CONCORDE SEARCH SOLUTION

Solution White Paper

Contents

1   Introduction

Over the years organisations and individuals have made a conscious decision to move to a "Paperless" information store. With the reducing cost of computer storage hardware and the increase in capacity of the media, the trend to store data electronically has been increasing exponentially. The usefulness of this data is only realised by the ability to search and retrieve the information of interest. The speed and accuracy of the result from search is critical for any application, whilst the data provided to query the repository cannot be relied on for its correctness.

Significant information uplift can be achieved from using good search technology. Proliferation of information in an enterprise can be across structured and unstructured data sources. The need for extracting key information for comparison, data harmonisation, de-duplication, fraud investigations and testing data quality is greater than ever.

Concorde Search Solution provides the capability of finding the relevant information across your organisation's disparate information repositories. The software provides a simple and effective interface to search using full-text (fuzzy) search, structured search and analytics. The management functions provide the capability to configure, secure and audit the utilisation of the search engine to meet the goals of the implementation.

Some of the key advantages offered by the Concorde Search Solutions are:

- Customisable Fuzzy Text Search Engine that provides a large selection of industry standard Fuzzy search & relevancy scoring algorithms, which can be configured without the need for programming

- Offers integration capability using RESTful web service interfaces

- Provides HTML5 web based administration functions to customise, configure, operate, tune and monitor the search engine implementation

- Off the shelf security integration with LDAP or Active Directory

- Offers Multi-Language support

- Runtime configuration of weightage and percentage of algorithm contribution per field

- Off the shelf dashboards and reporting, monitoring tools to monitor the health of the system and Data administration without the need of IT

- Simplified and unified installation process on all platforms, supporting scalable distributed setup

2 Concorde Search Solution

Nouveautech has developed a text search solution to provide a high-performance and feature rich search engine capable of:

- Distributed real-time data store where every field is indexed and searchable

- Distributed search engine that is simple to implement across a number of industry domains to meet specific business requirements

- Scalable (vertically and horizontally) implementation that can work with structured and unstructured data

- Access using RESTFul services

- Implementation managed via a simple administration user interface

- Supporting authentication schemes to major user stores such as LDAP & Active Directory

- Supporting industry standards web services security

- Audit feature with replay functionality

- Statistical dashboards to show results' accuracy and measure performance

- Runtime tuning using weightage and algorithmic rules to manipulate the score for specific implementation

- Ability to supply the field weightages as part of every request in a transaction

- Multilingual transliteration support

3 Concorde Search Engine

Concorde search engine offers a number of options to perform "Exact", "Fuzzy" and "Wildcard" comparisons, the following algorithms are built in to the standard capability:

- Phonetic

  - Double Metaphone
  - Metaphone
  - Refined Soundex
  - Soundex
  - Beider Morse
  - Caverphone 2
  - Cologne Phonetic
  - NYIIS
  - Koelner Phonetik
  - Haase Phonetik
  - Match Rating Approach
  - Daitch-Mockotoff Soundex

- Distance based

  - Damerau-Levenshtein Distance
  - Jaro-Winkler
  - Hamming

- String Sequencing

  - nGram
  - Edge nGram

- Synonyms

  - File based aliases

- Transliteration

  - Latin
  - Arabic
  - Armenian
  - Bopomofo
  - Cyrillic
  - Georgian
  - Greek (with UNGEGN variant)
  - Han
  - Hangul
  - Hebrew
  - Hiragana
  - Indic
  - Jamo
  - Katakana
  - Syriac
  - Thaana
  - Thai

The search results include a relevance score that grades the record according to the weightage assigned to the field in the search criteria. The engine uses one or a combination of the following algorithms to determine the score:

- Proportional Weightage Based Algorithm
- BM25 (also known as Okapi Best Match 25)
- TF/IDF
- LM Dirichlet similarity
- LM Jelink Mercer similarity
- DFR similarity
- IB similarity

Concorde Search Engine is robust, scalable and fast. An enterprise hardened solution that is simple to deploy and configure to meet the requirements of your business. A feature rich and powerful tool that employs a number of industry standard algorithms to search, find and present the data with the relevance score for each record in the result, to help the user determine the accuracy of the match.

4 Concorde Search Solution Architecture

Concorde Search Solution operates as a standalone, self-contained "black box" implementation, which can be deployed to scale vertically as well as horizontally to meet the non-functional requirements of the business.

The architecture design below (Figure 1) shows that the product is made up of several distinct components.
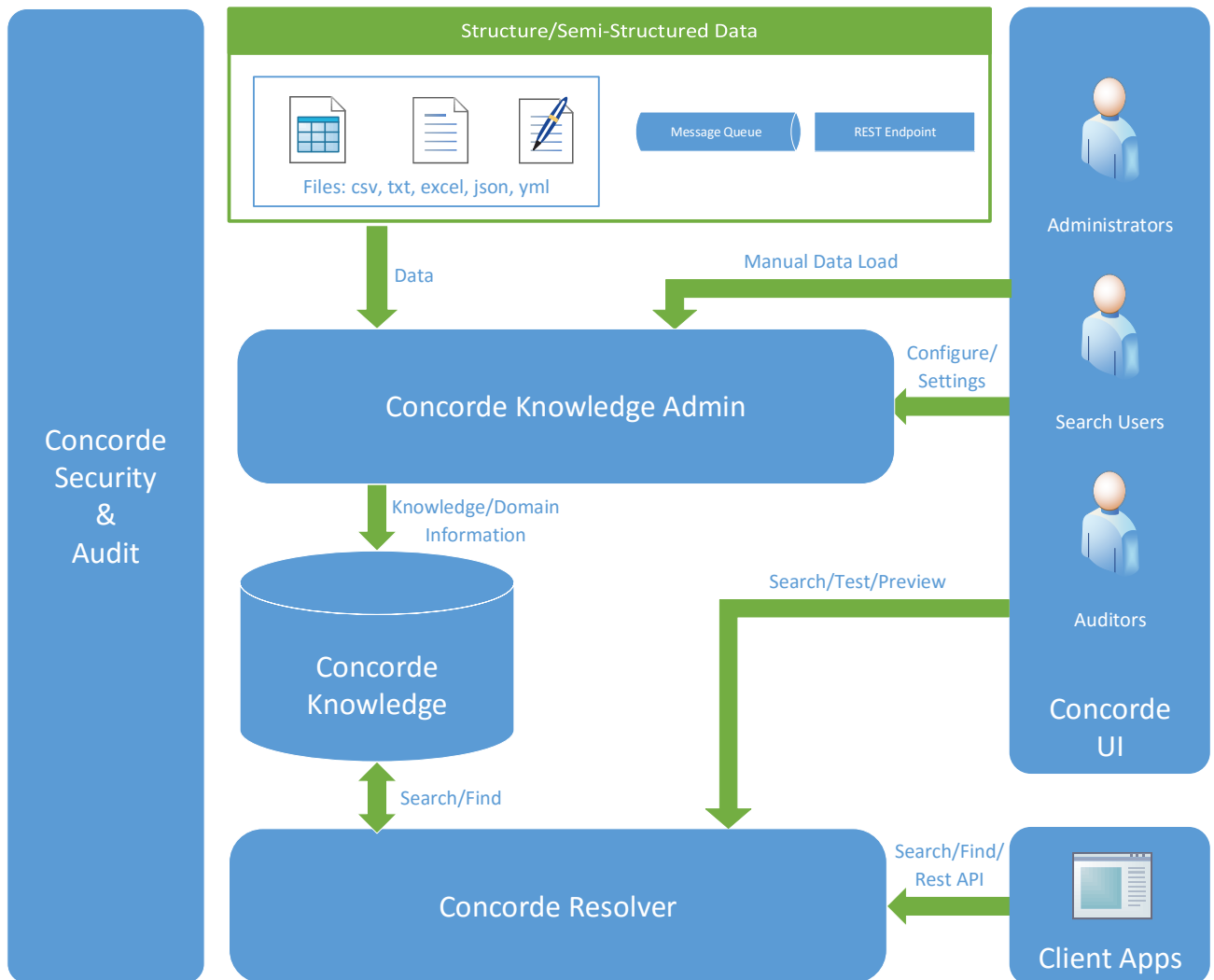

Figure 1

5 Concorde Solution Deployment View

The product is a platform agnostic collection of java components, designed to operate in a single server or a distributed virtual network comprising of individual nodes or cluster configuration.

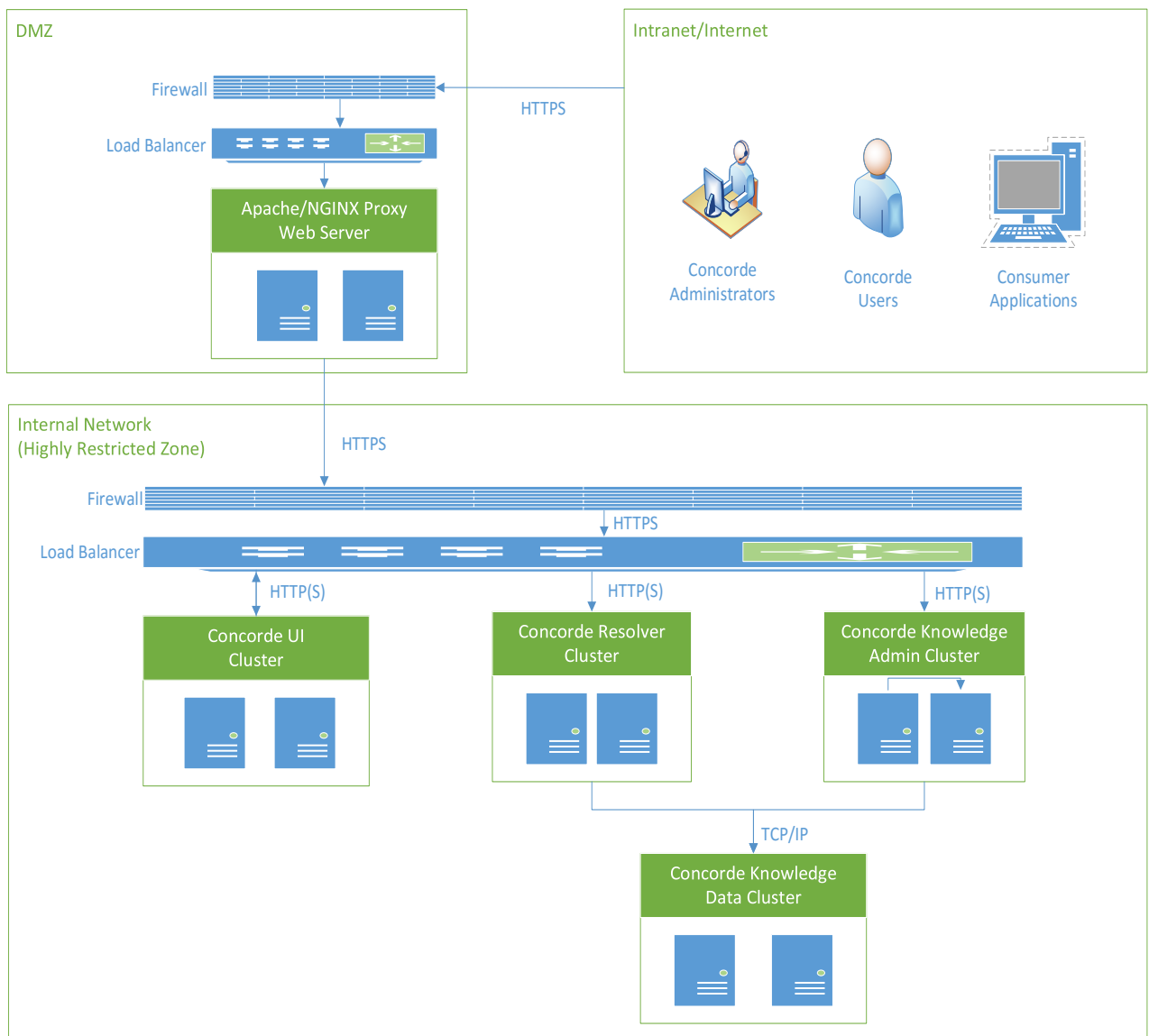The deployment view below (Figure 2) shows a sample implementation of the solution:

Figure 2

6 Conclusion

The need to make the most of the data available to an organisation is becoming increasingly apparent. The opportunities available to gain that edge over the competition need to be realised. The accuracy of data matching and retrieval is the differentiator and the selection of the tool that is simple, effective and straightforward is the critical success factor for each and every implementation.

Concorde Search Solution is developed to bring the technology to the end user. Concorde offers tools that simplify the adoption of complex algorithms and data searching techniques. The solution empowers IT users with the ability to configure, administrator and operate the implementation using the out-of-the-box features. The product provides well defined web service interfaces to aid uncomplicated integration using industry standard techniques. The ability to tune the engine using the experience from the organisation for its specific implementation not only improves the results, but allows the organisation to expand the use of this technology to get a better return for its investment.

For Further information and demonstration, please contact:

info@nouveautech.co.uk