Legacy System
Data Conversion and Migration

Key Management Group, Inc.

Fostering Partnerships

KMG

Proprietary Notice

All trademarks, services marks, trade names, logos, icons and other intellectual property rights (the "Intellectual Property") are proprietary to Key Management Group, Inc (KMG) or used by KMG as permitted through contractual arrangements with respective proprietors. The content of this testing process shall at all times continue to be the property of KMG/ proprietors and will be strictly protected by copyright and other applicable laws. Any usage, reproduction, retransmission, distribution, dissemination, sale, publication or circulation of the said content without the express prior written consent of KMG is strictly prohibited. Nothing in this document shall be construed as granting by implication or otherwise a license or a right to use such Intellectual Property, or indulge in reproduction or decompiling or sublicensing such Intellectual Property.

The users/clients expressly recognizes, admits and confirms that all the IPR rights in the development of this testing process is the sole property of KMG and the user /client shall not claim or in any way assist, aid, permit or abet others in claiming any right title or interest there in or thereto. If required by KMG the user/client shall cooperate in having KMG register the IPR in this documentation, if deemed necessary.

© Unless otherwise acknowledged, the entire IPR (Intellectual Property Rights) in this testing process and contents thereof absolutely vests in KMG. All rights reserved.

Table of Contents

1 Introduction

From time to time various Business organizations implement new Software Application System to **replace** the functionality currently delivered by one or more legacy systems. Complications arise when there is an attempt to take the information currently maintained by the legacy system and transform it to fit into the new system. More often, the data structure of the legacy systems is different from the new application being implemented, and that difference is not just limited to the table names, field names or attributes or sizes. The types of databases are different and diverse, or the entity relationships definitions in the new system are not compatible with the older legacy application. To the business organizations all the data being held in the legacy system remains critical for their business functions and decision making.

To bring the legacy system data to the new application some **Data Conversion** must take place, where an initiative, separate or concurrent with the implementation of the new application, is undertaken *to convert data from one structural form, used by the legacy application to the structural from required by the newer application* .

Often in a Data Conversion process, one would tend to think that any two similar systems that maintain the same sort of data, as they are doing very similar functions should map from one to another without much trouble. But that is not really the case as -

- In Legacy systems, historically, data integrity checks were not strictly enforced, leaving orphan data
- Theoretical design differences exist between hierarchical and relational systems.
- Legacy data may require some data cleansing.

Other Factors contribute to complexity of such projects are

- Need for a well-defined target data structure model.
- In-depth understanding of the functionality of the source data structure.
- Constant changes on the target model design have a knock-on effect on conversion process design.
- Source data quality, if poor, needs to be cleansed in order to be successfully migrated.
- Degree of complexity of the target model in relation to the source data model.
- The differences in task definitions between the source and target data structures.

Therefore it is important to have a sound, methodological approach by which organizations can undertake Data Conversion projects, which will help to confront unpleasant surprises on later stages and resolve those issues fast and effectively.

Key Management Group. Inc. (KMG) often helps various business organizations to implement the new Software Application, especially in the area of Property and Casualty (P&C) Insurance and Enterprise Resource Planning (ERP) package software. To help our business clients, KMG designed and developed a methodology and adopted that for such Data Conversion projects.

This document attempts to document that methodology, which it does by stating various processes to be performed at various stages of Conversion and Migration of Legacy System Data, and the Roles and Responsibilities of various stakeholders and participants in the projects.

Though over the years, this document was reviewed and refined and republished by KMG, based on the experience of many such projects, it is important to note that the methods and processes described within this document is generic in nature. Each project presents its own challenges and opportunities and KMG personnel engaged in the respective project design and develop, add or modify steps within the broad framework of this methodology, for the benefit of that specific project.

2 About KMG

KMG is a global software development, application maintenance and BPO operations company, which provides premium IT solutions worldwide using Microsoft, IBM & Java Technologies. In a wide-open sea of countless software development companies, KMG distinguishes itself as a company driven by excellence.
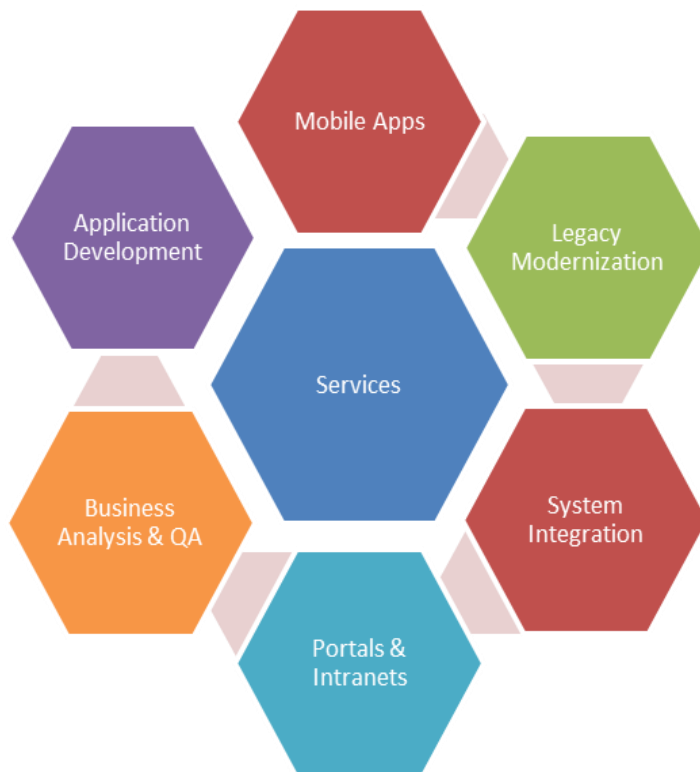
KMG was established in 1990 and is among Top 10 fastest growing Indian-owned companies in the US. It is also rated among top 50 software companies in India.

KMG has its headquarters in New York with multiple Offshore Development Centers in India. Our onsite-offshore model and industry expertise has enabled the company to enter into long-term, mutually beneficial strategic partnerships with many Fortune 500/1,000 companies. Unlike most other Indian-based software services firms, KMG maintains a large development team in the US. This team is used to interact with the client & provide a longer overlap to the users.

With a strength of approx. 60 professionals in the US supported by another 350 in India, KMG has resource expertise that covers Microsoft/.NET technologies, Java/J2EE, Mainframe, IBM iSeries (COBOL/RPG), Software Testing, BI, Open Source and other varied technologies.

KMG provides software maintenance, development and testing, and BPO solutions to large and medium sized insurance, banking, financial service, healthcare and government organizations throughout the world. Indeed, the company has aggressively captured a niche market in the property and casualty insurance sector in the USA.

KMG has a 600-seat state-of-art development center near Chandigarh (North West part of India). This center houses a large training center for providing training to in-house / external resources on various technologies (including legacy systems). The center also houses the teams for QA testing & BPO processes.

3 Overview

3.1 Conversion and Migration Process

Data Conversion project activities starts with Planning, leading to Analysis and Design, progressing to Conversion of Data, finishing in Migration - where converted data loaded in the target system database.

Following figure provides an overview of all the processes involved in legacy data conversion process
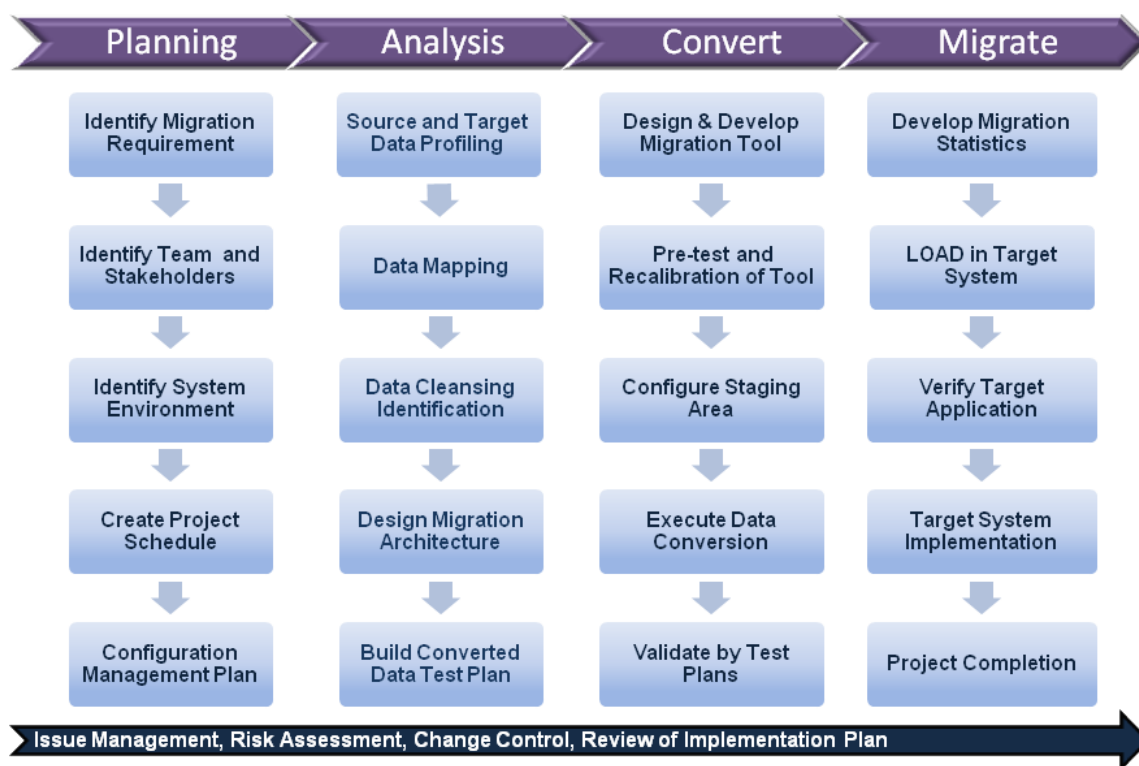

Figure 1

All the tasks to be conducted at various stages of a Legacy Data Conversion projects and the guiding principles for execution of those along with roles and responsibilities are described in the subsequent chapters.

3.2 Guiding Principles/Assumptions

This methodology is covering the automated process of Conversion and Migration of Legacy System Data, not the Manual Data conversion. Manual data loads will occur where data volumes are low and there is no time or cost benefit in automating the data loads.

The process of Automated Data Conversion could be through a pre-packaged Tool procured from one of the software vendors in the market or could be developed by KMG, based on the specific requirement of the Project. Benefits and costs of the both approach will be reviewed.

The Extraction and Conversion Process will be tested prior to the production conversion as per KMG Methodology, described later in this document. This is to forecast precise conversion timeframes and to validate/test programs/tools configuration.

This methodology is designed for Single tier and Single time Conversion and migration. Single-tier migration is a migration in which all source data is migrated onto one or more devices within a single tier—that is, primary storage.

This methodology does not define process for Continuous and policy-driven migration which, unlike all other data migrations, is a continuous process (requiring the installation of an archival application) that uses a set of user-defined policies to determine, in real time, when and where data needs to be moved.

4 Planning

To achieve the goal of a successful Legacy data migration through conversion, it is imperative that a lot of upfront planning happens prior to that move of data, irrespective of the complexity of the task.

In this stage KMG, in conjunction with its client, prepares a plan, which is also intended to shorten the duration of the Conversion/migration process and also reduces business impact and risk.

The **Migration Plan**, which is the end result of the planning process by KMG, defines

- Requirements i.e. what data is moved, where it is moved, how it is moved, when it is moved, and approximately how long the move will take

- Team i.e. the Users, Business Analysts, System/Data Analysts, Testers, Target Application Users

- Environment under which the Legacy (Source) and the Target Application system operates, What will be the Data Conversion Stage Area

- Schedules - and task distribution, staffing levels

- Configuration Management Plan – to maintain control over changes needed due to various reasons

The other objectives of this planning process by KMG is to design and document resolutions for events like

- Application downtime,

- Performance degradation,

- Technical incompatibilities, and

- Data corruption/loss.

KMG would initiate such project by deploying a Project Manager and an experienced Analyst to gather information and develop the Migration Plan

Location	Client – On-Site	
	KMG	**CLIENT**
Participants	Project Manager	CIO / CTO
	Business Analyst / SME	Project Manager
		Business Analysts
		Technical Lead (Legacy System)
		Technical Lead (Target System)
Output	The Migration Plan	

4.1 Identify Migration Requirement

KMG's Project Manager and the Analysts, gather information about the proposed conversion/migration task and define the Objectives and the Requirements for the project.

Following information is collected as part of Requirement definition -

a. Whether the project is part of the broader project of implementation of the target application or independent project?
b. What data is to be converted and moved – like the nature and usage of the data – more likely by its current Application / Function system and its environment?
c. How many source systems are involved?
d. What is the size of the legacy system data?
e. Where it is moved to – i.e. the target Application System?
f. Which are the more critical Data elements with respect to the Target Application System?
g. Whether any transformation or elimination will be part of this project?
h. How it is moved – if the client has already identified any tool and/or designed, developed any conversion application system?
i. When it is to be moved and how long the conversion/move could take?
j. Whether the objective is to convert all the Legacy application data at once or the process will be staggered phases?
k. Are there any production applications that may conflict with the migration?
l. Can the data in legacy system be migrated into the target system without data cleansing?
m. What is the level of data cleansing required to maximize the benefit of conversion and migration?
n. Whether Historical Data to converted or not – as it could be used by the Target system in future? Historical data is any transactional data that has completed its business cycle and is generally needed for inquiry purposes (statistical analysis, trending, comparison generation and legal implications). Historical data is typically not converted from legacy systems.
o. What are the major complications anticipated in the conversion process?

There could be, and more likely would be, few more questions based on the response to some of the questions listed above. It is also possible that all the questions listed above may not be answered in specific details (by the Client representatives) and some response could be generic. However those answers could still be useful to provide idea about the intended plan of the Client.

Responses to the above queries would be analyzed by KMG's team and then the Team would define the **Requirements** for the project and have that validated by the Client Representatives.

4.2 Identify Team and Stakeholders

KMG expects that the Top Management / CIO from Client's side would designate the following

a. **Client Project Owner** of the proposed Conversion/Migration project, who would have the absolute power to resolve all issues and take necessary decisions
b. **Client Project Coordinator/ Manager** – to assist KMG in all phases of the project and expected to be well versed with either or both Application Systems - Legacy and the Target Application system

Then the KMG Project Manager, in consultation with Client Project Owner and Client Project Coordinator / Manager, would identify various key personnel, who will be required to be part of the team for the project of Conversion and Migration of the Legacy System Data.

The table below defines the roles and responsibilities of various individuals – participating in Legacy System Data Conversion project from the Client side.

Roles	Responsibilities
Business Analysts	- Explains the Business Processes/Functions of the Legacy System and Target System - Identify Data elements / Groups related to various Business Process and Functions - Help to prepare / validate the Data Profile created by KMG's Business / Data Analysts - Validate Data Mapping document – developed by KMG's Business Analysts and Technical Leads for the Conversion and related reporting. - Analyze impacts of the Data Conversion processes and related actions (including Freezing of operation) on the Business
Technical Lead (Legacy System)	- Assist in KMG's team in learning/exploration of the Legacy System Environment - Help in developing the Statistics (File/Table Counts, Record/Rows count) for the Legacy System - Assist in Identifying the Legacy System Entities (in terms of File / Table name, Record Layouts) - Resolve queries during Analysis phases - Identify and extract relevant Legacy System Data for KMG's Technical Team for Testing of Conversion Programs/ Migration Tool
Technical Lead (Target System)	- Assist in KMG's team learn/explore the Target System Environment - Assist in Identifying the Target System Entities (in terms of File / Table name, Record Layouts) - Resolve queries during Analysis and Mapping phases - Makes technical decisions relating to the configuration of data exchange and transform features – after Data Mapping exercise is completed - Assist in determination Test results based on extracted relevant Legacy System Data for KMG's Technical Team for Testing of Conversion Programs/ Migration Tool
Database Administrator	- If RDBMS being used in either or both of the Legacy and Target system, DDL could be generated to identify Table and filed properties - Create and Maintain Test Database for the Target System and Staging Environment for the actual conversion

Roles	Responsibilities
	-
System / Network Administrator	- Provide help in understanding the Legacy System Environment and Target System Environment - Help to setup up Staging Area (Servers/ network) - Connectivity between Legacy System and Staging Area - Execute the process of DATA LOAD in target environment
Project Coordinator/ Manager	- Manages the Data Conversion Processes and repo/rting - Review and if required modify, develop the Migration Plan - Coordinate and Provide necessary resources – as per the Migration Plan - Assign the right Technical resources

KMG will assign Technical Resources to the Project team –

- Business Analyst, having relevant experience of the business processes
- Data Analyst / Subject Matter Expert (SME)
- Technical Lead – with experience in design and develop application system
- Developers, when required

In KMG - all projects are overseen by a Project Owner (Normally a AVP/VP of the company). Often in KMG project the PM and the Technical Lead have interchangeable skills and act as backups for each other.

The backup developers continuously shadow the main team members. They are as good as any other developer in the team and can take over from anyone at a day's notice. These people are used in cases of attrition as well handling spikes in the load (if any). These people also make sure that the work is not disrupted if any member goes on vacation. A fresh person is added to the team as backup the day any backup is absorbed into the team

4.3 Identify System Environment

There are a few key environmental factors that influence the migration methods selected i.e. how the data is moved.

KMG's Technical Team Leads would gather information about the environment in which the Legacy system operates

a. Operating System in which the Legacy System is hosted
b. Database Type – whether it is RDBMS or Hierarchical Database or Flat File with Indexed files (like C-ISAM)
c. Language used to develop the Legacy System
d. How and where data is stored, backed up, and archived.

Same details would be gathered for the Target System. Such information helps to determine whether the Legacy System data need to be converted to another format (like EBCIDIC to ASCII) which increases the complexity.

Apart from the ones mentioned above technical information is gathered with respect to the Network in which the Client organization hosts the Legacy System and Target System and operates those and other Hardware, servers used by the organization.

There other technology considerations, such as:

a. How old is the operating system(s) under which data is to be migrated? Some migration tools do not support legacy operating systems.
b. What staging area requirements are present, given current technologies and data migration requirements?
c. Whether Client would need or want the option to recover quickly from the source disk, or to fall back to the original storage device as a fail-over? This is to determine/design both procedural and technological ways to accomplish that.
d. Is a central console needed to manage data migrations across multiple servers?
e. Is there a need to control the data migration from a local server – or a remote server? If remote, which protocols must be supported?
f. Is there a requirement to throttle or control data flows between servers?
g. Which storage tiers are involved?
h. Whether any change in the Target System model is likely or not? Should that is possibility – that should be taken into account to analyze the consequences of the change in the eventual conversion process.

Based on the information gathered in response to the queries, listed above, KMG will enhance the Requirement definition. All these information will greatly influence in determining the Conversion Framework, method and ETL tools to be used or not.

4.4 Create Project Schedule

KMG understands that at this stage it would not be possible to prepare a detailed realistic schedule as there would be too many unknown parameters. Since this methodology uses a targeted search strategy, the amount of time required to cleanse data is directly proportional to the number and complexity of the mappings and transforms which must be applied to the source data.

An especially tricky part of doing any planning is that the target database is usually a "moving target" – it is being developed at the same time as the cleansing and conversion software, and is always subject to change. If the new software is not a custom application, but a commercially available package, then it makes the schedule creation process little easier even though that possibility still remains as most of the package system often heavily customized.

KMG will create a schedule – using a project plan and pert chart, showing tasks involved, dependencies between the tasks, and the staff required to perform each task. The schedule will include time to be spent on task of familiarizing with the legacy systems' operations and its data elements and MIS staff with the data cleansing effort, and the tasks required from them.

Key milestones in KMG methodology are

a.   The Migration Plan in place and agreed.
b.   Project team is formed.
c.   Completion of Source and Target System Data Profiling
d.   Completion of Data Mapping Operation and creation of Data Map Specification
e.   Completion of task to identification of Data Cleansing requirement
f.   Decision on usage of Automated Conversion tool or Development of Customized Solution
g.   Development of Acceptance criteria
h.   Identification and procurement of Conversion Tool – if a Tool to be used
i.   Design, Development and System Testing of Customized Solution – if that is decided
j.   Test plan is written.
k.   Creation of the Staging area
l.   Test run of Conversion Process and validation
m.   Execution of the Conversion Process
n.   Reconciliation / data checking reports are run.
o.   Acceptance criteria and performance metrics are evaluated.
p.   A go / no-go decision takes place.
q.   Data Load in the Target System
r.   Validation by the Target System
s.   Data issues may be resolved post-migration.

KMG Project manager develops the schedules in consultation with the Client and the Project team member, taking their experience and issues into consideration.

Project schedules would be revised based on the actual time taken to resolve complex issues and procurement of necessary hardware, software resources and team resources.

4.5 Configuration Management Plan

The Configuration Management (CM) Plan typically specifies how versions of the software will be managed.

KMG would include a detailed procedure for controlling changes to the

-         Mapping requirement and document
-         Software changes, in case customized solution to be deployed

A Change Control Board (CCB) could be constituted, consisting of KMG Project manager and Client Project Owner, to review any proposed database changes and its impact. This would be critical to ensure communication between the members of the project teams – whether software development team, Data Analysts/ SMEs.

KMG will review and recommend, if required, Business freeze in multiple areas – which could be a critical and required component of such Legacy Data conversion effort. Such requirement would be explained and reviewed with business data owners, management and auditors.

Business freeze requirements should be addressed in detail through a separate document and circulated well before the timelines - established for the cutover plan.

5 Analysis

In this methodology, the first stage of data conversion and migration is Data Classification, by creating Data Profile for the data elements used in Legacy System(s) and Target Application.
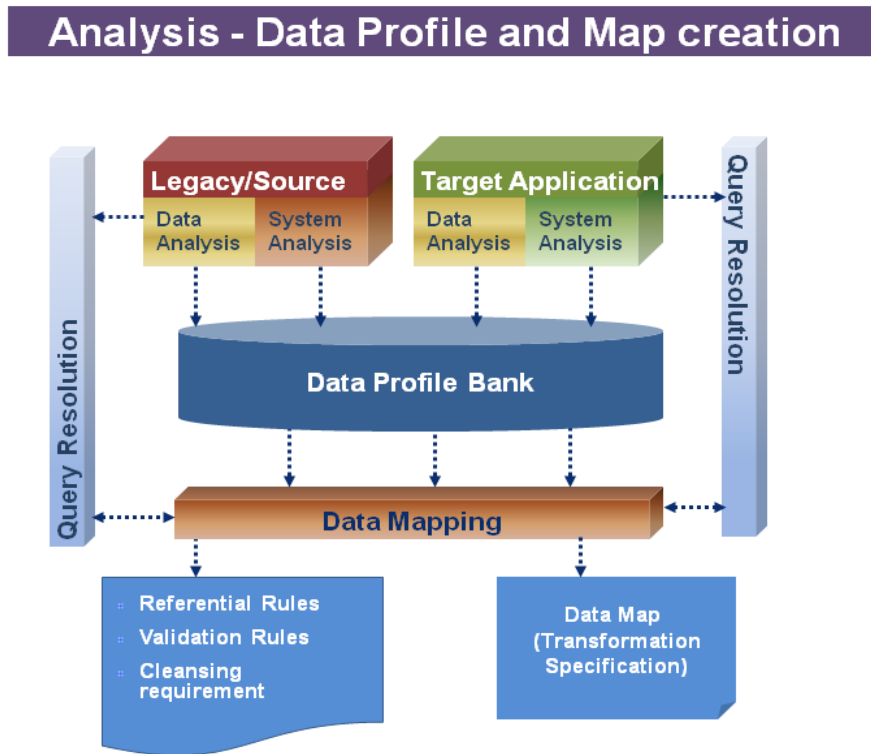

Figure 2

During the Analysis phase of the Project, KMG produces various documents which serves as the specification guideline for the Execution (Convert and Migrate) phases of the project.

Location	On-Site / Off-Site with remote access	
	KMG	**Client**
Participants	Project Manager	CIO / CTO
	Business Analyst	Project Manager
	Data Analyst / SME	Business Analysts
	Technical Leads	Technical Lead (Legacy System)
		Technical Lead (Target System)
Output	Data Profile Repository	Referential Rules
	Data Map – Transformation Specification	Validation Rules
	Migration Framework	Cleansing Requirement
	Test Plan	

5.1 Source and Target Data Profiling

In this stage, Data Structures model analysis involves an in-depth study (qualitative and quantitative in the case of a legacy system) of both the Legacy and Target systems.

The analysis is undertaken by the KMG Data Analysts/SMEs with the help of Client Business Analysts / Technical Leads, and documented in Data Profile repository. This repository, which integrates source and target data in mutual format, provides a profile of the various entities - visibility into that data's usage, capacity, and growth patterns. Various interfaces and reports utilizing the to-be-migrated data are also considered.

Also considered areas are

    a. Must data integrity be checked during (not just after) migration?
    b. How much data will be moving from point to point (server to server)?
    c. What is the estimation of the amount of transformation and cleansing needed?
    d. What are the data profiling and data validation rules/phases applicable to the data?

KMG's Data profiling process consist of three sequential steps with each step building on the information produced in the previous steps. Data sources are profiled in three dimensions: down columns *(column profiling)* ; across rows *(dependency profiling)*; and across tables *(redundancy profiling)*.

Column Profiling. Column profiling analyzes the values in each column or field of source data, inferring detailed characteristics for each column, including data type and size, range of values, frequency and distribution of values, cardinality and null and uniqueness characteristics. This step allows analysts to detect and analyze data content quality problems and evaluate discrepancies between the inferred, true meta data and the documented meta data.

Dependency Profiling. Dependency profiling analyzes data across rows comparing values in every column with values in every other column and infers all dependency relationships that exist between attributes within each table. Dependency profiling identifies primary keys and whether or not expected dependencies (e.g., those imposed by a new application) are supported by the data. It also identifies "gray-area dependencies" those that are true most of the time, but not all of the time, and are usually an indication of a data quality problem.

Redundancy Profiling. Redundancy profiling compares data between tables of the same or different data sources, determining which columns contain overlapping or identical sets of values. It looks for repeating patterns among an organization's "islands of information". Redundancy profiling identifies attributes containing the same information but with different names (synonyms) and attributes that have the same name but different business meaning (homonyms). It also helps determine which columns are redundant and can be eliminated and which are necessary to connect information between tables. Redundancy profiling eliminates processing overhead and reduces the probability of error in the target database.

KMG believes that developing an accurate profile of existing data sources is the essential first step in any successful data migration project. The most significant problem associated with this phase could be if there are frequent changes to the Target Application System Data model. Any change in the target system model would have to be taken into account to analyze the consequences of the change in the eventual conversion process. This renders the whole process to be iterative until a point wherein there is a freeze on the Target System Data structure model or a complete understanding of the legacy system has been reached.

5.2 Data Mapping

Data Mapping is the process in which each source data elements are assigned to one or more target data element. After having done the analysis, KMG's team of Business Analysts, Data Analysts / SMEs and Technical Leads undertakes data mapping process of identifying and documenting the target field for each of the fields in legacy system.

KMG's objective of this process is to produce a **comprehensive mapping** between Legacy and Target System. Every data field that is going to be migrated from the source system to the target system must be defined and examined to ensure compliance with field lengths, data types, domain values permitted, system rules, integrity checks and any other possible issues.

Data mapping process would yield the following results:

- Gaps in the design of the target system:

  a. Un-mapped Target System Data Entity – this are the cases where data requirement is not satisfied by Legacy/Source system. Specification could be developed for Deriving/defaulting data values for such cases – specifically when that data is mandatory field in Target System.
  b. Un Mapped Legacy System Data Entity - for which no Target Data element exists in the target system. This will lead to data loss, which might NOT be desirable in all cases.

  Data mapping process aims at identification and resolution of such gaps and such issues can be addressed in different ways.

  1. Ignore the data in legacy system if it is not used.
  2. If it is business critical data then request for modification of the target system.
  3. In cases wherein target system cannot be modified and it is business critical data that is absolutely essential to retain, consider some workaround to incorporate it as a part of some other field that can be easily understood and extracted with business consent.

- Data Map (Transformation specification): At the end of the data mapping process, a detailed document would be in place that would show the target field identified for each legacy field. Apart from identifying the target fields to which the legacy fields are mapped, mapping specifications define the rules to be applied in the conversion process. These rules are commonly known as *Transformation rules*.
- Referential Rules for Integrity check and domain values permitted,
- Validation rules – Application specific Business rules

Data mapping is an iterative process. For any change in the design of the target system or change in rule for setting a value of a particular field there is a need to amend the mapping specification reflecting the changes in Transformation rules.

The resulting Data Map (Transformation specification) document would be used later

a. In conjunction with third-party data migration tools to extract, scrub, transform and load the data from the old system to the new system or
b. Develop a customized Application system to convert the Legacy / Source system data into Target System data model. This will provide essential information to the programmers creating conversion routines to move data from the source to the target database.

5.3 Data Cleansing

The Data Cleansing process results in relevant and accurate data being converted from the Legacy systems into the Target System.KMG uses the Data Profile Repository prepared during Data profiling stage (by Data Analysis and System Analysis), to conduct the Data Cleansing – in which Legacy application data to verify that data are correct, complete, consistent and convertible, and it also includes process to identify :

- Unnecessary data records and specification for elimination of these
- Inaccurate data and design specification / process to correct these
- Duplicated data, (e.g., multiple vendor masters files, which will need to be consolidated for a single vendor) and design specification / method to merge those
- Data needs Reformatting and standardizing - so that it can be converted
- Mandatory Target System Fields without any related field in the Legacy/Source System and resolution / specification for each of those elements

KMG believes that Data Cleansing is critical to the success of any Data Conversion and Migration project. If not undertaken then business processes will not operate as designed. Data cleansing always takes more time and more resources than anyone anticipates. It is for this reason that data cleansing efforts will be launched as early as possible to make subsequent phase of the project easier and avoid delays.

Data cleansing can be accomplished in two different ways.

a. **Cleansing at the source:** This involves the cleansing directly in the production of the existing legacy system or systems. The main advantage of this approach is the exponential reduction in complications. It makes conversion process very simple.
b. **Cleansing through external means:** This type of cleansing is generally accomplished by spreadsheets. A report is sent to the persons responsible for data cleansing with data that is required to be cleansed. Care should be taken that the spreadsheet contains enough data to make the businessperson understand what they have to cleanse. It should also cater to the technical requirements so that it can be incorporated in the data conversion process with ease. Spreadsheet design is of utmost importance as it can be problematic, if at a later stage, a deficiency were to be identified with the basic design of the spreadsheet which results in an inability to accommodate it in the conversion process.

KMG, suggests that data can be cleansed in the source system or in a staging area. Due to audit requirements and for ease of cleansing, it is recommended that all data cleansing be performed in Legacy Systems unless not viable.

KMG's approach for data cleansing includes working with Client Integration/Functional teams to define / execute automated data cleansing – based on the findings of the analysis phase. Multiple cycles of cleansing could take place. Extracts are made periodically for validation of Cleansing activities and can be handed off to the Technical Development team for sample loads into Staging area.

Key Participating Groups
a. KMG Data Conversion and migration project team
b. Client IS team - Integration/Functional

Deliverables/Outputs
a. Modified source data that increases the success of automated data conversion
b. Control metrics
c. Data Cleansing Requirement Specification (/ Recommendation)

5.4 Design Migration Architecture

Conversion and Migration Framework

Data Conversion (/Migration) is, in general, performed by a set of Customized Programs/Scripts or by a Tool (configured or customized) that automatically transfer the data to Target Application database
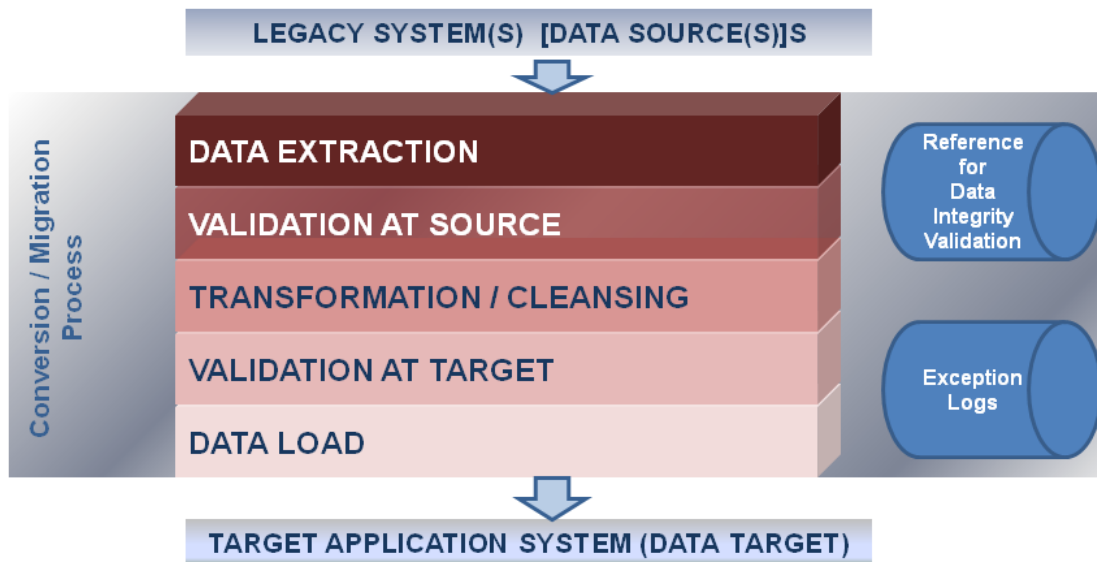
LEGACY SYSTEM(S) [DATA SOURCE(S)]S

Conversion / Migration Process

- DATA EXTRACTION
- VALIDATION AT SOURCE
- TRANSFORMATION / CLEANSING
- VALIDATION AT TARGET
- DATA LOAD

Reference for Data Integrity Validation

Exception Logs

TARGET APPLICATION SYSTEM (DATA TARGET)

Figure 3

The generic framework for a Data Conversion process consists of following steps

a. **Data Extraction** – Read and Gather data from source data store(s) into another storage, and if required converted to the data format of the Target System (e.g. EBCIDIC to ASCII) and loaded in Staging Area

b. **Validation and Cleansing** – to confirm content and structure of extracted data in light of business rules and fulfills integration rules based on the referential rules of Target System. Data Cleansing is performed at this time based on requirements identified during Analysis phase.

c. **Transformation** - convert the extracted data from its previous form into the target form. Transformation occurs by using Transformation Rules defined in Data Map (Transformation specification) and lookup tables.

d. **Validation -Target System** - confirm content and structure of transformed data is valid for target.

e. **DATA LOAD** - Write the data into the target database, either through script or copying data using system utilities.

Steps (a), (b), (c) and (d), shown in figure 3.0 above, are part of the "Convert" phase of the Data Conversion and Migration methodology. Step (d) i.e. DATA LOAD is part of the "Migrate" phase of the project.

KMG makes necessary changes, depending on the specific requirement of each project, to the tasks as defined in the various steps of the framework and create the executable process for Conversion of the Legacy Data

At this stage, KMG, in consultation with Client IT team, analyze and determines whether
- To use a data transformation tool to manage their data Conversion/migration effort? or
- Design and Develop an application system – specifically designed to fulfill that specific conversion requirement based on the Data Map (Transformation Specification) by a group of technical experts?

KMG also determines, based on some of the information and Client requirement – whether to execute the Conversion process in all at once or move data over through a controlled phase of multiple releases? KMG analyze pros and cons to both options, considering which approach will best fit for the project based on organization needs - to be evaluated on a variety of factors, like how much data involved and requirement of the Target System.

5.5 Build Converted Data Test Plan

After determining the Framework, execution model – i.e. tool route or development of the tailored solution, KMG develops Test Process and Plan based on the Data Map (Transformation Rules).

Testing process in Data Conversion and Migration project could be categorized into two :

- **Physical errors** are the result of syntactical errors of the scripts / programs, which can be easily identified and resolved.
- **Logical errors** – these are identified and resolved during Test phase. Such errors are result of the quality of the mapping effort. During Implementation/Testing scripts/programs developed based on the Data Map i.e. Transformation specification, are executed.

Based on the on the Data Map (Transformation Rules) KMG, creates Test Plans – in which KMG identifies Legacy System data element and determines the Target system element and the expected results based on set of extract to be used for testing. This test plan is to be prepared for each of the data element being converted.

Response to the following queries would be gathered and verified by Testing team

> a. How many records were expected to be created by the scripts being tested?
> b. Did the correct number of records get created? If not, why?
> c. Has the data been loaded into the correct fields?
> d. Is the data load complete – or are certain fields missing?
> e. Has the data been formatted correctly?
> f. Are any post-migration clean-up tasks in order?

The goal of a successful data migration is to keep the length of the deploy phase(s) to a minimum.

During Pilot /Testing, KMG would determine the quality of data mapping, by providing the populated target data structures to the users that assisted in the analysis and design of the conversion scripts/application system.

That would help Client Integration/Functional team to understand the data and would allow the user to physically interact with the new, populated data structures of the Target System.

5.5.1 Error Handling and Auditing Requirements

Data Extract process and Transformation (Conversion) process would produce reports from respective processing and those could be used to log and validate the results:

- Record counts from the legacy input data, records with errors, records without errors, and total records
- Record counts for the converted output data
- Counts for the anticipated number of transactions
- Date and time of the start and end of the run
- Any codes that were not found in the crosswalk table, and the record in which they were encountered
- Any fields missing or in error and not found in the conversion table
- Summarized dollar values comparable with legacy system subtotals/totals.

This will help maintain the Conversion tables, enable Client functional teams to verify that the data loaded match the data extracted. This will allow the KMG's project Team to estimate the execution time required for future conversion runs and provide an audit tracking mechanism.

6 Convert

6.1 Design & Develop Migration Tool

The data conversion process can be accomplished by the following methods:
- Using a data conversion tool.
- Scripts developed specifically for the purpose of conversion in the project.
- Manual data conversion and migration.

The choice of a right conversion tool for a given project is always debatable but the following significant factors form the basis of a rational decision

1. Cost involved in procuring a tool and having trained personnel to run the tool. Is this cost less than the cost of employing developers to script the conversion process?
2. Does the tool require any customization? If yes, then the cost and time scales of such a customization effort should be ascertained.
3. The volume of data to be converted and migrated. If the volume of data is much less, manual data conversion and migration is the best option.
4. Whether the Legacy System data and Target environment could be hosted on same environment as difference in operating system / data format would involve manual extraction and conversion to target environment

---

Data Transformation Tool Features

A data transformation tool offers a central repository to support the storage and ongoing maintenance of data mappings, lending itself to teams of all sizes. The most commonly used alternative tends to be spreadsheets, which could fail when one needs to map one source data element to one or more target elements. Spreadsheets typically prevent more than one person from making modifications at one time, resulting in a great deal of unnecessary administrative overhead.

A good data transformation tool offers the following features:

Flexible Reporting - Reporting is easily accomplished, and serves as a comprehensive communication device between the systems and business analysts.

Code Generation from Mapping Rules – Technical Leads/Analysts can generate the migration scripts directly from the mappings they have defined, reducing the overall programming requirements.

Script Scheduling Utilities - Some tools provide far more sophisticated options such as script compilation and scheduling utilities. Scheduling utilities can cut the overall script execution time of a migration project by more than one third in almost every case, by processing scripts in parallel as opposed to sequentially.

Automatic Integrity Violation Detection - Another important and useful feature of a data transformation tool is the automatic detection of integrity violations.

Tailored Application (Manual Programming) - The manual approach to data migration is a method – very useful when there are limited number entities involved and where Client wants to avoid significant investment in a data migration tool or large amounts of effort by one or more programmers.

6.2 Pre-test and Recalibration of Tool

Mock Migration. Conduct dress rehearsals for each planned release. Mock migrations may be partial or complete end-to-end cycles to verify migration procedures and benchmark the cycle times for each migration task.

Pilot Migration. Complete end-to-end migration in the pilot environment. Coordinate with business users in doing data validation, verify and evaluate the control mechanism and metrics.

Test Extract and Load. Once the Pilot Migration is completed, test the extract of the migrated data and test the load into the mock target environment.

Key Activities

- Create/verify data element mappings
- Run data extracts from current system(s)
- Create tables, scripts, jobs to automate the extraction
- Address additional data clean-up issues
- Execute application specific customizations
- Run mock migrations
- Load extracts into the new system using ETL tools or SQL loader with bulk loading functions
- Conduct internal data validation checks including business rules and referential integrity checks
- Report exceptions to client team
- Perform data validation

Key Participating Groups

- Data migration team
- Client IS team
- DBA team

Deliverables/Outputs

- Extracts from source system
- Data migration modules, jobs, scripts
- Application loaded with converted data
- Exceptions, alerts and error handling control points

6.3 Configure Staging Area

KMG's methodology proposes that the Data Conversion process is generally executed in a separate area known as the s*taging area*. This area is created, preferably on a separate server, to have the converted data – without disturbing the existing data in the production srver.

Figure below illustrates the various operations that are carried out in the staging area.
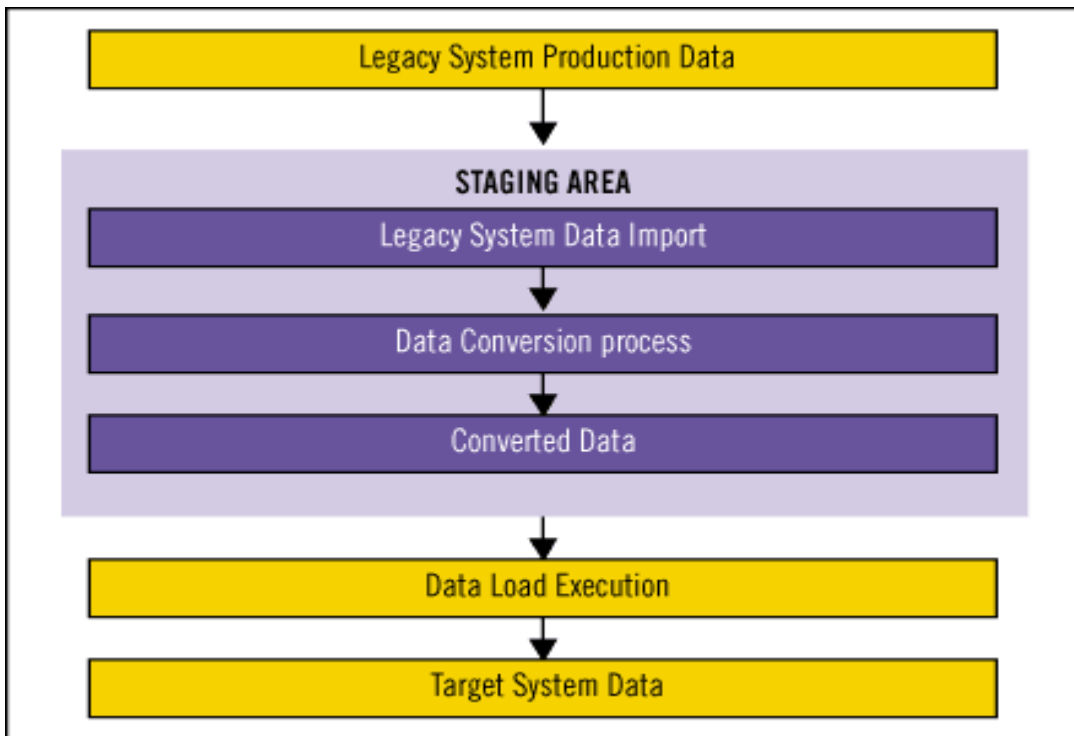
6.4 Execute Data Conversion

Conversion Process Design and Execution. This is the most important phase of data conversion process. Once the data mapping is complete, an overall data conversion process is designed to convert data into a structure required by the target system. Mapping specification serves as an input to this phase. All the rules detailed in mapping specification are incorporated in the conversion process design.

Technical Mapping Specifications. Technical mapping specifications are translation of mapping specifications from a technical standpoint prepared by technical team, and these serve as a primary input to code development. Care should be taken to incorporate each and every mapping rule defined in the mapping specification.

Overall Conversion Process Design. Though each data conversion process design varies depending on the system to be migrated the philosophy remains the same, which is explained below.

Legacy System(s) Data Import. Data from the production database of the legacy system is imported to the staging area.

Data Conversion Process. The data conversion process is executed on the imported legacy system data in the staging area.

Converted Data. The data conversion process gives rise to the converted data that is in the required target data structure model.

Final Extract and Load. Load the converted data into the target environment.

Key Activities

- Run final extracts from the current system(s)
- Execute specific customizations on target database and execute application specific customizations
- Run pilot migrations
- Load extracts into the new system using ETL tools or SQL loader with bulk loading functions
- Conduct internal data validation checks including business rules and referential integrity checks
- Report exceptions to client team
- Perform data validation

Key Participating Groups

- Data migration team
- Client IS team
- DBA team

Deliverables/Outputs

- Extracts from source system
- Data migration modules, jobs, scripts
- Application loaded with converted data
- Exceptions, alerts and error handling control points

6.5 Validate by Test Plans

Testing the converted data is an important activity of data conversion. The converted data, a result of the conversion process, based on mapping specifications should be tested for the following:

a. Is the entire set of mapping rules defined in mapping specifications document incorporated?

b. Does the data converted follow the data structure model of the target system?

c. Are all the fields mapped from the legacy system converted in the appropriate way in the target system?

d. Are all the constraints of the target system fulfilled or is there is any risk of data load rejections?

e. Did the correct number of records get created in the target system?

f. Was the cleansing data incorporated correctly in the conversion process?

g. Do the legacy and target system records reconcile?

7 Migrate

7.1 Develop Migration Statistics

There are specific goals associated with implementing an effective data migration strategy. Primarily, data must be migrated from the source platform to the target platform completely and accurately, and according to company and regulatory policies on information controls and security. This means no dropped or incomplete records, and no data fields that fail validation or other quality controls in the target environment. Another goal of data migration is that the process be done quickly, with as short a downtime window as possible. Finally, the cost of data migration must be manageable, in terms of technology and staff requirements.

There are many metrics that can measure the effectiveness and efficiency of data migrations:

- Number of customizations required
- Percentage of migrated records
- Percentage of migrated tables
- Percentage of data with quality problems
- Number of migration errors
- Migration impact on database size
- Downtime due to migration
- Required staging storage / hardware
- Percentage of reconciliation errors
- Percentage of cleansed data

7.2 DATA LOAD in Target System

Though the data in the staging area is present in the target data structure model, it is required to be loaded to target system. This is generally accomplished by data load scripts that pick data from the staging area and load it to the target system directly.

The data load scripts that load data into target system are tested after the loading is done to the target system.

Delta Data

Another important aspect of a data migration project is accounting for the data input to the production legacy system in between the last export of legacy system and the data conversion process run. This is commonly known as delta data. It has to be understood that the window of the conversion process is finite and the legacy system still operates during this window of conversion process execution. There are various methods to address this issue, notable of them are

1. If the volume of data input to production legacy system is very low, then a manual data conversion is a wise choice.
2. If the data volume does not permit for a manual data conversion then the conversion process is re-run on the delta data in a separate staging area and migrated to production.

Eliminating Delta Data Processing

It has to be appreciated that even though the delta data is migrated, there still exists another set of delta data unless the use of the legacy system is frozen in the window wherein the data conversion process is

executed. A data freeze on production legacy system is not always a very good idea in all the areas of business, as some of them cannot afford a freeze on their systems for a single day. So delta data conversion can be eliminated but such a decision is dependant on the nature of the business in which it is carried out.

Deployment phase

The first step in the design stage is to put together a project plan and structure. As part of this process there should be close analysis of any dependencies in data migrations; where possible, such dependencies and complexities should be reduced to better manage deployment risk. During the deploy phase the following occurs:

- Physical data structures are frozen on source and target.
- Interfaces and processing on source and target are brought down where required.
- Data is staged from the source location.
- Quality reports are run and any data errors or inconsistencies identified.
- Data quality issues are fixed in the staging area.
- A preliminary reconciliation takes place in the staging area, and any reconciling items are investigated and resolved.
- Data is migrated to the target location.

7.3 Validation - Target Application

- Reconciliation reports are run.
- Acceptance criteria are checked; if reconciliation errors or other criteria are not met, the system is rolled back to the original data source. Otherwise, interfaces and processing from the source is discontinued and then activated on the target.

The above may be conducted in phases, or as part of a parallel run or pilot depending upon the migration approach chosen.

7.4 Target System Implementation

Post-Migration Activities

Typical deliverables for the defined phases include:

- Data Migration Approach and Road Map
- Data Source Documents
- Infrastructure Planning and Metrics
- Technical Design Documents
- Failure Routines
- FMEA Document - Failure Mode Execution and Analysis
- Migration Status - Dashboards
- Data Migration Metrics and Control Charts

8 Risks and Mitigation

Every Data Conversion initiative comes with certain risks and such risks should be planned and mitigation steps need to be prepared.

Some risks worth mentioning are

- Application downtime/business freeze,
- Performance degradation,
- Technical incompatibilities, and
- Data corruption/loss.
- Migration might be done as part of a larger chain of dependencies (operating system upgrades, application upgrades or implementations, database structural changes, etc.) – thereby increasing complexity.
- Data requirements are not clearly defined - data rules for integrity, controls, security, availability and recoverability are often ill-defined. In the absence of such rules, data is migrated incorrectly.
- Migration acceptance criteria may not be defined.
- Data is often too distributed to be migrated easily (think end-user computing).
- Budgets may limit technology options for performing migrations.
- Expertise in data migration and management may not be present.
- Management attention might be insufficient (migrations are typically "routine" operations and not major attention-getting projects).
- There could be poor support from the storage vendor(s), making migrations all the more taxing.
- Not treating the Data Migration initiative as a project unto itself.

Each and every phase explained in the data conversion and migration process is iterative. No conversion and migration process is accomplished in a single go. It has to be refined as and when a change is triggered at the design level, generally for the target systems and occasionally for the legacy system (when a certain aspect of the legacy system is omitted in the preliminary analysis by oversight).

RISK MITIGATION

Rollback - When importing data into a target system, what happens if the data migration fails? Are we prepared to either utilize existing transaction rollback functionality or do we have capacity to design and build our own if none exists? How do we manage the client expectation in such cases? Do we have a mitigation plan in place? Have we discussed these with the client IS team and business users?

Answers to these questions gives us an additional layer of security and contributes a lot in terms of executing the project in time, within the budget as well as managing customer expectation all along.

Replication - The issue being: what happens in case of disaster or irrecoverable system failure? Migrating data to a backup system at the same time as a new target system should be seriously considered to add one more layer of security and ensure that the disaster recovery plan is in place.

Data Profiling - Gain a complete understanding of the content, structure, quality, and integrity of the data of the source system.

Data Mapping - Develop an accurate set of data mapping specifications from the source system to the target system.

Migration Approach and Architectural Considerations - Whether point-to-point migration or hub-and-spoke migration, this needs to be evaluated and carefully articulated.

Development - Selecting an ETL tool to automate the migration process and make it more scalable should be a high-priority item.

Quality Assurance - Conduct mock migrations, pilot migrations before the final migration run; this will ensure that the migration process is robust and trusted.