PANASAS



Dynamic Data Acceleration

A Better Way to Do HPC Storage



High performance computing sounds like it's all about the computing, but you need both the compute and the data to compute on to bring value to your organization. Panasas very nearly defined HPC storage with our parallel file system PanFS[®], and we've now introduced yet another innovation that moves the industry forward.

With the advent of new instruments such as Lattice Light Sheet (LLS) and Cryo-Electron Microscopy (CryoEM) in the Life Sciences space, as well as others in other disciplines, the amount of data and the pace at which it is generated continues to accelerate, putting enormous stresses on HPC storage systems.

It's very attractive to ask, "why can't we store our less frequently used data on lower cost storage?" The HPC market developed several techniques for doing just that, the most common of which is call "tiered storage". We saw the limitations of that approach and designed Dynamic Data Acceleration (DDA) to overcome them. DDA stores files in an entirely new way.

Let's take a look at both tiered storage and DDA.

Old-School Tiered Storage

Over the last few decades, tiered storage emerged as the most common method of balancing performance and cost. In practice, tiered storage systems usually move data between several tiers with different price/capacity ratios, based on recency of use. The theory is that if you haven't used a file in a while, you're not likely to use it for a while longer, allowing it to be stored on a lower-cost storage layer.

There are several challenges with this approach: the lower cost storage tiers are also lower performance, the customer must buy more storage performance overall to accommodate all the gratuitous movement of data between tiers, and since the lower tier(s) may not be directly accessible to the compute nodes (data must be moved to the hot tier before access) that hardware is not contributing to application performance.

The Genomics England case study is an interesting example of this (ask us for a copy). They purchased a traditionally tiered solution that had 3% of the total capacity on extremely fast (and expensive) NVMe storage and used an S3 archive for the other tier.

PANASAS

In other words, 97% of their data was slow. If their working set left that 3%, their performance either dropped dramatically, or the job had to wait while the data was migrated back to the hot tier, consuming hot tier bandwidth. An equivalently sized Panasas solution with DDA would have provided 3x the performance at an estimated half the cost, an overall 6x price/performance advantage as a result of our new Dynamic Data Acceleration architecture.

Dynamic Data Acceleration

Dynamic Data Acceleration (DDA) changes the status quo in two ways. First, it uses a file's size to decide where to store the file rather than recency of use. Second, each file is placed on the type of storage media best suited to that file.

DDA uses commodity storage drives differently, eliminating the complexity and costs of tiering, while delivering the highest possible performance at the lowest cost:

- Unmodified data and metadata are cached in DRAM
- Writes and metadata updates are committed to NVDIMM
- Metadata is stored on low-latency NVMe SSDs
- Files under about 1MB are stored on high IOPs flash SSDs
- Larger files are stored on low-cost, high-capacity, highbandwidth HDDs
- Small file accesses don't affect large file accesses, they're on different devices, and metadata doesn't wait for anything

All this is transparent to you, and usage of each type of drive is dynamically balanced, accelerating overall performance.



The result is that all the data being stored has the same high performance, increasing efficiency and reducing surprises. Everything a researcher wants to access is immediately available, using the aggregated performance of all the hardware in the storage subsystem.

With DDA, all the storage you buy will contribute to the performance you need.

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

© 2021 Panasas, Inc. All rights reserved. Panasas, the Panasas logo, ActiveStor, PanFS and DirectFlow are trademarks or registered trademarks of Panasas, Inc. in the U.S. and/or other countries. All other trademarks, registered trademarks, trade names, company names, and service marks are the respective properties of their holders. v2.11