Bielefeld University
Center for Biotechnology (CeBiTec)

Hierarchical Storage Management with SAM-FS

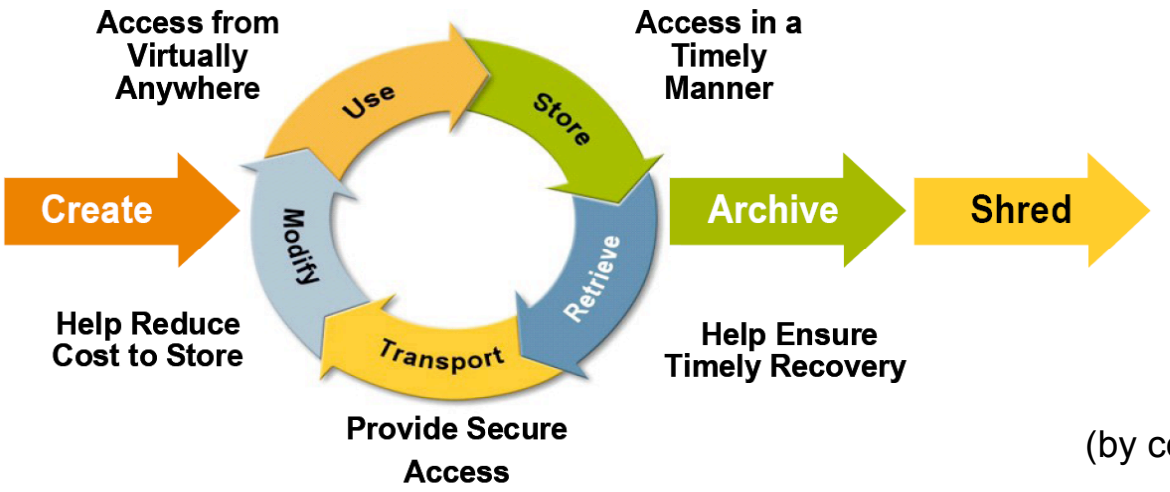
Achim Neumann
CeBiTec / Bielefeld University

Agenda

- Information Lifecycle Management
- Backup – Technology of the past
- Filesystems, SAM-FS, QFS and SAM-QFS
- Main Functions of SAM
- SAM and Backup
- Rating, Perspective and Literature

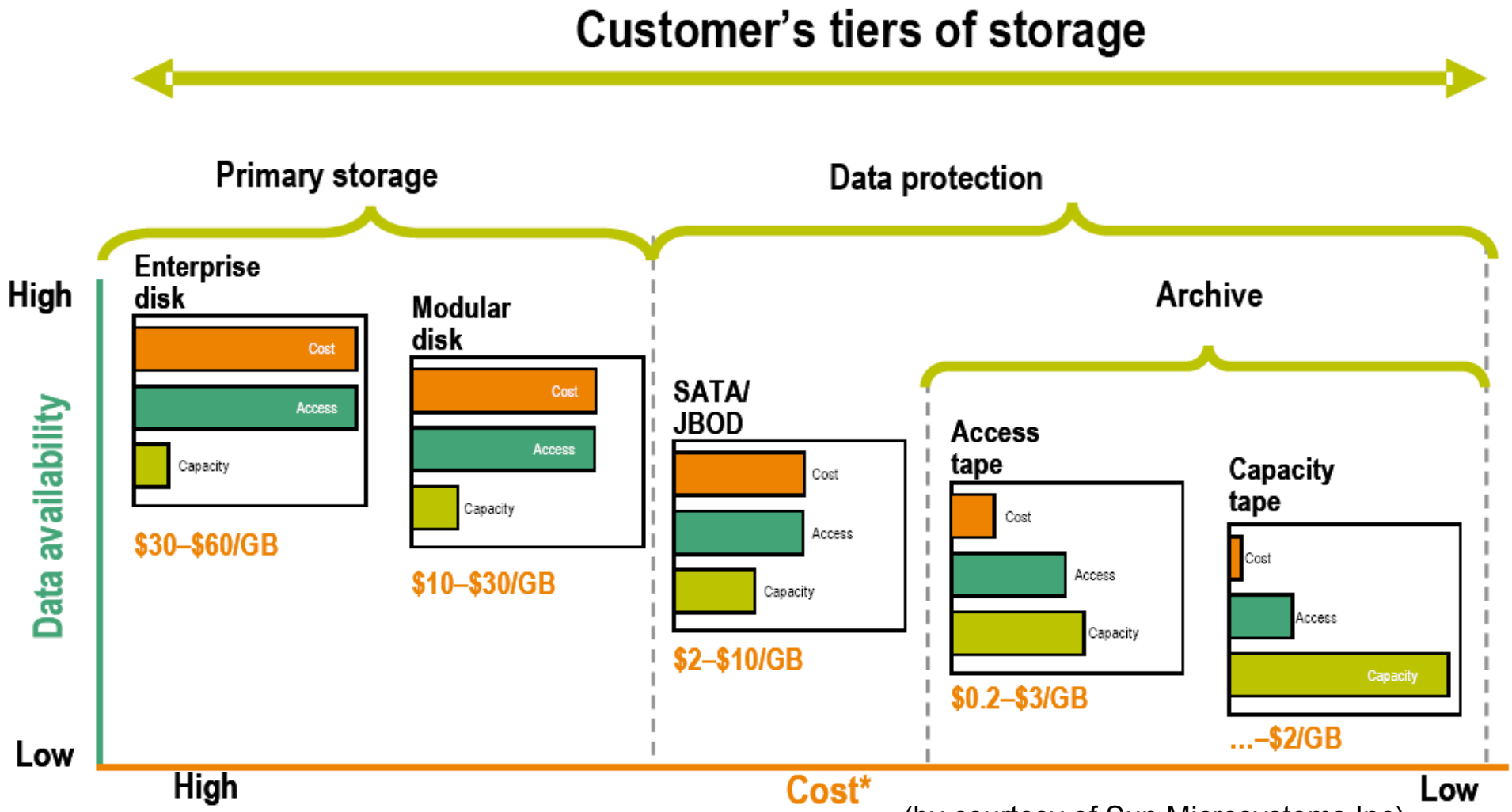
ILM – Information Lifecylce Management

- 5 Exabyte data per year (92 % digital, 8 % analog)
- 20% fixed media (disk), 80% removable media
- Managing data from Creation to Cremation (and everything that is between



(by courtesy of Sun Microsystems Inc).

Different classes of Storagesystems



(by courtesy of Sun Microsystems Inc).

Backup – Technology of the past

- Backup – protect data from data loss
- Backup is not the problem
- Ability to restore data when needed
  - Disaster
  - Hardware failure
  - User error
  - Old versions from a file

Backup – Technology of the past (cont.)

- Triple data in 2 – 3 years
- Smaller backup window
- ☞ more data in less time

#TB	Stream for 24 h	# Drives for 24 h	# Drives for 6 h
1	12,1 MB/s	2	5
10	121 MB/s	13	49
30	364 MB/s	37	146

Backup – Technology of the past (cont.)

- Limited number of tape drives in library
- Limited number of I/O Channels in server
- Limited scalability of backup software
- Inability to keep all tape drives streaming at one time

- At least every full backup has a copy of the same file
- It's not unusual to have 8 or 10 copies of the same file

☞ excessive media usage and costs

HSM – first (nice) try

- Data mover and Archiver
- Low and high watermark
- Archiving not until high watermark reached
  ☞ Showstopper
- Releasing until low watermark reached

HSM now

- Data mover on policy based rules (defined by an admin)
- Archiving (on WORM media)
- Different tiers of storage
  - Realtime
  - Online
  - Nearline
  - Offline
- Transparent for user and applications (not always)

Filesystems

- Stores named data sets with attributes (metadata)
- Local (UFS, VxFS, ZFS, Sam-FS, QFS (local))
- Shared (NFS, CIFS, PXFS, QFS (shared))
- Special (tmpfs, cachefs)
- Media-specific (UDF, ISO9660, PCFS (diskettes))
- Pseudo (procfs, devfs, xmemfs)

Sam-FS, QFS, Sam-QFS

- Designed and developed by LSC (midst of the 90's)
- OEM Product by StorageTek HFS
- Acquired by SUN (2002)
- Sun StorEdge Utilization Suite / Performance Suite
- Sun StorEdge SAM-FS / QFS
- Soon: Sun StorageTek SAM-FS / QFS

Sam-FS, QFS, Sam-QFS (cont.)

- Two disk filesystems:
  - Basic filesystem FS
  - High performance filesystem QFS
- A Storage Archive Management Component (SAM) can be combined with FS and QFS
- FS without SAM is unsupported
- Common kernel module: samfs

Basic and high performance filesystem

- Simplified meta data
  - Superblocks (at the beginning of the disk)
    - File system type (basic or high performance)
    - Hardware devices
    - DAU
    - shared
  - Inodes kept in a file .inodes (512 byte / inode)
  - Configurable DAU (Disk Allocation Unit)
  - Disk space usage (+ other meta data) in primary superblock

Basic and high performance filesystem (cont.)

- QFS
  - Separate data and meta data on different devices, only one superblock on QFS data disk
  - Parallel access on data and metadata
  - Optimal performance:
    - Metadata: mirrored solid state disks
    - Data: Hardware RAID 5
  - single writer, multiple reader (data and meta data access)
  - shared writer (metadata access only on master)
    data devices are architecture independent (x86 and sparc)

Basic and high performance filesystem (cont.)

- Disks can be Veritas Volumes or SVM Meta devices
- Integrated (wanna-be) Volumemanager
  - Striping
  - round robin
- **stripe width**
  - multiplier of DAU (the data, written to disk device (chunk))
  - is set automatically to a chunk size of 128 KB
- **disk groups**

Performance issues

- Read-Modify-Write on RAID 5 devices
  - Raid 5 device with 5 disks and 128 KB stripe size
  - optimal chunk: 4 x 128 KB = 512 KB
  - 512 KB / 64 KB (default DAU) = 8 (stripe width)
- EFI-Labels
  - the first 34 sectors are reserved
  - leave the first 512 KB (1024 sectors) free

Basic and high performance filesystem (cont.)

- Equipment Type Identifier for filesystems and disks
  - **ms**: basic file system
  - **ma**: high performance filesystem
  - **mm**: metadata device for QFS
  - **mr**: data device for QFS (single DAU)
  - **md**: meta data and data device for basic filesystem or data device for QFS (dual DAU)
- Dual Allocation Scheme
  - first 32 KB are written to 4 KB fragments and then DAU

Storage Archive Manager - Main Functions

- Archive: Transparent copy files to archive media
- Release: Disk Space Management, transform files to stub files on demand
- Stage: Restore archived files to disk cache on access
- Recycle: Consolidate tapes for reusing

- Daemons and configuration files for every function
- kernel module **samst** for robotic

SAM Daemons

Daemon	Started by:	Starts	Major Tasks
sam-fsd	init	sam-archiverd, sam-releaserd, sam-stagerd, sam-stagealld	starting and configuring storage management daemons
sam-archiverd	sam-fsd	sam-arfind, sam-arcopy	managing archiving, controlling archiver.cmd
sam-arfind	sam-archiverd		identifying files for archiving
sam-arcopy	sam-archiverd		copying files to archive media
sam-releaserd	sam-fsd		releases files
sam-stagerd	sam-fsd		stages individual files
sam-stagealld	sam-fsd		stages groups of files
sam-amld	sam-fsd	various daemons	controls removable storage media daemons

Configuration files

File	Location	Description
mcf	/etc/opt/SUNWsamfs/	Master configuration, Filesystems
defaults.conf	/etc/opt/SUNWsamfs/	logging, delays,
archiver.cmd	/etc/opt/SUNWsamfs/	Archive sets, global and filesystem archive attributes
releaser.cmd	/etc/opt/SUNWsamfs/	Releasing attributes
stager.cmd	/etc/opt/SUNWsamfs/	Stager attributes
recycler.cmd	/etc/opt/SUNWsamfs/	Recycler attributes
inquiry.conf	/etc/opt/SUNWsamfs/	Hardware recognition (Robotic, media drives)
vfstab	/etc/	Mount Options (stripe width)
system	/etc/	max_phys
st.conf	/kernel/drv	tape properties
(s)sd.conf	/kernel/drv	(s)sd_max_xfer_size

Logfiles

- archiver.log
- releaser.log
- recycler.log
- catalog (per library)
- **historian** (virtual catalog for exported media)
- dumps (**samfsdump**)

Archiving

- default: create one copy on archive media
- archive sets
  - global vs. per filesystem
  - admin defined attributes (for files): size, time, name, location
  - special archive sets: **no_archive**, **allsets**
  - default archive set for directories and not defined files
  - associative Archiving
  - up to 4 copies (different media, location and times)
    - copy on different conditions (time, size)

Archiving (cont.)

- file segmentation
  - parallel streaming
  - transparent for users and applications
  - only modified segmented parts will be archived again
- direct access on archive media (**request**)
- **no_archive** attribute on individual files or directories
- *scan mode* vs. *continuous archiving*

Releasing

- **min_residence_time**
- **archive_done** flag
- **no_release** flag
- high and low watermark
- partitial releasing (stub files, min 8kb, max 16kb)
- size and age attributes
- weight is configurable

Staging

- restore to disk cache
- configuration directives also in archiver.cmd
- stage on access
- associative staging
- **no_stage** flag, direct access to archive media

Recycle

- *stale* and *expired*
- no daemon, manually or by cron
- high watermark (default 95%)
- **minimum_gain** of expired space for a VSN: 50%
  data_quantity < 1 GB
- Consolidate tapes
- retention period
- **no_recycle**, **-ignore**, **-recycle_ignore**

Backup

- Disaster recovery
  - **samfsdump** and **samfsrestore**
  - meta data is enough to mount the filesystem
  - script to restore files again
  - restore only on SAM-FS or SAM-QFS
- individual restore
  - archiver log files
  - **request** and **star**
- tar format on archive media

Integration of Backup Software

- Backup on SAM-FS
  - daily / incremental backups only on disk (**no_archive**)
  - only full or weekly backups on archive media
  - **no_stage** flag
  - only Sam-FS control the library hardware
  - no partitioning of library hardware
  - save licence costs for backup software

(Dis-)Advantages

- Advantages
  - preferences
  - transparent for users and applications
  - API (CMS / SAP)
- Disadvantages
  - no online tuning
  - configuration complexity
  - no backup/restore interface / Third Party
    - RestoreME (Synstar)
    - SAM-FS Recovery Tools (it4future)

Perspective

- versioning for backup / restore
- (SAM integration in ZFS)

Literature

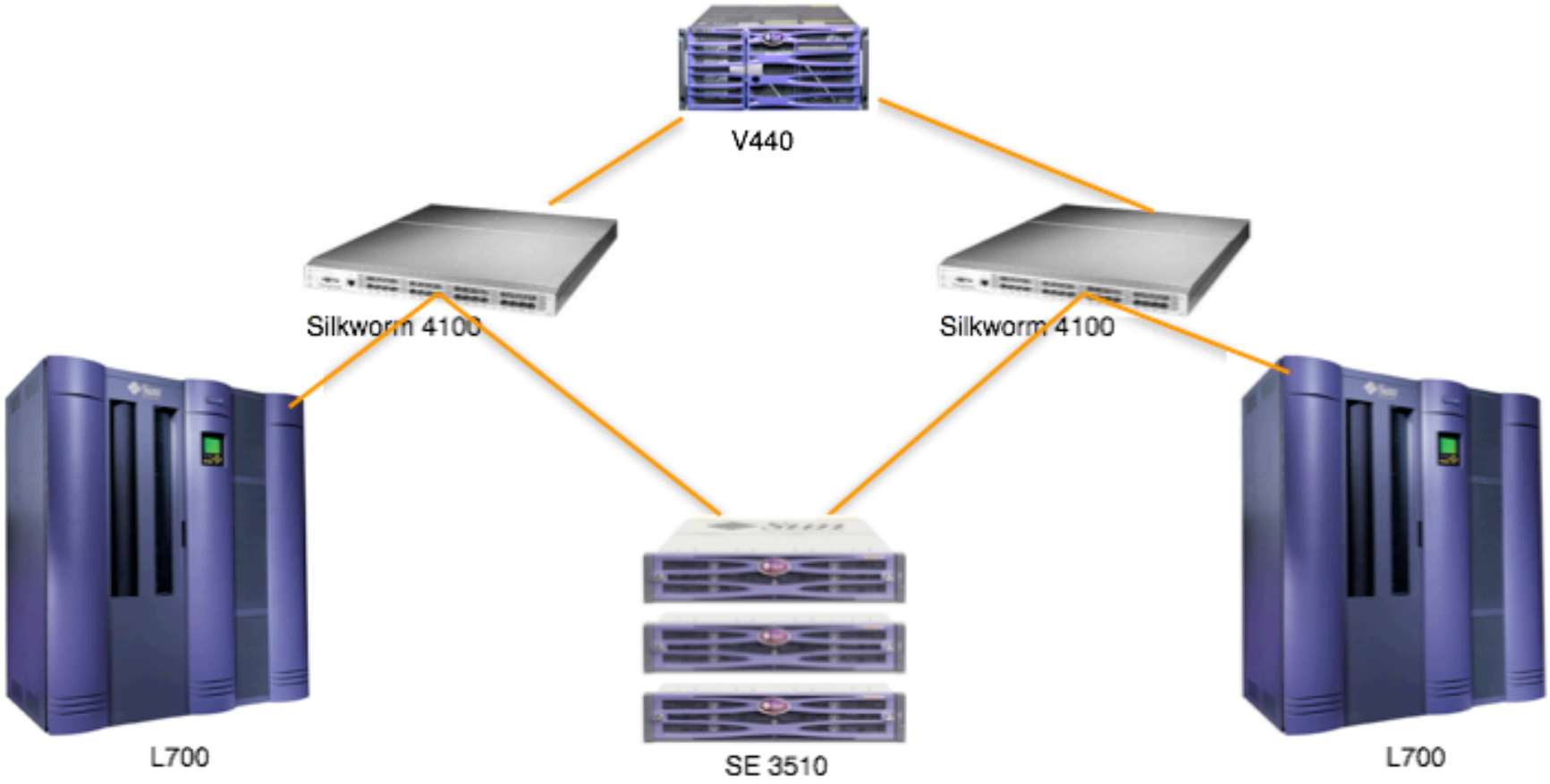
- Sun StorEdge SAM-FS 4.5 Documentation
- http://www.geocities.com/lynchmaryann/
- sam-managers Mailinglist Archive
- Design, Features, and Applicability of Solaris FileSystems (SUN BluePrints)
- SUN StorageTek White Papers

Thank you

Achim Neumann

<an@CeBiTec.Uni-Bielefeld.DE

CeBiTec (Test-)Environment



```
#
# smitty Sun StorEegde 3510 in V0-200S (LD 0 und 1)
#
GenoMik-Data                    10   ma   GenoMik-Data
/dev/dsk/c7t600C0FF000000000085A5875D2DA1A00d0s0  11  mm  GenoMik-Data  -
/dev/dsk/c7t600C0FF000000000085A5803A11CBD00d0s0  12  md  GenoMik-Data  -
#
# oook Sun StorEdge L700 in V0-200S
#
/dev/samst/c3t500104F000615E93u0       20  rb  eeek
/dev/rmt/4cbn                  22   tp   eeek   on
/dev/rmt/5cbn                  23   tp   eeek   on
/dev/rmt/6cbn                  24   tp   eeek   on
/dev/rmt/7cbn                  25   tp   eeek   on
#
#  /dev/samst/c3t500104F000615E93u0      20  s9  eeek
#   /dev/rmt/4cbn                 22   li   eeek   on
#   /dev/rmt/5cbn                 23   li   eeek   on
#   /dev/rmt/6cbn                 24   li   eeek   on
#   /dev/rmt/7cbn                 25   li   eeek   on
#
# eeek Sun StorEdge L700 in W01-211
#
/dev/samst/c4t500104F000615DD6u0       30  rb  oook
/dev/rmt/0cbn                  31   tp   oook   on
/dev/rmt/1cbn                  32   tp   oook   on
/dev/rmt/2cbn                  33   tp   oook   on
/dev/rmt/3cbn                  34   tp   oook   on
#
#  /dev/samst/c4t500104F000615DD6u0      30  s9  oook
#   /dev/rmt/0cbn                 31   li   oook   on
#   /dev/rmt/1cbn                 32   li   oook   on
#   /dev/rmt/2cbn                 33   li   oook   on
#   /dev/rmt/3cbn                 34   li   oook   on
```