Working with Cloud-Based NASA Earth Observations Data and Tools

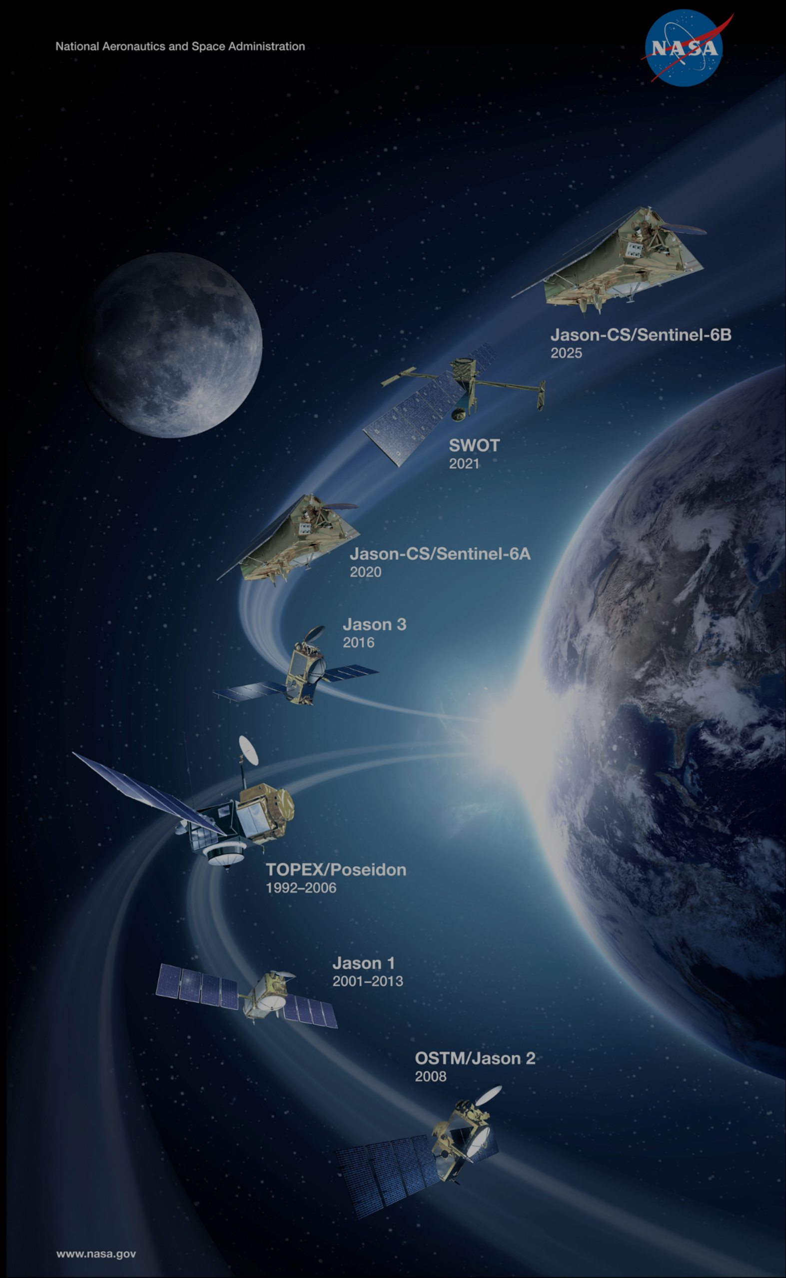
Presented by:

Physical Oceanography DAAC

National Snow and Ice Data Center DAAC

4 December 2020

AGU Fall Meeting Workshop



WELCOME!

Your Instructors & Presenters for the Workshop:

Catalina M. Oaida, NASA Jet Propulsion Laboratory, California Institute of Technology, PO.DAAC Applied Science Systems Engineer

Amy Steiker, NASA National Snow and Ice Data Center Distributed Active Archive Center (NSIDC DAAC), University of Colorado

Andrew P Barrett, National Snow and Ice Data Center, University of Colorado

Walt Meier, NASA National Snow and Ice Data Center Distributed Active Archive Center (NSIDC DAAC), University of Colorado

Jack McNelis, NASA Jet Propulsion Laboratory, California Institute of Technology, PO.DAAC Data Steward Engineer

Mike Gangl, NASA Jet Propulsion Laboratory, California Institute of Technology, PO.DAAC Project System Engineer

Luis Alberto Lopez, National Snow and Ice Data Center, University of Colorado

Stepheny Perez, NASA Jet Propulsion Laboratory, California Institute of Technology, PO.DAAC

Catalina Oaida

Amy Steiker

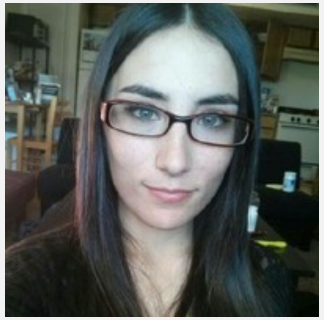
Andy Barrett

Walt Meier

Jack McNelis

Mike Gangl

Luis Lopez



Stepheny Perez

WORKSHOP MOTIVATION & SCOPE

Motivation

NASA's Earth Observing System Data and Information System (EOSDIS) is in the process of moving EOSDIS data to the cloud, motivated by rapid rate of data ingest into the EOSDIS archive. NASA remote sensing data from both upcoming (e.g. SWOT) and existing (e.g. Terra, Aqua, ICESat-2) missions will be available in the Earthdata cloud platform in the coming years. The paradigm shift from on-premise to cloud-based data distribution, and that from "download and analyze" to "analysis in place" present opportunities and challenges. Guiding users through this transition is of the upmost importance.

Scope

- Provide an overview of the EOSDIS evolution to the cloud, what it means to the Earth Observations data users, and
- Guide participants through interactive demos that highlight data discoverability, accessibility, and usability capabilities, on and off the cloud.

The demos run through science and applications user stories and workflows, which range across disciplines, including ocean, hydrology, cryosphere, and coastal.

Following the demos, participants will have the chance to further explore these cloud-based capabilities with Jupyter notebooks.

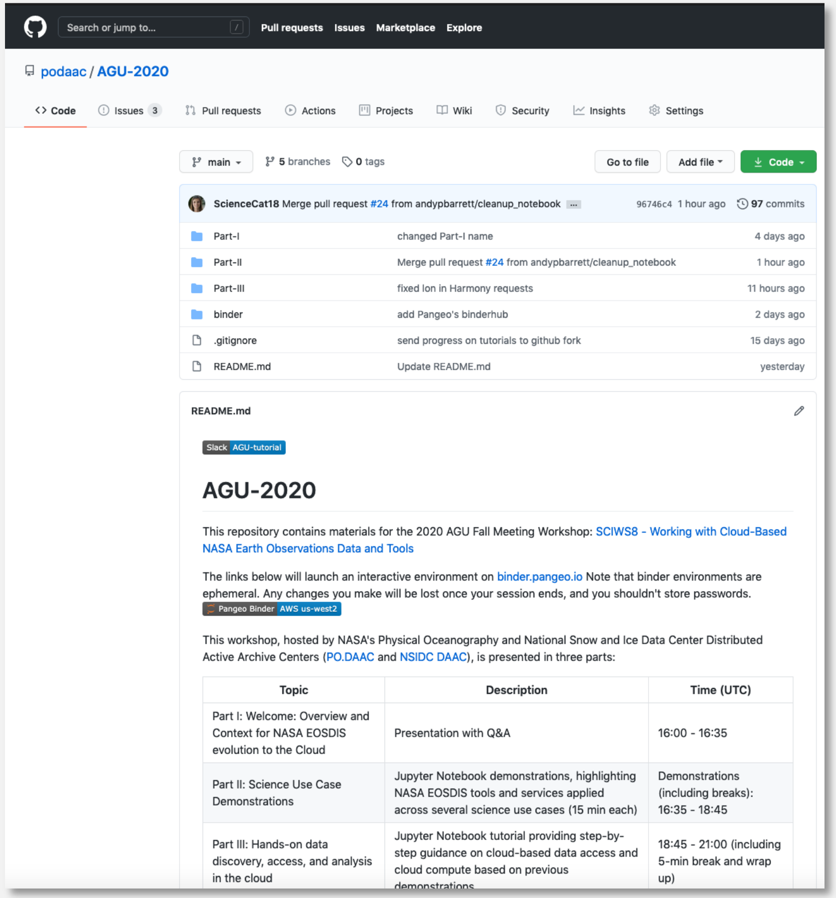
Participants should leave having a better understanding of how they can integrate cloud-based NASA data and services within their own work, whether working within or outside of the cloud

AGENDA – DETAILED

Topic	Duration (min)	Time (UTC)	Time (PST)
Welcome, Logistics	**5**	**16:00 – 16:05**	**8:00 – 8:05**
Part I: Overview of NASA EOSDIS evolution to the Cloud	**30**	**16:05 – 16:35**	**8:05 – 8:35**
Part II: Science User Stories Demos (18 min each, including Q&A)	**35**	**16:35 – 17:10**	**8:35 – 9:10**
Tutorial 1 – Explore Gulf of Mexico SSTs during the 2020 hurricane season (Search by user-defined geographic region of interest using shapefiles)	-	-	-
Tutorial 2 – Discover satellite derived sea surface temperature and ocean color along an ARGO drift track (Search for datasets coincident with a list of points)	-	-	-
Break	10	17:10 – 17:20	9:10 – 9:20
Part II: Science User Stories Demos (cont.) (18 min each, including Q&A)	**55**	**17:20 – 18:15**	**9:20 – 10:15**
Tutorial 3 – Discover, access and visualize ICESat-2 data using OpenAltimetry	-	-	-
Tutorial 4 – Coincident NASA remote sensing data over the Arctic ocean (Search and subset data by geographic bounding box from two data archives)	-	-	-
Tutorial 5 – Study Amazon Estuaries with Data from the Earthdata Cloud (Access cloud data without downloading locally)	-	-	-
Break	30	18:15 – 18:45	10:15 – 10:45
Part III: Hands-On workflow – *In-cloud data access and exploration*	**120**	**18:45 – 20:50**	**10:45 – 12:45**
Setup for Hands-On portion		18:45 – 19:00	10:45 – 11:00
Hands-On walk-through		19:00 – 20:15	11:00 – 12:15
Survey	10	20:15 – 20:25	12:15 – 12:25
Break (somewhere during hands-on, at good breaking point)	5	20:25 – 20:30	12:25 – 12:30
Continue Hands-On walk-through		20:30 – 20:55	12:30 – 12:55
Wrap-up, Q&A	**5**	**20:55 – 21:00**	**12:55 – 13:00**

LOGISTICS

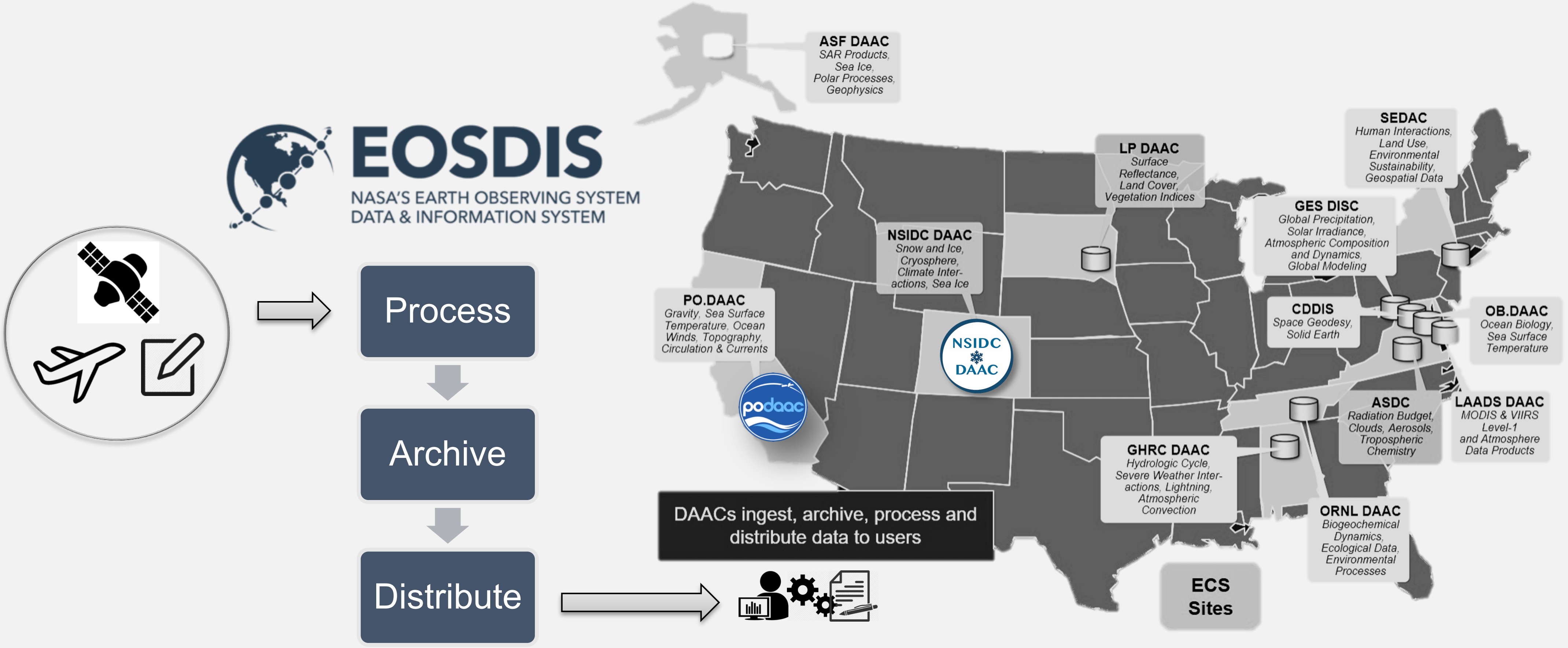
- If you haven't provided your Earthdata login *username* through the link in the email that was sent out, please provide it in the workshop slack channel and one of the instructors will add it to the data access list.
- Zoom platform for workshop (Q&A in Slack)
- Slack channel – *NASA DAAC AGU 2020 Tutorial*
  - Use for questions, discussion throughout workshop
  - There is a button in the github repo if you need quick access to our slack workspace
- Github repository: https://github.com/podaac/AGU-2020
- Hands-on portion (Part III) – python-based Jupyter notebooks
  - Cloud environment set up for you – Binder/JupyterHub
  - We will be walking through in-cloud data discovery & analysis workflow together, in phases
  - Will break out into ~5-person teams via Zoom Breakout rooms
    - Work within your group to work though the *Give it a Try*
    - Post any questions in the slack channel and an instructor will come in your breakout room to help resolve it.
  - We resume the hands-in workflow as a single group after each breakout



OVERVIEW & BACKGROUND

Part I

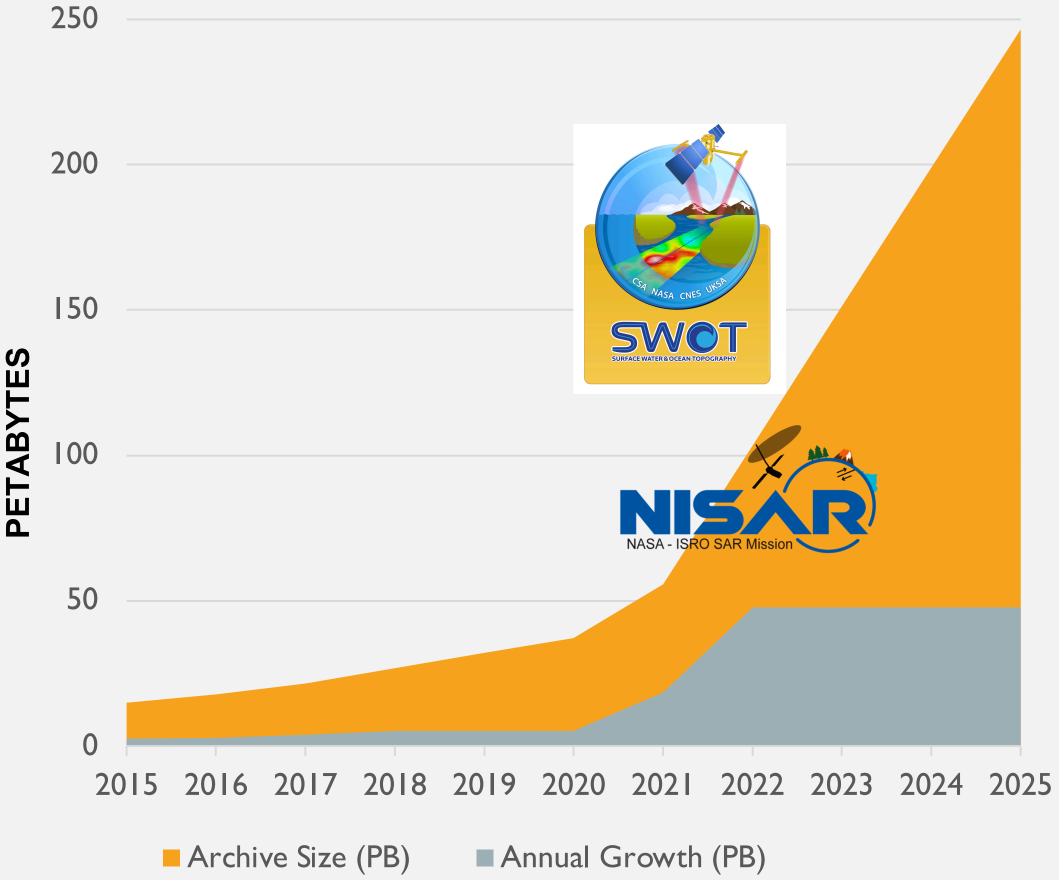
NASA DISTRIBUTED ACTIVE ARCHIVE CENTERS: PO.DAAC & NSIDC



PO.DAAC Mission:
Preserve NASA's ocean and climate data and make these universally accessible and meaningful.

NSIDC DAAC Mission:
Advancing knowledge of Earth's frozen regions, NSIDC manages and distributes scientific data, creates tools for data access, supports data users, performs scientific research, and educates the public about the cryosphere.

NASA EARTHDATA EVOLUTION TO THE CLOUD
BACKGROUND & MOTIVATION



	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
Archive Size (PB)	15	17.7	21.6	26.8	32	37.2	55.6	103.4	151.1	198.9	246.6
Annual Growth (PB)	2.6	2.8	3.9	5.2	5.2	5.2	18.4	47.7	47.7	47.7	47.7

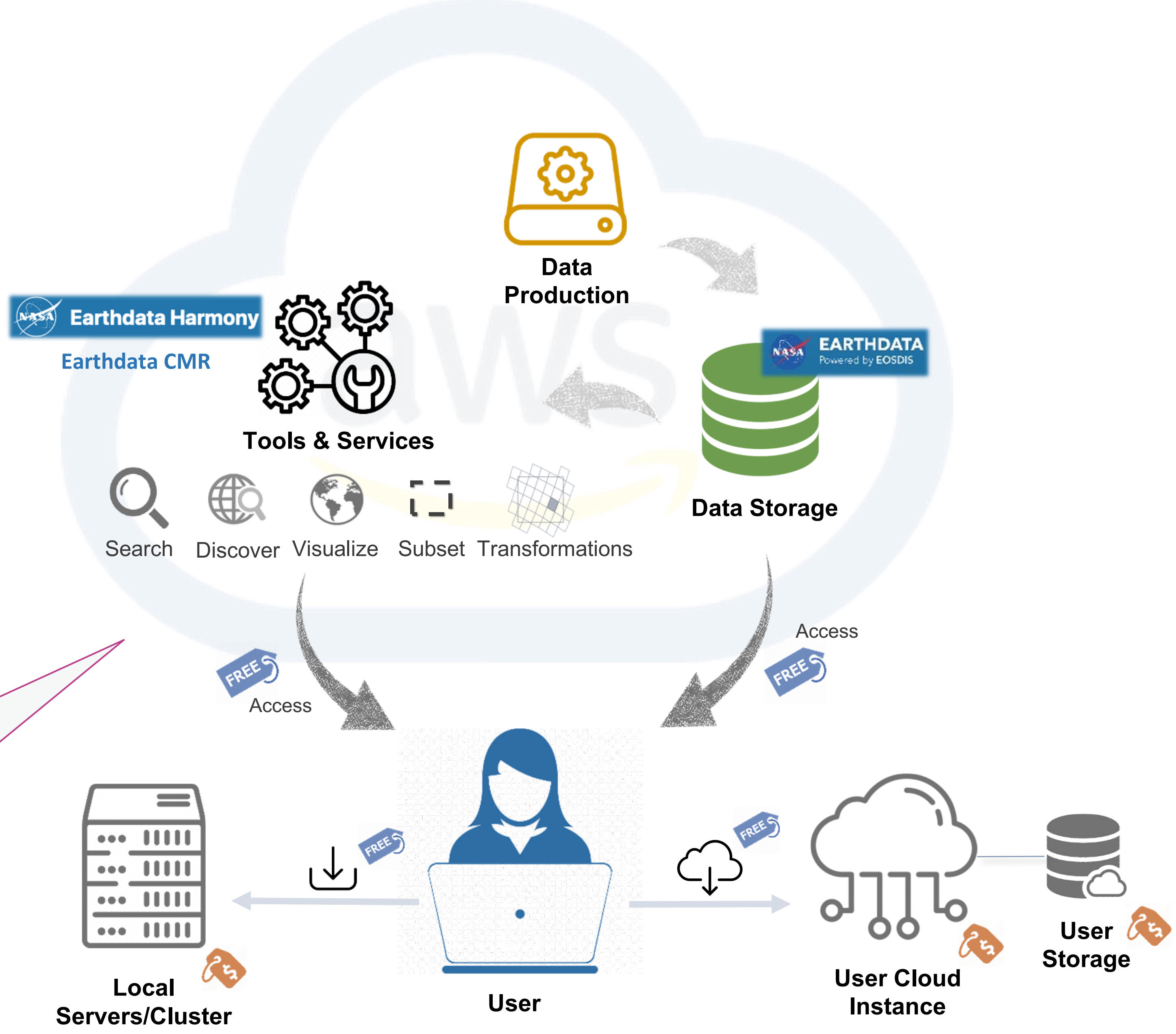
- Manage Large Data Volumes
- Performant and Scalable Services
- Co-locate synergistic datasets
- User analysis next to the data

But how will that affect you?

EARTHDATA CLOUD PARADIGM

Co-location

- Data
- Tools & Services
- Analysis



Big Data Challenge & Solutions

- Maintain DAAC level of service to user, by leveraging scalability
- Minimizes amount of data user needs to handle
- Makes data more analysis ready on behalf of user

EARTHDATA CLOUD PARADIGM

Which paradigm is right for me, as the data user?

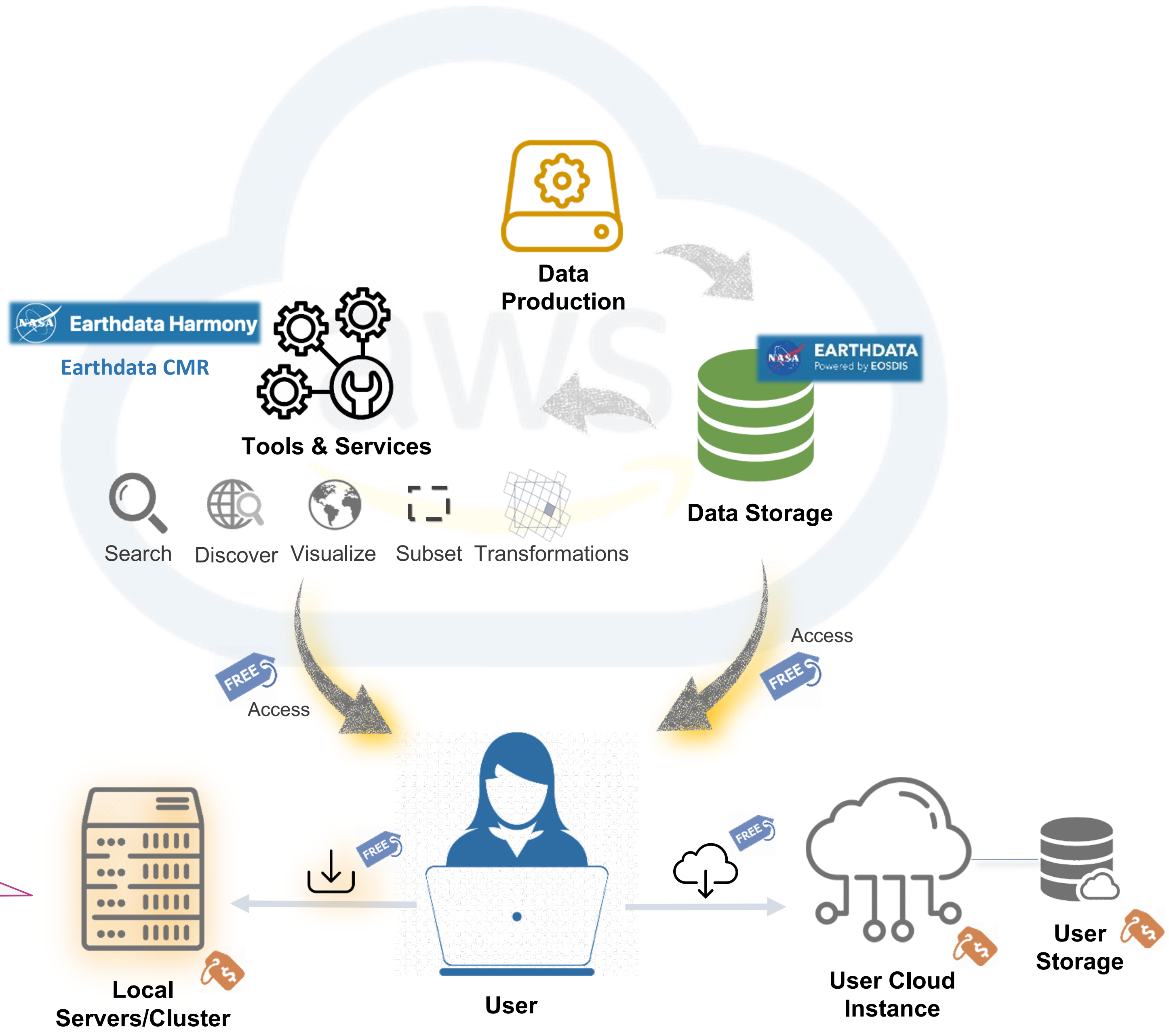
- Download and analyze
- Cloud computing

Think about your workflow:

- Which data do you need?
- What do you do with it?

!

- **Download time**
- **Storage**
- **Processing, compute power**

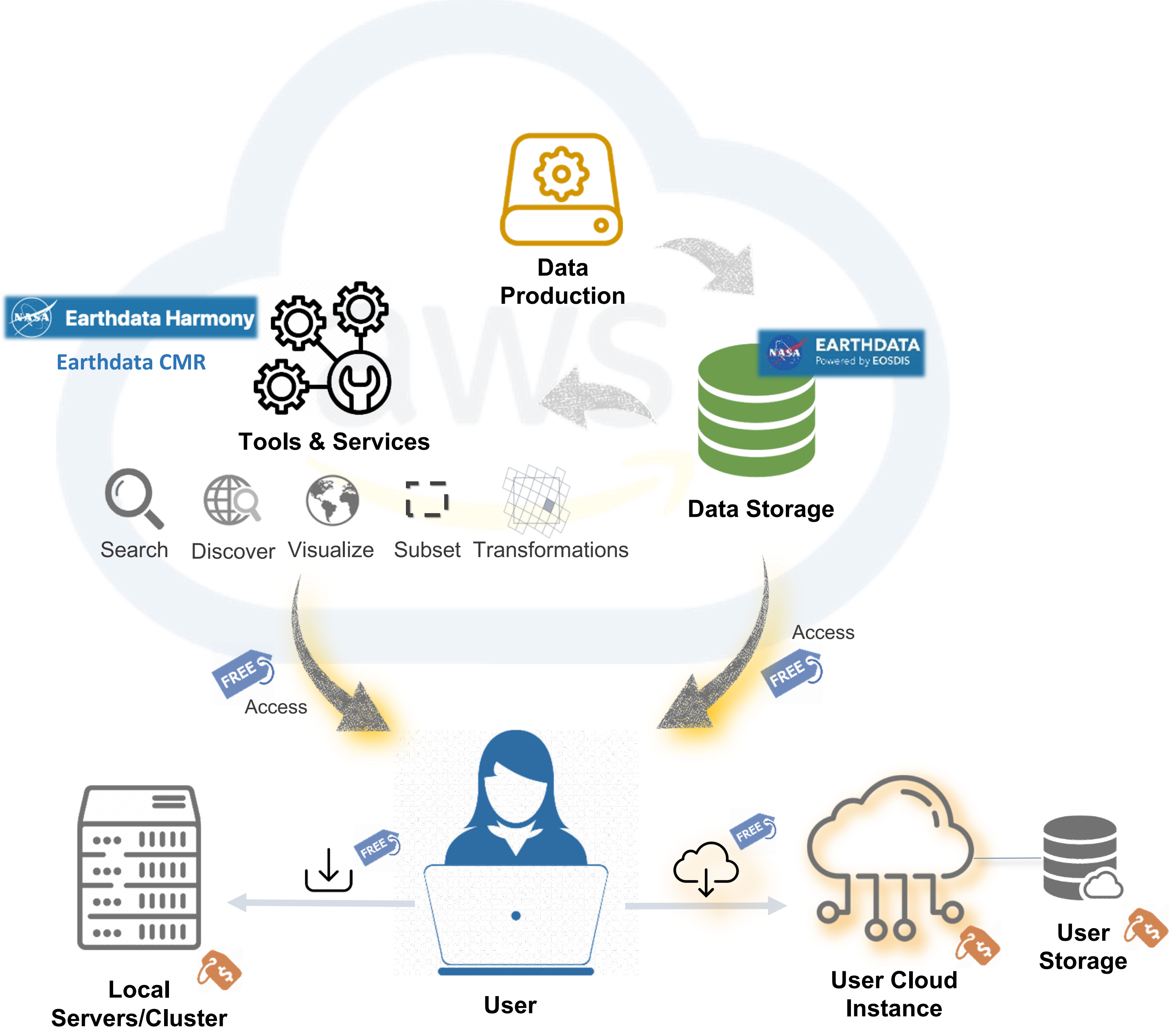


EARTHDATA CLOUD PARADIGM

Co-location

- Data
- Tools & Services
- Analysis

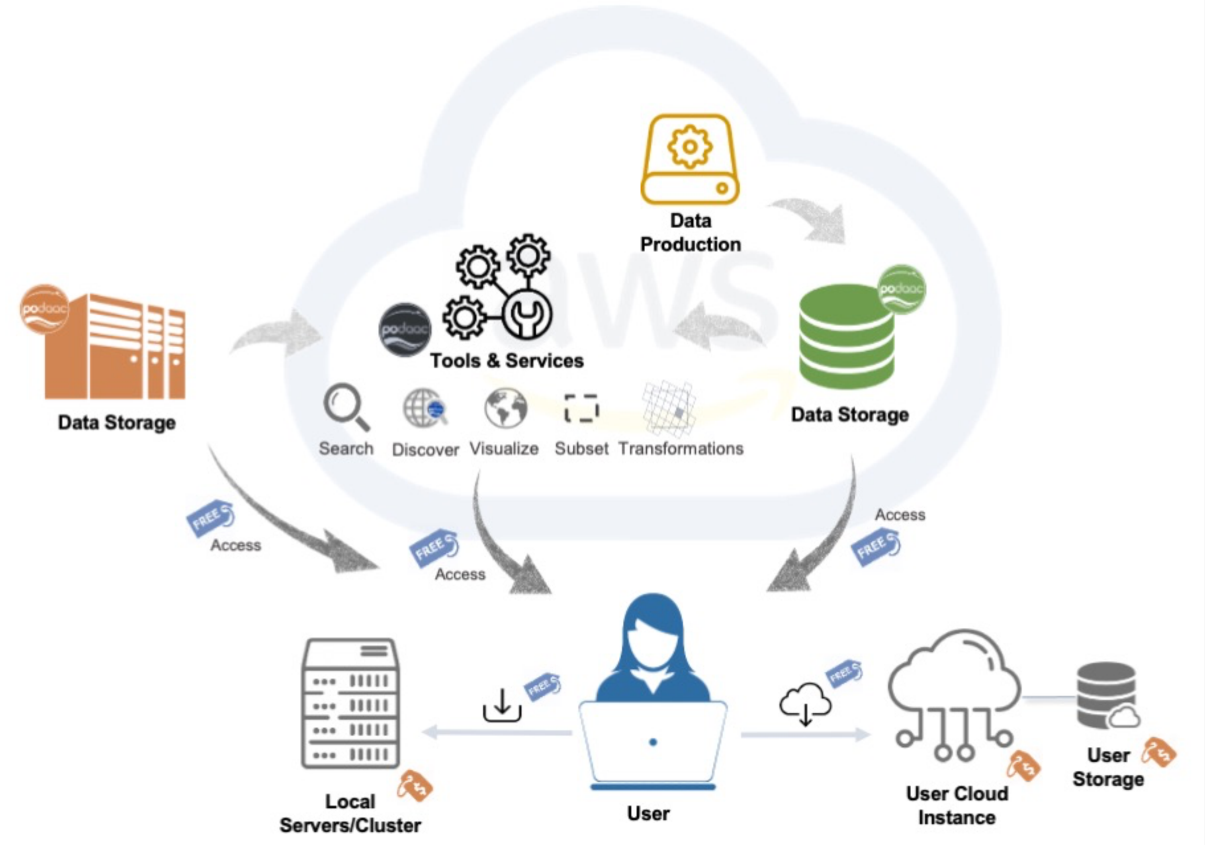
Offers opportunity to explore new ways of using Earth observations and perform science research and applications with big data



CLOUD PATHFINDER DATASETS

- Cloud Pathfinder Datasets are select PO.DAAC datasets that are among the first to be made available from the Earthdata Cloud.

- During the next few months (transition period), these Pathfinders are available both in the Operational archive (PODAAC provider) and in the Earthdata Cloud (POCLOUD provider).
  - To download data from PODAAC Operational archive, only need an Earthdata login account (anyone with an EDL can download).
  - To access and download data from POCLOUD archive, your Earthdata login (username) needs to be added to a restricted list for early access, for the time being (during migration period).
- For this workshop, your Earthdata login has been added* to the early access list, such that you have access to the Pathfinders in the Earthdata Cloud.
- Download from either archive is and will continue to be free, per NASA's free data policy.



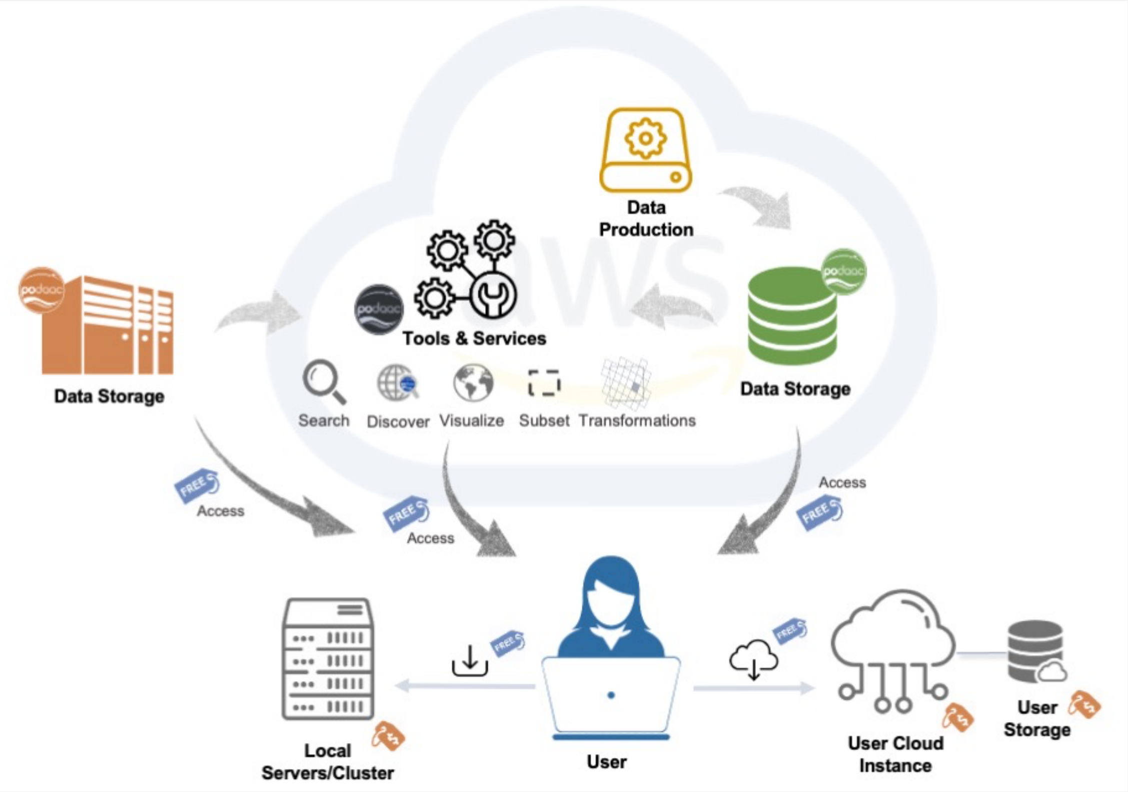
CLOUD PATHFINDER DATASETS

- PO.DAAC Cloud Pathfinder Datasets available now in the Earthdata Cloud:
  - SMAP L3 (e.g. Salinity)
  - GRACE-FO L3 (e.g. LWE)
  - MODIS Aqua/Terra L2 (e.g. SST)

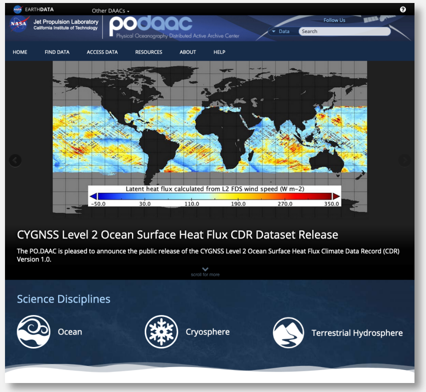
These are the datasets we'll be exploring today, along with ICESat-2 data from NSIDC DAAC outside of Earthdata Cloud

- Additional Pathfinders will become available in 2021, including
  - Merged Altimetry
  - Jason 1

- In 2021, a phased transition for all PO.DAAC datasets
  - Once operational in the Earthdata Cloud, these data will be able to be accessed and downloaded without the early access restrictions we're working with today.



DATA DISCOVERY & ACCESS

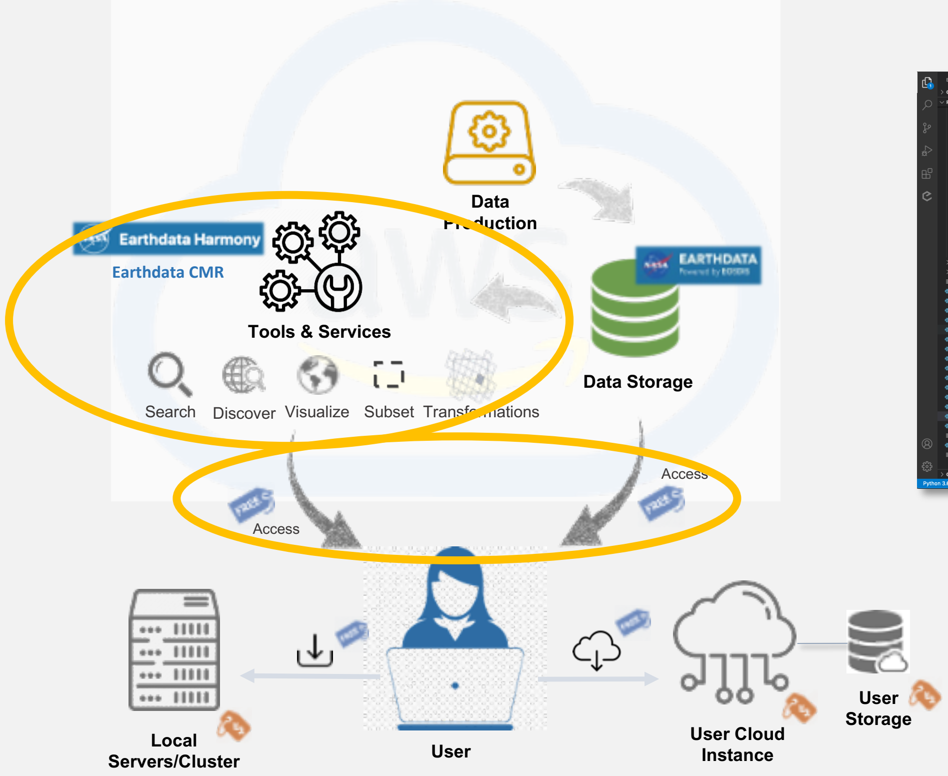


https://podaac.jpl.nasa.gov/

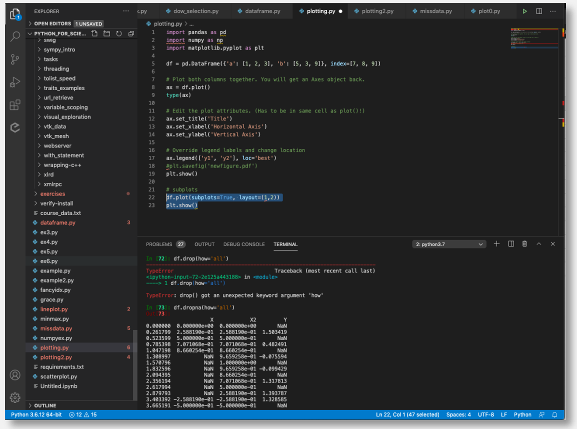
User Interface/ Portals

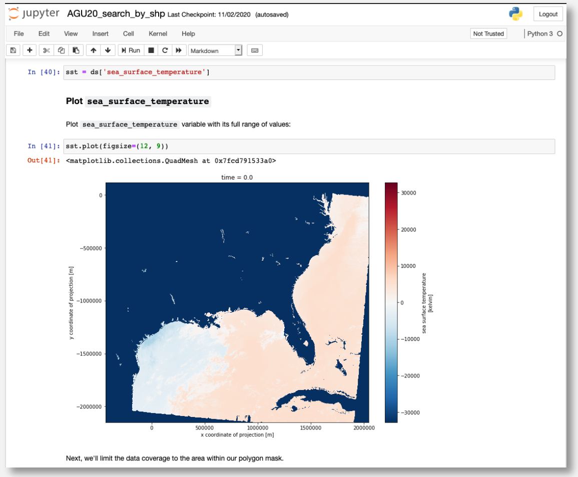
PO.DAAC Drive | OPeNDAP | THREDDS | Web Services | SOTO

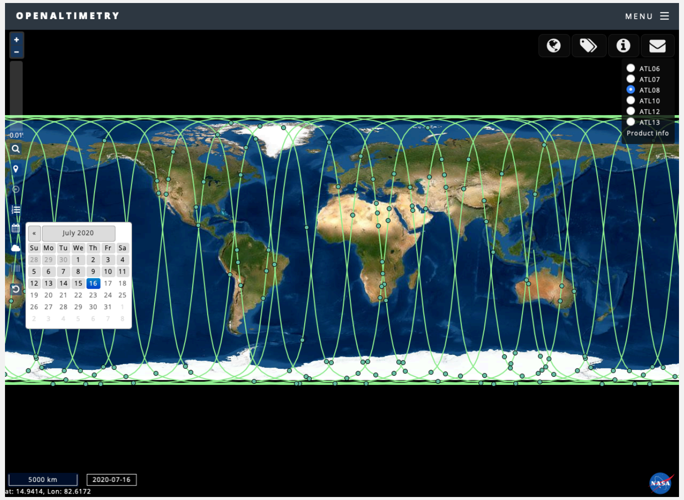
HiTIDE (L2 Subsetting) | LAS (L3 Subsetting) | PO.DAAC GitHub | Earthdata Search



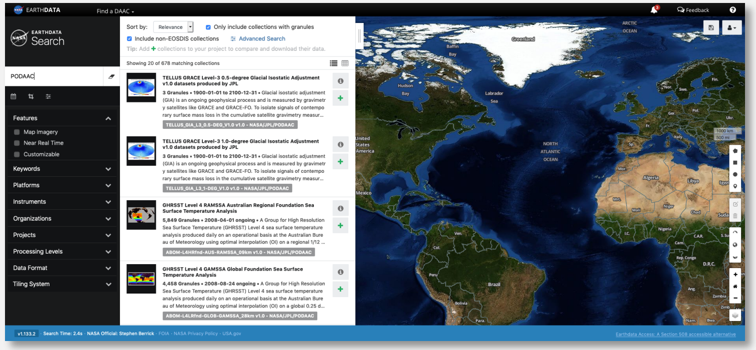
Programmatically APIs



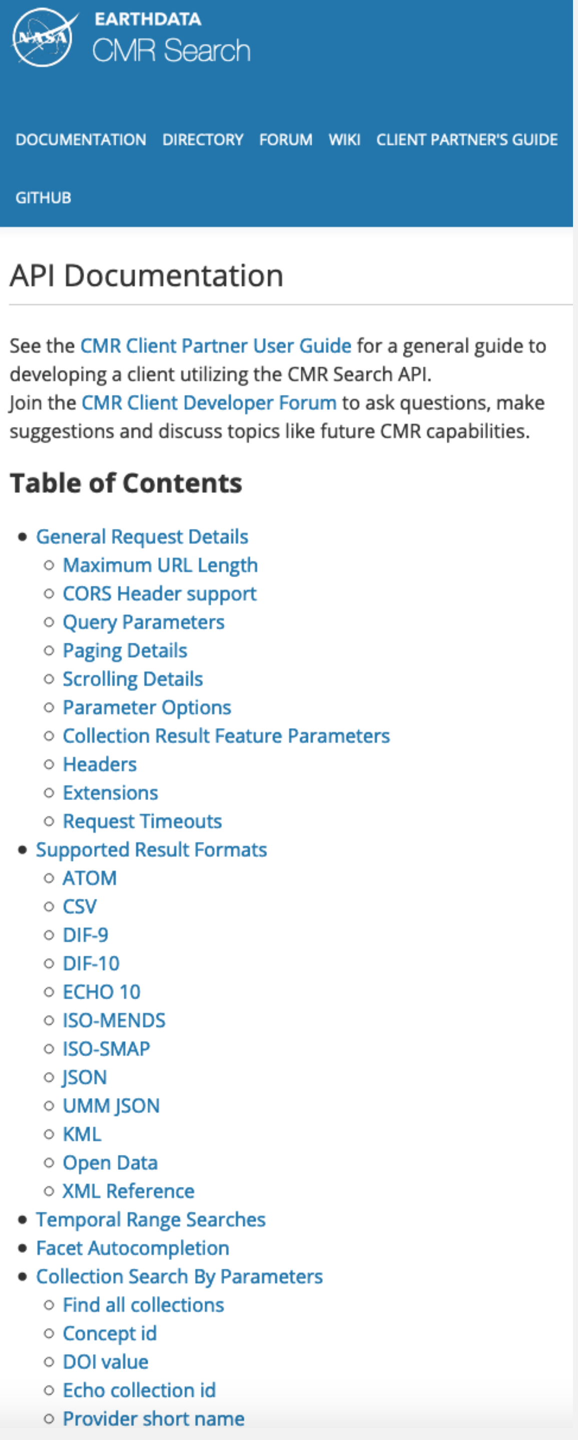




https://openaltimetry.org/



https://search.earthdata.nasa.gov/search



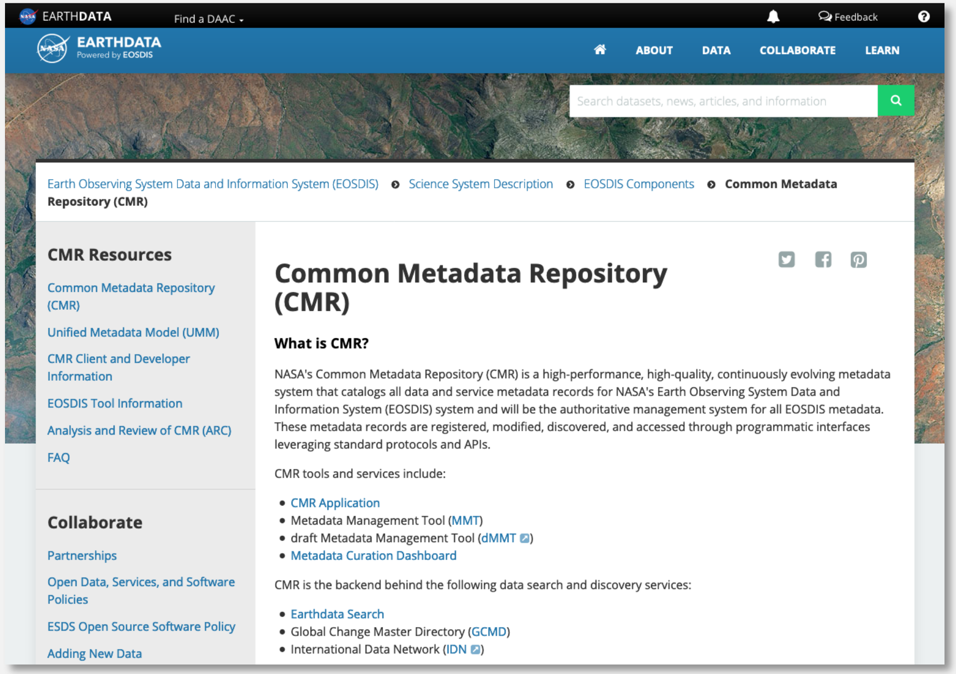
COMMON METADATA REPOSITORY (CMR)

What is CMR?

- NASA's Common Metadata Repository (CMR) is a high-performance, high-quality, continuously evolving metadata system that catalogs all data and service metadata records for NASA's Earth Observing System Data and Information System (EOSDIS) system and will be the authoritative management system for all EOSDIS metadata. These metadata records are registered, modified, discovered, and accessed through programmatic interfaces leveraging standard protocols and APIs.

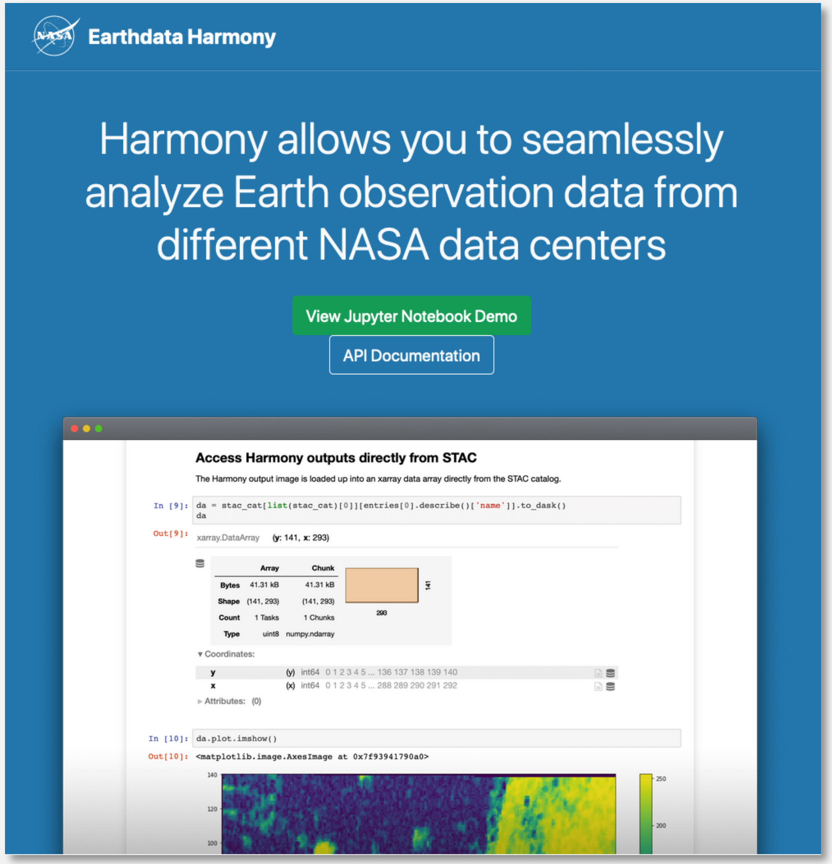
In today's workshop, we will be using the **CMR APIs to search** for datasets of interest, using several parameters, such as:

- Collection short name
- Temporal range
- Data center/provider
- Token (allows for Cloud archive early access)



https://earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/cmr

EARTHDATA HARMONY



https://harmony.earthdata.nasa.gov/

- ✓ **Consistent access patterns to EOSDIS holdings** make cross-data center data access easier
- ✓ **Data reduction services** allow users to request only the data they want, in the format and projection they want
- ✓ **Analysis Ready Data** and cloud access will help reduce time-to-science
- **Community Development** helps reduce the barriers for re-use of code and sharing of domain knowledge

Services

Users should be able to seamlessly analyze data from different NASA data centers in ways previously unachievable. Harmony aims to increase usage and ease of use of EOSDIS' data, focusing on opportunities made possible by cloud-accessible data.

Transformations can be performed using one of several Open Geospatial Consortium (OGC)-inspired APIs. Each API requires a collection concept ID CMR, and then transformations can be performed using `https://harmony.earthdata.nasa.gov/{collectionId}/ogc-api-coverages/1.0.0` or `https://harmony.earthdata.nasa.gov/{collectionId}/wms`
All users will need an Earthdata login account from Earthdata Login in order to perform transformations.

LEARNING OBJECTIVES

Upon completion of this workshop, you will have a better understanding of what the new cloud-based paradigm for data archiving, distribution, and particularly data access and use would mean for you, and your science or application workflow:

- Examine the changes, impacts, and opportunities provided by the Earthdata Cloud infrastructure, including cloud vendor (AWS) information, cost, and compute resources.
- Reflect on how the Earth and Space science, data science and informatics communities are evolving, including the acquisition, archiving, distribution, and use of big data, and how that evolution impacts scientific research and application of Earth observations.
- Investigate user stories across ocean, hydrology, and cryospheric science disciplines utilizing NASA EOSDIS capabilities within python-based Jupyter notebooks.
- Select and compare data transformation services and access methods within and outside of the Earthdata Cloud.
- Execute programmatic data access queries, basic GIS operations, plotting, and direct in-region cloud access using open source Python libraries.
- Identify where and when Earthdata Cloud components are implemented across data discovery, subsetting, access, and analysis/compute workflows.
- Develop new strategies for leveraging and integrating Earthdata Cloud capabilities within your own work.
- Identify resources, including the Earthdata Cloud Primer (https://earthdata.nasa.gov/learn/user-resources/webinars-and-tutorials/cloud-primer) for getting started with Amazon Web Services outside of the Workshop to access and work with data with a cloud environment.

SCIENCE USE CASE DEMOS

Part II

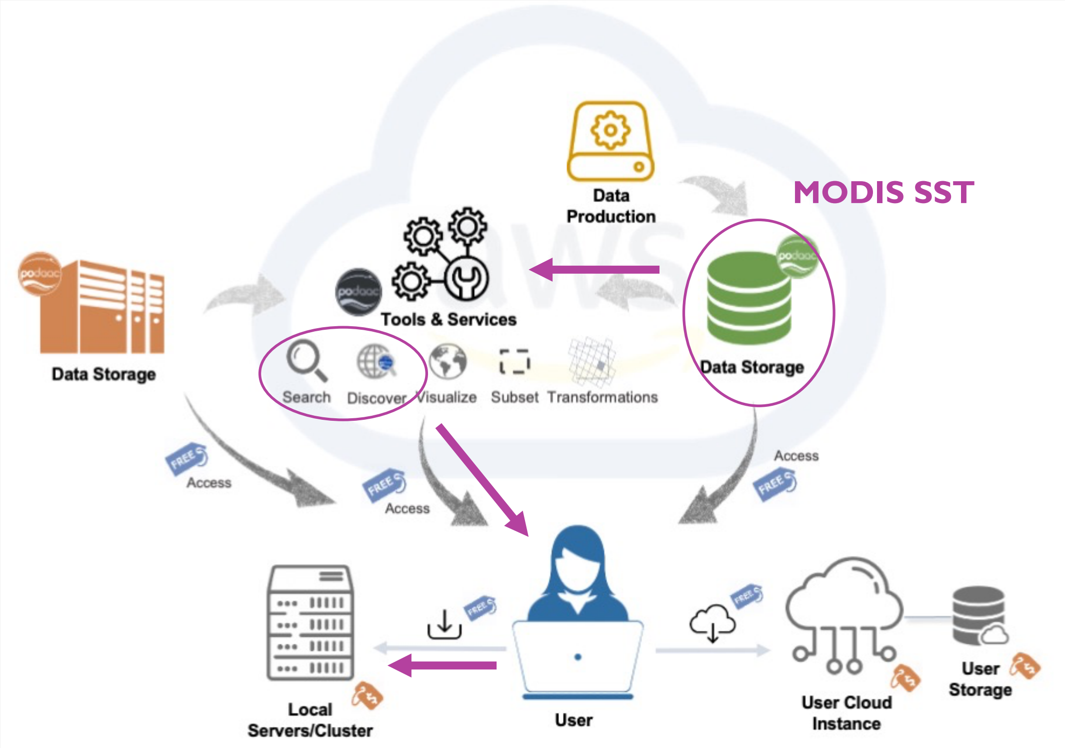
PART II - SCIENCE USE CASE DEMOS

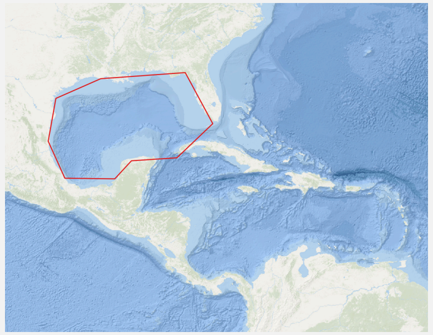
(18 min each, total of 90 min, plus 10 min break)

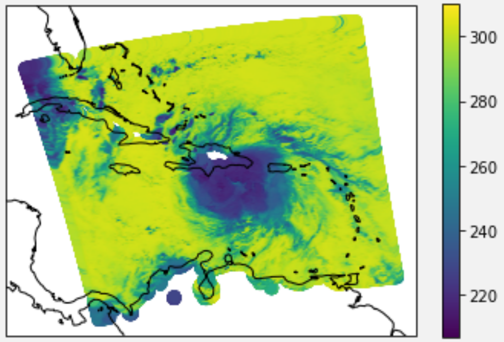
- **Tutorial 1**: Explore Gulf of Mexico SSTs during the 2020 hurricane season (Search by user-defined geographic region of interest using shapefiles)
- **Tutorial 2**: Discover satellite derived sea surface temperature and ocean color along an ARGO drift track (Search for datasets coincident with a list of points)
- *Break (10 min)*
- **Tutorial 3**: Discover, access and visualize ICESat-2 data using OpenAltimetry
- **Tutorial 4**: Coincident sea ice height and SST data over a melt pond in the Arctic ocean (Search and subset data by geographic bounding box from two data archives)
- **Tutorial 5**: Explore hydrology datasets in the Amazon river basin's estuary, in a cloud environment

TUTORIAL 1: Explore Gulf Of Mexico SSTs During The 2020 Hurricane Season

Tutorial 1:	**Explore Gulf of Mexico SSTs during the 2020 hurricane season**
User Story:	As an oceanographer, I want to explore the Gulf of Mexico SST response during the 2020 Atlantic hurricane season. I want to use my own shape file (polygon) to do the spatial search, and download only the data that meets my spatial and temporal request, so that I can perform further analysis locally.
Learning Objective	* Earthdata Login Authentication (for download access from Earthdata data archive) * Search CMR for collection and granule IDs, using the collection shortname and provider * Download a file from the PO.DAAC (Earthdata) cloud archive to local computer and preview the data * Search a collection by user-provided shapefile (ESRI shp) and temporal range * Download the first file (from the PO.DAAC cloud archive to local computer) and preview subset * Download all data (from the PO.DAAC cloud archive to local computer) based on shp and time search criteria
Functionality	Search by user-defined shapefile and download from cloud to local
Community of interest	Hydrology (e.g. watersheds), Oceanography, Coastal
Datasets	MODIS L2 SST
Service	CMR API (search by user-defined shapefile)
Cloud relevancy	Data & service operate in the cloud, on the backend. More efficient service to user on the front end (as we enter era of Big Data). Help user narrow down the amount of data needed for a given need. Especially relevant when working with big data volumes.
Resources	https://github.com/podaac/AGU-2020/blob/main/

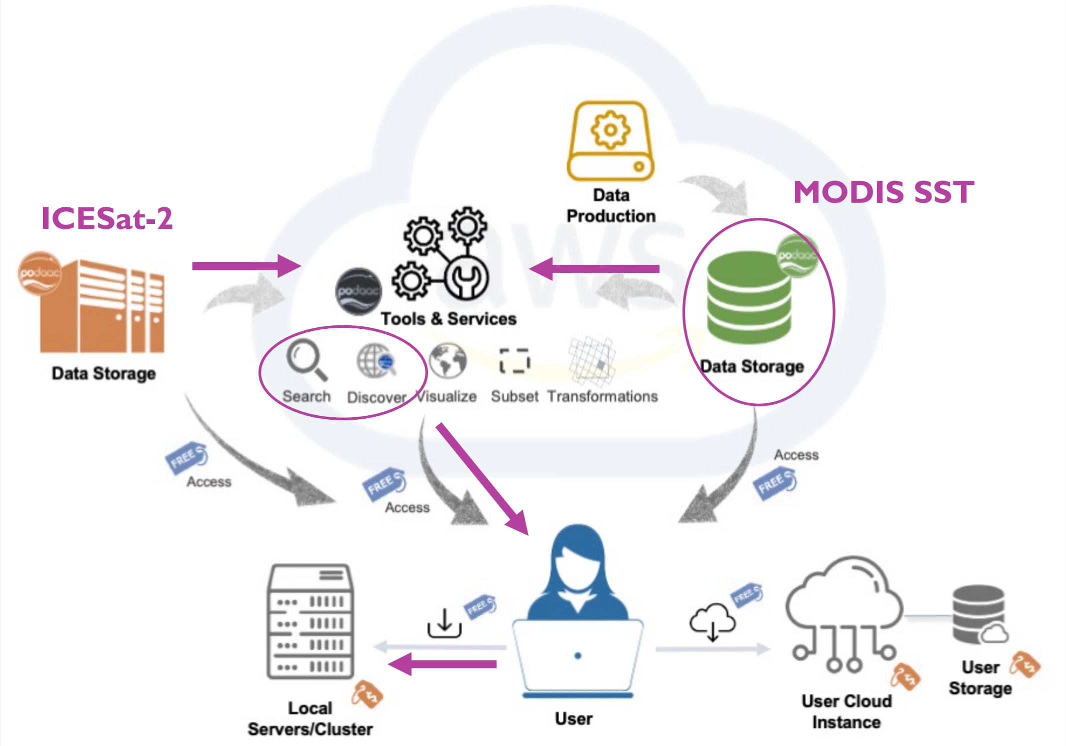


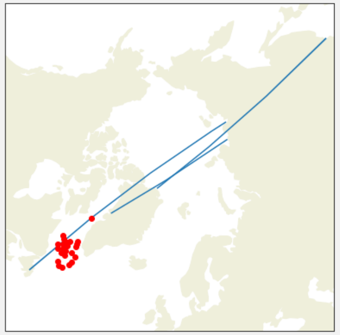




TUTORIAL 2: Sea Surface Temperature And Ocean Color Along An ARGO Drift Track

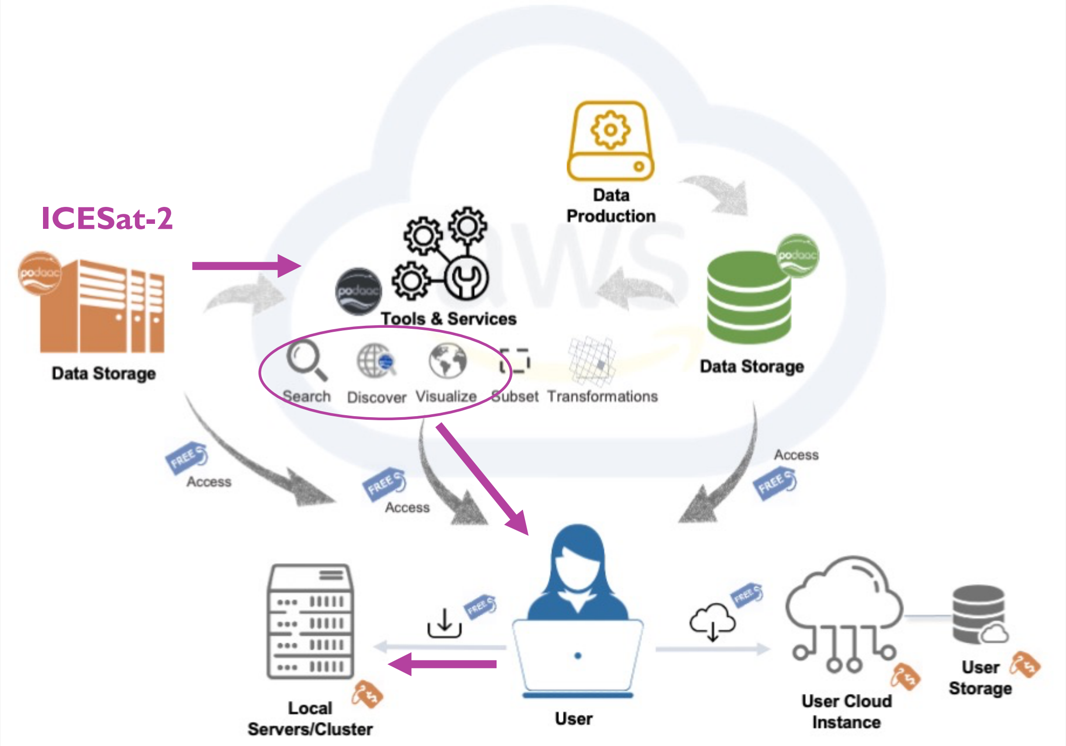
Tutorial 2:	**Sea surface temperature and ocean color along an ARGO drift track.**
User Story:	A physical oceanographer is interested in satellite derived sea surface temperature and ocean color along an ARGO drift track. Similar use cases would be to select data coincident with a cruise, with ice mass balance buoys in Arctic and Antarctic, or the MOSAIC experiment. Although this search could be done use EarthData search, this approach is not easily reproduceable, unless you save the search in some way such as taking a screenshot for example. Reproduceability is critical if you need to completely redo your analysis yourself, or if others want to recreate your reanalysis. By capturing the search in code, either in a notebook such as this one or in a script, you or anyone else can reproduce the search and any subsequent analysis.
Learning Objective	* Convert a list of coordinates into a GeoJSON file. * Write a query for the NASA CMR API. * Submit the query and interpret the response. * Order datasets returned by the query. * Visualize the results.
Functionality	Search by point-based locations.
Community of interest	Hydrology (e.g. river gauges), Oceanography (e.g. ARGO floats), Coastal (buoy)
Datasets	ICESat-2, (MODIS L2 SST)
Service	CMR API (search by user-defined shapefile)
Cloud relevancy	Data & service operate in the cloud, on the backend. More efficient service to user on the front end (as we enter era of Big Data). Help user narrow down the amount of data needed for a given need. Especially relevant when working with big data volumes.
Resources	https://github.com/podaac/AGU-2020/blob/main/





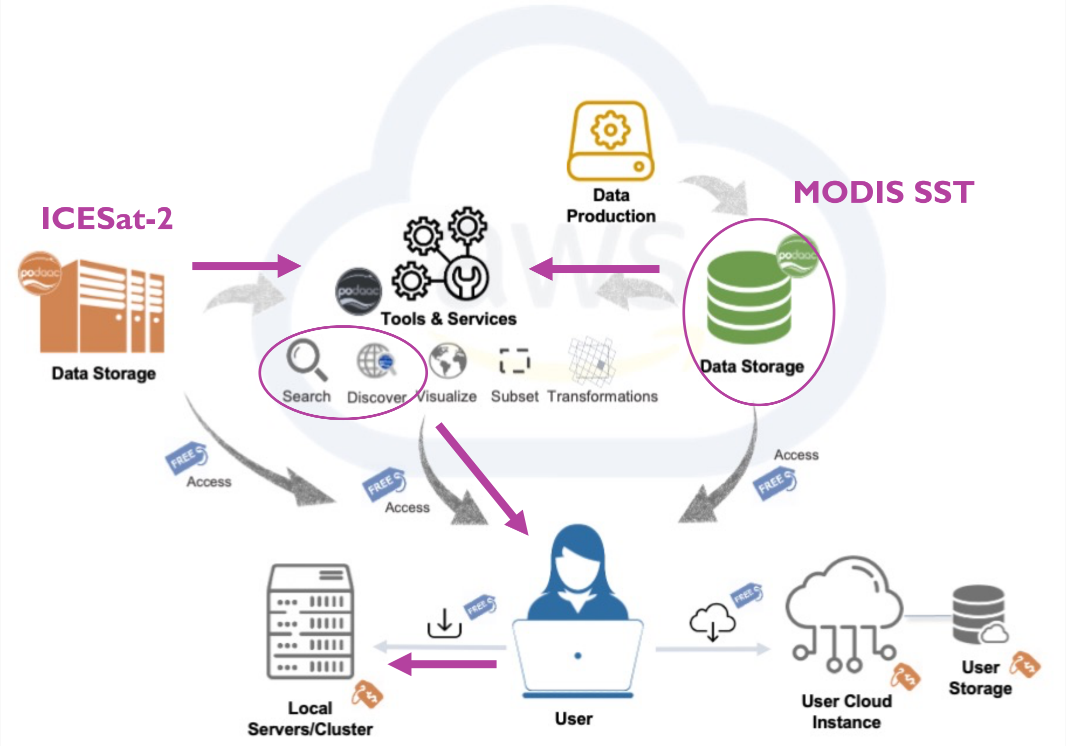
TUTORIAL 3: Open Altimetry

Tutorial 3:	**Exploring and visualizing ICESat-2 data with Open Altimetry**
User Story:	A Greenland ice sheet scientist is looking for ICESat-2 tracks that pass over their glacier of interest. She would like to look at potential tracks and quickly visualize the data to determine if the data will be useful.
Learning Objective	To familiarize oneself with the functionalities of Open Altimetry
Functionality	Data exploration and visualization
Community of interest	ICESat-2 users; researchers interested in sea ice, ice sheets, and land elevation, vegetation canopy height, and ocean and inland waters.
Datasets	ICESat and ICESat-2
Service	Open Altimetry server
Cloud relevancy	Exploration and visualization is done through the service, without any download (data from tracks can be downloaded if desired).
Resources	https://openaltimetry.org https://github.com/podaac/AGU-2020/blob/main/



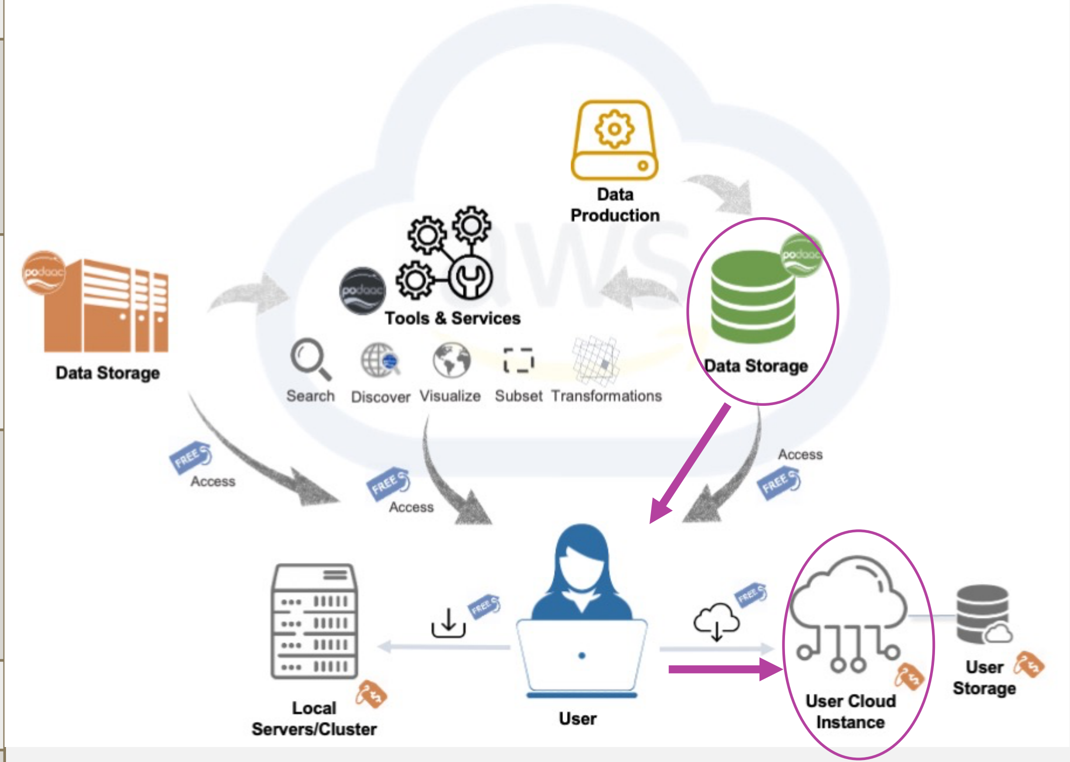
TUTORIAL 4: Coincident Sea Ice Height And SST Data Over A Melt Pond In The Arctic Ocean

Tutorial 4:	**Coincident sea ice height and SST data over a melt pond in the Arctic ocean**
User Story:	A climate modeler is interested in studying the effects of melt ponds on sea ice albedo and other changes in Arctic ocean heat balance and therefore needs to analyze multiple sea ice and ocean variables coincident over the same time and Arctic Ocean melt pond region.
Learning Objective	* Select NASA Earthdata sea ice and sea surface temperature datasets * Identify data file size and availability over time and geographic bounding box * Subset and download data, from both Earthdata Cloud and on-premise resources, using an Application Programmatic Interface (API) * Plot and compare coincident data values
Functionality	Search and subset by geographic bounding box across multiple data products
Community of interest	Cryosphere, Oceanography
Datasets	ICESat-2 Level-3A Sea Ice Height, MODIS Level-2 Sea Surface Temperature
Service	NSIDC subsetting API, PO DAAC Level-2 subsetter via Earthdata Harmony
Cloud relevancy	Help narrow down the amount of data needed for a given need. Especially relevant when working with big data volumes. Some data is cloud-based, some is still on-premise; you can work with both types of data regardless of whether they are in the cloud or not.
Resources	https://github.com/podaac/AGU-2020/blob/main/



TUTORIAL 5: Explore Hydrology Datasets In The Amazon River Basin's Estuary, In A Cloud Environment

Tutorial 5:	**Explore relationships between river height and land water equivalent thickness in the Amazon river basin's estuary.**
User Story:	This tutorial uses multiple satellite data products to analyze the relationships between river height and land water equivalent thickness in the Amazon river basin's estuary. Land water equivalent (LWE) thickness supports the observing of seasonal changes in water storage around the river. When discharge is high, the change in water storage will increase, pointing to a wet season.
Learning Objective	* Search for LWE (GRACE/GRACE-FO) and river discharge data (MEaSUREs Pre-SWOT) * Access LWE dataset in Zarr format from Earthdata Cloud (AWS) * Access discharge dataset from PODAAC on prem data pool * Subset both, plot and compare coincident data
Functionality	* Find metadata about NASA EOSDIS data holdings using the CMR Search API * Collections (datasets) and Granules (files) * Submit Zarr reformatter service requests to the Harmony API * Reformat native netCDF4 to Zarr, staged in S3 * Analyze data from GRACE/GRACE-FO and MEaSUREs Pre-SWOT using xarray * Draw static and animated plots of the data
Community of interest	Hydrology, Coastal
Datasets	JPL GRACE and GRACE-FO Mascon; SWOT Hydrology GRRATS Daily River Heights and Storage Version 2 (MEaSUREs Pre-SWOT)
Service	CMR API, Harmony API (zarr reformatter)
Cloud relevancy	* Leverage cloud processing services through the Harmony API * Work with cloud optimized formats such as Zarr. Open Zarr datasets staged on S3 in xarray
Resources	https://github.com/podaac/AGU-2020/blob/main/



HANDS-ON DATA DISCOVERY, ACCESS, AND ANALYSIS IN THE CLOUD

Part III

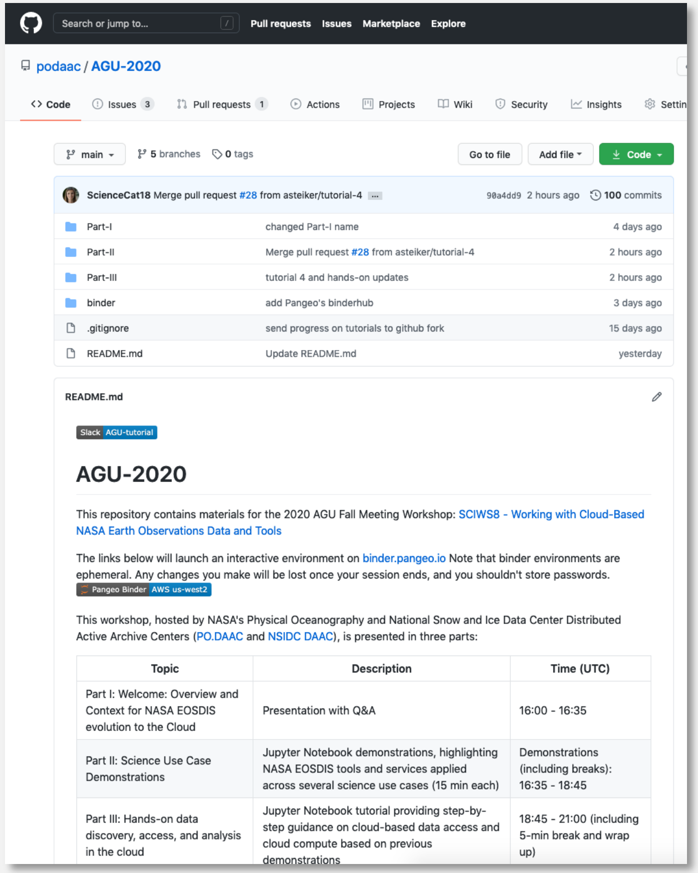
OUTLINE OF WORKFLOW
Amazon River Freshwater Discharge Impacts On Coastal Waters

Hands-on tutorial of AWS in-region access of NASA Earthdata products

- This notebook provides a basic end-to-end workflow to interact with data "in-place" from the NASA Earthdata Cloud, by accessing AWS s3 locations provided by NASA Harmony outputs without the need to download data. While these outputs can be downloaded locally, the cloud offers the ability to scale compute resources to perform analyses over large areas and time spans, which is critical as data volumes continue to grow.
- This workflow combines search, discovery, access, reformatting, basic analyses, and plotting components presented during Part-II. Though the example we're working with in this notebook only focuses on a small time and area to account for a large number of concurrent processing requests, this workflow can be modified and scaled up to suit a larger time range and region of interest.

Learning objectives:

- Understand the Pangeo BinderHub environment used during the workshop and how to execute code within a Jupyter Notebook
- Search for GRACE/GRACE-FO land water equivalent (LWE) and SMAP sea surface salinity (SSS)
- Execute programmatic data access queries, plotting, and direct in-region cloud access using open source Python libraries.
- Access data in Zarr format from Earthdata Cloud (AWS)
- Subset both, plot and compare coincident data.
- Identify resources, including the Earthdata Cloud Primer, for getting started with Amazon Web Services outside of the Workshop to access and work with data with a cloud environment.



https://github.com/podaac/AGU-2020

DATASETS USED

MODIS AQUA L2

GHRSST Level 2P Global Sea Surface Skin Temperature from the Moderate Resolution Imaging Spectroradiometer (MODIS) on the NASA Aqua satellite (GDS2)

Shortname
MODIS_A-JPL-L2P-v2019.0

Variable of interest: skin SST

Temporal Coverage: 2 July 2002 – present (ongoing)

Temporal Resolution:

Spatial Resolution: 1 km at nadir for L2P SST (36 spectral bands at a variety of spatial resolutions)

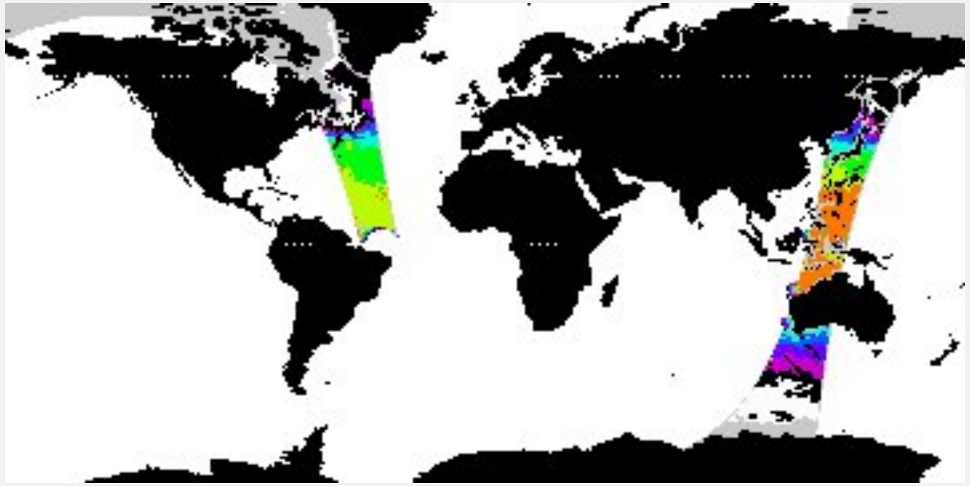
Native Format: NetCDF

Provider/Data Center (NASA Earthdata Cloud): POCLOUD

Collection ID: C1940473819-POCLOUD

https://podaac.jpl.nasa.gov/SMAP?sections=data

https://search.earthdata.nasa.gov/search?q=POCLOUD%20MODIS%20aqua&m=20.706409351350445!-114.2578125!3!1!0!0%2C2



DATASETS USED

ICESAT-2 ATL03

ATLAS/ICESat-2 L2A Global Geolocated Photon Data V003

Shortname: ATL03

Variable of interest: time-tagged photon height above ellipsoid

Temporal Coverage: 14 October 2018 – ongoing

Temporal Resolution: swath, 91-day repeat orbit

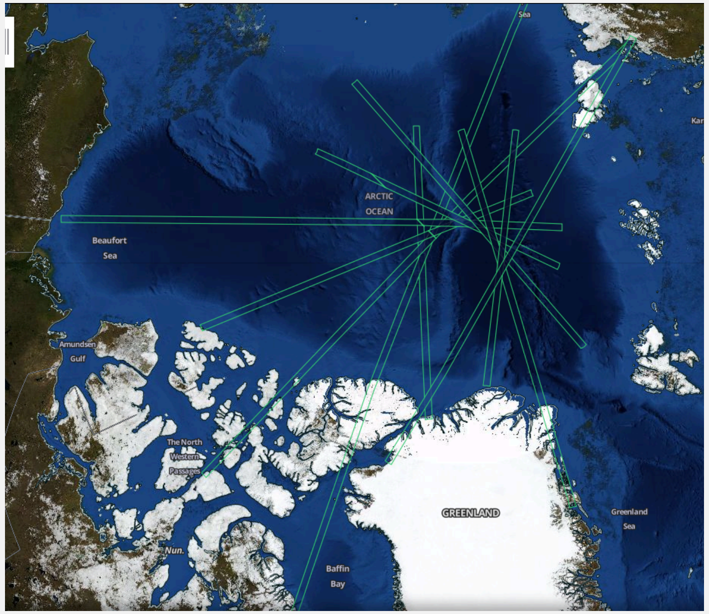
Spatial Resolution: one later pulse every 70 cm, actual resolution depends on surface reflectivity and cloud cover

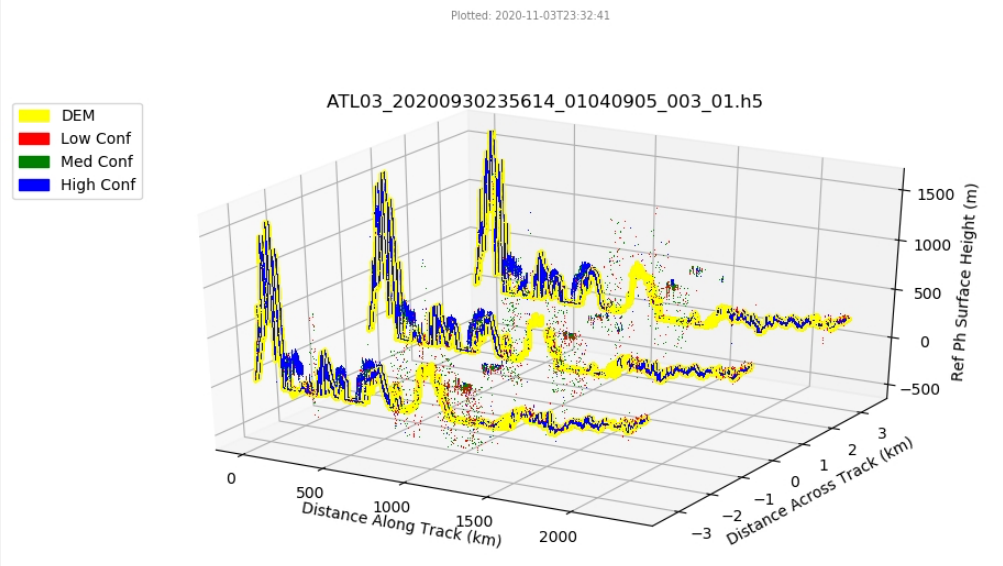
Native Format: HDF5

Provider/Data Center (NASA Earthdata Cloud): NSIDC DAAC

Collection ID: https://doi.org/10.5067/ATLAS/ATL03.003

https://nsidc.org/data/ATL03





ICESAT-2 ATL07

ATLAS/ICESat-2 L3A Sea Ice Height V003

Shortname: ATL07

Variable of interest: sea ice height

Temporal Coverage: 14 October 2018 – ongoing

Temporal Resolution: swath, 91-day repeat orbit

Spatial Resolution: variable depending on signal photon returns, up to 150 m

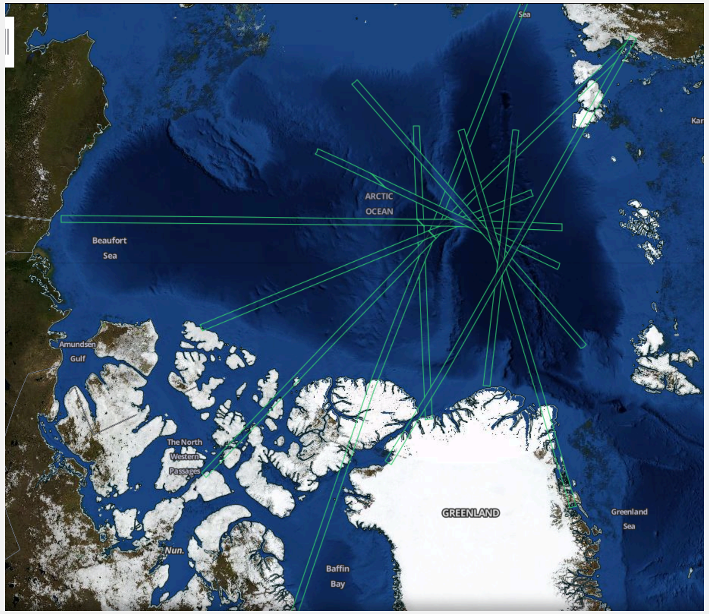
Native Format: HDF5

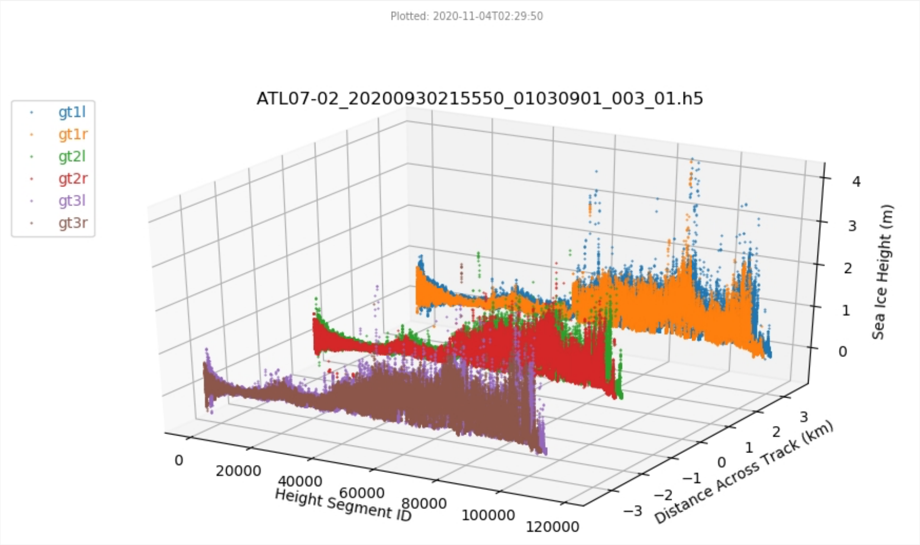
Provider/Data Center (NASA Earthdata Cloud): NSIDC DAAC

Collection ID: https://doi.org/10.5067/ATLAS/ATL07.003

https://nsidc.org/data/ATL07

DATASETS USED





DATASETS USED

SMAP L3

RSS SMAP Level 3 Sea Surface Salinity Standard Mapped Image 8-Day Running Mean V4.0 Validated Dataset

Shortname
SMAP_RSS_L3_SSS_SMI_8DAY-RUNNINGMEAN_V4

Variable of interest: Salinity

Temporal Coverage: 1 April 2015 – present (ongoing)

Temporal Resolution: Daily data files for this product are based on SSS averages spanning an 8-day moving time window.

Spatial Resolution: L3 products are global in extent and gridded at 0.25 degree x 0.25 degree

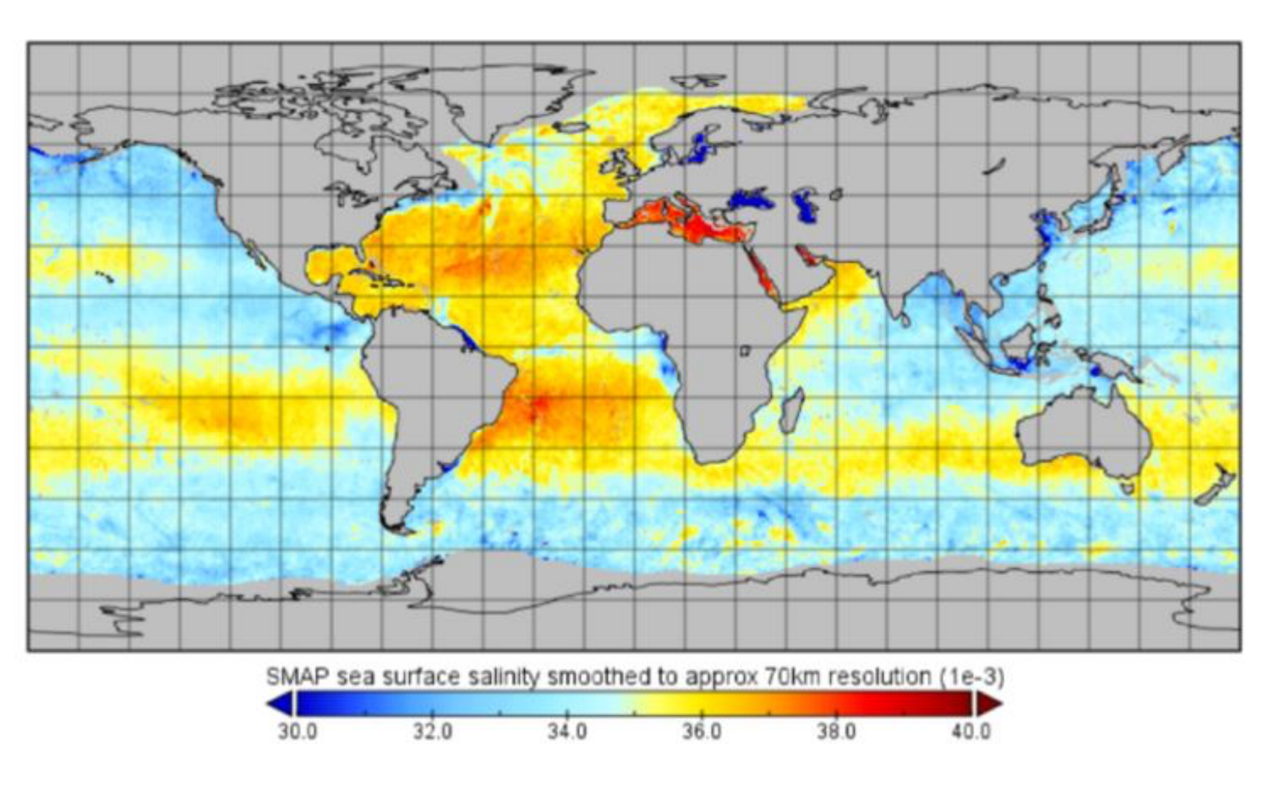
Native Format: NetCDF

Provider/Data Center (NASA Earthdata Cloud): POCLOUD

Collection ID: C1940468263-POCLOUD

https://podaac.jpl.nasa.gov/SMAP?sections=data

https://search.earthdata.nasa.gov/search?q=POCLOUD%20SMAP&m=20.706409351350445!-114.2578125!3!1!0!0%2C2



WORKSHOP SURVEY

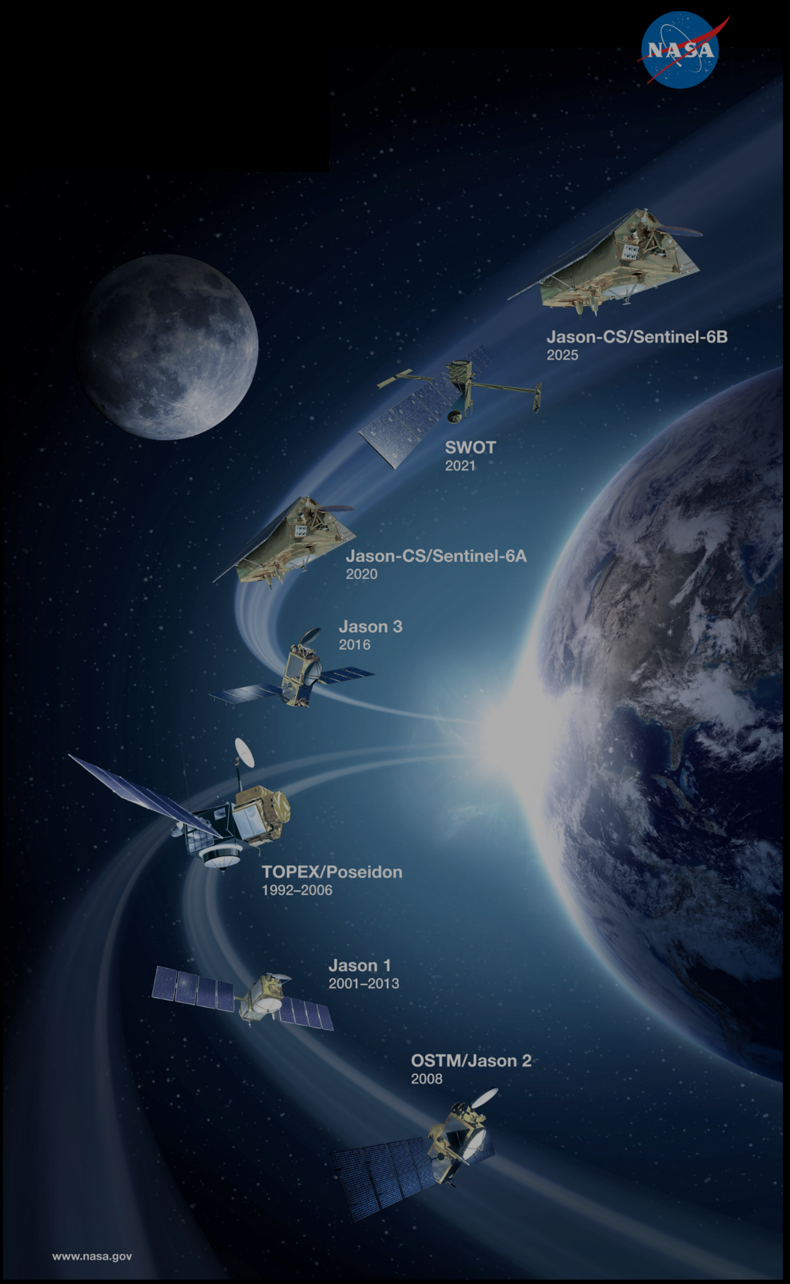
- Thank you for participating in this workshop!
- Please take a few minutes to let us know how the workshop went for you, what we could improve on, and what you would like to see more of (in terms of training and resources)
- https://docs.google.com/forms/d/e/1FAIpQLSeHv32kz6tNDNY-LLr_0Z4um1ga_8bzV1XnLzC9CJ_mOYBbEA/viewform?usp=sf_link

RESOURCES

Interesting demos and exercise during workshop. Now what?

RESOURCES

- AGU 2020 Workshop repo https://github.com/podaac/AGU-2020
- Earthdata Cloud Primer – Getting started with the AWS Cloud https://earthdata.nasa.gov/learn/user-resources/webinars-and-tutorials/cloud-primer
- Earthdata Harmony – https://harmony.earthdata.nasa.gov/
- STAC https://stacspec.org/



Thank you!

PO.DAAC

https://podaac.jp.nasa.gov

NSIDC

https://nsidc.org/

4 December 2020
AGU Fall Meeting Workshop

EXTRA SLIDES

OUTLINE

Welcome (5 min)
- Instructors & Presenters
- Workshop motivation & scope
- Agenda
- Logistics

Part I - Overview & Background (30 min)
- DAAC Introduction
- NASA Earthdata evolution to the cloud - background and motivation
- Cloud paradigm
- Datasets introduction - POCLOUD data - current restrictions
- CMR Introduction
- Harmony Introduction
- Ways to discover and explore Earthdata datasets (UI, API)
- Learning objectives (leading into demos)

Part II - Science Use Case Demos (90 min, with 10 min break)
- Tutorial 1: Search by shp
- Tutorial 2: Search by pt-based

Break (10 min)
- Tutorial 3: Open altimetry
- Tutorial 4: Subset ice and sst
- Tutorial 5: Access in the cloud, reformat to zarr

Break (30 min)

Part III - hands-on workflow: analysis in the cloud (120 min)
- Outline of workflow & logistics (slack, breakout rooms, team work)
- Go to github
  - Review readme (brief)
- Click on Binder button
- Jupyter Lab in AWS is spun up
  - Set up .netrc (EDL)
  - Generate harmony access key
- Begin running through notebook

Survey (10 min)

Break (5 min)
- Finish hands-on workflow

Wrap-up, Resources, Q&A

AGENDA

Topic	Description	Time (UTC)	Time (PST)
Part I: Overview of NASA Earthdata Cloud Evolution	Presentation with Q&A	16:00 - 16:35	8:00 – 8:35
Part II: Science Use Case Demonstrations	Jupyter Notebook demonstrations, highlighting NASA EOSDIS tools and services applied across several science use cases (15 min each)	16:35 - 18:45 (including breaks)	8:35 – 10:45 (including breaks)
Part III: Hands-on data discovery, access, and analysis in the cloud	Jupyter Notebook tutorial providing step-by-step guidance on cloud-based data access and cloud compute based on previous demonstrations	18:45 - 21:00 (including 5-min break and wrap up)	10:45 – 13:00 (including 5-min break and wrap up)

From Zero to Cloud: PO.DAAC & ESDIS Cloud Capability Timeline

Existing Levels of Service, New User Communities

2019	2020	2021	2022

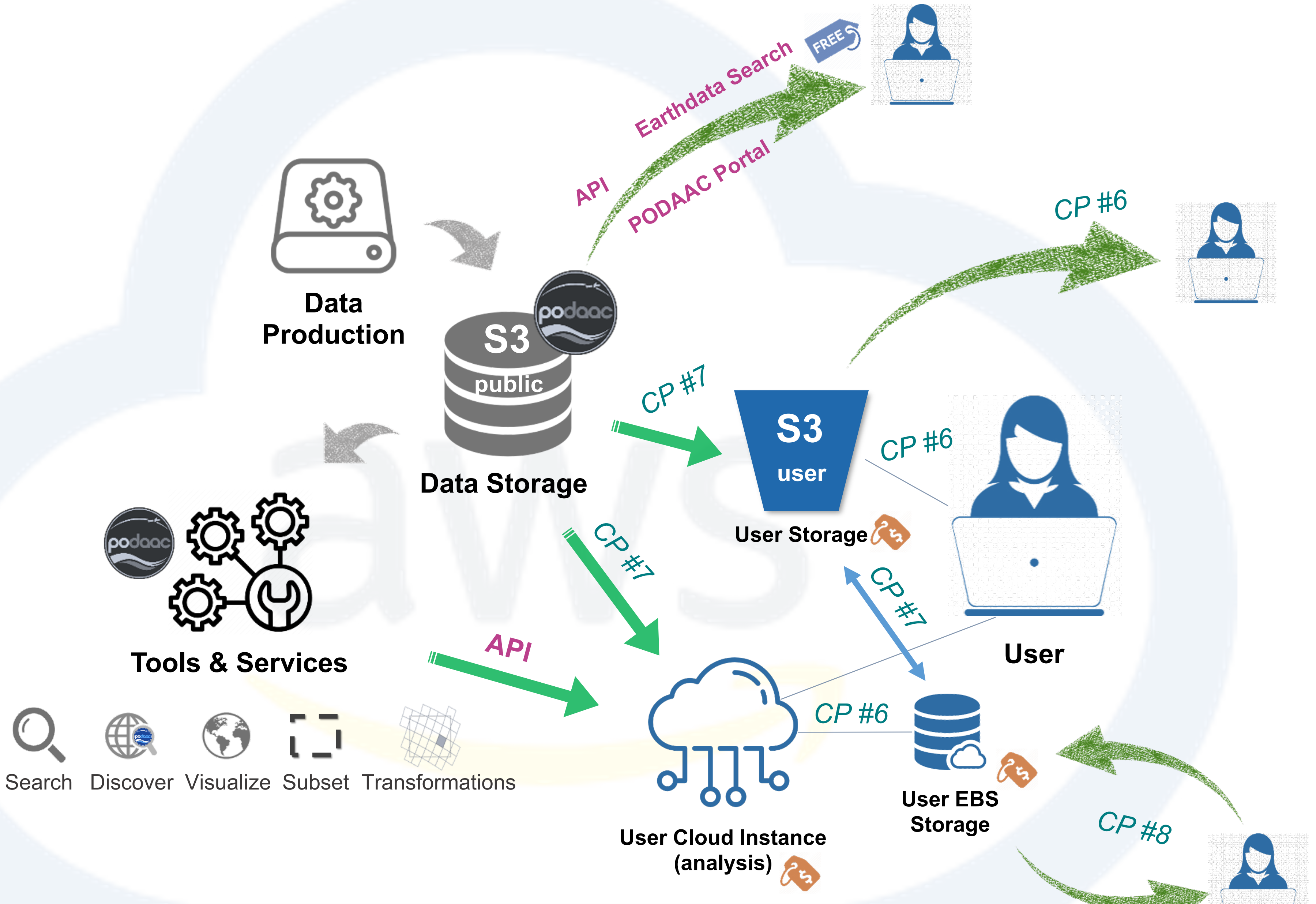
"Next Gen" Access Patterns

Search and Download
- Geometry Type
- Hydrological Unit Code
- Spatial and Temporal Search
- Subscriptions

Transformations
- Subset (space, time, variable, polygon, Quality)
- Reproject
- Regrid
- Reformat
- Data Product Specific: e.g. spatial/time averaging

Analysis
- Analysis in place (Cloud)
- Integration with other PO.DAAC, Earthdata datasets
- ZARR, cloud optimized geotiff (COG)

EARTHDATA CLOUD PRIMER – TUTORIALS GUIDE



https://earthdata.nasa.gov/learn/user-resources/webinars-and-tutorials/cloud-primer