



EGI-ACE HPC Integration Workshop

ENES pilot

Fabrizio Antonio

Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC)

Dissemination level: Public

Disclosing Party: CMCC

Recipient Party: EGI-ACE Project



EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.

ENES & IS-ENES

European Network for Earth System Modelling

A network of European groups in
climate/Earth system modelling
Launched in 2001 (MOU)

Ca 50 groups from academic, public
and industrial world

Main focus :
discuss strategy
**to accelerate progress in climate/
Earth system modelling and
understanding**



<http://enes.org/>



<http://is.enes.org/>

IS-ENES infrastructure projects

IS-ENES (2009-2013)

IS-ENES2 (2013-2017)

IS-ENES3 (2019-2022)

**Support WCRP internationally
coordinated climate model
experiments
(CMIP & CORDEX)**

**Support sharing of expertise
on**

climate models, tools & HPC
<https://is.enes.org>

ESGF and the CMIP data archive



IS-ENES provides the EU contribution to the ESGF

Climate analysis

Main challenges & issues

*Several **key challenges** and practical **issues** related to large-scale climate analysis*

- Input data from multiple models needed
- **Data download is a big barrier** for climate scientists
 - download can take a significant amount of time
 - network instability, dropped connections, etc.
- **Data analysis** mainly performed using **client-side & sequential approaches**
- **Tens/hundreds of analytics operators** required to analyse large datasets
- Installation and update of **data analysis tools** and **libraries** needed
- **Strong requirements** in terms of **computational** and **storage resources**

ENES Data Space

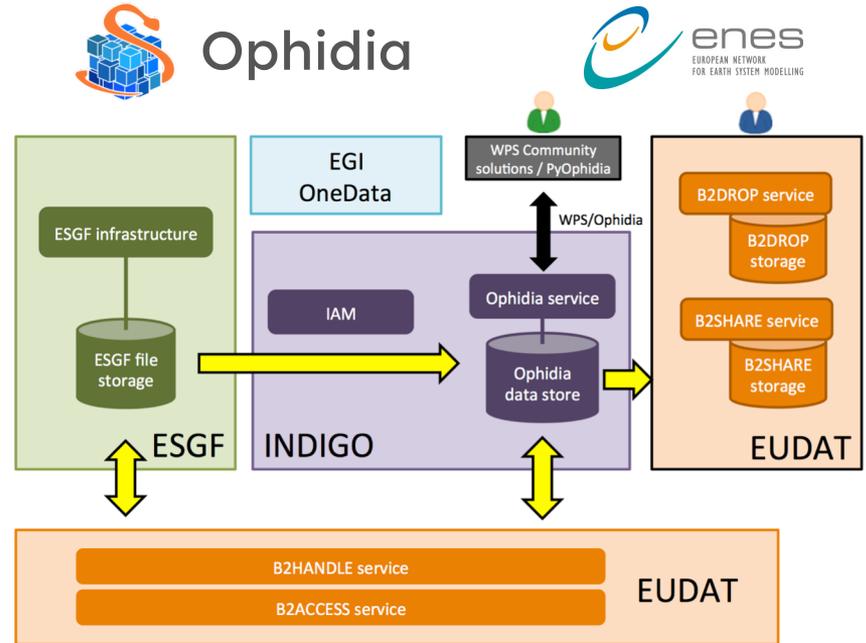
High-level architecture

The **ENES data space** will consist of three main layers:

- ❑ A **GUI** providing access to an Open (Data) Science environment for climate data analysis
 - **JupyterHub**
- ❑ **EOSC compute services** → to deploy the **ECAS** environment
- ❑ **Data collection** layer providing access to the **ESGF** data archive

ECAS: A bridge between ENES and EOSC

- ❑ The **ENES Climate Analytics Service (ECAS)** supports climate data analysis
- ❑ It is one of the **EOSC-Hub Thematic Services**
- ❑ ECAS builds on top of the **Ophidia big data analytics framework** with components from EGI, INDIGO-DataCloud and EUDAT
- ❑ ECAS provides a **multi-model data science environment** for CMIP-based analytics experiments in ESGF



ECAS is a Compute Service solution supported also by IS-ENES

Ophidia High-Performance Data Analytics Framework

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

- **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, HTC applications and large **workflows** on scientific datacubes, primarily in climate domain



Ophidia

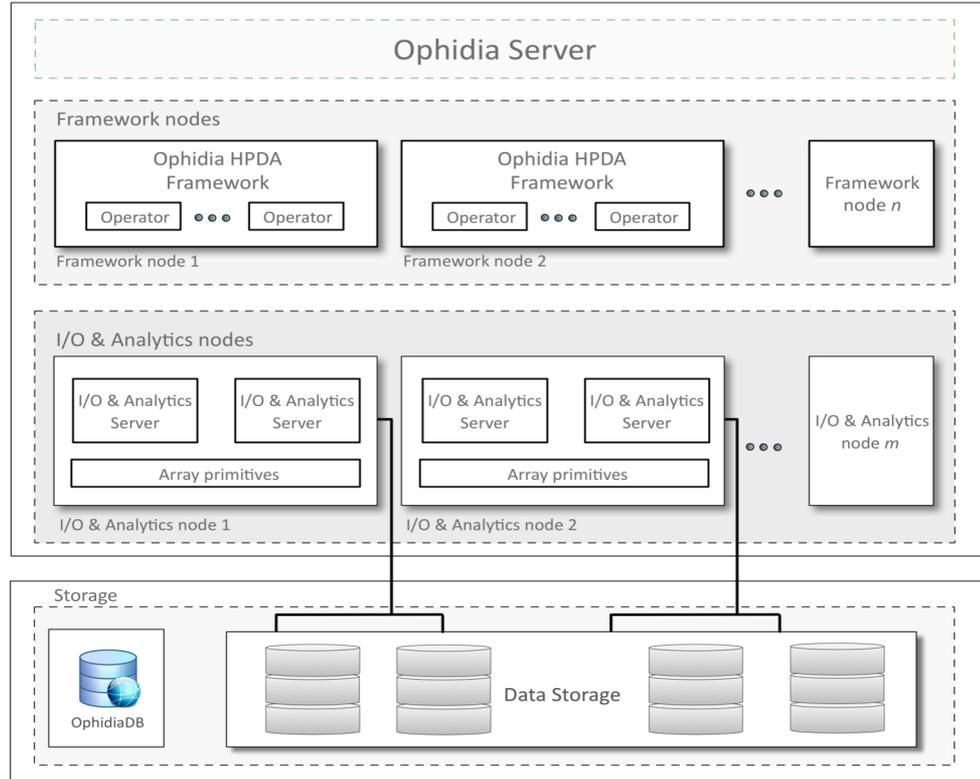
Ophidia multi-layered architecture

Modular and extensible software stack

The **Ophidia framework** dynamically executed in parallel on multiple nodes (multi-proc/multi-thread)

I/O & analytics servers can be deployed on-demand on multiple nodes based on requirements

Data partitioned in a hierarchical fashion and distributed over the storage



ENES pilot

Main objective and features

- ❑ **Goal:** provide a **container-based** Data Science environment for **data analytics** and **visualization** on top of a **HPC infrastructure**
- ❑ **Containerization** of the ECAS core components (i.e. **Ophidia**) to address a better **portability** of the platform on different infrastructures
- ❑ **Docker-based** deployment
 - JupyterHub, Dask, Python Data Science libraries, Onedata, Grafana
- ❑ HPC-oriented container technologies: **uDocker**
- ❑ Strong correlation between **T7.3** (HPC integration) and **T5.2** (ENES Data Space)

On-demand instantiation of an Ophidia cluster

Target environment: HPC cluster

Deployment of I/O & analytics servers

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

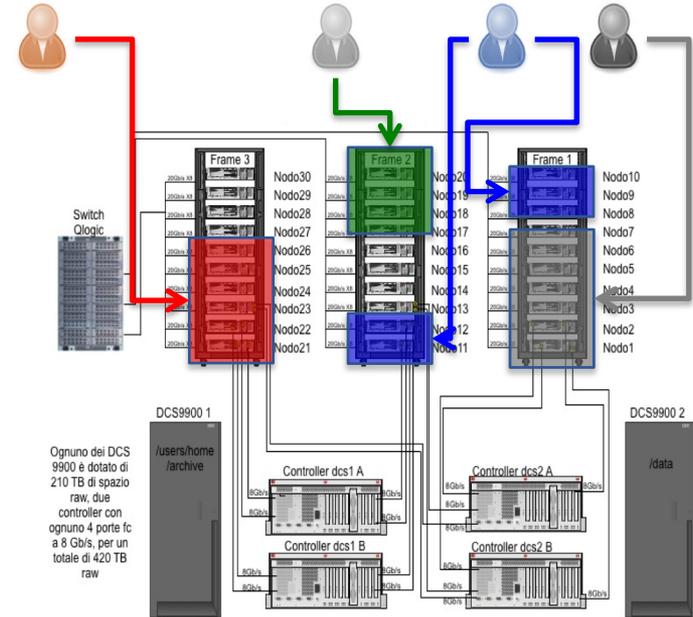
```
oph_cluster
action=undeploy;cluster_name=new;
```

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



Python and HPC infrastructure transparency

```
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
)
```

Dynamic I/O & Analytics
nodes allocation

```
In [ ]: 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')
```

I/O & Analytics nodes
undeployment

Links and references

- Ophidia website** <http://ophidia.cmcc.it>
- EGI-ACE** <https://www.egi.eu/projects/egi-ace/>
- EOSC-Hub** <https://www.eosc-hub.eu/>
- IS-ENES** <https://is.enes.org/>

These activities are supported in part by EGI-ACE and IS-ENES3 projects:

- *EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101017567*
- *IS-ENES3 is a project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 824084*



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





Thank you!

Contact: egi-ace-po@mailman.egi.eu

Website: www.egi.eu/projects/egi-ace



[EGI Foundation](#)



[@EGI_eInfra](#)



EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.