Terabyte IDE RAID-5 Disk Arrays

David A. Sanders, Lucien M. Cremaldi, Vance Eschenburg, Romulus Godang, Christopher N. Lawrence, Chris Riley, and Donald J. Summers

University of Mississippi

Donald L. Petravick

Fermilab
Introduction

• $2000 per Terabyte Storage is Available
• Scalable for use at both Small and Large Institutions — From 1 TB to 250 TB, the same as a $ Million tape silo.
• Fast Access to Data
• Redundant — RAID5
• Commodity Hardware
Definitions

- RAID — Redundant Array of Inexpensive Disks
- RAID level 0 — Concatenation
- RAID level 1 — Mirroring
- RAID level 4 — Parity
- RAID level 5 — Striped-Parity
- EIDE — Enhanced Integrated Drive Electronics
Why Use Commodity Hardware?

“Frankly sir, we’re tired of being on the cutting edge of technology.”
Hardware

- System Disk — 40 GB Maxtor
- Eight 160 GB Maxtor Disks
- 2 Promise Ultra133 PCI cards
- 24” EIDE Cables
- CPU — 1.6 GHz AMD Athlon
- Motherboard — Asus A7M266
- 512 MB DDR memory
- Second Power Supply (15A at 12V)
RAID5 Box for BABAR
RAID5 Box for CMS
## Disks

<table>
<thead>
<tr>
<th>Disk</th>
<th>RPM</th>
<th>$/GB</th>
<th>GB/platter</th>
<th>Amps@12V</th>
<th>Warranty</th>
</tr>
</thead>
<tbody>
<tr>
<td>160 GB Maxtor</td>
<td>5400</td>
<td>1.03</td>
<td>40</td>
<td>1.8</td>
<td>1 Yr.</td>
</tr>
<tr>
<td>250 GB Maxtor</td>
<td>5400</td>
<td>1.28</td>
<td>80</td>
<td>~1.1</td>
<td>1 Yr.</td>
</tr>
<tr>
<td>200 GB Western Digital (8MB cache)</td>
<td>7200</td>
<td>1.20</td>
<td>66</td>
<td>1.3</td>
<td>3 Yrs.</td>
</tr>
<tr>
<td>180 GB Hitachi/IBM</td>
<td>7200</td>
<td>1.34</td>
<td>60</td>
<td>2.0</td>
<td>3 Yrs.</td>
</tr>
</tbody>
</table>
Software

• Linux 2.4.17 Kernel (with >137 GB patch) (the latest stable kernel is the 2.4.20)
• raidtools available with most distributions
• Journaling File systems (ext3)
• NFS to mount on other computers (Linux, Sun Solaris, DEC Ultrix, Mac OSX)
• HDPARM speed test (~95 MB/s)
• Simple write test (95 MB/s)
High Energy Physics Data Analysis Strategy

• Use Parallel Processing
• Split data and store on many RAID5 PCs
• Analysis for a subset of data takes place locally on the PC where the data resides
• Network is only used to combine results
• Or use NFS to mount RAID5 array on many PCs (Less efficient due to network overhead)
High Energy Physics Cluster

Managed Gigabit Ethernet Switch

Gigabit Ethernet Switch

RAID5 CPU

RAID5 CPU

RAID5 CPU

RAID5 CPU

Gigabit Ethernet Switch

RAID5 CPU

RAID5 CPU

RAID5 CPU

RAID5 CPU
NFS Mounted Cluster

High Bandwidth Ethernet Switch

Gigabit Ethernet Switch

RAID5 CPU

CPU           CPU           CPU           CPU           CPU           CPU
Future Plans

Hardware:
• System Disk — 40 GB IBM
• Twelve 250 GB Maxtor Disks
• 3 Promise Ultra133 PCI cards
• CPU — Dual 2.0 GHz AMD Athlon
• Motherboard — Asus A7M266D
• Gigabit Ethernet Card
• Second Power Supply (15A at 12V)

Software:
• Try Stock Linux Kernel 2.4.20
• Test other Journaling File systems (ReiserFS, xfs)
## Commercial Systems

<table>
<thead>
<tr>
<th>Model</th>
<th>Capacity</th>
<th>Size</th>
<th>Price*</th>
<th>Price/GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple Xserve RAID</td>
<td>2.52 TB</td>
<td>3U</td>
<td>$10,999</td>
<td>$4.36</td>
</tr>
<tr>
<td>Dell EMC CX200</td>
<td>2.2 TB</td>
<td>3U</td>
<td>$30,000</td>
<td>$13.63</td>
</tr>
<tr>
<td>HP 7100</td>
<td>2.2 TB</td>
<td>Two 3U</td>
<td>$109,968</td>
<td>$50.21</td>
</tr>
<tr>
<td>IBM ProFibre DF4000R</td>
<td>2.2 TB</td>
<td>Two 3U</td>
<td>$43,974</td>
<td>$20.08</td>
</tr>
<tr>
<td>Sun StorEdge T3</td>
<td>2.64 TB</td>
<td>Three 3.5U</td>
<td>$144,300</td>
<td>$54.66</td>
</tr>
</tbody>
</table>

*Based on suggested retail prices on February 7, 2003
From Apple document L26325A_XserveRAID_TO.pdf
Summary

- $2000 per Terabyte RAID5 arrays of EIDE Drives tested, without tape backup.
- They are Scalable — Cost less/TB than a tape silo, but scalable down to 1 TB.
- Uses Commodity Hardware.
- Tested with 160 GB hard disks

Supported by the U.S. Department of Energy under DE-FG05-91ER40622 and DE-AC02-76CH03000.