Towards HPC and Big Data convergence for climate analysis at scale

7th ENES HPC Workshop
BSC, Barcelona
May 9-11, 2022

Donatello Elia

Advanced Scientific Computing Division,
Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC),
Lecce, Italy

cmcc
Centro Euro-Mediterraneo sui Cambiamenti Climatici

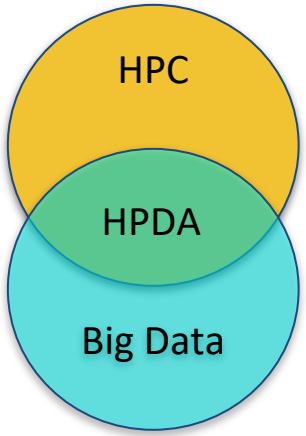
Outline

- ***Introduction to HPDA and data challenges in eScience***
- ***Overview of the Ophidia HPDA framework***
- ***Integration with HPC infrastructures***
- ***Support for Python applications***
- ***Extensions for I/O in the frame of ESiWACE***
- ***End-to-end ESM workflows***

The work shown in these slides include the effort from multiple people working on different projects at CMCC:
D'Anca, C. Palazzo, F. Antonio, S. Scardigno, M. Musio, G. Accarino, A.Giannotta,
V. Aloisi, F. Immorlano, E. Scoccimarro, D. Peano, S. Fiore and G. Alosio

Convergence of HPC and Big Data Analytics for HPDA

- **Exponential increase** in **data volumes** and **complexities** is causing a radical change in the scientific discovery process in several domains
- **High Performance Data Analytics** (HPDA) software solutions can effectively support **climate data analysis at scale** through HPC resources
- **Convergence** of **HPC** and **big data analytics** is a key factor for future scientific research and for enabling **HPDA** applications at **extreme-scale**
- Big Data and HPC software ecosystems have been developed mostly independently



- *Significant gaps in how the two ecosystems are designed (e.g., in terms of computing, networking and storage solutions, programming models, etc.)*
- *Several challenges must be addressed to **support HPDA at scale***

Challenges for HPDA in eScience

- **High-level interfaces** (Python-based) and easy-to-use environments to support scientists productivity (Jupyter)
  - *Abstracting from the data and infrastructure complexity and promoting scientist collaboration*

- **Reduce data movement** shifting the analysis closer to the data
  - *Move the analysis from the client-side to the server-side*
  - *In-flight, in-transit and in-situ analysis techniques*

- The complexity of the analysis leads to the need for **analytics workflow support**
  - *Able to manage hundreds of operators and parallelize their execution on large-scale resources*

- Easier **portability** and **deployment** of the HPDA software on HPC infrastructures
  - *Transparent use of scientific HPDA applications/workflows across different resources (HPC/Cloud)*

Ophidia HPDA framework

Ophidia (http://ophidia.cmcc.it) is a CMCC Foundation research project addressing data challenges for eScience, with a focus on climate science

- 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
- Support for **interactive analysis, complex experiments** and **workflows** on scientific data

Ophidia

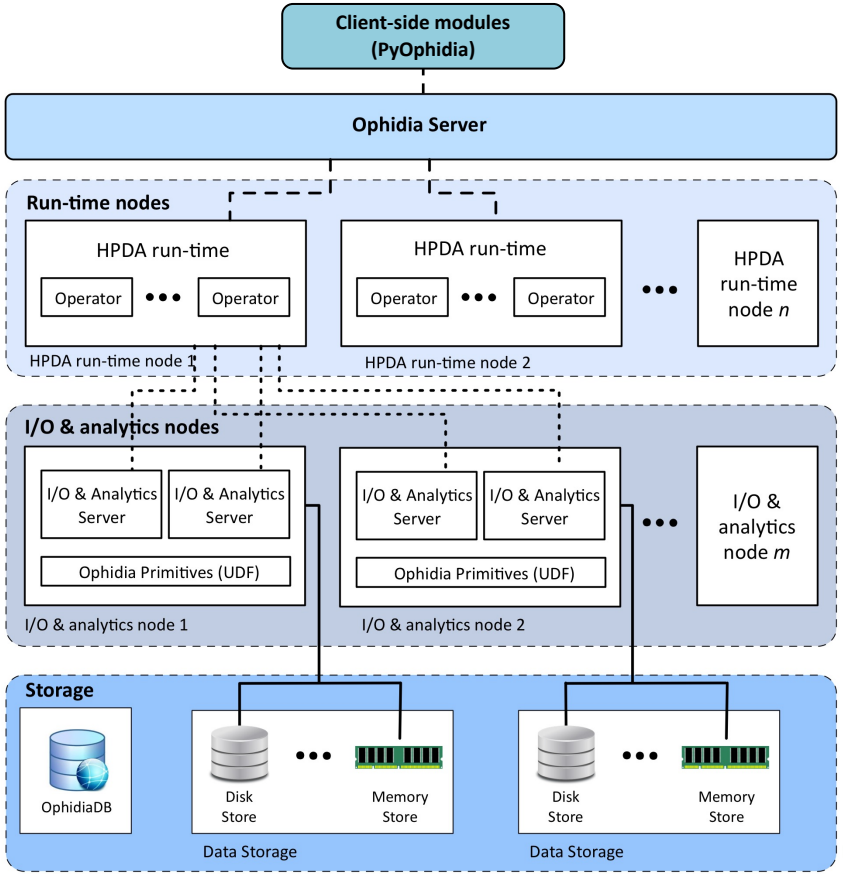
cmcc
Centro Euro-Mediterraneo
sui Cambiamenti Climatici

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

Ophidia architecture

The framework has been enhanced to support **large-scale HPDA use cases**:

- **Modular, extensible and scalable** software stack
- **User-friendly** Python interface (PyOphidia)
- **HPDA runtime** for executing parallel data operators
- Support for **in-memory analytics**
- Data partitioned in binary arrays and distributed across the I/O & analytics nodes using a **key-value approach**

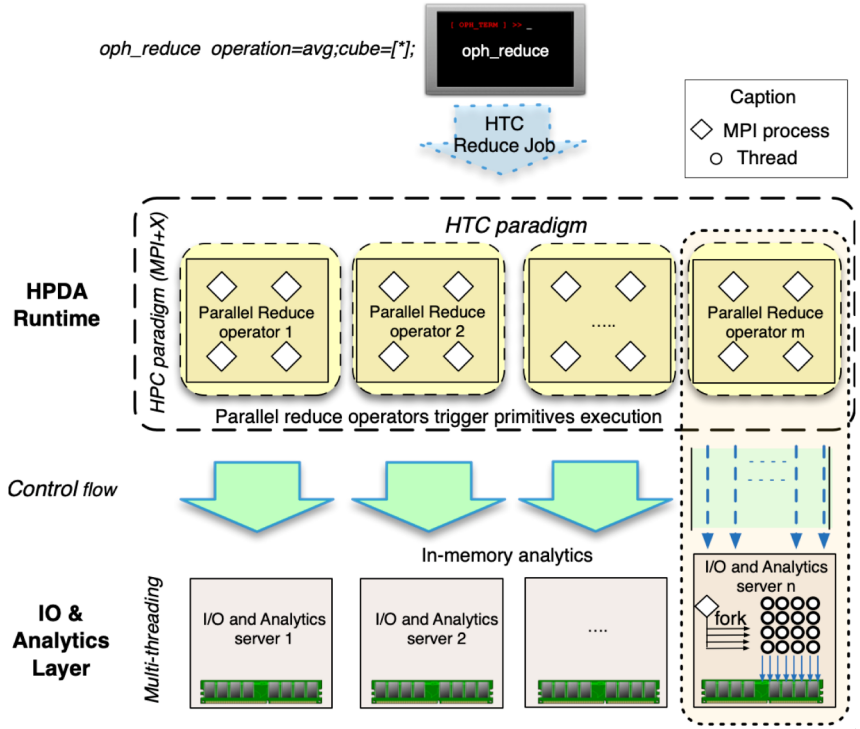


A parallel runtime for HPDA

Hierarchical parallel execution model for data analytics functions, with two levels of parallelism:

- **Datacube-level**: execute multiple operators on different input data (HTC paradigm)
- **Fragment-level**: MPI+X model for execution of single analytics operators on a datacube (HPC paradigm)
  - **Multi-thread** for intra-node parallelism
  - **MPI** to scale processing on multiple nodes

The analytics function on the data fragments are managed and executed by the I/O servers.

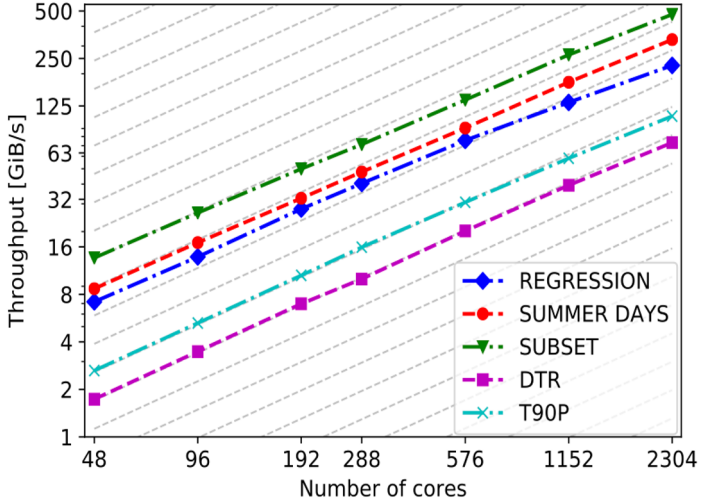


D. Elia, S. Fiore and G. Aloisio, "Towards HPC and Big Data Analytics Convergence: Design and Experimental Evaluation of a HPDA Framework for eScience at Scale," in IEEE Access, vol. 9, pp. 73307-73326, 2021

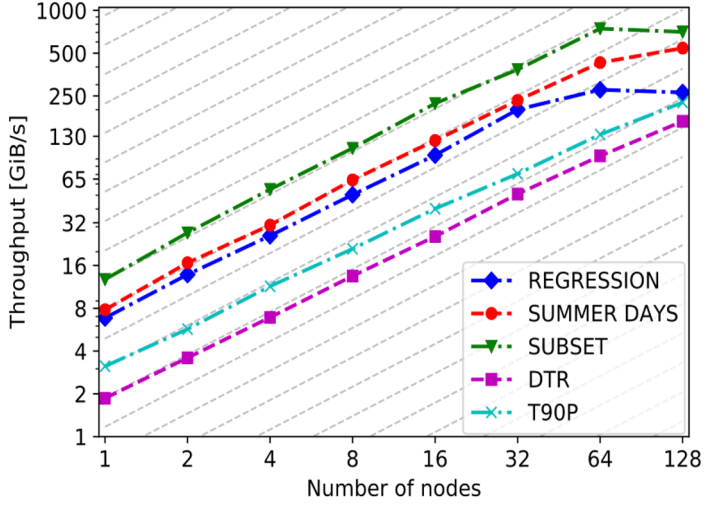
Ophidia HPDA framework benchmark

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

- **strong** and **weak scalability** tests
- different **representative operators** tested
- Good **scalability** in most cases **until ~3k cores**
- Other evaluations planned in future



Strong Scalability: size fixed to 3.2TiB



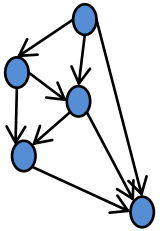
Weak Scalability: 67GiB/node
From 67GiB to 8.4TiB

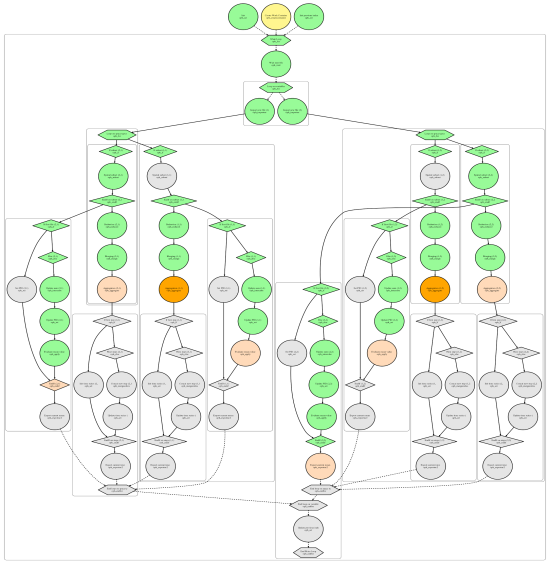
Benchmark performed using the core hours awarded by PRACE (Call 18), in the context of the ESiWACE CoE, on MareNostrum 4 at the Barcelona Supercomputing Center (BSC), Spain

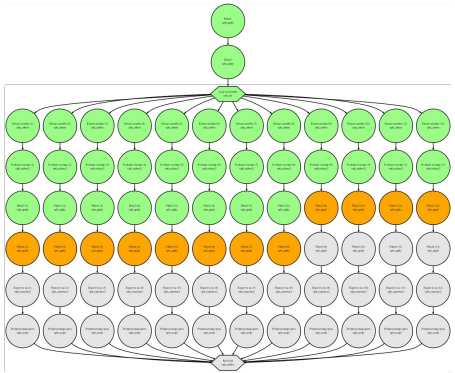
Analytics workflows

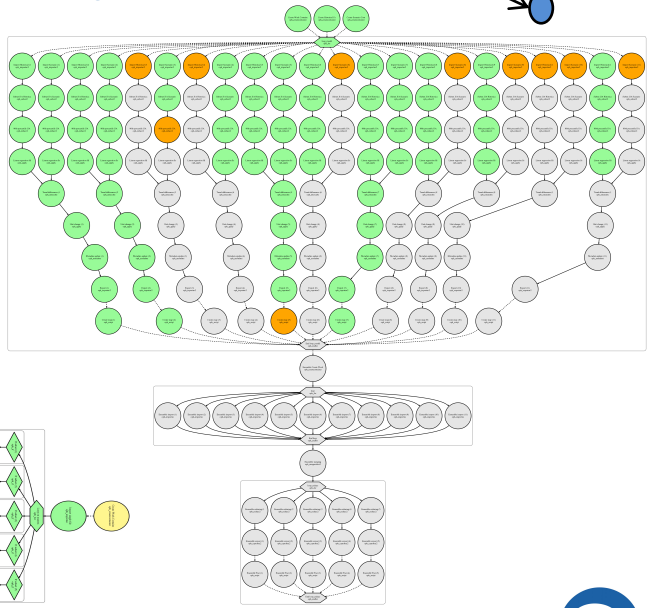
Ophidia supports the execution of complex workflows of analytics operators.

- Defines a **JSON representation** for the workflow DAG specification
- Supports different constructs: *dependencies; massive tasks; iterative (group of) tasks; parallel (group of) tasks; flow and error control*









C. Palazzo, A. Mariello, S. Fiore, A. D'Anca, D. Elia, D. N. Williams, G. Aloisio, "A Workflow-Enabled Big Data Analytics Software Stack for eScience", HPCS 2015, pp. 545-552

On-demand deployment on HPC infrastructures

Target environment: *HPC cluster*

On-demand execution of I/O & analytics servers

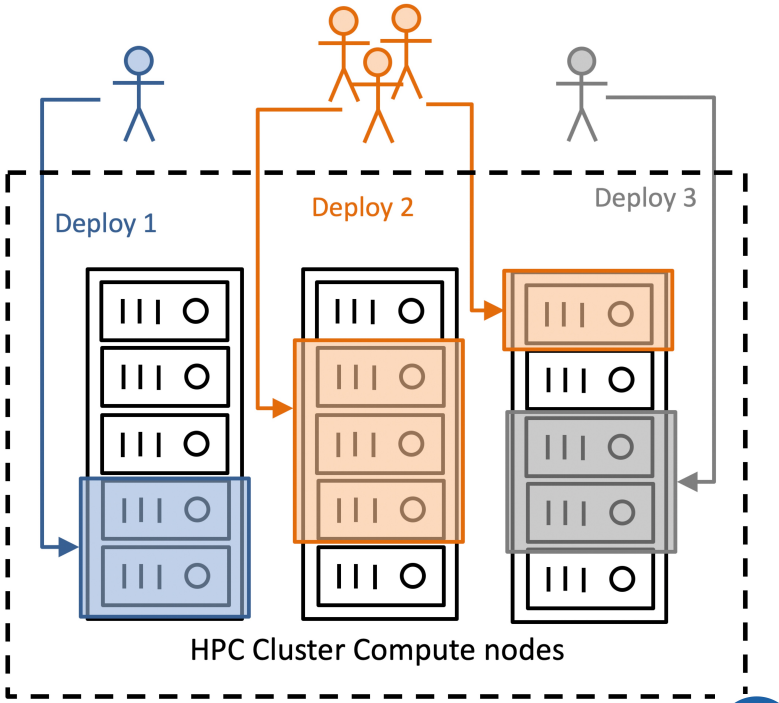
- `oph_cluster action=deploy;nhost=64;cluster_name=new;`
- `oph_cluster action=undeploy;cluster_name=new;`

Horizontal scalability of the framework components

Transparent interaction with batch scheduling systems

Zeus SuperComputer
348 nodes, 1.2 PetaFLOPS

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



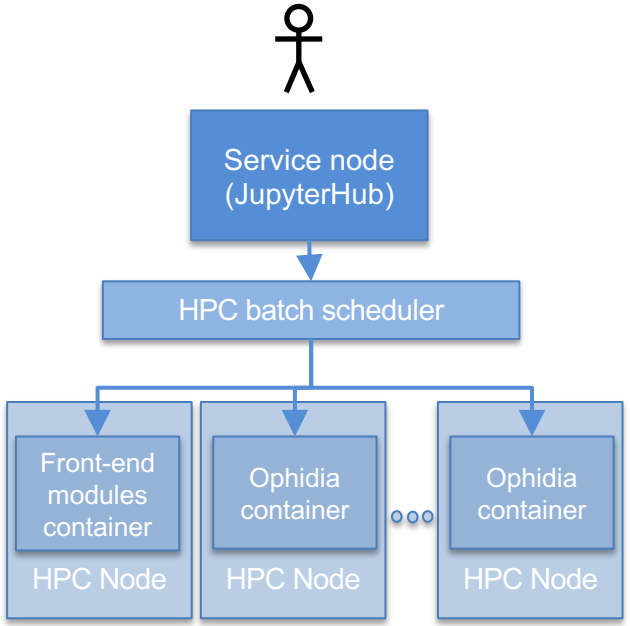
Portability on different infrastructures

Use of containers as a layer to enhance HPDA applications portability and usage across different infrastructures:

- **easy** and **transparent** portability and deployment of Ophidia on **HPC** (Singularity and udocker) and **Cloud** (Docker);
- ready-to-use **integration** of the framework **computing** components and **high-level** software (PyOphidia, Jupyter, visualization libraries) for user **productivity**;

Leading to novel service models: **HPDA as a Service**

- PoC carried out in the context of the HPC ENES Pilot in EGI-ACE



EGI-ACE *receives funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement number 101017567.*

High-level interface for data science applications

PyOphidia is a Python module to interact with the Ophidia framework

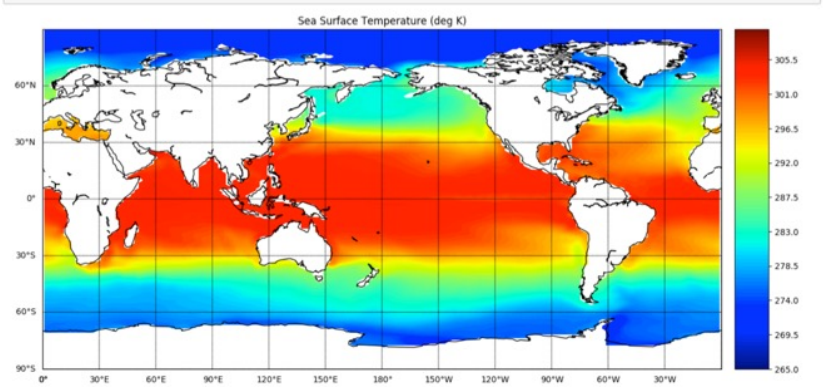
High-level and easy-to-use bindings for the HPDA framework:

- *Provides APIs to manage deployment, data distribution and computation parallelism*
- *Management of (remote) data objects in the form of datacubes*
- *Easy exploitation from Jupyter 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')
```

Python and HPC infrastructure transparency

PyOphidia hides the HPC infrastructure 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

Ophidia in ESiWACE2 project

Ophidia is one of the applications considered in the frame of the ESiWACE2 project:

- Addresses the **HPDA use cases** in the context of the Post-processing, Analytics and Visualisation (PAV) applications
- Improvements to support a **wider set** and **larger-scale** ESM datasets
- Integration with **Earth-System Data Middleware** (**ESDM)** for parallel I/O over heterogeneous storage systems (IMPORTESDM/EXPORTESDM operators)
- Extensions for **in-flight analytics** during data loading from the storage system (ESDM)

Ophidia

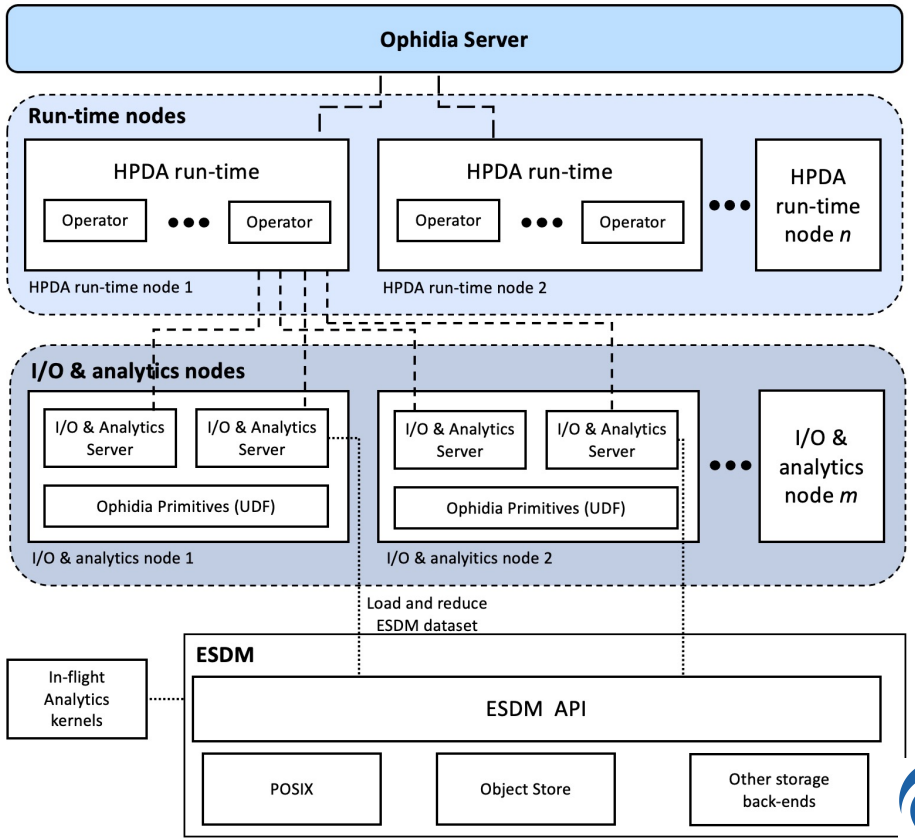
esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

ESiWACE2 *is a project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823988*

PoC Integration of novel I/O solutions

New operators implemented in Ophidia to interact with the ESDM:

- Support for parallel **load/store** operations in-memory from ESDM storage to Ophidia I/O servers (and viceversa)
  - ESDM supports **transparent access** to different storage back-ends
- Several **in-flight analytics kernels** available through the ESDM **read streaming interface** for the Ophidia load operation (statistical reductions, mathematical, subsetting)
  - **Reduce** the amount of **data moved** from the storage to the compute nodes

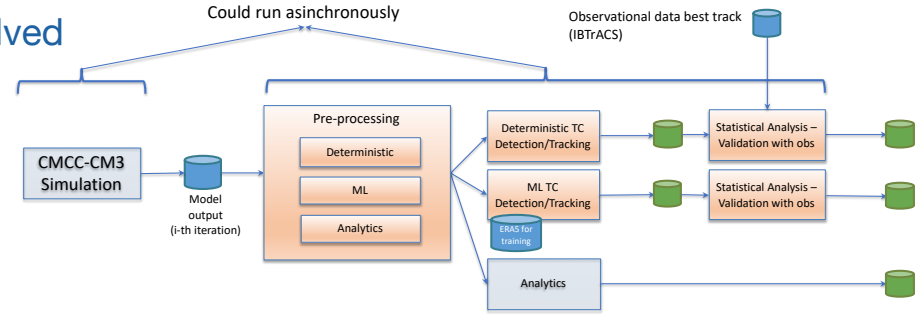


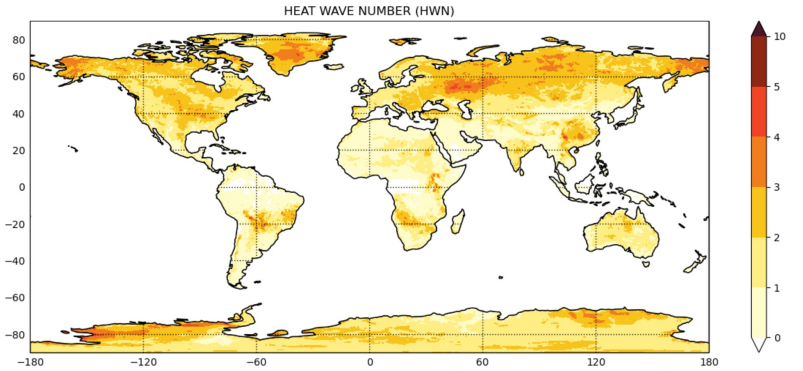
Ophidia for HPDA in end-to-end ESM workflows

In the context of Pillar II in eFlows4HPC CMCC is involved in the development of **a single end-to-end ESM workflow** integrating:

- High-resolution GCM simulation: CMCC-CM3
- HPDA for feature extraction: extreme climate indicators (e.g. Heat Wave index) with **Ophidia**
- ML algorithms: Neural Net for extreme events analysis (Tropical Cyclones detection)

eFlows4HPC aims at delivering a software stack integrating HPC, data analytics and ML frameworks to provide an overall workflow management system.





eFlows4HPC *this project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955558. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Spain, Germany, France, Italy, Poland, Switzerland, Norway*

Conclusions and future activities

Conclusions

- *HPDA applications can be effectively supported thanks to the convergence of HPC and Big Data software ecosystems. Several challenges to be addressed*
- *Parallel execution models, transparent integration with HPC, software portability and high-level Python interfaces exploited in Ophidia to support HPDA applications*
- *Integration with new I/O systems (ESDM) and use of in-flight analytics is promising*

Future activities

- *Testing in-flight on larger data/nodes and potentially with active storage kernels solutions*
- *Integration of the HPDA framework with other formats/storage solutions (Zarr, FDB) and technologies (NVRAM)*
- *Explore in-situ and in-transit analysis approaches closer to the data produced by ESM simulations*

Thank you for the attention!

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

ESiWACE2 *is a project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823988*

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

cmcc
Centro Euro-Mediterraneo sui Cambiamenti Climatici