Benchmark

Running Operational Workloads with Aerospike at Petabyte Scale in the Cloud on 20 Nodes

An Aerospike white paper developed in collaboration with Intel and Amazon Web Services

August 2021

Contents

Executive overview

Massive database scalability without massive server footprints? High availability without high maintenance? Exceptional performance without constant tuning? And total ownership costs unrivaled by alternate approaches?

It may be hard to believe that such technology exists, but a recently published petabyte benchmark and experiences from leading firms in finance, retail, telecommunications, advertising technology (AdTech), and other industries offer clear proof. Faced with growing data volumes, demanding read/write workloads and budgetary pressures, IT leaders around the globe rely on Aerospike to solve multiple operational database problems simultaneously that other solutions can't:

- Act instantly across billions of transactions while reducing server footprint up to 80% often saving $1-$10 million per application.
- Power real-time applications with predictable sub-millisecond performance up to petabyte scale.
- Maintain five-nines uptime with globally-distributed, strongly consistent data even during common failure scenarios and system upgrades.
- Deploy quickly on premises, in multi-cloud, or in hybrid configurations to support applications at the edge and at the core of enterprise architectures.

Achieving any of these goals can be tough. Achieving *all* may seem downright impossible. Indeed, many Aerospike users had previously deployed commercial or open-source NoSQL systems, commercial or open-source SQL DBMSs, multi-layer cache-based architectures, and even homegrown DBMSs to try to meet such goals. But each fell short in some way. Excessive operational costs, sprawling server footprints, unpredictable performance at scale, data consistency problems, and availability issues often violated service level agreements (SLAs) or hindered new business opportunities. That's when they turned to Aerospike, a distributed NoSQL DBMS platform that exploits the latest processor, storage, and networking technologies to support mission critical applications.

Time and again, Aerospike has delivered remarkable results. For example, Signal, a TransUnion company, reduced its server count from 450 to 60 nodes and expects to save $7 million in TCO over three years by migrating to Aerospike. ThreatMetrix cut its server footprint more than 70% and realized a four-fold performance improvement that was critical to its fraud prevention efforts. PayPal expects to save $9 million in hardware costs, cut TCO by 80% over 5 years, and enjoy better query response and data load times after replacing its in-memory DBMS with

Aerospike. Finally, a leading global brokerage firm offloaded work from its mainframe DBMS and replaced its caching platform with Aerospike to improve data access speeds while more than tripling its database size. Aerospike enabled the firm to deploy 90% fewer servers than originally anticipated, saving an estimated $10,000 per trading day.
Still not convinced? Consider the results of the Aerospike petabyte-scale benchmark with Intel and Amazon Web Services (AWS) that showed the platform delivering 4 - 5 million transactions per second (TPS) for read-only and mixed workloads with sub-millisecond latencies on a 20-node AWS cluster. Compared with alternative approaches (such as Apache Cassandra), Aerospike's footprint is remarkably small, driving substantial savings in TCO. As Figure 1 depicts, the 3-year TCO for running such a cluster in production is a modest $4.17 million, even with 15% annual growth in data volumes. By contrast, an Apache Cassandra-based architecture would require an estimated $15.29 million over the same 3-year period. That's $11 million in added expenses over the Aerospike implementation.

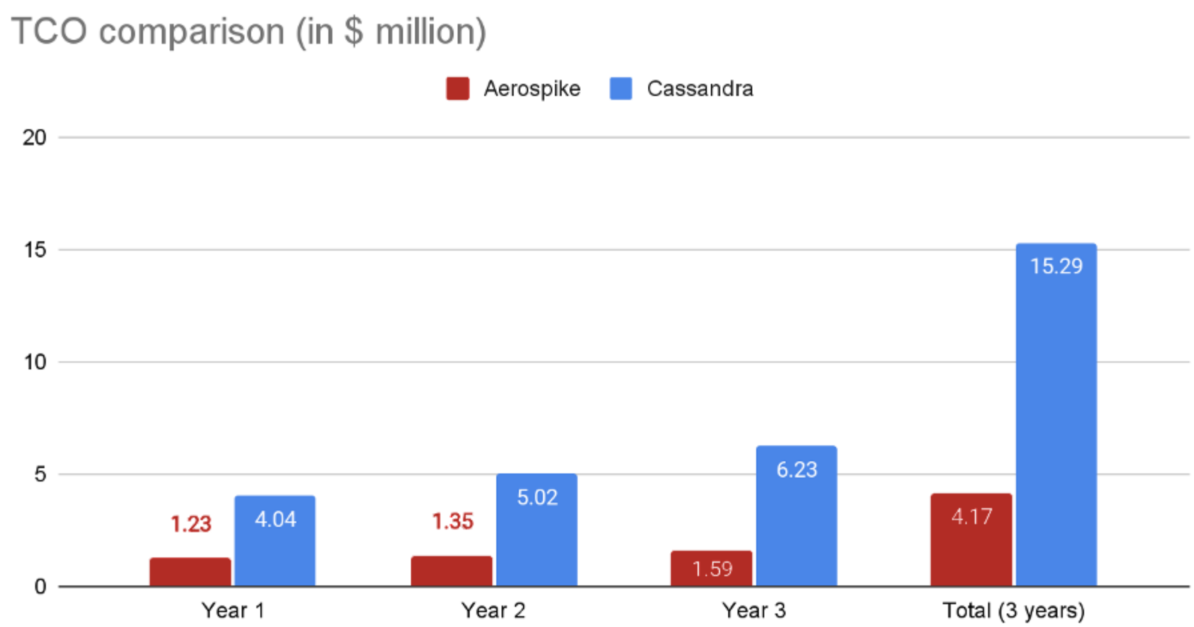

Figure 1: TCO for Aerospike, Cassandra over 3-year period for petabyte-scale workload

The sprawling server footprints required by other platforms to manage petabyte-scale operational data in real time drives up costs, even if the software is free. That's because infrastructure, maintenance, and operational expenses grow significantly when these platforms require hundreds (or even thousands) of nodes. Consider Figure 2, which compares the Aerospike and Cassandra server footprints for the petabyte-scale workload discussed later in this paper. Assuming a 15% annual growth rate in data volumes, Aerospike's footprint would grow from 20 to 26 nodes over a 3-year period. Compare that to Cassandra's, which would grow from 514 to 792 nodes. Which would you rather manage?

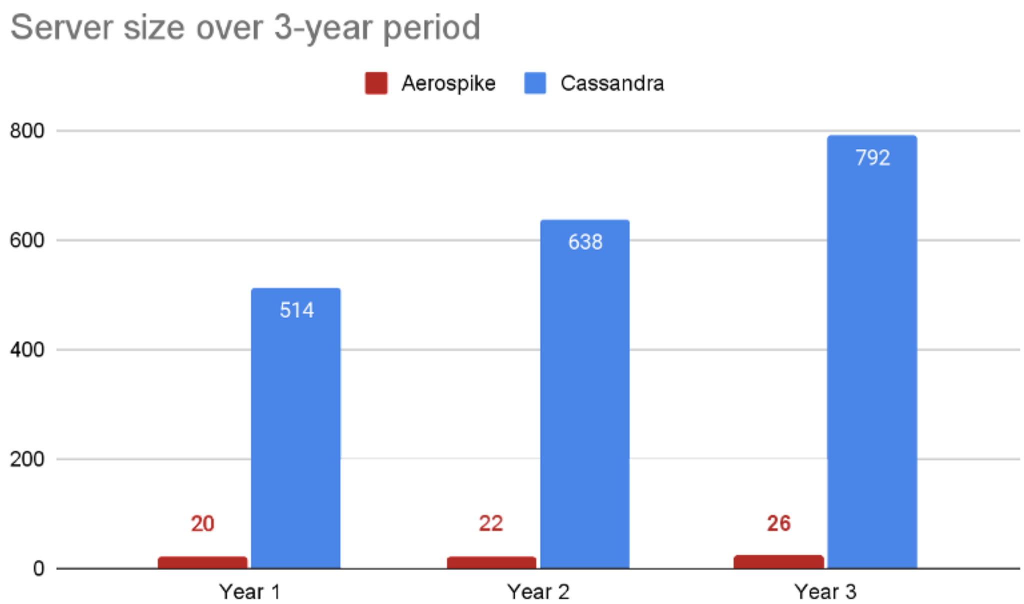

Figure 2: Server footprint for Aerospike, Cassandra over 3-year period for petabyte-scale workload

These benchmark results and TCO projections demonstrate what many Aerospike users already know: real-time petabyte-scale processing can be practical and affordable for most enterprises, not only those with $10+ billion in annual revenues and an abundance of sophisticated IT specialists.

This paper walks you through Aerospike's benchmark, a comparative TCO analysis, and the experiences of several Aerospike clients to help you explore the tangible benefits that Aerospike delivers every day, including a reduced server footprint that cuts costs and makes petabyte-scale processing practical. We'll start by considering real-world scalability challenges and how Aerospike addresses them.

Scalability in practice

Managing petabytes 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 scale up to petabyte-level processing for 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 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 to hundreds of terabytes or petabytes. Traditional caching systems may offer initial relief (e.g. up to 5 terabytes) but often exhibit erratic latencies at scale (e.g. 5+ terabytes), introduce additional application and operational complexities, and drive up TCO.

Could you be outgrowing your data management infrastructure? Ask yourself if . . .

- Your server footprint keeps expanding and you're worried about TCO.
- You're finding it harder and harder to meet SLAs.
- You struggle to manage peak or seasonal workloads.
- System stability and data availability are getting tougher – and more costly – to achieve.
- You can't afford to build new applications because your staff is too busy maintaining the status quo.
- You want greater standardization across your data architecture to simplify operations and minimize staff training.
- You want flexibility to deploy your database platform in the cloud, on premises, or in hybrid configurations, and you need to be able to configure each deployment as you see fit, leveraging advanced technologies such solid state drives (SSDs), Intel Optane persistent memory (PMem), data compression, and more.
- You're tired of compromising on data consistency to keep your data access speeds low.
- You're tired of constant tuning and dread the operational impact of software upgrades.
- The cost of onboarding a new customer and their data is less than the revenue received from the customer.

If you found yourself agreeing with several of these sentiments, you're not alone. That's how many current Aerospike customers became convinced it was time to evaluate their options. And that led them to Aerospike. Here's what a few are saying now:

". . . Very soon we reached a point that Redis couldn't scale any more. We moved to Aerospike. Aerospike is developer friendly. The scale problem was the first to be solved."

– Gil Shoshan, Software Engineer, ironSource

"Before Aerospike, we were spending more and more of our time on the care and feeding of Cassandra, and less and less time on the building of new product offerings. With Aerospike, we've now cleared the roadmap and we're just focused on adding new functionality to our platform for our customers."

– Jason Yanowitz, Executive Vice President, Chief Technology Officer, Signal

"With our previous database [Cassandra], we were operating at a request duration that would hurt our customers because they couldn't make a risk decision in the allocated time window... Aerospike helps us...to return in almost all cases with a very low latency so that all risk decisions can be made."

– Matthias Baumhof, VP Worldwide Engineering, ThreatMetrix

"With our past database [MongoDB], whenever there was a search in concurrent price updates from many services, we saw degradation in the buyer experience. Now with Aerospike, we can push through huge price changes while maintaining the same response time experience on the buyer's side – even with millions of buyers."

– Vice President of Engineering, Snapdeal

"Prior to Aerospike, we were using another in-memory data store and . . . running into challenges in terms of the cost of scaling...We moved to Aerospike for its hybrid memory architecture to leverage next generation memory and SSDs to their fullest advantage."

– Sai Devabhaktuni, Sr. Director of Engineering, PayPal

"Aerospike has done the unthinkable: they cut our server footprint by a factor of six while boosting our performance 300%,...saving us over $1 million a year.

– Guy Almog, Head of IT Engineering, Playtika

So how's that possible? Deployed to support systems of engagement or systems of record, Aerospike delivers exceptional scalability and runtime performance with low TCO because it is optimized to exploit modern hardware, including multi-core processors with non-uniform memory access (NUMA), non-volatile memory extended (NVMe) Flash drives, persistent memory (PMem), network application device queues (ADQ), and more. Such optimizations, coupled with Aerospike's multi-threaded architecture and other features, provide firms with distinct advantages for both on premises and cloud deployments. For example, Aerospike automatically distributes data evenly across its shared-nothing clusters, dynamically rebalances workloads, and accommodates software upgrades and most cluster changes without downtime. In the cloud (in this case, on AWS), Aerospike exploits caching on ephemeral devices, backing up data on Elastic Block Store (EBS) volumes.

Furthermore, Aerospike's flexible configuration options enable administrators to choose where to keep indexes and user data — all in traditional memory (DRAM), all in PMem, all on Flash (SSDs), or in hybrid configurations with indexes in DRAM or PMem and user data on Flash. Figure 3 illustrates a few possible configurations. Such flexibility enables firms to use Aerospike in different ways to support different business needs without incurring excess hardware costs or compromising on 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 [Introducing Aerospike's architecture.]

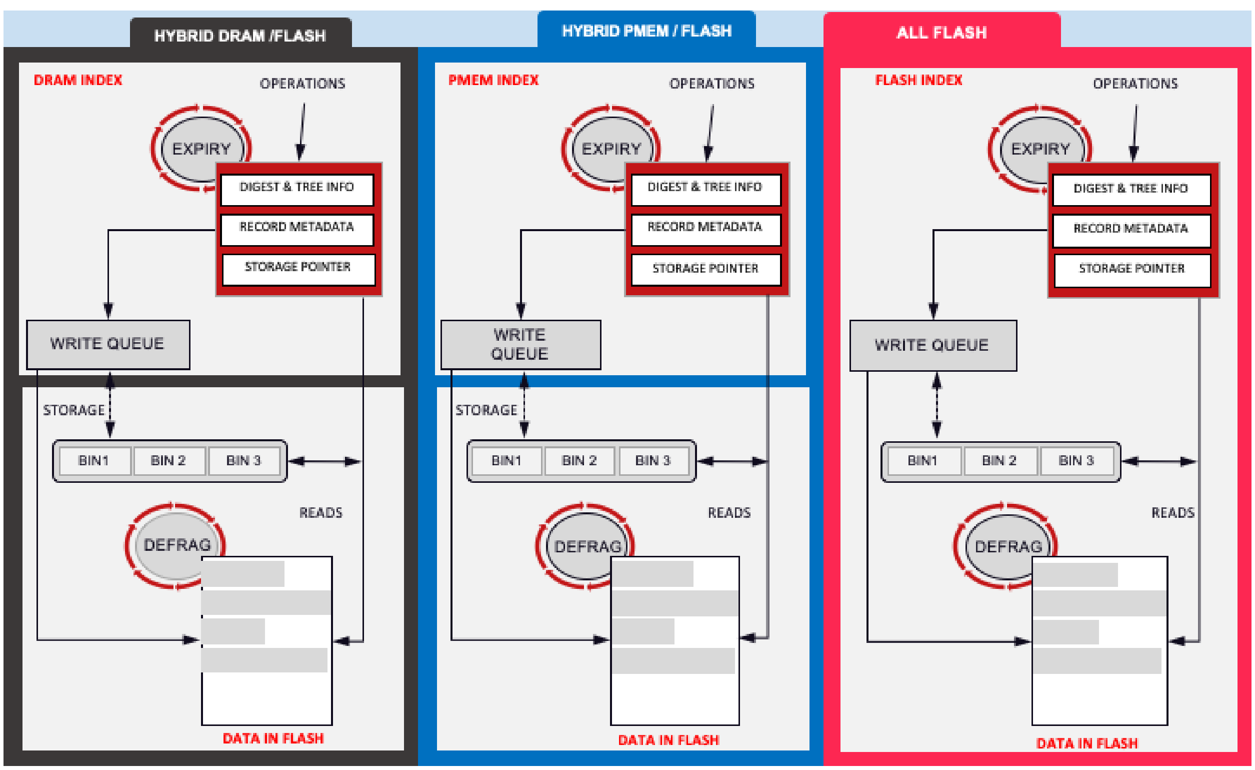

Figure 3: Three sample Aerospike configurations

Petabyte benchmark

Confident in its ability to deliver cost-efficient scalability and high performance for operational workloads, Aerospike put itself to a test, benchmarking its server platform on a 20-node AWS cluster managing 1 PB of unique user data. To date, no other NoSQL platform has openly conducted similar tests. The results prove that petabyte-scale processing is no longer the sole domain of deep-pocketed, technically elite firms, such as Facebook or Google. With Aerospike, petabyte-scale processing is now practical and affordable for a wide range of organizations.

This benchmark showed Aerospike delivering more than 5 million TPS for read-only operations and 3.7 million TPS for an 80/20 mix of read/write operations with server-side latencies of less than 1 ms in nearly all cases. Furthermore, during these benchmark runs, the same Aerospike cluster was managing a smaller system of record, concurrently processing almost 200,000 TPS with sub-millisecond latencies for a balanced mix of read/write operations against 1.5 TB of unique operational data.

In a moment, we'll delve into the details and results of the benchmark, beginning with a description of its workload. However, the TCO aspects of this benchmark are quite noteworthy: revealing that **Aerospike lowered costs by more than $11 million over a 3-year period compared with an anticipated Cassandra deployment**.

Let's review the Aerospike benchmark and its results, and then we'll walk through the TCO analysis.

Workload description

The workload for Aerospike's petabyte-scale benchmark was based on real-time application requirements common in many industries, including financial services, telecommunications, retail, AdTech, and others. From a database platform perspective, the requirements of such applications are extreme, calling for ultra-low data access latencies, 24x7 availability, high levels of scalability, and operational ease with low TCO.

The specific data and applications used in this benchmark were derived from the AdTech industry, where Aerospike usage has become the *de facto* standard for firms seeking to optimize the selection and delivery of online ads to target audiences through a real-time bidding (digital auction) process. With tens of billions of dollars of advertising revenues on the line, AdTech applications are technically sophisticated and demanding. Furthermore, competition among AdTech firms is fierce, with ever-increasing demands for faster speeds, lower costs, more information, and non-stop global services. Perhaps that sounds familiar. Indeed, firms in many industries are facing such pressures, which is why Aerospike developed its benchmark workload to reflect such demands.

User profile and campaign data

For readers unfamiliar with the AdTech industry, this section briefly introduces concepts related to user profile and campaign data, as Aerospike's benchmark involved both. User profile stores (which until recently had been mainly generated by storing cookie information) are maintained by AdTech demand-side platforms to help advertisers determine how much (if anything) they wish to bid on an available online advertising space. In effect, such stores contain anonymous user profiles modeling such data as an internal ID, market segments, geographic location, and other information gleaned from multiple sources, including various user devices (mobile phone, tablet, laptop), external data management platforms, etc. As you might imagine, data volumes for user profiles can grow quite large, often spanning several PBs or more. This user profile data is crucial to real-time ad purchasing decisions and must be accessible very quickly, as a demand-side platform may receive 10 million or more bid requests per second. A bid response is required in about 75 milliseconds and the auction is complete (win or lose) in about 100 milliseconds.

In a mixed workload, a separate namespace was established for campaign data simultaneously in the same database. Campaign data contains critical transactional information, including budgets and bidding data for digital ads, this data is managed in a system of record. Furthermore, campaign data more frequently involves a balanced mix of read/write operations, while user profile applications are read-centric.

Aerospike factored these varied requirements into its benchmark, which modeled user profile and campaign data in separate databases (namespaces) on a single Aerospike 20-node cluster. Aerospike wanted to evaluate how a real-world mix of requirements could be satisfied with one modestly sized cloud-based server. Aerospike's unique memory and storage architecture, data consistency support, compression options, and other features enabled each database to be tailored to fit the differing operational needs associated with different applications.

Benchmark operations and configuration overview

Aerospike's benchmark involved read-only and an 80/20 mix of read/write operations against a user profile database populated with 500 billion unique records of varying sizes to yield 1 PB of uncompressed user data. With replication, the total uncompressed size was 2 PB. In addition, Aerospike also benchmarked a 50/50 mix of read/write operations against a campaign database with 6 billion unique records totaling 1.5 TB uncompressed; with replication, the total uncompressed size was 3 TB. Workloads against both databases were run concurrently, as is typical in production scenarios.

The user profile and campaign databases shared these characteristics: both used speed-optimized Lz4 compression with a ratio of 4, and both used replication with a factor of 2. Compression reduces storage requirements, promoting cost efficiency and smaller server footprints. Cost-sensitive firms that can tolerate some added transactional overhead often rely on compression to reduce storage and operational costs. A replication factor of 2 promotes data availability in most failure scenarios and is often used in production Aerospike environments.

However, the user profile and campaign databases differed in other aspects to provide the most efficient, cost-effective support for the workloads unique to each. The user profile database kept all index and user data on Flash drives (SSDs) to minimize costs for the high data volumes involved. By contrast, the campaign database used a hybrid memory configuration with indexes in DRAM and user data on Flash to more evenly balance read/write throughout for this system of record. Furthermore, the user profile store employed the default consistency mode of AP (availability / partition tolerant), which is typical for analytical systems and systems of engagement. Since the campaign database was modeled as a system of record, it operated in SC (strong consistency) mode.

System configuration

As shown in Fig. 4, all benchmark tests were run on a 20-node AWS cluster with Aerospike Enterprise Server 5.4.0.2-1. Each EC2 i3en.24xlarge node with Intel® Xeon® Scalable processors featured 768 GB of DRAM and 8 x 7500 GB NVMe SSDs. For clients, the benchmark employed 40 Amazon EC2 c5n.9xlarge nodes using Intel® Xeon® Scalable processors with 96 GB DRAM each and EBS-only storage; operations were executed using Aerospike's C client.

Note that Aerospike selected the best storage-optimized instances currently available on Amazon EC2 to showcase the fewest number of servers for Aerospike Server nodes. All-Flash configuration does not need a lot of DRAM as the indexes and data are all stored in Flash. That said, I3en offers the highest storage density on EC2 and is good fit for this benchmark even we do not necessarily use all the DRAM available with I3en instances.

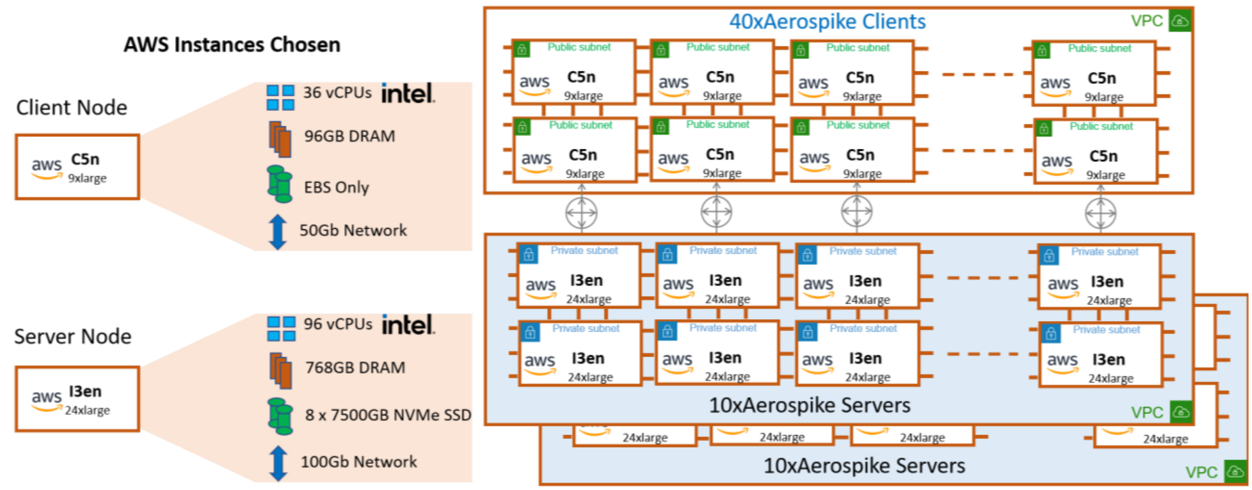

Figure 4: Overview of Aerospike's petabyte benchmark configuration. *(Note: Architecture can be readily configured to span multiple AZ's for additional resilience.)*

The Aerospike server was configured with compression that yielded a 4 times reduction in size for the user profile database, causing it to store 500 TB of compressed user data (250 TB of unique user data and 250 TB of replicated data). The campaign database also used a compression factor of 4, so it managed 750 GB of compressed user data (375 GB of unique user data and 375 GB of replicated data). Further details about the benchmark configuration are available in Appendix A.

Throughput and latencies

Aerospike loaded all necessary data and primed the environment by running a generic read/write workload for several days. This ensured that storage and other database resources were suitably randomized and that background jobs (such as defragmentation) would be triggered as they would in a production environment. Without priming, Aerospike might perform much better due to the localization effects of the initial workload.

Aerospike ran two sets of tests using the workloads described earlier to measure transaction throughput with sub-millisecond server-side latencies. Each test ran for 4 hours and featured concurrent operations spanning all keys in the user profile and campaign databases. Table 1 summarizes the results.

Test #	Data	Workload	TPS	Latency < 1 ms

1	User profile (1 PB unique, uncompressed)	Read only	5,009,980 reads	100%
1	Campaign (1.5 TB unique, uncompressed)	50/50 read/write	95,420 reads 95,420 writes	100% 100%
2	User profile (1 PB unique, uncompressed)	80/20 read/write	3,017,340 reads 754,160 writes	100% 99%
2	Campaign (1.5 TB unique, uncompressed)	50/50 read/write	95,800 reads 95,800 writes	100% 99%

Table 1: Benchmark results. User profile and campaign workloads ran concurrently.

As you can see from the Test 1 results, Aerospike processed more than 5 million read-only TPS with sub-millisecond latencies for user profile applications while also processing nearly 200,000 read/write TPS with sub-millisecond latencies for campaign applications. Test 2 featured an 80/20 mix of read/write mix operations run against the user profile database and a 50/50 read/write mix of operations run concurrently against the campaign database. Under those conditions, Aerospike delivered more than 3.7 million TPS for user profile applications, nearly all with sub-millisecond latencies; in addition, Aerospike also processed almost 200,000 read/write TPS for campaign applications, nearly all with sub-millisecond latencies.

The fast speeds and strong transaction volumes are particularly remarkable given the 20-node server size. Other NoSQL (or SQL) alternatives would require substantially larger footprints to manage similar transaction volumes and might still be unable to consistently deliver sub-millisecond latencies, as you will see in a moment. *(If you are interested in benchmarking a workload of your own, contact sales@aerospike.com.)* And if you're interested in learning how Aerospike performs with other workloads or other configurations, Aerospike's website contains a number of comparative in-house and third-party benchmark reports.

Cost

While operational throughput and data access latencies are two critical considerations for organizations seeking to manage real-time data at petabyte scale, cost is another. Indeed, for many firms, TCO is the biggest inhibitor to petabyte-scale deployments, which often require sprawling server footprints that lead to high infrastructure, maintenance, and operational costs. Because of this, Aerospike projected its TCO for running the benchmark system in production for 3 years, assuming a 15% annual growth in data volumes and transactions. Then, using publicly available capacity planning guidelines for Cassandra, Aerospike engaged a former

Cassandra solution architect to specify a configuration for Apache Cassandra 3.11.3 on AWS capable of managing the same workload for this benchmark, again assuming a 15% year-to-year growth rate. There's no guarantee that this Cassandra configuration could deliver data access latencies and throughput rates comparable to those Aerospike exhibited in its benchmark, as no tests were conducted, and many Cassandra customers report scalability issues. As such, the Cassandra configuration used in this analysis represents a best-case scenario for server footprint size.

Table 2 summarizes the results of this comparative TCO analysis. Aerospike's extremely efficient use of computing resources enables it to manage petabyte-scale workloads with a small fraction of the footprint that Cassandra would require. By Year 3, Aerospike would need a 26-node cluster to manage the target data volumes and workload, while Cassandra would need a footprint more than 30 times that size (792 nodes). Although Aerospike's configuration employed a more expensive type of server node from the standard AWS options available at the time of the benchmark, its efficient use of hardware resources would provide significantly lower infrastructure costs over 3 years ($4.11 million) when compared with Cassandra ($13.54 million). Larger server footprints drive up maintenance and support costs, too. As Table 2 shows, these 3-year costs were projected to be $61,200 for Aerospike versus $1.75 million for Cassandra. Finally, all those higher costs add up: Cassandra's 3-year TCO is estimated at $15.29 million, while Aerospike's is considerably lower at $4.17 million.

	Cassandra			**Total**	**Aerospike**			**Total**
	Year 1	**Year 2**	**Year 3**		**Year 1**	**Year 2**	**Year 3**	
Cluster size (Total servers)	514	638	792		20	22	26	
Cost per server ($ USD)	**$6,964**				**$60,409**			

Infrastructure ($ USD)	$3,579,599	$4,443,160	$5,515,646	**$13,538,405**	$1,208,179	$1,328,997	$1,570,633	**$4,107,809**
Fully burdened maintenance & support ($ USD)	$462,600	$574,200	$712,800	**$1,749,600**	$18,000	$19,800	$23,400	**$61,200**
TCO ($ Million USD)	**$4.04**	**$5.02**	**$6.23**	**$15.29**	**$1.23**	**$1.35**	**$1.59**	**$4.17**

Table 2: Comparative TCO summary for the petabyte-scale benchmark

Want to know the details behind those numbers? Both platforms were sized to manage 1 PB of unique user data with a compression factor of 4. Both employed the minimum recommended replication factors for their platforms to achieve high availability: a factor of 2 for Aerospike and a factor of 3 for Cassandra. Both were presumed to experience a 15% annual growth in data volumes and transactions. IT staffing costs for both were estimated at $180,000 per year per person, with one full-time person required for every 200 servers deployed.

As summarized earlier, Aerospike's benchmark was conducted on EC2 i3en.24xlarge server nodes, beginning with a 20-node cluster in Year 1; based on up-front AWS annual fees of $60,409 per node, this resulted in just over $1.2 million in infrastructure costs for the year. By contrast, Cassandra's Year 1 configuration was sized at a 514-node cluster of i3.4xlarge servers; based on up-front AWS annual fees of $6,964 per node, this yielded an infrastructure cost of nearly $3.58 million for Year 1. On the staffing side, Aerospike's 20-node cluster for Year 1 amounted to 10% of a full-time engineer's $180,000 annual cost, or $18,000. By contrast, Cassandra's 514-node cluster for Year 1 required more than 2.5 full-time engineers for support, resulting in an annual cost of $462,600.

Of course, annual expenses for both platforms increased to accommodate projected growth rates, but the growth in Aerospike's expenses was much lower than Cassandra's, as Fig. 5 depicts. Indeed, Aerospike would save an estimated $11.12 million in total costs over the 3-year period compared to Cassandra: $2.82 million in Year 1, an additional $3.67 million in Year 2, and an additional $4.63 million in Year 3.

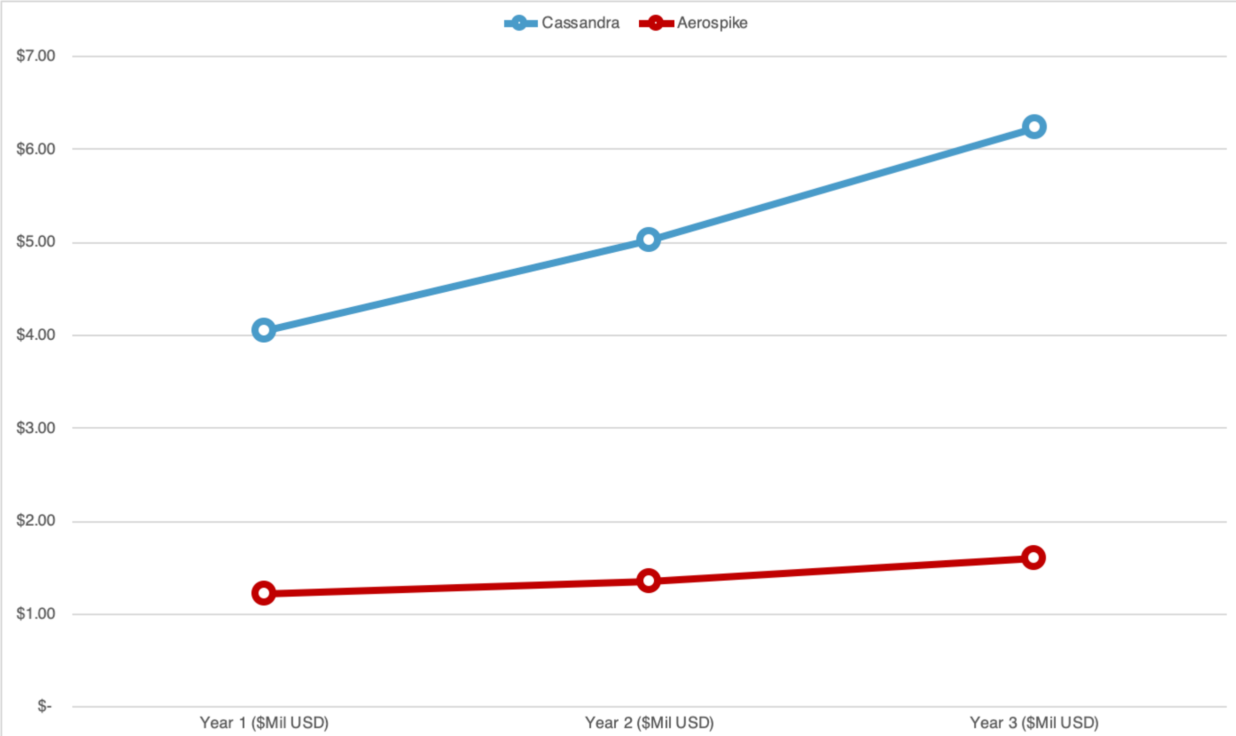

Figure 5: Total operational expenses by year for Cassandra and Aerospike

For further details about this TCO analysis, including specifics about the Cassandra configuration, see Appendix B: Comparative TCO details.

Client experiences

While the petabyte-scale benchmark results and comparative TCO analysis we just reviewed are quite compelling, the production experiences of Aerospike's clients are even more noteworthy. They prove that Aerospike's advantages are not limited to highly controlled, restrictive circumstances. Instead, Aerospike users have shown that the platform's performance, scalability, operational ease, and other features readily translate into tangible business benefits.

In most cases, clients turned to Aerospike after struggling with other solutions that fared well at modest levels of scale but simply could not meet more demanding business needs. As we step through some client scenarios, you will see how firms often struggled with overloaded legacy architectures, "freeware" that became too costly to operate at scale, and infrastructures that were too complicated, too expensive, and too unpredictable for future business needs.

Signal

An identity resolution firm, Signal faced latency problems with its existing Cassandra system as volumes grew. The platform's footprint spanned more than 450 servers and further server sprawl was looming due to projected growth in data volumes. Although Cassandra licensing was free, operational costs were becoming burdensome and taking resources away from higher-value projects. After testing Aerospike and seeing the potential reduction in operational costs and server footprint, Signal became convinced that the hidden costs of "freeware" were just too expensive to meet future business needs.

"One of the biggest problems we were running into was latency. And it was slowing down every element of our business processes. Compared to the other solutions that we were evaluating, the main driver that made Aerospike so attractive was its total cost of ownership was far lower than the competitive offerings that we had evaluated. (With Aerospike)…some processes that used to take 6 days now take 14 hours. Things that took 3 hours take 3 minutes now. And so, across the spectrum it's been much better."

– Jason Yanowitz, Vice President of Engineering, Signal

With Aerospike, the firm slashed its server footprint to 60 nodes and expects to save 68% in TCO over three years based on a detailed analysis. In addition, the firm "unwound a lot of weirdware we built to compensate for Cassandra's shortcomings." The result?

"In that rare engineering trifecta, we've found that it's [Aerospike's] faster, cheaper, and far more reliable."

– Jason Yanowitz, Vice President of Engineering, Signal

The Trade Desk

The Trade Desk is a demand-side AdTech firm that enables media buyers to purchase digital advertising. The firm once used Cassandra for storing "cold" data and Aerospike for caching "hot" data but ultimately standardized on Aerospike for both.

"In order to do the level of writes we needed to do (with Cassandra), we had to use compression, …there was a lot of tombstoning, there was a lot of CPU-

utilization needed relative to the size of the data that we were working on. And so in order to get to the throughput that we needed, we needed to scale the number of machines to a high number of machines with a lot of CPU compared to the disk that they had. Aerospike gave us another alternative approach…"

– *Matt Cochran, Director of Engineering, The Trade Desk*

The Trade Desk uses Aerospike at the edge for real-time bidding, where it receives up to 11 million queries per second or 800 billion queries per day. In addition, The Trade Desk uses Aerospike as a system of record on AWS to manage peak loads of 20 million writes per second in its cold storage of user profiles. By 2019, The Trade Desk had deployed 17 Aerospike clusters at 10 data centers to manage 400 - 800 billion objects, allocating a total of 3 PB for SSDs and 140TB for DRAM. Deployments spanned both cloud and on premises configurations.

"(We) run some of the largest (Aerospike) clusters in the world. Aerospike is the best No-SQL solution to support our millions of QPS workloads. (It's) amazingly simple to operate; dynamic cluster management and scalability is key."

– *Matt Cochran, Director of Engineering, The Trade Desk*

After nearly 10 years of using Aerospike, The Trade Desk has realized considerable business benefits:

"We have a ton of flexibility now… We can do all of these other things that we've been thinking about doing. We can take advantage of more opportunities in the market."

– *Matt Cochran, Director of Engineering, The Trade Desk*

Financial services firm

To improve customer service and accommodate projected growth in data volumes in a cost-effective manner, a leading worldwide brokerage firm began its data management modernization effort in 2013, embarking on a journey that ultimately led it to deploy Aerospike in multiple data centers to support applications that require a variety of sophisticated features, including ultra-low data access latencies, strong data consistency, rack awareness, and more.

Initially, the firm selected Aerospike to cache customer data spread across a mainframe DB2 system, a Teradata data warehouse, Microsoft SQL Server, and flat files. The company needed ultra-fast read access to this data for critical customer applications, including credit risk management, fraud management, and pricing. The firm used messaging and custom ETL processes to integrate data from its disparate data sources into an Aerospike caching layer, going live with its first production application on this modernized infrastructure in 2015 after carefully evaluating a number of solutions. Aerospike enables the company to run more than 13 billion queries a day with 99.9% of the requests returned in less than a millisecond.

Pleased with such results, the company subsequently leveraged Aerospike for write-centric work, creating an intraday trading application that relied on Aerospike as its system of record. To support legacy applications, intraday trading data stored in Aerospike is published to message queues, allowing its mainframe system to be updated asynchronously. This approach enabled the firm to improve the experience for its clients. One IT leader noted:

"We are migrating a lot of other applications onto this architecture… It [Aerospike] is working perfectly well and we have never had a customer complaint."

– *Vice President, Technology, Global Brokerage Firm*

The firm is now planning to expand its use of Aerospike even further, building new applications that will leverage it as a system of record. Aerospike's performance and cost-efficient scalability help make this feasible.

"We plan to store 500TB of… readily accessible data (in Aerospike). Surprisingly,…we are able to come up with a pretty small footprint to store that huge amount of data while we are getting all the benefits of low latency and able to keep up with all the SLAs."

– *Vice President, Technology, Global Brokerage Firm*

Verizon Media

Faced with a disparate, aging data management infrastructure, Verizon Media sought to consolidate and streamline its approach to managing mission critical real-time data. After extensive evaluations, the firm decided to base its global modernization effort on Aerospike to replace its current system-of-record (managing multiple petabytes of data) and its user data profile store tracking data from its 2 billion customers. The system of record supports a multi-tenant application for all key-value stores across the company, while the profile database serves as the key-value store for users across all Verizon Media domains.

Together, applications targeted for the new Aerospike infrastructure require low data access latencies, high scalability (at petabyte scale), ease of operations, strong data consistency, global active-active deployments spanning multiple data centers, agility, and a low server footprint. Verizon Media expects to deploy Aerospike in several worldwide data centers using cross-data center replication (XDR) as needed. With clusters of 100 nodes or less, many based on PMem, Verizon Media expects to achieve the scalability and performance it needs for 100+ applications. By replacing multiple open-source platforms (RocksDB, HBase) and some homegrown technology with Aerospike, Verizon Media expects to simplify its infrastructure, support emerging business requirements, prepare for future growth, and keep within its target budget. As one architect put it,

"Aerospike is the only product to beat our own technology in 12 years, and we've tested everything."

For more details, see Verizon Media's presentation at the Aerospike Digital Summit 2021.

Summary

Scaling operational databases from terabytes to petabytes isn't just about raw capacity. TCO, runtime performance during modest and peak workloads, data availability, deployment flexibility, and operational ease all determine a platform's suitability for managing real-time data at scale. With Aerospike, you don't have to compromise. Aerospike's aggressive exploitation of the latest storage, processor, and networking technologies provides exceptionally efficient use of computing resources to deliver ultra-fast performance on much smaller footprints than other alternatives. Beyond that, Aerospike provides self-managing and self-healing features, high availability, strong data consistency, and flexible deployment options, all of which combine to enable firms to scale their operational workloads easily and affordably.

A recently published petabyte-scale benchmark demonstrated how Aerospike delivers 4-5 million TPS with sub-millisecond latencies against more than 1 PB of unique user data on a mere 20-node cloud cluster. A competing NoSQL implementation could easily require hundreds of nodes, adding $11+ million in total operational costs within 3 years.

Aerospike's fast, predictable performance on a small server footprint makes petabyte-scale processing of real-time operational data pragmatic for a wide range of firms – not just those with extraordinary budgets and elite IT staff. Surprised? Contact Aerospike to explore how you, too, can enjoy high scalability without high costs.

Appendix A: Benchmark details

The petabyte-scale benchmark described earlier in this paper featured the following setup:

Aerospike Enterprise Server 5.4.0.2-1 on Redhat 7:

- 20 nodes of Amazon EC2 i3en.24xlarge instances with
  - Up to 3.1 GHz Intel® Xeon® Scalable (Skylake) processors with Intel® Advanced Vector Extension (or Intel® AVX-512) instruction set
  - 768 G of memory
  - 8 x 7500 GB NVME SSDs
  - 96 vCPU
  - 100 Gbps network
- Amazon Linux 2 OS
- Chrony for time synchronization
- Saltstack for benchmark provisioning and execution

Client systems:

- 40 nodes of Amazon EC2 c5n.9xlarge instances with
  - 3.0 GHz Intel Xeon Platinum processors with Intel Advanced Vector Extension 512 (AVX-512) instruction set
  - Sustained all core turbo frequency of up to 3.4GHz, and single core turbo frequency of up to 3.5 GHz
  - 96 G of memory
  - EBS-only (storage)
  - 36 vCPU 96
  - 50 Gbps network
- Aerospike C client
- Chrony for time synchronization
- Saltstack for benchmark provisioning and execution
- Bash scripting for local operations

Saltstack Master server:

- 1 Amazon EC2 c5n.4xlarge instance
  - 3.0 GHz Intel Xeon Platinum processors with AVX-512 instruction set
  - Sustained all core turbo frequency of up to 3.4GHz, and single core turbo frequency of up to 3.5 GHz
- Chrony for time synchronization
- Saltstack master, minion, and cloud

- Bash scripting to encapsulate Saltstack requests

Furthermore, the benchmark featured two Aerospike namespaces (databases), configured as follows:

User profile namespace:

- All Flash (index and data on SSDs)
- 500 billion unique keys
- Object sizes - 62.5 billion unique objects of each of these sizes (in bytes):
  - 256 B, 512 B, 768 B, 1024 B, 2048 B, 3072 B, 4096 B, and 5120 B
- Replication factor 2
- Lz4 compression; compression ratio of 4
- Total uncompressed data: 2 PB (1 PB primary copy and 1 PB secondary copy)
- Total compressed data: 500 TB (250 TB primary copy and 2 TB secondary copy)
- Consistency mode: AP (available/partition tolerant)

Campaign namespace:

- Hybrid memory (index in DRAM and data on SSDs)
- 6 billion unique keys
- Object size: 256 B (bytes)
- Replication factor 2
- Lz4 compression; compression ratio of 4
- Total uncompressed data: 3 TB (1.5 TB primary copy and 1.5 TB secondary copy)
- Total compressed data: 750 GB (375 GB primary copy and 375 GB secondary copy)
- Consistency mode: SC (strong consistency)

Benchmark workloads for the user profile namespace involved read-only and an 80/20 mix of read/write operations. The workload for the campaign namespace was a 50/50 mix of read/write operations.

Appendix B: Comparative TCO details

To help you understand the TCO analysis summarized earlier, this appendix details the Cassandra configuration used for the comparison and the rationale for its selection. You will also find detailed cost estimates for the Cassandra and Aerospike configurations based on AWS one-year up front pricing as of this writing.

Cassandra configuration over 3 years

To plan a hypothetical AWS deployment of Apache Cassandra 3.11.3, Aerospike relied on publicly available capacity planning guidelines, the experiences of current and former Cassandra customers, and the experience of a former Cassandra solution architect to determine the number and types of AWS servers needed to support the data volumes and transaction rates of the petabyte-scale benchmark for a read-only and 80/20 mixed read/write workload. Table 3 outlines the resulting configuration over a 3-year period, assuming a 15% year-to-year growth rate.

Cassandra

Name	Year 1	Year 2	Year 3
Cluster Size	514.0	638.0	792.0
# Data Centers	1.0	1.0	1.0
Total Servers	514.0	638.0	792.0
Unique Total Data footprint (TB)	1131.0	1300.6	1495.7
Data Capacity per node (TB)	3.8	3.8	3.8
Replication Factor	3.0	3.0	3.0
# of Keys (Bn)	506.0	581.9	669.2
Object size avg (bytes)	2048.0	2048.0	2048.0

Object Storage size (bytes)	2457.6	2457.6	2457.6
Read peak (ktps)	5095.0	5859.3	6738.1
Writes peak (ktps)	95.4	109.7	126.2
RAM per Node (GB)	122.0	122.0	122.0
Replication size (TB)	3393.0	3901.9	4487.2
Defrag (No GC) **Factor + Over Prov**	2.3	2.3	2.3
Compression Ratio	0.25	0.25	0.25
FTE Time Share %	257%	319%	396%

Table 3: Cassandra configuration based on AWS i3.4xlarge nodes with 122 GB DRAM and 3.8 TB storage

The following factors were considered when planning the Cassandra configuration:

- Cassandra specialists recommend using no more than 4 TB storage total per node, with approximately 2 TB for business data. Given the moderately high throughput rate of 5 million reads per second achieved in Aerospike's petabyte-scale benchmark, Cassandra would benefit from ample DRAM to promote caching. The i3.4xlarge nodes have about 3.8 TB storage and 122 GB DRAM – reasonably close to our target. (DRAM of 256+ GB per node would have been better, but we wanted to focus on the most reasonably priced configuration that could handle the target data volumes. A quick sizing using m5d.16xlarge nodes with 256 GB DRAM resulted in almost a 5x cost increase, so that configuration was rejected.)

- Given an average user object size of 2kB and the associated overhead Cassandra adds to objects (particularly when there are multiple columns), we estimated the average object size on storage at 2.4kB.

- As with Aerospike, we assumed a compression ratio of 0.25. We're not certain that Cassandra could actually achieve this level of compression for the user data employed in the Aerospike benchmark, but we chose to operate under the assumption that this was feasible, again to keep costs down and present an optimistic Cassandra scenario.

- Cassandra is a quorum-based system, so it needs an odd number of data copies. We assumed 3 copies of the data, which is consistent with many production deployments. (Aerospike is a multi-master system, so it can use any replication factor. In this benchmark, 2 copies were used; this is consistent with many production deployments.)

- Beginning with 506 billion keys in Year 1 at ~2.4kB per object, we added room for compactions, headroom, and so on to calculate the total storage needed. We then divided this by 3.8 TB storage per node to project the number of required nodes (514 in Year 1).

As mentioned, earlier in the paper, we didn't benchmark Cassandra. Therefore, we are not certain that this Cassandra configuration could readily deliver roughly 5 million reads and 95,000 writes per second, as Aerospike did. Our sizing is based purely on data volume and is the best-case scenario for Cassandra's number of nodes. Customers have told us that Cassandra scales non-linearly as data volumes grow and latency targets remain the same; indeed, they have reported an 8% addition to linear scaling, so we applied this to Years 2 and 3.

For both Cassandra and Aerospike cost estimates, we used the list price for one-year, advanced payment of the AWS nodes described. We also assumed a 15% annual growth rate in data volumes and throughput. Finally, we estimated that 1 full-time engineer could manage 200 cloud servers at a cost of $180,000 per year.

Aerospike configuration over 3 years

Since Appendix A details the Aerospike configuration employed in the benchmark, this section explores how this configuration evolved over the 3-year period used in the TCO comparison. Table 4 is Aerospike's equivalent to the data presented in the previous section, which profiled the Cassandra configuration used over a 3-year period in the TCO analysis.

Aerospike

Name	Year 1	Year 2	Year 3
Cluster Size	20.0	22.0	26.0

# Data Centers	1.0	1.0	1.0
Total Servers	20.0	22.0	26.0
Unique Total Data footprint (TB)	994.0	1143.1	1314.6
Data Capacity per node (TB)	60.0	60.0	60.0
Replication Factor	2.0	2.0	2.0
# of Keys (Bn)	506.0	581.9	669.2
Object size avg (bytes)	2048.0	2048.0	2048.0
Object Storage size (bytes)	2160.0	2160.0	2160.0
Read peak (ktps)	5095.0	5859.3	6738.1
Writes peak (ktps)	95.4	109.7	126.2
RAM per Node (GB)	768.0	768.0	768.0
Replication size (TB)	1988.1	2286.3	2629.2
Defrag (No GC) **Factor + Over Prov**	2.3	2.3	2.3
Compression Ratio	0.25	0.25	0.25
FTE Time Share %	10%	11%	13%

Table 4: Aerospike configuration based on AWS i3en.24xlarge nodes with 768 GB DRAM and 8 x 7.5 TB storage

As discussed earlier, Aerospike selected the best storage-optimized instances available at the time of this writing on Amazon EC2 to showcase the fewest number of servers. These instances included more DRAM than Aerospike needed to achieve its results with an all-Flash configuration, resulting in added costs. In the future, other AWS instances may become available that could provide even lower TCO to achieve the same performance. Furthermore, firms that prefer bare metal instances can customize their hardware configurations in various ways to optimize the TCO even further.

Comparative cost calculations

The Cassandra and Aerospike configurations presented in Tables 3 and 4 enabled us to project each platform's TCO. Infrastructure costs were derived from the list price of 1-year up front AWS fees for the respective instance types employed. Cassandra's i3.4xlarge nodes cost $6964.20 each, while Aerospike's i3en.24xlarge nodes cost $60,408.96 each. While the node type used for Aerospike was substantially more expensive, Aerospike also required substantially fewer nodes. This led to Aerospike's substantially lower TCO over the 3-year period.

As Tables 5 and 6 show, TCO for Cassandra was projected at $15.29 million over 3 years. During this same period, Aerospike's TCO was only $4.17 million – a savings of more than $11 million.

Cassandra TCO - Calculation

	Year 1 ($USD)	**Year 2** ($USD)	**Year 3** ($USD)
	Infrastructure Cost		
Annual Server Lease Cost	$3,579,599	$4,443,160	$5,515,646
	Maintenance, Support, Ops		
IT Personnel Cost	$462,600	$574,200	$712,800
	Total Infra Cost		
Total Cost of Ops (million $USD)	**$4.04**	**$5.02**	**$6.23**
	Total Software Cost		
Offset Factor	1.0000	1.0000	1.0000
	Total Cost of Operations		
(million $USD)	**$4.04**	**$5.02**	**$6.23**

Total 3YR cost (million $USD)	$15.29

Table 5: Cassandra 3-year TCO breakdown

Aerospike TCO - Calculation

	Year 1 ($USD)	**Year 2** ($USD)	**Year 3** ($USD)
	Infrastructure Cost		
Annual Server Lease Cost	$1,208,179	$1,328,997	$1,570,633
	Maintenance Cost		
IT Personnel Cost	$18,000	$19,800	$23,400
	Total Infra Cost		
Total Cost of Ops (million $USD)	**$1.23**	**$1.35**	**$1.59**
	Total Software Cost		
Offset Factor	1.0000	1.0000	1.0000
	TOTAL COST OF OPERATIONS		
	$1.23	**$1.35**	**$1.59**

Total 3YR cost (million $USD)	**$4.17**

Table 6: Aerospike 3-year TCO breakdown

About Aerospike

The Aerospike Real-time Data Platform enables organizations to act instantly across billions of transactions while reducing server footprint by up to 80 percent. The Aerospike multi-cloud platform powers real-time applications with predictable sub-millisecond performance up to petabyte scale with five-nines uptime with globally distributed, strongly consistent data. Applications built on the Aerospike Real-time Data Platform fight fraud, provide recommendations that dramatically increase shopping cart size, enable global digital payments, and deliver hyper-personalized user experiences to tens of millions of customers. Customers such as Airtel, Experian, Nielsen, PayPal, Snap, Verizon Media, and Wayfair rely on Aerospike as their data foundation for the future.

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

About Intel

About Intel

Intel (Nasdaq: INTC) is an industry leader, creating world-changing technology that enables global progress and enriches lives. Inspired by Moore's Law, we continuously work to advance the design and manufacturing of semiconductors to help address our customers' greatest challenges. By embedding intelligence in the cloud, network, edge and every kind of computing device, we unleash the potential of data to transform business and society for the better. To learn more about Intel's innovations, go to newsroom.intel.com and intel.com.

About AWS

Amazon Web Services (AWS) is the world's most comprehensive and broadly adopted cloud platform, offering over 200 fully featured services from data centers globally. Millions of customers—including the fastest-growing startups, largest enterprises, and leading government agencies—are using AWS to lower costs, become more agile, and innovate faster.