



DATA MIGRATION FOR PROJECT LEADERS

A Guidebook

Table of Contents

Introduction **3**

Part 1 The Route to Best Practice **4**

Part 2 Preparation **7**

Part 3 Executing the Project **16**

Part 4 Closedown **27**

Synopsis

The project leader plays a critical role in ensuring a successful data migration but exactly what responsibilities and activities should project leaders fulfill?

This guidebook provides the answers with a series of detailed sections including essential activities such as planning, forecasting, risk management, team selection, communication and collaboration.

Introduction

Data migration projects can pose significant challenges for project leaders.

As a discipline, data migration is still in its relative infancy. There is a noticeable lack of industry associations, formal education and published best practices compared to similar disciplines such as data warehousing, data quality and data integration.

There are many pitfalls awaiting the poorly skilled and inexperienced project leader and as a result far more data migration projects fail than succeed. The aim of this guidebook is to help reverse this trend by providing expert advice and a set of best practices to help chart your data migration project to future success.

Who Should Read This Guidebook?

This guidebook is of value to data migration project leaders but will also benefit:

- Project sponsors
- Data quality analysts
- Software developers
- Business analysts

In fact, anyone who needs to interact with the project will find the guidebook useful because it provides a thorough account of the activities and responsibilities required throughout.

Why Should You Read This Guidebook?

The practical techniques and advice found within the guidebook are based on many years of experience within the sector. As a result, this guidebook will help the reader effectively plan and react to the many obstacles and challenges a typical data migration project can present.

In short, this guidebook will help you create a more successful data migration project.

Part 1 The Route to Best Practice

Identifying and Resolving Misconceptions

As a project leader, one of the key problems you will face is convincing the organization to take a best-practice approach to data migration.

Too many organizations underestimate the challenges that data migration initiatives present. As a result, far too many data migration projects either fail or suffer major delays.

The cause of this failure is largely due to misconceptions about how a data migration should be delivered. An effective project leader must not only deliver the project successfully but also educate the organization so that these misconceptions do not impact the project implementation.

The good news is that this is certainly not a solo task. You must rely on the support of the data migration specialists within your team to help convince sponsors and stakeholders of the need for a best-practice approach.

By understanding the following misconceptions you will be in a much better position to identify them on your project and work proactively to prevent them negatively impacting your objectives.

Misconception: “We can fix the data in the new target system”

This is often heard when the migration deadline is looming (or has been and gone). You must explain to the business and technical teams that deferring data defects by

The end users of the new target system will be frustrated with the functionality problems that dirty data will invariably cause. Migration testers will find it harder to test and resolve issues at runtime as many data defects will result in processing failures.

The key here is to understand the defects in fine detail. Some data can be moved, irrespective of its quality but your team needs to carry out a full risk profile to assess what the possible impact may be. In the majority of cases it is far more cost-effective and time-efficient to resolve data issues before the data is migrated to the target.

Misconception: “We can just make use of our existing technology to migrate the data”

The majority of data migration projects nowadays still rely on hand-coded scripts to create the necessary migration logic. This practice is outdated, costly, extremely difficult to administer and should be avoided wherever possible.

Modern data migration software is far more suited for delivering particularly large data migration projects. Even on relatively small projects there are clear advantages of using advanced technology, so ensure your organization takes an open and pragmatic view of adopting new technology over outdated methods.

Misconception: “We don’t need to create a separate project for the data migration element of the program”

Data migration typically sits within a much larger implementation such as a merger and acquisition, a new system implementation or a system consolidation.

Many people over-simplify the task of data migration and feel that affording it the status of a separate project will lead to unnecessary expenditure and resource allocation. However, data migration requires a distinctive methodology and approach. It is therefore essential to manage it as a separate project, albeit closely tied to the main program.

Misconception: “Data migration is largely a technology-driven initiative”

It is easy to assume that with the word ‘data’ in the title the project will be concerned mostly with moving bits and bytes between two locations. In actual fact, the typical data migration project has a far greater demand on business engagement than most realize.

A best-practice data migration depends on business engagement at every phase of the project. As project leader it is your role to educate the business of its responsibilities at the earliest opportunity.

Misconception: “We need perfect quality data for our data migration”

Just as it is a mistake to believe that we can simply fix all our data defects in the target environment, it is also a mistake to assume we need perfect data in the first place.

All projects are constrained by time, money and resources, so it is rare to have perfect quality data for your project. While you don’t always need perfection, you do still need to plan for any substandard data quality. An accurate data quality assessment should be carried out so you can develop an effective plan. As project leader it is your role to work with the data quality assessment team to discover the impacts and adapt accordingly.

Misconception: “We don’t have any real data quality problems in our legacy systems, so there is no need for a big data quality effort”

This is a common problem that can cause havoc right at the end of the project when the data is being transformed and loaded into the target system.

Many sponsors will be unwilling to invest in data quality resources because they are reliably informed that the present data quality levels are acceptable.

The problem is that although this may be true in the old system, when we move this data to the new environment, the data may not be ‘fit for purpose’ based on the new demands expected of it.

Consider an inventory system that has been in existence for 7-10 years. If we migrated the inventory data to a new system that was designed to perform exciting new business services such as automation, advanced product search, enterprise

resource planning (ERP) integration and customer provisioning via the internet, we would almost certainly find that our data was being subject to a 'change in use'.

Another problem often encountered is the need to link disparate legacy systems together before migrating them. Each system may have excellent data taken as an independent system, but when we bring them together, major issues can be found.

The pressure of these demands on legacy information needs to be assessed in order to understand the impact and correct course of action. This is why every data migration requires a data quality strategy and the appropriate resources irrespective of data quality health in the old legacy systems.

Misconception: "We don't need a data migration methodology"

Most projects fail because they adopt an inappropriate or non-existent data migration methodology.

The responsibility for getting the methodology right lies squarely with the project leader. You must ensure that data migration specialists have an effective methodology at their disposal for this type of migration. If you have doubts, then seek out an expert to review your current approach and provide a second opinion.

Adopting the wrong approach may add months to the project duration so investing time and effort in creating the right methodology for your specific needs is absolutely vital.

Misconception: "We can't start building the migration code until the target application is well defined"

The reality is that on most projects the target system is poorly defined or incomplete. However, a great deal of work can still be completed. By defining the major business subject areas that need to be migrated, the migration team can still move ahead and perform a wealth of testing, data quality analysis and coding even though the target interface and models are not ready.

The key is good communication. If the target team continuously modifies the target environment it can result in a huge amount of coding rework by migration developers. As project leader, you must ensure that both teams are collaborating effectively. Try and integrate the target development team with your own development team, if possible, and ensure that any design changes are relayed to the relevant individuals in real-time.

Setting Expectations

Now that you understand some of the many misconceptions that can plague a data migration project, you are hopefully in a position to commence educating the organization as to how your project will be structured and what benefits a best-practice approach will offer.

The key thing to remember as a project leader is that you cannot plan the entire project in fine detail, despite the pressures that others may place on you. Data migration

is a lesson in discovery and adaptation. At each phase you must learn and change your plans by taking an agile approach.

Project leaders who map out a complete plan using a fixed waterfall type approach (often to placate their seniors or follow a company standard) can set themselves up for failure. Project sponsors may see any inevitable deviations and delays as a sign of poor leadership.

However, if you anticipate the risks documented in this handbook and put in place the necessary mechanisms to trap them, re-evaluate and move forward, then there is no reason why your project will not be a success.

You must set the correct expectations before the project starts. Creating a preparation phase before the main project launches can help you get started on the right foot.

The next section of this handbook will explain how to create a preparation phase that will help the sponsors visualize the type of challenges ahead as well as see the benefits of adopting a well-structured methodology that utilizes the appropriate skills and technology.

Part 2 Preparation

Overview

The data migration preparation phase is not compulsory. Many projects will simply wrap up the following activities into the initiation phase. However, when the migration project starts in earnest, you may be grateful that you adopted an earlier preparation phase to afford you the necessary time and resources to plan for the complex and challenging activities ahead.

As mentioned earlier, a preparation phase also helps the project leader convince the project sponsors of the need for a well-structured methodology and will almost certainly help to change many of the misconceptions people have surrounding data migration.

This part of the guidebook is designed to give you practical advice on how to prepare for a data migration so that when the project starts you build it on a firm foundation.

Reviewing the Business Case

One of the first activities for the project leader should be to review the original business case for the data migration. The answers to the following questions can help:

- What were the original business or technical drivers for the migration?
- Are these still valid?
- Are there new stakeholders or sponsors in charge of the migration?
- What is their yardstick for success?

Many data migrations fail to meet expectations, not because the migration was delivered poorly, but because the expectations were wildly optimistic.

A formal business case can be likened to an iceberg. There are the immediate benefits that are apparent on the surface, but also a vast number of hidden agendas lurking under the surface. There are great deals of politics with any data migration. As project leader, you are the prime target for resolving the many challenges that emerge from the business and technical communities.

Keep these key points in mind when reviewing the original business case:

- **Cast-iron project timelines and budgetary estimates:** Are these estimates viable without a feasibility study?
- **Selection of technology based on personal preferences or corporate standards:** Will these be suitable for the technical and business requirements of your migration?
- **Over-optimistic deliverables:** A business case is often the very first formal documentation the business communities will receive. It is important you understand what data and functionality may have been promised to the business in advance of any structured project discovery.
- **Assumptions:** The golden rule is that every assumption must be validated. Data migration misconceptions often manifest themselves as assumptions in the business case so be sure you manage them effectively using your risk strategy.

Identifying the Objectives

Once you have reviewed the business case it is worthwhile to outline the objectives of your project. As project leader you must set the goals and ensure that everyone is fully aware of what they are working towards. You may need to adapt your plan as you move into the later discovery phase, but it does help to get everyone on the same page before the project formally commences.

When outlining your objectives it is also important to emphasize that a further planning stage will almost certainly be required. The business will often demand a cast-iron, 'paint-by-numbers' style project plan that indicates exactly when the project will complete.

Remember that you are a project leader, not a fortune teller.

During the preparation phase you may have scant knowledge of the challenges that lie ahead. This information will not be available until you have completed the discovery and design phases. It is essential to educate the sponsors of this fact so they do not become frustrated later in the project if your plans, deliverables and timelines need to change.

Creating a Stakeholder Management Plan

Prior to the project, the project leader may not be aware of the various sponsors and stakeholders that are involved in a migration. This is certainly the case if you arrive from outside the organization. One of the key preparation activities is to therefore map out the various stakeholders and the roles you wish them to play in the project.

Some will require no more than simple reports at regular intervals but others, particularly those responsible for target or legacy data quality, may be actively integrated into the project team.

Importantly, the sooner you map out your stakeholders and begin constructive dialogue, the less change management issues you will face further into the project.

Developing a Risk Management Plan

There are many risks involved in leading a data migration project. It pays to take time during the migration preparation phase to understand and plan for these risks. It is often too late to take mitigating action during the live project, so make your preparations beforehand.

Here are some of the common types of project risks with some practical advice to help project leaders prepare in advance:

Lack of available expertise

Risk: Data migration demands specialist resource that is often lacking in-house or from external providers.

For instance, you will certainly need highly experienced business domain experts to interpret the many data quality anomalies you can expect to find securely entrenched in the legacy business.

Advice: Establish firm agreements for the staff required to deliver your project. The business is often guilty of retaining its best resources to ensure the smooth running of legacy services. You need to demonstrate the importance of both your initiative and the finished application. Escalate if required by using your stakeholder management plan.

If you are relying on external providers of resource then interview prospective staff well in advance to be confident of skill level and obtain binding agreements of their availability. It is not uncommon for some service providers to field their best resources during the pilot or tender phases only to substitute more junior staff or untested contract hire when the project commences.

Poorly executed scoping and budgetary analysis

Risk: The business case and financial projections for your migration may have been hastily put together. You may have inherited a project that is simply not feasible based on the earlier financial forecasts.

Advice: One of your first activities (if not already undertaken) is to create a short feasibility study. The preparation phase is an ideal time to undertake a feasibility study because the project investment will be at its lowest and the findings will directly benefit the direction of the project before too many assumptions are made. The feasibility study should consist of a prototype migration that assesses the cost and complexity of transferring a sample of data from source to target.

Lack of data quality management strategy and appropriate tools

Risk: You will almost certainly face data quality issues during your project. Because the extent and complexity of these may not be fully understood, a feasibility study is vital in this regard.

There may be misguided notions from the sponsor team that “our data is good enough” or “we don’t have the time for a big data cleanse.” This can often cause data quality to be firmly off the agenda before the project has even begun.

Advice: This is one of the most common, yet avoidable, causes of data migration failure. Put in place a data quality framework right from the outset. Explain the cost of project delay as a result of data quality and ensure your business case has a compelling case for investment. It is advisable to utilize data quality tools throughout the project. They will justify their investment and will reduce the overall duration of the project, helping solve the many data quality issues that will arise.

Lack of documentation and detailed knowledge of legacy and target environments

Risk: Most organizations have poor-quality or non-existent documentation of how their legacy systems and data were designed and are currently being used. Without this knowledge, mistakes in the extract, transform and loading (ETL) logic can cause data defects to flow into the target environment. Conversely, the target environments may not be documented because they are still in the early phases of design.

Advice: Finding the right business and technical domain expertise is critical here. Map out your source and target system landscape before the project commences as you identify relevant experts and documentation for the legacy and target environments. If no documentation exists, start to record the risks and ensure that you have a mitigation strategy in place before you start the project. It is better to understand the roadblocks before you start the project than to launch and realize you can’t move forward. In addition, ensure your data quality tools possess data discovery functionality such as data profiling and relationship discovery. This will perform a vital role when attempting to discover poorly documented system and data information.

Data migration methodology is insufficient or ignored completely

Risk: The method or framework required to coordinate all the activities and phases of a data migration is often poorly designed or incomplete leading to a lack of leadership and coordination, and ultimately project delay or failure.

Advice: The success of the data migration project is directly dependent on the quality of the data migration methodology. The importance of a mature, validated methodology cannot be overstated. During the preparation phase, ensure the methodology is rigorously vetted, ideally by a data migration expert.

Lack of collaboration between cross-party project teams

Risk: The different teams involved in a data migration may become separate islands of resource, all working to their own plan instead of a coordinated structure (often as a result of a poorly implemented data migration methodology).

Advice: This is a very common problem so plan early. It is important to adopt some practical solutions such as an integrated project repository and communication platform to coordinate the thousands of project materials, messages, issues and deliverables that are created throughout the project. Create a strong communication policy and a clear task framework as part of your methodology. Setting clear direction for deliverables and dependencies also helps to improve collaboration.

Poor choice of data migration technology

Risk: Many projects simply use whatever tools they possess 'in-house'; this often involves hand-coded scripts or inappropriate migration products.

Advice: It really pays to adopt the right technology for your particular type of migration. Using in-house scripts can cause major performance, auditing, configuration management and coordination headaches for the project leader. Using a well-proven data migration product that benefits from such features as fully integrated data quality functions, data integration technology, scheduling, real-time transformations and a wide range of interfaces can considerably reduce the cost and complexity of a migration project. This alone is typically enough to justify the investment.

Project delivery approach is inflexible

Risk: Many project managers adopt a 'waterfall' approach to data migration. While the analyze, build, test and deploy terminology is perfectly acceptable, running the entire project as a sequence of linear events is not always ideal. Doing so can create a rigid, inflexible project that fails to cope with the inevitable changes and problems that invariably arise.

Advice: An agile, iterative project delivery approach is often more suitable for data migration projects. An iterative cycle of learning and delivery makes it far easier for the project leader to plan the project in well defined segments as opposed to attempting to plan the entire project from start to finish, something that is frequently impossible given the unknown factors at the outset.

Go-live strategy is inappropriate for the needs of the business

Risk: Many projects adopt a big bang approach to moving their data whereby all the data is moved in one event, such as over a long weekend. This is increasingly seen as high risk technically and far less desirable in business terms.

Advice: The incremental data migration approach is often preferable to a one-off big bang. The reason is that incremental migrations move smaller segments of data, typically earlier and with less risk of failure. If there are issues then the entire business is not impacted and a fallback strategy is often easier to implement. An incremental migration also works particularly well with the agile project delivery approach discussed earlier.

Target application is constantly changing

Risk: It is not uncommon for a migration project to commence when the target environment is still evolving. This poses major challenges for the migration team who are effectively building against a 'moving target'.

Advice: Communication is vital. As project leader it is your role to ensure the target design and development team are complying with your communication strategy. They should be updating your team with all changes and updates as they happen so you don't incur scrap and rework costs because you are forced to redesign transformation and load components in your migration logic. Although the target model may change, you can still develop a large amount of migration functionality by agreeing to a 'common model' between the two environments as the conceptual and logical models should be fairly well defined. Your senior data migration architect or specialist should be able to explain this concept.

Receiving Agreement for the Project Security Policy

Data security has to be taken extremely seriously. Data migration projects can involve disparate teams often sourced from external integrators and even offshore partners. Data migration project team members may need to have close contact with sensitive data in order to verify data quality and discover business rules. If you have failed to create an adequate security framework prior to the formal launch it is not uncommon for the company security body to suspend the project.

The answer is to obviously specify the security measures, tools and training that will be enforced on your project and obtain formal agreement that these meet any corporate standards.

The company security team may audit your processes so it is essential to plan how regular internal audits will be carried out to ensure compliance. Now is the time to plan your security procedures. Waiting until the project is fully operational can cause major delays.

Defining the Data Migration Methodology

As already highlighted in the risk management section, without a robust data migration methodology in place, your goals of a coordinated and smooth running migration may be unfeasible.

It is advisable to define your methodology in advance so that the project is built on the right foundations.

Quite often the systems integrator or tool vendor will adopt their own migration methodology. As project leader, you will need to vet this rigorously and ensure it meets your needs. Ask for outside assistance if required and don't assume that a methodology is included in your agreement.

It is not uncommon for data warehousing, data integration, software development or even master data management (MDM) methodologies to be used as a makeshift data migration methodology. Pay close attention and get your service provider or data migration lead specialist to walk through every element of the methodology to ensure it will meet the specific needs of your project.

Planning the Staff Requirements

Allocating the appropriate blend of experience and skills is clearly paramount for a successful migration project. It really pays to commence planning your team in advance of the main project. You will invariably need to request resources from the business or technical teams that may be involved with other projects, so the earlier you make your requests, the more chance you have of acquiring the necessary resources.

Staff members are typically required to resource the following groups:

The Business Requirements Team

This team typically consists of members of the legacy and target business team. These are often business super-users who understand both environments intimately. They may have been with the company for many years. As a result they may be extremely difficult to extricate from their current roles to work on the project.

In this case, the role of the project leader is one of salesperson. You will need to convince senior managers why these resources are critical to your project. It helps to outline what type of work they need to perform and what utilization you will need from them. Again, preparation is vital. Ensure you have all the necessary information before you start to make demands for these critical staff members.

The Data Management Team

The data management team performs a hub role, interlinking the other teams. People in this team need to have good data analysis skills, be great communicators and stay extremely well organized. They will typically take input from the business requirements team and pass on requests to, the data quality team or testing team, for example. They are instrumental for documenting the various data mappings, rules and models that need to be known in order for the other teams to complete their tasks.

As this team interacts so heavily with all the other teams, and will be in the firing line right from the start, it is critical to hire the right skills and experience before the project commences.

The Data Migration Development Team

The developers typically come into the project mid-way through and convert the mappings provided by the data management team into some form of data migration software. This can pose a challenge for the project leader because many sponsors do not select their migration tool before the project has commenced. This is a risk as it can often be a challenge to find skilled developers on short notice. Ensure that you openly discuss this issue with all parties.

The developers work closely with the testing and data quality teams to ensure the logic they create will not only complete successfully at run time, but also protect the integrity and accuracy of any migrated data.

The Data Quality Management Team

Depending on the scale of the project these members may reside in the data management team but on larger projects it is not uncommon to have a dedicated team of data quality specialists that the project can draw on.

This team will engage with all the other teams, taking instructions and delivering insight in equal measure. The importance of this team cannot be overstated; migration projects frequently fail because there is no data quality team or data quality management process in place.

As a project leader, it is your role to help source the appropriate staff from the business community who are essential to the effectiveness of the data quality team. You may not need full-time commitment from these workers, but you will certainly need business users to help validate and define the data quality rules for cleansing data, identify defective data and generally supporting the other data quality practitioners within the team.

Sourcing skilled data quality specialists can also be a challenge. Many organizations have yet to implement a data quality capability so this often requires the project leader to find and recruit skills from outside the company. This is a non-trivial task so be sure to execute this activity well before the project formally commences.

The Data Migration Testing Team

This team plays a vital role in ensuring that the migration will deliver against your expectations. As a business leader, your primary concern is service continuity and any customer service level agreements.

Testing is often seen as a technical chore, so be confident that your advisors have fully vetted the quality of migration testing. The aim here is to ensure the final migration will be fit for the purpose of your business, not just the technicalities of migrating data.

The Data Migration Validation Team

The role of the migration validation team is to verify that the data migration has successfully transferred your business data to the target environment prior to the decommissioning of any legacy systems or applications.

This is an absolutely critical phase and you should allocate your most highly skilled domain experts to ensure that the migration has met both technical and business requirements.

The Data Migration Operations Team

Depending on the scale of the project, this optional team is responsible for the coordination of the migration execution and any resulting issues this can create. The physical act of migrating data can often take months – even years – so it is quite common to have a dedicated team that controls this process.

Carrying Out a Pre-Migration Assessment

It is important to perform some form of data migration impact or readiness assessment before the main migration project commences. This typically takes the form of a four- to 12-week data migration simulation where an indicative sample of data is assessed for its ability to migrate between source and target.

Although you won't have the full complement of resources, target systems or migration technology – in fact, some improvisation will almost certainly be required carrying out an assessment prior to the main migration has many advantages:

- Helps to clarify whether the migration is commercially or technically feasible: Given technical and commercial constraints it is not always possible to complete a migration in the allotted time or with the resources available. A pre-migration assessment will help you to identify and measure some of these risks more accurately.
- Starts to shape the go-live strategy: As we discussed earlier, the type of go-live execution strategy (e.g. big bang, incremental, etc.) can often be discovered during the pre-migration assessment. For a project leader, this is a huge benefit as it means you can start to communicate and plan for this as the project commences.
- Tests the data migration methodology: The pre-migration assessment is a cut-down version of your main migration methodology; any gaps or shortcomings in the methodology can often be uncovered during the assessment.
- Gathers vital data quality metrics: Perhaps the most important activity in the pre-migration assessment is to create a baseline of data quality across the sample dataset. This is critical as it will help you demonstrate the value of adopting a thorough data quality framework with the necessary tools and experience. It will also help you identify what type of issues to expect when the project enters into the main build and deployment phase.
- Validates the project forecasts: The assessment should help you to validate record counts and provide a rough measure of effort for a sample set of data. This can often be used to extrapolate figures for project duration and cost so you can gauge whether the earlier estimates are likely to be valid based on the complexity and challenges you faced in the assessment.
- Gets the team in advance of the main project: The pre-migration assessment is a perfect opportunity to get the team working closely together. It gives you, as project leader, an opportunity to verify who the natural leaders are and identify the most able members in the team. It is often far more difficult to adjust the team dynamics once the project is in full swing. Kicking off with a smaller, less pressured project also helps the team to build morale and prepare for the main project ahead.

Creating the Data Quality Framework

It is advisable to set up your data quality approach prior to the formal project launch. The reasons for this are varied:

- You will need a data quality capability during the pre-migration assessment so it is advisable to launch this capability early.
- Getting the stakeholders, technical and business resources to collaborate on data quality is the key to migration success; starting early can iron out any technical, commercial or political headaches early in the project lifecycle.
- It takes time to implement a data quality framework. The earlier you start the more chance you have of ensuring adoption.
- There are a multitude of templates, processes and procedures to create. It is far better to design these before the main project commences.

The data quality framework is a combination of people, processes and technology that will encompass the following activities during the project. These activities ideally need to be defined and agreed during the preparation phase.

- **Discovery:** Finding the data quality rules and data quality metrics within your legacy data sources is critical during the data migration. Rules will help you identify the relationships that bind the legacy data together – often across disparate systems. The data quality metrics will help you understand which data can be migrated ‘as-is’ and which data will require cleansing or transformation.
- **Assessment:** Where discovery will help you discover the rules, the assessment process will help you quantify the data quality levels found in these rules. This will almost certainly require some form of data quality technology to help you manage the complexity and quantity of rules that are generated on a typical data migration project.
- **Cleanse/Transformation:** It is highly likely you will need to modify the legacy data either in the source system or in some form of staging area in order to meet the demands of the target environment. Some people will refer to this as cleansing but most of the operations are simple transformations. Quite often the migration software will be unable to carry out these types of data quality functions without the use of a specialist tool.
- **Consolidation:** Data migration from multiple legacy systems typically requires the consolidation or aggregation of data records. For example, if you have inventory scattered across several parts databases then your team will need to create matching rules to create a consolidated view of that data.
- **Monitoring:** Data migration projects can take many months so it is advisable to monitor the quality of data in the source environment throughout this period. This helps you prevent failures during the migration execution phase.

Part 3 Executing the Project

With the preparation phase now complete, the migration project ramps up, and with it, your responsibilities as project leader. If you follow a structured methodology, utilizing the right tools and skills, you should lead your team to a successful outcome.

Important note regarding iteration: Data migration is ideally an iterative process. As project leader, it may be tempting to create a simple plan that links each phase in a linear, waterfall fashion. The reality is you may have to repeat several of the project phases many times before your migration is complete.

For example, you may find data quality ‘hotspots’ that are worth avoiding until the business has developed a remedial strategy. Rather than delay the project you may decide to migrate a segment of data that is fit for migration and postpone the migration of the lesser quality data until a later iteration.

Most projects have a set of phases that broadly follow this sequence:

- Prepare
- Discover
- Design
- Deploy
- Execute
- Validate
- Decommission

The preparation phase has been covered in detail in the previous section, so we will now explore the remaining phases. Before we do this, it is important to point out that these phases are merely suggestions. Every project is different and will have a unique set of project management challenges. Use these phases to help build your own project strategy, but always focus on what is necessary to deliver your particular type of data migration.

Discover

The discovery phase is the project leader's 'get out of jail card'.

If executed correctly it will help you create a far more accurate plan for the latter phases of the project. It is also invaluable in helping demonstrate to the business sponsors why you need a data quality management team and all the necessary technology.

Without the discovery phase you are 'flying blind'. You may not spot serious obstacles until well into the execution phase when it is often far too late to request major planning changes.

Ironically, there is often a considerable amount of reluctance to perform discovery and analysis activities. Many organizations will be struggling to keep the wider program on track at this point. It is common for project leaders to omit discovery in a bid to move quickly onto the execution phases.

This is where many projects ultimately fail. They struggle to understand the needs of the business and the complexities of the data. Data quality issues start to arise during the migration load phase and it may only be during the validation phase that the true extent of the problem is uncovered.

One of the first tasks as project leader is to therefore educate the organization as to why the data discovery element of the project is so critical. The best approach is to 'show-and-tell'. The more issues you discover and present to the business, the easier it becomes to demonstrate the value of your approach. For many data managers and sponsors it will be the first time their systems have been assessed, so there will probably be the occasional political minefield to traverse.

Here are some of the typical tasks that are common to the discovery phase:

Identifying Source Datasets

One of the earliest discovery phases is to identify the legacy or source data needed to perform the migration. There is often an assumption that this data will reside in legacy databases. However, this is not necessarily the case as business-critical information can also be found in local paper records, disparate spreadsheets and 'home-grown' data stores that the business has unwittingly created over the years.

The project leader must play an important role in ensuring that the business is fully engaged in this process to see that all possible data is identified, documented and, where applicable, accessed on the project.

Identifying Target Specifications

At this stage, you may lack a formal target data schema, but this does not mean significant progress cannot be made. By agreeing to high-level data specifications with the target application developers, you can still start to build some high-level mappings between source and target objects. You may lack the field-level specifications at this stage but quite often you will be able to identify major gaps, even at this stage.

Remember that the discovery and design phases are iterative; you will perform multiple cycles of discovery as the target environment begins to take shape.

Creating Accurate and Complete Project Models

The discovery phase is where many of the many data models that will be used throughout the project will be created. In an ideal world there would be complete and accurate models available. However, in the majority of cases, these will almost certainly need to be 're-engineered' by examining the data and its relationships.

Some of the typical models your migration will depend on are listed below:

- **Conceptual Model:** An extremely important model from the perspective as the conceptual model forms the high-level model of the entire data landscape. This model will contain the main subject areas that will be in scope for the project. As a communication and collaboration mechanism this will form one of the project's most useful assets.
- **Logical and Physical Models:** Most organizations pay lip service to logical and physical models and they will almost certainly be out of date. They hold the details of how the business perceives the data and any relationships plus information as to how the data is physically stored.
- **Common Models:** A common model attempts to link the source and target environment. It is useful for easily identifying gaps between the different data structures and relationships which can cause serious issues later in the project.
- **Business Function Model:** The business function model creates a registry of all the business functions that need to be implemented in the target environment. This can be extremely useful for identifying gaps in the legacy data that mean the new functions cannot be supported.

Data Profiling and Discovery

It is imperative that your data migration utilizes data profiling and discovery tools. Ideally, these will have been trialed and implemented during the preparation phase. You will also need them during the discovery phase of the migration. Data profiling tools will help you to automate the discovery of data quality rules that will need ongoing assessment throughout the project leading up to migration execution. Data discovery tools are typically used to re-engineer physical and logical models based on the actual values found within the physical databases. As such they can save the data management team considerable time and effort by dramatically improving the performance of an individual team member.

Detailed Data Quality Assessment

Data profiling and data discovery activities will help your team collate a large number of data quality rules. These rules are critical as they dictate what levels of quality are required in order to migrate the data successfully to the target system.

The assessment function is vital to the project leader as it will give you accurate information on the data quality levels within the data earmarked for migration. If you find major issues at this point you will have the necessary information to consult with the project sponsors to determine the appropriate course of action. Without this kind of activity, you are opening the project up to major risks in the latter phases.

Assigning Data Owners

If you followed the earlier suggestion of creating a stakeholder management plan, you may already know who your data owners are. In any event, there are always new data sources that come to light and, as project leader, you may need to explain to the data owners what role they and their team must play to create a successful data migration.

By assigning a data owner to each legacy data store you can also start to ask the question, "What validations and audits need to be completed before we can decommission each data store?" The data store owner should help you resolve all the business and technical dependencies involved in terminating an existing source of information after your migration has completed.

Refining Project Plan

Remember, there is no such thing as a complete data migration project plan, particularly at this stage of the project. The goal of the discovery phase is to help the project leader gather some hard facts so they can start to create a more realistic plan. The final project plan only really starts to take shape after the design phase is complete and you can be confident of your execution strategy.

Capturing/Validating Business Requirements

There may be some well-defined business requirements already in existence prior to this phase but you will need a more accurate view of the data quality levels and technical impacts of those business requirements. This will help you to define which of these requirements will be in scope and which ones may need to be deferred until a later iteration.

It is critical that the business requirements team works closely with the other teams in the project, and this is something the project leader must be wary of. Rather than create a set of requirements in isolation, there must be almost daily interaction between the teams, so that as each requirement is gathered the other teams (e.g. data quality/data management) can then assess the impact.

The use of an online project portal is an excellent tool for gathering and communicating business requirements allowing each team member simple access to the latest changes in business needs.

Creating the Project Data Dictionary

The data dictionary is an excellent means of sharing timely information discovered during this phase. One of the most important characteristics of a successful data migration is collaboration and it is the role of the project leader to ensure the appropriate tools are in place to facilitate this.

An increasingly popular approach is to use a wiki or project portal to store all the information that the various project teams depend upon. By definition, a project data

dictionary is simply a searchable resource where each team can add, update and notify any data resource information.

For example, the data discovery phase may define a relationship between two business entities. This information would be stored within the data dictionary, and the data quality management team would then assess the quality of this relationship throughout the project. The development team could then use the various profiling statistics gathered during the discovery phase to help design and code their migration logic.

The project data dictionary therefore becomes a critical asset during the project as it helps each team communicate more effectively.

Design

The discovery phase will have created a wealth of information detailing how the source and target environments are constructed. Now the project must move into design and prototype mode.

At this stage, the project may still not have purchased any tools to migrate the data, but this is not necessarily a problem. However, it is paramount that the project uses an appropriate data quality tool.

As the project moves into the design phase, we can use a data quality tool should be used to perform 'data migration simulation'. This will help to identify potential bottlenecks well before the build and execute phases take place. For the project leader, this is an invaluable technique as it significantly de-risks the project and affords you plenty of time to re-plan and create alternative solutions.

Creating the Source-to-Target Mapping Rules

The discovery phase will have created sufficient knowledge to help the data management team map out the source-to-target mappings. In the earlier iterations, this may simply be at an object-to-object level. However, as you cycle through multiple iterations, you can start to populate the field-level mappings.

Using a data quality tool, we can create a data quality assessment routine for each mapping rule. This will enable the assessment of the quality of data in the source environment well before moving into the build phase where the costs of changing the migration logic can be significantly higher.

Source-to-target mapping rules can be created in a variety of formats, but it is always preferable to document them before construction in the migration software. This enables all the other teams to observe any changes and perform their related functions. It also means that progress can be made without any purchase of a migration tool.

Building the Data Quality Cleansing/Transformation Rules

Throughout the project, data quality issues will be discovered that will require resolution. Some issues may be cosmetic and can be ignored, but most will require some form of remedial action.

There are several options for dealing with these issues:

- Cleanse in the source environment
- Cleanse in a staging area prior to migration
- Cleanse during the actual migration
- Cleanse in the target environment post-migration

As project leader, the data quality management process must be monitored on a daily basis. This is one of the key areas where most projects fail. Your team members can quickly get swamped with resolving data quality issues. Conversely, they can ignore issues which may come back to haunt you during migration execution and testing.

You will need a data quality management tool to perform this activity adequately. The advantages of using a tool far outweigh its relatively low cost. Modern data quality tools will allow your team to resolve issues at the source, in the staging area, during the migration or after the migration in the target area.

One of the key benefits of using a data quality tool on the project is that the data quality rules you create on the project can be used as an ongoing monitoring tool in the target environment. This is an extremely cost-effective means of launching a data quality initiative that will maintain data quality in your new target application.

Creating the Staging Area and Integration Model

A staging area is common to most data migration projects. It is a temporary storage and processing platform where legacy data can be transformed into a closer representation of the target environment.

Staging areas have many benefits:

- Makes it easier to perform an audit trail as changes to data can be tracked.
- Data quality improvements can be made without impacting the source systems and any other dependent data feeds.
- Data analysis in the staging area can take place without impacting the performance of the live environment.
- Migration simulations and test-runs can be easily performed without impacting either the source or target environments.

The integration model is a combination of staging area, source and target mapping rules and migration logic that integrates the environments into one cohesive design that will support the physical sequencing and migration of the data.

Designing the Fallback Process

In an ideal world, the migration will execute smoothly and the business users will cutover seamlessly to the target environment. As project leader, you have to plan for all eventualities – including worst-case scenarios – and the creation of a fallback process is strongly advised.

If the fallback process is ignored until the later build phases, you may find that it is actually impossible for the business to fall back to the source environment in the event of a migration failure.

Designing the Synchronization Process

Depending on the migration execution strategy, the migration architecture may require some form of synchronization to maintain data integrity between the source and target environments.

For example, if you intend to run both the source and target environments in parallel post-migration you may need to pass changes from the target environment back to the source environment and vice-versa.

This is a non-trivial task and careful consideration will need to be made as to whether the integration model and relationships across the source and target environments will support synchronization. In many cases a source-to-target mappings will be uni-directional and will not support backwards synchronization. As project leader, you need to discuss these issues early in the project lifecycle as delaying them can have a serious impact on the success of the project.

Designing the Validation Process

How will you know if the migration has successfully moved the data from the source to the target environment? It is not uncommon for a migration to be hailed as a complete success only for major segments of data to be found missing after the source system has been decommissioned.

The solution is to design a robust validation approach right from the outset. This will enable you to confirm that every data object, attribute and relationship has been migrated correctly.

There are various approaches which can be adopted, but it is preferable to perform an independent validation, often using a different team and software. Many projects simply use their existing migration tool and logic to validate that the data has migrated successfully. However, this can often hide serious data gaps. By using an independent team and tool you are more likely to find anomalies.

It is critical that the data mappings and data quality rules used to migrate the data are documented so that the independent team can implement their validation routines. As project leader, it is your role to ensure that a regular audit is carried out so that any documentation matches what was physically implemented by the various project teams.

Defining the Migration Strategy

In the design phase, you will have to make some major decisions regarding how the business will be transitioned during the migration. As we discussed earlier in the handbook, this may be big bang (move the business in one action) or incremental (a more gradual, iterative transition).

Each strategy will impact the migration design. For example, an incremental transition usually involves some of synchronization and this will require additional logic to be developed. Defining the migration strategy will require considerable involvement from both the technical and business communities.

On the technical side you will need to assess the optimal migration rate, i.e. how many objects can be physically migrated in a certain period.

Armed with this information, you can then consult with the business as to what period of downtime they can grant to allow the migration to take place. This 'window of opportunity' is critical. Many businesses now operate 24/7 so it is vital to have these discussions early in the project lifecycle. If you leave them to the build phase it may be far too late to re-design your migration architecture.

Confirming the Migration Toolset

During the migration design you will start to create a clearer picture of what type of migration technology is required. Many organizations will simply make use of whatever tools they already have available but it is wise to select a tool that can cope with your migration strategy.

- Will you need to run data quality transformations during the migration?
- Is synchronization important? Are there consolidations required in the legacy environment?
- Will you be running both environments in parallel post-migration?
- Do you need to provide an audit trail for compliance and regulatory approval?

There are many factors which will determine your choice of migration technology but it is advisable to use a custom product as opposed to hand-coded scripts. Selecting a product that can perform data discovery, profiling, integration, sequencing, migration, data quality improvement and validation all within an integrated working environment is also a clear advantage.

Presenting the Proposed Migration and Project Plan

At this stage you may have performed multiple discovery and design iterations and created a thorough design for your proposed migration. You will also have a good understanding of how the business will be transitioned during the physical migration. It is now time to present your migration plan to the business and technical communities.

For the program sponsors, this will come as a welcome relief. You will now be able to demonstrate clear progress and a visible route to the end of the project. It also gives each sponsor and stakeholder the opportunity to raise issues before you move onto the build phase, where change requests can become far more damaging to the project schedule.

If you can simulate the proposed migration at this stage it is a clear benefit. Using a staging area that mocks up the source and target environments you can demonstrate the physical migration of data and the clear benefits your structured approach has provided.

Deploy

The deployment phase is where the final design specifications take shape. As discussed earlier, there may be multiple deployments on the project. As a result, you may deploy different versions of logic to migrate specific segments of business data.

Implementing the Legacy Data Cleansing Rules

The deployment phase is the time to execute the data quality procedures to ensure your data is 'fit for migration'. In reality, data quality improvements may have taken

place right from project inception, but as the business is making changes (and mistakes) to the data every hour, it is important that data quality monitoring and improvement takes place right up to the point of execution.

Prioritization is important here and, as project leader, you may have to make decisions (with consultation) as to which segment of data and set of data quality rules will be in scope for each deployment.

Implementing the Live Migration Architecture

At this point you will build and implement the final migration software. Since you have been operating in a safe, offline environment, the deployment phase is when you want to start to integrate the migration architecture directly to the source and target interfaces so extreme care must be taken.

At this point, the source-to-target mappings, target interfaces, models and application design should be firmly in place. Any changes to the target at this stage can have major impacts so, as ever, good communication between the project team and target development teams is imperative.

Testing the Live Migration and Fallback Process

the testing and development teams will start to validate the migration logic and associated interfaces by extracting and loading data into source and target environments. Considerable testing and validation is also required to assess the integrity and quality of the migrated data.

This is extremely important as it will not only validate that the logic and functionality is correct but it will provide the end-to-end performance metrics that will give the project team accurate forecasts for how long the migration will take. The testing team must also assess whether the fallback process is viable.

Validating the Window of Opportunity for Migration and Fallback

On most migrations, the business can only afford a set period of downtime to permit the migration. The performance metrics gathered in the previous testing phase will help the project leader determine whether the migration execution will complete in the allotted window of opportunity.

One mistake that can be made is to ignore the time taken to execute the fallback process. Your window of opportunity must include sufficient time to revert any changes that are required on the legacy environment in the event of a migration failure.

Testing the Target Business Functionality

For the first time, the target applications will have legacy data to carry out functionality testing. The migration process may be a complete success, but it is not uncommon for design flaws in the target application design to be discovered which necessitate changes in the migration logic.

The key is to test the target environment using full extracts of legacy data. Test teams often make the mistake of using small samples of test data which pass basic functionality checks. Small samples are often insufficient to test the hundreds, often thousands, of business and data quality rules that exist in even the most modest of applications.

As project leader, you must push for full volumes of data to be migrated during the testing phases. Failure to do this prior to the final execution may result in a cry for the fallback process to be executed if serious issues are discovered and the migration has to be cancelled. Fallback can quickly become a nightmare scenario as the window of opportunity diminishes and the business needs to come back online to service their customers.

Testing the Business Transition Process

In addition to the physical migration of data, the project leader must also be confident that the business will be able to transition from the legacy to the target environment.

It will ultimately be the decision of the project leader to confirm that the migration is 'go for execution', so this is one area that will require your close involvement to ensure that the testing is an accurate representation of the live transition and that all teams are operating effectively.

Deploying the Data Integrity Validation Solution

Simply relying on the data migration software to detect data errors is not sufficient. It is strongly advisable to create a real-time independent validation solution that will be able to monitor data integrity between the source and target throughout the entire migration.

There are thousands of data and business rules in even the most modest of migrations, and the migration software will not necessarily report a failure if one of those rules fails.

An independent validation tool, such as a data quality or other data comparison mechanism, will ensure that the data has been migrated successfully and that it has been transferred to the target environment in accordance with the specified rules.

Execute

The execution phase of the migration project may run for several hours, days or even months. It is at this point where the business is directly impacted by the migration as business services effectively switch from the old legacy environment to the new target platform.

Data Migration Sequencing and Coordination

The data migration operations team will now work closely with the development, data quality and business representatives to coordinate the migration of data. Most migrations will require data to be migrated in a specific sequence to ensure the integrity of the data and any relationships. For example, customer accounts may need to be migrated before billing data. Inventory data may need to be migrated before service history, etc.

The operations team will closely monitor the migration of these data segments and halt or cancel the subsequent phases if issues are discovered. It is preferable to halt the entire migration and re-schedule as opposed to spending days and weeks trying to resolve data defects that may have occurred in the migration.

As project leader, you must closely monitor the quantity of data defects being created as this can cause serious resource issues to your entire team as they struggle to clear the backlog, particularly if executing over several weeks/months. It can also create tension with the business community as their service levels begin to drop as a result of the data anomalies. The fallback scenario must be avoided at all costs, so watch the error rate extremely closely.

Coordinating the Transition of Business Services

As the data is migrated, the operations team will have to collaborate with business user groups to switch over the provision of services from source to target applications. Excellent communication is vital here and, as we discussed earlier, this phase must be tested exhaustively to ensure each group is aware of their responsibilities.

There may be temporary measures in place to ensure the smooth transition of the workers. For example, all orders may be entered on the old legacy system as well as the new system.

Validating the Source-Target Data Integrity

Picture the scene: It has been a long weekend, the migration started at 6:00pm on Friday when the business granted your team a window of opportunity on the promise they would be back online with the new system 9:00am Monday morning.

Saturday and Sunday were frantic, some small issues were uncovered but, on the whole, the migration went smoothly. It is now 3:00am Monday morning and you are carrying out the final migration scripts. You have a five hour fallback window, so time is short.

The testing team start validation testing and they immediately observe anomalies. A number of inventory items have failed to migrate. There were no errors created by the migration software as the data was simply never extracted. As a result, several thousand customers have incomplete billing records. No error was recorded as technically there were no coding errors during the migration process.

The moral of this story is never to trust the migration software implicitly. No errors reported during the migration does not necessarily mean the data was successfully migrated. Sadly, it is not uncommon for project leaders to be informed of major data discrepancies even after a seemingly successful migration. At this point your options are limited. You cannot fallback because the window of opportunity has gone. Your only hope is that the issues can be resolved on the target system. The net effect is reduced confidence in the target system, some very disgruntled sponsors and a lot of data detective work to uncover exactly what has gone wrong.

It is at this stage you may wish you had invested in the real-time, independent data validation solution we covered in the deploy phase. This will give you, the project leader, extra confidence that the data has been migrated correctly and there will be no call for the dreaded fallback procedure.

One of the best solutions is to utilize your data quality tool to validate the transformation rules. This will allow you to validate that metrics such as record count, relationships, attribute populations and value distributions correspond between the source and target system.

Part 4 **Closedown**

As the migration draws to a close but there is still much to do. The goal of most data migration projects is to shut down legacy services gracefully while maintaining business continuity until the target environment becomes live.

As project leader, it is your role to manage this transition and ensure your project has completed all the necessary activities that will enable the business and technical communities to grant a final closedown authorization.

Execute Final Customer Acceptance Testing

Although there is testing throughout the deploy and execute phases, there will often be a round of customer testing where additional validations are performed, often by the business community themselves.

Your role will be to assign team members to support this process and work through any issues discovered.

Enact the Archiving Policy

It is common for the source system to be fully decommissioned which will result in full termination of the data. It is prudent to maintain some form of archive so that the business can access historical data in the event of a query or regulatory constraint. Many projects have identified serious data anomalies several weeks or months after the migration, so ensuring the legacy data is available is clearly advisable. This is another task that the project leader should add to their checklist and monitor its progress throughout the project.

Decommission the Legacy System

If the project sponsors are satisfied with the final customer acceptance testing and the archiving system is fully operational, the legacy system can now be fully decommissioned.

Consolidating silos of legacy systems is often a key strategic driver for the organization as it helps to reduce costs, increase performance and maintain profitability.

Transition Ongoing Data Quality Management Capability

During the data migration project, a thorough data quality management and data validation strategy has to be delivered. Now is the perfect time to continue managing the information on the target system as a high-value corporate asset. The expertise, documentation, technology and methodologies used on the data migration project should all enable the rapid implementation of an ongoing data quality management capability.

It is highly likely that new personnel would be responsible for any ongoing data quality management of the target environment. The final act of your tenure as data migration project leader should be to ensure that the knowledge and capabilities your team has developed over the course of the project are passed on.



THE
POWER
TO KNOW.

SAS Institute GmbH, In der Neckarhelle 162, 69118 Heidelberg, Germany
Phone +49 6221 415-123, Fax +49 6221 415-145 www.sas.de

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies. Copyright © 2012, SAS Institute Inc. All rights reserved.