Earthdata Cloud Analytics Project

Chris Lynnes* and Rahul Ramachandran*
NASA

*U.S. Civil Servant
Earth Observing System Data and Information System (EOSDIS)

data downlink

capture and clean

distribute

subset

archive

process

Research

Applications

Education

Users
Over time, EOSDIS archive volumes increase exponentially.
Distribution increases similarly to cumulative volume
How do we support user analysis of very large data volumes?
Solution: Data-proximal Analysis
Goals

1. Enable big compute next to big data
2. Encourage user adoption of cloud for analytics
3. Maximum analytics capability at minimum cost
   a. Use capabilities within NASA more effectively and efficiently
   b. Leverage analytics capabilities of external partners
Key Features

1. Satisfy a diverse user community
2. Support analysis in the cloud without egressing data
3. Facilitate multi-dataset comparison and fusion
4. Support batch, interactive and streaming modes
5. Support distributed filesystems and databases
6. Support cost constraints and cost-sharing
Earthdata Cloud Analytics Guiding Principles

1. Infusion- and innovation-friendly framework and building blocks
2. No monolithic systems
3. Open code and services
4. Interoperability and reuse
5. No unnecessary duplication ("undifferentiated heavy lifting")
Architectural Concept

Earth Science Data Analytics the Cloud-Native Way: Everything is a Service

This approach produces key important benefits for the user community and EOSDIS
Abstract Analytics Workflow

data → Extract Transform Load → Analyze → Visualize
Earthdata Cloud Analytics Reference Architecture

Cumulus Data Archive -> Preprocessing as-a-service -> AODS\(^1\) as-a-service -> Analysis as-a-service -> Visualization as-a-service

\(^1\) Analytics Optimized Data Store
Interactive Mode: Analytics-Optimized Storage

Cumulus Data Archive

Preprocessing as-a-service

AODS$^1$ as-a-service

Analysis as-a-service

Visualization as-a-service

$^1$ Analytics Optimized Data Store
Batch Mode

Preprocessing as-a-service

Analysis as-a-service

Visualization as-a-service

Cumulus Data Archive

AODS\(^1\) as-a-service

\(^1\) Analytics Optimized Data Store
Streaming Mode

Cumulus Data Archive

Preprocessing as-a-service

Event Analytics as-a-service

Visualization as-a-service
Open Pipeline Provides Outputs at Different Stages Appropriate for a Diverse User Base

Cumulus Data Archive → Preprocessing as-a-service → AODS¹ as-a-service → Analysis as-a-service → Visualization as-a-service

¹ Analytics Optimized Data Store
Open Pipeline Provides Outputs at Different Stages Appropriate for a Diverse User Base

Cumulus Data Archive

Preprocessing
as-a-service

AODS\(^1\)
as-a-service

Analysis
as-a-service

Visualization
as-a-service

End-User-Specific
Analysis

End-User
Cloud-Native
Analysis

\(^1\) Analytics Optimized Data Store
Open Pipeline Provides Outputs at Different Stages Appropriate for a Diverse User Base

Cumulus Data Archive → Preprocessing as-a-service → AODS¹ as-a-service → Analysis as-a-service → Visualization as-a-service

- End-User-Specific Analysis
- End-User Cloud-Native Analysis
- End-User Interpretation

¹ Analytics Optimized Data Store
Open Pipeline Provides Outputs at Different Stages Appropriate for a Diverse User Base

Cumulus Data Archive

Preprocessing as-a-service

AODS\(^1\) as-a-service

Analysis as-a-service

Visualization as-a-service

End-User-Specific Analysis

End-User Cloud-Native Analysis

End-User Interpretation

Data Exploration

\(^1\) Analytics Optimized Data Store
Open Pipeline Enables Integration with Other Data, Scripts, and Workflows

1 Analytics Optimized Data Store
Open Pipeline Enables Integration with Exploitation Platforms

Cumulus Data Archive

Preprocessing as-a-service

AODS\(^1\) as-a-service

Analysis as-a-service

Visualization as-a-service

\(^1\) Analytics Optimized Data Store