AN EFFICIENT AND SIMPLE TECHNIQUE OF DATA MIGRATION

Farid Ahmad*

Alighazi Siddiqui**

Mohammad Shabbir Alam**

Abstract

Every type of system may replace or enhance the functionality currently delivered by legacy systems to new system, regardless of the type of project/application; some data conversion may take place. Difficulties arise when we take the information currently maintained by the legacy system and transform it to fit into the new system. We refer to this process as data migration. Data migration is a common element among most system implementations. It can be performed once, as with a legacy system redesign, or may be an ongoing process as in storage of historical data in the form of a data warehouse. Some legacy system migrations require ongoing data conversion if the incoming data requires continuous cleansing. It should be that any two systems that maintain the same sort of data must be doing very similar things and, therefore, should map from one to another with ease. Legacy systems have historically proven to be far too lenient with respect to enforcing integrity at the atomic level of data. Another common problem has to do with the theoretical design differences between hierarchical and relational systems. In data migration one method apply in twice (i.e. automated and manual). This paper explores the steps to migrate date in form of manual, i.e. process of data migration without the help of any special tool those made for data migration. 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 in correct and efficient form.

Keywords: Legacy data, legacy system, data cleansing, data migration, source structure, target structure, field mapping.

* Research Scholar, CMJ University, Shillong, Meghalaya-793 003.

** Lecturer,College of Computer Science & Information Systems, Jazan University, Jazan, KSA.

Introduction:

The transfer of data from one format to another format or from one device to another device or from one system (Legacy System) to another system (New System) may be termed as Data migration. Data migration process comprises both the Data Conversion and the Data Take-On. In other words Data migration is the process of transferring data between storage types, formats, or computer systems. Data migration is usually performed programmatically to achieve an automated migration, freeing up human resources from tedious tasks. This is the process of importing legacy data to a new system. This can involve entering the data manually, moving disk files from one folder (or computer) to another, database insert queries, developing custom software, or other methods. The specific method used for any particular system depends fully on the systems involved and the nature and state of the data being migrated. It is required when organizations or individuals change computer systems or upgrade to new systems, or when systems merge (such as when the organizations that use them undergo a merger or takeover).

To achieve an effective data migration procedure, data on the old system is mapped to the new system providing 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.

Legacy data is the recorded information that exists in your current storage system. Legacy data can include database records, spreadsheets, text files, scanned images and paper documents. Data cleansing is the process of preparing legacy data for migration to a new system. Because the updated systems have different architecture and storage method, legacy data often does not meet the criteria set by the new system, and must be modified prior to migration.

Data Migration Steps:

Data migration completion steps are described below. Before initiating the process of data migration we require an extra database server for the process, it will contain the three schemas with required characteristics. One schema will be used to contain the legacy data, second to contain intermediate data and third for migrated data.

1. Analyze and define source structure:

The first step of data migration initiation is to analyze the structure of existing data in the legacy system in every object. It is not necessary that legacy system have data in correct and required format as you can expect. It is required to know usage of data at field level because some data may not be useful for new system. So that data is not required to migrate. Analyzing the source data is first and main step of data migration, so it should be completely concentrated upon in an efficient manner because correct data migration will be start from here and it will impact new system data.

This step or phase may take more time than expected because of the requirement of full analysis and understanding of the legacy system (if you are not aware or known about the legacy system).

Every system has some master tables, so it is required to analyze and define master and important tables.

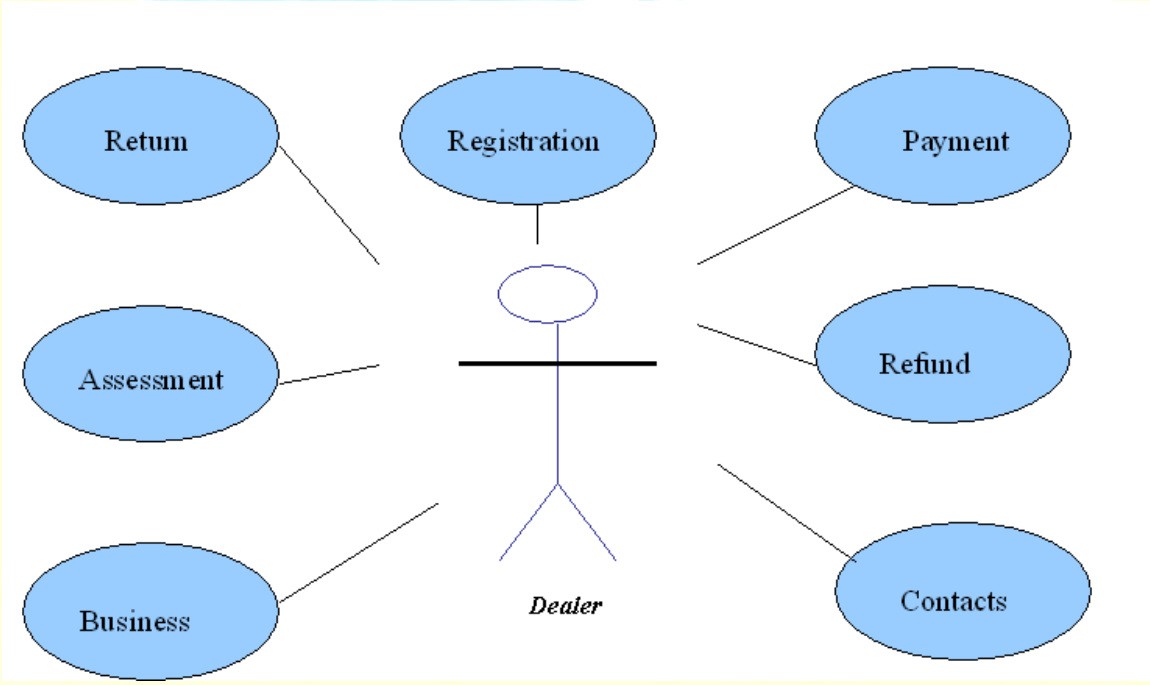

Figure: 1

The starting point for gathering information is in the existing documentation for each system. This documentation could take the form of the original specifications for the application, as well as the systems design and documentation produced once the application was completed. Often this information will be missing or incomplete with legacy applications, because there may be

some time lag between when the application was first developed and now. You may also find crucial information in other forms of documentation, including guides, manuals, tutorials, and training materials that end-users may have used. Most often this type of material will provide background information on the functionality exposed to end-users but may not provide details of how the underlying processes work.

As shown in the figure-1, diagram of a dealer linked with a taxation department is running his/her business and concerned respective modules. This is the example of legacy system i.e. it contains some module and every module has its own properties and description. So assume that the legacy system has a master table t_dealer. So it required to describe the table fields with data types and assign appropriate remarks to know the usability of database table and its columns as shown in below figure.

TABLE: T_DEALER		
FIELD NAME	*DATA TYPE*	*REMARKS*
TIN_SLNO	NUMBER(8)	Sr. no. of TIN
EDR_DATE	DATE	Effective date of Registration
TIN_VAT	NUMBER(11)	TIN of VAT Act
TIN_CST	NUMBER(11)	TIN of CST Act
TIN_ET	NUMBER(11)	TIN of ET Act
PROPRITER_NAME	VARCHAR2(100)	Name of Propriter
CIR_CODE	VARCHAR2(3)	Code of concerning Circle
BUSINESS_NAME	VARCHAR2(50)	Name of Business
ADD_HS_NO	VARCHAR2(15)	House number
ADD_MARKET	VARCHAR2(15)	Market name
ADD_LOCALITY	VARCHAR2(50)	Name of Locality
ADD_PO	VARCHAR2(15)	Name of Post Office
ADD_PS	VARCHAR2(15)	Name of police station
ADD_DISTRICT	NUMBER(2)	Name of district
ADD_PIN_CODE	NUMBER(6)	Pin Code
ADD_PHONE1	VARCHAR2(12)	Phone no. 1
ADD_PHONE2	VARCHAR2(12)	Phone no. 2
ADD_MOBILE	VARCHAR2(12)	Mobile no.
ADD_FAX	VARCHAR2(12)	Fax no,
ADD_EMAIL	VARCHAR2(30)	Email id
D_TYPE	VARCHAR2(3)	Type of dealer
LEGAL_STATUS	VARCHAR2(16)	Status of business
B_NATURE	VARCHAR2(16)	Nature of business
DATE_LIABILITY	DATE	Date of Liablity
PAN	CHAR(10)	PAN

Figure: 2

2. Analyze and define target structure:

Analyze and define target structure phase must be concentrated on data volume and data value of legacy system, after calculating both need to define the target structure. In the old system their might exist some discrepancies between application and database structure. So the target system must be in correct form and should be error free. Target system should be free from garbage objects. The important task to design the database for new system in respect of legacy system because the present application should run hassle free on this system. The analysis and define phase of target database in data migration should be scheduled to occur concurrently with the analysis phase of the legacy system. In some cases the same are expected to perform both analyses. This can be done; but this needs a clearly defined set of tasks for which they are responsible.

The main purpose of the analysis phase in data migration is to identify the target structure that legacy system data must be transported into the new system. Notice I stated data sources. Data sources are not limited to actual data processing systems, either. Inevitably, one will find the person that maintains files on their own workstations that they use to accomplish tasks that cannot be performed by their existing systems. Word processing documents, spreadsheets, desktop database packages and raw text files are just a few examples of data sources you can expect to uncover in the analysis phase. The next important part of the analysis phase involves getting acquainted with the actual data you have plan to migrate. In order to get a better sense, it is helpful to obtain reports that can provide row and column counts, and other statistics pertaining to your source data. This kind of information gives a rough idea of just how much data there is to migrate. The most common and simple data migration is to migrate the source data to the new system into data structures that are constructed identically to that of the source system but it happens rare cases. The design phase of data migration also happens in parallel with the analysis phase of the core project. This is done because each data element identified as a candidate for migration inevitably results in a change to the data model. The design phase is not intended to thoroughly identify the transformation rules by which historical data will be massaged into the new system; rather, it is essentially the act of making a checklist of the legacy data objects that we know should be migrated.

3. Perform field Mapping:

Perform field mapping means mapping between the source and target structure with data cleansing, if necessary. Data mapping maps data elements from the source to the destination and captures any transformation that must occur. Data element to data element mapping is frequently complicated by complex transformations that require one-to-many and many-to-one transformation rules. Data mapping is the process of creating data element mappings between two distinct data models. It is used as first step for a wide variety of data integration tasks. When merging several input datasets into a single output dataset, the field structure and contents are a consideration. Each input dataset will contain fields that also exist in other input datasets, as well as fields that are unique to only that dataset. Management of these fields determines the field structure and content in the output dataset. The field mapping allows you to define this output dataset field structure. Now that the source and target structures are defined, the mapping from the legacy to the target should fall into place fairly easily.

New System (Field Name/Default value)	Legacy System (Field Name/Default value)	Comments
Dealer_Master (Table Name)	**T_DEALER (Table Name)**	
ACK_NO,		To be auto generated
TIN_GRN,		To be auto generated
OWNER_FNAME,	PROPRIETER_NAME	
OWNER_MNAME,	PROPRIETER_NAME	
FIRM_NAME, OCCPNCY_STATUS,	BUSINESS_NAME **NULL,**	
STATUS_CD,	**CODE**	select code from lookup_code where description=business_stat
APPLN_RCPT_DT,	EDR_DATE	
RC_ISSD_DT,	EDR_DATE	
RC_EFF_DT,	EDR_DATE	
RC_RCPT_RA_DT,	**NULL,**	
RC_REFF_DT,	**NULL,**	
REG_STATUS,	**'REGD'**	
CREATED_BY,	**'VAT_MIG',**	
CREATED_DT,	**SYSDATE,**	
DEALER_TYPE,	**'BENT'**	
LOCATION_CD,	CIR_CODE	Circle code select from MIG_CIRCLE_MAP where CIR_CODE =Old_System
DEALER_ID,		
DLM_CATEGORY,	**'Q',**	
RC_ISSUE_FLAG,	**'N',**	
PA_REMARKS,	**'Y',**	
RA_REMARKS,	**'Y',**	
PA_RECOMMEND_REJECT_YN, Branch_count	**NULL,** **0**	

Figure: 3

Above mentioned figure shows the mapping of a table between legacy and new system, if there is need to generate data by any created procedure or sequence or other source table of legacy system then it is mentioned in column of comments (Bold face indicate Default values or taken assumption into above mentioned table). Mapping should include documentation that specifically identifies fields from the legacy system mapped to fields in the target system and any necessary conversion or cleansing. This allows in data migration to lock down the requirements to the end customer or user of the data. All input dataset fields will be mapped to the output dataset. When there is field duplication between all the inputs, the output dataset field will be a combination of each occurrence. All unique input dataset fields (those not found in other input datasets) are also mapped to the output dataset. It is possible for a field map's sub fields to be of varying data types. In this case, the output field's data type is set to the data type of the first input dataset and all other sub fields are cast to this type, e.g. the first input field is text, and the second input field (of the same name) is double. The output data type will be text, and the values in the second input field will be cast (converted) to this type.

4. Migration Process:

After completion of the analysis and mapping steps, the process of importing the data into the new system must be defined. Data migration solutions extract data from a source system, correct errors, reformat, restructure and load the data into a replacement target system. It sounds simple, but poorly managed data migration is the most common cause of failure in implementing a replacement system. The data is an immensely valuable asset, built over years of operations. The whole replacement project relies on successful migration. If the migration project runs into problems, the future of your company may be at stake. The data migration process will be completed here on manually defined process i.e. by writing a simple PL/SQL block.

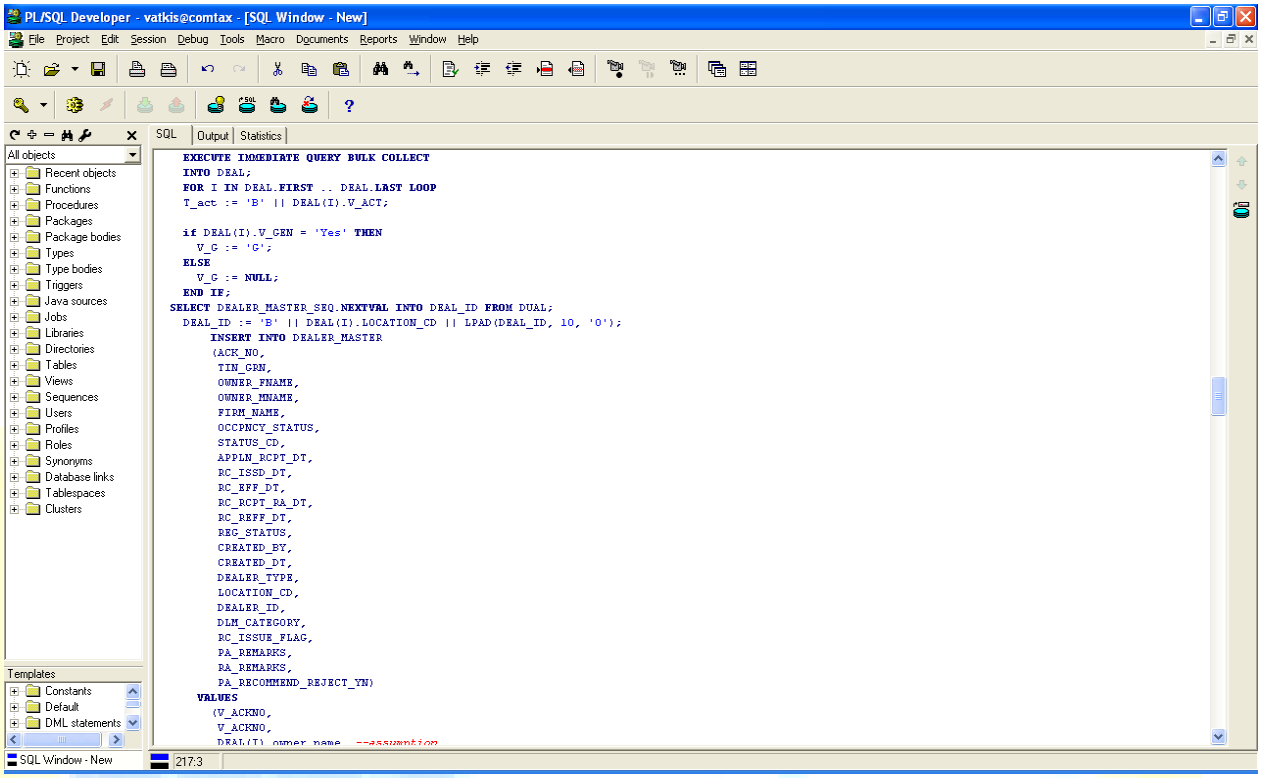

Figure: 4

The PL/SQL block (not showing completely) as mentioned in above figure, migrates the data in intermediate server i.e migration server after completion the process need to export the objects and then finally import the data in new system. Most data may be produce in flat files of legacy system. The files are then loaded either by a utility such as Oracle's SQL Loader or by vendor-supplied loading programs. Direct updates to the target database with constraints enabled are much slower.

Conclusion:

Data Migration is an important event that requires a significant amount of budget and occurs at regular intervals, consumes significant budget and labor and occurs very regularly. The combination of the frequency of and resources consumed in a data migration results in data migration taking a significant amount of the IT budget. With the increase in size and complexity of storage infrastructures, data migrations are becoming more complex, risky and labor intensive. Therefore organizations must focus on effectively managing this very crucial and significant portion of IT budget. When the Data migration projects are complex, the entire process includes

large-scale projects requiring many in-house and contractor personnel. Due to this market for labour, consultants, software and hardware for data migration has increased significantly. Size of data migrations market can be calculated by identifying amount of data migration activity that results in large data migrations. Data migration projects are crucial to the success of the initiatives that the migrations support; they impact business-critical data, applications and systems, and result in significant cost.Very detailed and careful planning needs to take place to clearly identify windows in which downtime is acceptable and ensure that no data is lost. For data migration projects that include mission-critical business data, the risk of impacting sales operations is high; the loss of availability or access to the data could directly impact the profit and loss of the business. Data migration projects are complex projects that possess significant cost and risk. To successfully complete a data migration project, organizations must develop a comprehensive plan encompassing people, processes and technology.

Different methods of Data migration offer ample scope for further research. Each method has different levels of cost and risk with varying advantages and disadvantages. An organization should choose that method which is optimal for its environment.

References:

1. Eric Anderson, Storage Systems Program, Hewlett-Packard Laboratories Palo Alto, In An Experimental Study of Data Migration Algorithms, CA 94304,Springer-Verlag Berlin Heidelberg 2001.
2. Joseph R. Hudicka, Dulcian, Inc., an Oracle consulting firm, In An Overview of Data Migration Methodology, Magazine - April 1998
3. Patrick Allaire, Justin Augat, Joe Jose and David Merrill, In Reduce Costs and Risks for Data Migrations, White Paper Feb-2012,Hitachi Data Systems
4. Gershon Pick, March 2001, Data Migration Concepts & Challenges