

A PROVEN APPROACH TO ENTERPRISE DATA CLEANSING AND MIGRATION

Objective

This paper provides an insight into Best Practices for assessing initial data quality, ability to cleanse the data prior to migration and assuring successful end-to-end data migration with a view of each step of the Software Development Life Cycle.

By Sid Azhar - Vice President, Chain-Sys Corporation

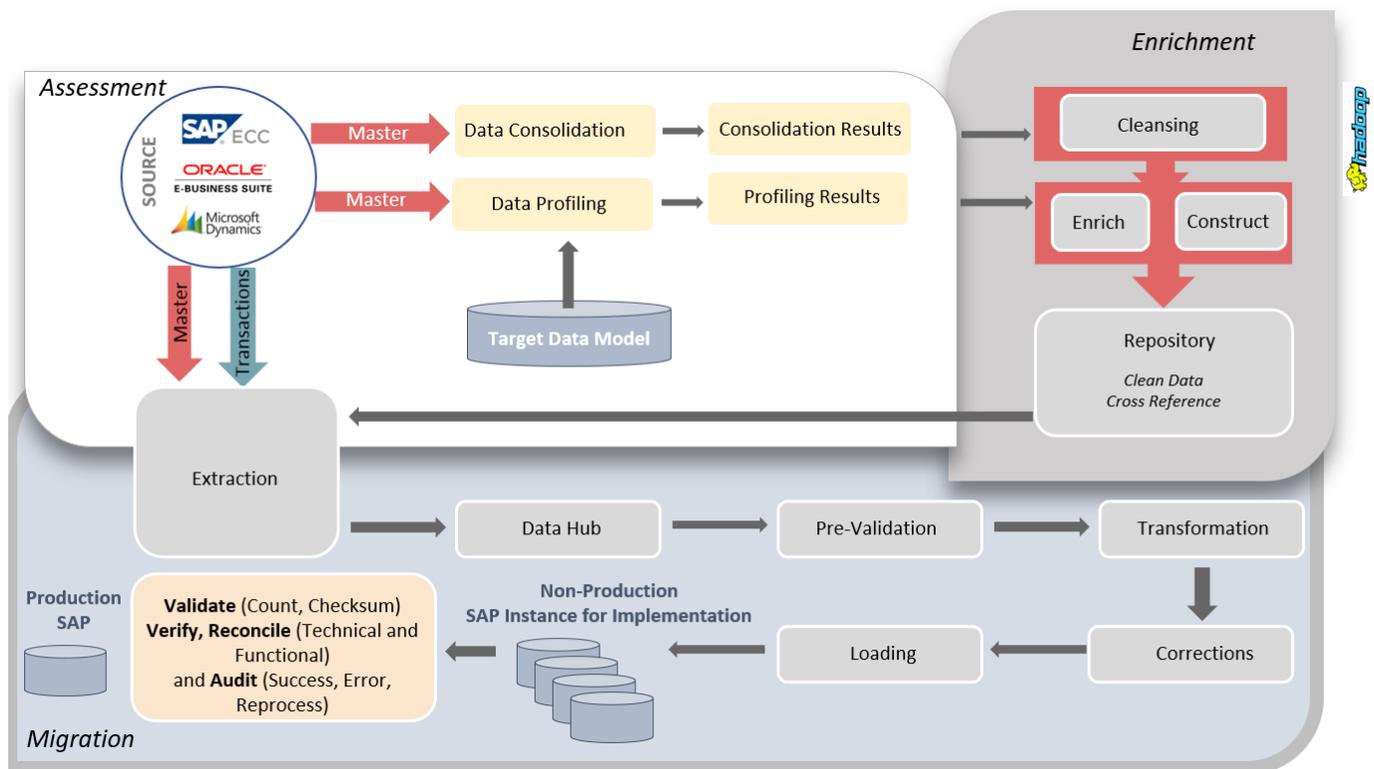
Chain-Sys Corporation, an Oracle Platinum partner & SAP partner, is a highly successful "Systems Solution Provider" that offers its own suite of Productivity solutions powering best of practice data interfaces between ERP and other external systems, master data management and data migration using pre-built adaptors.

A Proven Approach to Enterprise Data Cleansing and Migration

BUSINESS EXPECTATIONS

Enterprise customers face enormous data challenges when migrating data between various Applications, especially from older legacy applications to modern Applications. Business expectations are focused on:

- Assessment of Data Quality in the Source Systems
- Ability to clean the source data prior to migration
- Accommodate data transformation requirements between the Source and Target Applications
- Data readiness prior to data loading into Modern Target Applications
- Establish data governance for on-boarding data into Modern Target Applications
- Data Reconciliation between the Source and Target Applications
- Audit Trail that gives visibility into the data migration process



Data Assessment

Data Assessment is the first step of the data migration process and serves as a critical starting point in identifying the appropriate source systems and performing a health check on the source data. It is important to involve Business Users to identify all source systems and how the data is used, which helps with source data transformation requirements for the target systems. Once the appropriate data is extracted from the source system(s), the extracted data is subject to:

(1) Data Profiling

(2) Data Consolidation Identification

The output of these activities leads to Data Cleansing, Data Harmonization, Data Enrichment and helps build Data Construction requirements.

“Data Assessment” should also provide mappings between the source and target applications for all appropriate Master, Reference and Transactional data considered for migration.

Source System

Source systems are the systems from which appropriate data is extracted and migrated to the New Modern Target Applications. Source Systems can span across both on-premise and cloud environments. Typical source systems could be ERP, CRM, PLM, HCM, Supply Chain Planning, Project Management etc. Additionally, other sources such as Databases, Flat files, Excel, Unstructured Data and Machine Sensor Data may also exist.

It is important to involve both IT and Business Users to correctly identify all appropriate data sources.

Extraction

It is beneficial to utilize pre-defined adaptors for extracting Master data, Meta data and Transactional data from source Applications to expedite the extraction process and reduce reliance on developing custom code, which is cumbersome and often prone to errors.

Tool based adaptors support multiple connection methods such as Web Services (REST, SOAP), JDBC, JSON, XML, Excel, Flat Files, BAPI, OData, MQ,

JMS Queues etc., and can extract data from various databases (DB2, Oracle, Hana, Informix, Postgres, MS SQL, Progress, Sybase, and more) as well as Big data sources (MongoDB, Cloudera Impala, Hortonworks, Hadoop, Snowflake, Redshift, Teradata, Elastic Search etc.), and Sensor data.

Data Profiling

Data Profiling does an examination of the source data and gives an insight into the source data information. It helps identify how the data is being used, the cleansing requirements and how much data must be migrated. A Data Profiling engine can help identify all source data column attributes and table properties (Count, Nulls, Not nulls, Data types, Data patterns, Data values etc.), relationships between source tables and column attributes, and produces data mappings between the source and target applications such as SAP (S/4HANA , ECC, C/4HANA, Hybris), Oracle Cloud Applications (ERP, HCM, SCM, PPM, EAM, FIN, CPQ, Sales, MDM, PLM, QTR, Service), and Oracle On-Premise ERP (EBS, JDE and PSOFT).

Data Consolidation Identification (matching)

A consolidation engine should be considered that offers built-in Intelligent Automated Algorithms to identify potential duplicates across the master data spectrum. This helps with decisions related to (i) merging the records, (ii) keeping them as is or (iii) eliminating them as needed. This activity also helps to fix structural issues with the master data. Examples of master data generally considered for this activity are Materials, BOM, Routings, Customer, Supplier, Employees, Equipment's, Functional Locations, Chart of account etc.

Data Cleansing & Related Activities:

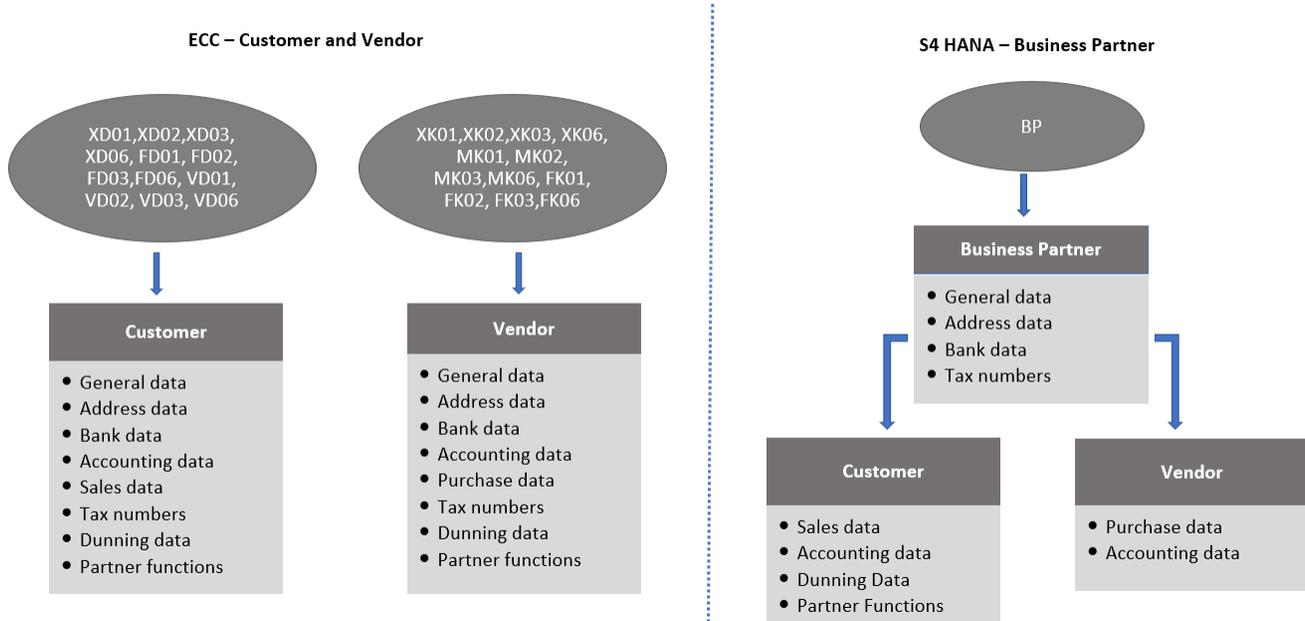
The “Data Assessment” output is utilized for assessing the requirements for additional activities such as Data Cleansing (De-duplication, Hierarchy etc.), Data Harmonization, Data Enrichment (adding new attribute columns as needed) and Data Construction (adding new data tables as needed).

A proven approach is to use “Repositories” for performing these activities. Repositories are configured using the ready-to-use “Data Models” available for SAP S/4HANA, SAP ECC, Oracle Cloud Application, Oracle

A Proven Approach to Enterprise Data Cleansing and Migration

EBS, Microsoft Dynamics etc. A Repository contains the end-to-end data cleansing and related activities as well as Data Governance to “Get the master data clean” and “Keep it clean”. Cleansed data is stored in the Repository. Data quality metrics are reported in the form of dashboards to get a view of the data and take appropriate action. Roles and responsibilities are also clearly defined, and progress monitored for the success of the project.

As an example, below is the Customer & Vendor master data structure in SAP ECC and the corresponding Business Partner master data structure in S/4HANA:



Data quality metrics are reported in the form of dashboards to get a view of the data and take appropriate action. Roles and responsibilities are also clearly defined, and progress monitored for the success of the project.

Data Cleansing

Data Cleansing helps eliminate duplicate master data records. This helps in re-organizing the master data against the right master and at the right levels (conforming to master data Hierarchy). Also, it helps create a ‘Single Source of Truth’ for Hierarchical representation of master data, which is very important for providing powerful analytics and insights.

Data Harmonization

Harmonization helps maintain cross referencing of merged child data against the parent data. It proves to be of critical importance when master data is merged

A Proven Approach to Enterprise Data Cleansing and Migration

during the data cleansing process. For example, if a Material Master record is merged, the underlying related transactional data records such as Sales Orders, Purchase Orders, Customer, Supplier etc. related to the particular Material Master record is crossed referenced appropriately.

Data Enrichment

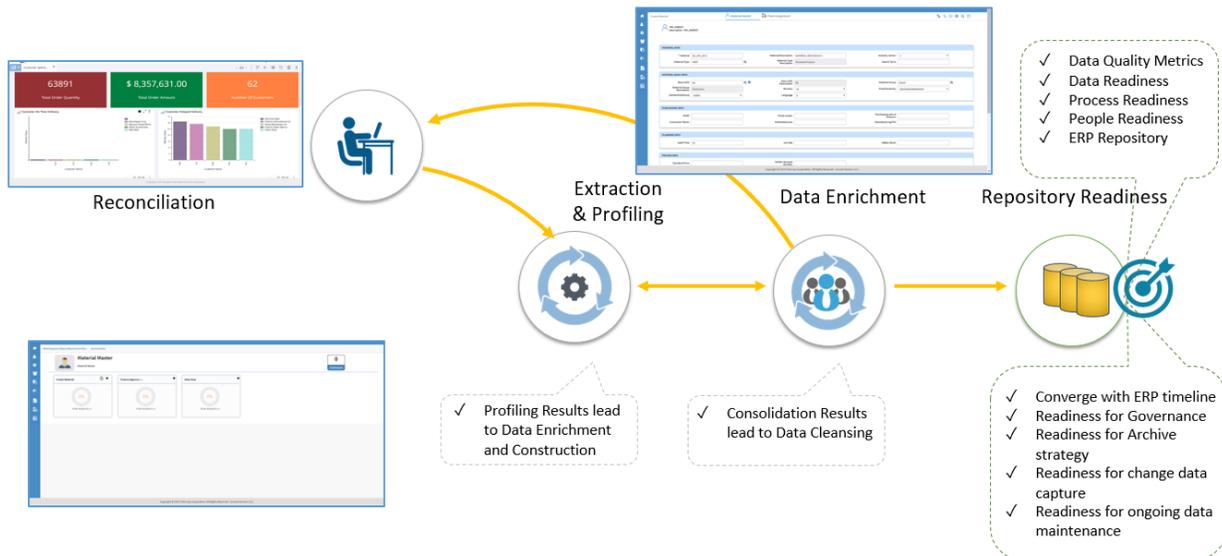
The Source master data, especially data from some legacy systems, may not have all the values needed for the New Modern Target Applications. This calls for additional Enrichments of attribute columns which can be very helpful for Improved Operations, Reporting, and Analytics.

Data Construction

The Source master data structure and the New Modern Applications data structures may not be the same. If new data tables must be added, this functionality can be utilized to construct new data tables that can be used for filling the data structures accurately. This is a valuable tool when migrating data from legacy systems.

Data Repository

It is very important to utilize a Repository to perform the Cleansing, Harmonization, Enrichment and Construction activities. Utilizing a configurable Repository with simplified screens and a governance mechanism is extremely valuable. The data output stored in the Repository is utilized for migrating the cleansed master data into the target application.



Data Migration:

The “Data Assessment” output provides standard mappings between the source and target applications for all appropriate Master, Reference and Transactional data considered for migration. Custom code development is discouraged for this phase, rather a tool-based approach can provide standard “data models” for target applications including SAP S/4HANA, SAP ECC, Oracle Cloud Applications, Oracle EBS, Microsoft Dynamics etc.

The output of the “Repository” that stores “clean master data” is used as the source for converting master data into the Target Applications. The Source Transactional data is directly used for migrating transactional data to the target applications. A migration tool should be utilized, which can look up the cross references built in the “Repository” and can apply the transactional data against the appropriate Master Data in the target applications.

Source Systems & Data Extraction	Source systems and Extraction are explained in the Data Assessment section.
Data Hub	Data is extracted into a Data Hub, which helps in viewing the extracted data in different forms. Extracted data is Profiled and Cleansing & Related Activities are performed in the Data Hub.
Pre-Validation	Pre-Validation helps to check the data for Functional Dependencies or missing Setups (uses check tables in SAP and Setups in Oracle). Pre-validation also checks for Technical data issues like: Constraints, Data Types, Length, Null check, Format mask for date and time, Check Text case, Duplicate data, Currency Rounding etc. The benefit of this is to understand issues before loading the data into the Target Application. Tools are available that provide hundreds of pre-defined pre-validations in each load Adaptors for SAP, Oracle, Microsoft and other major Applications, and can also enable users to configure additional Pre-validation rules based on business needs. Manual pre-validation approach is extremely cumbersome and often not reliable.
Transformation	Data Migration includes a Transformation program and not just Lift and Shift. Transformation rules engine enable users to configure Lookup and Expression based transformation rules. Internal or external data can be used for transformation logic. Available tools can provide pre-defined transformation logic for a majority of the source to target migration project needs.

Corrections	<p>Ideally the migration process should include Screens within the Data Hub to review all errors from Pre-validations and Application specific (SAP, Oracle, Microsoft etc.) load execution errors.</p> <p>These functionalities can be found within available tools. The errors can be corrected either online (Simple GUI) or offline (export and import using Excel). These are temporary corrections, which leads into transformation logic or corrections on the source applications to fix the problem permanently.</p>
Loading	<p>A tool-based approach is highly recommended since it not only automates the process but also helps expedite the data loading. An important consideration here should be to utilize a tool that offers high data volume loading within a short time and Change data capture (CDC) capability, which helps reduce the cut over timelines.</p>
Target Systems	<p>Target system is the destination for the data. An ideal data migration tool should support almost all Major ERP, CRM, HCM, PPM, PLM, MFG, FIN, MES, Databases, and Big Data Applications.</p>
Validation	<p>It is important to validate the data loaded into the target and report Object wise and table wise, Record Count and Check sum details.</p>
Verification	<p>It is important to use the converted data in the target application and perform 2 to 3 important transactions. For example: When Material master is converted, verify process will help in creating a BOM or Routing using the converted material, also perform Inventory transactions using the converted material.</p>
Reconciliation	<p>Multiple levels of reconciliation reporting should be utilized. (i) A pre-load reconciliation that compares data between the source system and the transformed data to confirm the right transformations have been applied. (ii) A field level reconciliation report between the source and target systems. This enables the project team to validate the successful migration of data elements. A functional level reconciliation report should also be generated. This essentially has comparison of functional data models between the source and target systems and enables the business users to validate the migration including the corresponding transformation.</p>

A Proven Approach to Enterprise Data Cleansing and Migration

Audit It is important to have an “audit trail” for all data migration executions, with number of records failed and successfully processed, and reasons for failure. Errored records should be re-processed after the necessary corrective actions through an iterative process.

An “Audit Trail Summary”, which gives a view of the percentage of data successfully migrated, errors, and the party responsible for fixing the errors is a must-have and gives a synopsis of the end-to-end data migration.

Summary:

The process outlined in this paper can be utilized to effectively ensure against the pitfalls of migrating unclesed data, which can lead to costly errors, frustration on the part of business users and impair the decision-making process.

Another important area that requires focused attention is Post Go-Live Data Maintenance. Available tools can be utilized for on-boarding new master data using pre-defined adaptors that help simplify the process and maintain data quality. It is also important to institute data governance policies to assure that existing master data stays clean. A tool-based approach is recommended since manual efforts are tedious and prone to failure, especially for large Enterprise environments.