

Kaminario Cuts Data Mining Process by Half

The Challenge

A leading corporation in web-based data mining collects, distills and disseminates data from the Internet to provide the knowledge needed to enable its customers to make smart business decisions. It harvests data (structured and unstructured) through web crawlers and maintains more than 50 million database files. Its applications include person/position/company data discovery and aggregation; sentiment analysis and market mapping; and brand awareness analysis.

In one of its applications, the company was matching and cleansing over 7 million records, the quantity of which was increasing by 10 percent per month. The program was providing large corporations with “identity resolution” and “data enrichment” of their customer files. Typically this data “cleansing,” which involves merging and purging the data as well as checking for errors, can take more than a week, an unacceptable turn-around time for most of the company’s customers. For the company to be competitive in the data mining market, it needed to shorten the cycle time required to process this data. Also, while processing the data, it needed to be able to find errors and make corrections quickly to meet its customers’ requirements. The company’s match-and-merge application runs multiple parallel processes, so any solution that it adopts to accelerate this application would need to perform well in that environment.

The Solution

The company chose Kaminario’s K2 DRAM storage appliance to reduce the processing time of its match-and-merge application. The Kaminario K2 quickly delivered extreme performance improvements to its parallel processes and data access. After one hour of installation and setup, the Kaminario K2 was in place and processing the customer files through the company’s application. The Kaminario K2 reduced the amount of time that it took to process the data and check for errors from more than one week to only three days. This 50 percent improvement in application processing time also cut data delivery time in half. Kaminario K2 supports industry-standard storage protocols that made the transition transparent to both the customer’s applications and its Microsoft® SQL Server® backup environment.

“Kaminario has shortened our data processing cycle by 50 percent, enabling us to deliver information faster to our customers.”

— CTO, a Leading Data Mining Company



Extreme Performance

The Kaminario K2 is the fastest storage performance appliance on the market today, allowing you to do more, faster. Extreme performance means low latency and high throughput, resulting in increased productivity and responsiveness as well as greater user and customer satisfaction.

- IOPS: 300,000 to 1,500,000 IOPS
- Throughput: 3.2 GB/s to 16.0 GB/s
- Consistent performance for any workload, with no degradation over time

High Availability

K2 boasts full N+1 hardware redundancy and thus maintains operational status and data availability with no single point of failure. Fast and automatic recovery means reducing costs related to data loss and unavailability, including wasted IT services time in rebuilding data stores.

- Redundant system hardware and uninterrupted power supply (UPS)
- Internal data protection with mirrored hard disk drives
- Automated hardware and data recovery while maintaining data integrity

Modular Scalability

K2 enables cost-effective growth – so you can start small and grow as you need more capacity and/or performance. K2 also maximizes storage utilization and extends the life of your storage infrastructure, increasing ROI and reducing TCO.

- Modular, scale-out architecture
- Scale performance and capacity independently
- Blade-based, open, high-end, industry-standard components

Easy Installation and Management

K2's industry-standard components make it easy to install and maintain. Its storage system management makes it easier for IT staff to manage and control. K2 enables IT organizations to reduce operations, maintenance and service costs.

- Standard block device, fully transparent — no change to applications, database or data
- Single system manageability
- Automatic data distribution quickly delivers extreme performance out of the box
- Fast deployment — less than one day to install



Kaminario K2

Kaminario K2 DRAM storage appliance delivers extreme performance for accelerating critical business applications. Based upon Kaminario's unique Scale-out Performance Architecture™ (SPEAR), the Kaminario K2 supports millions of IOPS and tens of GB/s of throughput, eliminating storage I/O bottlenecks. Organizations of any size can move their most strategic applications to Kaminario K2 appliances without modification or disruption. Kaminario is application, database and file-system agnostic. Kaminario K2 delivers enterprise-grade high availability, scales to businesses of any size, and is easy to deploy and maintain; installation and configuration can be done in less than one day. Kaminario is simply faster.

Kaminario, Inc.

275 Grove Street, Suite 2-400

Newton, MA, USA 02466

Tel: 1-877-98-CALL-K2

Email: info@kaminario.com

www.kaminario.com