WEHI brighter together + VAST

SUCCESS STORY

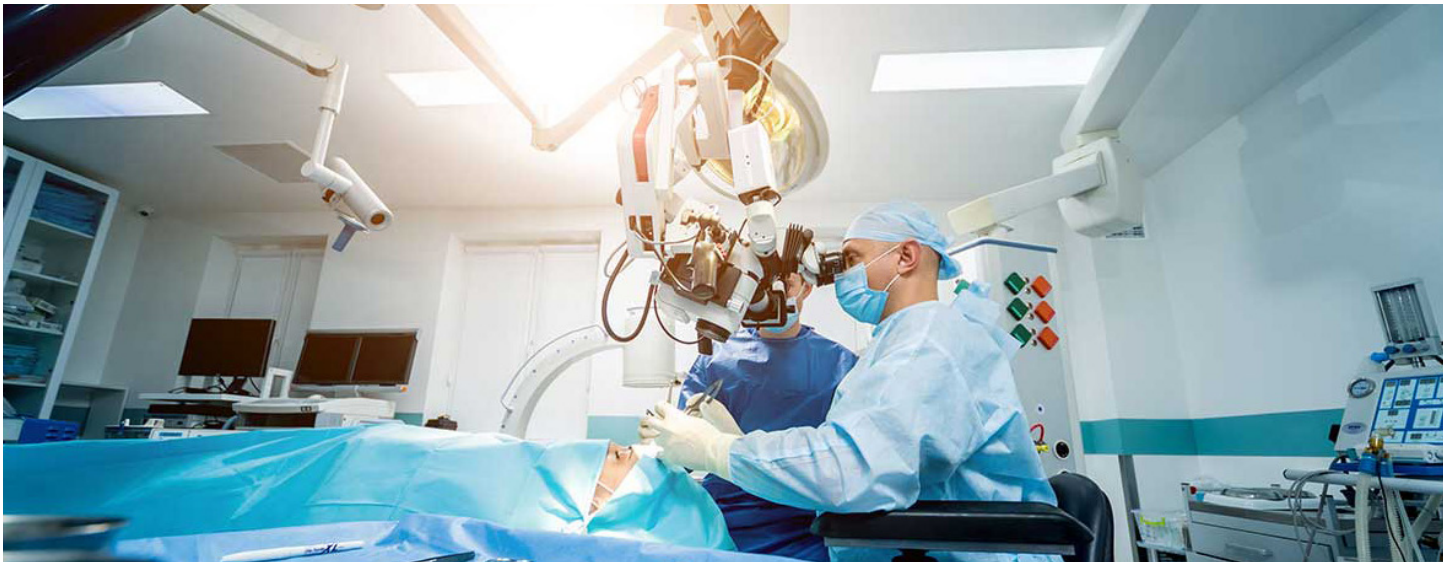
LIFE SCIENCES LEADER POWERS NEXT-GEN RESEARCH COMPUTING WITH UNIVERSAL STORAGE

OVERVIEW

WEHI (Walter and Eliza Hall Institute of Medical Research) is Australia's pre-eminent biomedical research institute. It has been serving the Australian community for more than 100 years, with teams of researchers committed to solving the most complex health problems, making transformative discoveries for cancer, infectious and immune diseases, developmental disorders and healthy aging.

Similar to other leading medical research facilities, WEHI is deeply engaged in the fields of genetics, structural biology and life sciences. Working in each of these fields involves acquiring, managing, analyzing and operating on huge amounts of data, where a single project may involve billions of files, both large and small.

WEHI's commitment to next-generation therapeutics techniques, such as Cryo-EM and other emerging applications, brought new challenges for managing and processing an enormous amount of data. WEHI's IT leaders envisioned a new approach to data infrastructure that would provide high-performance, scalable storage with exceptional ease of use and economics.



CHALLENGE

THE COMPLEX DEMANDS OF MEDICAL RESEARCH

Applications related to cryo-electron microscopy (Cryo-EM), protein folding (AlphaFold), and genomic analysis place unique and taxing demands on underlying computing technologies, particularly storage. Indeed, the majority of storage solutions—including both legacy hard drive systems and last-generation flash-based platforms—struggle (and often fail) to provide acceptable performance across all the datasets involved in such work. Thus, fast affordable flash storage has become a critical element in providing capable and usable application performance for researchers.

Cryo-EM datasets are particularly taxing due to the extraordinary variation between data sizes and access mechanisms used at each step in the processing pipeline. Though raw datasets are enormous, the process of cleanup and correction usually eliminates as much as half the original data in moving from one step to any next step. This changes the access behavior for datasets to progressively smaller, random IO patterns as they move through the pipeline from one end to another.

As well, WEHI has invested substantial time and effort into Deepmind's AI-based AlphaFold software and database. Protein folding is another complex process with demanding performance requirements across multiple pipeline stages; thus the same performance challenges with tiered storage arise given the massive amount of calculations AlphaFold delivers across a huge number of individual data points.

Legacy file systems rely on caching techniques to "pre-fetch" data from lower-cost hard disk drive (HDD) based storage tiers. However, it's nearly impossible to pre-fetch a random read in life sciences workflows or any application designed to extract value from random bytes across massive datasets. And when a cache miss occurs organizations can see performance reduced by upwards of 80 percent.

To harness the value of emerging research applications, WEHI needed a file system that could optimize I/O latency and performance to move data and images from intake all the way to final outputs. Only then could researchers undertake the real work of interpreting data from Cryo-EM 3D models and AlphaFold protein structures, unlocking and decoding the information they contain.

SOLUTION

UNIVERSAL STORAGE ACCELERATES DATA PROCESSING AT WEHI

VAST Data enables unlimited processing on exabyte-scale, affordable flash for bioinformatics and biomedical research communities. VAST has introduced the industry's first Disaggregated Shared Everything (DASE) architecture, built using modern end-to-end NVMe capabilities and a highly resilient container-based software architecture that provides 100% online operations and unprecedented fault-tolerance. VAST Data's Universal Storage combines low-cost, high-density flash with state-of-the-art data reduction capabilities to provide the most affordable all-flash form factor for researchers. Standard NFS, SMB and S3 access allows for simple multi-protocol storage for any scale-out application, in stark contrast to the complexity of legacy parallel file systems.

VAST accelerates data collection from Cryo-EM devices to make it available quickly to applications at scale.

Researchers report that they need no longer wait for data to make its way into applications for modeling and analysis. In fact, WEHI's Senior ITS Research Systems Engineer Tim Martin reports that:

- CryoSPARC, a Cryo-EM software platform, now runs five to eight times faster with VAST than its previous scratch file system
- Relion 2D classification - a cloud-based Cryo-EM data analysis platform - runs up to seventeen times faster than the previous solution
- Samtools, a suite of programs for interacting with high-throughput sequencing data, runs up to 3.5 times faster

VAST speeds up I/O-intensive pipelines.

Flash traditionally has been great for accelerating small storage transactions, such as database transactions. VAST eliminates the tradeoffs between cost and performance with an all-flash system for all data, enabling applications like Cryo-EM and AlphaFold that only get smarter when they can randomly read through the largest amounts of data possible. In fact, WEHI Cryo-EM now collects data as quickly and efficiently for large objects as it does for small ones.

VAST provides fast and reliable data access to researchers working on protein folding.

VAST enables full pipelines for AlphaFold and other protein folding application pipelines, ensuring researchers never have to wait for data.

VAST enables GPU-based processing.

Processing speeds for complex workloads run more quickly, and take better advantage of uniform and fast access speeds and enormous working sets for data in VAST Universal Storage. This is a noticeable boost in an environment like WEHI's, where big jobs were routinely scheduled months in advance, and where individual jobs took a year or more to complete.

RESULTS

OVERCOMING DATA ACCESS BOTTLENECKS IN MEDICAL RESEARCH

WEHI researchers reported after deploying VAST Data's Universal Storage on their networks, IO delays for read-intensive applications are a thing of the past. VAST has delivered significant improvements in job completion times and total workload handling capacity, and the system has yet to be pushed beyond its limits.

Says WEHI's Martin: "We are now getting over 30 GBs and more than 200,000 IOPS from VAST, a considerable improvement." The performance and stability of the system is adding impetus to migrate additional users and workloads onto VAST, Martin says, so that WEHI can realize further improvements in performance, scale, and resiliency.

> *"With VAST there is now consistent and repeatable performance with I/O no longer being a bottleneck. Our previous system used to top out at 50 simultaneous jobs; now, we can run over 1,000 jobs at the same time, and interactive users are not impacted even when the batch system is effectively running at capacity. We spend less time troubleshooting, and can spend more time helping researchers."*
>
> **– Tim Martin, Senior ITS Research Systems Engineer, WEHI**

ABOUT 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 aging.

ABOUT VAST

Headquartered in New York City, VAST Data is a storage company bringing an end to complex storage tiering and HDD usage in the enterprise. VAST consolidates applications onto a highly scalable all-flash storage system to meet the performance needs of the most demanding workloads, while also redefining the economics of flash infrastructure to finally make it affordable enough to store all of your data on flash. Since its launch in February 2019, VAST has established itself as the fastest selling storage startup in history. VAST's Universal Storage now powers several of the world's leading data centric computing centers.