Storage Tutorial For Content Creation and Distribution

Tom Coughlin
Coughlin Associates
www.tomcoughlin.com
Outline

• Content Value Chain
• Storage Demand for Entertainment Applications
• Storage Devices
• Storage Systems
• Digital Storage Applications for Entertainment Media
• Conclusions
STORAGE MAKES ME HAPPY!
Digital Content Value Chain

- Content Creation
- Content Archiving
- Content Distribution
- Content Reception
- Content Editing

- PVR/DVR/set-tops
- Game Machines
- Mobile Devices
- Streaming Media
- VOD
- PPV
- Tape
- ATA Disk Arrays
- Optical Jukeboxes
- Cameras
- Animation
- Field Editing
- Studio Editing
- Special Effects

© 2004 Coughlin Associates
Activities in Content Creation, Editing, Archiving and Distribution

**Content Creation**
- Acquire Content
- Field Editing
- Ingest
- Digitization
- Capturing Metadata
- Auto-Logging
- Proxy Creation

**Content Management**
- Library and Archive Storage
- Search Tools
- Edit and Re-Purpose material
- Metadata Management
- Rights Management
- Distribution Scheduling
- Content Import/Export
- Web access

**Content Distribution**
- Various Delivery Networks
- Point to Point & Point to Multipoint
- Push and Pull modes
- Real-time and Scheduled
- Dedicated Infrastructure
- Format/Destination Selection
- Caching and Edge Storage

© 2004 Coughlin Associates
## Uncompressed Video Production Storage Needs (Raw DPX 10 bit log files)

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Frames/sec</th>
<th>MB/second</th>
<th>Capacity/minute (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>1.7</td>
<td>38.4</td>
<td>2.3</td>
</tr>
<tr>
<td>1K</td>
<td>3.2</td>
<td>76.8</td>
<td>4.6</td>
</tr>
<tr>
<td>HD</td>
<td>8.2</td>
<td>197</td>
<td>11.8</td>
</tr>
<tr>
<td>2K</td>
<td>12.5</td>
<td>300</td>
<td>18.0</td>
</tr>
<tr>
<td>4K</td>
<td>50</td>
<td>1.2</td>
<td>72.0</td>
</tr>
</tbody>
</table>
Digital Production and Distribution Rules!

• Save more than a factor of 50 in video capture and editing costs vs. traditional film
• Special effects and editing possible with digital production can’t be matched with older analog techniques
• Save 80% on digital theatre distribution vs. film distribution
Adventures in Archiving

- Demand huge, and growing
- Long term storage formats
- Format obsolescence
- Need for format transfer planning—archiving will not be merely static
- Enormous need for good metadata tagging and data search and access improvements
Storage Devices
### Storage Hierarchy

<table>
<thead>
<tr>
<th>Positioning Parameters</th>
<th>Capacity Limit</th>
<th>Access Time/Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Price/GB</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>$100K–300K</td>
<td>±200 GB/sec</td>
</tr>
<tr>
<td>Immediate</td>
<td>&lt; $75K</td>
<td>20–40 ns</td>
</tr>
<tr>
<td>Online</td>
<td>$5K–10K</td>
<td>&lt; 0.1 ms</td>
</tr>
<tr>
<td>Disk Arrays</td>
<td>$40–70</td>
<td>4–10 ms</td>
</tr>
<tr>
<td>Solid-State Disk</td>
<td>$30–40</td>
<td>10–25 ms</td>
</tr>
<tr>
<td>Performance Disk</td>
<td>$2–20</td>
<td>&lt; 250 ms</td>
</tr>
<tr>
<td>High Capacity Disk/ATA Disk</td>
<td>$2–10</td>
<td>5–10 sec</td>
</tr>
<tr>
<td>Nearline Disk, MAID</td>
<td>$20–$4*</td>
<td>5–10 sec</td>
</tr>
<tr>
<td>Virtual Tape</td>
<td>&lt; $0.001*</td>
<td>min, hr, day</td>
</tr>
<tr>
<td>Nearline Tape</td>
<td>PB–EB</td>
<td>min, hr, day</td>
</tr>
<tr>
<td>Far-line Shelved Data/Archival Storage</td>
<td>+$20</td>
<td>min, hr, day</td>
</tr>
<tr>
<td>Analog Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Micrographics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Digital Images</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paper</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Based on recording technology

Source: Horison Information Strategies

© 2004 Coughlin Associates
Magnetic Rigid Disk Drives (HDD)

- Spindle Motor
- Disk
- Head Actuator
- Head Suspension
- 15k RPM FC Drive
- High Capacity ATA Drives (now up to 400 GB)
- 2.5 Inch Mobile Drive
- Toshiba 0.85 inch HDD

© 2004 Coughlin Associates
HDD Areal and Volumetric Density Growth

Storage areal density CGR starts to slow from 100% per year near 100 Gbit/in². Volumetric density follows at similar rate.

Progress begins to slow down due to technological challenges.

Lab demos (year 2002) at 130 Gbit/in²

Storage volumetric density has improved based on increased areal density, smaller form factors and closer packing of disks.

© 2004 Coughlin Associates

From Clod Barrara, IBM
March 2004
# SHIPPING PRODUCT DISK CAPACITY PROJECTIONS

<table>
<thead>
<tr>
<th>Year</th>
<th>95mm Mainstream Capacity Per Platter</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>40</td>
</tr>
<tr>
<td>2003</td>
<td>80</td>
</tr>
<tr>
<td>2004</td>
<td>120</td>
</tr>
<tr>
<td>2005</td>
<td>180</td>
</tr>
<tr>
<td>2006</td>
<td>270</td>
</tr>
</tbody>
</table>

By 2006 we could have four disk 3.5-inch disk SATA drives with storage capacities of over 1 TB.

© 2004 Coughlin Associates
The Universe Still Beats Us by Far in Information Capacity

- The holographic and universal information bounds are far beyond the data storage capacities of any current technology!
- Magnetic recording technology may allow up to 50 Tbpsi (50 X $10^{18}$ bpsi)

Source: Information in the Holographic Universe, August 2003 Scientific American

© 2004 Coughlin Associates
HDD Access Density

Desktop and Server Drive Performance

Sustained HDD I/O Rates per GByte

Year

© 2004 Coughlin Associates

From Clod Barrara, IBM
March 2004
HDD Reliability Trends

- MTBF/GB falling – drive rebuild times growing
- Multi-parity RAID a required aggregation technology

Mean-Time-To-Failure/GB of Storage

HDD MTBF Manufacturer Specifications

From Clod Barrara, IBM
March 2004

© 2004 Coughlin Associates
Popular Digital Tape Formats
(All ½ inch tape cartridge technologies)

- SAIT S-DLT
- LTO

Tape with Tape Drive

- Tape is still digital archive media of choice
- Tape data access is on the order of minutes vs. milliseconds or seconds for disk
- Tape media costs have been somewhat underwritten by VCR tape production, implications for future of tape costs
- ½ inch tape capacities of up to 10 TB projected

© 2004 Coughlin Associates
Active tape format CAGRs are about 40%. Disk Drive CAGRs are expected to be ~60%
Blue Ray Optical Disks and Drive


MultiMedia Object Size/Bandwidth

Holographic Disks

Source: Telcordia 3/03

© 2004 Coughlin Associates
Storage Systems
Digital Content Lifecycle in Production and Distribution.

<table>
<thead>
<tr>
<th>Content Stage</th>
<th>Frequent Changes</th>
<th>Frequent Accesses to Fixed Content</th>
<th>In-Frequent Accesses to Fixed Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production</td>
<td>Non-linear editing</td>
<td>Production Viewing</td>
<td>Archiving</td>
</tr>
<tr>
<td>Distribution</td>
<td>N/A</td>
<td>Distribution Viewing, PVR/DVR</td>
<td>DVDs, VHS, Content Downloads</td>
</tr>
</tbody>
</table>

© 2004 Coughlin Associates
SGI InfiniteStorage DMF
Data Life Cycle Management

Primary Storage
Online - high-performance disk

DMF manages data based on:
• age of file
• size of file
• type of file

Nearline Disk
High Capacity, Low cost, Lower performance

Promote
used last 24 hrs

Promote
used last 7 days

Demote
> 7 days < 365

Tape Libraries
Higher capacity, lower cost

Demote
> 1 Yr < 2 Yr

Archive
> 2 Yr

Gabrielle Broner, SGI,
Jan. 2004
© 2004 Coughlin Associates
Comparative Prices of Storage Systems

Performance Disk
Midrange Disk
Capacity Disk
Automated Tape

© 2004 Coughlin Associates
Comparison of Tape and ATA Disk Storage Economics

- Tape Drives
- Tape Drive + 100 Media
- Ghetto RAID
- Tape Media
- IDE Drives
- Sony SAIT Projection

© 2004 Coughlin Associates
RAID subsystem is a composite device made with many smaller devices.
RAID Systems

Volume Manager with RAID Capability or Host I/O Controller RAID Capability

Disk Drive
Disk Drive
Disk Drive

JBOD with each device addressed individually by host-resident RAID

RAID subsystem acting as a single virtual device

© 2004 Coughlin Associates
RAID Advantages

• Can allow for more reliable data and/or improved system performance
• A RAID requires fewer host I/O controller slots. Also a RAID can use a single network (e.g. SCSI) address rather than individual addresses for each drive
• By creating a virtual drive with 1 file system there is no need of the host to manage separate file systems on the individual drives

© 2004 Coughlin Associates
# Characteristics of RAID Levels

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>RAID 0</th>
<th>RAID 1</th>
<th>RAID 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usable disk space</td>
<td>100%</td>
<td>50%</td>
<td>67-93%</td>
</tr>
<tr>
<td>Parity and Redundancy</td>
<td>None</td>
<td>Duplicate data</td>
<td>Parity distributed over each drive</td>
</tr>
<tr>
<td>Minimum number of disks</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I/Os per Read</td>
<td>1 Read</td>
<td>1 Read</td>
<td>1 Read</td>
</tr>
<tr>
<td>I/Os per Write</td>
<td>1 Write</td>
<td>2 Write</td>
<td>2 Reads + 2 Writes</td>
</tr>
<tr>
<td>Performance</td>
<td>Best</td>
<td>Good</td>
<td>Worst</td>
</tr>
<tr>
<td>Fault Tolerance</td>
<td>Worst</td>
<td>Best</td>
<td>Good</td>
</tr>
<tr>
<td>Cost</td>
<td>Best</td>
<td>Worst</td>
<td>Good</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Best over all performance, but data is lost if any drive in the logical drive fails. Uses no storage space for fault tolerance.</td>
<td>Tolerant of multiple, simultaneous drive failures. Higher write performance than RAID 5. Uses the most storage capacity for fault tolerance.</td>
<td>Tolerant of single drive failures. Uses the least amount of storage capacity for fault tolerance.</td>
</tr>
</tbody>
</table>

© 2004 Coughlin Associates
RAID Reliability

• Redundancy
  – drives (hot spares)
  – power supplies
  – fans
  – controllers
• Automatic fail-over to spares
Direct Attached vs. Networked Storage

• In **DAS** (Direct Attached Storage) data storage can be incrementally added to a computer system and is subservient to the computer host.

• A **SAN** (Storage Area Network) is a “network storage” system in which storage is accessed through a separate storage network.

• A **NAS** (Network Attached Storage) is an independent aggregated system that can be attached to an existing LAN network in order to increase network available storage.
Network-attached storage (NAS) relies on a specialized file system that provides heterogeneous file sharing.
Switch Zoning

Diagram of a Fibre Channel Fabric with zones labeled Red, Blue, and Green. The diagram also shows JBOD, Loop 1, Loop 2, a hub, a server, an array, and virtual drives connected to the SAN switch.
Standard Device Management Interfaces – SMIS (SNIA Std.)
Examples of ATA-based Storage Systems
(Popular for Static Content Storage Systems)

Quantum DX30
The DX30 separates backup functions from archive functions to optimize the data protection process.

STK Bladestore product uses 3.5 inch drives on blade acting as one drive to a fibre channel output.

Nexsan ATABeast
Nexsan's 14 TB for 7 cents a MB.

(See new introductions at 2004 NAB)

© 2004 Coughlin Associates
MAID (Massive Array of Inactive Disks)

- Disks inactive most of the time (only about 25% active at any time)
- Can be RAID or JBOD
- Workload is mostly writes, seldom read
- Reduced costs since components shared
- Low power
- Field replaceable drives
- Start-ups ?? offering MAID systems
Virtual Tape Cache for Backup

Enterprise or midrange server
UNIX, Win2K, Linux

Direct I/O to disk arrays

Streaming I/O tape data path

Virtual library

Policy-based data movement

Embedded physical disk array
- Virtual Volumes 1, 2, ... n
- Disk or Tape device image
- Serves as a cache

Automated tape library
- Backup/recovery
- Long term archives

Source: Horison Information Strategies

© 2004 Coughlin Associates
Tape-based Digital Content Storage System
Content Software

- SGI
- Veritas
- Exanet
- SANbolic
- Kasenna
- Context Media
- Many Others
Connection Interfaces and Protocols

- SCSI
- Serial Attached SCSI
- Fibre Channel
- FATA (Seagate and HP)
- ATA
- Serial ATA
- TCP/IP and variations
- iSCSI
- FC over IP
- Infiniband
Each change represents intelligence moving from host to drive.
Each advancement was met with resistance.
Eventually advantages of new intelligence were compelling.
OSD: A New Standard Interface

Completes Device Abstraction

© 2004 Coughlin Associates

From Seagate Technology, 2003
Object Storage Systems

Expect wide variety of Object Storage Devices

- Disk array subsystem
  - I.e. LLNL with Lustre
- "Smart" disk for objects
  - 2 SATA disks – 240/500 GB
- Prototype Seagate OSD
  - Highly integrated, single disk

- Orchestrates system activity
- Balances objects across OSDs

Stores up to 5 TBs per shelf
- Battery-backed redundant power

© 2004 Coughlin Associates

From Seagate Technology, 2003
Applications for Entertainment
Content Storage
Professional Digital Camera
(Storage System for Content Capture)
Asynchronous packet switched architecture

From External Systems

Camera

Ingest Client

Non Linear Editor

Network Storage System

Production Switcher

Playout Client

WAN

To External Systems

To External Systems

© 2004 Coughlin Associates

Lowell Moulton, AF Associates
Jan. 2004
Material Exchange Format (MXF)

- International standard
- Designed to enable distribution of A/V files over IT infrastructures
- License free open source wrapper for video, audio and metadata
- Real time streams or non real time file transfers
- Wrapper can contain various metadata such as DRM
Material Exchange Format (MXF)

• Partitions enable files to be read while being written
• Files can also be tuned for file system
  – KLV Alignment Grid (KAG)
  – KAG specifies file system logical block size
• Standardized index tables
  – Enable fast access to edit units and partitions
Nonlinear System Design (Avid)

• The online real-time effects system
Workflow with File Sharing (SGI)

Near-instantaneous access for data-intensive workflows

File sharing means large files don’t have to be moved over the network—saving time, speeding workflow.

© 2004 Coughlin Associates

Gabriel Broner, SGI, Jan. 2004
Clustered Digital Content Storage

Traditional Storage Systems

- Separate islands of storage
- Complex & hard to grow
- Server performance bottlenecks
- Inherent single points of failure

Isilon IQ Clustered Architecture

- One single pool of storage
- Simple, easy, & modular to grow
- Cluster eliminates server bottlenecks
- No single points of failure

Acute Pain with Digital Content

- Separate islands of storage
- Complex & hard to grow
- Server performance bottlenecks
- Inherent single points of failure

Isilon IQ Eliminates Customer Pain

Brett Goodwin, Isilon, Jan. 2004

© 2004 Coughlin Associates
Conclusions

• Digital content creation and distribution will require large volumes of storage.
• Storage devices and requirements vary throughout the content value chain.
• Storage device and architecture development enables ever lower and more capable digital content creation and distribution!

Acknowledgement: Much of the material from this presentation was created while researching the 2004 Entertainment and Digital Media Storage Report, Authors: Tom Coughlin, Pat Hanlon, and Dennis Waid. For more information see www.tomcoughlin.com.
Digital Storage Will Entertain a New Generation!