The Ophidia framework: toward cloud-based data analytics for climate change

Sandro Fiore, Alessandro D′Anca, Donatello Elia, Marco Mancini, Andrea Mariello, Maria Mirto, Cosimo Palazzo, and Giovanni Aloisio

European Mediteran Center on Climate Change, CMCC, I-73100 Lecce, Italy

The Ophidia project is a research effort on big data analytics facing scientific data analysis challenges in the climate change domain. It provides parallel (server-side) data analysis, an internal storage model and a hierarchical data organization to manage large amount of multidimensional scientific data. The Ophidia analytics platform provides several MPI-based parallel operators to manipulate large datasets (data cubes) and array-based primitives to perform data analysis on large arrays of scientific data.

The most relevant data analytics use cases implemented in national and international projects target fire danger prevention (OFIDIA), interactions between climate change and biodiversity (EUBrazilCC), climate indicators and remote data analysis (CLIP-C), sea situational awareness (TESSA), large scale data analytics on CMIP5 data in NetCDF format, Climate and Forecast (CF) convention compliant (ExArch).

Two use cases regarding the EU FP7 EUBrazil Cloud Connect and the INTERREG OFIDIA projects will be presented during the talk. In the former case (EUBrazilCC) the Ophidia framework is being extended to integrate scalable VM-based solutions for the management of large volumes of scientific data (both climate and satellite data) in a cloud-based environment to study how climate change affects biodiversity.

In the latter one (OFIDIA) the data analytics framework is being exploited to provide operational support regarding processing chains devoted to fire danger prevention. To tackle the project challenges, data analytics workflows consisting of about 130 operators perform, among the others, parallel data analysis, metadata management, virtual file system tasks, maps generation, rolling of datasets, import/export of datasets in NetCDF format.

Finally, the entire Ophidia software stack has been deployed at CMCC on 24-nodes (16-cores/node) of the Athena HPC cluster. Moreover, a cloud-based release tested with OpenNebula is also available and running in the private cloud infrastructure of the CMCC Supercomputing Centre.