



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS  
DEPARTAMENTO DE ELETRÔNICA E SISTEMAS

MESTRADO EM ENGENHARIA ELÉTRICA

**UMA ESTRATÉGIA DE DESENVOLVIMENTO DE  
DATA WAREHOUSE GEOGRÁFICO COM  
INTEGRAÇÃO HÍBRIDA APLICADA AO  
MONITORAMENTO DE QUEIMADAS NA AMAZÔNIA**

ÁUREA HILÉIA DA SILVA MELO

DISSERTAÇÃO DE MESTRADO

RECIFE - PE  
30 de setembro de 2003

UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS  
DEPARTAMENTO DE ELETRÔNICA E SISTEMAS

ÁUREA HILÉIA DA SILVA MELO

**UMA ESTRATÉGIA DE DESENVOLVIMENTO DE DATA  
WAREHOUSE GEOGRÁFICO COM INTEGRAÇÃO HÍBRIDA  
APLICADA AO MONITORAMENTO DE QUEIMADAS NA  
AMAZÔNIA**

Trabalho apresentado ao Centro de Tecnologia e Geociências da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestre em Engenharia Elétrica com Ênfase em Redes de Computadores.

Orientador: Prof. Rafael Dueire Lins, PhD.

Co-orientador: Prof. Fernando da Fonseca de Souza, PhD.

RECIFE - PE  
30 de setembro de 2003

**ÁUREA HILÉIA DA SILVA MELO**

**UMA ESTRATÉGIA DE DESENVOLVIMENTO DE DATA WAREHOUSE  
GEOGRÁFICO COM INTEGRAÇÃO HÍBRIDA APLICADA AO  
MONITORAMENTO DE QUEIMADAS NA AMAZÔNIA**

Trabalho apresentado ao Centro de Tecnologia e Geociências da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestre em Engenharia Elétrica com Ênfase em Redes de Computadores.

Aprovado em 30 de Setembro de 2003.

**BANCA EXAMINADORA**

---

Rafael Dueire Lins, PhD  
Orientador

---

Profa. Fernanda Maria Ribeiro de Alencar, PhD

---

Prof. José Laurindo Campos dos Santos, PhD

À minha filhinha  
Ádria Melo Evangelista.

# Agradecimentos

*Primeiramente, agradeço a Deus por me fazer capaz, mas, sobretudo, pela força, coragem e fé para acreditar que seria capaz, vencendo assim os diversos, e não poucos, obstáculos surgidos ao longo deste trabalho.*

*À toda minha família, especialmente aos meus pais Francisco das Chagas Melo e Maria Valmira da S. Melo, cuja convivência comigo foi tantas vezes sacrificada em troca das minhas horas de estudo. Também aos meus filhos Gleice Ellen e Israel, principalmente, a você Ellen pelas privações de momentos de sua juventude para dar toda assistência em casa.*

*Ao meu amor Anderson Evangelista, pelo apoio nos momentos conclusivos deste trabalho e pelo seu amor, paciência e compreensão.*

*Às amigas de todas as horas Carla Oran e Nívia Menezes, pelo incentivo, convívio das madrugadas e pela amizade. E também a minha prima e amiga, Simone Melo, por muitas vezes assumir meu papel de mãe junto à minha filhinha Ádria.*

*À minha “cunhada predileta por parte do meu marido” Irninha, pelo apoio e amizade.*

*Às amigas Andréa e Andreza Mendonça, por jamais medirem esforços em ajudar-me nos momentos difíceis.*

*Ao professor Fernando Fonseca, pelo apoio, incentivo, e pelas muitas vezes que, tal como um pai, soube dizer as coisas certas nas horas certas. Muito obrigada!*

*Ao professor Rafael Dueire Lins, por TUDO.*

*Ao amigo pernambucano Unilton. Não há palavras que possam expressar meus sinceros agradecimentos por tudo que você fez por mim durante minhas estadias em Recife. Tens a minha eterna gratidão.*

*À família “amazonambucana” Barreto, pela hospedagem em sua “cobertura”, minimizando a saudade de casa. Mas, sobretudo, pelo apoio, incentivo e amizade.*

*Às Faculdades Objetivo, na pessoa da professora Maria Hercília Tribuzy, pela oportunidade de realizar este sonho.*

*Ao SIPAM, em especial a Marisa, a “garota do fogo”, por pacientemente atender-me em seu horário de expediente, fornecendo-me as informações necessárias para este trabalho. Ao amigo Augusto Queiroz, diretor do Centro Regional de Vigilância SIVAM Manaus, por ter tornado possível a realização do estudo de caso deste trabalho.*

*“É função do homem criativo perceber as relações entre os pensamentos, ou coisas, ou as formas de expressão que podem parecer diferentes, e ser capaz de combiná-los de novas maneiras - o poder de conectar o que aparentemente não seria possível.”*

—WILLIAM PLOMER

# Resumo

Os sistemas de suporte à decisão atravessaram as fronteiras dos negócios comerciais e atingiram, atualmente, as mais diversas áreas do conhecimento. Seja em instituições de ensino, governamentais, privadas com ou sem fins lucrativos, onde houver um processo decisório, haverá a necessidade de sistemas desse tipo. Nesta categoria de sistemas destaca-se o Data Warehouse (DW), que permite a geração de dados integrados e históricos possibilitando que seus usuários tomem decisões com base em fatos e não em especulações e, embora já exista há quase 20 anos, ainda não há uma metodologia formal para sua implementação. Tal fato estende-se também aos DW Geográficos (DWG), ou seja, um ambiente de DW com tratamento e armazenamento de dados geo-referenciados. Nesse contexto, o presente trabalho define uma estratégia de desenvolvimento de DWG, contemplando desde a fase de levantamento até a inserção dos dados, seguindo o método de integração híbrida e tendo como base, as metodologias de Ralph Kimball, Jonh A. Zachman e Vidette Poe para DW tradicionais. Além disso, é feito um estudo de caso do Sistema de Monitoramento de Queimadas da Amazônia Legal de forma a avaliar e validar cada uma das etapas da estratégia implementada.

**Palavras-Chaves:** Data Warehouse, Data Warehouse Geográfico, Desenvolvimento de Data Warehouse, Sistema de Informação Geográfica, Monitoramento Ambiental.

# Abstract

The decision support systems have crossed the boundaries of the commercial business and recently have reached the most diverse areas of knowledge. Whether in learning, government, private or non profitable institutions, wherever a decision process exists there is a need for such systems. In this system category we point out the Data Warehouse (DW), which allows the generation of integrated data and historical information that makes possible users to make decisions based on facts and not on assumptions and, although it already exists for over 20 years, there is still no formal methodology for its implementation. Such fact is also applied to the Geographical DW (GDW), i.e., a DW environment with geo-referenced data treatment and storage. In this context, this work defines a GDW development strategy wich envelops from requirements to data loading, following the hibrid integration method and using the Ralph Kimball, John A. Zachman and Vidette Poe methodology for traditional DW as basis. In addition, a case study of the Legal Amazon Burnning Monitoring System is made in order to evaluate and validate each one of the implemented strategy phases.

**Keywords:** Data Warehouse, Spatial Data, Data Warehouse Development, Geographic Information System, Monitoring Environmental.

# Sumário

<b>Capítulo 1—Introdução</b>	1
1.1 Objetivo Geral . . . . .	2
1.2 Objetivos Específicos . . . . .	2
1.3 Motivação . . . . .	3
1.4 Descrição do Problema . . . . .	3
1.5 Organização da Dissertação . . . . .	4
<b>Capítulo 2—Sistema de Informação Geográfica</b>	5
2.1 Conceitos e Características Principais . . . . .	5
2.2 Aplicações . . . . .	6
2.3 Dados Geográficos . . . . .	8
2.3.1 Classificação dos Dados Geográficos . . . . .	9
2.3.2 Modelos de Dados para SIG . . . . .	11
2.3.3 Modelos de Representação de Dados Espaciais . . . . .	12
2.3.4 Relacionamento entre Dados Geográficos e Dados Con- cionais . . . . .	16
<b>Capítulo 3—Data Warehouse</b>	18
3.1 Conceitos . . . . .	18
3.2 Características . . . . .	19
3.2.1 Orientado por temas (assuntos) . . . . .	19
3.2.2 Integrado . . . . .	20
3.2.3 Variante no tempo . . . . .	21
3.2.4 Não Volátil . . . . .	22
3.3 Tipos de Integração de Dados em DW . . . . .	22
3.3.1 DW Materializado . . . . .	22
3.3.2 DW Virtual . . . . .	23
3.3.3 DW Híbrido . . . . .	24
3.4 Construção de um DW . . . . .	26
3.4.1 Modelagem de Dados . . . . .	27
3.4.2 Extração, Transformação e Carga (ETL) dos dados . . . . .	30
3.5 Data Warehouse Geográfico . . . . .	32
3.6 Metodologia de Desenvolvimento de DWG . . . . .	33
3.6.1 Projeto Lógico de um DWG . . . . .	33
3.6.2 Dimensões em um DWG . . . . .	34
3.6.3 Medidas em um DWG . . . . .	35

<b>Capítulo 4—Metodologias para Construção de Data Warehouse</b>	<b>37</b>
4.1 Metodologia de Ralph Kimball . . . . .	37
4.1.1 Etapas do Desenvolvimento . . . . .	37
4.2 Proposta de Zachman . . . . .	41
4.3 Metodologia de Poe . . . . .	45
4.3.1 Ciclo de Vida de Suporte à Decisão . . . . .	45
4.4 Comparação entre as Metodologias . . . . .	50
<b>Capítulo 5—Estratégia para desenvolvimento de DWG</b>	<b>54</b>
5.1 Levantamento de Dados . . . . .	54
5.2 Classificação dos Dados . . . . .	58
5.3 Escolha do Modelo de Integração . . . . .	59
5.4 Modelagem dos Dados . . . . .	60
5.4.1 Identificar a tabela de fatos . . . . .	60
5.4.2 Identificar as Dimensões . . . . .	60
5.4.3 Identificar atributos da dimensão . . . . .	61
5.4.4 Associar Fatos e Dimensões . . . . .	61
5.4.5 Definir Atributos(medidas) da Tabela de fatos . . . . .	62
5.5 Extração, Transformação e Carga (ETL) . . . . .	62
5.6 Resumo . . . . .	65
<b>Capítulo 6—Implementação de um DWG para prevenção de Queimadas</b>	<b>66</b>
6.1 Descrição do Cenário . . . . .	66
6.1.1 O ArcSDE . . . . .	67
6.1.2 Considerações Gerais . . . . .	68
6.2 Aplicação da Estratégia Proposta . . . . .	68
6.2.1 Levantamento de Dados . . . . .	69
6.2.1.1 Identificação das Necessidades - O quê? . . . . .	69
6.2.1.2 Identificação dos dados - Onde? . . . . .	70
6.2.1.3 Granularidade dos Dados . . . . .	70
6.2.1.4 Estimativa do volume de Dados . . . . .	70
6.2.2 Classificação dos Dados . . . . .	71
6.2.3 Escolha do Modelo de Integração . . . . .	72
6.2.4 Modelagem dos Dados . . . . .	73
6.2.4.1 Identificação da Tabela de Fatos . . . . .	73
6.2.4.2 Identificação das Dimensões . . . . .	73
6.2.4.3 Identificação dos Atributos das Dimensões . . . . .	73
6.2.4.4 Associação entre a tabela de Fatos e as Dimensões . . . . .	74
6.2.4.5 Definição das medidas da tabela de Fatos . . . . .	74
6.2.5 ETL . . . . .	75
6.3 Análise dos resultados . . . . .	77
<b>Capítulo 7—Considerações Finais</b>	<b>79</b>



# Lista de Figuras

2.1	Visão de Camadas de um SIG . . . . .	6
2.2	Exemplos de aplicações SIG . . . . .	7
2.3	Objetos Vetoriais . . . . .	10
2.4	Objetos Matriciais . . . . .	10
2.5	Tipos de Geo-Campos . . . . .	11
2.6	Mapa da Amazônia Legal usando representação raster . . . . .	12
2.7	Mapa da Amazônia Legal no formato Vetorial . . . . .	13
2.8	Mapa Temático da Amazônia Legal: Tipo de Vegetação . . . . .	13
2.9	Mapa Cadastral . . . . .	14
2.10	Imagem da cidade de Manaus . . . . .	15
2.11	Exemplo de Redes aplicada ao sistema de luz elétrica [CM03]. . . . .	15
2.12	Exemplos de modelos numéricos de terreno [MEL03]. . . . .	16
2.13	Exemplo de ligação entre um geo-objeto e um objeto descritivo. . . . .	17
3.1	Característica Orientado por Assunto do DW . . . . .	20
3.2	Exemplo de Integração do atributo sexo . . . . .	20
3.3	Característica Variante no Tempo do DW . . . . .	21
3.4	Característica Não Volátil do DW . . . . .	22
3.5	Data Warehouse com Integração Materializada . . . . .	23
3.6	Data Warehouse com Integração Virtual . . . . .	24
3.7	DWG com Integração Híbrida . . . . .	25
3.8	Modelo Estrela . . . . .	28
3.9	Exemplo de Desnormalização no Modelo Estrela . . . . .	29
3.10	Dimensão Localização normalizada pelo modelo Floco de Neve . . . . .	30
5.1	Ciclo de Vida da estratégia de desenvolvimento de um DWG . . . . .	55
5.2	Relacionamento entre dimensões e a tabela de fatos . . . . .	62
5.3	Exemplo de consolidação das camadas do GIS . . . . .	64
6.1	Modelo de Armazenamento do ArcSDE . . . . .	68
6.2	Associação da tabela Fatos e as Dimensões . . . . .	74
6.3	Representação de Fatos e Dimensões no ErWin . . . . .	74
6.4	Modelagem do DWG . . . . .	75
6.5	Modelagem da DSA . . . . .	76

# Lista de Tabelas

2.1	Principais produtos SIG e seus fornecedores . . . . .	8
4.1	Matriz de Perspectivas e Dimensões de Zachman . . . . .	44
4.2	Resumo da Comparação entre as Metodologias de Kimball, Zachman e Poe. . . . .	53
5.1	Lista com informações sobre os dados requeridos . . . . .	63
6.1	Identificação dos Atributos . . . . .	71

# Lista de Abreviaturas

BD - Banco de Dados  
DM - Data Mart  
DSA - Data Staging Area  
DW - Data Warehouse  
DWG - Data Warehouse Geográfico  
ETL - Extração, Transformação e Carga  
GIS - Geographic Information Systems  
IBAMA - Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis  
INCRA - Instituto Nacional de Colonização e Reforma Agrária  
INPA - Instituto Nacional de Pesquisa da Amazônia  
INPE - Instituto Nacional de Pesquisas Espaciais  
FUNAI - Fundação Nacional do Índio  
OLAP - On-Line Analytical Processing  
OLTP - Processamento de Transações On-Line  
SAE/PR - Secretaria de Assuntos Estratégicos da Presidência da República  
SAD - Sistema de Apoio à Decisão  
SDW - Spatial Data Warehouse  
SGBD - Sistema de Gerenciamento de Banco de Dados  
SGBDOO - Sistema de Gerenciamento de Banco de Dados Orientado Objeto  
SGBDOR - Sistema de Gerenciamento de Banco de Dados Objeto Relacional  
SIG - Sistema de Informação Geográfica  
SIPAM - Sistema de Proteção da Amazônia Legal  
SIVAM - Sistema de Vigilância da Amazônia

# Capítulo 1

## Introdução

Sistemas de Informação Geográfica (SIG)[ARO89, BUR98, IR93, BER99] são aplicações que permitem a manipulação conjunta de dados convencionais e dados espacialmente referenciados, os quais costumam ser tratados com uma abordagem temática e orientada à aplicação, incorporando a dimensão geográfica à modelagem de dados convencional.

Essas aplicações são muito apropriadas à área de meio ambiente, em virtude de sua capacidade de produzir mapas e analisar eventos relativos ao globo terrestre. Atualmente caracterizam-se como um tipo bastante particular de sistema de apoio à decisão, em razão da incorporação de funcionalidades de manipulação (por exemplo, relacionamentos espaciais tais como adjacências, proximidades, cruzamentos, entre outros) e análise de dados geo-referenciados, com níveis de abstração e padrões de visualização intuitivamente mais próximos do mundo real.

A característica de apoio à decisão, nesse contexto, foi possível graças a mudança na tecnologia de armazenamento de dados, antes feito na forma de arquivos e, atualmente, utilizando Sistemas de Gerenciamento de Banco de Dados (SGBD). Este fato proporcionou aos usuários informações confiáveis auxiliando-os na tomada de decisão [CÂM00].

Assim como a tecnologia SIG, a de Banco de Dados (BD) também obteve um avanço significativo mediante a inclusão de novas tecnologias tais como *data mining* [HK00, HMS01], *data warehouse* (DW)[KIM96, INM96, MAC00, SIN01] e recuperação da informação [YRN99, LOP02] cujos mecanismos avançados de consultas proporcionaram aos usuários maior domínio na manipulação e extração de informações. A tecnologia de DW, em particular, está sendo amplamente utilizada como ferramenta de apoio à decisão, por contemplar a integração de dados provenientes de múltiplas fontes, manutenção de históricos, rapidez nas consultas, além de permitir, através das ferramentas OLAP (*On-Line Analytical Processing*) [THO02], análises estatísticas e simulação eficiente de novas associações entre os dados.

A combinação de dados geo-espaciais do SIG com a tecnologia de DW originou um novo seguimento de pesquisa denominado de *Spatial Data Warehouse* (SDW) [SLTC02] ou *Data Warehouse Geográfico* (DWG). Esse novo segmento torna-se bastante útil e atrativo, uma vez que pode proporcionar um histórico mediante o armazenamento de dados consolidados, característica esta limitada na tecnologia de SIG e necessária para

apoio à tomada de decisão em muitas áreas do conhecimento. Dentre elas, destaca-se o domínio ambiental, cujos problemas com o desflorestamento, ocupação e uso da terra, enchentes, unidades de preservação e conservação, terras indígenas e, em particular, a destruição de florestas proveniente de queimadas requerem informações precisas que auxiliem os gestores não apenas em suas ações corretivas, mas, sobretudo, em ações que possibilitem soluções preventivas quanto ao risco de degradação ambiental irreversível.

Nesse contexto é que o presente trabalho de dissertação está concentrado, uma vez que utilizou a tecnologia de DW geográficos de forma a prover informações necessárias aos gestores ambientais, para o planejamento e gerenciamento de ações essenciais e preventivas quanto ao risco de incêndios na Amazônia Legal (que compreende a Região Norte do Brasil, o estado do Mato Grosso e parte do estado do Maranhão).

## 1.1 Objetivo Geral

Descrever uma estratégia de desenvolvimento de DWG seguindo o modelo de integração do tipo híbrido, aplicada ao monitoramento de focos de calor para controle e prevenção de queimadas na Amazônia Legal.

## 1.2 Objetivos Específicos

- Discutir as abordagens de desenvolvimento de Ralph Kimball[K<sup>+</sup>98], John A. Zachman [IZG97] e Vidette Poe[P<sup>+</sup>98] para DW convencional, visto que esse autores, apresentam técnicas consolidadas, cujas referências são amplamente utilizadas tanto no meio acadêmico quanto no profissional.
- Verificar a existência de metodologias de construção de DWG;
- Verificar a aplicabilidade de etapas de desenvolvimento de DW convencional para DWG;
- Validar o uso de integração híbrida no desenvolvimento do DWG, como forma de minimizar o espaço em disco e a complexidade de manipulação de dados geográficos;
- Evidenciar a necessidade de um DWG como forma de auxílio na prevenção de queimadas na Amazônia Legal;
- Descrever uma estratégia de desenvolvimento de DWG, fundamentada nas metodologias de Kimball, Zachman e Poe;
- Implementar um DWG objetivando validar a aplicabilidade da estratégia proposta.

## 1.3 Motivação

A Amazônia é reconhecida como a maior floresta tropical existente, o que a torna uma grande fonte de recursos. Prever fogo na Amazônia é uma tarefa desafiadora, mediante sua imensa extensão, grande diversidade de tipos de floresta e de solo e pelo amplo espectro de práticas de uso de terra [NMA99].

Dado o tamanho e a riqueza da biodiversidade Amazônica, torna-se imprescindível o zoneamento ecológico-econômico e o monitoramento da região a fim de preservar os seus recursos naturais, em particular as florestas, responsáveis pela produção de oxigênio, absorção de dióxido de carbono da atmosfera e por apresentar uma diversidade de espécies arbóreas de grande valor econômico e ambiental [BAVU98]. Adicionalmente, as florestas são ainda imprescindíveis na conservação dos solos, da fauna e dos recursos hídricos [CS82].

Entretanto, o equilíbrio da biodiversidade proporcionado pela floresta Amazônica está ameaçado, na maioria das vezes, por ações antrópicas que estão substituindo as florestas por ecossistemas mais simples (por exemplo, a pecuária, agricultura e área de mineração) por meio de derrubadas e queimadas de espécies arbóreas.

Os órgãos ambientais dos diferentes níveis de governo, os quais possuem as missões de preservar, monitorar e conservar os recursos naturais são muitas vezes limitados no cumprimento de suas ações em razão da falta de integração dos dados que, na maioria das vezes, encontram-se departamentalizados, dado o domínio de atuação de cada um.

A falta de integração de dados para prover informações que auxiliem os gestores ambientais na tomada de decisões preventivas quanto ao risco de incêndios na Amazônia Legal, constitui-se numa ameaça à sustentabilidade dos recursos naturais, caracterizando-se como um problema que pode ser minimizado por meio da implantação de um DWG.

## 1.4 Descrição do Problema

O isolamento de dados no nicho de atuação de cada um dos órgãos responsáveis pelo tratamento de questões ambientais na Amazônia Legal, tais como IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis) responsável pelo licenciamento e fiscalização das atividades ambientais, o INCRA (Instituto Nacional de Colonização e Reforma Agrária) que realiza atividades de assentamentos, a FUNAI (Fundação Nacional do Índio) que identifica e regulariza as áreas de reserva indígena entre outras, constitui-se como um problema para o fornecimento de informações confiáveis sobre riscos de incêndio. Isto é, cada um desses órgãos mantém seus próprios dados armazenados com estrutura particular e desta forma possuem uma visão setorializada das informações, limitando a execução de suas atividades.

Além disso, as soluções tecnológicas para armazenamento, gerenciamento e manipulação dos dados geográficos não contemplam satisfatoriamente o fornecimento de dados históricos sobre as áreas ameaçadas por incêndio, fato este imprescindível no processo de prevenção.

No contexto da dispersão de dados, é inviável prover informações que subsidiem os gestores ambientais na tomada de decisão quanto à prevenção de queimadas de florestas. Podendo esse fato ser minimizado pela aplicação da tecnologia de DWG, visto que fornecem aos usuários uma base histórica e integrada, proporcionando assim, a elaboração de consultas mais eficientes, com resultados mais precisos e possibilitando uma melhor análise dos dados mais acuradas.

## 1.5 Organização da Dissertação

Esta dissertação está organizada em 7 capítulos. Além desta introdução, os Capítulos 2 e 3 apresentam, respectivamente, um breve estudo sobre Sistema de Informação Geográfica (SIG) e Data Warehouse, os quais contemplam os principais conceitos envolvidos pelo tema desta dissertação.

O Capítulo 4 discute as metodologias de desenvolvimento tanto de DWG quanto convencional, especificamente, as definidas por R. Kimball [KIM97], W. Inmon e Zachman [INM97] e Vidette Poe [P<sup>+</sup>98].

O Capítulo 5 descreve uma estratégia de desenvolvimento de um DWG fundamentada nas metodologias apresentadas no Capítulo 4 e no método de integração híbrida.

No Capítulo 6 será apresentado um estudo de caso utilizado como forma de avaliar e validar a aplicabilidade da estratégia proposta. A saber, o Sistema de Monitoramento e Detecção de Focos de Incêndio do Sistema de Proteção da Amazônia Legal (SIPAM). Posteriormente, serão destacadas algumas discussões e análise a respeito dos resultados obtidos.

No capítulo 7 são apresentadas as considerações finais deste trabalho, bem como propostas para trabalhos futuros.

# Capítulo 2

## Sistema de Informação Geográfica

Um SIG é um sistema baseado em computador que permite capturar, modelar, manipular, recuperar, consultar, analisar e apresentar dados geograficamente referenciados [ARO89, C<sup>+</sup>96]. Tais sistemas são poderosas ferramentas de software capazes de produzir mapas e analisar uma grande quantidade de eventos relativos ao globo terrestre. Suas múltiplas características, funcionalidades e aplicações permitem que possa ser inserido nas mais diversas áreas do conhecimento, como por exemplo, marketing, planejamento de uso do solo e análise de recursos ambientais. É o que trata este capítulo, cujo objetivo é apresentar um breve estudo sobre SIG e dados geográficos de forma a proporcionar uma base conceitual para escopo deste trabalho.

### 2.1 Conceitos e Características Principais

Os SIG são tecnologias de Geoprocessamento tais como Cartografia Automatizada, Processamento de Imagens Digitais e Sensoriamento Remoto que lidam com informações na forma de dados geográficos. Estes dados permitem que se conheça a estrutura geométrica de objetos espaciais (casa, rua, rio, parcela de solo e viatura), sua posição no espaço geográfico e seus atributos. Alguns SIG possuem ainda a capacidade de manipular relacionamentos espaciais (proximidade, adjacência, cruzamento, sobreposição, continência, entre outros). As análises desses objetos espaciais podem ser efetuadas por meio de diversas operações específicas de manipulação dos dados tais como projeção, seleção, união e sobreposição [RSV02]. São as ferramentas de manipulação que conferem aos SIG capacidade em apoiar processos de tomada de decisão sobre o domínio espacial, uma vez que aos decisores normalmente são apresentados mapas, imagens, gráficos, animações e tabelas.

Um SIG é uma base de dados computacional que permite aos usuários a organização de informação cartográfica de diferentes fontes. Essa informação depois de registrada no sistema torna-se consistente, permitindo que diferentes níveis de informação (temas, camadas, *layers*) possam ser combinados e visualizados de acordo com as necessidades de análise. A Figura 2.1 ilustra um exemplo no qual existem três camadas, que uma vez sobrepostas proporcionarão aos usuários informações sobre tipo de vegetação, tribos indígenas e unidades de preservação em determinadas áreas.

O conceito de SIG, depende da área seguida pelo autor. De um modo geral, os conceitos apresentados atualmente contemplam três visões diferentes: baseada em mapas, em bases de dados e na análise espacial [MAG91, ABR98].



**Figura 2.1. Visão de Camadas de um SIG**

- **visão baseada em mapas** - apresenta o SIG como sistemas para processamento e visualização de mapas [TOM91].
- **visão baseada em banco de dados** - enfatiza a importância dos SIG terem subjacente uma base de dados bem projetada e possuem um SGBD robusto [FRA88]. Nesse contexto, Smith [SMI87] define SIG como um banco de dados indexado espacialmente, sobre o qual opera um conjunto de procedimentos para responder a consultas sobre entidades espaciais;
- **visão baseada na análise espacial** - distingue a capacidade dos SIG para efetuarem análise espacial, defendendo a existência de uma ciência da informação espacial em alternativa à perspectiva tecnológica com que geralmente os SIG são abordados [GOO92, OPE91].

A partir da terceira perspectiva surge um outro conceito dado por Cowen [COW91] que define SIG como “um sistema de suporte à decisão que integra dados referenciados espacialmente num ambiente de respostas a problemas”.

Quanto às características de um SIG, segundo Câmara [CÂM00] as principais são:

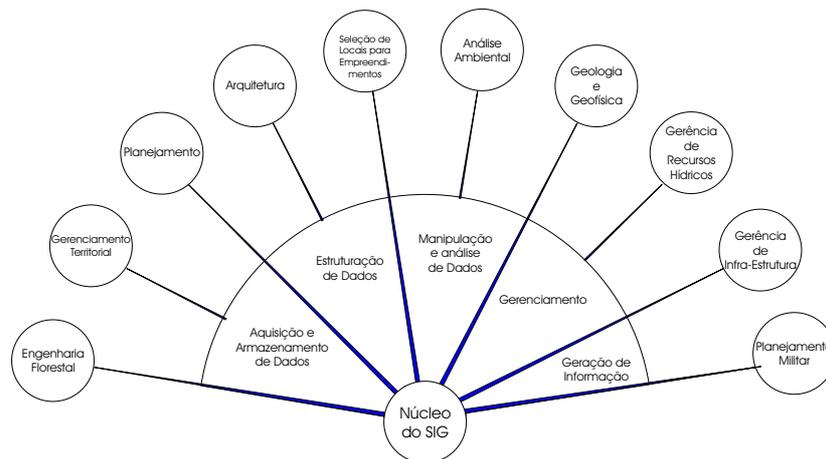
- integrar, numa única base de dados, informações espaciais provenientes de dados cartográficos, dados de censo e cadastro urbano e rural, imagens de satélite, redes e modelos numéricos de terreno;
- combinar as várias informações, através de algoritmos de manipulação, para gerar mapeamentos derivados;
- consultar, recuperar, visualizar e desenhar o conteúdo da base de dados geocodificados.

## 2.2 Aplicações

Considerando sua grande capacidade de manipular informações espaciais, os SIG possuem uma multiplicidade de aplicações nas mais diversas áreas do conhecimento, e que recebe, dependendo do autor, diferentes formas de classificação. Por exemplo, alguns

autores classificam as aplicações de SIG em: sócio-econômicas, ambientais e de Gerenciamento [C<sup>+</sup>96, M<sup>+</sup>93]. Outros como Aronoff [ARO89] descrevem aplicações representativas para as quais um SIG pode ser utilizado com sucesso, tais como agricultura e planejamento do uso da terra; silvicultura e gerenciamento da vida silvestre; arqueologia, geologia e aplicações municipais.

Essas classes de aplicações, não deixam de ser aplicações gerenciais, cujos principais objetivos são o gerenciamento de recursos naturais, agricultura, meio ambiente, entre outros. No Brasil, particularmente, órgãos como INPE (Instituto Nacional de Pesquisas Espaciais) e SIVAM (Sistema de Vigilância da Amazônia) utilizam aplicações SIG de forma a prover o monitoramento de queimadas, vôos clandestinos, áreas de preservação, entre outros, auxiliando os tomadores de decisão dos diferentes níveis de governo e aos demais órgãos a eles associados, tais como IBAMA, Polícia Federal, INPA (Instituto Nacional de Pesquisas da Amazônia) e FUNAI.



**Figura 2.2. Exemplos de aplicações SIG**

Nesse contexto, a Figura 2.2 apresenta algumas das diferentes áreas de aplicação de SIG, cujo crescimento constante e sistematizado permitiu que os sistemas se tornassem mais eficientes mais comuns e, conseqüentemente, tivessem um menor custo.

Importante salientar que o fator que mais contribuiu para o progresso dos SIG foi, indubitavelmente, o avanço das tecnologias de banco de dados e o suporte que as mesmas proporcionaram às aplicações SIG. Esse avanço vem sendo feito de forma sistemática, convergindo, quem sabe, para um SGBD (Sistema de Gerenciamento de Banco de Dados) totalmente espacial [RSV02], capaz de atender às necessidades de armazenamento de dados espaciais, com suporte completo a manipulação dos mesmos, proporcionando vantagens, principalmente, no desempenho das aplicações SIG visto que ter-se-á armazenado dados descritivos e geográficos numa única base de dados, possibilitando também a utilização dos mecanismos de integridade do SGBD tanto para os dados descritivos quanto para os

geográficos, tais como recuperação de paradas e falhas e controle de concorrência; e as manipulações mediante operações específicas, tais como sobreposição, união e concatenação de mapas. Essas operações, atualmente, não são possíveis via SGBD dada a arquitetura dual dos SIG, composta de um SGBD relacional que armazena os dados descritivos e um módulo espacial que contém os dados geográficos. Entretanto, a tendência atual é fazer uso dos SGBD que suportam as características espaciais, definindo-as como objetos complexos. Para isso, alguns SIG já estão incorporando Sistemas de Gerenciamento de Banco de Dados Orientado a Objetos (SGBDOO)<sup>1</sup> como o O2 da (*O<sub>2</sub> Technology*), *PostGres* (da Universidade da Califórnia) e outros estão optando por Sistemas de Gerenciamento de Banco de Dados Objeto-Relacional (SGBDOR) por exemplo, *Informix* (da *Informix*), DB2 (da IBM) e o *Oracle* (da *Oracle Corp.*) a partir da versão *8i*.

**Tabela 2.1.** Principais produtos SIG e seus fornecedores

Software	Fabricante	Site
ArcInfo, ArcView e ArcSDE	ESRI	<a href="http://www.esri.com">http://www.esri.com</a>
ERDAS	Leica Geosystems	<a href="http://www.erdas.com">http://www.erdas.com</a>
MapInfo	MAPINFO	<a href="http://www.mapinfo.com">http://www.mapinfo.com</a>
SPRING	INPE	<a href="http://www.dpi.inpe.br/spring">http://www.dpi.inpe.br/spring</a>

De um modo geral, convém dizer que as aplicações SIG são desenvolvidas, na maioria das vezes, com o uso de um software GIS. Por exemplo, ARC/VIEW e ARC/INFO (da ESRI), MGE (da Intergraph Corp.) entre outros. A Tabela 2.1 relaciona os principais produtos SIG e seus respectivos fornecedores. Através desses produtos vários projetos têm sido desenvolvidos no Brasil, particularmente pelos governos federal, estaduais e municipais, além de Organizações privadas. Um exemplo desses projetos já concretizados é o SIVAM (abordado no contexto desta dissertação), o qual foi concebido pela SAE/PR (Secretaria de Assuntos Estratégicos da Presidência da República), em conjunto com os Ministérios da Justiça e Aeronáutica, com o propósito de preservar a Amazônia Legal.

Uma vez que os dados geográficos compõem os principais elementos de um DWG, a seção a seguir tem como objetivo apresentar os conceitos, a classificação bem como os modelos de dados desses tipos de dados.

## 2.3 Dados Geográficos

Os SIG englobam dois principais tipos de atributos: atributos descritivos e espaciais. Os atributos descritivos correspondem ao conjunto de dados armazenados em um banco de dados convencional, importante para compreender e descrever os elementos

---

<sup>1</sup>Trabalhos como [Eng87, Dit88] lançaram os fundamentos para o uso de SGBDOO no desenvolvimento de SIG. Posteriormente trabalhos de caráter exploratório como [GS92] parecem confirmar a possibilidade de desenvolvimento de produtos baseados em SGBDOO com algumas vantagens.

espaciais. Enquanto que os atributos espaciais denotam qualquer tipo de dados que descreve fenômenos aos quais esteja associada alguma dimensão, por exemplo, as estruturas moleculares de um composto químico.

Os dados usados em SIG pertencem a uma classe particular de dados espaciais, são os dados geográficos ou dados geo-referenciados, os quais descrevem fatos, objetos e fenômenos do globo associados à sua localização sobre a superfície terrestre, em um certo instante ou período de tempo [C<sup>+</sup>96].

Segundo Aronoff [ARO89] e Medeiros [MP94] os dados geográficos possuem quatro características principais: posição geográfica (coordenadas), seus atributos (valores dos dados), seu relacionamento topológico e seus componentes temporais.

Uma vez que esses dados estão armazenados num SIG, eles são caracterizados a partir de 3 componentes fundamentais [ARO89, OII90, DAN90, PEU94]:

- 1) **Dados Convencionais** - atributos alfanuméricos armazenados pelos SGBD tradicionais;
- 2) **Dados Espaciais** - são os atributos que descrevem a geometria e a localização do fenômeno geográfico. Possuem propriedades geométricas (por exemplo, o tamanho) e topológicas (por exemplo, conectividade);
- 3) **Dados Pictoriais** - são os atributos que armazenam imagens (por exemplo, uma foto) e são gerenciados por funções de processamento de imagens.

### 2.3.1 Classificação dos Dados Geográficos

Os dados geográficos estão divididos em Objetos Vetoriais e Objetos Matriciais, os quais definem entidades geométricas fundamentais tais como pontos, linhas e áreas usadas para representar as características ambientais.

A seguir são definidas as terminologias padrões segundo o Comitê Nacional Norte-Americano para Padrões de Dados Cartográficos Digitais (NCDCDS), as quais são ilustradas nas Figuras 2.3. e 2.4.

#### Objetos Vetoriais

- **ponto** (*point*) - um objeto adimensional especificando localização geográfica por um conjunto de coordenadas;
- **nó** (*node*) - um objeto adimensional servindo como junção topológica a uma linha ou como ponto extremo (inicial ou final);
- **segmento Linear** (*line segment*) - uma linha reta entre dois pontos;

- **seguência Linear (*string*)** - uma sequência de segmentos lineares sem nós;
- **cadeia(*chain*)** - uma sequência de segmentos lineares com nó inicial e nó final;
- **anel (*ring*)** - uma sequência de cadeias ou sequências lineares que formam uma poligonal fechada;
- **polígono (*polygon*)** - um anel e sua área interna.

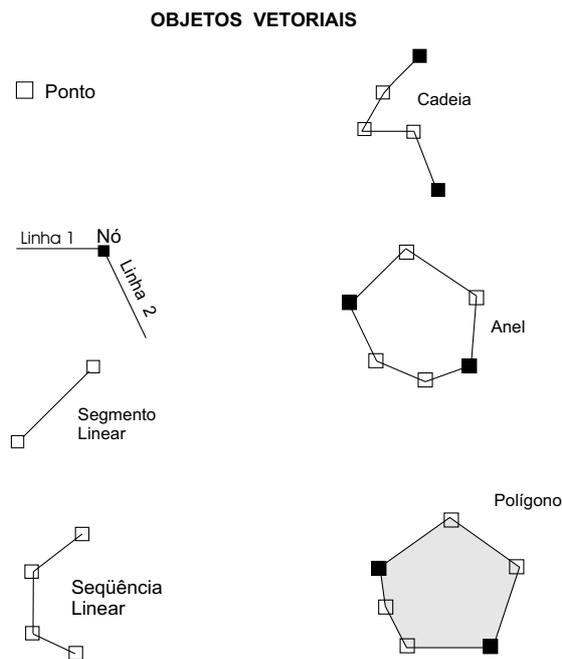


Figura 2.3. Objetos Vetoriais

### Objetos Matriciais

- **elemento Matricial (*pixel*)** - um elemento bidimensional que é a menor parte não divisível em uma imagem.
- **célula do Quadriculado(*grid cell*)** - um objeto bidimensional representado num elemento de uma tesselação regular da superfície.

### OBJETOS MATRICIAIS



Figura 2.4. Objetos Matriciais

### 2.3.2 Modelos de Dados para SIG

O mundo real, no contexto de aplicação SIG, pode ser modelado segundo duas visões: Campos e Objetos [COU92, FG90, G<sup>+</sup>92].

- 1) **Visão de Campo (geo-campos)** - representa as variações espaciais contínuas, usadas para representar grandezas distribuídas espacialmente, tais como solo e teor de minerais. Sua ênfase está no conteúdo das áreas e não nos seus limites, ou seja, este tipo de visão é utilizado principalmente nas situações de medidas imprecisas, como por exemplo determinar qual a temperatura de uma dada região (deve-se criar funções para determinar os valores intermediários aos que foram coletados). Esta visão possui seis tipos diferentes (ilustrados na Figura 2.5) que são: amostragem irregular de pontos, linhas de contorno ou isolinhas, polígonos adjacentes amostragem regular de pontos, grade regular de células e grade triangular.
- 2) **Visão Objetos (geo-objetos)** - são também chamados de objetos geográficos. Este tipo de dado espacial caracteriza-se por ter atributos não espaciais, e por poder estar associado a representações gráficas. Por exemplo: escolas, municípios e lotes de terra. É composto basicamente por uma identidade, uma parte descritiva e um componente espacial. A identidade, que o diferencia dentre os demais no espaço, e a sua parte descritiva constituem o seu componente não espacial. O componente espacial pode ser uma estrutura de dados - que descreve um vértice (ponto), uma aresta (linha) ou uma face (polígono ou região) - ou recursivamente uma coleção de outros geo-objetos. No primeiro caso, diz-se que o geo-objeto é atômico e, no segundo caso, tem-se um geo-objeto complexo.

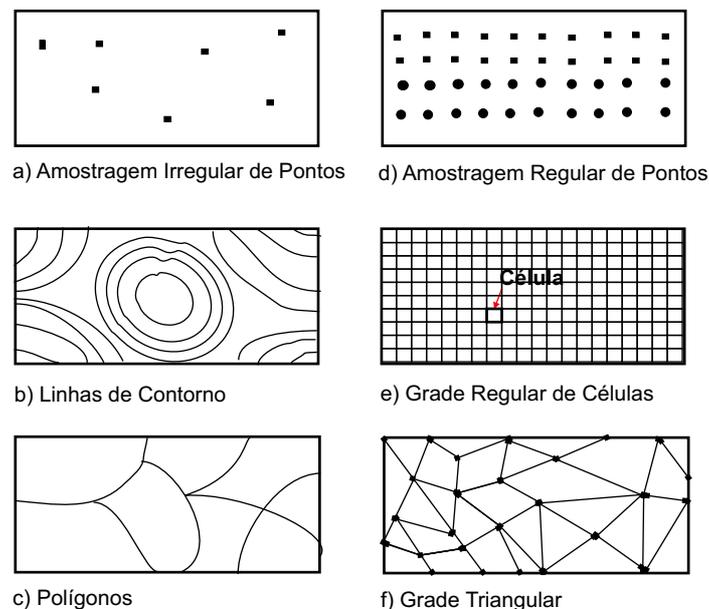


Figura 2.5. Tipos de Geo-Campos

### 2.3.3 Modelos de Representação de Dados Espaciais

Existem duas abordagens principais para representação dos componentes espaciais associados às informações geográficas que são: Representação Matricial (*raster* ou tesselação) e Representação Vetorial.

**Representação Raster** - é geralmente usada para representação de geo-campos. Caracteriza-se por uma matriz de células de tamanhos regulares, em que para cada célula é associado um conjunto de valores representando as características geográficas da região. O conjunto de células e respectivos valores constitui uma camada. Cada célula só pode armazenar um tipo de valor correspondente a um tema, dessa forma, uma área tem que ser representada por diferentes camadas. A figura 2.6 apresenta o mapa da Região Norte usando a representação *raster*, composto das camadas A, B, C, D e E.

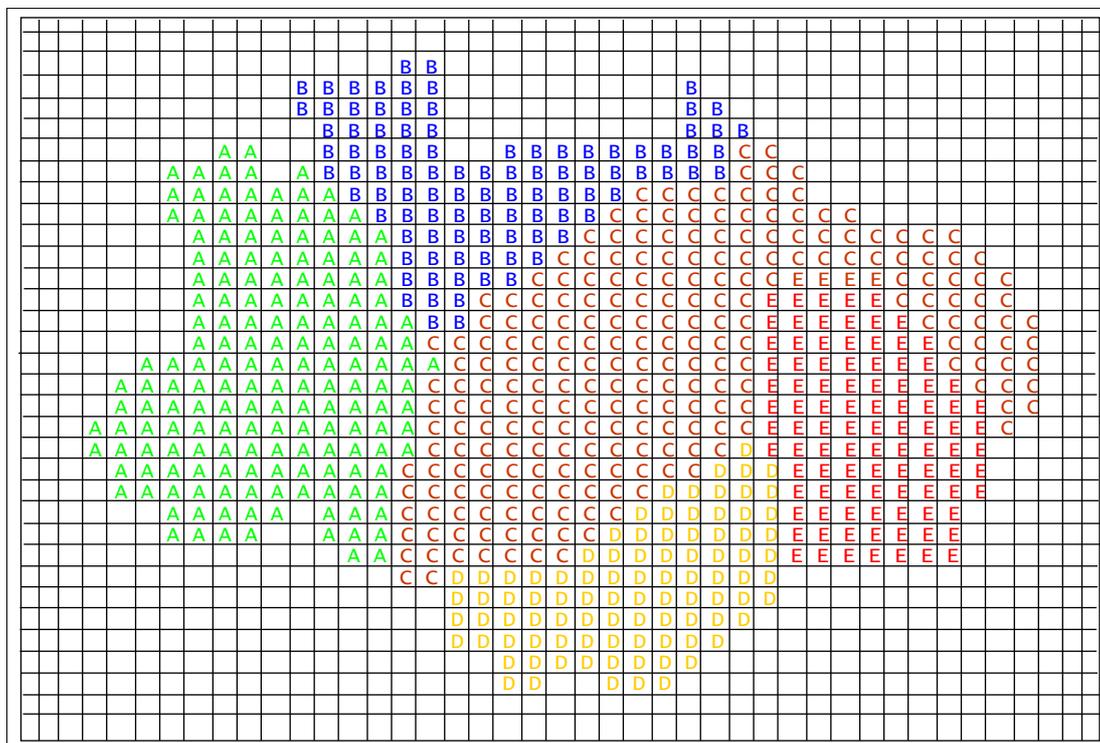
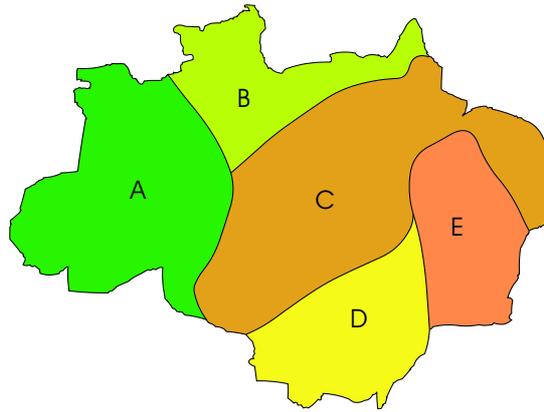


Figura 2.6. Mapa da Amazônia Legal usando representação raster

**Representação Vetorial** - usada frequentemente para representação de geo-objetos, utiliza pontos, linhas e polígonos para representar as fronteiras das entidades geométricas, tal como elas são desenhadas em um mapa. Neste tipo de representação, os dados são armazenados segundo a topologia arco-nó-região, nos quais os arcos estabelecem os limites das regiões e conectam-se entre si por meio dos nós [HAR97], conforme pode se visualizar na Figura 2.7.

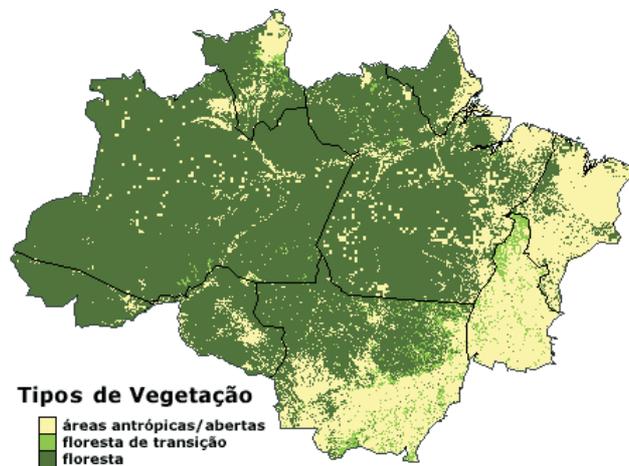


**Figura 2.7. Mapa da Amazônia Legal no formato Vetorial**

Geo-objetos e geo-campos possuem algumas classes para tratamento em Geoprocessamento [CÂM95], dentre elas estão os mapas temáticos, os mapas cadastrais, imagens, rede e modelos numéricos de terreno (MNT) e . A seguir será descrito cada uma dessas classes, embora este trabalho só englobe os três primeiros, dada a aplicabilidade desses no contexto de queimadas, áreas indígenas, entre outras.

### Mapas Temáticos

Conjunto dos objetos que possuem alguma afinidade estrutural ou funcional, por exemplo, para representar a hidrografia de uma região, estradas de rodagem, tipos de solos, entre outros [ARO89]. Ou seja, um mapa que representa um tema ou um conjunto de temas específicos. A Figura 2.8 ilustra um exemplo de mapa temático sobre o tipo de vegetação na Amazônia Legal.



**Figura 2.8. Mapa Temático da Amazônia Legal: Tipo de Vegetação**

## Mapas Cadastrais

Em um mapa cadastral, ao contrário do que ocorre em um mapa temático, cada um de seus elementos é um objeto geográfico, que possui atributos e pode estar associado a várias representações gráficas [C<sup>+</sup>96].

A maioria das aplicações que lidam com essas mapas precisa ter um relacionamento entre esses mapas e seus atributos no BD, uma vez que são feitas consultas ao BD para serem apresentados os resultados de forma simbólica [C<sup>+</sup>96].

A Figura 2.9 ilustra um exemplo de mapa cadastral (Fonte dos dados [IBG01]), o qual associa alguns Estados da Amazônia Legal com seus respectivos atributos descritivos, tais como população e o tamanho do território.



Figura 2.9. Mapa Cadastral

## Imagens

Com o desenvolvimento das áreas de processamento digital de imagens e sensoriamento remoto, o uso de imagens de satélites tem se tornado muito comum em SIG.

Uma imagem digital consiste em uma matriz de números digitais chamados de pixels (abreviação de *picture element*). Cada pixel corresponde a certas características físicas e/ou químicas da superfície da imagem original, não digital.

O processamento digital de imagens compreende um conjunto de operações de manipulações numéricas sobre imagens digitais. A aplicação de técnicas de análise de imagens pode permitir a identificação de qualquer fenômeno, a partir dos dados fornecidos por sensores remotos sobre determinada área [HAR97]. A Figura 2.10 ilustra uma imagem obtida via satélite da cidade de Manaus.

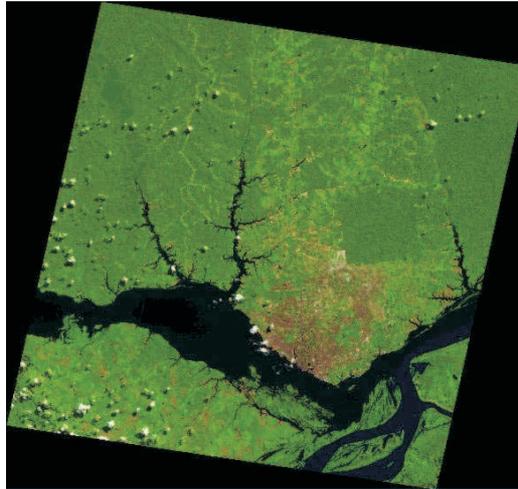


Figura 2.10. Imagem da cidade de Manaus

## Redes

Em Geoprocessamento, o conceito de redes está associado a informações referentes a serviços de utilidade pública, tais como água, luz ou telefone [HAR97].

Redes também são do tipo geo-objetos e, assim como os mapas cadastrais, caracterizam-se pela localização geográfica precisa e a associação de atributos contidos em um banco de dados convencional. A Figura 2.11 apresenta um exemplo de redes aplicado ao serviço de luz elétrica.

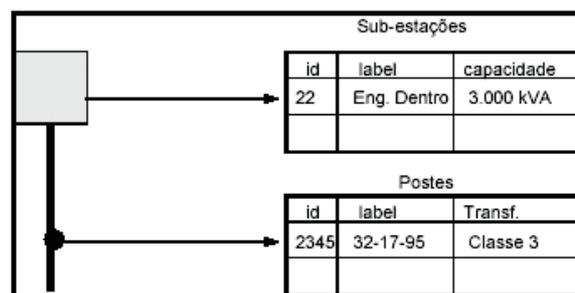


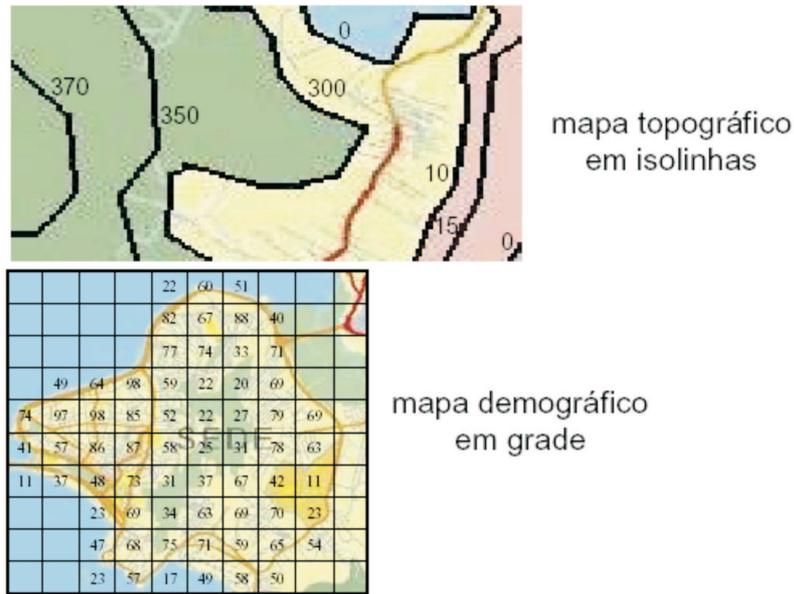
Figura 2.11. Exemplo de Redes aplicada ao sistema de luz elétrica [CM03].

## Modelos Numéricos de Terreno(MNT)

O termo MNT é usado para representar uma grandeza que varia continuamente no espaço. Entre as aplicações de MNT, pode-se incluir [CÂM95, HAR97]:

- armazenamento de dados de altimetria para gerar mapas topográficos;

- análises de corte-aterro para projeto de estradas e barragens; e
- análise de variáveis geofísicas e geoquímicas.



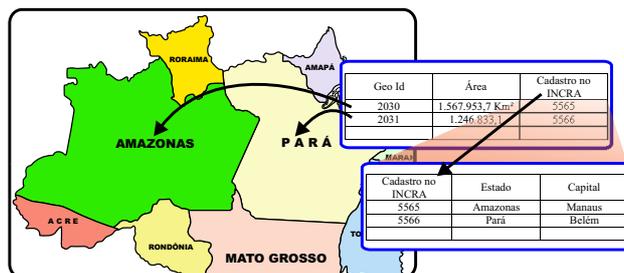
**Figura 2.12. Exemplos de modelos numéricos de terreno [MEL03].**

Existem três formas básicas de representação de modelos numéricos de terreno. São elas: grades regulares, grades triangulares e isolinhas [HAR97]. A Figura 2.12 apresenta exemplo de grades regulares e isolinhas.

### 2.3.4 Relacionamento entre Dados Geográficos e Dados Convencionais

Na sua forma mais usual, a integração entre informações geográficas e um ambiente de banco de dados, usa um SGBD para armazenar os atributos convencionais dos objetos geográficos e arquivos para armazenar representações geométricas [C<sup>+</sup>96]. Para cada instância de atributos descritivos, é destinado um identificador único (GeoId), chamado de rótulo, através do qual será feita a conexão lógica com as representações gráficas por eles descritas. A figura 2.13 ilustra um exemplo dessa ligação com o uso de rótulos. Ou seja, de acordo com a referida figura, os Estados do Amazonas e Pará possuem um GeoId o qual estará armazenado em uma tabela do banco de dados contendo os demais dados referentes a estes Estados. Esta tabela por sua vez poderá se relacionar outras tabelas do banco de dados.

Desta forma os identificadores tornam um SIG capaz de associar a representação gráfica dos geo-objetos com sua respectiva informação descritiva que está armazenada na forma de tabela no banco de dados. Além disso, permite computar novas informações e exibí-las na forma de mapas [C<sup>+</sup>96].



**Figura 2.13. Exemplo de ligação entre um geo-objeto e um objeto descritivo.**

De acordo com Câmara [CÂM95], para que a conexão entre dados espaciais e dados convencionais ocorra, três condições devem ser obedecidas:

- manter um relacionamento único entre dados geográfico sobre mapa e registros na tabela de atributos;
- manter uma ligação entre o dado geográfico e o registro através de um único identificador;
- manter o identificador armazenado fisicamente no arquivo que contém os dados geográficos e no arquivo que contém o correspondente registro da tabela de atributos.

Muitos conceitos abordados por este capítulo, tais como objeto vetorial, pontos, linhas e dados geográficos constituem a base de terminologias usadas no escopo desse trabalho.

Dando continuidade a esta parte conceitual, o capítulo a seguir trata das demais tecnologias contempladas pelo tema dessa dissertação.

# Capítulo 3

## Data Warehouse

Este capítulo aborda os principais conceitos, características, terminologias, aplicações e ciclo de vida da tecnologia de DW convencional e geográfico, necessários ao entendimento do contexto desta dissertação.

### 3.1 Conceitos

Geralmente, a maioria das organizações possui um grande volume de dados a respeito dos mais diversos serviços por elas prestados. Todavia, muitas vezes este imenso número de dados são sub utilizados, uma vez que não há mecanismos que possam explorá-los, convertendo-os em informações significativas, para auxiliar no processo decisório.

De forma a atender essas necessidades do âmbito empresarial, particularmente aos tomadores de decisões, surgiu o DW, definido como o processo de integração dos dados corporativos de uma empresa em um único repositório e a partir do qual os usuários finais podem facilmente executar consultas, gerar relatórios e fazer análises [SIN01].

A princípio, esse conceito parece não acrescentar nenhuma novidade, uma vez que nos sistemas convencionais pode-se também obter dados para relatórios e consultas. Porém, a principal diferença é o fato do DW permitir que os dados brutos necessários para uma análise profunda do negócio possam ser extraídos, mesmo que estejam em locais e formatos distintos, por exemplo banco de dados, planilhas e arquivos textos. Além disso, nos sistemas OLTP (processamento de transações on-line) [P<sup>+</sup>98] ou transacionais, embora seja possível a emissão de relatórios, muitas vezes tem-se que efetuar cálculos manualmente para que informações possam ser cruzadas e avaliadas. Tal fato não ocorre no DW, uma vez que essas operações são inerentes ao processo de criação do mesmo.

Pode-se elencar, então, três motivos básicos para justificar um DW:

- o primeiro está inserido nos dados corporativos, uma vez que existem padrões valiosos de informação que são muito importantes para dirigir o negócio;
- o segundo, é que esta informação irá formar a base de serviços únicos para clientes, descobrindo novas tendências, prevendo resultados de tal forma que possa transformar o entendimento que a Companhia pode ter do mercado;

- o terceiro diz respeito à diminuição da distância entre a identificação da estratégia e a sua execução. Isto permitirá um progressivo entendimento da companhia em relação a sua própria estrutura organizacional.

Dessa forma, os DW têm sua maior aplicabilidade em Sistemas de Apoio à Decisão (SAD) nas mais diversas áreas do conhecimento e, apesar de ser uma tecnologia ainda com o custo elevado, é possível para a empresa adotar mecanismos de implementação mediante a utilização de seus próprios recursos [SOU03].

Essa tecnologia traz, portanto, vários benefícios aos decisores, visto que tornou-se uma poderosa ferramenta competitiva, permitindo que seus usuários pudessem acessar uma representação integrada das múltiplas fontes de informações da empresa, garantindo a consistência das normas de gerenciamento e das convenções aplicadas aos dados.

Enfim, um DW proporciona aos seus usuários finais dados com qualidade que refletem as necessidades da empresa como um todo, mediante a disponibilização de informações mais precisas e eficientes, possibilitando decisões com base em fatos e não em intuições, fazendo com que novos mercados, novas oportunidades e novos produtos sejam descobertos, além de criar uma relação mais próxima com o cliente.

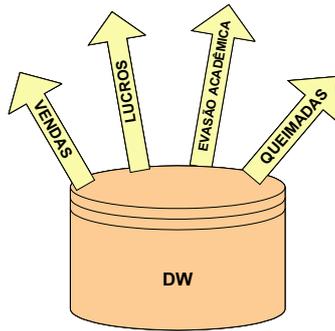
A seguir será mostrado um breve estudo sobre a tecnologia de DW, suas características, aplicações, tipos, seu ciclo de desenvolvimento, seu processo de construção e de que forma podem dar suporte de armazenamento a dados georreferenciados.

## 3.2 Características

Segundo W. H. Inmon [INM92], considerado o pioneiro no tema, um DW 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. Esta definição inclui as próprias características de um DW que são: Orientado por Tema, Integrado, Variante no Tempo e Não Volátil.

### 3.2.1 Orientado por temas (assuntos)

A primeira característica de um DW é que ele é projetado segundo um principal assunto (tema) da organização. Ou seja, a orientação por assunto, nada mais é do que o direcionamento aplicado para proporcionar aos usuários uma visão do negócio da empresa, por exemplo: numa empresa de telecomunicação, o principal assunto é o cliente, que podem ser residenciais, empresas e telefonia pública. Então, em um projeto de DW essas premissas devem ser consideradas.

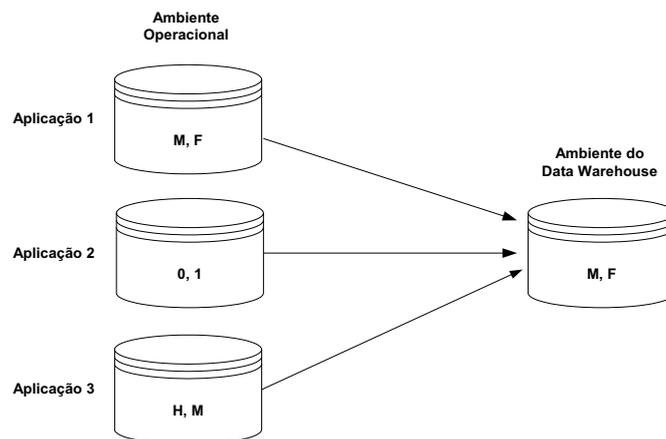


**Figura 3.1. Característica Orientado por Assunto do DW**

A Figura 3.1 apresenta alguns assuntos típicos de DW, inclusive de um DWG (queimadas).

### 3.2.2 Integrado

O mais importante aspecto do ambiente de DW é que os dados nele armazenados são, sem nenhuma exceção, sempre integrados. O processo de integração é responsável por sincronizar os dados de todos os sistemas existentes na empresa, e colocá-los em um mesmo padrão. Ou seja, a integração refere-se à consistência de nomes, das unidades das variáveis, dentre outros, no sentido de que os dados devem ser transformados até um estado uniforme.



**Figura 3.2. Exemplo de Integração do atributo sexo**

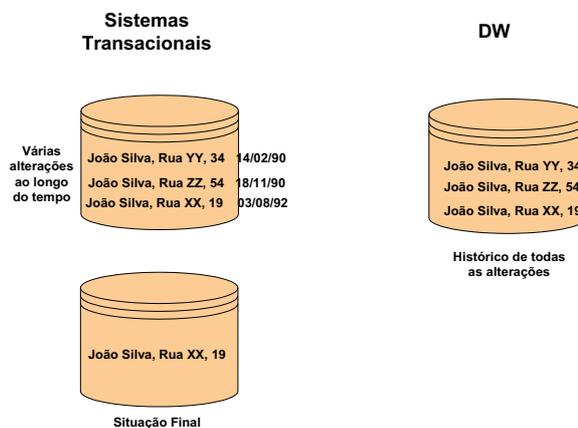
A Figura 3.2 mostra um exemplo clássico de integração referente ao atributo sexo: supondo-se um sistema A em que esse dado está armazenado no formato M para masculino, e F para feminino, já em um segundo sistema B, o mesmo dado está guardado,

como 0 para masculino e 1 para feminino e um terceiro C como H e M. Este fato gera algumas dificuldades na fase de análise, porém no processo de criação do DW todos os formatos são convertidos em um único padrão, o qual pode ser decidido com o usuário final. No caso do exemplo exposto, o padrão definido foi M para masculino e F para feminino.

### 3.2.3 Variante no tempo

Nos Sistemas transacionais o dado de produção é atualizado de acordo com mudanças em alguns requisitos, os quais refletem, geralmente, o estado do objeto no momento do acesso a um DW. A cada ocorrência de uma mudança, uma nova entrada é criada para marcar esta mudança. Ou seja, os dados no DW não são atualizáveis.

Esse tratamento de séries temporais apresenta características específicas, que adicionam complexidade ao ambiente do DW. 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 têm uma característica temporal, mas também os metadados, que incluem definições como as dos itens de dados, rotinas de validação e algoritmos de derivação. Sem a manutenção do histórico dos metadados, as mudanças das regras de negócio que afetam os dados no DW são perdidas, invalidando seu histórico.



**Figura 3.3. Característica Variante no Tempo do DW**

A Figura 3.3 apresenta um exemplo de dados que são alterados nos sistemas transacionais, ficando sempre a última alteração efetuada, enquanto que em um DW o histórico de todas as alterações são mantidas.

### 3.2.4 Não Volátil

Significa que o DW permite apenas a carga e consultas aos dados, não podendo os mesmos serem atualizados ou excluídos, caracterizando o chamado ambiente *load-and-access*. Pode-se afirmar que, os DW permitem somente as operações de Inserção e Consulta, conforme ilustra a Figura 3.4.

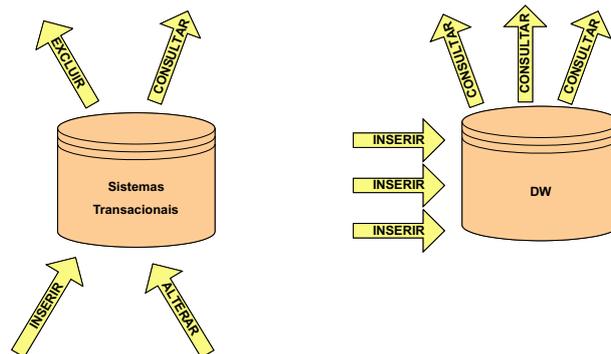


Figura 3.4. Característica Não Volátil do DW

A diferença desta característica de um DW para os sistemas convencionais é que nestes últimos, os dados são, em geral, atualizados registro a registro, em múltiplas transações. Esta volatilidade requer um trabalho considerável para assegurar integridade e consistência através de mecanismos de *rollback*, *commits*, recuperação de falhas e controle de concorrência. Esse grau de controle, típico dos sistemas orientados a transações, não são necessários no ambiente de DW [CRF97].

## 3.3 Tipos de Integração de Dados em DW

Esta seção aborda os mecanismos usados para fornecer um acesso integrado a diversas fontes de dados, dentre os quais o materializado e o virtual. Além destes será visto também a abordagem híbrida.

### 3.3.1 DW Materializado

Consiste em um DW tradicional, ou seja, nessa abordagem os dados passam pelas etapas de construção de um DW, visto que são previamente coletados, integrados e armazenados normalmente para posteriores consultas (Ver Figura 3.5). As consultas submetidas ao sistema de integração são processadas nesse repositório sem haver acesso aos mesmos às fontes de dados. Esta abordagem também pode ser aplicada a um DWG, no qual tanto os dados geográficos quanto os dados convencionais são fisicamente armazenados em um único ambiente.

O DW materializado apresenta a vantagem de proporcionar um bom tempo de resposta nas consultas, uma vez que permite acesso direto aos dados armazenados no DW. Além disso, mantém todas as características de um DW, especificamente a da não-volatilidade, visto que os dados armazenados não são modificados.

Por outro lado, um DW com este tipo de integração apresenta, como principal desvantagem, alto custo financeiro, complexidade de desenvolvimento, além de despesas com pessoal e recursos de *hardware*, especificamente os de espaço em disco.

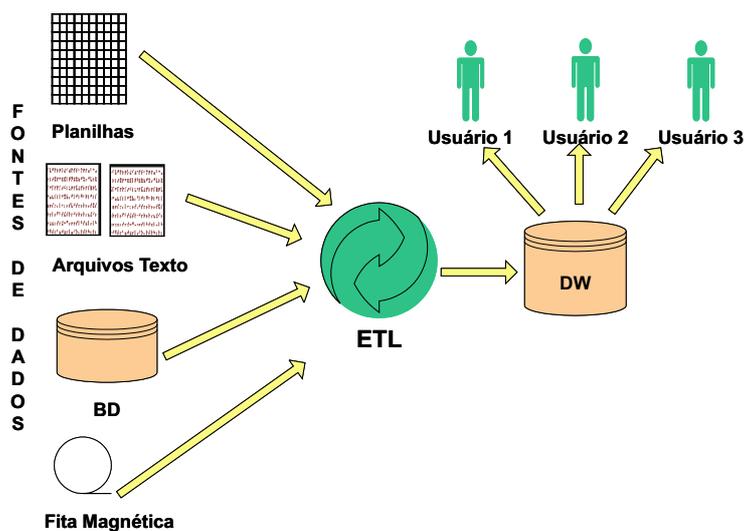


Figura 3.5. Data Warehouse com Integração Materializada

No que se refere aos dados geográficos, os mesmo ocupam, naturalmente, uma enorme área em disco. Conseqüentemente, criar um DWG materializado significa replicar muitos desses dados, o que torna a abordagem materializada não muito interessante para as organizações, principalmente pelo custo, visto que ter um DWG materializado significa ter pelo menos duas vezes a área que se tem reservada para o sistema OLTP. É evidente que isso não é uma regra, pode-se afirmar, inclusive, que se uma Instituição possui recursos suficientes para investir em um projeto de SIG, deve ter também para manter e estender esses sistemas. Contudo, algumas Instituições mesmo tendo condições de investir em recursos, ainda preferem a abordagem virtual ou até mesmo a híbrida, descritas a seguir.

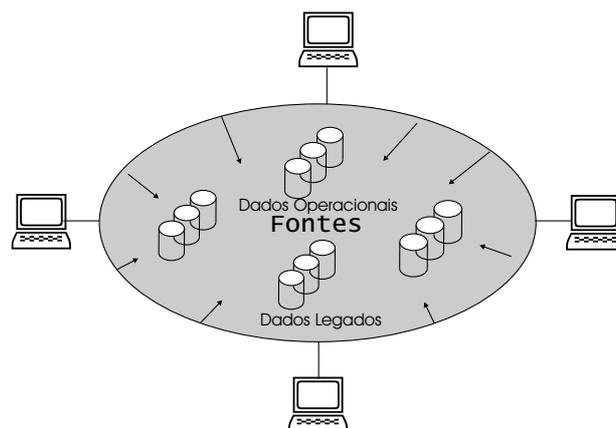
### 3.3.2 DW Virtual

A Figura 3.6 mostra um exemplo da abordagem virtual, a qual consiste em manter os dados em suas fontes, proporcionando ao usuário um acesso aos mesmos no momento de execução das consultas. Este tipo de abordagem é bastante atrativa, particularmente,

pelo seu menor custo e rapidez de desenvolvimento. Além disso, evita a replicação de dados.

Entretanto, a integração virtual se por um lado resolve o problema da replicação de dados, por outro afeta particularmente a característica da não-volatilidade do DW, visto que permanecendo os dados nas fontes originais, ficam vulneráveis às manipulações de atualização. Ou seja, se os dados forem atualizados fornecerão aos usuários finais diferentes informações para uma mesma consulta em períodos distintos. Além disso, caso o banco de dados operacional não esteja em funcionamento, o DW também não poderá funcionar, causando uma dependência entre ambos.

Uma outra desvantagem é que as consultas ao DW concorrem com as transações dos dados de produção, causando um prejuízo no que diz respeito ao desempenho; e como não há metadados (descrição da estrutura que compõe o DW) nem dados de sumário, todas as consultas devem ser repetidas, sobrecarregando o sistema. Acima de tudo não há um processo de depuração ou reciclagem, tornando as consultas muito complexas.



**Figura 3.6. Data Warehouse com Integração Virtual**

Portanto, as vantagens e desvantagens desse tipo de integração devem ser apresentadas aos usuários para que os mesmos tenham consciência e possam participar da decisão sobre o que é melhor para o projeto.

### 3.3.3 DW Híbrido

A integração híbrida envolve, em um único DW, tanto os aspectos virtuais quanto os materializados. Isto é, alguns dados são inseridos no DW (dados materializados), enquanto outros continuam nas fontes (dados virtuais). Sendo que no primeiro caso, os dados passam pelo processo completo de desenvolvimento de um DW (ver seção 3.4),

enquanto que no segundo, passam apenas por um tratamento para em seguida serem disponibilizados pra os usuários.

Este tipo de integração abrange, naturalmente, as vantagens e desvantagens de ambas abordagens. No entanto, tais desvantagens são um tanto quanto minimizadas pelo fato de os dados estarem divididos em ambas abordagens.

Um fator crítico no desenvolvimento de DW híbridos, consiste na decisão de quais dados serão definidos como virtuais e quais serão materializados. Uma sugestão é definir os virtuais, a partir de um levantamento estatístico sobre quais os dados que mais passam por alterações nos ambientes transacionais. Dessa forma pode-se diminuir a inconsistência.

Ressalta-se que este tipo de integração pode ser melhor aplicada a DWG, visto que ao contrário dos dados convencionais que são, geralmente, inseridos via digitação humana, a maioria dos dados geográficos são obtidos por equipamentos de grande precisão tais como satélites, *scanners* e mesas digitalizadoras, tornando-os menos suscetíveis a erros. Por esse motivo, os dados geográficos são os melhores candidatos a comporem a parte virtual do DW híbrido. A Figura 3.7 ilustra um DW a partir de uma integração híbrida, na qual tem-se a Fonte de Dados 1, cujos dados passam por todo o processo de ETL para serem inseridos no DW, e a Fonte de Dados 2 (virtual) que permanece com seus dados, disponibilizando-os para consultas.

A escolha do tipo de modelo de integração é uma tarefa exclusiva do desenvolvedor, baseada na fase de levantamento e, sobretudo, no apoio do usuário final, visto que é essencial que este tenha consciência das vantagens e desvantagens da abordagem a ser utilizada.

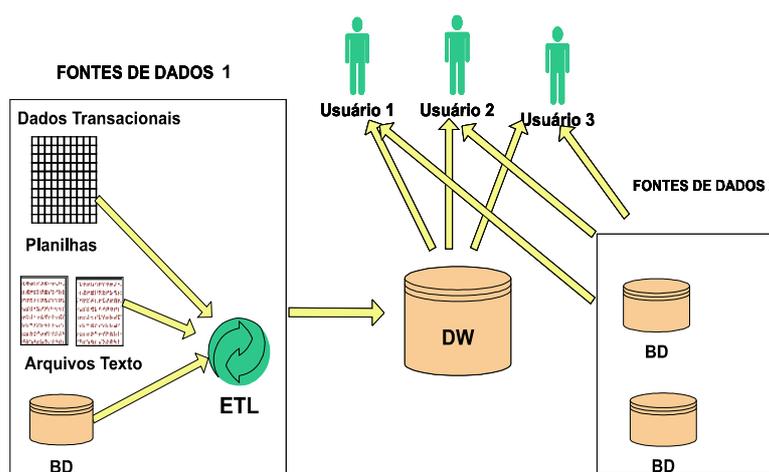


Figura 3.7. DWG com Integração Híbrida

A seguir será apresentado um ciclo de desenvolvimento para a construção de um DW. Ressaltando-se que por se tratar de um ciclo genérico, o mesmo não contempla a escolha do modelo de integração, sendo esse aspecto considerado somente na estratégia de desenvolvimento apresentada no capítulo 5.

### 3.4 Construção de um DW

O sucesso do desenvolvimento de um 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. Importante ressaltar que os DW não são construídos de uma só vez, ao contrário disso, eles são projetados e povoados passo a passo. Essa estratégia, chamada de evolucionária, é uma forma de diminuir os custos que uma implementação completa teria, além dos recursos consumidos e impactos no ambiente operacional da empresa [INM96].

Seguindo essa forma evolucionária muitas empresas iniciam o processo a partir de uma área específica, que normalmente é carente de informação e cujo trabalho é relevante para os negócios; cria-se então para essas áreas os chamados *data marts* (um DW departamental), para depois crescer de forma sistemática, seguindo uma estratégia *botton-up*, assunto-por-assunto ou ainda departamento por departamento. Salienta-se, entretanto, *data marts* (DM) também podem ser desenvolvidos como subconjunto de um DW maior, em busca de autonomia, melhor desempenho e simplicidade de compreensão [GW98, P+98].

Outra alternativa é selecionar um grupo de usuários, prover ferramentas adequadas, construir um protótipo do DW, e deixar que esses usuários façam experiências com pequenas amostras de dados. Somente após a concordância do grupo quanto aos requisitos e funcionamento, é que o DW de fato é carregado com dados dos sistemas operacionais na empresa e dados externos.

Na verdade é difícil indicar, atualmente, uma metodologia consolidada e amplamente aceita para o desenvolvimento de DW. De qualquer forma, a metodologia a ser adotada é ainda bastante dependente da abordagem escolhida em termos de ambiente, distribuição, arquitetura, entre outros. Isto é percebido com base no fato de que alguns autores definem diferentes critérios que devem ser considerados no desenvolvimento de um DW, mas que de um modo geral, contribuem para o sucesso do projeto. Por exemplo, Inmon [INM96] considera aspectos tais como: granularidade (nível de detalhe contido no DW), particionamento (repartição dos dados em unidades físicas separadas para serem tratadas independentemente), delimitação dos dados pesquisados, entre outras. Já Singh [SIN01] aborda indexação, particionamento, dados públicos e de detalhe permanente.

Uma contribuição interessante é dada em [DAL99, PER99, OLI02], onde são descritas algumas técnicas baseadas em software para incremento da performance do DW e em [CRF97] na qual são definidos algumas etapas de desenvolvimento para uso do modelo estrela.

O capítulo 4 apresenta uma breve descrição sobre as metodologias de desenvolvimento de DW, com base nos estudos de Poe[P<sup>+</sup>98], Kimball[K<sup>+</sup>98] e Zachman[IZG97], em face de seus trabalhos já estarem bastante consolidados e servirem, atualmente, como uma grande referência aos desenvolvedores e pesquisadores de DW.

Independentemente da metodologia a ser seguida, o desenvolvimento de um DW deve contemplar as seguintes etapas: Levantamento de Dados, Modelagem dos Dados e Extração, Transformação e Carga dos Dados (ETL), as quais constituem os passos básicos para construção de qualquer DW, havendo divergências de opiniões, na maioria dos casos na forma de condução das mesmas.

### **Levantamento de Dados**

Apesar de serem displicentemente ignoradas, as metodologias de levantamento de dados gerenciais são indispensáveis ao sucesso de um SAD que pretende atender às necessidades do usuário de negócio.

Atualmente, já existem metodologias de levantamento de dados Gerenciais, como a JAD (*Joint Application Design*)[IBM] e o DMD(*Dynamic Meeting Design*)[owg], que são baseadas em reuniões de trabalho, nas quais os participantes, orientados por um profissional com prática nesta etapa, extraem conhecimentos sobre o negócio.

Basicamente, as metodologias funcionam por meio de *Brain Storms* conduzidos por um facilitador. As missões dos Departamentos envolvidos, questões gerenciais e as informações que as respondem são levantadas, originando assim os objetos de negócio. Além dessas questões e respostas, também são levantadas a ciclicidade e o tempo de permanência dos dados na base do DW.

#### **3.4.1 Modelagem de Dados**

A especificação de requisitos do ambiente de suporte à decisão associado a um DW, é fundamentalmente diferente da especificação de requisitos dos sistemas que sustentam os processos usuais do ambiente operacional de uma empresa. Ou seja, é um erro pensar que técnicas de projeto que servem para sistemas convencionais serão adequadas para a construção de um DW.

Os requisitos para um DW não podem ser conhecidos até que ele esteja parcialmente carregado e já em uso. Dessa forma, a maioria das técnicas de modelagem é unânime no fato de que a aplicação completa da teoria relacional não é apropriada para DW.

A modelagem feita para DW é denominada de modelagem multidimensional, a qual é uma técnica de concepção e visualização de um modelo de dados de um conjunto de medidas que descrevem aspectos comuns de negócios. É utilizada especialmente para sumarizar e estruturar dados e apresentá-los em visões que suportem a análise dos valores dos dados [MAC00].

Para se fazer a modelagem de dados de um DW tem-se, atualmente, dois modelos mais usados: Modelo Estrela e Modelo Floco de Neve.

### Modelo Estrela (*Star Model*)

Esse modelo é composto por uma tabela central denominada tabela de fatos (*fact table*), a qual é ligada a um conjunto de tabelas menores chamadas de dimensões (*dimension tables*). Determinou-se o nome Estrela, justamente pela semelhança que a associação entre a tabela de fatos e as dimensões geram com o desenho de uma estrela (ver Figura 3.8).

A tabela de fatos representa um relacionamento muitos-para-muitos com as tabelas de dimensões, e sua chave primária é composta de todas as chaves estrangeiras das tabelas de dimensão.

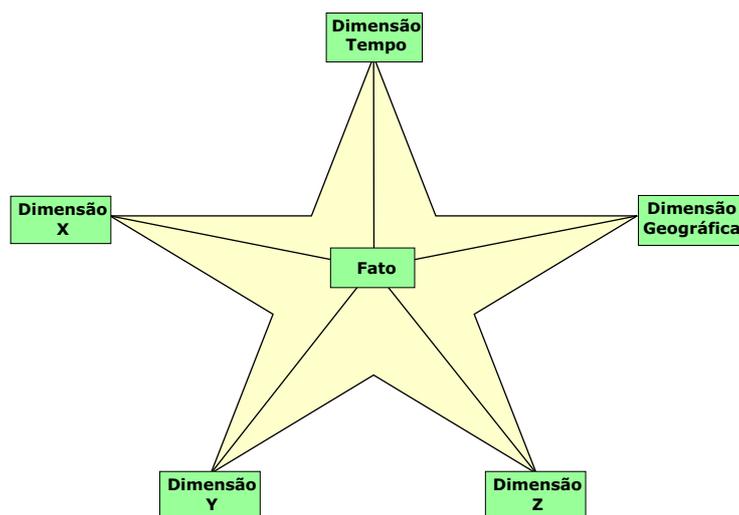
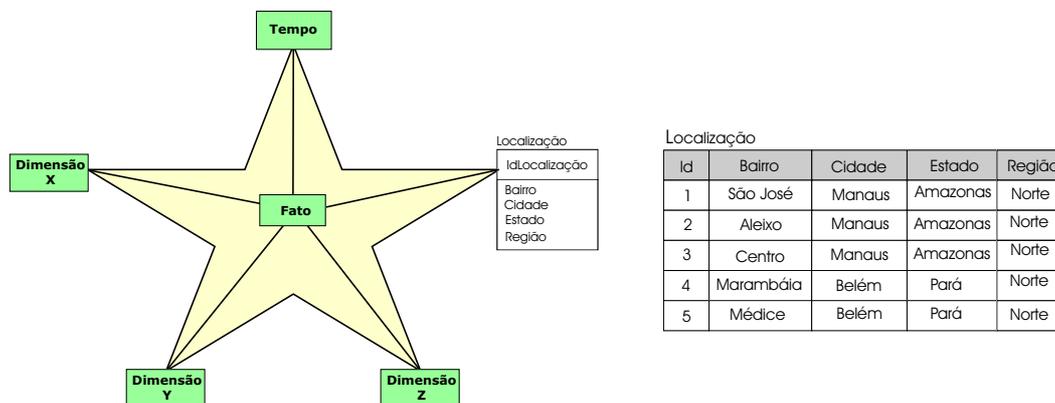


Figura 3.8. Modelo Estrela

Os atributos das tabelas de dimensão servem como fonte para restrições em uma consulta ou como cabeçalhos de linha no conjunto de resposta do usuário. Devido essas tabelas tenderem a usar tipos caracteres ao invés de numéricos, de um modo geral suas linhas são muito mais longas porém em menor quantidade, ocupando um pequeno percentual de espaço em disco. Por outro lado, a tabela de fatos é responsável pela utilização de até 95% da área destinada ao DW [OLI02].

Importante ressaltar que o modelo estrela corresponde a um modelo não normalizado, uma vez que o mais importante para o sistema como um todo é o *desempenho*, o qual é conseguido evitando-se *joins* (junções) entre várias tabelas já que os dados encontram-se (mesmo que de forma repetitiva) numa única tabela (dimensão). A Figura 3.9 mostra um exemplo no qual para cada cidade da dimensão Localização que pertença ao mesmo Estado, ter-se-á a repetição da descrição da região além da descrição do Estado.

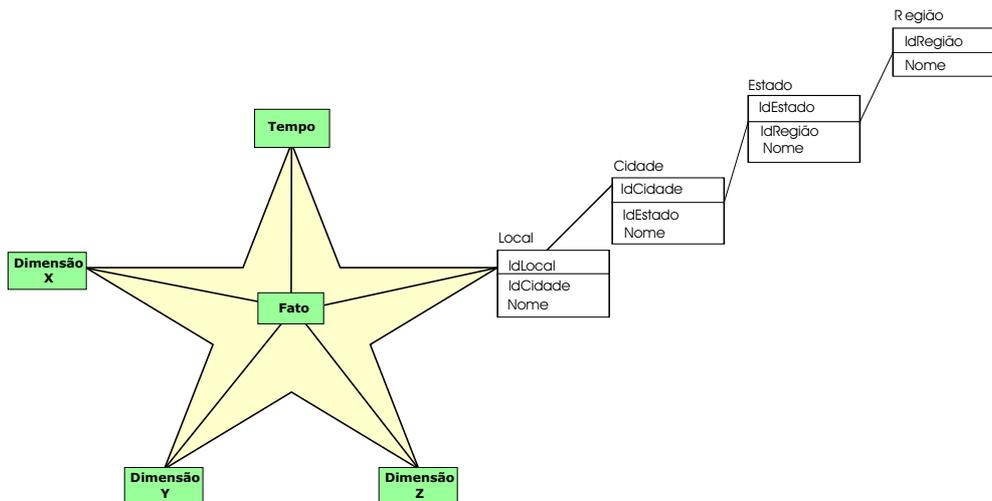


**Figura 3.9. Exemplo de Desnormalização no Modelo Estrela**

### Modelo Floco de Neve (*Snowflake Model*)

O Modelo Floco de Neve consiste em uma extensão do Modelo Estrela, no qual cada uma das pontas da estrela torna-se o centro de outras estrelas. Isto se deve ao fato de que esse modelo faz a normalização das dimensões. A Figura 3.10 apresenta a questão da Localização, anteriormente mostrada pela Figura 3.9, normalizada pelo modelo Floco de Neve.

Para sumarizar a arquitetura Floco de Neve, cada tabela dimensional armazena uma chave para cada nível da hierarquia da dimensão (isto é, para cada elemento da dimensão). A chave de nível mais baixo é combinada com a tabela dimensional na tabela central de fatos e na tabela de atributos que contém informações descritivas sobre o elemento da dimensão de nível mais baixo. As outras chaves são combinadas com a tabela dimensional nas tabelas de atributo correspondentes [SIN01].



**Figura 3.10. Dimensão Localização normalizada pelo modelo Floco de Neve**

A principal desvantagem do Floco de Neve em relação ao Estrela é a relativa complexidade da estrutura de dados normalizada. Caso os usuários estejam gerando consultas típicas e *ad hoc*, será mais difícil navegar através do Floco de Neve. Adicionalmente, os programas de carga e a manutenção geral ficam mais difíceis de administrar à medida que o modelo de dados se torna mais complexo.

### 3.4.2 Extração, Transformação e Carga (ETL) dos dados

A etapa de ETL é uma das mais críticas de um projeto de DW, pois um dado transferido erroneamente pode trazer conseqüências imprevisíveis nas fases posteriores. O objetivo desta fase é fazer a integração dos dados fontes, muitas vezes múltiplos e complexos. Sendo assim, os desenvolvedores devem despender uma atenção considerável nesse processo para que o sucesso do DW não seja comprometido.

Atualmente, embora existam várias ferramentas, por exemplo SUNOPSIS[Sun], *Data Stage* (da *Ardent/Informix*), o DTS (da *Microsoft*), *Sagent* (da própria *Sagent*) e ETI [ETI], com a finalidade de auxiliar na execução da ETL, esse ainda é um processo trabalhoso, complexo e cheio de detalhes, sendo responsável por mais de 50% do tempo gasto no desenvolvimento de um DW, e conseqüentemente, por ser o mais caro.

A seguir será descrito cada uma das fases do processo de ETL.

#### Extração

Inmon[INM96] define alguns mecanismos para se fazer extração de dados no desenvolvimento de DW. Isso envolve uma análise de quais dados (por exemplo, nome, en-

dereço, valor do pagamento, entre outros), seus tipos (por exemplo, *char*, *integer* e *date*) e, finalmente, sua localização. Note que a origem dos dados, tal como os formatos, pode ser bem variada. Isto porque pode-se encontrar desde sistemas transacionais das empresas até planilhas, arquivos textos e também arquivos DBF (*dBase*) ou do *Microsoft Access*.

Esta fase compreende um conjunto de rotinas implementadas, cuja finalidade é fazer a transferência dos dados dos locais de origem para o DW. Entretanto, um processo bem interessante e muito usado é a transferência dos dados para a DSA (*Data Stage Area*).

A DSA corresponde a uma área de armazenamento intermediário, cujo objetivo é facilitar a integração dos dados antes que sejam inseridos no DW. Essa facilidade consiste na redução da complexidade dos algoritmos de extração e transformação de dados mediante a divisão desse processo em duas etapas: a primeira responsável pela migração dos itens de dados das fontes para a DSA; e a segunda responsável pela transferência dos dados da DSA para o DW.

Os dados permanecem na DSA temporariamente, até que passem pelas devidas transformações, tais como integração, sumarizações e agrupamentos, ou até que o desenvolvedor possa validá-los no sentido de verificar se os dados extraídos são exatamente os requeridos. Ressalta-se que integrar, limpar e transformar os dados na carga da DSA ou a partir desta, ou até mesmo o uso ou não dela, é uma decisão de projeto.

## Transformação

Esta etapa está, geralmente, composta de outras sub-etapas, quais sejam: integração, limpeza e transformação dos dados.

- **integração** - Considerando que as fontes de dados podem ser bastante diversificadas, o processo de integração (ver subseção 3.2.2), é responsável por formatar os dados num único padrão. Por exemplo, se um elemento de dado é medido em centímetros em uma aplicação, em polegadas em outra, ele será convertido para uma representação única ao ser colocado no DW;
- **limpeza** - é necessária porque os dados, normalmente, advêm de fontes, em alguns casos, desconhecidas ou antigas, podendo conter além de dados inconsistentes, dados que foram inseridos de forma fictícia. Por exemplo: Uma empresa Administradora de Cartão de Créditos, na qual o vendedor está mais preocupado em vender o produto (cartão) que na qualidade de dados que está inserindo. Se o cliente não tiver o número do CPF na hora da venda, o vendedor cadastrará um número qualquer para agilizar a venda. Se for feita uma consulta posterior, levando-se em conta o número do CPF dos clientes, no mínimo informações estranhas aparecerão (algo como CPF número 99999999-99). Por isso, nessa fase, faz-se a limpeza desses dados, ou seja, é feito um tratamento dos mesmos para haver compatibilidade entre eles;

- **transformação** - é claro que quando os dados estão sendo integrados ou limpos, indiretamente eles estão sendo transformados. Dessa forma pode-se afirmar que toda integração e limpeza são uma transformação, mas o inverso não é verdadeiro. Uma transformação pode ser vista também como, por exemplo, um cálculo sobre um dado.

## Carga

Essa etapa consiste na transferência real dos dados das fontes para o DW. A carga dos dados é feita baseada em um período pré-definido.

Apesar da existência de ferramentas de ETL predefinidas, às vezes são necessárias criações de rotinas de carga para atender determinadas situações específicas que podem ocorrer. Todas têm os seus diferenciais e cada uma pode ser utilizada dependendo do caso de cada empresa. O mais importante é considerar que uma ferramenta de ETL tem grande valia, principalmente se os sistemas OLTP são muitos, pois elas constituem uma poderosa fonte de geração de metadados, e contribuem muito para a produtividade da equipe.

## 3.5 Data Warehouse Geográfico

A maioria das empresas geralmente possui em seus bancos de dados alguns atributos pertinentes a uma referência geográfica ou espacial, tais como endereços de fornecedores, distribuição de filiais, controle de rotas, entre outros. Dessa forma, o grau de relevância da informação georreferenciada e analisada espacialmente se torna cada vez maior na arquitetura de DW.

O componente espacial de um DW pode ser assimilado de três formas. Primeiramente, é necessária uma ferramenta que execute as agregações espaciais e as conversões geográficas (atribuição de coordenadas ou códigos que estabeleçam a referência espacial). Em seguida, um banco de dados de apoio ao DW deve ser modelado a partir de uma combinação entre as características de um banco de dados espacial e de um banco de dados voltado para DW, incluindo suporte a multidimensionalidade e escalabilidade. E finalmente, a análise estatístico-espacial dos dados e a apresentação alternativa dos resultados em mapas ou tabelas [7200].

Como se vê, as duas últimas possibilidades são viabilizadas somente se existir um suporte de SGBD adequado e uma ferramenta OLAP de apuração de dados contemplada com muitas facilidades de SIG. Por outro lado, ao se considerar um SIG que possa acessar uma base de dados modelada como DW, sentir-se-á falta da flexibilidade que aquele tipo de ferramenta possui para tratar dados analíticos no mesmo nível que uma ferramenta OLAP. Assim, as duas ferramentas, SIG e OLAP, se ressentem da falta de integração,

seja no ambiente de DW ou para uso geral no suporte à decisão. A apresentação de um produto denominado DWG ou Sistema de Informações Geográficas Orientado para DW passa obrigatoriamente pela composição dessas ferramentas, cujo sucesso depende do grau de transparência disponibilizados ao usuário e dos recursos oferecidos.

Deve-se considerar que um objetivo fundamental, será sempre permitir que sejam feitas consultas ao DW a partir dos atributos geográficos, uma análise espacial destes resultados, habilitar refinamentos sucessivos e agregar os resultados em áreas geográficas visualizadas instantaneamente [BER97].

## 3.6 Metodologia de Desenvolvimento de DWG

Os DW convencionais possuem diferentes metodologias de desenvolvimento defendidos por vários autores, embora ainda não exista nenhuma realmente consolidada.

Tal como os DW convencionais, os DWG também não possuem uma metodologia consolidada. Na verdade, não foi encontrada nenhuma metodologia destinada a este tipo de DW.

Portanto, a seguir serão descritas algumas considerações sobre o desenvolvimento de DW, as quais são baseadas no trabalho de Jiawei Han [HSK98].

### 3.6.1 Projeto Lógico de um DWG

Para modelar um DW espacial, o esquema do modelo estrela é ainda considerado como sendo uma boa escolha porque provê uma estrutura de DW concisa e organizada, além de facilitar as operações de OLAP e navegação. Todavia, a aplicação do esquema estrela em sua forma original pode gerar uma certa ineficiência na execução de um OLAP espacial. Dessa forma faz-se necessário um modelo diferente que possa atender às características de dados geográficos.

Como exemplo, suponha a existência de milhares de sondas meteorológicas espalhadas em uma determinada região, cada uma gravando a temperatura e precipitação diárias de uma determinada área pequena e transmitindo sinais para uma estação meteorológica estadual. Um usuário poderia querer ver padrões meteorológicos em um mapa por mês, por região e por diferentes combinações de temperatura e precipitação, ou até mesmo executar *drill-down* (operação que parte da análise dos dados mais resumidos para os mais detalhados) e *roll-up* (operação inversa ao *drill-down*) em qualquer dimensão para explorar padrões desejados, tais como regiões úmidas ou secas em um determinado mês.

Cada estação meteorológica está associada com uma área do mapa. É provável que áreas vizinhas tenham padrões meteorológicos similares. Além disso, quando um usuário executa *roll-up* em dimensões, a probabilidade de ter as mesmas descrições para áreas vizinhas aumenta. Conseqüentemente, alguém poderia obter uma quantidade de áreas compostas (agregadas). De uma perspectiva de decisão de suporte, estas regiões agregadas são medidas do negócio. Note que este tipo de medida é muito diferente das medidas tradicionais em DW tais como, total de vendas, quantidade de produtos, entre outras. Além disso, um usuário poderia querer executar operações OLAP diretamente em áreas do mapa. Nada disso poderia ser feito aplicando métodos convencionais de OLAP, pois eles lidam somente com medidas que são agregações numéricas.

A agregação de área é apenas uma das razões para explorar o novo modelo. Existe uma variedade de aplicações que lidam com múltiplos mapas temáticos, e a análise *on-line* de tais mapas envolve operações freqüentes de sobreposição. As áreas resultantes destas sobreposições são tratadas como medidas. Além disso, mapas podem conter objetos espaciais de diferentes tipos, e a sobreposição de tais objetos geralmente tem de ser calculado também. Note que em todos estes casos, um usuário pode querer executar *drill-down/roll-up* tanto em dimensões espaciais quanto em não espaciais.

Até agora foi apresentado apenas o modelo de um DWG. A seguir será apresentado que tanto dimensões quanto medidas podem conter componentes espaciais. Além disso, um cubo de dados espacial pode ser construído de acordo com as dimensões e medidas modeladas no DWG.

### 3.6.2 Dimensões em um DWG

Em um DWG podem ser identificados três tipos de dimensões:

- 1) **dimensão não-espacial** - consiste em uma dimensão contendo somente dados não espaciais. Por exemplo, duas dimensões não-espaciais, temperatura e precipitação, podem ser construídas para o DW do exemplo citado, cada um é uma dimensão contendo dados não espaciais cujas generalizações são também não-espaciais, como quente e úmido;
- 2) **dimensão espacial para não-espacial** - é uma dimensão que, no nível primitivo de dados, é espacial mas que em sua generalização, a partir de um nível mais alto, torna-se não-espacial. Como exemplo, a dimensão *Município* começaria espacial no nível primitivo; no nível imediatamente superior a dimensão *Temperatura Média* se tornaria não-espacial e assim continuaria na hierarquia com a dimensão *Clima*.
- 3) **dimensão espacial para espacial** - é uma dimensão em que o nível primitivo e todos os seus níveis mais altos de dados generalizados são espaciais. Por exemplo, as dimensões *Município*, *Estado* e *Região*, nesta ordem hierárquica, representam valores espaciais em todos os níveis;

Note que os dois últimos tipos de dimensão indicam que um atributo espacial, como município, pode ter mais que uma maneira de serem generalizados para conceitos de nível mais alto, e os conceitos generalizados podem ser espaciais, como mapas representando estado ou regiões, ou não espaciais, como área ou descrição geral da região.

### 3.6.3 Medidas em um DWG

Quanto as medidas, existem dois tipos em um DWG:

- 1) **medida numérica** - contém somente dados numéricos. Por exemplo, uma medida em um DWG poderia ser a revisão mensal de uma região, e a execução de um *roll-up* pode dar a revisão total por ano, por condado, entre outras. Medidas numéricas podem ser classificadas em distributivas, algébricas e holísticas. Uma medida é distributiva se pode ser computada por partição de cubo e agregação distribuída, como quantidade, soma, máximo; é algébrica se pode ser computada por manipulação algébrica de medidas distribuídas, como desvio padrão médio; e é holística se não existir um limite constante no tamanho do espaço de armazenamento necessário para descrever um sub-agregado, como mediano, mais frequente e *rank*;
- 2) **medida espacial** - contém uma coleção de ponteiros para objetos espaciais. Por exemplo, durante a generalização (ou *roll-up*) de um cubo de dados espacial do exemplo citado, as regiões com o mesmo intervalo de temperatura e precipitação são agrupadas na mesma célula, a medida então formada contém a coleção de ponteiros apontando para estas regiões. Uma medida computada pode ser usada com dimensão num DW, denominada medida consolidada. Por exemplo, a medida temperatura média mensal em uma região pode ser tratada como uma dimensão e pode ser mais tarde generalizada em um valor de intervalo ou descritivo, como frio. Além disso, uma dimensão pode ser especificada por especialistas/usuários baseada nos relacionamentos entre atributos ou valores de dados particulares, ou pode ser gerada automaticamente baseada nas técnicas de análise de dados espaciais, como agrupamento espacial, classificação espacial, ou análise de associação espacial.

Conforme descrito anteriormente, não existe nenhuma metodologia consolidada para a construção de um DWG. Dessa forma, o trabalho de Jiawei Han [HSK98] apresenta considerações sobre o projeto lógico, discussões e medidas, as quais constituem um auxílio para os desenvolvedores de DWG.

O Trabalho de Jiawei Han embora não represente nenhuma metodologia formal, proporciona aos desenvolvedores de DWG alguns aspectos que podem ser bastante úteis. Particularmente neste trabalho, foram utilizados os conceitos por ele definidos, principalmente pela limitação de material na literatura corrente.

O capítulo a seguir descreve algumas metodologias de desenvolvimento de DW, em que, a partir de algumas etapas das mesmas somadas às considerações de Jiawei Han entre outros aspectos, permitiram a definição da estratégia de desenvolvimento de um DWG, principal objetivo deste trabalho.

# Capítulo 4

## Metodologias para Construção de Data Warehouse

Indicar a melhor metodologia para se desenvolver um DW é uma difícil tarefa. Vários autores [K<sup>+</sup>98, P<sup>+</sup>98, DAL99] estabeleceram suas estratégias de desenvolvimento, abrangendo os mais diversos aspectos tais como: granularidade, particionamento, arquitetura, entre outros. Entretanto, ainda não há uma metodologia consagrada no meio acadêmico.

Este capítulo aborda, inicialmente, um resumo das metodologias definidas por Ralph Kimball [KIM96], John A. Zachman [IZG97] e Poe [P<sup>+</sup>98] *apud* [PER00] quanto ao desenvolvimento de um DW. Em seguida serão feitas algumas considerações sobre as metodologias apresentadas, enfatizando-se as vantagens e desvantagens, bem como as limitações de cada uma.

Importante salientar que não é escopo deste trabalho definir a melhor ou a mais adequada das metodologias, e sim discutí-las para que, a partir das etapas mais relevantes das mesmas, se possa definir uma estratégia para o desenvolvimento de um DWG.

### 4.1 Metodologia de Ralph Kimball

Kimball sugere em [KIM96] algumas etapas para a criação de um DW. Estas etapas não representam as fases de seu ciclo de vida [K<sup>+</sup>98], mas descrevem a “receita” para a sua construção, o que justifica incluí-las neste capítulo:

#### 4.1.1 Etapas do Desenvolvimento

Segundo Kimball [KIM96] existem nove etapas básicas para a construção de um DW, cada uma podendo ter sub etapas, cujas descrições encontram-se incluídas no referido trabalho.

**Etapas 1: Identificar os processos a serem modelados** Refere-se à identificação dos temas a serem modelados. Cada tema associado a, no mínimo, uma tabela de fatos na constelação<sup>1</sup> do DW.

---

<sup>1</sup>Múltiplas tabelas de fatos que compartilham tabelas dimensão, formando um grupo de estrelas.

## Etapa 2: Definir a granularidade de cada tabela de fatos para cada processo

Nesta etapa é especificado o nível de detalhe a ser representado pelos fatos. A granularidade da tabela de fatos não deve chegar a níveis muito detalhados, a ponto de extrapolar a capacidade de armazenamento do DW. Para verificar a sua viabilidade, pode-se estimar o seu tamanho, após um número de períodos, usando-se um dos fatos da tabela na seguinte expressão:

$$TamanhoDW(NumeroPeriodos) = \frac{\sum ValoresFatoPeriodo}{ValorMedioFatoPeriodo} \times NumeroPeriodos$$

**Etapa 3: Definir as dimensões de cada tabela de fatos** O DW deve conter dimensões que, ao serem cruzadas, ajudem o usuário na tomada de decisões<sup>2</sup>. Cada dimensão é descrita ou identificada por atributos em uma tabela específica, denominada tabela de dimensão. Se a dimensão apresenta uma hierarquia, esta é descrita em sua tabela e, possivelmente, por outras associadas aos demais níveis de seu detalhamento.

A criação de mais de uma tabela para representar a estrutura hierárquica da dimensão<sup>3</sup> não é recomendada por Kimball, que argumenta que a normalização das tabelas de dimensão, ou *snowflaking*, pouparia espaço desprezível em disco, quando comparado ao volume de dados da tabela de fatos, além de tornar as consultas ao DW mais complexas e lentas.

Por vezes, é possível ainda que uma dimensão apresente outras hierarquias além daquela explicitamente descrita. Essas *hierarquias embutidas* podem usar atributos de outras tabelas inclusive, o que não vale o esforço de separá-las em outras dimensões. Contudo, ferramentas de OLAP devem ser capazes de navegar por elas também.

## Etapa 4: Identificar os fatos

Nas dimensões com hierarquias, os fatos associados devem ser, preferencialmente, representados no menor nível de detalhe possível. Neste nível, o valor do fato pode ser identificado como *aditivo*, *semi-aditivo* ou *não aditivo*:

- Um *valor aditivo* é aquele em que pode ser aplicada uma função de agregação<sup>4</sup> a medida em que se sobe na hierarquia da dimensão;

---

<sup>2</sup>O tempo, embora nem sempre explícito como dimensão, é importante para manter o histórico de fatos significativos, o que permite prever alguns resultados a partir de estudos do passado e, conseqüentemente, acaba se tornando também uma informação importante para o processo decisório.

<sup>3</sup>Uma tabela para cada nível de detalhe, relacionada a outra de nível imediatamente superior através de associações 1:N.

<sup>4</sup>Somatório, média, máximo e mínimo entre outras funções agregadoras.

- *valores semi-aditivos* são valores que se agregaram a outros já contabilizados no DW, de tal maneira que uma nova agregação envolveria a contagem dupla de algum fato em seu resultado. O *saldo mensal* é um exemplo prático deste tipo de valor, uma vez que o seu somatório, dentro de um período maior na hierarquia como o trimestre, implicaria na recontagem de ítems que apareceram em mais de um mês;
- *valores não aditivos* são obtidos de cálculos que tornam sem sentido as suas agregações em qualquer nível de uma hierarquia. A exemplo de *cálculos percentuais*, não há coerência em somá-los quando se sabe que o percentual do total das grandezas envolvidas é diferente do total dos percentuais calculados individualmente. Valores estáticos que medem algum nível de intensidade também são identificados como não aditivos.

Os valores semi-aditivos e os não aditivos, por não poderem ser agregados como os aditivos, são limitados em algumas operações. Mesmo assim, Kimball defende que devem ser mantidos no DW por serem úteis em futuras manipulações e para economizar espaço na tabela de fatos.

### **Etapa 5: Analisar os atributos das dimensões, de modo a estabelecer descrições completas e terminologia apropriada**

A dimensão mais importante em um DW deve, geralmente, possuir muitos atributos, os quais aparecem geralmente nos rótulos de relatórios emitidos pelo sistema. No caso das demais dimensões de negócio, algumas questões precisam ser consideradas:

- mapear novamente as chaves para evitar a dependência da chave original, que pode ter sido reutilizada em entidades diferentes em ambientes OLTP, ou para reduzir chaves longas;
- gerar chaves genéricas para permitir alterações nas descrições das entidades sem que sejam refletidas nas suas chaves;
- prever chaves genéricas para associar tabelas de fatos em uma constelação;
- substituir códigos por descrições textuais ou associá-los a textos descritivos não abreviados;
- evitar variações em valores de campos textuais que deveriam ser iguais, bem como a igualdade destes valores quando estes deveriam ser diferentes.

## Etapa 6: Decisões sobre projeto físico

Neste momento, algumas questões inerentes ao projeto físico são consideradas:

- combinar duas dimensões numa só quando o relacionamento entre elas for predominantemente 1:N. Nos demais casos, quando o relacionamento é N:M, é preferível mantê-las em tabelas separadas;
- usar uma tabela para armazenar as numerosas tuplas de uma dimensão, indexando preferencialmente os seus atributos demográficos, de forma a manter o desempenho de navegação e a simplicidade da interface com o usuário;
- combinar os atributos que são muito usados em cruzamentos em grupos de 5 ou 6 e indexando-os separadamente, gerando um conjunto reduzido de tuplas, o qual desconsidera combinações improváveis e agrupa valores contínuos em intervalos. Cada grupo constitui uma *minidimensão* que não necessariamente substitui os seus atributos originários, os quais podem permanecer em suas respectivas dimensões para análises mais detalhadas, e nem tampouco exige que eles obedeçam um agrupamento lógico;
- pré-computar sumarizações em uma agregação de forma a otimizar o desempenho das aplicações de DW;
- Sumarizar, nas agregações, mais de 10 ou 20 ítems do nível inferior para evitar que o uso de agregações estoure a capacidade do DW. Isto porque, uma agregação pode ser implementada criando-se uma nova tabela de fatos agregados ou inserindo-se novos campos de indicação de nível de agregação nas tabelas de dimensão. Na primeira opção, o conjunto dos fatos agregados reside em outra tabela de fatos, enquanto que na segunda, os fatos originais e os agregados compartilham a mesma tabela, o que pode causar problemas de dupla contagem.

## Etapa 7: Preparar dimensões para suportar evoluções

Alguns atributos de dimensão, em geral os descritivos, evoluem lentamente ao longo do tempo. Quando modificados, podem perder informação do passado ao qual estavam vinculados.

A exemplo da dimensão pessoa, um determinado indivíduo que tenha evoluído de solteiro para casado perderia informações históricas se estas estivessem classificadas pelo seu estado civil. Neste caso, existem 3 possíveis soluções:

- modificar o atributo com o novo valor, sem manter o histórico da sua evolução;
- acrescentar, para o mesmo indivíduo, um novo registro com o novo valor na tabela de dimensões. Esta solução requer o acréscimo de um identificador de versão na chave da tabela, estabelecendo um vínculo com os fatos relacionados na linha do tempo.

- manter sempre o valor inicial e o valor atual no mesmo registro, junto com a data em que este último valor entrou em vigor. Neste caso, não há a necessidade de acrescentar um novo registro toda vez que uma atualização acontece.

A primeira solução é a mais simples e é geralmente recomendada para a correção de atributos. A terceira tem o controle mais complexo das três no que diz respeito à implementação dos programas, mas evita o armazenamento das várias tuplas que seriam criadas usando a segunda solução. Partindo destas considerações é, portanto, decisão do implementador escolher qual delas, ou mesmo qual combinação delas, resolverá o problema. Kimball adverte que, independente da solução adotada, ela não deve ser complexa sem que haja uma boa justificativa para o seu uso.

### **Etapa 8: Definir a previsão do histórico**

Esta fase visa determinar o intervalo de tempo em que os dados permanecerão armazenados no DW.

### **Etapa 9: Definir a frequência da carga do DW**

Nesta fase deve-se determinar qual a periodicidade com que os dados serão inseridos no DW.

As etapas discutidas até agora foram aplicadas, total ou parcialmente, em estudos de casos exemplificados por Kimball. Em cada um deles foram consideradas as suas peculiaridades e questões relativas aos seus problemas, além das soluções adotadas para solucioná-los. Apesar de não ser explícito, o autor não considera que estas etapas sejam seguidas em ordem e nem tampouco que uma etapa deva começar quando a anterior tenha sido terminada, mas apenas que elas sejam suficientes para a realização de praticamente todos os projetos envolvendo a construção de um DW.

Ressalta-se, que Kimball em [K<sup>+</sup>98] define um ciclo de desenvolvimento diferente do apresentado neste trabalho. Isto porque a metodologia aqui descrita tem como principal enfoque a construção do DW em si, enquanto que a descrita em [K<sup>+</sup>98] apresenta um ciclo de desenvolvimento como um todo, abrangendo etapas que incluem desde atividades de planejamento até as de manutenção e evolução.

## **4.2 Proposta de Zachman**

A proposta de Zachman [IZG97] descreve um esquema de classificação que permite desenvolver uma estratégia geral para o desenvolvimento de um DW.

O esquema se baseia na construção de uma matriz de classificação, na qual as linhas representam perspectivas (ou visões) e as colunas dimensões, que descrevem abstratamente como um produto deve ser observado segundo essas perspectivas. Sendo que o produto deve ser entendido como o resultado conceitual de um esforço de desenvolvimento, que necessite de uma abordagem evolutiva e disciplinada desde a sua concepção até a sua conclusão.

Zachman sugere que, para se definir completamente um produto, deve-se identificar as seguintes perspectivas:

- **planejador** - Envolve propostas de soluções alternativas ou melhoria do processo de negócio de acordo com as informações coletadas, escolha da equipe de analistas e projetistas e análise dos riscos de cada solução. Ao final define o escopo do projeto, identifica seus componentes principais e traz informações sobre os custos e benefícios de cada solução.
- **proprietário** - Enfatiza a praticidade do produto, identificando as limitações quanto ao seu uso de acordo com o escopo descrito na perspectiva anterior.
- **projetista** - Traduz as descrições em especificações técnicas do produto, segundo as limitações de uso e escopo identificados pela perspectiva anterior.
- **construtor** - Aborda a construção do produto a partir de um roteiro de atividades, que descreve a sequência de passos e as restrições impostas durante o desenvolvimento do DW. Como resultado, um plano de construção é elaborado contendo especificações de componentes para o subcontratado e instruções sobre o emprego de determinada tecnologia ou material, entre outros.
- **subcontratado** - Responsável pela execução da construção das partes específicas do produto a partir do plano elaborado pela perspectiva anterior.

Para cada perspectiva são feitas as seguintes interrogativas básicas: o quê, onde, como, por quê, quando e quem. Estas interrogativas representam respectivamente as dimensões entidades, localização, atividades, motivação, tempo e pessoas e abstraem a forma como um produto deve ser observado:

- **entidades** (o quê?) - são os itens de interesse, como dados de entrada e saída de um sistema de informação ou classe de objetos e itens de negócio que serão implementados, como vendas, pedidos e outros;
- **localizações** (onde?) - são os locais onde os componentes de um projeto devem ser armazenados ou utilizados;
- **atividades** (como?) - correspondem às descrições dos processos de transformação dos dados de entrada e às funções de suporte para as atividades operacionais do usuário;

- **motivação** (por que?) - descreve as razões que motivam a construção do novo sistema. Também pode descrever os resultados esperados com a sua implantação;
- **tempo** (quando?) - determina a seqüencia dos eventos e as suas interdependências.
- **pessoas** (quem?) - são os indivíduos, departamentos e organizações de interesse.

A estratégia de Zachman consiste em aplicar este esquema em um ciclo de desenvolvimento do DW de forma consistente, enfatizando o aspecto do dado de uma maneira menos ortodoxa e mais voltada ao objetivo final de um DW.

No ciclo de vida são identificadas as seguintes fases principais:

- 1) **confirmação para prosseguir com o projeto** - Começa pelo problema especificado, identificando na área de negócios aquilo que o sistema atual não atende e que poderia ser solucionado por um DW. Nesta fase é negociado o suporte financeiro para a execução do projeto e são elaboradas justificativas para este investimento, a partir dos benefícios que seriam alcançados ao adotar a solução proposta. O resultado desta fase é a definição do escopo do projeto, que contempla o modelo que irá retratar os negócios da organização e a identificação de um conjunto de requisitos;
- 2) **análise do Modelo de Dados** - Analisa as atividades atuais que suportam a tomada de decisões através da coleta de informações, tais como consultas e relatórios providos pelos sistemas atuais. Esta fase produz o modelo de negócios, que inclui o modelo de dados contendo os itens de negócio e os relacionamentos entre eles.
- 3) **breadbox analysis** - Estima o tamanho do DW para que sejam considerados o seu nível de granularidade, a necessidade de particionar os seus dados em diferentes servidores e os seus diferentes níveis de sumarização;
- 4) **avaliação técnica** - Identifica os servidores em que os dados serão armazenados e os locais de onde grupos de usuários irão acessar esses dados. Além disso, considera questões como suporte de rede, tempo de extração e carga do DW e segurança dos dados.
- 5) **análise das áreas de negócio de interesse** - Executa revisões do escopo, baseada nas restrições tecnológicas surgidas nas fase anteriores, e do modelo de dados para acomodar as alterações feitas nos requisitos dos processos de negócios;
- 6) **análise dos sistemas fontes** - Analisa os dados distribuídos pela organização para integrá-los segundo a perspectiva histórica de interesse.
- 7) **Preparação técnica do ambiente** - Define a arquitetura de hardware em função das deficiências do ambiente computacional levantadas na fase de *avaliação técnica*;
- 8) **projeto de DW (esquema)** - Define um esquema baseado nos modelos de dados desenvolvidos para cada área relacionada às decisões de negócio, as quais terão suporte no DW.

- 9) **especificação dos programas para transformação e integração dos dados** - especifica os programas necessários tanto na carga inicial como nas de atualização (ETL);
- 10) **codificação dos programas** - corresponde às atividades de implementação dos programas especificados na fase anterior, incluindo os testes e a definição do BD possivelmente distribuídos em vários servidores;
- 11) **carga no DW** - insere os dados no DW através dos programas de carga inicial e atualização implementados na fase anterior.

Cada uma das cinco perspectivas ou visões agrupa um subconjunto das fases definidas, observando-se nelas, quando aplicáveis, todas as dimensões já descritas. A Tabela 4.1 ilustra a aplicabilidade das dimensões em cada fase das perspectivas.

**Tabela 4.1. Matriz de Perspectivas e Dimensões de Zachman**

Perspectivas	Fases	Entidades	Localizações	Atividades	Motivação	Tempo	Pessoas
Planejador	Fase 1	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável
Proprietário	Fase 2	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável
Projetista	Fase 3	Aplicável	Aplicável	N/A	N/A	N/A	N/A
	Fase 4	N/A	Aplicável	N/A	Aplicável	Aplicável	Aplicável
	Fase 5	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável
Construtor	Fase 6	N/A	Aplicável	Aplicável	N/A	Aplicável	Aplicável
	Fase 7	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável
	Fase 8	N/A	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável
	Fase 9	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável
Subcontratado	Fase 10	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável
	Fase 11	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável	Aplicável

Terminada a matriz de classificação, tem-se uma visão geral para estabelecer uma estratégia que contemple cada fase do ciclo de vida do desenvolvimento do DW.

Zachman, contudo, adverte que o preço para se construir um esquema completo, tanto em perspectivas quanto em dimensões, é o provável descrédito da Organização com

relação a implantação do DW, uma vez que é exigido um tempo considerável para realizar os levantamentos necessários ao preenchimento de cada célula da matriz.

Uma alternativa menos cara e mais rápida seria a abordagem “*botom-up*”, no qual o DW é construído deixando-se evoluir naturalmente. Entretanto, é aconselhável usar esta abordagem para sistemas menores, em que o levantamento dos diferentes grupos de usuários não é tão relevante, ou como protótipo, para mostrar a evolução de todo o DW.

### 4.3 Metodologia de Poe

O trabalho de Poe [P<sup>+</sup>98] abrange, de forma completa e detalhada, toda a metodologia de desenvolvimento de um DW, enfatizando a importância de se estabelecer, antes da construção do DW, uma arquitetura de dados adequada, que realmente atenda às necessidades da corporação. A grande importância dessa abordagem quanto ao estabelecimento inicial de uma arquitetura de dados deve-se ao fato de que, em decorrência desta escolha, derivar-se-á a infra-estrutura técnica necessária para atender o DW, normalmente composta por tecnologias e itens tais como treinamentos em tecnologias de suporte à decisão, plataformas, bases de dados, ferramentas de conversão de dados, *hardware*, *software*, rede local e ferramentas de acesso aos dados. A arquitetura de dados está intimamente relacionada com a infra-estrutura, ou seja, os componentes e tecnologias da infra-estrutura a serem utilizados dependerão diretamente da arquitetura de dados escolhida. Em sua metodologia, Poe denomina as diversas fases que compõem o desenvolvimento do DW de “ciclo de vida de suporte à decisão”.

#### 4.3.1 Ciclo de Vida de Suporte à Decisão

As metodologias até então apresentadas descrevem especificamente as etapas de construção de um DW. A proposta de Poe, ao contrário, descreve etapas as quais compreendem fases relacionadas a atividades de pré-construção, ou seja, ao invés de restringir-se somente ao processo de desenvolvimento de um DW, Poe optou por uma metodologia mais completa e abrangente.

A metodologia de Poe abrange dez etapas, e embora não seja o enfoque deste trabalho abordar metodologias que não sejam exclusivas ao processo de desenvolvimento de um DW, essa metodologia foi, particularmente, escolhida com o objetivo de se demonstrar o desenvolvimento como um todo.

A seguir será descrito de forma concisa cada uma das etapas definidas por Poe.

## 1. Planejamento

A fase planejamento corresponde às seguintes atividades:

- definição do escopo do projeto;
- criação de um plano de projeto;
- definição dos recursos técnicos necessários (internos e externos);
- definição dos participantes do processo e suas responsabilidades;
- definição de tarefas e prazos de conclusão; e
- definição de cronogramas, incluindo o prazo final de entrega do projeto.

Segundo Poe não é conveniente construir um DW antes de estabelecer corretamente a arquitetura e a infra-estrutura, o que poderá acarretar insucesso no seu desenvolvimento e uso. Desta forma, caso a infra-estrutura não esteja pronta, os seguintes itens deverão também fazer parte do planejamento do DW: tecnologia LAN/WAN (*Local Area Network/Wide Area Network*), plataforma selecionada, conectividade com a base de dados, base de dados e seus utilitários, estações de trabalho devidamente configuradas e ferramentas de acesso aos dados.

Outras considerações técnicas, apesar de não fazerem parte da infra-estrutura, deverão ser incluídas no plano de projeto, tais como:

- estratégias de integração, arquivamento, reconsulta e atualização de dados: operação e agendamento de trabalhos e gerenciamento de metadados; e
- procedimentos para o usuário final acessar os dados arquivados.

## 2. Levantamento de requisitos e modelagem

Esta fase abrange dois tópicos principais: primeiro, o entendimento do negócio, exigências e necessidades dos usuários, e por último, a modelagem do universo de discurso, do qual fazem parte as referidas exigências e necessidades.

A seguir, essa fase será apresentada em duas etapas distintas: levantamento de requisitos e modelagem. Na primeira são realizadas atividades diversas tais como entrevistas, seções de facilitação e estudo de documentos e relatórios, de modo que seja possível a identificação dos seguintes pontos:

- como o usuário realiza as diversas atividades de negócio, quais os fatores que dirigem o negócio, quais atributos o usuário necessita e são absolutamente requeridos e quais são desejados;

- quais as hierarquias;
- quais dados os usuários utilizam e quais deveriam utilizar;
- qual o nível de detalhes ou sumário que o usuário necessita;
- qual tipo de ferramenta de acesso aos dados o usuário utilizará; e
- como o usuário espera ver o resultado de suas consultas.

A modelagem de dados refere-se ao processo de tradução dos conceitos do negócio em um modelo dimensional em formato diagramático que inclua fatos, dimensões, hierarquias, relacionamentos, chaves candidatas, entre outros. Parte dos componentes deste modelo dimensional futuramente se transformarão em objetos físicos na base de dados analítica, possibilitando o desenvolvimento do projeto de DW.

Poe apresenta em seu trabalho uma abordagem bastante interessante quanto à modelagem dimensional. Além dos tradicionais esquemas estrela e flocos de neve, normalmente adotados em modelagem dimensional, apresenta diversas variações, tais como: esquema estrela com múltiplas tabelas fato, esquema estrela com tabelas associativas, esquema estrela com tabelas externas e esquema multiestrela; defendendo a idéia de que projetos de DW não devem ser modelados seguindo um único modelo, mas que, caso seja necessário, pode vir a representar a adoção de modelos e técnicas híbridas, de modo a se obter os melhores resultados.

### **3. Projeto físico da base de dados e desenvolvimento**

Esta fase contempla as seguintes atividades:

- projeto e criação de objetos apropriados na base de dados, incluindo tabelas fato, tabelas dimensão, relacionamentos entre tabelas, entre outros;
- desnormalização de dados;
- estratégias de criação de índices;
- desenvolvimento de estratégias de agregação; e
- desenvolvimento de estratégias de particionamento.

#### **4. Determinação, integração e mapeamento das fontes de dados**

Essa fase é a que consome mais tempo de desenvolvimento, devido à necessidade de localizar os dados adequados dispersos em sistemas OLTP, analisar e entender os tipos de dados, implementar os processos de transformação necessários e mapear os campos das fontes de dados para os objetos da base de dados. Durante esta fase deve-se:

- identificar as possíveis fontes de dados e suas respectivas estruturas;
- executar a análise de dados para determinar as melhores fontes de dados e os processos de integração necessários;
- desenvolver programas que permitam realizar as conversões de dados para cada campo e refinar a estratégia de integração; e
- mapear as fontes de dados para os objetos da base de dados analítica.

#### **5. População do DW**

Embora o nome dessa fase dê a idéia de carga de dados, ela corresponde a definição e desenvolvimento de programas e rotinas, e até mesmo utilização de ferramentas ETL, para efetuarem o processo de extração, transformação e integração dos dados. Engloba as seguintes atividades:

- desenvolvimento de programas ou utilização de ferramentas para extrair e mover os dados;
- desenvolvimento de estratégias e procedimentos de carga de dados dentro do DW;
- desenvolvimento de programas ou utilização de ferramentas de conversão de dados de forma a integrá-los;
- desenvolvimento de estratégias que permitam consultar e atualizar dados;
- execução de programas e procedimentos de extração, transformação e carga de dados;
- desenvolvimento e realização de processos de validação e teste dos dados extraídos, transformados e carregados no DW; e
- desenvolvimento de procedimentos que permitam o tratamento de exceções, geração de estatísticas e garantia da qualidade dos dados.

É importante, nessa etapa, dedicar uma atenção especial em localizar e tratar adequadamente os tipos de dados sinônimos, homônimos e análogos, cuja identificação inexata acarretará baixa qualidade dos dados do DW ou até mesmo o fracasso do projeto por não proporcionar credibilidade aos seus usuários.

## **6. Automação dos processos de carga de dados**

Nessa fase são executadas as seguintes tarefas:

- automação e agendamento dos processos de extração, conversão e carga de dados;
- criação de procedimentos de *backup* e recuperação de dados; e
- realização de teste completo de todos os procedimentos automatizados.

## **7. Criação de conjunto inicial de relatórios**

Corresponde a realização das seguintes atividades:

- criação de conjunto de relatórios pré-definidos;
- teste dos relatórios;
- documentação das aplicações; e
- desenvolvimento de estruturas de navegação dos relatórios.

## **8. Validação e teste de dados**

A fase de validação e testes de dados compreende:

- validação de dados utilizando o conjunto inicial de relatórios;
- validação de dados utilizando processos padronizados; e
- modificações interativas dos dados.

## **9. Treinamento**

A fase de treinamento centra-se na criação de programas de treinamento para atender especificamente a comunidade de usuários e contempla as seguintes atividades:

- treinamento sobre o escopo do DW;
- treinamento nas ferramentas de acesso aos dados e de como acessar e navegar através dos metadados, para conseguir as informações desejadas no DW;
- treinamento em aplicações de suporte à decisão ou no conjunto de relatórios disponibilizados aos usuários finais; e
- contínuo treinamento e assistência aos usuários.

## 10. Produção

A última fase, denominada produção, inclui as tarefas necessárias para a disponibilização do DW e o correspondente suporte necessário aos usuários finais, podendo incluir as seguintes atividades:

- instalação de infra-estrutura física a todos os usuários (por exemplo, rede de comunicações, conectividade com a base de dados e configuração de estações de trabalho);
- disponibilização de aplicações de suporte à decisão;
- criação de procedimentos para adicionar novos relatórios e expandir as aplicações de suporte à decisão;
- desenvolvimento de procedimentos de backup para as aplicações de suporte à decisão e não somente o DW; e
- criação de procedimentos para investigar e resolver assuntos relacionados com a integridade de dados.

Poe considera que a interação contínua dos usuários com o DW possibilita o surgimento de novas exigências e, conseqüentemente, demanda a realização de modificações. Nesse caso, deve-se reiniciar o ciclo de desenvolvimento do DW, para que essas modificações sejam efetuadas.

Um outro aspecto levantado por Poe diz respeito ao desenvolvimento de um Projeto Piloto, cujo objetivo é, dentre outros, evitar um possível fracasso do projeto e possibilitar o ganho de experiência na construção do DW.

## 4.4 Comparação entre as Metodologias

Para fazer uma análise comparativa entre as metodologias descritas, foram definidos os seguintes critérios de avaliação:

- **terminologia Adequada** - *A terminologia usada deve ser de fácil entendimento e, de preferência, utilize termos comuns ao ambiente de DW*: esse critério avalia a terminologia adotada pelo autor no que se refere a clareza e termos comuns e de fácil compreensão;
- **praticidade** - *A metodologia deve ser prática, de forma a proporcionar etapas bem definidas e de fácil aplicação*: verifica se as etapas definidas são práticas e de fácil entendimento e aplicação, desde o levantamento inicial até a construção do DW abrangendo, inclusive, a definição das rotinas de ETL;

- **completude** - *A metodologia deve ser completa, abordando desde a fase inicial do projeto até a sua conclusão:* avalia a abrangência da metodologia, ou seja, se as etapas definidas são suficientes para o desenvolvimento do DW.
- **detalhamento** - *todas as fases devem ser bem detalhadas, facilitando a execução das diversas atividades:* Nesse critério é avaliado o nível de detalhe das etapas da metodologia;
- **aplicabilidade ao DWG** - avalia se a metodologia pode ser aplicada ao desenvolvimento de DWG, parcial ou integralmente.

A seguir é apresentado o resultado da comparação entre as metodologias descritas de acordo com os critérios definidos:

### **Terminologia Adequada:**

As metodologias de Kimball e Poe apresentaram uma terminologia bastante aceitável e dentro dos padrões usados no desenvolvimento de DW. Enquanto que a de Zachman, por sua vez, faz uso de termos, em sua maioria, específicos de sua metodologia.

### **Praticidade**

Neste critério o trabalho de Kimball foi considerado a melhor opção, visto que as etapas por ele definidas apresentaram-se bastante práticas, diretas e de fácil aplicação. Em seguida, tem-se a metodologia de Poe que, talvez, por ser mais abrangente que a de Kimball, não seja tão prática. Quanto a metodologia de Zachman, esta mostrou-se a menos prática, visto que, dentre outros aspectos, não é explicado como aplicar a matriz de classificação na modelagem e construção de um DW.

## Completude

O trabalho de Kimball juntamente com o de Poe sobressaem-se por serem os mais completos e abrangentes, sendo o de Poe até um pouco mais que o de Kimball, visto que sua abordagem abrange desde aspectos de levantamento de dados até evolução do DW. Embora o escopo definido no trabalho de Kimball aqui referenciado, seja menor que o de Poe, pode-se afirmar que seu trabalho é também completo. Já o trabalho de Zachman, apesar de considerar diversos pontos de vista em todo o ciclo de vida de desenvolvimento do DW, não é tão completo devido, principalmente, a falta de especificidade das etapas em relação a definição de dimensões, tabela de fatos, entre outras.

## Detalhamento

Neste quesito, tanto o trabalho de Kimball quanto o de Poe apresentam suas considerações de forma detalhada. Entretanto, pode-se afirmar que o de Kimball apresenta-se mais detalhado que o de Poe, devido seu escopo ser menor e mais específico ao desenvolvimento do DW.

Mais uma vez o trabalho de Zachman mostrou-se inferior aos outros dois. Claro que é um bom trabalho e tem a sua relevância, contudo foi definido de forma bastante genérica.

## Aplicabilidade ao DWG

Todos os trabalhos descritos podem ser aplicáveis ao desenvolvimento de um DWG. É óbvio que fica faltando, a todos eles, etapas de como definir e tratar dados geográficos. Mas, de um modo geral, as etapas descritas podem atender muito bem a um DWG.

Especificamente para esta dissertação, pode-se afirmar que as metodologias de desenvolvimento descritas atendem perfeitamente à construção de um DWG visto que o fato de estar sendo utilizado o método de integração do tipo híbrido, as manipulações são feitas como se todos os dados fossem convencionais.

A Tabela 3.2 apresenta um resumo da comparação entre as metodologias de Kimball, Zachman e Poe, segundo os critérios definidos.

Este capítulo teve como principal objetivo, descrever e avaliar as metodologias para o desenvolvimento de DW convencional de Ralph Kimball [KIM96], John A. Zachman [IZG97] e Poe [P<sup>+</sup>98] *apud* [PER00], de forma que as mesmas possam servir de base para algumas etapas da estratégia de desenvolvimento de DWG proposta por este trabalho, a qual será descrita no capítulo a seguir.

**Tabela 4.2.** *Resumo da Comparação entre as Metodologias de Kimball, Zachman e Poe.*

Critério	Kimball	Poe	Zachman
Terminologia Adequada	Ótimo	Ótimo	Razoável
Praticidade das Etapas	Ótimo	Bom	Razoável
Compleitude	Ótimo	Ótimo	Razoável
Detalhamento	Excelente	Ótimo	Não detalhada
Aplicabilidade ao DWG	Aplicável	Aplicável	Aplicável

# Capítulo 5

## Estratégia para desenvolvimento de DWG

Em se tratando de dados convencionais, já existem algumas metodologias e técnicas consolidadas, mas não formalizadas para construção de um DW. Entretanto, o mesmo não acontece quando se trata de dados geográficos, pois alguns autores abordam questões tais como custo, vantagens, benefícios e necessidades [ESRb, BER97, UB97, 4798], outros até mesmo apresentam DWG desenvolvidos ou produtos que auxiliam no desenvolvimento dos mesmos (Bar99, GeoServer), entretanto não é apresentada uma metodologia para o ciclo de desenvolvimento de um DWG.

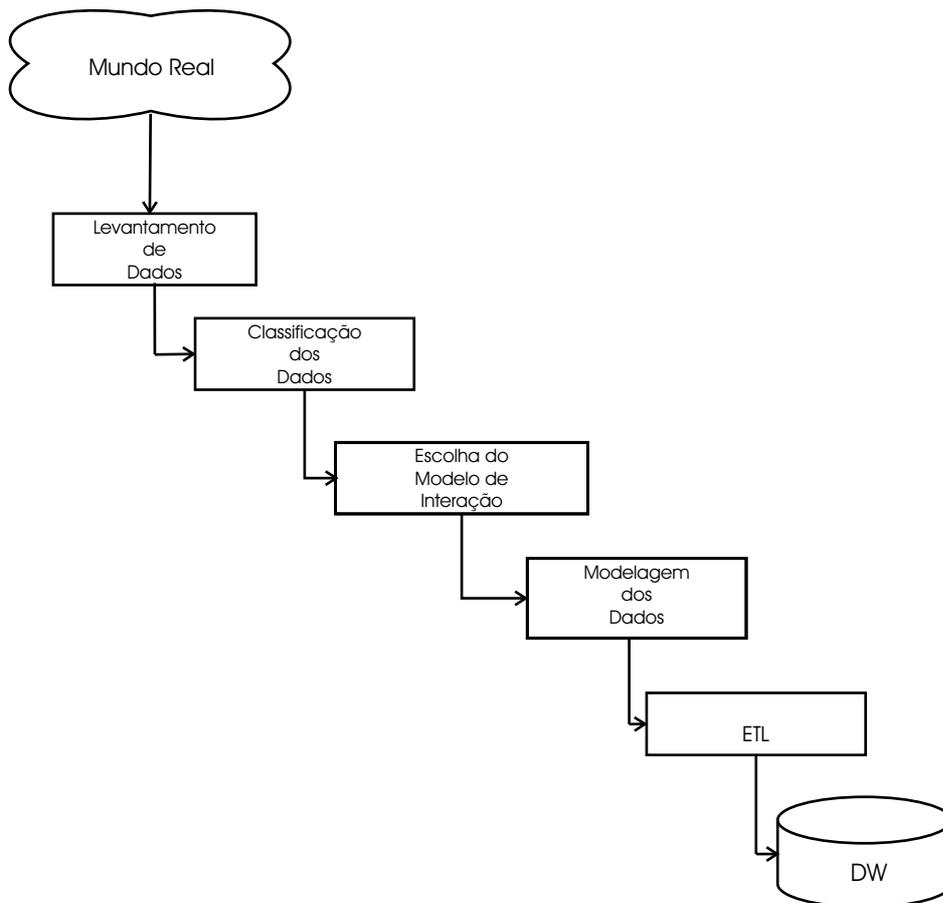
Dessa forma, tendo em vista a necessidade de se ter uma definição de métodos e técnicas para a construção de um DWG, este capítulo apresenta uma proposta de uma estratégia baseada nas abordagens de Kimball, Zachman e Poe (Capítulo 4), destinada a projetistas e desenvolvedores de DWG, de forma a dar-lhes suporte e auxílio nas decisões de projeto e tentar mostrar-lhes, pelo menos um caminho eficaz na execução dessa tarefa.

Conforme ilustra a Figura 5.1, a estratégia proposta está dividida em 5 etapas principais: *Levantamento de dados*, *Classificação dos dados*, *Modelo de Integração*, *Modelagem dos Dados* e a *ETL*. Nota-se que algumas dessas etapas já são de conhecimento comum porque são usadas em outras abordagens, como o ciclo de vida de desenvolvimento de Sistemas de Informação. Entretanto, o que difere nesse trabalho é o aspecto geográfico.

Vale ressaltar que tal estratégia utiliza o ciclo de vida de desenvolvimento clássico, embora não se tenha realizado nenhum estudo mais aprofundado sobre qual seria o mais adequado, visto que tal fato não foi considerado relevante para a realização deste trabalho.

### 5.1 Levantamento de Dados

Baseada na etapa 2 (Levantamento de requisitos e modelagem) da metodologia de Poe (seção 4.3 do Capítulo 4), o objetivo desta fase é a obtenção dos requisitos dos usuários finais, para uma avaliação das necessidades e conhecimento do negócio. E embora muitos desenvolvedores não dêem a devida importância a esta fase, ela é fundamental para o



**Figura 5.1. Ciclo de Vida da estratégia de desenvolvimento de um DWG**

sucesso de qualquer projeto, uma vez que todas as demais fases dependerão ou serão pensadas em função dessa primeira.

É importante ressaltar que essa fase não difere tanto assim dos desenvolvimentos de projetos já conhecidos, exceto pelo fato de que o sistema já exista e, particularmente, o tratamento de dados geográficos.

De forma a garantir um levantamento mais detalhado dos dados, esta fase foi subdividida em:

### 5.1.1. Identificação das Necessidades (O QUÊ?)

A finalidade desta etapa é, sobretudo, estudar as necessidades, expectativas e viabilidades do DWG. Para isso, deve-se fazer uso de técnicas tradicionais de entrevista e seções de facilitação, avaliação de documentos, questionários, entre outras [PF01, CC01].

Todavia, especificamente para um DWG, é imprescindível que seja abordado o aspecto geográfico. Ou seja, deve-se questionar quais informações os usuários desejam ver como informação textual e quais deseja ver na forma de imagem. É importante também nessa fase que, além de ser questionado quais informações o usuário quer ver, deve-se questioná-lo também sobre o quanto ele quer ver ou saber sobre uma determinada informação. Por exemplo, se um usuário quer visualizar, em forma gráfica, quais áreas foram atingidas por queimadas em um determinado mês, o projetista deve verificar que outras informações permeiam essa visualização do usuário, ou seja, que outras informações referentes às áreas queimadas seriam também importantes para ele, se interessar-lhe-ia saber o tamanho das áreas atingidas, a quantidades de focos identificados para as referidas áreas, a quantidade de áreas indígenas afetadas, entre outras. Tais questionamentos são imprescindíveis para todas as fases posteriores, particularmente, as fases de definição das dimensões e medidas.

Ressalta-se que tal etapa foi definida com base nas metodologias de Poe e Kimball, nas fases 2 (Levantamento de requisitos e modelagem) e fase 1 (Identificação dos processos a serem modelados) respectivamente.

### **5.1.2. Identificação dos Dados (ONDE?)**

Esta etapa baseou-se nas fases 6 (Análise dos sistemas fontes) e fase 4 (determinação, integração e mapeamento das fontes de dados) das metodologias de Zachman e Poe respectivamente. E tem como finalidade fazer um levantamento nos dados fontes para que se possa verificar e validar quais deles são descritivos (convencionais) e quais são os geográficos. Ou seja, deve-se obter em quais fontes de dados encontram-se as informações requeridas pelo usuário, e a partir daí identificar seus tipos, tamanhos, domínios, local de origem, formato, entre outros.

O produto dessa fase será uma lista contendo os atributos descritivos e geográficos e as demais informações supracitadas sobre os mesmos. Esta lista irá também auxiliar na definição dos metadados.

### **5.1.3. Granularidade dos dados**

Segundo Inmon [INM96], a granularidade é o aspecto mais importante do projeto. E a principal razão para isso é que ela afeta profundamente o volume dos dados que residem no DW e, ao mesmo tempo reflete o tipo de consulta que pode ser atendida.

O volume de dados de um DW é balanceado de acordo com o nível de detalhe de uma consulta. Ou seja, através da granularidade poderá se saber o quão detalhados deverão ser os dados e, conseqüentemente, como os mesmo influenciarão no tamanho do DW.

Dessa forma, esta etapa é de extrema importância por determinar o nível de detalhamento mais primitivo das dimensões de um DW, as quais devem ser suficientemente específicas para atender as necessidades obtidas durante a fase de levantamento dos dados.

Conseqüentemente, cada dimensão da tabela de fatos deve estar associada, no máximo, à unidade ideal de grandeza ou medida requerida por estas necessidades. Além disso, a unidade disponível nas fontes de dados determina geralmente um limite inferior que pode até prevalecer sobre a ideal se ela for superior a esta última.

Por exemplo, a análise mensal dos dados coletados sobre queimadas sugere que a unidade de tempo ideal seja o mês, embora estes dados sejam atualizados diariamente em suas fontes. Isto faz com que a unidade de tempo escolhida para a dimensão seja algo entre o dia e o mês, podendo, inclusive, ser semanal ou quinzenal. Supondo agora que os dados fossem atualizados a cada dois meses em suas fontes, e não diariamente como antes, não haveria como se fazer análises mensais e a unidade de medida de tempo passaria a ser então o bimestre.

Inmon [INM96] evidencia algumas regras para se definir o nível de detalhamento dos dados ou granularidade. Tais regras podem ser também aplicadas ao DWG.

Esta etapa baseia-se na metodologia definida por Kimball, especificamente na fase 2 (Definir a granularidade de cada tabela de fatos para cada processo).

#### **5.1.4. Estimativa do volume de dados (QUANTO?)**

Não se pode deixar de averiguar o volume de dados que será carregado no DW, ou seja, nessa etapa o desenvolvedor deverá analisar questões a respeito do quão grande o DWG ficará, tendo uma atenção especial para os dados geográficos, uma vez que um DW naturalmente já é extremamente volumoso.

Os dados geográficos ocupam um grande espaço em disco. Dessa forma, após terem sido devidamente identificados e localizados, deverão ser analisados em termos de tamanho, para que o desenvolvedor possa prevenir-se em relação às questões como desempenho e espaço em disco.

Via de regra, o volume de dados do DW cresce na proporção inversa ao tamanho do seu grão, o que implica dizer que a granularidade do DW interfere diretamente no tamanho de armazenamento necessário para comportá-lo num BD. Uma estimativa deste tamanho, ao longo de um determinado período de uso, permite decidir se o DW é viável ou não e, sendo esta última a possibilidade, rever a sua granularidade para diminuir o seu alto crescimento.

Pode-se estimar o tamanho do DW por quantidade de linhas ou *bytes*, o que for mais conveniente para o projetista. Basicamente, o cálculo se baseia na estimativa de quantidade de linhas (ou *bytes*), contabilizadas num período qualquer, multiplicada pela quantidade de períodos que se deseja considerar.

Esta etapa foi definida seguindo a fase 3 (*Bredbox Analysis*) da metodologia de Zachman

## 5.2 Classificação dos Dados

Classificar os dados significa incluí-los em categorias específicas de acordo com seus tipos ou funções. Esta etapa foi definida exclusivamente para esta estratégia, a qual determinou dois tipos básicos: Primário e Consolidado.

O tipo de dados Primário refere-se a qualquer dado que proporciona uma informação simples, tal como uma data, um nome, um endereço, entre outros. Esse tipo não se assemelha ao tipo simples ou monovalorado [DAT90, SKS99], porque o sentido de primário, no contexto aqui apresentado, diz respeito ao “valor origem”, ou seja, um atributo cujo valor não depende de nenhum outro, ou ainda, não é derivado ou gerado a partir de outro. Na verdade, o atributo primário pode originar ou produzir um outro atributo. Em outras palavras, seria o atributo como está no banco de dados fonte, e que ao ser transferido para o DWG não passou por nenhuma modificação. Ressalta-se que o dado primário até pode, no processo de ETL, passar por alguma modificação. Entretanto, essa modificação não deve alterar seu significado original, pois caso contrário ele será um atributo consolidado.

O Atributo Consolidado significa um atributo cujo valor foi gerado a partir de um processamento, cálculo ou mesmo da junção de atributos primários ou não. Este termo foi assim definido, para tornar-se mais próximo da terminologia usada pelos usuários de negócios nas mais diferentes áreas. Por exemplo, um atributo cuja função no DW é acumular valores, como valor total arrecadado no mês, ou percentual total de um determinado índice, ou ainda valor total de desconto.

Atributos do tipo consolidado são de extrema importância no contexto de DW, uma vez que esse tipo de sistema armazena, em sua maioria, dados acumulados ou totalizados. Dessa forma, tais atributos serão também fundamentais para a construção de um DWG.

Dados Geográficos do banco de dados fonte podem ser classificados tanto como primários quanto como consolidados. Por exemplo, supondo-se que diariamente são criados *Layers* dos focos de queimadas sobre a Região Norte do Brasil. Considerando que o nível de granularidade do tempo do DWG é o mês, um novo *layer* será gerado a partir da

combinação dos *layers* diários, este novo *layer* será classificado como consolidado e os que deram origem a ele serão primários.

### 5.3 Escolha do Modelo de Integração

Tal como a anterior, essa etapa também foi definida especificamente para esta estratégia, tendo como objetivo definir qual modelo de integração será usado.

Conforme visto no capítulo 3, um DW pode ser virtual, materializado ou híbrido.

Um DWG pode ser projetado segundo uma dessas três abordagens. Entretanto, para esta estratégia de desenvolvimento será considerada a abordagem híbrida, basicamente por três principais motivos:

- um DWG materializado exige um considerável recurso de memória disponível dificultando assim o seu desenvolvimento como forma de validar esta estratégia;
- um DWG virtual, embora resolvesse a questão de espaço, tornar-se-ia completamente vulnerável e dependente dos dados fontes; e
- um DWG híbrido, permite que sejam apresentadas ambas abordagens, oportunizando uma visão tanto do DWG materializado quanto do virtual. Deixando assim a critério do desenvolvedor a decisão de qual das três abordagens utilizar.

Um DWG híbrido exige que o projetista avalie a frequência de atualização dos itens de dados que serão contemplados no sistema, visto que os itens modificados menos frequentemente são, naturalmente, os maiores candidatos a tornarem-se virtuais. Para garantir o máximo possível de consistência no DWG, a sugestão dada por esta estratégia é a de se manter os dados geográficos nas fontes de dados e inserir no DWG os dados convencionais juntamente com as devidas referências aos dados geográficos. Essa sugestão baseia-se no fato de que esses últimos tipos de dados não são frequentemente atualizados, particularmente, pela forma com que tais dados são inseridos no sistema. Ou seja, dados convencionais são, geralmente, inseridos via digitação humana, tornando-os mais propensos a erros. Enquanto que dados geo-referenciados são sempre inseridos por equipamentos próprios tais como satélites, *scanners*, mesas digitalizadoras, câmeras digitais, entre outros.

Um aspecto a ser considerado, é que dados geográficos, por sua complexidade, apresentam maior dificuldade de manipulação que os dados convencionais. Dessa forma, manter os dados geográficos como virtuais, permite um desenvolvimento mais fácil e rápido do DWG.

## 5.4 Modelagem dos Dados

Os modelos de dados Estrela e Floco de Neve se adequam também a um DWG. Fica, portanto, a critério do desenvolvedor escolher qual modelo usar, de acordo com as características e necessidades do sistema.

Essa etapa é composta de 5 sub-etapas, sendo que as duas primeiras (5.4.1 e 5.4.2) foram baseadas nas fases 4 (Identificar os fatos) e fase 3 (Definir as dimensões de cada tabela de fatos) da metodologia de Kimball, enquanto as demais foram definidas pela estratégia proposta. A seguir será descrito cada uma delas.

### 5.4.1 Identificar a tabela de fatos

A partir do levantamento de dados feito, deve-se identificar os diferentes assuntos tratados. Por exemplo, queimadas, enchentes, desmatamento, ocupação e uso da terra, entre outros. Cada assunto identificado, inicialmente, corresponderá a uma tabela de fatos.

### 5.4.2 Identificar as Dimensões

Esta etapa foi definida, principalmente, porque a escolha das dimensões é o ponto chave no projeto, uma vez que cada dimensão pode ser vista como um ponto de entrada para tabela de fatos [OII90]. Entretanto, não há como se descrever uma “receita de bolo”, uma vez que cada projetista e desenvolvedor tem sua própria característica, perfil e forma de conduzir suas atividades. De qualquer forma, pode-se estabelecer os seguintes passos:

- O primeiro contato com as dimensões ocorre quando se está avaliando a granularidade da tabela de fatos, visto que a definição do nível de detalhe corresponde às informações que poderão ser avaliadas pelos usuários. Fazendo uma analogia com uma *query* na Linguagem SQL (*Structured Query Language*), tal como:

```
Select QtdeFocos
from Queimadas Q, Tempo T
where Q.IdTempo = T.IdTempo
Group by Q.IdLocal
```

- Uma outra alternativa para definir dimensões é fazer os seguintes questionamentos:

QUANDO?

ONDE?

QUEM?

Essa forma de identificação pode ser aplicada tanto para as dimensões espaciais quanto para as semi-espaciais e as não espaciais.

Geralmente, os atributos das cláusulas *where* e *group by* poderão ser vistos, a princípio como dimensões. Na *query* descrita anteriormente, pode-se identificar, então, as dimensões TEMPO e LOCAL.

### 5.4.3 Identificar atributos da dimensão

Os dois passos a seguir auxiliam na identificação dos demais atributos da dimensão, e se aplicam às categorias espaciais, semi-espaciais e não-espaciais.

- para cada dimensão definida o projetista deve se questionar que outras formas de visualização podem ser proporcionadas ao usuário. Ou seja, a partir do nível de detalhe especificado para cada dimensão deve-se verificar quais atributos devem ser vistos em níveis maiores ou menores. Por exemplo, uma dimensão Tempo cuja granularidade seja mês, pode ter como atributos complementares (níveis detalhe maiores) o bimestre, o semestre e o ano. Ou ainda, grãos menores de forma a se prevenir quanto a futuras solicitações tais como, quinzena e semana;
- outros atributos que podem ser identificados são aquele que dizem respeito a descrição do objeto da dimensão. Por exemplo, uma dimensão Local poderá conter atributos descritivos tais como Nome, Sigla (para o caso de Local referir-se a um Estado) quantidade de habitantes, entre outros. Além disso, para cada dimensão espacial ou semi-espacial, deverá se ter o atributo responsável por fazer a associação entre o atributo da dimensão e o seu respectivo dado geográfico no BD fonte.

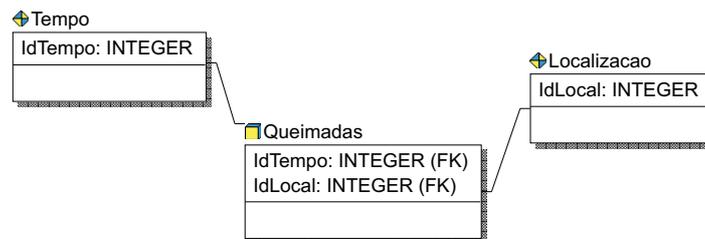
Além dos atributos acima descritos não se pode esquecer do atributo identificador, o qual corresponderá a chave primária (*primary key*- PK) [DAT90, SKS99, EN00] da tabela dimensão. Salienta-se que para as dimensões espaciais pode-se usar o seu próprio identificador geográfico como PK. Diferentemente dos projetos convencionais de BD, a PK não será determinada mediante um atributo identificador, ou seja, a PK corresponderá a um novo atributo, geralmente do tipo inteiro e seqüencial para cada instância da tabela.

### 5.4.4 Associar Fatos e Dimensões

Uma vez que a tabela de fatos e as dimensões estão definidas, deve-se fazer o relacionamento entre elas, para isso associa-se a PK da tabela dimensão com a tabela de

fato, tornando-se chave estrangeira (*foreign key* - FK) [DAT90, SKS99, EN00] na tabela de fatos.

A Figura 5.2 apresenta o relacionamento entre a tabela de fatos QUEIMADAS e as dimensões TEMPO e LOCAL.



**Figura 5.2. Relacionamento entre dimensões e a tabela de fatos**

A PK da tabela de fatos(QUEIMADAS) será composta pelo conjunto de FK, provenientes das dimensões TEMPO e LOCAL, ou seja, PK = IdTempo + IdLocal.

#### 5.4.5 Definir Atributos(medidas) da Tabela de fatos

Após a definição das dimensões e seus relacionamentos com a tabela de fatos, a próxima fase é complementar esta tabela mediante a identificação de medidas.

Partindo sempre das necessidades do usuário, o projetista deve verificar que dados permeiam as tomadas de decisões. Esses dados são, geralmente, valores numéricos chamados medidas e que correspondem, em sua maioria, aos dados consolidados.

As medidas representam os valores que deverão dar subsídios aos usuários para efetuarem análises e tomarem decisões. Por exemplo, em um DW convencional sobre vendas, uma medida seria o “valor total” arrecadado em um período. Já para um DWG sobre queimadas, por exemplo, uma medida poderia ser “área total atingida”.

Tal como muitas outras questões no desenvolvimento de um DWG, a definição das medidas também não possui uma regra. Ficando esta etapa dependente da experiência e capacidade do projetista. É claro que quanto melhor for explorada a fase de levantamento, mais fácil será a definição dessas medidas.

## 5.5 Extração, Transformação e Carga (ETL)

Muitos detalhes de implementação não são abordados por esta estratégia, em virtude de seu principal objetivo, que é disponibilizar um roteiro para construção de um DWG.

O leitor deve, portanto, buscar auxílio complementar nas mais diversas obras sobre DW na literatura corrente.

Ferramentas desenvolvidas para o processo de ETL também podem ser usadas para DWG com integração híbrida pelo fato deste manter os dados geográficos nas fontes de dados.

Para definir quais dados serão extraídos, o projetista deve ter em mãos a lista de dados definida na etapa 5.1.2. Em seguida deverá fazer a modelagem da DSA. Ou seja, projetar quais entidades e seus respectivos atributos pertencerão a DSA, bem como os relacionamentos existentes entre essas entidades. Após isso, deve-se implementar as rotinas necessárias para a transferência dos dados das fontes para a DSA. Conforme já citado anteriormente o uso da DSA não é obrigatório, entretando para esta estratégia foi considerado seu uso como forma de reduzir a complexidade das rotinas de extração.

## Transformação

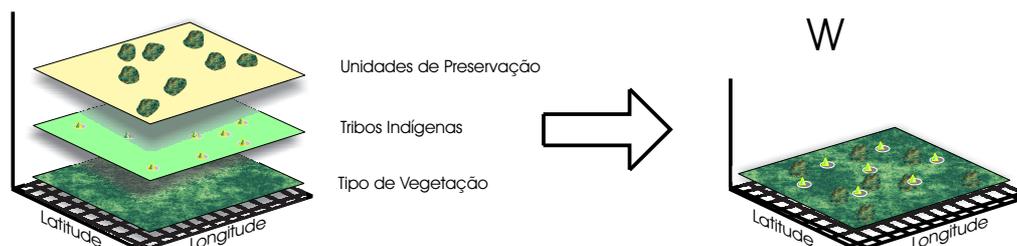
Conforme descrito na seção 3.4.2, nessa fase deverão ser executados os algoritmos responsáveis pela transformação de dados, efetuando mudanças nos dados primários, tais como cálculos, agrupamentos, somatórios, entre outros.

**Tabela 5.1.** Lista com informações sobre os dados requeridos

Atributo	Tipo	Tamanho	TipoOrigem	NomeOrigem	PK	FK	Obs
Codigo	Integer		Tabela	Clientes	S	N	
Nome	VarChar	50	Tabela	Clientes	N	N	
Idade	Integer		Tabela	Clientes	N	N	
DtaNiver	Char	5	Tabela	Clientes	N	N	
DtaCad	Date		Tabela	Clientes	N	N	

Especificamente para esta Estratégia sugere-se que transformações com pequenos graus de complexidade sejam executadas na primeira etapa da DSA. Por exemplo, o atributo idade da Tabela 5.1 deve ser transformado em Data do Nascimento para compor o DW. Dessa forma, o algoritmo deverá calcular pela data em que foi feito o cadastro (DtCad) o ano de nascimento do cliente, ou seja, se um determinado cliente possui o valor

25 no atributo idade, e na data de cadastro o valor 26/08/97, calcula-se seu ano de nascimento a partir da expressão:  $AnoNAsc = AnoCad - Idade$ . A partir daí deve ser feita a concatenação do Atributo DtaNiver, que contém somente o dia e o mês de nascimento, com o Ano de Nascimento encontrado. Note que, de certa forma, esta transformação não exige um nível de programação complexa, podendo, portanto, ser efetivada no momento da migração dos dados das fontes para a DSA.



**Figura 5.3. Exemplo de consolidação das camadas do GIS**

Os procedimentos executados para a obtenção da data de nascimento a partir da idade, e do mês e ano de aniversário, é um exemplo típico de transformação de dados convencionais. Para dados geográficos, entretanto, a maioria das alterações realizadas dizem respeito a transformação de dados geográficos primários em consolidados. A Figura 5.3 ilustra um exemplo de consolidação de dados geográficos, em que se tem as camadas (*layers*) Tipo de Vegetação, Tribos Indígenas e Unidades de Preservação; e a partir da combinação destes é gerado um *layer* W, o qual será armazenado no DWG. Salienta-se que a consolidação dos *layers* será feita no próprio GIS.

Devido essa fase incluir a etapa de limpeza dos dados, deve-se verificar quais dados tiveram seus valores inseridos de uma forma qualquer, não tendo nenhum significado para o sistema. Por exemplo, um CPF de um cliente que recebeu o valor 000.000.00-00 apenas para esse campo não ficar nulo. O ideal é que dados como esses não sejam inseridos no DWG, devendo o projetista, então, detectá-los e dar-lhes o tratamento necessário. Podendo esse tratamento ser definido, inclusive, com a ajuda do usuário final na decisão do que fazer.

### Carga

Nesta etapa o projetista deve, primeiramente, definir a periodicidade com que os dados deverão ser transferidos para o DWG. Salientando-se que esta periodicidade deve ser igual ou menor ao menor grão da dimensão tempo. Por exemplo, um DW cuja dimensão tempo possui como granularidade a semana, a carga deverá ser feita, de preferência, semanalmente. Sendo que, pode também ser feita diariamente e, nesse caso, deve-se efetuar algumas operações para calcular a informação desejada por semana.

Definir a carga em uma periodicidade com granularidade mais fina que a de tempo, embora seja uma forma do projetista prevenir o DWG quanto a futuras solicitações dos usuários, não é muito interessante visto que pode comprometer o desempenho das consultas ao DWG. Por exemplo, o DWG armazena dados referentes ao dia, porém os processa para mostrá-los por semana.

A partir da definição da periodicidade da carga do DWG, deve-se definir e implementar os programas responsáveis pela efetivação da carga de dados no DWG. Para tanto, pode ser desenvolvida uma aplicação a ser inicializada e manipulada pelo usuário, ou por uma rotina automática, inicializada pelo próprio SGBD na data e horário programado.

## 5.6 Resumo

Foi definida neste Capítulo uma estratégia de desenvolvimento de DWG, fundamentada nas metodologias de DW convencionais de Ralph Kimball [KIM96], John A. Zachman [IZG97] e Vidette Poe [P<sup>+</sup>98].

O uso dessas metodologias deve-se ao fato da inexistência de uma específica para DWG, em que são considerados dados geográficos também. Em função disso, os principais obstáculos encontrados na sua construção são:

- a dificuldade de manipulação dos dados geográficos; e
- o grande volume de dados ocupado pelos dados geográficos, o que implica na necessidade de uma grande área em disco para comportar o DWG.

Dessa forma, a estratégia descrita por este trabalho solucionou tais problemas mediante o uso da integração híbrida, a qual consiste na combinação das integrações virtual e materializada de um DW (descritas na seção 3.3 do Capítulo 3).

Um outro aspecto é que algumas etapas foram baseadas DW convencionais, e outras foram definidas de acordo com a experiência dessa autora somada às pesquisas e às necessidades do DWG usado como estudo de caso. Além disso, procurou-se, ao máximo, deixá-la com uma linguagem simples e concisa, deixando-a bem próxima do conhecimento comum dentro do ciclo de vida de desenvolvimento de um DW, de tal forma que qualquer pessoa com uma certa experiência em desenvolvimento de DW possa fazer uso dessa estratégia para desenvolver um DWG.

# Capítulo 6

## Implementação de um DWG para prevenção de Queimadas

Este capítulo objetiva descrever a implementação de um DWG seguindo a estratégia descrita no Capítulo 5, como forma de aplicar e validar a mesma mediante o estudo de caso do sistema de monitoramento de queimadas do SIVAM (Sistema de Vigilância da Amazônia).

### 6.1 Descrição do Cenário

O SIVAM possui uma infra-estrutura comum e integrada de meios técnicos destinados à aquisição e tratamento de dados ambientais. Esses meios abrangem o sensoriamento remoto, a monitoração ambiental e meteorológica, a exploração de comunicações, a vigilância por radares, recursos computacionais e meios de telecomunicações. Ressalta-se que tal sistema é considerado o maior projeto de controle e monitoramento do Brasil.

Para operacionalizar os produtos do SIVAM foi criado o SIPAM (Sistema de Proteção da Amazônia Legal), o qual é responsável pelo controle das aplicações que utilizam os dados coletados dos equipamentos de vigilância e proporcionam informações temáticas particulares às necessidades operacionais de cada usuário, quais sejam: uso de seus recursos hídricos e biodiversidade e ocorrência de desmatamentos e queimadas, assentamento e movimentação dos povos indígenas; das fronteiras terrestres; e o suporte à repressão ao contrabando, ao narcotráfico e à garimpagem ilegal.

Essas aplicações fazem o tratamento das imagens recebidas dos equipamentos de vigilância. Os analistas as verificam e, caso percebam alguma irregularidade, comunicam as Instituições Públicas responsáveis pela atuação direta em campo, tais como IBAMA, Polícia Federal, FUNAI, entre outros. Dentre essas aplicações tem-se a de monitoramento de focos de calor, que consiste na recepção diária das imagens do satélite NOAA-12. Essas imagens são tratadas por um algoritmo próprio que verifica e valida componentes referentes à temperatura, para em seguida identificar os focos de calor. O resultado do processamento é um arquivo texto contendo a localização geográfica (latitude e longitude) dos focos de calor detectados. O arquivo texto é importado em um SIG (ARC/VIEW 3.2)[ESRa] como objeto vetorial do tipo ponto (*point*), em que cada ponto representa a localização de um foco de calor detectado. Esse dado é associado a vários *layers* tais como áreas indígenas, mapa de vegetação e unidades de conservação, e a partir daí são

feitas análises sobre os focos identificados, sobretudo os detectados na Amazônia Legal. O *layer* final (contendo as sobreposições dos demais *layers*) é armazenado e identificado pela data. Uma das dificuldades em relação a funcionalidade do sistema é a ausência de um histórico, fazendo com que a maioria do trabalho feito atualmente seja destinado a combater e não prevenir queimadas.

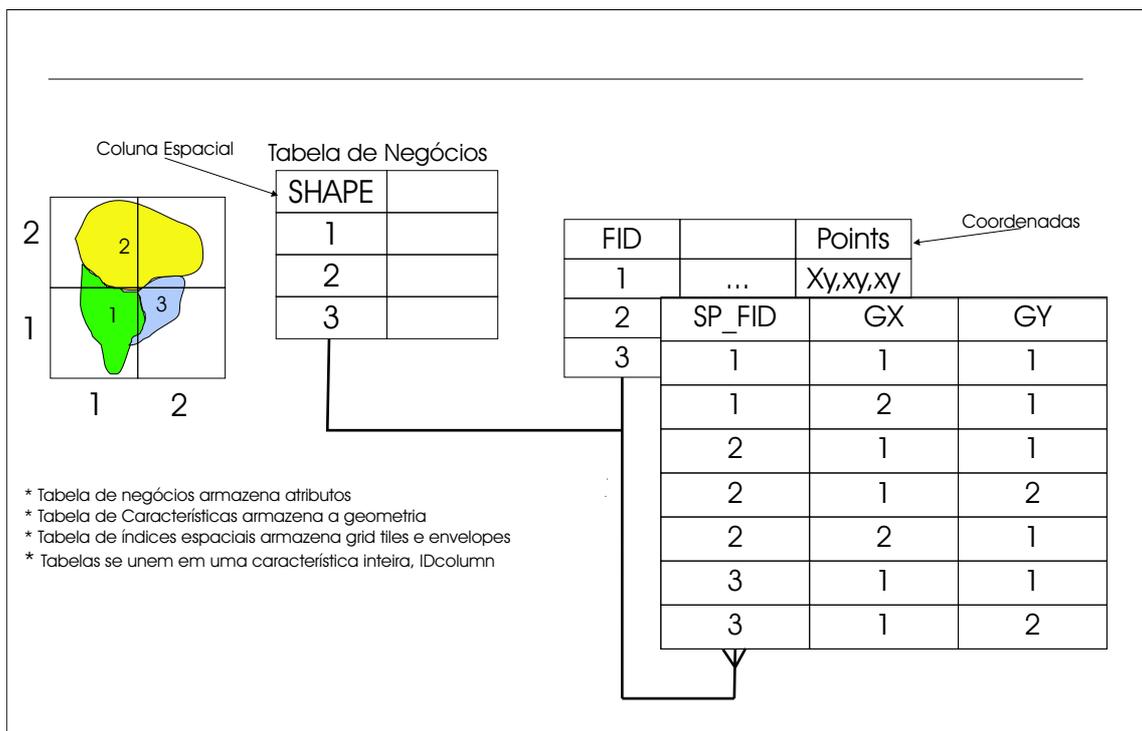
### 6.1.1 O ArcSDE

O SIPAM utiliza, para um melhor gerenciamento dos dados geográficos no banco de dados, o software ARC/SDE (da ESRI). Esse software atua em uma camada intermediária entre o GIS e o SGBD (*Oracle 8i*[URM00]).

Os dados geográficos armazenados pela aplicação de monitoramento de focos de calor, possuem representação de objetos vetoriais, especificamente pontos, linhas e polígonos. Esses dados são armazenados por meio do ARC/SDE, o qual gera para cada objeto geográfico 3 tabelas específicas, quais sejam:

- ***Business Table*** - A tabela de negócios contém os atributos da classe de objetos. Cada linha na tabela de negócio representa um objeto espacial (simples ou múltiplo), com um único identificador de forma. O único identificador de forma é armazenado na coluna espacial da tabela de negócio. O nome da tabela de negócio é o nome da classe de objeto, por exemplo *firelayer*. Cada registro é relacionado à linha correspondente na tabela de objeto, ou tabela F, através do identificador único.
- **Tabela F** - armazena a geometria numa coluna binária. Existe um relacionamento um-para-um entre a tabela de negócio e a tabela F responsável pelo armazenamento.
- **Tabela S** - a tabela de índice espacial, ou tabela S, armazena os *grid tiles* e os envelopes de objetos para a classe de objeto. Somente os *grid tiles* (células) que referenciam um envelope de objeto têm uma entrada na tabela. Os identificadores de objeto são também armazenados na tabela S para prover um relacionamento um-para-muitos das tabelas de negócio ou F para a tabela S.

As linhas das tabelas de negócio, F e S são conectadas através de um único identificador de objeto. O ArcSDE irá manter a integridade das tabelas de classe de objeto quando objetos forem adicionados, removidos ou modificados. Opcionalmente, tabelas relacionadas de atributos podem ser armazenadas no banco de dados. Essas tabelas seriam ligadas à tabela de negócio através de uma chave. A Figura 6.1 apresenta o exemplo de um dado geográfico pelo SDE nas tabelas F, S, *Business table*, e seus respectivos relacionamentos. Em que dada uma imagem composta de 3 *shapes* (1, 2 e 3) a mesma é dividida em *grid tiles* de forma a gerar uma matriz 2x2. A tabela de negócios contém os dados descritivos sobre cada um dos 3 *shapes*. O identificador dos mesmos são usados para fazer-se o relacionamento entre as tabelas F (FID), S (SP\_FID) e de negócios.



**Figura 6.1. Modelo de Armazenamento do ArcSDE**

A cada classe de objeto carregada no banco de dados é dado um único ID pelo ArcSDE, o SDE.Id. Esse ID é usado para nomear as tabelas F e S e relacioná-las com a tabela de negócio.

### 6.1.2 Considerações Gerais

Embora, esteja-se usando o termo DWG, na verdade a implementação aqui apresentada consiste em um DMG (*Data Mart Geográfico*), visto que o estudo de caso descrito refere-se a somente uma seção do setor de Vigilância Ambiental do SIVAM, para tornar-se um DWG seria necessário contemplar os sistemas da Organização como um todo.

Por questões de segurança não poderão ser apresentados os nomes reais de alguns atributos e tabelas.

## 6.2 Aplicação da Estratégia Proposta

Esta seção tem como objetivo apresentar a implementação de um DWG aplicado à previsão de queimadas, tendo como estudo de caso o Sistema de Monitoramento de Focos de Calor na Amazônia Legal do SIPAM. Ressaltando-se que o referido Órgão não possui,

atualmente, um sistema com essas características para atender as necessidades comuns dos usuários, havendo somente sistemas isolados atendendo de forma setorizada os grupos de usuários conforme descrito na seção 1.4 do Capítulo 1. Essa implementação seguiu a proposta de estratégia de desenvolvimento descrita no Capítulo 5. Adicionalmente, serão feitas algumas considerações no que diz respeito à aplicabilidade da estratégia, bem como as vantagens e desvantagens da mesma, e as dificuldades encontradas.

A seguir será apresentada cada uma das etapas definidas na estratégia de desenvolvimento de DWG, e suas respectivas considerações.

### **6.2.1 Levantamento de Dados**

Esta fase foi executada mediante entrevistas com funcionários do SIPAM, particularmente com a analista do Sistema de Monitoramento de Queimadas e os administradores do banco de dados, além de alguns gerentes de órgãos que utilizam seus serviços, no caso, IBAMA, FUNAI e Corpo de Bombeiros Florestal.

A seguir são descritas as principais necessidades levantadas.

#### **6.2.1.1 Identificação das Necessidades - O quê?**

Com base nas entrevistas, as necessidades identificadas foram divididas em Gerais e Específicas, conforme descrição a seguir:

##### **Gerais**

- a existência de um banco de dados que contemple numa única base dados geográficos e descritivos, de forma a otimizar o tempo de respostas das consultas;
- a integração de dados referentes às queimadas;
- o armazenamento de dados referentes a longos períodos de tempos (histórico), para prover análises e tomadas de decisões preventivas; e
- a criação de uma base de dados flexível para que novas necessidades possam ser acrescentadas no decorrer do tempo.

##### **Específicas**

- visualizar as queimadas por Município e por Estado em um determinado mês;
- obter de forma descritiva o total de queimadas por Município e por Estado em um determinado mês;

- visualizar graficamente as áreas indígenas e unidades de conservação afetadas por queimadas;
- obter a quantidade de áreas indígenas e unidades de conservação afetadas por queimadas;
- consultar a temperatura média de um determinado mês e ano; e
- analisar os índices pluviométricos e umidade relativa do ar.

#### **6.2.1.2 Identificação dos dados - Onde?**

A partir das necessidades reconhecidas, foi feito um segundo levantamento para se obter em quais tabelas, ou arquivos encontravam-se os dados para gerar as informações requisitadas. Vale ressaltar que por questões de segurança não será apresentado o modelo original do sistema de monitoramento de queimadas do SIPAM. A Tabela 6.1 apresenta a lista dos dados e seus respectivos tipos, obtida como resultado dessa fase.

#### **6.2.1.3 Granularidade dos Dados**

A análise de ocorrências de queimadas já é feita mensalmente, a partir da consolidação dos dados diários coletados. Conseqüentemente, usou-se o próprio mês como unidade para a dimensão tempo do DWG.

A unidade município foi adotada para a dimensão de localização por ser a menor subdivisão política disponível como objeto geográfico. O uso da subdivisão imediatamente maior, a de Estados, delimitava áreas muito grandes para análises que deveriam ser mais pontuais, envolvendo temperatura média, umidade relativa do ar entre outros, e portanto foi logo descartada.

#### **6.2.1.4 Estimativa do volume de Dados**

Considerando-se que o grão da localização de queimadas é o município e que a Amazônia Legal possui aproximadamente 729 municípios, tem-se, no máximo, 792 linhas inseridas no DWG em apenas um mês e 9.504 linhas após um ano.

Em um espaço de tempo de 5 anos, estima-se então que o DWG não conterà mais que 47.520 linhas, o que pode ser considerado um tamanho razoável para os padrões de sistemas transacionais e até mesmo modesto para os padrões de DW tradicionais.

Vale ressaltar que mesmo sendo essa estimativa calculada em *bytes*, ainda assim o tamanho obtido seria pequeno, em razão da utilização da integração híbrida que manteve os dados geográficos nas fontes.

**Tabela 6.1. Identificação dos Atributos**

Nome	Tipo	Tam.	Tipo da Fonte	Nome da Fonte	PK	FK	Observação
NomeMunicípio	VarChar2	60	Tabela	ES_MunicLUT			
NomeEstado	VarChar2	20	Tabela	ES_StateLUT			
NomeTerraInd	VarChar2	50	Layer	LimPolit_ADM	√		Polígono
NomeUnidConser	VarChar2	50	Tabela	ES_ConservUnity			
Data	Date		Tabela	Focos			
Ano	Char	4	Tabela	Focos			A partir da data do foco
SDE_IdEstado	Integer		Layer				Polígono
SDE_IdMunicípio	Integer		Layer				
SDE_ArcInd	Integer		Layer				
Cd.Tipo_Area	VarChar2	9	Layer				Se Munic. ou Estado
SD_IdUniConser	Integer		Layer	ES_ConservLayer			Polígono
SDE_Id	Integer		Layer	Focos			Ponto
Image_Data	Date		Layer	Focos			
Temperatura	Number		Arq.Texto	TempData			
IndPluviometrico	Number		Arq.Texto	TempData			
UmidRelativaAr	Number						

### 6.2.2 Classificação dos Dados

Os dados levantados para a construção do DWG foram classificados da seguinte forma:

#### Primitivos

- Nome do Município
- Nome do Estado
- SDE\_IdMunicípio

- SDE\_IdEstado
- IndicePluviometrico - menor do mês
- UmidadeRelativaAr - menor do mês

### **Consolidados**

- Mes - calculado a partir da data da queimada;
- Ano - tal como o mês também é calculado a partir da data da queimada;
- QdeQueimadas - calculada a partir do somatório dos foco diários;
- TemperaturaMedia - calculada a partir da média de temperatura diária em um determinado mês;
- TotalAreaIndigenas - calculado a partir do somatório de área indígenas afetadas por queimadas em um mês;
- AreasIndigenas - obtida através da concatenação do nome de todas as áreas indígenas afetadas por queimadas em um mês;
- TotalUnidadesConservação - calculado a partir do somatório das unidades de conservação afetadas por queimadas em um mês;
- UnidadesConservação - obtida através da concatenação do nome de todas as unidades de conservação afetadas por queimadas em um mês;
- SDE\_IdLayer - este atributo é obtido a partir da sobreposição dos *layers* diários de queimadas. Tal sobreposição é feita no GIS e em seguida convertida pelo SDE gerando assim o SDE\_IdLayer.

### **6.2.3 Escolha do Modelo de Integração**

A escolha do modelo de integração fica a critério do projetista. Entretanto, particularmente neste trabalho foi utilizada a abordagem híbrida, conforme anteriormente justificada.

O relacionamento do DWG com o dado geográfico que continuou nas fontes de dados, foi feito pelo atributo SDE\_Id, conforme anteriormente descrito, o qual foi gerado a partir de cada dado geográfico armazenado no banco de dado via SDE.

Uma grande vantagem do uso dessa abordagem é que o DWG também pode ser visualizado pelo próprio *software* GIS, embora este não possua todas operações de uma ferramenta OLAP. Além disso, questões sobre a economia de espaço de armazenamento e redução da complexidade de manipulação dos dados geográficos não podem deixar de serem citados.

## 6.2.4 Modelagem dos Dados

O modelo de dados utilizado neste estudo de caso foi o Modelo Estrela, por sua simplicidade e, pelo fato de o mesmo também poder ser aplicado a uma modelagem de dados espaciais [HSK98].

### 6.2.4.1 Identificação da Tabela de Fatos

Seguindo a característica “orientado por assunto” de um DW, a tabela de fatos definida foi QUEIMADAS. Ressalta-se, uma vez mais, que este estudo de caso corresponde apenas a um DMG e não a um DWG completo.

### 6.2.4.2 Identificação das Dimensões

Foram definidas as seguintes dimensões:

**Tempo** - principal dimensão de um DW, cujo objetivo é dar subsídios aos usuários para executarem consultas em relação aos intervalos de tempo em que determinados fatos ocorreram.

**Localização** - cuja função é prover informações aos usuários sobre onde os fatos de queimadas ocorreram.

### 6.2.4.3 Identificação dos Atributos das Dimensões

Para cada uma das dimensões definidas foram identificados os seguintes atributos:

#### **Tempo**

- Mês
- Ano
- Estação - ressaltando-se que na região amazônica as estações são divididas em apenas duas: da seca (verão) que inicia em meados do mês de abril e termina no início do mês de novembro quando começa a estação das chuvas (inverno).

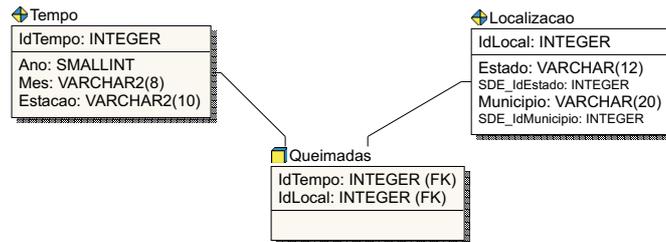
#### **Localização**

Responsável pelo armazenamento tanto dos nomes dos Municípios e Estados em que ocorreram as queimadas quanto pela referência dos mesmos nas tabelas do SDE.

- Município - corresponde ao nome do Município em que ocorreu a queimada;
- Estado - corresponde ao nome do Estado ao qual o Município em que ocorreu a queimada pertence;

- SDE\_IdMunicipio e SDE\_IdEstado - Identificadores responsáveis em efetuar o relacionamento com as tabelas do SDE (F, S e *business table*). Tal relacionamento tem como função capturar as imagens armazenadas dos municípios e estados nas tabelas SDE.

#### 6.2.4.4 Associação entre a tabela de Fatos e as Dimensões



**Figura 6.2. Associação da tabela Fatos e as Dimensões**

A Figura 6.2 apresenta a associação das dimensões com a tabela de fatos. Esta modelagem está representada através do modelo Estrela e foi feita utilizando-se a ferramenta ErWin (da Platinum) [LIM02] com modelagem multidimensional.

Note que a chave primária de cada uma das tabelas encontram-se na parte superior das entidades. Ou seja, para a dimensão Tempo a chave é o atributo IdTempo, para a dimensão Localização o atributo chave é IdLocal e, para a tabela de fatos Queimadas a chave é composta pelos atributos IdTempo e IdLocal os quais são Chaves estrangeiras das tabelas Tempo e Localização respectivamente.



**Figura 6.3. Representação de Fatos e Dimensões no ErWin**

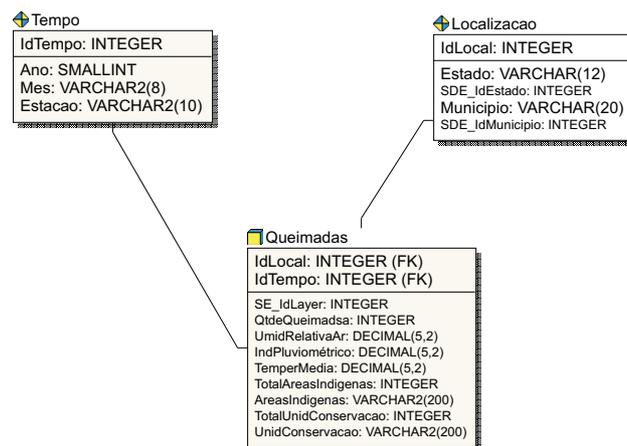
A Figura 6.3 ilustra as representações de fatos e dimensões usadas no modelo.

#### 6.2.4.5 Definição das medidas da tabela de Fatos

Os atributos medidas identificados para a tabela de Fatos Queimadas foram:

- SDE\_IdLayer - contém a referência para o *layer* composto por todos os focos de calor detectados em um mês;

- QdeQueimadas - armazena o total de focos de calor em em um mês;
- UmidadeRelativaAr - a menor do mês;
- IndicePluviometrico - o menor do mês;
- TemperaturaMedia - corresponde a média de temperatura diária em um mês;
- TotalAreaIndigenas - somatório de área indígenas afetadas por queimadas em um mês;
- AreasIndigenas - nome das áreas indígenas afetadas por queimadas em um mês;
- TotalUnidConservação - calculado a partir do somatório das unidades de conservação afetadas por queimadas em um mês;
- UnidConservação - nome de todas unidades de conservação afetadas por queimadas em um mês;



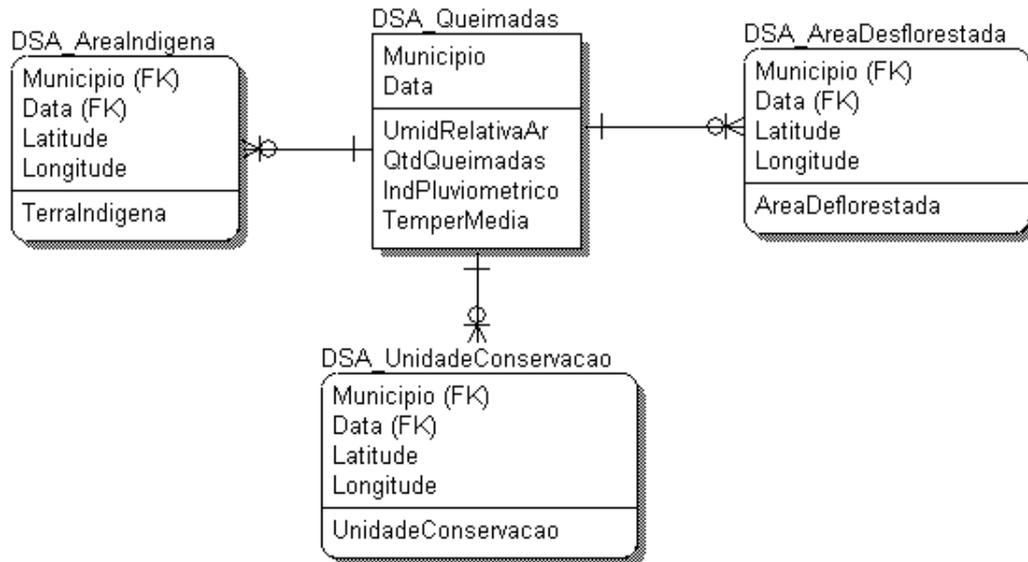
**Figura 6.4. Modelagem do DWG**

Tais medidas foram definidas com base nas necessidades levantadas junto aos usuários. A figura 6.4 apresenta o modelo completo, ou seja, fatos com suas respectivas medidas e sua associação com as dimensões.

### 6.2.5 ETL

Nessa fase foi modelado primeiramente a DSA (Figura 6.5), a qual contém todos os atributos que necessitaram passar por alguma transformação, como por exemplo a quantidade total de terras indígenas. Tal como o DWG a DSA também foi modelada mediante o uso da ferramenta ErWin. Em seguida foi feita a implementação da mesma, ou seja, a criação das tabelas da DSA. Como a DSA não contemplou todos os atributos do

DWG, alguns dados foram extraídos diretamente das fontes. A escolha por essa forma de implementação deu-se com o intuito de reduzir a complexidade dos algoritmos de ETL.



**Figura 6.5. Modelagem da DSA**

Após a implementação da DSA, foi definido e implementado, em seguida, o algoritmo responsável pela transferência dos dados tanto das fontes quanto da DSA para o DWG.

Além dos programas de ETL foi desenvolvido também uma rotina para executá-los automaticamente de forma que não se precisou criar uma aplicação *front-end* para que pudesse ser feita a carga do DWG.

A seguir é apresentado um trecho do código fonte responsável pela inserção de dados referentes às áreas indígenas no DWG.

```

PROCEDURE AgregaAreasIndigenas(p_Mes IN NUMBER, p_Ano IN NUMBER, p_IdLocal IN
INTEGER, p_TotalAreasIndigenas OUT INTEGER, p_AreasIndigenas OUT VARCHAR2) IS
CURSOR c_AreasIndigenas IS
SELECT
  TRIM(ArealIndigena) AS ArealIndigena
FROM
  DSA_ArealIndigena
WHERE
  TO_NUMBER(TO_CHAR(Data, 'MM')) = p_Mes AND
  TO_NUMBER(TO_CHAR(Data, 'YYYY')) = p_Ano AND
  Municipio = p_IdLocal
ORDER BY 1;
v_ArealIndigena c_AreasIndigenas%ROWTYPE;
BEGIN
  p_TotalAreasIndigenas := 0;
  p_AreasIndigenas := NULL;
  OPEN c_AreasIndigenas;
  FETCH c_AreasIndigenas INTO v_ArealIndigena;
  IF c_AreasIndigenas%FOUND THEN
    p_TotalAreasIndigenas := p_TotalAreasIndigenas + 1;
    p_AreasIndigenas := v_ArealIndigena.ArealIndigena;
    FETCH c_AreasIndigenas INTO v_ArealIndigena;
    WHILE c_AreasIndigenas%FOUND LOOP
      p_TotalAreasIndigenas := p_TotalAreasIndigenas + 1;
      p_AreasIndigenas := p_AreasIndigenas || ', ' || v_ArealIndigena.ArealIndigena;
      FETCH c_AreasIndigenas INTO v_ArealIndigena;
    END LOOP;
  END IF;
  CLOSE c_AreasIndigenas;
END AgregaAreasIndigenas;

```

### 6.3 Análise dos resultados

Quanto à aplicabilidade da estratégia proposta, pode-se afirmar que foi satisfatória para maioria das etapas. Apresentando uma certa redundância nas etapas “Identificando dados” e “Identificando Medidas da tabela de fatos”, visto que ao se identificar os dados, indiretamente já está se identificando também as medidas. É claro que, em um primeiro momento, não se tem a noção exata de quais dados serão medidas. Portanto, embora a etapa de definição de medidas dê ao desenvolvedor a impressão de que esteja repetindo uma atividade, na verdade, elas estão bem específicas, devendo ele apenas retirar da lista de atributos já identificados, aqueles que, efetivamente, serão medidas.

Quanto a aplicação do método de integração híbrida para DWG, o resultado foi também satisfatório, principalmente, devido a redução da complexidade no que diz respeito ao tratamento de dados geográficos, visto que eram suas referências e não seus valores que eram mantidos no sistema. Com isso, foi possível obter uma grande economia de espaço em disco.

Em consequência da aplicação desse método, a implementação do processo de ETL foi simplificada, uma vez que os objetos geográficos foram mantidos nas fontes de dados, e, portanto, suas referências foram tratadas como dados convencionais.

Considerando-se que esta implementação objetivou, particularmente, validar a estratégia proposta, usou-se a linguagem de especificação procedimental do SGBDR (Sistema de Gerenciamento de Banco de Dados Relacional) *Oracle* versão 8*i* denominada PL (*Procedural Language*) [FP97, OLI00].

# Capítulo 7

## Considerações Finais

Diferentemente de sistemas transacionais, nos quais questões sobre tempo de resposta, concorrência e otimização de espaço de armazenamento sempre foram consideradas com grande relevância, os DW (Data Warehouse) e suas aplicações apareceram com o foco na análise dimensional de fatos, obtidos a partir da extração, tratamento e carga de dados muitas vezes providos por esses mesmos sistemas transacionais.

Num outro segmento, o GIS (*Geographic Information System*) surgiu como uma ferramenta para processamento e análise de informações geo-espaciais, à qual foi progressivamente incorporada operações específicas para manipulação de mapas, inicialmente armazenados em arquivos com formato proprietário e que posteriormente passaram a residir também em bancos de dados. Desta forma, DW e GIS começaram seguindo caminhos bastante distintos em função dos seus objetos de análise e dos problemas que eles se propunham a resolver. Contudo, a tecnologia de banco de dados relacional permitiu que esses caminhos eventualmente se cruzassem<sup>1</sup>, dando origem a novas tecnologias como por exemplo o DWG.

Um DWG é essencialmente um DW com características espaciais, envolvendo o armazenamento, a recuperação e a execução de transformações (operações) em mapas ou objetos geográficos, os quais podem ser tanto dimensões como fatos. O seu desenvolvimento apresenta alguns obstáculos desafiadores, dentre eles podem ser citados:

- a inexistência de uma metodologia reconhecida e aplicada pela comunidade, que contemple todo o ciclo de desenvolvimento do DWG;
- o grande volume de dados ocupado por um DW convencional, o que implica no uso ainda maior de espaço em disco para o DWG, em função da necessidade de se armazenar também objetos geográficos e imagens; e
- a dificuldade de manipulação de dados geográficos, principalmente durante o processo de ETL (Extração, Transformação e Carga), quando ocorre a limpeza e transformação dos dados extraídos.

---

<sup>1</sup>A contribuição do banco de dados relacional foi permitir que dados geográficos (ponto, polígono, face e imagem) e não-geográficos (inteiro, real e texto) pudessem compartilhar a mesma área de armazenamento. Além disso, alguns fabricantes de SGBDR passaram a acrescentar, em seus produtos, tipos e operações primitivos específicos para dados geográficos.

Todavia, aliar o poder da análise histórica de fatos do DW com o de consulta, processamento e visualização de mapas traz enormes vantagens para o estudo e previsão de eventos. No caso do controle de queimadas da Amazônia, o cruzamento das dimensões de tempo e espaço dá ao analista a possibilidade de observar os fenômenos num contexto bem mais amplo que o de uma análise isolada em apenas uma destas dimensões.

Esses aspectos consistiram nos fatores primordiais de investigação deste trabalho, o qual teve como principal contribuição a definição de uma estratégia de desenvolvimento de DWG em que foi possível mostrar a integração de dados espaciais e não-espaciais para análise dimensional. Essa estratégia beneficia projetistas e desenvolvedores de DWG, no sentido de auxiliá-los em suas decisões de projeto, bem como servir de roteiro na criação de um DWG. Além disso, as etapas descritas pela estratégia podem ser complementadas por obras da literatura corrente, uma vez que foi embasada nas metodologias de desenvolvimento de DW convencionais existentes, particularmente as metodologias de Ralph Kimball, John A. Zachman e Vidette Poe, os quais são amplamente referenciados nesse contexto.

Outra contribuição a ser considerada foi a comprovação da possibilidade de se minimizar o uso de espaço de armazenamento em disco pelo DWG mediante o método de integração híbrida, visto que pela aplicação da estratégia proposta, em que se utiliza tal método, os dados geográficos foram mantidos nos sistemas transacionais e os convencionais transferidos para o DWG. Esse fato solucionou também o problema de manipulação de dados geográficos, visto que mantendo os mesmos nas fontes de dados, a manipulação foi toda feita somente sobre dados tradicionais, já que o que ficou armazenado no DWG foi apenas a referência ao dado geográfico e não ele próprio.

Cabe ainda salientar a relevância desse trabalho no que diz respeito as atividades em prol da preservação e conservação da Amazônia, uma vez que para avaliar e validar a estratégia proposta foi implementado um DWG para dar suporte a tomadas de decisões no que diz respeito a prevenção de queimadas na Amazônia Legal. Permitindo assim análises históricas dos dados e, conseqüentemente, o planejamento de ações estratégicas. Essa implementação foi feita utilizando-se o próprio ambiente e softwares do SIPAM composto de um SGBD *Oracle* versão 8i, O GIS ArcView 3.2, ArcSDE 8.0. A linguagem usada foi a PL/SQL da própria *Oracle* e a ferramenta usada para modelagem foi a ErWin 3.5 (da Microsoft).

Dentre as dificuldades encontradas na realização deste trabalho, as principais foram: O fato do SIPAM está operacional há apenas 1 ano e muitos dos seus produtos não estarem completamente implantados. Por exemplo, o módulo de desflorestamento, cujas informações são de suma importância para análises preventivas de queimadas; A falta de versões de produtos de GIS grátis, tais como o ArcView e ArcSDE, dificultando assim, não só parte da realização deste trabalho como também limitando os avanços de pesquisas na área. Finalmente, a maior das dificuldades que foi o fato de que os analistas atuantes hoje

no SIPAM não são, em sua grande maioria, os mesmos que participaram da concepção do projeto, fazendo com que a obtenção de informações sobre o armazenamento dos dados e seu contexto para queimadas fosse de difícil acesso.

## **Trabalhos Futuros**

Como proposta para trabalhos futuros sugere-se o seguinte:

- o desenvolvimento de um DWG seguindo a estratégia proposta, porém usando a abordagem materializada, visto que a abordagem virtual pode-se afirmar que seria facilmente aplicada;
- o desenvolvimento de uma ferramenta OLAP Espacial como forma de atender as características do DWG, proporcionando operações típicas para este tipo de dado;
- o desenvolvimento de um DWG completo para o SIPAM, ou seja, um DWG que contemple aos demais produtos (sistemas) como enchentes, metrologia, monitoramento de vôos clandestinos, entre outros; e
- finalmente, a complementação ou mesmo inserção de etapas na estratégia proposta neste trabalho, para consolidar uma metodologia de desenvolvimento de DWG;
- é importante também sugerir que seja feito um trabalho que dê suporte à metadados, visto que não foi considerado pela estratégia deste trabalho.

# Referências Bibliográficas

- [4798] Informativo Técnico N°. 47. Negócios e dados espaciais. Revista Unicamp, 1998. <http://www.revista.unicamp.br/infotec/informacao/inf54.htm>. Acesso em: 20/03/03.
- [7200] Informativo Técnico N°. 72. Conquistando a geografia. Revista Unicamp, 2000. <http://www.revista.unicamp.br/infotec/informacao/inf72.htm>. Acesso em: 17/04/03.
- [ABR98] Graça ABRANTES. Sistema de informação geográfica - conceitos, 1998. <http://www.isa.utl.pt/dm/sigdr/sigdr/SIGconceitos.html>. Acesso em: 25/03/03.
- [ARO89] Stan ARONOFF. *Geographic Information System: a Management Perspective*. WDL Publications, Canada, 1989.
- [BAVU98] P. BARRETO, P. AMARAL, E. VIDAL, and C. UHL. Costs and benefits of forest management for timber production in eastern amazonia. *Forest Ecology and Management*, 1998.
- [BER97] Jan Van BERKEL. Data warehouse: Where to locate GIS, 1997. <http://www.esri.com/base/common/userconf/proc97/PROC97/PAP650/P650.HTM>. Acesso em: 23/03/03.
- [BER99] T. BERNHARDSEN. *Geographic Information Systems: An Introduction*. Inc. John Wiley and Sons, New York, second edition edition, 1999.
- [BUR98] P.A. BURROUGH. *Principles of Geographical Information Systems*. Clarendon Press, Oxford, 1998.
- [C+96] Gilberto CÂMARA et al. Anatomia de sistemas de informação geográfica, 1996.
- [CÂM95] Gilberto Neto CÂMARA. Modelos, linguagens e arquitetura para bancos de dados geográficos, 1995. <http://dpi.inpe.br/dpi/teses/gilberto>.
- [CÂM00] Gilberto CÂMARA. Sistema de informações geográficas para aplicações ambientais cadastrais: Uma visão geral, 2000. <http://www.dpi.inpe.br/geopro/trabalhos/analise.pdf>. Acesso em:20/01/03.

- [CC01] Ariadne Rizzoni CARVALHO and Thelma CHIOSSI. *Introdução a Engenharia de Software*. Editora da UNICAMP, 2001.
- [CM03] Gilberto CÂMARA and Antônio M. V. MONTEIRO. Conceitos básicos da ciência da geoinformação. <http://www.dpi.inpe.br/gilberto/livro/introd/cap2-conceito.pdf>. Acesso em: 04/05/03, 2003.
- [COU92] H. COUCELIS. People manipulate objects (but cultivate fields): Beyond the raster-vector database in GIS. In *Proc. International Conference on GIS - From Space to Territory: Theories and Methods of Spatial Reasoning*. Springer Verlay LNCS, 1992.
- [COW91] David COWEN. What is GIS? In M. F. Goodchild and K. K. Kemp, editors, *NCGIA Core Curriculum, Introduction to GIS*, Santa Barbara, CA: National Center for Geographic Information and Analysis, 1991.
- [CRF97] Maria Luiza CAMPOS and Arnaldo V. ROCHA FILHO. Data warehouse. *XVII Congresso da Sociedade Brasileira de Computação*, 1997. <Http://genesis.nce.ufrj.br/dataware/tutorial/indice.html>. Acesso em: 23/03/03.
- [CS82] T. T. COCHRAN and P. A. SÁNCHEZ. Lan resources, soils and their management in amazonia: A state of knowledge report. In *Amazônia: agriculture and land-use research*. S. B. Hecht, 1982.
- [DAL99] Adriano DAL'ALBA. Um estudo sobre data warehouse, 1999. <http://www.geocities.com/siliconvalley/port/5072/>. Acesso em: 10/03/03.
- [DAN90] K. DANGERMOND. A classification of software components commonly used in geographic information systems. *Introductory Readings in Geographical Information Systems*, 1990.
- [DAT90] C. J. DATE. *Introdução a Sistemas de Bancos de Dados*. Campus, Rio de Janeiro-RJ, 4 edition, 1990.
- [Dit88] K. R. Dittrich, editor. *Multiple Inheritance and Genericity of the Integration of a Database Management System in an Object-Oriented Approach*. Springer Verlag, 1988. LNCS 334.
- [EN00] Ramez ELMASRI and Shamkant B. NAVATHE. *Fundamental of Database Systems*. Addison Wesley, 3 edition, 2000.
- [Eng87] *Object-Oriented Modeling GIS: Inheritance and Propagation.*, Baltimore, MD, 1987.
- [ESRa] <http://www.esri.com>. Acesso em: 20/06/2003.

- [ESRb] Gis by ESRI. Spatial data warehouse. an ESRI white paper. [www.esri.com/library/whitepapers/pdf/spatialdata.pdf](http://www.esri.com/library/whitepapers/pdf/spatialdata.pdf).
- [ETI] <http://www.eti.com>. Acesso em: 10/05/2003.
- [FG90] A. FRANK and M. GOODCHILD. Two perspectives on geographical data modelling. Technical report, National Center for Geographic Information and Analysis, 1990.
- [FP97] Steve FEUERTEN and BILL PRYBIL. *Oracle PL/sql Programming*. O'reilly & Associates, 1997.
- [FRA88] A. U. FRANK. Requirements of a database management system for a GIS. *Engineering and Remote Sensing*, 1988.
- [G+92] M. GOODCHILD et al. Integrating GIS and spatial data analysis: Problems and possibilities. *International Journal of Geographical Information Systems*, 6(5), 1992.
- [GOO92] F. GOODCHILD. Geographical information science. *International Journal of Geographical Information Systems*, 6, 1992.
- [GS92] S. C. GUPTILL and M. STONEBRAKER. The sequoia 2000 approach to managing large spatial object database. In *Proceedings of the 5th International Symposium on Spatial Data Handling*, Charleston, SC, 1992.
- [GW98] Paul Gray and Hugh J. Watson. *Decision Support in the Data Warehouse*. PTR, 1998.
- [HAR97] Lauro T. HARA. Técnicas de apresentação de dados em geoprocessamento, 1997. <http://www.dpi.inpe.br/spring/portugues/aplicacoes.html>.
- [HK00] Jiawei HAN and Micheline KAMBER. *Data Mining: Concepts and Techniques*. Morgan Kaufman, 2000.
- [HMS01] David J. HAND, Heikki MANNILA, and Padhraic SMYTH. *Principles of Data Mining (Adaptive Computation and Machine Learning)*. MIT Press, 2001.
- [HSK98] Jiawei HAN, Nebojsa STEFANOVIC, and Krzysztof KOPERSKI. Selective materialization: An efficient method for spatial data CubeConstruction. In *Research and Development in Knowledge Discovery and Data Mining, Second Pacific-Asia Conference, PAKDD'98*, 1998.
- [IBG01] Fundação IBGE. Sinopse preliminar do censo demográfico, 2001.
- [IBM] <http://www.ibm.com>. Acesso em: 20/05/2003.

- [INM92] William H. INMON. *Building the Data Warehouse*. John Wiley and Sons Inc., USA, 1992.
- [INM96] W. H. INMON. *Buiding Data Warehouse*. John Wiley & Sons, Inc, 1996.
- [INM97] William H. INMON. *Como Construir O Data Warehouse. Trad. 2a. Edição*. Ana M. Netto Guz, Rio de Janeiro, 1997.
- [IR93] Robert M. ITAMI and Robert J. RAULINGS. *What is a Geographic Information System?* DLSR, Melbourne, Australia, 1993.
- [IZG97] W. H. INMON, J. A. ZACHMAN, and G. J. GEIGER. *Data Stores, Data Warehousing and the Zachman Framework*. MacGraw-Hill, 1997.
- [K<sup>+</sup>98] Ralph KIMBALL et al. *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses*. John Wiley and Sons, Inc., New York, 1998.
- [KIM96] Ralph KIMBALL. *The Data Warehouse Toolkit*. John Wiley and Sons, Inc., New York, 1996.
- [KIM97] Ralph KIMBALL. Dimensional modeling manifesto. *DBMS*, agosto 1997.
- [LIM02] Adilson S. LIMA. *ErWin 4.0 Modelagem de Dados*. Érica, 2002.
- [LOP02] Ilza Leite LOPES. Estratégia de busca na recuperação da informação: Revisão da literatura. *Ci. Inf.*, v. 31(2), maio/agosto 2002.
- [M<sup>+</sup>93] David J. MAGUIRE et al. *Geographical Information Systems - Volume II - Applications*. John Wiley and Sons, 2 edition, 1993.
- [MAC00] Felipe N. R. MACHADO. *Projeto de Data Warehouse: Uma Visão Multidimensional*. Érica, São Paulo, 2000.
- [MAG91] Michael MAGUIRE, David J. and GOODCHILD. An overview and definition of GIS. 1, 1991.
- [MEL03] Ronaldo S. MELLO. Classes de dados geográficos, 2003. Disponível em: [http://www.inf.ufsc.br/~ronaldo//ine5342/10-BDG\(2\).pdf](http://www.inf.ufsc.br/~ronaldo//ine5342/10-BDG(2).pdf).
- [MP94] Cláudia B. MEDEIROS and Fátima PIRES. Databases for gis. *FALTA*, 23(1), 1994.
- [NMA99] Daniel C. NEPSTAD, Adriana G. MOREIRA, and Ane A. ALENCAR. *A Floresta Em Chamas: Origens, Impactos e Prevenção de Fogo Na Amazônia*. Programa Piloto para a Proteção das Florestas Tropicais do Brasil, Brasília, 1999.
- [OII90] Beng Chin OII. *Efficient Query Processing in Geogrraphic Information Systems*. Springer Verlag LNCS, 1990.

- [OLI00] Wilson José de OLIVEIRA. *Oracle 8i & PL/sql*. Visual Books, 2000.
- [OLI02] Wilson José OLIVEIRA. *Data Warehouse*. Visual Books. Visual Books, Florianópolis - SC, 2002.
- [OPE91] S. OPENSHAW. Developing appropriate spatial analysis methods for GIS. In D. J. MAGUIRE, M. F. GOODCHILD, and D. W. RHIND, editors, *Geographical Information Systems*, chapter 1 - Principles. Longman, London, 1991.
- [owg] <http://www.owg.com.br>. Acesso em: 20/06/2003.
- [P<sup>+</sup>98] Vidette POE et al. *Building a Data Warehouse for Decision Support*. Prentice Hall PTR, New Jersey, 1998.
- [PER99] Max Roberto PEREIRA. Data warehouse: Otimizando seu desempenho. *Developer's Magazine*, (32), abril 1999.
- [PER00] Walter A. L. PEREIRA. Uma metodologia de inserção de tecnologia de data warehouse em organizações. Master's thesis, Pontifícia Universidade Católica do Rio Grande do Sul - PUCRS, 2000.
- [PEU94] D. PEUQUET. It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. *in Annals of the Association of American Geographers*, 1994.
- [PF01] Wilson de Pádua PAULA FILHO. *Engenharia de Software: Fundamentos, Métodos e Padrões*. LTC, 2001.
- [RSV02] Philippe RIGAUX, Michel SCHOLL, and Agnès VOISARD. *Spatial Databases with Applications for GIS*. Morgan Kaufmann, USA, 2002.
- [SIN01] Harry SINGH. *Data Warehouse*. Makron Books, São Paulo - SP, 2001. Tradução: Mônica Rosemberg.
- [SKS99] Abraham SILBERSCHATZ, Henry F. KORTH, and S. SUDARSHAN. *Sistema de Banco de Dados*. 1999. 3.
- [SLTC02] S. SHEKHAR, C. T. LU, X. TAN, and S. CHAWLA. Map cube: A visualization toll for spatial data warehouse. Computer Science Departament, 2002.
- [SMI87] Terence R. SMITH. Requeriments and principles for the implementation and construction of large-scale geographic information systems. *International Journal of Geographic Information Systems*, 1, 1987.
- [SOU03] Carla O. F. de SOUZA. Uma proposta de redução de custos no processo de ETL de DW. Master's thesis, Universidade Federal de Pernambuco (UFPE), 2003. defendida em 30 setembo de 2003.

- [Sun] <http://www.sunopsis.com>. Acesso em: 10/05/2003.
- [THO02] Erik THOMSEN. *OLAP Solutions: Building Multidimensional Information Systems*. John Wiley & Sons, 2nd edition edition, 2002.
- [TOM91] C. Dana TOMLIN. Cartographic modelling. In D. J. Maguire, N. F. Goodchild, and D. W. Rhind, editors, *Geographical Information Systems*, chapter 1 - Principles. Longman, London, 1991.
- [UB97] L. W. X. URMAN and J. Van BERKEL. *GIS Mapping for the Data Warehouse*, 1997.
- [URM00] Scott. URMAN. *ORACLE 8i. Advanced PL/SQL Programming*. McGraw-Hill, 2000.
- [YRN99] Ricardo Baeza YATES and Berthier RIBEIRO NETO. *Modern Information Retrieval*. Addison-Wesley, 1999.