



Storage Technologies for SDM

Dr. Oliver Tennert, 02.07.2013

Director Technology Management & HPC Solutions transtec AG

Overview

- || Who is transtec?
 - || Storage Demands in HPC → parallel NFS
 - || HPC Storage Solutions from transtec
-

High Performance Computing at transtec

More than 30 Years of Experience in Scientific Computing

1980: transtec founded, a reseller for DEC equipment

1987: SUN/SPARC workstations in portfolio

1991: IBM RS/6000 systems in portfolio

2000: „Kepler Cluster“ in Tübingen installed (#215 in **TOP500**)

2003: transtec HPC cluster at **RRZ Erlangen** no. 317 in **TOP 500**

2005: transtec is **key supplier** for **CERN** in Geneva

2007: transtec HPC cluster at **KIT Karlsruhe** no. 104 in **TOP 500**

2010: transtec focusses on HPC as a **strategic business unit**

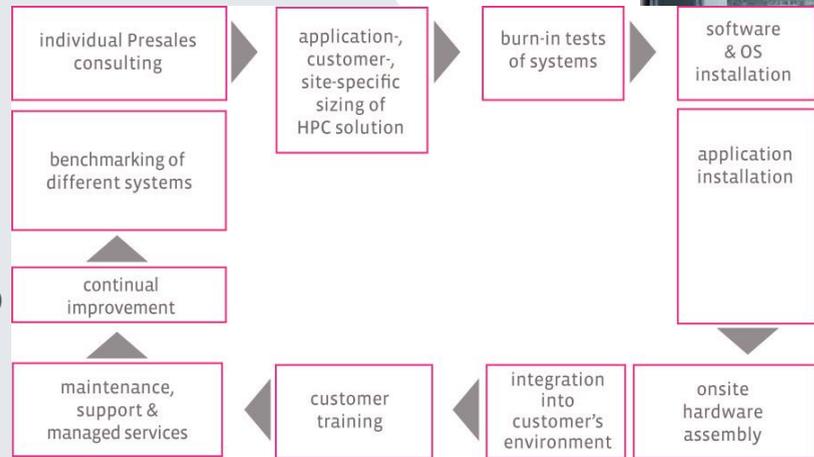
up to now: around **500 HPC installations** in Europe

High Performance Computing at transtec

Customer Care From A to Z

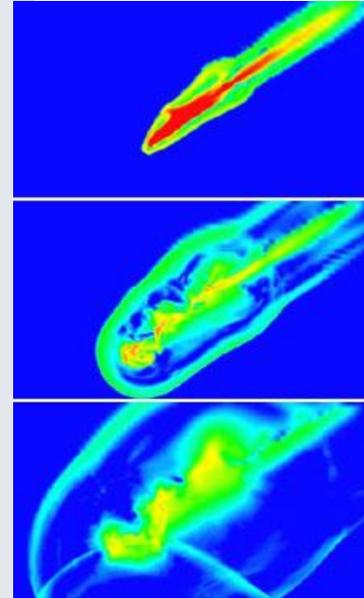
comprehensive solutions for High Performance Computing (HPC)

- **turnkey**
- caring for the customer during the whole **solution lifecycle**:
 - from **individual consulting** to **managed services**
 - customers have access to **transtec benchmark center**



Challenges With Simulation Data Management

- || **Creation and storage of large data in a highly-performing way**
- || Access to result data for post-processing

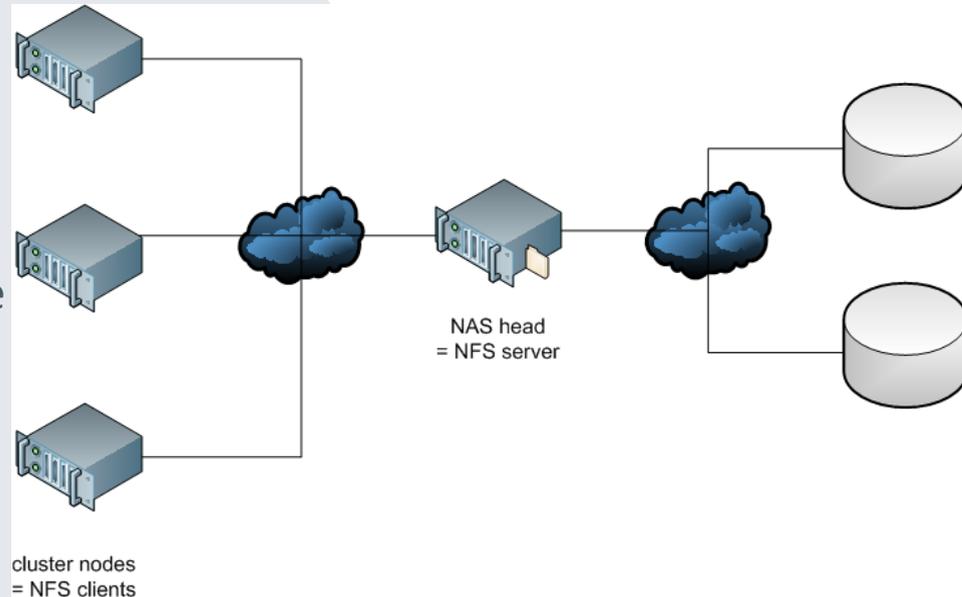


Storage Demands in HPC

- || need for **computing power**
 - due to need to run larger and more accurate models
 - more CPUs, more cores, more nodes, more RAM
 - || need for **network performance**
 - more highly parallellized jobs
 - high-speed interconnects (10GbE, InfiniBand,...)
 - **massive explosion of data sets**
 - demand for
 - **large storage capacity**
 - **high bandwidth**
 - **low latency**
-

Deficiencies of Today's Solutions

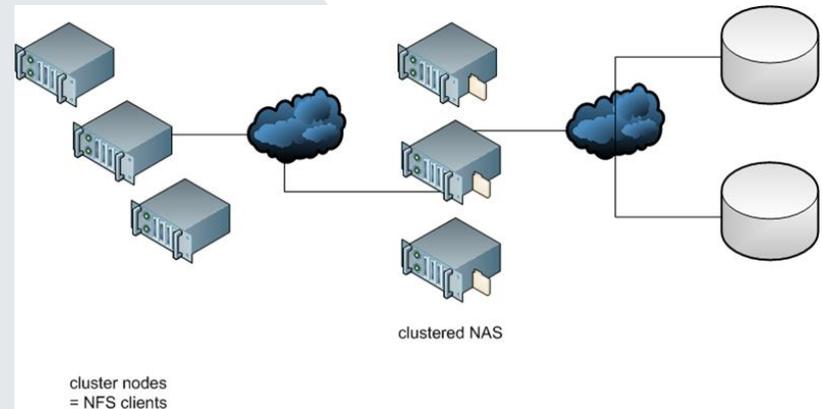
- II most widespread solution:
single NFS server
 - does not scale: NFS head is bottleneck
 - „high-speed“ NFS server will be bottleneck by tomorrow



Deficiencies of Today's Solutions

II „clustered NFS“: problematic

- either head-to-head synchronization limits scalability
- or manual partitioning of global namespace is cumbersome
- NFS is not suitable for dynamical load balancing (inherent state)



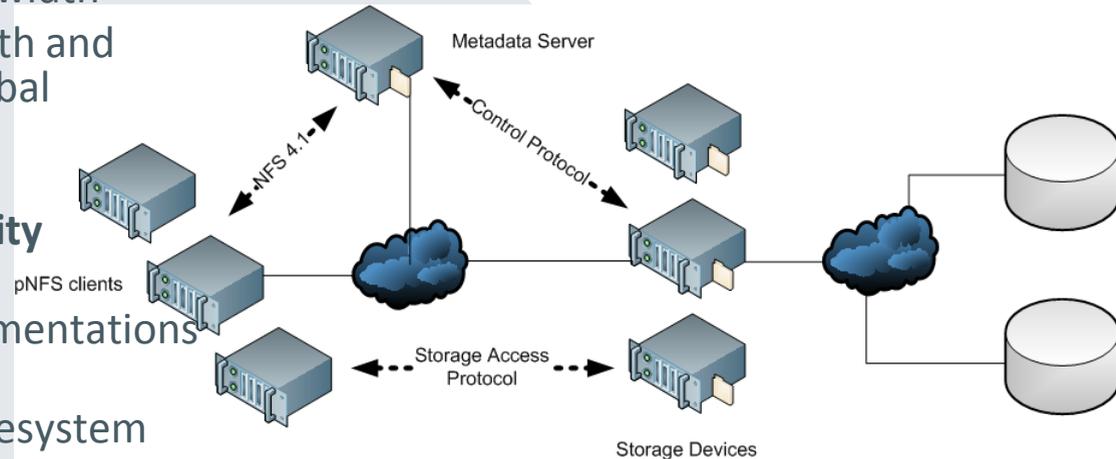
NFS 4.1 and Parallel NFS (pNFS)

- II **NFS 4.1:** idea to use SAN FS architecture for NFS originally from Gary Grider (LANL) and Lee Ward (Sandia)
 - II RFC 5661 in 2010
 - II major changes to NFS 4:
 - sessions
 - directory delegations
 - pNFS (optional feature)
 - referrals → federated filesystems
 - II **future standard NFS 4.2** (<http://tools.ietf.org/html/draft-ietf-nfsv4-minorversion2-19>)
 - server-side copy
 - application I/O advise
 - sparse files
 - space reservations
 - application data-hole support
 - labelled NFS
-

Parallel NFS (pNFS)

Generic Architecture 1

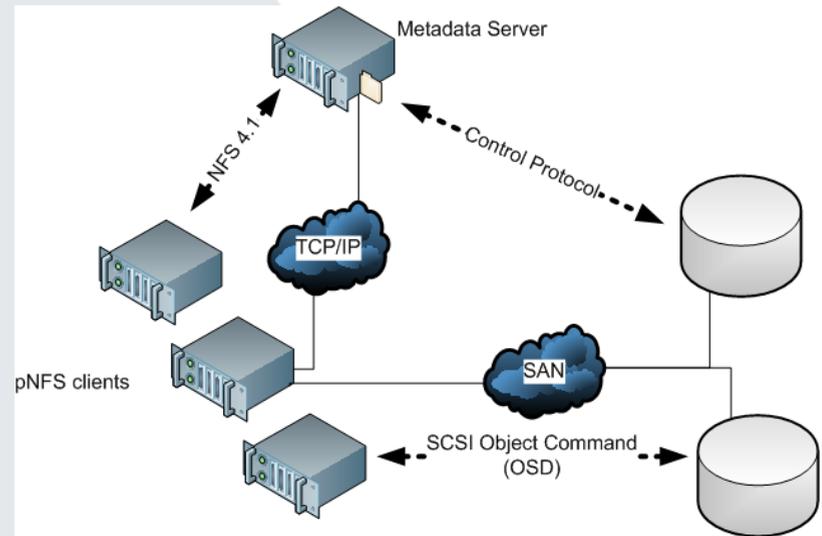
- II **scalable** capacity and bandwidth
- II **separation of metadata path and data path (out-of-band global namespace)**
- II built for **interoperability** and **backwards-compatibility**
- II **flexible** design allows for different storage implementations (**layouts**)
- II **global namespace** eases filesystem management and job flow



pNFS: Object Layout

Generic Architecture 2

- II Panasas' contribution, based on **NASD** design (Network-Attached Secure Disk) developed at Carnegie Mellon University, later evolved into forthcoming **SCSI OSD** standard (**object-based storage device**)
- II layout uses **SCSI object command** set
- II **space management** built into devices
- II designed for **secure access** and **high-performance data replication**
- II cryptographically secured credentials ("**capabilities**") needed to access storage devices

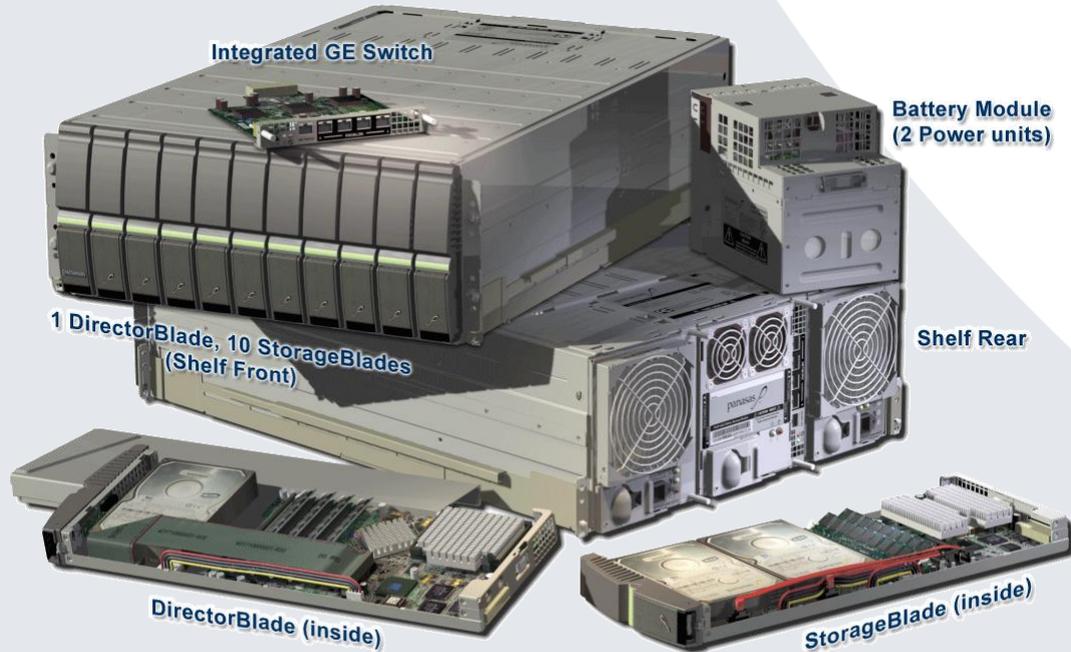


Panasas High Performance Storage from transtec



- || Blade-based parallel NFS appliance
 - || 11 blades per shelf = approx. 60 TB in 4 U
 - || Approx. 1.5 Gbyte/sec aggregate bandwidth per shelf
-

Panasas High Performance Storage from transtec

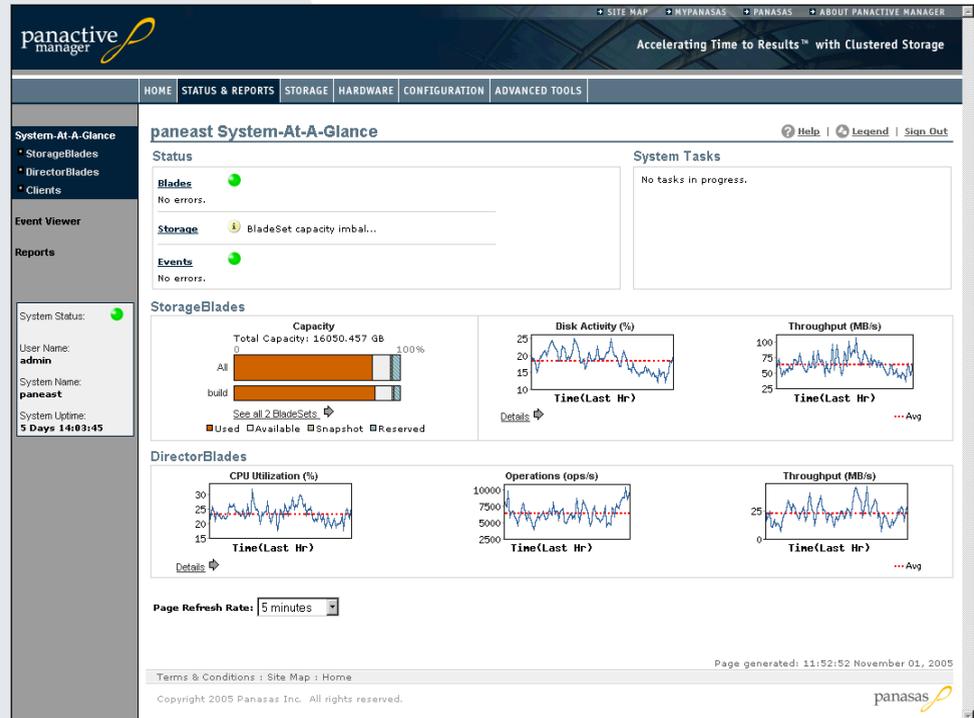


Panasas ActiveStor Specs

	ActiveStor 11	ActiveStor 12	ActiveStor 14	ActiveStor 14T
Product Focus	Balanced Capacity & Performance	Highest Performance	Price/Performance; high capacity and throughput	Highest performance for mixed workloads (large+small files)
Read Throughput (MB/sec)	1,150	1,500	1,500	1,500
Write Throughput (MB/sec)	950	1,600	1,600	1,600
File Creates/Sec. per Director Blade (Metadata Performance)	4,260	6,250	13,550	14,150
Capacity (TB)	40/60	40 / 60	80 PB	45 PB
Cache (GB)	40 + 8	80 + 12	92	72
High Availability Network Failover	Optional	Standard	Standard	Standard
Link Aggregation	No	Yes	Yes	Yes

Panactive Manager

- II Single Point of Management
- II Simple out-of-box experience
- II Seamlessly deploy new blades
- II Capacity & load balancing
- II Snapshots
- II 1-touch reporting capabilities
- II Scriptable CLI



Easy to Manage

- II Scriptable CLI
- II Easy web-based or CLI-based setup

```
0187ee222f0012# pancli
pancli -- Initial network setup

You may now configure the IP address, netmask, and default route
address of this system.

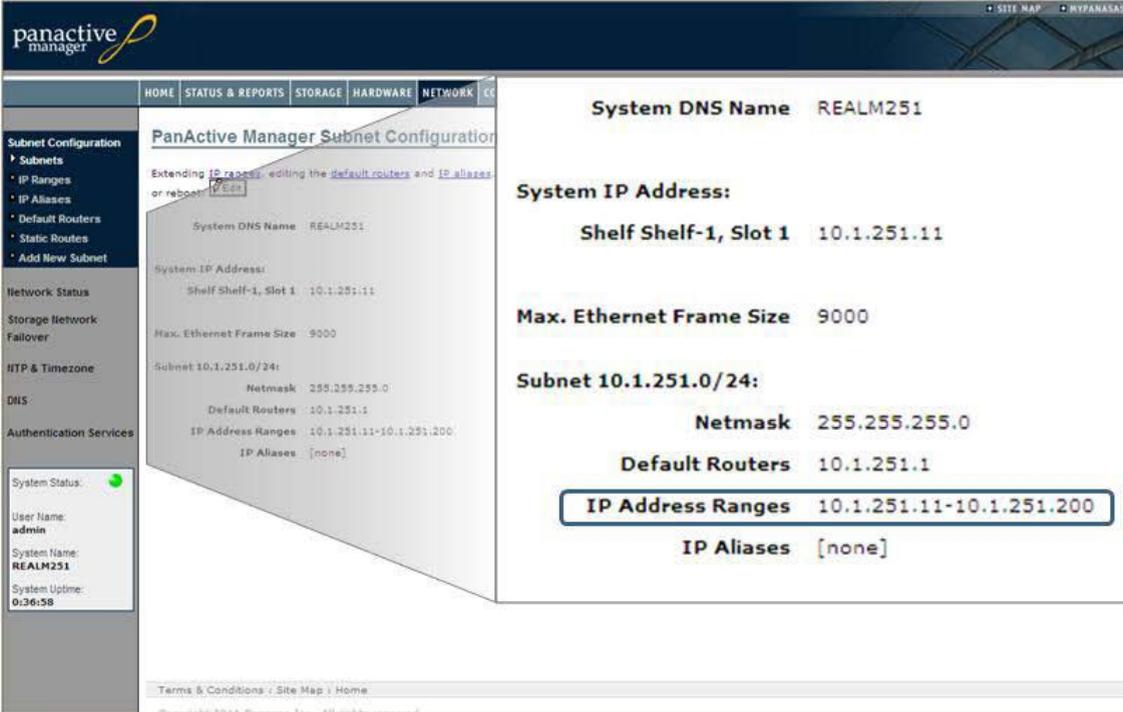
Would you like to setup the network now? (recommended)
[yes]

System IP Address: 10.1.251.11
Netmask: [255.255.255.0]
Default Router (type "none" to leave blank): [10.1.251.1]

Validating entries. Please wait...

Settings chosen:
System IP Address: 10.1.251.11
Netmask: 255.255.255.0
Default Router: 10.1.251.1

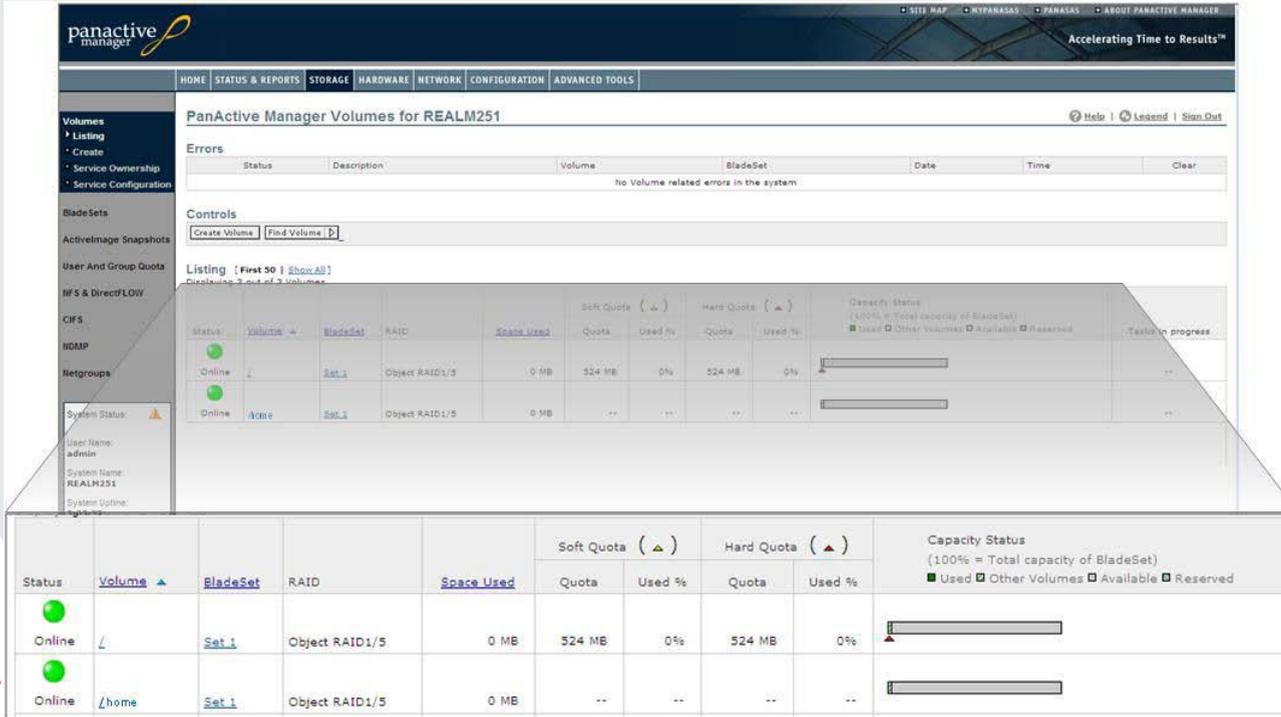
Save settings? [yes]
```



The screenshot displays the PanActive Manager web interface for Subnet Configuration. The interface includes a navigation menu on the left with options like Subnets, IP Ranges, IP Aliases, Default Routers, Static Routes, and Add New Subnet. The main content area shows configuration details for a subnet, including System DNS Name (REALM251), System IP Address (Shelf Shelf-1, Slot 1: 10.1.251.11), Max. Ethernet Frame Size (9000), Subnet (10.1.251.0/24), Netmask (255.255.255.0), Default Routers (10.1.251.1), IP Address Ranges (10.1.251.11-10.1.251.200), and IP Aliases ([none]). A summary table on the right lists these settings in a structured format.

System DNS Name	REALM251
System IP Address:	
Shelf Shelf-1, Slot 1	10.1.251.11
Max. Ethernet Frame Size	9000
Subnet 10.1.251.0/24:	
Netmask	255.255.255.0
Default Routers	10.1.251.1
IP Address Ranges	10.1.251.11-10.1.251.200
IP Aliases	[none]

Extensive Management and Monitoring Capabilities



The screenshot displays the PanActive Manager interface for REALM251. The top navigation bar includes links for HOME, STATUS & REPORTS, STORAGE, HARDWARE, NETWORK, CONFIGURATION, and ADVANCED TOOLS. The left sidebar lists various management categories such as Volumes, BladeSets, and User And Group Quota. The main content area shows the 'PanActive Manager Volumes for REALM251' page, which includes an 'Errors' section (reporting no errors), a 'Controls' section with 'Create Volume' and 'Find Volume' buttons, and a 'Listing' table.

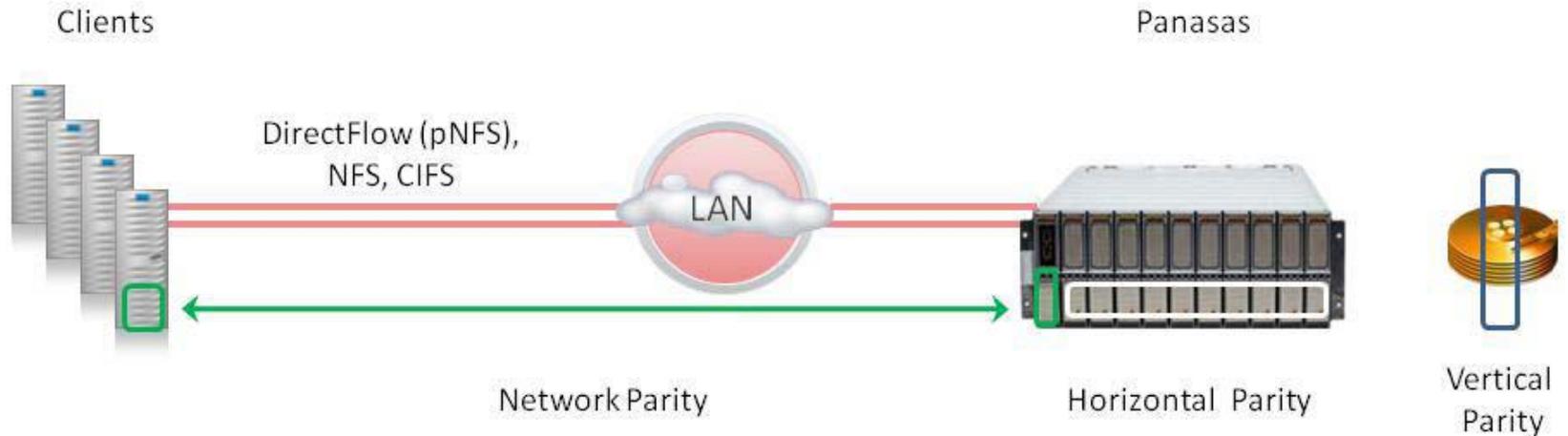
Listing (First 30 | Show All)
Displaying 3 out of 3 volumes

Status	Volume	BladeSet	RAID	Space Used	Soft Quota (▲)		Hard Quota (▲)		Capacity Status (100% = Total capacity of BladeSet) ■ Used ■ Other Volumes □ Available □ Reserved
					Quota	Used %	Quota	Used %	
Online	/	Set_1	Object RAID1/5	0 MB	524 MB	0%	524 MB	0%	
Online	/home	Set_1	Object RAID1/5	0 MB	

System Status: 
 User Name: admin
 System Name: REALM251
 System Uptime: 19 days

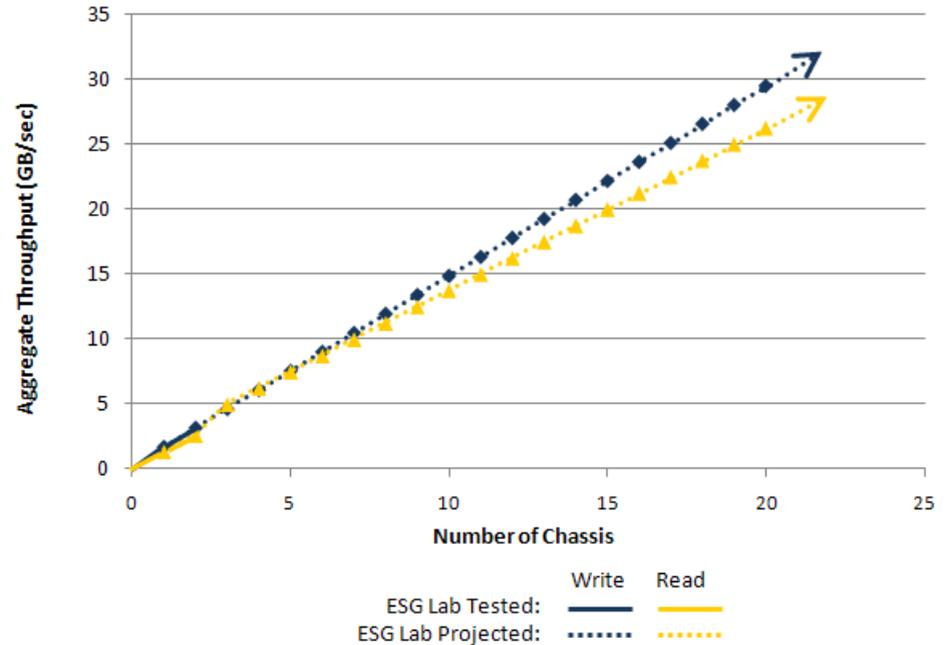
Tiered-Parity RAID

- || 3-level parity: network, horizontal, and vertical



Scalable Performance

II ESG Labs Performance Tests with ActiveStor14 chassis





transtec

accelerate productivity
