Web-based Visualization and Analytics of Petascale NASA Climate Data: Equity as a Tide that Lifts All Boats

Aashish Panta*, Xuan Huang*, Nina McCurdy†, David Ellsworth†, Amy Gooch*, Giorgio Scorzelli*, Gustavo Ovando-Montejo**, Hector Torres‡, Patrice Klein‡¶, Valerio Pascucci*
*SCI Institute, University of Utah, †NASA Ames, ‡NASA Jet Propulsion Lab, ¶Caltech, **Utah State University, Blanding

Introduction

Traditional tools are often unable to handle large scale datasets creating bottleneck in data analysis and visualization. We present a data model and a set of visualization dashboards designed to address these issues by enabling progressive streaming and visualization of scientific data.

1. Scalable Visualization Framework: An innovative, scalable dashboard framework that enable progressive visualization with advanced analytical tools.

2. Efficient Data Reduction and Optimization: A framework that enables efficient storage and transfer of large- scale data with analysis-ready cloud optimized format.

3. Data Democratization: Successfully converted and migrated more than a Petabyte of raw data from Pleiades, a NASA Supercomputer to Cloud, enhancing public access and collaborative opportunities.

Interactive Visualization and Dashboards

- We worked on two large-scale climate simulation datasets: DYAMOND dataset and LLC4320 Ocean Dataset including several fields such as: **ocean velocity, salinity, sea surface temperature**.
- We developed an innovative data conversion pipeline to convert more than **a Petabyte** of this data to OpenVisus IDX format, known for its fast and progressive streaming capabilities. We also tested various compression algorithms, both lossy and lossless, on these datasets
- In an effort to make the data publicly available, we uploaded all the data in **compressed streamable format** to the SealStorage cloud, a decentralized cloud storage service. We deployed several **dashboards and jupyter notebooks** fetching the data for fast analysis and rendering.

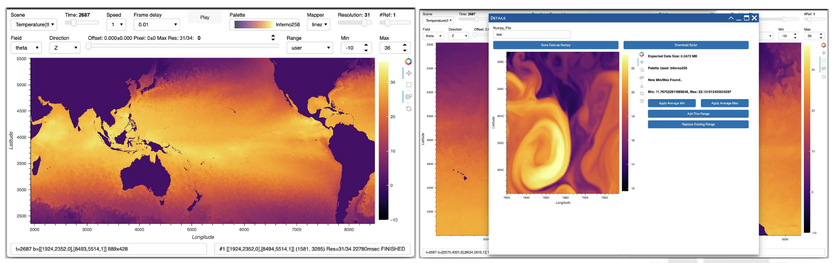

Fig. 1: Interactive visualization of sea surface temperature, along with inset for selecting the regions of interest (left). The dashboard provides the ability to directly download the data locally or to download a Python script that fetches the region from the cloud (right).

Workflow

1. Our dashboards are dynamically adjusted to fetch the data from the appropriate source as needed.
2. For NASA scientists who have access to Pleiades, the largest NASA supercomputer, the dashboard will access the data from their own file system.
3. Other users will be retrieve the data from the cloud and this data will be cached as needed to reduce networks data transmission overheads.

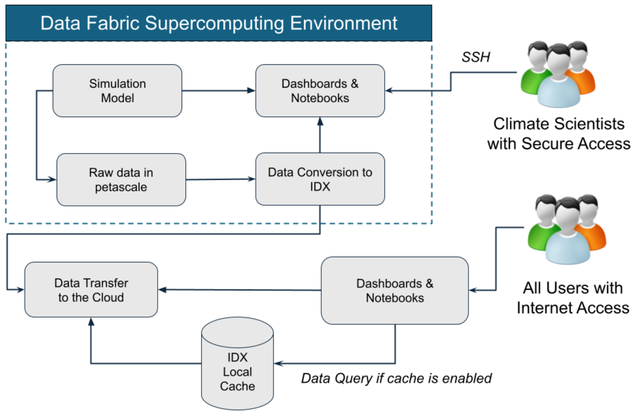

Fig 2: An example workflow showing data conversion and retrieval pipeline for Jupyter notebook and dashboards from a supercomputing environment. Simulation data, typically in raw format and limited due to credentials, is restricted due to large storage that impacts real-time capabilities. Our framework makes the data readily available for on-the-fly analysis and provide users with access to the data as needed through interactive dashboards and Jupyter notebooks.

Examples

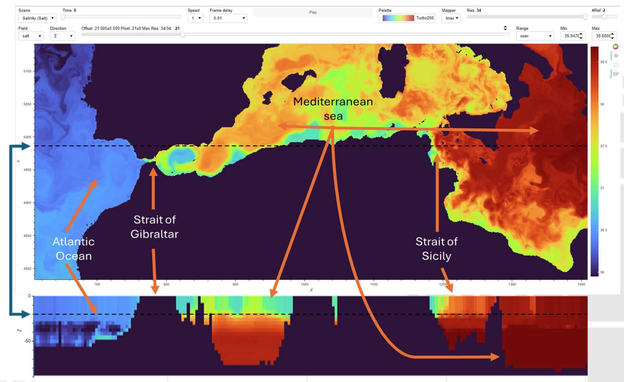

Fig 3: **Gibraltar Use Case.** Zoomed-in view of the general water circulation through the Strait of Gibraltar connecting the Mediterranean with the Atlantic. Low-salinity water enters the Mediterranean from the Atlantic through the Strait of Gibraltar; then the salinity increases, and the water starts to sink as the current moves east. This on-the-fly selection of interesting regions from a massive dataset and playing them through time facilitates a deeper understanding of complex climatic phenomena, which was not practically accessible before the implementation of our framework.

Fig4: Interactive features available form the dashboards include dataset selection, colormap, slices, timestep, resolution, data range, regions of interest extraction, and so on

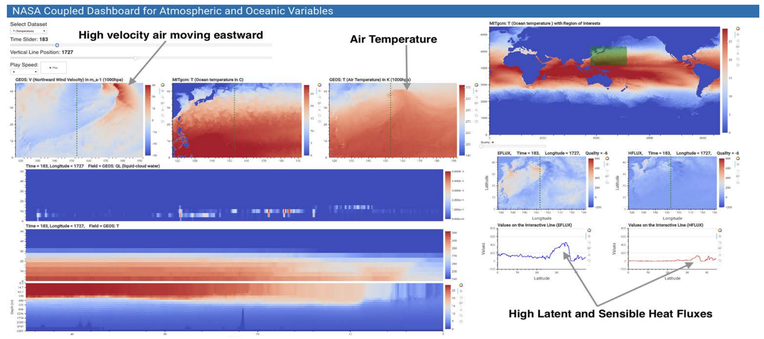

Fig. 5: **Kuroshio Use Case.** Scientific analysis shows how the increasing heat fluxes (two plots and images at the bottom right) and air temperature (top middle) create a high-velocity wind (top left) in the atmosphere moving eastward for the Kuroshio region

Data Democratization for Teaching at Utah State Blanding

- We successfully introduced the same dashboards and simplified training material in an undergraduate class on Geospatial Analysis in a minority-serving campus (Utah State Banding) with 69% of the Native American (mainly Navajo Nation) students, 86% being low-income.
- The students were initially struck by the realization and excitement that they could access these variables in full resolution, akin to methods used by NASA scientists.
- Overall, the use of dashboards proved to be an invaluable pedagogical tool for facilitating hands-on technical training within the realm of GIScience.

QR Codes to the dashboards

CEDMAV

CEDMAV

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu