

Accelerating Financial Services with High Performance Parallel Storage



ActiveStor® Ultra with PanFS® is Panasas' latest and most adaptable, total-performance HPC storage system. PanFS combines a distributed and clustered parallel file system into an integrated platform to deliver predictable high performance that automatically adapts to financial services workloads. What's more, Panasas ActiveStor Ultra has the enterprise-class reliability, ease of management, and great support other parallel file systems simply can't match.

ActiveStor Ultra provides incredibly fast access to very large datasets from many clients in parallel, directly to and from physical storage devices. This significantly reduces backtesting times and performance scales linearly, as capacity is added, with no bottlenecks. Thus, with ActiveStor Ultra, financial services firms and agencies can develop, test, and deliver the fastest and most accurate financial models and strategies through precise and reliable quantitative analyses validated across extensive and complex datasets.

Data Growth in Financial Services

The financial services industry collects and analyzes enormous sets of data using state-of-the-art computational tools and techniques involving mathematical and statistical modeling, machine learning, and big data analytics.

This enables investment banks, brokerage firms, exchanges, and government finance agencies to deliver leading-edge risk management, quantitative trading, compliance checking, pricing, and tax systems strategies. Investment banking and trading firms have deployed thousands of compute nodes and petabytes of storage in high performance computing (HPC) data centers to meet increasing strategy accuracy and reduced testing times requirements in order to implement new and updated strategies. Today, those HPC storage resources are under increasing strain from a legion of talented quant teams concurrently demanding fast, shared access to the extensive and ever-growing datasets.

Backtesting Bottlenecks

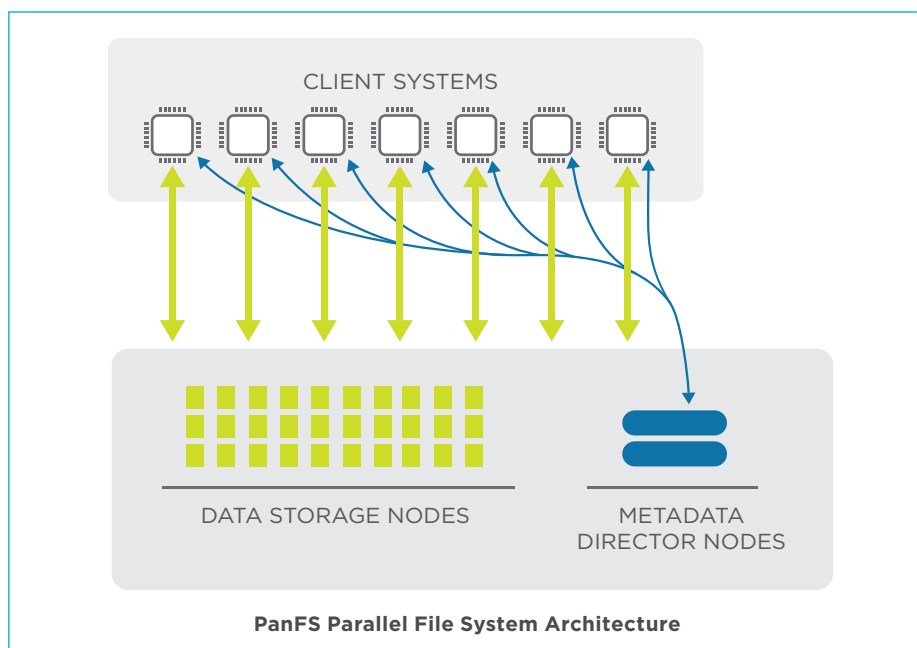
In order to determine how well financial services strategies perform, they are tested against historical data through a process called backtesting. Primarily conducted using the Monte Carlo Method, these simulations are iterated hundreds of thousands of times where completing 100K Monte Carlo simulation iterations can take several hours. One of the biggest constraints in using Monte Carlo simulations is storage access or I/O bottlenecks caused by the sheer volume of the highly parallel and data-intensive workload of the simulations themselves, coupled with concurrent storage access by multiple quant teams.

Solving Backtesting and Modeling Bottlenecks

One solution to the I/O bottleneck dilemma is a scalable, parallel, data storage system that provides excellent price-performance in a balanced architecture designed with financial services workloads in mind. The ActiveStor® Ultra solution running PanFS® is Panasas' latest and most adaptable, total-performance HPC storage system. It combines PanFS - a distributed and clustered parallel file system, volume manager, and software-based erasure-coding data protection into an integrated appliance platform. In addition to storage nodes, ActiveStor Ultra includes software stacks for the ActiveStor Director and PanFS DirectFlow® parallel data access protocol.

Enabling Direct Parallel Access to Data

Unlike traditional NAS or scale-out NAS architectures, ActiveStor Ultra uses three components working together to power the PanFS file system: Director Nodes, Storage Nodes, and the DirectFlow® Client driver. The Director Nodes and Storage Nodes are computer systems dedicated to running the PanFS software, and together they form the Panasas ActiveStor appliance.



PanFS explicitly separates the “control plane” from the “data plane” with the Director Nodes functioning as the control plane of the ActiveStor Ultra system, managing metadata services (e.g., directories, file attributes, etc.) instead of storing user data. In addition, the Director Nodes facilitate scalability and virtualizes data objects across all available storage nodes enabling the system to be viewed as a single, easily managed global namespace. Director Nodes can be scaled independently to scale metadata performance.

The Storage Nodes are the core of the data plane. For more capacity or more storage performance, more Storage Nodes can be added. In scale-out storage systems like PanFS, there simply is no maximum performance or maximum capacity and PanFS has been architected to provide linear scale-out; adding performance and capacity scale linearly as storage nodes are added.

The PanFS DirectFlow parallel data access protocol eliminates traditional scale-out NAS bottlenecks and hot spots by allowing compute clients to access all the Storage Nodes directly. As a result, hundreds of application clients have simultaneous and incredibly fast direct parallel access to very large financial

data sets. Direct parallel data access is important because while market data is generated in single streams, backtesting and analysis can be done in parallel with many clients reading and writing directly to storage.

PanFS also uses Network-Distributed Erasure Coding to ensure the highest levels of data integrity and reliability.

Framework for Unfettered Modeling

Technology infrastructure improvements can provide an improved framework allowing re-examination of constraints that necessitated model order reduction (MOR) including model truncation. With new storage architectures like ActiveStor Ultra with PanFS designed for parallel workloads, approaches for comprehensive, optimal higher order modeling and combinations of models become more viable, enabling increased accuracy and precision, more resilience to unanticipated market movements, and even decreased time to convergence.

Dynamic Data Acceleration

Pricing, market and credit risk management, fraud detection, operation risk and compliance, portfolio optimization and backtesting are key financial services applications that are computationally and data access intensive and can strain storage resources. Along with streaming large historical datasets for Monte Carlo simulations, backtesting can also be performed in combination with artificial intelligence (AI) including machine and deep learning or other neural network approaches and with Big Data analytics. These shared access mixed file size workloads require parallel infrastructure that is fast and extremely scalable, but adaptable to changing file sizes and mixed workloads without the need for tuning or manual intervention.

ActiveStor Ultra introduces the key to adapting to mixed workloads is a unique feature of PanFS called Dynamic Data Acceleration that eliminates the complexity, fragility, and manual intervention of tiered HPC storage. It maximizes the efficiency of all storage media in a seamless, all-hot system with a carefully balanced orchestration of DRAM for caching, extremely low-latency persistent memory, low-latency NVMe SSDs to store metadata, cost-effective SATA SSDs to store small files, and bulk high-bandwidth HDDs to store large files to provide a combination of excellent performance and low cost.

Built for Price/Performance, Engineered for Manageability

You can take PanFS from your dock to serving data in a day with a plug-and-play solution that is easy to install, manage and grow, and maintains the industry’s highest reliability and lowest total-cost-of-ownership (TCO) at any scale. With its modular architecture and building-block design, investment banks, trading firms, and exchanges, along with government financial agencies, departments, and services deploying ActiveStor Ultra can start small and scale linearly by adding metadata performance, or bandwidth and capacity, easing the strain on quantitative analyses backtesting and allowing improved modeling.