CLIMVIEWER: A SCALABLE, INTERACTIVE PLATFORM FOR CLIMATE & WEATHER DATA EXPLORATION

RUI YANG

SOFTWARE AND DATA MODERNISATION TEAM

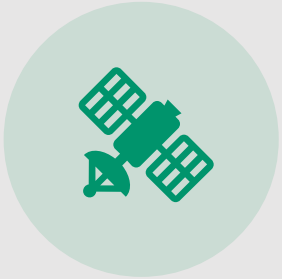
NATIONAL COMPUTATIONAL INFRASTRUCTURE (NCI)

eResearch Conference 2025

NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

NCI AUSTRALIA

DATA COLLECTIONS @ NCI



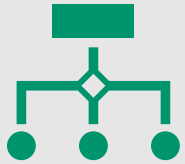
Australia's largest research data collection: climate, weather, Earth systems, environmental, satellite, geophysics, plus specialised domains (astronomy, genomics).

PBs of nationally and internationally significant datasets, co-located with HPC and analysis systems.

Prioritised with partners: ANU, Bureau of Meteorology, CSIRO, Geoscience Australia, ARC Centres of Excellence..

Data principles: FAIR, high-performance programmable access, open where possible, modern standards, transdisciplinary access.

EXISTING CLIMWX DATASET VISUALIZATION TOOLS



Local Applications

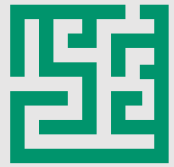
Ncview: Plot NetCDF files locally—basic and static

ECMWF Metview: Advanced desktop tool with scripting & interactive plots

Web Platforms

ec*Charts*: Interactive forecast visualization via dashboards

NOAA View: Predefined maps and interfaces with limited customization



Gaps

Local tools doesn't target at dataset; web tools limited in analytics/customizability

MOTIVATION & CHALLENGES

Growing high-res datasets

harder to explore & visualize

Existing tools

limited or inflexible

Need

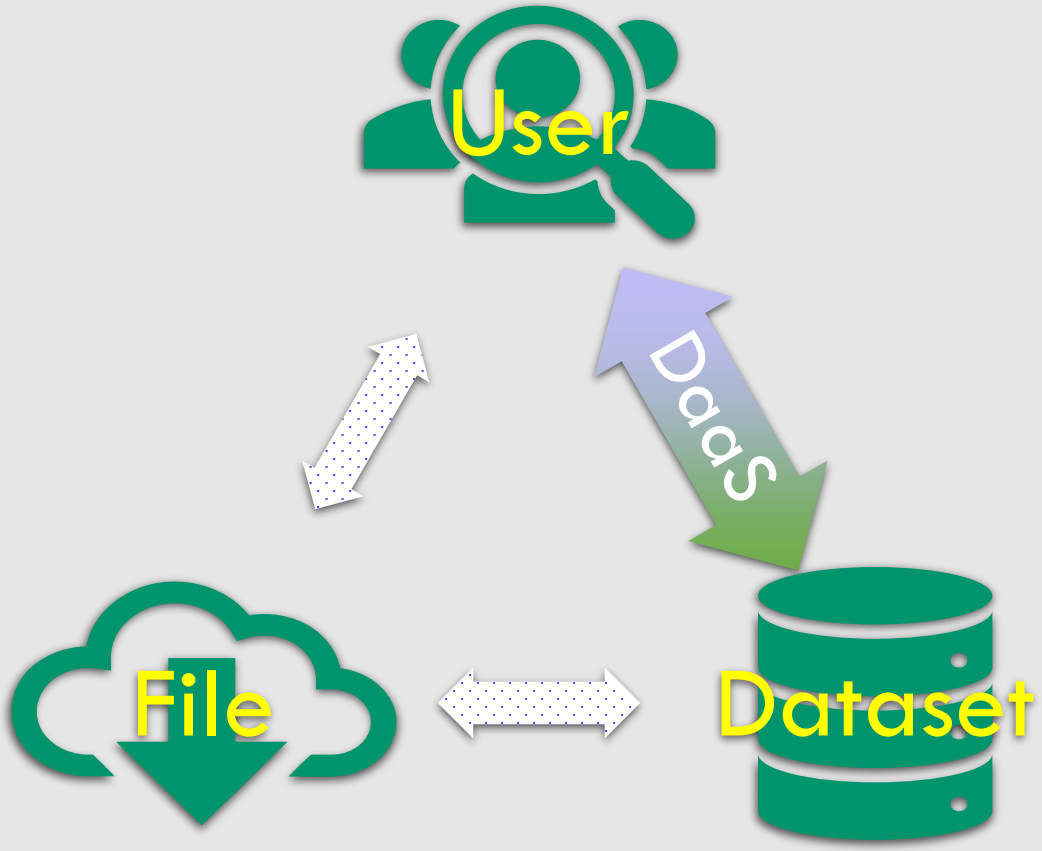
scalable, customizable, interactive viewer

Goal

visualize ClimWx dataset locally

DATASET AS A SERVICE

Dataset as a Service (DaaS) is a data delivery and management model where datasets are hosted, curated, and served on demand — usually through APIs, cloud platforms, or catalogs — so that users can easily discover, access, and use data without managing storage or infrastructure themselves.



NCI INTAKE DATA CATALOG

https://opus.nci.org.au/spaces/DAE/pages/275349609/Analysing+data+using+intake+indexes

Intake-esm is Python plugin to catalog Earth System Model datasets (CSV/JSON index), built on intake, pandas and Xarray

NCI uses these catalogs for climate/weather, geophysics satellite datasets

Provides fast metadata-based data discovery and dataset loading

INTAKE-ESM USAGE

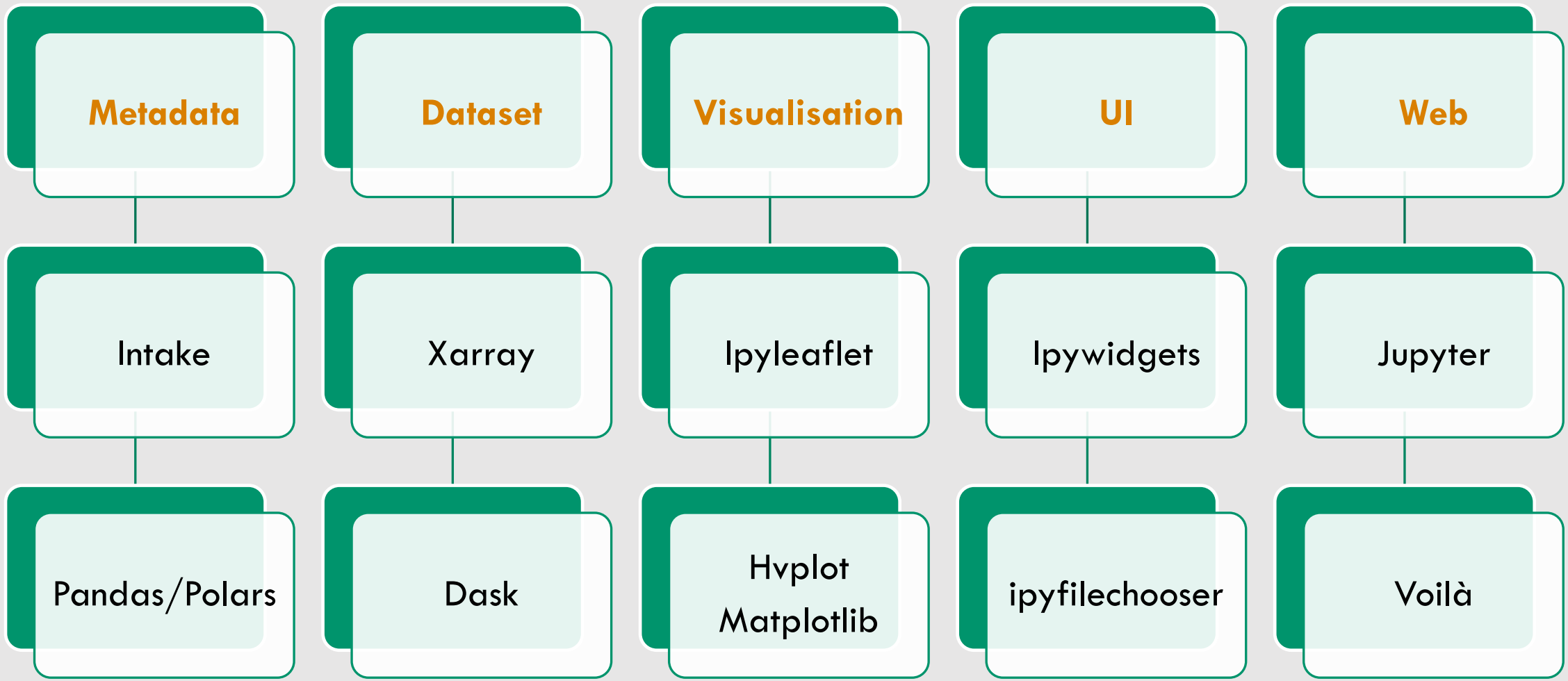
Without Intake Catalog	With Intake catalog

`base_dir = "/g/data/rr3/publications/CMIP5/output1/CSIRO-BOM/ACCESS1-3/historical/mon/atmos/Amon/r1i1p1/latest/tas"`
 	
`nc_files = sorted(glob.glob(os.path.join(base_dir, "*.nc")))`
 	
`ds = xr.open_mfdataset(nc_files, combine='by_coords')`
	`cat = intake.open_esm_datastore("cmip5_catalog.json")`
	
	`subset = cat.search(variable_id="tas", model="MPI-ESM-LR")`
	
	`ds = subset.to_dataset_dict()["CMIP5.MPI-ESM-LR.tas"]`

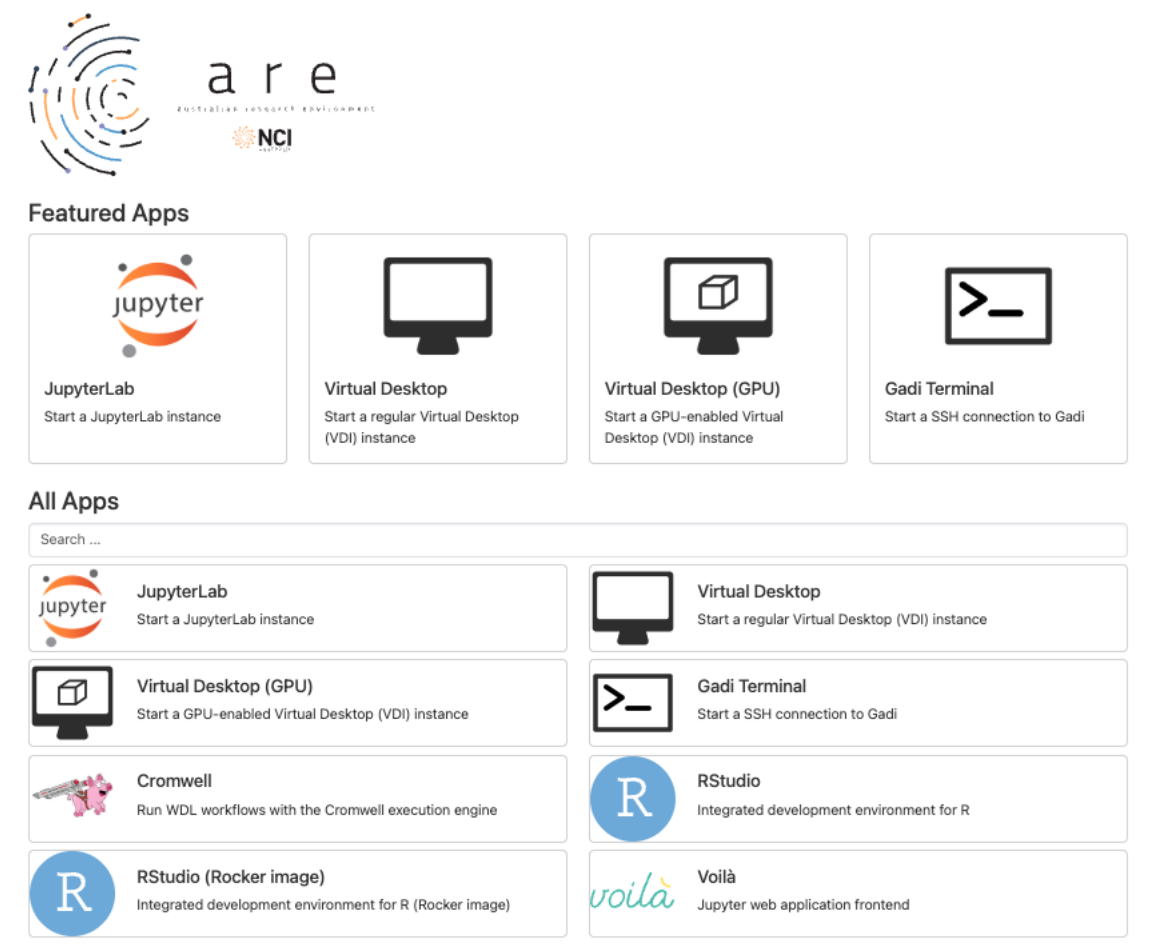
CLIMVIEWER FEATURES

- **Search and access** datasets using NCI (or user-generated) intake-ESM catalogs
- Support for **loading** NetCDF, Zarr, and related formats through Xarray
- **Animate** 2D,3D and 4D datasets along time or height sequential
- **Interactively zoom in&out for scope and details**
- Global and local **statistics for selected location** from map
- **Customized plot** such as color scheme, contours, labels etc.
- Dask-based **distributed processing** for performance and scalability
- Support massive **computing resources** across multiple cores or nodes

CLIMVIEWER ARCHITECTURE



NCI INTEGRATION VIA ARE

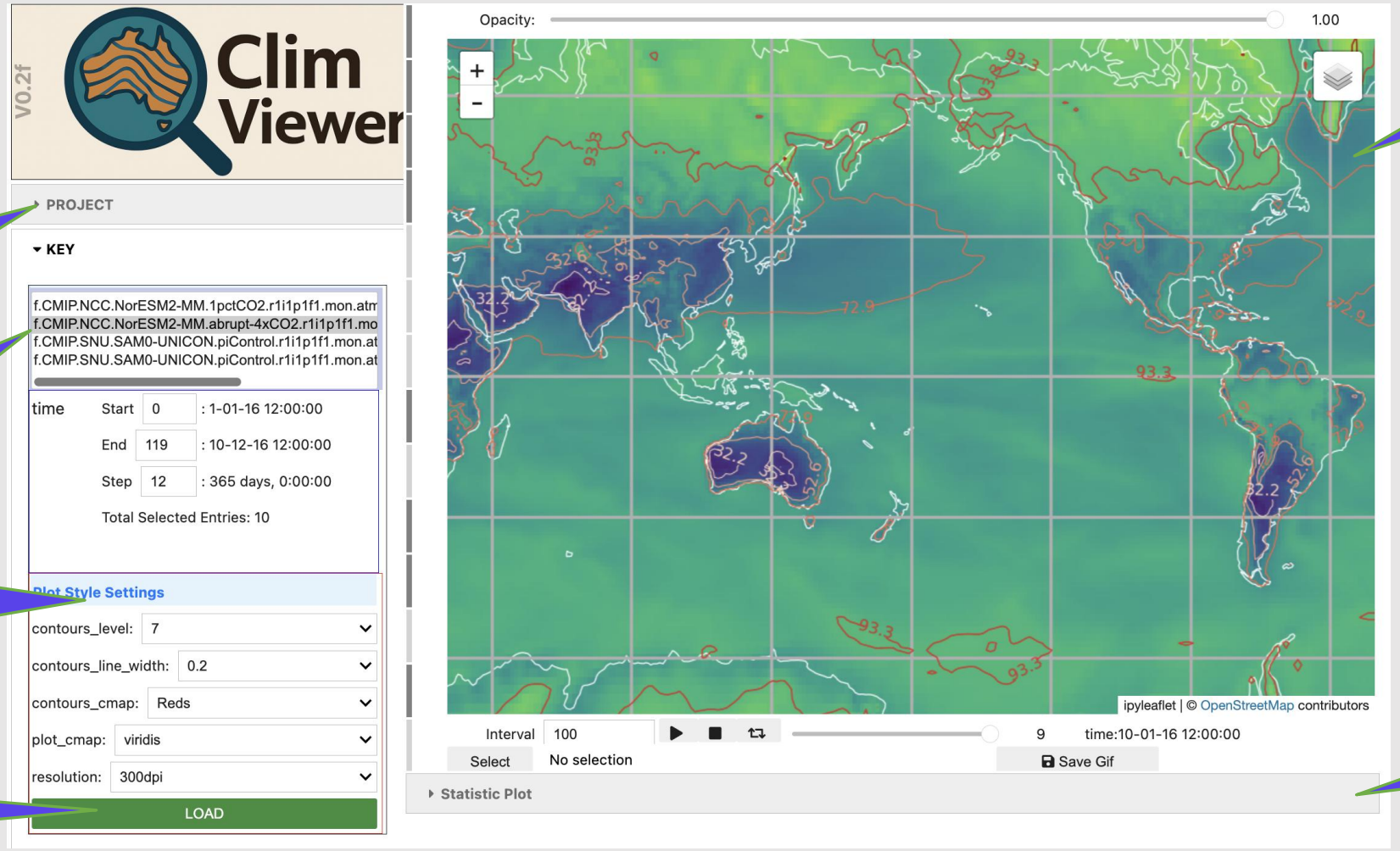


- Deployed and loaded via NCI ARE.
- Customized compute resources including single and multiple nodes
- Supports catalogs for multiple NCI local collections like CMIP5, CMIP6, ERA5, BARRA2, BARPA, CCAM, WeatherBench etc.
- Operating on web GUI interface via Voila

are.nci.org.au

https://opus.nci.org.au/spaces/DAE/pages/325387079/Voil%C3%A0+Apps+on+ARE

CLIMVIEWER INTERFACE



Step 1 Select data collection

Step 2 Select dataset

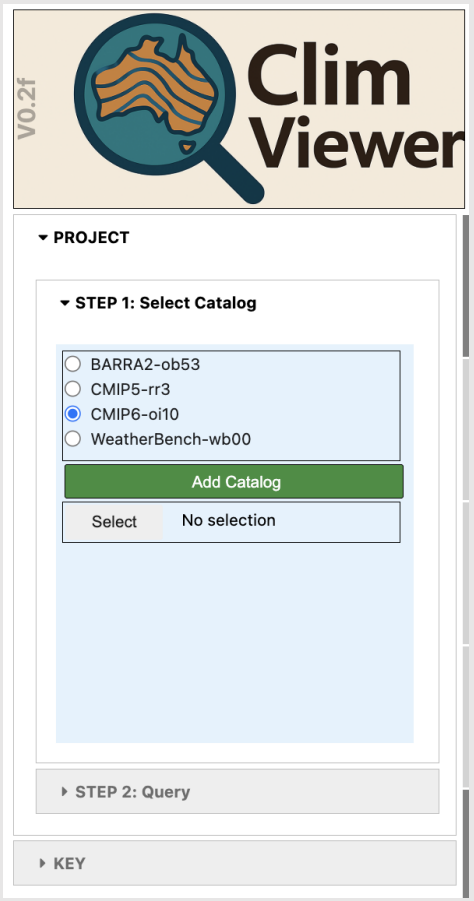
Step 3 Configure plots

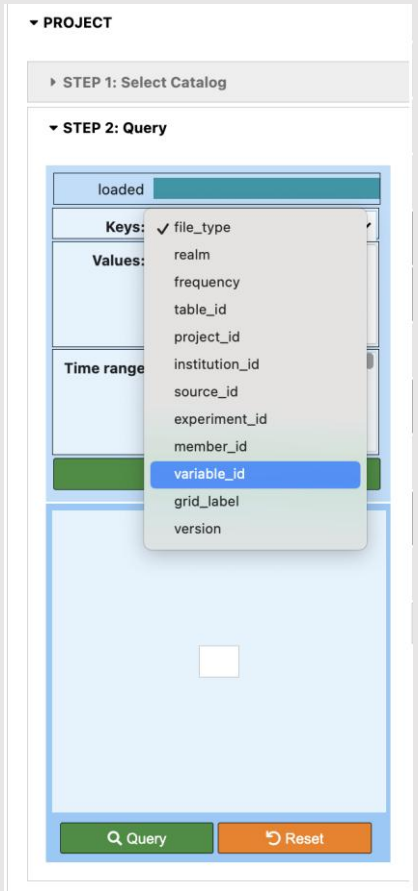
Step 4 Start

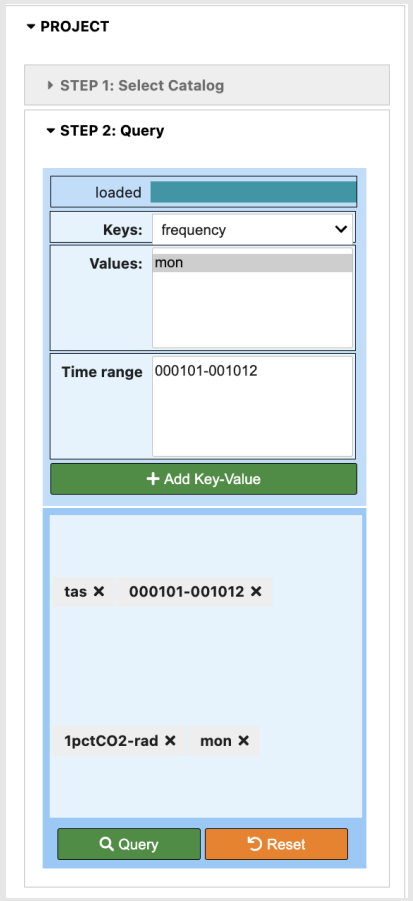
Step 5a Animate plots

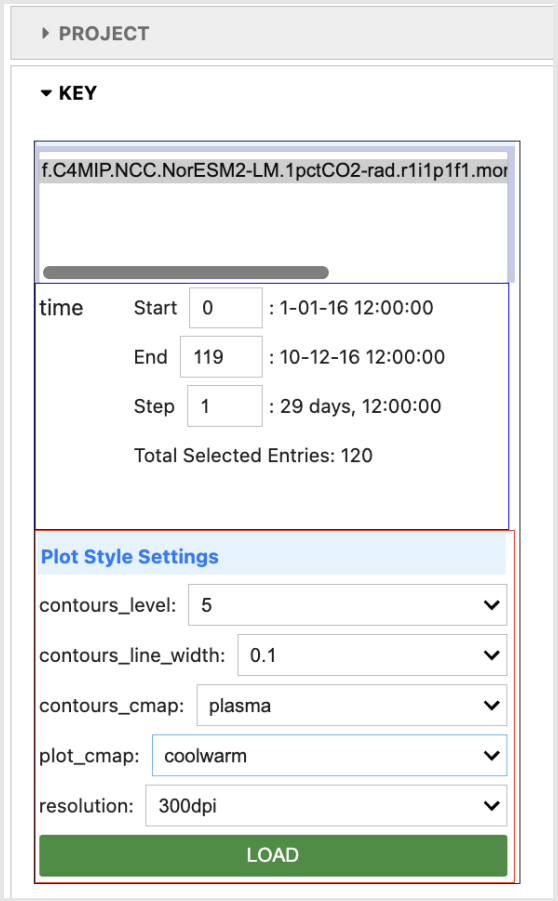
Step 5b Statistic analysis

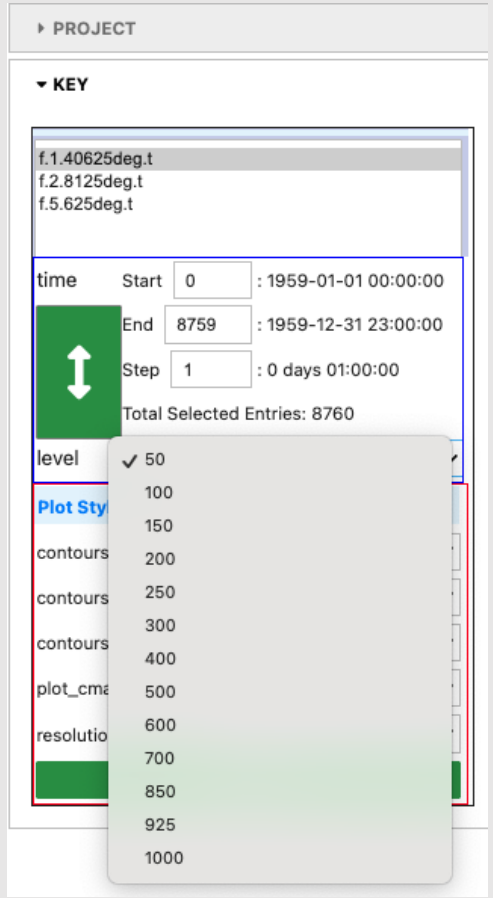
SEARCH METADATA AND LOADING DATASET





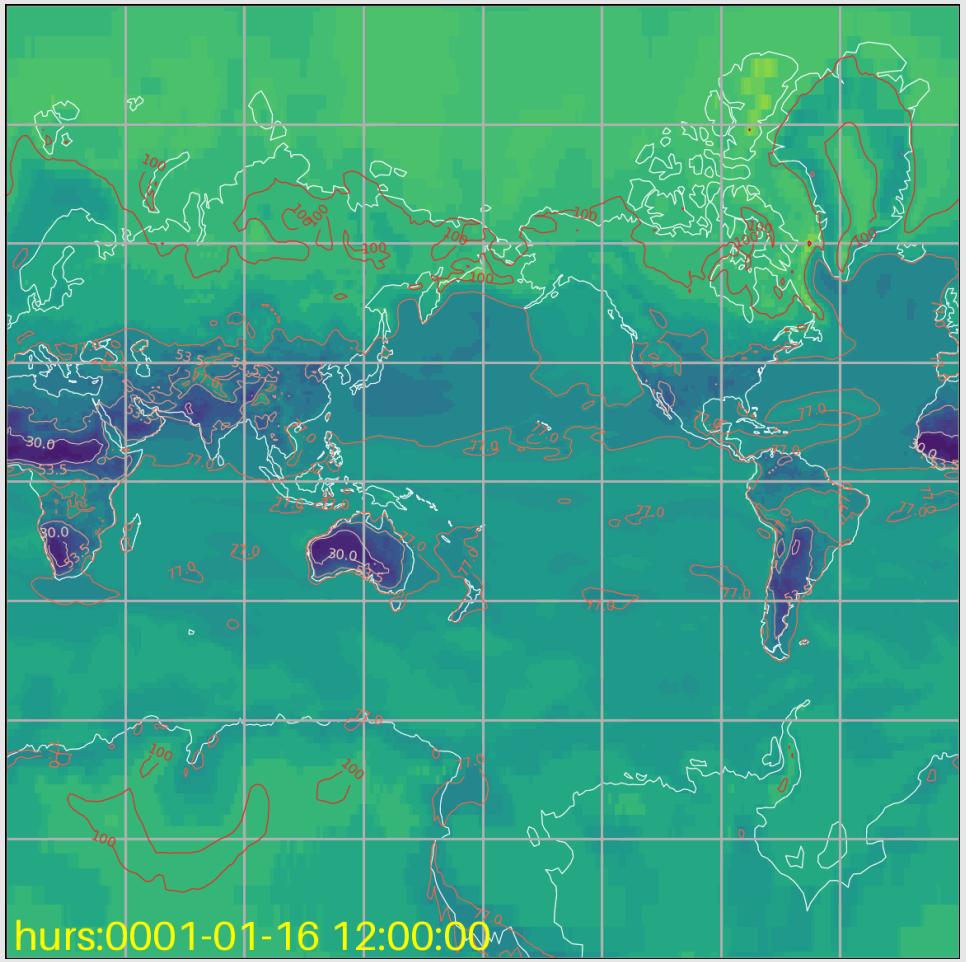




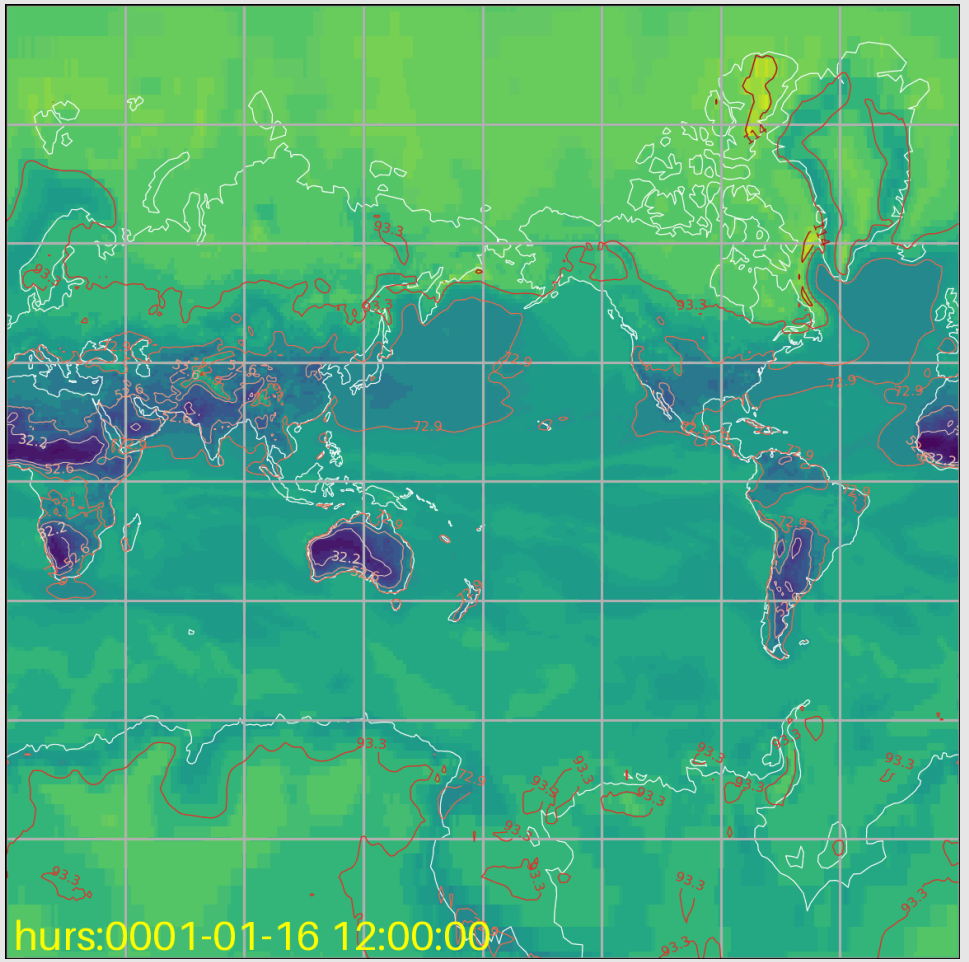


Select Data Collection => Search keys => Select series (3D) Select series (4D)

ANIMATE PLOTS



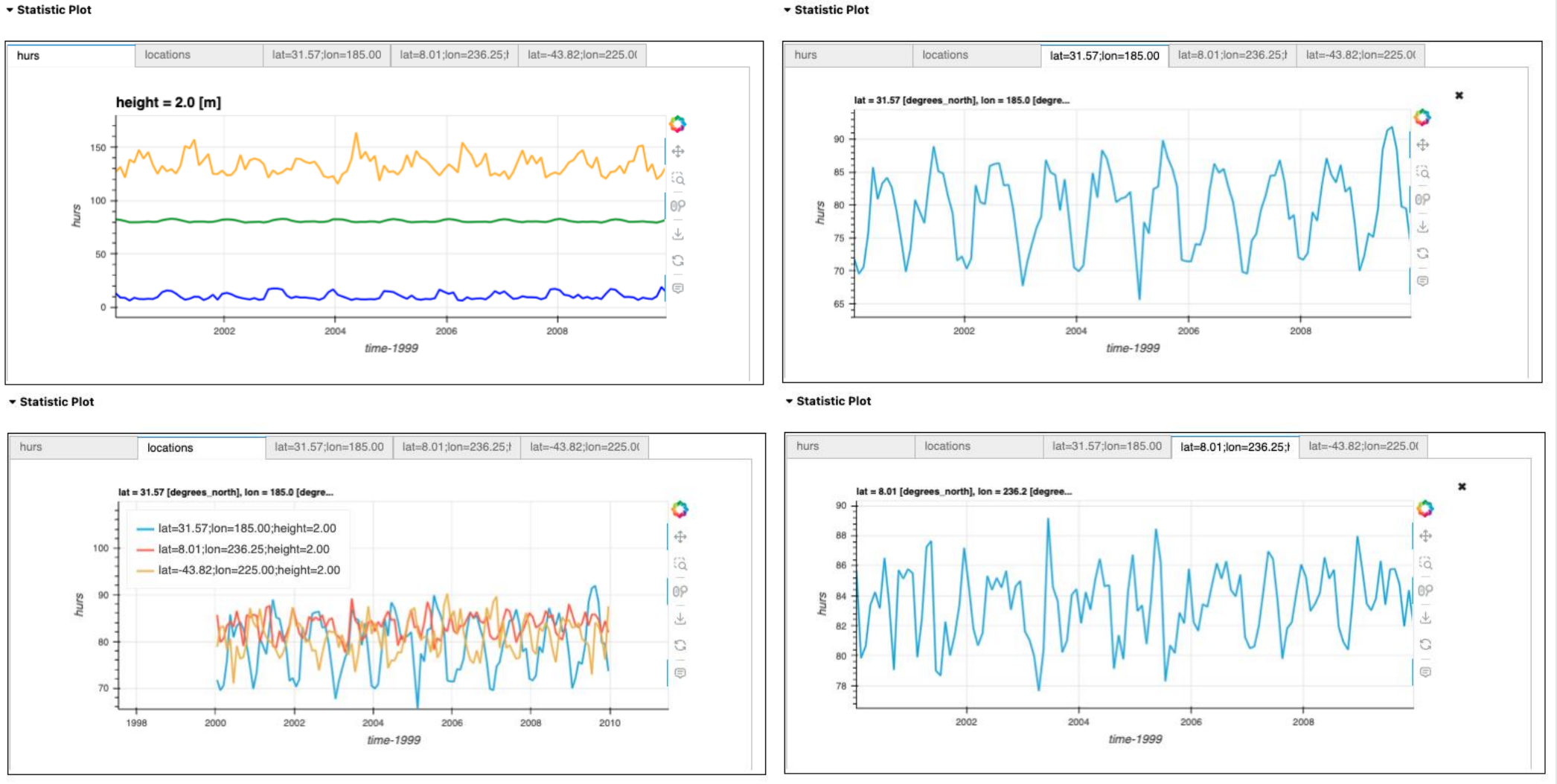
Sequential 12 months in 1year (nskip=0)



Fixed datetime in 10 years (nskip=12)

Relative Humidity at the Surface (in %)

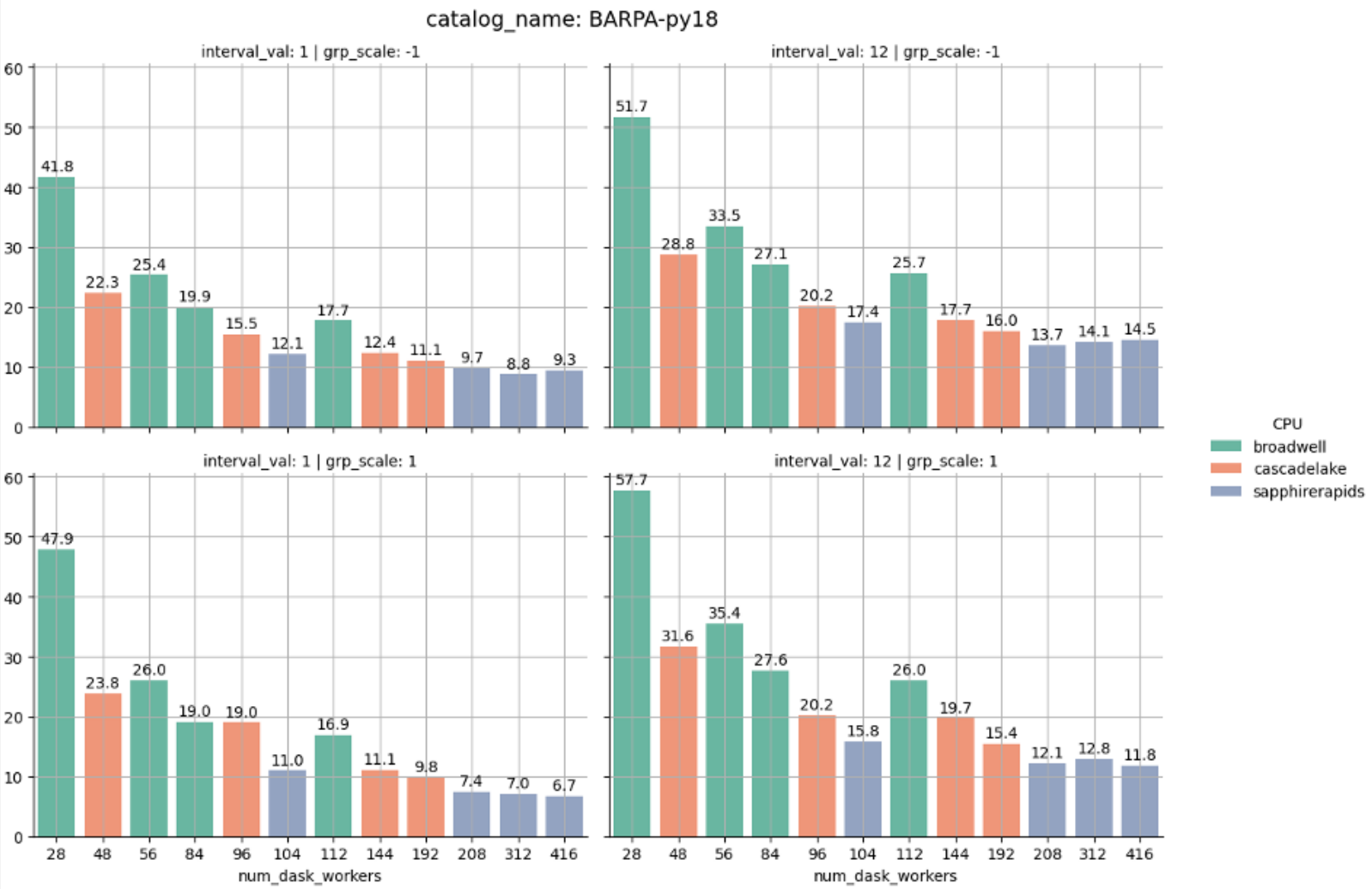
GLOBAL AND LOCALIZED STATISTICS



SCALABILITY & COMPUTE CONTROL

- Hotspots on statistics analysis and image rendering to overlays
- Access patterns may vary from sequential to non-contiguous slices
- Enhance scalability performance by tuning
  - Distributed computing platform covers multiple compute nodes and different CPU architectures
  - Load distribution over Dask cluster i.e. statically create 1 group per Dask worker or dynamically process 1 slice per Dask worker each time

SCALABILITY BENCHMARKING



CPU Architecture	CPU Model	CPU Freq(GHz)	Node Cores
Broadwell	E5-2690v4	2.6	2 x 14
CascadeLake	Platinum 8274	3.2	2 x 24
SapphireRapids	Platinum 8470Q	2.1	2 x 52

Total time slices= 1000

- **Scaling & Performance**
  - good scalability
  - Diminishing returns
- **Time Handling (interval_val)**
  - Sequential steps (interval_val=1) run faster.
  - Skipped slices (interval_val=12) add I/O & locality overheads.
- **CPU Architecture**
  - **Sapphire Rapids:** Best performance
  - **Cascade Lake:** Mid-tier
  - **Broadwell:** Slowest
- **Other Insights**
  - grp_scale: has minor impact compared to CPU & interval effects.
    - -1: Nchunk = number of slices
    - 1: Nchunk = number of cores
  - Hardware + workflow choices both are critical for efficiency.

SUMMARY & NEXT STEPS

- ClimViewer
  - fills the gap between local desktop tools and inflexible dashboards
  - offers powerful, interactive and scalable data visualization over large datasets
- Future work
  - Enhance search and custom analytics
  - Support additional formats or derived variables
  - Implement a wider range of data analysis workflow
  - Enable more scientific plot patterns

THANK YOU!