SNIA PERSISTENT MEMORY PM SUMMIT

JANUARY 18, 2017 | SAN JOSE, CA

Storage at Memory Speed and Amazing Future of Virtual Non-Volatile Memory

Rajesh Venkatasubramanian, Principal Engineer, VMware

Disclaimer

- This presentation may contain product features that are currently under development
- Features are subject to change, and must not be included in contracts, purchase orders, or sales agreements of any kind
- Technical feasibility and market demand will affect final delivery
- Pricing and packaging for any new technologies or features discussed or presented have not been determined

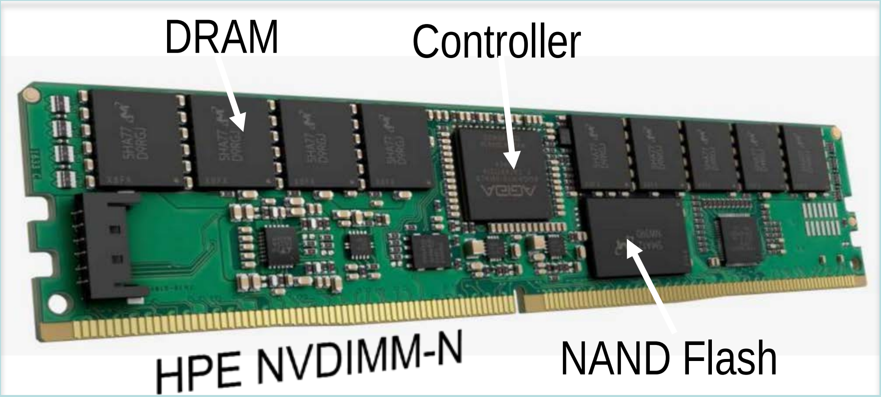
The information in this presentation is intended to outline our general product direction and it should not be relied on in making a purchasing decision. It is for informational purposes only and may not be incorporated into any contract.

Storage at Memory Speed

- Persistent Memory (PMEM or NVM)
  - Fundamental change in storage architecture happening now
  - e.g., 3D XPoint^{TM}, HPE NVDIMM-N
- Persistent Memory is Storage at…
  - DRAM latency: a few hundred nanoseconds latency
  - DRAM bandwidth: a few GBs of bandwidth
  - DRAM granularity: byte-level access
  - DRAM model for software: load/store instructions

CPU/Memory vs. Storage Speed

- Software Solutions
  - Use volatile DRAM as a large cache
  - Employ complex schemes to deal with power failure
  - Complicate code by using Asynchornous IO to hide latency
- Hardware Solutions (HPE NVDIMM, 3D XPoint^{TM})



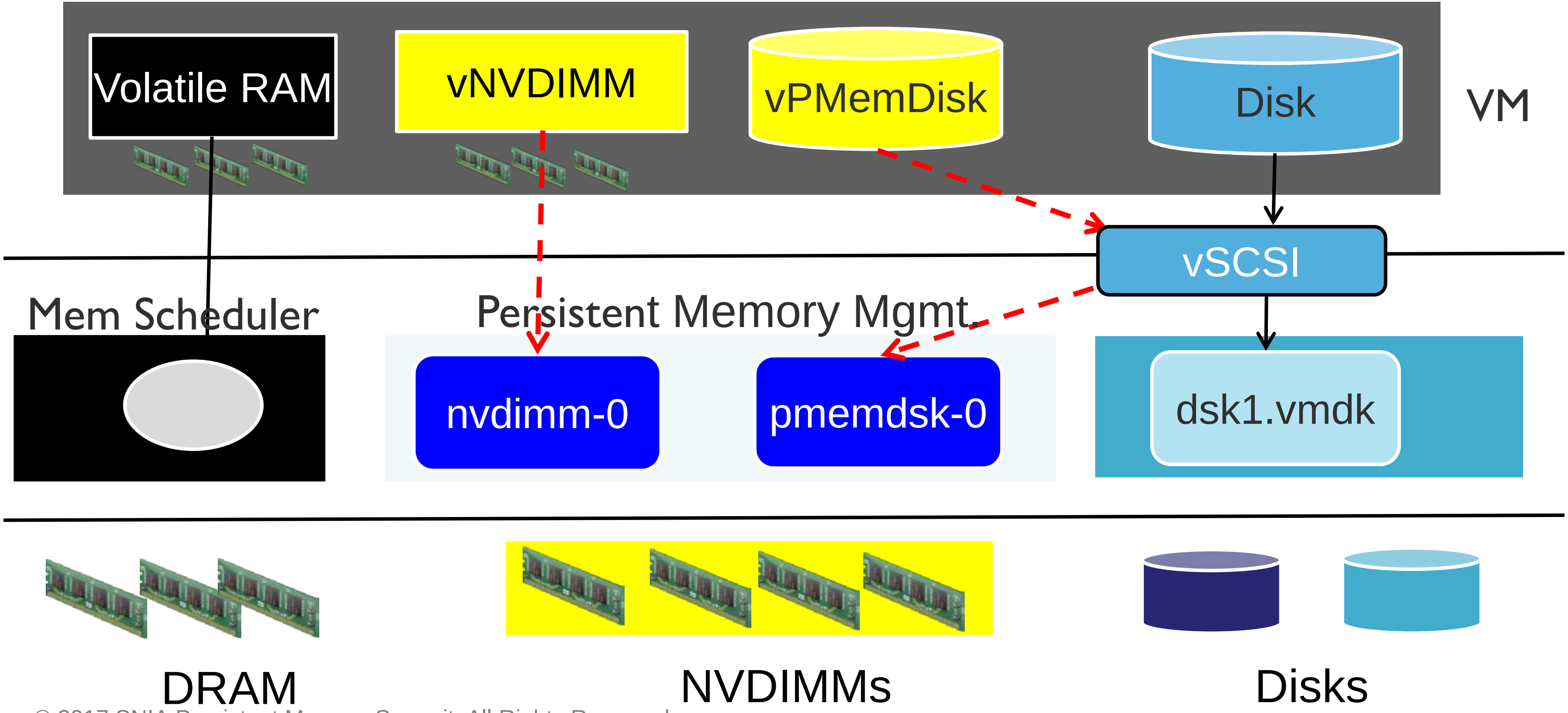
Operating Systems and Applications

- Operating Systems
  - Windows Server 2016, Fedora 24, RHEL 7.3, etc.
  - Provide Direct Access (DAX) mode access for applications
- Applications
  - Legacy: mount volume in non-DAX mode, use file access
  - PMem-optimized: use DAX mode, mmap file, byte-level updates
    - e.g., SQL Server 2016, PMem-aware Redis

Concept: vSphere Support for NVM

- Enable use of persistent memory hardware
- Support for legacy VMs
  - No change to guest OS/app – just a simple VM config change
- Expose virtual NVDIMMs
  - Byte-addressable with similar performance as physical NVDIMM
- Ease management by enabling vMotion and FT
- Help consolidation by intelligent cluster management

Where we are headed…



Physical NVDIMMs

- BIOS Requirements
  - NFIT, Namespace DSMs
  - Health information, ARS error reporting/clearing, MCEs
  - Block mode or BTT is not used
- Host-local Persistent Memory Datastore
  - Concatenates multiple namespaces (extents) to form a volume
  - Exposes a single persistent memory datastore per host
  - Plans to expose a single datastore even with different NVDIMMs

vPMemDisk and vNVDIMM

- Virtual Persistent Memory Disk
  - Guest accesses a regular vSCSI or vNVMe disk
  - Virtual disk is stored in persistent memory datastore
  - Provides atomic 512 byte block writes
- Virtual NVDIMM
  - Virtual BIOS exposes NVDIMM via NFIT, namespace DSMs, etc
  - PMem-aware OS and applications can run unmodified
  - Multiple virtual NVDIMMs can be attached to a VM
  - Almost zero overhead because ESXi avoids intercepts

Error Handling

- Physical Errors
  - Learns about errors via ARS records, MCEs, ACPI events
  - Avoids mounting if volume meta-data is corrupted
  - Marks data page errors permanently till poison clear or full write
- Exposing Errors
  - Exploring ways to expose virtual NVDIMM errors to guest
  - Clearing errors if vPMemDisk block IO covers error blast radius

Workload Mobility and Availability

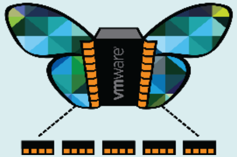
- Migration
  - Support migration of VMs with vPMemDisk and vNVDIMM
  - Changing host of a VM results in copying PMem contents
  - Virtual NVDIMM is always stored on a PMem datastore
  - vPMemDisk can be upgraded/downgraded from/to disk or SSD
- Availability
  - Synchronous replication of updates is costly, so no HA support
  - FT provides asynchronous replication and high availability

Cluster Management

- Distributed Resource Scheduler (DRS)
  - Helps to choose a host with sufficient PMem while creating VM
  - Migrates VMs for load-balancing and maintenance mode
  - Chooses a new host for migration based on PMem availability
- Replacing Physical NVDIMMs
  - Enter maintenance mode; move all (including powered-off) VMs
  - Power-off host and replace/reconfigure physical NVDIMMs
  - Power-on host, DRS will move VMs back to the host

VMware NVDIMM Program for ISVs

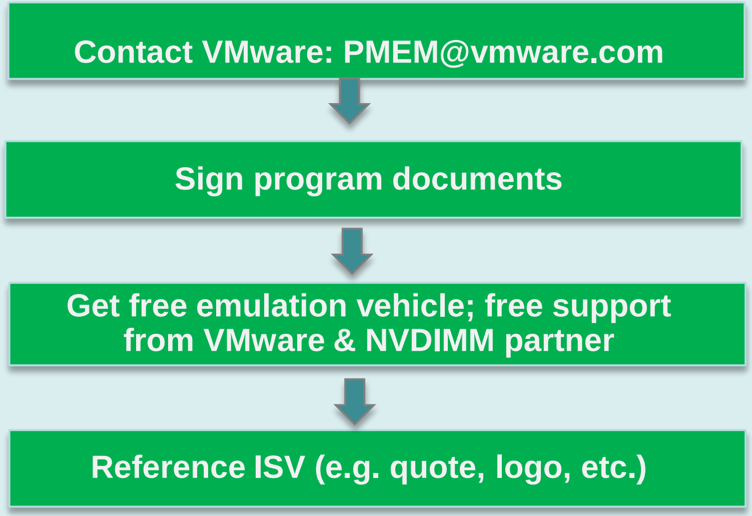
vSphere-based NVDIMM Emulation Vehicle



- Available Now
- Emulates all of the capabilities of NVDIMMs from different vendors
- Works with off-the-shelf commercial servers

To Get Emulation Vehicle

Join VMware NVDIMM Program



Contact VMware: PMEM@vmware.com

Sign program documents

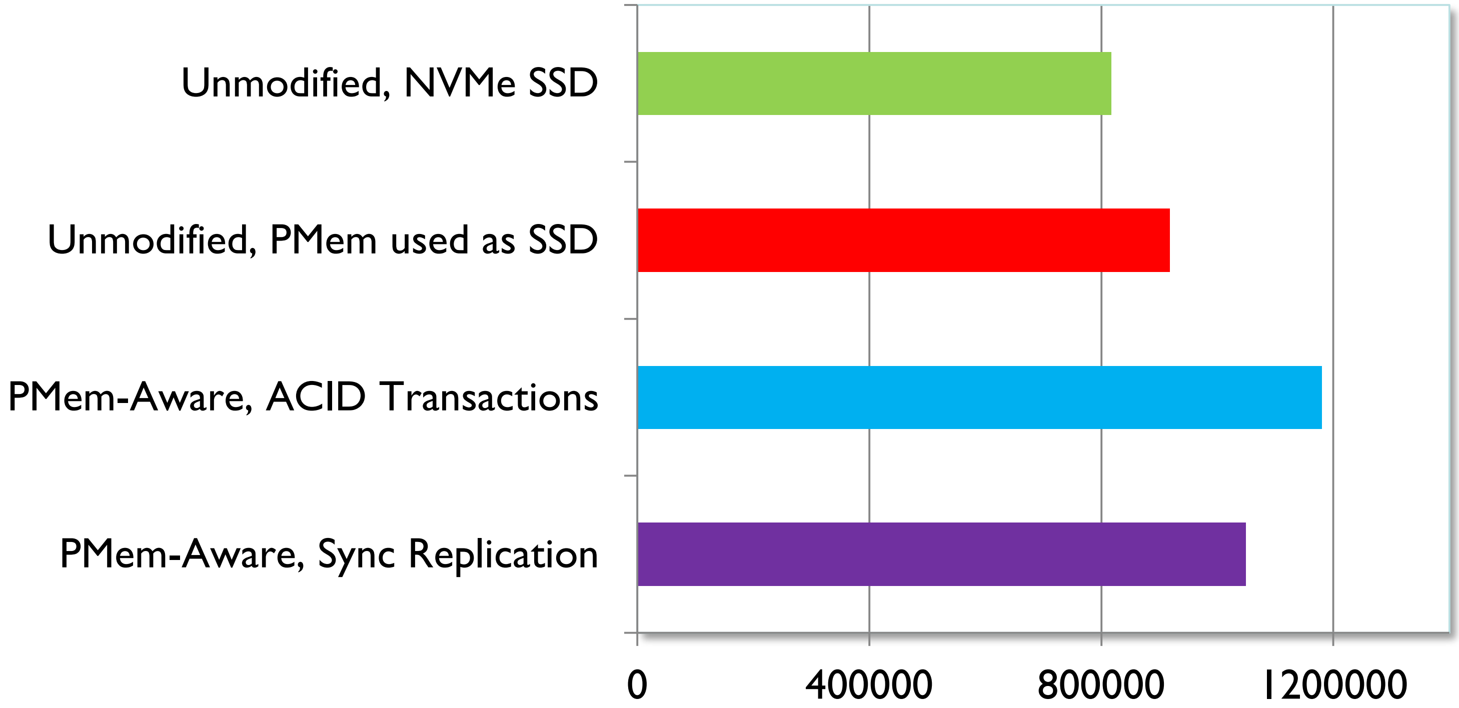
Get free emulation vehicle; free support from VMware & NVDIMM partner

Reference ISV (e.g. quote, logo, etc.)

A Tryst with Redis

- Redis
  - Keeps all data in memory, saves periodically to persistent media
  - Time for in-memory image warm up is proportional to DB size
  - Used in production by Github, Twitter, Pinterest, etc.
- PMem-aware Redis
  - Stores entire database directly in persistent memory
  - No background save thread any more
  - Nice performance improvement and instant restart after crash
  - Implemented synchronous replication for high availability

Performance of PMem-aware Redis



Summary

- Enable NVDIMM hardware
- Help legacy VMs with unmodified guest and applications
- Expose byte-addressable virtual NVDIMM to VM
- Simplify management of cluster of machines with NVDIMMs
- Preview NVDIMM performance using Redis