Efficient Data Storage

MSST 2008
Overview

- Atrato
- Efficiency: Physical Space, Energy Usage, and Performance in general
- Physical Space
- Storage Energy usage
- Secondary energy usage
- Performance
- Storage design considerations
- Why move to smaller disk drives
- Design Considerations
- Conclusions
Atrato

- Atrato was founded on the idea of addressing *Data Access* as opposed to just capacity and/or bandwidth.
- Startup formed in January 2004.
- Formerly Sherwood Information Partners, Inc., name changed Feb 2008.
- Based in Westminster, CO.
- Focused on:
  - Self-maintaining Array of Identical Disks (SAID).
  - High-density packaging of small-form-factor commodity disk drives.
  - Highly scalable storage controller – *Avenger*.
- Atrato, Inc. is named after the Rio Atrato in Colombia. The Rio Atrato discharges at least 175,000 cubic feet (5,000 cubic m) of water per second making it the fastest river in the world.
This is NOT a commercial for Atrato
It is intended to demonstrate some non-intuitive results of the application of small form-factor laptop-class disk drives in a massively parallel array
Atrato and a former division of Seagate are the only two companies working on this at the time of this writing
Physical Space

- Disk drive Form Factors are 3.5-inch low profile and 2.5-inch laptop
  - 3.5-inch = 147mm x 102mm x 26mm
  - 2.5-inch = 70mm x 40mm x 9.5mm
- Approximately 5.85:1 2.5-inch disks to 3.5-inch disks in terms of physical volume
- Practical packaging of 2.5-inch drives easily supports a volumetric ratio of 4:1
- Fail-in-place packaging model can boost the volumetric ratio to 10:1
  - 160 disk drives in a single 3RU enclosure versus 16
Energy Usage in General

- Data Center Power Consumers
  - Processors: The principle consumer in a data center – anywhere from 50% to 90%
  - Storage Devices – about 20%
  - Networks
  - Cooling units
  - Power Distribution units
  - Displays and Misc

- Focus here is on Storage Energy Usage, specifically disk drives (no tapes)
Storage Energy Usage

- Disk Drive is a primary energy consumer in a storage system
- Typical Storage System Components
  - Disk Drive
    - Motor
    - Electronics
    - Actuator
  - Infrastructure (enclosure, controllers, fans, …etc)
- Disk Drive Energy Usage Relative to Each Other
  - In a 3.5-inch disk it is about 33/33/33
  - In a 2.5-inch disk is it about 20/50/30
Secondary Energy Usage

- **Cooling**
  - For every watt used it takes 1 watt to remove the heat
  - Air cooling is currently the preferred method
  - Water is 4000 times more effective than air for cooling components
- Keep the heat out of the box
- Get the heat out of the rack
- Move the heat out of the data center
Performance

- **In General**
  - Instead of making a single disk run ever faster, just use a larger number of smaller disks
  - Common practice in CPU industry
  - Virtually unknown in Disk Storage industry

- **IOPS**
  - Smaller laptop-class disk drives are individually slower than an Enterprise-class drive – about 2:1 in favor of 3.5-inch disks
  - Can package an order of magnitude more laptop-class drives in an array
  - Aggregate IOP performance for an array of SFF drives is 5:1 in favor of the 2.5-inch disks

- **Bandwidth**
  - Same argument as above - about 2:1 in MB/sec in favor of 3.5-inch disks
  - Aggregate bandwidth for an array of SFF drives is 5:1 in favor of 2.5-inch disks
Why move toward smaller disks

- Power reduction is non-linear in favor of smaller form factor
- Cooling is simpler because of low power consumption
- Self induced (rotational) vibration modes are significantly reduced
- Pricing takes advantage of the commodity lap-top drives
- Reliability and data integrity is a different talk
- Aggregate performance is significant
- Aggregate head-count per unit space or volume is significantly higher than 3.5-inch packaging methods
### Why not Move Toward Smaller Disks

<table>
<thead>
<tr>
<th></th>
<th>3.5</th>
<th>2.5</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Capacity</strong></td>
<td>1TB</td>
<td>320GB</td>
<td>~1/3&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td>Per drive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Density</strong></td>
<td>1TB</td>
<td>1.87TB</td>
<td>87% higher</td>
</tr>
<tr>
<td>TB/unitvol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Space in units of 2.5” drives</strong></td>
<td>5.85</td>
<td>1</td>
<td>~1/6&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>IOPS</strong></td>
<td>77</td>
<td>59 / 236-354</td>
<td>Individually slower</td>
</tr>
<tr>
<td><strong>BW (MB/s)</strong></td>
<td>105</td>
<td>58 / 232-348</td>
<td>Individually slower</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Aggregate is much higher</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Aggregate is much higher</td>
</tr>
</tbody>
</table>
In terms of Power...

<table>
<thead>
<tr>
<th></th>
<th>3.5</th>
<th>2.5</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seek/R/W Power</td>
<td>12W</td>
<td>2W</td>
<td>1/6&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td>Idle Power</td>
<td>8W</td>
<td>0.6W</td>
<td>1/13&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td>Capacity</td>
<td>83.3 GB/W</td>
<td>160 GB/W</td>
<td>~ 2 X better</td>
</tr>
<tr>
<td>Density</td>
<td>83.3</td>
<td>312</td>
<td>3.75 X better</td>
</tr>
<tr>
<td>Power Density</td>
<td>2 UV/W</td>
<td>2 UV/W</td>
<td>Same</td>
</tr>
<tr>
<td>IOPS</td>
<td>6 IOPS/W</td>
<td>30 IOPS/W</td>
<td>5X better</td>
</tr>
<tr>
<td>BW (MB/s)</td>
<td>9 MB/s/W</td>
<td>29 MB/s/W</td>
<td>3X better</td>
</tr>
</tbody>
</table>
Storage Design Considerations

- **Performance**
  - Signal Aggregation
  - “We need more disks, not bigger ones” Gary Grider, NNSA

- **Tight packaging but you must get the heat out**
  - Heat is the #1 threat to disk drive life - maybe
  - Bigger disks produce more heat than smaller ones
  - Tight packaging can require sophisticated cooling

- **Vibration management**
  - 2.5-inch drives have virtually no rotational vibration

- **Maintenance** – Fail-in-place
A SAID – Self-maintained Array of Identical Disks
What it all comes down to

Atrato SAID
• 3RU
• 9 GB/sec
• 10,000 IOPS
• 50TB raw capacity
• 700Watts

3.5-inch standard packaging (16 drives per box)
• 30RU
• 16 GB/sec
• 12,320 IOPS
• 160 TB raw capacity
• 2500 Watts

Traditional 3.5-inch enclosures
Conclusions

- Small disks are non-intuitively better than 3.5-inch disks when it comes to power
  - Better performance/watt
  - Better capacity/watt
- Requires different engineering practices
- Requires different maintenance philosophy
Thank-you