

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM INFORMÁTICA

**Concepção e Validação de um Modelo Multidimensional para Data
Warehouse Espacial**

André Gomes de Sousa

(Mestrando)

Marcus Costa Sampaio, PhD

(Orientador)

Campina Grande – PB
Fevereiro de 2007

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM INFORMÁTICA

**Concepção e Validação de um Modelo Multidimensional para Data
Warehouse Espacial**

André Gomes de Sousa

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande, como parte dos requisitos necessários para obtenção do grau de Mestre em Informática.

Área de Concentração: Ciência da Computação

Linha de Pesquisa: Sistemas de Informação e Banco de Dados

Orientador: Marcus Costa Sampaio, PhD

Campina Grande – PB
Fevereiro de 2007

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

S725c Sousa, André Gomes de
2007 “Concepção e validação de um modelo multidimensional para Data Warehouse espacial”/ André Gomes de Sousa.- Campina Grande, 2007.
92f. : il.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática.

Referências.

Orientador: Prof. Dr. Marcus Costa Sampaio.

1. Data Warehouse. 2. Data Warehouse Espacial. 3. Dados Espaciais. 4. Modelo Multidimensional Espacial. 5. Sistemas de Informação Geográfica. I- Título.

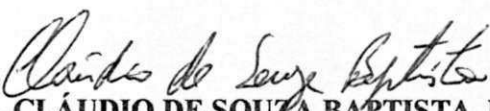
CDU – 004.625.5:004.512.3

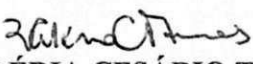
**“CONCEPÇÃO E VALIDAÇÃO DE UM MODELO MULTIDIMENSIONAL
PARA DATA WAREHOUSE ESPACIAL”**

ANDRÉ GOMES DE SOUSA

DISSERTAÇÃO APROVADA EM 26.02.2007


PROF. MARCUS COSTA SAMPAIO, Dr.
Orientador


PROF. CLÁUDIO DE SOUZA BAPTISTA, Ph.D
Examinador


PROFª VALÉRIA CESÁRIO TIMES, Ph.D
Examinadora

CAMPINA GRANDE – PB

Agradecimentos

Agradeço a Deus, pelo apoio em todos os momentos desse trabalho.

Agradeço a minha esposa Valéria, pelo apoio, carinho e paciência. E ao meu filho Pedro Henrique, que nasceu momentos antes da conclusão desse trabalho para aumentar ainda mais a minha motivação.

A minha mãe Josilda, por sempre ter apoiado os meus estudos e por todo o incentivo.

A minha tia Josivanda, por todo o apoio e dedicação.

A minha família, por estar presente em todos os momentos.

A Marcus Sampaio, meu orientador, que me deu a oportunidade de trabalhar com um tema de pesquisa muito interessante e me auxiliou bastante em todos os momentos do mestrado.

A Cláudio Baptista, que sempre esteve acompanhando a nossa pesquisa e ajudando com idéias e dicas relevantes.

A Fabiana, por ter iniciado as pesquisas desse trabalho.

Aos meus amigos de curso, que me ajudaram bastante, principalmente nas disciplinas do mestrado.

A Aninha e Vera, secretárias da COPIN, sempre dispostas a ajudar.

E todos que de alguma forma ajudaram neste trabalho e não tiveram seus nomes citados.

Sumário

Abreviações	vii
Lista de Figuras	viii
Lista de Quadros	ix
Lista de Quadros	ix
Resumo	x
Abstract.....	xi
Capítulo 1	1
Introdução.....	1
1.1 Data Warehousing	1
1.1.1 Data Warehouse.....	2
1.1.2 Operações e Interface OLAP	5
1.2 Motivação	6
1.4 Objetivos.....	7
1.4.1 Objetivos Gerais	7
1.4.2 Objetivos Específicos	8
1.4 Estrutura da Dissertação	8
Capítulo 2	9
Conceitos Básicos de Data Warehouse Espacial	9
2.1 Dados Espaciais	9
2.2 Data Warehouse Espacial	10
2.3 Dimensões em um Data Warehouse Espacial	10
2.4 Medidas em um Data Warehouse Espacial	11
2.5 Hierarquias em um Data Warehouse Espacial	11
2.6 Cubo de Dados Espacial	12
2.7 Operações OLAP espacial (SOLAP).....	12
2.8 Agregação Espacial	12
Capítulo 3	14
Trabalhos Relacionados.....	14
3.1 Abordagem Federada de Integração DW-SIG.....	14
3.1.1 Arquitetura de Matoušek et al.	14
3.1.2 Arquitetura de Ferri et al.	16
3.2 Abordagem de Integração Estreita entre DW e SIG.....	19
3.2.1 Arquitetura de Han et al.	20
3.2.2 Proposta de Shekhar et al.	21
3.2.3 Arquitetura de Fidalgo et al.	23
3.3 Operações OLAP Espaciais (SOLAP)	25
3.3.1 GEOMINER	26
3.3.2 Map Cube	28
3.3.3 SOVAT.....	29
3.3.4 ZHANG et al.	30
3.4 Outros Trabalhos	32
3.5 Considerações Finais	34
Capítulo 4	35
Modelo de Banco de Dados Multidimensional Espacial.....	35
4.1 Formalização dos Principais Conceitos do Modelo	35
4.2 Esquema Conceitual e Consultas SOLAP	37
4.3 Mapeamento do Modelo Conceitual para um Modelo Lógico Objeto Relacional.....	41

4.4. OR-OLAP Espacial	45
4.5 Considerações Finais	46
Capítulo 5	47
Otimização de Consultas Multidimensionais Espaciais	47
5.1 O Problema de Otimização Espacial	47
5.2 Otimização Lógica Espacial	49
5.3 O Algoritmo.....	51
5.3.1 Regras para Criação de Agregados Espaciais.....	52
5.4 Otimização de Consultas Utilizando Concatenação de Medidas Espaciais	53
5.5 Considerações Finais	55
Capítulo 6	56
Validação do Modelo	56
6.1 Implementação do Modelo Conceitual Multidimensional Espacial.....	56
6.1.1 Criação de Consultas para o DWE	57
6.1.2 Visualização dos Resultados	58
6.2 Análise do Desempenho de Consultas SOLAP	66
6.2.1 Indicador de Ganho de Performance (<i>Speedup</i>).....	72
6.3 Considerações Finais	73
Capítulo 7	74
Conclusão	74
7.1 Trabalhos Futuros	77
Referências Bibliográficas.....	79

Abreviações

DW - Data Warehouse (Armazém de dados)

SIG - Sistemas de Informação Geográfica

SAD - Sistemas de Apoio à Decisão

DWE - Data Warehouse Espacial

SGBD – Sistema de Gerenciamento de Banco de Dados

DBMS – *Database Management System*

Lista de Figuras

Figura 1 - Arquitetura Data Warehousing	2
Figura 2 - Representação de um cubo	3
Figura 3 - Exemplo de um esquema em estrela de Data Warehouse	4
Figura 4 - Formatos de representação de dados espaciais	9
Figura 5- Arquitetura GOLAP	15
Figura 6 - <i>Binding Attributes</i>	18
Figura 7 - DW com medidas e dimensões espaciais	21
Figura 8 - Visualização do Operador Map Cube	22
Figura 9 - Amostra de resultados do operador Map Cube sob dados censuais	23
Figura 10 - Arquitetura GOLAPA.....	24
Figura 11 - Operação Roll-up Espacial	28
Figura 12 - Interface do SOVAT	30
Figura 13 - Esquema CE Agrodistribuicao.....	38
Figura 14 - Esquema em estrela espacial OR Agrodistribuicao	43
Figura 15 - Divisão política do Estado da Paraíba	47
Figura 16 - Exemplo de agregado espacial.....	50
Figura 17 - Concatenação de Geometrias X União de Geometrias.....	54
Figura 18 - Arquitetura do <i>MapWarehouse</i>	56
Figura 19 - Interface do <i>MapWarehouse</i> , criando consultas.....	57
Figura 20 - Definindo atributos de uma consulta	59
Figura 21 - Definindo restrições para a dimensão Tempo.....	60
Figura 22 - Definindo restrições da dimensão espacial Localização.....	61
Figura 23 - Definindo janela espacial no mapa	61
Figura 24 - Definindo operação Roll-up Espacial	62
Figura 25 - Definindo agrupamento temporal	62
Figura 26 - Definindo funções de agregação.....	63
Figura 27 - <i>MapWarehouse</i> resultados no nível Microregião	64
Figura 28 - <i>MapWarehouse</i> resultados no nível Região	64
Figura 29 - <i>MapWarehouse</i> Informações no mapa por Microregião	65
Figura 30 - <i>MapWarehouse</i> Informações no mapa por Região.....	65
Figura 31 - Resultado dos experimentos para as abordagens de otimização.....	68
Figura 32 - Resultado dos experimentos para as abordagens de otimização (Histograma)	69
Figura 33 - Comparação entre abordagens de otimização de consultas	70
Figura 34 - Comparação entre as abordagens de otimização de consultas (Histograma)	70
Figura 35 - Comparação entre as abordagens COT e COC.....	71
Figura 36 - Comparação entre as abordagens COT e COC (Histograma)	72
Figura 37 - Curva mostrando o indicador de ganho de desempenho <i>speedup</i>	73

Lista de Quadros

Quadro 1 - Avaliação de abordagens federadas	19
Quadro 2 - Avaliação de abordagens integradas	25
Quadro 3 - Avaliação de abordagens SOLAP	32
Quadro 4 - Lacunas de pesquisa em DWE.....	33
Quadro 5 - Principais contribuições do trabalho	76

Resumo

Data Warehouse espacial pode ser caracterizado como um banco de dados multidimensional que inclui dimensões e medidas espaciais. Dependendo do contexto, informações espaciais são essenciais à análise de dados em Sistemas de Apoio à Decisão. Por exemplo, agregação dinâmica de mapas em diversos níveis, combinada com dados descritivos e numéricos correspondentes, deve facilitar sobremaneira a tomada de decisão.

Este trabalho de pesquisa propõe e define formalmente um modelo de dados multidimensional que integra estreitamente dados espaciais e dados não-espaciais – modelo multidimensional espacial. Esse modelo, apropriado para Data Warehouse espacial, é descrito em seus aspectos estruturais e comportamentais, tratando da modelagem multidimensional dos dados espaciais e analíticos e de operações OLAP espaciais. Também é discutida a implementação objeto-relacional (OR) de esquemas multidimensionais espaciais em SGBD OR que lidam com dados espaciais, levando em consideração questões relacionadas a performance de consultas que envolvem agregação espacial. Para garantir o bom desempenho dessas consultas, foram propostas algumas técnicas de otimização.

O modelo e as idéias propostas foram validados através da implementação de um protótipo que utiliza dados analíticos e espaciais baseados em um estudo de caso. O protótipo foi construído e testado utilizando o SGBD Oracle 10g. Uma avaliação experimental constatou a eficiência dos algoritmos de otimização de consultas propostos.

Abstract

Spatial Data Warehouse can be characterized as a multidimensional database that includes spatial dimensions and spatial measures. Depending on the context, spatial information are essential to analyze of data in Decision support systems. For example, dynamic aggregation of maps in diverse levels, combined with descriptive data and numerical correspondents, must facilitate the decision taking excessively.

This research proposes and define formally a multidimensional data model that tightly integrates spatial and non-spatial data – spatial multidimensional model. This model, suitable for Spatial Data Warehouse, is described in its structural and managing aspects, treating of the multidimensional modeling of the spatial and analytical data and spatial OLAP operations. Moreover, is discussed the object-relational implementation of space multidimensional schemas in OR DBMS that deals with spatial data, taking in consideration questions related to performance of queries that involve spatial aggregation. To guarantee the good performance of these queries, some enhancing techniques were proposed.

The model and the ideas proposed were validated through the implementation of an archetype that uses analytical and spatial data established in a case study. The archetype was constructed and tested using the Oracle 10g DBMS. An experimental evaluation evidenced the efficiency of the proposed enhancing algorithms.

Capítulo 1

Introdução

Sistemas de Informação em grandes corporações lidam com uma quantidade imensa de dados, potencialmente dificultando a extração de informações relevantes para a tomada de decisão em tempo satisfatório. Sistemas de Apoio à Decisão (SAD) são sistemas capazes de auxiliar na análise de informações de negócio, identificando tendências, comportamentos e possíveis problemas que possibilitem a tomada de decisões inteligentes por parte dos que desempenham o papel de gerência executiva (e.g. administradores, gerentes). Esses sistemas surgem devido às pressões empresariais, que exigem rapidez (aumentando a produtividade) e eficiência (focando nas informações relevantes) no ambiente corporativo para atender às necessidades e competitividades do mercado. Em geral, os dados primários que alimentam um SAD são muito dispersos, isto é, mantidos em diferentes lugares e por diferentes sistemas, resultando quase sempre em uma quantidade enorme de dados. Para atender aos requisitos de rapidez e eficiência de um SAD, uma arquitetura conhecida como Data Warehousing – armazenagem de dados, literalmente – vem se impondo.

Na subseção seguinte, são apresentados alguns conceitos e terminologia concernentes à abordagem de Data Warehousing, necessários para o entendimento do restante desta dissertação.

1.1 Data Warehousing

Numa arquitetura do tipo Data Warehousing, o elemento central é um Banco de Dados (BD) multidimensional, chamado de Data Warehouse – armazém de dados. Sistemas OLAP (On-Line Analytical Processing) (DATE, 2000) são operadores especiais que possibilitam a manipulação dinâmica dos dados do Data Warehouse (DW) (KIMBALL, ROSS e MERZ, 2002).

A Figura 1 (adaptada de (MALINOWSKI, 2003)) permite visualizar as três camadas principais da arquitetura: fontes, Data Warehouse e interface OLAP.

- Fontes: esta camada é representada pelos bancos de dados e/ou fontes externas, que contêm os dados operacionais resultantes das transações efetuadas diariamente pela empresa;

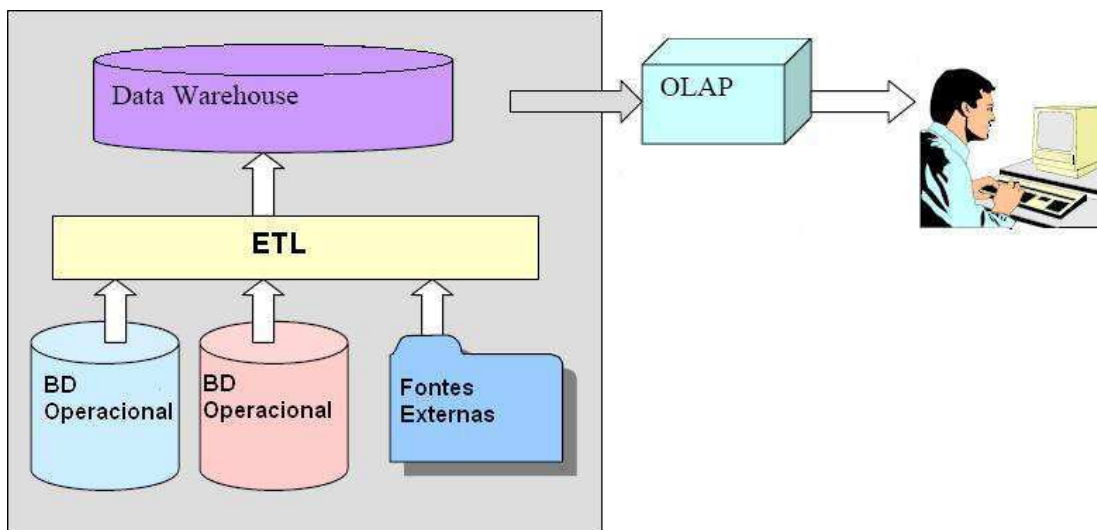


Figura 1 - Arquitetura Data Warehousing

- Data Warehouse: camada na qual os dados utilizados nas consultas são armazenados. Esses dados são extraídos das fontes transacionais, e depois de transformados e integrados, são armazenados em forma multidimensional, o que caracteriza o processo conhecido como ETL (*Extraction, Transforming and Loading*) (KIMBALL e CASERTA, 2004);
- OLAP (DATE, 2000): operações especiais sobre o DW.

A seguir, são detalhadas as camadas Data Warehouse e OLAP, respectivamente. O processo de ETL não é detalhado por estar fora do escopo do trabalho.

1.1.1 Data Warehouse

Data Warehouse é um banco de dados que armazena dados sobre as operações da empresa (e.g. vendas, compras), sendo os mesmos extraídos de uma ou mais fontes, e transformados em informações úteis, oferecendo um enfoque histórico, para permitir um suporte efetivo à decisão. O DW é uma visão materializada dos dados contidos nas fontes, que permite obter informações sumarizadas através de níveis de agregação maiores do que os das fontes, por exemplo: nas fontes (bancos de dados transacionais) são geradas constantemente transações de vendas feitas durante o dia em uma loja; e no DW, após uma sumarização dessas transações, são disponibilizadas as vendas totais de um dia, mês ou até ano, através de visões materializadas. Enfim, é um BD Multidimensional em que as dimensões são critérios de agregação para fatos (ou medidas numéricas) relativos a um certo tema de negócio. Por exemplo, um DW de vendas que armazena a quantidade de vendas (fato) por dia (dimensão Tempo), por cliente (dimensão Cliente) e por produto (dimensão Produto).

Um aspecto de extrema importância em DW é o temporal. Enquanto os dados nos sistemas transacionais são atualizados frequentemente, ocasionando a perda de informações históricas, a dimensão Tempo nos DW permite traçar a evolução e versão dos dados da organização. Para (KIMBALL, 1996), a dimensão temporal é essencial na definição de um DW, pois os negócios precisam separar seus dados por dias de trabalho versus feriados, ou por períodos fiscais, ou por estações, ou por outros eventos.

Uma das possíveis implementações para o modelo de dados de um DW pode ser feita a partir de um *cubo de dados*, com n arestas: cada aresta do cubo representa uma *dimensão*, enquanto as células do cubo contêm *medidas* ou *fatos* relacionados com valores bem determinados de cada uma das dimensões. Os fatos são atributos representando elementos específicos de análise, que são explorados a partir das dimensões, tais como: quantidade vendida, custo, número de clientes, e quantidade de acidentes. Em geral, eles podem ser agregados para facilitar o entendimento de aspectos particulares que venham a ser considerados. Por exemplo, o usuário poderia estar interessado em saber a quantidade de acidentes (fato) que ocorreram no ano de 2005 (dimensão Tempo) na cidade de “Brasília” (dimensão Localização); e para calcular essa quantidade seria necessário somar (agregação) todos os acidentes que ocorreram nos dias daquele ano.

Um exemplo de um cubo pode ser visto na Figura 2, na qual temos as dimensões *Tempo*, *Cliente* e *Produto*, e o fato “Quantidade vendida”; em destaque, podemos ver a localização no cubo do seguinte fato: a quantidade (Fato) de laranjas (Produto) vendidas no ano de 1999 (Tempo) ao supermercado “Super Super”(Cliente).

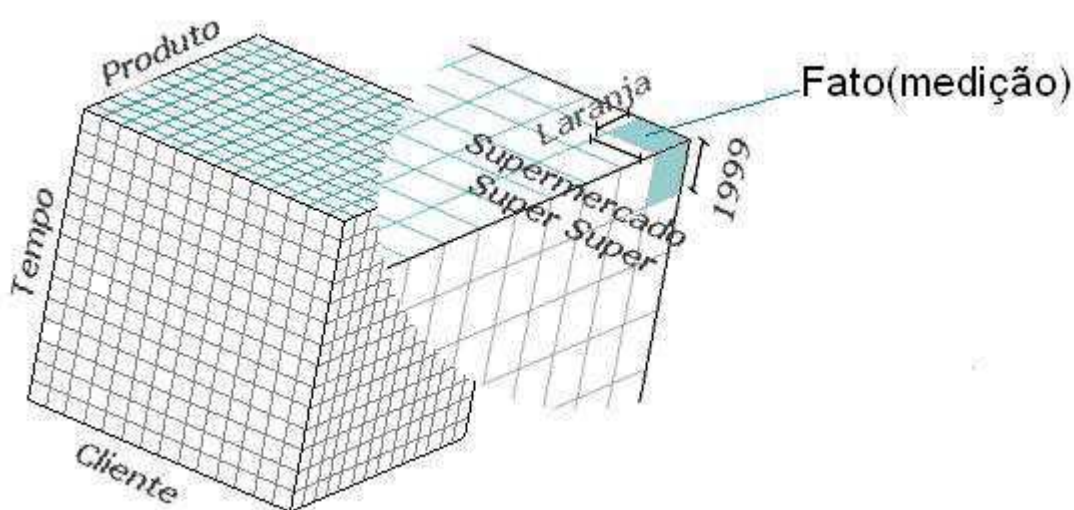


Figura 2 - Representação de um cubo

Outro conceito importante dos DW é o de hierarquia. Cada dimensão de um cubo pode ter uma ou mais hierarquias. Hierarquia pode ser definida como um conjunto de níveis de

agregação, de uma mesma dimensão, nos quais os dados podem ser analisados. Por exemplo, suponha que a dimensão “Loja”, indicando o local onde determinado produto foi vendido, tenha a seguinte hierarquia: Cidade → Estado → País. As hierarquias sempre obedecem ao relacionamento um-para-muitos do nível maior (ou mais geral) para o nível menor, como pode ser visto nessa hierarquia, na qual um País possui muitos Estados, mas um Estado só pode estar em um País. As hierarquias são bastante úteis, pois permitem que o usuário possa avaliar informações em diferentes níveis de agregação, como por exemplo: saber qual a quantidade de vendas de um determinado produto por cidade, por Estado e por País; sendo que para obter as informações de um País são necessários apenas os dados agregados de seus Estados (que podem já ter sido pré-calculados), e da mesma forma, para obter a quantidade de vendas de um Estado, basta somar a quantidade de vendas de suas cidades.

Na maioria dos casos, cubos de dados são implementados como um banco de dados relacional, na forma de um *esquema em estrela* (KIMBALL, ROSS e MERZ, 2002). Existem duas regras básicas para transformar um cubo de dados em um esquema em estrela: 1) As arestas do cubo são transformadas em tabelas de dimensão; 2) todas as células são armazenadas em uma única tabela de fatos. Além disso, as dimensões são conectadas à tabela de fatos através de chaves estrangeiras, como será visto no exemplo a seguir. A Figura 3 mostra um exemplo de esquema em estrela, consistindo de uma tabela de fatos, tabelas de dimensões e de hierarquias. Observe que no esquema em estrela, devido à pobreza semântica do modelo relacional, as hierarquias ficam obliteradas nas tabelas de dimensões, o que não acontece com o modelo conceitual, no qual as hierarquias são explicitamente definidas.

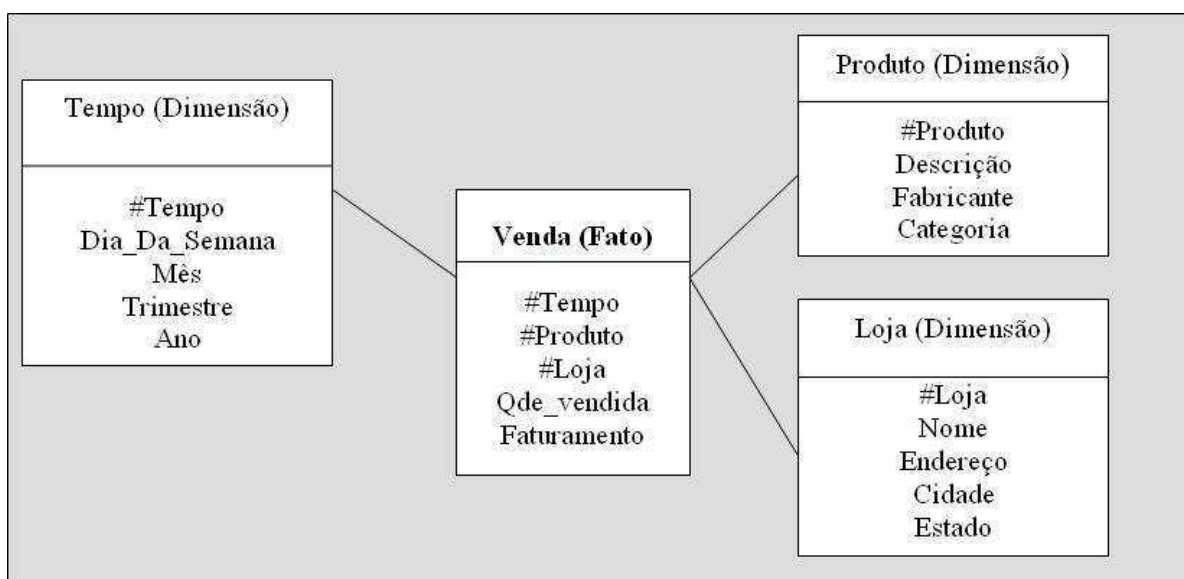


Figura 3 - Exemplo de um esquema em estrela de Data Warehouse

No modelo relacional, as tabelas de dimensões contêm atributos que se relacionam por dependência funcional, isto é, com cardinalidade de muitos-para-um entre um conjunto de atributos e outro (DATE, 2000); caracterizando uma forma de organização hierárquica, usada para prover a realização de operações de navegação nos dados através dos diferentes níveis de detalhamento. Isto reflete diretamente do conceito de hierarquia proveniente do cubo de dados.

1.1.2 Operações e Interface OLAP

Operações sobre as hierarquias e combinações de dimensões são realizadas pelas ferramentas OLAP. Estas fornecem um processo interativo de criação, gerência e análise dos dados nos níveis hierárquicos através de operações como (DATE, 2000), (POURABBAS, 2003):

- *Roll-up*: significa ir de um nível mais especializado (inferior) para um de maior granularidade (superior). Por exemplo, fazendo um Roll-up de mês para ano na dimensão Tempo da Figura 3 obtemos a agregação dos valores de vendas de todos os meses de um ano em um único valor, que representa a quantidade vendida por ano;
- *Drill-down*: tem efeito oposto ao anterior, ou seja, é feita a especialização de um nível superior para um de menor granularidade. Por exemplo, se temos a quantidade vendida por ano e queremos detalhar essas informações por mês, considerando a Figura 3, basta fazer um Drill-down de ano para mês na dimensão Tempo;
- *Slice*: omite dimensões do cubo e o resultado é um cubo definido pelas dimensões restantes. Por exemplo, na Figura 3, podemos fazer um Slice para visualizar as vendas considerando apenas as dimensões Tempo e Loja, omitindo a dimensão Produto;
- *Dice*: restringe o valor do domínio da dimensão do cubo conforme um predicado. Por exemplo, podemos fazer um Dice no esquema da Figura 3 para apresentar apenas as vendas que ocorreram no ano de 2002 e em uma determinada loja X;
- *Pivot*: permite a inversão das arestas do cubo, oferecendo a visualização dos resultados sobre diferentes perspectivas.

Esses operadores, em conjunto com algumas funções matemáticas e estatísticas (e.g. SUM, COUNT, AVERAGE, STANDARD DEVIATION) realizadas sobre as medidas de

negócio, oferecem uma gama de informações aos responsáveis pela tomada de decisão em bases com grande volume de dados. Por exemplo, no esquema da Figura 3 pode-se recuperar facilmente as quantidades vendidas (Fato) de um produto X no mês de Janeiro para obter o total de vendas desse mês. Em contrapartida, nos tradicionais sistemas orientados à transação, em que mudanças nos dados ocorrem constantemente, são fornecidos relatórios complexos que podem vir a dificultar o entendimento e a navegabilidade sobre os dados.

As consultas estratégicas são feitas ao longo das dimensões e/ou de suas hierarquias. Na dimensão Tempo da Figura 3, sua hierarquia pode ser definida como Dia_Da_Semana → Mês → Ano (sendo que “→” representa a dependência funcional entre os atributos) possibilitando fazer consultas sobre os dados considerando-os anualmente através da agregação de dias e meses, por exemplo. Da mesma forma, a dimensão Loja define a seguinte hierarquia: Endereço → Cidade → Estado, na qual o usuário poderia consultar informações de quantidade de vendas de um determinado produto X por cidade ou por Estado, por exemplo.

A subseção 1.2 contém uma discussão que motiva a inserção de informações espaciais em Data Warehouses convencionais.

1.2 Motivação

Dados representam o elemento principal do processo de tomada de decisão. São a matéria-prima da informação e conseqüentemente do conhecimento. Entretanto, um componente muito rico de informação, que é o componente espacial, não é explorado pelos esquemas em estrela tradicionais. Convencionalmente, dados espaciais são tratados de forma textual (uma descrição baseada em nomes de lugares), como pode ser visto na Figura 3 – cidade e Estado, na dimensão Loja – na qual as coordenadas geográficas das lojas são desconsideradas. Desta forma, não há como visualizar essas informações em mapas, e nem realizar consultas que necessitam de operações espaciais, como por exemplo: “Quantos produtos foram vendidos no ano de 2000 nas cidades que fazem fronteira com São Paulo?”.

Foi estimado que cerca de 80% dos dados corporativos possuem um componente de localização geográfica (FRANKLIN, 1992), para os quais a referência espacial pode ser descrita em termos de posição, forma, orientação e tamanho. Disponibilizar um fato usando o seu componente espacial ajuda o usuário a ter um melhor entendimento do mesmo, através da visualização de sua localização. O uso do componente espacial dos dados permite descobrir relacionamentos entre vários fenômenos geográficos, por exemplo, a correlação espacial entre a frequência de uma doença X e a média de emissão de um poluente Y. Daí surge à

necessidade de uma abordagem híbrida, composta por características multidimensionais (Data Warehouse) e espaciais (Sistemas de Informação Geográfica).

Sistemas de Informação Geográfica (SIG), também considerados sistemas de apoio à decisão, diferenciam-se dos demais sistemas por serem capazes de coletar, modelar, gerenciar, manipular, analisar e integrar dados referenciados geograficamente (ADAM e GANGOPADYAY, 1998); propiciando uma consistente análise espacial das informações, que pode ser aplicada em uma grande diversidade de áreas.

A maioria das ferramentas OLAP convencionais não são adaptadas para dados espaciais, pois os DW 'tradicionais' não exploram adequadamente o componente espacial dos objetos. Portanto, a necessidade de inserir o componente espacial nesse tipo de DW motiva a integração entre DW e SIG, surgindo assim um novo conceito, chamado de Data Warehouse espacial (DWE), que exige um novo modelo de DW composto de operações analítico-espaciais.

Data Warehouse espacial integra estreitamente BD Multidimensional e BD Espacial, criando um ambiente único de auxílio à tomada de decisão para possibilitar o processamento eficiente de consultas que consideram, conjuntamente, dados espaciais e numéricos, segundo diferentes critérios de agregação, e em diferentes níveis de agregação - análise de dados e mapas.

Diante desse contexto, os objetivos deste trabalho são apresentados na próxima seção.

1.4 Objetivos

Nesta seção, são apresentados os objetivos almejados por este trabalho, divididos em objetivos gerais e específicos.

1.4.1 Objetivos Gerais

- Propor um modelo multidimensional espacial orientado a objeto, adequado para Data Warehouse espacial, que integra estreitamente dimensões e dados espaciais e analíticos. O modelo também deve fornecer operações OLAP específicas para DWE;
- Definir formalmente os principais conceitos do modelo;
- Mapear o modelo proposto para um esquema Objeto Relacional (OR);
- Criar algoritmos de otimização de consultas multidimensionais espaciais;
- Validar o modelo, através da construção de um protótipo.

1.4.2 Objetivos Específicos

- Projetar uma arquitetura de SDW utilizando o sistema iGIS (internet Geographic Information System) (BAPTISTA et al., 2004) desenvolvido pelo grupo de Sistemas de Informação e Banco de Dados (SINBAD) da Universidade Federal de Campina Grande (UFCG);
- Estender um SGBD OR, o Oracle 10g, com operações OLAP específicas para DWE, para possibilitar a implementação de um esquema em estrela espacial Objeto Relacional;
- Conceber um otimizador de consultas baseado no cálculo antecipado de agregações espaciais;
- Validar o modelo proposto a partir do desenvolvimento de um protótipo de DWE no SGBD Oracle10g, utilizando a representação vetorial (COUCLELIS, 1992) para os dados espaciais.

A seguir, na seção 1.4, é apresentada a estrutura da dissertação.

1.4 Estrutura da Dissertação

Os capítulos seguintes estão organizados como segue. O capítulo 2 trata dos principais conceitos relacionados a Data Warehouse espacial. O capítulo 3 resume os trabalhos relacionados ao tema de integração entre Banco de Dados Multidimensional e Banco de Dados Espacial. O capítulo 4 propõe o modelo de Banco de Dados Multidimensional Espacial com operações de navegação em hierarquias espaciais e agregação de medidas espaciais. O capítulo 5 apresenta a implementação de um protótipo de software, com intuito de validar o modelo proposto no capítulo anterior. Para essa validação, é usado um estudo de caso que integra informações analíticas e espaciais. O capítulo 6 caracteriza o problema de otimização dessas consultas multidimensionais espaciais, apresentando o projeto e a implementação de um algoritmo genérico de otimização capaz de garantir o bom desempenho das mesmas, principalmente as que utilizam a operação de união de geometrias. O capítulo 7 demonstra a avaliação experimental do algoritmo de otimização espacial proposto. Finalmente, o capítulo 8 resume esse trabalho de pesquisa e discute questões para trabalhos futuros.

Capítulo 2

Conceitos Básicos de Data Warehouse Espacial

Neste capítulo, são apresentados os principais conceitos referentes à integração entre dados analíticos e espaciais, necessários para o entendimento do restante da dissertação.

2.1 Dados Espaciais

Existem duas abordagens principais de representação dos componentes espaciais associados às informações geográficas: a matricial (ou raster) e a vetorial (COUCLELIS, 1992), exemplificadas na Figura 4.

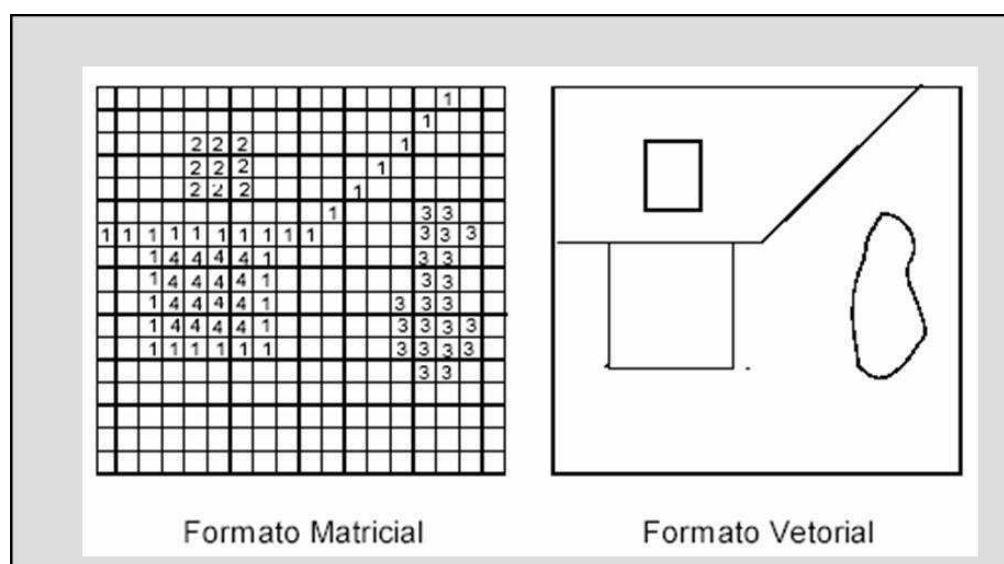


Figura 4 - Formatos de representação de dados espaciais

A representação matricial é feita através de uma grade de células, em cada uma delas armazena um valor correspondente ao tipo da entidade e sua posição é definida por linha e coluna. Já no formato de representação vetorial, as entidades geográficas são armazenadas espacialmente como ponto (e.g. poste), linha (e.g. rua) ou polígono (e.g. cidade), na qual cada objeto possui uma identificação própria. A posição destes objetos é dada em relação a um sistema de coordenadas previamente especificado. Os dados em um formato podem ser convertidos no outro.

Os principais componentes dos dados georeferenciados são (IOCHPE e LISBOA, 1996):

- Atributos descritivos: que armazenam aspectos não-geográficos das entidades espaciais, podendo ser tratados por SGBD convencionais;

- Atributos de localização geográfica: referem-se à geometria dos objetos e envolve conceitos de métrica, sistemas de coordenadas, posicionamento, funções de distância entre pontos e cálculo de áreas, entre outros;
- Componente tempo: oferece as características temporais, sazonais ou periódicas dos objetos;
- Componente gráfico: atributos que indicam como os objetos espaciais serão apresentados em mapas.

Um dado espacial (e.g. uma cidade, um rio, uma rua) só poderá ser localizado se for descrito em relação a outros objetos espaciais cujas posições sejam previamente conhecidas, ou se tiver sua localização determinada em uma rede coerente de coordenadas. Com a utilização de um sistema de coordenadas, é possível definir a posição de qualquer ponto na superfície terrestre (CAMARA et al., 1996). De acordo com CAMARA et al. (1996) um possível sistema de coordenadas espaciais seria através da representação de um ponto da superfície terrestre por um valor de latitude e longitude: longitude é a distância angular entre um ponto qualquer da superfície terrestre e o meridiano de origem; e latitude é a distância angular entre um ponto qualquer da superfície terrestre e a linha do Equador. Pontos que não correspondem à medição média dos oceanos podem ter também a altitude como terceiro parâmetro.

2.2 Data Warehouse Espacial

Data Warehouse espacial pode ser visto como a implementação de um modelo de dados multidimensional com medidas, dimensões e hierarquias espaciais e não-espaciais, resultante da integração entre sistemas de DW e SIG. Para tal, é utilizado um banco de dados multidimensional com suporte a dados espaciais. Desta forma, é criado um ambiente único de auxílio à tomada de decisão para possibilitar o processamento eficiente de consultas que consideram, conjuntamente, dados espaciais e numéricos, segundo diferentes critérios de agregação, e em diferentes níveis de agregação.

2.3 Dimensões em um Data Warehouse Espacial

Dimensões são critérios de agregação para medidas (ou fatos) de um domínio de negócio. Certos atributos das dimensões representam diferentes níveis de agregação dos dados, formando hierarquias.

Os tipos de dimensões identificados em (HAN, KOPERSKI e STEFANOVIC, 1998) são:

- Dimensão não-espacial: contém apenas dados não-espaciais ao longo de níveis de agregação da informação. Por exemplo, uma dimensão *Produto* que possui os atributos (ou níveis) descritivos *Modelo* e *Categoria*;
- Dimensão mista: contém tanto atributos analíticos quanto espaciais. O nível de agregação de menor granularidade é espacial, e à medida que generaliza ou aumenta a granularidade, surgem níveis não-espaciais. Por exemplo, uma dimensão *Localização* com um nível de agregação representado por atributo espacial *Município* (e.g. Geometria de Campina Grande), seguido de um nível de maior granularidade representado pelo atributo descritivo *Região* (e.g. nome 'Agreste');
- Dimensão espacial: todos os níveis de agregação possuem dados espaciais. Por exemplo, uma dimensão com os níveis de agregação definidos pelos dados geográficos *Geometria do Município*, *Geometria do Estado* e *Geometria do País*.

2.4 Medidas em um Data Warehouse Espacial

De acordo com STEFANOVIC (1997) existem dois tipos de medidas em um DW espacial:

- Medida numérica: contém apenas números. Por exemplo, a quantidade de vendas de uma empresa em uma região;
- Medida espacial: São objetos espaciais utilizados para representar fatos espaciais de um domínio de negócio, ou seja, quando um componente espacial é utilizado como objeto de análise para auxílio na tomada de decisões.

2.5 Hierarquias em um Data Warehouse Espacial

As dimensões de um DW convencional podem conter hierarquias, que são formadas por um conjunto de níveis de agregação dos dados. Por exemplo, um DW que armazena a quantidade de vendas de uma empresa por dia, por mês e por ano, na dimensão Tempo, formando a hierarquia Dia → Mês → Ano. Observe que os níveis da hierarquia são atributos de uma dimensão. O principal objetivo das hierarquias é de possibilitar que o usuário visualize os dados sobre diferentes níveis de agregação, usando operadores OLAP.

Nos DWE, temos as hierarquias espaciais. Uma hierarquia espacial é formada por diferentes níveis de agregação de dados espaciais ou não espaciais e deve conter pelo menos um nível espacial. No caso das dimensões espaciais, todos os níveis das hierarquias são

espaciais, por exemplo, a hierarquia Geometria_do_Município → Geometria_da_Microregião → Geometria_do_Estado, com todos os níveis representados por objetos espaciais. Já nas dimensões mistas, apenas alguns níveis da hierarquia são espaciais, por exemplo, a hierarquia Geometria_do_Município → Nome_da_Microregião → Nome_do_Estado, em que apenas o nível inicial é espacial e os outros são textuais.

2.6 Cubo de Dados Espacial

Pelo fato das medidas serem tipicamente multidimensionais, uma medida deve ser qualificada por cada dimensão para ter significado, mais precisamente, cada objeto espacial representando uma medida espacial é associado com um objeto e apenas um de cada uma das dimensões espaciais e não-espaciais.

Um cubo de dados espacial é definido pelas associações entre medidas espaciais e dimensões não-espaciais/espaciais, na qual as informações geográficas são armazenadas sobre diferentes perspectivas e níveis espaciais de agregação da informação.

2.7 Operações OLAP espacial (SOLAP)

As ferramentas OLAP comumente utilizadas para explorar os DW convencionais não são adequadas para analisar dados espaciais. Para explorar todo o potencial dos Data Warehouses espaciais, novas ferramentas são necessárias. Surge uma nova categoria de ferramentas, conhecida como SOLAP, que combinam GIS e OLAP em um novo tipo de aplicação. SOLAP pode ser definido como uma plataforma visual feita especialmente para fornecer suporte a análises espaços-temporais e exploração de dados com facilidade e rapidez, seguindo uma abordagem multidimensional composta de níveis de agregação disponíveis tanto em forma cartográfica quanto em tabelas (RIVEST, BÉDARD e MARCHAND, 2001).

2.8 Agregação Espacial

Nas operações SOLAP, são realizadas operações de agregação de objetos espaciais. O resultado de uma agregação espacial é um dado espacial representando a sumarização de dados geográficos. Por exemplo, suponha que o usuário deseja visualizar as áreas em que mais ocorreram acidentes de trânsito em uma cidade: nesse caso, todas as áreas em que ocorreram acidentes (objetos espaciais) seriam agregadas (e.g. unidos), formando um único objeto espacial que representa todas as áreas espaciais. Funções de agregação para objetos espaciais já foram definidas, incluindo *convex hull*, *geometric union*, *geometric intersection* (INDULSKA e ORLOWSKA, 2002), (SHEKHAR e CHAWLA, 2003): o resultado dessas

funções é definido por um único objeto espacial, que pode ser representado por geometrias simples ou complexas.

Capítulo 3

Trabalhos Relacionados

Neste capítulo, é discutido o estado da arte referente ao tema de integração entre Data Warehouse e Sistemas de Informação Geográfica. As vantagens e desvantagens de propostas de integração existentes são analisadas, visando discernir questões em aberto.

3.1 Abordagem Federada de Integração DW-SIG

A idéia de arquitetura federada ou integrada refere-se a como as informações estarão organizadas e como serão visualizadas pelo usuário, em um Data Warehouse espacial (DWE). Na arquitetura federada de integração de dados, as informações são extraídas a partir de esquemas globais integrados de bancos de dados independentes, mas cooperativos, conforme os requisitos das consultas sobre o sistema (ZISMAN e KRAMER, 1995). A integração entre dados analíticos e espaciais preserva a autonomia das bases locais de DW e SIG. Nas próximas subseções são analisadas algumas abordagens federadas.

3.1.1 Arquitetura de Matoušek et al.

Essa arquitetura, conhecida como GOLAP, propicia a realização de consultas analítico-espaciais segundo uma abordagem federada, e foi definida em (MATOUSEK, KOUBA e MIKSOVSY, 2000), (MATOUSEK e SVOBODA, 2000), (MATOUSEK, KOUBA e MIKSOVSKY, 2001) e (MATOUSEK, MORDACIK e JANKU, 2001). Ela faz parte do projeto GOAL (Geographical Information On-Line Analysis) que tem como objetivo encontrar formas de integrar DW com SIG (<http://krizik.felk.cvut.cz/goal>).

O principal componente da arquitetura é um mediador, denominado módulo de integração, responsável pela integração entre o DW e o SIG. Os componentes da arquitetura podem ser vistos na Figura 5 (MATOUSEK, KOUBA e MIKSOVSY, 2000). Durante o projeto dessa arquitetura, tentou-se identificar requisitos que possibilitassem soluções facilmente adaptáveis a diferentes sistemas OLAP e SIG. A seguir, são descritos os componentes da arquitetura GOLAP.

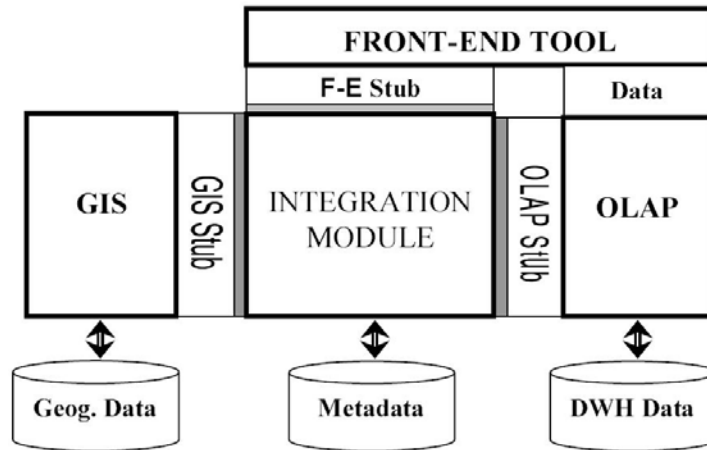


Figura 5- Arquitetura GOLAP

- Módulo de Integração Central (*Integration Module*): módulo independente de aplicação que faz a integração entre DW e SIG através de um repositório de metadados. Uma outra função é a transformação de dados provenientes de fontes externas;
- Subsistema OLAP (*OLAP*): representa o sistema OLAP/DW e se comunica com o Módulo de Integração através de um Stub específico à aplicação;
- Subsistema SIG (*GIS*): representa o Sistema de Informação Geográfica que se adapta à arquitetura através de um Stub específico à aplicação;
- Ferramenta Front-End (*Front-End Tool*): é uma ferramenta genérica de navegação dimensional e filtragem de dados, e é usada opcionalmente quando os subsistemas SIG ou DW/OLAP não poderem executar uma determinada funcionalidade;
- Stub OLAP (*OLAP Stub*): é um *driver*, que faz a adaptação entre a interface OLAP genérica e os requisitos específicos à aplicação;
- Stub SIG (*GIS Stub*): semelhante ao Stub OLAP, mas para adaptação da aplicação SIG;
- Stub Front-End (*F-E Stub*): faz ajustes entre a ferramenta de Front-End genérica e os requisitos específicos à aplicação, e trata-se de um componente opcional;
- Metadados: repositório que armazena o modelo de dados do Data Warehouse e da fonte de dados do SIG, scripts de transformação de dados, correspondência de classe e instância entre SIG e DW e definições de políticas de segurança dos dados.

Na integração entre o SIG e um servidor DW/OLAP, foi adotado um modelo de SIG orientado a objetos, que segundo FIDALGO, TIMES e SOUZA (2001) é composto por dois

elementos básicos: *Objetos de SIG*, que representam pontos, linhas e áreas; e *Classes de SIG*, para representar grupos de objetos abstratos do mesmo tipo, por exemplo, cidade e Estado. Nessa integração, foram feitas ligações entre os elementos (ou hierarquia) da dimensão geográfica do DW com a taxonomia de objetos de um SIG. Essas ligações são feitas a partir de três tipos de correspondência: 1) *Correspondência de classe*: mapeia um determinado nível de agregação da dimensão geográfica na correspondente classe da taxonomia SIG e vice-versa. Essa informação é armazenada estaticamente no repositório de metadados. 2) *Correspondência de instância*: associa instâncias dos níveis de agregação às instâncias das classes, ou seja, os objetos geográficos correspondentes. Esta parte é feita dinamicamente, na qual a integridade dos dados é garantida em tempo de execução, sendo a manutenção e propagação das mudanças dos sistemas feita pelo módulo de integração. 3) *Correspondência de ação*: assegura a consistência na navegação entre os níveis de granularidade da dimensão geográfica no DW e a taxonomia dos dados no SIG conforme as consultas são executadas.

Um protótipo foi construído para uma aplicação de Distribuição de Água no Oeste da Boemia, na República Tcheca, que já possuía em execução sistemas comerciais de Data Warehouse e de Informação Geográfica. A avaliação foi feita baseada nos sistemas GT Media 98 e ArcView by ESRI conjuntamente com MS SQL 7.0 e seus serviços OLAP.

Uma das desvantagens dessa abordagem é a utilização de um mediador (o módulo de integração) que diminui a transparência do sistema, visto que é necessária a intervenção humana, pelo menos na manutenção de metadados, aumentando os esforços para resolver os problemas semânticos. A necessidade de buscar dados em bases independentes aumenta muito o tempo de resposta a consultas analítico-espaciais.

GOLAP é proposta como uma arquitetura genérica, mas não se baseia em padrões abertos, oferecendo apenas suporte a servidores DW/OLAP baseados em tecnologias pré-estabelecidas. Além disso, o uso de *Stubs* sobrecarrega as tarefas do módulo de integração, pois são necessários ajustes incrementais para cada nova aplicação a ser adaptada.

3.1.2 Arquitetura de Ferri et al.

A cooperação de dados entre bancos de dados geográficos e multidimensionais, segundo a abordagem federada foi proposta em (POURABBAS, 2003) e (FERRI et al., 2000), na qual se explora o elemento chave *espaço* em comum, presente no Banco de Dados Geográficos na forma de classes e objetos geográficos e no Banco de Dados Multidimensional na dimensão referente a informações espaciais. Eles propõem uma abordagem que estende a estrutura de dados geográfica através da inclusão de atributos especiais, chamados *Binding*

Attributes, a fim de descrever o comportamento e as características representadas por Bancos de Dados Multidimensionais. Com essa solução, torna-se possível responder consultas OLAP dentro de Bancos de Dados Geográficos sem modificar a organização física dos dados em ambos os ambientes, caracterizando a arquitetura federada.

Para realizar a cooperação são necessários dois tipos de correspondência, baseadas em MATOUSEK, KOUBA e MIKSOVSY (2000): 1) *Correspondência de nível*, que mapeia uma classe da taxonomia SIG no nível correspondente pertencente à dimensão localização e vice-versa; 2) *Correspondência de instâncias*, que faz o mapeamento de uma instância genérica de uma classe geográfica em sua instância correspondente no nível da dimensão localização e vice-versa.

Um *Binding Attribute* é definido como um sub-cubo derivado de um determinado cubo em um Banco de Dados Multidimensional, em que o nível geográfico corresponde à classe de mesmo nome no Banco de Dados Geográfico. Esse atributo especial consiste de um par <nome_do_cubo, {nomes_das_variáveis_do_cubo}>, no qual nome_do_cubo corresponde ao fenômeno considerado (e.g. Vendas de carro por modelo, por mês, por município) e representa as medições do cubo; o segundo, nomes_das_variáveis_do_cubo, é composto por todos os nomes das variáveis do cubo (atributos/níveis das dimensões), exceto a variável geográfica, que está implícita na classe espacial (classe Município, por exemplo). Um exemplo deste sub-cubo em que se baseia os *Binding Attributes* é visto na Figura 6 (FERRI et al., 2000).

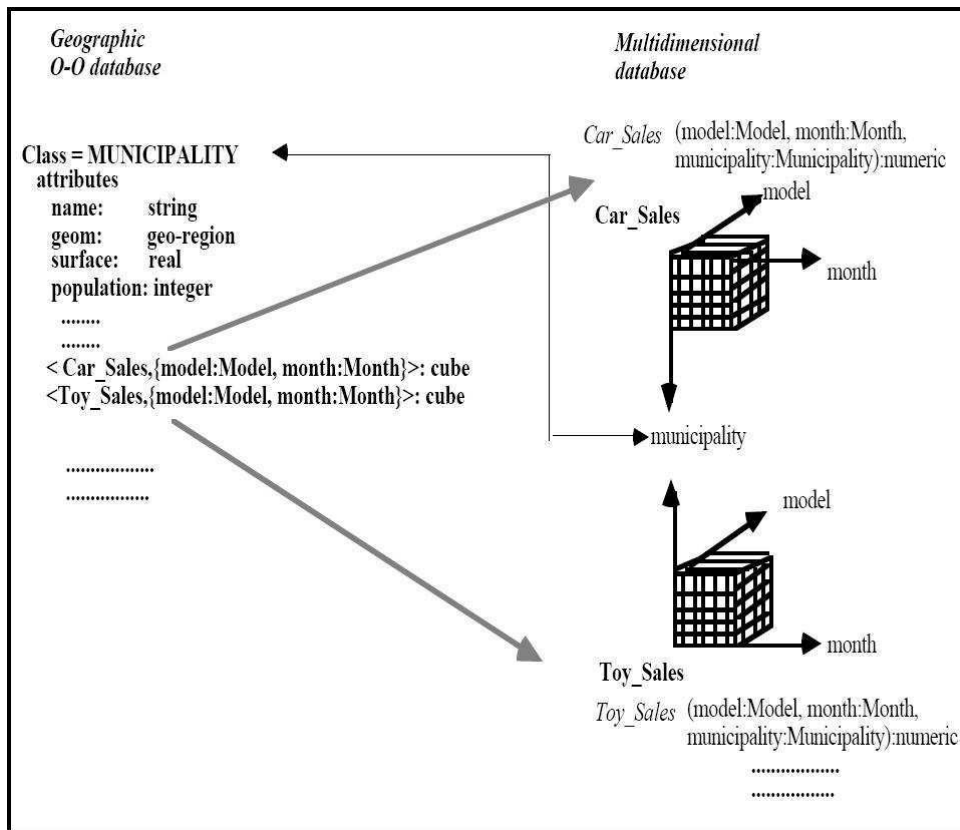


Figura 6 - Binding Attributes

A Figura 6 mostra que todo *Binding Attribute* na classe geográfica Município (*Municipality*) corresponde a um cubo do DW no qual a variável geográfica possui o nome da classe Município, ou seja, tem uma dimensão espacial contendo o nível Município. Os *Binding Attributes* $\langle \text{Vendas_de_Carros}, \{\text{modelo}, \text{mês}\} \rangle$, e $\langle \text{Vendas_de_Brinquedos}, \{\text{modelo}, \text{mês}\} \rangle$ são inclusos no esquema da classe geográfica *Município* devido à correspondência de nível feita com o nível de mesmo nome da dimensão geográfica dos cubos *Vendas_de_Carros* (*Car_Sales*) e *Vendas_de_Brinquedos* (*Toy_Sales*). Assim, a classe geográfica, que já continha atributos descritivos (e.g. nome/*name*, população/*population*) e de geometria, passa a obter o comportamento multidimensional, possibilitando a execução de consultas em função das demais dimensões do banco de dados (e.g. modelo e mês dos fatos venda de carros e venda de brinquedos). Para a correspondência de instâncias, segue-se o mesmo raciocínio, instanciando apropriadamente.

Com esta interligação dos *Bindings Attributes* é possível responder, por exemplo, à seguinte consulta analítico-espacial no Banco de Dados Geográfico: “Encontre todos os municípios vizinhos a Roma em que o número de carros (ou brinquedos) vendidos no mês de Outubro seja maior que 1000”. Todas as operações geográficas são executadas no Banco de Dados Geográficos e nos resultados são aplicados os operadores OLAP diretamente através

dos *Binding Attributes*. Além disso, é possível inferir o maior número de cubos a partir de um *Binding Attribute* básico, deduzindo novos atributos baseados em relações de pertinência e herança entre classes geográficas ou através da aplicação dos operadores OLAP: *roll-up*, *slice* e *dice*. São propostos um formalismo e um algoritmo para fazer as devidas inferências.

Quando os *Binding Attributes* são inferidos, podem ocasionar uma explosão de atributos devido à enorme combinação que os operadores OLAP podem gerar em cubos com muitas dimensões. Aliás, é deixada uma lacuna quanto ao prejuízo de desempenho causado por esta técnica para o processamento de consultas analítico-espaciais. Como também, tanto nessa proposta, quanto nas demais que utilizam abordagem federada, não há a completa integração das fontes de informação espaciais e convencionais num mesmo ambiente de DW. O que proveria melhor análise e concentraria mais esforços em operar de forma mais indutiva sobre um único modelo Multidimensional através de métodos e algoritmos a serem elaborados ou de ferramentas OLAP estendidas a um ambiente mais completo. Uma outra desvantagem dessa proposta é o fato de não apresentar nenhum protótipo de validação para as idéias propostas.

Para finalizar esta seção, resumimos no Quadro 1, os resultados da análise das abordagens apresentadas.

Quadro 1 - Avaliação de abordagens federadas

Critérios Propostas	Integração estreita entre DW e SIG?	Algum protótipo?	Otimização de consultas analítico-espaciais?
1) Matoušek et al.	Não	Sim	Não
2) Ferri et al.	Não	Não	Não

3.2 Abordagem de Integração Estreita entre DW e SIG

Este tipo de abordagem, também conhecido como abordagem integrada, utiliza integração completa, através de um esquema global ou de vários esquemas parciais dos diversos Bancos de Dados, a fim de gerar um ambiente capaz de fornecer uma visão única dos dados. Geralmente, os projetos desta abordagem aderem ao modelo de esquema estrela e fazem algumas adaptações para trabalhar com análise espacial. Estas soluções baseiam-se na visualização dos dados como a principal diferença entre Data Warehouses convencionais (tabelas sumarizadas, textos e estatísticas) e espaciais (mapas e relacionamentos topológicos),

e em realizar operações e métodos de agregação de tipos geográficos, incluindo um maior grau de complexidade na geração dos cubos de dados.

Neste contexto, serão apresentadas algumas abordagens integradas nas próximas subseções. Primeiro, descrevemos o projeto GeoMiner de integração de medidas e dimensões analíticas e espaciais. Logo em seguida, serão resumidas as características do operador Map Cube, análogo ao operador de cubo de dados agregado à funcionalidade de geração de mapas. Finalmente, apresentamos a abordagem integrada de FIDALGO, TIMES e SOUZA (2001).

3.2.1 Arquitetura de Han et al.

GeoMiner, proposto por HAN, KOPERSKI e STEFANOVIC (1997), é um projeto de Data Warehouse espacial, desenvolvido pelo Laboratório de Pesquisas de Sistemas de Bancos de Dados Inteligente da Universidade Simon Fraser do Canadá, que inclui os seguintes módulos: 1) Construção de cubo de dados espacial; 2) Processamento de consultas analíticas (OLAP) espaciais. Eles foram relatados em (HAN, KOPERSKI e STEFANOVIC, 1998) e na dissertação de mestrado de Stefanovic (STEFANOVIC, 1997).

Com a proposta de implementação de um ambiente para DW/OLAP geográfico, visando facilitar a análise de dados espaciais, foram identificados dois desafios principais: 1) *Complexidade de integração de dados espaciais de fontes e sistemas heterogêneos*: normalmente, são encontrados diferentes formatos de dados (e.g. raster vs. Vector, modelo orientado a objeto vs. relacional) e diferentes especificações de fornecedores SIG (e.g. ESRI, MapInfo); 2) *Realização de operações OLAP rápidas e flexíveis em um DWE*: que seria a implementação de algoritmos eficientes para construção de cubos de dados através da materialização (agregação) seletiva de conjuntos de objetos geográficos.

O modelo do DWE proposto por Han et al. é uma adaptação do modelo de esquema estrela para um contexto geográfico, que manipula tanto dimensões quanto medidas geográficas. Os tipos de dimensões identificados (ver Capítulo 2) são: dimensão não espacial, dimensão mista e dimensão espacial. Para complementar, distinguiram-se dois tipos de medidas (ver Capítulo 2) a serem modeladas num cubo de dados espacial: medida numérica e medida espacial.

A Figura 7, adaptada de (HAN, KOPERSKI e STEFANOVIC, 1998), mostra um modelo estrela para um Data Warehouse espacial de sensores meteorológicos distribuídos pelo Estado British Columbia do Canadá, consistindo de quatro dimensões: Temperatura, Precipitação, Tempo e Nome Da Região; e três medidas: Mapa Da Região, Área e Quantidade. Sendo que a dimensão Nome da Região é espacial, Tempo é não-espacial, e as

demais são mistas. Nessas dimensões mistas, é possível criar faixas de valores agrupados, por exemplo, no caso de Temperatura haveria uma classificação como *frio* para valores entre 0° e 18°, ou *muito frio* quando menores que 0°. Das três medidas, Mapa da Região é espacial e representa uma coleção de ponteiros para as regiões correspondentes, Área é numérica e armazena a soma total das áreas dos correspondentes objetos espaciais e Quantidade também é uma medida numérica que indica o número total de sensores acumulados em um determinado conjunto de regiões.

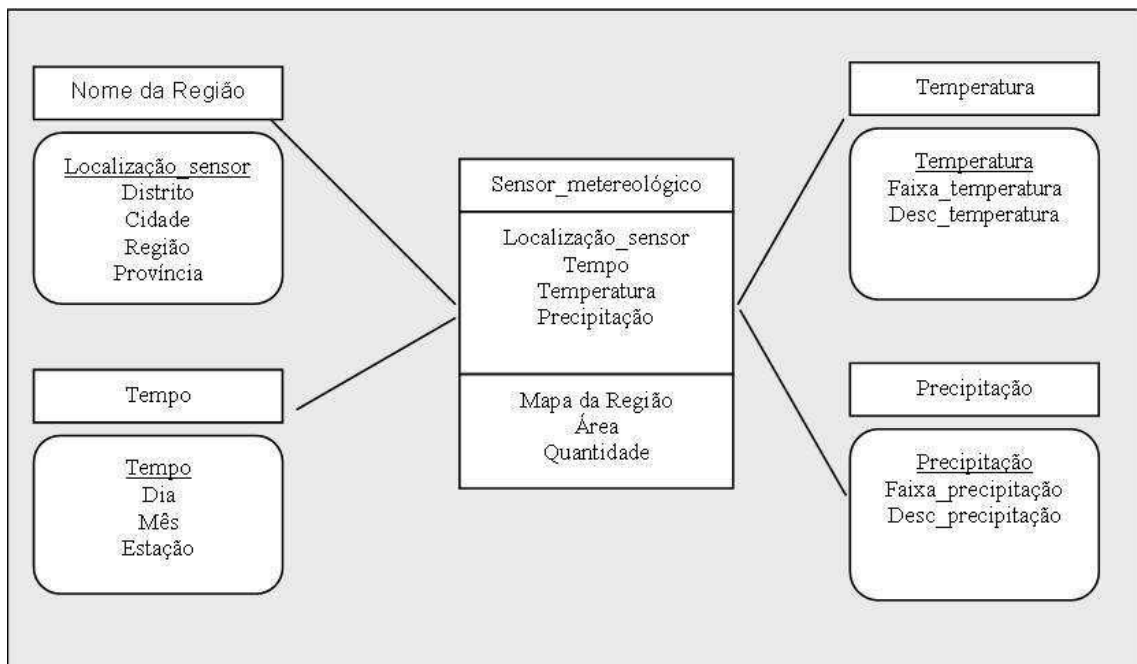


Figura 7 - DW com medidas e dimensões espaciais

As principais desvantagens dessa abordagem são: o protótipo não fornece uma interface gráfica de auxílio à criação de consultas multidimensionais espaciais; e não define formalmente o modelo multidimensional espacial. Outras questões serão discutidas posteriormente, quando serão apresentadas aplicações que utilizam operadores SOLAP.

3.2.2 Proposta de Shekhar et al.

O Map Cube, proposto por SHEKHAR et al. (2001), é um operador baseado no cubo de dados convencional estendido para Data Warehouse espacial, sendo possível visualizar os resultados tanto por tabelas como por mapas para o domínio espacial. O mesmo recebe como parâmetros um mapa “base”, associado às tabelas de dados e algumas preferências cartográficas, para produzir um álbum de mapas organizado através de hierarquias de agregação, que oferecerá apoio na exploração espacial através de operadores OLAP.

Um cubo de dados consiste de um aglomerado de visões, cada uma representando um determinado nível das hierarquias combinadas e fornecendo conjuntos de valores computados

a partir de funções de agregação. É feito um levantamento sobre os tipos de funções de agregação aplicando-as ao domínio numérico e ao espacial. São classificadas como: distributivas (e.g. Mínimo, Máximo, MBR, União, Intersecção), algébricas (e.g. Média, Desvio Padrão, Centróide) e holísticas (e.g. Mediana, Freqüência, Vizinhança).

A Figura 8 (SHEKHAR et al., 2001) faz uma comparação dos resultados do operador tradicional de cubo de dados e do Map Cube. O efeito da operação do cubo é a geração de tabelas definidas pela cláusula *Group By CUBE* na qual: se existem *n* atributos, serão obtidas 2^n tabelas. Portanto, para os dois atributos (de nome GPP e LPP) definidos na cláusula *Group By CUBE* do exemplo, haverá um total de quatro tabelas. Já o operador Map Cube não apenas gera as tabelas, mas também produz os mapas associados a cada tabela.

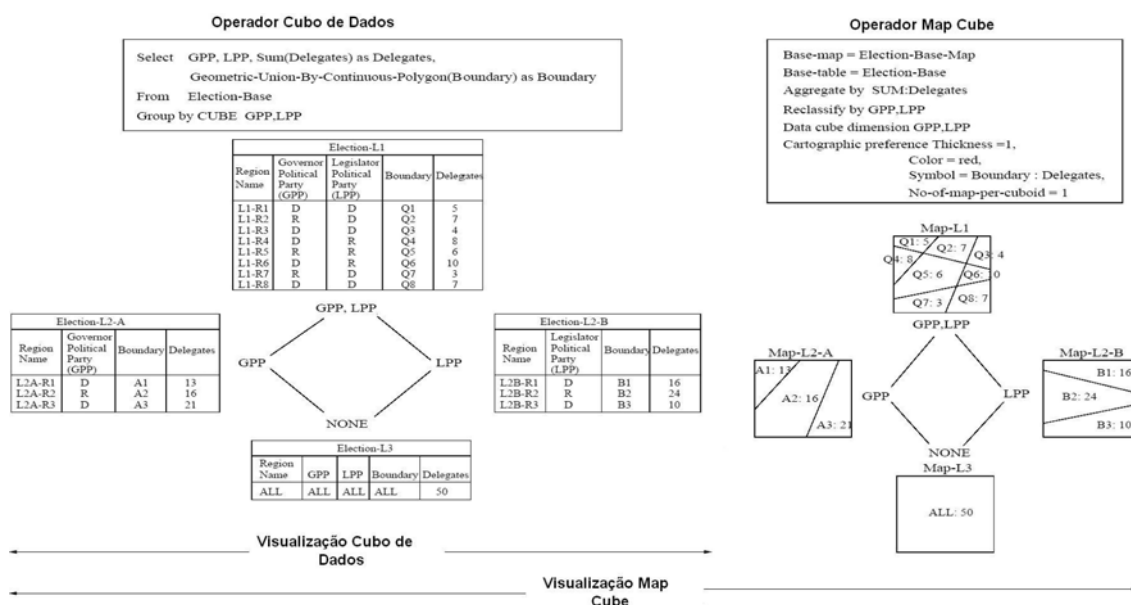


Figura 8 - Visualização do Operador Map Cube

Esta abordagem foi aplicada num estudo de caso para análise de dados de censo de regiões metropolitanas dos EUA. O Data Warehouse experimental possui as dimensões Localização Regional, Faixa Etária, Tipo de Renda e Categoria Racial que, ao serem combinadas, produzem a medida População. Ou seja, a dimensão espacial está embutida no mapa e possui seus limites definidos pela tabela associada. Através de combinações sobre grupo etário, rendimentos e etnias são gerados mapas de densidade combinados que servem de apoio à tomada de decisão.

No estudo de caso, quando combinadas duas dimensões, faixa etária com sete intervalos de classificação diferentes (e.g. menos de 25 anos, 25 a 34 anos) e nove categorias de rendimentos (e.g. menos que \$5.000, \$5.000 a \$9.999), juntamente a uma diversidade de regiões a serem ilustradas, foram produzidos sessenta e três mapas (sete intervalos de faixa

etária X nove categorias de rendimentos) com variações cartográficas em cada uma das sete regiões (amostra na Figura 9 (SHEKHAR et al., 2001)).

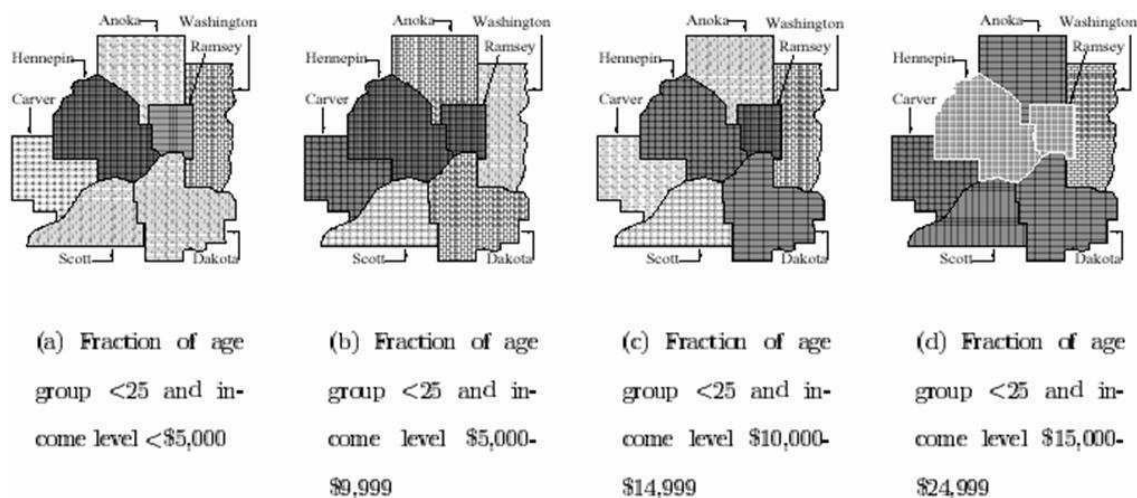


Figura 9 - Amostra de resultados do operador Map Cube sob dados censuais

As principais desvantagens dessa abordagem são as mesmas da proposta anterior: não fornece uma interface gráfica de auxílio à criação de consultas multidimensionais espaciais; e não formaliza o modelo multidimensional espacial. Além disso, esse operador ainda não foi integrado a nenhum DWE.

3.2.3 Arquitetura de Fidalgo et al.

Nessa abordagem, FIDALGO, TIMES e SOUZA (2001) propõem uma arquitetura de integração de ferramentas SIG e OLAP, denominada GOLAPA (*Geographic On-line Analytical Processing Architecture*). Essa abordagem é formada por cinco camadas (Figura 10 (<http://php.cin.ufpe.br/~golapa/architecture/>)), e apresenta um diferencial, pois leva em consideração a utilização de tecnologias abertas.

Existem três camadas para o processamento multidimensional espacial: camadas I, II e III, que disponibilizam respectivamente dados, serviços e interface com o usuário para um sistema multidimensional e espacial. As camadas restantes, ou seja, A e B, respectivamente lidam com fontes de aplicações transacionais e criam/mantêm a camada I (<http://php.cin.ufpe.br/~golapa/architecture/>). As camadas da GOLAPA são descritas a seguir.

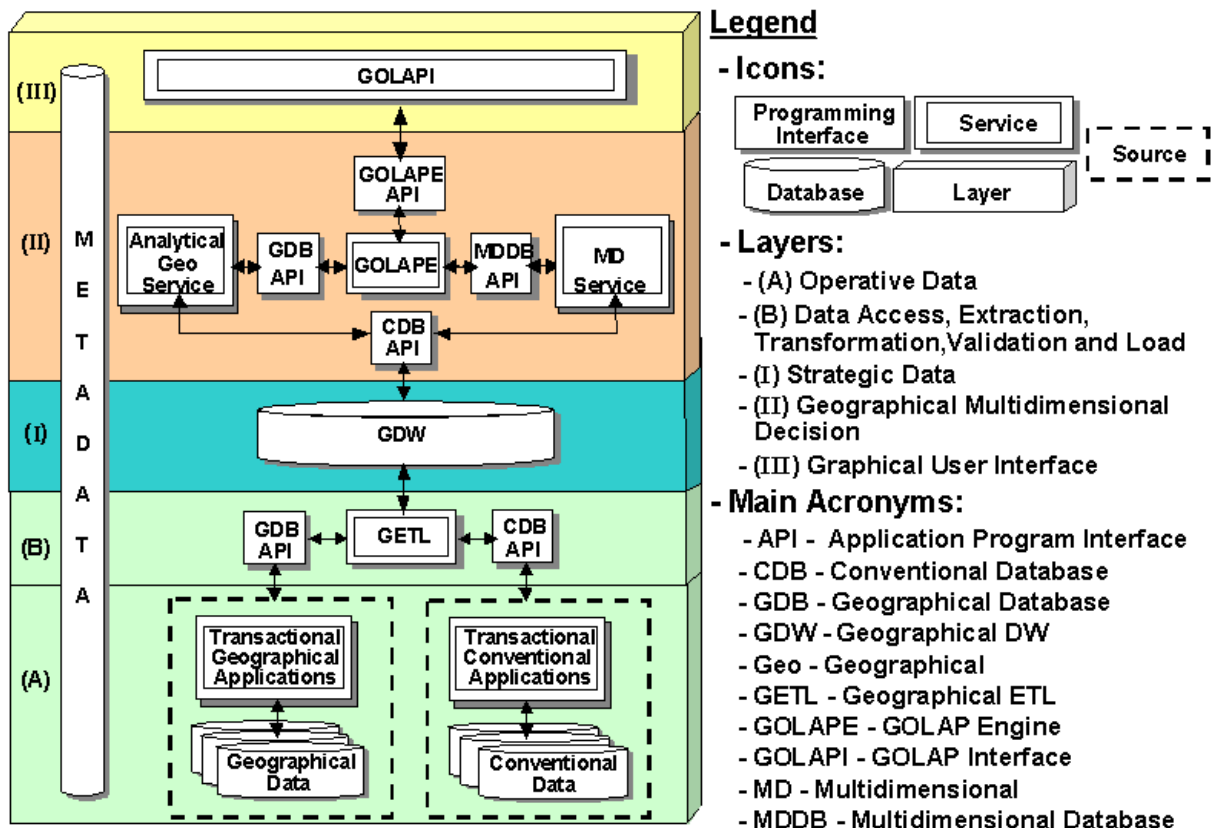


Figura 10 - Arquitetura GOLAPA

- (A) *Transactional Applications*: lida com ferramentas operativas. Corresponde aos sistemas de fontes de dados em uma arquitetura DW convencional;
- (B) *Data Access, Extraction, Transformation, Validation and Load* Acesso: extração, transformação, validação e carga de dados: camada responsável pela conversão do ambiente operacional para o de decisão;
- (I) *Strategic Data*: responsável por manter o DW espacial (*GDW*). Essa camada divide o ambiente transacional do ambiente de apoio a decisão, garantindo que qualquer operação que venha a ocorrer no ambiente transacional não seja automaticamente refletida no ambiente de apoio a decisão;
- (II) *Geographical Multidimensional Decision*: camada que manipula as funcionalidades geográficas e/ou multidimensionais;
- (III) *Graphical User Interface*: Camada de apresentação dos dados. Podendo ser tanto através de uma interface gráfica local quanto a partir de um cliente remoto ou browser Web.

SILVA, TIMES e SALGADO (2006) apresentam um avanço no desenvolvimento dessa arquitetura: trata-se de um *framework* baseado na Web e em tecnologias abertas para o

processamento geográfico e multidimensional; um Data Warehouse espacial com um esquema similar ao esquema em estrela tradicional (KIMBALL, ROSS e MERZ, 2002). É utilizado um repositório global de metadados para criação e manutenção da semântica dos dados por toda a arquitetura, cuja implementação é baseada em CWM OLAP, MOF (Meta Object Facility) (OMG, 2002) e nos metamodelos GAM (Geographical and Analytical Metamodel) e GeoMDM (Geographical Multidimensional Metamodel) apresentado em (FIDALGO et al., 2004). Também é apresentado um protótipo para validar a implementação dos módulos da arquitetura. O sistema de banco de dados utilizado para criar o DWE foi o *PostgreSQL* em conjunto com sua extensão espacial chamada *PostGis*. Foram inseridos dados multidimensionais e espaciais para o protótipo, através de *scripts* baseados na linguagem *PL/pgSQL* do *PostgreSQL*. Através da interface, o usuário pode realizar consultas multidimensionais, sendo os objetos espaciais plotados em um mapa, e os analíticos são apresentados através de tabelas de gráficos.

Uma possível desvantagem dessa arquitetura é o fato de não lidar com medidas espaciais, o que dificulta a realização de consultas de agregação de objetos espaciais. Além do mais, o modelo multidimensional espacial não é formalmente definido, e não existe uma interface para criação de consultas analítico-espaciais.

A seguir, no Quadro 2, são resumidas as principais características dessas abordagens integradas.

Quadro 2 - Avaliação de abordagens integradas

Critérios Propostas	Aderência ao modelo em esquema estrela?	Medidas espaciais?	Definição formal do Modelo Multidimensional espacial?	Interface para criação de consultas analítico-espaciais?
1) Han et al.	Sim	Sim	Não	Não
2) Shekhar et al.	Sim	Não	Não	Não
3) Fidalgo et al.	Sim	Não	Não	Não

3.3 Operações OLAP Espaciais (SOLAP)

Com as ferramentas SOLAP (ver capítulo 2), novos desafios surgem, entre eles: agregação de medições espaciais (alto custo de processamento), disponibilidade de resultados em mapas e tabelas sincronizados, navegação em hierarquias espaciais. No decorrer desta

seção, são apresentadas algumas soluções SOLAP existentes, detalhando suas principais características.

3.3.1 GEOMINER

O Geominer (HAN, KOPERSKI e STEFANOVIC, 1998) é uma abordagem integrada que adere ao modelo de esquema estrela e lida com dimensões e fatos espaciais, cuja arquitetura já foi apresentada na seção 3.2.1. Agora, focaremos nas características SOLAP.

Devido ao alto custo de agregação de objetos espaciais (e.g. através de *merge* ou *sobreposição*) e ao grande número de combinações possíveis dos níveis ao longo das dimensões, é necessária a pré-computação de algumas visões de um cubo de dados para serem armazenadas no BD como visões materializadas, facilitando a rápida geração de resultados pelas operações OLAP, apresentando mais um desafio para implementação das operações OLAP espaciais.

Para fazer o balanceamento entre o custo de computação dinâmica de um cubo e a sobrecarga de armazenamento de agregados espaciais foram propostas nessa abordagem técnicas de materialização seletiva de cubos de dados espaciais através dos algoritmos Interseção de Ponteiro e Conexão de Objeto.

Foram considerados alguns critérios na seleção de um cubo para materialização:

- A potencial frequência de acesso ao cubo gerado;
- O tamanho do cubo gerado;
- Benefícios para a computação de outros cubos.

O algoritmo Interseção de Ponteiros procura encontrar aqueles grupos de objetos espaciais no cubo de dados que são acessados mais frequentemente ou que executam muitas operações espaciais, e então pré-computam essas operações para tornar as consultas mais eficientes.

Já o algoritmo Conexão de Objetos diferencia-se do anterior pela manipulação e identificação de objetos conectados, garantindo a detecção de grupos de objetos espaciais que se repetem pela combinação de outros grupos. Isso permite a melhoria de desempenho na escolha das visões a serem pré-computadas em certas ocasiões, comprovado pelo teorema enunciado em (STEFANOVIC, 1997).

Contudo, apesar de fazer uma análise de desempenho sobre a qualidade dos algoritmos, e conseqüentemente, suas melhores aplicabilidades, não há uma proposta que integre ambos a fim de resolver o problema de materialização seletiva de objetos espaciais em diferentes situações. A idéia de utilizar medidas espaciais no esquema estrela, e que originou a

elaboração dos algoritmos, acrescenta complexidade a uma aplicação OLAP. A mesma medida poderia ser modelada ao longo de dimensões espaciais, concentrando esforços na melhor utilização dos operadores analíticos sobre as funcionalidades geográficas. Ainda, a informação de frequência de acesso necessária nos algoritmos nem sempre poderá estar disponível, visto que as mesmas devem ser estimadas, pois os cubos ainda não foram implementados. Ao invés de dispor de dois algoritmos com a mesma técnica, seria interessante usar algoritmos que exigissem diferentes critérios.

Nos Data Warehouses Espaciais, as operações OLAP tradicionais, como Slice, Dice, Drill-Down e Roll-up devem ser estendidas para lidar com informações espaciais. No caso do Geominer, é apresentada uma modelagem que considera essas novas operações OLAP espaciais. Por exemplo, a utilização de Drill-down ou Roll-up em um cubo de dados espacial, resulta em diferentes cubóides nos quais as células contêm a agregação de valores numéricos ou então uma coleção de apontadores para objetos espaciais. Se estes apontadores referenciarem objetos espaciais que estejam interconectados, pode ser feito um “merge” dessas geometrias, resultando em um novo objeto espacial que representa uma agregação espacial; caso contrário, o resultado será uma coleção de regiões isoladas.

Para melhor entender como funciona uma agregação espacial, vamos apresentar um exemplo de uma operação Roll-up espacial (Figura 11 (HAN, KOPERSKI e STEFANOVIC, 1998)): o modelo é o mesmo da Figura 7, que apresenta um modelo estrela para um Data Warehouse espacial de sensores meteorológicos distribuídos pelo Estado British Columbia do Canadá, consistindo de quatro dimensões: Temperatura, Precipitação, Tempo e Nome Da Região (dimensão espacial); e três medidas: Mapa Da Região, Área e Quantidade. À esquerda da Figura 11, podemos ver um mapa apresentando as localizações de todos os sensores meteorológicos; e a direita podemos ver o resultado de um Roll-up espacial para o nível de “Região”, sob critérios de uma determinada dimensão. Esse novo mapa apresenta regiões que são formadas a partir do “merge” de uma grande quantidade de áreas onde se encontram os sensores, sendo que diferentes critérios da operação Roll-up poderiam gerar diferentes mapas, já que os mesmos resultam da agregação de pequenos objetos espaciais. Os autores ressaltam que esse mapa só pode ser gerado em tempo aceitável se utilizar o recurso da pré-computação, ou seja, alguns “merges” espaciais já deveriam ter sido pré-calculados.

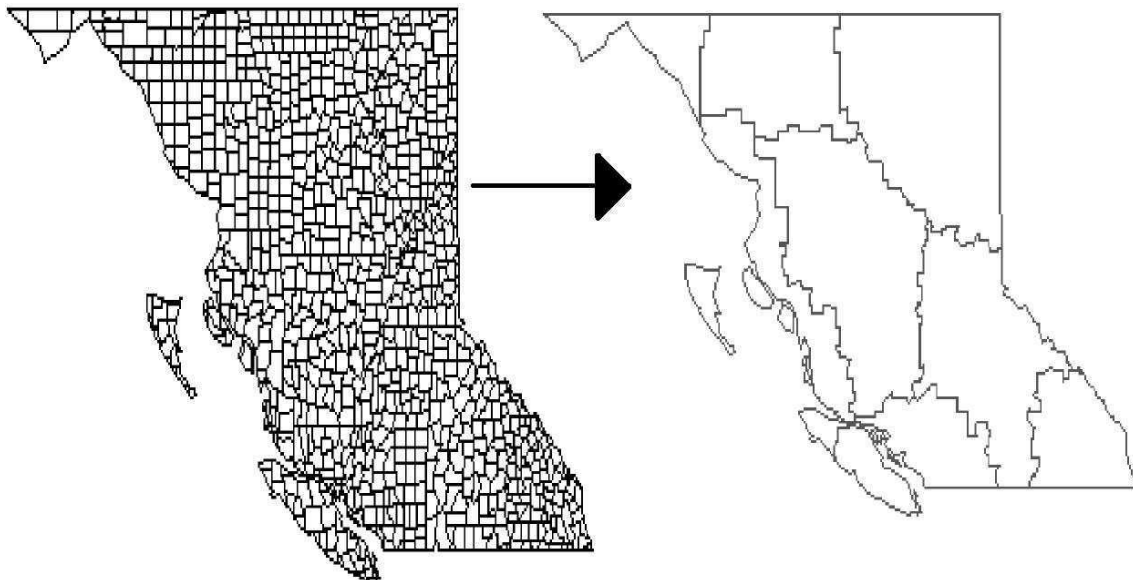


Figura 11 - Operação Roll-up Espacial

O escopo de interação das consultas não se restringe apenas a regiões pré-definidas do mapa, mas estas podem também ser determinadas dinamicamente (através de consultas ad-hoc), como exemplifica a seguinte questão: “Destaque as regiões dentro da janela de consulta (uma área específica do mapa definida pelo usuário) com maior precipitação pluviométrica entre os anos de 1999 e 2001”. Nestes casos, o agrupamento de objetos espaciais é definido arbitrariamente pelo usuário em tempo de execução. Portanto, não é possível prever a hierarquia espacial a priori, e conseqüentemente, nem propor materializações tradicionais. De forma a não provocar alterações no esquema dos dados, o mecanismo proposto por (PAPADIAS et al., 2001) deve ser estendido e incorporado para solucionar este problema de consulta. Apesar de fazer uma pré-computação através de materialização seletiva, o Geominer não leva em consideração as consultas ad-hoc, cujos critérios de agregação são definidos dinamicamente.

3.3.2 Map Cube

O Map Cube (SHEKHAR et al., 2001) é um operador espacial, cuja arquitetura já foi apresentada na seção 3.2.2. Ele permite visualizar os resultados das consultas através de um conjunto de mapas e também em forma tabular. Questões como a materialização de visões para otimizar a resposta a consultas e a inclusão de novas funções de agregação, principalmente espaciais, são levantadas como pesquisas futuras. E não é abordada a questão de consultas ad-hoc.

No Map Cube, observa-se que, apesar de prover uma forma mais dedutiva de análise espacial (através de mapas), a procura de padrões, quando da combinação de um grande

número de dimensões, pode ser inviável. Assim, seria possível imaginar os problemas agravados pelo aumento da quantidade de dimensões de um cubo a ser analisado visualmente. Além disso, esse operador ainda não foi integrado a nenhum DWE.

Essa proposta não trata operadores de agregação espacial como Roll-up e Drill-down espacial.

3.3.3 SOVAT

SCOTCH e PARMANTO (2005) desenvolvem um sistema chamado SOVAT (*Spatial OLAP Visualization and Analysis Tool*), ou ferramenta de análise e visualização OLAP para dados espaciais. Uma ferramenta que pode lidar com uma grande quantidade de dados, executar operações espaciais e cálculos estatísticos, e disponibilizar essas informações tanto em mapas quanto de forma descritiva, na mesma interface.

O desenvolvimento da interface do SOVAT consiste de dois componentes principais: 1) Uma interface gráfica com interação baseada em *drag and drop* (operações simples com o mouse) para esconder do usuário os detalhes inerentes à consulta; 2) Integração OLAP e SIG para criar um dispositivo único de visualização dos dados, e permitir consultas com diferentes representações visuais desses dados.

Eles fizeram um estudo de caso para analisar algumas taxas relacionadas à saúde da comunidade rural da Pensilvânia. As dimensões escolhidas para o DW foram: Idade, Sexo, Raça, Nível de educação, Região (urbana ou rural), Peso, Geografia e Ano. Na Figura 12 (SCOTCH e PARMANTO, 2005), pode ser visto o resultado de uma consulta através de mapas e também de gráficos referentes aos dados numéricos. Neste exemplo, é mostrada a taxa de mortalidade de uma doença (dimensão Diagnose) específica, para uma faixa etária específica, durante um determinado ano, para as localizações geográficas que foram selecionadas no mapa.

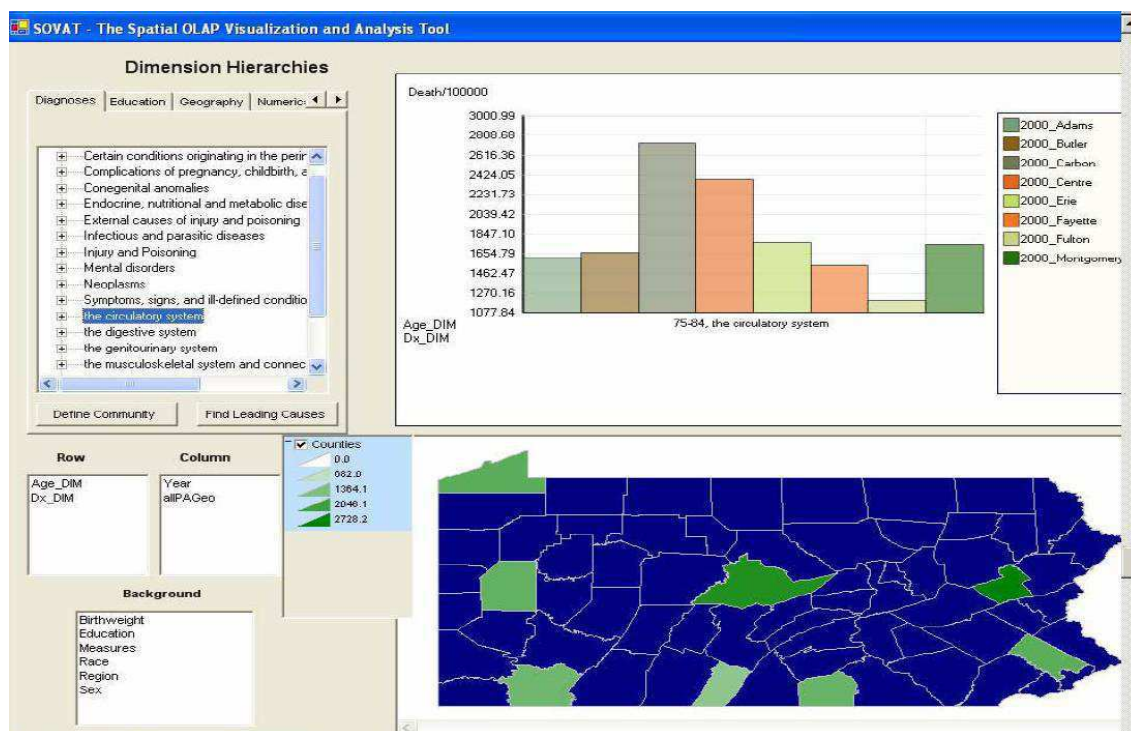


Figura 12 - Interface do SOVAT

A interface do SOVAT não é Web, mas permite que sejam realizadas operações OLAP Espaciais, como Drill-Down espacial, por exemplo, diretamente na sua interface, não tratando a questão de pré-computação para garantir eficiência neste tipo de operações e nem a realização de consultas ad-hoc.

3.3.4 ZHANG et al.

Por fim, temos a proposta de ZHANG et al. (2003a) e (2003b), que trata-se de uma abordagem de integração estreita entre DW e SIG, para permitir SOLAP através de agregação em hierarquias espaciais, que podem ser estabelecidas automaticamente a partir dos dados espaciais.

Nessa arquitetura, é preservado o esquema estrela do Data Warehouse, e existe uma indexação espacial baseada em R-trees (GUTTMAN, 1984), que possibilita relacionamento aninhado entre nodos de alto nível e de baixo nível, para modelar a hierarquia da dimensão espacial. O esquema é estendido pela introdução de predicados e funções espaciais, que expressam explicitamente o relacionamento espacial entre dados na tabela de fatos e dimensões. Um mecanismo de indexação espacial (Spatial Index Engine) é empregado para derivar a hierarquia espacial, pré-agregar resultados e materializar seletivamente, possibilitando respostas rápidas às consultas SOLAP.

O mecanismo de indexação espacial (*Spatial Index Engine*) é responsável pela recuperação de todos os nodos da R-tree, armazenada em *Spatial Index*, cujos MBR satisfazem os predicados espaciais extraídos pelo Processador de Consultas SOLAP (*Spatial OLAP Query Processor*) das consultas feitas pelo usuário. Com estes nodos retornados, o Processador de Consultas reescreve a consulta com os valores a serem enviados ao mecanismo relacional de processamento OLAP.

A exploração de tais árvores (R-tree), ao processar uma consulta OLAP espacial para encontrar o cubóide do DW, pode ser complexa, por isso também é proposto um método de pesquisa direcionado ao processamento analítico (OLAP-Favored) que retorna apenas os nodos intermediários em que todas as tuplas descendentes indexadas pelo nodo satisfazem à consulta, através de pré-agregação. Ainda, é proposto um método de pesquisa direcionado ao processamento analítico através de heurística que faz algumas estimativas do resultado final no caso de nodos intermediários que são sobrepostos parcialmente através de um fator de estimativa. Este método por heurística é mais eficiente que o primeiro, pois não necessita fazer tantas visitas quanto o método OLAP-Favored, no entanto, a acurácia é menor.

Essa pré-agregação dos dados pode responder às consultas ad-hoc, nas quais o usuário restringe uma área do mapa. Por exemplo, a seguinte questão poderia ser respondida sem problemas: “Qual o total de vendas de cada tipo de combustível para os clientes do tipo “taxista”, para todos os postos de gasolina dentro de uma *janela de consulta*, em outubro de 2002?”, sendo a *janela de consulta* um retângulo que o usuário desenha no mapa.

A atualização de dados exige novo processamento não só no próprio Data Warehouse como também na *Spatial index engine*, que apesar de ser um módulo facilmente integrado a qualquer ambiente de DW, pode acarretar em perda de desempenho do sistema como um todo. São tratados objetos espaciais apenas da forma ponto, mas é mencionado como trabalhos futuros, o processamento de linhas e polígonos. E apenas é dito que essa abordagem pode ser integrada a outros DW para fornecer análise espacial, mas não existe nenhuma interface gráfica.

Uma possível restrição desse modelo multidimensional é o fato de não lidar com medidas espaciais. Medida espacial é uma forma de introduzir informação espacial no processo de tomada de decisão, ou seja, usando-a como um objeto de análise ou como um fato. Assim, medida espacial pode ser analisada através de dimensões espaciais e analíticas.

A seguir, no Quadro 3, encontram-se as principais características das propostas apresentadas nesta seção.

Quadro 3 - Avaliação de abordagens SOLAP

Critérios	Pré-computação de mapas?	Consultas ad-hoc?	Drill-down e Roll-up espaciais?
Propostas			
Geominer	Materialização seletiva	Não	Sim
Map Cube	Não	Não	Sim
SOVAT	Não	Não	Sim
Zhang et al.	Pré-agregação de mapas, armazenando em índices R-Tree	Sim. Através de janelas de consulta.	Sim

3.4 Outros Trabalhos

A criação de operações OLAP em DWE é um grande desafio. Trabalhos recentes em DWE (BIMONTE, TCHOUNIKINE e MIQUEL, 2005), (DAMIANI e SPACCAPIETRA, 2006) e (MALINOWSKI e ZIMANYI, 2004) tratam apenas de modelos conceituais e não discutem aspectos relacionados com a implementação do modelo; em particular, a questão da implementação eficiente das operações Roll-up espacial e Drill-down espacial (operações OLAP espaciais) permanece em aberto. DAMIANI e SPACCAPIETRA (2006) apresentam um novo modelo multidimensional espacial para objetos espaciais. A inovação desse modelo é a representação de medidas espaciais em diferentes níveis de granularidade espacial. O modelo ainda inclui um conjunto de operadores OLAP capazes de navegar através das dimensões e dos níveis das medidas espaciais. BIMONTE, TCHOUNIKINE e MIQUEL (2005) definem os requisitos para um modelo de dados multidimensional e espacial, ou seja, que integra estreitamente medições analíticas e espaciais. O modelo multidimensional é capaz de lidar com objetos complexos como medidas e funções de agregação ad-hoc, com intuito de manipular dados geográficos. Entretanto, o modelo é apenas conceitual, o que dificulta uma validação do mesmo. MALINOWSKI e ZIMANYI (2004) apresentam um modelo multidimensional baseado no paradigma de modelagem Entidade Relacionamento. O modelo é uma extensão de um modelo conceitual multidimensional, que representa medidas espaciais, hierarquias espaciais e dimensões espaciais, entretanto foca apenas na representação das mesmas.

Visões materializadas podem garantir o bom desempenho no processamento de consultas, especialmente para consultas de agregação sobre bancos de dados extensos. Para tal, o otimizador de consultas deve saber como e quando tirar proveito das visões materializadas. GOLDSTEIN e LARSON (2001) apresentam um algoritmo rápido e escalável para determinar se parte ou toda a consulta pode ser reescrita em termos de visões materializadas e descreve como essa transformação pode ser incorporada pelos otimizadores. Infelizmente, o algoritmo apenas considera medidas numéricas. Manipulação de medidas espaciais é um desafio, já que objetos geométricos tendem a ser volumosos. Desta forma, está em aberto o problema de criar eficientemente visões materializadas para objetos espaciais sem gerar o problema de explosão de agregados (MALINOWSKI, 2003).

A seguir, no Quadro 4, é apresentado um resumo das principais lacunas de pesquisa encontradas na revisão bibliográfica.

Quadro 4 - Lacunas de pesquisa em DWE

Características / Propostas	Abordagem Federada X Integrada	Modelo Multidimensional Espacial	Definição formal de Modelo Multidimensional Espacial	Operações SOLAP	Medidas espaciais	Agregação Espacial	Otimização de consultas SOLAP	Protótipo de validação	Implementação em SGBD OR	Interface para criação de consultas	Cosultas ad-hoc
Matousek et al.	Federada	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não
Ferri et al.	Federada	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
Han et al.	Integrada	Não	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Não
Shekhar et al	Integrada	Não	Sim	Não	Não	Não	Sim	Sim	Não	Não	Não
Bimonte et al.	Integrada	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Não
Zhang et al.	Integrada	Não	Sim	Não	Não	Não	Sim	Sim	Não	Não	Sim
Fidalgo et al.	Integrada	Não	Sim	Não	Não	Não	Não	Sim	Não	Não	Sim

A maioria das propostas de integração entre DW e SIG não definem formalmente um modelo multidimensional espacial e suas respectivas operações.

Com relação às propostas que apresentam interface gráfica, nenhuma delas fornece uma interface gráfica que possibilite ao usuário definir consultas analítico-espaciais através da utilização de menus interativos, ou seja, de tal forma que o usuário não necessite saber detalhes de sintaxe da linguagem de consulta ao DWE. Poucas propostas possibilitam que o usuário realize operações SOLAP como Drill-down e Roll-up espacial, para navegar nas hierarquias espaciais.

Apenas duas propostas (ZHANG, 2003) (FIDALGO et al., 2004) possibilitam ao usuário definir consultas espaciais ad-hoc, nas quais critérios espaciais são definidos dinamicamente, por exemplo, quando o usuário deseja obter objetos espaciais que estejam dentro de uma janela espacial dentro de um mapa.

A maioria das propostas de integração entre DW e SIG não tratam o objeto espacial como objeto de análise (como medida espacial), mas apenas como critérios de agregação (dimensões espaciais) para medidas analíticas. Desta forma, não é possível a realização de operações de agregação espacial (e.g. união de geometrias), muito importantes no auxílio à tomada de decisão. Entre as propostas que tratam de medidas espaciais, poucas mencionam a questão de otimização de consultas espaciais, necessária para garantir o bom desempenho de consultas que fazem operações de agregação espacial.

Acreditamos que ainda não existe nenhuma proposta de integração entre DW e SIG com uma implementação baseada em um esquema em estrela espacial Objeto Relacional, que permite a utilização de referências para objetos, de tal forma que objetos espaciais volumosos não precisem ser duplicados na base de dados do DWE, aumentando assim a performance de consultas SOLAP.

3.5 Considerações Finais

Este capítulo apresentou o estado da arte referente ao tema de integração entre sistemas SIG e DW. Durante a análise das propostas de integração foram identificadas as principais características de cada uma delas, focando nas lacunas de pesquisa ainda existentes. No final do capítulo foi apresentado um quadro com o resumo das lacunas de pesquisa identificadas.

Capítulo 4

Modelo de Banco de Dados Multidimensional Espacial

Neste capítulo, apresenta-se a proposta de um modelo multidimensional para DW espacial, em que os principais conceitos são: *dimensão espacial*, *medida* (ou *fato*) *espacial*, e *hierarquia espacial*. Esses conceitos, informalmente discutidos no Capítulo 2, são agora rigorosamente definidos, com as dependências entre eles caracterizando uma estrutura chamada *cubo de dados espacial*. Os aspectos comportamentais consistem essencialmente em operações de análise multidimensional, no contexto de um cubo de dados espacial: “*roll-up*”/“*drill-down*” *espacial*, “*slice*”/“*dice*” *espacial* e “*drill across*” *espacial* (chamadas de operações SOLAP — “Spatial On-Line Analytical Processing”) (RIVEST, BÉDARD e MARCHAND, 2001). Essas operações são descritas de maneira informal. Ainda são definidas formalmente regras de mapeamento do modelo multidimensional espacial para um esquema em estrela espacial Objeto Relacional, que seja implementável em qualquer SGBD OR capaz de lidar com informações espaciais.

4.1 Formalização dos Principais Conceitos do Modelo

Os principais conceitos do nosso modelo multidimensional espacial orientado a objeto são as classes: *hierarquia espacial*, *dimensão espacial*, *medida espacial* e *cubo espacial*. Os mesmos serão definidos formalmente a seguir. Nas definições a seguir, $\{X\}$ denota um conjunto de classes X .

Uma classe *hierarquia espacial* é composta de uma ou mais classes e de seus relacionamentos topológicos. Formalmente:

Definição 1 - Hierarquia Espacial (HE)

Uma classe *hierarquia espacial* HE é uma tupla $(\{NE\}, NE_{Top}, NE_{Bottom}, \{NE_i > NE_{i-1}, i = (Bottom+1), \dots, Top\})$, na qual NE (nível espacial) é uma classe geométrica, com NE_{Top} e NE_{Bottom} sendo respectivamente as classes superior e inferior da hierarquia, e $>$ é um relacionamento topológico (*INSIDE* ou *CONTAINS*) $I:N$ entre duas classes NE, de NE_{Top} até NE_{Bottom} .

Uma classe *dimensão espacial* é um critério de agregação para classes *medida espacial* — ver definição 3 — e classes *medida numérica (não-espacial)*, com pelo menos uma classe de hierarquia espacial. Formalmente:

Definição 2 - Dimensão Espacial (DE)

Uma classe *dimensão espacial* DE é uma tupla ($\{HE\}, \{OC\}$), na qual OC é qualquer classe que não seja uma classe HE.

Definição 3 - Medida Espacial (ME)

Considere $\sim DE$ uma classe *dimensão não-espacial*, G uma classe geométrica, e GDF (*geometric-dimension function*) um relacionamento N:1 entre G e uma classe de dimensão $D \in \{DE\} \cup \{\sim DE\}$. Uma classe *medida espacial* ME é uma tupla ($G, \{GDF\}$).

Um cubo de dados espacial representa um *array* com n dimensões, no qual cada célula do array é um conjunto de classes *medida espacial* e de classes *medida não-espacial*, enquanto as dimensões do array são compostas por classes *dimensão espacial* e classes *dimensão não-espacial*. Formalmente:

Definição 4) Cubo de Dados Espacial (CE)

Considere $\sim ME$ uma classe *medida não-espacial*. Um cubo de dados espacial CE é uma tupla ($(\{DE\} \cup \{\sim DE\}), (\{ME\} \cup \{\sim ME\})$).

Um esquema CE é uma especificação formal de classes CE. Para tal, utilizamos o formalismo UML e ODL/ODMG. Em um esquema UML/ODL SDC, cada DE, $\sim DE$, ME e $\sim ME$ é especificado através de uma classe UML (formalismo gráfico) e uma classe ODL/ODMG (formalismo analítico). Formalmente:

Definição 5) Esquema UML/ODL de um CE

Considere UML/ODL $DE(\sim DE)(ME)(\sim ME)$ ser uma classe UML/ODL especificando uma classe CE $DE(\sim DE)(ME)(\sim ME)$, então, esquema UML/ODL $CE = \{UML/ODL DE\} \cup \{UML/ODL \sim DE\} \cup \{UML/ODL ME\} \cup \{UML/ODL \sim ME\}$.

Nosso paradigma SOLAP estende operadores OLAP ‘convencionais’ para funções de agregação de medições espaciais ao longo de dimensões espaciais ou não-espaciais. Os operadores SOLAP considerados, roll-up/drill-down espacial e slice/dice espacial, podem ser descritos informalmente:

Roll-up/Drill-down espacial A ação desta função é de gerar um conjunto de agregações espaciais ao longo de uma hierarquia espacial em um cubo de dados espacial.

Slice/Dice Espacial A saída desta função é um novo cubo de dados obtido a partir da remoção daqueles valores do cubo inicial que não satisfazem o predicado (restrição) de entrada sobre a dimensão espacial.

Funções de agregação para objetos espaciais já foram definidas, incluindo *convex hull*, *geometric union* e *geometric intersection* (INDULSKA e ORLOWSKA, 2002), (SHEKHAR e CHAWLA, 2003): o resultado dessas funções pode ser representado por geometrias simples ou complexas.

4.2 Esquema Conceitual e Consultas SOLAP

O modelo multidimensional espacial é a fundamentação teórica para a definição de um meta-esquema orientado a objeto (OO) de DW espacial, segundo o formalismo UML/ODMG. UML e ODMG se complementam da seguinte maneira nessa definição: UML é utilizada como uma linguagem gráfica OO para a definição de classes de objeto, enquanto o padrão ODMG de banco de dados OO é uma linguagem textual que contempla tanto os aspectos estruturais (classes ODL/ODMG), quanto os aspectos comportamentais (linguagem de consulta OQL/ODMG, sobre classes ODL/ODMG). Desta forma, um esquema conceitual adequado a DW espacial prescreve classes UML/ODL para representar dimensões, hierarquias e medidas espaciais, ao passo que operações SOLAP são definidas em termos de expressões OQL sobre o esquema UML/ODMG.

Para a validação dessas idéias, um estudo de caso sobre agricultura brasileira foi resolvido com um enfoque de DW espacial. A aplicação abrange tanto dimensões analíticas como espaciais, medidas numéricas e espaciais, bem como dados temporais. A Figura 13 mostra a parte UML do esquema de DW espacial para o estudo de caso, que seria a definição do esquema conceitual de um cubo de dados espacial, UML CE. Denotando (X) como um objeto da classe X, a Figura 13 apresenta um objeto de cada uma das classes: Estado, Mesoregiao, Microregiao, Localizacao.Geometria_Municipio e AreaDePlantacao. As classes UML do esquema da Figura 13 são as seguintes: DE = Localizacao, {~DE} = {Tempo, Plantacao, Solo, Pluviometria}, {NE}={Estado, Mesoregiao, Microregiao, Localizacao.Geometria_Municipio}, HE = Estado *contains* Mesoregiao *contains* Microregiao *contains* Localizacao.Geometria_Municipio (sendo *in* o inverso de *contains*), ME = (areaDePlantacao, {noTempo, dePlantacao, noSolo, comPluviometria, naLocalizacao}).

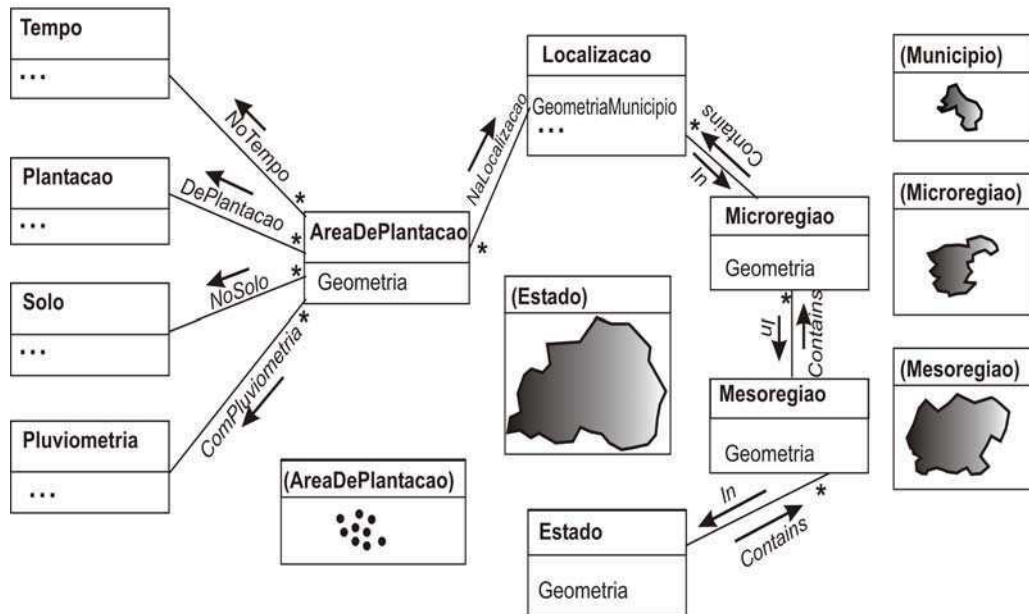


Figura 13 - Esquema CE AgroDistribuicao

Os scripts ODL/ODMG para definição das classes UML do esquema CE AgroDistribuicao são apresentados a seguir.

-- Criando a classe Tempo, para a dimensão tempo.

```
class Tempo
(extent TempoTable key tempo_id)
{
  attribute Date tempo_id;
  attribute String dia;
  attribute integer mes;
  attribute integer trimester;
  attribute integer semestre;
  attribute integer ano;
}
```

-- Criando a classe Solo, para a dimensão Solo

```
class Solo
(extent SoloTable key solo_id) {
  attribute integer solo_id;
  attribute String descricao;
  attribute String categoria;
}
```

-- Criando a classe Plantacao, para a dimensão Plantacao

```

class Plantacao
(extent PlantacaoTable key plantacao_id) {
attribute integer plantacao_id;
attribute String nome;
attribute String tipo;
}
-- Criando a classe Pluviometria, para a dimensão Pluviometria
class Pluviometria
(extent PluviometriaTable key pluviometria_id) {
attribute integer pluviometria_id;
attribute String faixa_indice;
attribute String classificacao;
}
-- Criando a classe Localizacao, para a dimensão espacial Localizacao
class Localizacao
(extent LocalizacaoTable key localizacao_id) {
attribute integer localizacao_id;
attribute String nome_municipio;
attribute Geometry GeometriaMunicipio;
relationship Microregiao In inverse Microregiao::Contains;
}
-- Criando a classe Microregiao, para o nível espacial Microregiao
class Microregiao
(extent MicroregiaoTable key microregiao_id) {
attribute integer microregiao_id;
attribute String nome_microregiao;
attribute Geometry Geometria;
relationship Set<Localizacao> Contains inverse Localizacao::In;
relationship Mesoregiao In inverse Mesoregiao::Contains;
}
-- Criando a classe Mesoregiao, para o nível espacial Mesoregiao
class Mesoregiao
(extent MesoregiaoTable key mesoregiao_id) {
attribute integer mesoregiao_id;
attribute String nome_mesoregiao;
}

```

```

attribute Geometry Geometria;
relationship Set<Microregiao> Contains inverse Microregiao::In;
relationship Estado In inverse Estado::Contains;
}
-- Criando a classe Estado, para o nível espacial Estado
class Estado
(extent EstadoTable key estado_id) {
attribute integer estado_id;
attribute String nome_estado;
attribute Geometry Geometria;
relationship Set<Mesoregiao> Contains inverse Mesoregiao::In;
}
-- Criando a classe AreaDePlantacao, para as medidas espaciais.
class AreaDePlantacao
(extent AreaDePlantacaoTable) {
--apontadores para as dimensões
relationship Tempo NoTempo;
relationship Solo NoSolo;
relationship Plantacao DePlantacao;
relationship Pluviometria ComPluviometria;
relationship Localizacao NaLocalizacao;
-- medidas
attribute Geometry geometria;
}

```

Depois de definido o esquema conceitual, vamos ver um exemplo de como podemos definir uma função SOLAP em OQL/ODMG sobre um cubo de dados espacial. Suponha a seguinte questão sobre o esquema CE AgroDistribuicao da Figura 13: *Recuperar as áreas de plantação de milho para cada microregião (mesoregião) e para cada mesoregião (microregião) do Estado da Paraíba, durante Maio de 2003*. Essa consulta pode ser interpretada como um Roll-up (Drill-down) espacial ao longo do subconjunto Mesoregiao Contains Microregiao da hierarquia espacial HE. Uma possível formulação OQL desta consulta seria como a seguir, denotando AreasDePlantacoes a extensão da classe AreaDePlantacao.

```

(Select Microregiao, Geometric_union (Select p.a From Partition p)
From AreaDePlantacao a

```

```

Where a.noTempo.Mes=5 And a.noTempo.Ano=2003
      /*Navegação pela hierarquia Tempo, não-espacial */
And a.dePlantacao.Nome='milho' And
a.naLocalizacao.In.In.Nome='Paraíba'
Group by Microregiao: a.naLocalizacao.In.Geometria)
      /* Navegação pela hierarquia espacial Localizacao */
UNION
(Select Mesoregiao, Geometric_union (Select p.a From Partition p)
From AreaDePlantacao a
Where a.noTempo.Mes=5 And a.noTempo.Ano=2003
      And a.dePlantacao.Nome='milho' And
      a.naLocalizacao.In.In.Nome='Paraíba'
Group by Mesoregiao: a.naLocalizacao.In.Geometria)

```

Observe que a consulta navega pela hierarquia espacial HE através dos seus N:1 relacionamentos inversos. A consulta também navega por duas hierarquias não espaciais no sentido inverso, Dia → Mês → Ano e Nome_Município → Nome_Microregião → Nome_Região → Nome_Estado. Finalmente, note que as agregações espaciais geradas pelo primeiro (segundo) Select e aquelas geradas pelo segundo (primeiro) Select, nesta ordem, correspondem a navegação do nível espacial Microregião (Região) para o nível espacial Região (Microregião), caracterizando um Roll-up (Drill-down) espacial, operação SOLAP, sobre o cubo de dados espacial.

4.3 Mapeamento do Modelo Conceitual para um Modelo Lógico Objeto Relacional

Para a efetiva implementação do esquema UML/ODMG conceitual, foi definido um conjunto de regras de transformação de elementos do meta-esquema UML/ODMG para elementos correspondentes do modelo objeto-relacional (OR), visando a derivar automaticamente um esquema Objeto Relacional (OR) implementável de um esquema conceitual UML/ODMG.

O mapeamento é definido da seguinte forma: cada classe UML/ODL de dimensão espacial (não espacial) é mapeada para um tipo OR distinto; o conjunto de classes UML/ODL de medições espaciais (não espaciais) é mapeado para outro tipo OR; e as extensões (*extents*) dos tipos tornam-se as respectivas tabelas OR. Esse mapeamento gera um esquema OR que é similar ao esquema em estrela relacional de Kimball (KIMBALL, ROSS e MERZ, 2002), ou seja,

composto de um tipo OR **tabela de fatos**, e suas dimensões definidas pelo tipo OR **tabela de dimensão**; seria um esquema em estrela espacial OR. Um esquema em estrela Objeto Relacional pode ser definido formalmente:

Definição 6: Esquema em Estrela Espacial Objeto Relacional

Um esquema em estrela espacial OR é construído a partir de um esquema UML/ODL de um CE da seguinte forma:

\forall classe UML/ODL de dimensão espacial (não-espacial) $D \in \text{UML/ODL CE}$

$f_{D\text{-mapeamento}}: D \rightarrow D\text{Type}$, $D\text{Type}$ é um tipo OR

\forall classe UML/ODL de medida espacial (não-espacial) $M \in \text{UML/ODL CE}$

$f_{M\text{-mapeamento}}: M \rightarrow M\text{Type}$, $M\text{Type}$ é um tipo OR

\forall função de dimensão espacial (não-espacial) $F \in \text{UML/ODL CE}$

$F_{F\text{-mapeamento}}: F \rightarrow \text{REF}$, REF é um tipo Referência (*Reference*) OR, que aponta para um $D\text{Type}$

$F_{\text{mapeamento-fatos}}: (\{M\text{Type}\} \cup \{\text{REF}\}) \rightarrow \text{FatoType}$, FatoType é um tipo OR

$\forall D\text{Type}$

$f_{\text{mapeamento-Dtable}}: D\text{Type} \rightarrow D\text{Table}$, $D\text{Table}$ é a respectiva tabela para o tipo $D\text{Type}$, ou seja, a tabela que representa uma dimensão.

$f_{\text{mapeamento-FatoTable}}: \text{FatoType} \rightarrow \text{FatoTable}$, FatoTable é a respectiva tabela para o tipo FatoType , ou seja, a tabela de fatos.

Na Figura 14, podemos ver o esquema em estrela espacial OR derivado do esquema UML/ODL CE AgroDistribuicao (Figura 13).

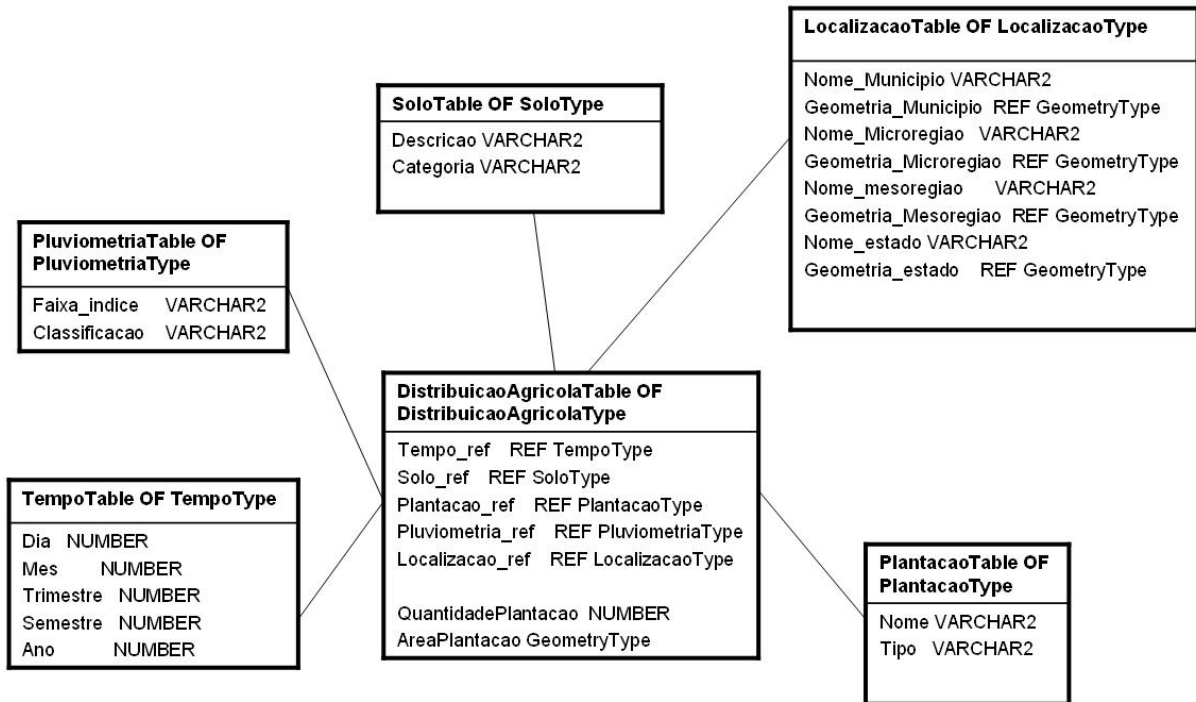


Figura 14 - Esquema em estrela espacial OR AgroDistribuicao

Para evitar redundâncias de objetos espaciais volumosos nas tabelas de dimensão espacial, assim como nas tabelas de fatos como *DistribuicaoAgricolaTable*, esses objetos geométricos são apenas referenciados (veja *LocalizacaoTable* e *DistribuicaoAgricolaTable*, na Figura 14): essa solução eficiente não pode ser implementada por esquemas em estrela relacionais.

Assim como nos esquemas em estrela relacionais, o esquema em estrela OR oblitera (é incapaz de mostrar) hierarquias espaciais e não-espaciais, além de funções espaciais e não espaciais. O fato é que a semântica do modelo OR ainda é pobre, quando comparada ao modelo UML/ODMG.

Um esquema em estrela espacial OR como esse da Figura 14 foi implementado no SGBD Oracle 10g. Áreas de plantação, assim como todas as informações espaciais do esquema Oracle OR, são representadas por objetos espaciais do Cartucho espacial da Oracle (Oracle, 2001). As informações espaciais foram associadas a um dos sistemas de coordenadas disponíveis no cartucho espacial do Oracle, representado por longitude/latitude relacionada a uma representação específica da terra. Desta forma, áreas de plantação podem ser localizadas dentro de municípios, geometrias de municípios podem ser localizadas dentro de microregiões e assim por diante.

O SGBD Oracle10g é o cerne do nosso projeto *MapWarehouse*, cujo protótipo será apresentado no próximo capítulo. A seguir apresentamos os scripts Oracle para definição do esquema em estrela OR da Figura 14.

```

-- Criando o tipo TempoType, para a dimensão tempo.
CREATE TYPE TempoType AS OBJECT(
  Dia NUMBER(2),
  Mes NUMBER(2),
  Trimestre NUMBER(1),
  Semestre NUMBER(1),
  Ano NUMBER(4)
)
/
-- Criando o tipo SoloType, para a dimensão Solo
CREATE TYPE SoloType AS OBJECT(
  Descricao VARCHAR2(60),
  Categoria VARCHAR2(20)
)
/
-- Criando o tipo PlantacaoType, para a dimensão Plantacao
CREATE TYPE PlantacaoType AS OBJECT(
  Nome VARCHAR2(30),
  Tipo VARCHAR2(20)
)
/
-- Criando o tipo PluviometriaType, para a dimensão Pluviometria
CREATE TYPE PluviometriaType AS OBJECT (
  Faixa_indice VARCHAR2(30),
  Classificacao VARCHAR2(30)
)
/
-- Criando o tipo Geometria, que é apenas uma referência para uma geometria do Oracle.
CREATE TYPE GeometryType as OBJECT(
  geom MDSYS.SDO_GEOMETRY);

-- Criando o tipo LocalizacaoType, para a dimensão espacial Localizacao
CREATE TYPE LocalizacaoType AS OBJECT(
  Nome_municipio VARCHAR2(100),
  Geometria_municipio GeometryType,
  Nome_microregiao VARCHAR2(100),
  Geometria_microregiao GeometryType,
  Nome_mesoregiao VARCHAR2(100),
  Geometria_mesoregiao GeometryType,
  Nome_estado VARCHAR2(100),
  Geometria_estado GeometryType
)
/
-- Criando o tipo DistribuicaoAgricolaType, para as medidas espaciais e não espaciais.
CREATE TYPE DistribuicaoAgricolaType AS OBJECT(

```



```

Tempo_ref REF TempoType,
Solo_ref REF SoloType,
Plantacao_ref REF PlantacaoType,
Pluviometria_ref REF PluviometriaType,
Localizacao_ref REF LocalizacaoType,

QuantidadePlantacao NUMBER, -- medida analítica
AreaPlantacao MDSYS.SDO_GEOMETRY -- medida espacial
)
/

-- Criação das tabelas para os tipos definidos acima
-- Criando a tabela TempoTable
CREATE TABLE TempoTable OF TempoType
/
-- Criando a table a SoloTable
CREATE TABLE SoloTable OF SoloType
/
-- Criando a tabela PlantacaoTable
CREATE TABLE PlantacaoTable OF PlantacaoType
/
-- Criando a tabela PluviometriaTable
CREATE TABLE PluviometriaTable OF PluviometriaType
/
-- Criando a tabela LocalizacaoTable
CREATE TABLE LocalizacaoTable OF LocalizacaoType
/
-- Criando a tabela DistribuicaoAgricolaTable, tabela de fatos
CREATE TABLE DistribuicaoAgricolaTable OF DistribuicaoAgricolaType

```

4.4. OR-OLAP Espacial

A consulta conceitual da seção 3.2, definida em OQL, pode ser mapeada para a linguagem de consulta SQL OR do Oracle10g, de acordo com o esquema em estrela espacial OR da Figura 14, da seguinte forma:

```

(Select a.Localizacao_ref.Geometria_Microregiao, SDO_AGGR_UNION(MDSYS.SDOAGGRTYPE(DEREF(a.AreaPlantacao), 0.005))
From DistribuicaoAgricolaTable a
Where a.Plantacao_ref.Nome = 'milho' And a.Tempo_ref.Mes = 5 And a.Tempo_ref.Ano = 2003 And
      a.Localizacao_ref.Nome_Estado = 'Paraíba'
Group by a.Localizacao_ref.Geometria_Microregiao)
UNION
(Select a.Localizacao_ref.Geometria_Mesoregiao, SDO_AGGR_UNION(MDSYS.SDOAGGRTYPE(DEREF(a.AreaPlantacao), 0.005))
From DistribuicaoAgricolaTable a

```

```
Where a.Plantacao_ref.Nome = 'milho' And a.Tempo_ref.Mes = 5 And a.Tempo_ref.Ano = 2003 And  
a.Localizacao_ref.Nome_Estado = 'Paraíba'  
Group by a.Localizacao_ref.Geometria_Mesoregiao)
```

Nós destacamos alguns pontos em relação a essa consulta:

1. A função conceitual `Geometric_union`, para agregação de geometrias, é mapeada para uma função do cartucho espacial Oracle: `SDO_AGGR_UNION(MDSYS.SDOAGGRTYPE(Geometry IN MDSYS.SDO_Geometry, Tolerance IN NUMBER)) RETURN MDSYS.SDO_Geometry`, que retorna um objeto geométrico representando a união topológica dos objetos geométricos agregados. O parâmetro `Tolerance` indica se o resultado vai ser uma geometria simples ou composta. Nessa consulta, o valor 0.005 indica a tolerância em metros;
2. No Oracle, critérios de agregação da cláusula `Group by` não podem ser do tipo `GeometryType` ou `LOBType`. Dessa forma, note que o critério de agregação de `Localizacao_ref.Geometria_Microregiao` e `Localizacao_ref.Geometria_Mesoregiao` na consulta são do tipo `REF GeometryType` (\neq `GeometryType`);
3. Atributos `REF` são essenciais para o desempenho de consultas na presença de objetos espaciais ou objetos de imagem.

4.5 Considerações Finais

Nesse capítulo foi apresentada a proposta de um modelo de dados multidimensional espacial para DW espacial. Os conceitos de dimensão espacial, medida espacial e hierarquia espacial foram definidos formalmente. Também foi mostrado que as dependências entre esses conceitos formam um cubo de dados espacial. Para validação do modelo, um esquema conceitual foi definido para o modelo utilizando um estudo de caso. Foram definidas algumas consultas SOLAP ao esquema conceitual, utilizando o formalismo OQL. Por fim, foi realizado o mapeamento de um modelo multidimensional espacial para um esquema em estrela espacial Objeto Relacional, o Oracle 10g, no qual também foram realizadas consultas SOLAP.

Capítulo 5

Otimização de Consultas Multidimensionais Espaciais

Neste capítulo, é utilizado o exemplo do estudo de caso do capítulo 4 para caracterizar o problema de otimização de consultas multidimensionais espaciais. Em seguida, é apresentado o projeto e a implementação de um algoritmo genérico de otimização capaz de melhorar o desempenho de consultas multidimensionais espaciais, principalmente as que utilizam a operação de união de geometrias. São apresentadas duas formas para agregar medidas espaciais: 1) através de uma operação de agregação espacial, a união de geometrias; 2) sem a necessidade de unir geometrias, mas simplesmente inserindo-as em um mapa no qual estarão “unidas” visualmente.

5.1 O Problema de Otimização Espacial

Para consultas SOLAP a um DW espacial (esquema espacial OR em estrela) o requisito tempo de resposta é fundamental. O entrave crítico ao desempenho de consultas SOLAP se dá principalmente em agregações de objetos espaciais, quando da navegação através de hierarquias (espaciais ou não); operações como união de medidas espaciais, por exemplo, têm normalmente um alto custo de processamento, se um esforço de otimização não for feito. Para tal fim, um modelo de otimização de operações de agregação de objetos espaciais é necessário.

Para melhor caracterizar o problema de otimização espacial para consultas SOLAP, consideremos, por exemplo, a geografia do Estado brasileiro da Paraíba, cuja divisão política pode ser vista na Figura 15.

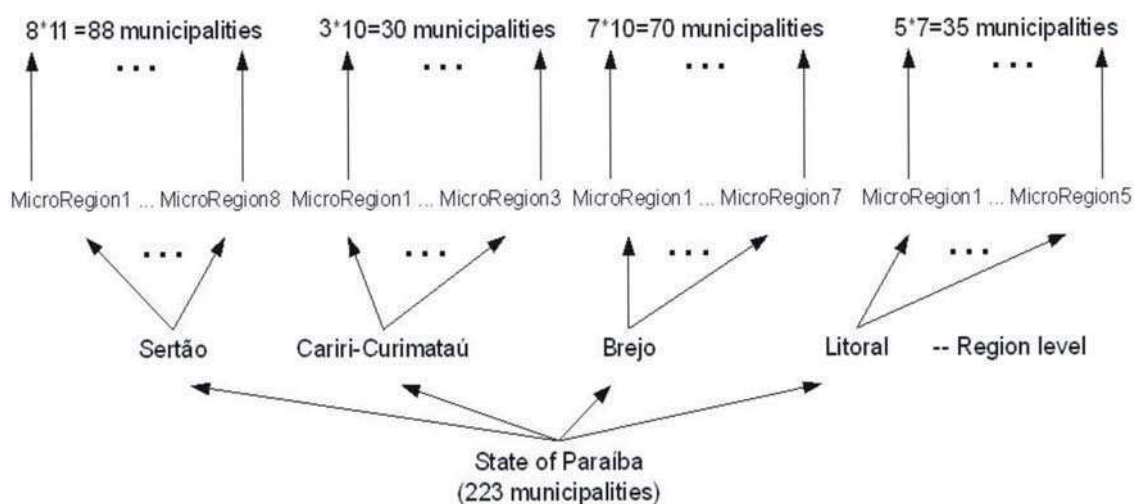


Figura 15 - Divisão política do Estado da Paraíba

Considerando o esquema em estrela espacial OR (AgroDistribuicao) da Figura 14, suponha que apenas o nível espacial Município é pré-armazenado, ou seja, são pré-agregadas todas as áreas de plantação (medida espacial) de cada Município do Estado da Paraíba. Agora suponha a seguinte consulta exemplo: *Recuperar as áreas de plantação de Milho para cada microregião (região) e para cada região (microregião) do Estado da Paraíba, durante Maio de 2003*. Essa consulta pode ser representada na linguagem OR do Oracle:

```
(Select d.Localizacao_ref.Geometria_Microregiao, SDO_AGGR_UNION(MDSYS.SDOAGGRTYPE(DEREF(d.AreaPlantacao), 0.005))
From DistribuicaoAgricolaTable d
Where d.Plantacao_ref.Nome = 'milho' And d.Tempo_ref.Mes = 5 And d.Tempo_ref.Ano = 2003 And
      d.Localizacao_ref.Nome_Estado = 'Paraiba'
Group by d.Localizacao_ref.Geometria_Microregiao)
UNION
(Select a.Localizacao_ref.Geometria_Regiao, SDO_AGGR_UNION(MDSYS.SDOAGGRTYPE(DEREF(d.AreaPlantacao), 0.005))
From DistribuicaoAgricolaTable d
Where d.Plantacao_ref.Nome = 'milho' And d.Tempo_ref.Mes = 5 And d.Tempo_ref.Ano = 2003 And
      d.Localizacao_ref.Nome_Estado = 'Paraiba'
Group by d.Localizacao_ref.Geometria_Regiao)
```

Dada essa consulta, considere que existem aproximadamente 50 plantações de milho por município e que as áreas de plantação sejam armazenadas na granularidade dia da dimensão Tempo: o que resultaria em aproximadamente $223 * 50 * 30 = 334.500$ áreas de plantação por mês (não consideramos aqui as demais dimensões para tornar o exemplo mais simples). O fator crítico para a performance da consulta são as uniões de todas as áreas de plantação dos municípios, realizadas dinamicamente para gerar objetos espaciais (sumarizações) representando as áreas de plantação nos níveis Microregião e Região, em uma operação de Roll-up espacial. O custo dessas operações é muito alto, inviabilizando a realização desse tipo de consulta.

O problema de otimização espacial pode ser representado genericamente como segue: *dada uma consulta espacial, quantas operações de agregação espacial devem ser previamente computadas, para que essa consulta seja executada em um tempo aceitável?* Isto nos levou a criar um modelo lógico de otimização, descrito na próxima subseção.

5.2 Otimização Lógica Espacial

O modelo de otimização lógica espacial do *MapWarehouse* é baseado no mecanismo de pré-armazenamento de agregados espaciais, uma extensão da noção de “prestored summary aggregates” de Kimball (KIMBALL, ROSS e MERZ, 2002). Definimos agregado espacial como uma medida espacial que representa a sumarização ou agregação de um conjunto de medidas espaciais que satisfazem a algum critério comum. Agregados espaciais são representados através de medidas espaciais em tabelas de fatos com níveis de agregação derivados a partir da tabela de fatos básica (que armazena fatos espaciais associados aos níveis de menor granularidade nas dimensões, por exemplo, nível *dia* na dimensão Tempo, nível *Município* na dimensão Localização). O agregado espacial também pode ser visto como um esquema em estrela espacial OR em que as agregações espaciais são obtidas a partir do esquema em estrela espacial OR básico.

Considere o esquema base *AgroDistribuicao* da Figura 14: nele, a menor granularidade das áreas de plantação é Município. Um agregado espacial (Figura 16), com um nível de granularidade Microregião, pode ser derivado do esquema básico, da seguinte forma:

```
CREATE TYPE MicroregiaoType AS OBJECT(  
  Nome_Microregiao VARCHAR2(100),  
  Geometria_Microregiao REF Geometry_Objtyp,  
  Nome_Mesoregiao VARCHAR2(100),  
  Geometria_Mesoregiao REF Geometry_Objtyp,  
  Nome_Estado VARCHAR2(100),  
  Geometria_Estado REF Geometry_Objtyp,  
);  
  
CREATE TABLE MicroregiaoTable OF MicroregiaoType;  
  
CREATE TYPE MicroregiaoDistribuicaoAgricolaType AS OBJECT(  
  Tempo_ref REF Time_Objtyp,  
  Solo_ref REF Soil_Objtyp,  
  Plantacao_ref REF Plantation_Objtyp,  
  Pluviometria_ref REF Precipitation_Objtyp,  
  Localizacao_ref MicroregiaoType,  
  QuantidadePlantacao NUMBER,  
  AreaPlantacao Geometry_Objtyp,  
);  
  
CREATE TABLE MicroregiaoDistribuicaoAgricolaTable OF MicroregiaoDistribuicaoAgricolaType;
```

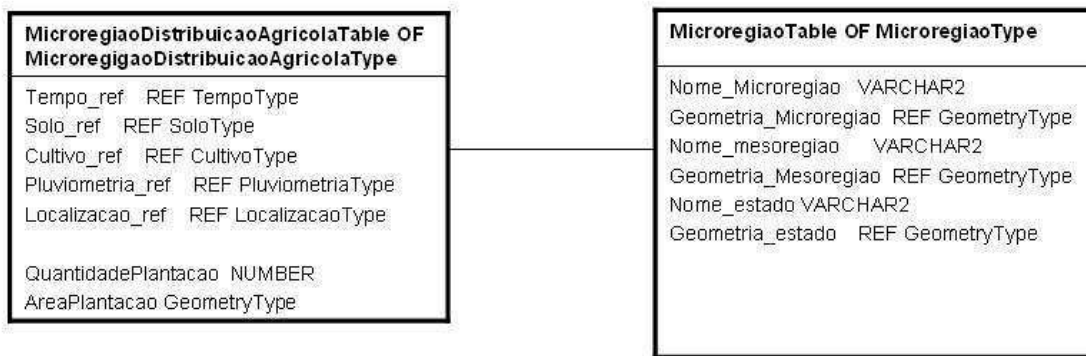


Figura 16 - Exemplo de agregado espacial

A seguir, mostramos dois exemplos de agregados espaciais:

- *Áreas de plantação por Microregião:* tabela de fatos MicroregiaoDistribuicaoAgricolaTable na Figura 16. A tabela de dimensão espacial MicroregiaoTable na Figura 16 é derivada da tabela de dimensão espacial LocalizacaoTable da Figura 14;
- *Áreas de plantação por Região e por Mês:* As tabelas de dimensões espaciais RegiaoTable e MesTable podem ser respectivamente derivadas das tabelas LocalizacaoTable e TempoTable da Figura 14.

O objetivo principal da otimização lógica espacial do *MapWarehouse* é fazer o melhor uso dos agregados espaciais pré-armazenados com intuito de melhorar a performance de consultas que realizam agregação espacial; na verdade, as consultas são automaticamente reescritas para acessar os agregados espaciais pré-armazenados de forma apropriada. Uma vez reescritas, as consultas submetidas para o otimizador de consultas específico do SGBD Oracle (otimização física) que explora índices espaciais como R-trees (GUTTMAN, 1984), entre outras técnicas de otimização física.

Considere que as tabelas MicroregiaoDistribuicaoAgricolaTable OF MicroregiaoDistribuicaoAgricolaType e MicroregiaoTable OF MicroregiaoType sejam respectivamente a tabela de agregados espaciais pré-armazenados e a tabela de dimensão espacial (Figura 16) — as outras dimensões são aquelas descritas no esquema da Figura 14 —; assim, a consulta exemplo seria reescrita pelo otimizador de consultas espaciais do *MapWarehouse* da seguinte forma:

```
(Select d.Localizacao_ref.Geometria_Microregiao, SDO_AGGR_UNION(MDSYS.SDOAGGRTYPE(DEFER(d.AreaPlantacao), 0.005))
From MicroregiaoDistribuicaoAgricolaTable d
Where d.Plantacao_ref.Nome = 'milho' And d.Tempo_ref.Mes = 5 And d.Tempo_ref.Ano = 2003 And
      d.Localizacao_ref.Nome_Estado = 'Paraíba'
Group by d.Localizacao_ref.Geometria_Microregiao)
UNION
```

```
(Select a.Localizacao_ref.Geometria_Regiao, SDO_AGGR_UNION(MDSYS.SDOAGGRTYPE(DEREF(d.AreaPlantacao), 0.005))
From MicroregiaoDistribuicaoAgricolaTable d
Where d.Plantacao_ref.Nome = 'milho' And d.Tempo_ref.Mes = 5 And d.Tempo_ref.Ano = 2003 And
      d.Localizacao_ref.Nome_Estado = 'Paraíba'
Group by d.Localizacao_ref.Geometria_Regiao)
```

Os dois itens em negrito na consulta acima são os únicos itens diferentes da consulta original! Ou seja, o otimizador lógico espacial simplesmente substituiu `DistribuicaoAgricolaTable` por `MicroregiaoDistribuicaoAgricolaTable`. Os ganhos de performance obtidos com essa simples transformação são descritos no próximo capítulo.

Observe que as novas tabelas de fatos e dimensão da Figura 16 mantêm os mesmos nomes de atributos das tabelas da Figura 14, das quais elas foram derivadas. Isto é essencial para garantir a simplicidade do algoritmo.

Em termos gerais, o algoritmo de otimização espacial do *MapWarehouse* faz uma busca sobre os agregados espaciais pré-armazenados, em tempo de consulta, para encontrar qual deles proporcionará mais ganho de performance para a consulta. Um fato importante é que os agregados espaciais (tabelas de fatos e dimensões) são conhecidos apenas pelo algoritmo: usuários finais e desenvolvedores de aplicações nunca vêm essas tabelas de agregados.

5.3 O Algoritmo

Nessa subseção, discutimos o algoritmo de reescrita de consultas do *MapWarehouse*, chamado *navegador de agregados espaciais* (NAE). NAE intercepta as consultas SOLAP requisitadas pelo cliente ao DWE, e quando possível, transforma as consultas feitas sobre esquema básico em consultas que acessam agregados espaciais pré-armazenados — otimização lógica.

Para realizar consultas espaciais de forma eficiente, são necessárias soluções eficientes para três problemas: 1) Determinar quais agregados espaciais serão materializados, incluindo a forma como os mesmos serão armazenados e indexados; 2) Manutenção dos agregados espaciais, no momento em que a tabela de fatos é atualizada; 3) Fazer o uso eficiente dos agregados espaciais existentes para garantir o bom desempenho de consultas SOLAP.

O algoritmo NAE apenas trata do problema 3, uma vez que os problemas 1 e 2 podem vir a ser tratados em trabalhos futuros. O tratamento do problema 3 é bem simples. Conceitualmente, consiste nos seguintes passos:

1. Intercepta a consulta requisitada pelo usuário;

2. Seleciona os agregados espaciais (visões materializadas) existentes e as ordena da menor para a maior, baseado na cardinalidade de tabela de fatos espacial. Escolhe o primeiro agregado espacial da lista;
3. Se todos os atributos da consulta SQL OR podem ser direta ou indiretamente encontrados no agregado espacial (ocorre um casamento entre a consulta e o agregado espacial), alterar a consulta original substituindo a tabela de fatos básica pela tabela de fatos do agregado espacial em questão; se não, escolhe o próximo agregado espacial da lista e volta para o passo 3;
4. Executar a consulta alterada.

No passo 3 do algoritmo pode ocorrer um casamento entre a consulta e um agregado espacial. Esse casamento pode ser de dois tipos:

- *Parcial*: quando existe uma pré-agregação em um nível (exceto o inicial) inferior ao nível da consulta. Por exemplo, uma consulta no nível Estado que usa um agregado espacial no nível Microregião;
- *Total*: quando existe uma pré-agregação no mesmo nível do que foi definido na consulta. Por exemplo, uma consulta no nível Microregião que usa um agregado espacial no nível Microregião.

Do ponto de vista de performance, fica claro que a consulta realizada sobre o agregado espacial (casamento parcial ou total) terá um tempo de resposta inferior ao que seria gasto se a consulta fosse executada sobre o esquema básico. Além disso, o casamento total é mais eficiente que o casamento parcial, já que nesse caso não seria necessária a realização de operações de agregação espacial. Note que o casamento parcial é mais provável de acontecer que o casamento total.

O algoritmo sempre termina, pois eventualmente, caso não haja casamento com nenhum dos agregados espaciais, a consulta termina sendo executada sobre o esquema básico, capaz de satisfazer qualquer consulta.

5.3.1 Regras para Criação de Agregados Espaciais

Note que a otimização apresentada na seção 5.1.2 foi feita a partir da reescrita da consulta, resumindo-se simplesmente a substituição do nome da tabela de fatos da consulta original. Entretanto, podemos afirmar que, apesar da simplicidade, essa técnica é genérica e pode ser utilizada para otimização de qualquer consulta SOLAP de agregação espacial feita no DWE. Para tal, alguns cuidados devem ser tomados no momento da geração dos agregados espaciais (também conhecidos como Visões Materializadas ou Estrelas Secundárias

(GOLDSTEIN e LARSON, 2001)). No momento de gerar uma estrela secundária, devem ser mantidos os mesmos nomes das colunas das tabelas originais, tanto para a nova tabela de fatos (e.g. `MicroregiaoDistribuicaoAgricolaTable`, na seção 5.1.2) quanto para as novas tabelas de dimensões (e.g. `MicroregiaoTable`, na seção 5.1.2). No caso das tabelas de dimensões, os atributos de menor granularidade são removidos (e.g. `Geometria_Municipio`, na seção 5.1.2); nas tabelas de fatos, são alterados os tipos das colunas que apontam para as novas tabelas de dimensão (e.g. `LocalizacaoType` por `MicroregiaoType`, na seção 5.1.2).

5.4 Otimização de Consultas Utilizando Concatenação de Medidas Espaciais

Todo o problema de otimização de consultas espaciais discutido na seção anterior se faz presente em consultas que contenham operações de agregação espacial, por exemplo, união de geometrias. Nesta seção, nós apresentamos uma nova técnica de otimização de consultas capaz de unir dados espaciais apenas visualmente, ou seja, sem a necessidade de unir geometrias, mas simplesmente selecionando-as e inserindo-as em um mapa, no qual ficam unidas visualmente. Essa abordagem pode ser chamada de *Otimização de Consultas por Concatenação de Geometrias*. Desta forma, apenas o problema de otimização analítica teria de ser resolvido, e esse já é tratado por grande parte dos SGBD que lidam com Data Warehouse.

A idéia dessa abordagem é bastante simples: quando o usuário fizer uma consulta de agregação espacial para um determinado nível espacial, basta selecionar, da tabela de fatos do esquema em estrela básico, as medidas espaciais que satisfazem às restrições da cláusula WHERE da consulta, formando uma coleção de geometrias (ou concatenação); em seguida, essas geometrias devem ser inseridas no mapa, ou seja, as geometrias do cartucho espacial do SGBD Oracle são convertidas para elementos da linguagem SVG; por fim, deve-se apresentar o mapa no respectivo nível espacial que foi definido na consulta, por exemplo, se o usuário estiver fazendo um agrupamento no nível espacial Microregião, para as áreas de plantação do Estado da Paraíba, devem ser apresentadas as geometrias de todas as microregiões da Paraíba, além de todas as áreas de plantação dessas microregiões. Além disso, devem ser feitas as agregações analíticas necessárias, caso existam medidas analíticas.

Para garantir o funcionamento desse processo, é necessário que exista pelo menos uma dimensão espacial, com mapas no nível espacial que o usuário deseja realizar a consulta. Por exemplo, uma consulta no nível espacial Município só pode ser executada se houver uma dimensão espacial com o nível espacial Município contendo o respectivo mapa para esse nível.

A grande vantagem da concatenação de geometrias é o bom desempenho de consultas espaciais, garantido pela não realização de operações de união de geometrias. Entretanto, a desvantagem é que para cada elemento de um nível espacial, por exemplo, para cada Microregião de um Estado, *não* será gerado um único objeto espacial representando a união de todos os objetos espaciais dessa Microregião, mas um conjunto (ou concatenação) de objetos espaciais dessa Microregião. Assim, não dá para associar apenas um objeto espacial para cada elemento do nível espacial e também não é possível associar uma agregação analítica (e.g. soma das quantidades de plantação de uma Microregião) com apenas um objeto espacial, como visto a seguir.

Na Figura 17 é mostrada uma comparação entre concatenação de geometrias e união de geometrias, através de um exemplo de visualização de áreas de plantação no nível espacial Município.

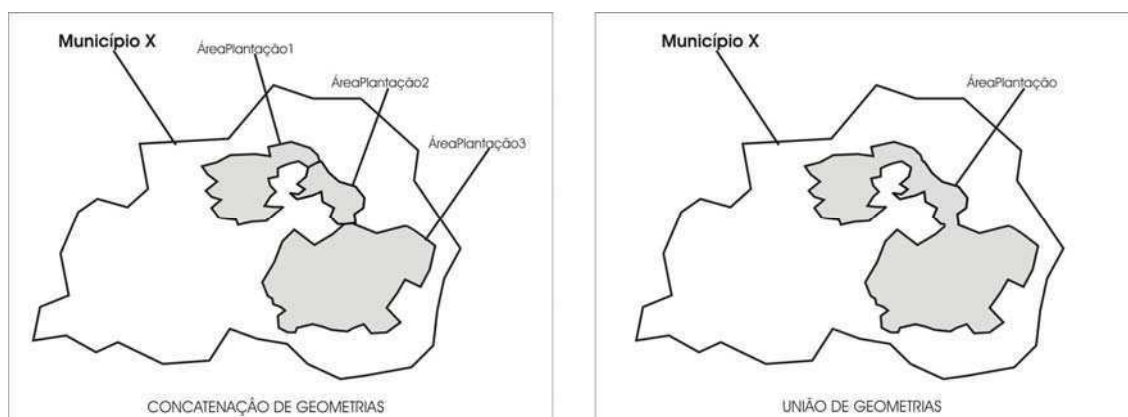


Figura 17 - Concatenação de Geometrias X União de Geometrias

Observe que na concatenação, temos as três áreas de plantação (ÁreaPlantação1, ÁreaPlantação2 e ÁreaPlantação3) do MunicípioX concatenadas no mapa, ou seja, gerando uma união visual. Já na união de geometrias temos apenas um objeto espacial (ÁreaPlantação) que representa a união das três áreas de plantação do MunicípioX. Caso houvesse uma medida numérica, o resultado analítico para as duas abordagens seria o mesmo: a soma das quantidades de plantação das três áreas de plantação. Mas, no caso da concatenação, essa soma (agregação analítica) estaria associada a três objetos e não a apenas um objeto espacial, como é o caso da abordagem que utiliza união de geometrias.

Para definir quais técnicas de otimização têm melhor desempenho e em que situações são preferíveis foram feitos testes comparando-as; uma avaliação dos resultados da comparação é tratada na seção 6.2 do próximo capítulo.

5.5 Considerações Finais

Este capítulo inicialmente caracterizou o problema de otimização de consultas multidimensionais espaciais. A seguir foi apresentado o projeto e a implementação de um algoritmo genérico de otimização capaz de melhorar o desempenho de consultas multidimensionais espaciais. Foram consideradas duas formas para agregar medidas espaciais: utilizando operações de agregação espacial e sem a utilização de operações de agregação espacial.

Capítulo 6

Validação do Modelo

Este capítulo apresenta a validação do modelo proposto no capítulo 4 e do algoritmo de otimização do capítulo 5. Inicialmente, é apresentada a implementação de um protótipo de software, visando a validar o modelo definido no capítulo 4. Para testar o protótipo, é usado o estudo de caso sobre plantações agrícolas no Brasil, que claramente integra informações analíticas e espaciais. Em seguida, é apresentada a avaliação experimental para a otimização espacial proposta no capítulo 5.

6.1 Implementação do Modelo Conceitual Multidimensional Espacial

Foi desenvolvido um protótipo chamado *MapWarehouse*, que implementa um esquema em estrela espacial OR e operações SOLAP utilizando o SGBD Oracle 10g. A arquitetura do mesmo pode ser dividida em três camadas, como visto na Figura 18.

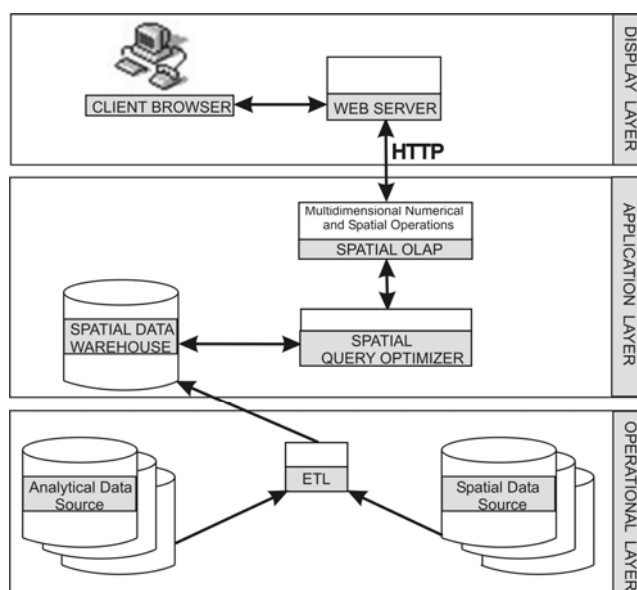


Figura 18 - Arquitetura do MapWarehouse

- *Camada Operacional (Operational Layer):* é composta por fontes de dados espaciais e fontes de dados analíticos, utilizadas para popular o Data Warehouse espacial — um banco de dados de acordo com o esquema em estrela espacial OR — através do processo ETL;
- *Camada de Aplicação (Application Layer):* é responsável pelo processamento de consultas SOLAP requisitas pelo usuário; a lógica de negócios do sistema foi implementada usando a linguagem de programação Java;

- *Camada de Apresentação (Display Layer)*: implementa a interface gráfica com o usuário, para permitir a definição de consultas SOLAP e visualização dos resultados em tabelas e mapas.

Para o desenvolvimento da interface gráfica foi utilizado um *framework* que permite o desenvolvimento rápido de aplicações de geoprocessamento na Web, chamado iGIS (BAPTISTA et al., 2004). As tecnologias utilizadas foram HTML, JSP/Servlet, SVG e JavaScript. Podemos distinguir dois pontos centrais em relação à interface: 1) Criação de consultas para o DWE; 2) Visualização dos resultados.

6.1.1 Criação de Consultas para o DWE

Para facilitar a vida do usuário, o *MapWarehouse* permite que o mesmo possa criar consultas analítico-espaciais sem necessitar conhecer a sintaxe da linguagem de consulta ao DWE; isso é feito através de um conjunto de menus e janelas interativos, como visto na Figura 19.

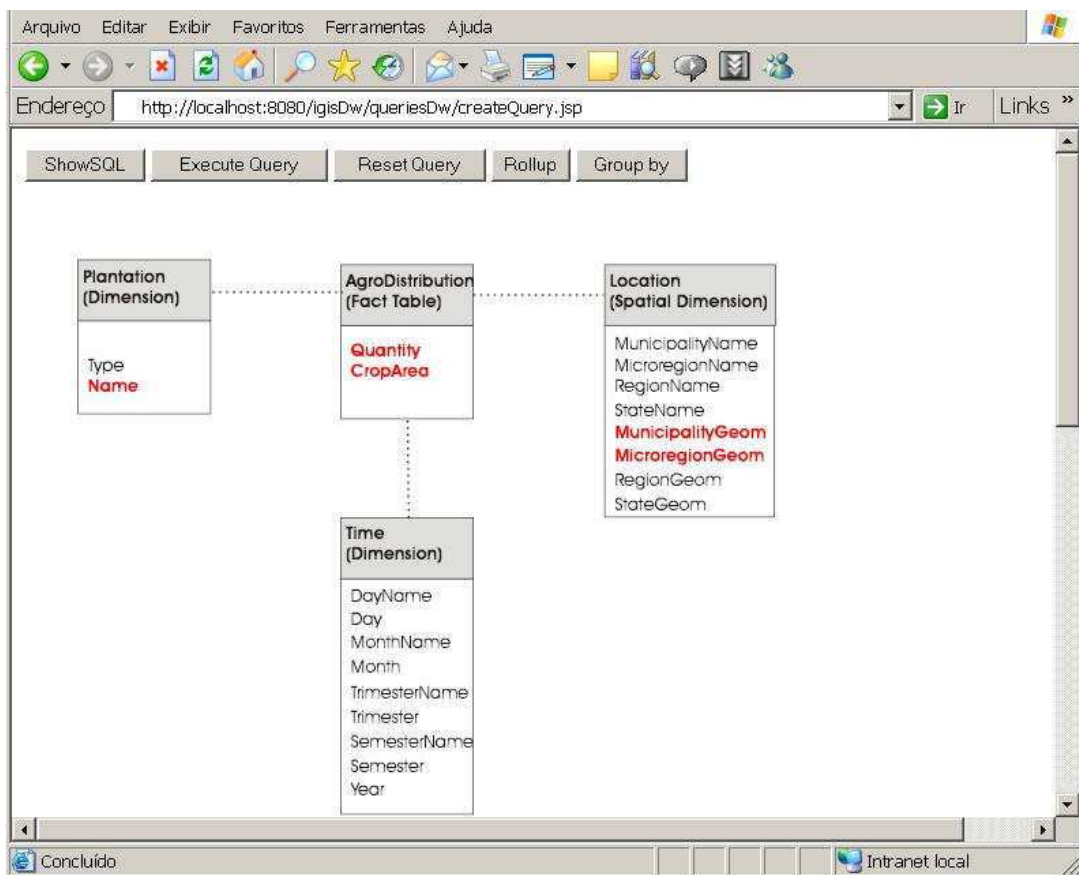


Figura 19 - Interface do *MapWarehouse*, criando consultas

Na tela inicial, é apresentada uma visão gráfica do esquema em estrela espacial OR, contendo a tabela de fatos e as dimensões. Nela, o usuário pode selecionar os atributos

(espaciais ou não) das dimensões ou as medidas da tabela de fatos que deseja visualizar, apenas através de um clique de mouse sobre o atributo; os atributos selecionados ficam em vermelho e os não-selecionados permanecem pretos. Para definir restrições (cláusula WHERE) da consulta basta clicar sobre a parte cinza de cada dimensão, quando é aberta uma nova janela para criação de restrições para a respectiva dimensão. Caso o usuário clique sobre a tabela de fatos, não serão definidas restrições, mas a função de agregação analítica (se uma medida analítica foi selecionada) e/ou a função de agregação espacial (se uma medida espacial foi selecionada).

Através da interface do *MapWarehouse* é possível realizar consultas ad-hoc, que possibilitam ao usuário definir no mapa uma janela espacial retangular, para que a consulta recupere apenas os objetos espaciais que estiverem dentro dessa janela. Para tal, basta que o usuário acesse as restrições da dimensão espacial e clique no botão específico para definir janela espacial. Lá, é apresentado o mapa referente à dimensão espacial e o usuário escolhe em que nível espacial deseja definir a janela, Região, por exemplo. Logo em seguida, o usuário clica em dois pontos desse mapa formando um retângulo que é desenhado no mapa; por fim, confirma a operação que define uma restrição espacial ad-hoc para a consulta que está sendo criada.

Depois de selecionar os atributos e definir as restrições da consulta o usuário pode definir as operações de Roll-up espacial e Drill-Down espacial. Para isso, basta clicar no botão “roll-up”, quando será aberta uma nova janela na qual o usuário informa sobre qual dimensão deseja fazer a operação SOLAP e quais os níveis inicial e final da operação.

Além disso, a qualquer momento o usuário pode clicar no botão “showSQL”, que monta e apresenta textualmente a consulta criada pelo usuário na sintaxe do Oracle. E por fim, para visualizar os resultados da consulta, basta clicar no botão “Execute Query”.

6.1.2 Visualização dos Resultados

Dependendo da consulta realizada podemos ter três tipos de resultados: apenas tabular (ou analítico), apenas espacial e analítico-espacial. O *MapWarehouse* executa a consulta e apresenta o resultado em uma dessas três formas. No primeiro caso, é apresentada apenas uma tabela com os resultados da consulta. Nos resultados espaciais o *MapWarehouse* gera os mapas em uma outra interface de saída. Esses mapas dinâmicos são sincronizados com informações analíticas (caso existam) e também são interativos, ou seja, quando o usuário passa o mouse sobre um objeto espacial são apresentadas informações referentes àquele objeto.

Para melhor entender a criação e visualização de resultados de uma consulta, vamos demonstrar os passos para realizar a seguinte consulta: *Quais as áreas de plantação (medida espacial), e suas respectivas quantidades (medida numérica), do Estado da Paraíba por Microregião e por Região (Rollu-up do nível Microregião para o nível Região) do mês de Janeiro de 2003 até o mês de Maio de 2003, mês a mês, que estejam dentro de uma janela espacial.*

Um ponto importante é que essa consulta apresenta uma outra funcionalidade do *MapWarehouse*, a realização de consultas com séries-temporais, ou seja, visualização de informações em diferentes pontos de um certo nível temporal e em diferentes mapas, por exemplo, mapas mês a mês, mapas ano a ano; e tudo isso podendo ser feito através de uma única consulta SOLAP.

Inicialmente são selecionados os atributos para a consulta (Figura 20). Foram eles: na tabela de fatos, a medida espacial área de plantação (*CropArea*) e a medida não espacial quantidade (*Quantity*); na dimensão espacial Localização, a descrição dos nomes das microregiões e das regiões (*MicroregionName* e *RegionName*) e as geometrias das microregiões e das regiões (*MicroregionGeom* e *RegionGeom*); na dimensão temporal, os atributos mês e ano (*Month* e *Year*); na dimensão Plantacao (*Plantation*), o nome da plantação (*Name*).

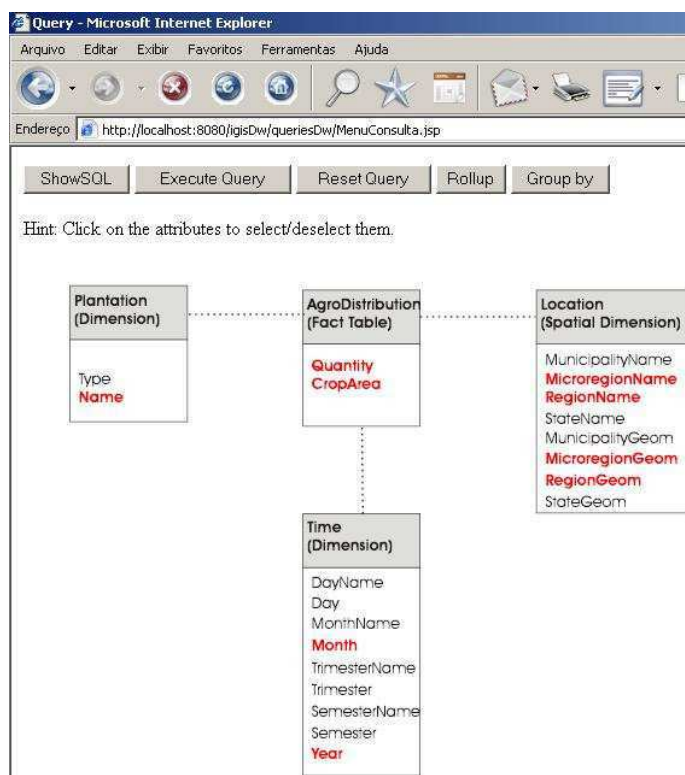


Figura 20 - Definindo atributos de uma consulta

A seguir, para definir as restrições de tempo, basta clicar na parte cinza da dimensão Tempo. Então, uma nova janela é aberta (Figura 21). Nela, o usuário define que o ano deve ser o de 2003 e que os meses estão no intervalo de Janeiro (1) até Maio (5) daquele ano. Depois basta clicar no botão “Ok” para retornar a tela principal.

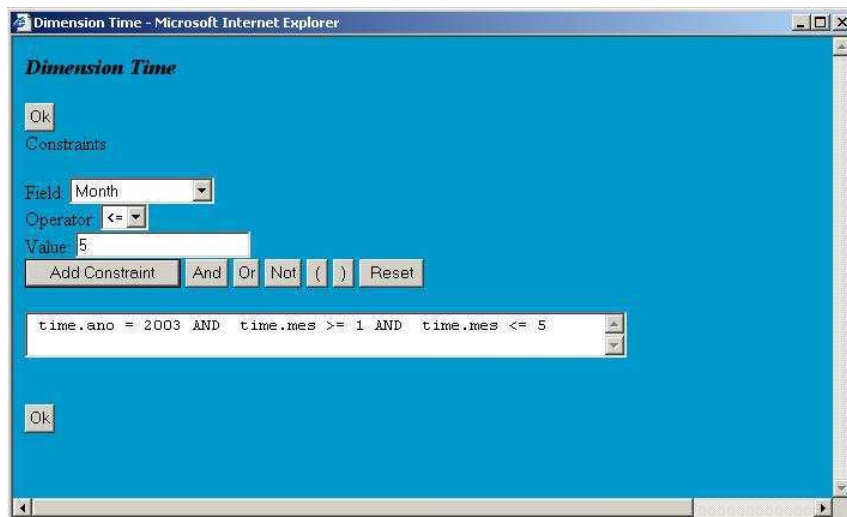


Figura 21 - Definindo restrições para a dimensão Tempo

O usuário segue o mesmo procedimento anterior para definir as restrições da dimensão espacial Localização (Figura 22). Nesse caso, o usuário determina que as medidas estejam dentro do Estado da Paraíba. E para definir uma janela espacial dentro do Estado da Paraíba, basta clicar no botão “Create Spatial window on map”, quando será aberta uma nova janela na qual o usuário pode definir uma janela espacial de forma bastante simples, como pode ser visto na Figura 23. Observe também que nessa janela o usuário pode escolher o nível espacial no qual deseja ver o mapa, no caso, foi definido Região.

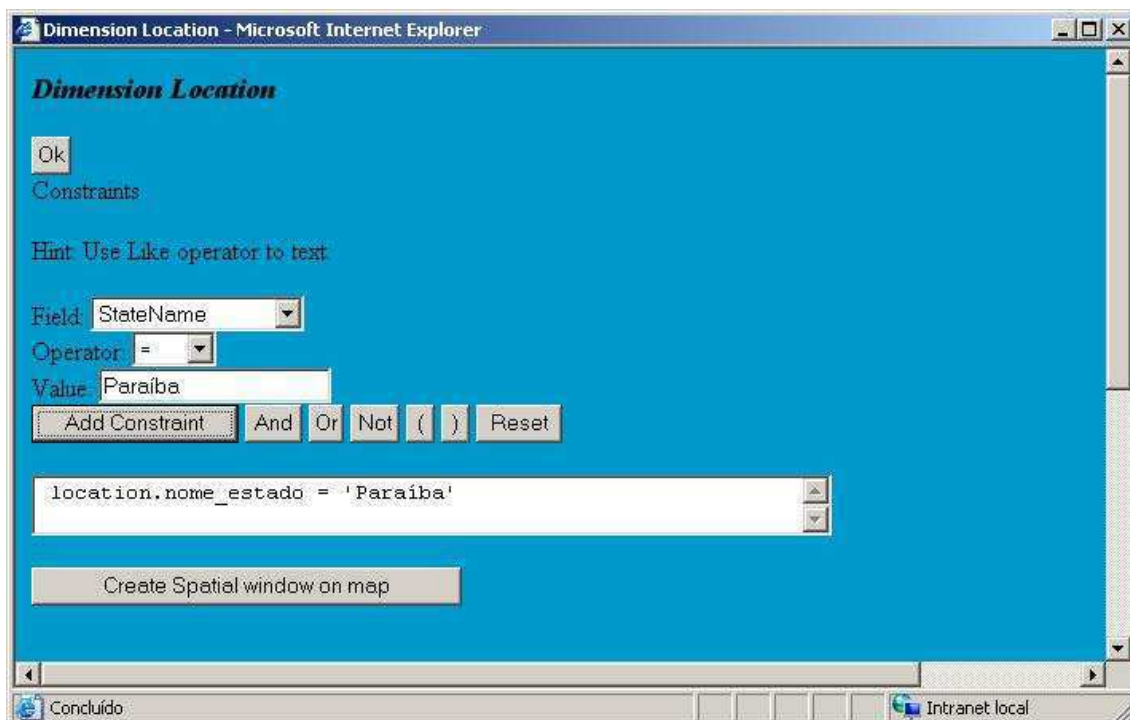


Figura 22 - Definindo restrições da dimensão espacial Localização

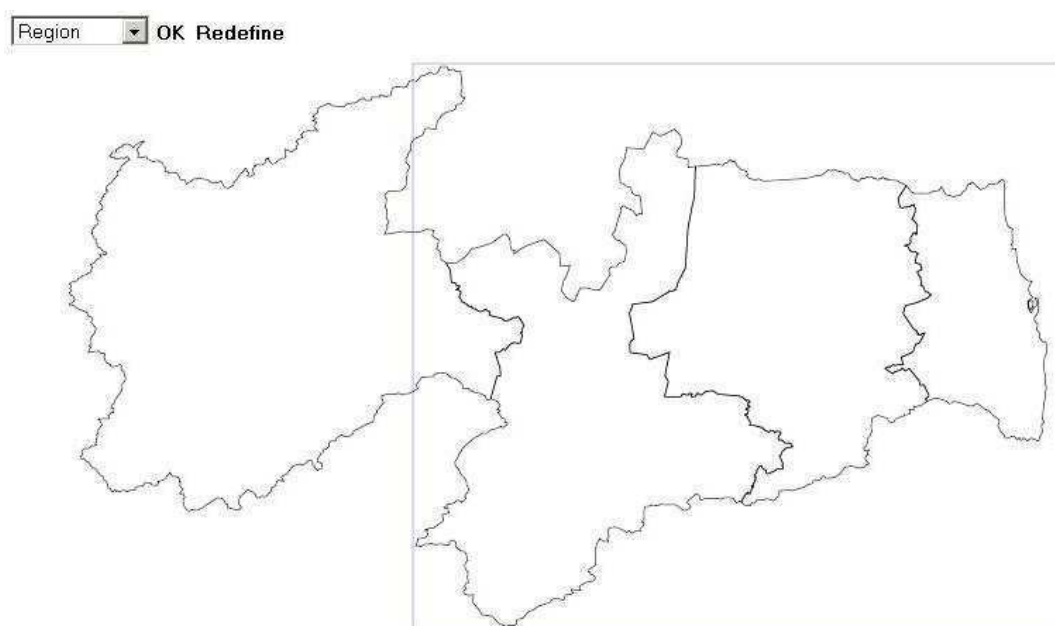


Figura 23 - Definindo janela espacial no mapa

Depois de definidas todas as restrições da consulta, o usuário clica no botão “Roll-up”, na tela inicial, para definir a operação SOLAP Roll-up espacial (Figura 24). Nessa janela, o usuário seleciona sobre qual dimensão do cubo de dados espacial deseja realizar a operação de Roll-up e depois define os níveis inicial e final da operação. Nesse caso, a dimensão

selecionada foi a dimensão espacial Localização e a operação vai do nível espacial Microregião para o nível espacial Região.

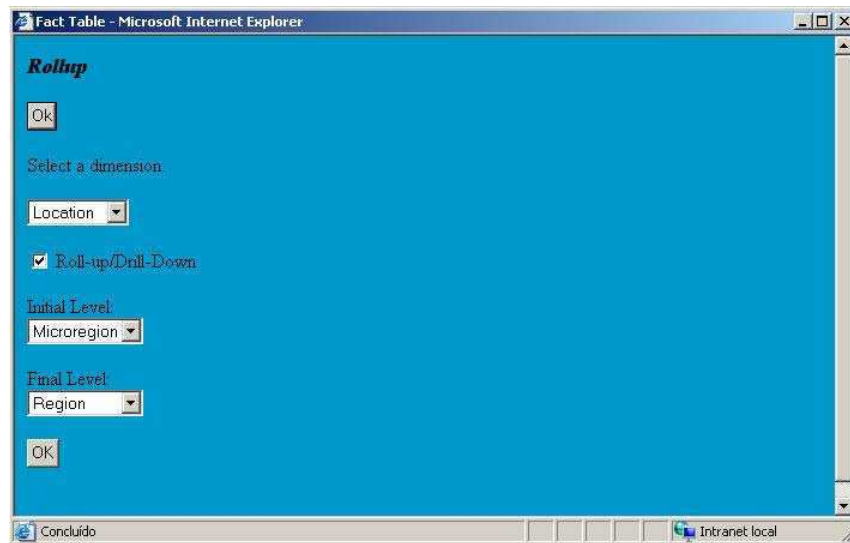


Figura 24 - Definindo operação Roll-up Espacial

Depois de definida a operação SOLAP, o usuário clica no botão “Group By” na tela principal. Será aberta uma nova janela (Figura 25) na qual o usuário pode definir o agrupamento na dimensão Tempo, necessário para definir a série temporal por mês. Observe que o usuário seleciona a dimensão Tempo e o nível Mês e como já havia definido que os meses devem estar entre Janeiro e Maio de 2003, conseguindo assim definir a série temporal: áreas de plantação de Janeiro a Maio de 2003, mês a mês.

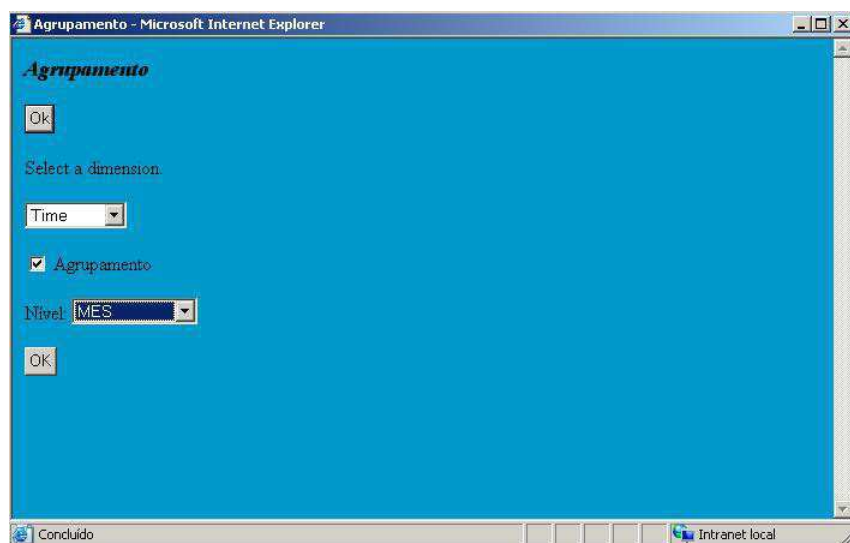


Figura 25 - Definindo agrupamento temporal

Pelo fato da consulta tratar de uma operação Roll-up e do usuário desejar obter tanto medidas espaciais quanto analíticas, devem ser definidas as funções de agregações espacial (para área de plantação) e analítica (para a quantidade de plantação). Para tal, basta o usuário clicar na parte cinza da tabela de fatos na tela inicial, surgindo uma janela (Figura 26) que permite ao usuário definir essas funções. Nesse caso, foi escolhida a função de agregação analítica soma (*SUM*) e a função de agregação espacial de união de geometrias (*Geometric Union*).

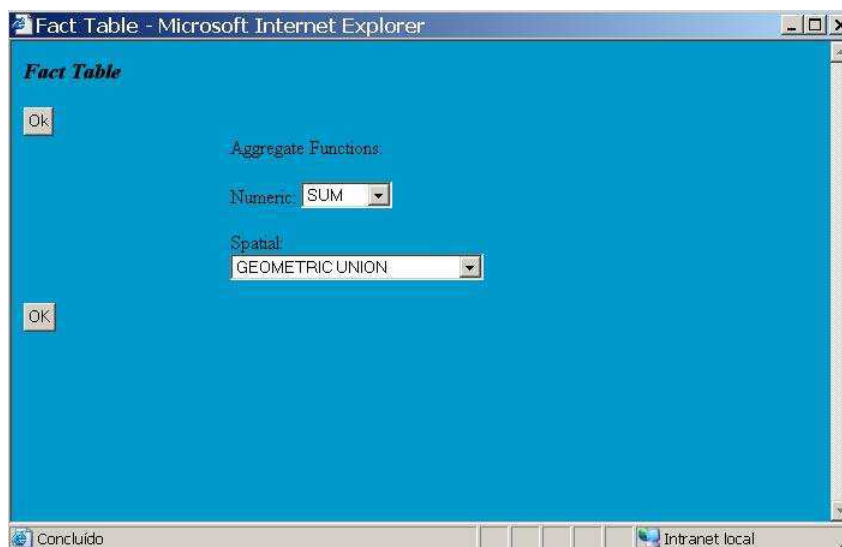


Figura 26 - Definindo funções de agregação

Pronto, a consulta acaba de ser definida. Agora, basta o usuário clicar no botão “Execute Query” na tela principal para obter os resultados.

A tela inicial de resultados pode ser vista na Figura 27, que apresenta tanto o resultado espacial no mapa quanto o analítico em tabela, de forma sincronizada. Observe que são apresentadas as áreas de plantação de milho (*corn*), feijão (*bean*) e algodão (*cotton*) dentro de uma janela espacial definida no Estado da Paraíba. Além disso, a tela inicial apresenta os dados no nível Microregião e para o mês de Janeiro (1). Se o usuário quiser ver as informações de outros meses basta clicar nos botões “Next” (ir para próximo mês) e “Previous” (ir para mês anterior) na parte superior da tela. E para navegar pelos diferentes níveis de agregação, basta escolher um dos níveis que estão disponíveis na parte superior da tela. No caso, apenas os níveis Microregião e Região estão disponíveis, de acordo com a consulta que foi realizada pelo usuário. Observe que se o usuário clicar no botão “Region” surgirá uma nova tela (Figura 28) contendo as áreas de plantação no nível Região.

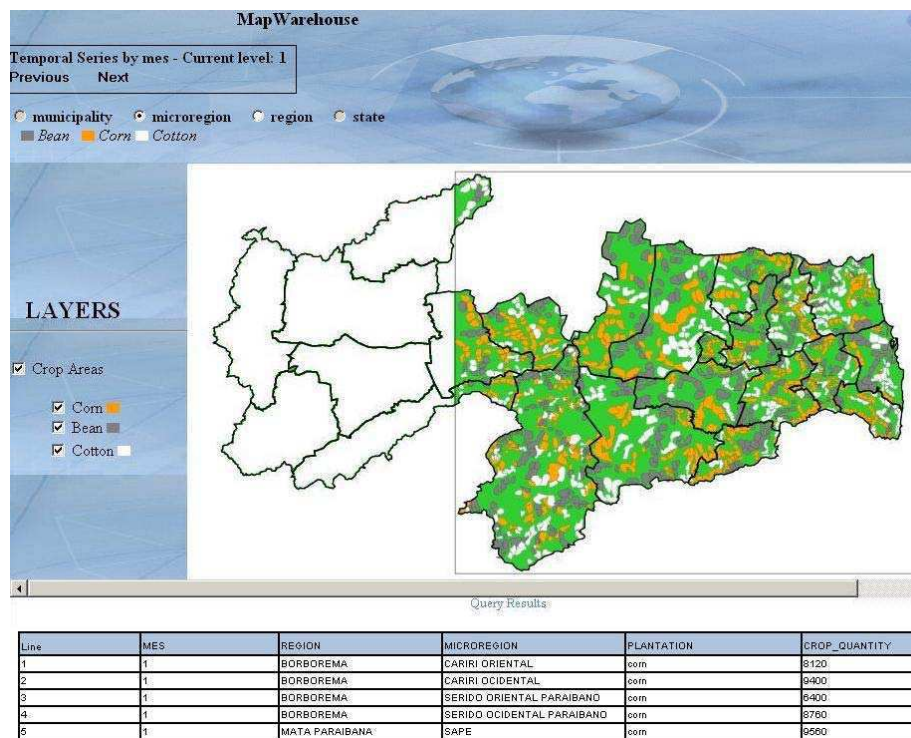


Figura 27 - MapWarehouse resultados no nível Microregião

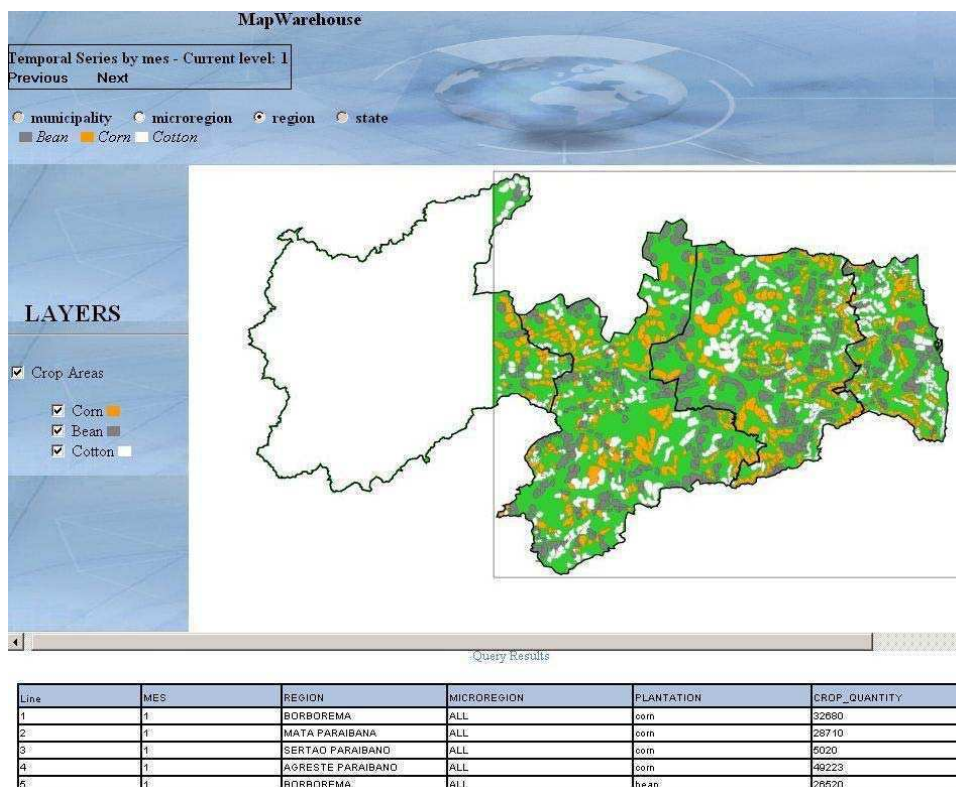


Figura 28 - MapWarehouse resultados no nível Região

Para facilitar a visualização de informações no mapa, o *MapWarehouse* apresenta os valores analíticos, referentes a objetos espaciais, no mapa. Para tal, basta o usuário passar o mouse sobre o objeto espacial, que serão apresentadas essas informações. Por exemplo, na

Figura 29 pode-se ver as informações referentes à Microregião do “Cariri Oriental”, indicando qual o total de plantação de milho, feijão e algodão nessa região. Essa operação funciona para qualquer nível espacial, como pode ser visto na Figura 30, no nível Região, no qual o usuário obtém informações no mapa da região da Borborema.

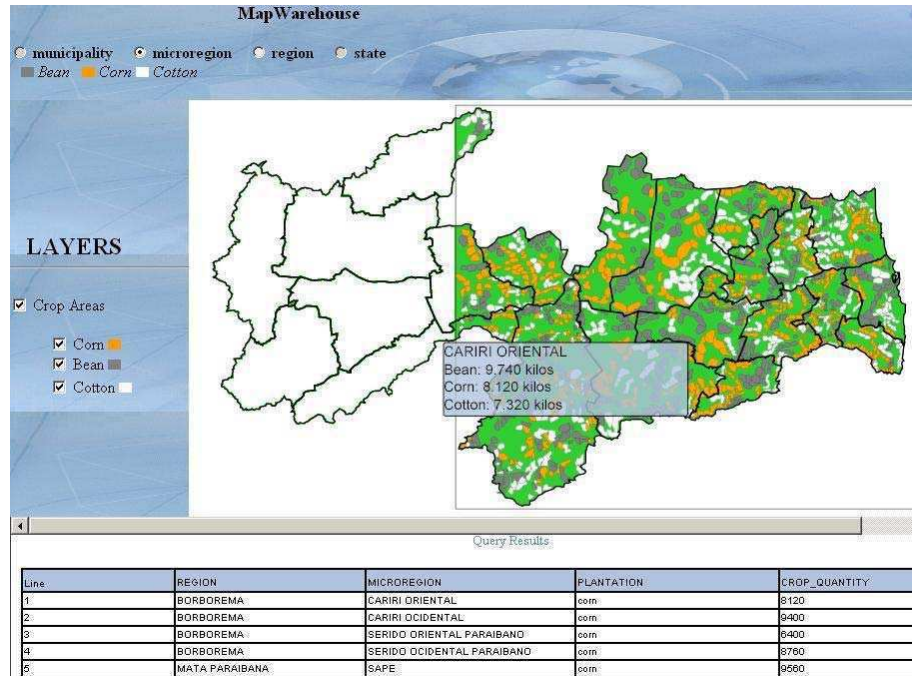


Figura 29 - MapWarehouse Informações no mapa por Microregião

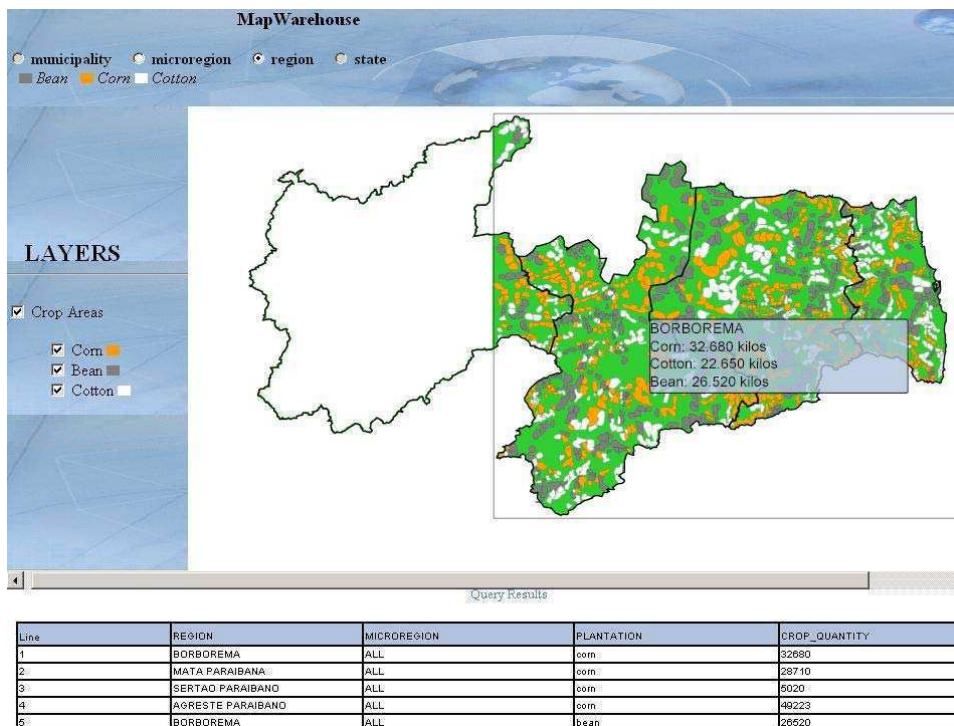


Figura 30 - MapWarehouse Informações no mapa por Região

6.2 Análise do Desempenho de Consultas SOLAP

Para validar a eficiência dos algoritmos de otimização do *MapWarehouse* foram realizadas inúmeras consultas baseadas em um plano básico de testes no qual cada consulta é executada utilizando quatro abordagens:

i) *Sem otimização espacial (SO)*: não utiliza nenhuma técnica de otimização espacial, ou seja, as consultas são realizadas sobre o esquema em estrela OR espacial básico, sendo as agregações obtidas de operações de união das geometrias armazenadas nos níveis de menor granularidade;

ii) *Com otimização espacial e casamento parcial com agregados espaciais (COP)*: ocorre nos casos em que o nível de agregação espacial solicitado na consulta é associado a um determinado agregado espacial em um nível intermediário, por exemplo, o usuário deseja obter os dados no nível Mesoregião e é utilizado um agregado espacial no nível Microregião;

iii) *Com otimização espacial e casamento total com agregados espaciais (COT)*: ocorre nos casos em que o nível de agregação espacial solicitado é exatamente o mesmo do agregado espacial ao qual a consulta foi associada, não necessitando de nenhuma operação de agregação espacial, mas apenas de recuperar os dados pré-armazenados. Por exemplo, o usuário deseja obter os dados no nível Microregião e é utilizado um agregado espacial no nível Microregião;

iv) *Com otimização espacial utilizando concatenação de geometrias (COC)*: nesse caso, nenhuma operação de agregação espacial é realizada, sendo apenas criada uma coleção de todas as geometrias que são apresentadas no mapa, gerando uma agregação meramente visual.

Para cada execução é calculado o tempo de resposta médio obtido utilizando cada uma das abordagens acima; desta forma, é possível avaliar qual o ganho obtido com a otimização espacial (COP, COT e COC) em relação a abordagem que não utiliza as técnicas de otimização (SO). Além disso, também é feita uma comparação entre as abordagens de otimização que utilizam união de geometrias (COP, COT) e a otimização por concatenação (COC) que apenas apresenta geometrias no mapa. Os resultados serão apresentados através de gráficos e tabelas.

Para os experimentos, dados espaciais e analíticos relacionados ao Estado da Paraíba, no Brasil, foram utilizados, de acordo com o esquema em estrela espacial OR da Figura 14, com dados históricos de 1999 a 2005. O tamanho total do banco de dados é de aproximadamente 30.000.000 tuplas (quase 12Gb).

O computador servidor utilizado foi um Athlon XP com processador de 2.1 GHz e 769MB de memória RAM. O browser Web para interface gráfica foi o Microsoft Internet Explorer 6.0 incrementado com o *plugin* Adobe SVG Viewer 3.0. O servidor de aplicação utilizado foi o Apache Tomcat/5.0.16.

A consulta base dos experimentos foi a seguinte: *Recuperar as áreas de plantação de milho do Estado da Paraíba, que estejam dentro de uma janela espacial, produzidas entre <mes1-ano1> e <mes2-ano2>, para os cinco primeiros meses de cada ano de 1999 a 2005, por mês e por Mesoregião.* Ou seja, uma operação que agrega as áreas de plantação de milho no nível de agregação espacial Mesoregião da dimensão espacial, sendo as informações distribuídas em séries temporais por mês. A seguir, essa consulta é definida na linguagem Oracle SQL OR, sendo MES1, MES2 os meses inicial e final da série temporal, respectivamente, e ANO, um dos anos entre 1999 e 2005:

```
Select d.Tempo_ref.Mes, a.Localizacao_ref.Geometria_Regiao, SDO_AGGR_UNION(MDSYS.SDOAGGRTYPE(DEREF(d.AreaPlantacao),
0.005)), SUM (d.QuantidadePlantacao)
From DistribuicaoAgricolaTable d
Where d.Plantacao_ref.Nome = 'milho' And d.Tempo_ref.Mes >= MES1 And d.Tempo_ref.mes <=MES2 And d.Tempo_ref.Ano
= ANO And d.Localizacao_ref.Nome_Estado = 'Paraíba' And SDO_INSIDE(d.areaPlantacao,
SDO_GEOMETRY(2003, 8307, NULL, SDO_ELEM_INFO_ARRAY(1,1003,3), SDO_ORDINATE_ARRAY(-37.1, -6.0, -34.00, -9.0))
) = 'TRUE'
Group by d.Tempo_ref.Mes, d.Localizacao_ref.Geometria_Regiao)
```

Com a variação dos parâmetros da consulta acima, foi calculado o tempo médio de resposta obtido com cada abordagem de otimização para consultas que consideram cada um dos cinco meses. Por exemplo, para obter o tempo de resposta médio dos dois primeiros meses de cada ano, são executadas cinco consultas considerando os anos de 1999 a 2005, com os parâmetros MES1=1, MES2=2 e ANO=1999; MES1=1, MES2=2 e ANO=2000; e assim por diante.

Para testar a abordagem SO executamos a consulta sobre o esquema em estrela OR espacial básico, no nível espacial Município e no nível temporal dia. Para a abordagem COP, utilizamos um agregado espacial (visão materializada) pré-armazenado no nível de agregação espacial Microregião, um nível intermediário entre Município e Mesoregião. No caso da COT, utilizamos o agregado espacial no nível de agregação espacial Mesoregião. Por fim, para testar a abordagem COC realizamos a consulta sobre o esquema básico, assim como na SO, mas sem realizar operações de agregação espacial.

Para os experimentos, consideramos que o tempo de resposta de uma consulta é obtido a partir da soma dos tempos: 1) tempo para executar a consulta no banco de dados; e 2) tempo para montar e exibir o mapa gerado como resultado da consulta.

Os resultados obtidos a partir desses experimentos podem ser vistos na Figura 31. Note que cada ponto representa o tempo médio de resposta obtido com uma determinada quantidade de meses. Por exemplo, se a série temporal considera apenas o mês de Janeiro então temos 1 mês, se considera os meses de Janeiro a Março temos 3 meses, e assim por diante. Considere, por exemplo, o ponto (3, ≈ 1350) da Figura 31: ele representa um tempo de resposta médio de aproximadamente 1350 segundos para uma série temporal que considera os três primeiros meses (Janeiro, Fevereiro e Março) de cada um dos anos de 1999 a 2005, sem utilizar técnica de otimização (SO). O mesmo vale para os outros pontos, ou seja, o ponto (i, *avg*) representa o tempo de resposta médio *avg* obtido para uma série temporal que considera os i primeiros meses de cada ano, entre 1999 e 2005, usando uma determinada técnica de otimização.

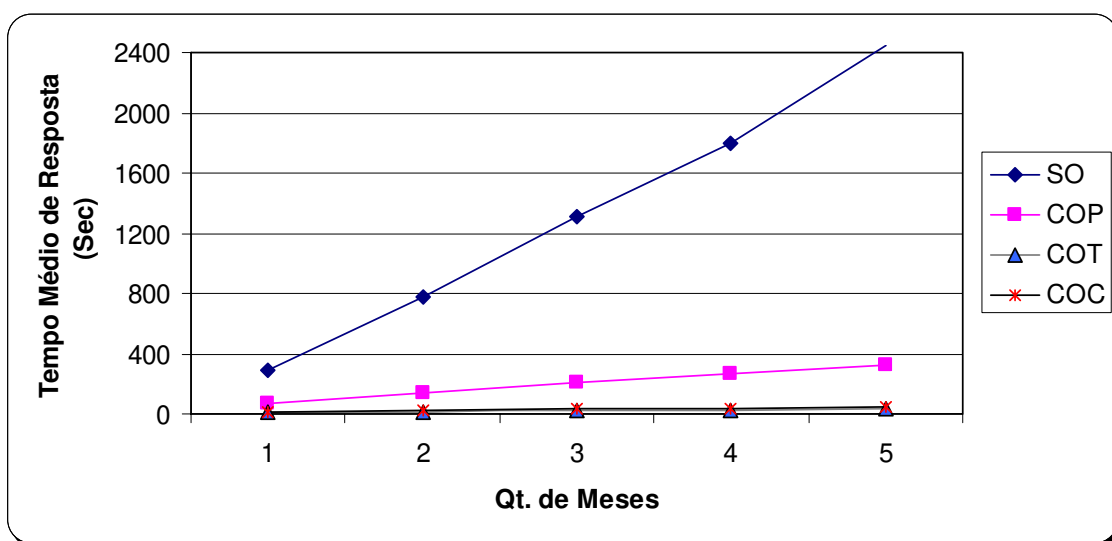


Figura 31 - Resultado dos experimentos para as abordagens de otimização

Por questões de visualização da informação, também apresentamos esse gráfico em forma de histograma, na Figura 32. O mesmo processo também será utilizado para os próximos gráficos.

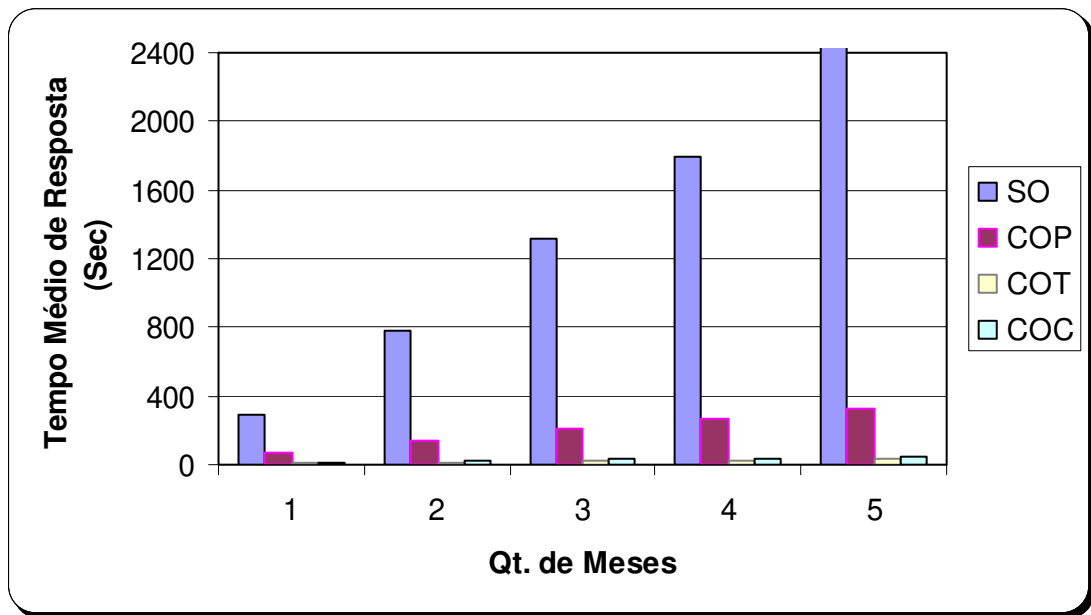


Figura 32 - Resultado dos experimentos para as abordagens de otimização (Histograma)

Nos gráficos das Figuras Figura 31 e Figura 32, fica claro que à medida que o número de meses vai aumentando (maior quantidade de dados consultados) o tempo de resposta obtido sem utilizar nenhuma forma de otimização vai aumentando muito rapidamente quando comparado com os tempos obtidos com o uso das técnicas de otimização. Nas situações em que a otimização de consultas foi utilizada, o tempo de resposta também cresce com a quantidade de meses, só que de forma menos acentuada. Note que a abordagem SO, sem utilizar otimização, apresenta um custo proibitivo para grandes quantidades de dados. Os tempos obtidos com as abordagens de otimização de consultas são melhor discutidos a seguir.

Nas Figuras Figura 33 e Figura 34, utilizamos os mesmos experimentos para fazer uma comparação apenas entre as abordagens que utilizam otimização.

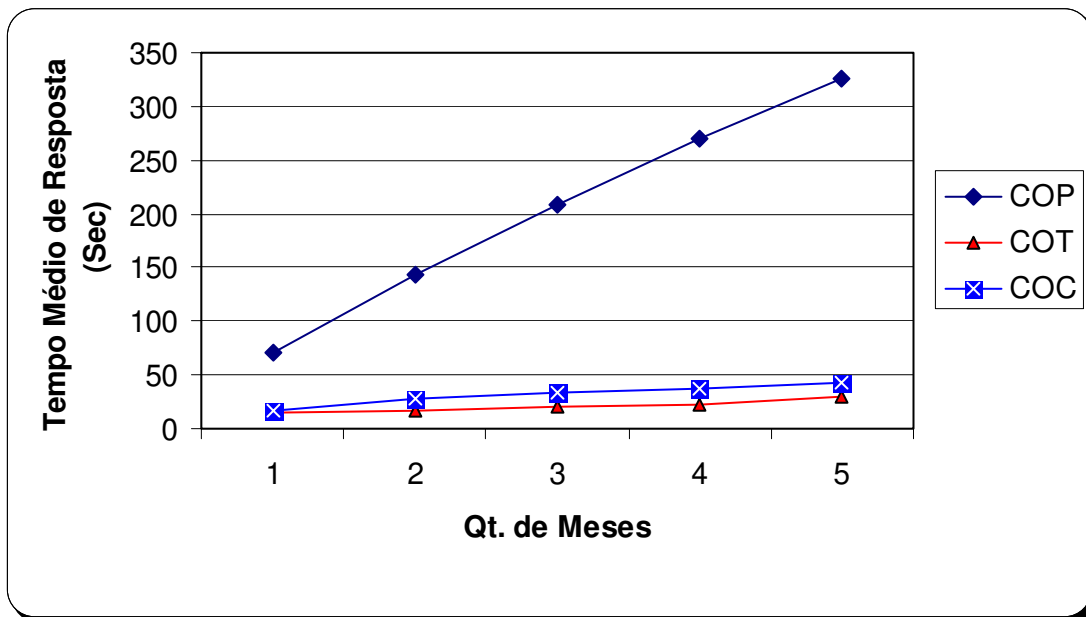


Figura 33 - Comparação entre abordagens de otimização de consultas

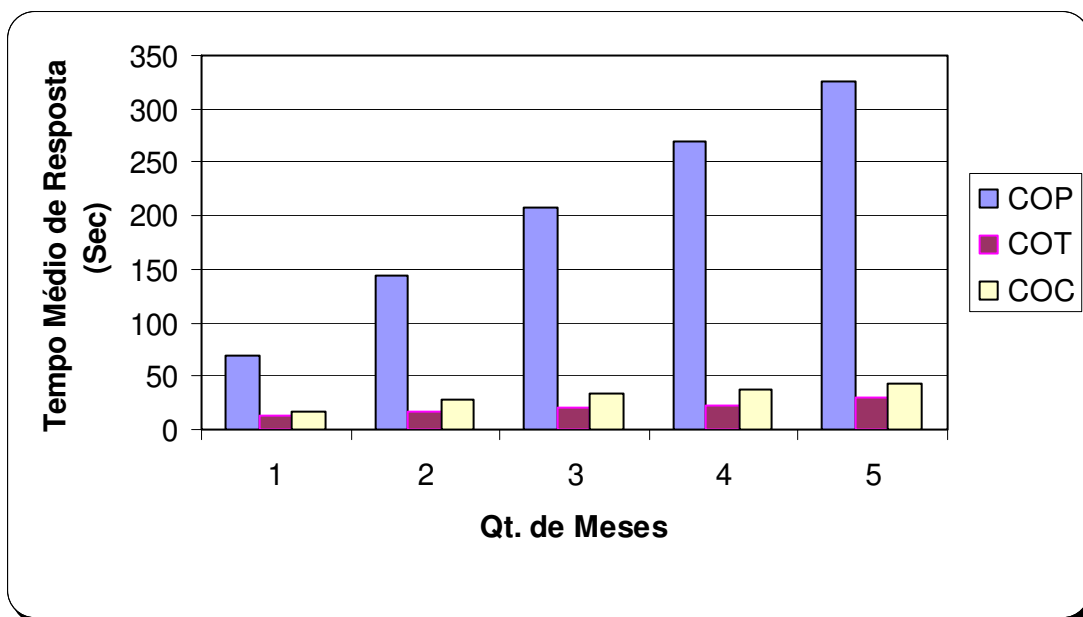


Figura 34 - Comparação entre as abordagens de otimização de consultas
(Histograma)

Note que a abordagem COP tem um crescimento mais acentuado que as outras abordagens de otimização, o que se justifica pelo seguinte fato: mesmo utilizando otimização de consultas com agregados espaciais, quando a quantidade de dados aumenta, também aumenta a quantidade de operações de agregação espacial e conseqüentemente o tempo de resposta. E isso não acontece para COT e COC pelo fato das mesmas não precisarem executar nenhuma operação de agregação espacial. Desta forma, apesar de COP apresentar um ganho

considerável em relação a abordagem SO, ela pode apresentar um custo elevado em consultas críticas com grandes quantidades de dados e operações de agregação espacial.

Observe na Figura 34 que para 1 (um) mês os tempos de resposta médio utilizando as abordagens COT e COC estão muito próximos. Podemos inferir desse gráfico que a partir de um certo ponto, com o aumento do número de meses, a técnica de otimização COT passa a ter um desempenho um pouco superior a COC; esse seria o ponto de inflexão a partir do qual COT passaria a ser mais vantajosa que COC.

Pelo fato das abordagens COT e COC terem tido um desempenho bastante parecido, decidimos comparar as mesmas entre si, o que pode ser visto nas Figura 35 e Figura 36. Nessas figuras, nota-se que o desempenho da otimização COT (otimização com casamento total) sempre, após o ponto de inflexão, tem um desempenho um pouco superior ao da abordagem COC (otimização por concatenação), isso pode ser explicado pelo seguinte fato: ambas não realizam nenhuma operação de união de geometrias, independentemente da quantidade de meses; entretanto, quanto maior a quantidade de geometrias, maior o tempo gasto para montar e exibir o mapa, o que representa uma desvantagem para o COC, que sempre seleciona as geometrias do esquema em estrela OR espacial básico. Já para o COT, a quantidade de geometrias necessárias para montar o mapa é bem menor, pois geometrias representando agregados espaciais já estão pré-armazenadas.

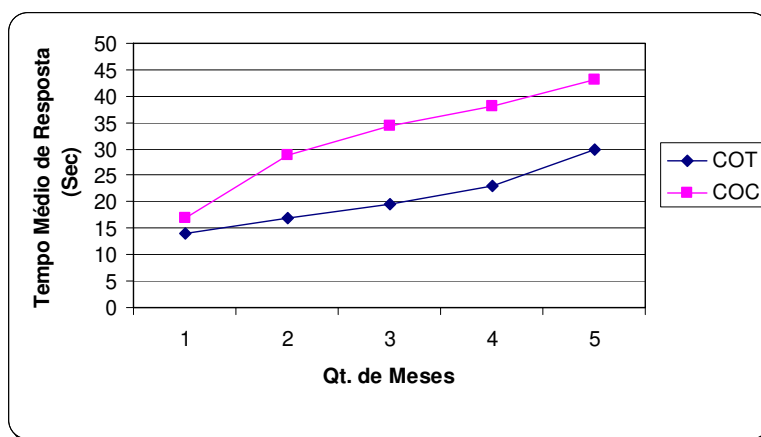


Figura 35 - Comparação entre as abordagens COT e COC

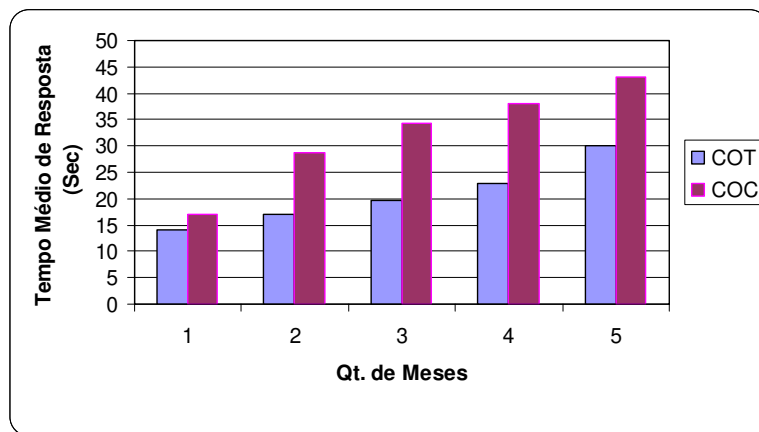


Figura 36 - Comparação entre as abordagens COT e COC (Histograma)

De uma forma geral, podemos afirmar que na maioria dos casos, a otimização por concatenação é a mais indicada, pois o tempo de resposta médio obtido com essa abordagem é sempre inferior ao tempo gasto com a otimização em casamento parcial, que representa a situação mais comumente encontrada nas consultas ao DWE. Para consultas em que ocorra casamento total, o tempo de resposta médio é ainda melhor que o gasto com otimização por concatenação. Mas, pelo fato da otimização com visões materializadas por casamento total acontecer mais raramente (não é realista) e das operações que utilizam visões materializadas serem mais complicadas e exigirem mais espaço para o pré-armazenamento de agregações, a abordagem de otimização por concatenação é indicada nas situações em que possa ser aplicada.

6.2.1 Indicador de Ganho de Performance (*Speedup*)

Com intuito de definir um indicador de ganho de performance obtido com as técnicas de otimização em relação a abordagem que não utiliza nenhuma técnica de otimização, nós medimos: *speedup1* entre COP e SO; *speedup2* entre COT e SO; e *speedup3* entre COC e SO, sendo:

$$Speedup1 = \text{tempo gasto com COP} / \text{tempo gasto com SO}$$

$$Speedup2 = \text{tempo gasto com COT} / \text{tempo gasto com SO}$$

$$Speedup3 = \text{tempo gasto com COC} / \text{tempo gasto com SO}$$

Note que um *speedup* maior que 1 implica que a técnica de otimização tem uma melhor performance. A partir dos resultados obtidos nos experimentos para os cinco primeiros meses de cada ano, geramos curvas *speedup* utilizando a técnica de interpolação polinomial quadrática; as mesmas são descritas na Figura 37. Nesse gráfico, fica claro que o *speedup* é

sempre maior que 1 e que aumenta de forma não-linear com o número de meses, sendo o melhor desempenho obtido com a técnica COT.

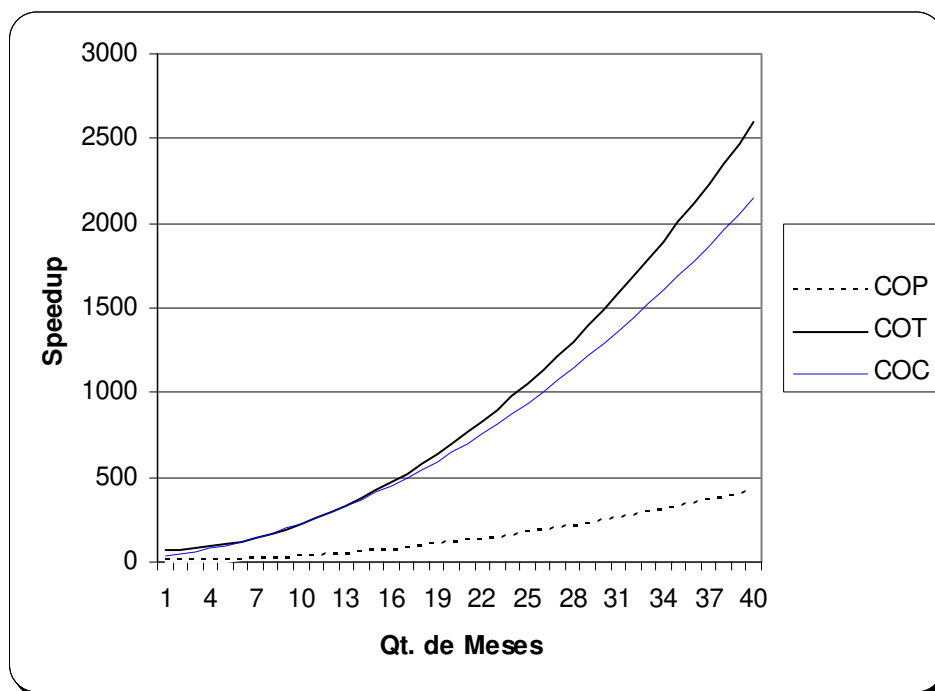


Figura 37 - Curva mostrando o indicador de ganho de desempenho *speedup*

De forma geral, notamos que com a utilização das técnicas de otimização de consultas baseadas em agregados espaciais pré-armazenados ou em concatenação de geometrias, o tempo de resposta total diminui bastante. Esse ganho torna-se notável nos grandes bancos de dados.

6.3 Considerações Finais

Neste capítulo, foi apresentado o protótipo *MapWarehouse* para a validação das idéias propostas nesse trabalho. A partir dele, é possível criar consultas SOLAP para um Data Warehouse espacial e visualizar o resultado das mesmas de forma tabular e espacial. Também foi mostrada uma avaliação experimental que constatou os ganhos de performance obtidos com a utilização dos algoritmos de otimização de consultas multidimensionais espaciais propostos.

Capítulo 7

Conclusão

Existem várias propostas na literatura visando à integração das funcionalidades e características pertinentes ao processamento de dados analíticos e geográficos. O principal objetivo é prover um ambiente único, com capacidades de processamento geográfico-multidimensional, para dar suporte ao processo de tomada de decisões estratégicas. Portanto, a necessidade de inserir o componente espacial nesse tipo de DW motiva a integração entre Data Warehouse e Sistemas de Informação Geográfica, surgindo assim um novo conceito, chamado de Data Warehouse espacial, que exige um novo modelo de DW composto de operações analíticas e espaciais.

Os conceitos básicos de DW e uma discussão que motiva a integração do mesmo com dados espaciais foram discutidos no capítulo de Introdução. Os novos conceitos de Data Warehouse espacial foram apresentados no capítulo 2, entre eles: dados espaciais, dimensão espacial, hierarquia espacial, medida espacial e agregação espacial.

A partir da análise feita no capítulo 3, foram identificadas algumas lacunas deixadas nas pesquisas referentes à integração entre dados analíticos e espaciais. As principais foram:

- a maioria das abordagens não definem formalmente um modelo multidimensional espacial (o modelo proposto nesse trabalho não define formalmente as operações SOLAP);
- poucas propostas tratam de operações SOLAP como Drill-down e Roll-up espacial;
- com relação as abordagens que apresentam interface gráfica, nenhuma delas fornece uma interface que possibilite ao usuário realizar consultas SOLAP utilizando menus iterativos e sem precisar saber detalhes de sintaxe da linguagem de consulta ao DWE;
- apenas duas propostas (ZHANG, 2003) (FIDALGO et al., 2004) possibilitam ao usuário lidar com consultas espaciais ad-hoc, com definição dinâmica de critérios espaciais para as consultas;
- nenhuma proposta trata de consultas espaciais com séries temporais;
- a maioria das propostas de integração entre DW e SIG não tratam o objeto espacial como objeto de análise (como medida espacial), mas apenas com critérios de agregação (dimensão espacial) para medidas numéricas, dificultando

a realização de operações de agregação espacial, muito importantes no auxílio à tomada de decisão.

Todas as propostas apresentadas, representam alguns dos esforços no sentido de amenizar essas carências, sendo que cada uma delas também apresenta algumas limitações, principalmente no que diz respeito ao aspecto espacial. As pesquisas de ambiente colaborativo entre as tecnologias Geográficas e Multidimensionais são recentes, o que justifica essas carências.

No capítulo 4, foi proposto um modelo conceitual multidimensional espacial, com dimensões espaciais, hierarquias espaciais e medidas espaciais, adequado para realização de operações de Roll-up espacial, Drill-down espacial e slice/dice espacial e que supre grande parte das lacunas de pesquisa identificadas nos trabalhos relacionados. Trata-se de um modelo de integração estreita entre DW e SIG que foi formalmente definido utilizando os formalismos UML e ODL/OQL/ODMG. Pelo fato de ter medidas espaciais, o modelo também tratou de questões relacionadas a operações de agregação de medidas espaciais, como a operação de união de geometrias. Foram definidas formalmente regras de mapeamento do modelo multidimensional espacial para um esquema em estrela espacial Objeto Relacional, implementável em qualquer SGBD OR capaz de lidar com informações espaciais.

A implementação do modelo foi feita através de um esquema estrela espacial OR, utilizando o SGBD OR Oracle 10g. Acreditamos que ainda não existe nenhuma proposta de integração entre DW e SIG com uma implementação baseada em um esquema em estrela espacial OR.

Para garantir o bom desempenho das consultas espaciais, foram utilizadas técnicas de otimização. O problema de otimização de consultas foi discutido no capítulo 5, que apresentou duas possíveis técnicas de otimização de consultas: otimização de consultas utilizando agregados espaciais e otimização de consultas utilizando concatenação de geometrias.

No capítulo seguinte, as idéias propostas no modelo foram validadas através de um protótipo chamado *MapWarehouse* e de uma avaliação experimental.

No protótipo, foi definida uma interface Web que possibilita o usuário realizar consultas SOLAP ao DWE utilizando menus interativos e sem a necessidade de lidar com detalhes de sintaxe da linguagem de consulta. O usuário pode realizar operações como Drill-down e Roll-up espacial, bem como realizar consultas espaciais ad-hoc, definindo janelas espaciais no mapa. Dependendo da consulta criada pelo usuário, o resultado pode ser apresentado de forma sincronizada em mapas e em forma tabular. As informações analíticas também podem ser

apresentadas no próprio mapa, no momento em que o usuário passa o mouse sobre objetos espaciais.

A avaliação experimental mostrou ganhos relevantes obtidos com a utilização das abordagens de otimização, tanto as baseadas em agregados espaciais pré-armazenados quanto a abordagem de concatenação de geometrias. Com a utilização da técnicas de otimização o tempo de reposta médio às consultas foi sempre inferior ao tempo que seria obtido se a consulta fosse realizada sem nenhum tipo de otimização e esse ganho torna-se notável em grandes bancos de dados, pois cresce com o aumento da quantidade de dados consultados. Os experimentos mostraram que a otimização por concatenação é a mais indicada, pois apresentou ganhos superiores aos obtidos utilizando a técnica de otimização com casamento parcial, que representa a situação mais realista de consultas a DWE. A técnica de otimização com casamento total obteve os maiores ganhos, mas não representa uma situação realista, na qual os dados estariam sempre pré-armazenados em todos os possíveis níveis espaciais de agregação espacial a serem consultados.

As principais contribuições desse trabalho encontram-se resumidas no Quadro 5.

Quadro 5 - Principais contribuições do trabalho

Características Propostas	Abordagem Federada X Integrada	Definição formal de Modelo Multidimensional Espacial	Operações SOLAP	Medidas espaciais	Agregação Espacial	Otimização de consultas SOLAP	Protótipo de validação	Implementação em SGBD OR	Interface para criação de consultas	Consultas ad-hoc
Matousek et al.	Federada	Não	Não	Não	Não	Não	Sim	Não	Não	Não
Ferri et al.	Federada	Não	Não	Não	Não	Não	Não	Não	Não	Não
Han et al.	Integrada	Não	Sim	Sim	Sim	Sim	Sim	Não	Não	Não
Shekhar et al.	Integrada	Não	Sim	Não	Não	Não	Sim	Não	Não	Não
Bimonte et al.	Integrada	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Não
Zhang et al.	Integrada	Não	Sim	Não	Não	Sim	Sim	Não	Não	Sim
Fidalgo et al.	Integrada	Não	Sim	Não	Não	Não	Sim	Não	Não	Sim
<u>Nossa proposta</u>	<i>Integrada</i>	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>

Os resultados mostram que o modelo multidimensional espacial proposto é bastante viável e que atendeu os objetivos almejados. As contribuições dessa pesquisa também foram reconhecidas através de duas publicações internacionais importantes:

- um artigo no *workshop* DOLAP 2006, intitulado *Towards a Logical Multidimensional Model for Spatial Data Warehousing and OLAP* (SAMPAIO, SOUSA e BAPTISTA, 2006);

- um capítulo no livro “*Intelligent Databases: Technologies and Applications*”, do *Idea Group*, intitulado *Enhancing Decision Support Systems With Spatial Capabilities* (SAMPAIO et al., 2006).

Esse trabalho representa mais um passo no avanço das pesquisas relacionadas a Data Warehouse espacial, preenchendo lacunas importantes. Entretanto, alguns pontos ainda podem ser pesquisados e desenvolvidos. Os mesmos, são descritos a seguir, juntamente com algumas sugestões para novas funcionalidades para a ferramenta *MapWarehouse*.

7.1 Trabalhos Futuros

As pesquisas relacionadas a integração entre Data Warehouse e Sistemas de Informação Geográfica ainda são incipientes, gerando margem para uma grande quantidade de trabalhos de pesquisa. Com a proposta de um novo modelo de integração estreita entre dados analíticos e espaciais e sua respectiva implementação, o *MapWarehouse*, identificou-se algumas questões importantes para o aperfeiçoamento do mesmo. As sugestões para trabalhos futuros são as seguintes:

Manutenção de agregados espaciais Durante o desenvolvimento do sistema, foram criados agregados espaciais (ou visões materializadas espaciais) de forma planejada, para serem úteis na realização de experimentos necessários para a avaliação de desempenho dos algoritmos de otimização propostos. Ou seja, os algoritmos de otimização partiram da hipótese de que os agregados espaciais já tinham sido criados previamente. Entretanto, a criação e atualização de agregados espaciais, que requer um esforço significativo por parte dos administradores de um DWE, não foi tratada nesse trabalho. O ideal, seria definir uma política de criação, armazenagem e indexação de visões materializadas para dados espaciais. Também seria necessário um tratamento específico para a atualização dos agregados espaciais.

MapWarehouse para outros estudos de caso O *MapWarehouse* poderia ser utilizado em outros estudos de caso de integração entre dados analíticos e espaciais, utilizando diferentes implementações do esquema em estrela espacial OR.

MapWarehouse em outros SGBD Sugerimos a implementação do modelo proposto em outros SGBD Objeto-Relacional que tenham suporte a dados e operações espaciais.

<i>WebServices</i>	Um outro trabalho interessante seria adaptar a arquitetura do <i>MapWarehouse</i> para uma arquitetura orientada a serviços, através de Web services (SILVA et al., 2004). Este é um ponto importante que vem sendo investigado pelo consórcio <i>OpenGeospatial</i> em uma iniciativa chamada Geo-Decision Support Services (GeoDSS) (http://www.opengeospatial.org). Desta forma, DWE distribuídos poderiam ser disponibilizados e SOLAP poderia se tornar um serviço automaticamente descoberto e utilizado.
<i>Interface do MapWarehouse</i>	A interface do <i>MapWarehouse</i> poderia ser melhorada em termos de usabilidade e também através da inclusão de outros operadores SOLAP e outras funções de agregação espacial.
<i>Representação dos dados espaciais</i>	Sugerimos a adaptação do <i>MapWarehouse</i> para que o mesmo também seja capaz de apresentar dados espaciais através da representação matricial (ou raster) (COUCLELIS, 1992).
<i>Experimentos adicionais</i>	Novos experimentos poderiam ser feitos para permitir a obtenção de resultados decorrentes de análise estatística.

Referências Bibliográficas

ADAM, N. R.; GANGOPADYAY A. Database Issues in Geographic Information Systems – Kluwer Academic Publishers. Boston/Dordrecht/London – Second Printing. (1998).

BAPTISTA, C. S. et al. Usando Tecnologias J2EE e SVG para disponibilização de mapas na web. In: GIS BRASIL 2004, 2004, São Paulo. Gis Brasil 2004. 2004.

BIMONTE, S.; TCHOUNIKINE, A.; MIQUEL, M. Towards a Spatial Multidimensional Model. In Proceedings of the Data Warehousing and OLAP Conference (DOLAP'05), 2005, 39-46.

CAMARA, G. et al. Anatomia de Sistemas de Informação Geográfica. Instituto de Computação, UNICAMP. Campinas, 1996.

COUCLELIS, H. People Manipulate Objects (but Cultivate Fields): Beyond the Raster-Vector Debate in GIS. In: Proc. International Conference on GIS - From Space to Territory: Theories and Methods of Spatial Reasoning. Springer Lecture Notes on Computer Science, vol. 639, pp. 65-77, 1992.

DAMIANI, M. L.; SPACCAPIETRA S. Spatial Data Warehouse Modeling. Processing and Managing Complex Data for Decision Support, Darmont, J. & Boussaid, O. (Eds), IDEA Group Publishing, 2006, p. 1-27.

DATE, C. J. An Introduction to Database Systems – 7ª Ed. Addison-Wesley (2000).

FERRI, F. et al. Extending Geographic Databases for a Query Language to Support Queries Involving Statistical Data, Proc. of the 12th International Conference on Scientific and Statistical Database Management (SSDBM'00), p. 220, 2000

FIDALGO, R. N. et al. Providing multidimensional and geographical integration based on a gdw and metamodels. In: Brazilian Symposium on Databases (SBBD), 2004.

FIDALGO, R. N.; TIMES, V. C.; SOUZA, F. F. GOLAPA: Uma Arquitetura Aberta e Extensível para Integração entre SIG e OLAP, Proc. GeoInfo, 2001

FRANKLIN, C. An Introduction to Geographic Information Systems: Linking Maps to databases. Database. 1992, 13-21.

GOAL. Disponível em: <<http://krizik.felk.cvut.cz/goal>>. Acessado em: 28/11/2006.

GOLAPA. Disponível em: <<http://php.cin.ufpe.br/~golapa/architecture/>>. Acessado em: 30/04/2006

GOLDSTEIN, J.; LARSON, P. Optimizing Queries Using Materialized Views: a Practical, Scalable Solution. In: Proceedings of the ACM SIGMOD, 2001, p. 331-342.

GUTTMAN, A. R-Trees: A Dynamic Index Structure for Spatial Searching, Proc. ACM SIGMOD International Conference on Management of Data, pp. 47-57, 1984

- HAN, J.; KOPERSKI, K.; STEFANOVIC N. GeoMiner: A System Prototype for Spatial Data Mining – Proc. ACM SIGMOD (1997).
- HAN, J.; KOPERSKI, K.; STEFANOVIC, N. Selective Materialization: An efficient method for spatial data cube construction – In PAKDD (1998).
- INDULSKA, M., ORLOWSKA, M. E. On Aggregation Issues in Spatial Data Management. Thirteenth Australasian Database Conference (ADC2002), 2002, p. 75-84.
- IOCHPE, C., LISBOA, J. Introdução a Sistemas de Informações Geográficas com Ênfase em Banco de Dados – Porto Alegre: CPGCC da UFRGS. (1996).
- KIMBALL, R.; CASERTA, J. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. John Wiley & Sons, 2004
- KIMBALL, R.; ROSS, M.; MERZ, R. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. John Wiley & Sons, 2002.
- KIMBALL, R. The Data Warehouse Toolkit – John Wiley & Sons, Inc. (1996).
- MALINOWSKI, E. Concepts and methodological framework for spatio-temporal data warehouse design. Diplôme d'Etudes Approfondies en Sciences Appliquées, Faculté des Sciences Appliquées, Université Libre de Bruxelles, Belgique, 2003.
- MALINOWSKI, E.; ZIMANYI, E. Representing Spatiality in a Conceptual Multidimensional Model. In: Proceedings of the 12th ACM International Symposium on Advances in Geographic Information Systems (ACM GIS 2004), 2004, p. 12-21.
- MATOUSEK, K.; KOUBA, Z.; MIKSOVSKY, P. On Data Warehouse and Gis Integration, Proc. 11th International Conference on Database and Expert Systems Applications (DEXA2000), Greenwich, UK, September, 2000
- MATOUSEK, K.; KOUBA, Z.; MIKSOVSKY, P. On Geographical On-Line Analytical Processing (GOLAP). ISAS-SCI (1): 201-205, 2001
- MATOUSEK, K.; MORDACIK, J.; JANKU, L. On Implementing the Data Warehouse - GIS Integration. ISAS-SCI (1) 206-210, 2001
- MATOUSEK, K.; SVOBODA, L. Extending GIS by Data Warehouse, Proc. of International Carpathian Control Conference. Vol. 1. TU FEI, Kosice, 2000
- Object Management Group (OMG). Meta object facility (mof) specification. Technical report, Object Management Group, 2002.
- OGC Web Services, Phase 4 (OWS-4), Geo Decision Support (GeoDSS), Disponível em: <<http://www.opengeospatial.org>>. Acessado em: 20/02/2007
- Oracle. Oracle Spatial User's Guide and Reference, Release 9.0.1. Oracle Corporation, 2001.
- PAPADIAS, D. et al. Efficient OLAP Operations in Spatial Data Warehouses. International Symposium on Spatial and Temporal Databases (SSTD), p. 443-459, 2001.

- POURABBAS, E. Cooperation with Geographic Databases. In: *Multidimensional Databases: Problems and Solutions*. Idea Group Inc., IGP/INFOSCI/IRM Press, Hershey, USA, 2003.
- RIVEST, S.; BÉDARD, Y.; MARCHAND P. Towards better support for spatial decision-making: Defining the characteristics of Spatial On-Line Analytical Processing (SOLAP), *Geomatica: The journal of the Canadian Institute of Geomatics*, 2001
- SAMPAIO, M. C. et al. Enhancing Decision Support Systems With Spatial Capabilities. In: Zongmin Ma. (Org.). *Intelligent Databases: Technologies and Applications*. : Idea Group Publishing, 2006, v. , p. 94-116.
- SAMPAIO, M. C.; SOUSA, A. G.; BAPTISTA, C. S. Towards a Logical Multidimensional Model for Spatial Data Warehousing and OLAP. In: *DOLAP, 2006, Arlington. ACM Ninth International Workshop on Data Warehousing and OLAP - DOLAP, 2006*.
- SCOTCH, M.; PARMANTO, B. SOVAT: Spatial OLAP Visualization and Analysis Tool. In: *proceedings of the 38th Hawaii International Conference on System Sciences*. Waikoloa, HI. 2005.
- SHEKHAR, S.; CHAWLA, S. (2003) *Spatial Database: a Tour*. Morgan Kaufmann.
- SHEKHAR, S. et al. Map Cube: A Visualization Tool for Spatial Data Warehouses – Chapter of *Geographic Data Mining and Knowledge Discovery*. Harvey J. Miller and Jiawei Han (eds.), Taylor and Francis, 2001
- SILVA, J. et al. Towards a web service for geographic and multidimensional processing. In: *VI Simpósio Brasileiro de GeoInformática, 2004, Campos do Jordão, SP, Brasil*. p. 2-17.
- SILVA, J.; TIMES, V. C.; SALGADO, A. C. An open source and web based framework for geographic and multidimensional processing. *SAC 2006*: p. 63-67
- STEFANOVIC, N. Design and Implementation of on-line analytical processing (OLAP) spatial data - M.Sc. Thesis, Simon Fraser University, Canada, 1997
- ZHANG, L. et al. An Approach to Enabling Spatial OLAP by Aggregating on Spatial Hierarchy. p. 35-44, *DaWaK, 2003*
- ZHANG, L. et al. Spatial hierarchy and OLAP-favored search in spatial data warehouse. p. 48-55, *DOLAP, 2003*
- ZISMAN, A.; KRAMER, J. Towards Interoperability in Heterogeneous Database Systems – Technical Report 11, Department of Computing, Imperial College of Science, Technology and Medicine. (1995).