Overview

The intensely competitive global economy is forcing firms to deliver highly personalized products and services in real time, around the clock, and around the globe, placing incredible demands on existing data management infrastructures. Existing data platform technologies include relational databases, first-generation NoSQL databases, and even mainframes—but they each have distinct shortcomings.

Relational databases, even distributed ones, don’t have the high performance (low latency millisecond responsiveness coupled with millions of data transactions per second) that NoSQL technologies do to meet this need. First-generation NoSQL technologies rely on volatile DRAM memory to get high performance but require high server counts to hold large amounts of data. Mainframes offer speed with accuracy and uptime but cannot be distributed globally at the Edge where users access data and are costly to operate.

Aerospike solves these challenges with its highly patented Hybrid Memory Architecture™, which delivers DRAM-like high performance with SSDs. As a result, Aerospike technology provides high performance, strong consistency, and availability while being globally distributed, reducing server footprint by up to 80% while your business grows—unlike any other data platform in the world.

Companies have built their foundation for the future with Aerospike. Some examples include:

- PayPal slashing its fraud exposure 30x.
- Wayfair increased its average cart size at checkout by double-digit percentages.
- Top 3 Global Brokerage processing trades five times faster.
The Aerospike architecture is unique
The Aerospike Real-time Data Platform is a shared-nothing, multi-threaded, multimodal data platform designed to operate efficiently on a cluster of server nodes, exploiting modern hardware and network technologies to drive reliably fast performance, often at sub-millisecond speeds across petabytes of data.

SSDs treated like DRAM
Aerospike has a patented Hybrid Memory Architecture™ (HMA), which places data on solid-state drives (SSDs) and indexes-only in DRAM. In addition, Aerospike treats SSDs as raw devices, writing data in large blocks using a highly efficient custom file format that avoids wear-leveling issues common with other providers.

Tiered storage options for optimal densities with high performance
Administrators can configure Aerospike’s storage usage to suit their cost, performance, availability, and scalability objectives. Options include keeping indexes and user data all in traditional memory (DRAM), all in Persistent Memory (PMem), all on SSDs, or in hybrid configurations (with indexes in DRAM or PMem and user data on SSDs). There is an advantage to having a single data platform that can execute in real time for various data set sizes: one codebase; no reprogramming; scale as you grow.

Node reduction by up to 80 percent
In-memory DRAM can only store a fraction of the hot data that Aerospike Hybrid Memory Architecture™ SSDs store per node. Figure 2 illustrates this node reduction. It compares an In-Memory system using AWS Memory Optimized instance nodes versus an Aerospike HMA system using Storage Optimized instances. The critical difference is in the addressable space: In-memory only has 122 GiB/node, whereas Storage Optimized nodes have 3.2 terabytes of addressable space per node - a whopping difference.
Predictable performance from terabytes to petabytes

Your apps have to perform consistently well no matter the circumstances - lest they damage your brand. Enter Aerospike’s Hybrid Memory Architecture™, designed to scale seamlessly and perform linearly at scale because it only needs to access indexes from volatile memory, not the data itself - unlike In-memory systems. No “cache misses,” no associated spikes in latency. Aerospike Hybrid Memory Architecture has low variation and low response times versus In-memory systems (see Figure 3).

Direct path to data (single hop)

Aerospike’s Smart Client™ is a first-class participant in the architecture and data fabric and communicates with every node. Thus, every node in the Aerospike system knows what data every other node has and can reach data in a single hop. The result is consistent, predictable access times, whereas other technologies often require multiple jumps from node to node to find data - increasing latency.
Hotspots avoided with intelligent auto-sharding

To avoid any single node receiving a disproportionate number of requests (thus making the system susceptible to either slowdowns or, worse, crashes), Aerospike automatically employs a deterministic, uniform data distribution scheme. As a result, Aerospike conducts load balancing continuously and automatically on all servers, even when scaling up or down or with cluster reconfigurations. The benefit: new use cases don’t require re-tuning, and DevOps teams can reuse the same scheme and algorithms.

This illustrates adding a fourth node (Z), initiated by a “heartbeat” communication, rebalancing data (25% per node), and the resulting roster communicated up to the Smart Client.

Node add/remove/update without disruption

Aerospike has a self-healing, auto-sharding, algorithmic cluster management system that adds, removes, or updates nodes without disruption/need to take the system down for maintenance. The result is high uptime as Aerospike has a “shared nothing” architecture, and there are no single points of failure, unlike other systems.

Real-time transactions without data loss

Aerospike provides strong consistency with high performance. For example, in case of a node failure or a network break to a node (known as a “partition”), it does not allow reads until that partition between two nodes gets healed, after which they can synchronize data, making the data consistent. Furthermore, no stale (old) nor dirty (non-consistent) reads are allowed.

Intercloud data replication with fine-grained control

Aerospike offers Cross-Datacenter Replication (XDR) with Expressions functionality that ships only the data you want, where you want it, asynchronously. The result is a high-performance global data hub with dynamic control for compliance needs, such as GDPR and CCPA, inventory, or logistics. In addition, Aerospike can route data captured at the Edge to non-Aerospike targets (e.g., ERP, CRM, or Inventory Systems via our Change Notification capability).

High availability upon site failures without intervention

Aerospike employs a feature called Rack Awareness, which pegs copies of data (either a Master or Replica) to specific racks across zones or data centers within a cluster for different hardware failure groups.

Many use cases

These use cases range from the simple to the previously unimaginable, from terabytes in size to petabytes:

- Caching (Edge-only)
- Cache/ODS consolidation
- Real-time data pipeline ingestion
- Mainframe Offloading
- Edge to System of Record (Warm store)
- Global real-time transaction processing
- Low latency store shared service.