NASA Earthdata Cloud

Mark McInerney
Deputy Project Manager - Technical
Earth Science Data & Information System (ESDIS)
June 2019

http://earthdata.nasa.gov
Earth Science Data Collections of the Earth Observing System Data and Information System (EOSDIS)

- Land
  - Cover & Usage
  - Surface temperature
  - Soil moisture
  - Surface topography

- Ocean Dynamics
  - Surface temperature
  - Surface wind fields & Heat flux
  - Surface topography
  - Ocean color

- Atmosphere
  - Winds & Precipitation
  - Aerosols & Clouds
  - Temperature & Humidity
  - Solar radiation

- Cryosphere
  - Sea/Land Ice & Snow Cover

- Human Dimensions
  - socioeconomic data (e.g., population, infrastructure, ..)

- Airborne
- International Space Station
- Field Campaigns
- International Partners

NASA Earthdata Cloud (EDC) – http://earthdata.nasa.gov
NASA’s EOSDIS provides end-to-end capabilities for managing NASA’s Earth science data from satellites, aircraft, field measurements, and various other programs.

EOSDIS is responsible for a data collection that is large in volume and projected to grow rapidly over the next several years.

NASA Earthdata Cloud (EDC) – http://earthdata.nasa.gov
NISAR Mission

- 80 TBs/day generation
- 400 TBs/day reprocessing
- 300 GB Granules
- 150 PBs @ 50 Gbps processing speed for months
Growth of Mission Data & Processing: Projected rapid archive growth and the need to effectively process significantly larger volumes of new mission data requires rethinking existing architectures.

Data Systems: More cost-effective, flexible, and scalable data system ingest, archive, and distribution solutions are needed to keep pace with new mission advancement.

Science Users: Significantly larger data volumes requires additional ways to access and utilize this data, with “Data Close to Compute” or Data Lake"
Earthdata Cloud (EDC)

- “Managed” commercial cloud for EOSDIS on AWS
- Improves the efficiency of NASA’s data systems operations – maintaining free/open data policy
- Designed for EOSDIS applications and mission data ingest, archive, distribution
- Increase opportunity for researchers and commercial users to access/process petabytes of data quickly without the need for data management.

NASA Earthdata Cloud (EDC) – http://earthdata.nasa.gov
Components of the Earthdata Cloud

NASA Earthdata Cloud (EDC) – http://earthdata.nasa.gov
Project Level Components and Core Elements
Project Level Components and Core Elements

NASA Earthdata Cloud (EDC) – http://earthdata.nasa.gov
Earthdata Cloud Platform Infrastructure

A multi-account, Infrastructure-as-a-Service (IaaS) cloud platform operating on AWS under a single top level “payer account”, providing shared cloud services and controls to EOSDIS.

- **Maximizes Autonomy**: Be a platform, not a gate; foster experimentation/innovation and support production needs of application owners.
- **Maximizes Flexibility**: Provide projects the freedom to implement solutions that fit their problem domains.
- **Shared Services & Controls**: Platform manages common shared services & controls to reduce duplication, system complexity, and cost across EOSDIS.

NASA Earthdata Cloud (EDC) – http://earthdata.nasa.gov
Common Services and Controls

1. **NASA-Approved Amazon Services** - vetted AWS and third-party SAAS services and process to add new. Focus is on using AWS cloud-native services.

2. **Code Deployment Services** - DevOps CICD Pipeline to security scan code, build, and deploy into EDC.

3. **Use of Infrastructure as Code** including re-useable template to define a multi-account ecosystem.

4. **cloudtamer.io Identity and Access Management** to:
   - Rotate AWS access keys
   - Apply session limits
   - Provide role-based access control.
User Access

1. User Authentication
   • Grant uniform access and experience to end users from multiple identity management systems (NAMS (SAML) / Internal directory / Active Directory)
   • Enforce and validate minimum authentication levels via two-factor authentication
   • Ensure that users only have access commensurate to the authentication type with cloudtamer.io
   • Secure PIV/Token login, NASA NAMS account provisioning

2. User Authorization
   • Control who views finances, who accesses AWS resources, and who manages finance
   • cloudtamer.io manages AWS IAM roles and policies at an organization level
Account Structure

- Single Payer Account using AWS Organization and Consolidated billing
- Multi-Account structure divided into NASA / Mission-defined organizational units
  - Isolation based on organizational units
  - Isolation based on application development, test, and production accounts
  - Isolation of Management and Security accounts from end user environment
  - Track AWS expenses to NASA organizations and funding sources
The Antideficiency Act (ADA) and Pay-as-You-Go

The ADA act prohibits federal agencies from obligations or expending federal funds in advance or in excess of an appropriation, and from accepting voluntary services.

Federal employees who violate the Antideficiency Act are subject to two types of sanctions: administrative and penal. Employees may be subject to appropriate administrative discipline including, when circumstances warrant, suspension from duty without pay or removal from office. In addition, employees may also be subject to fines, imprisonment, or both.
The cloudtamer.io account-level view

- Enforces individual AWS account-level budget through “budget caps”
- Provides account alert spend monitoring and budget control actions
- Allows for flexible access levels:
  - Top-level view for ESDIS management & business teams
  - Account view for APP-Admins & Developers as needed

NASA Earthdata Cloud (EDC) – http://earthdata.nasa.gov
Near-term Application and Data Onboarding

6 key projects in queue for production into the Earthdata Cloud

- **GHRC**: Global Hydrology Resource Center (GHRC) mission data into EDC production end of 2019
- **NISAR**: Alaska Satellite Facility (ASF) NASA-ISRO Synthetic Aperture Radar (NISAR) mission data into EDC production early 2022 (Dec 2021 launch)
- **SIR-C**: Alaska Satellite Facility (ASF) Spaceborne Imaging Radar-C (SIR-C) mission data into EDC production early 2020
- **SWOT**: Physical Oceanography (PO) Data Active Archive Center (DAAC) Surface Water Ocean Topography (SWOT) mission data into EDC production 2022
- **GITC** (Giovani +): NASA’s Goddard Earth Science Data and Information Services Center (GES DISC) Giovani visualization & analysis tool in EDC to support ESD analytics capabilities, pre-production environment 2019
- **GITC** (GITC +): ESDIS Global Imagery Browse Service in The Cloud (GITC), NGAP2.0 pre-production environment

Production

- **Sentinel-1**: ASF Sentinel-1 mission in production on Earthdata Cloud (NGAP2.0)
- **Common Metadata Repository**: ESDIS / EED Common Metadata Repository (CMR) in production on Earthdata Cloud (NGAP1.0)
- **Earthdata Search Client**: ESDIS / EED Earthdata Search Client (EDSC) in production on Earthdata Cloud (NGAP1.0)

ESDIS/DAAC dataset prioritization activity underway to determine mission data onboarding Oct 1, 2019

NASAs Earthdata Cloud (EDC) – http://earthdata.nasa.gov