SkillPages

SkillPages lets its 18 million registered users easily find skilled people and see how they are connected in real life. With skills that run the gamut from plumbing to computer programing to playing tennis, SkillPages helps users find skilled people who they share a mutual connection with, like an old school friend or former colleague. Based on the idea that people like to do business with people they know, this feature creates a trusted community. Privately held, the company was founded in 2010 and has enjoyed explosive growth; it serves users from more than 160 countries, and continues to add about 1 million registered users monthly. SkillPages aims to build a taxonomy of all the world's skills.

The Challenge: Fast Access, Huge Data.

SkillPages believes its best strategy for gratifying users is by being genuinely useful. For example, an entrepreneur might search for nearby software developers to help execute a business idea. SkillPages sorts and filters the results, overlaying them with data from the social networks of both the entrepreneur and the developers. To do that, SkillPages rapidly analyzes connections within its own network, as well as on Facebook and Google. Using this data and analysis, its SkillGraph returns results based on factors like geography, mutual connections, and relevance, so the entrepreneur could quickly find developers who have the necessary skills, live in the area, and have worked with people the entrepreneur knows and trusts.

Early on, with a user base in the tens of thousands, relational databases were equal to the task, but as the company came out of beta in 2010 and the user base began its explosive growth, hitting 3 million in 2011, SkillPages needed a much faster and much more scalable solution to keep up with actual and projected growth.

## The Solution: Supported Scalability.

SkillPages moved from its relational database solution to a commercial graph solution, which was better, but was going to be a vertical scale rather than a horizontal scale. Datasets were exceeding a billion rows. "We knew that being able to serve and process this amount of data on the fly was going to take something pretty special," says Mike McCarthy, CTO of SkillPages.

A NoSQL solution was clearly the best fit, and Cassandra became the top contender early on, leading SkillPages to DataStax. The company chose DataStax Enterprise which integrates an enterprise version of Cassandra with Hadoop, Solr, security features and an easy to use management console. The solution met SkillPages' need for scalability, availability, and ease of deployment. "We weren't the kind of company that would tie up two or three engineers on an open-ended project," McCarthy says. "We wanted something we could get support for, that had been tested in the field."

Keeping Up With Continual Growth.

SkillPages relies heavily on testing and quantification to make decisions about its solutions. In making its move to Cassandra, the team ran the new solution in parallel with the old one for six weeks. Having proved itself, Cassandra was on its own after that. To keep up with the demands of growth, SkillPages expanded its cluster from three nodes initially to six in 2011 and then to 12 nodes in 2012. Increasing the size of the cluster stayed ahead of the requirements for data management.

Figure 1: Registered Members
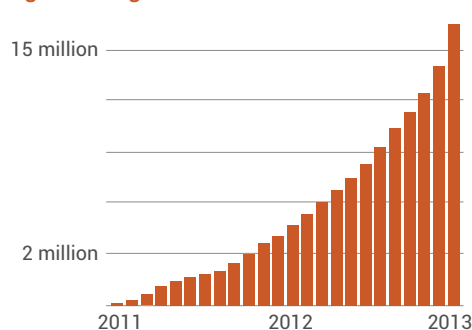
Company
SkillPages

Data Size
3B entities on the SkillPages skill relevance engine

18M registered users and growing

40GB of telemetry data recorded daily

Business Challenge
Quickly serve and process an enormous volume of data related to a user base experiencing explosive and ongoing growth

Solution
DataStax Enterprise allowed SkillPages to scale its product using intelligent, market-tested approaches

"We knew that being able to serve and process this amount of data on the fly was going to take something pretty special, and our DataStax solution has delivered."

Mike McCarthy
CTO
SkillPages

In 2012, SkillPages faced a high level of daily activity from Facebook and Google as it integrated data from those services with its own. An I/O bottleneck developed with Amazon related to the high traffic and volume. But the availability of SSD from Amazon allowed SkillPages to right size its clusters, reducing their size while running larger instances.

"During that transition we had a lot of help from DataStax. Working with experts saved us a great deal of time and effort," McCarthy says.

Solving users' problems and providing a compelling service has always been the primary focus for the SkillPages engineering team, so it tries not to get bogged down in building and maintaining infrastructure.

"Our relationship with DataStax has enabled us to keep this focus," McCarthy says.

The DataStax solution has supported a "graceful failure" approach, in which any temporary infrastructure issues are largely transparent to users.

SkillPages has been able to manage the transition to Cassandra without allocating additional engineering staff to the project.

"We find DataStax OpsCenter particularly powerful and insightful in detecting and resolving issues before they become problems," McCarthy says. "We are able to support and, more importantly, evolve the social underlay of our platform without adding additional engineering capacity to the project."

SkillPages is now achieving significantly greater write throughput than with its previous implementations. Although the primary use case was to power the company's SkillGraph, it has also migrated its URI encoding and mapping repository over to Cassandra and has seen benefits there as well.

Looking to the future, SkillPage has huge ambitions. As the user base continues to grow globally and languages expand beyond English and Spanish, Cassandra and DataStax will continue to support the company's needs. McCarthy and SkillPages look forward to exploring how DataStax and Cassandra can better support integrated Solr search and help expand analytics.

"From an engineering perspective we want to do more than just support growth," McCarthy says, "and also offer more functions for users that increase the value of the product."

"Our ethos has always been to focus on solving our users' problems and not necessarily get consumed with building and maintaining infrastructure. Our relationship with DataStax has enabled us to achieve just this."

Mike McCarthy
CTO
SkillPages

DataStax Headquarters

San Francisco Bay Area
777 Mariners Island Blvd #510
San Mateo, CA 94404
650-389-6000
www.datastax.com

ABOUT DATASTAX

DataStax powers the big data apps that transform business for more than 300 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™.

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