Advanced Lustre Infrastructure Monitoring

Torben Kling Petersen, PhD

Principal Solution Architect, HPC

Xyratex: A Leading Provider of Data Storage Technology to OEMs

- **> 4,000 Petabytes of storage shipped in 2011**

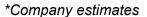- **Largest OEM Disk Storage System provider**

Enterprise Data Storage Solutions

- **~ 50%** of w/w disk drives are produced utilizing Xyratex Technology*

- **Largest independent** supplier of Disk Drive Capital Equipment

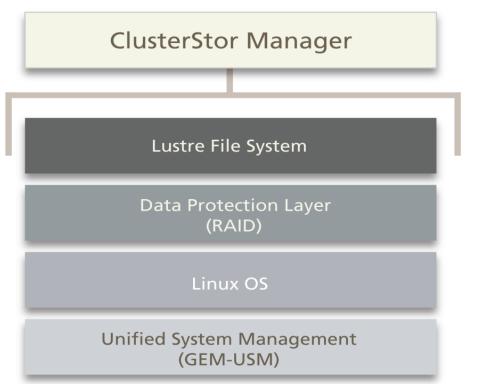*HDD Capital Equipment Solutions*

Company estimates

x y r a t e x

ClusterStor Design Philosophy

- Architected
- Integrated
- Tested
- Optimized
- Qualified
- Supported

- Factory integration
- Component and system testing
- System shipped to site, not built on site
- Single owner of entire stack
- Global Support capability

ClusterStor Manager

Lustre File System

Data Protection Layer (RAID)

Linux OS

Unified System Management (GEM-USM)

x y r a t e x

CS-2584 - Scalable Storage Unit (SSU) – Lustre OSS

- **Ultra HD - CS-2584 SSU - OSS**
  - ➢ 5U84 Enclosure – completely H/A
    - ▪ Two (2) trays of 42 HDD's each
    - ▪ Dual-ported 3.5" FatSAS & SSD HDD Support
    - ▪ 150MB/s SAS available bandwidth per HDD
  - ➢ Pair of H/A Embedded Application Servers
    - ▪ CS-3000: = 3.5 GB/sec IOR over IB
    - ▪ CS-6000: = 6 GB/sec IOR over IB
  - ➢ IB QDR/FDR or 10/40 GbE Network Link
  - ➢ Data Protection/Integrity (RAID 6, 8+2)
    - ▪ 2 OSS's per SSU
    - ▪ 4 OST's per OSS
  - ➢ 2x SSD OSS journal disks for increased performance
    - ▪ 2X Hot Spare HDD's
  - ➢ 64 Usable Data Disks per SSU
    - ▪ 1TB x 64 – 64TB usable per SSU
    - ▪ 2TB x 64 - 128TB usable per SSU
    - ▪ 3TB x 64 - 192TB usable per SSU
    - ▪ 4TB x 64 - 256TB usable per SSU

Only 5^0 C delta
with drawer open
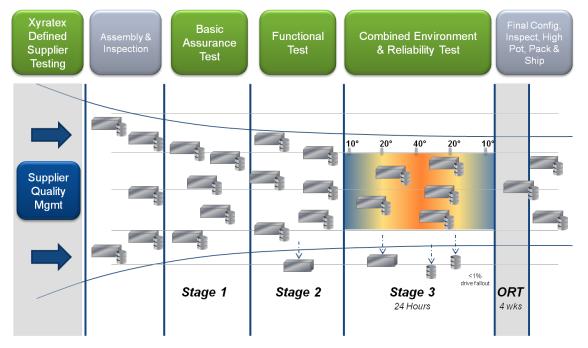
Embedded
server modules

xyratex

Extensive Testing = Reliability = System uptime

Integrated System Testing (IST) is a patented 3 Stage testing process embedded within manufacturing and designed to remove hidden quality problems

Features

Optimized 36 Hour Manufacturing & Test

Adaptable Test Automation

Standard Across the Globe

Benefits

- Reduces solution warranty and service costs

- Reduces Infant Mortality

- Up to 1.5X drive reliability improvement over 3 Yrs.

  o AFR Reduction to 1% or less

  o 67% less disk drive failures in first 3 months

- Accelerates time to market

5

Xyratex HDD Reliability : Failure Rate Comparison

Annual Failure Rate (AFR) by drive class

NetApp Study, 1.8M HDDs, 155K systems over 44 months, 99.99% reliability

Industry AFR experience

	3 Months	6 Months	1 Year	2 Years	3 Years	4 Years	5 Years
XYR ATA FR	0.92%	0.98%	1.04%	1.27%	2.19%		
XYR Enterprise FR	0.46%	0.51%	0.73%	1.32%	0.68%	1.04%	1.10%
Google paper Base ~AFR	2.80%	1.80%	1.75%	8%	8.70%	6%	7.40%

Product Under Test

- Up to 30-day 'Soak Test'
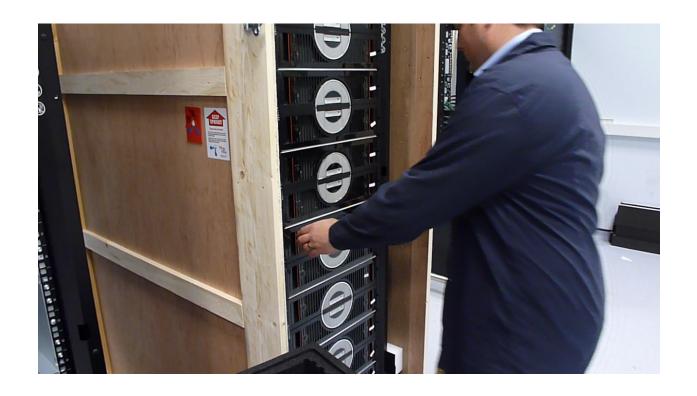
- Soak test measures:

  - I/O connectivity to (ClusterStor to Lustre clients)

  - I/O performance - read/write/rewrite (ClusterStor)

- Tests a system with significant load extended over a significant period of time

- Includes "adverse" conditions testing (running HA scenarios for ClusterStor systems)

xyratex

Drive Installation / Unloading Process

- The drives are removed from the unit with the use of a speed loader.
- The speed loader allows the user to rapidly remove and install 7 drives at a time.
- The packaging and loader compliment each other, thus significantly reducing the handling time.

xyratex

Racks are reinforced with an additional 32 rivets to ensure quality!

xyratex

- **Xyratex delivers a complete ready-to-run ClusterStor solution**
  - Sizing and Configuration optimization
    - Performance centric
    - Capacity centric
  - Factory Integration & Staging
    - Rack integration & Cabling
    - Entire storage software stack factory pre-installed and pre-configured
    - System soak test and benchmark testing area at Xyratex factory
  - Drive speed-loader reduces drive insertion time by 85%

x y r a t e x

ClusterStor CS-3000 & CS-6000

CS-3000 Overview

- Targeted >**24** GB/s per rack
- Overall Performance scalable to >100GB/s bandwidth
- Overall Capacity scalable to >30PBs

CS-6000 Overview

- Targeted >**42** GB/s per rack
- Overall Performance scalable to >1TB/s bandwidth
- Overall capacity scalable to >100PBs

ClusterStor a complete ready-to-run Lustre solution

- Up to 560 HDD's per rack (42RU)
- Up to 1.8PBs usable per rack (with 4TB HDD's)
- Up to 14 Application Controllers per rack
- Up to 14 high bandwidth Network connectivity ports/rack
- Factory Integration & Staging
  - Rack integration & Cabling
- Entire storage software stack factory pre-installed and pre-configured
  - System Burn-in and benchmark testing area at Xyratex factory
  - "Rack'n'Roll" installation – hours vs. days or weeks

ClusterStor™
xyratex

xyratex

- ## **Requirements:**
  - ➢ Compute system capable of at least 10 PFLOPs
  - ➢ Storage capable of doing 10% of Compute -> 1 000 GB/s
  - ➢ Energy efficient
  - ➢ Incredible reliability (well, let's settle for decent ....)
  - ➢ Supportable for 3-5 years …

Throughput reqs (GB/s)	**1000**	
Embedded Server	**CS6000**	
SSU Performance (GB/s)	**5**	
Volume requirements (TB)	**10 000**	
Disk size (TB)	**2**	
Rack size (42 or 48RU)	**42**	
Power (SSUs) kW	**2,08**	
SSUs per Rack (8 max)	**6**	

ClusterStor™ xyratex

	# SSUs	Total usable volume	Agg. throughput	IB Uplink ports	# Racks	# OSTs	# HHDs	Power reqs (kW)	Weight (T)	Floor space (m2)
Solution (performance)	210	26 880 TB	1050 GB/s	422	35	1 680	**17 220**	438,8 kW	40,25	42
Solution (Full racks)	279	35 712 TB	1395 GB/s	560	35	2 232	**22 878**	582,3 kW	40,25	42

xyratex

Managing a monster …

CLI - Worked in the past, works now, …. right ??

The first 50 …..
   (of 360 OSS nodes …)

"I've got my own scripts ..."

"CLI will always be faster.."
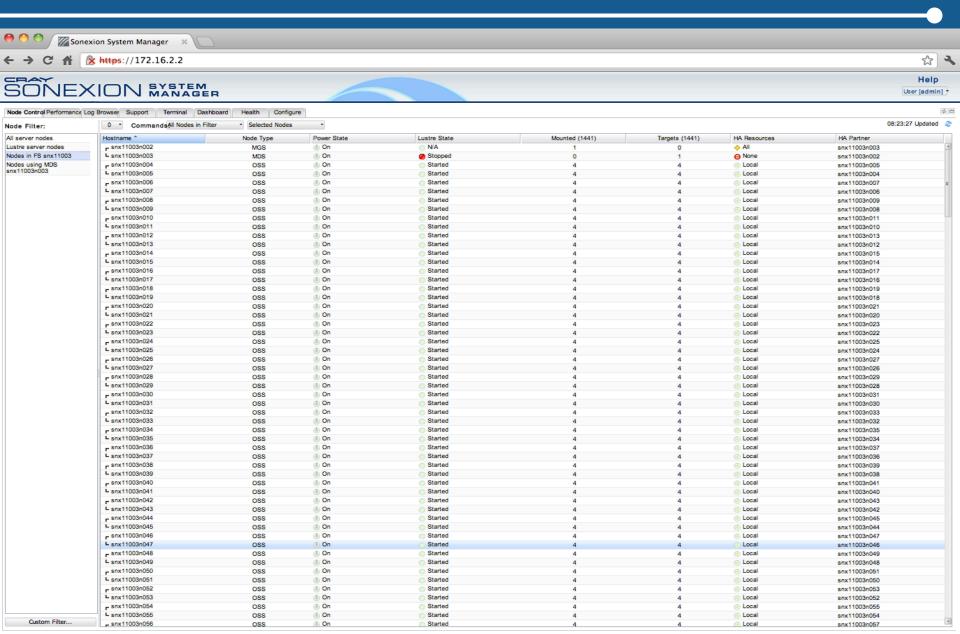
If it can't fit onto a single screen ….

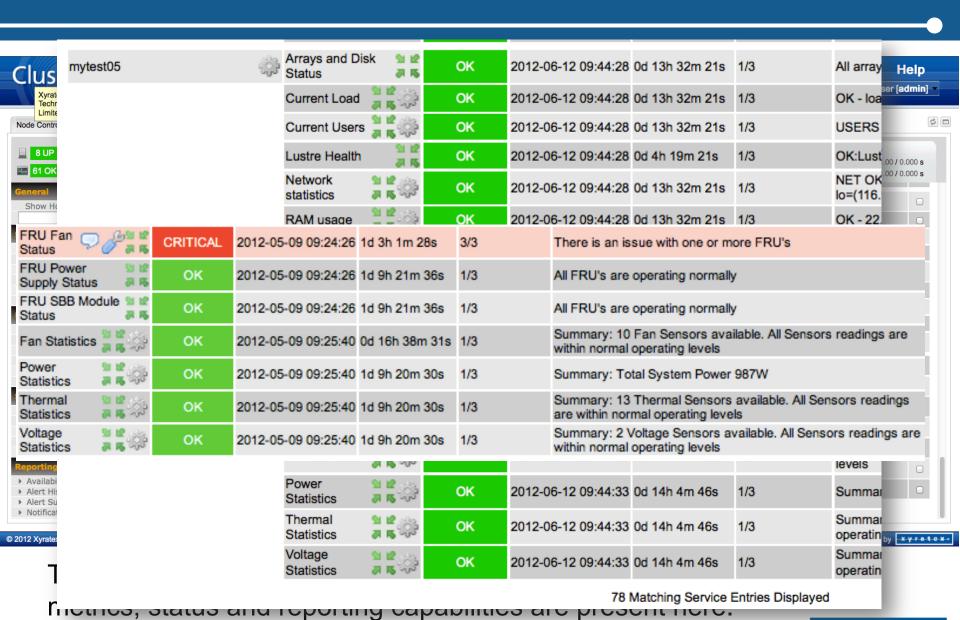Pipe it to something ..

```
#cscli show_nodes -c |
/dev/null
```

Are we having fun yet ???

```
[root@snx11003n000 admin]# date
Thu Jun  7 14:33:46 PDT 2012
[root@snx11003n000 admin]# /opt/xyratex/bin/cscli show_nodes -c snx11003n
-------------------------------------------------------------------------------------
Hostname        Node type     Power state    Lustre state    Targets    Partner         HA Resources
snx11003n000    mds           on             N/A             0 / 0      snx11003n001    None
snx11003n001    mds           on             Started         1 / 1      snx11003n000    Local
snx11003n002    oss           on             Started         4 / 4      snx11003n003    Local
snx11003n003    oss           on             Started         4 / 4      snx11003n002    Local
snx11003n004    oss           on             Started         4 / 4      snx11003n005    Local
snx11003n005    oss           on             Started         4 / 4      snx11003n004    Local
snx11003n006    oss           on             Started         4 / 4      snx11003n007    Local
snx11003n007    oss           on             Started         4 / 4      snx11003n006    Local
snx11003n008    oss           on             Started         4 / 4      snx11003n009    Local
snx11003n009    oss           on             Started         4 / 4      snx11003n008    Local
snx11003n010    oss           on             Started         4 / 4      snx11003n011    Local
snx11003n011    oss           on             Started         4 / 4      snx11003n010    Local
snx11003n012    oss           on             Started         4 / 4      snx11003n013    Local
snx11003n013    oss           on             Started         4 / 4      snx11003n012    Local
snx11003n014    oss           on             Started         4 / 4      snx11003n015    Local
snx11003n015    oss           on             Started         4 / 4      snx11003n014    Local
snx11003n016    oss           on             Started         4 / 4      snx11003n017    Local
snx11003n017    oss           on             Started         4 / 4      snx11003n016    Local
snx11003n018    oss           on             Started         4 / 4      snx11003n019    Local
snx11003n019    oss           on             Started         4 / 4      snx11003n018    Local
snx11003n020    oss           on             Started         4 / 4      snx11003n021    Local
snx11003n021    oss           on             Started         4 / 4      snx11003n020    Local
snx11003n022    oss           on             Started         4 / 4      snx11003n023    Local
snx11003n023    oss           on             Started         4 / 4      snx11003n022    Local
snx11003n024    oss           on             Started         4 / 4      snx11003n025    Local
snx11003n025    oss           on             Started         4 / 4      snx11003n024    Local
snx11003n026    oss           on             Started         4 / 4      snx11003n027    Local
snx11003n027    oss           on             Started         4 / 4      snx11003n026    Local
snx11003n028    oss           on             Started         4 / 4      snx11003n029    Local
snx11003n029    oss           on             Started         4 / 4      snx11003n028    Local
snx11003n030    oss           on             Started         4 / 4      snx11003n031    Local
snx11003n031    oss           on             Started         4 / 4      snx11003n030    Local
snx11003n032    oss           on             Started         4 / 4      snx11003n033    Local
snx11003n033    oss           on             Started         4 / 4      snx11003n032    Local
snx11003n034    oss           on             Started         4 / 4      snx11003n035    Local
snx11003n035    oss           on             Started         4 / 4      snx11003n034    Local
snx11003n036    oss           on             Started         4 / 4      snx11003n037    Local
snx11003n037    oss           on             Started         4 / 4      snx11003n036    Local
snx11003n038    oss           on             Started         4 / 4      snx11003n039    Local
snx11003n039    oss           on             Started         4 / 4      snx11003n038    Local
snx11003n040    oss           on             Started         4 / 4      snx11003n041    Local
snx11003n041    oss           on             Started         4 / 4      snx11003n040    Local
snx11003n042    oss           on             Started         4 / 4      snx11003n043    Local
snx11003n043    oss           on             Started         4 / 4      snx11003n042    Local
snx11003n044    oss           on             Started         4 / 4      snx11003n045    Local
snx11003n045    oss           on             Started         4 / 4      snx11003n044    Local
snx11003n046    oss           on             Started         4 / 4      snx11003n047    Local
snx11003n047    oss           on             Started         4 / 4      snx11003n046    Local
snx11003n048    oss           on             Started         4 / 4      snx11003n049    Local
snx11003n049    oss           on             Started         4 / 4      snx11003n048    Local
-------------------------------------------------------------------------------------
[root@snx11003n000 admin]#
```
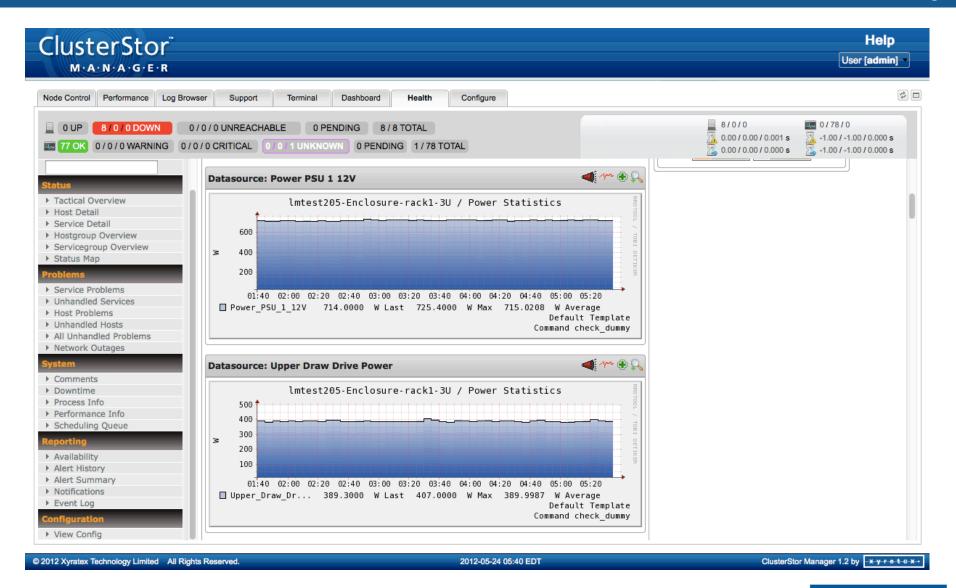
Trying the GUI instead (same 50 nodes ….)

mytest05	Arrays and Disk Status	OK	2012-06-12 09:44:28	0d 13h 32m 21s	1/3	All array
	Current Load	OK	2012-06-12 09:44:28	0d 13h 32m 21s	1/3	OK - loa
	Current Users	OK	2012-06-12 09:44:28	0d 13h 32m 21s	1/3	USERS
	Lustre Health	OK	2012-06-12 09:44:28	0d 4h 19m 21s	1/3	OK:Lust
	Network statistics	OK	2012-06-12 09:44:28	0d 13h 32m 21s	1/3	NET OK lo=(116.
	RAM usage	OK	2012-06-12 09:44:28	0d 13h 32m 21s	1/3	OK - 22

FRU Fan Status	CRITICAL	2012-05-09 09:24:26	1d 3h 1m 28s	3/3	There is an issue with one or more FRU's
FRU Power Supply Status	OK	2012-05-09 09:24:26	1d 9h 21m 36s	1/3	All FRU's are operating normally
FRU SBB Module Status	OK	2012-05-09 09:24:26	1d 9h 21m 36s	1/3	All FRU's are operating normally
Fan Statistics	OK	2012-05-09 09:25:40	0d 16h 38m 31s	1/3	Summary: 10 Fan Sensors available. All Sensors readings are within normal operating levels
Power Statistics	OK	2012-05-09 09:25:40	1d 9h 20m 30s	1/3	Summary: Total System Power 987W
Thermal Statistics	OK	2012-05-09 09:25:40	1d 9h 20m 30s	1/3	Summary: 13 Thermal Sensors available. All Sensors readings are within normal operating levels
Voltage Statistics	OK	2012-05-09 09:25:40	1d 9h 20m 30s	1/3	Summary: 2 Voltage Sensors available. All Sensors readings are within normal operating levels

	Power Statistics	OK	2012-06-12 09:44:33	0d 14h 4m 46s	1/3	Summar
	Thermal Statistics	OK	2012-06-12 09:44:33	0d 14h 4m 46s	1/3	Summar operatin
	Voltage Statistics	OK	2012-06-12 09:44:33	0d 14h 4m 46s	1/3	Summar operatin

78 Matching Service Entries Displayed

T

metrics, status and reporting capabilities are present here.

Monitor everything – Power usage

Compute and storage futures ??

Systems	2012	2015
System Peak	10 -15 Pflop/sec	100 - 200 Pflop/sec
Power	8 - 12 MW	20 MW
System Memory	1.5 PBs	16-32 PBs
Node Compute	~500 Gflop/s	2 - 4 Tflops/s
Node Memory BW	100 GB/s	1 - 2 TB/s
Node Concurrency	64	100 - 300
Total Node Interconnect BW	10 GB/s	50-100 GB/s
System Size (Nodes)	20,000	100,000
Total Concurrency	500.000	2.000.000
Storage	25 - 40 PB	200 - 400 PB
Object Storage Servers	400 – 600	5000
I/O	1 TB/sec	15 TB/sec

Increasing resiliency throughout the system

- **Disks WILL fail and RAID becomes a problem**
  - New algorithms are required (ZFS, Btrfs, FhGFS, RAID-X ….)
- **Silent disk errors is still a problem**
  - Solution T10-PI
- **Automatic Backups and Snapshots are required**
  - Not currently Lustre features, but current roadmap is delivering …
- **Single layer interconnect fabric**
  - Multi-rail IB with full LNET support is required ….
  - Dynamic re-routing (the holy grail revisited ….)
- **End to end monitoring**
  - Not just from the storage point of view ….

xyratex

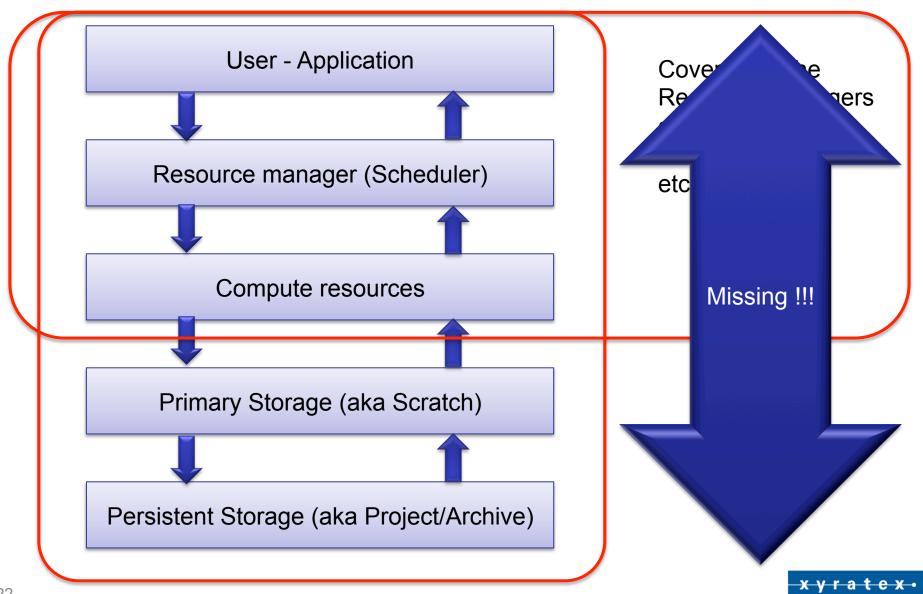# Adding more resiliency to the file system

- **T10-PI – End to End check summing in Lustre**

- **Additional checksum data described or carried in brw RPC**
- **Add PI and checking to data path**
- **For mmap'ed pages, early GRD failure implies data has changed, recompute from OSC**
- **Optional GRD checking on OSS can push all checksum load to HBA/disk hardware**
- **Disable bulk checksums**

User - Application

Resource manager (Scheduler)

Compute resources

Primary Storage (aka Scratch)

Persistent Storage (aka Project/Archive)

Cover... the
Re... ...rs
...
etc...

Missing !!!

xyratex

New times requires new tools …. (to mention a few …)

Tools to manage large installations need:

- **Instant feedback on issues and failures**

- **Instant help information for ANY issue**

- **Every function must be managed**
  - ➢ Individually
  - ➢ In groups
  - ➢ Manually
  - ➢ Automatically

- **Role based layouts**
  - ➢ Individual dashboard based on job description

- **Customizable arrangement of widgets**

- **Predictive maintenance**

- **Full inventory of current and replaced components**

- **Automatic support bundles**

- **Statistics and Analytics**

ClusterStor™ xyratex

xyratex

Summary and conclusions ….

- **Managing HPC storage is getting harder and harder …**
  - ➢ Current tools inadequate
  - ➢ The borders between multi-tiered storage are vanishing …
  - ➢ Scratch storage is not longer just scratch storage …

- **Higher levels of built in data integrity features are needed**
  - ➢ T10-PI for multilevel check summing
  - ➢ Data scrubbing and re-silvering of RAID systems
  - ➢ Data versioning of files and objects …

- **As a unified approach to storage is required, HPC storage needs to implement enterprise features.**
  - ➢ HSM
  - ➢ ILM toolkits
  - ➢ Snapshots and asynchronous backup for disaster recovery

- **Tools for predictive management and administrations**
  - ➢ Persistent database of FRUs, replacements, upgrades etc …

- **Lustre is a viable choice as the enterprise quality filesystem for weather, atmospheric and climate computational systems**

xyratex

Thank You - Questions?