An Automatic Tracking Risks System in Data Migration

Hatem A. Khater

Naval Research and Development Department, Alexandria, Egypt

Abstract: To achieve successful data migration process, data on the previous system is mapped to the novel system providing a design for data extraction and data loading. Converting and cleaning of data is commonly executed in migration to improve data quality, eliminate redundant or obsolete information and match the data schema of the target system. Risk Management (RM) should be part of every project, particularly a significantly large scale migration project. Risks associated to schedule, costs and scope should be clearly identified, documented and a plan must be put together to mitigate them. This paper will clarify what common problems and risks one might encounter on a Data Migration (DM) project. The main objective of this paper is to introduce a practical technique of automated workflow tracking for data migration risks to shrink or even get rid of these risks at any stage of the project, using indicators that can help in measuring of each risk. Therefore this technique can assist with the management of data migration projects, without needing to make the guess of the probability of the risk occurring.

Key words: Risk Management (RM) • Data Migration (DM) • Risk Indicator (RI) • Data Migration Tracking Form (DMTF)

INTRODUCTION

Risk management is increasingly seen as one of the main jobs of project managers. It involves anticipating risks that might affect the project schedule or the quality of the software being developed and taking action to avoid these risks. There are many philosophies and approaches for managing risks. The first step in risk management is to identify and prioritize the risk areas relevant to a project. Each project has different risks due to the unique characteristics that differ from one project to another [1]. The process of risk management is illustrated in Fig.1. It involves several stages:

- Risk identification: Possible project, product and business risks are identified.
- Risk analysis: The likelihood and consequences of these risks are assessed.
- Risk planning: Plans to address the risk either by avoiding it or minimizing its effects on the project are drawn up.
- Risk monitoring: The risk is constantly assessed and plans for risk mitigation are revised as more information about the risk becomes available.

The risk management process, like all other project planning, is an iterative process which continues throughout the project. Risk management is particularly important for software projects because of the inherent uncertainties that most projects face [1]. Consequently every data related project also needs a comprehensive risk management plan and data migration initiatives are no exception [2]. The need to migrate data from one (or many)
systems to another is a major headache for many organizations, large and small. While, the need to have the right data in the right place to support new application initiatives is fundamental, getting it there on time and on budget can all too easily become a recurring nightmare upon which enterprises spend vast sums of money. Moreover, it must be borne in mind that all data migrations are undertaken as a part of broader application; this means that data migration cost and time overruns can severely impact on the overall project. Ultimately, a failed data migration project means a failed application project [2]. For many projects, data migration is one of the last priorities and systems integrators will often defer to the client the completeness and integrity of the data file for loading into the system. Indeed data migration is often viewed as simply transferring data between systems. Yet the business impact can be significant and detrimental to business continuity when proper data management is not applied during the migration [3]. Therefore, the objective of this paper is to implement an automated technique to detect problems that might be faced the Project Manager during data migration project. This will be through the identification of all the indicators that can measure the risks that lead to the failure of the project. In this paper basic concept of Software Risk Management is introduced. In Section 2 we will introduce Data Migration Process and the difference between Data Moving and Data Migration; then in Section 3 a list of common risks that might face the data migration projects; then in Section 4 the use of indicators that could help to measure each risk separately; finally in section 5 the implantation of the DM risk tracking system are discussed in details. A typical example is also presented to demonstrate the application of our automated tracking system.

Data Migration: Data migrations are generally produced from the introduction of a new system. This may involve an application migration or consolidation in which one or more legacy systems are replaced or the deployment of an additional system that will sit alongside the existing applications. Whatever the specific nature of any data migration, the ultimate aim is to improve corporate performance and deliver competitive advantage [3]. Data migration is the process of transferring data between storage types, formats, or computer systems. Data migration is a key element to consider when adopting any new system, either through purchase or new development. Data migration stages (design, extraction, cleansing, load, verification) for applications of medium to high complexity are generally repeated several times before the new system is installed. In many situations this is complicated process, involving data in different formats, multiple databases and other data sources and active applications that rely on the data [4].

Data Moving or Data Migration: Technically migration means to move from one place to another but a distinction needs to be made between simply moving data and migrating data in a computing sense. For example, if we were to consolidate several copies of a single application (e.g. same version of a Customer Relationship Management (CRM) system) distributed across an organization into one single system and then this would be an example of data movement. However, if the versions of the CRM system being consolidated were from different vendors (heterogeneous systems) then this would be an example of data migration. This is because the data being consolidated will require modification in some way e.g. cleansing, transformation, removing duplicates, format changes etc (data transformation) before it can be migrated to the new system [5].

Data Migration Process: A data migration project focuses on the movement of data between legacy (source) data system(s) and a target system, including all necessary procedures for transferring and validating the data throughout the entire process, the high level of Data Migration process is presented in Fig. 2. Before data is moved, often it needs to be modified and/or transformed. This process is called Data Conversion. Planning and performing data conversion requires the development of transformation rules and procedures to implement the necessary changes. For example, if the legacy system stores date information in text format but the target system requires this information to be stored as date format, then a conversion of the legacy data is necessary prior to the data migration [6].

![Fig. 2: High-level data migration process.](image-url)
Data migration may sound simple. It isn’t. In fact, industry experience has shown data migration to be one of the most risky, complexes and time-consuming Information Technology (IT) projects. Cost overruns and delays are common. Business performance can suffer if data migration doesn’t support strategic objectives behind a merger or acquisition, legacy systems modernization, or an upgrade to a new Enterprise Resource Planning (ERP) or CRM application [7]. It is important to continue proactive issue management and proactive risk management throughout the lifecycle of the project. Each data migration project deals with different issues and risks and therefore each requires a risk mitigation Plan specific to the scope of the project. Similar issues and risks may have been identified and/or have occurred during other data migration projects of which the team has knowledge and thus the mitigation and resolution approaches from those projects should be used as guidance [6].

Some Common Data Migration Risks: There are many risks involved in leading a data migration project. It pays to take time during the migration preparation stage to understand and plan for these risks. It is often too late to take mitigating action during the live project, so preparations must be done beforehand [8]. Here are some of the common data migration risks:

- Organizing team (Lack of available expertise).
- Financial projection (budgetary analysis).
- Business needs (Poorly executed scoping).
- Project delivery approach is inflexible.
- Data cleansing - Data Quality (Lack of data quality management strategy and appropriate tools).
- Documentation (Lack of documentation and detailed knowledge of legacy and target environments).
- Tools efficiency (Poor choice of data migration technology).
- Target application is constantly changing.
- Go-Live strategy is inappropriate for the needs of the business.

Data Migration Risk Indicators: Indicators are the measures used to monitor and predict the risks of project. Risk Indicators (RI) are one of the key tools used to support risk assessment and risk monitoring. Risk indicators used to keep the risk management process dynamic and risk profiles current [9]. Table 1 gives some examples of indicators and measures that may be helpful in assessing the risks of data migration projects, which would calculate the probability of the risk. These indicators and equations are presented through studies and experiences in Data Migration projects.

Data Migration Tracking Form (DMTF) Application: The software programs and services used for DMTF system are all commercially available without any proprietary code and include the following:

- Microsoft InfoPath 2007: InfoPath is a commercial software program used to design user friendly form templates, published to and accessed from a common network location, allowing users to fill out forms for exportation to a database program. However it is based on open source standards such as XML and XSLT [4].
- Microsoft SQL Server 2008: MSSQL is a commercial Data-Base Management System (DBMS) program used to store data in a collection of tables with typed columns allowing for the retrieval of information through queries designed to organize and structure the data [10].
- SQL Server Reporting Services (SSRS) 2008: SSRS is a set of features included with Microsoft’s SQL Server product for the creation, management and execution of predefined reports [6].

Database Model: Microsoft SQL Server 2008 was used to develop the database module. This module contains different types of tables. They are, tables contain administrative data, tables contain setup data, tables contain Data Migration risks and its definitions, tables contain strategy plans for each risk and other for transactional data related to the Data Migration workflow process. Triggers, Views, Queries and Scalar functions are used to calculate each risk using the risk measures equations based on the indicators as shown in Table 2.

Sample of Legacy Data: The Hospital Management Information Systems (HMIS) is comprehensive, integrated information systems designed to manage the medical, administrative, financial and legal aspects of a hospital and its service processing. The proposed technique was applied on one of the Hospital Management Information Systems, which monitors the process of extracting the master data from old system (Source) and loading these data to new system [5]. The legacy system consists of several modules such as Front Office, which contain collection of tables that include a data related to Hospital structure, Investigations, Procedures and Patient Data. Back Office module that includes tables related to Warehouse Setup, Item Group Setup, Item Setup and Unit Setup. Human Resource module that includes tables related to Employee Data, Department Setup and Section Setup.
Table 1: Data migration risks indicators and measures.

<table>
<thead>
<tr>
<th>Risk</th>
<th>Indicators</th>
<th>Risk Measure Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of available expertise</td>
<td>Staff Turnover</td>
<td>(No. of employees leaving/ total No. of employees) *100</td>
</tr>
<tr>
<td>budgetary analysis</td>
<td>Total Cost</td>
<td>(Delay tasks cost / estimated cost) *100</td>
</tr>
<tr>
<td></td>
<td>Total Time</td>
<td>(Total Delay Tasks time / estimated time) * 100</td>
</tr>
<tr>
<td>Poorly executed scoping</td>
<td>Change of Requirements</td>
<td>No. of Change Requirements / total (No. of requirements) * 100</td>
</tr>
<tr>
<td>Project delivery approach is inflexible</td>
<td>Number of task dependencies</td>
<td>(No. of delayed tasks / total No. of tasks) * 100</td>
</tr>
<tr>
<td>Lack of data quality</td>
<td>Re-opened tasks because of lack of data</td>
<td>(No. of re-opened tasks because of lack of data quality / total No. of Tasks) * 100</td>
</tr>
<tr>
<td>Lack of documentation</td>
<td>Re-opened tasks because of lack of documentation</td>
<td>(No. of re-opened tasks because of lack of documentation quality / total No. of Tasks) * 100</td>
</tr>
<tr>
<td>Poor choice of data migration technology</td>
<td>Time of extracting and loading Data</td>
<td>(Delay of extract and loading / the estimated time for extracting and loading the data)*100</td>
</tr>
<tr>
<td>Target application is constantly changing</td>
<td>No. of re-opened tasks because of target changing</td>
<td>(No. of re-opened tasks because of target objects changing / total No. of Tasks) * 100</td>
</tr>
<tr>
<td>Go-Live strategy is inappropriate</td>
<td>Delay of Go-live Time</td>
<td>(Delay of Go-live Time / the estimated time for Go-Live)*100</td>
</tr>
</tbody>
</table>

Table 2: Difference between Initial Probability and Project Probability.

<table>
<thead>
<tr>
<th>Risks</th>
<th>Initial Probability</th>
<th>Project Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of available expertise</td>
<td>7</td>
<td>6.9</td>
</tr>
<tr>
<td>Financial projection</td>
<td>14</td>
<td>9.5</td>
</tr>
<tr>
<td>Poorly executed scoping</td>
<td>4</td>
<td>9.2</td>
</tr>
<tr>
<td>Project delivery inflexible</td>
<td>14</td>
<td>19.1</td>
</tr>
<tr>
<td>Lack of Data Quality</td>
<td>21</td>
<td>19.1</td>
</tr>
<tr>
<td>Lack of Documentation</td>
<td>12</td>
<td>11.5</td>
</tr>
<tr>
<td>Poor Migration Technology</td>
<td>9</td>
<td>11.5</td>
</tr>
<tr>
<td>Target Application Changing</td>
<td>11</td>
<td>7.6</td>
</tr>
<tr>
<td>Go-Live Strategy Inappropriate</td>
<td>9</td>
<td>5.7</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

**Proposed Model:** The proposed technique that described in this paper is based upon list of common data migration risks and the data migration workflow process. This technique can be used as virtual auditor for self risk assessment of data migration life cycle. A sample run is demonstrated in this section. The proposed application was developed as a three-tier architecture as shown in Fig. 3. The high-level process flow of the DMTF technique is presented in Fig. 4.

Upon execution of the technique, the main screen appears as shown in Fig. 5. The application gives the user the opportunity either to create a new project or to choose one created project from database. Then in case of creating new project, the user must to write the name of the project, specify the date the project was started, the end date which the project is scheduled to be completed and for Actual End Date which the date of the project was actually completed, this date will be determined automatically when the project status changed to closed.

![Fig. 3: Three-tier architecture of the proposed technique.](image-url)

Once the New project is created the user must specify the business scope and the system scope which represents the requirements that all the tasks will depend upon as shown in Fig. 6. Now the user must to create the initial data migration risk analysis based on the list of potential data migration risks, which are identified before on database to help user in the analysis process by assuming the probability of a risk and its impact as shown in Fig. 7.
PM Create New Project
- Specify the system & Business Scope.
- Add initial Risk Analysis
- Assign team roles.
- Define the team salary.

PM Add New Task
- Select the task type & Data migration object.
- Identify the task Duration.
- Assign the task to TM1.

TM1 receive the task
- TM1 working on the task.
- Change the task status to Initial Close.
- Send task to TM2 for reviewing.

TM2 receive the task
- Reviewing the task result.
- Is there any problem?
  - Yes
    - TM2 identify the Problem.
    - Change the task status to Re-Opened.
    - Send task to TM1 again.
  - No
    - TM2 Change the task status to Final Close.
    - Actual Date will be determined automatically.

PM closes the project upon completion of all tasks.

Fig. 4: DMTF technique process flow.

Fig. 5: DMTF Main Screen.

For more information about every risk like risk definition and strategy plan, the user can press the Data Migration Risk Setup as shown in Fig. 8. Else the PM can press back to assign the staff to this project by pressing on Assign Staff Roles button as shown in Fig. 6. Now the Project Manager (PM) can select the team member’s participants in the project and specify their salaries because it represents the cost of the project as shown in Fig. 9.

Now the PM can start to assign tasks to participants by backing to the main screen and pressing on New Task then Open button as in Fig. 10. After opening the created task the PM can assign this task to participant member then select the category of data objects that are specified before in project scoping as a requirements and choose one of the objects that are listed under the selected category, after that the PM can specify the task duration and the task type, next press on Save & Send the task as shown in Fig. 11. The team member will receive this task and work on it and it will appear as shown Fig. 12.

Fig. 6: Specifying the business and the system scope.
Fig. 7: Creating the Initial Data Migration Risk Analysis.

<table>
<thead>
<tr>
<th>Data Migration Risk</th>
<th>Probability</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organizing team (Lack of available expertise)</td>
<td>7%</td>
<td>High</td>
</tr>
<tr>
<td>Financial projection (budgetary analysis)</td>
<td>14%</td>
<td>Medium</td>
</tr>
<tr>
<td>Business needs (Poorly executed scoping)</td>
<td>4%</td>
<td>Low</td>
</tr>
<tr>
<td>Project delivery approach is inflexible</td>
<td>14%</td>
<td>Medium</td>
</tr>
<tr>
<td>Data Quality (Lack of data quality management strategy and appropriate tools)</td>
<td>21%</td>
<td>Medium</td>
</tr>
<tr>
<td>Documentation (Lack of documentation and detailed knowledge of legacy and target environments)</td>
<td>12%</td>
<td>High</td>
</tr>
<tr>
<td>Tools efficiency (Poor choice of data migration technology)</td>
<td>9%</td>
<td>High</td>
</tr>
<tr>
<td>Target application is constantly changing</td>
<td>11%</td>
<td>Low</td>
</tr>
<tr>
<td>Go-Live strategy is inappropriate for the needs of the business</td>
<td>0%</td>
<td>Medium</td>
</tr>
</tbody>
</table>

Fig. 8: Data Migration risks definition and strategy plan.

Fig. 9: Selecting the team participants and specifying their salaries.
Fig. 10: Adding new task.

Fig. 11: Assigning task to team member screen.

Fig. 12: Team member receiving task screen.
Once team member finishes the task, he must change the task status to Initial Close then select the responsible for reviewing the task and press send button. After reviewing the task, if there is any problem the reviewer will change the task status to Re-Open then identifies the problem type and choose one of the problems that are listed under the selected problem type, else the reviewer will change the task status to Final Close. Once the task status changed to final close the actual date of closing this task will determined automatically and saved to database. In parallel the PM has opportunity to monitor the project risks compared with the initial risk analysis by using the risk tracking report at any stage of a project as shown in Fig. 13. This report displays the risk analysis graph which illustrates the percentage of each identified risk. This analysis automatically produces by calculating each risk through the measurement equations that depends on the Data Migration workflow.

As shown in the Table 2, it can be concluded that there are many risks unpredictability in addition to they can’t be accurately determined, so that the dependence on the project manager’s experience in guess the probability of risk occurrence is inaccurate, but the dependence on the proposed technique which uses the equations that calculate the probability of risk based on the risk factors and indicators, gives more accurate results.

CONCLUSION

In this paper Data migration process and the difference between Data Moving and Data Migration are introduced. The common Data Migration risks and measuring equations which used to calculate each risk are listed. An automated technique of Data Migration Tracking Form is presented. This technique is designed to play the role of a virtual risk auditor, to help project managers to manage Data Migration risks and automatically calculate the probability of each risk and compare it with the initial risk analysis. With this application, the project manager can dispense with the risk management team. Therefore there is no need to guess the likelihood of the event occurring, because there are many threats unpredictability in addition to they can’t be accurately determined. The DMTF technique process flow, the software component, database model and the implementation of the DMTF technique are discussed. The benefits from this automated technique are: one can save time, money and accuracy of risk analysis will be high compared to the manual process and it is efficient with respect to time and follow-up tasks.
REFERENCES