Feeding the Pelican
Using archival hard drives for cold storage racks

Austin Donnelly†, Richard Black†, Dave Harper†, Ant Rowstron†, Aaron Ogus‡

†Microsoft Research, ‡Microsoft
Pelican

- 1152 Archive-grade HDDs
- Directly attached to 2 servers via HBA, PCIe
- Orthogonal power & cooling domains
- Spundown to meet power limit
- Schedule requests to get good performance
Archive drives

• New class of HDDs
• Optimised for minimum $/GB
• Targeting cold workloads:
  “The WD Ae hard drive is best suited for cold storage, backup and data archiving where data is stored on disk but rarely if almost never read again”
  -- WD6001F4PZ1 datasheet

• Workload is quantified as TB/year
• Lifetime affected by:
  POH  TB transferred  Spindown cycles
# Drive line-up

<table>
<thead>
<tr>
<th>Name</th>
<th>Technology</th>
<th>Spin up (s)</th>
<th>Capacity (TB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Auto SMR</td>
<td>10.1</td>
<td>8.0</td>
</tr>
<tr>
<td>A2</td>
<td>HA SMR</td>
<td>10.2</td>
<td>8.0</td>
</tr>
<tr>
<td>B1</td>
<td>PMR</td>
<td>7.9</td>
<td>4.6</td>
</tr>
<tr>
<td>B2</td>
<td>PMR</td>
<td>7.8</td>
<td>4.5</td>
</tr>
<tr>
<td>B3v1</td>
<td>PMR</td>
<td>9</td>
<td>4.9</td>
</tr>
<tr>
<td>B3v2</td>
<td>PMR</td>
<td>6</td>
<td>4.9</td>
</tr>
<tr>
<td>B4</td>
<td>PMR</td>
<td>6.4</td>
<td>6.1</td>
</tr>
<tr>
<td>C1</td>
<td>Auto SMR (?)</td>
<td>8.6</td>
<td>8.0</td>
</tr>
</tbody>
</table>
Spinup latency

![Graph showing CDF of spinup latency for different data sets A1, A2, B1, B2, B3v1, B3v2, B4, C1 with latency on the x-axis and CDF on the y-axis.](image-url)
Power draw

Austin Donnelly -- Feeding the Pelican
Is it ok to do all these spinups?

datasheet spec: 100K per year.

Number of spinups per drive

Austin Donnelly -- Feeding the Pelican
TBs transferred

datasheet spec: 60TB/year
Power On Hours

datasheet spec: 3120 POH/year (about 1/3\textsuperscript{rd} of a year)
Temperature

Austin Donnelly -- Feeding the Pelican
Datacenter environment

![Graph showing temperature and relative humidity over time for different locations: humidity, exh, inlet, ext. The graph spans from January 2015 to January 2016, with temperature ranging from -10 to 60°C and relative humidity from 15% to 65%.](image-url)
Failure rate

![Graph showing CDF against age of failed drives (days)]

AFR: 3.96%
Conclusion

• Archive drives are effective, provided workload is managed
• Spindowns do not seem to affect drive reliability
• Temperature and humidity are bigger factors

• Future:
  • Workload management not just for SSDs
  • Worth understanding drive performance at firmware level
  • Power control is critical
  • Background operations control is needed too (particular for SMR)