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

Alessandro D’Anca1, Donatello Elia1, Cosimo Palazzo1, Elisa Distante2, Sandro Fiore1, Giovanni Aloisio1,2

1 Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Lecce, Italy; 2Università 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.).

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