Design and Implementation of a GIS in the Cloud

George Chang
NASA/Jet Propulsion Laboratory

Use Cases

Hybrid Cloud Design for a GIS Portal
Hadoop and Cloud Computing for Image Tiling
Lunar Mapping and Modeling Portal

Aggregates lunar data from completed and current missions for use by scientists and the general public on consumer computers and devices.

Collaboration between various NASA centers and other government institutions.

Utilizes open standards to facilitate platform and application independent access.

Hybrid Cloud Architecture
## Data Products

<table>
<thead>
<tr>
<th>Data Products</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nomenclature</td>
<td>Gravity</td>
</tr>
<tr>
<td>Imagery</td>
<td>Temperature</td>
</tr>
<tr>
<td>Mineralogy</td>
<td>Rock &amp; Craters</td>
</tr>
<tr>
<td>Temperature</td>
<td>Slopes</td>
</tr>
<tr>
<td>Rock &amp; Craters</td>
<td>Contours</td>
</tr>
<tr>
<td>Rock &amp; Craters</td>
<td>Elevation</td>
</tr>
</tbody>
</table>

## Architectural Principles

<table>
<thead>
<tr>
<th>Principles</th>
<th>Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability</td>
<td>Tiered Architecture</td>
</tr>
<tr>
<td>Scalability</td>
<td>Service Oriented</td>
</tr>
<tr>
<td>Maintainability</td>
<td>Seamless Security</td>
</tr>
<tr>
<td>Open</td>
<td>Cloud-enabled</td>
</tr>
<tr>
<td>Data driven</td>
<td></td>
</tr>
<tr>
<td>Standardization</td>
<td></td>
</tr>
<tr>
<td>Reliability</td>
<td></td>
</tr>
<tr>
<td>Security</td>
<td></td>
</tr>
</tbody>
</table>
Data Clients

Available on multiple user interfaces
Loose coupling between UI and data services
Same URL for Cloud or Privately Hosted Data
**Security**

Multiple LDAP Sources
Exposes REST Wrappers
Single Sign-On Across All Domains

---

**Service Application**

GIS information provided by standard APIs, XML metacatalog for GIS sources

All features invoked via REST

https://dev.lmmp.nasa.gov/LMMP/rest/transform/translation/subset/stream/png?src=9a05d86b-ffa4-4d5d-be54-9a8e90b27bce&ulx=43.4268&uly=-29.7554&lrx=43.8398&lry=-30.0454

WebDAV, SFTP access to User Storage
**Data**

GIS Servers – JPL (T)WMS, ArcGIS

Data Converters – GDAL, Java ImageIO

Search Engine, Catalog – Solr, OODT, BerkeleyDBXML

Workflow – Jabber, Amazon SNS/SQS

Repository – Netapp, Amazon EBS/S3

**Key Takeaways**

LMMP system design allows for modularized component design, vendor independence

Utilizes cloud infrastructure where appropriate, make access invisible to the user

Security model transparently grants access to data and protects sensitive products
Image Tiling on the Cloud

The Challenge

How do we make these large images usable by desktop computers, mobile devices, and other memory constrained products?
Image Tiling Process

Starting from the full resolution image, divide images into small tiles.

Combine and shrink tiles to create tiles for the next zoom level

Continue process until the final zoom level has only one tile

Hadoop

How We Use It

*Map Function* – Takes image tiles and adds metadata about its position as well as scaling them for the next zoom level

*Reduce Function* – Collects the 4 adjacent tiles and merges them to create a new tile
Iterate several times and we’re done!
Cloud Implementation Using Amazon EC2

Test image, 2.77 gigabytes LRO LOLA (Lunar Orbiter Laser Altimeter) colorized digital elevation map which produced 9.1 gigabytes set of tiles.

Amazon EC2 is a cloud computing infrastructure allowing users to “rent” virtual machines. Tested the “Large” and “Cluster Compute” instances.

Installed Hadoop framework on a number of EC2 instances.

Output image files stored on Amazon S3, a cloud storage system.

Test Results

Performance Times (circa 2011)

20 EC2 “Large”
20 EC2 Large Instances (4 Compute Units ~ 4x1GHz Xeon)
7.5 GB RAM
860 GB Storage

20 EC2 Large

4 EC2 “CC”
4 EC2 Cluster Compute Instances (33.5 Compute Units)
Gigabit Interconnects
23 GB RAM
1.69 TB Storage

4 EC2 CC

Upload Time
Storage Time
Processing Time

Minutes

0 22 44 66 88 110

58
58
53
34
42
105

Copyright 2012 California Institute of Technology. Government sponsorship acknowledged.
Key Takeaways

Cloud computing provides an easily quantifiable cost on data product generation

Choice of cloud services dependent on the specific use

Hadoop framework provides a simple programmatic interface for developing distributed computing applications for problems that are parallelizable

Thank You!

http://www.lmmp.nasa.gov/

George Chang | @gchang | gchang@jpl.nasa.gov