

HEROES - Towards Eco-Responsible Hybrid HPC Platforms

Domain: HPC and Energy Transition

Abstract

Technology is evolving quickly. While artificial intelligence (AI) and machine learning (ML) are decades old, they are just now taking off. Therefore, bridging the gap between high-performance computing (HPC) and AI/ML is more important than ever. In this context, the HEROES¹ project [1], funded by the EuroHPC JU and national funding agencies from France, Spain and Italy, will develop a software solution enabling industrial and scientific user communities to easily advance complex simulation and ML workflows to HPC data centres and cloud infrastructures, with the goal to bring the project results to the market. This will allow user communities to choose the best platform to achieve their goals in time, within budget and with the best energy efficiency possible. The project focuses on workflows of strategic importance in the renewable energy domain and manufacturing applications where the design of more energy-efficient products, such as for energy-efficient vehicles, requires HPC.

The core of the project includes developing a decision module for workflow placement and the applying marketplace concepts to HPC; these embed from end-to-end the notion of eco-responsibility and will allow end-users to execute their simulations in the most energy-efficient way. For many years, it has been a topic only addressed from a hardware point of view, when HPC systems are designed. More recently, the rise of the electricity cost and the requirements for data centres to establish energy management as a global responsibility, made this topic not only a research topic but also and foremost a necessity for the future of the planet itself. Addressing this problem becomes even more challenging in a hybrid HPC environment.

HEROES platform embeds two complementary components that allow to monitor, predict, and act on the energy consumption of the workflows:

1. EAR [2] (Energy Aware Runtime) – an energy management framework for supercomputers that provides an easy-to-use and lightweight optimisation service, an energy and performance accounting and monitoring system, and global energy management.
2. OKA [3] – a data science platform for HPC environments that provides a centralized platform to gather information about on-premises or Cloud HPC platforms, analyse them, and provide predictions on their future behaviour, as well as a jobs' behaviour prediction module that is at the core of HEROES decision module.

Tightly integrated together, EAR and OKA enable the HEROES platform to optimise the energy consumption of the jobs by first selecting the best-fitted platform based on the job's requirements, the user's constraints, historical data, and current state of the available HPC and Cloud platform; and second by optimising the energy and power of a job during runtime by adapting the CPU, GPU and memory performances based on the actual needs of the jobs. Only minimum user intervention is needed: just providing high level goals (for energy optimization) such as minimize power or energy. This results in energy efficient jobs, thus lowering the carbon footprint of the workflows' execution.

¹ This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.