DATA MIGRATION SIMPLIFIED

AN INTRODUCTION TO MASTER AND TRANSACTION DATA MIGRATION

CHARLIE MASSOGLIA & RATHINA KANNAN

'A MUST READ FOR DATA ENTHUISIASTS' - AUSTIN DAVIS

About the Authors

Charlie Massoglia

VP & CIO, Chain-Sys Corporation

Former CIO for Dawn Food Products For 13+ years. 25+ years experience with a variety of ERP systems. Extensive experience in system migrations & conversions. Participated in 9 acquisitions ranging from a single US location to 14 sites in 11 countries. Author of numerous technical books, articles, presentations, and seminars globally.

Rathina Kannan

Product Manager, Data Migration and Integration Platforms

A Data Migration & Integration expert. Has designed and delivered solutions for various complex migration, conversion, and integration projects. Managing software products design & development.

Data Migration Simplified

An Introduction to Master and Transaction Data Migration

By

Charles L. Massoglia

VP & CIO

Chain●Sys Corporation

charles.massoglia@chainsys.com

and

Rathina Kannan

Product Manager

Chain●Sys Corporation

rathinakannan@chainsys.com

ISBN 978-0-9990073-7-2 (pbk); ISBN 978-0-9990073-6-5 (ebk)

10 9 8 7 6 5 4 3 2 1

Printed in the U.S.A.
First printing, September 2017

Table of Contents

Chapter 1 - Introduction

A Data Migration project could be a standalone movement of data from one system to another. It might be part of a software or database upgrade. It could be part of a conversion from one software system to another such as a conversion from a legacy system to SAP or Oracle. Or it might be part of a consolidation of databases from multiple standalone systems into a single enterprise system.

Data migrations are frequently the result of mergers, acquisitions, and divestitures.

Data migrations are typically complex projects requiring thorough planning and testing. These projects cannot be restricted to data migration alone. They should include Data Quality steps to ensure that only clean data gets migrated. Conducting a Data Migration project without taking into consideration data quality is likely to provide less than satisfactory results.

And further, Data Migration projects should be quickly followed by Master Data Management projects, in addition to

Data Quality Initiatives. Spending all the time and money to cleanse data for a Data Migration project only to let data quality deteriorate once the migration is complete just does not make sense.

As with any large project you need to use some type of formal project management methodology such as traditional, PMI, Agile, or any other methodology with which you are comfortable. The point is ad hoc or brute force approaches are less likely to succeed because of the complexity involved in Data Management.

The use of an ETL (Extract, Translate, Load) Tool can greatly facilitate the ease with which a Data Migration project can be accomplished and substantially reduce the time required. A Data Migration tool that predefines source and target system fields and the required transformations would be even better.

Data Migration Phases

Data Migration projects are typically comprised of a number of phases, with many of them overlapping. The phases themselves and how much time you spend on each of these phases will depend on the overall scope of the project and what type of tool(s) you are using. Perhaps the most complex and time-consuming Data Migration projects are likely to be migrations from a legacy system to a new enterprise system or migrations between different enterprise systems such as SAP to Oracle and vice versa.

The phases of a Data Migration project are not generally sequential. You don't have to wait until one phase is complete before starting the next. The phases will overlap and the process should be iterative.

1. Source System Data Analysis and Mapping Source to Target
2. Extraction of data from source system
3. Data Quality Management - Cleanse and Transform
4. Load into target system
5. Reconciliation of target to source system

Source Data System Analysis and Mapping Source to Target

Performing in depth data analysis on the source system is a critically important step. Investing time in this phase will save you much more time in subsequent phases and increase the likelihood of a successful migration.

Define the source(s) of information to be migrated. If you are migrating from an existing ERP or other enterprise system, you cannot assume that it is your only input. You probably have existing data interfaces to the Source System (in bound and out bound). The interfaced systems may or may not be on premise. So, you need to ensure you include Applications such as SuccessFactors, Kronos, Oracle Siebel CRM on Demand, Salesforce.com, and any other on premise or cloud based system you use above and beyond the core system.

Be thorough. IT may not even be aware of systems with which users exchange information. Don't forget Excel based data extracts and loads. You may even find word processing documents, Access Databases, and text files being used, that may not work well with the new Target System. Don't overlook any possible source of data utilized by business users.

Business users should be involved in all phases of the Data Migration project. They are most likely to know how the data is currently being used and what interfaces exist, beyond the knowledge of IT.

You will need to define the attributes of the source data and how that source data will be transformed into target data. At a minimum, you should include the following:

- Source Table Name (or File Name)
- Source Column Name (or Field Name)
- Source Data Type
- Source Data Validation Rules
- Source to Target Transformation Rules
- Data Owner
- Target Table Name
- Target Column Name
- Target Data Type
- Target Validation Rules
- Target Value Mandatory

Unless you are using a tool that has already defined your source system data, as data sources are identified you will need to identify the structure and attributes of all data to be migrated. Even with a tool that has pre-defined source system

data, there will probably be additional data sources that will need to be documented.

You also need to document how the source data is used including valid values and the meaning of those values.

As the structure and attributes of source data are identified, the source data should be mapped to target data. Special care should be given to identifying data that is mandatory in the target system but is optional in the source system. Data that the target system requires that does not exist in the source system may be a challenge. And you will need to decide what to do with data defined in the source system that is not defined in the target system.

Data Profiling is a process of examining source data and collection of statistics and information about that data. Statistical measures such as mean, median, average, minimum, maximum, sum, frequency distribution, etc. are determined. Data Profiling can help you get a handle on how the data is being used, how much data will be migrated, and the level of effort required to cleanse data.

If 99% of the values of a field, range from 0 to 14, but 1% of the values are greater than 700, this is a pretty good indication of a potential data quality problem without even knowing how the field is being used.

Data Profiling will also show you how often a field is populated with a value as opposed to blank or null.

Detailed transformation rules need to be defined. A wide variety of transformation rules might be utilized:

- Simple substitution such as change '1' to M' for Male and '2' to 'F' for Female or change blanks to zeroes.
- Variable substitution such as changing "Road" to "Rd".
- Table based substitution by looking up the source value in another table and retrieving the target substitution value.
- Arithmetic operations such as divide by 100 to change an implied two-digit decimal in an integer to floating point.
- Default values for fields that did not exist in the source system or were optional on the source system but are mandatory for the target system.
- Date format changes such as changing MMDDYY dates to YYYYMMDD format.
- Time format changes such as changing 6:00 PM to 24-hour time as 18:00:00.
- Parsing a title, first name, last name, and suffix from a single name field.

These are just a few examples of simple transformation rules. Other rules can be far more complex.

While some are tempted to track all of the source information and transformation rules in Excel, it is far more efficient to use

a tool created for that purpose – preferably part of a Data Migration Suite of tools.

Extraction of Data from Source System

Many Data Migration tools extract data from the source system to a staging area, work on the data in the staging area, and then load the cleansed and transformed data into the target system. One of the reasons for adopting this method is to preserve at least half of the investment made in a Data Migration project for future migrations.

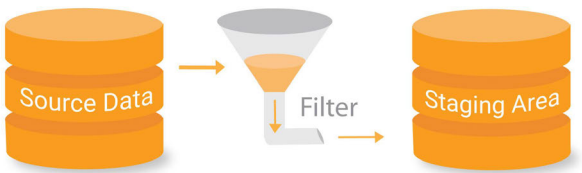
The source data extracted to the staging area might be in the original source system format or in a common intermediate format. Data is mapped from the source system to the staging area and then from the staging area to the target system.

If a common intermediate format is used and a second data migration is required from a different source, all that needs to be done is mapping from the source system to the staging area. Almost all of the work invested in cleansing and transforming the data in the staging areas and loading to the target system is preserved.

Further, extracting data to staging areas is less likely to have an impact on the source system while working on the data before loading to the target system.

Before migrating your data, you should attempt to limit the amount of data to be migrated by filtering out obsolete data to maximize performance and keep the target system clean

and nimble. The filtering could be done during the extraction of data from the source system to the staging area or during the load of data from the staging area to the target or both. There are some advantages of filtering when extracting from the source system to the staging area as it makes reconciliation to the target system easier and minimizes potential data cleansing issues with very old data.



If a customer has no business transactions with us for the past 10 years, is there a reason to retain the customer unless they were created recently? Or if an item with a zero on hand balance has not been ordered, used, or sold in 5 years and has not recently been created, do we really want to migrate that item? In both examples, dependent and historical data would also need to be omitted such as sales history or bills of materials.

Another way to reduce the amount of data being migrated is to purge or archive data that is not being used for analysis. If the data is unlikely to ever be used again, it can simply be purged. If access to the data may be required in the future, it

could be archived. Archiving of data could be done in the source system or could be done in the staging area or in a separate data warehouse. Retaining the archive in the source system would require you to keep the old system running just to do inquiries and reporting on the archived data. Keeping the archive in the staging area or a separate data warehouse would enable you to query historical information without having to retain the old system.

Data Quality Management (DQM) - Cleanse and Transform

Simply put, Data Quality Management is the process of ensuring data is fit for business use. Once the source data has been extracted into a staging area, data quality activity can take place. There are a number of processes that fall within the realm of Data Quality Management including the following:

- Data Standardization
- Data De-Duplication
- Data Cleansing
- Data Harmonization
- Data Consolidation
- Data Enrichment

Data Cleansing, also called Data Scrubbing, is the process of identifying and correcting errors in data. These errors might be invalid values in an individual field such as a field containing

the value 'X' when only '1' and '2' are valid values. They might be link errors such as an item that does not exist in the Material Master but is still used in a Bill of Materials or a sales order line with a customer number that does not exist in the Customer Master file. Some of the data errors can be corrected programmatically. Others will have to be corrected manually.

Missing data is also a Data Cleansing issue. Some legacy systems may treat a field as optional while the target system treats the field as mandatory. Or the source system may not store data required by the target system. Depending on how the data is used, a mass update of a default value may be done. In other cases, a default value may be generated based on the value of another field. Or the missing data may have to be manually entered.

Data Standardization is the process of ensuring that like data is stored in a common format. For example, the source system might store addresses as simple text fields while the target system might separate the values into street, city, state/province, zip/postal code, and so on.

Variations of street address, city, and state/province should be standardized into acceptable postal service format. Zip codes should also be standardized:

200 East Grand River Avenue, E Lansing, MI 48823

200 E GRAND RIVER AVE, EAST LANSING, MI 48823-2004

There are a number of services available that use external sources to standardize street addresses. These services can also be used to enrich data during the migration, e.g. adding zip+4 to US addresses. And National Change of Address utilities are available for the US and Canada to automate corrections of address when the party has moved and filed a change of address form.

Phone numbers should also be standardized. Some systems do not provide for a separate country code such as 49 for Germany or 44 for the UK. Others do not enforce any type of standardization so all of the following might be valid representations of the same phone number: 517-555-1212, (517) 555-1212, 517.555.1212, 1-517-555-1212, and 5175551212

Data De-Duplication is the elimination of duplicate records, e.g., elimination of duplicate customers, vendors, items, and other master data. De-duplication can be more easily accomplished once the fields used to match records are standardized. While standardization of addresses can facilitate detection of duplicate addresses, there are a number of tools which are much more effective using far more sophisticated algorithms. Once duplicates are detected, a decision can be made whether the addresses should be combined and if so, which address will be retained.

Standardization of names can also be used to detect duplicates and combine names. This process is much more difficult without a tool that has predefined language

dependent synonyms such as Robert and Bob which should be treated as a match. Again, once duplicates are detected, a decision can be made whether the names should be combined and if so, which version of the name should be retained?

When combining duplicate customers, vendors, item, and other master data records, it is important to retain a history of the records which were combined into a single master record, e.g. customer numbers 10456, 12033, and 18990 were merged with customer 17916. Of course, any transaction or other associated data that contains these merged customer numbers will need to be changed to the retained customer number. Special care will be required if summary tables are involved.

Data Consolidation can refer to the combination of data from multiple data sources. When consolidating customers, vendors, or other master data from multiple sources, you may encounter duplicate keys because master data might have been maintained independently on each source system. So, if you encounter the same customer number for two different customers in two different source systems, it is also important to retain a history of the change made to one of the numbers to make it unique. Of course, any associated data containing the changed customer number must also be modified. When two sources have same Customer record, based on criteria or Business Rules, Source “A” Customer record may be chosen over Source “B”, or vice versa, to become the Target record. Or the Target record may be constructed from parts of Source

"A" record and parts of Source "B", again based on Business Rules or manual intervention.

Data Enrichment is the process of collecting additional master data information. Enrichment of addresses to standardize them in approved postal service format has already been mentioned. Data Enrichment can also involve capturing data that is missing or not currently stored such as adding D&B number to customers and vendors. Geocoding of addresses would be another example. There are a number of external services that offer this type of service.

The Master Data Harmonization is used for improving overall quality of the data throughout the business landscape by doing:

1. De-duplication of data
2. Distribution of consolidated data to connected client systems.

Data Harmonization is the process of ensuring after de-duplication that all systems are consistent when the same data appears in multiple places. An example might be ensuring customer names are the same in an ERP system such as SAP, or Oracle as well as an associated CRM system such as salesforce.com.

All of the data quality issues should be addressed before data migration. Cleansing data after a migration is typically much

costlier. Data can be cleansed in source system or in the staging area.

Load Source to Target

This is the actual load of data to the target system. Transformation might take place from the source staging area to a temporary staging area. Or the data might be transformed as it is loaded into the target. The former process permits all validations to be done before anything is loaded into the target system.

Filters might be employed to reduce the amount of data to be migrated by not extracting obsolete data from the source staging area to the target system. As previously mentioned, there are advantages of filtering when extracting from the source system to the staging area. In any event, filters could be used in both steps if required.

Note also that sequence of loading becomes important. Obviously, you cannot migrate your Bills of Materials before migrating your Material Master.

Many software vendors provide standard interfaces to load data into their systems such as APIs for Oracle and BAPI s for SAP. It is almost always a better technique to use supported interfaces to load data than to load directly to tables in the target system.

Depending on how you do data validation, loading source to target may be an iterative process. General the target load is first done to a test environment. As you encounter load errors, you will need to go back into the staging area and make corrections or make configuration changes in the target system. Once corrections or changes have been made, the target load can be rerun.

If Data Quality techniques are properly applied, almost all of the errors should be detected and corrected before the production target load.

Reconciliation of Target to Source System

When performing a data migration, it is important to ensure that all data that should have been converted was converted. Depending on how you structure the data migration process, there are actually two reconciliations which should be done.

The first reconciliation is source to staging area to ensure all of the data that was extracted from the source system was successfully loaded into the staging area. This is not simply a record count. Numeric totals should also be reconciled, e.g.

total outstanding A/R in the source system matches total outstanding A/R in the staging area.

The second reconciliation is staging area to target to ensure all of the data in the staging area has been successfully loaded into the target system after de-duplication, cleansing, consolidation, and transformation. This reconciliation is made more complex because you will not be loading all of the data from the staging area into the target system if you have done de-duplication, purging, or archiving. So, you cannot do simple record counts. Numeric totals should still work because you would neither drop a customer with an open A/R balance nor drop an item with on-hand inventory. If you have obsolete records with non-zero balances that you do not want to migrate, you would first create an offsetting entry in the source system to zero out the balance.

Now that we have reviewed the phases of a Data Migration, let's review some of the challenges that are likely to arise and what can be done to mitigate the risks involved.

Chapter 2 - Data Migration Challenges

Data migrations present any number of challenges many of which apply to both a standalone migration of data from one system to another as well as a data migration which is part of a software system conversion such as conversion of a legacy system to SAP or Oracle. Being aware of and addressing these challenges in advance is critical to a successful data migration.

Treating a Data Migration as an IT Project

Data migration projects are not IT projects. It is the business users who really understand the data. Subject matter experts need to be identified and engaged from the onset of a data migration project.

The creation of data mapping and transformation rules should involve business users. They are most likely to understand the nuances of data in the source system.

Business users should be in integral part of data testing. After all, it is they who will be the ones using the data once it has been migrated and ultimately, they will decide if the data migration was successful.

All too often users do not report data issues because it sometimes takes an inordinate amount of time to get a response from IT while the users have jobs to get done.

Users are ingenious in coming up with ways of circumventing data issues. While a text field may theoretically be just a generic description, the users will know if the first 6 characters of a buy/sell item description represent the brand of the item. Or if the weight and pack size of an item are contained in the item description. Or that an unused field was used to store email addresses because the legacy system did not provide for this.

It is one thing to understand the technical aspects of the database in the source system including validation rules, which tables contain what data, dependencies between tables, etc. It is entirely another to understand how the users are actually using the data.

Understanding how the data is actually being used as opposed to how it was supposed to be used is critically important to data migration success.

Underestimating Resources Required

A data migration is not just a simple data conversion project. There is far more to this type of project than simply mapping fields and transforming data.

It can be very difficult to estimate the time required for a data migration before performing detailed analysis of the source system and identifying any gaps in data required by the target system. Further, until Data Quality is assessed, estimating the time required to cleanse data is extremely difficult.

There is no question using a Data Migration tool will reduce the resources required. But adequate resources still need to be allocated. Relying too much on a tool can lead to a host of problems one of which is an underestimate of time required.

As a general guideline estimating one day per column, field, or attribute to be converted should provide an order of magnitude estimate. Again, using the right tool can substantially reduce time required particularly in source system analysis and mapping source to target.

A large percentage of Data Migration projects come in over budget (as well as late) due in large part to underestimating the resources required. It is far easier to reduce an overestimate than it is to increase an underestimate.

Lack of Necessary Skills

It is extremely difficult to successfully plan and execute a Data Migration project without involving people with recent experience in such projects. Education will certainly help and should be done, but it is no substitution for experience. On a large Data Migration project such as a conversion from one system to another system, you should consider involving external contractors or even hiring someone with the requisite experience if internal personnel with recent experience are not available. If you do utilize external resources, a formalized plan for knowledge transfer should be in place when engaging those resources.

Often when converting from a legacy system to a new enterprise system, in depth technical knowledge of the legacy system will be lacking. Utilizing a tool that has predefined the mapping from the source to target system can go a long way to addressing this issue. And before assuming there is inadequate knowledge of the legacy system, don't forget to involve highly experienced business users. Users who are able to successfully create ad hoc reports are likely to have extensive knowledge of the source database.

Poor Data Quality

Poor data quality in the source system is a frequent problem in data migrations. Often this issue arises even if there is not a perception of any data quality issues in the source system. SAP and Oracle data validation rules are often more stringent than those in legacy systems. Data that worked fine in a legacy system may cause problems or unexpected results in SAP or Oracle.

Missing data is another issue that may have to be addressed. The source system may have defined a field as optional while the target system considers it mandatory. Or the source system may not even have a place to store information which is required in the target system. This can be a particular problem in conversions from legacy systems.

An "as is" assessment is the first step to determine the quality of data in the source system. Again, performing a detailed

assessment is essential to determine the resources required to successfully complete a data migration.

Data migrations from source systems in which IT has directly modified data without going through the application layer can be particularly troubling. If SQL was used to delete obsolete item master records so they could no longer be ordered but bills of materials, routings, or other references to those deleted items remain, issues will arise when attempting to migrate these data to SAP or Oracle.

Dates are another issue that will need to be addressed. Some legacy systems treat dates as numeric fields without necessarily enforcing date validation rules.

No matter how much time and money is spent, Data Quality will never be perfect. You have to set realistic goals for Data Quality. If you insist on perfection, you will never complete your Data Migration project. On the other hand, failure to adequately address Data Quality issues is likely to lead to a less than satisfactory Data Migration.

Again, there should be no standalone Data Migration projects without Data Quality projects that feed cleansed data to a Data Migration projects.

Deduplication and Data Consolidation

When consolidating data from multiple systems into a single target system, the issue of duplicate keys must be addressed.

If multiple divisions run their own instance of an ERP system and if each division has the ability to create new customers, vendors, items, or other master data, you will almost certainly encounter duplicates.

First, the potential duplicate master records will need to be identified. These duplicate records may or may not have the same key. Then a decision will have to be made regarding which master record is to be retained and which should be merged into the retained record. Associated master and transaction data of the merged records will need to be modified to reflect the key of the retained record.

So, if three customer records are identified as duplicates, the associated accounts receivable, sales orders, sales history, and any other record containing the customer number of the merged records will need to be changed to reflect to customer number of the retained record.

In a similar manner, if multiple item master records are identified as duplicates, the bills of material, recipes, sales order line detail, purchase order line detail, and any other record containing an item number of the merged records will need to be changed to reflect the item number of the retained record.

Potential duplicate records should be identified even if the data comes from a single source system. An additional step needs to be taken if the data comes from multiple systems.

If the master records have the same key, the records may not necessarily be duplicates. So, if customers are independently created on multiple systems, the same customer number could be assigned to two completely different customers. Once true duplicates are eliminated, the duplicate keys will have to be addressed.

All but one of the records with a duplicate key will need to be modified to change the key to a non-duplicate value. So, if two different customers created on different systems have the same customer number, one customer number will have to be changed to a non-duplicate custom number. And again, the associated master and transaction data will need to be modified to reflect the non-duplicate key.

To avoid mistakenly using the original customer number of a changed customer that now represents a different customer in the merged system, it may be prudent to change all of the keys involved. So, if customer 1001 represents Customer A from one source system and customer 1001 represents Customer B on another source system, customer 1001 representing Customer A might be changed to 10011 and Customer B might be changed to 10012 so the original customer number 1001 is not used at all in the consolidated system.

Historical Data

One of the keys to facilitating a smooth Data Migration is to migrate the minimum amount of data required by the business. Consolidating duplicate records is certainly one part of accomplishing this. But addressing obsolete historical data has even a greater impact in reducing the amount of data to be migrated.

The older the data the more likely data quality issues will arise. Why migrate data if it is so old and has no use unless there is a specific business need such as regulatory requirements, warranties, etc.? Many companies have spent an inordinate amount of time addressing data quality issues on historical data that will never be used.

Many companies do not routinely archive or purge historical data leading to much larger databases than are really necessary to provide the company with required information. Further, performance can be impacted by having to read thousands if not millions of records only to discard the vast majority of them. This can also lead to increasingly expensive hardware requirements both in terms of performance and storage.

To minimize the amount of data being migrated a decision needs to be made as to how many years of historical information are to be retained. Historical information older than what has been deemed useful should be purged on the

source system or dropped when loading the target system from the staging area.

Consideration should also be given to archiving information that is not frequently used. Again, the archiving could be done on the source system or in the staging area. But using a source system archive would require the source system to be retained to make the archived information available.

Failure to Identify All External Data Sources

When converting from one enterprise system to another, it is important to ensure you have identified the source of all data. Just because the data resides in the source system does not mean it was created in the source system. The data could have been uploaded from a manually entered Excel template or from another system such as Kronos.

The data in the source system can still be migrated no matter where it was created. But once the target system is in place, you will still need to modify the processes that import the data, reformatting the data from the external sources to meet the requirements of the target system.

This is one of the issues that can easily be missed until after the target system is in place. Involving the right business users can minimize the risk of missing a data source.

Failure to Identify all Data Integrations

Identifying all of the systems with which the source system or target system interacts is critical when converting from one enterprise system to another. If an enterprise system exchanges data with an external system such as salesforce.com, the interface will have to be modified to take into consideration the format of the data in target system.

And it is not just pointing the interface to the new target system. The meaning of attributes may have changed such as M=Male in source system while 1=Male in the target system.

Virtually every enterprise system in use today exchanges information with another system. You have to ensure you have the ability to continue to exchange that information once the data migration is complete.

Inadequate Testing of Data

Testing is another area that poses challenges for most Data Migration projects. First and foremost, testing is not a task IT can perform on their own. It is virtually impossible for IT to fully understand all of the nuances of the business aspects of the data. Business users must be intimately involved in all aspects of testing.

Failing to ensure an adequate sample of data migrated for testing can lead to unexpected errors when a full migration is

done. A full data set should be made available as soon as practical to facilitate thorough testing.

A formal test plan should be created spelling out the specific business functions that are to be tested including what data needs to be entered and what results are expected. It is important to record the results of each test case. A tool such as HP Quality Center could be used but even Excel will suffice if tools are not in place.

Any manual corrections that are made to data to complete the testing should made in such a manner that they do not have to be manually made again when loading to production. Using manual processes to correct data for testing exposes a risk if these manual processes have to be repeated in production.

Testing generally will require multiple test runs. Changing Customer, Vendor, Material, or other Master Data keys between one test run and another introduces major complexities. If Customers are being consolidated into a new generated Customer Number and that number changes from test run to test run, it becomes difficult to retest if the test plan does not specify the current Customer Number.

Source to Target System Reconciliation

It is important to confirm, if everything expected has in fact been migrated from the source system to the target system. This is more than just confirming record counts match.

General Ledger, Accounts Receivable, Inventory, and other balances must be compared and matched.

If a staging area is being used, one of the keys to effective reconciliation is ensuring the source data is reconciled with the staging area as data is extracted in addition to reconciling the staging area with the target data for final migration. Failure to do both complicates tracking down discrepancies because it may not be readily apparent whether the issue lies with the extraction of data from the target system or the transformation of data from the staging area to the target system.

Lack of an Audit Trail

Additional scrutiny can be expected in an annual IT audit conducted after a Data Migration especially if it involves conversion from one software system to another. This can become quite a challenge when duplicate customers, vendors, or other master data have been merged.

Simply showing the total outstanding Accounts Receivable in the target system matches what was in the source system will likely not be adequate. Being able to clearly show that the new Accounts Receivable balance for a customer in the target system is the sum of two specific customers in the source system should satisfy the auditors.

If a new customer is created, count on auditors asking for substantiation of who requested the new customer and who

approved it. Utilizing a Master Data Governance tool with audit reports should easily address this issue.

Criteria to Evaluate Data Migration Tools

It is important to define the criteria by which you will evaluate Data Migration software and techniques. Clearly the vendor needs to demonstrate their tool can meet the requirements of your initial scope.

You certainly want to ask the standard questions you would ask of any software vendor, e.g. size of company, support structure, availability of support for all of your locations and time zones, number of customers, customer references, how is pricing determined, etc.

But there are other requirements specific to data migration that you might want to consider that could save you a lot of time:

- ✓ Are predefined templates available for source and target systems or will you have to spend the time and money to manually define source fields and map them to the appropriate target field including creating transformation rules?
- ✓ If templates are used and you have some special requirements, can custom templates be created?
- ✓ Is Master Data Governance integrated into the tool and, if so, is automated workflow supported including the ability to have several users work concurrently on

providing required data and their approval before another level of individual approves the request (parallel approval)?

- ✓ What is the level of sophistication used in detecting duplicate master records? As an example, for Customers, what besides Customer name and address is used? And what happens if the name and address do not match exactly? Can names be matched phonetically? Are synonyms for names predefined such as “Robert” and “Bob” are the same name? Is a scoring methodology used to show how closely two records match? Are “Thomson” and “Thompson” shown as close matches?
- ✓ How are address standardization and validation handled? Can the tool use external data sources to validate and standardize addresses?
- ✓ Can DUNS number and other information can be added from external sources?
- ✓ Will the production target system validation rules be enforced when correcting data prior to migration to the target system?
- ✓ Can you define your own validation rules?
- ✓ Can you incorporate your own business rules into data transformation?
- ✓ If there is a discrepancy when reconciling source to target, are drill down capabilities available to narrow down to the transaction(s) causing the difference?
- ✓ Can you configure a display or update screen containing information from both the source and target systems at the same time?

- ✓ Can you configure a display or update screen containing information from multiple views?
- ✓ Is any programming required and if so, for what?
- ✓ Is the software certified by your enterprise software vendor, e.g. SAP®, Oracle®, Salesforce®, etc.

Once you have met with the tool vendors and taken a look at their products with a brief demo customized to your requirements, ask for a POC (proof of concept) proposal. How much a vendor is willing to work with you to give you a demo that is tailored to your requirements and to develop a POC that will prove their solution will meet your current and future requirements certainly says something about that vendor and how they are likely to work with you in the future.

Chapter 3 Chain•Sys Can Help

A word of caution about this Chapter: This is the start of a sales pitch but it will give you a basis for tool comparison.

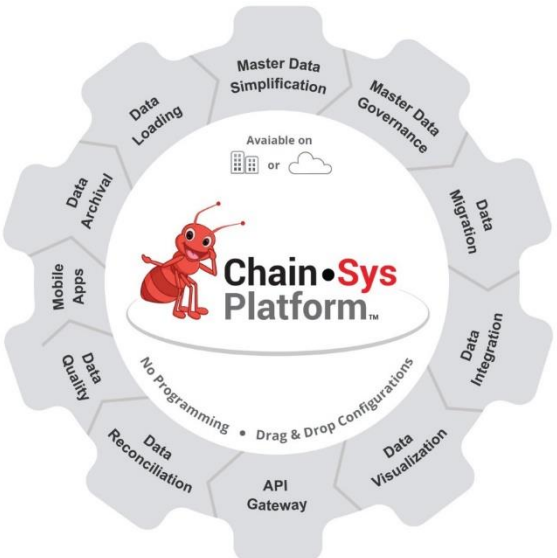
The Problem

IT and Business Users face a multitude of data challenges during SAP, Oracle, and other enterprise system implementations, re-implementations, upgrades, and on-going production maintenance including the following:

- Profiling and Cleansing of data in Legacy Systems and older SAP or Oracle versions and migrating that data to a current version.
- Reconciliation of migrated target with source data.
- Data integration of SAP ECC6, SAP S/4HANA, Oracle eBS, Oracle Fusion/Cloud Applications, and other enterprise systems with other enterprise and cloud applications.
- Maintain data cleanliness in SAP® ECC6, SAP® S/4HANA, Oracle eBS®, Oracle® Fusion/Cloud Applications and other enterprise systems.
- Simplify Clean Master Data Creation and establish Governance.
- Analytics, ad hoc reporting, and drill down dashboards.
- Bulk creation/updates of master/transaction data.
- How to get clean data into an SAP or Oracle environment in cases of data corruption.
- Testing resulting from Release Management (Monthly/Weekly Patches and Custom Code).

The Chain●Sys Solution

The Chain●Sys Platform Data Management Suite provides an integrated solution to address most of these challenges. You can utilize the entire suite for a complete end-to-end Data Management solution or pick and choose the components you require to address specific challenges.



Instead of having to integrate one solution for Data Migration and a separate solution for Data Governance, the Chain●Sys Platform components are tightly integrated while still

permitting you to use only the components you require. The tool is available in on-premise and cloud versions.

All of the capabilities of the platform can be utilized using drag and drop configuration with no programming required.

The Chain●Sys Platform Data Management Suite is a next generation tool providing more than 2,000 pre-configured templates that can save a tremendous amount of time in a Data Migration.

A large number of Enterprise Applications are supported using the template approach including the following:

- SAP® ECC6
- SAP® S/4HANA
- Oracle JDEdwards®
- Oracle eBS® R10, R11, & R12.
- Oracle® Fusion/Cloud® Applications
- Infor LX (BPCS), XA (MAPICS), LN (Baan), Lawson, and Aurora (JBA System21)
- Salesforce®
- Oracle Peoplesoft®
- PROCORE®
- Oracle Siebel®
- Microsoft Dynamics® NAV, AX, GP, and SL
- Excel, XML, Text and Queue

Why Chain●Sys?

Chain●Sys is an SAP Silver Partner and an Oracle Platinum Partner providing business organizations with global technology solutions and consulting services. Established

in 1998 with headquarters in Lansing, Michigan and Chennai, India, Chain●Sys is a fast-growing company with product solutions and services supported by a professional staff of more than 800 employees with operations in the US, Canada, UAE, Singapore, the Netherlands, and UK. Chain●Sys' success can be largely attributed to its focus on its SAP, Oracle, and other partnerships, industry and process knowledge, project management experience, SAP and Oracle Implementation expertise, and the Chain●Sys Platform of data management tools.

Chain●Sys has successfully performed over 200 Data Migrations and over 150 ERP Implementations for a wide variety of clients globally.

Examples of some of the types of clients who have utilized Chain●Sys are shown below:

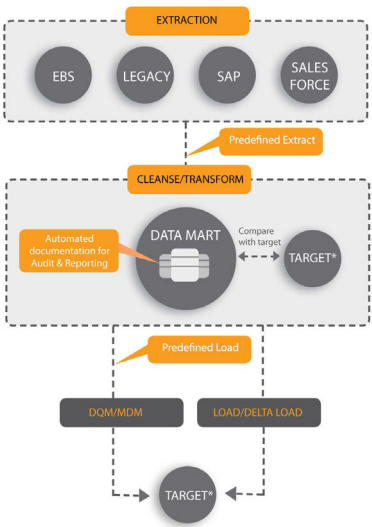
- Leading Imaging and Electronics
- Tier 1 Automotive Suppliers
- Document Management and Security
- Heavy Engineering and Mining
- Auto Giants
- Payroll Processor
- Construction Majors
- Leading Edge Pharmaceutical Companies
- Medical Devices and Equipments
- Lead Travel Sites
- Oil and Gas
- High Tech

appMIGRATE™

Let's briefly review some of the capabilities of the Chain●Sys Platform™ related to data migration. appMIGRATE™ provides an end-to-end data migration process, covering extraction from the source system, data quality components to clean, the rules engine to transform, loading capabilities to the target system, and reconciliation to provide full transparency and an audit on what transpired.

appMIGRATE

appMIGRATE™ facilitates many high-level business initiatives such as mergers, acquisitions, amalgamations, demerger, divestitures, transformation of chart of accounts, etc. No programming is required.



* SAP ECC, S/4HANA, Oracle Fusion Cloud Applications, E-Business Suite and other Enterprise Applications

Data Extraction and Data Mart

appMIGRATE™ performs data extraction to a data mart upon which all Data Migration activity will take place. Default templates allow you to quickly build a data mart. Multiple

domains are supported with default domains provided. Custom domains can also be built.

To simplify and speed the data migration process, appMIGRATE™ uses a template approach to build a central data hub and import data from one or more source systems. The templates will deal with extraction of both Master Data and Transaction Data (both open and historical).

The template approach can substantially reduce time required over other approaches because you don't have to manually map data from the source to the target system. With more than 2,000 predefined templates, you can build hubs in weeks instead of months and start testing data in a matter of weeks.

Custom templates can be configured to support data migration from/to virtually any system or file format. Again, no programming is required.

Data Profiling

appMIGRATE™ performs data profiling on the source data to reduce the data volume and explore data quality issues. Data profiling is done with the extracted data against the target application's reference data and setup data dependencies with predefined rules.

Profiling source data improves data quality and reduces data quantity. Profiling helps understanding data challenges early in any data intensive project, so that project surprises are

avoided during later stages of the data migration project thereby helping avoid project delays and cost overruns.

Data Profiling enables you to assess the data quality of each field in the data hub automatically analyzing for completeness, uniqueness, value distribution, range, and pattern with results shown on a Data Assessment report.

Data Maintenance

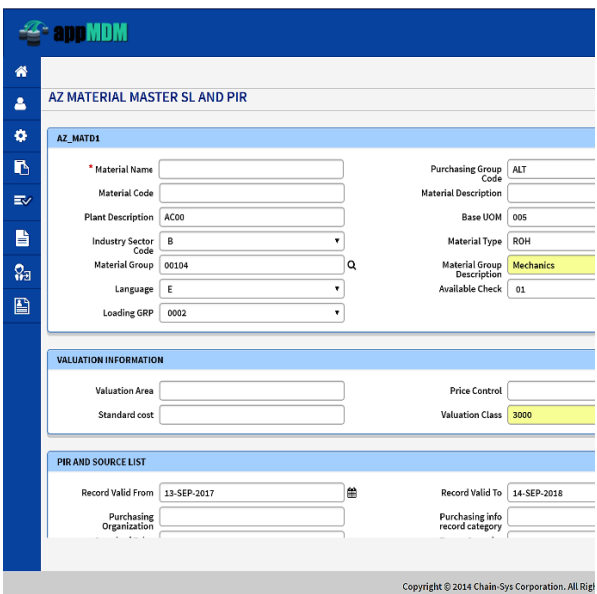
Data Maintenance is a key part of the appMIGRATE™ "Get Clean, Stay Clean" solution strategy. appMIGRATE™ provides over 2,000 predefined templates for maintaining Setups, Master Data and Transactional Data (Open and Historical Data).

Some of the key benefits of predefined templates are:

- Configurable adapters for specific business needs.
- Business users can perform data maintenance tasks, with no programming needed.
- Scalable architecture to handle very large volume data migration scenarios – 1 million records could be loaded in a single shot.
- Hub and Spoke architecture
- Capability to handle numerous data formats
- Uses robust Java/J2EE technology
- Automated audit trails for financial control
- Deep template coverage, supporting 100s of data flows

Master Data Simplification

To make things even easier, maintenance screens can be configured incorporating fields from multiple views with no programming required. The following screen shot shows an example of combining fields from Material Master basic view as well as PIR and Source List on a single screen.



Changes made can be pushed to the target system as they are made or at a scheduled time. In addition, source and target

fields can be configured on the same screen if desired. Validity checking from the target system is enforced. User defined business rules can also be configured.

Data Quality Management

Master Data Governance is tightly integrated into the solution to continuously ensure the quality of data using a "Get Clean – Stay Clean" philosophy. Workflow functionality supports data governance and approval processes.

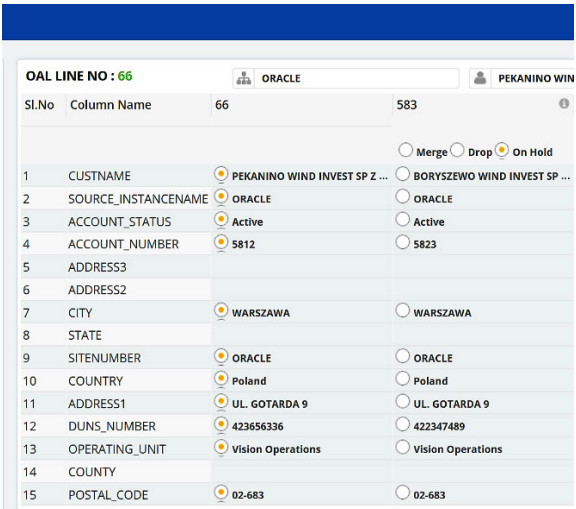
Data Cleansing (also called Data Scrubbing) can clean data based on user defined rules in a Business Rules Engine (BRE). Depending on the system you are using, validation rules from SAP® ECC, SAP S/4HANA®, Oracle eBS®, Oracle® Fusion/Cloud Applications and other Enterprise Systems can be enforced. Mass changes are supported.

Where necessary, maintenance programs can be configured to allow errors to be corrected manually if desired with no programming required.

Data Consolidation

Data Consolidation examines data from multiple sources using sophisticated algorithms to determine a potential list of matches.

Data Owners can classify them as "False Positive" or "False Negative" and perform merge, drop, or migrate options for each set of matches. Elimination of duplicates reduces the amount of data that needs to be evaluated and migrated.



In the above example, two customers with different names have been identified as potential duplicates.

Data Archiving

One of the basic guidelines for Data Migration is to limit the amount of data to be migrated leading to the requirement to purge or archive data. The templates handle the migration of historical data which can be migrated to a separate data mart, migrated into a target database, or purged. Instead of having to run an archive program in the source system which may never have been run before, the obsolete historical data can easily be archived in the data hub. Archived data can then be

made available using visualization tools (appVISUALIZE™) without requiring the source system to remain operational upon completion of the migration.

There can be a variety of reasons for filtering the extract from the source system. Data could be outdated on the source system. Or a divestiture or carve-out might require only a subset of data to be migrated.

Business rules can be established to filter data extracted from the source system. For example, an organization might not want to migrate obsolete data. A rule could be configured to filter out items that have a zero quantity on hand and have had no activity in the past two or more years. In a similar manner, customers and suppliers with a zero-balance due and no activity in a specified period of time could be filtered out.

These filtration rules can be configured via the front-end GUI of appMIGRATE™ with no programming required.

Data Reconciliation

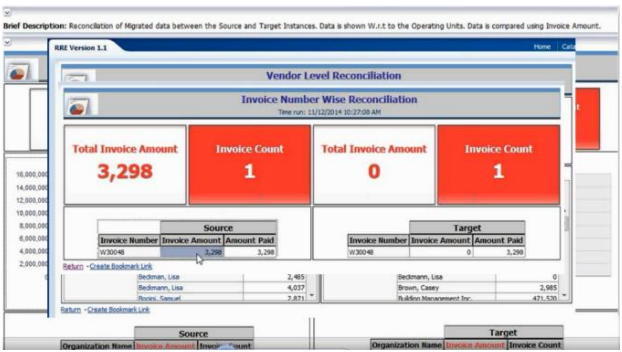
A key indicator for successful data migration is the reconciliation of the data between the Source and Target systems. The appMIGRATE™ reconciliation engine ensures and supports both functional and technical reconciliations as well as validation of the Data Migration results.

appRECON

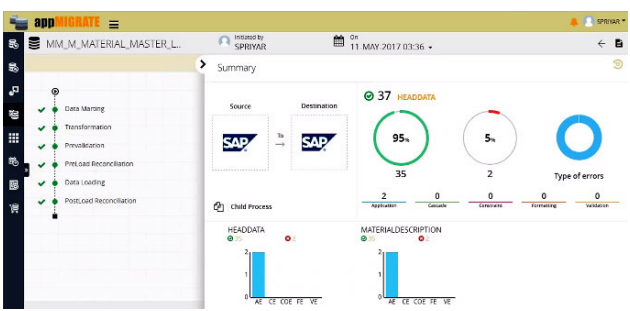
Reconciliation is done automatically from source to staging area and staging area to target. In the event there is any

discrepancy, appMIGRATE™ provides the ability to drill down into the discrepancy displaying the actual transactions causing the issue.

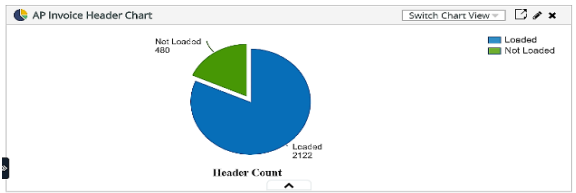
Payables Invoice Reconciliation – Drill down

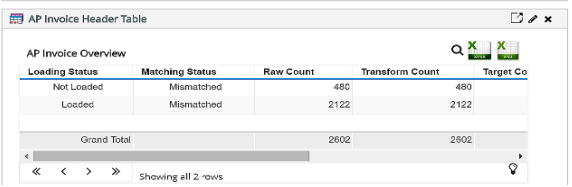


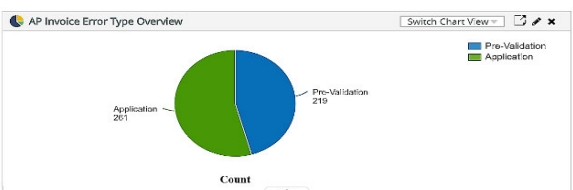
This drill down capability can substantially reduce the time needed to track down and correct discrepancies.



Reconciliation can be further augmented by using the data visualization feature, another component of the Chain●Sys Platform™, to further identify data discrepancies between Source and Target systems.

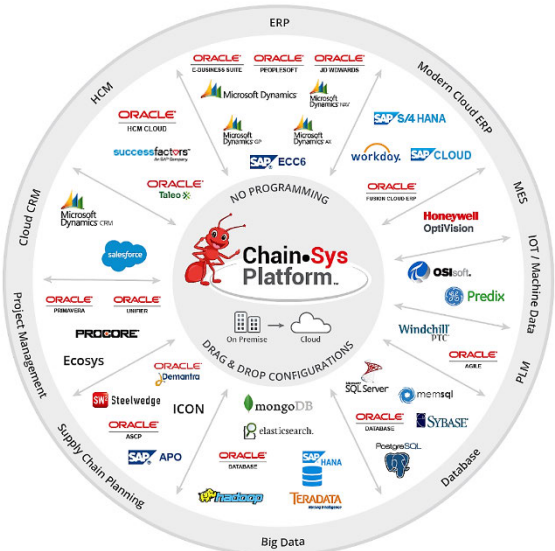






Data Integration

Integration of data with other systems is an all too often overlooked requirement in data migrations. It can be very time consuming and expensive to recreate interfaces based on the changes to the format and meaning of data resulting from a data migration.



Instead of rewriting integrations which have been cobbled together over the years, consider instead using the data integration capabilities of the Chain•Sys Platform™.

The cloud based appCONNECT ™ solution from Chain●Sys helps you to integrate multiple data sources to ensure the required data from each application flows to other applications to establish a seamless data flow across those applications giving a unified business view and benefits for the organization, thus saving unnecessary manual work and related pains and costly errors.

appCONNECT

Some of the capabilities of appCONNECT™ include:

- Supports connection to any ERP system and supports widely used Data Exchange Standards.
- Provides a huge set of pre-defined templates, that covers almost all the functions of an ERP system.
- Templates come with predefined pre-validation and mapping rules.
- Data Flow supports Validations & Transformations to be applied between Extracts and Loads.
- Work Flow supports combining multiple data flows using Activities and Constructs.
- Master Data Governance provides controls for maintaining Quality Issues.

So, you have the option of saving time by using predefined data interfaces instead of having to manually recreate them.

Summary

The three steps in a Data Migration scenario are:

1. Data Profiling and Cleansing (DQM)
2. Data Migration
3. Build/Adjust Data Integration/Interfaces

The templates configured in each step can be reused in subsequent steps.

The data migration components of the Chain●Sys Platform™ support its users, not only for data migration, but various key corporate level decisions like mergers, acquisitions, divestitures etc. Predefined templates save efforts and energy by providing readymade adaptors to extract and validate data from multiple data sources. Supporting data quality all along the data migration process and confirming the process using dashboard driven data reconciliation reports make this tool a single choice for all data management tasks. Continuous process management to ensure data quality management (DQM) reduces risks, ensures timeline gives total confidence to the users.

Remember, the Chain●Sys Platform™ is template based and requires no programming. Investing in the proper tool can facilitate data migration at a reasonable cost in a reasonable time period.

Key Features:

- Template Based Approach
- 2,000+ Templates
- 100s of Apps and Data Flows
- Hub and Spoke Architecture
- Data Maintenance
  - ✓ Master Data
  - ✓ Transactional Data
  - ✓ Historical Data
- Data Quality Management
- Data Profiling
- Supports Coexistence
- Supports Cross-referencing
- Workflow Driven Execution
- Smooth Cut-over
- Supports Setup Migration
- Multi-purpose Transformation
- Functional Reconciliation
- Configurable Platform

Key Benefits:

- No Programming
- Lower Ownership Cost
- High Quality Migration
- Reduced Data Risks
- Reduced Timelines & Budget
- Predictable on Time Cutover
- Quick Cloud Deployment
- Accelerated Conversion
- Data Validation Ensured
- Reconciled Audit
- Scalable & Repeatable
- Versatile Applications
- Setup Migration saves Money

Key Use Cases:

- Mergers & Acquisition
- Divestiture
- Data Conversion
- Setup Migration
- Code Migration
- Data Restructuring
- Data Cleansing
- Data Profiling
- Data Validation Ensured
- Data Archival, Purging & Retrieval.

“Master Data Simplification” Webinar

Getting control of Master Data can be quite difficult when multiple users in different locations are able to add or change Customers, Suppliers, Materials, and other master data. The problem gets worse when the master data is spread across multiple software systems, with some running in the Cloud. Users maintaining master data get confused when they are forced to go through dozens of fields on dozens of screens just to update the two or three fields for which they have responsibility. Or worse yet they have to maintain data in two different systems.

Chain●Sys Platform™ provides an end-to-end solution to the challenges of Master Data. Loading master data from multiple systems to a master hub can be simplified with over 2,000 pre-built templates for SAP® ECC, SAP S/4HANA®, Oracle® EBS, Oracle® Fusion/Cloud Applications, Salesforce®, and many other systems. And master data hubs can be built within weeks rather than months with no programming required! Show authorized users just the data they need to see or update – even if it comes from multiple systems. A complete audit trail is available for all changes. Provide a single integrated view of customers as well as consolidated reporting across multiple disparate systems. Data Governance utilizes automated workflow to simplify the master data add/change/delete process. Join us to see how Chain●Sys Platform™ supports high volume scalability and complex business validations to simplify your master data.

Please check for our webinars at www.chainsys.com.

“Data Migration” Webinar

Migrating data to SAP, Oracle, and other enterprise systems can be a daunting task for many companies. Mapping fields from the source system to the target system can be incredibly time consuming. Writing your own conversion programs can be painful and expensive. Providing data that is missing in the source system but required by the target system can be a tedious error prone task. Changes made manually in a test environment must be made again in the production environment with no guarantee the same change will be made. Scheduling cutovers that take days can have a serious impact on the business. And it can be difficult to prove to the business that the data has been properly converted and is in balance.

Chain●Sys Platform™ provides an end-to-end solution to the challenges of data migration. Data migration can be simplified with over 2,000 pre-built templates for SAP® ECC, SAP S/4HANA®, Oracle EBS®, Oracle® Fusion/Cloud Applications, Salesforce®, and many other systems. No programming is required! Supported incremental data loads can reduce cutover time to virtually zero. And reconciliation is automated with drill down capabilities. Join us to see how Chain●Sys Platform™ can ease your data migration pains.

Please check for our webinars at www.chainsys.com.

Chain●Sys Corporation
325 S. Clinton Street, Suite 205,
Grand Ledge, MI 48837, USA.
Telephone: 517.627.1173
Toll Free: 855.appLOAD
Dial Extension 202 or 205

Chain●Sys Asia Pacific Pte. Ltd.
#26-03, PSA Building,
460 Alexandra Road,
Singapore - 119 963

Chain●Sys Middle East (Branch)
#6WB-142, PO Box 371425
DAFZ, Dubai, UAE

Chain●Sys Europa B.V.
Startbaan 8
1185 XR Amstelveen
The Netherlands

www.chainsys.com

https://www.facebook.com/chainsyscorp

https://twitter.com/chainsys

https://www.linkedin.com/company/chain-sys

About us

CHAIN•SYS

A trusted data management innovator

Our suite of data management productivity tools - built on the Chain•Sys Platform™ - offers a complete, No-Programming solution for the entire data management lifecycle.

For over 19 years, customers have depended on Chain-Sys tools to eliminate the risk of data Governance, Simplification, Migration & Integration for more than 200 applications.

ORACLE Platinum Partner Cloud Standard

SAP Silver Partner

salesforce registered ISV partner

PROCORE

Demystify your Data Migration!

Many organizations continue to waste time and money struggling with Data Migration complexity. Learn how the tools and techniques described in this book enable organizations to successfully excecute Data Migration/Conversion along with Data Cleansing/Profiling in weeks instead of months!

Key takeaways

- Breaking the myth that Data Quality is difficult to manage
- No Programming setups, master data and historical data Migration
- Data and Application Coexistence. Drag and drop rapid data integrations/interfaces development
- Data Reconciliation and Visualization

CHAIN•SYS

ISBN: 978-0- 9990073-7- 2
Not for sale