam ser cada vez mais eficientes (enfocando o negócio) e menos instrumentais (enfocando a ferramenta, o *software*), dada a velocidade de atualização da tecnologia. Precisam, ainda, apoiar o aumento da flexibilidade da organização e facilitar a sua gestão do conhecimento (SERAFIM FILHO, 1999).

A contribuição que a EI pode trazer está no fato de que toda a modelagem da organização está registrada em um ambiente computadorizado. Todos os principais inter-relacionamentos podem ser revisados a partir do replanejamento em um tempo mais curto do que os métodos mais convencionais (SERAFIM FILHO, 1999). As fases da Engenharia da Informação utilizadas neste trabalho serão: o planejamento estratégico das informações e análise da área de negócios.

O resultado da fase do planejamento estratégico das informações é um modelo de negócio que conduz a organização a um entendimento comum sobre os seus aspectos críticos e, torna-se um instrumento de comunicação para todos os níveis (MARTIN, 1991).

Na fase de análise da área de negócios obtém-se um modelo de processos e um modelo de dados que formarão toda a base do *Data Warehouse*. Esta fase pode ser realizada em toda a corporação ou em parte considerada crítica. O processo de obtenção desses modelos é um trabalho em equipe em que participam pessoas das várias áreas da organização (MARTIN, 1991). Destaca-se que um modelo de dados bem elaborado torna o *Data Warehouse* muito mais consistente e robusto.

O próximo capítulo descreve conceitos sobre *Data Warehouse*, suas características gerais, a modelagem dos dados, metadados, e os processos associados ao povoamento do *Data Warehouse*.

### 3. Data Warehouse

*“O ritmo das mudanças não vai diminuir; por isso, a vantagem competitiva tende na direção de empresas que podem acelerar a sua capacidade de tomar decisões”.* (TANLER, 1998).

Segundo (CAMPOS, 2000), na organização, diariamente, os dados sobre os mais variados aspectos dos negócios são gerados e armazenados e passam a fazer parte dos recursos de informação dessa empresa. Entretanto, essas informações encontram-se, em geral, espalhadas em diversos sistemas e exigem um esforço considerável de integração para que possam dar suporte efetivo à tomada de decisão de gerentes e executivos. Isso significa que, apesar dos avanços da tecnologia de armazenamento e manipulação de dados, ainda se verifica, nos dias de hoje, uma grande deficiência na obtenção ágil de informações estratégicas.

Neste capítulo a seção 3.1. apresenta noções gerais sobre o *Data Warehouse*; a seção 3.2 descreve as características gerais do *Data Warehouse*; a seção 3.3 descreve a modelagem de dados; a seção 3.4 apresenta o papel dos metadados; a seção 3.5 mostra os processos associados ao povoamento do *Data Warehouse*; finalmente a seção 3.6 apresenta a conclusão.

#### 3.1. Conceito do *Data Warehouse*

Segundo (INMON, 1992), considerado um pioneiro no tema, um *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”.

De acordo com (SILBERSCHATZ, 1999),

um *Data Warehouse* é um repositório (ou arquivo) de informações coletadas em diversas fontes, armazenadas sob um esquema único, em um só local. Uma vez coletados, os dados são armazenados por um período longo, permitindo acesso a dados históricos. Assim, os *Data Warehouses* fornecem ao usuário uma interface única consolidada de dados, tornando mais fácil gerar consultas de suporte à decisão.

Em (SPENIK, 2001), um *Data Warehouse* é um recurso de armazenamento centralizado para diferenciar tipos de dados por toda organização. Esses dados são orientados por assunto, sensíveis ao tempo e organizados de tal maneira que ofereçam análise simplificada.

O *Data Warehouse* é um banco de dados contendo dados extraídos do ambiente de produção da empresa, que foram selecionados e depurados, tendo sido organizados para processamento de consulta e não para processamento de transações. Em geral, um *Data Warehouse* requer a consolidação de outros recursos de dados além dos armazenados em banco de dados relacionais, incluindo informações provenientes de planilhas eletrônicas, documentos textuais, etc (CAMPOS, 2000).

De uma forma geral, os *Data Warehouses* se constituem de um conjunto de programas que extraem dados do ambiente de dados operacionais da empresa, um banco de dados que os mantém, e sistemas que fornecem estes dados aos seus usuários (CAMPOS, 2000), como ilustrado na Figura 3.1.

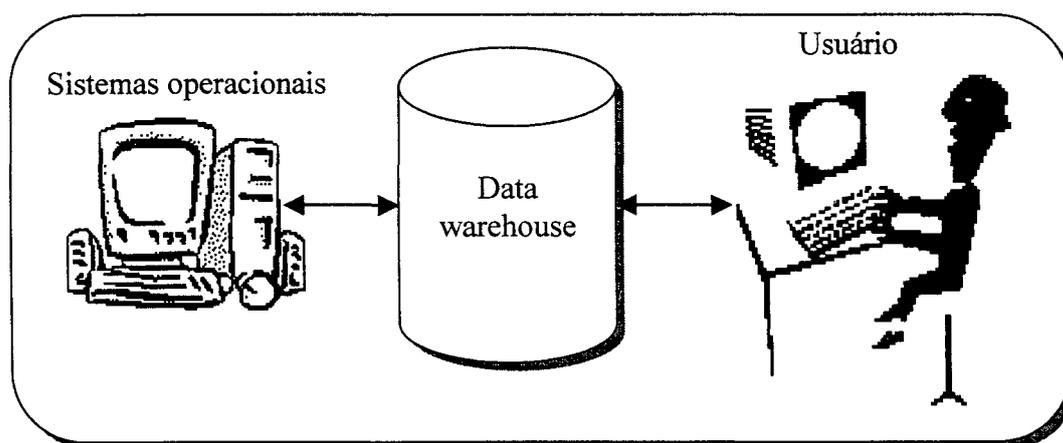


Fig. 3.1 – Dados extraídos de sistemas operacionais para o *Data Warehouse*.

É importante considerar, no entanto, que um *Data Warehouse* não contém apenas dados resumidos, podendo conter também dados primitivos. É desejável fornecer ao usuário a capacidade de aprofundar-se num determinado tópico, investigando níveis de agregação menores ou mesmo o dado primitivo, permitindo também a geração de novas agregações ou correlações com outras variáveis. É extremamente difícil prever todos os possíveis dados resumidos que serão necessários. Limitar o conteúdo do *Data Warehouse* apenas a dados resumidos significa limitar os usuários apenas às consultas e análises que eles possam antecipar frente a seus requisitos atuais, não deixando qualquer flexibilidade para novas necessidades (CAMPOS, 2000).

## 3.2. Características do *Data Warehouse*

O *Data Warehouse* tem como principais características ser orientado por temas, integrado, variável no tempo e não volátil. Essas características são descritas nas subseções a seguir.

### 3.2.1. Orientado por Temas

Refere-se ao fato do *Data Warehouse* armazenar informações sobre temas específicos importantes para o negócio da organização. Exemplos típicos de temas são: produtos, atividades, contas e clientes (CAMPOS, 2000).

Segundo (INMON, 1992), a implementação de um tema pode corresponder a um conjunto de tabelas relacionadas. Por exemplo, considerando informações sobre vendas de funcionários, podem existir tabelas contendo informações básicas dos funcionários (como `código_do_funcionário`, nome, endereço, sexo, `data_início`, `data_fim`, etc.), uma com dados do período 1990 a 1994, outra com dados para o período 1995-2000. Além dessas, existem tabelas acumulativas intermediárias com as atividades dos funcionários entre 1990 e 2000, contendo um registro resumo para as atividades de cada mês (contendo `código_do_funcionário`, mês, `número_de_transações`, `média_de_vendas`, `total_menor_venda`, `total_maior_venda`, `total_vendas_canceladas`, etc.). E finalmente, encontram-se ainda tabelas detalhadas de atividades para os períodos 1997-1998 e 1999-2000 (incluindo `código_do_funcionário`, `data_atividade`, `número_da_nota`, `número_pedido`, `quantia`, `cliente_id`, local, etc.).

Existem, portanto para o mesmo tipo de informação, diferentes níveis de detalhes. Note-se que todas essas tabelas contêm um identificador comum, o `código_do_funcionário`, além de um elemento temporal como parte da chave de cada tabela. Nem sempre todas estas tabelas seriam mantidas em disco, sendo possível que, em alguns casos, as informações mais detalhadas das atividades dos vendedores fossem mantidas em fita magnética, ficando acessíveis apenas quando solicitadas (INMON, 1992).

### 3.2.2. Integrado

Refere-se à consistência de nomes, das unidades das variáveis, etc., no sentido de que os dados foram transformados até um estado uniforme. Por exemplo, considere sexo como um elemento de dado. Uma aplicação pode codificar sexo como “M/F”, outra como “1/0” e uma terceira como “H/M”. Conforme os dados são trazidos para o *Data Warehouse*, eles são convertidos para um estado uniforme, ou seja, sexo é codificado apenas de uma forma (CAMPOS, 2000).

### 3.2.3. Variável no tempo

De acordo com (CAMPOS, 2000), o fato do dado em um *Data Warehouse* referir-se a algum momento específico, significa que ele não é atualizável. Em um *Data Warehouse*, a cada ocorrência de uma mudança, uma nova entrada é criada, para marcar esta mudança.

O tratamento de séries temporais apresenta características específicas, que adicionam complexidade ao ambiente do *Data Warehouse*. Processamentos mensais ou anuais são simples, mas dias e meses oferecem dificuldades pelas variações encontradas no número de dias em um mês ou em um ano, ou ainda no início das semanas dentro de um mês. Além disso, deve-se considerar que não apenas os dados tem uma característica temporal, mas também os metadados, que incluem definições dos itens de dados, rotinas de validação, algoritmos de derivação, etc (CAMPOS, 2000).

### 3.2.4. Não volátil

De acordo com (CAMPOS, 2000) significa que o *Data Warehouse* permite apenas a carga inicial dos dados e consultas a estes dados, o chamado ambiente de carga e acesso (*load-and-access*). Após serem integrados e transformados, os dados são carregados em bloco para o *Data Warehouse*, para que estejam disponíveis para acesso aos usuários.

No ambiente operacional os dados geralmente são atualizados registro a registro, em várias transações. Esta volatilidade requer um trabalho considerável para assegurar integridade e consistência por meio de atividades de *rollback* (desfaz todas as alterações executadas), recuperação de falhas, *commits* (indica que a transação foi concluída) e

bloqueios. Um *Data Warehouse* não requer esse grau de controle típico dos sistemas orientados a transações (CAMPOS, 2000).

### 3.3. Modelagem dos dados

Segundo (WELDON, 1997), “um modelo é uma representação abstrata de um objeto real ou de um ambiente. A modelagem de dados é a prática de elaborar um banco de dados usando modelos de dados já consagrados”.

O modelo E-R (Entidade-Relacionamento) divide os dados em diversas tabelas, que se relacionam entre si, formando um diagrama, como mostrado na Figura 3.2. Este tipo de diagrama é muito difícil de se interpretar e analisar, inclusive pelos próprios projetistas e não retrata a realidade dos negócios (BISPO, 1998).

Para (KIMBALL, 1995, 1997 e 1998), o modelo Entidade-Relacionamento não é adequado para se modelar os dados no ambiente gerencial, sendo o Modelo Dimensional (MD) o mais apropriado para este ambiente. Segundo o autor, esse modelo também é conhecido por esquema em estrela (*star join scheme*) ou simplesmente esquema estrela.

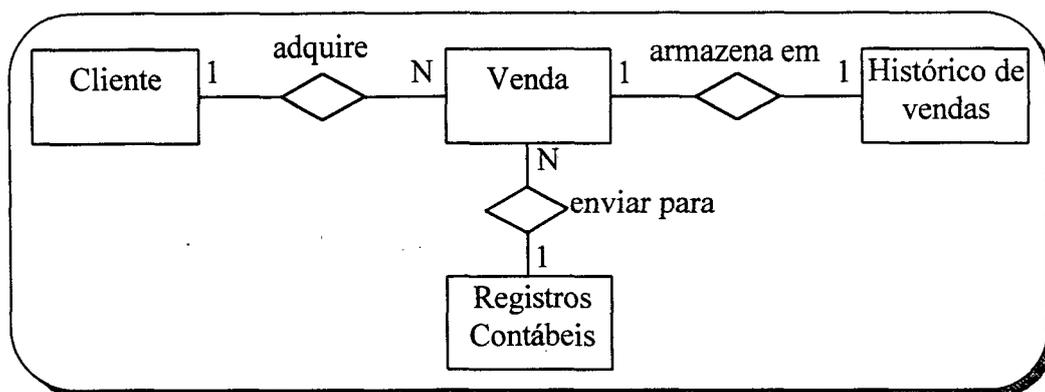


Fig. 3.2 - Diagrama de Entidade-Relacionamento.

Fonte: (YOURDON, 1990)

O Modelo Dimensional ilustrado na Figura 3.3 possui uma grande tabela que é a principal e está localizada no centro do diagrama. Possui outras tabelas secundárias ao seu redor que são menores e que se relacionam com a tabela principal. A tabela central é chamada de tabela de fato e as demais são chamadas tabelas de dimensão. A tabela de

fato armazena as medidas numéricas do negócio, por exemplo: unidades vendidas em cada transação, unidades produzidas em cada lote, etc. Esta tabela é composta por uma chave composta (BISPO, 1998).

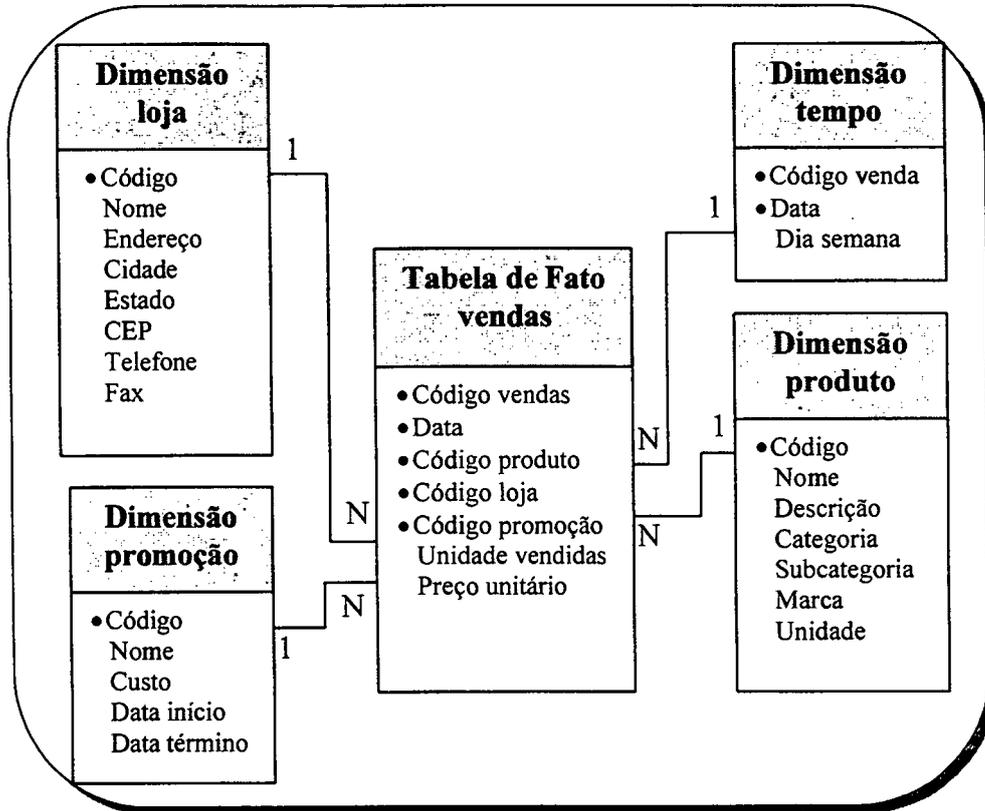


Fig. 3.3 - Diagrama de Modelo Dimensional.

Fonte: (BISPO, 1998)

As tabelas de dimensão armazenam as descrições textuais das dimensões do negócio, por exemplo, a dimensão produto possui os dados sobre a marca, a categoria, o tipo de embalagem, o tamanho do produto, etc. Cada tabela de dimensão tem uma única chave primária que corresponde aos componentes da chave composta da tabela de fato. O *Data Warehouse* possui várias tabelas de fatos, cada uma representando um assunto ou negócio diferente dentro da organização (BISPO, 1998).

A construção de um Modelo Dimensional é um processo *top-down*, primeiro identificam-se os processos empresariais que representem assuntos ou negócios da empresa e servem como fontes das tabelas de fato, depois povoam-se estas tabelas (KIMBALL, 1995).

### 3.4. Metadados

Os metadados são normalmente definidos como “dados sobre os dados” (HARRISON, 1998). Talvez uma definição mais exata seja que metadados é “uma abstração dos dados”, ou ainda, “dados de mais alto nível que descrevem dados de um nível inferior. Sem metadados, os dados não têm significado” (INMON, 1997).

Segundo (SINGH, 2001), “metadado é freqüentemente definido como dados sobre os dados, ou um pouco menos vago, como a informação necessária para tornar úteis dados científicos”.

De acordo com (SINGH, 2001), antes que o *Data Warehouse* possa ser acessado com eficiência, é necessário entender quais dados estão disponíveis e onde estão localizados. O metadado fornece a localização dos dados do *Data Warehouse* e ponteiros para esses dados. Além de ajudar a localizar os dados, o metadado pode conter:

- histórico de extração e transformação de dados;
- tamanho das tabelas o *Data Warehouse*;
- algoritmos de sumarização e modelagem de dados;
- estatísticas de uso de dados.

### 3.5. Povoamento de *Data Warehouse*

De acordo com (CAMPOS, 2000), a extração, limpeza, transformação e migração dos dados dos sistemas existentes na empresa para o *Data Warehouse* constituem tarefas críticas para o seu funcionamento efetivo e eficiente. Diversas técnicas e abordagens têm sido propostas, algumas bastante genéricas e outras especialmente voltadas para a manutenção da integridade dos dados num ambiente caracterizado pela derivação e replicação de informações.

Os processos associados com a migração de dados dos sistemas operacionais para o *Data Warehouse* incluem extração de dados, limpeza, transformação e carga no *Data Warehouse* (CAMPOS, 2000).

### 3.6. Obtenção de informações do *Data Warehouse*

Mesmo sabendo que a informação sobre o perfil do cliente ou do produto de sucesso de uma organização encontra-se de alguma forma entre os muitos *gigabytes* de dados de *marketing* e de vendas armazenados nos bancos de dados da organização, pode existir um longo caminho a ser percorrido até que esta informação esteja disponível. A extração eficaz da informação, de modo a subsidiar decisões, depende da existência de ferramentas especializadas que permitam a captura de dados relevantes e a sua visualização por intermédio de várias dimensões (CAMPOS, 2000).

Desse modo, as ferramentas não devem apenas permitir o acesso aos dados, mas também fornecer análises significativas, de tal maneira a transformar dados brutos em informação útil para os processos estratégicos da empresa. O sucesso de um *Data Warehouse* pode depender da disponibilidade da ferramenta certa para as necessidades de seus usuários (CAMPOS, 2000).

### 3.7. Conclusão

Um *Data Warehouse* é um sistema complexo que permite integrar um conjunto de componentes, tais como, diversos tipos de *software* e de *hardware*, redes de computadores, sistemas de comunicações de dados, servidores, *mainframes* e sistemas de administração de banco de dados, como também muitas pessoas de diferentes unidades organizacionais, com objetivos diferentes (LAMBERT, 1996).

Ao reunir informações de diferentes bancos de dados e plataformas, o *Data Warehouse* permite que sejam realizadas análises eficazes, transformando dados dispersos em informações estratégicas, que antes eram inacessíveis ou subaproveitadas (TAURION, 1997).

Segundo (HACKATHORN, 1997), a popularidade do *Data Warehouse* infelizmente obscureceu alguns de seus objetivos básicos. Existem negligências que estão levando à obtenção de resultados errôneos, por intermédio da execução de procedimentos errados. O *Data Warehouse* oferece uma grande variedade de modelos que tentam descrever os negócios, mas existe o problema da correta compreensão das informações geradas. Existe o desafio de manter a consistência e a credibilidade dos dados armazenados e dos dados gerados nas consultas e relatórios. Pode acontecer que

dois usuários interpretem os mesmos dados de maneiras diferentes, sendo que, na verdade, apesar de usarem os mesmos dados, estão realizando análises diferentes ou observando pontos de vista diferentes. Como o ambiente é extremamente dinâmico, uma mínima variação na consulta pode levar a resultados completamente diferentes, que podem ser confundidos por um usuário *distraído*. Os dados podem, também, ser interpretados por pessoas com pouca capacidade de análise dos negócios e estes, para superar suas dificuldades, podem fazer a culpa recair sobre o *Data Warehouse*.

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

O próximo capítulo apresenta duas metodologias utilizadas no desenvolvimento do *Data Warehouse*, as quais mostram os passos a seguir na manipulação da informação estratégica de uma organização de uma forma consistente e eficiente.

## 4. Metodologias de desenvolvimento do *Data Warehouse*

*“Os primeiros projetos de Data Warehouse muito provavelmente caminharam pelas trilhas metodológicas originadas basicamente em duas fontes de inspiração: Bill Inmon ou Ralph Kimball”.*(BARBIERI, 2001).

O sucesso do desenvolvimento do *Data Warehouse* depende fundamentalmente de uma escolha correta da estratégia a ser adotada, de forma que seja adequada às características e necessidades específicas do ambiente onde será implementado (WELDON, 1997).

Existe uma variedade de abordagens para o desenvolvimento do *Data Warehouse*, devendo-se fazer uma escolha fundamentada em pelo menos três dimensões: escopo do *Data Warehouse* (departamental, empresarial, etc), grau de redundância de dados, tipo de usuário alvo (WELDON, 1997). Em seguida, duas metodologias de desenvolvimento para o *Data Warehouse*, são apresentadas.

### 4.1. Metodologia segundo Ralph Kimball

O ciclo de vida para a implementação do *Data Warehouse* segundo Ralph Kimball, é ilustrado na Figura 4.1. Este diagrama descreve a seqüência de tarefas requerida para o projeto, desenvolvimento e aplicação do *Data Warehouse*.

O diagrama da Figura 4.1 mostra a metodologia de projeto na qual cada módulo serve como um indicador. As fases que compõem este diagrama são descritas a seguir, foram retiradas de (KIMBALL, 1998a).

#### 4.1.1. Planejamento de projeto

O ciclo de vida inicia-se com o planejamento de projeto. Esse dirige a definição e a abrangência do projeto do *Data Warehouse*, incluindo avaliação e justificação do negócio.

Essas são tarefas críticas, devido à alta visibilidade e custos associados com muitos projetos do *Data Warehouse*. Por isso, o planejamento de projeto enfoca os

recursos e o nível de habilidade requerida do pessoal, juntamente com o planejamento das tarefas associadas e a duração.

O planejamento de projeto identifica todas as tarefas associadas com o ciclo de vida empresarial e os comentários das partes envolvidas. Serve como base para a administração contínua do projeto do *Data Warehouse*.

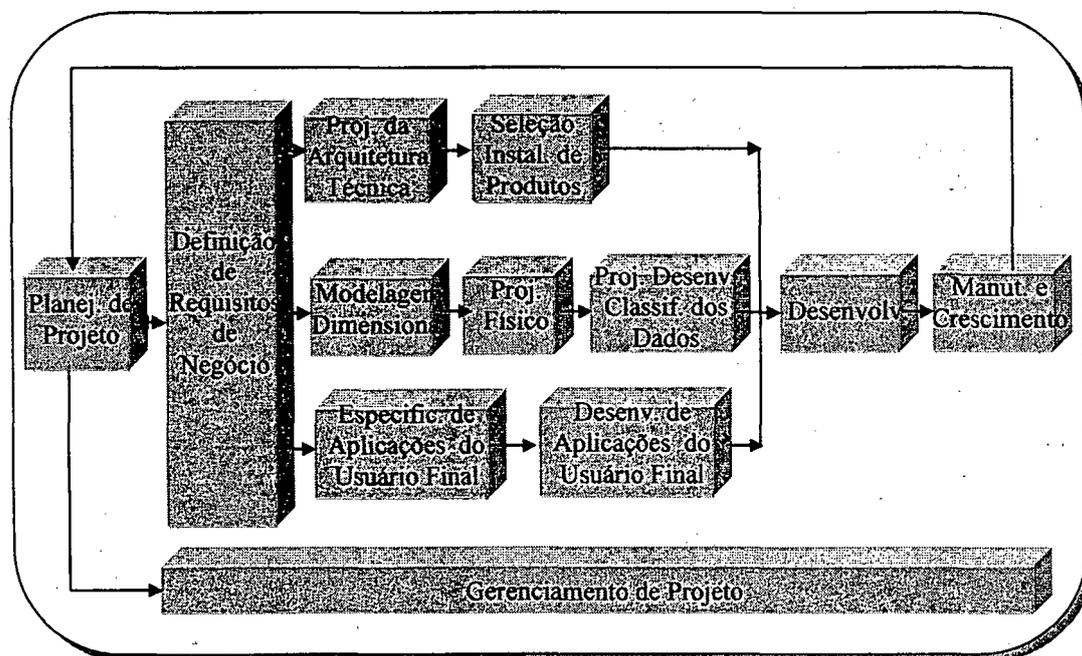


Fig. 4.1 – Diagrama do Ciclo de Vida.

Fonte: (KIMBALL, 1998a)

#### 4.1.2. Definição de requisitos de negócio

A probabilidade para o sucesso do *Data Warehouse* é aumentada quando se entende as exigências dos usuários. Projetistas do *Data Warehouse* têm que entender os fatores chaves que dirigem o negócio para determinar as exigências efetivamente empresariais e traduzir em considerações de implementação.

#### 4.1.3. Modelagem dimensional

A definição das exigências empresariais determina os dados que precisam ser armazenadas no *Data Warehouse* para atender às exigências dos usuários. Inicia-se construindo uma matriz que representa os processos de negócio e a dimensionalidade. A

matriz serve como uma fotocópia para assegurar que o *Data Warehouse* está extensível com o passar do tempo.

Dessa matriz uma análise de dados mais detalhada é realizada. A análise de dados é juntada com as exigências empresariais e então se desenvolve um Modelo Dimensional.

Este modelo identifica as dimensões associadas, atributos e fatos. O projeto lógico do banco de dados é completado com estruturas apropriadas e relacionamento entre as chaves.

#### **4.1.4. Projeto físico**

O enfoque principal do projeto de banco de dados físicos é definir as estruturas físicas necessárias para apoiar o projeto lógico do banco de dados. Os elementos primários deste processo incluem a definição da nomenclatura padrão e a construção do ambiente de banco de dados.

#### **4.1.5. Projeto e desenvolvimento da classificação dos dados**

O projeto e desenvolvimento da plataforma de classificação dos dados é tipicamente o mais subestimado no projeto do *Data Warehouse*. O processo de classificação de dados possui três passos principais: extração, transformação e povoamento. O processo de extração expõe questões de qualidade de dados que estiveram ocultos dentro do sistema operacional. Já que a qualidade dos dados tem um significativo impacto na credibilidade do *Data Warehouse*, necessita-se resolver esses problemas de qualidade durante a classificação dos dados. Para aumentar sua complexidade, necessita-se projetar e construir dois processos de classificação, um para o povoamento inicial e outra para as cargas regulares.

#### **4.1.6. Projeto da arquitetura técnica**

Ambientes do *Data Warehouse* requerem a integração de numerosas tecnologias. O projeto de arquitetura estabelece a arquitetura global.

Deve-se considerar três fatores: as exigências empresariais, ambientes técnicos atuais e o planejamento estratégico.

#### **4.1.7. Seleção e instalação de produtos**

Ao se usar o projeto de arquitetura, os componentes específicos, como a plataforma de *hardware* e o sistema de administração de banco de dados devem ser avaliados e selecionados.

Um processo de avaliação técnica é definido junto com fatores de avaliação específicos para cada componente arquitetônico. Uma vez que os produtos foram avaliados e selecionados, eles são instalados e testados para assegurar uma integração de fim-a-fim apropriada dentro do ambiente do *Data Warehouse*.

#### **4.1.8. Especificação de aplicações do usuário final**

Especificações de aplicações descrevem o modelo de relatório, parâmetros e cálculos exigidos. Estas especificações asseguram que a equipe de desenvolvimento e os usuários tenham um entendimento comum das aplicações a serem entregues.

#### **4.1.9. Desenvolvimento de aplicação do usuário final**

O desenvolvimento das aplicações do usuário envolve a configuração das ferramentas e a construção dos relatórios. Opcionalmente, aplicações que são construídas tendo acesso a ferramentas automatizadas, provêm maior produtividade para a equipe de desenvolvimento de aplicação. Além disso, oferece um poderoso mecanismo para os usuários facilmente modificarem os modelos de relatório existentes.

#### **4.1.10. Desenvolvimento**

Representa a convergência de tecnologia, dados e aplicações de usuário. Um planejamento é exigido para assegurar que tudo seja ajustado corretamente. Deve ser estabelecido apoio ao usuário e estratégias de realimentação antes que qualquer usuário tenha acesso ao *Data Warehouse*.

#### **4.1.11. Manutenção e crescimento**

Bastante trabalho segue o desenvolvimento inicial do *Data Warehouse*. É necessário continuar enfocando os usuários, proporcionando apoio contínuo e

treinamento. Deve ser medido e anotado a aceitação e desempenho do *Data Warehouse* com o passar do tempo.

As mudanças devem ser vistas como um sinal de sucesso, não de fracasso. Devem ser estabelecidos processos de prioridades para tratar esta demanda de modificações, para se obter uma evolução e crescimento.

Depois que as prioridades de projeto sejam identificadas, se regressa para o começo do ciclo de vida, construindo o que já foi estabelecido no ambiente do *Data Warehouse*, com um enfoque nas novas exigências.

#### **4.1.12. Gerenciamento de projeto**

O gerenciamento de projeto assegura que as atividades do ciclo de vida permaneçam dentro do esperado e em sincronização. Como mostra a Figura 4.1, as atividades de gerenciamento de projeto acontecem ao longo do ciclo de vida.

Estas atividades têm enfoque no monitoramento do estado do projeto, localização do assunto e controle das mudanças para preservar os limites do *Data Warehouse*.

### **4.2. Metodologia segundo Inmon**

A metodologia desenvolvida por Willian H. Inmon é composta por vários passos. Inicialmente é necessário que um modelo de dados tenha sido definido. Precisa ter identificado as principais áreas de interesse, definindo claramente as fronteiras do modelo, separando dados primitivos de dados derivados, identificando chaves, atributos, agrupamentos de atributos, relacionamento entre agrupamentos de atributos, dados repetitivos e tipos de dados para cada área de interesse. Os passos são ilustrados na Figura 4.2 e descritos nas sub-seções a seguir, retirados de (INMON, 1997).

#### **4.2.1. Análise do modelo de dados**

Este passo resulta na confirmação de que a organização construiu um sólido modelo de dados. Se o modelo não atender aos critérios especificados, o andamento do projeto deve ser interrompido até que o modelo seja elevado a um padrão aceitável de qualidade.

### 4.2.2. Dimensionamento

Uma vez que o modelo de dados tenha sido analisado e transferido a um nível de qualidade aceitável, o próximo passo consiste em realizar o dimensionamento. O dimensionamento é uma estimativa do ambiente do SAD (Sistema de Apoio à Decisão). Se o volume de dados vai ser um problema, é importante saber disso no início. O dimensionamento simplesmente projeta, em termos brutos, que quantidade de dados o *Data Warehouse* vai conter.

O resultado do dimensionamento é simples, se o *Data Warehouse* irá conter grandes quantidades de dados, é necessário levar em consideração a possibilidade de existência de vários níveis de granularidade<sup>1</sup>. Se o *Data Warehouse* não for destinado a conter uma enorme quantidade de dados, não existe a necessidade de planejar o projeto de vários níveis de granularidade.

### 4.2.3. Avaliação técnica

Os requisitos técnicos para o gerenciamento do *Data Warehouse* são muito diferentes dos requisitos e considerações técnicas para o gerenciamento de dados e processamento no ambiente operacional. Esse é o motivo pelo qual é tão comum a existência de um depósito central de dados SAD.

Quando apropriadamente conduzido, a definição técnica do *Data Warehouse* satisfaz aos seguintes critérios:

- capacidade de gerenciar grandes quantidades de dados;
- capacidade de permitir que os dados sejam acessados de modo flexível;
- capacidade de organizar dados de acordo com um modelo de dados;
- capacidade tanto de receber quanto de enviar dados para uma ampla variedade de tecnologias;
- capacidade de periodicamente carregar grandes quantidades de dados;
- capacidade de acessar um registro por vez.

---

<sup>1</sup> Granularidade: nível de detalhe contido em uma unidade de dados, quanto mais detalhe houver, mais baixo o nível de granularidade e quanto menos detalhe existir, mais alto o nível de granularidade.

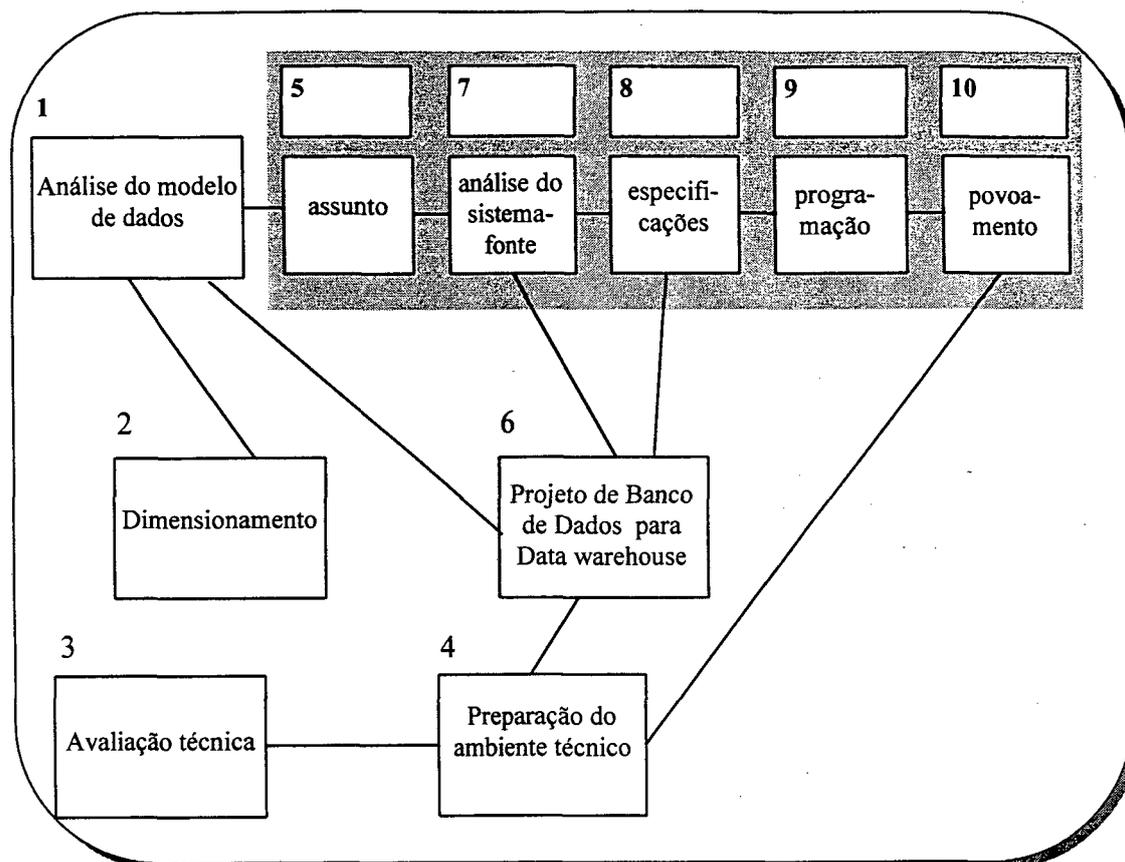


Fig. 4.2 – Desenvolvimento do *Data Warehouse*.

Fonte: (INMON, 1997)

#### 4.2.4. Preparação do ambiente técnico

Uma vez que a configuração da arquitetura do *Data Warehouse* tenha sido definida, o próximo passo consiste em identificar tecnicamente como a configuração pode ser acomodada. Algumas das questões típicas que precisam ser tratadas aqui são:

- a quantidade de DASD (Dispositivos de Armazenamento de Acesso Direto, por exemplo, o *winchester*) necessária;
- que enlace, quer fora da rede ou em rede, será necessário;
- o volume de processamento previsto;
- como diminuir ou atenuar conflitos de processamento entre programas de acesso concorrente;
- o volume de tráfego que será gerado pela tecnologia que controla o *Data Warehouse*;

- a natureza do tráfego, interrupções curtas ou longas, geradas pela tecnologia que controla o *Data Warehouse*.

Quando esse passo é apropriadamente conduzido, não existe impedimento técnico para o sucesso. Entre os componentes técnicos que devem estar instalados, alocados, ligados e prontos para receber dados encontram-se a rede, o DASD, o sistema operacional que gerencia o DASD, a interface do *warehouse*, o *software* usado para gerenciar o *Data Warehouse*.

#### **4.2.5. Análise das áreas de interesse**

Neste ponto é selecionada a área de interesse a ser povoada. A primeira área de interesse a ser selecionada deve ser suficientemente grande para ter sentido e suficientemente pequena para ser implementada. Se, por acaso, uma área de interesse for verdadeiramente grande e complexa, um subconjunto desta pode ser escolhido para sua implementação. O resultado deste passo é um escopo de empreendimento em relação a uma área de interesse.

Quando essa fase é conduzida corretamente, o resultado é uma definição de que área de interesse (ou assunto) deve ser povoada a seguir. Nos primeiros povoamentos, geralmente são selecionadas pequenas áreas de interesse. Quando tudo é feito de forma apropriada, o assunto selecionado para ser povoado atende às necessidades do nível corrente de desenvolvimento do *Data Warehouse*.

#### **4.2.6. Projeto do *Data Warehouse***

O *Data Warehouse* é projetado com base no modelo de dados. Algumas das características do projeto mais recente incluem:

- uma acomodação dos diferentes níveis de granularidade, caso existam vários;
- uma orientação de dados para os principais assuntos da empresa;
- a presença de somente dados primitivos e dados derivados públicos;
- a ausência de dados não SAD (Sistema de Apoio à Decisão);
- a variação em relação ao tempo de cada registro de dados;
- desnormalização física de dados onde aplicável, ou seja, onde a performance exigir;

- criação de artefatos de dados por meio dos quais, os dados que estiverem no ambiente operacional são passados para o *Data Warehouse*.

O resultado deste passo é um projeto físico de banco de dados para o *Data Warehouse*. Perceba que, no início, nem todo o *Data Warehouse* precisa ser projetado em detalhes. É recomendável inicialmente projetar as principais estruturas do *Data Warehouse* e, então, preencher os detalhes em um momento posterior.

#### **4.2.7. Análise do sistema-fonte**

Uma vez que o assunto a ser povoado é definido, a próxima etapa consiste em identificar, no ambiente de sistemas existentes, a fonte de dados para o assunto. É muito normal que existam várias fontes para os dados SAD. É neste momento que são tratadas as questões de integração. A seguir estão listadas as questões que devem ser tratadas aqui:

- estrutura de chave e escolha de chave à medida que os dados passam do ambiente operacional para o ambiente SAD (Sistema de Apoio à Decisão);
- o que fazer quando existe diversas fontes de dados para escolher, o que fazer quando não existe fontes de dados para escolher, quais transformações, codificação, decodificação, conversões, precisam ser feitas à medida que os dados são selecionados para a passagem para o *Data Warehouse*;
- de que modo a variação em relação ao tempo será criada a partir dos dados de valor corrente;
- como a estrutura de SAD será criada a partir da estrutura operacional;
- de que modo os relacionamentos operacionais serão representados no ambiente SAD (Sistema de Apoio à Decisão).

O resultado desse passo é o mapeamento de dados do ambiente operacional para o ambiente SAD.

#### 4.2.8. Especificações

Uma vez que a interface entre os ambientes operacionais e SAD tenha sido delineada, o próximo passo consiste em formalizá-la em termos de especificações de programas. Algumas das questões principais deste passo são as seguintes:

- Como saber quais dados operacionais varrer?
  - Os dados operacionais apresentam marca de tempo?
  - Há algum arquivo *delta*<sup>2</sup>?
  - Há *logs* de sistemas ou auditoria que possam ser usados?
  - O código-fonte e a estrutura dos dados existentes podem ser alterados para a criação de um arquivo *delta*?
  - É necessário utilizar imagens anteriores e posteriores?
  
- Como armazenar a saída, uma vez pesquisada?
  - Os dados SAD são pré-allocados e pré-formatados?
  - Os dados são acrescentados?
  - Os dados são substituídos?
  - São feitas atualizações no ambiente SAD?

O resultado desse passo é a descrição dos programas que serão usados para efetuar a passagem dos dados do ambiente operacional para o ambiente do *Data Warehouse*.

#### 4.2.9. Programação

Este passo inclui todas as atividades padrão de programação, como as seguintes:

- desenvolvimento de pseudocódigo;
- codificação;
- compilação;
- *walkthroughs*<sup>3</sup>;
- testes de unidade, de estresse, em suas diversas formas.

---

<sup>2</sup> Arquivo delta: arquivo histórico que registra as modificações do sistema.

<sup>3</sup> *Walkthroughs*: é a explicação verbal do código do programa na presença de pares de pessoas. O objetivo consiste em encontrar erros, ou qualquer outro problema, antes do teste.

Quando corretamente executado, o código que é gerado nesse passo é eficiente, documentado, com possibilidade de ser facilmente alterado, exato e completo.

#### 4.2.10. Povoamento

Este passo é a execução dos programas SAD (Sistema de Apoio à Decisão) anteriormente desenvolvidos. As questões aqui tratadas são as seguintes:

- frequência de povoamento;
- eliminação de dados povoados;
- obsolescência dos dados povoados, ou seja, a execução de programas de resumo de conferência;
- gerenciamento de vários níveis de granularidade;
- renovação dos dados de amostra viva, caso tenham sido construídas tabelas de amostra viva.

O resultado desse passo consiste em um *Data Warehouse* povoado e funcional.

### 4.3. Conclusão

Neste capítulo duas metodologias foram apresentadas para o desenvolvimento do *Data Warehouse*. A primeira metodologia descrita foi desenvolvida por Ralph Kimball, cuja abordagem tem um estilo mais simples e incremental. A metodologia apresenta o Esquema Estrela - *Star Scheme* (GALLAS, 1999), o qual utiliza projetos de *Data Marts* separados, que deverão ser integrados na medida da sua evolução. Tem como desvantagem a possibilidade de se produzir diversos *Data Marts*, sem uma perfeita integração entre eles, além de uma provável duplicação de esforços na fase de extração, preparação e carga dos dados (BARBIERI, 2001).

Segundo (BARBIERI, 2001) a essência da abordagem de Kimball está na etapa de projeto dos *Data Marts*. Centrada na modelagem dimensional, com o conceito de esquema estrela (*Star Schema*), essa abordagem transforma os dados em tabelas de fato (onde se encontram os dados de interesse, possíveis para manipulação numérica e estatística), e em tabelas de dimensão (tabelas que possuem as chaves de entrada do modelo, além das informações descritivas de cada dimensão).

Na segunda metodologia de Inmon, a abordagem se concentrou inicialmente no estilo mais tradicional de construção de Bancos de Dados, muito próximo daquele

surgido nos primeiros projetos de Bancos de Dados que buscavam uma forte integração entre todos os dados da empresa que habitavam áreas funcionais diferentes. Isso seria representado num modelo único, integrado e coeso, mas que, por vezes, mostrou-se rígido e de difícil consecução. O ciclo metodológico básico se iniciava pela análise de requerimentos de negócios, seguido pela Modelagem de Dados e pela Análise das fontes de Dados, fator preponderante no sucesso de qualquer *Data Warehouse* ou *Data Mart*. Seguia-se a fase de Projeto e *Design*, onde se faz o detalhamento do *Data Warehouse*, sua especificação e construção, o seu teste e validação e finalmente a sua implementação. Tudo muito parecido com a forma tradicional de projeto de Banco de Dados (BARBIERI, 2001).

Observa-se que essas duas abordagens apresentadas são ligeiramente diferenciadas mas, não são excludentes, do ponto de vista de início do projeto. Em uma, o *Data Warehouse*, segundo Inmon, deve ser um depósito central devidamente preparado a partir dos dados operacionais. E dele, deveriam nascer os *Data Mart*. Com o amadurecimento dessas abordagens pelas reais necessidades de mercado, Inmon hoje já incorpora na sua metodologia original os conceitos de *Data Mart*, originados diretamente dos dados operacionais, desde que devidamente integrados por um processo de integração sem rupturas (BARBIERI, 2001).

Na outra abordagem, a de Kimball, aceita a idéia de que os *Data Marts* possam ser derivados (de forma incremental) diretamente dos dados operacionais, sem a presença de grande depósito inicial que os origine. De qualquer maneira, os métodos de projeto de Kimball podem ser plenamente utilizados na metodologia de Inmon, desde que o foco seja ajustado devidamente (BARBIERI, 2001).

A convergência atual das duas abordagens segue a linha de que se deve iniciar a construção do *Data Warehouse*, pelos primeiros *Data Mart*. Uma fase inicial que define a compatibilidade das tabelas de Dimensão dos *Data Mart* e alguns padrões de dados das tabelas de fato, deve ser respeitada, pois são fatores fundamentais para diminuir possíveis rupturas. Os outros *Data Marts*, a seguir deverão ser integrados e manter a conformidade entre as dimensões (BARBIERI, 2001).

No desenvolvimento de nosso trabalho de pesquisa, portanto, utilizaremos o conceito de modelagem dimensional em esquema estrela (primeira metodologia), baseado nas informações bibliográficas consultadas e pelo tratamento dos dados

estratégicos que garantem ser confiáveis e seguros uma vez definidos no modelo do projeto. Integrou-se a essa metodologia a Engenharia da Informação, para o tratamento das necessidades de informação da organização. Com isso, eliminou-se um dos fatores críticos de sucesso no desenvolvimento de projetos de *Data Warehouse*. Foi possível ter uma melhor idéia dos resultados esperados do *Data Warehouse* desenvolvido.

O próximo capítulo descreve a aplicação da Engenharia da Informação no desenvolvimento do Modelo Dimensional do *Data Warehouse*, para que os dados acessados representem o espoco da informação real da organização e possibilitem a tomada de decisões.

## 5. Aplicação da Engenharia da Informação no desenvolvimento do *Data Warehouse*

*“Embora a Informática seja pródiga em produzir soluções de consistência espumante ... a proposta de se concentrar um arsenal de tecnologia em prol dos clientes é algo que parece também definitivo”.* (BARBIERI, 2001).

De acordo com (CAMPOS, 2000), o *Data Warehouse* surgiu nos últimos anos como uma promessa de solução ao maior problema das pessoas encarregadas de tomar decisões corporativas: o acesso a dados históricos, de forma segura e consistente, garantindo que a tomada de decisões seja realmente baseada na realidade da empresa, possibilitando representar uma arma tática para obter vantagem competitiva sobre os concorrentes. A informação tem que estar orientada em atender, na hora certa, aquilo que necessita na medida necessária.

A aplicação da Engenharia da Informação na organização foi realizada por intermédio da metodologia desenvolvida por (NETO, 1988), e não foram realizadas alterações ou complementos nessa metodologia.

Este capítulo está organizado como segue. A seção 5.1 apresenta os objetivos deste trabalho. A seção 5.2 mostra as etapas da aplicação da Engenharia da Informação na organização, tais etapas são o Planejamento Estratégico das Informações e a Análise das Áreas de Negócios da Empresa. A seção 5.3 apresenta a conclusão.

### 5.1. Objetivo

O trabalho tem como objetivo principal propor a aplicação da Engenharia da Informação no processo da modelagem dimensional do *Data Warehouse* por meio do desenvolvimento de um modelo orientado a uma organização de comércio eletrônico.

### 5.2. Breve histórico sobre a organização

A Paradigma ([www.pta.com.br](http://www.pta.com.br)) é uma empresa especializada no desenvolvimento de soluções de negócios via Internet. Começou suas operações em 1998, em Florianópolis, SC. Em apenas três anos de atuação no mercado, a empresa já aponta um crescimento superior a 1000% ao ano, o que lhe garantiu o 1º lugar no

*ranking* da Info 2000, na categoria de empresa de *software* que mais cresceu. Além disso, a empresa recebeu o prêmio de melhor solução de *e-commerce* do mundo pela *Microsoft Partners Fusion Awards 2001* e é também a primeira empresa brasileira a exportar *software* para *e-commerce*.

Existem várias ferramentas desenvolvidas por esta empresa, dentre elas algumas usando a abordagem da Inteligência aplicada em *Intelligent Agents*, *Data Miner*, *Data Warehouse*. Baseadas nestas abordagens algumas ferramentas foram desenvolvidas para possibilitar a gestão inteligente de um negócio, transformando-o em um agente pró-ativo baseado em *learning machines*, gerando recursos para o mundo do *e-Business* como *recommender systems* e *data miner*. Os diferenciais desenvolvidos para estas ferramentas estão no contexto de agência e nos mecanismos de aprendizagem conceituados em *Fuzzy Logic*.

A Paradigma, nos últimos quatro anos, vem trabalhando com *e-Knowledge Management* e *e-Competitive Intelligence*, composição do cenário geral de *e-Business Intelligence*. O objetivo deste trabalho é totalmente voltado ao segmento de Inteligência Empresarial Estratégica de um negócio afim, possibilitando nos momentos em que a tomada de decisão possa ser humana, de que a mesma venha a ser baseada em apontamentos precisos buscando um alvo de forma mais equilibrada. Quando a decisão não pode ser tomada por seres humanos, trabalha-se com agentes inteligentes para que possam ser executadas as ações, baseadas em regras mineradas num ambiente qualquer de *Data Warehouse*.

### **5.3. Estabelecimento das etapas da metodologia**

Este trabalho fará o uso das duas primeiras fases da Engenharia da Informação, o Planejamento Estratégico de Informações e a Análise das Áreas de Negócios da Empresa. Utilizará apenas a face dos Dados, pois esta “fornece a base de sustentação das informações necessárias para a sobrevivência da empresa” (NETO, 1988), que é o principal enfoque deste trabalho. A partir de informações bem planejadas e documentadas, será utilizada a modelagem dimensional segundo (BARBIERI, 2001), o Modelo Dimensional resultante será implementado de acordo com (SPENIK, 2001).

### 5.3.1. Planejamento estratégico de informações

Segundo (NETO, 1988), normalmente inicia-se o processo de Engenharia da Informação desenvolvendo um planejamento estratégico de informações. Este é composto dos objetivos e as metas da organização e também na forma pela qual o uso da tecnologia poderá criar novas oportunidades ou vantagens com a concorrência.

A seguir um breve resumo sobre a metodologia de Engenharia da Informação para a aplicação no planejamento estratégico de informações da organização.

#### a. Resumo da metodologia

Nessa fase, inicialmente, define-se a **missão da empresa**, que estabelece o propósito mais amplo da organização. A seguir, define-se os **objetivos da empresa**, que deverão ser a base de sustentação e apresentar total compatibilidade com a missão estabelecida (NETO, 1988).

Identificada a missão e os objetivos da empresa, deve-se elaborar o modelo organizacional de sustentação, de forma a identificar os executivos de níveis estratégico, tático e operacional que estabelecerão respectivamente os **objetivos da área funcional**, as **metas** e os **desafios** (NETO, 1988).

Os **objetivos da área funcional** definem o que deve ser atingido pela referida área, para satisfazer um ou mais objetivos da empresa (NETO, 1988).

As **metas** são identificados pelos executivos de nível tático e podem ser definidas como resultados quantificados, isto é, que se espera atingir para cada objetivo da área funcional à qual o executivo está alocado (NETO, 1988).

Os **desafios** são de responsabilidade dos executivos de nível operacional e tratam de quantificações e ações mais específicas das metas estabelecidas (NETO, 1988).

Definidos os objetivos, metas e desafios deve-se determinar os **fatores críticos de sucesso** que são aquelas poucas coisas que devem ir corretamente na empresa para que os objetivos sejam alcançados (NETO, 1988).

A seguir, deverão ser relacionados os **problemas** que estão dificultando alcançar o fator crítico de sucesso. Posteriormente, cada executivo relacionará suas **necessidades de informações** que levarão ao melhor atendimento dos seus fatores críticos de sucesso (NETO, 1988).

Identificadas as necessidades de informações, podemos obter as **entidades** e as respectivas **classes de dados** envolvidas (NETO, 1988). Em resumo, podemos visualizar a metodologia de forma geral na Figura 5.1.

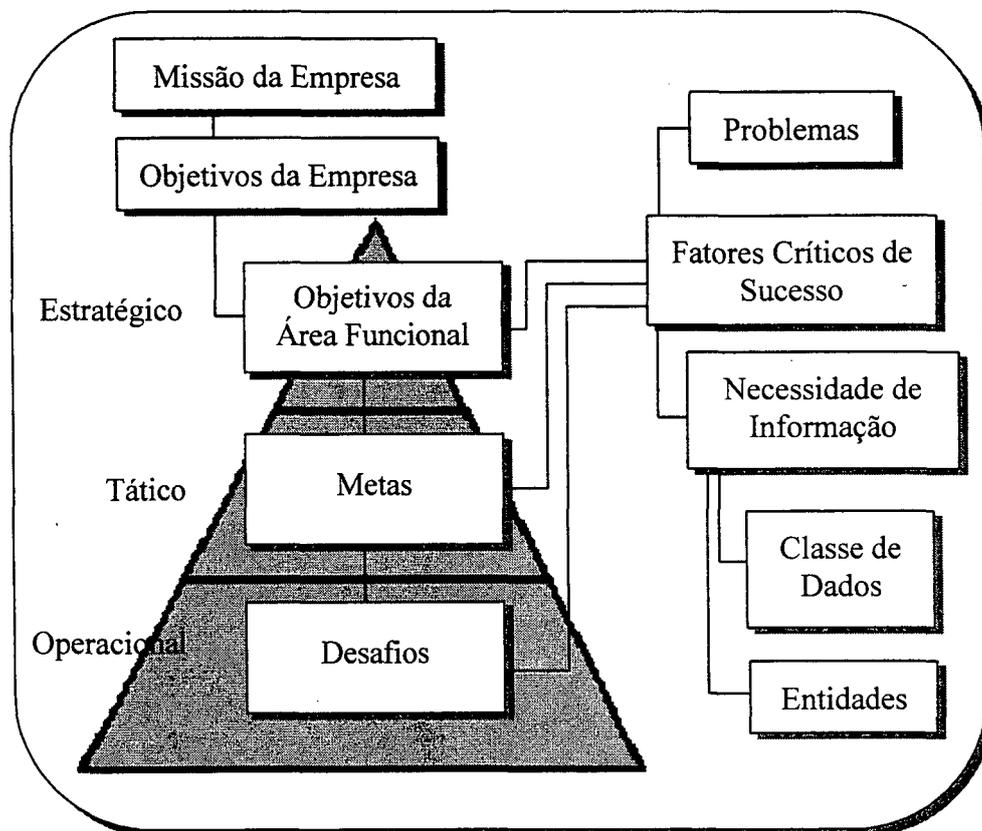


Fig. 5.1 – Resumo da metodologia.

Fonte: (NETO, 1988)

### b. Aplicação da metodologia

A aplicação da metodologia poderá resultar em desvios significativos e, por vezes, irrecuperáveis, caso não se apóie em fontes sólidas e confiáveis (NETO, 1988). Para tanto, foram realizadas várias reuniões de forma a esclarecer e compatibilizar todas as informações fornecidas pela organização.

A Figura 5.2 ilustra a seqüência a ser seguida da metodologia de Engenharia da Informação para o planejamento estratégico de informações, aplicada no nível estratégico da empresa.

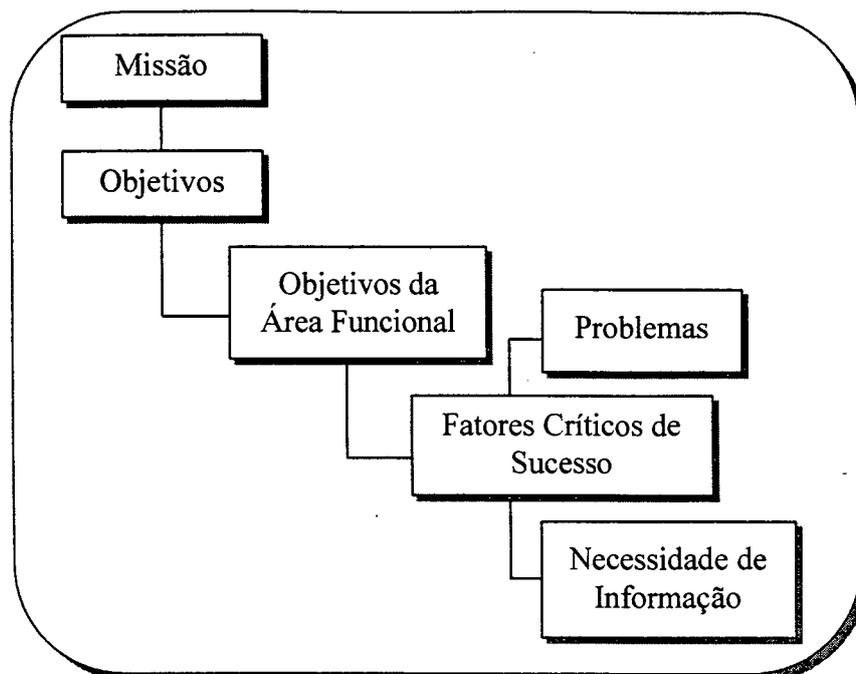


Fig. 5.2 – Planejamento estratégico de informações em nível estratégico.

Fonte: (NETO, 1988)

Seguindo os passos ilustrados na Figura 5.2, obteve-se os seguintes resultados:

- **Missão da empresa:** fornecer produtos em âmbito mundial;
- **Objetivo da empresa:** fornecer produtos de qualidade;
- **Objetivo da área funcional:** manter as características de diferenciação, exclusividade e atualização dos produtos;
- **Fator crítico de sucesso:** analisar as ações dos agentes (clientes) que impedem ou que substituem a ação de comprar;
- **Problema:** mudar o foco do fato, para poder avaliar ações negativas;
- **Necessidades de informação:** relatórios que especifiquem as seguintes informações:
  - compra do produto pelo cliente de determinada localidade;
  - produtos não comprados por determinados clientes de uma determinada localidade;
  - determinados produtos comprados por determinados clientes de determinadas localidades;
  - produtos não comprados por clientes de determinada localidade.

A partir das necessidades de informações identificadas, pode-se obter as entidades envolvidas, que são: *Agente*, *Produto*, *Transacao* e a classe de dados denominada *Localizacao*.

Pode-se verificar a presença de uma entidade *Fornecedor*, mesmo não sendo citada como uma necessidade, isto porque é necessário que alguém forneça os produtos para que a organização possa negociá-los.

Como este trabalho é voltado para a análise do comportamento das pessoas em um site de comércio eletrônico, vê-se a necessidade de armazenar algumas informações na base de dados relacional, então adicionou-se as entidades *Sessão* e *Página*. Uma sessão será aberta quando o agente *entrar* no *site*, dentro desta, estão contidas várias Páginas, por exemplo, uma página de cadastro de clientes, e o agente poderá realizar vários tipos de eventos nessas páginas, como, por exemplo, uma consulta de produto, um pedido, cadastro.

### c. Modelo de dados corporativo

O modelo de dados corporativo deverá conter todas as informações necessárias para que o negócio da empresa caminhe em boa forma, cumprindo a missão e os objetivos fixados, de forma que todos os propósitos sejam atingidos, sem que falte qualquer componente (NETO, 1988).

A Figura 5.3 ilustra o modelo de dados corporativo obtido pela aplicação da metodologia. Pode-se observar que o *Agente*<sup>4</sup> possui um endereço representado pela entidade *Localizacao*. Uma *Transacao* é composta por vários *Produtos*. Tais produtos são fornecidos por uma organização, representado pela entidade *Fornecedor*. Como foi dito anteriormente, um agente, sendo cadastrado ou não, quando visita o *site* de comércio eletrônico, abre uma *Sessao* contendo várias *Paginas* e podendo executar algumas ações como: consultar ou comprar um produto, consultar, incluir ou alterar seus dados de cadastro, consultar ou alterar seu pedido.

---

<sup>4</sup> Nomes de entidades (*Agente*, *Localizacao*, entre outras) e campos (*CodigoAgente*, *Endereco* e demais) são definidos sem acentuação no banco de dados.

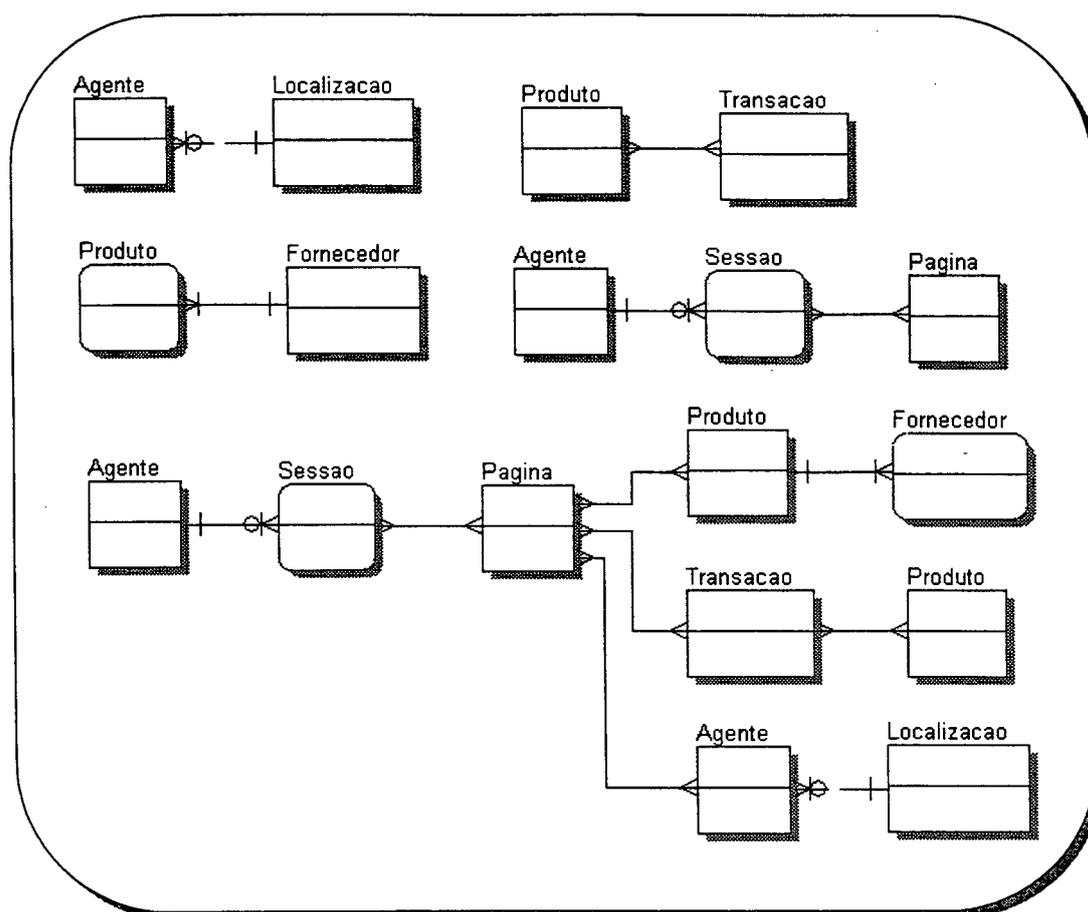


Fig. 5.3 – Modelo de dados corporativo.

### 5.3.2. Análise das áreas de negócios da empresa

Na análise das áreas de negócios, a face dos dados estará preocupada com a obtenção do modelo de dados da área escolhida. Com a base de entidades praticamente definida, e a identificação individual dos atributos de cada classe dos dados obtêm-se como resultado final um modelo estável porque as anomalias causadas principalmente pelas redundâncias e pela falta de normalização foram eliminadas (NETO, 1988).

#### a. Decompondo classe de dados em atributos

Inicia-se a fase de planejamento estratégico de informações que visa a face dos dados, com a identificação das classes de dados que suportarão as necessidades de informações. As classes de dados são grupos de atributos, que na fase de planejamento devem ser decompostas em atributos (NETO, 1988).

Após a identificação das necessidades de informação, decompõem-se as entidades *Agente*, *Produto*, *Fornecedor*, *Transacao* e a classe de dados *Localizacao* em atributos, chegando-se ao seguinte resultado, como mostra a Tabela 5.1.

Tabela 5.1 – Atributos das entidades identificadas.

<i>Entidades</i>	<i>Atributos</i>	<i>Atributos</i>	<i>Atributos</i>
<i>Agente</i>	CodigoAgente	Bairro	Cidade
	Estado	Pais	Endereco
	Numero	Complemento	Cep
	Email	FoneComercial	Fax
	Contato	Nome	Cpf
	Rg	FoneResidencial	
<i>Fornecedor</i>	CodigoFornecedor	Bairro	Cidade
	Estado	Pais	Endereco
	Numero	Complemento	Cep
	Email	FoneComercial	Fax
	Contato	RazaoSocial	NomeFantasia
	InscricaoEstadual	PercentualInvestidoMarketing	Cnpj
	TelefoneComercial		
<i>Produto</i>	CodigoProduto	NomeProduto	Preço
	NomeEmIngles	NomeEmEspanhol	DisponibilidadeEstoque
	Garantia	DataLancamento	
<i>Localizacao</i>	Bairro	Cidade	Estado
	Pais	BlocoEconomico	Moeda
<i>Transacao</i>	CodigoTransacao	TipoTransacao	DataTransacao
	DataEntrega	Quantidade	ValorUnitario
<i>Sessao</i>	CodigoSessao	Data	Hora
	StatusCorrenteAgente	TempoVidaSessao	
<i>Pagina</i>	CodigoPagina	TipoPagina	TipoEvento
	CaminhoPagina		

Segundo (NETO, 1988), recomenda-se que, nessa fase, não exista a preocupação de aplicar a normalização, deixando esta tarefa para as próximas fases. A tentativa de realizar várias tarefas em ocasiões inadequadas pode originar projetos confusos, com um conjunto de trabalhos inacabados e malfeitos.

## b. Simplificação do modelo

De acordo com (NETO, 1988), regularmente se definem tabelas perfeitamente normalizadas, mas que, no contexto do modelo de dados, apresentam incoerências, e até redundâncias que não são perfeitamente visíveis.

Observe na Tabela 5.1, existe a coluna com os atributos da entidade *Agente* e a outra coluna com os atributos da entidade *Fornecedor*. Alguns atributos são idênticos nas duas colunas. Podemos representar os atributos contidos nas entidades *Agente* e *Fornecedor* da Tabela 5.1 como observado na Figura 5.4.

Na Figura 5.4 observa-se que os atributos Bairro, Cidade, Estado, Pais, Endereco, Numero, Complemento, Email, FoneComercial, Fax e Contato são comuns nas duas entidades (*Agente* e *Fornecedor*). Para eliminar a redundância de atributos, foi realizado a simplificação nas entidades acima, tendo como resultado as entidades ilustradas na Figura 5.5.

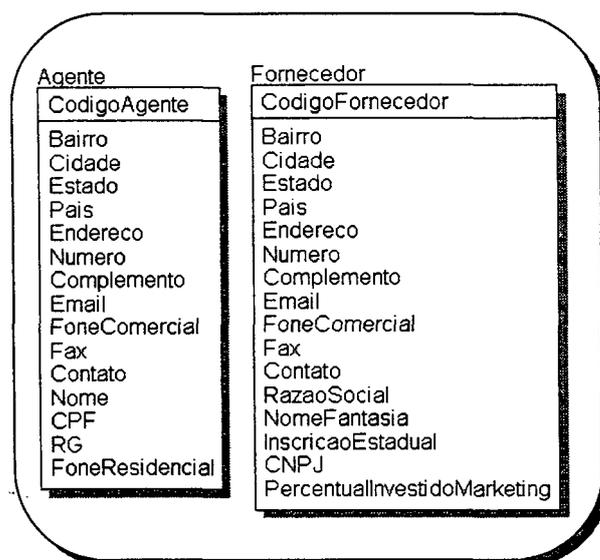


Fig. 5.4 – Representação gráfica das entidades *Agente* e *Fornecedor*.

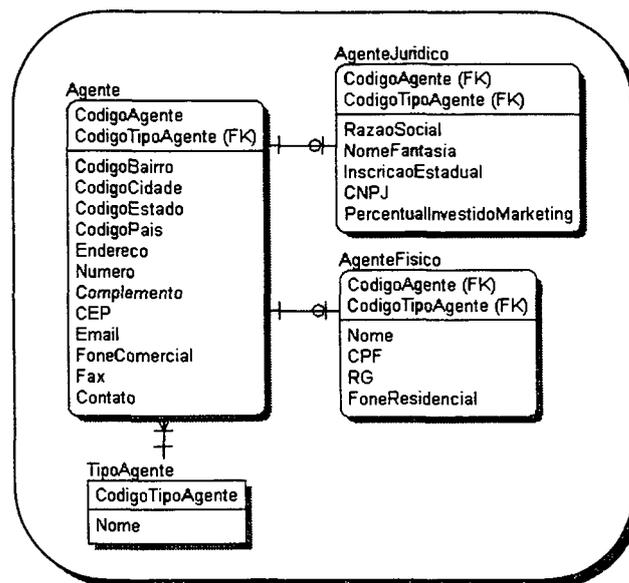


Fig. 5.5 – Entidade Agente Físico e Agente Jurídico.

A Figura 5.5 mostra que os atributos que são idênticos nas duas entidades (*Agente* e *Fornecedor*), foram agrupados na entidade *Agente*. Os atributos que não eram comuns as duas entidades foram distribuídos na entidade *AgenteJuridico* para os atributos dos fornecedores, e na entidade *AgenteFisico* para os dados dos clientes, com o objetivo de simplificar e evitar redundância no modelo de dados.

Criou-se uma nova entidade chamada de *TipoAgente* com os atributos *CodigoTipoAgente* e *Nome* para poder identificar se o agente cadastrado é um cliente ou um fornecedor. Por exemplo, se o campo *CodigoTipoAgente* na entidade *Agente* tiver valor um, então essa pessoa é um cliente. Se possuir valor dois, será um fornecedor. Essa nova entidade é muito útil, pois se no futuro a organização precisar incluir algum outro tipo de cliente ou fornecedor, até mesmo outro tipo de parceiro, o modelo de dados não precisará ser alterado, mantendo a estabilidade e a confiabilidade.

### c. Normalização

Em 1970 E. F. Codd introduziu o conceito de normalização. Esta técnica é um processo matemático formal e tem suas bases na teoria de conjuntos (MACHADO, 1996).

De acordo com (MARTIN, 1991a), a “normalização de dados refere-se ao modo como os itens de dados são agrupados em estruturas de dados.”

Segundo (NETO, 1988) a “normalização é um processo formal passo a passo que examina os atributos de uma entidade, com o objetivo de evitar anomalias observadas na inclusão, exclusão e alteração de tuplas específicas”.

Em (KERN, 1994),

a normalização serve para analisar tabelas e organizá-las de forma que a sua estrutura seja simples, relacional e estável, para que o gerenciamento possa ser também simples, eficiente e seguro. Os objetivos são evitar a perda e a repetição da informação e atingir uma forma de representação adequada para o que se deseja armazenar.

A Figura 5.6 mostra todas as entidades participantes do modelo de dados que foram normalizadas segundo (KERN, 1994), (MARTIN, 1991a), (MACHADO, 1996), (DATE, 1990).

A entidade *Produto*, como observado na Figura 5.6, já está normalizada. Esta entidade não possui nenhum atributo multivalorado<sup>5</sup>, sendo que não precisa ser alterada.

A classe de dados *Localizacao* precisa ser normalizada, os atributos Bairro, Cidade, Estado, Pais, Bloco econômico e Moeda são atributos multivalorados. Para normalizá-los foram criadas as seguintes entidades:

- *Bairro*: com os atributos *CodigoBairro* e *Nome*;
- *Cidade*: com os atributos *CodigoCidade* e *Nome*;
- *Estado*: com os atributos *CodigoEstado*, *Nome*, *TemperaturaMedia* e *Costumes*;
- *Pais*: com os atributos *CodigoPais* e *Nome*;
- *BlocoEconomico*: com os atributos *CodigoBlocoEconomico* e *Nome*;
- *Moeda*: com os atributos *CodigoMoeda*, *Nome* e *Simbolo*;
- *ConversaoDolar*: com os atributos *CodigoConversaoMoeda*, *DataFatorConversao* e *FatorConversao*.

Na entidade *Transacao* apenas o atributo *TipoTransacao* não está normalizado. Assim, criou-se uma nova entidade chamada de *TipoTransacao* com os atributos *CodigoTipoTransacao* e *Nome*.

---

<sup>5</sup> Atributo e Domínio multivalorado ou múltiplos podem assumir diversos valores em cada registro da tabela, por exemplo, os itens de produtos em uma nota fiscal (KERN, 1994).

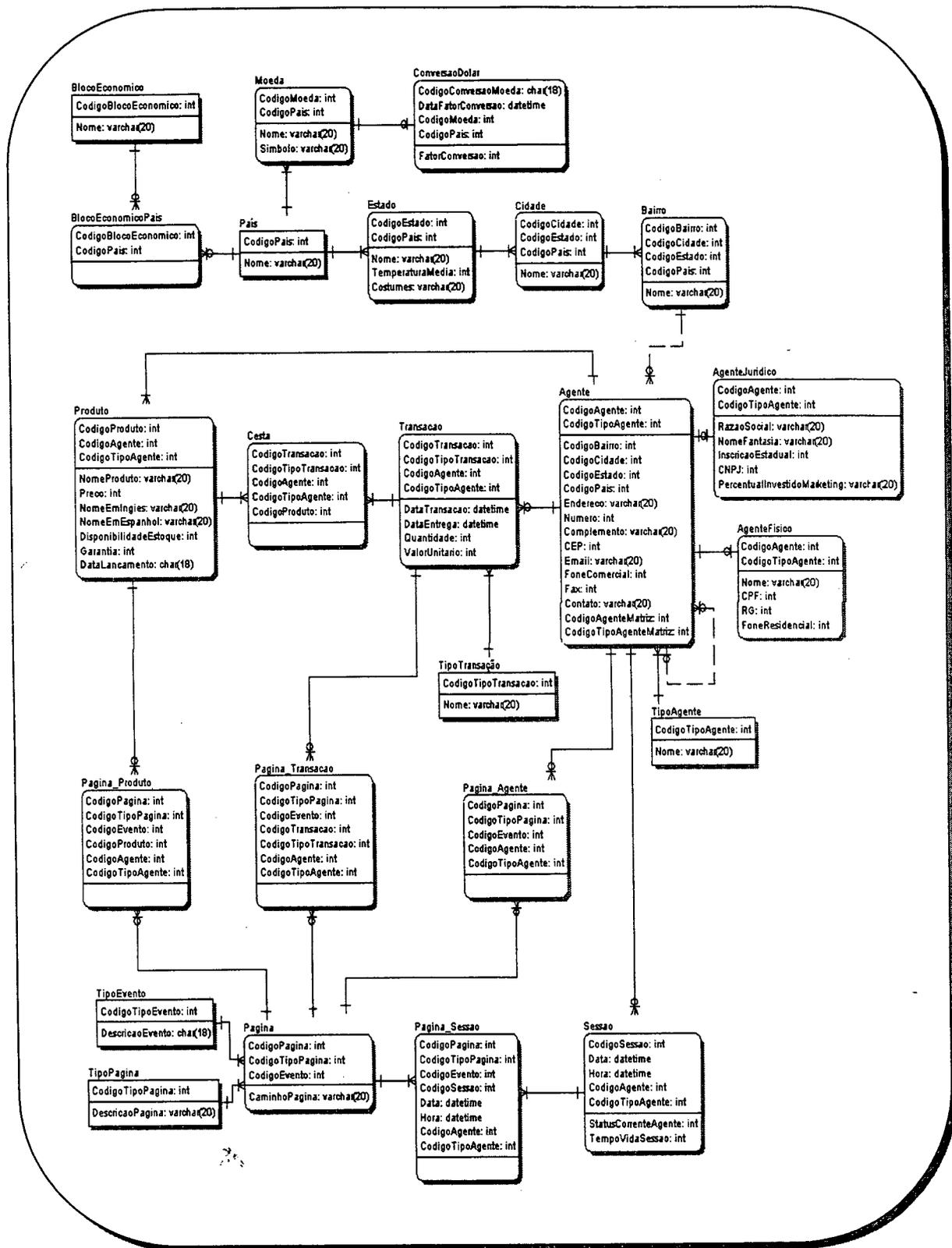


Fig. 5.6 – Modelo de dados normalizado.

A entidade *Sessao* já está normalizada.

A entidade *Pagina* possui dois atributos multivalorados, o atributo TipoPagina e o atributo TipoEvento. Foram criadas duas novas entidades, portanto, uma com o nome de *TipoPagina* com os atributos CodigoTipoPagina e DescricaoPagina, e outra com o nome de *TipoEvento* com os atributos CodigoTipoEvento e DescricaoEvento.

## 5.4. Conclusão

Ao se aplicar os dois primeiros níveis da Engenharia da Informação na organização, estabeleceu-se as metas e os objetivos do negócio de tal forma que as informações na base de dados ficaram organizadas, legíveis, documentadas, melhor identificadas, confiáveis, de modo a facilitar as próximas etapas da criação do *Data Warehouse*. As redundâncias dos dados foram eliminadas por intermédio da normalização do modelo de dados.

O modelo de dados relacional normalizado obtido pela aplicação da Engenharia da Informação na organização está implementado no banco de dados *SQL Server 2000*.

O próximo capítulo apresenta a modelagem baseada no esquema estrela e sua implementação por meio do Modelo Dimensional no *Analysis Services*, modelo este obtido pela aplicação da Engenharia da Informação na organização.

## 6. Modelagem e implementação do Modelo Dimensional

*“A necessidade do Data Warehouse surgiu do fato de que as organizações mantêm suas informações armazenadas em tantos lugares, que é difícil reportar todas as informações disponíveis”. (SPENIK, 2001).*

O capítulo anterior identificou e tratou das necessidades de informações que a organização necessita. Neste capítulo, é desenvolvido o Modelo Dimensional normalizado pelo esquema estrela por intermédio das necessidades de informação identificadas pela utilização da Engenharia da Informação na organização. A modelagem dimensional foi desenvolvida de acordo com (BARBIEIRI, 2001).

Segundo (SPENIK, 2001), o modelo de relacionamento de entidade tradicional utiliza uma abordagem normalizada para o projeto de banco de dados. A normalização remove a redundância do esquema relacional para minimizar o armazenamento dos dados. O *Data Warehouse* não está tão preocupado em economizar espaço, mas tem como prioridade fornecer simplicidade da perspectiva do usuário. Uma pequena quantidade de redundância é geralmente aceitável. Levando isso em consideração, a modelagem dimensional é a abordagem mais apropriada para o projeto de *Data Warehouse*. Esta abordagem envolve projetar o modelo de dados de tal forma que se consiga separar o negócio em eventos lógicos ou fatos (tabela central com o nome de *Transação*) e um conjunto de dimensões correspondentes (tabelas *Agente Físico e Jurídico, Produto, Bloco Econômico, País, Estado, Cidade, Bairro*, etc). Esse conceito é observado na Figura 6.1.

Este capítulo está organizado como segue. A seção 6.1 apresenta conceitos gerais sobre modelagem dimensional. A seção 6.2 mostra a comparação entre o modelo relacional e o Modelo Dimensional. A seção 6.3 apresenta a modelagem dimensional. A seção 6.4 apresenta a implementação do Modelo Dimensional. A seção 6.5 apresenta a visualização dos dados. E finalmente, a seção 6.6 apresenta a conclusão.

### 6.1. Modelagem Dimensional

De acordo com (BARBIERI, 2001) e (LI, 2000), a estrutura dimensional modifica a ordem de distribuição dos registros nas tabelas, permitindo uma formatação mais voltada para os dados das entradas, as dimensões, e menos para os dados granulares, os

fatos. Em uma estrutura dimensional os dados estarão em forma semi-estelar, onde várias tabelas de dimensão estarão se relacionando com poucas tabelas de fato, criando um modelo mais simples, legível e objetivo.

A tabela de Fato central normalmente consiste em eventos de negócio que podem ser registrados ao longo do tempo, como, por exemplo, transações bancárias, vendas, pedidos, visitas ao *site* da *web*, entre outros. As informações armazenadas na tabela de fato são normalmente estáticas, pois são informações históricas (SPENIK, 2001).

As tabelas de dimensão são compostas principalmente por informações textuais referentes aos registros da tabela de fato, como nomes de cliente, descrições de produtos, fornecedores e vendedores. As informações contidas nessas tabelas não são estáticas, pois, esses registros podem ser atualizados, como, por exemplo, o endereço do cliente talvez seja modificado no sistema fonte (SPENIK, 2001).

## 6.2. Comparação entre o modelo relacional e o Modelo Dimensional

A seguir, a tabela 6.1, sintetiza as diferenças entre os modelos relacionais e os modelos dimensionais.

Tabela 6.1 – Modelo Dimensional versus Modelo Relacional.

Fonte: (BARBIERI, 2001)

<b>Modelo Dimensional</b>	<b>Modelo Relacional – E/R</b>
Modelo mais fácil e legível.	Modelo mais complexo.
Tabela fato e tabelas dimensão.	Tabelas representam dados e relacionamentos.
Tabelas fato são normalizadas e, opcionalmente, as tabelas de dimensão.	Todas as tabelas são normalizadas.
Tabelas dimensão são os pontos de entrada de dados.	Todas as tabelas são entradas de dados.
Modelo mais facilmente <i>joined</i> .	Maior dificuldade de <i>join</i> pelo grande número de tabelas.
Compreensão mais fácil do modelo por usuários não especializados.	Grande dificuldade de compreensão pelo usuário não especializado.

O modelo relacional é usado para sistemas operacionais de transação cuja atualização no banco de dados é *on-line*, não sendo tão complexo e grande. No caso de *Data Warehouse* usa-se o Modelo Dimensional porque é um banco muito grande e complexo.

## 6.3. Modelagem do Modelo Dimensional

### 6.3.1. Definição da área de negócio

De acordo com (BARBIERI, 2001),

um projeto de *Data Warehouse* inicia-se igual aos outros projetos de Bancos de Dados do ambiente transacional. Uma área de negócio é escolhida, baseada nas prioridades da organização. Escolhida a área de negócio, parte-se para definir aqueles que serão os processos-alvo do projeto do *Data Warehouse*. Os processos que compõem o modelo deverão ser analisados por uma abordagem que permita o seu pleno entendimento, com o detalhamento de entidades de dados, relacionamentos, objetos, eventos, etc.

Esta etapa é muito favorecida com a aplicação da Engenharia da Informação, pois, a área de negócio **Vendas** já foi identificada, com seus processos alvos **Transação, Produto, Agente, Dados geográficos e Dados webográficos (Clickstream<sup>6</sup>)** sendo que os processos que compõem o modelo relacional foram analisados com a aplicação da Engenharia da Informação, levando a um maior esclarecimento, detalhamento e entendimento das entidades de dados e relacionamentos envolvidos.

#### a. Definição da granularidade

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

A granularidade do presente modelo é definida como: **Dados Geográficos** (compostos pelo Bloco Econômico, Unidade Federativa, País, Estado, Cidade, Bairro),

---

<sup>6</sup>*Clickstream* é um registro de todos “cliques” feitos por qualquer visitante a um *Website*.

**Agente, Produto, Dados Webográficos** (composto pela Sessão, Página, Eventos realizados na Página e Tipo de Página).

### 6.3.2. Definição das tabelas de dimensão

As tabelas de dimensão já foram definidas porque a escolha dos níveis de granularidade, realizada no passo anterior, depende delas. A Figura 6.1 mostra o Modelo Dimensional em esquema estrela com as tabelas de dimensão definidas.

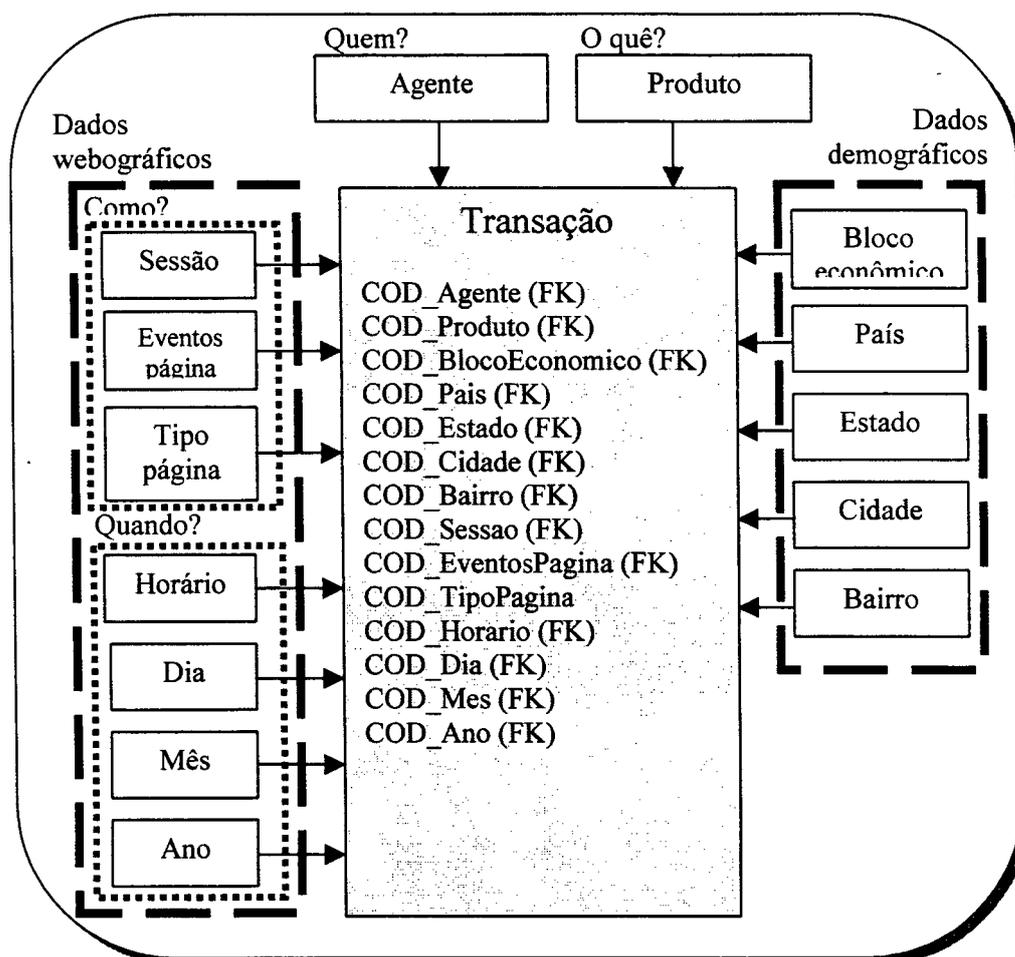


Fig. 6.1 – Definição das tabelas dimensões.

### 6.3.3. Normalização das tabelas de dimensão

Existem duas correntes diferentes com relação aos aspectos de normalização das tabelas dimensão (BARBIERI, 2001):

- *Star Schema* – Esquema Estrela: abordagem que recomenda a não normalização das tabelas de dimensão;

- *Snowflake Schema* – Esquema de Flocos de Neve: abordagem que recomenda a normalização das tabelas de dimensão.

Seguindo a orientação de (BARBIERI, 2001), faremos emprego do esquema estrela para a modelagem dimensional.

O emprego do esquema estrela (*Star Schema*) para a normalização do Modelo Dimensional é extremamente recomendável pelos aspectos de ganhos de performance, quando comparado com o esquema de flocos de neve (*snowflake*). A redundância contida na normalização por meio do esquema estrela será muito compensada pela redução de comandos de junção (*join*) que seriam necessários para recompor a informação desejada (BARBIERI, 2001).

#### 6.3.4. Relacionamentos de atributos das tabelas de dimensão

Segundo (BARBIERI, 2001), podemos classificar os relacionamentos entre as tabelas de dimensão da seguinte forma:

- as tabelas de dimensão de um modelo não possuem relacionamentos entre si, dessa forma são independentes;
- os atributos dentro da mesma tabela de dimensão possuem relacionamento hierárquico, isto é, 1:N;
- os atributos de uma dimensão possuem relacionamentos M:N. Por exemplo, imagine um Modelo Dimensional para controlar a arrecadação de direitos autorais, onde teríamos as dimensões **compositor**, **música** e **dia**. A dimensão **música** tem relacionamento M:N com **compositor**, ou seja, um compositor possui diversas músicas e uma música pode ter vários compositores. Essas dimensões poderão produzir uma nova tabela de relacionamento onde se teria o percentual de cada um nesta música.

#### 6.3.5. Definição dos atributos das tabelas de fato

Quando foi estabelecida a granularidade, as tabelas de fato foram parcialmente definidas. Na definição da granularidade foram definidas as chaves das dimensões. Falta agora definir as métricas que se deseja registrar para manter os processos de gerência. De acordo com (BARBIERI, 2001), são chamadas de métricas, pois normalmente são

somadas e trabalhadas nas diversas dimensões e geralmente estão relacionadas com valores, como, quantidade\_vendida, valor\_vendido, valor\_de\_custo, etc. São elas:

- **aditivas:** quando os valores somados ou com qualquer outro tratamento matemático possuem sentido em todas as dimensões, como por exemplo, valor\_vendido (BARBIERI, 2001);
- **semi-Aditivas:** quando o valor tiver sentido somente em uma dimensão, como por exemplo, quantidade-vendida. Esse valor teria sentido se acumulado na dimensão produto, mas na dimensão loja seria um valor de pouca utilidade. Não existe razão para ter uma somatória da quantidade vendida de todos os produtos da loja (BARBIERI, 2001);
- **não-aditivas:** quando não puder ser somado em qualquer dimensão, ou resultar em um valor sem nenhum sentido válido. Por exemplo, valor de percentagem de lucro  $((\text{valor\_vendido\_custo})/\text{valor\_vendido})$ , neste caso, não se faz necessário armazenar esse valor, pois ele pode ser obtido por intermédio de uma simples operação matemática entre duas variáveis (BARBIERI, 2001).

## 6.4. Implementação do Modelo Dimensional

### 6.4.1. Processamento Analítico On-line - OLAP

O Processamento Analítico On-line é uma tecnologia centralizada na análise dos dados. Os aplicativos de OLAP permitem aos usuários selecionar, visualizar e analisar dados transacionais de diversas fontes, permitindo as organizações extraírem informações adicionais de sistemas de Processamento Transacional On-line – OLTP tradicionais e de *Data Warehouses* (SPENIK, 2001).

As ferramentas OLAP são as aplicações que os usuários finais têm acesso para extraírem informações das bases de dados com as quais geram relatórios capazes de responder às questões gerenciais. Essas ferramentas surgiram com os sistemas de apoio a decisão para realizarem a extração e análise dos dados contidos nos *Data Warehouses* e *Data Marts*<sup>7</sup> (CIELO, 2001).

---

<sup>7</sup> O termo *Data Marts* significa depósito de dados que atende áreas específicas da organização e são voltados para o processo decisório gerencial (BARBIERI, 2001).

### 6.4.2. O Microsoft Analysis Manager

A aparência e o comportamento dessa ferramenta se tornam familiar muito rapidamente, segundo (SPENIK, 2001). Como mostrado na Figura 6.2 o *Microsoft Analysis Manager* possui uma representação hierárquica, em forma de árvore, no painel esquerdo, de todos os servidores e seus componentes. Pode-se examinar a maior parte da sua funcionalidade destacando um item na visualização da árvore e clicando com o botão direito do mouse sobre esse servidor ou componente.

O primeiro passo a ser realizado ao utilizar o *Analysis Services* é criar o banco de dados OLAP. Essa descrição passo a passo sobre a implementação será sobre o Modelo Dimensional definido na Figura 6.1.

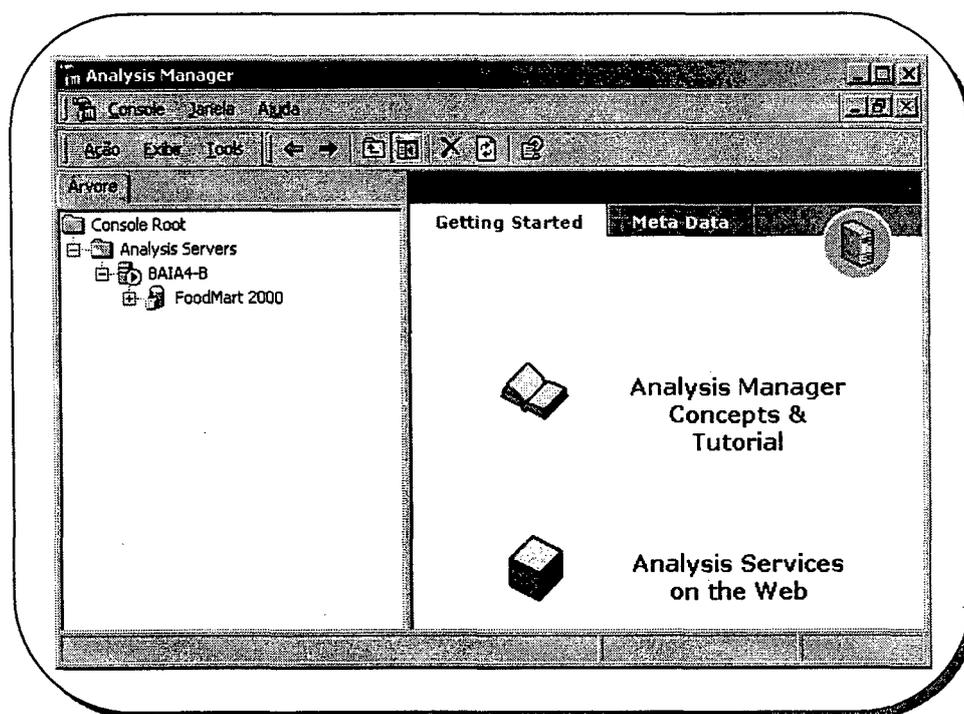


Fig. 6.2 – Interface do *Microsoft Analysis Manager*.

Para criar o banco de dados OLAP, deve-se destacar o ícone do servidor (BAIA4-B). Clique o botão direito do mouse sobre este servidor para exibir o menu de atalho. Escolha a opção *New Database* nesse menu de atalho. Vê-se a caixa de diálogo mostrada na Figura 6.3.

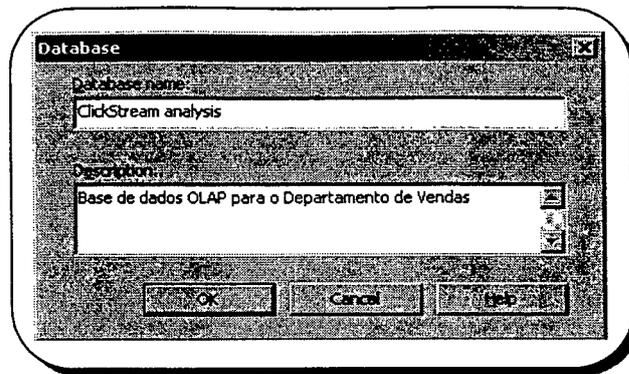


Fig. 6.3 – Caixa de diálogo *New Database*.

#### 6.4.4. Origem de dados

Para criar a origem de dados para o banco de dados de Processamento Analítico On-line - OLAP, no *Analysis Server Manager*, expanda o banco de dados criado anteriormente, e destaque a pasta *DataSource*. Clique com o botão direito do mouse sobre esta pasta (*DataSource*) para exibir o menu de atalho. Após a execução desses passos, vê-se a caixa de diálogo *Data Link Properties* mostrada na Figura 6.4.

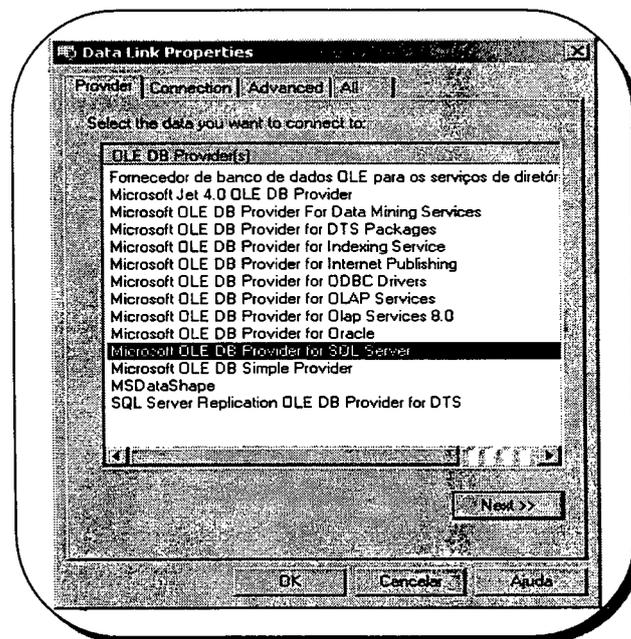


Fig. 6.4 – Caixa de diálogo *Data Link Properties*.

Para a implementação do Modelo Dimensional iremos utilizar o provedor *SQL Server*, pois a base de dados relacional desse trabalho está implementada nesse provedor.

### 6.4.5. Dimensões

Depois que o banco de dados OLAP foi criado, o próximo passo é definir as dimensões do Modelo Dimensional. Para criar uma nova dimensão, deve-se expandir o banco de dados criado anteriormente, e destacar a pasta *Shared Dimensions*. Exibir o menu de atalho, clicando com o botão direito do mouse sobre esta pasta. Escolher a opção *New Dimension* no menu de atalho. Pode-se escolher em utilizar um assistente ou o editor. Iremos utilizar o assistente. Após esses passos, deve-se clicar no botão de *next* na tela de introdução, então observa-se a caixa de diálogo mostrada na Figura 6.5.

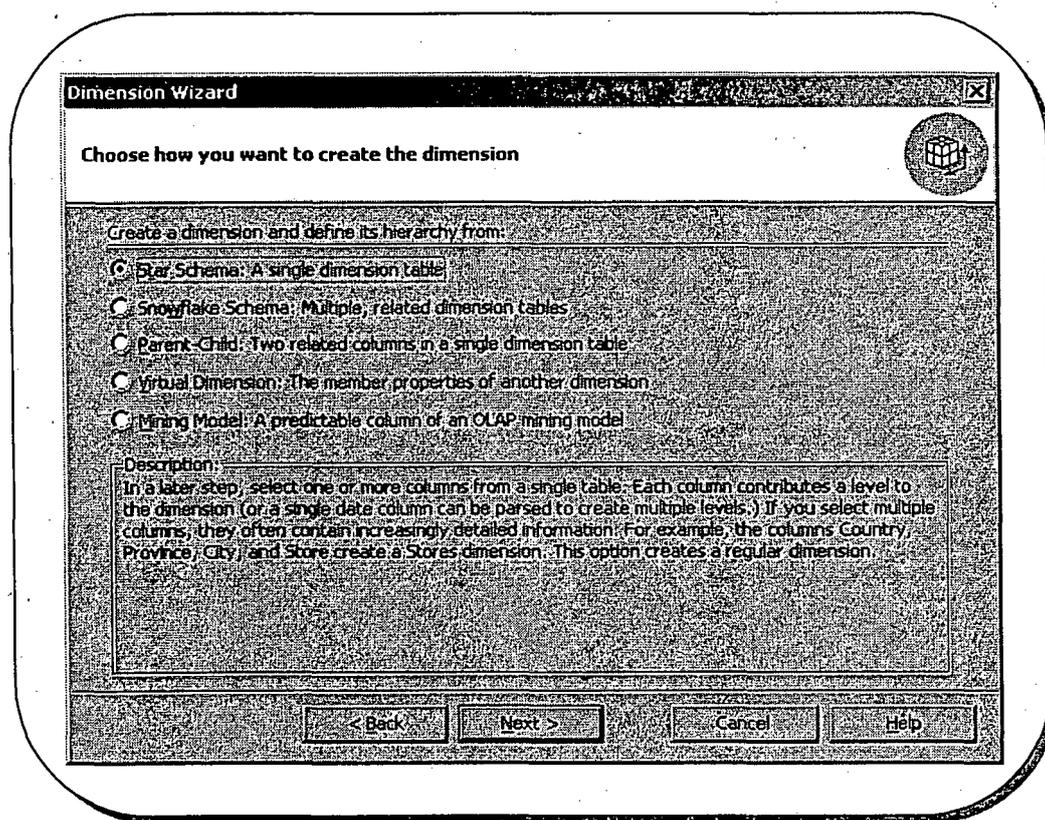


Fig. 6.5 – O *Dimension Wizard*.

O assistente solicitará o tipo de dimensão que se deseja criar. A escolha do tipo de dimensão deve ser baseada nas tabelas subjacentes de dados originais para sua dimensão. Foi escolhida a opção *Star Schema: single dimension table*. Essa é a dimensão básica que utiliza uma única tabela subjacente de banco de dados para definir suas características.

Depois de escolhido o tipo de dimensão que se deseja criar, o próximo passo na criação da dimensão é escolher a tabela de dimensão que fornecerá os dados originais para a dimensão. Para selecionar uma tabela de dimensão, clique no botão *next* na caixa de diálogo *Dimension Wizard*, como mostrada na Figura 6.6.

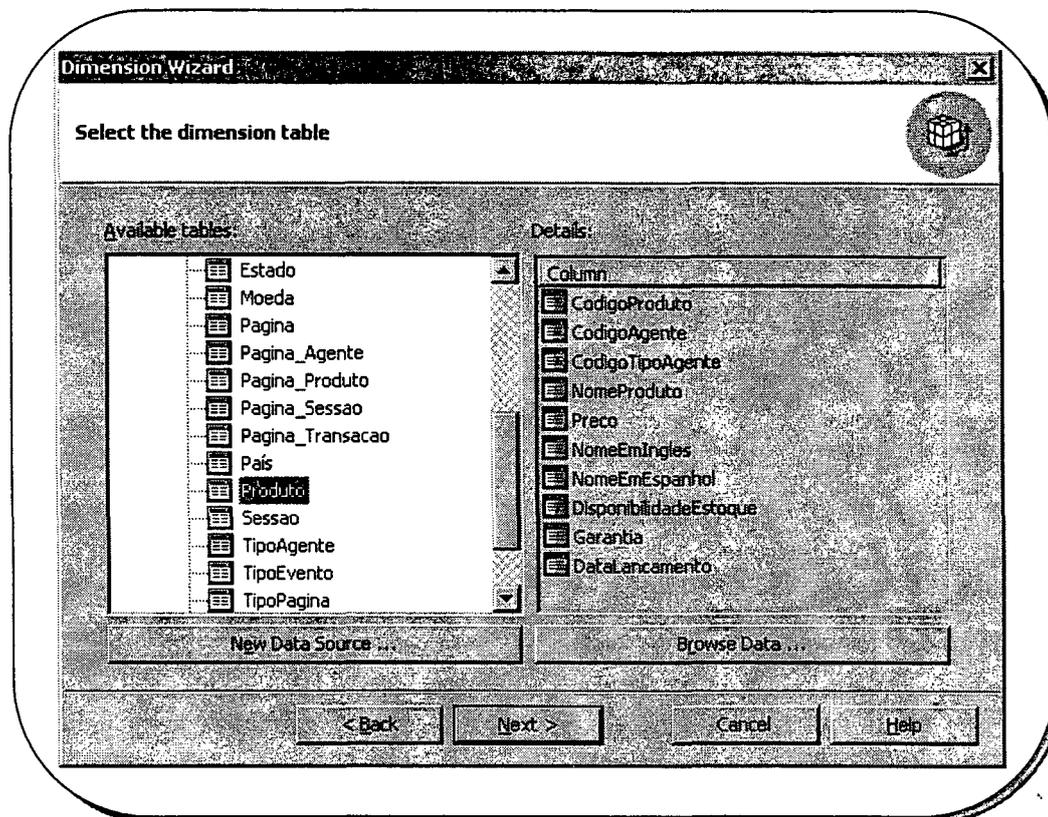


Fig. 6.6 – Caixa de diálogo *Select Dimension Table*.

Expanda a origem dos dados desejada para exibir a lista de tabelas disponíveis. Deve-se escolher a tabela mais adequada para a dimensão. Para o Modelo Dimensional proposto criaremos a dimensão **Produto**, então escolhemos a tabela **Produto** como fornecedora dos dados originais. Realizada a escolha, deve-se clicar no botão *next*.

Se a tabela de dimensão escolhida possuir um campo de data, o *Dimension Wizard* perguntará se deseja criar uma dimensão padrão ou uma dimensão de tempo. O *Analysis Services* pode criar automaticamente dimensões de tempo com base nas colunas de data utilizando as partes numéricas da data (SPENIK, 2001).

### 6.4.5.1. Níveis e membros de dimensão

Depois que se selecionou uma tabela de dimensão, o próximo passo é definir os níveis e membros dentro da dimensão. A caixa de diálogo *Dimension Wizard Select Levels* mostrada na Figura 6.7 é onde se deve definir os níveis dentro de uma dimensão. Deve-se escolher as colunas da tabela de dimensão que correspondam aos níveis que se deseja criar. Pode-se utilizar os botões *Move Up* e *Move Down* para organizá-los. Após a escolha dos níveis, deve-se clicar no botão *next*.

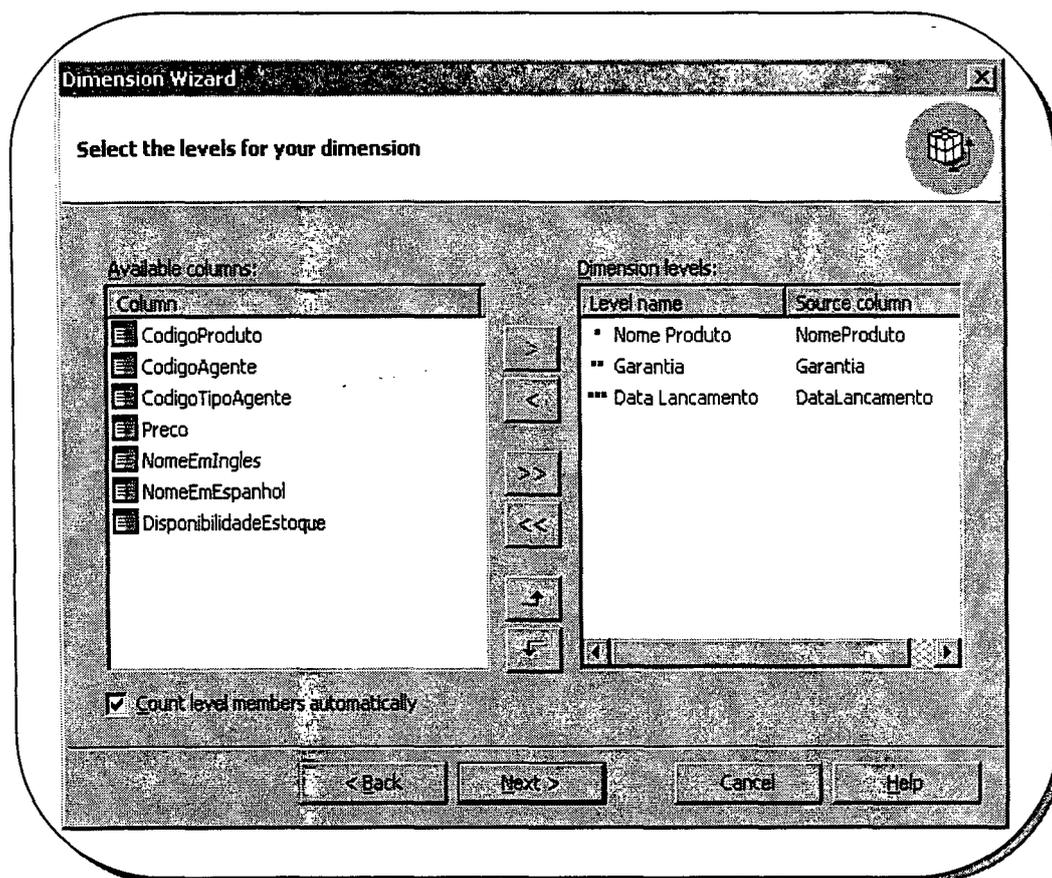


Fig. 6.7 – Caixa de diálogo *Dimension Wizard Select Levels*.

Depois que se definiu os níveis e os membros de dimensões, deve-se clicar no botão *next* nas duas próximas telas do assistente, pois essas telas fornecem algumas opções avançadas para criar as dimensões, sendo que a opção padrão é suficiente. Deve-se fornecer um nome para a dimensão, como mostrado na Figura 6.8, esta dimensão receberá o nome de *Produto*.

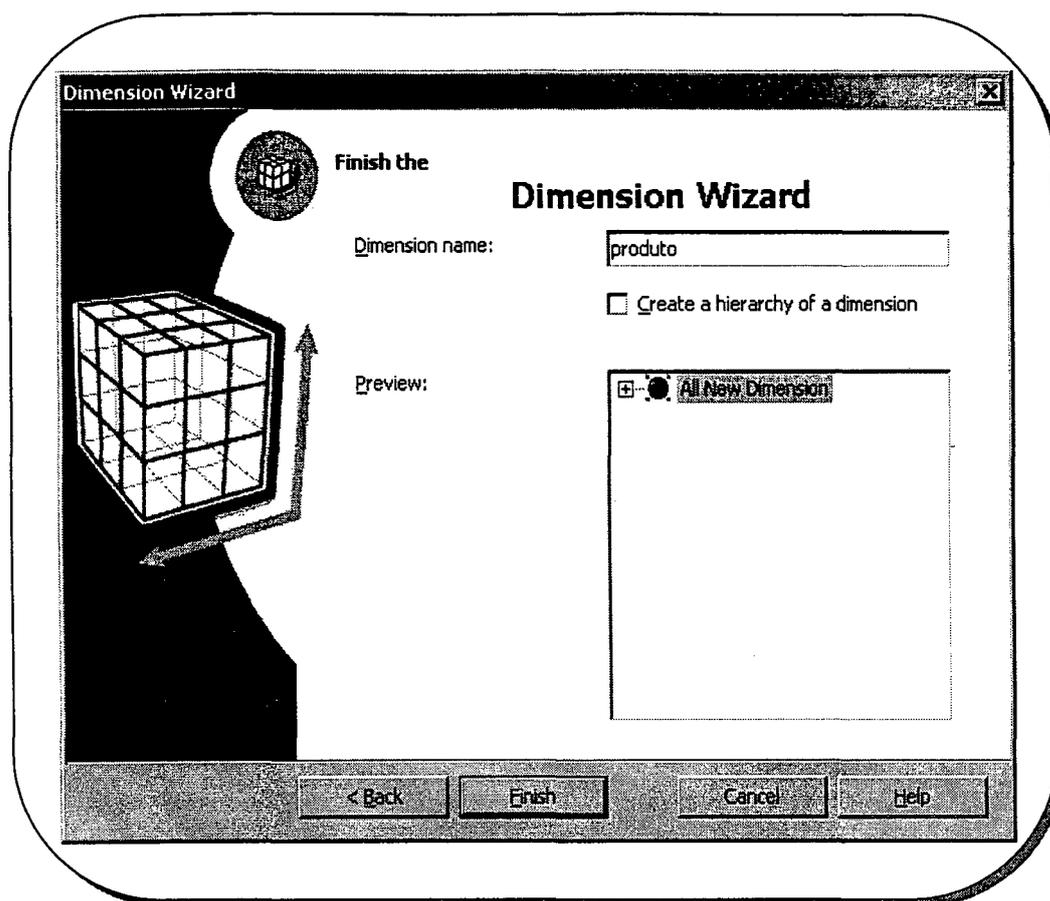


Fig. 6.8 – Caixa de diálogo *Dimension Wizard Finish*.

Na caixa de diálogo *Dimension Wizard Finish* deve-se clicar no botão *Finish* para finalizar a definição da dimensão, com isso retornará a caixa de diálogo *OLAP Manager*. Para cada nova dimensão que se deseja criar, deve-se repetir os passos descritos anteriormente.

#### 6.4.6. Cubo

Depois que se definiu as dimensões, o próximo passo é construir o cubo. Para criar um cubo, expanda o banco de dados criado na seção 6.4.2 e destaque a pasta *Cube*. Clique com o botão direito do mouse sobre esta pasta para exibir o menu de atalho, e escolha *New Cube*. Pode-se escolher utilizar o assistente ou o editor, utilizaremos novamente o assistente. Deve-se clicar no botão *next* na tela de introdução e observa-se a caixa de diálogo *Cube Wizard Select Fact Table* mostrada na Figura 6.9.

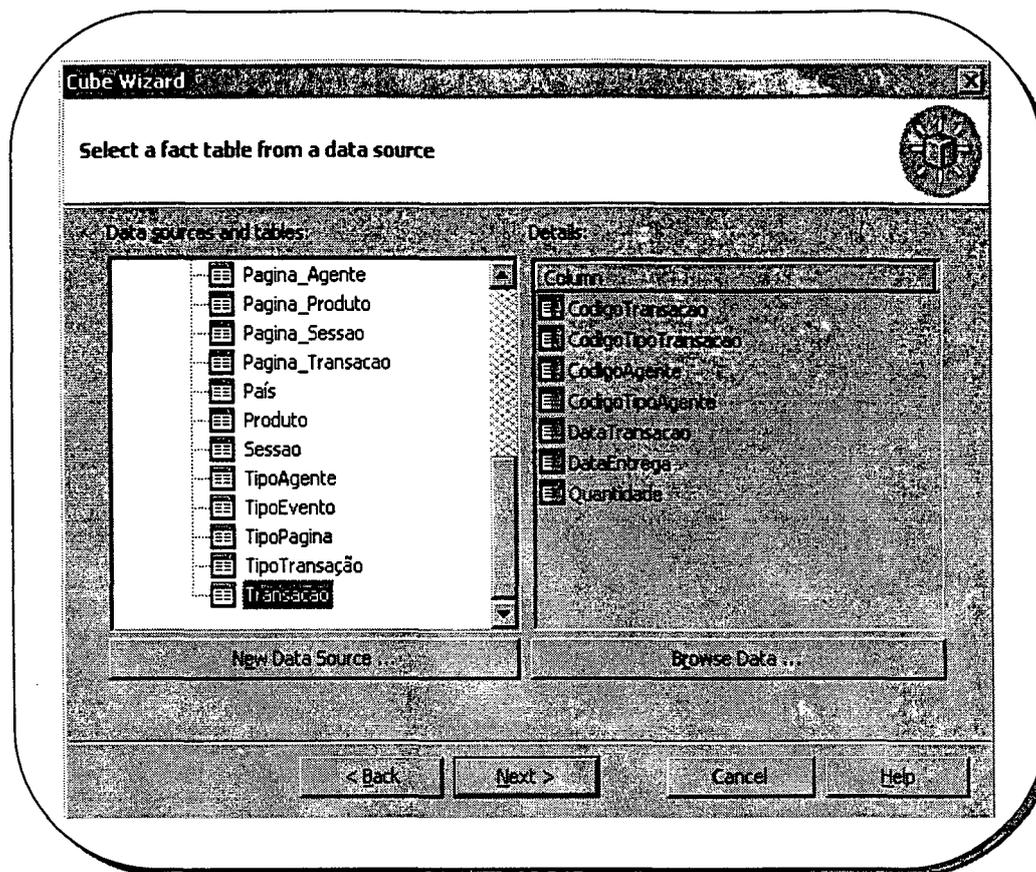


Fig. 6.9 – Caixa de diálogo *Cube Wizard Select Fact Table*.

Na caixa de diálogo *Cube Wizard Define Measures* mostrada na Figura 6.10, deve-se clicar no botão *next* para definir quais os dados dentro da tabela de fato se desejam analisar. Estes dados são conhecidos como as medidas ou métricas do cubo, descritos em maiores detalhes na seção 6.3.5. Nesta caixa de diálogo deve-se escolher as colunas que se desejam analisar e clicar no botão *next*. Pode-se adicionar quantas medidas ao cubo se desejar, desde que as colunas estejam presentes na tabela de fato.

Para o Modelo Dimensional proposto se deseja analisar a quantidade de vendas. Para fazer isso, escolhemos a coluna *quantidade* como medida do cubo.

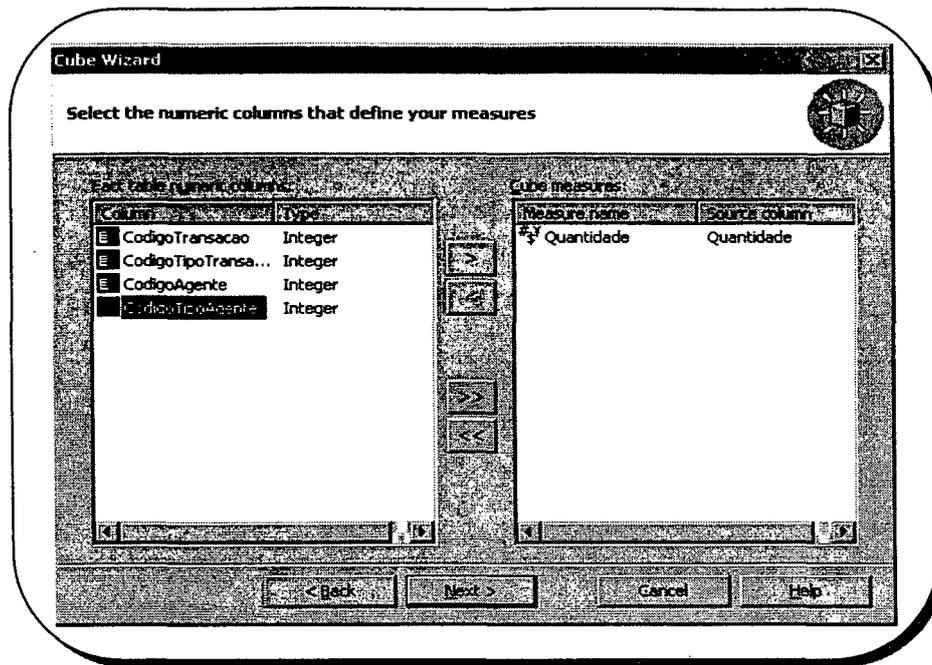


Fig. 6.10 – Caixa de diálogo *Cube Wizard Define Measures*.

Na caixa de diálogo *Cube Wizard Select Dimensions* mostrada na Figura 6.11, escolha as dimensões desejadas e clique no botão *next*.

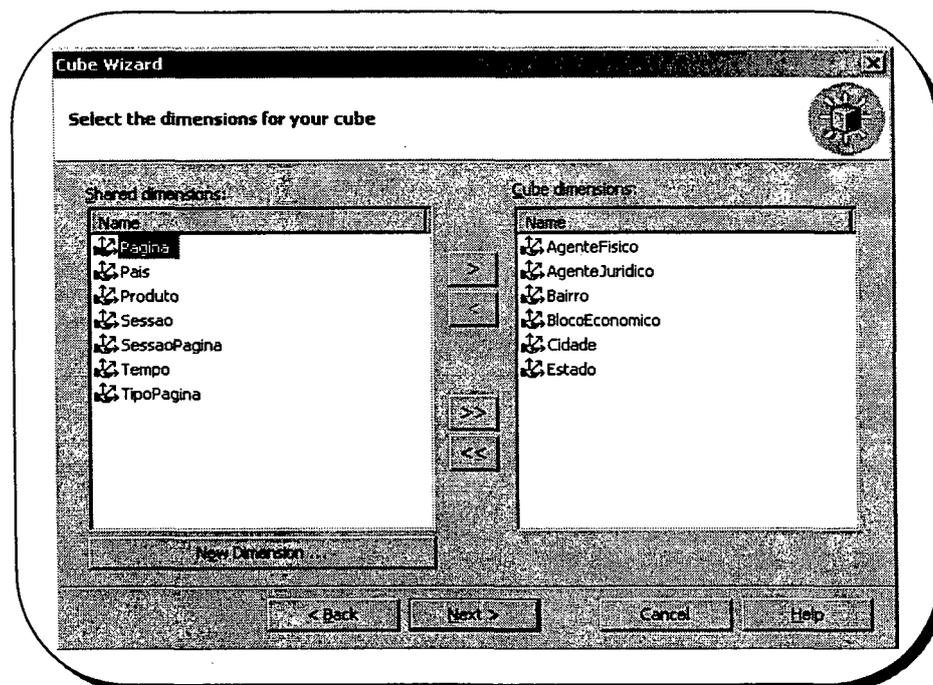


Fig. 6.11 – Caixa de diálogo *Cube Wizard Select Dimensions*.

### 6.4.7. Inclusão das dimensões no cubo

Ao se utilizar o assistente pode-se incluir qualquer dimensão que se definiu anteriormente no cubo. Para o modelo proposto utilizaremos as seguintes dimensões: *AgenteFisico*, *AgenteJuridico*, *Bairro*, *BlocoEconomico*, *Cidade*, *Estado*, *Pais*, *Pagina*, *Produto*, *Sessão*, *SessaoPagina*, *Tempo*, *TipoPagina*, sendo que essas dimensões adicionam significado à medida criada na seção 6.4.6.

Depois de definir as dimensões do cubo e clicar no botão *next*, deve-se finalizar o *Cube Wizard* clicando no botão *finish* na caixa de diálogo *Cube Wizard Finish* como mostra a Figura 6.12.

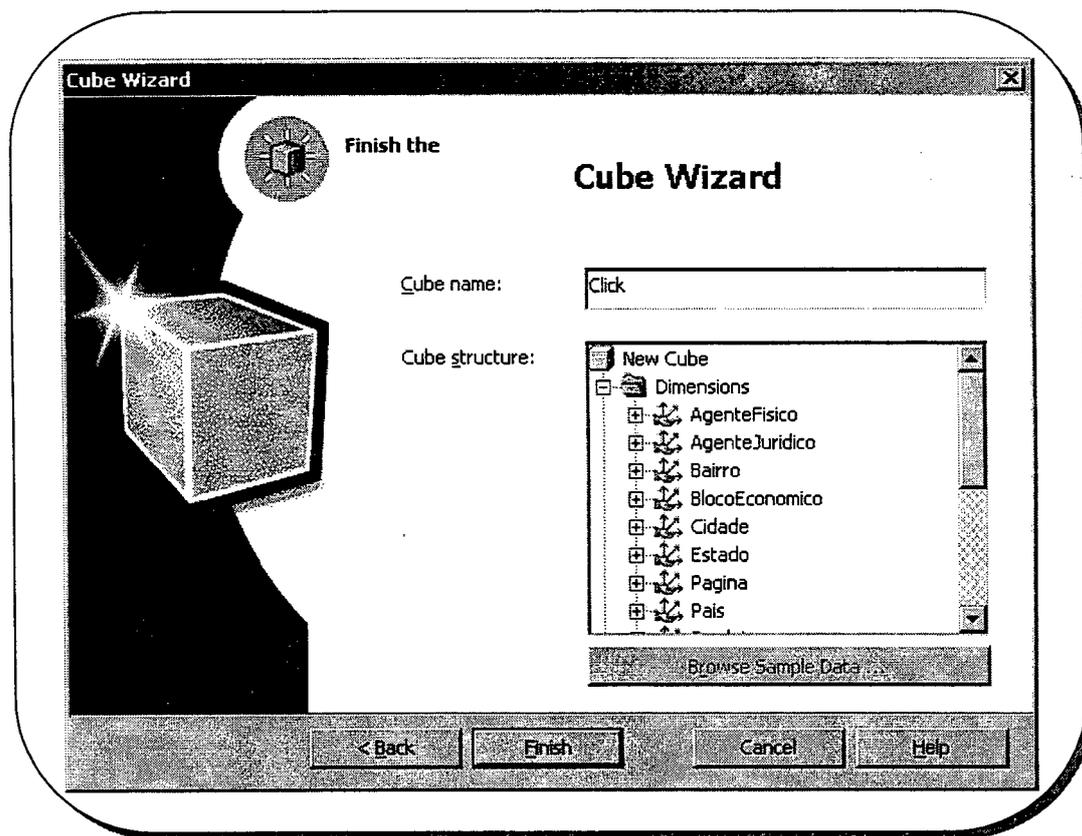


Fig. 6.12 – Caixa de diálogo *Cube Wizard Finish*.

Pode-se repetir os passos anteriores para criar quantos cubos desejar no banco de dados OLAP. A Figura 6.13 ilustra o Modelo Dimensional final com a tabela de fato central e suas dimensões.

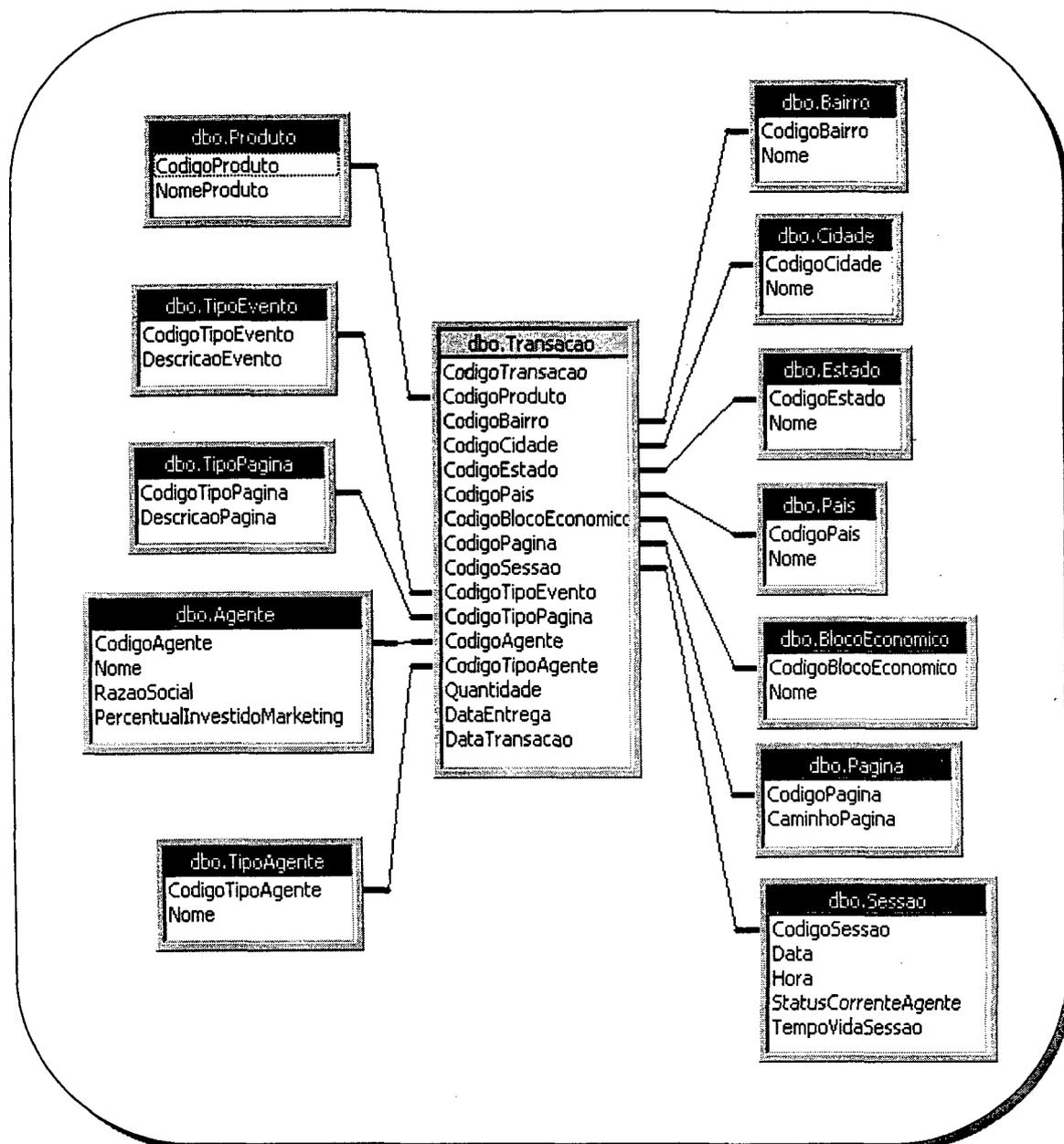


Fig. 6.13 – Modelo Dimensional final.

## 6.5. Visualização dos dados

A visualização dos dados contidos na tabela de fato e tabelas de dimensão pode ser realizada por meio do aplicativo chamado *FrontPage*. Assim, utilizou-se esse aplicativo para a apresentação dos dados ao usuário final. Para realizar a visualização dos dados, primeiramente deve-se inserir o componente chamado *Office PivotTable*, encontrado no menu inserir, componente.

Então, deve ser realizada uma conexão com o banco de dados. Para isso, clica-se no objeto Caixa de Ferramentas de Propriedades, localizado no objeto inserido. A Figura 6.14 ilustra a Caixa de Ferramentas.

Para realizar a conexão com o banco de dados, seleciona-se a opção conexão e editor de conexão na caixa de ferramentas. Na guia Provedor, escolher o provedor correspondente, para nosso exemplo, *Provider for OLAP Services* deve ser selecionado, pois estamos construindo essas visualizações pela base de dados *OLAP*.

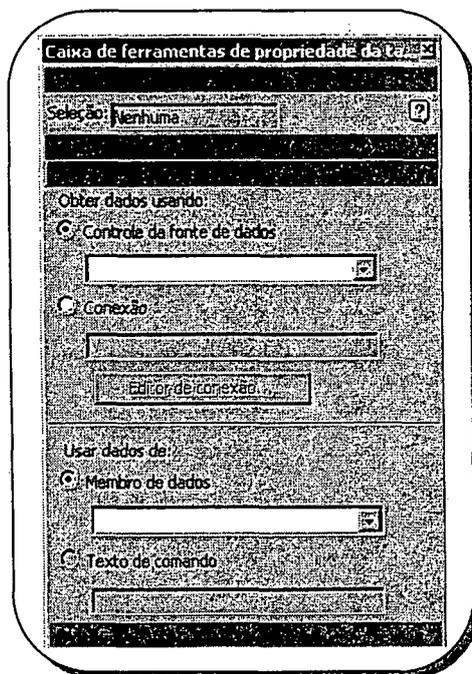


Fig. 6.14 – Caixa de ferramentas.

Na guia *Connection* especificar o local da fonte de dados, nome do usuário, senha e o cubo de dados que você quer visualizar e OK, como ilustrado na Figura 6.15.

Na caixa de ferramentas especificar o Membro de Dados, que no presente trabalho seria *Click*, completando a conexão com o banco.

No componente *Office PivotTable* clicar no objeto lista de campos para apresentar todas as tabelas de dimensão a tabelas de fato, correspondente ao cubo, como ilustrado na Figura 6.16. Deve-se arrastar os campos desejados das tabelas de dimensão ou de fato aos campos correspondentes no componente *Office PivotTable*.

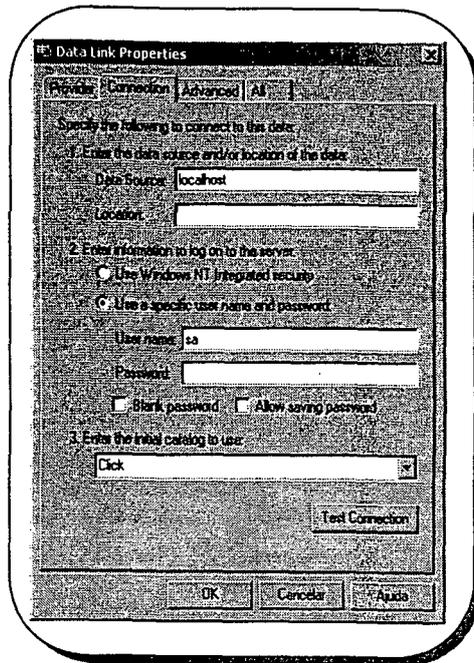


Fig. 6.15 – Caixa de diálogo *Data Link Properties*.

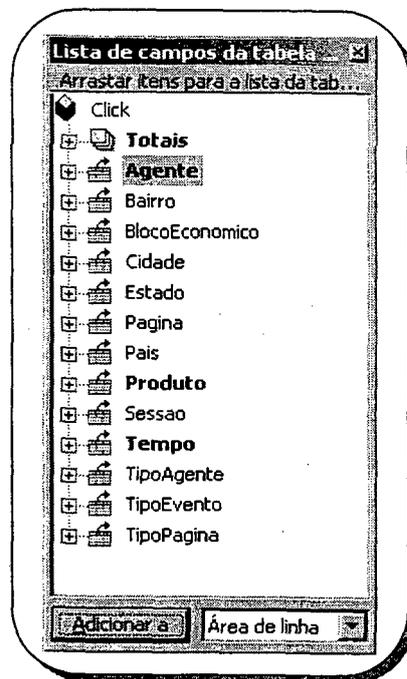


Fig. 6.16 – Lista de campos da tabela.

Finalmente, para visualizar os dados em forma de gráficos, deve-se inserir um componente *Office Chart*, encontrado no menu Inserir, Componente.

## 6.6. Conclusão

O sucesso da implantação de um *Data Warehouse* depende de um projeto muito bem elaborado. Os itens que farão parte devem ser bem delimitados e nessa questão a Engenharia da Informação desempenhou um papel muito importante.

Um dos pontos fundamentais na elaboração e implantação do *Data Warehouse* são os recursos humanos. Todo suporte computacional elaborado e implantado servirá para dar suporte aos usuários, ajudando-os a tomar decisões importantes para a organização. De nada vale realizar um grande investimento apenas no ambiente computacional e não investir nas pessoas que farão este ambiente ganhar utilidade.

As ferramentas utilizadas na implementação e validação do Modelo Dimensional obtido por intermédio da aplicação da Engenharia da Informação na organização foram: *SQL Server 2000* para a implantação do modelo relacional e o Modelo Dimensional, o *OLAP Services 2000* para a implementação das tabelas de dimensão e fato e o *Front Page* para a visualização dos dados. Todas foram de grande valia e sua funcionalidade nada deixa a desejar.

## 7. Conclusão e perspectivas futuras

A *Web* é uma forma de mercado diferente do mercado tradicional. Existem algumas questões que merecem atenção, como a distribuição geográfica e o anonimato dos usuários. Além de questões relacionadas à *Web* também temos questões referentes ao nível de exigência dos clientes e à constante evolução dos *Websites*, por isso análise de *clickstream* é importante em um mercado onde a interação *Data Warehouse – Clickstream* com os clientes é na maioria das vezes completamente virtual (VOELCKER, 2001).

Haja vista a importância do comércio eletrônico e a necessidade das empresas abrirem-se ao mercado é que se desenvolveu este trabalho, com o objetivo de fornecer uma ferramenta compatível aos anseios do comércio atual. Algumas constatações feitas são destacadas a seguir:

Com a utilização do Planejamento Estratégico das Informações obteve-se o modelo de dados corporativo. Com a Análise da Área de Negócios o modelo de dados corporativo foi decomposto em atributos, simplificado (para eliminar algumas redundâncias existentes) e normalizado. Resultou um sistema alinhado com os objetivos da organização onde foi desenvolvido o trabalho.

O processo de obtenção desses modelos é um trabalho em equipe em que participam pessoas de várias áreas da organização. Um modelo de dados bem elaborado e de acordo com os objetivos da empresa torna o *Data Warehouse* muito mais consistente e robusto.

A modelagem dimensional é um dos fatores críticos de sucesso em um projeto de *Data Warehouse*. Em um projeto de *Data Warehouse* os dados fundamentalmente importantes são aqueles consolidados nas dimensões específicas (BARBIERI, 2001). Existe duas formas de normalização do Modelo Dimensional, o esquema estrela que recomenda a não normalização das tabelas de dimensão e o esquema em flocos de neve que recomenda a normalização das tabelas de dimensão. No presente trabalho foi utilizado a normalização em esquema estrela, pois o modelo corporativo está de acordo com os objetivos da organização e normalizado.

A principal função da modelagem dimensional desenvolvida é fornecer ao analista de negócio subsídios que o ajudem a identificar e tentar entender o comportamento dos

agentes quando acessam o *Website*. Em momento algum, estamos tomando a posição do desse profissional na organização, nosso propósito é fornecer uma solução que o ajude a tomar suas decisões de uma forma mais clara e precisa.

Um *Data Warehouse* é um recurso de armazenamento centralizado para diferenciar tipos de dados por toda organização. Esses dados são orientados por assunto, sensíveis ao tempo e organizados de tal maneira que oferecem análise simplificada (SPENIK, 2001).

O *Data Warehouse* é um banco de dados extraído do ambiente de produção da empresa, que foram selecionados e depurados, otimizados para processamento de consulta e não para processamento de transações (CAMPOS, 2000).

Ao reunir informações de diferentes bancos de dados e plataformas, o *Data Warehouse* permite que sejam feitas análises bastante eficazes, transformando dados dispersos em informações estratégicas que antes eram inacessíveis ou subaproveitadas (TAURION, 1997).

A abordagem de Bill Inmon se concentrou inicialmente no estilo tradicional de construção de Bancos de Dados que busca uma forte integração entre todos os dados da empresa. Isso seria representado num modelo único, integrado e coeso que mostrou-se rígido e de difícil consecução (BARBIERI, 2001).

A abordagem de Ralph Kimball possui um estilo mais simples e incremental. Diferente da abordagem anterior, a metodologia *Star Schema*, aponta para projetos de *Data Marts* separados que deverão ser integrados na medida da evolução (BARBIERI, 2001).

Atualmente, a convergência das abordagens está no fato de que se inicia a construção do *Data Warehouse* pelos primeiros *Data Marts*. Em uma fase inicial se define a compatibilidade de dimensões dos *Data Marts* e alguns padrões dos dados que são fatores fundamentais para minimizar possíveis rupturas. Os outros *Data Marts*, a seguir, devem ser integrados, mantendo a conformidade entre as dimensões (BARBIERI, 2001).

Desse modo, com a integração da Engenharia da Informação à abordagem de Ralph Kimball para o tratamento das necessidades de informação da organização, base para o desenvolvimento deste trabalho, resultou em um Modelo Dimensional de acordo

com as reais necessidades da organização, obtendo-se do *Data Warehouse* desenvolvido os resultados esperados.

O modelo desenvolvido será aplicado em uma base de dados real que permitirá identificar alguns comportamentos dos agentes como uma possível intenção de desistência do agente, identificar se o anúncio na *Web* está funcionando, se as promoções são rentáveis, ajudar na tomada de decisão sobre quais serviços e produtos oferecer via *Web*, aumentar a eficiência do *Website*, facilitar a navegação no *site*; personalizar o *site* para cada agente, saber quais são as áreas do *site* mais representativas, valorizar cada área do *site* para venda de *banners*, saber as características dos produtos mais vendidos e as característica dos produtos menos vendidos, dentre outras particularidades.

## 7.1. Trabalhos futuros

Como trabalhos futuros poderão ser abordadas várias questões como:

- agregar mais entidades no modelo relacional, levando assim a criação de um modelo relacional mais completo e um Modelo Dimensional que contemple todos ou os principais departamentos da organização;
- a transferência de dados do ambiente operacional para o ambiente multidimensional é um dos maiores problemas, pois, pode-se encontrar diversas fontes para um mesmo dado. É possível desenvolver um mecanismo para transformar e carregar os dados na base de dados multidimensional;
- desenvolvimento de uma ferramenta para a mineração de dados (*data mining*), visto que a base de dados multidimensional já está modelada;
- comparar a questão de performance entre os conceitos de normalização da modelagem dimensional em esquema estrela e esquema flocos de neve;
- traçar um comparativo das ferramentas disponíveis no mercado para o desenvolvimento de *Data Warehouse*.

## 8. Referências Bibliográficas

- (BARBIERI, 2001) BARBIERI, C. **BI-Business Intelligence – Modelagem & Tecnologia**. Rio de Janeiro: Axcel Books do Brasil, 2001.
- (BISPO, 1998) BISPO, C.A.F. **Uma análise da nova geração de Sistemas de Apoio à Decisão**. São Carlos, 1998. Dissertação – Escola de Engenharia de São Carlos, Universidade de São Paulo. Disponível em: <http://cazarini.cpd.eesc.sc.usp.br/Bispo/DI/index.html> Acesso em: 19 abr. 2001.
- (CAMPOS, 2000) CAMPOS, M.L., ROCHA FILHO, A.V. **Data Warehouse**. Disponível em: <http://genesis.nce.ufrj.br/dataware/tutorial/home.html>. Acesso em: 14 out. 2000.
- (CIELO, 2001) CIELO, I. **OLAP**. Disponível em: <http://www.datawarehouse.inf.br/olap.asp>. Acesso em: 14 dez. 2001.
- (CRUZ, 2000) CRUZ, A. da C.; PEROTA, M.L.L.R.; MENDES, M.T.R. **Elaboração de referências (NBR 6023/2000)**. Rio de Janeiro: Interciência, 2000. ISBN 85-7193-043-0.
- (CURTY, 2000) CURTY, M.G.; CRUZ, A. da C. **Apresentação de trabalhos científicos: guia para alunos de cursos de especialização**. Maringá: Dental Press, 2000. ISBN 85-88020-01-7.
- (DATE, 1990) DATE, C.J. **Introdução a sistemas de bancos de dados**. Rio de Janeiro: Campus, 1990.
- (FURLAN, 1994) FURLAN, J.D., IVO, I. da M., AMARAL, F. P. **Sistemas de Informação Executiva – EIS**. São Paulo: McGraw Hill, 1994.
- (GALLAS, 1999) GALLAS, S. **Kimball Vs. Inmon**. DMDirect. setembro 1999. Disponível em: <http://www.dmreview.com/master.cfm?NavID=198&EdID=1400>. Acesso em: 03 maio 2001.
- (HACKATHORN, 1997) HACKATHORN, R. **Data warehousing's credibility crisis**. Byte, agosto 1997. Disponível em: <http://www.byte.com/art/9708/sec4/art1.htm>. Acesso em: 20 abr. 2001.
- (HARRISON, 1998) HARRISON, T.H. **Intranet Data Warehouse**. São Paulo: Berkeley Brasil, 1998.

- (INMON, 1992) INMON, W.H. **Building the Data Warehouse**. John Wiley & Sons Inc., USA, 1992.
- (INMON, 1997) INMON, W. H. **Como construir o Data Warehouse**. Rio de Janeiro: Campus, 1997.
- (KERN, 1994) KERN, V.M. **Bancos de Dados Relacionais: teoria e prática de projeto**. São Paulo: Érica, 1994.
- (KIMBALL, 1995) KIMBALL, R. **Is ER modeling hazardous to DSS?** DBMS Magazine, outubro 1995. Disponível em: <<http://www.Dbmsmag.com/9510d05.html>>. Acesso em: 23 abr. 2001.
- (KIMBALL, 1997) KIMBALL, R. **A dimensional modeling manifesto**. DBMS Magazine, agosto 1997. Disponível em: <<http://www.dbmsmag.com/9708d15.html>>. Acesso em: 23 abr. 2001.
- (KIMBALL, 1998) KIMBALL, R. **Data Warehouse toolkit**. São Paulo: Makron Books, 1998.
- (KIMBALL, 1998a) KIMBALL, R., REEVES, L., ROSS, M. THORNTHWAITE, W. **The Data Warehouse Lifecycle Toolkit: expert methods for designing, developing, and deploying Data Warehouses**. New York: John Wiley & Sons, 1998.
- (LAMBERT, 1996) LAMBERT, B. **Data warehousing fundamentals: what you need to know to succeed**. Data Management Review, março 1996. Disponível em: <<http://www.dmreview.com/master.cfm?NavID=55&EdID=1313>>. Acesso em: 25 maio 2000.
- (LESCA, 1994) LESCA, H., ALMEIDA, F.C.de. **Administração Estratégica da Informação**. Revista de Administração, São Paulo v.29, n.3, p.66-75, jul-set 1994.
- (LI, 2000) LI, R., SALZ, J. **Clickstream Data Warehousing**. ArsDigita Systems Journal. Agosto 2000. Disponível em: <<http://www.arsdigita.com/asj/clickstream/>>. Acesso em: 10 nov. 2001.
- (MACHADO, 1996) MACHADO, F. N. R., ABREU, M. P. de. **Projeto de Banco de Dados: uma visão prática**. São Paulo: Érica, 1996.
- (MARTIN, 1991) MARTIN, J. **Engenharia da Informação**. Rio de Janeiro: Editora Campus, 1991.

Editora Campus, 1991.

- (MARTIN, 1991a) MARTIN, J., McCLURE, C. **Técnicas Estruturadas e CASE**. São Paulo: Makron, McGraw-Hill, 1991.
- (NETO, 1988) NETO, A.F., FURLAN, J.D., HIGA, W. **Engenharia da Informação Metodologia, Técnicas e Ferramentas**. São Paulo: McGraw-Hill, 1988.
- (SERAFIM FILHO, 1999) SERAFIM FILHO, P. **A Engenharia da Informação na tomada de decisão**. Insight Informal, N. 020, 1999. Disponível em: <<http://www.informal.com.br/insight/insight20.html>>. Acesso em: 19/04/2001, 14:30.
- (SILBERSCHATZ, 1999) SILBERSCHATZ, A., KORTH, H.F., SUDARSHAN, S. **Sistema de Banco de Dados**. São Paulo: Makron Books, 1999.
- (SINGH, 2001) SINGH, H. **Data Warehouse**. São Paulo: Makron Books, 2001. ISBN 85.346.1034-7.
- (SPENIK, 2001) SPENIK, M., SLEDGE, O. **Microsoft SQL Server 2000 DBA: guia de sobrevivência**. Rio de Janeiro: Campus, 2001.
- (TANLER, 1998) TANLER, R. **Intranet Data Warehouse**. Rio de Janeiro: Editora Infobook. 1998.
- (TAURION, 1997) TAURION, C. **Data Warehouse: estado de arte e estado de prática**. Developer's Magazine, N.6, p. 10-11, fevereiro 1997.
- (TAURION, 1998) TAURION, C. **Data Warehouse: vale a pena gastar milhões investindo em um?** Developer's Magazine, N.18, p. 10-11, fevereiro 1998.
- (VOELCKER, 2001) VOELCKER, R. A. **Data Webhouse – Clickstream**. Disponível em: <[http://genesis.nce.ufrj.br/dataware/DataWarehouse/trabalhos/Trabs20002/Estudos\\_Avancados/DWClickstream.pdf](http://genesis.nce.ufrj.br/dataware/DataWarehouse/trabalhos/Trabs20002/Estudos_Avancados/DWClickstream.pdf)>. Acesso em: 15 dez. 2001.
- (WELDON, 1997) WELDON, J.L. **Warehouse Cornerstones**. Revista Byte, V.22, n.1, janeiro 1997. Disponível em: <<http://www.byte.com/art/9701/sec7/art2.htm>>. Acesso em: 20 abr. 2000.
- (YOURDON, 1990) YOURDON, E. **Análise Estruturada Moderna**. Rio de Janeiro: Campus, 1990.

## Índice Remissivo

---

### A

abordagens · 34, 37, 83  
 Aditivas · 68  
 ambiente de dados operacionais · 29  
 Análise da Área de Negócio · 18, 19  
 análise das áreas de negócios · 56  
 Analysis Manager · 69  
 Analysis Services · 14, 62, 69, 72  
 armazenamento · 28, 63, 65, 83

---

### B

banco de dados relacionais · 29  
 Blocos da Engenharia da Informação ·  
 22

---

### C

carga · 31, 34  
 carga e acesso · 31  
 CASE · 13, 15, 16, 23, 24  
 classes de dados · 18, 53, 56  
 Clickstream · 65, 82, 86, 87  
 comércio eletrônico · 13, 50, 55, 82  
 Construção · 20  
 cubos · 77

---

### D

dados históricos · 50  
 dados resumidos · 29  
 data warehouse · x, xi, 12, 13, 14, 27,  
 28, 29, 30, 31, 32, 33, 34, 35, 36, 37,  
 38, 39, 40, 41, 42, 43, 44, 45, 46, 47,  
 49, 50, 63, 65, 81, 82, 83, 84  
 Data Warehouse · ii, iii, x, xi, 28, 30,  
 34, 35, 37, 44, 50, 85, 86, 87  
 desafios · 18, 52, 82  
 desempenho · 41  
 diagramas de ação · 19  
 Diagramas de Ação · 18  
 Diagramas de Fluxo de Dados · 18  
 Dicionário · 24

---

### E

enciclopédia · 21, 24, 25, 26  
 engenharia da informação · x, 12, 13,  
 14, 15, 16, 17, 21, 23, 24, 25, 26, 27,  
 49, 50, 52, 53, 62, 65, 81, 87  
 entidades · 21, 53, 55, 56, 57, 58, 59,  
 60, 62, 65, 67, 84  
 esquema de flocos de neve · 67  
 Esquema de Flocos de Neve · 67  
 esquema estrela · x, 14, 32, 48, 62, 66,  
 67, 82, 84  
 Esquema Estrela · 66  
 extração · 34, 35, 39

---

### F

Fator crítico de sucesso · 54  
 fatores críticos de sucesso · 18, 52, 82  
 Front Page · 81

---

### G

granularidade · 42, 44, 47, 65, 66, 67

---

### I

I-CASE · 13, 15  
 informações estratégicas · 28, 35, 83  
 Integrado · 31

---

### L

limpeza · 34

---

### M

manutenção · 20, 23, 24, 26, 34  
 Manutenção · 16, 20, 40  
 metadados · 14, 27, 28, 31, 34  
 metas · 18, 52, 62  
 migração · 34  
 missão da empresa · 52  
 Missão da empresa · 54

modelagem de dados · 16, 21, 28, 32  
 modelagem dimensional · x, 13, 48, 50,  
 51, 63, 67, 82  
 modelo de dados · 13, 15, 18, 21, 26,  
 27, 41, 42, 44  
 modelo de dados corporativo · 55, 82  
 modelo dimensional · 14, 32, 33, 39, 49,  
 62, 63, 64, 65, 66, 69, 77, 81, 82, 84  
 modelo Entidade-Relacionamento · 32  
 modelo E-R · 32  
 modelo relacional · 63, 64, 65, 84

---

## N

Não Volátil · 31  
 Não-Aditivas · 68  
 Necessidades de informação · 54  
 necessidades de informações · 52, 53,  
 55  
 normalização · 26, 56, 58, 60  
 Normalização · 59, 66

---

## O

Objetivo da área funcional · 54  
 Objetivo da empresa · 54  
 objetivos da área funcional · 52  
 objetivos da empresa · 52, 82  
 OLAP · 68, 69, 70, 71, 74, 77, 79, 85  
*OLAP Services 2000* · 81  
 Orientado por Temas · 30

---

## P

pirâmide · 13, 15, 16, 17, 25  
 planejamento estratégico de  
 informações · 18, 21  
 Planejamento Estratégico de  
 Informações · 17, 51, 52  
 povoamento · 14, 27, 28, 39, 47  
 Problema · 54

problemas · 18, 20, 39, 52, 84  
 Projeto · 19, 20, 21, 23, 37, 39, 41, 44

---

## R

Ralph Kimball · 37, 47, 83  
 redundância · 37, 58, 59, 62, 63, 67

---

## S

Semi-Aditivas · 68  
 Simplificação do Modelo · 58  
*Snowflake Schema* · 67  
*SQL Server 2000* · 62, 81, 87  
*Star Schema* · 66, 67, 83  
 submodelo de dados · 19

---

## T

tabela de fatos · 32, 33, 75, 77  
 tabelas de Dimensão · 33, 64, 66, 67  
 tabelas de Dimensões · 78, 79, 81  
 tabelas de Fato · 67  
 transformação · 34, 39

---

## U

uso · x, 12, 18, 23, 27, 51, 52

---

## V

Variante no Tempo · 31  
 visualização dos dados · 63, 78, 81

---

## W

*Web* · 82, 83, 84  
*Websites* · 82  
 Willian H. Inmon · 41, 47