Ophidia: High Performance Data Analytics framework for Climate Change at scale

Donatello Elia[1,2]

[1] Advanced Scientific Computing Division, Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Lecce, Italy
[2] University of Salento, Lecce, Italy

IS-ENES3 Virtual workshop on requirements for a fast and scalable evaluation workflow

May 19, 2021

Outline

- *Climate data management and analytics challenges*
- *The Ophidia HPDA Framework*
- *Key aspects: execution model, parallelization, HPC integration*
- *Data and metadata operators*
- *Integration with Python*
- *Experimental results*
- *Summary*

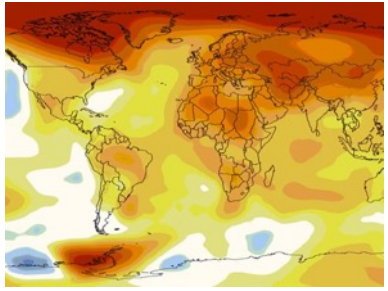
Climate analysis challenges & issues

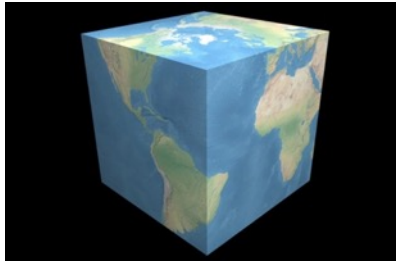
Key challenges and practical issues related to large-scale climate analysis

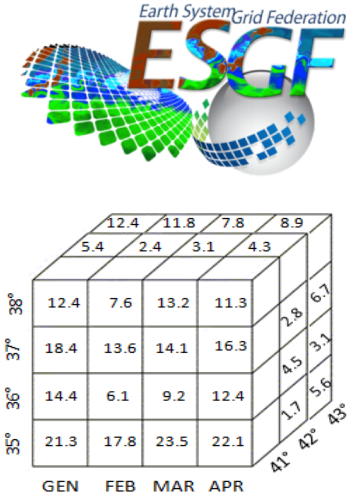
- Setup of a data analysis experiment requires the ***download of (multiple) input data***
  - *Reducing data movement is essential*
- The complexity of the analysis leads to the need for ***workflow support***
  - *Data analysis* mainly relies on *client-side* approaches
- Large data volumes pose **strong requirements** in terms of **computational** and **storage resources**
  - Need for novel HPDA solutions able to scale the analysis
  - Difficulties in exploiting data analysis solutions in HPC environments

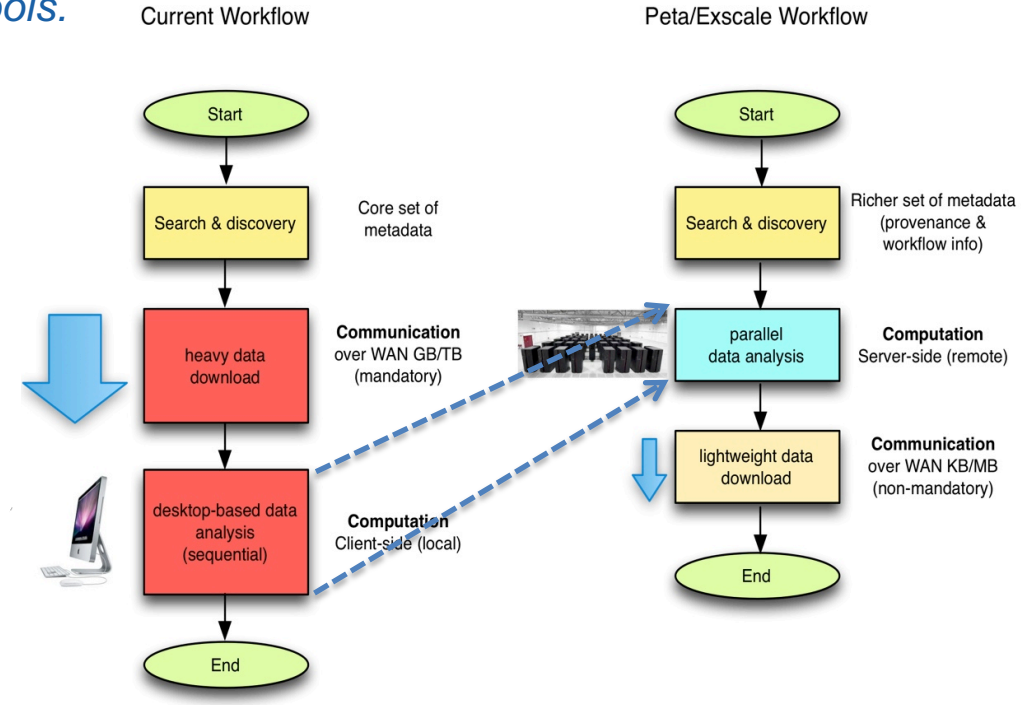
A paradigm shift

Volume, variety, velocity are key challenges for big data in general and for climate change science in particular. Client-side, sequential and disk-based workflows are three limiting factors for the current scientific data analysis tools.









S. Fiore, A. D'Anca, C. Palazzo, I. Foster, D. N. Williams, G. Aloisio, "Ophidia: toward bigdata analytics for eScience", ICCS2013 Conference, Procedia Elsevier, 2013

Ophidia HPDA framework

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

- A **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
- Multi-dimensional, array-based, storage model and partitioning schema for scientific data leveraging the **datacube** abstraction
- End-to-end mechanisms to support **interactive analysis, complex experiments** and **large workflows** on scientific data

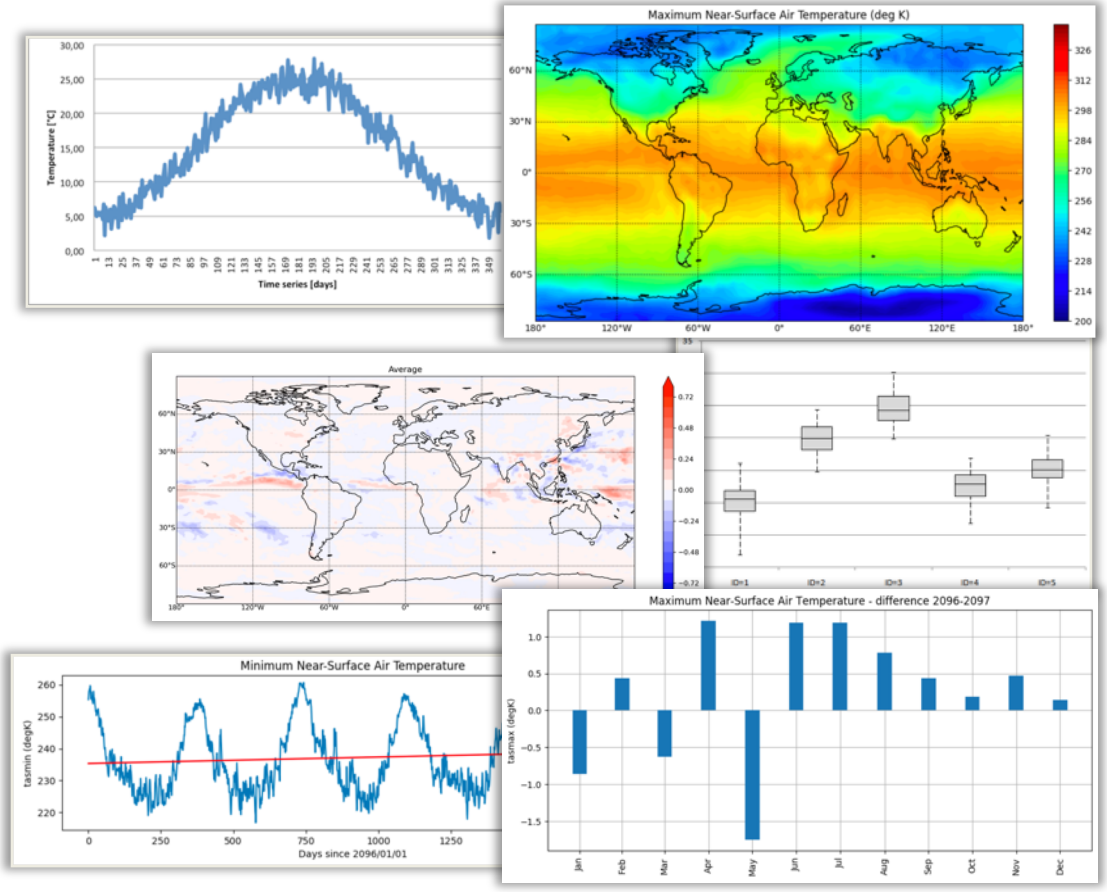
Ophidia

S. Fiore, D. Elia, C. Palazzo, F. Antonio, A. D'Anca, I. Foster, G. Aloisio, "Towards High Performance Data Analytics for Climate Change", ISC High Performance 2019, LNCS Springer, 2019

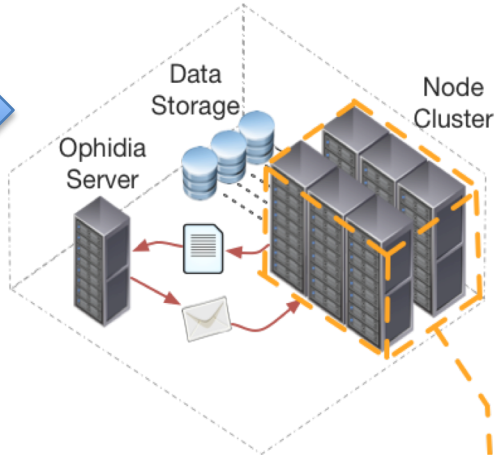
Data analytics requirements and use cases

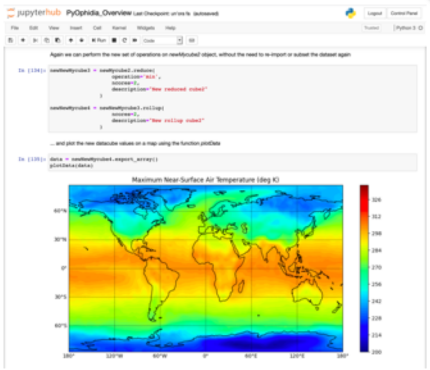
Requirements and needs focus on:

- *Time series analysis*
- *Data subsetting*
- *Model intercomparison*
- *Multi-model means*
- *Massive data reduction*
- *Data transformation*
- *Parameter sweep experiments*
- *Maps generation*
- *Ensemble analysis*
- *Data analytics workflow support*



Server-side paradigm and execution modes





Oph_Term: a terminal-like commands interpreter serving as a client for the Ophidia framework

PyOphidia: a Python interface for datacube management & analytics with Ophidia

Multiple execution modes:

- *Interactive data analysis*
- *Batch processing*
- *Python notebooks and applications*
- *Workflows of operators*

Ophidia Architecture

Multi-interface **interoperable** front-end

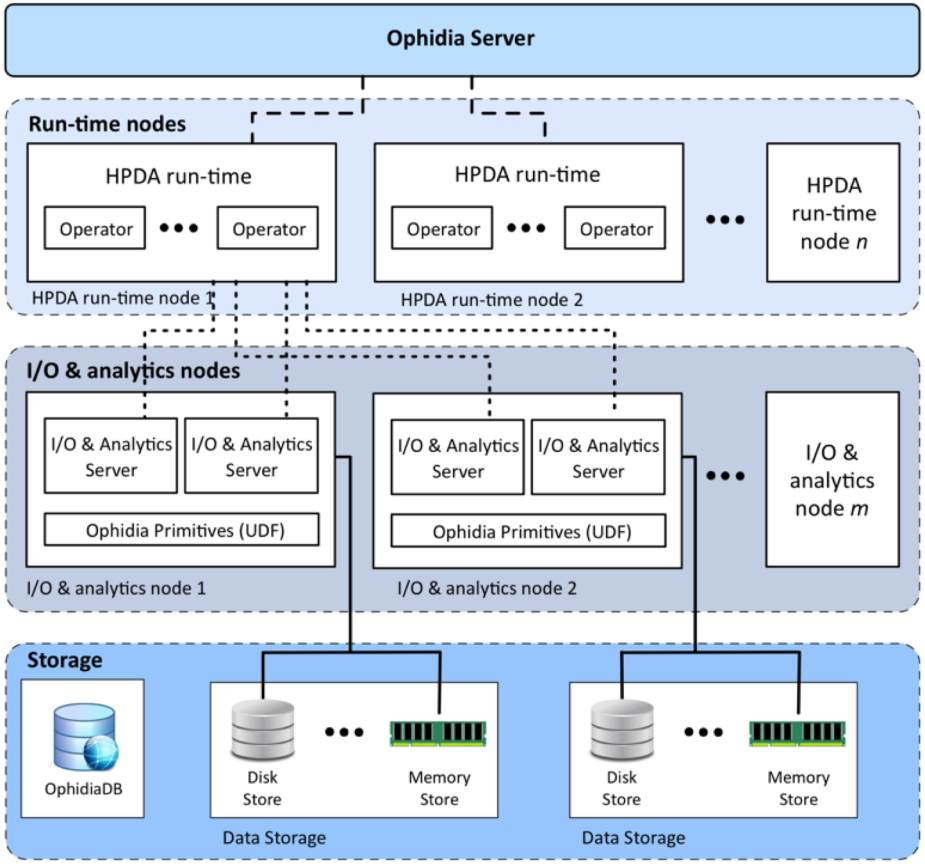
Modular and **extensible** software stack

Multi-level **parallel** execution model:

- Datacube-level (HTC-based)
- Fragment-level (HPC-based: MPI+X)

Support for **in-memory analytics** through array-based primitives (UDF kernels)

Data **partitioned** and **distributed** hierarchically over the storage and managed by the I/O & analytics nodes



D. Elia, S. Fiore and G. Aloisio, "Towards HPC and Big Data Analytics Convergence: Design and Experimental Evaluation of a HPDA Framework for eScience at Scale," in IEEE Access (Early Access), https://ieeexplore.ieee.org/document/9428012

On-demand deployment on HPC infrastructures

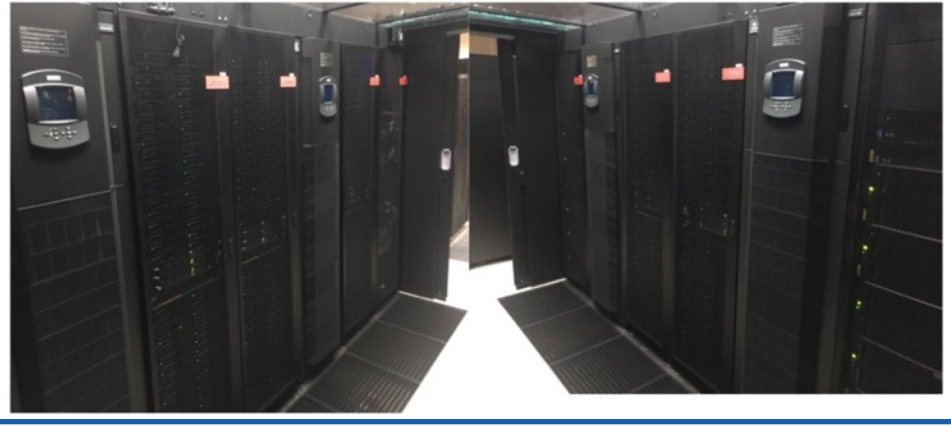
Target environment: *HPC cluster*

On-demand deployment of I/O & analytics servers

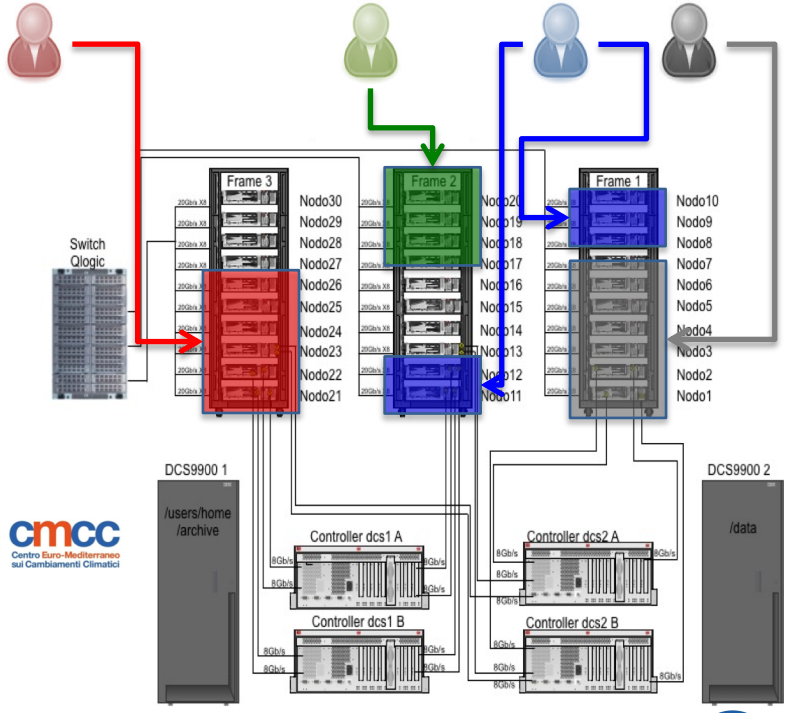
- `oph_cluster action=deploy;nhost=64;cluster_name=new;`
- `oph_cluster action=undeploy;cluster_name=new;`

Transparent interaction with scheduling systems

Zeus SuperComputer at CMCC: 1.2 PetaFlops, 348 nodes



Multiple isolated instances can be deployed simultaneously by different teams/users



Ophidia operators

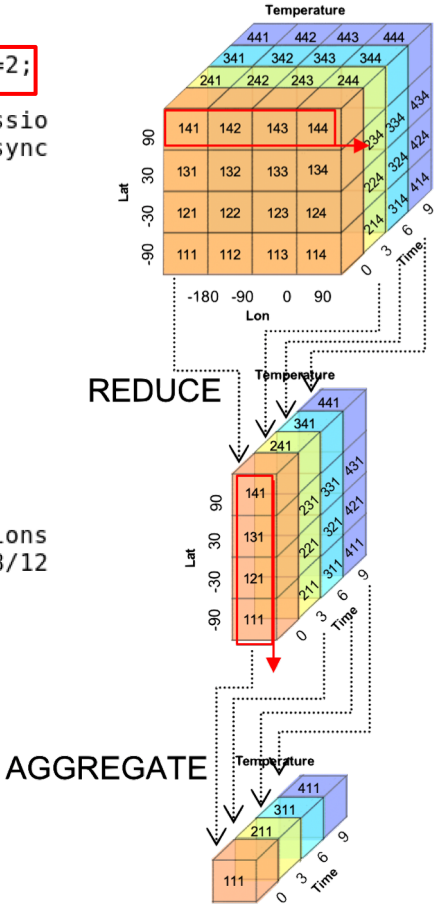
CLASS	PROCESSING TYPE	OPERATOR(S)
I/O	Parallel	OPH_IMPORTNC, OPH_EXPORTNC, OPH_CONCATNC, OPH_RANDUCUBE
Time series processing	Parallel	OPH_APPLY
Datacube reduction	Parallel	OPH_REDUCE, OPH_REDUCE2, OPH_AGGREGATE
Datacube subsetting	Parallel	OPH_SUBSET
Datacube combination	Parallel	OPH_INTERCUBE, OPH_MERGECUBES
Datacube structure manipulation	Parallel	OPH_SPLIT, OPH_MERGE, OPH_ROLLUP, OPH_DRILLDOWN, OPH_PERMUTE
Datacube/file system management	Sequential	OPH_DELETE, OPH_FOLDER, OPH_FS
Metadata management	Sequential	OPH_METADATA, OPH_CUBEIO, OPH_CUBESCHEMA
Datacube exploration	Sequential	OPH_EXPLORECUBE, OPH_EXPLORENC

About 50 operators for data and metadata processing

Ophidia operators documentation: http://ophidia.cmcc.it/documentation/users/operators/index.html

“data” operators

```
[12..3289] >> oph_reduce cube=http://127.0.0.1/ophidia/418/12717;operation=avg;ncores=2;nthreads=2;
[Request]:
operator=oph_reduce;cube=http://127.0.0.1/ophidia/418/12717;operation=avg;ncores=2;nthreads=2;sessio
nid=http://127.0.0.1/ophidia/sessions/127028404128222463341617004437753289/experiment;exec_mode=sync
;cwd=/;cdd=/;host_partition=auto;

[JobID]:
http://127.0.0.1/ophidia/sessions/127028404128222463341617004437753289/experiment?239#582

[Response]:
Output Cube
-----------

http://127.0.0.1/ophidia/418/12722

Execution time: 0.35 seconds
[12..3289] >> oph_aggregate operation=avg;ncores=2;nthreads=2;
[Request]:
operator=oph_aggregate;operation=avg;ncores=2;nthreads=2;sessionid=http://127.0.0.1/ophidia/sessions
/127028404128222463341617004437753289/experiment;exec_mode=sync;cube=http://127.0.0.1/ophidia/418/12
722;cwd=/;cdd=/;host_partition=auto;

[JobID]:
http://127.0.0.1/ophidia/sessions/127028404128222463341617004437753289/experiment?240#584

[Response]:
Output Cube
-----------

http://127.0.0.1/ophidia/418/12723

Execution time: 0.17 seconds
```

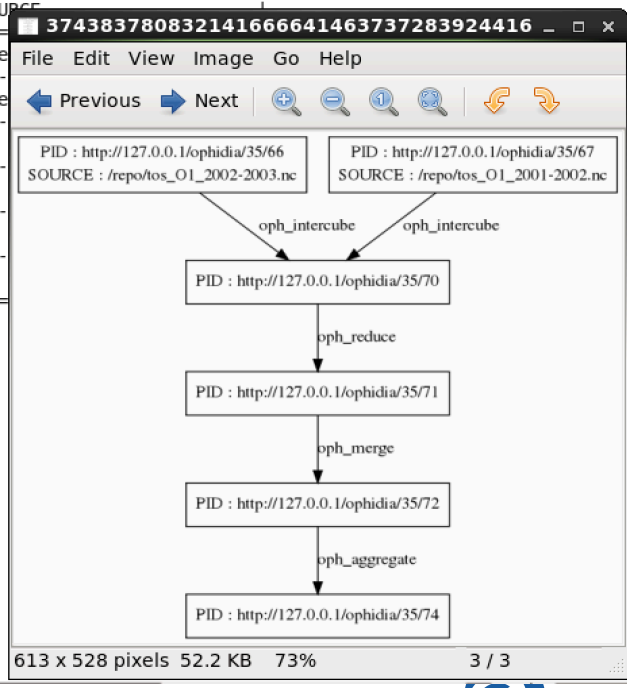


“data” operators

```
[37..4416] >> oph_explorecube cube=http://127.0.0.1/ophidia/35/67;subset_dims=lat|lon|time;subset_filter=39:42|15:19|1:275;show_time=yes;
[Request]:
operator=oph_explorecube;cube=http://127.0.0.1/ophidia/35/67;subset_dims=lat|lon|time;subset_filter=39:42|15:19|1:275;show_time=yes;sessionid=http://127.0.0.1/ophidia/sessions/3
7438378083214166664146373728392 4416/experiment;exec_mode=sync;ncores=1;cwd=/;

[JobID]:
http://127.0.0.1/ophidia/sessions/37438378083214166664146373728392 4416/experiment?106#224

[Response]:
tos
---
+===========+===========+=============================================================================================================================================+
| lat       | lon       | tos                                                                                                                                         |
+===========+===========+=============================================================================================================================================+
| 39.500000 | 15.000000 | 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20 |
|-----------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------|
| 39.500000 | 17.000000 | 287.3930664062, 286.8287048340, 286.5860595703, 286.9228210449, 288.5254516602, 292.3968200684, 295.8656921387, 297.2062072754, 295.7126464844 |
|-----------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------|
| 39.500000 | 19.000000 | 287.6926879883, 287.0508117676, 286.7896118164, 287.0781555176, 288.6802062988, 292.6882629395, 296.4769287109, 297.6632385254, 296.3418273926 |
|-----------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------|
| 40.500000 | 15.000000 | 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20 |
|-----------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------|
| 40.500000 | 17.000000 | 287.1098632812, 286.5683593750, 286.2949829102, 286.5216674805, 288.0316772461, 291.7698974609, 295.4139709473, 296.8489685059, 295.4132995605 |
|-----------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------|
| 40.500000 | 19.000000 | 287.4010009766, 286.7818298340, 286.4914245605, 286.7260742188, 288.3006286621, 292.1842346191, 296.0237731934, 297.2694702148, 295.9751892090 |
|-----------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------|
| 41.500000 | 15.000000 | 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20 |
|-----------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------|
| 41.500000 | 17.000000 | 286.5835876465, 286.0175781250, 285.7146911621, 285.9142761230, 287.4476623535, 291.1032104492, 294.7090454102, 296.0852355957, 294.7053222656 |
|-----------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------|
| 41.500000 | 19.000000 | 286.9717712402, 286.3946838379, 286.0617675781, 286.1446228027, 287.6101989746, 291.2955017090, 295.2700195312, 296.5146179199, 295.3194274902 |
+===========+===========+=============================================================================================================================================+

Summary
-------
Selected 9 rows out of 9
```

“metadata” operators

```
[37..4416] >> oph_cubeio
[Request]:
operator=oph_cubeio;sessionid=http://127.0.0.1/ophidia/sessions/37438378083214166664146373728392441 6/experiment;exec_mode=sync;ncores=1;cube=http://127.0.0.1/ophidia/35/74;cwd=/
;

[JobID]:
http://127.0.0.1/ophidia/sessions/37438378083214166664146373728392441 6/experiment?82#176

[Response]:
Cube Provenance
---------------

+-----------------------------------------------------------------+---------------+-------------------------------+--------------------------+
| INPUT CUBE                                                      | OPERATION     | OUTPUT CUBE                   | SOURCE                   |
+=================================================================+===============+===============================+==========================+
|                                                                 | ROOT          | http://127.0.0.1/ophidia/35/66 | /repo/tos_O1_2002-2003.nc |
|-----------------------------------------------------------------|---------------|-------------------------------|--------------------------|
|                                                                 | ROOT          | http://127.0.0.1/ophidia/35/67 | /repo/tos_O1_2001-2002.nc |
|-----------------------------------------------------------------|---------------|-------------------------------|--------------------------|
| http://127.0.0.1/ophidia/35/66 - http://127.0.0.1/ophidia/35/67 | oph_intercube | http://127.0.0.1/ophidia/35/70 |                          |
|-----------------------------------------------------------------|---------------|-------------------------------|--------------------------|
| http://127.0.0.1/ophidia/35/70                                  | oph_reduce    | http://127.0.0.1/ophidia/35/71 |                          |
|-----------------------------------------------------------------|---------------|-------------------------------|--------------------------|
| http://127.0.0.1/ophidia/35/71                                  | oph_merge     | http://127.0.0.1/ophidia/35/72 |                          |
|-----------------------------------------------------------------|---------------|-------------------------------|--------------------------|
| http://127.0.0.1/ophidia/35/72                                  | oph_aggregate | http://127.0.0.1/ophidia/35/74 |                          |
+=================================================================+===============+===============================+==========================+

Cube Provenance Graph
---------------------

Directed Graph DOT string :
digraph DG {
        node    [shape=box]

        0       [label="PID : http://127.0.0.1/ophidia/35/74\n"]
        1       [label="PID : http://127.0.0.1/ophidia/35/72\n"]
        2       [label="PID : http://127.0.0.1/ophidia/35/71\n"]
        3       [label="PID : http://127.0.0.1/ophidia/35/70\n"]
        4       [label="PID : http://127.0.0.1/ophidia/35/66\nSOURCE : /repo/tos_O1_2002-2003.nc\n"]
        5       [label="PID : http://127.0.0.1/ophidia/35/67\nSOURCE : /repo/tos_O1_2001-2002.nc\n"]

        1->0    [label="oph_aggregate"]
        2->1    [label="oph_merge"]
```



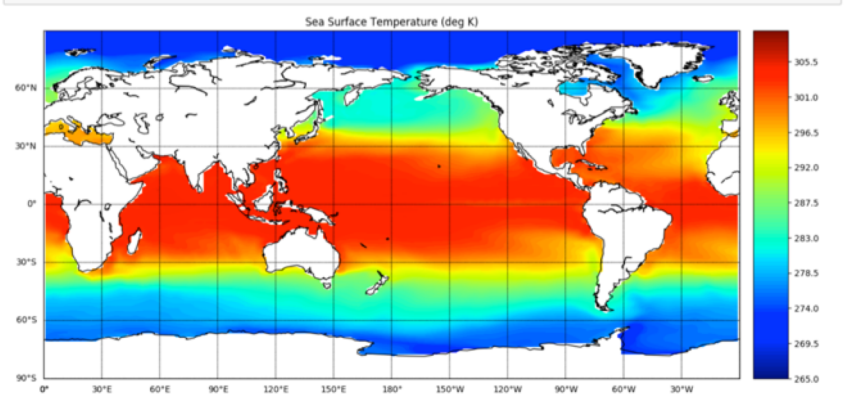
Programmatic support for data science applications

PyOphidia is a Python module to interact with the Ophidia framework.

It provides a programmatic access to Ophidia features, allowing:

- *Submission of commands to the Ophidia Server and retrieval of the results*
- *Management of (remote) data objects in the form of datacubes*
- *Easy exploitation from Jupyter Notebooks and integration with other Python modules*

```
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')
```



Export result to NetCDF file

```
]: mycube3.exportnc2(output_path='/home/' + cube.Cube.client.username,export_metadata='yes')
```

https://pypi.org/project/PyOphidia/
https://anaconda.org/conda-forge/pyophidia

Python and HPC infrastructure transparency

PyOphidia class hides the HPC environment complexity

```
In [ ]: from PyOphidia import cube, client
        cube.Cube.setclient(read_env=True)

In [ ]: cube.Cube.cluster(action='deploy',host_partition='test_partition',nhost=4)

In [ ]: myCube = cube.Cube(src_path='/work/ophidia/tests/tasmax_day_CMCC-CESM_rcp85.nc',
            measure='tasmax', import_metadata='yes', imp_dim='time', description='Max Temps',
            nfrag=16, nhosts=4,
            host_partition='test2',
            ncores=2, nthreads=8
            )

]: myCube2 = maxtemp.apply(
            query="oph_predicate('oph_float','oph_int',measure,'x-298.15','>0','1','0')",
            ncores=2, nthreads=8
        )

In [ ]: myCube3 = myCube2.subset(subset_filter=1, subset_dims='time

In [ ]: pythonData = myCube3.export_array(show_time='yes')

In [ ]: print(pythonData)

In [ ]: cube.Cube.cluster(action='undeploy',host_partition='test_partition')
```

Dynamic I/O & Analytics nodes allocation

Data partitioning and distribution

Framework operator parallelism

Ophidia-notebook data translation and transfer

I/O & Analytics nodes undeployment

ENES Climate Analytics Service (ECAS)

- *ECAS* is one of the *EOSC-Hub Thematic Services as well as one of IS-ENES3 compute services*
- ECAS builds on top of the *Ophidia HPDA framework,* integrated with components from INDIGO-DataCloud, EUDAT and EGI
- Integrates computing resources with datasets and data analytics tools
- It provides a user-friendly environment based on Jupyter Notebooks and well-known Python modules

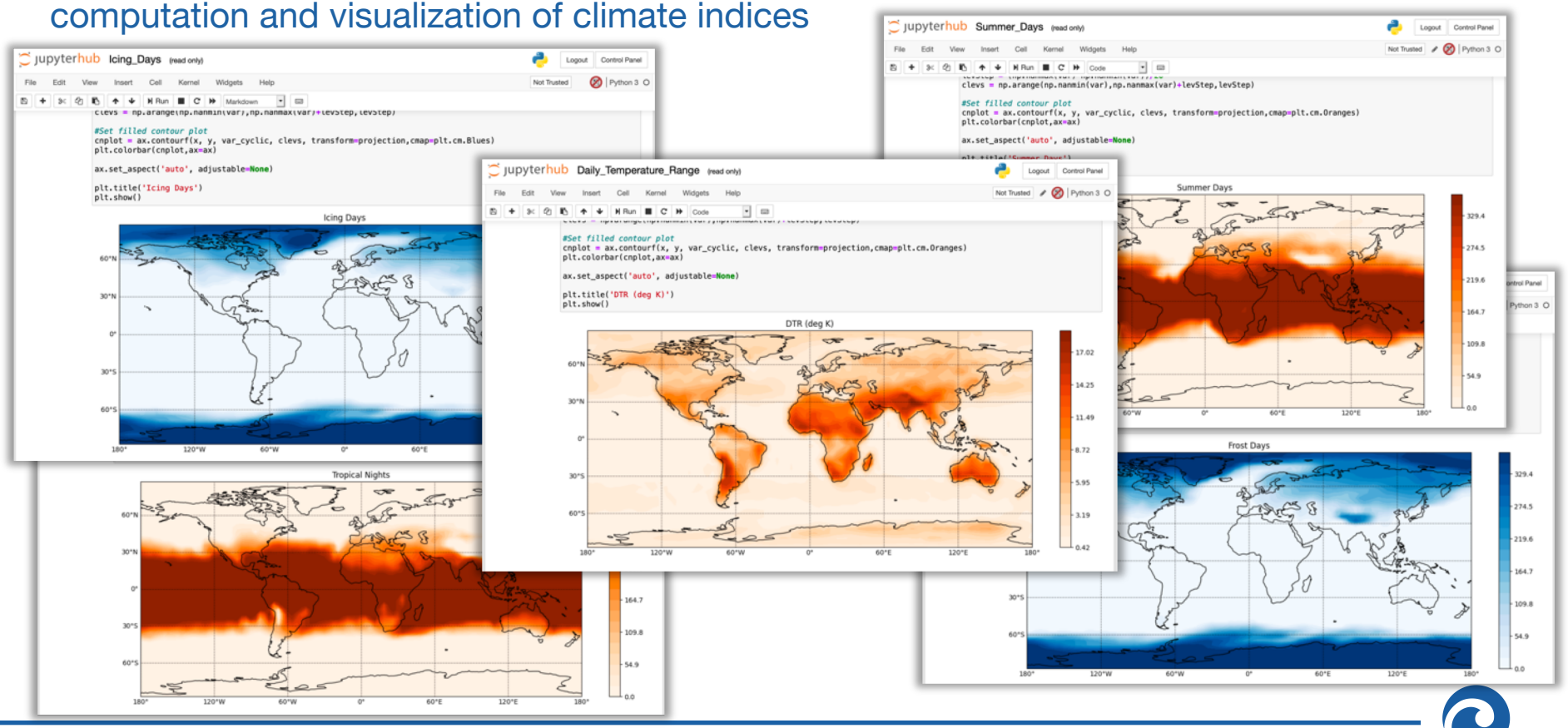
ECAS @ CMCC https://ecaslab.cmcc.it

EOSC Portal Catalogue: https://marketplace.eosc-portal.eu/services/enes-climate-analytics-service

Climate Analytics Service: https://portal.enes.org/data/data-metadata-service/climate-analytics-service

Computation of climate indices

PyOphidia can be combined with other Python libraries (e.g., cartopy/matplotlib) for the computation and visualization of climate indices



Ophidia in ESiWACE2 project

ESiWACE2 – Excellence in Simulation of Weather and Climate in Europe, Phase 2

(1) Enabling leading European weather and climate models to leverage the available performance of pre-exascale systems with regard to both compute and data capacity in 2021.

(2) Preparing the weather and climate community to be able to make use of exascale systems when they become available.

Ophidia is one of the applications considered in the frame of **ESiWACE2** (WP4, WP5):

- Integration with **Earth System Data Middleware** (**ESDM)** (https://github.com/ESiWACE/esdm) for I/O over heterogeneous storage systems
- One of the applications (HPDA) targeted by the **ESDM-PAV**
  - Extensions for **in-flight analytics** are being developed
  - Some HPDA **scientific use cases** defined (preliminary implementation)

Ophidia

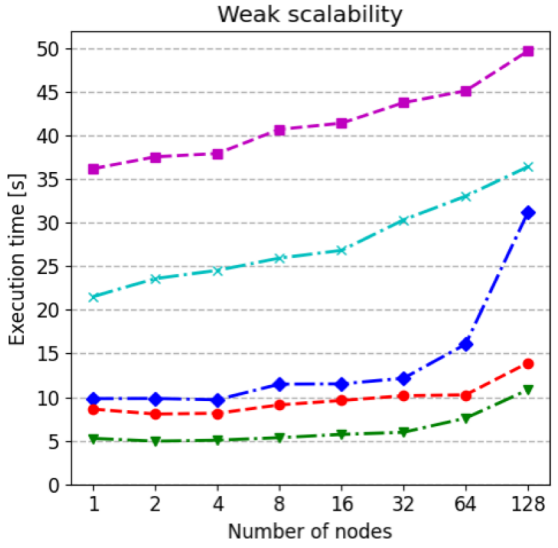
esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

Ophidia HPDA framework benchmark

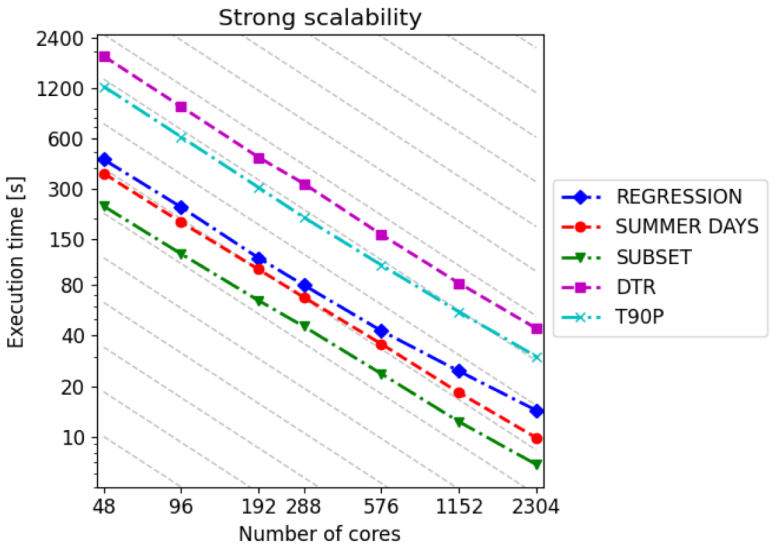
Goal: *benchmarking, tuning and optimisation over a large-scale HPC machine of the Ophidia HPDA framework*

Evaluate the scalability of Ophidia analytics kernels on a few thousands of cores:

- various **strong** and **weak scalability** tests performed
- **good scalability** in the majority of cases until **3k cores**



Data size per node: 67GiB



Data size fixed: 3.2TiB

We acknowledge PRACE for awarding access to MareNostrum 4 at Barcelona Supercomputing Center (BSC), Spain and the support provided by BSC (PRACE resources for CoE, in the context of ESiWACE).

D. Elia, S. Fiore and G. Aloisio, "Towards HPC and Big Data Analytics Convergence: Design and Experimental Evaluation of a HPDA Framework for eScience at Scale," in IEEE Access (Early Access), https://ieeexplore.ieee.org/document/9428012

Summary and future activities

Recap

- Ophidia HPDA framework address climate data challenges at scale
  - Tailored for HPC infrastructures
  - Provides (user-friendly) features for building complex analysis
  - Experimental results show scalability up to few thousand cores

Future activities

- Further benchmarks of the HPDA framework with large-scale applications/workflows
- Address portability with HPC-enabled containerization
- Full integration of Ophidia with the ESDM/ESDM-PAV (ESiWACE2 activities)

Thank you for the attention!

ESiWACE2 *has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823988*

IS-ENES3 *has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 824084*

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Ophidia website: *http://ophidia.cmcc.it*

Contact: *donatello.elia [at] cmcc.it*

cmcc
Centro Euro-Mediterraneo sui Cambiamenti Climatici