Geospatial Information System (GIS)

- Overview
  - Summary
  - Workflow
  - Data
  - Cloud platform
  - Issues/Gaps
- Spatial cloud computing 1
  - Summary
  - Workflow
  - Data
  - Cloud platform
  - Cloud performance
  - Issues/Gaps
- Massive spatial data processing model based on cloud computing model 2
  - Summary
  - Workflow
  - Data
  - Cloud platform
  - Cloud performance
  - Issues/Gaps
- Deployment of GEOSS clearinghouse on Amazon's EC2 3
  - Summary
  - Workflow
  - Data
  - Cloud platform
  - Cloud performance
  - Issues/Gaps
- Cloud computing platform for GIS image processing in U-city 4
  - Summary
  - Workflow
  - Data
  - Cloud platform
  - Cloud performance
  - Issues/Gaps
- A Cloud PaaS for High Scale, function, and velocity mobile applications 5
  - Summary
  - Workflow
  - Data
  - Cloud platform
  - Cloud performance
  - Issues/Gaps
- Studies on application of cloud computing techniques in GIS 6
  - Summary
  - Workflow
  - Data
  - Cloud platform
  - Cloud performance
  - Issues/Gaps
- In-Device spatial cloaking for mobile user privacy assisted by the Cloud 7
  - Summary
  - Workflow
  - Data
  - Cloud platform
  - Cloud performance
  - Issues/Gaps
- Geospatial editing over a federated cloud geodatabase for the state of NSW 8
  - Summary
  - Workflow
  - Data
  - Cloud platform
  - Cloud performance
  - Issues/Gaps
- GIS in the cloud: implementing a web map service on Google App Engine 9
  - Summary
  - Workflow
  - Data
  - Cloud platform
  - Cloud performance
  - Issues/Gaps
- PerPos: a platform providing cloud services for pervasive positioning 10
  - Summary
  - Workflow
  - Data
  - Cloud platform
  - Cloud performance
Overview

Summary

The main information technology challenges in geospatial sciences are 1) accessing and processing of large volumes of geospatial data, 2) dealing with massive concurrent users access, and 3) spatiotemporal intensive applications.

There is some review paper for this area


Workflow

The cloud computing workflow in geospatial sciences mainly involves:

- Data storage and processing
- Simulation and modeling

Data

In the papers, small data size are used for testing because the data size in GIS are usually with very big volume. For example, the dynamic real-time routing for a metropolitan region such as D.C. requires storage of 1TB for daily basis and 1PB for a yearly basis to retain historical records.

Cloud platform

- Currently, the most popular cloud service in GIS area is IaaS. Amazon AWS and OpenNebula are used since the users can conveniently implement OS-specified applications on the platform.
- PaaS are also used in GIS area such as Google App Engine. Some research group developed or proposed their own cloud platform such as PerPos.

Issues/Gaps

Common issues:

- Security. For example, the geospatial data of a country are usually very sensitive that it will raise concern when the sensitive data are stored in cloud provided by foreign organization.
- Privacy. Many location-based services involve the location and identity information of users, and location and identity privacies need to be considered.
- DaaS is essential to geospatial sciences because large volume of data such as geospatial images need to be stored and processed.

Spatial cloud computing

Summary

- This paper discusses how cloud computing could enable geospatial sciences and how spatiotemporal principles could be utilized to ensure the benefits of cloud computing.
- This is an overview paper about geospatial cloud computing, so there is no much detailed information about some specific workflow and cloud platform. The paper presents four research examples to analyze cloud computing requirements for different geospatial sciences applications.
- We can start from this paper to find scientific stories in GIS area.

Workflow

This paper does not investigate any specific workflow. Instead, the paper discusses four scientific and application scenarios.
• Data intensity scenario
  • A DaaS is developed based on spatial cloud computing. The DaaS is designed to maintain millions to billions of metadata entries.
  • The DaaS is developed and tested based on Microsoft Azure, Amazon EC2, and NASA Cloud Services for geospatial community.

• Computing intensity scenario
  • geospatial science phenomena are intrinsically computing-expensive to model and analyze
  • parameter extraction is required to execute complex geophysical algorithms
  • geospatial phenomena simulation is complex

• Concurrent-access-intensity scenario
  • Concurrent-intensive accesses (Google Earth) have access spikes, which needs to be responded by spatial cloud computing elastically invoke service instances from multiple locations.

• Spatiotemporal intensive scenario
  • Spatiotemporal indexing
  • spatiotemporal data modeling methods
  • Earth science phenomena correlation analyses
  • hurricane simulation
  • computer network
  • Example: real-time traffic routing

Data
See the "workflow" section for description. One example is for dynamic real-time routing for a metropolitan region such as D.C. which requires storage of 1TB for daily basis, 10TB for a weekly basis, 1PB for a yearly basis to retain historical records.

Cloud platform
• IaaS(e.g. Amazon EC2)
  • IaaS let users have full control over the virtualized machine. So IaaS users should have system administrative knowledge about OS.
• PaaS(e.g. Microsoft Azure and Google App Engine)
  • Parameter Extraction such as Vegetation Index (VI) or Sea Surface Temperature (SST) with PaaS since they involve a complex series of geospatial processes.
• SaaS(e.g. salesforce.com and Gmail)
  • Knowledge and Decision Support
  • Social Impact and Feedback. SaaS such as Facebook and email
• DaaS: DaaS supports data discovery, access, and utilization and delivers data and data processing on demand to end users.
  • Earth Observation (EO) data access with DaaS.

Cloud performance
See above sections. No detailed description was provided in the paper.

Issues/Gaps
• Spatiotemporal principle mining and extracting
• Important Digital Earth & complex geospatial science and applications
• Supporting the SCC characteristics
  • Amazon S3 suffered an outage lasting about two hours in 2008.
  • Cloud provider needs to provide services for perpetuity. (Coghead and "The Linkup" close business)
• Security
• Citizen and social science
  • Trustworthy: data and information authority
  • Privacy: open environment to provide or receive services
  • Ethical: GPS and location-based services.

Massive spatial data processing model based on cloud computing model

Summary
• A massive spatial data processing model is designed based on cloud computing model.
• Speed and processing capacity of stand-alone machine is bottleneck of immediate processing of spatial data (more than half a year to complete the production of Orthophoto Map for a medium-size city).

Workflow
• Workflow of the cloud spatial data processing model
  1. Preprocessing: deliver spatial data (remote sensing digital image) to cloud server for geometric correction, radiometric correction and other treatment.
  2. Processing: image data format conversion, image enhancement and equalization, band integration etc. The powerful computing resources of cloud system is used to achieve real-time image processing.
  3. Postprocessing: information extraction, classification, and thematic map production.
Data
- 6000 Fu DMC aerial digital images for a mid-size city.
- Each image is 7680X13824 pixels with image scale 1:12000.

Cloud platform
- DPGrid (Digital Photogrammetric Grid) by Wuhan University.
- There is actually no experiment on a real cloud computing system for the theoretical model.

Cloud performance
- The DPGrid (with 8 blade servers) can produce mosaic with the 6000 images in 15 days.
- Traditional modes requires 10 more staff members and need more than one year.

Issues/Gaps
- No discussion in the paper
- The paper presents a theoretical cloud model, but the platform they tested on is actually a parallel computing system (DPGrid).

Deployment of GEOSS clearinghouse on Amazon's EC2

Summary
- The GEOSS clearinghouse, a web based Geographic Metadata Catalog system, manages millions of the metadata of the spatially referenced resources for the Global Earth Observations (GEO).
- The GEOSS clearinghouse is deployed, maintained and tested on Amazon EC2.
- A very important CSW request for the GEOSS clearinghouse, GetCapabilities, is used to test the performance of different categories of Amazon EC2 instances.

Workflow
- Typical process of deploying GEOSS clearinghouse onto Amazon EC2
  1. A public Amazon Machine Image (AMI) is customized to launch an EC2 instance.
  2. Login to the instance through SSH.
  3. Database software Postgresql with PostGIS and the servlet container tomcat are installed. (Postgresql is to support spatial datasets which keep the data of GEOSS Clearinghouse, the servlet container tomcat is used to host GEOSS Clearinghouse)
  4. Setup a EBS volume to keep the data files of Postgresql.
  5. Transfer the GEOSS clearinghouse codes/data into the virtual server.
  6. Install Tomcat, Jetty or other servlet container.
  7. Start servlet container.
  8. Create a new AMI based on the running instance.

Data
- The corresponding average response times of concurrent request number are investigated.
- No other data description

Cloud platform
- Amazon EC2 is used.
- Amazon EBS is used to store restore the database on EC2 instances.
- Amazon SQS is used for GEOSS clearinghouse to scale up and scale down the instances automatically.

Cloud performance
- m1.large, m1.xlarge, m2.xlarge, m2.2xlarge, m2.4xlarge, and c1.xlarge have faster response times than the other two (m1.small, c1.medium).
- There is little difference between instances with different number of CPU cores and memory, because only one core of these six instances are used.
- Standard Linux Large instance (m1.large) costs approximately 0.34 and has relatively high performance.

Issues/Gaps
- The GetCapabilities response time depends on the number of metadata records in catalog. Using MapReduce for indexing might be a solution to improve the performance of GEOSS Clearinghouse.
- Geospatial middleware implementation for cloud computing is more difficult than other computing paradigms.
- Service interoperability.
Cloud computing platform for GIS image processing in U-city

Summary

- U-city is a city with ubiquitous information tech that citizens can access the converge information anywhere and anytime. So the large amount of data need to be processed in real-time.
- The paper presents a GIS image processing platform using cloud computing. The platform finds and selects optimal computing resources.
- A use case of parallel air pollution map generation is used to show the efficiency of the processing of the massive data of the platform.

Workflow

- Workflow of GIS image processing platform (Process image with GIS data automatically)
  1. Loads the virtual machines on the nodes
  2. submits the job from request of user to VMJM
  3. VMJM assigns job to job queue and request the computing resource suitable for the job to Cloud Infrastructure Manager (CIM)
  4. CIM reserves the VMs through Haizea
  5. Haizea schedules VM launching through OpenNebula
  6. OpenNebula launches VMs to optimal compute nodes and CIM returns the ready message to VMJM
  7. VMJM executes the job
  8. returns the result of the job to the user.

Data

- The air pollution map area of Cheonggyecheon, Seoul, Korea is selected.
- Data size: 250GB

Cloud platform

- Consists of OpenNebula/Haizea and Virtual Machine Job Manager (VMJM).
- Uses Hadoop to implement the generation of the parallel air pollution map.
- Different IaaS-style open source framework are compared (Eucalyptus, Nimbus, OpenNebula).
- OpenNebula is chosen because they can easily lease computing resources from other cloud platform and provide API for developers.

Cloud performance

- 1200 seconds with 10 VM nodes (8 nodes with Dual core Intel processor and 2 nodes with Quad core Intel processor, 4MB memory for each node)

Issues/Gaps

- The performance evaluation of the air pollution map generation is too simple (the authors set it as future job).

A Cloud PaaS for High Scale, function, and velocity mobile applications

Summary

The paper investigated the requirements of a PaaS necessary for supporting a car with a rich collection of mobile devices.

Workflow

The attributes of the applications of a "Connected Car":

- Used by 100's of millions of users
- Highly interactive. Applications serve as "sensors"
- Mobile

Data

Cloud platform

- N/A

Cloud performance

- N/A

Issues/Gaps
Specific requirements for a PaaS platform:

- Zoning in addition to Autonomic/Elastic management of Host, Storage and Network resource
- Self-scaling code container
- Event system including correlation
- Connection management
- Network enablers for communications

Studies on application of cloud computing techniques in GIS

Summary

- Studied applications of cloud computing techniques in GIS such as massive data storage, spatial data processing and analysis.
- Introduced a GIS platform solution based on open source architecture of cloud computing – SuperMap SGS.

Workflow

NA

Data

NA

Cloud platform

NA

Cloud performance

NA

Issues/Gaps

- Data security: sensitive geospatial data should not be stored in foreign country
- Supervision: government departments, third party organization, cloud provider?
- Regulations

In-Device spatial cloaking for mobile user privacy assisted by the Cloud

Summary

Spatial cloaking methods are used to protect identity privacy in location-based service. By using cloud, some third parties in the cloud will provide user density information, and the spatial cloaking is done in the mobile device. The cloud is used to provide user density information which the cloud collects from users.

Workflow

NA

Data

NA

Cloud platform

NA

Cloud performance

NA

Issues/Gaps

The paper proposed new spatial cloaking algorithms, there is no current cloud service used.
Geospatial editing over a federated cloud geodatabase for the state of NSW

Summary

- A versioned editing model for a geospatial cloud database environment is presented in the paper.
- The cloud database is used so that each government agency across the state would have access to the common shared data layer.

Workflow

Versioned editing workflow:

1. versioning
2. editing
3. reconciling
4. posting

In the paper, the editing workflow is taking place in the cloud.

Data

Federated geodatabase of Australia’s NSW state is used.

Cloud platform

- IaaS was explored for hosting the federated geodatabase
- Amazon EC2 is used in this case (An 1.7GB memory, 160GB storage single virtual core instance)

Cloud performance

By moving the geodatabase to the cloud:

- operate as if the data is in data center
- individual organizations can contribute to a statewide federated data editing
- individual organizations can benefit from editing made in other org
- reduce duplication, redundancy and minimizes the data editing overload
- increase the accuracy of statewide geospatial data

Issues/Gaps

GIS in the cloud: implementing a web map service on Google App Engine

Summary

Implementation of a Web Map Service for raster imagery within the Google App Engine environment.

Workflow

- Use Apache JMeter to set up and run a number of test scripts
- Each involved a single client machine making repeated requests to the server for images

Data

- 1G free-usage quota per day for GAE (the outgoing bandwidth from the server)
- Each test started with a single thread, ramping up linearly over at least 30 seconds to a maximum number of threads
- Each thread looped repeatedly through a set of 42 image requests in random order
- Each image was of size 256x256 pixels, representing three “zoom level” in WGS:84 latitude-longitude coordinates

Cloud platform

- Google App Engine (low cost and automated scalability)
- PaaS providing a custom hosting environment, applications have to be specially developed

Cloud performance

- Application scales well to multiple simultaneous users and performance will be adequate for many applications.
- "static files" configuration: serving 500 requests per second (limit of GAE’s default capability under free usage)
"self-caching" configuration:
- serving around 200 requests per second (see many failures due to the default GAE limit of 30 simultaneous dynamic requests)
- serve tiles to 50 simultaneous clients at latencies of ~300 ms

"fully-dynamic" configuration:
- 25 requests per second with 20 client threads (failed with over 10 client threads)
- serve 10 simultaneous clients with latencies of ~500 ms for 256x256 images
can serve around 100,000 JPEG images (10kB per image) per day at no cost
- or 10,000 PNG images (100kB per image) per day

Issues/Gaps
- There is a free-usage quota under heavy load for GAE
- Latency spikes degrade users’ experiences

PerPos: a platform providing cloud services for pervasive positioning

Summary
- A cloud platform for pervasive positioning (PerPos) cloud service is described
- The PerPos platform provides services for positioning and location-based applications
- Cloud platform allows for adding new features and improved algorithms and methods without requiring re-compile or re-distribution of new binaries to the users

Workflow

PerPos cloud services:
- Awareness of Positioning Quality
- Sensor Fusion
- Building Models
- Navigation Primitives
- Power Efficient Tracking
- Behavior Recognition

Data
NA

Cloud platform

PerPos http://galileo-platform.com/themes/Pages/Software-platform.aspx
PerPos is implemented using the OSGi framework

Cloud performance

Three different domains use case are described:
- Mission Critical Situation Awareness
- Livestock Behavior Awareness
- Indoor Navigation

Issues/Gaps
- New features of the European Galileo Satellite system will be integrated into the platform when Galileo becomes operational in a few years

Up in the air: adventures in serving geospatial data using open source software and the cloud

Summary
- Described a solution for GIS data posting and sharing on the web using open source map servers and open source GIS clients deployed on a virtual cloud server
- The solution based on cloud server is inexpensive and simple

Workflow
Data
NA

Cloud platform
- Cloud server: Amazon EC2
- Map server: GeoServer
- Map client: OpenLayers

Cloud performance
- Cost is low if the needs are to simply serve map data on the web
- Amazon EC2 is very reliable
- Scalability and ease of maintenance

Issues/Gaps
NA

General Review Papers
1. Geospatial Cloud Computing

Individual stories
2. Huang, Q. et al. HPDGIS (2010)