Flax Case Study:
Reed Specialist Recruitment
October 2011

Reed Specialist Recruitment is part of Reed Global, which also includes Europe's biggest jobsite - reed.co.uk; one of the world's leading Welfare To Work providers - Reed In Partnership; and the UK's learning provider of the year - Reed Learning.

Founded in 1960, Reed is a specialist provider of permanent, contract, temporary and outsourced recruitment solutions, and IT and HR consulting. REED operates in Europe, the Middle East and Asia Pacific and has more than 3,000 permanent employees working out of 350 offices worldwide across 30 specialisms.

reedglobal.com enables clients to interact with our recruitment services and jobseekers to search for opportunities across multiple industry sectors around the world. reed.co.uk receives over 1.5 million job applications per month.

In 2010 REED launched its most ambitious programme of IT development to date with a business-led strategy for technical innovation. Key to this strategy was the development of a raft of new search-based applications for use by staff both in local and central offices. Due to the many millions of items that would need to be searched, the applications would need to be backed by a powerful and scalable enterprise search architecture.

Due to the nature of the REED business, the ability to correctly match candidates to jobs is absolutely key to future profitability: thus the performance and accuracy of the search engine will directly drive business benefit.

After consideration of a number of commercial, closed-source search engines, REED engaged Cambridge-based search specialists Flax who recommended the open-source Apache Lucene/Solr platform as a solid foundation for the new system. Apache Lucene/Solr is a powerful, scalable enterprise search server written in Java and used across the world by companies including IBM, eBay, Ford and Boeing.

The first challenge encountered by Flax was to design an effective indexing scheme for the complex database structures in use at Reed. The database holds candidates' CVs (resumés) - sometimes many for each candidate, jobs, timesheets for temporary or contract employees and much more. Hundreds of millions of database rows were to be indexed. In addition, CVs are held external to the database as flat files in a multiplicity of different formats such as Microsoft Word and PDF. Most of the files are in English, but other languages include Polish, Arabic and Chinese.

Flax built a custom Indexer application in Java based on the Spring framework, which takes as its input a Context XML file defining Types for database data, Actions to be taken on various database views, and optional Processes that may transform the data. This allows a flexible, potentially complex but easy to configure pipeline between Reed's databases and the Solr search engine. The REED schema contains over 300 field definitions.

The Indexer also has the capability to verify the values of fields, ensuring that fields required by Solr are filled, so that for example, integer fields get integer values (rather than allowing a document to make it all the way to Solr and fail to be added), or verifying field values against configured regular expressions.

The indexing application uses Apache Tika to translate proprietary file formats into plain text for indexing. The language of the source file is automatically detected and used to send the text to a field configured with suitable tokenization and normalisation. Subsequently the SolrJ interface is used to add to and update the Solr indexes. The indexer is designed to scale efficiently with multiple threads and across multiple machines.

A separate Configuration Manager application reads the Context XML file and generates per-field entries for inclusion in the solr.xml configuration file, so that there is no need to manage this schema separately. In addition the Configuration Manager checks for and prevents any conflicts between multiple Actions and Processes on a field.

Once the searchable index is built, a number of front end applications use it to provide results. These include both a simple text search box and more specific search interfaces

for particular record types. Search features available include faceting and geospatial filtering (for example to restrict candidates to a particular distance from a potential job). The search applications are available to REED staff both at local offices and centrally.

Commenting on the high speed of delivering search results, a recruitment consultant at REED said: "It takes less than one second to get my results back, the same search used to take at least 30 seconds!"

Key to the REED requirement was the ability to filter and re-rank search results based on complex business rules. Solr's FunctionQuery has been used extensively to provide this as have custom boost options.

With a search application of this size and complexity it was important to be able to monitor performance as the system was developed and once it was released for use. Flax built a custom Performance Tester which allows for the average query time and number of results to be recorded for a concurrent batch of random queries based on terms generated from the index. It also allows for direct comparison between putative boost functions which may re-rank the results in hard to predict ways.

Flax worked closely with REED during an Agile development process, allowing the project to be delivered on time and within budget.

Flax is a partner of Lucid Imagination Inc., the commercial company for Apache Lucene/Solr open source enterprise search, who offers training, certification and support services for Apache Lucene/Solr. Flax was therefore able to offer a comprehensive 24-7-365 support package for the search application, backed by Lucid Imagination.

REED is already enjoying the benefits of the new recruitment search system. With the previous applications used by REED staff, searches could take some minutes to return results: with the new system results appear in around a second. The new Solr search server runs on only two virtual servers with some available headroom. REED now has a solid foundation for search applications, built on trusted open source code used by thousands of organisations across the world. By taking advantage of

the open source model REED is protected from vendor lock-in and pays no annual license fees.

Commenting on the project, Philip Pegden, IT Director said:
"The transition to Solr was the latest step in our strategy to develop a truly worldclass search application. We believe it provides a robust architecture that meets our future aims, it will scale economically and is a welcome addition to our existing suite of Open Source systems."

"Flax were a trusted partner throughout the work, brought significant expertise, and helped ensure we completed the project on time and to budget."

--- ENDS ---

About Flax

Flax is highly active in the information retrieval market with international clients from sectors including academia, public relations, e-commerce, government and private businesses. As a cross-section Flax's clients include Accenture, the Newspaper Licensing Authority (NLA), Durrants Ltd., The University of Cambridge and Mydeco. Flax delivers a cutting-edge enterprise search solution, using the power of open source software to drive down costs and provide world beating search performance with no software licence fees. Flax is an authorized partner of Lucid Imagination, the commercial company behind Apache Lucene/Solr. Flax's accolades include an award from the British Computing Society for its powerful, innovative search tools.

For further information please see http://www.flax.co.uk/

Editorial contacts:
Charlie Hull, Flax
+44 (0)8700 118 334
charlie@flax.co.uk

Lewis Campbell
Head of Research & Innovation
REED
020 7616 2347