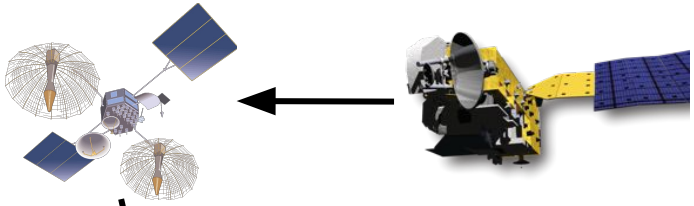
Assessing Applications of Cloud Computing to NASA's Earth Observing System Data and Information System (EOSDIS)

Chris Lynnes, Katie Baynes, Mark McInerney
NASA/GSFC
ESDIS

Earth Observing System Data and Information System (EOSDIS)



data downlink

capture and clean

process

archive

subset

distribute

EOSDIS

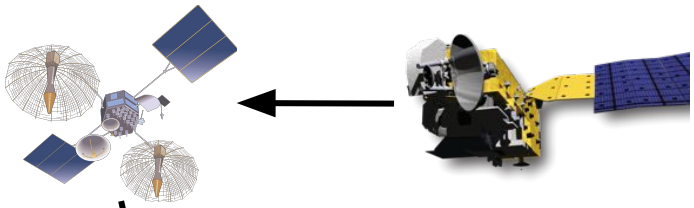
Research

Applications

Education

Users

Earth Observing System Data and Information System (EOSDIS)



data downlink

capture and clean

process

archive

subset

distribute

EOSDIS

Research

Applications

Education

Users

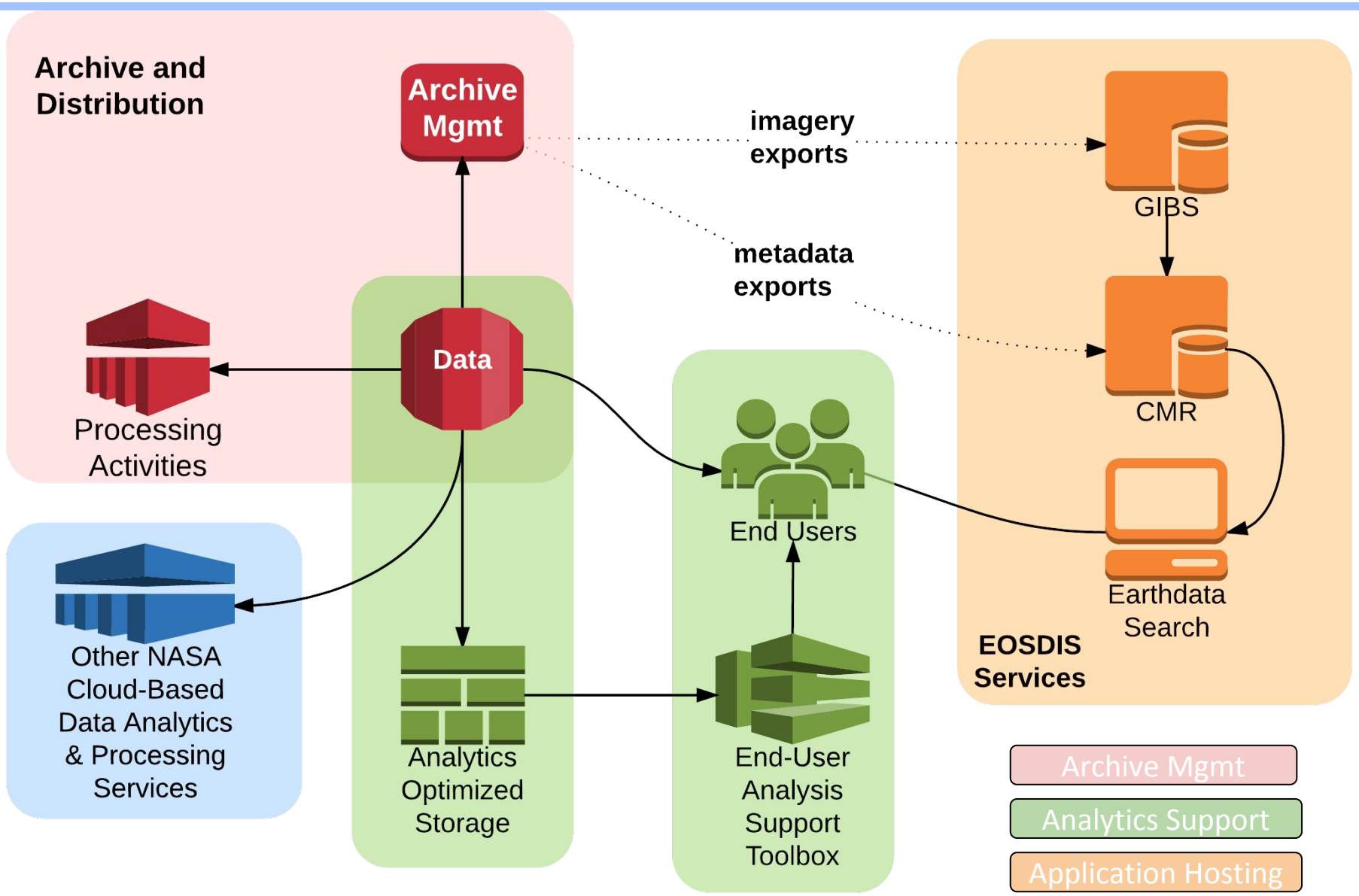
Opportunities from Cloud Computing

1. Data Volume
    a. Challenge for Archives
    b. Challenge for Users
2. Performance
    a. Load balancing and scaling
    b. Resiliency
3. Cost Reduction(?)

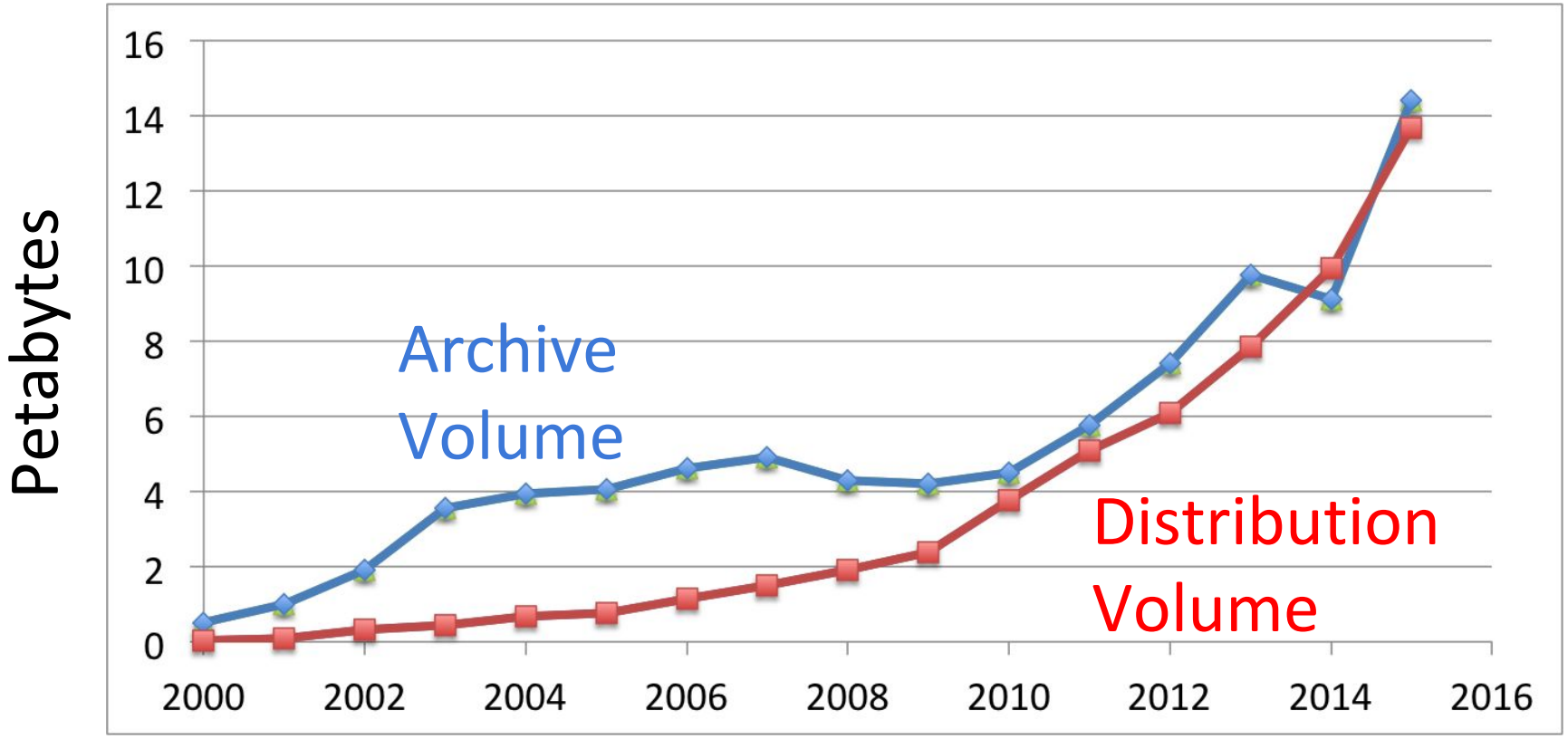
Overall Approach

- Program of prototypes
  - Demonstration prototypes: will this work?
  - Operational prototypes: how does this change the way we work?
- Focus on public clouds for max cost savings
- Leverage existing software and efforts…
- …But refactor software to be cloud-native
  - not “forklift migrations”

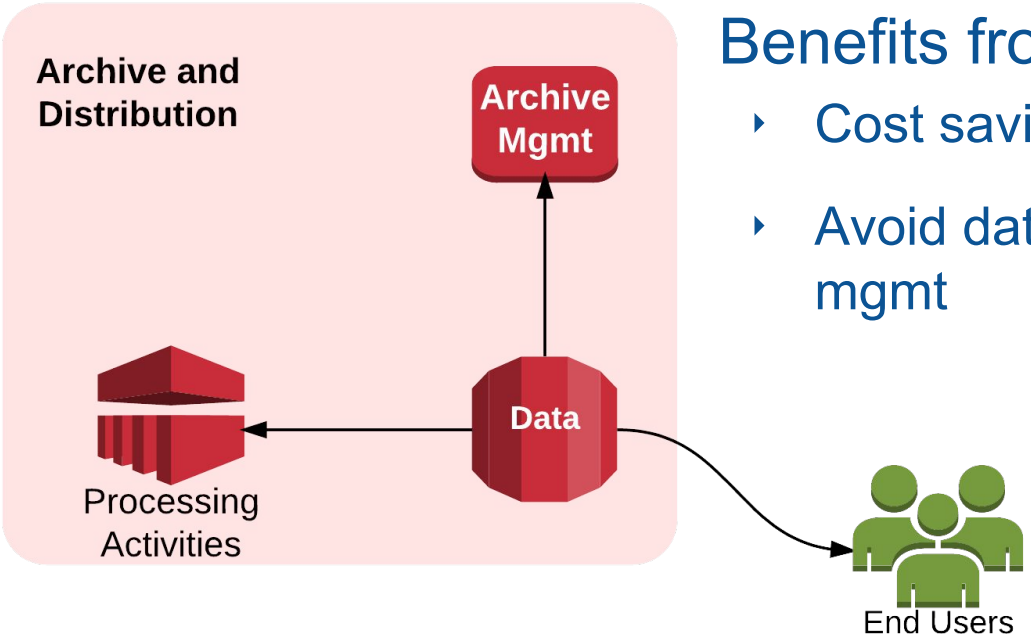
EOSDIS Cloud Prototypes



Data Volume Challenge for Archives



Archive Cloud Prototypes

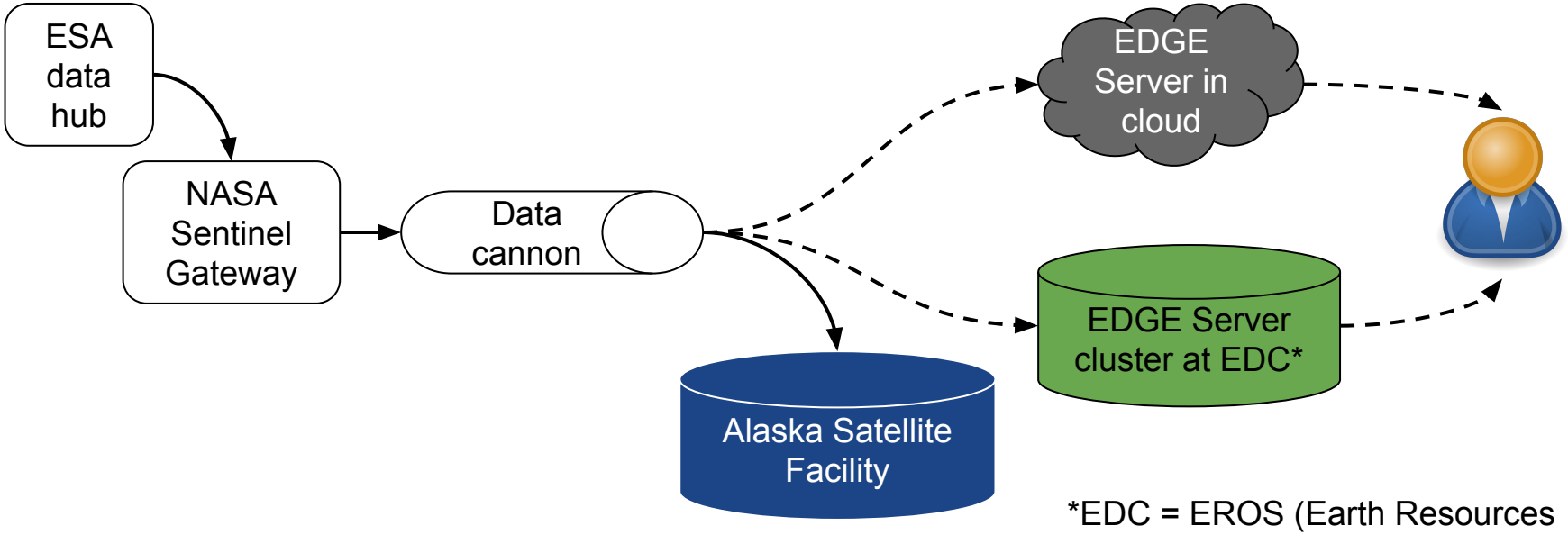


Benefits from Archive in the Cloud

- Cost savings for storage of Big Data?
- Avoid data downloading and local data mgmt

Archive Cloud Prototypes

- Alaska Satellite Facility Web Object Storage prototype
  - Distribute Sentinel radar data from Amazon storage



- Global Imagery Browse Service in the Cloud (AWS Lambda)
- Ingest and Archive management prototype (AWS Lambda)
- NISAR Mission Preparatory Prototype

Data Volume Challenge for the Users

One Solution:

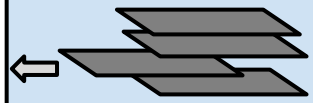
Move more analysis closer to the data

“Analysis”

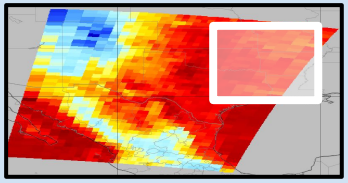
More Complexity

Subset

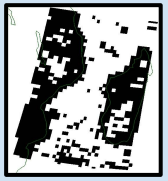
Data Variable



Spatial Area

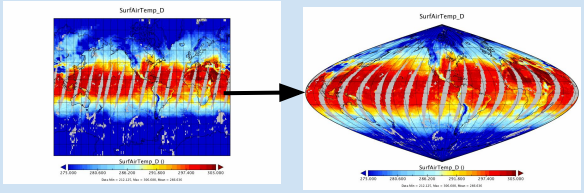


Quality Filter

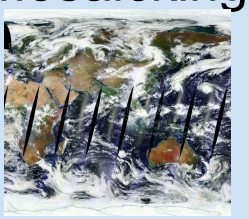


Transform

Reprojection

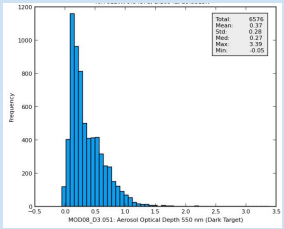


Mosaicking

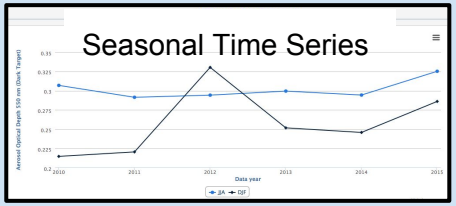


Analyze

Simple Stats

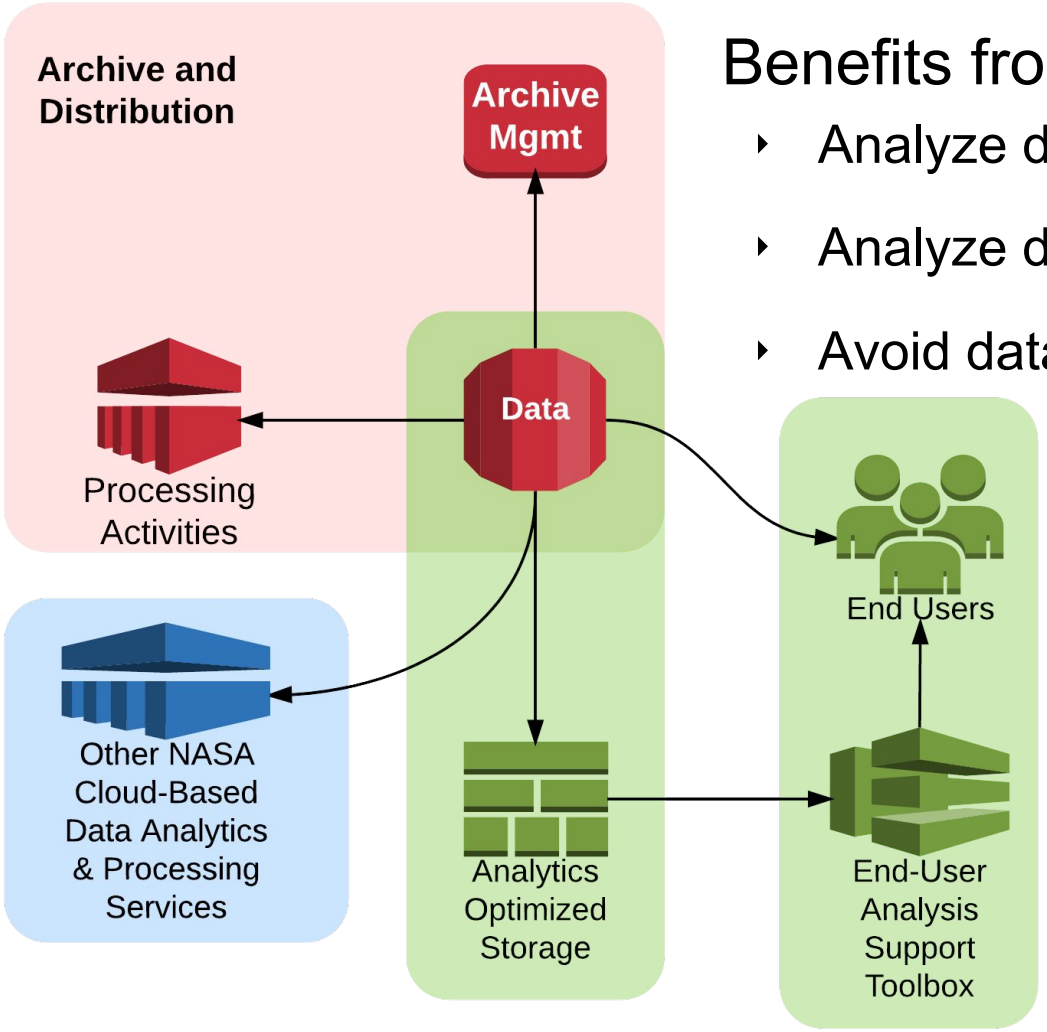


Complex Stats



End User’s Algorithm

Cloud Analytics Prototypes

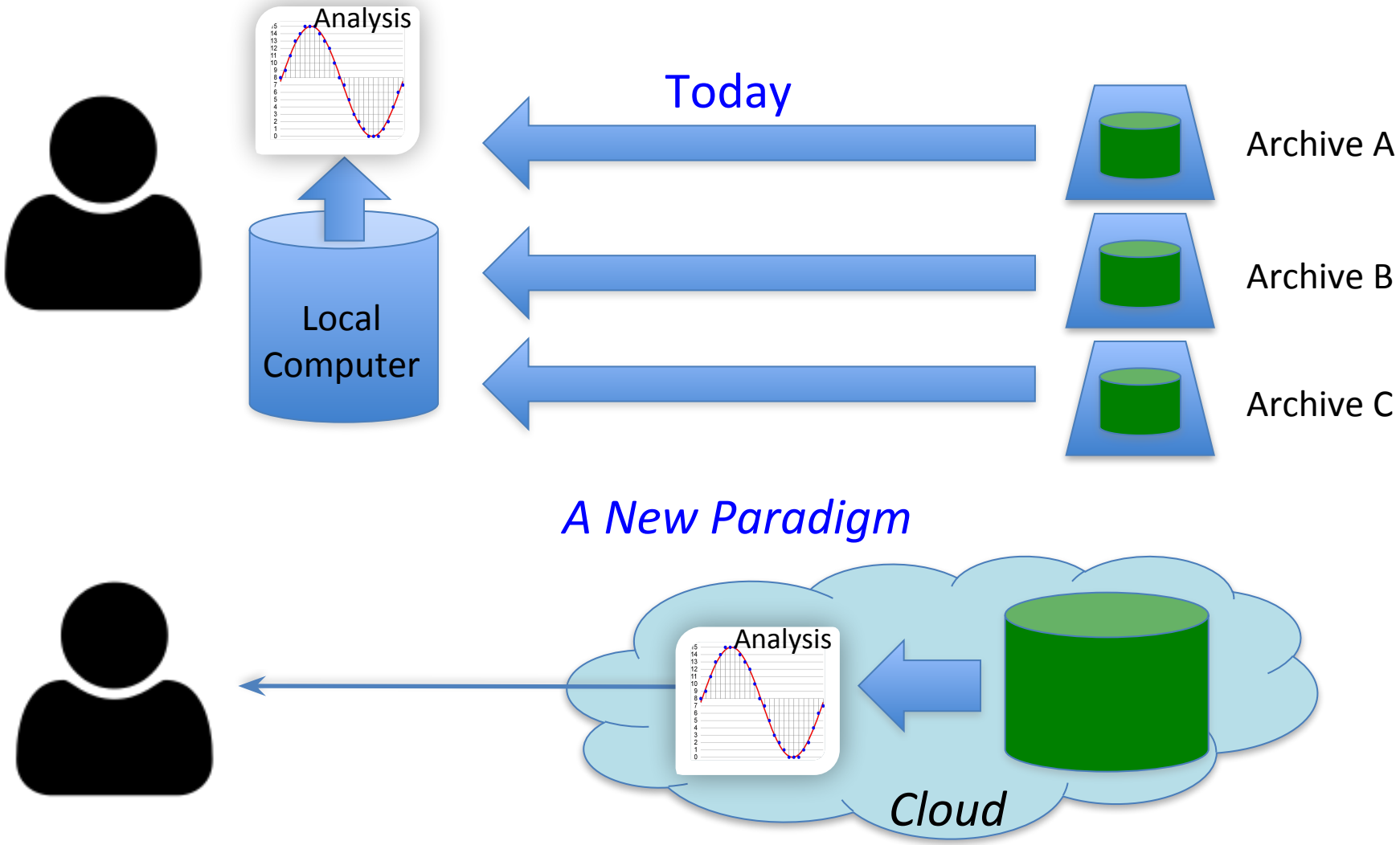


Benefits from Cloud Analytics

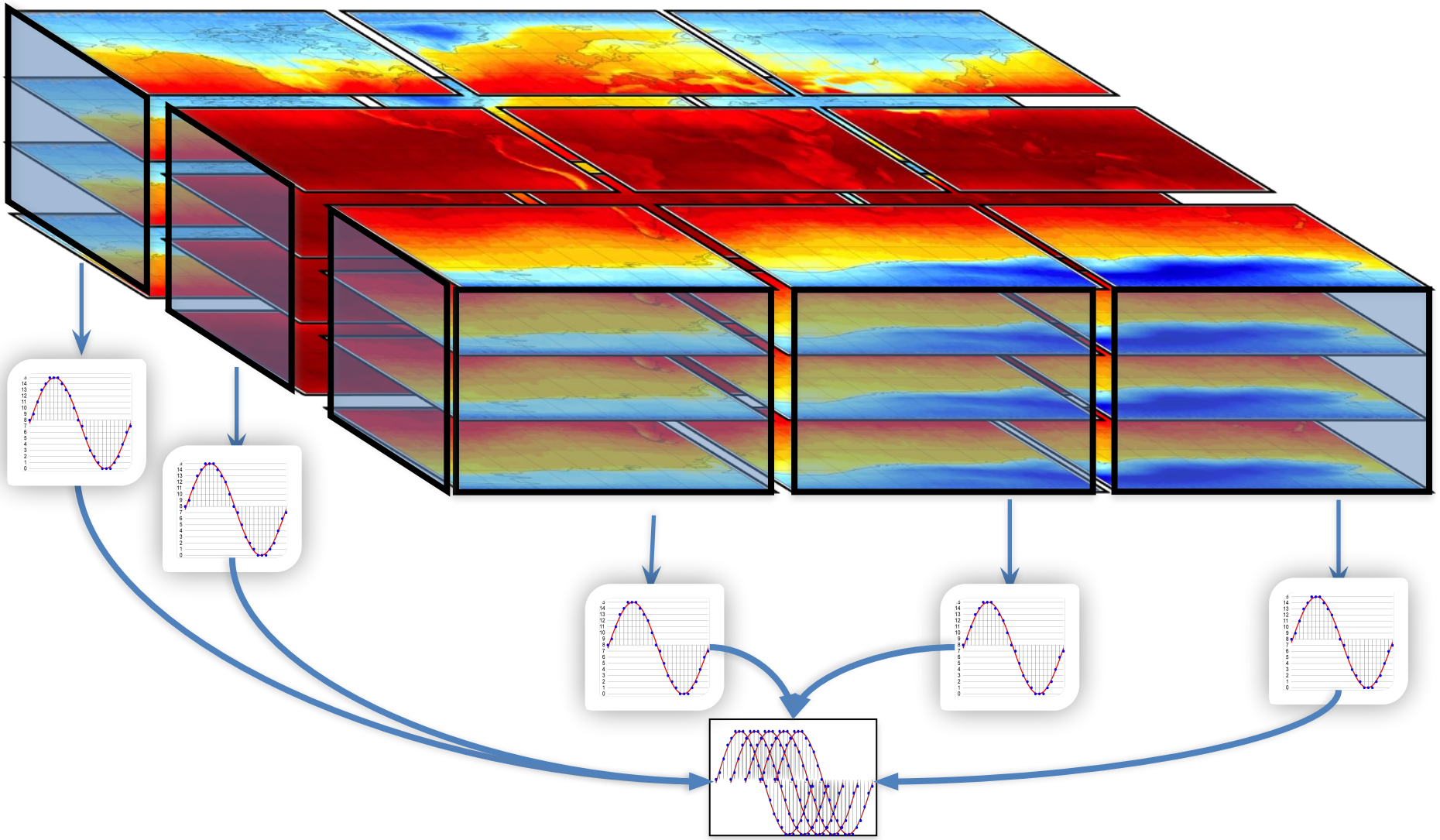
- Analyze data at scale
- Analyze datasets together easily
- Avoid data downloading and local mgmt

Users work with the data “in place”:

Avoid tedious cycles of download-preprocess-store-analyze

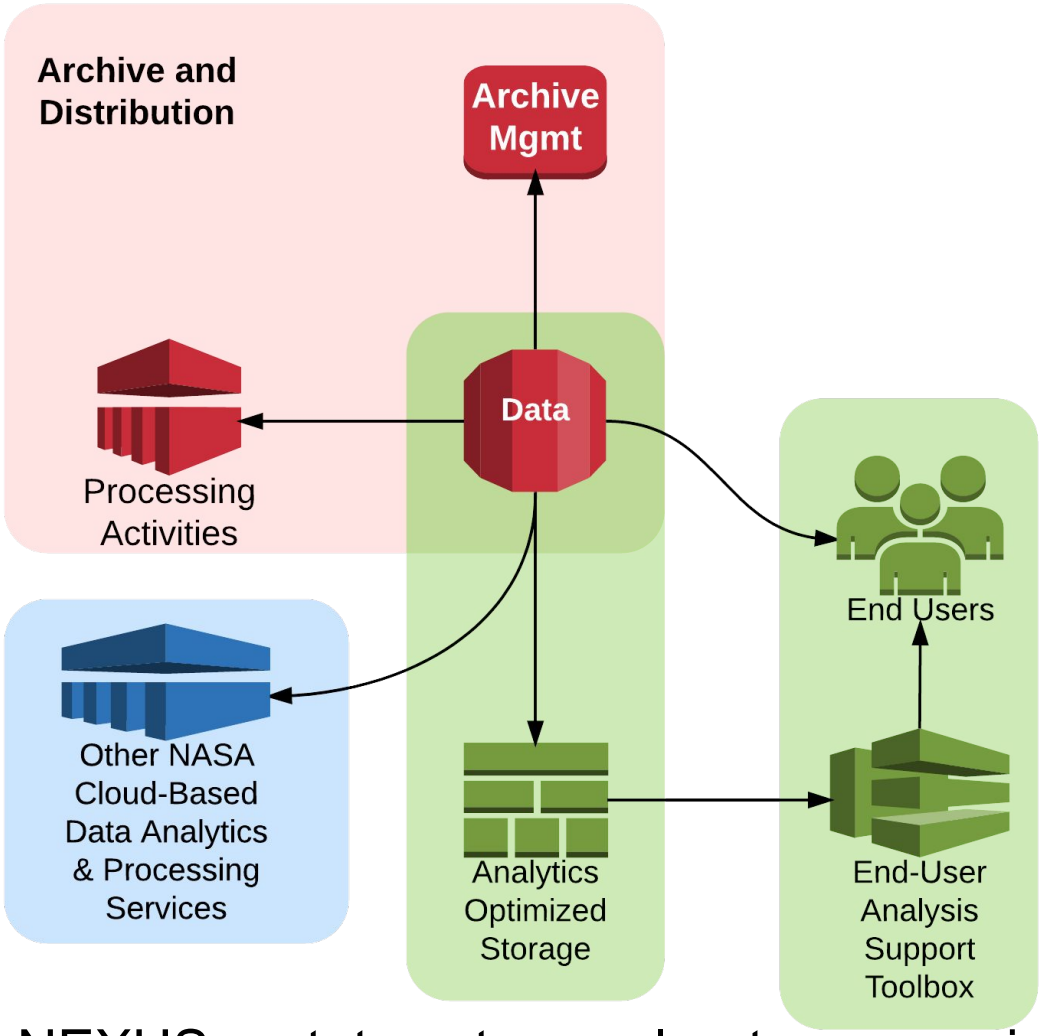


Distributing the data enables distributed, parallel computing.



Results are computed for chunks of the problem simultaneously and then recombined and reduced

Cloud Analytics Prototypes



Analysis support toolbox to attract users to cloud analytics

- Community open source tools
- DAAC-developed tools
- Cloud analytics examples and recipes

NEXUS prototype to accelerate processing

- Spark + Cassandra DB

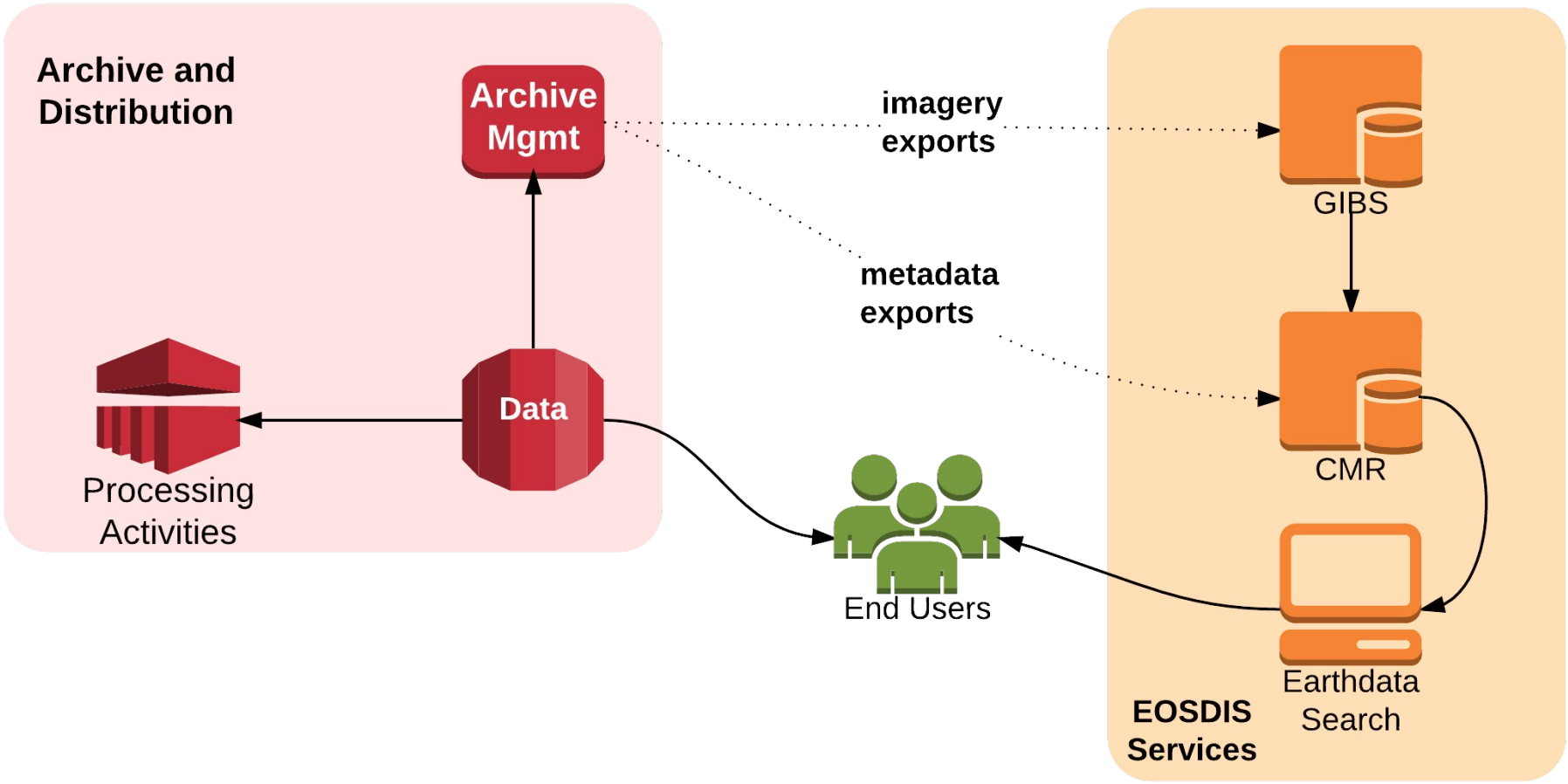
Analytics in the Cloud

Long Term Paradigm Shift:

How Scientists Work on Data

- Scientists will work on data “in place”
  - instead of downloading
- High-value data will (also) be in databases
- Pre-existing toolsets will be easy to find and use

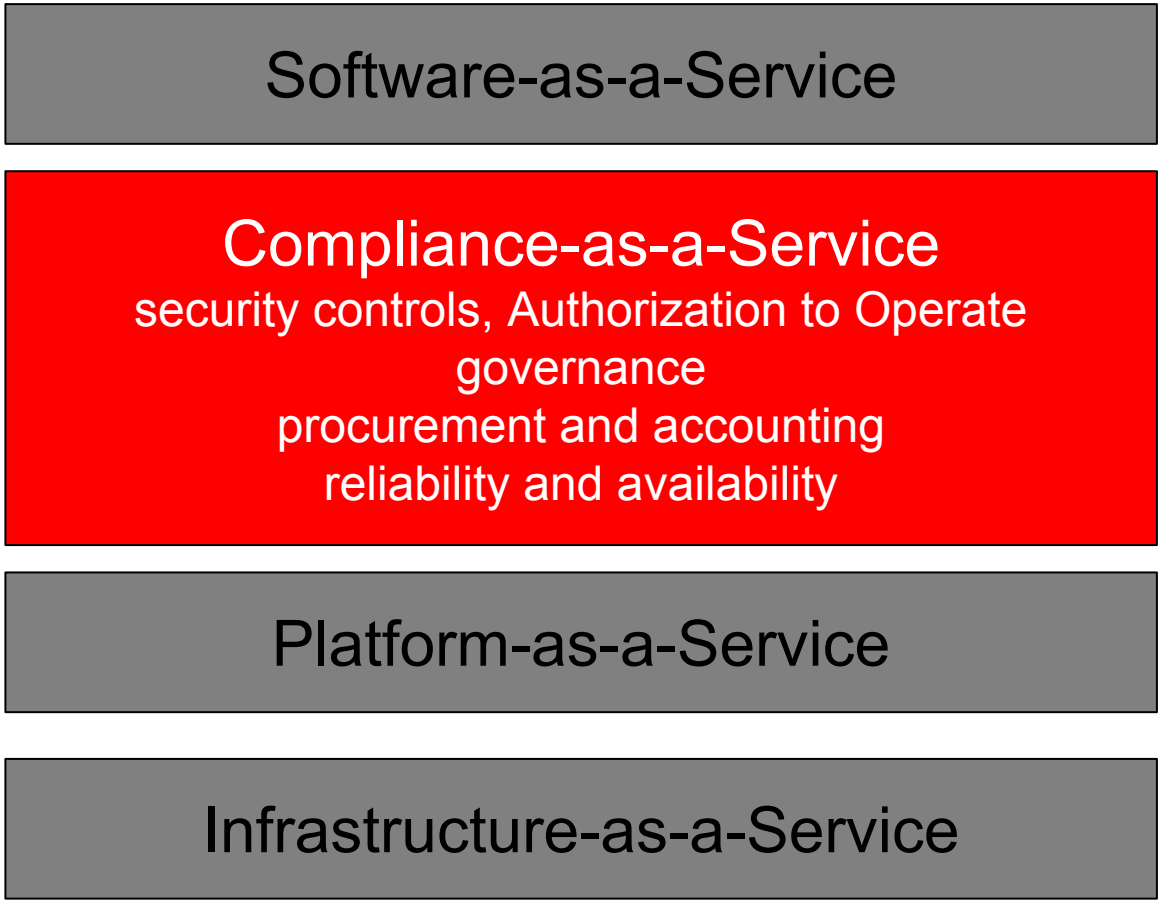
Application-Hosting & Processing Prototypes



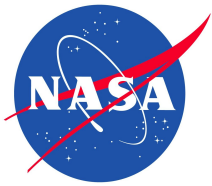
Cloud Application Deployments

- Port existing applications to public cloud
- Leverage NASA-Compliant General Application Platform

NGAP: “Compliance-as-a-Service”



Application-Hosting & Processing in the Cloud



- Paradigm Shift: How We Implement Systems
  - Off-the-shelf systems reduce effort for hardware procurement and deployment
  - Automate *everything*: testing, deployment, scaling, failover, ...
- Benefits
  - Science users get:
    - New capabilities sooner
  - NASA gets:
    - More reuse
    - Lower hardware costs

EOSDIS Cloud Progress so far...

Archive Prototypes

- Serving data to the public from Alaska Satellite Facility Web Object Storage prototype in AWS Simple Scalable Storage
- End-to-end lambda workflow has been demonstrated for the Ingest / Archive management prototype
- Global Imagery Browse Service is undergoing system testing

EOSDIS Cloud Progress so far...

Analytics Prototypes

- NEXUS analytics algorithms benchmarked vs. Giovanni
  - Roughly order-of-magnitude speedup achieved on a cluster
  - Now porting to cloud-native architecture

EOSDIS Cloud Progress so far...

Application Hosting Prototypes

- NASA-Compliant General Application Platform is now authorized to operate publicly
- Earthdata Search client is operational and accessible to the public in AWS
- Modified processes to account for new costing mechanisms

Terra Incognita

1. Vendor Lock-in
2. Future storage costs
3. Uncapped egress costs
4. Security Restrictions and Network trust