Flash Memory Summit 2018

Persistent Memory - NVDIMMs

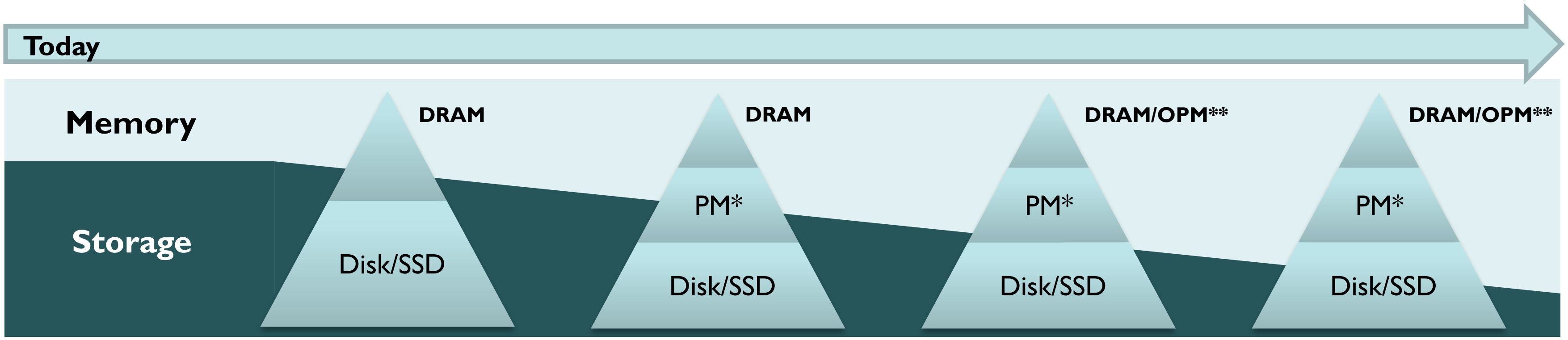
Contents

- Persistent Memory Overview
- NVDIMM
- Conclusions

Persistent Memory

Memory & Storage Convergence

- **Volatile and non-volatile technologies are continuing to converge**



*PM = Persistent Memory

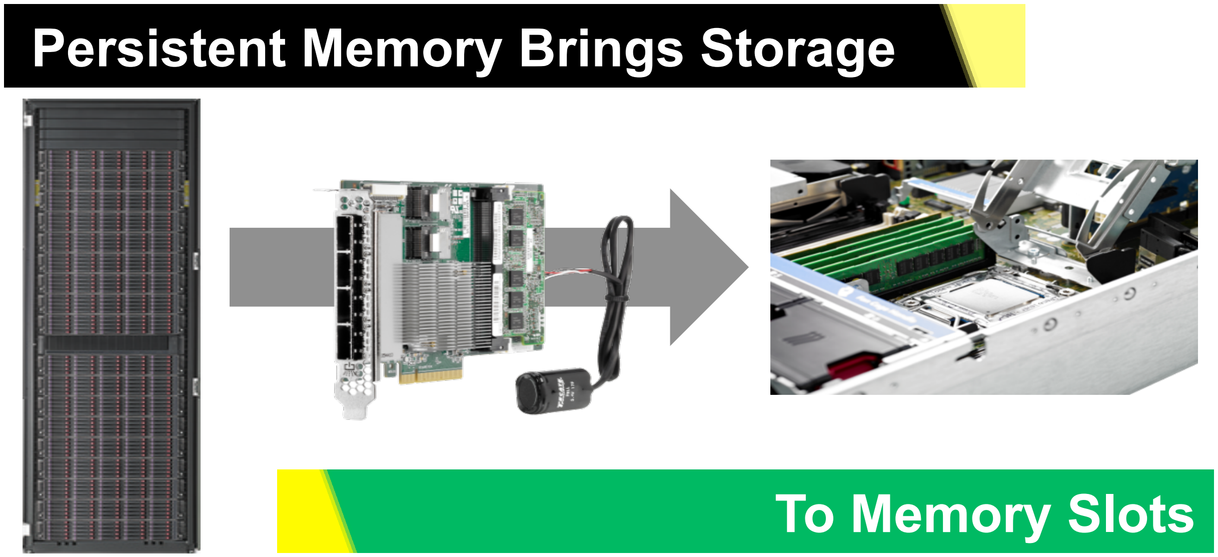
**OPM = On-Package Memory

New and Emerging Memory Technologies

HMC	3DXPoint^{TM} Memory	Low Latency NAND
HBM	MRAM	Managed DRAM
RRAM	PCM	

Source: Gen-Z Consortium 2016

Persistent Memory (PM) Vision



Persistent Memory Brings Storage

To Memory Slots

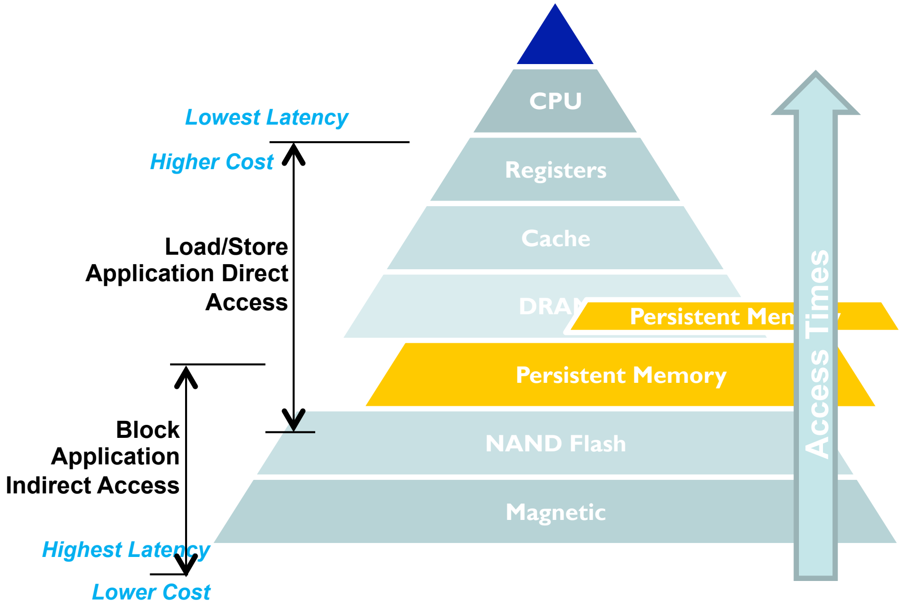
Fast
Like Memory

Persistent
Like Storage

- For system acceleration
- For real-time data capture, analysis and intelligent response

Persistent Memory

- Bridges the gap between DRAM and Flash
- Dramatically increases system performance
- Enables a fundamental change in computing architecture
- Apps, middleware and OSs are no longer bound by file system overhead in order to run persistent transactions



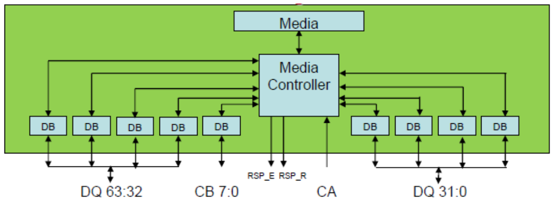
NVDIMM

Persistent Memory - NVDIMMs

NVDIMM-N

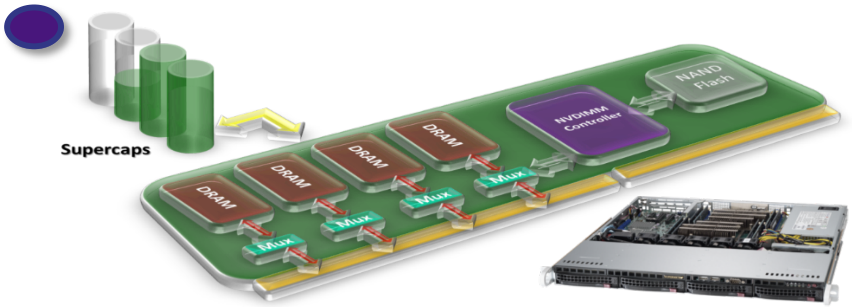
- Host has direct access to DRAM
- CNTLR moves DRAM data to Flash on power fail
- Requires backup power
- CNTLR restores DRAM data from Flash on next boot
- Communication through SMBus
- Byte-addressable DRAM for lowest latency with NAND for persistence backup

NVDIMM-P

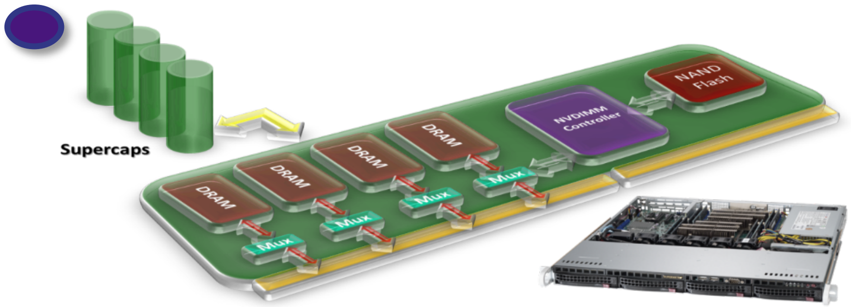


- NVDIMM-P interface specification targeting persistent memories and high capacity DRAM memory on DDR4 and DDR5 channels
- Extends the DDR protocol to enable transactional access
- Host is decoupled from the media
- Multiple media types supported
- Supports any latency (ns ~ us)
- JEDEC specification publication in 2018

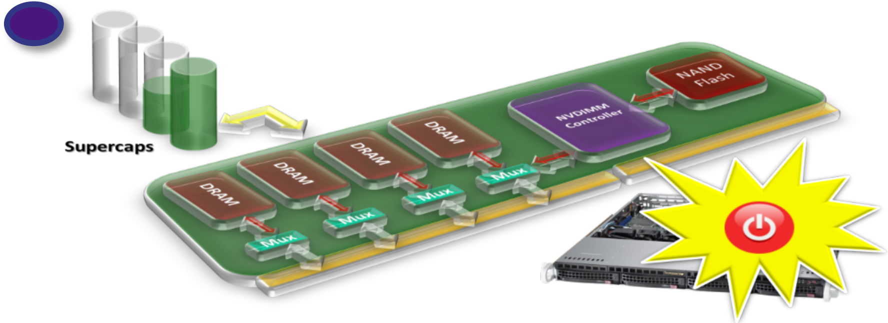
NVDIMM-N How It Works



- *Plugs into JEDEC Standard DIMM Socket*
- *Appears as standard RDIMM to host during normal operation*
- *Supercaps charge on power up*

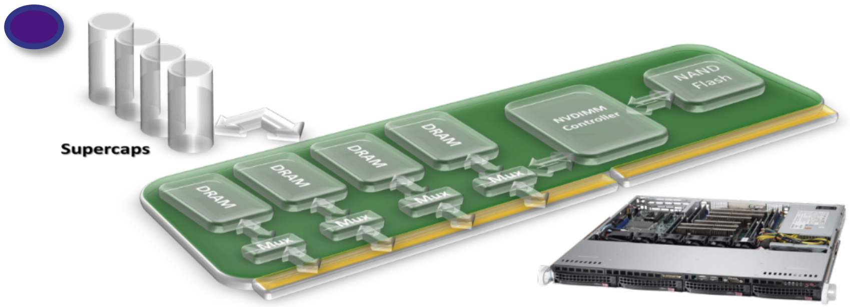


- *When health checks clear, NVDIMM can be armed for backup*
- *NVDIMM can be used as persistent memory space by the host*

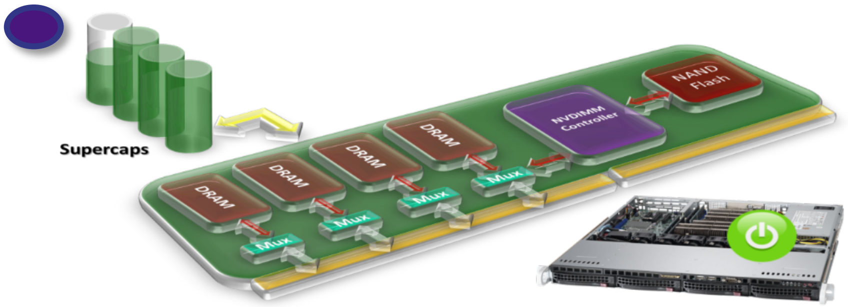


- *During unexpected power loss event, DRAM contents are moved to NAND Flash using Supercaps for backup power*

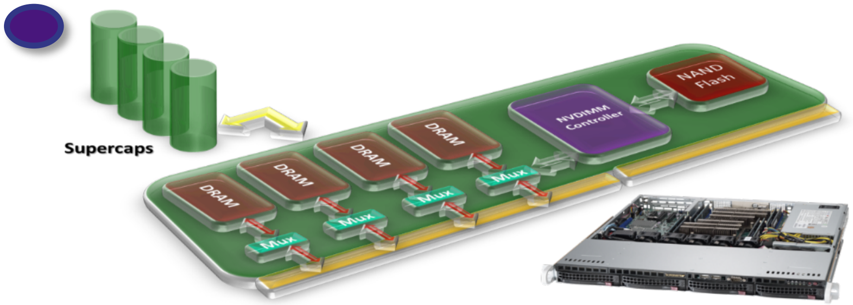
NVDIMM-N How It Works



- *When backup is complete, NVDIMM goes to zero power state*
- *Data retention = NAND Flash spec (typically years)*



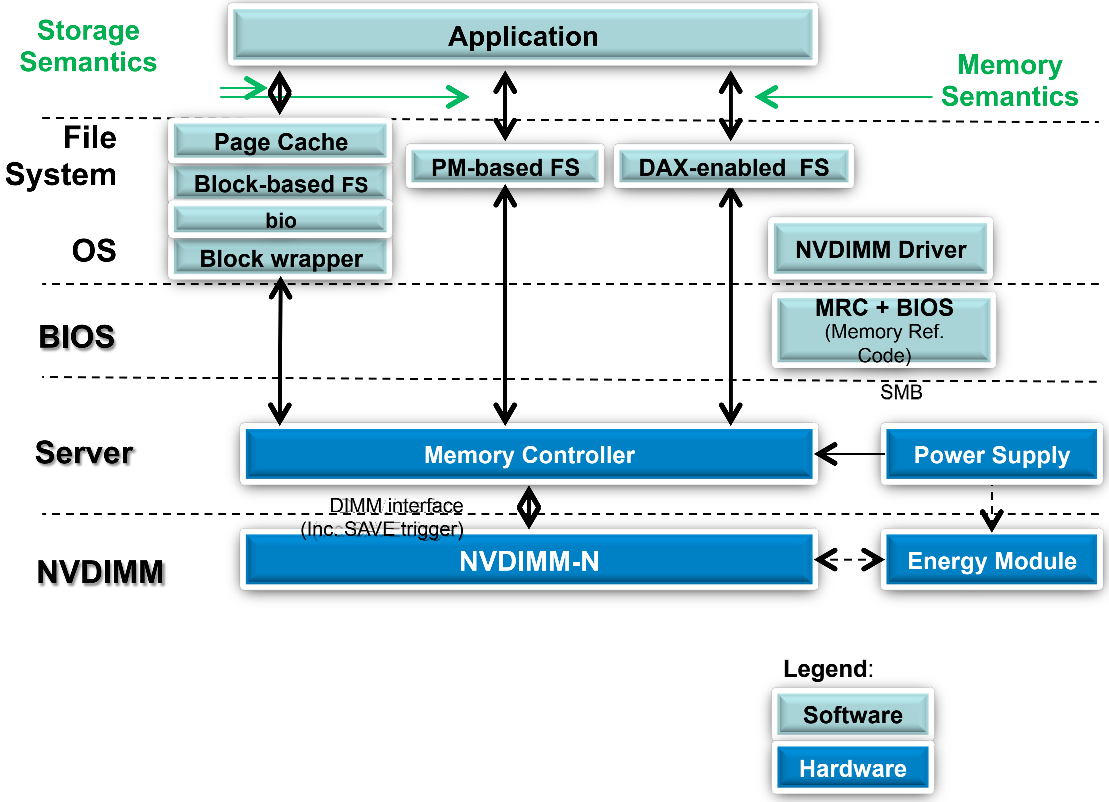
- *When power is returned, DRAM contents are restored from NAND Flash*
- *Supercaps re-charge in minutes*



- *DRAM handed back to host in restored state prior to power loss*

NVDIMM Ecosystem

- Standardized through NFIT and JEDEC
- Linux 4.4+ kernels have the software stack
- Open source library is available for applications



NVDIMM-N BIOS/MRC Support Functions

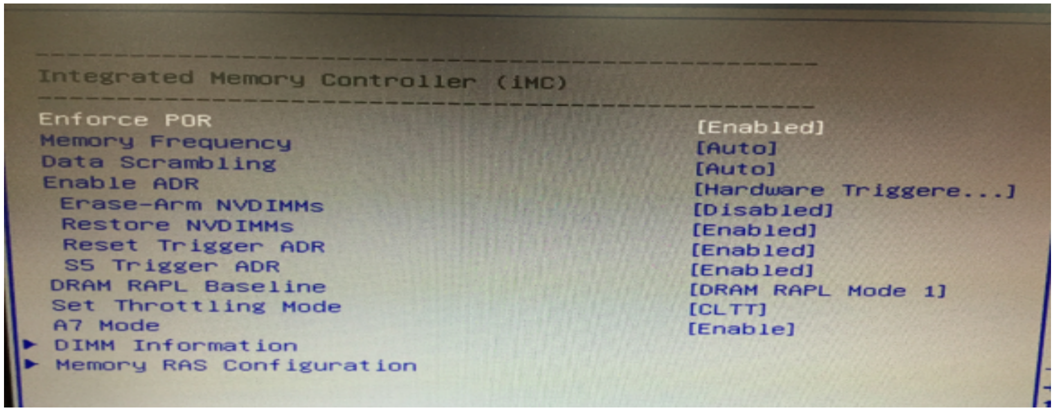
NVDIMMs BIOS/MRC (Memory Reference Code)

1. Detect NVDIMMs
2. Setup Memory Map
3. ARM for Backup
4. Detect AC Power Loss or BMC/CPLD Triggered ADR
5. Flush Write Buffers
6. RESTORE Data On Boot
7. Enable I2C R/W Access

Source: SNIA Persistent Memory and NVDIMM SIG

Additional BIOS Settings

- BIOS also presents various menu options to setup NVDIMM operation
- Configuration:
  - Erase-Arm NVDIMM
  - Restore NVDIMM
  - Reset Trigger ADR
  - S5 Trigger ADR



```
Integrated Memory Controller (iMC)

Enforce POR                                   [Enabled]
Memory Frequency                              [Auto]
Data Scrambling                               [Auto]
Enable ADR                                    [Hardware Triggere...]
  Erase-Arm NVDIMMs                           [Disabled]
  Restore NVDIMMs                             [Enabled]
  Reset Trigger ADR                           [Enabled]
  S5 Trigger ADR                              [Enabled]
 DRAM RAPL Baseline                           [DRAM RAPL Mode 1]
 Set Throttling Mode                          [CLTT]
 A7 Mode                                      [Enable]
▶ DIMM Information
▶ Memory RAS Configuration
```

Linux Kernel 4.4+
NVDIMM-N OS Support

- Linux 4.2 + subsystems added support of NVDIMMs. Mostly stable from 4.4
- NVDIMM modules presented as device links: /dev/pmem0, /dev/pmem1
- QEMO support (experimental)
- XFS-DAX and EXT4-DAX available

DAX

File system extensions to bypass the page cache and block layer to memory map persistent memory, from a PMEM block device, directly into a process address space.

BTT (Block, Atomic)

Block Translation Table: Persistent memory is byte addressable. Existing software may have an expectation that the power-fail-atomicity of writes is at least one sector, 512 bytes. The BTT is an indirection table with atomic update semantics to front a PMEM/BLK block device driver and present arbitrary atomic sector sizes.

PMEM

A system-physical-address range where writes are persistent. A block device composed of PMEM is capable of DAX. A PMEM address range may span an interleave of several DIMMs.

BLK

A set of one or more programmable memory mapped apertures provided by a DIMM to access its media. This indirection precludes the performance benefit of interleaving, but enables DIMM-bounded failure modes.

Windows
NVDIMM-N OS Support

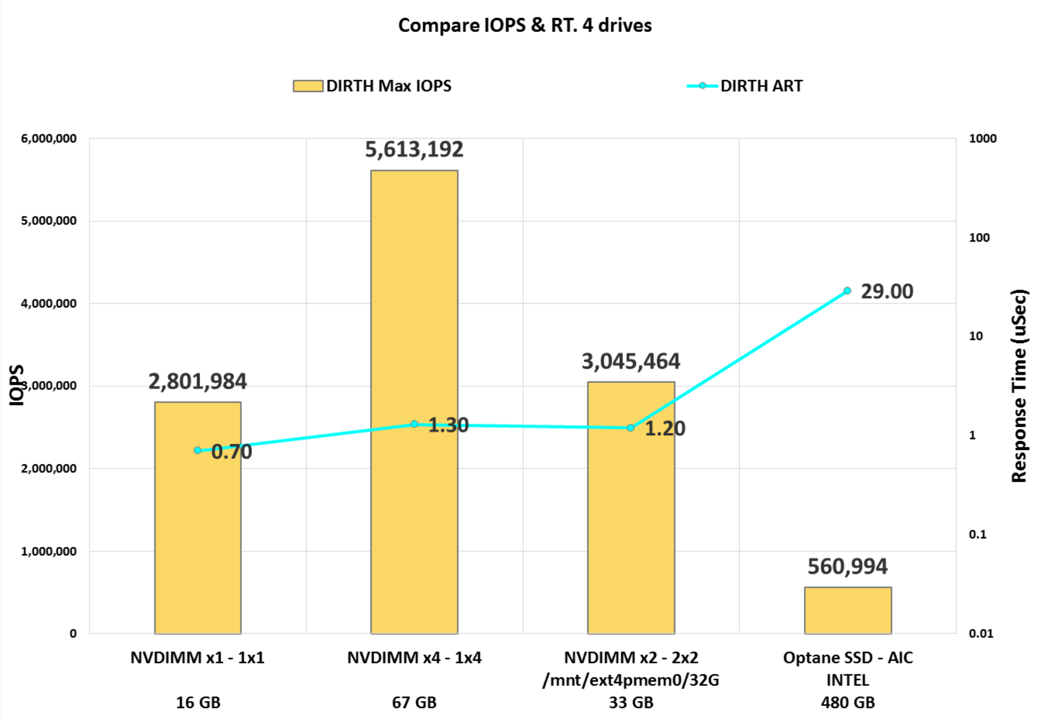
- Windows Server 2016 supports DDR4 NVDIMM-N
- Block Mode
  - No code change, fast I/O device (4K sectors)
  - Still have software overhead of I/O path
- Direct Access
  - Achieve full performance potential of NVDIMM using memory-mapped files on Direct Access volumes (NTFS-DAX)
  - No I/O, no queueing, no async reads/writes
- More info on Windows NVDIMM-N support:
  - https://channel9.msdn.com/events/build/2016/p466
  - https://channel9.msdn.com/events/build/2016/p470

4K Random Write	Thread Count	IOPS	Latency (us)
NVDIMM-N (block)	1	187,302	5.01
NVDIMM-N (DAX)	1	1,667,788	0.52

Technology Comparison

Technology	FeRAM	MRAM	ReRAM	PCM	3D Xpoint	NAND Flash	DRAM NVDIMM
Endurance	10^{12}	10^{12}	10^{6}	10^{8}	10^{6} - 10^{7}	10^{3}	10^{15}
Byte Addressable	yes	yes	yes	yes	yes	no	yes
Latency R/W	70ns-100ns	70ns/70ns	100ns/100µs	20ns/65ns	100ns/500ns	10µs/10µS	40-140ns
Power Consumption	Low	Medium/ Low	Low	Medium	Medium	Low	Medium
Interface	DRAM	DDR3 DDR4	Flash-like	Proprietary	Proprietary	Toggle ONPHI	DDR3 DDR4
Density Path	Low	Gigabit+	Terabit	64Gb+	64Gb+	Gigabit+	Gigabit+

NVDIMM Performance Comparison



- **Test Platform: Supermicro X11DRI 16GB DDR4 2400 Mhz RDIMM RAM, Intel XEON 8160 2.1 Ghz 24 core, 16 GB DDR4 JEDEC NVDIMM-N. 480GB Optane SSD**
- **Software: Ubuntu 16.04.3 LTS Linux 4.10.0-28; DAX File System**
- **Test Software: Calypso CTS 7.0 fe 1.26.25 be 1.9.317**

How NVDIMM-N’s Improve Performance

- NVDIMM-Ns are byte addressable. This allows databases to be built in memory
- With direct access to records this removes disk IO and all the overhead that involves
- A memcached structure is dramatically faster than even the best solid-state solution, with updates just requiring a register-to-memory computer instruction instead of the file stack and interface overhead
- Since this looks like DRAM to the system, using RDMA to create redundancy and cluster sharing is a given, with existing designs working just fine

NVDIMM-N Performance

- NVDIMMs provide 34 times the number of IOPS compared with standard SSDs, with 16 times the bandwidth and 81 times lower latency
- Streaming data applications can be architected to greatly benefit from this marriage of memory and storage

Source: Microsoft

NVDIMM Use Cases

- **In Memory Database**
  - Journaling, reduced recovery time, tables

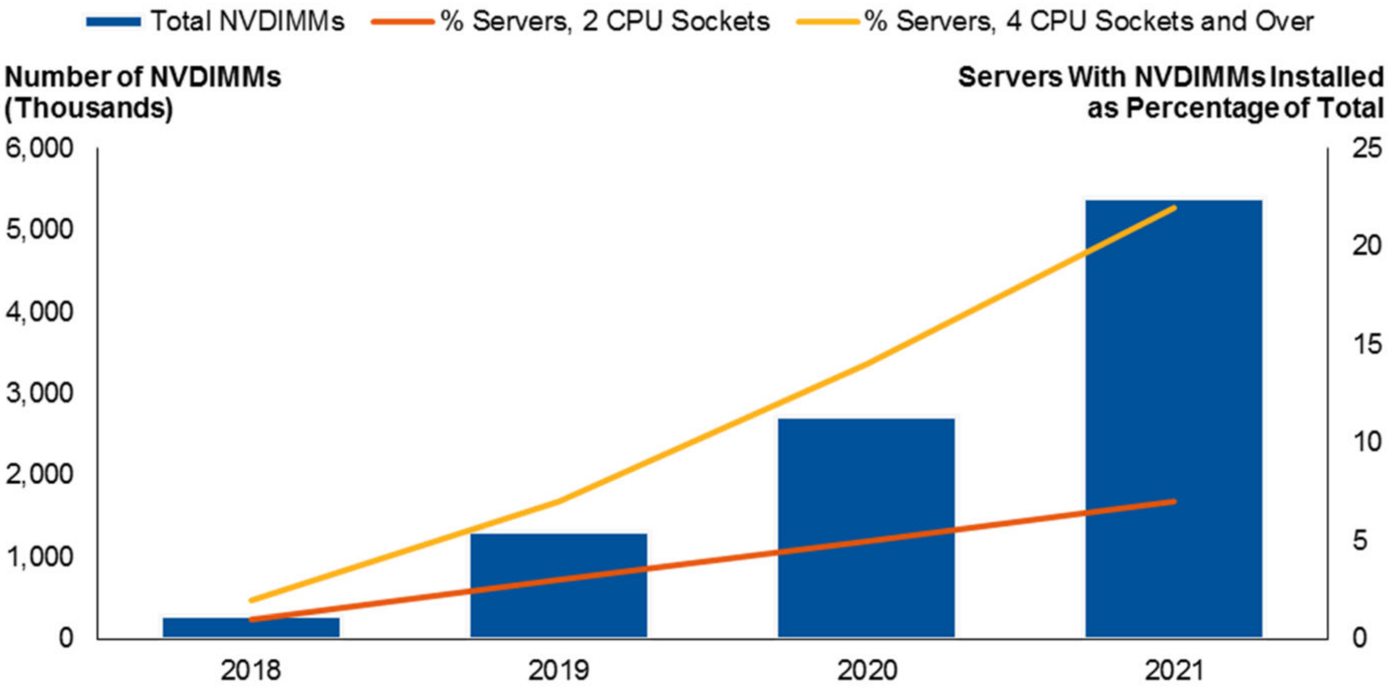
- **Traditional Database**
  - Log acceleration by write combining and caching

- **Enterprise Storage**
  - Tiering, caching, write buffering and meta data storage

- **High-Performance Computing**
  - Check point acceleration and/or elimination

What is the Outlook?

Nonvolatile Memory Shipments as a Percentage of Server Memory (PB) and NVDIMM Unit Forecast*



* Excludes NVDIMMs deployed in solid state arrays

Source; Gartner 2017

Infrastructure Changes

- Operating System
- File system changes for memory mapped files
- Memory Management software
- Hypervisors
- Containers
- Allocation of Persistent Memory to Guests
- Coordinating with Guest’s use of Persistent Memory
- User space libraries supporting Persistent Memory
- Support for legacy interfaces with Persistent Memory-aware implementations
- Securing application data in a multi-tenant environment

Persistent Memory Standards

- **JEDEC JESD 245, 245B: Byte Addressable Energy Backed Interface**
  - Defines the host to device interface and features supported for a NVDIMM-N

- **ACPI 6.2**
  - NVDIMM Firmware Interface Table (NFIT)
  - NVM Root and NVDIMM objects in ACPI namespace
  - Address Range Scrub (ARS)
  - Uncorrectable memory error handling
  - Notification mechanism for NVDIMM health events and runtime detected uncorrectable memory error

Source: SNIA Persistent Memory Summit January 2018

Encryption

- Estimated 10-20% of NVDIMM end-users require encryption
  - Financial – high-speed trading, OLTP
  - Public – DoD
  - Health – medical records
  - Private - corporate IT departments
- With block access NVDIMMs the controller chip can manage encryption in the same way as SSDs
- With byte access NVDIMMs the host memory controller needs to provide encryption support
- A key is supplied by the host to support backup with encryption (which could impact performance)
- During a system power loss, in-flight data written from the DRAM to the Flash will be encrypted

Key Takeaways

SNIA®

- Workloads are being re-architected to use large amounts of data placed in local memory
- Data reload times are significant, driving a need to retain data through a power failure
- More Persistent Memory technologies are emerging
- Applications help drive demand for Persistent Memory
- Standardization enables wider adoption of Persistent Memory-aware applications
- SNIA Persistent Memory and NVDIMM SIG is driving education and adoption

SNIA Persistent Memory and NVDIMM SIG

snia.org/pm

- Charter
  - To accelerate the awareness and adoption of Persistent Memories and NVDIMMs for computing architectures
- Activities
  - Educate on the types, benefits, value, and integration of Persistent Memories
  - Communicate usage of the NVM Programming Model developed to simplify system integration of current and future PM technologies
  - Influence and collaborate with middleware and application vendors to support Persistent Memories
  - Develop user perspective case studies, best practices, and vertical industry requirements
  - Coordinate with industry standards groups and promote industry standards related to PM and NVDIMM
  - Synchronize and communicate a common Persistent Memory taxonomy
- Membership
  - All companies with an interest in Persistent Memory and NVDIMM are welcome to join and participate – snia.org/join for more details.

Thanks for Attending

Questions?

Visit www.snia.org/pm

for Persistent Memory videos, webcasts, and presentations

Backup Slides