Support for CINCOMP project

EUROfusion spends significant resources in providing a dedicated supercomputer to the researcher of the consortium. Currently, this includes the MARCONI-Fusion supercomputer consisting of conventional CPUs (Intel Skylake) and a GPU (Nvidia Ampere) partition that is a dedicated share of the LEONARDO supercomputer. Both machines are hosted by CINECA, Bologna, Italy.

Whenever new hardware is taken into operation its assessment is done within the CINCOMP project. Depending on the resulting data, the Operation Committee decides whether or not the requirements from the Project Implementation Agreement (PIA) are fulfilled. In this process, it is important that the assessment is provided independently from CINECA in order to ensure an unbiased basis for decision-making. Following the planned replacement of both MARCONI-Fusion and LEONARDO, an assessment of the new hardware is due in the second half of the year 2024.

To make efficient use of the HPC resources it is necessary that the state of the hardware and software is constantly monitored. For this purpose the so-called “three code benchmark” is monthly executed. It allows to check the stability of MARCONI-Fusion and LEONARDO in terms of execution time for real production codes. It also provides the Operation Committee with objective metrics about the state of the two machines.

Furthermore, monthly meetings of the Ticket Committee take place to guarantee an efficient handling of the users’ support requests on EUROfusion HPC machine issues. In these meetings, the participants from the CINCOMP project have a bridging function between the users on the one side and the system and vendor personnel on the other side. It pursues performance, stability and software stack issues that affected all EUROfusion users’ applications, which has been particularly helpful for the recently installed GPU cluster LEONARDO. Performance degradation of MARCONI-Fusion and LEONARDO nodes, as well as software stack issues are spotted, leading to tens of tickets per year that are eventually investigated by CINECA, Intel, and Nvidia. Such tickets imply in-depth investigations, leading to an overall improvement in HPC knowledge among the EUROfusion community.

In addition, the performance of the individual codes has to be monitored routinely to identify “ailing codes”, i.e. codes which are particularly inefficient and can therefore only reach a significantly smaller fraction of the peak performance than expected. On MARCONI-Fusion, the top ten users typically consume half of the available node-hours. Therefore, it is of particular importance that the performance of these codes is monitored closely. The operating system kernels of MARCONI-Fusion and LEONARDO support the measurement of the FLOP rates via the hardware performance counters for all applications submitted to the respective batch systems. The *hpcmd* tool from MPCDF Garching, Germany, is able to produce derived metrics from theses counters, including e.g. their memory footprint. It was recently installed on MARCONI-Fusion and multiple tests were performed in order to determine its correctness. Several identified issues were successfully resolved together with the CINECA and MPCDF support teams.

It is foreseen that, in a first phase, any new *hpcmd* related issues arising on MARCONI-Fusion should be resolved in a timely manner. The developers of *hpcmd* from MPCDF offered support for debugging the software if necessary. In addition, CINECA has to dedicate human resources on their side. Coordination will be done by CINCOMP. This includes further testing until the software is mature. With the installation of the new supercomputer of EUROfusion, Pitagora, in the second half of the year, the *hpcmd* tool has to be implemented and tested on both its conventional and GPU partitions. Kinga Gal, from Program Management Unit (PMU) will be trained by CINCOMP to use the *hpcmd* tool. This will give the PMU the opportunity to use the performance data to identify ailing codes and to contact the code developers for improving the situation.

In a second phase, all ACH's should be involved in the improvement of ailing codes whenever they are using a significant fraction of the total node-hours on Pitagora. Already now each ACH is responsible for a list of codes provided by E-TASC. Whenever an ailing code falls under the responsibility of a certain ACH, it should make a profiling of the code followed by an assessment to identify the underlying reasons for the poor performance. A close collaboration with the code developers is mandatory. Finally, the responsible ACH should come up with suggestions about possible improvements of the situation. If the issue cannot be fixed on a short time scale, as e.g. when a parallelization concept has to be changed, or algorithmic adaptations are needed, a request from the developers for ACH support should be submitted at the next E-TASC allocation meeting.

Resources required

- 12 PM (now 9 PM) per year by Serhiy Mochalskyy and 3 PM per year by Roman Hatzky both from ACH-MPG
- One visit per year by Serhiy Mochalskyy of the ISC High Performance conference in Europe or similar: costs 3000 EUR
- 3 PM per year for each ACH to handle the requests of ailing codes.