Project Details

Name of project: Climate Observatory - Analyzing and Visualizing NOAA Satellite Climate Data Records on the Cloud

Project lead and contact details: Yuhan (Douglas) Rao, Cooperative Institute for Satellite Earth System Studies, affiliated with NOAA NCEI, yrao5@ncsu.edu

Project partners and contact details: Jessica Matthews, NOAA NCEI, jessica.matthews@noaa.gov

Proposed start and end date: November 2021 - May 2022

Project Outline

Project description:

NOAA's Climate Data Records (CDRs) are robust, sustainable, and scientifically sound climate records that provide trustworthy information, mostly derived from long-term satellite data archives, on how, where, and to what extent the land, oceans, atmosphere, and cryosphere are changing (https://www.ncei.noaa.gov/products/climate-data-records). NOAA CDRs were developed in response to the recommendations made by the National Research Council to embrace NOAA's mandate in understanding climate variability and change through national leadership in generating and managing satellite-based CDRs and ensuring long-term consistency and continuity for satellite CDR program.

Although CDR datasets are freely available from NOAA NCEI for nearly a decade, these datasets have been made publicly accessible via cloud platforms (e.g., Google Cloud and Amazon Web Services) through NOAA's Big Data Program (https://www.noaa.gov/information-technology/big-data) in 2021. Historically, users of these datasets were required to have advanced technical skills and large computing capabilities to access and analyze NOAA CDRs. However, the migration to the cloud platform alleviates some of these restrictions and provides the possibility to reach a larger user base who need reliable climate information for various purposes. Easy access to more than 40 NOAA CDR datasets has the great potential to enable different stakeholders to develop climate services and products to address the pressing climate change challenges.

To support the ease of use of these datasets we propose a project to build a variety of open source tools to simplify visualization and analysis of NOAA climate data records using cloud computing services. These tools include: 1) a climate data record viewing dashboard where on-the-fly user-defined climatology and anomaly maps can be created, and subsetted point and regional time series can be displayed, and 2) a library of interactive Jupyter notebooks of example scientific data analytics using NOAA CDRs to assist users for their own applications. As part of the development, we will also explore the impact of data format choice, comparing the efficiency of Network Common Data Form (netCDF) vs. cloud-optimized formats such as Zarr or Cloud-Optimized GeoTIFF (COG). The outcomes of this project will serve to inform efficient and effective user access approaches as NOAA transitions its data archive holdings to cloud platforms.

To make the scope of this pathfinder project practical within the proposed time frame and cloud computing resources, we plan to choose 2-3 representative NOAA CDR datasets as the prototype for the proposed project. The selection will be made in consultation with the NOAA CDR program and BDP based on the user need and statistics to prioritize the development of critical CDR datasets. In addition to

the proposed development, we will also incorporate educational activities to train future scientists for using NOAA data with cloud platforms via NOAA Hollings Scholarship and NC State University student internship opportunities.

Upon the completion of this pathfinder project, the developed tools and materials will be transitioned into the CISESS cloud computing account for further development with the expanded CDR datasets. We will also seek appropriate opportunities to transition the development to the future operational cloud environment with NOAA National Centers for Environmental Information to enhance the climate information and services provided to the public.

Project technical & learning objectives:
The project will be separated into three major phases listed below:

Phase 1 - Dataset selection. We will consult the NOAA CDR program and BDP to determine the prototype dataset for the development based on the historical data usage and user requests. We will select the top 2-3 representative CDR datasets to inform the tool and tutorial development.

Learning objective: Understanding user requirements and needs for NOAA CDR data based on historical data usage and how to design cloud-based services to expand potential users.

Phase 2 - Data format conversion and performance analysis. We will use this phase of the project to explore the performance of different data formats for cloud-based data analytics and visualization. The candidate data format includes the native CDR format (netCDF-4), Zarr, and cloud-optimized GeoTiff (COG). This exploration will build upon existing community efforts such as the ESIP Cloud Computing Cluster, Pangeo, and Unidata on how to produce cloud-optimized geospatial data. The resulted cloud-optimized CDR datasets will be available to the public. The “recipe”, or the computational workflow, for cloud-optimized data format conversion will be documented in interactive Jupyter notebooks that will be made publicly available as tutorials for potential users and developers.

Learning objective: Determining the suitability of various data formats for analyzing and visualizing satellite-based CDRs data on cloud computing platforms.

Phase 3 - Dashboard development and user feedback collection. Once the suitable data format is determined for the selected CDR datasets, we will develop the Climate Observatory dashboard to allow users to visualize climatology and anomaly maps as well as time series of user-defined geographical regions. The dashboard will be hosted via the CISESS website to provide public access to this cloud-based NOAA CDR data explorer. We will also provide associated Jupyter notebooks to allow users to develop their own analytics and applications based on the base development. Depending on the timeline of the development, we will seek user feedback through appropriate community events (e.g., ESIP meetings and scientific conferences).

Learning objective:

1. Learning how to provide a user-friendly data exploration dashboard to expand the user base.

2. Understand the cost of providing public CDR data explorer services via cloud platforms; this will help inform the plan and resource need of future transition into operational cloud environment.

Project significance & impact:

This project will ease the public access to NOAA CDR data and enable users to access NOAA CDR data on the cloud platforms and explore the data for different applications. This will further enhance NOAA's ability to fully achieve the recommendation from the National Research Council and amplify the value of future NOAA CDR development to address climate challenges.

The resulted cloud-optimized CDR datasets will also enable efficient climate research and climate-service related application development for the public.

Description of key project steps and timeline:

November 2021 - December 2021: Engagement with NOAA CDR and BDP to determine the prototype CDR datasets;

December 2021 - February 2022: Conduct performance comparison for different data formats using a subset of CDR datasets and implement the computational workflow to convert selected CDR datasets to the optimized data format;

February 2022 - May 2022: Develop the cloud-based dashboard based on the optimized data format for the selected CDR datasets in consultation with potential data users.

Outreach

1. Phase 1: we will engage with the NOAA CDR program at NCEI and BDP to determine user needs for the NOAA CDR data;
2. Phase 2: we plan to engage with ESIP Cloud Computing Cluster and the Pangeo community to evaluate the performance of different data formats for our cloud-based dashboard development;
3. Phase 3: we plan to engage with potential user groups, including the CMIP6 obs4MIP community, USGS Climate Adaptation Science Centers, NOAA Regional Climate Centers, and Group on Earth Observations (GEO) during the development of the CDR data explorer dashboard.
4. The project will also provide training and learning opportunities via NC State University internship and NOAA scholarship opportunities for students who want to learn more about developing cloud-based tools and services with NOAA climate data.