ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 1, January 2020

An Effective approach to Data Migration

Nidhi Jain, Sunita Kumari

Department of Computer Science, Kanya Gurukul Mahavidhyalaya, Dehradun, Uttarakhand, India

ABSTRACT: With the continuous change in technology and standards, business requirements are changing rapidly, which is one of the main reasons for migration. Data migration is a regular activity in IT organizations, but it is an activity that requires meticulous planning and execution with proper tools as failure of migrations can cost heavily to organizations and data loss or missing data can affect the organizations in a great deal. This paper focuses on data migration approaches based on an organization's specific business needs.

KEYWORDS: Data Migration, ETL, Legacy System, Migration Approach

I. INTRODUCTION

Data migration is a process where data is moved from one system or environment to another system or environment. Data migration projects are resource intensive and they require significant of IT budget for implementation. According to Gartner [1] "83% of data migration projects either fail or exceed their budgets and schedules". There are many risks involved with data migration and without proper planning and business support, the possibility of a data migration project's success is low. A successful data migration project is complex and challenging as it requires a team of people with required and adequate knowledge and a thorough understanding of the current system, its business process and various dependencies among the data sources. It is important that the New DB/ Target DB has all the required functionality to accommodate the source data, metadata and the features and the functionalities of the source system.

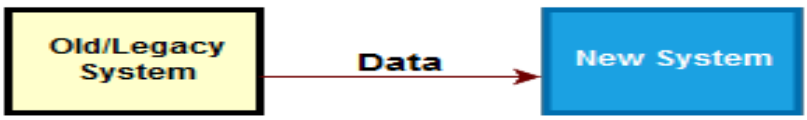

Figure 1: Data Migration

A successful data migration requires a strong quality assurance process where the migration process is being tested at every stage of the migration, before and after the migration to build the confidence that the data has been migrated as required and functionalities are intact and working as expected. It is essential to have the proper requirement specifications which describes the information to be migrated and how the new DB will be implemented.

II. NEED FOR DATA MIGRATION

The requirement of data migration arises whenever a new system is introduced, or a new system replaces an old system in an organization. Sometimes, a legacy system is replaced or additional new systems or deployed to support the business needs which results in data migration.The goal of any data migration is to improve performance and to provide competitive edge in an organization. Following are some of the reasons for data migration –

1. There can be a requirement to change the existing process due to the merging of business units or among organizations that results in data migration.
2. Software and applications require constant upgrades for better performance, scalability and efficiency. Theserequire data migrations to support the upgrade activities in an organization.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 1, January 2020

3. To provide better customer service and to be competitive in the market or to be compliant with the regulations, it may require data migrations.
4. Companies bear heavy operational costs to keep the systems up and running and as in effective cost saving measure, it is always recommended to focus on enhancing the efficiency by removing the bottlenecks in the software application through data migration to unify the data sources.

III. TYPES OF DATA MIGRATION

Data migration can fall under in one the following four categories

A. *Storage Migration:*

This kind of migration involves migration of physical blocks of data from the source hardware to the destination hardware.

B. *Database Migration:*

In this type of migration, a complete database is migrated from one vendor to another or complete upgradation of DB software takes place.

C. *Application Migration:*

This type of migration modernizes any environment by migrating a on premise application to the cloud or migration from one to another cloud platform.

D. *Business Process Migration:*

This kind of migration takes place when there are changes of business process at the event of merging or acquisition of other companies.

IV. AN APPROACH TO MIGRATION

In this paper, we are going to propose a methodology and the best practices to be followed to avoid the risks and challenges in the migration process.

Following are the key activities (Figure2) during a migration process:

V. SCOPING AND PLANNING

To accomplish a successful migration, it is very important to create a detailed plan is before starting of the actual data migration. The plan layouts the detailed process, its complexity its potential issues and concerns along with the timeline.

Scoping and planning can be divided in the following subprocesses:

A. *Requirement analysis:*

Activities such as business requirement gathering, and functional specification documentation are part of requirement analysis process. Scope of the projectcan be determined accurately by thorough review and understanding of the requirements. Requirements should be well captured and need to becarefully documented, with signoffs from stakeholders wherever required.

B. *Identify stakeholders/teams:*

In the next step, it is required is to identify the key partners and team members. The relevant stakeholder, functional and technical teams must be includedin order to eliminate the risk of gaps in the target system. To set the expectationsat an early stage of the project, thus setting expectations for all the stakeholders. To set the expectations for all the stakeholders, a RACI matrix can be used to delegate the roles and responsibilities at an early stage of the project

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 1, January 2020

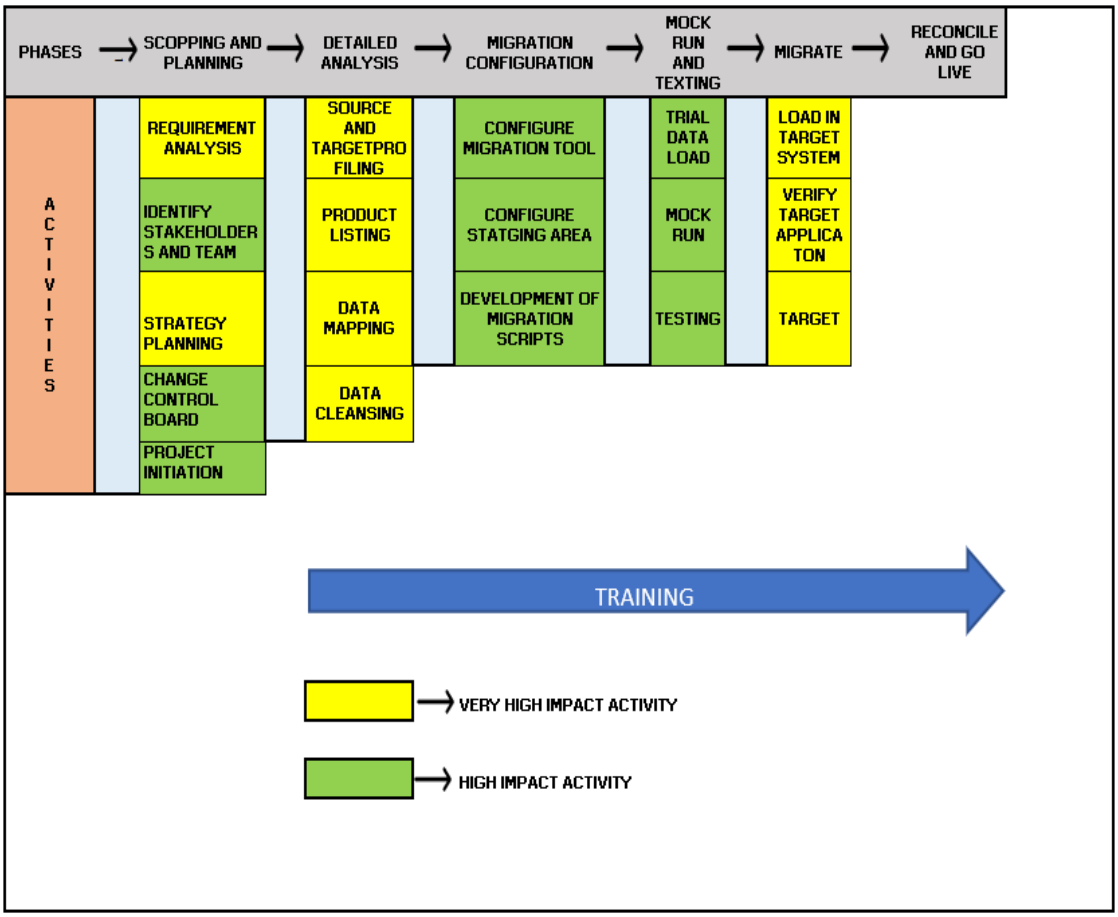

Figure 2: Migration Approach

C. *Strategy planning:*

The importance of stage planning is that it can help to identify the risks and issueswhich may occur later in the project. This must be addressed and planned up-front in the rollout strategy.

Following are the basic strategies to consider for a roll out plan

i) **Big Bang:**This type of migration is accomplished in a single operation which normally is taken care of in the weekends. This type of migration is preferred for low data volumes.

ii) **Phased:**In this type of migration, Data is moved to the target system in several. For new environment, records are created directly in the target system.

iii) **Parallel run:**In this type of data migration, transactions happen parallelly on both the source and target system until the migration completes entirely. Reconciliation is done at the end of each day until all the

iv) data is migrated.

D. *Change management board:*

This management board takes the initiatives to establish an effective change management process. The board must examine all the required changes and schedule them for implementation based on the analysis.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 1, January 2020

E. *Project schedule and initiation:*

After accumulating all the information and size of the project, it is required to implement a proper projectgovernance structure. Contingency can be handled effectively if there is an efficient project delivery structure. The structure must include factors such as project timelines, milestones and deliverables.

VI. DETAILED ANALYSIS

Detailed analysis of data is an important step as proper analysis of data makes the migration process smoother as duplicate and unwanted data can be removed based on these analyses and that makes the data migration process more efficient. Following are some of the activities which can ensure the sanctity of the data:

A. *Data profiling:*

Data profiling is the process of testing and analysingdata whichhelps to identify the quality of data where the profiled data determines theof inconsistencies, redundancies and inaccuracies. Data structure and relationships are discovered by profiling the source and the target data.

B. *Product listing:*

In this process mapping is done on the product between the new system to the legacy system. In some cases, the products are rationalized to fit the supported products in the new target system.

C. *Data mapping:*

It is the process of mapping the data elements from the source system to the target system. Data fields are checked to confirm data types, field length, integrity checks before the migration of data. Data mapping is an iterative process and mapping specification must be updated if there is any change in the design or rule in the system.

D. *Data cleansing:*

The main objective of data migration project is to migrate clean data to the target system. Detailed planning is needed here which requires cleansing legacy data. Data cleansing detects and fixes incorrect records in a data setand there provides a strong platform for the entire ETL process

VII. MIGRATION CONFIGURATION

A. *Configure migration tool:*

The data migration tool can be considered as a central repository to support storage and maintenance of data mapping. It is important to select a good ETL tool for smooth migration and which is capable of auto transfer of data and should be scalable and flexible. The tool should be able to be customized as per the target system.

B. *Configure staging area:*

Staging area is the section where extracted files are stored after transformation. This area acts as an intermediate section to store the results after ETL. It is used to store the intermediate results of the ETL process. The target tables should defineaccurately to the final database, which would ensure faster migration to the target system.

C. *Execute data conversion:*

Data is migrated from legacy system to the staging area and finally to the target system.Data conversion happens through ETL and as per the requirement of the target system.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 1, January 2020

VIII. MOCK RUNS & TESTING

A. *Trial data load:*

As a pilot effort, small samples of data are uploaded to ensure load sequence, determine duration and accuracy, and correct any loading errors. This is iterated till the load process completes successfully. Initially, to ensure the accuracy and to check load error, a small sample of data are migrated as a pilot effort. These trial data loads reaffirm that bulk loading of data with accuracy and any load errors.

B. *Mock run:*

This is a mock-up of the entire ETL process to check for any load errors during the process. In case of any error, it is reiterated with corrections till the process completes successfully. If no error is found, then the team can go for final migration.

C. *Testing:*

Testing is an important part for a successful data migration. It reduces the risks associated with migration and issues and errors can be identified which could be the causes of failure and delay.

Following are the phases ofTesting:

i. **Unit testing:**Unit testing comprises of scope verification, data mapping, target system requirements and tools to be used for migration. In unit testing, eachunit will be tested as part of the functional end-to-end strategy.
ii. **Post-migration testing:**This type of testing is conducted after the completion of migration. It involves testing the throughput of the migration process and reconciliation.
iii. **UAT:**User approval testing is the testing where migrated data in the target system is tested based on therequirement specifications

IX. MIGRATE

It is the final phase of the migration process where data is migrated into the target system. To execute the entire migration process, a cutover period is defined. This time window is the time period between the shutting down of the legacy system and the new target system going live. Data in the source legacy system is frozen and then extracted using an ETL tool during this period. Following is the process outlined below for a successful migration:

A. *Load on to target system:*

Data load scripts and migration tool are used to load data from staging area to the target system in sequential manager.

B. *Verify target application:*

Reconciliation checks are done to verify the accuracy of data once the data is loaded to the target system. If there are any discrepancies identified, roll back is requested to bring the system to original state.

C. *Target system implementation:*

At this phase, ETL process completes and target system becomes ready to go live. Business rules and Acceptance criteria are verified and KPIs are evaluated.

X. RECONCILE & GO-LIVE

This is the post-migration stage where the target system is ready for use. The source system needs to be brought to a logical accounting stage during the cutover period. During the migration process, no new transactions must be entered into the legacy system and data must be frozen. After completion of data migration, reconciliation checks need to be performed to confirm that there is no discrepancy before the system go live. Target system is monitored for any

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 1, January 2020

improvement once it goes live.Migration is conducted during a weekend, keeping a minimum impact on the business as it requires a downtime of the system and all the stakeholders are notified about this planned downtime.

XI. CONCLUSION

In today's dynamic environment, organizations spend billions of dollars on data migration and surprisingly 75% of the new systems fail to meet the standards. Data migrations are expensive and there are lot of hidden challenges and cost associated with it. Hence, data migrations need to be given the required attention and a proper migration approach is required for a successful migration and which could help to avoid pit falls such as over budget, exceed the allotted time or a complete failure.

REFERENCES

1. S.Sarmah "Data Migration", Scientific & Academic Publishing, 2018, [Online], Available: http://article.sapub.org/10.5923.j.scit.20180801.01.html
2. Zhang, G., Chiu, L., & Liu, L. (2010, July). Adaptive data migration in multi-tiered storage based cloud environment. In 2010 IEEE 3rd International Conference on Cloud Computing (pp. 148-155). IEEE.
3. Delsing, J., Eliasson, J., Kyusakov, R., Colombo, A. W., Jammes, F., Nessaether, J., ... & Diedrich, C. (2011, November). A migration approach towards a SOA-based next generation process control and monitoring. In IECON 2011-37th Annual Conference of the IEEE Industrial Electronics Society (pp. 4472-4477). IEEE.
4. Behm, A., Geppert, A., & Dittrich, K. R. (1997). On the migration of relational schemas and data to object-oriented database systems (pp. 13-33). Universität Zürich. InstitutfürInformatik.
5. King, A., & Chiu, D. C. (2003). U.S. Patent No. 6,530,004. Washington, DC: U.S. Patent and Trademark Office.
6. O'conor, G. W. (2008). U.S. Patent No. 7,457,807. Washington, DC: U.S. Patent and Trademark Office.
7. Kenley, G., Ericson, G., Fortier, R., Holland, C., Mastors, R., Pownell, J., ... & Webber, N. (1994). U.S. Patent No. 5,276,867. Washington, DC: U.S. Patent and Trademark Office.
8. Lu, C., Alvarez, G. A., & Wilkes, J. (2002, January). Aqueduct: Online Data Migration with Performance Guarantees. In FAST (Vol. 2, p. 21).
9. Morris, J. (2012). Practical data migration. BCS, The Chartered Institute.
10. Davis, A. B. A., & Mansell, P. (2003). U.S. Patent No. 6,546,404. Washington, DC: U.S. Patent and Trademark Office.
11. Dasgupta, K., Jain, R., Sharma, U., & Verma, A. (2006). U.S. Patent Application No. 11/011,861.
12. Spivak, D. I. (2012). Functorial data migration. Information and Computation, 217, 31-51.