



UERJ – Universidade do Estado do Rio de Janeiro
Programa de Pós-Graduação em Eng. de Computação
Área de Concentração: Geomática

Designing Service Architectures for Distributed Geoprocessing: Challenges and Future Directions

Disciplina: SigWeb Avançado

Prof^o.: João Araújo

Aluno: Wesley Silva Fernandes

Apresentação

- O artigo apresenta uma proposta para utilizar uma Infraestrutura de Dados Espaciais (IDE) como base para distribuir serviços de Geoprocessamento;
- Como caso de uso será desenvolvida uma aplicação que calcula as estatística de incêndio em uma determinada área, onde foi demonstrado que a IDE facilitou o desenvolvimento rápido da aplicação;
- Serão utilizados serviços de WMS, WFS, Gazetter, Serviço de catálogo e serviços de Geoprocessamento;

Introdução

- A nova geração de Plataforma de Computação Distribuída chama-se computação orientada a serviços, pertencente ao modelo de arquitetura SOA (Arquitetura Orientada a Serviço;
- Arquitetura é um paradigma de organização e utilização de capacidades distribuídas;

Introdução

- Assim um SOA específico fornece estrutura e regras para descrição e descoberta de serviços, interação entre fornecedores e consumidores e o respectivo ambiente de execução;
- Atualmente o exemplo mais proeminente de SOA são os *webservices* com seus protocolos e mecanismos para sua descrição, descoberta e invocação;

Arquitetura Orientada a Serviços (SOA)

- A SOA é aberta e interoperável;
- Ao contrário dos sistemas tradicionais em uma SOA são distribuídos serviços, que os usuários podem utilizar conforme a necessidade;
- Evita a inconsistência de dados, possuindo um repositório centralizado;
- Serviços oferecem algoritmos que podem ser reutilizados por várias aplicações diferentes, evitando redundância de implementação;
- Assim, pode-se desenvolver sistemas mais flexíveis, adaptáveis as mudanças de tecnologia e necessidades dos usuários, e mais fáceis de manter que sistemas *standalone*;
- Desvantagens: novos paradigmas necessitam de novas competências e capacidades e necessidade de um consentimento sobre uma arquitetura e metodologia de desenvolvimento sobre um portfólio de serviços

Informações Geográficas (IG)

- Consolidação de especificações de interoperabilidade, sobretudo da OGC, que seguem a idéia geral da SOA;
- Migração das aplicações *standalone* para aplicações distribuídas, especializadas e interoperável sobre serviços de IG;
- Apoio Governamental para criação de IDE's nacionais como INSPIRE na Europa e Geoconnectios no Canadá;

Objetivo

- O objetivo da pesquisa é conhecer até que ponto as tecnologias (especificações, padrões e produtos) disponíveis para SOA atualmente, são adequadas para desenvolver aplicações GIS, principalmente distribuindo serviços de geoprocessamento;

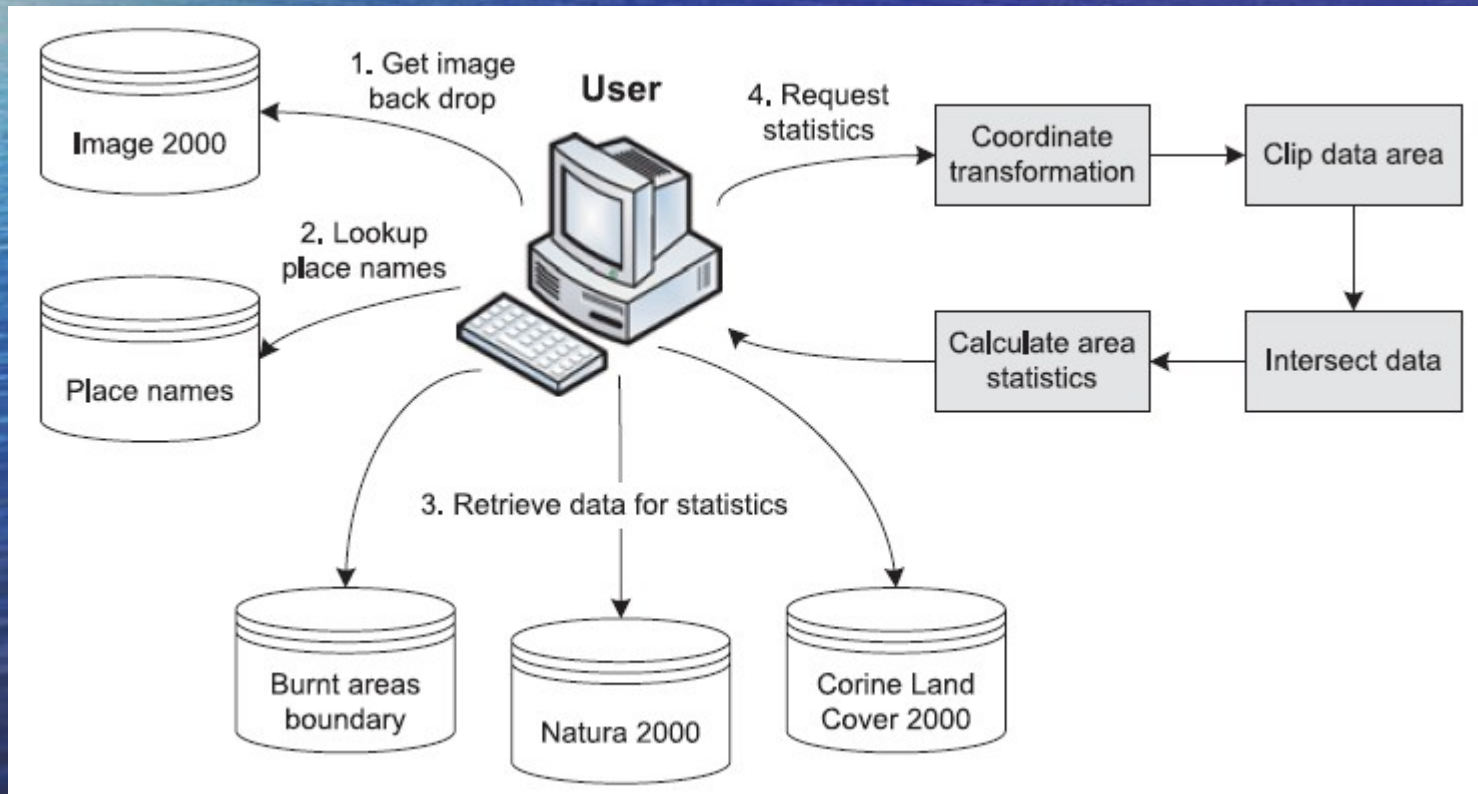
Objetivo

- Uma aplicação de cálculos estatísticos de incedios florestais foi desenvolvida para demonstrar o desafio que existe na distribuição de serviços de Geoprocessamento;
- A aplicação deve ser desenvolvida em um ambiente distribuído e interoperável;
- O objetivo principal da aplicação é exemplificar escalabilidade, flexibilidade e reusabilidade dos componentes de uma IDE;
- Serão discutidas alternativas para contornar os problemas encontrados;

A Aplicação Desenvolvida

- Propósito: calcular as estatísticas de incêndio florestal de uma determinadas área baseado em dados de cobertura do solo;

Fluxo de Trabalho



A Aplicação Desenvolvida

- São utilizados dois conjuntos de dados para auxiliar ao usuário;
 - Imagens de Satélite;
 - Nomes Geográficos;

A Aplicação Desenvolvida

- Para o cálculo das estatísticas são utilizados os seguintes conjuntos de dados:
 - Natura 2000, para descrever áreas especiais de conservação, utilizados para avaliar o crescimento de áreas protegidas atingidas pelo fogo;
 - Corine Land Cover 2000, para identificar quais classes de cobertura do solo são afetadas, quando se está avaliando o prejuízo de incendio florestal;

A Aplicação Desenvolvida

- No sistema o usuário identifica a área de interesse com o auxílio da imagem de satélite e de nomes geográficos, e os submete ao sistema que calcula as estatísticas, seguindo uma seqüência de operações. As operações são, nesta ordem:
 - Transformação de coordenadas quando necessário;
 - Seleção dos dados, através de um recorte com auxílio da área de interesse;
 - O conjunto de dados temáticos selecionado pelo usuário é intersectado por áreas de queimadas e as estatísticas são calculadas;
 - A estatística pode ser calculada de acordo com o conjunto de dados temáticos, ou por classes de uma dada especificação;

Arquitetura

- Além de acesso aos dados o sistema disponibiliza um serviço de catálogo, para o usuário consultar os dados disponíveis para a região de interesse;
- Atualmente estão disponíveis dois conjuntos de dados, Corine e Natura;
- O catálogo é utilizado devido as origens dos dados serem desconhecidas pelo usuário, facilitando entre outras coisas, a sua descoberta;
- Os passos necessários para o cálculo das estatísticas foram reunidos em um serviço de geoprocessamento, por questão de praticidade e performance, porém isto torna a aplicação altamente específica impossibilitando a reutilização do serviço em outras aplicações, o que será discutido mais tarde;
- Finalmente, há clientes que disponibilizam a interface e executam o fluxo de trabalho da aplicação;

Seqüência de operações

- Cliente solicita imagem, fornecida por um serviço WMS;
- Usuário seleciona área e tempo de interesse para receber dados apropriados;
- Área de interesse é visualizada em imagem de satélite;
- Depois, o catálogo é utilizado para pesquisar e selecionar o conjunto de dados que serão utilizados como origem (ou mascara) e alvo;
- Neste cenário a mascara é uma camada específica de áreas queimadas e o alvo poderia ser dados de cobertura, como Corine ou Natura;
- Após isso, o serviço de estatística é solcitado;
- O serviço recebe os dados, transforma as coordenadas, quando necessário, calcula as estatisiticas e retorna o resultado ao final, em forma de tabela;

Serviço de Estatísticas

- Foi implementado segundo a especificação WPS (Web Processing Service) da OGC;
- Foi implementado utilizando Java 1.5 e a API Geotools 2.1

Especificação WPS

- Possui 3 operações principais:
 - `getCapabilities`, comum a todos serviços OGC, retorna os metadados do WPS;
 - `describeProcess`, que descreve o processamento suportado pelo WPS em questão. Não foi implementado no serviço de estatísticas;
 - `execute`.

Especificação WPS

- Suporta solicitações GET e POST;
- GET é uma solicitação HTML contendo todos os parâmetros necessários;
- Em POST um documento XML é enviado, contendo os parâmetros em tags XML
- A operação execute aceita os dois;
- O serviço de Estatística utilizou a solicitação GET;

O serviço de Estatísticas

- Parametros:
- Service:WPS;
- Processo: há somente um, AreaStatistics;
- Mask: uma URL para os dados de máscara;
- Masktypename: tipo de feição;
- Os mesmo parametros são fornecidos para os dados alvo;
- Bbox: área limite sobre a qual serão realizados os cálculos;
- Attribute: especifica o atributo que será utilizado, caso presente o cálculo estatístico será realizado por classes temáticas;
- Totalarea diz se o total da área queimada deve ser dada, o padrão é *true*

Exemplo de solicitação WPS

```
http://naturegis.h07.jrc.it:8090/StatisticsService/Process?  
SERVICE=WPS&  
REQUEST=Execute&  
VERSION=0.3.0&  
PROCESSNAME=AreaStatistics&  
STORE=false&  
MASK=http://naturegis.h07.jrc.it:8090/geoserver/wfs/&  
MASKTYPENAME=INSPIRE:ba2003&  
TARGET=http://naturegis.h07.jrc.it:8090/geoserver/wfs/&  
TARGETTYPENAME=INSPIRE:clc&  
ATTRIBUTE=CLC_00&  
BBOX=-9,39.33,-8.56,39.7
```

O serviço de Estatísticas

- Em uma solicitação ao serviço de estatística, o mesmo chama uma operação execute que utiliza os parâmetros da solicitação;
- Para receber as feições da área de interesse, um getFeature é lançado a um serviço WFS;
- Enfim, as estatísticas são calculadas e retornadas em um documento XML;

Serviços de Mapeamento e Feições

- Foram utilizados as especificações OGC WMS 1.3 e WFS 1.0;
- Para disponibilizar o serviço WFS foi utilizado o Geoserver 1.3.0;
- Para servir o WMS foi utilizado o ArcGis 9.1;
- Para o Gazetter foi utilizado o Ionic Software RedSpider Studio

Catálogo

- Para o serviço de catálogo foi utilizado o *Software* terraCatalog da con terra;
- Suporta os perfis ISO 19115/19119 e utiliza os serviços de catalogo CSW 2.0;
- Para a pesquisa o cliente recebe como parâmetro o título, área abrangente e ano;
- O cliente suporta dois protocolos diferentes, Z39.50 e CSW, sendo utilizado pela aplicação o CSWCatalogClient desenvolvido a partir do CSW;
- São retornados o título e um serviço URL, e os dados de preferência podem ser selecionados no cliente da aplicação, sendo a URL passada como parâmetro para o serviço de estatísticas;

Cliente da aplicação

- O cliente foi desenvolvido utilizando DHTML, diminuindo o tempo de resposta utilizando a técnica AJAX;
- O cliente possui funções de navegação como ZOOM e PAN;
- Depois de selecionar um ano e palavras chaves para pesquisar áreas queimadas e dados alvo um relatório de estatística é gerado

Cliente da aplicação

http://naturegis.h07.jrc.it:8080 - Forest Fire Damage Area Assessment - Microsoft Internet Explorer

Forest Fire Damage Area Assessment

Year: 2003

Burnt area keyword:

FIND

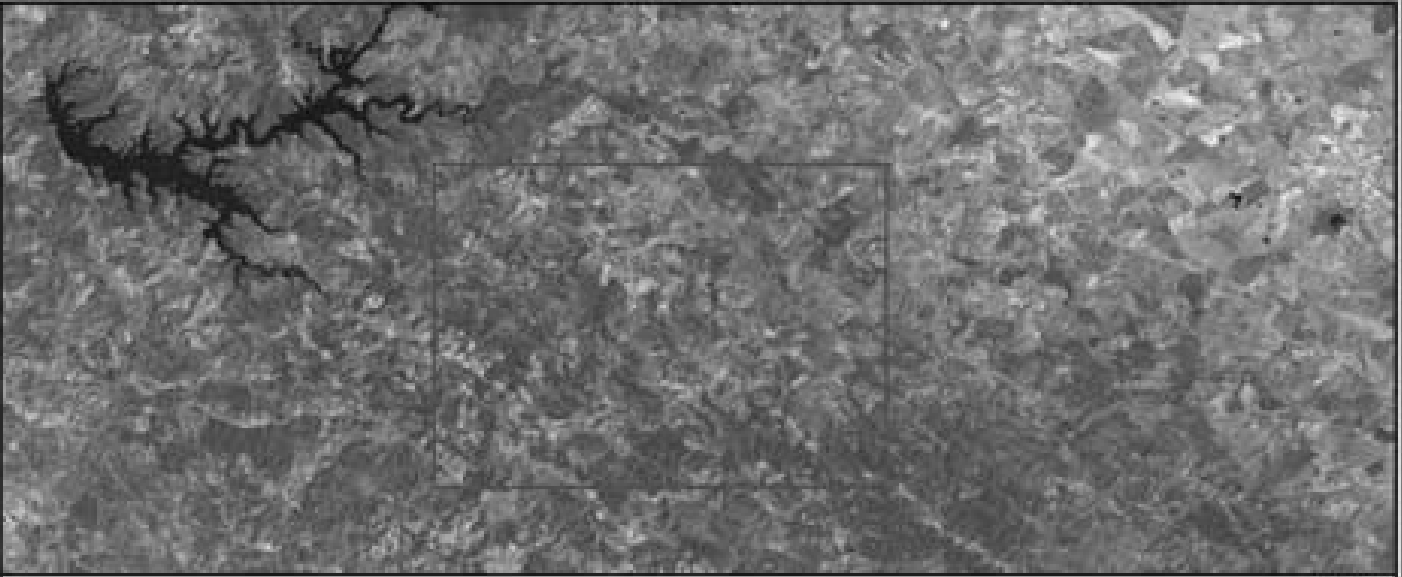
Service Settings

QUICK MAP SEARCH

Find a Place: (city, country, region, administrative boundary)

GNSE

FIND



-8.28337763 - 37.5548015

Total area affected by fire: 13.75km²
Intersected area for layer 'INSPIRE:clc': 13.75km² (100.0%)

class	area	pct %
Non-irrigated arable land	0.0	0.0
Agro-forestry areas	2.69	19.58
Sclerophyllous vegetation	0.94	6.85
Broad-leaved forest	9.37	68.16
Land principally occupied by agriculture, with scattered areas of natural vegetation (non-forestry) areas	0.33	2.39

Done Internet

Questões da Arquitetura Apresentada

- O exemplo mostra que é possível utilizar os serviços OGC para distribuir serviços de geoprocessamento em uma IDE;
- A padronizações torna fácil a combinação de diferentes serviços para visualização, distribuição e processamento de dados espaciais;
- Torna flexível o uso do geoprocessamento em IDEs;
- Entretanto surgem questões fundamentais de performance em função de limitações desta técnica e do projeto arquitetural limitando a usabilidade e escalabilidade da arquitetura presente;

Questões da Arquitetura Apresentada

- Principal Problema: o transporte dos dados na cadeia de serviço, em função dos mesmos estarem distribuídos na IDE;
- Entende-se por cadeia de serviço (ISO) uma sequência de processamento, sendo necessário o fim de um processamento para o início do subsequente;
- Os dados transportados normalmente são GML e dados binários de imagens (GeoTiff)
- Um GML da base Corine abrangendo Portugal e Espanha tem aproximadamente 1 Gb, ou 1/3 disto quando comprimido;
- Um Geotiff da mesma área tem aproximadamente 350 Mb, no caso de uma imagem de 100 m de resolução espacial e 8 bits de resolução radiológica;
- O transporte de dados desta magnitude pode facilmente causar problemas;

Questões da Arquitetura Apresentada

- Outro problema foi a combinação de muitos serviços em uma operação, com a intenção de melhorar a performance, pois é necessário apenas uma solicitação para atender o requisito, porém isto torna muito específica esta operação, impossibilitando sua reutilização;
- Uma alternativa seria criar uma operação para cada processamento necessário;

Questões que precisam ser investigadas

- Operações de geoprocessamento que demandam muito tempo devem utilizar chamadas assíncronas? Que passos adicionais estariam faltando?
- Como torna as operações de geoprocessamento mais flexíveis e eficientes?

Abordagens Alternativas

- Usuários de internet esperam respostas rápidas, ao contrário do que acontece em Gis Desktop, onde os usuários estão acostumados a longos períodos de processamento;
- Solução: Mensagens Assíncronas, em solicitações que demandam muito tempo, assim o usuário não precisa esperar o fim do processamento, sendo o sistema responsável pelo monitoramento do processo;

Abordagens Alternativas

- Para isso será necessário que cada chamada ao serviço tenha um identificador único, permitindo sua recuperação mais tarde;
- A solicitação estar habilitada a monitorar o status do processo, recuperando o resultado ao final;
- O usuário deverá ter a alternativa de abortar o processo;

Abordagens Alternativas

- Existe especificação OGC que trata desta abordagem, a WNS (Web Notification Service);
- A especificação WPS prevê a utilização de mensagens assíncronas através de um mecanismo de 'pull' de mensagens, sendo uma solicitação respondida imediatamente com um reconhecimento da mesma. Um documento XML é retornado com uma URL para verificar o status do execute, assim caso o usuário tente recuperar as estatísticas e o processo não estiver finalizado, será retornado um documento indicando o tempo restante para a conclusão do processo. Quando o processo estiver finalizado será retornado um documento com os resultados. Este modo é mais conveniente para o usuário, uma vez que o processo pode ser monitorado.

Abordagens Alternativas

- Outra abordagem seria o mecanismo de mensagens assíncronas chamado 'push', adotado pela especificação WNS;
- Este mecanismo permite que o usuário seja notificado da ocorrência de um evento, sendo necessário se registrar informando seus dados e um alvo de notificações;
- É proposto a integração deste mecanismo de mensagem ao serviço de estatística, para que o usuário seja notificado no fim da operação com uma mensagem contendo o local do resultado da operação, etc.

Abordagens Alternativas

- Chegou-se a conclusão que o mecanismo de mensagens 'pull' seria melhor, criando um mecanismo de comunicação entre máquinas, sendo mais prático para o usuário;
- Ainda há uma questão sobre como o resultado de chamadas assíncronas seria tratado, pois de modo geral não há políticas para isso. (Nota-se aqui uma preocupação com a padronização)

Melhorando a eficiência dos serviços de processamento

- É proposto a subdivisão do serviço de estatística, em serviços mais elementares e passando a responsabilidade de criação do fluxo de trabalho ao usuário, utilizando técnicas de encadeamento de serviços;
- Para isso seria necessário realizar adaptações na especificação WPS, extendendo o conceito de execução, permitindo criar uma cadeia de operações e também mudanças na sintaxe;
- Outra abordagem a ser investigada é o processamento na origem dos dados, sendo implementado com a especificação WFS, surgindo uma variação desta, o WFS-P;
- Esta abordagem diminuiria o tamanho dos dados solicitados, sendo enviado apenas o resultado do processamento;

Conclusões

- IDE's podem ir além dos serviços de descoberta e visualização de dados, utilizando-as para desenvolver aplicações distribuídas;
- É possível obter a interoperabilidade utilizando soluções de diferentes fornecedores;
- Foi esboçada uma especificação de processamento web para dar suporte a implementação de serviços de geoprocessamento distribuído e interoperável;
- Foi visto que catalogos de metadados e serviços são a espinha dorsal de aplicações que utilizam dados distribuídos;
- Foram apresentados os benefícios de um ambiente de geoprocessamento distribuído;
- Áreas que necessitam de pesquisa: distribuição de algoritmo, ao invés de dados;
- Separação entre geometria e atributo (uso de Geolinking);
- Utilização das tecnologias WSDL, SOAP, UDDI e WS-BPEL;