

The Second “Big Bang”

A NetSearch White Paper

**Taking advantage of the Data Explosion and providing
scalable Business Analytics for End-Users**

Created: 15/12/2008

Version: 6.0

Created by: Richard Lewis

Ref: DataExplosionWhitePaperV6.doc

Private, Confidential and Commercially Sensitive

The Second “Big Bang”

Over the past decade there has been an explosion of data borne out of people’s insatiable appetite for accessing information on just about anything and everything.

Huge increases in hardware capacity, bandwidth, processor power and the improved accessibility of information through media such as the Internet, TV, loyalty cards, SMS and email have fuelled an unprecedented growth in global information.

The world-wide popularity of Wikipedia and the social networking phenomena of Facebook, YouTube, Digg, Twitter and LinkedIn are just the proverbial tip of this data iceberg. These applications may have caught the public imagination and become embedded in people’s lives but the real value of the data explosion lies in the much deeper commercial application of information.

Every day, a plethora of data – both structured and unstructured - is being gathered and stored on every one of us.

For companies with the right know-how and resources, this information is seriously valuable. With forensic insight it can reveal everything about us - from our individual identities, lifestyle habits, behaviours and pastimes to our spending patterns, consumer tastes, personal preferences and much, much more.

Some of this data is used to drive service improvement and enhance the ‘customer experience’ - such as suggesting related products and services when making a purchase on-line.

Other information is being used to help generate demand and streamline business processes. But there remains a mountain of information that that is not being utilised at all – data that is too difficult to analyse due to the sheer volume, complexity and cost involved in mining the data through traditional Business Intelligence (BI) and Data Warehousing tools.

Information as an Asset

Information is only an asset if it can be used directly by the business to drive more effective decision-making.

Business users must be able to identify and understand patterns in data. They need to join and link information together - defining relationships that will point to previously unseen opportunities for service improvement or revenue generation. But all of this is easier said than done...

Today’s volumes of information are huge and growing almost exponentially. Many modern applications deal in petabytes of data – a scale and complexity that traditional BI solutions were never designed to handle. Slow, clunky and cumbersome they have left frustrated business users searching around for a new solution... a tool capable of turning complex, multi-dimensional, data into a usable information asset capable of delivering real bottom-line benefit and driving significant service improvements.

The snag is, high-end tools are often cost-prohibitive and most require the ongoing support of a development infrastructure. Business users are, therefore, straight-jacketed. They have to

Private, Confidential and Commercially Sensitive

request specific data from IT departments that often have to create new data sets and write new programs which can take weeks to complete.

This is hugely inhibiting because it denies business people the chance to explore and to experiment with data. When there is no scope to play with information, there is precious little chance of a 'Eureka' moment – the serendipity of stumbling on a hidden truth that can literally change the fortunes of business.

That is why Connexica believes that a new approach to data analysis and reporting is an absolute imperative. An approach that harnesses the technologies which have grown out of the information explosion to directly address the issues inherent in mining and refining vast amounts of randomly stored data¹.

Introducing NetSearch

NetSearch has been built specifically as a 'next generation' Data Analytics tool .It combines the best from two very different technologies - Search and Business Intelligence – by moulding them into one integrated solution.

The product utilises the raw power of search technology in order to assemble data for rapid querying and filtering. However, unlike standard search engine functionality, it has been enhanced to provide comprehensive data mining and analytics.

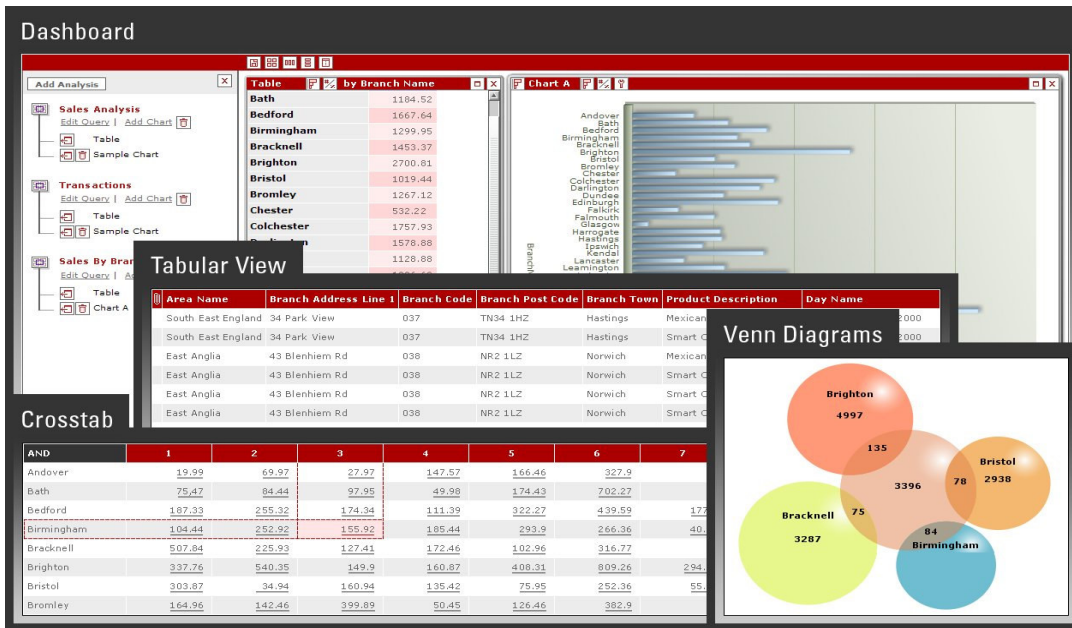
Search technology provides the technical infrastructure to support extremely large volumes of data² and delivers this through an intuitive, fast, user friendly and familiar query language - *natural language search* - which is already familiar to most business users.

NetSearch delivers information to the user through an intuitive, business-orientated front-end. This permits the end-user to run both simple and sophisticated queries against all or part of the organisation's data without the need for intervention or help from IT personnel.

This is seriously empowering. There's no more reliance on IT. No more pre-defining of information requirements. No time-consuming need to assemble information into tightly structured data-sets. Instead, business users can explore and interrogate huge bundles of randomly sourced information themselves. They can use enhanced, real-time search technology to weave strands of seemingly unrelated data into an instant context.

¹ Structured relates to fixed data layouts such as those defined in a relational schema. Unstructured relates to variable content such as that held in XML, documents and emails.

² NetSearch is able to index both structured and unstructured data. Structured data is information contained in one or more databases - where as unstructured data relates to information such as web pages, word documents and spreadsheets.



Data Analysis with NetSearch

Mining and querying data

Traditionally, one of the biggest restrictions for reporting and database technologies has been their inability to support real-time ad-hoc analysis across large volumes of data.

Analysis has become more and more difficult as data volumes grow and information requirements become increasingly dynamic.

For example, an advertising campaign or a major sponsorship programme could easily trigger a sales explosion that will blow an almighty hole in traditional stock forecasting models. In such situations, historical trend analysis will be of little value to buyers and merchandisers. They need a flexible mining and querying solution that analyses the very latest information to provide instant answers to ad-hoc questions.

Unfortunately, ad-hoc enquiries are not a strong suit for conventional technologies. They require the user to write code (such as SQL for relational databases or MDX for OLAP cubes).

This demands high levels of technical expertise. It also calls for a comprehensive knowledge of how data is physically stored and linked – only then can the correct information be swiftly and efficiently accessed.

In the case of SQL, the problems become even more complex. The skill required to query data increases as the volumes rise. Inefficient or badly written SQL can take hours or days to run. In turn, this can easily consume all of the machine's resource, radically reduce accessibility, impair performance for other users and potentially cause far-reaching damage to businesses.

Private, Confidential and Commercially Sensitive

NetSearch addresses these issues by:

- Providing an interface that allows business users to perform ad-hoc enquiries against the data without reverting to code
- Holds the information in a search engine which allows searches against the entire data set(s) to be returned with sub-second response times without the risk of running badly formed, inefficient code
- Delivering dynamic analytics to enable the user to view the results of a search as lists, tables, charts, VENN diagrams or GIS maps.

NetSearch is written to specifically address the issues of large and potentially volatile data volumes. Through the use of advanced indexing and search technology, Connexica has developed a highly intuitive solution that allows people in the business to directly query and mine information in real-time.

Joining information together

Most BI solutions use “rows” as the default currency for report creation. Reports are based on filtering and joining rows of data together through the use of the SQL inner, outer, left and right join constructs.

Other than the performance and scalability issues referred to earlier, these join constructs are not enough to answer what might, to a business user, be considered fairly straight-forward questions.

For example, an event-based dataset can provide entrepreneurial business leaders with an extensive and eminently minable transaction record. Shrewd managers will want to do much more than merely query the number of events in a given time-span. Typically, they might decide to target specific customer groups, such as:

“Those customers who have purchased from a specific store more than ‘n’ times and spent, on average, in excess of ‘n’ pounds”

They might then decide to drill down for a far deeper analysis of this prime customer profile.

Unfortunately, this is another instance where traditional IT departments struggle to provide instant, end-user insights. IT teams would have to resort to time-consuming programming because the data needs to be linked together by a set of business rules; not SQL joins.

These shortcomings crystallise both the problem and the need. New approaches and techniques to data enquiry must revolve around business-users and their need to ask questions across large, dynamic data sets. Traditional, row-based querying techniques simply cannot deliver the deep, dynamic Intelligence that management now needs.

NetSearch specifically addresses these business issues. It empowers business users. It allows them to dynamically define their own data clusters based on business rules (as opposed to SQL) which can then be searched and analysed directly by the end-user.

As well as linking data on a time-line and the occurrence of one or more key events, NetSearch also allows you to:

- Create new composite search engines by physically combining one or more datasets
- Pass contextual search information from one index to another. This allows filtered data in one dataset to be used as the basis for a search in a different data set. For example: "Find me all the customers that have bought X and Y in dataset 'A' and then cross-relate them with dataset 'B' to find out if any of them has a bad credit rating"
- Search across different datasets containing structured and unstructured data. For example, you can find a customer by name and post code and then locate documents and other literature related to that customer from an almost infinite range of unrelated source data

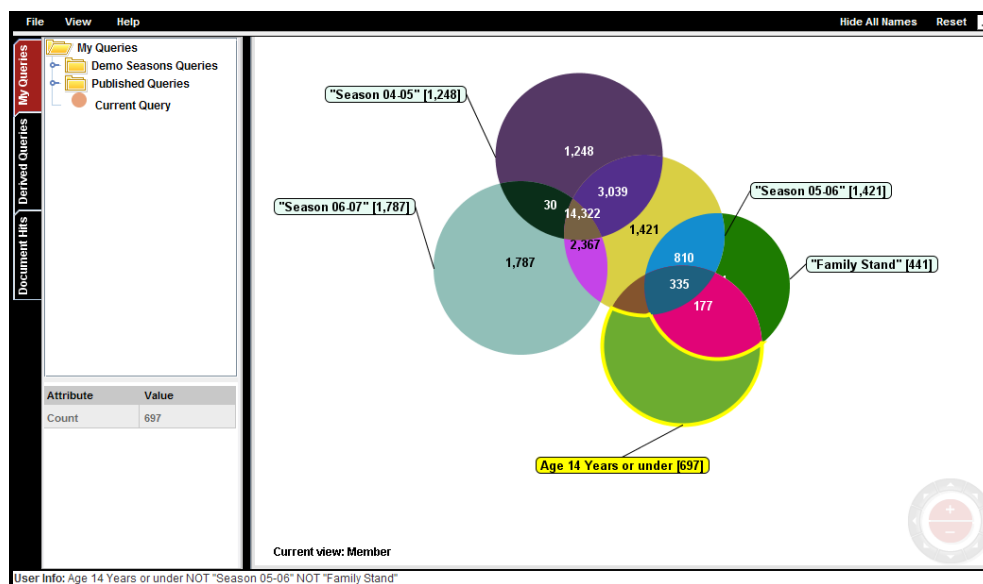
NetSearch not only makes the process of joining and linking data across datasets transparent but also uses this ability to open up new methods and techniques for ad-hoc querying that are not easily replicated in other technologies.

Segmenting and clustering data

Statistical packages - such as SPSS and SAS - provide the functionality to segment and cluster data. However, for large and complex datasets, these facilities are often only available through the generation of proprietary code and the production of mini data extracts which are then imported into specialist segmentation and clustering tools.

NetSearch provides real-time segmentation and clustering capabilities as standard through the use of VENN diagrams where users can:

- Create segments or clusters depending on their transactional behaviour in real-time
- Predict future behaviour based on previous activity
- Identify correlations between different attributes



Private, Confidential and Commercially Sensitive

NetSearch allows business users to identify relationships in data and drill directly down to the underlying data without the need to create off-line data extracts or write proprietary code.

Conclusion

The massive increase in data volumes and the sheer scale of the services that are being deployed over the Internet and by large corporate organisations are now creating real issues as well as huge opportunities for service improvement and revenue generation.

Traditional BI tools are creaking under the weight of data. New technologies have to be developed and fully utilised to address this pressing problem.

The larger the data volume, the harder it is to access. It is also undeniably true that larger data volumes also offer far greater potential value. The key is to make information easily and instantly accessible to the right people – the people at the very heart of the business.

With the advent of search technology has come the development of new products that not only leverage the power of search but also bring together the best elements of BI to provide a new generation of affordable analytical tools.

NetSearch has been written, from the ground up, to allow business users to rapidly analyse and interrogate up-to-date information, drawn from both structured and unstructured data sources.

At Connexica, we firmly believe that the adoption of tools such as NetSearch will transform the world of real-time analytics, data mining and guided search.

By developing technology that empowers business users, it will grow to meet the dynamic needs of management. It will maximise the competitive value that companies derive from their information assets and give decision-makers the instant, in-depth Intelligence they desperately need.