Aerospike

WHITE PAPER

Achieve price-performance gains for real-time workloads with Aerospike and AWS Graviton

Contents

Executive overview

Aerospike, Inc., the real-time data platform leader, demonstrated compelling price-performance results for Aerospike 6 running on Amazon Web Services (AWS) Graviton2 processors. As Fig. 1 shows, Aerospike observed 63% better price-performance using a Graviton2 cluster when compared with an equivalent x86 cluster. Higher transaction throughput rates and lower annual cluster costs drove this price-performance advantage.

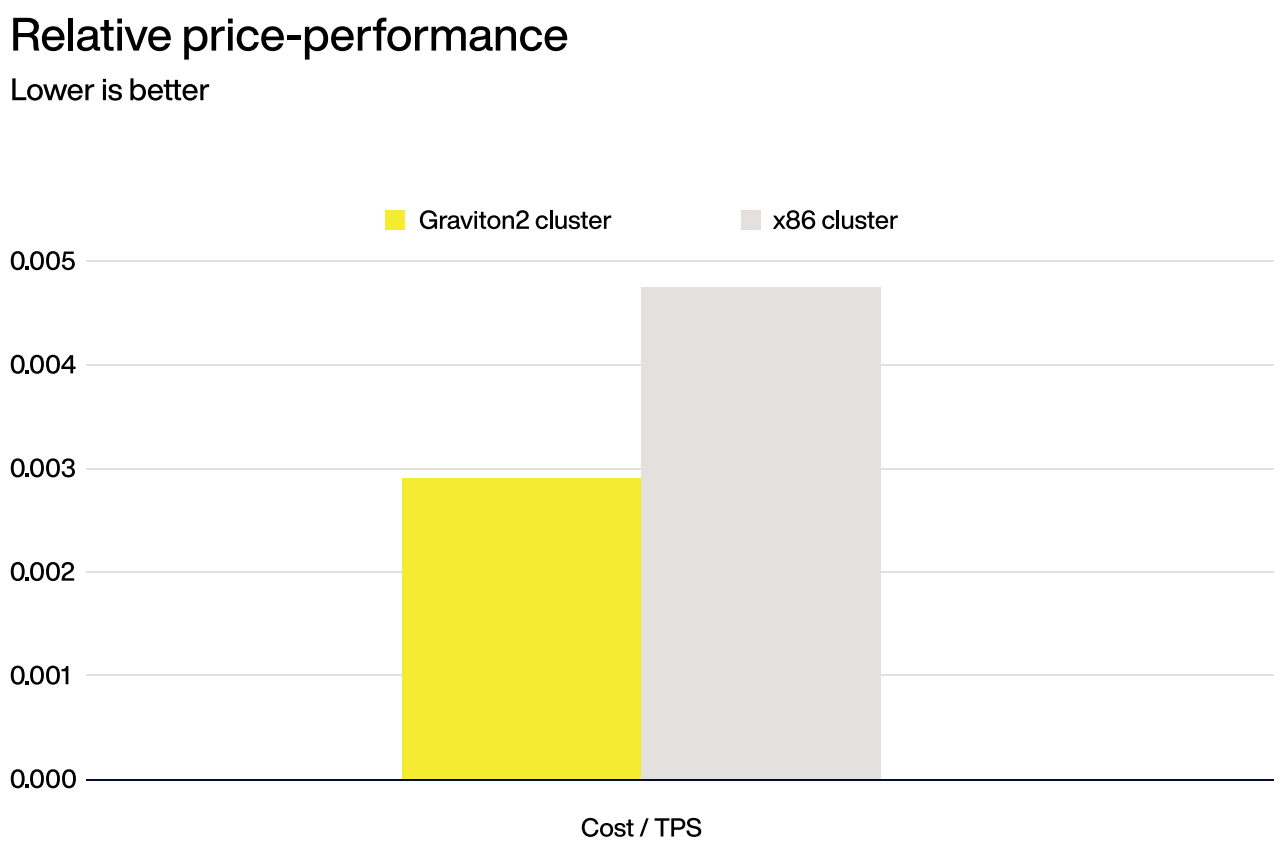


Figure 1: Price-performance (estimated annual cluster cost / number of transactions processed per second) was 63% better on the Graviton2 cluster

The performance test compared Graviton2 and x86 clusters with the same number of virtual CPUs running a read-only, real-time, typical Ad Tech customer workload on Aerospike. Both clusters completed 99% of all transactions in less than 1 millisecond. The Graviton2 cluster processed 25 million transactions per second (TPS), while the x86 cluster processed 21.1 million; this translated to an 18% higher throughput rate for the Graviton cluster, as shown in Figure 2. In addition, the annual cost for the Graviton2 cluster was 27% less, as this paper explains later.

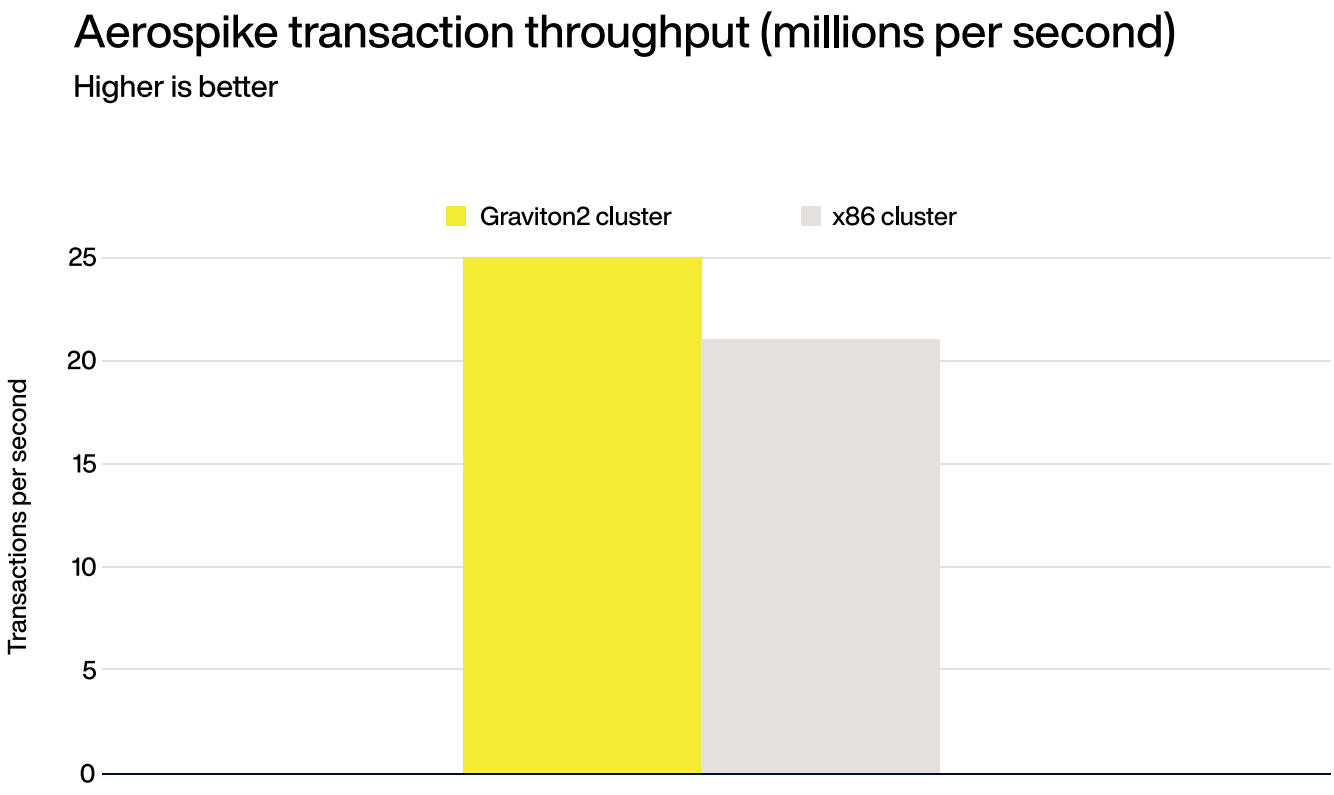


Figure 2: Transaction throughput (TPS rate) was 18% better on the Graviton2 cluster

Given previously published benchmarks that demonstrate Aerospike's ultra-fast and predictable performance at scale, this benchmark illustrates how cost-efficient and practical it can be to analyze, manage, and process high volumes of real-time data using Aerospike. The technology advances pioneered by Aerospike and AWS can also help firms cut their carbon footprints to support green goals. For example, AWS estimated that the Graviton environment used in this performance test cut carbon emissions by an estimated 49% over over the x86 environment while fulfilling the benchmark's aggressive transaction throughput and data access latency targets.

This paper presents benchmark results that help you compare the price-performance and carbon footprint of Aerospike on Graviton to x86 deployments. It also summarizes key features of Aerospike on Graviton to help you understand the technologies behind these results.

Is your current real-time data management infrastructure delivering comparable price-performance in a climate-friendly manner? If not, perhaps it's time to explore Aerospike on Graviton as an alternative.

Aerospike on AWS Graviton benchmark

To demonstrate the cost-efficiency for operational workloads on AWS Graviton2 processors, Aerospike benchmarked its server platform on two EC2 topologies: one using Graviton processors and another using x86 processors. The goal was to explore how Aerospike's leveraging of Graviton CPUs translates into tangible price-performance benefits by conducting a side-by-side comparison. The results were revealing.

Aerospike extracts considerable efficiencies from processor, storage, and networking improvements from its hardware partners. For the purposes of this benchmark, however, we focused specifically on CPU performance. As this benchmark

demonstrated, Aerospike with AWS Graviton delivered an impressive 63% improvement in price-performance. AWS also estimated that the Graviton cluster cut carbon emissions 49% per transaction per second compared with the x86 alternative. Both clusters were tasked with delivering their maximum TPS with the 99 percentile of client-side data access latencies of less than 1 ms. Furthermore, each cluster had the same number of vCPUs, 192.

Workload and instances

Aerospike, in conjunction with AWS, ran a CPU-intensive workload with 300 asbench[1] processes connecting to Aerospike database version 6.2. Each of the databases contained 2 billion unique records. Benchmark clients methodically ramped up the number of transactions executed on the clusters and executed more than 20 million read-only transactions per second. While the TPS was increased, each cluster was monitored to determine the point at which the 99 percentile latencies for those transactions exceeded 1ms. This recorded "TPS under the 1ms SLA" was used for comparison of the two clusters.

Each cluster was run in a single AWS availability zone within US East. And both clusters contained the same total number of vCPUs. The Graviton2 cluster consisted of 3 `c6gn.16xlarge` nodes that each contained 64 vCPUs, 128 GiB memory, and a network bandwidth of 100 Gbps. The non-Graviton cluster consisted of 2 `m5n.24xlarge` nodes that each contained 96 vCPUs, 384 GiB memory, and a network. Note that the tests were conducted entirely in-memory, so any unused memory for the same data size would not affect the processing, only the vCPUs (which were held constant).

On each cluster, Aerospike was configured for in-memory storage (i.e., to retain user data and index data in memory). This is one of several Aerospike deployment options and the one best suited to create CPU-heavy workloads, as the focus of this test was on the CPU processing capabilities.

Each Aerospike system used a replication with a factor of 2, which provides high data availability in most failure scenarios and is often used in production Aerospike environments, but the read only workload means that replication factor does not influence the TPS.

Results

The 3-node Graviton2 cluster averaged 25 million TPS with 99% of those transactions completing within the 1ms SLA, while the 2-node x86 based cluster averaged 21.1 million TPS for the same workload with 99% of those transactions completing within the 1ms SLA.

Using Amazon's online pricing calculator and other publicly available data, Aerospike and AWS sought to arrive at a reasonable cost comparison of the two environments. To do so, we considered the hourly cost per node used in each cluster based at prevailing US East rates using the 1-year upfront Linux on-demand pricing structure. For each Graviton2 node, this was $2.7648 per hour. For each x86 node, this was $5.7120 per hour. Assuming round-the-clock daily usage for each cluster, this resulted in an estimated annual cost of $72,659 for the Graviton2 cluster and $100,074 for the x86 cluster. To quantify the price-performance of each cluster, the estimated annual cost was divided by the transaction throughput rate each supported. This yielded the following results:

- The Graviton2 cluster reduced annual costs by an estimated 27%.
- The Graviton2 cluster delivered an 18% higher TPS rate under 1ms SLA.
- The relative price-performance of the Graviton2 cluster was 1.63x better than the x86 cluster (i.e., the Graviton2 cluster improved price-performance by 63%).

For details on these calculations, see Appendix A.

[1] asbench is an open source tool to benchmark the Aerospike Database.

The carbon footprints of each approach are noteworthy, too. AWS estimated emissions produced by each cluster based on the instance types, workload, and performance results associated with this benchmark and determined that Aerospike saved 49%, confirmed by AWS, on carbon emissions when running its workload on the 3-node Graviton2 cluster (using C6gn.16xlarge instances) rather than the 2-node non-Graviton cluster (using the M5n.24xlarge instances).

Technology leadership

To help you better understand the technologies behind these benchmark results, this section introduces you to Aerospike's architecture and how it exploits advanced AWS technologies to deliver exceptional price-performance and other advantages.

Aerospike architecture

Aerospike is a multi-model, real-time data platform that supports multi-cloud, large-scale JSON, and SQL use cases. It offers ultra-low data access latencies with predictable performance at any scale, provides exceptional uptime, and requires up to 80% less infrastructure than alternative solutions. It achieves its high scalability and exceptional price-performance in part by exploiting modern hardware, including non-volatile memory extended (NVMe) Flash Drives, and now AWS Graviton processors.

Such optimizations, coupled with Aerospike's multi-threaded architecture and other features, provide a distinct price-performance advantage, having saved production users $1 to $10 million per application in total cost of ownership (TCO) over other solutions. (For one cost savings example, see this global brokerage firm profile.) Aerospike's architecture also enables the system to automatically distribute data evenly across its shared-nothing clusters, dynamically rebalance workloads, and accommodate software upgrades and most cluster changes without downtime. On AWS, Aerospike exploits caching on ephemeral devices, backing up data on Elastic Block Store (EBS) volumes.

Aerospike's flexible storage configuration options and its Hybrid Memory Architecture™ enable administrators to choose where best to keep indexes and user data. As mentioned earlier, in this benchmark, Aerospike kept all index and user data in memory (DRAM)— a configuration common for real-time fraud detection and online bidding applications in the AdTech industry. Other configuration options include storing all index and user data on Flash (SSDs) or in hybrid configurations with indexes in DRAM and user data on Flash. Aerospike's Hybrid Memory Architecture is further enhanced on AWS Nitro SSDs due to Nitro's predictable performance and ability to handle higher throughputs coupled with their low variance for low latency, high I/O use cases.

Such flexibility enables firms to use Aerospike in different ways to support different business needs without incurring excess infrastructure costs or compromising application requirements. It also helps firms standardize on a common database platform at the edge and at the core to simplify operations and reduce overall expenses. For further details, see The Aerospike Difference or Introducing Aerospike's architecture.

Aerospike and AWS

Aerospike partners with key cloud and hardware vendors to ensure its platform can leverage new technologies as they emerge. With Amazon, this includes exploitation of Graviton processors.

Based on Arm architecture, AWS Graviton processors feature custom silicon and 64-bit Neoverse cores, delivering lower power consumption, stronger price-performance, lower latencies, and better scalability than other alternatives. Well suited for high performance computing, machine learning, in-memory caches, and other applications, Graviton is a natural fit for Aerospike customers seeking to maximize cost efficiency without compromising on aggressive SLAs or inhibiting future business growth.

Although not showcased in this benchmark, Aerospike also leverages Amazon's latest Nitro SSD technology (im4gn and

is4gen), which can deliver up to 60% lower latencies and up to 75% lower latency variability than AWS i3 and i3en instances. For applications better suited to an all-SSD or hybrid configuration of Aerospike (with indexes in DRAM and user data on SSDs), Aerospike's ability to efficiently use Nitro SSD technology provides added performance and cost benefits. For more details, see this 2021 presentation from AWS and Aerospike.

Finally, energy-conscious firms may find that running Aerospike on AWS can cut carbon emissions significantly compared with other alternatives. Indeed, a recent IEEE paper that explored infrastructure and energy costs of Aerospike and Cassandra deployed on AWS calculated that Aerospike's software efficiencies can lower costs and carbon emissions by 80%. Furthermore, just moving from an on-premises infrastructure to a cloud infrastructure can result in substantial energy savings. By one estimation, an AWS infrastructure is 3.6x more energy efficient than the median of US enterprise data centers.

Where other approaches fall short

Cost-effective management of operational data places incredible demands on database infrastructures and IT organizations. Performance, operational ease, elasticity, availability, data consistency, enterprise integration, and cost efficiency are common—and vexing—pressure points.

Many open source and commercial solutions simply can't manage high-volume mixed workloads without critical shortcomings surfacing in one or more essential areas. For example, relational DBMSs often integrate well with other software and provide strong data consistency guarantees but can't deliver ultra-fast performance at scale with low total cost of ownership (TCO). Certain open source and commercial NoSQL systems offer faster, less expensive alternatives than relational DBMSs but suffer from operational complexity, unpredictable performance, and sprawling server footprints as databases grow. Traditional caching systems may offer initial relief but often exhibit erratic latencies at terabyte scale (and beyond), introduce additional application and operational complexities, and drive up TCO.

Summary

The latest benchmark from Aerospike and AWS set a new bar for price-performance for real-time workloads. Aerospike on AWS Graviton2 processors delivered 63% better price-performance compared with x86 environments while processing 21 - 25 million read transactions per second (TPS); 99% of these transactions completed in less than 1 ms. Furthermore, running Aerospike on Graviton can even result in a substantial reduction in carbon emissions compared with other alternatives. For this benchmark scenario, Aerospike saved significantly on carbon emissions by running the workload on the Graviton cluster rather than the x86 cluster.

If your real-time data management infrastructure can't deliver comparable price-performance in a climate-friendly manner, perhaps it's time to explore the Aerospike alternative. Visit Aerospike's web site or contact Aerospike directly to schedule a briefing.

Appendix A: Calculating price-performance

Price-performance data cited earlier in this paper are based on calculations included in this appendix. Table 1 offers a comparative summary.

	Graviton2 cluster (C6gn.16xlarge)	x86 cluster (M5n.24xlarge)	Notes
Hourly cost per node	$2.7648	$5.7120	
Number of nodes	3	2	192 vCPUs per cluster (64 per Graviton2 node, 96 per x86 node)
Annual cost (est'd)	$72,659	$100,074	27% lower cost with Graviton2
TPS @ 99% < 1ms	25,000,000	21,100,000	18% higher TPS with Graviton2
Price-perf ratio (Lower is better)	.00291	.00474	63% better price-perf with Graviton2

Table 1: Comparative data for Graviton2 and x86 clusters

Let's step through the calculations leading to the data shown in the final three rows of Table 1. The annual cost for each cluster was calculated as follows:

```
Annual cost = Hourly rate per node * 24 hours * 365 days * # of nodes

Graviton cluster annual costs = $72,659
x86 cluster annual costs = $100,074
Cost differential = 27% lower cost with Graviton cluster
```

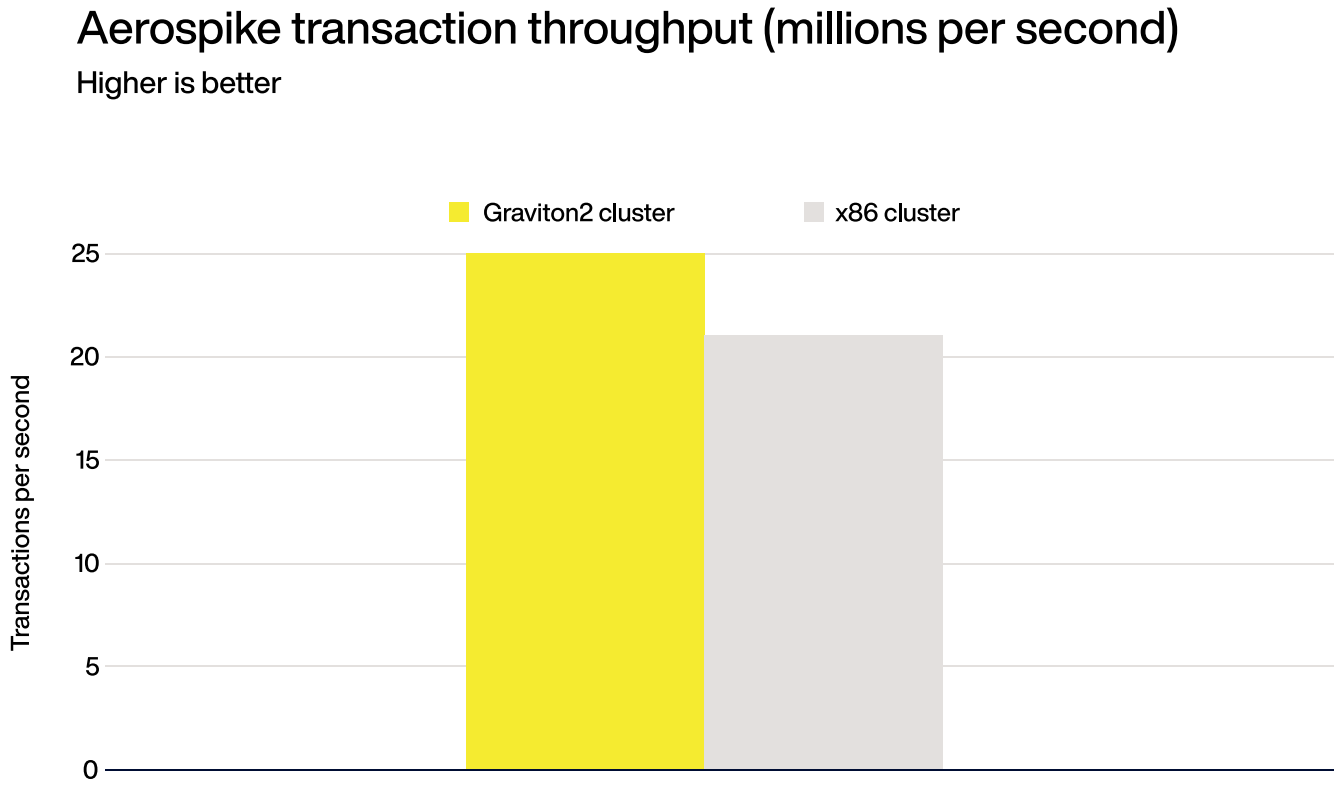
To derive price-performance ratios, the annual cost of each cluster was divided by its throughput rate (TPS). A lower ratio is better, as it indicates a lower cost per transaction:

```
Price-performance ratio = annual cost / TPS

Graviton cluster TPS (throughput) = 25,000,000
Graviton cluster annual cost = $72,659
Graviton price-performance ratio = .00291

x86 cluster TPS (throughput) = 21,100,000
x86 cluster annual costs = $100,074
x86 price-performance ratio = .00474
```

Finally, to compare the relative price-performance of the two clusters, the price-performance ratio of each cluster was divided by the price-performance ratio of the cluster with the lowest ratio (the Graviton cluster, in this case). Again, lower is better.

```
Graviton relative price-performance = .00291 / .00291 = 1
x86 relative price-performance = .00474 / .00291 = 1.63
```

This final calculation indicates **63% better price-performance with the Graviton2 cluster.**



About Aerospike

Aerospike is the real-time database built for infinite scale, speed, and savings. Our customers are ready for what's next with the lowest latency and the highest throughput data platform. Cloud and AI-forward, we empower leading organizations like Adobe, Airtel, Criteo, DBS Bank, Experian, PayPal, Snap, and Sony Interactive Entertainment. Headquartered in Mountain View, California, our offices include London, Bangalore, and Tel Aviv.

For more information, please visit https://www.aerospike.com.