PROWESS

Technical Research Report

Do Customers Benefit from Intel® In-Memory Analytics Accelerator (Intel® IAA), and at What Cost?

Testing by Prowess Consulting demonstrated high performance and less complexity when deploying systems with AMD EPYC™ processors for database workloads, compared to systems with Intel® Xeon® processors using Intel IAA accelerators.

Executive Summary

In testing by Prowess Consulting, an "out-of-box" configuration built with AMD EPYC™ processors offered high performance and efficiency for a database workload with far less configuration complexity than an Intel® processor–powered system using Intel® In-Memory Analytics Accelerator (Intel IAA®). For example, our engineers were unable to configure ClickHouse to make use of Intel IAA for testing despite hours spent trying to resolve numerous errors, compiler version issues, and other hurdles. No special configurations were required to benefit from the higher core count in the AMD EPYC–processor powered system. In addition, our testing with a RocksDB workload demonstrated higher performance and a higher performance per watt from the AMD processor–powered system, compared to an Intel processor–based system using four instances of Intel IAA. Finally, the AMD EPYC 9734 processor has a 43% lower manufacturer's suggested retail price (MSRP) than the Intel Xeon Platinum 8490H processor, making the AMD option a compelling prospect for businesses looking to achieve optimum efficiency and price/performance for their analytics and in-memory database (IMDB) workloads, without the added administrative burden required by Intel IAA.

Highlights

Compared to a system built on Intel® Xeon® Platinum 8490H processors, a system powered by AMD EPYC™ 9734 processors provides:

>>

1.65x better out-of-box performance*

2.13x better performance per watt, with Intel® IAA enabled on the Intel system*

43% lower cost

*Running dbbench against a RocksDB workload

The analysis in this document was done by Prowess Consulting and commissioned by AMD.

Introduction

Organizations are looking to achieve high performance for analytics and IMDB workloads while minimizing power usage, but there are different ways to achieve these goals. Businesses might use Intel Xeon processors with Intel IAA to speed up the processing of targeted workloads and to offload work from cores. In theory, this should result in highly optimized performance and reduced energy use. An alternative approach is to forego accelerators and instead use a larger number of efficient cores. Users can take this latter approach by deploying systems powered by AMD EPYC processors, which provide higher core and thread counts than comparable Intel Xeon processors.

Which approach is most efficient and cost-effective for organizations running analytics and IMDB workloads? And which option requires the least effort to deploy and configure? To find out, Prowess Consulting, sponsored by AMD, compared the performance and efficiency of Intel Xeon processors using Intel IAA versus AMD EPYC processors without accelerators but equipped with a higher core count.

Configurations and Testing

For our testing, we compared a system powered by two Intel Xeon 8940H CPUs to one powered by two AMD EPYC 9734 CPUs, as shown in Table 1. See the methodology report for full details.

Table 1 | Configurations used in testing

	Intel® Xeon® Platinum 8490H Processor–Based Configuration	AMD EPYC™ 9734 Processor–Based Configuration
CPU	Intel Xeon Platinum 8490H	AMD EPYC 9734
Number of CPUs	2	2
Cores/Threads per CPU	60/120	112/224
Cores/Threads Total	120/240	224/448
CPU Frequency	3.5 GHz	3.0 GHz
Installed Memory	1 TB	1.5 TB
Memory DIMM	64 GB Micron® Technology	64 GB Samsung®
Memory Speed	DDR5 (4,800 megatransfers per second [MT/s])	DDR5 (4,800 MT/s)
Number of Memory DIMMs	16	24
Operating System (OS)	Ubuntu® 22.04.3 LTS	Ubuntu 22.04.3 LTS
OS Kernel	5.15.0-91-generic	5.15.0-86-generic
Hardware Accelerator	Intel® In-Memory Analytics Accelerator (Intel IAA®)	Not applicable
MSRP[2]	$17,000	$9,600

As a starting point, we looked at three claims made by Intel regarding Intel IAA:

- 1.12x average performance per watt efficiency improvement for ClickHouse (Intel IAA vs. LZ4 compression)[3]
- 1.26x average performance per watt efficiency improvement for ClickHouse (Intel IAA vs. ZTD compression)[3]
- 2.01x average performance per watt efficiency improvement for RocksDB (Intel IAA vs. ZTD)[3]

We began by attempting to verify the claims regarding ClickHouse, a high-performance, column-oriented SQL database management system (DBMS) for online analytical processing (OLAP). To make use of Intel IAA with ClickHouse, administrators must first configure the Intel® Query Processing Library (Intel® QPL), an open-source library designed to provide high-performance query processing operations on Intel CPUs.

Unfortunately, we encountered several errors while attempting to configure Intel QPL using documentation provided by ClickHouse.[4] After numerous attempts, the log files indicated that ClickHouse was falling back to using software compression, rather than making use of Intel IAA hardware acceleration, as indicated by this error message:

```
DeflateQplJobHWPool: Initialization of hardware-assisted DeflateQpl codec failed, falling back to
software DeflateQpl codec. Failed to Initialize qpl job
```

Over the course of several days, our engineers attempted various build configuration flags, different compiler versions, multiple versions of ClickHouse software, and even multiple versions of the underlying QPL library. A quick search on GitHub revealed several threads from other users reporting issues with building the Intel QPL library, and a response from ClickHouse indicating that it does not officially support Intel QPL.[5] Without full configuration details or methodology from Intel, our engineers were unable to determine how Intel successfully configured ClickHouse to make use of Intel IAA acceleration.

Our research also noted that Intel's published claim regarding Intel IAA with ClickHouse provides results from only a single query. The Star Schema Benchmark (SSB) that Intel used for testing runs multiple queries (using simplified star schema data based on the TPC-H benchmark) against the database. A full list of results or a geomean score of all query results would be more useful than a single query to organizations assessing the performance benefits of Intel IAA for database workloads. We were more successful in configuring and testing with RocksDB, a high-performance, open-source storage engine used in common database applications. To align with Intel's performance claim, we used dbbench, a benchmarking tool distributed with RocksDB.

We ran dbbench against RocksDB with different read/write configurations and gradually increased the load on the system. As the workload increased, the core/thread counts needed to perform the work increased accordingly. We then measured performance by recording the sum of the total random read operations and the total random read/write operations. On the Intel system, we tested with zero, one, and the maximum of four Intel IAA accelerators.

Test Results

First, we compared performance for both systems using out-of-box configurations, without acceleration (see Figure 1). In this scenario, the AMD system performed 1.65x more operations, using the same number of cores as the Intel system without Intel IAA. Note that the Intel system is using all available cores (2 CPUs, with 60 cores per CPU), whereas the AMD system is using only 120 out of 224 total cores (2 CPUs, with 112 cores per CPU).

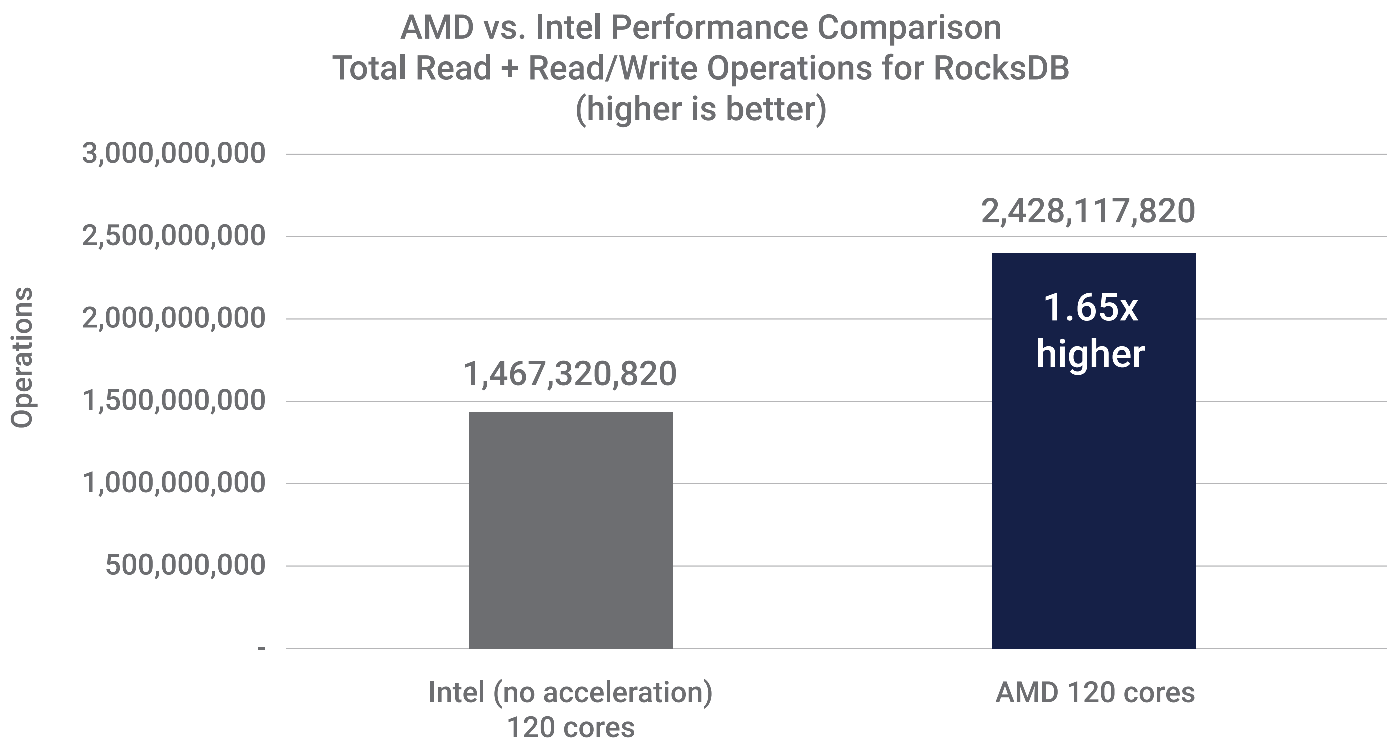

Figure 1 | RocksDB performance on a two-socket Intel® system (no acceleration enabled) with all 120 total cores compared to an AMD system using 120 out of 224 total cores

Next, we looked at performance with one and four accelerators (respectively) configured on the Intel system and compared this to the same AMD configuration with 120 cores (see Figure 2). Intel IAA boosted performance, as expected. However, as with the non-accelerated scenario, the Intel system core count is now maxed out, while the AMD system still has 104 cores available. Some or all of those extra cores can be applied to this workload to increase performance or can be used to support other workloads. This is an important consideration because organizations typically don't dedicate a server to only one workload.

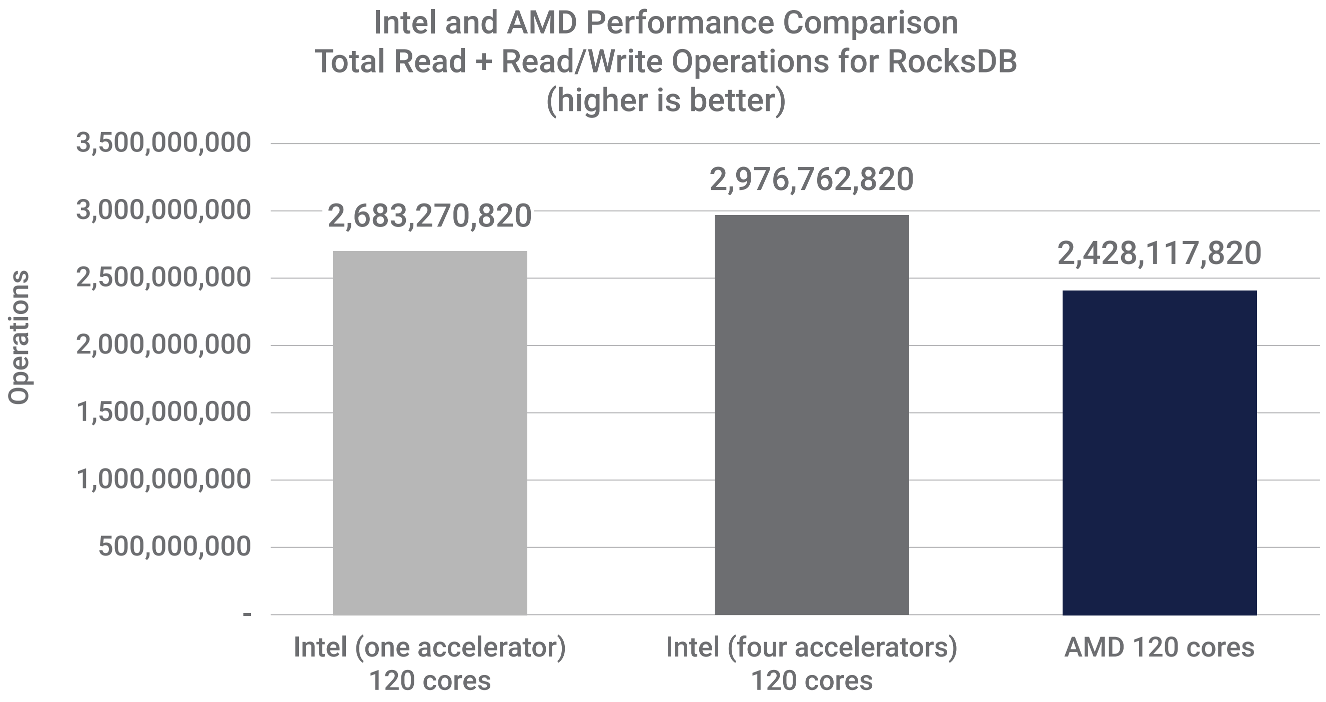

Figure 2 | RocksDB performance on a two-socket Intel® system with all 120 total cores using one and four Intel® IAA accelerators (respectively) compared to a two-socket AMD system using 120 out of 224 total cores

In addition to assessing performance, we looked at efficiency for the systems by measuring performance per watt. Note that this is a measurement of total system watts, rather than just CPU thermal design power (TDP). Figure 3 compares efficiency for the Intel system using all 120 cores, both with and without accelerators, to the AMD system using 120 out of 224 total cores.

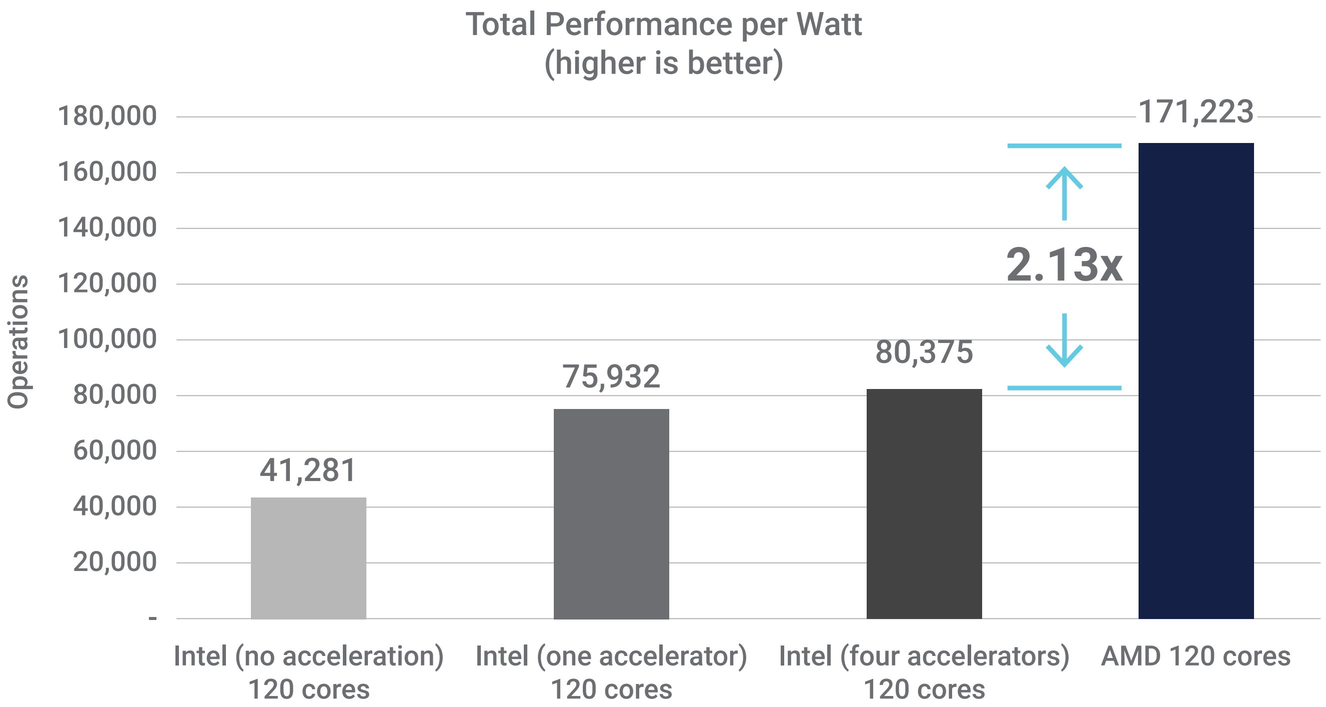

Figure 3 | Performance per watt for the two-socket Intel® system with all 120 total cores using zero, one, and four Intel® IAA accelerators (respectively) compared to the two-socket AMD system using 120 out of 224 total cores

This chart demonstrates two findings. First, enabling four Intel IAA accelerators on the Intel system demonstrated a performance-per-watt increase of 1.94x, compared to the non-accelerated Intel configuration. This result is close to Intel's stated claim of a 2.01x efficiency improvement.

More notably, however, the AMD system showed a performance-per-watt result 4.14x higher than the non-accelerated Intel instance and 2.13x higher than the Intel instance configured with four Intel IAA accelerators.

Efficiency for the AMD system increased when we increased the number of operations in the workload to make use of more cores. With 156 cores in use, the AMD system demonstrated 2.64x higher performance per watt compared to the Intel system using all available cores (120) and with four Intel IAA instances. Even with this increased load on the system, the two-socket AMD system still has 68 additional total cores available to boost performance or run other workloads.

Assessing the Test Results

According to Intel, Intel IAA accelerates database queries and optimizes performance by offloading compression, decompression, and other compute-intensive tasks from the CPU.[6] As our testing demonstrates, Intel IAA did help performance and efficiency for the RocksDB workload. However, while Intel IAA accelerates only specific functions, such as compression and decompression, AMD's higher core count offers a broader solution for accelerating workloads. By providing more overall cores, AMD EPYC processors offer the potential to enhance performance and efficiency gains across the board, beyond just compression/decompression algorithms. Our results offer strong evidence for this claim by demonstrating significant gains in RocksDB performance and performance per watt by increasing the core count without relying on discrete accelerators.

Organizations looking for the best value should also consider pricing. The AMD EPYC 9734 processors we tested offer nearly double the core count of the Intel Xeon Platinum 8490H processors, yet the AMD processors are priced $7,400 less than the Intel Xeon CPUs ($9,600 versus $17,000).[2]

In addition, the AMD system works essentially "out of the box" because no special configuration or programming is required to reap the performance and efficiency benefits we uncovered. To make use of Intel IAA, the applications must support the accelerators. In addition, plugins (like Intel QPL) might be required, contributing to the complexity and configuration overhead required by developers and IT admins.

Furthermore, although Intel IAA does offload some of the primary workload from the CPU, the Intel IAA process itself must be managed, resulting in some potential resource overhead on some of the cores on the Intel Xeon CPU.

Because of these factors, organizations considering server upgrades should evaluate the administrative burden and complexity inherent with configuring workloads to use Intel IAA. AMD systems offer a more plug-and-play experience for delivering high performance and greater efficiency for workloads. The higher core count in AMD systems also gives organizations the freedom to run multiple workloads on the same system while still meeting their performance needs. By performing more work on fewer servers, businesses can potentially lower their total cost of ownership (TCO).

Conclusion

Based on our testing, Intel IAA can provide some performance and efficiency benefits for analytics and IMDB workloads. However, organizations are likely to realize greater benefits with far lower complexity or administrative burden by simply using the higher core counts provided by AMD EPYC processors. In addition, the AMD EPYC CPUs we tested provide a significant price/performance advantage, coming in at a 43% lower cost, despite offering significantly more cores per CPU.

Learn More

For complete configurations and testing details, see the methodology report.

[1] Based on testing performed by Prowess Consulting, completed on April 19, 2024.

[2] Manufacturer's suggested retail pricing (MSRP) of $9,600 for AMD EPYC™ 9734 (112C/224T) obtained from "AMD Server Processor Specifications" on June 9, 2024. MSRP of $17,000 for Intel® Xeon® Platinum 8490H (60C/120T) obtained from "Intel® Xeon® Platinum 8490H Processor" on June 9, 2024.

[3] See [E1] at www.intel.com/performanceindex.

[4] ClickHouse. "Build Clickhouse with DEFLATE_QPL." Accessed April 19, 2024.

[5] GitHub. "QPL can no longer be compiled" and "Failed to build ClickHouse with QPL." https://github.com/ClickHouse/ClickHouse/issues/42982#issuecomment-1304841888.

[6] Intel. "Intel® In-Memory Analytics Accelerator (Intel® IAA)." Accessed June 4, 2024.