Containers and Persistent Memory

Jeff Chang
Co-Chair
SNIA Persistent Memory & NVDIMM Special Interest Group
AgigA Tech, Inc.
SNIA-At-A-Glance

185 industry leading organizations

2,000 active contributing members

50,000 IT end users & storage pros worldwide

Learn more: snia.org/technical  @SNIA
What SNIA Does

20 YEARS of Standards Development

✓ ISO & ANSI Standards
✓ Storage Standards
✓ Best Practices & Security
✓ Interoperability & Conformance Testing
Areas of Focus

PHYSICAL STORAGE
- Solid State Storage
- Hyperscaler Storage
- Object Drives
- Connectors, Form Factors & Transceivers

DATA MANAGEMENT
- Protection
- Integrity
- Retention

DATA SECURITY
- Storage Security
- Privacy and Data Protection Regulations

DATA IN THE CLOUD
- Data Orchestration
- Data into and out of the Cloud

PERSISTENT MEMORY
- NVDIMMs
- Non-Volatile Memory Programming Model

POWER EFFICIENCY MEASUREMENT
- SNIA Emerald™ Power Efficiency

NEXT GENERATION DATA CENTER
- Software Defined Storage
- Composable Infrastructure
- Next Generation Storage Management API

NETWORKED STORAGE
- Data Access Protocols
- Networking Technologies for Storage

STORAGE MANAGEMENT
- Device and System Management
A Confession...
Terms of Art

PERSISTENT MEMORY
Terms of Art

PERSISTENT MEMORY

Data Survives Over Power Loss
Terms of Art

PERSISTENT MEMORY

Data Survives Over Power Loss

Load/Store Semantics

Byte Addressable
Terms of Art

PERSISTENT MEMORY

- Data Survives Over Power Loss
- Byte Addressable Load/Store Semantics

128B Magnetic Core Memory - Circa 1955
(source: Wikipedia, courtesy of Konstantin Lanzet)
Terms of Art

NVDIMM = Non-Volatile Dual Inline Memory Module
A computer memory module that maintains data persistence across power cycles

SSD : NAND Flash :: NVDIMM : Persistent Memory

There Are Multiple NVDIMM Types,
More On That In A Minute
Compute Architecture: The Last Few Decades

- **Latency (ns)**
  - CPU: $10^0$
  - DRAM: $10^1$
  - Performance SSD: $10^4$
  - Capacity SSD: $10^5$
  - HDD: $10^6$

In Human Observable Terms
- Seconds
- Hours
- Days
- Weeks

#ContainerWorld
@ContainerWrld
https://tmt.knect365.com/container-world/
Compute Architecture: Why PM Changes Everything!

- Eliminate/Reduce Latency Gap (depending on PM type)
- Memory Channel Speed >> Storage Interface Speed
- PM = Byte Addressable Storage = Block Access
- PM eliminates burden of significant SW latency

Latency (ns)

- CPU: $10^0$
- DRAM: $10^1$
- NVDIMM: $10^2$
- Performance SSD: $10^4$
- Capacity SSD: $10^5$
- HDD: $10^6$

In Human Observable Terms:
- Seconds
- Hours
- Days
- Weeks

https://tmt.knect365.com/container-world/
JEDEC Defined NVDIMM Types

**NVDIMM-N**
- Host has direct access to DRAM
- Flash is only used for backup (requires energy source)
- Capacity = DRAM (10's - 100's GB)
- Latency = DRAM (10's of nanoseconds)
- Endurance = DRAM (infinite)
- No impact to memory bus performance
- Standards completed and released

**NVDIMM-P**
- Host is decoupled from the media (agnostic to PM type)
- Extends DDR protocol to enable non-deterministic timing
- Extended addressing for higher capacities
- Capacity = PM Media (100's GB+)
- Latency = PM Media (>10's of nanoseconds)
- Endurance = PM Media (may be finite)
- Standards still under definition

**NVDIMM Types Are Complementary, Servicing Different Workloads**
Many Flavors Of PM

Address granularity

Endurance

Write latency

Read latency

Cost per bit

Density

Power
Maturity
Retention

Carbon Nanotubes
Graphene Memories
CeRAM
Polymeric ferroelectrics
FTJ
FeFETs
MeRAM
Racetrack Memory

Source: ARM, 2019 PM Summit
What About 3DXPoint?
Persistent Memory Use Cases

**USE CASES**
- Log Acceleration
- Storage Tiering
- Fast Caching
- SSD Wear-Out
- Higher VM Consolidation
- More Virtual Users/System
- Fast IOPS Workloads
- In-Memory Processing
- Byte-Level Data Processing
- Metadata Store
- Low Latency Look-Up & Processing

**Database**
- Storage
- Virtualization
- Big Data
- Cloud Computing/ IoT
- Artificial Intelligence

The same factors driving NAND Flash adoption apply to NVDIMMs: IOPS, Latency, Performance AND it looks just like DRAM!

Where Do Containers Fit?
Performance Comparison vs SSD

IOPS RND 4K Writes & Reads: NVDIMM-N v U.2 v SAS

IOPS = I/O Operations Per Second

Higher is Better

Source: Calypso Systems

All data taken from PTS-E v1.1 DIRTin Tests using CTS test software. SAS and U.2 SSDs tested on Calypso RTP Intel i52600GCE, Dual 2667W 8 core 3.2-GHz, 32GB DDR3 RAM. Four NVDIMM-N Modules tested on SuperMicro X10DR, Dual E5 2670v3, 32GB DDR4 RAM with Intel Open Source NVDIMM-N Development Block I/O Driver and CTS test Software.
Reducing Core Count for DB Applications

Database throughput
SAS SSDs with 32 Cores vs NVDIMMs with 16 Cores

NVDIMM with 16 cores performed better than SAS SSDs with 32 cores, reducing DB licenses by 50%
“A Hard Drive with an NVDIMM is faster than SSD”

- Microsoft SQL Server Benchmark*

**SQL Server “Tail of the Log”**

**Note: Log scale**

**HDD + NVDIMM Performance >> SSD**

- SQL Server 2016
- Writes updates to SQL log through NVDIMM first (“hardened log”)
- Uses memory instructions to issue log updates to NVDIMM directly

Pairing NVDIMM with SSD provides the best performance AND reduces SSD wear out

Doesn’t Such A Fundamental Change Require An Ecosystem? YES!

- PM support native in Windows Server, Linux, vSphere
- Others still adopting

- Multiple vendors shipping NVDIMM-N today
- Successful demonstrations of interoperability among vendors

- JEDEC JESD245: Byte Addressable Energy Backed I/F
- JEDEC JESD248: NVDIMM-N Design Standard
- uEFI ACPI NVDIMM Firmware Interface Table
- SNIA NVM Programming Model
- NVDIMM Special Interest Group

- All major OEMs shipping platforms with NVDIMM support
- Requires hardware and BIOS mods

#ContainerWorld
@ContainerWrld
https://tmt.knect365.com/container-world/
The SNIA NVM Programming Model

• A vendor/media/platform agnostic programming model
• It is not an API
  • Describes application visible behaviors: attributes, actions and use cases
  • Allows API’s to align with OS’s
• Block and File modes use IO
• Volume and PM modes enable Load/Store
The SNIA NVM Programming Model

APPLICATION

FILE SYSTEM

DISK DRIVER

STORAGE

APPLICATION

PM-AWARE FILE SYSTEM

NVDIMM DRIVER

NVDIMM

USER SPACE

KERNEL SPACE

HARDWARE

BLOCK ACCESS

BYTE ACCESS

#ContainerWorld
@ContainerWrd
https://tmt.knect365.com/container-world/
Containers and Persistent Memory

“46% of containers will require persistent storage within the next 3 years” – IDC Infobrief, Sep’18

- Persistent Memory is exposed to container engines through the OS
- There’s no such thing as a stateless architecture. State in your application is stored somewhere.
- Application data (e.g. databases, message queues, instrumentation) will require dedicated persistent storage AND guaranteed performance
- Application provisioning: creating persistent volumes and policies for stateful data and management