

Library of Congress HSM Panel

Lance Evans

Chief Architect, SGI Storage

2012-09-21



A Little Recent History



StorageTek 4400

- ~4000-6000 cartridges (typical ~5000)
- STK 4490 tape drives (IBM 3490E equivalent)



A Little Recent History

- On the anniversary of the 200th 4400 shipped:
 - Steve:
“Well, I think STK has solved the capacity problem.”

A Little Recent History

- On the anniversary of the 200th 4400 shipped:
 - Steve:
“Well, I think STK has solved the capacity problem.”
 - Lance:
“Maybe, but the new problem is keeping track of all that data.

The future industry will be around finding, accessing, searching, and making use of all that data, as information.”

Where We Are Now



- HGST Deskstar 7K4000
- 4TB capacity
- Desktop drive (Nearline 4TB not yet available)

Where We Are Now

- Current densest data center disk footprint
 - ~24,000 more dense than that 4400 library
 - at ~33x the price (not inflation adjusted)
- The “new” problem is keeping track of all the data, and the metadata referencing it.

HSM Industry Challenges

- Scaling Up (each instance)
 - Billions of objects to manage behind one HSM
 - Must manage metadata reliability too
- Scaling Out (hundreds or thousands of instances)
 - Sharing common back-end infrastructure for cost amortization
 - Making such large systems practical to configure and use
- As for the media wars... I really don't care
 - As long as there's more than one type! (and there must be)
 - Constant media migration, old-to-new
- Data management software must outlive many hardware generations AND the data it manages.

HSM Industry Challenges

- Guarding against bit rot
- Guarding against transmission data transformation
- Firmware or software bugs
- Mechanisms to prevent user error
- Managing access latency
- Keeping throughput high
- Continuous operation via HA and clustering
- Security, permissions, roles, encryption
- Disaster protection of both data and metadata
- Constant hardware technology evolution
- Adding the easy button
- Etc. etc. etc.

HSM Industry Challenges

- Good news! Tape is still not dead
 - Still the least expensive medium for enormous archives
 - Still the most reliable long-term (deep / offline) storage medium

HSM Industry Challenges

- Good news! Tape is still not dead
 - Still the least expensive medium for enormous archives
 - Still the most reliable long-term (deep / offline) storage medium
- Bad news! Tape is still not dead :)
 - Time to first byte has not improved much
 - Even slower than spinning up a RAID array
 - Very low concurrency of access
 - A few tape drives shared across thousands of cartridges
 - Difficult to manage latency expectation in the Google era

My Laptop Hard Disk



My Laptop Hard Disk



- Two of about 1.7 Billion Google Garage Images
- **NOBODY IS THROWING ANYTHING AWAY**

Final Question, Commentary

- If your employer cut your budget by half this year, then half again next year,,, what would you do?
 - What data would you decide not to transition forward to new media?
 - What data would you decide not to collect and archive?
 - HOW WOULD YOU DECIDE WHAT WAS IMPORTANT?
- Semantic access to data is critical
 - Searching and culling rich metadata is key
 - Setting and enforcing policies including retention limits on specific data
 - Software facilities to MANAGE data lifecycle, not just live in denial



Trusted Leader in Technical Computing