The Benefits of Flash in Enterprise Storage Systems

David Dale, NetApp
The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.

Member companies and individual members may use this material in presentations and literature under the following conditions:
- Any slide or slides used must be reproduced in their entirety without modification
- The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.

This presentation is a project of the SNIA Education Committee.

Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.

The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.
Agenda

Why flash in the datacenter? Why now?

Memory, cache and storage

Flash in enterprise storage today
- Hybrid arrays; SSD storage tier
- Storage controller-based cache
- Flash in host systems
- All-flash arrays

What’s next

Conclusion
Why Flash in the Data Center?

Why flash?

- Capacity efficiency versus DRAM
  - ~5x better $ per GB
  - ~40x better power per GB

- IOPS efficiency versus HDDs
  - ~40x better $ per IOPS
  - ~600x better power per IOPS

Why now?

- Period of rapid density advancements led to HDD-like bit density at lower $/GB than DRAM
- Innovations in SSD and tiering technology
IOPS Efficiency

Mixed 8KB IOPS efficiency

- DDR3 DIMM: O(1000)
- SLC SSD: O(100)
- MLC eSSD: O(100)
- 15K SAS: O(1000)
- 7200 SATA:

Mixed IOPS per $  vs  Mixed IOPS per mW
Capacity Efficiency

![Capacity Efficiency Chart](chart.png)

- DDR3 DIMM
- SLC SSD
- MLC eSSD
- 15K SAS
- 7200 SATA

**Capacity Efficiency**

- **MB per $**
- **MB per mW**
Read Latency

Read latency to first 512 bytes

- DDR3 DIMM
- SLC SSD
- MLC eSSD
- 15K SAS
- 7200 SATA

- O(100)
- O(1000)
Assuming that the cost of a cache is dominated by its capacity, and the cost of a backing store is dominated by its access cost (cost per IOPS), then the breakeven interval for accessing a page of data in cache is given by:

$$\text{Break-Even-Interval} = \frac{\text{Backing-Store-Cost-Per-IOPS}}{\text{Cache-Cost-Per-Page}}$$

1987: Disk $2,000 / IOPS; RAM $5 / KB ➜ 1 KB breakeven = 400 seconds ≈ 5 minutes
Five Minute Rule, 2010: DRAM & HDD

- **Disk** $1 / IOPS (2,000x reduction)

- **DRAM** $25 / GB (200,000x reduction)

- $100 KB breakeven ~= 5 minutes
- $8 KB breakeven ~= 1 hour
- $1 KB breakeven ~= 10 hours as Gray predicted

- $200,000x / 2,000x = 100-fold decrease in breakeven access rate for a DRAM cache page backed by disk \( \Rightarrow \) much bigger DRAM caches
Five Minute Rule, 2010: DRAM & Flash

- MLC eSSD ~$0.10 / mixed 8 KB IOPS

- DRAM $25 / GB

- 8 KB breakeven ~= 8 minutes (1/10th DRAM)

Adding flash between DRAM and HDD reduces the breakeven access interval for DRAM by 10x, indicating that DRAM capacity could be reduced to hold working sets for data accessed 1/10th as often.
An IOPS Density View

Highest MB per $ at a given IOPS density is best

Mixed 8KB IOPS per GB

- DDR3 DIMM
- SLC SSD
- MLC eSSD
- 15K SAS
- 7200 SATA
NAND vs HDD History

Price per Gigabyte

$10^4$

$10^3$

$10^2$

$10^1$

$10^0$

$10^{-1}$

$10^{-2}$

$10^{-3}$

$10^{-4}$

20X the Price

Source: Objective Analysis

Understanding the NAND Market

Benefits of Flash in Enterprise Storage

© 2014 Storage Networking Industry Association. All Rights Reserved.
Categories of Flash Systems

- **Host-based Flash**
  - Flash hardware, stores persistent data
  - May be combined with s/w to form cache
  - May act as internal DAS

- **Flash in Storage Controller**
  - Flash hardware and s/w
  - “Behind wire”
  - E.g. Flash Cache

- **Host-side Flash S/W**
  - Software only, may be tied to particular flash hardware

- **Flash-based Virtual Storage Appliance**
  - Software

- **Hybrid (Flash/HDD) Array**
  - Mixed flash / HDD
  - Automated Storage Tiering
  - Traditional HDD arrays w/SSDs
  - Custom-designed Flash/HDD systems

- **All-Flash Array**
  - Traditional arrays configured only w/ SSDs
  - Flash only designed specifically for SSDs
(A) Hybrid Arrays

- Mixing SSD and HDD for a particular workload will probably be the most cost-efficient use of SSDs in over the next few years
- SSD and HDD tiers accommodated in storage shelves
- Issue is to dynamically map workload to appropriate media
- Automated data placement and movement is essential
  - Automated storage tiering (AST)
  - Policy-based
  - No administrator overhead imposed
  - Some vendors refer to this as tier-less storage
AST Considerations

❖ **Media configurability**
  ◆ Virtual pool to LUN/aggregate mapping

❖ **Management granularity**
  ◆ Automatic or Policy-engine based? Per LUN?

❖ **Operational flexibility**
  ◆ Dynamic or batch migration? Predictable?

❖ **Management granularity**
  ◆ Block size (smaller, better)? Dedupe? RAID limitations?

❖ **Workload sizing**
  ◆ Predictive cache? Online workload tool? Best guess? Spreadsheet?
Practical Use Cases

Database acceleration solution
- Entire database on SSD tier, or
- Hot random access files on SSD and rest of database on standard disk
  - Indexes and temp space

Large scale virtual machine environments
- Solves “boot storm” problem for large numbers of virtual machines
- Deduplication of VM data, e.g. virtual desktops
  - Reduces capacity requirements, increasing IOPS density, potentially making SSD economical
(B) Controller-based Flash Cache

- Functions as an intelligent read cache for data and metadata
- Automatically places active data where access can be fast
- Provides more I/O throughput without adding high-performance disk drives to a disk-bound storage system
- Effective for file services, OLTP databases, messaging, and virtual infrastructure
Reduce Latency with Flash Cache

Cache hits can reduce latency by a factor of 10 or more
Use case: Scale Performance of Disk-bound Systems

Add Spindles
- Use more disks to provide more IOPs
- May waste storage capacity
- Consumes more power and space

Starting Point:
Need More IOPs
- Performance is disk-bound
- Have enough storage capacity
- Random read intensive workload

Add Flash Cache
- Use cache to provide more IOPs
- Improves response times
- Uses storage efficiently
- Achieves cost savings for storage, power, and space

Benefits of Flash in Enterprise Storage
© 2014 Storage Networking Industry Association. All Rights Reserved.
SATA HDD plus Flash Cache Example

**Benchmarked Configurations**

- **FC Baseline Configuration**
  - 224 FC drives
  - 64TB

- **SATA + Flash cache Configuration**
  - 96 SATA drives
  - 96TB

**50% More Capacity**

**SPECsfs2008 Performance**

- **FC Baseline**
- **SATA + Flash cache**

<table>
<thead>
<tr>
<th>Throughput (k-ops/sec)</th>
<th>Response Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>1.0</td>
</tr>
<tr>
<td>12</td>
<td>1.5</td>
</tr>
<tr>
<td>18</td>
<td>2.0</td>
</tr>
<tr>
<td>24</td>
<td>2.5</td>
</tr>
<tr>
<td>30</td>
<td>3.0</td>
</tr>
<tr>
<td>36</td>
<td>3.5</td>
</tr>
<tr>
<td>42</td>
<td>4.0</td>
</tr>
<tr>
<td>48</td>
<td>4.5</td>
</tr>
<tr>
<td>54</td>
<td>5.0</td>
</tr>
</tbody>
</table>

- Purchase price is **39% lower** for SATA + Flash cache compared to FC baseline
- SATA + Flash cache yields **66% power savings** and **59% space savings**


SPEC® and SPECsfs2008® are trademarks of the Standard Performance Evaluation Corp.
(C) Host-based Flash

- Flash card on PCI bus in host system
  - Can support SCSI semantics or device driver model
- Acts as Tier 0 storage (IOPS tier) in front of networked storage (capacity tier)
  - Requiring no data movement (caching)
  - Requiring data movement (AST)
- Multiple implementations in development:
  - High performance DAS
  - Shared storage RAID subsystem in VM on host
  - Shared storage OS in VM on host
- Area of intense industry and standards activity
Typical Use Cases

- **High-performance DAS workloads** which entirely fit into host Flash (typically OLTP or hot virtual server applications)
  - For data and workloads which need to take advantage of shared storage data protection, use host flash with shared storage RAID subsystem in VM on host
  - For data and workloads which need to take advantage of shared storage data protection, data management and/or deduplication use host flash with shared storage operating system in VM (Virtual Storage Appliance) on host
Two categories today:

- Hardware design center, focused on high performance and price/performance, lightweight data management, small footprint, low power usage, designed for scale-up
- Software design center, balance of performance and features, robust data management with built-in efficiency and protection mechanisms, designed for scale-out

Ultra high performance and sustained low latency for real-time OLTP, VDI, tech apps

The next-generation tier 1 storage?
System Evolution

1980
- Logic: CPU
- Memory: RAM
- Active Storage: DISK
- Archival: TAPE
  - Fast, synch
  - Slow, asynch

2011
- Logic: CPU
- Memory: RAM
- Active Storage: DISK/FLASH SSD
- Archival: TAPE

2017+
- Logic: CPU
- Memory: RAM
- Storage Class Memory*: DISK
- Archival: TAPE

* e.g. Phase Change Memory
  - MRAM/STTMRAM
  - RRAM

Benefits of Flash in Enterprise Storage
© 2014 Storage Networking Industry Association. All Rights Reserved.
Summary

- Solid state technologies are having a profound impact on enterprise storage
- It’s not just about replacing mechanical media with solid state media
- The architectural balance of memory, cache and persistent storage is changing
- Today’s solid state implementations in enterprise storage demonstrate these changes
- It’s still early days in this discontinuity… with persistent non-volatile memory on the horizon
The SNIA Education Committee would like to thank the following individuals for their contributions to this Tutorial.

**Authorship History**

- David Dale; Spring 2010
- **Updates:**
  - David Dale; Fall 2010
  - David Dale; Spring 2011
  - David Dale; Fall 2011
  - David Dale; Spring 2012
  - David Dale; Spring 2013
  - David Dale, Spring 2014

**Additional Contributors**

- Jim Handy
- Jeff Kimmel
- Chris Lionetti
- Phil Mills
- Amit Shah
- Mark Woods
- Alan Yoder
- Bharat Bagmane
- Alex McDonald