

NetDocuments Next Generation Platform is a Success – Thanks to Azul Platform Prime and Apache Solr

Azul Platform Prime eliminates unacceptable application pauses that impact the user experience

NetDocuments was one of the first Software-as-a-Service (SaaS) companies. Their service is used by legal, real estate, financial services, health services and a variety of other businesses who wish to have organizational control and universal access to their documents. The NetDocuments service is a true multi-tenant system that can scale to serve large organizations with millions of documents and thousands of users who manage tens of thousands of clients or projects. The company leverages a cloud delivery model to offer the world's most feature-rich, efficient document management service, including collaboration on work in progress documents and emails, document images, and documents as records.

Built on Search

The NetDocuments service is built around search. Their users upload, manage and collaborate on nearly 1 billion documents across 4 global datacenters. These documents must be easy to find, not only by title, author and date, but also based on the full content regardless of the file type.

Migrating from Microsoft® FAST to Apache Solr™

NetDocuments' SaaS infrastructure originally used Microsoft FAST as a key component, but when Microsoft's roadmap no longer matched NetDocuments' high-scale system needs, the company decided to switch to Apache Solr, an open source high performance Java-based search platform.

Each group of a million documents in the NetDocuments system corresponds to about 1 gigabyte of metadata search index. The full text index is about 10 times larger than the metadata index. The system uses both a metadata index and a full text index, since both types of searches are critical.

With FAST, NetDocuments was experiencing extended indexing latency in the document processing pipeline as the store of documents grew. In contrast, Solr is well known for handling massive indices and is built as an open source application. NetDocuments decided to move to a Solr-based search platform which would scale, deliver near-real-time search capabilities and allow for further customization. They created a single-node, multi-core, multi-tenant index, in order to keep the process of feeding new documents simple, allowing the company to complete the migration to Solr quickly and minimize latency for both indexing and search.

Unexpected Roadblock

As the service was deployed in beta testing, the monolithic index experienced unpredictable pauses during a search, lasting about 1 minute. When searching for a critical legal document speed is key, and any delay is unacceptable to their users.

Benefits of Azul Platform Prime for NetDocuments

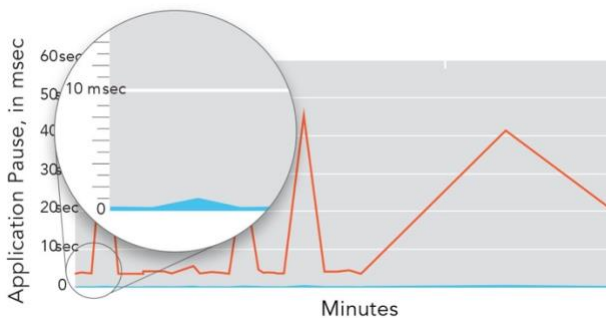
- Allows the company to meet a key goal – moving the data onto a high-scale search platform
- Provides a better user experience with faster page rendering and search results
- Supports future content growth
- Eliminates application pauses and stalls due to Java garbage collection
- Allows each instance to use twice as much memory – over 140 gigabytes
- Provides a faster compiler for increased throughput

NetDocuments Next Generation Platform is a Success – Thanks to Azul Platform Prime and Apache Solr

The Search for a Solution

With further testing the company was able to identify the source of the pauses. When memory fills in a Java-based system, the Java Virtual Machine (JVM) eventually needs to free up blocks of memory. When that happens, it pauses application processing to garbage collect - freeing up memory that is no longer in use.

At an Apache Solr training session in Palo Alto, the NetDocuments team heard about a blog post by Michael McCandless, an Apache Lucene™ committer and PMC member. Michael had tested Lucene with Azul Platform Prime and discovered that it removes pauses, even when utilizing more than 70 gigabytes in the RAM Directory, which is just what NetDocuments was looking to accomplish.



Azul Platform Prime reduces peak pauses from over 40 seconds to under 5 milliseconds

Azul Platform Prime Eliminates Pauses

After some additional online research, the NetDocuments team requested a trial copy of Azul Platform Prime. Azul Platform Prime took just five minutes to install, required no changes to their application, and eliminated application pauses in the first test. They also noted Azul Platform Prime's faster compilers that improved overall throughput.

The company's IT team used Azul's production-time monitoring, diagnostic and tuning tool to monitor CPU usage for each Java class to find and eliminate memory leaks. This helped speed the production launch and improved overall system reliability.

With testing complete, NetDocuments launched the new metadata index on Apache Solr using Azul Platform Prime. In production they are able to support 150 gigabyte heaps for each instance without application pauses. In summary, Mou Nandi, Search Engineer and Architect noted, "Azul Platform Prime is the ideal solution for us and has proven to be critical for search performance. Best of all, it was really straightforward to implement."

"Azul Platform Prime delivers exactly as promised. It took just 5 minutes to install, and we could see in the first test that the application stalls were gone."

Mou Nandi,

Search Engineer and Architect NetDocuments

Problem: The company was migrating from Microsoft FAST to Apache Solr but experienced long application pauses during Beta testing

Solution: Azul Platform Prime made it possible for them to use Apache Solr in production

"Without Azul Platform Prime we would not have been able to deploy Apache Solr for our production system. Our customers could have experienced long pauses when searching for critical documents."

Mou Nandi,

Search Engineer and Architect NetDocuments

Contact Azul

385 Moffett Park Drive, Suite 115

Sunnyvale, CA 94089 USA

+1.650.230.6500

www.azul.com