

# Accelerating Time to Discovery

## Frustration-Free Data Storage For Scientific HPC Workloads

HPC User Forum, Lugano  
October 8<sup>th</sup>, 2019

Steve Perkins  
Systems Engineer – Panasas, Inc.



FRUSTRATION-FREE STORAGE

# FOCUS ON WHAT REALLY MATTERS TO YOU

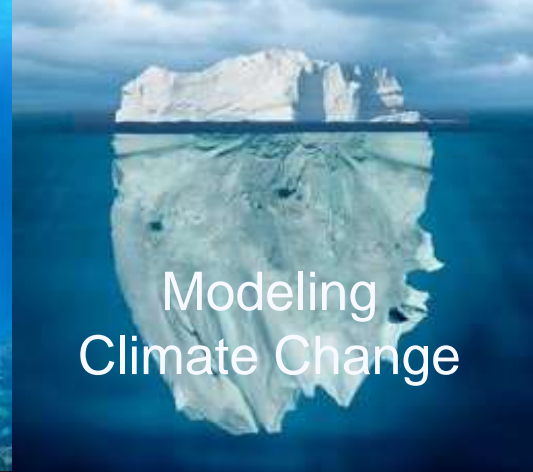
Curing Disease



Resource  
Exploration



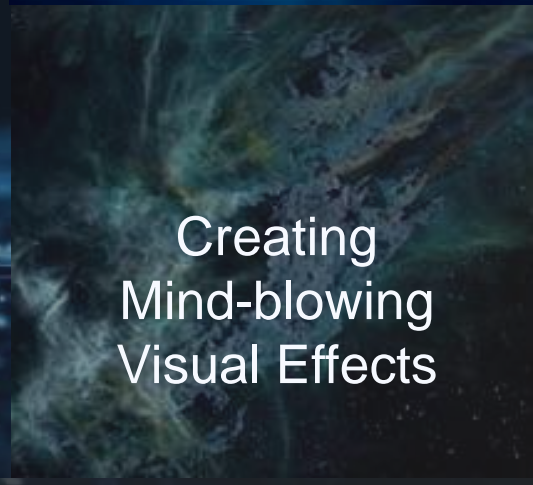
Modeling  
Climate Change



Enabling  
Artificial  
Intelligence



Creating  
Mind-blowing  
Visual Effects



Building the next  
Dreamliner



Winning  
Formula One Races





# HPC Storage Decision Factors

## Highest Performance

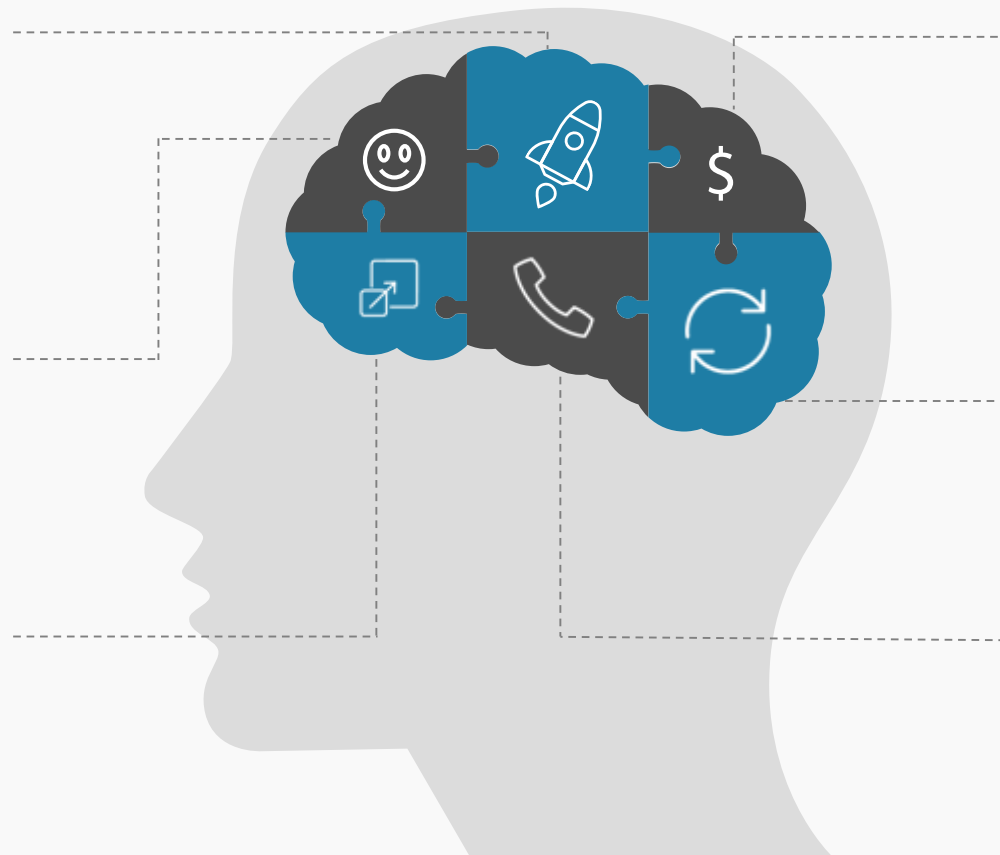
Consistently fast regardless of complexity.

## Frustration-Free

Set it and forget it.

## Flexibility

Dynamically adapts to changing workloads.



## Cost-Effective / Lowest IT Support Cost

Competitive acquisition cost combined with lowest cost of operation.

## Reliability

Uninterrupted user experience.

## Timely, Quality Support

Single vendor problem resolution.

# PanFS<sup>®</sup> – The Intelligent Filesystem

## Object Backend

Massive scale-out performance

## POSIX Filesystem

Standard APIs & ease-of-use

## Parallel Filesystem

Peak transfer performance

## Automated Recovery

Simplicity of self-managed failure recovery

## Per-File Striping

Spread & defuse all hotspots



## Single GUI/CLI

Simple to manage at any scale

## Erasure Coding

Distributed for extreme reliability

## Ethernet, InfiniBand,

## OmniPath\*

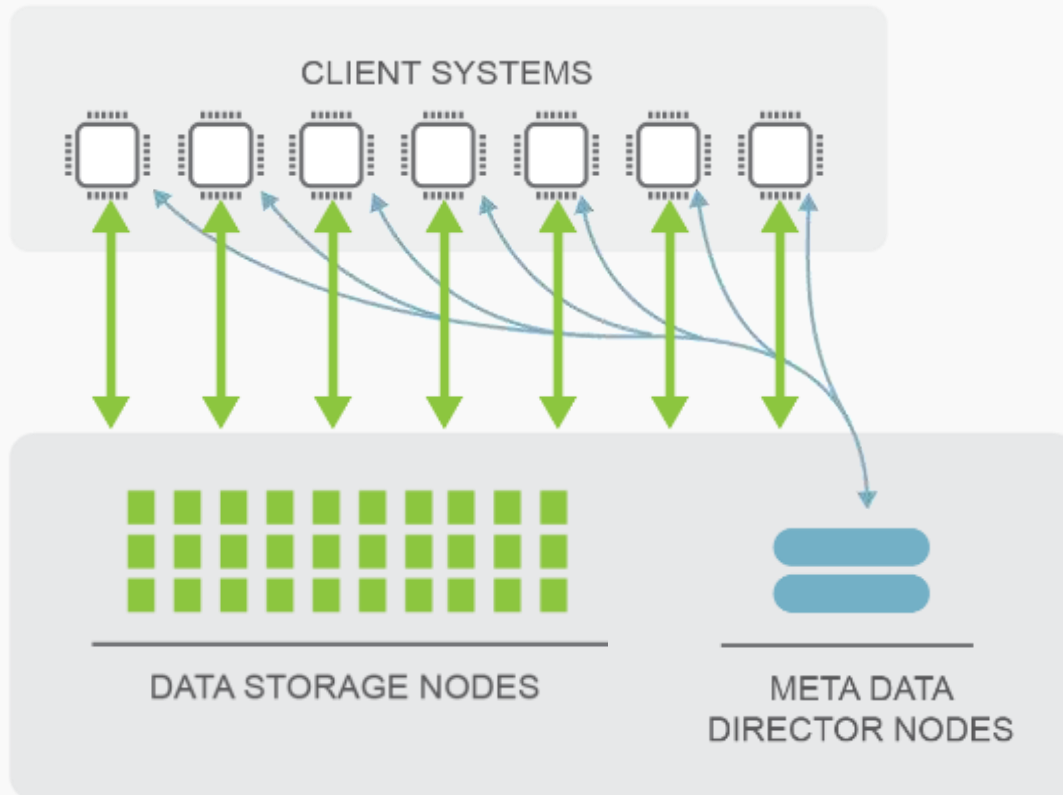
Flexible connectivity options

## Portable

Industry standard H/W and always the latest technology

# The Panasas Parallel Architecture

Designed for Linear Performance Scalability



## You Get

- Linear scalability of performance and capacity
- Limitless performance as you scale-out
- Reliability that improves with scale
- Excellent mixed workload performance
- Consistent user experience

## You Avoid

- Bottlenecks and hot spots
- Tuning/re-tuning when workloads change
- Noisy neighbor problems
- Productivity impact from failures
- Frustrated application users

# Panasas ActiveStor Product Evolution

Proprietary hardware to COTS with PanFS 8

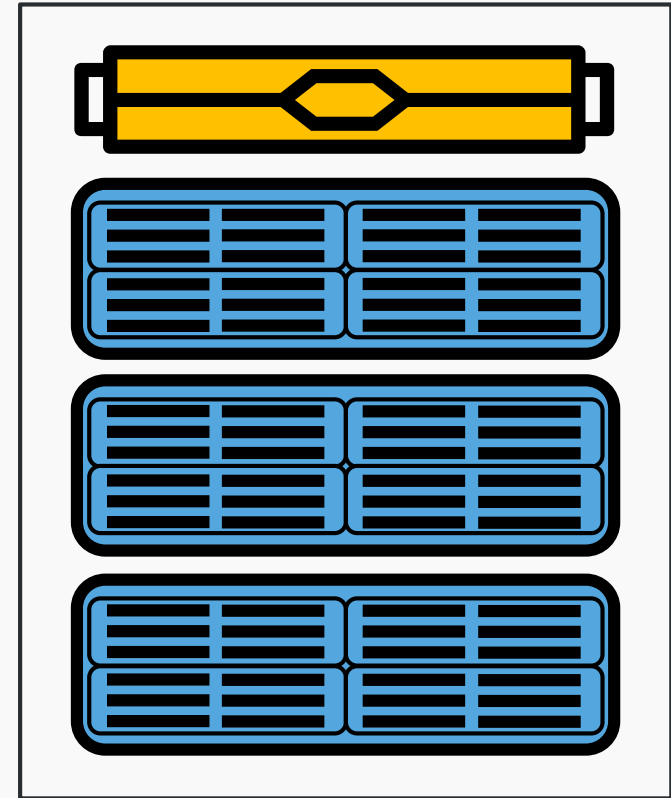
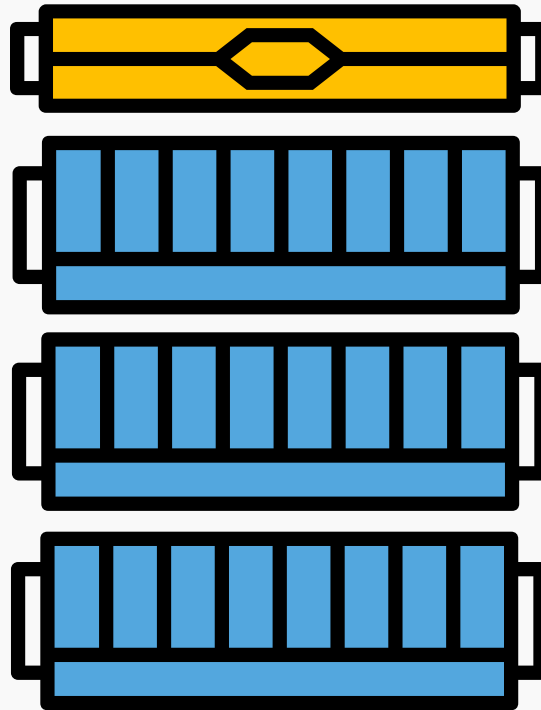
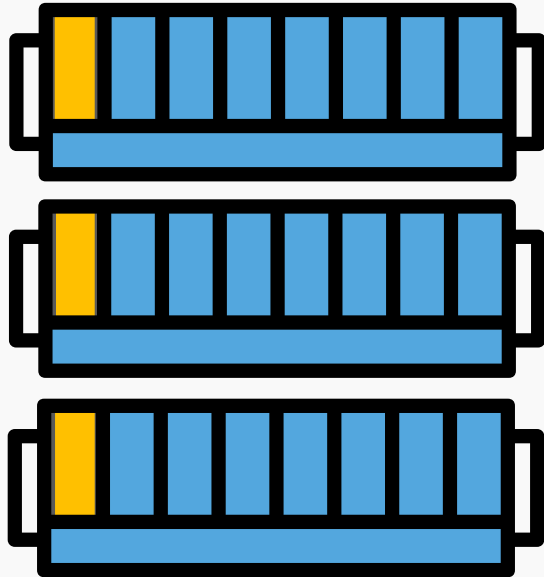
**ActiveStor Classic**  
formerly ActiveStor 18/20



**ActiveStor Prime**  
formerly ActiveStor Hybrid



**ActiveStor Ultra**



# ActiveStor Ultra

COTS Advantages Without the Headaches

## PanFS 8 ported to COTS hardware

- Continues the legacy of PanFS
- 25/40GbE and Infiniband options

## Scalable performance and capacity

- 4 GB/s building blocks
- Up to 300TB / 4U

## Flexible

- Designed for multiple workload types

## Simple to Manage

- Low administration overhead
- Single point of support for complete system

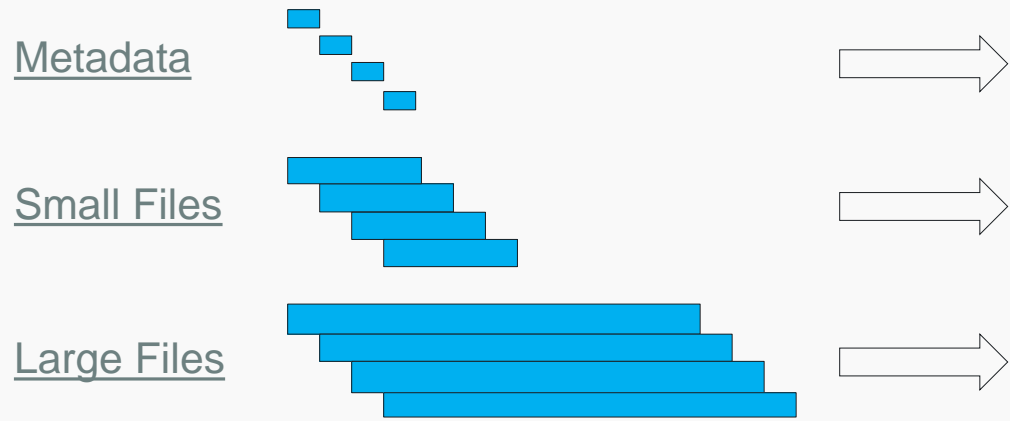


# ActiveStor Ultra

Performance for Mixed and Variable Workloads

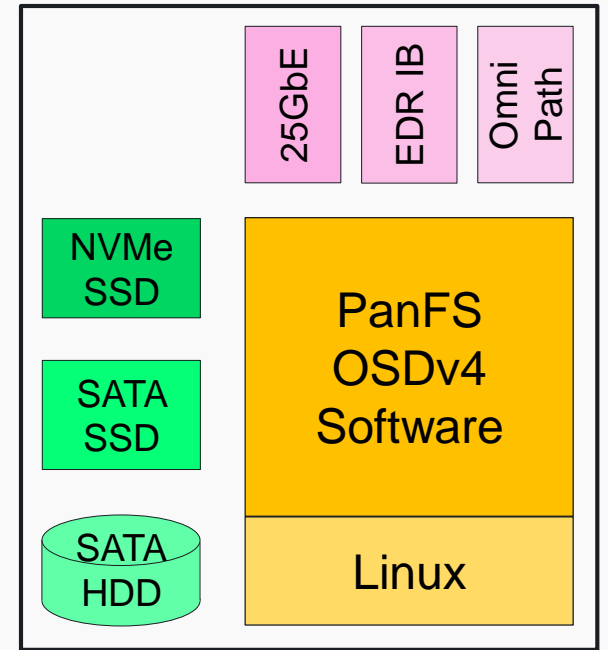
- Separation of metadata from files
- Minimise task-by-task tuning
- Fully API compatible with prior generations of OSDs
- Uses a high-performance COTS platform

New 3-Tier data placement maps items being stored to the right type of storage device



Storage Workload

Stored On



ActiveStor Ultra with OSDv4



# PanFS – Ease of Management at Scale

Storage that “just works”

## Easy to deploy

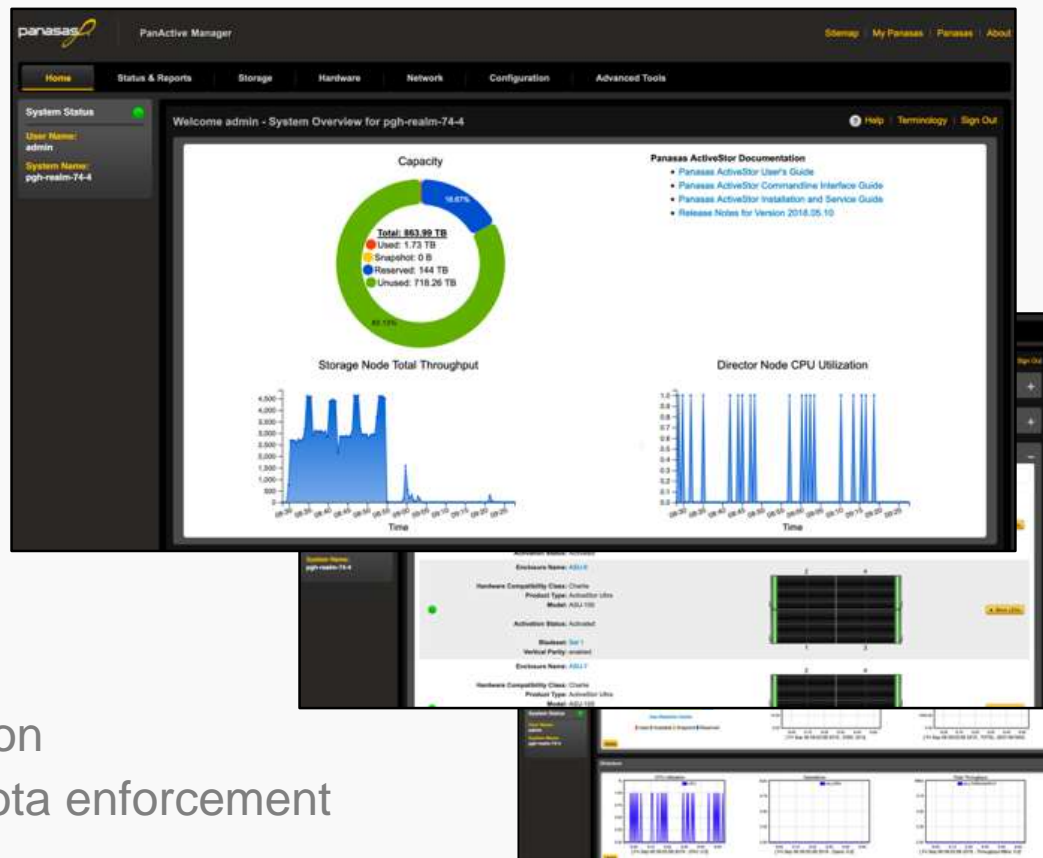
- Same day install for a PB-class system from dock to data
- Automatic recognition of new storage simplifies scaling out

## Easy to use

- Single GUI/CLI manages the entire cluster/namespaces
- Online capacity/load rebalancing

## Easy to manage at any scale

- Automated workflows such as new load balancing, reconstruction
- Enterprise data services such as reporting, snapshots, user quota enforcement



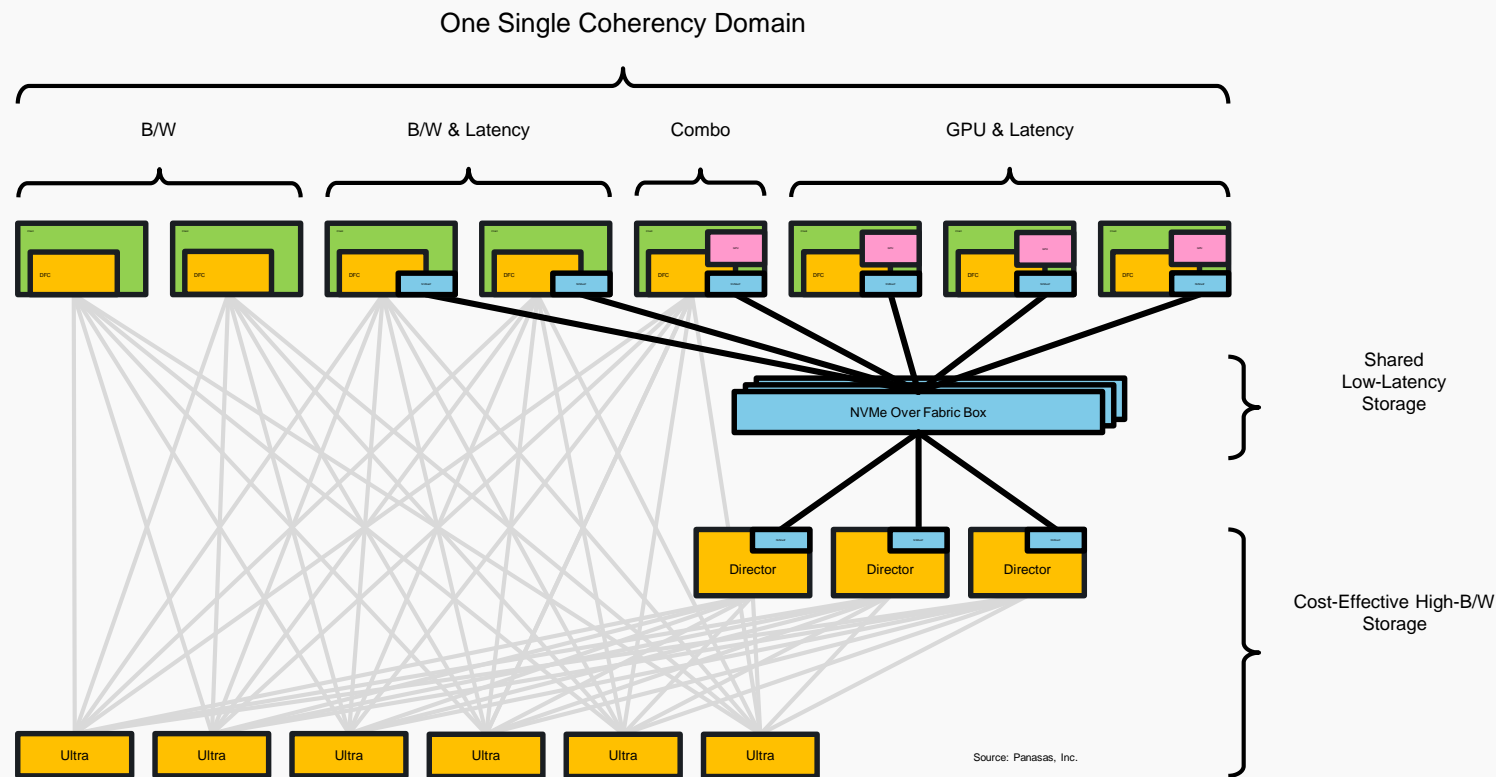
# Ongoing Developments

2  
7  
7  
0  
6  
7  
0  
3  
7  
2  
7  
3  
0  
6  
7  
0  
3  
7

# PanFS Ludicrous Mode Extreme Acceleration

Adding Support For Latency Sensitive Workloads

- Coherent NVMeoF low latency tier
- Integrated with Panasas environment
- Directors fill/drain NVMeoF boxes of their contents – freeing clients
- Use unmodified COTS NVMeoF hardware
- Hardware administration free



# ActiveStor Ultra **Delivers**

Limitless Performance Scaling and Management Simplicity

## **FAST**

4GB/s building blocks

## **FRUSTRATION FREE**

Simple admin

## **RELIABLE**

Always available

## **COST-EFFECTIVE**

*“Meeting our commitments to researchers requires extremely high computational power that is available 24/7. Panasas lives up to its promise of terrific performance with negligible maintenance and administration time.”*

– Dr. Warren Kaplan, Chief of Informatics, Garvan Institute of Medical Research



Thank you



[www.panasas.com](http://www.panasas.com)



# PanFS Application Considerations

“Out of the box” Performance for Mixed Workloads

---

## PanFS defaults typically fine for most applications

- Large Files striped across HDDs as RAID6+
- Small files stored on SSD as RAID 1
- Metadata on SSD or NVMe

## Default Volume creation is RAID6+ for resilience

- Optional RAID5+ volumes
- Users can dynamically assign RAID5 / 10 to their own directories or files within a volume
- Users can assign RAID stripe parameters (stripe unit, width, depth)

## Concurrent Write Mode

- Per-file or directory option
- Puts onus on client to manage write coherency – this is a responsibility!
- Performance “tweak” for e.g. write rarely / read often

# Storage Aware Applications

Optimising System Performance

---

## **Panasas SDK for programmatic control of storage layouts**

- ROMIO patches to integrate PanFS with MPI-IO
- e.g. HDF5 should be aligned to stripe boundaries

## **Huge numbers of small files can be inefficient for space and performance**

- Larger cache RAM, SSDs and NVMe metadata on Panasas systems
- OpenFOAM 5 / 17.12 introduce “collated” file format

## **Single large files can be inefficient**

- Cache coherency issues can stall writers
- Concurrent Write (CW) mode allows application ownership of write coherency