



# CASE STUDY: World Vital Records and Perfect Search

## WORLDVITALRECORDS.COM

### Paul Allen and MyFamily.com Needed Help with Massive Queries

#### The Problem

Paul Allen knew the challenge he was facing. After all, he was one of the founders of the world's largest genealogy companies, MyFamily.com. MyFamily (now The Generations Network) has over six billion genealogical records and requires thousands of query servers to handle customer searches. These thousands of servers cost millions to purchase and millions per year to operate with energy and maintenance costs. Paul was now launching World Vital Records, a new genealogy company that was rapidly acquiring hundreds of millions of genealogical records. His goal was to become the second largest genealogical company. The question was how could he provide sub-second search to his customers without building the costly and massive infrastructure that MyFamily had built to support the query load from customers?

#### The Choice

Paul Allen had the following choices:

- Build his own proprietary search engine as MyFamily had done
- Build his own search platform using open source solutions such as Lucene
- Buy a solution from an Enterprise Search Vendor

As a new start-up, World Vital Records was constrained with limited resources. Paul wanted to use the majority of his resources towards data acquisition, new social networking features, and marketing—not towards his search platform. Building his own proprietary search engine would take multiple man-years to develop and was not feasible.

As World Vital Records acquired more data, solutions from the current Enterprise Search Vendors became too cost prohibitive. World Vital Records had over 800 million genealogical records with a mix of both structured and unstructured data in over 10,000 files. Solutions from existing enterprise search vendors ranged into the millions of dollars, well beyond the reach of World Vital Record's limited budget.

With a limited budget and some talented developers, World Vital Records decided to use Lucene, a free, open source search engine, as their initial search platform, providing exact and near-exact search. Lucene worked well with low traffic and the initial small data sets, but as the data and the

traffic grew, the Lucene platform started to have its own performance and financial costs.

As the data sets grew in size, indexing times became a bigger and bigger hassle. It would take over 880 hours of processing time to index 40 gigabytes of data. The Lucene system could handle about one query per second per server. To meet the traffic demands and to keep their index stored in cache, they partitioned their index across six servers and utilized a Collation Server to distribute the query demand. In spite of their load balancing attempts customer query volume frequently peaked and query response times slowed to unacceptable levels and would sometimes time out during a customer query. The CPUs were often maxed at 100% utilization trying to process this volume of traffic and query load.

To continue to grow with Lucene, World Vital Records would need to continue to add servers to handle additional data and additional queries. Paul Allen was continuing to aggressively add genealogical data, with plans to add 50% more records to the 800 million existing records.

### The Solution

World Vital Records accepted a proposal from Perfect Search Corporation to test Perfect Search's new search engine in a parallel system. The requirements were to replace Lucene; to match existing business rules; to incorporate exact and near-exact search; to match or improve results; to perform on fewer servers; and to provide query results back to World Vital Records in the same format as expected from Lucene.

### The Result

Perfect Search was able to reduce the server requirement of World Vital Records from 7 servers to 1 server, while handling a 60% growth in data. Today, more than 1.3 billion records exist on 1 server, with an additional server for redundancy. The same 40 gb of data that took over 880 hours to index on Lucene is now indexed in about 8 hours by Perfect Search. Wait times were reduced so query response times were sub-second, even with an increase of traffic and data size. CPU utilization seldom exceeds 10% at peak query loads. Perfect Search's solution delivered the following benefits to World Vital Records:

- Reducing indexing processing time to 1/100 of the Lucene times
- Reducing query servers from 7 to 1 server
- Reducing query times to sub-second
- Allowing for continued dramatic data growth without significant server expansion
- Allowing World Vital Records to compete with the market leaders at a fraction of the server capitalization and maintenance costs.

*“Perfect Search has blistering speeds and massive capacity to handle large datasets. We are extremely happy with Perfect Search.”*

- Paul Allen, Founder  
World Vital Records

## About Perfect Search Corporation

Founded in 2007, Perfect Search Corporation is an enterprise search solution delivering 10 to 1000 times faster and more scalable search across exponentially growing data sets. Through software innovation, the company enables the world's fastest search, requiring far fewer hardware resources, so companies can quickly and cost-effectively access and use information from structured and unstructured data sets all across the enterprise. Offering both search engine SDK for OEM and search appliance solutions, Perfect Search enables enterprises to find the information they need when they need it. Learn more at [www.perfectsearchcorp.com](http://www.perfectsearchcorp.com).