Cloud onboarding with NGAP

Cloud Onboarding Session
ESIP Summer Meeting 2017
11am July 27th 2016

Doug Newman (NASA Earthdata – Raytheon)

The material is based upon work supported by the National Aeronautics and Space Administration under Contract Number NNG15HZ39C
Agenda

• Three important things
• How things do (and don’t) change
  – Deployment
  – Configuration
  – Security
  – External communications
  – Scaling
  – Logging
  – Monitoring
  – Metrics
  – Contingency and Recovery
• How did it go with the Common Metadata Repository?
THREE IMPORTANT THINGS
1. What is NGAP?

• Next Generation Application Platform
• NGAP is the NASA Compliant General Application Platform. It provides a cloud-based Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS) for EOSDIS applications.
Earth Science Instruments on ISS:
RapidScat, CATS, LIS, SAGE III (on ISS), TSIS-1, OCO-3, ECOSTRESS, GEDI, CLARREO-PF
12 Discipline Oriented DAACs

ASF SDC
SAR Products,
Sea Ice,
Polar Processes,
Geophysics

LP DAAC
Surface
Reflectance,
Land Cover,
Vegetation Indices

GES DISC
Global Precipitation,
Solar Irradiance,
Atmospheric Composition
and Dynamics,
Global Modeling

PO DAAC
Gravity, Sea Surface
Temperature, Ocean
Winds, Topography,
Circulation & Currents

NSIDC DAAC
Snow and Ice,
Cryosphere,
Climate Interactions, Sea Ice

CDDIS
Space Geodesy,
Solid Earth

LaRC ASDC
Radiation Budget,
Clouds, Aerosols,
Tropospheric Chemistry

ORNL DAAC
Biogeochemical
Dynamics, Ecological Data,
Environmental Processes

SEDAC
Human Interactions,
Land Use,
Environmental Sustainability,
Geospatial Data

MODAPS/ LAADS
MODIS Level-1 and Atmosphere Data Products

www.nasa.gov
EOSDIS Archive Growth Estimate (Prime + Extended)

Lots of assumptions in this chart. Subject to change...
ExCEL Efforts and Project Prototypes

**NGAP**
NASA Compliant General Application Platform (NGAP), an operational, dev-ops, and sandbox AWS cloud based operating environment.

**ASF WOS Prototype**
AWS/NGAP Web Object Storage (WOS) prototyping large volumes of mission data dynamically between AWS S3, S3-IA, and Glacier object storage. Managed out of Alaska Satellite Facility

**Earthdata Search Client to Cloud**
NASA Earth Science data search by keyword and advanced filters such as time and space

**Cumulus**
Prototype addressing core EOSDIS capabilities including data ingest, archive, management, and distribution of large volumes of EOS data.

**Getting Ready for NISAR (GRFN)**
Integrated prototype of science product generation and delivery from a DAAC system focused on coupling ASF DAAC and JPL ARIA systems.

**CATEES**
Easy-to-use Python tools packaged to support EOSDIS cross-DAAC science workflows and analytics over large volumes of EOS data in AWS.

**ECC to Cloud Study**
Earth Code Collaborative (ECC) study to determine cloud ready capabilities to migrate into AWS/NGAP platform.
ExCEL Efforts and Project Prototypes Continued

**GIBS in the Cloud**
Migrating GIBS to the AWS/NGAP Cloud based on recommendations made in the “GIBS in the Cloud Study”

**Earthdata Login to Cloud Study**
Study to determine and recommend migrating the Earthdata Login into AWS/NGAP cloud environment

**CMR to Cloud**
Migration of the Common Metadata Repository, into the AWS/NGAP platform based on recommendations made in the CMR to Cloud study.

**OPeNDAP/HDF Cloud Studies**
Study to determine and recommend a cloud native integration of OPeNDAP accessing HDF5 and netCDF4 data on AWS/NGAP platform.

**NEXUS**
Prototype to accelerate end-user analysis of remote sensing data, highly parallel to better enable science discovery

**Network Prototypes**
Network prototypes to support to test security, monitoring, logging, and to perform R&D testing to support all ExCEL project prototypes.
ExCEL Go/No-Go

(01) Full Scale Deployment (Q)

Full scale enterprise deployment of EOSDIS services and infrastructure to the cloud

(02) Partial Deployment (Q)

Select deployment of EOSDIS services and/or infrastructure to the cloud

(03) Cloud Stand-down (Q)

No EOSDIS services or infrastructure operationally migrated to the cloud

(04) Decision Point (Q)

More prototyping required, or cloud hybrid, or other next steps based on ExCEL prototyping and business analysis results

Determining Project Success

Project success is determined by viable outcomes of fully completed project prototypes and business analysis.

- or -

Technical and business results of the ExCEL project needed for strategic decision on EOSDIS and the cloud.
NGAP as a Platform

NGAP Services
(Monitoring, Logging, Security, Autoscaling, Billing, etc.)

NASA’s Office of the Chief Information Officer
(AWS Reseller)
A Rough Look at Separation

Policy
- Budgeting
- Security
- Usage

Technology
- Hosting
- Storage
- Services

NGAP Services

OCIO GP-MCE
NGAP as a Platform

<table>
<thead>
<tr>
<th>NGAP Services</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Monitoring, Logging, Security, Autoscaling, Billing, etc.)</td>
<td></td>
</tr>
<tr>
<td>NASA’s Office of the Chief Information Officer</td>
<td></td>
</tr>
<tr>
<td>(AWS Reseller)</td>
<td></td>
</tr>
</tbody>
</table>
Layer security throughout the architecture

- ESDIS “blessed” component

Usable cloud “platform”

NASA Office of the CIO (AWS Reseller)

NGAP Services (Monitoring, Logging, Security, Autoscaling, Billing, etc.)

Application

Application Builder

NGAP Secured Components

NASA Development Infrastructure

App Source Code

Application

EOSDIS
2. Instances are ephemeral

NGAP deployments follow a blue-green deployment process.

To maximize the availability and performance of our applications, a deployment is spun up in parallel with the existing deployment. When the secondary deployment is ready it swaps with the existing deployment which is then discarded.

NGAP application instances are not available in perpetuity.
3. No ssh

To preserve the integrity of an application instance, ssh (secure shell) access is limited to NGAP personnel
HOW THINGS DO (AND DON’T) CHANGE
Deployment

**Bamboo** is used to perform deployments

Production and UAT deployments are tightly controlled by the **DEVOPS** team

SIT deployments are controlled by the **development** team

Earthdata Operations maintain a **Deployment Doctrine** that is publicly available
Configuration

12-factor-app practices encourage the storage of configuration with the environment.

We developed the Earthdata Environment Configuration Service (EECS) to configure our applications.

EECS provides an API to read and write JSON-formatted configuration for our application on a per-environment basis.

If an implementer chooses not to use EECS then configuration should be externalized from code.
Security (1 of 2)

The responsibility for identifying and resolving security issues and software patches rests with the GP-MCE.

They will release Amazon Machine Instances (AMIs) to NGAP.

NGAP will release that AMI to NGAP PROD after SIT and UAT testing.

The application team will deploy the new AMI with any deployment of their applications that exist in NGAP PROD.

This approach has a number of elements that need to be allowed for.
Security (2 of 2)

1. Not all applications have a presence in NGAP SIT and UAT
2. Once an AMI hits NGAP PROD all deployments there will use the new AMI*

*We plan to mitigate this by giving an operator choices in AMI at certain points
External communications

On-premises solutions generally have a static set of IP addresses that an external entity can expect traffic from.

NGAP instances are ephemeral.

NGAP applications have a range of possible IP addresses.

Stick to standard ports if possible. Amazon Web Services (AWS)/GP-MCE/NGAP do not block outgoing traffic to standard ports.
Manual scaling is extremely simple to achieve via the ngap-cli application.

> bundle exec ngap ps:scale <app name> 2
Logging (1 of 2)

No ssh access. No log files.

NGAP automatically generates all needed artifacts to analyze application and access logs with **Splunk**

Use Splunk
Logging (2 of 2)
External monitoring (1 of 2)

External monitoring strategy will not be affected by the transition to NGAP

Monitoring of public APIs and applications do not change
External monitoring (2 of 2)

A screenshot of a web interface showing external monitoring status. The table displays monitors with their types, last check times, checkpoints, total times, and status indicators for active and alerting active.
Internal monitoring strategies may be affected by the transition to NGAP.

No ssh access.

Monitoring must be done using one of the following methods,
Internal monitoring (2 of 3)

1. NGAP provides an API to obtain a list of instance IP addresses for our applications
2. NGAP generates metrics, alarms and notifications for our ephemeral instances
   – Disk utilization
   – CPU utilization
   – Memory utilization
3. Custom alarms may also add alarms to metrics associated with static AWS resources
Internal monitoring (3 of 3)
Metrics

Metrics can be obtained using the following applications aligned with NGAP,

• Splunk
• AWS CloudWatch*

And external applications such as,

• Google Analytics
• Uptrends

These can be leveraged during issue triage, reporting and performance analysis

*We are looking into piping AWS metrics into Splunk
Contingency & Recovery

• NGAP can currently deploy our applications to multiple availability zones (AZ) within the US-East region for Platform as a service (PaaS) applications
• In the future, we could support deployment across multiple regions (within CONUS)
• If one AZ goes down the other one is still there. Our applications keep working
• We expect to be able to leverage recovery capabilities provided by the cloud and NGAP
HOW DID IT GO WITH CMR?
CMR?

‘The Common Metadata Repository (CMR) is a high-performance, high-quality, continuously evolving metadata system that catalogs Earth Science data and associated service metadata records’

- 33K collections
- 380 million granules
- 95% of queries are resolved in less than 1 second
- 12 node elastic search cluster (1.4 TB) for search
- Oracle Relation Database Service (RDS) for metadata persistence
- 14 micro services
  - On premises – 5 hosts (1 instance on each)
  - NGAP – 42 application instances (varying numbers of redundancy)
Performance (1 of 2)

Collection Search Performance

- Collection 95th Percentile
- Collection 99th Percentile
- NGAP cutover
- Collection Search Count
Performance (2 of 2)
Stability

- 2017 Prod uptime on-premises: 99.70%
- 2017 Prod uptime on-cloud: 99.93%

- 2017 UAT uptime on-premises: 99.76%
- 2017 UAT uptime on-cloud: 99.95%

- 2017 SIT uptime on-premises: 96.76%
- 2017 SIT uptime on-cloud: 99.79%

Cut over to cloud - 041917
Scalability (1 of 2)

- New functionality in CMR has required the re-indexing of our granule inventory. This is a time-consuming process.
- While on-premises are only recourse was to intelligently distribute the load of re-indexing across our 5 instances.
- On the cloud we can, and have, spun up additional, temporary processing instances to reduce the time taken.
Scalability (2 of 2)

• Re-indexing granules on premises: 7 days
  – 5 workers

• Re-indexing granules on cloud: 3 days
  – 1 worker per provider (normally 5)
Miscellaneous

• CMR uses Uptrends for external monitoring
• CMR uses Uptrends, Google Analytics and Splunk for metrics
• CMR SIT, UAT and PROD are only deployed to NGAP PROD
This material is based upon work supported by the National Aeronautics and Space Administration under Contract Number NNG15HZ39C.