



CHALLENGES IN DATA MIGRATION IN SUPER SPECIALITY TERTIARY CARE HOSPITAL: A CASE STUDY

Amal Krishna Sarkar

Department of IT

BB Ambedkar University, Lucknow, India

R A Khan

Department of IT

BB Ambedkar University, Lucknow, India

C M Pandey

Department of Biostatistics & HI

Sanjay Gandhi PGIMS Lucknow, India

Abstract: Data migration is the movement of data from a single source or multiple sources to a new target database. It may be driven by a range of initiative starting from application up-gradation to replacement. The other areas may be the need to consolidate data within a data warehouse in an organization. The main objective of this study is to make a roadmap for the migration of healthcare data in super speciality tertiary care hospital and research organization. The purpose of the work presented in this is to also to study the challenges in migration of healthcare data. With the goal of providing a comprehensive solution and improve the process of data migration and also meet out the challenges in migrating complex data, a strategy has been made. Considering the fact that a tertiary care super specialty hospital, the work requires in depth understanding of its unique requirements and the complexities involved in healthcare data. A formal framework for refining transformations occurring in the process of data migration will be developed. Hence it is an essential that the processes are efficient, and suitably enabled by latest technology.

Keywords: Hospital Information System, migration, Picture Archival Communication System, Electronic Health Record

I. INTRODUCTION

The science and practice of health or medical informatics changed radically in the late 1970s and early 1980 when computer use began to become increasingly common in healthcare environments. Since then, improvements in the speed and processing power of computers, computer networks and the internet has led to increased accessibility and availability of information for healthcare professionals to support their decision making process. A revolution is taking place in the healthcare field with information technology playing an increasingly important role in its delivery [1]. As infrastructure has a limited life cycle and technology is getting changed and also the volume of structured and unstructured data is increasing at a very fast speed in any super speciality health care setup, the organization must find the efficient and effective way of upgrading their hospital information system and strategy for migrating the HIS and PACS data. Data migration, as a fundamental aspect of projects on modernizing legacy systems, has been recognized to be a difficult task that may result in failed projects as a whole [2]. Clinical and administrative procedures are generating numerous records for each patient, thus complicating storage, analysis and retrieval of information required to support hospital functions. Integration is achieved by implementing an expert system. This means that there are tools that offer solution to the problem of increasing accumulation of patient data. It has to be noted that some of the features discussed for the future already have been seen in the past and have vanished in the present. They may come back in the future in a 'recycling' development as technology advances [3]. The implementation of HIS is therefore a major challenge in the

healthcare setting. Acknowledgement of this has led to a need for understanding the match between HISs and existing IT infrastructure, organizational structure and established routines. Tertiary hospitals deliver complex medical and related services, such as allied health care, inpatient, outpatient and emergency services, within a single organization [4].

II. PROBLEM DEFINITION

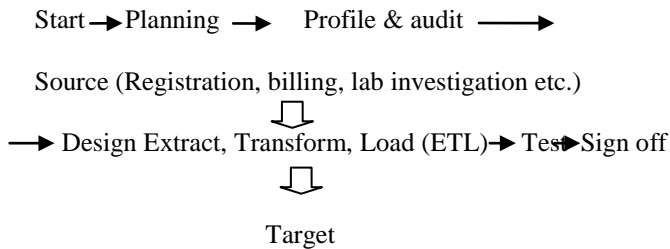
It focuses following problem areas

- What are the different types of risks and challenges in the migration of healthcare data in a super speciality tertiary care hospital and research organization?
- How to overcome the challenges, minimize transition time and improve the data migration process?

III. HEALTHCARE AND DATA MIGRATION

Data migration is the process of translating data from one digital format to another. Successful data migration is the key for ensuring accurate, timely and quality data which become the foundation for the digital healthcare system, whether in the form of electronic health records, computerized physician order entry or picture archival and communication system. In hospital environment specifically in super speciality tertiary care medical hospital and research organization, data migration is an important and challenging task because of its complex nature of data. Access to accurate clinically relevant healthcare information is imperative for quality of care. The whole migration process should be prepared carefully and thoroughly

investigated with regards to required time, manpower, and equipment resources [5]. A basic understanding of data migration is crucial to IT manager and higher administrative authorities of any organization. Knowing the data and the quality of data provides the understanding need to formulate a plan to provide the new EHR system [6]. Any healthcare organization wish to migrate their patient data may also require some external expertise to help in planning process.



A. Data challenges

Exploding data volume: The data size in most of the healthcare organization is large in size and may contain both structured and unstructured data, digital images and video contents. Every healthcare organization, therefore, should find an efficient way to store, manage and protect healthcare data.

Content Management: Healthcare data may be of two types. It may be generated internally within the system or from outside the system and the data generated outside may not be structured in nature. Proper management of contents of such data is very important.

Data privacy and security: Data privacy and security in patient care data is an imperative aspect. Security breaches in the privacy of health care data may cause very serious impact. Patient portal and mobile access present new challenges for maintaining patient privacy and data security and accessibility.

B. Migration planning stages:

Virtualization: In case of data migration, any organization should take into consideration moving to virtualization of servers and storage system and even opt for private cloud storage if feasible. This will reduce the cost and increase the efficiency. Virtualization of storage smoothen the data migration process and also reduce the downtime during the migration. Proper disaster recovery setup is also an important aspect this process. In this step, it has to be determined what data should be transferred; further, the possible “clearout” of data should be determined. This requires defining methods and rules as to how to determine which data are obsolete or incorrect and thus should be removed or corrected during the data transfer process. Analysis of legacy PACS stored information may also require an investigation of the archived data to estimate how much of the data will need cleanup [7].

Advanced hardware and protocols: To smooth the progress of enhanced performance and improved volume of storage healthcare organization should opt for new possible products and protocols.

Data storage archiving: As data volume grows significantly day by day, it becomes very expensive and not easily manageable to store and handle large volume of patient care data on storage. Therefore all healthcare

organization should plan and decide how much and how long data will be stored on expensive online storage and how much can be transferred to archive and comparatively lower performance storage tiers. Now a day, automated tiering solution is also available with IT players like EMC, IBM and some other brands / vendors.

Downtime: In any data migration process, downtime and cost are very important aspects. IT decision-makers must determine how much downtime the organization can allow and make a migration schedule considering the downtime tolerance into account. Migrating different applications or parts of applications and data at different times may be desirable. Some applications and data may be migrated during night hours when only emergency services are operating and can get by with read-only access. Other data and applications may have to be migrated live as they’re still being accessed by users.

C. Planning and design

The next step is a careful design of the migration steps, methodology, scheduling and tools. Planning carefully for the long term allows IT system managers to draw up a possible plan that works for everyone. Knowledge regarding the data enables the organization to populate new electronic health information system with clean and trusted data to support patient, clinicians and other paramedical staff of healthcare organization. Without proper planning, there is a high risk that the cost of the project will go higher and exceed the allotted time, or even fail completely [8]. This is where to determine if third party migration tools will be needed.

Preparation: In this phase, one should prepare the old environment for migration and install while configuring the new environment. In case of moving the storage geographically, one should strongly consider doing the migration onsite and then moving the equipment, because migration across a WAN will take more time and also it is error-prone.

Migration: If the plan and design are effective, this should proceed in a somewhat knowable fashion. This step shows the plan which will be used to do the actual migration of data. It is estimated that as much as 90% of the specifications initially provided for data migration projects change significantly during the life of the project; and over 25% of the specifications will change more than once (in many cases several times) before the project is complete [9].

Validation: This is one of the most important steps in this process. Migration teams should consult with various healthcare departments to decide exactly when to cut over to new systems and storage once they have been migrated. It makes sense to do so at the end of a shift or a weekend or during holiday when the user load minimum. Also it needs a best possible plan for effective validation of systems and data, by key IT staff and medical professionals, before the final cutover takes place.

D. Data migration tools

There are number of ETL tools available in the market, with different functionality and facilities. Most of the tools have almost similar functionality, but each has its own merits and demerits. Some tools have developed a more suitability and popularity for specific types of data and

application. Following are some of the most common tools which have gained popularity.

Apex Data Loader: It is a data loader app, with a basic interface and with the help it import and export of data can be done very easily. Apex Data Loader comes with the option of updating existing records also. When a similarity/match is found, the record is updated else a new record is created. This is a very handy and useful when there are frequent updates of information from an external system. It also has some other advantages which generate a pair of success and error CSV files with each operation. It has an Automap button which speeds up field mapping process significantly. There is a "Mass Delete" option for unlimited records and all tables. The main drawbacks of this tool are that the app supports the import of only CSV files, and has a very limited option to transform the data.

JitterBit Data Loader: This is a useful free ETL tool. It is best suited for quick and ad-hoc tasks. The app scores in its easy management and wide range of functionality. It not only offers insert, update and "upsert" features, but also provides options to query, delete and bulk load. It is a wizard-based graphical point and click configuration that makes the app very easy to use. This app can be obtained freely although it has paid versions also. The free version is restricted in terms of some specific facility and functionality. For example, the free version makes it very difficult to manage specific ETL projects or share projects with other users.

Progress Data Direct: This tool is a very highly reliable and strong ETL tool that combines conventional and new interfaces to achieve better data connectivity with Salesforce. Some of the other features are its efficiency and ease of use. Built-in cache of this tool resides entirely within the driver, without requiring an intermediate database. This improves performance and also restricts web service calls to the Salesforce API. As this tool uses only a single connector to Salesforce across all applications hence the administration is comparatively easy. It is possible to configure the driver without the changes in application code to load strategic and tactical data.

Starfish ETL: This ETL is popular due to its flexibility, power and speed. It can connect almost all data sources. It also functions by bringing together multiple databases. It uses pre-defined data maps and allows modification of such data maps. This app can perform some multipurpose actions such as converting data types and checking of duplicates and can run scheduled jobs.

Midas: Midas is a flexible ETL tool. It is a cloud based ETL tool with high performance and it can deliver seamless bi-integration. It supports databases such as MySQL, SAP and other ERPs also. The tool minimizes implementation time, effort and cost. This tool is very suitable for handling large volumes of data with considerable ease in the cloud, makes it suited for considerably big data stacks. The processing takes place in double quick time, without affecting the quality and at the same time with reduced costs.

Apartar: This offers an advanced ETL toolset that works flawlessly on most of the third-party applications and databases. Its versatile interface makes it possible to carry out complex integration tasks without even a single line of coding or design. The suite creates native SQL automatically. It is so powerful that it can process data

spread over several data sources and files. It remains lightweight with consumption of small memory and minimum CPU utilization. This can be installed very fast, taking just a matter of seconds. The tool records run and error logs and prepares diagnosis and troubleshooting easy. The main advantage is that the full functional-no limitations Apartar Community Edition is a free tool.

MassEffect: This tool is available with support for importing and exporting advanced file formats like CSV, MDB and UDL, and it has many other exceptional features such as support for international characters.

E. Data migration risks

Similar with all other technology operations, data migration also does have some risks. These include:

- Scope creep - additional machines
- Schedule slip – it may happen due to network outage/bandwidth issues
- Planned outage/maintenance window limitations
- Resource backup/unavailability – sometimes resource backup of data may not be available
- Hardware failure – it is also one of the main failure aspects. During the migration process some of the hardware components involved in the migration process may face failure
- Poor qualification/due diligence/ design (storage data sizing/capacity)
- Data corruption – It is also an important aspects of failure of data migration. If the data to be migrated is corrupted the whole migration process may fail. A study by The Standish Group found that "72% of all migration and consolidation projects suffer significant overruns or failure" [10].

IV. SCOPE

With the goal of providing a comprehensive solution, the scope includes the estimating the data challenges and risks involved in migration, planning and design, implementation of a Data migration work. Considering the fact that a tertiary care super specialty referral hospital, the project requires in depth understanding of its unique requirements and the complexities of data involved. This challenging project seeks to transform the current environment to a more effective and efficient one. Hence it is an imperative that the processes are efficient, and suitably enabled by technology. For this purpose, migration of healthcare data at Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow has been taken as case study.

V. DATA MIGRATION AT SGPGIMS, LUCKNOW

A. Challenge

The migration of OLD HIS data to new HIS was complex because of the following reasons

- Live environment without affecting the patient care.
- Different Source and target (9i → 10g)
- The source had around 360 tables and the target had around 1800 tables. Table -1 shows the classification of tables in the old database.
- Different lifecycles of records (0 days-13 yrs).
- Complex patient care/mapping rules.
- Large Volume of data(500 million records)

- Missing/incorrect data in various source tables
- Improper linkage between different modules.

Classification of tables

Table – 1.Tables can be classified into following types

Type	Description	Tables
Masters	All data required for the master configuration comes under this category (table -2).	62
Transactions	All patient transactions are stored in the database comes under this category. For example registration, appointment, investigation orders, billing etc (table-3)	277
Archive	Data which are in zip or compressed format comes under this type	12
Temporary	Tables store the temporary information that was reset after completion of process.	9
	Total	360

B. Migration Strategy

- Different migration approaches like incremental migration, Difference approach and ETL tool were tried out.
- Six trials of full data migrations and mock run were done before final migration based on ETL.
- A cold back up of the OLD HIS data was taken and loaded into a similar database structure with similar platform outside the system. Validation of data done and then loaded into the new system.

C. Validation strategy

- ▶ Using **ETL** Tool.
- ▶ Using **Script**:
A Script was written for validating the below things.
 - Number of Source records and Number of Target records
 - Table having zero records and Count records of each table.
 - The total collections and refund was compared for a particular date range for source and target tables

Elements of ETL: ETL stands for Extract Transform and Load. Architecture of the ETL is shown in figure – 1.

- *Extracting* data from primary source active or archive systems, data warehouse (eg. database, excel sheet, csv file etc)
- *Transforming* the data - which may involve cleaning, filtering, validating and applying business rules.
- *Loading* the data into a data warehouse or any other database or application.

Above tool was used for migrating the OLD HIS to new HIS tables.

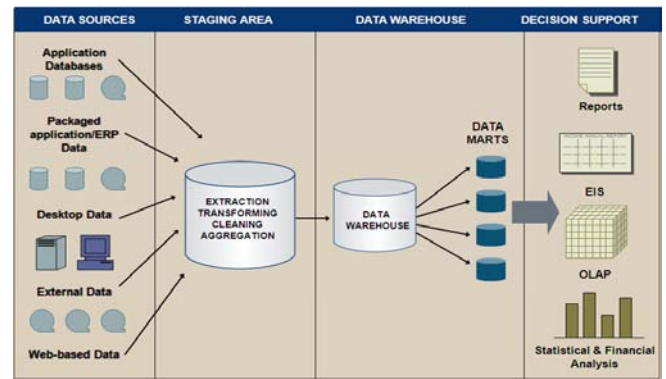


Figure-1: ETL Architecture

Table -2.Sample Summary of master data

Table Name	Category	No. Of Rows	Indexes	Key in Table
BILL_ACCT_MAS	Billing	6086	3	1
BILL_ACCT_MAS_01	Billing	3617	2	1
MAS_TAB_01	OPD	328	1	1
MAS_TAB_02	OPD	37	0	0
MAS_TAB_05	Appointment	62	0	0
MAS_TAB_08	Billing	1379	1	1
MAS_TAB_08_01	Billing	2161	3	2

Table –3.Some Important tables used by Investigation Module

Table Name	Category	No. Of Rows	Indexes	Key in Table
MAS_TAB_08	Master	31	1	1
MAS_TAB_09	Master	26	1	2
MANDATORY_VALU	Transaction	40	1	1
TRAN_05_03	Transaction	22972000	4	3
TRAN_05_04	Transaction	6863301	2	2

VI. FUTURE PATH FOR FURTHER MIGRATION

Based on experience of this migration exercise the hospital information system (HIS) should have been provided with several export and backup facilities:

- Cold backup
 - Data pump
 - Rman
 - Export (csv,xls,pdf)
 - ▶ Optimization of backup time – Backup and restoration should be minimized
 - ▶ Provision of Incremental migration.
- All tables have primary key and date fields.

VII.CONCLUSION

Ultimate aim of data migration should be to improve the data availability, performance of healthcare information

system. To succeed, data migrations must be given the attention they deserve, rather than simply being considered part of a larger underlying project. Following are the key areas which should be addressed in fully integrated migration environment:

Proper understanding: Thorough checking and auditing of various data sources from adequate samples can eliminate unexpected scenarios during the migration process.

Improving data quality: Where ever the poor quality data source is there those should be addressed before or during the migration process. Data quality software can be used to restructure, standardize, cleanse, enrich, find duplicate and reconcile the data.

Projecting and maintaining data quality: Over the time migrated data degrades naturally until it becomes a problem again. Maintaining and improving the quality of this data is vital to increasing the value that can be derived from the information. Healthcare data needs to be protected from degradation due to human and other possible errors, incompleteness, or duplication. Implementing a data quality firewall to police data feeds—in both batch and real time—is critical to maintaining the integrity and therefore the value of the application.

Proper governing: Tracking and publishing data quality metrics regularly to a dashboard enables senior executives or users to monitor the progress of data migration projects or data quality initiatives.

Proper and structured methodology can reduce the problem of managing a complex data migration, but the correct choice of technologies will go a long way to promote a successful outcome. A range of software from different suppliers, plus a lot of technical know-how, has been necessary to successfully accomplish a data migration, but the architecture becomes difficult to manage and performance deteriorates with numerous different interfaces between applications. The ideal solution is a software tool that supports the whole data migration lifecycle, from profiling and auditing the source(s) through transformation, cleansing, and matching with the population of the target.

VIII. ACKNOWLEDGEMENT

The authors like to extend very sincere thanks to the reviewers and acknowledge their valuable suggestions and also sincere thanks to IT team at Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow who has supported the work.

IX. REFERENCES

- [1] E. S Smith, JHP Eloff, "Security in healthcare information system – current trends," International journal of medical informatics- 54-(1999).
- [2] Bernhard Thalhin, Quing Wang, "Data Migration: A theoretical prospective," Data and Knowledge Engineering (2013), 87, 260-278.
- [3] Peter L. Reichertz, "Hospital Information System – Past, present, future," International journal of medical informatics (2006) 75, 282-299.
- [4] Jaunita I Fernando, Linda L Dowson, "The health information system security threats: an informatics theory," International journal of medical informatics (2009) 78, 815-826.
- [5] Jung H, Kim H-J, Kang W-S, "Migration of medical image data archived using mini-PACS to full-PACS," J Digit Imaging (2004).
- [6] Joy Ales, "Maximizing Legacy Data Migration to new EHR," white paper, August 2015.
- [7] Peter M. A. van Ooijen · Kadek Yota Aryanto André Broekema · Steven Hori, "DICOM data migration for PACS transition: procedure and pitfalls," Int. J CARS (2015) 10:1055–1064.
- [8] "Successful Data Migration" An Oracle White Paper October 2011.
- [9] "Rapid Application Development (RAD) for Data Migration" —White Paper Solutions by Premier International, http://www.premier-international.com/pdf/Applaud_White_Paper.pdf 2004.
- [10] "Comparing Applaud to ETL Tools" — White Paper Solutions by Premier International, http://www.premier-international.com/pdf/Comparing_Aplaud_to_ETL_Tools.pdf.