Vol. 19, n.° 2, pp. 85-88, 2020
Revista UIS Ingenierías
Página de la revista: revistas.uis.edu.co/index.php/revistauisingenierias
Universidad Industrial de Santander

Evolution of the maintainability of HPC facilities at CIEMAT headquarters

Evolución de la sostenibilidad de las infraestructuras HPC en la sede central del CIEMAT

Antonio Juan Rubio-Montero [a], Angelines Alberto-Morillas [b], Rosa De Lima Herrera-Insua [c], Pablo Colino-Sanguino [d], Jorge Blanco-Yagüe [e], Manuel Giménez [f], Fernando Blanco-Marcilla [g], Esther Montes-Prado [h], Alicia Acero [i], Rafael Mayo-García [j]

Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT),
Av. Complutense 40, 28040, Madrid, España. Email: [a] antonio.rubio@ciemat.es, [b] angelines.alberto@ciemat.es,
[c] r.herrera@ciemat.es, [d] pablo.colino@gmail.com, [e] jorge.blanco@ciemat.es, [f] manolo.gimenez@ciemat.es,
[g] fernando.blanco@ciemat.es, [h] esther.montes@ciemat.es, [i] alicia.acero@ciemat.es, [j] rafael.mayo@ciemat.es

Received: 20 August 2019. Accepted: 29 December 2019. Final version: 16 March 2020.

Abstract

Since its establishment in 1951, CIEMAT has been continuously boosting the use of computation as a research method, deploying innovative computing facilities. Hence, Vectorial, MPP, NUMA, and distributed architectures have been managed at CIEMAT, resulting in an extensive expertise on HPC maintainability as well as on the computational needs of the community related to international projects. Nowadays, the evolution of HPC hardware and software is progressively faster and implies a continuous challenge to increase their availability for the greater number of different initiatives supported. To address this task, the ICT team has been changing towards a flexible management model, with a look toward future acquisitions.

Keywords: resilience; management practices; history of computing.

Resumen

Desde su creación en 1951, el CIEMAT ha estado impulsando continuamente el uso de la computación como un método de investigación, desplegando plataformas de cómputo innovadoras. De esta manera, arquitecturas vectoriales, MPP, NUMA y otras completamente distribuidas han sido gestionadas en el CIEMAT, acumulando un extenso conocimiento sobre su sostenibilidad y sobre las necesidades de las comunidades científicas relacionadas con proyectos internacionales. Actualmente, la evolución del hardware y el software para HPC es cada vez más rápida e implica un desafío constante para aumentar su disponibilidad debido al número de iniciativas que el centro apoya. Para abordar esta tarea, el equipo TIC ha estado cambiando su gestión hacia un modelo flexible, con una mirada puesta en las adquisiciones futuras.

Palabras clave: adaptación; buenas prácticas; historia de la computación.

ISSN Printed: 1657 - 4583, ISSN Online: 2145 - 8456, **CC BY-ND 4.0**
How to cite: A. J. Rubio-Montero, A. Alberto, R. Herrera, P. Colino, J. Blanco, M. Giménez, F. Blanco, E. Montes, A. Acero, R. Mayo-García, “Evolution of the maintainability of HPC facilities at CIEMAT headquarters,” *Rev. UIS Ing.*, vol. 19, no. 2, pp. 85-88, 2020. doi: 10.18273/revuin.v19n2-2020009

1. Introduction

Good practices in maintainability continue being the basis for improving the reliability to reach the desired availability of any system [1].

This is especially important in HPC environments, where the failure of a critical component could waste many CPU-hours and prevent restarting calculations for days. Formally, availability, i.e. the uptime percentage, results from reliability, the mean time between fails and their recovery. Maintainability (or serviceability) is the combination of steps to reduce these times. They include guaranteeing the power supply, redundancy, data recovering, securitize the access and use, cooling, re-scheduling computational work and shutdowns, monitoring, predict failures, among others.

However, the perception of system availability is moving from the traditional measuring of its uptime time into the capacity to compute any kind of calculation for which its hardware is suitable. In other words, while administrators are usually devoted to only improve reliability to increase the global throughput of the system, users are worried about the feasibility of their projects, due to configuration issues. Of course, the design of each HPC platform will determine its suitability for certain type of processing. Nevertheless, why can some calculi become unfeasible in certain HPC facilities when they are straightforward in similar architectures? The reason is the different configuration in these systems (OS, middleware, specific software, and licenses). This issue is an important availability aspect from the user's point of view. If the configuration doesn't match his computational requirements, this system is actually unavailable for the user. This is not a new idea, for example NIST [2] defines availability as "Ensuring timely and reliable access to and use of information." If the configuration hampers processing data, the system doesn't accomplish the previous statement.

Provisioning different configurations is a current trend in the HPC industry, but it was not a priority until the last decade. In the past, an HPC system was usually purchased for a small number of users, with similar computational requirements. However, since the introduction of commodity hardware and the open source in HPC, the evolution of these platforms is progressively faster, as well as number of collaborations that a research center must support. This is the case of CIEMAT.

CIEMAT, *Centro de Investigaciones Energéticas Medioambientales y Tecnológicas*, is a Spanish R&D Public Organism which has been carrying out research and technological development focused on energy and environment since 1951. From the beginning, CIEMAT has been continuously boosting the use of computation as a research method, deploying innovative computing facilities. The institution was originally focused on nuclear fission development, as it was denoted with its first name, JEN (*Junta de Energía Nuclear*). This type of research and the early computing knowledge determined the facilities implemented in these pioneering decades. Progressively, other areas related were incorporated, firstly high energies and nuclear fusion, then radiotherapy, biotech, and environmental forecasting, among others. Thus, Vectorial, MPP, NUMA, and distributed architectures have been managed in CIEMAT, resulting in an extensive expertise on the computational needs of the community related to international projects, and not only to its own scientific staff.

This paper describes this transition as follows: firstly the facilities set before 2008 were listed, then current HPC clusters were described and finally, future plans were explained.

2. Background (60's – 2008)

2.1. From monolithic to commodity clusters

The first computer devoted to scientific calculations in Spain was purchased by JEN (the former name of CIEMAT) in 1959. This machine was an UNIVAC SOLID STATE UCT, from the Remington Rand Company. Later on, three UNIVAC machines, 1106, 1110, and 1110/81, were consecutively bought by leasing in the 70's, being the latter operated by terminal and not by punched cards. All these computers were available for any researcher from other national institutions such as Ministries, INTA (*Instituto Nacional de Técnica Aeroespacia*l), CSIC (*Consejo Superior de Investigaciones Científicas*), Universities, etc.

In 1985, CIEMAT launched its first ICT internal programme for distributing tasks among several terminals connected in a LAN and a new computer, an IBM 3090/150 with a vectorial processor, was purchased. Posteriorly, three CRAY computers were incorporated to the CIEMAT Data Center: XMS, YMP-EL and J90 (16 processors). In those days, the needs of the thermonuclear fusion group required vectorial processing, but the ratio cost-flexibility were changing to MPP platforms. In this sense, the CRAY T3E computer arrived in the late 90's, with a configuration suitable for other calculations.

A step beyond was taken in the following decade with the SGI/Origin 3800 (160 MIPS processors R14000, 600 MHz) computer with cc-NUMA architecture and IRIX64 as OS. Its performance was upgraded with two SGI/Altix machines of 96 (1.3 GHz) and 64 (1.5 GHz) Itanium 2 processors, counting on Linux. Although these were shared memory systems, their access was yet non-uniform and their processing associated units were similar to nodes. Thus, the batch scheduler (in this case, LSF) and its policies gained importance, being an essential part of the HPC management. However, these systems can still be considered as unique and immutable entities based on licensed software, but not yet monolithic.

Finally, a Beowulf cluster, *Lince*, was installed in several rounds. This system counted on 88 nodes with two 32-bit Xeon processors. Network was a mixed Infiniband - Ethernet interconnection. The purpose was using as much as possible open source to reduce costs and to enable software upgrades. Therefore, Scientific Linux, PBS/Torque, and free monitoring tools were used.

2.2. To the DPC establishment

Previous subsection illustrates the general evolution in computing: from a centralized model allowing few users without external connectivity, to an interrelated environment with multiple distributed hardware elements, services, administration domains, users, and uses, exposed to Internet. Consequently, possible failure points and security risks exponentially grew decade by decade. This imposed the need of creating a data processing center (DPC), the design and management policies of which were the foundation to maintain the systems running.

Focused on HPC, three main increases implied the DPC establishment: consumption and heat; the number of independent devices; and, the amount of data. To accomplish the first one, a building was kitted out with forced cooling and batteries in 1986. On the other hand, the evolution of LANs enabled the distribution of data through different RAID NAS, in detriment of local SANs. Consequently, storage was continuously enlarged and the backup could not be performed by local tapes. Thus, the first tape robot was purchased in 2000, a StorageTek L9310 (10 drives, 1500 cartridges, 30 TB).

Therefore, although the management based on simply ordering punched cards and support required to the manufacturer when something failed were overcommitted, usual practices did not allow required reliability. For example, monitoring was mainly performed writing specific scripts in a way in which many warnings were sent by e-mail. Additionally, few elements were designed for high-availability. Failure anticipation was not possible in many cases.

3. Current HPC facilities and practices

Through the last decade, CIEMAT headquarter incorporated three Infiniband-enabled x86-64 clusters in different administration domains.

The first one, *Euler* (2008 and later upgrade in 2010), is the main HPC facility to the date, with 240 nodes, 1920 Xeon cores and an Rpeak of 23TFlops. Moreover, a departmental mini-cluster, Dirac, (2012, counting on 11 nodes, 132 cores, Rpeak of 1.27TFlops), is used for specific applications in the nuclear fusion area. Both facilities are similarly managed, batch scheduling relies on PBS/Torque and the base software is unchanged (except security updates) since their installation.

On the other hand, *ACME* (2015 and 2017) is mainly used for research on HPC efficiency, considering the requirements of Exascale Era [3]. The objective is to experiment with new job scheduling mechanisms. Consequently, developments are carrying on Slurm, taking account check-pointing [4] and virtualization tools too [5]. Additionally, it is used for calculations of advanced users. Consequently, it software is completely updated every year. The cluster counts on 24 nodes, 720 cores with an Rpeak of 40,6 TFlops. Additionally it has 4 Tesla P100 GPU cards with an Rpeak of 18.8 TFlops (double precision).

Taking account number of nodes, users (350 in 30 research groups, 100 of them external), and data generated, the most important maintainability improvements have been done in these areas:

Service redundancy. With *Euler*, high availability was set for HPC clustering: at least two paths per switching stacks, Ethernet network as spare of Infiniband, active-backup for the batch scheduler, and active-active for the user interfaces and storage. Although originally *Euler* counted on Lustre filesystem, currently a NetApp cDOT facility maintains the high availability, but exporting NFS.

Three level backup. The great amount of HPC data generated and lately managed (1,5 PB) resulted in their distribution among several RAID machines (16), all of them serving NFS. According to the importance of these data, they can be copied by three backup systems with different policies. Firstly, cDOT facilities enable hourly snapshots maintained every three weeks. Secondary storage servers daily make differential rsync copies until two months. Both copies are accessible by users through read-only mounted directories. Long-term backup is performed by a IBM TS3584 tape library (18 drives, 1581 cartridges, 4.42 PB), where daily incremental copies are maintained for next three months.

Uninterruptible power supply, efficient cooling and whole monitoring. In December 2017, a diesel engine (1,000 KVA) was connected to DPC. This avoids the need of stopping computing jobs when most electrical issues happen. Additionally, the data center was remodeled to efficiently guide air flow only through racks. Moreover, several Nagios are progressively used for monitoring all aspects in DPC, i.e. not only the status of nodes, services or RAIDs, but also batteries and power.

4. Manageability in future acquisitions

Nowadays, the center has opened a bid for the acquisition of the *Euler* replacement. Experience obtained with current clusters [6, 7] taught us that we can't rely forever on the same OS release, as well as the hardware became obsolete from the third year. Therefore, the new cluster will adopt a constellation design, where a yearly update cycle of software and the 25% of computing hardware, will be performed. First phase (the first quarter of cluster) will count on 44 nodes with two Xeon gold 6148 processors, i.e. 1760 cores, Rpeak > 135 TFlops and > 672 TB based on Lustre filesystem.

On the other hand, there is the capacity of Slurm for integrating check-pointing and provisioning customized user configurations. Advantages are clear in big clusters [6], they increase reliability, enable backward compatibility and reproducibility. However the flexibility provided is sometimes prioritized over efficiency. Provisioning implies underperformance due to virtualization (virtual machines, containers), time loss in reconfigurations (reinstall nodes), and it will only be justified if the system will support many kinds of processing tasks. Automatic check-pointing also give users the temptation of submitting never-ending jobs. Additionally, some availability and security requirements are currently restricted by Law [8], especially when these resources will be accessible through the international supercomputing initiatives [9, 10]. The consequent solution should be offering a limited number of environments transparently to users, but supervised and validated by administrators.

References

[1] M. Y. Hsiao, W. C. Carter, J. W. Thomas, W. R. Stringfellow, "Reliability, Availability, and Serviceability of IBM Computer Systems: A Quarter Century of Progress," *IBM Journal of Research and Development*, vol. 25, no. 5, pp. 453-468, 1981. doi: 10.1147/rd.255.0453

[2] *United States Code - Definitions (44 U.S.C., Sec. 3542) and NIST Glossary*, [Online]. Available: https://csrc.nist.gov/Glossary/?term=3103.

[3] F. Cappello, "Fault tolerance in petascale/exascale systems: current knowledge, challenges and research opportunities," *Int. J. High Perform. Comput. Appl.*, vol. 23, no. 3, pp. 212-226, 2009. doi: 10.1177/1094342009106189

[4] J. A. Moríñigo, M. Rodríguez-Pascual, R. Mayo-García, "On the Modelling of Optimal Coordinated Checkpoint Period in Supercomputers," *J. of Supercomputing*, vol. 75, no. 2, pp. 930-954, 2019. doi: 10.1007/s11227-018-2621-1

[5] A. J. Rubio-Montero, E. Huedo, R. Mayo-García, "Scheduling multiple virtual environments in cloud federations for distributed calculations," *Future Generation Computer Systems*, vol. 74, pp. 90-103, 2017. doi: 10.1016/j.future.2016.03.021

[6] D. Stanzione *et al.*, "Stampede 2: The Evolution of an XSEDE Supercomputer," in *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact*, vol. Part F1287, pp. 1–8. doi: 10.1145/3093338.3093385

[7]J. A. Moríñigo, P. García-Muller, et al. "Benchmarking LAMMPS: Sensitivity to Task Location under CPU-based Weak-scaling," *5th Latin American Conference on High Per-formance Computing (CARLA2018), Comm. Comp. Inf. Sci.*, vol. 979, 2019. doi: 10.1007/978-3-030-16205-4_17

[8] Spanish Official Bolletin (BOE-A-2010-1330) *R. D. 3/2010, de 8 de enero, por el que se regula el Esquema Nacional de Seguridad en el ámbito de la Administración Electrónica*.

[9] E. Mocskos, C. J. Barrios, H. Castro, et al. "Boosting advanced computational applications and resources in Latin America through collaboration and sharing," *Comp. Sci. & Eng.*, vol. 20, no. 3, pp. 39-48, 2018. doi: 10.1109/MCSE.2018.03202633

[10] PRACE Homepage. [Online]. Available: http://www.prace-ri.eu/