OphidiaLab: a user-friendly, integrated environment for scientific data analytics @ CMCC

Alessandro D'Anca[1], Donatello Elia[1], Cosimo Palazzo[1], Elisa Distante[2], Sandro Fiore[1], Giovanni Aloisio[1,2]

[1]Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Lecce, Italy, [2]Università del Salento, Lecce, Italy

{alessandro.danca, donatello.elia, cosimo.palazzo, sandro.fiore}@cmcc.it - giovanni.aloisio@unisalento.it

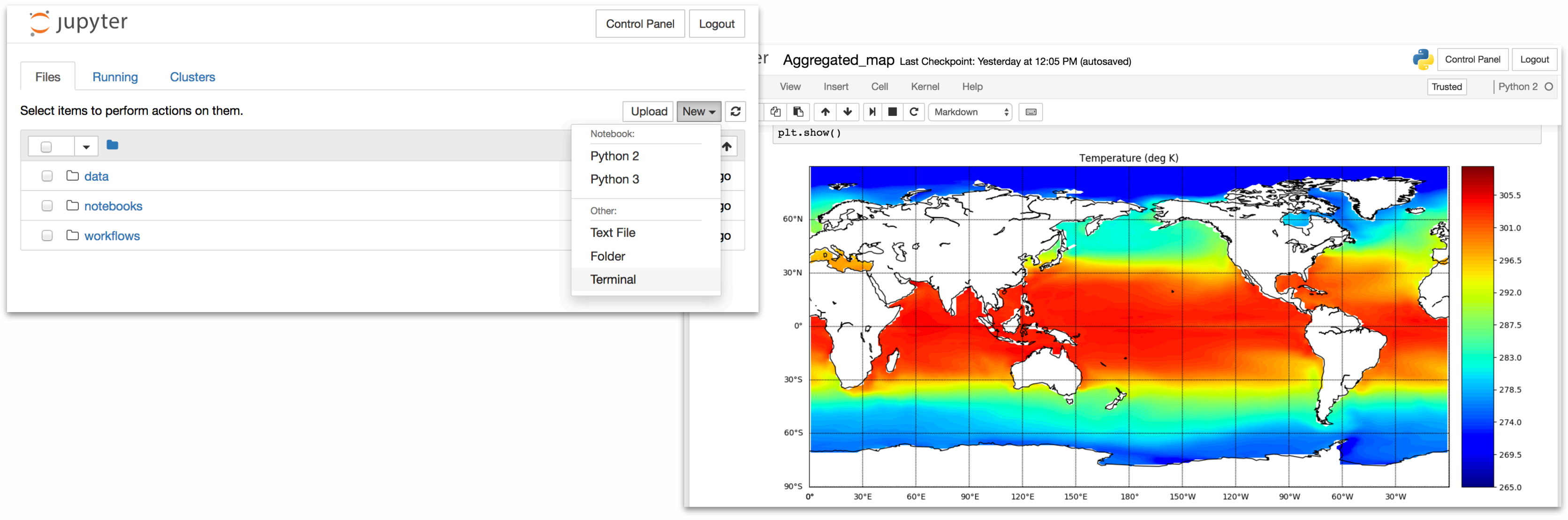
ENVIRONMENT FOR DATA ANALYSIS AND VISUALIZATION

OphidiaLab is a user-friendly scientific data analysis environment deployed at CMCC SuperComputing Center integrating data and analysis tools to support scientists in their daily research activities. The environment joins the features of the *Ophidia data analytics framework* [1] with a large set of Python libraries for running data manipulation, analysis, and visualization.

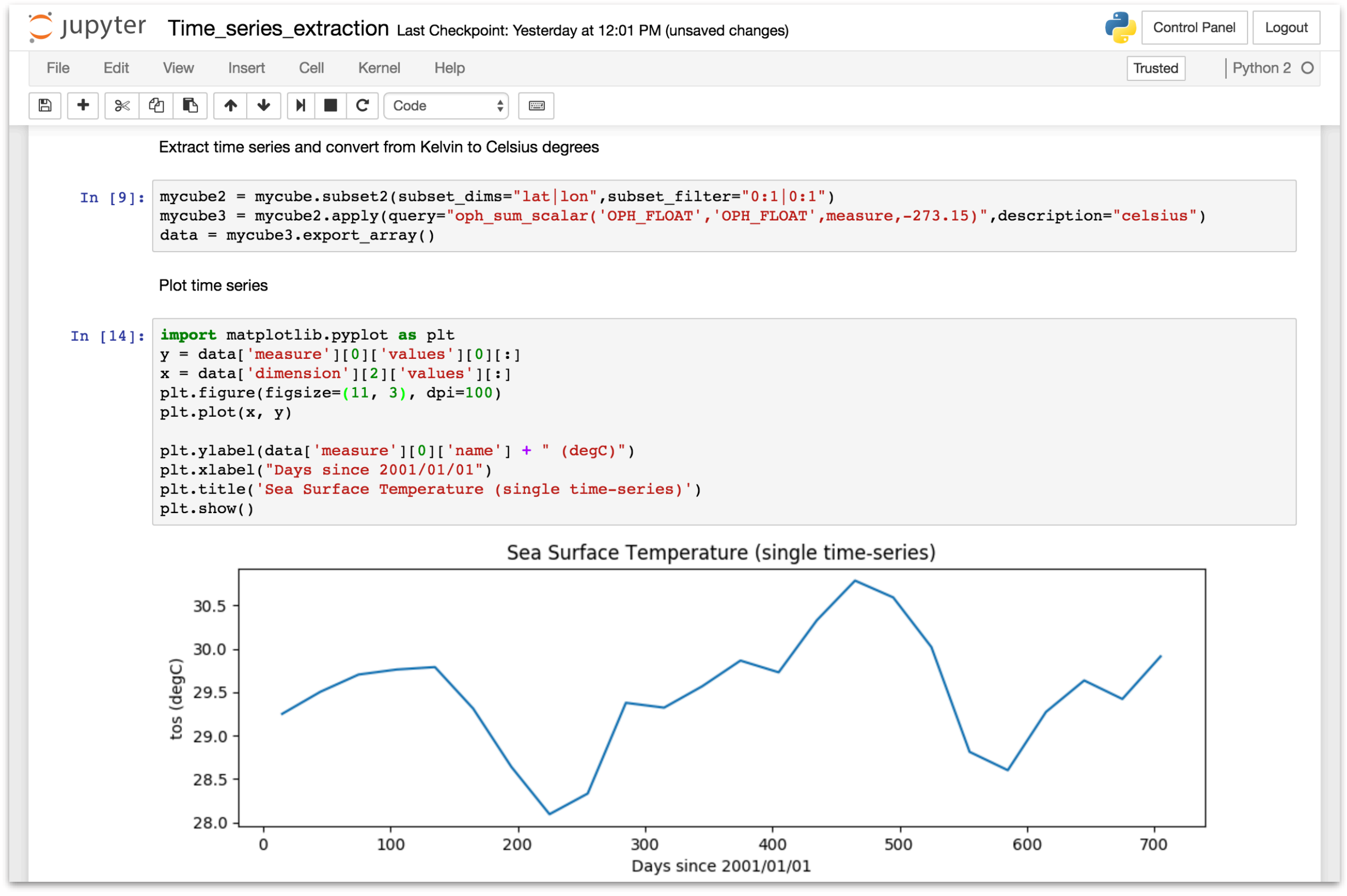
The OphidiaLab environment integrates the following services:

- a **JupyterHub** instance providing the user with a web-based system for creating, executing and sharing **Jupyter notebooks** (Python-based) supporting live-coding and visualization;
- a multi-node **Ophidia framework** instance with WPS-enabled interface accessible through the Ophidia Terminal and any WPS-compliant client;
- a GUI for interactive workflow composition - the **Ophidia Experiment Editor**;
- a monitoring system based on **Grafana**.

Moreover, the environment provides access to a set of datasets (also through a **THREDDS Data Server**), a number of example Jupyter notebooks and real-world workflows describing indicators from several use cases. Experiment output can be exported in the user space or on the publicly accessible data server, whereas JupyterHub provides the features to update files and navigate the file system.



The features of the Ophidia framework can be directly exploited in the notebooks to run data analytics tasks on big datasets and plot the results on charts and maps using well-known Python libraries in a user-friendly interface. **PyOphidia** - the Ophidia Python bindings - allows to easily interact with Ophidia and other Python-based modules (e.g. Matplotlib, NumPy, etc.).


jupyter
Time_series_extraction Last Checkpoint: Yesterday at 12:01 PM (unsaved changes)
Control Panel
Logout
File
Edit
View
Insert
Cell
Kernel
Help
Trusted
Python 2
Code

Extract time series and convert from Kelvin to Celsius degrees

```
In [9]: mycube2 = mycube.subset2(subset_dims="lat|lon",subset_filter="0:1|0:1")
        mycube3 = mycube2.apply(query="oph_sum_scalar('OPH_FLOAT','OPH_FLOAT',measure,-273.15)",description="celsius")
        data = mycube3.export_array()
```

Plot time series

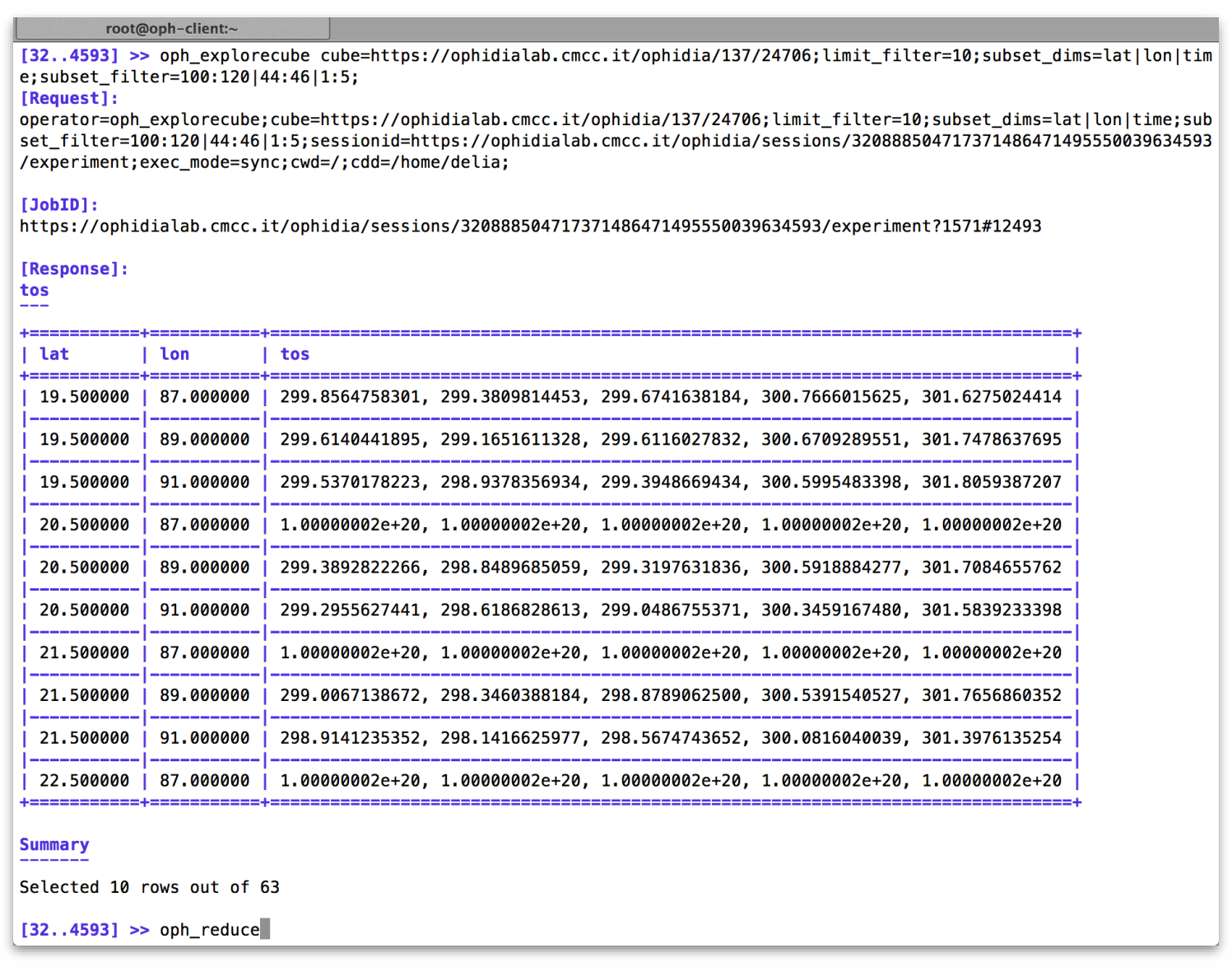
```
In [14]: import matplotlib.pyplot as plt
         y = data['measure'][0]['values'][0][:]
         x = data['dimension'][2]['values'][:]
         plt.figure(figsize=(11, 3), dpi=100)
         plt.plot(x, y)

         plt.ylabel(data['measure'][0]['name'] + " (degC)")
         plt.xlabel("Days since 2001/01/01")
         plt.title('Sea Surface Temperature (single time-series)')
         plt.show()
```

OPHIDIA BIG DATA ANALYTICS FRAMEWORK

Ophidia represents a complete software stack developed by CMCC for data analytics in multiple eScience domains, such as climate change, astrophysics, etc.

In terms of end-user features, the framework provides, among the others:

- *data reduction and subsetting;*
- *data intercomparison;*
- *metadata and provenance management;*
- *time series analysis with a wide set of array-based primitives (around 100);*
- *interactive data analysis;*
- *workflows of tasks;*
- *rapid setup of operational chains.*

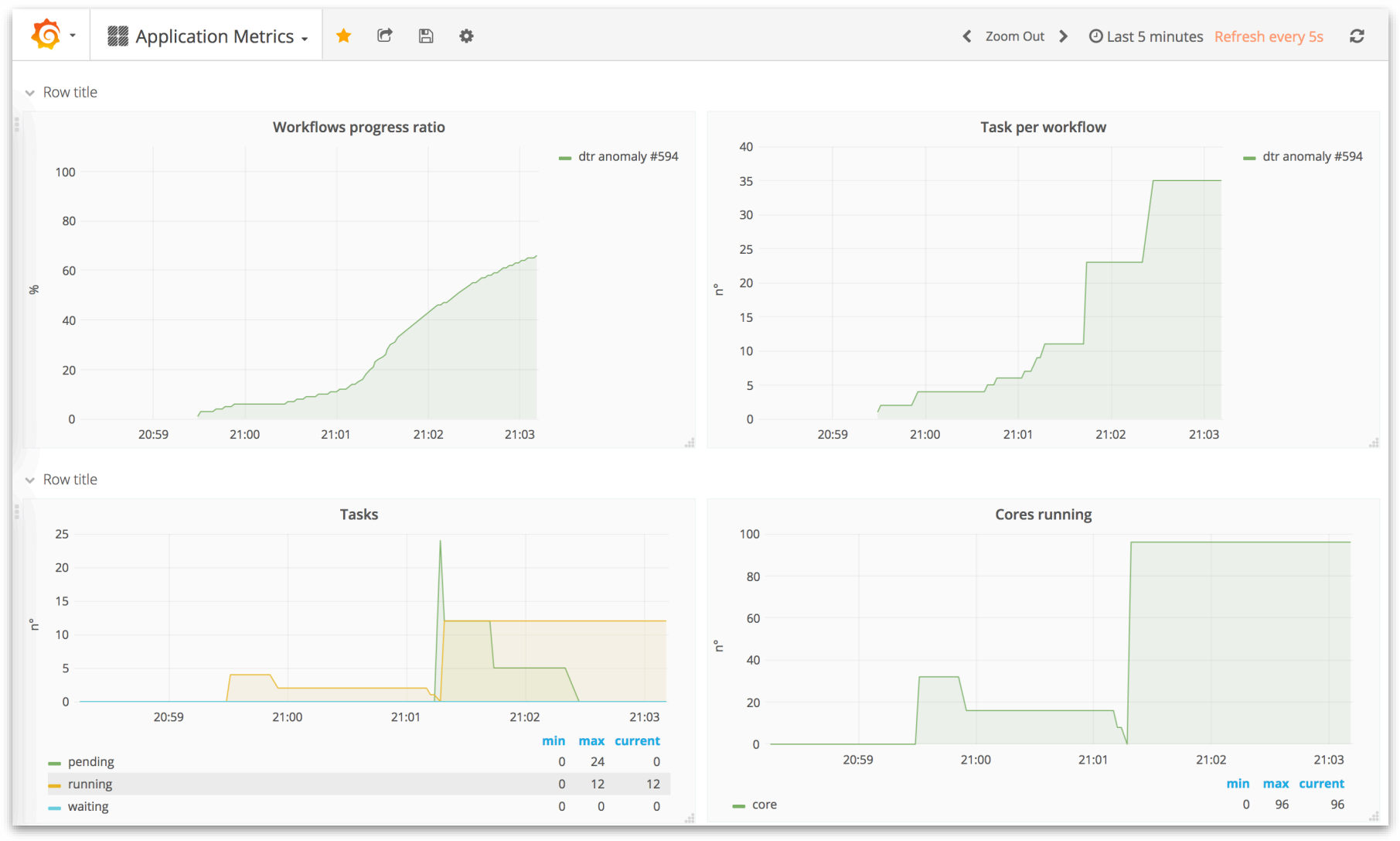


```
[32..4593] >> oph_explorecube cube=https://ophidialab.cmcc.it/ophidia/137/24706;limit_filter=10;subset_dims=lat|lon|tim
e;subset_filter=100:120|44:46|1:5;
[Request]:
operator=oph_explorecube;cube=https://ophidialab.cmcc.it/ophidia/137/24706;limit_filter=10;subset_dims=lat|lon|time;sub
set_filter=100:120|44:46|1:5;sessionid=https://ophidialab.cmcc.it/ophidia/sessions/320888504717371486471495550039634593
/experiment;exec_mode=sync;cwd=/;cdd=/home/delia;

[JobID]:
https://ophidialab.cmcc.it/ophidia/sessions/320888504717371486471495550039634593/experiment?1571#12493

[Response]:
tos
---

+===========+===========+==============================================================================+
| lat       | lon       | tos                                                                          |
+===========+===========+==============================================================================+
| 19.500000 | 87.000000 | 299.8564758301, 299.3809814453, 299.6741638184, 300.7666015625, 301.6275024414 |
|-----------|-----------|------------------------------------------------------------------------------|
| 19.500000 | 89.000000 | 299.6140441895, 299.1651611328, 299.6116027832, 300.6709289551, 301.7478637695 |
|-----------|-----------|------------------------------------------------------------------------------|
| 19.500000 | 91.000000 | 299.5370178223, 298.9378356934, 299.3948669434, 300.5995483398, 301.8059387207 |
|-----------|-----------|------------------------------------------------------------------------------|
| 20.500000 | 87.000000 | 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20 |
|-----------|-----------|------------------------------------------------------------------------------|
| 20.500000 | 89.000000 | 299.3892822266, 298.8489685059, 299.3197631836, 300.5918884277, 301.7084655762 |
|-----------|-----------|------------------------------------------------------------------------------|
| 20.500000 | 91.000000 | 299.2955627441, 298.6186828613, 299.0486755371, 300.3459167480, 301.5839233398 |
|-----------|-----------|------------------------------------------------------------------------------|
| 21.500000 | 87.000000 | 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20 |
|-----------|-----------|------------------------------------------------------------------------------|
| 21.500000 | 89.000000 | 299.0067138672, 298.3460388184, 298.8789062500, 300.5391540527, 301.7656860352 |
|-----------|-----------|------------------------------------------------------------------------------|
| 21.500000 | 91.000000 | 298.9141235352, 298.1416625977, 298.5674743652, 300.0816040039, 301.3976135254 |
|-----------|-----------|------------------------------------------------------------------------------|
| 22.500000 | 87.000000 | 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20 |
+===========+===========+==============================================================================+

Summary
-------

Selected 10 rows out of 63

[32..4593] >> oph_reduce
```

TASKS AND CLUSTER MONITORING

OphidiaLab environment features are provided through a set of various types of nodes like **computing nodes**, a **server front-end** and **client/services machines**.

In this environment, Ophidia allows the execution of **single operators**, **massive tasks** and **workflows of tasks**. Hence, an instance of the **Grafana** monitoring system is included to monitor the cluster resource usage and activity from an **infrastructural** and an **application-level** point of view.



	min	max	current
pending	0	24	0
running	0	12	12
waiting	0	0	0

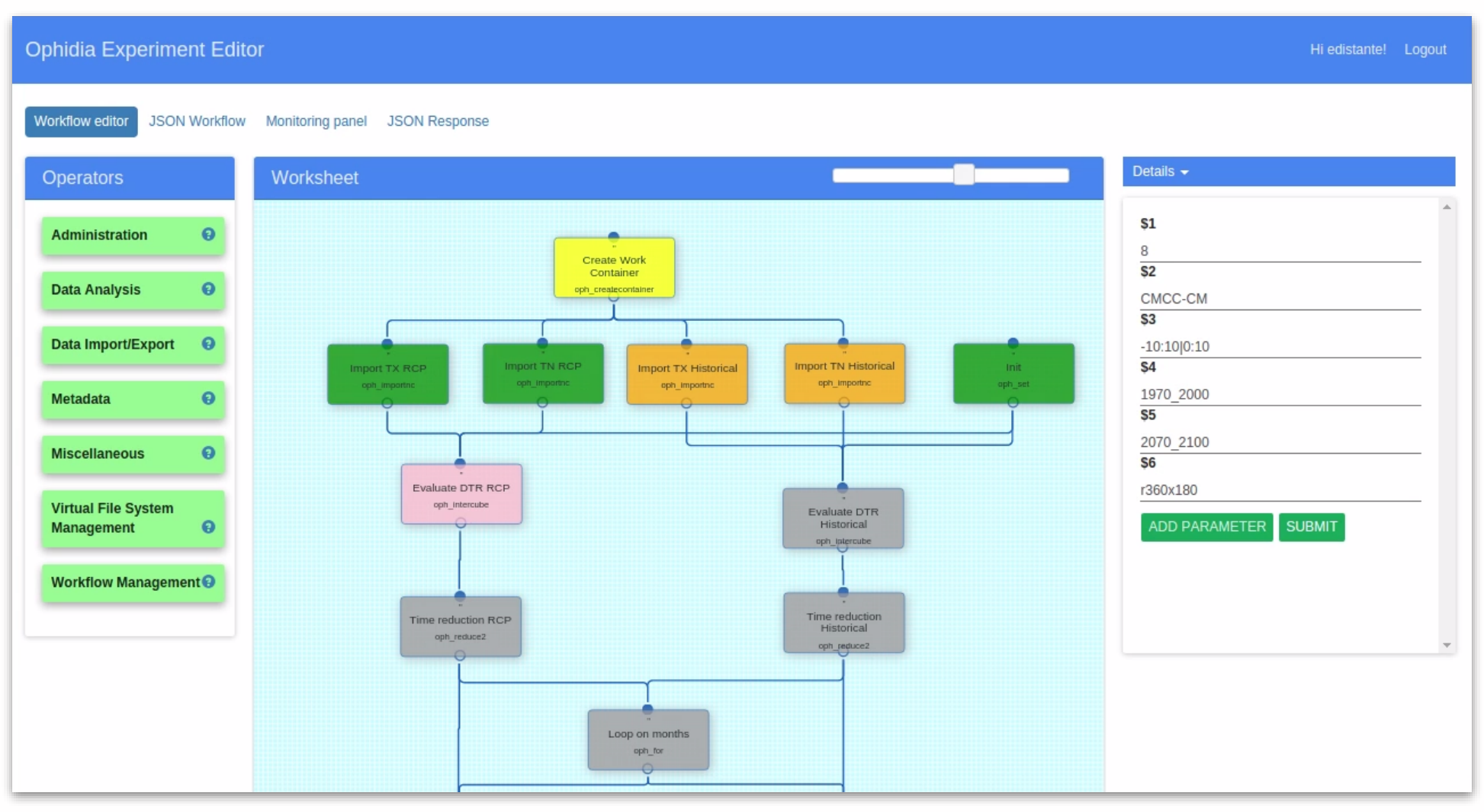
	min	max	current
core	0	96	96

INTERACTIVE WORKFLOW COMPOSITION

The **Ophidia Experiment Editor** supports the user in the creation and update of workflow-based experiments through an easy-to-use graphical web interface. The main panel provides a unified view with a drag-and-drop working area to combine the Ophidia operators and define complex dependencies between tasks.

The workflow editor implements the features to govern the whole workflow life-cycle, allowing:

- *workflow composition;*
- *task creation and update;*
- *dependency management;*
- *workflow submission;*
- *execution monitoring.*



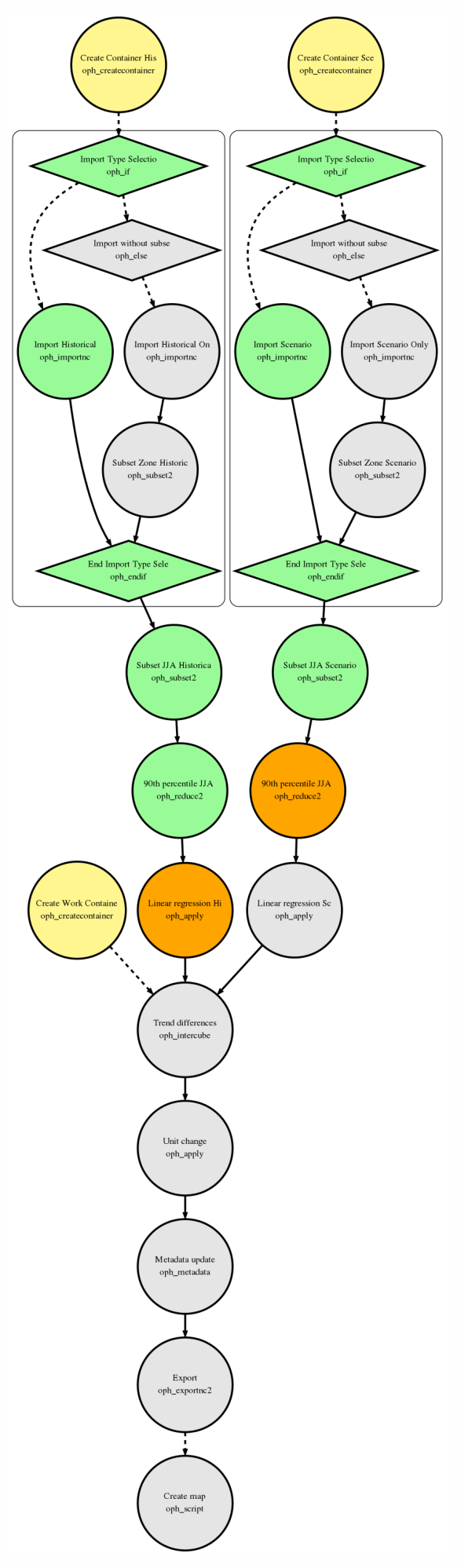
$1
8
$2
CMCC-CM
$3
-10:10|0:10
$4
1970_2000
$5
2070_2100
$6
r360x180

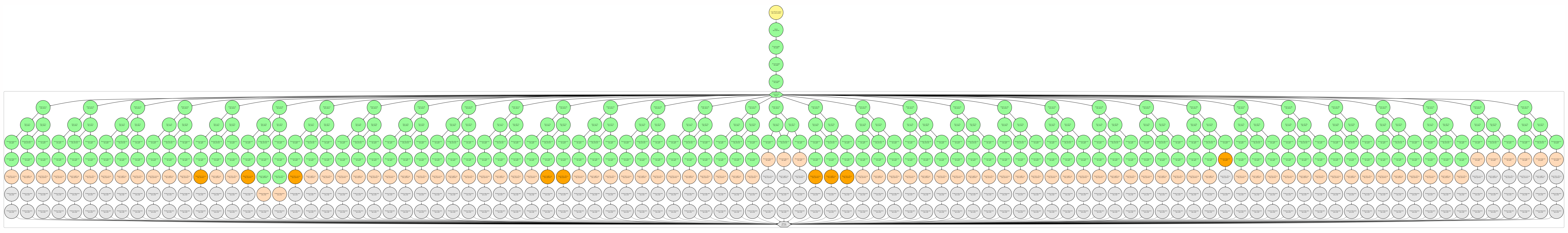
ANALYTICS WORKFLOWS FOR SCIENTIFIC INDICATORS

Ophidia embeds an analytics **workflow manager** designed to make more flexible the platform, to help in reducing the complexity of scientific experiments, to increase the re-usability, and fully exploit the available computational resources.

In the climate change context, several workflows for real-world use cases have been defined. By writing down a **simple** task graph including the basic operations to be executed, the user is able to quickly process large input datasets and evaluate one or more **indicators** like *sea surface temperature anomaly*, *precipitation trend* [2] (workflow on the right), *snow season statistics* [3] (workflow on the bottom), *climatological averages, unusual warm events.* Furthermore, the parallel workflow interface allows to easily replicate the same set of operations over different input datasets to compute complex data analysis with no effort (as shown in the picture below).

By using Ophidia a number of workflows have already been defined to perform experiments also in other scientific domains (e.g. astronomy, seismology, biology).





REFERENCES

[1] S. Fiore et al., "Ophidia: Toward Big Data Analytics for eScience", ICCS 2013, June 5-7, 2013 Barcelona, Spain, ICCS, volume 18 of Procedia Computer Science, pp. 2376-2385. Elsevier, 2013.

[2] S. Fiore et al., "Distributed and cloud-based multi-model analytics experiments on large volumes of climate change data in the earth system grid federation eco-system," IEEE BigData 2016, Washington, DC, pp. 2911-2918.

[3] A. D'Anca et al., "On the Use of In-Memory Analytics Workflows to Compute eScience Indicators from Large Climate Datasets", CCGrid2017, Madrid, Spain, May 14-17, 2017, pp. 1035-1043.