

# UC San Diego Center for Microbiome Innovation Breaks Data Bottlenecks, Accelerating Discovery of New Methods to Improve Human Health and the Environment

Using Panasas ActiveStor, CMI Develops Novel Tools and Methods to Analyze and Manipulate Microbiomes

## SUMMARY

### CUSTOMER

UC San Diego Center for Microbiome Innovation (CMI)

### SIZE

More than 150 labs across campus

### LOCATION

San Diego, California

### INDUSTRY

Life Sciences

### CHALLENGE

- Support data-intensive research that requires analysis of dynamically growing volumes of data
- Execute growing number of I/O-intensive tasks without compromising system performance
- Enable the use of advanced scientific technologies to support innovative research methods

### SOLUTION

- Supports rapid analysis, storage, and retrieval of huge volumes of data with consistently high performance
- Enhances control of storage resources with flexible workload processing
- Simplifies storage maintenance and optimization with frustration-free manageability

### RESULTS

- Improves the user experience, removing storage bottlenecks and system delays for researchers
- Speeds data exploration and discovery, supporting scientific innovation
- Enhances control and usability of the compute environment, ensuring resources are available when needed
- Simplifies storage management for administrators



“Panasas technology supports the mission of the Center because it never limits our exploration.”

#### Yoshiki Vázquez-Baeza

Associate Director of Bioinformatics Integration  
University of California San Diego  
Center for Microbiome Innovation

Microbiomes are the distinct, diverse communities of bacteria, viruses, and other microorganisms that live in, on, and around us. They help us digest and process nutrients, interact with our immune systems, and play surprising roles in our lives and environment – many of which are still being discovered.

The Center for Microbiome Innovation (CMI), part of the University of California San Diego, works to accelerate the research and understanding of microbiomes. CMI offers advanced technologies and expertise in clinical medicine, bioengineering, computer science, and biological and physical sciences. It also supports core facilities for data generation, bioinformatics, and data analysis.

Using many of the latest technologies – such as genomics and metagenomics – the center processes hundreds of

thousands of human, animal, food, and built-environment samples annually. These advanced technologies, combined with scientific instruments such as DNA sequencers and mass spectrometers, help CMI researchers, graduate students, and post-doctoral researchers in various research labs develop novel tools and methods to improve human health and benefit the environment.

Performing genomic sequencing and analysis of microbiomes requires huge volumes of data. Within CMI, each of the nearly 300 users is assigned local storage across two clusters. However, those legacy resources are becoming insufficient to process high I/O tasks such as DNA sequence processing or analysis of other multi-terabyte data sets. Attempts at processing current data volumes with traditional storage technologies resulted in degraded performance for the computational workflows, which compromised the pace of discovery.

Three years ago, one of the main research groups at the Center began performing large-scale metagenomics sequencing, an advanced technology for microbial surveying. “The scale of that type of computing makes things break left and right,” adds Yoshiki Vázquez-Baeza, associate director of bioinformatics integration at CMI. “That was the catalyst. We knew then that we needed to make a change.”

“It’s not unusual for our researchers to concatenate several hundred files into one large file for analysis,” explains Jeff DeReus, systems administrator for CMI. “Trying to do so many reads at the same time would bring other storage systems to their knees. There wasn’t enough disk readability to handle that kind of workload in addition to the other tasks that execute on the cluster each day.”

### Data-intensive support for innovation

After considering other storage technologies and evaluating vendors, CMI deployed a Panasas ActiveStor® high-performance storage solution. DeReus, the systems administrator, had worked with Panasas storage technology at previous academic institutes and was familiar with its superior functionality, manageability, and scalability. “I had zero complaints about the choice of Panasas,” he says.

The new storage solution helped CMI create a foundation for continued research excellence. “Our teams often benchmark several methods or rework the benchmarking with new data sets,” says Vázquez-Baeza. “If we have to limit the benchmark methods because of storage concerns, we wouldn’t be able to explore the full breadth of scientific options. Panasas technology supports the mission of the Center because it never limits our exploration.”

### Rapid Performance

With Panasas, CMI researchers can rapidly store, retrieve, and analyze unprecedented volumes of data. The fast, efficient PanFS® parallel file system accelerates performance at every stage of the computational research process.

“By staging the data into a high-speed parallel file system like PanFS, we’re able to alleviate the load on the storage cluster itself,” says DeReus. “People can generate their intermediate files and do the analysis they need. That reduces the impact of these larger datasets on the other system users.”

It’s clear that job throughput has increased. “We know that researchers are able to crunch the data faster,” says DeReus.

With consistently high performance, regardless of the workloads being processed, the system prevents bottlenecks and allows CMI researchers to get their work done in a timely manner. This is especially critical when researchers are working to meet paper submission deadlines, applying to present their results at industry and academic conferences, or submitting grant applications.

### Flexibility and Control

It’s common for CMI researchers to need varying volumes of storage, depending on their projects and the current stage of the research. The Panasas ActiveStor solution allows CMI to adapt to changing workload requirements without the need for labor- or skill-intensive tuning and administration efforts.

“We have more people that want to get work done, using data sets that are orders of magnitude larger than anything we’ve seen,” says DeReus. “Having technology that takes the pain away from sluggish file systems or longer run times is really important.”

“A lot of our work is collaborative,” says Vázquez-Baeza. “If we cannot share the data between users or with other partners, that creates a roadblock. Having a reliable storage resource like Panasas ActiveStor facilitates a lot of the creative work that happens here. The Panasas technology supports the mission of CMI.”

### Simplified Storage

Researchers working on the next big scientific discovery aren’t typically aware of the storage infrastructure that supports their work – unless it’s inhibiting their progress. A lack of complaints is a reliable indicator that the ActiveStor solution is meeting user demand, according to DeReus.

With previous storage solutions, CMI researchers faced roadblocks to completing their research. During deadline-critical periods, some users were even unable to get a prompt to log onto their machines. This frustrated scientists, slowed discovery, and increased the complaints to IT.

By eliminating these problems, the Panasas solution simplified storage at CMI. “I don’t see the slow downs, which means

I don’t have to track the jobs that are crushing the file system throughput,” says DeReus. “There’s no need to try to remove a load from the system so we can bring everyone else back to life.”

System management tasks are essentially off the to-do list. “The graphical interface reports on all aspects of performance, so we don’t have to get down into the lower-level system details to diagnose issues,” he says.

### Accelerating Microbiome Insight

Looking ahead, CMI plans to continue using the latest generations of scientific technologies to understand more about microbiomes. With a focus on converting the possibilities of the research into solutions, CMI appreciates technology solutions that don’t distract researchers from the scientific problems they aspire to solve.

By freeing researchers from worries about storage, CMI can accelerate the adoption of advanced scientific technologies.

“We strive to ensure that technology never gets in the way of cool science,” says Vázquez-Baeza. “Being able to control the compute environment from beginning to end is critical to us. Panasas gives us the flexibility to do this without any impact to the user community.”