

WHITE PAPER

Fueling real-time AdTech: Unified data infrastructure for buy-side precision and sell-side scale

Contents

Introduction	3
AdTech ecosystem overview	3
The sell-side	5
The scenario	5
The buy-side needs to optimize ROAS	6
Programmatic advertising is a real-time, data-driven process	6
The importance of commerce media	7
Commerce media: A special case	8
Scenario highlighting graph and vector search needs	8
Why Aerospike for commerce media	10
Offline data needs	10
User profiles and audiences are preeminent	10
The customer journey should be captured in a graph	10
Product catalogs are best served by both document and vector models	12
Audience creation	14
Types of audiences/segments	14
Aerospike is ideal for AdTech real-time use cases	16
One database for AdTech	16

Introduction

In the world of programmatic advertising, milliseconds matter. From the moment a user lands on a page to the instant a decision is made to serve an ad—or not—the entire AdTech supply chain must operationalize data at speed and scale. This requires real-time access to diverse data sources across buy-side and sell-side platforms, stitched together with precision to support [identity resolution](#), contextual relevance, [fraud detection](#), and attribution. Every impression is an opportunity—and a challenge—to deliver value and relevance, all while operating at the outer limits of performance.

To meet these demands, AdTech leaders are increasing personalization at scale by enhancing how they make decisions, target users, attribute outcomes, and integrate disparate data signals. Real-time streaming look-ups power smarter, faster decisioning engines that optimize every transaction. On the buy-side, this means selecting the right user at the right time with the right message. On the sell-side, it means maximizing yield while maintaining integrity and user trust. Across both, signal aggregation is key, hydrating applications with the critical data they need to assess attention, viewability, context, and fraud risk in the blink of an eye.

Aerospike delivers the unified platform to make this possible, combining the speed and resilience of the [Aerospike Database](#), the relationship intelligence of the [Aerospike Graph](#), and the contextual precision of [Aerospike Vector Search](#). With a globally distributed architecture, data can be accessed and streamed instantly to where it's needed, keeping pace with modern programmatic ecosystems. Whether you're preventing fraud, personalizing experiences, or making high-stakes decisions in real time, Aerospike helps AdTech businesses live life a millisecond at a time.

AdTech ecosystem overview

In the AdTech ecosystem, the buy-side and sell-side play crucial roles in the [real-time bidding \(RTB\)](#) process.

The **buy-side** comprises advertisers and brands looking to purchase ad inventory to reach their target audience. Often, advertisers and brands are represented by an agency. They use [demand-side platforms \(DSPs\)](#) to configure campaigns, bid on impressions in real time, and measure campaign performance.

The **sell-side** comprises publishers and retailers selling advertising inventory, such as banner ads on web pages, video ads in video streaming, device app ads, audio ads in audiobooks and smart speakers, etc. They use [supply-side platforms \(SSPs\)](#) to handle the sale of the advertising space based on factors like deals done with DSPs, direct bidding between retailers and brands, retargeting of users, etc.

Commerce media is a specialized case of both the buy-side and sell-side, focusing specifically on retailers as publishers and brands as advertisers. Think about your shopping experience at Walmart, Booking.com, or Carrefour, for example. You search for products and are provided with products that match your semantic query. You also receive recommendations for similar products and complementary products. The goal of this AdTech system is to have you add items to your cart and checkout.

Commerce media platforms provide:

- **The customer journey:** The user's path through the retailer's site.
- **Recommendations:** The brands and products the user would like or use
- **Product catalogs:** Normalized catalogs of brands and products
- **Retargeting:** Namely, of the products viewed, the items in one's cart (Onsite of the retailer and offsite in the broader DSP ecosystem)
- **Dynamic content optimization:** Involves refactoring (resizing, recoloring, reskinning) product creatives to make them more appealing to the user in real time

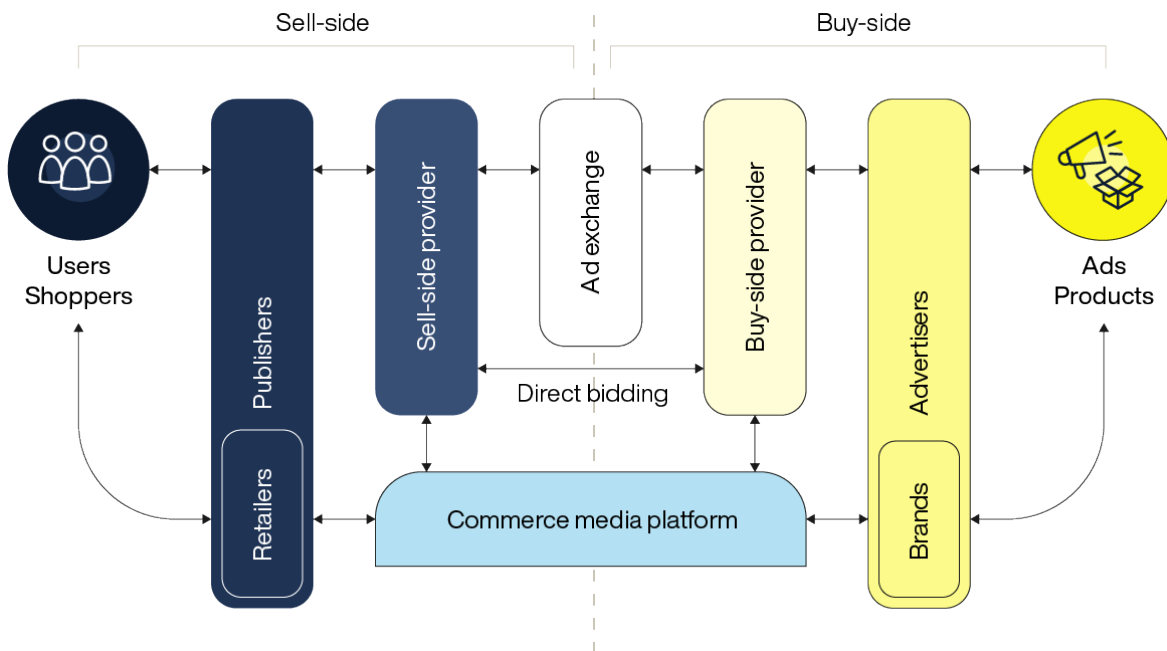


Figure 1: Real-time bidding process high-level flows.

Offline processing creates and updates user profiles, audience segments, etc. It typically aggregates and analyses large datasets from various sources like first-party systems, ad interaction events, transaction logs, and third-party data providers. This data undergoes cleaning, de-duplication, and transformation to ensure accuracy and consistency. Advanced AI and machine learning models are then applied to identify patterns, preferences, and behaviors that help define audience segments. These profiles are enriched with demographic, geographic, and psychographic attributes, enabling marketers to target specific groups more effectively in campaigns while maintaining compliance with data privacy regulations.

The sell-side

Publishers or retailers aim to secure the highest price possible for their inventory (ad spaces) on their sites and video content apps, done in real time to remain responsive to the user.

They use the services and technology of SSPs to manage first-party data, deals between publishers and advertisers, the preferred DSPs and ad exchanges, and direct bidding that bypasses the auction/ad exchange.

The scenario

When a user visits a page, views a video, or opens an app, the publisher or retailer uses the SSP to:

1. Retrieve, enhance, and anonymize the publisher’s user profile
2. Fetch deals associated with the publisher or retailer
3. Prepare either a direct request or initiate an auction
4. Forward this request to an ad exchange and on to DSP or directly to a deal’s ad server
5. Receive and display the ad

This process must happen in the blink of an eye for a seamless user experience and optimal value from advertising inventory.

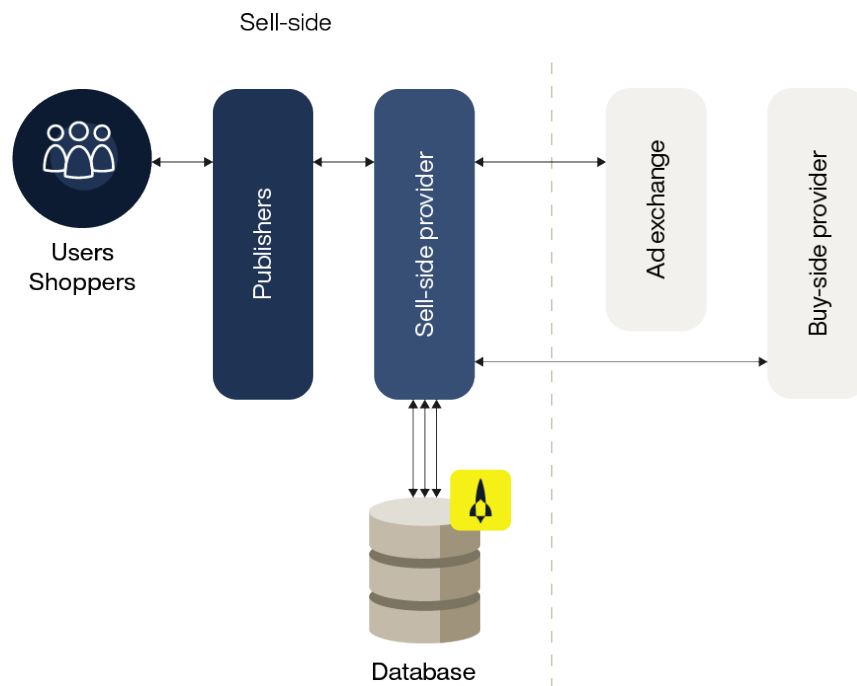


Figure 2: Sell-side providers store profiles and deals.

Storing sell-side user profiles and deals requires swift data access to ensure a responsive web page/app. Aerospike Database is an excellent choice for this due to its low latency, comprehensive data model, scalability, and integration capabilities.

The buy-side needs to optimize ROAS

Advertisers or brands consider their target audiences and how to reach them with the most effective ads when setting up campaigns. They define their campaign objectives, goals, targeting options, budgets, and bidding strategies.

Campaign configuration involves defining unique parameters for each ad creative, ensuring the right message is delivered to the right audience at the right time. A campaign is subdivided into smaller execution plans called Lineitems. Each Lineitem has its own configuration with connections to creatives (ads) and audiences.

The **creatives**, which can encompass images, videos, or HTML5 banners, are carefully crafted to resonate with target audiences, while Lineitems provide the necessary ad format and delivery options.

Audiences, or more specifically, audience segments, categorize users based on their interests, behaviors, or demographic characteristics, ensuring that ads reach the most relevant audience. Segments are essential for a collection of user profiles that meet specific criteria. By leveraging these profiles, ads are tailored to particular segments of users, increasing the likelihood of engagement and conversion.

Measurement and reporting are critical components of a successful RTB campaign. They provide insights into ad performance, user behavior, and return on ad spend (ROAS). The DSP provides a wealth of data-driven intelligence, informing future campaign optimization and driving long-term success in the ever-evolving digital landscape.

Programmatic advertising is a real-time, data-driven process

Advertisers and brands want to present their ads to internet users interested in the lowest price in real time. The best prices are struck employing the most data on each side inside SLAs.

1. **Bid request:** The process begins when a user visits a website or opens an app. The publisher (the SSP) sends a bid request to the DSPs and ad exchanges. This request contains valuable information such as user data, location, device type, and the context of the page or app. It's like the SSP announcing to advertisers: "Hey, we've got a user here! Who wants to bid for an ad spot?"
2. **Bid response:** Advertisers or DSPs analyze the bid request data to decide whether they want to compete for that ad impression. If they do, they send back a bid response. This includes their bid amount and details about the ad creative they want to display. It's their way of saying, "We'll pay X amount to show this ad to the user."
3. **Bid win:** After collecting all bid responses, the SSP evaluates them and selects the highest (or most relevant) bid. This is the winning bid. Essentially, one advertiser triumphs, securing the right to serve their ad to the user.
4. **Serve creative:** Once the bid is won, the creative (the actual ad content) associated with the winning bid is delivered to the user. This could be a banner, video, or native ad format. The user sees the ad, and the content resonates with them.

It's a swift, data-driven process that takes milliseconds and happens in real time as the user loads a page or interacts with an app. The cycle is programmatic advertising.

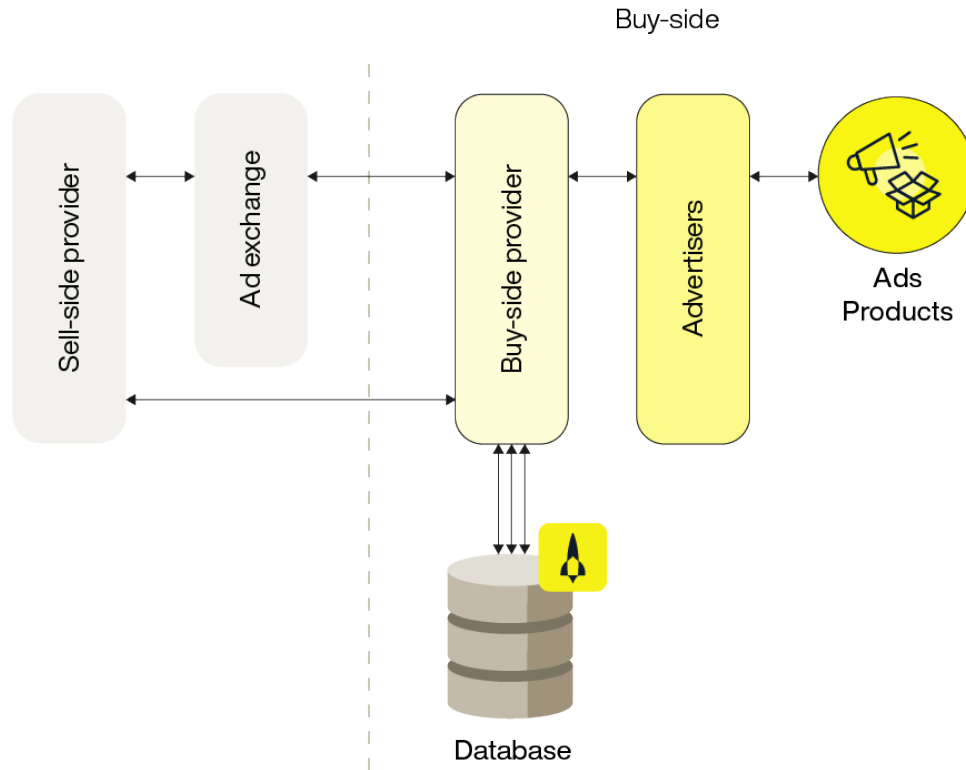


Figure 3: Buy-side providers store profiles, Lineitems, segments, and creatives.

Each bid request requires retrieving the matching user profile, candidate campaigns/line items, selecting the “best” campaign/line item, selecting the “best” creative (ad) to display, and calculating the bid price, all in less than 100ms. A DSP will receive and respond to tens of millions of bid requests in the blink of an eye.

In technical terms, that is a throughput of tens of millions and a bid latency of less than 100ms. Responding to a bid request will require up to hundreds of database operations (though ideally 10-20); in Aerospike, that would be a total latency of 9-12ms, leaving 85-50ms for the DSP’s proprietary bid processing.

The importance of commerce media

We are familiar with search media (Google, YouTube, etc.) and social media (Facebook, Instagram, TikTok, etc.). Commerce media refers to advertising on e-commerce platforms or applications that attract buyers during their purchase journey. It involves displaying ads to shoppers who are actively browsing or transacting in these online marketplaces. Early adopters of this strategy were primarily retailers, who have since coined the term “retail media” for advertising on their e-commerce platforms.

Retail media is a subset of commerce media, focusing on advertising within e-commerce platforms operated by retailers. Early adopters, including Amazon and Walmart, generated significant advertising revenue. By 2025, commerce media is expected to be worth \$290 billion in total ad spend, with retail media accounting for \$130 billion. This indicates the growing importance of these advertising formats within the digital marketing landscape.

Commerce media is revolutionizing digital advertising in the following ways:

1. It combines online and offline data, allowing for real-time optimization and measurement. This is a significant improvement over the current state, where measuring ROAS can take months, and optimizations can't be made on the fly.
2. Retailers and brands integrating with a commerce media platform have the potential to unlock billions of dollars currently sitting in brand marketing budgets.
3. Commerce media provides high-fidelity data for a closed-loop environment, enabling advertisers to measure and optimize simultaneously. This contrasts with current digital marketing, where data from different sources are often disjointed and difficult to reconcile.

Commerce media: A special case

Commerce media platforms, on behalf of their retailers, can leverage capabilities, such as Aerospike Graph or Aerospike Vector Search (AVS), to identify and engage with the most likely users to complete a purchase. For example, those with items in their cart but haven't purchased yet, or those who showed interest in your brand, but didn't take the next step.

User journeys through retailers' sites along with their past buying patterns create behavior graphs of product relationships and product similarities. Commerce media platforms, as a result, can create highly personalized campaigns and recommendations for retailers that resonate with customers.

By analyzing the buying patterns of loyal customers, you can identify users with similar interests and preferences and recommend products from affiliated brands or categories that are likely to appeal to them. AVS lets you quickly find the most relevant products and promotions. With graph queries, you can understand the underlying behavior and patterns that drive user engagement and find products that have an affinity with the product selected by the user.

Scenario highlighting graph and vector search needs

Justina, a 32-year-old Danish marketing professional with a master's degree, is an individual who embodies the perfect blend of style and sophistication. When she's not working or exploring new travel destinations, you can find her hitting the dance floor at salsa and bachata clubs or enjoying a refreshing winter bath to rejuvenate her senses. As an animal lover, Justina also spends quality time with her dog. With an annual income between 50k-85k euros, she's in a prime position to indulge in travel experiences that bring her joy.

Justina is attending a salsa festival in Greece and needs to sort out the logistics of her trip. She's shopping online for flights, accommodation, and an Athens tour to make the most of her time in the ancient city. With a passion for dance, Justina is also on the hunt for new gear to fuel her salsa passion, specifically, shoes, skirts, and tops that will allow her to move freely and confidently on the dance floor.

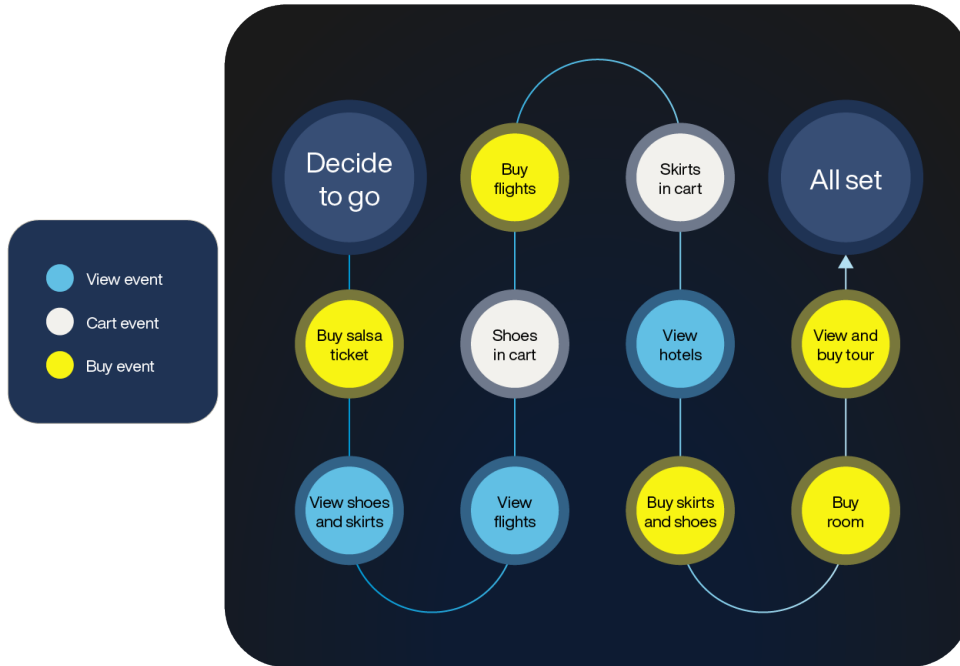


Figure 4: The e-commerce journey is an opportunity for advertisers at many steps.

She browses e-commerce websites; this is her journey (Figure 4).

Here is where Aerospike’s graph and vector search capabilities can come into play. These solutions enable retailers to offer targeted product recommendations based on Justina’s interests and preferences—in this case, highlighting popular Latin dance wear brands or suggesting complementary travel accessories that align with her festival-going spirit.

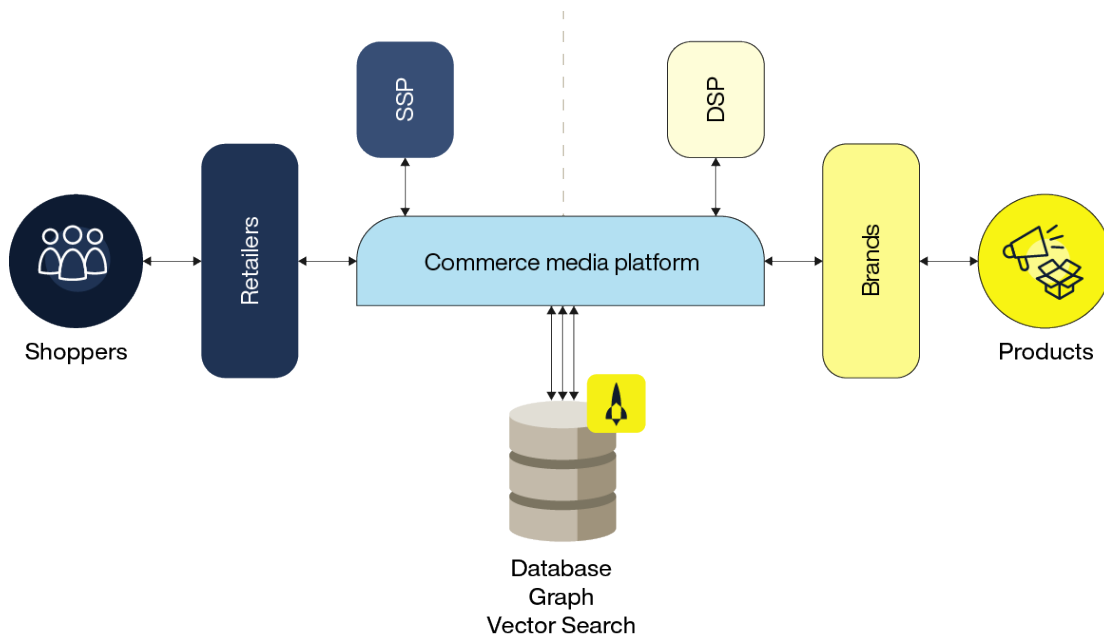


Figure 5: The commerce media pipeline.

Why Aerospike for commerce media

With Aerospike Graph and Vector Search, retailers can now unlock opportunities to connect with Justina and other like-minded individuals who share similar interests and preferences, such as a passion for travel and fitness. By leveraging targeted marketing campaigns that speak to her lifestyle and interests, businesses can tap into this valuable demographic and drive conversions that align with her spending power.

For Justina, she is targeted and recommended products that she is interested in right now, not something she has already bought or that is of no interest to her. With Aerospike, this can be done in real time as she browses e-commerce sites.

For retailers, leveraging the capabilities of a commerce media platform that uses Aerospike Database, Aerospike Graph, and AVS creates a more seamless and personalized shopping experience that drives loyalty, retention, and revenue growth.

Offline data needs

Offline processing of user behavior data from the internet is done by artificial intelligence (AI), and other data science techniques process vast amounts of data to create or update user profiles, customer journeys, and audience segments.

Processing is usually every six hours, every day of the year, to ensure that inferences from data are fresh and relevant. It processes petabytes of data at a time and involves complex data pipelines with several intermediate stages.

User profiles and audiences are preeminent

Segmentation has long been foundational to real-time AdTech, powering everything from targeting and bidding to personalization. While traditional techniques still apply, vector search now supplements them with a more powerful and flexible approach to segmentation, along with other classification use cases. It uses embeddings, which represent users or behaviors as multi-dimensional numeric representations based on recent actions and signals.

These embeddings are stored in Aerospike alongside the user's real-time profile and audience membership. A user profile store may begin with basic data—browsing history, purchases, location, device, and demographics—but becomes more valuable when enriched with derived insights and vector representations that capture behavioral patterns and intent.

Because Aerospike offers both sub-millisecond read latency and vector search in the same database, applications can instantly retrieve a user's current profile while finding similar users or segments using [approximate nearest neighbor \(ANN\) search](#). This enables marketers to scale campaigns with fresh, high-affinity audiences in real time without waiting for offline pipelines or duplicating data across systems.

User profiles and audience segments are stored in Aerospike, along with embeddings that classify the user or segment. Aerospike allows both real-time access to the user profile and segment (in under 1 ms) and vector search of similar profiles and segments for audience creation, all in the one database, eliminating stale data.

The customer journey should be captured in a graph

A customer journey typically starts with awareness, where a potential customer discovers a brand or product through various channels such as social media ads, search engine results, or word-of-mouth recommendations. This is followed by consideration, during which the customer researches and evaluates options, comparing features, benefits, and reviews. Next comes the purchase stage, where the customer decides to buy, often influenced by personalized offers or seamless user experiences. After the purchase, the journey continues with post-purchase engagement, including follow-ups, customer support, or loyalty programs that build trust and encourage repeat business.

A customer journey is a graph of activity. Over time, it represents a user's interests in products and the relationships between products; it is a knowledge graph. Customer journey graphs become interconnected via products viewed or purchased and sites visited. This shopper graph is rich in detail and can be used for product recommendation, audience segment creation, and campaign analytics.

Justina's journey as a graph is rich with data describing interests, related websites, related products, and related behaviors.

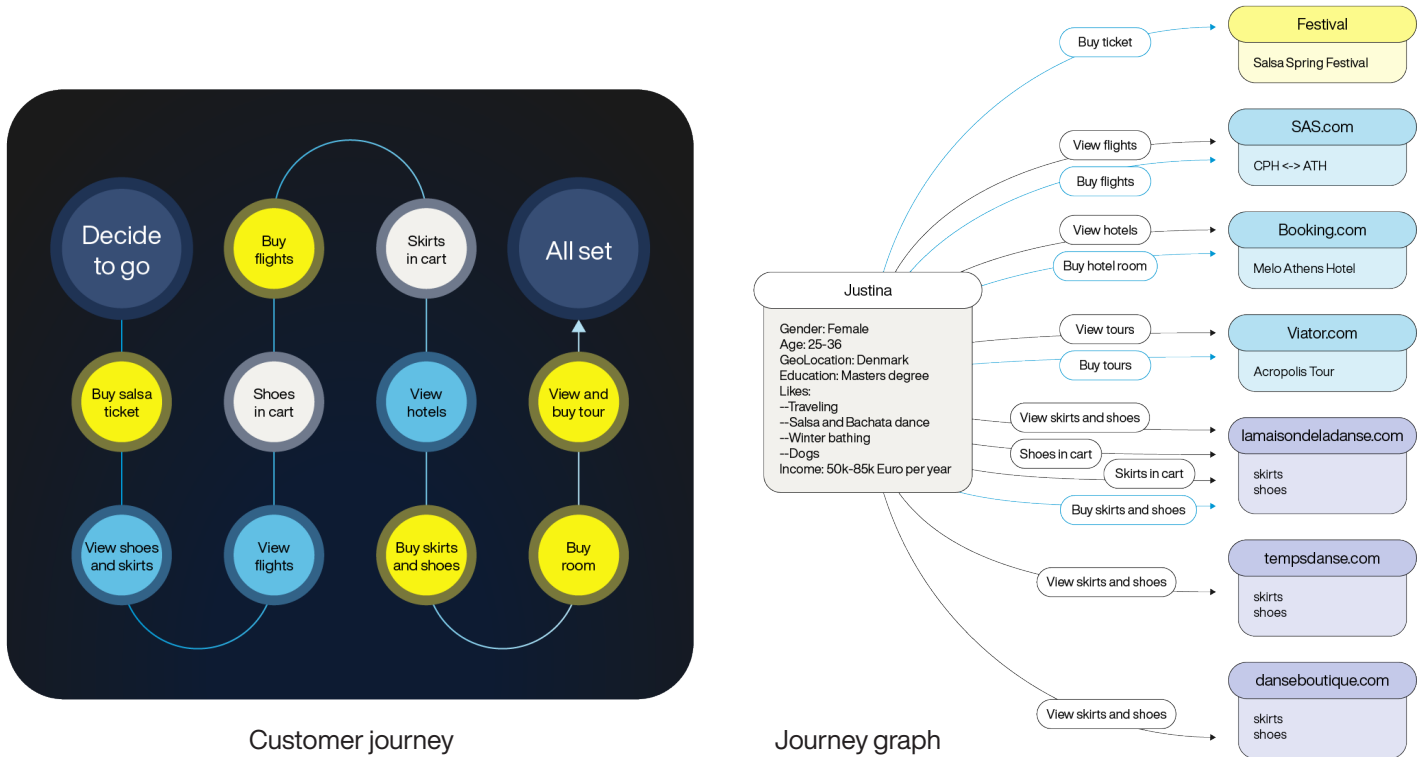


Figure 6: Journey graph based on customer journey.

Aerospike Graph can capture a customer's journey in real time on a massive scale. Moreover, billions of individual customer graphs become linked into a comprehensive graph of shoppers. Coupled with the Aerospike Database, you have the ultimate [user profile store](#).

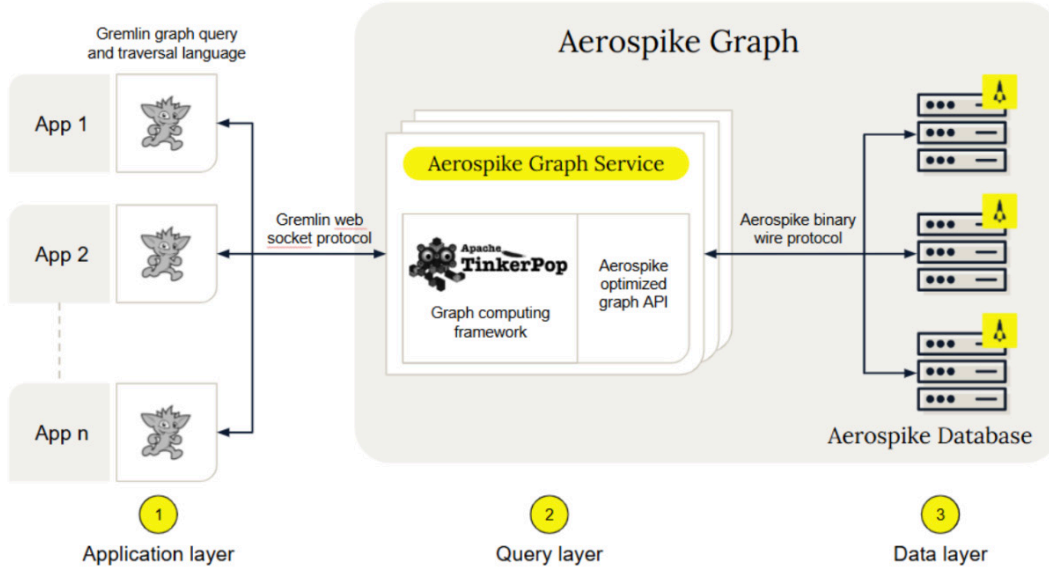


Figure 7: Aerospike Graph architecture with a scalable query layer to add more compute to increase throughput separate from the data layer.

Product catalogs are best served by both document and vector models

Commerce media platforms usually have a normalized product catalog. That is, products from all brands are stored in a master catalog, and access is constrained by cohorts of brands and vendors. The number of products is in the billions.

Products in the catalog have data about the items’ names, descriptions, colors, etc., as well as embeddings of product attributes for the semantic search of products. The catalog uses both a document store and vector indexing.

Product similarity can be determined by comparing embedded representations, which capture structured data like style, size, and category as well as unstructured data like product images. For example, a user viewing a red women’s shoe might be shown visually similar options, even if they’re not explicitly labeled “red,” thanks to vector embeddings.

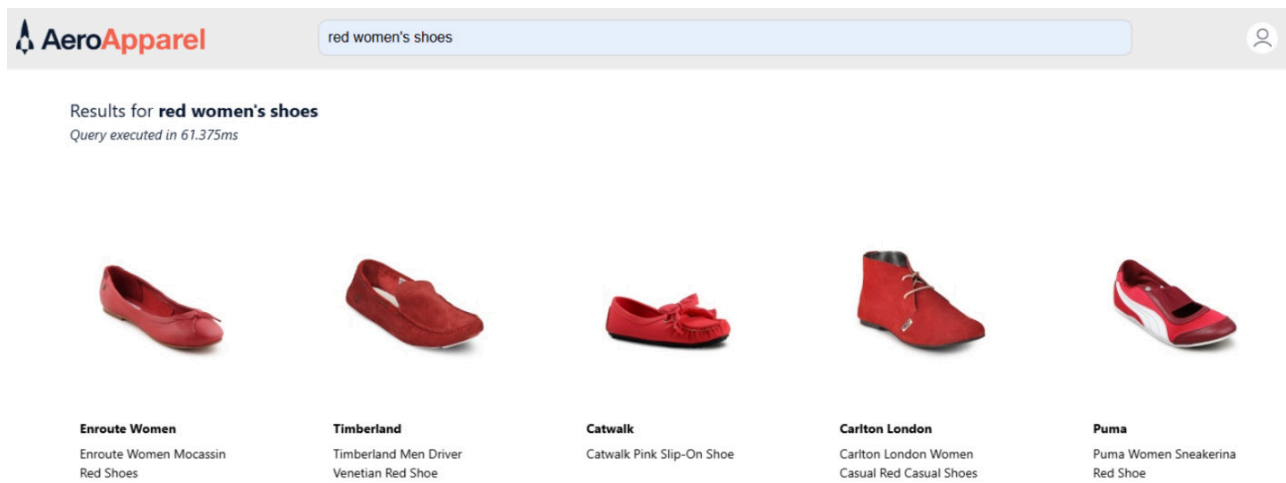


Figure 8.1: Aerospike Vector Search demonstration with apparel data showing “Catwalk Pink Slip-On Shoe” for a “red women’s shoe” as a match based on vector embeddings of the image showing, indeed, how similar it is in appearance to the search term despite its name.

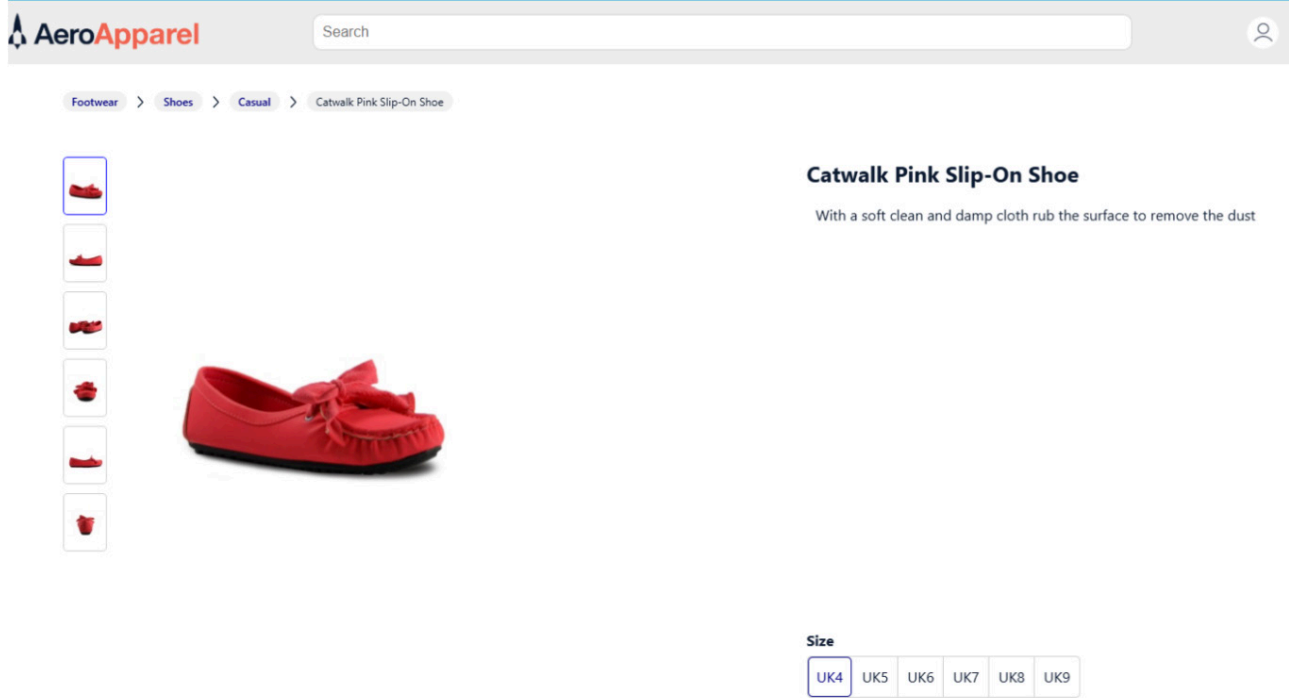
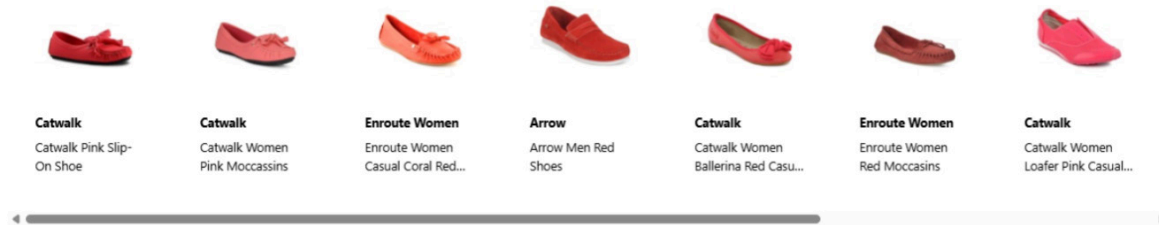


Figure 8.2: Aerospike Database with document-stored data being served up: name, description, image, sizes.

Product affinity is a graph query. People “also bought” these shoes, as well as the ones you’re looking at.

Similar items



Also bought

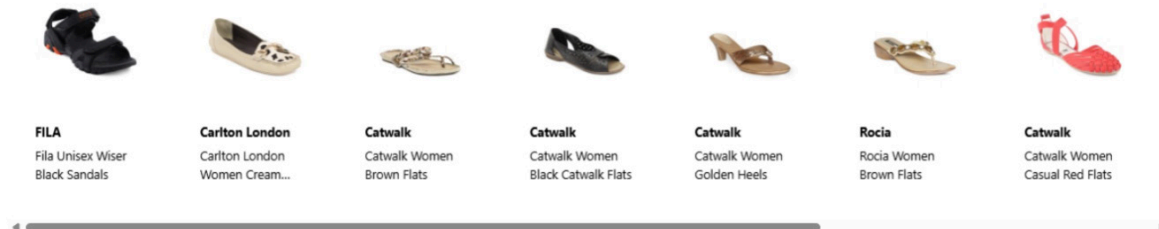


Figure 8.3: AeroApparel demonstration has ‘Similar items’ served up via Aerospike Vector Search and ‘Also bought’ served up via Aerospike Graph based on other consumers’ purchases.

Aerospike provides all three needs: document, vector, and graph.

Audience creation

An audience is a logical grouping of audience segments used for campaign targeting. Segments are user profiles with similar demographics and/or interests.

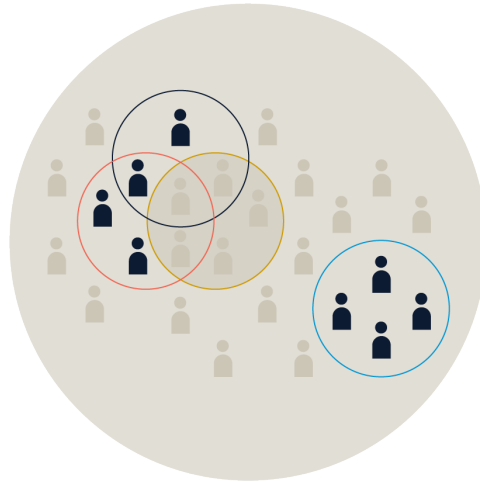


Figure 9: Audience groupings.

Segments are groups of similar users, and the vector or embeddings of user profiles are used to determine similarity. For example, Justina has a friend who is in the same age range and salary bracket and is interested in Latin dance; a vector search based on demographics and interests would likely group them into the same segment.

A combination of demographics and interests and each user’s individual “shopper graph” can be embedded and used to enhance accurate vector search, ensuring the accuracy of the segmentation.

Types of audiences/segments

- **Similar:** The most basic segment of similar user profiles
- **Retargeting:** An audience of one user profile (e.g., used to retarget a user with items in a cart)
- **Contextual:** A collection of websites that have similar content or context. When a user visits a site in the contextual segment, they are targeted with ads based on the context
- **Commerce:** Based on a set of rules
- **Look-a-like:** A seed collection of prototypical users expanded to find the k nearest (similar) neighbours (kNN)

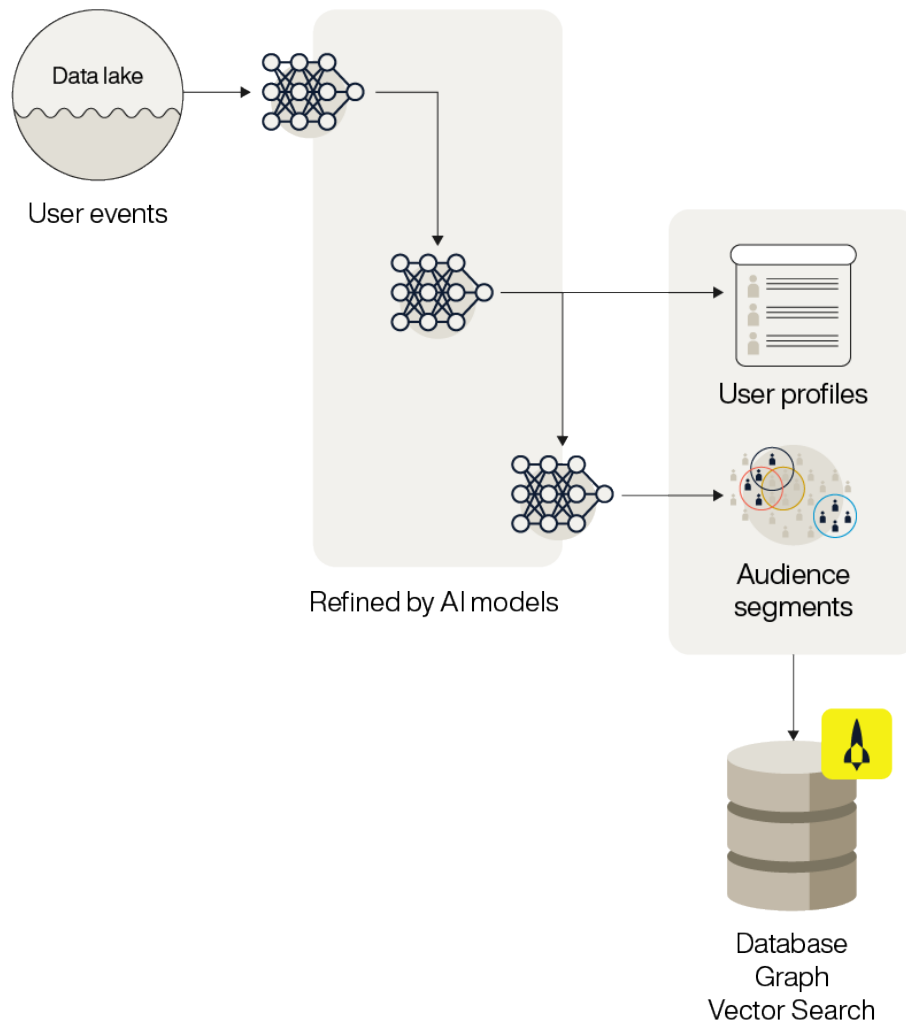


Figure 10: Audience segments and user profiles.

Aerospike Database, Vector Search, and Graph provide ideal storage and retrieval for user profiles and audience segments. As new data arrives, indexes or graphs do not have to be rebuilt, and Aerospike can update data without affecting real-time operations.

Aerospike is ideal for AdTech real-time use cases

Aerospike's unique architecture and features make it the perfect database for AdTech. Here are some key reasons:

1. **Low latency:** Aerospike's [Hybrid Memory Architecture](#) ensures sub-millisecond read and write latencies, enabling real-time decision-making.
2. **High throughput:** With its ability to handle millions of transactions per second, Aerospike can effortlessly manage the high volume of bid requests and frequency capping in RTB on the buy-side. On the sell-side, Aerospike can quickly supply data to enrich impressions sent to ad exchanges and publisher networks.
3. **Scalability:** Aerospike's linear scalability allows it to grow with the increasing demands of buy-side and sell-side, ensuring consistent performance.
4. **Reliability:** Aerospike's [strong consistency](#) and fault-tolerance features ensure data integrity and availability, even in the face of hardware failures.
5. **Graph:** Graphs for users' journeys through a website and product purchases create valuable knowledge graphs for product recommendations and targeting in real time.
6. **Vector search:** Similarity search is essential for targeting and recommendation in real time.

One database for AdTech

A single database incorporating document stores, vector databases, and graph databases offers numerous advantages. One primary benefit is that it eliminates the need for separate, standalone databases for each use case, reducing complexity and increasing scalability. With a unified database, you can store and manage data of various types (e.g., documents, vectors, and graphs) within a single repository, streamlining site reliability engineering and reducing the cost of maintaining multiple databases.

The advantages of having all these features in one database extend to availability, real-time performance, massive scale, and data consistency. With a unified database, you can ensure that your data is always available, as the entire system can continue to operate even if individual components fail.

Moreover, this setup allows for real-time data access and updates, making it ideal for applications requiring instantaneous data processing and analytics. Furthermore, the massive scale of such a database enables it to handle large volumes of data efficiently, eliminating the need for data copying or creating separate databases.

Strong consistency is also ensured, as all updates are propagated uniformly across the system, guaranteeing that data remains in sync across all nodes. This integrated approach eliminates the hassle and costs of maintaining multiple databases, making it an attractive solution for applications requiring a wide range of data management capabilities.