

# Data Migration

## NetSearch White Paper

***How NetSearch can be used to improve the efficiency and accuracy of data migration***

Created: 10/12/2008

Version: 2.0

Created by: Richard Lewis

Ref: NetSearchForDataMigrationV2.doc

Private, Confidential and Commercially Sensitive

## Data Migration

---

Data migration is the process of moving data from one or more multiple systems to another.

Data migration is normally carried out in the context of databases when a system is being replaced, decommissioned or data from multiple systems are being consolidated into one to provide a single, common reporting environment for the organisation.

The process of data migration typically involves the use of Extract, Transformation and Load (ETL) tools to take data from one or more source systems, apply business rules to transform and cleanse data, then move into new consolidated table structures e.g. a data warehouse.

Data is often held in multiple systems and formats so the process of unifying the data and removing duplicate values is difficult and requires a great deal of manual intervention and assistance from different areas of the business.

Data migration cannot be a process that is restricted to the IT departments as the people that are most likely to understand the authenticity and accuracy of data are business orientated employees entering and using the information on a daily basis.

Data migration projects need to be able to address the following issues;

- Legacy contracts and client information that are no longer utilised by the business. Is this information still required by the business or should it be erased?
- Which legacy information needs to be retained in its original form for risk, compliance and audit purposes?
- Where contracts or data has been novated due to acquisitions and mergers, how is the information to be referenced and should the data be changed to reflect the change in ownership?
- Coding issues where temporary codes were entered into the system before a revised code was made available. Can this information now be reconciled?
- The validity of information held against a product or customer. For example which offers can or have been applied to a product or service and what are the business rules that govern what can and cannot be sold or purchased by a client?

There are many such examples where the migration of data becomes a business process issue as well as a technical one. The key challenge is how to get business users involved in the data migration process to help ensure technical staff understand the issues inherent in the legacy data and have sufficient knowledge and understanding of the business to construct the right ETL rules.

## Introducing NetSearch

---

NetSearch has been built specifically as a 'next generation' Business Intelligence (BI) tool. The product utilises the raw power of search technology in order to assemble data for rapid querying and reporting purposes.

NetSearch allows users of any skill level to search, enquire, assemble and analyse information stored anywhere in your organisation through a single, easy to use, browser based interface.

An intuitive, business-orientated front end provides the end user the ability to run both simple and sophisticated queries against all or part of the organisations data without the need for intervention or help from IT personnel.

## Using NetSearch to support Data Migration

---

*"Traditional tools, on their own, are not enough. We need data migration methodologies; we need experienced professionals, and we need special-purpose tools; or at least add-ons to conventional tools that have been specifically designed to support data migration." - Bloor Research*

NetSearch allows business users and IT staff to gain a rapid insight into the information assets of an organisation by making data accessible through a concise, fast, easy to use search interface.

By gaining a better understanding of what issues exist within your information stores, IT departments are able to enhance the data migration process and provide a consistent and consolidated view of the organisation.

Through the enablement of business users to rubber stamp business rules and make decisions around data quality and coding, the chances of creating an accurate representation of the business in the new database is massively improved.

### Index data directly from your source systems

Data contained in any of the organisational databases and repositories can be accessed and stored directly by NetSearch in a highly optimised search engine<sup>1</sup> that provides a single interface for exploring information that originated from multiple systems. Users can perform simple searches on the data, perform record counts and understand the data linkages across the various systems.

---

<sup>1</sup> A search engine is a collection of one or more NetSearch indexes. An index is a highly optimised copy of the source data stored separately from the originating systems. Search engines allow rapid free text and analytical searching across vast amount of data with sub-second response times.

## Data discovery through data mining

*"...there are the three things that you need to do: "profile your data, profile your data and profile your data". You need to know where your data is (if necessary going through a discovery exercise) and you need to understand it. If you understand your data fully (which is the role of profiling) then the remainder of any data migration project should be relatively straightforward" - Bloor Research*

Without the need to involve IT, business users can freely mine and explore the data and feed back to IT informed knowledge about which data should be taken into the new migrated database and what can be ignored or archived.

The ability to access information via a simple and intuitive interface allows business users and IT staff to better understand;

- a) Which systems contain key information and data items
- b) Where data appears duplicated
- c) Where the most accurate version of the data exists
- d) Which data quality issues exist in the data e.g. invalid coding, missing data, mis-spelled names and addresses etc...
- e) Gain rapid understanding of data issues
- f) Help construct business rules based on an informed view on the quality and availability of data
- g) Share the ownership of the data migration process with the IT department.

## Data Consolidation

*"In terms of actually moving the data, traditional techniques assume that you are moving database tables (or their equivalent in a non-relational environment). However, a better approach is to migrate business entities (for example, a customer together with all of her orders). Amongst other things, this is easier to understand at a business level, enables incremental migration, facilitates parallel running and makes fallback easier to implement." - Bloor Research*

Seeing information from multiple data sources through a single interface allows you to have a better and more complete picture of the information available to your business.

Related information is likely to exist in multiple systems where the business understands the links and associations between the different data entities; however the systems are treating the various data stores as individual silos of information. These disconnected systems cannot be joined together easily by traditional Business Intelligence technologies.

NetSearch has a number of features that allow you to merge and link data from different indexes together to provide a consolidated and integrated view of you data.

NetSearch allows you to easily consolidate your data together and provides you with a single view that can be centred around key business entities and inter entity relationships as opposed to the restricted views exposed through multiple legacy systems and databases.

Private, Confidential and Commercially Sensitive

Building a consolidated view of your data reduces duplication and allows you to bring together associated data. This means that whenever you make an enquiry you can easily navigate from one entity to another without having to be concerned about logging into multiple systems and re-establishing the context of your search.

## Historical Analysis and Archiving

One of the big advantages search engine technology has over traditional database technologies is the scale to which a search engine can grow and the performance that can be achieved when performing ad-hoc queries against data.

Data that would traditionally be backed up to an off-line storage system can now be kept on line by storing the data as a NetSearch index.

Current and historic data can be stored in separate or combined indexes allowing you to quickly access up-to-date or historic data with sub-second response times. Reducing the need to lose or archive information enables you to:

- a) Streamline the data requirements for the new system without losing historical data
- b) Create a copy of the source data which can be kept on-line indefinitely
- c) Use a copy of the data as a source for historical analysis and validation against the migrated data
- d) Retain historic coding for compliance and investigative analysis
- e) Compare data volumes and ratios between entities from the original and migrated data and use this information to analyse performance improvement by retaining legacy data for comparative analysis

## Conclusion

---

Data Migration is a task that needs to involve business users, not just technical IT staff.

Successful data migration requires collaboration between the business community and IT to ensure that the quality and content of the information transferred into the new system supports the current business model and provides an enhanced environment for integrated reporting and analysis.

NetSearch enriches the data migration process by allowing people from within the business to directly participate in the data migration process to streamline the process and improve the accuracy and completeness of the final system.

*References have been taken from an article by Philip Howard of Bloor Research*

*<http://www.bloor-research.com/research/market-update/952/data-migration.html>*