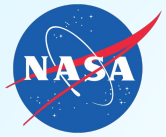
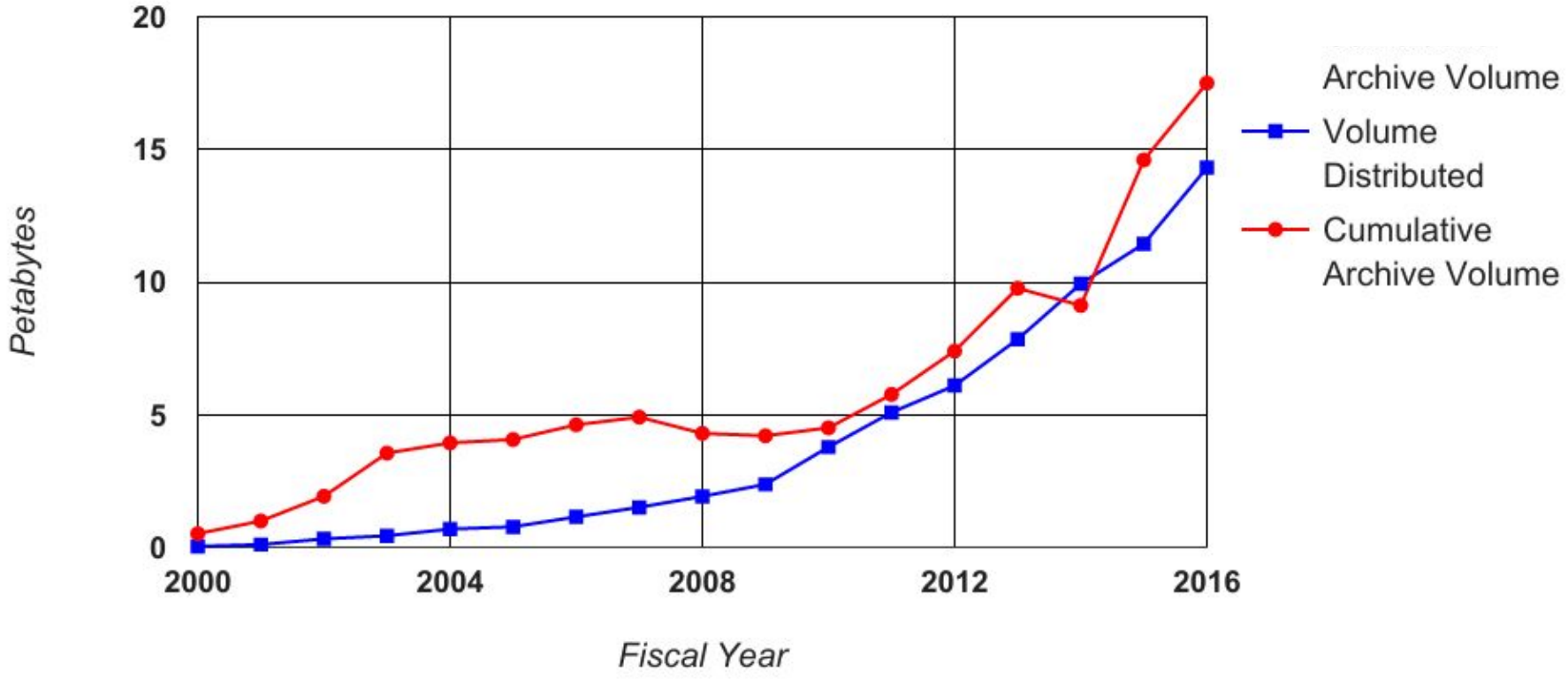


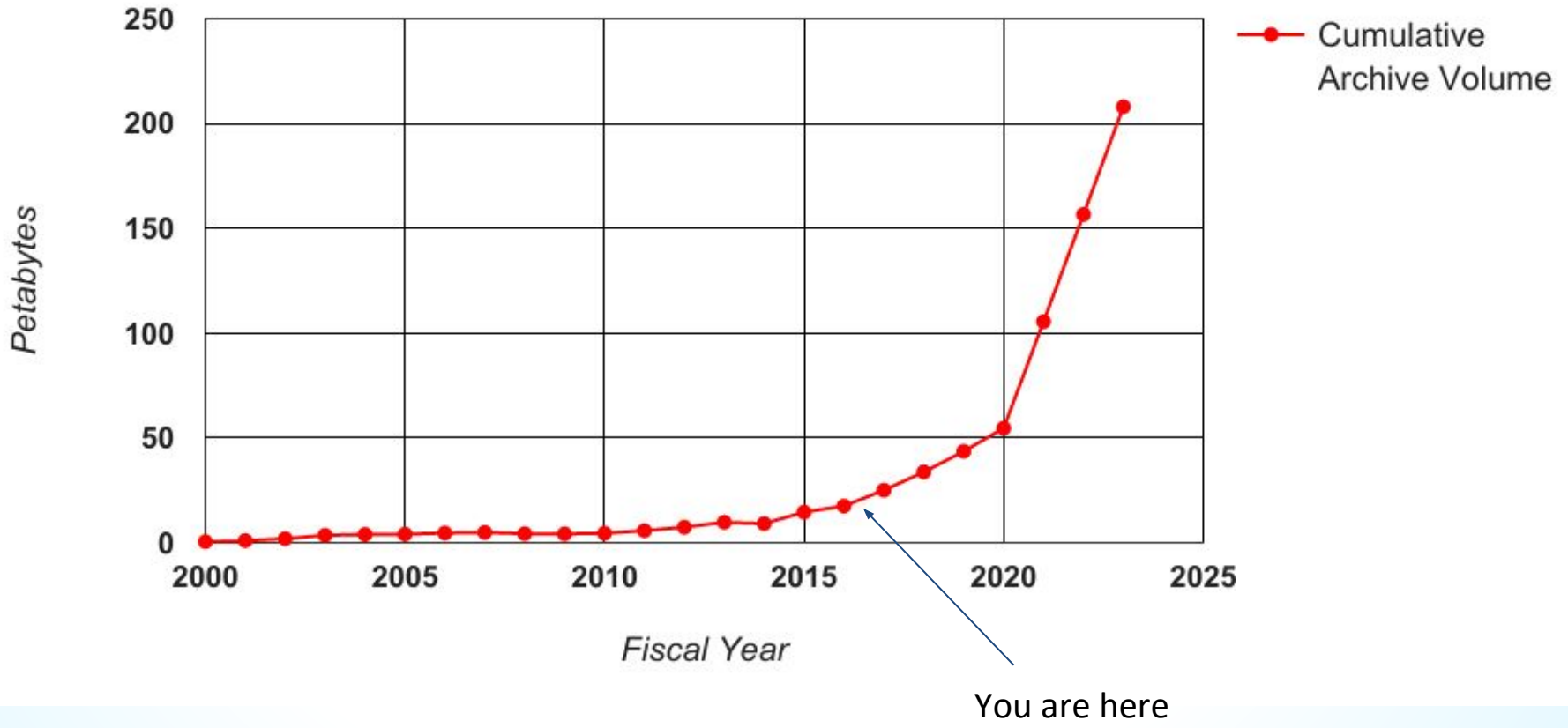
NASA EOSDIS
Cloud Prototype Systems

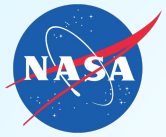
Chris Lynnes, Mark McInerney, Katie Baynes
NASA

Annual distribution is on the same order of magnitude as the total archive volume



And our archive is scheduled to grow dramatically





If these high volumes are a challenge for robust, experienced data centers to manage...

...what about the science end users?

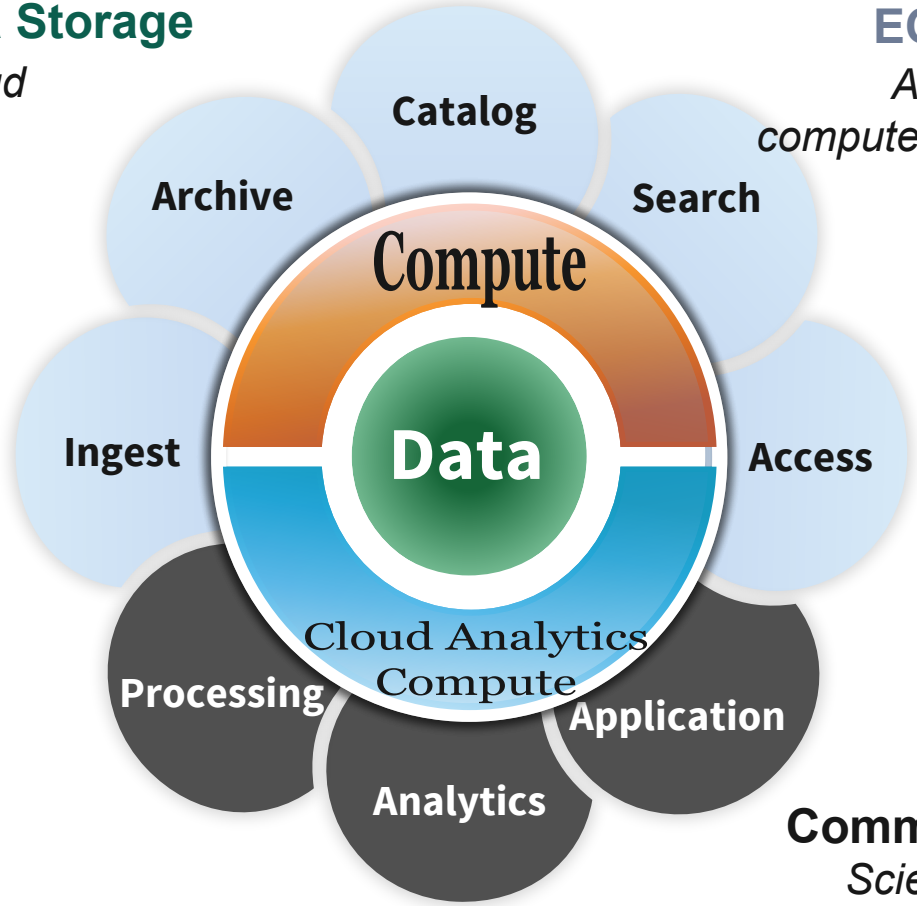
Conceptual "Data Close to Compute"

Large Volume Data Storage

Datasets stored in cloud object storage

EOSDIS Applications

Applications using cloud compute, storage, and services



Scalable Compute

Provision, use, and terminate as needed

Community Applications

Science community brings algorithms to the data.

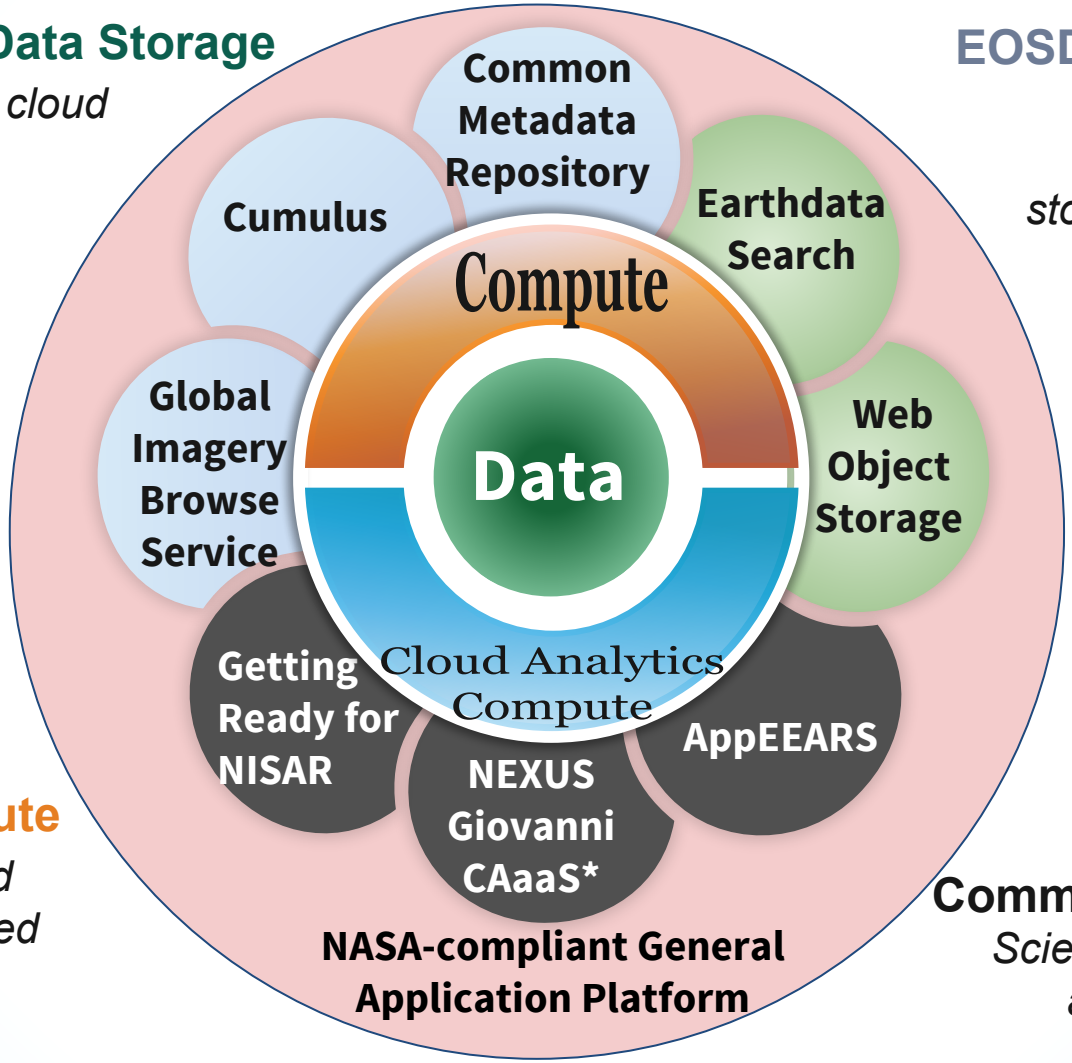
Conceptual "Data Close to Compute"

Large Volume Data Storage

Datasets stored in cloud object storage

EOSDIS Applications

Applications using cloud compute, storage, and services



Scalable Compute

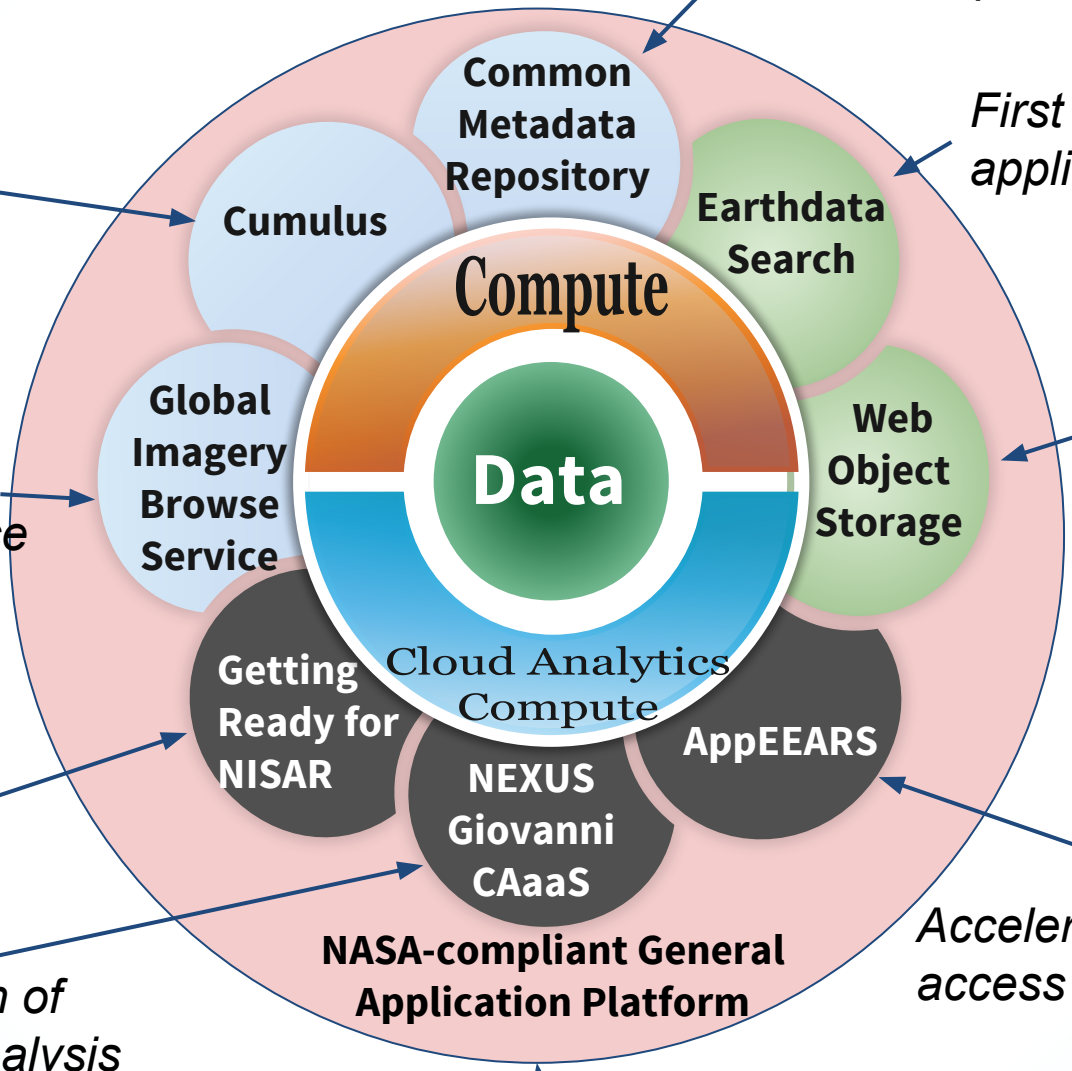
Provision, use, and terminate as needed

Community Applications

Science community brings algorithms to the data.

*CAaaS = Climate Analytics as a Service

Specific Benefits



Avoidance of expensive hardware purchase

First public EOSDIS application in the cloud

Serverless architecture demonstration

High-bandwidth edge server for radar data

Improvement of ingest performance

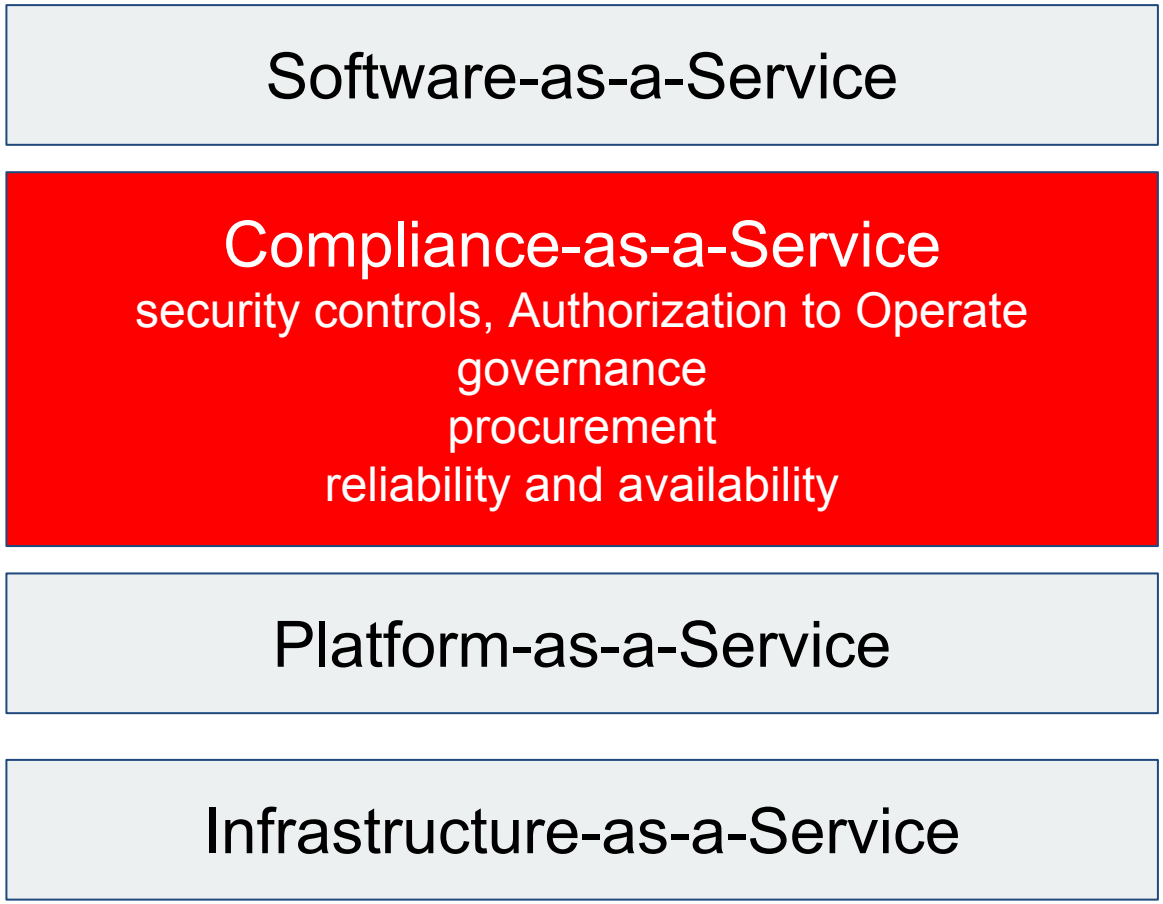
On-demand regeneration of virtual products

Acceleration of OPeNDAP access to data

Acceleration of end-user analysis

Approval to operate publicly

NASA-compliant General Application Platform: “Compliance-as-a-Service”



EOSDIS

8

Key Cloud Benefits for Science Users

Immediate direct access to data
Immediate availability of compute

Lower Barrier to Entry

Dynamic scaling for bursting
Variety of CPU and GPU

Scalable Performance

Purchase compute as needed
Spot market for cost control
No need for local copy of data

Cost

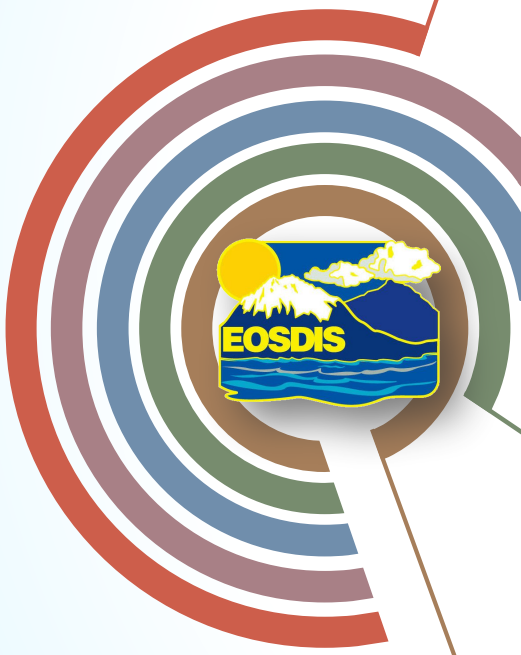
Experiment replication enabled
Sharing of complex analysis workflows

Repeatability & Sharing

Decrease in startup time for projects
Analysis at scale for anyone
Rapid iteration enabled

Promoting Science Discovery

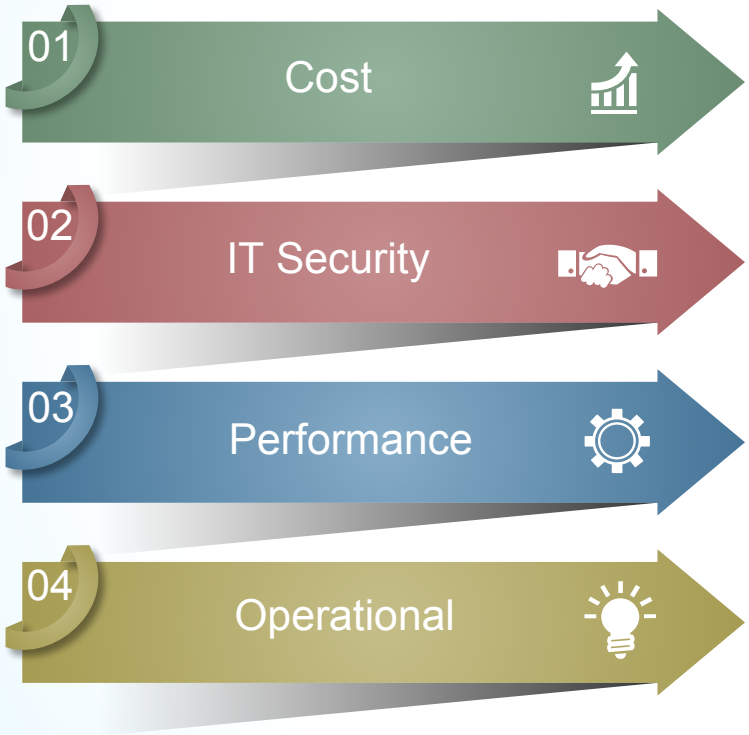
Key Cloud Benefits for Data Systems (EOSDIS)



- **Cost-Effective**
  - Costs only for the compute and storage used
  - Separate accounting for project/function
- **Scalable Performance**
  - Auto scaling, elastic load balancing
  - Adjustable duty cycle
- **Flexibility**
  - Storage type, CPU size
  - Operating system, languages
- **Deployment Speed**
  - Simplified procurement + provisioning
  - DevOps automation
- **Accessibility**
  - Data next to compute
  - All data in the same virtual “place”

Decision Considerations

High level decision consideration for individual project prototypes and capabilities to operationalize into AWS (commercial cloud)

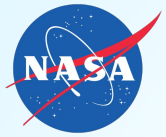


Is AWS (commercial cloud) affordable?.

Is NASA IT Security compliance and tactical operations achievable in AWS (commercial cloud)?

Is performance equal to or better than current on-premises solutions?

Can we operate “Operationally” in AWS (commercial cloud), technical and business?



Most Obvious-yet-surprising Lesson: Savings from Test Environments

If N is the number of Virtual Machines:

- Full-workload testing: $N_{perftest} = N_{ops}$

Most Obvious-yet-surprising Lesson: Savings from Test Environments

If N is the number of Virtual Machines:

- Full-workload testing: $N_{perftest} = N_{ops}$
- On-Premise: $(N_{ops} + N_{perftest}) \times 24hr \times 7d = 336$ N-hr

Most Obvious-yet-surprising Lesson: Savings from Test Environments

If N is the number of Virtual Machines:

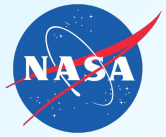
- Full-workload testing: $N_{perftest} = N_{ops}$
- On-Premise: $(N_{ops} + N_{perftest}) \times 24hr \times 7d =$ **336** N-hr
- Cloud - Operate test environment during work hours:
  $(N_{ops} \times 24\ hr \times 7\ d) + (N_{perftest} \times 10\ hr \times 5\ d) =$ **218** N-hr

35% Savings from just the Test Environment!

Most Important Lesson: Cloud-Native Architecture

Cloud-native re-architecture is often cheaper than "forklift" (also known as *lift-and-shift*)

- Take full advantages of key cloud capabilities
  - Auto-scaling
  - Resiliency
  - Blue/Green Deployment
- Rearchitecture cost may be lower than you think
  - E.g., Serverless architecture



Cloud Architecture Game Changers

- Serverless architecture (Lambda, Step Function)
  - Reduces code to just business logic
- Containerization (Docker)
  - Portability (cross-platform, -OS, -cloud, …)
  - Deployability
- S3-aware/S3-native architectures?
  - Surprising efficiencies
  - Hmmm...