Maximizing Performance in the Archive

Matt Starr
<table>
<thead>
<tr>
<th>BW Cost $/ (GB/s)</th>
<th>$10 (CPU included!)</th>
<th>$110</th>
<th>$200</th>
<th>$2K</th>
<th>$30K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity Cost $/GB</td>
<td>$8</td>
<td>$0.3</td>
<td>$0.05-0.15</td>
<td>$0.01-0.02</td>
<td></td>
</tr>
<tr>
<td>Node BW (GB/sec)</td>
<td>1 TB/s</td>
<td>100 GB/s</td>
<td>20 GB/s</td>
<td>5 GB/s</td>
<td></td>
</tr>
<tr>
<td>Cluster BW (TB/sec)</td>
<td>1 PB/s</td>
<td>100 TB/s</td>
<td>5 TB/s</td>
<td>100 GB/s</td>
<td>10’s GB/s</td>
</tr>
<tr>
<td>Software</td>
<td>Language level</td>
<td>Language level HDF5 / DAOS</td>
<td>DDN IME Cray Data Warp</td>
<td>Parallel FS Lustre</td>
<td>Archive HPSS/DMF</td>
</tr>
<tr>
<td>Key features</td>
<td>transparent computation</td>
<td>transparent computation ultra-fast storage apps</td>
<td>name space scientific formats FS style container</td>
<td>bulk data movement - many files subtrees of MD</td>
<td></td>
</tr>
</tbody>
</table>
Disk 100-200MB/s per brick @ $250-800 per brick

Tape 300-360MB/s per brick @ $7,000 – 15,000 per brick
TFinity ExaScale

TFinity Delivers

• Up to 641 PB raw with LTO-8
• Over 1.5 Exabytes with Compression
• Flexible Configurations
  • 3 to 44 frames
  • 100 to 53,460 LTO / 40,680 TS slots
  • 2 to 144 tape drives
• LTO-6, LTO-7, LTO-8, TS1150, TS1155 Technology
• T10K legacy support for migration
• Dual robotics for availability and performance
• 24 x 7 x 365 operations
TFinity ExaScale
Hardware / Software interface

Do it faster
or
Do less
TFinity – High Performance Transporter

• High Performance Transporter
  • Evolutionary Progression in performance, and reliability.
  • Completely new robotics and firmware.
  • Designed for 2,000,000 MCBF.

<table>
<thead>
<tr>
<th></th>
<th>First Generation Transporter</th>
<th>High Performance Transporter</th>
</tr>
</thead>
<tbody>
<tr>
<td>TeraPack Drawer Pick/Put</td>
<td>7 seconds</td>
<td>&lt; 1 second</td>
</tr>
<tr>
<td>Slot Pick</td>
<td>3 seconds</td>
<td>.5 seconds</td>
</tr>
<tr>
<td>Slot Put</td>
<td>3 seconds</td>
<td>.5 seconds</td>
</tr>
<tr>
<td>Picker Head Rotation</td>
<td>1 second</td>
<td>.25 seconds</td>
</tr>
</tbody>
</table>
Zoning
TFinity Zoning Optimizations

• The TFinity Bluescale interface has been modified to include zoning information in the Read Element Status response.

• This allows the storage software to keep both robots working optimally in their zone without any robotic contention.
TFinity TeraPack® Affinity

• Do as many tape moves as possible from the same TeraPack to increase overall robotic performance.

• By giving the storage software the TeraPack barcode associated with each tape barcode the software can sort the internal move queue to consolidate all tape moves within a given TeraPack to happen in order.
ISV + Spectra Logic Mount Rate percentage

No library knowledge: 100%
Zoning: 189%
Terapack Affinity: 223%
TAOS™
Read Performance

Time-based Access Order System
(patent pending)
Evolution of reading a disk drive

- **Reads were done one cylinder at a time.**
  - Usually one read per rotation
- **Disk driver/OS Strategy Routine**
  - Disk drivers began to understand the disk layout and optimize the order of the read.
  - Disk Driver Strategy Routine would re-orders read to optimizes heads/Sectors/Cylinders
- **Strategy Routine moved into the disk drive**
  - Chained commands given to drive
  - Drives buffers the data and does multiple reads per rotation
  - No knowledge of disk geometry
Why block file locate is important

• ISV (HPSS/DMF/Versity) software see tape as one long sequences of blocks

• Modern tape uses wraps
  • Each tape generations changes number of wraps (Single Pass)
  • Track count per band also changes

• An LTO 8 tape is almost 200 km long when read from beginning to end

• Recommend Access Order
  • Read in the best restore order is only available on Enterprise drives, Not on LTO
Example of 6 files recalled

LTO7: 112 wraps, 28 Per Band
LTO8: 208 wraps, 52 Per Band
Linear recalls are very inefficient
Ordered recalls based on LPOS much more efficient
Crossing Bands Is Considered
300; 100 MB files restored
Normal ~2 hour and 45 minutes
TAOS ~55 minutes
Media and Drive Wear Reduction Testing with TAOS

• A test was performed with 100 files with a file size between 1-100MB:
• TAOS Meters of Tape Across the Drive Head: 2,470
• Unordered Meters of Tape Across the Drive Head: 31,878
• That's a ~13x reduction in meters of tape. Or 8.4% of the original meters of tape.
• Spectra currently estimates that TAOS will reduce tape and drive wear by ~10X on media read operations when used in conjunction with HPSS 7.5 or later
• Releasing with major archive packages
Thank you