PREPARE FOR THE NEXT GENERATION OF MEMORY:

IS YOUR APPLICATION A GOOD CANDIDATE?

Jackson Marusarz
Senior Technical Consulting Engineer
Compute Performance & Developer Products Division

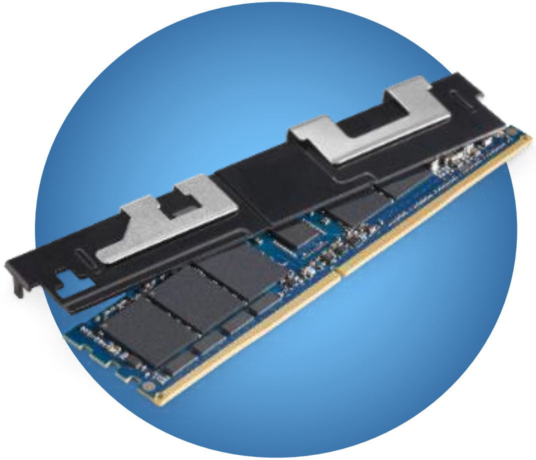
Agenda

- Brief introduction to Intel® Optane™ DC Persistent Memory
- Concepts and tools for enabling
  - Workload suitability analysis
  - Profiling and tuning workloads running with Intel Optane DC Persistent Memory
- Case Studies

THE BEST OF BOTH WORLDS WITH INTEL® OPTANE™ DC PERSISTENT MEMORY

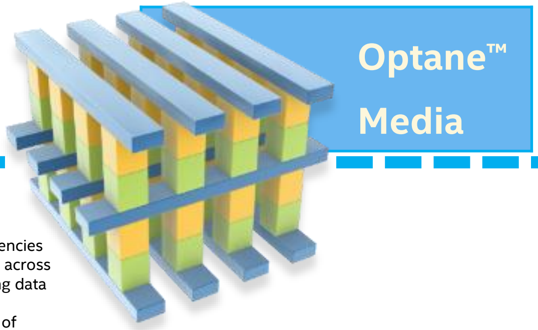
DRAM ATTRIBUTES

Performance comparable to DRAM at *low latencies*[1]



NAND SSD ATTRIBUTES

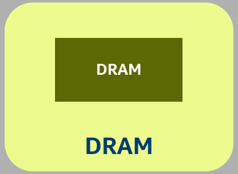
Data persistence with higher capacity than DRAM[2]



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.

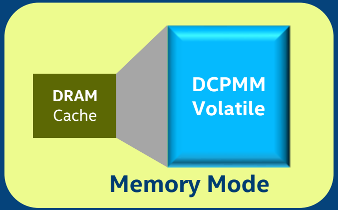
PROGRAMMING MODELS

Legacy
(DRAM)



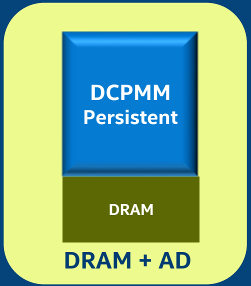
DRAM
1st Level Memory

Memory Mode
(MM)



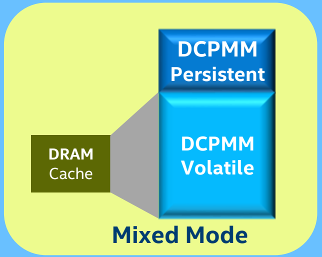
Memory Mode
(2nd Level Memory)
Volatile with DRAM caching

DRAM + App Direct
(AD)



App Direct (AD)
Persistent programming model
with no caching

Mixed Mode
(MM+AD)



Volatile	Volatile	Persistent or Volatile	Volatile or Mixed

Will I benefit from Intel® Optane™ DC Persistent Memory?

- Do you need: Big Memory - Fast Storage - Memory Resilience ?
- Are your workloads running out of DRAM memory?
- Is Disk I/O a large portion of your overhead?
- Does your warmup/data population phase takes a long time?

Use analysis tools to determine:

- How your system and applications may benefit from Intel® Optane™ DC Persistent Memory
- The best and easiest ways to take advantage of Intel® Optane™ DC Persistent Memory

Software Tools For Intel® Optane™ DC Persistent Memory

Intel® VTune™ Amplifier – Performance Analysis

- Platform Profiler – find configuration issues and potential for larger memory
- Memory analysis – design data structures for hot/warm/cool memory
- Memory analysis – tune use of DCPMM memory
- Storage analysis – are you CPU or I/O bound?

Intel® Inspector – Persistence Inspector

- Finds missing/redundant cache flushes, PMDK logging errors, and more

Use cases for tools with Intel® Optane™ DC Persistent Memory

Before you have hardware (workload suitability)

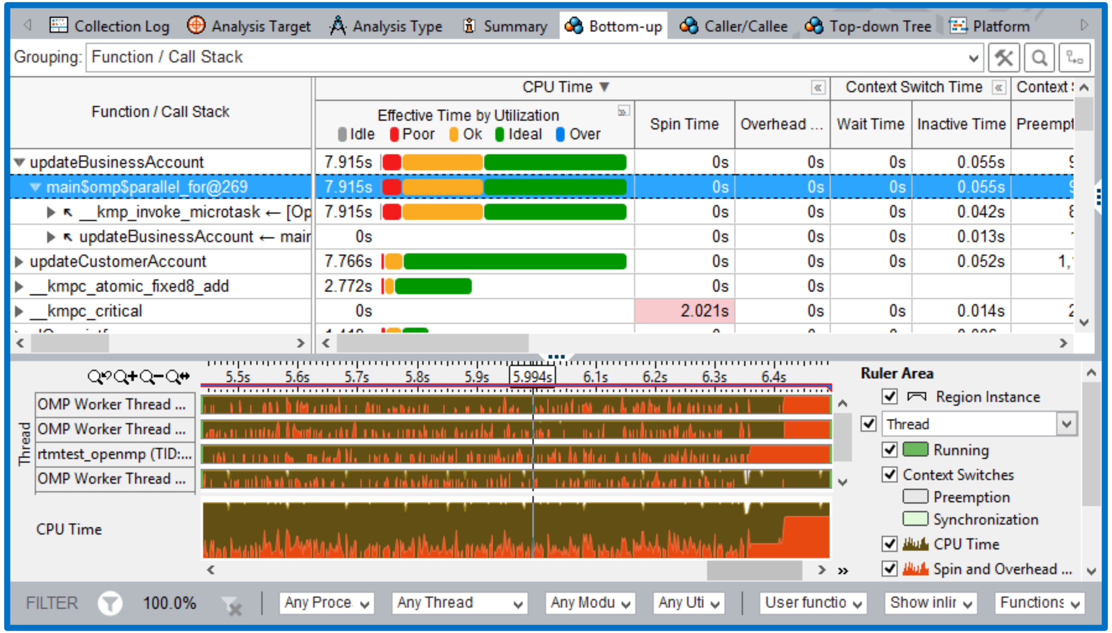
- Transitioning from DRAM-only to Memory Mode
- Transitioning from DRAM-only to App Direct (non-persistent mode)
- Transitioning from DRAM-only to App Direct (persistent mode)

After you have hardware

- Tuning existing Intel® Optane™ DC Persistent Memory usages

Intel® VTune™ Amplifier - Performance Profiler

Analyze & Tune Application Performance & Scalability

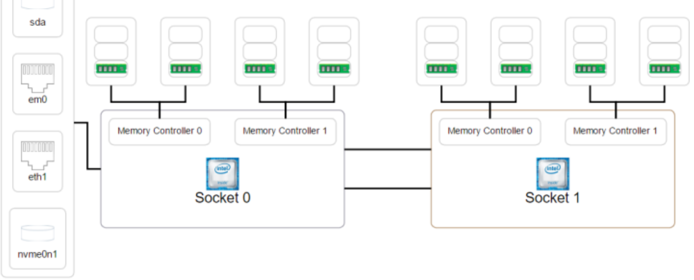


Faster, Scalable Code, Faster

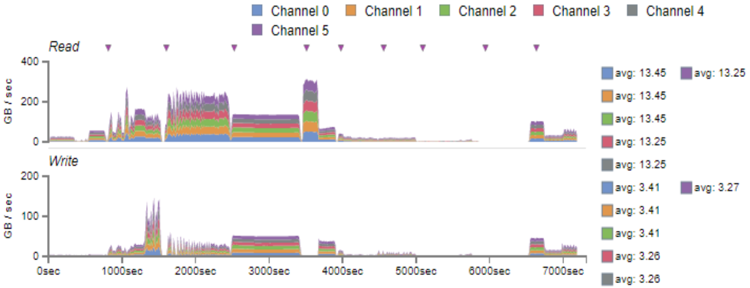
- Accurately profile C, C++, Fortran*, Python*, Go*, Java*, or any mix
- Optimize CPU/GPU, threading, memory, cache, MPI, storage & more
- Save time: rich analysis leads to insight
- Data displayed on the source code
- Easy set-up, no special compiles
- Cross-OS support and IDE integration

Intel® VTune™ Amplifier - Platform Profiler

Server Topology Overview



Traffic Patterns



Performance metrics on system topology

- Display current configuration
- Socket → Core → Internal Caches
- Socket → Memory Link → Memory Module

Identify system configuration issues

- Inefficient memory module placements
- Need for faster storage
- Need for larger/faster memory

Identify potential software issues

- Low CPU utilization
- NUMA-related issues (near vs. far memory accesses)
- Inefficient usage of memory/storage resources

Compare different system configurations

BEFORE YOU HAVE HARDWARE

- Transitioning from DRAM-only to Memory mode
- Transitioning from DRAM-only to AppDirect (non-persistent mode)
- Transitioning from DRAM-only to AppDirect (persistent mode)

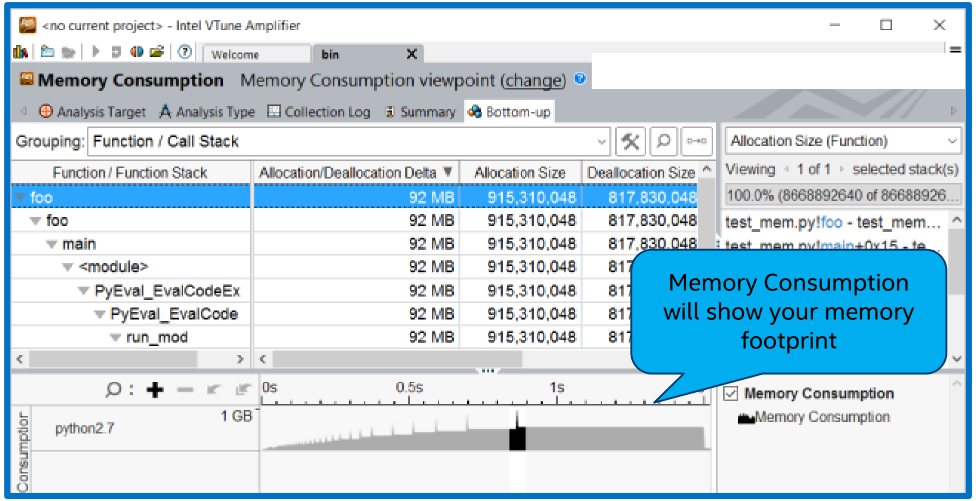
Intel® VTune™ Amplifier

DRAM-only -> Memory Mode (Big Memory - no code modification)

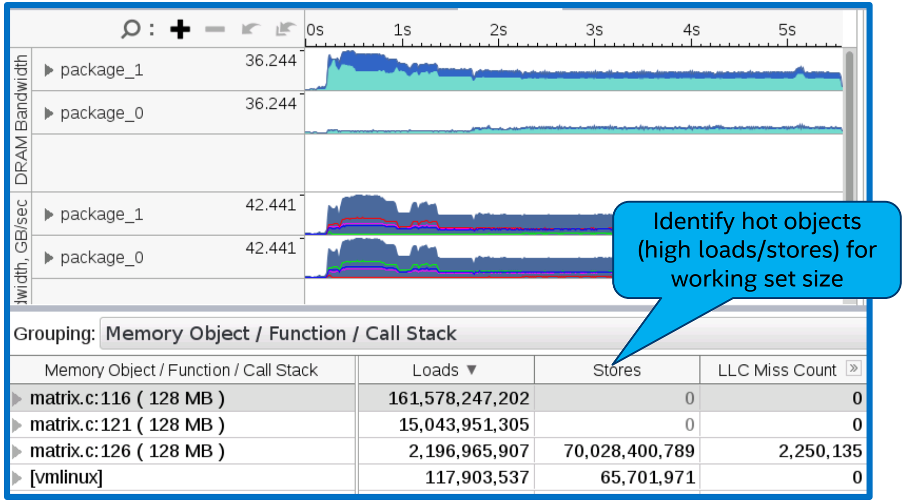
Memory Consumption

Memory Access

Look for applications with a memory footprint larger than DRAM but a hot working set size smaller than DRAM

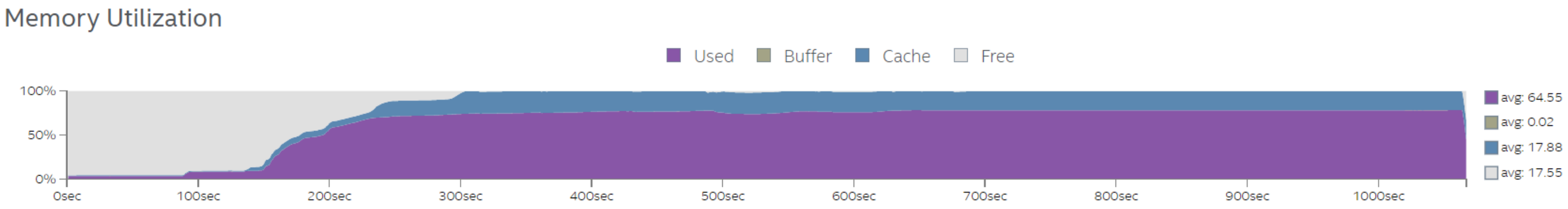


Memory Consumption Analysis



Memory Access Analysis + Dynamic Memory Object Analysis

Memory Consumption with Platform Profiler



- Profile system wide
- Longer running workloads
- Correlate with other platform profiler metrics

Memory Access

Intel® VTune™ Amplifier

DRAM-only ->App Direct (volatile) Mode (Big Memory - code modification required)

Identify objects to allocate in Intel® Optane™ DC Persistent Memory:

- For objects larger than DRAM – allocate in Intel® Optane™ DC Persistent Memory
- If an object is smaller than LLC – allocate in Intel® Optane™ DC Persistent Memory because it will likely be cached

Grouping: Bandwidth Domain / Bandwidth Utilization Type / Memory Object / Allocation Stack

Bandwidth Domain / Bandwidth Utilization Type / Memory Object / Allocation Stack	CPU Time	Memory Bound	Loads	Stores	LLC Miss Count	Average Latency (cycles)
DRAM, GB/sec	9.703s	64.3%	6,517,0 ...	4,141,26 ...	191,811,508	92
High	4.253s	56.8%	2,345,0 ...	2,111,23 ...	119,007,140	115
lin_stream.cpp:100 (152 M			910,002 ...	887,613, ...	51,453,087	119
lin_stream.cpp:99 (152 MB			826,002 ...	770,011, ...	39,902,394	91
lin_stream.cpp:98 (152 MB			609,001 ...	452,206, ...	27,651,659	142
[Unknown]			0	1,400,021	0	0
Medium	.880s	70.3%	2,765,0 ...	981,414, ...	52,853,171	83
Low	.571s	71.6%	1,407,0 ...	1,048,61 ...	19,951,197	57

Object Size

Hottest objects in Cache/DRAM

Warm objects in Intel Optane DC persistent memory

Cold objects on disk

Memory Access Analysis + Dynamic Memory Object Analysis

Memory Access

Intel® VTune™ Amplifier

DRAM-only ->App Direct (volatile) Mode (Big Memory - code modification required)

Intel® Optane™ DC Persistent Memory reads are faster than writes:

- Put load heavy objects in Intel® Optane™ DC Persistent Memory
- Put store heavy objects in DRAM

Identify load/store ratio

Grouping: Bandwidth Domain / Bandwidth Utilization Type ... cation Stack

Bandwidth Domain / Bandwidth Utilization Type / Memory Object / Allocation Stack	CPU Time	Memory Bound	Loads	Stores	LLC Miss Count	Average Latency (cycles)
DRAM, GB/sec	9.703s	64.3%	6,517,0 ...	4,141,26 ...	191,811,508	92
High	4.253s	56.8%	2,345,0 ...	2,111,23 ...	119,007,140	115
lin_stream.cpp:100 (152 MB			910,002 ...	887,613, ...	51,453,087	119
lin_stream.cpp:99 (152 MB			826,002 ...	770,011, ...	39,902,394	91
lin_stream.cpp:98 (152 MB			609,001 ...	452,206, ...	27,651,659	142
[Unknown]			0	1,400,021	0	0
Medium	2.880s	70.3%	2,765,0 ...	981,414, ...	52,853,171	83
Low	2.571s	71.6%	1,407,0 ...	1,048,61 ...	19,951,197	57

Memory Access Analysis + Dynamic Memory Object Analysis

Intel® VTune™ Amplifier

DRAM-only ->App Direct (non-volatile) Mode
Persistent Memory - code modification required

Input and Output

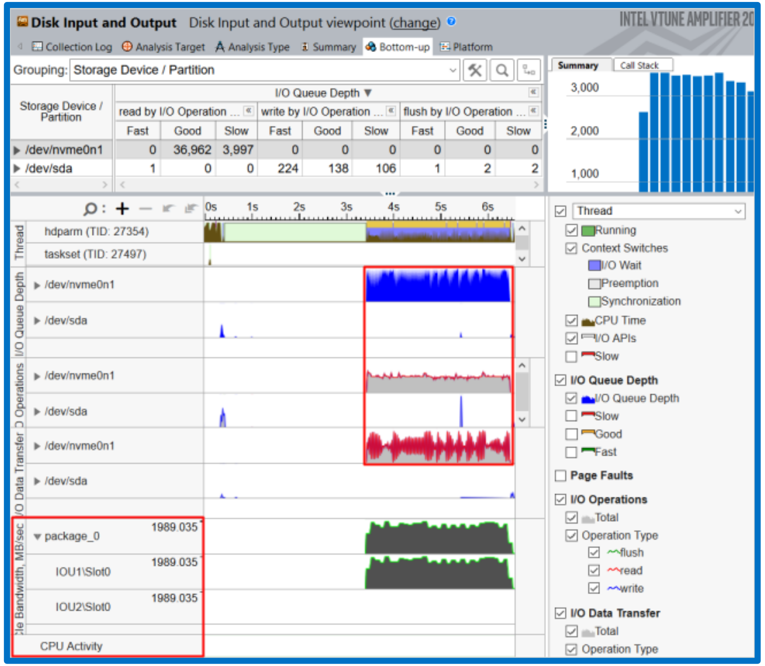
Identify disk related performance issues with Input and Output Analysis

Elapsed Time: 6.571s

I/O Wait Time:	2.113s
CPU Time:	**1.406s**
Instructions Retired:	2,391,093,000
CPI Rate:	1.109
CPU Frequency Ratio:	1.260
Total Thread Count:	263
Paused Time:	0s

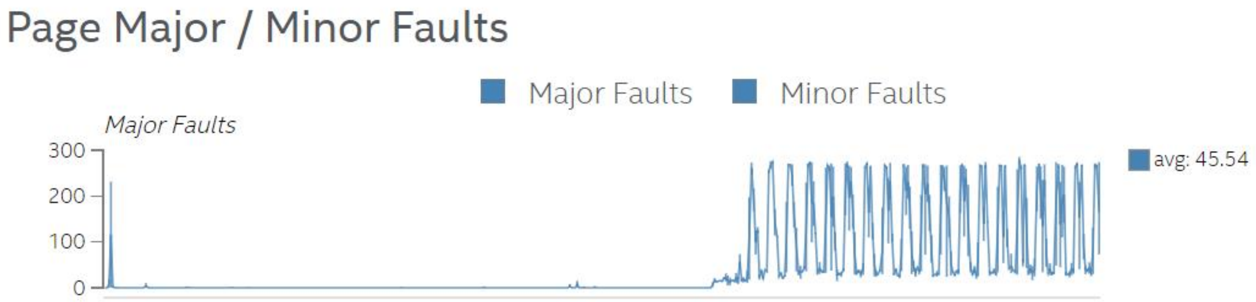
Time wasted waiting for disk

Correlate I/O stalls with software stats



Using Intel® Optane™ DC Persistent Memory for faster storage may alleviate these Disk-related performance issues

Disk Issues with Platform Profiler



AFTER YOU HAVE HARDWARE

Tuning existing Intel® Optane™ DC Persistent Memory usages

Tuning Intel® Optane™ DC Persistent Memory Systems

Memory Access

Elapsed Time: 22.976s

CPU Time:	22.612s	
Memory Bound:	**69.6%**	**of Pipeline Slots**
L1 Bound:	12.5%	of Clockticks
L2 Bound:	0.4%	of Clockticks
L3 Bound:	2.9%	of Clockticks
DRAM Bound:	**0.0%**	**of Clockticks**
Persistent Memory Bound:	**49.0%**	**of Clockticks**
Persistent Memory Bandwidth Bound:	0.0%	of Elapsed Time
Local Persistent Memory:	100.0%	of Clockticks
Remote Persistent Memory:	0.0%	of Clockticks
Loads:	10,519,515,576	
Stores:	4,292,228,763	
LLC Miss Count:	**36,902,214**	
Average Latency (cycles):	41	
Total Thread Count:	16	
Paused Time:	0s	

**N/A is applied to metrics with undefined value. There is no data to calculate the metric.*

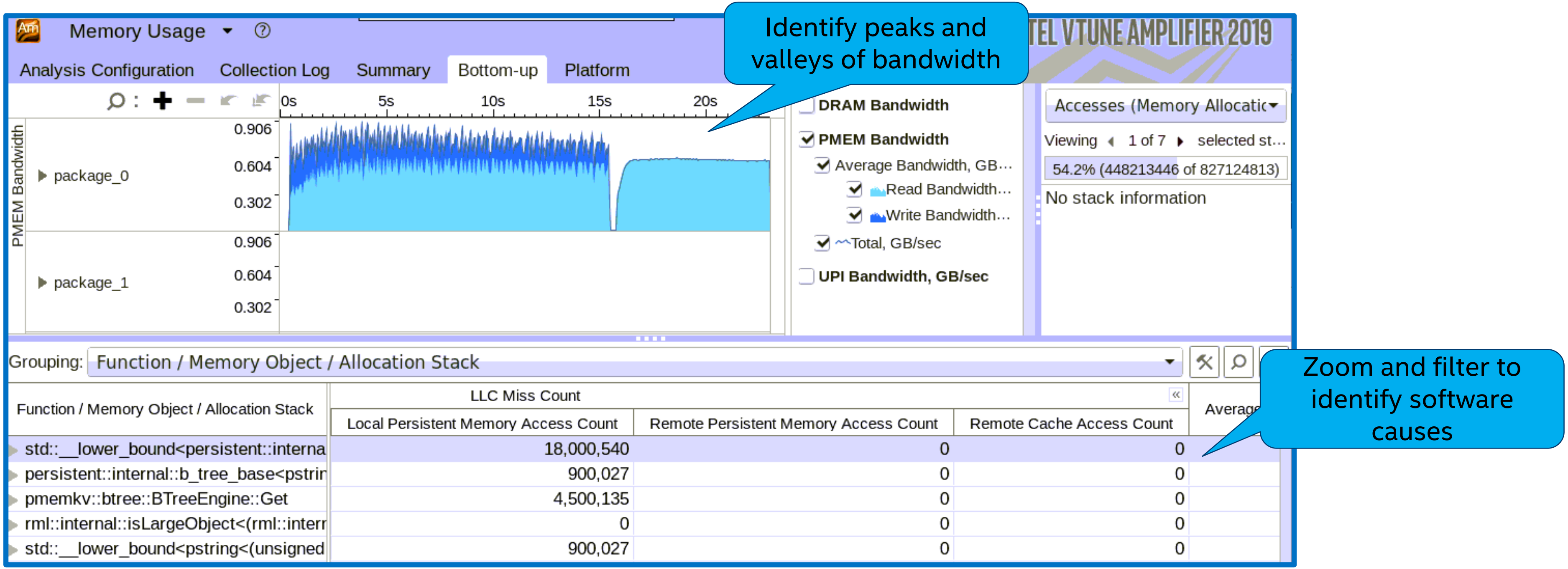
Bound by Intel® Optane™ DC Persistent Memory?

Tuning Intel® Optane™ DC Persistent Memory Systems

Intel® VTune™ Amplifier (Cont'd)

Memory Access

- View Intel® Optane™ DC Persistent Memory bandwidth over time
- Correlate data with CPU metrics and source code information

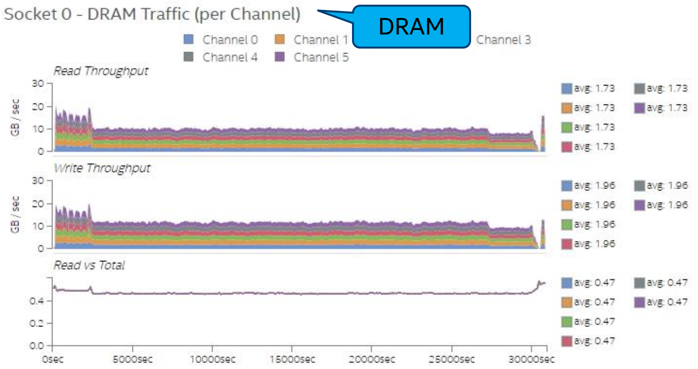


Function / Memory Object / Allocation Stack	LLC Miss Count			Averag
	Local Persistent Memory Access Count	Remote Persistent Memory Access Count	Remote Cache Access Count	
std::__lower_bound<persistent::interna	18,000,540	0	0	
persistent::internal::b_tree_base<pstrir	900,027	0	0	
pmemkv::btree::BTreeEngine::Get	4,500,135	0	0	
rml::internal::isLargeObject<(rml::interr	0	0	0	
std::__lower_bound<pstring<(unsigned	900,027	0	0	

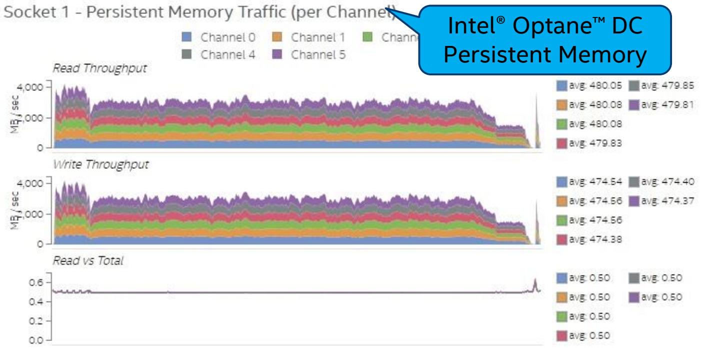
Tuning Intel® Optane™ DC Persistent Memory Systems

VTune™ Platform Profiler

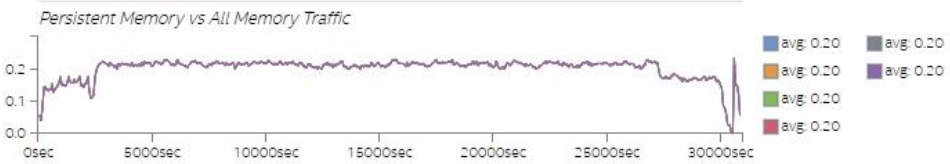
For Memory mode systems – make sure DRAM Bandwidth is much higher than Intel® Optane™ DC Persistent Memory Bandwidth



DRAM



Intel® Optane™ DC Persistent Memory



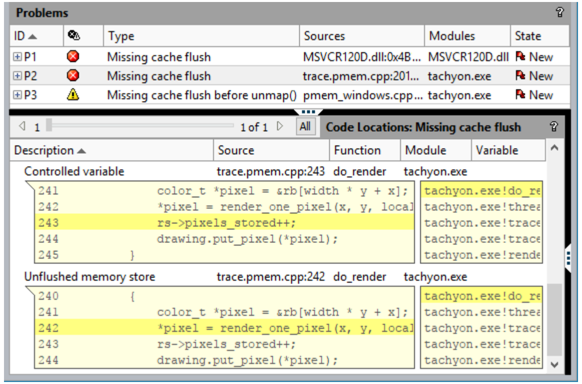
This ratio should be small

Find Missing/Extra Flushes/Commits

Intel® Inspector - Persistence Inspector

Target User

- Persistent memory programmers



What does it do?

- Finds persistent memory programming errors
- Detects:
  - Missing / redundant cache flushes
  - Missing store fences
  - Out-of-order persistent memory stores
  - PMDK transaction redo logging errors
- As a design tool, it finds places to insert flushes
- As a performance tool, it finds redundant flushes

[How to Use Intel® Inspector - Persistence Inspector]

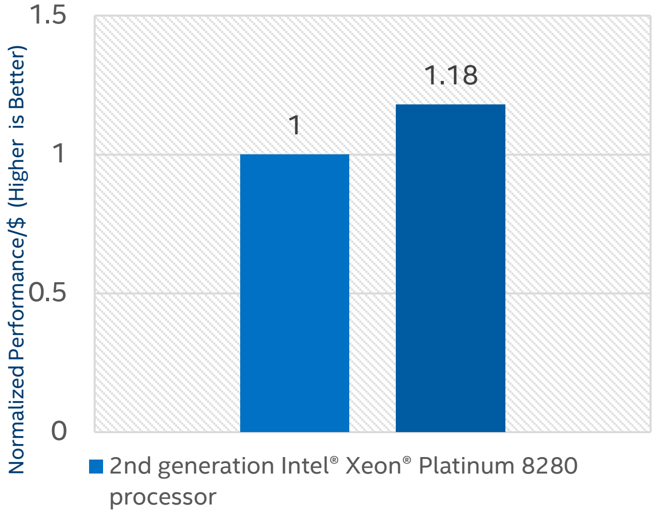
Note: PMDK = Persistent Memory Developer Kit (formerly NVML)

CASE STUDIES

http://www.sas.com

SAS (VIYA 3.4)*

Memory Mode



Performance Metric: Completion Time for 3 Concurrent Logistics Regression Tasks (400GB Datasets), Per $TCO (i.e., Perf/$TCO)[2]

APPLICATION

SAS* is a world leader in analytics and Artificial Intelligence. SAS Viya* provides a unified, open analytics platform replete with cutting-edge algorithms and AI capabilities. SAS Viya is a cloud- enabled, in-memory analytics engine that provides quick, accurate, and reliable analytical insights.

CUSTOMER CHALLENGES

- Customers are currently limited by memory capacity, which restricts the volume of datasets that can be stored close to the CPU, thereby limiting the potential to improve query response times. Expanding the memory footprint to overcome this challenge is often cost-prohibitive for customers.

SOLUTION

- With 2nd generation Intel® Xeon® Scalable processors and Intel® Optane™ DC persistent memory (Memory Mode), SAS can take advantage of larger available memory capacity per system, while making it a more cost-effective solution for customers.
- Customers can now keep multiple large datasets used for gradient boosting models in memory, with **little to no performance degradation**, and **at a reduced cost** (see chart showing up to 18% performance improvement for a given cost)[1].

VALUE PROPOSITION

- **Better performance at similar cost** - SAS customers can benefit from improved analytics response times, **with better TCO[1], and while meeting performance expectations**.

1 - Performance results are based on testing by Intel and SAS on 02/15/19 and may not reflect all publicly available security updates. No product or component can be absolutely secure. For complete testing configuration details, see Configuration Section.
2 - Pricing Guidance as of March 31, 2019 & valid until Jun 29, 2019. Intel does not guarantee any costs or cost reduction. You should consult other information and performance tests to assist you in your purchase decision.

Configuration Details

SAS® Viya®*; In-memory Analytics: SAS® Viya 3.4 VDMML application. Workload: 3 concurrent logistic regression tasks each running on 400GB datasets. Testing by Intel and SAS completed on February 15, 2019. Pricing Guidance as of March 31, 2019 & valid until Jun 29, 2019. Intel does not guarantee any costs or cost reduction. You should consult other information and performance tests to assist you in your purchase decision.

BASELINE: 2S Intel® Xeon® Platinum 8280 processor, 2.7GHz, 28 cores, turbo and HT on, BIOS SE5C620.86B.0D.01.0286.011120190816, 1536GB total memory, 24 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 1x 800GB, Intel SSD DC S3710 OS Drive + 1x 1.5TB Intel Optane SSD DC P4800X NVMe Drive for CAS_DISK_CACHE + 1x 1.5TB Intel SSD DC P4610 NVMe Drive for application data, CentOS Linux* 7.6 kernel 4.19.8.

NEW: 2S Intel® Xeon® Platinum 8280 processor, 2.7GHz, 28 cores, turbo and HT on, BIOS SE5C620.86B.0D.01.0286.011120190816, 1536GB Intel Optane DC persistent memory configured in Memory Mode(8:1), 12 slots / 128GB / 2666 MT/s, 192GB DRAM, 12 slots / 16GB / 2666 MT/s DDR4 LRDIMM, 1x 800GB, Intel SSD DC S3710 OS Drive + 1x 1.5TB Intel Optane SSD DC P4800X NVMe Drive for CAS_DISK_CACHE + 1x 1.5TB Intel SSD DC P4610 NVMe Drive for application data, CentOS Linux* 7.6 kernel 4.19.8.

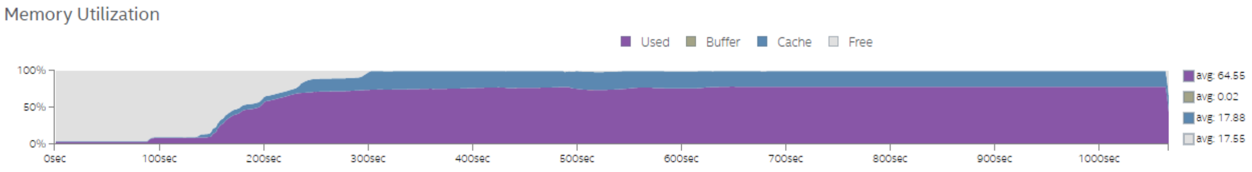
GraphX Workload Analysis

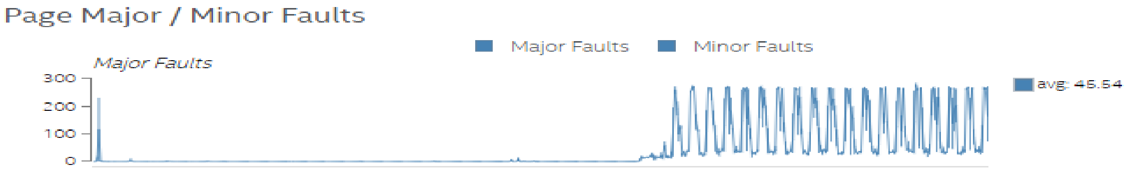
Runs with large datasets are failing

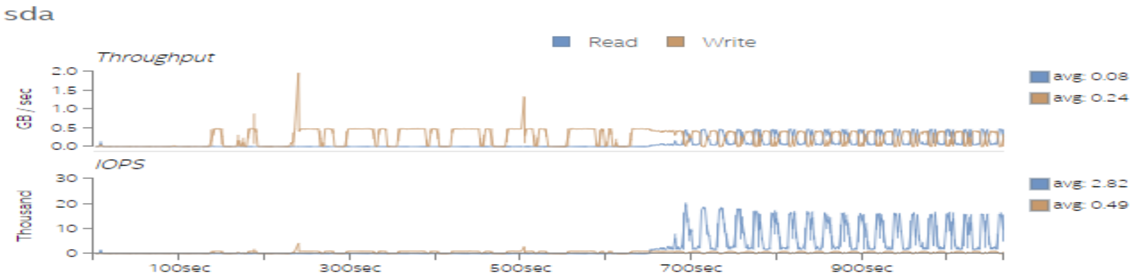
Application is using all available memory

Overflow causes page faults

Page faults require expensive disk reads







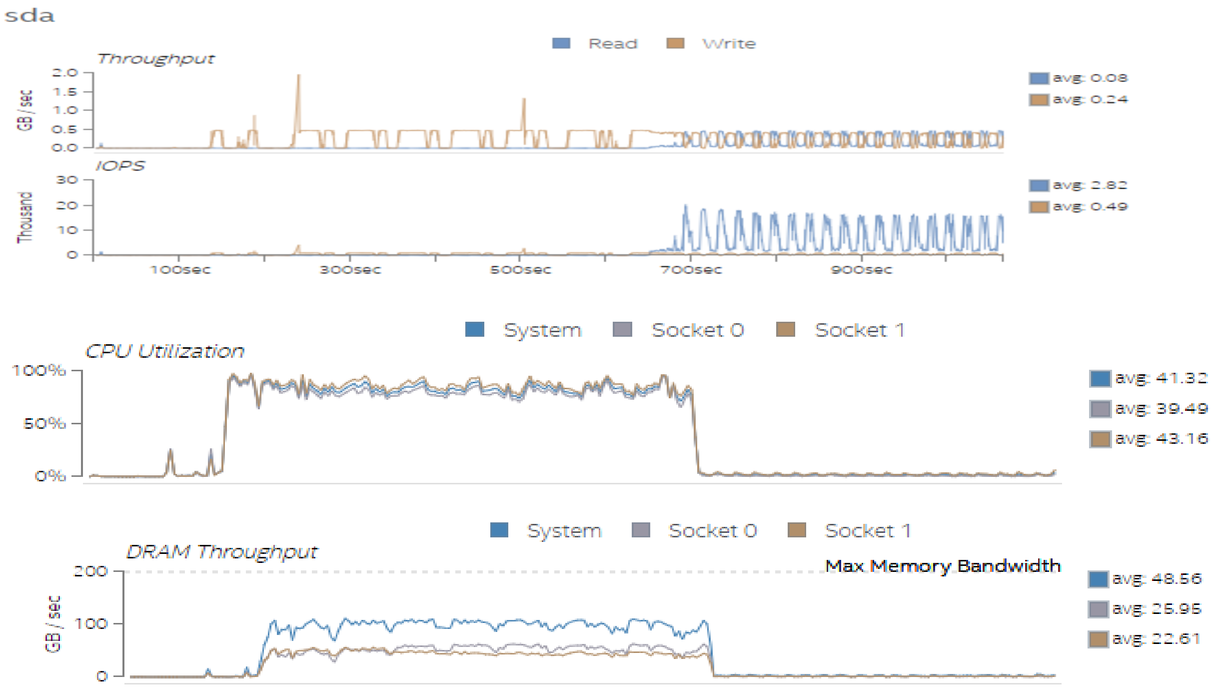
GraphX Workload Analysis

Runs with large scale factor are failing

Page faults and disk IO correlate on timeline

CPU Utilization drops

DRAM Throughput drops



Conclusion:

- Application is using all available memory and showing high disk and paging activity.
- The application is not bound by the memory bandwidth or CPU saturation. This is an indicator that application is bound by the capacity of the memory.
- Adding Intel® Optane™ DC Persistent Memory in Memory Mode allowed GraphX to scale to these larger datasets.

It's not just a raw performance calculation – it's TCO, performance, and scalability

Summary

- Intel® VTune™ Amplifier brings its best-in-class performance profiling and tuning to Intel® Optane™ DC Persistent Memory systems
- Including use cases:

  Before you have hardware

  - Transitioning from DRAM-only to Memory mode
  - Transitioning from DRAM-only to AppDirect (non-persistent mode)
  - Transitioning from DRAM-only to AppDirect (persistent mode)

  After you have hardware

  - Tuning existing Intel® Optane™ DC Persistent Memory usages

Legal Disclaimer & Optimization Notice

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.

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