SCALAC

Advanced Computing System for Latin America and the Caribbean

High-Performance Computing Robust Systems Report in Latin America and the Caribbean November 2024

Version 2.3

Authors:
Carlos Jaime BARRIOS HERNANDEZ, PhD.
Nicolás WOLOVICK, PhD.
Luis Alejandro TORRES NIÑO, MSc. Eng.

Revision:
Philippe Olivier NAVAUX, PhD.
Harold Enrique CASTRO BARRERA, PhD.
Esteban MENESES, PhD.

Content

Summary

High-performance computing (HPC) Infrastructure supports advanced computing applications and research. Countries in Latin America and the Caribbean have invested in HPC infrastructure via institutions, universities, and special projects. Usually, these investments are made with national funds that are not only their own but also constitute common funds through alliances by partners or foreign programs.

In this reality, investments in robust infrastructure, while not continuous, hold the potential to support a wide range of needs. These needs go beyond simulation and traditional scientific computing, extending to uses such as data analysis, artificial intelligence, and quantum computing. This potential opens up exciting possibilities for the future of computing in Latin America and the Caribbean.

Starting with Deliverable 2 of the RISC-2 project[1], the HPC Observatory has continued for Latin America and the Caribbean, considering the robust HPC infrastructures mapped in these regions and with the infrastructure in which the information is public. This version of the document High-Performance Computing Robust Systems Report presents the state of these infrastructures in Latin America and the Caribbean, observing the information collected and compared and other parameters for this report. This second version, conserving some important information, considers the confrontation of information from the past published report 1.3, complementing and updating information given previously. On the other hand, in this report, we start some critical discussion points about the use and deployment of robust infrastructures for HPC in Latin America and the Caribbean.

The report is organized as follows: Section 1 introduces this report, describes the context and methodology used, and provides a general description of the potentialities in infrastructure in the region. In this version, the document includes energy, type of cooling, and connectivity to a high-speed collaborative network via an NREN or RedCLARA. Section 2 shows the list of infrastructure considered in this document following the proposed method, now with performance discrimination; Section 3 shows an analysis of the use and downtime of the HPC top verified Infrastructures. Finally, Section 4 contains the final remarks and recommendations for the next version of this report.

[1] Deliverable 2.1., RISC-2 Project (www.risc-2.eu) White Paper on HPC RDI in LATAM https://www.risc2-project.eu/wp-content/uploads/2022/09/RISC2_Deliverable-2.1-White-Paper-on-HPC-RDI-in-LATAM_v1.1-1.pdf

1. Introduction

HPC Infrastructure investment in Latin America has been a topic of discussion. According to an Inter-American Development Bank (IDB) report[2], massive infrastructure transformation is needed in Latin America and the Caribbean. The report emphasizes efficiency, digital technologies, and a focus on the quality and affordability of consumer services. Latin American and Caribbean countries should invest around 6.2% of their annual Gross Domestic Product (GDP) in infrastructure to meet their needs. However, the actual investment in infrastructure needs to catch up to this target, ranging from 2.3% to 4% of GDP on average[3] [4].

Although Latin American HPC infrastructures have appeared in the Top500[5] for more than ten years, they are mainly located in two main countries: Brazil and Mexico. Taking only from 2010 to 2024, in June 2010, Brazil managed 3 supercomputers in the Top500 hosted by Petrobras[6], INPE[7], and a consortium of three entities NACAD/COPPE/UFRJ[8] and in November 2012, Mexico arrived with an HPC infrastructure hosted by the UNAM[9]. After observing the Top500, several milestones were presented in the region after these entries: in June 2015, ABACUS I[10], a supercomputer hosted by the CINVESTAV in Mexico, became the first new-generation supercomputer installed by mixing CPU and GPUs, and the National Laboratory of Scientific Computing (LNCC)[11] introduced the Santos Dumont[12], the first petaflop machine in the region; in November 2023, Argentina enters the Top500 with Clementina XXI, hosted by the SMN[13]; and finally in June 2024, Brazil positions eight (8) supercomputers in the Top500, having supremacy in HPC in Latin America and the Caribbean, and it is in position 12 worldwide, below Poland and above the United Arab Emirates.

---

[2] https://impact.economist.com/new-globalisation/infrascope-2024/en/

[3] https://www.cepal.org/en/news/infrastructure-investment-latin-american-and-caribbean-countries-remains-below-needs-region

[4] https://www.wilsoncenter.org/article/latin-america-must-prioritize-infrastructure-spur-economic-growth

[5] https://top500.org/

[6] https://petrobras.com.br/

[7] https://www.gov.br/inpe/pt-br

[8] http://www.nacad.ufrj.br/

[9] https://www.unam.mx/

[10] https://www.abacus.cinvestav.mx/

[11] https://www.gov.br/lncc/pt-br

[12] https://sdumont.lncc.br/

[13] https://www.smn.gob.ar

Brazil's positioning is thanks to both Brazilian public investment (which causes a significant percentage of GDP to go towards this type of investment by law) and private investment. Likewise, thanks to an important scientific activity that not only demands this type of infrastructure but also uses it, creating a meaningful community that maintains a shared vision and collaborative policies, national access to HPC infrastructures (via SINAPAD)[14] even including other countries, as is the case with SCALAC.

Outside of the Top500, however, the region has had an interesting dynamic. Without going into details of the GRID projects of the 90's and the beginning of this century, only in the same observation window (from 2012 to date) have there been very interesting referenced achievements: for example, it is no secret to anyone in the HPC community that the ASIC infrastructures for Astronomy hosted by the National Laboratory on HPC (NLHPC)[15] of Chile are among the most powerful infrastructures in the world[16]. In 2010, Colombia, thanks to the High Performance and Scientific Computing Center (SC3UIS)[17] at Universidad Industrial de Santander[18] in Bucaramanga, in an architectural co-organization and collaboration with HP (today HPE[19]), NVIDIA[20], and INTEL[21], decided to put into operation the first highly dense infrastructure for supercomputing in the region with a continental use: GUANE[22]. This machine (still in operation as continuous integration for testbed addressed to sustainability) became the first infrastructure with 8 NVIDIA HPC GPUs per node for 128 NVIDIA HPC GPUs (and 2 and 6 CPUs by node) and became a reference case for both the integrator (HP) and NVIDIA. The impact was so huge that investments in supercomputing in Colombia practically skyrocketed after GUANE, and it became a global test infrastructure for HPE. In 2016, the High-Performance Computing Center at Universidad Nacional de Córdoba in Argentina (CCAD-UNC)[23] led the development of clusters for general scientific and academic use, seeking low-cost and sustainable and integrating them into Argentina's national High-Performance Computing system (SNCAD)[24]. In the last report (1.3), three achievements in the region were mentioned:

[14] https://www.lncc.br/sinapad/

[15] https://www.nlhpc.cl/

[16] The Top500 only considers general-purpose machines, but if this were not, these supercomputers in Chile would be on that list.

[17] https://www.sc3.uis.edu.co

[18] https://www.uis.edu.co

[19] https://www.hpe.com/

[20] https://www.nvidia.com/

[21] https://www.intel.com/

[22] https://es.wikipedia.org/wiki/Guane-1_(supercomputador)

[23] https://ccad.unc.edu.ar/

[24] https://www.argentina.gob.ar/ciencia/redes/sistemasnacionales/cad

the first one, in 2018, the University of Guadalajara in Mexico created a unit whose primary purpose is data analysis (leaving the supercomputing part as a surname), named Center for Data Analytics and Supercomputing (CADS-UDG)[25] installing a first Fujitsu[26] supercomputer for this purpose, Leo Átrox[27]; second, the National Research and Education Network (NREN) in Ecuador, CEDIA[28], launches the first HPC infrastructure based on NVIDIA DGX for uses in Artificial Intelligence (IA) exclusively on a continental level, via SCALAC and RedCLARA[29]; and finally, in 2021, the Brazilian SENAI-CIMATEC[30] together with Atos (today Eviden) also puts into operation and continental use the first quantum simulator in the region (the Learning Quantum Machine Kuatomu). For this report, two achievements have been added in the region: first, the Brazilian company SiDi offers for the first time an infrastructure for the productive environment mainly dedicated exclusively to artificial intelligence, and two, ABACUS in Mexico and Santos Dumond in Brazil, were the first reported large-scale infrastructures that have implemented the use of water cooling in the region (reported). Other efforts have also been made in Uruguay, Costa Rica, and Panama.

Although in the previous summary, outside the Top500, the main milestones documented by the community were mentioned, some other investments and implementations have not been taken into account, either due to their low real impact on the specialized community or the false sizing of the owners or promoters of the infrastructure projects, or because these are private projects that cannot be publicized, such as those of some oil, banking, and military companies, which, to the knowledge of the authors of this document, would even be within the Top500 said infrastructures.

With this information, the HPC Observatory for Latin America and the Caribbean, now performed by SCALAC in collaboration with RedCLARA, has proposed conducting an inventory of strategic infrastructure using the methodology presented below.

1.1. Methodology

The list was built mainly based on the information sent by the technology heads, directors, or technological administrators of different sites, information pre-collected

[25] http://cads.cgti.udg.mx/
[26] https://www.fujitsu.com/
[27] https://es.wikipedia.org/wiki/Leo_Atrox
[28] https://cedia.edu.ec/
[29] https://www.redclara.net
[30] https://www.senaicimatec.com.br/

by the RISC-2 HPC Observatory, and information collected by SCALAC and RedCLARA. The information requested included not only technical aspects and real and theoretical measurements (estimated according to methodologies given by the manufacturers and integrated) in terms of computing but also storage and connectivity (basically, if a National Academic Network connected it). The methodology includes feedback (from the previous 1.3 published report) and confrontation given by the specialized community and confirmed interest groups.

The sellers obtained the information confidentially, based on equally confidential information from the manufacturers and integrators. In other cases, where the information was considered to have potential bias, tests and exact data were requested from independent administrators and engineers.

The information then requested to compile the list was the following:

- Institution
- Country
- City
- Institution (Including Type of Institution)
- Web (URL)
- Type of HPC Infrastructure (Cluster or other)
- Manufacturer
- Technological Description
- Cores CPU/GPUs
- CPU/GPU Tech Processor Type
- # Processors Cores by Processors (CPUs and GPUs)
- Interconnection
- RAM per Node (GiB)
- Total Memory (GiB)
- Year of implementation
- # Nodes
- # Racks
- Cooling Type
- Power consumption (kW)
- Operating System
- Theoretical TeraFLOPS (R_{peak})
- Real TeraFLOPS (Only with HPL, R_{max})
- Applications
- Number and Type of Users
- Contact
- Seller
- Storage Information
    - Manufacturer
    - Description
    - Capacity (TB)

Once the information has been processed, its visibility is verified (for example, if the owning institutions publicly display the information and if there are doubts, it is requested directly from the entities that own the machine). Then, the information is confronted, and the measured and analyzed results are collected to publish in different SCALAC reports linked with the HPC Observatory for Latin America and the Caribbean and used in further analysis.

Considering this information, information was collected from about 127 robust infrastructures in Latin America and the Caribbean for the last report and 25 newest for this report. First, those infrastructures whose information, despite being verified, could not be public (as mentioned above, military projects, private companies, and strategic sectors, among others) were eliminated. Likewise, it has been decided to eliminate from this report infrastructure that is not in minimum operation. However, it was in the installation testing phase at the time of the previously published report (1.3). In addition, obsolete infrastructures and those not installed or updated before 2012 were eliminated[31]. Technically, a relationship was measured between the number of nodes, number of processors per node, number of cores per node, and minimum performance, whether theoretical or measured, and verified observing the technology was considered. This listing only considers compute-related measurements and is limited to a minimum theoretical performance of 50 TeraFlops (TFlops).

For this report, the metrics and technical features presented in the HPC Infrastructure List are:

- Institution
- Country
- Institution type (Public, Private or mixed)
- Web (URL)
- Manufacturer
- Cores CPU
- Number of GPUs
- GPU Technology
- Processor Type
- Interconnection
- Year of Implementation
- Theoretical TFlops (GPU (FP32) + CPU)
- Measured TFlops (HPL using LinPack[32])

[31] In this sense, it is important to mention that in the region there are a dozen infrastructures installed between 2010 and 2012 that have been continually updated and strengthened in the observation window with considerable use.

[32] https://top500.org/project/linpack/

This report includes new infrastructures and updates that increased or decreased the installed capacities in the previous document.

It is also important to remember that the list only considered information that could be verified. The different entities often provided erroneous information (critically), so the technical committee decided to remove it from the list. In the same way, although there were institutions that sent information on infrastructure tenders and in the purchasing process, for this report, only installed capacity until September 1, 2024, was considered, as were upgrades of the information in the 1.3 document.

1.2. About SCALAC

SCALAC is the advanced computing system for Latin America and the Caribbean, created on March 1, 2012 with the *Declaration of Bucaramanga*, as a regional alliance with the support of national education and research networks grouped by RedCLARA. In 2018, it was formalized as an international civil society with legal headquarters in Costa Rica.

Strategically, SCALAC is the alliance that combines advanced computing capabilities and knowledge for Latin America and the Caribbean. Its vision is to be a non-profit organization that supports and promotes the development of advanced computing (called supercomputing, quantum computing) in Latin America and the Caribbean to guarantee not only the reduction of gaps, technological autonomy, and data sovereignty but also collaboration around regional needs and expectations and equal integration with global partnership as peers.

1.2.1 Organization of SCALAC

SCALAC is a non-profit organization (Non-Profit Association in Costa Rica (Affiliation Number: 3-002-788193) mainly comprised of two governance bodies: the board of chairs and the council.

The Council or (**directive council)** is the governing body with the authority to issue directives or mandates with civil and legal responsibility. The council has the power to make decisions that must be followed (Decision-Making Authority, Implementation Focus: Their role often involves ensuring that specific goals or objectives are met through the directives they issue and Accountability)

The 2024-2026 Council is made up of the following members:

- Philippe O. A. Navaux (Brazil), President
- Salma Jalife (México), Vicepresident
- Alvaro de la Ossa (Costa Rica), Secretary
- Carla Osthoff (Brazil), Treasurer
- Harold Castro (Colombia), First Vocal
- Nicolas Wolovick (Argentina), Second Vocal
- Luis Eliecer Cadenas (Venezuela - RedCLARA), Fiscal

The board of chairs (**board**) is the executive and leadership body responsible for overseeing its management and execution and ensuring that it operates in the best interests of its shareholders or stakeholders. This board plays a crucial role in setting SCALAC's strategic direction and policies.

The 2024-2026 Board is made up of the following members:

- Carlos J. Barrios (SC3UIS, Colombia), General Chair
- Esteban Meneses (CENAT-CNCA, Costa Rica), Vice General Chair
- Ginés Guerrero (NLHPC, Chile), Strategic Infrastructure and Services Chair
- Lizette Robles (CADS-UDG, México), Communication Chair
- Esteban Mosckos (UBA, Argentina), Training, Outreach and Education Chair
- Isidoro Gitler (ABACUS-CINVESTAV, México), I+D+i Chair
- Carlos González (RedCLARA, Colombia), RedCLARA Liason

Finally, the steering board committee to support the action of the board is made by:

- Sergio Nesmachnow (UdR, Uruguay)
- Robinson Rivas (UCV, Venezuela)
- Antonio Tadeu Gomes (LNCC-SINAPAD, Brazil)
- Moisés Torres (RedMexSu – CUDI, México)
- Claudio Chacon (CEDIA, Ecuador)
- Dennis Cazar (UNQ, Ecuador)

1.3. Organization, Design and Publication

The design and published graphical edition of this version of the report is lead by Lizette Robles Dueñas and the CADS-UDG Team.

2. HPC Robust Infrastructure List

The following list does not rank capabilities, performance, or supremacy. It only presents the capacities by country, according to the methodology explained above and the guidelines followed by the SCALAC team. This document organized the infrastructure into two main groups: 1. robust (verified) infrastructures with high-performance computing capabilities above 100 TFlops. and 2. robust infrastructures worth mentioning but below 100 TFlops. On the other hand, we have also created a table that considers two additional values: energy consumption and access to the advanced technology network connected by an NREN or by RedCLARA.

Infrastructures Type 1.

Institution type	Platform Name /iD	Manufacturer	Description	Cores CPU	# GPUs	GPU Tech	Processor Type	Interconnexion	Year	Theoretical Tflops (GPU (FP32) + CPU)	TFlops (HPL)
Argentina											
Servicio Meteorológico Nacional											
Public	CLEMENTINA XXI	LENOVO	Lenovo ThinkSystem SD650 V3, Xeon Max 9462 32C 2.7GHz, Infiniband NDR 400	43008	296	Intel Data Center GPU Max Series (Ponte Vechio)	Xeon Max 9462 32C 2.7GHz	Infiniband NDR400	2023	3715,89	3973,12
CCAD-UNC											
Public	MendietaF2	Supermicro	Hybrid Cluster	440	44	NVIDIA A30	Intel Xeon E5-2680v2	Infiniband 40	2021	463,05	
	Serafín	Supermicro	Cluster de CPU AMD	4096	0	None	AMD EPYC 7532	Infiniband 100	2021	314,57	
Brasil											
Petróleo Brasileiro S.A											
Mixte	PÉGASO	EVIDEN	Hybrid Cluster	233856		NVIDIA A100	AMD EPYC 7513	Infiniband HDR	2022	43008	19527,68
	DRAGÃO	EVIDEN	Hybrid Cluster	188224		NVIDIA TESLA V100	Xeon Gold 6230R	Infiniband EDR	2021	14346,24	9195,52
	GAIA	DELL EMC	Hybrid Cluster	84480		NVIDIA A100	AMD EPYC 74F3	Infiniband	2023	14059,52	7137,28
	ATLAS	EVIDEN	Hybrid Cluster	91936		NVIDIA TESLA V100	Xeon Gold 6240	Infiniband EDR	2020	9062,4	4485,12
	GEMINI	DELL EMC	Hybrid Cluster	42240		NVIDIA A100	AMD EPYC 74F3	Infiniband	2023	7024,64	3952,64
	FÊNIX	EVIDEN	Hybrid Cluster	60480		NVIDIA TESLA V100	Xeon Gold 5122	Infiniband EDR	2019	5498,88	3235,84

Institution type	Platform Name /iD	Manufacturer	Description	Cores CPU	# GPUs	GPU Tech	Processor Type	Interconnexion	Year	Theoretical Tflops (GPU (FP32) + CPU)	TFlops (HPL)
SiDi											
Private	IARA	NVIDIA	Hybrid Cluster	24800		NVIDIA A100	AMD EPYC 7742	Infiniband	2021	4229,12	37053,84
Software Company MBZ											
Private	NOBZ1	LENOVO	Multicore Cluster	80640			Xeon Platinum 8280	100G Ethernet	2022	7137,28	3635,2
	A16A	LENOVO	Multicore Cluster	61440			Xeon Gold 6252	100G Ethernet	2021	4229,12	2140,16
Laboratório Nacional de Computação Científica - LNCC											
Public	Santos Dumont	Eviden	Hybrid Cluster. Multicore and many-core architecture: CPUs, GPUs and MICs. CUDA-cores and 5120 Tensor-cores. NVIDIA Link interconnexion GPUs	4752	198	NVIDIA K40	Intel Xeon E5-2695v2	Infiniband FDR 56	2014 2019	942,63	4
				1296	54	XEON PHI	Intel Xeon E5-2695v2			133,74	
				240	0	None	Intel Xeon Ivy Bridge				
				11808	0	None	Intel Xeon Cascade Lake Gold 6252			793,49	
				1728	0	None	Intel Xeon Cascade Lake Gold 6252			116,12	
				4512	376	NVIDIA V100 32 GB	Intel Xeon Cascade Lake Gold 6252			5567,20	
				40	8	NVIDIA V100 16 GB	Intel Xeon Skylake Gold 6148			115,07	
				12096	0	None	Intel Xeon E5-2695v2			232,24	
			TOTAL	**36472**	**636**					**7900,52**	**51122,4**
Centro Nacional de Processamento de Alto Desempenho em São Paulo											
Pubic	USP 1	DELL	Hybrid Cluster	9376	10	NVIDIA A100	AMD EPYC 7662	Infiniband	2021	235,96	
					0	None	AMD EPYC 7662			475,136	
					0	None	AMD EPYC 7H12			21,2992	
					0	None	AMD EPYC 7443		2023	96,3072	
			TOTAL	9376	10					**828,70**	

Institution type	Platform Name /iD	Manufacturer	Description	Cores CPU	# GPUs	GPU Tech	Processor Type	Interconnexion	Year	Theoretical Tflops (GPU (FP32) + CPU)	TFlops (HPL)
Chile											
Laboratorio Nacional de Supercomputación/Universidad de Chile - NLHPC											
Public	Lefrararu -EPU	LENOVO	Multi Hybrid Cluster (One common access)	7104	0	None	AMD EPYC 9754	InfiniBand NDR 400	2024	497,66	270
					12	AMD Instinct MI210	AMD EPYC 9224			278,88	
					0	None	AMD EPYC 9224			7,68	
	Guacolda	DELL		2852	4	NVIDIA V100 32 GB	Intel Xeon Gold 6152		2019	61,91	196
					2	AMD Instinct MI100	AMD EPYC 7713			54,39	
					0	None	Intel Xeon Gold 6152			141,92	
					0	None	Intel Xeon Gold 6152			26,61	
	Lefrararu - ONE	HPE		2640	0	None	Intel Xeon E5-2660 v2		2014	0,352	44
			TOTAL	**9956**	**18**			(Including all partitions)		**1069,41**	**510**
Colombia											
Policía Nacional de Colombia											
Public	POL 1	HPE Cray	Hybrid Cluster	1024	16	NVIDIA H100 80 GB	EPYC AMD 9354	Infiniband 100	2024	1178,49	
Seguridad Nacional Colombia											
Public /Militar	SNA 1	HPE Cray	Hybrid Cluster	1024	16	NVIDIA L40S 48 GB	EPYC AMD 9354	Infiniband 100	2024	1572,09	
Universidad de Ibagué*											
Public	1	HPE Cray	Hybrid Cluster	2048	16	NVIDIA H100 80 GB	EPYC AMD 9354	Infiniband 100 Switch de 40 puertos	2024	1284,99	

Institution type	Platform Name /iD	Manufacturer	Description	Cores CPU	# GPUs	GPU Tech	Processor Type	Interconnexion	Year	Theoretical Tflops (GPU (FP32) + CPU)	TFlops (HPL)
Universidad de Cartagena											
Public	PACCA	HPE	Hybrid Cluster	1328	1	NVIDIA A100	Intel Xeon-Gold 5315Y	Infiniband	2023	67,37	
						None	Intel Xeon-Gold 5320				
						None	Intel Xeon-Gold 5317				
SC3UIS											
Public	GUANE-1	HPE	Hybrid Cluster	1440	40	NVIDIA Tesla M2050 3GB	Intel Xeon CPU E5645	Infiniband 40	2014	42,352	**105**
					24	NVIDIA Tesla M2050 3GB	Intel Xeon CPU E5640			25,23072	
					64	NVIDIA Tesla M2075 5G	Intel Xeon CPU E5645			67,7632	
	FELIX	HPE	Hybrid Cluster	32	2	NVIDIA GeForce GTX Titan X 12GB	Intel Xeon CPU X7560	Ethernet	2019	22,09056	
	KTHOR	HPE	Hybrid Cluster	64	2	NVIDIA TESLA K20	Intel Xeon CPU E7-8867 v3	Ethernet	2016	10,46	
	YAJE	HPE	Hybrid Cluster	6	1	NVIDIA GeForce GTX Titan X 12GB	Intel Xeon CPU E5-2609 v3	Ethernet	2018	11,0824	
	EXA	DELL	Clúster AMD	128	1	AMD Instinct MI210	AMD EPYC 9554		2023	35,2976	27,716
		Supermicro		128	2	AMD Instinct MI210	AMD EPYC 9534		2023	57,8976	47,781
	GUANE-SC3 ALL		**TOTAL**	**1798**	**136**	(Including Kthor, Yaje and Felix as node partitions)				**272,17**	**180,497**

Institution type	Platform Name /iD	Manufacturer	Description	Cores CPU	# GPUs	GPU Tech	Processor Type	Interconnexion	Year	Theoretical Tflops (GPU (FP32) + CPU)	TFlops (HPL)
BIOS											
Private	Asimov	HPE	Multicore Nodes	144	0		Intel(R) Xeon(R) CPU E5-2670	Infiniband	2013	2,9952	
			Big Memory Node	32	0		Intel(R) Xeon(R) CPU E5-4650			0,6912	
			Hybrid Cluster	32	4	GPU Nvidia Tesla K20	Intel(R) Xeon(R) CPU E5-2670			16,4656	
			Hybrid Cluster	60	0		Intel Xeon Phi Coprocessor 5110P			4,032	
				16	0		Intel(R) Xeon(R) CPU E5-2670			0,6656	
	Tayra	Inspur	Multicore Nodes	560	0		Intel Xeon 6132	Giga Ethernet	2019	46,592	
			Big Memory Node	176	0		Intel Xeon E7-8880 v4			6,1952	
			MIC Nodes	256	0		Intel Xeon PHI 7210			10,6496	
			GPU Nodes	84	12	TESLA P100 NVLINK	Intel Xeon 6132			134,1888	
			TOTAL	**1360**	**16**					**222,47**	**0**
Costa Rica											
Centro Nacional de Alta Tecnología (CeNAT)											
Mixte	Kabré	DELL	Hybrid Cluster	2608	4	NVIDIA K40	Xeon Silver	10 Gb	2016	17,16	100
					4	NVIDIA V100 32 GB	Xeon Gold			56	
					0	None					
					0	None					
		Supermicro			0	None					
					0	None					
					0	None					
					0	None					
			TOTAL	2608	8				(Including all partitions)	73,16	

Institution type	Platform Name /iD	Manufacturer	Description	Cores CPU	# GPUs	GPU Tech	Processor Type	Interconnexion	Year	Theoretical Tflops (GPU (FP32) + CPU)	TFlops (HPL)
Ecuador											
CEDIA											
Private	1	NVIDIA	3 nodos de computo DGX NVIDIA A100	384	24	NVIDIA A100 40 GB	AMD EPYC 7742	Ethernet	2019	495,64	
México											
Universidad Nacional Autónoma de México - UNAM											
	1	HPE Cray	Hybrid Cluster	8192	64	NVIDIA H100 80 GB	AMD EPYC 9354	Infiniband 100	2024	5139,96	
Centro de Análisis de Datos y Supercómputo/Universidad de Guadalajara (UDG-CADS)											
Public	Leo Atrox	FUJITSU	Hybrid Cluster	5400	2	NVIDIA TESLA P100	Intel Xeon-6154	Intel Omni-Path	2018	537	
					0	None	Intel Xeon-6154				
					0	None	Intel Xeon-6154				
ABACUS - Laboratorio de Matemáticas Aplicada y Cómputo de Alto Rendimiento del CINVESTAV											
Public	ABACUS 1	SGI Silicon Graphics Inc	Hybrid Cluster	7680	100	NVIDIA K40m	Intel Xeon 2697v3	Infiniband FDR	2016	800,40	429
Universidad de Sonora											
Public	Ocotillo	DELL	Hybrid Cluster	1168	6	None	AMD	Mellanox Infiniband FDR Connectx-3	2018		51
							Intel		2018		
Centro de Investigación Científica y de Educación Superior de Ensenada* - CICESE											
Public	1	HPE	Multicore Cluster	656	0	None	Intel con 20 núcleos	InfiniBand	0		
	2	Supermicro	Multicore Cluster	720	0	None	Intel Xeon con 24 cores	InfiniBand	0		

Institution type	Platform Name /iD	Manufacturer	Description	Cores CPU	# GPUs	GPU Tech	Processor Type	Interconnexion	Year	Theoretical Tflops (GPU (FP32) + CPU)	TFlops (HPL)
Instituto Nacional de Investigaciones Nucleares											
Public	ININ 1	NVIDIA		40	8	NVIDIA V100 32 GB	Xeon E5-2697v3		0	113,16	
Centro Nacional de Supercómputo del IPICYT											
Public	1	BULL ATOS	Multicore Cluster	2736	16	NVIDIA P100	Intel Xeon Skylake 6130	Gigabit	2017	333,73	
Laboratorio Nacional de Supercómputo del Sureste - BUAP*											
Public	Cuetlaxcoapan	FUJITSU	Multicore Cluster	2568	0	None	Intel Xeon Haswell	Infiniband FDR 56	2015	0	
	Centepetl	FUJITSU	Hybrid Cluster	3060	8	NVIDIA GTX 1080 Ti 11GB	Intel Knights Landing MIC	Omnipath	2015	135	
Universidad Autónoma del Estado de México*											
Public	1	TYAN	MultiCore Cluster	0	0	None	Intel Xeon Gold 6148	Infiniband/ Ethernet	2022	18,43	
	2	DELL	Hybrid Cluster	0	19	NVIDIA	None	Infiniband/ Ethernet	2024		
Uruguay											
Centro Nacional de Super computación											
Public	CNS-1	HPE	Hybrid Cluster	2240	29	NVIDIA P100	Xeon Gold 6138	Ethernet 10 Gbps	2018	326,02	
					2	NVIDIA A100	Xeon Gold 6138			44,12	
					0	None	Xeon Gold 6138			7,68	
	CNS-2	DELL			4	NVIDIA A40	AMD EPYC 7642	Ethernet 10 Gbps	2018	177,8624	
			TOTAL	**2240**	**35**			(Including all partitions		**555,68**	

Infrastructures Type 2.

Institution	Country	Institution type	Name /iD	Manufacturer	Description	Cores CPU	# GPUs	GPU Tech	Processor Type	Interconnexion	Year	Theoretical Tflops (GPU (FP32) + CPU)	TFlops (HPL)
Universidad Mayor de San Simón	Bolivia	Public	UM 1	DELL	Hybrid Cluster	480	2	NVIDIA TESLA	None	Infiniband 100	2020	28	7
Universidade Federal Do Pará	Brasil	Public	UFP 1	HPE	HPE DL380 Gen10 (head-node), HPE Apollo 4510 (Storage-node), GPU NVidia Tesla V100 32 GB e HPE Apollo r2200 (Compute-node)	812	1	NVIDIA V100 32 GB	Intel Xeon-Gold 5119T	Infiniband, Gigabit Ethernet	2021	35,87	
Laboratório Multiusuário de Computação Científica LCC CENAPAD-MG*	Brasil	Public	LNCC_CP 1	No Provided	No Provided	424	0	None	Intel Xeon Quadcore X5335	Infiniband Voltaire ISR 2012 20	2012		9
Universidad de los Andes Colombia	Colombia	Private	1	DELL	Dell EMC PowerEdge R740	64	6	NVIDIA RTX6000	Intel Xeon Gold 6226R	Infiniband, Switch Mellanox Quantum HDR	2022	1,9	
					Dell EMC PowerEdge R6525	144		None	AMD EPYC 7402			7,32	
					Dell EMC PowerEdge R6525	96		None	AMD EPYC 7402			4,88	
					Dell EMC PowerEdge R6525	48		None	AMD EPYC 7402			2,44	
					Dell EMC PowerEdge R640	400		None	Intel Xeon Gold 6242R			13,75	
					TOTAL	**752**				(Including all partitions)		**30,29**	

Institution	Country	Institution type	Name /iD	Manufacturer	Description	Cores CPU	# GPUs	GPU Tech	Processor Type	Interconnexion	Year	Theoretical Tflops (GPU (FP32) + CPU)	TFlops (HPL)
SC3UIS	Colombia	Public	1	HPE	FELIX	32	2	NVIDIA GeForce GTX Titan X 12 GB	Intel Xeon CPU X7560	Ethernet	2019	13,35	
					THOR	64	2	NVIDIA TESLA K20	Intel Xeon CPU E7-8867 v3		2016	8,13	
					YAJE	6	1	NVIDIA GeForce GTX Titan X 12 GB	Intel Xeon CPU E5-2609 v3		2018	6,89	
Fuerza Aérea Colombiana	Colombia	Public / Militar	1	HPE Cray	Cluster de Alto desempeño de 8 nodos de Cray XD2000´s, con doble procesador 9354 de AMD por nodo	512	0	None	EPYC AMD 9354	Slingshot 200. Switch Slingshot	2023	52	

(Platforms with (*) must be verified)

3. Use and Critical Downtime in HPC Infrastructures

While specific usage percentages can vary depending on the institution, political decisions, and physical infrastructure available, the demand for HPC resources is driven by the need to process large datasets and perform complex simulations and a mix of these factors.

While exact statistics on the usage percentage may not be readily available, the overall trend indicates a robust and growing reliance on HPC infrastructures across multiple sectors, reflecting their critical role in modern computational tasks. For the next table we use the following SLURM commands:

```
sreport cluster Utilization -p -t hours start=2023-01-01 end=2024-01-01
```

```
sreport cluster UserUtilizationByAccount start=2023-01-01 end=2024-01-01 format=u -Pn -t hours | awk -v threshold=1024 '$1 > threshold {count++} END {print " " threshold " horas-core:", count}'
```

Country	Institution	Infrastructure Name	% Use	% Downtime	Active Users
Type 1 Infrastructure					
	CCAD-UNC	Serafín	91.49%	20.98%	155
	CCAD-UNC	MendietaF2	83.38%	7.90%	62
	Exactas-UBA	CeCAR	21.07%	32.96%	29
	CIMEC	Pirayú	23.63%	64.81%	16
	LNCC	Santos Dumont	90.38%	3.81%	2939
	NLHPC	Leftraru	77.75%	1.52%	322
	UNIAndes	Hypatia	51.78%	0.54%	99
	UniCartagena	PACCA	31.17%	1.53%	3
	UIS	Guane-1	69.19%	11.50%	39
	UniGuadalajara	Leo Átrox	42.83%	46.14%	34
	CNS	ClusterUY	36.75%	2.05%	107
Infrastructure Type 2					
	CCAD-UNC	Eulogia	91.87%	19.76%	63
	CCAD-UNC	Mulatona	69.79%	37.57%	35

Although this list is really sparse in terms of the clusters listed before, it gives us a good idea of the use given to the clusters, and some of the common problems found in the installations. Some clusters are still in very low usage with very few users, while others show a good degree of usage, above 70%, with a fair amount of people actively using it. This measure strives to give a message to policy-making people behind the operations of the cluster. The infrastructure should be down as little as possible. From the remaining up-time, it has to be used intensively by as many scientists and technologists as possible. Measuring these quantities is a first step towards establishing a baseline for quality in the LAC cluster installations. In 2024 we will start collecting these measurements comprehensively and systematically to know where we are and where we are heading.

3.1 Bad Measurements

Despite the efforts and advice on the measurements of the last parameters to analyze the use of the platforms, we found several cases whose names we have omitted, first at the request of the reviewers of the document and second because the reading of them can be misinterpreted in this table.

We show three examples of 5 erroneous data collected, which are still mentioned because they show not only a bad configuration, but also the need to adjust the slurm-based usage metric.

The following first example shows the inconsistency between the time of use and the downtime for the number of users measured, this clearly indicates a bad configuration of the slurm.

~~XXX~~	~~XXXX~~	~~XXXX~~	~~94.14%~~	~~81.19%~~	~~47~~

The second following example has two interpretations of the analysis: one that despite the number of users, they are sending jobs outside the queue manager, which prevents the usage from being correctly measured, and second that the time down the platform for the measurement given does not correspond to the use.

~~XXX~~	~~XXXX~~	~~XXXX~~	~~14.04%~~	~~11.19%~~	~~107~~

And finally, the last example presented shows a wrong configuration of the slurm.

~~XXX~~	~~XXXX~~	~~XXXX~~	~~-9.04%~~	~~110.39%~~	~~7~~

These results raise a worrying situation, which is discussed a little more in the next section and concerns the need to monitor the configuration of the architectures in order to carry out transparently and adequate usage monitoring that allows, for example, not to underutilize infrastructure and sustain new investments.

4. Final Remarks

The list shows the verified high-performance computing capabilities in Latin America and the Caribbean, which is quite important, as the document shows. Beyond the machines recognized as Top500 (today, according to the June 2024 list, concentrated in Brazil and with the inclusion of the Argentine Clementina XXI machine), in the region, there are interesting infrastructures that mainly support hybrid research needs in scientific computing, data analytics, and artificial intelligence. In that order of ideas, machines to support Artificial Intelligence were deployed in the region for enterprises and universities.

The information and numbers provide a perspective on some problems of LAC clusters that are worth mentioning, in the hardware, software, and usage of these supercomputers.

Internal cluster fragmentation or excessive heterogeneity is an issue. Typical examples are in CeNAT (Ecuador) and SC3UIS (Colombia). This generates an infrastructure that is harder to maintain and precludes any possibility of scaling the apps running to the whole cluster, and simultaneously it ties the binaries to the lowest common denominator in terms of ISA for all the platforms involved in the cluster. Both characteristics are undesirable for producing good science and technology. Grants come in small amounts, usually meaning buying and adding what is available to the cluster.

Measuring basic uptime and usage numbers brought another dimension of problems. In many cases the installed resource management software was not set up or the cluster was not even functional. This shows that the act of measuring is already pushing some good practices to be implemented in the LAC supercomputers. Most of them could measure downtime, usage and amount of active users in 2023. Some supercomputers are on par with world-class installations in terms of <5% downtime and >70% of usage, while others show there is room to improve both quantities. Active users are also a good measure of utilization, and here we can see clearly closed vs open community clusters.

We are planning to bring more metrics for 2024, and it is important to remark that they are not tied to how big in terms of TFLOPS are the clusters (they are not), but more on how they are used and how many person-hours is behind. We understand that some of these numbers showing big downtimes, poor utilization or even small number of users had a good explanation in terms of the amount and frequency of grants, cluster infrastructure, and quantity of human resources dedicated to running and promoting the use of supercomputing to improve science and technology of LAC. We want these numbers to tell where we are and from that, to move forward.

Although there is no clear roadmap for implementing robust infrastructure for HPC in Latin America and the Caribbean, there is an essential investment in technology, which, as is the global trend, comes from the public sector. Equally important is the percentage of private investments. It is important to mention that although this list considers implementation, the relationship of performance and technology implemented, and the destination of use (academic, private), the effective use of the installed HPC platform and its relationship with performance will be considered in a third report.

Finally, while the information comparison yielded few surprises, it is concerning that performance may be inflated, real capabilities may be estimated or overvalued, and the method for calculating and measuring the performance of an HPC platform may need to be clarified. This issue is of particular concern to SCALAC, as misinformation can undermine the credibility of investment decision-makers and governments, erode the trust of the specialized international community in a specific country, and lead to erroneous investments and facilities that do not contribute to the responsible formation of robust capabilities to support high-performance computing.

5. Acknowledgments

The authors of this document express their gratitude to all the centers, universities, institutions, and companies that have responded to the request. Their contributions have been invaluable. The authors also extend their thanks to the integrators, technology providers and manufacturers, in alphabetical order: AMAZON, AMD, ATOS (EVIDEN), DELL, FUJITSU, GOOGLE, HPE, HUAWEI, IBM, INTEL, INSPUR, LENOVO, MICROSOFT, NVIDIA, SAMSUNG, SIEMENS and SUPERMICRO. They also thank the former individual and institutional members of the RISC-2 project and the Americas HPC Collaboration Alliance, mainly DOE-ARNL, Compute Canada Alliance, and DOE ORNL.

We also appreciate feedback from readers of previous reports, members of the HPC community, administrators of the platforms mentioned in this document, and interest groups.