NASA's EOSDIS Cumulus

:::

Ingesting, Archiving, Managing, and Distributing Earth Science Data from the Commercial Cloud

Katie Baynes - Civil Servant, NASA Goddard
Rahul Ramachandran - Civil Servant, NASA Marshall
Dan Pilone, Patrick Quinn, Jason Gilman - Element 84, Inc
Ian Schuler, Alireza Jazayeri - Development Seed
EOSDIS in Context

Capture and Clean

Distribute

Transform*

Archive

Process

Research

Applications

Education

*Subset, reformat, reproject

satellite, airplane, compass icons by Nook Fulloption, database, transformation, decision, process, data mining, customer community by Becris all from thenounproject.com (CC 3.0)
EOSDIS Holdings and Projected Growth

Fiscal Year vs Petabytes

- Distributed Volume
- Archive Volume
- Archive Growth
Past 24 Months: Focused on evaluation and planning for a cloud migration in 4 areas

- Compliance, Security, Cost Tracking
- Core Archive Functionality and Processing
- End-User Application Migration
- Pursuing Cloud Partnerships
Past 24 Months: Focused on evaluation and planning for a cloud migration in 4 areas

- Compliance, Security, Cost Tracking
- Core Archive Functionality and Processing
- End-User Application Migration
- Pursuing Cloud Partnerships
Example Step Function for GIBS in the Cloud
Cumulus

Collaboration
- Data Streams and Onboarding
  - Metrics Prototyping (GHRC)
  - Analytics
  - OPeNDAP
  - OGC
  - SDPS Tools
  - SDPS Requirements Review (EED2)

- Services Working Group (EED2)
  - SDPS Working Group (LPDAAC)

Core Development & Deployment
- Technical Debt
- Performance Testing (DevSeed and EED2)
- GHRC
- NGAP On-boarding

Operational Transition Activities
- Decision Making Framework Document
- GIBS and ASTER
- Operator Dashboard (DevSeed)
- Forrester

Cost Modeling
- PODAAC
- GITC
- Code Governance (ESDIS)

Documentary
- Container Registry
- Documentation

Code Governance (ESDIS)
- PodAAC
- ASF

Disaster Recovery
- On Prem Archive

Lock-In Avoidance
- Preservation Working Group (LPDAAC)

Lock-In Avoidance
Looking at the Past 18 Months

- **Cumulus Kick-off**
  - June 23, 2016

- **Phase 1 - GHRC Complete**
  - October 18, 2016

- **Phase 2 - LPDAAC Complete**
  - May 12, 2017

- **Phase 3 - PO.DAAC Complete**
  - September 29, 2017

- **Phase 4 End Date**
  - December 22, 2017

- **Open Source approval**
  - Today

- **Scaled Agile Begins**
  - Actual Open Sourcing of Code Base
Cumulus Phased Methodology

1. Determine objectives
2. Identify and resolve risks
3. Development and Test
4. Plan the next iteration

Cumulative cost
Progress

Review
Implementation
Release

175x289
79x343
416x110
416x103
540x13
692x19

https://en.wikipedia.org/wiki/Spiral_model
So, what does this mean in practical terms for EOSDIS?
Cumulus prototyping worked on distributing responsibility
Earthdata Cloud 2021
Strategy for DAAC Migration

Migrating birds, get it?
Using SAFe®* to align Earthdata Cloud 2021 Migrations

*Yes, I am aware of the huge number of buzzwords on image above. And it is less “agile”. But this project involves 100s of people and this system has proven extremely useful.

http://www.scaledagileframework.com/program-increment/
DAAC Cloud Migration Organization January

EOSDIS
Ingest, Archive, and Distribution

- Product Manager
- System Architect
- Release Train Engineer

Cumulus Core Development
GHRC Migration Team
LPDAAC Migration Team
GIBS Migration Team
NSIDC Prototyping Team
ASDC Prototyping Team
PO.DAAC and ASF Prototyping Teams
DAAC Team Organization

DAAC-owned Teams
- Led from internal DAAC resources
- Development, operators, testing
- Requests external support from Core team for specific tasks

Cumulus Platform Core
- Mainly ESDIS contracted development/design resources
- Working on smoothing out the delivery, versioning, on-boarding
- Aiding DAAC teams at specific intervals (development help, face to face meetings, etc) “We need help integrating xyz, etc”
Ingest, Archive, and Distribution is only one of 3 three potential scaled agile teams.

Data Usage (Clients and Services)

Ingest, Archive, Distribution

Cloud Hosting Platform
DAAC Cloud Migration PI Roadmap

**OCT - DEC**
- Planning Activities
- Process alignment
- NSIDC Initial Engagement
- GiBS Performance Testing

**JAN - MAR**
- System Evolution
- Collaboration Coordination
- NSIDC New Data Streams

**APR - JUN**
- Operational Data Onboarding Begins (LPDAAC, GHRC, GiBS)
- Service Integration Begins (LPDAAC)*
- ASDC Initial Engagement

**JUL - SEP**
- Operational System Testing (LPDAAC, GHRC, GiBS)
- Continued Service Integration
- ASDC New Data Streams
All migration teams are able to start onboarding data and metadata **no later than April 2018** in order to meet Earthdata Cloud 2021 milestones.

All migration teams are capable of **standing up and operating Cumulus independently** in order to scale development capabilities.

All migration teams have **detailed migration roadmap** for reaching identified milestones for FY 2018.
Q1 and Q2 2018 Migration Goals

- All migration teams are capable of contributing and testing code/tasks/services to Cumulus core in order to scale development capabilities (March 2018)

- Cumulus Core Evolution and Cross-DAAC System Engineering Activities
  - Updated, cloud-native requirements documentation
  - Cloud-native metrics system
  - Backup and recovery procedures for archived data, workflows, and tasks
  - Distribution API and Egress Shaping
Ingest/Archive/Distribution in FY18

1. Determine objectives
2. Identify and resolve risks
3. Development and Test
4. Plan the next iteration

Cumulative cost
Progress

- Review
- Requirements plan
- Concept of operation
- Development plan
- Concept of requirements
- Verification & Validation
- Test plan
- Verification & Validation
- Test
- Implementation
- Release
- Detailed design
- Code
- Integration
- Operational prototype
- Prototype 1
- Prototype 2

https://en.wikipedia.org/wiki/Spiral_model
Overview


DAAC/GIBS Data stream status: http://bit.ly/2j1JZNq


Cumulus Docs (very much in work) https://cumulus-nasa.github.io/

Cumulus Code Base: https://github.com/cumulus-nasa
What about Code Governance?

Because this is an *organizational* shift, not just a technology shift
Drafting a contribution guide

1. Provide working definitions of the high-level components of the Cumulus system, including specifying which of those components are governed by this document.

2. Establish roles and responsibilities for contributions to Cumulus NASA’s EOSDIS

3. Identify key communication flows, as well as information on documentation, testing and deployment paradigms

4. Outline high-level process expectations for Cumulus contributions and provide example process flows for these contributions
As we scale (up or down) we can adjust/combine/tailor these roles. We can adapt to other projects/systems.
Questions?
katie.baynes@nasa.gov
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAAC</td>
<td>Distributed Active Archive Center</td>
</tr>
<tr>
<td>SIPS</td>
<td>Science Investigator-led Processing System</td>
</tr>
<tr>
<td>EOSDIS</td>
<td>Earth Observation System Data and Information System</td>
</tr>
<tr>
<td>CMR</td>
<td>Common Metadata Repository</td>
</tr>
<tr>
<td>EMS</td>
<td>EOSDIS Metrics System</td>
</tr>
<tr>
<td>NGAP</td>
<td>NASA Compliant General Application Platform</td>
</tr>
<tr>
<td>ESDIS</td>
<td>Earth Science Data and Information System</td>
</tr>
<tr>
<td>GIBS</td>
<td>Global Imagery Browse Services</td>
</tr>
<tr>
<td>GHRC</td>
<td>Global Hydrology Resource Center</td>
</tr>
<tr>
<td>LP DAAC</td>
<td>Land Processes DAAC</td>
</tr>
<tr>
<td>PO DAAC</td>
<td>Physical Oceanography DAAC</td>
</tr>
<tr>
<td>ASDC</td>
<td>Atmospheric Science Data Center</td>
</tr>
<tr>
<td>NSIDC</td>
<td>National Snow and Ice Data Center</td>
</tr>
<tr>
<td>SAFe*</td>
<td>Scaled Agile Framework*</td>
</tr>
<tr>
<td>ASF</td>
<td>Alaska Satellite Facility</td>
</tr>
</tbody>
</table>