Data Migration for Legacy System Retirement

A discussion of best practices in legacy data migration and conversion.
TABLE OF CONTENTS

The Importance of Legacy Data Migration 3
Facing Reality 4
Checklist 5
  1 | Business Stakeholder Participation 5
  2 | Estimating Effort 5
  3 | Data Validation Plan 6
  4 | Managing the Specification 7
  5 | Refresh and Regression Testing 8
  6 | Go-Live Strategy 8
  7 | Ongoing Data Governance 9
  8 | Process Trumps Mechanics 9
MDX Overview 10
Conclusion 11
About Gaine Solutions 12
The Importance of Legacy Data Migration

Many times we look for new applications to meet our business needs – requirements are defined, software is evaluated, software selected, software configuration begins – Ready Set Go! More like “Ready, Set, ‘but what about the data’?” Much of this data is sitting in legacy applications and must be converted and migrated to the new application – often with limited or outdated documentation for the legacy systems or dependent on resources that are no longer with the company.

As companies undertake the transformation of their core systems it is crucial they pay sufficient attention to the migrating of legacy data and conversion of historical information into the new applications and analytic platforms. The overwhelming majority of projects that involve the migration of legacy data to a new platform are plagued by costly overruns or even project failures. Recent research by Gartner confirms the challenges of data migration but the quantum is astonishing.

In this paper we discuss the most important considerations of data migration projects to avoid becoming part of the unfortunate majority.
Facing Reality

The harsh reality for the majority of data conversion migration projects from legacy systems is the project team moves from an initial comfort with their approach, to a realization that the data conversion is not going well and finally into an extended period of remediation. It is not uncommon that the costs of remediation substantially exceed the original data conversion budget as resources are thrown at the problem in an attempt to reach the go-live date through brute force.

What is certain is that none of the 83% believed they would fall victim to the “harsh reality” when they were planning their projects. The data migration effort is underestimated for a variety of reasons:

Data Quality Problems – Even organizations that are aware they have data quality challenges fail to understand the effort required to fix all that is incomplete, inaccurate or inconsistent in their information.

Delaying Functional Testing – Getting data loaded into a new platform is not a measure of success. The data loaded into new applications may be valid but not correct which only becomes apparent during functional testing. Functional testing should happen early in the plan and continue in parallel with data migration.

Lack of Flexibility and Specification Changes – Gaps in Master Data specifications and inaccurate assumptions of data quality/availability, combine with other changes to business requirements to create a “moving target” which in turn creates a huge volume of specification changes. Traditional data migration methods, based on a waterfall project approach are ill suited to the churn of a data migration effort.

Data Validation and Audit – Validating the data post-transformation extends beyond merely looking at data in Excel spreadsheets. The data validation process must combine legacy data with the configuration specification and the new platform in a reporting capability that will provide checksums, record counts and like-for-like comparisons between the old and the new data models. Balancing and validation of financial and statistical data can be a significant undertaking.

Lack of Data Knowledge – Documentation of legacy systems is typically inaccurate, incomplete or missing entirely. An organization rarely has the people with the time and knowledge of existing data to fill these knowledge gaps. Many insurance companies with aging systems may have re-used fields with different context for different time frames adding to the difficulty of conversion activities.
Checklist

The problems that create project overruns are typically only visible in the later stages of a project when it is too late or very costly to do anything about them. Organizations should carefully assess their preparedness prior to starting their data migration in order to avoid the pitfalls.

Here we discuss eight key areas that should be given close attention before embarking upon a data migration project.

1 – Business Stakeholder Participation
An organization should be realistic about the time commitment required from business subject matter experts (SME’s). The very people that are indispensable to day-to-day business operations are typically the same resources required throughout the system transformation effort to make important decisions and to provide key insights. As painful as it is, an organization must be prepared to make key resources available to the project team throughout the migration effort. Failing to secure the business resources required to define, clarify and validate the conversion will result in the technical team having to guess at business decisions. This guesswork by the project team typically remains hidden until the impact is felt during functional testing in the later stages of the project.

- Identify the business SME’s by name and role.
- Agree business ownership of the data with key business stakeholders.
- Secure dedicated time commitments from SME’s throughout the project duration.

2 – Estimating Effort
Data transformation effort is driven by complexity more than volume. Basic measures such as the number of data objects or number of records to be converted are poor indicators of the time and effort required for migration. Estimates based on hours-per-transformation become meaningless when a single issue in a complex business process may give rise to hundreds of hours of additional
effort. Even apparently simple mappings can be complicated by obscure data issues relating to granularity, integrity or consistency of the legacy data.

During the scoping and requirements phase, and most definitively before committing to a go-live date, an organization should undertake a data profiling exercise to assess the “as-is” data against the “to-be” model. These insights will help the project team make a more realistic and accurate plan for the conversion effort.

3 - Data Validation Plan

The project team should have a detailed plan of how the new data will be validated against the specification. It is not sufficient to say that users will validate data in Excel reports without giving consideration to the inventory of reports that will be required. This inventory of reports and the data required to populate these reports will drive the design of the underlying data store to capture the key metrics and transformation steps throughout the process.

It is not uncommon for a project team involved in a system transformation initiative to produce upwards of 200 validation reports during the testing phase of the project. Many of these reports reply upon interim counts and temporary calculations; if you don't have a plan upfront to capture this process information, then discovering this during validation is too late.

- Complete the “to-be” model and basic legacy data mapping.
- Profile legacy data against the new model.
- Identify gaps and dependencies in legacy data in collaboration with business SME’s.
- Create an inventory of validation reports and how they will be used.
- Identify all data required for validation and create a data model to store this data.
- Allocate project time to produce the reports and support the validation process.
4 – Managing the Specification

Throughout the course of a system transformation project the data conversion specification may undergo hundreds of changes – both large and small. These changes are driven by evolving business requirements, dealing with unknown issues in legacy data, and managing newly identified use-cases that arise during functional testing.

It is not sufficient to merely control updates to the specification documents – although for many this is a challenge in itself. You should be able to determine the impact of each change to the specification so that you understand how many transformations or data objects are impacted by each change. This helps you plan how to deal with the change and how to manage the impact upon the project plan.

Another consideration is that within a “to-be“ data specification a single data element may be included within multiple areas and, as such, a change in one area may cause unforeseen problems in another. It is extremely difficult, if not impossible, to reliably manage these interdependencies using a manual process and spreadsheets.

A final thought on managing the specification is that each requirement expressed within the specification should be linked to one or more validation reports. When the specification changes so too should the associated validation process.

The specification should be tightly managed, preferably within a database where the integrity can be enforced and changes are more easily analyzed. Remember, the specification is not just documentation artifact; it is the central point of control throughout the project.

- Define a structured process to store and manage changes to the specification.
- Provide a method to report on the impact of changes to the specification.
- Create a link between the specification and the validation reports to test the specification.
5 – Refresh and Regression Testing

At some point in the migration project, the source data will need to be refreshed. Typically data is refreshed prior to each testing cycle and certainly prior to go-live.

Two challenges present themselves during a refresh. The first is to ensure that all of the transformations and data manipulations applied to-date are re-applied in the correct sequence; the second is to validate that the refresh has not introduced (or re-introduced) any data errors.

The ability to reload the data and to test each requirement in the specification against the results is considered a test harness and is a critical element of the migration approach.

6 – Go-Live Strategy

Whether as part of a phased implementation or a “big bang” go-live strategy, it is important to plan for business operations during, and immediately following the period of go-live.

There is inevitably a period of time between the final refresh of legacy data and completing the load and validation of the new platform. In a large project this period may stretch into several weeks. During this time either the legacy systems are “frozen” and no new updates are allowed, or a process to capture these legacy updates and apply them to the new application platform is required.

Post go-live attention quickly shifts to the auditors who are required to ensure that the data loaded into the new systems can be reconciled with the retired
legacy systems. Auditors use a combination of high-level aggregations and specific detailed tests to satisfy their analysis. The migration team should not overlook the need to collect detailed history and lineage of all the new data to support this audit process. If this detailed audit trail is not readily available then the audit process will become an unwelcome distraction to the business as it adjusts to the use of the new application.

7 – Ongoing Data Governance
The transformation of legacy systems presents an ideal opportunity to introduce formal data governance processes to an organization. The data migration project will require the collaboration of IT and business stakeholders to define data rules and policies. These same rules and policies may be used as the initial basis for data governance within the organization.

No application system is infallible; at the root cause of bad data are poor process controls and a lack of standards and training for system users. These same underlying factors will continue to introduce errors in the new system and dilute the investment in data quality throughout the migration.

8 – Process Trumps Mechanics
Be wary of an over-emphasis of the ability to manipulate data and connect systems rather than the ability to manage the migration process. Technical staff may present a compelling plan for the transformation of legacy data but the mechanics of the transformation are just a small part of the overall recipe for success.

Concentrate the project planning on the first seven points in this section before you get concerned with the tool that you will use to handle database connections and transformations.

- Ensure that you have a good estimate of the time required to load and validate the new systems
- Define how data will be synchronized between systems during any period of parallel operations.
- Get agreement from the project steering group on what criterion define the end of the pilot phase.
- Include ongoing data governance as part of the core system transformation plan.
- Ensure that you have a process-based strategy not a tools-based strategy.
MDX Overview

It is no coincidence that the MDX data migration platform developed by Gaine Solutions helps tick all the boxes discussed in the previous section. MDX is a purpose-built platform to address the challenges of legacy data migration. Since 2007 MDX has been used to retire more than 2,000 legacy applications in support of core system transformation initiatives and continues to help companies manage risk and reduce cost within data migration projects.

MDX combines multiple components in an integrated environment to support a best-practice methodology for data migration.

Process control is at the heart of the MDX solution. With MDX all aspects of the migration can be coordinated and integrated and throughout the migration no change to either the specification or business data goes unmanaged. This level of control allows data to be refreshed in an automated manner and it provides a central point of control for the project team.

MDX provides not only cleansing, transformation and matching capabilities but can also integrate with third-party tools or custom programs if these are available. The MDX platform provides data profiling tools to assess the legacy data against the new data model as well as a library of validation reports that are supported by a validation data mart. In addition, the MDX specification repository provides control of the specification, impact analysis and a link to validation reports.

Data governance workflows are provided to enable ongoing management of data quality. MDX provides templates for an extensive number of data governance events as part of the platform. MDX data governance templates include workflows for managing changes to key attributes on the master record, how to handle conflicts between matching records, resolving suspect duplicates between systems, resolving inconsistent relationships between records and many others.

MDX should not be compared to ETL tools or data cleansing tools. MDX does not replace the transformation or connectivity provided by an ETL tool but it does enable an ETL tool to be used in a controlled and effective manner. The MDX platform provides you with an integrated environment ideally suited to the churn of legacy data migration without the need for you to design and build this capability while you have plenty of other things calling for your attention. Furthermore, it eliminates the need for the user to build out all of these processes thereby saving significant time and money.
Conclusion

Legacy data migration is more difficult than it initially appears and getting it wrong is expensive and painful. Don’t let the mechanics of how legacy data will be transformed overshadow the need to manage the process.

Give careful and objective scrutiny to the checklist in section two of this document and, if you have any doubt about your approach, be prepared to pause. It is far cheaper and easier to fill a gap up-front than to dig your way out of trouble later in the project.

Data migration is a specialist discipline and requires the coordination of a lot of moving parts. It’s not enough to say “we have smart people and good tools”, you need experienced people and integrated tools to be successful.

The MDX platform is much more than just an accelerator to a legacy migration project; While MDX relieves the project team from the overhead of designing and building a migration capability, what is more important is that MDX encourages, even enforces a best-practices approach. The project team will find that each component is seamlessly integrated with the next, and very soon everything from upfront data profiling to data validation during testing and ongoing data governance are working together to provide visibility and control to all stakeholders.
About Gaine Solutions

Gaine is an Enterprise Data Management specialist, creating value for its clients through a range of business services delivered in an on-demand commercial model. The unique Gaine approach accelerates time-to-value and minimizes the time, cost and risk inherent in data intensive initiatives. Gaine is the developer of the widely deployed Master Data eXchange (MDX) platform and provides EDM services to some of the world’s largest and most respected Global 2000 organizations through its offices in San Francisco, Dallas, New York and Cape Town.

For more information please visit
www.gainesolutions.com