Reference Code: TA002012BIN

Publication Date: November 2010

Author: Mike Thompson, Balachandar Ganesh, Richard Edwards

TECHNOLOGY AUDIT

NetSearch

Connexica

SUMMARY

IMPACT

One of the frustrating aspects of the business intelligence market continues to be the limited permeation of BI tools and applications throughout large organizations. This limits the return on BI investments generally, and offers the power of analysis and data mining only to "the few" rather than "the many". Connexica's NetSearch combines enterprise search techniques with BI technologies to deliver desktop and enterprise-wide search and query facilities across databases, documents, directories, email servers, intranets and enterprise applications. It also combines traditional analytics with search-engine technology to provide a fast and easy-to-use ad hoc enquiry and reporting solution. Connexica has an established presence in the UK healthcare sector and other sectors where NetSearch provides a real alternative to the building of complex data warehouses.

- The Google-esque look and feel of NetSearch provides the typical business user with the information analysis and discovery tools necessary for today's "new world of work".
- NetSearch is relatively easy to set up and deploy, and is well proven in the UK healthcare and early successes in other sectors including retail and local government.
- NetSearch presents a simple search interface that enables end users to find, filter, and view business information without the need to know SQL.

KEY FINDINGS

Strengths:	✓ The ability to use Venn diagrams to visually understand complex relationships. ✓ Dashboards are easy to build and are highly visual, with the ability to incorporate Google maps.
Weaknesses:	✗ NetSearch is not a pixel-perfect BI reporting system. ✗ NetSearch would benefit from integration with enterprise applications such as SAP.
Key Facts:	i The solution is a Java EE-compliant web application. i Connexica recommends 64-bit Windows for NetSearch.

OVUM VIEW

There is little doubt that many organizations are coming across the limitations of traditional BI solutions in terms of deployment cost, ease-of-use, and cost of ownership. Moreover, because very few organizations have rolled out their BI investments to an enterprise-wide audience, the return on investment has not developed to a level that Ovum believes it should.

NetSearch addresses these issues with a novel approach using open-source search technologies and "hiding" the complex structure of queries from end users through the use of an intuitive filter-centric search interface. At a superficial level, NetSearch could be seen as just another user-friendly way of accessing corporate data and information, but the technology and capabilities developed by Connexica go a lot deeper, and Ovum believes NetSearch cannot be categorized as "just another BI tool". It's fundamentally different in the way it works and this deserves to be recognized.

The value proposition of NetSearch is found in the way the product provides the end user with the ability to search, analyze, and view data. NetSearch comes with the usual views associated with data analysis solutions such as tabular, cross-tabs, and graphs, all of which are fully interactive in that NetSearch allows the user to drill down into the source data. Sophisticated Venn diagrams enable complex data relationships to be easily viewed and analyzed, and NetSearch can be configured to deliver visually rich dashboards.

Connexica has a strong track record in the UK healthcare and other sectors in the UK and Internationally with NetSearch, and Ovum believes that this offering will be of interest to many different types of organizations in other sectors and geographies. NetSearch also presents opportunities from an OEM point of view, and so application vendors should take a closer look at the product's capabilities.

Recommendations

- NetSearch is worthy of closer examination by any organization looking to carry out ad hoc data analysis across a large diverse set of data and information sources.
- Healthcare organizations and medium-sized companies and institutions suffering from data overload will find real and immediate value in NetSearch.
- NetSearch is not a replacement for existing enterprise-grade BI solutions, but it could provide added benefit and value to departments or business units with more ad hoc BI requirements.

SOLUTION OVERVIEW

NetSearch is classed as a "next-generation" BI tool, and was created by combining two distinct technologies, taking the best that each has to offer and removing the limitations that exist within them. These two technologies are "search" and "standard" BI, and to better understand NetSearch it is relevant to look at the strengths and weaknesses of both.

Search technology provides a number of benefits, the most significant of which is the ability to handle large distributed repositories of primarily unstructured information. The ability to search for information using keywords or phrases rather than cryptic SQL statements means that typical business users can find relevant information with relative ease. However, the downside of enterprise search solutions has always been the lack of analysis features to handle the search result set to find the precise piece of information that the user is seeking.

In many ways BI tools and platforms are like the flipside of enterprise search in that they enable business analysts and management reporting teams to query and report against large structured data sets and databases. These BI systems take as their source structured repositories and transactional applications that are usually located in the corporate datacenter. Expertise and technical insight are usually required to maintain efficacy, and complex queries must be crafted to return the right level of information. The net result of this focused attention is the detailed analysis of a business-critical metric or exploration of a data-centric business-related insight.

Both technologies can handle large data volumes. In the case of BI, data usually has to be extracted, transformed, and loaded using ETL technology. This is time-consuming and gives a degree of latency to the available data. Search solutions must also acquire, parse, index, and rank data, but this is usually a continuous process and is not critical to the outcome of the query.

NetSearch creates an environment through a user-friendly graphical interface that provides the power, flexibility, and ease of use to mine data, together with the ability to analyze that data in any form and in any way the user might wish. This is another important aspect of the NetSearch technology. Unlike many traditional BI tools there is no requirement to predefine data cubes or create any form of pre-query structure for the data.

Although the background technology to NetSearch is important in that it provides the engine that makes the solution work, the really interesting part from the user point of view is the interface. Figure 1 shows a typical interface as it might be implemented.

SOLUTION ANALYSIS

Visualization capabilities

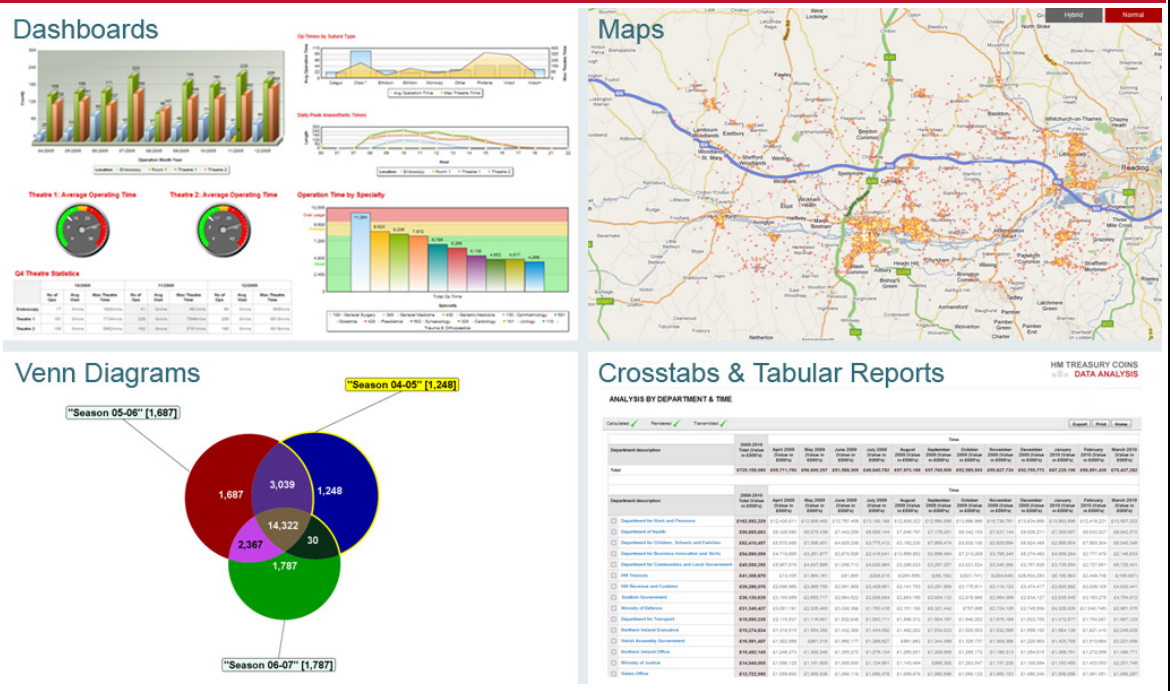
Figure 1 illustrates a few of the possible data-representation techniques delivered via NetSearch. There are a number of views available from standard dashboards that can represent data from a returned result set in both tabular and graphical form including bar charts, pie diagrams, and graphical point lines.

This is an example, and the interface can be completely customized to the user's requirements. The really interesting representation is the inclusion of a Venn diagram, and the best way of describing how this might be used is by a simple example such as if the user was a magazine publisher then they could query the data for subscriptions over the last three years, say 2007 to 2009. The intersections of the Venn diagram would indicate the number of people that had taken out subscriptions for all of the last three years, but a different intersection would show the numbers that had taken out subscriptions in 2007 and 2008, but not in 2009. All of this is provided by NetSearch out-of-the-box without the need to pre-aggregate data or write extremely complex SQL.

At this level the data analysis would be of interest and of some value, but NetSearch provides greater functionality. If the publisher was interested in discovering details about who those people were that had subscribed for the first two years but not the third, it could simply click on the relevant intersection to provide the drill-down details of those individuals. This result set could then be used to carry out a targeted marketing campaign.

Although it would not be too onerous a task to write an SQL query to abstract this data, it does have some presuppositions. First, a user would need to understand the schema of the underlying data sources (in this simple example one could assume a single data source, but NetSearch can carry out the same type of analysis across multiple sources). Next, there has to be IT intervention to create the SQL query, which would be likely to create latency between the data analysis requirement and the implementation of the query. Last, if users see some correlation between queried data then they can simply restructure the query to include or exclude other variables. Moreover, NetSearch can do this kind of analysis over millions of records in real-time. Connexica believes that doing this in SQL would not work, which is why there are no SQL-based real-time VENN applications available on the market today.

Figure 1: User Interface Representation



Source: Connexica

OVUM

Administration

As with an enterprise-grade solution, NetSearch offers a number of capabilities that help IT administrators configure the solution before it is rolled it out to the user base. The solution offers a number of indexing modes depending on the number and type of information artefacts in the enterprise. For example, the solution supports:

- Complete indexing: in this scenario the index is rebuilt every time the index schedule runs. This is recommended for indexes containing a small number of records.
- Incremental indexing: this indexing mode involves adding only those records to the index that have been changed/modified. This is particularly useful where there are a large number of records to deal with.

- Snapshot indexing: this involves updating the index with effective “from” and “to” dates. The snapshot indexing method is particularly useful for point-in-time and/or churn analysis on the data.
- Continuous indexing: this involves incrementally updating the index at a predefined time interval. It is often used to provide virtually real-time indexes where the index needs to be updated within 1 to 15 minutes of the source data.
- Amalgamated: this indexing mode involves joining together two or more indexes based on a specified set of filters and join keys. This is often used when combining data from multiple systems into a single index.

NetSearch also ships with a number of security capabilities to help enforce role-based access policies and configure the data sets and source systems that the solution is capable of scanning. Supported security mechanisms include filters and redirectors (an API that prompts the user to enter his/her credentials before they gain access to certain pages within NetSearch), single sign-on, single user accounts (password-based), group access rights, and feature security where user authentication and security policies can be applied to each piece of functionality within NetSearch. NetSearch uses the concept of realms to enforce fine-grained security policies. Realms are filters that are automatically applied when a user searches against an index. Each realm can include or exclude data based on one or more conditions such as table view, export, load, search details, or Venn.

Reporting capabilities

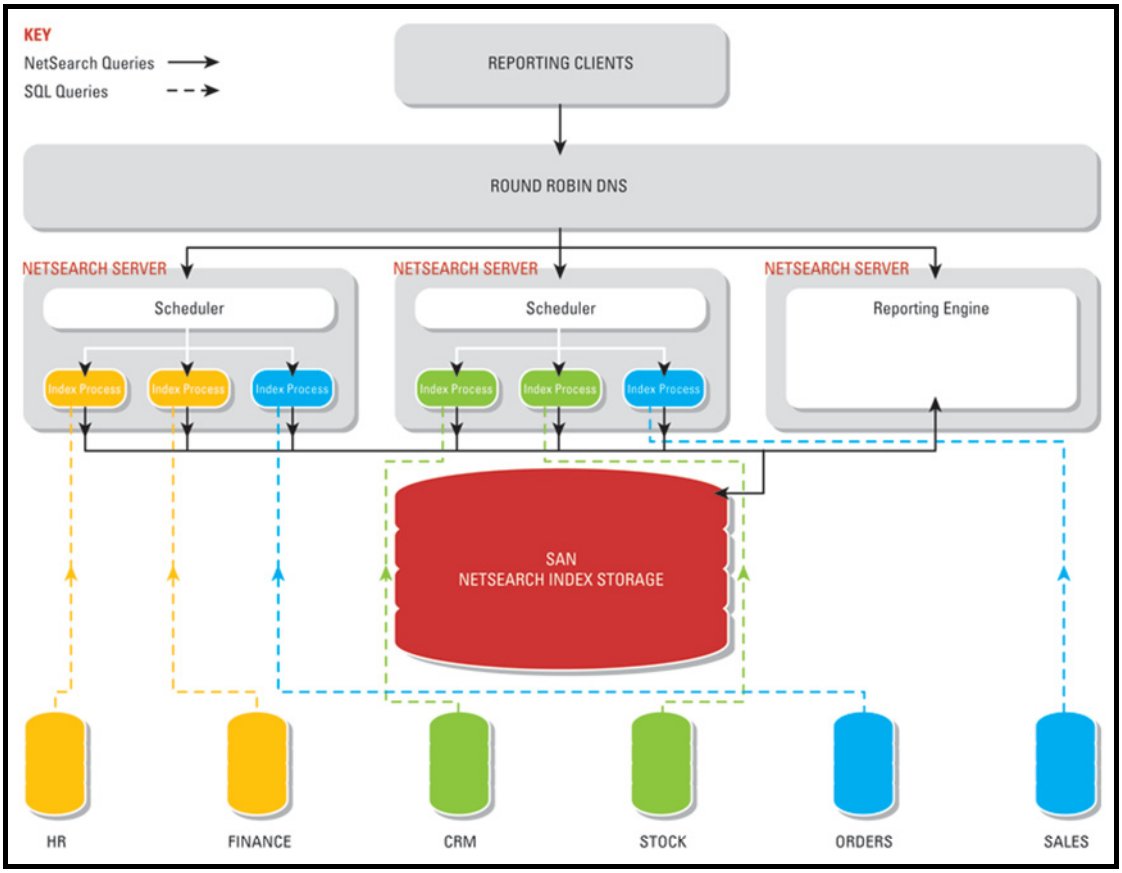
From Connexica’s point of view, reporting is just another visualization option available in NetSearch. However, the solution offers three distinct custom visualization/reporting capabilities: aging reports, churn analysis, and point-in-time analysis. Aging reports are usually “date”-based and let users build in calculations and generate reports based on operations such as current date, dates before, or dates after a specific date. The solution enables users to produce rolling date reports that can run on a daily basis without the user having to manually change the parameters used to filter the data. Churn analysis is a reporting mechanism that helps users compare and study data between two different points in time. Point-in-time analysis is a variation of the churn analysis where users can see snapshots of data at a specific point in the past.

Architecture

NetSearch is a Java EE-compliant web application. The core application code runs within a Java Server Pages (JSP) container on a web server, and serves static HTML and Java Script (with Ajax) back to the client. All data transformations and business logic are applied on the server with the exception of charting (including maps) and Venn diagrams that use Flash and Java Applets respectively. The solution implements web services to allow interoperability with other programming languages including PHP, .NET, C, C++, and Python. Configuration and security information is persisted in an embedded database. Search engine data is persisted on disk as fast-access index files.

Figure 2 shows a sample configuration of a NetSearch cluster for high-speed indexing and access to indexes containing large volumes of data sourced from multiple data repositories.

Figure 2: Configuration



Source: Connexica

OVUM

Performance and Scalability

NetSearch is designed to support massive concurrency, high data volumes, and high query demands. NetSearch can be configured as a stand-alone server, clustered, or as a server farm. Each server has a series of thread pools that are either application or user-based and can be adjusted depending on usage and index sizes. The recommended operating system for NetSearch is 64-bit Windows, which allows the application access to increased RAM. Connexica recommends that the indexes are written to span multiple disks configured with RAID 1+0.

This configuration allows users to scale up each server in terms of RAM and IO. Doubling up the servers or increasing RAM and IO by adding additional disks and controllers will effectively double capacity. In terms of performance, NetSearch has indexed over 22 million records in an hour on a dual quad-core Windows 2008 server, and is able to handle over 1,000 queries per second.

NetSearch uses flat files on disk that can be secured at any point. The product also contains an import/export function to allow the backup/restore of the configuration and indexes in a logical structure.

PRODUCT STRATEGY

Market Focus

The target market for Connexica's NetSearch is vertical-agnostic. Connexica says NetSearch is suitable for all organizations that have large amounts of data, a large and diverse set of user communities, and a large number of disparate data sources. While Ovum tends to agree with Connexica's stance, we would like to qualify it one step further and state that NetSearch has the potential to bring about benefits to organizations that have a requirement to carry out data analysis on an ad hoc basis. For these organizations, investing in a traditional BI solution might not be a viable option simply because of the overheads involved in deploying and managing it. Although Connexica's NetSearch is a vertical-agnostic offering, NetSearch's initial customers were mainly in the healthcare sector, and for this reason Connexica will continue to build on its position in this domain.

Connexica has a growing partner network, with 80% of its sales through the channel. Its partners include Clinical Solutions, Migrationware, and Systemation.

There has been a mix of OEMing and straight reselling, sometimes "white-labeled".

Licensing and implementation costs

An average cost for a typical implementation would be about £50,000, which would include a £5,000 service cost. Annual maintenance is 20% of the implementation cost.

Product releases

Connexica's product release cycle for NetSearch includes two major releases per year and quarterly point releases. Bug fixes and patches are rolled out on an as-needed basis.

Future developments are based on two key drivers:

- To enable users, regardless of role or job function, to access information locked away in disparate systems and use that information through a flexible and user-friendly GUI.
- To extend the reach and use of data analysis.

To do this Connexica will continue to provide new ways of visualizing data to make analysis simpler and more intuitive.

IMPLEMENTATION

Because NetSearch works by creating indexes against the underlying data sources, there is an initial requirement for some (minimal) SQL skills to create the indexes used by NetSearch. There is also a requirement for an administrator to manage security. These roles would typically be fulfilled by one person.

Connexica has a clear six-week implementation plan that it follows to ensure speedy delivery of the system to the end user:

- Week 1: supply of software components and project initiation.
- Week 2: requirements analysis and systems design.

- Week 3: building of core solution elements and sample data indexing.
- Week 4: production data indexing.
- Week 5: user acceptance testing and system sign-off.
- Week 6: delivery of training courses and roll-out to all users.

In a typical deployment new indexes would be added over a period of time, usually at a departmental level. This would be done by the end user because Connexica expects customers to be self-sufficient within a week. Training requirements are also minimal. The only courses required are one-day end-user training, and a one-day administrator course.

There is ongoing support available via a helpdesk, although times for this are limited to GMT 9am to 5.30pm Monday to Friday.

NetSearch runs on Microsoft Windows, Linux, and Solaris. Both 32-bit and 64-bit processing is supported, but the recommendation is for the latter. Clients access the solution from a web browser, which has to be IE (v7 or above) or Fire Fox v3 or above.

NetSearch is delivered as a single product that has certain functionality disabled based on the license, which can be user or server-based. There is a long list of items that can be switched on or off within NetSearch to enable end users to tailor the product and the cost to their own specific requirements. For example, although NetSearch will perform Internet searches as well as searches against corporate data sources, this is optional and allows the end user to choose not to pay for unwanted functionality.

Table 1: Contact Details

Connexica Limited
United Business Centres
16 Parker Court
Staffordshire Technology Park
Stafford ST18 0WP
UK
Tel: +44 (0)1785 246777
Fax: +44 (0)1785 221780
Email: info@connexica.com
www.connexica.com

Source: Connexica

OVUM

Headquarters
Shirethorn House,
37/43 Prospect Street,
Kingston upon Hull,
HU2 8PX, UK
Tel: +44 (0)1482 586149
Fax: +44 (0)1482 323577

Australian Sales Office
Level 46, Citigroup Building,
2 Park Street, Sydney,
NSW, 2000,
Australia
Tel: + 61 (02) 8705 6960
Fax: + 61 (02) 8705 6961

End-user Sales Office (USA)
245 Fifth Avenue,
4th Floor, New York,
NY 10016,
USA
Tel: +1 212 652 5302
Fax: +1 212 202 4684

For more information on Ovum's Subscription Services please contact one of the local offices above.