

## CyberColombia: a Regional Initiative to Teach HPC and Computational Sciences

Esteban Hernandez<sup>1</sup>, Carlos E. Alvarez<sup>2</sup>, Carlos Alberto Varela<sup>2</sup>, Juan Pablo Mallarino<sup>3</sup> and Jose J. De Vega<sup>4</sup>

<sup>1</sup>Universidad Distrital, Bogotá, Colombia

<sup>2</sup>Universidad del Rosario, Bogotá, Colombia

<sup>3</sup>Universidad de los Andes, Bogotá, Colombia

<sup>4</sup>Earlham Institute, Norwich, United Kingdom

\* Corresponding author/ Autor principal: [ejhernandezb@udistrital.edu.co](mailto:ejhernandezb@udistrital.edu.co)

## CyberColombia: Una iniciativa regional para enseñar Cómputo de Alto desempeño y ciencias computacionales

### Abstract

The series *Summer School HPC Colombia* is an initiative to extend highperformance computing-related knowledge in Colombia, and more widely in Latin America, and integrate expertise and research from academia and industry in the same event. This year's edition, which is the third in the series, was carried out entirely online due to the outbreak of the COVID 19 pandemic during the first half of the year 2020. In this paper, we summarise the aims, development, deployment, and results of the *Summer School HPC Colombia 2020* event. It is an example of the potential that the use of virtual tools and environments has to grow education for HPC.



**Keywords:** HPC Training, Biocomputation training, Parallel programming

### Resumen

La serie de escuelas de Verano en HPC Colombia es una iniciativa para extender la capacidad y conocimiento en Colombia relacionado al cómputo de alto desempeño, esta iniciativa pretende tener impacto en Latinoamérica integrando experiencias de investigación, academia e industria en un mismo evento.

Dado la pandemia relacionada al COVID 19, desde la mitad del año 2020, el evento desde su tercera edición se ha desarrollado estrictamente en línea. En este paper hemos descrito de manera general el propósito, el desarrollo, el despliegue y los resultados obtenidos de la escuela de verano HPC Colombia 2020. Este evento es un ejemplo del potencial del uso de herramientas y ambientes virtuales para desarrollar y hacer crecer la educación relacionada al cómputo de alto desempeño (HPC)

**Palabras clave:** Entrenamiento HPC, Computación Paralela, Programación Paralela, Biocomputación

Licencia Creative Commons  
Atribución-NoComercial 4.0



Editado por /  
Edited by:  
Dennis Cazar

Recibido /  
Received:  
04/19/2021

Aceptado /  
Accepted:  
07/09/2021

Publicado en línea /  
Published online:  
15/12/2021



## MOTIVATION

Exploring the large data-sets generated with today's highly instrumented data collection practices stretches the capacity of single groups or institutions, and requires interdisciplinary partnerships across research domains, particularly between scientists and computational engineers.

As the computational power at our disposal increases, the possibility to solve larger problems opens to us. High performance computing (HPC) is a vehicle that can foster scientific innovation and knowledge-driven economic growth in Colombia.

Several sectors important for Colombia's economy are data-intensive, including: Drug development [1], weather prediction [2], oil and gas exploration [3], astrophysics [4], biodiversity genomics [5], development of new materials [6, 7] and AI [8].

The power of HPC systems mainly relies on the use in parallel of many processors, which implies the management of distributed and/or shared resources, and the communication between different threads or processes. This is a computational model most programmers are not used to work with, which presents an access barrier for many programmers. There is consequently an opportunity to grow education and training to acquire skills in parallel computing in Latin America. Filling this knowledge gap is important for the development of the region.

## THE CYBERCOLOMBIA INITIATIVE

Cybercolombia is an interdisciplinary partnership across research domains, particularly between scientists and computational engineers, resulting from the coordination of several independent projects with shared objectives, including the *Summer School HPC Colombia*.

Cybercolombia aims to develop the critical skills, strategic planning and networking required to make available and maintain a highperformance digital infrastructure (or cyberinfrastructure) for the analysis of large data-sets in Colombia and Latin America. However, an efficient data infrastructure not only consists of an advanced set of computational tools, so the partnership also aims at influencing sustainable data policies, as well as fulfilling Colombia's needs for experts with the technical skills necessary to execute and share those resources, services and tools in a sustainable, secure and interoperable way.

Cybercolombia's objectives are to:

- Improve the development and availability of tools and services for data-intensive science.
- Facilitate advanced skills and competencies in data management and analysis.
- Promote best practice and influence policies for data access and management.



A primordial event was the *C3Biodiversidad* workshop organised in Bogota in June 2018. In this workshop, experts from sixteen Colombian institutions, and a panel of international infrastructure maintainers and tool developers from the UK and USA, carried out an analysis of the needs to promote a cyberinfrastructure for the analysis of Colombia's biodiversity data [5].

In a later event held at University de Los Andes (Bogota, Colombia) in 2019, thirty two stakeholders from industry, University and Government in the areas of big data and bio-economy in Colombia analysed the challenges and opportunities for the big data sector in Colombia, and the tentative role of this sector in the socio-economic growth of Colombia.

## THE HPC SUMMER SCHOOL ALONG THE YEARS

The HPC summer school events have been a regular yearly series up to the present. The events consisted of two main parts: (i) Informative talks given by international speakers in various HPC related topics and (ii) practical workshops supervised by both teachers from academia and outreach staff from industry. The events are usually more focused on hands-on training, which we consider closer to the core purpose of a *summer school*. However, the balance between the number of talks and workshops has changed from year to year due to the availability of speakers, the prioritisation of recent developments in the field, or external factors, as the recent COVID-19 pandemic.

The aim of the informative talks is twofold. First, they aim to introduce the attendees to basic-to-intermediate level HPC-related topics with the intention of leveling up the field for all participants, who may come from very different backgrounds and possess different knowledge on the field. Second, it attempts to motivate the attendees and create an atmosphere of curiosity by presenting cutting-edge relevant topics.

On the other hand, workshops are designed to give a hands-on introduction to the technical aspects of the subject. By directly interacting with the different methods and technologies related to HPC, the attendees acquire an understanding of topics that allow them to link core concepts, such as parallel programming, management of shared and distributed memory, use of container technologies, etc.; to the user-cases in their respective fields. This should facilitate options to enhance their productivity and explore novel approaches that were previously out of reach due to a need for large computing resources or data size.

## FIRST HPC SUMMER SCHOOL

The first iteration of the summer school took place in 2018 at Universidad de los Andes, Bogotá, Colombia. The participants were mainly students at the undergraduate (60%) and graduate (20%) level, along with a fraction of the participants coming from private and public institutions (20%). This initial event focused mainly on hands-on workshops and the talks, at the beginning of each day, were aimed at providing a background for the workshops held later in the day.

The topics presented in where the following:



- Introduction to HPC (talk)
- Introduction to C++, cython and the torque scheduler (workshop, C. Alvarez, V. Arias, J.P. Mallarino, U. de los Andes, U. del Rosario)
- OpenMP (Workshop on Hybrid MPI/OpenMP programming on Intel platforms, S. Stanzani, )
- Introduction to accelerators (talk, P. Cruz Silva, Nvidia)
- Cuda/OpenACC (talk and workshop, P. Cruz Silva, Nvidia)
- Singularity containers (talk and workshop)

In this iteration a challenge that consisted in accelerating a particular application using the tools learned in the school.

## SECOND HPC SUMMER SCHOOL

The second iteration was held at Universidad del Rosario, Bogotá, Colombia from June 5th to 9th 2019. On this event the attendees were mainly students at the undergraduate (50%) and graduate (40%) level, as well as academic/teaching staff (10%). The focus on this occasion continued to be on hands-on workshops with introductory talks to each topic. A talk and workshop day about distributed memory computing with MPI was also added.

The topics presented during this iteration were the following:

- Introduction to HPC and cloud computing (talk and workshop, K. Jorissen, AWS)
- Introduction to C++ (workshop, J. Rincón, U. del Rosario)
- OpenMP (talk and workshop, J.P. Mallarino, U. de los Andes)
- OpenACC (talk and workshop, P. Cruz Silva, Nvidia)
- MPI (talk and workshop (C. Alvarez, U. del Rosario))

This second iteration also included a challenge to parallelize code using the tools learned in openACC.

## THIRD HPC SUMMER SCHOOL

The year 2020 was marked by the outbreak of COVID 19, which among other things, affected in many aspects the way in which events could be developed. Our third iteration of the summer school was no exception and the organization presented new challenges as all the aspects of the summer school had to be moved to virtual mode.



The balance of topics in the 2020 iteration of the summer school leaned towards the talk presentation mode more than the previous versions. This was principally because the presentations done via streaming allowed the speakers to remain at their home towns, which meant less traveling expenses and time consumption for them. This facilitated their participation. Likewise, the levels of attendance increased, presumably due to the same factors.

On the other hand, the deployment of the workshops was presented with some new technical challenges, as the attendees had to be able to participate remotely and have access to advanced machines and software which was not available at their home machines. How these challenges were met is the topic of the next section.

The topics presented during this last iteration were the following:

- Building HPC systems (keynote, J. Moreno, IBM)
- Convergence of HPC and Big Data (talk, S. Caí no-Lores, U. Tennessee)
- Nvidia for Healthcare (talk, P. Cruz Silva, Nvidia)
- New directions in AI-driven research (keynote, P. Buitrago and N. Nystrom, PSC)
- Biology at true resolution (talk, A. Suarez, 10x Genomics)
- HPC on the cloud (talk, K. Jorissen, AWS)
- AWS Graviton2 processors (talk, A. Petitpiere, AWS)
- Scalability on bio-inspired computational models (talk, D. Dematties, U. Buenos Aires and G. Thiruvathukal, Loyola U. Chicago and S. Rizzi, ANL)
- Tensorflow (talk, F. MARTINEZ, PSL)
- HPC against COVID 19 (talk, D. Bhowmik, ORNL)
- OpenACC (workshop, J. Monsalve, U. Delaware)
- MATLAB for biomedical applications (workshop, L. Walker-Hannon, MathWorks)  
Parallel programming with MATLAB (workshop, S. Obando, MathWorks)

## TECHNOLOGICAL CHALLENGES FOR THE 2020 SUMMER SCHOOL

Due to the *COVID-19* pandemic, the main obvious challenge for this year summer school was to bring the school up in a virtual fashion. To do that we used several tools to leverage virtual sessions as well as interactions among organizers and participants. For the virtual sessions Zoom conference rooms were enabled. Each day a different Zoom room was appointed and participants were informed via e-mail in advance.



In order to keep communications going, a slack channel was enabled so participants and organizers could keep updates and questions.

Hands on practical sessions are of paramount importance for events such as summer schools. It is there where the participants get their hands on and learn from practice and experience. As a key component of the 2020 *Summer School HPC Colombia*, two main workshop were carried out: *OpenACC* and *Mathworks*. Each of them presenting their own challenges to become operative and functional. The speaker Presentations were given through Zoom virtual rooms while practical tutorials were performed in parallel through the available remote platforms. Below we present a short overview of the main sessions carried out in this year summer school.

- *Mathworks session*: This session developed two different tracks. The first one being related to *deep learning* and aiming to appoint data science and artificial intelligence in MATLAB for Biomedical applications. While the second one focused on parallel programming principles using OpenACC[9]. Both practical session were sponsored by Mathwoks[12]. They provided a private virtual platform where summer school participants could access via web to execute the exercises remotely. Users were required to register in advance to apply for a software valid licence.
- *OpenACC session*: This session was organized using the OpenACC Official Training Material [9]. Since the material was designed to execute in a single instance of docker per user. Modifications had to be done in order to adjust jupyter notebooks multi-user execution in the Centauro HPC cluster at Unviersidad del Rosario.

To do that, a jupyterhub[10] server was set on the master node. So multiple users could access simultaneously through a friendly jupyter notebook web-based interface. The training OpenACC material was taken out from the docker instance and make it available to user working space. In this way users could execute notebooks on the master node. Since the tutorial targeted the use of GPU's, an integration with slurm workload manager[11] had to be implemented in order allow users to allocate computer resources equipped with GPU capabilities within the cluster. In this way, users could independently access notebooks from the jupyterhub server to later on execute each notebook through the submission of jobs via slurm[11] to compute nodes available. The architecture is depicted in Figure 1:



**Figure 1.** jupyterhub slurm integration architecture for OpenACC training session



## CONCLUSION AND FUTURE PERSPECTIVES

The HPC Summer School series of events, as part of the Cybercolombia partnership, has successfully trained dozens of students and staff from academic institutions and industry in the usage and development of applications for HPC. In the last edition in 2020, we developed an entirely virtual *summer school*. From this experience we concluded several differences compared with our previous events :

- The number of speakers increased, as the financial and time costs of presenting in a virtual environments are significantly lower than that of live presentations.
- The involvement and questions from the attendees were similar to that in previous events.
- Nevertheless, opportunities of networking were probably affected as no interaction with the speakers was possible outside of the programme.
- The workshops required more preparation from a technical point of view, but once these aspects were covered, we were able to deploy them without further complications.
- The workshop full filled their aims, i.e. it was possible for the students to perform the exercises and interact with the tutors.
- The use of breakout rooms to aggregate the attendees into smaller groups proved to be a successful strategy to focus the time and attention of the tutors.

For future events we plan to use the lessons learned in face-to-face as well as remote events in order to offer possibilities for remote as well as in-person participation, broadening the reach and scope of the event.

## ACKNOWLEDGEMENTS

Supported by GROW Colombia ([www.growcolombia.org](http://www.growcolombia.org)) and the Royal Academy of Engineering's Newton Fund.

## AUTHORS' CONTRIBUTIONS

Esteban Hernandez, Carlos Alvarez, Juan P. Mallarino and Jose De Vega conceived the idea and execution of the workshop. Carlos Varela was responsible for the technical aspects of the HPC services.

## CONFLICT OF INTEREST

The authors whose names are listed immediately below certify the following:



- They have NO affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.
- All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.
- This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

Esteban Hernández, Carlos Eduardo Álvarez, Carlos Alberto Varela, Juan Pablo Mallarino, Jose de Vega.

## REFERENCES

- [1] Allen, J. (2020). HPC-driven collaboration accelerates drug discovery. LLNL. <https://computing.llnl.gov/newsroom/hpcdriven-collaboration-accelerates-drugdiscovery>
- [2] Tracy, S. (2018). Why HPC matters: Weather and climate. CIO. <https://www.cio.com/article/3254168/whyhpc-matters-weather-and-climate.html>
- [3] Tyler, T. (2019) The surge in Oil & Gas supercomputing. Enterprise AI. <https://www.enterpriseai.news/2019/01/21/the-surge-in-oil-gas-supercomputing/>
- [4] Guedes, J.; Callegari, S.; Madau, P.; & Mayer, L. (2011). Forming realistic late-type spirals in an  $\Lambda$ CDM universe: the Eris simulation. *The Astrophysical Journal* 742, 76. doi: <https://doi.org/10.1088/0004-637X/742/2/76>
- [5] De Vega, J. J.; Davey, R. P.; Duitama, J.; Escobar, D.; Cristancho [U+2010] Ardila, M. A.; Etherington, G. J.; ...; & Di Palma, F. (2020). Colombia's cyberinfrastructure for biodiversity: Building data infrastructure in emerging countries to foster socioeconomic growth. *Plants, People, Planet* 2, 229-236. doi: <https://doi.org/10.1002/ppp3.10086>
- [6] Pickard, C.J.; & Needs, R.J. (2006). Highpressure phases of silane. *Physical Review Letters* 97, 045504. doi: <https://doi.org/10.1103/PhysRevLett.97.045504>
- [7] Schusteritsch, G.; & Pickard, C.J. (2014). Predicting interface structures: From SrTiO3 to graphene. *Physical Review B* 90, 035424. DOI: <https://doi.org/10.1103/PhysRevB.90.035424>
- [8] Schneibel, A. (2020). New data analysis software tools aim at addressing the convergence of HPC and AI. ORNL. <https://www.olcf.ornl.gov/2020/04/29/newdata-analysis-software-tools-aim-ataddressing-the-convergence-of-hpc-and-ai/>
- [9] OpenACC (2020). [Training Material]. Retrieved August 24, 2020, from <https://ngc.nvidia.com/catalog/containers/hpcopenacc-training-materials>
- [10] Jupyterhub (2020) Jupyter Project [multi-user notebook version]. <https://jupyter.org/hub>
- [11] Slurm (2020) Slurm Workload Manager [Version 21.08]. <https://slurm.schedmd.com/>
- [12] Mathworks (2020) [Mathematical computing software for engineers and scientists] Retrieved August 27, 2020, from <https://www.mathworks.com/>