



Proposta de uso de computação em nuvem em aplicações de zoneamento

Sérgio Aparecido Braga da Cruz ¹
João dos Santos Vila da Silva¹
Sandra Mara Alves da Silva Neves²

¹ Embrapa Informática Agropecuária
Av. André Toselo, 209 - Caixa Postal 6041
13083-886 - Campinas - SP, Brasil
{sergio.cruz, joao.vila}@embrapa.br

² Universidade do Estado de Mato Grosso - UNEMAT
Av. dos Ingás, 3001 - Jardim Imperial
78555-000 - Sinop, MT
ssneves@unemat.br

Resumo. A demanda por produtos de geoprocessamento de melhor qualidade, com processos de produção ágeis e rastreáveis baseados no uso de dados com maior escala e resoluções espacial e temporal têm estimulado a adoção de novos paradigmas de produção de dados georreferenciados. A abordagem tradicional baseada no uso de ferramentas *desktop* tem sido complementada com o uso de infraestruturas virtuais para armazenamento e processamento de grande volume de dados. Este artigo apresenta as bases deste novo paradigma de construção de aplicações em geoprocessamento, e ilustra a sua aplicação na produção de um zoneamento do cultivo da soja no Mato Grosso.

Palavras-chave: computação em nuvem , zoneamento agrícola.

Abstract. The demand for better quality GIS products with agile and traceable production processes based on the use of data with higher resolution and scale have stimulated the adoption of new georeferenced data production paradigms. The traditional approach based on the use desktop tools have been supplemented by the use of virtual infrastructure for storing and processing large amounts of data. This article presents the foundations of this new application building paradigm in geoprocessing, and illustrates its application in the production of a soybean crop zoning in Mato Grosso.

Key-words: cloud computing, crop zoning.

1. Introdução

Imagens de alta resolução do sistema *RapiEye* estão disponíveis para órgãos públicos a partir da assinatura do convênio com o governo federal, que possibilitou a sua utilização por estes órgãos de forma gratuita principalmente na realização do cadastro ambiental rural (lei 12.651 de 2012) (Mengatto Júnior e Silva, 2014). Dados de altíssima resolução obtidos a partir de sensores de radar do tipo SAR (*Synthetic Aperture Radar*) estão disponíveis no mercado para suporte nos mais diferentes estudos e análises geoespaciais (Antunes e Siqueira, 2013). A produção de dados georreferenciados nos mais diferentes temas por usuários especialistas ou leigos tem aumentado rapidamente a partir do surgimento de sensores acoplados à dispositivos móveis.

Associando o processamento destes dados à análises de séries temporais, integração com dados de diferentes domínios (bioinformática, modelagem climática, etc. ...), e análise de grandes áreas, tais como aquelas necessárias para obtenção de levantamentos, caracterizações e monitoramentos de recursos naturais, análises socioeconômicas, zoneamentos e avaliação de cenários sobre grandes biomas brasileiros atingimos um cenário que supera a capacidade de processamento de laboratórios tradicionais de geotecnologias apoiados em aplicações *desktop*. Ainda podemos considerar a urgência na produção de resultados em situações de catástrofes ambientais, agrícolas e sociais, associadas a requisitos mais rígidos quanto a precisão espacial e de classificação exigidas por aplicações de impacto econômico, planejamento e tomada de decisão. A execução de atividades de análise espacial e geoprocessamento neste contexto implica no aumento crescente de gastos na compra de recursos computacionais para sua realização.

Porém este aumento de capacidade no processamento de dados geoespaciais pode ser obtido alternativamente por meio da utilização de infraestruturas virtuais de processamento ou computação em nuvem. Estas infraestruturas possibilitam o processamento e integração remota de dados complementando a capacidade de computação de infraestruturas de processamento locais. Neste artigo descrevemos alguns tipos de infraestruturas disponíveis e exemplificamos como uma aplicação de geoprocessamento desenvolvida de forma tradicional com o uso de ferramentas instaladas em *desktop* pode ser migrada para este novo paradigma de construção de aplicações.

2. Objetivo

Neste artigo são apresentados alguns conceitos sobre a computação em nuvem na área de processamento de dados geoespaciais e como aplicações reais poderiam ser realizadas nesta infraestrutura.

3. Material e Método

3.1 Computação em Nuvem

A computação em nuvem possibilita o acesso a recursos computacionais de forma remota como uma prestação de serviço. A especificação dos serviços a serem utilizados neste ambiente é

dinamicamente configurável apoiada em um ambiente de computação distribuída formado por componentes que implementam funções específicas e que podem ser combinados para atendimento da demanda do usuário. Neste contexto, componentes podem ser vistos como computadores físicos, dispositivos para armazenamento de grande volume de dados (*storages*) e de dispositivos para organização de redes. Sobre estes componentes físicos são implementados máquinas virtuais, que correspondem a programas simuladores de computadores reais que podem ser **conFigurados**, executados e interrompidos de acordo com a demanda por processamento. Estes ambientes podem ser fornecidos externamente por empresas terceirizadas (nuvem pública), podem ser totalmente implantados internamente em uma instituição (nuvem privada) ou podem ser implantadas de forma a combinar estas duas possibilidades (nuvem híbrida), onde o uso de recursos terceirizados é realizado em momentos de demanda extraordinárias de processamento.

Os serviços de computação em nuvem podem ser classificados em 4 tipos segundo Yang et al (2011):

IaaS (Infrastructure as a Service): infraestrutura como serviço, possibilita o acesso a recursos computacionais físicos (computadores, recursos de rede, recursos de armazenamento, etc. ...) de forma virtual. O usuário do serviço tem autorização para **conFiguração** destes recursos de acordo com a sua necessidade, podendo instalar sistemas operacionais e softwares mais adequados a realização de suas tarefas.

PaaS (Platform as a Service): na plataforma como serviço, além do acesso aos recursos computacionais físicos, é fornecido também acesso a uma série de ferramentas de programação que possibilitam desenvolvedores de software implementarem suas aplicações de forma remota.

SaaS (Software as a Service): No software como serviço, o usuário final tem acesso as funcionalidades de um software de forma remota, normalmente através de um navegador Web.

DaaS (Data as a Service): no dado como serviço, operações relativas a busca, localização, acesso aos dados e a sua integração são providas de forma remota sob demanda para os usuários finais. A localização física tanto dos usuários consumidores quanto dos produtores dos dados é irrelevante.

3.1 Serviços de computação em nuvem para dados geoespaciais

Google Earth Engine

O *Google Earth Engine* integra um conjunto de dados geoespaciais com cobertura global, a capacidade de armazenamento e processamento gigantescos correspondentes ao ambiente de computação em nuvem da *Google*, e um ambiente de desenvolvimento integrado suportando a implementação de algoritmos de análise e processamento de dados geoespaciais nas linguagens *JavaScript* e *Python* (Google Earth Engine Team, 2015).

Dentre os dados geoespaciais disponíveis no catalogo da *Google* podemos destacar vários produtos com cobertura regional e/ou global dos sensores *ASTER*, *MODIS*, *LANDSAT*, *NOAA* e produtos *MeaSURES*. Além dos dados já disponíveis na infraestrutura da *Google*, o usuário pode enviar os seus dados para a integração/processamento remoto. O *Google Earth Engine* pode ser classificado como uma *PaaS*, que fornece para especialistas, um ambiente de programação

e um conjunto de dados para implementação de aplicações geoespaciais, principalmente para dados *raster*.

Este ambiente se encontra em estágio experimental e o acesso é limitado e sem custos para usuários avaliadores previamente cadastrados e aprovados para desenvolvimento de aplicações de pesquisa, educação ou sem fins lucrativos. A necessidade de conhecimento em linguagens de programação *JavaScript* ou *Python*, apesar de possibilitar construção de aplicações geoespaciais genéricas, também representa uma sobrecarga inicial de aprendizagem, o que pode limitar a utilização deste ambiente.

ArcGIS Online

É um ambiente de computação em nuvem do tipo *SaaS*, que oferece para seus usuários acesso a ferramentas e programas para geoprocessamento da empresa *ESRI*. O preço de acesso ao ambiente varia de acordo com o número de usuários e de serviços que serão acessados. Neste ambiente são oferecidos serviços para visualização, análise e soluções de processamento configuráveis para atender aos diferentes tipos de clientes da plataforma. O serviço de *IaaS* da Amazon (Amazon Web Services, 2016) fornece infraestrutura física para execução das ferramentas *ArcGIS*. O ambiente *ArcGIS Online* fornece soluções especializadas que podem ser diretamente adotadas por empresas ou usuários que já possuam fluência no conjunto de ferramentas para geoprocessamento da empresa *ESRI*, porém esta abordagem limita e dificulta o uso compartilhado de ferramentas especializadas externas, utilizadas na área de pesquisa, por exemplo.

Soluções de computação em nuvem privadas

Atualmente, podemos encontrar um grande número de ferramentas, inclusive to tipo *open source*, que permitem a construção de um ambiente de computação em nuvem corporativo. As diferentes camadas da infraestrutura podem ser implementadas utilizando, por exemplo, o conjunto de ferramentas indicado na **Tabela 1** abaixo.

Tabela 1. Camadas de serviços de ambiente de computação em nuvem

DaaS	SaaS	Serviços de dados e de processamento com suporte aos padrões de descrição e intercâmbio de dados geoespaciais (Padrões OGC, GeoJSON, KML, etc).
PaaS		- Ambientes de workflow científico e dataflow (Kepler, Galaxy, Apache NiFi, etc...); - Ferramentas e linguagens de processamento especializadas (JavaScript, Python, Java, etc. ...); - Plataforma <i>Cloud Foundry</i> (The Cloud Foundry Foundation, 2016); - Plataforma <i>OpenShift</i> (Red Hat, 2016).
IaaS		<i>OpenStack</i>

A base para construção de um ambiente de computação em nuvem está na capacidade de gerenciamento dinâmico de recursos computacionais. Plataformas para Gerenciamento de Nuvens (*CMP - Computing Management Platforms*) como a ferramenta *OpenStack* possibilitam a gestão de recursos computacionais instalados em um *datacenter*. Nela um administrador define a quantidade máxima e a configuração de recursos a serem utilizados por um usuário final, de acordo com requisitos estabelecidos para prestação do serviço *IaaS*. A ferramenta controla o uso dos recursos pelo usuário de forma que permaneçam dentro dos limites estabelecidos (Rackspace Cloud Computing, 2016) e possibilitem o uso otimizado dos recursos disponíveis.

A camada de *PaaS*, sobre a camada *IaaS*, pode ser implementada de várias formas. Ferra-

mentas como a *Cloud Foundry* e *OpenShift* possibilitam a construção de ambientes virtuais para desenvolvimento de aplicações genéricas com suporte à linguagens de programação *Python*, *PHP*, *Java* e *Ruby*, dentre outras. Esta camada pode também ser implementada a partir do zero através da instalação e configuração de ferramentas que possibilitem a construção de um ambiente virtual para desenvolvimento de aplicações. Ambientes de *workflow* científico e *dataflow*, tais como *Kepler* (Kepler Project, 2016), *Galaxy* (Galaxy Team, 2016) e *Apache Nifi* (Apache NiFi, 2016) podem ser utilizados como *PaaS*, provendo uma visão de alto nível para construção de aplicações para domínios especializados. Um exemplo de infraestrutura que adota esta estratégia esta sendo implementado pelo projeto da Embrapa denominado *GeoCloud* (Cruz, 2015). Este projeto tem como objetivo implantar uma infraestrutura virtual para computação distribuída no domínio geoespacial com foco no processamento de grandes volumes de dados geoespaciais.

4 – Resultados

4.1 – Workflow de zoneamento do cultivo da soja no Mato Grosso

Para avaliar a viabilidade de utilização de um ambiente de computação em nuvem em aplicações práticas foi realizado uma prova de conceito que consistiu no mapeamento para este ambiente das atividades realizadas em um zoneamento de soja na região do Mato Grosso (Kreitlow et al, 2014). Para realização desta tarefa assumimos que todos os dados necessários para o zoneamento estão disponíveis na forma digital com acesso por meio de serviços de dados no padrão OGC (OGC, 2016). As atividades necessárias ao zoneamento foram mapeadas para um ambiente de *dataflow*, disponível na plataforma *GeoCloud*. Este ambiente funciona como uma *PaaS*, fornecendo ferramentas para construção de aplicações geoespaciais utilizando linguagem gráfica baseada em *workflow*. O especialista no domínio define um fluxo de tratamento de dados, por meio da edição de um *workflow*. Esta edição corresponde a seleção e configuração de componentes disponíveis em um catálogo de componentes que implementam funções para acesso e processamento de dados geoespaciais. O *workflow* gerado ao final da edição representa um processo para geração de produtos geoespaciais, que pode ser parametrizado e executado repetidamente.

O *workflow* gerado para reprodução das atividades realizadas durante a produção de um zoneamento de soja da região sudoeste do Mato Grosso está representado na **Figura 1**.

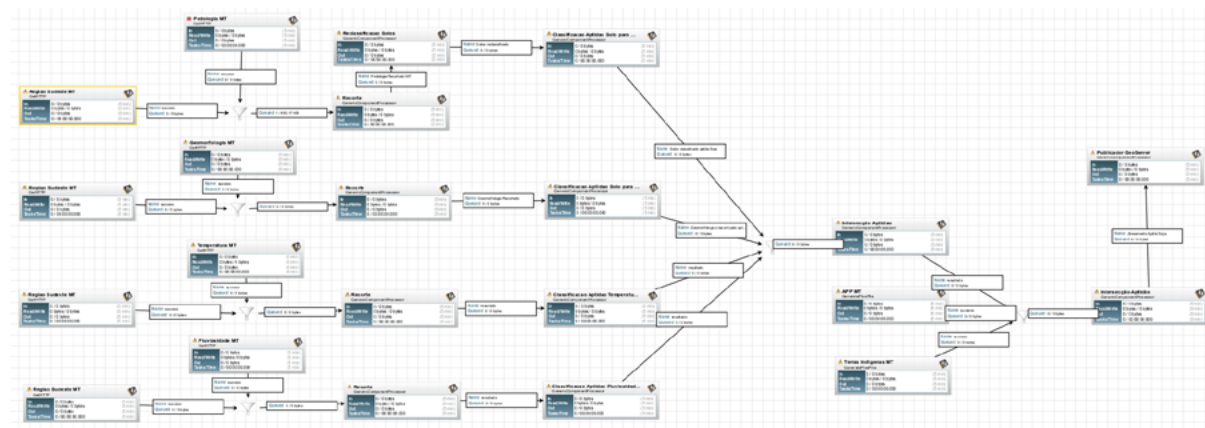


Figura 1. Workflow representando o processo de produção de zoneamento de soja

A **Figura 2** apresenta de forma mais clara partes deste processo. Inicialmente os dados básicos para a realização do zoneamento são obtidos de serviços do tipo OGC utilizando componentes *GetHTTP*. Dado o endereço *URL* dos serviços OGC, estes componentes permitem a consulta e integração dos dados solicitados ao ambiente de *workflow* (**Figura 2a**). Os dados especificados no *workflow* correspondem a máscara da região sudoeste do Mato Grosso e dados sobre pedologia, geomorfologia, temperatura e pluviosidade da região. No passo seguinte o *workflow* indica componentes para recorte dos dados geoespaciais de acordo com a máscara da região de estudo (**Figura 2b**). Os dados de solo recortados são reclassificados e um mapa de aptidão para a cultura da soja pode ser gerado. Mapas de aptidão baseados nos dados de geomorfologia, temperatura, e pluviosidade devem então também serem gerados. Na sequência, os mapas de aptidão são cruzados e, do resultado, são excluídas regiões relativas a APP's (Áreas de Proteção Permanente) e Terras Indígenas. O mapa resultante deve então ser publicado em um servidor de mapas do tipo *GeoServer* (OSGeo, 2016).

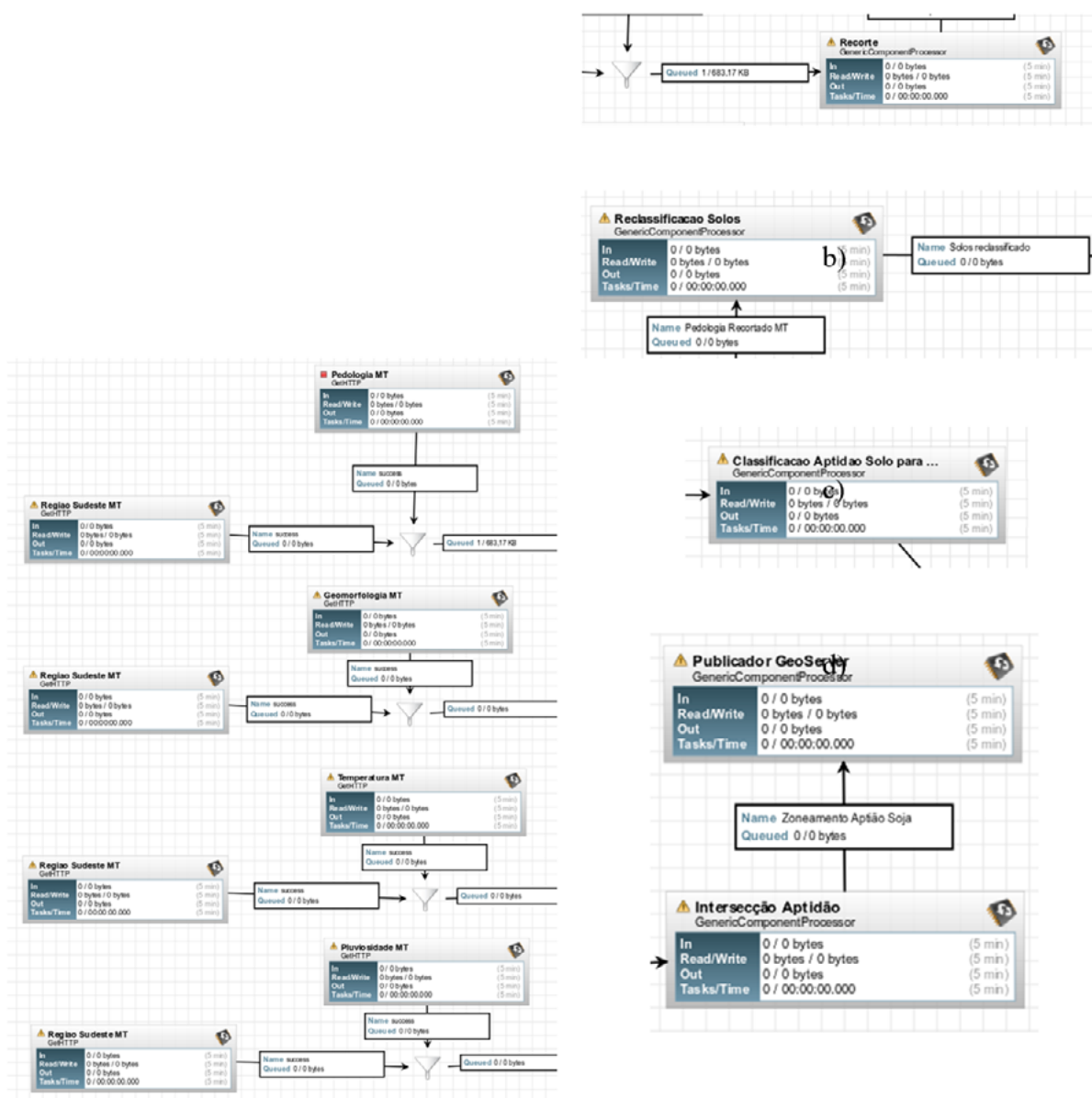


Figura 2. *Workflow* de zoneamento em detalhes a) componentes para entrada de dados; b) recorte; c) reclassificação de solos; d) classificação quanto a aptidão; e) cruzamento e publicação dos resultados.

5 - Conclusões

Existem várias possibilidades para uso e/ou implantação de ambientes de computação em nuvem. Soluções mais gerais e poderosas, porém que exigem maior conhecimento sobre programação, como a representada pelo *Google Earth Engine* até soluções mais específicas representadas pelo acesso à ferramentas para geoprocessamento e análise espacial proprietárias como a solução apresentada pela *ESRI*. Entre estas duas estratégias de acesso a serviços de computação em nuvem, soluções intermediárias podem ser implementadas com base nas diversas ferramentas *open-source* disponíveis. A adoção de *PaaS* baseados em *workflow* científico, como a que esta sendo implementada pelo projeto *GeoCloud*, permite a distribuição adequada do esforço para construção de aplicações geoespaciais entre dois tipos de usuário. O usuário com perfil de desenvolvedor pode implementar componentes de *workflow*, usando linguagens de programação mais adequada. Estes componentes podem ser compartilhados e reutilizados em diferentes *workflows*. Especialistas no domínio geoespacial, tais como, geógrafos, biólogos e ecologistas constroem suas aplicações utilizando uma linguagem de mais alto nível representada pelos *workflows*, a qual facilita a interação do usuário com o ambiente e aumenta sua produtividade. Além desta vantagem, o registro de processos de tratamento de dados na forma de *workflow* facilita a sua reprodução, ajuste e compartilhamento dentro de uma comunidade de especialistas, contribuindo em uma maior possibilidade de validação e crítica destes processos. Além dos aspectos relativos a construção de aplicações geoespaciais, os ambientes de computação em nuvem possibilitam, uma maior capacidade de processamento de dados e possibilidade de compartilhamento de dados entre usuários. Desta forma estes ambientes viabilizam a realização de atividades de forma remota envolvendo grande volume de dados e de esforço computacional por usuários possuindo computadores de baixo custo. No caso das atividades de um zoneamento como o descrito neste artigo, todo o trabalho pode ser realizado sem que nenhum dado geoespacial seja transmitido para o computador do usuário. Também nenhuma ferramenta é exigida do usuário além do navegador Web e de uma conexão de Internet.

6 – Referências

- Amazon Web Services, Inc. Amazon Web Services (AWS) - Cloud Computing Services. Disponível em : <<https://aws.amazon.com>>. Acesso em: jun. 2016
- Antunes, M. A. H.; Siqueira, J. C. S. Características das imagens RapidEye para mapeamento e monitoramento e agrícola e ambiental. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 16. (SBSR)., 2013, Foz do Iguaçu. **Anais...** São José dos Campos: INPE, 2013. p. 547-554. DVD, Internet. ISBN 978-85-17-00066-9 (Internet), 978-85-17-00065-2 (DVD).
- Apache NiFi, **Apache Nifi** Disponível em: <<https://nifi.apache.org/>> Acesso em: julho, 2016.
- ArcGIS, **ArcGIS Online**. Disponível em <<http://www.esri.com/software/arcgis/arcgisonline>>. Acesso em: julho 2016.
- Cruz, S. A. B. **Infraestrutura virtual para processamento e análise integrada de dados geoespaciais** Campinas: Embrapa Informatica, 2015. 27 p. (Embrapa. Macroprograma 5 - Desenvolvimento Institucional, Linha Tematica DataExp - Armazenamento e Processamento de de Dados Experimentais da Embrapa Projeto 05.14.03.003.00.004). Projeto em execução GeoCloud
- Galaxy Team, **The Galaxy Project: Online bioinformatics analysis for everyone** Disponível em: <<https://galaxyproject.org/>>. Acesso em: julho, 2016.
- Google Earth Engine Team, **Google Earth Engine: A planetary-scale geo-spatial analysis platform** Disponível em: <<https://earthengine.google.com/>>, dez. 2015.
- Kepler Project, **The Kepler Project** Disponível em: <<https://kepler-project.org/>>. Acesso em: julho, 2016.
- Kreitlow, J. P.; Neves, S. M. A. da S.; Silva, J. dos S. V. da; Neves, R. J.; Paiva, S. L. P. de. Zoneamento do culti-

vo da soja (*Glycine max L.*) na região Sudoeste de planejamento do estado de Mato Grosso. In: **SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL**, 5., 2014, Campo Grande, MS. Anais... São José dos Campos: INPE, 2014. p. 876-885. 1 CD-ROM. Geopantanal 2014.

Mengatto Júnior, E. A.; Silva, J. dos S. V. da. Imagens de alta resolução (rapideye) para elaboração de mapas para planejamento e gerenciamento: estrada parque do Pantanal, MS. In: **SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL**, 5., 2014, Campo Grande, MS. Anais... São José dos Campos: INPE, 2014. p. 513-522. 1 CD-ROM. Geopantanal 2014.

OGC, **OGC® Standards and Supporting Documents**, Disponível em: <<http://www.opengeospatial.org/standards>>. Acesso em: 10 mai. 2016.

Open Source Geospatial Foundation - OSGeo, **GeoServer** Disponível em: <<http://geoserver.org/>> Acesso em 10. maio, 2016.

Rackspace Cloud Computing. **OpenStack Open Source Cloud Computing Software** Disponível em: <<http://www.openstack.org/>> Acesso em:jun 2016.

Red Hat, **The Open Source Container Application Platform** Disponível em: <<https://www.openshift.org>>. Acesso em: jun. 2016.

The Cloud Foundry Foundation, **Cloud Foundry | The Industry Standard For Cloud Applications**. Disponível em: <<https://www.cloudfoundry.org>>, acesso em: jun. 2016.

Yang, C.; Goodchild, M.; Huang, Q.; Nebert, D.; Raskin, R.; Yan Xu, Bambacus, M.; Fay, D. Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing?, **International Journal of Digital Earth**, 2011. 4:4, 305-329, DOI: 10.1080/17538947.2011.587547