Driving IMC Computing Efficiency with Flash Extended Memory
Dan Baigent, Sr Director Strategic Partner Ecosystems
June 30, 2015
Forward-Looking Statements

During our meeting today we will make forward-looking statements.

Any statement that refers to expectations, projections or other characterizations of future events or circumstances is a forward-looking statement, including those relating to market growth, products and their capabilities, performance and compatibility, cost savings and other benefits to customers. Information in this presentation may also include or be based upon information from third parties, which reflects their expectations and projections as of the date of issuance.

We undertake no obligation to update these forward-looking statements, which speak only as of the date hereof.
Mega Trend: Legacy Storage I/O Bottleneck

Non-Volatile Memory (NVM)

- **Today:** NAND Flash
  - Capacity: 100s of GB to 100s of TB per device
  - Trends: Higher capacity, lower cost/GB, lower write cycles, SLC->MLC->3BPC
  - IOPS: 100K to millions, GB/s of bandwidth

- **Tomorrow:** Non-Volatile Memory technologies (Phase Change Memory, MRAM, STT-RAM, etc.)
  - Potential for orders of magnitude performance improvement
Why Do Applications Need Optimization for Flash?

- Many applications assume high latency storage (some even optimize for read/write head positioning)
- Flash is different from disk
  - Performance, endurance, addressing
  - Getting more different over time
- Flash-focused application acceleration
  - Shifting bottlenecks (Compute to I/O to Network to Application)
  - Managed writes = greater device lifetime (wear leveling, endurance)
  - Improved system efficiency (TCO and TCA)
  - Even lower power and cooling costs

<table>
<thead>
<tr>
<th>Area</th>
<th>Hard Disk Drives</th>
<th>Flash Devices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read/Write Performance</td>
<td>Largely symmetrical</td>
<td>Heavily asymmetrical.</td>
</tr>
<tr>
<td>Sequential vs Random Performance</td>
<td>100x difference.</td>
<td>&lt;10x difference.</td>
</tr>
<tr>
<td>Background ops</td>
<td>Rare</td>
<td>Regular</td>
</tr>
<tr>
<td>Wear out</td>
<td>Largely unlimited</td>
<td>Limited writes</td>
</tr>
<tr>
<td>IOPS</td>
<td>100s to 1,000s</td>
<td>100Ks to Millions</td>
</tr>
<tr>
<td>Latency</td>
<td>10s milli sec</td>
<td>10s-100s micro sec</td>
</tr>
<tr>
<td>Addressing</td>
<td>Sequence, Sector</td>
<td>Direct, byte addressable</td>
</tr>
</tbody>
</table>
Becoming “Flash Aware”: SanDisk NVMFS

Non-Volatile Memory File System – Optimized for Flash and Beyond

Value

- Increase life expectancy of flash devices
- Consistent low latency
- Consistent high performance

How

- Reducing MySQL™ Writes to flash
- Optimize IO Write path for flash
- Applications leverage enhanced I/O interface

Available today!

Eliminating Duplicate Logic and Leveraging New Primitives for Optimal Flash Performance and Efficiency
In-Memory Database and Persistence

Checkpoints, Logs and Data Tiering, Oh My!

Persistent Storage

State Changes

Mem Image Checkpoints

Logs

In-Memory Compute
(transactions, processing, analytics)

“Hot Data”

Data Tiering

Restore Data into memory

Backup and Archiving

Backup Systems

Possible wait conditions

Transactions over Time

Transaction Commit

Warm Data

Persistent Snapshots

Logs
Emergence of Flash Extended Memory

Decoupling of the physical infrastructure (memory) from the management and utilization of that infrastructure through a software interface

**Flash Extended Memory**

Memory Addressable and I/O Addressable

Storage Device Technologies:
- Memory
- Persistent Memory
- Flash Devices
- HDDs

“Flash as Memory”

“Flash as Disk”

Traditional Block I/O
Flash Extended Memory Example: MongoDB – Lowering TCO through transparent DRAM/Flash tiering

**MongoDB Throughput (operations/second)**

- 24 GB Node
- 12 GB Node: -26%
- 8 GB Node: -33%

**Throughput Deviation (operations/second)**

- 24 GB Node
- 12 GB Node: 3x Improvement
- 8 GB Node

YCSB Workload A

DRAM reductions of 2x and 3x yield 26% and 33% throughput degradations respectively. Throughput predictability actually improves with less DRAM. Latency shows similar trend.

Source: Based on internal testing by SanDisk; Nov 2014
Storage Memory Convergence: Rethinking the Memory Hierarchy

Accessed Like Memory and Managed Like Storage

New Memory-Storage Hierarchy

L1, L2, L3 CPU Caches

Main Memory System

DRAM

Persistent Memories

High Performance I/O System

Flash

Hard Drive

Range of persistence – I/O becomes Load-Store

ACCESS DELAY

Nanoseconds

| Microseconds |

Milliseconds

CYCLES TO WAIT

2 cycles

4 million cycles

2 cycles
Implications for Applications

- Extremely low latency transaction commits
  - 10-100us reduced to <1us
- Accelerate logs, indexes etc.
- Convergence of in-memory and disk approaches
  - New data structures to optimize directly for memory access
- Rich variety of programming models
  - Traditional I/O will still be available
  - Mmap directly to persistent memory is alternative model: use pointer operations to manipulate persistent data
  - SNIA Non-Volatile Memory Programming TWG – standardize mmap for persistent memory
- New challenges
  - CPU cache management
  - Atomicity?
Technology Preview: Database Acceleration Through Flash Extended Memory

- A “transparent” Software Defined Memory layer can provide accelerated I/O over a “baseline” unaware interface
- But “flash-aware” applications can optimize that acceleration
Technology Preview Example Configuration

“Baseline”
- Fusion ioMemory™ PCIe-based flash
- Standard MySQL database
- Standard Flash Acceleration
- HP ProLiant DL380 Gen8

“Transparent”
- Fusion ioMemory™ and persistent memory with NVMFS
- Standard MySQL database
- NVMFS and ACM
- HP ProLiant DL380 Gen8

“Flash Aware”
- Fusion ioMemory™ and persistent memory with NVMFS
- Optimized MySQL database
- NVMFS and ACM
- HP ProLiant DL380 Gen8

Flash Extended Memory Enabled
Performance Results

Latency (lower is better)

Throughput (higher is better)

Source: Based on internal testing by SanDisk; Nov 2014
Advantages & Benefits

- Improve “Baseline” MySQL throughput performance by roughly 60% via “Transparent” acceleration (no software mods)

- Optimize MySQL throughput performance by over 3x with “Flash Aware” acceleration (modified software)

- Improve “Baseline” MySQL latency by roughly 2.3x (Transparent) and optimized latency by more than 4x (Flash Aware)

- Uses “Flash-as-Memory” byte-addressable architecture and interface

- Seamless deployment – add ioMemory and NVMFS/ACM software to Linux

- Increase performance and capacity in flexible configurations

Source: Based on internal testing by SanDisk; Nov 2014
In-Memory Computing – New Approach to I/O

In-Memory Compute
(transactions, processing, analytics)
“Hot Data”

Transactions over Time

Flash Extended Memory

Flash Devices
POSIX File System

Memory
Persistent Memory

Persistence via CPU Load/Store

Backup Systems
Restore Data into memory
Backup and Archiving
Thank You!

@BigDataFlash
#bigdataflash
ITblog.sandisk.com
http://bigdataflash.sandisk.com

©2015 SanDisk Corporation. All rights reserved. SanDisk is a trademark of SanDisk Corporation, registered in the United States and other countries. InfiniFlash, SanDisk ION Accelerator, Fusion ioMemory, ioDrive, Cloudspeed MAX, Optimus MAX and Lightning Ultra are trademarks of SanDisk Enterprise IP LLC. All other product and company names are used for identification purposes and may be trademarks of their respective holder(s).