

UNIVERSIDADE CÂNDIDO MENDES
CAMPOS DOS GOYTACAZES
MESTRADO EM INFORMÁTICA APLICADA

ALAN CARVALHO GALANTE

SISTEMA DE SUPORTE À DECISÃO NO PLANEJAMENTO URBANO
MUNICIPAL: UM ESTUDO DE CASO DO MUNICÍPIO DE MACAÉ

CAMPOS DOS GOYTACAZES
2004

ALAN CARVALHO GALANTE

SISTEMA DE SUPORTE À DECISÃO NO PLANEJAMENTO URBANO MUNICIPAL: UM
ESTUDO DE CASO DO MUNICÍPIO DE MACAÉ

Dissertação apresentada ao Curso de Mestrado em
Informática Aplicada da Universidade Cândido Mendes,
como requisito parcial para a obtenção do grau de Mestre.
Área de Concentração: Sistemas de Informação e Apoio à
Decisão.

Orientador: Prof. Jorge Luís Nunes e Silva Brito, PhD

Campos dos Goytacazes
2004

ALAN CARVALHO GALANTE

SISTEMA DE SUPORTE À DECISÃO NO PLANEJAMENTO URBANO MUNICIPAL: UM
ESTUDO DE CASO DO MUNICÍPIO DE MACAÉ

Dissertação apresentada ao Curso de Mestrado em
Informática Aplicada da Universidade Cândido Mendes,
como requisito parcial para a obtenção do grau de Mestre.
Área de Concentração: Sistemas de Informação e Apoio à
Decisão.

Aprovada em _____

BANCA EXAMINADORA

Professor Jorge Luis Nunes e Silva Brito, PhD – Orientador
Universidade Cândido Mendes

Professora Sahudy Montenegro Gonzalez, DSc
Universidade Cândido Mendes

Professor Jugurta Lisboa Filho, DSc
Universidade Federal de Viçosa (UFV)

Campos dos Goytacazes
2004

AGRADECIMENTOS

A Deus, que sempre esteve ao meu lado possibilitando o término deste trabalho e colocando pessoas importantes para me acompanhar por toda esta jornada.

Aos meus pais, Rute e Joel, que me apoiaram desde o início dando-me todo o incentivo não me deixando desanimar e desistir. Aos meus irmãos, Aline e Jocimar, que sempre estiveram juntos me ajudando no possível.

A minha namorada, que compreendeu minhas limitações de horários de estudo sempre com carinho e companheirismo.

Ao meu orientador, Prof. Nunes pelos ensinamentos que possibilitaram o desenvolvimento deste trabalho, a paciência com o qual fui conduzido, a amizade e confiança que aumentou nossa cumplicidade e o estímulo que foi o essencial para me manter neste caminho.

Aos meus companheiros do mestrado, principalmente a Giovanni e Sérgio, pelo fato de nos tornamos mais do que companheiros de turma e sim amigos inseparáveis. Devo um destaque especial ao Giovanni que muito contribuiu para o desenvolvimento deste trabalho, auxiliando-me independentemente de hora ou local em que se encontrava.

A Prefeitura Municipal de Macaé, pela atenção em todos os momentos em que foi solicitada por minha pessoa. Em particular a equipe do projeto Geo-Macaé, que disponibilizou todo o material necessário para a realização deste trabalho. Um agradecimento especial ao Sr. José Manuel que sempre me ajudou, apoiou e incentivou na realização deste projeto.

Aos professores Tanaka e Thomé que também acrescentaram muito a este projeto, através dos ensinamentos, amizade e incentivo.

SUMÁRIO

1 INTRODUÇÃO	13
1.1 MOTIVAÇÃO.....	15
1.2 OBJETIVO DA DISSERTAÇÃO	16
1.3 ESTRUTURA DA DISSERTAÇÃO	17
2 FUNDAMENTAÇÃO TEÓRICA.....	18
2.1 SISTEMAS DE SUPORTE À DECISÃO	18
2.2 DATAWAREHOUSE	19
2.3 “ON-LINE ANALYTICAL PROCESSING” (OLAP).....	23
2.4 MINERAÇÃO DE DADOS (“DATAMINING”).....	26
2.5 SISTEMA DE INFORMAÇÃO GEOGRÁFICA (SIG).....	26
2.5.1 <i>A análise sobre dados geográficos</i>	29
2.6 O CONSÓRCIO “OPEN GIS”	30
2.7 MODELAGEM.....	33
2.7.1 “ <i>Unified Modeling Language</i> ” (UML)	34
2.7.2 <i>A utilização de estereótipos na modelagem</i>	35
2.7.3 <i>Modelando um Datawarehouse com características espaciais</i>	38
2.7.3.1 Definição da tabela fato	38
2.7.3.2 Definição das dimensões	39
2.7.3.3 Definição de dimensões geográficas no Datawarehouse.....	40
2.7.3.3.1 Dimensão Geográfica Agregável	40
2.7.3.3.2 Dimensão geográfica não agregável	41
2.8 FORMATO “ARCVIEW SHAPE FILE”	42
2.9 POSTGRESQL	43
2.9.1 <i>PostGIS</i>	44
2.9.1.1 Tipos geográficos do PostGIS	45
2.9.1.2 Funções geográficas do PostGIS	45
2.10 POSTGEOOLAP.....	47
2.10.1 <i>Ferramentas OLAP Espacial</i>	47
2.10.2 <i>Características do PostGeoOLAP</i>	48

2.11 O SOFTWARE PLANET GIS	50
3 O MUNICÍPIO DE MACAÉ	51
3.1 A PREFEITURA MUNICIPAL DE MACAÉ	52
3.2 O PROGRAMA MACAÉ CIDADÃO	52
3.3 O PROJETO GEO-MACAÉ	53
4 PROJETO DO SISTEMA DE SUPORTE A DECISÃO DE MACAÉ.....	54
4.1 A MODELAGEM DO SSD DO MUNICÍPIO	56
4.1.1 A Modelagem do Espaço Geográfico.....	56
4.1.2 As classes geográficas não agregáveis	62
4.1.3 As dimensões do Datawarehouse	63
4.1.4 O modelo do Datawarehouse do município de Macaé.....	68
5 A IMPLEMENTAÇÃO DO MODELO	70
5.1 O AMBIENTE DE DESENVOLVIMENTO DO SSD	70
5.2 A GERAÇÃO DO BANCO DE DADOS MULTIDIMENSIONAL GEOGRÁFICO	71
5.3 A MIGRAÇÃO DOS DADOS	71
5.3.1 Migração das dimensões não-agregáveis	72
5.3.1.1 Tratamento dos dados convencionais de Escolas	72
5.3.1.2 Tratamento dos dados convencionais de Unidades de Saúde.....	73
5.3.1.3 Georreferenciamento das dimensões não-agregáveis	74
5.3.2 Migração da dimensão setorcensitario	76
5.3.2.1 Migração dos dados convencionais	76
5.3.2.2 Migração dos dados geográficos	77
5.3.2.3 Atualização dos totalizadores das dimensões geográficas não-agregáveis.....	79
5.3.3 Migração das dimensões agregáveis.....	80
5.3.3.1 Migração da dimensão característica	82
5.3.3.2 Migração da dimensão condicaomoradia	83
5.3.3.3 Migração da dimensão educacao	85
5.3.3.4 Migração da dimensão trabalho	86
5.3.3.5 Migração da dimensão deficienciafisica	88
5.3.3.6 Migração da dimensão problemasaude	89
5.3.3.7 Migração da dimensão queixasistemaude	90

5.4 O TRATAMENTO DOS MAPAS	91
5.5 A PREPARAÇÃO DO CUBO OLAP	92
6 TESTES E ANÁLISE DOS RESULTADOS	98
6.1 FUNCIONAMENTO DA FERRAMENTA OLAP	98
6.2 REALIZAÇÃO DE CONSULTAS	103
6.2.1 <i>Quantidade de pessoas por Setor Administrativo que possuem mais de 45 anos e estão desempregadas.....</i>	<i>103</i>
6.2.2 <i>Quantidade de pessoas sem estudo que estão com problema crônico de saúde, tem mais de 50 anos e estão desempregadas, exibindo o resultado por sexo.....</i>	<i>104</i>
6.2.3 <i>Quantidade de pessoas que recebem educação especial e moram a mais de 1000 metros de alguma Unidade de Ensino Público</i>	<i>105</i>
6.2.4 <i>Quantidade de brancos e outras raças, que estejam cursando o 3º grau, estejam ou não trabalhando</i>	<i>106</i>
6.2.5 <i>Quantidade de mulheres, por bairro, menores de 18 anos que não trabalham, não estudam e têm algum problema de saúde.....</i>	<i>107</i>
6.2.6 <i>Número de pessoas que possuem deficiência física que estão fora do mercado de trabalho, por setor administrativo.....</i>	<i>108</i>
6.2.7 <i>Quantidade de pessoas que não têm saneamento na residência (fossa, esgoto) e apresentam problemas de saúde, por bairro.....</i>	<i>109</i>
6.2.8 <i>Quantidade de pessoas que não sabem ler e escrever que estão trabalhando, por bairro.</i>	<i>110</i>
6.2.9 <i>Quantidade de pessoas que não trabalham e declaram renda acima de 5 salários mínimos, por bairro.....</i>	<i>111</i>
6.2.10 <i>Quantidade de pessoas que moram no Setor Vermelho e utilizam os serviços públicos de saúde.</i>	<i>112</i>
6.3 CONSIDERAÇÕES SOBRE OS TESTE REALIZADOS	113
7 CONCLUSÃO.....	114
7.1 CONTRIBUIÇÕES DO TRABALHO	116
7.2 SUGESTÕES PARA TRABALHOS FUTUROS.....	116
8 REFERÊNCIAS	118

FIGURAS

Fig.1	O Processo Completo do Datawarehouse.....	20
Fig.2	Processo de extração, transformação e carga de dados no Datawarehouse	21
Fig.3	Exemplo de Modelo Lógico Multidimensional	22
Fig.4	Exemplo de Modelo Lógico “Snowflake”.....	22
Fig.5	Exemplo de hierarquia da dimensão tempo	25
Fig.6	Hierarquia de Classes Geométricas – OpenGIS	30
Fig.7	Exemplos de Polígonos.....	31
Fig.8	Exemplos de Não-Polígonos.....	31
Fig.9	Exemplos de LineString.....	32
Fig.10	Exemplos de MultiLineString.....	32
Fig.11	Exemplo de representação com estereótipos - comparação da representação da hierarquia com a representação com estereótipos.....	38
Fig.12	Exemplo de modelo multidimensional – fato e dimensão	39
Fig.13	Exemplo de dimensão geográfica agregável.....	41
Fig.14	Exemplo de dimensão geográfica não-agregável	42
Fig.15	Tipos de dados geográficos e suas definições no OpenGIS.....	45
Fig.16	Visualização dos Setores Censitários do Município de Macaé	56
Fig.17	Visualização dos Bairros do Município de Macaé	57
Fig.18	Visualização dos Setores Administrativos do Município de Macaé	58
Fig.19	Representação do espaço geográfico do município de Macaé	59
Fig.20	Representando as classes geográficas com o auxílio dos estereótipos	60
Fig.21	Modelo Físico das Classes Geográficas	61
Fig.22	Classes Geográficas não agregáveis	62
Fig.23	Dimensões condicaomoradia e característica	63
Fig.24	Dimensões tempo e trabalho	65
Fig.25	Dimensões educacao e queixasistema	66
Fig.26	Dimensões problemasaude e deficienciafisica	67
Fig.27	O Modelo Lógico Dimensional.....	68
Fig.28	Modelo Multidimensional Físico	69

Fig.29 Diagrama de Distribuição.....	70
Fig.30 Arquivo disponibilizado pela Prefeitura de Macaé com o cadastro de escolas	72
Fig.31 Arquivo de cadastro de escolas formatado para exportação para o PostgreSQL....	73
Fig.32 Cadastro de unidades de saúde disponibilizado pela Prefeitura de Macaé	74
Fig.33 Mapa com as Escolas, Unidades de Saúde e Setores Censitários.....	75
Fig.34 Tabelas de Setores Censitários disponibilizadas pela Prefeitura	77
Fig.35 Comandos SQL criados pelo shp2pgsql – geração dos Setores Censitários.....	78
Fig.36 Consulta geográfica retornando um totalizador da dimensão setor censitário para escolas privadas.....	79
Fig.37 Ferramenta de Migração dos dados das dimensões agregáveis.....	80
Fig.38 Atributos dos dados dos Municípios.....	81
Fig.39 Origem da dimensão caracteristica	82
Fig.40 Origem da característica Possui Plano de Saúde	83
Fig.41 Origem da dimensão condicaomoradia	84
Fig.42 Origem da dimensão educacao	86
Fig.43 Origem da dimensão trabalho	87
Fig.44 Origem da dimensão deficienciafisica	88
Fig.45 Origem da dimensão problemasaude	89
Fig.46 Origem da dimensão queixasistemasaude	90
Fig.47 Mapa completo com todas as cinco camadas geográficas	92
Fig.48 Esquema de conexão com o banco de dados multidimensional	93
Fig.49 Definição da tabela-fato	94
Fig.50 Definição das hierarquias da dimensão tempo	94
Fig.51 Definição das hierarquias da dimensão deficienciafisica	95
Fig.52 Definição das hierarquias da dimensão caracteristica	95
Fig.53 Comparativo entre a hierarquia do cubo e a hierarquia dimensional conceitual ...	96
Fig.54 Tela principal da ferramenta OLAP.....	99
Fig.55 Operações de “slice” e “dice” sobre o cubo.....	100
Fig.56 Resultado da consulta da quantidade de moradores de um bairro	100
Fig.57 Resultado da consulta da quantidade de moradores de um bairro por sexo.....	101
Fig.58 Exemplo de consulta retornando resultados geográficos.....	101
Fig.59 Exemplo de consulta usando função geográfica	102
Fig.60 Pessoas com mais de 45 anos, desempregadas por Setor Administrativo	104

Fig.61 Pessoas sem estudo, com problema crônico de saúde, mais de 50 anos, desempregadas por sexo.....	105
Fig.62 Setores Censitários que possuem pessoas com educação especial morando a mais de 1Km de alguma Unidade de Ensino Público	106
Fig.63 Pessoas cursando o 3º grau por raça e por estar ou não trabalhando	107
Fig.64 Quantidade de mulheres menores de 18 anos que não trabalham, não estudam e têm algum problema de saúde, por bairro.....	108
Fig.65 Número de pessoas que possuem deficiência física e não trabalham, por setor administrativo	109
Fig.66 Quantidade de pessoas sem saneamento na residência, com problemas de saúde, por bairro	110
Fig.67 Pessoas que não sabem ler e escrever por bairro e trabalham.....	110
Fig.68 Quantidade de pessoas que não trabalham e declaram renda acima de 5 salários mínimos, por bairro	111
Fig.69 Quantidade de pessoas que moram no Setor Vermelho e utilizam os serviços públicos de saúde.	112

TABELAS

Tab. 1 -	Comparação do modelo Open GIS com o modelo UML com estereótipos ...	36
Tab. 2 -	Principais funções do PostGIS provenientes do OpenGIS	46
Tab. 3 -	Descrição dos atributos da dimensão condicaomoradia	64
Tab. 4 -	Descrição dos atributos da dimensão trabalho	65
Tab. 5 -	Descrição dos atributos da dimensão educacao	66
Tab. 6 -	Descrição dos atributos da dimensão problemasaude	67
Tab. 7 -	Descrição dos atributos da dimensão deficienciafisica	67
Tab. 8 -	Migração dos dados da dimensão setorcensitario	76
Tab. 9 -	Migração dos dados de característica	83
Tab. 10 -	Migração dos dados de condicaomoradia	84
Tab. 11 -	Migração dos dados de educacao	86
Tab. 12 -	Migração dos dados de trabalho	87
Tab. 13 -	Migração dos dados de deficienciafisica	89
Tab. 14 -	Migração dos dados de problemasaude	90

RESUMO

O trabalho tem por objetivo desenvolver um Sistema de Suporte à Decisão e de Gestão do Conhecimento para a Prefeitura Municipal de Macaé. Para tal, integram-se tecnologias de DataWarehouse, OLAP e Sistemas de Informação Geográfica. A estratégia básica consiste na união dos dados geográficos municipais com as diversas bases de dados transacionais, integrando-se, assim, as informações que cada Órgão ou Unidade municipal pode oferecer. Isto possibilitará ao gestor municipal o cruzamento e a análise dos dados geográficos e transacionais, com o propósito de extrair informações de cunho gerencial para a sua tomada de decisão.

Como resultados e benefícios práticos do sistema ora desenvolvido, destaca-se a possibilidade do usuário realizar consultas “on-line”. Esta interface proporciona respostas também amigáveis, por intermédio de mapas, gráficos, planilhas, listas e relatórios. Espera-se que, uma vez de posse de uma ferramenta como esta, o município de Macaé seja capaz de desenvolver um trabalho de boa qualidade, atendendo de forma muito mais efetiva, isto é, com uma melhor relação custo-benefício, às necessidades dos seus munícipes, além de ter também um instrumento de gestão de conhecimento de suas Unidades ou Órgãos.

ABSTRACT

This work aims to develop a Decision Support and Knowledge-Management System dedicated to the Urban Planning of the Macaé County. For doing that, the technologies of Datawarehouse, OLAP and GIS were integrated accordingly. The basic strategy is to join both geographic and conventional data, thus integrating the information of many county management departments. Such a strategy has paved the way for implementation of data analysis procedures. Those procedures led to the extraction of information for county management purposes. The traditional techniques and procedures of System Analysis and Design, as well as the Object Oriented approach, were used for system development and implementation.

One of the main results and potential benefits of this research are the on-line queries using a user-friendly interface.

An important contribution of this work is to offer a low-cost Decision Support System to the Macaé county Management Department to better attend to the community.

One expects that, once the Macaé county Management Department uses this system, it could better attend to the necessities of its inhabitants.

1 INTRODUÇÃO

No mundo de hoje não é exagero afirmar que “as empresas morrem de sede por conhecimento em um oceano de dados”. Com a crescente informatização da área operacional das empresas, foram gerados dados para o controle das mesmas; em meio a todos esses dados, muitas vezes, estão ocultas informações de extrema importância para o gestor. Daí decorre o papel dos Sistemas de Suporte a Decisão: extrair informações implícitas de dados explícitos, fornecendo, desta forma, uma valiosa ferramenta para o decisor tornar sua empresa mais competitiva.

Dentro do campo do conhecimento dos Sistemas de Suporte a Decisão existem três grandes sub-sistemas que auxiliam no processo decisório: Sistemas baseados na tecnologia Datawarehouse (DW) com Processamento Analítico On-line (OLAP), Data mining e Sistemas de Informação Geográfica (SIG).

O Datawarehouse é, na verdade, uma grande massa de dados centralizada que tem como característica principal a dependência do tempo. Os dados que a originam podem provir de diversas bases de dados geograficamente distantes. Os bancos de dados relacionais, que são usados na grande maioria das empresas, tendem a seguir a normalização, que é um conjunto de regras que se propõe a minimizar as redundâncias, e a facilitar operações de inserção, exclusão e atualização. Como o Datawarehouse tem que ser preparado para consultas e não para operações de manipulação de dados, então se tende evitar as regras de normalização de tabelas. O Datawarehouse é não volátil, ou seja, suas informações sofrem poucas mudanças ao longo do tempo; o que existe é uma atualização periódica dos dados.

A obtenção dos dados do Datawarehouse é feita através de um processo de pré-processamento “inteligente” de forma que os dados operacionais sejam particionados visando à facilitação das consultas sobre essa base de dados.

Uma vez de posse dos dados centralizados e organizados de forma agregada, há que se conseguir extrair as informações dessa base de dados. Os bancos de dados em geral oferecem uma tecnologia baseada em consultas, tal como a SQL. Mas esse tipo de consulta restringe-se às operações da álgebra relacional, que se apóiam na seleção de colunas de tabelas ou na junção de tabelas relacionadas, baseadas em campos comuns. Como a tecnologia do Datawarehouse permite ver a mesma informação sobre múltiplas dimensões, o ideal é o uso de uma tecnologia capaz de minerar ou descobrir novas informações a partir de regras ou padrões aplicados a uma base de dados.

O conjunto de ferramentas que possibilita efetuar a exploração dos dados contidos num Datawarehouse é conhecido como Processamento Analítico On-Line (“On-Line Analytical Processing” - OLAP), que trabalha com análise multi-dimensional, ou seja, representação dos dados como dimensões, ao invés de tabelas, sendo possível tirar informações do tipo: comparações entre períodos, percentual, médias, somas, contagens, funções estatísticas e financeiras, etc.

Sobre o Datawarehouse é possível ainda aplicar um processo de mineração de dados (Datamining), que é o processo de extrair informação válida, previamente desconhecida e de máxima abrangência, a partir de extensas bases de dados. O Datamining vai muito além da simples consulta a um banco de dados, no sentido de que permite a extração de informações não tão óbvias como predição, associação, classificação e clusterização, com o uso de diversos algoritmos ou estratégias, tais como: redes neurais, algoritmos genéticos, lógica nebulosa, árvores de decisão, regressão linear e não linear, etc.

Dentre os Sistemas de Suporte a Decisão, destacam-se os Sistemas de Informação Geográfica (SIG), os quais englobam diversas disciplinas e tecnologias como geografia, cartografia, sensoriamento remoto, fotogrametria, posicionamento espacial e levantamentos topográficas.

Esses tipos de sistemas têm como principal característica o armazenamento, manipulação e análise de objetos e fenômenos cuja principal característica é a localização geográfica.

Observa-se, portanto, que o Datawarehouse, em conjunto com o Datamining, trata os dados convencionais com uma forte ligação com a variável tempo, enquanto que o SIG trata

dados onde a variável preponderante é a geográfica. Segundo (FERREIRA, 2002), essa diferenciação entre as ferramentas SIG e DW se dá pelo fato de que “durante vários anos estas foram usadas em domínios específicos e por usuários com diferentes perfis”.

1.1 MOTIVAÇÃO

A região Norte Fluminense tem como uma das principais atividades econômicas a exploração do petróleo na plataforma continental da bacia de Campos. Essa atividade econômica tem trazido inúmeros benefícios aos municípios da região, motivados principalmente pela concessão dos “royalties” do petróleo.

Dentre os municípios de maior importância na região ora em tela, cita-se o de Macaé, cuja Prefeitura Municipal experimenta uma série de reformulações e atualizações tecnológicas, particularmente com o emprego dos recursos de Informática. Nesse contexto, destacam-se os projetos Geo-Macaé e Macaé-Cidadão.

Quanto ao primeiro projeto, trata-se de um Sistema de Informações Cartográficas (SIC), que tem como proposta georreferenciar as informações do município, de modo que se obtenha uma base cartográfica digital com a estrutura territorial da região a partir da base de setores censitários do IBGE. Este projeto integra-se ao projeto Macaé-cidadão, que consistiu de uma coleta de campo de dados sócio-econômicos da população daquele município.

A Prefeitura de Macaé possui diversas Secretarias Municipais sob seu comando. Essas Secretarias possuem diversos programas aplicativos, por intermédio de bancos de dados denominados transacionais ou convencionais, que controlam seus diversos Órgãos e Unidades administrativas. Entretanto, tais aplicativos não são integrados uns aos outros, de forma que se observa a geração de verdadeiras “ilhas” entre os diversos órgãos municipais. Tal fato gera a ocorrência de redundâncias e inconsistências nas informações fornecidas ao planejador municipal, o que resulta em prejuízos para a comunidade Macaense.

1.2 OBJETIVO DA DISSERTAÇÃO

Este trabalho tem por objetivo desenvolver um sistema de suporte a decisão municipal para a cidade de Macaé. Para tal, pretende-se unir os dados geográficos com as diversas bases de dados convencionais aplicando-se em conjunto, as técnicas de Datawarehouse, OLAP e SIG. Com isso, o planejador municipal terá em mãos informações de cunho gerencial para a tomada de decisões, de forma a melhor gerir os recursos públicos.

A partir da metodologia de Datawarehouse é possível centralizar os dados convencionais e tê-los dispostos de forma a facilitar a busca de informações. Na geração desse Datawarehouse serão realizadas operações de transformação, extração e carga dos dados dos diversos Órgãos ou Unidades Municipais.

Aplicando técnicas de Datamining, OLAP e SIG sobre o Datawarehouse ter-se-á em mãos uma poderosa ferramenta de suporte à decisão. A título de exemplificação com uma ferramenta como esta pode-se obter os seguintes resultados:

- Com o OLAP e SIG em conjunto, pode-se extrair do Datawarehouse quais os locais de ocorrência de casos de dengue ou outras enfermidades no município. De posse dessa informação, o gestor poderá verificar as regiões que devem receber algum tratamento especial para minimizar o problema em questão.
- Com o OLAP e SIG em conjunto, pode-se extrair do Datawarehouse o número de crianças que não estejam estudando, numa determinada faixa de idade, por local. De posse dessa informação, o gestor poderá, por exemplo, detectar a necessidade de construção de uma escola nestes locais, visando ao atendimento da demanda.
- Com o OLAP, o planejador será capaz de verificar o número de contribuintes de IPTU com idade acima de determinado valor, com algum tipo de cardiopatia. Por exemplo, é possível identificar nesta consulta a presença de bases de dados diferentes (Secretaria de Saúde e Secretaria de Fazenda), que é possibilitado pelo Datawarehouse. De posse dessa informação, o gestor poderá, por exemplo, conceder isenção de IPTU para essas pessoas.

Com esse Sistema de Suporte à Decisão, o gestor municipal terá em mãos uma ferramenta para desenvolver um trabalho racional em prol do município.

1.3 ESTRUTURA DA DISSERTAÇÃO

Esta dissertação está organizada em 7 capítulos, incluindo-se o presente.

No capítulo 2, são detalhados os fundamentos teóricos julgados necessários para a pesquisa, procurando-se apresentar as referências e trabalhos utilizados para a realização do presente projeto.

O capítulo 3 versa sobre o município de Macaé, objeto do estudo de caso desta dissertação, os trabalhos que foram ou estão sendo desenvolvidos pela prefeitura e que serão de suma importância para o desenvolvimento da presente pesquisa.

No capítulo 4 está descrito o processo utilizado no desenvolvimento do modelo conceitual e físico do Datawarehouse. O processo baseia-se nas informações disponibilizadas pela Prefeitura Municipal de Macaé, oriundas dos projetos GeoMacaé, Macaé Cidadão e alguns dados transacionais disponibilizados pela Prefeitura Municipal de Macaé.

O capítulo 5 trata do processo de implementação do modelo. Pode-se destacar neste capítulo a geração do banco de dados multidimensional, o processo de migração dos dados das diversas bases de dados da prefeitura de Macaé, o tratamento dos mapas e a criação do cubo OLAP.

O capítulo 6 aborda as consultas realizadas sobre o projeto para a validação do modelo ora proposto. Neste capítulo foram realizadas consultas em parceria com a prefeitura municipal de Macaé, de forma que fossem apresentados resultados correspondentes às reais necessidades municipais.

O capítulo 7 traz as conclusões acerca do trabalho e inclui sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 SISTEMAS DE SUPORTE À DECISÃO

No dia a dia das empresas, sejam estas públicas ou privadas, são “produzidas” grandes quantidades de dados fornecidos pelos softwares de controle e acompanhamento da área operacional, além de dados gerados por comunicação entre pessoas internas ou externas à empresa, dados provenientes da internet, etc.

O quê, na verdade, diferencia as empresas, é o processo de seleção e organização desses dados, de forma a fornecer informações mais apuradas ao decisor da organização. Daí vem o objetivo dos Sistemas de Suporte à Decisão: retirar da grande quantidade de dados de uma organização informação útil para que os profissionais das áreas estratégica, tática ou operacional possam fazer seus trabalhos de forma racional e otimizada.

Elmasri e Navathe (2000) definiram Sistemas de Suporte à Decisão como um sistema que leva dados aos decisores de alto nível de uma organização para apoiar decisões complexas e importantes. Come (2001) definiu apoio a decisão como o “processo de agrupar, estruturar, manipular, armazenar, acessar, apresentar e distribuir informações de negócios de maneira oportuna, ou seja, a informação certa no momento certo e na quantidade certa”. Esses sistemas possuem conhecimentos específicos sobre o assunto e, mediante as normas de avaliação introduzidas pelo usuário, apresentam solução otimizada para o problema em questão, auxiliando o processo decisório.

Os sistemas de suporte à decisão podem ser divididos em dois tipos:

a) Sistemas Orientados a Modelos, os quais se baseiam em modelos teóricos e na capacidade de análise e agregação de dados estatísticos e históricos. Por exemplo, os sistemas usados nas bolsas de valores pelos investidores que, a partir de estatísticas históricas sobre preços, tomam suas decisões de compra ou venda de ações.

b) Sistemas Orientados aos Dados: são aqueles que, a partir de totalizações, médias e distribuições, em diferentes perspectivas, são capazes de fornecer variados elementos para a interpretação do decisor ou gerente.

Pode-se visualizar duas grandes classes de Sistemas de Suporte à Decisão Orientados a Dados: os sistemas baseados em Datawarehouse e On-Line Analytical Processing, e os Sistemas de Informação Geográfica.

2.2 DATAWAREHOUSE

Segundo Kimball (1998), Datawarehouse “é o lugar onde as pessoas podem acessar seus dados”. Inmon (1992) caracterizou o termo como “uma coleção de dados orientada por assuntos, integrada, não volátil, variante no tempo, e que tem por objetivo dar suporte aos processos de tomada de decisão”. O Datawarehouse é orientado por assunto porque contém informações sobre temas específicos importantes para a empresa, tais como: produtos, vendas e clientes.

A integração é um dos pontos mais importantes do Datawarehouse, pois possibilita ter-se bases de dados diferentes, tanto tecnologicamente quanto geograficamente, num mesmo repositório de dados, de forma que a área gerencial da empresa tenha uma visão não somente deste ou daquele setor, mas da organização como um todo. O processo de integração não corresponde somente à colocação de todos os dados em uma única base de dados, mas também num pré-tratamento destas informações, de forma que os dados sejam consistentes. Por exemplo, se no contexto de uma empresa estiver sendo criado um Datawarehouse que irá integrar três bases de dados diferentes, e em cada uma dessas bases o “sexo” é tratado de forma diferente (M/F, Masculino/Feminino, Verdadeiro/Falso), o pré-tratamento terá de ser capaz de preparar esses dados de forma que sejam armazenados de maneira única no Datawarehouse.

A propriedade de não volatilidade se refere ao fato de que o Datawarehouse recebe apenas uma carga de dados inicial e, então fica pronto somente para consultas. Nos bancos de

dados operacionais a atualização ocorre a cada transação; no Datawarehouse a atualização é periódica e incremental, de acordo com uma política pré-definida de acordo com o seu propósito.

A variabilidade se refere ao fato de que o propósito do Datawarehouse é mostrar de que forma as informações contidas nele se comportam em função do tempo. O Datawarehouse tem a propriedade de mostrar como os dados variam ao longo do tempo.

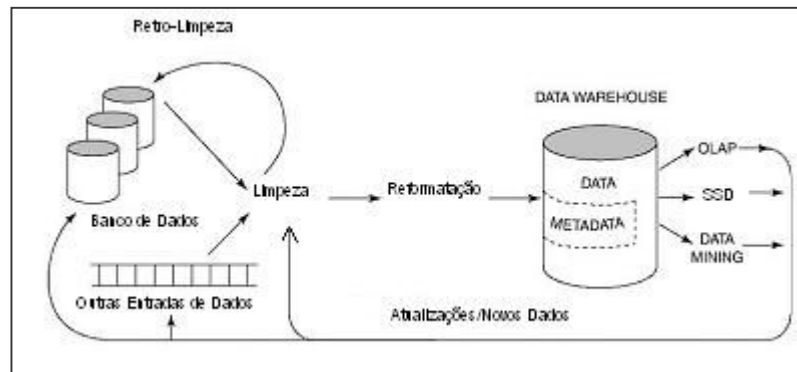


Fig.1 O Processo Completo do Datawarehouse
Fonte: (KIMBALL, 1998)

A figura 1 representa o que é um Datawarehouse e qual o seu processo: a partir de diversas bases de dados vindas do dia-a-dia de uma empresa e até mesmo externas à empresa, é realizado um processo de limpeza, de forma que somente se utilize as informações úteis ao propósito do Datawarehouse para aquela empresa. Após essa limpeza, esses dados sofrem um processo de pré-formatação, de modo a torná-los consistentes em seus resultados, ou seja, apresentem o mesmo formato no final. De posse dessas informações pré-tratadas é que se pode então armazená-las nesse grande depósito que é o Datawarehouse.

Quanto ao pré-tratamento, pode-se destacar três conceitos de extrema importância (ELMASRI; NAVATHE, 2000):

- Extração: Consiste na retirada dos dados dos sistemas transacionais, a partir dos sistemas fonte, banco de dados, arquivos ou ainda de fontes externas, que podem ser dos mais variados formatos.

- Transformação dos dados: inconsistências de formatos podem existir em um único banco de dados e são praticamente certas quando vários bancos de dados são utilizados. Uma vez que as regras de transformação são estabelecidas, definições podem ser criadas e incluídas em rotinas de transformação. A transformação ocorre ao nível de registro ou ao nível de atributo.

- Carga: usa os dados criados no processo de transformação e transporta-os para o Datawarehouse de forma que possam ser utilizados.

O pré-tratamento dos dados para serem armazenados no Datawarehouse é demonstrado na figura 2.

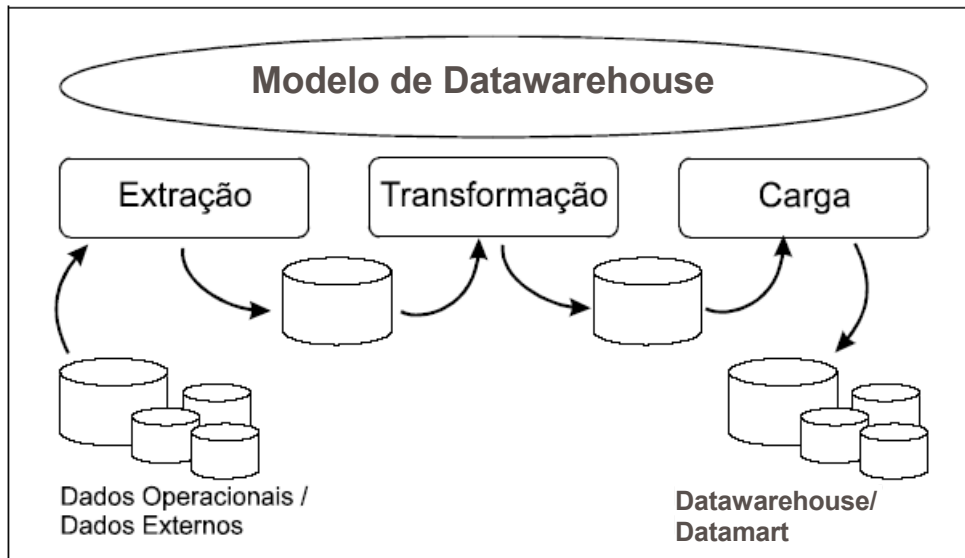


Fig.2 Processo de extração, transformação e carga de dados no Datawarehouse

O Datawarehouse utiliza o modelo multidimensional, que tira proveito das relações de herança de dados para povoá-los em matrizes multidimensionais denominadas “cubos de dados”. Um exemplo de modelo multidimensional é uma matriz de produtos vendidos com os períodos; ter-se-ia nesse caso um modelo de duas dimensões (produtos vendidos e período). Se for acrescentada mais uma dimensão, como cliente, ter-se-ia agora um cubo composto de três dimensões (produtos vendidos, período e cliente).

O modelo multidimensional possui basicamente dois tipos de tabela, denominados, respectivamente, de “tabela dimensão” e “tabela fato”. Nas tabelas dimensão armazenam-se as informações referentes às dimensões do modelo, enquanto que na tabela fato ficam armazenadas as medições relativas ao negócio da empresa.

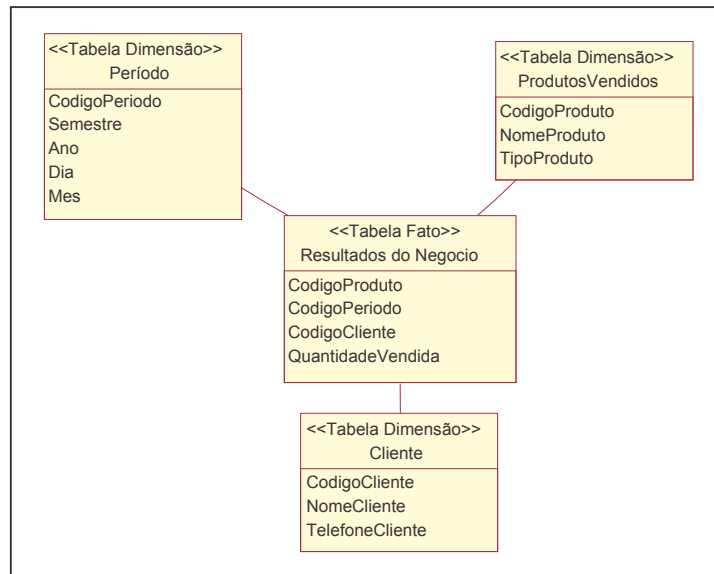


Fig.3 Exemplo de Modelo Lógico Multidimensional

Existem basicamente dois esquemas de modelos multidimensionais: esquema estrela e esquema “snowflake”. O modelo estrela tem como característica a existência de uma única tabela fato para todas as dimensões, conforme pode ser visto na figura 3. O modelo “snowflake” (figura 4) é uma variação do modelo estrela em que as tabelas dimensão são decompostas, formando hierarquias.

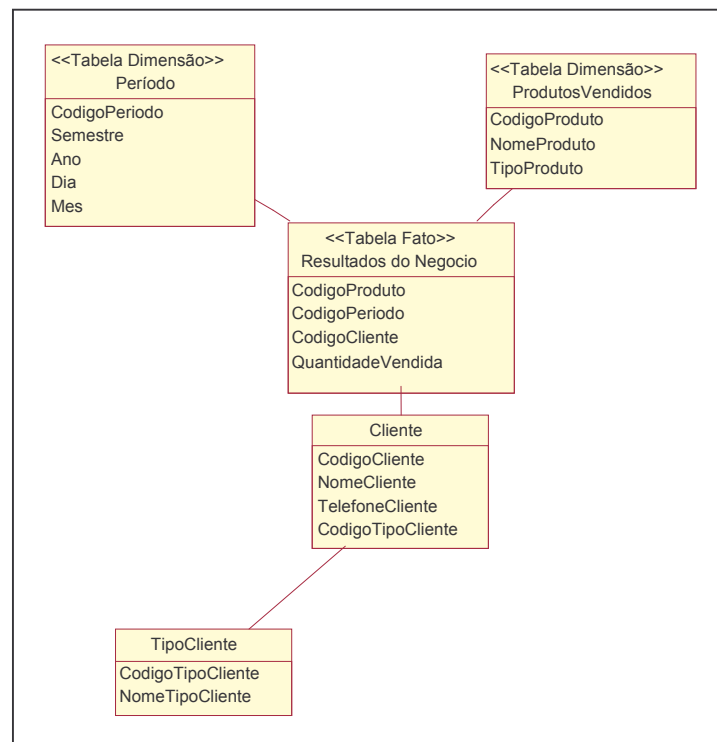


Fig.4 Exemplo de Modelo Lógico “Snowflake”

Para a realização de consultas sobre Datawarehouse são utilizadas duas técnicas: a On-Line Analytical Processing e o DataMining. Estas técnicas estão descritas a seguir.

2.3 “ON-LINE ANALYTICAL PROCESSING” (OLAP)

OLAP se refere a um conjunto de ferramentas que, de forma amigável e flexível, fornece ao usuário acesso e análise de dados ad-hoc, permitindo que armazéns de dados sejam usados efetivamente para análise “online”, provendo respostas rápidas e interativas em consultas analíticas complexas (ELMASRI; NAVATHE, 2000). Com o modelo de dados multidimensional o OLAP é capaz de aplicar suas técnicas de agregação, organização e sumarização em grandes volumes de dados, provendo ainda uma visualização “online” com o uso de ferramentas gráficas.

O OLAP é fundamentado nas seguintes características (ELMASRI; NAVATHE, 2000):

- **Consultas ad-hoc:** é a possibilidade que o usuário tem de criar suas próprias consultas de acordo com as suas necessidades, cruzando informações de diferentes formas e com métodos que permitam a descoberta da informação desejada. Com isso, o OLAP vai além das consultas pré-definidas.

- **Slice-and-Dice:** é a capacidade que a ferramenta OLAP tem de fornecer informações sobre diversos pontos de vista. O termo “slice” vem de fatiamento ou seleção, onde se pode registrar filtros sobre determinadas consultas, tal como a busca de todos os produtos vendidos, dado um determinado período. O termo “dice” se refere à projeção, em que se faz a busca de determinados atributos sobre o DW, como por exemplo a exibição somente da quantidade vendida de um determinado produto.

- **Drill-Down/UP:** Consiste na busca da informação em diferentes níveis de detalhamento, até se alcançar o menor grau de granularidade. Por exemplo buscar as informações de produtos vendidos num ano, num mês, numa semana ou mesmo num único dia.

Quando se fala em granularidade, tenta-se referenciar o nível no qual os dados são sumarizados. Considera-se o grão como o nível mais detalhado do dado. Quanto mais o grão se aproximar de um nível mais detalhado, torna-se possível realizar agregações em qualquer nível. Deve-se lembrar, contudo, que, quanto mais baixo o nível, maior o volume de dados

armazenados, o que reduz a performance das consultas. Por outro lado, quanto maior a granularidade, menos informações de agregação são obtidas.

Apesar da ferramenta OLAP ter seus conceitos e finalidades bem definidos, existem diversas formas de ferramentas OLAP implementadas no mercado, que podem ser agrupadas quanto ao armazenamento e quanto à interface:

Quanto ao armazenamento:

- **MOLAP (“Multidimensional OLAP”)**: neste caso, os dados são armazenados em um banco de dados multidimensional. O usuário trabalha diretamente no servidor. O armazenamento utiliza um espaço em disco menor que seria usado em um banco de dados relacional. Os dados são mantidos em estruturas de dados do tipo “array”, aumentando o desempenho do mesmo. Uma outra vantagem deste tipo de OLAP é o rico e complexo número de funções de análise presentes, além do seu suporte a um grande número de usuários sem degradação do desempenho. Por outro lado, pode-se destacar como desvantagem a falta de padronização, uma vez que esse armazenamento normalmente é proprietário.

- **ROLAP (“Relational OLAP”)**: é o uso da tecnologia de banco de dados relacionais para o armazenamento. Tem como característica a possibilidade de fazer qualquer consulta, uma vez que não está limitado ao que foi definido no cubo. A capacidade de navegação pelos dados é de forma fácil. Apesar da grande vantagem de se estar trabalhando sobre uma tecnologia bem conhecida, aberta e padronizada, pode ser considerada como grande desvantagem a limitação da quantidade de funções que se pode executar sobre os dados e o baixo desempenho da linguagem SQL na execução de consultas complexas.

- **HOLAP (“Hybrid OLAP”)**: Esta arquitetura tenta integrar a capacidade e escalabilidade das ferramentas ROLAP com o desempenho superior dos bancos de dados multidimensionais. De acordo com o volume de dados a ferramenta OLAP executaria sua consulta sobre o banco de dados multidimensional ou no relacional. Para volumes de dados extensos as consultas devem ser realizadas sobre bancos de dados multidimensionais e, em se tratando de dados de menor escalabilidade, sobre bancos de dados relacionais.

Quanto à interface com o usuário:

- **DOLAP (“Desktop OLAP”)**: ferramenta para usuários que possuam uma cópia da base de dados multidimensional ou de um subconjunto dela ou que queiram acessar localmente

uma base de dados. A grande vantagem desta estrutura está na redução da sobrecarga no servidor do banco de dados. Sua desvantagem está na limitação do tamanho do cubo.

- **WOLAP (“Web OLAP”)**: corresponde à utilização de uma ferramenta OLAP tendo como interface um browser. Como vantagem dessa arquitetura pode-se destacar: a distribuição da ferramenta, o acesso remoto aos dados e a utilização de uma ferramenta independente da plataforma.

Uma outra característica do cubo OLAP é a “hierarquização das dimensões”. Esta característica vem para resolver os diferentes níveis de consolidação que se pode ter ao longo de uma dimensão. Por exemplo, numa dimensão tempo com os atributos dia, mês, ano, semestre e trimestre, pode ser definido como hierarquia: ano, semestre, trimestre, mês e dia. Esta situação pode ser visualizada na figura 5.

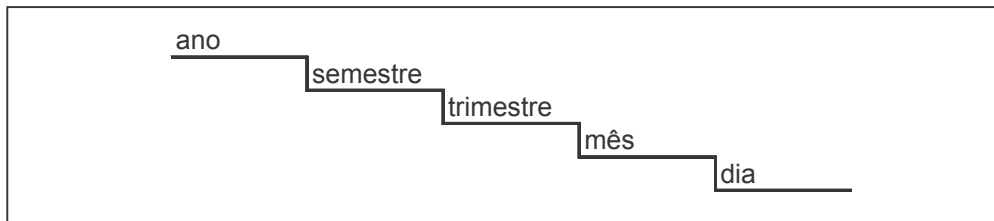


Fig.5 Exemplo de hierarquia da dimensão tempo

Esta informação é útil para a geração de agregações do cubo, pois assim a ferramenta OLAP “sabe” que não há necessidade de agregar, por exemplo, os dados de trimestre e ano, pois os dados trimestrais já foram agregados no nível hierárquico anterior.

Além de criar facilidades de consultas aos usuários, um outro problema que a ferramenta OLAP se propõe a resolver é o tempo de resposta, uma vez que manipula grandes volumes de dados. A solução é o uso da pré-agregação de dados em totais e subtotais. Esse é um processo muito natural, uma vez que se pode facilmente notar a tendência de agrupar regiões geográficas tais como municípios, estados, países, etc; produtos por tipo, e assim sucessivamente. O OLAP irá pré-calculer os resultados possíveis para as agregações, de forma a responder rapidamente as questões do usuário. Obviamente quando o usuário quiser fazer um refinamento da consulta (drill-down) sobre uma certa informação que estiver sendo analisada, haverá uma demora na resposta.

2.4 MINERAÇÃO DE DADOS (“DATAMINING”)

O objetivo do Datamining é retirar informações implícitas de dados explícitos. Elmasri e Navathe (2000) citam que “Datamining ajuda a extrair padrões novos e significativos que não podem ser encontrados meramente consultando ou processando dados ou metadados num Datawarehouse”.

O Datamining é a mineração ou busca por novas informações, em termos de padrões ou regras, a partir de bases de dados. O Datamining pode ser aplicado a qualquer tipo de base de dados, mas seu uso em Datawarehouse pode obter resultados mais precisos, uma vez que o Datawarehouse pode ser planejado levando em conta quais os tipos de buscas irão ser realizadas com o “Datamining”.

O Datamining utiliza diversas técnicas e algoritmos para a análise e obtenção das informações desejadas. Dentre esses algoritmos, encontra-se: estatísticas, árvores de decisão, redes neurais, regressão linear, lógica nebulosa, algoritmos genéticos, etc.

O uso do Datamining não irá retornar consultas simples como quantidade de clientes, quantidade de vendas por bairro, etc. Ao contrário, o processo destina-se a responder a consultas complexas.

2.5 SISTEMA DE INFORMAÇÃO GEOGRÁFICA (SIG)

Rocha (2000) define SIG como “um sistema com capacidade para aquisição, armazenamento, tratamento, integração, processamento, recuperação, transformação, manipulação, modelagem, atualização, análise e exibição de informações digitais georreferenciadas, topologicamente estruturadas, associadas ou não a um banco de dados alfanuméricos.”

No contexto desta dissertação, o SIG será entendido como um Sistema de Suporte à Decisão que tem como principal característica de análise o dado geográfico. Para isso, a tecnologia SIG utiliza-se de um conjunto de disciplinas e técnicas como geografia, cartografia, sensoriamento remoto, fotogrametria, modelagem de dados, etc. As técnicas utilizadas em SIG são denominadas de “geotecnologias”.

Os dados geográficos caracterizam-se pelo componente posicional, relativo a um referencial terrestre. Este referencial pode ser cartesiano, esférico ou elipsóidico, conforme o modelo matemático adotado para o espaço geográfico de interesse. Os dados podem ser obtidos de diversas fontes, tais como: fotografias aéreas, imagens de satélite, mapas, e pesquisas de campo. Para a aquisição das mesmas podem ser usados dispositivos diferentes, em formatos distintos, requerendo funções de análise e apresentação especializadas, não presentes em sistemas de banco de dados convencionais, e que fornecem ao SIG características de peculiaridade, muitas vezes chamados de sistemas de informação não convencional ou sistemas que manipulam objetos complexos ou longos.

Outra peculiaridade dos SIG é que estes, além de manipular itens de dados chamados “convencionais” (números, textos, booleanos, etc), também manipulam dados espaciais (geometria e relacionamentos topológicos das entidades geográficas) e dados pictóricos. Estes dados podem ser representados em dois formatos (ARONOFF, 1995):

- Vetorial: representam objetos geométricos por intermédio de pontos, linhas, e polígonos. Nesse formato, um lago pode ser representado como um polígono, um rio por uma série de segmentos de linha, e assim por diante.
- Matricial ou Raster: que se caracteriza por uma matriz de pontos, onde cada ponto representa o valor de um atributo fornecido por uma localização geográfica do mundo-real.

Os dados em um SIG sofrem vários tipos de análise. Por exemplo, em aplicações tais como a detecção do potencial de erosão de solo deve-se analisar o impacto ambiental, e para este tipo de descoberta será necessário o uso de diversas ferramentas de análise dos SIG.

Uma das características desejáveis do SIG é a integração, pois um Sistema de Informação Geográfica tem que ser capaz de manipular dados do tipo vetor e raster ao mesmo tempo e de diversos tipos de fontes, dados esses que são originados de sistemas de coordenadas diferentes, requerendo transformações apropriadas, para manter a consistência da análise espacial.

Num SIG, os dados são organizados em camadas ou planos de informação. Aronoff (1995) conceituou camada de dados como “um conjunto de características e atributos geográficos relacionadas logicamente”. Desta forma, pode-se encontrar camadas do tipo rodoviária, ferroviária, hidrográfica, distritos, bairros, ruas, etc.

Segundo Rocha (2000), um SIG deve possuir as seguintes características:

“- Ter capacidade para coletar e processar dados espaciais obtidos a partir de fontes diversas, tais como: levantamentos de campo (incluindo o sistema GPS), mapas existentes, fotogrametria, sensoriamento remoto e outros;”

“- Ter capacidade para armazenar, recuperar, atualizar e corrigir os dados processados de forma eficiente e dinâmica;”

“- Ter capacidade para permitir manipulações e a realização de procedimentos de análise dos dados armazenados, com possibilidade de executar diversas tarefas, tais como, alterar a forma dos dados através de regras de agregação definidas pelo usuário, ou produzir estimativas de parâmetros e restrições para modelos de simulação e gerar informações rápidas a partir de questionamentos sobre os dados e suas interrelações;”

“- Ter capacidade para controlar a exibição e saída de dados em ambos os formatos, gráfico e tabular”.

Para Elmasri e Navathe (2000) um SIG deve ser capaz de realizar as seguintes operações espaciais: interpolação, interpretação, análise de proximidade, processamento de imagens e análises de rede.

Segundo Aronoff (1995), interpolação é o “procedimento de prever valores desconhecidos usando valores conhecidos em localizações vizinhas”. Um exemplo, seria a reconstituição de pontos perdidos através dos pontos existentes. Para a realização dessa operação, pode-se usar regressão polinomial, séries de Fourier, etc.

A edição se preocupa com operações sobre terrenos tais como: suavização, redução de detalhes, aumento, realce de bordas de feições do terreno e fusão.

A análise de proximidade se refere aos algoritmos que realizam operações de distância entre objetos. Aronoff (1995) informa quatro parâmetros necessários para medir proximidade: um alvo, uma unidade de medida, uma função de cálculo de proximidade e a área a ser analisada.

Quanto ao processamento de dados modelados matricialmente, pode-se distinguir duas categorias:

- Mapa Algébrico: integra características geográficas em camadas de mapas diferentes para produzir um novo mapa. Esta operação é também como “Análise de Overlay” e fundamenta-se na lógica booleana;
- Análise de Imagens: analisa uma imagem digital para a descoberta de características ou extração de objetos geográficos. Por exemplo, encontrar estradas numa imagem em nível orbital.

Segundo Aronoff (1995), as análises de redes são normalmente usadas para “mover recursos de um local para outro. As ruas de uma cidade, uma malha de linhas de transmissão elétrica, uma rota de serviço aéreo, ou os fluxos de drenagem de uma bacia são exemplos de

redes.“ Um SIG pode ser usado para executar três tipos principais de análises de redes: predição do carregamento na rede, otimização de rotas, e distribuição de recursos de centros. Um exemplo típico da distribuição de recursos é a alocação de alunos a escolas, em função da localização de suas residências e de seu nível de escolaridade.

2.5.1 A análise sobre dados geográficos

Segundo Dias, Oliveira e Câmara (2002) “além da percepção visual da distribuição espacial do problema, é muito útil traduzir os padrões existentes com considerações objetivas e mensuráveis”. Para isso, podem ser modelados ou determinadas as seguintes visões:

- Epidemiológicas: distribuição da doença por um espaço geográfico.
- Policiais: distribuição espacial de crimes.
- Geológicas: estimativa e estudo de uma reserva de algum recurso mineral.
- Agrícola: distribuição de cultura de acordo com tipo de solo, vegetação, relevo, etc.

Um dos fatores com que se deve ter muita cautela no tratamento das informações, segundo Dias, Oliveira e Câmara (2002) é a preservação dos dados individuais. Para resolver esse problema usa-se a agregação dos dados, neste caso, dados geográficos.

O Censo utiliza como menor unidade de agregação o **setor censitário**, que para o Instituto Brasileiro de Geografia e Estatística (IBGE) corresponde à capacidade de levantamento do recenseador, variando por região em torno de 200 a 400 domicílios.

Quanto mais alto o nível de agregação menor a precisão da informação. Por exemplo, se consideradas partições no espaço, tais como: setor censitário, bairro, distrito e município, o setor censitário tende a ser mais preciso. Quanto maior a área, menor será a correlação entre as variáveis.

Este trabalho irá se apoiar sobre este conceito para possibilitar ao usuário a navegação entre as diversas agregações (geográficas ou não). Nesta dissertação o termo “nível de agregação” está sendo denominado “granularidade”, uma vez que estes conceitos são próximos, sendo que o primeiro vem do SIG e o segundo vem do DW.

2.6 O CONSÓRCIO “OPEN GIS”

Buehler e Mckee (2003) definem o “Open GIS Consortium” (OGC) como uma organização internacional que está criando novas padronizações técnicas e comerciais para garantir interoperabilidade em SIG. Fundada em 1994 por fornecedores de software, companhias de telecomunicações, universidades, provedores de informação e órgãos governamentais, entre outros, a OGC busca criar uma especificação de software e novas estratégias empresariais, a fim de tornar os sistemas de geoprocessamento abertos e integrar completamente os dados geográficos e as operações necessárias para manipulá-los.

Parte dos membros da OGC têm uma visão positiva de uma infra-estrutura de informação global em que dados geográficos e recursos de geoprocessamento são distribuídos gratuitamente. Esses recursos estão completamente integrados com as mais recentes tecnologias de computação distribuída, acessíveis para o público em geral, habilitando uma grande variedade de atividades que, atualmente, estão fora do domínio de geoprocessamento. Entre alguns dos membros desse consórcio estão empresas como a Microsoft e a Oracle.

O OpenGIS criou uma especificação, a fim de definir um esquema padrão de armazenamento, recuperação e atualização de dados espaciais. Essa especificação será descrita a seguir:

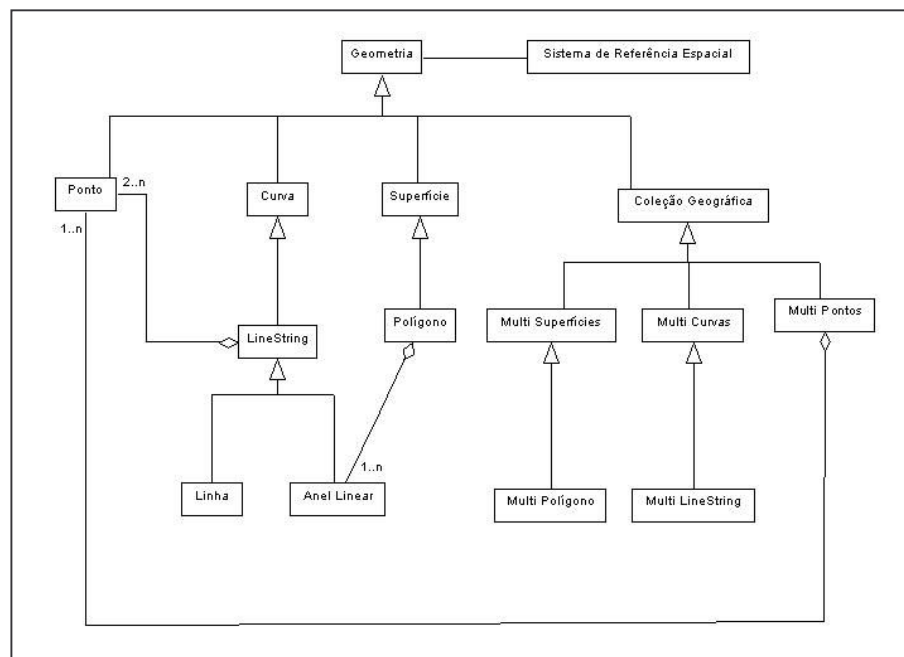


Fig.6 Hierarquia de Classes Geométricas – OpenGIS

Fonte: (OGC, 1999)

A Figura 6 especifica a coleção de classes de zero, uma ou duas dimensões nomeadas, respectivamente, por Multi-Pontos, Multi-Linhas e Multi-Polígono. Esta coleção é utilizada para a modelagem de objetos geográficos. Multicurva e MultiSuperfície são introduzidos como superclasses abstratas.

Como classes primitivas existem as classes ponto, curva e superfície. O ponto representa uma localização geográfica única. A curva é um objeto unidimensional representado por uma seqüência de pontos. A superfície é um objeto geográfico bidimensional.

Um polígono é uma superfície plana, definida por um limite exterior e um ou mais limites interiores. Cada limite interior define um buraco no polígono.

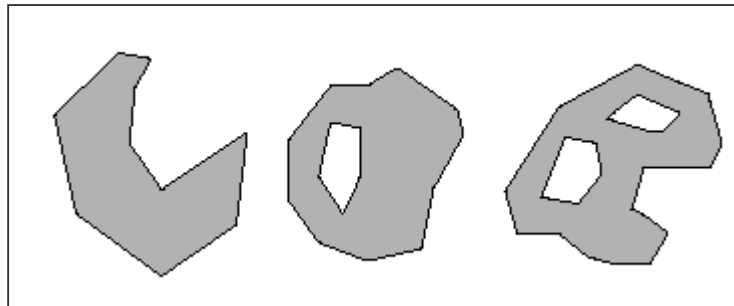


Fig.7 Exemplos de Polígonos
Fonte: (OGC, 1999)

A figura 7 representa exemplos de polígonos. No primeiro exemplo, o polígono só possui o limite externo; no segundo caso existe o limite externo e um limite interno (com um único buraco), e no terceiro exemplo, além do limite externo, existem dois limites internos (dois buracos).

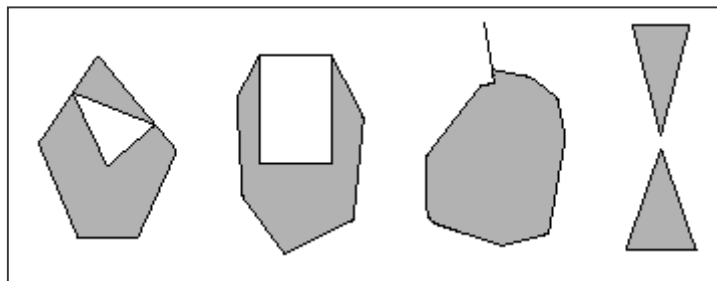


Fig.8 Exemplos de Não-Polígonos
Fonte: (OGC, 1999)

A figura 8 exibe figuras que não são consideradas pelo OGC como polígonos simples. O primeiro e o quarto exemplo representam, na verdade, dois polígonos distintos. O segundo exemplo não consiste um polígono, uma vez que existe uma linha que liga dois de seus vértices. No terceiro exemplo, existem pontos de um limite externo do polígono que não estão presentes no polígono.

A curva é um objeto unidimensional geralmente armazenado com uma seqüência de pontos. O LineString é uma curva com interpolação linear entre pontos. O LineString possui dois tipos: linha e anel linear, conforme pode ser visto na figura 9.

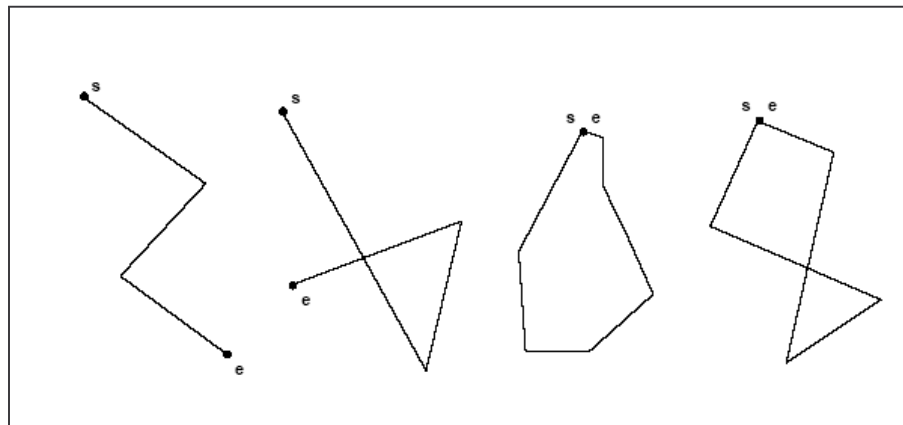


Fig.9 Exemplos de LineString
Fonte: (OGC, 1999)

A Linha é um LineString com exatamente dois pontos. O Anel Linear é um LineString fechado e simples. Entende-se como fechada a curva que possui o ponto inicial igual ao ponto final. A característica de simplicidade se dá quando não existe passagem pelo mesmo ponto duas vezes.

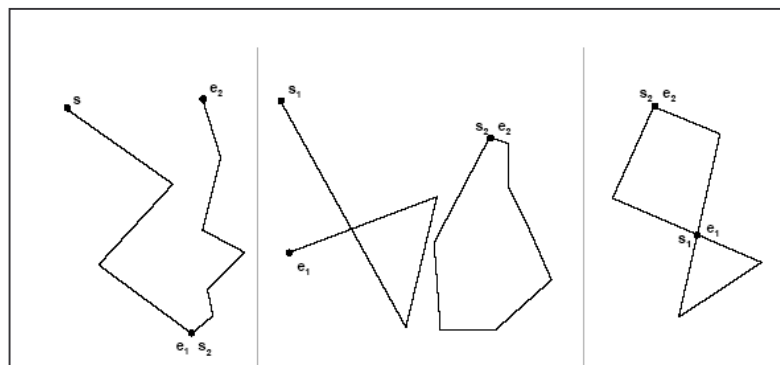


Fig.10 Exemplos de MultiLineString
Fonte: (OGC, 1999)

O Multiponto é uma coleção geométrica zero-dimensional, restringido a formação por pontos. Os pontos não são conectados ou ordenados. A Multi-curva é uma coleção unidimensional de curvas. O MultiLineString (figura 10) é uma MultiCurva formada por LineString. A Multi-Superfície é uma coleção geométrica bidimensional de superfícies. Os interiores de qualquer duas superfícies em uma coleção Multi-Superfície podem possuir interseção. O Multi-Polígono é uma Multi-Superfície formada por polígonos.

É importante lembrar que o Consórcio Open GIS define como tipo de dados instanciáveis apenas: ponto, linha, anel linear, polígono, multi-polígono, multi-linestring e multiponto. Serão esses os tipos geográficos que serão tratados no modelo de dados multidimensional deste trabalho.

2.7 MODELAGEM

Um sistema é representado por diversos modelos que são elaborados durante o processo de desenvolvimento. Os modelos projetam o sistema em diferentes dimensões e níveis de abstração. Um modelo é uma representação metafórica do mundo real, enfatizando algumas características e desprezando outras. Existem diversas razões para modelar, tais como: facilitar a compreensão e a visualização do problema ou fenômeno; facilitar a comunicação; validar a custo baixo, e documentar um fato.

A dimensão de um modelo corresponde ao conjunto de características que o modelo enfatiza ao representar um fenômeno ou fato. Basicamente as dimensões de um sistema são as seguintes: dado, função e controle. No processo de desenvolvimento de um sistema de informação, normalmente, são utilizados os seguintes níveis de abstração: **conceitual**, cujas características dependem apenas do mundo real; **tecnológico**, onde as características dependem da tecnologia utilizada; e **interno**, no qual as características estão em um nível de detalhe compreendido pelo computador (bits e bytes) (YOURDON, 1990).

Metodologia é a maneira de se utilizar um conjunto coerente e coordenado de métodos para atingir um objetivo. Em outras palavras, a metodologia deve definir quais as fases de trabalho previstas no desenvolvimento de sistemas. **Método** é um procedimento a ser adotado para se atingir um objetivo. **Técnica** é um modo apropriado de se investigar sistematicamente

um determinado universo de interesse ou domínio de um problema. Ex: análise estruturada, análise essencial e projeto estruturado, e análise orientada a objetos (FOWLER; SCOTT, 2000).

Este trabalho irá desenvolver um modelo de dados capaz de suprir as necessidades de documentação e extração de informações sobre as diversas bases de dados da Prefeitura Municipal de Macaé. Segundo Lisboa (2002) “Na fase de projeto conceitual é necessário definir o modelo de dados a ser utilizado, o que vai facilitar a comunicação entre projetistas e/ou usuários envolvidos.” No caso desta dissertação, como se tem uma modelagem mesclada entre Dados Convencionais e Dados Geográficos, é uma preocupação o uso de uma ferramenta de análise capaz de suprir essa necessidade: modelagem geográfica e convencional.

2.7.1 “Unified Modeling Language” (UML)

A UML foi desenvolvida por Grady Booch, James Rumbaugh, e Ivar Jacobson que são conhecidos no meio técnico científico como “os três amigos”. Eles possuem extenso conhecimento na área de modelagem orientada a objetos. Na realidade três das mais conceituadas metodologias de modelagem orientada a objetos foram desenvolvidas por esses autores. Nesse contexto, a UML é a junção do que havia de melhor nestas três metodologias, gerando uma poderosa ferramenta de modelagem.

A UML aborda o caráter estático e dinâmico do sistema a ser analisado levando em consideração todas as futuras características do sistema em relação à utilização de pacotes próprios da linguagem, o uso do banco de dados, bem como as diversas especificações do sistema a ser desenvolvido, de acordo com as métricas finais desejadas (FOWLER; SCOTT, 2002).

A orientação a objetos é uma tecnologia para a produção de modelos que especifiquem o domínio do problema de um sistema. Quando construídos corretamente, os sistemas orientados a objetos são flexíveis a mudanças, possuem estruturas bem conhecidas e provêm oportunidade de criar e implementar componentes totalmente reutilizáveis. Os sistemas OO são implementados convenientemente utilizando uma linguagem de programação orientada a objetos. A engenharia de software orientada a objetos é muito mais que utilizar mecanismos de sua linguagem de programação; é saber utilizar da melhor forma possível todas as técnicas da modelagem orientada a objetos.

A UML é uma tentativa de padronizar a modelagem orientada a objetos de uma forma que qualquer sistema, seja qual for o tipo, possa ser modelado corretamente, com consistência, facilidade de comunicação com outras aplicações, simplicidade de atualização e compreensão (FOWLER; SCOTT, 2000). Existem várias metodologias de modelagem orientada a objetos que, até o surgimento da UML, causavam uma “guerra” entre a comunidade de desenvolvedores. A UML acabou com esta “guerra”, trazendo as melhores idéias de cada uma destas metodologias e mostrando como deveria ser a migração de cada uma para a UML.

Os objetivos da UML são:

- A modelagem de sistemas em geral e não apenas de software, usando os conceitos da orientação a objetos;
- O estabelecimento de uma união, fazendo com que métodos conceituais sejam também executáveis;
- A criação de uma linguagem de modelagem usável tanto pelo homem quanto pela máquina.


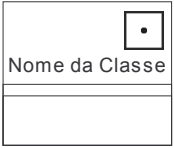
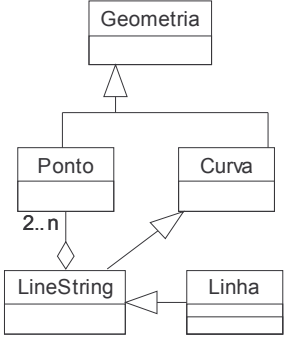
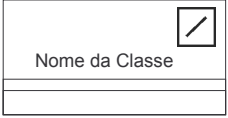
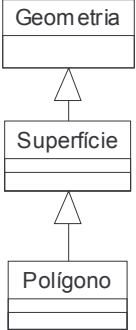
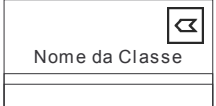
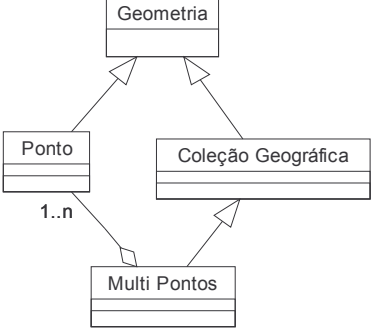
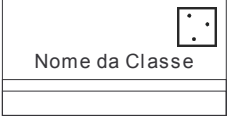
A UML está destinada a ser dominante: a linguagem de modelagem comum a ser usada nas aplicações práticas. Ela está totalmente baseada em conceitos e padrões extensivamente testados, provenientes das metodologias anteriormente existentes, sendo também muito bem documentada, com toda a especificação da semântica da linguagem representada em meta-modelos.

2.7.2 A utilização de estereótipos na modelagem

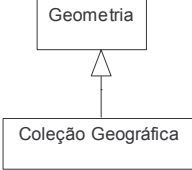
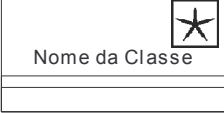
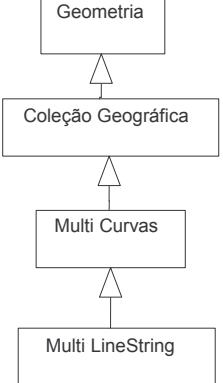
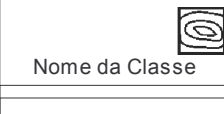
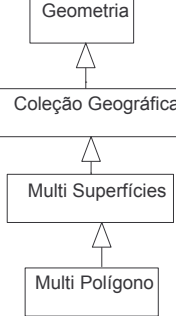
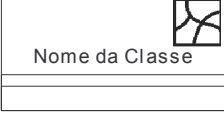
Segundo Colonese (2004), “o emprego da orientação a objetos na modelagem de Datawarehouses traz também a importante possibilidade do uso de estereótipos, que simplificam sobremaneira a representação de hierarquias extensas de objetos, substituindo a reprodução integral da hierarquia por um símbolo ou ícone representativo da superclasse com a qual o objeto corrente possui uma relação”.

Baseado nesta idéia, além do uso de todos os construtores da UML, serão usados estereótipos que deixarão o modelo mais simples e com um alto nível de representatividade. Desta forma, ao invés de representar toda a linha de dependências vista na figura 4 para a

representação de qualquer classe do tipo geográfica, usa-se os estereótipos para simplificar essa representação, conforme representado na tabela 1.

Tipo de Dado	Padrão Open GIS	Classe com estereótipo
Ponto		
Linha		
Polígono		
Multiponto		

Tab. 1 - Comparação do modelo Open GIS com o modelo UML com estereótipos (continua)

Tipo de Dado	Padrão Open GIS	Classe com estereótipo
<p>Coleção Geométrica</p>	 <pre> classDiagram class Geometria class ColeçãoGeográfica["Coleção Geográfica"] Geometria < -- ColeçãoGeográfica </pre>	
<p>Multilinha</p>	 <pre> classDiagram class Geometria class ColeçãoGeográfica["Coleção Geográfica"] class MultiCurvas["Multi Curvas"] class MultiLineString["Multi LineString"] Geometria < -- ColeçãoGeográfica ColeçãoGeográfica < -- MultiCurvas MultiCurvas < -- MultiLineString </pre>	
<p>Multipolígono</p>	 <pre> classDiagram class Geometria class ColeçãoGeográfica["Coleção Geográfica"] class MultiSuperfícies["Multi Superfícies"] class MultiPolígono["Multi Polígono"] Geometria < -- ColeçãoGeográfica ColeçãoGeográfica < -- MultiSuperfícies MultiSuperfícies < -- MultiPolígono </pre>	

Tab 1 - Comparação do modelo Open GIS com o modelo UML com estereótipos (Continuação)

A figura 11 é um exemplo comparativo da representação com a hierarquia de dados e o modelo com o uso de estereótipos.

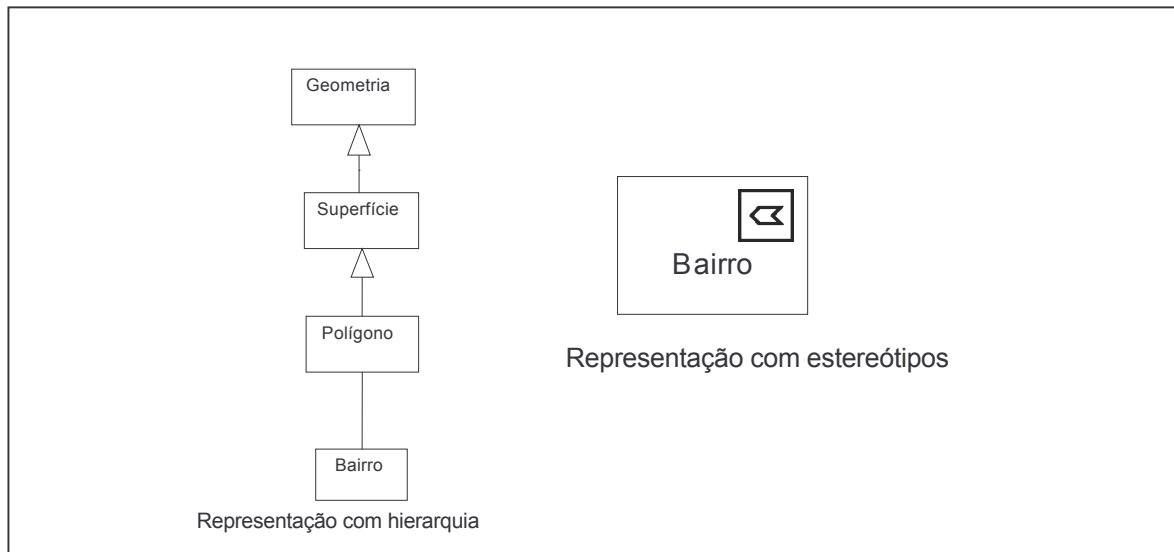


Fig.11 Exemplo de representação com estereótipos - comparação da representação da hierarquia com a representação com estereótipos

2.7.3 Modelando um Datawarehouse com características espaciais

Na verdade a modelagem de um Datawarehouse com características espaciais não apresenta muitas diferenças em relação à modelagem de um Datawarehouse sem essas características.

No caso do presente trabalho será apresentado um modelo estrela, deixando toda a estrutura de hierarquia para a característica de hierarquização de dimensão geográfica.

2.7.3.1 Definição da tabela fato

Antes de falar da tabela fato, deve-se primeiro abordar o significado do termo “fato” no contexto de um Datawarehouse. Um fato é uma coleção de itens de dados relacionados, consistindo de métricas e contexto de dados (NAVATHE; ELMASRI, 2001). É, na verdade, o dado sobre o qual se deseja fazer análises. Se pensar, por exemplo numa venda de produtos, um

fato poderia ser a quantidade de produtos vendidos. De maneira geral, cada fato representa um item ou transação de negócios.

A tabela fato representa uma coleção de fatos. Segundo Kimball (1998), as tabelas fato mais úteis contém fatos numéricos e aditivos que são importantes porque as aplicações de Datawarehouse normalmente recuperam muitos registros de uma única vez e uma tendência é somá-los para análise, o que não deixa de ser uma "compactação" que trará ganhos de performance.

Para definir essa tabela, deve-se então ter a preocupação com o foco do Datawarehouse, ou seja: o que se deseja realmente analisar? Que pontos devem ser analisados? O que se deseja somar, de acordo com as características? O que deve ser contado, de acordo com as características?

2.7.3.2 Definição das dimensões

A dimensão é uma coleção de membros ou unidades do mesmo tipo de visões. São características sobre as quais se deseja visualizar o fato.

Por exemplo, a quantidade de produtos vendidos da categoria X no segundo semestre de 2003, conforme pode ser visto na figura 12. Neste caso, o fato é a quantidade de produtos vendidos (sumarização do tipo soma) e a categoria é uma dimensão sobre a qual se aplica um foco de visão sobre o fato. Da mesma forma, o tempo pode ser analisado como uma dimensão a partir da qual é possível realizar um filtro sobre os dados da tabela fato.

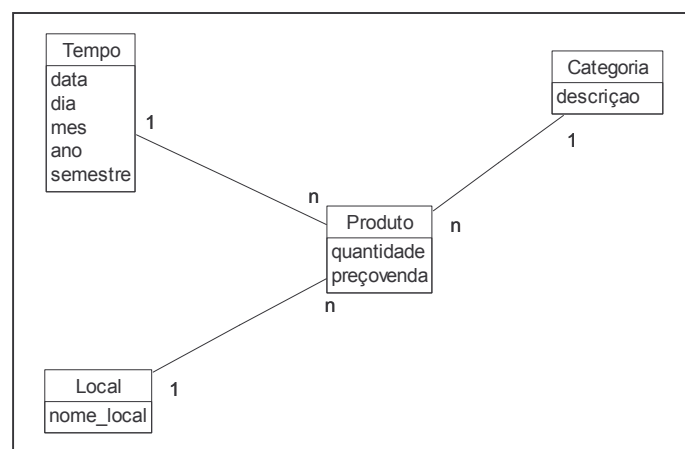


Fig.12 Exemplo de modelo multidimensional – fato e dimensão

No diagrama, uma dimensão é usualmente representada por um eixo. Em um modelo dimensional estrela, cada ponto de dado em uma tabela **fato** é associado com um e somente um membro de cada uma das múltiplas dimensões.

As tabelas dimensionais armazenam as descrições textuais das dimensões do negócio. A tabela **fato** possuirá como chave primária o conjunto de todas as chaves estrangeiras das dimensões, enquanto que as tabelas correspondentes às dimensões apresentam um código identificador único como sendo sua chave primária.

2.7.3.3 Definição de dimensões geográficas no Datawarehouse

A proposta deste trabalho é a definição de um modelo multidimensional estrela que terá como principal característica a existência de dimensões do tipo geográfico. Logo, baseado nos tipos de dados preconizados pelo Open GIS, poderão haver dimensões do tipo ponto, linha, polígono, multipolígono, multilinha e coleção geográfica.

Colonese (2004), define como tipos de dimensões geográficas possíveis no modelo multidimensional e que serão utilizadas neste trabalho: dimensão geográfica agregável e dimensão geográfica não agregável.

2.7.3.3.1 Dimensão Geográfica Agregável

Entende-se por dimensão geográfica agregável, toda dimensão de um modelo que seja um dos pontos de análise sobre a tabela fato. A figura 13 mostra um exemplo de dimensão geográfica do tipo polígono (Bairro), que é um dos pontos de análise sobre o indivíduo, como por exemplo onde determinado indivíduo habita, quais indivíduos moram próximos, etc.

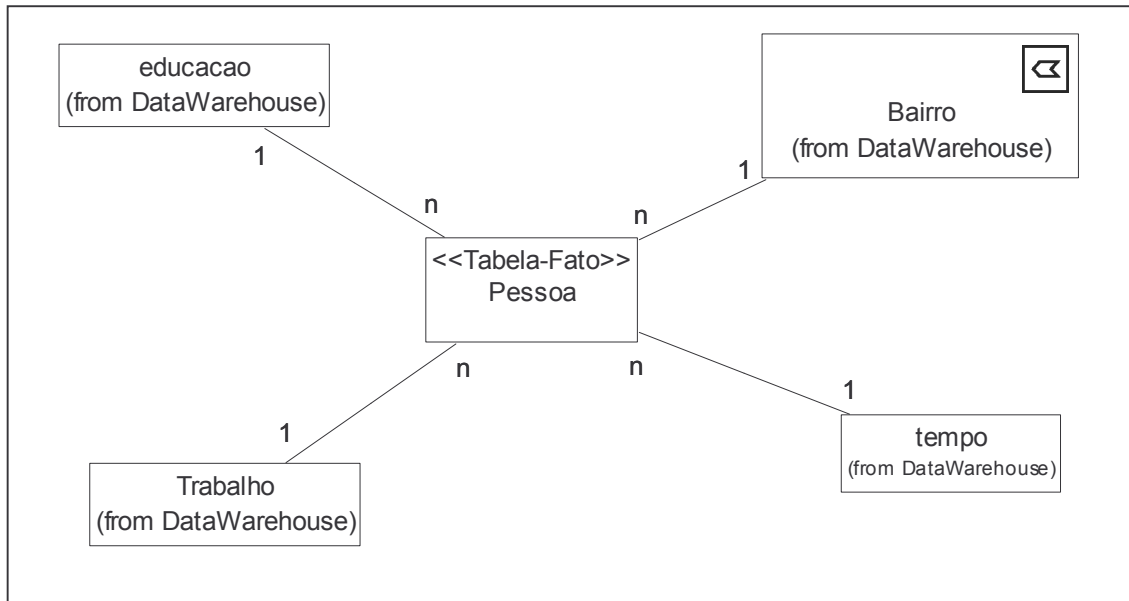


Fig.13 Exemplo de dimensão geográfica agregável

Observa-se, no exemplo da figura 13, que se pode avaliar as condições de educação num determinado intervalo de tempo, sobre uma pessoa. Além disso pode-se analisar quais as condições de trabalho dessa pessoa. Ainda existe a hipótese de consulta sobre o bairro no qual uma pessoa mora. Desta forma, o Bairro pode ser denominado uma dimensão geográfica agregável, uma vez que representa um dos ângulos de análise ou visão em ponto de vista da tabela fato.

2.7.3.3.2 Dimensão geográfica não agregável

A dimensão geográfica não agregável se refere a dimensões que têm como característica o posicionamento geográfico, mas que não possuem correlação direta com a tabela fato do modelo multidimensional. Ou seja, a dimensão não representa um dos pontos de vista de análise sobre os dados da tabela fato, mas sim um comparativo com alguma dimensão geográfica agregável.

Se no exemplo da figura 13 fosse acrescentada uma dimensão do tipo escola, esta não teria nenhum tipo de ligação com a pessoa. Qual a ligação entre escola e pessoa? Diretamente nenhuma, mas pode-se verificar quantas pessoas moram longe de uma escola. Esta consulta seria implementada a partir do relacionamento entre o local onde a pessoa mora (Bairro –

Dimensão geográfica agregável) e a Escola (Dimensão geográfica não agregável). Este exemplo pode ser visto na figura 14.

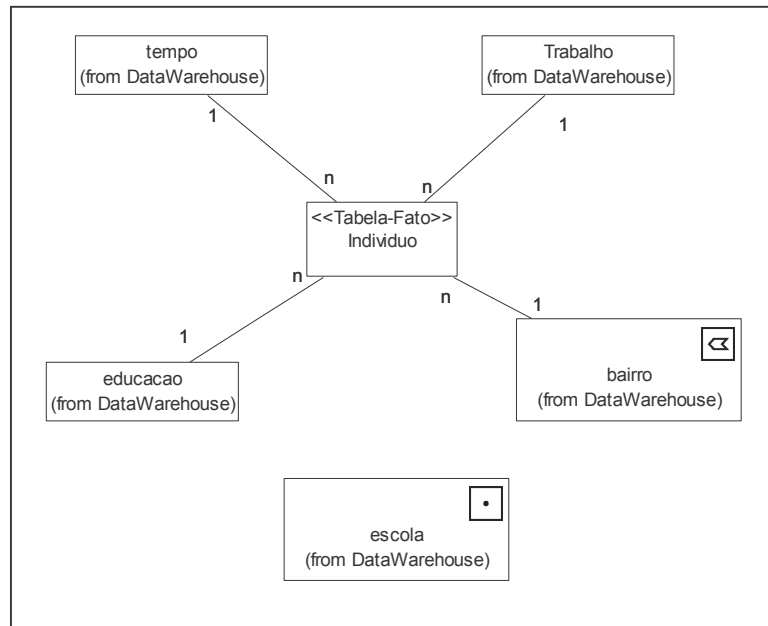


Fig.14 Exemplo de dimensão geográfica não-agregável

É importante lembrar que, apesar de aparentemente não haver relacionamento entre escola e bairro, na verdade existe um relacionamento topológico em nível de abstração diferente: no posicionamento geográfico. Um ponto pode estar contido em um polígono ou não; pode estar distante de um polígono ou não, etc. Desta forma, o relacionamento entre a dimensão geográfica agregável e a dimensão geográfica não agregável se dá por intermédio do posicionamento geográfico relativo entre a pessoa e a escola.

2.8 FORMATO “ARCVIEW SHAPE FILE”

O “shapefile” é um formato desenvolvido pela Environmental Systems Research Institute Inc (ESRI) capaz de armazenar informações geométricas não topológicas e atributos para uma feição espacial em um conjunto de dados. A geometria de uma feição é armazenada num arquivo comprimido com um conjunto de vetores de coordenadas.

O ESRI (1998) diz que, graças ao fato desse formato não ter processamentos complexos sobre sua estrutura de dados topológica, apresenta como vantagem a facilidade e velocidade no

desenho do mapa e a habilidade na edição do mesmo. Além disso, requer pouco espaço em disco e facilidade na leitura e na escrita.

Este formato suporta pontos, linhas e polígonos. Seus atributos são armazenados em arquivo no formato dBASE. Cada atributo do arquivo “shapefile” possui um relacionamento de um-para-um com os registros do arquivo dBASE.

Obviamente, este formato oferece poucos recursos de operação sobre mapas, mas praticamente se tornou um padrão de troca de informações geográficas, principalmente pelo tamanho de seus arquivos que ocupam pouquíssimo espaço em disco. São muitos os softwares de SIG que suportam trabalhar com este arquivo, além de exportar mapas desse padrão e importar mapas para o formato proprietário do software em questão.

2.9 POSTGRESQL

O PostgreSQL (POSTGRESQL.ORG, 2004) corresponde a um banco de dados objeto-relacional Espacial, ou seja, é um banco de dados capaz de armazenar dados convencionais e dados geográficos. Quanto aos dados geográficos este Sistema de Gerência de Banco de Dados (SGBD) é capaz não só de armazenar estes tipos de dados, mas também de executar funções geográficas sobre esses dados.

O PostgreSQL se tornou um banco de dados muito interessante para ser utilizado neste trabalho uma vez que, além de todas essas capacidades ditas anteriormente, é uma ferramenta de distribuição gratuita o que viabiliza o seu uso não só para a aplicação corrente mas também para muitas aplicações do mesmo tipo.

O Datawarehouse e as ferramentas OLAP estão se tornando ferramentas fundamentais para uma boa administração de uma empresa, seja pública ou privada, mas que se inviabilizam pelo seu custo de implantação. Usando ferramentas gratuitas ou de distribuição livre, permite-se que tais tecnologias se tornem mais acessíveis a uma grande variedade de empresas.

O PostgreSQL é um SGBD desenvolvido para funcionar também em ambientes Linux. Isto proporciona mais uma grande economia na implantação de um sistema com um banco de dados como este, uma vez que o Linux é um sistema operacional gratuito e livre com grande aceitação no mercado nos dias de hoje. Este reconhecimento se dá em termos de

funcionalidades, gerenciamento de redes de computadores, e facilidade na interface com o usuário no que se refere ao uso nas estações de trabalho.

É importante ressaltar que, apesar do PostgreSQL ser desenvolvido para operar em sistemas operacionais Linux, nada impede que estações com sistemas operacionais Windows (2000, XP, 98, etc) possam acessar os dados armazenados no mesmo. Existe uma versão disponível para Windows, que até a conclusão desta dissertação, ainda se encontrava em fase de testes.

Uma outra característica que viabilizou o uso deste SGBD para a realização deste trabalho é a característica geográfica (armazenamento e funções). Os SGBDs com estas características são poucos no mercado e ao mesmo tempo muito caros. Dentre eles pode-se citar o SQL Server e o Oracle.

2.9.1 PostGIS

O PostgreSQL é um SGBD que suporta extensões em sua funcionalidade, adicionando a ele as seguintes flexibilidades: criação de tipos de dados, funções, operadores, funções agregadas, métodos de indexação e utilização de linguagens procedurais.

Na verdade, o PostgreSQL não possui nativamente todos os tipos e operações geográficas contempladas pelo Open GIS. Esses tipos e operações são implementados através de uma extensão do PostgreSQL conhecida por PostGIS (POSTGIS, 2004). Logo, o PostGIS implementa as definições do OpenGIS para a SQL, permitindo desta forma que a SQL tenha funções geográficas.

O PostGIS, foi desenvolvido por Refractions Research Inc, como uma tecnologia de banco de dados espacial. Sua proposta é a implantação de um conjunto de importantes funcionalidades de SIG, incluindo o suporte completo ao OpenGIS, construtores topológicos avançados (coberturas, superfícies e redes), ferramentas de interface com o usuário para visualização e edição de dados geográficos e ferramentas de acesso baseado em Web (OGC, 1999).

2.9.1.1 Tipos geográficos do PostGIS

A versão 0.8 do PostGIS, contempla os seguintes tipos de dados: Ponto, Linha, Polígono, MultiPonto, MultiLinha, MultiPolígono e Coleção Geométrica, ou seja, todos os tipos geográficos instanciáveis contemplados pelo Open GIS (OGC, 1999). A figura 15 mostra a definição de cada tipo de dado preconizado pelo PostGIS.

<ul style="list-style-type: none"> • POINT(0 0 0) • LINESTRING(0 0,1 1,1 2) • POLYGON(((0 0 0,4 0 0,4 4 0,0 4 0,0 0 0)),(1 1 0,2 1 0,2 2 0,1 2 0,1 1 0)) • MULTIPOINT(0 0 0,1 2 1) • MULTILINESTRING(((0 0 0,1 1 0,1 2 1)),(2 3 1,3 2 1,5 4 1)) • MULTIPOLYGON((((0 0 0,4 0 0,4 4 0,0 4 0,0 0 0)),(1 1 0,2 1 0,2 2 0,1 2 0,1 1 0)),((-1 -1 0,-1 -2 0,-2 -2 0,-2 -1 0,-1 -1 0))) • GEOMETRYCOLLECTION(POINT(2 3 9),LINESTRING((2 3 4,3 4 5)))

Fig.15 Tipos de dados geográficos e suas definições no OpenGIS
Fonte: PostGIS Manual, 2004

É importante lembrar que o OpenGIS suporta tanto dados bidimensionais quanto dados tridimensionais. Alguns exemplos da figura 13 estão em 3 dimensões outros em 2 dimensões. O ponto (POINT) está definido em formato tridimensional, já a linha (LINESTRING) está definida como bidimensional.

2.9.1.2 Funções geográficas do PostGIS

Além dos tipos geográficos do OpenGIS, o PostGIS também implementou as funções geográficas. A tabela 2 exhibe as principais funções. É importante lembrar que a tabela não possui a especificação de todas as funções geográficas contemplada pelo PostGIS. Apenas foram destacadas as funções julgadas mais relevantes no contexto desta dissertação.

Função	Definição
AddGeometryColumn (varchar, varchar, varchar, integer, varchar, integer)	Adiciona uma coluna geográfica a uma tabela já existente.
Dimension (geometry)	Retorna a dimensão de um objeto.
IsEmpty (geometry)	Retorna se determinada geometria está ou não vazia.
Equals (geometry)	Compara dois objetos geográficos.
Disjoint (geometry, geometry)	Retorna se dois objetos geográficos estão disjuntos.
Touches(geometry,geometry)	Retorna se um objeto geográfico toca um outro objeto geográfico.
Crosses(geometry,geometry)	Verifica se um objeto geográfico cruza um outro objeto geográfico.
within(geometry,geometry)	Verifica se um objeto geográfico está dentro de outro objeto geográfico.
Overlaps(geometry,geometry)	Retorna se um objeto geográfico sobrepõe outro objeto geográfico.
Contains(geometry,geometry)	Retorna se um objeto geográfico contém outro objeto geográfico.
intersects(geometry,geometry)	Retorna se um objeto geográfico intercepta outro objeto geográfico.
intersection(geometry,geometry)	Retorna uma figura geométrica que representa o conjunto de pontos de uma figura geométrica em relação a outra figura geométrica.
GeomUnion(geometry,geometry)	Retorna uma figura geométrica que representa o conjunto união de pontos entre dois objetos geográficos. Possui duas variações: GeomUnion(geometry,set) e memGeomUnion(geometry, geometry).
difference(geometry,geometry)	Retorna um conjunto de pontos simetricamente diferentes entre dois objetos geográficos.
X(geometry)	Exibe a coordenada X do primeiro ponto de um objeto geográfico.
Y(geometry)	Exibe a coordenada Y do primeiro ponto de um objeto geográfico.
Z(geometry)	Exibe a coordenada Z do primeiro ponto de um objeto geográfico.
NumPoints(geometry)	Retorna o número de pontos existentes na primeira linha de um objeto geográfico.
PointN(geometry,integer)	Retorna o enésimo ponto da primeira linha de um objeto geográfico.
ExteriorRing(geometry)	Retorna o anel exterior do primeiro polígono de um objeto geográfico.
NumInteriorRings(geometry)	Retorna o número de anéis interiores de um objeto geográfico.
InteriorRingN(geometry,integer)	Retorna o enésimo anel interior do primeiro polígono de um objeto geográfico.
IsClosed(geometry)	Verifica se uma forma geográfica inicia e termina no mesmo ponto.
IsRing(geometry)	Verifica se uma dada curva é fechada ou não.
GeometryN(geometry,int)	Retorna o enésimo objeto geográfico de uma coleção geométrica.
Distance(geometry,geometry)	Retorna a distância cartesiana entre dois objetos geométricos.
AsText(geometry)	Retorna a definição em ‘Well-Known Text’ de um dado objeto geográfico (conforme figura 13).
GeometryFromText(varchar, integer)	Converte um objeto definido em ‘Well-Known Text’ em um objeto geográfico.
Centroid(geometry)	Retorna o ponto centróide de uma forma geográfica.

Tab. 2 - Principais funções do PostGIS provenientes do OpenGIS

Uma outra função do PostGIS que merece destaque é o “shp2pgsql”, que é executado em interface texto no sistema operacional Linux, sendo capaz de, a partir de um arquivo no formato “Arcview Shape File”, gerar um arquivo texto com comandos no formato WKT (“Well-Known Text”). Este arquivo executado no banco de dados, gera uma tabela com os atributos geográficos definidos no arquivo “shape”. Em resumo, ele realiza um mapeamento do arquivo “shape” para o banco de dados geográfico PostgreSQL.

2.10 POSTGEOOLAP

O objetivo deste trabalho é o de construir uma ferramenta de consultas do tipo OLAP sobre um banco de dados espacial. Essas consultas devem ser realizadas através de uma software que possibilite ao usuário buscar as suas informações de maneira fácil e dinâmica. Além dessas características, existe a necessidade que esta ferramenta também seja capaz de realizar consultas geográficas. Conforme será mostrado a seguir, o PostGeoOLAP se enquadra muito bem dentro destas características.

2.10.1 Ferramentas OLAP Espacial

Existem muitas ferramentas OLAP disponíveis, mas a característica geográfica é um fator que dificulta. Existem diversos trabalhos com a proposta de união entre essas duas tecnologias (SIG e OLAP), mas a grande maioria trabalha sobre esses dois conceitos separadamente, e gera um módulo integrador capaz de “filtrar” os resultados entre as duas ferramentas. Como exemplo desses tipos de ferramentas, (FERREIRA, 2002) destacou:

- GOAL (“Geographical Information On-Line Analysis”): trabalha com um método integrador, que une as tecnologias de SIG dos softwares GT Media98 e ArcView com o MSOLAP, para as consultas OLAP (KOUBA; MATOUSEK; MIKSOVSKY, 2000).
- SIGOLAPA: integra o OLAP e o SIG utilizando um modelo de integração dividido em camadas, das quais possui como proposta a “independência da tecnologia”. Para isso trabalha em três camadas que continuam integrando as duas tecnologias que trabalham de modo “independente”. Ou seja, para quem já possui uma aplicação em SIG e uma OLAP, o SIGOLAP irá desenvolver uma camada de integração entre essas aplicações, constituindo três camadas: SIG, OLAP e integração (FIDALGO; TIMES; SOUZA, 2001).
- GOLAPA (Geographical On-Line Analytical Processing Architecture): com a mesma proposta de integração das tecnologias, destaca-se pelo uso de tecnologias abertas, tais como Java, XML e Web Services (FIDALGO; TIMES; SOUZA, 2001).

Essas tecnologias são interessantes para aqueles desenvolvedores que possuem pelo menos uma das duas ferramentas (SIG e OLAP) já desenvolvidas e que pretendam aumentar a potencialidade de suas ferramentas de suporte à decisão.

Para o caso do presente trabalho estes tipos de ferramentas, que trabalham com módulos integradores, gerariam dificuldades na elaboração do modelo conceitual e do projeto, uma vez que haveria necessidade na geração de aplicações distintas para, então, utilizar esses módulos integradores como “front-end” do usuário.

O uso de uma ferramenta, que possibilite o desenvolvimento desde o modelo conceitual até o modelo físico das duas técnicas de SIG e OLAP integradas, apresenta-se como alternativa muito mais encorajadora. Algumas ferramentas com essa característica foram citadas por Colonese (2004):

- GeoMiner: ferramenta capaz de realizar consultas do tipo OLAP sobre um modelo estrela com dados georreferenciados. Seu maior propósito é a realização de Datamining (STEFANOVIC, 1997).
- MapCube: ferramenta capaz de realizar uma consulta num banco de dados multidimensional, realizando consultas sobre dados geográficos e retornando tanto tabelas quanto mapas (Shekhar et al., 2000).
- PostGeoOLAP: ferramenta de consultas OLAP sobre um banco de dados multidimensional capaz de gerar agregações com características geográficas. Opera sobre o banco de dados PostgreSQL para o armazenamento do cubo e, para a manipulação dos mapas utiliza o software gratuito Planet GIS. É uma ferramenta disponibilizada gratuitamente.

Para a confecção do presente trabalho foi selecionada a ferramenta PostGeoOLAP, pelo fato de integrar ferramentas de distribuição gratuita, além de permitir a integração das ferramentas SIG e OLAP por completo, numa única solução.

2.10.2 Características do PostGeoOLAP

O PostGeoOLAP foi desenvolvido num projeto de dissertação de mestrado da Universidade Cândido Mendes, cujo propósito, segundo Colonese (2004) é “a possibilidade de tornar mais simples a modelagem e implementação de um sistema de suporte à decisão que

integre características analíticas de um Datawarehouse e geográficas de um SIG, fazendo com que tais conceitos possam coexistir no mesmo modelo, tanto em nível conceitual quanto em nível da implementação e, mais ainda, que possa ser o mais direto possível tal mapeamento entre os citados níveis de abstração”.

Na verdade, o PostGeoOLAP é uma ferramenta OLAP que permite consultas espaciais sobre dimensões geográficas. Não existe sumarização sobre os dados geográficos, ou seja, não é contemplada a possibilidade de se ter uma tabela fato com dados geográficos. Tem-se apenas dimensões (sejam agregáveis ou não agregáveis) com este tipo de característica.

Uma característica que esta ferramenta contempla é a da hierarquização de uma dimensão. Este conceito se refere ao fato de se poder definir diferentes níveis de consolidação para uma mesma dimensão. Por exemplo, considerando-se a dimensão *tempo*, com os atributos dia, mês, semestre e ano, poder-se-ia definir como nível hierárquico diferente o *ano*, que corresponde a um agrupamento de semestres; um semestre, por sua vez, é um agrupamento de meses; um mês é um agrupamento de dias. Com a ferramenta PostGeoOLAP, seriam definidas as hierarquias por números, dos quais o nível 9 (nove) é o nível sem consolidação (agregação) até o nível 0 (zero) que é o de maior consolidação. Baseado no exemplo em questão, ter-se-á: ano (nível 6), semestre (nível 7), mês (nível 8) e dia (nível 0).

O PostGeoOLAP é uma ferramenta ROLAP capaz de gerar agregações, em forma de tabelas no próprio banco de dados PostgreSQL. Além disto, utiliza os operadores do OpenGIS para a realização das consultas geográficas. Como uma característica da ferramenta ROLAP, o PostGeoOLAP requer muito espaço em disco para a realização das agregações, mas ganha em desempenho nas consultas sobre o cubo.

Uma dificuldade dessa ferramenta é o tratamento dos dados. Deve haver um primeiro momento onde os dados armazenados no SGBD devem ser também armazenados no banco de dados multidimensional (PostgreSQL). Existe também um segundo momento, em que os mapas devem estar preparados, dentro de uma ferramenta denominada “Planet GIS”. Esta ferramenta é gratuita, com um potencial geográfico, mas com pouca ajuda disponível sobre como utilizá-la.

Deve-se destacar a incrível facilidade no uso da ferramenta para o usuário final. As consultas on-line são realizadas de forma tranquila, sendo baseadas em filtros, tanto sobre dados transacionais, quanto sobre dados geográficos, permitindo o drill-down ou o roll-up sobre o cubo.

2.11 O SOFTWARE PLANET GIS

O Planet GIS se propõe ser uma ferramenta que reduz lacuna existente entre as pessoas que necessitam usar uma ferramenta SIG e as ferramentas dispendiosas disponíveis no mercado. Suas vantagens são: gratuidade, facilidade no uso, velocidade no processamento gráfico e habilidade na manipulação de conjuntos extensos de dados (PLANETGIS, 2004).

Suas principais características são:

- Interpretação dos principais formatos de troca (SHP, MIF, DXF, DGN, FEA, SDL, E00).
- Aquisição e edição de todos os tipos de entidades geográficas, inclusive no formato raster.
- Produção de mapas com grande potencial estético.
- Criação de mapas temáticos de acordo com os atributos desejados.
- Edição e visualização de atributos vinculados com um banco de dados.
- Vinculação das entidades geográficas a uma tabela em diversos tipos de banco de dados, possibilitando exibir características completas de cada entidade. Por exemplo, além de exibir uma entidade escola, é capaz de mostrar seu nome, o tipo de escola, a quantidade de alunos, etc. O dado geográfico é armazenado no próprio Planet GIS e as demais informações vêm de um banco de dados transacional.

Percebe-se diante dessas características, que o Planet GIS é uma ferramenta SIG de grande potencial. Contudo, destaca-se como dificuldade de seu uso o fato da ausência de ajuda. Muitas de suas características de operação devem ser descobertas por curiosidade ou entrando em contato com o desenvolvedor.

Apesar de ter o seu formato próprio de armazenamento de arquivo (Planet GIS format – “.map”), permite o trabalho com o formato ArcView Shape File para importação e exportação de dados. Isto vem a viabilizar o trabalho com esta ferramenta em praticamente todos os ambientes, uma vez que a maioria das ferramentas de manipulação de dados interpretam e exportam para esse padrão de arquivo.

3 O MUNICÍPIO DE MACAÉ

O município de Macaé está localizado na região Norte Fluminense, interior do estado do Rio de Janeiro distante 182 quilômetros da capital e teve grande projeção nacional após a instalação da sede da Petrobrás, que trouxe para a região um grande número de empresas prestadoras de serviços na área petrolífera.

Graças aos recursos provenientes principalmente dos “royalties” e da presença da Petrobrás, Macaé hoje é considerado o quarto município em qualidade de vida no Estado, tendo, portanto, grande potencial para investimentos.

Segundo (PREFEITURA MUNICIPAL DE MACAÉ, 2004) “uma pesquisa realizada pelo Instituto de Pesquisa Econômica Aplicada (IPEA), após ter examinado o comportamento da economia em aproximadamente cinco mil municípios brasileiros, apontou Macaé como a sétima cidade que mais cresceu entre os anos de 1970 e 1996, levando-se em conta o Produto Interno Bruto (PIB) “per capita”.

Localizada entre duas importantes capitais de Estado, Rio de Janeiro e Vitória, Macaé conta com boas malhas rodoviária e ferroviária, um aeroporto e um porto - hoje operado pela Petrobrás. Macaé tem como acessos principais a BR-101 e a Rodovia Amaral Peixoto (RJ 106). Estima-se sua população em cerca de 132.461 habitantes de acordo com o Censo Demográfico realizado em 2000 pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

3.1 A PREFEITURA MUNICIPAL DE MACAÉ

A Prefeitura de Macaé atualmente encontra-se organizada administrativamente em Órgãos e Unidades, dentre elas: Gabinete do Prefeito, Gabinete do Vice Prefeito, Gerência Municipal, Procuradoria Geral, Secretarias Municipais, Guarda Municipal de Macaé, Instituto de Previdência dos Servidores do Município de Macaé, Fundo Municipal de Saúde, Fundação Educacional de Macaé, Fundação Macaé de Cultura, Empresa Pública Municipal de Turismo, Instituto Macaé de Metrologia e Tecnologia, Fundação de Ação Social, Fundo Municipal de Defesa dos Direitos da Criança e Adolescente, Fundo Municipal dos Direitos da Pessoa Portadora de Deficiência, Fundo Municipal do Idoso e, Fundo Municipal de Assistência Social.

Cada Órgão ou Unidade Administrativa possui a liberdade de desenvolver ou contratar empresas para o desenvolvimento de sistemas para o controle e gerenciamento de suas atividades cotidianas.

Além das informações já existentes em seus órgãos ou Unidades Administrativas, a Prefeitura Municipal, com o propósito de levantar dados específicos, avaliar os serviços prestados pela Administração Municipal e direcionar os futuros investimentos do governo, desenvolveu um programa denominado “Macaé Cidadão”. Além desse programa, a Prefeitura Municipal está desenvolvendo o Projeto Geo-Macaé, que tem como propósito adequar o Município de Macaé às novas tecnologias no trato da Informação, e georreferenciar as informações do município.

3.2 O PROGRAMA MACAÉ CIDADÃO

O Programa Macaé Cidadão consiste numa pesquisa de campo realizada pela própria prefeitura, cujo propósito é o levantamento de informações sobre a população, com o objetivo de enriquecer as informações vindas do IBGE.

Este programa foi motivado pela necessidade da Prefeitura de Macaé obter de sua população informações sobre diversos temas, tais como: educação, saúde, infra-estrutura, trabalho e renda, esporte e lazer, indicadores sociais, indicadores urbanos, território, divisão administrativa, dados domiciliares e dados demográficos. A consolidação dos dados desses

diversos temas serve como base para elaboração de tabelas capazes de identificar as carências e necessidades da população.

Os dados foram obtidos por bairro, região administrativa, distritos e municípios, tais que possibilitassem a tomada de decisão pelo poder municipal que, ao saber dos problemas individuais de cada morador pode, de modo seguro, adotar procedimentos para corrigir e melhorar as condições de vida da população macaense como um todo.

3.3 O PROJETO GEO-MACAÉ

A proposta do projeto Geo-Macaé é a de adequar o município de Macaé às novas tecnologias no trato da informação. Trata-se de um Sistema de Informações Cartográficas (SIC), que tem como proposta georreferenciar as informações do município, de modo que se obtenha uma base cartográfica digital com a estrutura territorial da região, a partir da base de setores censitários do IBGE e do programa Macaé-cidadão.

Para a realização deste projeto, a prefeitura contratou uma empresa de consultoria em geoprocessamento. Este projeto tem como objetivos: estruturação e treinamento técnicos para dominar as técnicas de Geoprocessamento, elaboração de uma base cartográfica digital com a estrutura territorial do Município a partir da base de Setores Censitários do IBGE e geração de Dados sobre domicílios.

Com o projeto Geo-Macaé ficam disponibilizados um conjunto de Mapas, com a Estrutura Territorial e Administrativa do Município, além de Mapas Temáticos.

Entretanto, ao que se tenha investigado nesta pesquisa, o projeto Geo-Macaé ainda não oferece aos usuários a possibilidade de realizar operações de consulta e análise de dados ad-hoc. Este fato deixa uma lacuna, a ser ocupada pela presente dissertação.

4 PROJETO DO SISTEMA DE SUPORTE A DECISÃO DE MACAÉ

A proposta deste trabalho é o desenvolvimento de um Sistema de Suporte à Decisão (SSD), usando as tecnologias de Datawarehouse, OLAP e SIG como ferramenta de auxílio à tomada de decisão.

A tecnologia do Datawarehouse é extremamente pertinente ao caso do município, uma vez que este possui diversas Unidades Administrativas e Órgãos que se encontram geograficamente distantes, além de terem sido implantados com recursos tecnologicamente diferentes, o que dificulta, em termos computacionais, a obtenção de informações genéricas do município.

Com um trabalho de pré-processamento de dados e criação de um Datawarehouse ter-se-á uma ferramenta de suporte na qual será possível a extração de informações de todas Unidades e Órgãos, além dos dados da Pesquisa Macaé Cidadão e Projeto Geo-Macaé.

Com o OLAP e o SIG será possível a realização de consultas sobre o Datawarehouse, além do que as duas tecnologias separadamente já são capazes de realizar. Por exemplo, uma consulta que retorne crianças e as escolas que mais se aproximam de suas casas.

Em entrevista realizada junto a Prefeitura Municipal levantou-se, como um dos objetivos do SSD, a capacidade não só de mostrar informações de cada Órgão ou Unidade Administrativa, mas que mostre informações do município como um todo.

Seguem-se alguns exemplos de consultas apontadas pela prefeitura como de grande interesse prático:

- Quantas pessoas que possuem mais de 45 anos e estão desempregadas?

- Quantas pessoas sem estudo estão com problema crônico de saúde, tem mais de 50 anos e estão desempregadas?
- Quantas pessoas que recebem educação especial e moram a mais de 1000 metros de alguma Unidade de Ensino Público?

É importante lembrar que essas consultas são apenas alguns exemplos citados pela Prefeitura. O objetivo do SSD é ir além dessas perguntas, uma vez que as consultas serão realizadas de acordo com as necessidades do usuário em determinado instante (ad-hoc). Essas questões são exemplos de informações das quais a Prefeitura necessita, mas cuja única forma de obtê-las é através da leitura de vários relatórios, ou mesmo através de uma pesquisa de campo.

A partir dessas necessidades da prefeitura, uma primeira questão que surgiu foi qual seria a menor granularidade, em termos geográficos, sobre a qual as consultas poderiam ser feitas. De antemão, sabe-se da disponibilidade das seguintes regiões ou unidades: setores censitários, bairros e setores administrativos. Essas unidades serão detalhadas posteriormente. A escolha foi o setor censitário, que segundo Dias, Oliveira e Câmara (2004) “no Brasil, corresponde à capacidade de levantamento do recenseador, variando por região em torno de 200 a 400 domicílios”. Qualquer região menor que essa, segundo a Prefeitura, não ajudaria a título de gerência do Município para a realização dos seus objetivos, pois aí chegar-se-ia praticamente a consulta do indivíduo, o que não mostraria uma tendência ou característica, e sim um caso isolado.

Quanto às áreas de atuação ou interesse imediato foram selecionadas para o desenvolvimento deste trabalho as seguintes: Saúde e Educação, por serem julgadas como as áreas de maior interesse para a Prefeitura Municipal. As informações sobre saúde provêm do Projeto Macaé Cidadão (dados da população) e da Secretaria de Saúde (Postos de Saúde). As informações sobre a Educação são oriundas do Projeto Macaé Cidadão (dados da população) e da própria Secretaria Municipal de Educação (Escolas). Além dessas informações, deverão ser usadas as informações dos cadastros mobiliário e imobiliário que também serão retirados do Projeto Macaé Cidadão. Todas essas informações serão unidas através do Projeto Geo-Macaé, que tem várias dessas informações georreferenciadas. Desta forma, o SSD estará atendendo ao propósito de trazer à tona as necessidades da população e os resultados já obtidos pelos programas implantados.

4.1 A MODELAGEM DO SSD DO MUNICÍPIO

4.1.1 A Modelagem do Espaço Geográfico

Conforme dito anteriormente, a menor unidade de análise a ser usada no modelo é o setor censitário. Segundo informação obtida no Projeto Geo-Macaé, o município de Macaé dispõe de 195 Setores Censitários dispostos conforme a figura 16.

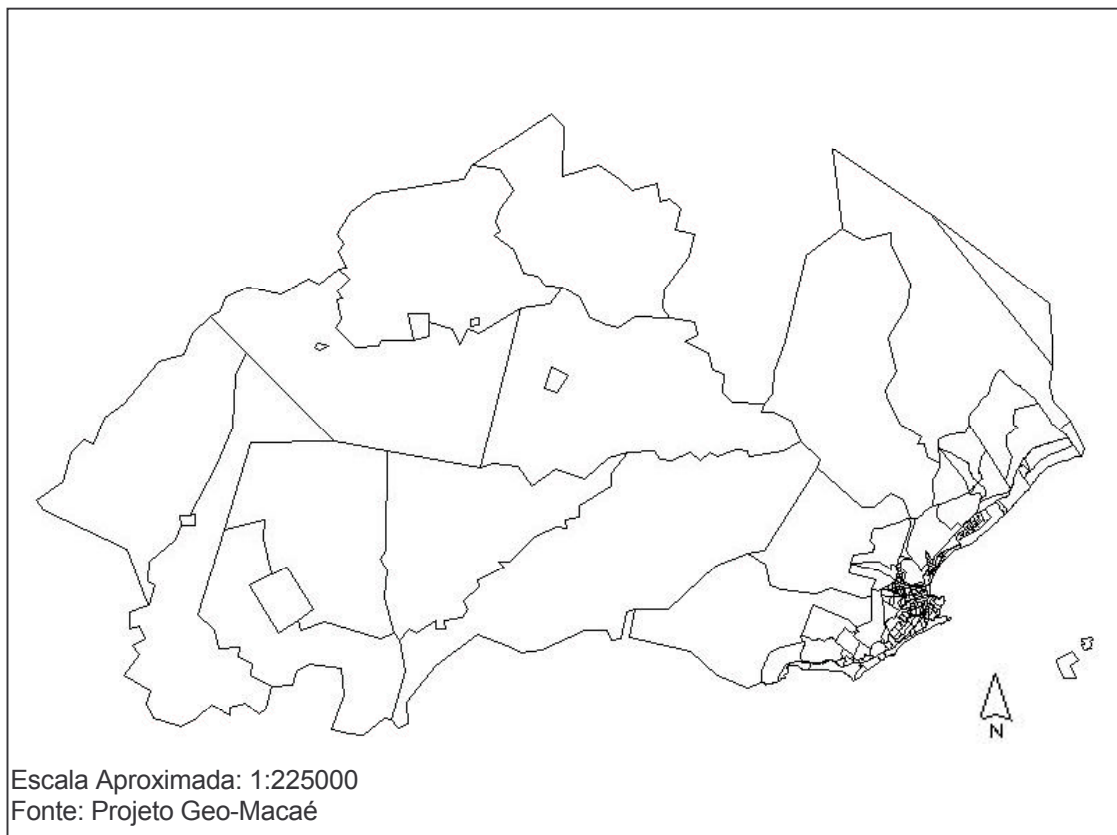


Fig.16 Visualização dos Setores Censitários do Município de Macaé

Pode-se considerar o Bairro como um conjunto de Setores Censitários, que também possuiria informações sobre tendências tão importantes quanto o Setor Censitário. Com esta camada geográfica é possível obter informações importantes a cerca da população macaense. Pelos dados informados pela prefeitura, o município de Macaé possui hoje 27 bairros, conforme pode ser visto na figura 17.

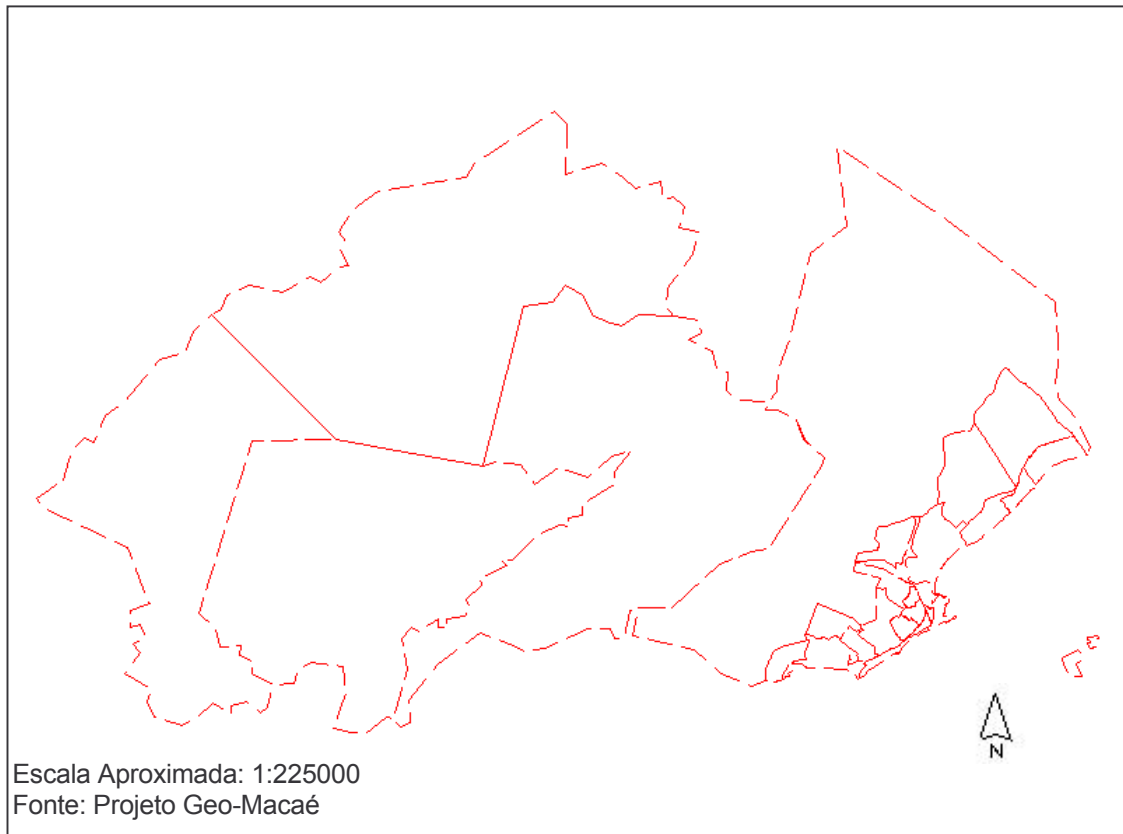


Fig.17 Visualização dos Bairros do Município de Macaé

Uma outra camada geográfica importante administrativamente para o município de Macaé denomina-se “Setor Administrativo”. Esta foi criada para facilitar a administração do município. Trata-se dos setores gerenciais do município. Conforme pode ser visto na figura 18, o município de Macaé possui 10 setores administrativos, nomeados por cores.

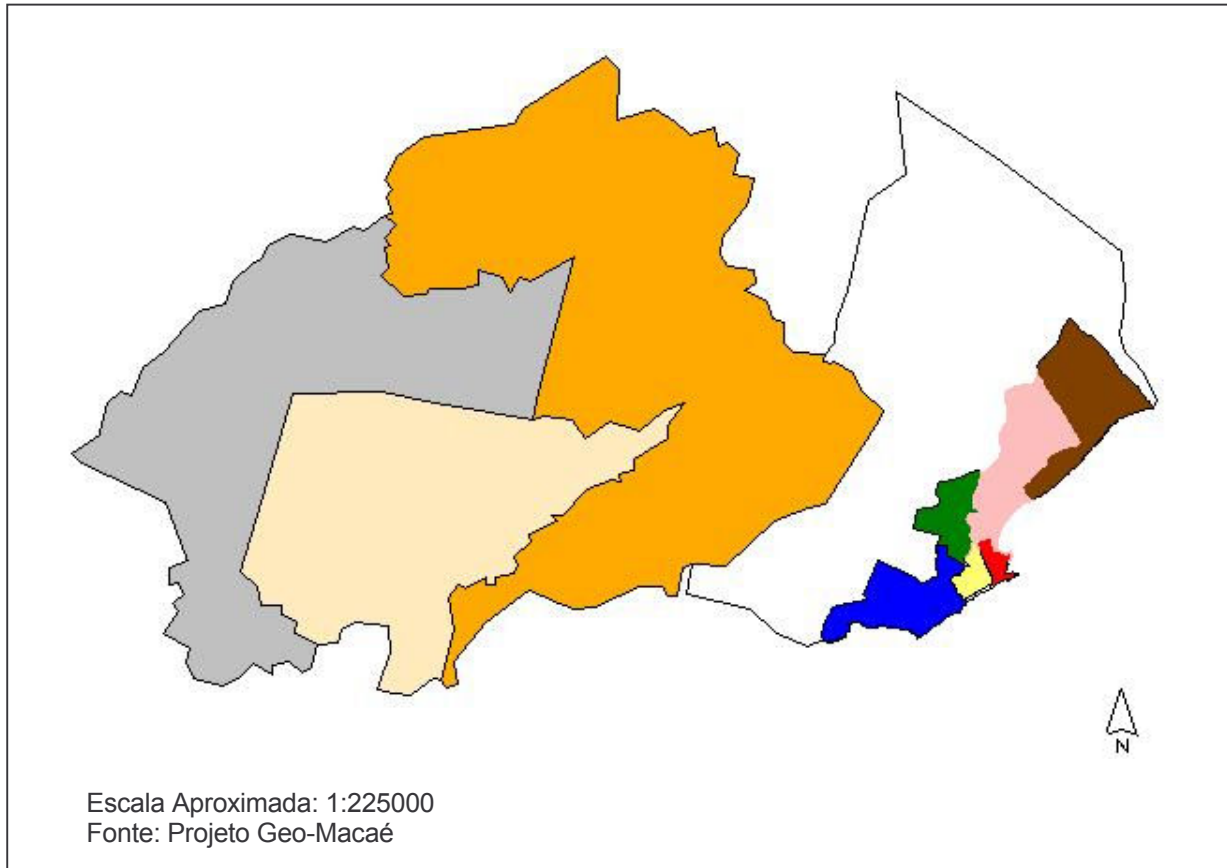


Fig.18 Visualização dos Setores Administrativos do Município de Macaé

Numa visão mais geral, é possível visualizar a relação “bem comportada” entre essas três características geográficas do município, o que levou a ser denominado de hierarquização do espaço geográfico do município de Macaé. Esse fato está representado graficamente na figura 19, mostrando que a menor granularidade é o setor censitário. Um conjunto de setores censitários corresponde ao bairro. Seguindo-se o mesmo raciocínio, um conjunto de bairros corresponde ao setor administrativo.

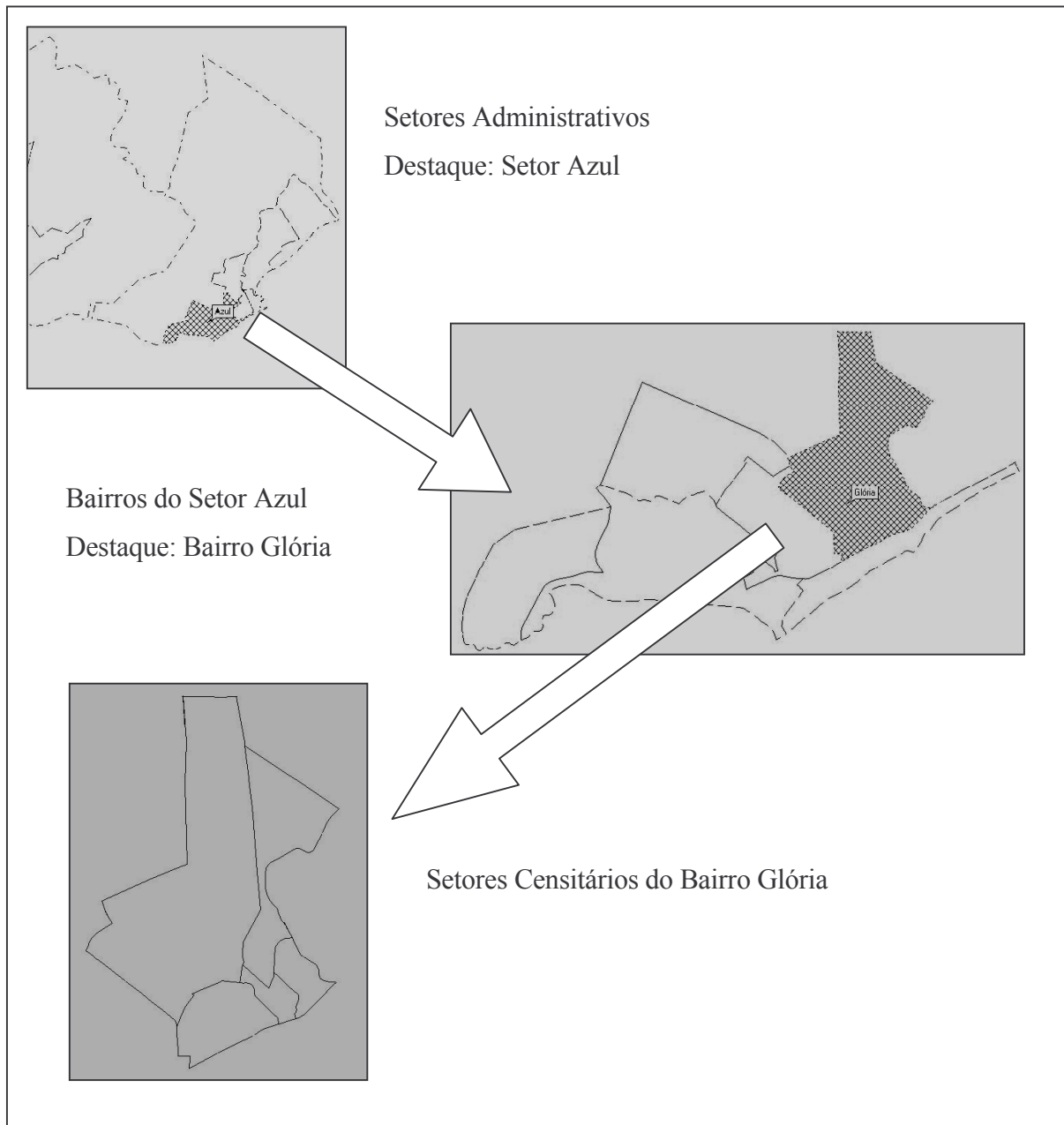


Fig.19 Representação do espaço geográfico do município de Macaé

Usando a UML, as três camadas geográficas seriam representadas conforme a figura 20, onde o Bairro é composto de vários setores censitários, e o setor administrativo composto de vários bairros.

Conforme o tratamento usado no Projeto Geo-Macaé, considera-se todas as três camadas como figuras geométricas do tipo polígono. Usando os estereótipos para a representação das classes geográficas, obter-se-á um diagrama, conforme a figura 20.

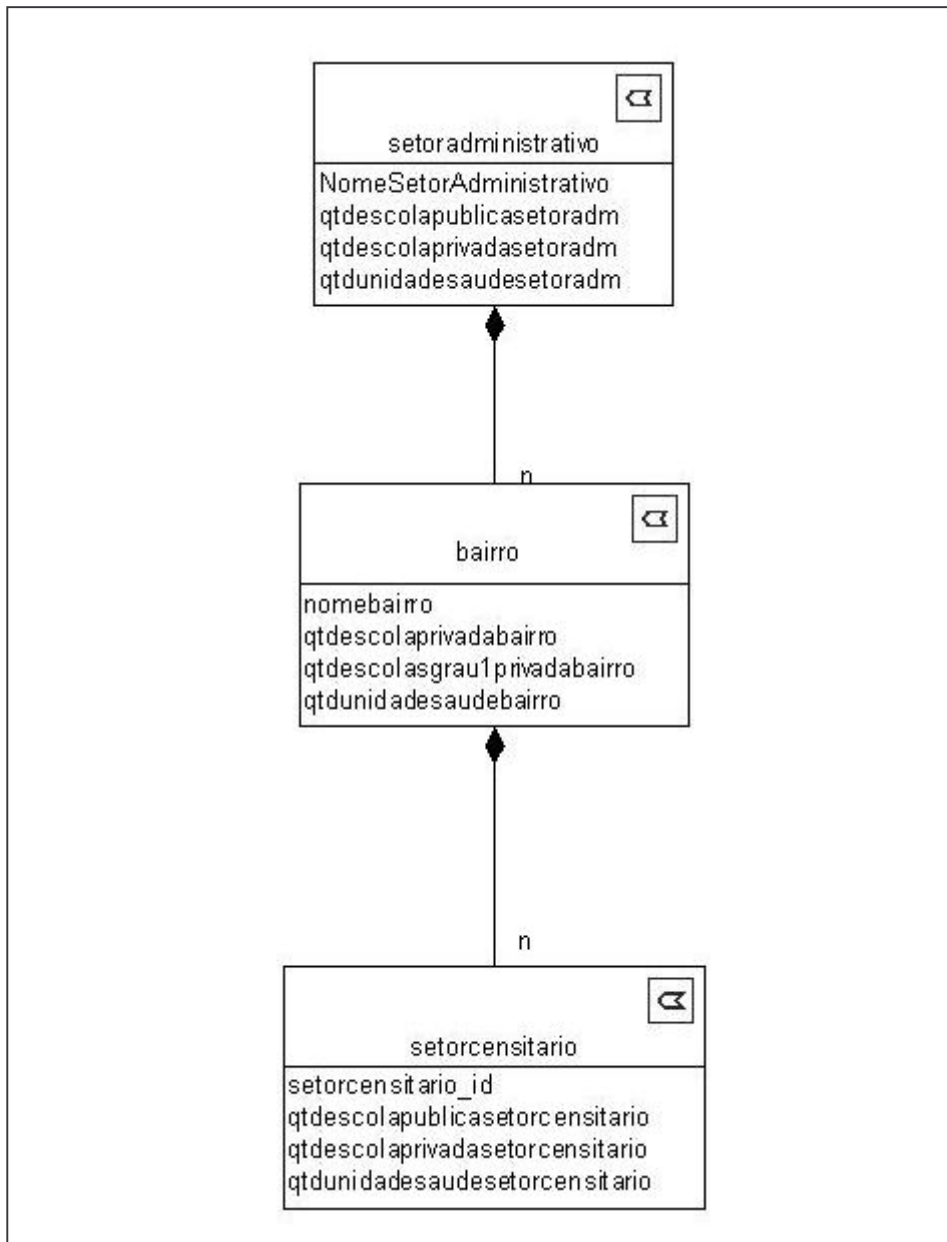


Fig.20 Representando as classes geográficas com o auxílio dos estereótipos

Além dos estereótipos, a figura 20 exibe os atributos de cada camada geográfica. Por exemplo, a camada Setor Censitário possui o atributo `setorcensitario_id`, que representa o próprio código gerado pelo IBGE; no Bairro, o atributo `nomebairro` e no Setor Administrativo o atributo `NomeSetorAdministrativo`. Além desses atributos, foram colocadas em cada classe as quantidades de escolas públicas e privadas, além das Unidades de Saúde. A repetição dos mesmos atributos em todas as classes geográficas se dá pela facilidade de consulta por todos os níveis da hierarquia. Embora com certa tendência, julga-se oportuno lembrar que não está sendo

modelado um banco de dados transacional e sim um Datawarehouse, onde a redundância é permitida, desde que facilite o processamento de consultas.

Para se evitar complexidades no modelo do Datawarehouse e evitar degenerações nas dimensões, que poderiam tornar o modelo pouco eficiente em termos de consulta, foi adotada a união de todas as classes em uma única classe como recurso para a construção do modelo físico da dimensão geográfica. Apesar desse recurso gerar repetição de uma mesma informação várias vezes nesta classe, deve-se novamente lembrar que o propósito não é a facilidade de inclusão e exclusão como num modelo para sistemas transacionais, e sim a eficiência das consultas.

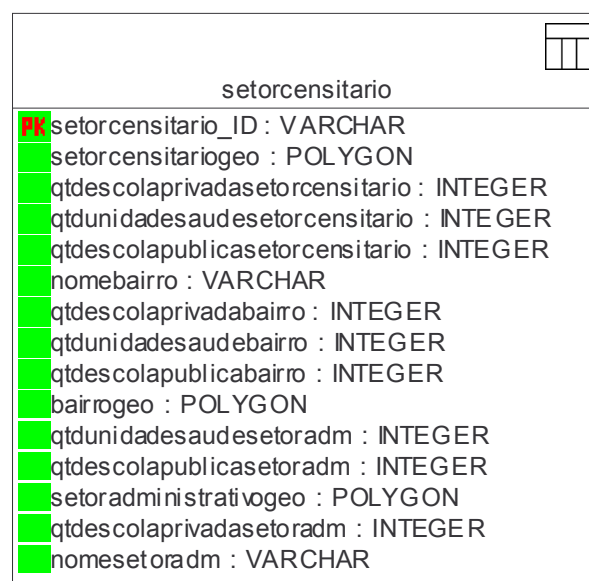


Fig.21 Modelo Físico das Classes Geográficas

É possível observar na figura 21 que, fisicamente, as três dimensões geográficas foram incorporadas a uma única classe até mesmo pelo comportamento das mesmas: Setor Administrativo é composto de Bairros; um bairro é composto por Setores Censitários. Se for usado no modelo físico as três classes em separado, a grande quantidade de operações do tipo “JOIN” entre elas seria grande, sempre que se desejasse fazer um drill-down ou um drill-up. Para agilizar as consultas, as operações de “JOIN” entre as tabelas serão realizadas no processo de armazenamento de dados, guardando todas as informações em uma única tabela deixando para o momento das consultas somente as operações de “slice” e “dice”, conforme foi ilustrado na figura 14, página 42.

Outro ponto que merece destaque é o surgimento de três atributos: setorcensitariogeo, bairrogeo e setoradministrativogeo. Estes três atributos correspondem ao armazenamento em banco de dados dos polígonos referentes a cada uma das tuplas da dimensão **setorcensitario**.

4.1.2 As classes geográficas não agregáveis

Consideram-se classes geográficas não agregáveis as classes geográficas que não têm ligação direta com o modelo dimensional do Datawarehouse. São classes que não terão ligação com a tabela fato. Essas classes, na verdade, estarão associadas a uma dimensão que seja geográfica, uma vez que a “chave” de ligação entre a dimensão geográfica e a classe geográfica não agregável é o posicionamento geográfico.

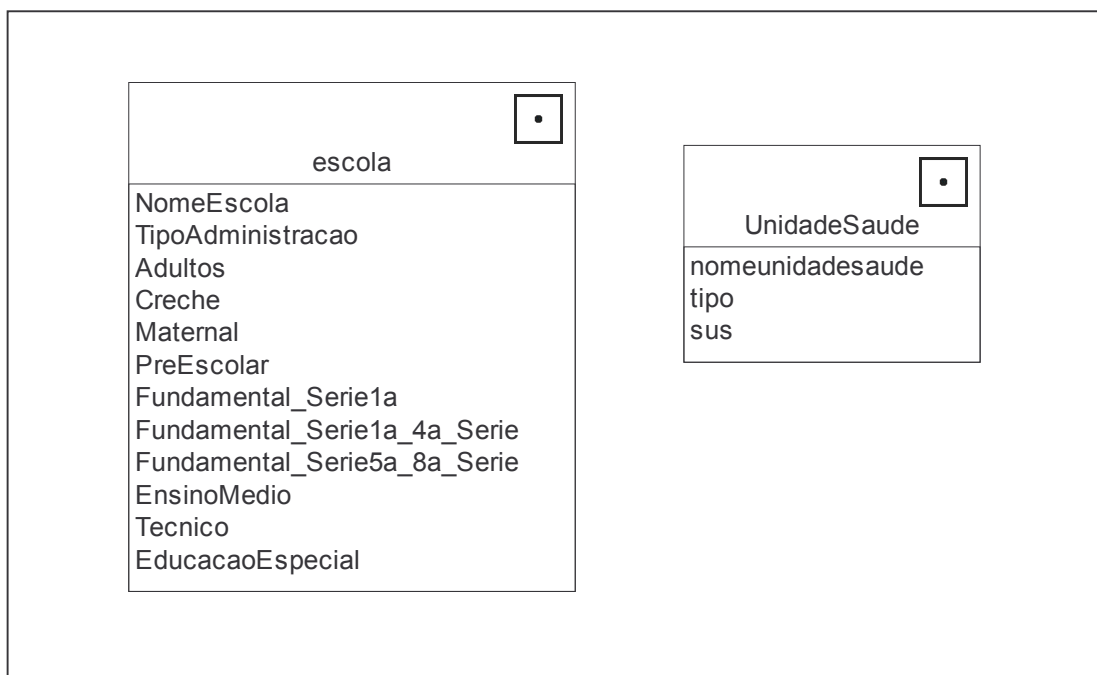


Fig.22 Classes Geográficas não agregáveis

Usando o exemplo da escola a existência dessa classe é factível ao modelo, uma vez que pode-se desejar saber quais pessoas menores de 18 anos não estudam e não possuem escolas num raio de 1Km de distância de sua casa. É necessário observar nesta pergunta que as características “idade” e “não estuda” pertencem à pessoa, da mesma forma o local onde mora (“sua casa”) o indivíduo. Já a escola não está relacionada à pessoa e sim ao local onde ela mora. Aí está o motivo para que a classe escola não seja uma dimensão do modelo dimensional, e sim

uma dimensão geográfica que está associada com a classe **setorcensitario** através dos atributos geográficos. No exemplo, foi usado o termo distância para representar a sua ligação com a classe **setorcensitario**.

No modelo usado para o Datawarehouse implementado para a Prefeitura de Macaé, foram usadas as seguintes classes geográficas não agregáveis: **escola** e **UnidadeSaúde**. Observando na figura 20 os estereótipos das classes não agregáveis, as Escolas e as Unidades de Saúde estão sendo consideradas do tipo ponto. Essas duas classes estão associadas ao modelo através da dimensão geográfica denominada **setorcensitario**.

4.1.3 As dimensões do Datawarehouse

Além da dimensão geográfica **setorcensitario** outras dimensões de análise foram criadas para possibilitar as consultas multidimensionais do município de Macaé. Dentro dos pontos de análise solicitados pelo gestor municipal, conforme visto no início deste capítulo.

A dimensão **condicaomoradia** (figura 23) serve para exibir as condições nas quais o cidadão está morando no município. Seus atributos são descritos na tabela 3.

Os atributos constantes na dimensão denominada **característica** são apresentados na figura 23 e deixam de ser aqui explicitados por serem auto-explicativos.

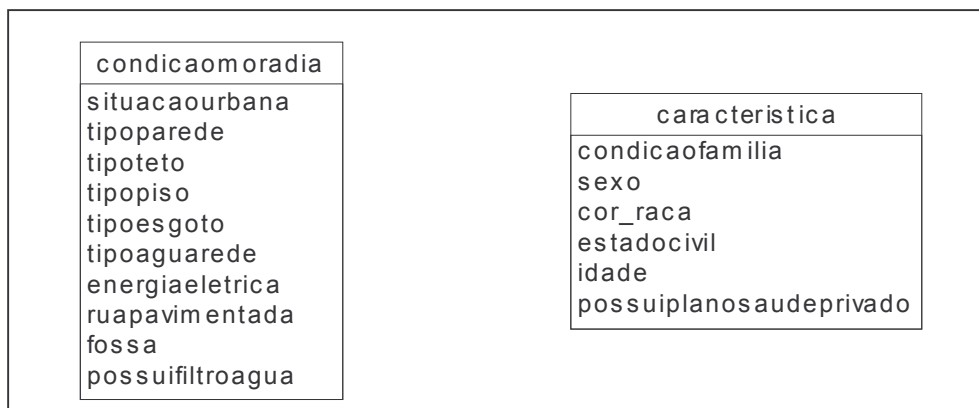


Fig.23 Dimensões **condicaomoradia** e **característica**

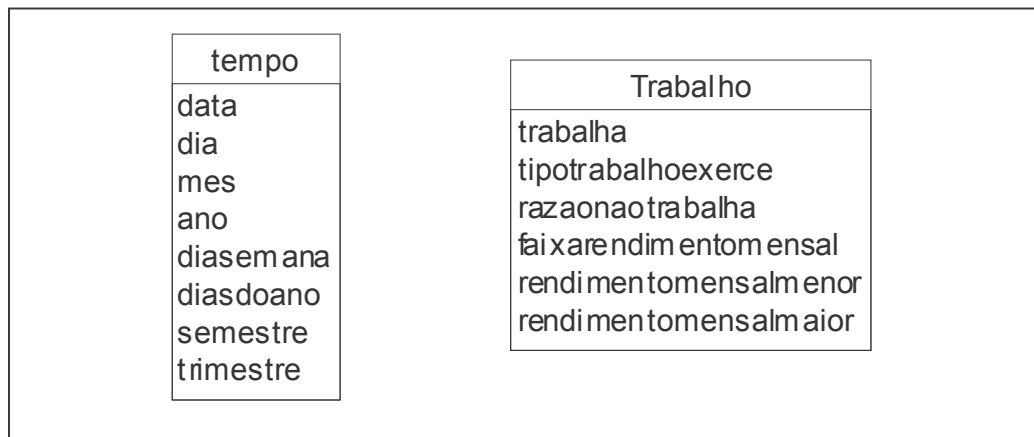
Nome	Semântica	Descrição
situacaourbana	Situação Urbana	Indica se o cidadão mora num setor urbano ou num setor rural
tipoparede	Tipo de Parede	Indica qual o tipo de parede que a moradia do cidadão possui, tal como: Alvenaria, Madeira Aparelhada, Madeira Aproveitada, Palha, etc
tipoteto	Tipo de Teto	Indica qual o tipo de teto que a moradia do cidadão possui, tal como: Laje de Concreto, Madeira Aproveitada, Telha, Zinco, etc.
tipopiso	Tipo de Piso	Indica qual o tipo de piso que a moradia do cidadão possui, tal como: Carpete, Cimento, Piso Frio, Terra, etc.
tipoesgoto	Tipo de Esgoto	Indica qual o tipo de tratamento de esgoto a pessoa possui em sua casa, tal como: Rede Geral de Esgoto, Sem Esgotamento Sanitário, etc.
tipoaguarede	Tipo de Água da Rede	Especifica o tipo de água que a pessoa recebe em sua casa, tais como: Água de Rede Geral, Água de Poço, Água de Nascente, etc.
energiaeletrica	Energia Elétrica	Indica se a pessoa possui ou não energia elétrica em sua casa.
ruapavimentada	Rua Pavimentada	Indica se a rua na qual a casa do cidadão se encontra é ou não pavimentada.
fossa	Fossa	Indica se a moradia do cidadão possui ou não fossa.
possuifiltroagua	Possui Filtro Água	Indica se a pessoa usa filtro para beber casa em casa ou não.

Tab. 3 - Descrição dos atributos da dimensão **condicaomoradia**

A dimensão **tempo**, figura 24, pertinente à praticamente todos os Datawarehouse, existe nesse modelo com o propósito de armazenar o tempo em que as informações foram carregadas na base de dados. Com essa dimensão será possível avaliar as mudanças no município ao longo do tempo. Nesta dimensão existe basicamente um único valor: a data. Todos os demais atributos são “explosões” deste atributo, como: dia, mês, ano, dia semana (o dia da semana em que este dado foi carregado para o sistema), dias do ano (há quantos dias do ano a informação foi importada), Semestre e Trimestre.

A dimensão **trabalho** (figura 24) exhibe as informações do cidadão no que se refere a sua característica trabalhista, conforme descrito na tabela 4.

Nome	Semântica	Descrição
trabalha	Trabalha	Indica se a pessoa trabalha ou não
razaonaotrabalha	Razão não Trabalha	Caso o cidadão não trabalhe, indica o motivo pelo qual ele não trabalha, tal como: Afazeres Domésticos, Aposentado, Doente, Desempregado por falta de oportunidade, etc.
tipotrabalhoexerce	Tipo Trabalho Exerce	Caso a pessoa trabalhe, indica qual o seu ambiente de trabalho, tal como: Autônomo, Empregado, Empregador, etc.
faixarendimentomensal	Faixa Rendimento Mensal	Indica de forma textual a faixa de rendimento da pessoa, sempre representando um intervalo em salários, tal como: Entre 1 e 2 Salários Mínimos, Acima de 20 Salários Mínimos, Até 1 Salário Mínimo, etc.
rendimentomensalmenor	Rendimento Mensal Menor	É uma “explosão” de Faixa de Rendimento Mensal que terá o menor valor da faixa indicada em Faixa de Rendimento Mensal.
rendimentomensalmaior	Rendimento Mensal Maior	Também é uma “explosão” de Faixa de Rendimento Mensal que indicará o maior valor da faixa indicada em Faixa de Rendimento Mensal.

Tab. 4 - Descrição dos atributos da dimensão **trabalho**Fig.24 Dimensões **tempo** e **trabalho**

A dimensão **educacao**, figura 25, abrange características do ponto de vista de estudos do cidadão, conforme pode ser visto na tabela 5.

Nome	Semântica	Descrição
sabeler	Sabe Ler	Indica simplesmente se o cidadão sabe ou não ler.
frequentaescola	Freqüenta Escola	Indica simplesmente se a pessoa freqüenta ou não a escola.
justificanaofrequentaescola	Justifica não Freqüenta Escola	Caso a pessoa não freqüente a escola, neste atributo será relatada a razão pela qual o problema se dá, tal como: trabalha, Por falta de vaga, não existe escola perto de casa, etc.
cursoquefrequenta	Curso que Freqüenta	Indica qual o curso que a pessoa que está estudando freqüenta, tal como: Primeiro Grau do Ensino Fundamental, Segundo Grau do Ensino Médio, etc.
tipoescola	Tipo Escola	Indica qual o tipo de escola que a pessoa que estuda está freqüentando, como: Particular, Público Estadual, Público Municipal, etc.

Tab. 5 - Descrição dos atributos da dimensão **educacao**

A dimensão **queixasistemaude**, figura 25, exhibe quais as principais queixas que um cidadão possui em relação à área de saúde do município. Todos os seus atributos são meras respostas Sim ou Não, tais como: Possui Queixa, Dificuldade Atendimento, Atendimento Insatisfatório, Carência de Recursos Materiais, Carência de Recursos Humanos, Carência de Especialistas, Falta de Unidade Próxima de sua Casa, Falta de Informação ao Usuário.

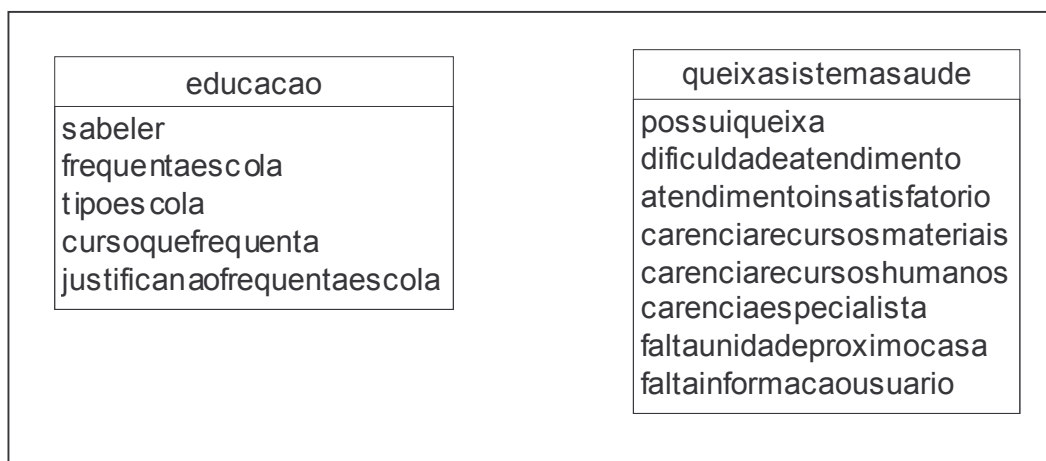


Fig.25 Dimensões **educacao** e **queixasistemaude**

Na dimensão **problemasaude**, figura 26, estão representadas as informações referentes aos problemas de saúde dos munícipes. Os atributos dessa dimensão são explicados na tabela 6.

Nome	Semântica	Descrição
possuiproblemacronicosaude	Possui Problema Crônico de Saúde	Informa se a pessoa possui ou não problema de saúde
tipoproblemasaude	Tipo de Problema de Saúde	Caso a pessoa possua problema de saúde, qual problema ela possui, tal como: Alergia, Coração, Câncer, Diabetes, etc.
fazacompanhamentomedico	Faz Acompanhamento Médico	Indica se a pessoa que possui problema de saúde faz algum tipo de acompanhamento médico.
localacompanhamentomedico	Local do Acompanhamento Médico	Para aqueles que fazem acompanhamento médico, qual o local que este acompanhamento é feito, tal como: Clínica Particular conveniada, Hospital Público, Posto de Saúde, etc.

Tab. 6 - Descrição dos atributos da dimensão **problemasaude**

Na dimensão **deficienciafisica**, figura 26, estão representados os dados referentes as deficiências físicas mais comuns entre os munícipes. Os atributos dessa dimensão são descritos na tabela 7.

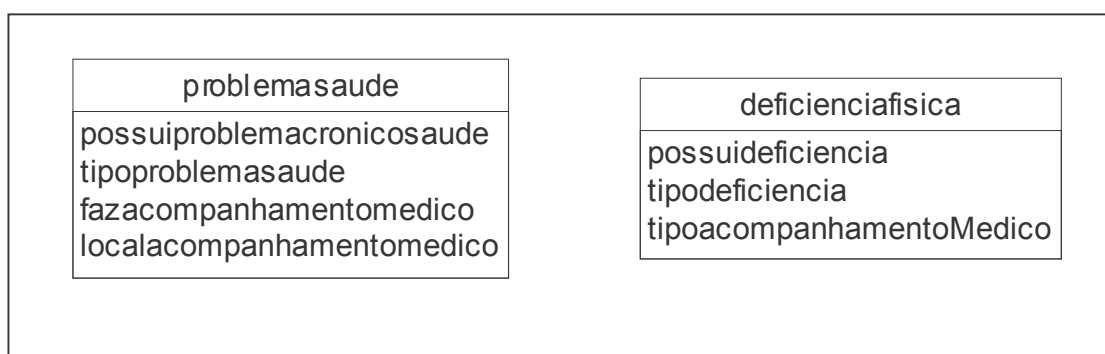


Fig.26 Dimensões **problemasaude** e **deficienciafisica**

Nome	Semântica	Descrição
Possuideficiencia	Possui Deficiência	Informa se a pessoa possui ou não alguma deficiência física
Tipodeficiencia	Tipo Deficiência	Informa qual o tipo de deficiência que a pessoa possui.
tipoacompanhamentomedico	Tipo Acompanhamento Médico	Informa onde a pessoa faz acompanhamento médico para a sua deficiência, tais como: Sem Local, Acompanhamento Público Permanente, Acompanhamento Particular Permanente, etc.

Tab. 7 - Descrição dos atributos da dimensão **deficienciafisica**

4.1.4 O modelo do Datawarehouse do município de Macaé

Após terem sido definidas todas as dimensões do modelo do Datawarehouse, sejam estas agregáveis ou não agregáveis, o modelo lógico multidimensional é mostrado na figura 27.

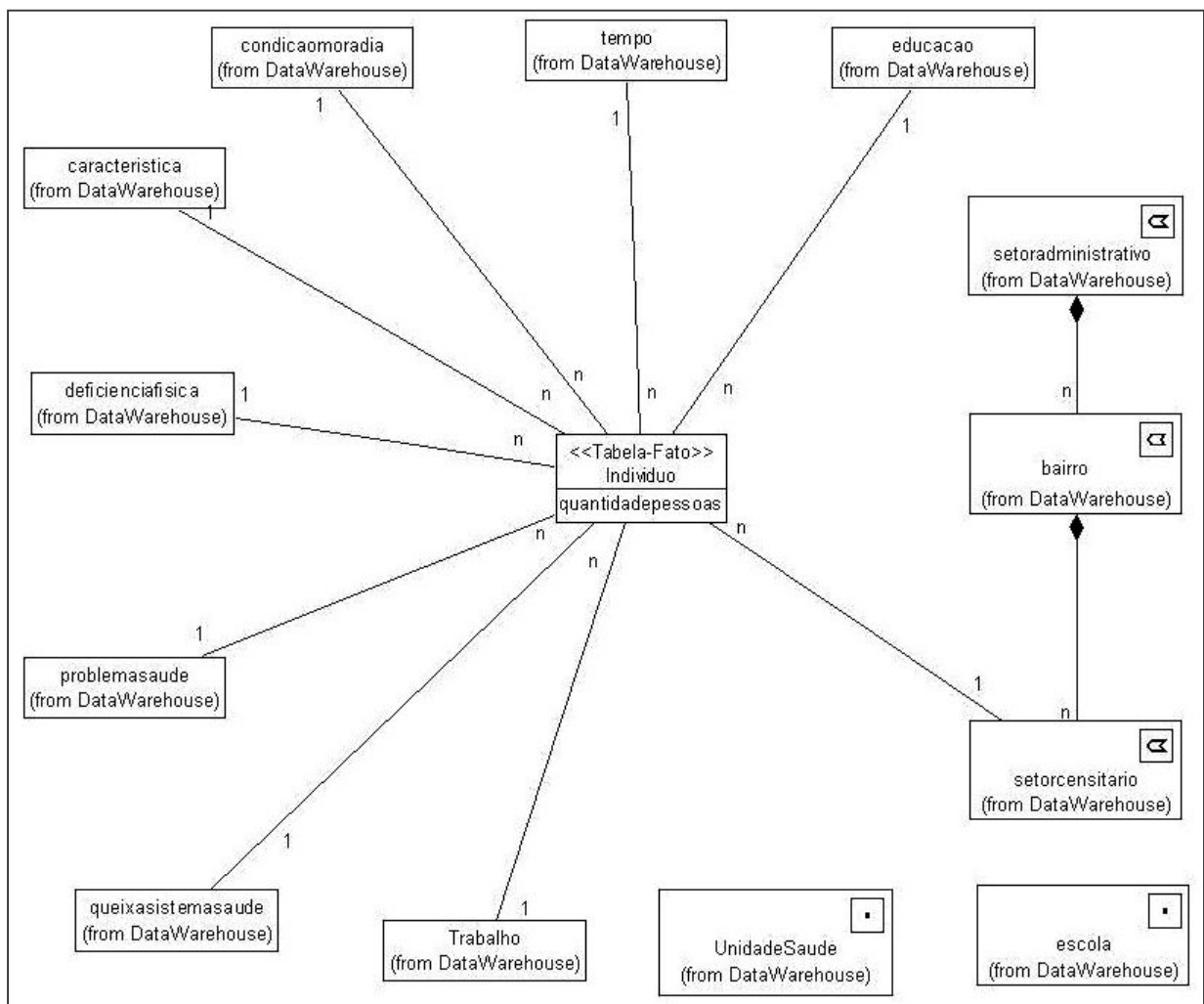


Fig.27 O Modelo Lógico Dimensional

No modelo estão evidentes todas as dimensões tratadas anteriormente, e aparece a tabela fato denominada **Individuo**. É importante salientar que a tabela fato só possui um atributo – quantidadepessoas. A princípio, a tabela fato havia sido projetada sem nenhum atributo, imaginando a impossibilidade de existência de duas ou mais pessoas com todas as mesmas características (condições de moradia, características, deficiência física, problema saúde,

educação, trabalho, queixa sistema saúde). No entanto, numa primeira análise dos dados, foi possível verificar a existência dessa possibilidade. Assim, na hipótese da ocorrência de mais de um indivíduo com as mesmas características, será incluído um contador. Na maioria dos casos a quantidade de ocorrências será de apenas 1, mas também existem casos com 2, 3 e até mesmo 4.

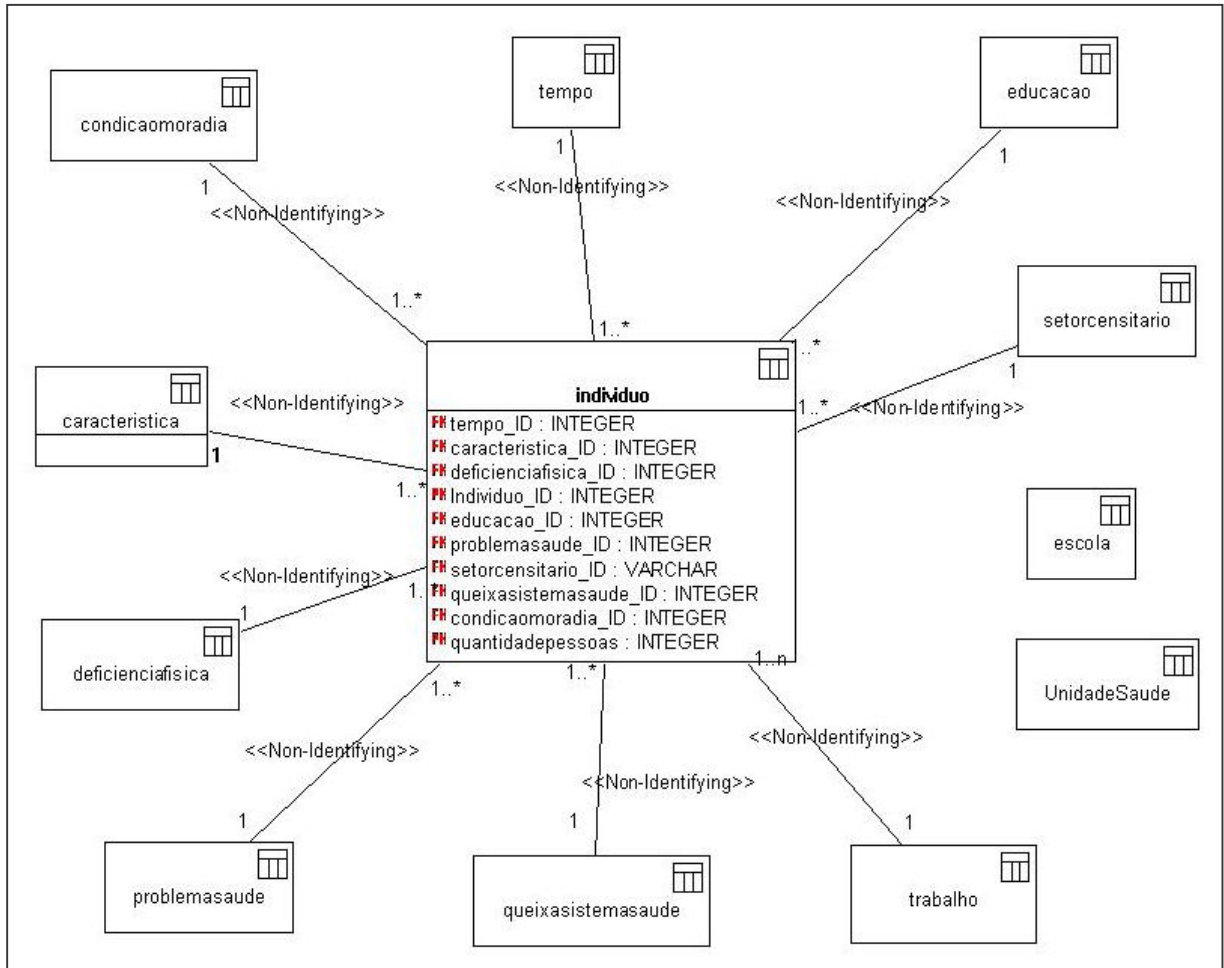


Fig.28 Modelo Multidimensional Físico

A figura 28 exibe o modelo multidimensional físico, onde se deve dar destaque a tabela fato **indivíduo** que tem como sua chave primária as chaves de todas as dimensões agregáveis do modelo e a quantidade de pessoas.

De posse da modelagem do Datawarehouse, o próximo passo é a geração do banco de dados e a migração dos dados das diversas bases de dados da Prefeitura de Macaé. Estas etapas estão mostradas no próximo capítulo desta dissertação.

5 A IMPLEMENTAÇÃO DO MODELO

5.1 O AMBIENTE DE DESENVOLVIMENTO DO SSD

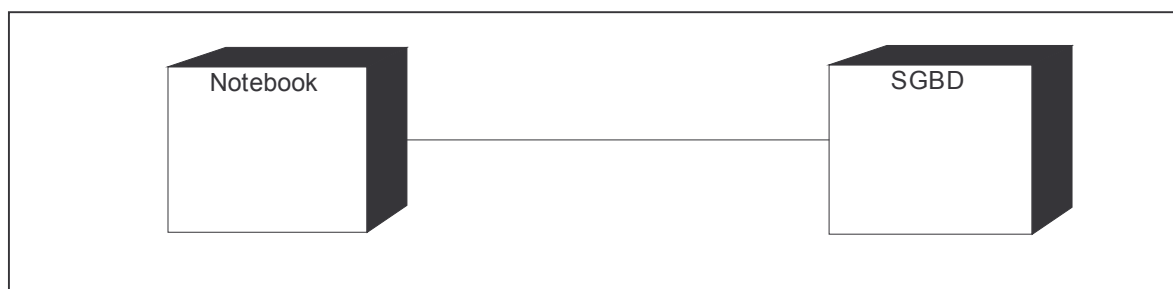


Fig.29 Diagrama de Distribuição

O diagrama de distribuição da figura 29 está apresentando a estrutura de hardware usada no desenvolvimento desse estudo de caso. Um Notebook Celeron 1.2 GHz, 512MB de memória RAM, usando como sistema operacional o Microsoft Windows XP e as aplicações PG Admin III, PostGeoOlap e PlanetGis.

O Servidor de Gerenciador de Banco de Dados é um Duron 1 GHz, com 640MB de memória RAM, e um Disco Rígido de 40 GB, usando como sistema operacional Conectiva Linux 9.0 e o SGBD PostgreSQL 7.4, e o PostGIS 0.8.1. A comunicação entre esses computadores se dá por meio de uma rede TCP/IP com velocidade de 100Mbps.

O computador denominado SGBD tem como único propósito o gerenciamento da base de dados do Datawarehouse. Todo tipo de processamento é realizado pelo Notebook.

5.2 A GERAÇÃO DO BANCO DE DADOS MULTIDIMENSIONAL GEOGRÁFICO

Alguns passos foram seguidos para a geração do banco de dados multidimensional, uma vez que não foi encontrada uma ferramenta CASE para a realização desta dissertação que gere automaticamente uma base de dados relacional e geográfica para o banco de dados PostgreSQL.

A base de dados relacional foi gerada normalmente, através de um script com comandos SQL (DDL), de forma a gerar as tabelas e índices num banco de dados denominado *dwprefeitura*. É importante lembrar que, para agilizar o processo de consultas, foram criados índices para cada atributo de cada tabela. Cada tabela possui um número de índices sempre igual ao número de atributos mais um, pois além de existir um para cada atributo, também existe a chave primária.

5.3 A MIGRAÇÃO DOS DADOS

Os dados convencionais das dimensões agregáveis (característica, condição de moradia, deficiência física, educação, problema de saúde e queixa do sistema de saúde) são provenientes do Projeto Macaé Cidadão. Este trazia informações de uma pesquisa de campo que buscava dados sob diversos pontos de vista diretamente com o município, inclusive os abordados no projeto dessa dissertação.

Os dados referentes às dimensões não-agregáveis (escolas e unidades de saúde) vieram das próprias secretarias ou órgãos municipais responsáveis. Esses dados, apesar de no modelo multidimensional serem definidos como geográficos, não estavam disponibilizados na prefeitura como dados georreferenciados. Desta forma, o georreferenciamento foi realizado manualmente sobre as camadas geográficas disponibilizadas pela prefeitura, conforme será explicado adiante.

As informações espaciais (setor censitário, bairros e setores administrativos) são provenientes do Projeto GeoMacaé, que foi o projeto responsável pelo georreferenciamento do município, conforme já visto. Os atributos não-geográficos dessas dimensões foram obtidos das dimensões não-agregáveis.

5.3.1 Migração das dimensões não-agregáveis

Para o tratamento dos dados das Unidades de Saúde e das Escolas houve duas preocupações: o tratamento das informações geográficas e das informações não-geográficas ou transacionais. As informações não-geográficas foram provenientes dos órgãos responsáveis, que disponibilizaram os dados, no formato Microsoft Excel. Quanto às informações geográficas, não existia qualquer informação disponível até o momento da construção da ferramenta ora desenvolvida nesta dissertação, fazendo parte da implementação deste trabalho.

5.3.1.1 Tratamento dos dados convencionais de Escolas

As informações sobre as escolas macaenses foram disponibilizadas pela prefeitura no formato Microsoft Excel, o qual contém as seguintes informações: Nome da escola, Endereço e o foco de ensino da escola (educação infantil, pré-escolar, ensino fundamental, etc.). Essas informações estavam presentes em três planilhas distintas, uma das quais apresentava dados de escolas particulares, outra de escolas públicas municipais e outra de escolas públicas estaduais (figura 30).

SETOR	Nº	ESCOLAS	DIRETOR
	1	E.E.M. JACYRA TAVARES DUVAL (Educação Infantil - Pré-Escolar/ Ensino Fundamental 1ª a 4ª séries/ Rua: Passárgada, 34 Bairro: Novo Cavaleiros - CEP: 27930-300 Tel: 2770-2236	Marisete Faria de Oliveira
	2	E.M. LIONS (Educação Infantil - Pré-Escolar / Ensino Fundamental 1ª a 4ª séries) Rua: Irene Meirelles, s/nº Bairro: da Glória - CEP: 27910-000 Tel: 2770-2210	Therézinha Maria dos Santos Arenari
		E.M. DOLORES GARCIA RODRIGUEZ (Educação Infantil-Pré-Escolar/Ensino Fundamental-1ª a 4ª séries) Rua: W 29, 152	

Fig.30 Arquivo disponibilizado pela Prefeitura de Macaé com o cadastro de escolas

Manualmente, foi gerada uma planilha única com todas as escolas, e criado um atributo que identificava qual o tipo de escola (público municipal, público estadual ou particular). Nesta mesma planilha, foi feita uma análise sobre os focos de ensino das escolas e identificados os seguintes atributos: adultos, creche, maternal, pré-escolar, fundamental 1ª. Série, Fundamental 1ª. a 4ª. Série, Fundamental 5ª. a 8ª. série, Ensino Médio, Técnico e Educação Especial. Desta forma, foi feita uma leitura linha a linha de cada escola, atribuindo-se sim ou não para cada uma dessas características. Uma vez formada essa planilha contendo 138 registros (figura 31), bastou usar a própria aplicação do Microsoft Excel para exportar os dados para o PostgreSQL usando o ODBC para acessar os dados no servidor de banco de dados.

A	B	C	D	E	F	G	H	I	J	K	L
nomeescola	tipoescola	adultos	creche	maternal	pre_escolar	fundamental_serie1a	fundamental_serie2a	fundamental_serie3a	ensino_medio	tecnico	educacao_especial
2 C. E. Luiz Reid	Público Estadual	Sim	Não	Não	Não	Não	Sim	Sim	Sim	Não	Não
3 C.E. PEDRO ADAMI	Público Municipal	Não	Não	Não	Não	Não	Sim	Sim	Sim	Não	Não
4 C.E.M. RAUL VEIGA (Escola-Pólo)	Público Municipal	Sim	Não	Não	Sim	Não	Sim	Sim	Não	Não	Não
5 C.M. PEDRO ADAMI - SEÇÃO I	Público Municipal	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não
6 C.M. PROFESSOR SAMUEL BRUST	Público Municipal	Não	Não	Não	Não	Não	Não	Sim	Sim	Não	Não
7 C.M. PROFESSORA MARIA ISABEL DAMASCENO S	Público Municipal	Sim	Não	Não	Não	Não	Sim	Sim	Sim	Não	Não
8 C.M. ZELITA ROCHA DE AZEVEDO	Público Municipal	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não
9 CEMEAES - CENTRO MUNICIPAL DE EDUCAÇÃO E	Público Municipal	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não
10 CEMEAES - CENTRO MUNICIPAL DE EDUCAÇÃO E	Público Municipal	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não
11 CEMEAES - CENTRO MUNICIPAL DE EDUCAÇÃO E	Público Municipal	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não
12 CEMEAES - CENTRO MUNICIPAL DE EDUCAÇÃO E A	Público Municipal	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não
13 Centro Educacional Ativo	Particular	Não	Não	Não	Sim	Não	Sim	Sim	Não	Não	Não
14 Centro Educacional Crescer	Particular	Não	Não	Não	Sim	Não	Sim	Sim	Não	Não	Não
15 Centro Educacional São José	Particular	Sim	Não	Não	Não	Não	Sim	Sim	Sim	Não	Não
16 Centro Educacional Souza	Particular	Sim	Não	Não	Sim	Não	Sim	Sim	Sim	Não	Não
17 Centro Educacional Tia Gi	Particular	Não	Não	Não	Sim	Não	Sim	Sim	Não	Não	Não
18 CES Othon Barroso de Carvalho	Público Estadual	Sim	Não	Não	Não	Não	Não	Não	Sim	Não	Não
19 CIEP 068 M. OSCAR CORDEIRO	Público Municipal	Não	Não	Não	Sim	Não	Sim	Sim	Não	Não	Não

Fig.31 Arquivo de cadastro de escolas formatado para exportação para o PostgreSQL

5.3.1.2 Tratamento dos dados convencionais de Unidades de Saúde

O procedimento usado no tratamento dos dados das Unidades de Saúde não foi diferente do adotado em relação às Escolas. As informações foram disponibilizadas numa planilha do Microsoft Excel constando o nome da Unidade de Saúde, o tipo (Ambulatórios Especializados, Núcleos de Atenção à Saúde, Unidades de Saúde da Família e Unidades Básicas de Saúde, Mista, Coordenadorias, Laboratórios, Ambulatórios, Especializado, Hospitais), se é ou não associado ao SUS e o endereço (figura 32).

Id	SUS	Tipo	Nome	Rua
1	N	Ambulatórios Especializados	Centro de Saúde Dr. Jorge Caldas	Praça Washington Luiz, s/nº
2	N	Ambulatórios Especializados	Centro de Especialidades Médicas Dr. Moacyr Santos	Rua Marechal Rondon, 390
3	N	Ambulatórios Especializados	Serviço Municipal de Hemoterapia	Rua Dr. Bueno, 50
4	N	Núcleos de Atenção à Saúde - Municipal	Núcleo de Atenção à Saúde da Criança	Rua Caetano Correa reis, s/nº
5	N	Núcleos de Atenção à Saúde - Municipal	Núcleo de Atenção à saúde da Mulher e Criança	Rua Francisco Alves Machado, 235
6	N	Núcleos de Atenção à Saúde - Municipal	Núcleo Ginecológico e Obstétrico da Barra	Rua Eurico Barbosa de Souza, s/nº
7	N	Núcleos de Atenção à Saúde - Municipal	Núcleo Municipal de Higiene e Saúde Mental	Rua Dr. Bueno, 190
8	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Aterrado do Imbuuro	Rua Principal, s/nº
9	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Engenho da Praia	Rua 11, nº 274
10	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Fronteira	Rua M, nº 133
11	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Módulo Ajuda de Baixo	Av. 06, s/nº
12	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Nova Holanda	Rua 04, nº 237
13	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Bicuda Grande	Rua Principal, s/nº
14	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Cabeceira do Sana	Rua principal, s/nº
15	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Córrego do Ouro	Rua Principal, s/nº
16	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Frade	Rua Adelino José Martins, s/nº
17	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Lagomar	Rua W 14, s/nº
18	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Sana	Rua principal, s/nº
19	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Bicuda Pequena	Rua principal, s/nº
20	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Botafogo	Rua Antônio Bichara Filho, 264
21	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Catalunya em Missão	Rua Antônio Bichara Filho, 264
22	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Morro São Jorge	Rua general Craveiro Lopes, 203
23	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Malvinas A	Rua Principal, s/nº
24	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Malvinas B	Rua principal, 525
25	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Trapiche	Rua Comandante Gerson, s/nº
26	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Praia Campista	Rua Professor Gusmão, 120
27	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	PSF Areia Branca	Rua Principal, s/nº
28	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	UBS Colônia dos Pescadores 23	Rua Dr. Júlio Olivier, 148
29	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	UBS Visconde de Araújo e Miramar	Rua Ivair Itagiba, 118
30	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	UBS Morro de Santana	Rua Leopoldina Neves pinheiro, s/nº
31	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	UBS Glicério	Rua Camil Tanus, s/nº
32	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	UBS Novo Cavaleiro	Rua Nazareno, 300
33	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	UBS Imboassica	Praça São Benedito, 80
34	N	Unidades de Saúde da Família e Unidades Básicas de Saúde	UBS Barra	Estrada do Aeroporto, s/nº

Fig.32 Cadastro de unidades de saúde disponibilizado pela Prefeitura de Macaé

Para a exportação da planilha disponibilizada pela prefeitura (figura 32) foi usada a ferramenta de exportação de dados do Microsoft Excel que, através de um driver ODBC, levou os 62 registros para o modelo multidimensional.

5.3.1.3 Georreferenciamento das dimensões não-agregáveis

O próximo passo corresponde a transformar os “endereços” que estavam registrados nas planilhas em pontos do espaço geográfico, para que esses dados estivessem georreferenciados. Como a menor granularidade geográfica usada no modelo desta dissertação é o setor censitário, a preocupação era que as escolas e unidades de saúde estivessem posicionadas corretamente dentro do setor censitário respectivo.

Usando os mapas disponibilizados pela Prefeitura Municipal de Macaé, o software Planet GIS e as localizações das dimensões não-agregáveis dadas no formato convencional de

endereço, cada escola e unidade de saúde foram localadas manualmente uma a uma no mapa, gerando, mais duas camadas geográficas: UnidadesSaude e Escolas.

O modelo do mapa usado no software Planet GIS passou a possuir cinco camadas geográficas distintas: Setor Administrativo, Bairro, Setor Censitário, Escolas e Unidades de Saúde. Os três primeiros foram disponibilizados pela Prefeitura Municipal de Macaé no formato ArcView Shape File, enquanto que os dois subseqüentes foram localados e inseridos manualmente integrando um único arquivo no formato do software Planet GIS.

A figura 33 mostra as camadas geográficas: setores censitários, unidades de saúde e escolas. As unidades de saúde são representadas por um símbolo no formato de cruz na cor vermelha, enquanto que as escolas são representadas por um símbolo no formato de cruz na cor azul.

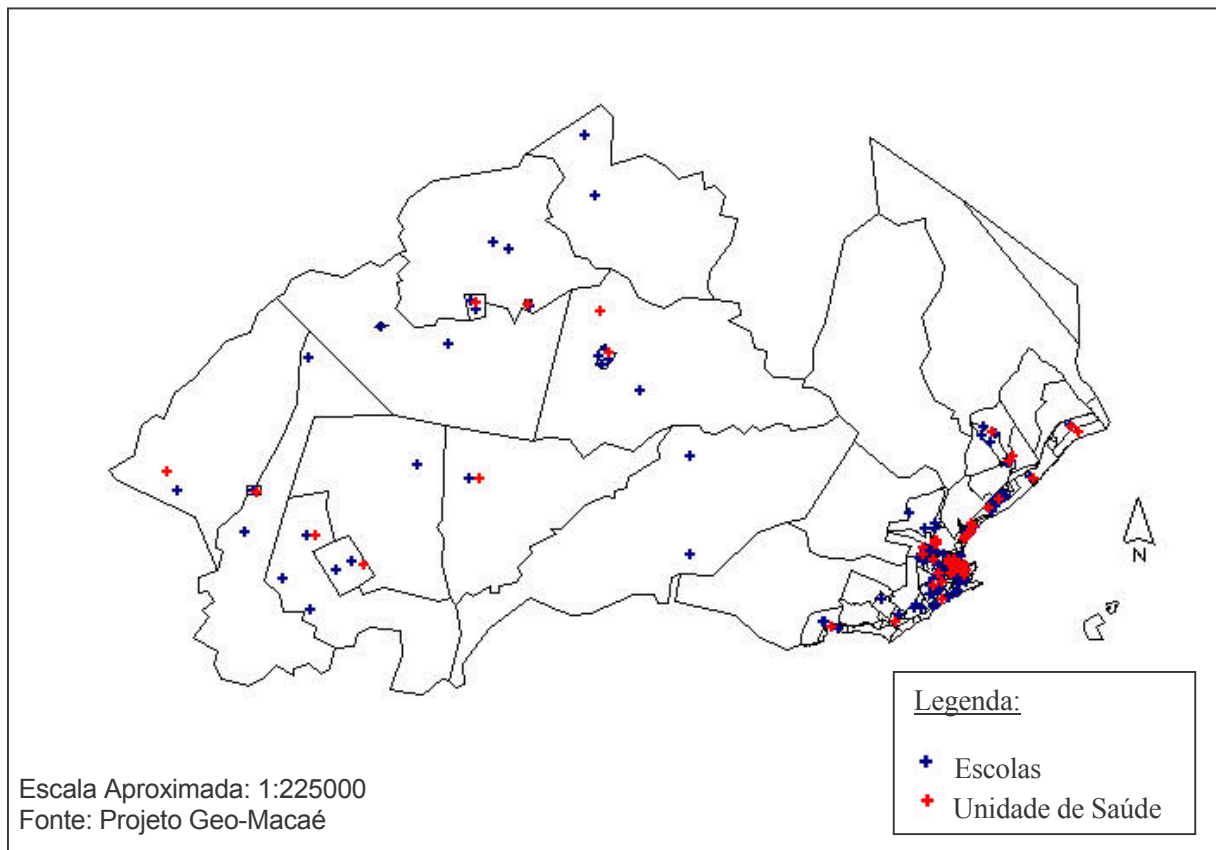


Fig.33 Mapa com as Escolas, Unidades de Saúde e Setores Censitários

O próximo passo foi levar as informações geográficas geradas no Planet GIS para o Datawarehouse. Para isso, foi usada uma ferramenta do Planet GIS que exporta uma camada geográfica para o formato de banco de dados Dbase IV. Dois arquivos Dbase foram gerados,

um com as Unidades de Saúde e outro com as Escolas. Nestes arquivos, consta o nome de cada ponto e os valores de X e Y, correspondentes ao posicionamento geográfico planimétrico em coordenadas cartesianas.

Usando o Visual Basic 6.0, foi criado um programa que lê esses arquivos e gera um arquivo texto com comandos SQL capaz de criar uma tabela e preencher essa tabela com o nome e o ponto geográfico usando os X e Y's existentes no arquivo no formato Dbase IV.

Uma vez executados esses comandos SQL, duas tabelas temporárias foram geradas no PostgreSQL com as informações necessárias. Com uma operação SQL simples foi possível a atualização das dimensões não-agregáveis escola e unidadesaude usando os nomes como chave de comparação para ambas.

5.3.2 Migração da dimensão setorcensitario

A dimensão setorcensitario foi migrada em três etapas: os dados convencionais genéricos, os dados geográficos dos polígonos de cada setor, e os totalizadores das dimensões geográficas não-agregáveis.

5.3.2.1 Migração dos dados convencionais

Para a migração dos dados convencionais foi usada a base de dados disponibilizada pela prefeitura em Microsoft SQL Server, em três tabelas diferentes. O tratamento dos atributos de origem e destino está mostrado na tabela 8.

Origem (Microsoft SQL Server)		Atributo de Destino (Datawarehouse) PostgreSQL
Tabela	Campo	
TAB_SetoresCensitarios	Setor	setorcensitario_id
TAB_Bairros	NomeBairro	nomebairro
TAB_SetorAdministrativo	NsetorAdm	nomesetoradm

Tab. 8 - Migração dos dados da dimensão **setorcensitario**

A figura 34 mostra a visão que foi criada no Microsoft SQL Server. A partir desta visão, foi utilizada a ferramenta de exportação do próprio SQL Server, que via ODBC, gravou os dados no modelo multidimensional, conforme tabela 8.

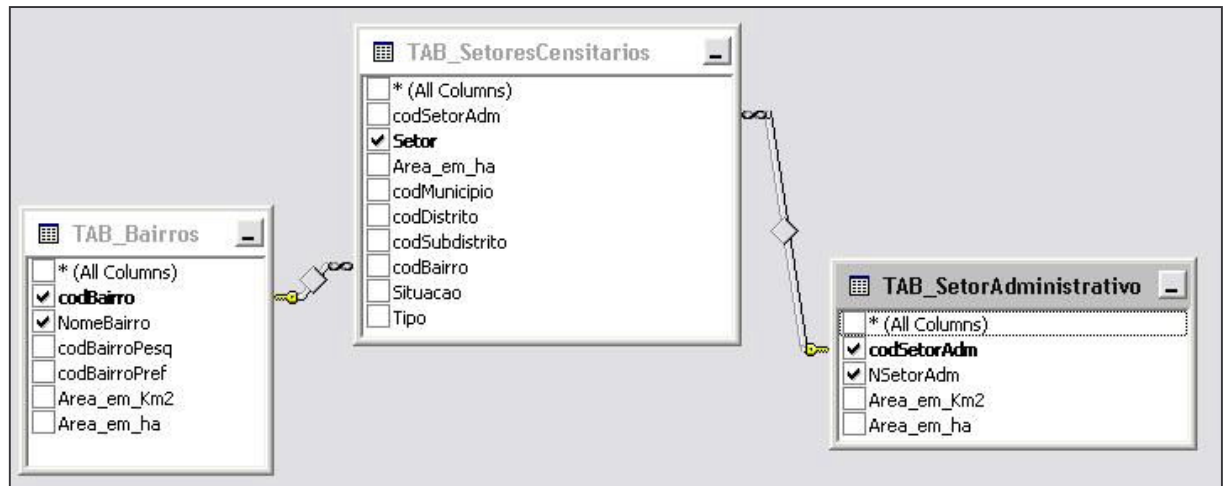


Fig.34 Tabelas de Setores Censitários disponibilizadas pela Prefeitura

5.3.2.2 Migração dos dados geográficos

Setor Administrativo, Setor Censitário e Bairro são dimensões geográficas do tipo polígono. Os mapas geográficos de cada um foram disponibilizados no formato “ArcView Shape File” pela Prefeitura Municipal de Macaé, oriundos do Projeto Geo-Macaé.

Para incorporar as informações geográficas contidas nesses arquivos ao Banco de Dados multidimensional, foi usado o aplicativo “shp2pgsql” do próprio PostGeo, capaz de, a partir de um arquivo shape, criar um arquivo com comandos SQL (figura 35) para construir e povoar uma tabela no PostgreSQL e com os dados contidos no arquivo ArcView Shape. Três arquivos com comandos SQL foram criados: um para Bairro, outro para Setor Administrativo e mais um para Setor Censitário. Esses comandos SQL, executados no PostgreSQL, geraram tabelas temporárias.


```

CREATE TABLE setcens
(gid serial, OBJECTID_1 numeric, OBJECTID numeric, ID_ varchar, OBJECTID_2 numeric, SHAPE_LEN
numeric, SHAPE_AREA numeric, FREQ_ESC_P numeric, FREQTOU_ES numeric, N_V_DEST_A numeric,
ESGOTO_INE numeric, AGUA_INEX_ numeric, SERV_PAV_R numeric, SERV_ILUM_ numeric,
SERV_COL_L numeric, POP_FUNC_P numeric, RENDA_ABAI numeric, RENDA_3_A_ numeric,
CASADO_E_V numeric, N_TRAB_DES numeric, POP_DOMPRO numeric, POP_SE_TRA numeric,
LAZER_AMB_ numeric, TOT_DOMPRO numeric, POP_TOTAL numeric, CRIAN€AS_0 numeric);

select AddGeometryColumn('','setcens','the_geom','22524','MULTIPOLYGON',2);

begin;

Insert into setcens
(gid,OBJECTID_1,OBJECTID,ID_,OBJECTID_2,SHAPE_LEN,SHAPE_AREA,FREQ_ESC_P,FREQTOU_
ES,N_V_DEST_A,ESGOTO_INE,AGUA_INEX_,SERV_PAV_R,SERV_ILUM_,SERV_COL_L,POP_FUNC
_P,RENDA_ABAI,RENDA_3_A_,CASADO_E_V,N_TRAB_DES,POP_DOMPRO,POP_SE_TRA,LAZER_A
MB_,TOT_DOMPRO,POP_TOTAL,CRIAN€AS_0,the_geom)

Values
('0','1.000000','394.000000','330240305000002','221.000000','570537089493.000000','840607339364.000000','0.
000000','0.000000','0.000000','0.000000','0.000000','0.000000','0.000000','0.000000','0.000000','0.000
000','0.000000','0.000000','0.000000','0.000000','0.000000','0.000000','0.000000','0.000000','0.000000',GeometryFromText('
MULTIPOLYGON(((208702.088405216 7548373.43813486, ...

```

Fig.35 Comandos SQL criados pelo shp2pgsql – geração dos Setores Censitários

Deste ponto, apenas foram executados comandos SQL de atualização que comparavam a tabela temporária com a tabela setorcensitario. Primeiro, foram migrados os Setores Censitários, comparando a tabela temporária com a tabela setorcensitario através da chave setorcensitario_id. Depois, foram migrados os Bairros, comparando a tabela temporária com a tabela setorcensitario, através da chave nomebairro. Por fim, foram migrados os Setores Administrativos, comparando-se a tabela temporária com a tabela setorcensitario através da chave nomesetoradministrativo.

Assim, os atributos da dimensão setorcensitario setorcensitariogeo, bairrogeo e setoradmgeo foram instanciados com o polígono que materializa a sua imagem geométrica.

5.3.2.3 Atualização dos totalizadores das dimensões geográficas não-agregáveis

A última etapa da atualização dos Setores Censitários é a atualização dos valores referentes às dimensões geográficas não agregáveis. Na verdade, esses totalizadores existem somente para agilizar o processo de consulta.

Para isso, foram utilizadas funções geográficas presentes no PostgreSQL que associam os pontos das Escolas e Unidade de Saúde com os Setores Censitários, Bairros e Setores Administrativos.

Com o comando SQL da figura 36, é retornado o número de escolas particulares existentes dentro de cada Setor Censitário. O comando “contains(polígono, ponto)” retorna se o ponto está ou não contido no polígono.

```
SELECT COUNT
(escola.nomeescola), setorcensitario.setorcensitario_id

FROM
escola, setorcensitario

WHERE

CONTAINS(setorcensitario.setorcensitariogeo, escola.escolageo) = TRUE AND
escola.tipoadministracao = 'Particular'

GROUP BY
setorcensitario.setorcensitario_id
```

Fig.36 Consulta geográfica retornando um totalizador da dimensão setorcensitario para escolas privadas

Seguindo essa estrutura, foram realizadas as atualizações dos seguinte atributos da dimensão setorcensitario, a partir de **escola**: qtdescolapublicasetorcensitario, qtdescolapublicabairro, qtdescolapublicasetoradm, qtdescolaprivadasetorcensitario, qtdescolaprivadabairro, qtdescolaprivadasetoradm.

Da mesma maneira, foram atualizados os atributos: qtdunidadesaudesetorcensitario, qtdunidadesaudebairro, qtdunidadesaudesetoradm, utilizando como base de consulta a dimensão não agregável **unidadesaude**.

5.3.3 Migração das dimensões agregáveis

A base de dados do projeto Macaé Cidadão foi disponibilizada em SQL Server em um banco de dados denominado “macae”, de onde as informações dimensionais foram importadas. Faz parte dessa dissertação o desenvolvimento de uma ferramenta, na linguagem de programação Visual Basic 6.0, que realiza todo o procedimento de limpeza, formatação dos dados, e armazenamento no Datawarehouse apenas das dimensões agregáveis.

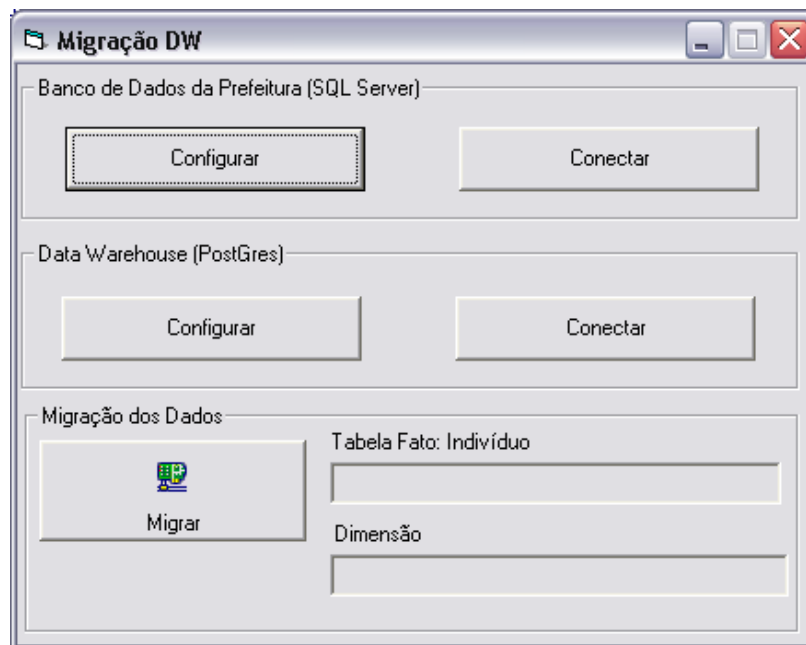


Fig.37 Ferramenta de Migração dos dados das dimensões agregáveis

A figura 37 apresenta a tela principal do programa responsável pela migração dos dados do SQL Server banco de dados “macae” para o banco de dados PostgreSQL no servidor de dados.

O objetivo deste software é fazer uma iteração pelas pessoas identificadas no Projeto Macaé Cidadão, oriundos da base de dados da Prefeitura de Macaé, observando suas características em todas as dimensões do modelo do Datawarehouse (Condição de Moradia,

Característica, Trabalho, Educação, Saúde, Queixa do Sistema de Saúde, Deficiência Física). Caso essa pessoa possua alguma característica que ainda não havia sido armazenada em alguma dimensão, ele cria aquela característica no modelo multidimensional para só então armazenar na tabela fato todas as chaves estrangeiras dessas dimensões e atribuir à quantidade de pessoas com aquelas características o valor 1.

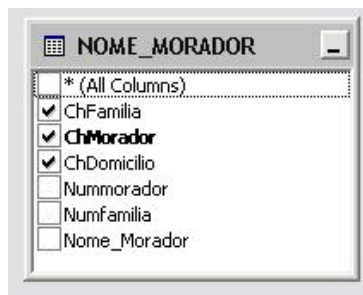


Fig.38 Atributos dos dados dos Municípios

A figura 38 mostra a tabela responsável pelo armazenamento de todos os habitantes do município de Macaé disponibilizado pela Pesquisa Macaé Cidadão. Nesta tabela, cada indivíduo possui alguns campos de identificação. Um morador tem a sua identificação única (ChMorador), a identificação de sua família (ChFamilia) e a identificação de seu domicílio (ChDomicilio). O *software* usa como base de busca das informações do indivíduo o item de dado denominado ChMorador quando forem características únicas dele ou ChDomicilio, quando forem características comuns a todos que moram naquele domicílio; entretanto, mas as informações serão armazenadas por indivíduo (granularidade mínima do Datawarehouse).

É importante salientar que na dimensão **tempo** não existe uma análise a ser feita, uma vez que os indivíduos importados num dado instante vão receber o mesmo valor de tempo. O parâmetro tempo informa somente o instante em que os dados foram migrados, para que se possa fazer uma análise das modificações sofridas pelo município ao longo do tempo. O usuário simplesmente informa a data de importação e este valor é atribuído a todos que estão sendo migrados naquele instante. A data informada é “quebrada” em diversos valores diferentes dentro da dimensão tempo, tais como: dia, mês, ano, diasemana (segunda, terça-feira, quarta-feira, etc), diasdoano (quantos dias no ano indicado), semestre (1 ou 2) e trimestre (1, 2, 3 ou 4), com o propósito de facilitar as consultas e comparações ao longo do tempo.

Na dimensão Setor Censitário, também não há uma análise a ser realizada, pois o preenchimento dessa dimensão foi realizado previamente, conforme já visto. A única preocupação foi a de informar ao *software* o código do setor censitário no qual o indivíduo

mora, para que fosse possível fazer a ligação dos dados entre a tabela fato e a dimensão geográfica denominada Setor Censitário.

No caso da ocorrência de uma pessoa coincidir com todas as características com de alguma outra pessoa, na tabela fato ocorre apenas uma edição do campo quantidadepessoas somando-se 1 ao seu valor atual. Este fato ocorreu em 2521 registros da tabela fato, onde em 2200 casos foram de 2 pessoas, 255 casos de 3 pessoas, 49 casos com 4 pessoas, 11 casos com 5 pessoas, 3 casos com 6 pessoas, 2 casos com 7 pessoas e 1 caso com 9 pessoas. Esta situação pode, a primeira vista, indicar uma possível duplicação dos dados. Entretanto, a investigação desse fato foge ao escopo da presente dissertação, razão pela qual foi omitida.

5.3.3.1 Migração da dimensão **característica**

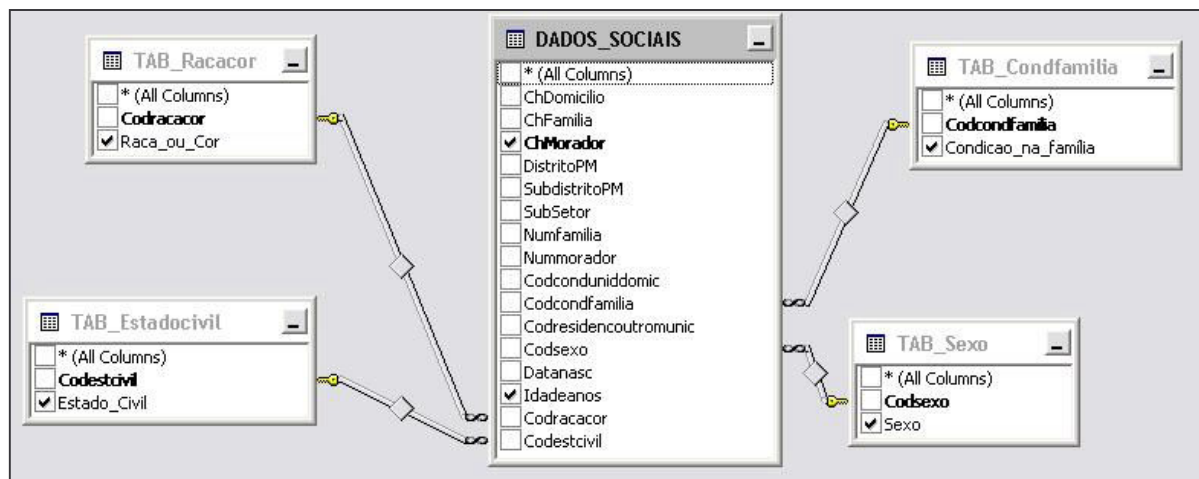


Fig.39 Origem da dimensão **característica**

A figura 39 exibe a origem dos dados da dimensão **característica**. Nesta dimensão ficam armazenadas as informações sobre diversas características que cada cidadão possui, tais como: idade, cor, estado civil, etc. Desta forma, foi usado como chave principal da análise o campo ChMorador (Chave Morador), que está na tabela DADOS_SOCIAIS que representa o código único identificador de cada morador do município gerado na Pesquisa Macaé Cidadão.

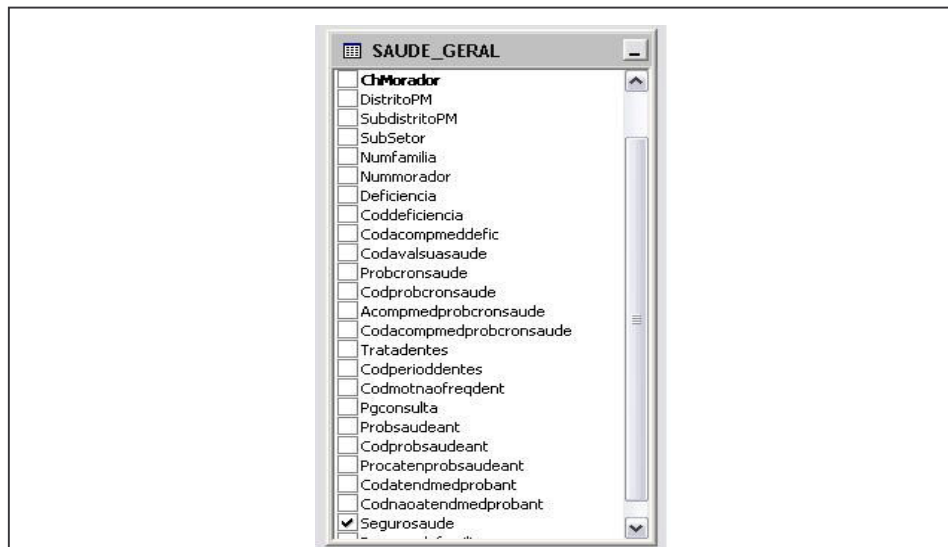


Fig.40 Origem da característica Possui Plano de Saúde

A única característica que não é oriunda da tabela vista na figura 40 é se o indivíduo possui plano de saúde, que na verdade, vem da tabela SAUDE_GERAL. A busca dessa informação é bem simples, uma vez que essa tabela possui o campo ChMorador. A tabela 9 mostra toda a comparação entre origem e destino dos dados relativos à dimensão **característica**.

Origem (Microsoft SQL Server)		Atributo de Destino (Datawarehouse)
Tabela	Campo	PostgreSQL
TAB_Condfamilia	Condicao_na_familia	Condicaofamilia
TAB_Sexo	Sexo	Sexo
TAB_racacor	Raca ou Cor	cor_raca
TAB_Estadocivil	Estado_Civil	Estadocivil
DADOS_SOCIAIS	Idadeanos	Idade
SAUDE_GERAL	Segurosaude	Possuiplanosaude

Tab. 9 - Migração dos dados de **característica**

5.3.3.2 Migração da dimensão **condicaomoradia**

Esta dimensão é responsável pelo armazenamento de informações pertinentes às diversas condições de moradia na qual o munícipe habita, tais como: tipo da parede de sua casa, tipo do teto de sua casa, tipo de água, se possui filtro de água, etc. Desta forma, para esta

dimensão, a chave de busca foi ChDomicilio (chave do domicílio), uma vez que esses dados não estão vinculados, na Pesquisa Macaé Cidadão, a um único município, mas a um domicílio.

O Datawarehouse trata como menor granularidade a pessoa, de forma que há a necessidade de expandir as informações do domicílio para cada pessoa que nele habita.

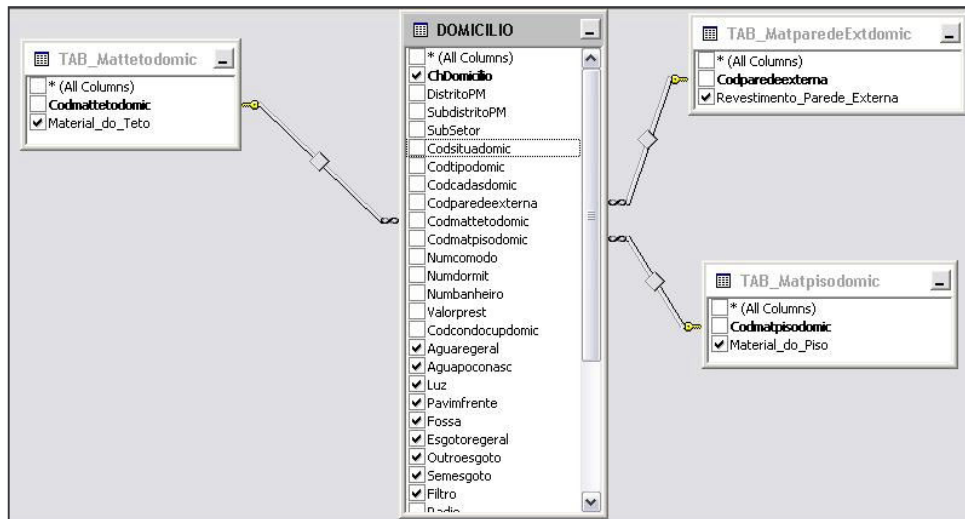


Fig.41 Origem da dimensão **condicaomoradia**

A figura 41 exhibe consulta de origem dos dados para esta dimensão. Alguns dados foram armazenados diretamente da consulta para o Datawarehouse; outros dados receberam algum tratamento antes de seu armazenamento, uma vez que foram tratados de forma diferenciada. A comparação entre os dados de origem e de destino encontra-se na tabela 10.

Origem (Microsoft SQL Server)		Atributo de Destino (Datawarehouse)
Tabela	Campo	PostgreSQL
DOMICILIO	Codsituadomic	situacaourbana
TAB_MatparedeExtdomic	Revestimento_Parede_Externa	tipoparede
TAB_Mattetodomic	Material_do_Teto	Tipoteto
TAB_MatPisodomic	Material_do_Piso	Tipopiso
DOMICILIO	Esgotoredegeral	Tipoesgoto
	Outroesgoto	
	Semesgoto	
DOMICILIO	Aguaredegeral	tipoaguarede
	Aguapoconasc	
DOMICILIO	Luz	energiaeletrica
DOMICILIO	Pavimfrente	ruapavimentada
DOMICILIO	Fossa	Fossa
DOMICILIO	Filtro	possuifiltroagua

Tab. 10 - Migração dos dados de **condicaomoradia**

Deve-se destacar os casos diferentes para três atributos: *situacaourbana*, *tipoesgoto* e *tipoaguarede*. O item *situacaourbana* que indica se o município habita numa área urbana ou rural. Este item de dados foi extraído do campo Codsituadomic, conforme tabela 10; sua semântica é: 1, se o município residir em área urbana; 0, se em área rural.

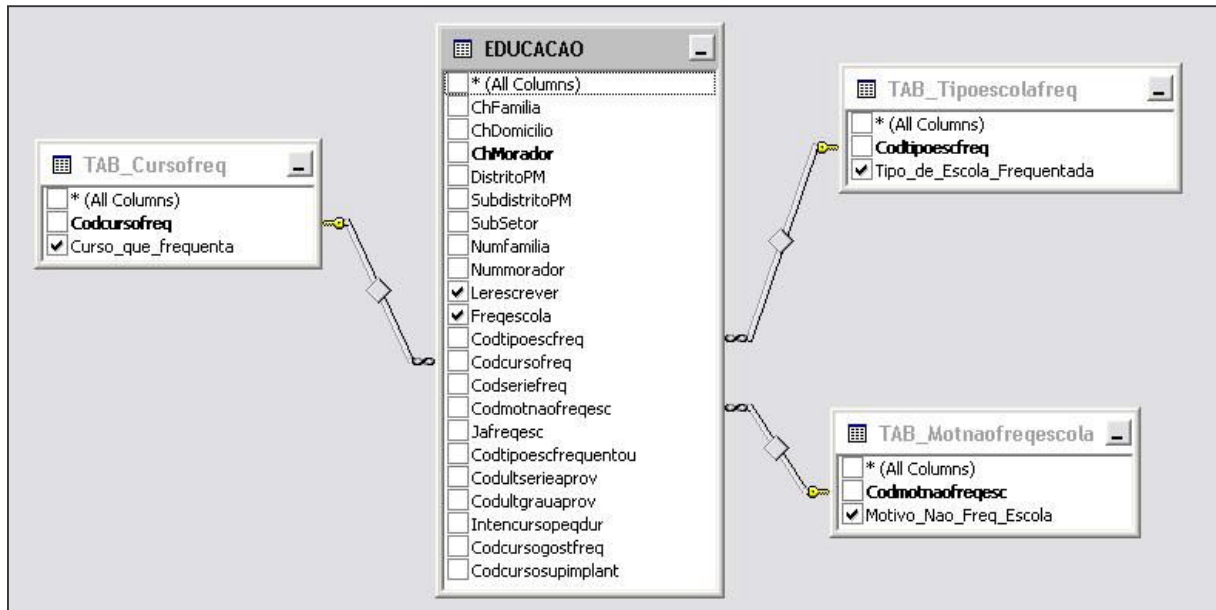
O atributo tipoesgoto é formado por três campos diferentes da tabela DOMICILIO: Esgotoredegeral, Outroesgoto e Semesgoto, conforme a tabela 10. Todos esses campos são do tipo booleano e, caso o primeiro seja verdadeiro, armazena-se no Datawarehouse “Rede Geral de Esgoto”, no segundo caso “Outra forma de esgotamento sanitário”, senão “Sem esgotamento sanitário”.

Um caso semelhante ao anterior foi usado com o atributo *tipoaguarede*, no qual havia dois campos para análise: Aguaredegeral e Aguapoconasc, conforme a tabela 10. Se o primeiro caso fosse verdadeiro, o software armazena: “Água da Rede Geral”, caso o segundo: “Água de Poço ou Nascente”.

5.3.3.3 Migração da dimensão **educacao**

Esta dimensão é responsável pelos dados referentes à educação de cada morador do município, tais como: se sabe ler, se frequenta a escola, porque não frequenta a escola, etc. Para esta dimensão a chave de busca foi ChMorador (chave do morador), uma vez que os dados referentes à educação foram armazenados individualmente pelo Projeto Macaé Cidadão.

A figura 42 mostra a consulta criada no Microsoft SQL Server que serviu como origem dos dados vindos do Projeto Macaé Cidadão. Para esta dimensão não houveram complicações na transição dos dados. Essas informações de origem e destino são mostradas na tabela 11.

Fig.42 Origem da dimensão **educacao**

Origem (Microsoft SQL Server)		Atributo de Destino (Datawarehouse) PostgreSQL
Tabela	Campo	
EDUCACAO	Lerescrever	Sabeler
EDUCACAO	Freqescola	Frequentaescola
TAB_Tipoescolafreq	Tipo_de_Escola_Frequentada	Tipoescola
TAB_cursofreq	Curso_que_frequenta	Cursoquefrequenta
TAB_Motnaofreqescola	Motivo_Nao_Freq_Escola	justificanaofrequentaescola

Tab. 11 - Migração dos dados de **educacao**

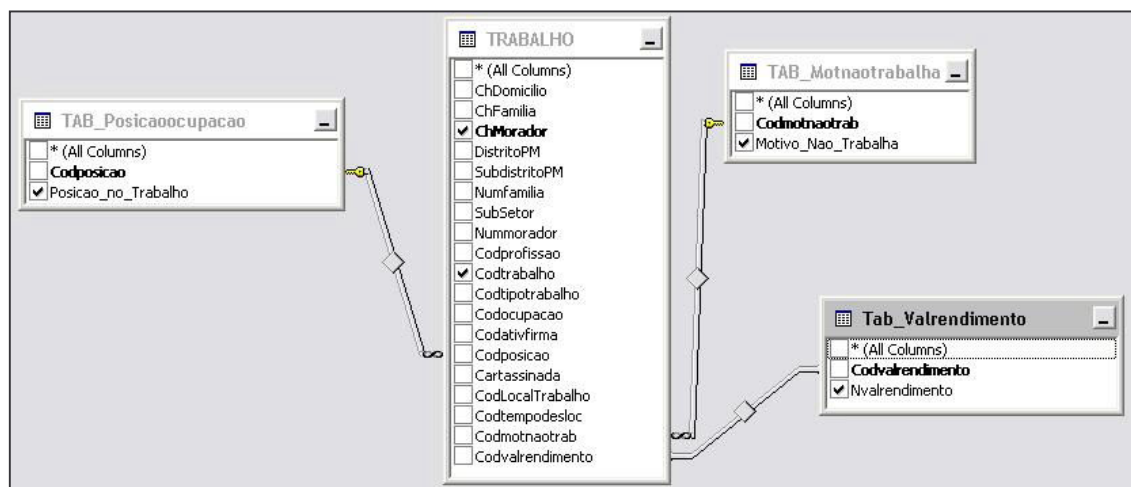
5.3.3.4 Migração da dimensão **trabalho**

Esta dimensão mostra as diversas informações referentes à situação de emprego (trabalho) e renda de cada pessoa da população do município de Macaé. Essas informações podem ser: trabalha ou não, qual a renda mensal, se trabalha em empresa pública, privada, etc. A chave de busca usada foi ChMorador (chave do morador), uma vez que os dados referentes ao trabalho e renda foram armazenados individualmente pelo Projeto Macaé Cidadão.

Origem (Microsoft SQL Server)		Atributo de Destino (Datawarehouse) PostgreSQL
Tabela	Campo	
TRABALHO	Codtrabalho	Trabalha
TAB_Posicaoocupacao	Posição no trabalho	tipotrabalhoexerce
TAB_Motnaotrabalha	Motivo_Nao_Trabalha	razaonaotrabalha
TAB_Valrendimento	Nvalrendimento	faixarendimentomensal
		rendimentomensalmenor
		rendimentomensalmaior

Tab. 12 - Migração dos dados de **trabalho**

A tabela 12 exhibe todo o processo de migração de dados. É importante salientar que houve casos, durante o processo de migração, de um morador não possuir informação a cerca do trabalho, situação em que a busca de um Morador na estrutura exibida na figura 43 não retornava registro. Para este caso o atributo “trabalha” recebia o valor “falso”; o atributo “tipotrabalhoexerce” recebia o valor “Sem Resposta”; o atributo “faixarendimentomensal” recebia o valor "Sem Declaração ou Rendimento" e os atributos “RendimentoMensalMenor” e “RendimentoMensalMaior” recebiam o valor zero.

Fig.43 Origem da dimensão **trabalho**

Os atributos faixarendimentomensal, rendimentomensalmenor e rendimentomensalmaior, conforme pode ser visto na tabela 12, vêm de um único atributo: o Nvalrendimento. O atributo faixarendimentomensal é exatamente o mesmo valor do campo Nvalrendimento. Já os atributos redimentomensalmenor e rendimentomensalmaior são, na verdade, uma análise sobre o atributo faixarendimentomensal, no qual são extraídos os valores do menor e do maior rendimento. Por exemplo, se a faixarendimentomensal for “Entre 1 a 2

Salários Mínimos”, obviamente durante a transição será armazenado em rendimentomensalmenor o valor 1 e em rendimentomensalmaior o valor 2.

5.3.3.5 Migração da dimensão **deficienciafisica**

Nesta dimensão, a responsabilidade é quanto às pessoas que possuem algum tipo de deficiência física. Desta forma, será encontrada nesta dimensão: se a pessoa possui algum tipo de deficiência física, qual o tipo de deficiência física que ela possui, se faz algum tipo de acompanhamento médico, etc. Para isso, chave de busca usada é ChMorador (chave do morador), uma vez que essas informações são pertinentes a cada munícipe, de forma independente.

A figura 44 exibe os dados de origem provenientes do Microsoft SQL Server, enquanto que a tabela 13 mostra o comparativo entre os atributos de origem e os campos de armazenamento no modelo multidimensional.

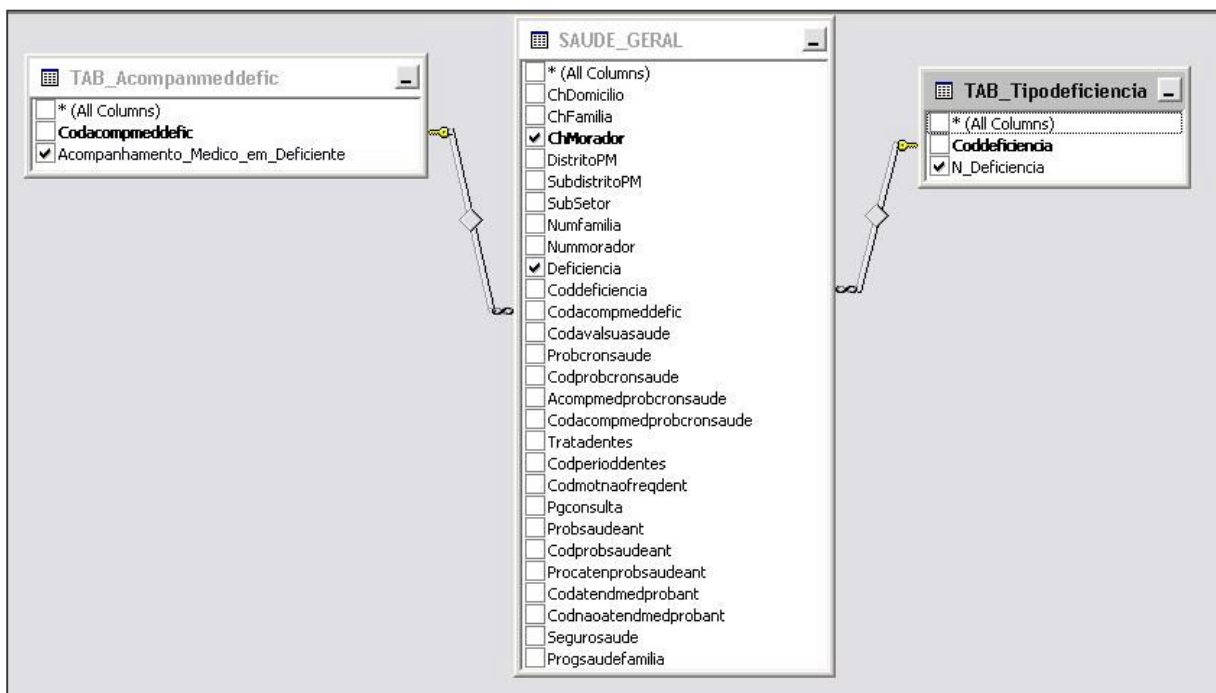


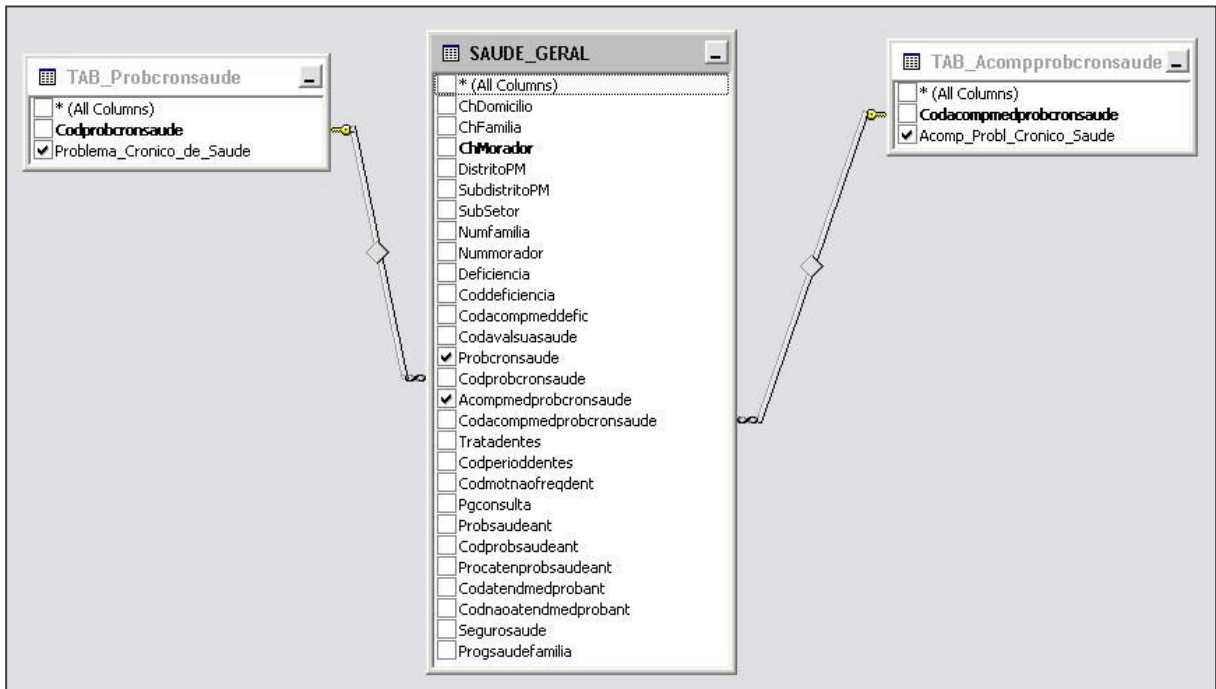
Fig.44 Origem da dimensão **deficienciafisica**

Origem (Microsoft SQL Server)		Atributo de Destino (Datawarehouse) PostgreSQL
Tabela	Campo	
SAUDE_GERAL	Deficiencia	possuideficienciafisica
TAB_Tipodeficiencia	N_Deficiencia	tipodeficiencia
TAB_Acompanmeddefic	Acompanhamento_medico_em_Deficiencia	tipoacompanhamentomedico

Tab. 13 - Migração dos dados de **deficienciafisica**

5.3.3.6 Migração da dimensão **problemasaude**

Nesta dimensão a preocupação está com os problemas de saúde que podem inclusive necessitar de algum tipo de acompanhamento médico, tais como: hipertensão, problema respiratório, problema ósseo ou muscular, diabetes, etc. Desta forma, a chave de procura tem que ser o indivíduo (ChMorador). A origem dos dados é mostrada na figura 45 e o processo de transição da origem para o destino está exibido na tabela 14.

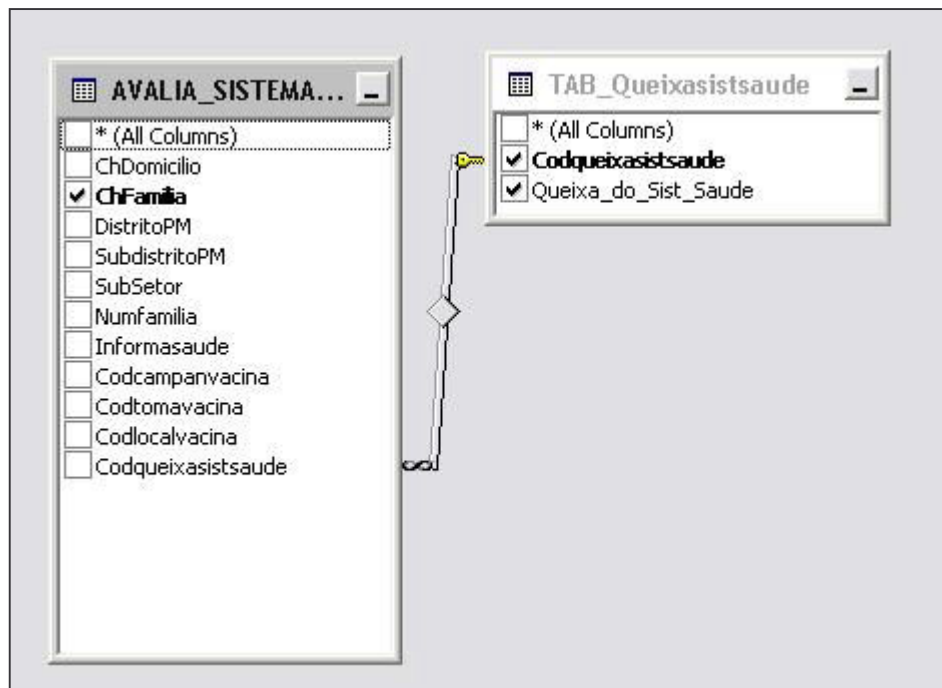
Fig.45 Origem da dimensão **problemasaude**

Origem (Microsoft SQL Server)		Atributo de Destino (Datawarehouse) PostgreSQL
Tabela	Campo	
SAUDE_GERAL	Probcronsaude	possuiproblemacronicosaude
TAB_Probcronsaude	Problema_Cronico_de_Saude	tipoproblemasaude
SAUDE_GERAL	Acompmedprobcronsaude	fazacompanhamentomedico
TAB_Acompprobcronsaude	Acomp_Probl_Cronico_Saude	localacompanhamentomedico

Tab. 14 - Migração dos dados de **problemasaude**

5.3.3.7 Migração da dimensão **queixasistemasaude**

Nesta dimensão serão armazenados os dados referentes às principais queixas que o município fez na Pesquisa Macaé Cidadão, referentes à saúde no município. Essas queixas são avaliadas por pessoa; desta forma, tem-se como chave de busca o item ChMorador.

Fig.46 Origem da dimensão **queixasistemasaude**

As formas de armazenamento do Datawarehouse e do banco de dados da Pesquisa Macaé Cidadão são completamente diferentes. Enquanto que na Pesquisa Macaé Cidadão foram armazenados textos contendo as possíveis queixas de cada cidadão, no Datawarehouse houve uma análise das respostas possíveis (obtidas no questionário usado), e cada resposta possível se

transformou em um atributo do tipo booleano. O primeiro atributo é se o município possui ou não queixa, que é descoberto quando do retorno dos registros sobre a consulta apresentada na figura 44: caso não haja retorno de registros, é armazenado o valor lógico “Falso”; caso contrário é armazenado “Verdadeiro”.

Sobre todos os demais atributos é realizada uma análise sobre a resposta e enquadrada de acordo com o atributo booleano correspondente. É importante lembrar que os valores armazenados na base de dados da Pesquisa Macaé Cidadão estão padronizados, não gerando qualquer tipo de problema neste processo de migração.

5.4 O TRATAMENTO DOS MAPAS

As informações georreferenciadas do município de Macaé foram obtidas de duas formas diferentes:

- Mapas obtidos diretamente da Prefeitura Municipal de Macaé, gerados pelo Projeto Geo-Macaé. Nesse sentido foram disponibilizados os seguintes mapas: de setor censitário, setor administrativo e bairros. Todos esses mapas encontram-se no formato ArcView Shape File.
- Dados que foram georreferenciados nesta dissertação. Neste tópico se encontram as escolas e as unidades de saúde. Para cada escola e/ou unidade de saúde foram observados os seus endereços e inseridos ponto a ponto. A preocupação foi colocar uma determinada escola ou unidade de saúde no setor censitário correto.

O primeiro passo foi a integração dos mapas vindos do Projeto Geo-Macaé em um único mapa no formato do software Planet GIS, e cada um dos mapas se tornaram camadas de um mapa denominado “macae”.

Neste mapa, foram criadas mais duas camadas (unidade saúde e escola), ambas do tipo ponto, definidas por símbolos do tipo cruz, onde as unidades de saúde são representadas em cor vermelha e as escolas com cruces de cor azul. Os pontos de cada escola e unidade de saúde foram locados sob a visualização dos setores censitários e por comparativo com mapas impressos, conforme exemplificado na figura 47.

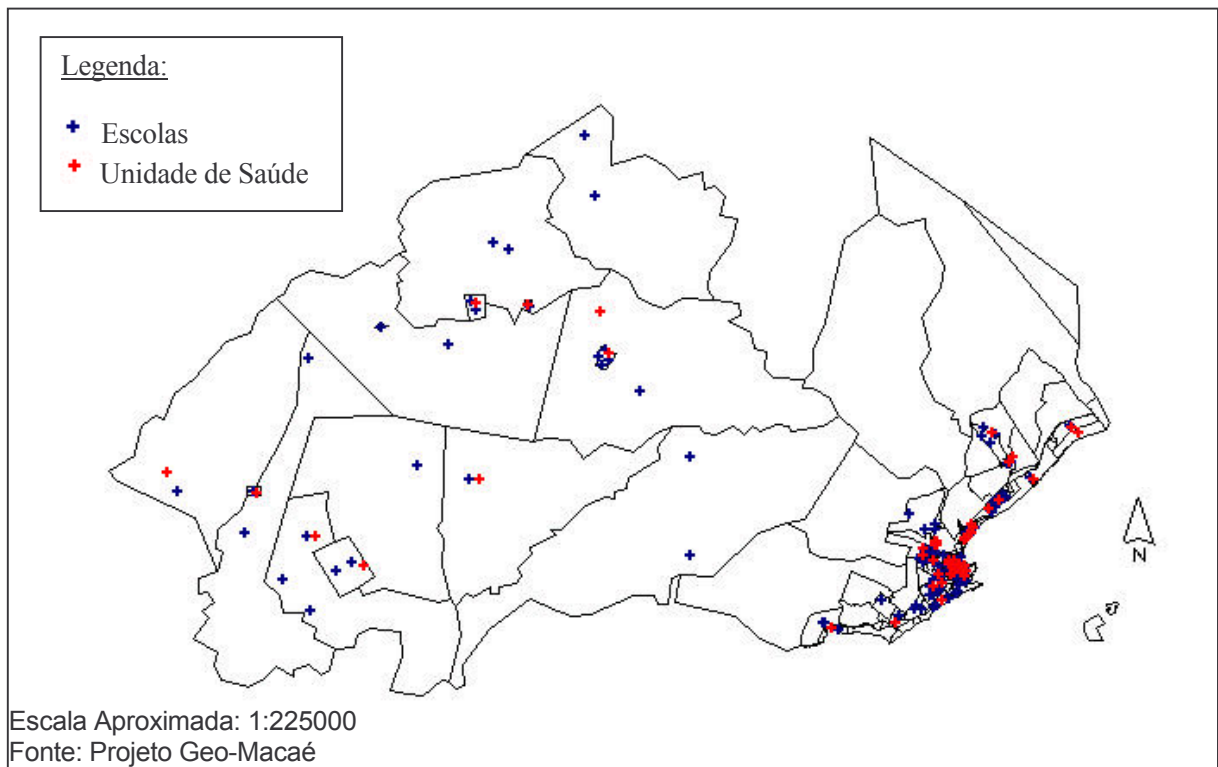


Fig.47 Mapa completo com todas as cinco camadas geográficas

5.5 A PREPARAÇÃO DO CUBO OLAP

Uma vez preparado o esquema do Datawarehouse, migrados os dados, preparados e integrados todos os mapas, o próximo passo é a preparação da ferramenta OLAP para que seja possível a realização de consultas.

Para as consultas OLAP e Geográfica foi usada a ferramenta PostGeoOlap. Trata-se de uma ferramenta que trabalha sobre o banco de dados PostgreSQL, conforme visto no capítulo 2. O objetivo deste tópico é mostrar todos os processos de configuração do cubo OLAP para a realização de consultas “ad hoc”.



The image shows a Windows-style dialog box titled "Esquema". It has a standard title bar with minimize, maximize, and close buttons. The dialog contains several text input fields arranged vertically. The first field is labeled "Servidor" and contains the IP address "192.168.0.100". The second field is labeled "Usuario" and contains "postgres". The third field is labeled "Senha" and is empty. The fourth field is labeled "Database" and contains "dwprefeitura". The fifth field is labeled "Mapa" and contains the file path "C:\Documents and Settings\Alan Carval". The sixth field is labeled "SRID do Mapa" and contains the number "22524". At the bottom of the dialog, there are two buttons: "Conectar" on the left and "Cancelar" on the right.

Fig.48 Esquema de conexão com o banco de dados multidimensional

A figura 48 se refere ao processo de conexão ao banco de dados onde está armazenado o modelo multidimensional com as dimensões geográficas. Pode-se observar na figura 48 a configuração do banco de dados e a referência ao mapa.

O próximo passo é a configuração da tabela fato, das dimensões agregáveis e das dimensões geográficas não-agregáveis. Para isto, foi criado um cubo denominado “pmm”. Neste cubo foram adicionadas as dimensões, com a preocupação na definição dos níveis de consolidação que se pode ter ao longo de cada dimensão. Esta operação foi denominada “hierarquização”.

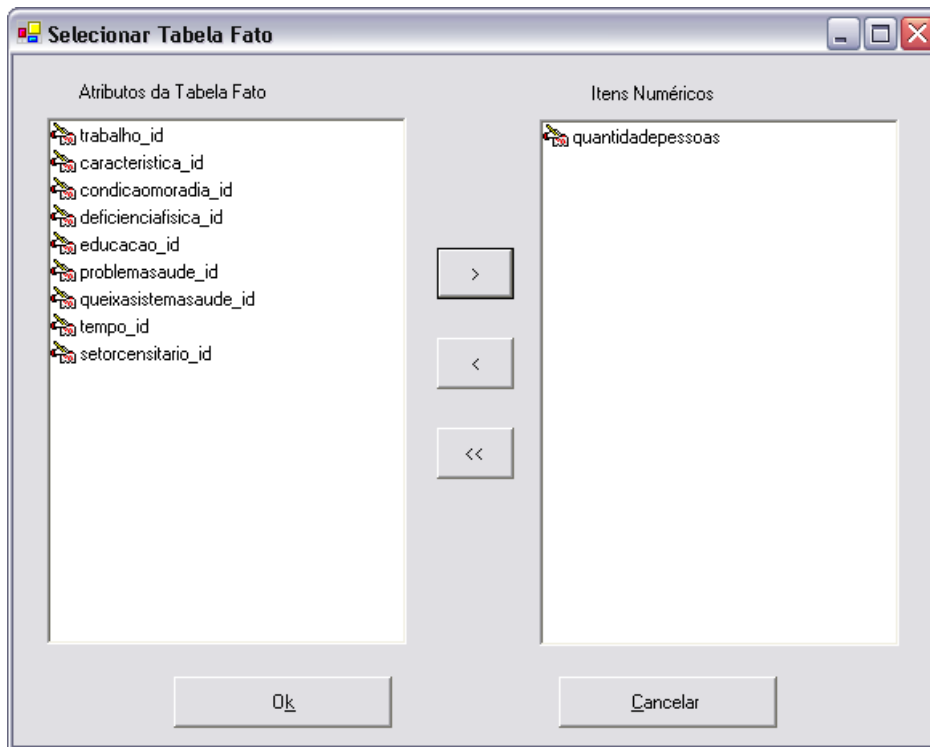


Fig.49 Definição da **tabela-fato**

A figura 49 mostra a definição da tabela fato usando o PostGeoOlap. O software define como Item Numérico todo atributo da tabela fato sobre o qual serão realizadas operações, tais como: soma, média, etc. No caso do cubo pmm, a tabela fato possui um único item numérico que é denominado “quantidade pessoas”. Este atributo representa a quantidade de pessoas que possuem uma determinada característica. Sobre esse atributo o cubo deve ser capaz de realizar operações de soma.

Nome	Nivel Hierar
ano	6
semestre	7
trimestre	8
diasdoano	9
mes	9
dia	9
diasemana	9
data	9
*	

Fig.50 Definição das hierarquias da dimensão **tempo**

Após a definição da tabela fato, o próximo passo é a definição das hierarquias internas de cada dimensão. No caso da dimensão **tempo**, a figura 50 demonstra como as hierarquias foram definidas.

Na dimensão **deficienciafisica** foi definido o atributo possuideficiencia como sendo de maior nível hierárquico e os demais atributos (tipoacompanhamentomedico e tipodeficiencia) num nível imediatamente abaixo, conforme figura 51. Na dimensão **educacao** o atributo sabeler ficou definido em um nível hierárquico superior. Logo abaixo dele, vem o frequentaescola e, abaixo deste último, encontram-se os demais atributos (justificativanaofrequentaescola, cursoquefrequenta e tipoescola).

Nome	Nivel Hierarq
possuideficiencia	8
tipoacompanhamentomedico	9
tipodeficiencia	9

Fig.51 Definição das hierarquias da dimensão **deficienciafisica**

Na dimensão **problemasaude**, o atributo possuiproblemacronicosaude foi definido como de nível hierárquico mais alto, deixando fazacompanhamentomedico, localacompanhamentomedico e tipoproblemasaude imediatamente abaixo. Na dimensão queixasistemasaude, foi definido como atributo de maior nível hierárquico o atributo denominado “possuiqueixa”, deixando todos os demais abaixo deste.

Na dimensão **condicaomoradia**, foi definido o atributo situacaourbana como único atributo de nível hierárquico mais alto. Para a dimensão **trabalho** não foram definidas hierarquias, uma vez que fica imperceptível algum tipo de agrupamento dentro desta dimensão.

Nome	Nivel Hier
cor_raca	8
possuiplanosaudeprivado	9
idade	9
estadocivil	9
sexo	9
condicaofamilia	9

Fig.52 Definição das hierarquias da dimensão **caracteristica**

Na dimensão **característica** (figura 52) o único agrupamento definido foi em relação a `cor_raca`, e desta forma, ter-se-ia todas as demais características baseadas na cor ou raça de cada indivíduo.

Finalmente, tem-se a dimensão **setorcensitario**, que tem os seus níveis hierárquicos já definidos no modelo conceitual visto na figura 53. Todos atributos da dimensão **setoradministrativo** foram definidos como atributos de nível hierárquico mais alto. Todos os atributos da dimensão **bairro** foram definidos como atributo da dimensão **bairro**. E, por fim, todos os atributos da dimensão **setorcensitario** foram definidos como os atributos de menor hierarquia.

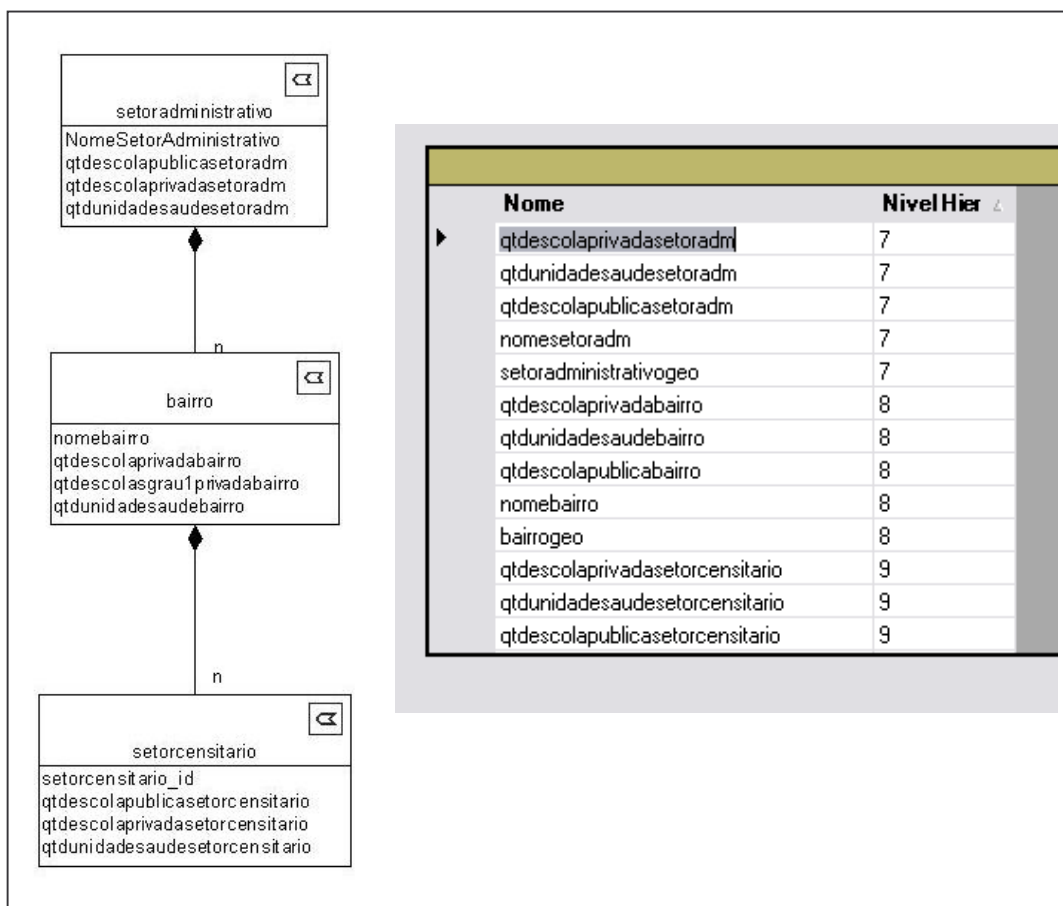


Fig.53 Comparativo entre a hierarquia do cubo e a hierarquia dimensional conceitual

Após a definição da tabela fato e das dimensões, o próximo passo é a definição das dimensões geográficas não agregáveis (**escola** e **unidadesaude**). Em seguida, realizam-se as pré-agregações, baseado na operação definida (soma sobre o atributo **quantidadepessoas**), em todas as dimensões com suas devidas hierarquias. Isto permite que o software desenvolvido

realize a geração de tabelas com as previsões de consultas sobre os dados para agilizar o processo de busca de informações do usuário.

De posse dessas pré-agregações realizadas, a ferramenta OLAP espacial está pronta para a realização de consultas “on-line” para a validação e busca de informação para o gestor municipal.

6 TESTES E ANÁLISE DOS RESULTADOS

Este capítulo tem como propósito realizar algumas consultas para testar a validade do sistema desenvolvido, tanto sobre a ótica da funcionalidade quanto da importância na obtenção dos resultados.

Será mostrado de que forma o tomador de decisão pode realizar suas consultas, exibindo-se a estrutura do sistema.

6.1 FUNCIONAMENTO DA FERRAMENTA OLAP

A figura 54 exibe a tela principal do sistema OLAP do município de Macaé. Na parte esquerda encontram-se a tabela fato, as dimensões agregáveis e as dimensões não agregáveis. No lado superior direito encontra-se o mapa com os displays (maneira com que o Planet GIS se refere às feições geográficas visíveis). A parte inferior da tela é o setor onde serão exibidos os resultados convencionais.

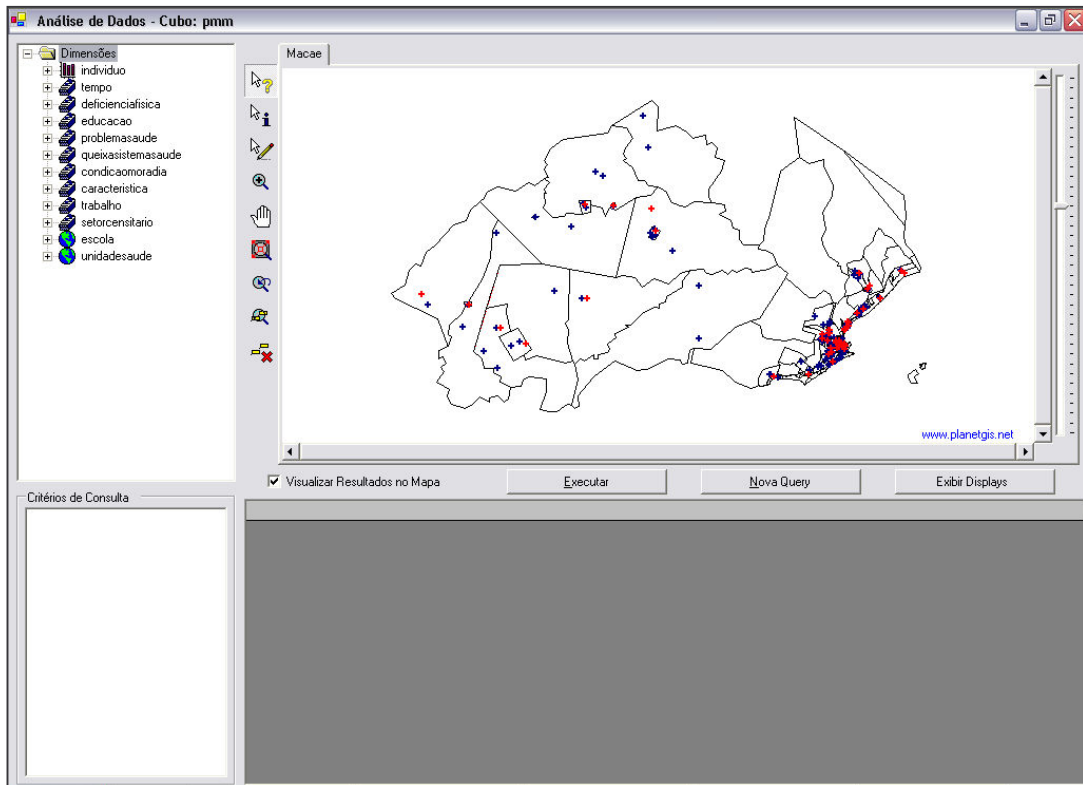


Fig.54 Tela principal da ferramenta OLAP

Existem dois caminhos que o usuário pode tomar: selecionar resultados (“slice”) ou realizar filtros (“dice”). Por exemplo, se quiser saber quantas pessoas moram no bairro do “Parque Aeroporto”, a operação de “slice” é “quantas pessoas”, ou seja, o campo que se deseja mostrar, enquanto que a operação de “dice” será “moram no bairro do Parque Aeroporto”, ou seja, o filtro que se deseja.

Como propriedade do cubo, as operações de soma são realizadas somente sobre o campo “quantidadepessoas” da tabela fato (individuo). Todos os demais atributos, se selecionados, serão para a realização de agrupamento de resultados.

A figura 55 exemplifica o caso da seleção de moradores no “Parque Aeroporto” querendo-se saber o total de pessoas que lá residem. Observa-se que, na parte inferior da tela, o atributo quantidadepessoas já está selecionado, o que significa que sobre ele será realizado uma totalização em função dos demais atributos selecionados.

Um outro fato mostrado na figura 55 é a operação de “slice”, em que o usuário está selecionando “Parque Aeroporto”.

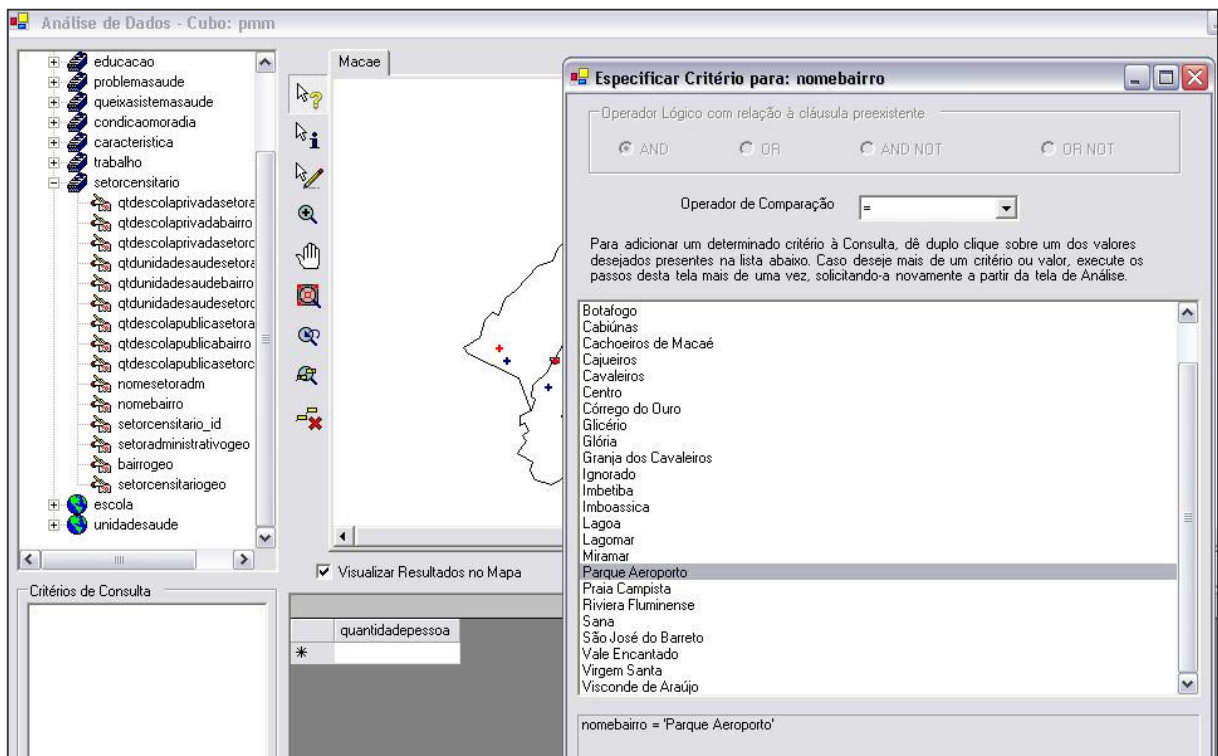


Fig.55 Operações de “slice” e “dice” sobre o cubo

A figura 56 exibe o resultado da consulta: são 17.055 pessoas habitando o bairro do Parque Aeroporto. Observa-se que a ferramenta totalizou a quantidade de pessoas.

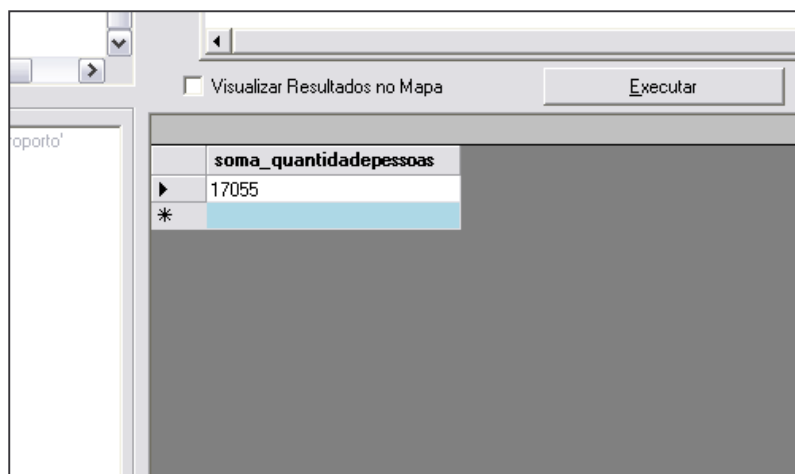
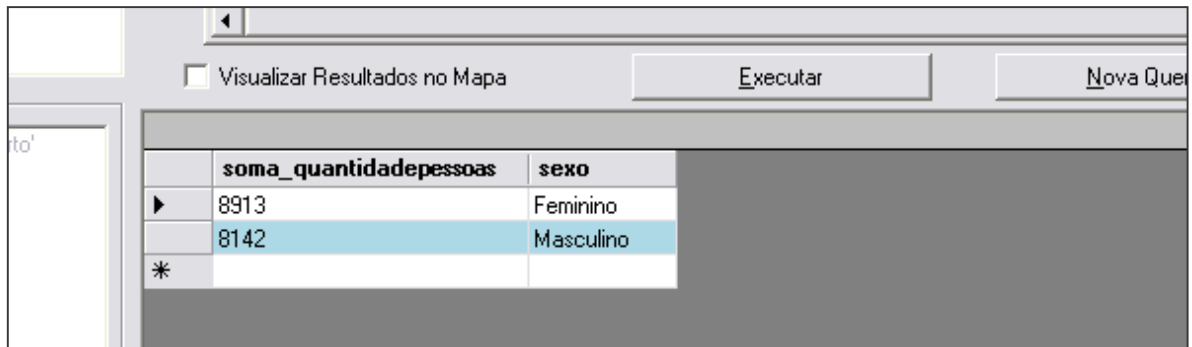


Fig.56 Resultado da consulta da quantidade de moradores de um bairro

As consultas podem ser agrupadas por um determinado atributo. Suponha-se que se deseja descobrir a quantidade de pessoas moradoras do bairro “Parque Aeroporto” por sexo.

Neste caso, basta solicitar a ferramenta que também seja exibido o campo “sexo” da dimensão **característica**. Como este atributo não é um atributo de sumarização, este então irá definir um atributo de agrupamento de resultados, conforme pode ser visto na figura 57.

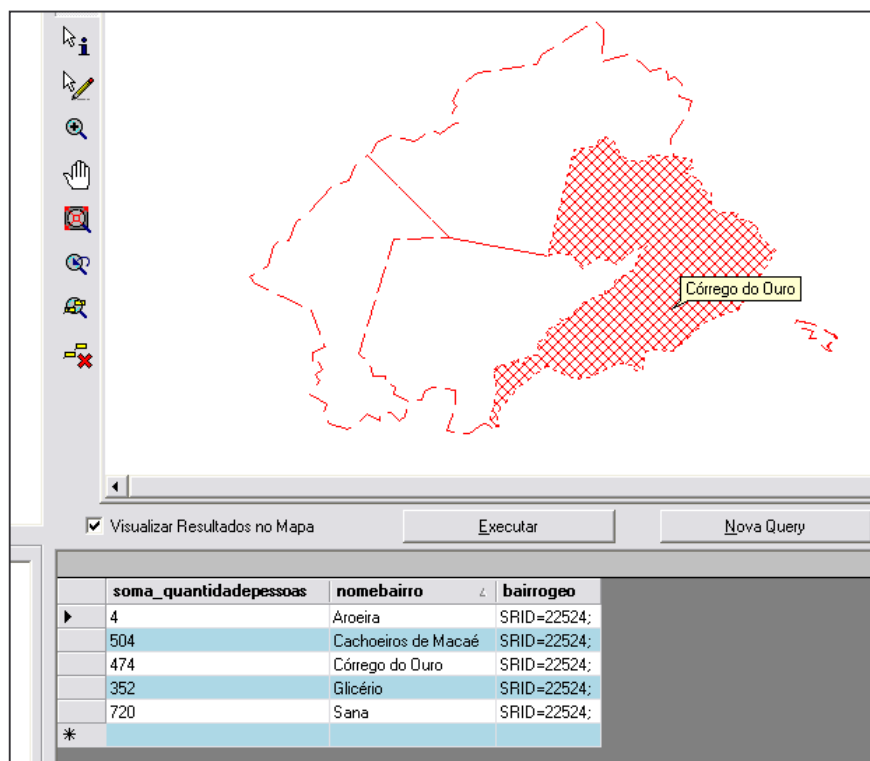


	soma_quantidadepessoas	sexo
▶	8913	Feminino
	8142	Masculino
*		

Fig.57 Resultado da consulta da quantidade de moradores de um bairro por sexo

Caso se deseje exibir os resultados geográficos, basta selecionar como atributo a ser exibido algum atributo geográfico, tal como: setorcensitariogeo, bairrogeo e setoradmgeo.

A figura 58 mostra o resultado de uma pesquisa sobre a quantidade de pessoas do sexo masculino que moram em área rural. O software exibe o resultado na listagem mostrando os valores e apresentando no mapa os locais onde essas pessoas moram.



	soma_quantidadepessoas	nomebairro	bairrogeo
▶	4	Aroeira	SRID=22524;
	504	Cachoeiros de Macaé	SRID=22524;
	474	Córrego do Ouro	SRID=22524;
	352	Glicério	SRID=22524;
	720	Sana	SRID=22524;
*			

Fig.58 Exemplo de consulta retornando resultados geográficos

Uma outra funcionalidade que pode ser utilizada é o filtro sobre as dimensões geográficas. Para isso, deve-se utilizar as dimensões geográficas não agregáveis. A figura 59 exhibe um exemplo para este caso. Estão sendo exibidos os Setores Censitários de um bairro onde se encontra a escola “C. E. Luiz Reid”. Observa-se que o termo “onde se encontra” é uma consulta geográfica do tipo “contains” (tabela 2, capítulo 2), feita entre o atributo bairrogeo que se encontra na dimensão agregável **setorcensitario** e o atributo escolageo, que se encontra na dimensão não agregável denominada **escola**.

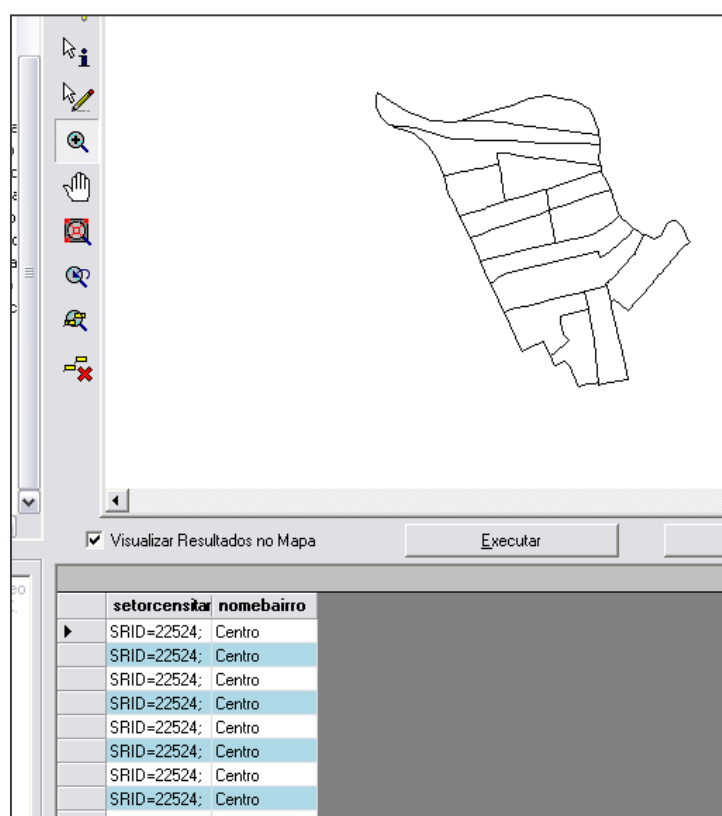


Fig.59 Exemplo de consulta usando função geográfica

Uma vez vistas as potencialidades da ferramenta de consulta, o próximo passo será realizar consultas que já validem o uso da ferramenta com o propósito de busca de informações gerenciais.


6.2 REALIZAÇÃO DE CONSULTAS

Serão abordadas neste item algumas consultas, com o propósito de validar a finalidade do sistema. Para isso foram levantadas algumas questões junto à Prefeitura de Macaé, para poder dar mais autenticidade e valor aos testes aqui realizados. Para responder a cada questão, primeiramente serão explicados quais atributos de cada dimensão estão envolvidos e, em seguida, apresentada a resposta dada pelo sistema. Vale ressaltar que os mapas apresentados nas figuras a seguir destinam-se a apenas à visualização dos resultados das consultas. Não houve esforço maior na contextualização cartográfica dos mesmos, em face do escopo desta dissertação.

6.2.1 Quantidade de pessoas por Setor Administrativo que possuem mais de 45 anos e estão desempregadas.

Para a realização desta consulta, estão envolvidas as dimensões **característica**, **trabalho**, **setorcensitario**, além da tabela fato.

Sobre a dimensão **característica** deve-se informar que o atributo idade deve ser maior que 45 anos. Sobre a dimensão **trabalho**, deve-se informar que trabalha possui o valor lógico falso. Na dimensão **setorcensitario** não será realizada qualquer seleção, exceto a escolha do atributo nomesetoradm para a exibição, já que assim será realizado o agrupamento dos dados. Finalmente, na tabela fato, é selecionado o atributo quantidadepessoas para a realização da operação de totalização. O resultado desta consulta é exibido na figura 60.



	nomesetoradm	soma_quantidadepessoas
▶	Amarelo	2893
	Azul	977
	Bege	193
	Cinza	513
	Laranja	658
	Marron	2251
	Rosa	1790
	Verde	2162
	Vermelho	3192
*		

Fig.60 Pessoas com mais de 45 anos, desempregadas por Setor Administrativo

6.2.2 Quantidade de pessoas sem estudo que estão com problema crônico de saúde, tem mais de 50 anos e estão desempregadas, exibindo o resultado por sexo.

Para a realização desta consulta, estão envolvidas as dimensões **característica**, **trabalho**, **educacao**, **problemasaude** e a tabela fato.

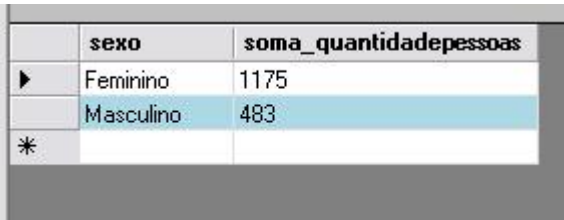
Sobre a dimensão **característica** deve-se informar que o atributo idade deve ser maior ou igual a 50 anos, além de solicitar a exibição do sexo, para que o sistema possa fazer o agrupamento desejado.

Sobre a dimensão **trabalho**, deve-se informar que trabalha é falso para buscar as pessoas que estão desempregadas.

Na dimensão **problemasaude** será realizada a busca somente das pessoas que possuem algum tipo de problema crônico de saúde, bastando informar que o atributo possuiproblemacronicosaude é verdadeiro.

Para buscar somente as pessoas que não possuem estudo deve-se, na dimensão **educacao**, informar que o atributo sabeler é falso.

Selecionando para a exibição o atributo quantidadepessoas da tabela fato indivíduo, tem-se o resultado exibido na figura 61.



	sexo	soma_quantidadepessoas
▶	Feminino	1175
	Masculino	483
*		

Fig.61 Pessoas sem estudo, com problema crônico de saúde, mais de 50 anos, desempregadas por sexo

6.2.3 Quantidade de pessoas que recebem educação especial e moram a mais de 1000 metros de alguma Unidade de Ensino Público

Alguns cuidados devem-se ter na realização desta consulta, que envolve a tabela fato e as dimensões **escola**, **setorcensitario** e **educacao**, uma vez que somente o seu enunciado não traz todos os filtros realmente existentes.

Usa-se o atributo cursoquefrequenta para informar que devem ser envolvidas nesta consulta somente as pessoas que recebem educação especial.

Na busca das pessoas que moram a 1000 metros de alguma unidade de ensino público têm-se duas dimensões envolvidas: a dimensão agregável **setorcensitario** e a dimensão não-agregável **escola**. Sobre o atributo geográfico setorcensitariogeo realiza-se a operação de comparação de distância sobre o atributo escolageo da dimensão **escola**, desde que as escolas possuam o atributo educacaoespecial verdadeiro e o atributo tipoadministracao seja diferente de particular, uma vez que o tipo de administração pode ser Público Municipal, Público Estadual ou Particular.

Para engrandecer o resultado desta consulta, foram exibidos no mapa os setores censitários envolvidos, além da quantidade de pessoas da tabela fato, conforme pode ser visualizado na figura 62.

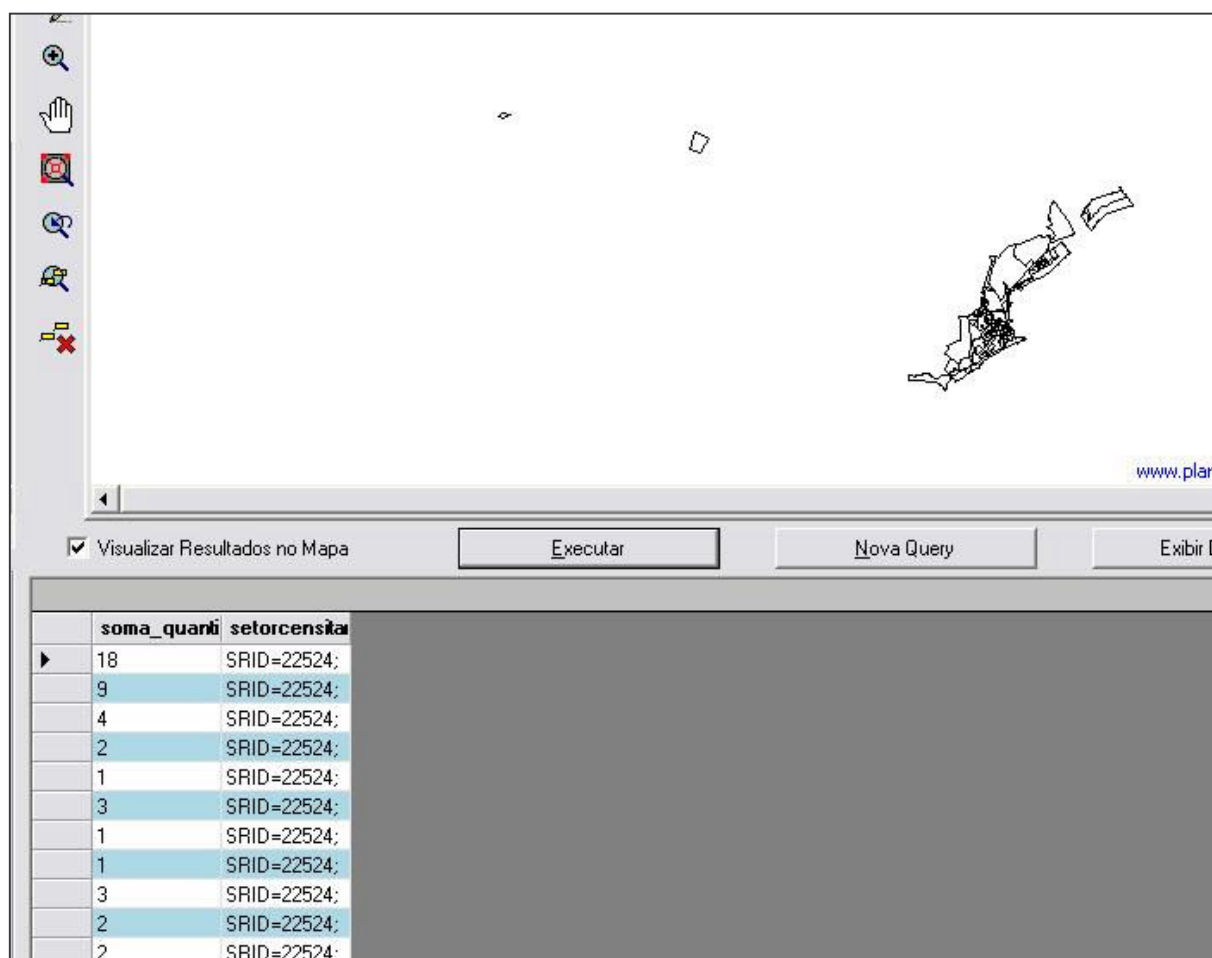


Fig.62 Setores Censitários que possuem pessoas com educação especial morando a mais de 1Km de alguma Unidade de Ensino Público

6.2.4 Quantidade de brancos e outras raças, que estejam cursando o 3º grau, estejam ou não trabalhando

Esta é uma consulta que envolve a tabela fato e as dimensões **educacao**, **caracteristica** e **trabalho**.

A afirmativa “que estão trabalhando e os que não estão trabalhando” deve ser vista como uma forma de agrupamento do resultado pelo atributo trabalha da dimensão **trabalho**. Além desse agrupamento, entende-se que as raças também são formas de agrupamento.

Desta forma, devem ser exibidos no resultado os atributos quantidadepessoa da tabela fato, o atributo trabalha da dimensão **trabalho** e o atributo cor_raca da dimensão **característica**.

Sobre o atributo cursoquefrequentada da dimensão **educacao** será selecionado somente o 3º grau ensino superior, obtendo-se o resultado exibido na figura 63.

	cor_raca	trabalha	soma_quantidadepessoas
▶	Amarela	1	2
	Amarela	0	1
	Branca	0	939
	Branca	1	1052
	Parda	0	191
	Parda	1	377
	Preta	1	76
	Preta	0	19

Fig.63 Pessoas cursando o 3º grau por raça e por estar ou não trabalhando

6.2.5 Quantidade de mulheres, por bairro, menores de 18 anos que não trabalham, não estudam e têm algum problema de saúde

Esta consulta envolve as dimensões **caracteristica**, **educacao**, **trabalho**, **setorcensitario** e **problemasaude**.

Para resolver essa consulta, deve-se aplicar filtros sobre a dimensão **característica** dizendo que o sexo é o feminino e que a idade é menor que 18 anos. Em **educacao** deve-se executar o filtro sobre o atributo estuda buscando somente aqueles que têm o valor falso. Em **trabalho** o slice deve ser aplicado sobre o atributo trabalha, dizendo que é falso.

Por fim, deve-se buscar na dimensão **problemasaude** apenas aquelas pessoas que possuem o atributo possuiproblemacronicosaude verdadeiro. Para exibição deve-se escolher o nomebairro da dimensão **setorcensitario** e quantidadepessoas da tabela fato, obtendo o resultado apresentado na figura 64.

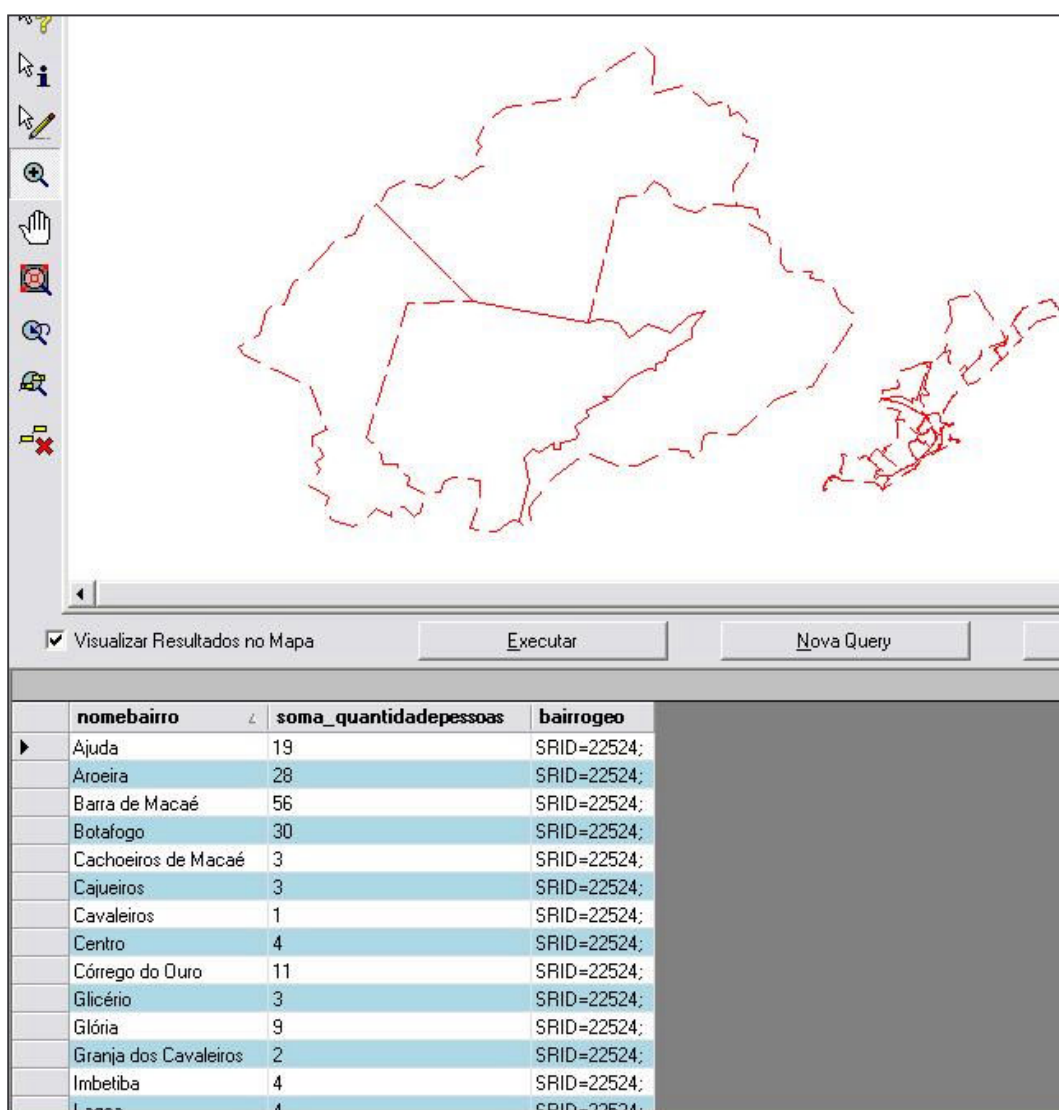


Fig.64 Quantidade de mulheres menores de 18 anos que não trabalham, não estudam e têm algum problema de saúde, por bairro

6.2.6 Número de pessoas que possuem deficiência física que estão fora do mercado de trabalho, por setor administrativo

Nesta consulta, devem ser consideradas as dimensões **deficienciafisica**, **trabalho**, **setorcensitario**, além da tabela fato **individuo**.

A dimensão **deficienciafisica** é somente para a seleção das pessoas que possuem algum tipo de deficiência física. A dimensão **trabalho** é somente para buscar as pessoas que estão fora do mercado de trabalho, ou seja, aquelas cujo atributo trabalha seja igual a falso.

O resultado exibido na figura 65 mostra que foram usadas a quantidadepessoas da tabela fato e o nomesetoradm da dimensão **setorcensitario** como respostas ao usuário.



	nomesetoradm	soma_quantidadepessoas
▶	Verde	276
	Cinza	45
	Amarelo	315
	Azul	92
	Rosa	324
	Laranja	72
	Marrom	254
	Vermelho	242
	Bege	14
*		

Fig.65 Número de pessoas que possuem deficiência física e não trabalham, por setor administrativo

6.2.7 Quantidade de pessoas que não têm saneamento na residência (fossa, esgoto) e apresentam problemas de saúde, por bairro.

Para responder a esta pergunta, deve-se buscar as informações nas dimensões **setorcensitario**, **condicaomoradia** e **problemasaude**.

A figura 66 mostra que foram usados para exibição os atributos quantidadepessoas da tabela fato e o atributo nomebairro da dimensão **setorcensitario** (agrupamento).

Para buscar as pessoas sem saneamento na residência, deve-se levar em conta os atributos tipoesgoto e fossa que devem ser “Sem Esgotamento” e falso, respectivamente. Além disso, deve-se levar em consideração que o atributo possuiproblemacronicosaude da dimensão **problemasaude** deve ser verdadeiro.

	nomebairro	soma_quantidadepessoas
▶	Ajuda	27
	Ároeira	30
	Barra de Macaé	140
	Botafogo	32
	Cachoeiros de Macaé	5
	Cajueiros	10
	Cavaleiros	2
	Centro	8
	Córrego do Ouro	6
	Glicério	8
	Glória	5
	Granja dos Cavaleiros	2
	Imbetiba	46
	Imbetiba	1

Fig.66 Quantidade de pessoas sem saneamento na residência, com problemas de saúde, por bairro

6.2.8 Quantidade de pessoas que não sabem ler e escrever que estão trabalhando, por bairro.

Na realização desta consulta só existem dois filtros: “não sabe ler” e “não estar trabalhando”. Para isso, basta informar na dimensão **educacao** que o atributo sabeler é falso e que na dimensão **trabalho** o atributo trabalha é verdadeiro.

Para exibir, além de quantidadepessoas de **individuo**, mostra-se o nomebairro da dimensão **setorcensitario**, conforme pode ser visto na figura 67.

	nomebairro	soma_quantidadepessoas
▶	Ajuda	84
	Ároeira	187
	Barra de Macaé	631
	Botafogo	331
	Cabiúnas	2
	Cachoeiros de Macaé	64
	Cajueiros	39
	Cavaleiros	6
	Centro	27
	Córrego do Ouro	173
	Glicério	148
	Glória	31
	Granja dos Cavaleiros	25
	Imbetiba	22

Fig.67 Pessoas que não sabem ler e escrever por bairro e trabalham

6.2.9 Quantidade de pessoas que não trabalham e declaram renda acima de 5 salários mínimos, por bairro

Esta consulta, que utiliza as dimensões **trabalho** e **setorcensitario**, tem como filtro o atributo rendimentomensalmenor maior que 5 e o atributo trabalha como falso. Para a exibição, são usados os atributos quantidadepessoas da tabela fato, nomebairro e bairrogeo da dimensão **setorcensitario**, conforme pode ser visto na figura 68.

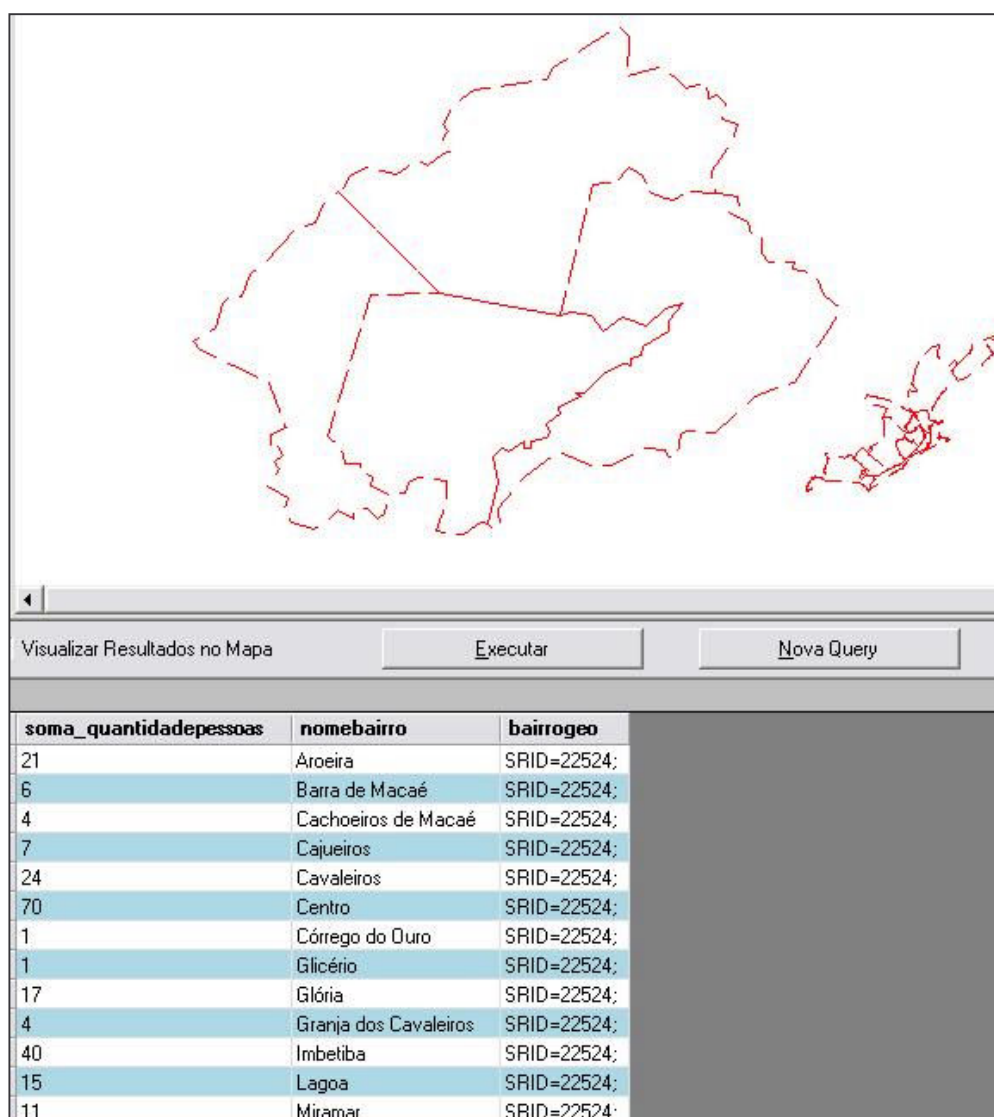


Fig.68 Quantidade de pessoas que não trabalham e declaram renda acima de 5 salários mínimos, por bairro

6.2.10 Quantidade de pessoas que moram no Setor Vermelho e utilizam os serviços públicos de saúde.

Para a realização desta consulta, serão necessárias as dimensões **setorcensitario**, **problemasaude** e **deficienciafisica**, além da tabela fato, para totalizar as pessoas.

No **setorcensitario** serão procurados somente aqueles que tiverem o nomesetoradm Vermelho. Para que se saiba quem recebe serviços públicos de saúde, tem-se que buscar esses dados nas duas dimensões que tratam desse assunto: **problemasaude** e **deficienciafisica**. Em **problemasaude** o atributo localacompanhamentomedico = “Hospital Público” ou “Posto/Centro de Saúde”. Em **deficienciafisica** o atributo tipoacompanhamentomedico igual a “Acompanhamento Público Permanente”.

Para uma maior comodidade na exibição dos dados, além de quantidadepessoas, também foi exibido o setoradministrativo, para que o setor administrativo Vermelho fosse exibido no mapa, conforme pode ser visto na figura 69.

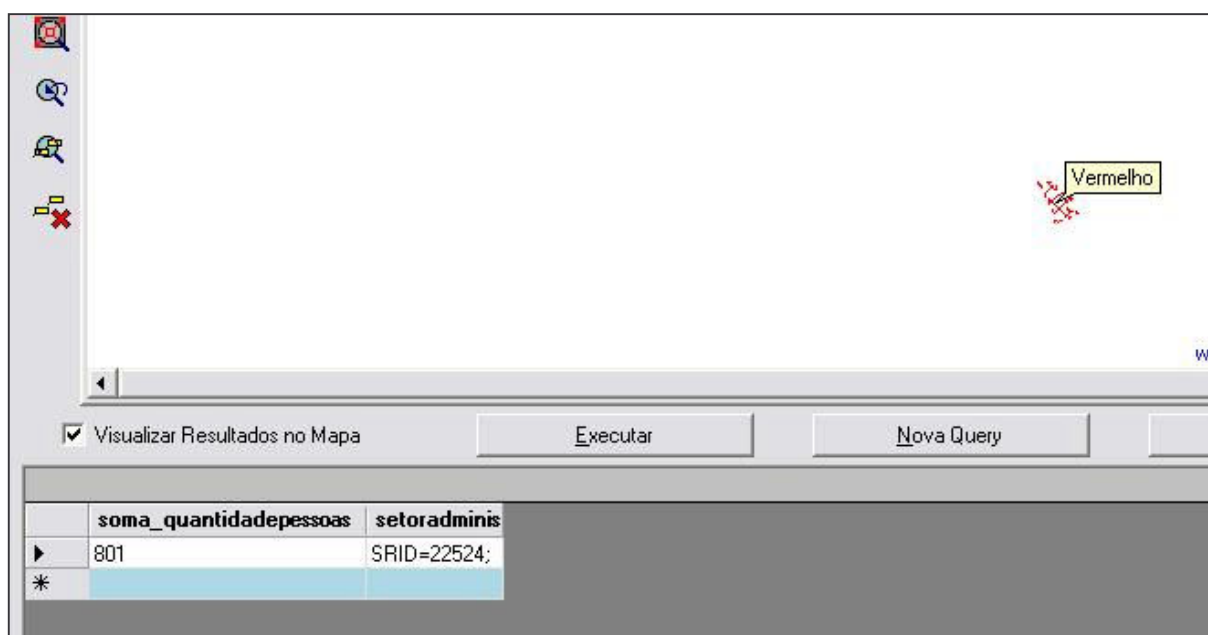


Fig.69 Quantidade de pessoas que moram no Setor Vermelho e utilizam os serviços públicos de saúde.

6.3 CONSIDERAÇÕES SOBRE OS TESTES REALIZADOS

Com a carga de dados atualmente disponível para a realização desta dissertação, não foi possível realizar operações sobre a dimensão **tempo**, uma vez que haveria a necessidade de efetuar pelo menos duas cargas de dados para que pudessem ser realizadas as operações de comparação ao longo do tempo. As comparações no tempo permitem enxergar melhorias ou pioras sobre a população, dando uma maior potencialidade nas consultas do sistema.

É importante salientar que o sistema é de extrema prática no que se refere ao uso do administrador municipal. Para por este sistema em prática, basta dar ao(s) administrador(es) municipal um treinamento que o deixe envolvido com as nomenclaturas usadas nos atributos e também para o aprendizado no uso do PostGeoOLAP.

Uma dificuldade no emprego desta ferramenta está na migração dos dados, uma vez que qualquer alteração realizada nos dados do Projeto Macaé Cidadão repercutirá em mudanças na ferramenta de migração dos dados. Mas esse é um trabalho transparente ao usuário final. O importante, neste tipo de ferramenta é a facilidade para este usuário do sistema que não necessita ter conhecimentos profundos de informática para ter as suas informações em mãos.

As consultas realizadas neste capítulo destinam-se somente para a validação do sistema desenvolvido. Espera-se que, com o uso diário, os usuários apresentem mais necessidades que um sistema com tamanha flexibilidade é capaz de lhes oferecer.

7 CONCLUSÃO

O município de Macaé se tornou nos últimos tempos, um dos lugares de maior importância da região Norte Fluminense, principalmente pela atividade de exploração de petróleo. Verifica-se amplamente que o município apresenta-se em constante crescimento ao longo do tempo.

Por conta desse desenvolvimento, a sua prefeitura vem tentando realizar trabalhos em favor do município e de seus munícipes. Dentre esses trabalhos, pode-se destacar o Projeto Macaé Cidadão que é uma pesquisa realizada com toda população macaense, tentando buscar informações sobre cada uma pessoa sobre diversos aspectos. Um outro projeto de grande importância é o Projeto GeoMacaé, preocupado com o georreferenciamento e com a cartografia digital do município.

Este trabalho se propôs a utilizar o Projetos Macaé Cidadão e GeoMacaé, além de dados vindos dos diversos núcleos administrativos do município, a fim de gerar um Sistema de Suporte a Decisão baseado nas tecnologias de Datawarehouse, OLAP e SIG. Desta forma, dando uma ferramenta mais acessível, simples que produzisse resultados rápidos para a melhor tomada de decisão dos gestores municipais.

Para o desenvolvimento desta ferramenta, foi utilizada uma modelagem baseada em objetos capaz de representar tanto os dados convencionais quanto os dados geográficos em um único ambiente. Desta forma foi possível partir do modelo conceitual, passar pelo modelo físico e chegar ao banco de dados sem modificações drásticas.

Um conceito da orientação a objetos que foi aproveitado no desenvolvimento desta dissertação foi o emprego de estereótipos na representação dos tipos geográficos instanciáveis preconizado pelo OpenGIS.

O banco de dados utilizado para o armazenamento do Datawarehouse foi o PostgreSQL que é um excelente SGBD, de distribuição gratuita e que tem como possibilidade a integração do PostGIS que se constitui de um conjunto de recursos para operações geográficas sobre o PostgreSQL baseado no Consórcio OpenGIS.

Para a migração dos dados das bases disponibilizadas pela prefeitura de Macaé foi desenvolvida neste trabalho uma ferramenta capaz de ler os dados provenientes da base do Projeto Macaé Cidadão e gravá-la no modelo multidimensional no PostgreSQL. As informações geográficas foram importadas dos mapas disponibilizados pelo Projeto GeoMacaé no formato ArcView Shape File. Os dados das dimensões geográficas não agregáveis foram migrados manualmente da base de dados disponibilizada no formato Microsoft Excel. Seus dados foram georreferenciados manualmente, tendo-se como preocupação de certificar-se sobre sua inserção nos setores censitários corretos.

Para a realização das consultas sobre o modelo multidimensional, foi utilizado o software PostGeoOlap, que é uma ferramenta de distribuição gratuita capaz de realizar consultas OLAP integrado a operações GIS sobre um Datawarehouse armazenado em PostgreSQL.

Uma vez construído o Datawarehouse e configurada a ferramenta OLAP e SIG, foram realizadas consultas baseadas em perguntas feitas por representantes da prefeitura de Macaé. O uso destas perguntas teve como propósito a valorização da proposta da ferramenta: responder on-line a questões sobre diversos aspectos do município de Macaé. Respondendo essas questões o gestor municipal pode desenvolver um trabalho de maior qualidade, atendendo de forma muito mais efetiva às necessidades dos seus munícipes, além de ter também um instrumento dinâmico de controle de suas Unidades ou Órgãos.

7.1 CONTRIBUIÇÕES DO TRABALHO

Espera-se que a pesquisa possa apresentar diversas contribuições, dentre as quais pode-se destacar:

- Os benefícios para o município de Macaé, em particular, que terá nas mãos de seus administradores uma ferramenta capaz de responder muitas das perguntas das quais eles necessitavam de respostas de forma rápida, prática e eficiente.
- Mostrar que um Sistema de Suporte a Decisão tem grande importância no âmbito municipal, principalmente na organização de informações “espalhadas”, para que decisões a favor do município e de seus habitantes sejam tomadas de forma mais inteligente, precisa e consciente.
- A utilização de ferramentas de distribuição gratuita no desenvolvimento de um Sistema de Suporte à Decisão, tentando colocar essa tecnologia cada vez mais disponível para as mais diversas empresas sejam de pequeno ou de grande porte, pública ou privada.
- A integração de ferramentas de suporte a decisão, dando um valor muito maior à informação. SIG, OLAP e Datawarehouse juntos, podem trazer informações mais precisas e completas ao tomador de decisão.
- A utilização da modelagem orientada a objetos para acompanhar todo o processo de desenvolvimento de um Datawarehouse de forma prática e sem dificuldades na leitura do modelo.
- O georreferenciamento das escolas e unidades de saúde do município que até o momento da realização deste trabalho ainda não estava disponível pelo Projeto Geo-Macaé.
- O uso de uma ferramenta de interface amigável para o usuário que não tem necessidade de conhecimento tecnológico para a busca de informação que até então seria de difícil acesso e baseada em conhecimento profundo de informática.

7.2 SUGESTÕES PARA TRABALHOS FUTUROS

Seguem aqui algumas sugestões de trabalhos a serem desenvolvidos para um melhor desenvolvimento do trabalho para a Prefeitura de Macaé:

- O desenvolvimento de um SSD usando os dados disponibilizados pelo IBGE, de forma a gerar um modelo genérico acessível a qualquer município, independente de realizações de pesquisas locais, reduzindo o custo de desenvolvimento. Em outras palavras, é possível desenvolver um SSD para qualquer município, utilizando-se somente os dados censitários disponibilizados pelo IBGE.

- O desenvolvimento de Datamining capaz de descobrir informações que não estão tão claras, ou seja, fazer com que o sistema seja capaz de buscar informações baseadas em algoritmos que conseguissem de alguma forma classificar a população ou procurar tendências baseados nos dados.

- Desenvolvimento de uma ferramenta capaz de validar os dados trazidos da Prefeitura de Macaé, a fim de encontrar inconsistências ocorridas talvez no período de digitação das informações principalmente do Programa Macaé Cidadão.

- Inclusão de mais dados geográficos, tais como rios e estradas, a fim de buscar informações mais precisas de distância, saneamento básico, saúde, etc.

- O desenvolvimento de uma ferramenta OLAP capaz de trazer dados principalmente estatísticos sobre o município de Macaé, como por exemplo, o percentual de pessoas adultas que não estudam ou são analfabetas.

- Transformação de dados armazenados em papel ou outros formatos não digitais que se encontram nas diversas unidades administrativas, de forma a trazer mais informações para o tomador de decisão.

8 REFERÊNCIAS

ARONOFF, S. *Geographic Information Systems: A Management Perspective*. Ottawa, Canadá: WDL Publications, 1995.

BUEHLER, K.; MCKEE, L. *The OpenGIS Reference*. Massachusetts, USA. Disponível em: <http://www.opengis.org/info/orm/03-040.doc>. Acessado em: set 2003.

CAMPOS, Maria Luiza; ROCHA FILHO, Arnaldo V. *Data Warehouse*. Disponível em <<http://genesis.nce.ufrj.br/dataaware/tutorial/home.html>>. Acessado em 13/02/2004.

COLONESE, Giovanni; TANAKA, Astério K.; CARVALHO, Rogério A. *Mapeando Modelos Conceituais Dimensionais OO para Modelos Lógicos Dimensionais Relacionais*. (Artigo aceito para publicação nos anais do CACIC 2003, Argentina).

COLONESE, Giovanni. *Uma Ferramenta aberta de desenvolvimento integrado de sistemas de informação para processamento analítico e geográfico*. Dissertação de Mestrado. Campos dos Goytacazes, Universidade Cândido Mendes, 2004.

COME, G. *Contribuição ao Estudo da Implementação de Data Warehousing: Um Caso no Setor de Telecomunicações*. Dissertação de Mestrado. São Paulo: FEA/USP, 2001.

DORSEY, Paul. *UML Class Models as a Data Modeling Tool*. Disponível em <<http://www.datawarehouse.com/iknowledge/articles/article.cfm? ContentID=2685>>. Acessado em 10 de maio de 2004.

DIAS, Taciana de Lemos; OLIVEIRA, Maria da Piedade Gomes de; CÂMARA, Gilberto; CARVALHO, Marília de Sá. *Problemas de Escala e a Relação Área-Indivíduo em Análise Espacial de Dados Censitários*. *Informática Pública*, Ano 4, Número 1, 2002. Disponível em: <<http://www.ip.pbh.gov.br/sum0401.html>>. Acesso em 23 de maio de 2004.

ESRI. *ESRI Shape File Technical Description. An ESRI White Paper*. Julho de 1998. Disponível em: <www.esri.com/library/whitepapers/pdfs/shapefile.pdf>. Acesso em 23 de maio de 2004.

FERREIRA, A. C. F. *Um Modelo para Suporte à Integração de Análises Multidimensionais e Espaciais*. Dissertação de Mestrado. Rio de Janeiro, UFRJ/IM/NCE, 2002.

FIDALGO, R.N.; TIMES, V. C.; SOUZA, F. F. *GOLAPA: Uma Arquitetura Aberta e extensível para Integração entre SIG e OLAP*. *GeoInfo 2001*, III Workshop Brasileiro de Geoinformática. p.111-118. Instituto Militar de Engenharia, Rio de Janeiro. 4 e 5 de outubro de 2001.

FOWLER, Martin; SCOTT, Kendall. *UML Essencial: Um breve guia para a linguagem-padrão de modelagem de objetos*. Bookman, 2002. Segunda Edição.

KOUBA, Z.; MATOUSEK, K.; MIKSOVSKY P. *On Data Warehouse and GIS Integration*. In *Proceedings of DEXA2000*. Greenwich, Inglaterra: DEXA2000, 2000.

LISBOA, Jugurta; PEREIRA, Marconi de Arruda. *Desenvolvimento de uma Ferramenta CASE para o Modelo UML-GeoFrame com Suporte para Padrões de Análise*. IV Simpósio Brasileiro de Geoinformática 2002, Caxambu, MG.

INMON, W.H. *Building the Data Warehouse*. New York, USA: Wiley, 1992. Terceira Edição.

INPE – INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. *Spring – Sistema de Processamento de Informações Georeferenciadas*. Disponível em <<http://www.dpi.inpe.br/spring/portugues/index.html>>. Acesso em 10/08/2003.

KIMBALL, R. *Data Warehouse Toolkit: Técnicas para construção de Datawarehouses Dimensionais*. São Paulo: Makhron Books, 1998.

NAVATHE, S. B.; ELMASRI, R. *Fundamentals of Database Systems*. USA: Addison-Wesley, 2001.

OGC: *Open GIS Consortium*. Disponível em <<http://www.opengis.org>>. Acessado em 17/10/2003.

OGC. *OpenGIS Simple Specifications for SQL Revision 1.1*. OpenGis Project Document 99-049. Publicado em 05/05/1999.

PLANETGIS. *Documentação do software PlanetGIS, versão 2.29*. Disponível em <<http://www.planetgis.co.za/>>. Acessado em 20/10/2003.

POSTGIS. *Documentação da extensão espacial PostGIS ao PostgreSQL, versão 0.8*. Disponível em <<http://postgis.refrations.net/>>. Acessado em 05/02/2004.

RAMSEY, Paul. *PostGis Manual*. Disponível em <<http://postgis.refrations.net>>. Acessado em 05/02/2004.

POSTGRESQL.ORG. *Documentação do SGBD PostgreSQL 7.4*. Disponível em: <<http://www.postgresql.org>>. Acessado em 05/02/2004.

PREFEITURA MUNICIPAL DE MACAÉ. *Prefeitura de Macaé*. Disponível em <<http://www.macaerj.gov.br/municipio>>. Acessado em 25/07/2003.

ROCHA, C. H. B. *Geoprocessamento: tecnologia transdisciplinar*. Juiz de Fora, MG: Ed. do Autor, 2000.

SHEKHAR et al. *Map Cube: A visualization Tool for Spatial Data Warehouse*. Disponível em <<http://www.cs.umn.edu/research/shashi-group/mapcube.htm>>. Acesso em 20/01/2004.

STEFANOVIC, N. *Design and Implementation of On-Line Analytical Processing (OLAP) of Spatial Data*. Dissertação de Mestrado. Disponível em <<http://gunther.smeal.psu.edu/3070.html>>. Universidade de Belgrado, 1997. Acesso em 20/09/2003.

YOURDON, Edward. *Análise Estruturada Moderna*. Editora Campus, 1990.