Archive Management of NASA Earth Observation Data to Support Cloud Analysis

Christopher Lynnes, Kathleen Baynes, Mark McInerney
NASA Goddard Space Flight Center

Big Data from Space 2017, 29 Nov 2017
NASA flies a variety of Earth Observing satellites

- **Hydrosphere**
  - Physical Oceanography
  - Hydrology
- **Atmosphere**
  - Composition
  - Dynamics
- **Biosphere**
  - Land
  - Oceanic
- **Cryosphere**
  - Ice
  - Snow
- **Lithosphere**
- **Anthroposphere**
Earth Observing System Data and Information System (EOSDIS)

data downlink

capture and clean

EOSDIS

distribute

subset

archive

process

Research

Applications

Education

Users
Over time, EOSDIS archive volumes increase exponentially.
Distribution increases similarly to cumulative volume
Three things that will become more difficult and tedious:

1. Provisioning big data storage and supporting utilities
   - floorspace, cooling, power, network

2. Moving big data to local storage and managing them

3. Analyzing (really) big datasets on local compute
Cloud Computing Benefits

1. Infinite data storage available
2. Infinite computing power available
3. Computing power “next to” data storage
4. Rapidly evolving ecosystem of high-performance analytics tools
EOSDIS “Cumulus” Science Data Archive

Orchestrated by Step Function service
Cumulus Dashboard

Dashboard

- 21 Errors
- 6 Collections
- 13 Granules Processed in the Past Hour
- 18.7s Average Processing Time
- 0 Pending Tasks
- 6 Running Tasks
- 4 Queued Messages

Granules Updated Today

- 0 Granules Ingesting
- 0 Granules Processing
- 0 Granules Updating CMR
- 0 Granules Completed
- 0 Granules Failed
Services for Data In Web Object Storage

- Web Object Storage
- Elastic Block Storage
- Cache
- OPeNDAP Server
The Big Win: Data-proximal Analysis
Catch #1: Web Object Storage
Catch #2: Data Processing Paradigm
A Cloud-Native Generalized Analytics Platform

1. Cumulus Web Object Storage
   - Common Metadata Repository
2. Transform as a Service
   - End-User-Specific Analysis
3. Analysis-ready Data as a Service
   - Cloud-Native Analysis
4. Summary Statistics as a Service
   - Visualization as a Service
   - End User Interpretation
   - End User Exploration
Different end users get data from different parts in the value chain based on specific needs and capabilities.
End users can import data from other sources (including their own)
Service-based abstraction enables mix-and-match reuse vs. monolithic apps