Expanding Horizons: Advancing HPC Education in Colombia through CyberColombia's Summer Schools

Aurelio Vivas
Universidad de los Andes
Argonne National Laboratory
aa.vivas@uniandes.edu.co

Carlos E Alvarez
Tecnológico de Monterrey
carlose.alvarez@tec.mx

Jose M Monsalve Diaz
Argonne National Laboratory
jmonsalvediaz@anl.gov

Esteban Hernandez
CyberColombia
esteban.hernandez@cybercolombia.org

Juan G. Lalinde-Pulido
Universidad EAFIT
jlalinde@eafit.edu.co

Harold Castro
Universidad de los Andes
hcastro@uniandes.edu.co

ABSTRACT

High-performance computing (HPC) is an important tool for research, development, and the industry. Moreover, with the recent expansion of machine learning applications, the need for HPC is increasing even further. However, in developing countries with limited access to the HPC ecosystem, the lack of infrastructure, expertise, and access to knowledge represents a major obstacle to the expansion of HPC. Under these constraints, the adoption of HPC by communities presents several challenges. The HPC Summer Schools are an initiative of CyberColombia that has taken place over the past 5 years. It aims to develop the critical skills, strategic planning, and networking required to make available, disseminate, and maintain the knowledge of high-performance computing and its applications in Colombia. Here we report the results of this series of Summer Schools. The events have proven to be successful, with over 200 participants from more than 20 institutions. Participants span different levels of expertise, including undergraduate and graduate students as well as professionals. We also describe successful use cases for HPC cloud solutions, namely Chameleon Cloud.

1 INTRODUCTION

Supercomputers are of paramount importance in solving critical challenges in many fields. Some of these are atmospheric simulation, genome sequencing, and cybersecurity, to name a few [15]. In Colombia, this area is developing at a slower pace in comparison to leading countries such as the USA, Japan, and European countries. Colombia only spent 0.29% of its Gross domestic product (GDP) on Research and Development according to the UNESCO statistics [23]. This level of investment is 0.92% behind South American countries such as Brazil, which is one of the leading Latin American countries in supercomputing technology. Differences in investment levels on the development of supercomputers [10] and adoption of HPC are some of the factors that heavily influence this gap. On the other hand, Colombia is facing technological, economic, and educational challenges despite the efforts made by universities, and public and private institutions to promote the development of this area. However, initiatives that provide access to HPC resources such as SCALAC [12], RedCLARA [12], and HPC Americas Collaboration [9] still thrive.

Education in HPC is not only a critical concern in Colombia, but also in supercomputing leading countries [10, 19]. There is a gap between the speed at which new technologies are developed compared to the speed at which they are adopted. The HPC Summer Schools presented in this paper are an initiative of CyberColombia[1] [16], that aims to develop critical skills in HPC, strategic planning for HPC centers, and provide networking opportunities for the national community. We focus on these objectives as we consider them to increase the availability of knowledge of high-performance computing and its applications in Colombia. We also make a great effort in the establishment of collaborations and virtual organizations around common practices, tools, and data usage. Learning from the experience of already developed HPC communities worldwide, we emphasize multi-institutional collaborations, considering actors from industry, academia, and the international community.

The upcoming sections are organized as follows: Section 2 presents the main Colombian High-Performance Computing education and training event, the High-Performance Computing Summer School (HPCSS). Here we briefly describe (i) the main goals of this initiative, (ii) how the initiative has managed to persist during the pre-pandemic, pandemic, and post-pandemic periods, and (iii) what makes our initiative different to already existing ones. Section 3 describes (i) the timeline of events and decisions consolidating the HPCSS and statistics depicting the progression of the initiative. The latest version of the initiative, HPCSS 2023, which we consider the most successful, is detailed in Section 4. This work concludes in Section 5, where learned lessons and future directions are discussed.

2 SUMMER SCHOOLS

The High-Performance Computing Summer School (HPCSS) initiative was primarily motivated by already established non-local HPC education and training programs such as the Supercomputing Camp (SC-Camp) [22] and the Argonne Training Program on Extreme-scale Computing (ATPESC) [7]. Nevertheless, we have created a

https://doi.org/10.22369/issn.2153-4136/15/1/4

[1] https://cybercolombia.org

program that closely matches the country's current ecosystem and difficulties and it is tailored to resolve them. Over time, we have adopted methodologies, such as the multi-site and virtual delivered training, established on scientific environments like XSEDE [24] (nowadays ACCESS [6]) to enhance the effectiveness and reach of the Summer Schools program.

2.1 Objectives and Differences

Our summer schools have mainly two objectives: (i) bring awareness in Colombia of the broad array of research and career opportunities in areas supported by HPC, and (ii) promote cooperation between **academia**, **industry**, and **international entities**. In a nutshell, our program seeks to generate HPC expertise needed in Colombia [14] while broadening the community.

There are currently several HPC initiatives worldwide that target HPC education. The National Science Foundation in the United States of America created the Engineer Discovery Environment (XSEDE) [24]. This program ran from 2011 to 2022, and its focus was on sharing education and infrastructure for advanced services such as supercomputers. XSEDE offered a variety of online lectures and other training distributed across XSEDE sites. This program was then replaced in 2022 by the Advanced Cyberinfrastructure Coordination Ecosystem: Services and Support (ACCESS) [6], which currently has a similar idea, but it has been focused on increasing HPC resources and community development. Despite the success of these programs, they have been created mostly to tackle the needs of the community in the United States. Likewise, the US Department of Energy's leadership computing facilities (e.g., NERSC [4], ALCF [1], OLCF [18], etc.) run a series of training programs that include long training sessions (e.g., ATPESC). However, these programs are made to increase the knowledge of already existing users of their facilities, as well as encourage users with applications that are ready to be ported to these machines to apply for allocations. Unfortunately, these criteria create a really high bar for many researchers in Colombia who are just learning to use these systems.

Worldwide, there are also other supercomputing training programs. The International SuperComputing Camp [22], a non-profit event for training HPC, is held annually in different parts of the world. The International HPC Summer School [3], sponsored by different organizations around the world, has been an excellent option since 2010. More recently, the Latin American Introductory School on Parallel Programming and Parallel Architecture for High-Performance Computing organized by the International Center for Theoretical Physics (ICTP), focuses on parallel and distributed computing for scientists. Additionally, a large number of initiatives across multiple other countries and institutions [2, 5].

Despite the large number of options currently in the market, our program has a fundamental difference that makes it valuable. First, most of the already existing programs are created for those scientists who already have a need for HPC. Many of these participants have already been exposed to the idea of High-Performance Computing and are looking to expand their knowledge. As previously mentioned, the low investment in HPC in Colombia has resulted in limited exposure to the area of HPC to the large majority of scientists, students, and professionals. Furthermore, there is a lack of a strong HPC community in the country that does not have the necessary grounds to grow. Our focus is to tackle these problems. Our low entry bar allows many to participate.

Furthermore, we strongly encourage the interaction of members from different institutions. Thus, helping to increase the visibility and sense of community for these individuals, providing an experience that they would not have in their own institutions. The emphasis on networking aims to increase connections across institutions and individuals that could result in longer-term collaborations. In general, we see ourselves as facilitators of HPC in Colombia, and we expect that our participants can, in the future, take part in more advanced programs as those mentioned earlier.

2.2 Summer School Evolution

The annual HPCSS series program started in 2018. It represents the main high-performance computing informative and training event organized and delivered by CyberColombia. Our organization works in collaboration with local universities serving as hosts for the events and national (Colombian) and international institutions such as Universidad de los Andes (academia), Argonne National Laboratory (research), and NVIDIA (industry) participating as speakers or mentors.

The lockdown generated as a response to the SARS-CoV-2 pandemic posed great challenges to the development of activities of many organizations, and CyberColombia was not an exception. However, it also opened the opportunity to test communication technologies that we had not considered up to that moment and expand our international collaborations. For this reason, we will divide the development of the HPCSS into 3 consecutive periods of time: From 2018-2019 (pre-pandemic), from 2020-2022 (pandemic), and from 2023 onward (post-pandemic).

During the pre-pandemic period, the summer schools consisted of two main parts: (i) informative talks given by international speakers on various HPC-related topics, and (ii) practical workshops supervised by both professors from academic institutions and outreach staff from research institutions or industry. These first iterations of the summer school were 5-day events. Each day, we started with a keynote speaker who would help to motivate the materials presented during the day. Following, a series of hands-on tutorials were conducted. The topics of the tutorials were mostly introductory rather than driven by particular use cases. Among some of the topics that were covered are: introduction to HPC infrastructure, introduction to C and C++, introduction to job scheduling technologies (e.g., Torque, Slurm, PBS, etc.), introduction to parallel programming (e.g., OpenMP, OpenACC, Pthreads, etc.), introduction to accelerators programming (e.g., CUDA, OpenACC), and introduction to distributed Programming (e.g., MPI).

In the pandemic period, the events were switched from an in-person modality to a fully virtual experience. However, this transition did not come free of challenges when switching the talks and tutorials to a virtual environment. First, selecting a proper video conferencing software. Not only there were personal preferences in the attendees and speakers, but also different organizations and participating institutions had their own set of rules and preferences with respect to this matter. As a result, different sessions required us to switch from one videoconferencing software to another. Additionally, we also noticed that it was more difficult to keep the audience engaged in a virtual environment. Consequently, tutorials had lower interaction and participation from the attendees, diminishing the learning process. Another challenge brought by a virtual event was a reduction of networking opportunities between attendees. In-person sessions force participants to be face-to-face during several parts of the event. In contrast, virtual attendees do not require an active engagement, thus reducing the incentive to discuss ideas in a casual way or to engage in random conversations. Despite our efforts, most participants did not have the initiative to participate. Finally, virtual sessions proved to be more exhausting for participants and speakers. Maintaining the same position in front of a device considerably increased fatigue.

However, the virtual settings opened certain opportunities that were not possible before. The lack of a venue, catering, and travel expenses for speakers makes virtual meetings less costly to organize. For participants, it is also easier to attend, as the time and costs of traveling and accommodation are no longer present. These factors were reflected in lower registration fees, a factor that was particularly important for attendees with low-income levels in the country. As a result, the 2020, 2021, and 2022 versions of the HPCSS featured an increased level of international participation from both attendees and speakers/mentors. Furthermore, the scholarships, usually offered to a few participants who could not afford the registration, now were available for a larger number of people. Additionally, under normal circumstances, high-impact researchers and professionals would have a harder time scheduling a multi-day trip to the country. The lower time and personal commitment of speakers increased the chance to secure more impactful content. Finally, prior to the pandemic, we used different university venues that offered their facilities for our event. However, this reduced the visibility of CyberColombia as the main organizing partner. It also meant that other institutions would be more cautious about advertisement and participation. By removing dependencies on physical venues, CyberColombia, and our initiatives, were decentralized, allowing us to be a more neutral actor in the country and further increasing our collaborations with a wider range of institutions.

During the pandemic, the 5-day format was maintained, but the topics of the talks shifted from HPC core concepts to its applications. From this point on summer schools were more application oriented. We maintained the HPC focus, but, by considering a primary topic and a hot or emerging topic, we increased participation from domain-specific sectors. For the 2020 version, the main topic was HPC in Data Science and Artificial Intelligence, while the emerging topic was HPC against COVID. In 2021, a substantial amount of talks and tutorials were conducted in Convergence HPC, IA, and Big Data; but some of them were treated as emerging topics such as Quantum Computing. The 2022 version was concentrated on Space exploration and High-performance computing. While most of the tutorials during this period remained focused on the basics of HPC, new topics like the use of Matlab for HPC and quantum computing were also present.

Lastly, the post-pandemic era and the latest edition of the HPCSS was held in 2023. It will be addressed in more detail in Section 4.

3 SUMMER SCHOOLS IN NUMBERS

The progression of CyberColombia events over time has shaped what nowadays is called the HPCSS. The commitment to introduce Colombians to the world of HPC started in 2016 with a workshop held at Universidad Distrital de Colombia. In 2017, Universidad Distrital de Colombia welcomed students from multiple universities, fostering a diverse, inclusive, and cooperative learning environment. In 2018, the initiative was formally constituted as the HPCSS series, with the first edition taking place at Universidad de los Andes.

Also in 2017 the Earlham Institute, Colciencias, BRIDGE Colombia, and GROW Colombia launched the C3 Biodiversidad (C3) initiative; “aiming to promote a research cyberinfrastructure in Colombia for the analysis of the natural and agricultural biodiversity” [13]. In 2019, C3, Grow Colombia, and organizers of early versions of the Summer School series joined efforts to create CyberColombia. The result was a more formalized version of the HPCSS program that helped us tackle challenges in data-intensive science in Colombia from critical research areas including computational biology [16] [14]. From 2020, the HPCSS turned into a virtual event as explained in section 2. The following sections present statistics characterizing the estimated impact of the above-mentioned events in the progression of the HPCSS over time. Metrics on participation in attendance and mentoring, as well as some demographic data through the years, are described.

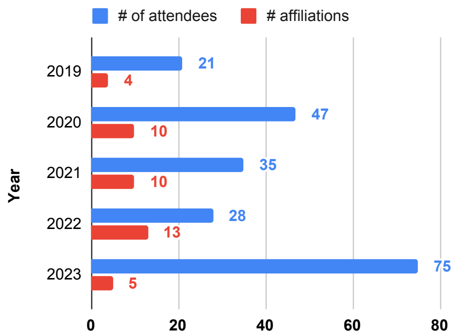
3.1 Participation Over Time

Figure 1 illustrates attendance and participation of speakers/mentors over time, the number of participating organizations (affiliations) in every case is also depicted. Note that in 2019, the event took place in its traditional, in-person format, whereas in 2020-2021 (pre-pandemic) and beyond (starting from 2022, post-pandemic), it transitioned to a virtual format.

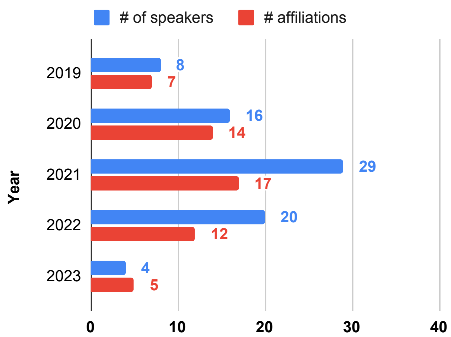
As seen in Figure 1a, the beginning of virtual experiences in 2020, encouraged attendance at a wide variety of events given the ease of access, savings in transportation times, and cost reductions. Nevertheless, the transition to the virtual experience had three associated drawbacks, impacting participation during periods 2021 and 2022. First, we lost the student-university interaction. Second, the large exhaustion of virtual meetings among participants discouraged people from taking part in these events. Third, our focus on specialized topics forced us to open new doors thus limiting our ability to reach out to a larger number of participants.

Despite receiving virtual support from the hosting university, challenges persisted due to students not being in an optimal learning environment and the inability to effectively gauge students’ emotions and progress during talks and tutorials.

More so, the full-day week-long program turned out to be difficult for participants. We noticed a high dropout rate in the last days of training, especially during the 2021 and 2022 versions. We attributed the high desertion rate to the considerable burden of commitments acquired by the participants given the virtual experience and the psychological effects generated by the confinement situation during the pandemic. The above-mentioned situation led us to reformulate our program from one week to a three days comprised format and maintain the virtual experience for the latest version of the event, HPCSS 2023.



(a) Number of attendees over time



(b) Number of speakers over time

Figure 1: Participation over time

Figure 1b shows the number of speakers and speakers' affiliations. The variation across versions can be attributed to different factors. The virtual experience also encouraged speakers' and mentors' participation and promoted a larger diversity of academia, industry, and public and private institutions. Moreover, the change in focus area across the different versions increased or reduced the need for different speakers as well as our ability to recruit impactful ones.

Virtual events also provided a better opportunity to include more speakers and institutions, as mentioned in section 2. However, this virtual-only mode also significantly reduced the speakers-attendees interaction, finally reducing the impact of their participation in the attendee's professional development and networking opportunities.

We made significant progress on strengthening the speakers-attendees relationship on the last version of the event, HPCSS 2023 which is detailed in Section 4.

3.2 Participation Over Time per Sector

Tables 1 and 2 illustrate the number of HPCSS attendees and the number of institutions participating with speakers and/or mentors for tutorials. Note that both the total number of attendees and participating institutions over time are disaggregated across sectors. *Academia* includes universities; *research* includes research laboratories and institutions; *industry* includes national and multi-national companies; and *other* includes other public, private, and non-profit organizations.

Table 1: Attendance per Sector Overtime

Year\Sector	# academia	# research	# industry	# other
2019	14	-	-	7
2020	30	5	-	12
2021	21	-	1	13
2022	26	1	1	-
2023	75	-	-	-

Table 2: Institutions per Sector Overtime

Year\Sector	# academia	# research	# industry	# other
2019	6	-	1	-
2020	5	3	6	-
2021	5	4	7	1
2022	2	4	6	
2023	5	-	-	-

When comparing Tables 1 and 2, it is worth noting the difference in participation across sectors in attendance against institutions giving talks or mentoring tutorials. While institutions across sectors are more diverse, i.e., the attendance force is highly concentrated in academia.

Although our future goal is to make a more diverse attendance across sectors, academia is where we have found potential for the HPCSS program to make a meaningful and far-reaching impact. This is because academia is where future research and industry workforce develop fundamental technical and critical skills, in the realm of a learning environment, to face more specific, and complex problems in research laboratories and the industry.

Finally, the diversity in participating institutions has served two vital purposes: firstly, it has reaffirmed the sector's commitment to developing education in HPC in Colombia. Secondly, it has enabled us to keep the school's curriculum updated and relevant, incorporating cutting-edge tools and real-world applications.

3.3 Inclusivity

In order to incentivize the participation of underrepresented communities and guarantee inclusivity across underrepresented groups, we offer open-to-all scholarships for attendance on every version of the summer school. Acceptance was granted to participants (general public) who demonstrated in their applications how the summer school program would benefit their interest in career opportunities or strengthen ongoing research. We also consider the need to increase the participation of underrepresented groups in STEM, such as women in science communities, and low-income communities in the country. Although we slightly increased female attendance, the gap between men and women who participate in this kind of event is still high, as shown in Figure 2. Nevertheless, we understand that gender equality in science is a progressive transformation process.

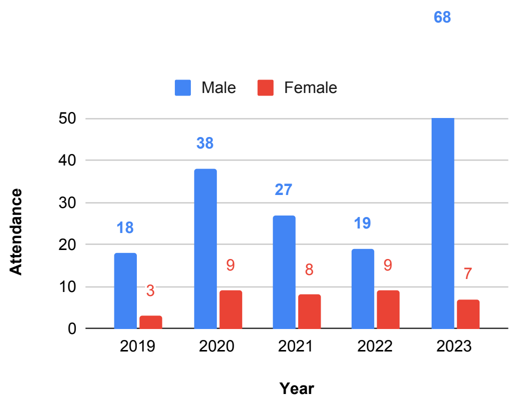


Figure 2: Attendance gender

3.4 Influence on Attendance and Collaboration

The Summer School has impacted more than 200 students in more than 20 institutions, as shown in Figure 3. This achievement owes itself to the collaborative effort of 27 organizations comprising universities, research institutions, and the industry, as described in Figure 4 and Table 3.

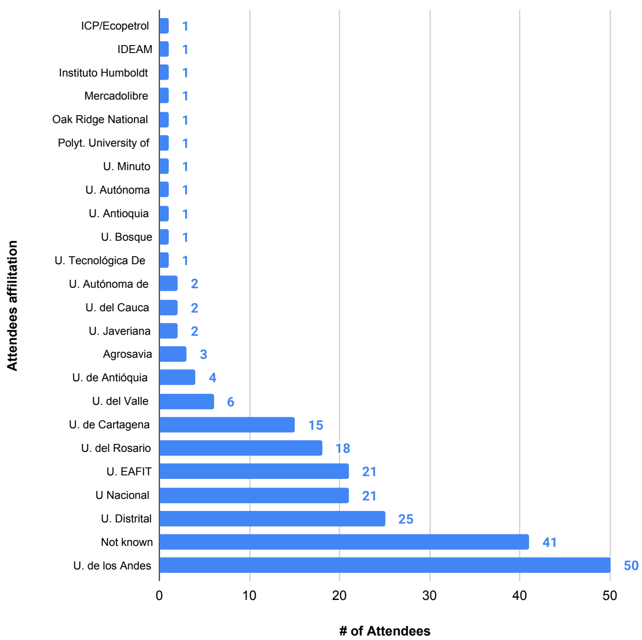


Figure 3: Number of attendees per institution (2019-2023)

4 HPC SUMMER SCHOOL 2023: A DISTRIBUTED EVENT

As a result of our experiences before and after the SARS-CoV-2 lockdowns, we were able to use both remote and in-person tools to plan and carry out the HPCSS initiative in 2023. The possibility of creating a hybrid remote-in-person event offered a way to scale the initiative by including more actors, particularly universities. These multiple institutions were willing to act as decentralized venues for the event, connecting all of them remotely but maintaining the in-person character of the workshops within each venue. Similar approaches have been taken by XSEDE and related programs.

A big challenge for this modality is to procure centralized access to the hardware and software used to run the tutorials. Decentralization has the potential to reduce coherence and coordination,

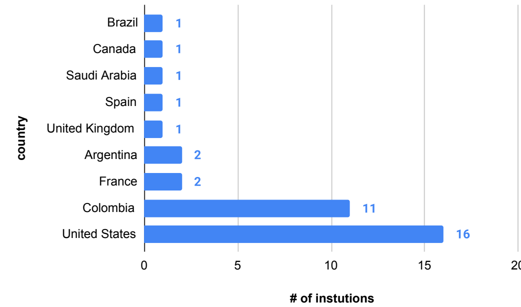


Figure 4: Number of Institutions Participating (with Speakers or Mentors) per Country (2019-2022)

Table 3: List of Institutions Participating (with Speakers or Mentors) per Country (2019-2022)

Institutions	Country
UNESP Center for Scientific Computing	Brazil
NVIDIA	United States
Universidad de los Andes	Colombia
Universidad del Valle	Colombia
Universidad Industrial de Santander	Colombia
KAUST Supercomputing Core Laboratory	Saudi Arabia
University of Delaware	United States
IBM	United States
University of Tennesse	United States
Pittsburgh SuperComputer Center	United States
AWS	United States
10x Genomics	United States
PSL	Colombia
Oak Ridge National Laboratory	United States
Argone National Laboratory	United States
University of Buenos Aires	Argentina
Loyola University	United States
MathWorks	United States
Argonne National Laboratory	United States
Coiled	United States
Universidad del Rosario	Colombia
National Aeronautics and Space Administration (NASA)	United States
Universidad Distrital	Colombia
Cybercolombia	Colombia
Renata	Colombia
Earlham	United Kingdom
Microsoft	United States
Barcelona Supercomputing Center	Spain
Universidad Industrial de Santader	Colombia
ATOS	France
McMaster University	Canada
iMMAP Colombia	Colombia
USGS Geologycal Survey	United States
UbiHPC	Colombia
European Space Agency (ESA)	France
DYMAXION LABS	Argentina

ultimately diminishing the experience for participants. After evaluating multiple options including in-house solutions, request allocations to leadership computing facilities, and cloud, we decided to use Chameleon Cloud [2].

Chameleon Cloud [17] is a *"large-scale, deeply reconfigurable experimental platform built to support Computer Sciences systems research"* [8], and it provides testbeds as instruments for research, as well as education. CyberColombia's HPCSS initiative, being an educational project, benefits greatly from the resources offered by Chameleon Cloud, providing the centralization of computing and software that the distributed modality requires.

In this section, we describe the methodology behind this last version of the Summer School. Here we explain the general organization, the program at a glance, supporting material, the challenges encountered, and how we managed to solve them.

4.1 Organization

Figure 5 provides a visual representation of the general structure of the Colombian HPCSS 2023. For this version, we partnered with five universities located in Colombia's primary metropolitan areas (Bogota, Medellin, and Cartagena). Faculty professors at these universities, referred to as *leaders*, were responsible for (1) encouraging participation in the event at their respective universities; (2) selecting students (up to 15) as well as a group of *volunteers* ranging from one to eight; and (3) providing a suitable space at their institution for students to receive the training and interact with each other.

CyberColombia, as the primary organizer, conducted two meetings with the leaders prior to the event. In the first meeting, the methodology and organization of the event were explained. In addition, individual sessions were scheduled to carry out network tests for the transmission (i.e., audio and visualization) and connection to the HPC platform. In the second meeting, the aforementioned tests were carried out in the rooms designated for the student's training. These technical tests enabled us to assess the transmission delay between universities and identify any connection issues from the universities to the HPC on Cloud infrastructure.

Furthermore, CyberColombia provided training materials and facilitated access to the HPC on Cloud platform in advance for volunteers, ensuring that these participants were well-informed about the content and adequately prepared to address minor issues and questions.

During the event, a team of four CyberColombia speakers and mentors, alongside two highly experienced volunteers, delivered both theoretical and hands-on content via Zoom, while also offering continuous support through Slack. Active support for students was presented by leaders of each institution and volunteers in the host sites along with the remote team. Professors at host sites not only reinforced the online-delivered content in person but also gauged students' comprehension and involvement with the material. During the event, they provided valuable feedback on what was the environment at each site, allowing us to adapt our progress accordingly. To say it simple, faculties and volunteers were our eyes on site.

[2] https://chameleoncloud.org/

Other responsibilities of each institution included: catering, collection of the registration fee, dealing with administrative procedures and permits, and generation of the necessary certificates.

4.2 Program at a glace

Table 4 depicts the HPCSS 2023 Daily schedule. The summer school program featured two discussion forums along with five introductory training that included both theory and hands-on tutorials.

Table 4: HPCSS 2023 - Daily schedule

TIME	DAY 1	DAY 2	DAY 3
8:00 - 900	Welcome	Colombia in HPC: Discussion Forum	Landscape of Supercomputing in Colombia: Discussion Forum
9:00 - 9:30		Coffee break	
9:30 - 12:30	Track 0 Parallel Computing Fundamentals	Track 2 Paralle Programmin with OpenMP	Track 4 Parallel and Distributed Programming with Python
12:30 - 13-30		Lunch	
13:30 - 16:30	Track 1 C++ 101	Track 3 Distributed Programming with MPI	Networking / Closing

4.3 Tutorial topics and material

The objective of this iteration of the summer school was to introduce the participants to the general concepts of parallel programming, as well as the basic hardware and software tools that are used. The tutorial is planned for 20 hours, divided into 3 days. The topics presented to the participants are divided into sections as follows:

Introduction to HPC. To motivate the whole program, the participants were first introduced to Moore's proposal for the growth of computational power in time and its limits. We described limitations posed by energy consumption, heat dissipation, and the physical limits of miniaturization of the transistor. The concepts of concurrent, parallel, and distributed computing were then introduced as solutions to the computational scaling problem. The concepts of process and thread were also presented here, along with common metrics used for assessing the success of parallel solutions such as speed-up, weak scaling, and strong scaling. This introductory section was then finalized with some examples and exercises using Python's threading module. We have chosen Python as it is a widely known language with a simple syntax, while the threading module allows the participants to manually create and destroy threads, consolidating the concept of the fork/join model as a basis for many parallel tasks.

Introduction to C++. The basic frameworks commonly used in HPC are OpenMP and MPI. However, before we can discuss them, we need to introduce C++, a language with solid and up-to-date implementations of these frameworks as well as a staple in the field of HPC. Many new programmers have not been introduced to this language. Thus, we used this tutorial section to introduce C++'s syntax and the basic structure of code. In particular, this section introduced the strong-typed nature of C++, the basic syntax for conditionals, loops, and functions, as well as the concepts of references and pointers. These latter are frequently used in memory copying operations in HPC, especially during accelerator programming.

Parallel computing with shared memory. In this section, the fork/join model was explored further. This section explained how to

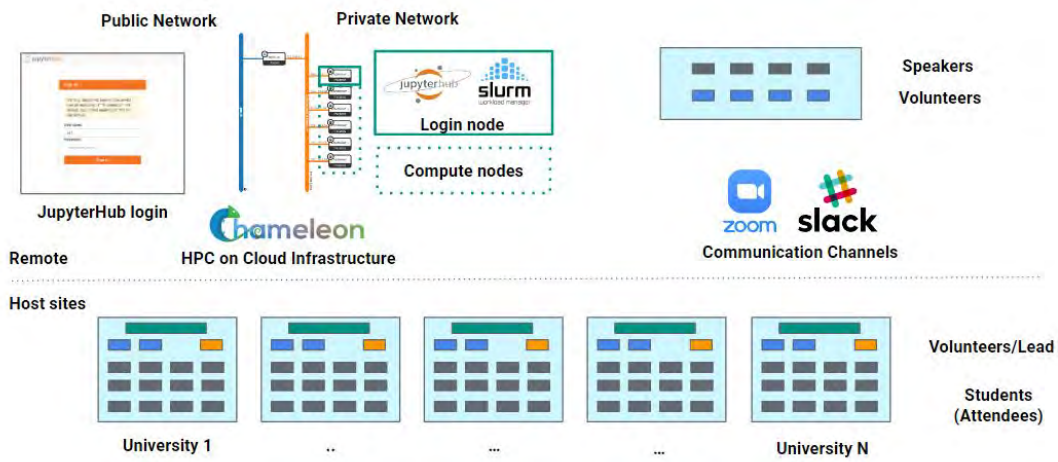


Figure 5: Colombain Organization of the High-Performance Computing Summer School 2023

implement parallel solutions in C++ using the OpenMP framework and deploy them on multi-core CPUs. The participants learned how to fork and join the flow of the program by creating threads with the #pragma omp parallel directive, as well as creating parallel loops and tasks. Several examples and exercises were provided to consolidate these concepts. The participants are encouraged to estimate the speedup of their parallel solutions. The tutorial used for this section is publicly available at [11].

Distributed computing. The concept of distributed computing was explored using C++ and MPI. The participants learned that to use several machines or nodes concurrently, a message protocol between nodes is needed. Using examples and hands-on exercises, participants learned how to initialize MPI to create the processes and perform point-to-point communications to send and receive messages between individual nodes.

Parallel computing with Python. Data science (DS) and artificial intelligence (AI) have gained great predominance. A great amount of the computational work in this field is done using Python. In this final module, the participants had the opportunity to work with Dask [21], a Python framework that allows the scaling of popular Python modules used for DS and AI. Dask performs the scaling by parallelizing the Python code, allowing the programmer to work with large amounts of data.

4.4 Technical Aspects

Teaching fundamental skills in parallel and distributed computing, the building blocks of High-Performance Computing, presents unique pedagogical challenges [19, 20]. These challenges extend beyond just delivering content to students; they also encompass the crucial task of ensuring access to the appropriate technological tools for a comprehensive and effective learning experience integrating both theory and practice.

When it comes to enabling the appropriate technological tools, the primary concern resolves around access to a flexible, configurable, and user-friendly HPC infrastructure. On this matter, our teaching infrastructure has shifted over different HPC solutions, providing considerable improvements every year.

Back in 2018, our main education infrastructure was based on an in-house university cluster. This required direct coordination and constant support from the facilities team. Challenges like account management, network restrictions, and unsupported software were common. Other solutions involved shared time in cloud computing services such as Amazon AWS and Google Cloud. However, the credits were often not enough. Access required additional steps for the users and the final result was not representative of leadership computing facilities worldwide (e.g., no batch scheduling).

Most recently, we deployed all our infrastructure using Chameleon Cloud [17]. This solution has proven to be the right middle point between resembling HPC ecosystems and having enough control over the configuration. Chameleon Cloud enabled us to allocate bare-metal HPC computing nodes located in leading HPC computing facilities such as Argonne National Laboratory and Texas Advanced Computing Center. The allocated instances are provisioned using ready-to-use Ansible playbooks, developed and maintained by CyberColombia for the HPCSSs. The provisioned software stack typically includes Slurm (workload management system), C/C++ and C-lang (compilers) with support for OpenMP, CUDA, OpenMPI, JupyterHub (for user-HPC interaction), and Jupyter notebooks (to deliver theoretical and practical content). Recently, we updated our software stack to include data analytics frameworks such as Dask; comprising interfaces like Dask-jobqueue, enabling the interaction of the framework with the workload management system.

Jupyter notebooks have garnered significant momentum as a powerful tool for interactive "human-in-the-loop" based research. However, care should be taken in the way they are used for teaching, particularly because the format does not resemble a genuine user-HPC system interaction. Nonetheless, for training and education, they substantially reduce the learning barriers.

As the HPC system is managed with the Slurm scheduler to provide isolation between different executions and deliver the performance expected in parallel and distributed applications, we did not write executable code directly into the notebook's cells. Instead, the notebooks provide the guidelines to edit pre-filled source files, then source files are compiled and sent to the scheduler through customized commands inserted on notebooks' cells. These commands provide interactive job submissions, timing, and memory and CPU usage tracking. This approach allowed us to maintain an interactive learning experience without delving too deeply into the intricacies of job submission. The primary goal of the summer school is for students to grasp the potential of parallel and distributed computing.

5 LESSONS LEARNED AND FUTURE DIRECTIONS

From the experience conducting HPC training events in Colombia, we concluded the following lessons learned: First, although virtual events stimulate participation in both speakers/mentors and attendees, this approach presents several drawbacks in the long term such as the lack of interaction among participants (student to student, students to mentors, and students to university), the lack of an appropriate learning environment, the lack of proper supervision of progress, among others. To overcome these challenges, we reintroduced in-person events while simultaneously sustaining remote content delivery, allowing us to expand our reach across multiple universities. Professors, on-site volunteers, and remote experienced volunteers played essential roles in providing students with a real-time, supportive learning experience. Secondly, in prior editions of the event, we observed that a significant number of participants lacked a fundamental understanding of HPC, leading to feelings of overwhelm and frustration when exposed to advanced materials and use cases that were beyond their current understanding. To address this concern, we reformulated our program into five tracks: Track 0. Parallel Computing Fundamentals, Track 1. C++ 101, Track 2. Parallel Programming with OpenMP, Track 3. Distributed Programming with MPI, and Tack 4. Parallel and Distributed Programming in Python. Finally, HPC in cloud solutions such as Chameleon Cloud played a paramount role in decentralizing our training efforts and amplifying our overall impact.

Future works will focus on tracking students' progress beyond the summer school and incentivizing their participation in the initiative, enabling them to contribute to the expansion of our programs across various locations and sectors, in different modalities (monthly talks, short tutorials, work-in-progress showcases).

ACKNOWLEDGEMENTS

We extend our gratitude to the numerous institutions, speakers, and mentors whose support has been instrumental in advancing CyberColombia's initiatives, along with all attendees who have contributed to our endeavors. Special appreciation goes to the Red Nacional Académica de Tecnología Avanzada (RENATA) for their invaluable partnership in the development of our latest version of the HPCSS. Furthermore, we express our thanks to Chameleon Cloud, backed by the National Science Foundation, for supporting our last two events. Finally, this research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy (DOE) Office of Science and the National Nuclear Security Administration. This research used resources from the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357.

REFERENCES

[1] Argonne Leadership Computing Facility . 2023. Training and Events | Argonne Leadership Computing Facility. https://www.alcf.anl.gov/events. (Accessed on 09/08/2023).

[2] Association for Computing Machinery (ACM) . 2023. HPC. https://europe.acm.org/seasonal-schools/hpc. (Accessed on 09/08/2023).

[3] IHPCSS . 2021. International HPC Summer School (IHPCSS). https://www.ihpcss.org/. (Accessed on 09/08/2023).

[4] National Energy Research Scientific Computing Center . 2023. NERSC Training Events. https://www.nersc.gov/users/training/events/. (Accessed on 09/08/2023).

[5] Partnership for Advanced Computing in Europe (PRACE). 2019. Training portal. https://training.prace-ri.eu/. (Accessed on 09/08/2023).

[6] ACCESS. 2023. ACCESS: Advancing Innovation. https://access-ci.org/. (Accessed on 09/08/2023).

[7] Argonne Training Program on Extreme-scale Computing. 2013. Argonne Training Program on Extreme-Scale Computing. https://extremecomputingtraining.anl.gov/. (Accessed on 09/06/2023).

[8] Chameleon Cloud. 2023. About Chameleon. https://www.chameleoncloud.org/about/chameleon/. (Accessed on 12/07/2023).

[9] AMERICAS HPC Collaboration. 2022. AMERICAS HPC Collaboration. https://americashpccollaboration.net/. (Accessed on 08/09/2022).

[10] ASC Community. 2018. *The Student Supercomputer Challenge Guide: From Supercomputing Competition to the Next HPC Generation.* Science Press.

[11] Jose M Monsalve Diaz CyberColombia. 2023. OpenMP Tutorial in Jupyter Notebooks. https://github.com/josemonsalve2/openmp_tutorial. (Accessed on 08/09/2023).

[12] SCALAC: Sistema de Computación Avanzada para Latino América y el Caribe. 2022. RedCLARA Home. https://scalac.redclara.net/index.php/es/. (Accessed on 08/09/2022).

[13] Jose De Vega, Monica C Munoz-Torres, Jorge Duitama, Narcis Fernandez-Fuentes, Graham J Etherington, and Robert P Davey. 2019. C3-Biodiversidad Consortium: A Cyberinfrastructure to Accelerate the Understanding and Preservation of the Colombian Biodiversity. In *Plant and Animal Genome XXVII Conference (January 12-16, 2019).* PAG.

[14] Jose J De Vega, Robert P Davey, Jorge Duitama, Dairo Escobar, Marco A Cristancho-Ardila, Graham J Etherington, Alice Minotto, Nelson E Arenas-Suarez, Juan D Pineda-Cardenas, Javier Correa-Alvarez, et al. 2020. Colombia's cyberinfrastructure for biodiversity: Building data infrastructure in emerging countries to foster socioeconomic growth. *Plants, People, Planet* 2, 3 (2020), 229–236.

[15] S.L. Graham, M. Snir, and C.A. Patterson (Eds.). 2005. *Getting Up to Speed: The Future of Supercomputing.* The National Academies Press.

[16] E. Hernández, C.E. Álvarez, C.A. Varela, J.P. Mallarino, and J. De Vega. 2021. CyberColombia: a Regional Initiative to Teach HPC and Computational Sciences. *ACI Avances en Ciencias e Ingenierías* 13, 2 (2021), 9.

[17] K. Keahey, J. Anderson, Z. Zhen, P. Riteau, P. Ruth, D. Stanzione, M. Cevik, J. Colleran, H. Gunawi, C. Hammock, J. Mambretti, A. Barnes, F. Halbach, A. Rocha, and J. Stubbs. 2020. Lessons learned from the Chameleon testbed. In *Proceedings of the 2020 USENIX Annual Technical Conference (USENIX ATC '20).* USENIX Association, Berkley, CA, USA, 219–233. https://doi.org/10.5555/3489146.3489161

[18] Oak Ridge Leadership Computing Facility . 2023. Training – Oak Ridge Leadership Computing Facility. https://www.olcf.ornl.gov/for-users/training/. (Accessed on 09/08/2023).

[19] Rajendra K Raj, Carol J Romanowski, John Impagliazzo, Sherif G Aly, Brett A Becker, Juan Chen, Sheikh Ghafoor, Nasser Giacaman, Steven I Gordon, Cruz Izu, et al. 2020. High performance computing education: Current challenges and future directions. In *Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education.* 51–74.

[20] Diego A. Roa Perdomo, Paige C. Kinsley, Jose M. Monsalve Diaz, and Michael E. Papka. 2023. DEMAC – A Platform for Education in High-performance Computing, Bridging the Gap Between Users and Hardware. (2023). [Manuscript submitted for publication].

[21] Matthew Rocklin et al. 2015. Dask: Parallel computation with blocked algorithms and task scheduling. In *Proceedings of the 14th python in science conference*, Vol. 130. SciPy Austin, TX, 136.

[22] Supercomputing Camp (SC-camp). 2023. Supercomputing Camp 2023. https://sc-camp.org/2023/index.html. (Accessed on 09/06/2023).

[23] UNESCO Institute for Statistics. 2022. Research and development expenditure. https://datos.bancomundial.org/indicator/GB.XPD.RSDV.GD.ZS. (Accessed on 08/09/2022).

[24] XSEDE. 2022. XSEDE: Extreme Science and Engineering Discovery Environment. https://www.xsede.org/. (Accessed on 09/08/2023).