

Danilo da Silveira Figueira

# **Arquitetura de alto desempenho e escalabilidade para álgebra de mapas**

Brasil

30 maio de 2017

Danilo da Silveira Figueira

# **Arquitetura de alto desempenho e escalabilidade para álgebra de mapas**

Universidade Federal de Minas Gerais

ICEX

Ciência da Computação

Brasil

30 maio de 2017

# Resumo

**Palavras-chaves:**

## Lista de ilustrações

## Lista de tabelas

# Sumário

<b>1</b>	<b>Introdução</b>	<b>6</b>
<b>2</b>	<b>Trabalhos relacionados</b>	<b>9</b>
<b>3</b>	<b>Metodologia</b>	<b>11</b>
<b>4</b>	<b>Cronograma</b>	<b>13</b>
	<b>Referências</b>	<b>15</b>

# 1 Introdução

Essa pesquisa nasceu da necessidade de oferecer os resultados das análises feitas sobre a floresta amazônia em um projeto com o banco mundial. Nesse projeto avaliou-se muitas variáveis, por exemplo, a produção de borracha, a biodiversidade, o risco de incêndio, entre outros. Para o levantamento desses dados, toda a amazônia foi dividida em pequenas regiões, cada uma com área de um  $\text{km}^2$ , para que cada uma dessas regiões fossem analisadas independentemente. A amazônia por sua vez possui 5217423  $\text{km}^2$  e por sua vez a análise terá essa mesma quantidade de regiões. Então, o problema é que apesar dessa enorme quantidade de dados gerados, foi necessário disponibilizar esses mapas via internet para consultas, além de somente a visualização dos dados. Essas consultas permitiram combinar os mapas resultantes das pesquisas já realizadas em um novo mapa usando a lógica definida pelo usuário, esse resultado deveria ser gerado quase instantaneamente para permitir a visualização interativa da consulta. Ainda assim, o servidor deveria suportar múltiplos usuários sem diminuir a qualidade do serviço. Então, uma forma de aplicar os cálculos seria usar o processamento do cliente, para isso era preciso enviar os dados de maneira mais simplificada possível para permitir o processamento rápido daqueles dados pelo cliente.

Para começar a resolver esse problema optou-se pela utilização de dados matriciais no formato raster, com a aplicação de operações segundo um conjunto de regras algébricas. A álgebra de mapas, é um conjunto de operações para a manipulação de dados geográficos no formato raster proposto por [Tomlin \(1990\)](#). De maneira simplificada, rasters podem ser entendidos como mapas georreferenciados, estruturados em formato matricial, onde cada célula representa o valor de uma área do mapa. Essas operações são divididas em quatro grupos: Locais, Focais, Zonais e Globais. Cada um desses grupos avaliam diferentes quantidades de células, no entanto, o resultado é sempre um valor por célula. A solução proposta, foi usar o dado visível enviado para visualização do mapa, já num formato que pudesse servir para a aplicação das operações desejadas pelo usuário. Por exemplo as operações Locais, que usam somente as células da mesma região, podem facilmente ser aplicadas no mapa visível sem erro, uma vez que o dado necessário para a conta é o mesmo que o dado para a visualização. Porém, para permitir a aplicação dos outros três grupos de operações, é preciso o uso de dados além dos dados de visualização. Por exemplo, na pesquisa sobre a amazônia, uma das consultas, foi permitir calcular a produção total de borracha por município na amazônia. Nesse caso, somar as células visíveis iria gerar um resultado com erro, então foi necessário gerar uma tabela no servidor dizendo qual era o resultado de cada município, e então ao executar a consulta, essa tabela era consultada e o resultado apresentado interativamente. Portanto deseja-se definir um subconjunto de

operações da álgebra de mapas, que sejam passíveis de se executar de maneira rápida pelo cliente, que os resultados sejam reutilizáveis em outras expressões, assim, permitir combinar os mapas no cliente segundo uma expressão definida por ele e que gere o mapa dinamicamente.

OGC (Open Geospatial Consortium), uma organização internacional, definiu vários padrões para a disponibilização de mapas pela internet [OGC \(2017\)](#). No ano de 1999 publicou a primeira versão de um dos padrões de maior sucesso, o WMS (Web Map Service), que permite a visualização de mapas georreferenciados pela internet de maneira rápida. Esse serviço funciona de maneira simples, oferecendo uma versão re-amostrada do dado para cada escala, assim para uma determinada escala, todos os mapas são exibidos com tamanho de célula iguais, independente da resolução original, assim o tamanho do mapa fica limitado à porção visível na escala atual. Esse serviço apesar de simples e muito eficiente é bastante limitado pois permite somente a visualização do dado. Porém, devido aos vários avanços da tecnologia e o aumento da utilização de serviços Web, é cada vez mais imprescindível que além da visualização, sejam criados serviços que permitam o processamento desses dados. Nesse sentido a OGC definiu dois padrões para cálculos de dados no servidor: o WPS (Web Processing Service) e o WCPS (Web Coverage Processing Service), que permitem processar os dados no servidor disponibilizando os resultados em formato bruto. Foram definidos também os padrões Web Feature Service (WFS) e Web Coverage Service (WCS), que permitem transmitir via internet os dados brutos no formato vetorial e rasterizado, respectivamente, permitindo assim ao cliente usar os serviços Web para analisar os dados no servidor ou em seu próprio computador. Esses padrões por sua vez foram implementados em várias ferramentas de SIG (Sistemas de Informação Geográfica) tais como ARCGIS, QGIS, dentre outras. Para finalidades em casos gerais esses padrões se mostraram suficientes. Porém, se mostram insuficientes para a construção de um serviço de cálculo de mapas pela internet, principalmente por dois motivos: a superutilização do poder de processamento do servidor e a necessidade do envio de grandes quantidades de informação para a visualização do resultado. Ambos os problemas são críticos para serviços Web com dados geográficos, pois os dados comumente possuem vários gigabytes de tamanho, o que dificulta a transferência para cada cliente, e ainda o processamento desses dados pode levar horas ou até mesmo dias.

Por isso, pretende-se discutir a criação de um padrão que permita a análise dos mapas utilizando um conjunto de operações ao se combinar o processamento em partes no servidor e no cliente. Pretende-se em nosso trabalho mudar a abordagem do problema em dois pontos. O primeiro é criando uma variante do WMS para disponibilizar não mais os dados já com as legendas aplicadas, mas uma representação crua do dado adequada para a escala visível. Já o segundo ponto seria combinar o WPS com o WMS alterado, ao se permitir pré-processar os dados “brutos” no servidor através de operadores pré definidos, para calcular algumas operações, e utilizar esses resultados no cliente ao se combinar os



resultados usando a álgebra de mapas.

O objetivo desse trabalho é permitir a aplicação de cálculos interativamente utilizando uma arquitetura cliente/servidor, pretende-se definir um framework de consulta aos dados geográficos. O cliente acessará os dados através de uma API, essa requisição será executada em dois passos, uma no servidor e outra no cliente. Ao receber a requisição, o servidor analisa quais os dados necessários, e gera esses dados em um formato conhecido pelo cliente. Em seguida o cliente recebe os dados pré-processados e os combina de acordo com as operações desejadas, permitindo uma execução muito mais rápida. Assim a carga de processamento fica dividida entre o cliente o servidor, e permite também a possibilidade do uso de cache pelo servidor para que um único resultado possa ser reutilizado múltiplas vezes.

## 2 Trabalhos relacionados

A OGC define a maioria dos padrões que iremos utilizar para sustentar a nossa proposta de arquitetura, pois, é importante manter compatibilidade ao máximo com os padrões já estabelecidos. Dentre os serviços de disponibilização de mapas estão os: WMS, WFS, WCS, WPS e WCPS (OGC, 2017). Por sua vez os formatos para a transmissão dos dados também possuem seus próprios padrões, alguns permitem a transmissão de dados brutos: GML, KML, SVG (W3C, 2017), GeoJSON (BUTLER et al., 2016), já outros entregam os mapas renderizados, como por exemplo: JPEG, PNG, GIF, dentre outros. Temos alguns estudos que questionam os padrões OGC tais como Peng e Zhang (2004) e Vinhas et al. (2016). No primeiro estudo são apontados vários pontos positivos do WFS porém é exemplificada a ineficiência no transporte dos dados nesse padrão. Já no segundo trabalho é apontado que nenhum dos padrões permite a análise temporal de maneira simples e eficiente já que a arquitetura necessita ir de mapa por mapa para pegar o pixel na mesma posição e para isso é proposto um padrão chamado WTSS (Web Time Series Service) que propõe a inspeção 3D dos mapas ao invés da visão 2D atual. Além disso existem estudos da aplicação desses padrões em softwares desktop GIS como em Vatsavai et al. (2006) que discute as peculiaridades dos serviços de disponibilização dos mapas WMS, WFS e WCS e propõe uma arquitetura de software onde automaticamente esses serviços seriam trocados de acordo com o uso do cliente. Nesse sentido o usuário poderia usar o WMS para visualizar o mapa, e ao executar uma consulta o software usaria o WFS ou o WCS automaticamente, fazendo que os dados brutos só sejam enviados em caso de necessidade. Por tanto, o simples fato de se combinar esses padrões de maneira iterativa permite uma maior eficiência na visualização dos dados e menor uso desnecessário do servidor. Alguns trabalhos publicados tratam da arquitetura de serviços visando o processamento e análise de mapas via internet como em Vatsavai et al. (2006) e em (SAMPLE et al., 2008). No primeiro é proposta uma arquitetura balanceada de carga entre servidor e cliente que se mostrou mais adequada por permitir maior escalabilidade e disponibilidade do servidor. Essa proposta arquitetural por sua vez diferencia da grande maioria dos serviços de Web-GIS que utilizam estruturas “cliente leve e servidor pesado” ou “servidor leve e cliente pesado”. Já no segundo, são discutidas decisões arquiteturais visando a escalabilidade do servidor através da utilização dos serviços WFS e WCS. Assim, o processamento dos dados ficariam por conta dos clientes, o que melhorou o problema da escalabilidade do servidor mas que devido à necessidade do envio de grandes quantidade de dados se mostrou uma solução insuficiente. Então a solução proposta foi a criação de um WPS distribuído, que permita o processamento de grandes quantidade de dados sem a necessidade de transferência dos arquivos e que resolva os problemas da escalabilidade pois

permite que o crescimento da infra estrutura do servidor possa acompanhar a quantidade consultas múltiplas já que seria necessário apenas o aumento do número de servidores.

## 3 Metodologia

O trabalho pode ser dividido em cinco partes: estudo do problema, escolha das operações, definição dos operadores, definição da arquitetura e validação da arquitetura.

Na primeira parte, estudo do problema, iremos fazer o levantamento das características dos serviços da OGC visando um entendimento mais profundo de cada serviço de interesse, além disso iremos buscar outras propostas não oficiais de serviços de dados geográficos. Esse levantamento deverá servir como base para saber as peculiaridades, pontos positivos e negativos, de cada padrão para podermos aproveitar as ideias e melhorar onde for possível. Ainda nessa etapa pretende-se também buscar mais detalhes sobre as peculiaridades sobre a implementação dos padrões em cada ferramenta para melhor compreender as dificuldades e problemas técnicos em cada abordagem.

A segunda parte, escolha das operações, pretende-se estudar os problemas mais comuns e as soluções mais populares para cada problema, para assim poder definir quais operações deveriam ser implementadas para atender uma maior quantidade de usuários. Além disso pretende-se também entender as peculiaridades de cada operação para permitir criar os operadores de maneira mais geral e assim criar uma API mais abrangente possível e com o menor conjunto de operadores possível para que cada operação seja reutilizada no máximo de expressões possível. Dessa maneira pretende-se minimizar a carga no servidor e ainda assim permitindo combinar os resultados em mapas completamente diferentes com operações pixel a pixel executadas no cliente.

Na etapa de definição dos operadores, iremos avaliar quais operações são viáveis de serem implementadas, e também analisar como minimizar a troca de dados para cada operação. Para isso iremos avaliar os dados de entrada e saída esperados para cada operação e tentar encontrar para cada operação a melhor forma de organização do resultado. Pretende-se saber qual dado deverá ser enviado, e em qual formato por exemplo em alguns casos espera-se que seja mais conveniente retornar o resultado através de mais de um mapa ou em outros casos até mesmo no formato de tabela.

Na quarta etapa será a definição da arquitetura, onde pretende-se definir a forma da troca de informações entre o servidor e cliente especificando o protocolo de comunicação. Espera-se também definir a melhor forma de organizar e interpretar as requisições dos clientes, com isso deseja-se garantir a criação do cache das operações e permitir a escalabilidade do sistema. Outro passo importante será definir a forma que o cliente deverá interpretar os dados de resultados para execução da expressão algébrica.

Já na ultima etapa pretende-se avaliar os resultados obtidos, as métricas avaliadas serão: a velocidade de execução das consultas, a quantidade de dados transmitidos

em relação aos dados originais, o consumo de poder computacional e o consumo de armazenagem em disco. Os testes serão executados executando as consultas em diferentes arquiteturas. Com base nesses resultados pretende-se tomar as decisões finais, fazendo as ultimas melhorias e assim entregar a definição arquitetural proposta.

## 4 Cronograma

O plano de atividades proposto é o seguinte:

1. Estudo do problema
  - a) Levantamento bibliográfico da OGC;
  - b) Levantamento bibliográfico não oficial;
  - c) Identificação dos pontos positivos e negativos de cada abordagem;
  - d) Identificação de dificuldades técnicas de implementação;
2. Escolha das operações
  - a) Identificação dos problemas e soluções populares;
  - b) Identificação de caracteres gerais de cada operação;
3. Definição dos operadores
  - a) Avaliar viabilidade de implementação de operador;
  - b) Listagem dos dados necessários a cada operador;
  - c) Definir-se formato apropriado para cada operador;
4. Definição da arquitetura
  - a) Definição do protocolo cliente-servidor;
  - b) Organização do cache dos resultados;
5. Validação da arquitetura
  - a) Testes
  - b) Avaliação dos resultados;
  - c) Validação;
6. Defesa da dissertação

Atividades	05/17	07/17	09/17	11/17	12/17	02/18	04/18	06/18
1.a								
1.b								
1.c								
1.d								
2.a								
2.b								
3.a								
3.b								
3.c								
4.a								
4.b								
5.a								
5.b								
5.c								
6								

# Referências

- BUTLER, H. et al. *The geojson format*. 2016. Disponível em: <<http://www.rfc-editor.org/info/rfc7946>>.
- MENNIS, J.; VIGER, R.; TOMLIN, C. D. Cubic map algebra functions for spatio-temporal analysis. *Cartography and Geographic Information Science*, Taylor & Francis, v. 32, n. 1, p. 17–32, 2005.
- OGC. *Implementation Specifications*. 2017. Disponível em: <<http://www.opengeospatial.org/standards/specifications/list>>.
- PENG, Z.-R.; ZHANG, C. The roles of geography markup language (gml), scalable vector graphics (svg), and web feature service (wfs) specifications in the development of internet geographic information systems (gis). *Journal of Geographical Systems*, Springer, v. 6, n. 2, p. 95–116, 2004.
- SAMPLE, J. T. et al. *Geospatial services and applications for the internet*. [S.l.]: Springer Science & Business Media, 2008.
- TOMLIN, C. D. *Geographic information systems and cartographic modelling*. [S.l.]: New Jersey, US: Prentice-Hall, 1990.
- VATSAVAI, R. R. et al. Umn-mapserver: A high-performance, interoperable, and open source web mapping and geo-spatial analysis system. *Lecture Notes in Computer Science*, Springer, v. 4197, n. 2006, p. 400–417, 2006.
- VINHAS, L. et al. Web services for big earth observation data. In: *GeoInfo*. [S.l.: s.n.], 2016. p. 166–177.
- W3C. *Scalable Vector Graphics (SVG) 1.1 (Second Edition)*. 2017. Disponível em: <<https://www.w3.org/TR/SVG/>>.