



Applying the Science of Search to Display Advertising with Aerospike Real-Time NoSQL Database

By Apurva Dalal, Vice President of Engineering at Komli Media

Display and search are the clear leaders in driving Internet advertising revenues. In fact, *The Economic Times* has reported that marketers across India, Australia, South Korea, Japan, and China are expected to grow their investments in advertising via online display and search media to nearly \$60 billion by 2017¹.

At Komli Media, where we deliver Asia Pacific's leading real-time digital technology platform, we believe there is an opportunity for display advertising to surpass search ads. However, doing so requires closing a significant gap in click-through rate (CTR) that exists today between the two ad types. Google gets approximately a 2% CTR on search while the average CTR for display is only a 0.1% CTR.

As one of the few companies in the world to provide an integrated real-time display, mobile, video, social, and search platform for advertisers, we are working to close that gap. In the process, we are innovating on cutting-edge technologies—on top of Hadoop, NoSQL, Netty, Kafka, machine learning frameworks—to implement infrastructure that processes tens of thousands of requests per second, delivers analytics on terabytes of transactional data, and matches the user with the best possible ad for the most optimal return on ad spend for the advertiser.

One product that is playing an increasing role in ensuring that we can deliver digital advertising at scale is the Aerospike real-time NoSQL database. To understand its role, it is helpful to begin by looking more broadly at how we are working to make display advertising as effective as search advertising.

Bringing the Science of Search to Display Ads

At Komli, we innovate to deliver on sophisticated algorithms, the power of analytics and data, and real-time technologies as a state-of-the-art system that revolutionizes how the display ad world works. Some of the most notable and complex components are:

- **Real-time bidding (RTB)** – Similar to a stock exchange, hundreds of buyers are bidding in real-time to present an ad to a consumer on a Web page or mobile site. We will look more closely at RTB in the next section.
- **Ad price prediction** – Hand-in-hand with RTB is the need to decide within milliseconds the predicted value of an ad unit; you want to maximize on the spend return for the advertiser as well as ensure that the ad is most suitable for the user.
- **Big data** – At Komli, we generate terabytes of data each day including stats around ad impressions, clicks, impressions, bids handled, etc. This data is aggregated and crunched at scale to generate real-time insights for the advertisers.
- **Web-scale** – Internet advertising demands systems that can handle billions of requests per day and hundreds of thousands of requests per second.
- **Redundancy** – We strive to attain a 99.999% (“five nines”) availability goal and to that extent are replicated within and across multiple data centers for critical functionality.

The Central Role of Real-Time Bidding

Our goal is to help marketers reach millions of the right consumers across the Internet, at the right time, and our business model reflects that commitment. We are all about delivering performance advertising for our advertisers.

Real-time bidding enables us to take in inventory signals of a given ad bid, combine them with data about what these users are doing now/have done in the past, the audience segment these users belong to, and then serve them targeted advertising. Understanding our audience better than competing bidders and placing a bid in real-time based on what we know about each individual is the key. In some measure this is a data analytics/insights problem at a massive scale.



Identifying a Database for RTB Demands

The demands of real-time bidding translate into a specific set of requirements for the storage layer that stores real-time user, ad campaign and inventory signals. Because we're working with a range of data sources, we need sparse, distributed multi-dimensional data storage capable of handling both structured and unstructured data. We are working with tremendous amounts of data, so the ability to scale is key in terms of reads/writes throughput from this storage infrastructure. We also need sub-millisecond latency to read/write this data and bid better than competitors in real time, using the data. Last, but certainly not least, we have to satisfy the 99.999% availability

requirement across many data centers that are handling bid requests. The other criteria we included in our benchmarking exercise across data storage systems were around metrics for robustness, manageability, development effort, and community support.

Given that a large part of our stack is built with open source technologies, we looked at the established open source NoSQL database offerings in the market. We also evaluated Aerospike, a commercial-licensed NoSQL database, and at the end of the day, Aerospike emerged as the database that could best meet our demands for sub-millisecond latency, scale and reliability. For each of the incoming bid request that we want to handle, there is a need to look up past information around the user/inventory/campaign signals—any NoSQL DB would have to handle five-plus lookups in a 1-2 millisecond SLA. Now add the fact that we handle many billions ad requests in a given day. We found Aerospike met this set of stringent requirements better than all other open source NoSQL offerings.

“Also unique to Aerospike is the database’s architecture, which features native support for flash memory and solid-state drives (SSDs), as well as DRAM. This translates into the ability to achieve higher performance while also reducing the number of servers required by roughly ten-fold. That is an important factor as we continue to scale our system to meet growing demand.”

Apurva Dalal
Vice President of Engineering,
Komli Media

Solution Architecture for Internet Display Ads

The success of targeted display advertising relies heavily on the ability to link historical data and trend analysis with real-time data, such as user clicks, cookies, or geo-location data, in milliseconds to serve up highly targeted and relevant ads.

To this end, our solution architecture for display ads is an integrated set of transactional and analytical data management systems. On the analytics side, we have implemented an Apache Hadoop-based data pipeline cluster along with HP Vertica online analytical processing (OLAP) cluster(s) primarily used for reporting. On the transaction side, sitting at the edge is the Aerospike cluster, which combines the real-time data it directly receives with information and analysis from Vertica and Hadoop to respond to queries.

In the past, we ran all our software on bare metal hardware. Starting in 2012, we have been building out a private virtual cloud infrastructure. By having a single box with several virtual machine instances, we have been able to squeeze the maximum throughput out of our hardware while optimizing on the CPU/memory footprints. This has allowed us to build a hybrid environment with disaster recovery capabilities as well as the ability to stay agile (with rolling upgrades) in terms of our weekly/daily code deployments.

Ensuring 24x7 Availability

Earlier I mentioned the need for redundancy to ensure 24x7 high availability. Therefore we have implemented our solution architecture in multiple data centers: Asia, North America and soon Europe. The cross-data center replica-

tion in the Aerospike database enables us to maintain redundant data through two-way replication. So if there is a failure affecting one data center, we can immediately switch to the other to ensure continuity.

There are also some features within each of the two Aerospike database clusters to protect against downtime. These include replication among the database nodes in a cluster, automatic re-balancing across a cluster, and fault-tolerance.

Because the **Aerospike database is self-tuning**, it is maintenance-free. It just works. During upgrades we have reached out to the Aerospike support team, and their responsiveness has always pleasantly surprised us. Since there are fewer demands for database maintenance, we can devote more of our time and resources to enhancing and expanding on our services.

Conclusion

We believe that display advertising can begin to approach the click-through rates enjoyed by search ads today. Central to that effort will be the effectiveness of the real-time bidding platform and the data management software that supports it, including both analytical and transactional systems.

On the transactional side, databases must meet the demands of Internet advertising for real-time lookups, big data capacity, high throughput on writes as well as reads, extreme scalability, and 24x7 availability. Additionally, the diversity of structured and unstructured data is more effectively supported by newer NoSQL databases rather than traditional relational database management systems.

At Komli, the data storage infrastructure that is fulfilling these requirements is Aerospike, and we are already seeing a clear return on our investment.

¹*The Economic Times, March 29, 2013, "Online Advertising in Asia Pacific to Grow to \$60 billion by 2017," by Shelley Sing.*

Read more customer success stories @ www.aerospike.com/customers

2525 E. Charleston Rd #201 | Mountain View | CA | 94043
+1 408.462.AERO (2376) www.aerospike.com | sales@aerospike.com



AEROSPIKE