



The Top Three Data Migration Risks (and how to avoid them)

EXECUTIVE SUMMARY

Many current approaches to data migration suffer from a consistently low success rate. A major part of the problem is the unplanned issues that often occur at the later stages of the project.

The top three areas of uncertainty are:

- ▶ Unrealistic estimates of data quality
- ▶ Inaccurate, missing, or out of date source system documentation
- ▶ Inability to reconcile the target system data to the source system

An unexpected problem in any of these areas can cause project costs and timelines to be significantly compromised.

This paper outlines a “data driven methodology” that allows you to eliminate these uncertainties at the earliest stages of your project, so that the planning is more accurate and there are fewer surprises later.

Using this methodology, the project team can access the source data as soon as it is identified. This allows direct query into that source data, enabling data quality issues to be accurately quantified at the outset. Next, as an extension of the query capabilities, much of the required metadata can be generated directly from the source data, reducing the reliance on possibly suspect or absent source documentation.

Finally, the extensive query, file comparison and drill-down capabilities allow source and target differences to be quickly identified, and resolved.

According to studies, 70% - 90% of data migration projects fail

Data Migration and conversion can be difficult and daunting tasks, but you might be surprised to learn how difficult. Depending on the study, between 70 and 90 percent of data migration projects either fail outright or run over budget, with an average cost overrun of ten times the original estimate.

Industry analysts estimate that at any given time two thirds of Fortune 1000 companies are engaged in some form of data migration or data conversion. This translates into billions of dollars of unexpected expenditures. For projects that fail completely, these costs are totally wasted.

THE TRADITIONAL PROCESS

These are almost unbelievable statistics, but when you understand the standard process, it is easy to see how this might occur.

Typically, a data migration project starts with a thorough review of the source materials, including source documentation, COBOL copybooks and other metadata. The process is a costly one, typically involving a large team comprising administrators, designers, business analysts, subject matter experts and others. In some cases the data itself is examined, but where legacy data is involved, the tools available are either limited, or involve very complex handling. As a result, the data is seldom examined rigorously in advance.

This stage is meant to ensure a complete understanding of the data, but unfortunately, seldom achieves this goal. The problem is that the source materials are often out of date, inaccurate or incomplete. As a result, the team gains an incomplete, or worse, inaccurate understanding of the data.

Based on the understanding that has been developed, the data elements are mapped to the target environment, appropriate transformations and cleansing routines developed, and the target system is loaded.

Comprehensive data analysis only takes place once the data has been loaded into the target database or some intermediate source, as this is the first opportunity where most tools can access the data.

It's easy to see why the loaded data doesn't initially meet expectations. In some cases the data was much harder to access than expected, in other cases the estimates of data quality were significantly inaccurate, and in other cases still, the inevitable inaccuracies in the documentation have resulted in an imperfect mapping.

Sometimes these various problems result in errors that can at least be specifically identified, but in many cases they just result in differences between the two systems that are very difficult to reconcile. The result is that most projects are forced into a cycle of rework, until a successful result is achieved or the project is abandoned.

Given this scenario, the failure statistics start to make sense. What is needed is a way to eliminate these uncertainties at an early stage, to minimize or eliminate the wasted effort that results.

A DATA CENTRIC APPROACH

*The data itself
doesn't lie*

We at Arbutus like to approach the problem of data migration from the opposite direction. It is recognized that source materials are almost certainly inaccurate to some degree and data quality estimates are difficult to quantify in advance, so why start there? Instead, we remove the barriers to accessing the source legacy data. While documentation can be, and often is, inaccurate and estimates are never more than estimates, the data itself doesn't lie.

Rather than being forced to rely on inaccurate documentation, you instead start by directly querying the source system data. You develop metadata that is based on the actual data, and therefore cannot be inaccurate or out of date. This knowledge allows you to form a picture of the data before reviewing the documentation.

As well, once you are querying the data, you also have the ability to query any data quality issue you wish to test. This means that rather than estimating your data quality you can actually quantify it. Using additional capabilities of Arbutus' tools you can also dynamically correct many of the problems you identify.

...significantly improved results on the first data load

For other types of data quality problems, other tools may be available that can correct the problem. In these situations, Arbutus can supply the source data directly to those tools.

Depending on the size of the project and the quality of the metadata created, you may be able to reduce or even eliminate efforts spent reviewing the source documentation.

For the documentation you do review, you can immediately test the information against the data itself. This allows you to quickly identify inaccuracies, and prevent wasted efforts.

By improving your understanding of the source data from the outset, you can create a more accurate mapping and transformation of the data elements. The result is significantly improved results on the first data load. This translates directly into a lower need for rework, much less effort to achieve, and most importantly, a dramatically reduced risk of failure or overruns.

Once you have loaded the data into the target system, your problems typically aren't over. It is likely that there will still be differences that must be identified and addressed. These represent gaps in our understanding that still remain, despite our best efforts.

These differences fall into two categories: critical ones that must be addressed and other non-critical differences (that are often missed due to limited access to the source data).

Arbutus's ability to access the source data and target data is an invaluable aid in identifying and correcting both these classes of differences. For critical differences, Arbutus's ability to drill down into the data allows quick identification of the problems.

For non-critical areas, our extensive source data query capabilities allows a much more thorough comparison of the two systems than would normally be possible. The result is a much quicker, and more thorough, comparison than is typically available.

In data migration projects, the worst things are surprises, and surprises are particularly costly at the latter stages of a project. By starting with the actual data, you get fewer surprises, and they typically happen very early in the process, when their impact is minimal.

ARBUTUS MIGRATE ENABLES A DATA CENTRIC APPROACH

To be able to start with the data, you need a tool set that exceeds the capabilities you have come to expect. This is where Arbutus Migrate comes in.

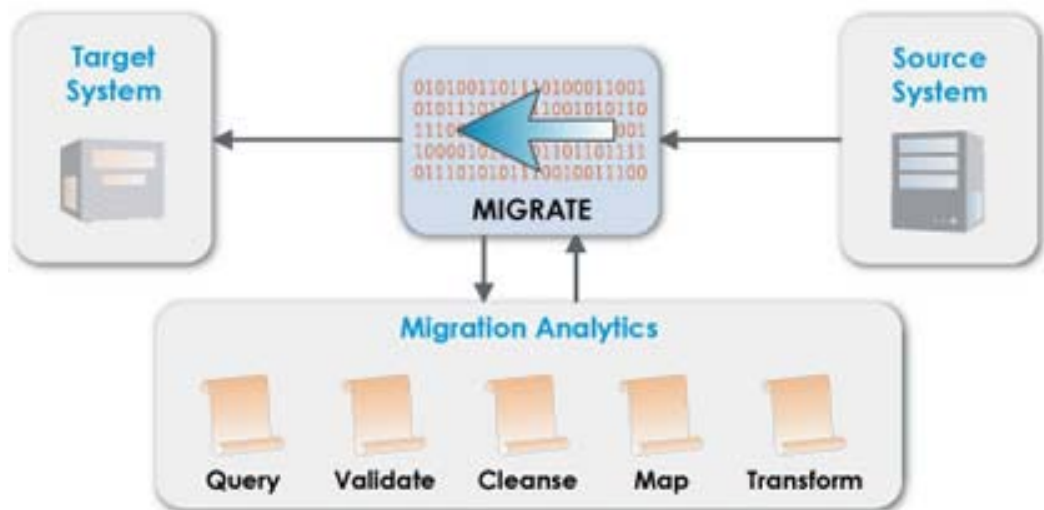
Arbutus Migrate uses technology that has been proven through over 25 years of use in the audit departments of Fortune 100 and other large companies. In serving the ever increasing needs of the audit community, the technology has matured to the point where it has the capability to directly access virtually any legacy and non-legacy data source, regardless of complexity.

This ability to access even the most complex files, including mainframe and AS/400 legacy files, is what distinguishes Arbutus Migrate from other alternatives, and what makes a data driven methodology possible.

When you start by accessing the source data, you need access to all the data. When only a portion is accessible, the value that can be derived drops dramatically. Arbutus Migrate can directly access virtually any mainframe data set, including VSAM, ISAM and QSAM sequential flat files. It can directly process variable length records, multiple record types, and every data type supported, including custom date field coding and bit level fields as well as overloaded fields and OCCURS clauses. In almost 20 years of use we have yet to encounter a mainframe file that couldn't be accessed.

As well as mainframe legacy data sets, Arbutus Migrate can also access any DB2 tables or IMS segments. Each of these sources can be freely combined, so that DB2, IMS, VSAM and sequential data sets can be used to create a virtual data model that is built on the source data itself.

*...proven through
over 25 years
of use*



Where mainframe resources are at a premium, Arbutus Migrate also allows you to easily stage the actual data in a Windows environment, taking the processing off-line.

Arbutus Migrate offers similarly complete data access for AS/400 platforms as well. DB2, IFS, Flat files and even spool files can be accessed, combined with each other, and combined with data from mainframe platforms.

All of this data is presented in a tabular format that can be queried, analyzed and transformed by the tools supplied with Arbutus Migrate. In addition, you can also access this same data with any other Windows application or SQL tool, via the LegacyLink™ ODBC driver included with the package. This allows you to apply the tools of your choice to profile, analyze and transform the data.

ARBUTUS MIGRATE & ARBUTUS ANALYZER

Migrate uses the Analyzer Windows application to provide a powerful set of analytic capabilities to complement its data access. Using a simple point-and-click interface you can create an almost limitless variety of analyses.

*...powerful and
flexible analytic
capabilities*

Exception reports can be generated based upon virtually any criteria you can conceive. You also have the ability to sort, summarize, subset and otherwise process the data to better highlight other data relationships or problems.

You can check the referential integrity of your data (even across disparate sources), scan for blank or invalid data values, check for inconsistent data, confirm data formatting, and even identify every unique code value used within a column. Fuzzy comparisons can easily be performed to de-duplicate or harmonize data.

Analyzer includes a wide range of column statistics, stratification, sequence, gaps, duplicates and exception testing as well as cross-tabulation. These can be combined with our powerful and flexible transformation and manipulation capabilities to support virtually any analytic requirement.

Arbutus Migrate also provides procedures and virtual columns to implement any business rule or transformation, allowing you to leverage the data access to support your mapping and transformation requirements. For most situations just this one tool will support your entire data migration project.

If you choose a traditional approach to migration, Arbutus Migrate can provide invaluable insight for your existing processes because it directly accesses the source data.

- Visit www.ArbutusSoftware.com/form-eval to request a 30-day free trial of Arbutus Migrate.
- Request a **20-minute web demo** to see Arbutus Migrate in action. Email us today or call toll-free.



#270 - 6450 Roberts Street
Burnaby, British Columbia
Canada V5G 4E1

Toll-free: 877.333.6336
Direct: 604.456.6336
Fax: 604.437.7872
info@ArbutusSoftware.com
www.ArbutusSoftware.com

Based on 25 years of innovation excellence, Arbutus delivers the very best in purpose-built audit analytics technology to meet the exacting demands of today's business environment. Auditors, business analysts, and fraud investigators rely on Arbutus to enhance their testing, analysis, and compliance capabilities.

The data universe is wide and varied. One of our core strengths as a technology firm is the ability to easily work with all types of data, both legacy and non-legacy. Arbutus solutions allow auditors, IT, and business professionals to overcome many of their current constraints in areas such as data migration, data quality, fraud detection, and data analysis.

Arbutus Audit Analytics, our flagship product suite, is a proven solution used by auditors, business professionals, IT, and management all over the world. With outstanding customer service, strong product support, and flexible licensing, Arbutus Software offers the best value for advanced data conversion, migration, and analysis solutions.

Arbutus Software Inc. is a privately held company based in Greater Vancouver. For more information about our company or products, please contact us.

