



High-Performance Computing Training by Skills for Advanced Computing Ecosystems

Entrenamiento por Competencias de Computación de Alto Rendimiento para Ecosistemas de Computación Avanzada

Carlos J. Barrios H. ^{1a}, Robinson Rivas ², Xavier Besson ³, Gilberto Diaz ^{1b}, Yiannis Georgiou ^{4a}, Pedro Velho ^{4b}

¹ Grupo de investigación en Cómputo Avanzado y de Gran Escala (CAGE), Escuela de Ingeniería de Sistemas e Informática, Supercomputación y Cálculo Científico UIS (SC3UIS), Universidad Industrial de Santander (UIS), Colombia. Orcid: [0000-0002-3227-8651](https://orcid.org/0000-0002-3227-8651) ^a, [0000-0001-8188-5784](https://orcid.org/0000-0001-8188-5784) ^b. Emails: cbarrios@uis.edu.co ^a, gjdiaz@uis.edu.co ^b

² Departamento de Computación, Facultad de Ciencias, Universidad Central de Venezuela (UCV), República Bolivariana de Venezuela. Orcid: [0000-0001-7900-0764](https://orcid.org/0000-0001-7900-0764). Email: rrivas@ciens.ucv.ve

³ Faculté des Sciences, des Technologies et de Médecine, Département Ingénierie, Université de Luxembourg, Luxembourg. Orcid: [0000-0002-7783-2185](https://orcid.org/0000-0002-7783-2185). Email: xavier.besson@uni.lu

⁴ Ryax Technologies, France. Orcid: [0000-0003-1264-7234](https://orcid.org/0000-0003-1264-7234) ^a, [0000-0002-9767-468X](https://orcid.org/0000-0002-9767-468X) ^b. Emails: yiannis.georgiou@ryax.tech ^a, pedro.velho@ryax.tech ^b

Received: 23 July 2023. Accepted: 30 January 2024. Final version: 16 March 2024.

Abstract

High-Performance Computing (HPC) is one of the pillars of developing modern science and disruptive technologies, uniting computer architectures and parallel programming into multidisciplinary interactions to face domain-specific problems. That is why different areas of knowledge require their future professionals (scientists or not) to acquire skills in using HPC. The Super Computing and Distributed Systems Camping School, SC-Camp, is a non-profit activity that proposes a series of courses about HPC with an important focus on practical sessions (more than half of the time) addressed to undergraduate and graduate students who could benefit from HPC by demand. It is an itinerant school, bringing the HPC knowledge to a different place every year, focusing on diversity, sustainability, and humanity.

Keywords: Education; HPC; Training; Outreach.

Resumen

La computación de alto rendimiento (HPC) es uno de los pilares del desarrollo de la ciencia moderna y las tecnologías disruptivas, uniendo arquitecturas informáticas y programación paralela en interacciones multidisciplinarias para enfrentar problemas específicos de dominio. Es por ello que distintas áreas del conocimiento exigen a sus futuros profesionales (científicos o no) adquirir habilidades en el uso de HPC. El Camping Escuela de Super Computación y Sistemas Distribuidos, SC-Camp, es una actividad sin ánimo de lucro que propone una serie de cursos sobre HPC con un importante foco en sesiones prácticas (más de la mitad del tiempo) dirigidas a estudiantes de grado y posgrado que podrían beneficiarse de HPC según la demanda. Es una escuela itinerante, que lleva el conocimiento de HPC a un lugar diferente cada año, con un enfoque importante en la diversidad, la sostenibilidad y la humanidad.

ISSN Online: 2145 - 8456

This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 License. [CC BY-ND 4.0](https://creativecommons.org/licenses/by-nd/4.0/)



How to cite: C. J. Barrios, R. Rivas, X. Besson, G. Diaz, Y. Georgiou, P. Velho, "High-Performance Computing Training by Skills for Advanced Computing Ecosystems," *Rev. UIS Ing.*, vol. 23, no. 1, pp. 175-188, 2024, doi: <https://doi.org/10.18273/revuin.v23n1-2024014>

Palabras clave: educación; HPC; entrenamiento; extensión.

1. Introduction

Supercomputing is one of the pillars of developing modern science and supporting technology development and knowledge. Skills in HPC are important for various professional fields and crucial in scientific and technology areas. For example, accurate and timely weather forecasts are based on detailed models requiring large-scale parallel simulations. In data analytics and artificial intelligence, the growth in applications and uses depends on the many abilities and capacities necessary to exploit parallelism and accelerators, support large-scale data, guarantee scalability, and perform concurrency in processing. Several initiatives, such as the HPC Certification Forum [1] and the Special Interest Group in High-Performance Computing (SIG HPC) [2] Education of the ACM (From the French acronym Conseil Européen pour la Recherche Nucléaire, European Organization for Nuclear Research), present an organization of the skills necessary in HPC by goals and roles. The two visions, built from discussions between the specialists and users, propose three main actors: HPC systems engineers, HPC software engineers, and computer scientists.

These actors are easily related to the general computer science players: HPC system engineers integrate, monitor, tune, optimize, and evaluate the performance of each element and process of the HPC hardware and software environment. HPC software engineers develop parallel programs, debug, and analyze the performance of the applications, optimizing the cycle of the HPC software. The latter, computer scientists, can be associated with users. But in the case of HPC, they need specific skills: coding, running, and deploying parallel programs efficiently using HPC Environments. These three actors are in continuous interactions and must share common knowledge.

Besides the formal courses in different universities around the world, the HPC community has been proposing for years summer schools, seminars, tutorials, workshops, and certifications. In the interest of the scientific community, the specialists (among them the engineers) and the industry were involved in developing training on computer architectures, applications, tools, methodologies, and programming paradigms. The deployment of different specific schools around the world allows for democratizing knowledge and the creation of previously non-existent collaborations in multicultural, multidisciplinary, and inclusive ways. By the beginning of the 2000's years, international projects motivated by high-energy physicists, like the CERN HPC

summer schools, were made around the world to support collaborations related to the treatment of the large volume of data generated by CERN facilities (CERN projects that still exist today, as the case of the Worldwide Large Hadron Collider Computing Grid - WLCG). Federated around the well-known concept of Grid Computing, this computer community leveraged distributed computing, parallelism, and connectivity with common goals [3]. Inspired by these initiatives, different institutions and associations created summer schools that included skills at different levels for scientists, programmers, and infrastructure administrators involved in developing science and technology projects using HPC. Initiatives such as the International Summer School in Grid Computing (ISSGC) [4] supported by the International Collaboration to Extend and Advance Grid Education (ICEAGE) [5] led by M. Atkinson and D. Ferguson in the context of the Enabled Grids for E-Science (EGEE) [6] project officially ended in 2010, brought together experts in the three mentioned HPC skills with students and trained many of the people who today lead similar initiatives such as the International Summer School on HPC Challenges in Computational Sciences (ISHPCSS) [7] or the International Supercomputing and Distributed Systems Camping School (SC-Camp) [8].

In this paper, following the main goals for HPC competencies, we present an original non-formal school for training and development focused on skills and collaboration. Developed by the *International Supercomputing and Distributed Camping School*, or SC-Camp, to keep it short, it ran, for now, for (14) fourteen years in several cultural and economic contexts and adapted its approach due to the COVID-19 pandemic situation. The contribution is organized as follows: the current section generalizes the HPC summer schools motivated by the Grid Computing projects. The following section presents the advanced computing context, introduces the HPC ecosystem, and identifies the necessary skills for the actors involved in HPC projects. Section III shows the relevant aspects of the Supercomputing and Distributed System Camping School (SC-Camp) and how the methodology and activities of the SC-Camp improve the skills of the different participants. Finally, we present our conclusions and further work.

2. The Advanced Computing Context

Complex scientific computational projects from theoretical and practical angles coincide with using the term *Advanced Computing* [9]. This involves hardware

and software technologies, especially HPC, but also methodologies for using supercomputers and HPC architectures in the computing continuum to support large-scale data processing, including visualization. Indeed, advanced computing projects need large storage, networking, interaction, and processing capacities. Additionally, to guarantee good results in predictable times, they require specialized execution and deployment and efficient exploitation of computing resources. These elements constitute the stakeholders' competencies in these projects: the HPC infrastructure operators, research or HPC software developers, and HPC scientists.

2.1. The Involved Actors and Necessary Skills

Following the cloud computing visibility model, we introduce (3) three actors involved in advanced computing projects corresponding to (3) three well-known levels: HPC technology and operations engineers, HPC (or research) software developers, and HPC scientists. All of them are in continuous interaction to achieve the goals and ensure the best deployment (including energy efficiency) and performance of the applications of the advanced computing projects. Figure 1 shows the different elements impacted by each one of the roles in the HPC environment at levels from infrastructure to applications, as said before, taking the cloud visibility model. From those roles, the responsibilities and actions of each actor are visible, allowing the identification of the involved knowledge. For example, exploiting the elements related to infrastructure requires skills in computer architectures and networks, as well as the provision of services and monitoring.

Elements related to the support of application execution and development require skills such as the deployment of environments or development services. In the case of application use or visualization, abilities such as data repository manipulation, customization of applications, or access to resources are joined with frameworks, repositories, and access services. Of course, some competencies and actions overlap over many roles for which the actors must share standard know-how. With those elements, it is possible to classify the involved actors into three categories.

2.2. HPC Technology and Operations Engineers

Nowadays, the role of HPC Technology and Operations is given to computer or systems engineers with competencies in computer architecture, system administration, organization, and networking. However, for HPC, the necessary competencies should be more specific towards architectures and infrastructures with parallel processing support (undoubtedly multicores and multithreading), advanced networking and I/Os (high bandwidth interconnection, latency optimization, parallel communication strategies), hybrid integration (accelerators, multiple processing units), scalability (many cores, large scale systems, data-massive and data-intensive storage, and transfer), memory configuration (NUMA, DMA), file systems (including parallel file systems), large scale visualization and distributed services (user interaction, on-premise or hybrid/cloud services).

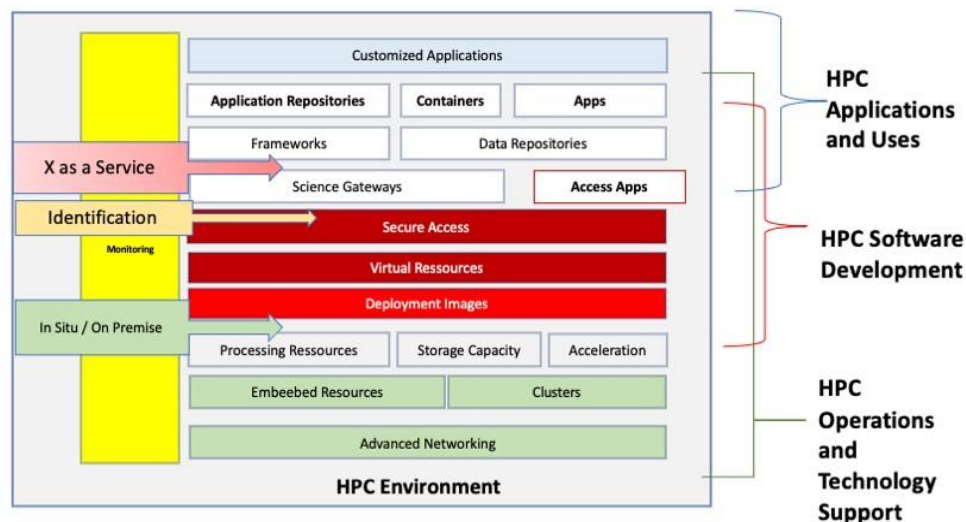


Figure 1. HPC Roles in the HPC Environment.

On the other hand, software-related competencies are also required, or at least strongly needed. They focus on analysis and performance evaluation techniques, monitoring, debugging, profiling, fault tolerance and checkpoint processes, system integration and organization, operating systems, deployment, and frameworks. Of course, all of this implies expertise in parallel runtime, execution support, and development environments (libraries, programming language support).

2.3. HPC or Research Software Developers

It is well known that developing scientific software (or HPC software) is more than just writing code. A specific community of software engineers involved in writing, maintaining, and managing software inside research institutes and universities proposes a formalization of the necessary role and good practices to develop HPC and scientific computing applications as part of this formalization appears the Research Software Engineering (RSE) concept and an important growing community clearly differentiated from HPC users (scientists or not). The formalization process has been proposed and highlighted in [10]. RSEs are responsible for the entire software cycle (planning, definition of requirements, designing and prototyping, software development, testing, deployment, operations, and maintenance. Depending on the methodology and approach, some authors propose other definitions and stages, adding for example and evolution stage or an analysis stage. An interesting review of the different methodologies and development techniques can be seen in [11]). However, the RSE community includes other responsibilities and tasks to guarantee good and sustainable software: community interaction, training, and defining policies to support the community and recognize the impact of research software. These activities and tasks are organized into four pillars that manage the RSE Skills: Software Development, Community Networking, Training and Outreach, and finally, Policies Influence [12].

2.4. HPC Scientists (or Users)

HPC scientist is the end user, who will use the advanced computing environment to analyze large-scale data, compute, simulate, or visualize looking high capabilities and high computing performance. In the HPC ecosystem, limiting their tasks to simply running an application or launching a script code is impossible. This actor must prepare the launch and the deployment of their programs but also participate in the development and analysis of HPC applications execution. It often leads the HPC scientists or users to code or experience in software

development, so HPC scientist needs to know about parallel computing. The technical competencies can be listed in three groups: First, related to coding and scripting, as the use of shell scripting and parallel programming (knowing the different parallel programming paradigms); second, related to the HPC environment interaction, as the use of the command line interface, knowledge of the main commands of the cluster operating system (and the use of tools to measure the system performance), analysis and performance evaluation (including debugging, profiling and tuning of code) using tools and techniques, research and selection of software environments; and finally a third group related with data managing, visualization and analysis of results, as the data treatment (including version control, data structure analysis), data visualization, reproducibility, and results release. However, as HPC Software Developers presented before, community networking and influence in policy decisions by part of stakeholders are non-technical yet desired competencies.

In practice, it might be difficult for a single actor in each category to master all the competencies highlighted. Instead, the actor will specialize, create comprehensive expertise on a few specific topics, and acquire basic knowledge of the other skills. Working teams covering a specific role are usually built around this principle. For example, the HPC Infrastructure and Operations team will gather people with different expertise (e.g., hardware and networking or software compilation and tuning) to cover all the competencies required for the role. Additionally, some competencies clearly overlap between roles in the Advanced Computing Ecosystem.

The collective goal of building and running better sustainable software applications, optimizing the use of resources, and achieving a significant impact drives the development of training proposals in the HPC community. Then, it is possible to create common teaching spaces to leverage interaction and multidisciplinary competencies, as shown in the next section.

3. Integrate Skills Actors in HPC Ecosystems

In the Introduction, we mentioned schools and training events focusing on developing HPC skills, mainly addressing developing parallel programs. In Section 2., we showed how all actors of the HPC ecosystem have common skills, and there is a continuous and necessary interaction between them to improve their capabilities and the performance and impact of their work. Observing the span of technical skills, we identify common spaces of training and development, contextualizing the ecosystem. Figure 2 shows three main integrated spaces

associated with the actors of the ecosystem, with their own and common skills that can guide the development of courses or practical sessions of HPC schools through the participation of all involved actors; likewise, the relationship between the actors/elements of the HPC ecosystem, as side skills (i.e. communication abilities). Then, the three development spaces are **HPCTOE**, **HPCRSE**, and **HPCUSE**. The acronym **HPCTOE** refers to the HPC Technical and Operation Engineers. **HPCRSE** refers to the HPC Research Software Engineers (or Developers), and **HPCUSE** for the HPC Scientific Users or simply users.

Seeing the skill set in the last section, shared or common knowledge, core skills (represented in pink boxes), and side skills are identified for each role. **Figure 2** highlights how the three functions of HPCUSE, HPCRSE, and HPCTOE share some common skills, e.g., monitoring, analysis, and performance evaluation (this skill includes debugging, profiling, benchmarking and testing of applications or the HPC environment, including platform and infrastructure) but also a set of specific skills or core knowledge for the role. All the technical skills are shared between the different actors in the sense that the core skills of one role are side skills for the other roles. Side skills show competencies that are not required to perform the main tasks of the roles but for which basic knowledge is desirable to understand the complexity of the system properly and some abilities, as the case of communication or teamwork mentioned before. It points

out clearly how much the competencies and abilities of the three types of actors overlap and why it makes sense to train them in interaction with each other.

Motivated by these characteristics, we proposed a special training event to integrate elements and actors following joint research or application interests, the *Supercomputing and Distributed Systems Camping School* or SC-Camp.

3.1. The SC-Camp

SC-Camp is a training event that can be understood as a summer school about Supercomputing and distributed systems. It is a non-profit event addressed to all students, including those with a lack of financial support. Also, places are sought where the diversity and participation of students can be guaranteed and with common development interests. It proposes a series of courses around the theme of High-Performance Computing with an important focus on practical sessions (more than half of the time). SC-Camp targets undergraduate and graduate students in the field of Computer Science, Engineering, and any other fields that could benefit from HPC (Physics & Material Sciences, Biology/Bioinformatics, Finances, etc.). However, it is also open to the participation of researchers, senior engineers, or end users involved in non-profit activities (such as teaching at different levels).

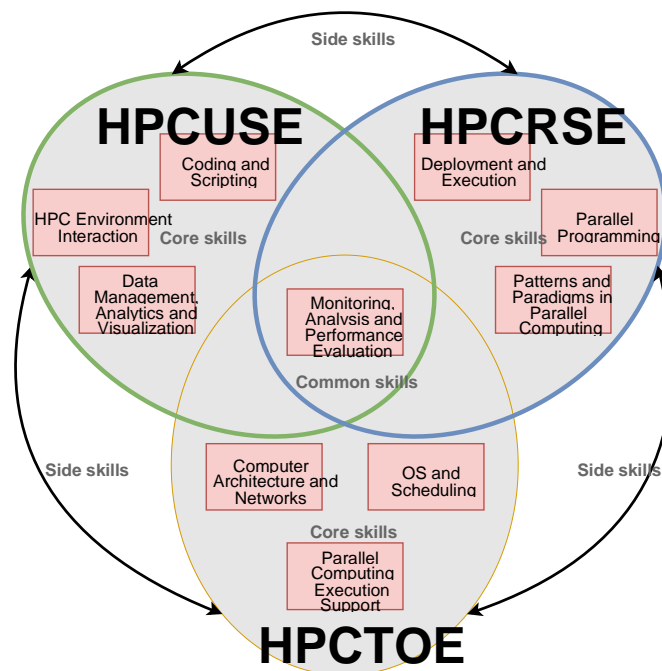


Figure 2. HPC Technical Skills and Spaces of Development.

Taking advantage of the Internet and high-speed networks available today, one can exploit high-performance computing infrastructures from anywhere, even in the middle of nature. These aspects are fundamental because SC-Camp seeks to observe social and ecological impact, responsibility, sustainability, and inclusion. This initiative of researchers inspired by these ideas offers undergraduate and master students state-of-the-art lectures and practical programming sessions on high performance and distributed computing topics. It is an itinerant school, bringing the HPC knowledge to a different place every year.

3.1.1. Organization

Founded by young researchers and PhD students in 2010, since the beginning, the SC-Camp has been organized

yearly by a general and a local committee. The general committee is composed of researchers and teachers from different countries. The members are mainly from research institutes and universities however there are allies from the industry (mainly from the high technology industry). The general committee is supported also for a steering committee composed by global very-known researchers and personalities related with HPC. The local committee is made up of the applicants responsible for the SC-Camp. These applicants are teachers at universities, research managers of institutions, or scientific network coordinators. Table 1 summarizes the different SC-Camp schools by year, showing the local site and the hosting institution (the official name of the institution is given in English or national language to facilitate the identification of the institution).

Table 1. Host Institutions

Year	Country and Site	Host	Type
2010	Catay ¹ , Colombia	University: Universidad Industrial de Santander (UIS) Catay is a path of the Piedecuesta municipality in Colombia.	On-Site
2011	Turrialba, Costa Rica	University: Universidad de Costa Rica (UCR)	On-Site
2012	Vega Sol ² , Venezuela	University: Universidad de los Andes (ULA) Vega Sol is a path of the Merida State in Venezuela.	On-Site
2013	Colima, Mexico	University: Instituto Tecnológico de Colima (ITC)	On-Site
2014	Manizales, Colombia	Institute: Bionformatics and Computational Biology Center (BIOS)	On-Site
2015	Toluca, Mexico	Institute: Mexican National Advanced Research Center (CINESTAV)	On-Site
2016	Santiago de Cali, Colombia	University: Universidad del Valle (UNIVALLE)	On-Site
2017	Cadiz, Spain	University: Universidad de Cadiz (UCA)	On-Site
2018	Da Nang, Vietnam	University: Da Nang International Institute of Technology (DNIIT)	On-Site
2019	Madrid, Spain	Institute: Centro Nacional de Investigaciones Energeticas, Medioambientales y Tecnológicas (CIEMAT)	On-Site
2020	Virtual, Colombia	Network: Red Nacional Academica de Tecnología Avanzada (RENATA)	Virtual
2021	Virtual, Chile	Network: Cooperación Latino Americana de Redes Avanzadas (REDCLARA)	Virtual
2022	Hybrid, Colombia, and Mexico	Network: Cooperación Latino Americana de Redes Avanzadas (REDCLARA) and Advanced Computing System for Latin America and Caribe (SCALAC), Site 1: Universidad Industrial de Santander, Colombia (UIS) and Site 2: Universidad de Guadalajara, México	Hybrid
2023	Cartagena de Indias, Colombia	University: Universidad de Cartagena, Colombia	On-Site

Table 1 shows that SC-Camp took place in different countries (and continents) following the general goals of the school but also particular requirements that we will present later. The first SC-Camp was made as a summer camping in the middle of nature in 2010, bringing together mainly undergraduate students of systems engineering and computer science from Colombia and Venezuela. This camping school mainly was directed towards programmers and administrators in undergraduate formation. The interesting thing about this first SC-Camp is that it laid the foundations to reflect on the roles we discussed in this article and matured the idea of interests. It also strengthened aspects such as inclusion and multidisciplinary participation. The second SC-Camp 2011 was held in Costa Rica and sought to include people from different backgrounds and levels (undergraduate and graduate), mainly related to computer science, biology, physics, and mathematics. The site was the Atlantic campus of the Universidad de Costa Rica in Turrialba, at the Turrialba Volcano National Park.

Still, the main focal point was bioinformatics and the environment. Students came from Costa Rica, Nicaragua, Panamá, México, Jamaica, and Cuba. Most SC-Camp schools were generally developed in universities, non-capital cities, or secondary places.

Candidates hosting sites for SC-Camp present an application evaluated by a committee composed of the founders of the SC-Camp, an international steering committee integrated by researchers, and the former guest keynotes. The committee decides where the following two SC-Camps will be held according to the following criteria:

- **Hosting:** There is a preference for non-developed countries and secondary or less-known cities. Cadiz and Madrid, Spain, were selected as exceptions in 2019 because they guaranteed the participation of students from African-Mediterranean countries. The natural environment is also observed following the first criteria proposed by the founders of SC-Camp, as well as the type of entity to host the camping school with a special interest in public universities. For example, in 2011, Turrialba in Costa Rica was selected due that in this place is the Atlantic site of the University of Costa Rica, and in 2023, Cartagena de Indias in Colombia as the leading site of the University of Cartagena; both are public universities in Latin America.
- **HPC Resources and Internet Access:** Good Internet quality is important to access HPC resources remotely during the courses and practical sessions. A preference is given to use the HPC facilities proposed by the hosting

site to train the students on the platform they are most likely to use. However, if needed, the Universidad Industrial de Santander [13] in Colombia via the High Performance and Scientific Computing Center (SC3UIS) [14] and the University of Luxembourg [15] via the HPC@Uni.lu team [16] gracefully provide access to their HPC clusters for the duration of the school. Both sites were used including the SC-CAMP 2018 in Da Nang, Vietnam.

Academic Interests and Multidisciplinary Participation: Interests in developing competencies in the context of particular goals in science and technology are influential in selecting the SC-Camp host. For example, a host institution might want to build bioinformatics or data analysis competencies for forecasting and weather. Another would like a more general formation of thinking in different application areas. Generally, it is observed that there is the participation of the three possible actors and of different domains of knowledge. Additionally, the committee stimulates the involvement of students of various levels as part of the SC-Camp (undergraduate or graduate students and professors or researchers). It is noted that in three editions of SC-Camp, the committee decided to focus on participants with the same level of formation: Ph.D. students in 2016 and graduate students (master's and Ph.D.) in 2017 and 2019. Returning to in-person SCCAMP, during 2023, participants are also students and junior teachers (or teachers assistants).

Inclusion and Diversity of Participants: **Figure 3** shows the SC-Camp participation identifying the participation of women, men, and the total of attendees. Although it is requested that the involvement of both women and minorities be guaranteed to exceed 10%, it has been tough to comply with it due to the sociocultural context of the hosts. However, it has been achieved that there is always the participation of minorities and of neighboring countries or regions. For example, in SC-Camp 2013, students from Mexican indigenous communities who lived in both the state of Colima and Jalisco were invited to participate; the same was true for undergraduate students of different areas residing in regions of the Han River in SC-Camp 2018 and in 2023, students from the Caribbean region from Costa Rica to Venezuela, principally females.

- **Sponsorship and Funding:** SC-Camp is a non-profit activity, but economic funds are necessary for its development. All instructors and keynotes are volunteers, and travel expenses are mostly borne by the SC-Camp association, also a French non-profit association that channels resources from organizations associated with science and technology, mainly European. However, a host must guarantee local

expenses for all the instructors and keynotes and low-cost or free charges for student participation. Then, local and general committees drive resources given by companies and national and international entities supporting the initiative, such as the case of RedCLARA (Latin American Cooperation of Advanced Networks) and SCALAC (Advanced Computing System for Latin America and the Caribbean) [17] when the SC-CAMP is developed in Latin America region. A company that has accompanied the initiative since its creation is NVIDIA.

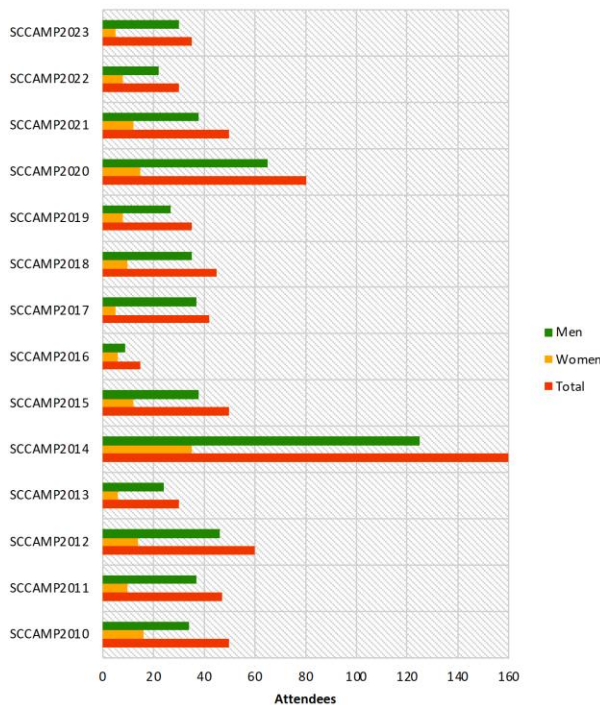


Figure 3. Diversity among registered students.

3.1.2. Attendees

Attendees are the student participants in the SC-Camp. The kind of attendees is mixed: as actors and roles in the HPC or Advanced Computing Ecosystem, and they must be diverse in different aspects, such as age, education level, culture, ethnicity, or gender (a clear deficiency in this aspect is observed in Figure 3 with the participation of women and men in the SC-Camp. Among the participant applications, the number of women is significantly lower than the number of men). Table 1 and Figure 3. show information about the type of host entity and number of students by year. This information is important because each camping school had objectives that generated guidelines for both the type and number of participants.

For example, relating the data in Table 1 with Figure 3 highlights two events with the maximum and minimum

number of participants. In 2016, during the SC-CAMP in Cali, Colombia, only 15 doctoral students who participated directly in projects related to the physics of materials were selected due to the interests of a research center in a growing process that implies skills in HPC. On the contrary, in 2014, in Manizales, Colombia too, 160 attendees were selected to generate a critical mass of students with HPC skills among computing sciences, biology, and engineering undergraduates through more than ten educational institutions in a wide region of a country and to support a newly created national computational biology center.

Depending on the targeted audience, some background knowledge in computer programming is required, as well as a general understanding of English (proficiency enough to follow a technical talk). However, it is sought in the same way to develop some talks in the official language of the host. Table 1 clearly shows that most of the countries in which it has been developed are Spanish speaking. However, in 2018, for the SC-Camp organized in Vietnam, some sessions were held in French and dual Vietnamese English in partnership with the local collaborators.

3.1.3. Instructors and Keynotes

Instructors and keynote speakers are researchers, scientists, or engineers with important expertise in a subject related to HPC or its applications. They are, of course, part of the Advanced Computing Ecosystems covering the roles of the HPC Technical and Operations activities, Research Software Engineers, and HPC Scientists. We have already said that depending on the academic interest, specific activities are carried out both with guest instructors and lecturers. Typically, and as will be detailed later in Figure 4, the technical program of SC-Camp covers six thematic days that require at least three specific skills coordinated by at least four core instructors, usually members of the SC-Camp association. If well, there is a general academic chair by SC-Camp edition, for each day, there exists a leader, assumed by one core instructor, and the other instructors support the activities of the day, mainly the hands-on sessions. Then, in each camping school, the core team is joined by two specialized instructors and two guest lecturers. At this point, it is important to say that the guest lecturers are specialists (usually local or regional scientists) who fulfill an inspiring mission toward the students by showing the contributions of their work using HPC.

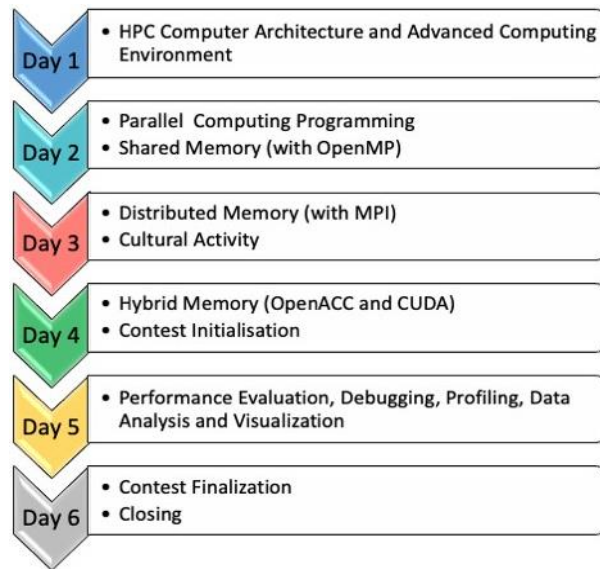


Figure 4. SC-Camp Scheduling

The core instructors guarantee good communication with the attendees and generate a trusting relationship with the students. At the same time, as specialists who develop their work in the competitions given in the SC-Camp, they not only facilitate the interaction and training processes but also the resolution of questions.

3.1.4. Development

SC-Camp has six days of scientific sessions, with several programming practical sessions and a final collaborative project as a programming contest by teams of students. The schedule is organized by days in which, following specific topics and all participants from the three HPC roles are defined in Section 2.1. participate in acquiring skills according to their role.

Usually, all the participants (instructors, keynotes, students) stay in the same place during all the SC-Camp days. The proposed initial program is usually adapted to the level of the students and the local constraints. Additionally, the daily work schedule is completed with continuous interactions, homework, and recreative activities, during which the participants work, discuss, and socialize together.

Figure 4 shows a typical program proposed during the development of SC-Camp. Each day is a space of training, starting on Day 1 with a theoretical and practical day introducing the camping school, the HPC computer architecture features (parallelism from the infrastructure point of view), and the advanced computing environment (including aspects of interaction and connection such as batch scheduling, scripting, general monitoring, etc.).

Days 2 and 3 introduce the good practices for scientific computing programming before focusing on the shared and distributed memory paradigms (with OpenMP [18] and MPI [19], [20]). Day 4 to deal with hybrid memory management, OpenACC [21], and CUDA [22]. Day 5 aims to develop monitoring, analysis, and performance evaluation of an HPC environment and debugging, profiling, and optimizing applications to achieve good performance, as well as the use of specific compilers, interpreters, and batch/scripting techniques. Finally, Day 6 is proposed to close the SC-Camp, including a co-evaluation of the participants (students, instructors, keynotes, committee, and guests) and the conclusion of the collaborative project.

There are some special features in the schedule of the SCC-amp: Day 3 is usually used to promote cultural activities to generate social exchange and non-technical interaction. On Day 4 starts, the collaborative project, which usually takes the form of a parallel programming contest. Finally, on the last day, the contest is finished with a discussion of proposed solutions and best results. We have sought to preserve a similar timeline in all versions (and the day schedule), including the virtual versions made during the pandemic and the SC-Camp 2022 version in hybrid mode.

SC-Camp is organized by day with a generic structure illustrated in Figure 5. It balances the theoretical lectures and practical sessions to hold the students' attention and guarantee the interactions and coverage of the subjects of the day.

Figure 5 shows how sessions are placed and organized in activities performed at different times of the day with a maximal duration. Each SC-Camp Day has an instructor responsible (or a couple of them) for all day. This main instructor (in blue on the Figure) has expertise in the technical subject of such a day and conducts three of the five theoretical sessions and all the hands-on activities. Each journey starts with an opening session in which the different activities are introduced. As each instructor is equally an expert in the area, the primary instructor presents a technical talk that motivates the knowledge acquisition of the day based on their activities and experiences as a researcher, professor, or engineer. After this introduction, a scientific keynote will take place by a guest with the competencies described above (in green in Figure 5). Then, the instructor gives the theoretical session explaining the day's theoretical content. The hands-on sessions are typically divided into two sessions and are conducted by the main instructor and an assistant instructor (in white in the Figure).

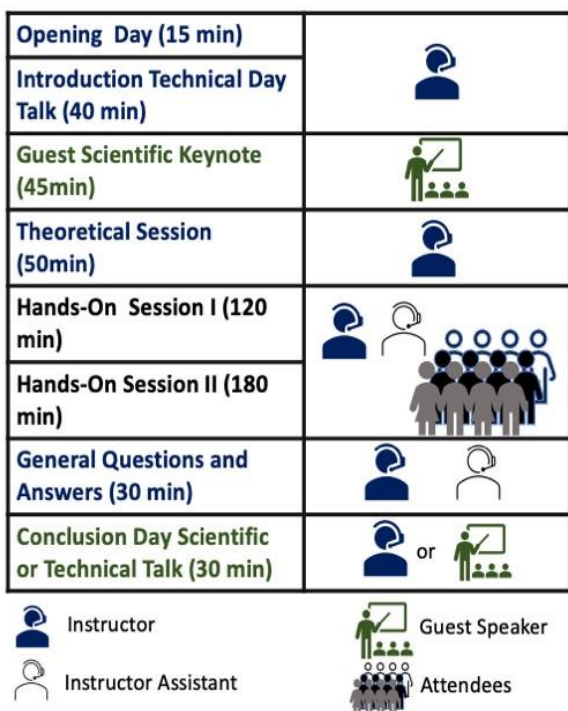


Figure 5. Structure of a SC-Camp Day.

The assistant is just another instructor with technical skills in the subjects of the day, and generally, it is one of the other instructors of the week. At the same time, the attendees participate actively in the hands-on sessions, sharing experiences, collaborating with their peers, and contributing to the technical discussion, questions, and answers. Then, the attendees actively participate in the session's development.

At the end of the day, they are two important activities: the general questions and answers session led by the instructor and the instructor assistant, where different questions are proposed by the attendees (but also for the instructors to assess the degree of learning and collect feedback). A final scientific or technical lecture given by one of the instructors or a guest speaker, following the topics of interest of that year (Bioinformatics, Aerospace, Finance, Computer Architecture, etc.). When sponsors are of special interest in the subjects of the SC-Camp, they sometimes participate in a technical talk.

Although breaks and mealtimes are not shown in Figure 5, there are 15- to 20-minute breaks to guarantee friendly interaction. On the other hand, there is permanent "coffee" during the hands-on because the practical sessions are designed to be continuous and collaborative work. This time is considered equally

important and allows you to get to know all the participants better.

Finally, the teaching activities of SC-Camp are completed with online materials provided to the students before each SC-Camp. These materials contain the slides of the presentations, the source code of examples, the instructions for the practical sessions, etc. These materials are updated progressively with the school's development to include new lectures and provide the solution to the exercises after corrections. Example materials for SC-Camp 2019 are available at [23]. Likewise, all material from all versions of the SC-Camp remains online and open to the public; continuous consultation by alumni has been monitored.

3.2. Actors Enrolment and Skills Spreading

Every year, SC-Camp aims to improve the skills of the attendees in accordance with the host institution's interests. Once the hosting site is selected, the local and general committees work together on a tailored technical program and select adequate instructors. They also invite guest speakers from the local scientific community to give keynotes and technical talks. The number of attendees is established and motivated by the host but should be in line with the school's objective. It affects the number of instructors necessary to develop good training.

Concerning the student attendees, the committees organize a selection process. The students can apply online and are selected by the local committee according to pre-defined criteria: adequate English proficiency, technical or scientific background in conformity with the goals of the host institution, educational level according to the host profile, and motivation (using a letter of motivation joined to the application). Other aspects as ethnicity, minority, or gender representation, are guaranteed by the local organizing committee and overviewed by the general committee. Furthermore, the general committee recommends that the local organizing committee achieve a minimum percentage of the three main actors mentioned in Section 2.1. For example, taking the nomenclature used in Figure 2., if an SC-Camp has major participation of HPCRSE and HPCUSE attendees, the general committee recommends at least 10% of attendees with an HPCTOE profile. For Master students whose role is not clearly defined, their profile is inferred by the committee based on their current study program and the content of their motivation letter.

As presented before, the instructors and keynote speakers are scientists, engineers, or experts with a deep experience in the HPC and Advanced Computing communities also observing communication abilities and

they are from different countries. Instructors are selected from a pool by the general committee with the agreement of the local committee. The pool of instructors is also approved by the steering committee, and all of them are volunteers. As said before, most part of the instructors are part of the SC-Camp Association, a non-profit association registered in Lyon, France.

As well as for the attendees, each SC-Camp seeks a balance in the expertise of instructors. Also, these aspects are considering selecting and inviting the keynotes, further diversity, ethnicity, gender, and professional background. Likewise, local speakers' participation is promoted to highlight the value of local experts and motivate local students.

With these participants: student attendees, instructors, and keynote speakers, the skills spreading are proposed following integrated spaces of development shown in [Figure 6](#). Each triangle corresponds to a specific skill development space. The idea of spaces of development is inherited from the Software Engineering community emphasizing good practices as a central element. Indeed, the main object of knowledge is the development of parallel computing codes, which implies specific design and implementation spaces. These spaces are discussed in detail in [\[24\]](#), and the interest is the integration of competencies and the mandatory interaction between all the actors represented in [Figure 6](#) by specific icons.

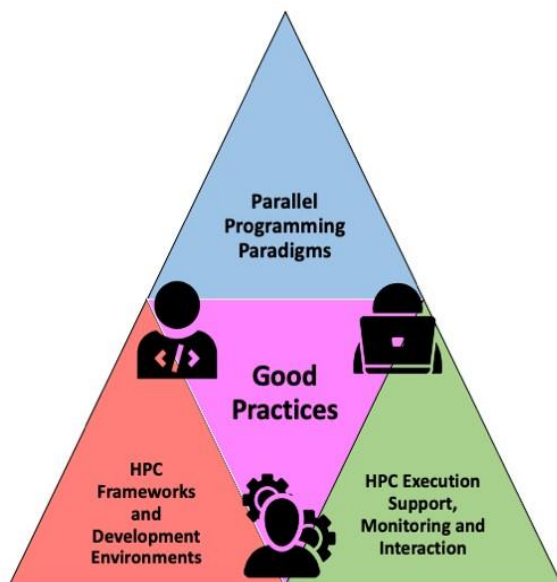


Figure 6. Skills Development Spaces. The three main actors are represented by icons clockwise from the top left: HPCRSE, HPCUSE, and HPCTOE.

Depending on the role and interests of each one of the actors, the depth of knowledge sought in each development space varies, but the relationship with the others is mandatory. For example, if we consider the scheduling, execution, and deployment of application environments (or application themselves) on a cluster, the three role share common characteristics: reservation instructions, launch commands, etc. However, how to install or attach the scheduler to the infrastructure is more related to the role of HPC execution support. Similarly, how to perform efficient deployment images or efficiently exchange data between a collection of applications is tied to the role of an HPC developer.

3.3. Collaborative Project and Co-Evaluation

A collaborative project is a project or contest carried out by teams of students, organized, and supervised by the instructors. In this project, the students must solve a problem using the knowledge acquired during the SC-Camp. It is presented at the end of the fourth day, and students work on it during the remaining days of school, progressively integrating the new elements learned during the following sessions. For example, they incorporate aspects related to debugging and profiling, optimization, data analytics, or data visualization. Teams can interact with each other, and the main value is the collaborative work and the innovative approaches proposed by the students to achieve better solutions to the given problem. During the whole time, teams interact with instructors and guests to achieve the best possible solution in the given time.

On the last day, each project is evaluated by the instructors and reviewed by another team of students. The last day is the occasion to give a detailed project correction and present a typical HPC solution. Additional care is taken to discuss some of the solutions proposed by the students, highlight the common mistakes and comment on the best approaches.

4. The Virtual SC-CAMP

Without a doubt, the years 2020 and 2021 represented a transformation of different human activities due to the COVID-19 pandemic. One of them was the constraints induced by the face-to-face training, which directly impacted the organization of SC-Camp. Despite the fear and general pessimism in the first part of COVID-19 due to disease reports and the impact on human life [\[25\]](#), a large part of the population felt the need to acquire knowledge in science and technology. Education was (and still is) one of the best ways to maintain social contact with (physically) isolated persons. In the case of the academic community, the interest of students from

different parts of the world in contributing something useful increased their curiosity in topics related to HPC, bioinformatics, and data analytics following different reports [26]. On the other hand, the coronavirus pandemic forces digital acceleration, increasing access to Internet connections and digital resources significantly. In the case of training, the offer of remote activities increased enormously.

Nowadays, the SC-Camp organization foresees where the camping school will be held for the next two years. However, in the first semester of 2020, the SC-Camp committee decided to implement the SC-Camp in a virtual mode, so-calling it #SCCAMPV (Supercomputing and Distributed Systems Camping School in Virtual). The virtual mode of the 2020 and 2021 versions allowed the development of activities over two weeks, in live theoretical sessions of 30 to 45 minutes, and personal practical activities with the possibility to interact all day via Slack [27] and Discord [28]. Also, several scientific talks were scheduled and recorded so that attendees and the public could see them at any time. The collaborative project was performed by a single overall team in constant (online) interaction.

To develop these # SCCAMPVs, we joined efforts with institutional allies. In 2020, #SCCAMPV took place with the support of the Colombian national advanced technology network, RENATA [29], and in 2021 with the support of the Latin American consortium of advanced networks, RedCLARA [30], which provided the infrastructure (in particular video conference platforms and collaborative tools) for the events. These two organizations are responsible for the academic networks. RENATA in a national scope in Colombia and RedCLARA in an international consortium of NRENs (National Research and Education Networks) in Latin America and the Caribbean. The main objective of 2020 SC-Camp-V was to integrate the different actors of the nascent RedCCA (Colombian Advanced Computing Network), and for the 2021 SCCAMPV to promote the integration of new users to the services and center joined by SCALAC (Advanced Computing Services for Latin America and the Caribbean), despite the possibility of extending participation to hundreds of people, same that for the previous versions, the student was selected by the steering committee to ensure follow-up and participation in future joint projects.

The experience of these two virtual events led to a reflection on the development of SC-Camp in a post-pandemic context, which is still in progress by the organizing committee at the time of writing this paper. The economic impact, carbon footprint, coverage, and the need for human interaction, digital transformation,

and other questions arise and evolve considering the situation.

5. The Hybrid SC-CAMP

The COVID-19 pandemic forced new ways to implement the activities of the SC-Camp. We presented how the #SCCAMPV was performed during the 2020 and 2021 years, contributing to the digital acceleration and maintaining the initial spirit and objectives of the camping school.

After consideration of issues related to the economic and environmental impact, the inclusion and participation of remote students using the Internet, the SC-Camp committee proposes to organize test versions of a Hybrid SC-Camp (called #SCCAMPH). In 2022, the organizing committee developed an in-situ training program in two selected sites for one day for six weeks, but with the possibility of participating online remotely. Results of 2022 are analyzed, and the hybrid mode comes with additional challenges, such as how to get online and physical participants to interact. The experience in 2022 was interesting, managing it in two physical sites, the High Performance and Scientific Computing Center (SC3UIS) at Universidad Industrial de Santander in Colombia and the Advanced Data Science Center (CADS) at Universidad de Guadalajara in Mexico, for in-situ participation. RedCLARA provided the remote support, and SCALAC granted the HPC provision. As an exercise, there were 6 and 12 participants in each site, respectively, and remotely, 22 attendees linked using RedCLARA provided connection. The particularity of this exercise was that the on-site participants of each one of the places where remote at one time. The #SCCAMPH starts in Bucaramanga, Colombia, in the first days and finishes in Guadalajara, México. The instructors were on-site and remote with minimal support of one (1) instructor assistant by site given by each institution.

The hybrid version of the SC-Camp contributes to guaranteeing cost reduction and the widest participation of participants. However, the coordination of two sites and the idea of maintaining a common activity and constant interaction cannot be handled in a virtual or hybrid way.

6. Conclusion And Further Work

In this paper, we presented how we identified different roles in an advanced computing ecosystem and described how common and specific skills are distributed among the different actors of HPC. The interaction between them and common actions following good practices and communication are important to perform better HPC

development, seek to build better and sustainable applications, optimize resource use, and achieve a big impact in the community.

Following these aspects and strongly aimed at democratizing knowledge, we develop the Supercomputing and Distributed System Camping School, SC-Camp, to create common teaching spaces and leverage interaction, enhance technical and scientific multidisciplinary competencies, and promote the appropriation of knowledge. In effect, the most notable scope of the SC-Camp has been the increase in collaboration between actors, even from different places and knowledge areas. Equally important is the impulse to appropriate knowledge by the participants. In practice, the follow-up of the alumni was not developed in this paper (and is difficult to organize). However, from their own initiative, the students have created groups in different social networks and continue to interact (i.e., it is possible to follow the hashtag #SCCAMP or #SCCAMPStudents).

Funding acquisition

Deeply appreciate the contributions of sponsors, companies, governmental and non-governmental entities, and associations, without which it would not be impossible to sustain the activity, as is the case of the RedCLARA, SCALAC, the Academic French Diplomacy Bureau, the RISC-2 Project (www.risc2-project.eu), among many others that can be consulted on the site www.sc-camp.org. We must highlight the special technological support of two institutions: the Universidad Industrial de Santander in Bucaramanga, Colombia, which via SC3UIS, provides HPC access platforms for all the practices and operations support, and the technical provision of the University of Luxembourg and its HPC team for adding the experience of qualified instructors and teaching assistants.

Author Contributions

C. J. Barrios: Conceptualization, Formal Analysis, Investigation, Writing – review & editing. R. Rivas: Conceptualization, Investigation, Methodology, Validation, Writing – review & editing. X. Bessonon: Conceptualization, Investigation, Methodology, Validation, Writing – review & editing. G. Diaz: Conceptualization, Formal Analysis, Investigation, Writing – review & editing. Y. Georgiou: Conceptualization, Investigation, Methodology, Validation, Writing – review & editing. P. Velho: Conceptualization, Investigation, Methodology, Validation, Writing – review & editing.

All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Acknowledgements

The authors thank all the countries, institutions, and centers that hosted the Supercomputing and Distributed Systems Camping School, SC-Camp, in its different versions, instructors, guest speakers, local organizers, and the distinguished members of the steering committee who have voluntarily participated in the SC-Camp.

Finally, we thank all the SC-Camp students (now SC-Camp alumni), instructors, keynote guests, and volunteers from whom much more has been learned.

References

- [1] HPC Certification Forum. [Online]. Available: <https://www.hpc-certification.org/>.
- [2] ACM SIGHPC Education. [Online]. Available: <https://sighpceducation.hosting.acm.org/>
- [3] I. Foster, Carl Kesselman, *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998.
- [4] International Summer School in Grid Computing (ISSGC). [Online]. Available: <https://www.dma.unina.it/~murli/SummerSchool/>.
- [5] M. Atkinson, “International Collaboration to Extend and Advance Grid Education (ICEAGE)”. [Online]. Available: <https://www.research.ed.ac.uk/en/projects/international-collaboration-to-extend-and-advance-grid-edu>.

- [6] Enabling Grids for E-science project (EGEE). [Online]. Available: <https://eu-egee-org.web.cern.ch/>.
- [7] International Summer School on HPC Challenges in Computational Sciences (IHPCSS). [Online]. Available: <https://www.ihpcss.org/>.
- [8] International Supercomputing and Distributed Systems Camping School (SC-Camp). [Online]. Available: <https://www.SC-Camp.org>.
- [9] National Research Council, *Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering 2017-2020*. 2016, doi: <https://doi.org/10.17226/18972>
- [10] A. Brett, M. Croucher, R. Haines, S. Hettrick, J. Hetherington, M. Stillwell, C. Wyatt, “Research software engineers: State of the nation report 2017”, 2017. [Online]. Available: <https://zenodo.org/record/495360>, doi: <https://doi.org/10.5281/ZENODO.495360>
- [11] N. B. Ruparelia, “Software development lifecycle models,” *ACM SIGSOFT Software Engineering Notes*, vol. 35, no. 3, pp. 8–13, 2010, doi: <https://doi.org/10.1145/1764810.1764814>
- [12] J. Cohen, D. S. Katz, M. Barker, N. C. Hong, R. Haines, C. Jay, “The four pillars of research software engineering,” *IEEE Software*, 38(1):97–105, 2021, doi: <https://doi.org/10.1109/ms.2020.2973362>
- [13] Universidad Industrial de Santander (UIS). [Online]. Available: <https://www.uis.edu.co>.
- [14] Supercomputación y Cálculo Científico at Universidad Industrial de Santander (SC3UIS). [Online]. Available: <http://www.sc3.uis.edu.co>.
- [15] S. Varrette, H. Cartiaux, S. Peter, E. Kieffer, T. Valette, A. Olloh, “Management of an Academic HPC & Research Computing Facility: The ULHPC Experience 2.0.,” In *6th High-Performance Computing and Cluster Technologies Conference*, 2022, doi: <https://doi.org/10.1145/3560442.3560445>
- [16] High-Performance Unit at the University of Luxembourg. [Online]. Available: <https://hpc.uni.lu/>.
- [17] Advanced Computing System For Latin America and the Caribbean (SCALAC). [Online]. Available: <https://scalac.redclara.net>
- [18] OpenMP API Specification Version 5.1. [Online]. Available: <https://www.openmp.org/spec-html/5.1/openmp.html>.
- [19] MPI Forum. [Online]. Available: <https://www.mpi-forum.org/>.
- [20] L. Clarke, I. Glendinning, R. Hempel, “The MPI Message Passing Interface Standard,” In *Programming Environments for Massively Parallel Distributed Systems*, pages 213–218, 1994, doi: https://doi.org/10.1007/978-3-0348-8534-8_21
- [21] OpenACC. [Online]. Available: <https://www.openacc.org/>.
- [22] NVIDIA CUDA. [Online]. Available: <https://developer.nvidia.com/cuda-toolkit>.
- [23] Teaching Materials for SC-Camp 2019. [Online]. Available: <https://gitlab.uni.lu/SC-Camp/2019>.
- [24] T. Mattson, B. Sanders, B. Massingill. *Patterns for Parallel Programming*. Addison-Wesley Professional, 2004.
- [25] World Health Organization (WHO), Coronavirus disease (COVID-19) Situation Reports. [Online]. Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.
- [26] Organization for Economic Co-operation and Development (OECD). An assessment of the impact of COVID-19 on job and skills demand using online job vacancy data, 2021. [Online]. Available: <https://bit.ly/3IP1h82>
- [27] Slack. [Online]. Available: <https://slack.com>.
- [28] Discord. [Online]. Available: <https://discord.com/>.
- [29] Red Nacional Académica de Tecnología Avanzada (RENATA). [Online]. Available: <https://www.renata.edu.co/>.
- [30] Consorcio Latinoamericano de Redes Académicas Avanzadas (Red-CLARA). [Online]. Available: <https://www.redclara.net/>.



Available in:

<https://www.redalyc.org/articulo.oa?id=553781612014>

How to cite

Complete issue

More information about this article

Journal's webpage in redalyc.org

Scientific Information System Redalyc
Diamond Open Access scientific journal network
Non-commercial open infrastructure owned by academia

Carlos J. Barrios H., Robinson Rivas, Xavier Bessonon,
Gilberto Diaz, Yiannis Georgiou, Pedro Velho
**High-Performance Computing Training by Skills for
Advanced Computing Ecosystems**
**Entrenamiento por Competencias de Computación de
Alto Rendimiento para Ecosistemas de Computación
Avanzada**

Revista UIS ingenierías
vol. 23, no. 1, p. 175 - 188, 2024
Universidad Industrial de Santander,
ISSN: 1657-4583
ISSN-E: 2145-8456

DOI: <https://doi.org/10.18273/revuin.v23n1-2024014>