

## SUCCESS STORY

### ACCELERATE: LIFE SCIENCES

Van Andel Research Institute Optimizes HPC Pipeline, Driving Research Discoveries and New Drug Therapies with End-to-End DDN Storage Solution.



# VAN ANDEL INSTITUTE

## CHALLENGES

- Fragmented storage pools for research and instrumentation data were costly, cumbersome and lacked sufficient safeguards
- Addition of high-powered, cryo-electron microscopy was anticipated to quadruple existing storage
- Ever-increasing requirement to ingest, process, store, archive, and share research
- Parallel file system was needed to address storage needs while providing enterprise data management capabilities

## SOLUTION

DDN's GS7K<sup>®</sup> parallel file system appliance and WOS<sup>®</sup> object storage appliance

## RESULTS

- Simple, fast access to 2PB of storage for research and instrument data to address exponential storage growth and active archive requirements
- Infinitely scalable storage capacity available to keep pace with increased use of Cryo-EM and next-gen sequencing technologies

### VAN ANDEL INSTITUTE HAS A 20-YEAR LEGACY OF BIOMEDICAL RESEARCH AND SCIENTIFIC

education with a focus on improving health and enhancing the lives of current and future generations. Formed in 1996 by Jay and Betty Ann Van Andel, the Grand Rapids, Michigan-based organization has evolved into a premier center for research and education that supports more than 360 scientists, educators, and staff.

At Van Andel Research Institute (VARI), a dedicated team works tirelessly to determine the epigenetic, genetic, molecular and cellular origins of cancer, Parkinson's and other diseases. Using state-of-the-art technologies and instrumentation, the Institute's scientists are working to translate discoveries into highly innovative and effective diagnostics and treatments.

For example, scientists in VARI's Center of Epigenetics are shedding light on the mechanisms that control how genes are regulated to determine what happens when a cell transitions from a normal state into a rapidly dividing cancer cell. This field, called epigenetics, is rapidly emerging and an important area of cancer research, according to Dr. Scott Rothbart, assistant professor in VARI's Center of Epigenetics.

"We're answering key questions about the molecular mechanisms of gene regulation," he explains. "These findings give us new vision for how to target cancers with innovative drug therapies. Of course, this research generates more data than ever before and requires more computing and storage than previously available."

## THE CHALLENGE

In the Rothbart Laboratory, scientists perform time- and data-intensive molecular dynamics simulations to model protein behavior in silico. "We employ molecular dynamics in part to simulate and predict protein-protein and protein-drug interactions," Rothbart says. "This technique is rapidly accelerated by high-performance computing."

Zachary Ramjan, the Institute's Research Computing Architect, joined the organization in 2014 to ensure there would be ample compute and storage power to continually push the research envelope. "I was given the freedom to build a high-performance compute and storage solution from the ground-up that would meet both current and future research needs," he explains. "Our goal was to create a progressive computing platform anchored by powerful, scalable storage."

In setting the stage, Ramjan sought to replace the organization's fragmented storage silos with primary shared storage harboring instrument and other research data. "Centralized storage produces major cost savings," he adds. "It also provides an extra measure of protection by moving irreplaceable research and instrumentation data from individual hard drives onto a single system."

## BUSINESS BENEFITS

- Projected savings of hundreds of thousands of dollars by centralizing storage for state-of-the-art scientific instruments
- Researchers are forming the future of chemistry through increased molecular dynamics simulations
- Cryo-EM technology will push research boundaries and scientific breakthroughs with near atomic-level images

“DDN makes it incredibly easy for us to put data where it best belongs, all within the context of a single system. The ability to store data in the most performance- and cost-efficient place gives us flexibility to grow as research needs dictate.”

**Zachary Ramjan**  
Van Andel Institute's Research Computing Architect

“The new electron microscopes will definitely push the boundaries of research while enabling us to see things at much higher resolutions. Having sufficient compute and storage capacity is critical, as these high-end microscopes generate about 20TB of data every three days. Using cryo-EM, we're advancing studies on GPCR receptors and DNA modifications in ways we couldn't before,” adds Zhao. The storage needs to keep up, as each cryo-EM project can produce 40TB of data and we'll likely support a dozen projects simultaneously.”

**Dr. Gongpu Zhao**  
Cryo-EM Core Manager

Centralized storage and computing accommodate major growth, including the Institute's expanding structural biology research program, which is home to a suite of cryo-electron microscopes (cryo-EM). This technology, which enables scientists to see the structure of molecules that are one-tenth the width of a human hair, is transforming investigations of small yet vital components of human biology. “The installation of highly specialized electron microscopes quadrupled our current storage capacity,” Ramjan says. “Performance and speed are crucial for accommodating these data-hungry instruments.”

Approximately 30 researchers are getting trained on the Institute's FEI Tecnai Spirit G2 BioTWIN, FEI Talos Arctica and FEI Titan Krios microscopes. “The new electron microscopes will definitely push the boundaries of research while enabling us to see things at much higher resolutions,” says Dr. Gongpu Zhao, Cryo-EM Core manager. “Having sufficient compute and storage capacity is critical, as these high-end microscopes generate about 20TB of data every three days.”

Cryo-EM can produce images at near-atomic level resolutions, offering many unique advantages over other methodologies, such as X-ray crystallography. The resulting strain on compute and storage can present challenges, especially since multiple users need to access images simultaneously. “Using cryo-EM, we're advancing studies on GPCR receptors and DNA modifications in ways we couldn't before,” adds Zhao. “The storage needs to keep up, as each cryo-EM project can produce 40TB of data and we'll likely support a dozen projects simultaneously.”

## THE SOLUTION

As a result, the organization began a thorough evaluation of next-gen HPC and storage solutions, including cluster and cloud computing, as well as parallel file and object storage. On one hand, VARI wanted to take advantage of the performance and scalability delivered by a traditional GPFS-based parallel file system. On the other hand, the organization also wanted the flexibility to implement a private cloud for storing and crunching large amounts of data efficiently and cost-effectively.

“I had a gut instinct that a hybrid on-premises and cloud computing solution was the best path to the future, which was proven out as we deployed our hybrid solution, says Ramjan. “This approach gave us the flexibility to grow and scale computing on demand.”

The storage decision was driven by the need for a parallel file system and object storage to handle data ingest, processing, storage, collaboration, and archiving. The team reviewed several file-system solutions before choosing GPFS (now known as IBM® Spectrum Scale™), which offered all the desired features in an easily expandable platform. VARI also looked closer at its existing EMC Isilon storage before determining it would be too cost prohibitive to scale the platform in keeping pace with rapid storage growth.

In the decision process, Ramjan drew upon extensive experience with DataDirect Networks (DDN) from working on the next-generation sequencing team at University of Southern California. “We were really happy with the performance and scalability of our DDN parallel file system storage appliance,” he says. “Because of that experience and its Spectrum Scale data management capabilities, we chose DDN's GRIDScale® GS7K parallel file system appliance.”

DDN's GS7K® solution is designed to minimize total cost of ownership. The purpose-built, all-in-one appliance includes a scale-out parallel file system, high-performance storage, and enterprise-class data management capabilities. “DDN's GS7K gives us a big data storage solution that blends the performance of a Spectrum Scale parallel file system with enterprise-class capabilities we needed, including snapshots and rollbacks, replication, seamless cloud connectivity, and data tiering,” says Ramjan.

Selecting DDN's implementation of Spectrum Scale also meant that VARI could take advantage of DDN's unique capability to present a single, federated namespace across file and object. This was a key consideration in VARI's selection of DDN's WOS® Object Storage as both an active archive for storing ever-increasing amounts of unstructured data, and as a research collaboration solution that facilitates data sharing and research collaboration.

## TECHNICAL BENEFITS

- DDN GS7K with IBM® Spectrum Scale™ delivered performance, scalability and enterprise features, including snapshots, rollback and replication
- OpenStack® driver simplified storage integration
- WOS object storage ideally suited for active archiving and data tiering

“

*We now have the technology backbone to support novel research that translates into groundbreaking therapeutic interventions. The work we're doing in molecular dynamics, which is supported by DDN, is informing the future of chemistry.* ”

**Dr. Scott Rothbart**

Assistant Professor, VARI's Center of Epigenetics

“

*We'll save hundreds of thousands of dollars by centralizing storage on DDN for data-intensive research and a dozen data-hungry scientific instruments. We can elevate the standard of protection, increase compliance and push the boundaries of science on a single yet highly scalable storage platform. That's why DDN is core to our operation and a major asset to our scientists.* ”

**Zachary Ramjan**

Van Andel Institute's Research Computing Architect

“The storage was actually one of our easiest decisions,” Ramjan adds. “With DDN's GS7K and WOS, we could offer HPC users a place to work on their data and archive results while also ingesting massive amounts of instrument data from dozens of next-gen sequencers and electron microscopes.”

## THE BENEFITS

Betting on DDN has paid off several times for VARI, starting with the flawless implementation of two GS7Ks with 1.5PB and WOS with a half-petabyte of capacity. “Our DDN systems were very easy to deploy,” says Ramjan. “In particular, DDN's OpenStack® driver support was a welcome surprise and a huge bonus because it significantly streamlined our OpenStack storage integration.”

VARI took advantage of DDN's deployment flexibility to implement storage at two sites for increased redundancy and protection of the institute's most sensitive datasets. The addition of WOS further simplified data movement, tiering, protection and replication that are all handled automatically across file and object. “DDN makes it incredibly easy for us to put data where it best belongs, all within the context of a single system,” Ramjan notes. “The ability to store data in the most performance- and cost-efficient place gives us flexibility to grow as research needs dictate.”

With the Institute's Cryo-EM now online, expectations are that up to 13TB of data will be generated each day through new ways of conducting single-particle analysis and tomography. “Extremely large datasets are automatically transferred to storage without our scientists having to worry about it at all,” says Zhao. “It's quite amazing that we can keep our data there and everyone can access it at the same time. This will allow our scientists to conduct even more groundbreaking research and accelerate the pace of major scientific discoveries.”

Additionally, scientists can collaborate more effectively with other research centers as DDN's storage scales seamlessly. “I just had a request from a researcher who is pulling down a 100TB RNASeq dataset from The Cancer Genome Atlas project,” says Ramjan. “Now we can accommodate this kind of request easily without causing storage constraints for anyone else.”

In fact, DDN's scale-out capabilities can accommodate ongoing spikes in computation and data collection generated by a host of workflows, including GROMACS and AMBER for molecular dynamics simulations and GATK for next-gen genomics toolsets. “We now have the technology backbone to support novel research that translates into groundbreaking therapeutic interventions,” says Rothbart. “The work we're doing in molecular dynamics, which is supported by DDN, is forming the future of chemistry.”

Major cost savings will be realized with DDN as the primary storage for more than a dozen state-of-the-art scientific instruments, including Illumina NextSeq and iScan sequencers, Aperio slide scanners, and the new crop of electron microscopes. “We'll save hundreds of thousands of dollars by centralizing storage on DDN for data-intensive research and a dozen data-hungry scientific instruments,” says Ramjan. “We can elevate the standard of protection, increase compliance, and push the boundaries of science on a single yet highly scalable storage platform. That's why DDN is core to our operation and a major asset for our scientists.”

In the Rothbart lab and other labs within VARI, the institute's powerful HPC and storage environment is ready to serve an increasing cadre of scientists with diverse research demands and aggressive project timelines. “Science and technological advances are making it possible for us to ask different questions, run longer simulations, share more insight, and solve problems as part of our mission to improve health for future generations,” says Rothbart.

Unexpectedly, word spreading of VARI's newly enhanced HPC and storage environment has resulted in the organization having a new and powerful recruiting tool: More scientists want to join their ranks. “It's interesting to see how important storage is to researchers,” concludes Ramjan. “They recognize it's pointless to have all the latest scientific instruments if you can't capture, collect, analyze, and share the research data safely and reliably.”

## **ABOUT DDN®**

DataDirect Networks (DDN) is the world's leading big data storage supplier to data-intensive, global organizations. For more than 18 years, DDN has designed, developed, deployed, and optimized systems, software, and solutions that enable enterprises, service providers, universities, and government agencies to generate more value and to accelerate time to insight from their data and information, on premise and in the cloud. Organizations leverage the power of DDN technology and the deep technical expertise of its team to capture, store, process, analyze, collaborate, and distribute data, information, and content at largest scale in the most efficient, reliable, and cost effective manner. DDN customers include many of the world's leading financial services firms and banks, healthcare and life science organizations, manufacturing and energy companies, government and research facilities, and web and cloud service providers. For more information, visit our website [www.ddn.com](http://www.ddn.com) or call 1-800-837-2298.