



designed. engineered. results.

DMF Roadmap

Agenda

- Recent Feature Review
- Upcoming Features
- Roadmap

Recent Feature Review

Recent DMF Releases

- 4.0: Parallel DMF, Web Administration
- 4.1: Licensing (capacity, capability)
- 4.2: Scheduler and Audit enhancements
- 4.3: Monitoring, Configuration
- 4.4: SLES 10 SP3, LTO-5*, async ACSLS LCP*
- 5.0 SLES 11, Performance Improvements

* Delivered via patch

Parallel DMF (pDMF)

- ISSP 1.5 / DMF 4.0
- Dedicated data mover nodes
- Advantages:
 - Scale throughput to/from the tape drives
 - Provide redundancy
 - Lower cost
 - Smaller DMF server
 - Add data mover nodes as needed
- Details in upcoming session

DMF Web-based Administration

- Introduced in ISSP 1.5 / DMF 4.0
- View/Modify configuration
- Respond to DMF alerts
- Usage and throughput history
- Tape library management
 - Add/ejecting cartridges
 - verify data integrity
 - merge “sparse tapes”
 - Enable/disable tape drives

Tape ID	Volume Group	Data Left	Data Written	Hold Flags (hflags)	UP Age
000035	tapes	0.000000	0.000000	e-----	5h
000014	volume1	0.000000	167.202690	.f-----	1d
000015	volume1	167.202690	167.202690	-----	
000013	tapes	0.000000	0.000000	e-----	
000012	tapes	0.000000	0.000000	o-----	
LTO202	volume2	167.202690	167.202690	-----	
LTO201	tapes	0.000000	0.000000	-----	
LTO203	tapes	0.000000	0.000000	-----	
LTO204	tapes	0.000000	0.000000	-----	

Select	Time	Priority	Alert Message	Count
<input type="checkbox"/>	03:01:02	Warning	run_scan_logs: Log errors on thud	1
<input type="checkbox"/>	01:10:02	Info	run_remove_journals completed successfully. Deleted 2 journals	1
<input type="checkbox"/>	01:00:08	Info	run_remove_logs on thud completed successfully. deleted 0 logs, 0 transaction logs and 0 alerts	1
<input type="checkbox"/>	01:00:08	Info	run_remove_logs on thud completed successfully. deleted 8 logs, 0 transaction logs and 0 alerts	1
<input type="checkbox"/>	01:00:00	Info	run_tape_merge completed successfully.	2

DMF Licensing Change

- ISSP 1.6 / DMF 4.1
- Capacity licensing
 - 10TB, 100TB, 1PB
- Capability licenses
 - DMF server
 - Parallel Data Mover
 - High Availability
 - TMF/OV/APD:
 - Individual product licenses removed
 - Require a DMF server license

DMF I/O Scheduler

- ISSP 1.7 / DMF 4.2
- Dynamically adjusts to load and faults
 - Load balanced across data movers
 - Adjusts to changing load conditions
- Bandwidth-aware scheduling
 - avoids over-subscription of fibre channel
 - avoids over-subscription of mover nodes

DMF Database/Filesystem Audit

- ISSP 1.7 / DMF 4.2
- Improvements in:
 - Data-structures for lookups/sorting of records/inodes
 - Page cache hints to prevent page cache thrashing

Daemon Performance

- ISSP 2.0 / DMF 5.0
- Daemon request processing
 - largely single-threaded
 - built around a non-blocking model
 - slow/expensive work done in background threads
- Freeing of data blocks can be slow
 - Previously done in main daemon thread
 - Reduced daemon responsiveness
 - Could prevent data movers from getting enough work
- Data blocks now freed in background thread

Candidate Selection

- ISSP 2.0 / DMF 5.0
- Performance improvements in dmfsfree

Upcoming Features

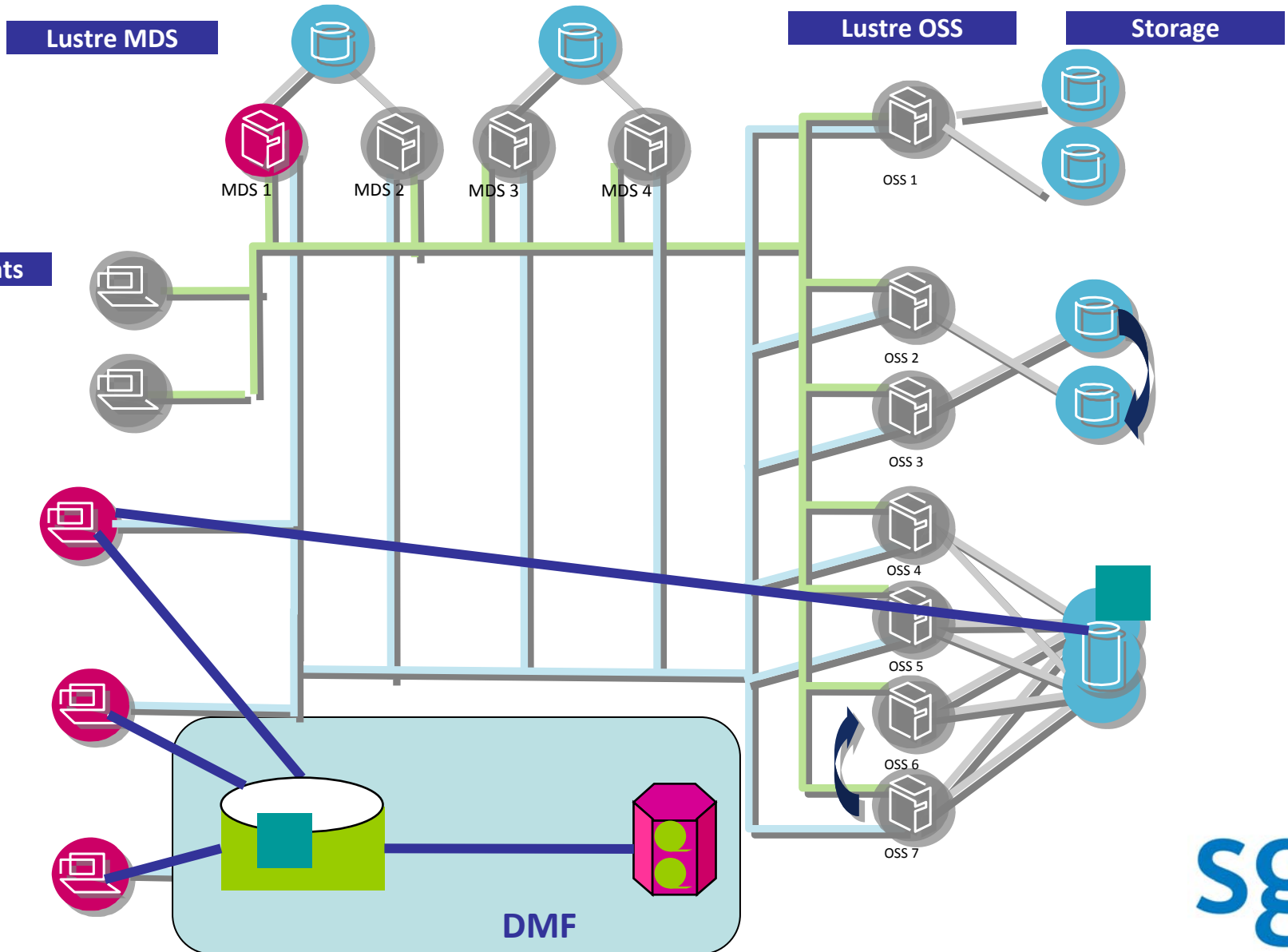
DMF + COPAN VTL

- Q2 2010 release
 - patch for ISSP 1.9 and 2.1
- COPAN MAID + VTL:
 - Massive array of idle disks
 - Dense, lower-power RAID
 - Tape library/drive virtualization software
 - DMF sees it as 1+ tape libraries
- DMF enhancements for:
 - Flexible migration policy
 - Optimized I/O
- Details in upcoming session

DMF as an Archive

- Q3 2010 / ISSP 2.2
- Primary work being done outside DMF
- DMF filesystems used as an archive
 - Files copied from primary filesystem to DMF
 - Migrated & freed immediately
 - Later recalled, copied to primary, freed again

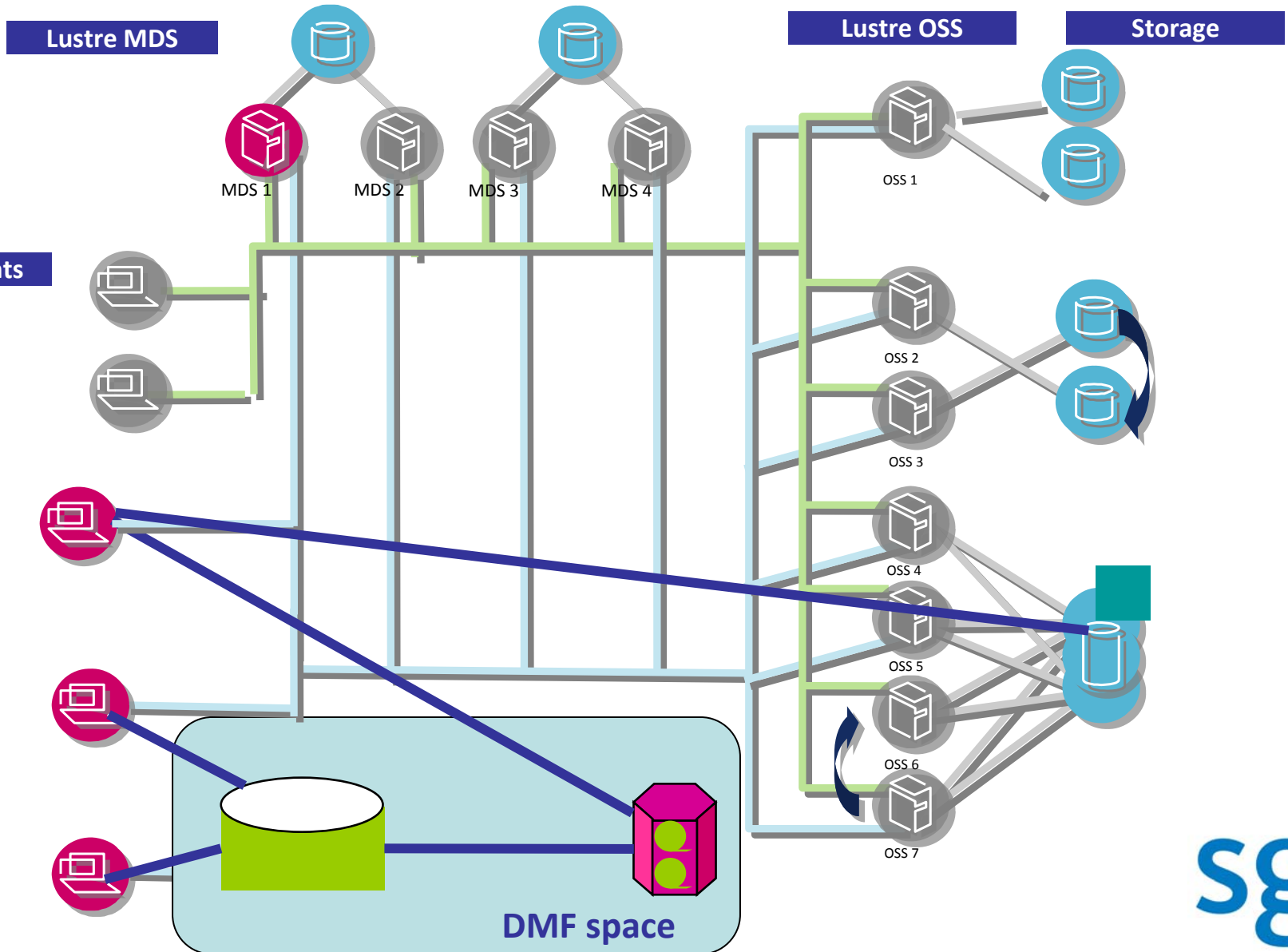
DMF as an Archive: Current Data Flow



DMF as an Archive

- Direct-to-tape
 - Data moved directly from primary filesystem to
 - Tape, Disk/FTP MSP, DCM
 - DMF filesystem used only as a namespace
 - Low capacity & bandwidth requirements
 - Primary filesystem can be any POSIX filesystem
- Direct-from-tape
 - dmcoppy can now copy to non-DMF filesystems

DMF as an Archive



Integration with Lustre HSM

- Availability: Lustre 2.1 (2011)
- CEA in France is developing an HSM interface for Lustre
 - Project endorsed officially by Oracle/SUN
- SGI working with CEA to integrate DMF
- Lustre HSM manages free space
 - Submits requests to a “copytool agent”
 - Offline files remain visible in Lustre (just like DMF today)
- DMF manages offline data
 - DMF provides a copytool agent implementation
 - Migrates / Recalls data on behalf of Lustre HSM
 - Data goes direct to/from tape
 - Parallel DMF compatible

DMF Roadmap

ISSP 2.1 – Q2 2010

- New OpenVault ACSLS LCP
 - Allows DMF to simultaneous control each robot
 - Substantially faster mounts
 - Also available in ISSP 1.9 patch
- SOAP Interface for User Commands
 - XML-based RPC transport over HTTP
 - Alternative to using C++ libdmfusr.so
 - Programming language / OS independent
 - Authentication through Apache
 - A limited set of calls at this time:
 - dmattr
 - dmget
 - dmput
 - dmtag
 - API not locked down yet -- subject to change

ISSP 2.1 (continued)

- DMF High-Availability for SLES11
- Async Recall Option
 - default dmget is to block until data is recalled from tape, this could take hours to complete
 - new 'no wait' option which queues the request and returns

ISSP 2.1 (continued)

- New library support
 - SpectraLogic T-Finity (T950E)
- New tape drive support
 - LTO-5: 175 MB/s | 1.5 TB cartridges
- New DMF clients
 - SLES11 SP1
 - RHEL 5.5

ISSP 2.2 – Q3 2010

- SLES11 SP1 support
- Enhanced Performance Tracking
 - Identify performance bottlenecks
 - PCP-based metrics and graphs
 - quickly identify existence of a problem
 - look at history of performance
 - look at data in different ways to isolate problem
- Archive (direct-to-tape)

ISSP 2.2 (continued)

- DMF/COPAN VTL Integration
- DMF supported on non-SGI hardware
- Option to keep tape mounted for write to archive workloads, few reads
- RHEL 6 client

- DMF Lustre HSM integration
 - Tentative: dependent on Lustre 2.1 release
- OpenVault configuration
 - dynamic tape path discovery
 - DMF Manager (Web Interface) for configuration
 - Central OV Configuration for pDMF

ISSP 2.3 (continued)

- Async Control of SpectraLogic Robots
 - Openvault will be able to get multiple robot arms to concurrently move cartridges in a single library partition
 - Tentative: based on SpectraLogic functionality

- DMF Database Replacement
 - Faster DMF DB Maintenance
 - Scale to billions of records
 - Flexible
 - Add new tables/fields without extended downtime during a DMF upgrade
 - Many desired features require new DB
 - Daemon request queue management
 - Multiple files with same bfid
 - End-to-end verification

DMF Futures

- Continued focus on
 - Scalability improvements
 - Ease-of-use
- Parallel Disk MSP / DCM
 - Today's pDMF limited to tapes
- Technology Upgrades

Technology Upgrades

- Problem:
 - Moving data from old technology to new
 - Time and bandwidth requirements to move data
- Proposal:
 - Provide tools to make the task easier
 - Provide better control or use of resources necessary to perform technology upgrades
 - Prioritizing workload with normal workload

Non Plan of Record Ideas

- End-to-end data verification
- Data deduplication
- Load-balanced recall
- DMF Path Migration Policy
 - Requires XFS Parent Inode Pointers
- Native COPAN Integration



www.sgi.com

©2002 SGI. All rights reserved. SGI, IRIX, and the SGI logo are registered trademarks and SGI SAN Server, XFS, and CXFS are trademarks of Silicon Graphics, Inc., in the U.S. and/or other countries worldwide. MIPS is a registered trademark of MIPS Technologies, Inc., used under license by Silicon Graphics, Inc. UNIX is a registered trademark of The Open Group in the U.S. and other countries. Intel and Itanium are registered trademarks of Intel Corporation. Linux is a registered trademark of Linus Torvalds. Windows is a registered trademark or trademark of Microsoft Corporation in the United States and/or other countries. All other trademarks are the property of their respective owners.

(10/02)