

# SA-CERN 15-year celebration

*High-Performance Computing  
in South Africa*

*Advancing South Africa's Research and  
Innovation Capacity*

21 January 2025

A national initiative of the Department of Science  
and Innovation and implemented by the CSIR



science & innovation  
Department  
Science and Innovation  
REPUBLIC OF SOUTH AFRICA



# Importance of Advancing South Africa's Research and Innovation Capacity

- South Africa's ability to address urgent social and economic challenges—such as climate change, healthcare inequalities, and food security—depends on its innovation capacity.
- High-performance computing (HPC) is crucial for scientific discovery due to its ability to process large amounts of data, perform complex calculations and simulate real-world phenomena at unmatched speed and scale

## **South Africa faces the challenge of unequal access to advanced and emerging technology.**

- Infrastructure Costs and a shortage of skill
  - High-performance computing (HPC), artificial intelligence (AI), and big data analytics require substantial infrastructure investment.
  - Universities and Research centres, especially those that have been historically disadvantaged, struggle to afford world-class resources.
  - A significant lack of expertise in HPC and AI exists, particularly in rural and economically disadvantaged areas.
  - Limited training and exposure hinder individuals and institutions from maximising potential, worsening the digital divide and stalling research and innovation.

# CHPC Role: Advancing South Africa 's research and Innovation Capacity through Democratizing and Converging Emerging and Disruptive Technologies

Formula:  $R\&I\ Capacity = [(D \times E) + (C \times S)] \times I$

Where:

- **R&I Capacity:** Research and Innovation Capacity
- **D:** Democratization of Technologies (ensuring equitable access and inclusion)
- **E:** Empowerment (training and capacity building for researchers and institutions)
- **C:** Convergence of Technologies (integration of emerging and disruptive innovations)
- **S:** Synergy (collaborative efforts among academia, industry, and government)
- **I:** Infrastructure and Resources (HPC, funding, tools, and platforms)

Formula: **R&I Capacity = [(D × E) + (C × S)] × I**

- 1. Infrastructure and Resources: CHPC provides researchers, academic institutions, and industry stakeholders with world-class computational power, storage, and data processing capabilities.**
2. The democratisation of Technology: Through free or subsidised access, CHPC ensures that Universities and research institutions across South Africa and the Rest of Africa can access advanced technologies.
3. Bridging the Digital Divide: CHPC actively supports historically disadvantaged institutions and underserved communities by providing them with the tools and training to participate in cutting-edge research.
4. Capacity Building: The centre offers extensive training programs to build expertise in HPC and emerging technologies, creating a skilled workforce ready to drive innovation.
5. Catalyst for Collaboration: By acting as a hub for collaboration between academia, government, and industry, CHPC facilitates partnerships that lead to impactful solutions.

# CHPC is part of the Cyber-Infrastructure Triangle

**National Integrated Cyberinfrastructure System  
(NICIS)**

**SANReN  
(Networking Services)**



**CHPC  
(Computing Services)**

**DIRISA  
(Data Services)**

# CHPC Mandate under NICIS to support national research initiatives.

- ❑ CHPC is **national HPC facility** funded by the SA Government
- ❑ Department of Science and Innovation (**DSI**)
- ❑ **National Integrated Cyber-Infrastructure System (NICIS)**
- ❑ Administered by the **CSIR**
- ❑ Started operations in **June 2007** and based in Cape Town
- ❑ Until recently hosted the **largest HPC system in Africa**
- ❑ CHPC has total number of ~**35** employees
- ❑ CHPC has **Research, Technical** and **Operational** divisions

# 1. World-Class Infrastructure and Resources:

- **CHPC provides world-class computational power, storage, and data processing capabilities to researchers, academic institutions, and industry stakeholders**
  - High-Performance Cluster
  - GPU Cluster
  - Localised Cloud Platform
  - Tier 2 Node Facility (CERN - ALICE)
  - Highly Skilled and Dedicated Team

# 1. World-Class Infrastructure and Resources

- ❑ Lengau cluster since **7 March 2017** (First phase June 2016)
- ❑ Africa's first Petascale system  $\Rightarrow$  **1.029 PFlops = 1 029 TFlops**

(121<sup>st</sup> on Top500\* – June 2016)  
(127<sup>th</sup> on Top 500\* – June 2017)

## Lengau (HPC Cluster)

- 1386 nodes
- 32 832 cores
- 30 V100 GPUs
- 56Gbps IB interconnect
- 4PB Lustre storage
- ~1Pflop/s HPL



**Deployed in 2016/17**  
**Debut at 121 on Top500**

## Special Nodes

- 5 High memory nodes
- 30 GPU nodes

# 1. World-Class Infrastructure and Resources



**2008:** IBM Blue Gene/P; 0.8 GHz;  
4096 cores; **11.5 Tflops**



**2010:** GPU Cluster; 2.4 GHz; 96 cpu cores; 22 GPU's  
**16 Tflops**; 14 TB file system



**2011:** Tsessebe Cluster  
Upgraded 2.9GHz (4032 cores)  
**61.6 Tflops**

**2007:** IBM  
2.5  
cores  
(L)

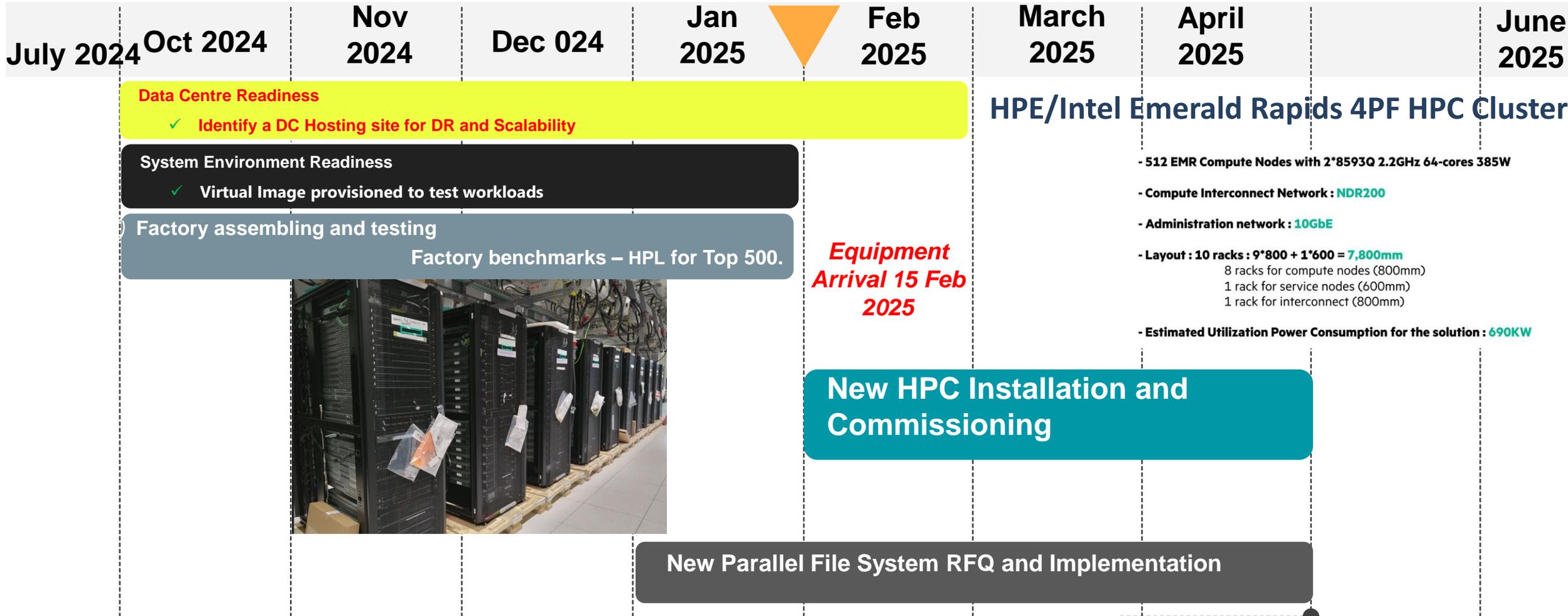


**2009 :** Sun Constellation Cluster; 2.9/3.0 GHz; 2684 cores  
**27 Tflops**; 480 TB Lustre File System

# 1. World-Class Infrastructure and Resources – New HPC Cluster Project Status

PROJECT  
START  
DATE

WE ARE  
HERE



## HPE/Intel Emerald Rapids 4PF HPC Cluster

- 512 EMR Compute Nodes with 2\*8593Q 2.2GHz 64-cores 385W
- Compute Interconnect Network : **NDR200**
- Administration network : **10GbE**
- Layout : **10 racks : 9\*800 + 1\*600 = 7,800mm**  
 8 racks for compute nodes (800mm)  
 1 rack for service nodes (600mm)  
 1 rack for interconnect (800mm)
- Estimated Utilization Power Consumption for the solution : **690KW**

PROJECT  
END DATE

# 1. World-Class Infrastructure and Resources

- ❑ **Graphical Processor Unit (GPU) Cluster: 30 NVIDIA V100's**
  - ❑ 6 nodes (36 Intel Gold 6150 CPU's; 2 V100's; PCIe)
  - ❑ 2 nodes (40 Intel Gold 6150 CPU's; 4 V100's; NVLink)
- ❑ Since **September 2018**
- ❑ Demand from CHPC Users for GPU Resources:
  - ❑ **Chemistry Molecular Dynamics (MD)**
  - ❑ **Resources for Machine Learning (ML)**
- ❑ Policies:
  - ❑ Dedicated GPU Queues through PBS Pro
    - ❑ gpu\_1, gpu\_2 and gpu\_4
  - ❑ 1 GPU Hour  $\approx$  40 CPU Core Hours (Allocation Accounting)



# 1. World-Class Infrastructure and Resources

## *Localized Managed Cloud*



### **OpenStack (Kayobe)**

- 1 monitoring node
- 3 controller nodes
- 32 compute nodes
- 8 SSD CEPH nodes
- 12 HDD CEPH nodes
- 2 Mellanox 100GbE SN2700 (Spine switches )
- 4 Mellanox 100Gbe SN3700 (Leaf switches)
- 2 Mellanox 1Gbe switches

### **Compute**

- 2944 CPU threads
- 12 TB RAM

### **Storage (CEPH)**

- 3.61 PB raw (60TB NVMe)
- 3-copy pools for OpenStack

### **Network**

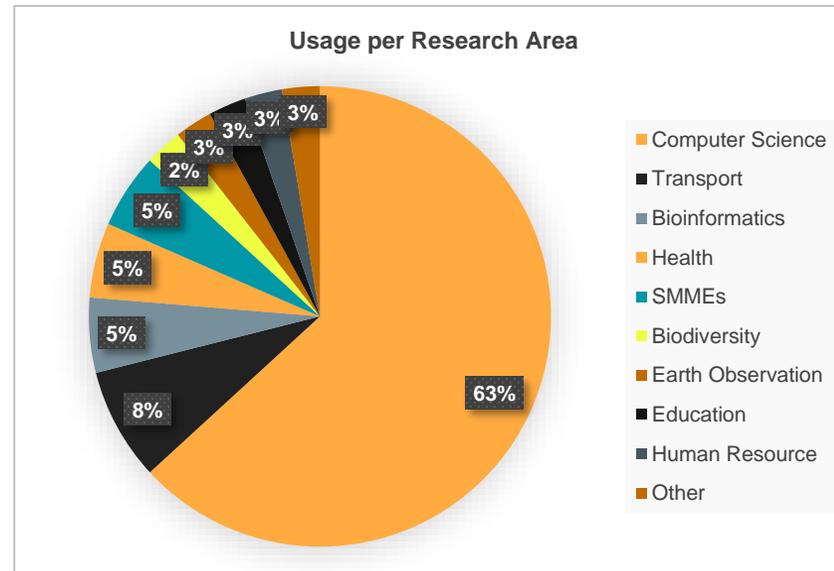
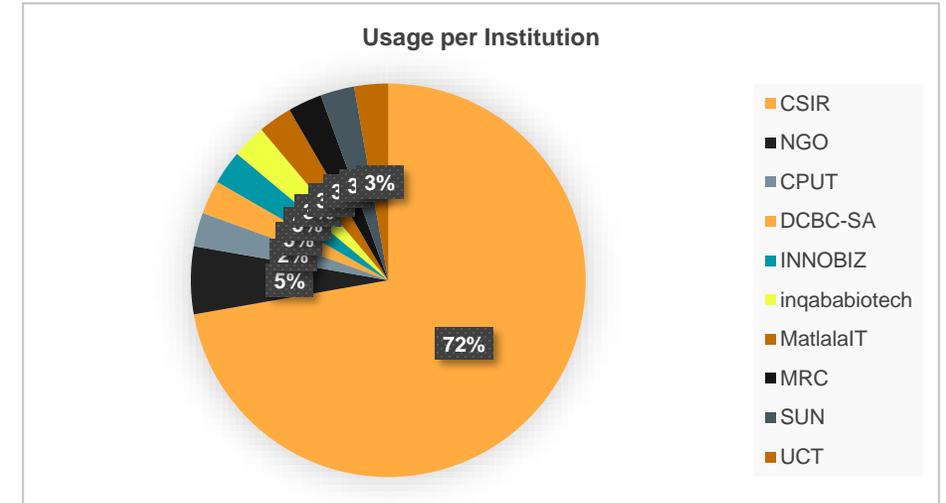
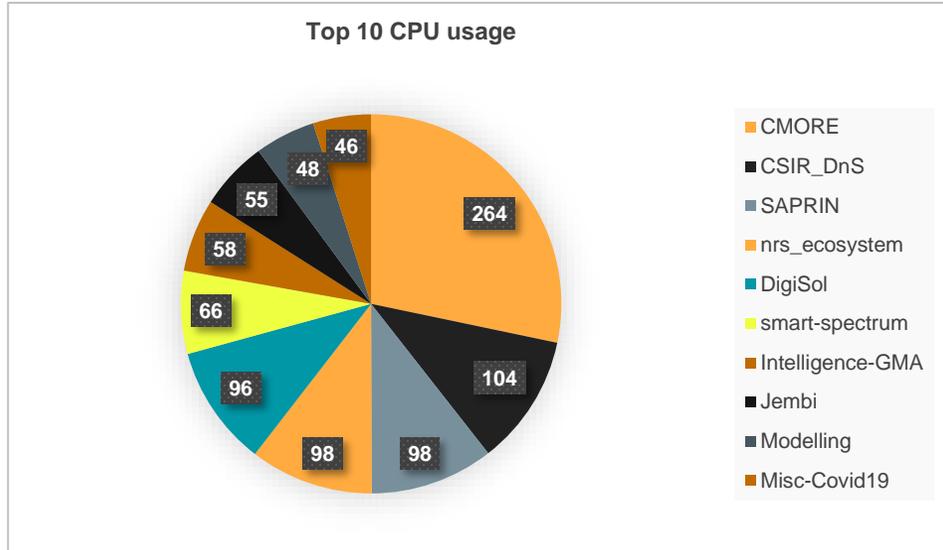
- 2x 25GbE per host
- 100GbE spine



**openstack**®

# Sebowa Production Cloud

There is a total of 55 approved projects hosted on Sebowa



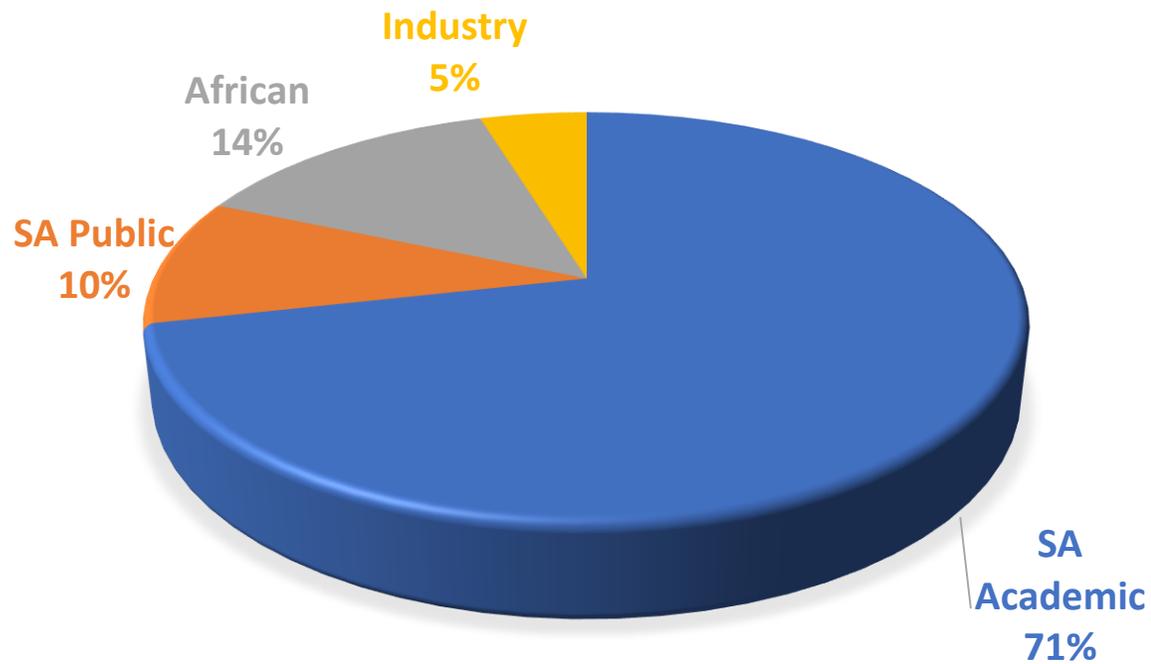
A national initiative of the Department of Science and Innovation and implemented by the CSIR.

Formula: **R&I Capacity = [(D × E) + (C × S)] × I**

1. Infrastructure and Resources: CHPC provides researchers, academic institutions, and industry stakeholders with world-class computational power, storage, and data processing capabilities.
2. **The democratisation of Technology: Through free or subsidised access, CHPC ensures that Universities and research institutions across South Africa and the Rest of Africa can access advanced technologies.**
3. **Bridging the Digital Divide: CHPC actively supports historically disadvantaged institutions and underserved communities by providing them with the tools and training to participate in cutting-edge research.**
4. **Capacity Building: The centre offers extensive training programs to build expertise in HPC and emerging technologies, creating a skilled workforce ready to drive innovation.**
5. Catalyst for Collaboration: By acting as a hub for collaboration between academia, government, and industry, CHPC facilitates partnerships that lead to impactful solutions.

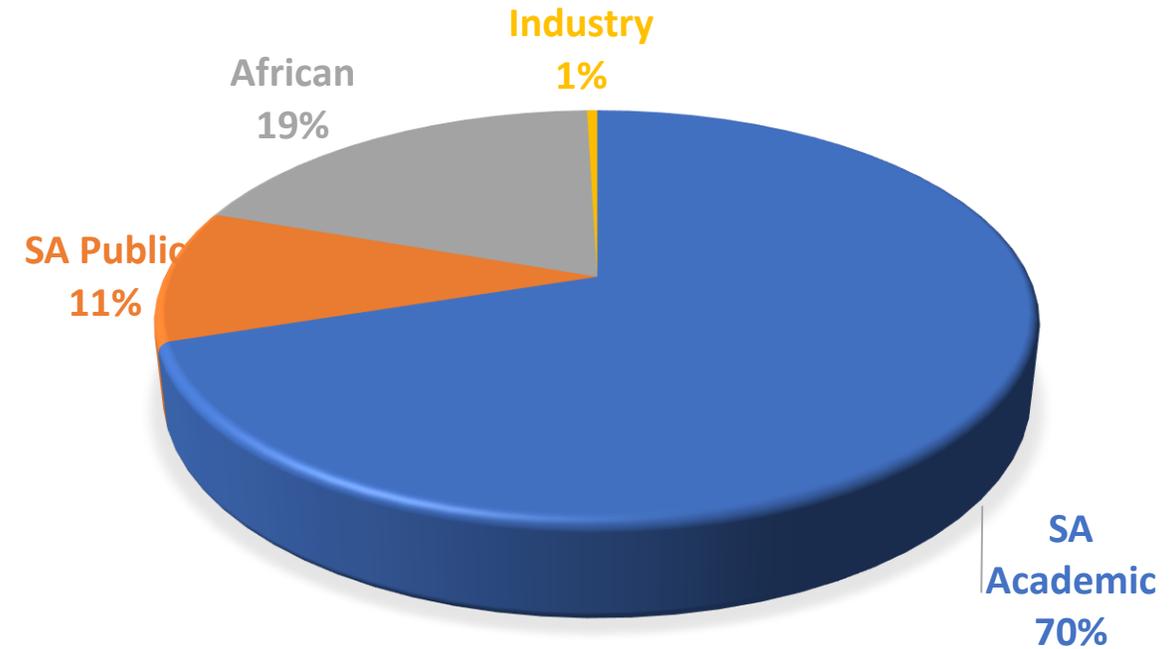
## 2. Democratising Technology: Free or Subsidised Access to HPC Resources to Academic Institutions and Research Centres

### CHPC User Categories *Total Active\* Programmes*



**# Active Programmes**

**Total: 320**

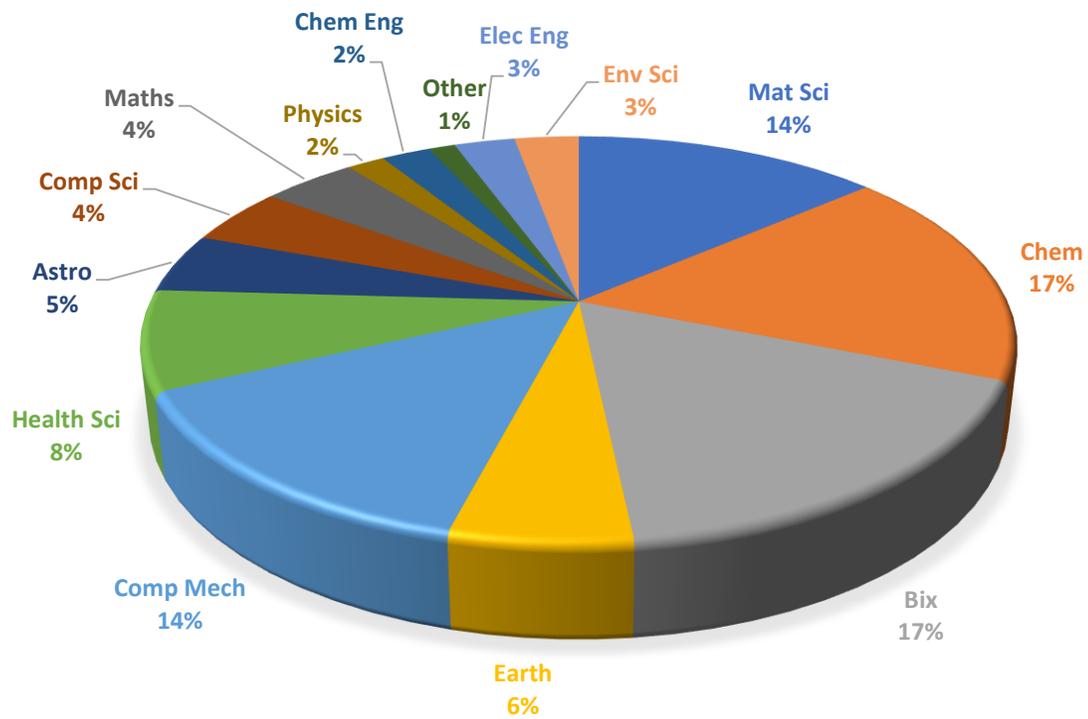


**# Hours Used (Million)**

**Total: 222 million**

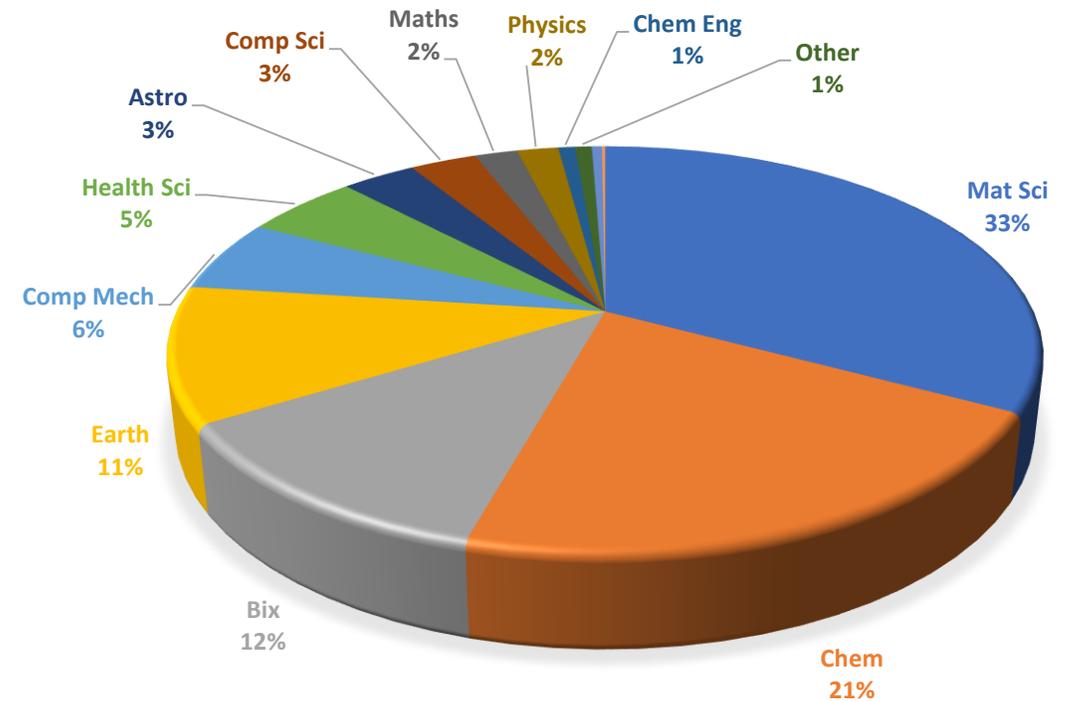
*\*Active refers to at least 1000 compute hours used over the relevant period.*

# CHPC Use: Disciplines *Total Active\* Programmes*



**# Active Programmes**

**Total: 606**

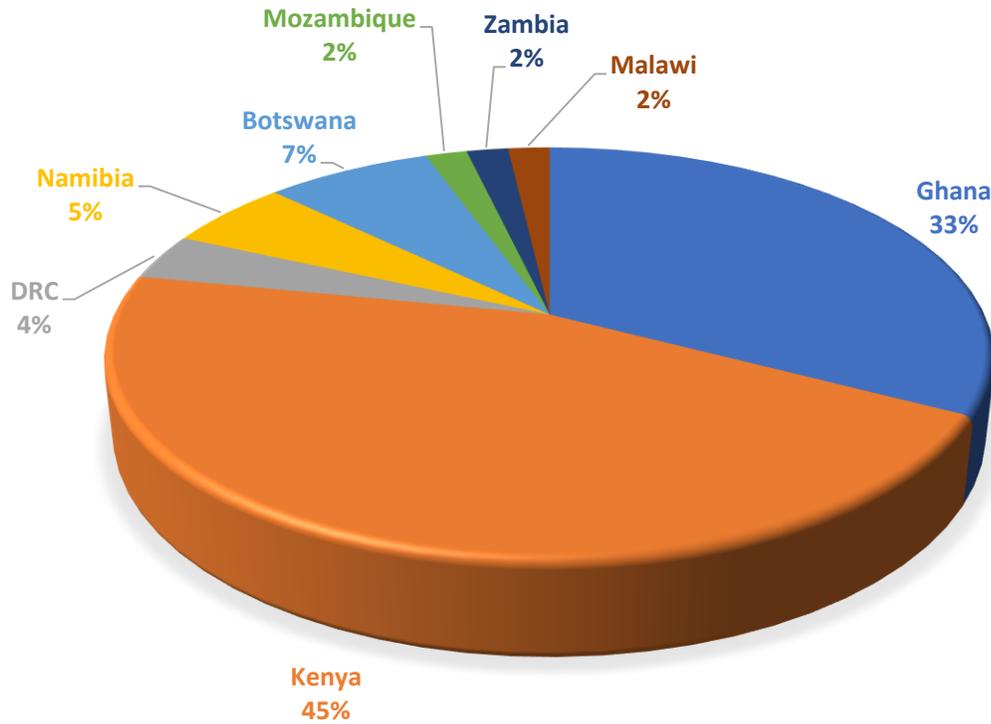


**# Hours Used (Million)**

**Total: 1419 million**

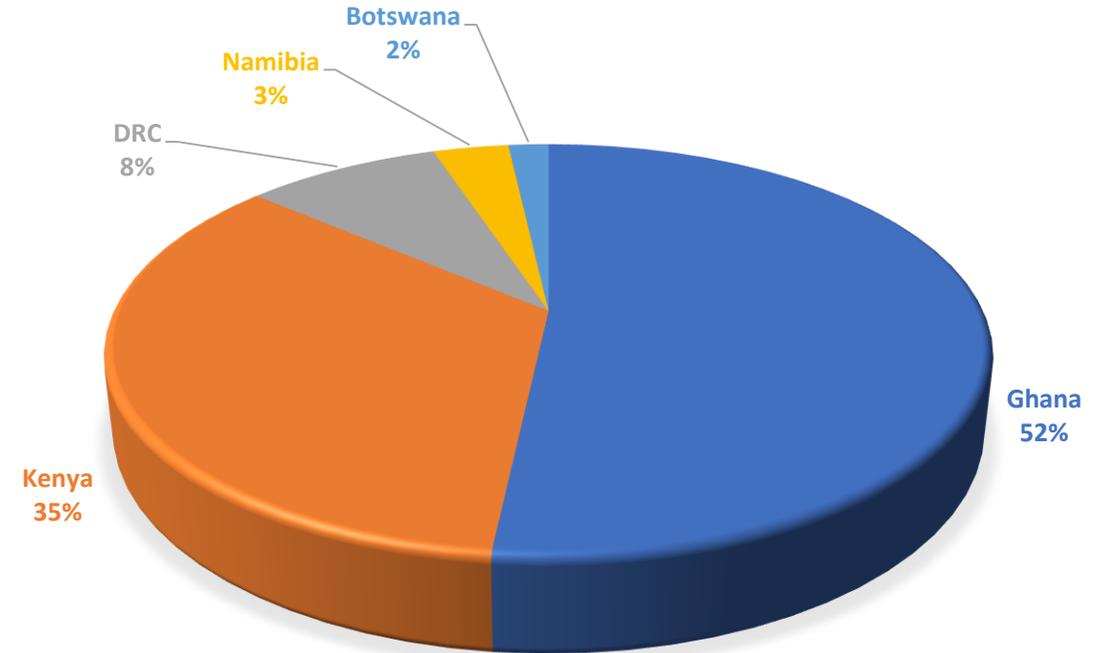
*\*Active refers to at least 1000 compute hours used over the relevant period.*

# CHPC Use: African Partners **Total Active\* Programmes**



**# Active Programmes**

**Total: 55**



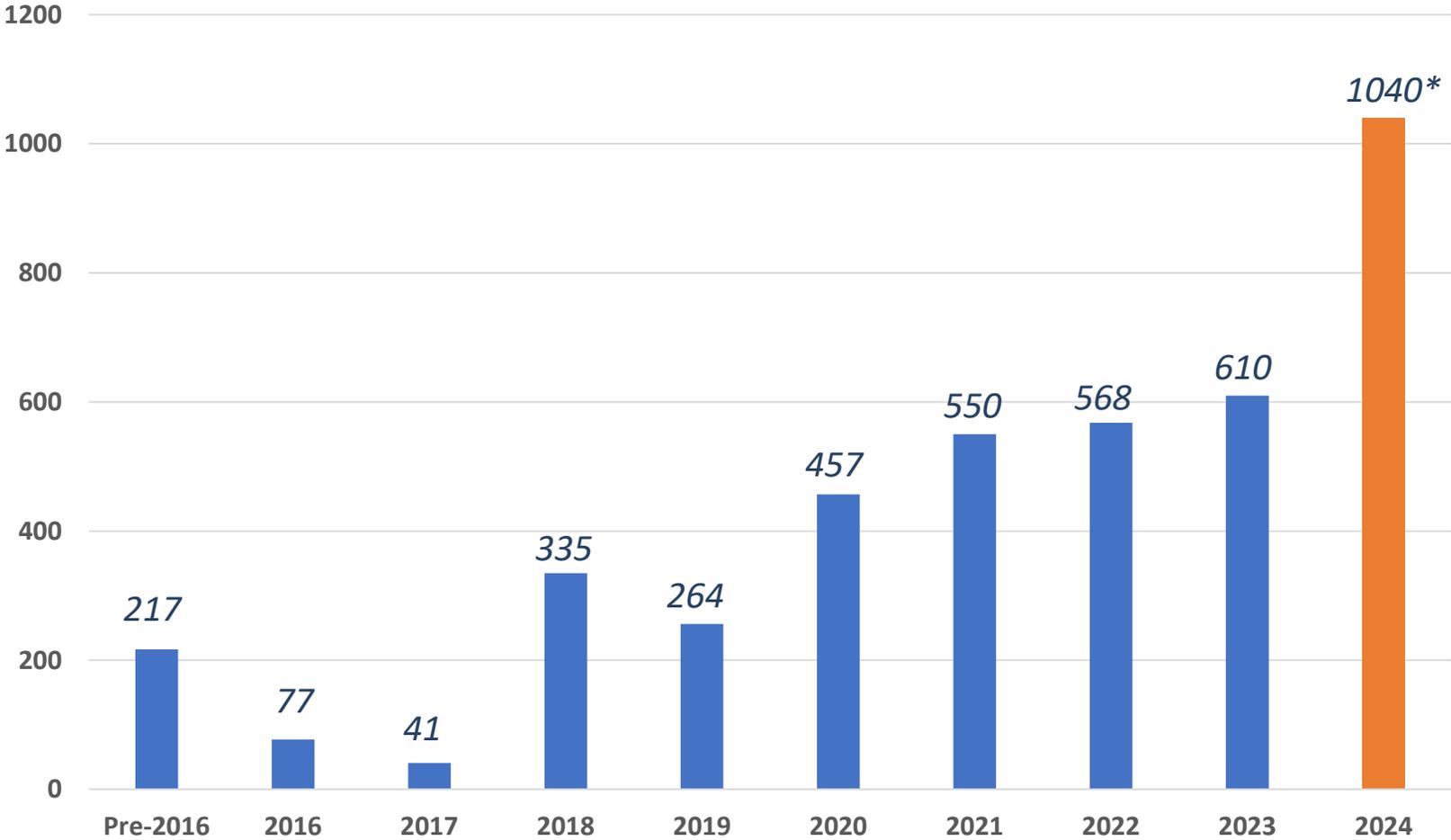
**# Hours Used (Million)**

**Total: 174 million**

*\*Active refers to at least 1000 compute hours used over the relevant period.*

# 2. Democratising Technology: Free or Subsidised Access to HPC Resources to Academic Institutions and Research Centres.

## Total Research Outputs\* Count Per Calendar Year

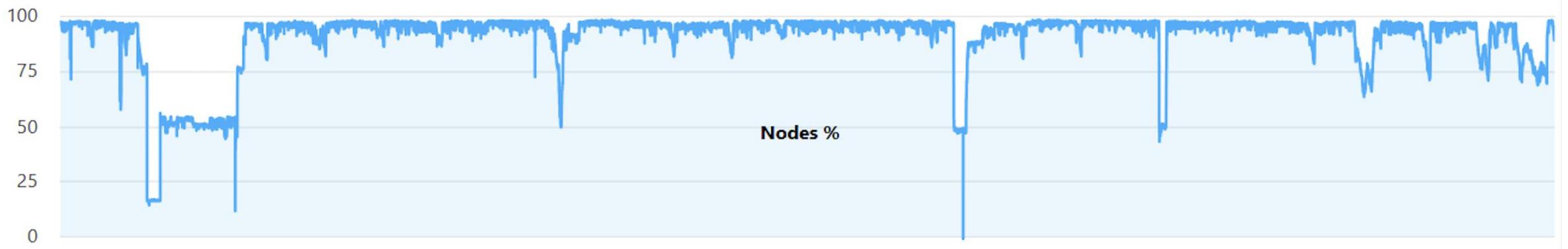


Total Number of Research Outputs\*:  
**4151**

\*Research Outputs = Journal Article, Book, Thesis (PhD, Master's, Hons), etc.

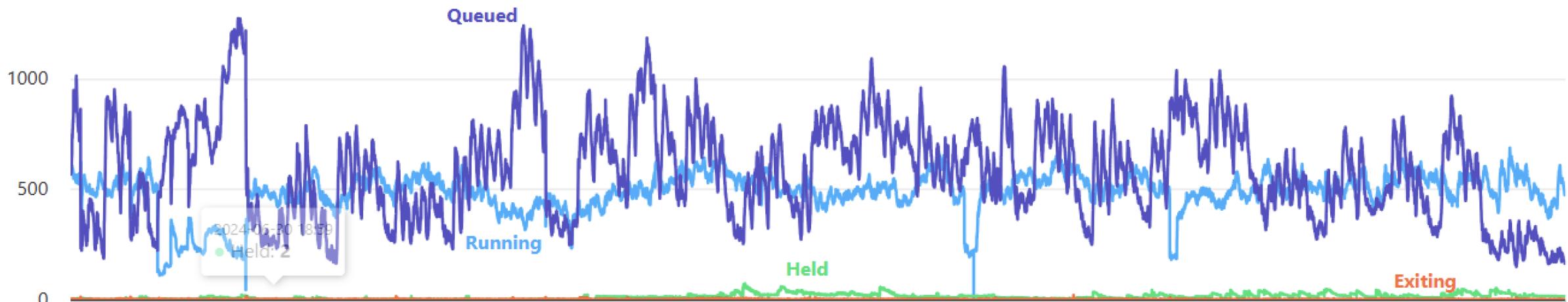
## 2. Democratising Technology: Free or Subsidised Access to HPC Resources to Academic Institutions and Research Centres

### CHPC Lengau Usage (Past 6 Months): CPU Cluster



4 June 2024

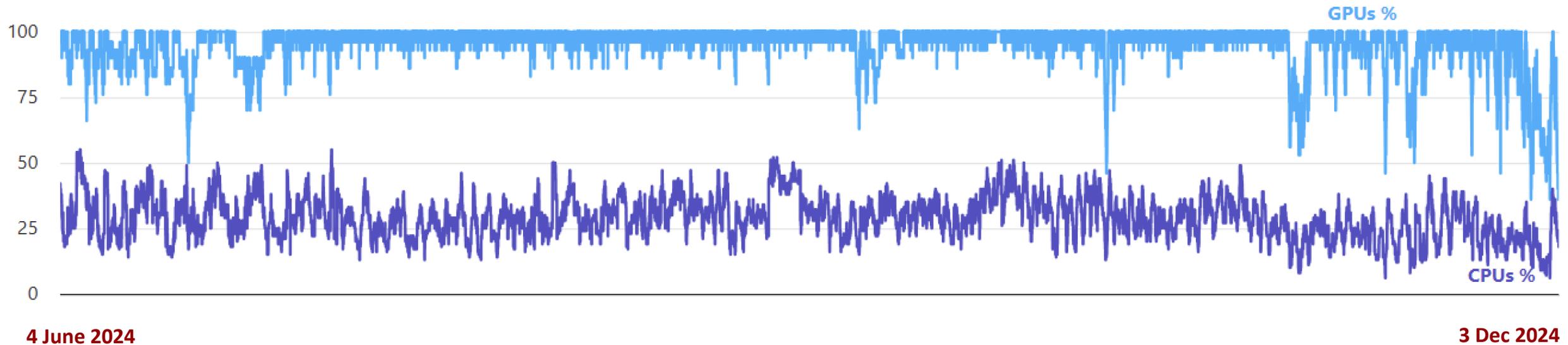
3 Dec 2024



2024-06-20 18:59  
Held: 2

## 2. Democratising Technology: Free or Subsidised Access to HPC Resources to Academic Institutions and Research Centres

### CHPC Lengau Usage (Past 6 Months): GPU Cluster



**3/4. Bridging the digital divide and Capacity building:** CHPC offers extensive training programs to build expertise in HPC and emerging technologies, creating a skilled workforce ready to drive innovation.

- Dedicated CHPC Training events
  - CHPC Coding Summer School
  - CHPC Winter School in Practical HPC
  - HPC Ecosystems Project
  - Student Cluster Competition (SCC)
  - Domain-specific workshops / Ad Hoc Training
- CHPC National Conference
- Women in HPC (WHPC) Program
- Outreach Programmes



**Our Student Cluster competition produced the most significant upset in years. Several milestones were achieved: the first-time SCC entrant, CPUT, outperformed nine other universities across South Africa. The competition also saw the first all-female team participate, along with several new universities entering, to name a few.**

27 January 2025 to 7 February 2025

Africa/Johannesburg timezone

Overview

Registration

Surveys

Contact

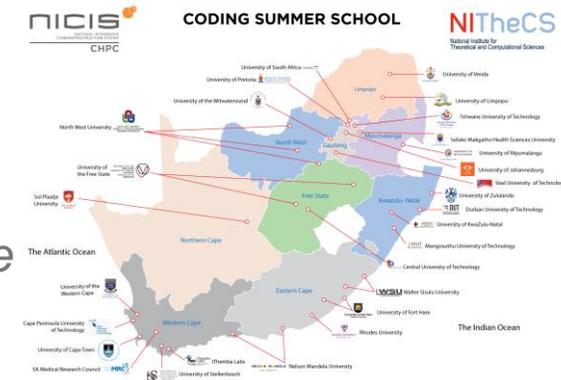
✉ [helpdesk@chpc.ac.za](mailto:helpdesk@chpc.ac.za)

*Train researchers across South Africa and Southern Africa the fundamentals of scientific computing and data science."*

## The 15th CHPC & 7th NITheCS Coding Summer School on Data Science and Computational Thinking

27 January - 7 February 2025

The Coding Summer School takes place in a physical form at various university and research locations around South Africa and Southern Africa. Students are required to attend one of the designated locations.



Formula: **R&I Capacity = [(D × E) + (C × S)] × I**

1. Infrastructure and Resources: CHPC provides world-class computational power, storage, and data processing capabilities to researchers, academic institutions, and industry stakeholders.
2. Democratisation of Technology: Through free or subsidized access, CHPC ensures that institutions across South Africa, regardless of their resources, can leverage advanced technologies.
3. Bridging the Digital Divide: CHPC actively supports historically disadvantaged institutions and underserved communities by providing them with the tools and training needed to participate in cutting-edge research.
4. Capacity Building: The center offers extensive training programs to build expertise in HPC and emerging technologies, creating a skilled workforce ready to drive innovation.
5. **Catalyst for Collaboration: By acting as a hub for collaboration between academia, government, and industry, CHPC facilitates partnerships that lead to impactful solutions.**

**5. Catalyst for Collaboration:** By acting as a hub for collaboration between academia, government, and industry, CHPC facilitates partnerships that lead to impactful solutions.

**Collaboration:** Facilitates partnerships between academia, government, and industry to foster innovation.

- Large - Scale Science Projects
  - SKA and SADC HPC Ecosystem Project
  - CERN ALICE Scientific Research Programme
  - SARA0 (South African Radio Astronomical Observatory)
  - SAWS (South African Weather Service)
  
- Establish Strategic Partnerships with Emerging Technology Communities
  - Establish CHPC as a Quantum Computing Hub for South Africa
  - Establish CHPC as the Strategic Partner in the Machine Learning / AI Community

**5. Catalyst for Collaboration:** By acting as a hub for collaboration between academia, government, and industry, CHPC facilitates partnerships that lead to impactful solutions.

## SKA and SADC: HPC Ecosystem

- The HPC Ecosystems Project is responsible for distributing and commissioning HPC resources
- Decommissioned Tier-1 HPC systems are repurposed into smaller, mid-tier HPC systems and allocated to partner institutions within three broad categories:
  - Square Kilometre Array partner countries - (Ghana, Kenya, Mozambique, Madagascar, Mauritius, Zambia, Botswana, Namibia, South Africa);
  - Southern African Development Community (countries in the Southern African region); and
  - South African Historical Disadvantage Institutions

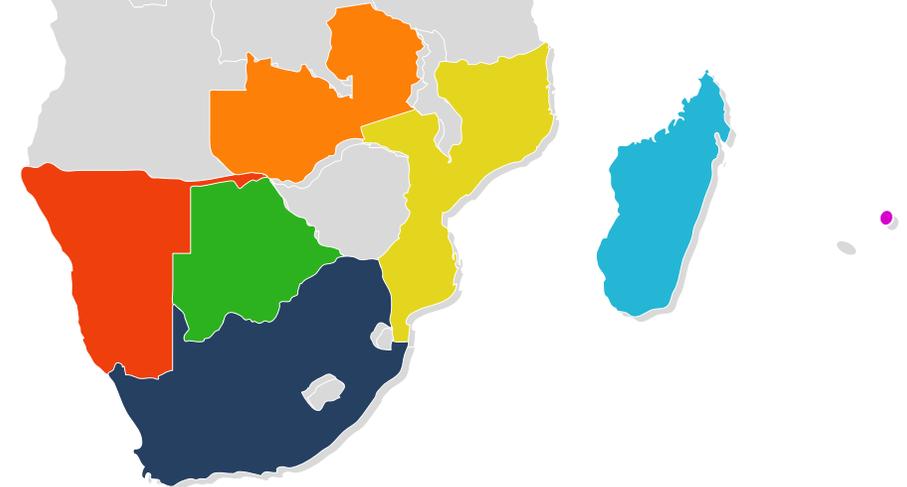
# HPC Ecosystem Project / SKA



- ❑ Repurposing of decommissioned HPC Systems
- ❑ Within South Africa and SKA Partner Countries
- ❑ Building of Institutional Capacity
- ❑ System Administrators Training and Users

## “Stampede”

- Dell C8220 (Sandy Bridge)
- 10 chassis \* 4 nodes
- 2 \* Xeon E5-2680 8C 2.7Ghz (640C)
- 32GB RAM (~1.28TB RAM)
- 250GB HDD (~10TB)



**5. Catalyst for Collaboration:** By acting as a hub for collaboration between academia, government, and industry, CHPC facilitates partnerships that lead to impactful solutions.

- **SARAO (South African Radio Astronomical Observatory)**
  - CHPC provides Data Centre Facilities for the MeerKAT radio telescope tier 1 facility research infrastructure
  - CHPC provides System Administration services to their Cluster
- **SAWS (South African Weather Service)**
  - CHPC provides continuity infrastructure services to SAWS
  - CHPC Platform serves as a failover system for the SAWS HPC system.
  - CHPC provides HPC training to SAWS staff

**5. Catalyst for Collaboration:** By acting as a hub for collaboration between academia, government, and industry, CHPC facilitates partnerships that lead to impactful solutions.

- Establish CHPC as a Quantum Computing Hub for South Africa
  - The RFQ process is being finalised, and the winning bidder will be announced soon.
  - The CSIR/NICIS QC Special Interest Group was established to consolidate QC specialists and streamline QC requests.
- The year 2025 is designated as the Year of Quantum Technologies Africa.
- The CSIR/NICIS QC event will be announced soon.

**5. Catalyst for Collaboration:** By acting as a hub for collaboration between academia, government, and industry, CHPC facilitates partnerships that lead to impactful solutions.

- Strategic Partner in the Machine Learning / AI Community
  - Adapting to changing times - from the scientific community to local Users of HPC Services (grass roots tech communities)
  - Engaging with Deep Learning Indaba and African Computer Vision, grassroots communities

**5. Catalyst for Collaboration:** By acting as a hub for collaboration between academia, government, and industry, CHPC facilitates partnerships that lead to impactful solutions

## CERN ALICE Scientific Research Program

- As of 1 May 2024, the ALICE Experiment granted CSIR/NICIS Associate Membership, reshaping the collaborative journey that began several years ago.
- Under the new Associate Membership Framework, CSIR/NICIS will contribute to the goals of the ALICE Scientific Research Programme in the following ways:
  - Providing a Tier 2 high-performance computing node for data analysis, storage, and simulation
  - Supplying a dedicated resource responsible for the operations, maintenance, and support of the Tier 2 node and
  - Provide a NICIS collaboration team with advanced skills in high-performance computing and infrastructure, cloud computing, artificial intelligence, and machine learning.



# CERN ALICE Scientific Research Program - Providing a Tier 2 high-performance computing node for data analysis, storage, and simulation



1. Worker Nodes
2. Obsolete Worker Nodes
3. Management Servers
4. EOS Storage

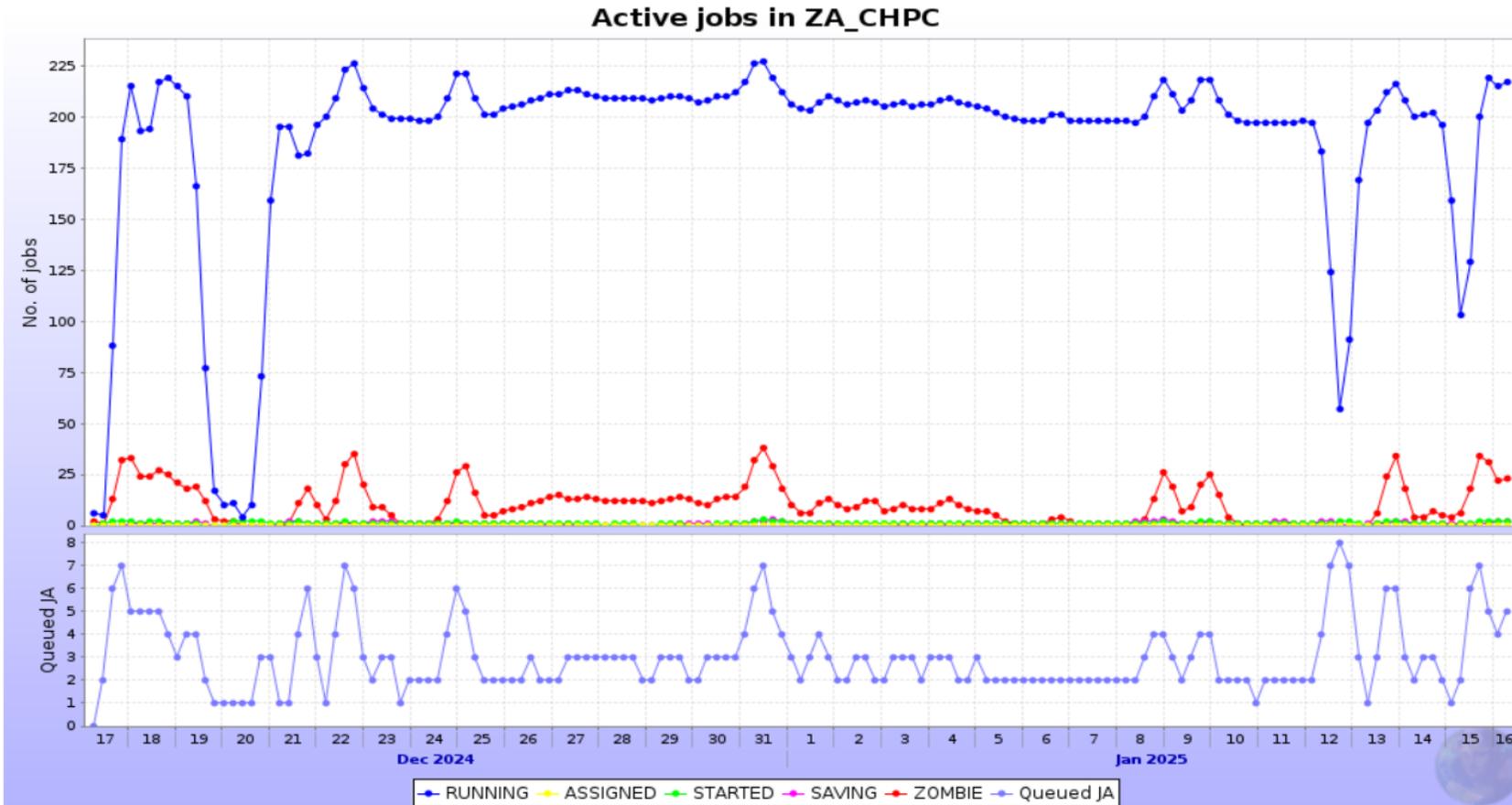
## Purpose

- To provide computing and storage resources for the Worldwide LHC Grid (WLCG) ALICE experiment.



# CERN ALICE Scientific Research Program - Providing a Tier 2 high-performance computing node for data analysis, storage, and simulation

Graph depicting active jobs for the **last 30 days** taken on 16 January 2025



# CERN ALICE Scientific Research Program - Providing a Tier 2 high-performance computing node for data analysis, storage, and simulation

- CHPC Tier2 Node Status
  - Our site is running Monte Carlo (MC) jobs due to the absence of functional storage.
  - The Tier2 Node will achieve full operational status once the storage filesystem is online.
  - CHPC will host Dr. Martin Vala from CERN for Tier 2 Node Technologies training sessions from February 14 to 24, 2025.
  - This training will upskill the CHPC technical team to manage and maintain the Tier 2 Node.
  - It will also enhance their skills in troubleshooting, optimising, and securing the EOS Storage system within the ALICE Experiment Grid computing framework.



**CERN ALICE Scientific Research Program** – Provide a NICIS collaboration team with advanced skills in high-performance computing and infrastructure, cloud computing, artificial intelligence, and machine learning

The CHPC currently **supports** the Alice experiment through the registered programme: **PHYS1669: ALICE Physics at the LHC**

- The principal investigator for the research programme is Prof Edith Zinhle.
- Current students registered include: Joshua Browne (UCT), Lunga Mandlazi (UCT), and Stephan Potgieter (UCT) – additional students are expected to register in 2025.
- This grants Alice's researchers complete access to the CHPC's Lengau Cluster, as well as to its user support.
- Specific tools have been requested and installed for their research, namely: Geant, ROOT, and Pythia.
- Dedicated documentation has been created to assist researchers, for example: <https://wiki.chpc.ac.za/research:root>.



**CERN ALICE Scientific Research Program** – Provide a NICIS collaboration team with advanced skills in high-performance computing and infrastructure, cloud computing, artificial

Recommended CHPC training includes:

- Cluster Induction Course (Self-study): Fundamentals of the CHPC Lengau Cluster and job management.
- HPC Winter School (~August 2025): Two-week introduction covering job scripting, workflows, software installation, and performance scaling.
- Parallel Programming School (TBC): Core concepts and techniques for efficient parallel computation.
- Sebowa Cloud Training (Self-study): Accessing and managing virtual machines.
- Participation at the CHPC National Conference (December 2025).



# CERN ALICE Scientific Research Program – Provide a NICIS collaboration team with advanced skills in high-performance computing and infrastructure, cloud computing, artificial

- One of the limitations students faced when running O2 Physics simulations was the inadequate resources available on their laptops.
- Owing to the root access requirements and the complex installation processes for O2 Physics, executing their simulations on Lengau was not practicable.
- An ideal solution was proposed: establishing a dedicated Sebowa Cloud project for the Alice research programme.
- A working demonstration was presented at the last CHPC Alice Seminar in November 2024, showcasing O2 Physics operating on an Ubuntu VM on Sebowa.
- It was requested that students participating in the research programme receive access to a VM with a minimum of 16 vCPUs and 24GB of RAM.



**Formula: R&I Capacity = [(D × E) + (C × S)] × I**

**Thank You!**

A national initiative of the Department of Science  
and Innovation and implemented by the CSIR



**science & innovation**  
Department  
Science and Innovation  
REPUBLIC OF SOUTH AFRICA

