KALRAY
Customer Success

PURDUE UNIVERSITY

Cuts workload times by 25%

Challenge
Purdue needed a reliable and efficient way to manage massive datasets across multiple research labs.

Solution
They chose Ngenea to orchestrate data movement, optimizing storage tiers and ensuring data access for researchers working in multiple locations.

Results
Ngenea accelerated high-end applications, scaling effortlessly while reducing costs. This optimization reinforced Purdue's status as a leader in innovation and scientific discovery.

Accelerating Discovery with Supercomputing

Purdue University uses data management to drive research forward, faster.

Purdue University is widely regarded as a technology powerhouse, renowned for its top-ranked engineering programs, cutting-edge research in computing, and a strong emphasis on innovation. The university excels in aerospace engineering and actively contributes to the fields of semiconductors, cybersecurity, and supercomputing. With a culture of entrepreneurship and a commitment to translating research into real-world applications, Purdue stands as a beacon of technological excellence.

Students and faculty come to Purdue’s Rosen Center of Advanced Computing from over 60 departments, every academic college, and three campuses, and pursue innovation in fields including fluid dynamics, aeronautics...one researcher is even using AI to count the trees on planet earth.

Ngenea Delivers Data On-Demand for over 600 Research Labs

Over 1600 active users now benefit from accelerated simulations and research, with significant improvements in data management and cost efficiency.

Performance and Efficiency
Ngenea integration has reduced bioinformatics workload runtime by 25%, significantly accelerating Purdue's high-end modeling and simulation applications, allowing researchers to achieve faster results.

Data Accessibility
Ngenea's efficient data orchestration ensures seamless, on-demand access to datasets from labs, desktops, and supercomputers, enabling Purdue to maintain its rapid pace of innovation and research productivity.

Scalability
The Research Data Depot, powered by Ngenea, scales effortlessly to meet Purdue's growing data demands, maintaining high performance and accessibility without disruption.

Data Protection
Ngenea provides robust data redundancy and protection, ensuring the reliability of Purdue's research data infrastructure and safeguarding against hardware failures and accidental deletions.

Cost Optimization
By tiering less critical data to cost-effective storage, Ngenea optimizes Purdue's total cost of ownership, ensuring high-speed resources are available for the most demanding applications.

The challenge
As a leading research institution, Purdue is home to more than 1900 faculty, over 600 of whom store extremely large and diverse datasets – from source code to climate models to bioinformatics. Considered a high-performance computing hub in the Midwest, Purdue wanted to create hyper-efficient workflows across several locations that would operate without slowing their pace of discovery.

Since data is generated and accessed by multiple labs, student desktops, scientific instruments, and supercomputers, Purdue needed a performant and reliable mechanism to connect data from a central repository to application sites. Their tool of choice would need to contain, secure, and protect researchers' data while offering compatibility with a full range of research applications. It had to be fast, user-friendly, and above all – reliable.

A previous solution struggled with volume, often leaving researchers without access to their data when and where they needed it. Slowed timelines delayed important research, so to keep up with their track record of innovation, the scientists at Purdue built a centralized storage hub with reliable data access called the Research Data Depot.

The solution
The Data Depot offers usable capacity for important projects to thousands of researchers. It provides data storage in increments of 1TB and uses the Storage Scale filesystem to facilitate joint work on shared files across research groups.

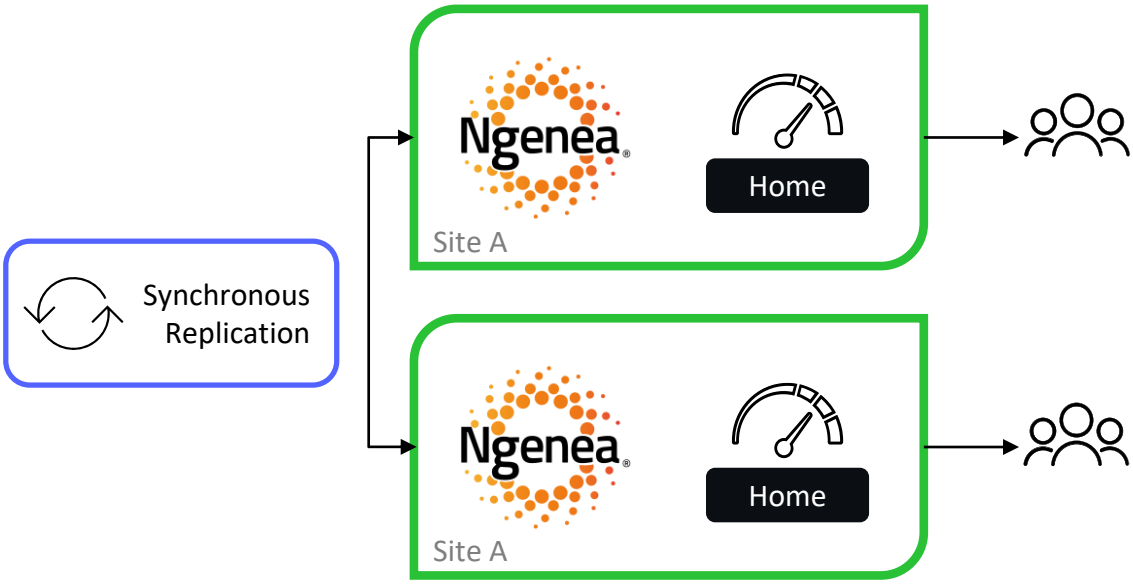
> **"We turned to Kalray for a proven high-performance software stack that allows us to take data from all these places on the edge and get it onto the supercomputers where they can be used for high-end modeling and simulation."**
>
> *Preston Smith*
> *Executive Director, Rosen Center for Advanced Computing*

They came to Kalray for high-performance software to coordinate data movement from research locations into university supercomputers to be used for high-end modeling and simulation. Kalray Ngenea was installed first to act as a management layer, on top of Purdue's existing storage stack. Ngenea could seamlessly move data to the appropriate storage tier, freeing up higher tiers for the most important workloads, and moving less critical data to less-expensive storage. That data orchestration and automation gave Purdue more time and people-power to continue a bigger build: a complete storage and data management solution.

Because Purdue serves thousands of researchers across hundreds of labs, it was important that Kalray could serve their data from the same storage platform, over multiple network technologies – from InfiniBand to Ethernet – to different compute islands. Kalray's network-agnostic approach gave Purdue the flexibility to serve their entire data landscape.

Preston Smith, Executive Director of the Rosen Center for Advanced Computing said, "We turned to Kalray for a proven high-performance software stack that allows us to take data from all these places on the edge and get it onto the supercomputers where they can be used as well for high-end modeling and simulation."

Using Kalray Ngenea storage, the Data Depot is redundant and protected against hardware failures and accidental deletion. All data is mirrored at two different sites on campus to provide for greater reliability and to protect against physical disasters.



The results

Today, 200 faculty members and over 1600 active supercomputer users use the university's Data Depot to run and store active simulations as they are engaged in solving some of the world's most complex problems.

And the experience with the Data Depot has allowed Purdue to explore the art of the possible. With support from the National Science Foundation, Purdue deployed a supercomputer called Anvil to support scientists across the country, again partnering with Kalray to deliver the storage subsystem. With a high-performance NVMe layer with petabytes of hard drive below, Purdue takes advantage of a high-speed flash tier to accelerate modeling and simulation applications and AI jobs while then safely tiering off to a lower cost spinning disk tier, which optimizes TCO and resource utilization for the nation's science and engineering community.

"The results of that capability for Anvil had been really profound. In our early user period on the system, one of the early adopters reported that Anvil was the fastest supercomputer that he had used for his bioinformatics workload. And that it had shaved 25% off the runtime for his workload. And a lot of that is through that data management tier made possible with Ngenea."

- Preston Smith, Executive Director, Rosen Center for Advanced Computing