Plans to support cloud migration

Manil Maskey
Cumulus

• Lightweight, cloud-native framework for data ingest, archive, distribution and management

• Goals:
  • Provide core DAAC functionality in a configurable manner
    • Data acquisition
    • Data ingest (Validation, Preprocessing)
    • Metadata harvesting, creation, publication into the catalog
    • Data archiving and distribution
    • Metrics publication
    • Enable DAACs to help each other with re-usable components
    • Enable DAAC-specific customizations

• Timeline:
  • Completion of prototype phase – FY17
  • Start of transition phase - FY18
• Phase 1: 06/2016-10/2016
  • Provided data streams for HS3 Field Campaign and AMSR2 datasets
  • Interacted closely with Cumulus development team
  • Developed test plan
  • Tested Cumulus for HS3 data streams

• Phase 3: 06/2017-09/2017
  • Developed metrics prototype
  • Developed transition framework document
  • Tested Cumulus for AMSR2 data streams
FY18 Planned Activities

• Continued support for Cumulus test efforts

• GHRC Data migration to cloud

• Cumulus Metrics Prototype

• Continuous Integration for NGAP

NGAP is the NASA General Application Platform. It provides cloud-based Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS) for ESDIS applications, services, and data.
Cumulus Metrics Prototype

Objective:

*Develop a cloud-native, scalable, cost effective metrics system for Cumulus*

- Implemented a prototype using AWS Athena + Tableau
- Utilized EMS-style files
Cumulus Metrics Prototype
Move all data to cloud – FY18
  - AWS Glacier is already used for offsite backup

Use Cumulus for all forward processing

Adapt tools to use Cumulus APIs

Train staff to operate and develop on NGAP

Develop a migration process document
Forward processing for Field Campaign Datasets

On Premise

Field Campaign Data (from PI) → FC Data Staging → FC Portal Access

Field Campaigns

Rename Files → Create netcdf → Copy to public & archive → Create Metadata → GHRC Storage

DAPPeR → Dream Factory (postgres) → HyDRO → CMR

NGAP Version

Field Campaign Data (from PI) → FC Data Staging → FC Portal Access

DAPPeR

NGAP

Rename Files → Create netcdf → Create Metadata → S3 Storage

Dream Factory (postgres) → Metadata Sync → CMR

Legend:
- GHRC
- Cumulus
Forward processing for Real-Time Datasets

On Premise

- ISS POIC Data Ingest (via TCP ports)
- Port listening script
- Create 2 minute files
- Copy to dev/test env's
- Digest processes (wait for all 4 files)
- Science Processing
- Metadata

ISS LIS

- CRON: daily processing
- CRON: calendar creation
- CRON: NRT/I-K webpages
- CRON: rolling archive
- DAPPeR
- Dream Factory (postgres)
- DMR
- HyDRO

NGAP Version

- LIS Data rebroadcast via PCCC
- Ingest Processing
- Digest processes (wait for 4 files)
- Science Processing
- Create Metadata
- S3 Storage

NGAP

- Scheduled Task: Rolling Archive
- Scheduled Task: Daily Processing

Legend:
- GHRC
- Cumulus

- DAPPeR
- Dream Factory (postgres)
- DMR
Discussion