

White Paper

AEROSPIKE REAL-TIME DATA PLATFORM

Facing a future where data could consume 20% of global energy by 2030, can we make our data platforms greener, cheaper and faster? It's time to confront the challenge.

Harvard researchers expect that by 2030 information and computing technology will account for as much as 20% of global energy demand.



Table of CONTENTS



DATA TRIBES

Unlock real-time performance and reduce carbon footprint by 80% with ultra low latency at any scale.

01

Speed - Processing data at scale without compromising latency has become key in competing with global markets. This section investigates Aerospike's unique architecture, which enables extreme speed at scale.

02

Carbon footprint - EU has ruled to reduce greenhouse gas emissions by >55% and companies are required to report ESG metrics as early as 2023. As efficient software lowers company CO₂ emissions significantly, we analyse Aerospike's CO₂ footprint.

03

Costs - If managed incorrectly, infrastructure can quickly become cost prohibitive, so much so that a business's profitability is eroded. In this section, we discuss how Aerospike helps avoid costly mistakes on your path to growth.

Introduction to AEROSPIKE

Why Aerospike?

The intensely competitive global economy is forcing firms in banking, retail, financial markets, telecommunications, and other industries to deliver highly personalised products and services in real time around the clock, placing incredible demands on data management infrastructures. Aerospike's NoSQL operational database platform leverages hardware innovations to deliver high-scale, extreme low latency transaction processing in a cost-effective manner, enabling users to shrink their server footprints up to 90% and cut operational costs while fulfilling aggressive service level agreements (SLAs).



Industry leaders such as Airtel, Banca D'Italia, Nielsen, PayPal, Snap, Verizon Media and Wayfair leverage Aerospike for mission-critical applications involving recommendation engines, personalised customer engagement services, fraud detection, payment processing, machine learning, and more. Such applications require Aerospike to operate at the edge, at the core, and across multiple data centres or cloud regions. These deployments demand predictable, ultra-fast data access at scale from a database platform that is always on, always available, and always consistent.

This paper introduces Aerospike's capacity to improve performance at scale, while reducing carbon footprint and the total cost of ownership (TCO).



"Very early on, we started working with Aerospike on this particular situation. It was very clear that the technological advantage was there. But as we started using it, it wasn't just about the latency, it wasn't about the tool we could get - it was also about the reliability."

Sandeep Nawathe, Senior Director of Engineering for Adobe Experience Platform, Adobe



32%
annual growth

20TB
of data replicated
across data centres

2ms
latency @95th
percentile

4MM
transactions/second
at edge

Aerospike & SPEED

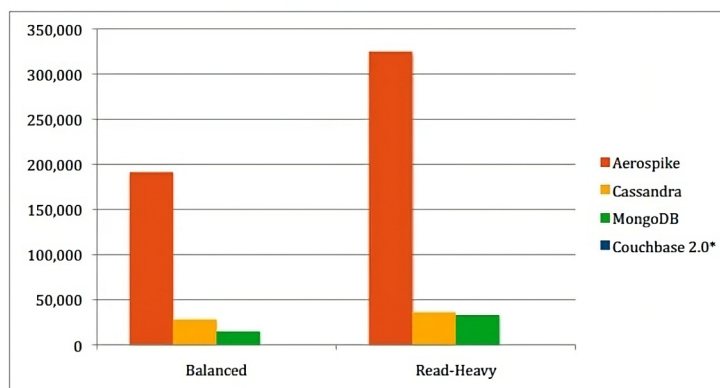
How does Aerospike guarantee better speed?

Collecting, filtering, and analysing streaming data is vital in today's highly demanding business landscape. Leveraging these vast amounts of data to make more accurate, timely business decisions can yield substantial tangible results. This requires an infrastructure that integrates disparate data sources and platforms in real time and makes that data available to applications without delay. Aerospike has developed critical technologies to help firms process, persist and protect petabytes of streaming data on modest server footprints. Aerospike's highly performant, scalable interface to streaming and messaging platforms enables its real-time data platform to ingest massive data volumes captured at the edge to power operational and transactional applications. Users can combine this streaming data with other transactional, historical data stored. In addition, Aerospike's change data notification capabilities promote the sharing of new or modified data with downstream platforms.

The Aerospike Real-time Data Platform architecture features massive parallelism, support for modern hardware architectures, intelligent workload processing, strong data consistency, and self-managing features. Powered by Starburst Enterprise, Aerospike SQL provides an enterprise-grade implementation of the open source Trino project that is installed, managed and supported by Aerospike, enabling SQL programmers to access streamed data stored in Aerospike for immediate query and analysis.

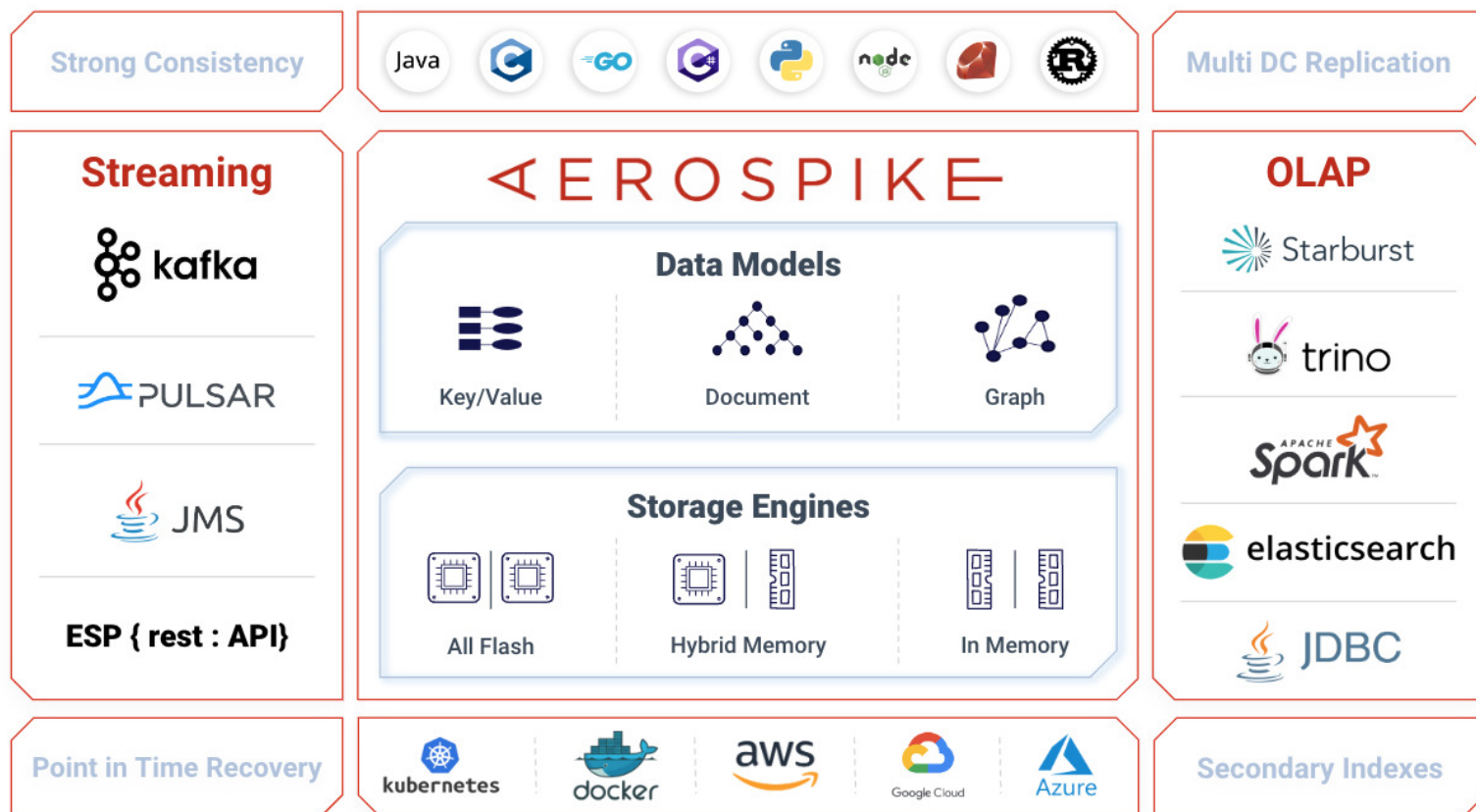
Aerospike delivers exceptional availability and runtime performance with dramatically smaller server footprints through deep exploitation of modern hardware, including multi-core processors with non-uniform memory access (NUMA), non-volatile memory extended (NVMe) Flash drives, persistent memory (PMem), network application device queues (ADQ), and more.

1. Maximum Throughput - SSD-backed Data Set



*For Couchbase 2.0, SSD throughput numbers are based on smaller sample size and asynchronous. Couchbase 1. was unable to load even the reduced data set.

Other features that distinguish Aerospike include its ability to automatically distribute data evenly across shared-nothing clusters, dynamically rebalance workloads, intelligently route application requests to appropriate nodes for fast performance, and accommodate software upgrades and most cluster changes without downtime.



Aerospike stands out in the field of data management with its unique features and innovative architecture. The platform is designed to automatically distribute data evenly across shared-nothing clusters, dynamically rebalance workloads, and intelligently route application requests to appropriate nodes. This ensures fast performance and allows the system to accommodate software upgrades and most cluster changes without experiencing downtime.

Adding to these distinguishing features, Aerospike's patented Hybrid Memory Architecture™ (HMA) brings a new level of efficiency and performance to data storage. HMA treats solid-state drives (SSDs) as raw devices, managing input/output operations directly to bypass common wear-levelling issues. HMA enables Aerospike to get sub-millisecond response times from PMem and SSDs. This innovative approach enhances the system's performance and longevity.

Furthermore, Aerospike has optimised the use of SSDs and DRAM to strike a balance between cost and performance. By storing data on SSDs and indexing on DRAM, Aerospike provides the speed of DRAM with the cost-effectiveness and reliability of SSDs.

This combination of unique features and a breakthrough architecture enables Aerospike to offer a high-performance, cost-efficient and reliable real-time data platform.



"PayPal is innovating deep analytics to rapidly respond to emerging fraud patterns, then deploying into an event-driven, fast data, in-memory architecture to accelerate detection, reduce losses and achieve near-continuous availability"

Mikhail Kourjanski PHD, Lead Data Architect, Risk and Compliance Management Platform, PayPal



30X

reduction in fraud transactions missed

15X

reduction in server footprint

\$9M

hardware cost savings projected

5X

throughput improvement

Reduce your FOOTPRINT

Climate change and environmental degradation pose existential threats to both Europe and the world. In order to tackle these challenges, Europe has introduced the Green Deal, which aims to achieve climate neutrality. Additionally, the Corporate Sustainability Reporting Directive (CSRD) ensures that companies report their ESG (Environmental, Social, and Governance) impact as early as 2024.



Harvard researchers (Gupta et al., 2020) expect that by 2030 information and computing technology will account for as much as 20% of global energy demand. Minimising carbon footprint in your server can thus result in significant environmental sustainability gains.

The *Minimising CO₂ Production from IT Systems* White Paper provides a comparison of CO₂ emissions between Apache Cassandra and Aerospike. This article summarises the findings of the estimates and the CO₂ emission differences of nominally similar technologies.

The study compared 1 PB use case:

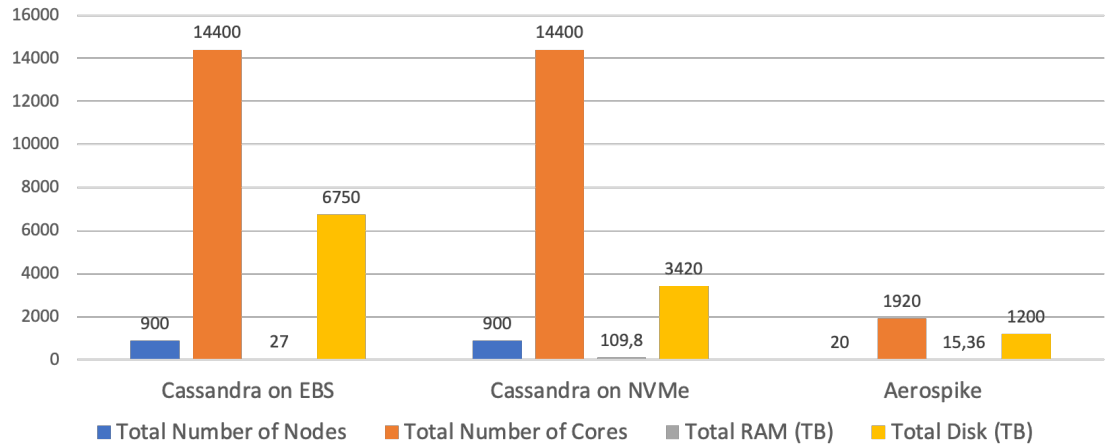
- 1. Physical Resource Estimate:** Minimum recommended hardware if using AWS.
- 2. Emissions Estimate:** This includes emissions produced during the hardware manufacturing process and emissions generated by operational use. (Scope 1, 2, and 3)
- 3. Operational Cost:** The monetary cost of deploying versus the AWS platform.

1. Carbon footprint by Solution

Solution	Number of Nodes	Total Annual EC2 Manufacturing (kgCO ₂ eq)	Total Annual Energy Consumption (kgCO ₂ eq)	Total Annual Storage Manufacturing (kgCO ₂ eq)	Total Annual Storage Energy Consumption (kgCO ₂ eq)	Total Annual Emissions (kgCO ₂ eq)
Cassandra on EBS	900	105,646	329,156	224,497	18,578	677,878
Cassandra on NVMe	900	130,874	320,785	87,496	7,241	546,397
Aerospike	20	14,972	60,849	30,701	2,541	109,017

The results are surprising. As presented in Table 2, the choice of hardware can reduce the CO₂ emissions of a solution by 20% (Cassandra: 678 → 546). The amount of resources each solution requires is illustrated in the Plot 2.

2. Physical Resource Estimation



Sometimes efficiency has a price. For example, in the case of Apache Cassandra, the solution with the least amount of CO₂ emissions is more expensive, and the cheaper solution is more polluting than the alternative.

2. CO₂ Emissions & TCO Estimation

Solution	Total Annual Emissions (kgCO ₂ eq)	Annual TCO
Cassandra on NVMe	546,397	\$10,848,384
Aerospike	109,017	\$2,102,400

However, the efficiency of platforms based on different underlying technologies can be dramatically different. As we saw in this comparison, a

more efficient software solution can be many times cheaper and less polluting. More recently, Aerospike has also been adapted to support AWS Graviton. Deploying Aerospike on AWS Graviton-based instances reduces the TCO and carbon footprint of solutions even further.

Although the purpose of the Carbon Footprint section is not to compare the "performance" of these technologies, it is worth mentioning that the expected latency of Apache Cassandra is in the region of a single digit millisecond (<10ms), whereas Aerospike works in less than a millisecond latency range (>1ms). Yet another substantial improvement.



"Aerospike is by far the best for what we do...It makes a real big difference for our ability to execute when we're trying to follow a customer along the customer journey and they're transitioning between states. If you're an extra 50 milliseconds behind, it can cost you putting the right message in front of them."

Dave Pickles, Founder and CTO, The Trade Desk



800B
queries per day

9X
reduction in node count

<8ms
reduction in data access

10M
queries per second

Reduce COSTS

In today's economy, inflation continues to drag on business profits, supply chain glitches remain, and the demand by investors for accountability in tech spending only grows. A proven solution for your digital transformation initiatives can be the greatest tool for saving money while growing in a competitive environment.



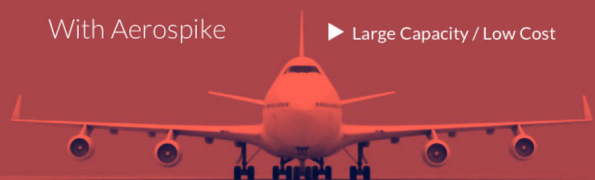
Total Cost of Ownership (TCO) doesn't usually get much consideration early in the application development life cycle. But as a business becomes more successful, its applications require a dramatic expansion in scale to handle demands from an increasingly mobile Internet constituency. With this upsurge in scale comes an increase in costs, often exponentially.

There are two types of costs to consider—tangible and intangible. **Tangible costs** are easy to measure and include things like hardware and software. **Intangible costs**, though harder to measure, are equally real and include things like downtime (scheduled and unscheduled) and poorly performing websites. At best, these occurrences simply frustrate users. At worst, they result in damage to a brand and users abandoning apps altogether. Other intangible costs such as those for DevOps and IT support are even harder to measure; yet, in terms of impact, these can be substantial—lost opportunity, employee burnout, phone calls at 3AM.

Without Aerospike ▶ Small Capacity / High Cost



With Aerospike ▶ Large Capacity / Low Cost



Since 2010, Aerospike has been deployed in some of the most rigorous production environments at an unbeatable TCO. Its dynamic (self-healing) cluster management allows Aerospike to easily scale out, with uptimes of five 9s or better, and rebalance when a node is lost, without requiring extensive, hands-on management. Aerospike's hybrid memory system allows for a dramatic reduction in node count, usually **requiring fewer than 1/5th the number of servers** typical of NoSQL and relational database implementations. This alone can save your enterprise millions in capital expenditures, not to mention enable reduced server sprawl, easier maintenance, and less OpEx.

In today's economy, businesses face inflation, supply chain glitches, and increasing demands for tech spending accountability. Embracing a proven digital transformation solution like Aerospike can save money and foster growth. Aerospike's unbeatable TCO, dynamic cluster management, and hybrid memory system offer substantial cost savings, improved performance, and enhanced operational efficiency for enterprises.

CONCLUDING REMARKS:

In today's competitive and sustainability-focused landscape, the Aerospike Real-time Data Platform emerges as a powerful tool for organizations. It provides the necessary speed, scalability, and cost efficiency while minimizing environmental impact. By embracing Aerospike, businesses can unlock the full potential of their data, confidently navigate the future, and drive positive change.

The developers of Aerospike recognize that efficiency is not only about speed. Efficient software also requires less hardware, uses less energy, and provides a lower Total Cost of Ownership (TCO). By leveraging hardware innovations and parallelism, Aerospike achieves exceptional availability and run-time performance while significantly reducing the number of servers required, resulting in substantial savings in capital expenditures and operational expenses as well as lowering carbon emissions.

In conclusion, Aerospike's real-time data platform offers a compelling solution that integrates performance, sustainability, and cost-effectiveness. By adopting Aerospike, organizations can achieve higher speed, lower total cost of ownership, and reduced carbon footprint. This strategic choice not only ensures a competitive edge but also contributes to a greener and more sustainable future for data management.

Hear more about AEROSPIKE'S CUSTOMERS:





DATA TRIBES

Creating data-driven organisations that make a difference.

Get in Touch



www.datatribes.nl



getintouch@datatribes.nl



+31 (0)20 2101269



www.instagram.com/datatribesnl



www.linkedin.com/company/datatribesnl



www.twitter.com/datatribesnl



www.meetup.com/data-thursday

