N - I - 9

Panasas:
High Performance Storage
for the Engineering Workflow

E. Jassaud, W. Szoecs

Panasas / transtec AG

panasas®

The leader in high performance parallel storage for business-critical applications

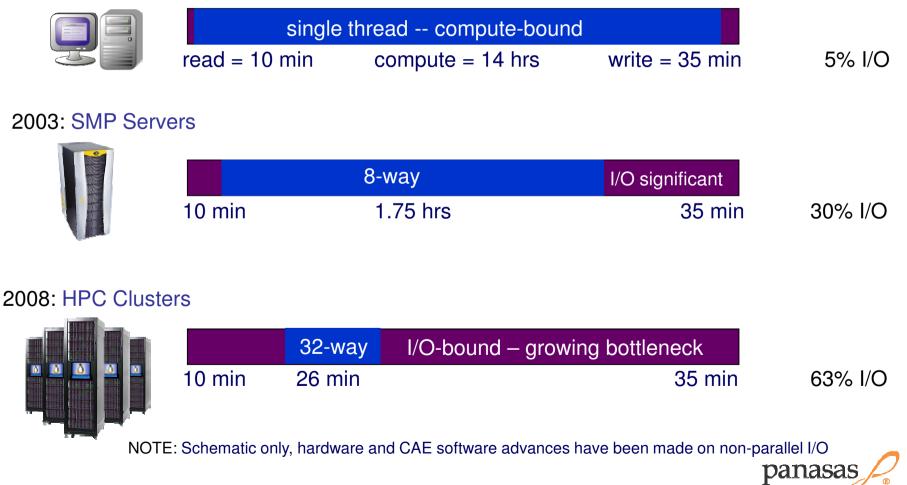**High-Performance Storage for the Engineering Workflow**
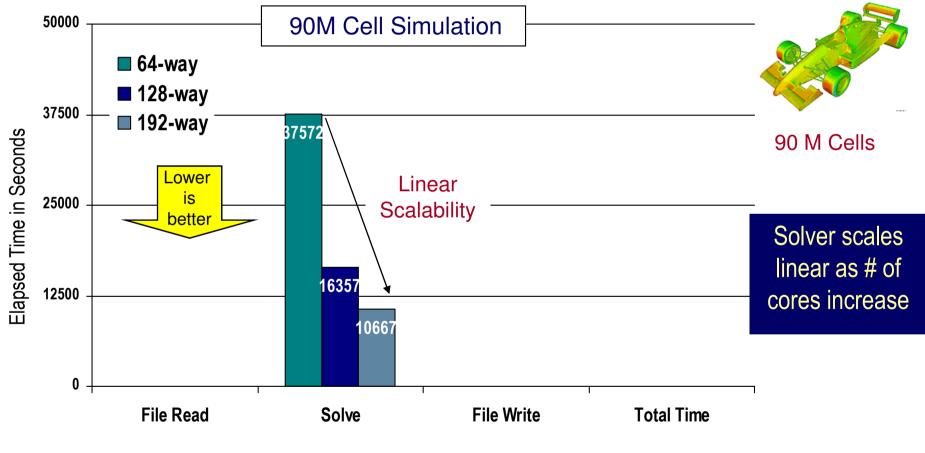
Terry Rush

Regional Manager, Northern Europe.

Why is storage performance important?

Example - Progression of a <u>Single</u> Job CAE Profile with non-parallel I/O
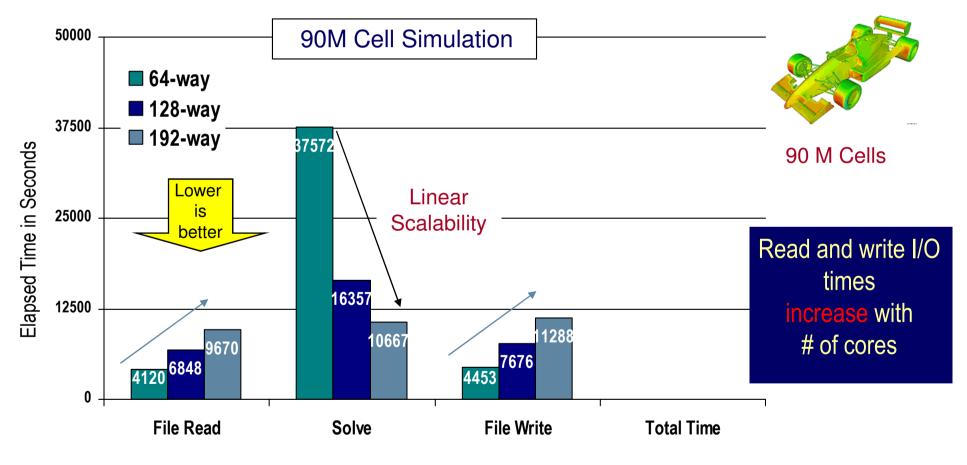
1998: Desktops

single thread -- compute-bound	
read = 10 min compute = 14 hrs write = 35 min	**5% I/O**

2003: SMP Servers

8-way	I/O significant	
10 min 1.75 hrs 35 min		**30% I/O**

2008: HPC Clusters

32-way	I/O-bound – growing bottleneck	
10 min 26 min 35 min		**63% I/O**

NOTE: Schematic only, hardware and CAE software advances have been made on non-parallel I/O

panasas

The I/O problem

90M Cell Simulation

Elapsed Time in Seconds

- ■ 64-way
- ■ 128-way
- ■ 192-way

Lower is better

37572

16357

Linear Scalability

10667

90 M Cells

Solver scales linear as # of cores increase

| File Read | Solve | File Write | Total Time |

Source: Barb Hutchings Presentation at SC07, Nov 2007, Reno, NV

panasas

Read & Write Time increases with Core Count

90M Cell Simulation

- 64-way
- 128-way
- 192-way

Elapsed Time in Seconds

Lower is better

Linear Scalability

90 M Cells

Read and write I/O times increase with # of cores

	File Read	Solve	File Write	Total Time
64-way	4120	37572	4453	
128-way	6848	16357	7676	
192-way	9670	10667	11288	

panasas

Bigger cluster does not mean faster jobs

90M Cell Simulation

- **64-way**
- **128-way**
- **192-way**

Elapsed Time in Seconds

Lower is better

Linear Scalability

90 M Cells

File Read
- 4120
- 6848
- 9670

Solve
- 37572
- 16357
- 10667

File Write
- 4453
- 7676
- 11288

Total Time
- 46145
- 30881
- 31625

Y-axis values: 0, 12500, 25000, 37500, 50000

Source: Barb Hutchings Presentation at SC07, Nov 2007, Reno, NV

panasas

Serial verses Parallel

Serial I/O Schematic

Parallel I/O Schematic

Head node

Storage
(Serial)

Compute
Nodes

MPI I/O

Storage
(Parallel)

Compute
Nodes

panasas

Parallel storage I/O returns scalability

90M Cell Simulation

64-way
128-way
192-way

90 M Cells

Lower is better

PanFS gains vs. serial I/O about 3x on 192 cores

Elapsed Time in Seconds

50000
37500
25000
12500
0

46145
30881 31625
38656
17442
11759
~2x
~3x

panasas

Two Measures of Performance

- Single wide-parallel job vs. multiple instances of less-parallel jobs
  - Often referenced in HPC industry as capability vs. capacity computing
  - Both important, but multi-user capacity computing more common in practice
    - Example: parameter studies and design optimization usually based on capacity computing since it would be economically impractical with capability approach
- File system performance characteristics for capability vs. capacity
  - Capability: Many cores writing to a large single file
  - Capacity: A few cores writing into a single file, but multiple instances
  - PanFS scales I/O for the capability job, and provides multi-level scalable I/O for the capacity jobs (each job parallel I/O -- multiple jobs writing to PanFS in parallel)
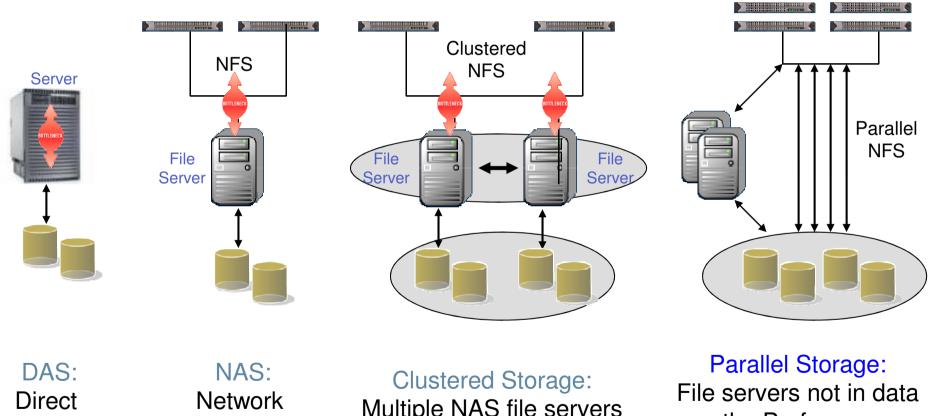 - NFS has single, serial data path for all capacity and capability solution writes

Capability Computing   Capacity Computing

Compute nodes

PanFS and storage   Single Large Job   Multiple Job Instances

panasas

What is Parallel Storage?

Parallel NAS: Next Generation of HPC File System

NFS

Clustered
NFS

Parallel
NFS

Server

File
Server

File
Server

File
Server

DAS:
Direct
Attached
Storage

NAS:
Network
Attached
Storage

Clustered Storage:
Multiple NAS file servers
managed as one

Parallel Storage:
File servers not in data
path. Performance
bottleneck eliminated.

panasas

Typical Complex Parallel Storage Configuration

Linux Clients

GE/10GE/IB
Network

10GE/IB

FC

OSS
Object Storage
Servers

Controllers

FC
SAN

OSS
Object Storage
Servers

Controllers

FC
SAN

MDS Metadata Server
(Active/Passive)

OSS
Object Storage
Servers

Controllers

FC
SAN

panasas

Panasas Solution

Linux Clients

GE/10GE/IB Network

Integrated solution
-Panasas File System
-Panasas HW and SW
-Panasas Service/Support
-Single management point

Simplified Manageability
-Easy and quick to install
-Single Fabric
-Single Web GUI/CLI
-Automatic capacity Balancing

Scalable Performance
-Scalable Metadata
-Scalable NFS/CIFS
-Scalable Reconstruction
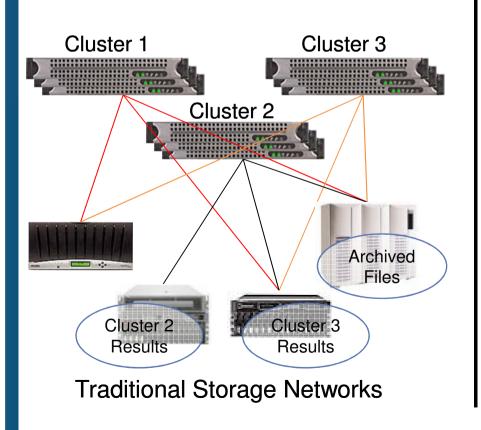-Scalable Clients
-Scalable BW

High Availability
-Metadata failover
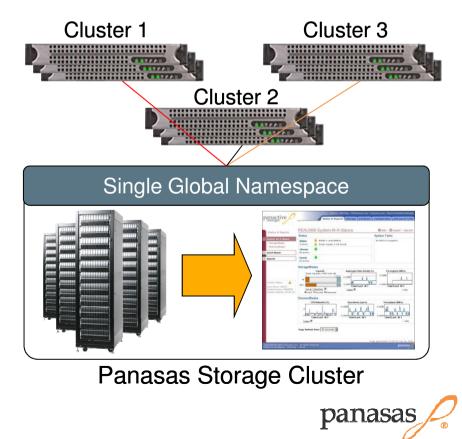-NFS failover
-Snapshots
-NDMP

panasas®

What else besides fast I/O?
Single Global Namespace for Easy Management
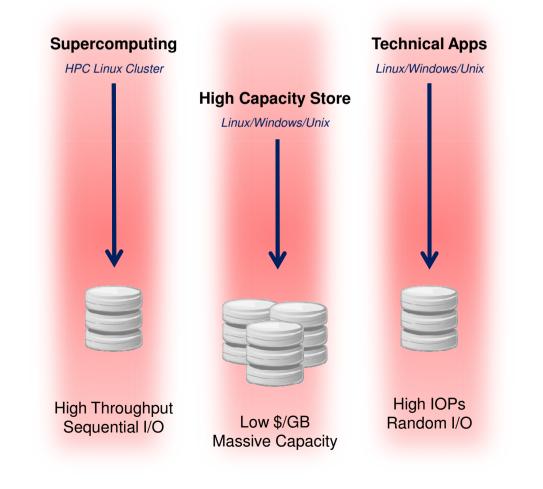
Remove artificial, physical and logical boundaries

- Eliminates need to maintain mount scripts or move data

Cluster 1   Cluster 3

Cluster 2

Archived
Files

Cluster 2
Results

Cluster 3
Results

Traditional Storage Networks

Cluster 1   Cluster 3

Cluster 2

Single Global Namespace

Panasas Storage Cluster
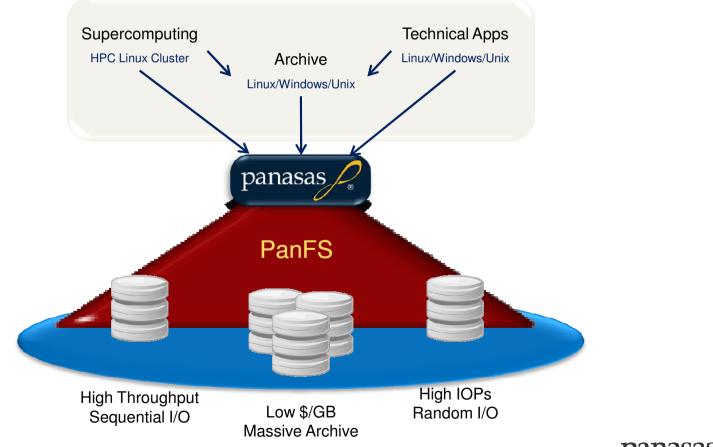
panasas

Removing Silos of Storage

- Historic technical computing storage problem
  - Multiple applications and users across the workflow creates costly and inefficient storage silos by application and by vendor
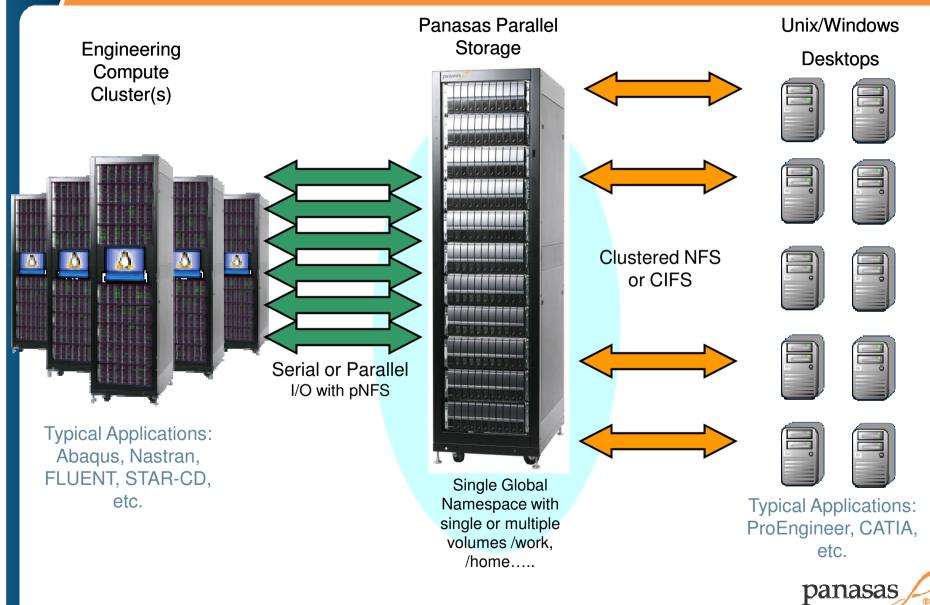
Supercomputing

HPC Linux Cluster

High Capacity Store

Linux/Windows/Unix

Technical Apps

Linux/Windows/Unix

High Throughput
Sequential I/O

Low $/GB
Massive Capacity

High IOPs
Random I/O

panasas

Removing Silos of Storage – single global file system

- ## High performance scale-out NAS solutions
  - PanFS Storage Operating System delivers exceptional performance, scalability & manageability
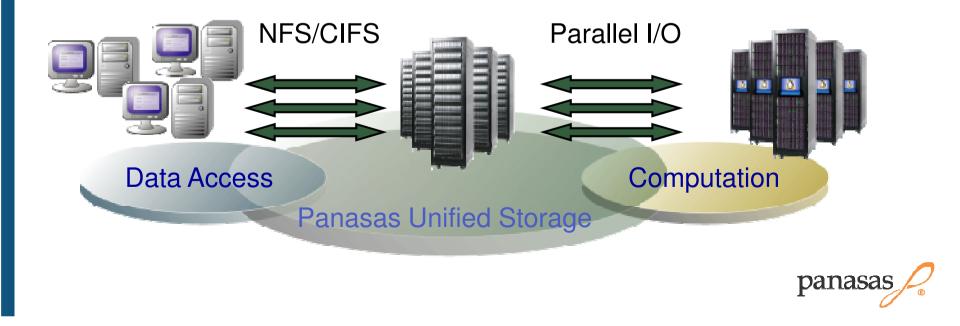
The Technical Enterprise

Supercomputing

HPC Linux Cluster

Archive

Linux/Windows/Unix

Technical Apps

Linux/Windows/Unix

panasas

PanFS

High Throughput
Sequential I/O

Low $/GB
Massive Archive

High IOPs
Random I/O

panasas

Typical Panasas Parallel Storage Workflow:

Engineering
Compute
Cluster(s)

Panasas Parallel
Storage

Unix/Windows

Desktops

Serial or Parallel
I/O with pNFS

Clustered NFS
or CIFS

Typical Applications:
Abaqus, Nastran,
FLUENT, STAR-CD,
etc.

Single Global
Namespace with
single or multiple
volumes /work,
/home…..

Typical Applications:
ProEngineer, CATIA,
etc.

panasas ®

Unified Parallel Storage for the Engineering Workflow

- Panasas Unified Parallel Storage for Leverage of HPC investments
  - <u>Performance</u> that meets growing I/O demands
  - <u>Management</u> simplicity and configuration flexibility for all workloads
  - <u>Scalability</u> for fast and easy non-disruptive growth
  - <u>Reliability</u> and data integrity – for long term storage

NFS/CIFS   Parallel I/O

Data Access   Computation

Panasas Unified Storage

panasas ®

Panasas Overview

- Founded April 1999 by Prof. Garth Gibson, CTO

- Locations
  - US: HQ in Fremont, CA
            Development center in Pittsburgh, PA
  - EMEA: Sales offices in UK, France & Germany
  - APAC: Sales offices in China & Japan

- First Customer Shipments October 2003, now serving over 400 customer sites with over 500,000 server clients in twenty-seven countries

- 13 patents issued, others pending

- Key Investors

- Speed with Simplicity and Reliability - Parallel File System

  and Parallel Storage Appliance Solutions

Broad Industry and Business-Critical Application Support

Energy

Seismic Processing
Reservoir Simulation
Interpretation

Government

Imaging & Search
Stockpile Stewardship
Weather Forecasting

Universities

High Energy Physics
Molecular Dynamics
Quantum Mechanics

Aerospace

Fluid Dynamics (CFD)
Structural Mechanics
Finite Element Analysis

Finance

Credit Analysis
Risk Analysis
Portfolio Optimization

Industrial Mfg

EDA Simulation
Optical Correction
Thermal Mechanics
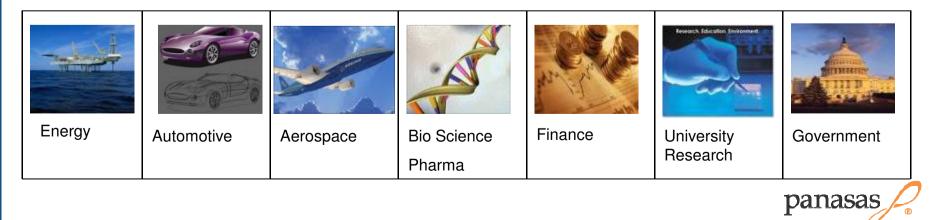
Automotive

Crash Analysis
Fluid Dynamics
Acoustic Analysis

Bio/Pharma

BioInformatics
Computation Chemistry
Molecular Modeling

panasas

Customer Success Highlights

- Almost half the Formula One teams use Panasas

- Five of the top six Oil & Gas companies in the world use Panasas

- Every Intel chip since 2006 has been designed using Panasas

- The world's first Petascale system, RoadRunner, uses Panasas

- The world's three largest genomic data centres use Panasas

- The largest aircraft manufacturer in the world uses Panasas

- Leading Universities including Cambridge, Oxford, Stanford & Yale use Panasas

- The world's largest Hedge Fund uses Panasas for risk analysis

Energy	Automotive	Aerospace	Bio Science Pharma	Finance	University Research	Government

panasas

Some Panasas Customers

Panasas Differentiation

- Leader in high performance parallel storage for business-critical applications
  - Optimized for demanding storage applications in the energy, government, finance, manufacturing, bioscience and core research sectors

- Panasas® PanFS™
  - Parallel file system with integrated RAID protection
  - Most scalable performance in the industry
  - Appliance-like scalable hardware architecture for large-scale storage deployments
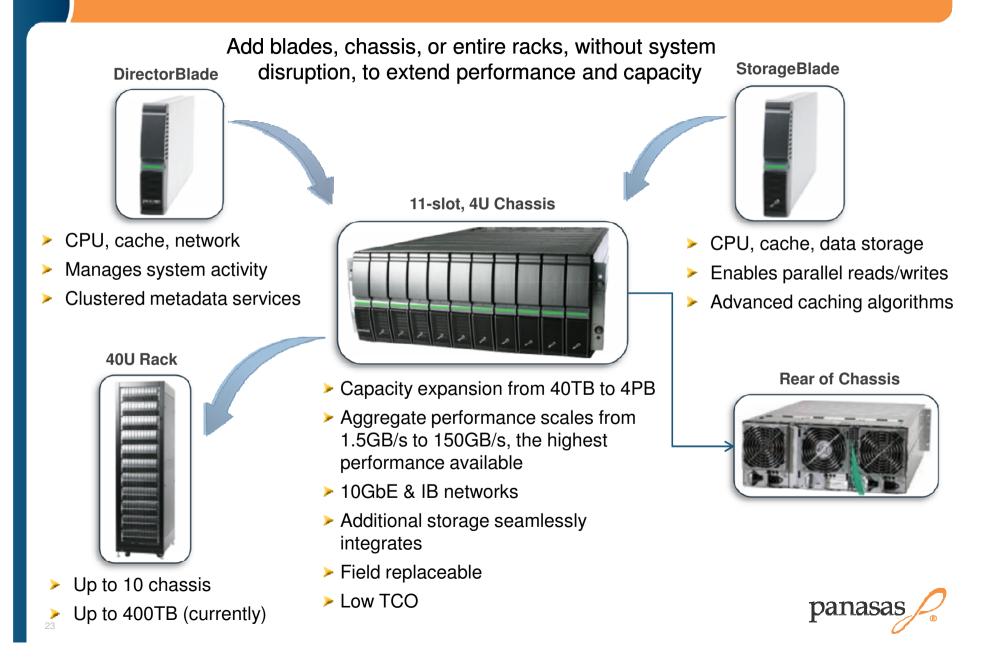 - Highly integrated management suite

panasas

Integrated RAID Reliability

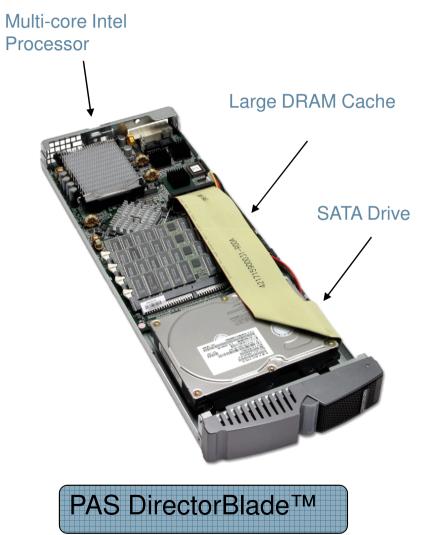| RAID Implementation	Benefit
Object RAID	✓ System intelligently assigns RAID levels based on file size ✓ Automatic transitioning from RAID 1 to RAID 5 without re-striping
Per File RAID	✓ High performance file reconstruction (vs. drive sector reconstruction) ✓ Rebuild in hours ✓ Parallel rebuild – all blade sets participate in RAID rebuilds
Horizontal (Blade) and Vertical (Disk) Parity RAID	✓ RAID within the individual drive as well as across drives ✓ Improves internal ECC capabilities ✓ Predicatively solves media errors ✓ Significantly lowers drive failure probability

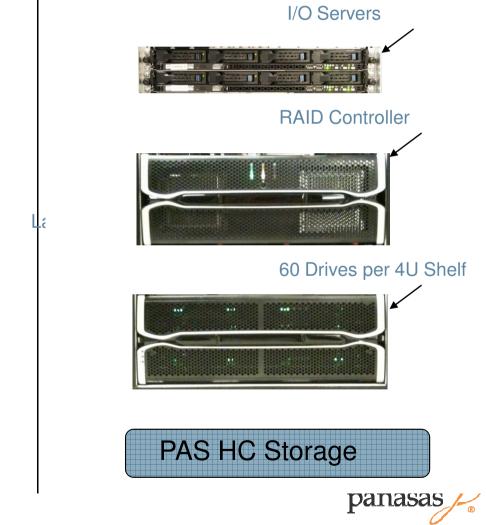∴ Only Panasas includes RAID data protection as a component of its file system

panasas

Appliance-Like, Modular Hardware Design

Add blades, chassis, or entire racks, without system disruption, to extend performance and capacity

DirectorBlade

StorageBlade

11-slot, 4U Chassis

- CPU, cache, network
- Manages system activity
- Clustered metadata services

- CPU, cache, data storage
- Enables parallel reads/writes
- Advanced caching algorithms

40U Rack

Rear of Chassis

- Capacity expansion from 40TB to 4PB
- Aggregate performance scales from 1.5GB/s to 150GB/s, the highest performance available
- 10GbE & IB networks
- Additional storage seamlessly integrates
- Field replaceable
- Low TCO

- Up to 10 chassis
- Up to 400TB (currently)

23

panasas

Panasas Hardware Technology

Multi-core Intel
Processor

Large DRAM Cache

SATA Drive

I/O Servers

RAID Controller

60 Drives per 4U Shelf

PAS DirectorBlade™

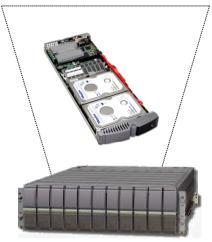**PAS HC Storage**

panasas

Panasas Product Family

Same Panasas software, management interface, single name space

PAS HC

(Per 42U Rack)

- 960TBs Raw Capacity
- 5GB/sec Aggregate B/W
- RAID 6
- Unlimited Scalability
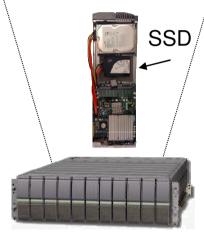- Entry Level $/GB 500TBs+

PAS 7

(Per 4U Shelf)

- 40TBs Raw Capacity
- 350MB/sec Aggregate B/W
- RAID 1, 5 & 10
- Unlimited Scalability
- Entry Level $/GB Sub 500TB

PAS 8

(Per 4U Shelf)

- 40TBs Raw Capacity
- 600MB/sec Aggregate B/W
- RAID 1, 5 & 10
- HA & Snapshots
- Unlimited Scalability

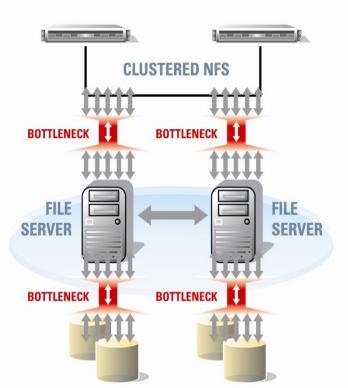## PAS 9

(Per 4U Shelf)

- 600MB/sec Aggregate B/W
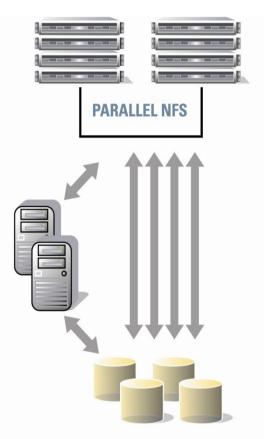- Unlimited Scalability
- RAID 1, 5 & 10
- HA & Snapshots
- 21,000 IOPS

SSD

NAS
Network Attached Storage.
Serial I/O bottleneck.

CLUSTERED STORAGE
Multiple NAS file servers
managed as one.
Multiple serial I/O bottlenecks.

PARALLEL STORAGE
File server not in data path.
Performance bottleneck eliminated.
Fastest and easiest to manage.