



**Amazon Hosted ESRI GeoPortal Server
GeoCloud Project Report**

Description of Application

Operating Organization

The USDA participated in the FY 2011 Federal Geographic Data Committee (FGDC) GeoCloud Sandbox Initiative by hosting the ESRI GeoPortal in the cloud environment.

Esri Geoportal Server is a free, open source product that enables discovery and use of geospatial resources including datasets, rasters, and web services. It helps organizations manage and publish metadata for their geospatial resources to let users discover and connect to those resources.

Our objectives for the project were to prove that the open source GeoPortal could be hosted in a cloud computing environment and successfully serve geographic data for end users. In the process we created a step by step installation manual for other agencies wanting to take advantage of cloud computing technologies coupled with open source software applications for serving geospatial data.

Community of Interest

The community of interest for this project was the USDA, NRCS, and the public GIS user community. The public GIS user community consists of private companies; students; researchers; tribal, state, and local governments, and the public at large.

Operating System and Software Requirements

A CentOS operating system (OS) and support for the following software was required:

GCC

C++

Apache Web Server

PostgreSQL

ESRI GeoPortal

Gedit

Operational Requirements

The USDA/NRCS normally hosts the ESRI GeoPortal application in-house with a single server housed at the Fort Worth Federal Center and is one of numerous systems managed by the USDA/NRCS staff. The in-house host server utilizes a Windows OS. The host has access to 60 GB of storage that is shared between

multiple applications and products besides the ESRI GeoPortal application. The ESRI GeoPortal takes 267 MB of storage space. The Geoportal is a very light weight application that is being run on a light weight virtual machine server.

The USDA host server has the following hardware configuration:

- **CPU:** there are four (4) processors with model name “AMD Opteron™ 2.41GHz” However, the Virtual Machine is only utilizing a single core.
- **Memory:** 2GB
- **Disc Space:** GeoPortal application resides on file system that has 60GB volume
- **Operating System:** Windows Server 2003 R2 operating system. This could easily be implemented on a Linux OS to replicate our effort in the cloud.

Deployment in the Cloud

The USDA/NRCS management and GeoCloud team determined that it was feasible to participate in GeoCloud Sandbox Initiative due to low risk considering security and business operations as well as significant overall benefits. Benefits included gaining experience with the cloud environment (OS development and hardening, system configuration, application deployment, and administration), and obtaining actual data on the performance, and reliability of cloud services.

The USDA/NRCS used an existing CentOS 5.5 image, ami-48bf4421 released for production 6/24/2011.

For our sandbox testing iterations of the OS build and patch and application installations, we used the micro instance.

The final AWS configuration for the GeoPortal data and download application was as follows

T1 Micro Instance

- 613 MB memory
- 2 EC2 Compute Units
- EBS storage only
- 64-bit platform
- I/O Performance: Low
- API name: t1.micro

The AWS environment could have been configured for auto scaling and load balancing. Auto-scaling could have provided the ability to add system-resources automatically as needed, based on threshold configuration for memory, CPU, and bandwidth. The USDA/NRCS did not use the capability because the demand for system-resource was not anticipated to fluctuate significantly over time. We were unsuccessful in implementing a startup script, as well as adding multiple users to the instance. We successfully generated new users and keypairs for those users on the CentOS instance, however, the instance did not respond as expected to the credentials for all users once new keypairs were generated. Once the GeoPortal was online, we were easily able to add numerous users with varying credential levels to the application. We were able to create ebs volumes, but due to security constraints, we were unable to boot from the volumes. Once we had configured the instance to our satisfaction, we were limited in not being able to create an image (EBS AMI) from the instance due to security credentials. This generated a fault code: Unauthorized Operation, and details explain the code stated "You are not authorized to perform this operation."

Operations in the Cloud

The GeoPortal application did not require maintenance once it was deployed on AWS platform. Our team used the instances to develop the application, configure multiple users, keypairs, and test startup scripts. The AWS platform provided the ability to expand available system resources if necessary, although it was not utilized. We had 2-3 instances in the cloud at a time, serving as development, test, and control environments. The control environment was being run for 24 hours per day, 7 days a week for a duration of approximately 5 months. At a constant charge of \$.02/hr. the total cost was \$72. The test and development environments were online approximately 75% of the time our control environment was, making the total running cost of both environments \$108. The USDA/NRCS incurred a grand total of \$180 in running costs during the project. The internet data transfer pricing sheet states the first Gigabyte going out from the site per month is free. We did not anticipate any more data than this being utilized in any given 30 day period.

Operational Cost Comparisons

The low monthly costs, and small size of the T1 micro instance made it the most attractive option for the USDA/NRCS. Had we continued the project, we would have down sized to two micro instances, and had them online for the entire year. This would make the annual operating costs for the two on demand instances approximately \$350. Due to our limited experience of Linux, and S3, there was a learning curve in time spent deploying the application. If we were to continue with the cloud hosted GeoPortal, we anticipate no more than two days spent on

implementation and configuration. Given that our internal operation costs and server deployment costs are confidential, we are unable to disclose the exact cost of what we anticipate spending on an internal deployment of the GeoPortal. We can say the Amazon hosted instance of the ESRI GeoPortal is cheaper option for the USDA/NRCS for the first few years of the cloud deployment. Eventually the annual cloud hosting costs would exceed the internal hosting costs given the hardware life expectancy rates.

Issues and Lessons Learned

Initially the USDA/NRCS had planned on having a Windows based instance running ArcGIS Server and GeoPortal working in unison. The GeoPortal would have been consuming services from ArcGIS Server, however, this never came to fruition. Towards the end of the project in mid November, a Windows 2008 R2 instance running ArcGIS 10 service pack 3 became available. We had already completed the installation and implementation of the ESRI Geoportal in the Amazon cloud. During the process we created an installation manual that would allow any Amazon Cloud user to compile an open source Geoportal following the exact same procedure we went through on our CentOS instance. The manual addresses the complete application installation. During the development process we were able to replicate the majority of the USDA/NRCS styling, layout, and content on our Amazon instance. We felt the customization of the final user interface was best left to the end user to suit individual business needs, thus we did not document the changing of the aesthetics.

We included the step by step procedure for the GeoPortal creation in the following section. The USDA/NRCS has stopped hosting the GeoPortal in the Amazon Cloud to prevent any unnecessary billing expenses. During the application installation, we found little to no documentation for the process online, proved an open source geospatial portal can be created in a cloud environment, and believe this reference is a good guide for other agencies wanting to follow our proof of concept.

The extreme low cost and small size of the Amazon instance make cloud hosting a better alternative to hosting the GeoPortal internally.

Installation Guide

The installation guide is available through the GeoCloud Community Portal.