

spinTwo: Advancing Life Science Discoveries



In the highly competitive field of life sciences and drug discovery, cutting-edge infrastructure is key to accelerating breakthroughs. spinTwo and VAST Data empower scientists to focus on groundbreaking research by minimizing data management complexities. This enables faster delivery of new treatments and cures to diseases for the patients who need them.

About spinTwo

spinTwo also offers a comprehensive suite of services, from building and maintaining custom HPC systems to developing efficient AI models and delivering cutting-edge data analysis. By employing former scientists with strong technical backgrounds, spinTwo's team possesses a deep understanding of the computing requirements of the scientific community, enabling them to build and maintain infrastructure that powers scientific breakthroughs. Hugo Hernandez, co-founder of spinTwo who oversees scientific computing infrastructures, explained, "We are not just providing a solution from an infrastructure perspective. We go further by working closely with scientists on how to use the infrastructure, how to use different environments to push the boundaries of scientific discovery."

Accelerating Discovery

One spinTwo customer required a robust computing infrastructure to handle billions of small files while integrating seamlessly with their existing HPC systems—crucial for accelerating research in immune analysis and drug discovery. In one experiment, scientists captured 1.4 billion images of mice brains to study their neurological activity. Managing and processing this data was a monumental challenge. "This analysis is not possible if we don't have the right infrastructure. The right infrastructure includes not only the computing power but also, most importantly, the right file system. A file system that allows us to deal with a lot of very small files, and by a lot, I mean billions of files," commented Hernandez.

After evaluating multiple options capable of integrating with traditional HPC systems, spinTwo selected VAST Data on behalf of their customer. VAST Data's architecture stood out for its ability to handle massive amounts of small files while delivering high performance and providing flexible data access. "We found that traditional parallel file systems couldn't handle the large number of small files and performance requirements we needed," said Hernandez. "With VAST Data, not only did we overcome these limitations, but we also gained a file system that could present data not only to HPC systems but also to high-performance workstations and critical servers controlling scientific instrumentation and experiments."

Why the VAST Data Platform?

The VAST Data Platform is a highly scalable and affordable all-flash data system that allows you to run AI-scale analytics at less than half of the cost of traditional all-flash solutions.

DASE Architecture

With its unique Disaggregated and Shared-Everything architecture, VAST breaks tradeoffs deliver significant savings and make flash affordable for all of your data.

VAST DataStore

The VAST DataStore is an enterprise all-flash NAS platform built to meet the needs of today's powerful AI computing architectures and beyond.

VAST DataBase

The VAST DataBase is a key strategic piece of the Data Platform targeting structured data. It provides transactional and analytical data warehouse functionality on top of the scalable, performant, all-flash base layer to ingest and process data at best-in-class speeds.

VAST Data Catalog

The VAST Data Catalog leverages the Database and SQL to provide a built-in metadata index that allows you to search and find data easily – structured or unstructured – at scale and in a fraction of the time of traditional methods.

The results transformed their research capabilities. Scientists accessed their data faster than ever before. In one notable example, a data analysis project estimated to take seven years using traditional computational techniques was completed in just a few weeks with the optimized infrastructure integrating VAST Data as a core element of their HPC cluster.

Results and Benefits

By leveraging VAST's infrastructure, the customer significantly improved its data processing speed and ability to scale complex life sciences research. The system allows scientists to focus on their research, free from the burden of data management and infrastructure logistics.

Hernandez summed up the impact: "We helped our customer accelerate science. With VAST Data, we process large datasets much faster, allowing scientists to spend more time on research rather than waiting for data to process. What used to take years, can now be completed in weeks, and we're working to reduce that to hours."



"With VAST Data, we process large datasets much faster, allowing scientists to spend more time on research rather than waiting for data to process."

Hugo Hernandez, Co-Founder spinTwo

spinTwo integrated the VAST Data Platform into a scalable, future-proof system that will continue to support all AI and ML initiatives, large-scale data analytics, and long-term scientific projects. As Hernandez put it, "VAST is an important component of the infrastructure we built for our customer, particularly for data analytics and scientific operations."

Government Agencies

The spinTwo team has played a major role in driving insights in collaboration with the National Artificial Intelligence Research Resources (NAIRR) Secure Pilots, co-led by the Department of Energy (DOE) and the National Institutes of Health (NIH). The team has been pivotal in integrating computational resources and AI-based workflows across environments within both NIH and DOE organizations.

Notably, spinTwo facilitated the integration of large language model (LLM)-based workflows with the Frontier supercomputer at Oak Ridge National Laboratory (ORNL) and helped integrate JupyterHub to launch notebooks as traditional jobs running on an HPC cluster. VAST has also been actively supporting NAIRR projects across its customer base at research universities and federal agencies.

"One of the main objectives of the NAIRR Secure Pilots is to identify challenges and opportunities for interoperability between tools, infrastructure, software, and other resources across NAIRR Secure enclaves. This enables not only scientific but also technical collaborations across the Federal Government. These pilots have proven that such collaborative efforts can be achieved, ensuring interoperability and providing scientists with enhanced capabilities to accelerate their research projects—all while maintaining essential security measures," said Hernandez.

Looking Forward

In the next phase, spinTwo plans to expand VAST Data's capabilities to more scientific research projects with their customers, including efforts around artificial intelligence and machine learning to support enhanced analysis and data management. As Hernandez explained, "We have laid out the infrastructure foundation required to accelerate discoveries in life sciences."

Dave Reese, Director of the Department of Defense and Federal Healthcare, at VAST Data explained, "spinTwo has been a driving force behind innovation, particularly in advancing AI, ML, and HPC capabilities within the scientific community. As legacy, siloed storage systems give way to intelligent, ZTA-Data Pillar compliant platforms, we are witnessing a fundamental transformation in how data is managed and leveraged for scientific research across institutions."

"The introduction of the VAST InsightEngine with NVIDIA marks a major milestone in accelerating groundbreaking discoveries while enabling real-time global data sharing. With NVIDIA NIM™ microservices now seamlessly integrated into the VAST Data Platform, we are equipping the scientific community with next-generation tools to push the boundaries of discovery and innovation. The timing of the VAST InsightEngine enables real-time agentic AI, empowering autonomous agents to process, adapt, and act on dynamic data streams with minimal human oversight. This is vital as efficiency becomes a driving force in federal healthcare." noted Reese.

The collaboration between spinTwo and VAST Data transforms scientific research. By providing a scalable, high-performance data platform, VAST has enabled scientists to accelerate their research, allowing for quicker analysis and more impactful discoveries in life sciences.

About VAST Data

VAST Data is the leading data platform software company bringing businesses into the AI era. The VAST Data Platform delivers scalable performance, simplifies data management, and enhances productivity by accelerating time-to-insight for workload-intensive applications. Launched in 2019, VAST has become history's fastest-growing data infrastructure startup.

Looking to simplify your data center and discover insights in all of your data? **Contact us at hello@vastdata.com**