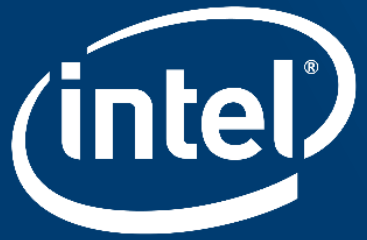


OPEN SOURCE DATABASES AND NON-VOLATILE MEMORY

Frank Ober

@fxober @IntelStorage

May 29, 2019

AGENDA

Primer on Intel® Optane™ technology and Intel® NVM

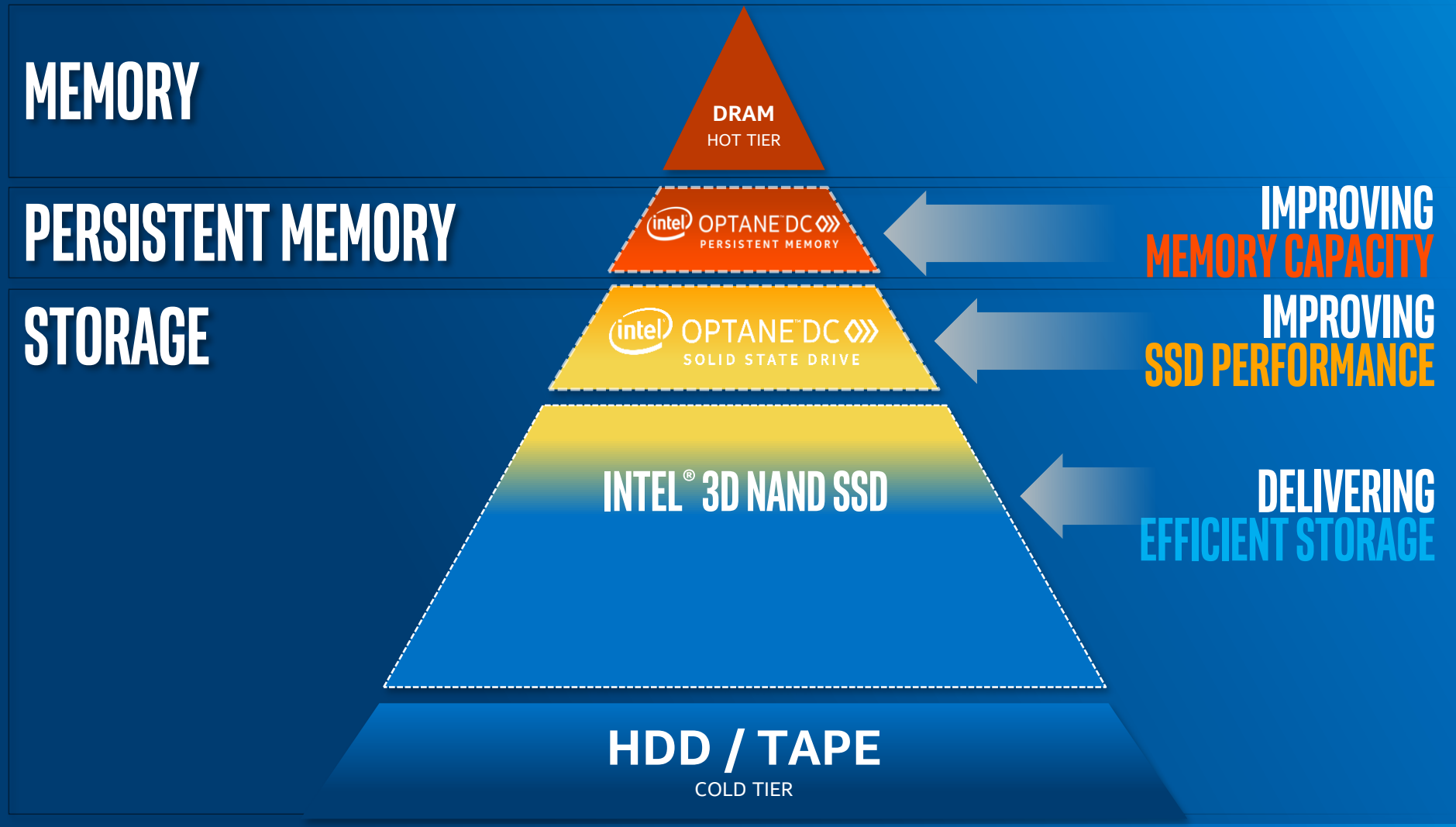
Benchmarks, Tools and Process

What can you do with it?

WHAT IS INTEL DOING WITH NVM?

INTEL® OPTANE™ TECHNOLOGY AND INTEL® 3D NAND SSD

CHANGING THE HIERARCHY



Latency in Hardware and Human Terms

- **Memory Operations** – 64B cache line – like getting an apple from the refrigerator
- **Storage Operations** – 512B or 4096B typical storage block, which is like getting a case of apples from the cellar
- Most storage solutions are unfortunately in the milliseconds still today. Applications are typically characterized in milliseconds.

A blink of an eye is about 100 milliseconds

Sources – Other – these numbers will and can vary by generation of hardware, this Is purely an exercise in understanding relative performance at the synthetic Hardware level.

HW Operation	Latency
L1 CPU Cache	1 ns (nanoseconds)
L2 CPU Cache	3 ns
L3 CPU Cache	10-20 ns
DRAM	70-100 ns
Intel® Optane™ DC PMM	< 1 us
Intel® Optane™ SSD	< 10 us (micro secs)
SSD SLC NAND TLC/QLC is slower	25 us
7200RPM HDD	8 - 10 ms
Network Latency New York to Europe	40 milliseconds

INTRODUCTION TO PERSISTENT MEMORY

- What is Persistent Memory?
  - Byte Addressable
  - Cache Coherent
  - Load /Store Access
    - No page caching
  - Memory-like Performance

- Why Does it Matter Now?
  - Adds a new tier between DRAM and Block Storage (SSD/HDD)
  - Large Capacity, High Endurance, Consistent low latency
  - Ability to do in-place persistence
    - No Paging, No Context Switching, No Interrupts, No Kernel Code
  - Ability to do DMA & RDMA

- What's the Impact on the Applications?
  - Need Ways to Enable Access to New High Performance Tier
  - May need to re-architect to unlock the new features and performance

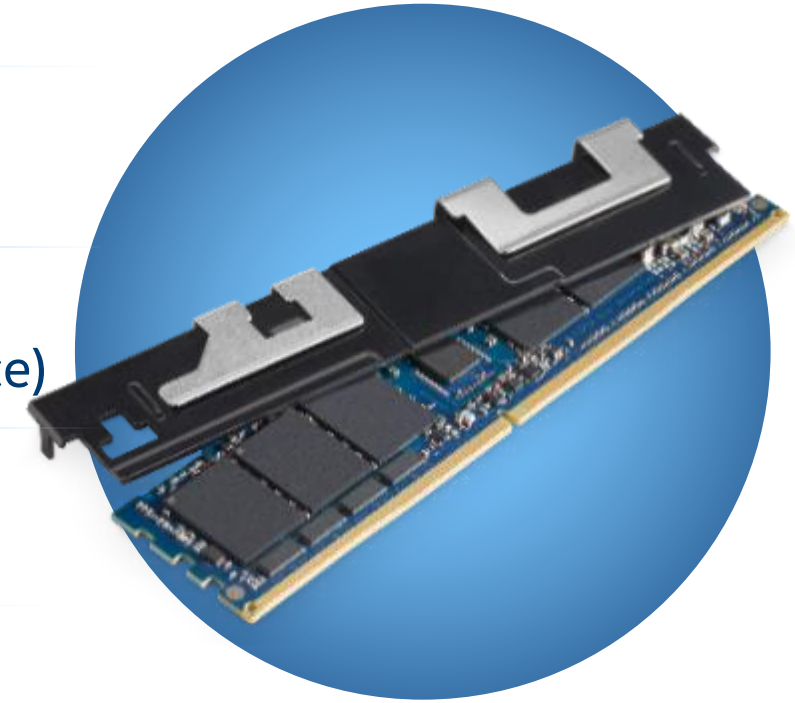
http://pmem.io/

INTEL® OPTANE™ DC PERSISTENT MEMORY – INTENDED VALUES & BENEFITS

Plenty and affordable memory

High performance storage (latency, bandwidth, QoS, endurance)

Application managed memory



More and extended VMs
Targeting > 1.2X VM at cost parity†

Capacity for In-Memory Database
at near-DRAM performance

Super-fast storage
Targeting > 3X NVMe* performance

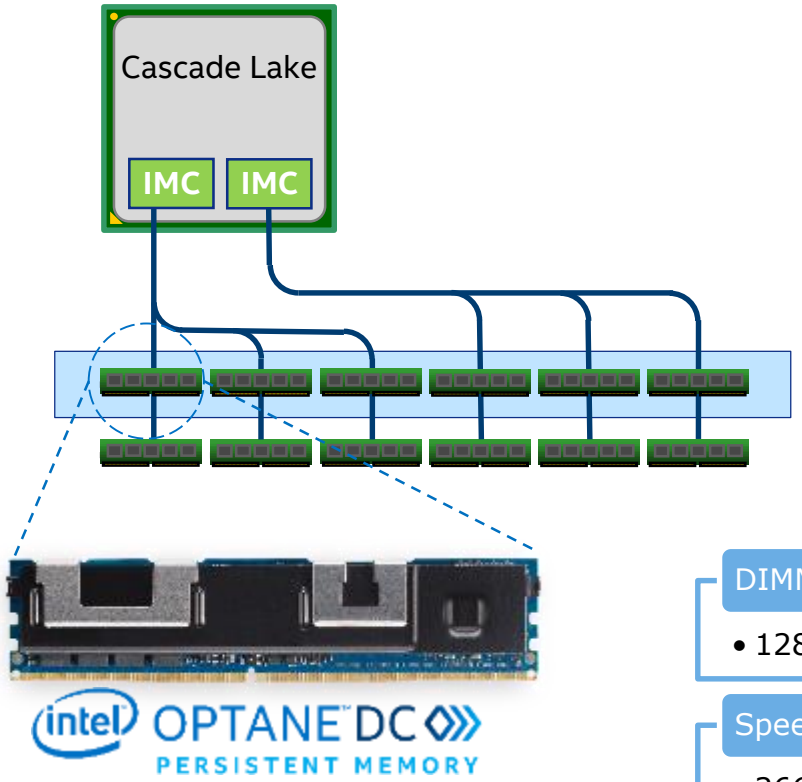
Targeting larger memory pools
Up to 3TB (not including DRAM)

PROVIDING LOWER AND CONSISTENT LATENCY WITH MORE CAPACITY PER DOLLAR

*other names and brands my be claimed as the property of others.
† Not all VM's are same and pending on usage and application, the results may differ.

INTEL® OPTANE™ DC PERSISTENT MEMORY - PRODUCT OVERVIEW

(Intel® Optane™ SSD-based Memory Module for the Data Center)



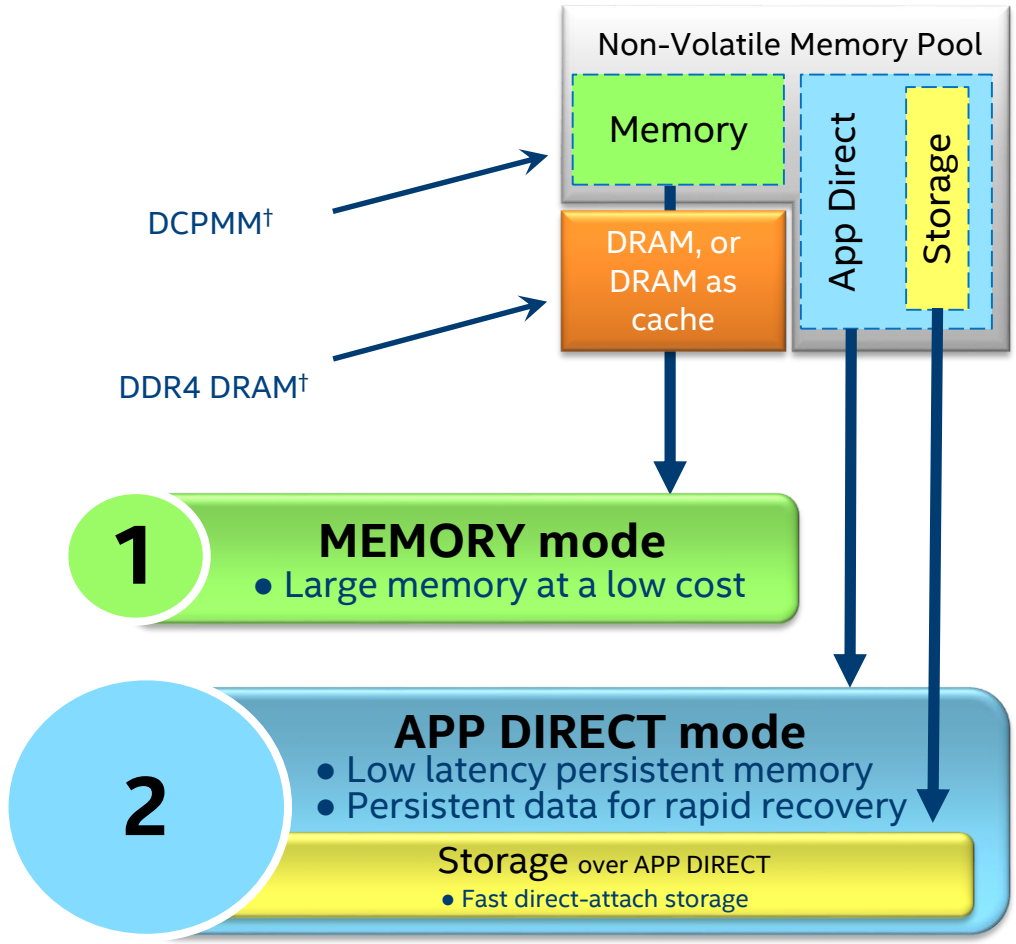
- DDR4 electrical & physical
- Close to DRAM latency
- Cache line size access

DIMM Capacity
- 128, 256, 512GB

Speed
- 2666 MT/sec

Capacity per CPU
- 3TB (not including DRAM)

Flexible, Usage Specific Partitions



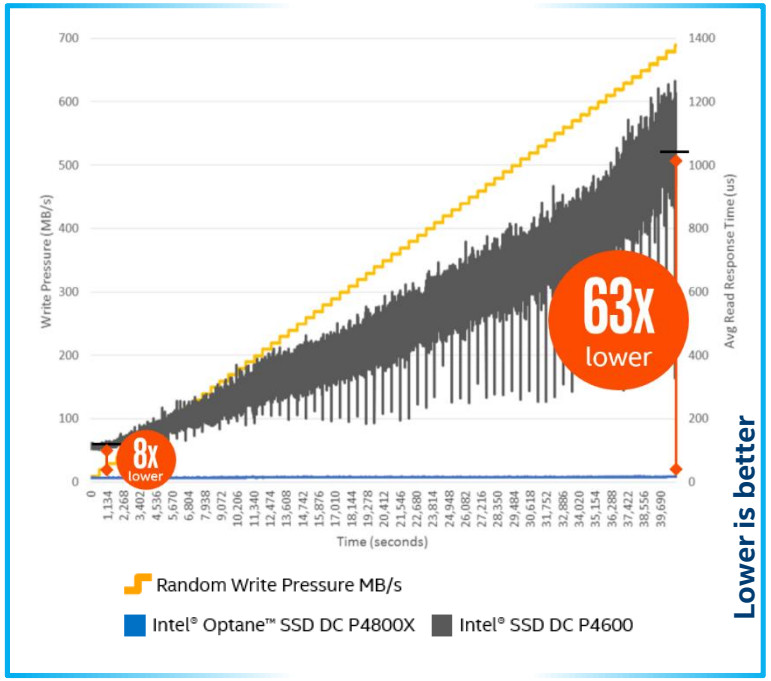
† DIMM population shown as an example only.

NOW FOR INTEL® OPTANE™ SSD

CACHING: INTEL® OPTANE™ SSD DC P4800X. THE IDEAL CACHING SOLUTION.

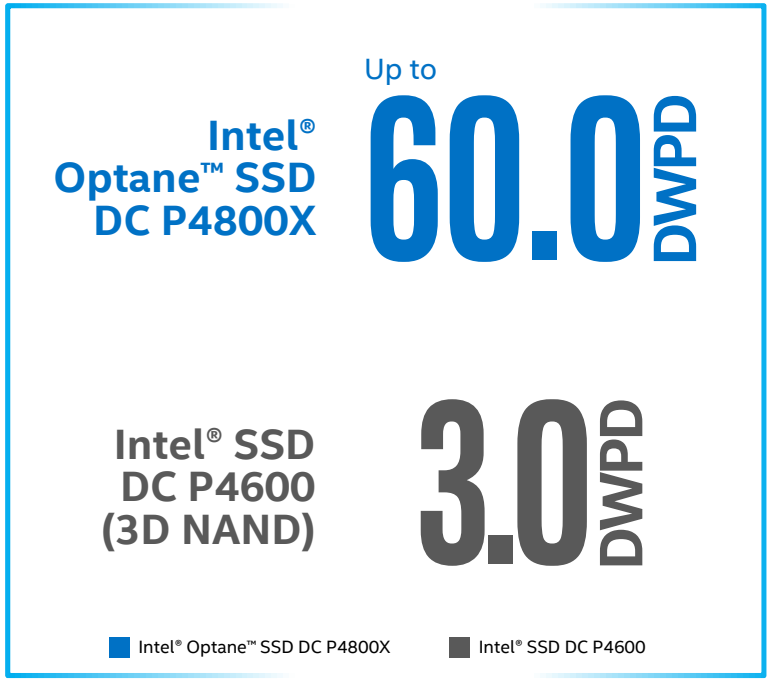
LOWER & MORE CONSISTENT LATENCY

Average Read Latency under Random Write Workload[1]



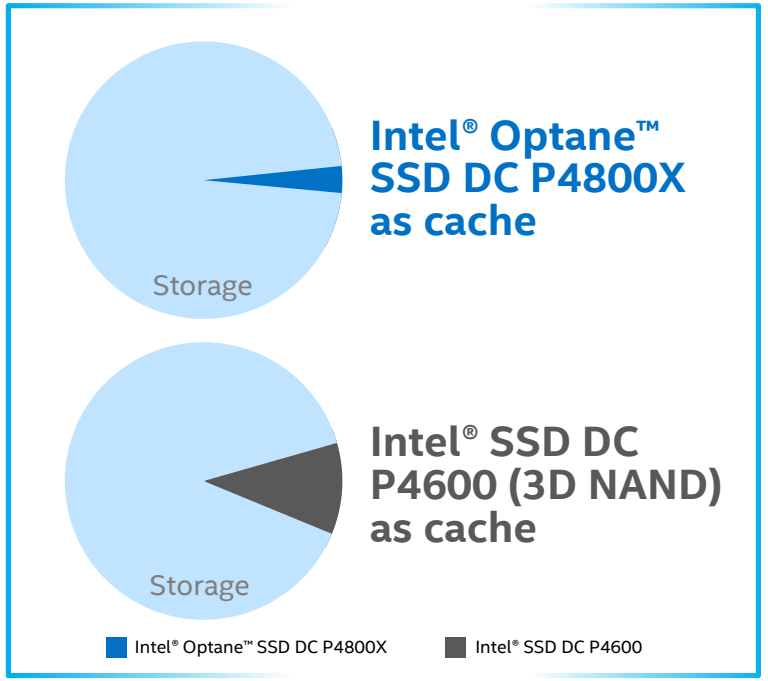
HIGHER ENDURANCE

Drive Writes Per Day (DWPD)[2]



MORE EFFICIENT

Cache as a % of Storage Capacity[3]



Low latency + high endurance **greater SDS system efficiency**

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.

1. Source – Intel-tested: Average read latency measured at queue depth 1 during 4k random write workload. Measured using FIO 3.1. Common Configuration - Intel 2U Server System, OS CentOS 7.5, kernel 4.17.6-1.el7.x86_64, CPU 2 x Intel® Xeon® 6154 Gold @ 3.0GHz (18 cores), RAM 256GB DDR4 @ 2666MHz. Configuration – Intel® Optane™ SSD DC P4800X 375GB and Intel® SSD DC P4600 1.6TB. Latency – Average read latency measured at QD1 during 4K Random Write operations using FIO 3.1. Intel Microcode: 0x2000043; System BIOS: 00.01.0013; ME Firmware: 04.00.04.294; BMC Firmware: 1.43.91f76955; FRUSDR: 1.43. SSDs tested were commercially available at time of test. Performance results are based on testing as of July 24, 2018 and may not reflect all publicly available security updates. See configuration disclosure for details. No product can be absolutely secure.
2. Source – Intel: Endurance ratings available at https://www.intel.com/content/www/us/en/products/docs/memory-storage/solid-state-drives/data-center-ssds/optane-ssd-dc-p4800x-p4801x-brief.html
3. Source – Intel: General proportions shown for illustrative purposes.

USAGE MODELS FOR OPTANE IN DATABASES

Create a Caching drive over SATA or QLC SSDs (Intel® Cache Acceleration Software)

Tier in Intel® Optane™ SSDs for specific cases

Write Logs or High Ingest drive

Temp Data drive

Temp, Log, Metadata	Cache	Tiering	OS Memory Extension	App Memory Persistence

BREAK TIME: WHAT ABOUT FSYNC ON INTEL® OPTANE™ SSD?

PERCONA

› Tools › Percona Germany › Percona France › Percona Live › About Us ›

Services Products Solutions

Home / InnoDB, MySQL, Insight for DBAs, Hardware and Storage / Fsync Performance on Storage Devices

Fsync Performance on Storage Devices

08 Feb 2018

Fsync Performance on Storage Devices

By Yves Trudeau Hardware and Storage, InnoDB, Insight for DBAs, MySQL

fsync, fsync performance, InnoDB, MySQL, rotating drives, SSD, Storage 8 Comments

While preparing a post on the design of ZFS based servers for use with MySQL, I stumbled on the topic of fsync call performance. The fsync call is very expensive, but it is essential to databases as it allows for *durability* (the "D" of the ACID acronym).

Let's first review the type of disk IO operations executed by InnoDB in MySQL. I'll assume the default InnoDB variable values.

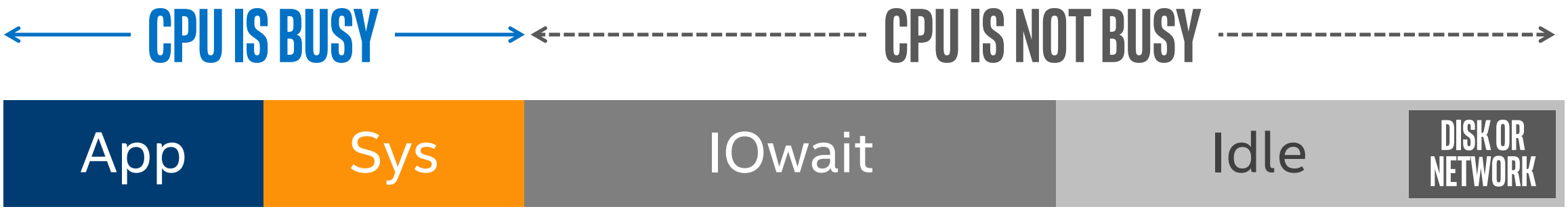
The first and most obvious type of IO are pages reads and writes from the tablespaces. The pages are most often read one at a time, as 16KB random read operations. Writes to the tablespaces are also typically 16KB random operations, but they are done in batches. After every batch, fsync is called on the tablespace file handle.

Source: Percona – used with permission of Percona , *Other names and brands may be claimed as the property of others.

SYSBENCH AND OTHER TESTING

IDENTIFY STORAGE-BOUND WORKLOADS TO OPTIMIZE YOUR PLATFORM

Where average data center workloads spends their time:



- Running the application
- Running the OS/VMM
- Waiting for Disk
- Limited application parallelism
- Imbalances
- Waiting for network
- Indirectly waiting for network (hidden IOwaits)

INTEL VTUNE PLATFORM PROFILER 2019 (FREE WITHOUT SUPPORT)

TPCC ON NVME* DEVICES…. INTEL® SSD DC P4510 (NVME NAND)

```
1  [||||||||                                  23.0%]   19 [||||||||                                  20.8%]   37 [||||||||                                  22.7%]   55 [||||||||                                  23.3%]
2  [||||||||                                  22.7%]   20 [||||||||                                  20.3%]   38 [||||||||                                  20.8%]   56 [||||||||                                  22.7%]
3  [||||||||                                  22.7%]   21 [||||||||                                  22.4%]   39 [||||||||                                  21.1%]   57 [||||||||                                  22.0%]
4  [||||||||                                  21.9%]   22 [||||||||                                  23.8%]   40 [||||||||                                  21.7%]   58 [||||||||                                  21.1%]
5  [|||||||                                   20.1%]   23 [||||||||                                  20.8%]   41 [||||||||                                  20.7%]   59 [||||||||                                  19.9%]
6  [|||||||                                   18.7%]   24 [|||||||                                   20.5%]   42 [||||||||                                  21.9%]   60 [||||||||                                  22.7%]
7  [||||||||                                  20.5%]   25 [|||||||                                   19.3%]   43 [||||||||                                  22.0%]   61 [|||||||                                   19.3%]
8  [||||||||                                  22.9%]   26 [||||||||                                  23.2%]   44 [|||||||                                   21.5%]   62 [||||||||                                  21.7%]
9  [|||||||                                   20.1%]   27 [||||||||                                  22.2%]   45 [||||||||                                  21.5%]   63 [||||||||                                  23.8%]
10 [||||||||                                  21.7%]   28 [|||||||                                   20.0%]   46 [|||||||                                   20.1%]   64 [||||||||                                  21.2%]
11 [||||||||                                  22.9%]   29 [||||||||                                  21.7%]   47 [|||||||                                   18.1%]   65 [||||||||                                  22.5%]
12 [||||||||                                  23.0%]   30 [|||||||                                   19.9%]   48 [||||||||                                  21.7%]   66 [||||||||                                  22.7%]
13 [||||||||                                  22.1%]   31 [|||||||                                   20.5%]   49 [||||||||                                  22.9%]   67 [||||||||                                  23.8%]
14 [||||||||                                  20.3%]   32 [|||||||                                   20.5%]   50 [|||||||                                   19.9%]   68 [||||||||                                  20.7%]
15 [||||||||                                  22.5%]   33 [||||||||                                  22.5%]   51 [||||||||                                  21.2%]   69 [||||||||                                  21.2%]
16 [||||||||                                  22.4%]   34 [||||||||                                  21.1%]   52 [||||||||                                  21.5%]   70 [|||||||                                   19.2%]
17 [||||||||                                  23.8%]   35 [||||||||                                  23.3%]   53 [||||||||                                  22.0%]   71 [||||||||                                  22.1%]
18 [||||||||                                  24.2%]   36 [|||||||                                   19.9%]   54 [||||||||                                  20.5%]   72 [||||||||                                  20.0%]
Mem[|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||34.5G/188G]   Tasks: 141, 583 thr; 36 running
Swp[                                                                      0K/0K]   Load average: 28.23 7.47 2.94
                                                                                   Uptime: 2 days, 05:10:39

  PID USER      PRI  NI  VIRT   RES   SHR S CPU% MEM%   TIME+  Command
43482 mysql      20   0 40.5G 32.5G 18760 S 1407 17.3 13:15.25 /usr/sbin/mysqld --daemonize --pid-file=/var/run/mysqld/mysqld.pid
43557 root       20   0 4306M 61276  6624 S 185.  0.0  1:47.93 sysbench tpcc --threads=64 --time=300 --tables=10 --percentile=99 --db-driver=mysql --mysql-db=tpc
```

```
avg-cpu:  %user   %nice %system %iowait  %steal   %idle
          27.15    0.00    5.85   20.04    0.00   46.96

Device:         rrqm/s   wrqm/s     r/s     w/s    rMB/s    wMB/s avgrq-sz avgqu-sz   await r_await w_await  svctm  %util
nvme1n1           0.00     0.00    0.00    0.00     0.00     0.00     0.00     0.00    0.00    0.00    0.00   0.00   0.00
nvme0n1           0.00     0.00    0.00    0.00     0.00     0.00     0.00     0.00    0.00    0.00    0.00   0.00   0.00
nvme2n1           0.00     0.00    0.00    0.00     0.00     0.00     0.00     0.00    0.00    0.00    0.00   0.00   0.00
nvme3n1           0.00 14468.00 27727.00 29082.00   433.23   704.10    41.00    41.63    0.40    0.24    0.56   0.02 101.10
```

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 visitwww.intel.com/benchmarks.

[1] System configuration: Intel Sourced: Tested: October 3, 2018. Server Intel® Server System , 2x Intel® Xeon® Scalable 6154, 384 GB DDR4 DRAM, database drives- 3x Intel® SSD DC P4800X Series (375 GB) and 1x Intel® SSD DC P4510 Series, CentOS 7.5 (kernel 4.18.8 (from elrepo)), BIOS: SE5C620.86B.00.01.0014.070920180847 system product type: S2600WFT
Percona MySQL Server 5.7.23, Sysbench 1.0.15 configured TPCC / sysbench per github project https://github.com/Percona-Lab/sysbench-tpcc 100GB database size. 30% Database Memory provided to MySQL (30GB). Performance results are based on testing as of Oct 3, 2018 and may not reflect all publicly available security updates. See configuration disclosure for details. No component or product can be absolutely secure.
*Other names and brands names may be claimed as the property of others

TPCC ON INTEL® OPTANE™ DC SSDs (NVME* INTEL® OPTANE™ MEMORY MEDIA)

```
1  [||||||||||||||||||||||||||||80.1%]   19 [||||||||||||||||||||||||||||80.4%]   37 [||||||||||||||||||||||||||||78.9%]   55 [||||||||||||||||||||||||||||79.9%]
2  [||||||||||||||||||||||||||| 77.3%]   20 [||||||||||||||||||||||||||||78.0%]   38 [||||||||||||||||||||||||||||79.2%]   56 [||||||||||||||||||||||||||||80.0%]
3  [||||||||||||||||||||||||||||79.3%]   21 [||||||||||||||||||||||||||||79.7%]   39 [||||||||||||||||||||||||||| 77.2%]   57 [||||||||||||||||||||||||||| 78.5%]
4  [||||||||||||||||||||||||||||78.0%]   22 [||||||||||||||||||||||||||| 77.2%]   40 [||||||||||||||||||||||||||| 77.9%]   58 [||||||||||||||||||||||||||| 78.0%]
5  [||||||||||||||||||||||||||||80.4%]   23 [||||||||||||||||||||||||||||80.1%]   41 [||||||||||||||||||||||||||||78.0%]   59 [||||||||||||||||||||||||||||78.7%]
6  [||||||||||||||||||||||||||| 77.9%]   24 [||||||||||||||||||||||||||||79.1%]   42 [||||||||||||||||||||||||||||80.1%]   60 [||||||||||||||||||||||||||||80.0%]
7  [||||||||||||||||||||||||||||80.0%]   25 [||||||||||||||||||||||||||| 80.4%]   43 [||||||||||||||||||||||||||||78.8%]   61 [||||||||||||||||||||||||||| 77.3%]
8  [||||||||||||||||||||||||||| 77.5%]   26 [||||||||||||||||||||||||||| 78.4%]   44 [||||||||||||||||||||||||||| 77.4%]   62 [||||||||||||||||||||||||||||78.1%]
9  [||||||||||||||||||||||||||||80.1%]   27 [||||||||||||||||||||||||||| 78.5%]   45 [||||||||||||||||||||||||||| 77.5%]   63 [||||||||||||||||||||||||||| 78.4%]
10 [||||||||||||||||||||||||||| 78.9%]   28 [||||||||||||||||||||||||||||80.1%]   46 [||||||||||||||||||||||||||||78.5%]   64 [||||||||||||||||||||||||||| 78.5%]
11 [||||||||||||||||||||||||||||79.6%]   29 [||||||||||||||||||||||||||||79.6%]   47 [||||||||||||||||||||||||||| 78.3%]   65 [||||||||||||||||||||||||||| 78.3%]
12 [||||||||||||||||||||||||||||78.3%]   30 [||||||||||||||||||||||||||||79.3%]   48 [||||||||||||||||||||||||||| 77.9%]   66 [||||||||||||||||||||||||||||79.3%]
13 [||||||||||||||||||||||||||| 78.4%]   31 [||||||||||||||||||||||||||| 79.5%]   49 [||||||||||||||||||||||||||||77.6%]   67 [||||||||||||||||||||||||||| 78.0%]
14 [||||||||||||||||||||||||||||79.5%]   32 [||||||||||||||||||||||||||||80.3%]   50 [||||||||||||||||||||||||||||78.8%]   68 [||||||||||||||||||||||||||| 77.6%]
15 [||||||||||||||||||||||||||||80.4%]   33 [||||||||||||||||||||||||||||79.5%]   51 [||||||||||||||||||||||||||| 78.2%]   69 [||||||||||||||||||||||||||| 78.9%]
16 [||||||||||||||||||||||||||| 79.9%]   34 [||||||||||||||||||||||||||||79.5%]   52 [||||||||||||||||||||||||||| 77.6%]   70 [||||||||||||||||||||||||||| 77.9%]
17 [||||||||||||||||||||||||||||79.9%]   35 [||||||||||||||||||||||||||||80.1%]   53 [||||||||||||||||||||||||||| 78.8%]   71 [||||||||||||||||||||||||||| 78.4%]
18 [||||||||||||||||||||||||||||80.8%]   36 [||||||||||||||||||||||||||| 80.1%]   54 [||||||||||||||||||||||||||| 78.0%]   72 [||||||||||||||||||||||||||| 77.6%]
Mem[|||||||||||||||||||||||||||||||||||||||||||||||||||||||||34.4G/188G]   Tasks: 139, 582 thr; 72 running
Swp[                                                          0K/0K]   Load average: 29.25 16.79 13.09
                                                                       Uptime: 2 days, 05:21:56

  PID USER      PRI  NI  VIRT   RES   SHR S CPU% MEM%   TIME+  Command
44077 mysql      20   0 40.5G 32.5G 18760 S 5066 17.3 26:39.81 /usr/sbin/mysqld --daemonize --pid-file=/var/run/mysqld/mysqld.pid
44140 root       20   0 4304M 60876  6944 S 659.  0.0  3:31.89 sysbench tpcc --threads=64 --time=300 --tables=10 --percentile=99 --db-driver=mysql --mysql-db=tpc
```

```
avg-cpu:  %user   %nice %system %iowait  %steal   %idle
          64.29    0.00   11.54    1.13    0.00   23.04

Device:         rrqm/s   wrqm/s     r/s     w/s    rMB/s    wMB/s avgrq-sz avgqu-sz   await r_await w_await  svctm  %util
nvme1n1           0.00  9300.00 14917.00 19332.00   233.08   470.99    42.10     1.64    0.00    0.00    0.00   0.02  54.00
nvme0n1           0.00  9344.00 14801.00 19229.00   231.27   470.10    42.21     1.58    0.00    0.00    0.00   0.02  56.10
nvme2n1           0.00  9204.00 14714.00 19224.00   229.91   470.96    42.29     1.60    0.00    0.00    0.00   0.02  55.50
nvme3n1           0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00    0.00    0.00    0.00   0.00   0.00
sda               0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00    0.00    0.00    0.00   0.00   0.00
sdb               0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00    0.00    0.00    0.00   0.00   0.00
md0               0.00     0.00 44425.00 88374.00   694.14  1419.52    32.60     0.00    0.00    0.00    0.00   0.00   0.00
```

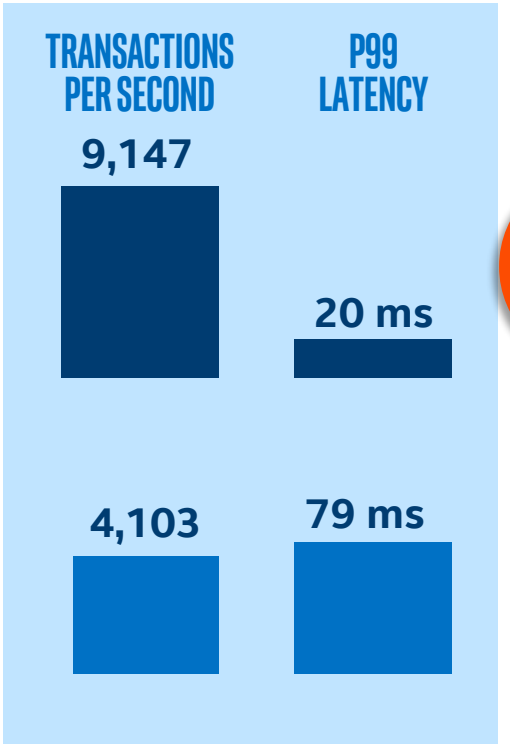
Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may 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. 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 visitwww.intel.com/benchmarks.

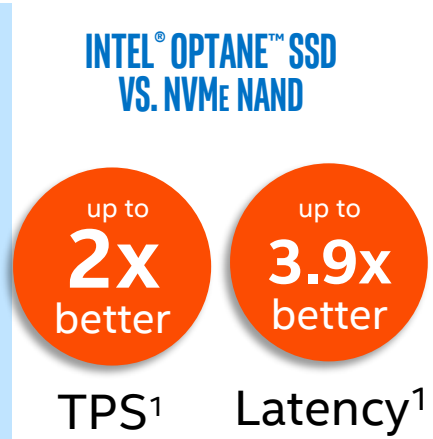
[1] System configuration: Intel Sourced: Tested: October 3, 2018. Server Intel® Server System , 2x Intel® Xeon® Scalable 6154, 384 GB DDR4 DRAM, database drives- 3x Intel® SSD DC P4800X Series (375 GB) and 1x Intel® SSD DC P4510 Series, CentOS 7.5 (kernel 4.18.8 (from elrepo)), BIOS: SE5C620.86B.00.01.0014.070920180847 system product type: S2600WFT

Percona MySQL Server 5.7.23, Sysbench 1.0.15 configured TPCC / sysbench per github project https://github.com/Percona-Lab/sysbench-tpcc 100GB database size. 30% Database Memory provided to MySQL (30GB). Performance results are based on testing as of Oct 3, 2018 and may not reflect all publicly available security updates. See configuration disclosure for details. No component or product can be absolutely secure. *Other names and brands may be claimed as the property of others.

PERCONA* MySQL* 5.7 TPCC TEST (64 THREADS, 10 TABLES, SCALE 100)

		STORAGE	CPU	DATABASE MEMORY
OPTANE	Meet latency SLA, trade-off DRAM for cheaper storage at high scale.	Intel® Optane™ DC SSD Intel® Optane™ SSD DC P4800X	Intel® Xeon® Intel® Xeon® Scalable (HT-on 72 core)	DRAM 30GB
NAND	Great consolidation benefits with NVMe*.	Intel® 3D NAND SSD – PCIe* Intel® SSD DC P4510	Intel® Xeon® Intel® Xeon® Scalable (HT-on 72 core)	DRAM 30GB





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 visitwww.intel.com/benchmarks.

[1] System configuration: Intel Sourced: Tested: October 3, 2018. Server Intel® Server System , 2x Intel® Xeon® Scalable 6154, 384 GB DDR4 DRAM, database drives- 3x Intel® SSD DC P4800X Series (375 GB) and 1x Intel® SSD DC P4510 Series, CentOS 7.5 (kernel 4.18.8 (from elrepo)), BIOS: SE5C620.86B.00.01.0014.070920180847 system product type: S2600WFT

Percona MySQL Server 5.7.23, Sysbench 1.0.15 configured TPCC / sysbench per github project https://github.com/Percona-Lab/sysbench-tpcc 100GB database size. 30% Database Memory provided to MySQL (30GB). Performance results are based on testing as of Oct 3, 2018 and may not reflect all publicly available security updates. See configuration disclosure for details. No component or product can be absolutely secure.

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

ORACLE* MySQL* 8.0 OLTP RW TEST (32 THREADS, 8 TABLES X 50M ROWS)

INTEL® CACHE ACCELERATION SOFTWARE (INTEL® CAS) FOR LINUX 3.6

CACHE Intel® CAS Cache using an Intel® Optane™ SSD and coalesce writes to NAND. Accelerate reads with lower latency device.

SATA SATA is a baseline configuration for many.

	STORAGE	CPU	DATABASE MEMORY	TRANSACTIONS PER SECOND	AVG LATENCY
CACHE	Intel® Optane™ DC SSD + D3-S4510 20G / 100G	Intel® Xeon® Scalable (HT-on 72 core)	Database Memory 30GB	10,007	3.2 ms
SATA	Intel® 3D NAND SSD – SATA Intel® SSD D3-S4510 100GB DB	Intel® Xeon® Scalable (HT-on 72 core)	Database Memory 30GB	3,060	10.45 ms

INTEL® OPTANE™ SSD AS A CACHE DRIVE VS. SATA

up to 3x better TPS[1]

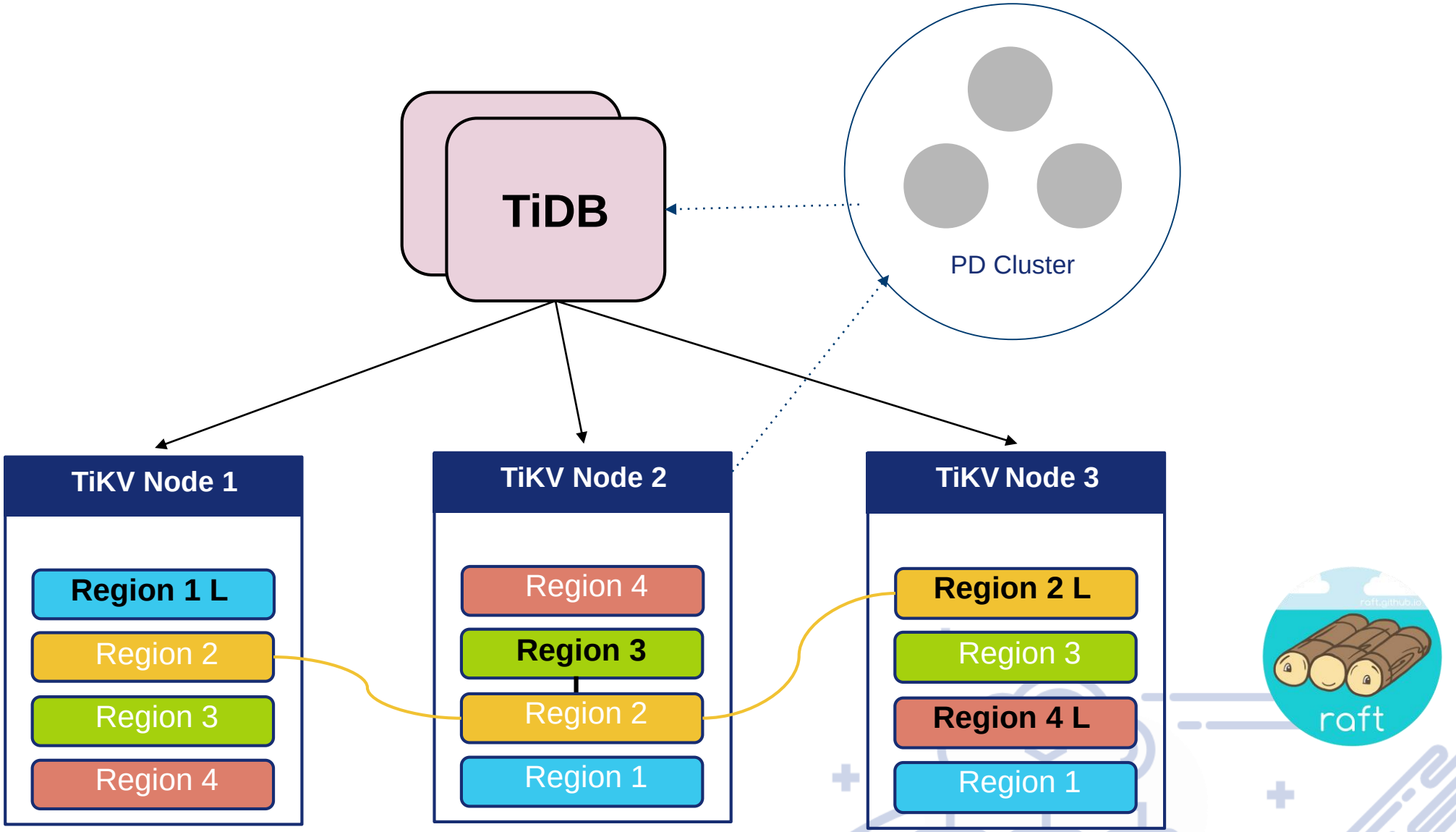
up to 3x better Average Latency[1]

TESTED WITH PARETO 80/20 RANDOMIZATION

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. 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 visitwww.intel.com/benchmarks.

[1] System configuration: Intel Sourced: Tested: October 20, 2018. Server Intel® Server System , 2x Intel® Xeon® Scalable 6154, 384 GB DDR4 DRAM, database drives- 2x Intel® SSD DC P4800X Series (375 GB) and 1x Intel® SSD DC P4510 Series, 1x Intel® SSD DC S4510 Series, CentOS 7.5 (kernel 4.18 (elrepo)), BIOS: SE5C620.86B.00.01.0014.070920180847 system product type: S2600WFT, Intel version of Intel CAS used version 3.6 set in write back mode over a single S4510 SSD.
MySQL Server 8.0.13, Sysbench 1.0,15 configured for 70/30 Read/Write OLTP transaction split using a 100GB database. 30% Database Memory provided to MySQL (30GB). Performance results are based on testing as of [Oct 20, 2018] and may not reflect all publicly available security updates. See configuration disclosure for details. No component or product can be absolutely secure. *Other names and brands names may be claimed as the property of others

TIDB – LOOKING AT AN OPEN HTAP DB

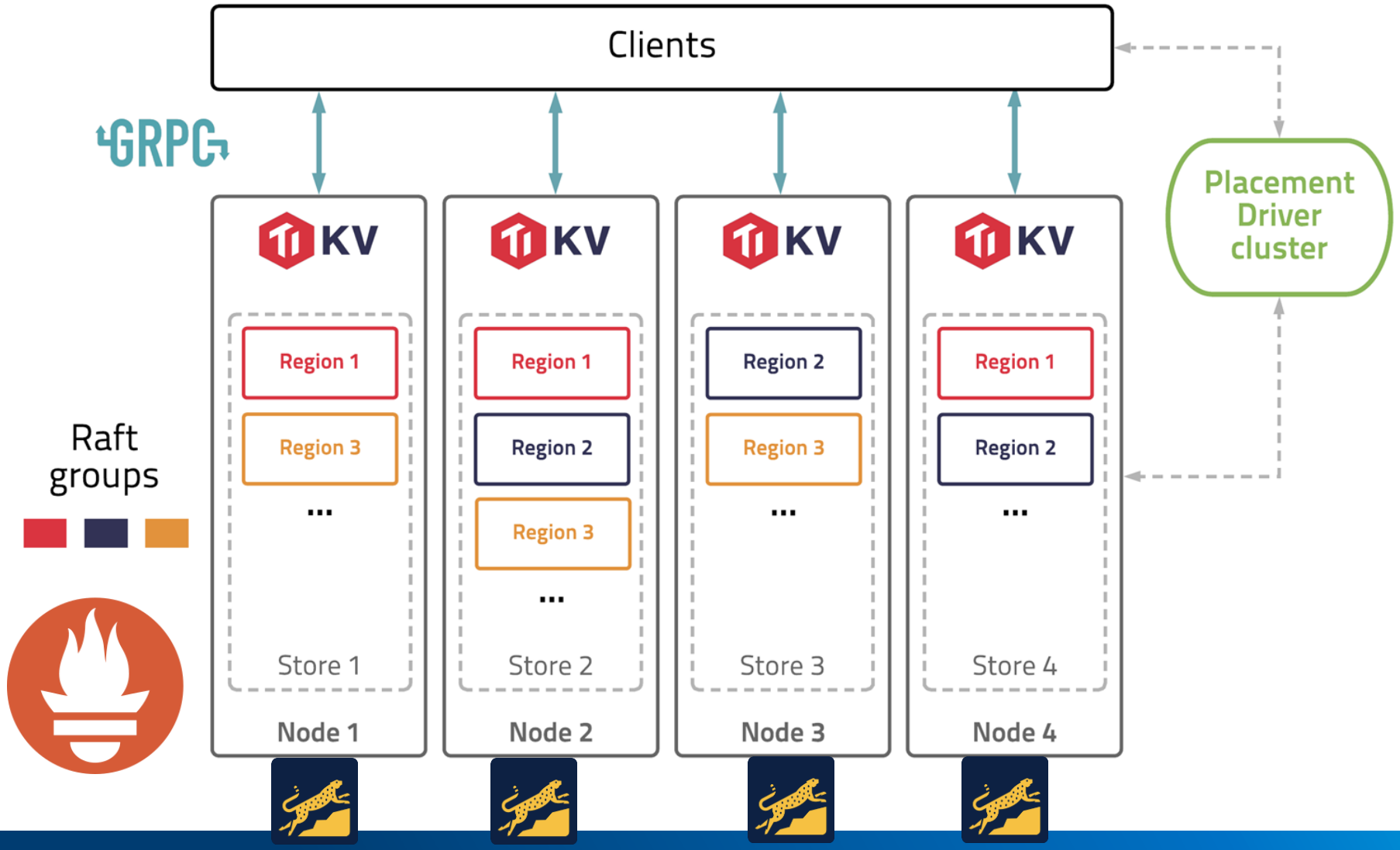


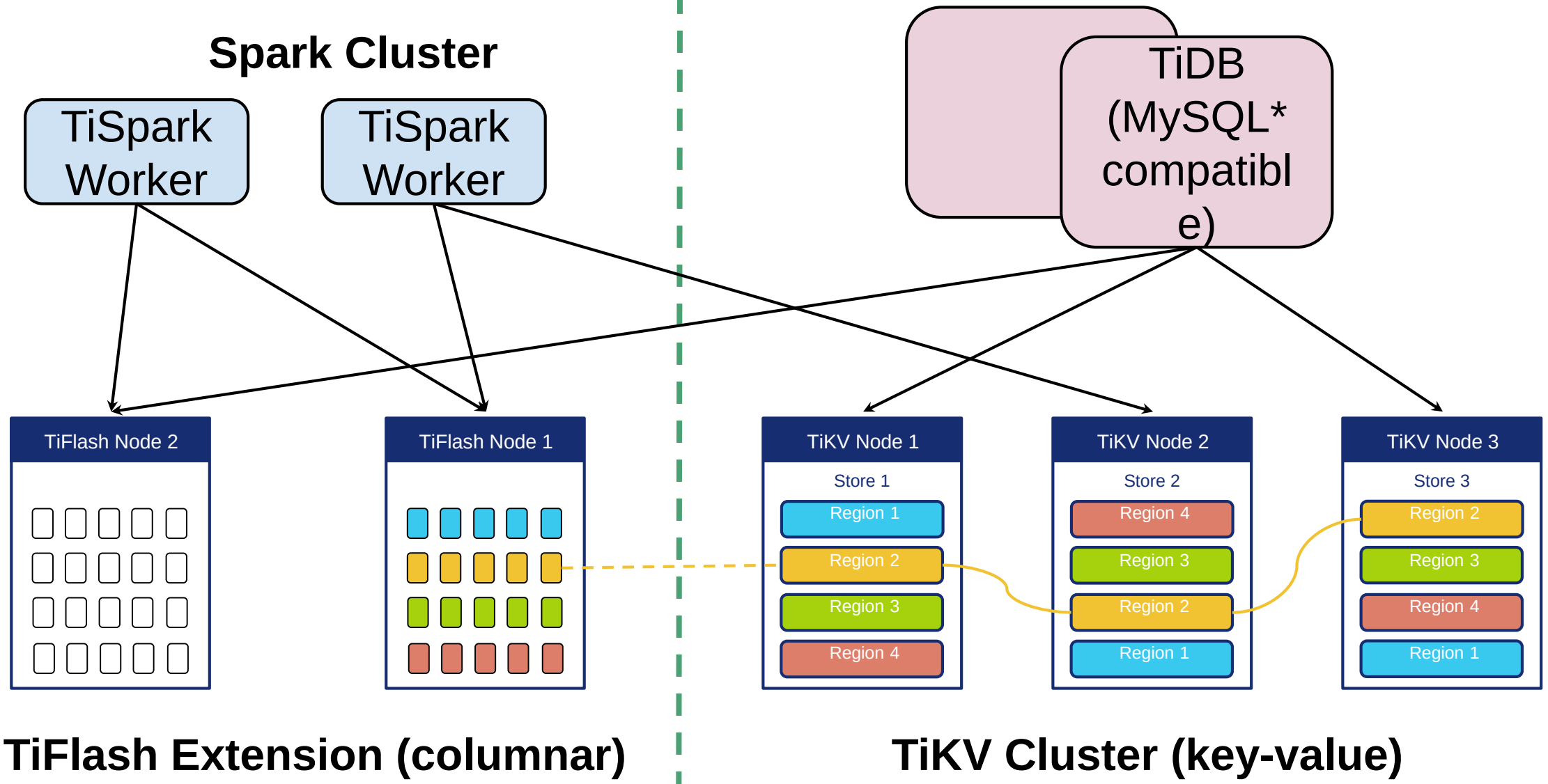
Source: TiDB – used with permission of TiDB *Other names and brands may be claimed as the property of others.

TiKV Cluster

TiKV Architecture

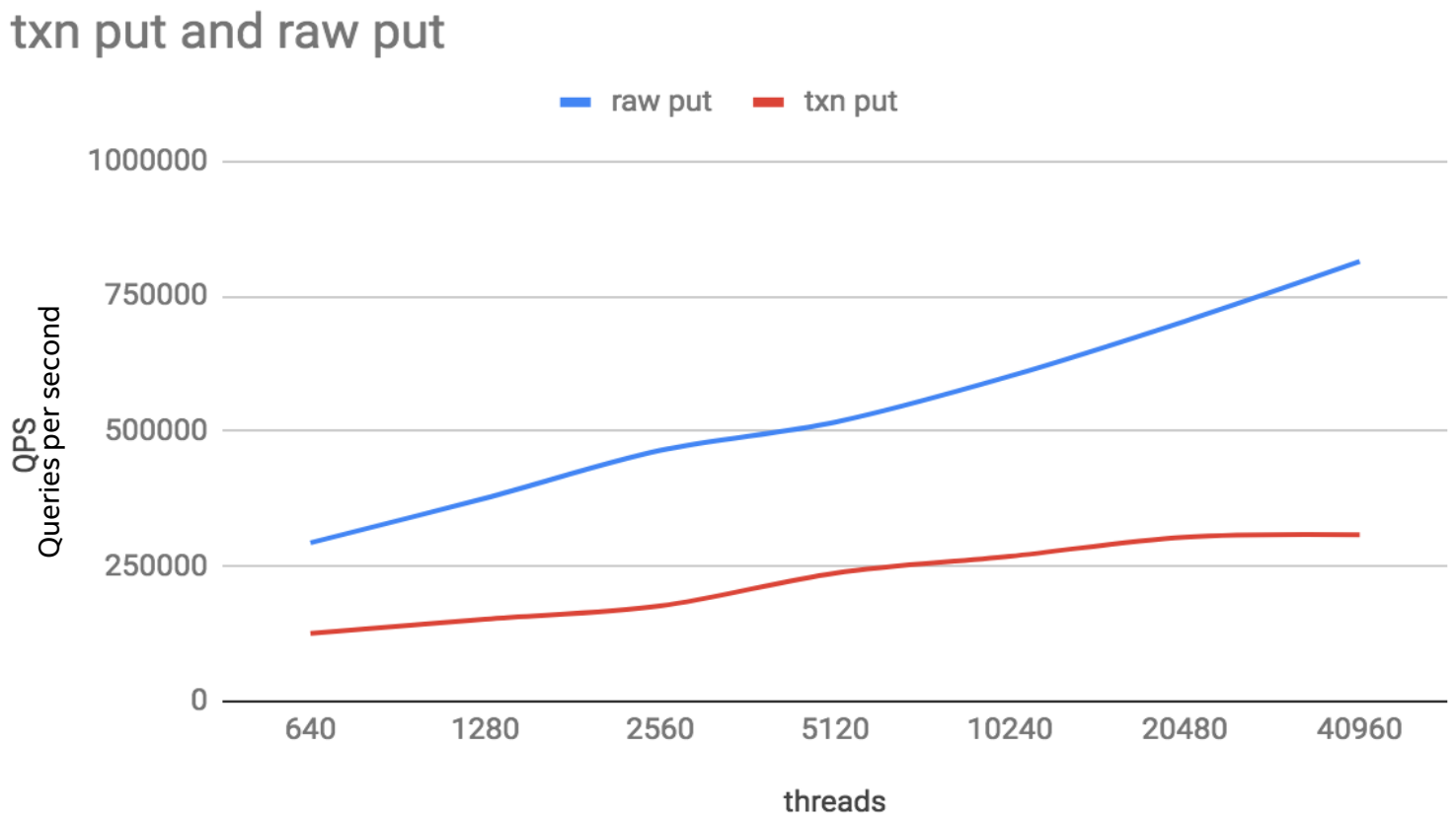
Source: TiDB – used with permission of TiDB. *Other names and brands may be claimed as the property of others.





Source – TiDB – used by permission from TiDB - *Other names and brands may be claimed as the property of others.

Performance of TiKV (Txn Gets and Puts)



Source – TiDB – used by permission from TiDB - *Other names and brands may be claimed as the property of others.

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. 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 visitwww.intel.com/benchmarks.

[1] System configuration: PingCap Sourced: Tested: November 20, 2018. 5 Server Cluster - Server Intel® Server System , 2x Intel® Xeon® Scalable 6142 (64 vcpus per server), BIOS: SE5C620.86B.00.01.0014.070920180847 system product type: S2600WFT 384 GB DDR4 DRAM, database drives- 2x Intel® SSD DC P4800X Series (375 GB) CentOS 7.5 (kernel 4.18.5 (elrepo)), TiDB Server v3.0.0-beta-204-gc0d4632fc, 3 PD Servers in the cluster, 3 DB engines across 5 Nodes (15 DBs)

ACCESS TO INTEL® OPTANE™ SSDs AND OTHER TOOLS

WHAT BENCHMARKS MY TEAM USES…

sysbench -

- https://github.com/akopytov/sysbench

HammerDB - https://www.hammerdb.com/download.html

Redis we focus mostly on using memtier from https://github.com/RedisLabs/memtier_benchmark

What about YCSB ?? Sure, also capable …

Use your own top 10 queries in SQL… and data schema that matches your goals

TOOLS AND GUIDES…

How to efficiently test Intel® Optane™ SSDs (Intel® Optane™ SSD Optimization Guide)

- https://itpeernetwork.intel.com/tuning-performance-intel-optane-ssds-linux-operating-systems/

How to start with SPDK and VHOST (Virtualization with low latency SSDs)

- http://spdk.io/

More info CAS-Linux and Intel® Memory Drive Technology

- http://intel.com/cas and the Open Source version, https://open-cas.github.io/

Persistent Memory Programming - http://pmem.io/

Performance Analysis Tool - https://github.com/intel-hadoop/PAT

Thanks for listening!

frank.ober@intel.com
@fxober
http://intel.com/optane

INTEL® VTUNE AMPLIFIER - PLATFORM PROFILER 2019

Finds

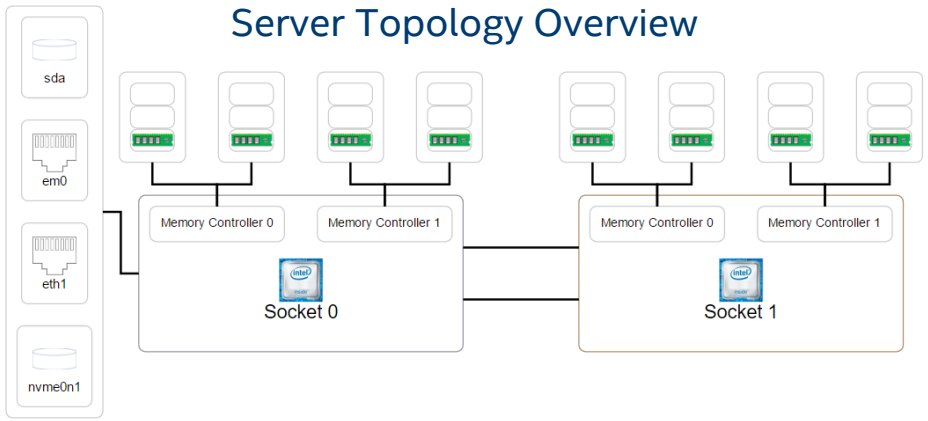
- Configuration issues
- Poorly tuned software

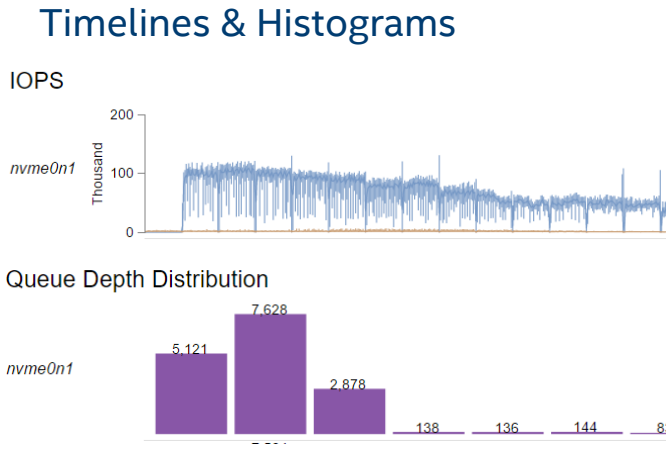
Target Users

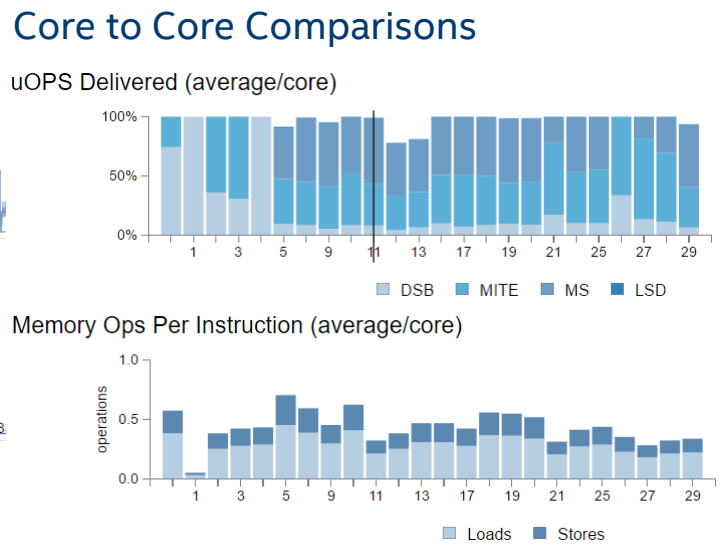
- Infrastructure Architects
- Software Architects & QA

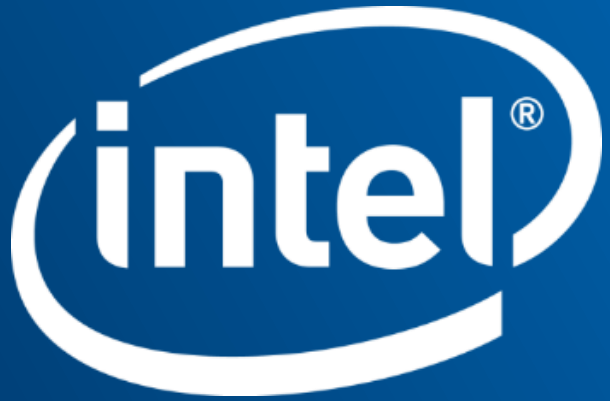
Performance Metrics

- Extended capture (minutes to hours)
- Low overhead – coarse grain metrics
- Sampling OS & hardware performance counters
- RESTful API for easy analysis by scripts








intel