Advancing Geographic Place-Based Understanding for Decision-Making

FGDC GeoCloud Incubator

Initiative Outcomes, Lessons Learned and Best Practices

Eldrich “Rich” Frazier
Technical Advisor
Federal Geographic Data Committee
July 12, 2016
Background on GeoCloud Sandbox Initiative

- Initiated as an Architecture and Technology Working Group activity in December 2009
- Five-year project to test and monitor externally-hosted Cloud data and service solutions for the geospatial domain, to support the Geospatial Platform activity
- Initiative Sponsors:

  FGDC.GOV
  FEDERAL GEOGRAPHIC DATA COMMITTEE
  GSA
  U.S. DEPARTMENT OF HOMELAND SECURITY
Cloud Community Platform Initiative – Mission Needs

Business Drivers
- Growing pool of identified agency applications seeking cloud benefits
- Quantify savings, identify risks, ease migration

Platforms
- Construct a discrete subset of secure supporting platforms
- Develop processes for rapid application migration
- Leverage certifications and platform commonalities
- Develop chargeback mechanisms for self-sustaining funding

Best Practices
- Capture lessons learned
- Document techniques, issues and solutions
- Document cost models and benchmarks
- Build a cloud platform support community
- Maintain platforms, scripts and updates for client agencies

Results
GeoCloud Programmatic Goals, Objectives, Activities and Outcomes

Architectures
- Identify requirements-driven solution architectures and platforms for various sized deployments of geospatial data and services

Cost Models
- Document and assess cost models to support scalability, reliability, and redundancy

Certification
- Expedite FISMA (security) certification and accreditation for agency adoption of packaged solution architectures

Comparisons
- Support and collect cost comparison information from agencies for existing and externally-hosted Cloud solutions
  - Document lessons learned and best practices

Acquire, compose, document, and deploy reference platform cloud that support Geospatial Platform standards

Monitor costs, loads, issues and options in support of OMB IT project document guidance

Certify Geospatial Solution Packages to facilitate re-use

Document Best Practices and guides to agencies on adoption of geospatial Cloud infrastructure
GeoCloud Outcomes – Development of a Cloud Services Framework

**Service Architecture**

- **Basic Image**
  - Windows Server Image.NET, IIS
  - Windows Platforms
  - Open Source Platforms
    - Base Linux 64 (of unknown provenance)
- **Base Platforms**
  - Open Source Additions
    - Java, PHP, PostgreSQL
    - Tomcat, FLEX, Apache
- **Specialized Platforms**
  - ESRI ArcGIS Geospatial Platform
  - Additional Open Source
    - GlassFish3, Axis, SunMQ
  - Additional Geospatial
    - THREDDS, More

**Roles & Responsibilities**

- **Deployers**: Assemble Platforms and Apps
  - Install and harden the OS
  - Install and configure Platform Enablers (Postgres, Java, ...)
  - Rapid Iteration, Shared Change Code, No Cloud Mgmt.
- **Managers**: Manage Applications on Platforms
  - Launch and stop project applications
  - Monitor Usage and Issues
  - View, Transform, Agency-Change Calls, Move Apps.

**ConOps**

- **GeoCloud Project**
  - GeoCloud Project
  - GeoCloud Scripts
    - CloudFormation Template
    - Installation Script
  - Centrally Maintained Platform
    - OSS Platform
    - ArcGIS Platform

**FGDC Platform**

- **Maintainers**
  - FGDC Platform
    - Create, Compose, Harden Platform
    - Document and Script
    - Research and Layer Enablers
    - Security Update
    - Release
    - Update Platform

- **Consumers**
  - FGDC Platform
    - Gather Changes
    - Configuration Script
    - Application Code Base
    - redeploy
    - execute a launch run the project application on the platform

**Operating GeoCloud Project**

- Platform Updates
  - Project developers configure project to install on the platform
  - Platform maintainers create and periodically update the platform
GeoCloud Technical Goals and Objectives – Community Platform Sandbox

Goals

• Prototype/Pilot applications for use on a GeoCloud Platform
• Determine the application’s deployment and scalability requirements
• Move to production account, operate and document metrics and experience
• Graduate to full operation as appropriate

Objectives

• Identify and document the steps required to install the application
• Document any new application platform requirements and receive an enhanced platform from the platform development team
• Create and test application code bundles and installation scripts
• Test application for scalability and determine deployment strategy
• Migrate application to a production account and document
GeoCloud Community Platform Concept and Value

Pool of geospatial projects identified for cloud migration

FGDC App 1

FGDC App 2

8 More ...

GeoCloud Community Platform

Platforms Amplify IaaS Savings

Federal Platform Reference Implementations

Apps Need Platforms

GeoCloud Incubator

Infrastructure as A Service Savings

Hardware Savings Operations Savings Scalability Savings

Platform Savings

Platform building time & effort Reduced maintenance costs Faster deployment Cost effective development / test
GeoCloud Sandbox History

<table>
<thead>
<tr>
<th>Project Kickoff</th>
<th>Project Year 1 – General Purpose</th>
<th>Project Year 2 – Add ArcGIS</th>
<th>Project Year 3 – Focus on A16 Datasets</th>
<th>Final Project Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Build manual platform</td>
<td>• Build limited platform automation</td>
<td>• Increase GeoSpatial focus</td>
<td>• Revise and update OSS and ArcGIS platforms</td>
<td>• Introduce platform.mu with full automation</td>
</tr>
<tr>
<td>• Develop business processes</td>
<td>• Add OpenGeo to OSS platform</td>
<td>• Add GeoNode to OSS platform</td>
<td>• Initial CKAN prototype</td>
<td>• Add Ubuntu OS</td>
</tr>
<tr>
<td>• Build community portal</td>
<td>• 11 projects across 6 agencies</td>
<td>• Initial hardened ArcGIS platform</td>
<td>• 4 projects, 4 grandfathered A16 projects</td>
<td>• Migrate 2 new services – GeoNode and GeoShape</td>
</tr>
<tr>
<td>• Find agency candidates</td>
<td></td>
<td>• 15 agency projects across 10 agencies</td>
<td></td>
<td>• 6 new projects</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Beta Platform</th>
<th>Initial Automation</th>
<th>Add ArcGIS Platform</th>
<th>A16 Datasets</th>
<th>Full Automation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Kickoff</td>
<td>Project Year 1</td>
<td>Project Year 2</td>
<td>Project Year 3</td>
<td>Project Year 4</td>
</tr>
<tr>
<td>2010</td>
<td>2011</td>
<td>2012</td>
<td>2013</td>
<td>2014</td>
</tr>
</tbody>
</table>
GeoCloud Community Platform – Participating Organizations

- Federal
- State
- Academia
- Tribal
- International

*Sample Participating Organizations

Projects

<table>
<thead>
<tr>
<th>Year</th>
<th>Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
</tr>
<tr>
<td>Platform</td>
<td>Based On</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Wetlands Mapper</td>
<td>Windows/IIS/ESRI</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>National Elevation</td>
<td>Windows/TomCat/ORACLE/ESRI</td>
</tr>
<tr>
<td>Data Set</td>
<td></td>
</tr>
<tr>
<td>National Map</td>
<td>Windows/TomCat/ORACLE/ESRI</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>EPA Lakes &amp; Ponds</td>
<td>Windows/PostGres/ESRI</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic Wiki +</td>
<td>Linux</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>NOAA Particles</td>
<td>Linux/Java/Axis/MQ</td>
</tr>
<tr>
<td>GEOSS GeoNetwork w.</td>
<td>Linux/TomCat/PostGres</td>
</tr>
<tr>
<td>additions</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>TIGER Shapefiles</td>
<td>Linux / Apache / Perl/Python</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>IOOS Registry +</td>
<td>Linux/TomCat/PostGres</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>ERDAP</td>
<td>Linux/TomCat/THREDDS</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic Content Mgt -</td>
<td>Linux LAMP</td>
</tr>
<tr>
<td>Drupal 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>HHS Connect</td>
<td>Linux/EJB3 (GlassFish V3)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Emergency Web Response</td>
<td>Linux LAMP</td>
</tr>
<tr>
<td>Site</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>OpenStreetMap</td>
<td>Linux/MySQL/Ruby On Rails</td>
</tr>
</tbody>
</table>
GeoCloud – Pilot/Prototype Reports

National Center for Education Statistics

Amazon Hosted ESRI ArcGIS Servers
Project Final Report

NSDI Cooperative Agreements Program
Category 7: Geospatial Platform Cloud Service Test bed
Interim Report

Date: November 1, 2012
Agreement Number: G12AC20138
Project title: Ca Implementation
Organization: Indiana University
V. 317-274-2455
F. 317-278-1830
http://www.iupui.edu
Principle Investigator: David J Bodenham

Table 7. Cost comparison between AWS services and local server

<table>
<thead>
<tr>
<th></th>
<th>AWS EC2</th>
<th>Local Server</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed (one-time) cost</td>
<td>None</td>
<td>Around $2000 to purchase the computing instances</td>
</tr>
<tr>
<td>Network cost</td>
<td>$1 for data transferring / month</td>
<td>None (included in the maintenance fee)</td>
</tr>
<tr>
<td>Storage</td>
<td>$22 /month</td>
<td>None (Included in the first purchase)</td>
</tr>
<tr>
<td>Computing power</td>
<td>Not reserved</td>
<td>$255 / month</td>
</tr>
<tr>
<td>Reserved</td>
<td>$76 / month</td>
<td>$76 / month</td>
</tr>
<tr>
<td>Monthly maintenance cost</td>
<td>$200 / per month</td>
<td>(Assuming $100 for cooling, network, and room fee, and $100 for paying a sys admin to check and maintain the server)</td>
</tr>
<tr>
<td>Yearly Total</td>
<td>$1104</td>
<td>$2604</td>
</tr>
</tbody>
</table>

Note: The cloud computing cost is based on the standard large instance with 2 CPU cores of EC2 to host GEOSS clearinghouse. If more advanced instance types are selected, the computing power cost would increase.

Figure 1: AWS Console Monitoring Interface

Figure 1 August 2012 Indiana Harvest Parcels
GeoCloud Technical Services and Resources Tested

<table>
<thead>
<tr>
<th>GeoSpatial</th>
<th>Management</th>
<th>App Servers</th>
<th>Platform Tools</th>
<th>Operating Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeoShape*</td>
<td>Confluence**</td>
<td>Apache Tomcat</td>
<td>JAVA J2EE</td>
<td>CentOS</td>
</tr>
<tr>
<td>ArcGIS</td>
<td>JIRA</td>
<td>django*</td>
<td>Microsoft SQLServer</td>
<td>ubuntu</td>
</tr>
<tr>
<td>GeoServer</td>
<td>WIKI</td>
<td>nodeJS</td>
<td>ORACLE DATABASE</td>
<td>windows Server System</td>
</tr>
<tr>
<td>GeoNode*</td>
<td>Wordpress</td>
<td>RAILS</td>
<td>PostgreSQL</td>
<td></td>
</tr>
<tr>
<td>GeoNode*</td>
<td>Drupal</td>
<td>Jenkins</td>
<td>Clustner</td>
<td></td>
</tr>
<tr>
<td>splunk**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>splunk**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>splunk**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ckan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Examples
- EPA How’s My Waterway

EPA’s mobile app and website, How’s My Waterway, helps people find information on the condition of their local waterways using a smartphone, tablet, or desktop computer. It makes science-based water quality information accessible and understandable for everyone.

For decades, EPA has collected information on healthy and polluted waterways that states and territories report under the Clean Water Act. Tens of thousands of polluted waterways have been identified nationally, and much has been done to reduce pollution risks to people’s health, the economy, and the environment. However, many Americans care the most about their local lake or stream rather than the thousands of waterways described in national statistics. People want an easy way to learn about their local waterways, pollution problems, why they matter, and what’s being done to restore and protect them. Rather than sifting through baffling scientific information stored in complex databases, an average citizen might say, “All I really want to know is, how’s my waterway? And please tell me in words I understand.” EPA developed How’s My Waterway to provide answers quickly and clearly, anywhere at anytime.

**How It Works**

- **SEARCH:** Use a smartphone to find out about a lake, river, or stream while standing right at the water’s edge. Check any location in the US by entering a zip code or place name on your computer, smartphone or tablet.
- **RETRIEVE:** Instantly receive a list of waterways within about five miles of the search location. Each waterway is identified as unpolluted, polluted, or unassessed, along with the year its condition was reported. A map option offers a view of the search area with the waters color-coded by assessment status. Zoom in for more details or pan across the map to check on new areas and new water.
- **DISCOVER:** Once you select a specific waterway from the map or list of waters, the app and website offer more detailed results, including the type of pollution reported and what has been done by EPA and the states to reduce it. Technical users can follow links to detailed scientific assessment reports.
- **LEARN:** Read simple, non-technical descriptions of each type of water pollutant. These include what the pollutant is, where it comes from, how it can harm the environment, human health, or valuable economic uses of the waterway, and what you can do to help.
- **FIND:** Looking for even more? The related links page connects you to popular water information on beaches, drinking water, fish habitat projects, and more!

Visit How’s My Waterway at: www.epa.gov/mywaterway
For questions or comments, contact: http://water.epa.gov/contactus.cfm.
Examples

– USGS BISON Tool
Examples
– HUD and EPA

NEPAnode provides a suite of geospatial enriched shared web services, software, and GIS tools to assist practitioners, resource managers and decision makers in reporting and understanding environmental assessments and permitting.

The HUD Storefront unifies public access to HUD’s portfolio of applications and geospatial data resources, including the 11 National Geospatial Data Asset (NGDA) datasets that HUD is currently contributing to the Federal geospatial data portfolio.
• **Collaboration:** Community of Practice was a positive method for this effort

• **Cost Efficiency:** Cloud offers strong potential for operational savings

• **Architecture:** Application re-engineering generally necessary to take optimize use

• **Security:** Securing Cloud Services very challenging
GeoCloud – Recommendations

• Continue Collaborative R&D of Emerging Cloud Services
• Education and Awareness
• Share Lessons Learned
• Establish Resources via the GeoPlatform
• Security Architecture Development and Efficiency
Summary

Questions?

Rich Frazier
efrazier@usgs.gov