



WEHI Tackles Data Inflammation with VAST

VAST Storage system from XENON accelerates scientific research at WEHI

WEHI (Walter and Eliza Hall Institute of Medical Research) is where the world's brightest minds collaborate and innovate to make life-changing scientific discoveries that help people live healthier for longer. WEHI's medical researchers have been serving the community for more than 100 years, making transformative discoveries in cancers, infectious and immune diseases, developmental disorders and healthy ageing.

In 2017, WEHI worked with XENON to implement a private cloud solution for their High Performance Computing (HPC) environment. At that time, approximately 90 bioinformaticians and biologists were processing research data on the HPC cluster. The XENON private cloud implementation allowed the researchers to 'containerise' their workloads, making their experiments

more repeatable. The private cloud approach also allowed for more efficient scheduling of jobs and delivered better compute resource allocations.

The HPC cluster utilised a traditional hierarchical storage management system (HSM) for data storage. This included a tier of high-performance hybrid disks (flash and hard drives), used as scratch cache, and a tier of capacity hard drives with automated tiering to a large tape library, used for archive and long-term storage. The advantage of this tiered system was that researchers were allocated working storage space both within scratch space and within the capacity tier – this presented as 'unlimited' storage capacity to the researcher, with older data tiered to tape.





DATA GROWTH AND NEW RESEARCH TOOLS

Since 2017, WEHI has introduced new research tools and instruments, like the cryo-electron microscope (aka cryo-em). The cryo-em processes allow researchers to see biological actions as they happen, frozen in time at the molecular level. To achieve these results and this level of insight, the cryo-em researchers process large, high-resolution images to extract micro-graphs and 3D interpretations of their data. This process alone requires storage with a high level of streaming data read/write performance, and also a high level of input/output processing per second (IOPS).

The amount of genomic sequencing data generated at WEHI also increased massively over the last five years.

“It’s in the order of hundreds of millions of files. Lots of sequencing data; some

researchers have directories with over 200,000 individual files,” stated Tim Martin, Senior ITS Research Systems Engineer. Tim also said that a lot of file systems struggle with this scale.

Miguel Esteva, Senior ITS Research Systems Engineer noted, “We have a little bit of everything. We have massive imaging files, and some divisions that create millions of tiny files.”

This variation is a key design issue in storage and data management.

While the HSM storage environment provided large capacity, it had performance limitations as the system was originally intended for managing the large data sets of the researchers and providing data protection through the replication to tape.

▲ New instruments create more data, faster than before.

THE HUNT FOR PERFORMANCE

The WEHI team needed not only high-performance scratch storage that would meet their streaming requirement; it also needed to cope with its expanding IOPS load requirements and provide for growth into the future. With XENON, they explored the range of offerings from leading storage vendors.

Tim said, “We were really attracted to the architecture of VAST. It was clearly fresh, simple and scalable. Other vendors had legacy architecture issues like east-west traffic, or scalability issues – we have a reputation for pushing storage and compute to its breaking point due to our ever-growing data sets. If there is a limit, we will find it.”

Tim added, “We also wanted to avoid drivers and go with a more flexible open standard

protocol. We’re in a constant battle updating drivers and clients between current storage and filesystems.”

Future scalability was another key consideration.

“VAST also offered a seamless expansion path to add capacity and performance without disrupting users,” Tim said.

While VAST was new in Australia, and a new storage company, the recommendation from XENON helped.

“XENON has always supplied quality that has been proven in the test of time,” explained Miguel.

SCIENCE DISCOVERIES ARE NOW VAST

WEHI procured an initial 676 TB of VAST storage from XENON and this was implemented as new HPC scratch storage. This was presented to applications as a cache area, or as a cache and data area, depending on the application requirements. The VAST storage was initially provided to applications requiring high IOPS with a random data access pattern, like the cryo-em team's processing needs for their image files.

Initial site acceptance testing was showing a dramatic acceleration of the new research tools that required high IOPS scratch storage. CryoSPARC was running 5-8x faster when using VAST as a cache, while Relion 2D classification was 10-17 times when using VAST as a cache and data area. Samtools sort was 2-3.5x faster.

With VAST, IOPS from the storage is no longer a bottleneck for these high-performance research applications. The VAST storage is

achieving close to network saturation at 50Gb/s between compute nodes and the storage, and hitting over 300,000 IOPS. These results have encouraged a wider adoption of VAST, with the cryo-em users moving across to use VAST to support their work.

Tim noted that "feedback from users is that it has greatly improved the performance of their workloads".

Importantly, from a research perspective, Tim stated, "Time to result has been reduced, and the scale of research we can run has massively increased. And from our team's perspective we have more visibility into what the storage is doing, we have really good support from VAST and XENON, and we've never had any kind of issues. We spend less time troubleshooting problems with the filesystem, and that's time we can spend on other things that might help the researchers more."

VAST STORAGE is achieving close to network saturation at

50Gb/s

between compute nodes and the storage, and hitting over

300,000 IOPS

FURTHER ENHANCEMENTS

Tim and Miguel recently rolled out a snapshot feature to specific use cases. Researchers have to manage their data within their VAST allocation, and snapshots are being used in case of accidental deletions.

Tim explained, "The system allows users to do this kind of data recovery themselves, using the snapshots".

The WEHI team has also expanded their VAST storage, adding another 676 TB, and found this a painless experience – just adding the node and, according to Tim "the system auto-balanced the metadata and it just worked for the end users instantly".

With no clients to update or manual data migration, expansion has been seamless with the full 1.3PB available without tuning, migration or manual interventions.

Now that VAST has provided a performance storage platform, the next challenges ahead for the WEHI team are in the realm of data management: providing further innovation to the researchers who are creating, storing and protecting the data that is the foundation of modern medical research.

Learn more about



[WEHI](#)



[XENON](#)



[VASTDATA](#)