



Success Story

The Salk Institute Embraces AI and Big Data with VAST Data

Nestled along the Pacific coastline, the iconic buildings of the Salk Institute for Biological Studies have housed scientific visionaries and trailblazers for over 60 years. The Institute is dedicated to improving the quality of life worldwide with advancements in neuroscience, cancer research, aging, immunobiology, plant biology, computational biology, and more.

Since its founding in 1960 by renowned physician and virologist Jonas Salk, the independent, non-profit research organization has made transformative breakthroughs – from fundamental discoveries about the underlying causes of cancer to unraveling the mysteries of how the brain processes information.

Behind these life-changing innovations lies an organization deeply committed to pushing the boundaries of human knowledge through pioneering research fueled by ingenuity and enabled by cutting-edge infrastructure. One innovator at the frontier is Talmo Pereira, Salk Fellow and Principal Investigator. Today, Talmo focuses on quantifying motion using deep learning and computer vision to discover behavioral biomarkers of disease – understanding how movement patterns indicate disease. As Talmo explains, “There have been many discoveries about how the brain processes information, controls movements of the body and processes social information.”

Transitioning to VAST Data

Talmo's journey to finding the proper data infrastructure took a lot of work. “I was bewildered, essentially going through our incumbent purchasing workflow and requisitions,” he recalls. “That’s when I started looking around. Investing all this money on GPUs, it made sense to reevaluate the rest of the stack. That’s where we got to seriously thinking about a data platform.”

The challenges compounded as Talmo's data needs grew. “I needed a bit more data passed around. There wasn’t a good standard solution. The refurbished system I shared with several others wasn’t cutting it. We spent more time troubleshooting and working around it than having it be a layer we don’t think about.”

About Us

VAST is the data platform for simplicity at scale. VAST engineered the first new scale-out architecture in 20 years to take you beyond the conventional limits of scale, resilience and cost.

↳ Looking to simplify your data center and unlock insights from all of your data? Contact us at hello@vastdata.com

”

“I needed a bit more data passed around. There wasn’t a good standard solution. The refurbished system I shared with several others wasn’t cutting it. We spent more time troubleshooting and working around it than having it be a layer we don’t think about.”

– Talmo Pereira,
Salk Fellow and Principal Investigator



This shifted when Talmo learned about VAST Data from trusted partners. “When we heard from our partners at Cambridge Computers about VAST, that it was kind of the hot new thing we should check out – that’s when we saw this would fit the bill. It simplifies things and would fulfill any needs we could foresee.”

The Vision: Harnessing AI’s Power

Even trailblazing institutes must periodically modernize systems and workflows to stay at the top. Robust data pipelines and shared AI/ML infrastructure empower teams to unlock AI’s full potential.

Recognizing AI/ML’s broad potential and critical need across research disciplines, Talmo embarked on an ambitious BioComputing Initiative to evolve the scientific and research computing infrastructure. Continuing pioneering work required an infrastructure supporting data-intensive artificial intelligence workloads beyond what existing systems delivered.

“AI and deep learning integration into research methodologies is not just a trend but a necessity. Every area of scientific research now has a large component of high-resolution data that is impossible to make sense of without appropriate computational tools,” explains Talmo. Through his lab’s specialization in applying deep learning and computer vision to study biology computationally, Talmo witnessed firsthand AI/ML’s transformative impact.

The Journey to VAST Data

In reimagining its research data architecture, the Salk Institute turned to VAST Data, the industry-leading platform for AI-powered computing. “We wanted to future-proof the Institute’s data capabilities,” Talmo notes. After conducting in-depth due diligence evaluating solutions on performance, scalability, and total cost of ownership, VAST Data emerged as the clear choice. Talmo’s lab runs on premier infrastructure comprising many AI/ML enabling technologies, including VAST, NVIDIA, and Run.ai. Run.ai offers fair-share scheduling, allowing the Salk data science and AI team to quickly and automatically share clusters of GPUs without memory overflows or processing clashes.

”

“It was all white glove,” Talmo recalls about the deployment. “VAST came in, they installed everything, and it was pretty much turnkey from day one.”

– Talmo Pereira,
Salk Fellow and
Principal Investigator



Crafted to tackle complex computational workloads, the VAST Data platform is optimized for high performance. It makes 100% of an organization’s data AI-ready with revolutionary data reduction technology that shrinks research data sets by up to 90%. This powerful combination accelerates and simplifies data pipelines across the spectrum of research workflows – from ingest to training to analysis.

System fragmentation ceases to pose barriers with VAST Data’s unified scale-out architecture, which can consolidate large data lakes into a single namespace while delivering high performance. The built-in global namespace enables seamless accessibility from anywhere across standard protocols, facilitating secure data sharing between collaborators.

The VAST Disaggregated Shared-Everything (DASE) architecture allows for independent scaling of capacity and performance. This flexibility, combined with enterprise features like QoS policies, enables the Salk team to adapt to new workloads without the complex tuning and downtime required by legacy solutions.

“It was all white glove,” Talmo recalls about the deployment. “VAST came in, they installed everything, and it was pretty much turnkey from day one.”

Understanding Data: Accelerating Discovery

The team has developed a large-scale phenotyping system with VAST Data as an integral component. “It’s a system that allows us to capture the behavior of our preclinical rodent models,” Talmo explains. “We have mice that live in their home cage at the bottom of these boxes. We can record high-quality video 24/7 at 50 frames per second and capture all their movements. From this, we can derive phenotypes like tremors or motor coordination that give us insights into different diseases.”

This high-frequency, high-resolution video footage is stored on VAST and serves as input into AI models. The video is transformed into quantitative data points to characterize animal behaviors over time. Enabled by VAST’s low-latency, high-throughput performance, Talmo’s team can efficiently process these massive datasets. “We can iterate through the data, do the AI inferencing, then pass it to a probabilistic model to characterize the structure – essentially the body language of the observed animals based on their trajectories.”

The VAST Data platform rollout has delivered rapid results. Talmo observes tangible improvements in his lab’s AI/ML pipelines, highlighting noticeable latency reductions in streaming research image and video datasets from centralized data tiers to GPU compute nodes. This accelerated data ingestion enables faster model development cycles.



“The increased throughput is a huge help for our deep learning workloads where high I/O is critical,” he explains. “Being able to mount and operate seamlessly across platforms without file permission or flag issues makes our lives much easier. There’s zero IT overhead regarding management, since VAST handles that for us.”

– Talmo Pereira, Salk Fellow and Principal Investigator



Talmo’s research recently expanded beyond earth to include space as well. NASA is funding his lab to study and to protect astronaut health on long space missions. NASA is pursuing the best methods and technologies to support safe, productive human space travel of which Talmo’s lab is contributing via the lab’s research.

By taking a strategic, forward-looking approach to modernizing its research data infrastructure, the pioneering Salk Institute continues pushing boundaries – poised to accelerate the next wave of scientific breakthroughs.