

Uma Ferramenta em Software Livre para o Desenvolvimento de Sistemas de Suporte à Decisão

Giovanni Colonese¹, Rodrigo Soares Manhães^{1,2}, Sahudy Montenegro González², Alan Carvalho Galante¹, Rogério Atem de Carvalho³, Asterio Kiyoshi Tanaka⁴

¹Faculdade Salesiana Maria Auxiliadora (FSMA)

²Universidade Candido Mendes – Campos (UCAM-Campos)

³Centro Federal de Educação Tecnológica de Campos (CEFET-Campos)

⁴Universidade Federal do Estado do Rio de Janeiro (UNIRIO)

colonese@lagosnet.com.br, rmanhaes@cefetcampos.br,
alancarvalho@gmail.com, sahudy@ucam-campos.br,
ratem@cefetcampos.br, tanaka@uniriotec.br

Resumo. *Este trabalho descreve PostGeoOlap, uma ferramenta livre e de código aberto para suporte à decisão que integra tecnologias OLAP (On-Line Analytical Processing) e SIG (Sistemas de Informação Geográfica) em uma única aplicação. Além de descrever a ferramenta e seu ambiente, procurou-se demonstrar como esta pode ser empregada na obtenção de um efetivo suporte à decisão sobre uma plataforma de baixo custo, adequada para uso em pequenas e médias empresas e unidades públicas de pequeno porte.*

1. Introdução

As tecnologias da chamada terceira revolução industrial, a revolução da microeletrônica iniciada com a invenção do transistor, têm transformado a face do planeta. Embora diferentes autores discordem das suas conseqüências, é consenso o entendimento de que a generalização do microcomputador e o avanço das telecomunicações impulsionam o mundo a um novo patamar [1].

Neste contexto, a informação adquire um imenso valor, seja para, em se tratando de empresas, estar em melhores condições para a concorrência com seus pares, seja para, falando-se do setor público ou organizações sem fins lucrativos, melhor desempenhar suas tarefas e atender as demandas de seu público-alvo. Assim, percebe-se que a informação está, cada vez mais, moldando a maneira como a sociedade trabalha, produz e pensa [1].

Especificamente no que tange ao setor de tecnologia da informação, as diversas instituições têm acumulado grandes quantidades de dados em seus sistemas. Estes dados, porém, para significar algo mais do que meros cadastros e registros operacionais, precisam tomar a forma, mais útil e adequada, de informação. Para isto, surgiram os Sistemas de Suporte à Decisão, que integram tecnologias como data warehousing, data mining, OLAP (On-Line Analytical Processing), SIG (Sistemas de Informação Geográfica), KDD (Knowledge Discovery in Databases) e IR (Information Retrieval), entre outras.

O objetivo final destas tecnologias é permitir ao decisor, normalmente alguém que atua nos níveis estratégico ou tático de uma instituição, visualizar, sob várias perspectivas, diversos tópicos pertinentes à sua área de atuação. A capacidade de análise de dados agregados e integrados a partir de diversas fontes faz do data warehouse (DW) e de uma aplicação OLAP que lhe forneça processamento analítico ferramentas indispensáveis aos decisores, permitindo observações e análises em tempo real do domínio de atuação sob diversas perspectivas. A estas tecnologias, somam-se os SIGs, que utilizam mapas para prover a seus usuários a análise de

dados que possuam atributos referentes ao posicionamento espacial, algo praticamente indispensável para quem possui alguma dimensão espacial em seu domínio de atuação [2].

Inicialmente tecnologias distintas, OLAP/DW e SIG têm, nos últimos anos, sido alvo de esforços de inúmeros pesquisadores, empreendendo trabalhos no sentido de sua integração e unificação, de forma a que os Sistemas de Suporte à Decisão possam contar, de modo unívoco, com funcionalidades de processamento analítico sobre dados convencionais e geográficos.

Neste contexto, propomos uma ferramenta livre de suporte à decisão, chamada PostGeoOlap, que trate as tecnologias OLAP e SIG de forma integrada desde o início. PostGeoOlap faz parte do projeto GeoOlap, que propõe uma técnica para modelagem de aplicações desde sua concepção inicial, nas quais a coexistência entre conceitos espaciais e temporais é essencial. O objetivo da ferramenta PostGeoOlap é permitir a construção de “sistemas de suporte à decisão onde, desde a fase de modelagem conceitual, seja possível tratar aspectos analíticos e geográficos” [3]. Assim, PostGeoOlap é uma ferramenta que realiza consultas analíticas sobre um *data warehouse* contendo dimensões com atributos geográficos, não contemplando operações OLAP automatizadas.

1.1. Trabalhos Relacionados

Há vários trabalhos relacionados ao projeto GeoOlap e à ferramenta PostGeoOlap. Estes trabalhos possuem diferentes abordagens em relação à integração de tecnologias OLAP e SIG.

O projeto GeoMiner [4] permite operações OLAP em cubos com dados georeferenciados e o projeto MapCube [5] propõe operações sobre um cubo que podem retornar mapas. Ambas as propostas prevêm processamento analítico de mapas, mas sem a utilização de um SIG.

Os projetos GOAL (Geographical Information On-Line Analysis) [6], SIGOLAP [7] e GOLAPA (Geographic On-Line Analytical Processing Architecture) [8] não utilizam um modelo unificado com conceitos analíticos e geográficos, mas tratam estas duas tecnologias separadamente, propondo módulos de integração que mapeiam requisições e dados de um lado a outro.

Os trabalhos em [9], [10] e [11] são os mais próximos à abordagem trazida neste artigo (embora nós não consideremos atributos geográficos nas medidas), mas não propõem qualquer técnica de modelagem de sistemas como um todo, desde seu nível conceitual, como faz o projeto GeoOlap.

1.1. Motivação

Como é fato que o custo de aquisição ou licenciamento de sistemas de suporte à decisão como aplicações OLAP ou SIG é proibitivo a pequenas e médias empresas, a presente iniciativa torna acessíveis a um público muito maior estas tecnologias, normalmente restritas à esfera das grandes organizações. No setor público, cujos custos com tecnologia da informação em boa parcela são creditados a licenciamentos de *software*, também possui grande importância o uso do Software Livre, no sentido da economia de recursos, da independência de fornecedores e, principalmente, da melhoria das condições sociais e da qualidade de vida das populações, seguramente resultantes da honesta utilização das citadas tecnologias de suporte à decisão.

É importante frisar que principalmente em países pobres ou ditos “em desenvolvimento” como o Brasil, é ainda mais importante a adoção de *software* sem restrições de uso ou distribuição e de código aberto, como instrumento para a democratização da tecnologia e da informação [12, 13].

A configuração atual da economia mundial tem por característica a concentração da tecnologia na esfera das grandes empresas e corporações, até que esta se torne obsoleta e, só então, seja repassada às pequenas e médias empresas e aos países ditos em desenvolvimento.

Neste sentido, o Software Livre é um instrumento de democratização tecnológica, indo na contramão destas tendências centralizadoras, na medida em que permite que o acesso às

tecnologias mais recentes seja simultâneo para pequenas e grandes organizações, para países pobres e ricos.

Num contexto mais localizado, pode-se facilmente vislumbrar a imensa gama de projetos envolvendo suporte à decisão baseado em OLAP e SIG que empresas e instituições que antes não possuíam acesso a tais tecnologias, principalmente devido aos custos proibitivos de licenciamento, poderão desenvolver com o apoio do PostGeoOlap

A ferramenta vem sendo utilizada em uma empresa distribuidora de publicações de editoras nacionais, em pontos de venda no Norte Fluminense, e, no contexto de gestão municipal, em aplicações espaciais nos municípios de Macaé – RJ [14], e Cachoeiro do Itapemirim – ES. Outras aplicações estão sendo consideradas e propostas, inclusive internacionalmente, a partir da inclusão do projeto no PGFoundry, a incubadora de projetos do SGBD PostGreSQL, no endereço <http://postgeoolap.projects.postgresql.org>.

2. Descrição da Ferramenta

PostGeoOlap é uma ferramenta para a criação de soluções OLAP espaciais sobre o SGBD PostGreSQL e o PostGIS, sua extensão espacial. O nome PostGeoOlap tem origem na integração entre funcionalidades geográficas, tecnologia OLAP e PostGreSQL.

2.1. Princípios de Projeto

Para a construção da ferramenta PostGeoOlap, adotou-se o paradigma de Software Livre como princípio de projeto. Assim, todas as APIs, frameworks e softwares de persistência de dados utilizados neste projeto são open-source. Deste modo, pode-se atingir um dos objetivos do projeto, que é prover uma plataforma de baixo custo para OLAP espacial, permitindo a pequenas e médias organizações o desenvolvimento de aplicações de data warehousing e SIG, o que outrora, lhes seria proibido devido aos altíssimos custos de licenciamento de soluções proprietárias para estes fins.

Para o projeto de PostGeoOLAP adotou-se ROLAP (Relational OLAP) como modelo de armazenamento do data warehousing para obter as vantagens da maturidade da tecnologia de SGBDs objeto-relacionais, que possuem a capacidade de armazenar tanto dados convencionais quanto geográficos e usar funções de agregação e funções espaciais sobre eles, além de permitir a adição customizada de funcionalidades. Assim, ambas as pesquisas, geográficas ou analíticas convencionais, serão processadas e respondidas pelo SGBD PostGreSQL e todos os dados, desde o nível base até as agregações, serão mantidos no modelo relacional.

2.2. Modelo e Arquitetura

O diagrama de classes na Figura 1 representa os metadados utilizados pelo PostGeoOlap para a manutenção dos dados no data warehouse. A classe Esquema representa o banco de dados (database) no PostGreSQL que contém o data warehouse de interesse. A classe Cubo representa um hipercubo, ou, no nível do usuário, cada perspectiva de negócio pela qual um Esquema pode ser analisado. A classe Tabela representa todas as relações existentes no data warehouse. A classe Dimensao é uma subclasse de Tabela e se refere a todos os componentes de um cubo, tanto tabelas fato quanto dimensões. A classe Atributo representa os dados existentes em cada Tabela. O auto-relacionamento indica que um atributo convencional está funcionando como rótulo para um ou mais atributos geográficos. A classe Agregacao representa as várias hierarquias para os atributos de uma dimensão. A classe ItemHierarquia aloca cada Atributo a uma Hierarquia.

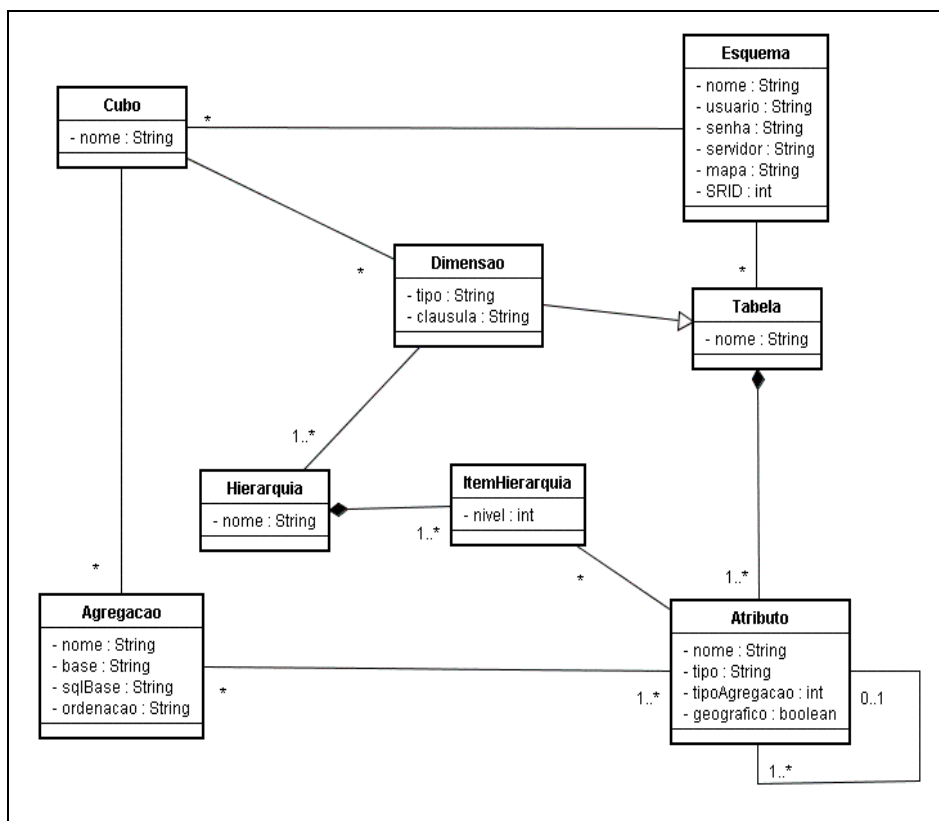


Figura 1. Modelo conceitual do PostGeoOlap.

O SGBD PostGreSQL é utilizado pelo PostGeoOlap para a realização das funções convencionais de agregação padrão SQL (e.g. sum, max, min, avg, count) e todas as funções geográficas definidas pelo Open Geospatial Consortium, nas especificações OpenGIS [15].

2.2 Funcionalidades

Os casos de uso do PostGeoOlap estão resumidos na Tabela 1. Os casos de uso *Criar esquema*, *Criar cubo* e *Adicionar dimensão* trabalham na construção da estrutura dos cubos que serão verificados e processados pelo PostGeoOlap.

Após a definição do esquema e do cubo, com a adição das dimensões, a ferramenta está pronta para o processamento do cubo. Nesta operação, explicitada pelo caso de uso *Processar cubo*, o desempenho das pesquisas no cubo é inferido e, caso caia para abaixo de um patamar mínimo pré-estabelecido, a aplicação cria uma nova estrutura de agregação representada por uma tabela. A agregação é realizada em três passos:

- criação da tabela que conterá os dados agregados;
- transferência dos dados agregados para a nova tabela;
- e criação dos índices para a nova tabela, utilizando estruturas de árvore-B para atributos convencionais e GiST (*Generalized Search Tree*) [16] para os atributos espaciais. Ambas as estruturas de indexação são oferecidas pelo PostGreSQL.

CASO DE USO	DESCRIÇÃO
Criar esquema	Cria uma conexão com um banco de dados (<i>database</i>) no PostgreSQL e persiste os metadados necessários.
Criar cubo	Cria um cubo dentro de um esquema, selecionando uma tabela fato e definindo seus itens numéricos e as operações desejadas sobre estes itens.
Adicionar dimensão	Cria uma perspectiva para a análise dos dados contidos na tabela fato previamente definida, selecionando uma das tabelas relacionadas ao esquema. Também define a hierarquia dos atributos da dimensão, atribuindo um nível numérico para cada um.
Processar cubo	Verifica a massa de dados referenciada pelo cubo e tenta inferir o desempenho de consultas em tempo de execução. As consultas avaliadas como tendo um baixo desempenho são otimizadas através de agregações. Isto permite que o cubo possa ser analisado através de qualquer das dimensões, dentro de um tempo de resposta aceitável.
Adicionar dimensão não agregável	Cria uma dimensão que, apesar de não constituir uma perspectiva de análise e, assim, não participando do processamento do cubo, serve como referência para comparações geográficas com outras dimensões que possuam atributos geográficos.

Analisar dados	Oferece uma interface visual que permita a seleção de atributos para consulta utilizando restrições convencionais e/ou geográficas. Os resultados são visualizados tanto na forma tabular quanto na forma cartográfica, caso necessário.
----------------	--

Tabela 1. Lista de casos de uso do PostGeoOlap

2.3. Sistema Gerenciador de Banco de Dados utilizado

O SGBD utilizado pelo PostGeoOlap é o PostgreSQL, que 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 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 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.

Na verdade, o PostgreSQL não possui nativamente todos os tipos e operações geográficas contempladas pelo OpenGIS, que é uma organização internacional responsável pela criação de novas padronizações técnicas e comerciais para garantir interoperabilidade em SIG, buscando 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.

Esses tipos e operações são implementados através de uma extensão do PostgreSQL conhecida por PostGIS. Logo, este implementa as definições do OpenGIS para a SQL, permitindo desta forma que a SQL tenha funções sobre tipos geográficos.

2.4. Visualização e Tratamento de Mapas

Para a visualização e tratamento de mapas e dados geográficos, foi necessária a inclusão, na ferramenta, de algum tipo de biblioteca ou componente que pudesse realizar estas funções. JUMP (acrônimo recursivo para JUMP Unified Mapping Platform) é uma aplicação GUI (Graphical User Interface) para visualização e processamento de dados espaciais, incorporando funcionalidades existentes em muitos outros SIGs direcionados a análise e manipulação de dados geoespaciais. O JUMP também oferece um framework altamente extensível para o desenvolvimento e a execução de aplicações customizadas de processamento de dados espaciais.

Esta última característica interessa diretamente ao desenvolvimento do projeto PostGeoOlap, pois neste se utilizam classes de visualização e processamento de dados geográficos fornecidas pelo framework JUMP. O JUMP dispõe, assim, de uma API (Application Programming Interface) que possibilita pleno acesso, via programação, a todas as funções, visualização de mapas e operações espaciais. As consultas espaciais realizadas pelo PostGeoOlap, deste modo, utilizam-se de classes de visualização e processamento do JUMP para apresentar os resultados, cartograficamente, ao usuário.

O JUMP é inteiramente escrito na linguagem Java, sendo distribuído através da licença GPL (General Public License), o que o torna também aderente aos conceitos do Software Livre. Em termos de compatibilidade, JUMP é aderente às especificações do OpenGIS, através do uso de classes de um co-projeto, o JTS Topology Suite (JTS), que fornece funcionalidades de predicados espaciais, operações topológicas, e outras características, implementando a especificação OpenGIS de Feições Simples para SQL [15].

O JUMP implementa, também, um vasto conjunto de funcionalidades e as oferece inteiramente ao acesso programático, tais como rotulação por atributo com resolução de conflitos, colorização por atributo, estatísticas por camada, estatísticas por feição, operações geométricas como interseção, união, diferença e diferença simétrica, sobreposição de mapas, validação geométrica e outras.

Posteriormente, a empresa responsável pelo desenvolvimento do JUMP descontinuou o projeto, que, por ser Software Livre, pôde ter seu desenvolvimento levado adiante pela comunidade, sob o nome de OpenJUMP.

3. Estudo de Caso

Nesta seção será apresentado um estudo de caso de aplicação do modelo de integração e da ferramenta propostos no presente trabalho, com vistas a produzir “do zero” um novo sistema de suporte à decisão que integre funcionalidades analíticas e geográficas, desde o nível mais alto de concepção, destinado ao auxílio à tomada de decisão nos setores de Educação e Saúde da Prefeitura e Macaé.

O município de Macaé está localizado na região Norte do estado do Rio de Janeiro, distando 182 quilômetros da capital. Este município 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.

Localizado 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. O Projeto Geo-Macaé

Para o desenvolvimento do SSD, foram usados os dados das Unidades Administrativas do município, do Projeto Macaé Cidadão e do Projeto Geo-Macaé.

Quanto às áreas de atuação ou de interesse imediato foram selecionadas a Saúde e a Educação, por terem sido julgadas como as áreas de maior interesse pela 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, também foram usados os cadastros mobiliário e imobiliário que também foram retirados do Projeto Macaé Cidadão. Todas essas informações foram 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.

3.2. A divisão geográfica do Município

A menor unidade de análise a ser usada neste trabalho é 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. Uma visão da estratégia para a modelagem do espaço geográfico macaense é apresentada na figura 1.

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 à cerca da população macaense. Pelos dados informados pela prefeitura, o município de Macaé possui hoje 27 bairros.

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 dos quais existem 10 setores, nomeados por cores.

3.3. Modelagem do SSD

Nesta seção, será mostrada a modelagem do Sistema de Suporte à Decisão.

3.3.1. Modelando as dimensões geográficas agregáveis

Entende-se por dimensão geográfica agregável toda dimensão do modelo multidimensional do data warehouse que possui ligação direta com a tabela fato e, além disso, possui algum atributo geográfico.

Usando a UML, as três camadas geográficas são representadas conforme a Figura 2, onde um Bairro é composto de vários setores censitários, e um setor administrativo composto de vários bairros. Conforme o tratamento usado no Projeto Geo-Macaé, consideram-se todas as três camadas como figuras geométricas do tipo polígono. Além disso, usa-se o conceito de estereótipos para representar o tipo de figura geométrica que cada classe representa.

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 um setorcensitario.

3.3.2 Modelando as dimensões geográficas não agregáveis

Consideram-se como classes geográficas não agregáveis as classes que não têm ligação direta com a tabela fato do modelo multidimensional. 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.

No modelo usado para o DW implementado para a Prefeitura de Macaé, foram usadas as seguintes classes geográficas não agregáveis: escola e UnidadeSaúde. 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.

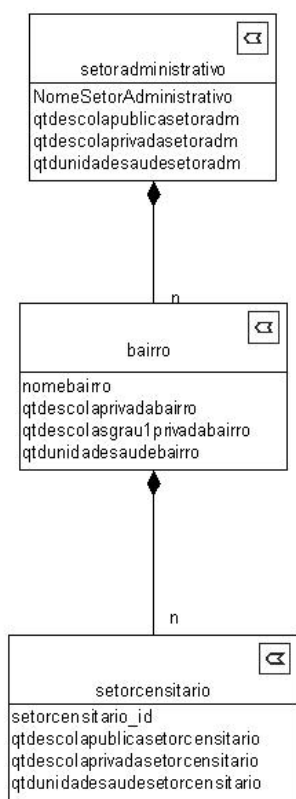


Figura 2. Representação com UML do espaço geográfico do município de Macaé

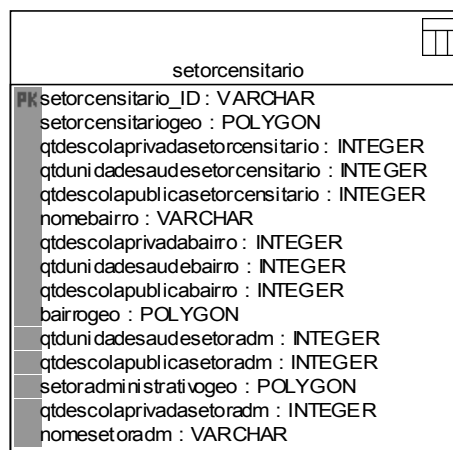


Figura 3. Modelo físico das classes geográfico

3.3.3 Modelando as dimensões não geográficas agregáveis

Além das dimensões geográficas agregáveis, as dimensões geográficas não agregáveis e a tabela fato, existem as dimensões que estão ligadas à tabela fato e que não possui nenhuma relação direta com o espaço geográfico. Essas dimensões serão explicadas a seguir.

A dimensão condicaomoradia serve para exibir as condições nas quais o cidadão está morando no município, tais como a moradia possui filtro de água, saneamento básico, etc. A dimensão tempo, pertinente à praticamente todos os DW, 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. A dimensão trabalho exibe as informações do cidadão no que se refere a sua característica trabalhista, tais como: se trabalha, se é funcionário público, etc. A dimensão educacao, abrange características do ponto de vista de estudos do cidadão, tais como se ele estuda, qual o grau de escolaridade, etc. A dimensão queixasistemaude, exibe quais as principais queixas que um cidadão possui em relação à área de saúde do município. Na dimensão problemasaude, estão representadas as informações referentes aos problemas de saúde dos munícipes. Na dimensão deficienciafisica, estão representados os dados referentes as deficiências físicas mais comuns entre os munícipes.

3.3.4 O Modelo Multidimensional

A figura 4 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, que corresponde a um contador, quando da ocorrência de indivíduos com as mesmas características.

3.3.5 A preparação do cubo OLAP

O próximo passo é a configuração do cubo OLAP na ferramenta PostGeoOLAP, que foi 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 é denominada “hierarquização”.

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 “quantidadepessoas”. 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.

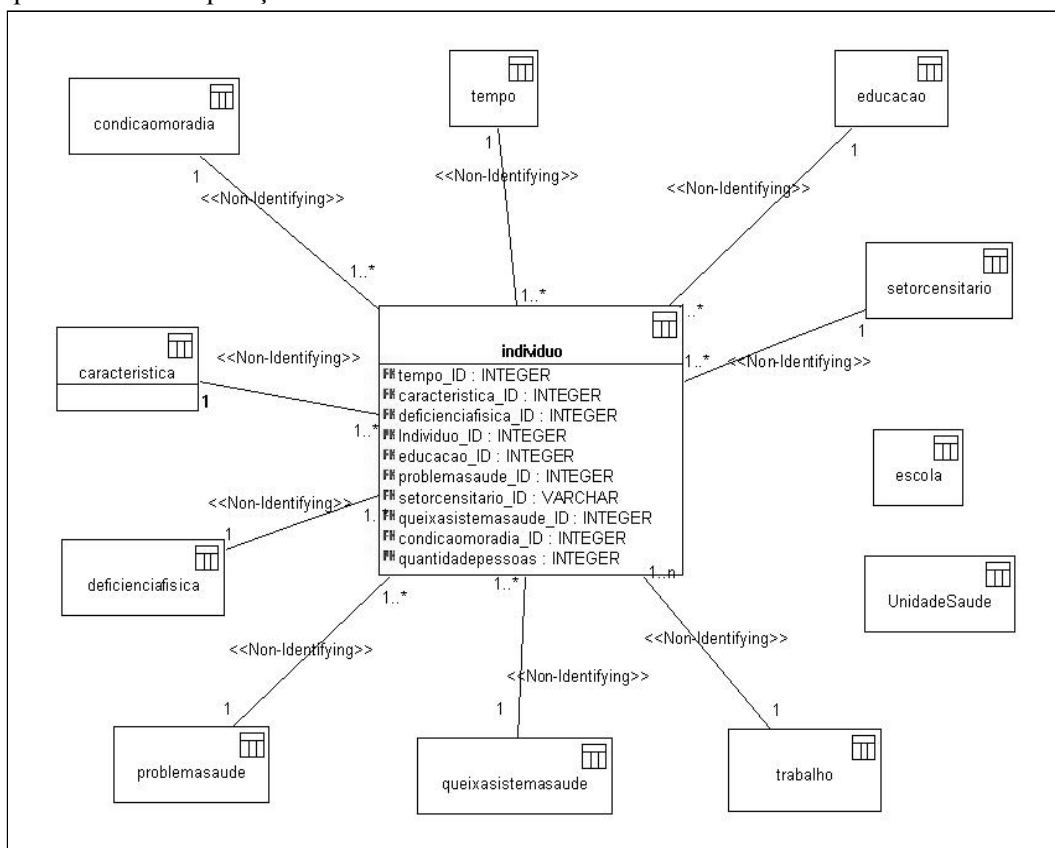


Figura 4. Modelo Multidimensional Físico

Depois de definida a tabela fato, passou-se a definir as dimensões agregáveis, juntamente com as hierarquias internas de cada dimensão. Por exemplo, na dimensão setor censitario foram definidos os atributos relativos a setor administrativo como de nível hierárquico mais alto, os de bairro como de nível hierárquico intermediário e o nível mais baixo com os atributos de setor censitário.

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.

3.4. Testes e Análise dos Resultados

Serão abordadas neste item algumas consultas, com o propósito de validar o SSD proposto. Para isso foram levantadas algumas questões junto à Prefeitura de Macaé, para poder dar mais autenticidade e valor aos testes aqui realizados. Para responder à cada questão, primeiramente serão explicados quais atributos de cada dimensão estão envolvidos e, em seguida, apresentada a resposta dada pelo sistema.

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

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

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

Fig. 5 - Pessoas com mais de 45 anos, desempregadas por Setor Administrativo.

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 setor censitário 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 pode ser visto na figura 5.

3.4.2 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, característica 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.

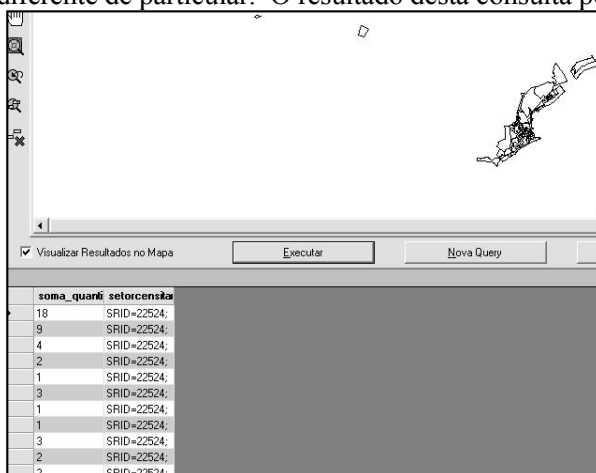
Desta forma, devem ser exibidos no resultado os atributos quantidadepessoa da tabela fato, o atributo trabalha da dimensão trabalho e o atributo cor_raca da dimensão característica, conforme a figura a seguir.

	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. 6 - Pessoas cursando o 3o grau por raça e por estar ou não trabalhando

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

Esta consulta exige alguns cuidados envolvendo 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. O resultado desta consulta pode ser visto na figura 7.



soma_quarã	setorcensita
18	SRID=22524:
9	SRID=22524:
4	SRID=22524:
2	SRID=22524:
1	SRID=22524:
3	SRID=22524:
1	SRID=22524:
1	SRID=22524:
3	SRID=22524:
2	SRID=22524:
2	SRID=22524:

Fig. 7 - Setores Censitários que possuem pessoas com educação especial morando a mais de 1Km de alguma Unidade de Ensino Público

4. Conclusões

Uma ferramenta OLAP em Software Livre com suporte à manipulação de dados geográficos, tal qual se propõe aqui, possui um elevado potencial de aplicações no setor público, no que se refere ao uso pelo administrador municipal. Além disto, uma das principais características de uma ferramenta de suporte à decisão, a facilidade de uso, é oferecida pelo PostGeoOlap, dispensando conhecimentos profundos de informática.

O presente trabalho traz as seguintes contribuições: a demonstração da utilização de ferramentas de distribuição gratuita no desenvolvimento de um Sistema de Suporte à Decisão, demonstrando que esta tecnologia está disponível para as mais diversas empresas, sejam de pequeno ou de grande porte, públicas ou privadas; a integração, desde o nível conceitual, de diferentes tecnologias de suporte a decisão, como SIG e OLAP. Estas tecnologias, juntas, permitem ao decisor um alcance analítico bem maior do que permitiram caso utilizadas de modo isolado.

Referências

- [1] CASTELLS, Manuel. *A Era da Informação: Economia, Sociedade e Cultura*. Volume 1: A Sociedade em Rede. São Paulo: Paz e Terra, 1999.
- [2] CÂMARA, Gilberto; DAVIS, Clodoveu. “Conceitos Básicos da Ciência da Geoinformação”. In: CÂMARA, Gilberto et al. *Introdução à Ciência da Geoinformação*. Disponível em <http://www.dpi.inpe.br/gilberto/livro/introd>, acesso em 30/04/2006. São José dos Campos: DPI/INPE, 2001.
- [3] COLONESE, Giovanni. *Uma Ferramenta para Integração de Sistemas de Bancos de Dados Analíticos e Geográficos*. Dissertação de Mestrado. Campos dos Goytacazes: Universidade Candido Mendes, 2004.
- [4] STEFANOVIC, Nebojsa. *Design and Implementation of On-Line Analytical Processing (OLAP) of Spatial Data*. Dissertação de Mestrado. Belgrado, Iugoslávia: Universidade de Belgrado, 1997.
- [5] SHEKHAR, Shashi et al. “MapCube: A Visualization Tool for Spatial Data Warehouse”. In: MILLER, Harvey; HAN, Jiawei (orgs.). *Geographic Data Mining and Knowledge Discovery*. Taylor and Francis, 2001.
- [6] KOUBA, Zdeněk; MATOUŠEK, Kamil.; MIKŠOVSKÝ, Petr. “On Data Warehouse and GIS Integration”. In: *Proceedings of 11th International Conference on Database and Expert Systems Applications (DEXA 2000)*. Greenwich, Inglaterra, 2000.
- [7] FERREIRA, Ana Cristina F.; CAMPOS, M.L.; TANAKA, Asterio K. “An Architecture for Spatial and Dimensional Analysis Integration”. In: *Proceedings of World Multiconference of Systemics, Cybernetics and Informatics 2001*. Vol. XIV, Computer Science Engineering, Part II, p. 392-395. Orlando, EUA: SCI, 2001.
- [8] FIDALGO, R. N.; TIMES, V. C.; SOUZA, F. F. “GOLAPA: Uma Arquitetura Aberta e Extensível para Integração entre GIS e OLAP”. In: *Anais do III Workshop Brasileiro de Geoinformática*. Rio de Janeiro:GeoInfo, 2001.
- [9] RIVEST, Sonia; BÉDARD, Yvan; MARCHAND, Pierre. “Toward Better Support for Spatial Decision Making: Defining the Characteristics of Spatial On-Line Analytical Processing (SOLAP)”. In: *Geomatica*. Volume 55, nº 4. 2001.
- [10] HAN, Jiawei; STEFANOVIC, Nebojsa; KOPERSKI, Krzysztof. “Selective Materialization: An Efficient Method for Spatial Cube Construction”. In: *Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 1998.
- [11] PAPADIAS, Dimitris et al. “Efficient OLAP Operations in Spatial Data Warehouses”. In: *Proceedings of the 7th International Symposium on Advances in Spatial and Temporal Databases*, págs. 443-459. ACM Records, 2001.
- [12] SILVEIRA, Sérgio A. *Inclusão Digital, Software Livre e Globalização Contra-Hegemônica*. Disponível em http://www.softwarelivre.gov.br/softwarelivre/artigos/artigo_02, acesso em 08/04/2006. SoftwareLivre.gov.br, 2003.
- [13] BRANCO, Marcelo. *Software Livre*. Disponível em http://www.vecam.org/edm/article.php?id_article=55, acesso em 09/08/2005. Enjeux de Mots, 2005.
- [14] GALANTE, Alan C. *Sistema de Suporte à Decisão no Planejamento Urbano Municipal: Um Estudo de Caso no Município de Macaé*. Dissertação de Mestrado. Campos dos Goytacazes: Universidade Candido Mendes, 2004.
- [15] OPENGIS Consortium Inc. *OpenGIS Simple Features Specification for SQL: Revision 1.1*. OpenGIS Project Document 99-049. Disponível em <http://www.opengeospatial.org/docs/99-049.pdf>, acesso em 11/04/2006. OpenGIS Consortium, 1999.
- [16] HELLERSTEIN, Joseph; NAUGHTON, Jeffrey; PFEFFER, Avi. “Generalized Search Trees for Database Systems”. In: *Proceedings of 21st International Conference on Very Large Databases (VLDB'95)*, págs. 562-573. Zurich, Suíça: Morgan Kaufmann, 1995.