

General Aspect of (Big) Data Migration Methodologies

V.Rathika

Assistant Professor in Computer Applications
Idhaya College for Women, Kumbakonam
Tamil Nadu, India

Dr.L.Arockiam

Associate Professor in Computer Science
St. Joseph's College, Trichy
Tamil Nadu, India

ABSTRACT

The process of moving vast amount of data from one place to another is called big data migration. This is bundled with number of issues and challenges. It is important when upgrade and relocation of existing systems. It has unique tools, techniques, architectures, algorithms and methodologies. Methodology explains the steps which are followed to do successful migration from source to destination. Businesses are creating significant data management challenges by increasing volumes of data. They should be able to access and organize volumes of data stored in a variety formats. This paper is going to provide important and common aspects of some methodologies.

General Terms

Big Data, Migration, Methodology

Keywords

Big Data, Migration, Methodology, General aspects.

1. INTRODUCTION

Big data migration is happening when user needs to move large quantity of data from source system to destination system for updating the techniques. It has the proper design, data extraction, data loading, data verification and data cleansing to accomplish the process [3]. Methodologies explain the parts of migration and its significance. Each and every step or part is unique in the entire process. It has to be completed based on the hierarchy. A good methodology can achieve the expected migration in short time even data is huge.

2. BIG DATA

Structured, semi structured and unstructured data are collectively called big data. Social Medias are the origin of big data. Every day data volume is increasing in constant. It has 5 characteristics such as Volume (vast amount), Velocity (speed), Variety (different types), Veracity (quality) and Value. Medicine, biology, physics, healthcare monitoring, astrometry, transportation and manufacturing are some examples of big data applications.

3. METHODOLOGIES

A methodology is an organized set of methods to provide solution to a problem. It explains the entire process in modules. It gives overview of the process also. This paper is going to present some methodologies for successful data migration.

3.1 First Methodology

This methodology has eight steps like business impact analysis, discovery, mapping & design, creation of migration plan, provisioning, pre migration test, migration & cutover and migration validation.

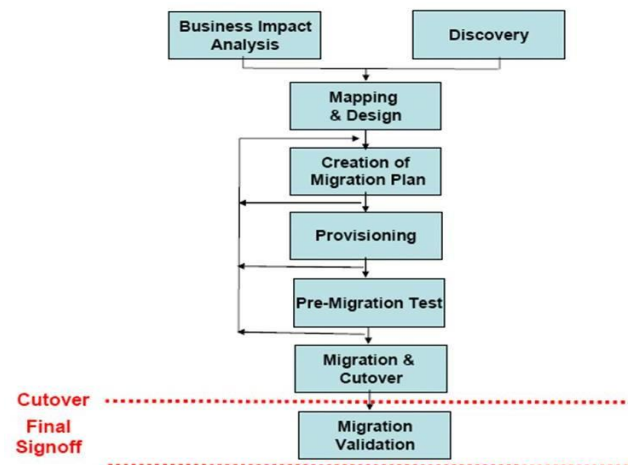


Figure 1 - First Methodology

3.1.1 Business Impact Analysis

It identifies the business and operational requirements that impacts the migration process.

3.1.2 Discovery

Details about the migration hardware and software environment are collected during the discovery step. It identifies what data is going to be moved.

3.1.3 Mapping and Design

It identifies where the data is going to be moved.

3.1.4 Creation of Migration Plan

It functions as the blueprint for the migration implementation. It is specifying customer expectations. It is defining project deliverables.

3.1.5 Provisioning

It prepares the destination storage environment for the data move.

3.1.6 Pre-Migration Testing

Before any data is moved, it is important that some portion of the migration be tested and validated.

3.1.6 Migration and Cut over

Data are moved from source to destination. After completion the entire process switches to target.

3.1.7 Migration Validation

It validates the post migration environment and confirms that all expectations have been met prior to committing.

3.2 Second Methodology

This methodology was introduced by Seagate Recovery Services in 2013. It consists of evaluate, review, restore and migrate. It ensures that data is properly evaluated, reviewed and restored before it is migrated to new, more accessible and cost effective media. It assures that data is safeguarded from beginning to end [4].

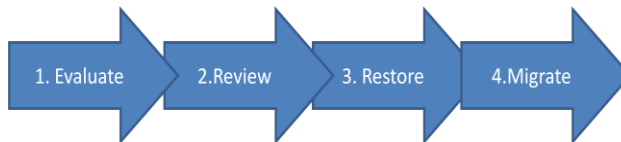


Figure 2 - Second Methodology

3.2.1 Evaluate

It evaluates media (source) type, state of media and estimated effort.

3.2.2 Review

It evaluates customer requirements and criteria. It determines the targeted media also.

3.2.3 Restore

It identifies and restores individual files. It recovers data before the data extraction process.

3.2.4 Migrate

It indexes and de-duplicates the restored data. It removes unnecessary files to reduce storage requirements. It migrates data to new media which is delivered to customer.

3.3 Third Methodology

This was introduced by DB Best. It was designed specially to handle database migrations. It has the methods such as migration readiness stage, migration stage, ETL stage and application stage.

3.3.1 Migration Readiness Stage

It plans the infrastructure includes hardware and software. It evaluates risk, creates business case and assess database environment. It defines engagement model also.

3.3.2 Migration Stage

It analyzes existing database and design SQL Server databases. It builds prioritized migration roadmap based on business objectives. It moves database objects such as tables, views, triggers. It transfers data, verifies database schema and validates that data was migrated correctly.

3.3.3 ETL Stage

It designs and develops ETL packages to handle parallel data load. It test data load performance. It validates technical accuracy, functionality and security.

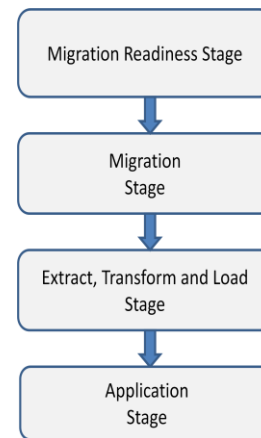


Figure 3 - Methodology

Third

3.3.4 Application Stage

It performs integration testing with the applications to ensure the correctness. It does user acceptance testing also. It implements production environment.

3.4 Fourth Methodology

This was introduced by Gershon Pick. It has also four steps.

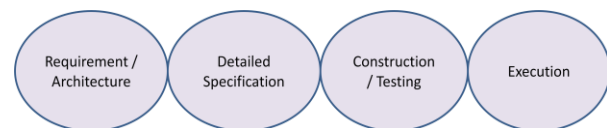


Figure 4 - Fourth Methodology

3.4.1 Requirement / Architecture

It defines high level requirements such as data, performance requirements, and detailed project plan.

3.4.2 Detailed Specification

It defines validations, transformations and structure changes.

3.4.3 Construction / Testing

It builds the migration solution and tests it works.

3.4.4 Execution

Solution is executed and results are verified. It ensures target system brought up in production.

3.5 Fifth Methodology

This was introduced by Dell. It has five steps as planning, pre – production testing, migration, validation and cutover.

3.5.1 Planning

It defines migration goal and requirement. It defines HW / SW and Tools. It creates Migration plan.



Figure 5 - Fifth Methodology

3.5.2 Pre – Production Testing

It creates testing migration environment. It validates HW / SW and Migration Tools. It collects and validates data. It updates final migration plan.

3.5.3 Migration

It installs migration software. It Performs migration based on plan.

3.5.4 Validation

It collects migration statistics. It verifies the completion of the migration. It creates migration report.

3.5.5 Cutover

It moves application to the target storage environment. It creates final report.

3.6 Sixth Methodology

This methodology has eight steps. It is general methodology which explains the common and necessary steps are involved in the migration process.

3.6.1 Analysis of business impact

Data is the basic and important term of each and every business. It should be analyzed to enhance the business process. So, this step analyzes customer's requirements.

3.6.2 Information gathering

It collects the details about software and hardware. It finds source and destination systems. It has two methods to collect information are manual and automatic.

3.6.3 Mapping, Designing

It maps source and destination systems. It has two types of design layouts. One to one layout explains source and destination layouts are same. Relay out layout explains source and destination layouts are different.



Figure 6 - Sixth Methodology

3.6.4 Plan of migration

This is the plan of all information like business, operational constraints to migrate, data and its attributes, tools to be used during the process.

3.6.5 Provisioning

It copies former structure of files, data volumes, and attributes so that the environment could be ready for receiving the actual streams of data.

3.6.6 Test Migration

It checks all the presumptions are correct and tools are selected properly. It minimizes the risk of losses such as time and money wastes.

3.6.7 Migration

It determines the moving of data from source to destination. It has two possibilities that moving data out of path and in the path.

3.6.8 Validation

After all the process it has to check everything went in an ordered way. It checks about hidden errors as

well as data access, file permissions, structure of directories and the work of applications.

4. GENERAL ASPECTS

Migration is the process to move data from one environment to another. It has the above mentioned methodologies to do successful moving. But all the methodologies are having some common steps such as requirements, planning, migration and validation. So who ever is willing to develop a methodology to migrate data they have to concentrate on these steps must. Requirements refer the customer needs only. Planning can create the entire plan for migration such as new and old environments, tools and technologies and all. Migration is the stage data is moving from existing system to new system. Validation checks the entire process and the associated risks of migration. Finally these steps ensure the successful migration.

5. CONCLUSION

Migration is important factor in business organizations. They have to update and introduce new systems to meet their competitors and achieve their goal also. Each and every day new concepts are coming to enhance the existing systems. So it ensures migration is the greatest and unavoidable event in the businesses. Migration has verity of methodologies, tools and technologies, frameworks and architectures. This paper is provided the common features of methodology to consider and follow while development of new methodologies for data migration.

16. REFERENCES

- [1] Anuradha Bhatia and Gaurav Vaswani. 2013. Big Data – An Overview. International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 8.
- [2] Klaus Haller.2008. Data Migration project management and standard software experiences in analog implementation projects. Proceedings of DW2008 Conference.
- [3] C.Bizer, P.Bonez, M.L.Bordie and O.Erling. 2011. The meaningful use of Big Data : Four Perspective – Four Challenges. SIGMOD, Vol. 40, No.4.
- [4] Philip Russom. 2006. Best Practices in Data Migration. TDWI (Informatica).
- [5] Seagate Recovery Services white paper, “What is Data Migration”, 2013.
- [6] Cisco Data Center. 2011. Cisco Data Center Migration Service. Cisco Public Information.
- [7] Klaus Haller. 2009. Towards the industrialization of data migration: concepts and patterns for standard software implementation projects. Proceedings of CAISE.
- [8] Ch.Sai Krishna Manohar. 2013. A Greener Approach to Cloud Computing using Virtual Migration. International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 8.
- [9] Dylan Jones. 2009. Some Common Data Migration risks (and how to avoid them). Data Migration.
- [10] Atul Srivastav. 2010. Data Migration - Case Study. www.enterprise data migration.com.

- [11] Bing Wu, Deirdre Lawless, Jesus Bisbal, Jane Grimson. 1997. Lagacy System Migration: A Legacy Data Migration Engine. 17th International Database Conference.
- [12] Gershon Pick, “Data Migration Concepts and Challenges”, White Paper, 2001.
- [13] Martin Wagner, Tim Wellhausen. 2011. Patterns for Data Migration Projects. www.TNGTECH.com.
- [14]A Dell Technical’s White paper, “Methodologies for Data Migration to Dell Fluid Architecture”.
- [15] An Oracle’s White Paper. 2011. Successful Data Migration. www.oracle.com.
- [16] Kam Woods, Geoffrey Brown. 2008. Migration performance for legacy Data Access. The International Journal of Digital Curation, Issue 2, Volume 3.
- [17] Nasuni’s White Paper. 2012. Bulk Data Migration in the Cloud. www.nasuni.com.