



Ludymilla Marianne Guimarães de Andrade

***Data Warehouse Geoespacial aplicado à  
Gestão estratégica na Prefeitura de Belo  
Horizonte***

XIII Curso de Especialização em  
Geoprocessamento  
2011



UFMG  
Instituto de Geociências  
Departamento de Cartografia  
Av. Antônio Carlos, 6627 – Pampulha  
Belo Horizonte  
cartog@igc.ufmg.br

**LUDYMILLA MARIANNE GUIMARÃES DE ANDRADE**

***Data Warehouse* Geoespacial aplicado à Gestão estratégica na  
Prefeitura de Belo Horizonte**

Monografia apresentada como requisito parcial à obtenção do grau de Especialista em Geoprocessamento. Curso de Especialização em Geoprocessamento. Departamento de Cartografia. Instituto de Geociências. Universidade Federal de Minas Gerais.

Orientador: Prof. Dra. Karla Albuquerque de Vasconcelos Borges

**BELO HORIZONTE**

**2011**

A554d Andrade, Ludymilla Marianne Guimarães de.  
2011 Data Warehouse Geoespacial aplicado à gestão estratégica na  
prefeitura de Belo Horizonte [manuscrito] / Ludymilla Marianne  
Guimarães de Andrade. – 2011.  
ii, 47 f. : il. (algumas color.)

Monografia (especialização em Geoprocessamento) – Universidade  
Federal de Minas Gerais, Instituto de Geociências, 2011.

Orientadora: Karla Albuquerque de Vasconcelos Borges.

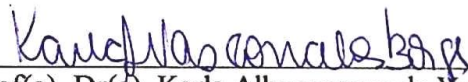
Bibliografia: f. 46-47.

1. Sistemas de informação geográfica. 2. Armazenamento de dados.  
3. Sistemas de informação gerencial. I. Borges, Karla Albuquerque de  
Vasconcelos. II. Universidade Federal de Minas Gerais, Instituto de  
Geociências. III. Título.

CDU: 528

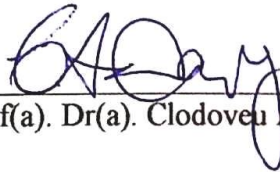
Aluno (a) Ludymilla Marianne Guimarães Andrade

Monografia defendida e aprovada em cumprimento ao requisito exigido para obtenção do título de Especialista em Geoprocessamento, em 24 de novembro de 2011, pela Banca Examinadora constituída pelos professores:



---

Prof(a). Dr(a). Karla Albuquerque de Vasconcelos Borges



---

Prof(a). Dr(a). Clodoveu Augusto Davis Júnior

## RESUMO

A utilização de uma ferramenta de *Data Warehouse* Geográfico (DWG) apresenta vantagens estratégicas quando comparada à ferramenta de *Data Warehouse* convencional. Através do DWG é possível analisar informações georreferenciadas, o que traz inúmeros benefícios ao tomador de decisão, as informações apresentadas em forma de mapas torna a análise mais intuitiva. No âmbito da administração pública, essas informações têm grande valor para os gestores, pois é possível ter uma visão global da informação distribuída por todo o município. A Prefeitura de Belo Horizonte é carente de uma ferramenta de DWG, atualmente o processo de geração de informações estratégicas espacializadas é trabalhoso, visto que a ferramenta de DW existente não possui extensão espacial. Além de utilizar uma ferramenta de *Data Warehouse* para manipular as informações alfanuméricas, é necessária também, a utilização de ferramentas de SIG para georreferenciar, processar e analisar os dados. A ferramenta utilizada nesse trabalho é a Pentaho, *software* de *Business Intelligence Open Source*. Porém essa ferramenta não é intuitiva, demanda certo tempo para aprender a manuseá-la, além da escassez de documentação e informações, principalmente no que se refere aos componentes com extensão espacial. Ressalta-se que, devido à limitação quanto ao uso da ferramenta, alguns objetivos específicos, propostos nesse trabalho, não puderam ser alcançados. Contudo, foi possível observar que, uma ferramenta de DWG que integre, em um único ambiente, funcionalidades de manipulação de dados alfanuméricos e geográficos, sendo capaz de gerar mapas, gráficos e relatórios e que seja de fácil manipulação pelos próprios gestores, que serão capazes de gerar informações estratégicas de forma dinâmica e confiável, é de suma importância.

## SUMÁRIO

|  |    |
|--|----|
| <b>CAPÍTULO 1 - INTRODUÇÃO</b> .....   | 1  |
| 1.3 Objetivos .....  | 3  |
| 1.3.1 Geral .....  | 3  |
| 1.3.2 Específico .....   | 4  |
| <b>CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA</b> .....                                      | 5  |
| 2.1 Business Intelligence - BI .....   | 5  |
| 2.2 Data Warehouse e Data Warehouse Geográfico .....                                 | 5  |
| 2.3 OLTP versus OLAP .....   | 13 |
| 2.4 Estrutura do Data Warehouse.....   | 15 |
| 2.5 Data Mart.....   | 17 |
| 2.6 Granularidade .....  | 17 |
| 2.7 Projetando um Data Warehouse/Data Warehouse Geográfico .....                     | 18 |
| 2.8 Modelagem Dimensional .....  | 19 |
| 2.8.1 Modelo Estrela ( <i>Star Schema</i> ).....                                     | 20 |
| 2.8.2 Modelo Floco de Neve ( <i>Snowflake</i> ).....                                 | 22 |
| 2.8.3 A tabela de Fatos .....  | 23 |
| 2.8.4 Tabela de Dimensões .....  | 24 |
| 2.9 Sistemas de Informação Geográfica (SIG) e DWG .....                              | 25 |
| <b>CAPÍTULO 3 - ESTUDO DE CASO - SACWeb</b> .....                                    | 27 |
| 3.1 SACWeb.....  | 27 |
| 3.2 Geração de informações estratégicas especializadas através do SACWeb na PBH .... | 28 |
| 3.3 Obtenção dos dados para a execução do trabalho .....                             | 34 |
| 3.4 Pentaho .....  | 34 |
| 3.5 Procedimentos para a execução do trabalho .....                                  | 36 |
| <b>CAPÍTULO 4 - RESULTADOS E DISCUSSÕES</b> .....                                    | 43 |
| <b>CAPÍTULO 5 - CONCLUSÕES</b> .....   | 45 |
| <b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....  | 46 |

## LISTA DE FIGURAS

|  | <u>Pág.</u> |
|--|-------------|
| Figura 1: Arquitetura de um DW   | 8           |
| Figura 2: Arquitetura de um DWG  | 8           |
| Figura 3: Orientação por assunto.  | 10          |
| Figura 4: Integração de dados.   | 12          |
| Figura 5: Não volátil no DW.   | 12          |
| Figura 6: Estrutura e fluxo de dado do <i>Data Warehouse</i> .                       | 15          |
| Figura 7: Cubo tridimensional com o total de vendas                                  | 20          |
| Figura 8: Representação do modelo Estrela.   | 21          |
| Figura 9: Representação do modelo Floco de Neve.                                     | 23          |
| Figura 10: Exemplo de tabela de Fatos com chave composta e Dimensões.                | 24          |
| Figura 11: Exemplo de Sobreposição de Camadas.                                       | 26          |
| Figura 12: Dados do SACWeb, com o campo de coordenadas geográficas                   | 28          |
| Figura 13: Mapa gerado no Arcgis com os pontos representando a coordenada geográfica | 30          |
| Figura 14: Mapa gerado no Arcgis   | 31          |
| Figura 15: Mapa com a localização das árvores em logradouros públicos de BH          | 32          |
| Figura 16: Gráfico gerado no Arcgis  | 33          |
| Figura 17: Pentaho Geoespacial.  | 35          |
| Figura 18: Tela do pgAdmin - PostGres  | 37          |
| Figura 19: Tela do Pentaho   | 38          |
| Figura 20: Tela do <i>Workbench</i>  | 38          |
| Figura 21: Tabelas criadas no <i>Workbench</i>                                       | 39          |
| Figura 22: Modelo do <i>Data Mart</i> Geográfico                                     | 40          |

## LISTA DE SIGLAS E ABREVIATURAS

|        |   |
|--------|---|
| BD     | -Banco de Dados                                   |
| BI     | - <i>Business Intelligence</i>                    |
| DM     | - <i>Data Mart</i>                                |
| DSS    | - <i>Decision Suport System</i>                   |
| DW     | - <i>Data Warehouse</i>                           |
| DWG    | - <i>Data Warehouse Geográfico</i>                |
| EA     | - <i>Enterprise Architect</i>                     |
| EIS    | - <i>Enterprise Information System</i>            |
| ETL    | -Extração, Transformação e Carga                  |
| GAIS   | - Gerência de Arquitetura de Informações          |
| GPS    | - Sistema de Posicionamento Global                |
| KML    | - <i>Keyhole Markup Language</i>                  |
| MDX    | - Multidimensional Expressions                    |
| MER    | -Modelo Entidade Relacionamento                   |
| OLAP   | -Processamento Analítico em Tempo Real            |
| OLTP   | - Processamento de Transações em Tempo Real       |
| OMT-G  | - Técnica de Modelagem de Objetos Geográficos     |
| PBH    | - Prefeitura de Belo Horizonte                    |
| PDI    | - <i>Pentaho Data Integration</i>                 |
| SACWeb | - Sistema Informatizado de Atendimento ao Cidadão |
| SGBD   | -Sistema Gerenciador de Bando de Dados            |
| SOLAP  | - Processamento Analítico Espacial em Tempo Real  |
| SIG    | -Sistema de Informação Geográfica                 |
| TI     | -Tecnologia de Informação                         |



# CAPÍTULO 1

## INTRODUÇÃO

### 1.1 Apresentação

Desde os tempos antigos os gestores de qualquer organização, principalmente na administração pública, trabalhavam com um grande número de dados referentes aos variados negócios da organização, porém esses dados eram armazenados em papel, o que dificultava muito a manipulação dos mesmos.

Com o avanço da tecnologia e o surgimento dos computadores, houve uma evolução e os dados anteriormente arquivados em papel passaram a ser arquivado em computadores, facilitando um pouco mais a vida dos tomadores de decisões, contudo ainda era um processo muito lento, pois eles não dispunham de ferramentas que trabalhassem e disponibilizassem esses dados de maneira eficaz e eficiente.

Para solucionar esse problema, surgiu os Sistemas Gerenciadores de Banco de Dados (SGBD), que tem como objetivo principal gerenciar o acesso e a correta manutenção dos dados armazenados em um banco de dados (BD), que de acordo com (INMON, 1997) corresponde a uma única fonte de dados para todo o processamento.

Com o BD e o SGBD foi possível melhorar a gestão dos dados, contudo a quantidade de dados foi aumentando, da mesma forma que a necessidade por obter informações estratégicas também aumentava. O BD solucionou o problema de armazenamento dos dados, porém não é capaz de gerar informação estratégica.

É sabido que, um fator primordial que os sistemas computacionais devem apresentar para que os gestores possam obter melhores resultados sobre suas decisões, é a capacidade de transformar dados em informação. Os dados em sua forma primitiva não fornecem nenhuma informação.

De acordo com (SOBRAL; SOUZA, 2006) “O processo realizado com o intuito de explorar e analisar informações estruturadas e específicas de um domínio para enxergar

tendências ou padrões e, através disso, produzir percepções e tirar conclusões sobre o mercado é denominado *Business Intelligence (BI)*”.

Surgiu então, decorrente da necessidade do armazenamento eficiente de grande quantidade de dados históricos de uma organização, sobre os quais serão feitas análises estratégicas para facilitar as tomadas de decisões, o *Data Warehouse (DW)*. Com essa tecnologia é possível oferecer aos tomadores de decisão maior precisão com relação aos dados que estes manipulam.

E com o passar do tempo as organizações foram expandindo seus horizontes, começando a trabalhar com dados geográficos, ou seja, dados que possuem localização na superfície da terra. No âmbito da administração pública trabalhar com dados geográficos é fundamental, todas as tomadas de decisões geradas sobre informações espacializada, tornam-se mais completas, pois oferece ao gestor uma visão geral do município, facilitando a formulação e monitoramento de políticas públicas.

Com a crescente demanda na utilização de dados geográficos surgiu os Sistemas de Informações Geográficas (SIG). Essa ferramenta possibilita a manipulação de dados geográficos e permite a realização de operações complexas de análises espaciais. Cerca de oitenta por cento de todos os dados armazenados em bases de dados das organizações tem um componente espacial (BADARD, 2009). Diante disso, os Bancos de Dados e os *Data Warehouses* convencionais foram estendidos para suportarem dados espaciais. Passando a existir os BDGs (Banco de dados Geográficos) e os DWGs (*Data Warehouses* Geográficos).

De acordo com (QUEIROZ, 2005) “a evolução dos SIGs credita-se ao fato dos mesmos passarem a ser usados em ambientes corporativos”. A utilização dos SGBDs esta consolidada nos SIGs permitindo o armazenamento tanto da geometria (ponto, linha, polígono) como dos atributos (dados alfanuméricos) dos objetos. Sendo assim, os Bancos de Dados Geográficos atualmente fazem parte de qualquer SIG, como parte de um nível mais interno do sistema, com a função de armazenar e recuperar as informações.

Diante do cenário acima descrito, torna-se necessário uma ferramenta de BI/DWG que ofereça instrumentos analíticos para que os dados alfanuméricos e geográficos sejam

armazenados de maneira a proporcionar consultas estratégicas, para que seja possível tomar decisões fundamentadas, tanto táticas quanto estratégicas. Viabilizando assim, um melhor atendimento ao cidadão, além de garantir uma visão global da situação da cidade para o prefeito e seus assessores, facilitando a gestão do município.

Portanto, o trabalho proposto baseia-se em avaliar o uso de uma ferramenta DWG, com o intuito de verificar o quão essa ferramenta pode auxiliar os gestores públicos na tomada de decisão estratégica, verificando se essa solução é viável para resolver o problema existente na Prefeitura de Belo Horizonte (PBH), que é a falta de uma ferramenta que integre informações estratégicas, alfanuméricas e geográficas, em um único ambiente.

## **1.2 Justificativa**

A Prefeitura possui um *Data Mart* do SACWeb (Sistema Informatizado de Atendimento ao Cidadão), que contém informações sobre solicitação realizadas pelos munícipes. Porém o DM do SACWeb não consegue disponibilizar de forma ágil, as informações estratégicas especializadas para a sala de situação, que atende diretamente às necessidades do prefeito. O DM não oferece as informações especializadas solicitada pelo prefeito. Todas as informações especializadas inerentes ao SACWeb, passam por um processo demorado e nem sempre confiável. A informação é extraída do *Data Mart*, um especialista utiliza um SIG para mapeá-las e fazer análises espaciais pertinentes, depois são elaborados relatórios, gráficos e mapas para serem entregues ao prefeito.

Diante dessa deficiência na PBH, com relação à falta de uma ferramenta de DWG, surgiu a ideia desse trabalho.

## **1.3 Objetivos**

### **1.3.1 Geral**

Este trabalho tem como objetivo geral avaliar a utilização de uma ferramenta de BI/DWG, com o intuito de verificar a capacidade desta ferramenta na geração de informações estratégicas especializadas que auxiliarão a tomada de decisões na PBH.

### **1.3.2 Específico**

Como forma de alcançar o objetivo geral foram definidos os seguintes objetivos específicos:

- Definir uma ferramenta de DWG para a execução do trabalho;
- Avaliar a viabilidade de implantação de uma solução de DGW no ambiente corporativo da PBH;
- Avaliar as funcionalidades suportadas pelo DWG, verificando se estas atendem às necessidades da PBH, referentes à obtenção de informações estratégicas especializadas ou não em uma única ferramenta;
- Avaliar se a utilização de uma ferramenta de DWG traria benefícios para a administração municipal.

## CAPÍTULO 2

### FUNDAMENTAÇÃO TEÓRICA

#### **2.1 *Business Intelligence* - BI**

O atual interesse pelo BI (Inteligência de Negócios) vem crescendo à medida que sua utilização possibilita às organizações realizar diversas análises e projeções, de forma a agilizar os processos relacionados às tomadas de decisão.

O termo *Business Intelligence*, segundo (Primak, 2008), faz referência ao processo inteligente de coleta, organização, análise, compartilhamento e monitoramento de dados contidos em um *Data Warehouse/Data Mart*, gerando informações para o suporte à tomada de decisões no ambiente de negócio.

Com a evolução natural da tecnologia o termo *Business Intelligence* aumentou sua abrangência e importância, embutindo várias ferramentas como:

- *Data Warehouses*
- Ferramentas de ETL
- *Data Marts*
- *Data Mining*
- Geradores de relatórios e consultas
- Ferramentas OLAP / EIS (Enterprise Information System) / DSS (Decision Support System)

Entre outras, que tem como objetivo dinamizar a capacidade de tomar decisões. É comum encontrarmos na literatura autores definindo BI e DW como sendo a mesma coisa.

#### **2.2 *Data Warehouse* e *Data Warehouse Geográfico***

Um *Data Warehouse* (armazém de dados) é composto por dados originários de diversas fontes de informação. Os *Datas Warehouses Geográficos* (DWGs) se comportam da mesma maneira de DWs tradicionais, o que os difere é o fato do DWG suportar objetos geográficos. Segundo (DEGGAU, 2010), um DWG combina dados e operadores do modelo dimensional com representações geométricas e operações para manipulação de dados geográficos. Por isso, todas as citações sobre DW, encontradas nesse trabalho,

aplicam-se também ao DWG, destacando algumas particularidades que estarão descritas no decorrer desse trabalho.

O DW é um sistema de suporte à decisão composto por um conjunto de ferramentas que centralizam, armazenam, gerenciam e extraem informações históricas da empresa, as informações são entregues aos usuários em forma de relatório, gráficos, ou mapas no caso de DWGs. Dessa forma, o tomador de decisão, apenas ao olhar o resultado da consulta, pode tomar alguma decisão sem ser necessária uma análise profunda. A informação é entregue de forma customizada.

Levando em consideração o DWG, torna-se necessário ter uma ferramenta robusta, que integre funcionalidades espaciais aos dados modelados no *Data Warehouse* Geográfico, sem essa ferramenta seria impossível obter informações estratégicas especializadas sem ter que gerar, manualmente, essas informações através de um SIG, o qual será detalhado mais adiante.

Deve-se dar importância, de acordo com (OLIVEIRA, 1998), a determinados princípios que devem ser observados na construção de um DW que também são aplicados ao DWG:

- O *Data Warehouse* deve oferecer acesso imediato aos dados da corporação ou organizacionais. Os gerentes e analistas devem conseguir estes dados através de seus PC's, de acordo com sua necessidade, e com alta performance. As ferramentas de extração de dados devem ser amigáveis.
- Os dados em um *Data Warehouse* devem ser consistentes. Duas requisições, por exemplo, sobre compras de fevereiro na unidade sul devem retornar o mesmo número. As definições dos dados estão disponíveis e são compreensíveis. Os analistas devem receber um aviso se os dados recuperados estão incompletos.
- Os dados em um *Data Warehouse* devem poder ser separados e combinados de todas as formas possíveis.
- O *Data Warehouse* não deve ser somente um conjunto de dados, mas também um conjunto de ferramentas para pesquisar, analisar e apresentar informações.

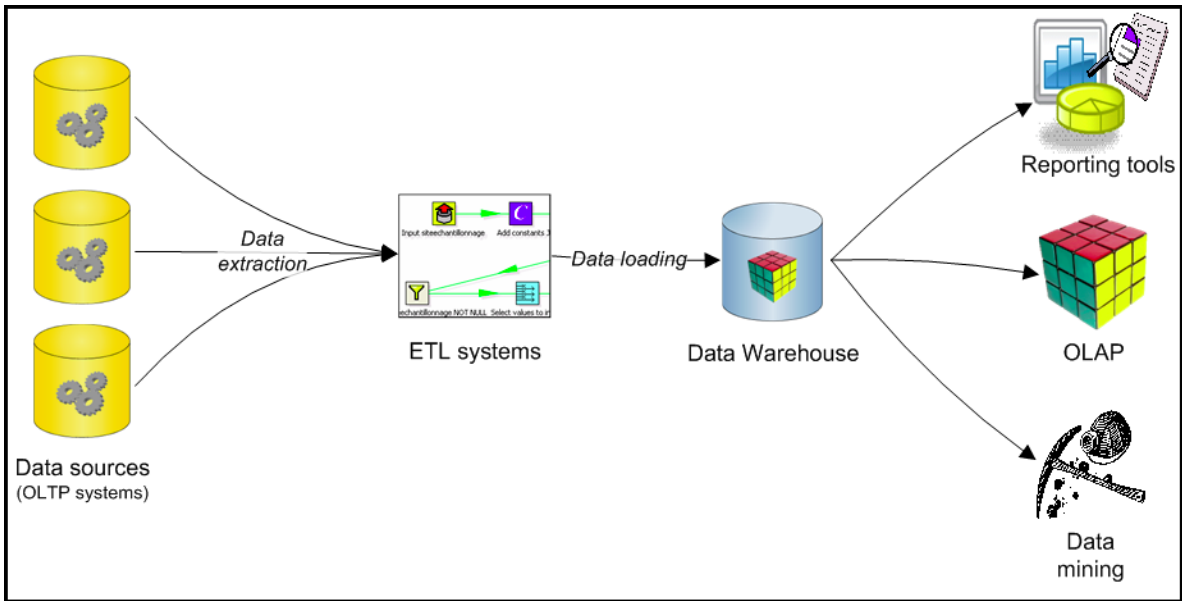
- Os dados devem ser gerados de uma variedade de fontes organizacionais, limpos, com qualidade assegurada, e recuperados apenas se servem para uso.
- A qualidade dos dados em um *Data Warehouse* deve servir como um direcionador para a reengenharia dos negócios. O *DW* não pode conter dados de qualidade pobre.

Segundo (INMON; HACKATHORN, 1997), os elementos de dados brutos que entram no *Data Warehouse* são oriundos, em quase todos os casos, do ambiente operacional não integrado. Portanto, os dados encontram-se em diferentes formatos, isso implica que estes deverão ser transformados e padronizados antes de serem carregados no *Data Warehouse*, esse processo é chamado de ETL (*Extract Transform Load*), ou seja, extração, transformação e carga no DW.

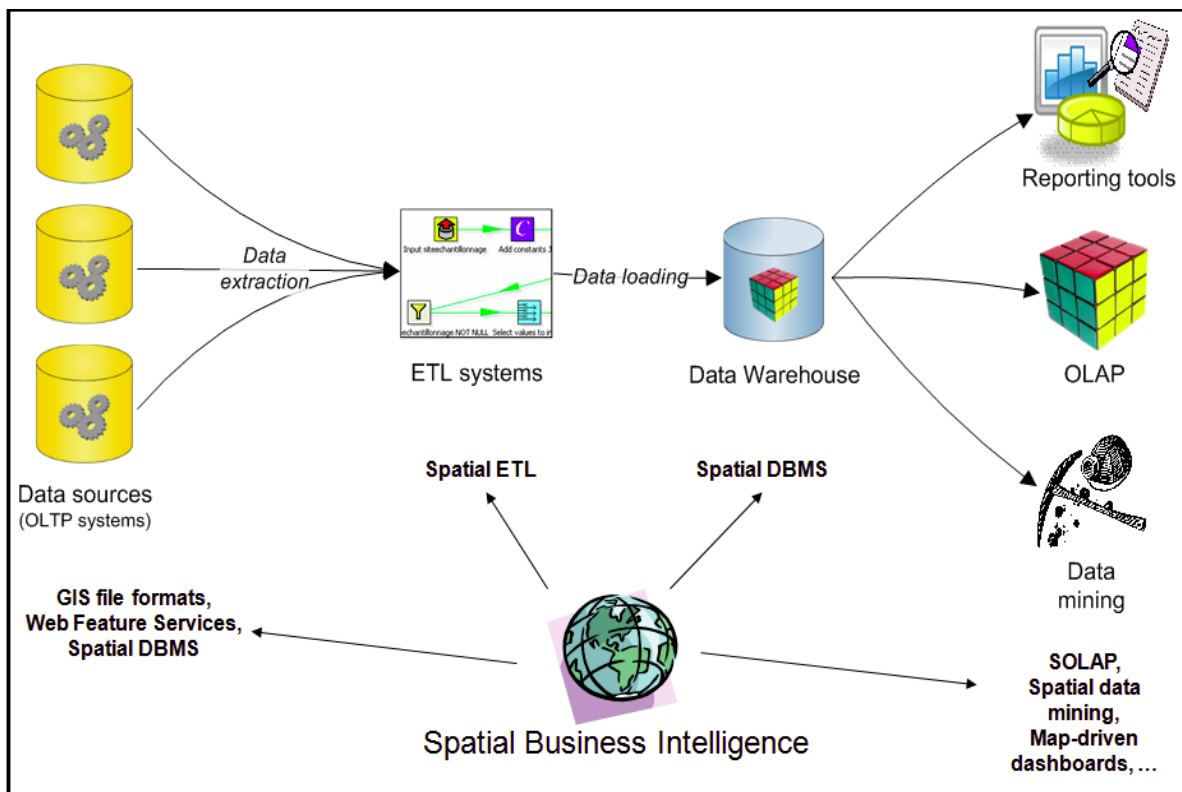
Um dos fatores que influenciam o sucesso dos *Data Warehouses* é a questão da integridade dos dados, é fundamental que os dados estejam padronizados antes de serem carregados no DW, afim de que as respostas obtidas através de uma consulta no banco, não estejam inconsistentes.

O processo ETL é o mais complexo e demorado na construção do *Data Warehouse*, o qual será explicado posteriormente. Depois da limpeza e padronização dos dados, eles serão carregados para uma única fonte de dados consolidada, o *Data Warehouse*, facilitando e tornando as consultas mais confiáveis e eficientes.

A figura 1 e 2 representam a visão geral de um DW e um DWG, respectivamente. Existem diferentes bases de dados operacionais, de onde são extraídos os dados necessários ao DW, são utilizadas ferramentas que realizam a fase de ETL, após as extração e a transformação dos dados, estes são carregados no *Data Warehouse* e então estará disponível para a realização de consultas. Esses procedimentos ocorrem tanto no DW convencional, quanto no DW geográfico, a diferença é que para cada ferramenta do DW convencional temos sua correspondente voltada para operações espaciais.



**Figura 1: Arquitetura de um DW**  
**Fonte: Adaptado de Badard, 2009**



**Figura 2: Arquitetura de um DWG**  
**Fonte: Adaptado de Badard, 2009**



As principais justificativas de se implantar a tecnologia de DW em uma organização é a existência de alguns fatores na empresa (MACHADO; ZORZIN , 2006), tais como:

- A existência e utilização de várias plataformas de *hardware* e *software*;
- As constantes alterações nos sistemas transacionais corporativos;
- As dificuldades acentuadas na recuperação de dados históricos em períodos superiores ao ano atual em que está sendo operadas as informações;
- A existência de programas de fornecedores diferentes;
- A falta de padronização e integração dos dados existentes dos diversos programas utilizados pela empresa;
- A falta de documentação e segurança no armazenamento dos dados;
- A dificuldade de reunir informações dos diversos sistemas existentes nas empresas.

Cenários que podem ser visualizados com a utilização de DW, de acordo com (SOUZA, 2005):

- Estudar o rumo que a empresa deve tomar (como se focar em um produto ou serviço, A ou B);
- Analisar ameaças eminentes (como fortes concorrentes já conhecidos ou possíveis novos concorrentes entrantes no mercado);
- Avaliar oportunidades escondidas (como padrões de venda de produtos no clássico exemplo das vendas de cervejas e fraldas no *Walmart*);
- Realizar comparações históricas (como vendas no mês de maio de 2005 comparadas às vendas de maio 2004);

Cenários que podem ser visualizados utilizando um DWG:

- Em qual região está concentrado o maior índice de roubos a mão armada, no ano de 2010;

- Exibir a população dos estados brasileiros, ordenando estes em relação à menor distância tomando como referência algum outro estado;
- Obter a população de cada sub-região do estado brasileiro de maior área no ano de 2010;
- Obter o número de nascidos vivos e de óbitos maternos nos municípios adjacentes a região metropolitana de Belo Horizonte;

Segundo (INMON, 1997), *Data Warehouse* é um conjunto de dados baseado em assuntos, integrado, não volátil, e variável em relação ao tempo, que servem de apoio às decisões gerenciais.

- **Orientado ao assunto:** faz referência aos assuntos primordiais da organização, por exemplo: clientes, produtos, atividades, endereço. Cada assunto será implementado como tabelas, chamadas de dimensões, que será explicada posteriormente. A figura 3 mostra que em bancos de dados transacionais, a orientação por assunto é feita de acordo com as aplicações existentes, já no *Data Warehouse* tem-se o banco de dados divididos pelos principais assuntos da empresa.

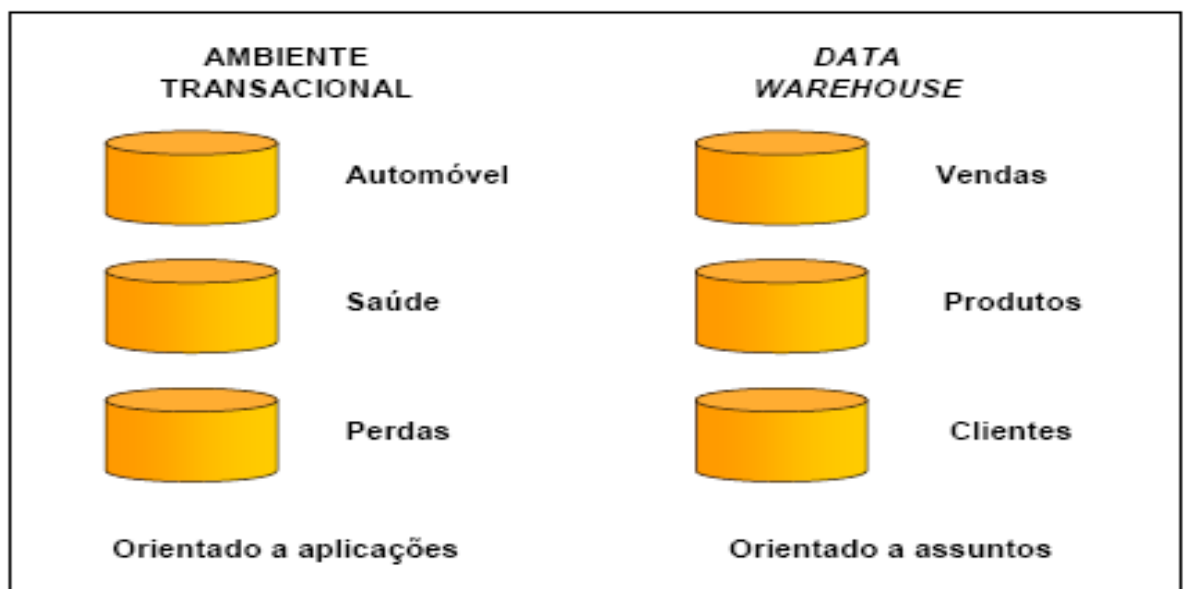


Figura 3: Orientação por assunto.  
Fonte: Adaptado de Machado *apud* Zorzini 2006

- **Integrado:** Como foi dito, o DW precisa ter os dados padronizados, para que haja a integração dos mesmos. É necessário que estes sejam previamente tratados antes de serem armazenados de forma definitiva. Exemplos de critérios que podem ser executados para conseguir essa consistência dos dados são:
  - Padronização no formato de dados, ou seja, quando se vai referenciar o sexo de um indivíduo, deve-se escolher uma forma padrão para descrever isso, utiliza-se masculino e feminino, homem e mulher, ou ainda M para Masculino e F para Feminino.
  - Determinar medidas comuns<sup>1</sup>. Existem medidas de atributos, como por exemplo, um caminho qualquer, em metros, polegadas, pés e jardas, pode-se definir como forma padrão, que a medida de todo caminho será em metros.

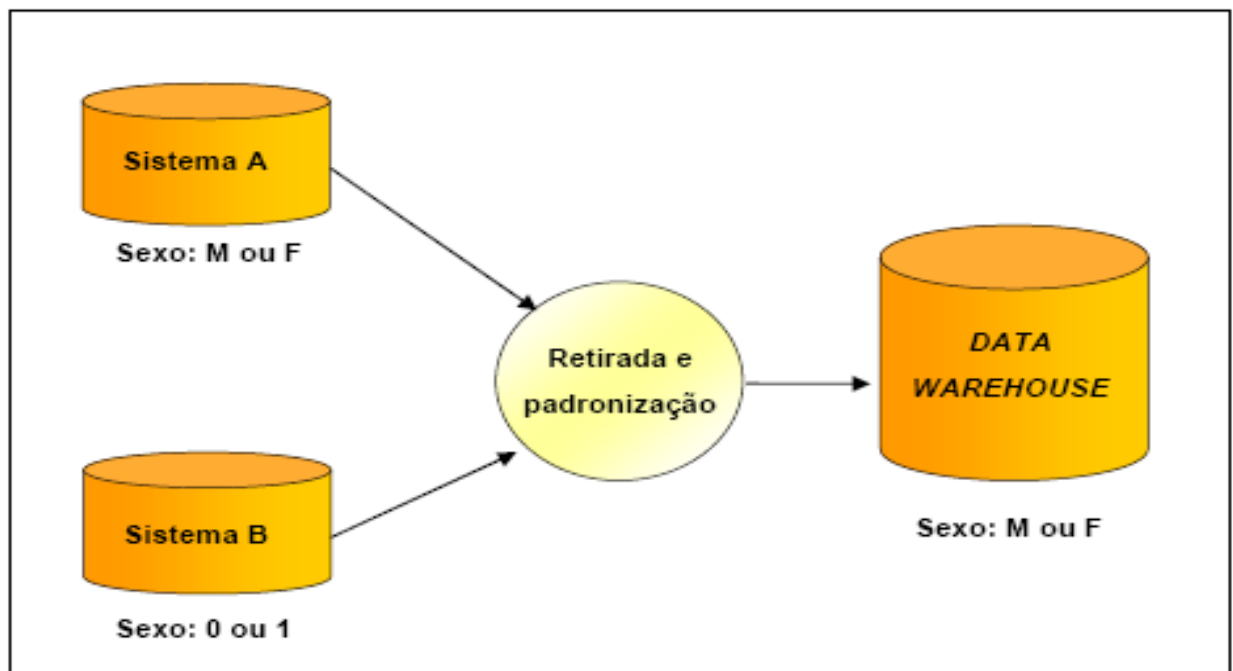
Assim, toda forma de padronização de dados deve ser utilizada com o intuito de manter a consistência e integração dos mesmos.

A figura 4 demonstra a situação de integração, onde os dados referentes ao sexo dos clientes são armazenados nos bancos de dados transacionais em formatos diferentes.

No sistema A, masculino e feminino está representado pelas letras “M” ou “F” respectivamente e no sistema B, utiliza-se “0” ou “1”. É por falta de integração como estas, que os dados precisam ser tratados e padronizados, a fim de serem carregados no *Data Warehouse* de forma integrada e concisa.

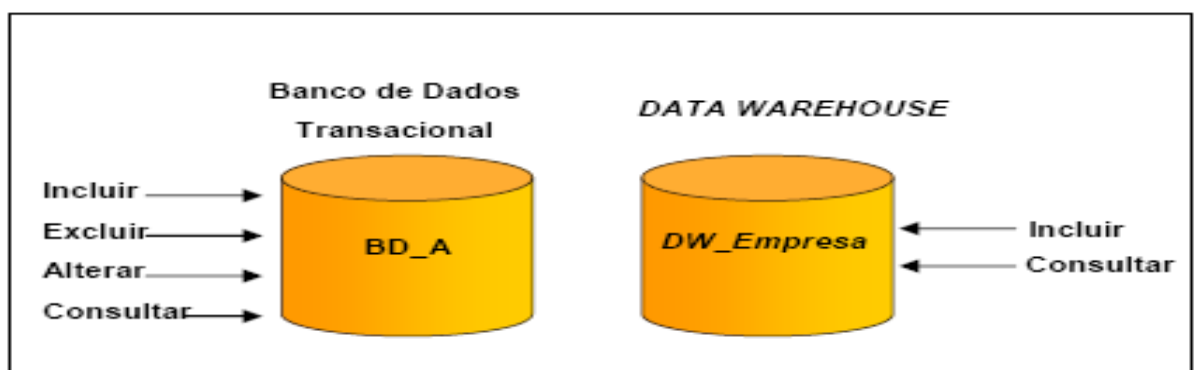
---

<sup>1</sup>Medidas comuns: determinar ou avaliar a grandeza, extensão ou quantidade de alguma coisa.



**Figura 4: Integração de dados.**  
 Fonte: Adaptado de Machado *apud* Zorzin, 2006

- **Não volátil:** Dizer que um dado é não volátil significa que uma vez que esse dado é carregado no DW ele nunca mais será apagado ou alterado, umas das características do *Data Warehouse* é que os dados somente são consultados e carregados, isso permite a afirmação da principal funcionalidade de um DW, a de ser um banco de dados histórico e de possuir um grande volume de dados. A figura 5 representa as operações realizadas em bancos de dados transacionais onde é permitido fazer inclusões, exclusões, alterações e consultas e em um *Data Warehouse*, onde somente é permitido a inclusão e consulta ao banco.



**Figura 5: Não volátil no DW.**  
 Fonte: Adaptado de Machado *apud* Zorzin, 2006

- **Variável com relação ao tempo:** O elemento tempo é fundamental em um DW. Esse aspecto permite aos tomadores de decisão analisarem alguma informação, em um determinado momento no tempo, estando essas mesmas informações com os mesmos resultados em qualquer outra consulta que fosse realizada posteriormente. Vale lembrar que essa façanha não é possível em um sistema de dados operacionais baseados em transações conhecida como OLTP (*On Line Transaction Processing*), no item 2.3 será descrito com mais detalhes esse sistema.

*“Um Data Warehouse é variável com o tempo devido ao fato de que os dados contidos no mesmo referem-se a algum momento específico, o que significa dizer que o dado quando carregado recebe, na sua chave, uma unidade de tempo e nunca é atualizado. Este tipo de armazenamento permite que os analistas de negócios façam análise de tendências, podendo visualizar as variações de uma determinada informação ao longo do tempo”* (INMON,GONÇALVES, 2003).

### **2.3 OLTP versus OLAP**

OLTP – *On Line Transaction Processing* (Processamento de Transação em Tempo Real) são sistemas encarregados de armazenar todas as transações efetuadas em determinada operação dentro da organização, assim os dados utilizados pelo sistema OLTP são transacionais.

Porém, ao utilizar esse processo, a cada nova demanda de informações que algum usuário precisasse efetuar, e que fossem diferentes das que existiam nos relatórios pré-definidos, tornava-se necessário a programação, por parte dos responsáveis pela área de TI (Tecnologia da Informação), de novos relatórios, que pudessem satisfazer às necessidades dos usuários.

Para tornar esse processo mais produtivo e eliminar essa necessidade de envolver os técnicos e analistas a cada nova demanda que surgia com relação à geração de relatórios, e também com o advento do *Data Warehouse*, surgiu o chamado OLAP – *On Line*

*Analytical Processing* (Processo Analítico em Tempo Real) que são ferramentas que concedem aos usuários finais a capacidade deles próprios gerarem relatórios de acordo com a necessidade de cada operação.

A ferramenta OLAP utiliza-se de recursos chamados “cubos” que transformam os dados em informações rápidas e confiáveis. A figura do cubo é capaz de representar várias dimensões de dados inter-relacionados, própria de um sistema multidimensional, como lembra (ROSA, 2004).

É esse conceito de cubo que permite aos usuários a facilidade de construir seus próprios relatórios. A aplicação torna-se mais dinâmica, pois o usuário tem a possibilidade de enxergar através de um cubo todas as informações que deseja. Diferentemente do OLTP, o OLAP utiliza dados agregados, ou seja, dados consolidados, característicos de um *Data Warehouse*, isso é primordial, pois melhora a performance das consultas, já que diminui o número de registros que terão de ser recuperados.

Idêntica à ferramenta OLAP, no *Data Warehouse* Geográfico criou-se a Spatial OLAP. Segundo (BARROS, 2007), a ferramenta SOLAP sintetiza as características presentes nas ferramentas OLAP, DW e SIG - Sistema de Informação Geográfica (item 2.9).

*“O objetivo final de uma ferramenta OLAP é transformar dados em informações capazes de dar suporte às decisões gerenciais de forma amigável e flexível ao usuário e em tempo hábil”* (THOMSEM; ROSA 2004).

Outra forma que os usuários têm para visualizarem os dados contidos no DW, além da mencionada acima, é utilizar outra ferramenta de apoio a decisão chamada *Data Mining* que visa buscar correlações escondidas em altos volumes de dados. O *Data Mining* também pode ser aplicado ao DWG.

*“Ferramentas OLAP são ferramentas de caráter descritivo, ou seja, permitem a apresentação dos dados do negócio sob diferentes perspectivas e implementam funcionalidades de agregação, razão e produto, entre outras. Ferramentas de Data Mining, por outro lado, são ferramentas de caráter explicativo e implementam*

funcionalidades de regressão, extração de associações e agrupamentos entre outras” (THOMSEN; SELL 2006).

## 2.4 Estrutura do Data Warehouse

Em um *Data Warehouse* existem diferentes níveis de detalhes. Existe um nível de detalhe mais antigo, um nível corrente de detalhe, um nível de dados levemente resumidos - o *Data Mart* é um nível de dados altamente resumidos (INMON, 1.997). A figura 6 ilustra essa estrutura.

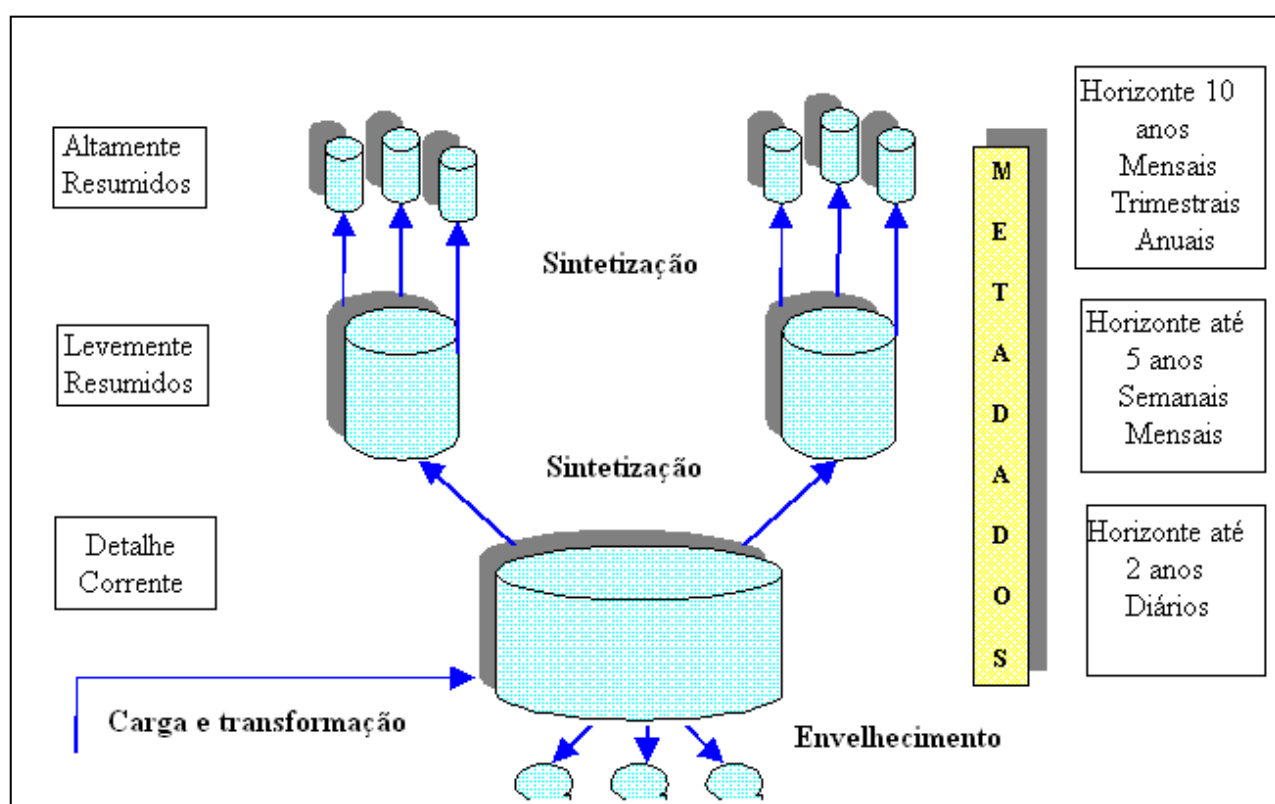


Figura 6: Estrutura e fluxo de dado do *Data Warehouse*.  
Fonte: adaptada de Filho, 2008

Segundo (INMON, 1.997), os dados que são considerados antigos, passam do detalhe corrente para o detalhe mais antigo. Assim à medida que os dados são resumidos, passam do detalhe corrente para os dados levemente resumidos e, finalmente, os dados levemente resumidos passam para os dados altamente resumidos.

(GONÇALVES, 2003) descreve melhor esses componentes:

- Dados antigos: ou dados históricos, são acessados com uma baixa frequência, portanto ficam armazenados em nível de detalhe consistente com o detalhe dos dados atuais;
- Dados atuais (correntes): são os que necessitam de maior atenção, pois são os dados que diz respeito às operações realizadas mais recentemente pela organização. Eles ocupam um espaço maior, porque estão armazenados no menor nível granularidade;
- Dados levemente resumidos: são encontrados no nível atual de detalhe e extraídos do nível mais baixo;
- Dados altamente resumidos: são compactados e de fácil acesso.

O metadado é um componente importante, é ele que ajuda a encontrar os demais dados em um *Data Warehouse*.

*“Metadados são as informações que descrevem os dados utilizados por uma organização, isto envolve informações como descrições de registros, comandos de criação de tabelas, diagramas Entidade Relacionamento (ER), dados de um dicionário de dados. É necessário que exista uma grande variedade de metadados no ambiente de arquitetura do Data Warehouse para que ele mantenha sua funcionalidade e os usuários não precisem se preocupar onde residem os dados ou a forma com que eles estão armazenados” (FELÍCIO, 2005).*

Um *Data Warehouse* é composto por dados de acordo com os principais assuntos de uma organização (por exemplo, cliente, produto, transações), nesse sentido, (FELÍCIO, 2005) diz que a principal área de interesse acaba tornando-se, fisicamente em um DW, uma série de tabelas relacionadas. É importante salientar que para os DWG, pelo menos uma dimensão precisa possuir um campo geográfico. Os *Data Warehouses* não são construídos de uma só vez, não se consegue construir um DW com todas as informações que são necessárias, os custos são elevados e os recursos de informações não são suficientemente precisos quando se visualiza um *Data Warehouse* de maneira global.



Levando isso em consideração, torna-se viável a construção de *Data Marts* (DM).

## **2.5 Data Mart**

Implementar um ambiente por inteiro de *Data Warehouse* é muito complexo e também causa um impacto expressivo no ambiente operacional, por essa razão muitas organizações optam por essa arquitetura de *Data Marts*.

De acordo com (ZIULKOSKI, 2003), “um *Data Mart* é um DW de tamanho reduzido e de baixo custo, projetado para atender apenas a uma área da organização, mantendo informações sobre apenas um assunto (vendas, por exemplo)”.

*“Muitas organizações iniciam o processo a partir de uma área específica, que normalmente é uma área carente de informação e cujo trabalho seja relevante para os negócios da empresa, criando o chamado Data Mart, que pode ser representado como um Data Warehouse departamental, para depois ir crescendo aos poucos”* (LONEY; FELÍCIO, 2005).

Existem dois tipos de implementação de *Data Marts*:

- O top-down: é quando a organização cria um *Data Warehouse* e divide-o em *Data Marts* orientado à assunto (departamentos) e;
- Bottom-up: cria-se *Data Marts* referentes aos assuntos (departamentos) da organização de maneira a expandir até virar um *Data Warehouse* que abrange todos os assuntos da mesma.

Da mesma forma que temos um DW geográfico, podemos ter também, de acordo com a necessidade, o *Data Mart* Geográfico.

## **2.6 Granularidade**

A granularidade é um ponto fundamental que deve ser observado cuidadosamente na hora de se projetar um *Data Warehouse* ou *Data Marts*. “A granularidade diz respeito ao nível

de detalhe ou de resumo contido nas unidades de dados existentes no *Data Warehouse*” (INMON, 1.998). Quanto mais detalhes, mais baixo o nível de granularidade e vice versa. Recebe o nome de grão o nível mais detalhado do dado.

A granularidade é uma questão delicada em um projeto de DW, pois ela determina o volume de dados que o *Data Warehouse* vai possuir, afetando assim os tipos de consultas que podem ser realizadas. Segundo (INMON, 1.998), quando existe um nível muito baixo de granularidade, torna-se possível responder a qualquer consulta. Entretanto, com um alto nível de granularidade, o número de questões que os dados podem satisfazer é limitado.

## **2.7 Projetando um *Data Warehouse/Data Warehouse Geográfico***

Para a criação de um DW não é suficiente que se faça apenas a extração de dados das bases de dados operacionais e a inserção dos mesmos no *Data Warehouse*. Deve-se lembrar que os dados operacionais não são integrados, pois as aplicações existentes antes do grande advento dos DWs, não havia a necessidade de integração dos dados, não se imaginava que algum dia ocorreria essa integração e que esta seria tão importante.

Portanto, não é tão simples, como se pensa, fazer essa migração de dados. Como dito anteriormente, um mesmo dado pode estar descrito em diferentes lugares, com diferentes nomes. Fazer essa extração de dados, estando eles em diferentes lugares e sem nenhuma integração é muito complexo. Essa fase é um tormento para os projetistas de DW.

E pensando em facilitar o trabalho dos projetistas, empresas voltadas para a área de Tecnologia de Informação, desenvolveram ferramentas capazes de realizar esse processo de maneira mais eficiente e eficaz, tornando o processo menos desgastante, como ele o é quando era realizado sem o auxílio de alguma ferramenta.

No *Data Warehouse* exclui-se a idéia, que se tinha dos projetos de bancos de dados relacionais utilizando o modelo entidade – relacionamento (MER), ou seja, distribuição de campos por entre tabelas. Agora existe como modelagem lógica para um *Data Warehouse* a chamada modelagem dimensional.

## 2.8 Modelagem Dimensional

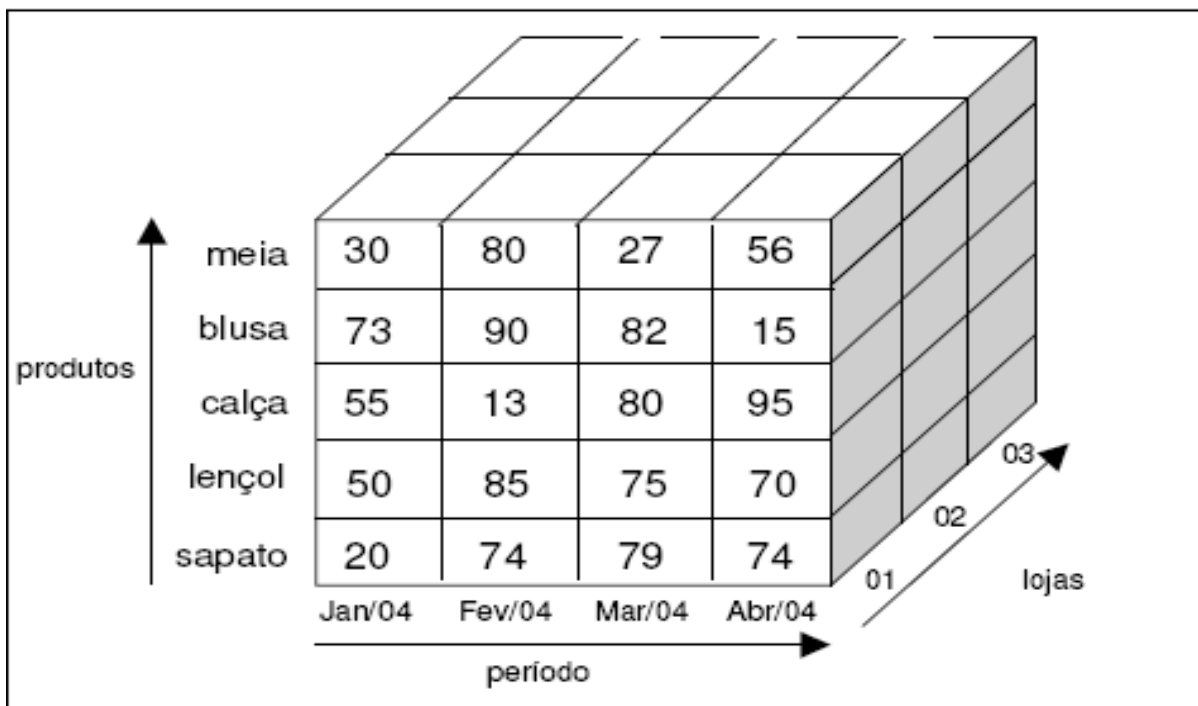
O modelo dimensional busca padrões para apresentação de dados que seja de fácil visualização por parte do usuário final e que tenha bom desempenho e agilidade nas consultas, o que torna o modelo muito mais vantajoso para o projeto de um *Data Warehouse*, como lembra (SILVA, 2006).

Como os dados do MER são divididos em tabelas, que são normalizadas para terem o mínimo de redundância, o tempo de resposta é relativamente alto às consultas realizadas no banco, o modelo acima citado torna-se inviável para um projeto de *Data Warehouse*, visto que o mesmo preza por um bom desempenho e agilidade em suas consultas.

Pelo fato de *Data Warehouse* possuir um grande volume de dados, o tempo de resposta a alguma requisição feita pelo usuário é maior. Portanto para melhorar o desempenho das consultas, planeja-se redundância dos dados, criando-se então uma estrutura simples, com modelos que simulam o processo de análise de negócios (HOKAMA, 2004).

Tanto no *Data Warehouse* convencional quanto no *Data Warehouse* geográfico a representação dos dados é estruturada como um cubo, onde cada célula do cubo armazena os valores que foram medidos e os lados do cubo são as dimensões. A idéia do cubo é utilizada para facilitar a visualização do ambiente multidimensional do DW. As consultas são feitas de acordo com a combinação de diferentes e quaisquer dimensões. Pode-se ter mais de três dimensões em um cubo, estes passarão a ser chamados de Hiper cubos.

A figura 7 ilustra o exemplo de um cubo com três dimensões: loja, produto e período.



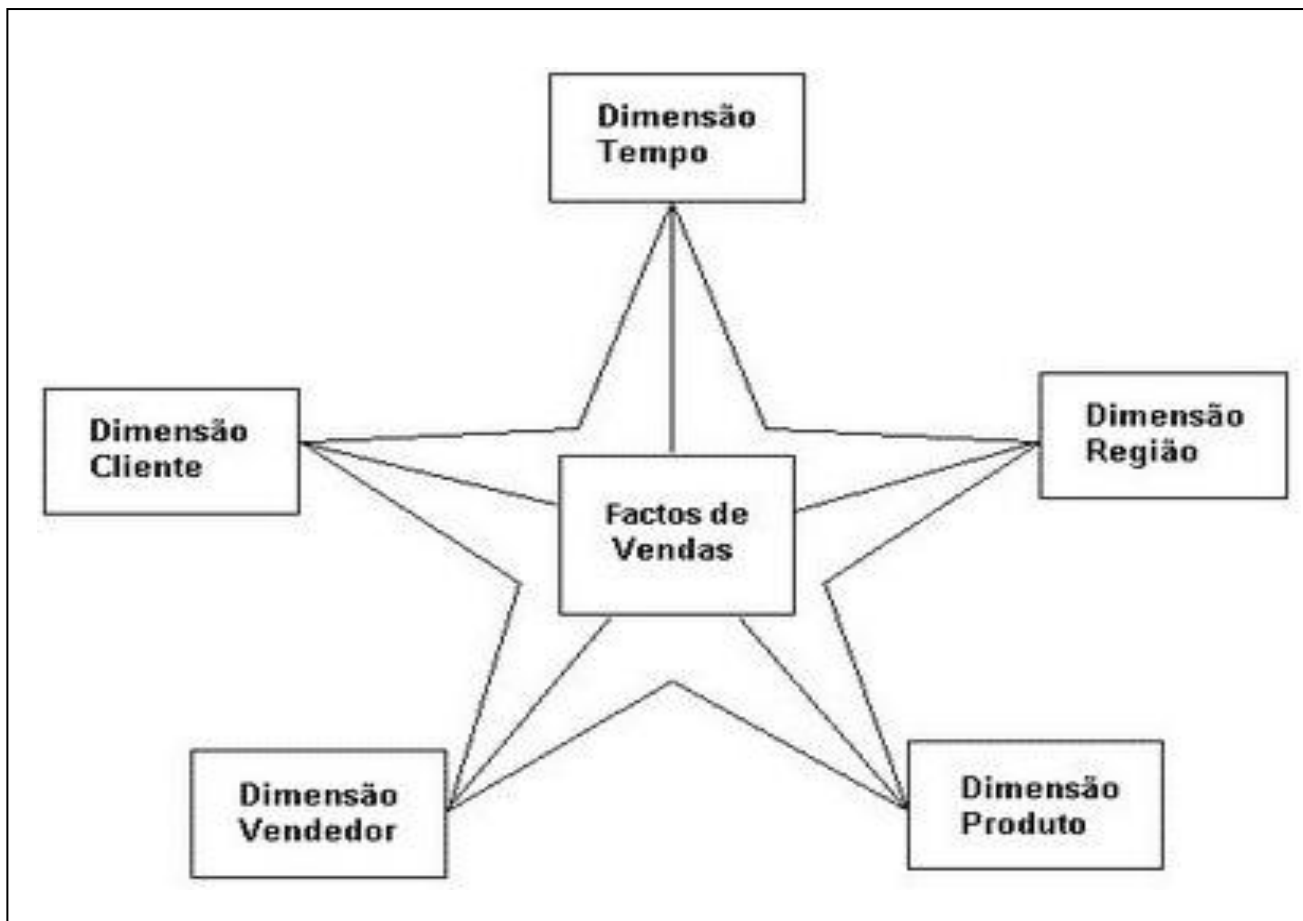
**Figura 7: Cubo tridimensional com o total de vendas**  
 Fonte: adaptada de Campus *apud* Araujo & Batista, 2008

Abaixo será apresentado dois dos modelos mais utilizados para se modelar os dados em um *Data Warehouse* convencional e geográfico.

### 2.8.1 Modelo Estrela (*Star Schema*)

De acordo com (ZORZIN, 2006) esse modelo, por ser mais rápido em suas consultas, é o mais utilizado na modelagem de um DW. O fator que justifica essa agilidade nas consultas, é que as dimensões que compõem o modelo são desnormalizadas, ou seja, ela permite que os dados sejam repetidos na base de dados, dessa forma garante-se a simplicidade e o desempenho nas consultas. Então é normal que se repita dados em diversas dimensões, garantindo assim um maior desempenho.

O Modelo Estrela vai conter uma tabela denominada Fato, que fica ao centro do modelo e diretamente ligada a essa tabela Fato temos as Dimensões, formando assim uma estrela. A figura 8 representa o modelo citado.



**Figura 8: Representação do modelo Estrela.**  
**Fonte: adaptado de Afonso, 2008**

Na figura, logo acima, temos uma tabela Fato central denominada venda, o conteúdo dessa tabela responde a perguntas do tipo:

- Valor da venda realizada e prevista;
- Quantidade de itens realizados e previstos;
- Preço médio de vendas;
- Margem de venda.

E ligadas a essa tabela Fato vendas, temos as dimensões discriminadas: Tempo, Região, Produto, Vendedor e Cliente. Essas dimensões reúnem os atributos que serão utilizados para qualificar as consultas e cujos valores vão servir para agrupar as métricas que formarão a tabela Fato.

A tabela Fato contém atributos chaves e métricas ou fatos numéricos, ligadas as tabelas dimensões através das chaves. Deve-se verificar se cada métrica da tabela Fato se relaciona com todas as dimensões definidas. Para povoar a tabela Fato faz-se a seguinte pergunta: Quais informações serão analisadas? E para as tabelas Dimensões: Como serão analisadas?

Vale lembrar que o modelo pode ter mais de uma tabela Fato.

O modelo Estrela oferece algumas vantagens importantes, segundo (SINGH; ZORZIN, 2006):

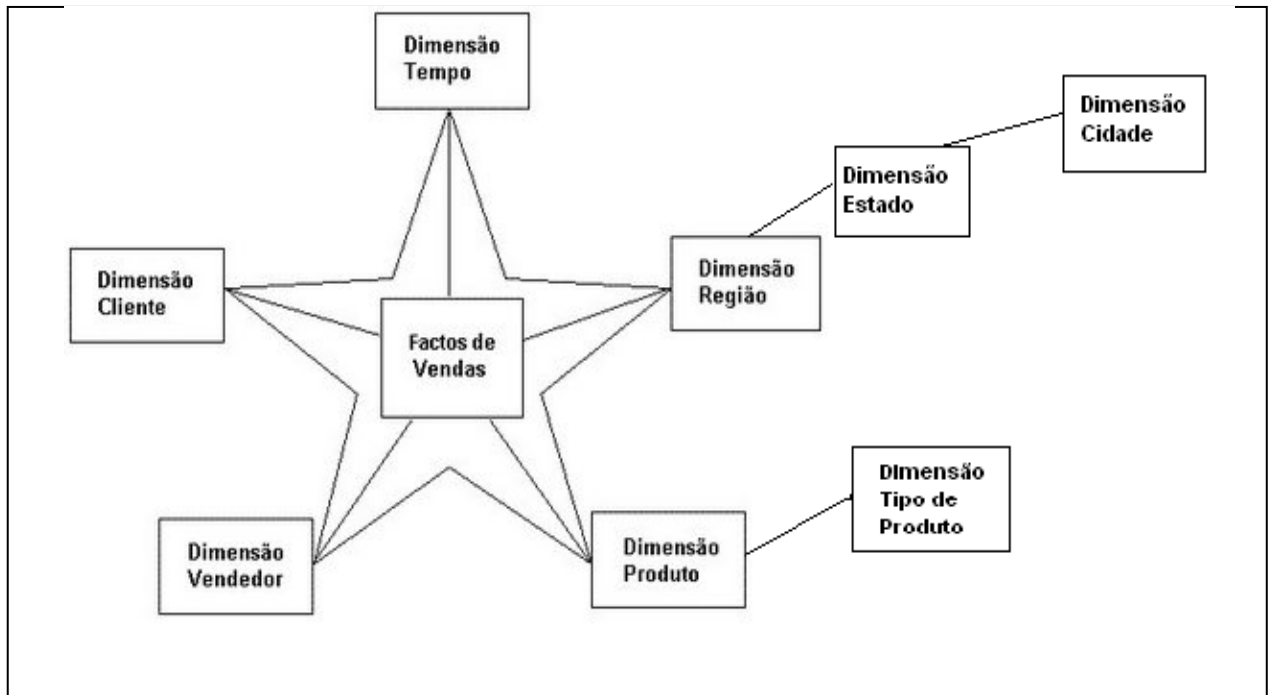
- Permite criar uma estrutura multidimensional complexa com um modelo de dados simples;
- Facilita a definição de relacionamentos hierárquicos dentro de cada dimensão e facilita as consultas entre muitas tabelas;
- Reduz a quantidade de consultas feitas entre tabelas fisicamente, melhorando o desempenho;
- Ao simplificar a visualização do modelo de dados, diminui a chance de os usuários fazerem consultas erradas que consomem a máquina e trazem resultados duvidosos;
- Permite a expansão e a evolução do DW com pouca manutenção.

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

Segundo Silva (2006), o modelo Floco de Neve é a decomposição de uma ou mais dimensões, que utiliza como referência o modelo Estrela, usado para demonstrar os seus relacionamentos hierárquicos. Se você normalizar os dados das tabelas dimensionais de um modelo Estrela, este se transformará no modelo Floco de Neve. Esse modelo possui algumas desvantagens com relação ao modelo Estrela, segundo (SINGH; ZORZIN, 2006):

- A estrutura dos dados normalizada torna-se mais complexa;

- A navegação executada sobre o modelo se torna mais difícil, por ter que passar por muitas tabelas para chegar ao resultado;
- A carga e a manutenção dos dados no DW acabam sendo mais difícil de ser executado e dificulta a administração desse processo.



**Figura 9: Representação do modelo Floco de Neve.**  
 Fonte: adaptado de Afonso, 2008

Através desse modelo, acima representado, podem-se definir relacionamentos entre dimensões o que não é possível no modelo Estrela, porém a perda de performance por causa do aumento do número de tabelas faz esse esquema ser menos viável do que o modelo Estrela.

### 2.8.3 A tabela de Fatos

Segundo (SILVA, 2006), “A tabela de fatos é a tabela que armazena os fatos das dimensões para possibilitar a análise das informações”. Esta tabela é constituída por chaves compostas, é a única que possui esta chave. É pela chave composta que é feita a ligação entre a tabela Fato e as Dimensões (Figura 10). Para povoar a tabela Fato faz-se a seguinte pergunta ao cliente (usuário): Quais informações serão analisadas?

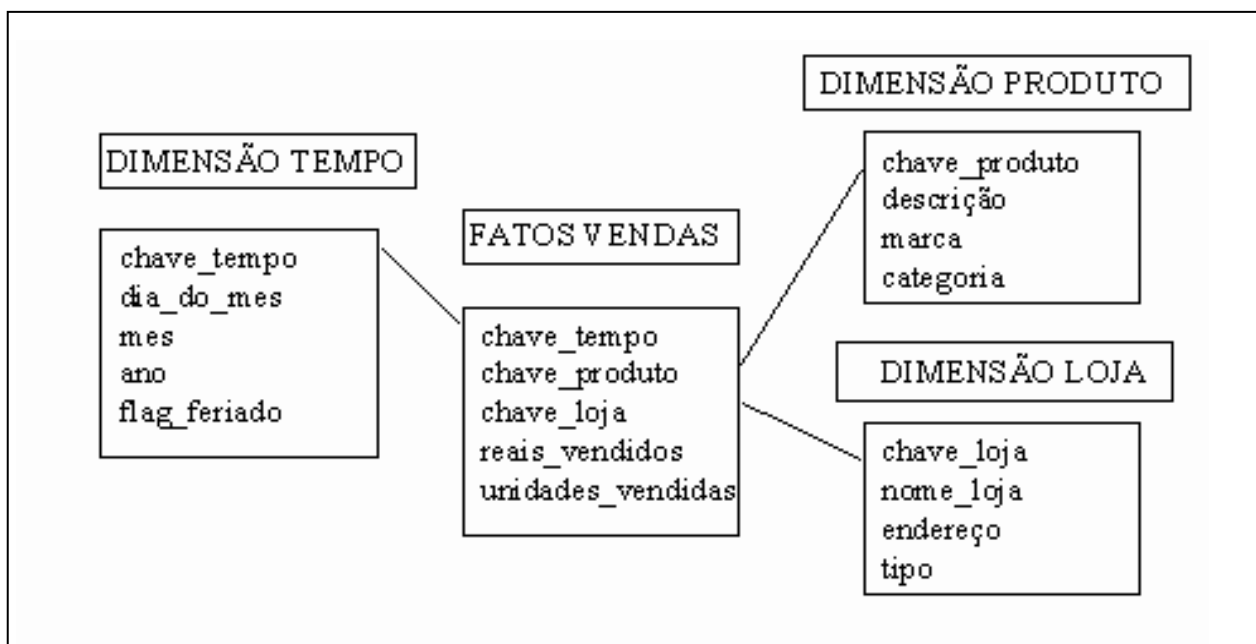


Figura 10: Exemplo de tabela de Fatos com chave composta e Dimensões.  
 Fonte: adaptada de shammas, 2008

#### 2.8.4 Tabela de Dimensões

Essas tabelas descrevem lugares, objetos, pessoas, períodos relacionados a um negócio. Cada dimensão possui sua granularidade, é ela quem determina o nível de detalhes que uma consulta pode ter. As Dimensões possuem chaves primárias que se relacionam à chave composta correspondente da tabela Fato.

Uma Dimensão importantíssima e que nunca pode faltar em um modelo de *Data Warehouse/ Data Warehouse Geográfico* é a Dimensão Tempo, pois é ela que permite a validação de uma das principais características de um DW, que é o fato de ser variante no tempo. Vale ressaltar que para um DW ser geográfico, deve existir um campo espacial em pelo menos uma dimensão.

Pode-se fazer a seguinte pergunta para encontrar dados que serão colocados nas dimensões: Como as informações serão analisadas?



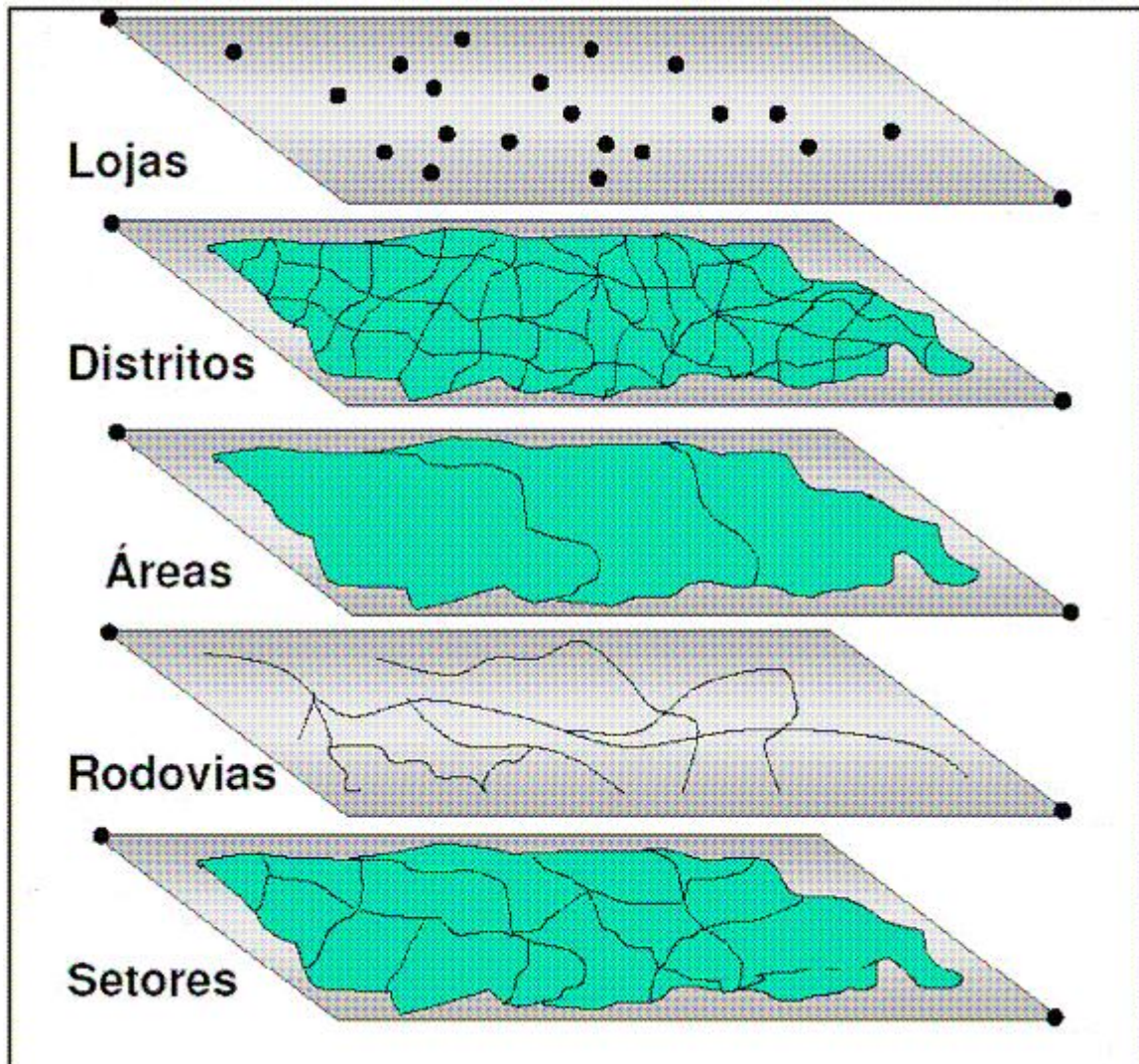
## 2.9 Sistemas de Informação Geográfica (SIG) e DWG

Os Sistemas de Informações Geográficas podem ser entendidos como instrumento de análise, visualização e interpretação de dados geográficos. A principal diferença de um SIG para um sistema de informação convencional é sua capacidade de armazenar tanto os atributos descritivos como as geometrias dos diferentes tipos de dados geográficos, ou seja, dados que representam objetos e fenômenos em que a localização geográfica é uma característica inerente à informação. “Os dados geográficos são abstrações geométricas e topológicas dos fenômenos do mundo real. Ponto, linha e polígono são exemplos de tipos de dados espaciais” (SIQUEIRA, 2009).

De acordo com (CÂMARA, 2005), as principais características dos SIGs são:

- Inserir e integrar, numa única base de dados, informações espaciais provenientes de meio físico-biótico, de dados censitários, de cadastros urbano e rural, e outras fontes de dados como imagens de satélite e GPS;
- Oferecer mecanismos para combinar as várias informações, através de algoritmos de manipulação e análise, bem como para consultar, recuperar e visualizar o conteúdo da base de dados geográficos.

Em geral, o comportamento das feições geográficas é analisado visualmente com o auxílio de mapas. Segundo (BARROS, 2007), “Mapa é a denominação dada à representação de uma seleção de características abstratas da superfície terrestre, em uma dada escala e sobre uma superfície plana. Um recurso que os SIGs utilizam para visualizar diferentes tipos de informações geográficas é a de sobreposição de temas ou camadas. Com este recurso, cada espaço geográfico pode ter diversos temas de dados georeferenciados, sendo um para cada tipo de informação espacial a ser representada”, como mostra a figura 11.



**Figura 11: Exemplo de Sobreposição de Camadas.**  
**Fonte: Adaptado de Barros 2007.**

A tecnologia de DWG reúne, em uma só ferramenta, todas as características de um DW convencional e um SIG. Inúmeros são os benefícios dessa ferramenta, no que diz respeito ao apoio às tomadas de decisões gerenciais. O DWG resume-se em uma base de dados que herda características do DW, SIG e OLAP. Desta forma o *Data Warehouse Geográfico* possibilita consultas analíticas multidimensionais com atributo geográfico.

## CAPÍTULO 3

### ESTUDO DE CASO - SACWeb

Neste capítulo é apresentado o processo atual de geração de informações estratégicas relacionadas ao SACWeb na PBH, além dos resultados obtidos e as dificuldades enfrentadas no decorrer do trabalho.

A metodologia adotada para atingir os objetivos desse trabalho, segue os seguintes passos:

- Definição e estudo da ferramenta DWG;
- Instalação da ferramenta e seus componentes;
- Instalação do Banco de Dados Postgres/Postgis;
- Conexão da ferramenta de DWG com o Postgis;
- Escolha do estudo de caso e solicitação da amostra de dados aos órgãos responsáveis;
- Importação dos dados no Postgis;
- Elaboração do modelo de dados do *Data Mart* Geográfico do SACWeb;
- Execução da parte prática do projeto.

#### 3.1 SACWeb

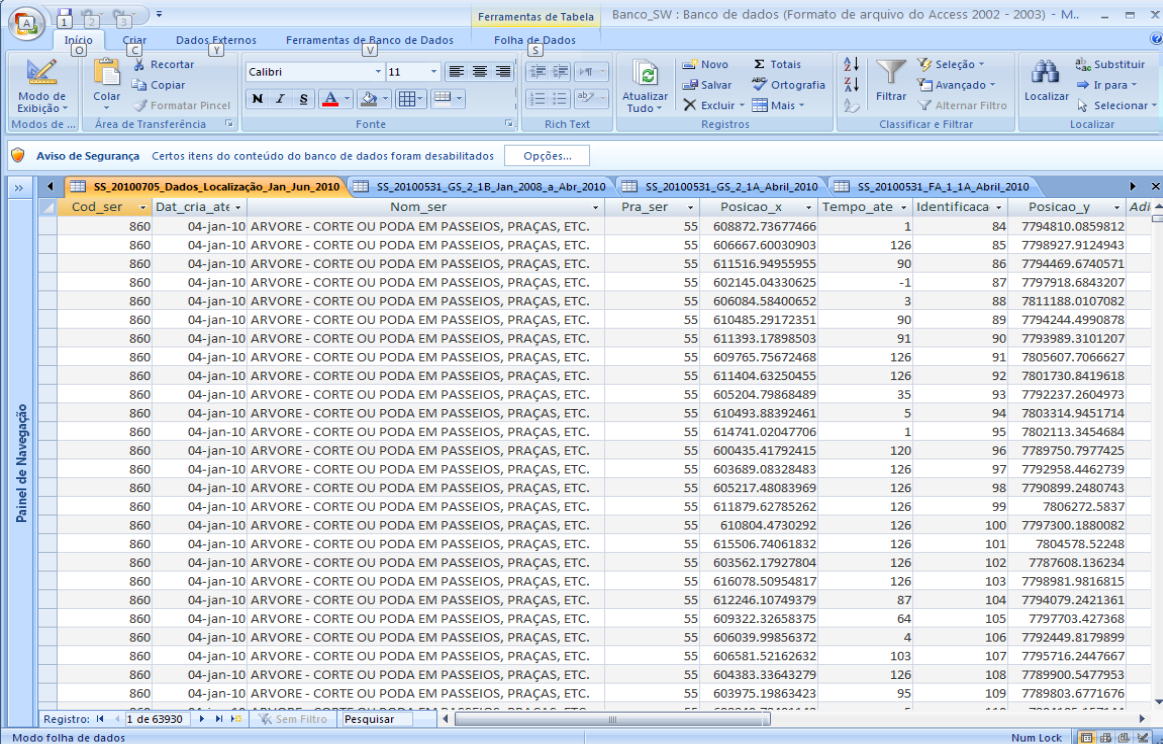
A PBH tem implantado o SACWeb - Sistema Informatizado de Atendimento ao Cidadão. Esse sistema é responsável pelo registro e gestão da captação, andamento, tramitação, conclusão e da resposta às solicitações de serviços pelo cidadão. Também já tem implantado o *Data Mart* do SACWeb que foi construído utilizando o modelo Estrela.

O item 3.2 apresenta como é realizado, atualmente, a geração de informações estratégicas especializadas através do SACWeb.

### 3.2 Geração de informações estratégicas espacializadas através do SACWeb na PBH

Todos os dados referentes ao SACWeb estão armazenados no banco de dados Oracle, a ferramenta de BI utilizada é a Discoverer, porém, como essa ferramenta não possui extensão espacial, o trabalho de geração de informações georreferenciadas é feito em outra ferramenta, tornando o processo mais trabalhoso e suscetível a erros.

O processo de geração das informações é o seguinte: cada demanda feita pelo prefeito, relativa às informações relacionadas ao SACWeb, passa pelo setor responsável pela manutenção do SACWeb, onde serão extraídos os dados, que atendem à determinada demanda. A figura 12 mostra uma tabela com dados extraídos do Discoverer (ferramenta de BI da Oracle), que foi aberta no Microsoft Access. Estas tabelas são enviadas ao setor responsável por fazer o georreferenciamento das informações.



| Cod_ser | Dat_cria_ate | Nom_ser  | Pra_ser | Posicao_x       | Tempo_ate | Identificaca | Posicao_y       | Adi |
|---------|--------------|--|---------|-----------------|-----------|--------------|-----------------|-----|
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 608872.73677466 | 1         | 84           | 7794810.0859812 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 606667.60030903 | 126       | 85           | 7798927.9124943 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 611516.94955955 | 90        | 86           | 7794469.6740571 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 602145.04330625 | -1        | 87           | 7797918.6843207 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 606084.58400652 | 3         | 88           | 7811188.0107082 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 610485.29172351 | 90        | 89           | 7794244.4990878 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 611393.17898503 | 91        | 90           | 7793989.3101207 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 609765.75672468 | 126       | 91           | 7805607.7066627 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 611404.63250455 | 126       | 92           | 7801730.8419618 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 605204.79868489 | 35        | 93           | 7792237.2604973 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 610493.88392461 | 5         | 94           | 7803314.9451714 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 614741.02047706 | 1         | 95           | 7802113.3454684 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 600435.41792415 | 120       | 96           | 7789750.7977425 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 603689.08328483 | 126       | 97           | 7792958.4462739 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 605217.48083969 | 126       | 98           | 7790899.2480743 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 611879.62785262 | 126       | 99           | 7806272.5837    |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 610804.4730292  | 126       | 100          | 7797300.1880082 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 615506.74061832 | 126       | 101          | 7804578.52248   |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 603562.17927804 | 126       | 102          | 7787608.136234  |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 616078.50954817 | 126       | 103          | 7798981.9816815 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 612246.10749379 | 87        | 104          | 7794079.2421361 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 609322.32658375 | 64        | 105          | 7797703.427368  |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 606039.99856372 | 4         | 106          | 7792449.8179899 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 606581.52162632 | 103       | 107          | 7795716.2447667 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 604383.33643279 | 126       | 108          | 7789900.5477953 |     |
| 860     | 04-jan-10    | ARVORE - CORTE OU PODA EM PASSEIOS, PRAÇAS, ETC. | 55      | 603975.19863423 | 95        | 109          | 7789803.6771676 |     |

Figura 12: Dados do SACWeb, com o campo de coordenadas geográficas  
Fonte: Prodabel

Todas as tabelas enviadas passam por um filtro, onde são selecionadas as informações que farão parte de um mapa temático. Essa operação é muito trabalhosa e demanda muito tempo, já que é feita através de consultas SQL, que têm que ser geradas manualmente.

Todo o processo de geração dos mapas é feito em um SIG. Destaca-se a importância de recursos humanos especialistas na ferramenta e em análise espacial. Somente é possível construir mapas se as tabelas possuírem campos com coordenadas geográficas. A figura 13 exemplifica este trabalho no mapa, representando a densidade acumulada do serviço de corte e poda de árvore através de pontos que são as coordenadas geográficas de cada solicitação.

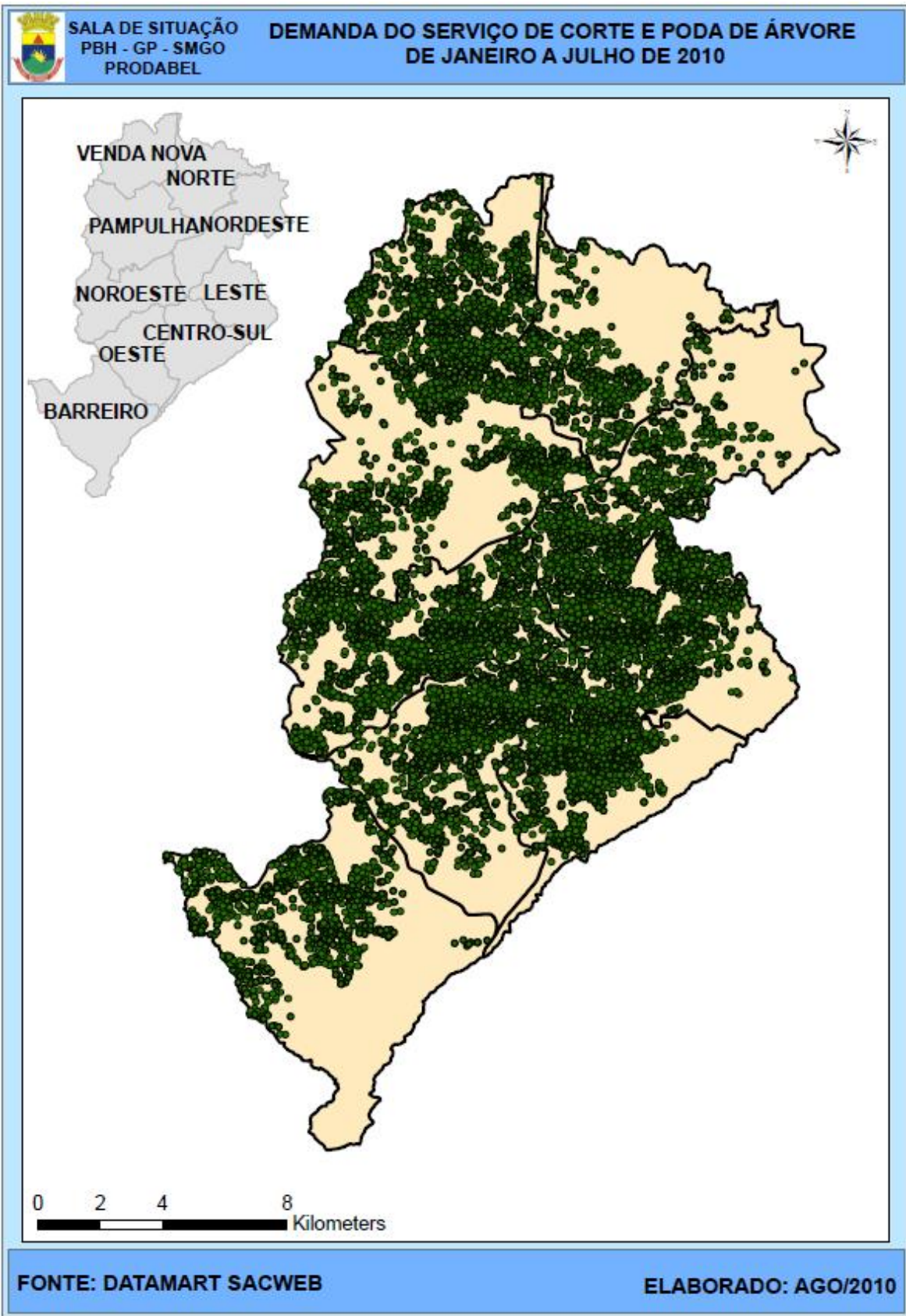


Figura 13: Mapa gerado no Arcgis com os pontos representando a coordenada geográfica  
Fonte: Prodabel

A figura 14 apresenta os mesmos pontos gerados no mapa acima, porém foi utilizada uma operação do Arcgis (Densidade de Kernel) para facilitar a análise das informações contidas no mapa.

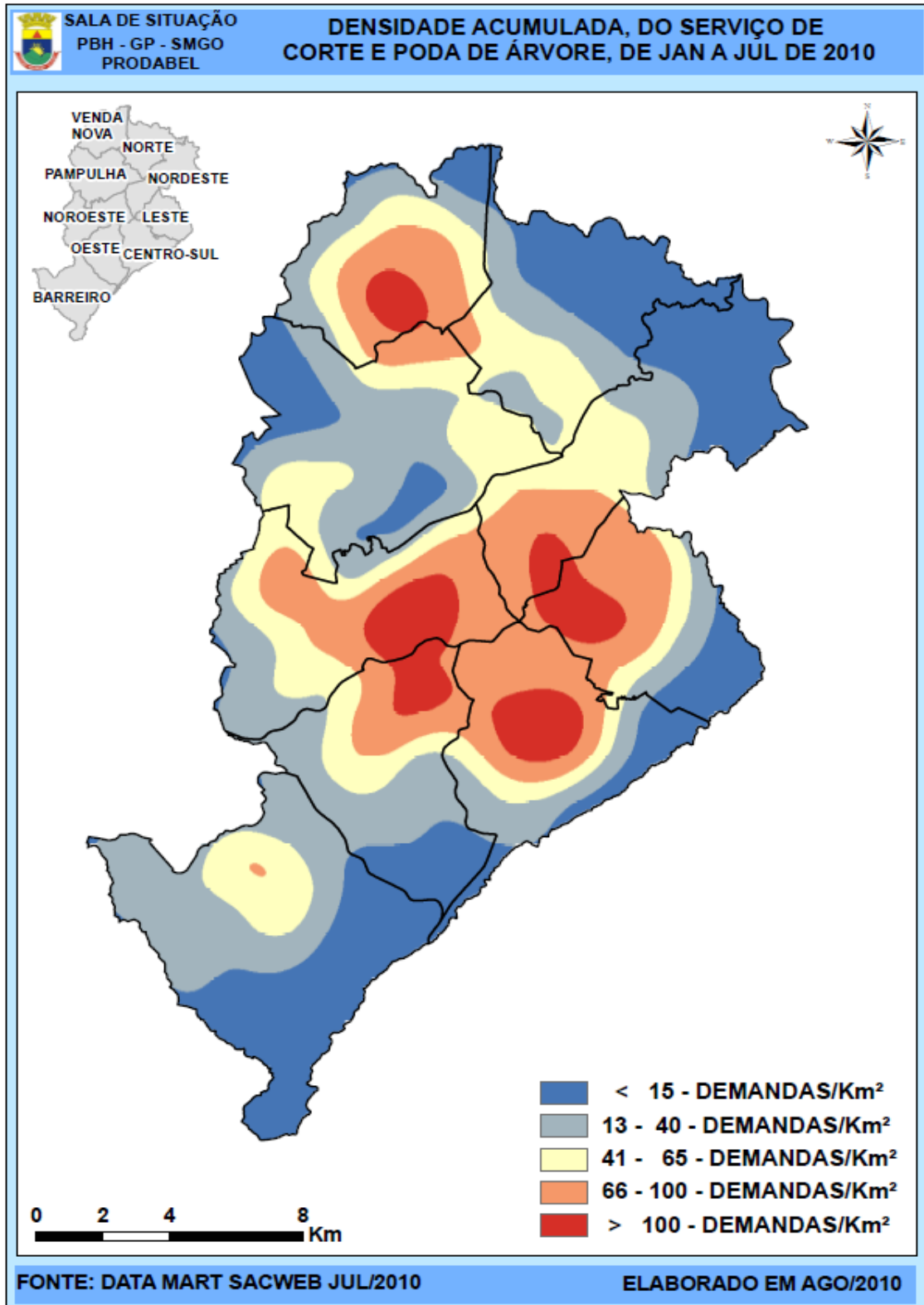


Figura 14: Mapa gerado no Arcgis  
Fonte: Prodabel

A título de curiosidade a figura 15 mostra um mapa de localização das árvores em logradouros públicos de Belo Horizonte.

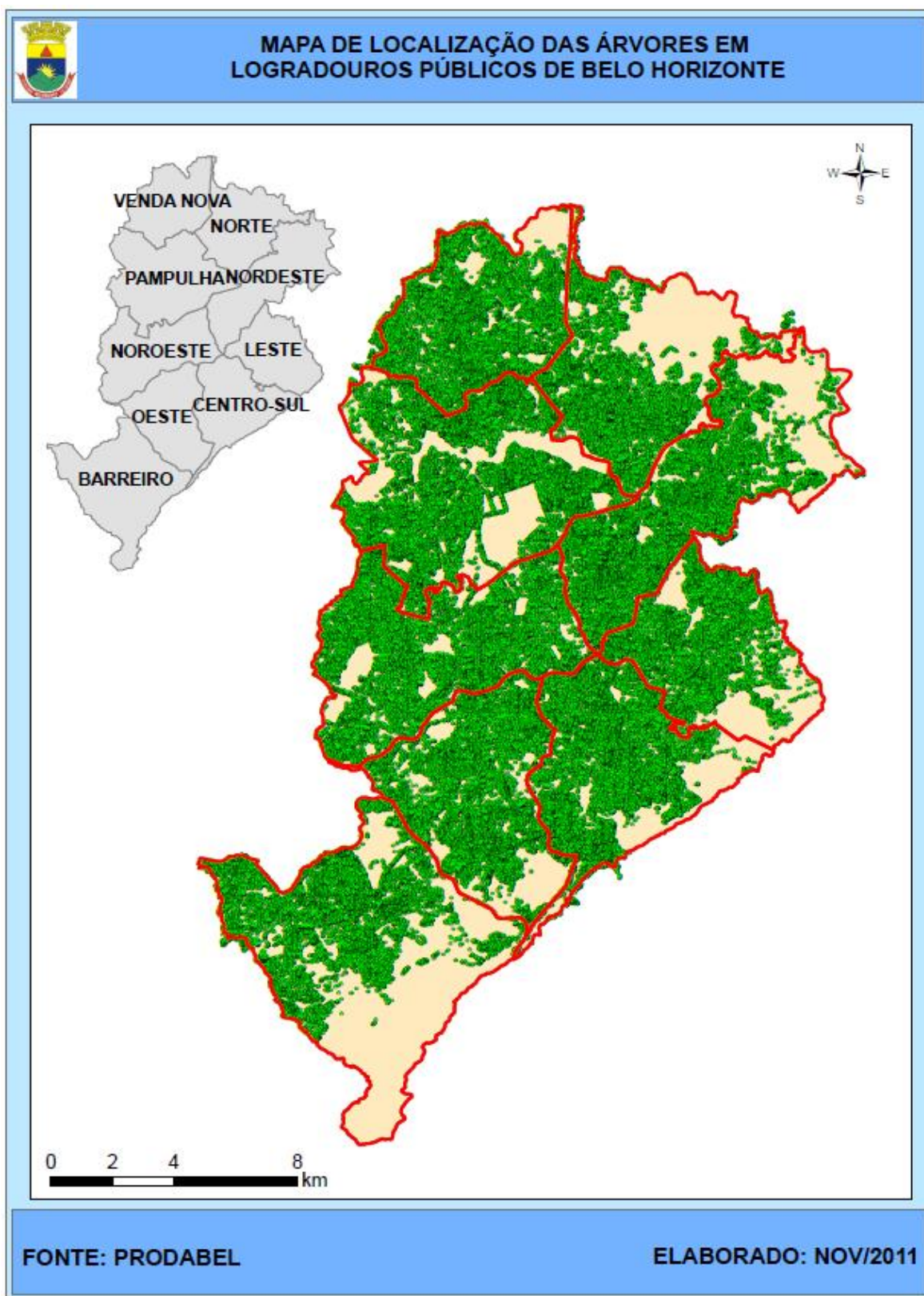
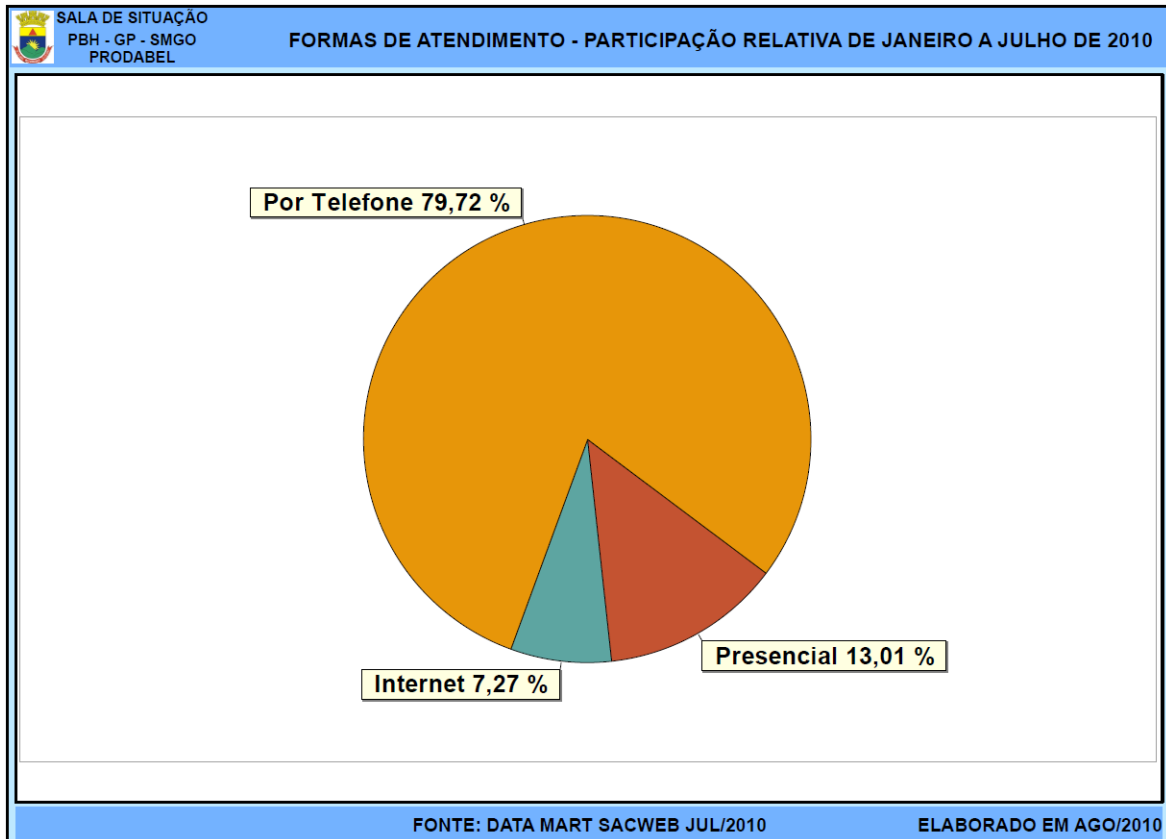


Figura 15: Mapa com a localização das árvores em logradouros públicos de BH  
Fonte: Prodabel



Também são construídos gráficos. A figura 15 mostra um gráfico de pizza com as formas de atendimentos e as porcentagens de cada forma, em um determinado período de tempo.



**Figura 16: Gráfico gerado no Arcgis**  
**Fonte: Prodabel**

É importante salientar que, os mapas ou gráficos gerados atualmente, são estáticos, ou seja, os gestores só terão a visão da informação da forma que elas foram geradas, não sendo possível gerar nenhuma informação dinamicamente. Para cada nova informação, é preciso passar por todo o processo novamente. Todas as informações constantes no SACWeb são atualizadas mensalmente.

Utilizando uma ferramenta de DWG, as informações poderiam ser vistas da mesma maneira que foram mostradas nas figuras 13,14 e 15, porém, esse trabalho contínuo e demorado que é feito hoje, seria evitado. Além, é claro, da ferramenta possibilitar várias outras formas de análises e visualizações das informações estratégicas reunidas em um só local.

### 3.3 Obtenção dos dados para a execução do trabalho

Os dados utilizados para a realização desse trabalho são oriundos do SACWeb da PBH. Foi solicitado à Secretaria Municipal Adjunta de Modernização a autorização para a utilização dos dados do *Data Mart* do SACWeb, posteriormente, a Gerência de Arquitetura de Informações – GAIS, que mantém esses dados, foi contatada para que fosse fornecido a amostra dos dados. Foi disponibilizada uma amostra de dados dos meses de novembro e dezembro de 2010. Os dados adquiridos já tinham passado pelo processo de transformação, ou seja, os dados já foram saneados e padronizados. Não será necessário utilizar a ferramenta de ETL.

### 3.4 Pentaho

Para atingir os objetivos propostos nesse trabalho, optou-se pela ferramenta Pentaho, por ser uma ferramenta de BI *Open Source* que possui componentes que suportam dados geográficos capazes de dinamizar a capacidade de tomar decisões.

Soluções proporcionadas pelo Pentaho:

- Geração de relatórios empresariais;
- Processos de ETL (Extração, Transformação e Carga);
- Análise de informações (OLAP);
- Painéis (*dashboards*) para controle gerencial;
- Mineração de dados (*Data Mining*);
- Workflow.

Os módulos mais importantes do Pentaho são:

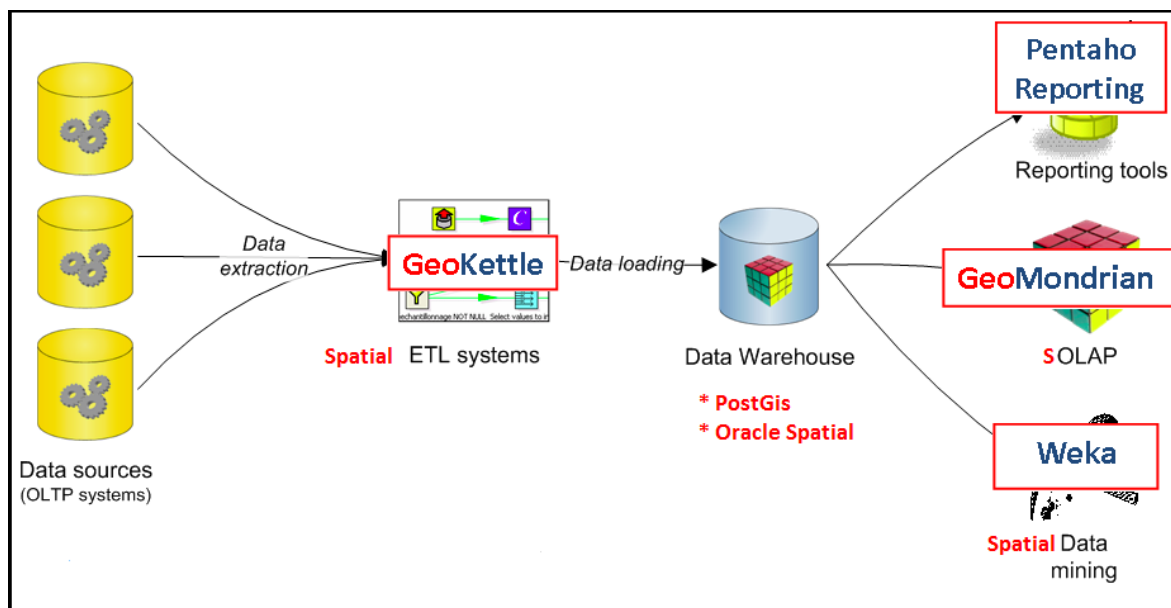
- *Pentaho BI Platform*: suíte para criação de soluções de *Business Intelligence*;
- *Pentaho Data Integration* (PDI): módulo que executa a extração, transformação e carga de dados no Pentaho. Também conhecida como *Kettle*;
- *Analysis View*: ferramenta para visualização de dados dimensionalmente;
- *Mondrian*: ferramenta para geração do modelo dimensional do DW;

- *Pentaho Reporting*: capaz de gerar vários tipos de relatórios;
- *Weka*: ferramenta para realizar a mineração de dados.

Esses módulos são todos independentes, podendo ser instalados e utilizados integrados à outra ferramenta de BI.

Na versão 3.5, utilizada nesse trabalho, o *Pentaho BI Platform*, o *Analysis View* e o *Pentaho Reporting* vêm integrados no mesmo pacote "bi-server". Para adquirir as outras ferramentas é necessário o *download* individual e a configuração do ambiente. O Pentaho pode ser utilizado no Windows, Linux, Mac, BSD, Solaris, etc.

O Pentaho possui módulos específicos para se trabalhar com dados geográficos, são as mesmas ferramentas convencionais oferecidas no mercado, porém elas são estendidas para dados espaciais e recebem outros nomes conforme figura 16.



**Figura 17: Pentaho Geoespacial.**  
**Fonte: Adaptado de Badard, 2009.**

Segue abaixo a descrição de cada componente espacial do Pentaho:

O módulo GeoKettle realiza a Extração, Transformação e Carga (ETL) de dados alfanuméricos e geográficos de sistemas OLTPs, como shapefile, tabelas do Access, KML, Excel, etc.

O módulo GeoMondrian é uma máquina para suportar consultas SOLAP (Spatial OLAP), escrita em Java, executando consultas a ele submetidas na linguagem MDX (multi-dimensional expressions ou expressões multidimensionais) lendo dados de um SGBD com suporte a dados geográficos, como por exemplo, PostGis e Oracle Spatial.

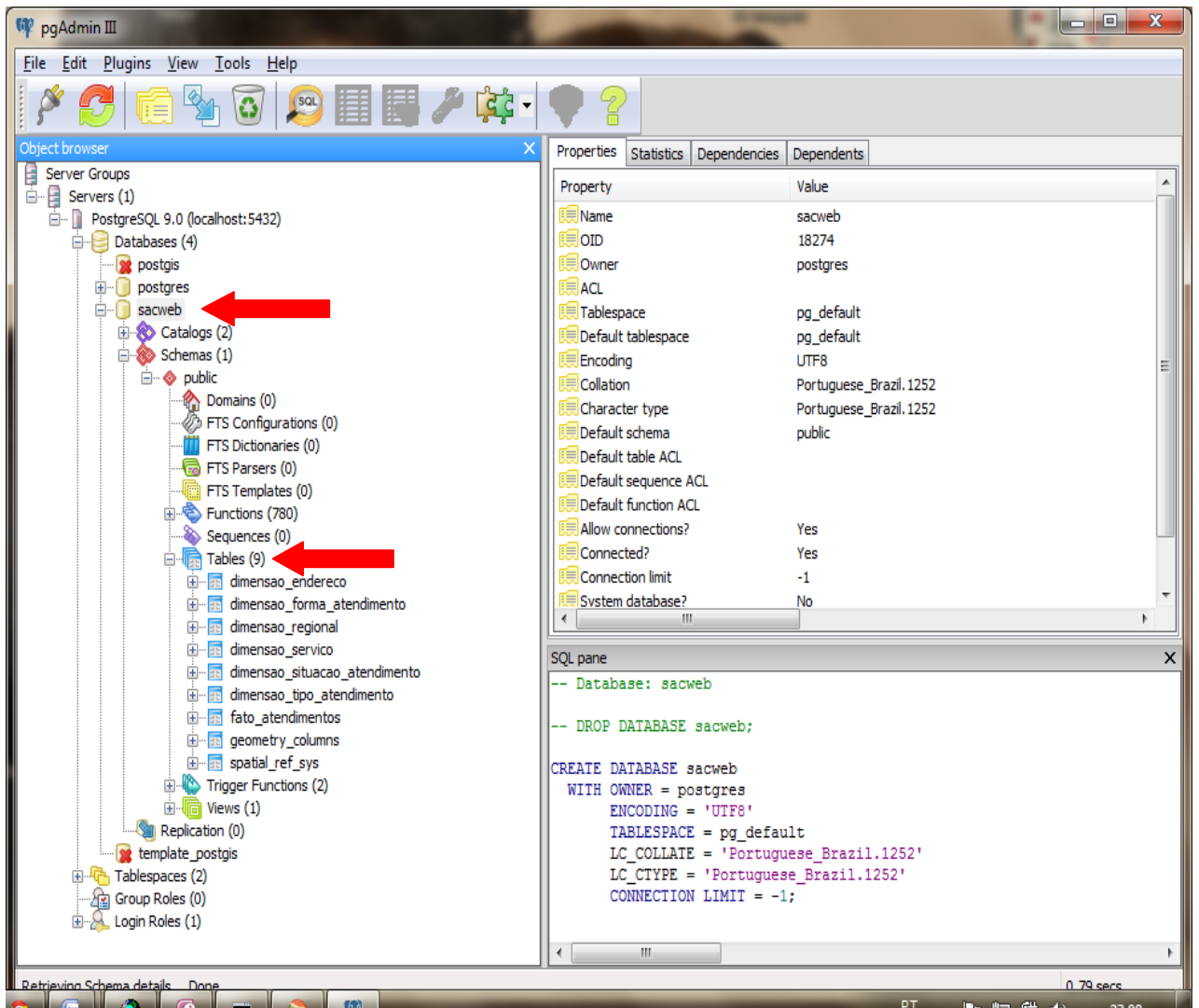
O módulo Weka realiza a mineração de dados alfanuméricos ou geográficos.

O módulo Pentaho *Reporting* é utilizado na geração de relatórios acerca das informações obtidas nas consultas ao DWG/DM existente.

Maiores informações sobre a ferramenta Pentaho podem ser encontradas no link: <http://www.pentaho.org.br>

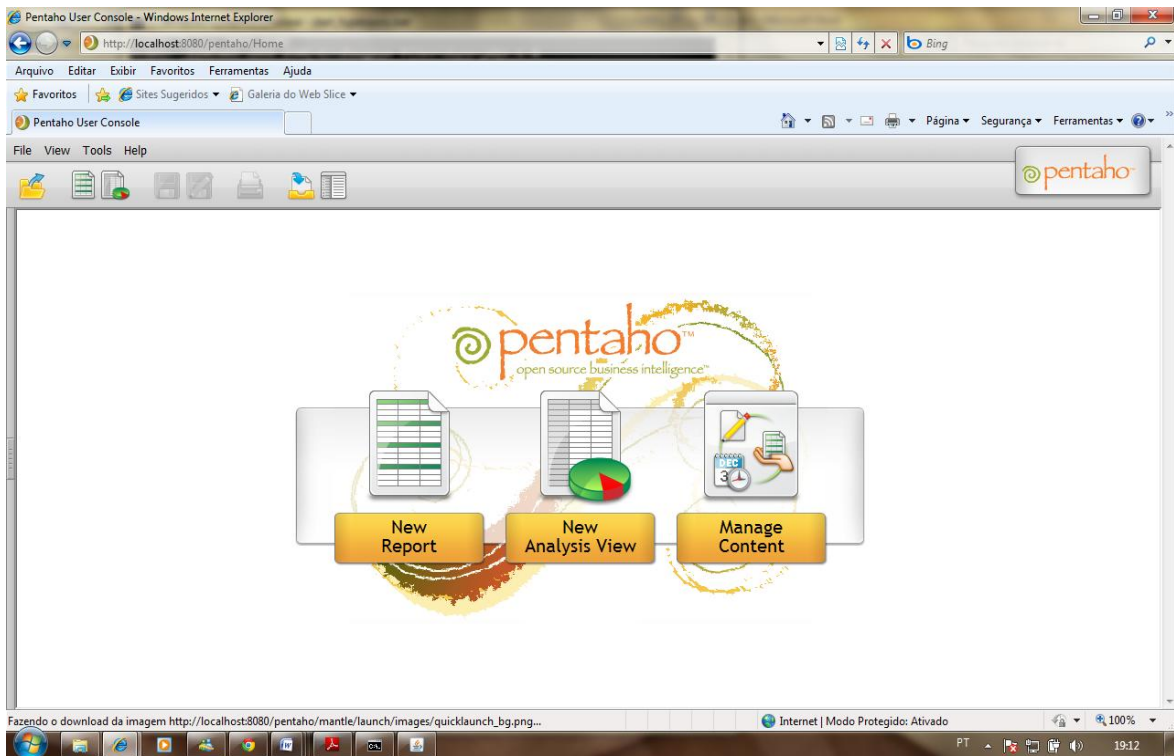
### **3.5 Procedimentos para a execução do trabalho**

Optou-se pela utilização do banco de dados Postgres, por ser um SGBD *Open Source* e possuir extensão para dados geográficos (Postgis). A ferramenta Pentaho não precisa estar conectada a nenhum SGBD, como dito anteriormente, ela é capaz de importar dados em diversos formatos. Porém, visando eliminar a necessidade de utilizar a ferramenta de ETL, foi utilizado o Postgres/Postgis, já que o script de criação das tabelas foi enviado pela GAIS. A figura 17 mostra o Postgres criado e povoado.

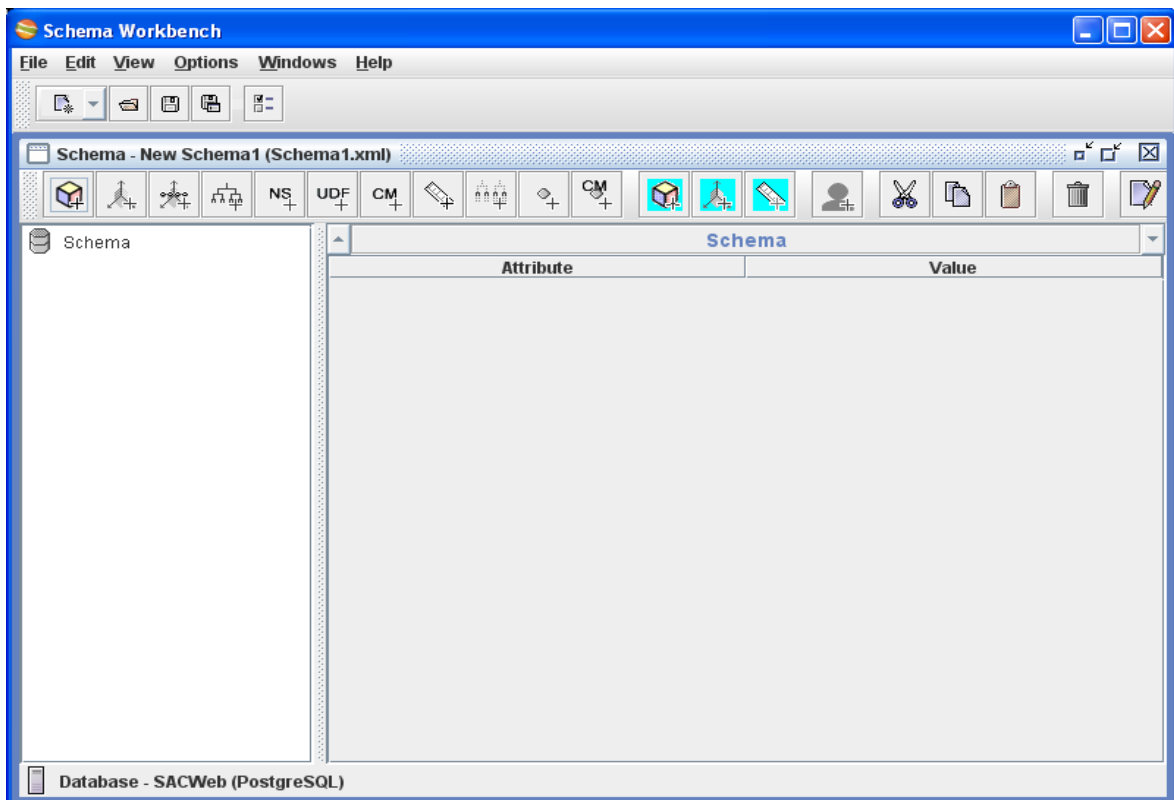


**Figura 18: Tela do pgAdmin - PostGres**  
**Fonte: Própria**

Foi instalada, como dito anteriormente, a versão 3.5 do Pentaho, o pacote bi-server (*New Report, New Analysis View e Manage Content*) bem como o componente *mondrian/workbench*. Como pode ser visto nas figuras 18 e 19.



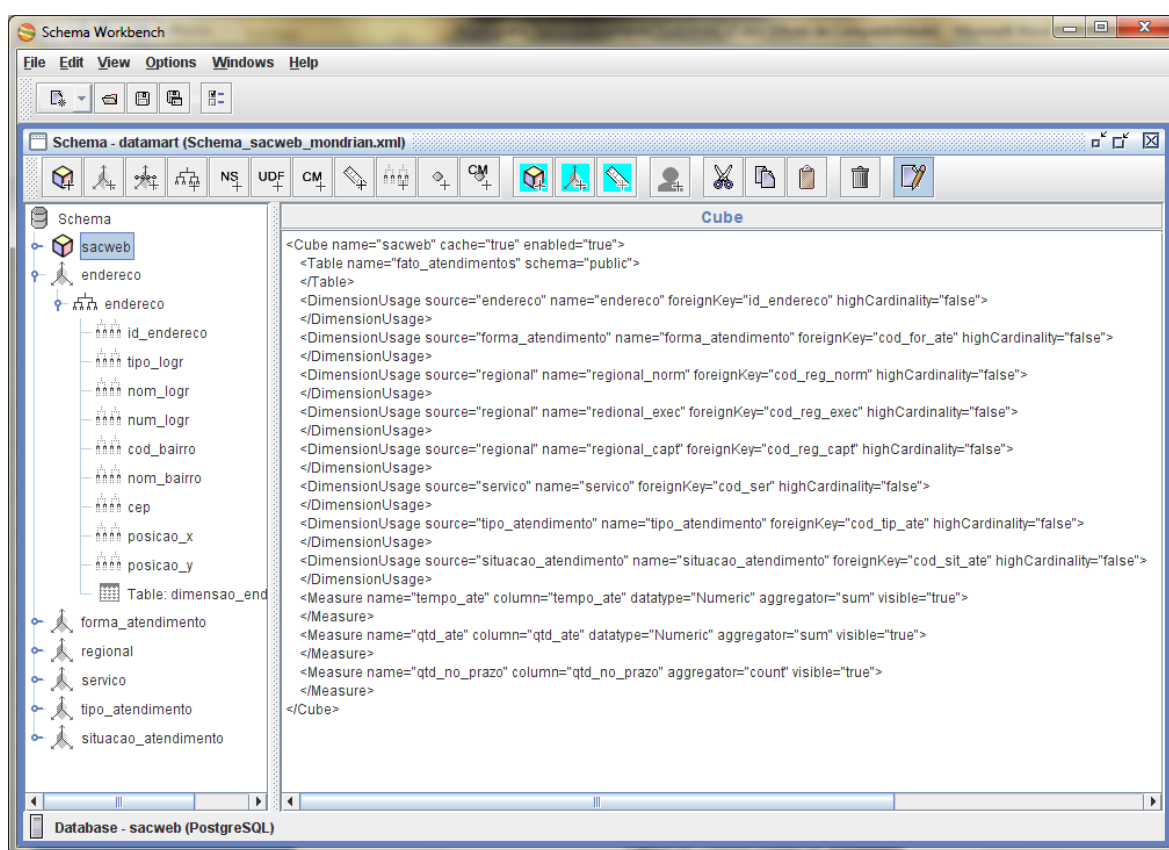
**Figura 19: Tela do Pentaho**  
Fonte: Própria



**Figura 20: Tela do Workbench**  
Fonte: Própria

Depois da criação das tabelas do SACWeb no Postgres/Postgis, foi feita a conexão entre o Pentaho e o Postgres.

Após tudo instalado e configurado, foi necessário a utilização de uma ferramenta para modelar o *Data Mart* geográfico do SACWeb. Pois, o modelo do banco dimensional (Postgres) não é suficiente para o Pentaho identificar os cubos e dimensões. É necessário criar um arquivo XML descrevendo quais são os cubos e dimensões e suas respectivas tabelas associadas. Foi utilizado o *workbench*, já que ele possui integração com o Pentaho e faz a publicação do XML automaticamente, conforme figura 20.

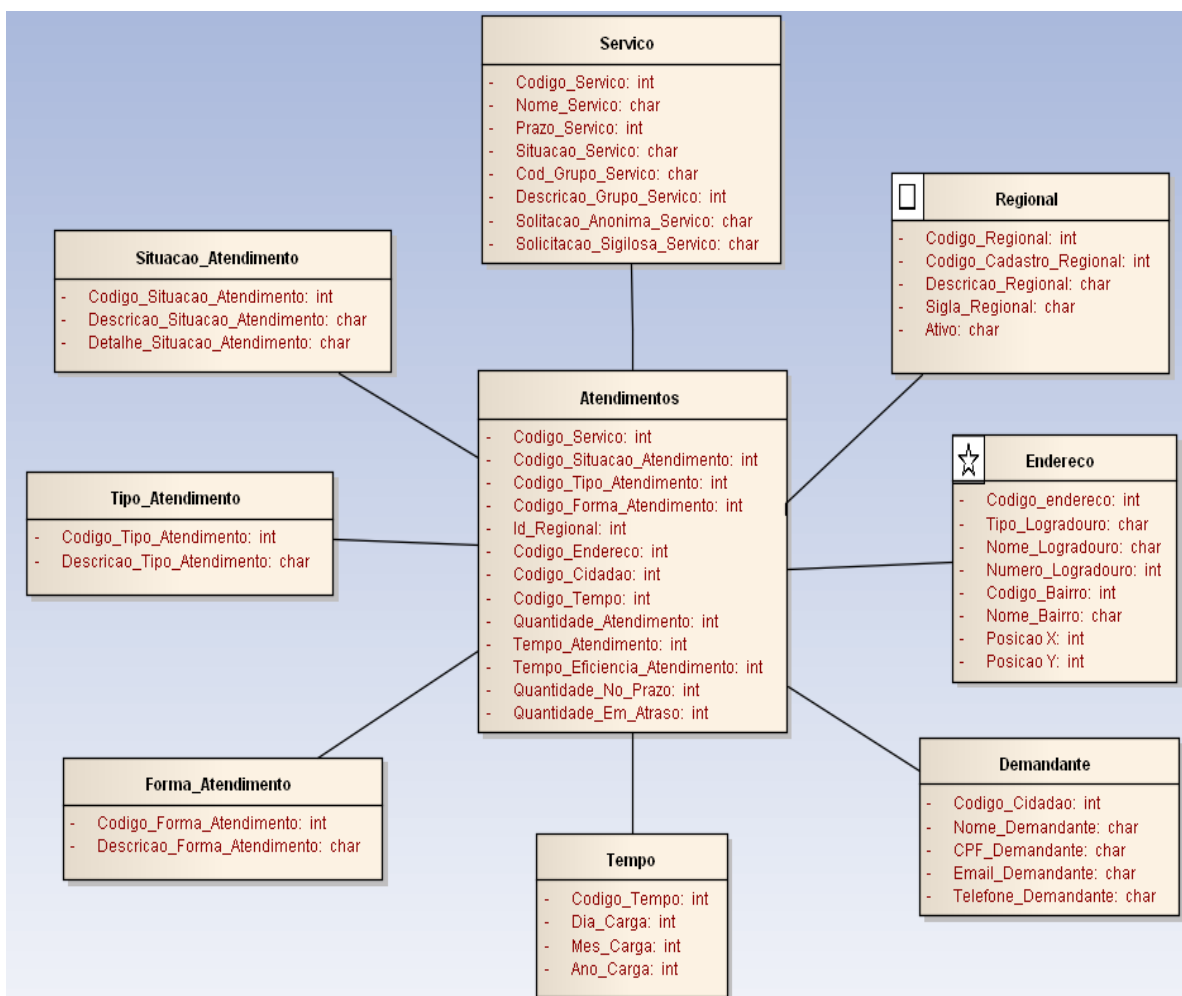


**Figura 21: Tabelas criadas no Workbench**  
Fonte: Própria

Depois do esquema criado, deve-se publicá-lo para então visualizar o modelo no pentaho – *New Analysis View*. Porém, esta etapa não pode ser concluída. Ocorreram diversos erros na importação do modelo pelo *New Analysis View* e apesar dos esforços empreendidos neste trabalho não foi possível solucioná-los devido ao curto tempo para a execução do mesmo e

a pouca documentação e informação sobre as ferramentas escolhidas. Assim, não foi possível atingir todos os objetivos específicos propostos.

Diante deste problema, o modelo de dados do *Data Mart* Geográfico foi criado utilizando a ferramenta *Enterprise Architect* (EA), que foi customizada com o perfil OMT-G (NALON, F.R et AL, 2011). O modelo multidimensional Estrela (Figura 21) foi o escolhido para modelar o *Data Mart* Geográfico do SACWeb.



**Figura 22: Modelo do Data Mart Geográfico**  
Fonte: Própria

O modelo criado é composto pela tabela fato ATENDIMENTOS e pelas tabelas dimensões SITUAÇÃO ATENDIMENTO, TIPO ATENDIMENTO, FORMA ATENDIMENTO, SERVIÇO, REGIONAL, ENDEREÇO e TEMPO.



Segue a descrição de cada tabela do modelo:

Atendimentos: é a tabela Fato do modelo, nela estão armazenadas todas as chaves compostas das tabelas dimensões a qual a tabela Fato esta associada e é nesta tabela que ficam armazenados dados que irão possibilitar análises de informações;

Situação atendimento: pode conter valores como, captado, em andamento, concluído-indeferido ou concluído-executado;

Tipo atendimento: pode ser sugestão, solicitação ou ainda uma reclamação;

Forma de atendimento: pode ser presencial, por telefone, ofício, internet ou outros (forma não listada no modelo);

Serviço: são vários serviços disponíveis, como por exemplo, poda de árvore, capina de lote vago, etc.;

Regional: esta tabela contém o nome e o polígono correspondente a cada regional do município de Belo Horizonte;

Endereço: contém as coordenadas geográficas de onde deverá ser realizado o atendimento de uma determinada solicitação, além de conter o endereço de forma descritiva.

Tempo: através dessa tabela é possível obter informações em diferentes momentos do tempo;

Demandante: contém informações sobre o cidadão que está fazendo alguma solicitação.

O que caracteriza esse DM como geográfico é a dimensão ENDEREÇO e REGIONAL. Dessa forma é possível realizar consultas espaciais, cruzando todo tipo de informação constantes nas 7 (sete) dimensões.

Pelo fato de não ter mais tempo para continuar a execução desse projeto, ele é finalizado com a criação do modelo acima descrito, o qual está pronto para ser implementado em qualquer ferramenta de *Data Warehouse* Geográfico.

## CAPÍTULO 4

### RESULTADOS E DISCUSSÕES

Devido ao tempo gasto na tentativa de eliminar os erros encontrados na hora de publicar o modelo no Pentaho, a pouca documentação e informação sobre as ferramentas escolhidas e ao curto tempo para a execução desse trabalho, não foi possível atingir todos os objetivos específicos propostos. A instalação e configuração das ferramentas foram tranquilas, porém a modelagem do *Data Mart* utilizando o *Workbench*, não é muito intuitiva, necessitando de tutoriais que auxiliem nessa criação. Contudo, esses materiais ainda são escassos, existem poucas informações sobre essas ferramentas, principalmente no que se refere aos componentes com extensão espacial. Também não foram encontradas respostas para os problemas, na comunidade do Pentaho no Brasil.

Um ponto que deve ser analisado é com relação à suíte paga do Pentaho e a suíte/componentes *Open Source*, pois a suíte paga dispõe de suporte técnico e a documentação é completa e robusta, o que não é encontrado na suíte/componentes *Open Source*. Isto pode inviabilizar o uso da ferramenta open source, dependendo da complexidade do projeto a ser desenvolvido.

Foi possível verificar que o trabalho, hoje executado, para extrair informações estratégicas, ainda é muito manual, pois a ferramenta de BI utilizada na PBH não possui extensão espacial, não permitindo as análises de informações georreferenciadas, que é primordial para as tomadas de decisão do prefeito. É importante salientar, que o resultado final é obtido, porém de forma muito mais trabalhosa, demorada e passível de erros.

É fato que os objetivos específicos não puderam ser alcançados, não sendo possível afirmar que com a ferramenta de BI geoespacial esse problema seria eliminado. Mas é sabido que, de acordo com o propósito de uma ferramenta de BI geoespacial, seria eliminado todo o processo de geração de informação estratégica, anteriormente descrito, possibilitando o manuseio da ferramenta pelo próprio prefeito ou mesmo seus assessores. Estes poderiam executar suas consultas de forma dinâmica, obtendo informações estratégicas a tempo e à hora, facilitando a gestão do município.

É importante ressaltar que outros trabalhos, disponíveis na internet, obtiveram êxito na criação de modelos utilizando o *Workbench* e na publicação do mesmo no *New Analysis View*. Fica claro que não se pode dizer que a ferramenta não funciona, apenas precisa-se de um conhecimento maior acerca da ferramenta o que não é conseguido apenas com os tutoriais disponíveis na internet.

Diante da dificuldade de trabalhar com o componente *Workbench*, o modelo de dados do *Data Mart* Geográfico do SACWeb, que era um objetivo específico desse trabalho, foi elaborado utilizando a ferramenta EA. Um trabalho futuro poderá executar os objetivos específicos que não foram alcançados, utilizando o modelo criado no EA para implementação de um *Data Warehouse/Data Mart* Geográfico, utilizando a mesma ferramenta aqui proposta.

## CAPÍTULO 5

### CONCLUSÕES

Visto como é o processo de geração de informações estratégicas atualmente, conclui-se que a necessidade de uma ferramenta de BI com extensão espacial é enorme. A PBH precisa dessa ferramenta, para facilitar a manipulação dos dados referentes ao Sistema de Atendimento ao Cidadão, tornando a gestão das informações e as tomadas de decisões ainda mais precisas.

Mesmo que este trabalho não tenha apresentado na ferramenta Pentaho, o modelo do *Data Mart* Geográfico criado, o modelo conceitual foi elaborado podendo ser implementado utilizando qualquer outra ferramenta de BI que tenha suporte à dados espaciais.

É sabido que, apesar de não atingir todos os objetivos específicos aqui propostos, ter uma ferramenta que integre em um só lugar, todas as características necessárias para implementar um *Data Warehouse* Geográfico, bem como funcionalidades de geração de relatórios, gráficos e mapas é fundamental para tornar mais eficaz e eficiente o processo de geração de informações estratégicas no âmbito da administração municipal de Belo Horizonte.

## REFERÊNCIAS BIBLIOGRÁFICAS

BADARD, Thierry. **Open source geoespacial business intelligence (BI) in action**. 2009, Laval University – Canada.

BARROS, Diego Martins Vieira. **Especificação de testes funcionais para o processador de consultas da arquitetura Golapa**. 2007. Graduação em Engenharia da Computação – Universidade Federal de Pernambuco.

BORBA, Sueli de Fátima Poppi. **Metodologia para Implementação de Modelos Multidimensionais em Banco de Dados Orientado a Objetos**. 2006. Tese de Doutorado - Universidade Federal de Santa Catarina.

CAMARA, Gilberto. **Representação Computacional de dados geográficos**. 2005. Instituto Nacional de Pesquisas Espaciais.

DEGGAU, Renato. **Enriquecendo Data Warehouse Espaciais com Descrições Semânticas**. 2010. Programa de Pós Graduação em Ciência da Computação – Universidade Federal de Santa Catarina.

FELÍCIO, Fernando Luís. **Desenvolvimento de um software Gerenciador de Sistemas SAD**. 2005. 139p.: il. Monografia (Graduação) – Universidade do Oeste Paulista – UNOESTE: Presidente Prudente.

FERREIRA Manuele; RAMOS Gustavo; BERNARDO Luis; SILVA Robson; MAIOR Bruno. **Plataforma Pentaho de Business Intelligence**. Manual de Utilização – Universidade Federal da Bahia – Departamento de Ciência da Computação.

GONÇALVES, Márcio. **Extração de Dados para Data Warehouse**. Editora Axcel Books do Brasil. 2003.

HARRISON, Thomas H. **Intranet Data Warehouse**. São Paulo, Berkeley Brasil, 1998.

INMON, William H. **Como construir o Data Warehouse**. Rio de Janeiro: Campos, 1997.

INMON, William H; HACKATHORN, Richard D. **Como usar o Data Warehouse**. Rio de Janeiro, Infobook, 1997.

KIMBALL, Ralph. **Data Warehouse Toolkit**. São Paulo: Makron Books, 1998.

NALON, Felipe Ribeiro; FILHO, Jugurta Lisboa; BRAGA, Luiz José; BORGES, Karla Albuquerque de Vasconcelos; Andrade, Marcus Vinicius Alvim. **Applying the Model Driven Architecture Approach for Geographic Database Design using a UML Profile and ISO Standards**. Journal of Information and Data Management. Vol 2, No 2, 2011.

OLIVEIRA, Adelize Generini. **Data Warehouse: Conceitos e Soluções**. SFO Gráfica e editora. 1998.

PRIMAK, Fábio Vinícius. **Decisões com B.I.** Editora Ciência Moderna, 2008.

QUEIROZ, Luis Gustavo Carneiro. **Geoprocessamento e Data Warehouse: Uma aplicação para apoio à tomada de decisão em Segurança Pública**. 2005. Monografia (Graduação) – Universidade Federal da Bahia.

ROSA, Kátia Liane. **Análise de Ferramentas OLAP Open Source e Freeware com o acesso à Banco de Dados Multidimensionais e Relacionais**. 2004. Monografia (Graduação) – Centro Universitário Feevale.

SELL, Denilson. **Uma Arquitetura para Business Intelligence baseada em tecnologias semânticas para suporte a aplicações analíticas**. 2006. 210f.: il. Tese (doutorado) - Universidade Federal de Santa Catarina. Programa de Pós Graduação em Engenharia de Produção.

SIQUEIRA, Thiago Luíz Lopes. **DWE: Uma estrutura de indexação para Data Warehouse Espacial**. 2008. Programa de Pós Graduação em Ciência da Computação – Universidade Federal de São Carlos.

SOBRAL, Fábio Monteiro; SOUZA, Marco Antonio Motta. **Proposta de Data Mart para o Plano de Saúde do Senado Federal**. 2006. Monografia (Graduação). Brasília.

SOUZA, Tiago Marafane Lins. **Análise, Projeto e implementação de um Data Mart para Ambientes de Business Intelligence**. 2005. Monografia (Graduação). Universidade Federal de Pernambuco.