Scaling capacity and performance without compromise using SGI® DMF™

Capacity, Performance & Reliability
Agenda

- Hierarchical Storage Management
- Lustre Scalability with DMF (HSM)
- Tiered Data Management
- DMF – Start small and grow
- DMF Direct Archiving
- JBFS Fast Mount Cache
- Summary
HSM | Data Migration Facility (DMF)

- **Data life cycle management**
  - DMF manages the placement of data within multiple tiers of storage

- **Automated data migration**
  - From expensive, production disk to 2nd or 3rd tier storage

- **Transparent to user**
  - All data appears on line all the time

- **Key Benefits**
  - DMF reduces tier 1 disk investment
  - DMF reduces power consumption
  - DMF protects data long term

- **SGI® DMF™ 25 years** in production

---

Hierarchical Storage Management

Transparently migrate data to Tape, MAID or Cloud

- Lustre*
- Cloud

* = Some names and brands may be claimed as the property of others
Scalability without compromise

Capacity, Performance & Reliability

Lustre* Filesystem

DMF Managed HSM Environment

DMF Direct Archiving

Scale Here

Scale Here

Scale Here

©2014 SGI

SGI Company Proprietary

* = Some names and brands may be claimed as the property of others
Seamless Tiered Data Management

The most recent and active data is “live” in Lustre* and mirrored within DMF. ALL DATA APPEARS ONLINE to users.

“Overflow” data is stored and protected within DMF on various cost-correct media types.
Tiered Data Management

HSM perspective: regular file
User perspective: online file

Before migrating

Managed filesystem tier-1

Lustre*

Destination: Lustre*
high-performance disk

inode
data

Fast-mount cache tier-2

Destination: DMF
JBFS

Secondary storage tier-3

Destination: DMF
cloud/FTP/tape

DMF

* = Some names and brands may be claimed as the property of others
Tiered Data Management

HSM perspective: dual-state file

User perspective: online file

After migrating
Tiered Data Management

HSM perspective: **offline** file
User perspective: **online** file

After freeing space

![Managed filesystem diagram]

- **Managed filesystem tier-1**
  - Destination: Lustre*
  - high-performance disk
  - inode

- **Fast-mount cache tier-2**
  - Destination: DMF
  - JBFS

- **Secondary storage tier-3**
  - Destination: DMF
  - cloud/FTP/tape
  - data

* Some names and brands may be claimed as the property of others
Tiered Data Management

HSM perspective: **unmigrating** file
User perspective: **online** file

Recalling file data from cache
**Tiered Data Management**

HSM perspective: **unmigrating** file

User perspective: **online** file

Recalling file data from cache
DMF Evolution
Start small and grow
DMF Direct Archiving

DMF starts small and grows with you...

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS

Remote disk target via DMF, FTP, NFS, Cloud

High Availability

Lustre* HSM with DMF, direct to Tier 2/3

Lustre* OSS/OST Building Block

Lustre* MDS/MDT Building Block

NFS/CIFS
DMF Direct Archiving | Data Flow

DMF Servers

Lustre* Clients

Lustre* MDS

DMF Managed Data

DMF Data Mover

DMF MDS 1

DMF MDS 2

Lustre* OSS

Storage

Primary Storage

Metadata

Data

Logical Path

Physical Path

* = Some names and brands may be claimed as the property of others
Lustre* HSM | Communication & Data Flow

* = Some names and brands may be claimed as the property of others

Parallel Data Mover Option
- Data migration from multiple parallel servers
- Scales I/O performance
- Add Additional data movers as required
JBFS | The OpenVault VTL for DMF

- JBFS is an acronym for JBOD File System
- JBFS provides mounting services
  - Serialised access to disk media
  - Independent from Linux disk mounts and file systems
- Disks treated like tapes mounted in tape drives
- The primary advantages
  - Mount performance
  - Low-cost scalable data throughput performance
  - Power management via ZeroWatt™
## JBFS | Why A New File System?

<table>
<thead>
<tr>
<th>XFS (or any typical file system)</th>
<th>DMF Preferred</th>
<th>JBFS Provides</th>
</tr>
</thead>
<tbody>
<tr>
<td>A large number of objects</td>
<td>Small number of objects</td>
<td>✓ Small number of objects</td>
</tr>
<tr>
<td>Object sizes change</td>
<td>Object sizes fixed</td>
<td>✓ Object sizes fixed</td>
</tr>
<tr>
<td>Flexible object organisation</td>
<td>Fixed object organisation</td>
<td>✓ Fixed object organisation</td>
</tr>
<tr>
<td>Primarily random access</td>
<td>Primarily sequential access</td>
<td>✓ Primarily sequential access</td>
</tr>
<tr>
<td>Bursty access</td>
<td>Sustained access</td>
<td>✓ Sustained access</td>
</tr>
<tr>
<td>Mount/dismount infrequently</td>
<td>Mount/dismount frequently</td>
<td>✓ Mount/dismount frequently</td>
</tr>
</tbody>
</table>
JBFS | Additional Benefits

- Recoverability
- Data Assurance
- High Performance
- Flexibility
- Power Management with Zero-Watt™
- JBFS API (Same as SGI Copan)

OpenVault includes a new DCP and a new LCP to manage JBFS volumes
Preparing a disk device for use with JBFS consists of three basic steps:

1. Apply the GPT labels
2. Apply the XVM labels
3. Apply the JBFS format

XVM volume name must be "JBFS_{lib}_{PCL}"
- JBFS – Fixed
- {lib} – OpenVault Library Name
- {PCL} – unique 6-character value [0-9A-Z]
Lustre* Native Clients

Lustre* Filesystem
- Lustre* OSS/OST Building Block
- Lustre* OSS/OST Building Block
- Lustre* MDS/MDT Building Block
- Lustre* OSS/OST Building Block

Lustre* Network
- pDMF Data Mover
- pDMF Data Mover
- pDMF Data Mover
- pDMF Data Mover

FC Switch
- TAPE
- TAPE
- TAPE
- TAPE

SAS Switch
- JBFS
- JBFS
- JBFS
- JBFS

* = Some names and brands may be claimed as the property of others
Summary and Key Points

- New Lustre* and DMF features allow cost effective scalability without compromising performance
- SGI DMF provides a high performance parallel HSM for Lustre* with direct archiving to tier 2/3 storage targets
- SGI DMF – JBFS delivers a tier 2 fast mount cache with built in power management$ capabilities
- The Result:
  - Cost effective capacity, reduced TCO (low cost/power storage tiers)
  - Proven long-term data protection (DMF – 25 years in production)
  - Improved operational procedures (simplified access to data)
  - Scalable performance within archive tiers (parallel DMF)

$ = on supported hardware

* = Some names and brands may be claimed as the property of others
Questions & Responses

Robert Mollard
Senior Storage Specialist, Asia Pacific
rmollard@sgi.com