Storage Technologies for SDM

Dr. Oliver Tennert, 02.07.2013
Director Technology Management & HPC Solutions transtec AG

transtec

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

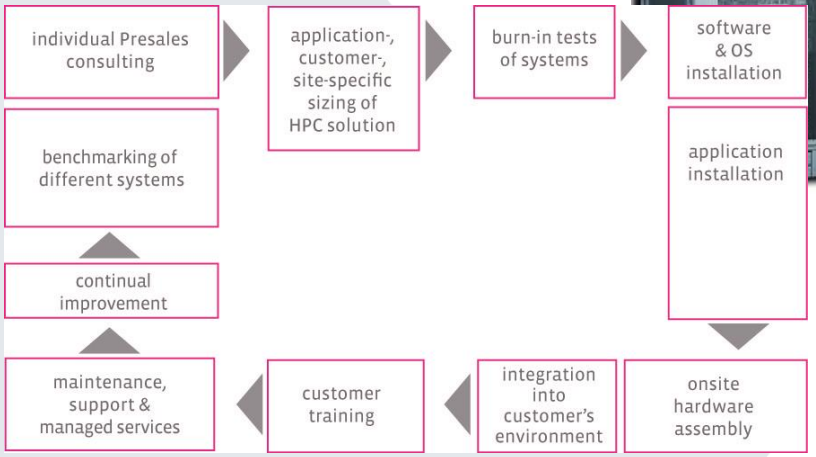
1980: transtec founded, a reseller for DEC equipment
1987: SUN/SPARC workstations in portfolio
1991: IBM RS/6000 systems in portfolio
2000: „Kepler Cluster“ in Tübingen installed (#215 in **TOP500**)
2003: transtec HPC cluster at **RRZ Erlangen** no. 317 in **TOP 500**
2005: transtec is **key supplier** 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

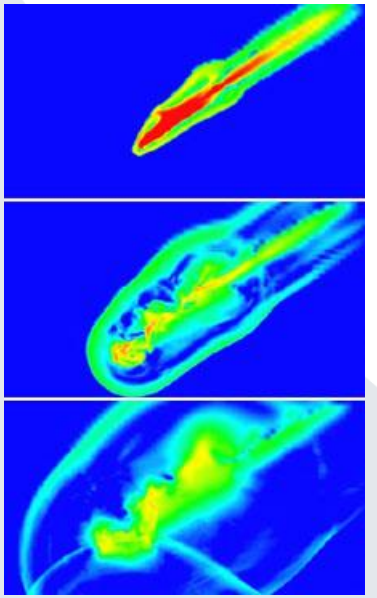
comprehensive solutions for High Performance Computing (HPC)

- **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



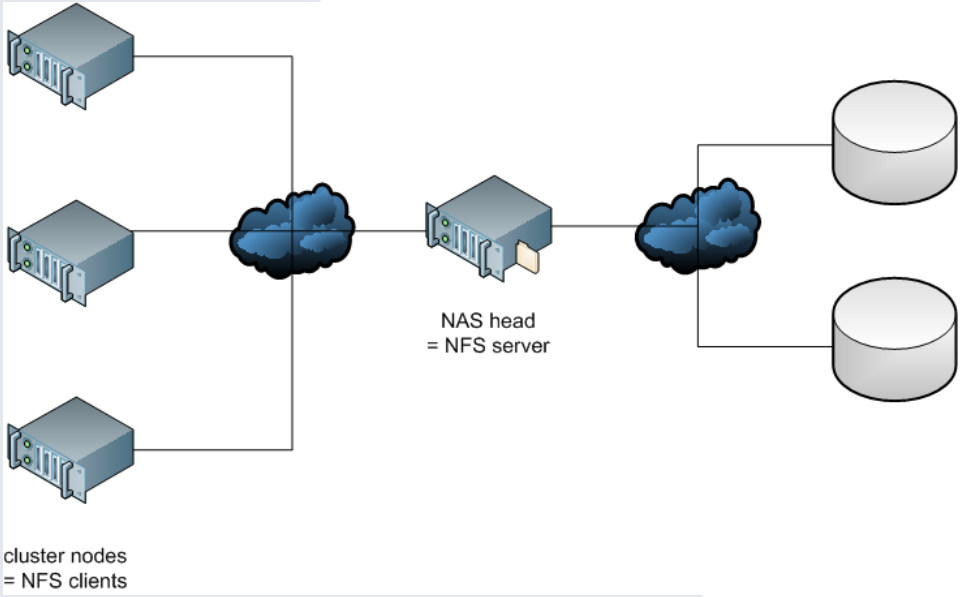
Storage Demands in HPC

- 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 paralellized jobs
  - high-speed interconnects (10GbE, InfiniBand,...)

→ **massive explosion of data sets**
→ demand for
  - **large storage capacity**
  - **high bandwidth**
  - **low latency**

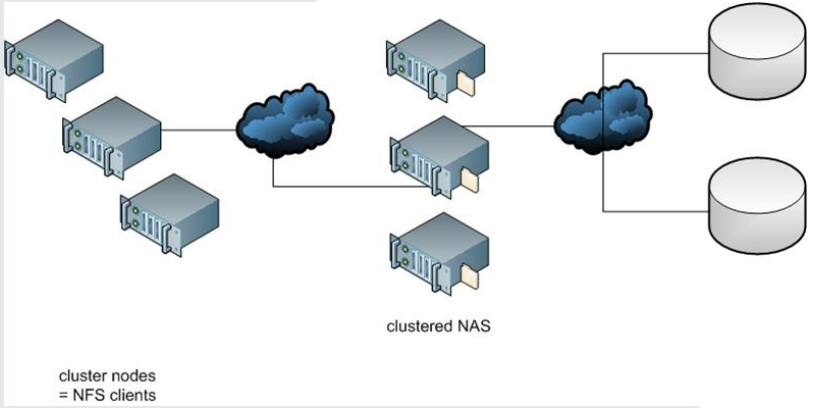
Deficiencies of Today‘s Solutions

- 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

- **„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)



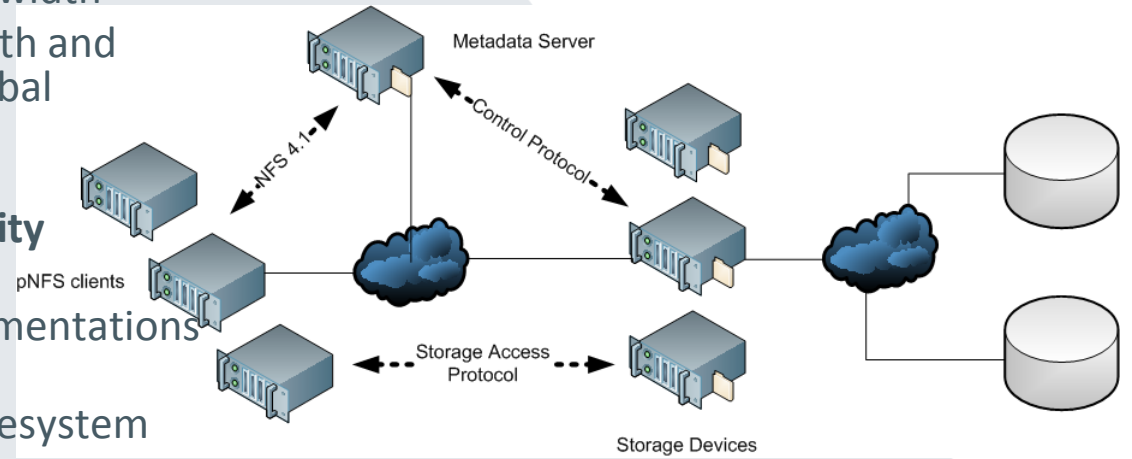
NFS 4.1 and Parallel NFS (pNFS)

- **NFS 4.1:** idea to use SAN FS architecture for NFS originally from Gary Grider (LANL) and Lee Ward (Sandia)
- RFC 5661 in 2010
- major changes to NFS 4:
  - sessions
  - directory delegations
  - pNFS (optional feature)
  - referrals → federated filesystems
- **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

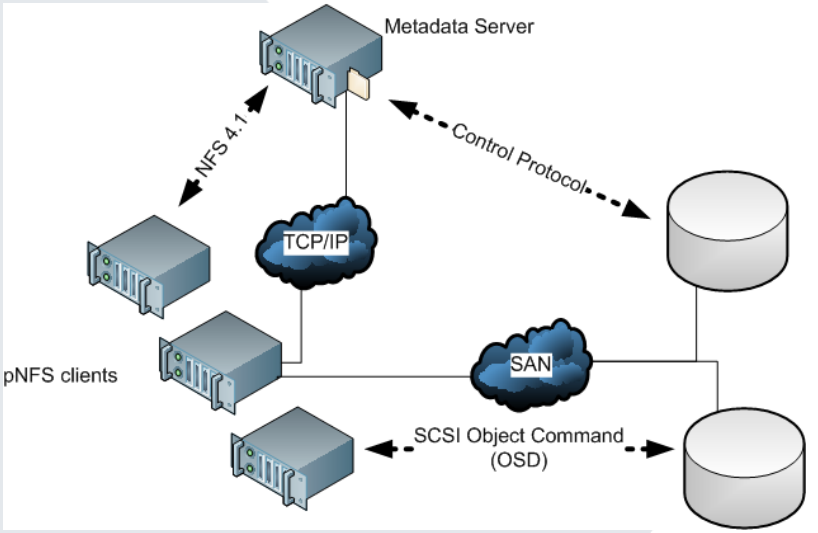
- **scalable** capacity and bandwidth
- **separation** of **metadata** path and **data** path (**out-of-band** global namespace)
- built for **interoperability** and **backwards-compatibility**
- **flexible** design allows for different storage implementations (**layouts**)
- **global namespace** eases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



pNFS: Object Layout

Generic Architecture 2

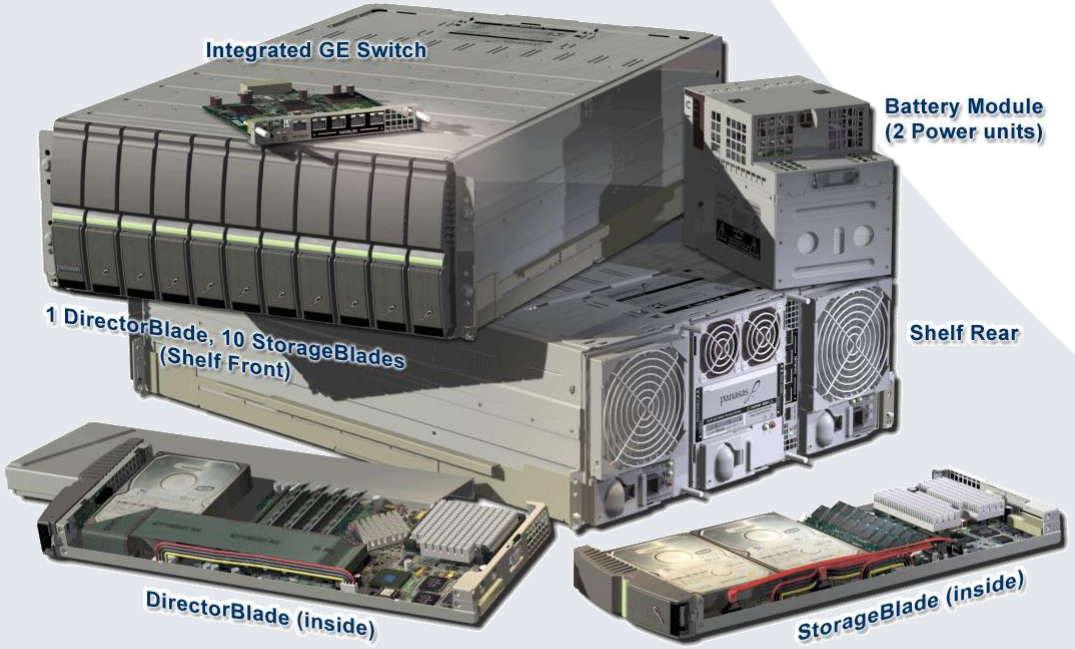
- Panasas' contribution, based on **NASD** design (Network-Attached Secure Disk) developed at Carnegie Mellon University, later evolved into forthcoming **SCSI OSD** standard (**object-based storage device**)
- layout uses **SCSI object command** set
- **space management** built into devices
- designed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

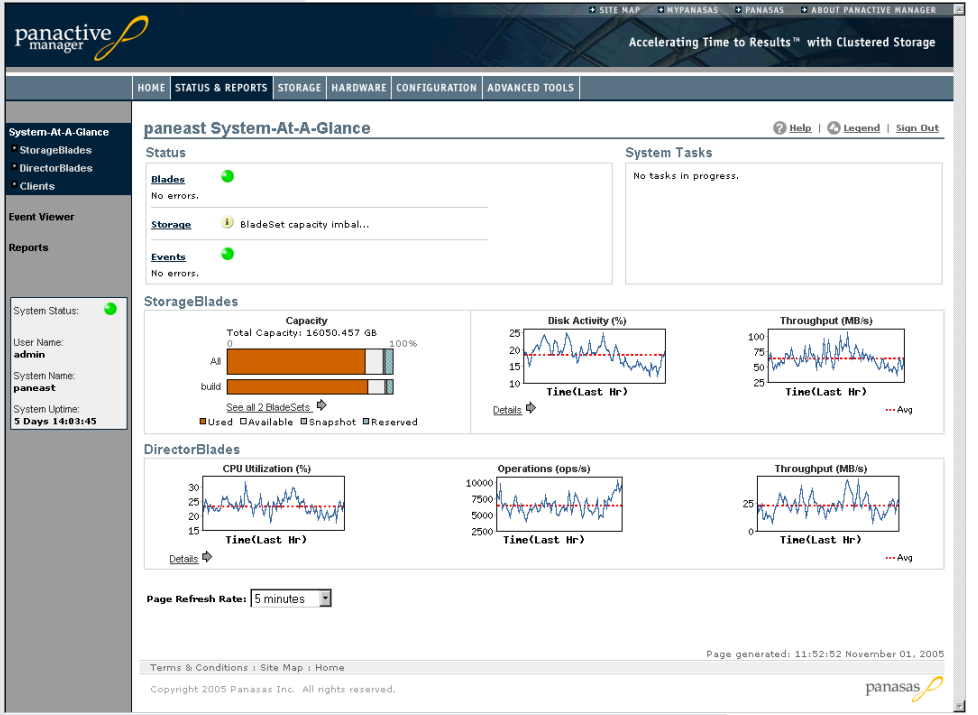


Panasas ActiveStor Specs

	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

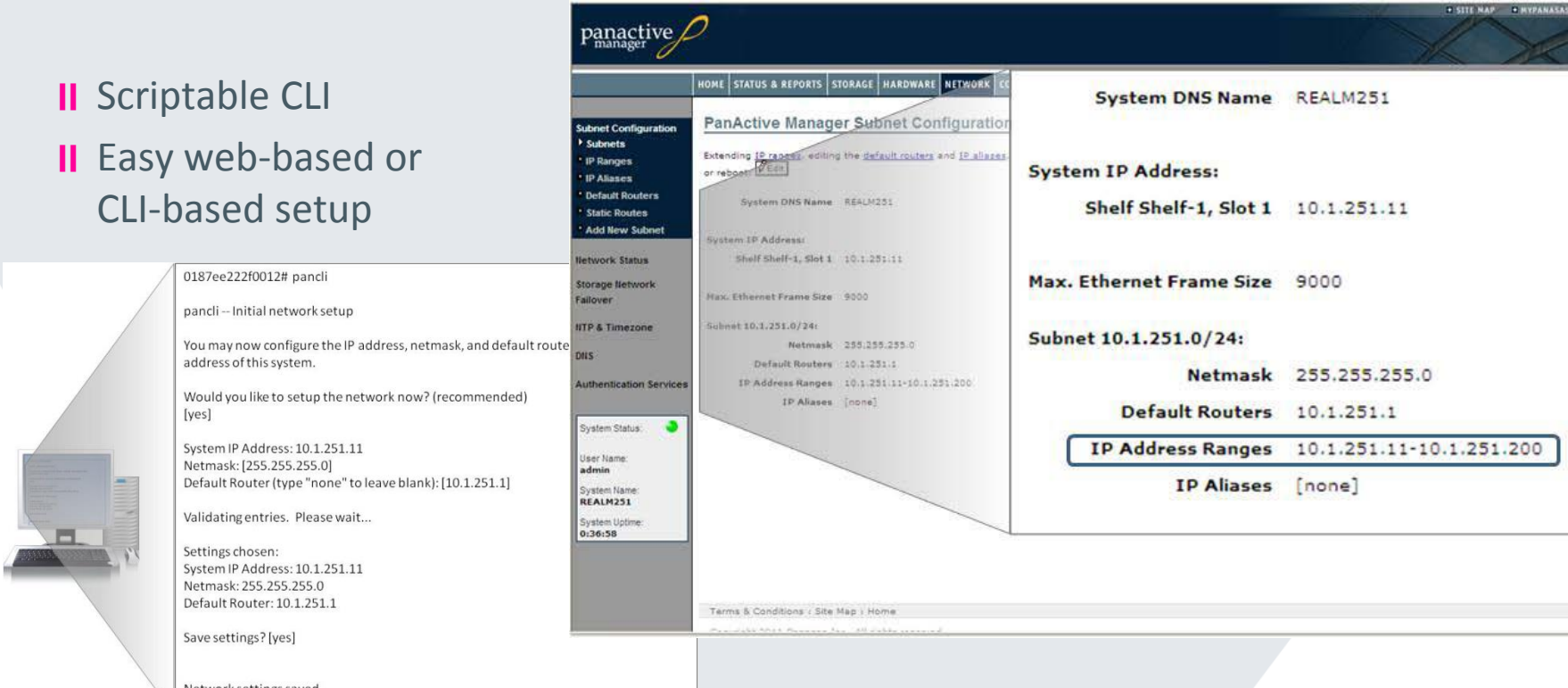
Panactive Manager

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

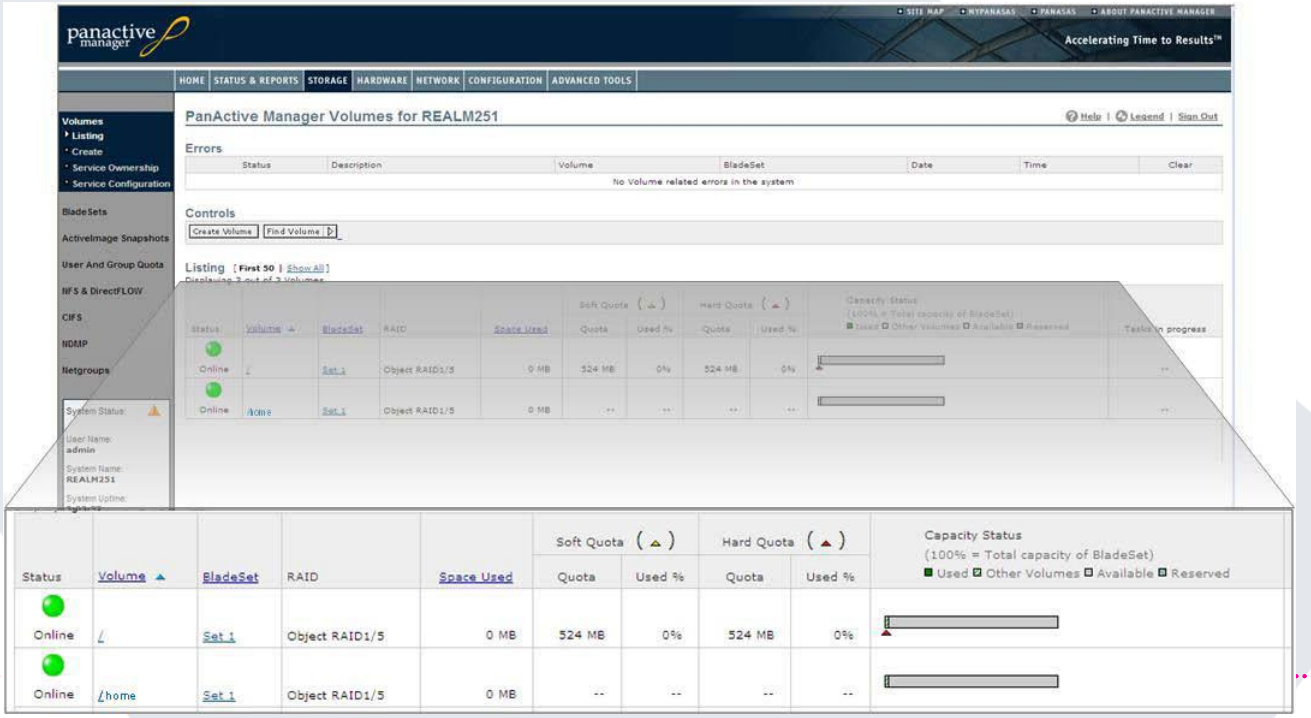


Easy to Manage

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

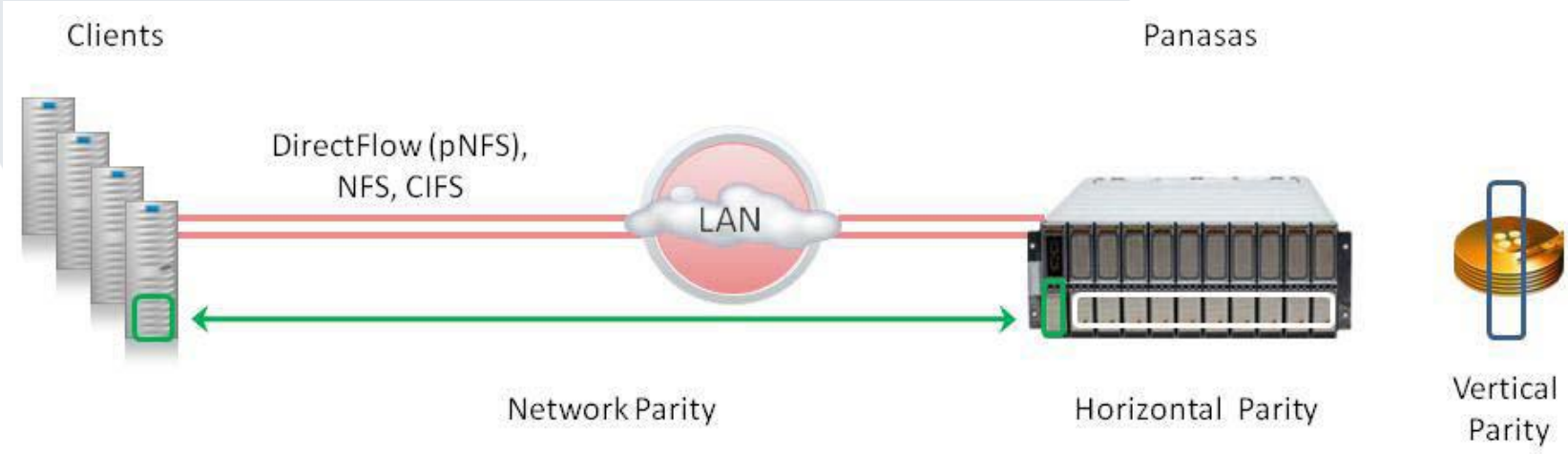


Extensive Management and Monitoring Capabilities



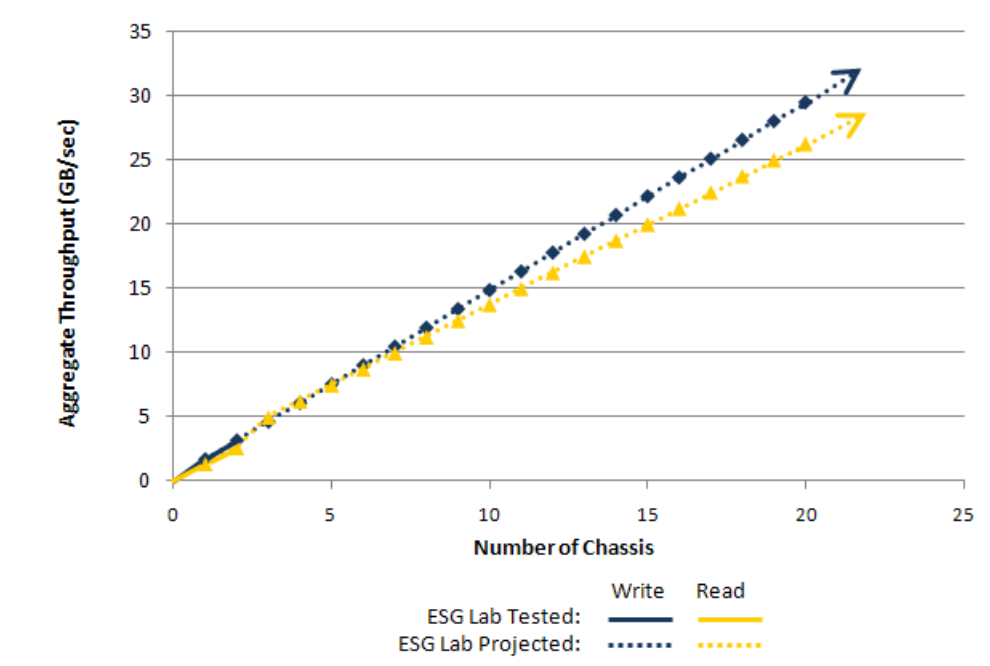
Tiered-Parity RAID

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



Scalable Performance

- ESG Labs Performance Tests with ActiveStor14 chassis



transtec

accelerate productivity