EOSDIS System of Systems Architecture circa 2018

- Distributed Active Archive Center System Functions
  - Ingest and Archive – Acquire and validate data products from multiple data providers, store multiple versions, ensure archive integrity.
  - Distribution – Enable users to learn about data products, format, theoretical basis. Enable users to search for and access data products. Provide subsetting, reproject and format conversion.
  - Data Providers
    - Satellite Mission Science Data Systems, Aircraft and Ground Validation Campaigns, Principle Investigators
    - Collect and deliver data products to DAAC for archive and distribution.
  - Enterprise Functions
    - Data Search and Discovery
    - Data Visualization

Concept of Operations and Purposes Objectives
- Describe how the DAACs will operate utilizing commercial compute services circa 2021-2026:
  - Data Visualization
  - Data Search and Discovery
  - Satellite Mission Science Data Systems, Aircraft and Ground Validation Campaigns, Principle Investigators
  - Collect and deliver data products to DAAC for archive and distribution.

Science Data Systems Operations Tenets
- Fully automated ingest, archive & distribution with 24x7 system availability; 8hr weekday support staffing.
- Efficient add/insertion of missions, campaigns and products through automation.
- More data services cross product domains.
- Integrated EOSDIS-wide responses for help with products, services & tools.

Roles and Responsibilities
- Personnel that will operate or control system resources and data.
- Activities envisioned to require authorization to access.
- Access may be restricted to specific system environments or data to limit exposure, reduce risk, and protect system integrity.

EOSDIS Component Systems and External Interfaces
- Data Provider: may be internal to EOSDIS such as SIPS or DAAC, or external such as a NASA Principle Investigator; may be from on-premise or in-cloud facility; DAAC on-premise facility may serve as proxy science data source.
- Data Consumer: may be internal to EOSDIS such as SIPS or DAAC, or external such as public user, may be on-premise or in-cloud. The interface is agreed to in advance and can represent a one-time or an ongoing active data transfer.
- Common Metadata Repository (CMR): catalogs all DAAC science data and services; metadata records are registered, modified, and accessed via tools such as the MMT or via standard APIs.
- Metadata Management Tool (MMT): allows metadata authors (data owners) to create, update, publish, view, delete and manage their metadata records in CMR.
- Earthdata Search Client: Web application allowing users to search, discover, visualize, and access NASA Earth science data products using CMR and GIBS.
- Earthdata Login: single sign-on user registration and user profile management system for public users to get Earth science data form any of the DAACs.
- Global Imagery Browse Services: provides standard image services from DAAC science data.
- Worldview: public Web tool for interactively browsing global imagery from GIBS.
- Data Preservation Archive: Secure copy of unique and irreplaceable science data and information at a separate location.
- Cumulus: a framework for data ingest and archive management configurable to perform acquisition, ingest, validation, pre/post-processing, metadata harvesting & creation, publication to CMR, data distribution and metrics reporting.

Major Functional Concepts and Scenarios

Data Management:
- Setup and retirement of ingest, archive and distribution workflows & services for data collections.
- Metadata curation, product guide development, Website landing page development.
- Setup test environments, build/test/checkout metadata and data service workflows.
- Cumulus Dashboard used to add/delete new dataset collections or individual granules.
- MMT or discipline-specific metadata tool used to add collection metadata to the CMR.
- Worldview used to test/check product imagery generation and delivery to GIBS.
- Test/checkout dataset and documentation backup and restore configuration.
- Control data access based on guidance from the Principle Investigator or science team.
- Monitoring, troubleshooting workflow interrupts and checking quality of data services.
- Auditing product granule availability, checking for data gaps, removing bad/duplicate data.
- Augmenting collection and granule metadata as necessary

Distribution:
- Transfers to and from user’s on-premise facility.
- The user is authenticated and authorized via Earthdata login.
- The user requests and downloads data products.
- Network throttling or circuit breakers may be implemented to control egress costs.
- Transfer rates are managed similar to DAAC on-premise bandwidth to the internet.
- Direct access to DAAC AWS archive via external AWS account within same AWS region.
- The user is authenticated and authorized via Earthdata login.
- Temporary access is granted with no egress costs in the cloud-to-cloud transfer.
- Utilization is monitor & controlled by data collection or by registered user.

Maintenance:
- DAAC Application Owners run their own instance of Cumulus on their own NGAP account.
- Deployments to operations accounts occur without downtime.
- User Acceptance Test environment is separate isolated NGAP account.
- Used by DAAC operators and selected Data Providers and Consumers.
- Acceptance test new releases on realistic workflows.
- Development teams will have NGAP sandbox accounts to work on new code.
- DAAC Application Owners ensure integration of the latest releases of NGAP and Cumulus.