WHITE PAPER

DATA REMEDIATION IN FINANCIAL SERVICES

Triggers, execution, controls and continuous improvement

'Raising the value of the data within an organisation'

AUTHOR: MILAN CIRIC, CONSULTANT

MAY 2019

WE HAVE A SLEW OF DORMANT CUSTOMER DATA ISSUES AND WE KEEP LOADING MORE DATA ON TOP OF THEM
WHAT ?!
ACROSS THOUSANDS OF CUSTOMERS OVER SEVERAL YEARS
DATA
I AM COMFORTABLE THAT OUR CUSTOMER DATA MEETS INDUSTRY STANDARDS
DATA

INTRODUCTION

Data remediation activities in financial services will never cease. The best that can be achieved is significantly reducing the frequency and scope of remediations over time.

This paper is a brief discussion of the triggers, execution and controls associated with data remediation events and can be applied to superannuation, wealth management, banking and insurance.

Remediation can be described as an ongoing process which brings correctness and order to the underlying data within an organisation. It does not necessarily indicate a negative financial impact to a customer, but it *does* indicate a negative financial impact to the organisation.

Once remediation is underway, it usually garners significant internal attention, given that it often exposes a dormant issue. A lot of pressure is placed into assessing scope, product impact, legal and regulatory breaches in a very short timeframe.

This is one of the elements which make remediation unique. It crosses over many areas of the business and the requirements are often not all understood at the onset.

The paper discusses:

- Triggers: the event that instigated the remediation
- Execution: the process of fixing the current state
- Controls: the process to stop the issue from recurring

DATA REMEDIATION TRIGGERS

Stopping the bleeding *versus* finding the root cause

The types of data remediation projects in financial services vary significantly, however they all share the same origin, a latent defect. This defect could have been introduced by anything from an administrative mistake to a system issue. The exposure of the defects triggering the need for the data remediation will fall into one of two categories:

- *Reactive trigger*: ad-hoc or accidental identification of a wider issue, for instance, an issue identified by a customer or regulatory body.
- *Focused trigger*: resulting from a cyclic and ongoing data quality assessment, instigated by a controlled data quality processes, part of a wider data management system.

Regardless of the trigger, it must go through the same stage gates and analytic processes. This however does mean the trigger can affect the quality of the implementation, given remediations instigated by a controlled process will be far better prepared.

Reactive remediations tend to be run with tight deadlines and budgets. Teams are usually stretched with their efforts and the possibility of errors introduced during the remediation are increased, meaning the quality of the final implementation can suffer. Here, there is a stronger focus on quarantine or '*stopping the bleeding*', rather than isolating the root cause.

"*Reactive* remediations tend to be run with tight deadlines and budgets. Teams are usually stretched with their efforts, and possibility of errors introduced during the remediation are increased,"

Focused remediation, in comparison, can be well structured, scoped and budgeted. The fact you have existing data quality rules, indicates thought has already gone into the business impact, if this rule is broken. But more importantly, this approach also drives towards a root cause analysis. For instance, if an attempt is made to simply 'stop the bleeding', the data quality rule will raise the same issue when next executed and will keep doing so until the underlying problem is fixed.

Focused remediations require a mature data management system, and the reality is that most financial services organisations are still on that journey, and most remediations are still very much reactive.

The next section covers some key elements in executing remediations.

"*Focused* remediation, in comparison, can be well structured, scoped and budgeted. The fact you have existing data quality rules, indicates thought has already gone into the business impact, if this rule is broken."

DATA REMEDIATION EXECUTION

What doesn’t kill you will make you stronger

The mantra, ***‘what doesn’t kill you will make you stronger’*** is very applicable when it comes to the execution / implementation of remediation projects in financial services. It is often a demanding process, with short deadlines, typically requiring multiple attempts and internal reviews to get right.

Additionally, it is not uncommon to uncover new insights during the execution that alter the course of the remediation. As such, it’s critical to have a structured process ready when entering this phase, covering everything from data ingestion to analysis and reporting.

The first step is acquiring the correct data. To do this, a deep dive with businesses subject matter experts (SME’s) must be undertaken with the objective of understanding what key data sets are required. During the data collection phase there are at least three things to consider:

SCOPE

Data extraction can be a costly process requiring time and resources. Limiting the data set to strictly cover your scope will save both the time required for analysis and any subsequent re-extractions. Ensure this scope is clearly communicated with the technology team providing the extracts. For instance, many business units interchangeably use member number, client number and account number, yet in the back-end database, these are all very different data elements.

SOURCES

Identify all the sources of the required data sets. Typically, this will be various groups within the technology department that may provide source system extracts or set up direct access to database replicas. It could also be project managers that have relevant data stored as flat files, or SME’s that may hold data around product or business rules. Having an overview of this will aid in scoping out the work involved in centralising all the data in one structured database.

AUDITABILITY

Remediation activities will almost always be subject to either an internal or external audit. As such, it’s important to consider the steps which will need evidencing and or attestations, that prove correct/production data was used in analysis. This would apply to any data received in the form of extracts, or data referenced through specific product rules.

These attestations need to be clear about the data being provided, and any known limitations or filters applied to the data sets. It’s always easier to request these upfront by the stakeholder providing the data. For instance, in case of technology extracts, it allows them to save the logic they used for extraction, at the time they performed that task.

DATA LOAD GOALS

Once the data has been sourced and attested to, it should be pulled together within one structured data set through the data load and transform process. The main goals here are to ensure:

NO DATA LOSS

For example, use a loading process that reports on the number of lines pulled into the database from the data extract and compare to lines of data in extract.

DATA INTEGRITY

For example, ensure values retrieved are sensible. Common areas are checking data types, date ranges, duplicated records, missing or null values in key fields, truncation of leading zeroes and so forth.

DATA COMPLETENESS

For example, does the data set meet the scope requirement. Does a simple trend analysis show large unexpected gaps in the data? Investigate relationships between data sets which are meant to be related and see if they converge to the same story.

Once the data is loaded, the above tests must be carried out and documented, as they are likely to form part of a review/audit. It is imperative extensive testing is done on the data once loaded, as this will form the foundation of the data analysis to be performed for the remediation.

Keep in mind that data integrity issues, outside of the one causing the remediation, are likely to exist in the extracts provided, and can add a significant effort to the validations being performed.

"It is imperative extensive testing is done on the data once loaded, as this will form the foundation of the data analysis to be performed for the remediation. Keep in mind that data integrity issues, outside of the one causing the remediation, are likely to exist in the extracts provided, and can add a significant effort to the validations being performed."

CORE DATA ANALYSIS

Following on from these validations, is the core data analysis and where much of the complexity lies. Some of the key steps listed also apply to broader analytic projects:

KEEP THE BUSINESS INFORMED

Keep the business informed about any findings that are unexpected. It could point to another core issue or a mis-interpretation of the requirement. Try to take the business SME on the data analytic journey, as their input and feedback will help cut a lot of noise inherently present in all data sets and keep them on the same page.

KEEP A RECORD OF ALL CHANGES

Keep a record of all changes made to the analytic script or logic being developed. Particularly if your logic is hard coding anything or implementing specific product requirements. The logic implemented here will be under heavy scrutiny and review, so all decisions made here need to be clearly documented. It is good idea to do this as you develop, when all design decision is fresh and supporting evidence is likely on hand.

DEVELOP THE DATA VISUALISATION TO HELP WITH THE COMMUNICATION

The analysis being done will need to be communicated back to other stakeholders, so it is good practice to try and envisage what form this data visualisation will take early on. It could be a MECE (Mutually Exclusive Collectively Exhaustive) tabled output, logical or hierarchical data groupings etc. Developing this view early in the analysis piece also allows the data analyst to have a good overall understanding of the data.

When using the data sets at hand, it is recommended to establish a rough baseline of the size and scope for the remediation at the very beginning of the project. In the early stages, this will likely constitute several broad level assumptions (which must be explicitly stated), with the goal of developing a heuristic for the member numbers and dollar impacts at play.

The process of diving into the data to develop the baseline might expose data sets originally not thought to be required. As the remediation progresses, this can also serve as a sanity check that numbers are still in the right ballpark. If there are major differences, then it is worth taking a closer look at the area with the discrepancy and being able to justify it either quantitatively or qualitatively.

> “The process of diving into the data to develop the baseline size and scope might expose data sets originally not thought to be required. As the remediation progresses, this can also serve as a sanity check that numbers are still in the right ballpark.”

DATA REMEDIATION CONTROLS

How to prevent recurring issues

Typically, the final steps within a remediation project are implementing controls, to ensure the same issue does not manifest itself again. These controls can take many forms, and some are very much influenced by the limitations of the system on which the controls reside.

However, this article only explores controls through the data lens. Data controls are well suited to sitting across a relational database, as data quality rules. This allows precise code to be developed to flag the scenarios that gave rise to the remediation. This output would still require monitoring and an escalation process to stop the issue flowing through once detected.

Attributes can be assigned to the rules to aid in the post-detection process, such as:

1. Frequency: how often the check needs to be carried out
2. Severity: defined by a risk assessment of this scenario
3. Accountability: entity that oversees the monitoring and reporting of this rule

This is really what would constitute part of the organisations data management system, and ties right back into the development of *'focused triggers'*, arising from an established data quality process, discussed at the beginning of the article.

It is worth re-iterating that once the data rules are in place, and executing on a recurring basis, it must not operate as a set and forget function, if they are to be of real value. Maintenance and reporting are two key elements required to keep the rules meaningful.

For instance, regarding maintenance, there may be product or regulatory change which affect the logic of existing rules, or how often then need to be run and reported on.

Regular reporting is arguably the most critical step, one that can be the single point of failure in this phase. Outputs of the ongoing data quality rules must be made very visible. The entity accountable for the output of these rules must have clear reporting lines to a data steward or another role within the data governance framework.

If these functions are still under development within the organisation, it is possible to report the high-level outputs of these rules to operation risk. The end goal is to have the insights gained from the data quality rules, actioned and not ignored, otherwise the vast amount of work done up to that point will definitively go to waste.

"Once the data rules are in place and executing on a recurring basis, it must not operate as a set and forget function, if they are to be of real value. Maintenance and reporting are two key elements required to keep the rules meaningful."

DATA REMEDIATION AND CONTINUOUS IMPROVEMENT

Raising the value of data within an organisation

If new data is entering a system, remediation activities will not be going away, nor should they. Data is inherently 'dirty' and must undergo a systematic quality process.

Remediation activities themselves are not necessarily an indication of poor controls, but what gave rise to these remediations is. If most remediations are instigated by ad-hoc triggers and external parties, then there is a clear lack of a reliable data quality process.

A mature data management system can minimise the number of ad-hoc triggers and remediation activities overall. While it does not completely eradicate them, it will certainly allow project teams to be better prepared for the resulting remediation, minimise risk of introducing new issues, and will show the controls put in place are effective.

When done correctly, data remediation contributes to the important cycle of continuous improvement, raising the value of the data within any financial services organisation.

"If new data is entering a system, remediation activities will not be going away, nor should they. Data is inherently 'dirty' and must undergo a systematic quality process."

QMV

Where to now?

If your organisation needs assistance with data remediation, QMV can help.

QMV has performed hundreds of data remediation projects. We utilise an innovative and flexible approach to assist clients identify and create visibility of data quality issues. Our holistic methodology promotes accuracy of data throughout the data life cycle, leading to trust and confidence in operations and decision making.

QMV's extensive work in data quality management has led us to develop *Investigate* our data quality software solution. QMV identified the need for rigorous and systematic data quality management in financial services because poor data quality costs financial institutions millions each year.

Please reach out to QMV for further information.

Contact QMV today

p +61 3 9620 0707

e sayhi@qmvsolutions.com

w qmvsolutions.com