CAPCO

THE CAPCO INSTITUTE JOURNAL OF FINANCIAL TRANSFORMATION

DATA INTELLIGENCE

Data quality imperatives for data migration initiatives: A guide for data practitioners

GERHARD LÄNGST | JÜRGEN ELSNER
ANASTASIA BERZHANIN

DATA ANALYTICS

50TH EDITION | NOVEMBER 2019

THE CAPCO INSTITUTE

JOURNAL OF FINANCIAL TRANSFORMATION

RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

CONTENTS

DATA MANAGEMENT

DATA ANALYTICS

DATA INTELLIGENCE

DEAR READER,

Welcome to the milestone 50th edition of the Capco Institute Journal of Financial Transformation.

Launched in 2001, the Journal has covered topics which have charted the evolution of the financial services sector and recorded the fundamental transformation of the industry. Its pages have been filled with invaluable insights covering everything from risk, wealth, and pricing, to digitization, design thinking, automation, and much more.

The Journal has also been privileged to include contributions from some of the world's foremost thinkers from academia and the industry, including 20 Nobel Laureates, and over 200 senior financial executives and regulators, and has been co-published with some of the most prestigious business schools from around the world.

I am proud to celebrate reaching 50 editions of the Journal, and today, the underlying principle of the Journal remains unchanged: to deliver thinking to advance the field of applied finance, looking forward to how we can meet the important challenges of the future.

Data is playing a crucial role in informing decision-making to drive financial institutions forward, and organizations are unlocking hidden value through harvesting, analyzing and managing their data. The papers in this edition demonstrate a growing emphasis on this field, examining such topics as machine learning and AI, regulatory compliance, program implementation, and strategy.

As ever, you can expect the highest caliber of research and practical guidance from our distinguished contributors, and I trust that this will prove useful to your own thinking and decision making. I look forward to sharing future editions of the Journal with you.

Lance Levy, **Capco CEO**

FOREWORD

Since the launch of the Journal of Financial Transformation nearly 20 years ago, we have witnessed a global financial crisis, the re-emergence of regulation as a dominant engine of change, a monumental increase in computer processing power, the emergence of the cloud and other disruptive technologies, and a significant shift in consumer habits and expectations.

Throughout, there has been one constant: the immense volume of data that financial services institutions accumulate through their interactions with their clients and risk management activities. Today, the scale, processing power and opportunities to gather, analyze and deploy that data has grown beyond all recognition.

That is why we are dedicating the 50th issue of the Journal of Financial Transformation to the topic of data, which has the power to change the financial industry just as profoundly over the coming 20 years and 50 issues. The articles gathered in this issue cover a broad spectrum of data-related topics, ranging from the opportunities presented by data analytics to enhance business performance to the challenges inherent in wrestling with legacy information architectures. In many cases, achieving the former is held back by shortcomings around the quality of, and access to, data arising from the latter.

It is these twin pillars of opportunity and challenge that inform the current inflection point at which the financial industry now stands. Whilst there is opportunity to improve user experiences through better customer segmentation or artificial intelligence, for example, there are also fundamental challenges around how organizations achieve this – and if they can, whether they should.

The expanding field of data ethics will consume a great deal of senior executive time as organizations find their feet as they slowly progress forward into this new territory. In my view, it is critical that organizations use this time wisely, and do not just focus on short-term opportunities but rather ground themselves in the practical challenges they face. Financial institutions must invest in the core building blocks of data architecture and management, so that as they innovate, they are not held back, but set up for long-term success.

I hope that you enjoy reading this edition of the Journal and that it helps you in your endeavours to tackle the challenges of today's data environment.

Guest Editor
Chris Probert, **Partner, Capco**

DATA QUALITY IMPERATIVES FOR DATA MIGRATION INITIATIVES: A GUIDE FOR DATA PRACTITIONERS

GERHARD LÄNGST | Partner, Capco
JÜRGEN ELSNER | Executive Director, Capco
ANASTASIA BERZHANIN | Senior Consultant, Capco

ABSTRACT

This article is based on the experiences gained through a large data migration and business process outsourcing project in 2019. Examining static data linked to approximately 12 million customer records spread across over 10 source systems led to the early identification of unclean data in approximately 10% of the golden source data and resulted in large-scale data remediation efforts that were necessary prior to data migration. Key takeaways and lessons learned about data quality on a financial institution's customer data are summarized here for the data practitioner, with an emphasis on applicable methods and techniques to gain transparency about an institution's overall current state of data.

1. AN EARLY EFFORT IS CRITICAL

In a data migration project, risk, in the form of target system (load) failures, usually surfaces very late. These risks are often the result of poor data quality and poor understanding of the data. Most data migration projects rely on documentation of the current state of data landscape or conversations with source data owners and business experts. This approach is insufficient, as data transformations based on assumptions that may only be valid in certain circumstances will result in data issues and load failures, even with minimal deviation from these assumptions.

Our experience of working at data warehouses of Fortune 500 financial institutions confirms the importance of early and analysis-driven data profiling in the context of data migrations. Data quality issues are bound to surface through profiling activities, and the sooner corrupt, duplicate, incomplete, or incoherent data is identified, the more time remains to cleanse and remediate source data, or to turn to altogether new sources of data that are better fit for the purpose.[1]

Indeed, Gartner research has found that organizations believe poor data quality to be responsible for an average of U.S.$15 million per year in losses.[2] In addition, the Data Warehousing Institute warns that "two percent of records in a customer file become obsolete in one month because customers die, divorce, marry, and move. The problem with data is that its quality quickly degenerates over time."[3] In addition to the immediate impact on data migrations, failing to address bad data will result in long-term loss of trust in information, an organization's most valuable asset in the wake of digitization efforts. As stated in the Gartner Data & Analytics Summit 2018, "as organizations accelerate their digital business efforts, poor data quality is a major contributor to a crisis in information trust and business value, negatively impacting financial performance."[4]

[1] Matthes, F., C. Schulz, and K. Haller, 2011, "Testing and quality assurance in data migration projects," 27th IEEE International Conference on Software Maintenance (ICSM)

[2] Judah, S., and T. Friedman, 2018, "How to create a business case for data quality improvement," Gartner, June 19, https://gtnr.it/2P9VpOI

[3] Eckerson, E., "Data quality and the bottom line," Proceedings of the Seventh International Conference on Information Quality (ICIQ-02)

[4] Friedman, F., 2018, speech at Gartner Data & Analytics Summit 2018 in Frankfurt, Gartnet, https://gtnr.it/2myYjet

Reviewing source data for content, quality, relationships, interdependencies, and fitness for purpose is thus a crucial step for gaining insights into the state of a company's data, making it possible to discover the potential data gaps through descriptive data analysis across multiple systems. This process is referred to as data profiling and will be the focus of this paper. Each of the methods and examples that are mentioned in the following sections are based on practical application of data profiling initiatives and have been evaluated to be imperatives for a successful data migration, be it in the context of a business process outsourcing, a cloud data migration, a merger or acquisition, a system rationalization or retirement, or any other similar data management initiative.

2. DATA PROFILING BEYOND STATISTICS

What is the distribution and frequency of occurrence of qualifiers in SWIFT messages over the past year? How many customers have opened trading accounts on a weekend over the past month? Does the value of a securities portfolio on the customer side reconcile with the value on the custodian side? What is the cause of discrepancies in transactions booked on either side? These are some important questions that can be answered with the help of data profiling analyses, beyond the traditional interpretations of 'data profiling'.

Traditionally, data profiling is thought of as a statistical report that identifies natural keys, foreign keys, distinct values, missing data, and distribution and frequency of a column's attributes in a relational database. Business analysts and data analysts perform these statistical analyses to be able to assess:

- Whether the required information is available and complete in the source table
- Whether the source data format adheres to the necessary standards of the target data
- What transformations are necessary to accommodate each occurrence of the source data.

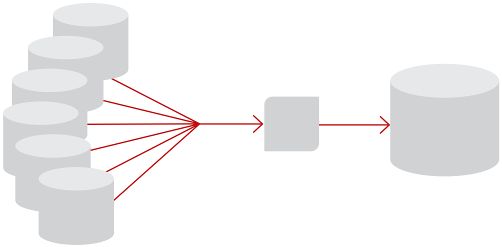
While traditional statistical reports based on a particular column-set are a helpful and necessary start, we recommend data profiling that goes beyond statistical data quality analyses on a column level. The real value of data profiling is derived through the analysis of cardinality, relationships, and interdependencies of data among related datasets and across systems. These attributes are the focus of this article.

3. DATA PROFILING PREREQUISITES

Several factors need to be determined before a data profiling project is started, from the perspective of technology tools, infrastructure, data security, data synchronization, and data sources. Specifically, before a company can embark on a data profiling project, the following activities need to be performed:

- **Agreement with data owners on how source data will be used and which measures will be taken to ensure data security:** it is critical that data profiling exercises are performed on production data (often sensitive data containing customer's personal information) to ensure that the data being analyzed is representative of the true state and quality of the data. Agreements with data owners may include anonymizing sensitive information prior to testing with the data or specific instructions for encrypting data when it needs to be shared with other parties.
- **Setup of a common repository in which synchronized source data extracts will be stored:** it is imperative that disparate datasets are analyzed in conjunction with each other, rather than independently. For this reason, the earliest stages of a data profiling exercise should include coordinating a synchronized date and time of pulling the various datasets and loading the disparate datasets onto a common staging area. Once the datasets are staged, they can be loaded onto a designated data mart for analysis.
- **Setup of a data mart outlining all data in scope for the data migration:** data in scope for data profiling needs to be aggregated onto a designated database, with appropriate keys, links, and referential integrity to allow for a comprehensive analysis of the complete dataset. This may require the availability of additional key mapping tables to allow for a cohesive data model to be built, with linkage to and from all source systems. The data model

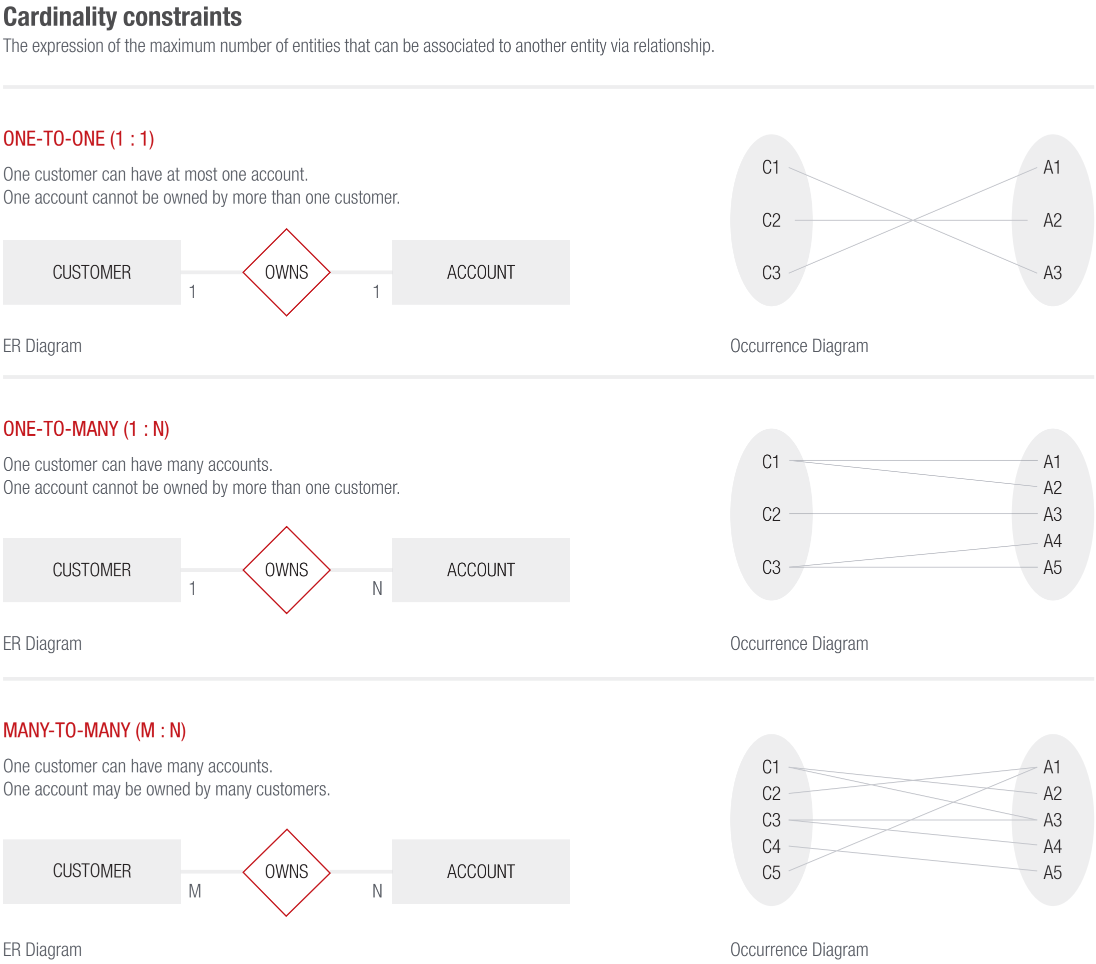
Figure 1: Data profiling architecture



should allow for easy transformations between the source and target state. For this reason, a setup that adheres to the target, as opposed to the legacy system's data model should be sought. To facilitate easier analysis, datasets should be restricted as much as possible to the data that is necessary for the transformations. This may require irrelevant, i.e., historical or out of scope, data to be deleted as part of the setup measures. Alternatively, additional tables need to be created that will flag keys among the various datasets as in scope versus out of scope for the exercise. With appropriate 'write' and data load privileges, the data setup exercise can be performed by a data analyst. Alternatively, this step requires the assistance of a database administrator.

- **Infrastructure and tools selection for data quality assessment:** the process of uncovering data quality exceptions, duplicates, or missing records is fast and reliable with the implementation of proper data quality profiling tools connected to a database, such as Informatica Data Quality, Oracle Data Profiling, or SAS DataFlux. These modern enterprise tools can reduce time to insights and allow for quick and easy results in a data quality assessment. They include the following types of analyses, as summarized by Panoply:
  - **Completeness analysis:** detection of null or blank values
  - **Uniqueness analysis:** validation of uniqueness in primary keys and duplicates detection
  - **Values distribution analysis:** distribution of records across different values for a given attribute
  - **Range analysis:** minimum, maximum, average, and median values found for a given attribute
  - **Pattern analysis:** formats and structures for a given attribute, e.g., phone number or area code

Figure 2: Cardinality constraints


Cardinality constraints
The expression of the maximum number of entities that can be associated to another entity via relationship.
ONE-TO-ONE (1 : 1)
One customer can have at most one account.
One account cannot be owned by more than one customer.
CUSTOMER
OWNS
ACCOUNT
1
1
ER Diagram
C1
C2
C3
A1
A2
A3
Occurrence Diagram
ONE-TO-MANY (1 : N)
One customer can have many accounts.
One account cannot be owned by more than one customer.
CUSTOMER
OWNS
ACCOUNT
1
N
ER Diagram
C1
C2
C3
A1
A2
A3
A4
A5
Occurrence Diagram
MANY-TO-MANY (M : N)
One customer can have many accounts.
One account may be owned by many customers.
CUSTOMER
OWNS
ACCOUNT
M
N
ER Diagram
C1
C2
C3
C4
C5
A1
A2
A3
A4
A5
Occurrence Diagram

Source: Encyclopedia of Database Systems. Springer

A shortcoming of such tools, and traditional interpretations of data profiling, is that the focus remains on an isolated dataset and attribute. We believe that these tools are best leveraged in conjunction with an SQL or SAS data mart to enable discovery of deeper insights related to the interdependencies of the various datasets. For this reason, we have added traditional data profiling methods in the form of a data quality assessment as prerequisites for deeper and more advanced data profiling methods.

Based on the number of resources available to perform these activities, approximately six to eight weeks should be dedicated to the data profiling prerequisite activities mentioned in the steps above. Assuming that the data and infrastructure are established, it is then time to move beyond the basics.

4. DATA PROFILING FOCUS AREAS

In our experience, issues related to data cardinality and miscommunication across different systems are just as pervasive as isolated data quality issues. A tax system, for example, may identify a particular customer as eligible for tax reporting, while the customer data system might perceive this same individual to be free of tax reporting requirements. Another scenario is where a securities static data system may point to a security that is actively traded on an account, which is flagged as inactive in the corresponding account management system. For this reason, data extracts across disparate datasets need to be analyzed in conjunction with each other, and, ideally, target state transformations should be performed in the context of data analysis, in the form of a prototype of the target state transformations. The goal should be to combine and analyze source datasets in a way that will mimic how the source data fits into the new system. The following areas warrant special attention in such an endeavor.

4.1. Cardinality and referential integrity

As mentioned, a key focus of data analysis should be on the relationships among the various datasets. A data profiling assessment cannot underestimate the requirements of validating cardinality constraints. Data design and architecture decisions, after all, are based on data cardinality. If the target system, for instance, expects a legal entity to be unique, with a 1:1 relationship to a legitimate address that is used for tax purposes, then the same validation needs to occur on the source data. It may appear logical that each legal entity is depicted as a unique customer record, however, like many other general assumptions about data, there will likely be exceptions to this rule. If a business customer has set up multiple accounts, with the same tax ID for validation, and has different legitimate addresses keyed with each account, then the customer record for a legal entity will no longer be unique. At this point, design decisions need to be made on how to address duplicate records, so as to avoid technical load issues that result from failure to adhere to the target data model. Figure 2 visually depicts data cardinality in form of an entity relationship diagram.

Examples of issues to look out for when assessing entity relationships include the following:

1:N relationships between source data and expected target data:

- **Source data:** several tax exemption clauses are active for a customer during a given time period (e.g., if the inactive clauses have not been terminated yet).
- **Target data:** the target system will only accept one tax exemption clause per customer.
- **Potential remediation:** identify which tax exemption clauses are truly active, or in the case of multiple active clauses, which are in scope for the target state data needs. Terminate or identify the rest as 'out of scope'. Remediate the data to adhere to the 1:N relationship requirements.

N:1 relationships between source data and expected target data:

- **Source data:** a security links to a CpD (Conto pro Diverse), trust, or custodian account, meaning that a conglomerate of securities link to one and the same account.
- **Target data:** a security needs to be associated with a unique account. An account cannot be set up on multiple occasions with more than one security.
- **Potential remediation:** identify all cases of securities linked to CpDs, custodians, or trusts and parametrize corresponding accounts in a unique setup that adheres to the N:1 relationship requirements of the target system.

Both examples are taken from the context of an ongoing business process outsourcing project in one of Germany's biggest banks, in 2018. Despite strict reporting standards for large financial institutions, the complexity and intricacy of source enterprise data systems, combined with the vast array of errors that can arise in the source data, mean that large financial institutions are not exempt from inconsistencies in data cardinality. While performing data analysis and data profiling activities, the focus needs to be on outlining, understanding, and documenting data cardinality for every constellation of source data. A data analyst that is well-scripted in SQL can perform this analysis easily with grouping, partitioning, or modeling statements. For visualizing and identifying relationships between datasets, we recommend building a matrix of keys, which outlines each combination or occurrence of a key value (Table 1).

Table 1: Data relationship assessment

CUSTOMER	ACCOUNT	#ACCOUNTS	#OPEN TRANSACTIONS
ABCDX	08972	3	6
BZODU	14952	3	6
08972	08972	2	1

Implies an many-to-many (M:N) relationship between customers and accounts, as the same customer (identifier BZODU) owns multiple accounts and at the same time this customer's account (08972) is associated with a further customer, presumably the spouse or another form of a joint owner.

ACCOUNT	#CUSTOMERS IDENTIFIED	#OPEN TRANSACTIONS
14952	1	6
08972	2	7

In this way, outliers or inconsistencies to expectations in cardinality can be addressed with adequate time to react. Updates to a data model require significantly more time to address, as these changes need to be discussed with data architects and business experts. For this reason, it is imperative that this type of analysis is performed with adequate lead time, to allow for necessary reconciliations in the data model.

4.2 Data miscommunication across systems

A further focus of data analysis is to identify the most suitable source of information, or the 'golden source' of the source data. In many cases, multiple systems appear suitable and yet when comparing data among the different source data extracts, conflicting conclusions can be drawn. Examples to look out for include, but are not limited to the following two scenarios:

- Discrepancies in values within custodian account portfolios indicate that the number of source tables taken into account is inadequate. Transactions in transit are stored in an altogether different table, which additionally needs to be referred to in order to reconcile account portfolios.

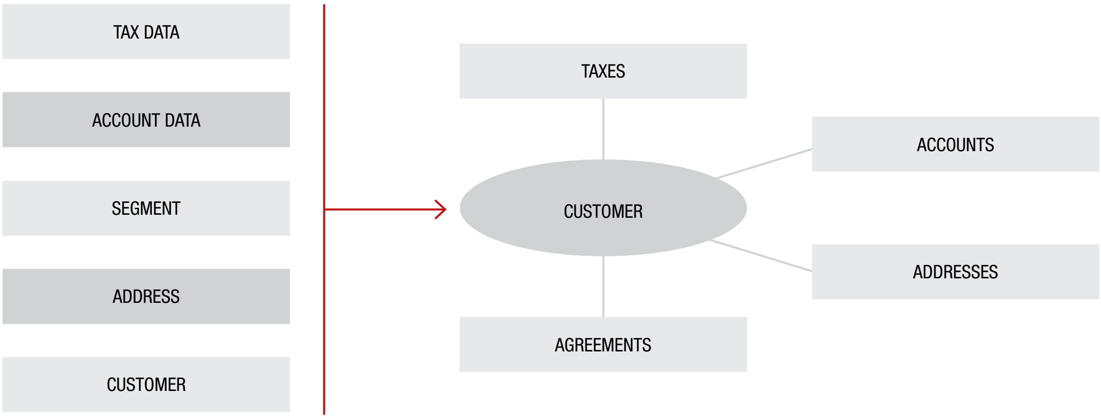
  Overall takeaway: discrepancies in aggregated data are a sign that source data extracts that have been considered are insufficient, unsynchronized, or contain booking errors that need to be reconciled.

- An external account referenced in a SWIFT message is not linked to an external account listed in a company's account management system. The target state system requires an external account to be listed as declared on the SWIFT message, while the preferred source of the data is a structured, relational account management system.

  Overall takeaway: inconsistencies in the same data across different source systems can signify that a customer can be associated with several external accounts, based on the settlement instructions at a given point in time. The alternative and more probable assumption, however, is that one of the entered accounts is invalid.

Miscommunication across different systems is a classic issue that surfaces frequently during data profiling exercises that compare data across multiple datasets. Such an issue is difficult to identify in the context of isolated data quality assessments that are performed on a column level. Investigations of this sort require SQL or SAS to draw connections across multiple tables and various sources. When such a level of analysis is already performed in the context of data quality investigations, there is significant return on investment in going that last extra step and prototyping the source data into the target state, with the required transformations. A data analyst with deep skills in SQL or SAS can build a prototype of the target state transformations in a far shorter time frame than is required

Figure 3: Target state transformations



by a development team to build the enterprise software for a data migration project. In doing so, further gaps that may have not been considered will easily surface and can be addressed prior to the build of the software that will take the project live. This leads to the last focus area of a thorough data profiling exercise:

4.3 Prototyping of target state transformations

An ideal use-case can be developed from the perspective of a business outsourcing project, in which customer static data needs to be migrated to a provider in an all-encompassing customer-based view. Customer data may be spread across more than ten enterprise source systems, but this disparate data needs to be clustered together and sent to the provider as a package for each customer. Performed correctly, the segregated data depicted in the source state needs to be transformed into an all-comprising cluster, as presented in Figure 3.

At first glance, this activity may seem simple, as if only a series of connections across the varying tables is required. In practice, this is not the case. In the process of prototyping these target state transformations, all the constellations that stray from a typical customer will become apparent. Examples could be a legal entity corporation with subsidiaries spread across several countries, who may share legal addresses or tax information with other subsidiaries, or a married couple with several joint accounts as well as individual accounts, or an LLC that belongs to a customer, with the account comprising both personal data and tax agreements associated with their corporation. How do you transform each of these specific cases, and multiple others that will arise during data analysis investigations? By prototyping each constellation, the gaps between the current state of the data in its legacy form, and the future state data requirements will become apparent. The devil, after all, is in the detail. The value of this exercise is the recognition of such data clusters that are different from the norm. A further example that could highlight the benefits of performing target state transformations in the context of data profiling is a scenario of two different corporations, with different legal addresses, that share an account and securities. This is not a normal setup of a client and it points to a trust relationship.

The target state of the data would require these corporations to be depicted as a joint organization, with their ties outlined in further agreements. The present state of the data, however, does not suggest any partnership between the two corporations. How do you then transform such a constellation? Such obscure cases, which most business analysts and data experts may not consider initially, are precisely the constellations that need to be discussed and prototyped early in trials and tests. This would enable a sound development 'backbone' that accommodates the full spectrum of constellations that need to be transformed. It would also prevent subsequent load failures of clusters that do not adhere to the norm.

Overall takeaway: inconsistencies in the same data across different source systems can signify that a customer can be associated with several external accounts, based on the settlement instructions at a given point in time. The alternative, and more probable assumption, however, is that one of the entered accounts is invalid.

This exercise of prototyping the target state based on current state production data will likely take an analyst between a few weeks to several months, depending on the complexity of the data and transformations. This time investment is strongly recommended to avoid pitfalls in the later development of the software that will bring the project live. It is easier for developers and architects to set up the basis of the technology with a knowledge of the full spectrum of constellations that need to be accounted for, rather than realize that their development framework is insufficient and does not allow for successful transformation of exceptional data constellations during development.

Indeed, the greatest value derived from data profiling activities is in this third step, which examines and raises questions about less conventional constellations that stray from the expected target state data clusters. Having concluded these steps, a representative view of a company's data will be achieved, with the inevitable identification of the most critical data quality issues. It is now time to engage with the business stakeholders and assess the impact and potential mitigation measures for the data quality issues found.

5. MITIGATION MEASURES UPON IDENTIFYING DATA QUALITY ISSUES

Upon identifying data quality issues, the magnitude and size of the problem needs to be assessed by data owners, upper management, system owners, architects, and business owners alike. A large-scale data cleansing effort is difficult, if not impossible, to initiate without a sound basis of facts, grounded by business context and an impact assessment. For this reason, we strongly suggest that the data analyst validates each identified data quality issue with technology and business owners and discusses root causes and potential workarounds with the appropriate parties, prior to the onset of the data cleansing efforts.

In the context of unsynchronized data across systems, perceived data quality issues may simply turn out to be the result of late data loads. In the context of missing customer information, for example missing countries of citizenship, such issues may already be addressed in later program updates with default values. A tax division, for example, may re-key an empty record to the country in which the customer is based prior to submitting tax information to government authorities.

It is also important to discuss data with the target system owners. There may be a workaround for data quality issues that the target system can suggest. It may, for example, turn out that certain requested data is not required for any future process and can be altogether excluded from the data migration. Alternatively, the use of technically feasible default keys may be recommended. Data analysis without appropriate conversations with data owners and business consumers of this data will likely lead to some false findings. Enterprise data systems are highly complex and multi-faceted and even an expert data analyst cannot grasp the full context with a mere data-driven approach to analysis.

Upon validating the pool of identified data quality issues that need to be remediated, the next step for a data analyst is to consolidate each record containing data quality issues, with appropriate links and keys that a business unit can work with, into a single report. Large-scale data cleansing efforts will likely be initiated by a company's senior management, such as a Chief Data Officer (CDO) or Chief Technology Officer (CTO), and statistics, such as the range and frequency of issues per source system (or customer, business unit, etc.), are likely to be requested.

The design of these reports must be well thought out and contain key information that each business unit needs for further investigations. If certain units require the customer's name, while others require the customer account number, both pieces of information need to be available.

Management will need to understand the overall impact prior to arranging for large-scale data remediation efforts to take place. Assessing the overall impact may require calculating the average number of data quality issues per client in a particular business division, or the number of customers that cannot be migrated due to data quality issues associated with their accounts. As a result, the report needs to be designed in such a way that will make these statistics possible to retrieve. For this reason, it is important to discuss with all relevant stakeholders their requirements for receiving erroneous data records in a way that enables them to initiate data cleansing initiatives.

At this point, regular meetings will need to be set up with the appropriate units that will be responsible for data cleansing. The ideal scenario is a data cleansing effort that does not require manual intervention of personnel to investigate each customer record. The goal is an automated means of data cleansing. This, however, will not always be possible and for this reason it is important to plan ahead and account for time and resources required for manual investigations. The data analyst needs to be available to assist business owners in this step, through further data assessments and data-driven analysis. System experts familiar with communication across different source systems also need to be brought on board to identify the source systems where the data remediation initiative should be performed, for best results. In our experience, the stage of analysis of the root cause of data quality issues, in an effort to determine the best approach for data cleansing measures, will require approximately a month, depending on the number of data quality issues that needs to be addressed.

Upon identifying root causes and methods of remediation, the data cleansing initiatives in each source system can begin. The timeline for data cleansing efforts may differ per source system and per data quality issue, and the number of records cleansed may not be completed the first time round. For this reason, the data analyst will need to engage in ongoing data checks to assess data cleansing success. Data remediation, after all, is an ongoing effort, as new data is collected every day. In order to prevent new 'bad' data from entering IT systems, it is advisable at this stage to discuss and assess the feasibility of setting up front-end controls with plausibility checks that prohibit inaccurate data entry records. Only if the data profiling effort has been completed early enough will there be sufficient time for the completion of a multi-month data remediation effort prior to the data migration.

6. CONCLUSION

The advantages of a comprehensive data profiling effort in the early stages of a transformation project are vast and far reaching, facilitating on-time and in-budget project delivery, while making possible the following benefits:

- Risk reduction through early feedback and early testing of planned data transformations
- Identification of data dependencies and inconsistencies
- Early recognition of potential gaps and problems presented by data, to allow for sufficient time for remediation, with reduced effort
- Early quality checks of both source data and target data against the initial conceptional design and documentation
- Consistent and comprehensive understanding of source production data
- Adequate time for any required data scrubbing, cleansing, and remediation activities
- Discovery of unanticipated business rules and data constellations

The overall benefit of this initiative is in the early risk detection and consequent prevention of wasted costs and project extensions that arise from risk detection in the migration test stages.

While highlighting the benefits, it is important to remember that a data profiling effort needs to be performed at the early stages of a project, ideally before the start of larger-scale developments. A data profiling exercise that occurs concurrently with development or in isolation from development is less effective and will not be as advantageous in terms of the risk reduction goals. The earlier the onset of data profiling activities, the higher the benefits and gains from this effort.

This document was produced for information purposes only and is for the exclusive use of the recipient.

This publication has been prepared for general guidance purposes, and is indicative and subject to change. It does not constitute professional advice. You should not act upon the information contained in this publication without obtaining specific professional advice. No representation or warranty (whether express or implied) is given as to the accuracy or completeness of the information contained in this publication and The Capital Markets Company BVBA and its affiliated companies globally (collectively "Capco") does not, to the extent permissible by law, assume any liability or duty of care for any consequences of the acts or omissions of those relying on information contained in this publication, or for any decision taken based upon it.

ABOUT CAPCO

Capco is a global technology and management consultancy dedicated to the financial services industry. Our professionals combine innovative thinking with unrivalled industry knowledge to offer our clients consulting expertise, complex technology and package integration, transformation delivery, and managed services, to move their organizations forward.

Through our collaborative and efficient approach, we help our clients successfully innovate, increase revenue, manage risk and regulatory change, reduce costs, and enhance controls. We specialize primarily in banking, capital markets, wealth and asset management and insurance. We also have an energy consulting practice in the US. We serve our clients from offices in leading financial centers across the Americas, Europe, and Asia Pacific.

WORLDWIDE OFFICES

APAC
Bangalore
Bangkok
Hong Kong
Kuala Lumpur
Pune
Singapore

EUROPE
Bratislava
Brussels
Dusseldorf
Edinburgh
Frankfurt
Geneva
London
Paris
Vienna
Warsaw
Zurich

NORTH AMERICA
Charlotte
Chicago
Dallas
Houston
New York
Orlando
Toronto
Tysons Corner
Washington, DC

SOUTH AMERICA
São Paulo

FSC®
www.fsc.org
MIX
Paper from
responsible sources
FSC® C001648

WWW.CAPCO.COM

CAPCO