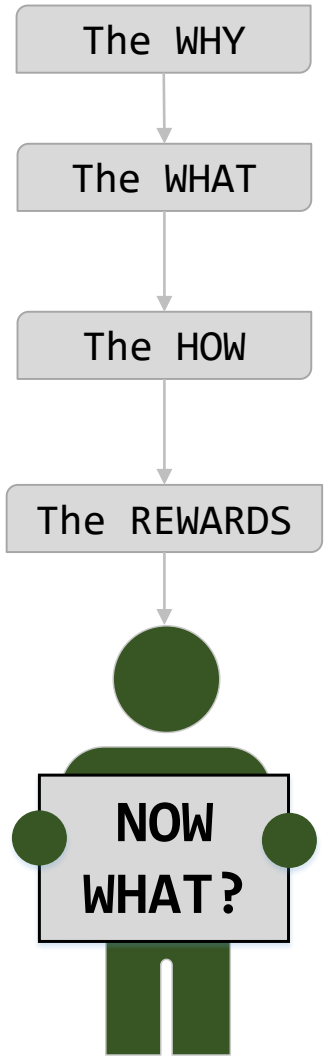

AEROSPIKE
SUMMIT '19
intel
Aerospike with Intel® Optane™ DC persistent memory
Ginger Gilsdorf
Software Engineer
Intel

Agenda



- **Setting the stage**
- **Intel Optane DC persistent memory**
- **Operating modes**
- **SNIA programming model**
- **Aerospike with persistent memory**
- **Benefits**
  - Lower cost
  - Faster restarts
  - DRAM-like performance
- **How to get started**

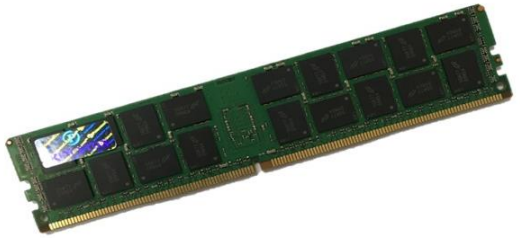
Compromising or settling?

We've been settling!

vs.

vs.

vs.

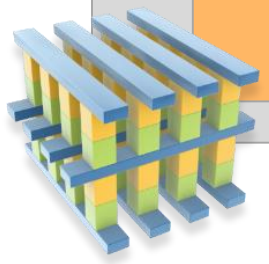


vs.

The technology and form factors*

Intel® Optane™ Technology

3D XPoint™ Memory Media + Intel Memory and Storage Controllers + Intel Interconnect IP + Intel Software



Intel Optane DC persistent memory

- DIMM form factor
- “pmem”; non-volatile memory (NVM)
- DCPMM (DC persistent memory modules)
- Launched with 2nd Gen Xeon® Scalable Processors

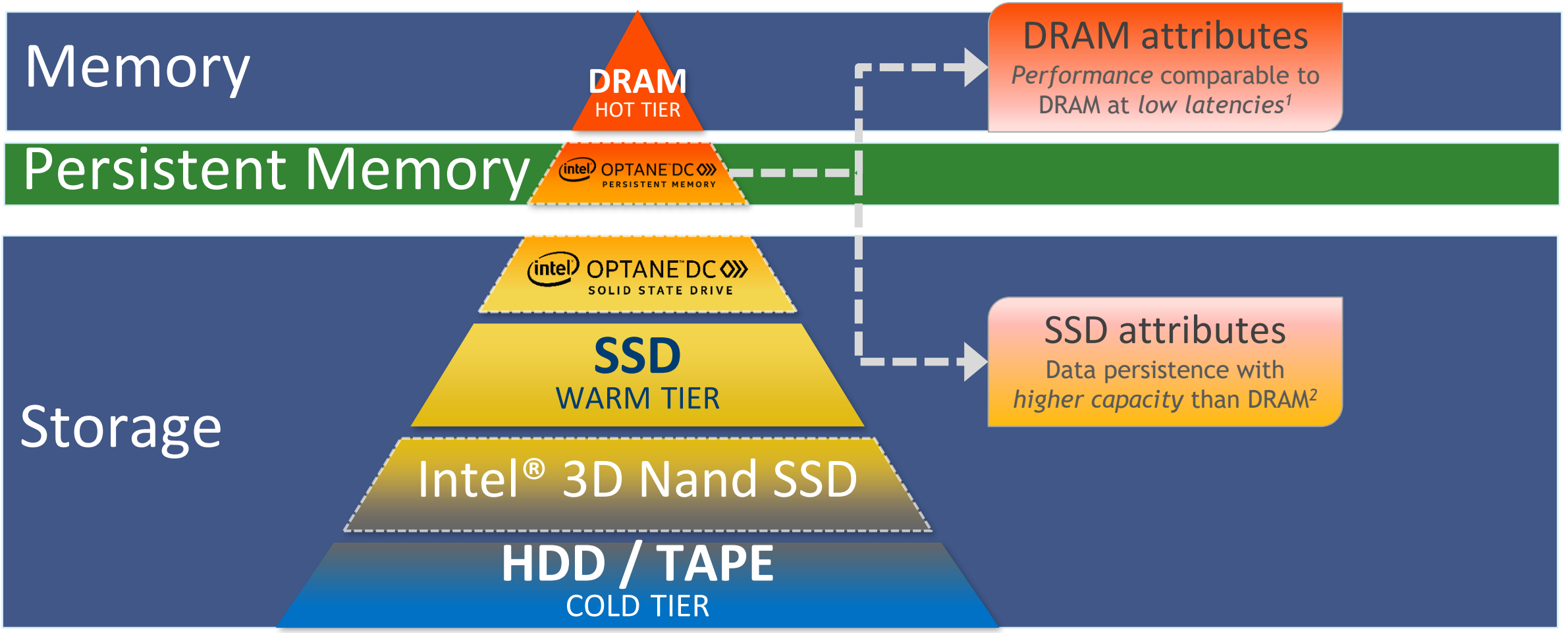
Intel Optane DC SSDs
P4800X and P4801X Series

Intel Optane SSDs
8 and 9 Series

Intel Optane Memory
Memory Drive Technology

* Images may not reflect all current offerings. See https://www.intel.com/content/www/us/en/architecture-and-technology/intel-optane-technology.html for details.

Re-architecting the memory-storage hierarchy



1. "*Fast performance comparable to DRAM*" - Intel persistent memory is expected to perform at latencies near DDR4 DRAM. Benchmarks and proof points forthcoming. "*low latencies*" - Data transferred across the memory bus causes latencies to be orders of magnitude lower when compared to transferring data across PCIe or I/O bus' to NAND/Hard Disk. Benchmarks and proof points forthcoming.
2. Intel persistent memory offers 3 different capacities – 128GB, 256GB, 512GB. Individual DIMMs of DDR4 DRAM max out at 256GB.

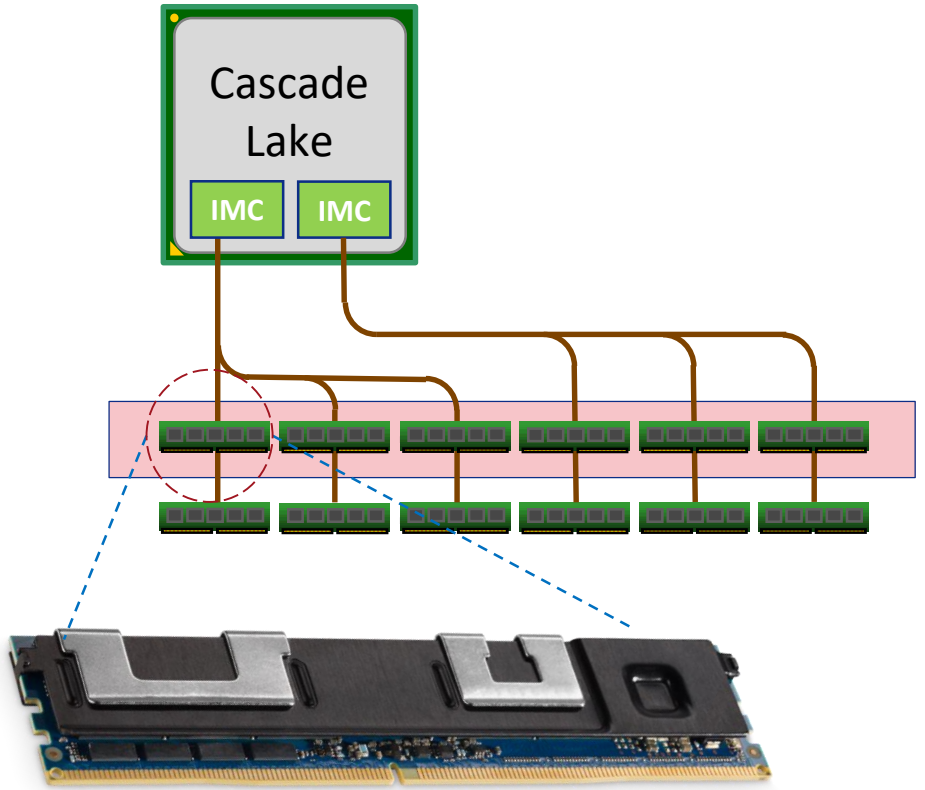
Product details

DIMM Capacities

- 128, 256, 512 GB

Per CPU

- 6 channels
- 2 DIMMs per channel
- Max one DCPMM/channel
- Max capacity 3TB
  (not including DRAM)



Speed

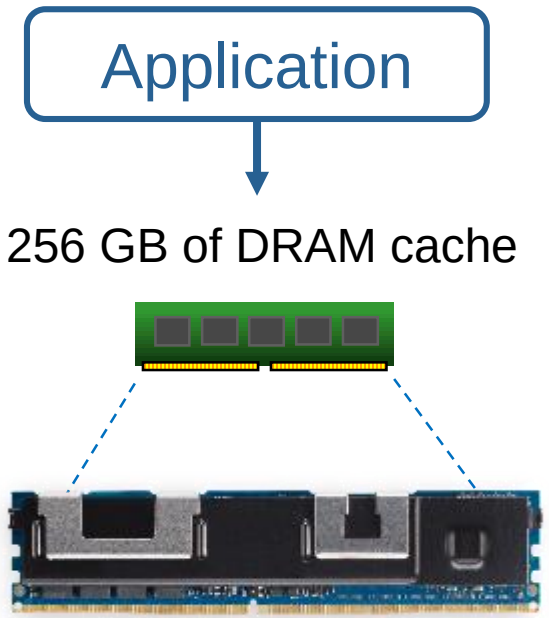
- 2666 MT/sec

Details

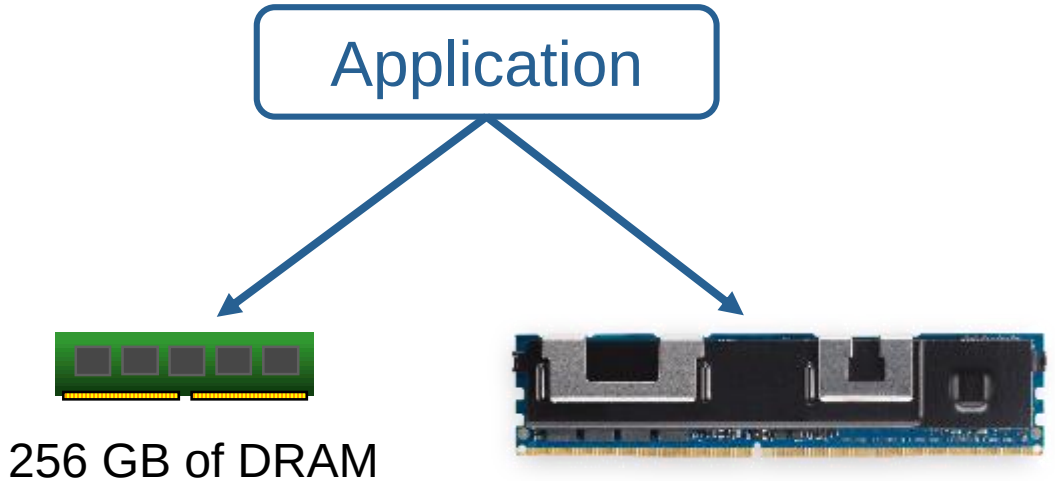
- DDR4 electrical & physical DIMM compatible
- Close to DRAM latency
- Cache line size access

BYTE ADDRESSABLE

Operating modes



1 MEMORY mode
- Large memory at lower cost
- Transparent to OS/app

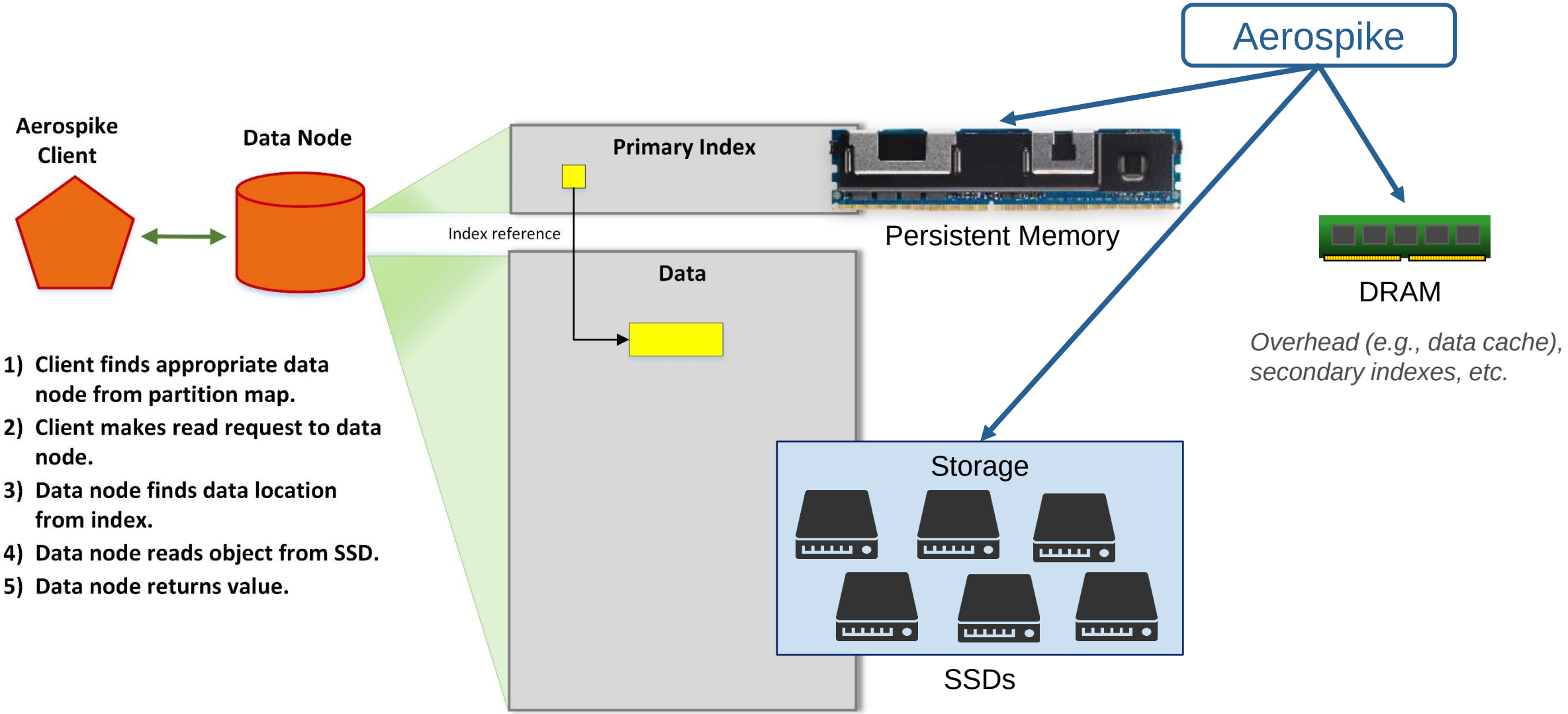


2 APP DIRECT mode
- Low latency persistent memory
- Persistent data for rapid recovery
- Requires software modifications

Storage over APP DIRECT
- Fast direct-attach storage

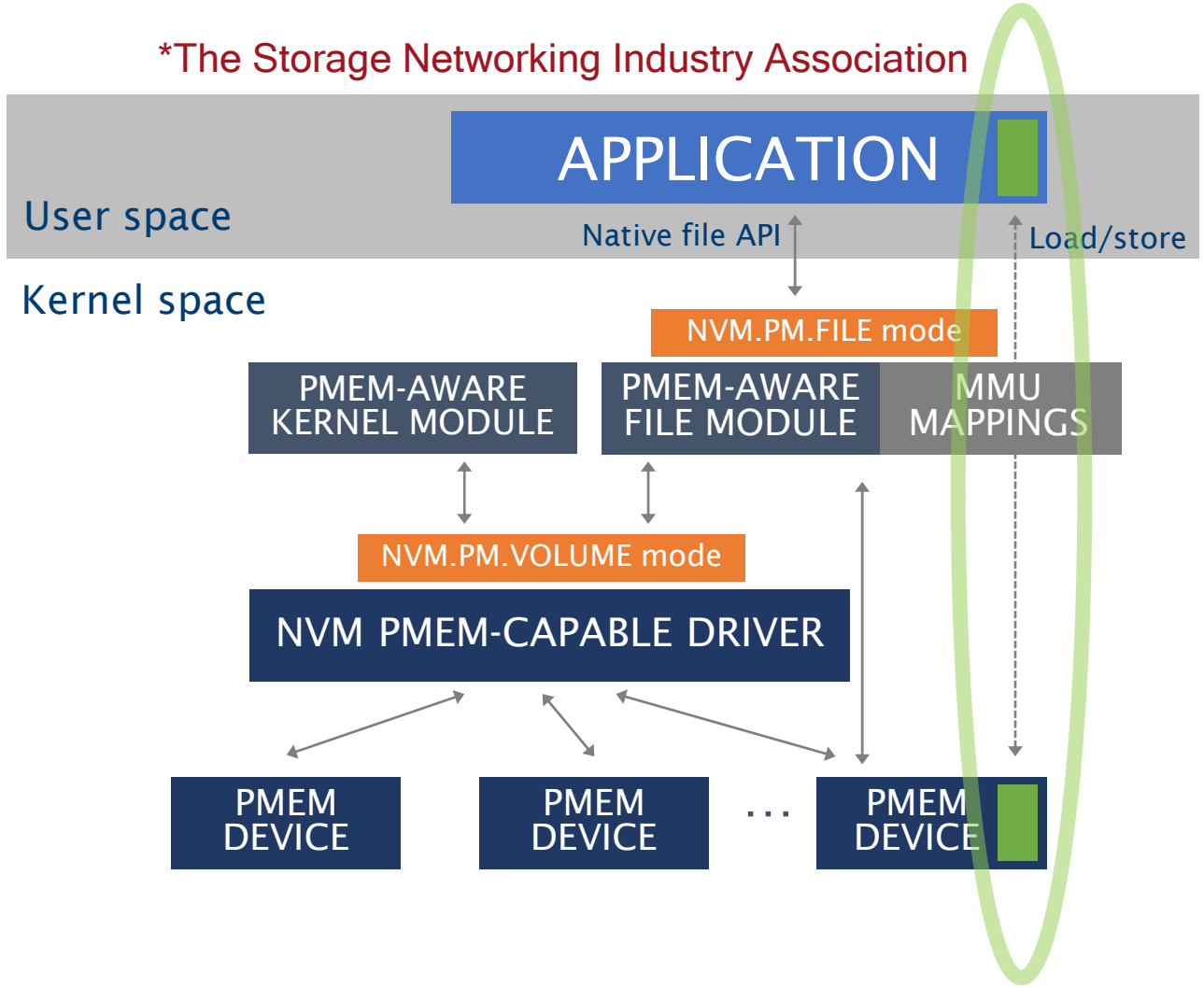
* Encryption key discarded on application shutdown.

Aerospike with persistent memory in app direct mode



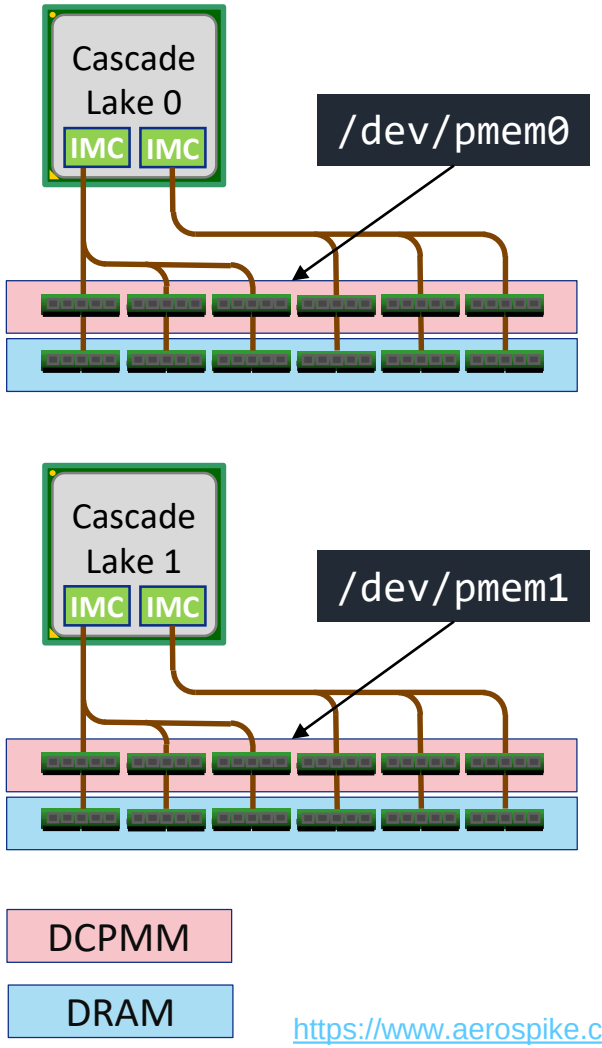
SNIA* NVM programming model

- **Persistent memory-aware file system manages access to DCPMM**
  - File system uses DAX mode
- **Kernel maps persistent memory to application address space**
- **Application has direct access to DCPMM**
  - Load/store without buffering in DRAM or kernel involvement
  - Bypass traditional I/O stack
- **The Persistent Memory Development Kit (PMDK) is built on this model**
  - Set of software libraries



ACPI 6.0 NFIT definition: https://uefi.org/sites/default/files/resources/ACPI_6_3_final_Jan30.pdf; SNIA NVM Programming Model: https://www.snia.org/tech_activities/standards/curr_standards/npm

DCPMM utilities and configuration



- **Provision DCPMMs in app direct mode**
  - ipmctl and daxctl utilities
  - Each CPU is a separate region (namespace) for DCPMM
  - Aerospike's `auto-pin numa` setting is optimized for this
- **Mount a file system with DAX (direct access)**

```
$ mkfs.ext4 /dev/pmem0
$ mount -o dax /dev/pmem0 /mnt/pmem0
```

```
$ mkfs.ext4 /dev/pmem1
$ mount -o dax /dev/pmem1 /mnt/pmem1
```

- **Configure Aerospike to run in DCPMM**

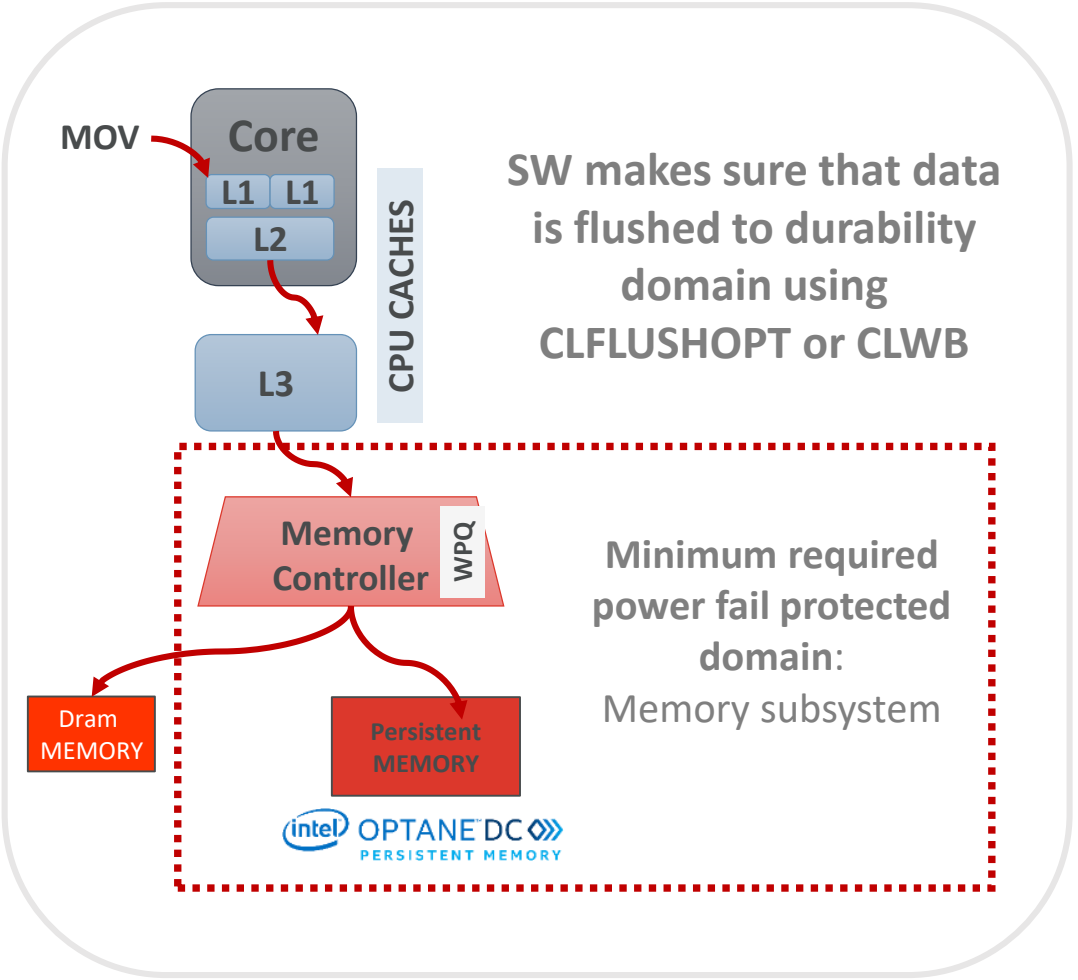
```
...
index-type pmem {
    mount /mnt/pmem0
}
...
```

```
...
index-type pmem {
    mount /mnt/pmem1
}
...
```

https://www.aerospike.com/docs/operations/configure/namespace/index/index.html

Making the index persistent

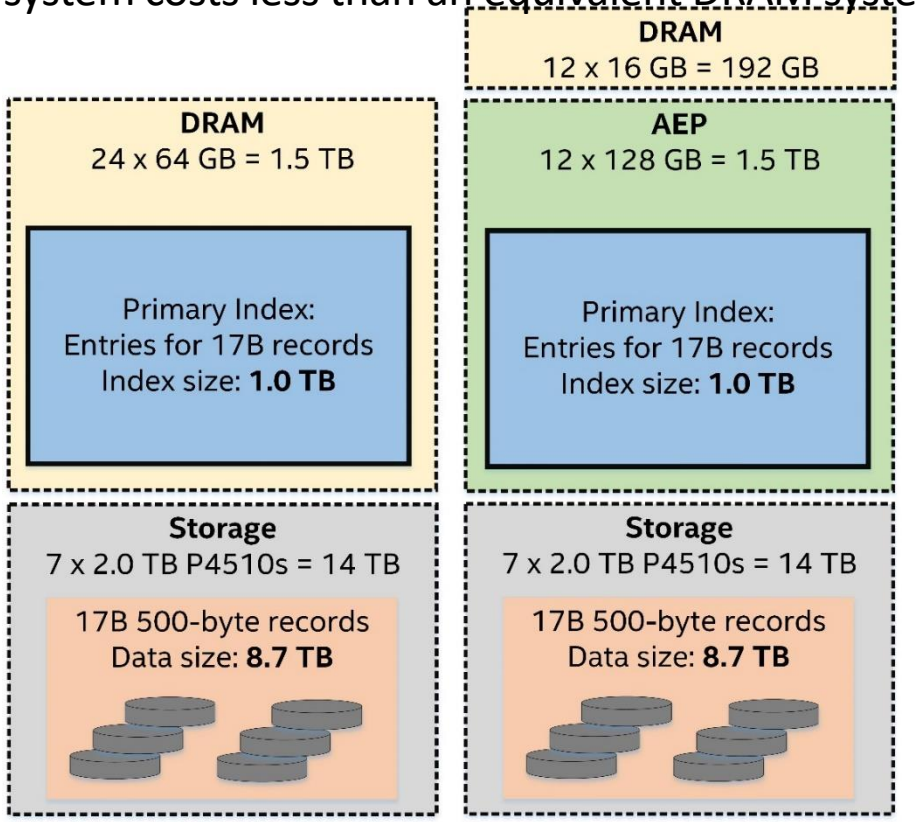
- **Run Aerospike as usual**
- **Shutdown Aerospike as usual**
  - Aerospike uses PMDK's libpmem to flush all index blocks to persistence
- **Restart Aerospike as usual**
  - Warm start even after reboot!
  - Restart attaches to existing index blocks (instead of rebuilding the index)



Aerospike user benefit: Lower cost

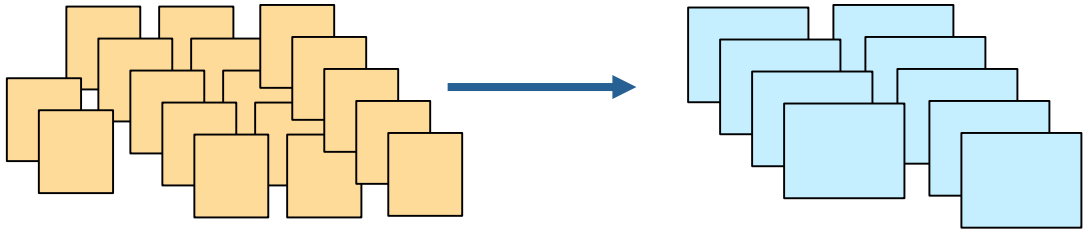
Use Case 1: Single Node TCO Reduction

- Same index capacity at a lower price/GB, so an AEP system costs less than an equivalent DRAM system



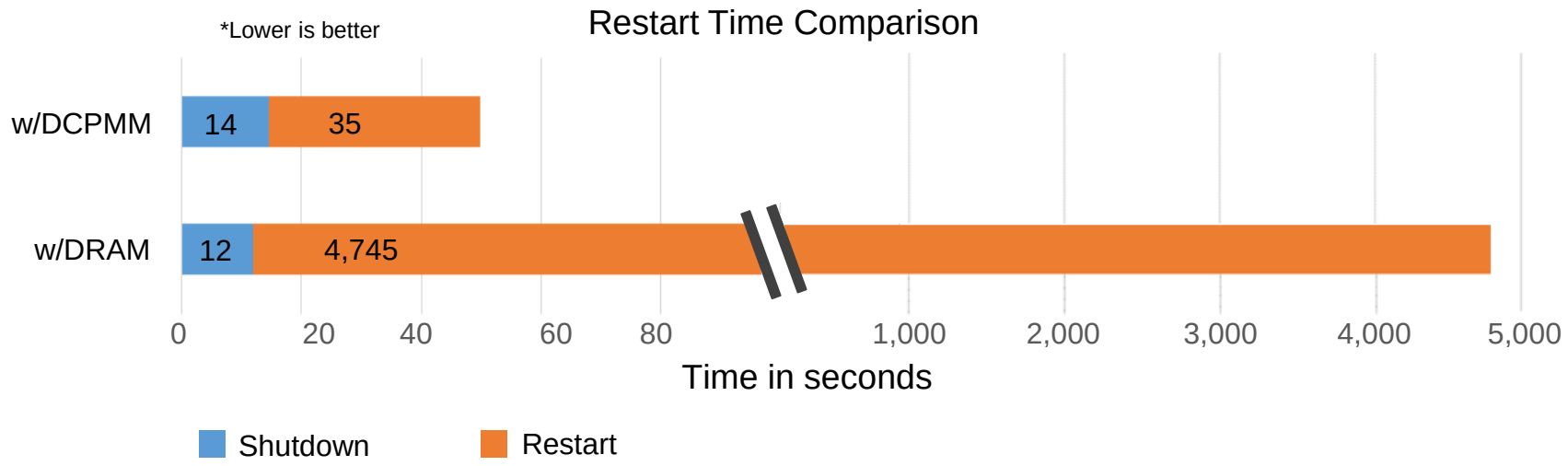
Use Case 2: Node Consolidation

- Greater index capacity allows you to store more data per node, so you need fewer nodes per cluster
- *Performance has not been tested yet*



Number of records primarily limited by index capacity

Aerospike user benefit: Faster restarts

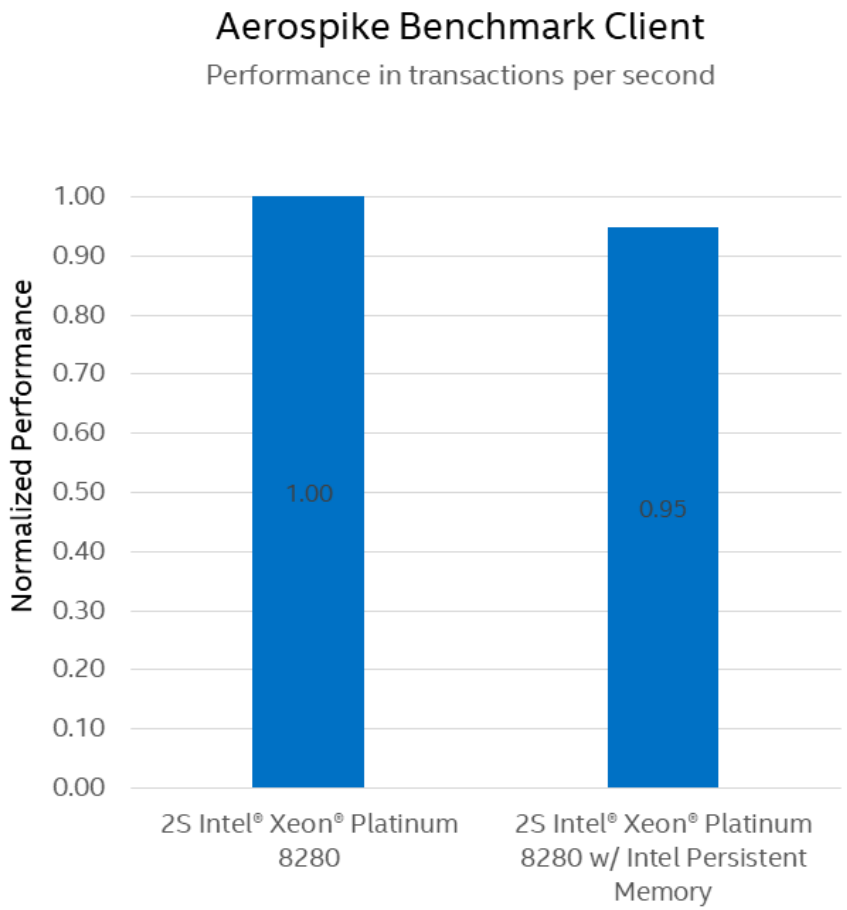


Restart Aerospike Enterprise Edition 4.5 **135x** faster after a planned system reboot with the primary index stored in Intel® Optane™ DC persistent memory.

- DCPMM system reattaches to existing index
- DRAM system scans data on SSD to rebuild index

1 - Testing conducted on ISV* software comparing Intel® Xeon® Platinum 8280L processor to Intel® Xeon® Platinum 8280L processor with Intel(r) Optane™ DC persistent memory. Testing done by Intel. For complete testing configuration details, **see slide 20**.

Aerospike user benefit: DRAM-like performance



Performance can vary by workload, but is DRAM-like in many use cases

1 - Testing conducted on ISV* software comparing Intel® Xeon® Platinum 8280L processor to Intel® Xeon® Platinum 8280L processor with Intel(r) Optane™ DC persistent memory. Testing done by Intel. For complete testing configuration details, **see slide 20**.

Resources

- Aerospike Enterprise Production Support
- Check with your OEM/CSP for availability
- Check with OS for requirements – version, kernel,…
- Quick start guide
  - https://software.intel.com/en-us/articles/quick-start-guide-configure-intel-optane-dc-persistent-memory-on-linux
- Intel Developer Zone
  - https://software.intel.com/en-us/persistent-memory
- Persistent Memory Program (links to PMDK)
  - http://pmem.io/

Aerospike perspective and Q&A

Brian Bulkowski

Aerospike Founder and Advisor

Andy Gooding

VP of Engineering, Aerospike

AEROSPIKE SUMMIT '19

Backup

Abstract

The wait is over: Intel's breakthrough non-volatile memory technology is now available for purchase. Optane DC persistent memory is faster and more durable than NAND while also more affordable and more dense than DRAM. Aerospike Enterprise Edition version 4.5+ already includes support for persistent memory, so what are you waiting for?

This session provides an overview of persistent memory, describes the different operating modes, and explains how Aerospike users benefit from storing the primary index on large capacity, affordable persistent memory instead of DRAM.

Disclaimers

http://legal.intel.com/Marketing/Pages/notices-disclaimers-chart.aspxonly the bullets

- FTC Disclaimer: For Performance Claims
  - Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/benchmarks. *Configurations: 2-socket comparison of DRAM-only system vs. DRAM + Intel Optane DC persistent memory. Test by Intel on 02/27/2019.*
- FTC Short Disclaimer
  - For more complete information about performance and benchmark results, visit www.intel.com/benchmarks. *Configurations: 2-socket comparison of DRAM-only system vs. DRAM + Intel Optane DC persistent memory. Test by Intel on 02/27/2019.*
- FTC Optimization Notice:
  - Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Notice Revision #20110804
- Security Disclaimer
  - Performance results are based on testing by Intel as of 02/27/2019 and may not reflect all publicly available security updates. See configuration disclosure for details. No product or component can be absolutely secure.
- Technology Disclaimer
  - Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at [intel.com].

Configuration details

Aerospike Enterprise Edition 4.5*: OS: CentOS Linux* 7.4, kernel 4.19.8. Testing by Intel and Aerospike completed on 02/27/2019. Security Mitigations for Variants 1, 2, 3 and L1TF in place.
BASELINE: 2nd generation Intel® Xeon® Platinum 8280 processor, 2.7GHz, 28 cores, turbo and HT on, BIOS 01.0286, 1.5TB total memory, 24 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 1 x 800GB, Intel® SSD DC S3700 + 7 Intel® SSD P4510 2TB 2.5" PCIe, CentOS Linux 7.4 kernel 4.19.8
NEW: 2nd generation Intel Xeon Platinum 8280 processor, 2.7GHz, 28 cores, turbo and HT on, BIOS 01.0286, 192GB total memory, 12 slots / 16GB / 2666 MT/s / DDR4 RDIMM and 12 slots/ 128 GB / Intel® OptaneTM DC persistent memory, 1 x 800GB, Intel SSD DC S3700 + 7 Intel SSD P4510 2TB 2.5" PCIe, CentOS Linux 7.4 kernel 4.19.8

