The role of medium size facilities in the HPC ecosystem: the case of the new CRESCO4 cluster integrated in the ENEAGRID infrastructure

G. Ponti, F. Palombi, D. Abate, F. Ambrosino, G. Aprea, T. Bastianelli, F. Beone, R. Bertini, G. Bracco, M. Caporicci, B. Calosso, M. Chinnici, A. Colavincenzo, A. Cucurullo, P. Dangelo, M. De Rosa, P. De Michele, A. Funel, G. Furini, D. Giammattei, S. Giusepponi, R. Guadagni, G. Guarnieri, A. Italiano, S. Magagnino, A. Mariano, G. Mencuccini, C. Mercuri, S. Migliori, P. Ornelli, S. Pecoraro, A. Perozziello, S. Pierattini, S. Podda, F. Poggi, A. Quintiliani, A. Rocchi, C. Sciò, F. Simoni, A. Vita

Technical Unit for the Development of IT and ICT Systems
ENEA — Italian Agency for New Technologies, Energy and Sustainable Economic Development
Lungotevere Thaon di Revel 76 – 00196 Rome, Italy
e–mail: giovanni.ponti@enea.it

INDUSTRIAL POSTER PAPER

***Abstract*—Medium size HPC clusters play an important role in the HPC landscape in that they provide both the training environment for system scalability and a flexible production field for a large class of numerical problems. In this poster we present CRESCO4, the latest medium size HPC cluster purchased by ENEA, in operation since few months. CRESCO4 is part of a family of HPC systems, all integrated within ENEAGRID, a large infrastructure for cloud computing, which includes all the computational facilities installed at several ENEA sites in Italy.**

***Keywords*—HPC, supercomputing, Linux clusters**

I. Role of Medium Size HPC clusters

Since a long time it has been established that performance data populating the TOP500 list tend to follow Zipf distributions [1], e.g.

$$R_{\max} = \frac{A}{k^b} \tag{1}$$

with k denoting the rank in the list and A and b representing time depending constants. In particular, it is known that A tends to increase exponentially in time as a reflection of the Moore's law. More subtle is the trend of the exponent b: while displaying a slight decrease over the years 1999–2007 [2], it shows a slow growth since 2009 up to date, thus signaling an increase in time of the unbalance of computing power between higher and lower ranks in the list[a]. More precisely, Fig. 1 shows the trend of the power–law parameters of $R_{\max}$ as obtained from numerical fits to the yearly data, with interpolations neglecting the lowest 100 items in the lists, as they show a slightly steeper decrease than the rest. In the absence of additional data enabling a robust differential

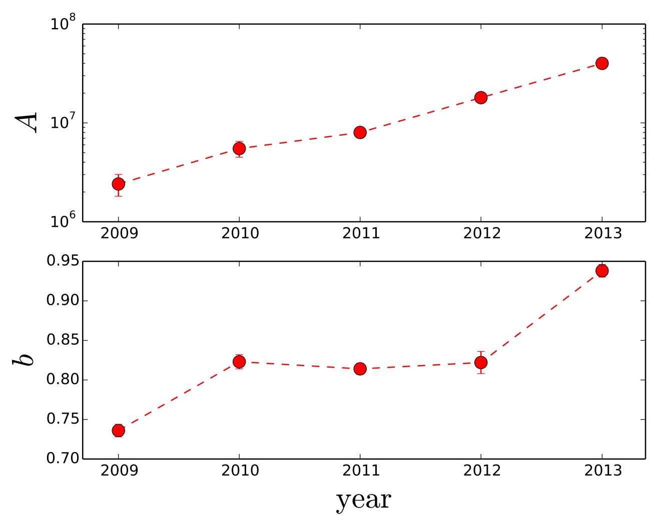

Figure 1. 5–year trend of the power–law parameters characterizing $R_{\max}$ as obtained from numerical fits to TOP500 lists from 2009 to 2013

analysis per geographic area, it is reasonable to expect that eq. (1) holds not only on the global stage but also at a national level. If this is correct, we can speculate that both global and local HPC ecosystems have been recently evolving towards a concentration of computing power within larger and larger facilities, aimed at reaching the exa-scale computing, whereas — by contrast — intermediate machines are suffering from a less focused attention.

[a]Notice however that the shape of the distribution depends to some extent upon the specific benchmark chosen so far in order to measure $R_{\max}$, namely the HPL one. The introduction of the new benchmark HPCG could affect non-trivially the size of the exponent b and its time dependence.

Although such a concentration of computing resources allows forefront researchers to dare more and more challenging computational projects (which could never be run on small to medium size facilities), it is true as well that accessing and successfully exploiting large facilities rises several crucial issues, from both technical and management points of view.

On the technical side, high scalable applications must be tailored to the specific hardware solutions adopted by the leading edge machines, such as special communication networks and/or specific co–processor architectures. Moreover, the access rules to large size machines usually require a heavy amount of bureaucracy which, though well justified by the cost and value of the resources made available, result in hampering fast exploration of new ideas and methodologies. As a consequence, medium size clusters and flexible access can represent an answer for scalability training and developing.

In addition, medium size clusters are optimal for numerical problems which do not intrinsically require top class machines (partitioning large machines in a plethora of small subsections often results in a waste of precious resources). For this reason, it is not rare that scientific and technological communities prefer to invest R&D funds to purchase their own machine, even though this ends up in occupying the lowest ranks of the TOP500 list (if ever at all...). In this context the Italian HPC ecosystem is highly concentrated and rather different from other European countries, such as Germany, France and United Kingdom, as most of the Italian TOP500 resources are polarized in a limited number of large machines and institutions, totalling five TOP500 entries compared to the 20–30 of the other cited countries.

The above picture sets the background to discuss the case of CRESCO4, the latest Linux cluster at ENEA, in operation since few months, at the service of a national community of applied researchers, working in collaboration with academic and industrial partners.

II. CRESCO4 Linux Cluster

The CRESCO computing laboratory at ENEA consists of a family of Linux Clusters. As the name suggests, CRESCO4 is the fourth (and newest) one in the family [3]. It is made of six racks, out of which five are occupied by computing nodes and one gathers all network devices, see Fig. 2. The peak computing power of the machine amounts to $R_{\text{peak}} = 101\,$Tflops and its measured HPL efficiency is 85%.

Here are some technical specs of the machine. The system is composed by 38 Supermicro F617R3-FT chassis, each hosting 8 dual CPU nodes. Each CPU, specifically an Intel E5-2670, hosts in its turn 8 cores, for a total number of 4,864 cores. These operate at a clock frequency of 2.6 GHz. The machine is provided with a RAM memory of 4 GB per core. The computing nodes access a DDN S2A9900 storage system, for a total storage amount of 480 TB. Finally, the computing nodes are interconnected via an Infiniband 4xQDR QLogic/Intel 12800-180 switch (432 ports, 40 Gbps). The single node architecture is summarized by Fig. 3.

Figure 2. CRESCO4 Linux cluster

The cluster is located in a new computer room, where the cooling is provided by Climaveneta i-AFO 50 air units and the cold air flows through the floor to a confined cold aisle in front of the computing nodes. These units are dual cooling systems, featuring a direct expansion circuit with inverter and a secondary chilled water circuit connected to external dry coolers. This technology permits to take advantage of free-cooling when the external conditions are favourables. In the qualification tests of the system, the HPL benchmark yielded 85 Tflops with a computing electrical power consumption of 109 kW, a cooling power consumption of 35 kW and no free–cooling during the benchmark. The PUE (Power Usage Effectiveness) was 1.32 with 0.78 TFlops/kW. The optimization of the free-cooling system settings is currently underway and global results will be available after at least one year of data collection.

CRESCO4 has been purchased in the framework of the TEDAT project (Advanced Technologies and Diagnostics for transport policies). The project aims at supporting and strengthening the use of existing facilities for R&D, and at promoting the upgrade of equipment and technologies for the analysis and diagnostics of materials and components, in the framework of transport policies. Its targets, perfectly consistent with the mission of ENEA and the specific programmes of the technical units involved, are:

- to strengthen hardware facilities and technical know–how in support of the national productive system, by upgrading the already operating laboratories and starting–up new ones;
- to create a national centre for advanced diagnostics and 3D characterization of materials and surfaces of interest for transport policies;
- to support and develop technologies for the production of materials and surfaces used in transport;
- to promote the identification and synthesis of new materials for catalyzers and electrodes for batteries, replacing

old critical materials;
- to foster the development of new technologies for energy–saving light sources, lighting systems, traffic signs, etc. in the framework of transport policies;
- to promote IT technologies for the management and the remote access of complex experimental facilities, modelling, simulation and experimental data mining.

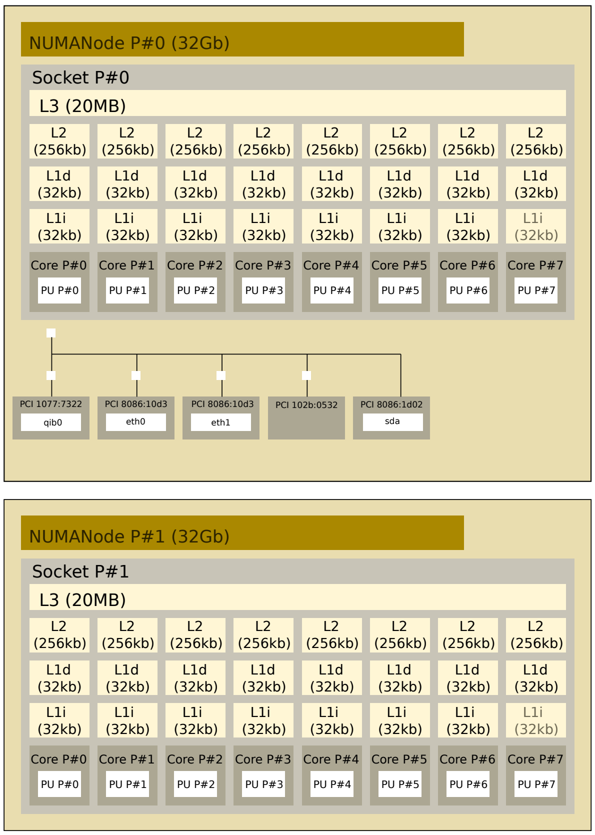

Figure 3. CRESCO4 single node architecture

III. Integration within ENEAGRID

As well as its predecessors CRESCO1/2/3 (which are still in operation at ENEA), CRESCO4 is also integrated within ENEAGRID, a large infrastructure for cloud computing, which includes all the ENEA computing resources installed at the various ENEA research centres in Italy, altogether 13 sites distributed across the whole country, as shown in Fig. 4. ENEAGRID is characterized by solid structural components, offering reliability and easy management, and by web interfaces which have been developed in–house and customized so as to ensure a friendly user environment:

- authentication via Kerberos v5;
- geographic filesystems: AFS/OpenAFS;
- GPFS parallel file system also on WAN among different clusters and sites;
- resource management via LSF Multicluster;
- system monitoring via Zabbix;
- web access via in–house FARO interface, see [5];
- user management via in–house WARC interface, see [3].

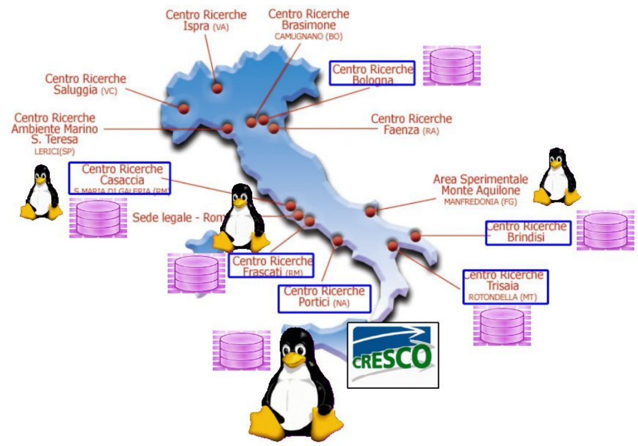

Figure 4. Geographical map of the ENEAGRID sites

The computing resources currently provided to the users are x86–64 Linux systems (the CRESCO HPC gather $\sim 10,000$ cores) and dedicated systems (e.g. GPU systems). The computing resources spread essentially across four ENEA research centres:

- ENEA Portici: CRESCO1 (672 cores), CRESCO2 (2,720 cores), CRESCO3 (2,016 cores), CRESCO4 (4,864 cores)
- ENEA Frascati: CRESCOF (480 cores)
- ENEA Casaccia: CRESCOC (192 cores)
- ENEA Brindisi: CRESCOB (88 cores)

Specifically, CRESCO4 has been placed in the ENEA Portici site, where the largest fractions of computing and man power concentrate. ENEA Portici is connected to the Internet through the PoP GARR of Napoli–Monte S. Angelo thanks to two 1 Gbps links (GARR is the Italian Research & Education Network, planning and operating the national high-speed telecommunication network for University and Scientific Research, see [4]). In 2012 the in–/out– data transfers amounted to 230 Tbyte, equivalent to an average bandwidth of 60 Mbps, while the transfer peak value reached 80% of the available bandwidth. In 2013 an increase of 30% of the overall network traffic was observed with respect to 2012, due to the entrance of new users and to changes in software applications.

IV. Numerical Research Projects

Since its start-up in February 2014, CRESCO4 is running numerical projects covering several research areas, from Material Science (storage of Hydrogen, European flagship on graphene, materials for nuclear technologies, organic-inorganic adhesion) to Technologies for Energy and Industry (Thermo-acoustic instabilities in combustion, turbulence/ combustion interaction, sub-grid LES models, development of numerical algorithms, extension to supercritical fluxes), Environmental Modeling (climate models for the Mediterranean Sea, air pollution dispersion) and Nuclear Fusion (simulation of Alfven waves, physics and design technologies for nuclear fusion devices).

These projects are well suited for a medium size machine for various reasons. The exploratory nature of some of them takes advantage of the flexibility in the access to the computational resources while for others the limited scalability of the software could hardly benefit from running on a very large machine. In some important cases the limit in scalability is due to the number of available licenses of proprietary codes which, in some contexts such as Computational Fluid Dynamics, are an essential tool for technological development activities. In the latter case, the proprietary codes are often available for standard computing platforms only, as the market driving forces discourage software companies from developing code versions specific to the very specialized architectures of some of the top–class HPC systems. For all these reasons the choice of a medium size cluster such as CRESCO4 fits optimally with a large fraction of our community needs.

The user community of CRESCO4 amounts to about 50 entitled users, out of which 20 make a full–time use of large chunks of the machine. This group of users is actually a subset of the wider CRESCO community, amounting to about 200 researchers. Initial users have been selected among those in CRESCO running jobs with the highest core numbers and foreseeing in the short term a computational demand exceeding the resources made available by the older CRESCO1/2/3. In order to have a clear picture of the research fields which are currently and will be in the future covered by CRESCO4, it is interesting to look at the relative weight of the various research fields pursued by the wider CRESCO community. This is illustrated in Fig. 5 *(top)*, referring to 2012 (latest data currently available, see [6]).

In Fig. 5 *(bottom)* we show instead the relative weight of the software packages run by the CRESCO users in 2012. As can be seen, the leading fields are Material Science and Computational Fluid Dynamics.

V. Conclusions

As the time evolution of the TOP500 distribution discussed in sect. I suggests, the recent trends of the HPC world are towards an overall concentration of computational resources in larger and larger facilities, at the expense of medium size infrastructures. When facing the strategical dilemma shared by several scientific institutions, which is whether to invest money in buying CPU time from external partners or in buying a medium size in-house machine, the ENEA experience is that in many circumstances the latter choice may fit better the needs of average users running standard scientific/technological software applications. From a wider perspective, the presence and diffusion of medium size machines across the world of R&D turns out to be of great advantage even for the development of newer and more powerful top-level counterparts, in that it favours a progressive and smooth migration of entire communities of users towards more advanced and complex standards of hardware and software solutions. Our belief is that CRESCO4 is a good compromise between the current state-of-the-art of the HPC and the concrete needs of scientific computational practices.

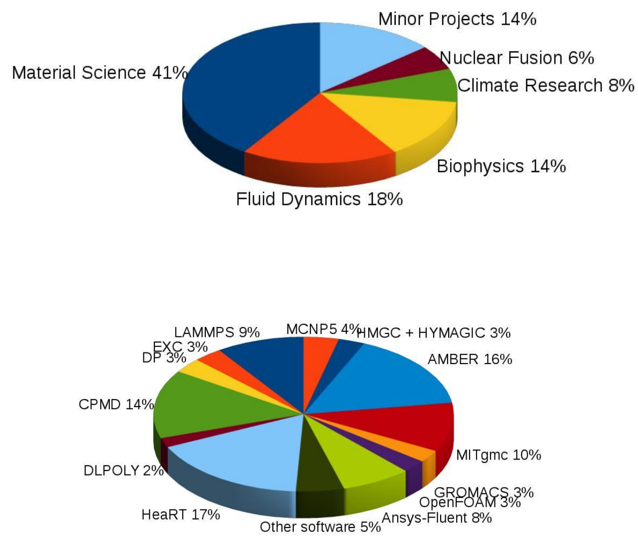

Figure 5. *(top)*: relative weights of numerical research fields in CRESCO; *(bottom)*: relative weights of software packages used by CRESCO users

Acknowledgments

CRESCO4 has implemented in the framework of the TEDAT project, an endeavour co-funded by ERDF within the programme "PON Ricerca e Competitività" by the Italian Ministry for Education, University and Research. Special thanks go to the TEDAT project leader L. Tapfer, head of the Technical Unit for Material Technologies - ENEA Brindisi Research Centre, and to the whole project staff.

References

[1] D.G. Feitelson, *On the Interpretation of TOP500 data*, International Journal of High Performance Computing Applications, Summer 1999, 13:146–153.

[2] M. Ripeanu, *Note on Zipf distribution in TOP500 supercomputer list*, IEEE Distributed Systems Online, 2006

[3] Complete information can be found at http://www.cresco.enea.it

[4] http://www.garr.it

[5] A. Rocchi, S. Pierattini, G. Bracco, S. Migliori, F. Beone, A. Santoro, C. Sciò, S. Podda, *FARO - The Web portal to access ENEAGRID Computational Infrastructure*, Procs. of the International Workshop on Science Gateways (IWSG 2010), 20-22 SEPTEMBER 2010, Catania (IT), R. Barbera, G. Andronico and G. La Rocca (Eds.), Consorzio COMETA (2010). ISBN 978-88-95892-03-0

[6] High Performance Computing on CRESCO Infrastructure: research activities and results 2012. ISBN: 978-88-8286-302-9. Web: http://www.utict.enea.it/it/rapporti-tecnici