

Além do Acesso a Dados Distribuídos: Estendendo uma Arquitetura de Mediação Orientada a Serviços para Prover Serviços de Processamento Estatístico e Apresentação de Dados Espaciais

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. *The importance of sharing spatial data has been promoting the development of strategies that facilitate access and integration of distributed data. One of these strategies is the mediation architecture of distributed bases. It permits that data can be recovered from multiple sources with queries to a single access point. Accessing is this strategy's focus. But often treatment on data is necessary to extract information from it. This treatment can be provided at access moment, turning faster the knowledge production. This paper proposes to extend this architecture, including a component that mediates statistical processing and data preparation services for non-trivial presentations, always involving distributed spatial data.*

Resumo. *A importância de se compartilhar dados espaciais têm promovido o desenvolvimento de estratégias que facilitam acesso e integração de dados distribuídos em diferentes localidades. Uma dessas estratégias é a arquitetura de mediação de bases distribuídas. Ela permite que dados sejam recuperados de múltiplas fontes a partir de consultas a um único ponto de acesso. O foco dessa estratégia é o acesso. Contudo, quase sempre é necessário um tratamento sobre o dado para que dele seja extraída informação. Este tratamento pode ser fornecido já no acesso, agilizando a produção de conhecimento. Este trabalho propõe estender a arquitetura de mediação adicionando componentes que mediam serviços de processamento estatístico e preparação de dados para apresentações não convencionais, sempre envolvendo dados espaciais distribuídos.*

Palavras Chave

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

1. Introdução

Diversas áreas do conhecimento humano lidam com eventos que possuem um componente de localização. Dados espaciais, são parte dos processos de funcionamento

das organizações. Uma análise de dados espaciais quase sempre requer dois ou mais tipos de dados, geralmente produzidos em instituições diferentes. Tal fato leva à busca de estratégias que permitem integração dos dados em tempo de execução. Uma das abordagens, citadas por Casanova et al. [2005], é a estratégia de mediação [Wiederhold, 1992] [Gupta et al., 1999] [Essid et al., 2004] [Souza, 2008] [Xavier 2008], que é baseada numa arquitetura em três camadas: aplicações, fontes de dados com adaptadores e mediadores.

Após a fase de aquisição do dado espacial, processos podem ser necessários para extrair informações dele ou tornar sua apresentação diferente. Softwares livres e proprietários estão a disposição dos analistas, que podem processar o dado adquirido como desejarem. Uma alternativa mais interessante é adquirir, já no acesso, o dado com um nível de tratamento, agilizando o ciclo de produção do conhecimento.

Este trabalho parte da premissa de que um estágio de tratamento do dado espacial no momento do acesso agiliza o processo de análise de dados espaciais e apoiam a tomada decisão. Propõe-se estender a arquitetura de mediação orientada a serviços para que, além de acessar os dados, seja capaz de mediar serviços de processamento estatístico e apresentação de dados espaciais distribuídos.

2. Trabalhos Relacionados

O monitoramento ambiental é uma das maiores motivações do desenvolvimento de sistemas de disseminação de dados espaciais. Xavier [2008] apresenta um sistema de monitoramento ambiental participativo na Amazônia baseado em uma arquitetura mediada. A comunicação entre os componentes é feito por meio de serviços padronizados pelo *Open Geospatial Consortium* (OGC)¹. É, portanto, um sistema baseado em SOA (Arquitetura Orientada a Serviços).

Fora do escopo dos dados espaciais, Aldana et al. [2005] apresenta um sistema de mediação de dados biológicos. Este sistema possui serviços cujos resultados são gráficos e tabelas, e não o dado original. Ou seja, fornece serviços de apresentação de dados.

3. Incorporando novas funcionalidades a uma Arquitetura de Mediação Orientada a Serviços

A arquitetura distribuída e mediada proposta em outros trabalhos [Wiederhold, 1992] [Gupta et al., 1999] [Essid et al., 2004] [Souza, 2008] [Xavier 2008] promove o acesso a dados de diferentes fontes. Contudo, não é percebida a preocupação com o tratamento e apresentação dos dados.

Propõe-se neste trabalho incorporar dois novos componentes à arquitetura: um mediador de serviços de processamento estatístico e um mediador de serviços de apresentação em adição ao mediador de acesso aos dados. Agindo em conjunto, os mediadores terão condições de recuperar dados dos servidores de acesso e fornecê-los como entrada aos servidores de processamento e apresentação. A figura 3.1 ilustra um protótipo da SOA distribuída e mediada com os componentes de processamento e apresentação pretendidos.

Nada impede que os usuários requisitem apenas um serviço de processamento ou apresentação, fornecendo como entrada dados de sua propriedade e abrindo mão de uma consulta às bases de dados distribuídas por meio do mediador de acesso.

¹ <http://www.opengeospatial.org>

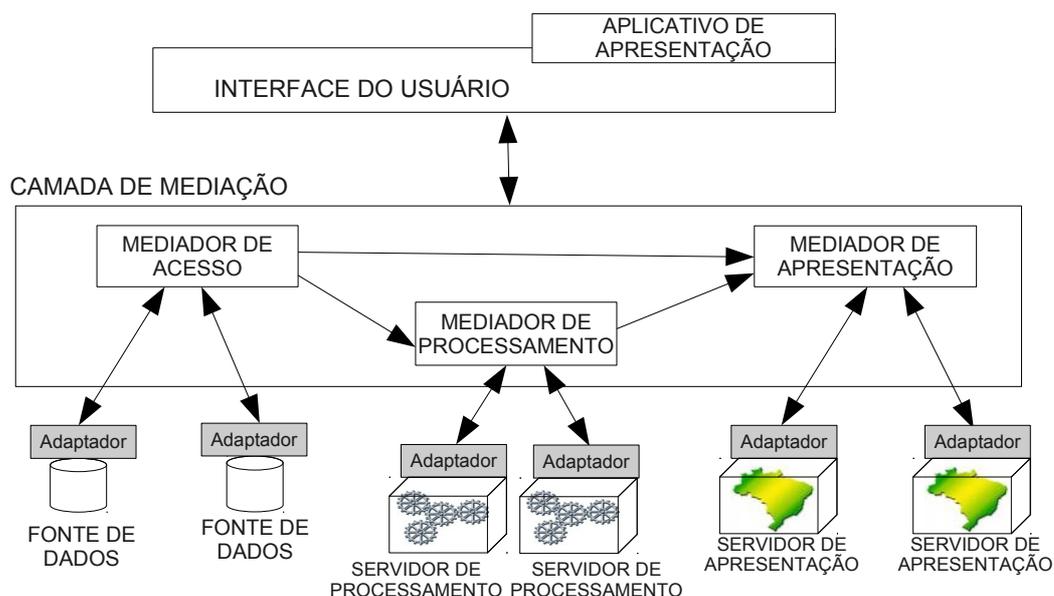


Figura 3.1. - Arquitetura proposta

4. Uma prova de conceito: Serviços de Processamento Estatístico Espacial e Apresentação

O projeto *Observatorium* [INPE, 2011], iniciativa do Instituto Nacional de Pesquisas Espaciais (INPE) e da Fundação Oswaldo Cruz (Fiocruz), foi concebido com objetivo de analisar as interações existentes entre clima, saúde e meio ambiente. É um exemplo de sistema que exige a manutenção de uma arquitetura distribuída e mediada. O projeto é uma extensão da arquitetura proposta por Xavier[2008] e, atualmente, é construída sobre o TerraOGC², extensão da biblioteca Terralib³ que implementa servidores e clientes OGC sobre um componente mediador e sobre bancos Terralib. O *Observatorium* será a plataforma base para demonstrar o funcionamento dos componentes propostos neste trabalho.

Como prova de conceito para o mediador de processamentos, dois serviços de análise espacial serão implementados: o cálculo do Índice Local e Global de *Moran*, para análise de autocorrelação espacial de um conjunto de regiões, e o cálculo da superfície de saída de um Estimador de Intensidade (*Kernel Estimation*) para uma distribuição de pontos. A interface em linguagem C do R [R Development Core Team, 2010] será experimentada como ferramenta de apoio à implementação desses serviços, visto que o R é um amplo pacote de ferramentas matemáticas e estatísticas e possui, também, funções de análise de dados espaciais.

O adaptador dos servidores de processamento será implementado a partir da especificação WPS (*Web Processing Service*) do OGC, um padrão de interoperabilidade para requisição de processamentos via rede.

Como prova de conceito para o mediador de serviços de apresentações será testada uma solução baseada no aplicativo *Gapminder*⁴, que possui uma forma de

2 <http://www.terralib.org/php/dow.php?body=Extensions>

3 <http://www.terralib.org/php/about.php?body=AboutTL>

4 <http://www.gapminder.org>

exibição não convencional de dados. A forma com que o *Gapminder* processa e apresenta os dados será estudada e inserida na arquitetura. Apresentações em formatos usuais, como SVG (*Scalable Vector Graphics*) e os utilizados no padrão WMS (*Web Map Server*) do OGC poderão ser acoplados à arquitetura sem esforço.

5. Conclusões

A arquitetura proposta está em fase de estudo e implementação. Possíveis dificuldades podem surgir quanto à performance do processamento de dados de grande dimensão. Este problema ainda é comum a esse tipo de estrutura de *software*. Outro ponto de dificuldade são as especificações do WPS, que ainda são novas, com pouca informação difundida e poucas implementações no mercado e no meio acadêmico.

Espera-se que, ao término deste trabalho, uma evolução da atual arquitetura do *Observatorium* seja alcançada, visando aumentar o nível da informação adquirida dos servidores por meio do portal de acesso .

Referências

- Aldana, J. F.; Hidalgo-Conde, M.; Navas, I.; Roldán, M. M.; Trelles, O. **Bio-broker: a biological data and services mediator system**. IADIS International Conference on Applied Computing, 2005, p. 527-534.
- Casanova, M. A.; Brauner, D. F.; Camara G.; Júnior, P. O. L. **Integração e interoperabilidade entre fontes de dados geográficos**. INPE, 2005. Disponível em <<http://www.dpi.inpe.br/livros/bdados/capitulos.html>>. Acesso em 25 fev 2011.
- 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.
- INPE. **Observatorium: Observatório Nacional de Clima e Saúde**. INPE, 2011. Disponível em: <<http://www.dpi.inpe.br/observatorium/>>. Acesso em: 01 ago 2011.
- R Development Core Team. **Writing R Extensions – Version 2.12.0**. R Project, 2010. Disponível em: <<http://cran.r-project.org/manuals.html>> Acesso em: 10 dez 2010.
- 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.