SCALING ORACLE DATABASE PERFORMANCE ON LINUX TO FABULOUS NEW LEVELS

Hubert Nueckel
Principal Engineer,
Intel

Doug Nelson
Technical Lead, Intel

September 2017

Legal Disclaimer

ORACLE OPEN WORLD

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.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessor 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 about performance and benchmark results, visit www.intel.com/benchmarks.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase.

For more complete information about performance and benchmark results, visit http://www.intel.com/benchmarks

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

Agenda

ORACLE OPEN WORLD

- Intel and Oracle – How we work together
- The New Intel® Xeon® Scalable Processor (Formerly Skylake-SP)
- OLTP performance improvements over time
- NVME Storage can take performance to the next level
- In-Memory Database improvements
- Future Work
- Performance Monitoring Tools
- Questions

INTEL AND ORACLE DATABASE COLLABORATION

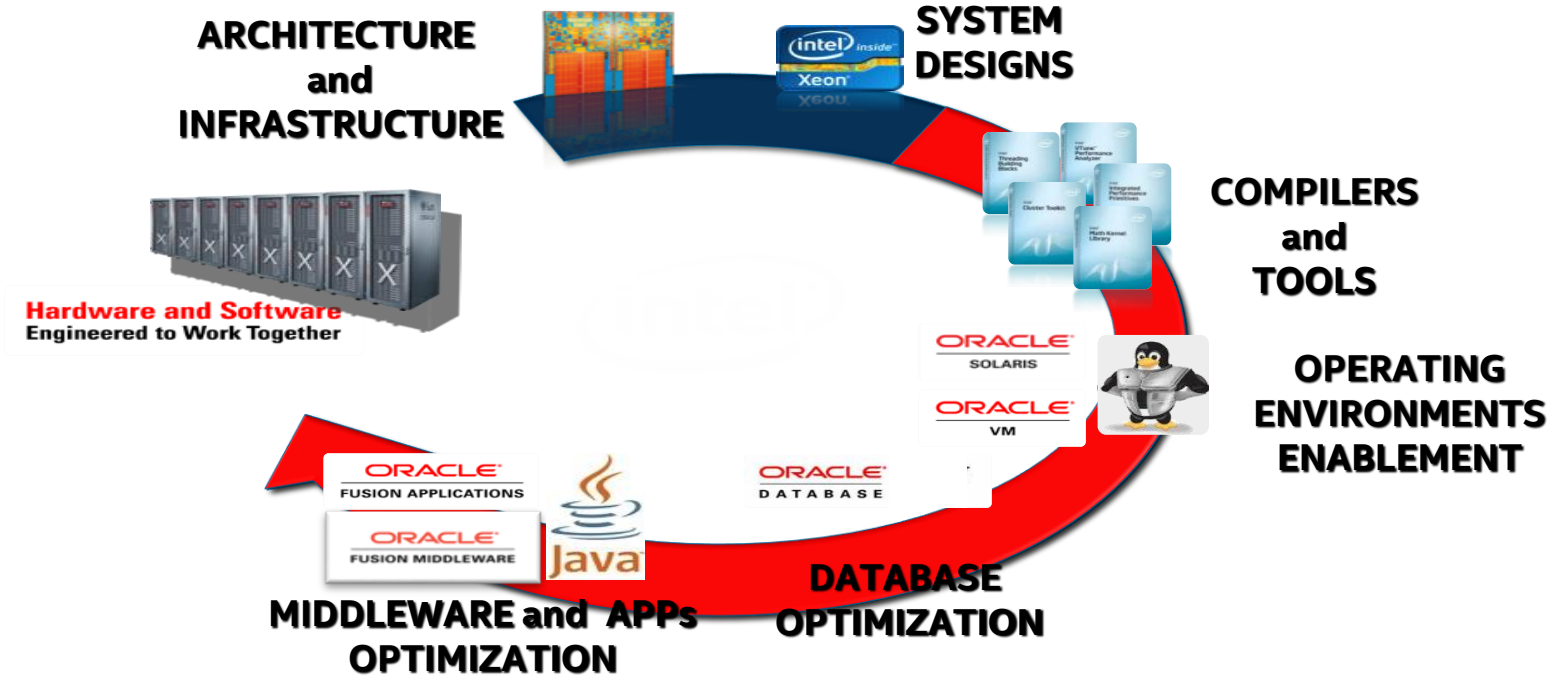
Collaboration Introduction

Oracle Open World 2013

Intel started close collaboration with Oracle late ‘90s

Both companies realized potential of engagement

Intel/Oracle Collaboration Model



Intel/Oracle Collaboration

ORACLE OPEN WORLD

- How can software leverage hardware features
- Any gaps between hardware and software
- Regular architecture meetings
  - Where does Oracle go with regards to their software and engineered systems
  - Where does Intel go with processor, platform and infrastructure
  - Several years in advance planning
- Close and open collaboration

Intel and Oracle Collaboration

ORACLE OPEN WORLD

- Test each CPU generation with a variety of database workloads
- Feedback to CPU architects influences future CPUs
- Work together to implement new CPU/platform features
- Test the whole system with new storage, memory, CPUs, network to give you the best possible performance… and catch any issues before product release

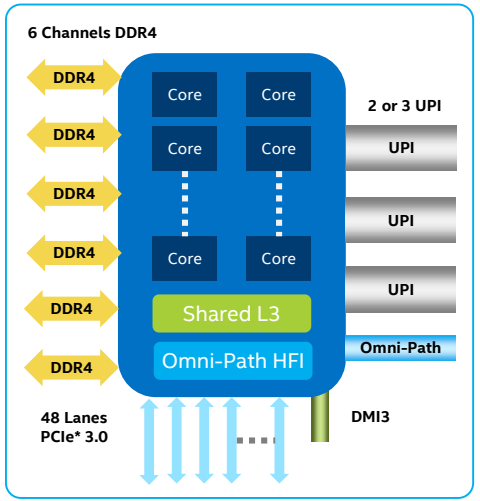
THE NEW INTEL® XEON® SCALABLE PROCESSOR
(FORMERLY SKYLAKE-SP)

Intel® Xeon® Scalable Processor Overview

Re-architected from the ground up

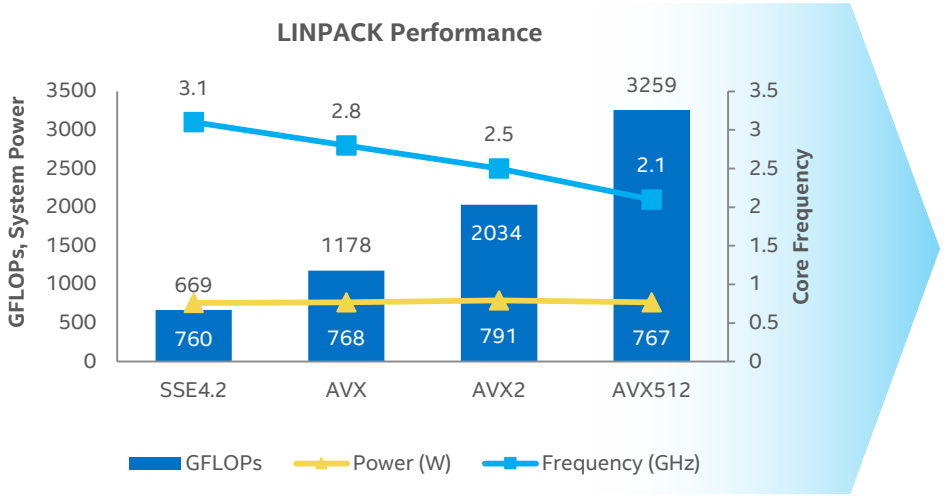
- Skylake core microarchitecture with data center specific enhancements
- Intel® AVX-512 with 32 DP flops per cycle per core
- Datacenter optimized cache hierarchy – 1MB L2 per core, non-inclusive L3
- New mesh interconnect architecture
- Enhanced memory subsystem
- Modular IO with integrated devices
- New Intel® Ultra Path Interconnect (Intel® UPI)
- Intel® Speed Shift Technology
- Security & Virtualization enhancements (MBE, PPK, MPX)
- Optional Integrated Intel® Omni-Path Fabric (Intel® OPA)

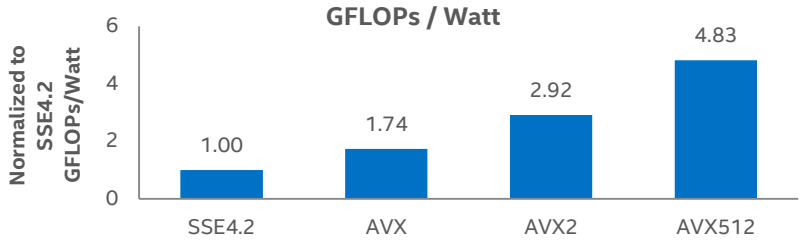
Features	Intel® Xeon® Processor E5/E7 v4	Intel® Xeon® Scalable Processor
Availability	Q2 2016	Q3 2017
Cores Per Socket	Up to 22 or 24	Up to 28
Threads Per Socket	Up to 44 or 48 threads	Up to 56 threads
Last-level Cache (LLC)	Up to 55MB or 60 MB	Up to 38.5 MB (non-inclusive)
QPI/UPI Speed (GT/s)	2x or 3x QPI channels @ 9.6 GT/s	Up to 3x UPI @ 10.4 GT/s
PCIe* Lanes/ Controllers	40 / 10 / PCIe* 3.0 (2.5, 5, 8 GT/s)	48 / 12 / PCIe 3.0 (2.5, 5, 8 GT/s)
Memory Population	4 channels of up to 3 RDIMMs, LRDIMMs, or 3DS LRDIMMs	6 channels of up to 2 RDIMMs, LRDIMMs, or 3DS LRDIMMs
Max Memory Speed	Up to 2400	Up to 2666
TDP (W)	55W-145W	70W-205W

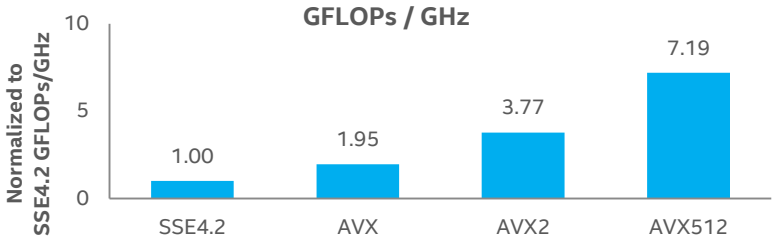


Intel® Advanced Vector Extensions 512 (Intel® AVX-512)

512-bit wide vectors, 32 operand registers, 8 64b mask registers, Embedded broadcast & rounding







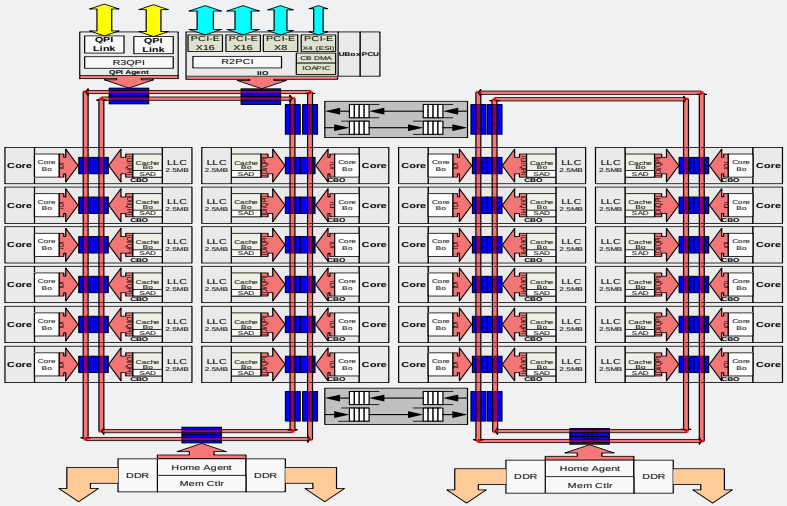
INTEL® AVX-512 DELIVERS SIGNIFICANT PERFORMANCE AND EFFICIENCY GAINS

Source as of June 2017: Intel internal measurements. 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. Configuration Summary: 1-Node, 2 x Intel® Xeon® Platinum 8180 Processor on Purley-EP (Lewisburg) (S2600WF) with 384 GB (12x32GB DDR4-2666) Total Memory, Intel S3610 800GB SSD, BIOS: SE5C620.86B.01.00.0471.040720170924, 04/07/2017, RHEL Kernel: 3.10.0-514.16.1.el7.x86_64 x86_64, Benchmark: Intel® Optimized MP LINPACK

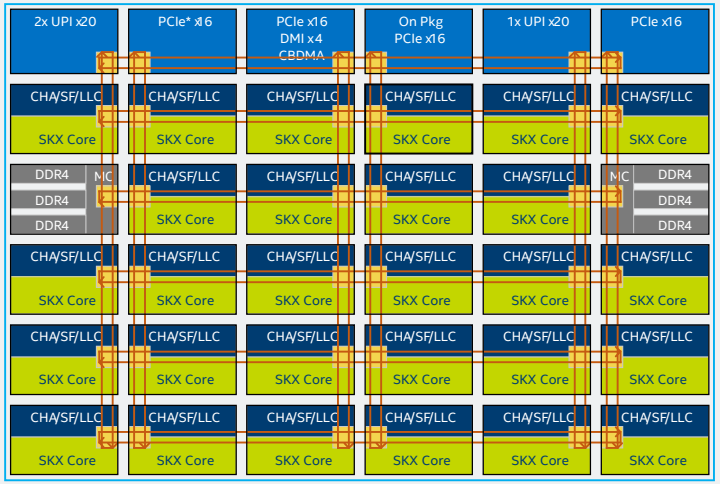
New Intel® Mesh Interconnect Architecture

ORACLE OPEN WORLD

2016: Intel® Xeon® Processor E7 v4, 14nm (Broadwell EX 24-Core Die)



2017: Intel® Xeon® Scalable Processor, 14nm (Skylake-SP 28-Core Die)



CHA– Caching and Home Agent, SF– Snoop Filter, LLC– Last Level Cache
SKX Core- Skylake Server Core, UPI – Intel® UltraPath Interconnect

Intel® Mesh Improves Scalability with Higher Bandwidth and Reduced Latencies

Intel® Xeon® Scalable Processor Summary

ORACLE OPEN WORLD

Architectural innovations to unlock data center performance

- **Up to 60% increase** in compute performance per socket with Intel® AVX-512
- **Improved performance and scalability** with Mesh on-chip interconnect
- L2 and L3 cache hierarchy **optimized for data center workloads**
- Improved memory subsystem with **up to 60% higher memory bandwidth**
- Faster and more efficient Intel® UPI interconnect for **improved scalability**
- Improved integrated IO with **up to 50% higher aggregate IO bandwidth**

Balanced Hardware Configuration

ORACLE OPEN WORLD

- Always populate all memory channels
- Make sure you have sufficient storage for IOPs/BW and redundancy (not just capacity)
- Make sure you have enough CPU power
- Run an OS that supports your CPU

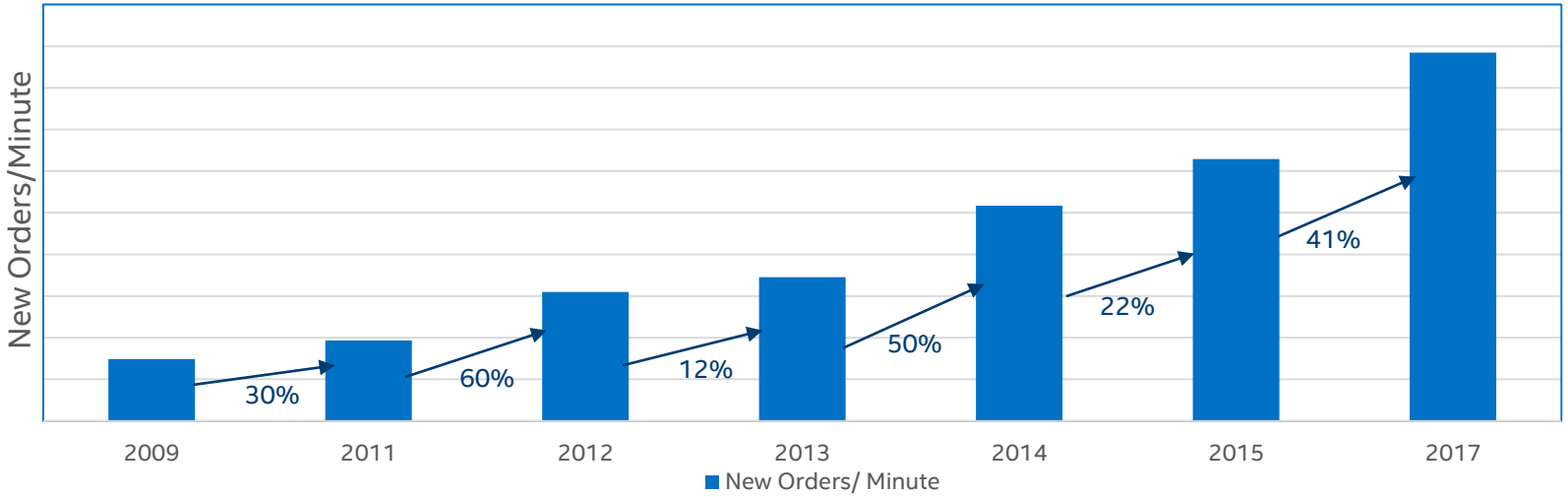
User controllable things you can check

ORACLE OPEN WORLD

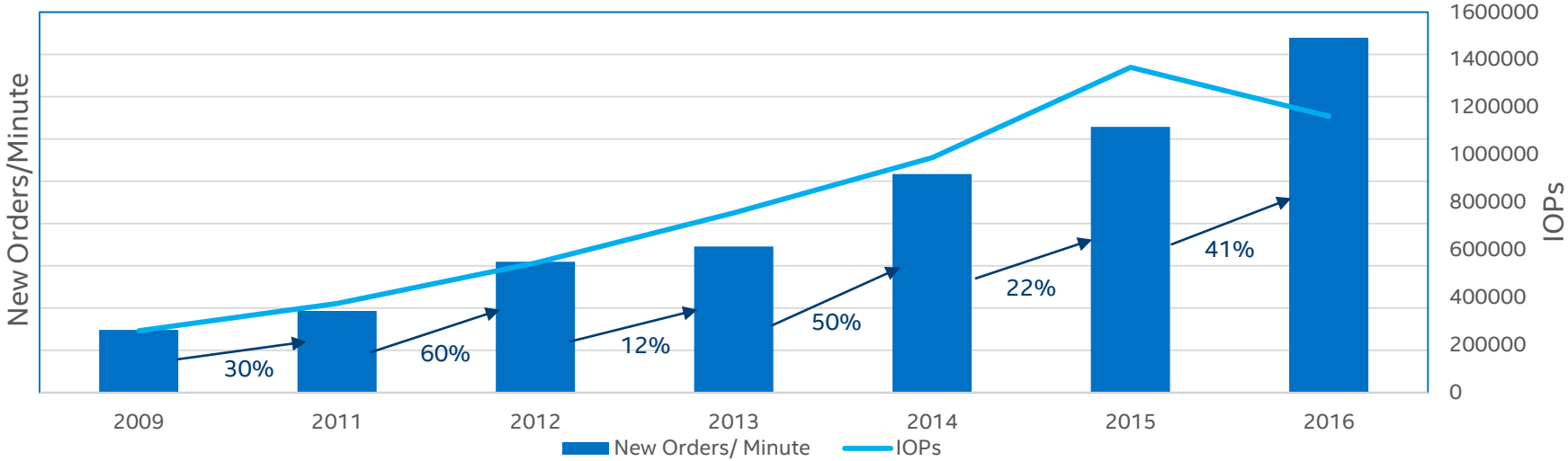
- Intel® Memory Latency Checker
- Use HT
- Use 2MB large pages
- Use process binding for better numa locality
- check queue depth and latency on storage
- Check c-states
  - Always a battle between power saving and maximum performance
- Check that you are using clocksource tsc (hpet can kill your perf)

OLTP PERFORMANCE IMPROVEMENTS OVER TIME

Generational OLTP Performance



Generational OLTP Performance (Disk IO scaling)



Device Type	Intel SSD X25-E	Intel SSD X25-E	Intel SSD 710	Intel SSD 710	Intel DC S3700	Intel DC S3700	Intel DC P46xx
# SSDs	272	325	200	384	96	96	14
# rack unit	44	48	24	48	24	24	2

ORACLE OPEN WORLD

NVME STORAGE CAN TAKE PERFORMANCE TO THE NEXT LEVEL

Intel P46xx NVMe storage

ORACLE OPEN WORLD

- Intel P4608 NVMe device
  - 6.4TB storage capacity
  - x8 PCIe add in card ➔ 6GB/sec read bandwidth
    - Actually measured!

Intel® Optane™ Technology: 3D XPoint™ Memory Media

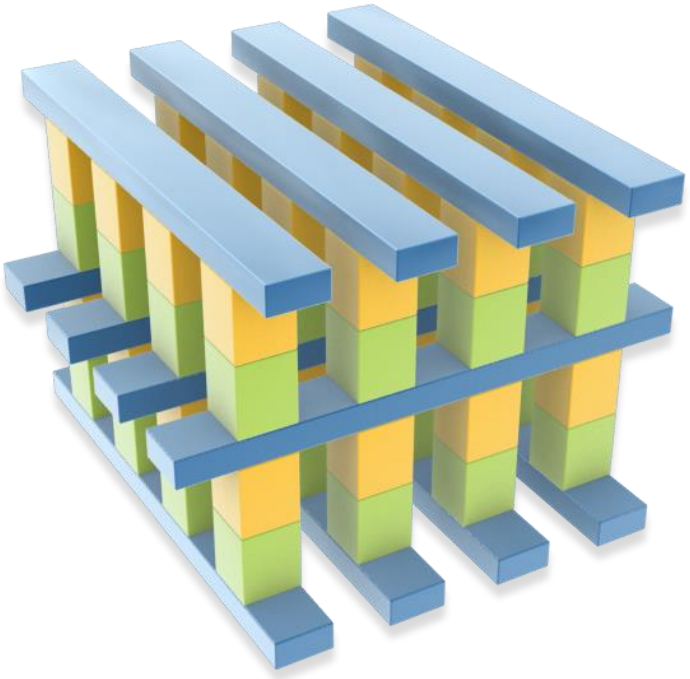
ORACLE OPEN WORLD

Cross Point Structure

Selectors allow dense packing and individual access to bits

Scalability

Memory layers can be stacked in a 3D manner



Breakthrough Material Advances

Compatible switch and memory cell materials

High Performance

Cell and array architecture that can switch states much faster than NAND

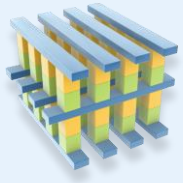
Intel® Optane™ Memory for Clients

ORACLE OPEN WORLD

Intel® Optane™ Memory

Fast Media

3D XPoint™ Memory Media



Intel Controller, Hardware and Interconnect IP

Standard M.2 connector Memory Module

PCIe* Gen3 x2 M.2 2280 & M.2 2242, NVMe*

Intelligent Software

Intel® Rapid Storage Technology

End-User Value

Accelerate System Performance

Accelerate day-to-day tasks with intuitive software that continually customizes and improves computer experience

Keep Your HDD Capacity

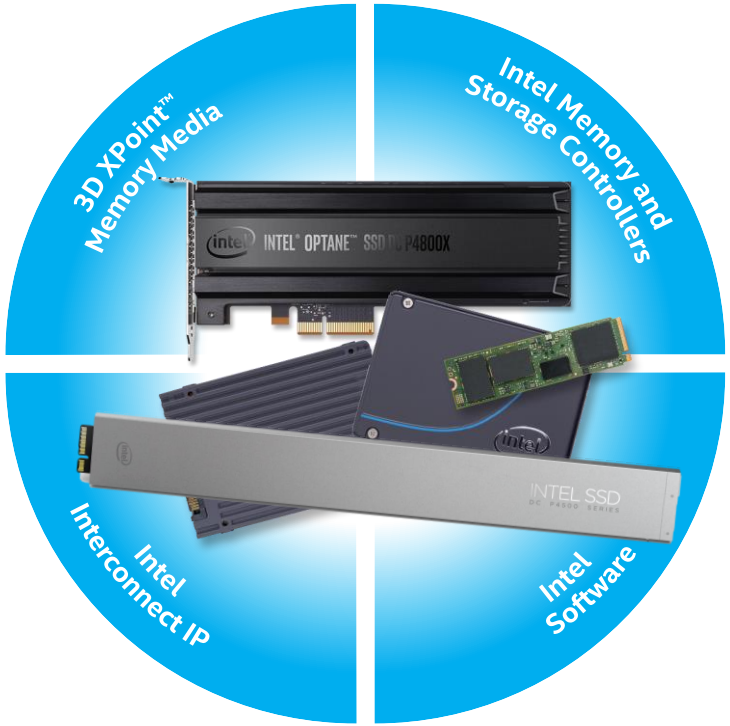
Pair Intel® Optane™ memory with an HDD for SSD-like speed and HDD capacity

Platform Confidence

The performance of a 7th Gen Intel® Core® processor coupled with the speed of Intel® Optane™ memory delivers amazing desktop computer responsiveness

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

Intel® Optane™ SSDs for Data Center



=

Ultra-high Endurance

Responsive Under Load Low Latency

Predictably Fast Service QoS

Breakthrough Performance IOPS

Technology claims are based on comparisons of latency, density and write cycling metrics amongst memory technologies recorded on published specifications of in-market memory products against internal Intel specifications. Intel® Optane™ SSD prototype compared to the Intel® SSD DC P3700 Series (NAND)

INTEL® OPTANE™ TECHNOLOGY

PCIe*/NVMe* Delivered in Many Form Factors

ORACLE OPEN WORLD

M.2

U.2, 2.5 in

Add-in Card

"Ruler" form factor

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

ORACLE OPEN WORLD

BUILDING AN OPEN-MEMORY-CENTRIC COMPUTING ARCHITECTURE USING INTEL OPTANE

Tuesday, Oct 03, 11:30 a.m. - 12:15 p.m. | Moscone West - Room 3003

In-Memory Database improvements

In-memory collaboration

ORACLE OPEN WORLD

- Intel Xeon vector instructions
- Oracle utilizes SSE4.2 and now AVX, AVX2, AVX512
- Why is that important?
  - Vector instructions enable Oracle to operate on multiple data elements with 1 instructions instead of multiple
- Transparent to customer, advantages with Intel Xeon processors
  - Wider registers, more operations per instruction

In-memory collaboration

ORACLE OPEN WORLD

- Doesn't stop there
  - Intel architects working with Oracle
    - Modify current algorithms to work better with vector instructions
    - Identify gaps in instructions and investigate additional vector instructions in future processors
- Great collaboration leads to great results

Performance Monitoring Tools

ORACLE OPEN WORLD

- Performance Co-Pilot
  (http://www.pcp.io/)

- Intel® Storage Snapshot tool
  (https://software.intel.com/sites/products/snapshots/storage-snapshot/)

- Intel® Memory Latency Checker
  (https://software.intel.com/en-us/articles/intelr-memory-latency-checker)

Performance Monitoring Tools (cont.)

ORACLE OPEN WORLD

- Linux Perf
- toplev tool (Top Down methodology with linux perf) (https://github.com/andikleen/pmu-tools/wiki)
- Turbostat
- Intel® Vtune Amplifier
- Intel® PCM

Future Work

ORACLE OPEN WORLD

- Persistent memory
- Acceleration of database features
- Fabric, including NVMEoF
- Security (very important in Cloud usage)

Acknowledgements

Frank Ober: (Intel) Optane and Intel SSD expertise.

Questions?


intel
ORACLE