CDATWeb: A Client/Server Model for the Ultrascale Visualization Climate Data Analysis Tools (UV-CDAT)

Jonathan Beezley[1], Aashish Chaudhary[1], Charles Doutriaux[2], Matthew Harris[2], Dean N. Williams[2]

[1]Kitware
[2]Lawrence Livermore National Laboratory

Backgound

UV-CDAT is an integrated framework that provides an end-to-end solution for management, analysis, and visualization of ultrascale data sets generated for current and future DOE climate data repositories and for the climate science community at large.

UV-CDAT is based on a client-server architecture and is integrated within the Earth System Grid Federation (ESGF) framework, allowing UV-CDAT to take advantage of the advanced data management mechanisms of ESGF. In this way, UV-CDAT provides regridding, reprojection, and aggregation tools directly as a component of the ESGF data node, eliminating or substantially decreasing data movement. The UV-CDAT server provides a turnkey application for building complex data analysis and visualization workflows by interacting with one or more UV-CDAT backend servers. These workflows may use predefined components for data transformation and analysis, data collection from disparate data sources outside ESGF, visualization, and user-defined processing steps. Extended provenance information can be captured on the backend server for sharing and collaboration.

Access

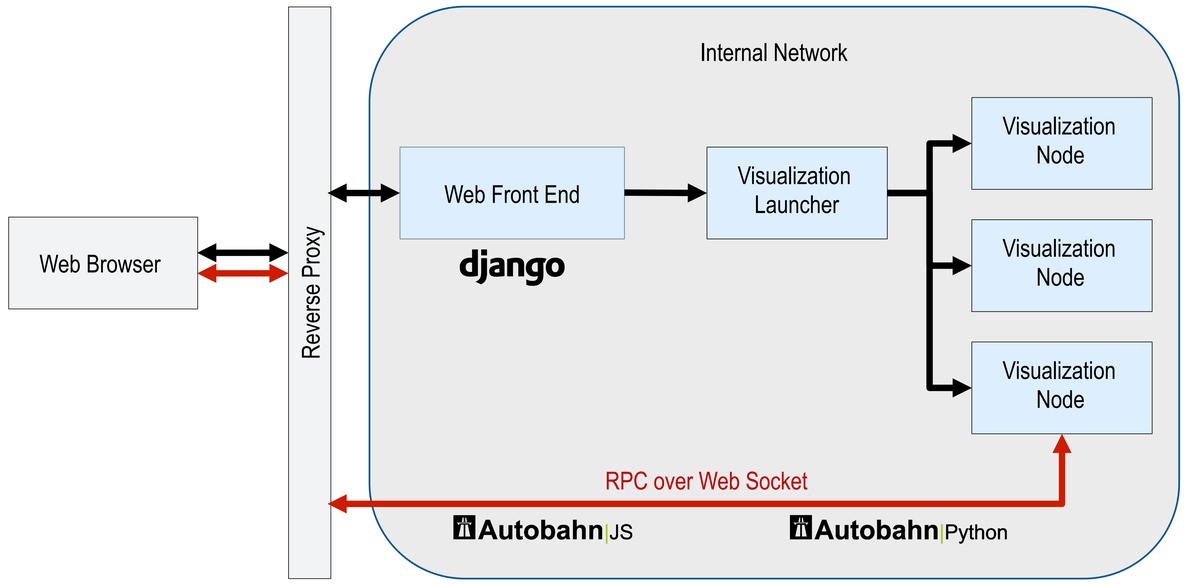
We are providing three different ways for users to access UV-CDAT web informatics (i.e., CDATWeb):
(1) import data from ESGF;
(2) import an existing UV-CDAT project that has been previously saved; and
(3) import data sets from local machines (prior to ESGF publication).

Import Data from ESGF. A user can simply go to ESGF to find their desired data sets and import these data sets into the CDATWeb informatics interface via the UV-CDAT hyperlink that is associated with either a single data set or the entire data cart.

Existing Project. Our web interface provides the option to open an existing project that has been previously saved. Also, we intend to provide the option of recovering an existing project in the case where a client browser crashes in the middle of an active session.

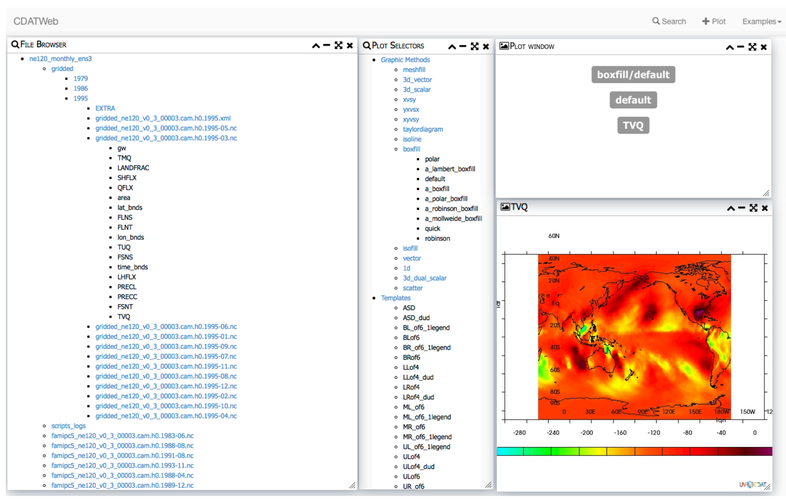
Local File. A user can simply open a directory containing all the data sets and upload to the UV-CDAT server for remote data analysis.

Architecture

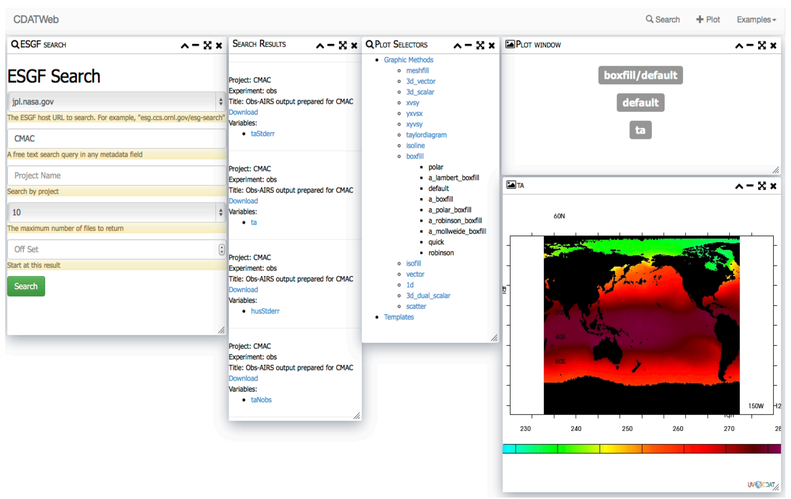


Prototype

Data from CADES at ORNL
Visualization Server in CADES at ORNL (File server mounted on visualization server)
CDATWeb Client running in development environment at LLNL



Data from JPL Data Node
Visualization Server in CADES at ORNL (File server not mounted on visualization server)
CDATWeb Client running in development environment at LLNL



Components

Next Steps

High-resolution models, ensemble analysis, derived data product generation, and intercomparison of model results and observations require substantial data resources regarding storage infrastructure (e.g., high-performance parallel input/output (I/O) environments) and server-side (or remote) data processing. Remotely manipulating these data sets—from generation to transformation and fusion to archiving—requires a robust infrastructure that supports multiple models, storage systems, operating systems, and compute environments.

Test bed simulations, such as the DOE's Accelerated Climate Modeling for Energy (ACME) project, require parallel I/O and parallel compute environments that provide a structured data model suitable for parallelizing and automating ensemble analysis and intercomparison of models and observational data sets remotely. Observational data sets come in a variety of formats and use numerous metadata models. Supporting the server-side requirements of these data sets—and data products derived from these data sets—requires decoupling the data format from data access and exploratory/parallel analysis.

To accommodate ACME better in the future, the following next steps must occur:

- Data centers provide better compute and visualization hardware for ESGF/CDATWeb server-side computation;
- CDATWeb visualization server runs on a local cluster or node at any of the DOE leadership compute facilities (i.e., ALCF, OLCF, NERSC);
- Mount point for data be near data center's storage facilities for fast access and data movement;
- CDATWeb Django front-end installed alongside and or integrated with the ESGF CoG interface;
- Produce PNG files on full rotating 3D plots;
- CDATWeb fully intergraded with ESGF's Solr search engine for better search capability; and
- Mimic the UV-CDAT desktop application's graphical user interface, so users become accustomed to similar widgets.

This work is supported by the ESGF consortium: DOE, NASA, NOAA, IS-ENES, NCI.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DEAC52-07NA27344. LLNL-POST-679928