

Success Story

# The GPU Tsunami and the IO Bottleneck

Success Story

VAST | SciNet

SciNet, the supercomputing center at the University of Toronto, is one of Canada's five centralized HPC facilities and home to the TOP500 supercomputer Trillium. The center serves the entire academic community in Canada, providing researchers with the resources and expertise necessary to perform diverse, large-scale research, ranging from cutting-edge bioinformatics to climate science and astrophysics. Their mission demands a general-purpose, highly available system capable of handling the largest parallel compute jobs.

With the installation of their new, third-generation supercomputer, a system boasting 240,000 CPU cores and 250 NVIDIA H100 GPUs, SciNet faced a critical challenge: ensuring the underlying data infrastructure could keep pace with the massive compute power and, crucially, the burgeoning demands of AI workloads. "We support virtually every scientific discipline, from traditional fields like physics, high-energy physics, and engineering to emerging areas like bioinformatics, social sciences, and AI/statistics. Computing is essential for all academic work." explains Daniel Gruner, CTO, SciNet.

The previous architecture, which ran for over seven years, relied on a complex, two-tiered system: ~18 petabytes on GPFS for bulk storage and a separate flash burst buffer dedicated to handling the IOPS-intensive small-file and temporary data loads. Gruner was clear about the limitations of this model.

This multi-tiered complexity was an inhibitor to effectively utilize the new GPU resources. AI/ML training and inference workflows generate a deluge of small files and metadata operations, requiring constant, high-speed, low-latency access to data. SciNet needed a platform that could deliver this extreme performance uniformly across the entire capacity, eliminating the need for complex workarounds and separate storage partitions. The ability to efficiently handle the increasing demand for AI workloads was non-negotiable for their new system.

> *I was adamant I didn't want to do the same thing again. We are adventurous, if you will, but not in a silly sense. There's a good reason for looking to do our work differently and better.*
>
> **Daniel Gruner,**
> CTO.
>
> SciNet

## A Unified Data Foundation for AI and HPC Convergence

SciNet's search for a unified, high-performance solution led them to VAST AI Operating System. After vetting alternatives, the VAST proposal stood out for its unique DASE architecture, which promised to deliver the simplicity and speed of an all-flash system without the traditional cost barriers. Daniel noted the revolutionary nature of VAST's design:

> *VAST is a completely different beast for us. We were always waiting for a proper parallel NFS and it never materialized. VAST is the future way of doing things and that's why we brought them on.*
>
> **Daniel Gruner,** CTO, SciNet.

Following a successful Proof of Concept (POC), where a VAST system was subjected to intense stress testing by being plugged into the old 2,000-node cluster with "no problem," SciNet moved forward. They installed the VAST Data Platform with 30 petabytes capacity, achieving a core goal of ending their reliance on spinning disk and the associated risks of rebuilds and performance degradation.

## Enabling CPU & GPU Workloads with Extreme Performance

The VAST deployment provided SciNet with immediate, structural advantages critical for their accelerated science programs:

- **Eliminating the Burst Buffer Concept:** The VAST platform renders the burst buffer obsolete. The entire 30 petabytes acts as a single, consistent performance tier. This move simplified the data path dramatically and ensures that the system is not IO-bound when running large-scale jobs, such as single-run ocean simulation which produces a petabyte of output.
- **Powering AI Workflows:** By choosing VAST, SciNet ensured their GPUs were fully fed with data. The VAST architecture efficiently handles the chaotic IO patterns characteristic of AI/ML, whether it's the parallel IO from large model training or the intense metadata operations from complex data lakes.
- **Preparing for Massive AI Expansion:** SciNet is already planning the next phase of their AI capability, which requires an increase in storage capacity of 30 petabytes. The proven scalability of the VAST platform will allow them to increase their capacity seamlessly, proving it as the scalable foundation for their long-term AI strategy.

> *We are greatly expanding our AI capability and VAST will be central to this, inclusive of AI agents.*
>
> **Daniel Gruner,** CTO, SciNet.

# The Impact: AI Readiness, Resilience, and Advanced Services

SciNet now operates with a more resilient and flexible data environment. Three months into full production, the system has run flat out without a single issue of saturation or complaint, which is a testament to the platform's stability. Furthermore, the VAST AI OS is enabling several advanced strategic initiatives:

## 1 Operational Excellence and I/O Quality of Service (QoS)

- **Administrator Happiness:** VAST's management interface has been a welcome change for the system administration team, who previously dealt with the lack of friendly tools of the GPFS world. This operational efficiency frees up time to focus on supporting researchers.
- **Data Efficiency:** VAST is already achieving a ~2:1 data reduction ratio, exceeding the expected ratio, delivering greater savings than first thought.
- **QoS for AI:** The data platform environment makes implementing I/O QoS a reality. As they bring their substantial new AI capability online, QoS will be critical to dedicate resources for big AI jobs to ensure that they're not IO bound.

## 2 Enabling Secure, Multi-Tenant Research

VAST's architecture is foundational to SciNet's development of a secure data facility. By leveraging VAST's built-in multi-tenancy, SciNet can effectively partition the file system, a requirement for working with sensitive data such as private health information and industrial intellectual property. The ability to eventually use VAST's encryption capabilities to assign encryption by tenant with different keys will be central to offering a high-performance, secure environment for Canada's most sensitive research.

Working with VAST, SciNet has moved beyond the constraints of past architectures, creating a unified, high-performance, data platform that efficiently fuels both their massive traditional HPC and their rapidly expanding AI initiatives, solidifying their position as an innovative leader in scientific computing.


©2025 VAST Data, Inc. All rights reserved. All trademarks belong to their respective owners.
11/2025