Master Data Managements Role in Data Migration - Part1

Applies to:

SAP NetWeaver 2004s/ MDM 5.5 SP 06.
For more information, visit the Master Data Management homepage.

Summary

This article in 2 part series discuss the role of MDM in a data migration project and how MDM can provide an effective data cleansing solution , which is a mandatory and important phase before data migration. In absence of which, the data will be just “garbage in and garbage out”.

Author: Deepankar Pandit

Company: Satyam Computer Services Ltd.

Created on: 06 September 2008

Author Bio

He has been associated with Satyam for 2 years and has been a part of MDM practice since February 2007.He also has ABAP and data modeling experience and completed his Bachelor’s degree in Electronics and Telecom Engineering.

Table of Contents

Introduction

Master data is data that describes an organization's key business entities, such as customers, products, vendors and employees. A master data management (MDM) program combines data expertise focused business processes and specialized technology to ensure consistency and accuracy of data across organizational and business process lines. Once in place, an MDM program greatly simplifies application, process and data integration, enables competitive agility and responsiveness, and facilitates analytical accuracy and real-time reporting.

Having said that, it's out of denial that bad data is poison, clogging operations, increasing costs, prompting poor decisions and, in the worst-case scenario, may lead to business downfall. Gartner estimates that more than 25% of critical data within large businesses is somehow inaccurate or incomplete. Of 750 IT professionals and business executives surveyed by the Data Warehouse Institute in 2005, 53% claim their companies have suffered losses or increased costs because of poor data.

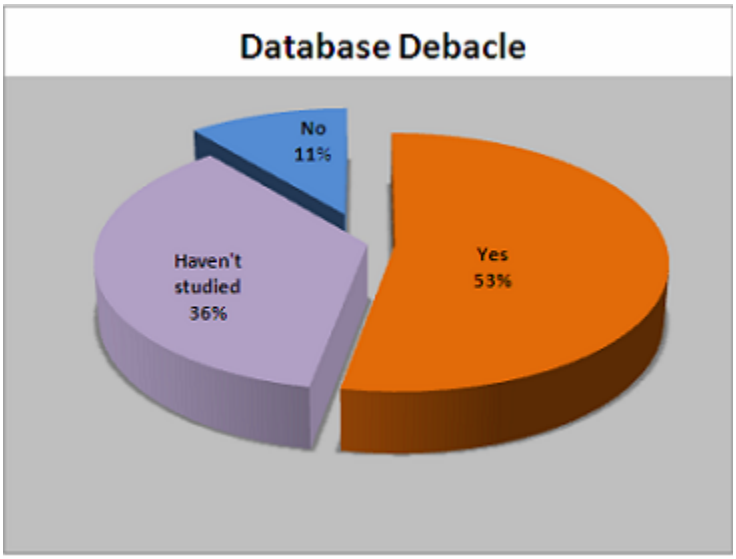

Figure 1: Response to query for companies that incurred losses due to bad data.

Data migration is the process of translating data from one format to another. Data migration is necessary when an organization decides to use a new computing systems or database management system that is incompatible with the current system. Typically, data migration is performed by a set of customized programs or scripts that automatically transfer the data. Even though the structure of the new system might differ from that of the old system, it is important to retain the essential meaning of the relationships between data. Data migration usually involves some element of data analysis to ensure that information captured by the data is not lost.

Data Cleansing or **data scrubbing** is the act of detecting and correcting (removing) inaccurate, incomplete or erroneous records from a record set, table, or database which involves identifying the prior defined attributes in a data and replacing, modifying or deleting them.

An organization in a data-intensive field like banking, insurance, retailing, telecommunications or transportation have to use a data cleansing tool to systematically examine data for flaws by using rules, algorithms and validations. Data cleansing is a complex activity which not only requires tool specialist but also process owners and data analysts who have complete understanding of data. Using a cleansing tool can save a database administrator a significant amount of time and can be less costly than fixing errors manually.

Landscape – Legacy System Retirement

There are good amount of projects running across the globe where data is getting migrated from one Legacy system to another Legacy or SAP system and eventually the prior is decommissioned. These are huge data migration projects where data cleansing is a mandatory activity before incorporating the data in to the target system. The presence of data alone does not ensure that all the management functions and decisions can be taken smoothly. There is a compulsive requirement for the data to be meaningful which pertains to data issues such as

- Accuracy
- Integrity
- Cleanliness
- Completeness
- Correctness
- Consistency

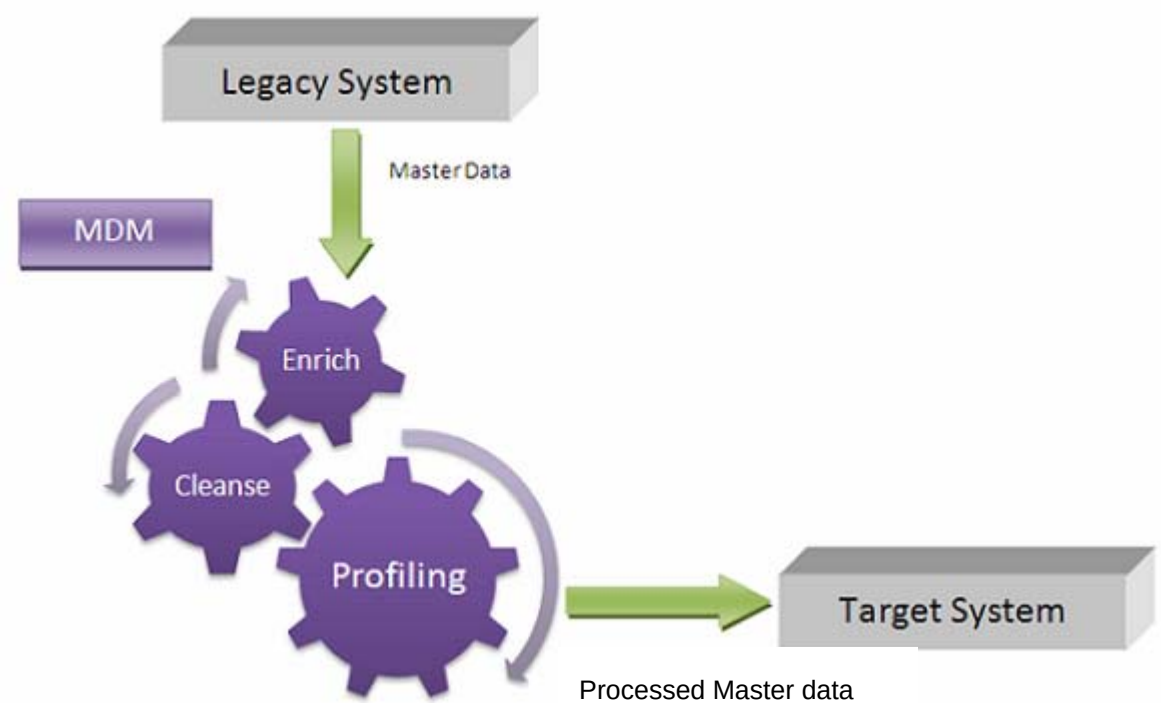

Figure 2: Landscape Overview

As a migration is planned, there can be potentially additional issues apart from data which is raised due to differences in data model in the source and the target system. These error types have to be categorized for cleansing. In such cases legacy data update is not performed as it would conflict with the data model of the legacy system.

Often the existing data has no consistent format being derived from many sources. It may also contain duplicates records and may have missing or incomplete descriptions. Data has to be normalized in cases where the standards are not matching in the source and target system. For e.g. lets consider a legacy system maintaining materials where the unit of measure have been kilograms, tones or liters etc where in SAP it's maintained as codes in base unit of measure like KG, TO or L .These kind of values have to be standardized so name of each attribute is consistent.

Activities

In a Data migration project the main objective is to process and migrate clean data to the target system to make it"system ready" in relevance to the target system. Data cleansing can be an elaborative process and has to be planned carefully to achieve the objective of elimination of dirty data and can be achieved through a methodology which is discussed below.

Error identification Phase (Data Audit): The foremost phase identifies and categorizes the various errors in the legacy system. This requires a deep study of the functionality and process of the legacy system and is carried out with the help of Business analysts. A data audit will provide:

- Error types that need cleansing categorized as critical errors. They can't be carried forward to the target system.
- Error types that can be ignored safely and are not business critical.
- Data Volume of each of the critical error types.

Data Cleansing Phase: This is the phase where MDM will play the crucial role. This is typically a batch process for correcting the errors based on the error types. Based on the above phase the data owners will suggest the changes implementing best practices. These change requests will be attended by a MDM consultant .These phase will involve steps like **data modeling, data profiling, and data cleansing** which will be acquired using MDM's features and modeling capabilities. This will be discussed in broad light in the phase 2 of this article.

Post Cleanse Data Audit: Data cleansed in batches is verified by re-run of error identification process. This is to ensure the successful completion and flawless functioning of the data cleansing process.

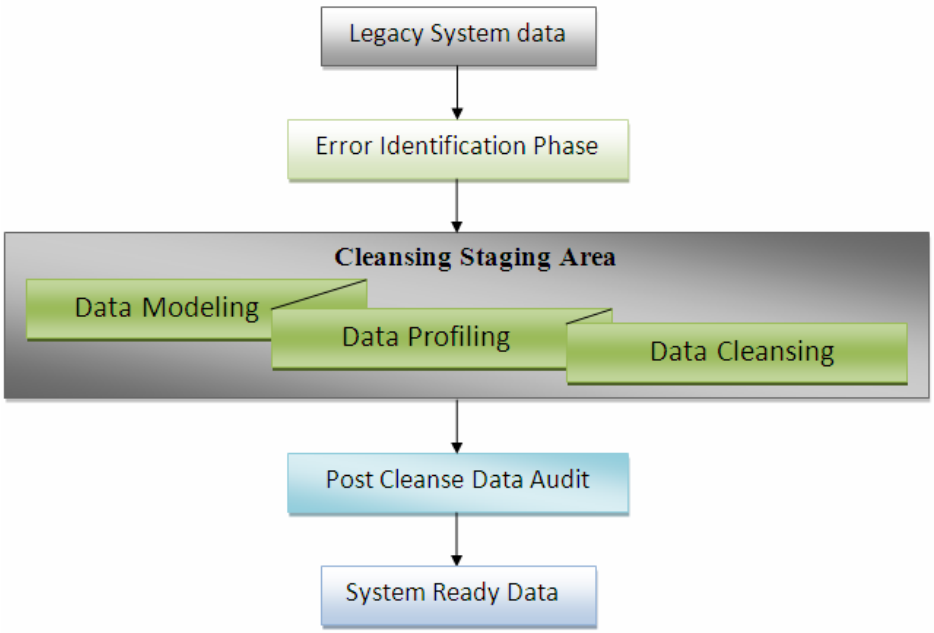

Figure 3: Process Chart

Achieving 100 percent data cleansing in reality is a bit difficult. Despite best efforts made there still exits a certain percentage of dirty data. This residual unclean data should be reported stating the reasons for failure.

Note: In projects wherein data cleansing is the only motive and the legacy system is to be used, a legacy system update is performed. Moreover in such cases an effort is to be made with respect to the origin of these error types. Investigations may almost lead to change in process or rectification of its identified vulnerabilities**.**

Realization in MDM

The cleansing phase is realized using MDM and there are components in MDM architecture which supports this phase. The Console of MDM has data modeling capabilities and repository can be designed using it, from scratch as per the requirement. The **Import manager** has specific data conversion and mapping capabilities which may come handy in data profiling for e.g. if the data in the source contains the material number and description together which has to be separated in to two different fields separately in the target system .

The **Data Manager** has amazing data cleansing features through which data can be validated, values can be assigned, and even duplicates can be sorted out from the source data set.

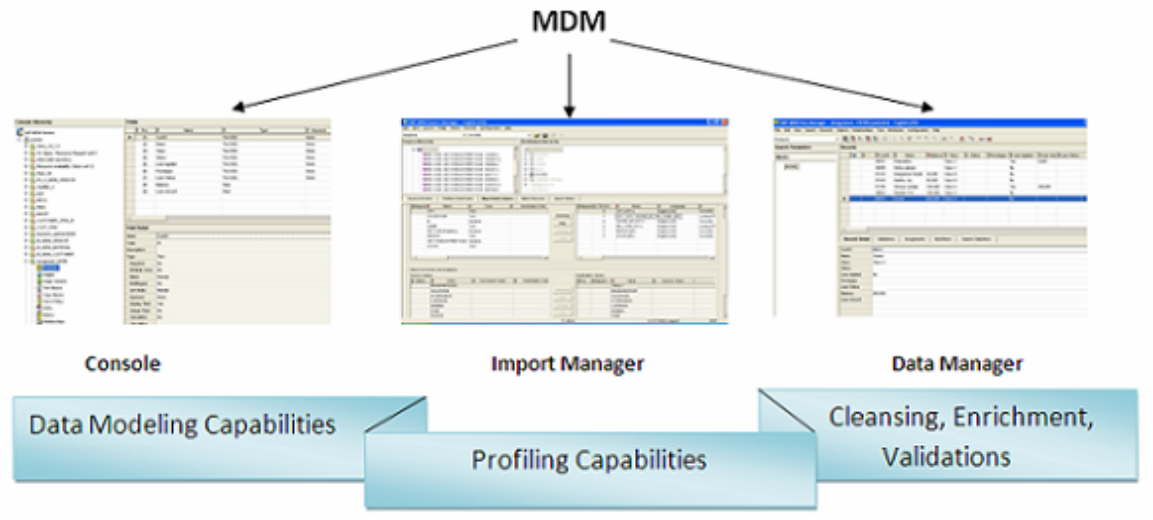

Figure 4: MDM Architecture modeled for data handling

These features are described in detail in the second part of this article. The MDM architecture provides an ideal platform which can efficiently handle data related issues and make the data system ready for the target system.

In the cleansing phase the MDM consultant has to work very closely and collaboratively with the data stewards for designing rules, strategies and validations as per the scenario and landscape. Achieving cent percent error free data is a bit difficult but the strength of the cleansing determines the closeness to that magic figure.

Summary

The quality of data must be evaluated before data migration to ensure its usability and effectiveness in the new system since master data is concerned. The success of operational data in the new system highly depends on the quality of the data feed to it and MDM can be an efficient tool to handle this. Systems doesn't have intelligence to make data audits thus unless it's backed up by good data cleansing methodologies and business strategies its going to be garbage in and garbage out.

Related Content

https://www.sdn.sap.com/irj/sdn/thread?threadID=162559

https://www.sdn.sap.com/irj/sdn/weblogs?blog=/pub/wlg/10453.

www.sdn.sap.com/irj/sdn/mdm-elearning.

For more information, visit the Master Data Management homepage.

Disclaimer and Liability Notice

This document may discuss sample coding or other information that does not include SAP official interfaces and therefore is not supported by SAP. Changes made based on this information are not supported and can be overwritten during an upgrade.

SAP will not be held liable for any damages caused by using or misusing the information, code or methods suggested in this document, and anyone using these methods does so at his/her own risk.

SAP offers no guarantees and assumes no responsibility or liability of any type with respect to the content of this technical article or code sample, including any liability resulting from incompatibility between the content within this document and the materials and services offered by SAP. You agree that you will not hold, or seek to hold, SAP responsible or liable with respect to the content of this document.