Geophysical Research Abstracts
Vol. 20, EGU2018-18720, 2018
EGU General Assembly 2018

Ophidia: a datacube oriented framework for large-scale OLAP-based scientific data analysis experiments

Giovanni Aloisio (1,2) and Sandro Fiore (1)

(1) Euro-Mediterranean Center on Climate Change (CMCC Foundation), Lecce, Italy, (2) University of Salento, Italy

The Ophidia project is a research effort facing big data analytics challenges in multiple eScience domains, such as climate change, astrophysics, life sciences, etc. It aims at addressing scientific use cases related to the analysis and mining of large-scale datacubes.
A core component of the Ophidia software stack is its analytics framework. It (i) represents the analytical (OLAP-based) engine of the system and (ii) supports in-memory datacube analytics; (iii) yet, it has been developed to be extremely flexible and extendible, allowing the execution of a comprehensive set of datacube operators covering both data and metadata aspects.
The datacube operators are designed to work in parallel on large hyper-cubes thanks to an internal multidimensional data model. The range of available datacube operators goes from mathematical and statistical processing to multidimensional data sub-setting (e.g. slicing an dicing), data reduction, roll-up, drill-down, exploration. Recently OLAP-based data mining and machine learning algorithms have been also integrated to support predictive analytics on climate change and weather forecast datacubes. Besides data analysis, both metadata management (e.g. search, validate) and file system-like support for datacubes (e.g. datacube import, export, copy, delete, move, find) are also available to properly manage datacubes in the user space.
From a computational science point of view, the framework exploits the advantages offered by a distributed data-warehouse approach jointly with the use of HPC paradigms and in-memory techniques to address both fast and big datacube analysis.
Implementing a server-side approach, Ophidia exposes all the available operators through a standards OGC Web Processing Service interface. Additionally it (i) supports Python binding through PyOphidia - a very lightweight client library also available on conda-forge – and (ii) provides a workflow native interface to implement large-scale complex experiments involving hundreds of operators.
From an end-user perspective, the Ophidia framework has been applied to support the implementation of real use cases on multi-model climate data analysis, climate change and biodiversity indicators, as well as processing/visualization workflows for operational environments in different EU FP7 and H2020 projects such as: CLIP-C, EUBRAZILCC, EUBra-BIGSEA, INDIGO-DataCloud. Additionally, Ophidia is now part of the H2020 EOSC-Hub project service portfolio as a core component of the ENES Climate Analytics Service, a thematic service aiming at serving the climate community at large in the context of the Earth System Grid Federation as well as researchers from different domains interested in implementing datacube-based analytics experiments.