



NASA Cloud Data Access

Christopher Lynnes*

*U.S. Civil Servant at NASA

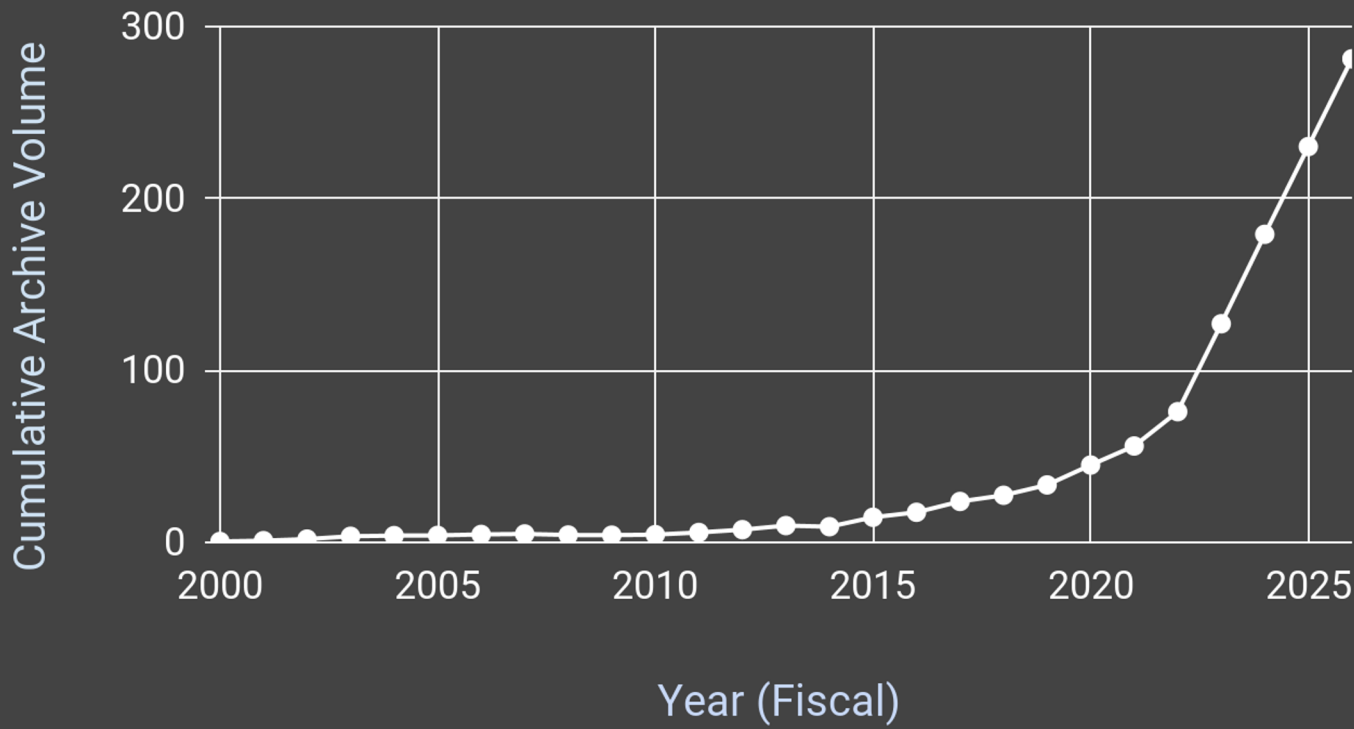


Overview

- Hosting data in the Cloud enables users to analyze the data in place, with no data movement or management required.
- Hosting the data in one virtual place facilitates multi-dataset studies and scientist collaboration.
- NASA is developing a framework to support both analysis-in-place and traditional data download access to data.



Projected Data Volume Rise





Data Usage Benefits of Cloud Computing for End Users

1. Scale

- a. Availability of compute power
- b. Non-necessity to move or manage (lots of) data

2. Togetherness

- a. Datasets
- b. Users



The main benefit from the Enterprise perspective:

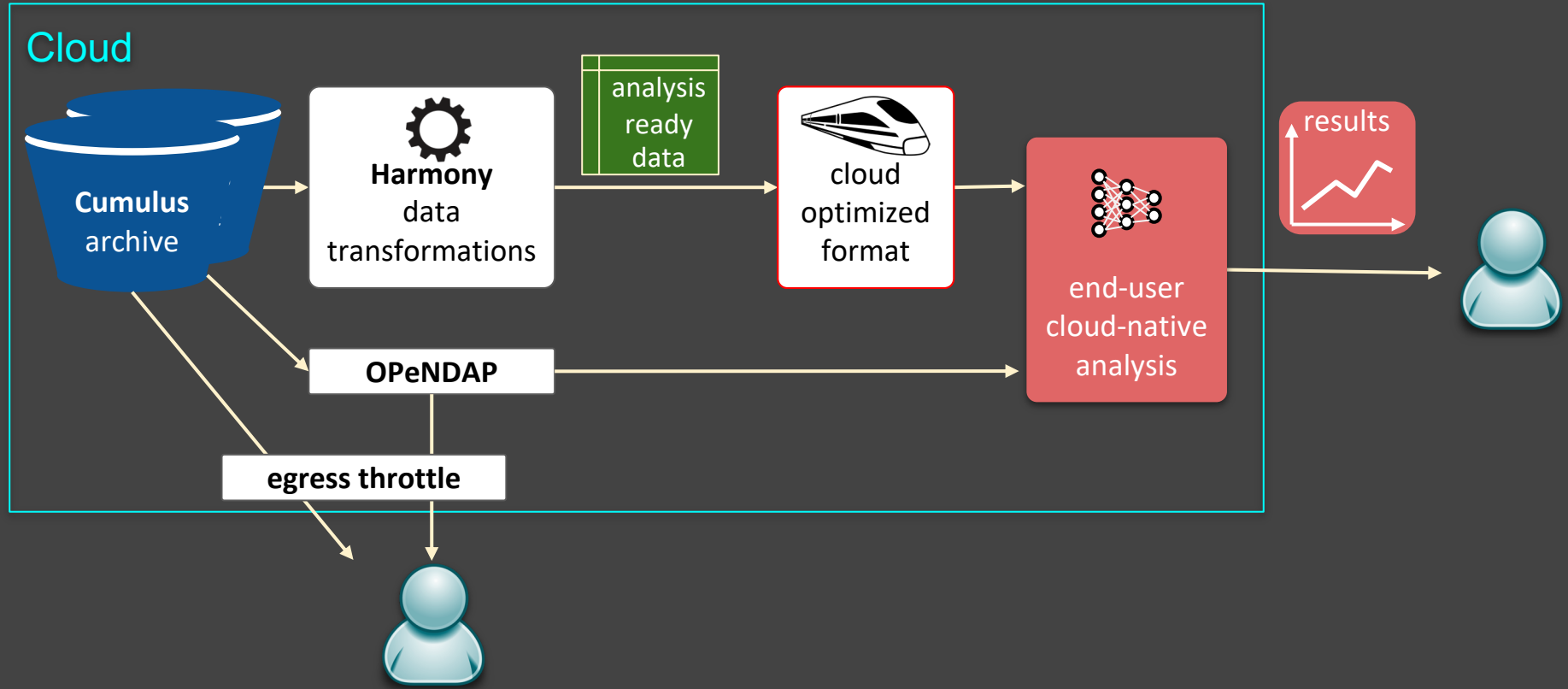
Hosting data in the cloud allows users to Analyze in Place

- **Power:** Any user can access big processing power “next to” Big Data.
- **Performance:** Data can be offered in a form enabling high-performance analysis.
- **Freedom from Data Transfers:** Users need not move Big Data.
- **Freedom from Data Management:** Users need not store and manage Big Data.
- **Data Co-location:** Users can easily work with multiple EOSDIS datasets together, without moving them to a common location.
- **Choice:** Users can still download data if they prefer.

**Specific benefits will vary with use case.*

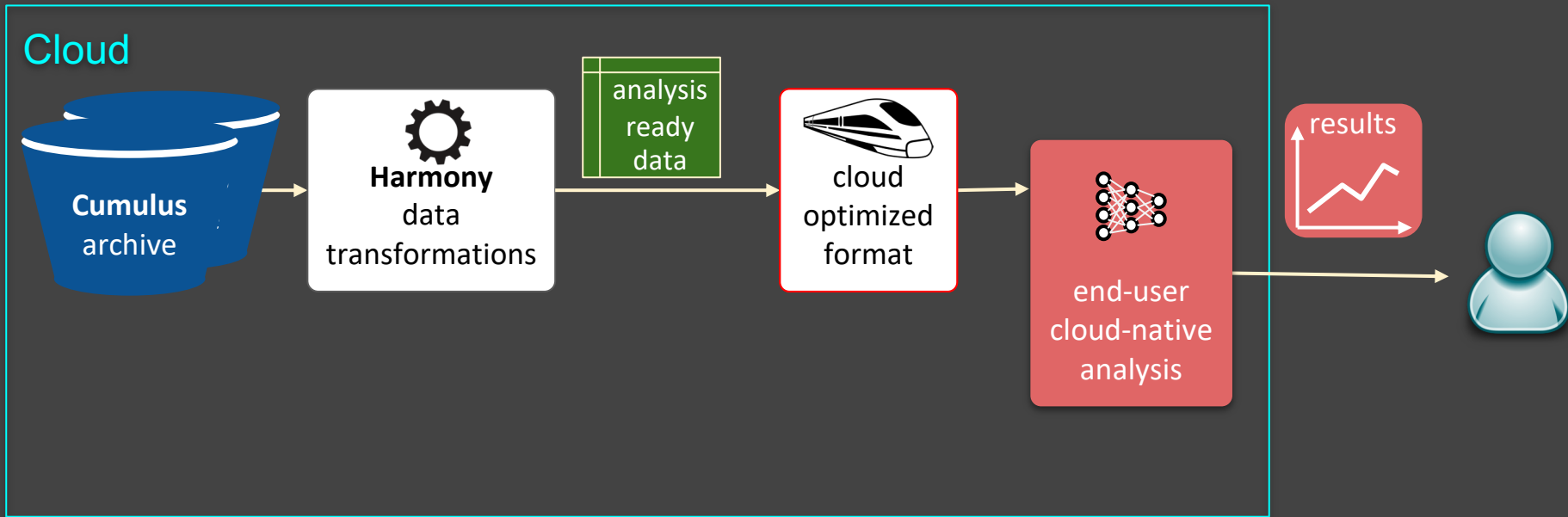


EOSDIS Data Access



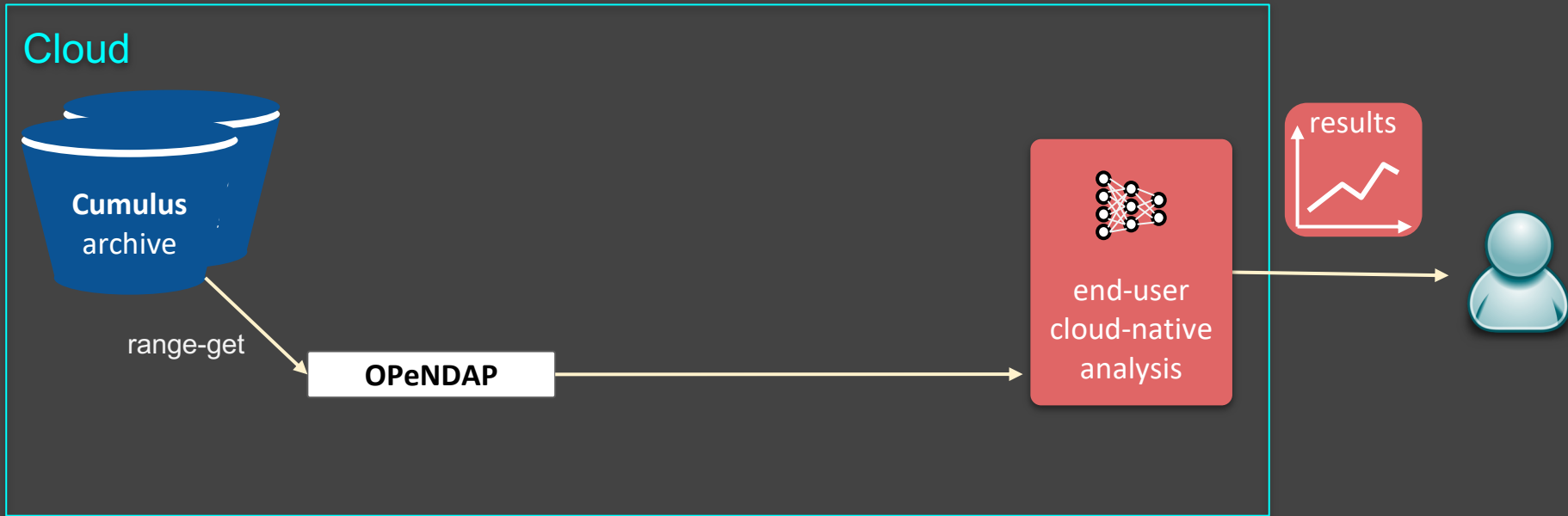


Analysis In Place



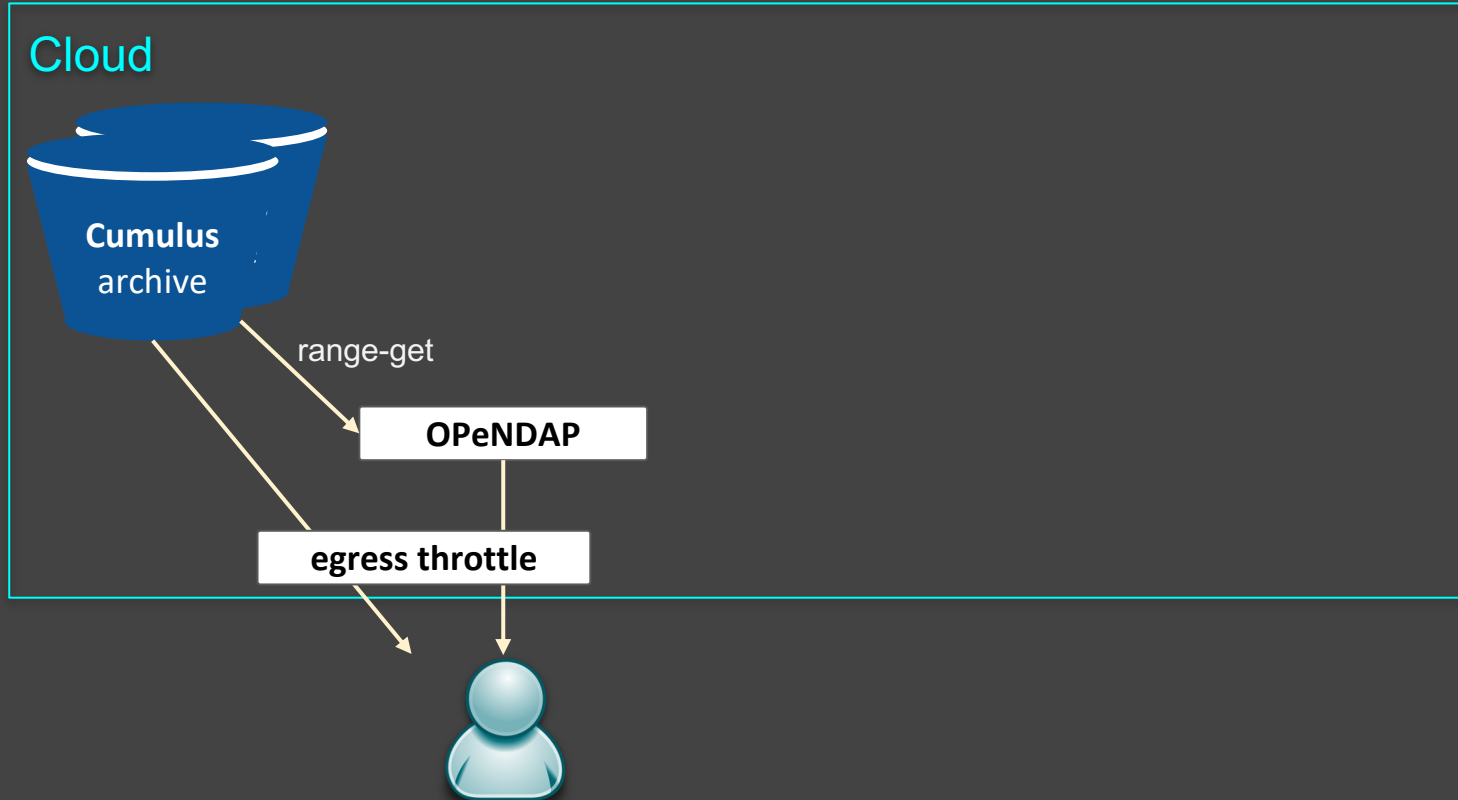


Analysis In Place





Download Scenario





Cloud Facet in Earthdata Search

Search for collections or topics

📅

🔗

☰

🔼 Filter Collections

Categories

Features

☒ Available from AWS Cloud

🏠

700 Matching Collections

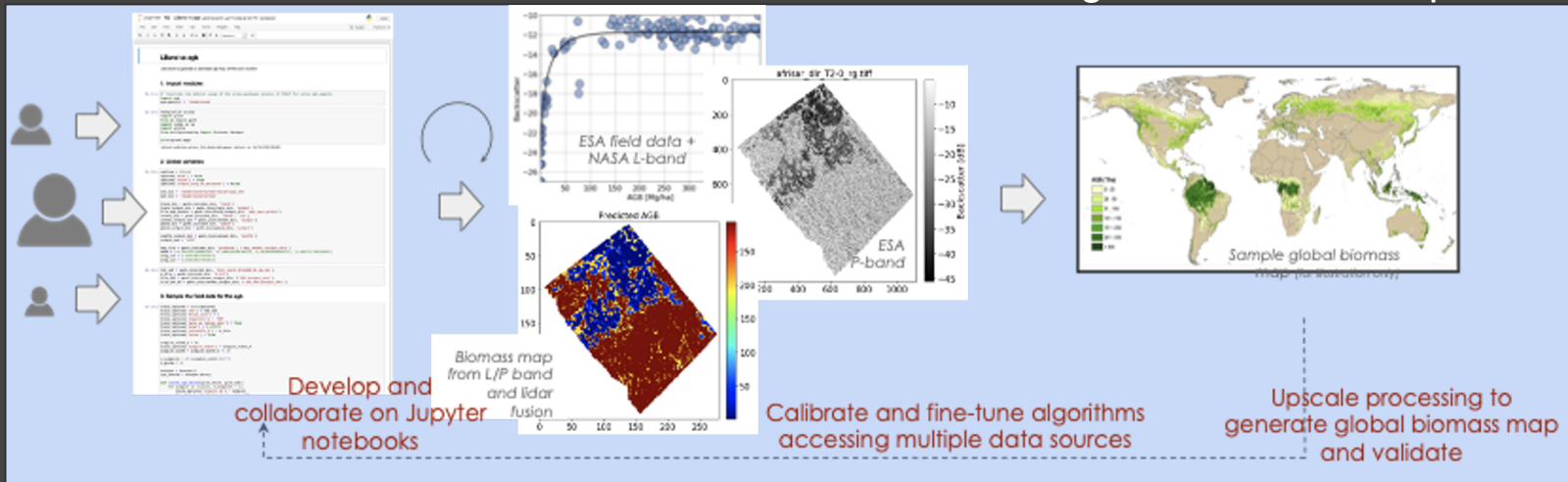
Showing 20 of 700 matching collections

Collection		Version
SENTINEL-1A_SLC	i +	1
SENTINEL-1B_SLC	i +	1



Multi-Mission Algorithm and Analysis Platform (MAAP)

- Prototype innovative data system to enable open science in cloud
 - Implemented initially for Biomass: Lidar, Synthetic Aperture Radar
 - Adaptable to other disciplines
- Provide seamless access to satellite, airborne and field data from European Space Agency and NASA
- Scale user algorithms from small regions to global scale
- Enable collaboration on calibration and validation of higher-level science products





NASA Cloud Data Access Status

Cumulus Archive	Operational: 1 complete archive, 4 partial archives 685 datasets
Egress throttle / cutoff	Operational
Cloud-Optimized Datasets	Two Harmonized-Landsat Sentinel in Cloud-optimized GeoTIFF
OPeNDAP	Operational in mid-April
Direct Access URLs in catalog	In progress
Harmony	Operational for a few datasets Convert to zarr, spatial subset, regrid
End-User Training	Primer, workshops, Openscapes underway
MAAP	Almost operational



Lessons

1. Commercial cloud works...
 - a. ...mostly
 - b. Keep a catalog in the same region as the data
2. Putting datasets in the cloud enables:
 - a. Processing at scale (MAAP, Pangeo)
 - b. Combination dataset usage (EOSDIS)
 - c. Collaboration on data transformation services (EOSDIS)
 - d. End-user science collaboration (MAAP)
3. Users could benefit from some assistance in moving their analysis to the cloud