IN21B-1726
Fall AGU Meeting
December 12, 2016
San Francisco, CA

Technical Challenges and Lessons from the Migration of the GLOBE Data and Information System to Utilize Cloud Computing Service

John F Moses[1], Nargess Memarsadeghi[1], David Overoye[2], Brian Littlefield[2]
[1] NASA Goddard Space Flight Center, Greenbelt, MD, United States,
[2] Science Systems and Applications, Inc., Pasadena, CA, United States

Global Learning and Observations to Benefit the Environment (GLOBE)

- Worldwide hands-on primary and secondary school-based science and education program
- Promotes and supports students, teachers and scientists to collaborate on inquiry-based investigations of the environment and Earth system
- Environmental measurements in Atmosphere/Climate, Hydrology, Land Cover/Biology, Soils, and Phenology
- Began Earth Day 1995 as partnership between NASA, NOAA, NSF
- Grown to more than 48,000 trained teachers, 24,000 schools, 1.5 million students

GLOBE Data and Information System (DIS)

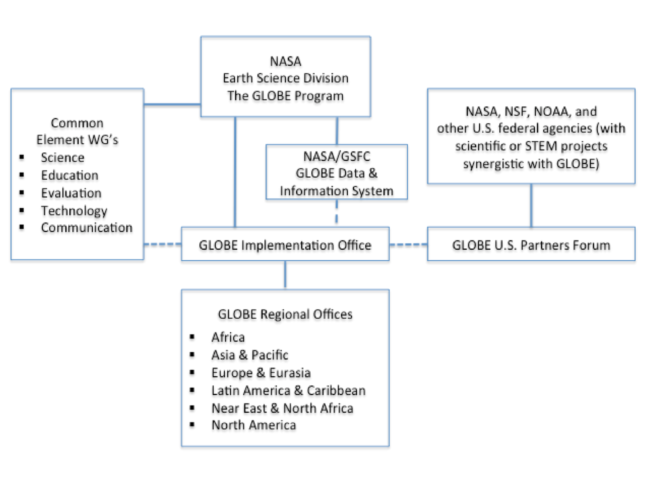
Top Functional Requirements

1. Provide training and material on the use of GLOBE, including measurement protocols
2. Receive, check, accept member observation entries compliant with protocols
3. Ingest data from automated weather stations and other data centers
4. Maintain records across protocols (over 127 Million since 1995)
5. Visualize observations on a map
6. Provide data analysis tools, graphs and data access tools via search and filtering
7. Export observations recorded across protocols
8. Enable data entry from mobile Apps
9. Recognize contributions of GLOBE students and schools
10. Administer workshops, science Blog, Help Desk, Email, Mass Mailing

GLOBE DIS Modernization

- In 2010, NASA Goddard was asked to lead the evolution of the GLOBE DIS to a new architecture; evolving software and internet technologies to achieve program goals.
- The Goddard GLOBE DIS Team was formed and partnered with UCAR to developed a new enterprise portal, web application framework, and modern visualization and graphing features
- Migration to new system completed July, 2012

GLOBE Team and Organization circa 2013



UCAR GLOBE System Development and Operations

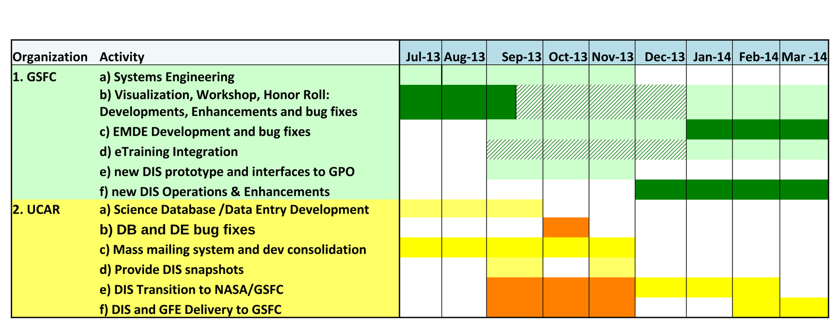
UCAR Role	Plan (FTE)
Task Lead/Architecture	1.0
Database Admin	1.0
Developer (Data Transfer)	1.0
Developer (Data Entry)	1.0
System Developer (Design)	1.0
Bug Triage and Test	1.0
Developer (Web app Data Entry)	1.0
System Admin	1.0
Systems Admin	0.5
Developer (Data Entry)	1.0
Software Engineer	1.0
Developer (Database)	0.5
total	11.0

GSFC GLOBE Software Engineering Services Task

GSFC Contract Role	Plan (FTE)
Developer (RubyOnRails)	0.9
Developer (Vis/GIS)	1.0
Developer (SysArch)	1.0
Developer (Liferay)	0.9
Developer	0.9
Bug Triage and Test	0.5
Environment/CM/Deployment	0.4
Ops Lead	1.0
Task Lead	0.4
System Admin	0.5
total	7.5

GLOBE Transition Objective, Tasks, Milestones

In 2013, NASA Goddard Science Data Systems was asked to transition the GLOBE DIS from UCAR to ensure its long-term stability, integrity and continued improvement.



Critical Milestones

- September:
  - UCAR provides snapshots of the as-is Data Entry protocols and database (2.d).
  - GSFC suspends enhancements, begins prototyping the new DIS (1.b, 1.c, 1.e).
- November:
  - UCAR completes implementation and bug fixes of high priority DE protocols in Ruby-on-Rails (2.a-b). New snapshots provided (2.d).
  - GSFC prepares the new DIS for the switchover (1.e). Migrates portal and non-Data Entry capabilities for test. Production readiness review verifies all equipment ready, all images functional. Pending final security review.
- February:
  - UCAR completes all transitions (2.e).
  - GSFC launches new environment (1.f).

www.nasa.gov

Infrastructure Architecture Trade, Estimates, Drivers

- The UCAR GLOBE DIS hardware was near end-of-life and a trade study was prepared to evaluate alternate infrastructures

480 Hours of estimated labor for both solutions involve similar tasks

Rack and Local Facility

- Initial Rack Configuration and Wiring
- Configuration of VMware and image transfer prep
- VM Snapshot transfer and reconfiguration
- Liferay configuration and tuning
- Initial Test and Validation
- Failover planning, testing and validation

Production & Staging	Year 1	Year 2	Year 3	Total
Hardware	$190,000	$5,000	$5000	$200,000
Software*	$71,000	$49,000	$49000	$169,000
Data Center	$16,400	$9,600	$9,600	$35,600
Installation labor	480 hours	N/A	N/A	480 hours
Total	$327,400	$63,600	$63,600	$454,600

- Pros
  - Ownership and control
  - Service Level Agreement claimed 100% availability
- Cons:
  - Recurring Data Center Costs`
  - Hardware/software maintenance and renewal costs
  - Needs additional backup
- Risks:
  - Small business: recovery, performance and history unknown

Amazon Web Services

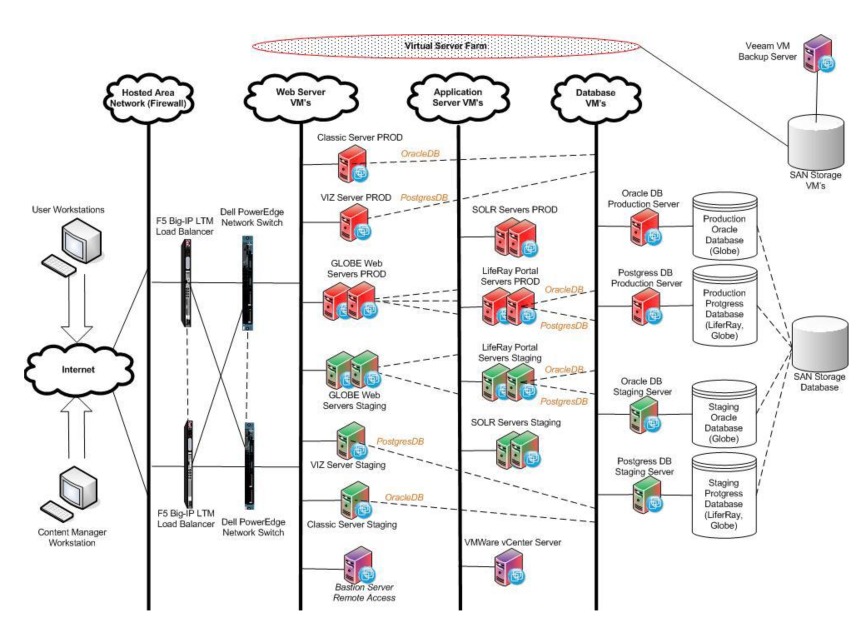
- Initial server instance request, setup & configuration
- Configuration of VMware and image transfer prep
- VM Snapshot transfer and reconfiguration
- Liferay configuration and tuning
- Initial Test and Validation
- Failover planning, testing and validation

Production & Staging	Year 1	Year 2	Year 3	Total
Hardware	$40,000	$40,000	$40,000	$120,000
Software	$43,000	$43,000	$43,000	$129,000
Data Center	$0	$0	$0	$0
Installation labor	480 hours	N/A	N/A	480 hours
Total	$133,000	$83,000	$83,000	$299,000

- Pros
  - No data center costs ($10k/year)
  - No hardware refresh costs (~$150K)
  - No VMware license, no F5 support costs ($11K/year).
  - Annual availability (Service Level Agreement): 99.95%
- Cons
  - Not one accessible infrastructure
  - Shared/distributed data ownership/security
- Risks
  - Not able to enforce/comply with new government/NASA policies
  - Outages beyond control of GLOBE System Administrator

GLOBE Data and Information System

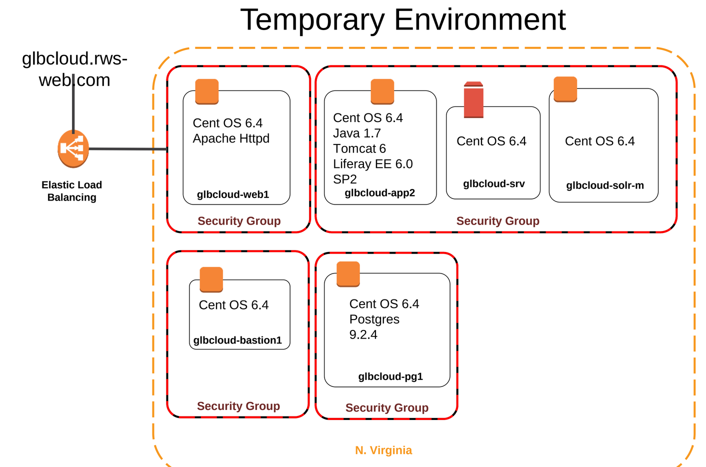
Architecture Before Transition to GSFC



- Bandwidth In: 600 GB/month
  - Full system (files only) 150Gb + 10% nightly incremental backup (15Gb/night) + 50 Gb/month for user interaction (data entry, collaborations, uploads)
- Bandwidth Out: 600 GB/month
  - 120,000 users per month for page sizes of 2-5 MB based on Google Analytics
- Data Storage: 150 GB
  - Postgres Database: 125 GB
  - Liferay files: 15 GB
  - Apache cached files: 20 MB
  - Data storage: 10 GB
- Backup size
  - One full backup: 150 GB
  - 2 full backups + 6 months of incremental backups <= 3.0 TB
- Backup frequency
  - Nightly, with increments of less than 10%
- Rack Power: 20 amp at 120v

Development (build & Test) Path

Temporary Environment



- Phase 1 – Amazon Single Instance
  - Liferay runs on CentOS 6.4, Java 1.7, clean startup
  - Content migration from production snapshot
  - Liferay runs with Postgres 9.2.4, index functions
- Phase 2 – Load Balancing & Performance Tuning
  - Liferay App Server – shared file storage
  - Liferay App Server Clustering, Load balanced
  - HTTP Servers load balanced
  - SOLR Master & Slave setup, index test
  - Postgres VIZ Server – configure and setup
  - Grinder setup for external load testing
- Phase 3 – Load Testing and Performance Tuning:
  - 2 external agents, 50 worker processors, 100 runs
  - Review security, verify failover, backup and restore
- Approval to purchase Amazon Production & Staging Environments

Data Entry/Ruby Failover Test:

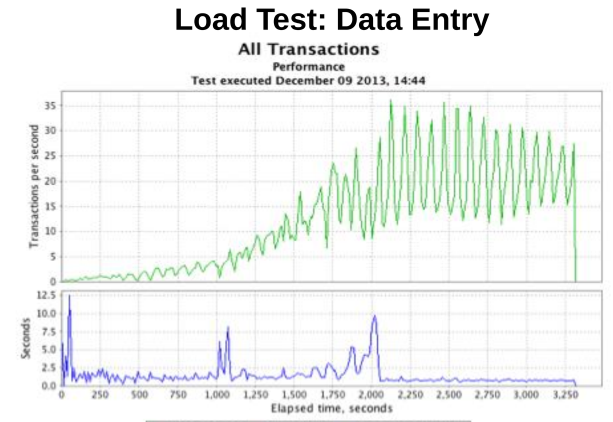
1. Two Ruby Servers running under load balancers
2. Stopped Ruby 2 Application Server
3. Data Entry continued to function on Ruby 1
4. Started Ruby 2 back up
5. Data Entry continue to function as normal
6. Stopped Ruby 1 Application Server
7. Data Entry continued to function on Ruby 2

Performance & Load Test Results

Adjustments and Adaptations

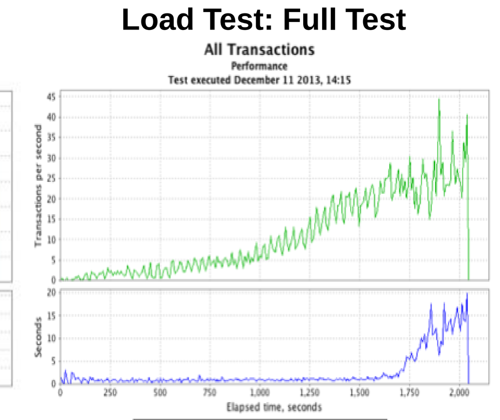
- Some VIZ system queries were re-written to leverage a new table to prevent the need to join multiple times.
  - improved response from 2 minutes to less than 15 seconds.
- Discovered areas to improve database performance.
- Applied software performance improvements in Liferay code in managing connections to the database.
- Added autoscaling to Apache, Data Entry, Vis.
  - Autoscale not possible with Liferay because of licensing.
- Postgres CPU Utilization improved to handle more users.
- Upgrade Postgres to next larger size.
- Upgrade IOPS from 2K to 4K.
- Upgrade Ruby servers to next larger size.
- Eliminate 2 training ruby servers.

Load Test: Data Entry



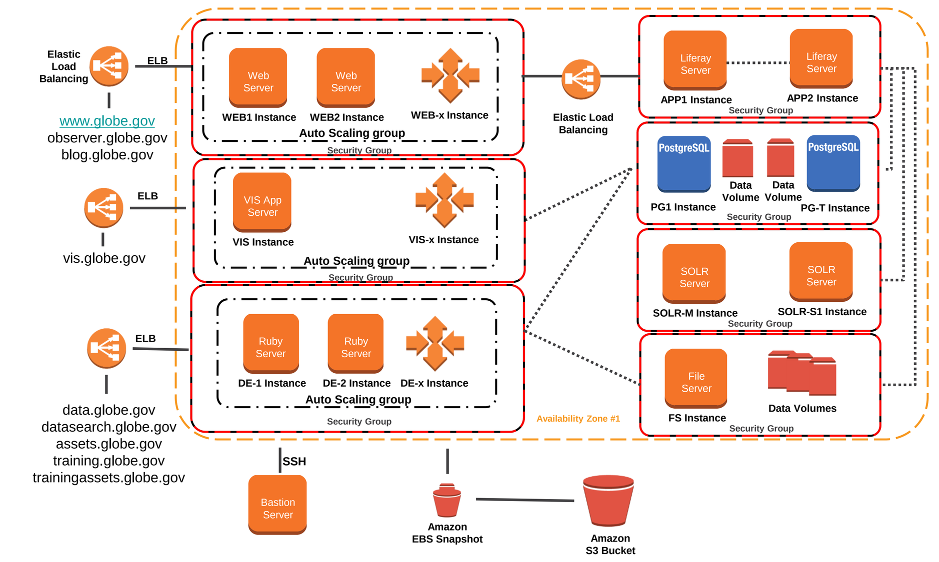
- 200 Concurrent Users.
- System stressed under heavy load at 150 users.
- Servers still function.

Load Test: Full Test



- 500 Concurrent Users.
- Ruby Servers show same results when multiple data entries are simulated.
- System still operates under normal parameters.
- Some deterioration as full load is reached.

Final Production Architecture



Name	Instance Type	Application
glbprod-app1	m3.xlarge	Liferay 6.2 w/Tomcat
glbprod-app2	m3.xlarge	Liferay 6.2 w/Tomcat
glbprod-bastion	m1.small	Bastion SSH
glbprod-pg1	m3.2xlarge	PostgreSQL
glbprod-pgt	m3.2xlarge	PostgreSQL Training
glbprod-pgmon	m3.medium	PGbadger/Pgclu monitoring
glbprod-ruby1	m3.large	Ruby Passenger w/Apache
glbprod-rubyX	m3.large	Ruby Passenger w/Apache
glbprod-solr	m3.large	Apache SOLR w/Tomcat
glbprod-solr-s1	m3.large	Apache SOLR w/Tomcat
glbprod-srv	m3.xlarge	File Server / ADAT
glbprod-vis	m3.large	Post GIS w/Tomcat
glbprod-web1	m3.medium	Apache HTTP
glbprod-webX	m3.medium	Apache HTTP

Launch Readiness Checklist and Responsibilities

	Task	Date	X
1	Complete Staging Environment Setup	1/14/2014	X
2	Lock Down Cloud Environment	1/15/2014	X
3	Get Detailed List of cronjobs	1/16/2014	X
4	Identify deltas to prepare for Cloud Staging	1/16/2014	X
5	Notify Users of Scheduled Downtime	1/17/2014	X
6	Development Environment Setup	1/20/2014	
7	Freeze all new development on Cloud and Migration	1/24/2014	
8	Migration Fire Drill on Cloud	1/27/2014	
9	Validate Cloud system and Load Test (Grinder)	1/29/2014	
10	Backup of Colorado Production Postgres Database	1/30/2014	
11	Shutdown Production Cloud Servers	1/30/2014	
12	Replication to Development System	1/31/2014	
13	Suspend all cronjobs on PostGres and Application Servers	2/3/2014	
14	Disable Login to Liferay on GLOBE.gov	2/3/2014	
15	*Enable Cloud Weather Bug Collection Servers (Cron)	2/3/2014	
16	*Final Oracle to Postgres Data Migration	2/4/2014	
17	*Migrate Final Assets to the Cloud	2/4/2014	
18	*Migrate Postgres Database to Cloud Postgres Databases	2/4/2014	
19	*Copy and Migrate over Liferay Data Folder to the Cloud	2/4/2014	
20	*Backup and Image Full Cloud System	2/5/2014	
21	*Application Server Configuration Updates	2/5/2014	
22	*Restart both App Servers	2/5/2014	
23	*Re-Index Solr Indexes	2/5/2014	
24	*Validate System is Up and Working	2/6/2014	
25	*Data Entry verification	2/6/2014	
26	*Enable Cloud Email Data Entry Collection Servers (Cron)	2/6/2014	
27	*Test and Validate	2/6/2014	
28	Go / No Go Decision for DNS Change Over	2/7/2014	

Role	Go / No Go
Chief Tester	
Project Lead	
Data Entry / Database	
Weather Bug / Liferay	
Vis System	
System Administrator	
Data Entry / Rails	
UCAR Tech Lead	
Liferay Technical Lead	
Program Manager	
Chief Scientist	
GLOBE Manager	

- Full team review of environment, sizing and approach
- Begin purchase – verifying ability to "reflag" environment as long-term.
- Step 1 – "Reflag" current environment as Production
- Step 2 – Move new production AMI's to a similar staging
  - All servers in production are re-created in staging
  - All(verify) servers are one size smaller in staging that production
- Step 3 – verify backups
- Step 4 – quick functional tests

GLOBE went live on February 7th, 2014

Amazon Specific Findings

IPV6

- Classic Elastic Load Balancers (ELB) only; no support for newer Virtual Private Cloud architecture.

Load balancer properties are set by Amazon

- Want to extend timeouts to handle long queries.
- Customizing limited – modifiable by Amazon.
- DNS domain root requires static IP, Amazon ELB IP may change at any time – impact to hard-coded social media external users; difficult to trace.
- HTTP://globe.gov gets redirected to www.globe.gov however HTTPS://globe.gov does not, requires separate server redirect.

Hardware is retired occasionally at any time

- 1 to 2 week notice, takes a few hours.
- Requires recreating instances based on image.
- Includes IP changes and DNS changes.
- Database instance takes longer because of size.

Amazon has default limits dependent on service level

- Number of volumes, servers, IP addresses.
- Not always aware you're up against a limit.

Availability and Uptime

No quarterly 1 day power cycle outage (previous UCAR facility)
Homepage uptime including planned (code migration) outages

- Jul 2015 – Aug 2016 (13.5 months: 99.9996% uptime
- Feb 2016 – Aug 2016 (7 months) : 99.9999% uptime

Aug 2015- Aug 2016 (12 months): 1 server crash not env related

Server Instance Maintenance

10-Feb-14	-	25-Feb-14	Instance scheduled retirement	staging web server
10-Feb-14	-	26-Feb-14	Instance scheduled retirement	PROD ruby server
6-Oct-14	-	24-Oct-14	Instance Maintenance down for 2 hours and rebooted	PROD ruby2 server
9-Oct-14	-	24-Oct-14	Instance scheduled retirement	PROD ruby server
9-Nov-14	-	17-Nov-14	Instance Maintenance down for 2 hours and rebooted	staging server
27-Nov-14	-	2-Dec-14	Instance Maintenance down for 2 hours and rebooted	staging server
3-Dec-14	-	8-Dec-14	Instance Stop Required	PROD ruby2 server
8-Apr-15	-	23-Apr-15	Instance scheduled retirement	staging server
22-Apr-15	-	29-Apr-15	Instance Maintenance down for 2 hours and rebooted	PROD bastion server
1-Jun-15	-	17-Jun-15	Instance scheduled retirement	staging web server
29-Apr-16	-	14-May-16	Instance scheduled retirement	PROD postgresql server
5-Jun-16	-	20-Jun-16	Instance scheduled retirement	staging web server
29-Jun-16			SES Bounce Email Probation (organization change)	
12-Jul-16	-	26-Jul-16	Instance scheduled retirement	PROD bastion server
14-Feb-14			EC2 instance limit reached	increase requested
18-Feb-14			EC2 elastic ip limit reached	increase requested
30-Oct-15			EC2 elastic ip limit reached	increase requested
12-Feb-16			EC2 instance limit reached	increase requested

Changes from the Original Plan

Labor Hours Revised

- 30% over original labor estimate (480 hours grew to 625).
  - System upgrades, additional load tests, AMI from prototype to production.

Changes to System Environment:

- Next larger instances for database, application servers in production, smaller sizes in staging (net +1.5K/year).
- Guarantee bandwidth for key instances (database, app servers) via EBS Volume Optimized (+$4.7K/year).
- Add DB & Ruby instances dedicated to training (+5K/year).
- Weekly backup data center transfers ($600/year).
- Estimate 2 months of prototype environment ("pay by the drink") prior to final production environment commitment ($10K).
- Cost changes over time reduced annually 3-10% depending on service, typically come with increase in capabilities. e.g., M3 server cost more than M4.

Staff Skills & Level

- UCAR support ended Feb 2014.
- No significant change in GSFC staffing due to transition.

Skills change in transition to GSFC

- Developers
  - Maintain knowledge of AWS Services.
  - Focus on Postgres and Liferay performance.
  - Focus on Ruby-on-Rails and Data Entry.
- System Administration
  - DNS service through GSFC.

Skills change in transition to AWS

- System Administration
  - Respond to Amazon instance retirements.
  - Monitor and analyze instance performance.
  - Respond to IP and instance limits.
  - Plan architecture service changes in anticipation of growing loads.

Sponsor: NASA Science Mission Directorate