Accelerating Research Applications: Enterprise Storage for HPC Environments

04 May 2007

NAS is best?

- NAS is most applicable architecture, invented for simultaneous access from many clients, easy to use
- *Performance and scalability bottlenecks of "legacy NAS" are not trivial*

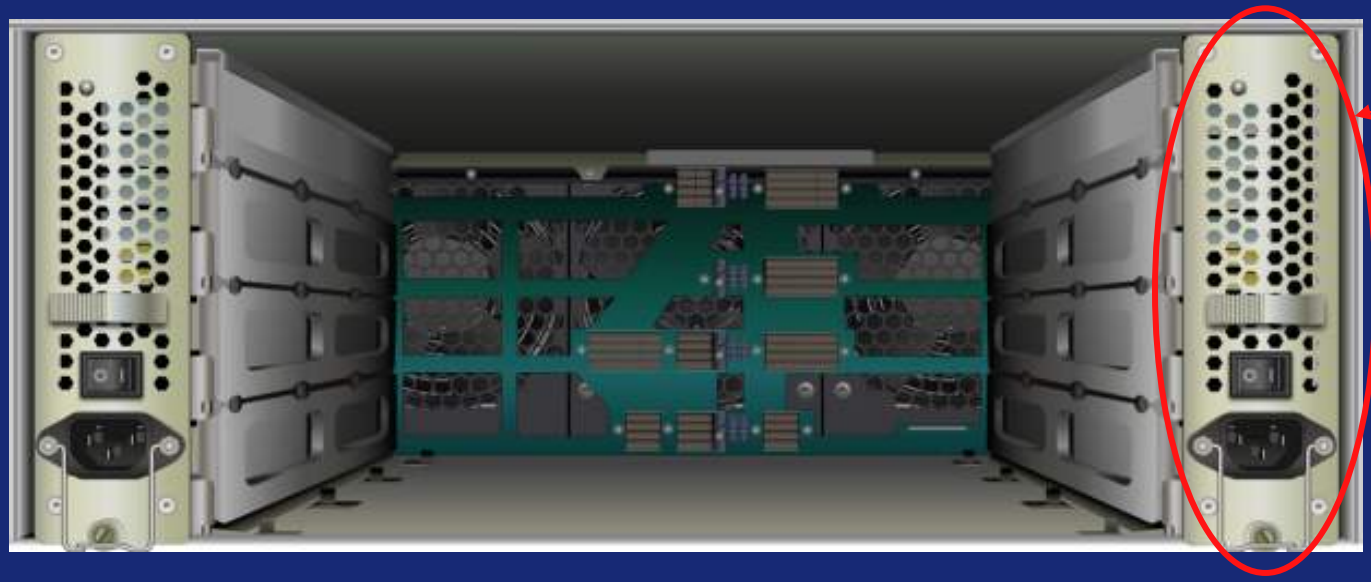
- Clustered DAS attractive, *but offers neither scalability of SAN nor ease of NAS*
- Is Clustered DAS cheap? *Not as inexpensive as it sounds when total costs are considered*

- SAN scales well, performs well, *but largely proprietary, very expensive, and requires specific skill set*
- SAN access physically controlled, *not good at sharing data with many clients or across large distances*

Architecture & Technology

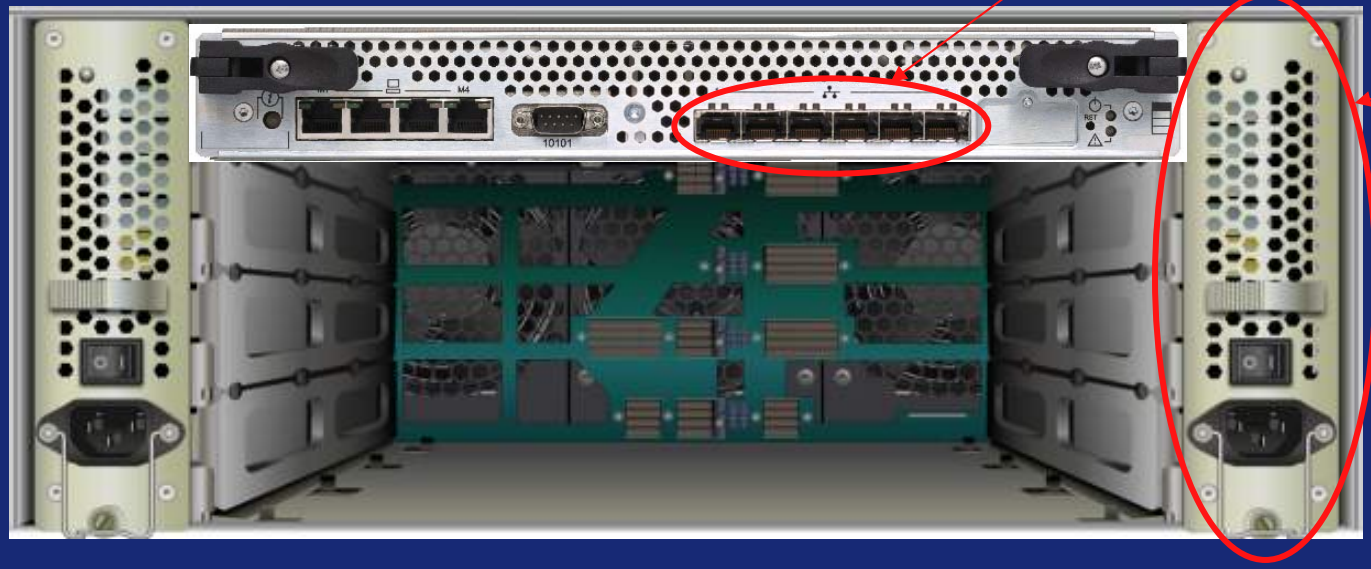
BLUEARC

Titan 2000 Chassis



Fully redundant power supplies

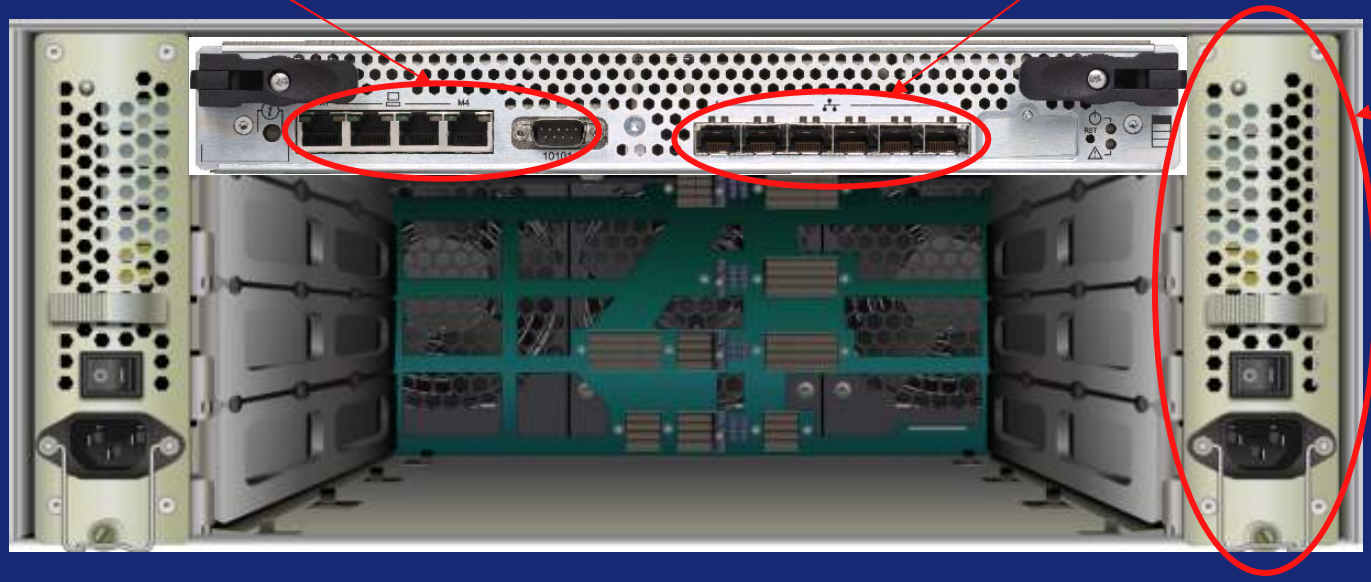
Titan 2000 Chassis



Titan 2000 Chassis

Dedicated 10/100 management LAN,
DB9 console port

6 x Gigabit Ethernet ports

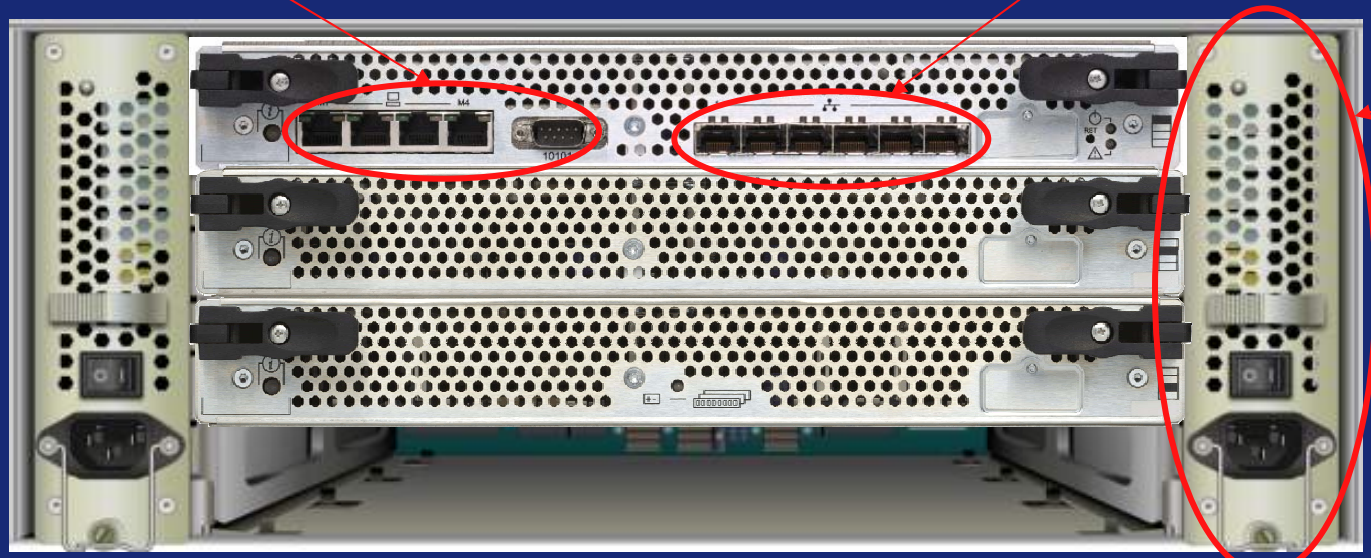


Fully redundant power supplies

Titan 2000 Chassis

Dedicated 10/100 management LAN, DB9 console port

6 x Gigabit Ethernet ports

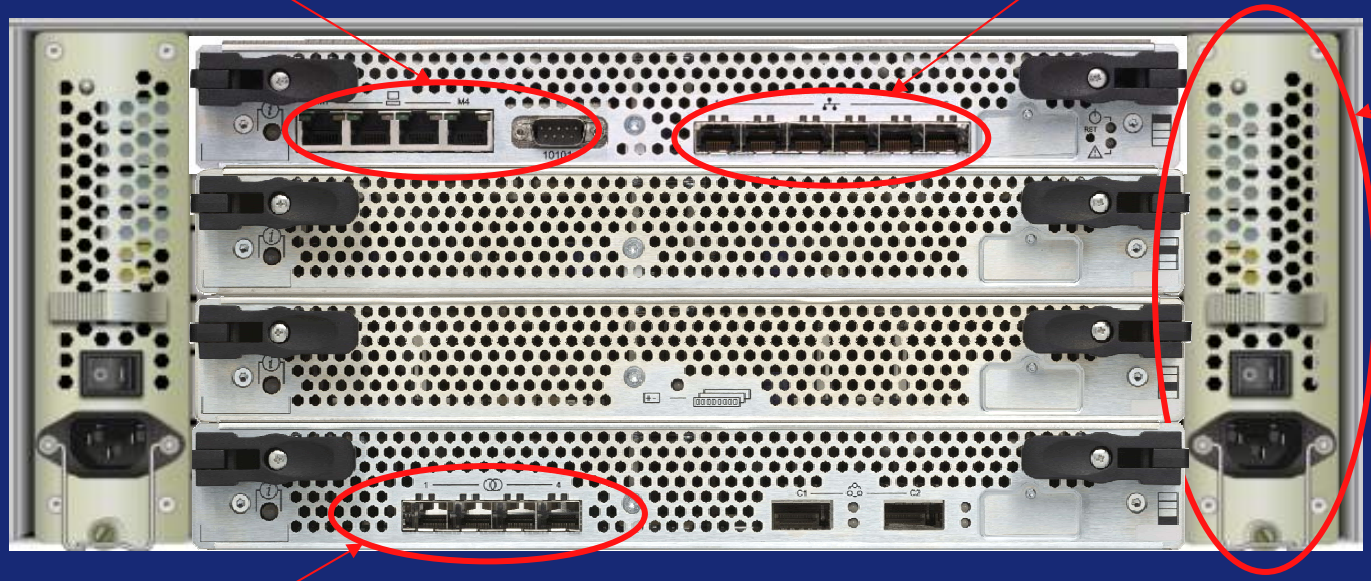


Fully redundant power supplies

Titan 2000 Chassis

Dedicated 10/100 management LAN, DB9 console port

6 x Gigabit Ethernet ports



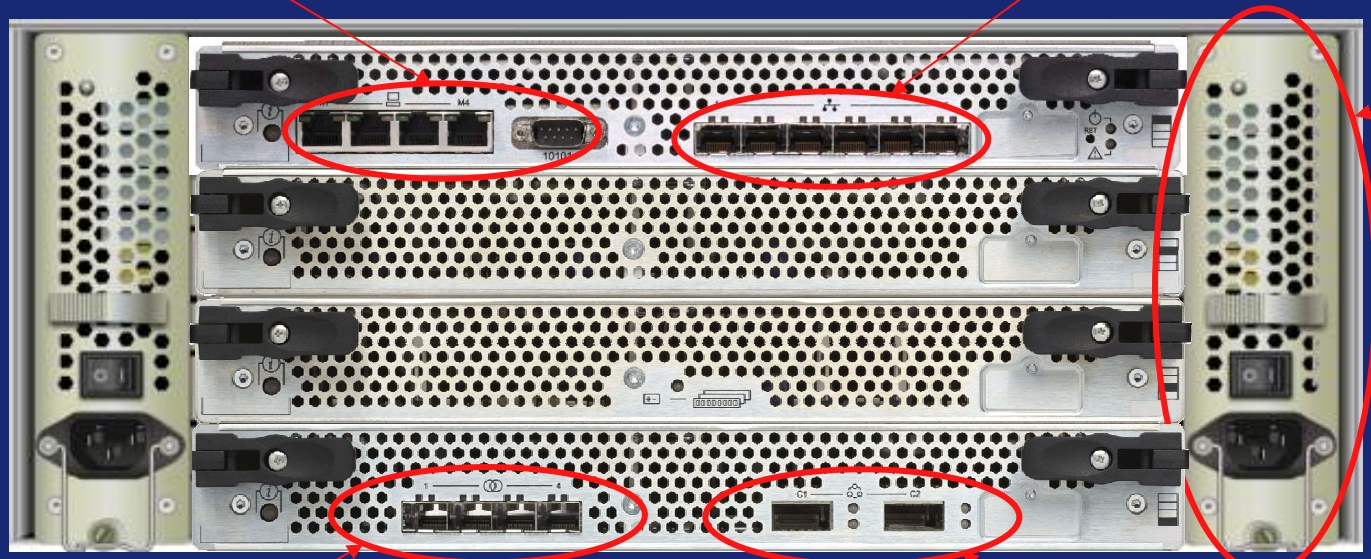
Fully redundant power supplies

4 x 4Gb FC ports

Titan 2000 Chassis

Dedicated 10/100 management LAN, DB9 console port

6 x Gigabit Ethernet ports



Fully redundant power supplies

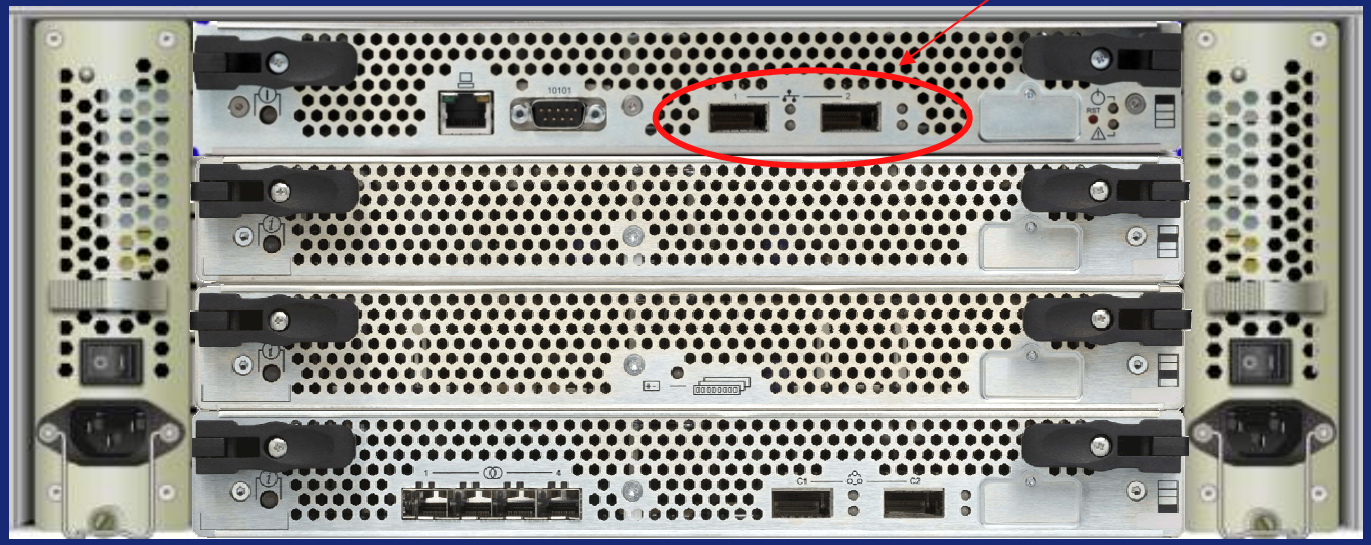
4 x 4Gb FC ports

2 x 10GbE XFP ports, dedicated HCI

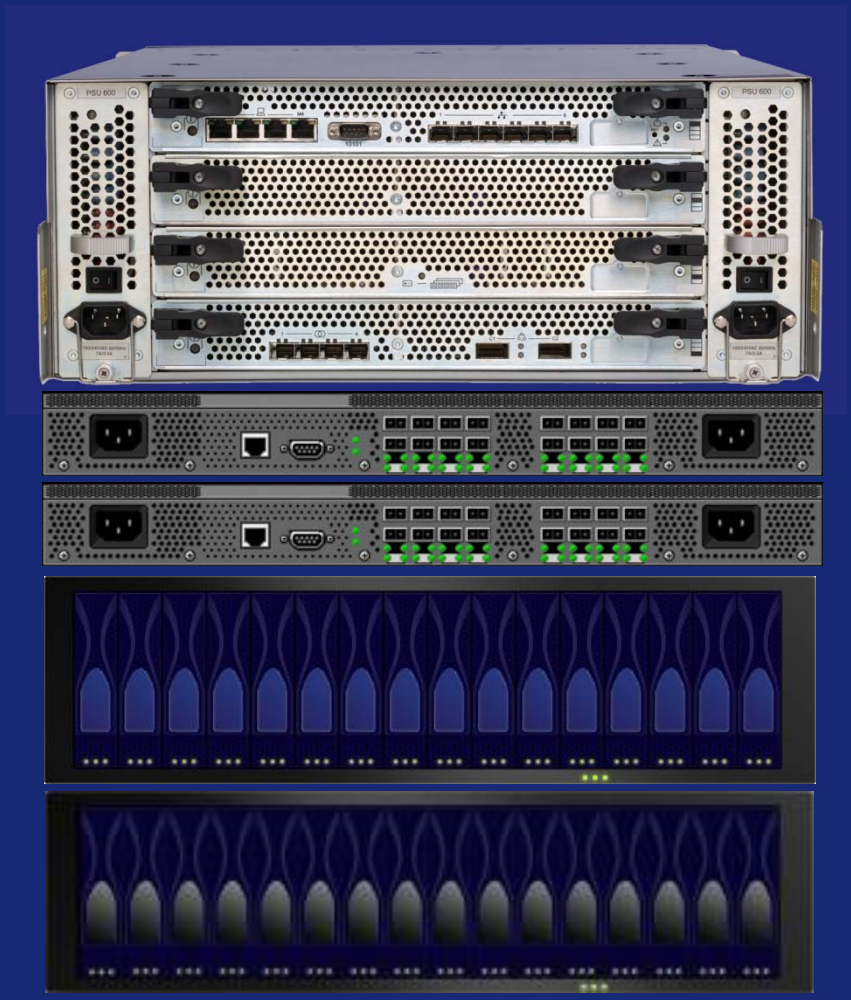
BLUEARC

Titan 2500 Chassis

2 x 10GbE XFP ports



Titan Storage Solution



Hybrid NAS-on-SAN architecture

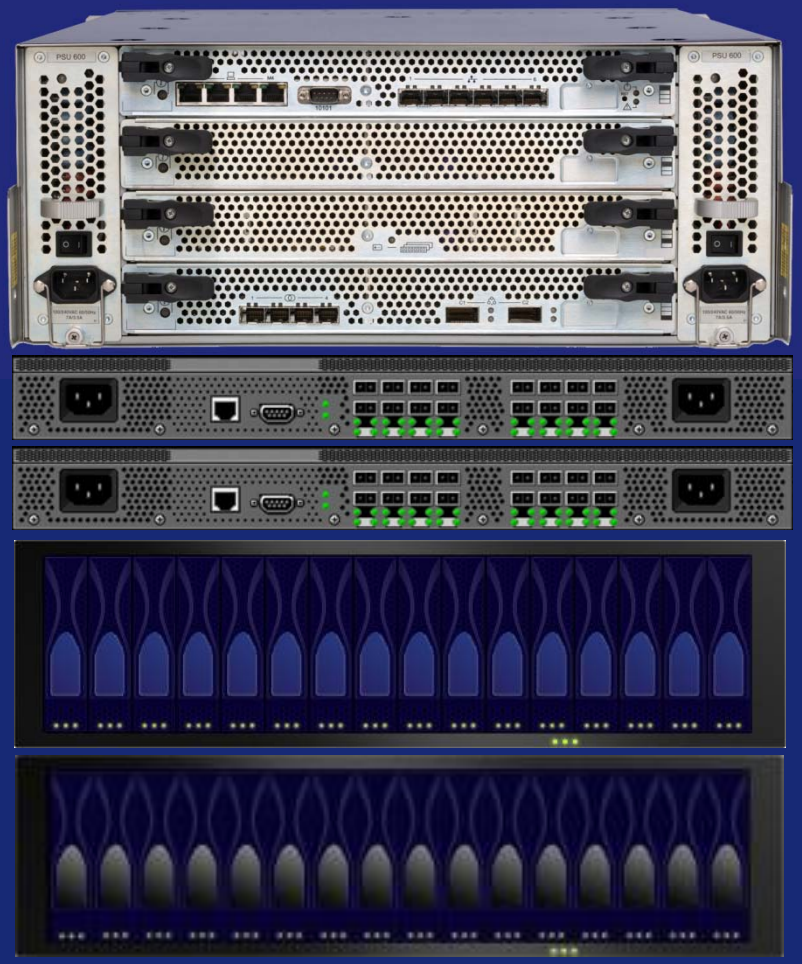
- Fault tolerant solution with redundant components
- Upgradeable, “future-proof” blade design

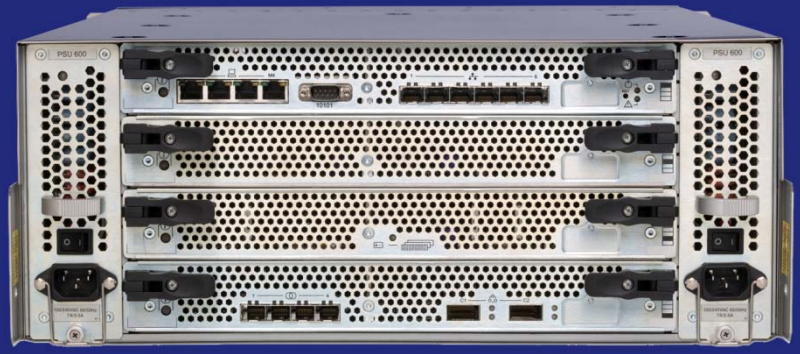
- Full 4Gb FC switched SAN fabric back-end

- Hardware-based RAID controllers
- Parallel RAID striping for increased performance
- Global hot spares for high availability
- Three levels of media error detection
- On-the-fly LUN expansion (optionally automated)

- Simultaneous FC and SATA disk options for true multi-tiered storage architectures, lower TCO
- Disk-to-Disk-to-Tape archiving

Titan Storage Solution





Clustered solutions too!

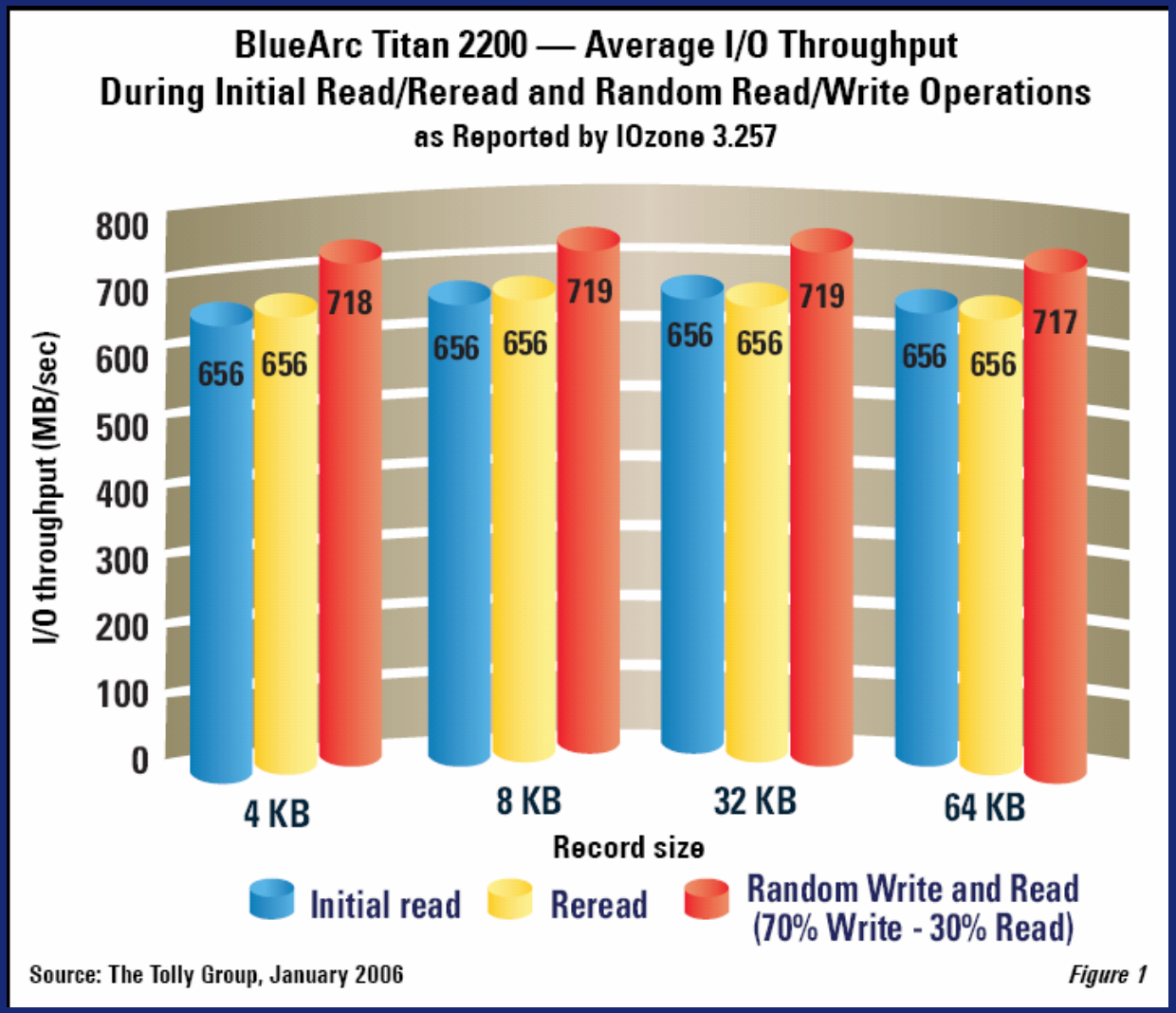
- Active-Active clustering with CNS™
- Automated failover
- Advanced replication options
- Extreme scalability

BLUEARC

Documented Performance

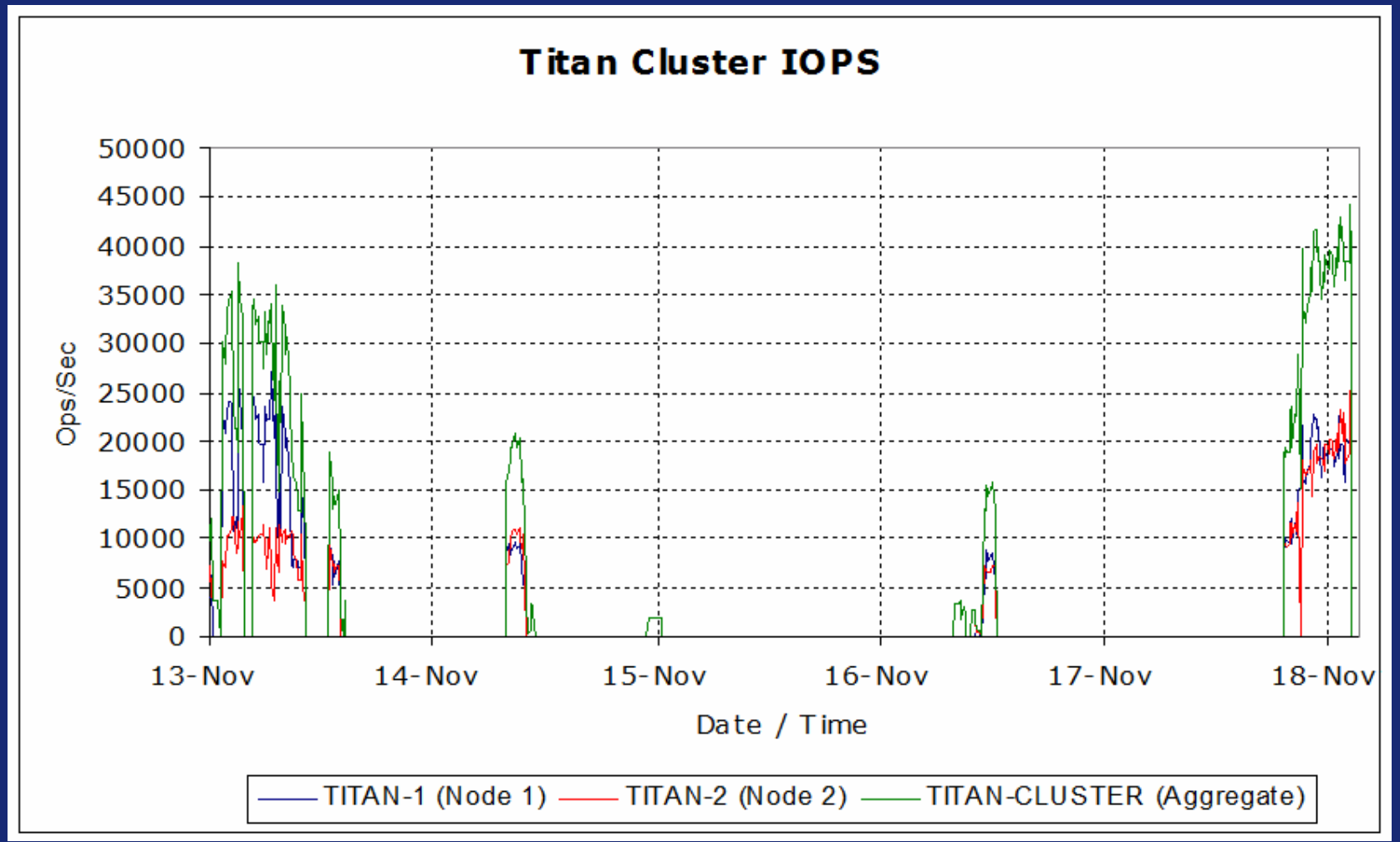
BLUEARC

The Tolly Group: Independently Verified

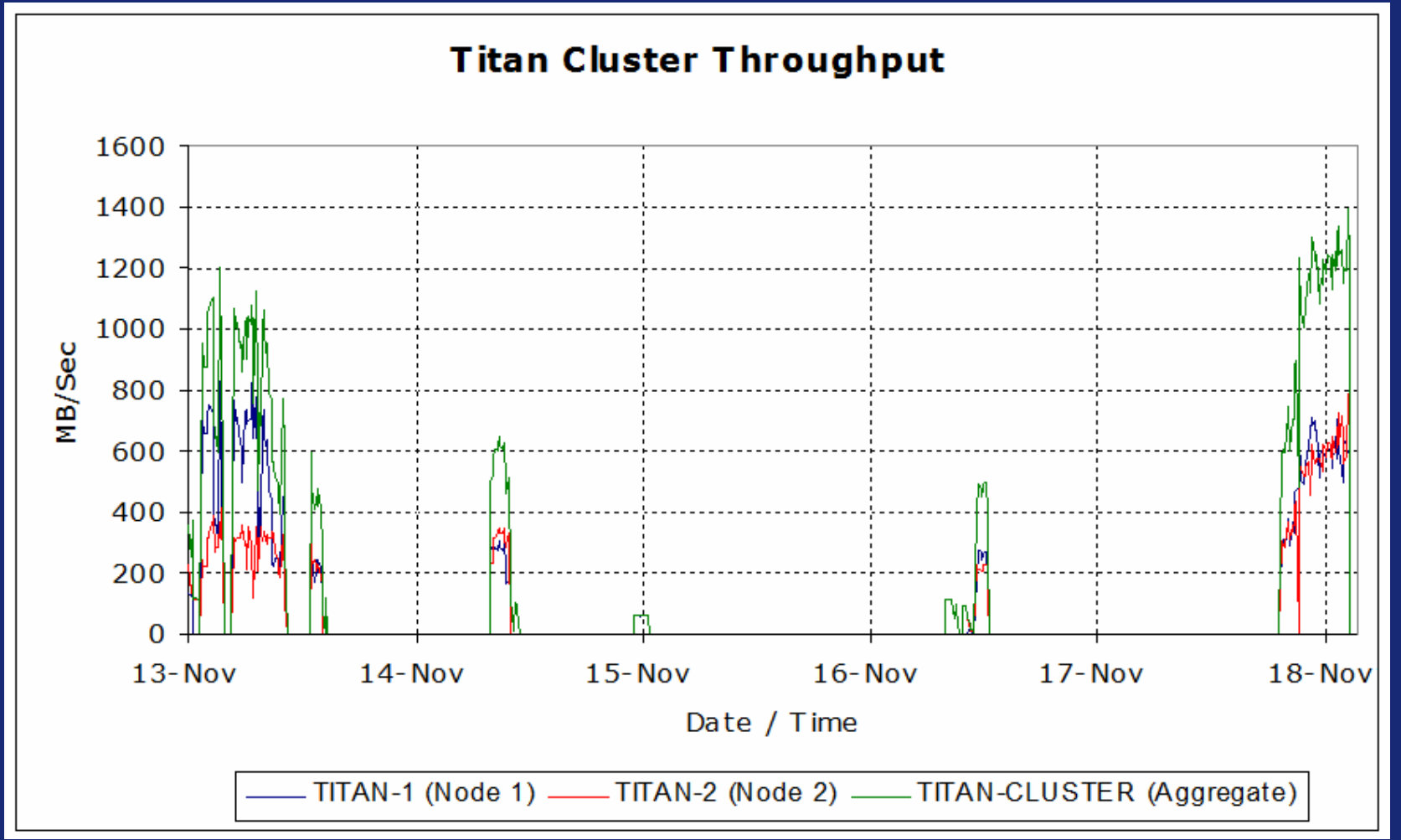


- Tolly Certified Results
  - >717 MB/s random mixed read/writes
  - Single Titan performance
  - 48 clients, GigE network, no jumbo frames
  - 146GB FC disk, 35TB usable
  - Throughput invariant with record size

SC|06 Bandwidth Challenge



SC|06 Bandwidth Challenge



SC|06 Bandwidth Challenge

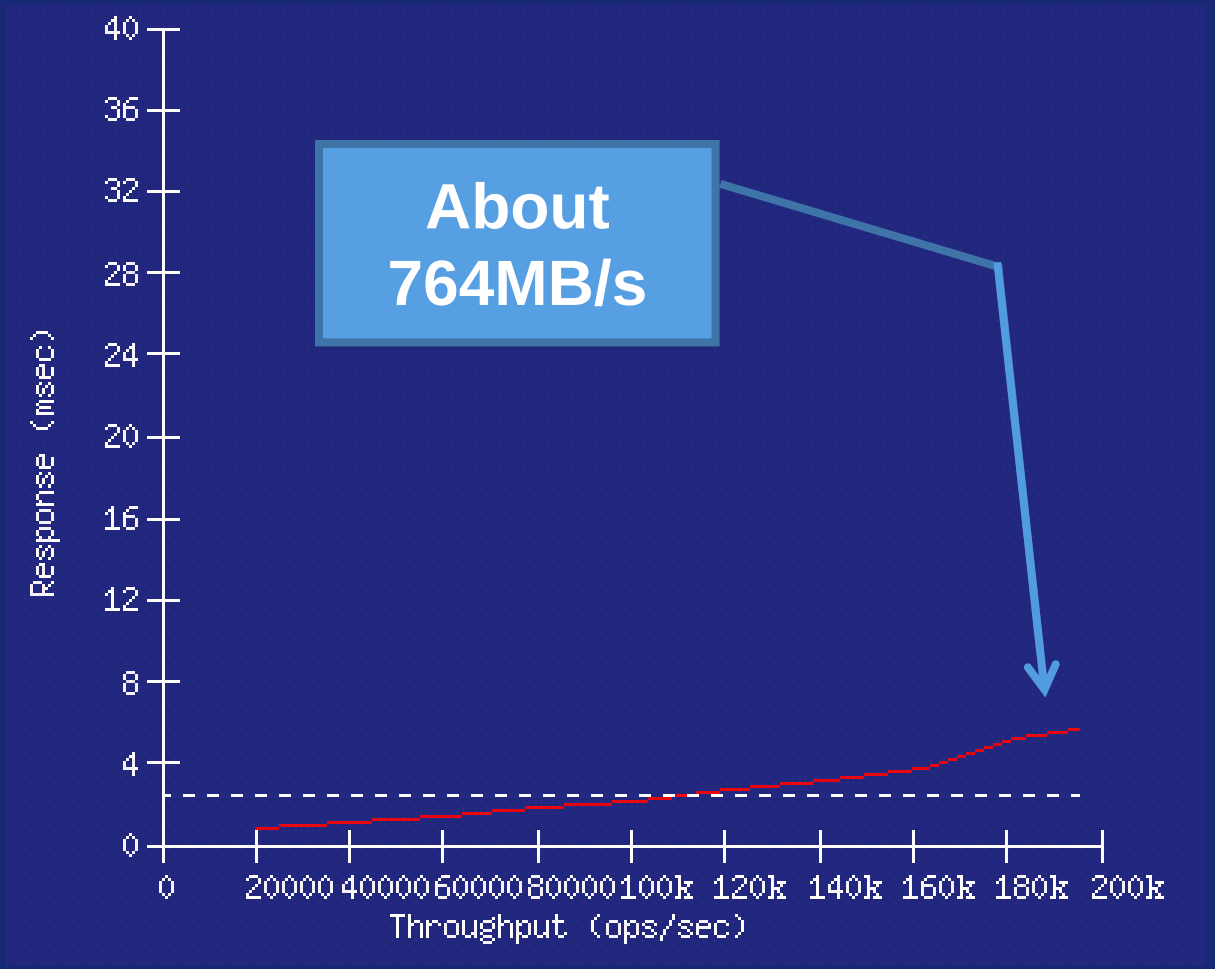
- 32kB block size
- 44,221 IOPS peak concurrent value
- Max concurrent R/W bandwidth delivered: 1382MB/s

- 2-node Titan 2200 cluster, 12 GbE links
- 3 filesystems per Titan node
- Single namespace (common mount-point)
- 140 146GB 15k FC disks

- 36 compute nodes (144 cores), 8 processes each
- 50% R/W mix; NFSv3, no particular optimizations

SPECsfs Results: Titan Cluster

Throughput (ops/sec)	Response (msec)
20208	0.8
40214	1.1
60618	1.4
80937	1.8
101499	2.2
121719	2.7
142104	3.1
162759	3.7
181640	5.2
195502	5.6



SPECsfs97_R1.v3 results, AA clustered Titan 2200s, single namespace showing Titan cluster response time as a function of delivered I/O performance

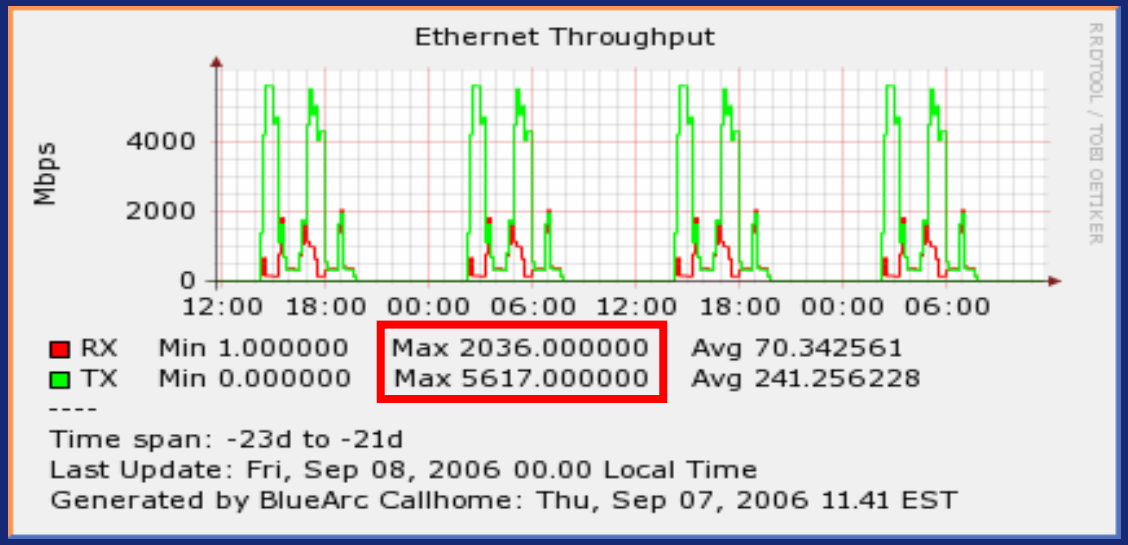
SpecSFS to BWC comparison

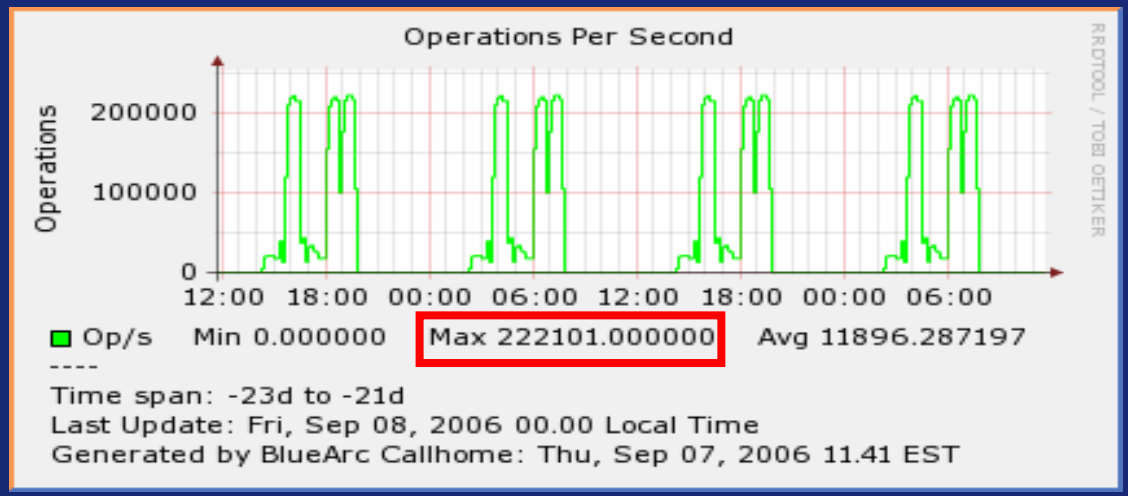
- 4kB block size
- 195,502 IOPS peak concurrent value
- Max concurrent R/W bandwidth delivered: 764MB/s

- 2-node Titan 2200 cluster, 12 GbE links
- 8 filesystems per Titan node
- Single namespace (common mount-point)
- 416 73GB 15k FC disks

- 28 compute nodes (56 cores), 32 processes each
- Spec R/W mix; NFSv3, no particular optimizations

Real World Performance





- 48 hour cyclical workload, single Titan
- Large computational Linux cluster
- High performance
  - Peak throughput over 700MB/sec Reads, 250MB/sec Writes
  - Over 222k aggregate peak IOPs

Multi-Tiered Storage

146 GB Fibre Channel 15K RPM

146 GB Fibre Channel 15K RPM

300 GB Fibre Channel 10K RPM

300 GB Fibre Channel 10K RPM

500 GB SATA 7.2K RPM

500 GB SATA 7.2K RPM

750 GB SATA 7.2K RPM

750 GB SATA 7.2K RPM

- Match disk to application and price requirements
- Any combination of FC and SATA
- Immediately benefit from new disk technology and prices
- Disk-to-disk-to-tape eliminates the backup window
- LAN-free data migration

Tape Library

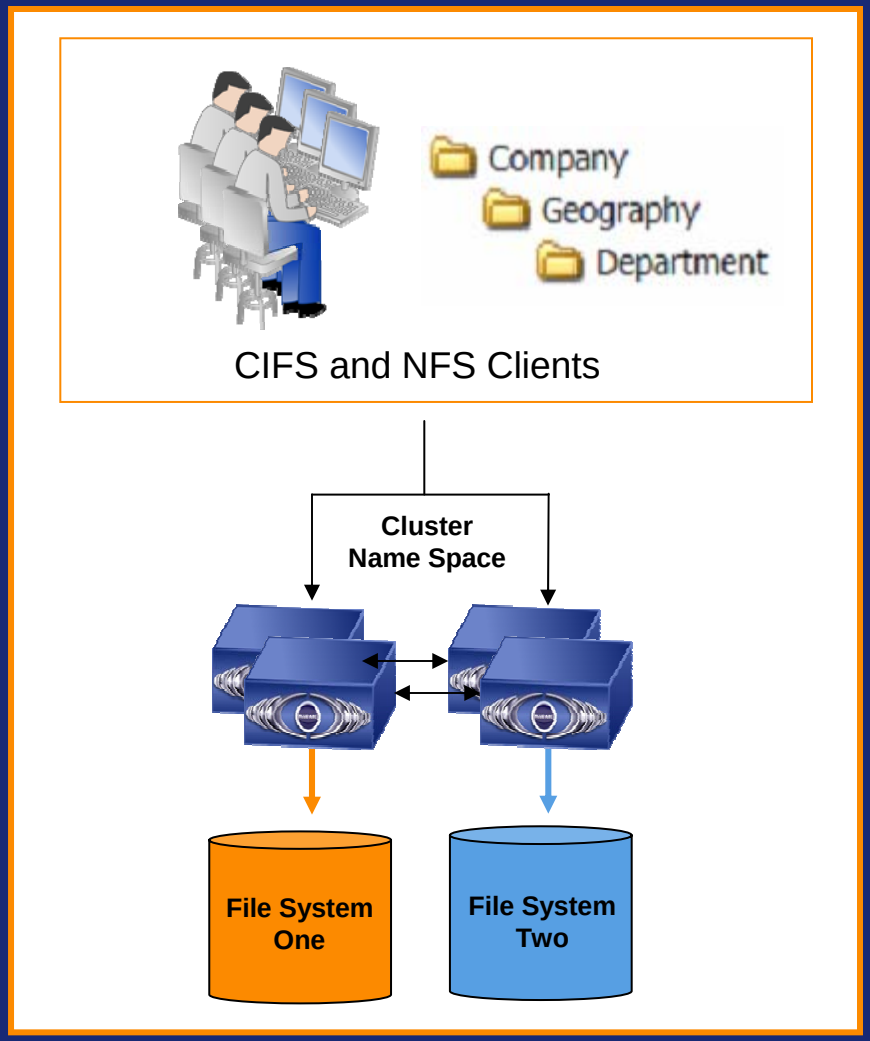
BLUEARC

Multi-tiered storage design

- 146GB FC, 15k – max spindles for max IOPS
- 300GB FC,10k – good for IOPS, bandwidth, many tasks or unknown storage I/O profiles
- 500GB → 750GB → 1TB SATA – cost effective, high-density, decent bandwidth

- 4kB blocks to maximize IOPS
- 32kB blocks for flexible storage I/O needs
- Larger blocks to maximize bandwidth

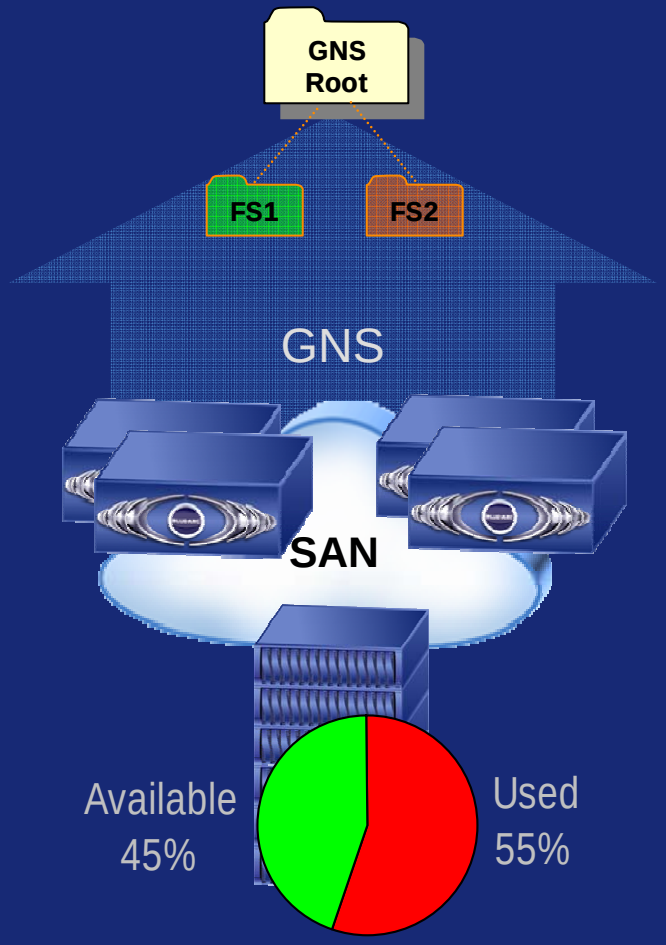
Cluster Name Space



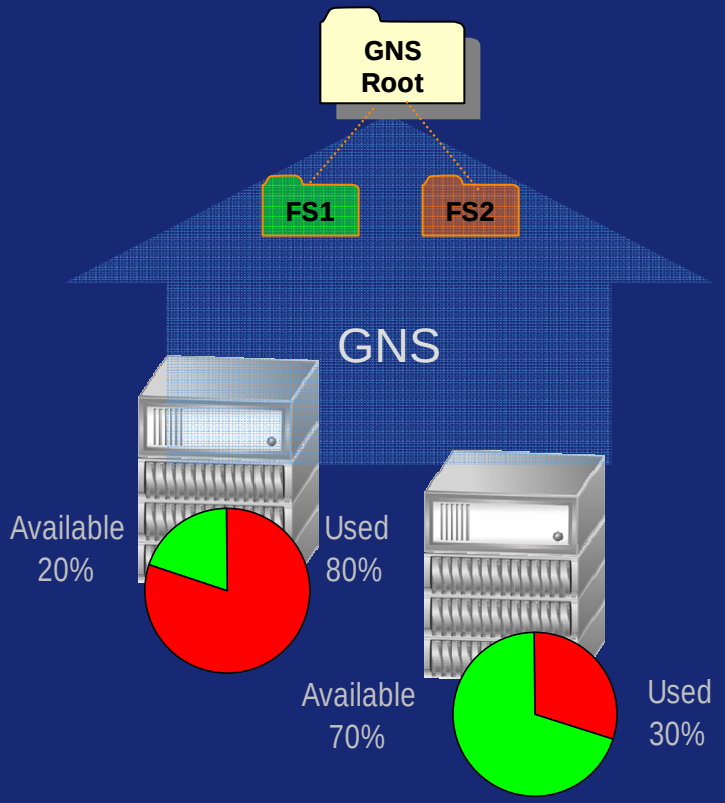
- Features:
  - Single name space
  - NFS and CIFS Support
  - Dual 10GbE Cluster Interconnect
  - Request Redirection in HW
  - Multi-node Read Caching
- Benefits:
  - Single mount point and file system for simplified user administration
    - Universal CNS Access
    - Unified Directory Structure
  - Load balancing

Shared GNS Cluster Storage Pool

Balanced & efficient storage use



BlueArc Shared Storage Pool



Separate unbalanced storage

DAS-like inefficiency

Balanced, efficient, scalable storage

- Titan not particularly optimized for small-file, high-IOPS nor large-file, high-bandwidth profiles, just happens to be good at both
- Design tiers to meet just about any I/O profile
- Add more Titans for higher performance without needing to add capacity
- Add more disk for higher capacity if needed
- True separation-of-function in storage design


Life Sciences
Entertainment
Internet Services
BLUE ARC®
Any Questions?