

Krux

Founded in 2010, Krux delivers data fabric for the consumer web. Krux's data management platform (DMP) helps companies protect, manage, and monetize across screens and sources. With Krux, companies give their consumers cooler, safer, faster, smarter web experiences for more than 700 million users worldwide. With Krux, consumers gain confidence that their favorite websites are operating under the plain light of day. Companies across the US, Europe and Asia have adopted Krux technology, including *The New York Times*, Bloomberg, Meredith, NBC Universal, Forbes Media, Triad Retail Media, Wikia, *Financial Times*, NBCNews.com, Apartments, and *The Wall Street Journal* Digital Network. Find out more at www.krux.com.

The Challenge: Always On, and Always Fast.

Krux provides real-time insight into its clients' audiences while protecting privacy, enabling them to deliver web pages faster and share relevant content to their viewers while ultimately delivering more lucrative advertising. Krux needs to work fast behind the scenes while managing a large amount of data that includes more than one terabyte of client data and 50,000 requests per second.

Over time, the company began maxing out the single machine that handled its old data store, even after upgrading the memory. Krux's application requires high availability and low latency, so the company needed a database solution that would allow it to deliver fast, actionable, profitable insights to clients while scaling its business. With a growing customer base requiring it to store more customer data, Krux needed a storage system that could expand horizontally to every new device and browser — especially as mobile platforms continue to expand rapidly.

The DataStax Solution: A Robust Fit.

Looking for compatibility with its Amazon EC2 infrastructure, Krux decided on a big data solution using Apache Cassandra and DataStax Enterprise after competing options, including an Apache HBase solution from Cloudera, could not demonstrate successful deployments.

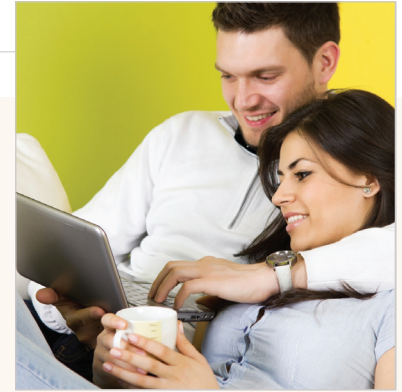
Krux emphasized the need for a robust database, and DataStax outperformed other solutions, particularly with its continuous availability, which is key to Krux's application. DataStax's tunable consistency model also fit better with Krux's infrastructure design philosophy, allowing developers to set the level of consistency required (from very strong to eventual consistency) depending on individual use cases on a per-operation basis. This enables developers to flexibly tailor their systems according to their unique business needs.

"We deployed test clusters of the solutions from both DataStax Enterprise and Cloudera to get a feel for the operational characteristics of both," said Paul Lathrop, senior infrastructure engineer, Krux. "DataStax Enterprise's performance and availability proved to be a much better match for the needs of our application and was easier to manage as well."

User Data Store and Beyond: DataStax Supports Three Key Use Cases.

Krux uses DataStax Enterprise to support three key use cases — its user data service, its ETL-like system for offline analysis and reporting, and a dynamic audience generation service that allows its clients to perform content targeting.

Krux's user data service, its most important use case, organizes the full range of audience information and lets clients dynamically model and manage their audience segments for better ad revenue and content targeting. Krux processes about 50,000 requests per second and can turn a user action on a client web site into an actionable segment in less than 75 milliseconds.



Company

Krux

Data Size

*1 Terabyte
50,000 Requests per second*

Business Challenge

Provide high availability and low latency in a scalable data store for Krux's user data service

Solution

DataStax Enterprise allowed Krux to deploy a robust data store solution compatible with its Amazon EC2 infrastructure and support other key use cases

"DataStax provides a valuable tool for storing large quantities of data that don't fit in memory, but still serving that data with low latency."

*Paul Lathrop
Senior Infrastructure
Engineer
Krux*

DataStax Enterprise serves as the backend data store for the User Data Service. Krux allows its customers to define audience segments that are processed at varying frequencies throughout the course of a given day using regularly scheduled map-reduce jobs. Each such job results in multiple, concurrent updates of varying sizes (millions of rows on the low end for one job to 100s of millions of rows on the high end) to the Cassandra data store.

"On the read side, Krux's front-end infrastructure (JavaScript or iOS/Android mobile SDK) calls a real-time web-service that queries the User Data Store when a user visits a client property (web page, mobile web page, mobile App)," said Paul Lathrop, senior infrastructure engineer, Krux. "We take advantage of Cassandra's key cache and the file system cache to ensure we mostly serve data out of RAM...Latency is very important on the read side and Cassandra delivers very fast performance."

Krux also operates an ETL-like system used for offline analysis and reporting. DataStax Enterprise's Apache Hadoop integration ensured that the platform was responsive enough to provide dynamic generation of audience segments for clients using content targeting. The robust, scalable back end and service provided by DataStax allows Krux to deliver many benefits to its customers. Clients can monitor how third parties collect data and ensure practices meet the terms and conditions of contracts, or block unwanted data leakage while quantifying its cost for page latency.

"DataStax support has been very helpful navigating challenges around capacity planning and tuning," said Lathrop. "We can get connected with a support engineer immediately, and DataStax has kept up with our needs in terms of a large data set and low latency reads."

Krux clients can boost operational efficiency by managing tag lifecycles centrally, leveraging audience profiles to improve revenue, and also increase page performance by more than 30 percent. Using a holistic audience databank managed on the Krux platform, clients can taxonomize their audience information such as registration data, page visits, clicks, and path-to-site. This allows clients to dynamically model and define their audience segments in order to receive a true view of their media and audience assets.

"Overall we've been very pleased with DataStax Enterprise," Lathrop says. "It has more than kept up with our data storage needs as our customer base continues to grow and as new devices and mobile platforms come online. DataStax is a valuable tool for managing large amounts of data in an online application with low latency."

"We can get connected with a [DataStax] support engineer immediately without needing to escalate."

Paul Lathrop
Senior Infrastructure
Engineer
Krux

About DataStax

DataStax powers the apps that transform businesses. DataStax powers big data apps for more than 300 companies, including start-ups and 20 of the Fortune 100, with its flexible and massively scalable big data platform built on Cassandra, through multi-data centers.

DataStax Enterprise delivers enterprise-ready Cassandra, then goes one step further by integrating the best of breed big data technologies — Apache Hadoop for analytics, and Apache Solr for search across multiple datacenters and the cloud.

Top companies such as Adobe, eBay, Netflix, and HealthCare Anytime, rely on DataStax to transform their businesses. Based in San Mateo, Calif., DataStax is backed by industry-leading investors: Lightspeed Venture Partners, Crosslink Capital and Meritech Capital Partners. For more information, visit <http://www.datastax.com/> and Follow @DataStax.



About Mainstay

Research and analysis for this business impact study was conducted by Mainstay, a leading provider of independent value assessment and IT strategy services. For more information, visit www.mainstaycompany.com.

© 2013 Datastax. (07.13)



777 Mariners Island Blvd #510
San Mateo, CA 94404
650-389-6000

DataStax powers the big data apps that transform business for more than 250 customers, including startups and 20 of the Fortune 100. DataStax delivers a massively scalable, flexible and continuously available big data platform built on Apache Cassandra™. DataStax integrates enterprise-ready Cassandra, Apache Hadoop™ for analytics and Apache Solr™ for search across multi-datacenters and in the cloud.

Companies such as Adobe, Healthcare Anytime, eBay and Netflix rely on DataStax to transform their businesses. Based in San Mateo, Calif., DataStax is backed by industry-leading investors: Lightspeed Venture Partners, Crosslink Capital and Meritech Capital Partners. For more information, visit DataStax.com or follow us on Twitter @DataStax.