

A Python-oriented environment for climate experiments at scale in the frame of the European Open Science Cloud

EGU General Assembly 2020
Online | 4-8 May 2020
D806 | EGU2020-17031

D. Elia^{1,3}, F. Antonio¹, C. Palazzo¹, P. Nassisi¹, S. Bendoukha², R. Kwee-Hinzmann², S. Fiore¹, T. Weigel², H. Thiemann², and G. Aloisio^{1,3}

¹ *Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Lecce, Italy*

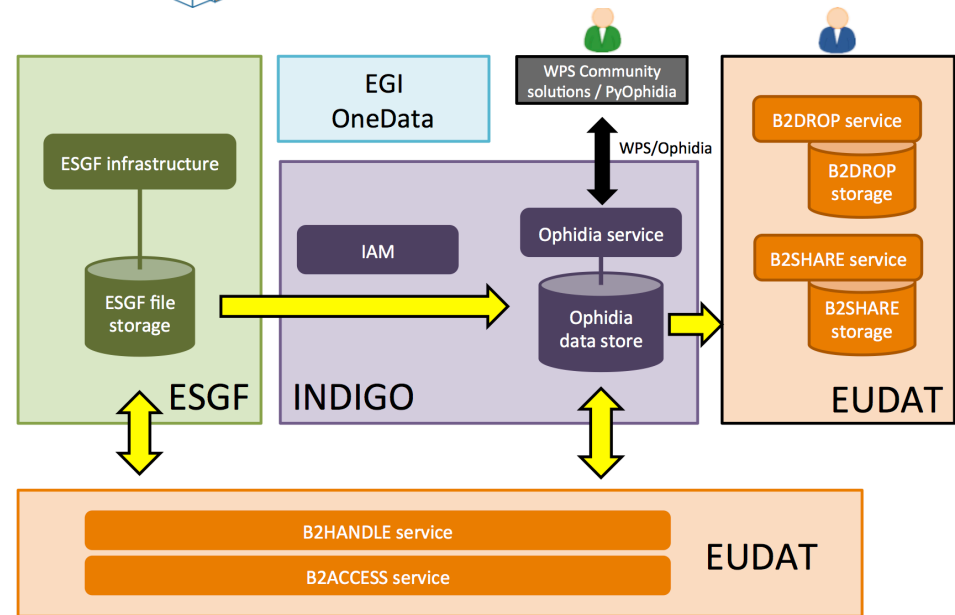
² *Deutsches Klimarechenzentrum (DKRZ), Hamburg, Germany*

³ *University of Salento, Lecce, Italy*



ENES Climate Analytics Service (ECAS)

- ✓ The *ENES Climate Analytics Service (ECAS)*, proposed by CMCC & DKRZ in the *EU H2020 EOSC-Hub project*, supports climate data analysis
- ✓ It is one of the *EOSC-Hub Thematic Services*
- ✓ ECAS builds on top of the *Ophidia HPDA framework* integrated with components from INDIGO-DataCloud, EUDAT and EGI



The European Commission launched the European Open Science Cloud (EOSC) Initiative to capitalise on the data revolution. EOSC will provide European science, industry and public authorities with world-class digital infrastructure that bring state of the art computing and data storage capacity to the fingertips of any scientists and engineer in the EU.

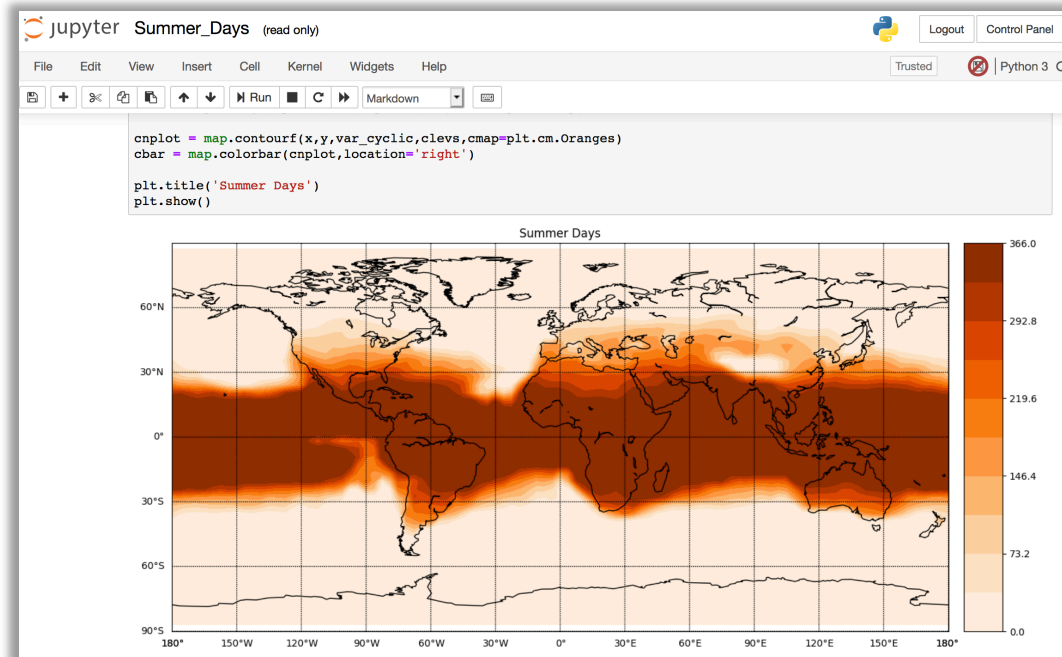
ECASLab: a Python environment for data analysis

ECASLab provides a user-friendly environment for scientific analysis based on:

- ✓ The ECAS integrated service
- ✓ A JupyterHub instance providing a graphical environment for user's experiments
- ✓ Bundled with a wide set of Python scientific modules for data manipulation, analysis and visualization, such as PyOphidia, NumPy, Pandas, Dask, Matplotlib, basemap, Cartopy
- ✓ A set of ECAS usage example notebooks (<https://github.com/ECAS-Lab/ecas-notebooks>)

Two major instances are hosted by:

- ✓ CMCC <https://ecaslab.cmcc.it>
- ✓ DKRZ <https://ecaslab.dkrz.de>



The Ophidia project

Ophidia (<http://ophidia.cmcc.it>) is a CMCC Foundation research project addressing data challenges for eScience¹

It provides:

- ✓ a *High Performance Data Analytics* (HPDA) framework for multi-dimensional scientific data joining HPC paradigms with scientific data analytics approaches
- ✓ in-memory and server-side data analysis exploiting parallel computing techniques and database approaches
- ✓ a multi-dimensional, array-based, storage model and partitioning schema for scientific data leveraging the datacube abstraction
- ✓ end-to-end mechanisms to support complex experiments and large workflows on scientific datacubes, primarily in climate domain



Ophidia



cmcc
Centro Euro-Mediterraneo
sui Cambiamenti Climatici

Programmatic access through the PyOphidia class

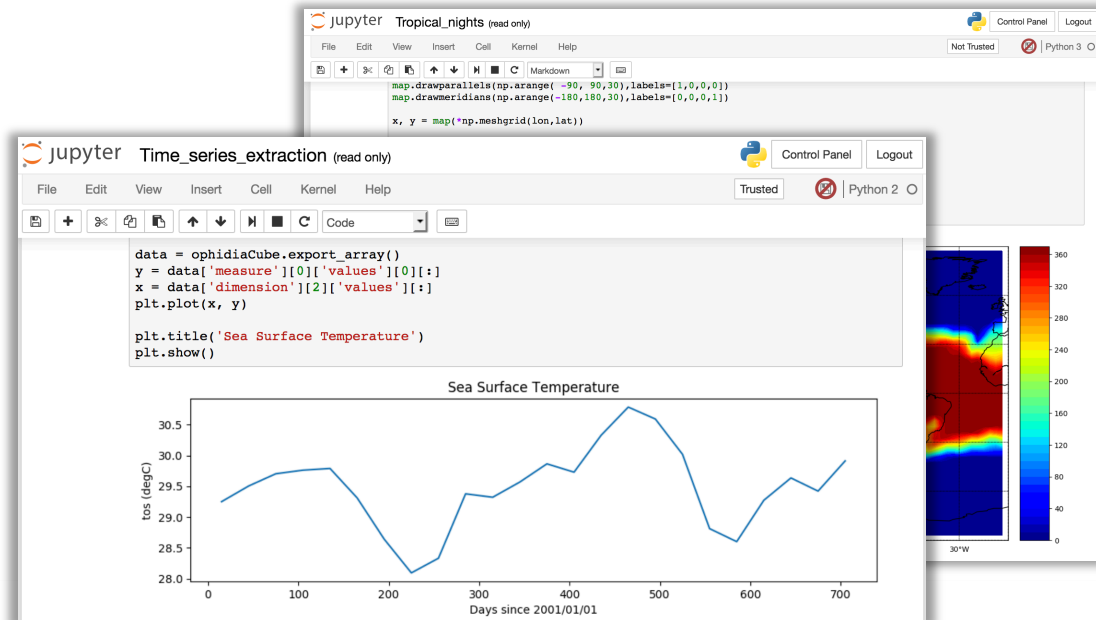
- ✓ *PyOphidia* provides a Python interface to submit commands to the Ophidia Server and to retrieve/deserialize the results (e.g. in Jupyter Notebooks)
- ✓ Two modules available:
 - ✓ *Client class*: supports the submissions of Ophidia commands and workflows, as well as the management of session
 - ✓ *Cube class*: provides the datacube type abstraction and the methods to manipulate, process and get information on cubes objects

```
from PyOphidia import cube, client
cube.Cube.setclient(read_env=True)

mycube =
cube.Cube.importnc(src_path='/public/data/ecas_training
/file.nc', measure='tos', imp_dim='time',
import_metadata='yes', ncores=5)
mycube2 = mycube.reduce(operation='max', ncores=5)
mycube3 = mycube2.rollup(ncores=5)
data = mycube3.export_array()

mycube3.exportnc2(output_path='/home/test',
export_metadata='yes')
```

<https://github.com/OphidiaBigData/PyOphidia>
<https://pypi.org/project/PyOphidia/>
<https://anaconda.org/conda-forge/pyophidia>



ECAS on the EGI Federated Cloud Infrastructure

ECAS has been integrated into the EGI FedCloud, considering two scenarios¹:

- ✓ A ready-to-use ECAS single-node VMI is available from the EGI AppDB

<https://appdb.egi.eu/store/vappliance/ecas>

- ✓ A multi-node ECAS cluster can be dynamically provisioned on the EGI FedCloud through the Elastic Cloud Computing Cluster (EC3)

The screenshot shows the EGI Applications Database interface. At the top, there are navigation tabs: Home, Software Marketplace, Cloud Marketplace, and People. The main content area displays the details for the ECAS virtual appliance. It includes a title 'ECAS', a description 'This Virtual Appliance relates to ECAS, the ENES Climate Analytics Service', and a category 'Application Servers • Big Data'. There are also tags for 'big data', 'climate', 'data analytics', 'eScience', and 'FAIR'. A 'Browse & Download Images' button is visible, along with information about the latest version (1.0) and supported hypervisors (VMWare).

The screenshot shows the EC3 Elastic Cloud Computing Cluster interface. The main heading is 'MANAGE YOUR CLUSTERS DEPLOYED WITH EC3'. Below this, there is a section for 'CONFIGURE YOUR CLUSTER'. The 'CLUSTER CONFIGURATION' section is active, and a dropdown menu is open, showing the selection of 'ECAS' as the LRMS (Local Resource Management System) for the cluster. Other options in the dropdown include SLURM, Torque, SGE, Mesos + Marathon + Chronos, Kubernetes, OSCAR, and Nomad. A 'NEXT' button is visible at the bottom right of the configuration panel.

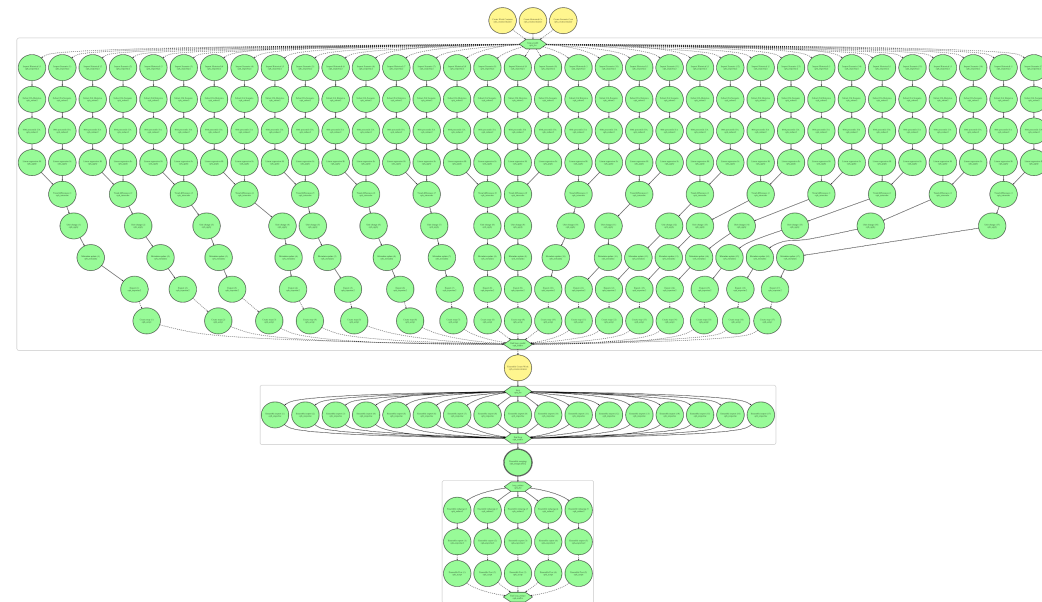
1. Elastic deployment of ECAS on EGI: <https://www.egi.eu/about/newsletters/elastic-deployment-of-ecas-on-egi/>

A complete environment for climate experiments

ECASLab provides a complete environment for supporting scientist in their daily research activities with a focus on those from the climate change domain

- ✓ It represents a single entrypoint to *analysis tools*, *scientific datasets* (e.g., from ESGF data archive) and *computing resources*
- ✓ It provides the capabilities for the implementation and execution of both interactive and complex experiments (workflows), such as *multi-model CMIP-based data analysis*¹

ECAS is also one of the *compute services* made available to climate scientists by the *EU H2020 IS-ENES3 project*



1. S. Fiore, D. Elia, C. Palazzo, A. D'Anca, F. Antonio, D. N. Williams, I. Foster, G. Aloisio, "Towards an Open (Data) Science Analytics-Hub for Reproducible multi-model Climate Analysis at Scale", 2018 IEEE International Conference on Big Data (Big Data)

Free access to computing platforms for multi-model climate data analyses for CMIP6 and CORDEX !



The European Research Infrastructure for climate modelling IS-ENES offers a new charge-free server-side computing service:

- ✓ Reduce data transfer issues
- ✓ Direct access to petabytes of model data
- ✓ Run your own scripts and Jupyter notebooks
- ✓ Speed up the computational analysis

Information on the call:

<https://portal.enes.org/data/data-metadata-service/analysis-platforms>

Learn more at EGU sessions:

- ✓ CL5.7 Climate Services - Underpinning Science , 05 May, 10:45–12:30, EGU 2020-19121: <https://meetingorganizer.copernicus.org/EGU2020/session/36737>
- ✓ CL2.6 Detecting and attributing climate change: trends, extreme events, and impacts, 07 May, 08:30–10:15, EGU 2020-19340: <https://meetingorganizer.copernicus.org/EGU2020/session/36768>

Acknowledgment



These activities are supported in part by EOC-Hub and IS-ENES3 projects:

- ✓ *EOSC-hub receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 777536*
- ✓ *IS-ENES3 is a project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 824084*