Multi-Terabyte EIDE Disk Arrays running Linux RAID5

by

University of Mississippi: David A. Sanders, Lucien M. Cremaldi, Vance Eschenburg, Romulus Godang, Michael D. Joy and Donald J. Summers and Fermilab: Donald L. Petravick

Presented at Computing in High Energy Physics 2004 (CHEP04) 27th September - 1st October, 2004, Congress Centre, Interlaken, Switzerland

Introduction
- $2000 to Terabyte Storage is Available
- 10 times cheaper than Sun StorEdge 6120
- Scalable for use at both Small and Large Institutions
- From 2 TB to 250 TB, the latter at a $3 Million cost
- 250 TB is 1 month of LHC data
- Redundant — RAID5
- Commercial Hardware

Hardware — Base System
- System Disk — 128-GB Western Digital
- 24" IDE IDE Cables
- CPU — Dual 2.06 GHz AMD Athlon
- Motherboard — MSI K7T29 Master MPX
- 10/100 MB EDDI network
- Second Power Supply (115V or 12V)
- Gigabit Ethernet

RAID5 Test Box

32 Disk Array — upper bay

RAID5 Test Box

Results — Software RAID5
- Base writes spend 20 MB/s
- Speed of 24 MB/s for 2 concurrent writes (17% overhead)
- 15 MB/s read (using cp to system disk)
- 33 MB/s write (using cp from system disk)
- 10-15% CPU overhead but on other CPU.

Definitions
- RAID — Redundant Arrays of Inexpensive Disks
- RAID level 0 — Concatenation
- RAID level 1 — Mirroring
- RAID level 4 — Parity
- RAID level 5 — Strip Parity
- EIDE — Enhanced Integrated Drive Electronics

Additional Hardware
Software RAID5
- Eight 250 GB Western Digital disks at 800-GB Master disks
- 2 Promise 1lu/33 PCI cards

Hardware RAID5
- Twelve 250 GB Western Digital disks
- 2ware 12 disk RAID controller 7506-12

Why Use Commodity Hardware?
- Disk:

<table>
<thead>
<tr>
<th>Disk</th>
<th>RPM</th>
<th>Capacity</th>
<th>512B Block</th>
<th>MB/s Read</th>
<th>MB/s Write</th>
<th>MB/s Read (512B)</th>
<th>MB/s Write (512B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7200</td>
<td>3.5&quot;</td>
<td>120 GB</td>
<td>46</td>
<td>45 MB/s</td>
<td>35 MB/s</td>
<td>13 MB/s</td>
<td></td>
</tr>
<tr>
<td>7200</td>
<td>3.5&quot;</td>
<td>73 GB</td>
<td>46</td>
<td>44 MB/s</td>
<td>42 MB/s</td>
<td>10 MB/s</td>
<td></td>
</tr>
<tr>
<td>7200</td>
<td>3.5&quot;</td>
<td>50 GB</td>
<td>46</td>
<td>43 MB/s</td>
<td>40 MB/s</td>
<td>8 MB/s</td>
<td></td>
</tr>
<tr>
<td>7200</td>
<td>3.5&quot;</td>
<td>30 GB</td>
<td>46</td>
<td>41 MB/s</td>
<td>38 MB/s</td>
<td>6 MB/s</td>
<td></td>
</tr>
<tr>
<td>7200</td>
<td>3.5&quot;</td>
<td>20 GB</td>
<td>46</td>
<td>39 MB/s</td>
<td>35 MB/s</td>
<td>5 MB/s</td>
<td></td>
</tr>
</tbody>
</table>

Problems and Solutions

Data Storage Cake

Tests
- Measured speed and CPU overhead of Software RAID5
- Measured speed and CPU overhead of Hardware RAID5
- Measured speed of network file transfers
- Tested journaling file system ext3 and ReiserFS

Future Recommendations

Commercial Systems

- Sun Microsystems
- IBM System/390
- HP StorageWorks
- IBM and Sun RAID10

Summary
- $2000 to $10,000 RAID arrays of EIDE Drives tested, normal tape included
- They are Scalable — Consume less TB than a tape drive but not as cheap
- Use Commercial Hardware
- Tested Hardware not laboratory, but in 50% of 300+ GB hard drives
- Tested Hardware