

SeaMicro SM10000 Servers Increase eHarmony's Available Compute While Reducing Total Cost of Ownership

eHarmony (www.eharmony.com) is a pioneer in using relationship science to match singles seeking long-term relationships. Its service, available in the United States, Canada, Australia, the United Kingdom and Brazil, presents users with compatible matches based on key dimensions of personality that are scientifically proven to predict highly successful long-term relationships. An average of 542 eHarmony members marry daily in the United States as a result of being matched on the site*.

Each day, eHarmony sifts through data provided by its millions of members to find the best possible matches. Cormac Twomey is responsible for implementing the matching algorithms and manages the extensive compute infrastructure required to find compatible matches among eHarmony's users. The data set is large and complex, and the algorithms used are extremely sophisticated.

To achieve matching on this massive scale, the application developers at eHarmony turned to Apache Hadoop. Hadoop is an open source technology that allows many small independent servers to work together as if in a cluster to solve big complex problems. Hadoop enables applications to work with thousands of compute nodes and petabytes of data. Initially, to meet its compute requirements, eHarmony turned to a leading cloud provider and rented compute by the hour. The cloud environment provided a flexible but very expensive solution for eHarmony's Hadoop cluster.



As eHarmony's membership grew exponentially, it sought a more cost-efficient solution for its Hadoop work. After extensive testing, eHarmony purchased the SeaMicro SM10000™ high density, low power server with a configuration that enabled its Hadoop application to complete its run in exactly the same time that it had been taking in the cloud. In comparison with its previous four hours per evening cloud-based rental, eHarmony reduced its operating expenses by tens of thousands of dollars a month, and its TCO by more than 74 percent.

"Each month, we were paying huge amounts for a few hours per night of compute in the cloud," said Twomey. "We brought the SeaMicro system in-house, and this allowed us to save tens of thousands of dollars each month, complete our Hadoop run in the same elapsed time, and use the rest of the day of compute for free."

In addition to the TCO savings, eHarmony benefited from the following SeaMicro capabilities:

- **Additional compute:** For less than they were paying for a few hours of cloud compute, eHarmony was able to purchase the SeaMicro SM10000 and have it running 7 x 24 in its facility. This gave eHarmony the opportunity to run its Hadoop jobs more often without any additional costs.
- **Reliable completion time:** In the cloud, the amount of time it took Hadoop to complete the run varied widely, making it difficult to predict when the matching data could be delivered to other parts of the organization. This variability stems from the fact that in the cloud, CPUs, DRAM, disk, and network bandwidth are shared among customers competing for resources. On the other hand, inside the SeaMicro SM10000, resources were dedicated to eHarmony, and the system was able to reliably replicate run times. This allowed the operations team

*2009 U.S. survey conducted for eHarmony by Harris Interactive®

to keep its commitments to other parts of the organization and even offer SLA's to colleagues.

- **Eliminated data upload charges:** Cloud providers charge customers a fee to upload data. Loading data into the SeaMicro system was quick and easy and without charge.
- **Accelerated application performance:** eHarmony was able to leverage SeaMicro's 512 compute nodes and its high performance fabric to reduce data retrieval latency between its Hadoop and memcached applications by running both in the same SeaMicro system.
- **Simple management:** With 512 Intel® Atom™ processors inside the SeaMicro SM10000, the management, cabling, and switching hassles of building a cluster were removed. The SM10000 provided a single box Hadoop cluster that was easy to install and simple to manage.

Twomey concluded, "In the end, it seemed like a no-brainer. We purchased SeaMicro servers and immediately reduced our operating expenses, delivered our Hadoop work more reliably, impressed our colleagues with an internal SLA, and had more compute time available to refine and improve our methodologies. I wish more decisions were this easy and paid dividends this quickly."



SeaMicro SM10000 without covers illustrates its compact design.