Cloud Storage Innovations with Intel® Optane™ and Intel® 3D NAND SSDs

Jack Zhang
SSD Enterprise Architect
Intel Non-Volatile Memory Solution Group

法律声明

英特尔技术特性和优势取决于系统配置，并可能需要支持的硬件、软件或服务得以激活。产品性能会基于系统配置有所变化。没有计算机系统是绝对安全的。更多信息，请见intel.com，或从原始设备制造商或零售商处获得更多信息。
在特定系统中对组件性能进行特定测试。硬件、软件或配置的任何差异都可能影响实际性能。请进行多方咨询，以评估您考虑购买的系统或组件的性能。关于性能及基准数据的更完整的信息，敬请登陆：http://www.intel.cn/content/www/cn/zh/benchmarks/intel-product-performance.html。
在性能检测过程中涉及的软件及其性能只有在英特尔微处理器的架构下方能得到优化。诸如 SYSmark 和 MobileMark 等测试均系基于特定计算机系统、硬件、软件、操作系统及功能，上述任何要素的变动都有可能导致测试结果的变化。请参考其它信息及性能测试（包括结合其它产品使用时的运行性能）以对目标产品进行全面评估。　更多信息请访问 http://www.intel.cn/content/www/cn/zh/benchmarks/intel-product-performance.html。
描述的成本降低方案旨在作为举例，说明指定的英特尔架构产品在特定环境和配置下，可能如何影响未来的成本和提供成本节省。环境将有所不同。英特尔不保证任何成本或成本降低。
本文包含尚处于开发阶段的产品、服务和/或流程的信息。此处提供的信息可随时改变而毋需通知。联系您的英特尔代表，了解最新的预测、时间表、规格和路线图。
本文件不构成对任何知识产权的授权，包括明示的、暗示的，也无论是基于禁止反言的原则或其他。
本文中涉及的本季度、本年度和未来的英特尔规划和预期的陈述均为前瞻性陈述，包含许多风险和不确定性。英特尔 SEC 报告中包含关于可能影响英特尔结果和计划的因素的详细讨论，，包括有关 10-K 报表的年度报告。
所有涉及的所有产品、计算机系统、日期和数字信息均为依据当前期望得出的初步结果，可能随时更改，恕不另行通知。所述产品可能包含设计缺陷或错误（已在勘误表中注明），这可能会使产品偏离已经发布的技术规范。英特尔提供最新的勘误表备索。
英特尔不对本文中引用的第三方基准数据或网站承担任何控制或审计的责任。您需要访问参考网站以确认所引用数据是否准确。
英特尔、英特尔标识、Intel.Experience What’ s Inside 标识是英特尔公司在美国和/或其他国家的商标。
* 其他的名称和品牌可能是其他所有者的资产。

扫码观看大会视频

Agenda

- Today’s SSD solutions at CSPs
- Re-architect cloud storage with Intel® Optane™ and Intel® 3D NAND SSDs
- Summary and Q&A

Todays' SSD solutions at CSPs

CSP Goals

Leadership principles:

1. Innovation to deliver differentiated services, enhance customer experience and grow customer base.
2. Contain and lower costs across the business.

Objectives

- Differentiated Services
- DC/Service Optimization

Workloads

Operational DB

Cassandra, redis, mongoDB, MySQL, Neo4j

Analytics

hadoop, Spark

Scale Out Storage

ceph, openstack

IaaS/PaaS

KVM, Xen, docker

SSD Use Cases Evolving!

Temp	• Hadoop*, Microstrategy*, Lustre* at Intel® Endeavour, SAS Analytics*
Tier	• HBase*, Marklogic*, Ceph*, MemSQL*
Cache	• Vmware* Virtual SAN, VMware* vSphere Flash Read Cache*, Intel® CAS, B-Cache*, etc…
Boot/Swap	• <200GB seems next large Enterprise target

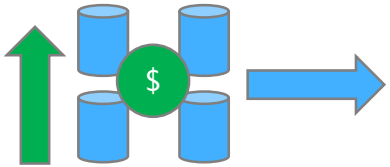
*Other names and brands may be claimed as the property of others.

MySQL* Database OLTP Performance Improvements (Percona Live 2016)

PROBLEM/CHALLENGE:

- MySQL* scale-out is well known using sharding, but this does not drive consistency of transactions or simplicity of operations.
- Modern CPU's can be starved for data and measurably slow or idle, oftentimes generating unnecessary iowait time.

Choose cost effective scale out, but scale up first!



NVMe and MySQL can provide better scale up per node.

SOLUTION BENEFITS[1]:

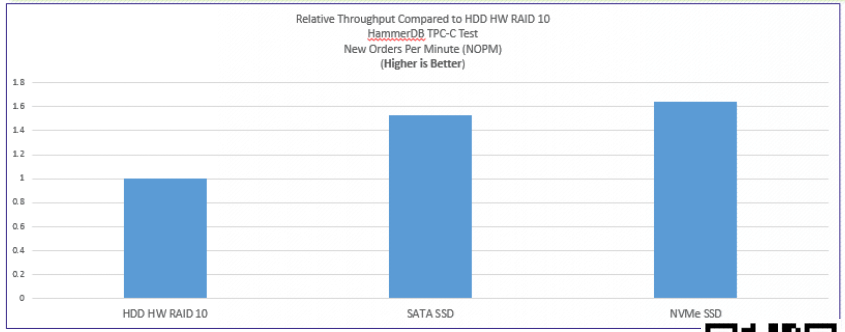
- Up to 64% performance increase of NVMe SSD over HDD
- Up to 53% performance increase of SATA SSD over HDD
- High consistency SSDs help your operational DB goals
- More performance per node reduces operational expenses

SOLUTION INGREDIENTS:

- Intel® SSD DC – NVMe* for performance

New Orders per Minute data, higher is better!

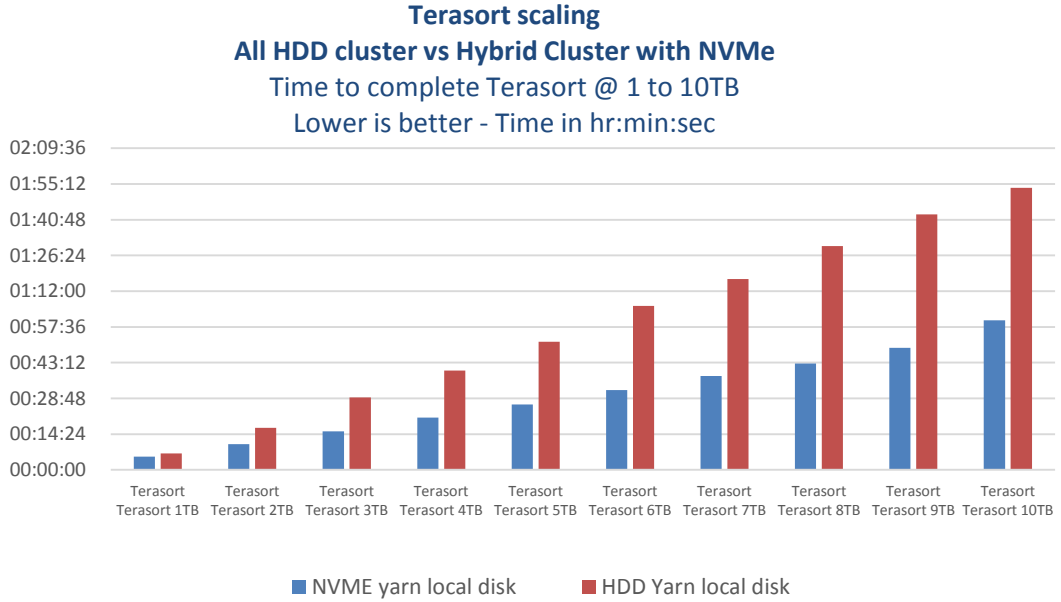
HammerDB* TPC-C: Results[1]



More info: MySQL Database OLTP Performance (Percona Live 2016)

[1]Testing Source: Intel Corporation. Intel Server System R2208WT2YS server, Intel® Xeon® E5-2699 v3 18 core at 2.3Ghz dual socket, Intel Server Board S2600WT2. Cent l 3.10.0-327.el7.x86_64 , 128GB DDR4 RAM, Intel Data Center SSD S3710 Series boot drive, RAID Controller: Intel RAID Controller RS3DC080. Percona Server Server 5.7/1 SysBench 0.5, HammerDB 2.19 (64-bit), mdadm v3.3.2, Inbox NVMe* and RAID Controller Drivers, XFS file system.

Big Data Hadoop*: Advantages of NVMe* for MapReduce* on Yarn*



Solution Ingredients:

- Intel® SSD DC P3500 Series
- Intel® SSD DC S3700 Series
- Intel® Xeon® Processor E5-2697 v3
- Intel® Ethernet Controller XL710-AM2

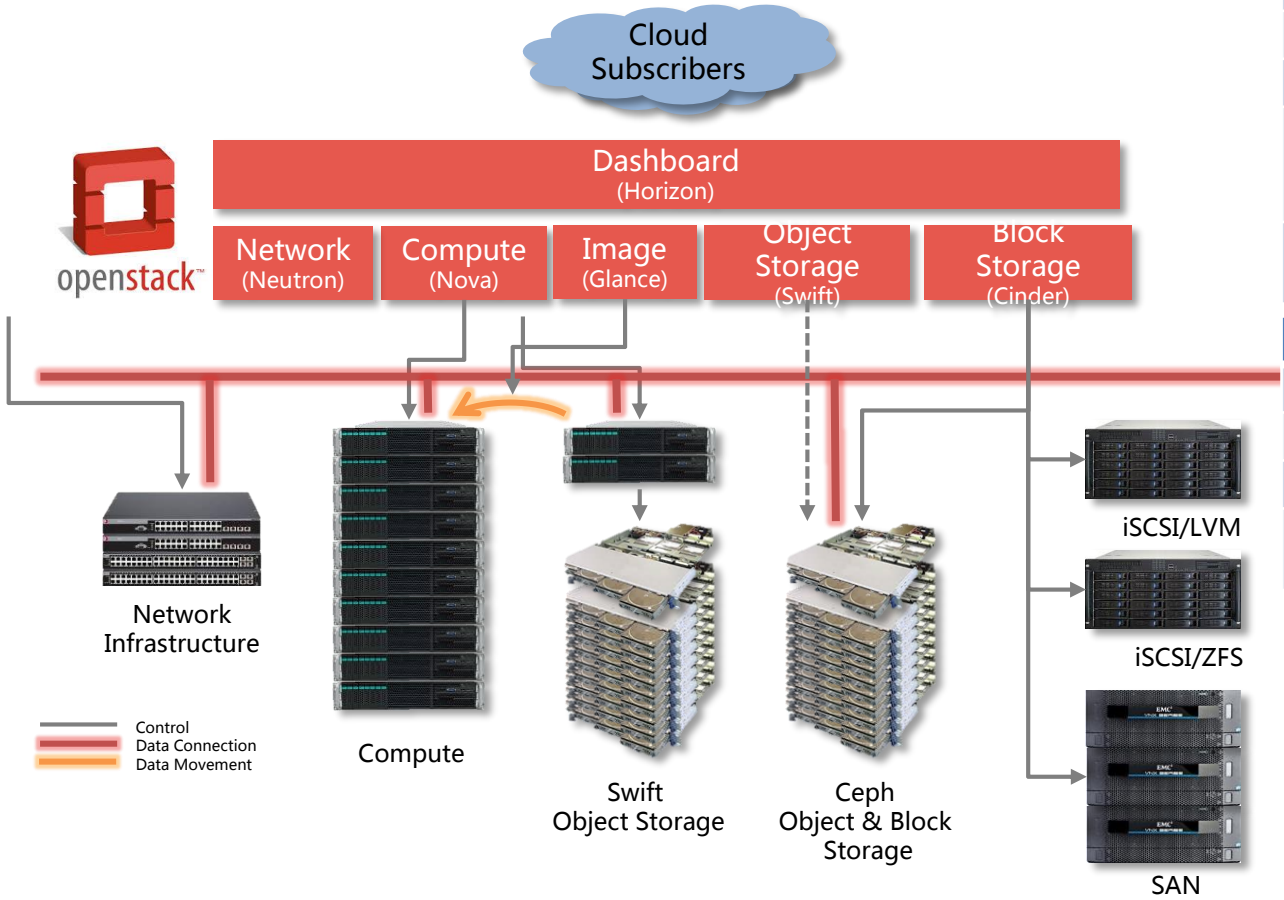
Benefits:

- Improved overall platform scalability for this workload type
- Enabled workloads to increasingly complete faster at scale compared to HDD **without any code or significant infrastructure changes.**

Adding a single NVMe* to each data node resulted in performance improvements of up to 110%

† Test configuration: MapReduce Cluster of 16 servers:1)Chassis - Intel® S2600WT 2)Processor - 2x Intel® Xeon® E5-2697 3)HDD - 12x Seagate* Enterprise Ca 2.5" HDD (2 TB) 4)NVMe SSD - 1x Intel® SSD DC P3500 Series (1.2 TB) 5)Network - Intel® Ethernet Converged Network Adapter XL710-AM2 6)Boot SSD - Int DC S3700 Series (400 GB)

OpenStack™, Swift, & Ceph



Ceph* storage node --Go	
CPU	Intel(R) Xeon(R) CPU 2650v3
Memory	64 GB
NIC	10GbE
Disks	1x 1.6TB P3700 + 12 x 4TB HDDs (1:12 ratio) P3700 as Journal and caching
Caching software	Intel(R) CAS 3.0, option: Intel(R) RSTe/MD4.3

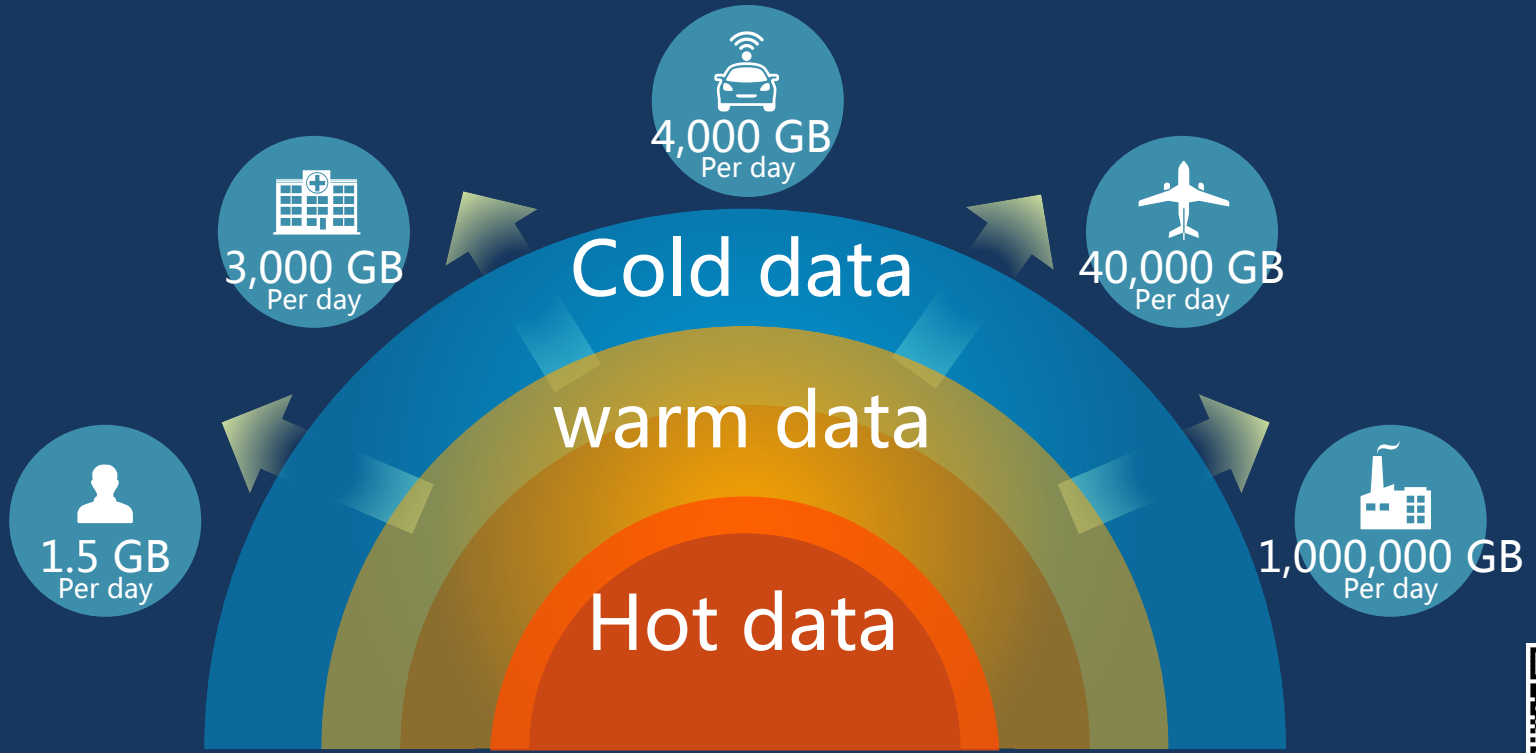
Ceph* Storage node --Better	
CPU	Intel(R) Xeon(R) CPU E5-2690
Memory	128 GB
NIC	Duel 10GbE
Disks	1x Intel(R) DC P3700(800G) + 4x Intel(R) DC S3510 1.6TB

Ceph* Storage node --Best	
CPU	Intel(R) Xeon(R) CPU E5-2699v3
Memory	>= 128 GB
NIC	2x 40GbE, 4x dual 1
Disks	4 to 6 x Intel® DC

扫码观看大会视频

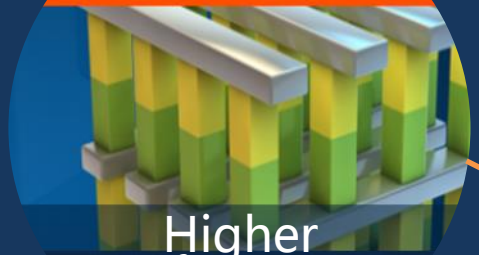
Re-architect cloud storage with Intel® Optane™ and Intel® 3D NAND SSDs

Tier by different usages



Intel investment: Two technologie

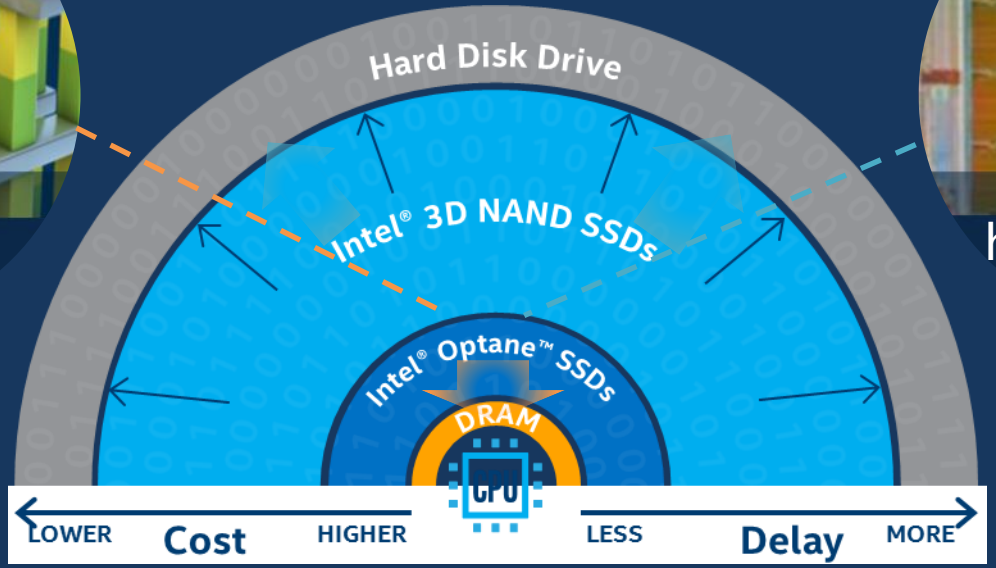
Intel® Optane™ Technology



Higher performance

Intel® 3D NAND

lower cost & higher density



Intel® Optane™ TECHNOLOGY

Size and Latency Specification Comparison

HDD
Latency: ~10 MillionX
Size of Data: ~10,000X

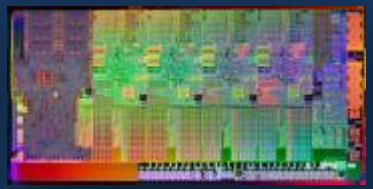
MEMORY

NAND SSD
Latency: ~100,000X
Size of Data: ~1,000X

Intel® Optane™ Technology
Latency: ~100X
Size of Data: ~1,000X

DRAM
Latency: ~10X
Size of Data: ~100X

SRAM
Latency: 1X
Size of Data: 1X

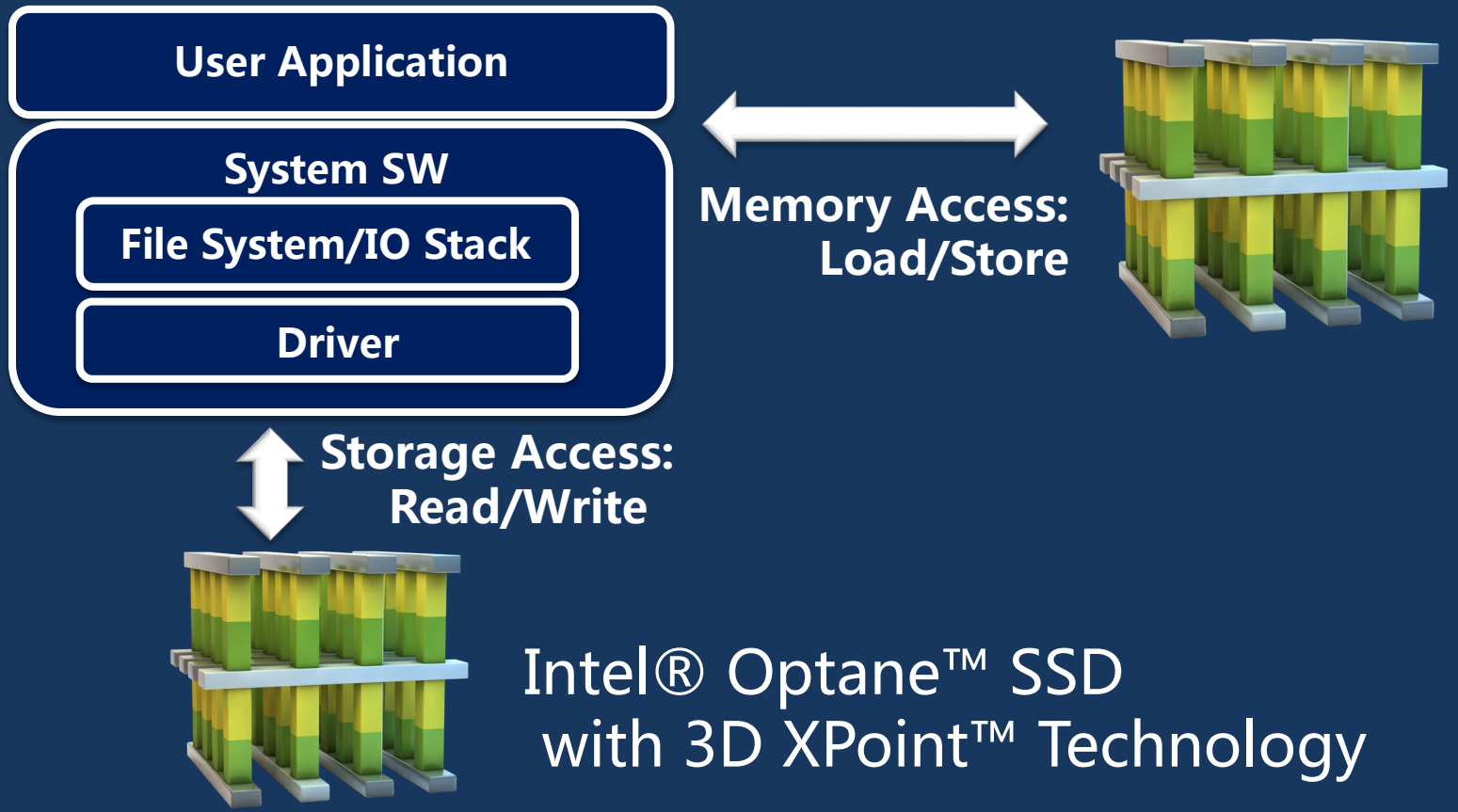


STORAGE

Technology claims are based on comparisons of latency, density and write cycling metrics amongst m… technologies recorded on published specifica… market memory products against internal Int… …ns.

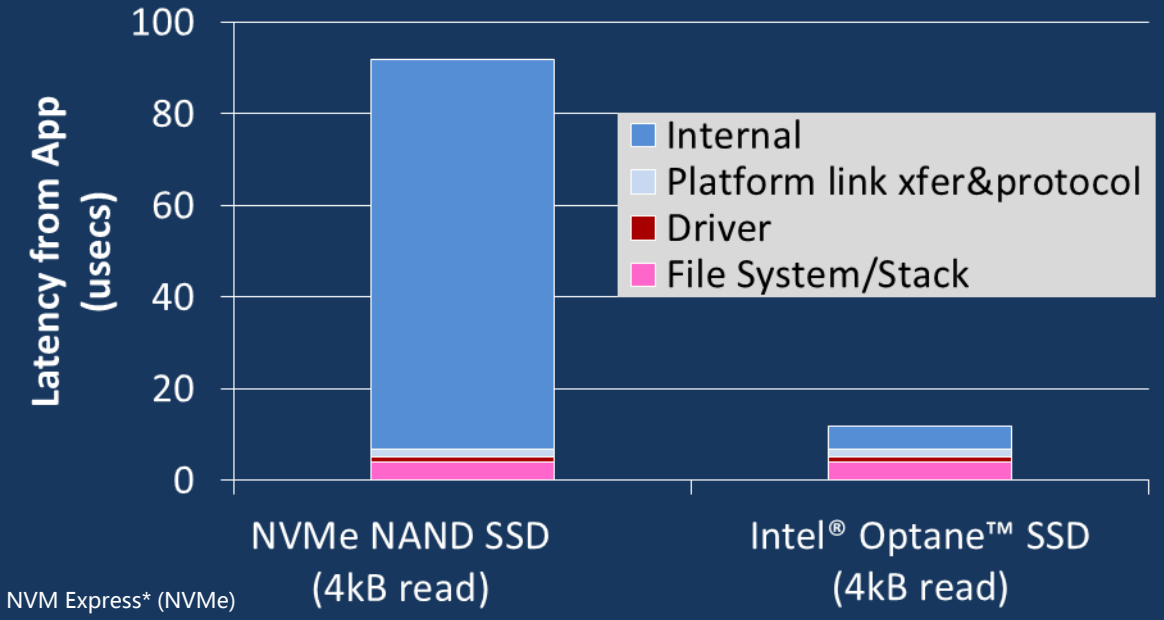
Lower latency is faster

Memory and Storage Platform Connection



Intel® Optane™ SSD
with 3D XPoint™ Technology

Storage System Interconnect

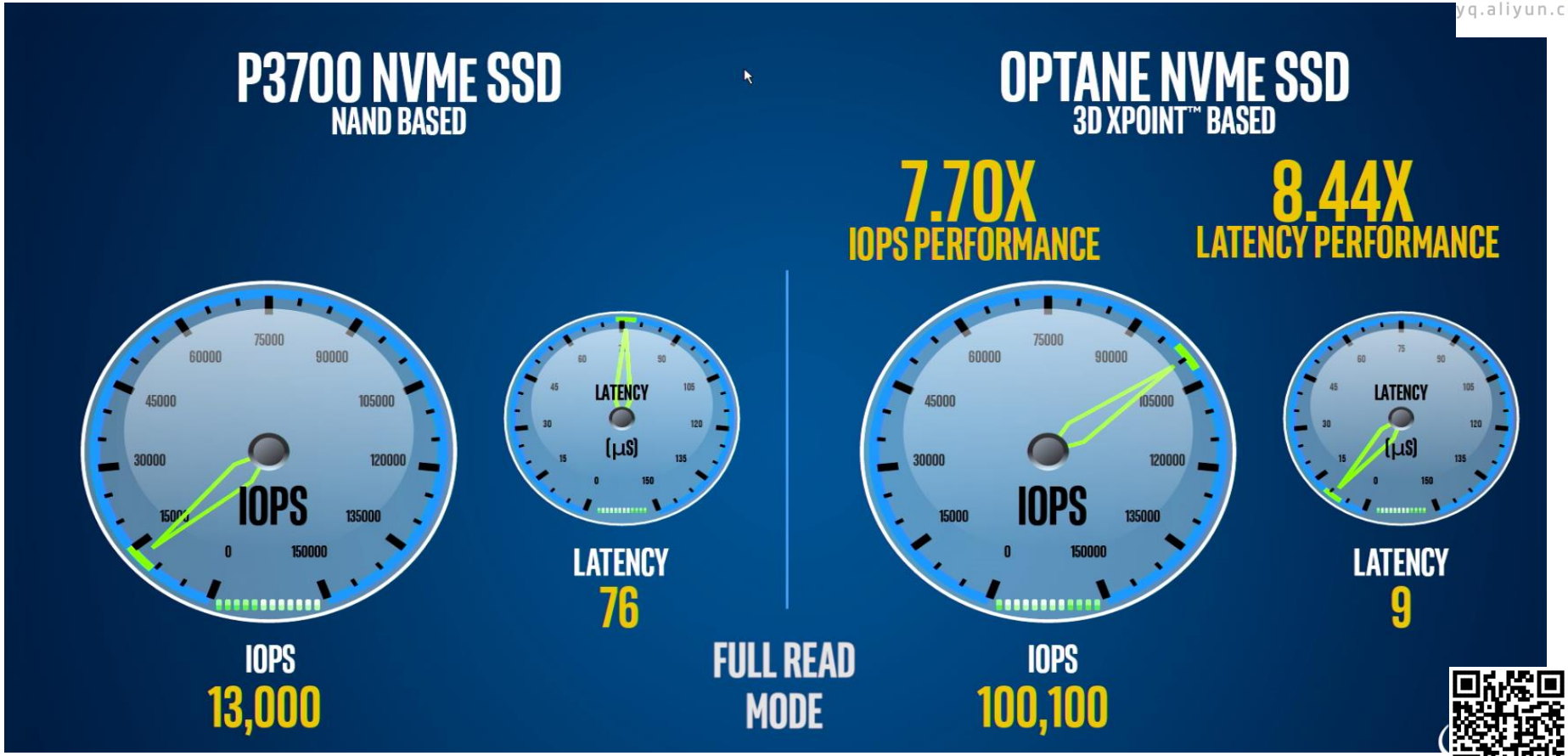


NVM Express* (NVMe)

Low Latency of Intel® Optane™ SSDs *is* accessible by User Applications on today's systems

Sources: "Storage as Fast as the rest of the system" 2016 IEEE 8th International Memory Workshop and measurement, Intel® Optane™ SSD measurements and Intel SSD P3700 Series measurements

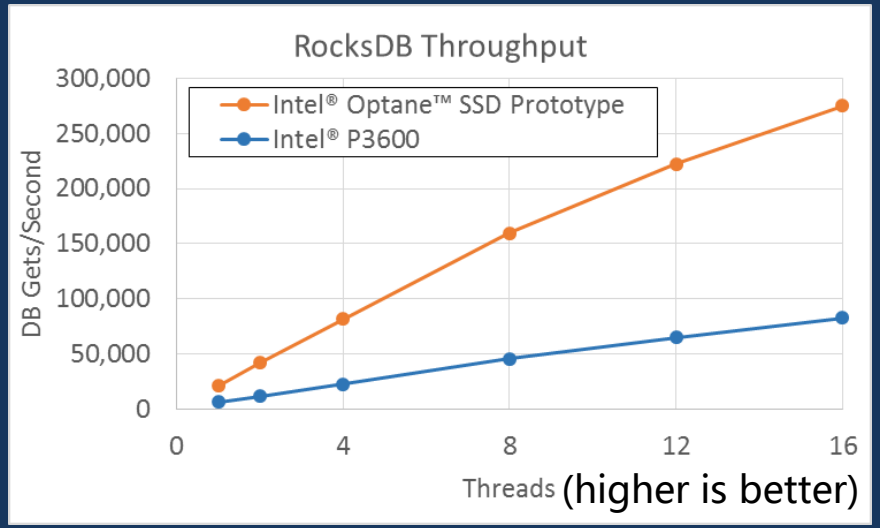
Intel® Optane™ (prototype) vs Intel® SSD DC P3700 Series at QD=1

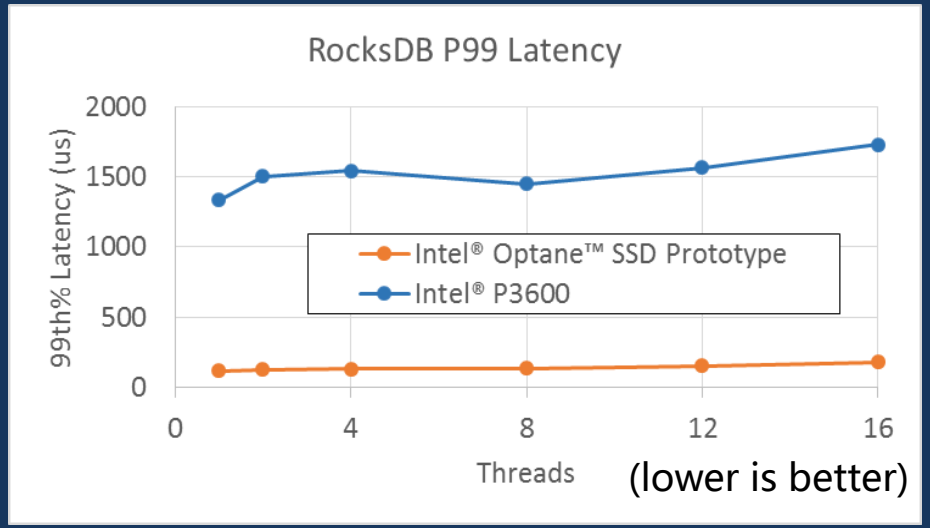


Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sour
information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit http://www.intel.com/performanc
Configuration: 2x Intel® Xeon® E5 2690 v3 NVM Express* (NVMe) NAND based SSD: Intel P3700 800 GB, 3D Xpoint based SSD: Optane NVMe OS: Red Hat* 7.1

Data Center: RocksDB Perf on Test5 (from rocksdb.org

- Open source persistent key-value store
- All threads randomly reads keys, one writer thread updates up to ~80K keys/second





~3x Throughput advantage

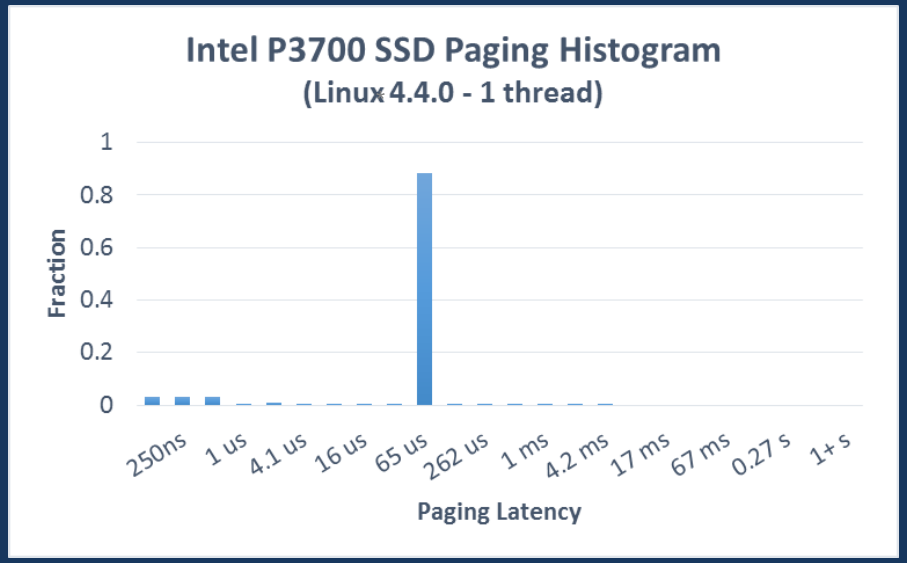
~10x Latency advantage (99th percentile)

Increased persistent key-value store throughput with better QoS

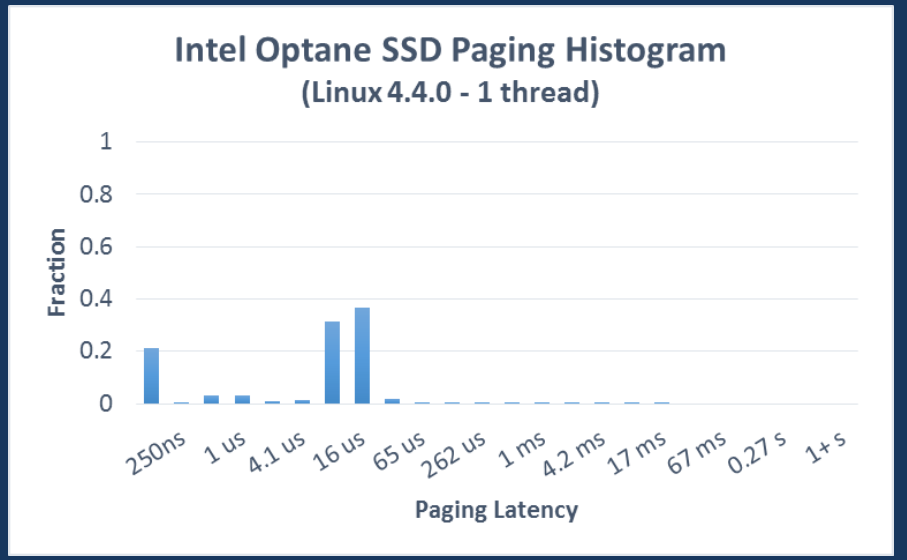
RocksdB setup based on published tests at rocksdb.org: 1B Key Database used, 8 "Shards" of 25M Key/Values each, 20 byte keys, 800 byte values, 50% compression, ~100 GB on-disk. Read: All threads randomly read all keys. Read/Write: All threads randomly reads keys 1 writer thread updates up to ~80K keys/second.

Quanta Leopard base board, 2x Intel CPUs (2.5 GHz, 12 core, HT Enabled, 8 DDR4 DIMMs, 256GB, 32GB Used, CentOS* 7.2, no OS changes X build/mount opts, TRIM enabled, P3700 (50% capacity used) and Intel Optane Based Prototype (75% capacity used).

Paging Performance



Average paging time = 88 usecs



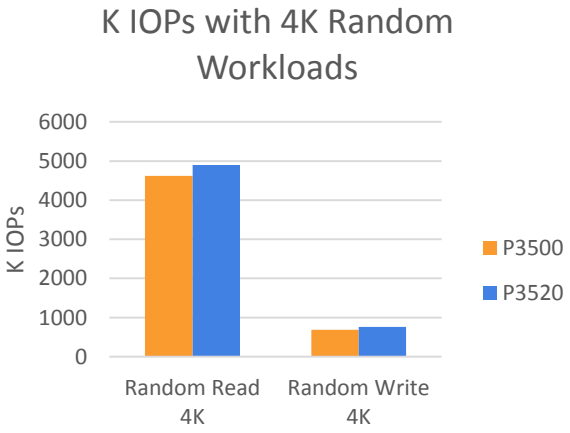
Average paging time = 15 usecs

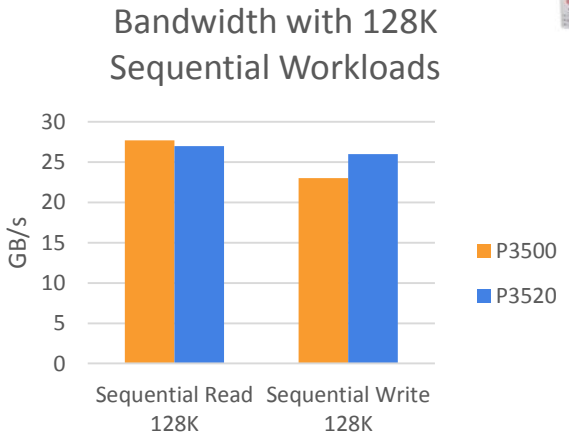
Paging to extend system memory now a viable strategy

PMBench 0.71 2GB file w/random access, 1 thread, Ubuntu* server 4.4.0-31,
third-party platform, 1.5GHz, 32GB DDR4 – 200 limited to 1GB with GRUB, Neon Coty MB

High Density Data Storage on 3D NAND SSDs

New Storage Criteria: **IOPS/TB**



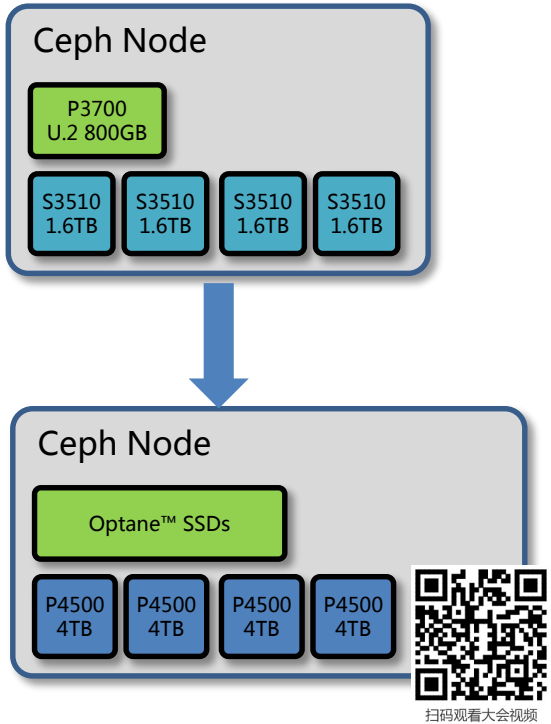


40% Lower cost than P3500

P3500 vs. P3520 Comparison. Results based on running FIO under Centos 7.2 on an Intel® DC P3500 2TB versus Intel® DC P3520 2TB 24xU.2 NVMe on SMC model SYS-2028U-TN24R4T+.

Intel® Optane™ & Intel® 3D NAND SSD High performance & cost effective solutions

- Enterprise class, highly reliable, feature rich, and cost effective AFA solution:
  - NVMe as Journal, 3D NAND TLC SSD as data store
    (performance) (capacity)
- Enhance value through special software optimization on filestore and bluestore backend



Summary

- Intel® 3D NAND is the building block for high capacity, low cost SSDs
- Intel® Optane™ technology delivers high performance, low latency and fills the gap between DRAM and Flash SSDs
- Intel® Optane™ + Intel® 3D NAND SSD together enable high performance and cost effective storage innovations

2016 The Computing Conference

THANKS