

Case Study



High Performance Computing
2nd Generation Intel® Xeon® Gold Processors
Intel SSD DC drives

South Africa CHPC's OpenStack Production Cloud Enables COVID-19 Research and Programs

CHPC's private cloud cluster built on Intel HPC technology reflects the growing popularity of hybrid infrastructures that converge HPC with data analytics, artificial intelligence/machine learning to address a wider range of user needs

CEPH OpenStack storage cluster highlights:

- Supermicro TwinPro servers (phase 1)
- Dell PowerEdge R640 servers (phase 2)
- 2nd Gen Intel® Xeon® Gold Processors
- Intel SSDs



Executive Summary

The [South Africa Center for High Performance Computing \(CHPC\)](#) has provided large-scale research computing and data storage for over a decade. CHPC's [Lengau](#) supercomputer is the fastest on the continent. Since its installation in 2016, CHPC users have steadily grown as the center expanded its offerings to other researchers and industries throughout Africa. In 2017, CHPC joined the [Square Kilometer Array \(SKA\)](#) project to provide computational capacity for SKA's Science Data Processor (SDP). Part of the SKA is being built in South Africa.

Within three days of the CHPC OpenStack Production Cloud going live, the country went into lockdown due to COVID-19. The new private cloud was overwhelmed by many government agencies' needs to provide research and support for their activities.

Over the years, a growing number of CHPC research and industry users have needed non-HPC compute and storage services. The combination of increasing need for non-HPC and SKA's SDP computational resources led CHPC architects to develop a private cloud. The cloud was built on OpenStack and OpenStack CEPH storage software using 2nd Gen Intel® Xeon® Scalable processor-based Supermicro TwinPro servers and Intel SSD drives. Within three days of the CHPC OpenStack Production Cloud going live, the country went into lockdown due to COVID-19. The new private cloud was overwhelmed by many government agencies' needs to provide research and support for their activities. CHPC turned to Intel and Dell to upgrade their brand-new cloud system. Using servers built on 2nd Generation Intel Xeon Scalable Processors and Intel SSD DC drives, CHPC fulfilled the growing need for resources and met the demands of the pandemic.

Challenge

As a key center for large-scale computing in Africa, South Africa CHPC supports both academic and industry research. CHPC's 1.3 petaFLOPS Lengau cluster and its Lustre parallel file system cluster have been used on several flagship projects with supercomputing-level resources. These include [regional coupled ocean-atmospheric modeling at high resolutions](#), [energy storage materials](#), and the [MeerKAT array](#), among others. It has also contributed resources to commercial projects to support efforts through the [South Africa Development Co-Operative \(SADC\)](#) and in other countries in Africa, including Ghana and Kenya.

CHPC's user demand for computing and data resources over the last few years has accelerated—but in differing directions.



After supporting supercomputing with the Lengau cluster and more general-purpose users with individual VMs, CHPC deployed an OpenStack private cloud built on Supermicro servers to replace their virtual environment. (Photo courtesy CHPC)

"In addition to supercomputing, researchers also needed non-HPC, general purpose computing support. They wanted to store their data remotely, so they needed a more typical processing and storage environment rather than Lengau and the Lustre parallel file system," said Dora Thobye, Technical Manager for HPC resources.

CHPC created a VMware-based environment in a cluster called IT-Shop to deploy individual virtual machines (VMs). Storage was still provided by the Lustre parallel filesystem. As non-HPC workloads expanded, VM support grew in complexity. Storage demand overburdened the Lustre storage system, degrading storage performance for supercomputing by 30 to 40 percent, according to CHPC.

Then, in 2017, MeerKAT was joined with the [Square Kilometer Array \(SKA\)](#) project, and CHPC joined the SKA to provide computing and storage resources for its Science Data Processor (SDP). The growing demand for general-purpose computing and storage services, and the need to support SKA with a cloud environment, led CHPC into a new direction. The center began research into a converged cloud and HPC data center infrastructure that would support automated orchestration of compute and storage along with supercomputing.

A growing number of HPC centers around the world are creating hybrid infrastructures. Compute-intensive, parallel performance clusters are converging with data analytics, artificial intelligence/machine learning (AI/ML), and private cloud architectures to address a wide range of user needs under one infrastructure umbrella. The UK Science Cloud at Cambridge University built on OpenStack is one example. CHPC referenced the Cambridge University OpenStack solution in their implementation.

"Much like data from the Large Hadron Collider's Atlas detector, computation for SDP data will be shared across many countries and users," explained Dr. Happy Sithole, CHPC's director. "OpenStack provides a transparent environment for users around the world to analyze SDP data. And OpenStack offers a foundation for our existing needs and for our future converged infrastructure."

CHPC worked with [StackHPC](#) and [Linomtha ICT](#) to design the CHPC OpenStack Production Cloud to replace the existing VMware environment. The new private cloud was built on Supermicro TwinPro servers with 2nd Gen Intel Xeon Scalable processors and 3 TB of memory per chassis. 1.5 petabytes of mechanical disks and more than 220 TB of Intel SSD drives created a CEPH storage cluster with a hierarchical storage architecture for short- and long-term storage.

"The new cloud system was designed to support many virtual jobs related to ongoing research, such as custom workflows, pleasingly parallel workloads, and web hosting," commented Thobye.

The IT department began migrating existing users to the OpenStack Production Cloud on March 23, 2020. Three days later, everything changed, and the new production cloud was quickly overwhelmed.

Solution

On March 26, 2020, South Africa went into lockdown due to the impacts of the COVID-19 pandemic across the country. As CHPC began migrating users off the previous VM environment, the COVID pandemic drove additional need for cloud computing and storage. The government turned to CHPC for support. Government programs originated by the Department of Health required enormous computing and storage resources for processing population tracking and tracing and other data. Demand for resources to support emerging remote education, artificial intelligence, and other services related to the virus also increased. DNA sequencing of the virus required massive amounts of data storage.

"Because of the pandemic and all the new users it brought to us, we were running out of compute and storage resources," explained Thobye.

With support from two major universities in the country, plus Dell EMC and Intel's Pandemic Response Technology Initiative, CHPC was able to expand the OpenStack Production Cloud. The two universities involved were University of Cape Town and the University of the North West (Potchefstroom).

The OpenStack Production Cloud expansion included the following:

- 15 new compute nodes using Dell PowerEdge R640 servers with dual Intel Xeon Gold 6230R processors for a total of 780 cores providing performance of 33.285TFlops
- 3 new storage nodes using Dell PowerEdge R740XD2 servers with dual Intel Xeon Gold 6226 processors
- 80 TB of hot data storage using Intel SSD DC drives
- 480 TB of HDD storage (3 x 160 TB copies)

The expansion was completed in mid-2020 and went into production with a total capacity of 780 compute cores, 480 TB of cold storage, and 60 TB of hot storage (Intel SSDs). With more storage and compute capacity, users are experiencing a much more capable system.

"Instead of being far overprovisioned with continuous 100 percent utilization," commented Dr. Sithole, "workloads now consume from 60 to 100 percent of the compute capacity, depending on the activities."

Result

“OpenStack provides a different offering for users of the data center,” said Sithole. “This implementation is a step in the right direction to revolutionize our data center as a converged environment. We see this as a continuum between compute-intensive and data-intensive computing. It allows us to easily support both HPC research and general-purpose cloud computing in the same infrastructure.”

With the original Supermicro cluster and the Dell EMC expansion, the expanded cloud can now support ongoing pandemic-related activities by the Department of Higher Education and Training, Department of Health, university research, and other public and private projects to address needs from the pandemic. Compute- and data-intensive projects include sequencing and virus research, remote education and online learning, bandwidth analysis of remote communities who need remote learning, television whitespace analytics, analytic epidemiology (including track and tracing), and others. The discovery of the South Africa variant of COVID-19 was accomplished using CHPC resources.

“This implementation is a step in the right direction to revolutionize our data center as a converged environment. We see this as a continuum between compute-intensive and data-intensive computing. It allows us to easily support both HPC research and general-purpose cloud computing in the same infrastructure.”

—Dr. Happy Sithole, CHPC Director

According to Dr. Sithole, the larger cloud also brings many new tools that will allow users to take advantage of the new environment. Intel AI technologies, machine learning (ML) libraries, containerization, and other resources will help users who want to implement artificial intelligence (AI) and explore new approaches to their scientific problems.

“The cloud platform further enables CHPC to gather the necessary technical and operational expertise to develop, provision, and operate a national federated OpenStack platform,” stated Thobye. “It will allow for global connectivity in a virtual environment for mega projects, like the Square Kilometer Array and similar in stature.”

Before the pandemic struck South Africa, CHPC was piloting other Intel HPC technologies, such as Intel Optane persistent memory and Intel Optane storage. CHPC expects these technologies can improve large-memory processing performance and efficiencies by keeping more data closer to the processing platform. Such proximity is important with workloads that interact with huge amounts of data like the SKA. These technologies can also accelerate genome sequencing and assembly.

Once the population has been vaccinated and the virus under control, CHPC's OpenStack Production Cloud will be able to support many other activities. More members of SADC can take advantage of easy access to computing and storage resources. New weather models are being explored that will help Africa understand and deal with its unique weather events and the effects of climate change.



Within three days of going live, CHPC's new system was overwhelmed due to the pandemic. Dell EMC and Intel assisted CHPC to expand their OpenStack Production Cloud to address the emerging needs. (Photo courtesy CHPC)

“Once COVID is beyond us,” concluded Dr. Sithole, “we have different challenges in Africa. The OpenStack platform gives us AI and other tools that will help find solutions for Africa's unique problems. One of those challenges is the issue of communicable diseases. Ebola, for example, but Ebola is not the worst disease that Africans face. And what we have learned with COVID is that you cannot solve such problems alone. There has to be a concerted effort from everybody together to find cures for the problems that we have. Hopefully that will accelerate the uptake of the CHPC platform so we can find solutions for those unique African problems as well.”

Solution Summary

With a growing user base and expanding role beyond traditional supercomputing resources across Africa, CHPC needed to evolve its computing environment. After supporting supercomputing with the Lengau cluster and more general-purpose users with individual VMs, CHPC deployed an OpenStack private cloud built on Supermicro servers to replace their virtual environment. Within three days of going live, the new system was overwhelmed due to the pandemic. Dell EMC and Intel assisted CHPC to expand their OpenStack Production Cloud to address the emerging needs. The new expanded cloud environment is allowing the country to address the disease and the outcomes of it with easy access to compute- and data-intensive processing and storage resources. The OpenStack Production Cloud is CHPC's next step in their journey to a converged HPC/cloud datacenter.

Where to Get More Information

Learn more about the [Center for High Performance Computing](#).

Explore the capabilities of the [2nd Generation Intel Xeon Scalable processors](#) with integrated Intel Deep Learning Boost capabilities for accelerated AI inferencing.

Solution Ingredients

- Supermicro TwinPro servers (phase 1)
- Dell R640 PowerEdge servers (phase 2)
- 2nd Gen Intel Xeon Gold Processors
- Intel SSDs



¹ <https://www.chpc.ac.za/index.php/resources/lengau-cluster>

Performance varies by use, configuration and other factors. Learn more at www.intel.com/PerformanceIndex.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.

Your costs and results may vary.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy. Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.