IJASC 21-2-24

A Design of the Improved Data Conversion Process for System Upgrade Project

[1]Hee Wan Kim

[1]*Professor, Division of Computer Science & Engineering, Sahmyook Univ., Korea*
E-mail hwkim@syu.ac.kr

Abstract

Data conversion refers to the process of extracting the data existing in the existing system, that is, the past data accumulated by the old information system or other methods and transferring it to the improved table of the new system. The person in charge of data conversion refers to the entire process of converting to the final destination table according to the rules designed/planned in advance. In most cases, data conversion design should be consider when the old system replace or the data of another existing system is converted and applied to a newly constructed information system. The goal of data conversion is to understand the current database system of operating environment, understand the characteristics of the DBMS in use, maintain the optimal database structure, and make the new system perform at its best. Data conversion methods are largely divide into a method using a tool and a conversion method using a program preparation. In this paper, we examine the advantages and disadvantages of the data conversion method, and try to derive the problems of the existing data conversion method. Based on this, an improved data conversion method for the system upgrade project was proposed, and verified through a questionnaire of an IT expert to prove its effectiveness

Keywords: data conversion, information system, data conversion method, database system, system upgrade project

1. Introduction

Data conversion is the process of changing the format, structure, or values of data. For data analytics projects, data may be transform at two stages of the data pipeline. Organizations that use on-premises data warehouses generally use an ETL (extract, transform, load) process, in which data conversion is the middle step. Today, most organizations use cloud-based data warehouses, which can scale compute and storage resources with latency measured in seconds or minutes. The scalability of the cloud platform lets organizations skip preload conversions and load raw data into the data warehouse, then transform it at query time — a model called ELT (extract, load, transform).

Data conversion refers to the process of extracting the data existing in the existing system, that is, the past

data accumulated by the old information system or other methods, and moving it to the improved table of the new system (ex. next generation system, new system, etc.). It can easily said that it is similar to the process of moving from the old, cramped house where you lived to the new spacious house. The person in charge of data conversion refers to the entire process of converting to the final target table according to the previously designed/planned rules (Mapping, Conversion Rule) [1][2].

In addition, the goal of data conversion is to understand the operating environment of the current database system, understand the characteristics of the DBMS in use, maintain the optimal database structure, and make the new system perform at its best. In spite of this importance, specific studies on effective methods of performing data conversion and integration processes or data verification methods are insufficient [3]. Therefore, in this paper, we propose a concrete method for efficient data conversion and integration process for the system upgrade project.

2. Related Work

2.1 Importance of data conversion

If the task be built in the SI project is a completely new task that has never applied information system before, or if there is no need to utilize existing computerized data, the data conversion design step can be omitted. However, in most cases, data conversion design should be consider when the old system needs to replace or the data of other existing systems must be converted and applied to a newly built information system.

Depending on the project, data conversion may occupy a nucleus in the work of the entire project. In particular, data conversion needs to be carefully considered because the amount of data held for a long period of time, such as the replacement of the new information system of the financial information system or the administrative information system of a state agency, is enormous and the importance of the data is very high [4].

2.2. Kind of data conversion project

When the project is in progress, the data conversion stage must carry out, and there are projects related to data conversion as following <table 1>.

Table 1. Types of data conversion projects [3]

Type of Project	Contents
DBMS version-up	1:1 data conversion due to DBMS version up
Heterogeneous DBMS change	1:1 data conversion due to heterogeneous DBMS change
Hardware change	1:1 data conversion due to data center movement or hardware change
Improvement of Data model	Data conversion by changing data model structure compared to AS-IS like next-generation SI project

Data conversion due to DBMS version up/DBMS heterogeneous change/hardware change is converted to AS-IS table structure, so conversion mapping design is simple. However, if the data model needs to be changed significantly copared to the AS-IS due to the improvement of the data model, not only the case of a simple AS-IS:TO-BE table of 1:1, but also conversion to 1:N. In this case, not only conversion implementation but

also conversion mapping design and conversion verification tasks must be included in the project WBS. Therefore, in this study, we would like to suggest a more efficient way to convert data.

2.3 Data conversion methods

Data conversion methods are divide into a method using a tool and a conversion method using a program (SQL, Java, C#, etc.). In the method using the tool, the implementation logic can be easily implement as a top-down method in the integrated GUI environment where the platform is independent of the data format. In addition, there is an advantage in that the total development time and required man power can be shortened by utilizing the automatic debugging function. By continuously classifying source data errors by type and managing them systematically, it is possible to eliminate the cause of errors in the data or applications of the source system. In addition, when a data update error occurs, execution can be stop and the error can be restart after correcting the error. However, it is possible to respond immediately. The implementation program should be automatically play. It is possible to design to prevent data inconsistency by implementing a function to monitor data update status at all times. However, when data inconsistency occurs, the record of inconsistency occurrence and its cause must be provide through the monitoring log.

The method through program creation takes a lot of manpower and time to develop and debug implementation programs according to each platform and data format, and a lot of effort is required for testing and verifying results. Since there is no data verification system through monitoring, efforts to improve functions other than the verification standards implemented by the initial program logic are required. Since the batch processing method is adopt, it is difficult to immediately respond to errors that occur during data update, and when the requirements of the source data and target DB change, it is necessary to identify the related data implementation program and change the logic manually. In addition, in order to check whether the update to the target DB is correct, it is necessary to periodically check both data values by metabolizing them, and it is difficult to find the cause when the data is inconsistent [5].

3. Problems with the existing data conversion method

Existing traditional mapping design work uses Excel, which manages table mapping, column mapping, and code mapping with a single Excel. When the conversion designer writes a mapping definition using Excel, the conversion developer writes a conversion program based on the mapping definition. As for the working method using Excel, the problem of the mapping definition that changes according to model change and mapping change is 1) change management, 2) difficulty in aggregation of mapping definitions, 3) omission of parameters, and 4) unstandardized conversion logic technology [6].

<Table 2> is a survey of 30 database experts and 16 IT experts with 10 to 20 years of IT experience on the methodology performed when converting general data that implements SQL based on mapping design and mapping design using Excel documents [5][6].

Table 2. Problems of the existing data conversion method

No	Issues	Remarks
1	Documents subject to management due to document-oriented	
2	High probability of errors such as consistency between documents and input mistakes	
3	Mapping (conversion) rule expression method that is not unified	
4	Inconsistency between mapping design (Excel) and conversion program (SQL)	
5	Implementation of non-standard transition programs	
6	Difficulties in aggregate management such as mapping design progress	
7	Difficulty in detecting change lines in mapping design due to DB structure change	
8	Difficulty in creating a conversion design that meets both TO-BE and AS-IS	
9	Possibility of missing parameters for data conversion	
10	Difficulty in managing mapping design history	
11	Efforts to write verification programs and possibility of omission of verification implementation	

The problem was as documents subject to management due to document-oriented, high probability of errors such as consistency between documents and input mistakes, mapping (conversion) rule expression method that is not unify, and inconsistency between Excel mapping design and conversion SQL program.

4. Proposal of data conversion plan

4.1 Data conversion execution, validation process

The following is a detailed process derived by tailoring the execution/verification stage of the data conversion process for each SW development methodology and further subdividing it. In the data loading/pre-work stage, after loading the AS-IS data to the conversion staging DB Server, decrypt the encrypted column before conversion and check the index for character set/conversion performance. In the data pre-verification stage, if you want to adopt a new KEY from TO-BE or change the code, it verifies whether the AS-IS and TO-BE KEY are covered. When KEY/CODE mapping coverage cannot be secure, data conversion must perform after mapping data correction before data conversion is perform. In the conversion program pre-check phase, it is checked whether there is an unmapped column, and the table conversion method is verified when there is no mapping rule or conversion rule even though TO-BE is a NOT NULL column. The entire program regenerate to prevent the case where the conversion program is not regenerate after the conversion design is change. Since the conversion program may be invalid due to DB object change, the entire program is recompile and corrected to maintain valid status. And, if the new PK of the TO-BE table is an artificial key, SEQUENCE initialization is performed before starting during conversion programming. The data conversion program deletes all data because there may be data in the conversion table in the TO-BE table. In case of pre-conversion due to table size, change of previously converted data should be check at the time of main conversion. For the conversion result, the conversion execution time required, success status, and error message when an error occurs should be included. In the verification step after data conversion, the duplication of newly acquired keys should be check and the uniqueness of the replacement identifier that is not use as PK such as resident

registration number/business number/corporation number but must be unique should be checked. Data referential integrity between parent and child should be verify, and the entire raw data should be compare and verify. AS-IS vs. using aggregate functions such as COUNT/SUM/MIN/MAX. TO-BE data must be compare, and whether it conforms to the domain rule of the date/amount type column must be verify. Finally, after saving the Biz Rule in the Biz Rule Repository, converting the data and performing batch verification.

4.2 The whole process of data conversion

The following < table 3> shows the data conversion process tailored based on the data conversion process for each SW development methodology. Tasks to be perform during data conversion analysis, design, implementation and testing, pre-conversion/pre-open/main conversion, and implementation on the day of opening and the required products, and the precedence and successor relationships between each task were derived. The following is the main tasks of the process and the output of the work that must be perform at each stage during data conversion [1][4][7][8][9][10].

Table 3. Data conversion process and the output

Job Phase	Job Name	Job Details	Outputs
Plan	Data Plan	Data conversion schedule / conversion method / verification method	Data conversion plan
Analysis	Data conversion environment analysis and definition	Data conversion environment analysis and conversion date, range, and data loading plan	
	AS-IS data analysis	When switching to a new system, the data type classified so that there is no missing data.	
Design	Table mapping Design	AS-IS vs. TO-BE table mapping design	Table mapping definition
	Column mapping design	AS-IS vs.TO-BE column mapping design	Column mapping definition
	Data conversion environment construction	AS-IS DB staging server / TO-BE conversion system / AS-IS data loading plan	Data conversion architecture diagram
Implementation	Conversion program development	Developing conversion program for each TO-BE table using SQL	Conversion SQL
	Conversion program turning	Transition Program SQL Tuning to improve Performance	
Transformation	Conversion consistency verification	Verifies the consistency during conversion.	Conversion verification Report
	Conversion result report	Report results on conversion.	Data conversion report

4.3 Verification of data conversion project suitability

To verify the suitability of the entire data conversion process, a survey conduct with 32 IT-related experts, and the results analyze. Each check item has a full score of 5, indicating a score of 4.0 or higher for all items. Most of the items related to planning/analysis/design/implementation/transition performance showed a high distribution of 4.3 to 4.7. Therefore, the results of the experts' survey on the entire data conversion process find to be suitable for the proposed process as following <table 4>.

Table 4. Suitability of data conversion process and output

Job Phase	Job Name	Job Details	Results	Standard Deviation
Plan	Data Plan	Is the schedule / conversion method / verification method well written in data conversion plan?	4.44	0.178
Analysis	Data conversion environment analysis and definition	Has the data conversion environment been analyze normally?	4.63	0.180
	AS-IS data analysis	Have you identified the table and method of conversion to AS-IS?	4.63	0.180
Design	Table mapping Design	Are there any missing TO-BE table mapping designs for AS-IS conversion target tables?	4.63	0.180
	Column mapping design	Are there any missing column mappings for the conversion table?	4.63	0.180
	Data conversion environment Construction	Does the data conversion architecture design by reflecting system environment?	4.63	0.180
Implementation	Conversion program development	Are there any missing conversion programs for TO-BE table and column mapping design?	4.38	0.179
	Conversion program turning	Has the conversion program been implemented by mapping design?	4.63	0.180
Transformation	Conversion consistency verification	Have you performed the Nth conversion rehearsal, and have the results of the conversion drawn?	4.63	0.180
	Conversion result report	Are required items included in the conversion result without omission?	4.63	0.180

5. CONCLUSION

Data is more important today than ever before. This reality coincides with the rapidly developing IT flows such as big data, the 4th revolution, public data, and data analysis, and corporate data is soon identified as an asset. As the operating system that operates these data develops technology, business changes, customer needs, and IT technology day by day, information system upgrade projects continue to occur periodically. In this process, the demand from users to convert data, which is an asset of the company, without any problem in consistency, is increasing. In spite of the importance of data conversion, productivity and efficiency decrease

due to the existing Excel document mapping design and separate SQL implementation, high possibility of human error, and a lot of effort must put into it. There is a way to use a famous ETL solution that can replace the data conversion solution, but it is necessary to know how to use it because the original purpose is not for data conversion for next-generation projects, but for building Data Warehouse and DataMart. In addition, there are disadvantages such as additional time input and high cost due to the fact the process of using the solution is not optimize for data conversion. Therefore, in this paper, an improved data conversion method for the system upgrade project proposed by improving the disadvantages of the data conversion method, and to prove its effectiveness, it was verified through the questionnaire of IT experts.

REFERENCES

[1] Korea Database Agency, Data Industry Status Survey Results Report, Seoul: Korea Database Agency, 2015

[2] Data Streams, Data migration, data migration issues, http://datastreams.co.kr/consulting-service/data-migration/

[3] Shin, S.H.(2014), Design and consistency verification of science and technology information content data migration model, Korea Institute of Science and Technology Information

[4] C.H. Jung, Way to the SI Projects Experts, Youngjin.com, 2010

[5] Encore, Migration procedure and method for data integration, Korea Database Agency, 2009

[6] Korea Database Agency, Data Industry White Paper, Seoul: Korea Database Agency, 2017

[7] National Information Society Agency, CBD SW Development Standard Output Management Guide, Seoul: National Information Society Agency, 2011

[8] Ministry of National Defense, Defense CBD Methodology V2.0 User Guide Volume 2, Ministry of National Defense, 2010

[9] Korean Intellectual Property Office, Korean Intellectual Property Office Software Development Methodology, Seoul: Korean Intellectual Property Office, 2014

[10] J. K. Nho, The role of DA for a successful build the next generation of systems, Encore, 2014 http://www.dator.co.kr/encore/textyle/1542745