Use of commercial cloud for NASA Earth science data

C. Bond, L. Booker, D. Crumly, K. Gergely, L. Goldman, K. Heightley, J. Kovarik, A. Leon, S. Leslie, M. Savoie, A. Steiker
National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder

Overview

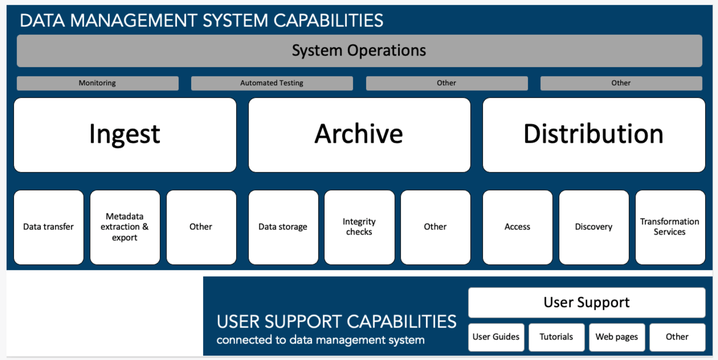
- NASA's Earth Science Data Systems (ESDS) Program is implementing a strategic vision to utilize commercial cloud environments for data management and access.
- The NASA Distributed Active Archive Center at NSIDC (NSIDC DAAC) is supporting this vision by establishing new cloud-based data systems and migrating data to the cloud.
- NSIDC DAAC manages NASA's snow, ice, and related data as one of 12 domain-focused DAACs across the country.

Cloud 101

- "The cloud" refers to servers that are accessed over the Internet, and the software and databases that run on those servers.
- By using cloud computing, users and organizations don't have to manage physical servers themselves or run software applications on their own machines.
- A number of commercial companies provide cloud services to support cloud computing.
- Amazon Web Services (AWS) is the commercial cloud used for NASA Earth science data – called Earthdata Cloud.

More than moving data

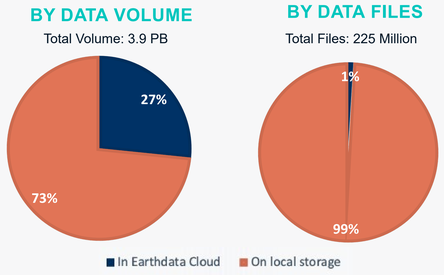
- Utilizing the commercial cloud for NASA Earth science data requires more than moving the data into the Earthdata Cloud.
- NSIDC DAAC must re-establish data management systems ingesting, archiving, and distributing data.
- Because the DAAC manages the authoritative data archive, the systems must:
  - Be robust and operationally reliable
  - Perform data integrity verification
  - Support a wide range of data: satellite, airborne, field measurements, and derived
- NSIDC DAAC must also modify our user support.
  - User support staff must learn new technologies and workflows
  - User resources must be developed or updated



Data management system capabilities which must be established in the NASA Earthdata Cloud

NSIDC DAAC data in the cloud

- Publicly released 34 data products from two NASA satellite missions in the Earthdata Cloud.
  - Ice, Cloud, and land Elevation Satellite-2 (ICESat-2)
  - Ice, Cloud, and land Elevation Satellite (ICESat)
- Additional data will be migrated in the coming months and years.



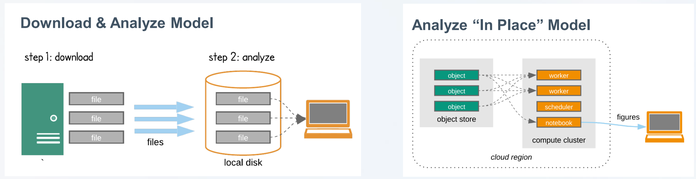
User adoption of cloud computing

- Will occur at different times for different users.
  - Some are already working in cloud computing environments
  - Some will use cloud computing once systems are provided to abstract away technical complexities
  - Some may never use cloud computing
- Requires learning new technologies and changing existing workflows and code
- Involves new cost models which need to be understood, budgeted for, and funded
- NSIDC DAAC is supporting users transitioning science workflows to the cloud through skill-building and community engagement efforts.
  - NASA Openscapes
  - Earthdata Cloud Cookbook (https://nasa-openscapes.github.io/earthdata-cloud-cookbook/)

Why is NASA utilizing commercial cloud

1 Easy access to data

- Data users will be able to access and analyze data directly in the cloud, making the need to download large volumes of data unnecessary.
- Users will continue to be able to freely download NASA data from Earthdata Cloud.



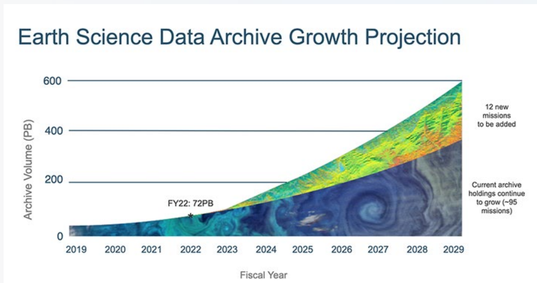
Credit: Open Architecture for scalable cloud-based data analytics. From Abernathey, Ryan (2020): Data Access Modes in Science.

2 Rapid deployment

- Users can bring their algorithms and processing software to the cloud and work directly with the data in the cloud.
- Cloud-managed services can be used to rapidly develop and scale analysis.

3 Scalability

- Earth science is and will continue to experience an exponential growth in data volumes.
- The size and use of cloud-hosted data archives can be expanded easily and rapidly as needed.



4 Reduced redundancy

- The use of a common infrastructure with shared cloud native services will reduce redundant tools and services.

5 Cost effectiveness

- NASA ESDS pays only for the storage and services used.
- Users rent large amounts of compute for short periods of time.