Innovent Solutions
Phone: 949-754-2940
300 Spectrum Center Drive
Suite 400, Irvine, CA 92618

Search
Business Intelligence
eCommerce

Spotlight on
Client Success

Client Profile

Name:

TMP Worldwide Advertising

Industry:

Advertising & Communication

Need:

Create a search solution that processes tens-of-thousands of job records very quickly, provides quality search results with flexible linguistic abilities, and can easily scale as business grows.

Results:

Delivered a Solr-based search solution which was extremely fast to index, returned search results very quickly and of high quality, was capable of many flexible forms of linguistic manipulation, and was easily scaled into any number of applications internally.

Background

TMP Worldwide Advertising is the largest independent firm focusing on recruitment advertising and communications. TMP offers a depth of resources and services including digital expertise, brand development as well as traditional creative and targeted media planning. TMP deploys a search-engine-optimized index of searchable jobs and has enjoyed strong growth of its service offering to its many client companies. However, the proprietary search solution employed for the public-facing site began to show some strain, occasionally causing serious problems and partial outages of service. As such, TMP decided that a more robust and powerful search solution was necessary.

TMP processes XML feeds of job data provided by their client companies. This data is extracted and loaded into a Microsoft SQL Server database and serves as the primary content for job data on the public-facing web site. During the processing of this data, the existing proprietary search data was generated and inserted into the database. At the time of this project the total number of jobs available averaged around 20,000.

Business Challenge

The project tasks were essentially organized around four themes: process client data quickly with minimal impact to production resources, improve the quality of search results, provide a highly-available and scalable system, and take ownership of an easy-to-grow and maintainable solution by staff.

Process Client Data Quickly with Minimal Impact

The critical issue facing TMP was a growing burden on the production database caused by the creation of the search data. The process involved tokenizing all of the searchable terms found in the job data and manually building an inverse index of tokens to Job IDs. This was a very expensive operation to run that would take many hours, and cause significant load to the database.

Reducing the strain on the database to process the search data was a primary driver in seeking a new search solution. The job data extraction processing was taking more than eight hours to complete, pushing the database load to the point of total resource contention, and rendering the web site unable to yield results while data was being processed.

Improve the Quality of Search Results

Another important issue to TMP was the quality of search results and the business rules applied to them to assist in creating a good search result. Search results from the proprietary process were quite slow, taking upwards of 2-6 seconds per search. The search solution would perform a simple SQL 'Like' clause query against the data, sorting by the weights applied to each token that were defined by the source field the token was derived from. In this way, each individual token not only related to its source document, but to its source field and thus a great deal of data was necessary to process and store the searchable tokens.

Provide a Scalable, Highly-Available System

The search solution needed to help alleviate any condition where a failure of a single resource could cause the critical functionality of the web site services to be unavailable. An important goal of the project was to have a scalable solution that allowed multiple points of redundancy. Additionally, the ability to scale the redundancy was an important consideration.

Business Challenge (continued)

Take Ownership of an Easy-to-Grow and Maintainable Solution

TMP desired to be directly involved in all stages of the process to correctly understand the solution. Being capable of growing the solution to extend into further business areas was a vital goal of the project, giving great value for the initial investment over time and allowing TMP to branch additional projects into the same install base.

Solution

Process Client Data Quickly with Minimal Impact

Innovent began the project by determining the correct configuration of Solr to support the desired goals. A data schema was developed to provide the right balance between multi-field searched data and data to return for website rendering. Solr was configured to return results based upon defined business rules, emphasizing specific fields for boost using the Disjunction Maximum query parser.

Innovent provided a lightweight framework that extracted the data from TMP's Microsoft SQL Server database, threading the requests to take maximum advantage of the processing server's capacity to handle load, while streamlining the database queries to put minimal load on the database server. Information was extracted, transformed, and loaded into a Solr-compatible XML schema that was defined for TMP's specific search needs. The framework also provided the means of loading data directly into the Solr indexing server, optionally allowing the bypass of the XML write step. This streamlined process reduced the time to create a searchable index from 3 hours down to less than 20 minutes.

Improve the Quality of Search Results

Extensive Quality of Search and load testing was performed on the search data to ensure the blended mix of Solr boost equations was suitable for the business requirements defined for search. Use case testing was performed to help identify outlier search issues, address the needs of synonyms or other natural language processing techniques, and adjust the data processing steps for potential problems in search. Encoding problems, removing HTML tags, multiple-value field modifications, and other search-quality issues were identified and resolved during the course of testing. Additionally, load testing was performed to ensure the solution was robust enough to handle the estimated peak traffic in the coming two years.

Provide a High-Availability and Scalable System

Innovent worked with TMP to install Solr with a distributed index and search server setup, creating an indexing server that was segmented in the network away from the web servers and allowed a headless indexing of data. Two search servers were installed to pull data from the Master, allowing a load balancer to deliver stateless search traffic from any of the three web servers to either of the Solr search servers. Replication between servers occurred over HTTP and was taking less than 60 seconds for a full index of data.

Take Ownership of an Easy-to-Grow and Maintainable Solution

During delivery of the solution Innovent took great care to provide a hands-on, step-by-step walkthrough of the configuration and code. An extensive knowledge transfer setup TMP to be fully self-sufficient and capable of extending the reach of their search solution to many of the additional search features Solr provides. Over the course of a week, every aspect of the solution was detailed, with alternate scenarios tested to show the flexibility of Solr.

The Results

TMP partnered with Innovent for its search expertise to create and implement a Solr-based search solution. Innovent prepared a project plan to integrate TMP's job data into a Solr-based schema and provide a high-availability, redundant search solution that was capable of meeting TMP's growing needs.

Innovent delivered the project on time and on budget. Search results were reduced from 2-6 seconds on average to less than 250ms on average. This led directly to the search results having a highly-noticeable difference in speed, and being "lightning fast". The complex and flexible field boosting capability and natural language processing features of Solr provided excellent tools to swiftly address interesting search scenarios for its diverse client base. Most importantly, the strain on the database to process and provide search results was eliminated, moving search processing time down from 3 hours to less than 20 minutes and eliminating the highly-contentious machine processing necessary to provide search.

Innovent Solutions
300 Spectrum Center Drive, Suite 400
Irvine, CA 92618

Phone: 866.466.6868
Fax: 562.262.2835
Email: info@innoventsolutions.com

ENABLING INTELLIGENT DECISIONS®

Find out more:

www.innoventsolutions.com

Innovent Solutions provides consulting, training and support services and solutions for Search, Business Intelligence and eCommerce technologies. We build systems that enable our clients to:

- Find the information they want and need
- Understand the context and meaning of information
- Trust the information to be accurate and timely

Our focus on customer goals and objectives combined with deep technical depth and experience, allow us to build successful, high visibility solutions that deliver immediate value to our clients.

Innovent Solutions is a privately held company headquartered in Irvine, California with offices in Minneapolis.

How can we help you?

Innovent provides consulting, architecture, development, training and support services to organizations using for enterprise, eCommerce and Big Data search applications. Our services include:

- Requirements Scoping and Analysis
- System architecture
- Installation and configuration
- Data preparation services to optimize indexing and quality of search
- Connectivity to specific content systems
- Search result analysis and tuning
- Performance Tuning and Load testing
- In-depth reporting and analysis of search information