

Data Migration Process Strategies

Win Zaw latt

latt.winzaw@gmail.com

Abstract

Today, data migration is important to the organizations to transform the data sources of old systems to the new systems. Most companies do data migration to update their systems. Data migration is the process of moving data from one location to another, one format to another, or one application to another. Generally, this is the result of introducing a new system or location for the data. The business driver is usually an application migration or consolidation in which legacy systems are replaced or augmented by new applications that will share the same dataset. These days, data migrations are often started as firms move from on-premises infrastructure and applications to cloud-based storage and applications to optimize or transform their company. This paper describes strategies and techniques in data migrations in organization perspective.

1. Introduction

Data migration means that numerous operational systems and external data sources that includes Relational Data Base (RDB) which includes schema translation and data transformation, personal files from various sources have to transform from old system to new systems. The core reason for need of migration is upgrading the existing system into developed system according to the industry requirements. The impending problem is redefining of existing database and storage system in terms of complex language. Based on the industry rule, it is common that during the migration, the source and target databases are structurally different, or data is inconsistent across multiple data sources. Due to this problem, several researches and development of migration tools are emerged continuously [12].

2. Literature Review

The research of Housel et al.'s 1974. [9] paper at IBM is one such example where some of the main issues in data migration are summarized and future researches reasons include 1) a change in the hardware system, 2) conversion from one system to another, 3) a change in the structure or program as a result of

modified application requirements, and 4) the addition of an application to an existing database. Other significant early contributions include Fry [6] and McGee [6], who suggest the creation of a data and storage structure definition language.

Young and Ku [20] provide a concise but rather insightful overview of many of the main issues of data migration along with some helpful examples. The article contains many of the primary issues that should be considered during migration. They initially define the process of migration as moving data from a source to a target database – which can actually consist of multiple databases in each case. What makes this process difficult and, therefore, worthy of research and analysis, is the fact that the source and target are rarely identical. Specifically, this means that in most cases, decisions will be made along the way to identify which structures and data will be needed in the target system.

Hudicka [11] also provides a good overview of the phases for data migration. His breakdown may be slightly different than Young and Ku's [20], but both articles provide useful starting points. Hudicka [11] points out that in the case of migrating from legacy systems which are based on hierarchical databases, the migration process needs to be planned especially carefully, since many of these systems do not enforce referential integrity, while two cornerstones of this older structure – de-normalization and redundancy – are in precise contradiction to more modern relational theory. He then argues for a number of different phases each of which should be completed before proceeding to the next phase. In the pre-strategy phase, the project manager should identify the number of legacy systems and count their data structures. Interfaces should also be identified at this point, if possible.

The design phase should include the mapping of key constraints and performing data mappings from the logical to the physical model. The subsequent testing phase should deal with both logical and physical (syntactical) errors. Once test data has been migrated, basic questions should be addressed:

- How many records were supposed to be created?
- How many were actually created?
- Did the data migrate to the correct fields?
Was the data formatted correctly?

3. Data Migration Process

To achieve an effective data migration procedure, data on the old system is mapped to the new system utilising a design for data extraction and data loading. The design relates old data formats to the new system's formats and requirements. Programmatic data migration may involve many phases, but it minimally includes data extraction where data is read from the old system and data loading where data is written to the new system.

After loading into the new system, results are subjected to data verification to determine whether data was accurately translated, is complete, and supports processes in the new system. During verification, there may be a need for a parallel run of both systems to identify areas of disparity and forestall erroneous data loss. Automated and manual data cleaning is commonly performed in migration to improve data quality, eliminate redundant or obsolete information, and match the requirements of the new system. Data migration phases (design, extraction, cleansing, load, verification) for applications of moderate to high complexity are commonly repeated several times before the new system is deployed (Figure 1).

The first two categories of migration are usually routine operational activities that the IT department takes care of without the involvement of the rest of the business. The last two categories directly affect the operational users of processes and applications, are necessarily complex, and delivering them without significant business downtime can be challenging. A highly adaptive approach, concurrent synchronization, a business-oriented audit capability and clear visibility of the migration for stakeholders are likely to be key requirements in such migrations [19].

Data from
different sources

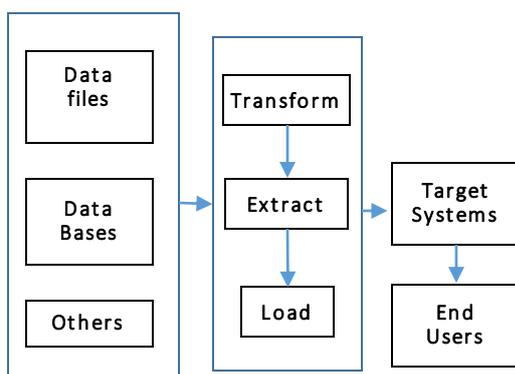


Figure 1: Data Migration Process

3.1. Types of Data Migration

3.1.1. Schema migration

Similarly, it may be necessary to move from one database vendor to another, or to upgrade the version of database software being used. The latter case is less likely to require a physical data migration, but this can happen with major upgrades. In these cases, a physical transformation process may be required since the underlying data format can change significantly. This may or may not affect behavior in the applications layer, depending largely on whether the data manipulation language or protocol has changed – but modern applications are written to be agnostic to the database technology so that a change from Sybase, MySQL, DB2 or SQL Server to Oracle should only require a testing cycle to be confident that both functional and non-functional performance has not been adversely affected [19].

3.1.2. Application migration

Changing application vendor – for instance a new CRM or ERP platform – will inevitably involve substantial transformation as almost every application or suite operates on its own specific data model and also interacts with other applications and systems within the enterprise application integration environment. Furthermore, to allow the application to be sold to the widest possible market, commercial off-the-shelf packages are generally configured for each customer using metadata. Application programming interfaces (APIs) may be supplied by vendors to protect the integrity of the data they have to handle [19].

3.1.3 Business process migration

Business processes operate through a combination of human and application systems actions, often orchestrated by business process management tools. When these changes they can require the movement of data from one store, database or application to another to reflect the changes to the organization and information about customers, products and operations. Examples of such migration drivers are mergers and acquisitions, business optimization and reorganization to attack new markets or respond to competitive threat [19].

4. Legacy Migration

Legacy migration has been classified into well-defined interfaces, applications, and database services. For legacy migration, user and system interfaces are separate modules, at the same time applications and database services are not separable. Legacy migration strategies are easy to apply, fast to implement, and can be widely applied to industry software projects. An important issue of this system is that it is very difficult to incorporate with newer system such as open source operating systems as having of non-extensibility, incompatibility, and less-openness of the underlying hardware and software of the legacy systems [22], figure 2.

Legacy migration life cycle includes the following procedures:

- Before Migration : Plan, assess and prepare
Assess hardware, software and network readiness and plan for future
Clean up by eliminating useless data, consolidating resources, monitoring everything
- During Migration : Prototype, pilot and deploy migration
Use powerful database modeling to simulate migration, resolving issues before commit. Track migration
- After migration : Maintain and manage new environment

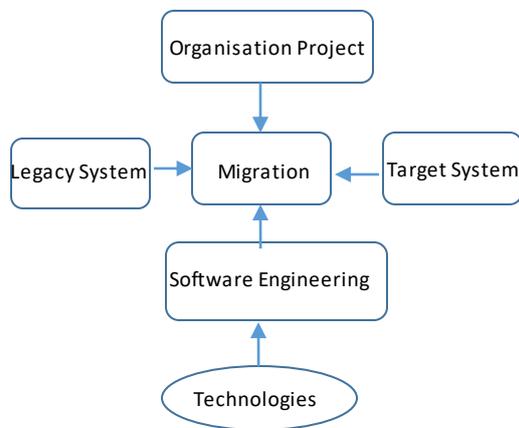


Figure 2: Framework for legacy system (Bisbal et al., 1999 [22])

5. Migration strategies

Senior researchers Bin Wei, and Tennyson X. Chen [5], developed Data Migration Tool (DMT) with five criteria for US National Oceanic and Atmospheric Administration (NOAA) that need to be considered when evaluating a DMT [22].

The five criteria's are

1. Types of database the DMT supports, If does not support the database from or to which users need to perform the migration.
2. How the database transfer is configured through the DMT's interface. This configuration will determine whether the data transfer can be executed repeatedly.
3. DMT should check database integrity before executing a data transfer. Sometimes database migration operation fails because of database integrity violation among the data.
4. How well the DMT incorporates customized data transfer requirement.
5. Ensuring the correctness of a completed database migration operation

In the organization, Data Base expert planning a data migration should consider what type of migration is suitable for their needs. Migration includes data files, different types of operation systems and platforms, personal files and numerous sources. They can choose from several strategies which depends on the project requirements and available resources. There are two migrations: Big Band Migrations and Trickle Migrations.

6. Big bang migrations

Big bang migrations involve completing the entire migration and defined processing of window. The system migration that involves downtime of the systems while the extracting the data from the source system(s), processed, and loading to the target system, followed by the switching of processing over the new environment. This approach is short-possible time, but it carries the risks. Some organizations can live with an old core system for long time, so there is intense stress on the migration and verification of data and sign-off are on the critical path.

Organizations need to aware that business adopting this approach should prepare plan to run of the migration before the live event and plan a contingency data for the migration in case of failure of first attempt. Few organizations ever do this that process. Big band migrations are most often planned as a one-off that is requiring a shutdown over a weekend or night time or public holidays, meaning that the quality of the migrated data is often compromised.

7. Trickle migrations

Trickle migrations is an incremental approach to migrating data, rather than to complete the whole event in a short time window. This migration involves running the old and new systems in parallel and same time migrating the data in phases. This method the zero downtime that mission-critical applications requiring 24/7 operation is needed. This migration can be implemented with real-time processes to move data and these processes can be used to control and maintain the data by passing future changes and validations to the new target system.

Using the trickle approach needs to add some complexity to the design, because it must be possible to track that what types of data has been migrated. This is the part of a system migration, it means that source and target systems are operating in parallel, with users must switching between them, depending on where the information that they need is currently situated. Alternatively, the old system(s) can continue to be operational, until the entire migration is completed and transferred, before users are switched to the target system(s). In such a case, any changes to data in the source system(s) must trigger remigration to the target systems is updated correctly.

8. Data Mapping Route

Data mapping is the important priority to ensure that all physical transfer of data from primary source(s) to the target system without disruption of the daily business processes. To ensure focus on preparing specific source(s) of data routes from the existing systems to the target systems in safe physical transfer of data. All data, files, documents must be transferred to the target systems. Understanding and knowledge of data types, data source(s), files types, types of documents that need to be ensured and transferred to the target systems.

Before starting the project, data migration experts and team must ensure fully understand the data sources before starting the project. Understanding the nature of data, types of data and sources of data are important. To achieve a complete profile and audit of all sources of data must ensure and understand in early stage and it can deliver tangible benefits. In initial stages, team need to ensure to identify and address all potential problems before starting the data migration project. Team need to develop the rules for migration plans, data mapping methods, testing methods that can be based on thorough analysis of sources of all data

rather than small data set. Making decisions can be based on proven facts rather than assumptions.

9. Validation and Implementation of Data

Following key attributions should be done:-

1. Relevance: Is it relevant to all sources of data?
2. Accuracy: Can data be accurate?
3. Integrity: Does it logical structure?
4. Consistency: Is it understand and consistent?
5. Completeness: Does it ensure to complete?
6. Validity: Is it valid for business processes?
7. Timeliness: Is it updated?
8. Accessibility: Can it be accessed and reliable?

Compliance: Does it comply to standard?

10. Testing and verifying Data

Before the conversion of data, teams need to ensure unit, system, volume, online application teste need to be carried out before the sighed off by the business agreement. Each unit of work, the full-volume upload and online application test stage as early as possible to test and ensure of several units of work may need to be completed and online test can be done in the initial stages. This stage helps avoiding and preventing storing up issues until too late in the development stage, and that cause expensive to rectify. Another issue is that data is changing in the source systems, when migration process is in development. Creating a profile and auditing of the sources, that is possible to rerun the audit to access any changes and taking appropriate necessary action. That has to be done before the milestones of all major projects stages, to facilitate to continue, stop and necessary decisions.

There are many complexities involved in carrying out data migration testing and verifying, a small miss that can cause in any aspect of verification during testing will lead to the risk of failure of the data migration project. It is important to consider carrying out and careful through study and analysis of the system before starting the project and after the migration. Planning and designing the effective migration strategy with the robust tools along with the experts, skilled and trained users to test the systems.

Migration has huge impact on business processes, quality of the applications, data, files, databases, programs. The teams to verify the entire systems in all aspects like accuracy, validation, verification, performance, usability, integrity, reliability, completeness, compatibility, consistency,

relevance and timeliness which will turn to ensure “Data Migration” successfully.

11. Methodology

A pilot conversion for software migration method that involves rolling out the new system to a small group of users for testing and evaluation. The pilot study is the important for the data migration project before implement the actual systems in an organization. During the pilot implementation, the test group users can provide valuable feedback on the system to make the eventual rollout to all users go more smoothly. Once the test group has approved the system, it can be rolled and out across the organization.

The pilot study includes following steps

Step 1. Big Bang Adoption: Complete and simultaneous implementation of the new system across an organization.

Step 2. Phased Adoption: rolling out the new system incrementally

Step 3. Parallel Adoption: running both old and new systems simultaneously until implementation of the new system which is ready to be used and successful.

12. Conclusion

The data migration is ultimate objective to improving performance of organization and deliver competitive advantages. To successful data migration in organization, to give the attention that is deserve, rather than underlying project to be being considered. Proper planning is the important requirement of the data migration project, there is a risk that can effect the over budget of the project, lost of data, exceed the deadline and or even fail in complete.

Team members of data migration to be flexible and highly skills, require minimal technical knowledge, and be intuitive so the business and technical staffs can work collaborative. Users should be able to understand, implement complexity of business rules for data migration or quality assurance of data.

References

- [1] Andreas Meier et al., (1994). ‘Hierarchical to Relational Database Migration’, IEEE Software, Vol. 11 Issue 3, pp 21-27.
- [2] Abu-Hamdeh, R., Cordy, J. & Martin, P. (1994). Schema translation using structural transformation. Proceedings of the 1994 conference of the Centre for

- Advanced Studies on Collaborative research, (pp. 123-43), IBM Press.
- [3] Agrawal, R., Evfimievski, A. & Skrikant, R. (2003). Information sharing across private databases. International Conference of Management of Data. Proceedings of the 2003 ACM SIGMOD international conference on Management of data. (pp. 86-97). New York: ACM Press.
- [4] Alan, Robert. (2002). The Serials Data Migration Dilemma. Technical Services Quarterly 20 (4), 29-38.
- [5] Bin Wei and Tennyson X. Chen (2012), ‘Criteria for Evaluating General Database Migration Tools’, [online] Research Report, RTI Press Publication No. OP-0009-1210. Research Triangle Park, NC: RTI Press. <http://www.rti.org/pubs/op-0009-1210-chen.pdf> (Accessed 28 August 2015)
- [6] Fry, J. P. (1970). Introduction to Storage Structure Definition. ACM SIGFIDET Workshop on Data Description and Access.
- [7] Ganti, N. & Breyman, W. (1995). Transition of Legacy Systems to a Distributed Architecture. John Wiley & Sons.
- [8] Griswald, W. G. (1991). Program Restructuring as an Aid to Software Maintenance. (PhD Dissertation, University of Washington, 1991.
- [9] Hausel, Barron C., Lum, V. & Shu, N. (1974). Architecture to an Interactive Migration System. Proceedings of the 1974 ACM SIGFIDET (now SIGMOD) workshop on Data description, access and control (pp. 157-69) New York. ACM Press.
- [10] Henrard, J, J. Hick, M, Thiran, P. & Haimaut, J –L. (2002). Strategies for data engineering. Proceedings. Ninth Working Conference on Reverse Engineering. (pp. 211-220). Los Alamitos, California.
- [11] Hudicka, J. R. The Complete Data Migration Methodology. Retrieved September 2, 2004. From <http://www.dulcian.com/papers/The%20Complete%20Data%20Migration%20Methodology.htm>
- [12] Moriarty, T. & Hellwege, S. (1998). Data migration. Database Programming & Design, 11-14.
- [13] Philip Howard, “Data Migration: A White Paper”, Bloor Research, September 2007
- [14] Philip Howard and Carl Potter, "Data Migration in the Global 2000- Research, Forecasts and Survey results", Bloor Research, September 2007 [3] The Great Migration, Oracle Magazine, May/June-2002 Volume XVI, Issue 3, page-57-65
- [15] Sibley, E. H. & Taylor, R. W. (1970) Preliminary Discussion of a General Data-to-Storage Structure Mapping Language. ACM SIGFIDET Workshop on Data Description and Access, 368-80.
- [16] Smith, D. P. (1971). An Approach to Data Description and conversion. (Doctoral. Dissertation, University of Pennsylvania, 1971.
- [17] Weiderman, Nelson H. Bergey, John K. Smith, Dennis B. Tilley, Scott R (1997) Approaches to Legacy System Evolution, [online] Technical Report CMU/SEI-97-

TR-014 ESC-TR-97-014, Software Engineering Institute, Carnegie Mellon University, Pittsburg, PA 15213.

- [18] Weiderman, N., Smith, D. & Tilley, S. Approaches to Legacy System Evolution.
- [19] Retrieved October 1, 2004 from <http://www.sei.cmu.edu/publications/documents/97.reports/97tr014/97tr014abstract.html> Weikipedia

[20] Young, W. (1970). A Procedural Approach to File Translation. ACM SIGFIDET Workshop on Data Description and Access, 1970.

[21] Zaniolo, C. (1979). Design of relational views over network schemas. Proceedings. ACM SIGMOD

[22] Bisbal, J., Lawless, D., Bing, W., "An Overview of Legacy Information System Migration", Proceeding of the 4th International Computer Science Conference and 4th Asia Pacific Software Engineering Conference 1999 Publisher IEEE