2ND GEN INTEL® XEON® SCALABLE PROCESSORS INTEL® OPTANE™ DC PERSISTENT MEMORY

ENABLED APPLICATIONS MARKETING GUIDE

(V2.0)

Notices & Disclaimers

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer. No computer system can be absolutely secure.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.

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

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.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.
Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Intel, the Intel logo, Atom, Xeon and others are trademarks of Intel Corporation in the U.S. and/or other countries. *Other names and brands may be claimed as the property of others.

© 2019 Intel Corporation.

2ᴺᴰ GEN INTEL® XEON® SCALABLE PROCESSORS – APPS MARKETING GUIDE

- Commercial Software Ecosystem
- Platform Value
- 'Real World' Software Proof Points:
  - Intel® Optane™ DC persistent memory
  - Intel® DL Boost
  - Gen-to-Gen CPU-centric Performance
- Configurations & Disclaimers
- Solution Briefs:
  - Affordable Real-time Computing at Petabyte Scale
  - Accelerating Telecom Business Software
  - Break the Cost and Capacity Barrier
- Video Testimonials:
  - Gigaspaces Drives Data Intensive Use Cases with Intel® Optane™ DC persistent memory
  - Virtuozzo Scales Service Delivery with Intel® Optane™ DC persistent memory

SAMPLE COMMERCIAL SOFTWARE ECOSYSTEM SUPPORT*

(FOR INTEL® XEON® SCALABLE PROCESSORS & INTEL® OPTANE™ DC PERSISTENT MEMORY)

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

4

PLATFORM VALUE PROPOSITION

2ND GENERATION INTEL® XEON® SCALABLE PROCESSOR

INSIGHT PROPELLED (PERFORMANCE)

NEW

- **Enhanced Workload Optimized Performance**
  - **Higher CPU frequency + Improved Turbo profiles** (vs. prior-gen Intel Xeon Scalable processors)
  - Intel® Mesh Architecture
  - Intel® AVX-512
- **Intel® Deep Learning Boost** –
  - embedded AI Inference acceleration

Platform Enhancements

- **Intel® Optane™ DC persistent memory***
- **Intel® Ethernet 800 Series** (**100GbE**)
- Intel® Optane™ DC **SSD (Dual Port)**
- Intel® QLC SSD (**new "ruler" form factors**)
- **Increased DDR4 Memory Speed & Capacity**
  - **Up to 2933MT/s****
  - **16Gb based DIMM** support**

BUSINESS RESILIENCE (HW-ENHANCED SECURITY)[1]

NEW

- **Intel® Security Libraries for Data Center (Intel® SecL)**
  - **Secure** (platform attestation) and **Protect** (data sovereignty) via Intel® Trusted Execution Tech (Intel® TXT)[1]
- **Intel® Threat Detection Technology (TDT)**
- **Intel® Select Solution for Hardened Security**[1] **with Lockheed Martin**

AGILE SERVICE DELIVERY

NEW

- **Intel® Infrastructure Management Technologies (IMT)**
  - **Enhanced** Intel® Resource Director Technology (**Intel® RDT**)
- **Application Device Queues (ADQ)** on Intel® Ethernet 800 Series adapters
- **New workload *specialized* CPUs**
- **Intel® Speed Select Technology (SST)***
  - **SST-PP (Performance Profile)**
  - **SST-BF (Base Frequency)**

* On some SKUs
** on some SKUs, 1 DIMM per channel

[1] No product or component can be absolutely secure.

Features/Frameworks in **BOLD** are new to the 2nd Generation

intel XEON PLATINUM inside™

6
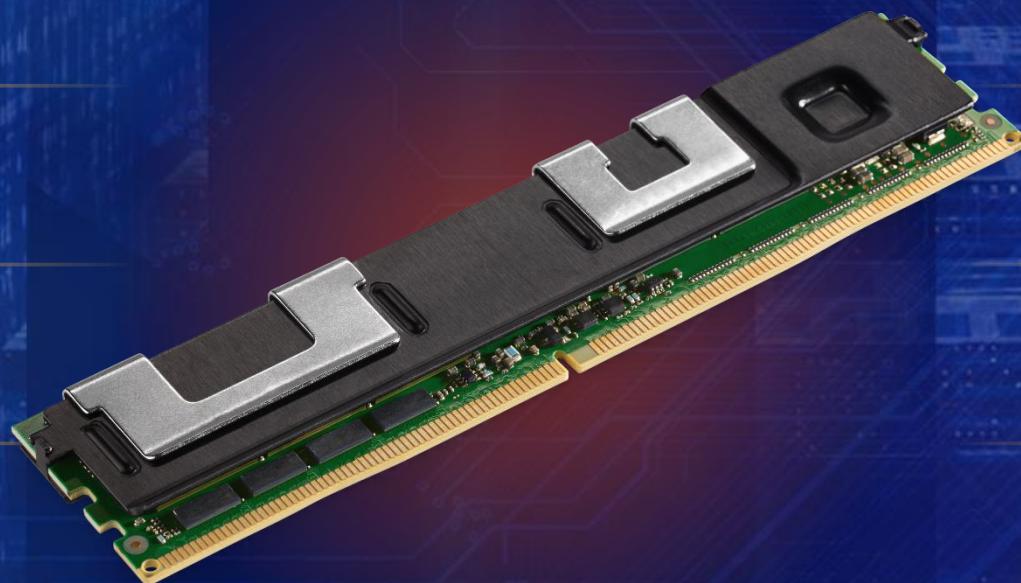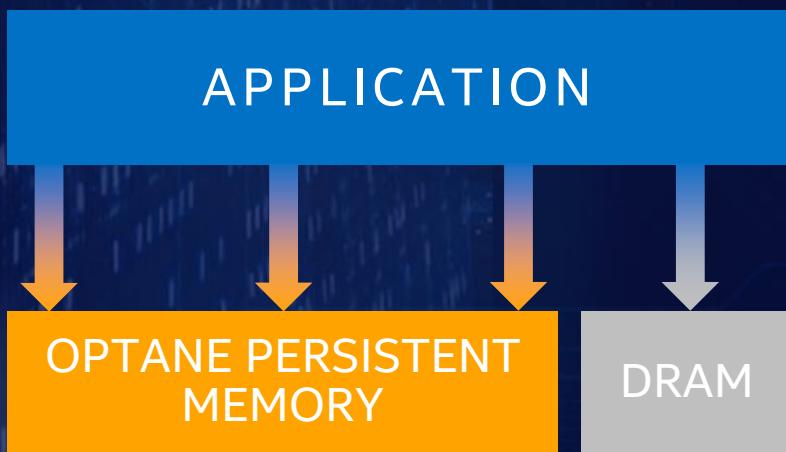
INTEL® OPTANE™ DC PERSISTENT MEMORY: SOFTWARE PROOF POINTS

AEROSPIKE

intel **OPTANE DC** PERSISTENT MEMORY

intel **XEON** PLATINUM inside

http://www.aerospike.com

App Direct Mode

79.09

135X reduction in restart time

0.59

2nd Gen Intel® Xeon® Platinum 8280L

2nd Gen Intel® Xeon® Platinum 8280L + Intel Optane DC persistent memory

Restart Time in Minutes (Lower is Better)

Performance Metric: Restart Time (Minutes)
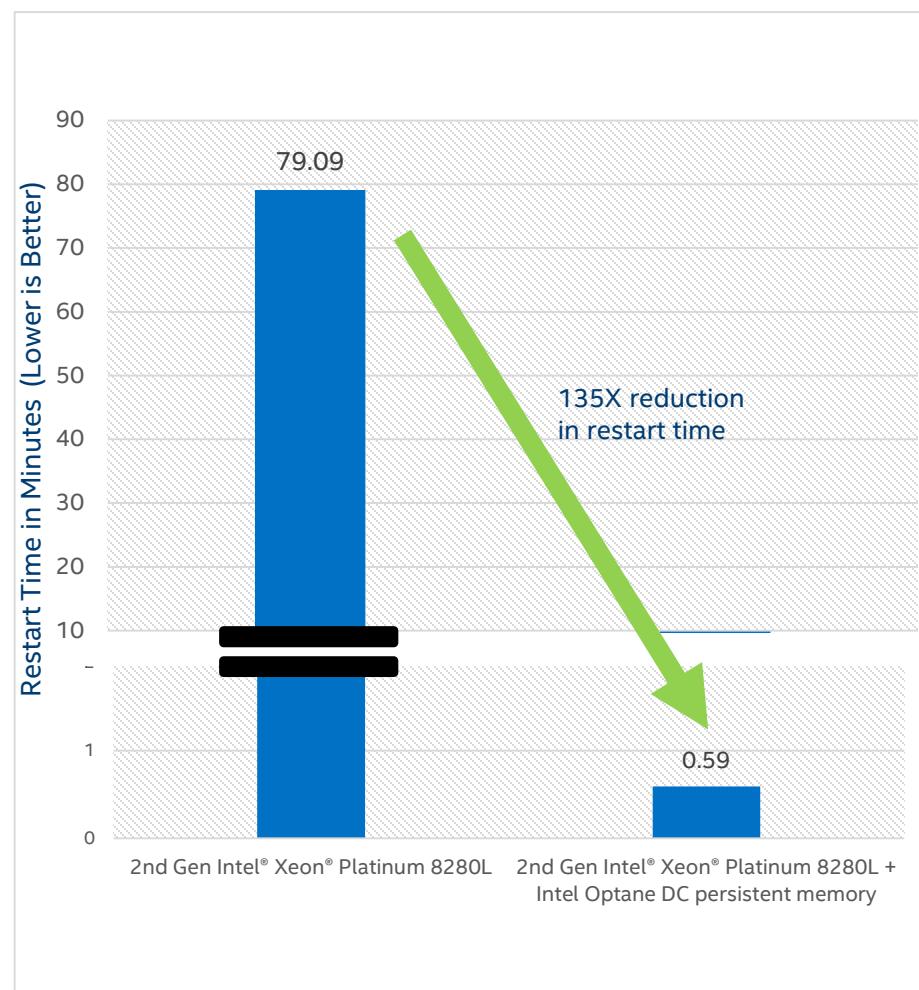
APPLICATION

- Aerospike Enterprise Edition 4.5 is a NoSQL key value database, with indexes typically stored in DRAM and data stored on SSD. Deployments are typically in 2S node clusters. Key customers are in the FSI (Fintech) and Ad Tech segments..

CUSTOMER CHALLENGES

- Each record stored in Aerospike requires a corresponding index entry in DRAM. Customers who run out of DRAM capacity are forced to either purchase larger DIMMs (if available) or scale out to more nodes.

- Any system reboot erases the DRAM index, which must be rebuilt by scanning all records in the database. This process can take hours for large systems, and many customers avoid rebooting unless absolutely necessary.
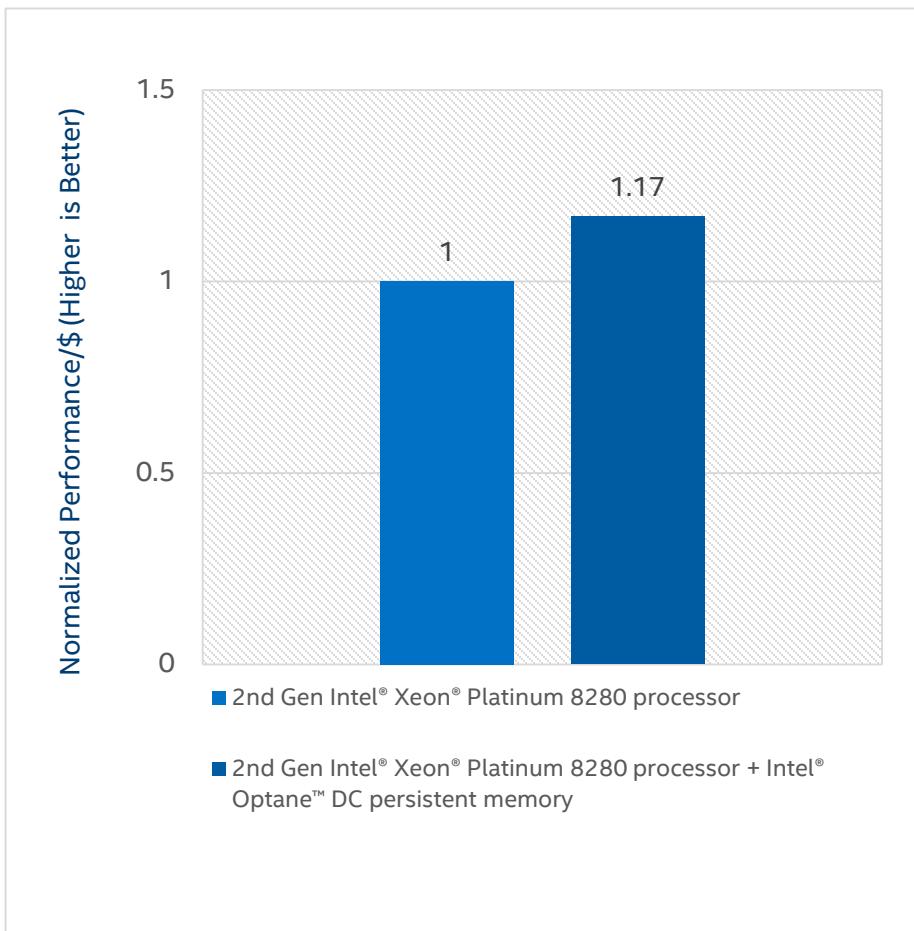
SOLUTION

- Storing the index on large capacity Intel® Optane™ DC persistent memory allows customers to store more data per node, therefore reducing the need to scale out, and at near-DRAM performance.[1]

- "By using a persistence layer for indexes, full restarts of Aerospike are possible without primary index rebuilds." – Brian Bulkowski, Aerospike Co-founder and CTO

VALUE PROPOSITION

- Maintain performance SLAs at lower cost per GB and higher capacity.

- Restart application **within seconds instead of hours** (up to 135X reduction in restart time – see chart)[1] after a planned system reboot., allowing for more frequent software and security updates, while significantly reducing disruption.

ALTIBASE

www.altibase.com

intel OPTANE DC PERSISTENT MEMORY

intel XEON PLATINUM inside

Memory Mode

NEW

Performance Metric: Transactions Per Second, per $TCO (i.e. Perf/$TCO)[2]

Normalized Performance/$ (Higher is Better)

- 2nd Gen Intel® Xeon® Platinum 8280 processor
- 2nd Gen Intel® Xeon® Platinum 8280 processor + Intel® Optane™ DC persistent memory

APPLICATION

Altibase 7.1* is an in-memory Relational Database Management System that provides fast data processing speeds for online transactional processing and online analytical processing. The workload measures transactions per second (TPS) using SQLCLI mode based TCP.

CUSTOMER PAIN POINTS

Customers are currently limited by memory capacity, which requires them having to scale out and have a larger hardware footprint. Scaling up to increase memory capacity is often cost-prohibitive.
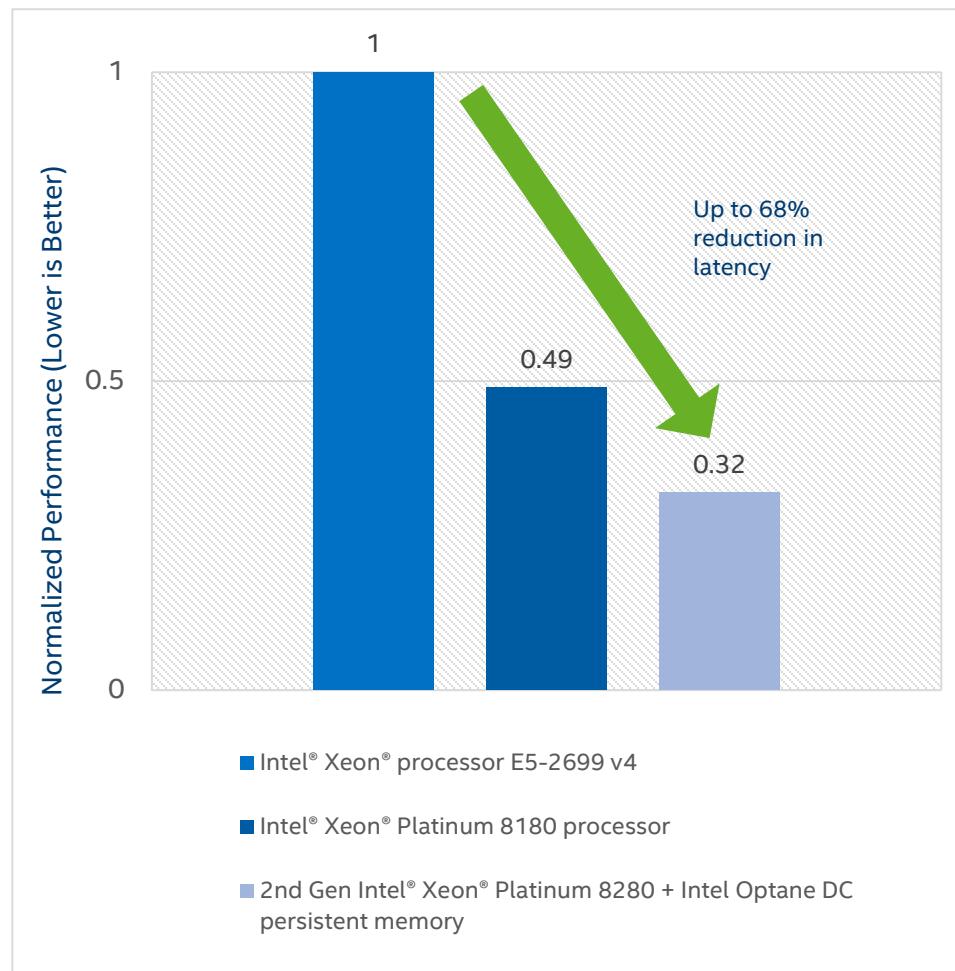
SOLUTION

- With 2nd generation Intel® Xeon® Scalable processors and Intel® Optane™ DC persistent memory (Memory Mode), Altibase can take advantage of the larger memory capacities per socket, while making it a more cost-effective solution at the same time.
- Customers can now keep more data for transaction processing closer to the CPU, with negligible performance impact[1] **at a reduced cost** (see chart showing up to 17% performance improvement for a given cost).[2]

VALUE PROPOSITION

Better performance at similar cost - Alitbase customers can benefit from being able to process more transactions per second, for a given cost [1,2], and while meeting SLA targets for the response time.

ASIAINFO (TELCO BUSINESS SUPPORT SYSTEM)

http://www.asiainfo.com.cn

Chart

Normalized Performance (Lower is Better)

- 1 — (Intel® Xeon® processor E5-2699 v4)
- 0.49 — (Intel® Xeon® Platinum 8180 processor)
- 0.32 — (2nd Gen Intel® Xeon® Platinum 8280 + Intel Optane DC persistent memory)

Up to 68% reduction in latency

Legend:
- ■ Intel® Xeon® processor E5-2699 v4
- ■ Intel® Xeon® Platinum 8180 processor
- ■ 2nd Gen Intel® Xeon® Platinum 8280 + Intel Optane DC persistent memory

Performance Metric: Query Response Time (Latency)

APPLICATION

AsiaInfo is the largest BSS (Business Support System) provider in PRC, and its Telco BSS is a fundamental application for telecom carriers. The benchmark is developed by AsiaInfo as a proxy to simulate a commonly occurring business analytics scenario. In this scenario, where 9 queries are executed on a mock dataset. Key customers include China Mobile.

CUSTOMER CHALLENGES

Complex queries take a long time to execute due to amount of data processing and inability to store large volumes closer to the CPU, due to limited, affordable memory capacity.

SOLUTION

Using Intel® Optane™ DC persistent memory (App Direct (Storage over App Direct) mode), improves performance by allowing customers to store more data in memory (reduce spillover to SSDs) **and** by reducing frequent accesses to disk (see chart for normalized reduction in latency, up to 68%, compared to older generation)[1].

VALUE PROPOSITION

Better performance at similar cost - AsiaInfo's telecom customers will benefit from improved performance (i.e., lower latency/response time when executing queries) at comparable (ISO) cost. They can process more data in less time.

http://en.baosight.com/

App Direct Mode
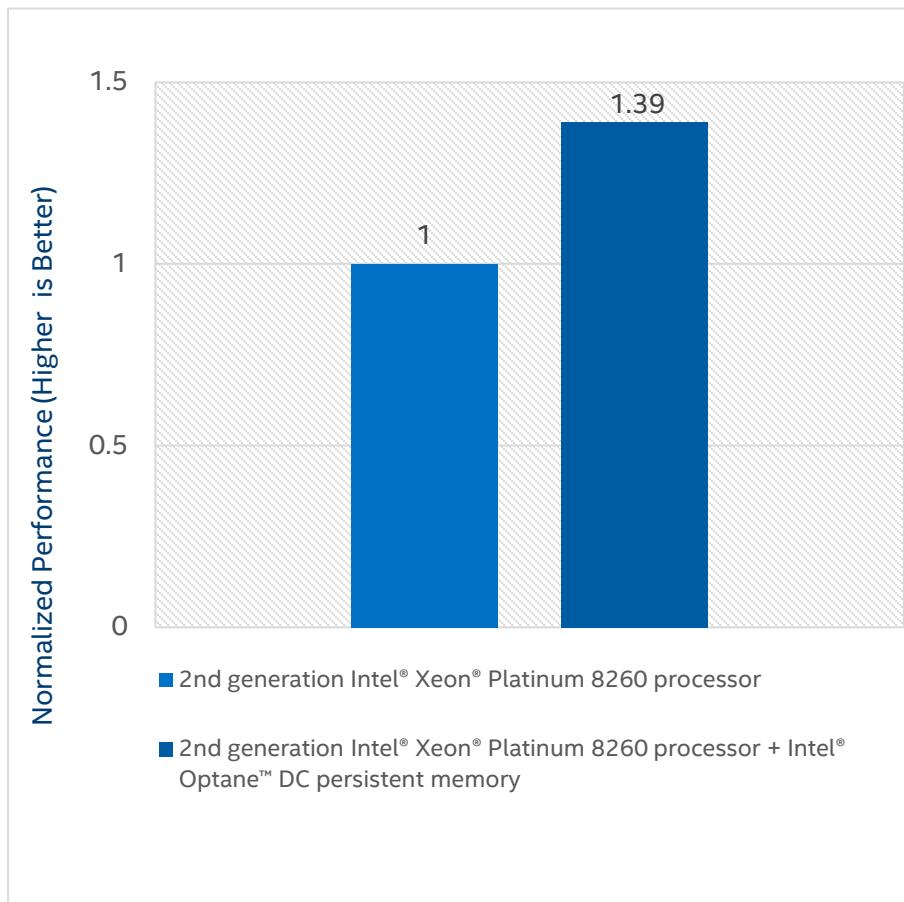
Performance Metric: Throughput (Transactions Per Second), While Maintaining SLA of a Response Time Less Than 10ms.

Legend:
- 2nd generation Intel® Xeon® Platinum 8260 processor
- 2nd generation Intel® Xeon® Platinum 8260 processor + Intel® Optane™ DC persistent memory

Y-axis: Normalized Performance (Higher is Better)
Bar values: 1, 1.39

APPLICATION

Baosight xInsight is an industrial big data platform which is deployed in manufacturing, transportation and government fields. It is used for storage, query and analysis of big data. Major customers include Bao Steel and Chongqing (Shanghai) Metro.

CUSTOMER CHALLENGES

Customers experience an I/O bottleneck due to limited main memory today, which underutilizes the CPU and limits throughput that can be achieved. Adding additional memory is often cost prohibitive, especially for large deployments.
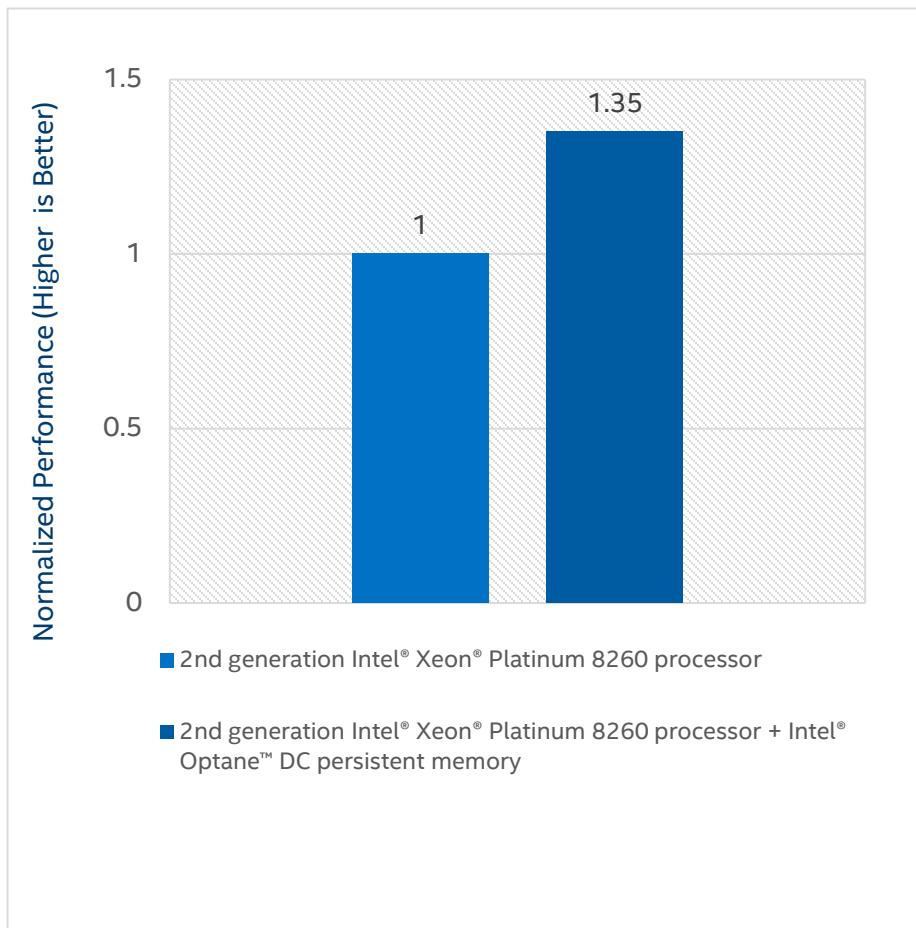
SOLUTION

- Baosight xInsight application is limited by available memory, resulting in an I/O bottleneck. Intel® Optane™ DC persistent memory (App Direct mode) delivers significantly larger, affordable memory capacities, as well as the flexibility to allocate data structures to specific memory tiers.
- This allows customers to have larger datasets in memory, therefore **delivering better performance with more transactions per second** (see chart for performance improvement)[1].

VALUE PROPOSITION

Better performance at similar cost – Baosight customers will benefit from improved performance (i.e., being able to run more transactions per second), **at comparable cost, and while meeting SLA target** of a response time less than 10ms.

1 - Performance results are based on testing by Intel and Baosight on 01/08/19 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.

GBASE (8M DATABASE)*

http://www.gbase.cn

Normalized Performance (Higher is Better)

- 1
- 1.35

■ 2nd generation Intel® Xeon® Platinum 8260 processor

■ 2nd generation Intel® Xeon® Platinum 8260 processor + Intel® Optane™ DC persistent memory

Performance Metric: Throughput (Transactions Per Second), While Maintaining ISV-specified QoS/SLAs

APPLICATION

GBase 8m is an in memory database targeted for high throughput and low latency OnLine Transaction Processing (OLTP) scenarios. This test workload is developed by GBase – it simulates an Online Charging System (OCS) that is used in telecom applications for performance benchmarks.

CUSTOMER CHALLENGES

Customers experience a capacity bottleneck due to limited, affordable main memory options available today. Since memory needs to be allocated for each online account, GBase is not able to store the desired number of accounts in-memory (and is therefore capped on throughput). Adding additional memory requires scaling out, which is often cost-prohibitive.
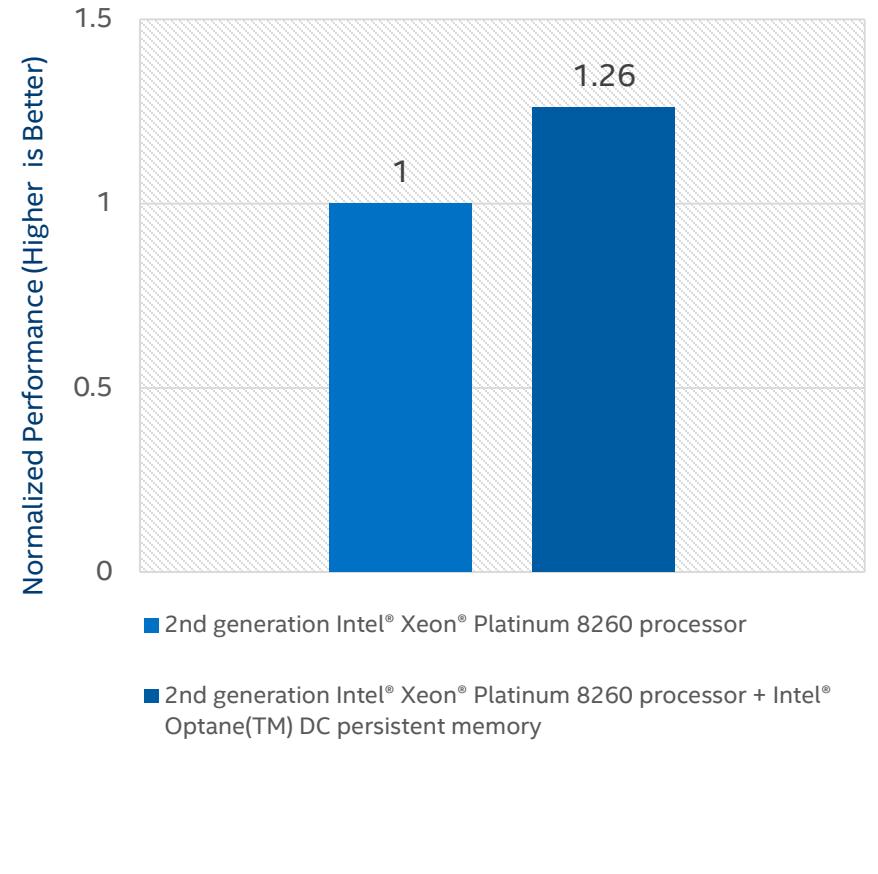
SOLUTION

- Intel® Optane™ DC persistent memory (Memory Mode) delivers significantly larger, affordable memory capacities, in a single system, which reduces the need to scale out and therefore reduces cost.
- With more memory, greater number of online accounts can be stored in memory, thereby enabling the system to process greater throughput i.e., **more transactions per second** (see chart for performance improvement)[1].

VALUE PROPOSITION

- **Better performance at similar cost** - GBase customers will benefit from being able to store more online accounts per server, thereby reducing hardware footprint. In addition, throughput is also improved **at comparable cost, and while meeting SLA targets for the response time**.

1 - Performance results are based on testing by Intel and GBase on 12/28/18 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.

http://www.hisign.com.cn

Memory Mode

Performance Metric: Queries Per Second (No. of Fingerprints Processed)

APPLICATION

Hisign is one of the key biometric companies in China. They provide biometric authentication solutions for Foreign Affairs and Public Security. Hisign's fingerprint verification service, MCH, is a cloud service – a fingerprint dataset is loaded into memory, compared to the target fingerprint sent by the clients and a correlation returned to the backend component of the system. The database is usually split into multiple data sets for multi-instance processing.

CUSTOMER CHALLENGES

Hisign application is memory intensive – expanding memory capacity to alleviate this bottleneck can often be cost-prohibitive for customers.
Higher latencies can impact customers' user experience and productivity.
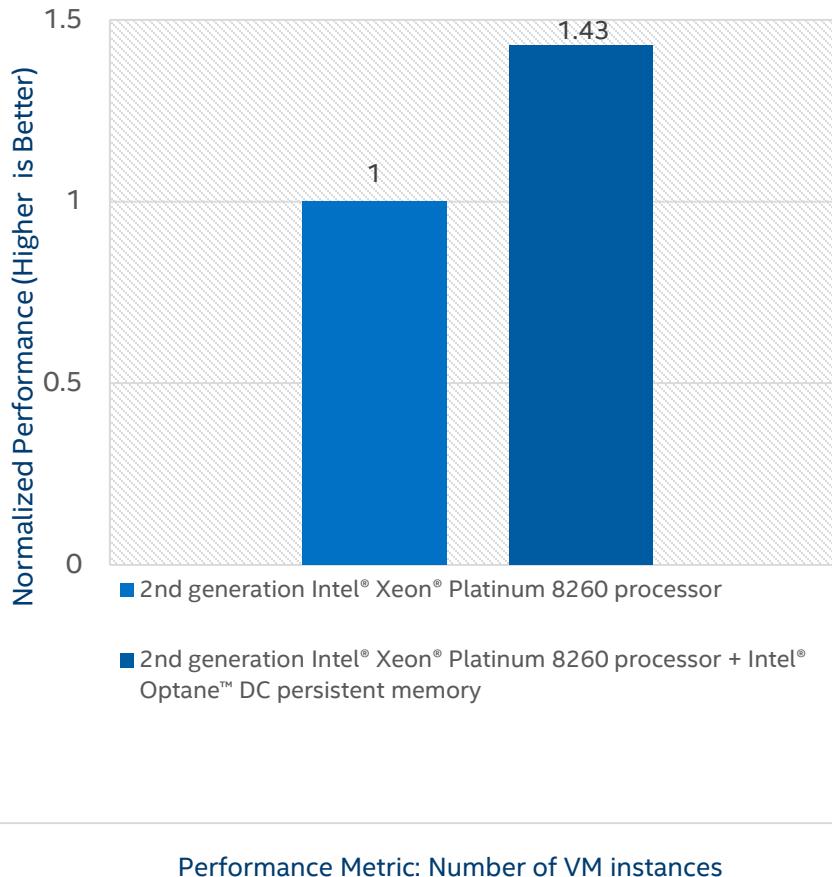
SOLUTION

- With Intel® Optane™ DC persistent memory (Memory Mode), more fingerprint datasets are stored in memory for a similar cost. Performance is improved due to being able to **run more tasks in parallel**, **while maintaining the same latency SLA** (see chart for performance improvement)[1].
- Higher memory bandwidth and larger memory capacity, compared to the older generation baseline, contributed to the performance.

VALUE PROPOSITION

Hisign fingerprint verification service customers will benefit from improved performance (i.e., being able to **run more tasks in parallel at comparable cost, and without impacting latency**. They can process more data in the same time, delivering a better user experience.

Memory Mode

https://e.huawei.com/en/

Normalized Performance (Higher is Better)

1.5

1.43

1

0.5

0

1

■ 2nd generation Intel® Xeon® Platinum 8260 processor

■ 2nd generation Intel® Xeon® Platinum 8260 processor + Intel® Optane™ DC persistent memory

Performance Metric: Number of VM instances

APPLICATION

FusionSphere* of OpenStack* version is Huawei's commercial release with a built-in HUAWEI Kernel-based Virtual Machine (KVM) virtualization engine based. The workload uses the SysBench tool to benchmark MySQL* performance on FusionSphere; it is I/O and Memory bound, but not CPU bound. The software is used in Enterprise private cloud, carrier NFV and public CSP environments.

CUSTOMER CHALLENGES

- Memory and I/O bound applications running FusionSphere limit the number of virtual machines that can be instantiated.
- Being able to support more VM instances on the same hardware drives better utilization of hardware resources, transparent to the user.

SOLUTION

The MySQL with SysBench benchmark running on HUAWEI FusionSphere is I/O and memory bound. Higher memory capacity offered by Intel® Optane™ DC persistent memory delivers improved performance by being **able to support more VM instances** (see chart for performance improvement)[1].

VALUE PROPOSITION

Better performance at similar cost - Huawei will benefit from improved performance, i.e., being able to have more VM instances per server, at comparable cost, and while meeting SLA targets for both latency and throughput (transactions per second). This also translates to better, more efficient hardware utilization.

www.kingbase.com.cn

App Direct

Performance Metric: Query Response Time, per $TCO (i.e. perf/$TCO)[2]
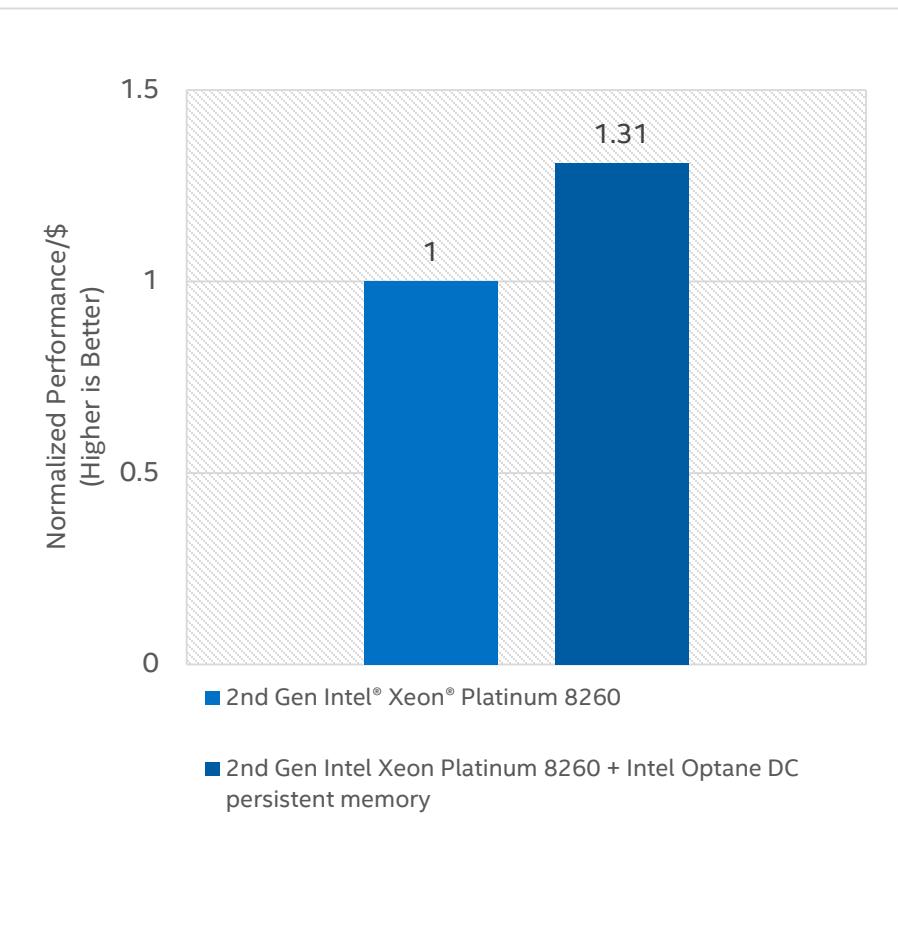
APPLICATION

Kingbase Analytics Database (KADB) is a distributed OnLine Analytics Processing (OLAP) database. The workload is specified by Kingbase to simulate a critical business scenario that needs near-real time processing, and runs concurrent SQL queries on a 1.5TB dataset.

CUSTOMER CHALLENGES

The queries in KADB are usually I/O-bound. When customers desire high performance for near-realtime processing, the best way is to cache all the needed data in memory to reduce IO. However, this approach is currently limited by memory capacity, and can be cost-prohibitive for large datasets.
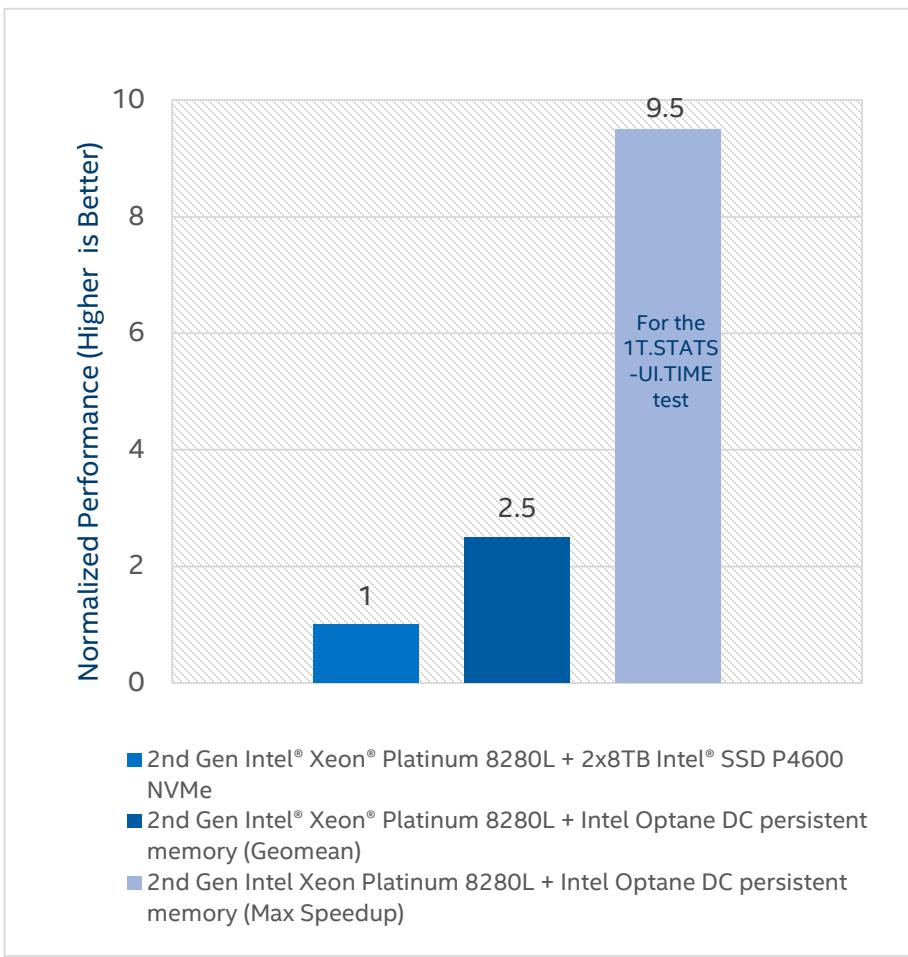
SOLUTION

With Intel® Optane™ DC persistent memory (App Direct Mode), Kingbase can take advantage of the larger memory capacities per socket to store data closer to the CPU and eliminate the I/O bottleneck, while achieving better price-performance (see chart)[1,2].

VALUE PROPOSITION

Improved price-performance: Kingbase customers will benefit from similar performance (being able to process almost the same number of transactions per second), **at lower cost[2], and while meeting SLA targets for the response time**.

http://www.kx.com

intel OPTANE DC PERSISTENT MEMORY

intel XEON PLATINUM inside

App Direct

Chart — Normalized Performance (Higher is Better):

- 2nd Gen Intel® Xeon® Platinum 8280L + 2x8TB Intel® SSD P4600 NVMe: **1**
- 2nd Gen Intel® Xeon® Platinum 8280L + Intel Optane DC persistent memory (Geomean): **2.5**
- 2nd Gen Intel Xeon Platinum 8280L + Intel Optane DC persistent memory (Max Speedup): **9.5** (For the 1T.STATS-UI.TIME test)

Performance Metric: Query Latency Speedup (STAC Antuco benchmark)

APPLICATION

KDB+ is a time series database built for the Financial Services Industry to handle historical data used by 20 out of 21 top global banks. The STAC Antuco benchmark tests multiple aspects of database and HW performance on a typical customer workload. It is an IO-stressing benchmark with various types of access (random-like, sequential like, with overlap, etc.)

CUSTOMER CHALLENGES

Quick access to historical data is absolutely critical to electronic trading to get a competitive advantage. The dataset size is often too large to be stored in DRAM. Adding additional memory is often cost prohibitive, resulting in the need to scale out and increasing HW footprint.
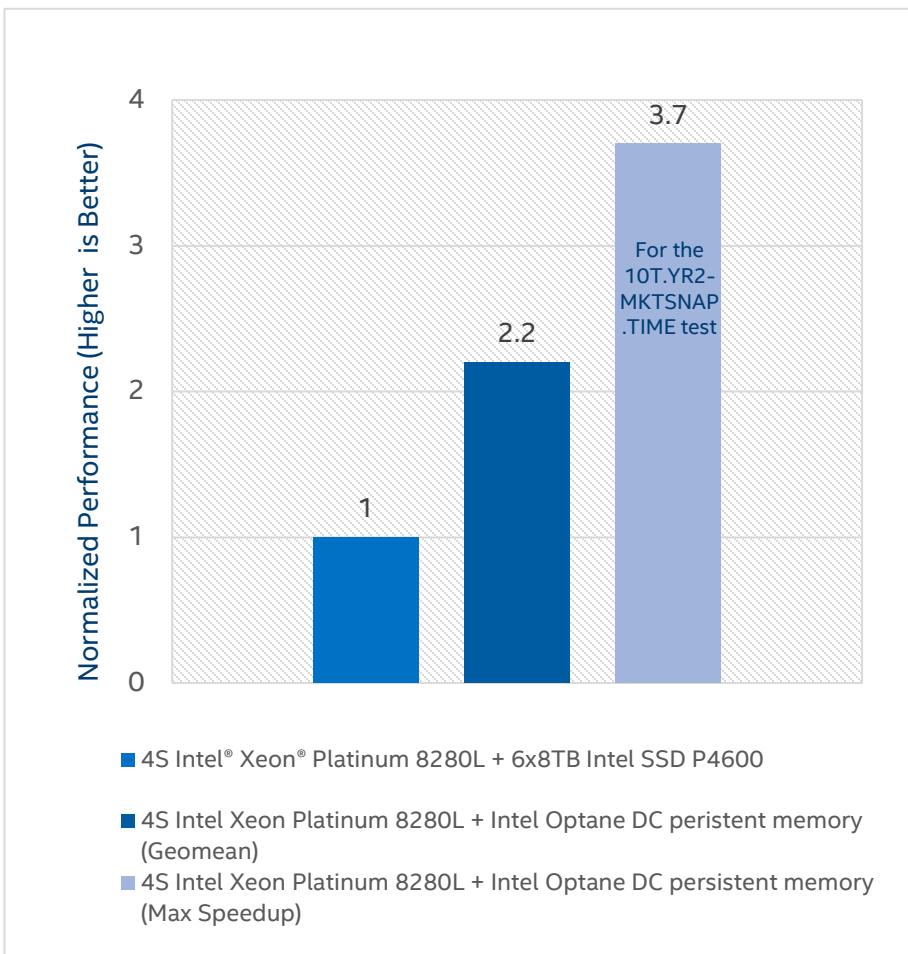
SOLUTION

Higher memory capacities of Intel® Optane™ DC persistent memory (App Direct Mode (Storage over AD)) facilitates storing larger datasets and more historical data closer to the CPU, delivering significant performance improvements (see chart)[1], compared to using SSDs (current solution).

VALUE PROPOSITION

Significantly better performance – Kx customers will benefit from being able to access historical data much faster than currently available solutions.

intel

KX SYSTEMS (KDB+ 3.6)*

App Direct

Normalized Performance (Higher is Better)

- 1
- 2.2
- 3.7 — For the 10T.YR2-MKTSNAP.TIME test

■ 4S Intel® Xeon® Platinum 8280L + 6x8TB Intel SSD P4600

■ 4S Intel Xeon Platinum 8280L + Intel Optane DC peristent memory (Geomean)

■ 4S Intel Xeon Platinum 8280L + Intel Optane DC persistent memory (Max Speedup)

Performance Metric: Query Latency Speedup (STAC Kanaga benchmark)

APPLICATION
KDB+ is a time series database built for the Financial Services Industry to handle historical data used by 20 out of 21 top global banks. The STAC Kanaga benchmark tests multiple aspects of database and HW performance on a typical customer workload. It is an IO-stressing benchmark with various types of access (random-like, sequential like, with overlap, etc.)

CUSTOMER CHALLENGES
Quick access to historical data is absolutely critical to electronic trading to get a competitive advantage. The dataset size is often too large to be stored in DRAM. Adding additional memory is often cost prohibitive, resulting in the need to scale out and increasing HW footprint.

SOLUTION
Higher memory capacities of Intel® Optane™ DC persistent memory (App Direct Mode (Storage over AD)) facilitates storing larger datasets and more historical data closer to the CPU, delivering significant performance improvements (see chart)[1], compared to using SSDs (current solution).
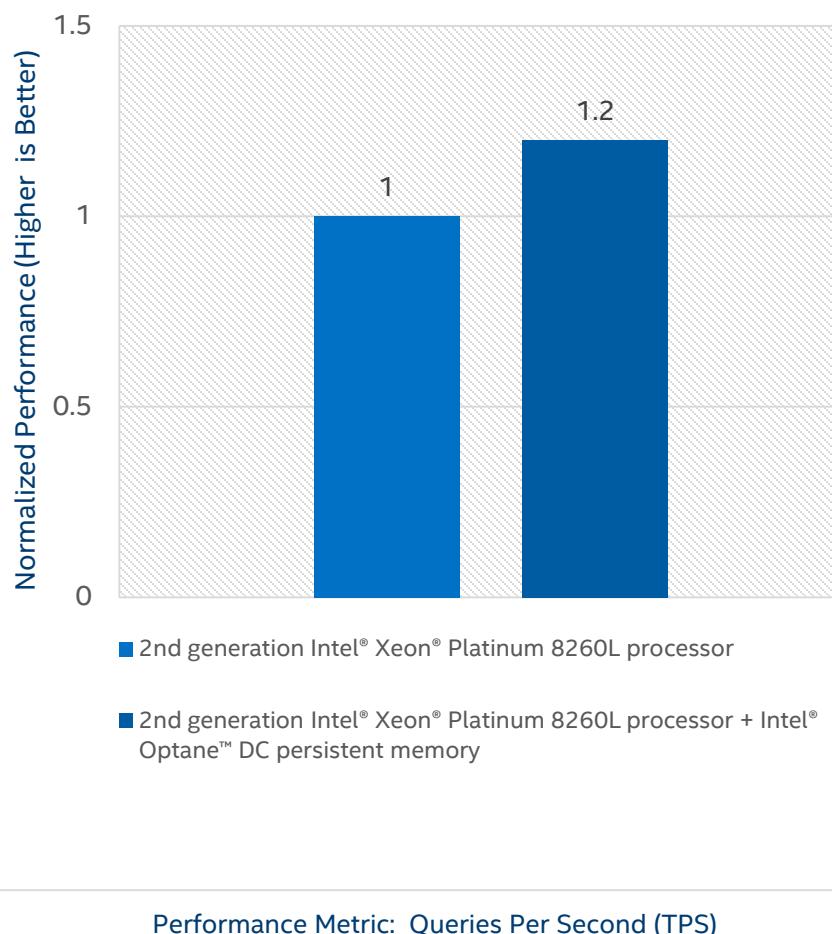
VALUE PROPOSITION
Significantly better performance – Kx customers will benefit from being able to access historical data much faster than currently available solutions.

NARI 南瑞集团 NARI GROUP CORPORATION
信息系统集成分公司 INFORMATION SYSTEM INTEGRATION COMPANY

http://www.sgepri.sgcc.com.cn/

intel OPTANE DC PERSISTENT MEMORY

intel XEON PLATINUM inside

Memory Mode

Normalized Performance (Higher is Better)

1.5

1.2

1

0.5

0

■ 2nd generation Intel® Xeon® Platinum 8260L processor
■ 2nd generation Intel® Xeon® Platinum 8260L processor + Intel® Optane™ DC persistent memory

Performance Metric: Queries Per Second (TPS)

APPLICATION

NARI* is a leading solution provider of power and automation technologies in China, and its Power Consumption Analysis system is a fundamental application for the country's National Grid. Power Consumption Analysis system is used to collect and store data from thousands of geo-distributed sensors, and provides the analysis & dispatching function. The workload is a self defined benchmark to simulate typical customer analysis scenarios.

CUSTOMER CHALLENGES

- Best performance with NARI Power Consumption Analysis application is achieved when all data tables are loaded into memory for analysis. Limited memory capacity today results in some tables being compressed and then having to be decompressed in serial fashion, which limits achievable performance. Expanding memory capacity to alleviate this bottleneck can often be cost-prohibitive for customers.
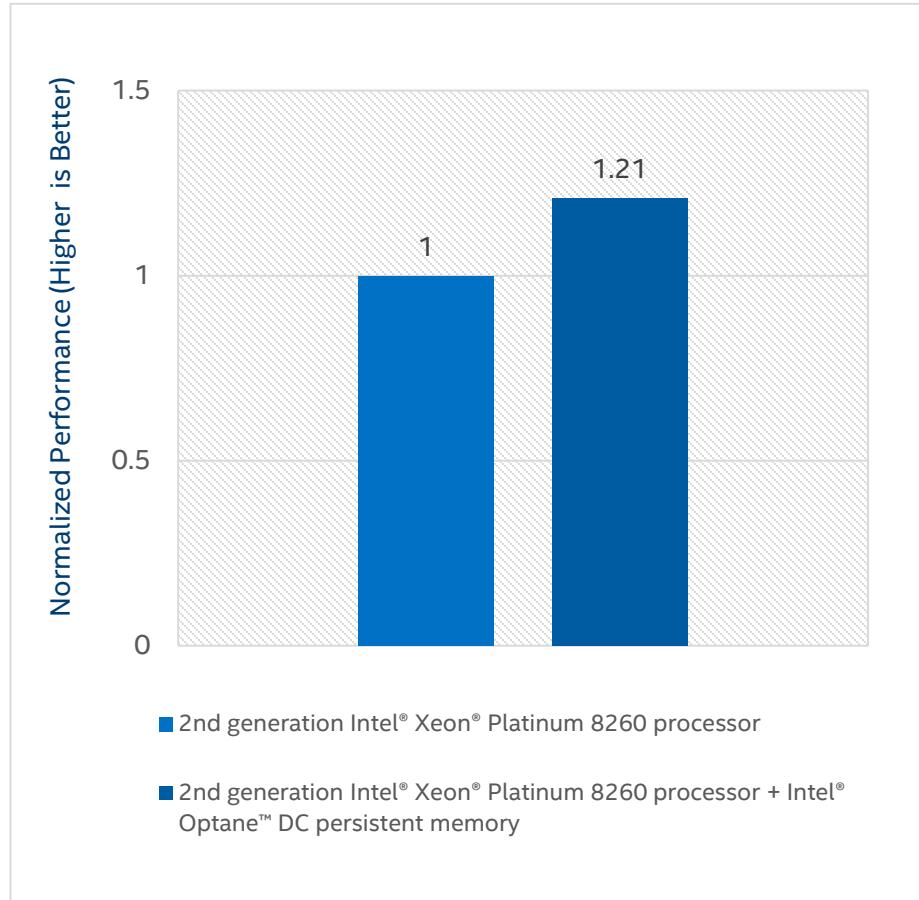- Not being able to process as many queries per second (QPS) as desired.

SOLUTION

Higher memory capacity of Intel® Optane™ DC persistent memory (Memory Mode) helps reduce the in-memory table compression ratio for NARI Power Consumption Analysis . This allows customers to maintain large amounts of data in an uncompressed format, thereby alleviating the added step of decompression before analyzing the data. As a result, it **delivers better performance by processing more queries per second** (see chart for performance improvement)[1].

VALUE PROPOSITION

Better performance at similar cost – NARI customers will benefit from being able to process up to 20% more queries per second, at comparable cost.

1 - Performance results are based on testing by Intel and Nari on 12/28/18 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.

Neusoft®
http://www.neusoft.com

intel OPTANE™ DC
PERSISTENT MEMORY

intel XEON PLATINUM inside™

Memory Mode

Normalized Performance (Higher is Better)

Bar chart values: 1, 1.21

Legend:
- 2nd generation Intel® Xeon® Platinum 8260 processor
- 2nd generation Intel® Xeon® Platinum 8260 processor + Intel® Optane™ DC persistent memory

Performance Metric: Transactions Per Second (TPS)

APPLICATION

Neusoft Aclome* Cloud is a visual, automatic and agile cloud management environment that enables customers to deploy, monitor and manage cloud applications flexibly and conveniently. Aclome DB*, used for this self-defined workload, is a core component of the Aclome package.

CUSTOMER CHALLENGES

Customers experience an I/O bottleneck due to limited cache, which underutilizes the CPU and limits throughput that can be achieved. Adding additional memory is often cost prohibitive, especially for large deployments.
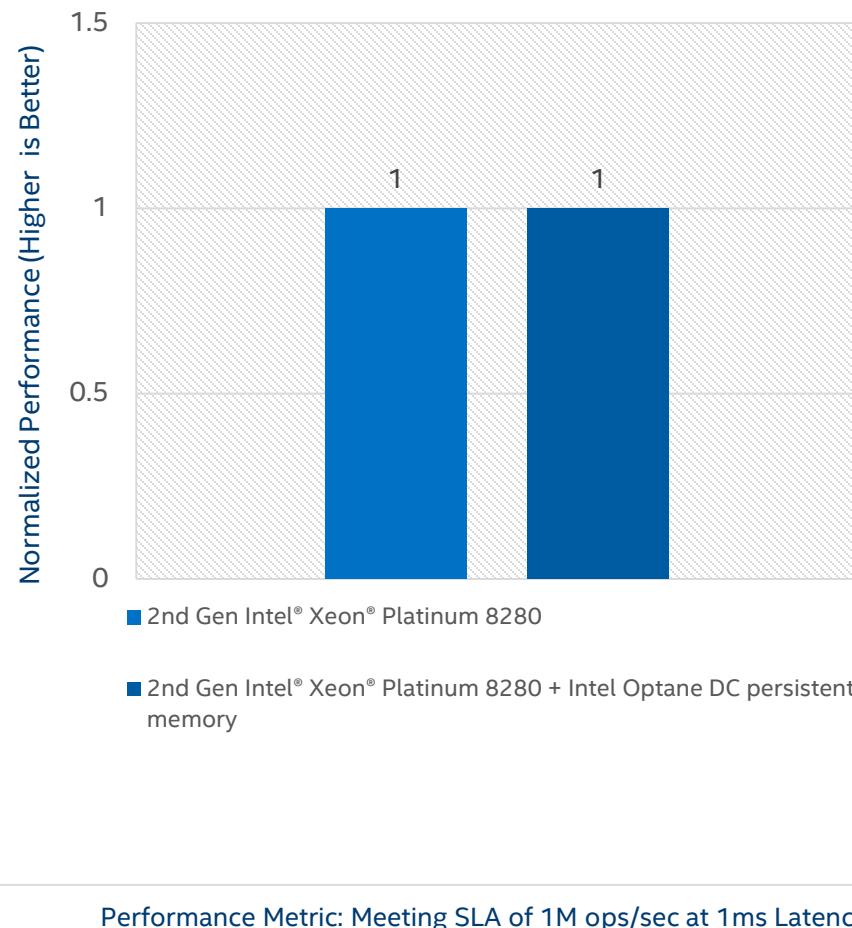
SOLUTION

- Neusoft Aclome DB application is limited by memory capacity, resulting in an I/O bottleneck. Higher memory capacities of Intel® Optane™ DC persistent memory (Memory Mode) can provide a larger caching layer..
- This allows customers to have larger datasets in memory, therefore **delivering better performance with more transactions per second** (see chart showing up to 21% performance improvement)[1].

VALUE PROPOSITION

- **Better performance at similar cost** - Neusoft customers will benefit from improved performance (i.e., being able to run more transactions per second), **at comparable cost, and while meeting SLA targets for the response time.**

Memory Mode

Normalized Performance (Higher is Better)

- 2nd Gen Intel® Xeon® Platinum 8280
- 2nd Gen Intel® Xeon® Platinum 8280 + Intel Optane DC persistent memory

Performance Metric: Meeting SLA of 1M ops/sec at 1ms Latency

APPLICATION

Redis Enterprise* is the commercial offering of Redis (an in-memory database and one of the most popular NoSQL databases The workload is called "memtier" and it evaluates the ability of a server (in terms of throughput and latency) to service get/put requests.

CUSTOMER CHALLENGES

- Complex, multi-tiered solutions (with hot data on DRAM and warm data on SSDs).
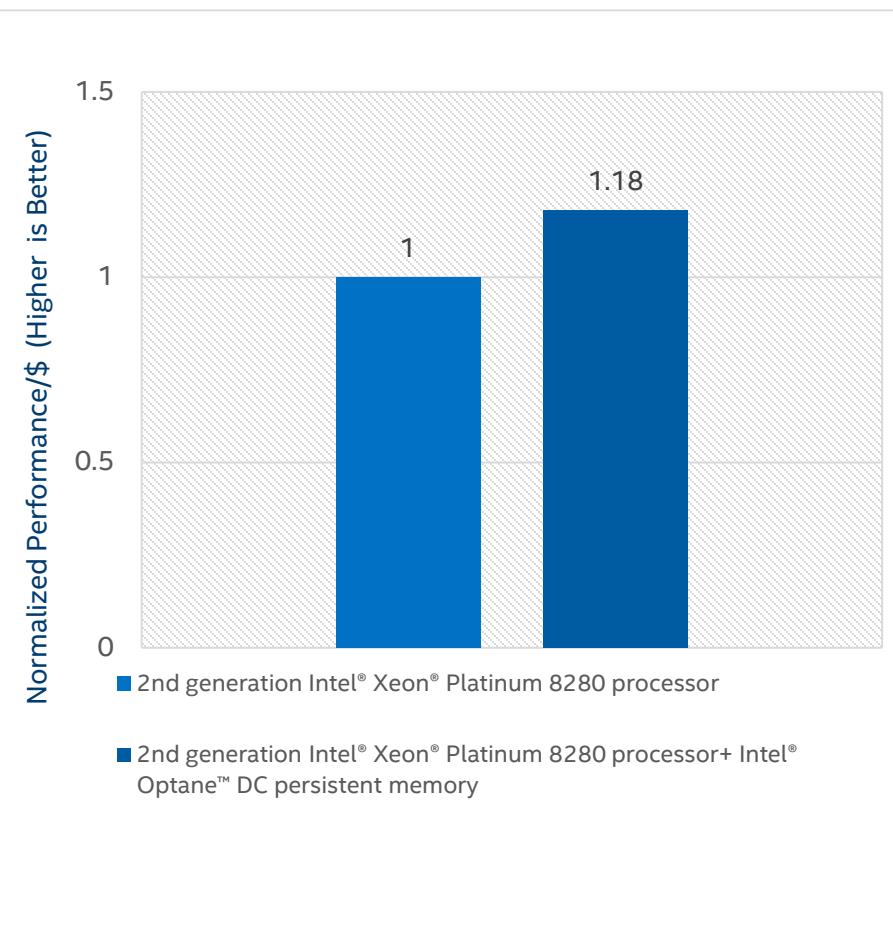- Scaling to very large data-sets, with high performance in a cost effective way.

SOLUTION

- With larger available memory capacities, Intel® Optane™ DC persistent memory allows the entire dataset to be deployed in a **single** in-memory database (vs. splitting data sets across servers), thereby reducing costs and hardware footprint.
- As shown in the performance chart, the SLA of driving 1M ops/second at sub-millisecond response times is able to be met with this configuration.

VALUE PROPOSITION

- Maintain typical customer SLAs (1M ops/sec @ 1ms latency) at reduced hardware costs.
- Accommodate larger capacity deployments with fewer servers, reduced hardware and operational costs, while maintaining required throughput and latency metrics.

SAS (VIYA 3.4)*

intel OPTANE DC PERSISTENT MEMORY

intel XEON PLATINUM inside

Memory Mode

Normalized Performance/$ (Higher is Better)

1.18

1

- 2nd generation Intel® Xeon® Platinum 8280 processor
- 2nd generation Intel® Xeon® Platinum 8280 processor+ Intel® Optane™ DC persistent memory

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.
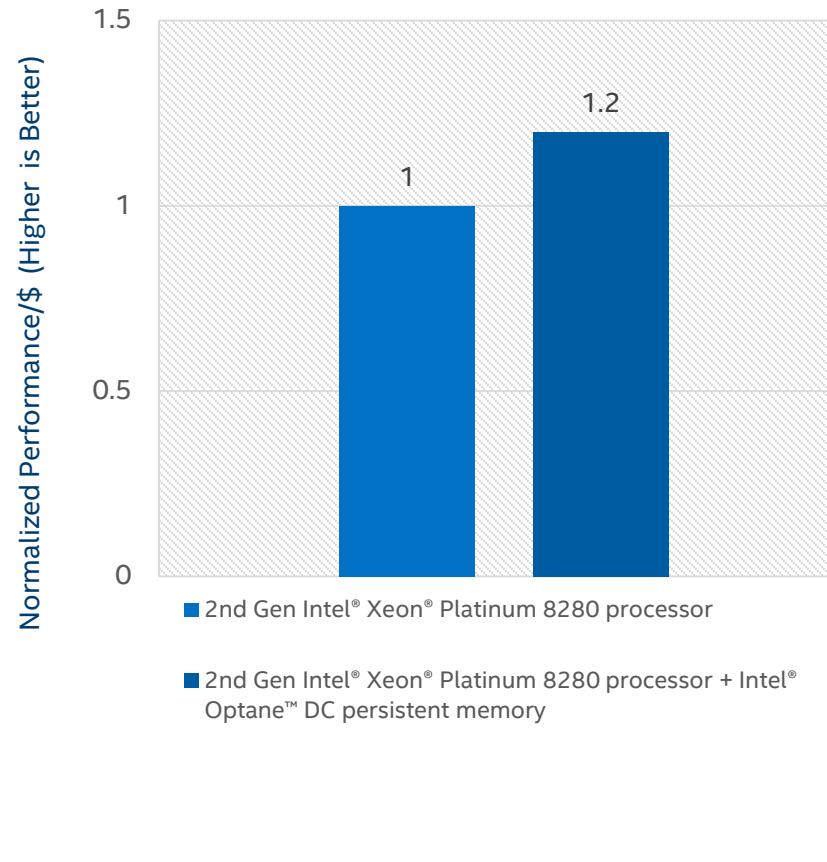
SOLUTION

- 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**.

intel OPTANE DC PERSISTENT MEMORY

intel XEON PLATINUM inside

Memory Mode

NEW

Normalized Performance/$ (Higher is Better)

- 2nd Gen Intel® Xeon® Platinum 8280 processor
- 2nd Gen Intel® Xeon® Platinum 8280 processor + Intel® Optane™ DC persistent memory

Performance Metric: Transactions Per Second (TPS), per $TCO (i.e. Perf/$TCO)[2]

APPLICATION

Sunjesoft Golilock v3.2.0* is an in-memory database with high scalability and low latency performance. The workload measures transactions per second (TPS) with multi-client Direct Access (DA) node.

CUSTOMER PAIN POINTS

Customers are currently limited by memory capacity, which requires them having to scale out and have a larger hardware footprint. Scaling up to increase memory capacity is often cost-prohibitive.
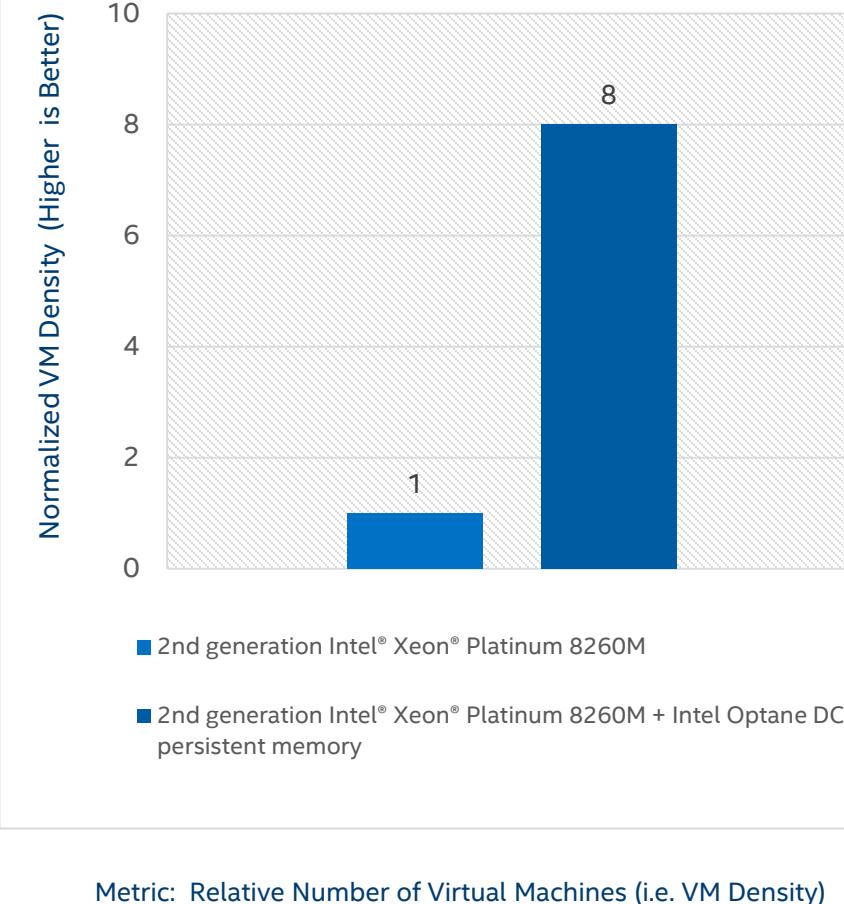
SOLUTION

- With 2nd generation Intel® Xeon® Scalable processors and Intel® Optane™ DC persistent memory (Memory Mode), Altibase can take advantage of the larger memory capacities per socket, while making it a more cost-effective solution at the same time.
- Customers can now keep more data for transaction processing closer to the CPU, with negligible performance impact[1] **at a reduced cost** (see chart showing up to 20% performance improvement for a given cost)[2].

VALUE PROPOSITION

Better performance at similar cost - Sunjesoft customers can benefit from being able to process more transactions per second), for a given cost[1,2] and while meeting SLA targets for the response time.

VIRTUOZZO (VIRTUOZZO 7)*

Memory Mode

NEW

Metric: Relative Number of Virtual Machines (i.e. VM Density)

APPLICATION

Virtuozzo is used by over 700 Cloud Service Providers (CSPs) and enterprises to enable over 5 million virtual environments running mission-critical cloud workloads. The workload involved spawning a suite of three Consolidated Stack Units (CSUs) of Virtual Machines (VMs), running synthetic workloads (spejbb, sysbench, webbench). The test mimics a typical customer scenario, where a Java app, a Database and a Webserver may run on a host, with gradual increase number of VMs.

CUSTOMER CHALLENGES

CSPs typically over-provision number of VM instances, but are often limited on the number of VMs that can be supported in a given system currently, due to limited memory capacity. Adding additional memory is often cost prohibitive, resulting in the need to scale out and increasing HW footprint.

SOLUTION

Higher memory capacities of Intel® Optane™ DC persistent memory (Memory Mode) facilitates supporting more number of VMs on a single host (see chart, showing up to 8X improvement in VM density)[1], while maintaining similar geomean performance. This approach thereby provides a more cost-effective alternative.

VALUE PROPOSITION

Greater VM density at similar cost – Virtuozzo customers will benefit from being able to have more VMs available on demand, to meet spikes in user activity, thus eliminating the considerable waiting time needed to spawn a VM (and enhancing revenue potential for customers).

25

Case Study: Cloud – in-memory database

Kingsoft Cloud
KSCLOUD

"We chose Intel Xeon Scalable processors and Intel Optane DC persistent memory for the combined value and ability to deliver improved TCO for our Cloud Redis Service. We can now maintain our business requirements and meet specified SLAs at lower cost, while ensuring we're delivering superior level of service our customers expect." *Liu Tao, Co-Partner & GM of Public Cloud BU, Kingsoft Cloud**

RESULT
1.3X[1] TCO BENEFITS

Over baseline testing using Intel® Xeon® Scalable 8260 processors and Intel® Optane™ DC persistent memory

intel XEON PLATINUM inside

intel OPTANE™ DC PERSISTENT MEMORY

Customer: Kingsoft Cloud*, provides online cloud computing, storage and distribution services, ranking among the top CDN/video, public cloud in enterprise, government, and gaming industries and so on.

Challenge: Kingsoft Cloud* needs to differentiate itself with new cloud services at optimized cost and needs the agility to be able to address customer demand with scalable cloud solutions. Kingsoft Cloud wanted to deploy more instances which requires higher memory capacity, but the cost of DRAM at large capacities was prohibitive.

Solution: Kingsoft Cloud* chose 2nd generation Intel® Xeon® Scalable processors and Intel® Optane™ DC persistent memory operating in Memory Mode versus a DRAM-only solution for the company's Cloud Redis* Service and was able to achieve 1.3X TCO benefit[1] while meeting SLA requirements.
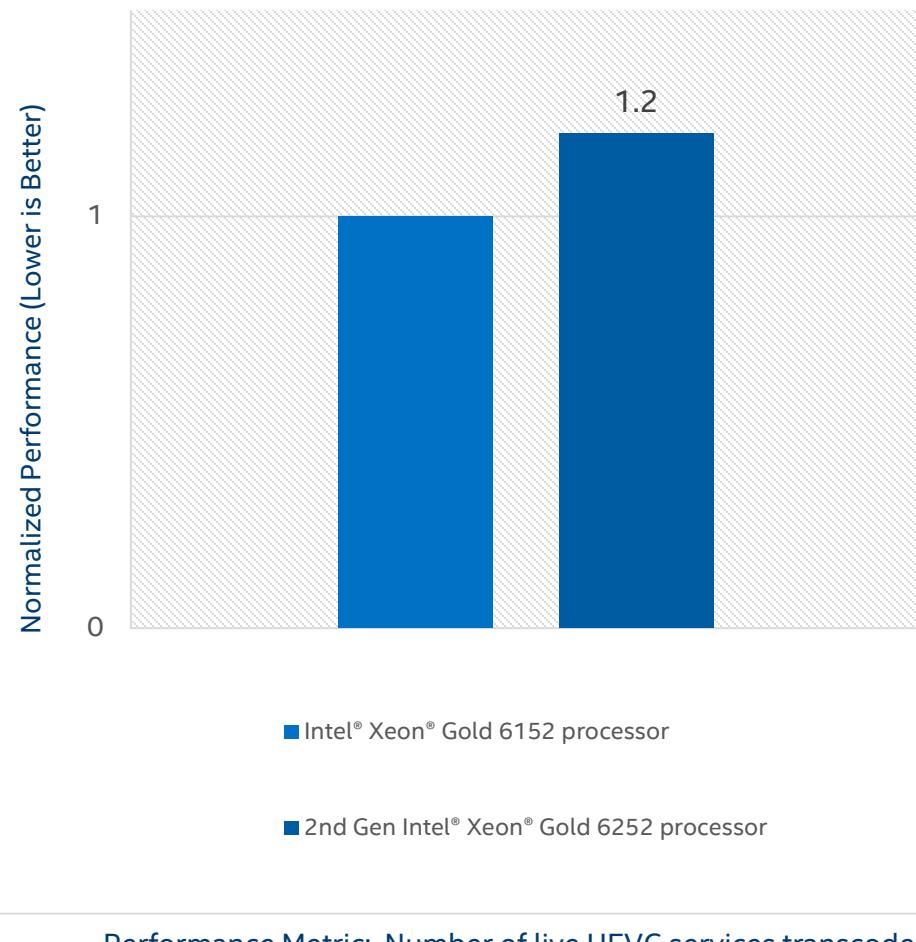
INTEL® DL BOOST: SOFTWARE PROOF POINTS

ANEVIA (GENOVA LIVE)*

http://www.anevia.com/

NEW

Normalized Performance (Lower is Better)

1.2

1

0

■ Intel® Xeon® Gold 6152 processor

■ 2nd Gen Intel® Xeon® Gold 6252 processor

Performance Metric: Number of live HEVC services transcoded

APPLICATION

Anevia is one of the leading software providers for delivery of live TV, time shifted TV, streaming and video on demand services. Their Genova Live application is a real-time HEVC video transcoder for Internet TV. This is a core bound application..

CUSTOMER PAIN POINTS

Anevia is impacted by the cost of hardware infrastructure and quality of video being delivered. As such, any improvement in performance translates to being able to transcode (1) more video channels in parallel at the same quality or (2) same number of video channels at lower bit rate, which saves costs. Hence customer is always seeking to get the best possible performance out of their hardware.

SOLUTION

With 2nd Gen Intel® Xeon® Scalable processors, Anevia is able to achieve up to 22%[1] more live HD services compared to previous generation, due to improved memory bandwidth and the new Vector Neural Network Instructions. These new instructions helped more HEVC services to be transcoded, than without.

VALUE PROPOSITION

- Higher density and lower hardware cost per TV service.
- Lower bitrate, translating to lower network bandwidth usage and lower cost, for the same video quality and density.

1 - Performance results are based on testing by Intel and Anevia on Apr 9, 2019 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.
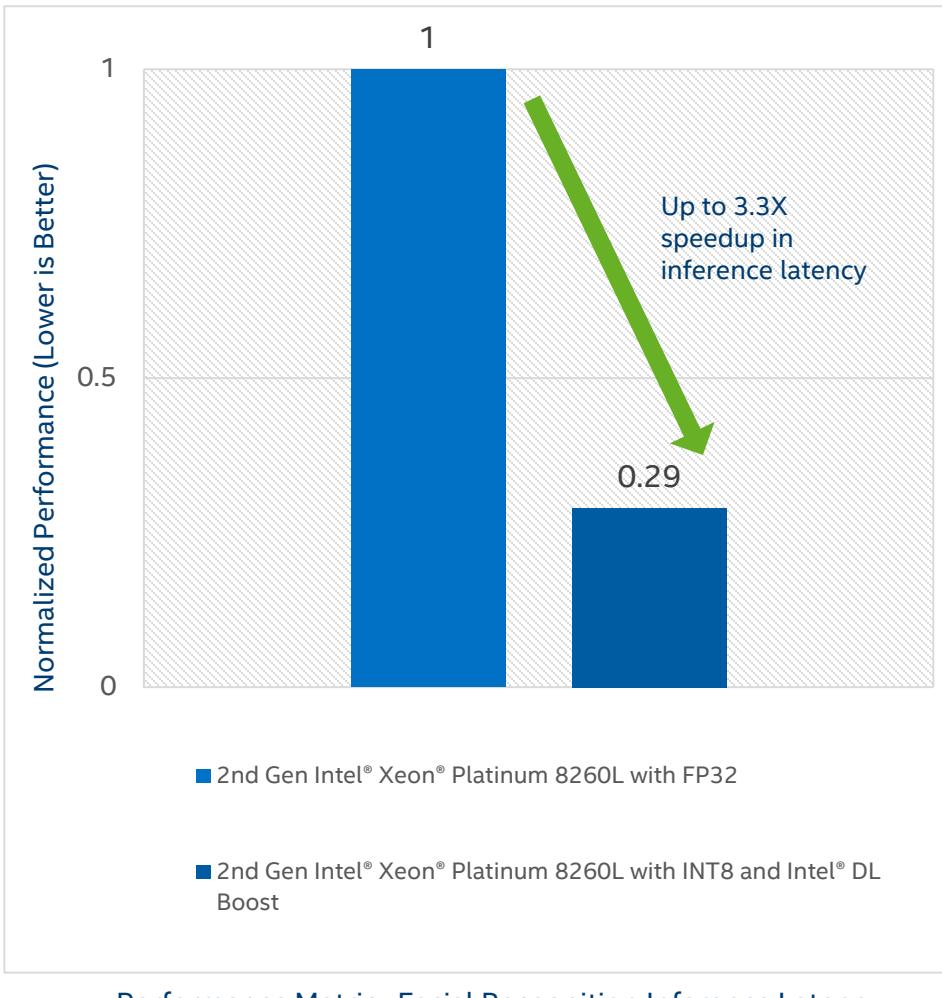
CLOUDWALK (IMAGE RECOGNITION)*

http://www.cloudwalk.cn/

Up to 3.3X speedup in inference latency

1

0.29

Normalized Performance (Lower is Better)

■ 2nd Gen Intel® Xeon® Platinum 8260L with FP32

■ 2nd Gen Intel® Xeon® Platinum 8260L with INT8 and Intel® DL Boost

Performance Metric: Facial Recognition Inference Latency

APPLICATION

Cloudwalk is one of Top 3 computer vision technology companies in China, providing solutions and services for the public security and finance sectors. Cloudwalk facial recognition application is used for customer recognition in banking, security, transportation, remote ID recognition and so on. Cloudwalk uses customized Resnet50 for online facial recognition.

CUSTOMER CHALLENGES

Deploying facial recognition solutions in bank, security government or police station faces two bottlenecks: network bandwidth and computing capabilities, that negatively impact deep learning inference throughput and latency, thereby resulting in less than optimal user experiences.

SOLUTION

For facial recognition, a customized Resnet50 (FP32 and INT8) model was optimized on Intel Caffe. Compared to FP32, Intel® DL Boost (delivered by Vector Neural Network Instructions (VNNI)/INT8) optimizations helped achieve a **speedup of 3.3X in inference latency**, for the same batch size and same instance (see chart)[1]. The application also meets Cloudwalk's desired accuracy requirement (of accuracy loss being less than 0.03%).

VALUE PROPOSITION

Significantly reduced deep learning inference latency, delivering better user experience – Cloudwalk customers will benefit from improved performance (i.e., lower latency), while maintaining SLAs for accuracy loss.

ASIAINFO (AURA FACIAL RECOGNITION)*

http://www.asiainfo.com

2nd Gen Intel® Xeon Scalable processors with Intel® DL Boost

Chart

Up to 2.08X speedup in inference latency

0.48

Normalized Performance (Lower is Better)

1

0.5

0

■ 2nd Gen Intel® Xeon® Platinum 8260L with FP32

■ 2nd Gen Intel® Xeon® Platinum 8260L with INT8 and Intel® DL Boost

Performance Metric: Facial Recognition Inference Latency

APPLICATION

AsiaInfo provides big data and AI solutions to all three telecom carriers in China. As the key AI product, Asiainfo Aura is a Machine Learning (ML) and Deep Learning (DL) development platform that is driven by telecom sector data. Within this platform, the function of Face Recognition uses the Keras VGG16 model for image recognition to identify the human face online.

CUSTOMER CHALLENGES

Deploying facial recognition solutions often faces two bottlenecks: network bandwidth and computing capabilities, that negatively impact deep learning inference throughput and latency, thereby resulting in less than optimal user experiences.
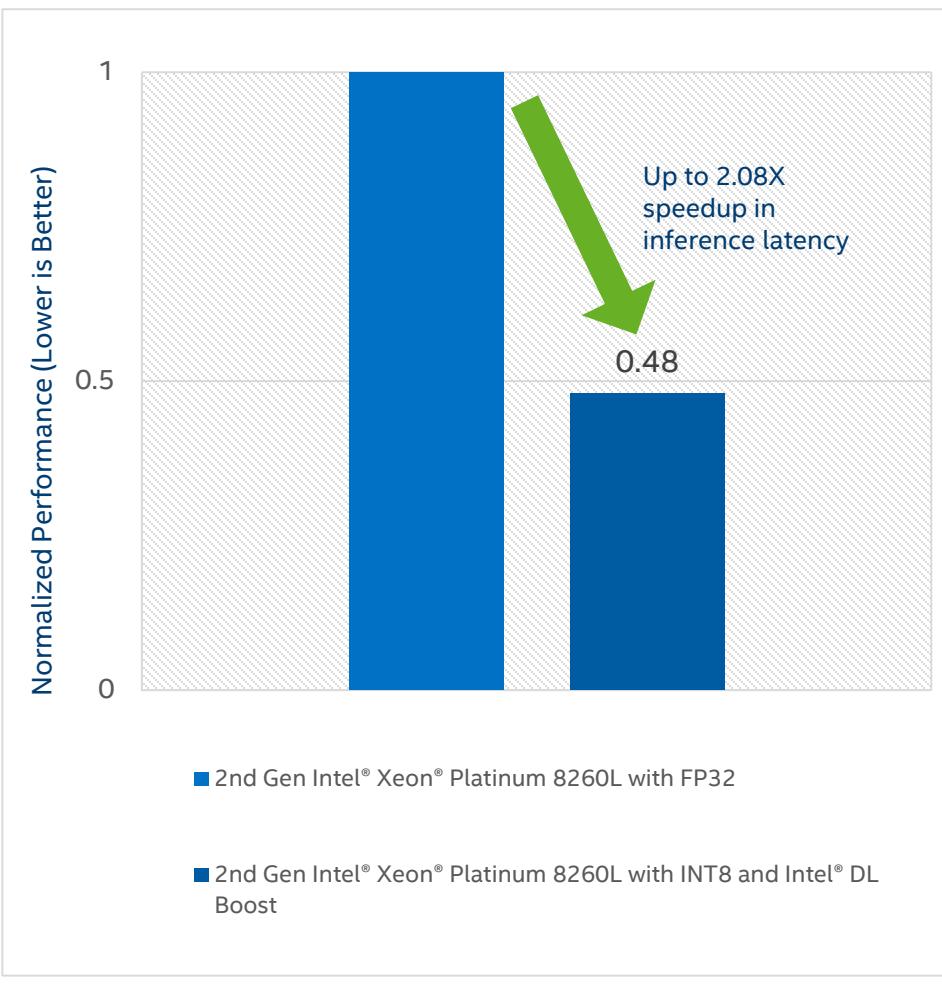
SOLUTION

For facial recognition, the VGG16 model was optimized on Intel Caffe. Compared to FP32, Intel® DL Boost (delivered by Vector Neural Network Instructions (VNNI)/INT8) optimizations helped achieve a **2.08X speedup in inference latency**, for the same batch size and same instance (see chart)[1], while keeping to the desired accuracy requirements.

VALUE PROPOSITION

Significantly reduced deep learning inference latency, delivering better user experience – AsiaInfo customers will benefit from improved performance (i.e., lower latency), while maintaining SLAs for accuracy loss.

HISIGN (IMAGE RECOGNITION)*

http://www.hisign.com.cn/en-us/index.aspx

Normalized Performance (Higher is Better)

2.18

- 2nd Gen Intel® Xeon® Platinum 8260L with FP32
- 2nd Gen Intel® Xeon® Platinum 8260L with INT8 and Intel® DL Boost

Performance Metric: Inference Throughput, while meeting partner-specified latency SLA

APPLICATION

Hisign is one of the key biometric companies in China. They provide biometric authentication solutions, including fingerprint, voice and face recognition - for Foreign Affairs and Public Security. Their solutions are used for national ID card systems, e-passports, and for access control systems in enterprise. Hisign Face Recognition software uses customized Resnet32 to accurately identify the human face.

CUSTOMER CHALLENGES

Deploying facial recognition solutions often faces two bottlenecks: network bandwidth and computing capabilities, that negatively impact deep learning inference throughput and latency, thereby result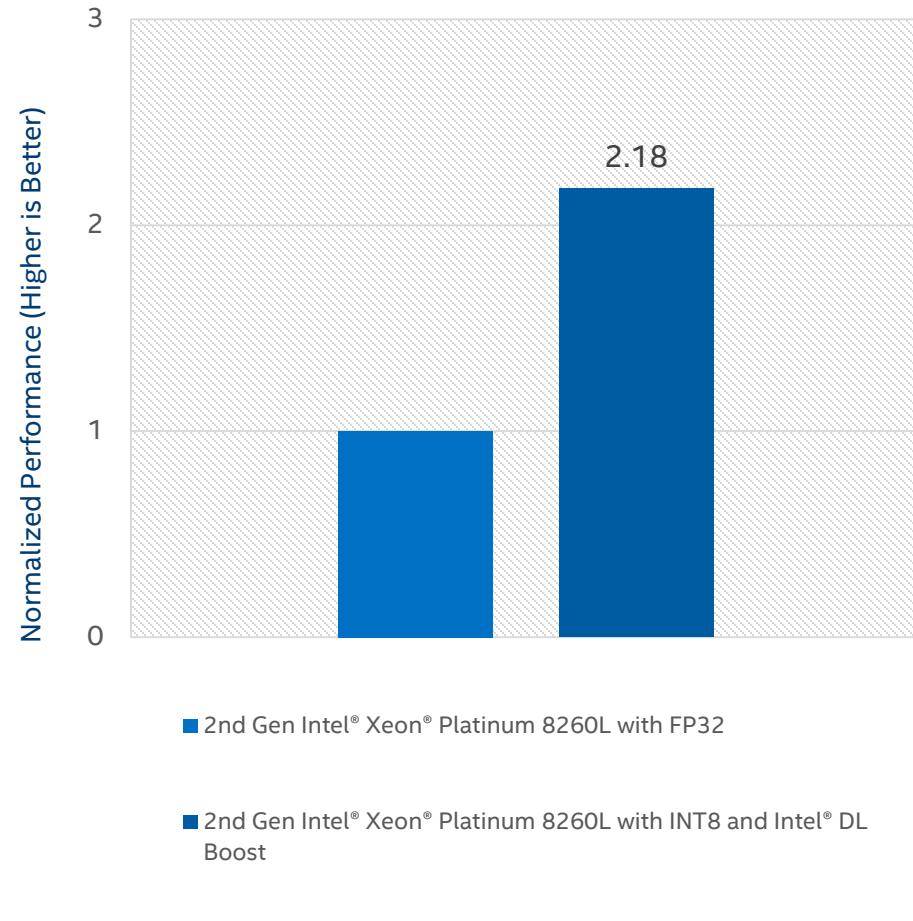ing in less than optimal user experiences. Additionally, achieving improvement in throughput often requires scaling out, resulting in added deployment costs & complexity.

SOLUTION

For facial recognition, a customized Resnet32 model was optimized on Intel Caffe. Compared to FP32, Intel® DL Boost (delivered by Vector Neural Network Instructions (VNNI)/INT8) optimizations helped increase deep learning inference throughput by 2.18X (see chart)[1], while meeting the partner-specified latency of less than 10ms.

VALUE PROPOSITION

Significantly improved deep learning inference throughput, without impacting latency, thus delivering better user experience .

Neusoft®
http://www.neusoft.com

2nd Gen Intel® Xeon Scalable processors with Intel® DL Boost

intel XEON PLATINUM inside™

Normalized Performance (Higher is Better)

2.02

- 2nd Gen Intel® Xeon® Platinum 8260L with FP32
- 2nd Gen Intel® Xeon® Platinum 8260L with INT8 and Intel® DL Boost

Performance Metric: Inference Throughput

APPLICATION

Neusoft is one of the largest software solution & service provider in China. CareVault is a Smart Medical Cloud Platform, providing AI tools to assist medical and scientific research. One of the major applications of CareVault is Down's Syndrome medical image classification. The application is based on Intel® Caffe and customized Alexnet to predict Down's Syndrome, and the medical image inference processing runs offline.

CUSTOMER CHALLENGES

Image recognition and classification applications are often compute-bound. This negatively impacts deep learning inference throughput, thereby resulting in less than optimal user experiences..
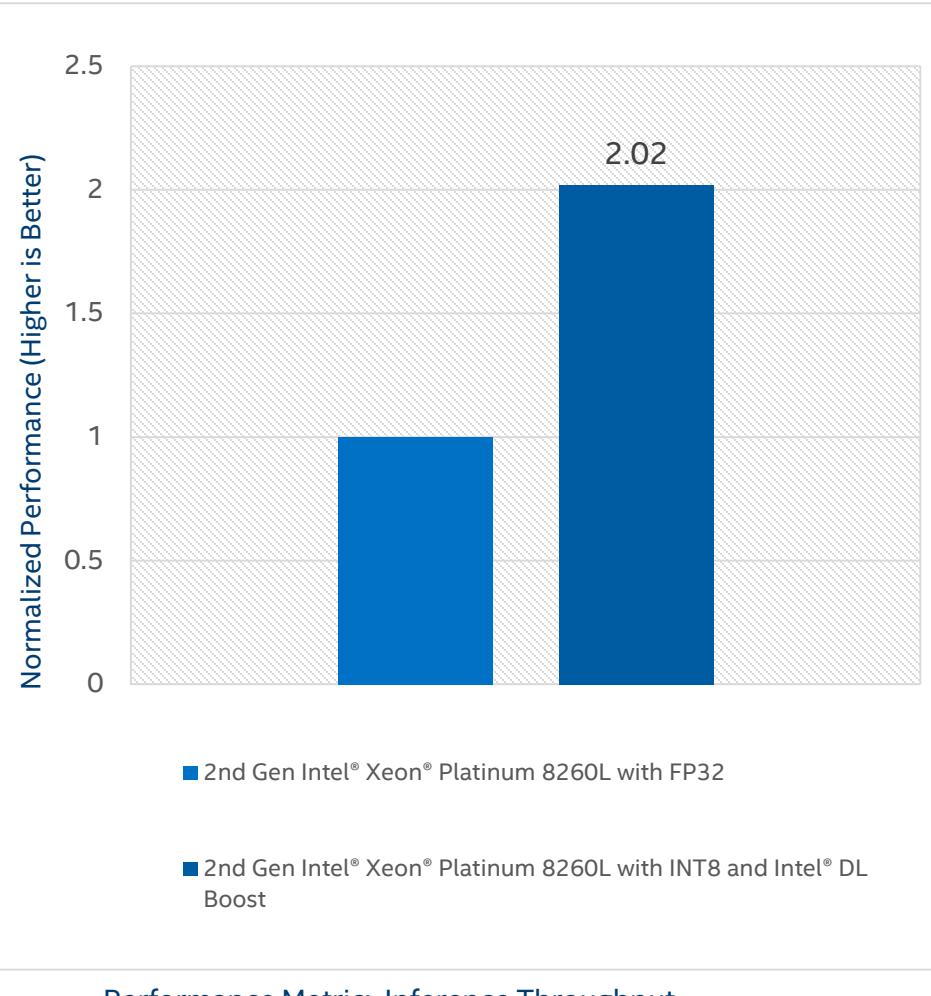
SOLUTION

For medical imaging, a customized Alexnet model was optimized on Intel Caffe. Compared to FP32, Intel® DL Boost (delivered by Vector Neural Network Instructions (VNNI)/INT8) optimizations helped increase deep learning inference throughput by 2.02X (see chart)[1], while meeting Neusoft's accuracy requirements.
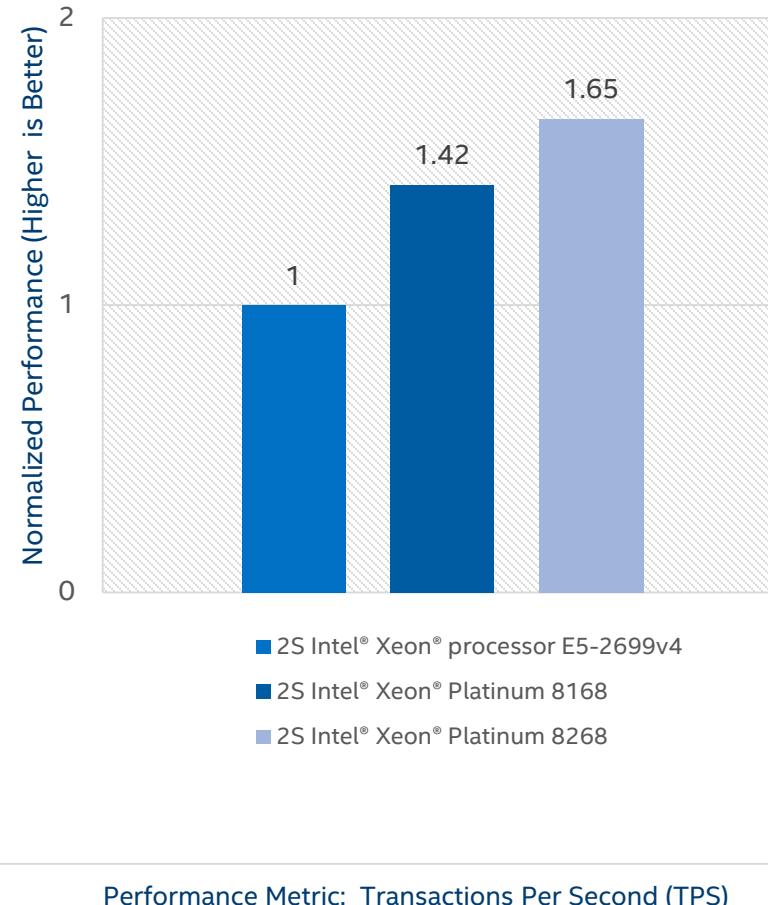
VALUE PROPOSITION

Significantly improved deep learning inference throughput, while meeting accuracy requirements, thus delivering better user experience .

(intel)

2ND GEN INTEL® XEON® SCALABLE PROCESSOR: GEN-TO-GEN SOFTWARE PROOF POINTS

Performance Metric: Transactions Per Second (TPS)

Chart — Normalized Performance (Higher is Better):
- 2S Intel® Xeon® processor E5-2699v4: 1
- 2S Intel® Xeon® Platinum 8168: 1.42
- 2S Intel® Xeon® Platinum 8268: 1.65

APPLICATION

Altibase 7.1* is an in-memory Relational Database Management System that provides fast data processing speeds for online transactional processing (OLTP) and online analytical processing (OLAP). The workload measures transactions per second (TPS) using SQLCLI mode based TCP. It is aiming for real-time access to time critical data.

CUSTOMER BENEFITS

- Accelerates time to insights and analytics, due to faster data processing speeds for OLTP and OLAP
- Greater throughput from improved core scalability and faster memory speed

PERFORMANCE CONTRIBUTORS

- With 2nd generation Intel® Xeon® Platinum 8268 processor, Altibase can take advantage of the higher core count and faster memory speed, compared to the 3-year old baseline.
- Intel® Advanced Vector Instructions 512, available with Intel Xeon Scalable processors, also contributed to the performance gain of 65% compared to older generation processor (see chart)[1].

VALUE PROPOSITION

Significantly higher throughput compared to older systems – with the latest 2nd Gen Intel® Xeon® Platinum 8268 processor, Altibase can achieve analytics results more rapidly, while benefiting from faster transactions while storing and manipulating data.

1 - Performance results are based on testing by Intel and Altibase on 1/31/19 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.

SUNJESOFT (GOLDILOCK V3.2.0)*

Performance Metric: Transactions Per Second (TPS)

Legend:
- 2S Intel® Xeon® processor E5-2699v4
- 2S Intel® Xeon® Platinum 8168
- 2S Intel® Xeon® Platinum 8268

Chart values: 1, 1.39, 1.43

Y-axis: Normalized Performance (Higher is Better)

APPLICATION

Sunjesoft Goldilock v3.2.0* is an in-memory database with high scalability and low latency performance. The workload measures transactions per second (TPS) with multi-client Direct Access (DA) node..

CUSTOMER BENEFITS

- Deliver faster response times and serve more customers
- Greater throughput from improved core scalability and faster memory speed
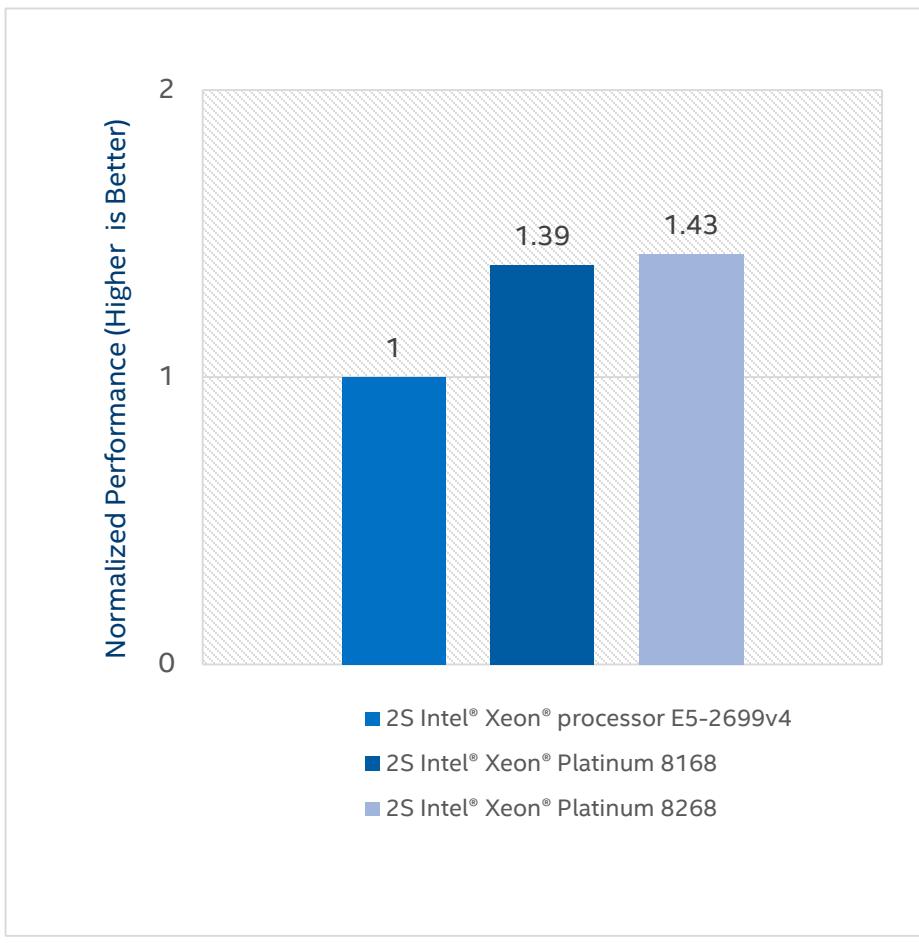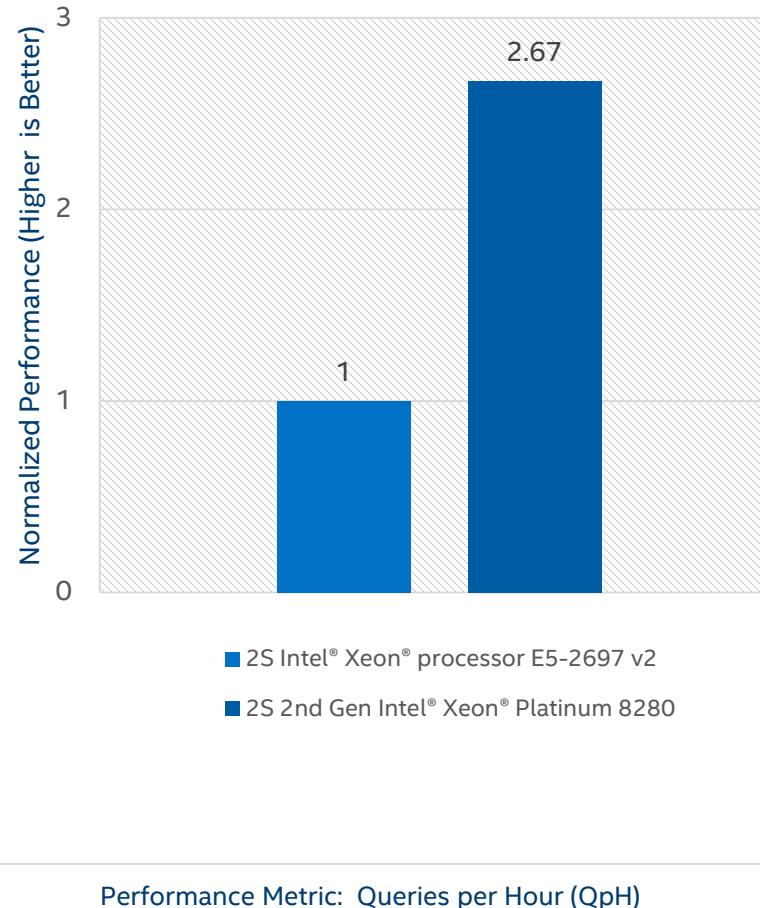
PERFORMANCE CONTRIBUTORS

- With 2nd generation Intel® Xeon® Platinum 8268 processor, Sunjesoft can take advantage of the higher core count and faster memory speed, compared to the 3-year old baseline.
- Intel® Advanced Vector Instructions 512, available with Intel Xeon Scalable processors, also contributed to the performance gain of 43% compared to older generation baseline (see chart)[1].

VALUE PROPOSITION

Higher throughput compared to older systems – with the latest 2nd Gen Intel® Xeon® Platinum 8268 processor, Sunjesoft customers benefit from faster transactions.

1 - Performance results are based on testing by Intel and Sunjesoft on 1/29/19 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.

Normalized Performance (Higher is Better)

2.67

1

■ 2S Intel® Xeon® processor E5-2697 v2
■ 2S 2nd Gen Intel® Xeon® Platinum 8280

Performance Metric: Queries per Hour (QpH)

APPLICATION

Db2® is IBM's flagship database product which supports in-memory column store tables for analytics workloads. The proprietary IBM Big Data Insights Workload (BDInsights) is a multi-user data warehousing workload based on a retail environment. The workload as configured uses a 300GB scale factor and 12 concurrent users. It is compute-bound in the tested configuration

CUSTOMER BENEFITS

- Accelerates response time for analytics queries
- Higher throughput to support more concurrent users for customers

PERFORMANCE CONTRIBUTORS

- Greater number of cores/threads and improved memory bandwidth of Intel® Xeon Scalable processors
- Intel® Advanced Vector Instructions 512, available with Intel Xeon Scalable processors, also contributed to the performance gain of 65% compared to older generation processor (see chart)[1.]

VALUE PROPOSITION

Significantly higher throughput compared to older hardware – with the latest 2nd Gen Intel® Xeon® Platinum 8280 processor, IBM Db2 users can achieve analytics results more rapidly, while supporting more concurrent users.

CONFIGURATIONS & DISCLAIMERS

(Listed in the order that the slides appear in the deck)

Configurations for software <u>proof points highlighting Intel® Optane™ DC persistent memory optimizations & performance</u>

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® Optane™ 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

Altibase 7.1* (self-defined workload). OS: CentOS 7.6 kernel 3.10.0-957.el7.x86_64. Testing by Intel and Altibase completed on Jan 31, 2019. Security Mitigations for Variants 1, 2, 3, 3a, 4 and L1TF in place.
BASELINE: 2nd Gen Intel® Xeon® Platinum 8280 processor, 2.4 GHz, 24 cores, turbo and HT on, BIOS SE5C620.86B.0D.01.0134.100420181737, 1.5TB total memory, 24 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 1x480GB SSD(Intel SSD DC S3500)
NEW: 2nd Gen Intel® Xeon® Platinum 8280 processor, 2.4 GHz, 24 cores, turbo and HT on, BIOS SE5C620.86B.0D.01.0134.100420181737, 192GB total memory, 12 slots / 16GB / 2666 MT/s / DDR4 LRDIMM and 1.5 TB DCPMM, 12 slots/ 128 GB / 2666 MT/s Intel® Optane™ DC persistent memory, 1x480GB SSD(Intel SSD DC S3500)
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.

AsiaInfo Telco Business Support System* 3.1.1 + self defined workload. OS: Red Hat Enterprise Linux* 7.5 kernel 3.10.0-957.1.3.el7.x86_64. Testing by Intel and AsiaInfo completed on Dec 28, 2018. Security Mitigations for Variants 1, 2, 3 and L1TF in place.
BASELINE: 2S Intel® Xeon® processor E5-2699v4, 2.5GHz, 18 cores, turbo and HT on, BIOS 251R01, 256GB total memory, 32 slots / 32GB / 1600 MT/s / DDR4 LRDIMM, 7 x 800GB, Intel SSD DC S3700 + 4 2TB Intel® SSD Data Center Family for NVMe*
NEXT GEN: Intel® Xeon® Platinum 8180 processor, 2.5GHz, 28 cores, turbo and HT on, BIOS x0007, 768GB total memory, 32 slots / 32GB / 1600 MT/s / DDR4 LRDIMM, 7 x 800GB, Intel SSD DC S3700 + 4 2TB Intel SSD Data Center Family for NVMe.
NEW: 2nd Gen Intel® Xeon® Platinum 8280 processor, 2.7GHz, 28 cores, turbo and HT on, BIOS 1.018, 192GB total memory, 12 slots / 16GB / 1600 MT/s / DDR4 LRDIMM and 8 slots/ 128 GB / Intel® Optane™ DC persistent memory, 7 x 800GB, Intel SSD DC S3700 + 4 2TB Intel SSD Data Center Family for NVMe

Shanghai Baosight xInsight* v2.0 (self-defined workload); OS: CentOS* 7.5 Kernel 3.10.0-957.1.3.el7.x86_64. Testing by Intel and Baosight completed on Jan 8, 2019. Security Mitigations for Variants 1, 2, 3 and L1TF in place.
BASELINE: 2nd Gen Intel® Xeon® Platinum 8260L processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.0180, 768GB total memory, 32 slots / 32GB / 2666 MT/s / DDR4 LRDIMM, 1 x 480GB / Intel SSD DC S4500 + 2 x 1TB / Intel SSD DC P4500
NEW: 2nd Gen Intel® Xeon® Platinum 8260L processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.0180, 192GB total memory, 12 slots / 16GB / 2666 MT/s / DDR4 LRDIMM and 1024GB DCPMM, 8 slots/ 128 GB / 2666 MT/s Intel® Optane™ DC persistent memory, 1 x 480GB / Intel SSD DC S4500 + 2 x 1TB / Intel SSD DC P4500

GBase 8m* V8.5.1 + self defined workload. OS: CentOS* 7.5 kernel 3.10.0-957.1.3.el7.x86_64. Testing by Intel and GBase completed on Dec 28, 2018. Security Mitigations for Variants 1, 2, 3 and L1TF in place
BASELINE: 2nd Gen Intel® Xeon® Platinum 8260 processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.018, 768 GB total memory, 12 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 1 x 480GB / Intel SSD DC S4500 + 1 x 1TB / Intel SSD DC P4500
NEW: 2nd Gen Intel® Xeon® Platinum 8260 processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.018, 192GB total memory, 12 slots / 16GB / 2933 MT/s / DDR4 LRDIMM and 8 slots/ 128 GB / Intel® Optane™ DC persistent memory, 1 x 480GB / Intel SSD DC S4500 + 1 x 1TB / Intel SSD DC P4500

HiSign Fingerprint* MCH v4. OS: Red Hat Enterprise Linux* 7.5 4.19.3-1.el7.elrepo.x86_64. Testing by Intel and HiSign completed on Jan 7, 2019. Security Mitigations for Variants 1, 2, 3 and L1TF in place.
BASELINE: 2nd Gen Intel® Xeon® Platinum 8260 processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.018, 768 GB total memory, 12 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 1 x 480GB / Intel SSD DC S4500 + 1 x 1TB / Intel SSD DC P4500
NEW: 2nd Gen Intel® Xeon® Platinum 8260 processor, 2.3GHz, 24 cores, turbo and HT on, BIOS 1.018, 192GB total memory, 12 slots / 16GB / 2933 MT/s / DDR4 LRDIMM and 8 slots/ 128 GB / Intel® OptaneTM DC persistent memory, 1 x 480GB / Intel SSD DC S4500 + 1 x 1TB / Intel SSD DC P4500

Huawei FusionSphere 6.3.1* MySQL (5.7.24) SysBench (1.0.6) workload; OS: FusionSphere HyperV, Kernel 3.10.0-514.44.5.10_96.x86_64. Testing by Intel and Huawei completed on Jan 11, 2019. Huawei confirmed security mitigations for Variants 1, 2, 3 and L1TF in place.
BASELINE: 2nd Gen Intel® Xeon® Platinum 8260 processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.0180, 768GB total memory, 16 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 1 x 480GB / Intel SSD DC S4500 + 3 x 2TB / Intel SSD DC P4500
NEW: 2nd Gen Intel® Xeon® Platinum 8260 processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.0180, 384GB total memory, 12 slots / 32GB / 2666 MT/s / DDR4 LRDIMM and 12 slots/ 128 GB / Intel® Optane™ DC persistent memory, 1 x 480GB / Intel SSD DC S4500 + 3 x 2TB / Intel SSD DC P4500

Kingbase KADB* V3R2 (self-defined workload); OS: CentOS* 7.5 Kernel 4.19.13. Testing by Intel and Kingbase completed on Jan 10, 2019. Security Mitigations for Variants 1, 2, 3 and L1TF in place.
BASELINE: 2nd Gen Intel® Xeon® Platinum 8260L processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.0180, 1536GB total memory, 24 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 1 x 240GB / Intel SSD DC S4500 + 1 x 2TB / Intel SSD DC P3520
NEW: 2nd Gen Intel® Xeon® Platinum 8260L processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.0180, 192GB total memory, 12 slots / 16GB / 2666 MT/s / DDR4 LRDIMM and 1536GB DCPMM, 12 slots/ 128 GB / 2666 MT/s Intel® Optane DC persistent memory, 1 x 240GB / Intel SSD DC S4500
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.

Kx Systems KDB+* v3.6 (STAC* Antuco benchmark); OS: Red Hat Enterprise Linux* 7.6 kernel 3.10.0-957.el7.x86_64. Testing by Intel and KX, and auditing by STAC, completed on March 25, 2019.
BASELINE: Lenovo ThinkSystem SR650: 2S Intel® Xeon® Platinum processor 8280L, 2.7GHz, 28 cores, turbo and HT on, BIOS IVE135N-2.10, 768GB of DDR4, 12 slots / 64GB / 2933 MT/s / DDR4 LRDIMM, 2 x 8TB Intel SSD P4600 NVMe, Red Hat Enterprise Linux* 7.6 kernel 3.10.0-957.el7.x86_64.
NEW: Lenovo ThinkSystem SR950?: 2S Intel® Xeon® Platinum processor 8280L, 2.7GHz, 28 cores, turbo and HT on, BIOS IVE135N-2.10, 768GB of DDR4, 12 slots / 64GB / 2933 MT/s / DDR4 LRDIMM, 12288GB of DCPMM, 24 slots / 512GB / 2666 MT/s, 2 x 8TB Intel SSD P4600 NVMe, Red Hat Enterprise Linux* 7.6 kernel 3.10.0-957.el7.x86_64.

Kx Systems KDB+* v3.6 (STAC* Kanaga benchmark); Workload consists of historical data for 2011-2012; OS: Red Hat Enterprise Linux* 7.6 kernel 3.10.0-957.el7.x86_64. Testing by Intel and KX, and auditing by STAC, completed on March 25, 2019..
BASELINE: Lenovo ThinkSystem SR950?: 4S Intel® Xeon® Platinum processor 8280L, 2.7GHz, 28 cores, turbo and HT on, BIOS PSE121C-1.50, 1532GB of DDR4, 24 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 6 x 8TB Intel SSD P4600 NVMe, Red Hat Enterprise Linux* 7.6 kernel 3.10.0-957.el7.x86_64.
NEW: Lenovo ThinkSystem SR950?: 4S Intel® Xeon® Platinum processor 8280L, 2.7GHz, 28 cores, turbo and HT on, BIOS PSE121C-1.50, 1532GB of DDR4, 24 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 12288GB of DCPMM, 24 slots / 512GB / 2666 MT/s, 6 x 8TB Intel SSD P4600 NVMe, Red Hat Enterprise Linux* 7.6 kernel 3.10.0-957.el7.x86_64

NARI Power Consumption Analysis (PCA) System* 2018v6 (self-defined workload); OS: Fedora 25* 7.5 Kernel 4.21.0-0.rc0.git6.1.fc30.x86_64. Testing by Intel and NARI completed on Dec 28, 2018. Security Mitigations for Variants 1, 2, 3 and L1TF in place.
BASELINE: 2nd Gen Intel® Xeon® Platinum 8260L processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.0180, 768GB total memory, 12 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 1 x 480GB / Intel SSD DC S4500 + 2 x 1TB / Intel SSD DC P4500
NEW: 2nd Gen Intel® Xeon® Platinum 8260L processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.0180, 192GB total memory, 12 slots / 16GB / 2666 MT/s / DDR4 LRDIMM and 1024GB DCPMM, 8 slots/ 128 GB / 2666 MT/s Intel® Optane DC persistent memory, 1 x 480GB / Intel SSD DC S4500 + 2 x 1TB / Intel SSD DC P4500

Neusoft Aclome DB* v9.3 (self-defined workload); OS: Red Hat Enterprise Linux* 7.5 Kernel 4.19.3-1.el7.elrepo.x86_64. Testing by Intel and Neusoft completed on Dec 27, 2018. Security Mitigations for Variants 1, 2, 3 and L1TF in place.
BASELINE: 2nd Gen Intel® Xeon® Platinum 8260L processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.0180, 768GB total memory, 12 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 1 x 480GB /Intel SSD DC S4500 + 2 x 1TB / Intel SSD DC P4500
NEW: 2nd Gen Intel® Xeon® Platinum 8260L processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.0180, 192GB total memory, 12 slots / 16GB / 2666 MT/s / DDR4 LRDIMM and 1024GB DCPMM, 8 slots/ 128 GB / 2666 MT/s Intel® Optane DC persistent memory, 1 x 480GB / Intel SSD DC S4500 + 2 x 1TB / Intel SSD DC P4500

Redis Labs* Multi-modal Database; Workload is memtier benchmark (6 billion keys , 100B values, 50/50 read/write ratio, random access). Redis Enterprise 5.4, CentOS 7.5 kernel 4.19.8. Testing by Intel, completed on 2/14/2019. Security Mitigations in place. BASELINE (DRAM-only system): 2S Intel® Xeon® Platinum 8280L processor, 28 cores, turbo and HT on, BIOS SE5C620.86B.0D.01.0286.011120190816, 1.5TB of total memory, 24 slots / 64GB 2666MHz DDR4 DIMM, 2 x 40Gb bonded XL710 networking.
NEW (DRAM + Intel® Optane DC persistent memory): 2S Intel® Xeon® Platinum 8280L processor, 28 cores, turbo and HT on, BIOS SE5C620.86B.0D.01.0286.011120190816, 192GB of DRAM (12 x 16GB 2666MHz DRAM) and 1.5TB of DCPMEM (12 x 128GB DCPMEM) in Memory Mode, 2 x 40Gb bonded XL710 networking.
CLIENT: (same client used to drive the workload for both baseline and memory mode): 2S Intel® Xeon® processor 8180 2.5GHz, 28 cores, turbo and HT on, BIOS SE5C620.86B.00.01.0009.101920170742, 384 GB of total memory, 12 slots / 32GB 2666MHz DDR4 DIMM, 1 x 40Gb XL710 networking.

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.
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.

Sunjesoft Goldilock* v3.2.0 (self-defined workload). OS: CentOS 7.6 kernel 3.10.0-957.el7.x86_64. Testing by Intel and SUNJESOFT completed on Jan 31, 2019. Security Mitigations for Variants 1, 2, 3, 3a, 4 and L1TF in place.
BASELINE: 2nd Gen Intel® Xeon® Platinum 8280 processor, 2.7 GHz, 28 cores, turbo and HT on, BIOS SE5C620.86B.0D.01.0134.100420181737, 1.5TB total memory, 24 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 1x480GB SSD(Intel SSD DC S3500)
NEW: 2nd Gen Intel® Xeon® Platinum 8280 processor, 2.7 GHz, 28 cores, turbo and HT on, BIOS SE5C620.86B.0D.01.0134.100420181737, 192GB total memory, 12 slots / 16GB / 2666 MT/s / DDR4 LRDIMM and 1.5 TB DCPMM,   12 slots/ 128 GB / 2666 MT/s Intel® OptaneTM DC persistent memory, 1x480GB SSD(Intel SSD DC S3500)
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.

Virtuozzo Virtuozzo* 7; Workload: Spawn Consolidated Stack Units (CSUs) of Virtual Machines in sets of 3, each running synthetic workloads of SpecJBB, SysBench, WebBench. This is a density load test where CSU performance is being measured when increasing quantity of CSUs, while memory is available (1,4,8,16, … 72, 80, 100, 120,…,300,320,340 CSUs). OS: Vzlinux 7.5.0-844, Kernel: vzkernel 3.10.0-862.11.6.vz7.64.7; 2S Intel® Xeon® processor Platinum 8260M, 2.3GHz, 24 cores, turbo and HT on, BIOS 0D.01.0286, Testing by Intel and Virtuozzo completed on Mar 10, 2019. Security Mitigations for Variants 1, 2, 3 and L1TF in place.
BASELINE (DRAM): 2S Intel® Xeon® Platinum processor 8260M 2.3GHz, 24 cores, turbo and HT on, BIOS 0D.01.0286, 384GB total memory, 12 slots / 32GB / 2666 MHz / DDR4 Micron, 400GB, Intel SSD DC S3700 + 1.6 2TB Intel SSD Data Center Family for NVMe, OS: Vzlinux 7.5.0-844, Kernel: vzkernel 3.10.0-862.11.6.vz7.64.7
NEW: 2S Intel® Xeon® Platinum processor 8260M 2.3GHz, 24 cores, turbo and HT on, BIOS 0D.01.0286, 384GB total memory, 12 slots / 32GB / 2666 MHz / DDR4 Micron, + 12 slots AEP /512GB / 2666MHz / / (Only 3.9TB of AEP visible due to CPU limiting capacity) FW: 01.02.00.5310 800GB, Intel SSD DC S3700 + 4 1.6TB Intel SSD Data Center Family for NVMe, OS: Vzlinux 7.5.0-844, Kernel: vzkernel 3.10.0-862.11.6.vz7.64.7

Kingsoft Cloud REDIS service* (self-defined workload); OS: Red Hat Enterprise Linux* 7.5 4.18.8-x86_64. Testing by Intel and Kingsoft Cloud completed on Jan 10, 2019. Security Mitigations for Variants 1, 2, 3 and L1TF in place.
BASELINE: 2nd Gen Intel® Xeon® Platinum 8260 processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.018, 1536GB total memory, 12 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 1 x 480GB / Intel SSD DC S4500 + 1 x 1TB / Intel SSD DC P4500
NEW: 2nd Gen Intel® Xeon® Platinum 8260 processor, 2.3GHz, 24 cores, turbo and HT on, BIOS 1.018, 1536GB total memory, 12 slots / 16GB / 2933 MT/s / DDR4 LRDIMM and 12 slots/ 128 GB / Intel® OptaneTM DC persistent memory, 1 x 480GB / Intel SSD DC S4500 + 1 x 1TB / Intel SSD DC P4500
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.

Configurations for software <u>proof points highlighting Intel® DL Boost optimizations & performance</u>

Anevia Genova Live* v5.4; OS: CentOS Linux* 7.6 kernel 3.10.0-957 Testing by Intel and Anevia completed on April 9, 2019. Security Mitigations for Variants 1, 2, 3, 3a, 4 and L1TF in place.
BASELINE: 2S Intel® Xeon® Gold processor 6152, 2.1GHz, 22 cores, turbo and HT on, BIOS 1.0286, microcode: 0x2000057, 192GB total memory, 12 slots / 16GB / 2666 MT/s / DDR4 dual ranked RDIMM, CentOS Linux* 7.6 kernel 3.10.0-957
NEW: 2S Intel® Xeon® Gold processor 6252, 2.1GHz, 24 cores, turbo and HT on, BIOS 1. 0395, microcode: 0x400001c, 192GB total memory, 12 slots / 16GB / 2933 MT/s / DDR4 dual ranked RDIMM, CentOS Linux* 7.6 kernel 3.10.0-957

Asiainfo Face Recognition* (self-defined workload); OS: Red Hat Enterprise Linux* 7.4 Kernel 4.19.3-1.el7.elrepo.x86_64. Testing by Intel and Asiainfo completed on Jan 5, 2019. Security Mitigations for Variants 1, 2, 3 and L1TF in place.
TEST SYSTEM CONFIG: 2nd Gen Intel® Xeon® Platinum 8260L processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.0180, 384GB total memory, 24 slots / 16GB / 2666 MT/s / DDR4 LRDIMM, 1 x 800GB / Intel SSD DC S4500; Intel® Optimization for Caffe

Cloudwalk Facial Recognition* (self-defined workload); OS: CentOS* 7.5 Kernel 3.10.0-957.1.3.el7.x86_64. Testing by Intel and Cloudwalk completed on Jan 5, 2019. Security Mitigations for Variants 1, 2, 3 and L1TF in place.
TEST SYSTEM CONFIG: 2nd Gen Intel® Xeon® Platinum 8260L processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.0180, 192GB total memory, 12 slots / 16GB / 2666 MT/s / DDR4 LRDIMM, 1 x 480GB / Intel SSD DC S4500 + 1 x 1TB / Intel SSD DC P4500; Intel® Optimization for Caffe

HiSign Face Recognition* (self-defined workload); OS: Red Hat Enterprise Linux* 7.5 Kernel 4.19.3-1.el7.elrepo.x86_64. Testing by Intel and HiSign completed on Jan 5, 2019. Security Mitigations for Variants 1, 2, 3 and L1TF in place.
TEST SYSTEM CONFIG: 2nd Gen Intel® Xeon® Platinum 8260L processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.0180, 768GB total memory, 12 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 1 x 480GB / Intel SSD DC S4500; Intel® Optimization for Caffe

Neusoft Medical Image Classification* (self-defined workload); OS: Red Hat Enterprise Linux* 7.5 Kernel 4.19.3-1.el7.elrepo.x86_64. Testing by Intel and Neusoft completed on Jan 5, 2019. Security Mitigations for Variants 1, 2, 3 and L1TF in place.
TEST SYSTEM CONFIG: 2nd Gen Intel® Xeon® Platinum 8260L processor, 2.3 GHz, 24 cores, turbo and HT on, BIOS 1.0180, 768GB total memory, 12 slots / 64GB / 2666 MT/s / DDR4 LRDIMM, 1 x 480GB / Intel SSD DC S4500; Intel® Optimization for Caffe

----
Configurations for <u>gen-to-gen</u> software performance proof points based on 2nd Gen Intel Xeon Scalable processors

Altibase Altibase 7.1* (self-defined workload). OS: CentOS 7.6 kernel 3.10.0-957.el7.x86_64. Testing by Intel and Altibase completed on Jan 31, 2019. Security Mitigations for Variants 1, 2, 3, 3a, 4 and L1TF in place.
BASELINE: 2S Intel® Xeon® processor E5-2699 v4, 2.2GHz, 22 cores, turbo and HT on, BIOS GRRFSDP1.86B.0275.R01.1603300531, 128GB total memory, 8 slots / 16GB / 2400 MT/s / DDR4 LRDIMM, 1TB HDD(SATA 7200rpm)
NEXT GEN: 2S Intel® Xeon® Platinum 8168 processor, 2.7GHz, 24 cores, turbo and HT on, BIOS SE5C620.86B.0D.01.0286.011120190816, 192GB total memory, 12 slots / 16GB / 2666 MT/s / DDR4 LRDIMM, 480GB SSD(Intel SSD DC S3500)
NEW: 2S Intel® Xeon® Platinum 8268 processor, 2.9GHz, 24 cores, turbo and HT on, BIOS SE5C620.86B.0D.01.0286.011120190816, 192GB total memory, 12 slots / 16GB / 2933 MT/s / DDR4 LRDIMM, 480GB SSD(Intel SSD DC S3500)

SUNJESOFT Goldilocks* v3.2.0 (self-defined workload). OS: CentOS 7.6 kernel 3.10.0-957.el7.x86_64. Testing by Intel and SUNJESOFT completed on Jan 29, 2019. Security Mitigations for Variants 1, 2, 3, 3a, 4 and L1TF in place.
BASELINE: 2S Intel® Xeon® processor E5-2699 v4, 2.2GHz, 22 cores, turbo and HT off, BIOS GRRFSDP1.86B.0275.R01.1603300531, 128GB total memory, 8 slots / 16GB / 2400 MT/s / DDR4 LRDIMM, 300GB HDD(SAS15000 rpm)
NEXT GEN: 2S Intel® Xeon® Platinum processor 8168, 2.7GHz, 24 cores, turbo and HT off, BIOS SE5C620.86B.0D.01.0286.011120190816, 192GB total memory, 12 slots / 16GB / 2666 MT/s / DDR4 LRDIMM, 480GB SSD(Intel SSD DC S3500)
NEW: 2S Intel® Xeon® Platinum processor 8268, 2.9GHz, 24 cores, turbo and HT off, BIOS SE5C620.86B.0D.01.0286.011120190816, 192GB total memory, 12 slots / 16GB / 2933 MT/s / DDR4 LRDIMM, 480GB SSD(Intel SSD DC S3500)

IBM® Db2®* v11.1.4.4. The IBM Big Data Insights Internal Heavy Multiuser Workload (BDInsights) is a multi-user data warehousing workload based on a retail environment. The workload is comprised with a mix of complex and intermediate queries. The scale factor for the workload is 300GB with 12 users. Testing by Intel completed on February 1, 2019. Security Mitigations for Variants 1, 2, 3, 3a, 4 in place.
BASELINE: 2S Intel® Xeon® processor E5-2697 v2, 2.7GHz, 12 cores, turbo on, HT on, BIOS 02.06.0007, 192GB total memory, 12 slots / 16GB / 1600 MT/s / DDR3 DIMM, 1 x 400GB, Intel® SSD DC S3700, Red Hat Enterprise Linux® 7.5, kernel 3.10.0-862.el7.x86_64.
NEW: 2S Intel® Xeon® Platinum 8280, 2.7GHz, 28 cores, turbo on, HT on, BIOS 0D010299, 192GB total memory, 12 slots / 16GB / 2666 MT/s / DDR4 LRDIMM, 1 x 375GB, Intel® Optane™ SSD DC P4800X, Red Hat Enterprise Linux® 7.5, kernel 3.10.0-862.el7.x86_64.

(intel)