

# Making Better Matches with Intel® Xeon® Processors

**Intel® Xeon® Processor E5 Family**  
**Big Data**  
Cloud Computing



eHarmony®

“With this new Spark\* programming framework, a new version of Cloudera CDH\*, and a robust infrastructure built on the latest Intel® Xeon® processors, we can do more personalized affinity matching and produce better results for more people.”

– Jonathan Morra,  
Principal Machine Learning Engineer,  
eHarmony

Founded in 1997 by clinical psychologist and marriage counselor Dr. Neil Clark Warren, eHarmony depends on cutting-edge technology to help produce long-lasting, meaningful relationships. The company uses robust cloud environments based on the Intel® Xeon® processor E5 family and running Cloudera CDH\* software to analyze a tremendous volume and variety of data so it can deliver potential matches to millions of people every day. The company upgraded its version of Cloudera and refreshed servers to the latest Intel Xeon processors to support a new software framework. This cloud environment should accommodate more complex analyses and ultimately produce more personalized matches that improve the likelihood of relationship success.

## Challenges

- **Increase performance.** Accelerate analytics so eHarmony can deliver millions of new matches every day.
- **Support innovation.** Upgrade software and hardware in the cloud environment to support a new software framework that will help boost performance and enhance matching capabilities.
- **Improve scalability.** Accommodate a rapidly growing volume and variety of data to produce better relationship matches.

## Solution

- **Private cloud based on Intel Xeon processors.** eHarmony uses a cloud environment based on Cloudera CDH software. The infrastructure includes servers equipped with the Intel Xeon processor E5 v2 family and Supermicro\* servers with the Intel Xeon processor E5 v3 family.

## Technology Results

- **Strong performance.** By using a new software framework with the latest Cloudera software running on servers equipped with Intel Xeon processors, eHarmony can analyze more data faster than before and deliver improved matches within overnight processing windows.

## Business Value

- **More personalized results.** The upgraded environment supports new software capabilities that will help deliver more personalized analyses and improved matches.
- **Ready for growth.** eHarmony is better equipped to handle the rapidly rising volumes and variety of user information used for matching millions of users every day.



## Intel® Xeon® processors give eHarmony the performance and memory capacity to drive innovation

### Lessons Learned

Through the process of incorporating a new software framework into the company's matching process, eHarmony developers have recognized the value of engaging with the Apache Hadoop\* software community. "Being on the forefront of technology can be challenging," says Jonathan Morra, principal machine learning engineer at eHarmony. "Our work on the Spark\* framework taught us the importance of staying up to date on the latest technologies, trends, and ideas. If you can incorporate new ideas into your work, developing cutting-edge solutions can be very exciting."

### Matchmaking in the Cloud

eHarmony uses a scientific approach to matchmaking that is clearly successful. On average, 438 people in the United States get married every day because of the company. That's the equivalent of nearly four percent of new marriages. In all, there are approximately 600,000 married couples who met through eHarmony. And every day, millions more people around the world begin new relationships through matches made by eHarmony.

How does eHarmony make those matches? The company uses sophisticated data analytics that combine "compatibility" and "affinity" matching. "Compatibility matching uses responses from our relationship questionnaire to determine psychological compatibility and romantic potential," explains Jonathan Morra, principal machine learning

engineer at eHarmony. "Affinity matching brings together people who have similar interests. For that type of matching, we analyze a wide variety of information provided to us by users, including demographic data, images, communication history, and more."

The IT group relies on a robust cloud environment to process affinity matching and to conduct machine learning research that helps refine the matching process. The cloud is built on Cloudera CDH software—an Apache Hadoop\* distribution that enables scalable storage and distributed computing while providing a user interface and a range of enterprise capabilities. "Using Cloudera lets us stay focused on our core capabilities," says Morra. "The Cloudera team also has been very responsive and helpful in enabling our developers to write better software for our cloud environment."

The eHarmony development team recently began writing code with Apache Spark\*—an Apache code framework designed to replace MapReduce\*. "Spark brings the goal of massively parallel computation closer so we can capitalize on the compute power of thousands of machines operating as one," says Morra.

The new framework takes a fresh approach to accessing data that should enable faster matching for eHarmony. "While MapReduce uses disk storage, Spark stores as much data as possible in memory to improve performance," says Morra. "But to capitalize on the benefits of Spark, we needed to upgrade our version of Cloudera and refresh our servers, adding memory capacity."

New infrastructure also had to deliver outstanding compute performance and scalability. "We need to process more than 10 million matches every

night so results will be waiting for users first thing in the morning," says Hav Mustamandy, director of systems operations for eHarmony. "And even as the volume and variety of data continue to grow, we still have to meet the same deadlines."

### Finding a Match in the Intel Xeon Processor E5 v3 Family

The existing cloud infrastructure included servers equipped with the Intel Xeon processor E5 v2 family. The IT group first upgraded the Cloudera software and then implemented new Supermicro servers based on the Intel Xeon processor E5 v3 family. "We've looked at competing processing architectures, but the Intel Xeon processor E5 family delivers the best performance, plus the large-scale memory capacity that we need to support our memory-intensive cloud environment," says Morra.

eHarmony software developers use Intel® Threading Building Blocks (Intel® TBB) to help optimize new code. "We developed facial recognition software based on academic code we discovered, and we used Intel TBB to replace the Pthreads that were used in the original code," says Morra. "We saw significant performance improvements as a result."

### Enabling Innovation and Enhancing Matching Performance

The new environment gives eHarmony the foundation for improving the quality and speed of matches. "With this new Spark programming framework, a new version of Cloudera CDH, and a robust infrastructure built on the latest Intel Xeon processors, we can do more personalized affinity matching and produce better results

for more people,” says Morra. “Previously, our matching models reflected the preferences of our general user population, and they became stale quickly since our user base is continuously changing. Now we have the power we need to incorporate more user-specific preferences into our matching models and retrain those models daily so they are better tuned for current users.”

New servers with the latest-generation Intel Xeon processors also help accelerate the nightly matching process while maintaining a consistently strong user experience. “We anticipate that moving to the latest Intel Xeon processors will help us improve matching performance so we can process more data each night and continue to meet our strict deadlines,” says Mustamandy.

### Improving Scalability

The infrastructure refresh gives the eHarmony team better scalability to accommodate continued growth. “We now have the scalability to support more eHarmony customers and analyze a larger volume and wider variety of data,” says Morra. “We’re also looking to expand the use of this cluster and offer these resources to additional business groups at eHarmony.”

### Sparking Relationships and Maintaining a Competitive Edge

By supporting analysis of more and different types of data, the refreshed cloud should also help eHarmony retain its competitive edge over a growing number of dating and matchmaking services. “At eHarmony, we measure success by the percentage of users

who act on their match results and communicate with other users,” says Morra. “Analyzing more variables lets us improve our matching process and deliver results that are more likely to result in communication. That’s great for our users, because it will put them on the path of building meaningful, long-lasting relationships. And it’s great for us, because those successes help to demonstrate the value of eHarmony to new potential users.”

Find the solution that’s right for your organization. Contact your Intel representative, visit Intel’s [Business Success Stories for IT Managers](#), and check out [IT Center](#), Intel’s resource for the IT industry.



This document and the information given are for the convenience of Intel’s customer base and are provided “AS IS” WITH NO WARRANTIES WHATSOEVER, EXPRESS OR IMPLIED, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NONINFRINGEMENT OF INTELLECTUAL PROPERTY RIGHTS. Receipt or possession of this document does not grant any license to any of the intellectual property described, displayed, or contained herein. Intel® products are not intended for use in medical, lifesaving, life-sustaining, critical control, or safety systems, or in nuclear facility applications.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations, and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to <http://www.intel.com/performance>

Intel does not control or audit the design or implementation of third-party benchmark data or Web sites referenced in this document. Intel encourages all of its customers to visit the referenced Web sites or others where similar performance benchmark data are reported and confirm whether the referenced benchmark data are accurate and reflect performance of systems available for purchase.

© 2015, Intel Corporation. All rights reserved. Intel, the Intel logo, Intel Inside, the Intel Inside logo, and Xeon are trademarks of Intel Corporation in the U.S. and other countries.