

Scale Up Storage with No Downtime

Gain Capacity with High Performance, Reliability, and Stability



Until recently, many Life Sciences organizations architected their compute and storage to handle primarily genomics-only workflows. But now, analysis of imaging from new and existing instruments—along with the growth of artificial intelligence and machine learning—are bringing an explosion in data storage requirements, new algorithms, and the need for higher raw processing capability. These expanding compute clusters and increasingly mixed application workloads will stress your storage infrastructure in ways you may not even imagine.

As you adapt to the changing research landscape, you need to plan for a data storage foundation that delivers high performance in a scalable, adaptable, and reliable way. This will insulate you from some of the changes and will release your IT staff from administering and maintaining infrastructure to let them become true partners in scientific discovery.

The Panasas File System (PanFS®) solution delivers key advantages that ensure a superior solution for scaling.

Parallel and Direct File Storage

Parallel file systems are the only file storage technology that can ensure fast enough data access for the deluge of new Life Sciences workloads as you scale up.

Panasas continues to innovate in the file system space, since defining the first modern parallel architecture 20 years ago. Each file individually striped across many Storage Nodes.

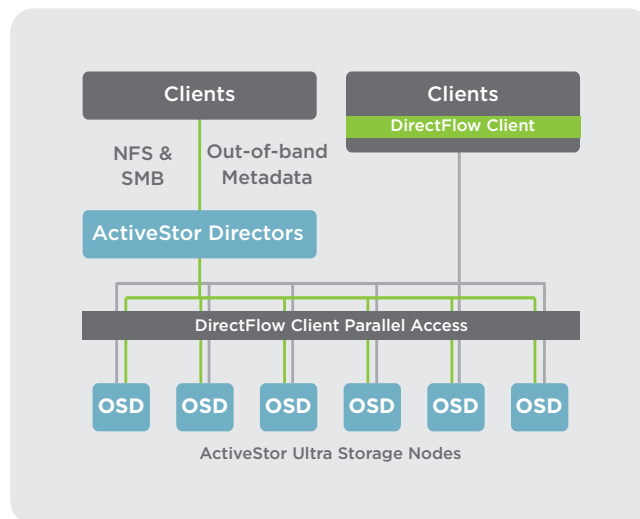


Figure 1. Parallel and Direct Transfers

That allows the file's component pieces to be read and written in parallel, increasing per-file performance. PanFS is also a direct file system, where the compute nodes talk over the network directly to all the Storage Nodes holding a file's data. There are no bottlenecks like you would find in a traditional "NAS head" architecture.

One PanFS customer reports having zero unplanned downtime in over 8 years of deployment of PanFS.

Seamless Placement of Metadata, Small Files, and Large Files

PanFS uses a carefully selected combination of Storage Class Memory (NVDIMMs), SSDs, and HDDs. By using just the right amount of each type of technology, and using it for what it does best, PanFS can deliver twice the performance from a given overall capacity as other products. All file metadata is stored on low-latency NVMe SSDs, and large files are stored on cost-effective, high bandwidth HDDs. This efficiently supports the mixed workloads in Life Sciences that include a wide range of file sizes and access patterns, that change small over time, supporting new instruments and algorithms.

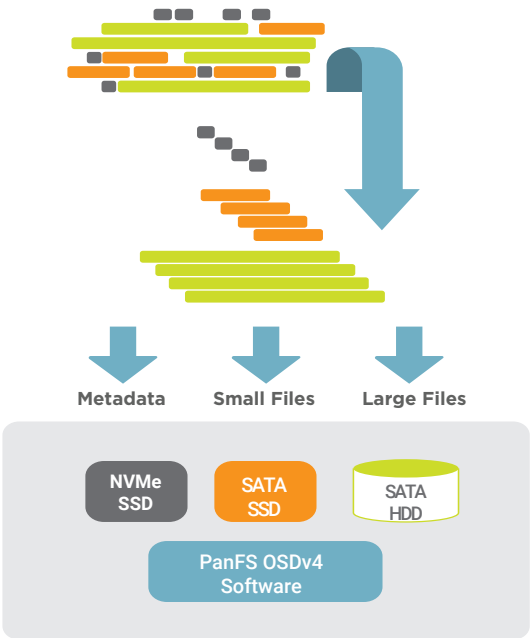


Figure 2. Seamless Placement of Metadata, Small Files, and Large Files

Transparent Growth and Automatic Balancing

Adding more ActiveStor enclosures when you need more space is transparent and immediately results in linear gains in performance and capacity, with no maximum upper limit. Scalable performance depends on spreading all your files evenly across the pool of Storage Nodes so PanFS uses two types of automatic balancing across the ActiveStor

enclosures. Balancing happens seamlessly and continuously, without any administrator intervention.

Greater Reliability with Scale-up

Performance means little if researchers face constant downtime. In other storage systems, more drives means more drive failures, so reliability typically goes down as scale goes up. PanFS uses software-based erasure codes to individually protect each file. That distributed file-level protection enables all the Storage Nodes to cooperate to recover after a hardware failure, reducing the time your data is at less than full protection levels. Continuous background scrubbing of the erasure coding ensures ongoing confidence your data is safe and protected.

Full-Scale Solution

With its modular architecture, PanFS running on Panasas ActiveStor helps you manage the explosion in Life Sciences data by easily scaling up capacity without losing performance, reliability, or stability. Parallel and direct file access eliminates hotspots or bottlenecks. A range of storage technologies, intelligently used, supports mixed workloads and a high price/performance ratio. Finally, the Panasas solution provides another advantage - the more data you add, the stronger your storage solution becomes.

One PanFS customer had a user community of several thousand researchers, each running their own applications. Over several years they gradually grew their PanFS filesystem to over 1,500 Storage Nodes and 150 Director Nodes. They saw linear growth in performance and increased reliability over that time, but no added administrative costs.

About Panasas



Panasas builds a portfolio of data solutions that deliver exceptional performance, unlimited scalability, and unparalleled reliability – all at the best total cost of ownership and lowest administrative overhead. The Panasas data engine accelerates AI and high performance applications in manufacturing, life sciences, energy, media, financial services, and government. The company's flagship PanFS® data engine and ActiveStor® storage solutions uniquely combine extreme performance, scalability, and security with the reliability and simplicity of a self-managed, self-healing architecture. The Panasas data engine solves the world's most challenging problems: curing diseases, designing the next jetliner, creating mind-blowing visual effects, and using AI to predict new possibilities.

Worldwide Office 1-888-PANASAS info@panasas.com	Panasas Headquarters San Jose, CA, USA Panasas Research & Development Pittsburgh, PA, USA	Panasas EMEA Oxford, United Kingdom emeainfo@panasas.com	Panasas APAC Sydney, Australia apacinfo@panasas.com	Panasas China Shanghai, China chinainfo@panasas.com
--	--	---	--	---