

Processamento Estatístico e Apresentação de Dados Espaciais Distribuídos Via Web

Leonardo Arthur Esteves Lourenço¹

Antônio Miguel Vieira Monteiro²

Lúbia Vinhas³

¹Programa de Mestrado em Computação Aplicada – CAP
Instituto Nacional de Pesquisas Espaciais – INPE

^{2,3}Divisão de Processamento de Imagens – DPI
Instituto Nacional de Pesquisas Espaciais – INPE

{lourenco,miguel,lubia}@dpi.inpe.br

Abstract. *Spatial data are used by many different sectors of society and for different purposes. The Federal Government has dedicated efforts to facilitate access to public data collections. Therefore, sharing of spatial data is a strategic action of the State. However, the methods with digital spatial data are published, integrated and processed are still subject of much study and debate in the scientific community. The complexity of the problem increases when interesting data are spread in several databases and managed by different actors, this is the most common case. This article presents one of the strategies used for publication and integration of distributed spatial data, which is the Distributed Basis Architecture Based on Mediators, and a form to use this solution to perform processing and statistical operations with distributed spatial data, besides preparing these data to be presented to the requester. Part of the implementation, which is based on Web Services, is shown here with two services: a kernel estimator and a map slicing.*

Resumo. *Dados espaciais são utilizados pelos mais diversos setores da sociedade e para os mais diversos fins. O governo federal tem empenhado esforços em facilitar o acesso aos acervos de dados da administração pública. O compartilhamento de dados espaciais é, portanto, uma ação estratégica de Estado. No entanto, os métodos com que dados espaciais digitais são publicados, integrados e processados ainda são alvo de muita discussão e estudo na comunidade científica. A complexidade do problema aumenta quando os dados de interesse estão espalhados em diversas bases de dados e administrados por gestores diferentes, e este é o caso mais comum. Este artigo apresenta uma das estratégias utilizadas para publicação e integração de dados espaciais distribuídos, que é a Arquitetura de Bases Distribuídas Baseada em Mediadores e uma maneira de utilizar esta solução para realizar processamentos e operações estatísticas com dados espaciais distribuídos, além de preparar a forma com que estes dados são apresentados ao*

requisitante. Parte da implementação, que é baseada em Web Services, é mostrada aqui por meio de dois serviços: um estimador de intensidade kernel e um fatiamento de mapa.

Palavras Chave

Mediação, integração, dados distribuídos, dados espaciais, web services, processamento estatístico, apresentação de dados.

1. Introdução

Dados geográficos constituem importante ferramenta de apoio a decisões. Portanto, são parte dos processos de funcionamento das organizações, seja como objeto de estudo científico, fonte de informação ou base de soluções de problemas. Por essa razão, soluções que viabilizem acesso a dados espaciais são foco de muitos trabalhos encontrados literatura.

Às vezes, o dado por si só não atende as necessidades de quem o adquire. Nestes casos, o processamento e a apresentação (formatação) dos dados se tornam fases tão importantes quanto sua aquisição, pois apenas após um tratamento prévio consegue-se extrair conhecimento de um dado bruto.

Este trabalho apresenta parte da implementação de uma solução para processamento de dados geográficos distribuídos. A Arquitetura de Bases Distribuídas Baseada em Mediadores, que possibilita a aquisição e integração dos dados, foi utilizada como plataforma. Baseando-se nela, foram implementados dois serviços que operam sobre os dados adquiridos e que estão em fontes diferentes: um estimador de intensidade kernel e um fatiamento de mapas, mostrando a possibilidade de mediação de processos, e não apenas de dados, já que cada operação ocorre num servidor diferente.

2. Conceitos e Trabalhos Relacionados

2.1. Arquitetura de Bases de Dados Distribuídas

Bases de dados distribuídas são bases que não são partilhadas fisicamente, podendo estar separadas por grandes distâncias. As bases são independentes e podem ser gerenciadas por diferentes sistemas. Contudo, se há um sistema que consome e administra mais de uma base com essas características, ele possui uma Arquitetura de Bases de Dados Distribuídas, podendo realizar transações envolvendo uma ou mais bases simultaneamente e gerenciando todo o acesso remoto necessário.

Existem várias tecnologias e protocolos que implementam técnicas de acesso remoto entre sistemas. Neste artigo, é apresentada uma implementação baseada em *web services*.

2.2. Arquitetura Orientada a Serviços

No início da última década, o crescimento dos *Web Services* e dos padrões de suporte à integração automatizada de sistemas deram origem a uma tecnologia considerada um

avanço no domínio da interoperabilidade, a Arquitetura Orientada a Serviços (SOA), que visa atender aos requisitos de baixo acoplamento, padronização e independência entre os componentes participantes de uma aplicação (Papazoglou e Heuvel, 2007). A padronização de interfaces reduz o esforço dos desenvolvedores em compreender os requisitos de compartilhamento. A única necessidade é saber que o outro componente implementa a linguagem e as interfaces de intercâmbio.

As vantagens da abordagem SOA são observadas na facilidade de integração dos componentes dos sistemas, reuso de sistemas legados e tolerância a mudanças de tecnologia, o que agrega valor ao negócio das organizações e evita que recursos de TI sejam desperdiçados.

2.3. Arquitetura Baseada em Mediadores

O conceito de um componente que centraliza a busca por dados armazenados em bases distribuídas não é recente (Wiederhold, 1992; Gupta et al., 1999; Essid et al., 2004; Souza, 2008; Xavier 2008). Mediador é um componente que

- realiza os procedimentos necessários para manter as interfaces com as fontes de dados e com as aplicações funcionando;
- detém o conhecimento das estruturas que regem a transformação do dado em informação e
- realiza qualquer armazenamento intermediário que seja necessário.

A arquitetura mediada apresentada por Wiederhold (1992) possui três camadas: a de aplicações independentes presentes nas estações de trabalho dos usuários, a de mediadores e a de bases de dados. Os mediadores formam uma camada intermediária que torna as aplicações dos usuários independentes das bases de dados.

Mediadores devem implementar interfaces simples e orientadas à comunicação máquina-máquina. Além disto, uma linguagem geral e padronizada deve ser estabelecida entre os componentes da arquitetura, tornando desnecessário que uma aplicação ou um banco de dados específico possua um protocolo próprio de comunicação com o mediador.

2.4. Serviços de Processamento Estatístico Espacial e Apresentação

Bailey e Gatrell (1995) declaram que há três tipos básicos de análise de dados espaciais: os que se concentram em visualização, os exploratórios e os que são especificados a partir de modelos estatísticos e estimação de parâmetros. Neste trabalho, será mencionada uma das técnicas de análise exploratória, mais especificamente uma das fases da análise de padrões de pontos: o Estimador de Intensidade (*Kernel Estimation*).

Segundo Hara (1997), existem diversas formas de representar graficamente um dado geográfico que possui atributos: barras, áreas, pizzas, dispersão, mapas coropléticos etc. Aqui, foi implementado um serviço de fatiamento de mapas a partir de uma grade numérica.

2.5. Tecnologias envolvidas

Neste trabalho, foi utilizado o TerraOGC¹, extensão da biblioteca Terralib², como servidor e mediador de dados espaciais. Ele implementa *web services* nos padrões especificados pelo OGC³. Para a invocação das operações de kernel e fatiamento via *web services*, foi implementado um código em C++ que utiliza a biblioteca Rserve⁴, cuja tarefa é invocar operações do pacote estatístico R⁵. Este pacote disponibiliza diversas operações matemáticas e estatísticas, inclusive sobre dados espaciais.

3. Processando Dados Distribuídos

Neste trabalho, é apresentada a implementação de dois serviços: um estimador de densidade kernel e um fatiamento de mapas, representando, respectivamente, um serviço de processamento estatístico e um de apresentação de dados.

O estimador de densidade kernel é utilizado em distribuições de pontos de um espaço bidimensional que representam as ocorrências de um evento em estudo. Ele fornece uma visão geral da distribuição de primeira ordem desse evento. Trata-se de um indicador de fácil uso e interpretação (Drucks et al., 2004). O resultado deste processo é uma matriz de valores reais georeferenciada que apresenta valores mais elevados nas regiões onde há maior concentração do evento em questão.

O fatiamento consiste na subdivisão de uma grade numérica em faixas de valores (classes), a cada qual é, geralmente, atribuída uma legenda.

Os serviços foram implementados em C++, cujo código invoca funcionalidades do pacote estatístico R por meio da biblioteca Rserve.

Dois conjuntos de pontos foram produzidos e colocados em duas fontes de dados diferentes. O mediador do TerraOGC é responsável por reunir os pontos que provém de fontes diferentes. Um mediador de processos é encarregado de reunir entradas de processamentos que estão em origens diferentes. Em nosso caso específico, a entrada do processamento kernel é o conjunto de pontos adquiridos pelo mediador de dados do TerraOGC. O resultado desta operação é uma grade numérica georreferenciada, que pode ser representada por uma imagem geotiff. Na figura 3.1, ela está representada em tons de cinza.

Esta grade, resultado do primeiro serviço, é o parâmetro de entrada para o serviço de fatiamento, que dividirá a grade em faixas de valores e a apresentará ao requisitante com uma legenda (figura 3.2).

Genericamente, outros serviços de processamento e apresentação podem ser registrados na arquitetura e invocados pelo mediador de processos, de acordo com o desejo do gestor do servidor.

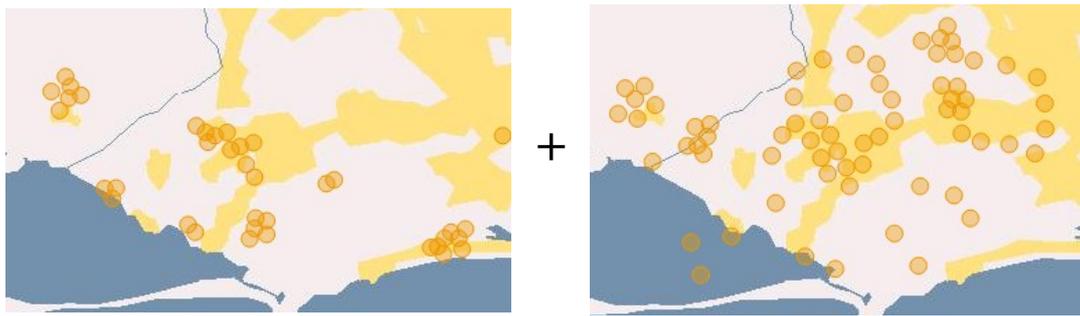
1 www.terralib.org/php/dow.php?body=Extensions

2 www.terralib.org

3 www.opengeospatial.org

4 <http://www.rforge.net/Rserve/>

5 www.cran.r-project.org/



↓
KERNEL

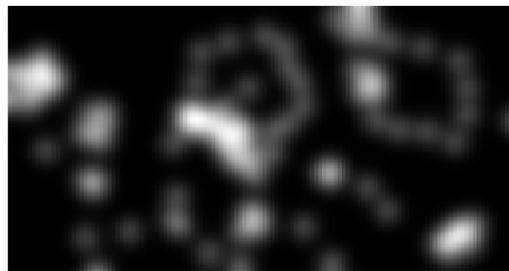


Figura 3.1. - Operação Kernel



↓
FATIAM.

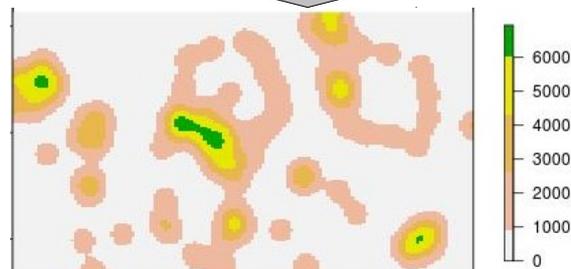


Figura 3.2. - Operação de Fatiamento

4. Conclusões

Este trabalho demonstra a implementação de uma arquitetura de mediação de dados e processos, e ressalta a utilidade dos *web services* como ferramenta de análise de dados espaciais. Para trabalhos futuros, ainda é necessária uma adequação dos serviços mostrados aos padrões da OGC, em especial, o WPS⁶ e o WCS⁷, o que é pretendido em trabalhos futuros.

Referências

- Bailey, T. C.; Gatrell, A. C. **Interactive Spatial Data Analysis**. Longman Scientific & Technical. 1995.
- Drucks, S.; Carvalho, M.S.; Câmara, G.; Monteiro, A.M.V. (eds). **Análise Espacial de Dados Geográficos**. Brasília, EMBRAPA, 2004 (ISBN: 85-7383-260-6).
- Essid, M.; Boucelma, O.; Colonna, F. M.; Lassoued, M. **Query processing in a geographic mediation system**. In: Annual ACM International Workshop on Geographic Information Systems, n. 12, 2004, Washington DC, USA. New York: ACM Press, 2004. p. 101-108.
- Gupta, A.; Marciano, R.; Zaslavsky, I.; Baru, C. **Integrating GIS and imagery through XML-based information mediation**. In: International Workshop on Integrated Spatial Databases: Digital Images and GIS, 1999, Portland, USA. Berlin: Springer, 1999. p. 211-234.
- Hara, L. T. **Técnicas de apresentação de dados em geoprocessamento**. 1997. Dissertação de Mestrado - INPE.
- Oliveira, P. F. A.; Oliveira, P. A.; Junior, C. A. D. **Arquitetura orientada a serviços (SOA): um estudo de caso em sistemas de informação geográficos**. Bol. Ciênc. Geod., sec. Artigos, Curitiba, v. 16, no 2, p.295-308, abr-jun, 2010.
- Papazoglou, M. P.; Heuvel, W. **Service oriented architectures: approaches, technologies and research issues**. The VLBD Journal, 2007, v. 16, p. 389-415.
- Souza, V. C. O. **Geoportal global para centros de imagens de sensoriamento remoto**. 2008. Dissertação de Mestrado – INPE.
- Wiederhold, G. **Mediators in the architecture of future information systems**. Computer, 1992, v. 25, n. 3, p. 38-49.
- Wiederhold, G.; Genesereth, M. **The conceptual basis for mediation services**. IEEE Expert, 1997, v. 12, n. 5, p. 38-47.
- Xavier, E. M. A.; **Serviços geográficos baseados em mediadores e padrões abertos para monitoramento ambiental participativo na Amazônia**. 2008. Dissertação de Mestrado – INPE.

6 www.opengeospatial.org/standards/wps

7 www.opengeospatial.org/standards/wcs