



## **The Ophidia Project: Towards Large Scale Climate Data Analytics**

S. Fiore (1), G. Aloisio (1), and I. Epicoco (2)

(1) Euro-Mediterranean Centre for Climate Change and University of Salento, Lecce, Italy (sandro.fiore@unisalento.it), (2) CMCC, Italy

By 2020, climate scientists are expected to generate hundreds of exabytes of data, distributed across several countries. The large volume of data and the time needed to locate, access, analyze, and visualize this data will greatly impact the scientific productivity. Significant improvements in the data management field will be critical to increase research productivity in solving complex scientific problems. I/O issues will become more and more critical in the management of large volumes of data. Moreover, parallel I/O solutions (both from a hardware and software point of view) will be key to implement efficient data management platforms.

Working on more efficient “data kernels” (i.e. data reduction operators) can represent an interesting and valid perspective to provide higher level software and libraries to enable post-processing activities, with stronger I/O performance.

In this context the Ophidia project (a research effort started in 2010 at the University of Salento and the Euro-Mediterranean Centre for Climate Change) combines together high performance computing and database management systems to provide users with an efficient and climate-oriented data analytics platform.

The talk presents the main goal of this project giving a complete overview of the proposed system both at an architectural and infrastructural level. Data compression, distribution and partitioning are also presented and discussed as enabling factors for a large scale data management system. Technical aspects about the implementation are also presented and discussed highlighting the main differences in terms of I/O performance, between a server-based solution and an embedded one.