



About Customer

Ranked amongst the world's top universities by QS, [Nanyang Technological University](https://www.ntu.edu.sg) (NTU) is a research-intensive public university in Singapore. NTU has over 33,000 students and is home to world-class autonomous institutes and research centers. www.ntu.edu.sg

Geo
APJ

Industry
State and Local Government and Education

Solution Area
Activate Real-time Analytics

Products in Use
Pure Storage® FlashBlade®

FlashBlade Expands NTU's Supercomputing Power

As a global institution with a strong multidisciplinary research culture, the High Performance Computing Centre (HPCC) under NTU's Centre for IT Services (CITS) worked with Pure Storage to fulfill a new scratch space requirement and cater to a growing demand for supercomputing resources.

With Pure's FlashBlade and Evergreen™ delivering significant improvements, researchers have been able to accelerate their investigations and process multiple streams of high-volume data, without disruptive, costly upgrades. These improvements have also allowed the HPCC team to offer additional resources and support to even more research teams.

"FlashBlade's ability to scale up performance while lowering I/O latency has exceeded our expectations. Evergreen's non-disruptive upgrades have also simplified the management of our infrastructure dramatically."

ALVIN ONG,
CHIEF INFORMATION
OFFICER, NTU SINGAPORE

Impact on NTU HPCC



Able to support a wider range of projects with more complex requirements



Accelerate completion of projects and free up computing resources



Optimize use of rack space, with improved power and cooling efficiencies

Challenges



Number of species analyzed was almost 10x larger than past projects



More computational power needed to accelerate the genome assembly process



Larger capacity required to analyze multiple species simultaneously

Expanding NTU's Supercomputing Infrastructure

Central to the university's research the HPCC is a large-scale computing facility that processes data for projects ranging from genome sequencing and weather analysis to supporting the university and its partners. The team required a storage architecture that can scale up to at least a million Input/Output Operations Per Second (IOPs) linearly and easily, given that they can run up to eight different workloads at any point in time. This led to the deployment of FlashBlade to fulfill a new scratch space requirement.

Cataloging the Botanical Diversity

Researchers from NTU's School of Biological Sciences were among the first to leverage FlashBlade. Their goal is to analyze the genomes of over 1,000 plant species.

Conducting the first project of this scale globally, the research team has been working with DNA samples to map the genetic codes, also known as assemblies, of every species in their study. To build each assembly, the teams compare DNA sequences for overlaps and predict the genes in the sequences, piece by piece. With over a trillion possible combinations, genome assembly is one of the most computationally intensive processes the HPCC supports.

FlashBlade: A Game-changer for Genetic Research

A key priority from the start was the ability to run multiple analyses concurrently. "If we looked at one genome a week, it would take us 20 years to cover 1,000 species, so it was important for us to do things in parallel and moving to FlashBlade has enabled us to significantly speed up the process," shared project lead, Assistant Professor Jarkko Salojarvi. With FlashBlade enabling them to run up to four jobs in parallel, the team has since crossed the project's halfway mark, completing more than 550 assemblies in just 18 months, compared to less than 100 assemblies within the same time span if examined one by one.

Once this initial research phase is complete, the team hopes to expand the study to other areas and dive deeper into specific species and groups. The data will also help other researchers studying environmental conservation and climate change as well as the plants' potential for medical and pharmaceutical use.

Results



Able to efficiently support studies with significantly larger data sets



Drastically improved IOPs, with no more sleeping processors



Capable of parallel processing without stalling or compromises on speed

purestorage.com

800.379.PURE

