Hyperscale: Challenges and Solutions

Ross Stenfort
Hardware System Engineer, Facebook
Facebook @ Scale

1 Billion

1.3 Billion

2.7 Billion
Where Does Hyperscale Use Flash Today?

M.2s
Hyperscale Evaluation Scorecard

Important

• Scalable & Flexible
• High volume & Low cost
• Power & Thermal Efficiency
• Hot-swappable & Serviceable
• Performance per TB & Quality of Service
• Security

Less Important

• Backwards compatible
• Support for non-NVM media
• Maximum density
• Peak Performance (Peak IOPs/BW)
Hyperscale Evaluation Scorecard

Important

• Scalable & Flexible
• High volume & Low cost
• **Power & Thermal Efficiency**
• Hot-swappable & Serviceable
• Performance per TB & Quality of Service
• Security

Less Important

• Backwards compatible
• Support for non-NVM media
• Maximum density
• Peak Performance (Peak IOPs/BW)
Power & Thermal Efficiency

Power and thermal efficiency are important

- Limited airflow and power is available in datacenters
- Temperature increase across servers is large (delta T)
- Operation expense is important

M.2s are used today however the LFM/ W is a challenge which is driving to new form factors.
Form Factor Thermal Comparison

- M.2
- M.2 with FB Heat Sink

Too large if there is a processor in storage/compute box

2U tall which does not scale across the data center echo system due to 1U is the standard building block which things are built from

E1.S 25W provides excellent power/thermal/size trade offs for hyperscale
## Hyperscale Evaluation Scorecard

<table>
<thead>
<tr>
<th>Important</th>
<th>Less Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Scalable &amp; Flexible</td>
<td>• Backwards compatible</td>
</tr>
<tr>
<td>• High volume &amp; Low cost</td>
<td>• Support for non-NVM media</td>
</tr>
<tr>
<td>• Power &amp; Thermal Efficiency</td>
<td>• Maximum capacity density</td>
</tr>
<tr>
<td>• Hot-swappable &amp; Serviceable</td>
<td>• Peak Performance (Peak IOPs/BW)</td>
</tr>
<tr>
<td><strong>Performance per TB &amp; Quality of Service</strong></td>
<td></td>
</tr>
<tr>
<td>• Security</td>
<td></td>
</tr>
</tbody>
</table>
IOPs scales with capacity

*Basic assumptions: 4TB SSDs @ 300k 4k IOPs and 600k IOPs SATA limitation
Industry Trends

Flash and CPU continue to diverge

Flash Capacity Growth
CPU Performance
Dark Flash

Flash capacity utilization trend vs. target

Growing gap of underutilized Flash (Dark Flash)

Note: Includes 25% generation over generation performance improvements
Scalable Performance with NVM Sets

Compute client A

Compute client B

Compute client C

Compute client D

Storage server

Improves Utilization

NVM Set 0
NVM Set 1
NVM Set 2
NVM Set 3

NVM Set 0
NVM Set 1
NVM Set 2
NVM Set 3

NVM Set 0
NVM Set 1
NVM Set 2
NVM Set 3

NVM Set 0
NVM Set 1
NVM Set 2
NVM Set 3

NVMe SSD

NVMe SSD

NVMe SSD

PCIe
Hyperscale Challenges and Solutions

Important

- Scaleable and Flexible
- High volume and Low Cost
- Power and Thermal Efficiency
- Hotswap and Serviceability
- Performance per TB & Quality of Service
- Security

Less Important

- Backwards compatible
- Support for non-NVM media
- Maximum capacity density
- Peak Performance (Peak IOPs/BW)

There may be many challenges, but innovative, standardized solutions are the key to scaling for the future!