



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

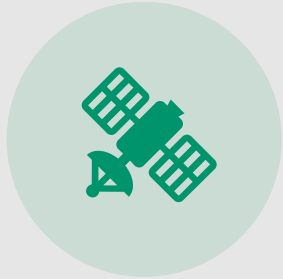
RUI YANG

SOFTWARE AND DATA MODERNISATION TEAM
NATIONAL COMPUTATIONAL INFRASTRUCTURE (NCI)

eResearch Conference 2025



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.

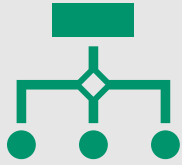


Prioritised with partners: ANU, Bureau of Meteorology, CSIRO, Geoscience Australia, ARC Centres of Excellence..



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

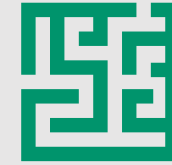
ECMWF Metview: Advanced desktop
tool with scripting & interactive plots



Web Platforms

ecCharts: 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

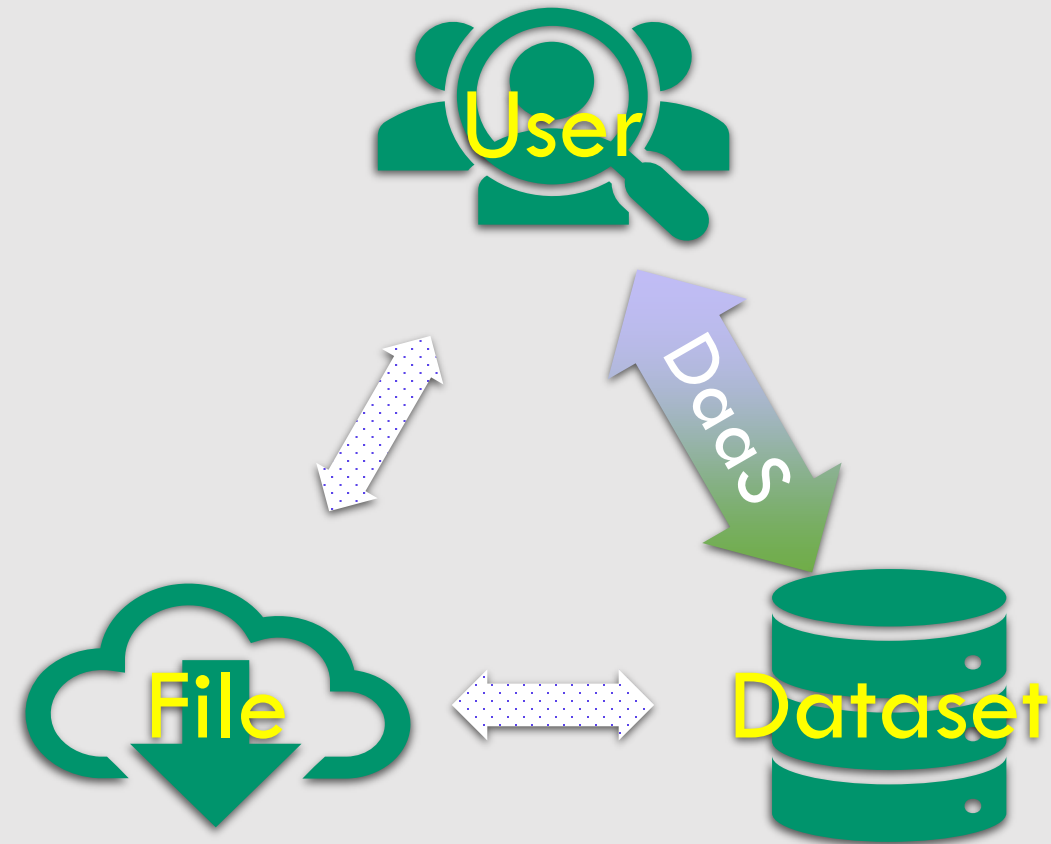
Existing tools
limited or inflexible

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
<pre>base_dir = "/g/data/rr3/publications/CMIP5/output1/CSI RO-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')</pre>	<pre>cat = intake.open_esm_datastore("cmip5_catalog.j son") subset = cat.search(variable_id="tas", model="MPI-ESM-LR") ds = subset.to_dataset_dict()["CMIP5.MPI- ESM-LR.tas"]</pre>

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**

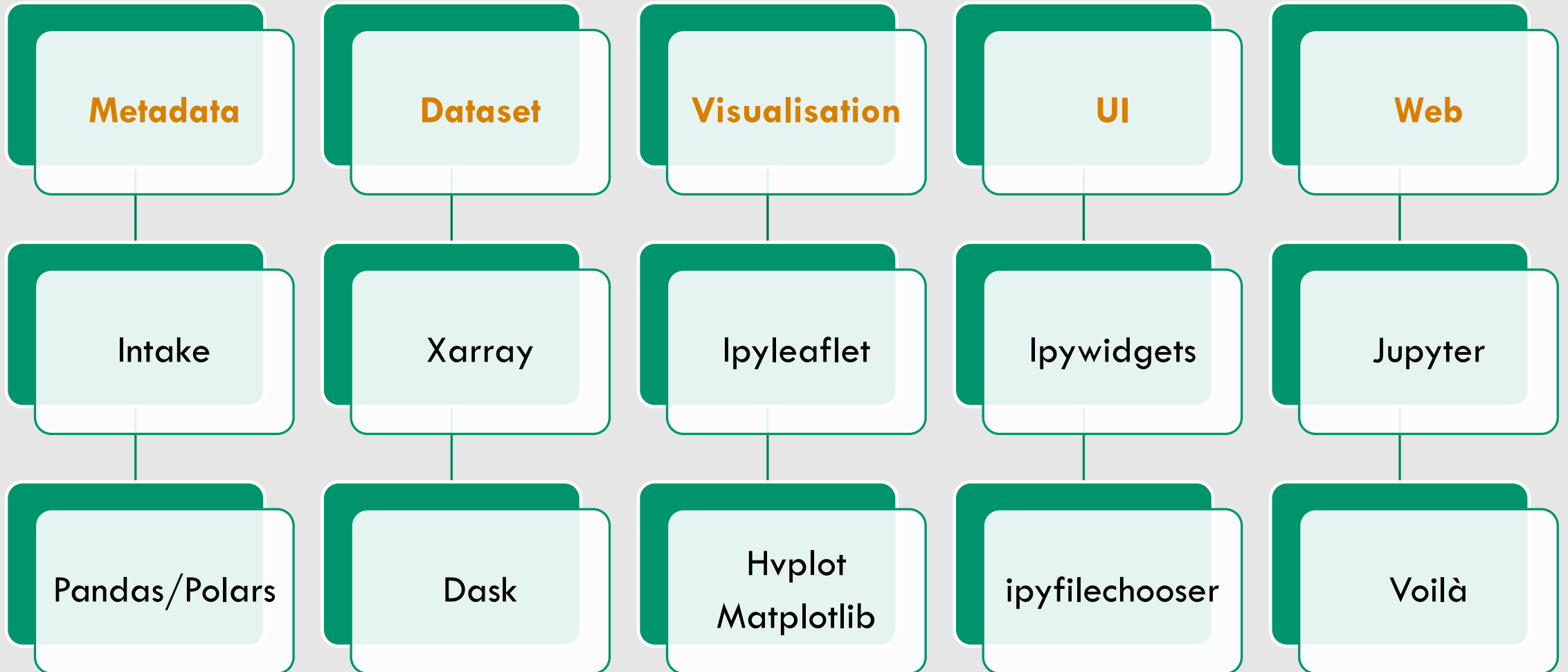
Global and local **statistics for selected location** from map

Customized plot such as color scheme, contours, labels etc.

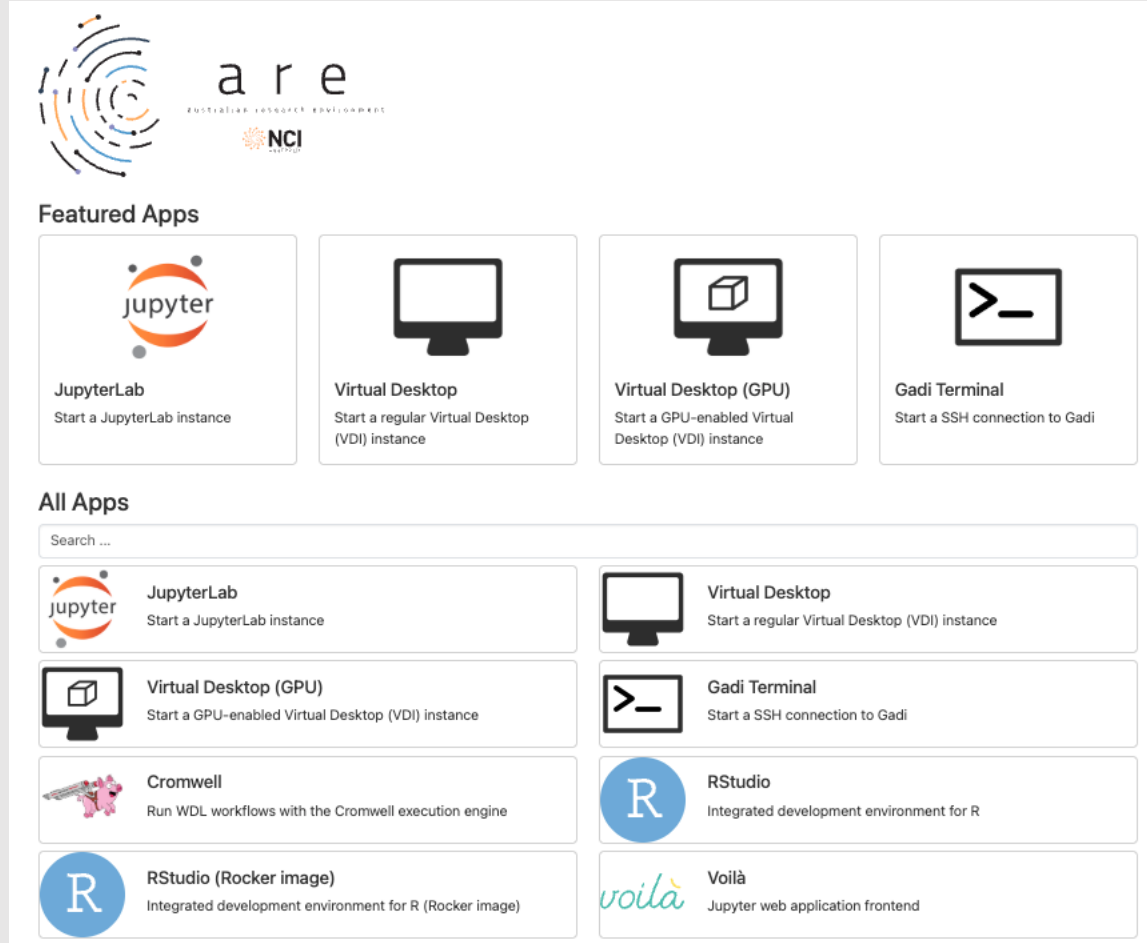
Dask-based **distributed processing** for performance and scalability

Support massive **computing resources** across multiple cores or nodes

CLIMVIEWER ARCHITECTURE



NCI INTEGRATION VIA ARE



The screenshot displays the ARE interface. At the top left is the ARE logo (Australian Research Environment) and the NCI logo. Below the logo is the text "are AUSTRALIAN RESEARCH ENVIRONMENT". The main content is divided into two sections: "Featured Apps" and "All Apps".

Featured Apps:

- JupyterLab:** Start a JupyterLab instance
- Virtual Desktop:** Start a regular Virtual Desktop (VDI) instance
- Virtual Desktop (GPU):** Start a GPU-enabled Virtual Desktop (VDI) instance
- Gadi Terminal:** Start a SSH connection to Gadi

All Apps:

Search ...

- JupyterLab:** Start a JupyterLab instance
- Virtual Desktop:** Start a regular Virtual Desktop (VDI) instance
- Virtual Desktop (GPU):** Start a GPU-enabled Virtual Desktop (VDI) instance
- Gadi Terminal:** Start a SSH connection to Gadi
- Cromwell:** Run WDL workflows with the Cromwell execution engine
- RStudio:** Integrated development environment for R
- RStudio (Rocker image):** Integrated development environment for R (Rocker image)
- Voilà:** Jupyter web application frontend

- 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 Voilà

are.nci.org.au

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

CLIMVIEWER INTERFACE



PROJECT

KEY

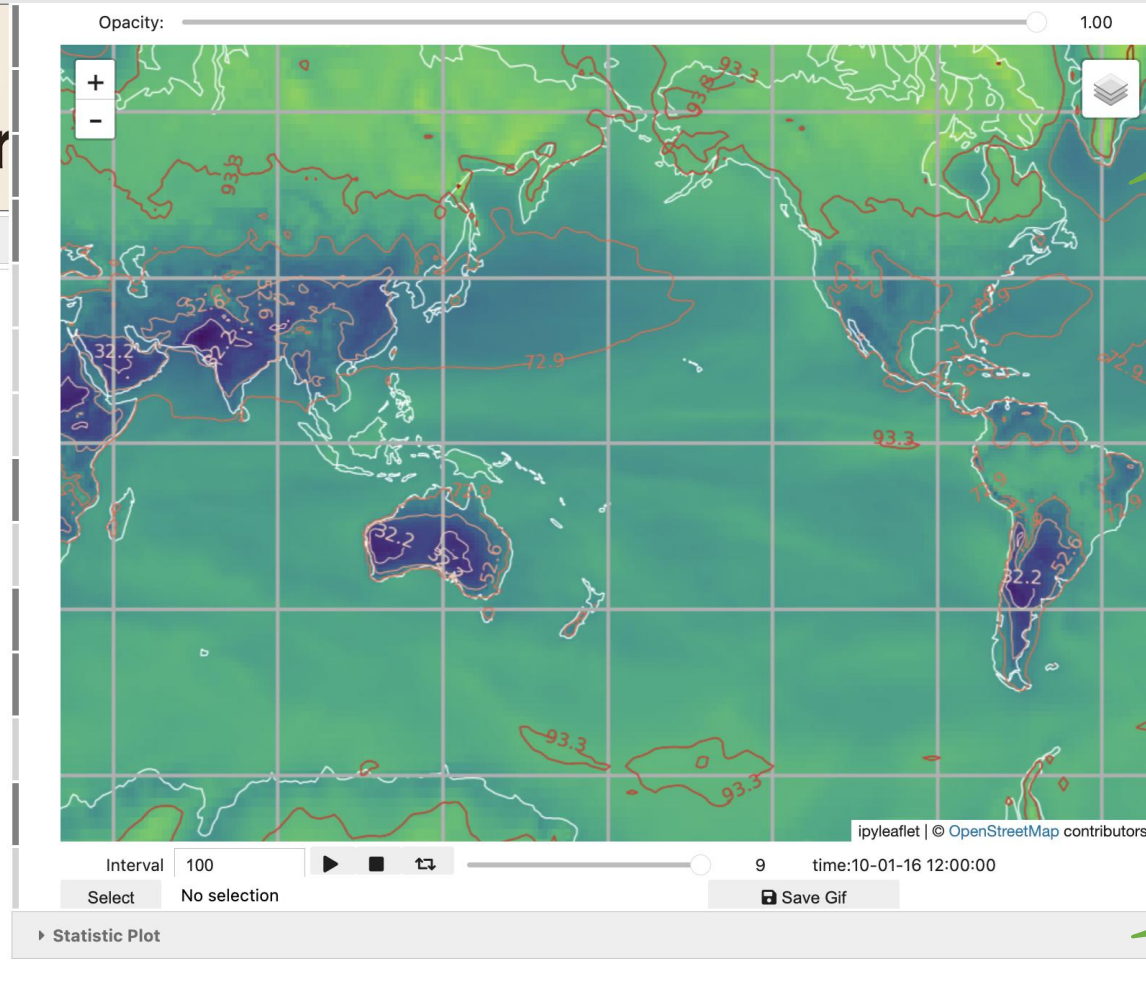
f.CMIP.NCC.NorESM2-MM.1pctCO2.r1i1p1f1.mon.atr
f.CMIP.NCC.NorESM2-MM.abrupt-4xCO2.r1i1p1f1.mo
f.CMIP.SNU.SAM0-UNICON.piControl.r1i1p1f1.mon.at
f.CMIP.SNU.SAM0-UNICON.piControl.r1i1p1f1.mon.at

time Start 0 : 1-01-16 12:00:00
End 119 : 10-12-16 12:00:00
Step 12 : 365 days, 0:00:00
Total Selected Entries: 10

Plot Style Settings

contours_level: 7
contours_line_width: 0.2
contours_cmap: Reds
plot_cmap: viridis
resolution: 300dpi

LOAD



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



PROJECT

STEP 1: Select Catalog

- BARRA2-ob53
- CMIP5-rr3
- CMIP6-oi10
- WeatherBench-wb00

Add Catalog

Select No selection

STEP 2: Query

KEY

Select Data Collection

PROJECT

STEP 1: Select Catalog

STEP 2: Query

loaded

Keys: file_type

Values: realm, frequency, table_id, project_id, institution_id, source_id, experiment_id, member_id, **variable_id**, grid_label, version

Time range

+ Add Key-Value

tas x 000101-001012 x

1pctCO2-rad x mon x

Query Reset

Search keys

PROJECT

STEP 1: Select Catalog

STEP 2: Query

loaded

Keys: frequency

Values: mon

Time range 000101-001012

+ Add Key-Value

tas x 000101-001012 x

1pctCO2-rad x mon x

Query Reset

=>

PROJECT

KEY

f.C4MIP.NCC.NorESM2-LM.1pctCO2-rad.r1i1p1f1.mor

time Start 0 : 1-01-16 12:00:00

End 119 : 10-12-16 12:00:00

Step 1 : 29 days, 12:00:00

Total Selected Entries: 120

Plot Style Settings

contours_level: 5

contours_line_width: 0.1

contours_cmap: plasma

plot_cmap: coolwarm

resolution: 300dpi

LOAD

Select series (3D)

PROJECT

KEY

f.1.40625deg.t
f.2.8125deg.t
f.5.625deg.t

time Start 0 : 1959-01-01 00:00:00

End 8759 : 1959-12-31 23:00:00

Step 1 : 0 days 01:00:00

Total Selected Entries: 8760

level 50

Plot Style Settings

contours_level: 100

contours_line_width: 150

contours_cmap: 200

contours_resolution: 250

plot_cmap: 300

resolution: 400

500

600

700

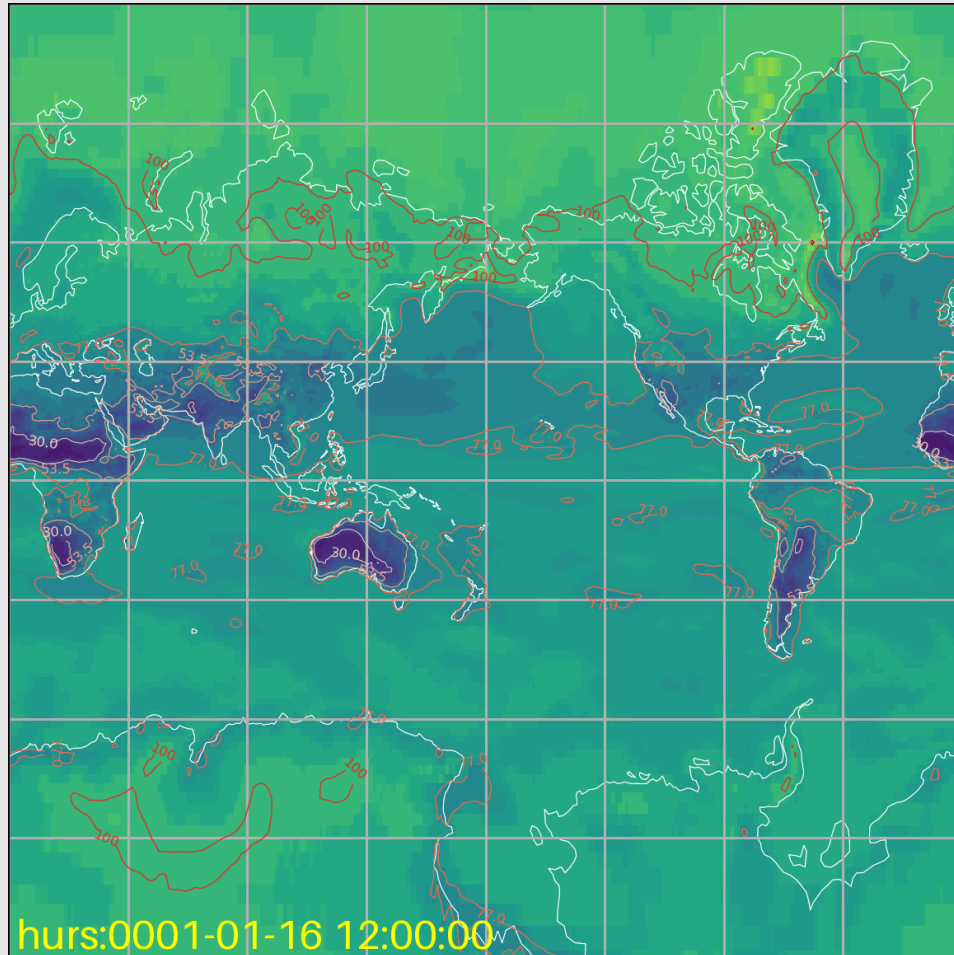
850

925

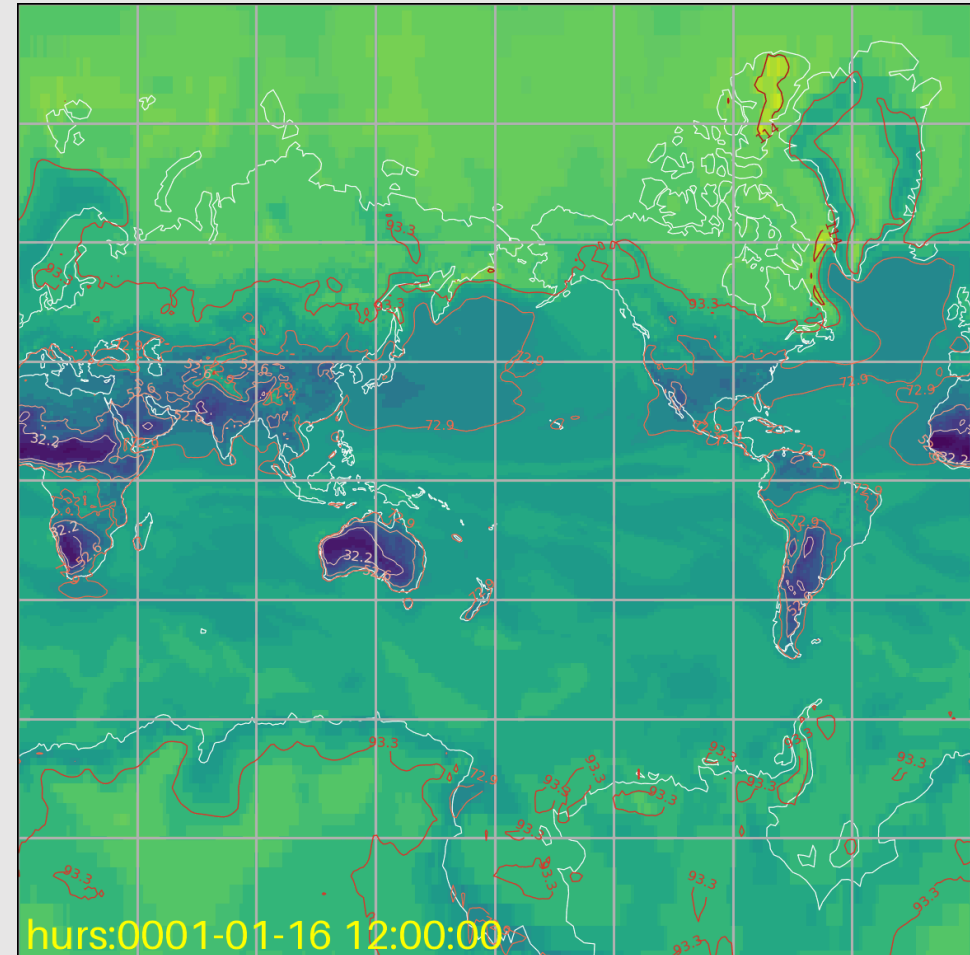
1000

Select series (4D)

ANIMATE PLOTS



Sequential 12 months in 1 year (nskip=0)

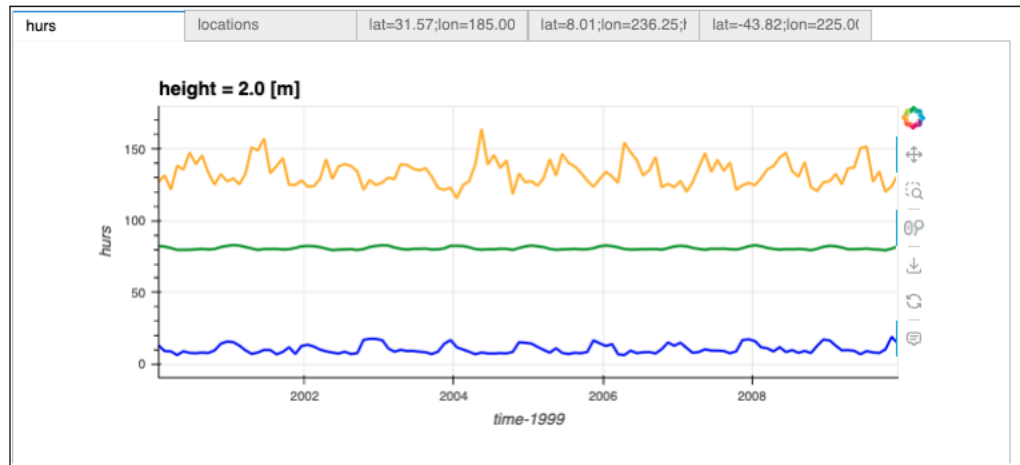


Fixed datetime in 10 years (nskip=12)

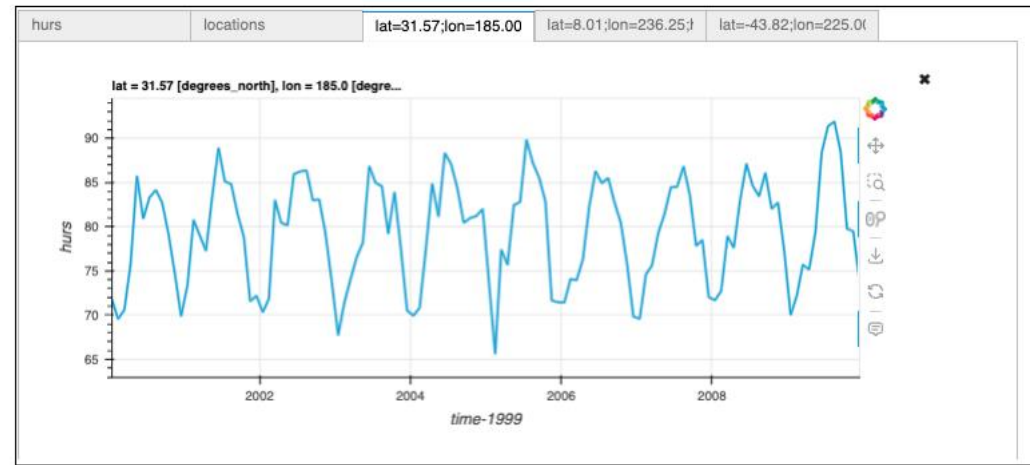
Relative Humidity at the Surface (in %)

GLOBAL AND LOCALIZED STATISTICS

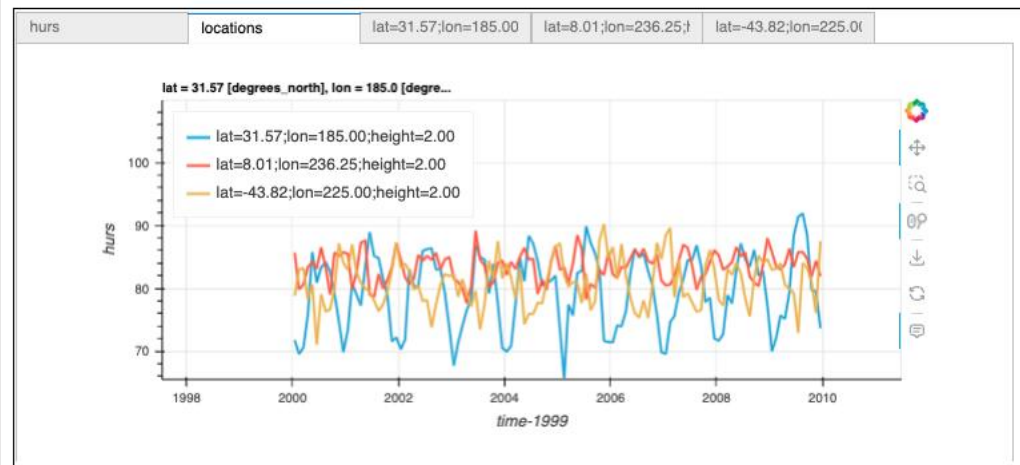
Statistic Plot



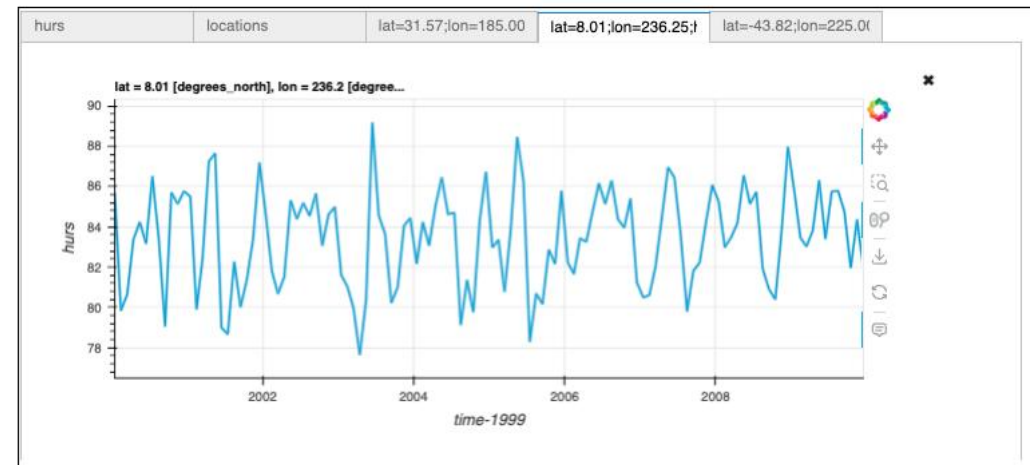
Statistic Plot



Statistic Plot



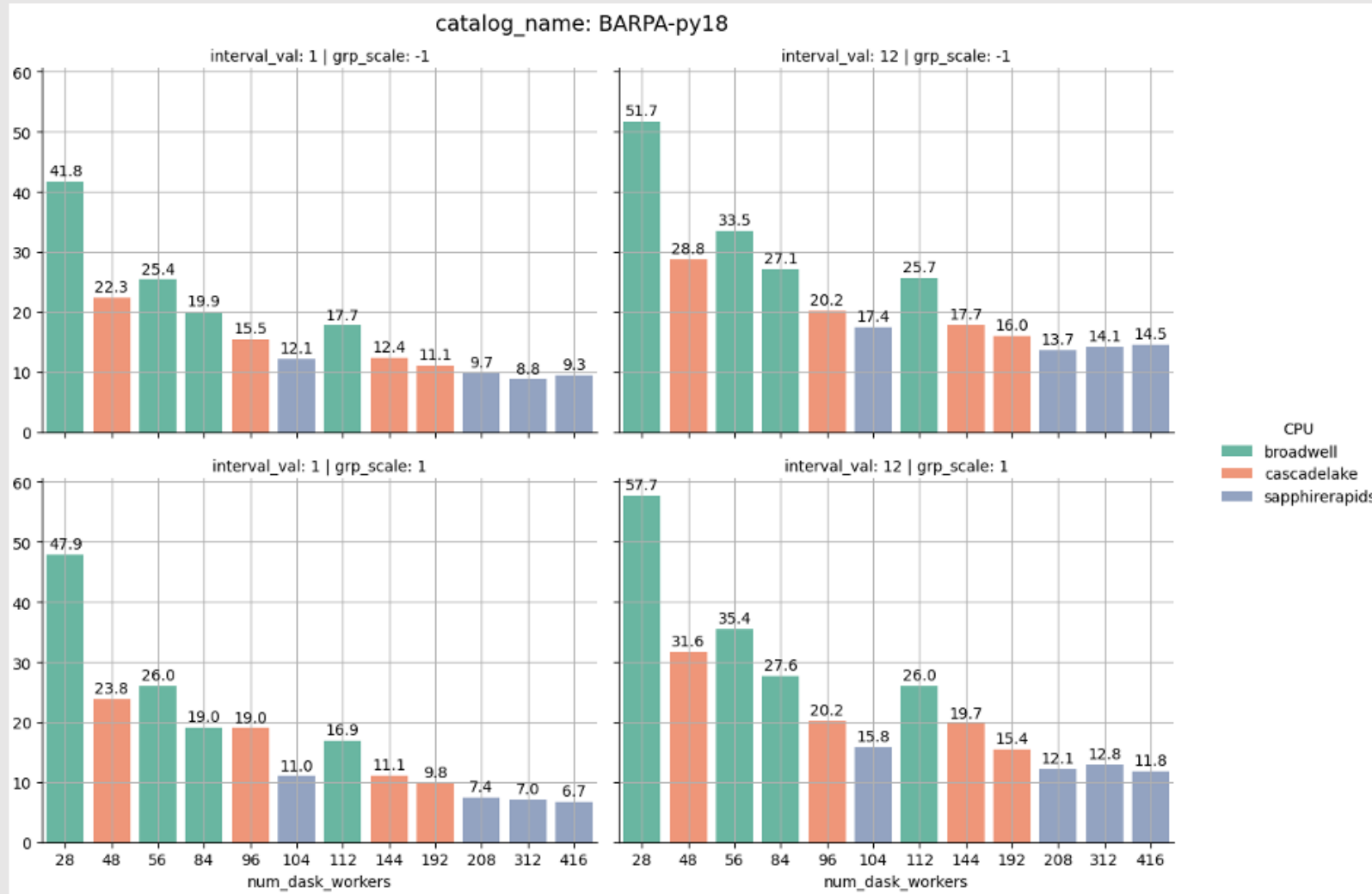
Statistic Plot



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!