International Journal of Finance and Investment
ISSN: 3078-6002 (Print), ISSN: 3078-2295 (Online) | Vol. 6, No. 1, 2026

Analysis of Key Points in Data Cleaning and Migration Implementation for Financial Institution System Upgrades

Wenxin Li

School of Business, University of New South Wales, Sydney, Australia

Abstract: The quality of historical data migration during the upgrade of financial institutions will directly affect the stable operation of the new system after it is released. The new-generation core system of Guizhou Rural Credit Cooperative and the centralised trading system of Guolian Minsheng Securities are chosen as the subjects of this study. The former had more than 90 million current accounts and over 25 million fixed-term accounts. After 520 days of hard work and 22 hours, 64 system switches have been set up to complete the core data migration without any errors. The latter had connected 23 different systems, combined the data of about 1.92 million brokerage customers, and merged information from 74 branch offices. On the first day after the switch, 640,000 transactions had reached 9.476 billion yuan. The five contents of the implementation plan for the migration project are as follows: data cleaning, migration plan, consistency verification, risk response and organisation cooperation. Data quality assessment at the source end serves as the foundation for the cleaning operation. The Migration Strategy should balance the window for business interruption with the risk of the project. Whole-field comparison serves as the first-line verification method. The entire-process risk prevention and multi-departmental cooperation offer the final guarantee. Based on the public project data, discussions will be held to provide a reference for the design of a similar financial system migration plan.

Keywords: Financial institutions, System upgrade, Data migration, Data cleaning, Consistency verification.

1. Introduction

Under the promotion of the information-technology innovation strategy and business-transformation, core systems in the financial sector have been upgrading at an accelerated pace. In March 2025, the People's Bank of China's technology work conference put forward deepening data governance and advancing self-reliant and controllable financial infrastructure. In the same year, the "Data Security Management Measures for Banking and Insurance Institutions" were issued, and a data full-life-cycle protection mechanism was required to be established. In light of policies and business development, some institutions have begun system renovations in the areas of accounting, settlement and credit.

Linkage of the old and new systems depends on data cleaning and migration. Customer information, account balances and transaction records under the old system need to be cleansed, mapped and validated before they can be used by the new system; this is not insignificant work in itself. Guizhou Rural Credit Cooperative's migration of the new-generation core system has over 90 million current accounts and more than 25 million fixed-term accounts, and the switch of Guolian Minsheng Securities Trading System includes 1.92 million brokerage clients. The scale of the data migration may result in accounting errors, transaction failures and other business problems.

The three problems of migration are: The old system is mainly based on a traditional relational database for the data model, and the new system is distributed. Field mapping is item-by-item review. There are problems with the quality of the data in a long-running system, such as null values in fields, inconsistent formats and broken foreign keys. The abnormal data that the old system can tolerate will cause a write failure in the new system directly. The institution will not allow for a prolonged service interruption, and the migration needs to be finished within a short time. Multiple constraints raise the migration to the level of engineering and require a system of rules.

The five contents of this paper are the directions for data migration: cleaning strategies, migration plans, consistency verification, risk prevention and control, and organisational support [1].

2. Engineering Implementation of Data Profiling and Cleaning Rules for the Source System

The first step in cleaning data is to understand the basic information of the source system well, and then start the actual work.

Systems that have been in use for a long time generally have data problems everywhere: mandatory fields are empty, enumeration fields have undefined codes, the amount column contains negative numbers or excessively large values, and foreign key relationships among related tables are frequently broken. These problems are usually detected by the exception-handling mechanism at the application level, and the business continues normally. However, in the new system, the verification rules are relatively strict; all the exceptions fail, and thus the migration cannot proceed. A full data quality report should be prepared before cleaning to list the types and quantities of anomalies in each table and field, and based on this, set a priority order and allocate resources [2].

Cleaning rules cannot be set independently by the technicians. Technicians can determine whether the data format is correct and whether the types match by themselves; however, the meaning and value range of the field need to be specified by business personnel. For example, the customer type field is shown as 1, 2 and 3 for individuals, enterprises

and peers in one system, but in another system it is shown as A, B and C. It may be shown in the new system by another set of codes. The technician cannot determine it from the business end, and only after being approved by the business department and signed can it be formally mapped.

In the course of building the new cross-border payment settlement system for Industrial Bank, a project team has been analyzing a heterogeneous database with over 600 tables and almost 10,000 fields. Manual verification would have taken several weeks, so the team created an automatic script to generate a data model comparison report and reduced the execution time from 15 hours to 5 hours.

The amount of data is also becoming a problem. Processing hundreds of millions of records one by one with stored procedures is slow and can affect production. Some projects have moved the cleaning logic to the database engine layer and employed parallel computing for sharding processing of the data [3]. During the migration of the core system for Guizhou Rural Credit Cooperative, which has more than 90 million current accounts and 25 million fixed-term accounts, a cleaning step was taken at the same time by dividing the database into separate tables to ensure quality and meet the planned schedule. Another way is to do it in stages: prioritise cleaning for the main account table and transaction detail table, and for non-core historical log-type data, a lower standard can be applied or it may be left in its original format before migration and managed later.

Wash the samples, and from among them, take some to verify whether the rule has been applied correctly. The modified record needs to save the complete log for audit purposes. The preparation of the data is complete, and now a migration plan will be chosen.

3. Selection Logic of Migration Plans and Management of Switching Windows

The selection of a migration plan needs to balance the length of the business interruption with migration risk [4].

The most direct way is to turn off the old system, move all the data to the new system for verification, and then go live. This way has fewer operating steps and simpler rollback, but the interruption time will be relatively long. The new core system of Fusion Bank was released in this way, and the switch took 15 hours to complete after a 6-hour migration window. For institutions with concentrated business hours, this solution is still feasible if all the data export, conversion, loading and basic verification can be carried out within the time window.

The Case of Institutions Providing 24-hour Services is Different. Every extra minute of downtime is more time that customers have to wait and another minute of accumulated transaction instructions. Guizhou Rural Credit Cooperative has finished the 22-hour switch of 64 systems and is now 10 hours ahead of schedule. However, the switch will simultaneously migrate many systems, including credit and debit, as well as core systems from abroad. At this time, the parallelism of the migration tool and the accuracy of the operating steps were both pushed to their limits. To reduce the downtime window, the project team adopted a "full data + incremental data" synchronization method: they began transferring the full historical data several weeks in advance, allowed the old system to continue running and continuously received and loaded data in the new system, and on the day of the official migration, only the newly generated data during this period needed to be synchronized; thus, the volume of data transfer for the final step was significantly reduced [5].

The above discussion is about the migration of the same kind of database. If the databases at the source and destination are different types, then the solution design will be even more complicated. Between the traditional commercial databases and the domestic distributed databases, a clear column-by-column data-type mapping needs to be performed, and objects such as stored procedures, triggers, and user-defined functions also need to be rewritten or adapted. Only the different SQL dialects need to trigger extensive code modifications. Some institutions select to deploy middleware or data synchronization tools in a transitional area, encapsulate the heterogeneous conversion logic within the synchronization link, and minimise changes to the application code as much as possible.

After the merger, 23 heterogeneous systems in the central trading system of Guolian Minsheng Securities needed to be connected, the data of 74 branches and 69 exchange seats had to be centralised, and 1.92 million brokerage clients needed to be served normally on the first trading day after the switch. There are constraints here, and the sole condition for the migration plan is that customers should not be aware of it. After several rounds of drills and verification, the parallel switching plan was finally finalised, and the whole switch was completed in the non-trading period.

4. Consistency Verification: from Sampling Verification to Full-Field Comparison

First of all, move the data in this way. Are the data in the new system the same as in the old system? One needs to find out what it is.

Early projects generally used sampling verification by extracting some records from the tables and manually comparing them, with limited coverage. Many unexamined abnormal records did not show up until after the system went live. Recently, the trend has been to develop full-scale, full-field automatic comparison that compares record by record and field by field to set the difference target to zero.

The feasibility of the full-scale comparison is determined by the Design of the comparison tool [6]. The migrated data is spread across many databases and servers. The comparison tool must connect to both the source and the target at the same time, retrieve their records one by one using the primary key or unique index, and then compare the fields. Without primary keys for historical log tables, a compound conditional matching logic needs to be used. Complete the comparison and then output a difference report that lists all inconsistencies. When the new generation core system of Guizhou Rural Credit Cooperative was released, the main data achieved "one-time zero-deviation migration". Based on the results of multiple rounds of full-scale comparison, after each round of testing and migration, a complete comparison process was carried out; differences were traced back to the source data or the conversion logic was modified, and iterations were repeated until full consistency was achieved.

Do not carry out a single verification at the end of the full migration; instead, integrate verification at all times during the migration process. The first full-scale migration stage can begin comparing after each batch of data has been loaded. The earlier the problems are identified, the better, to avoid

accumulation until the end [7]. The data in the incremental synchronisation stage is in an unstable state. The verification strategy should be changed to incremental comparison based on timestamps or log sequence numbers to ensure that each modification at the end of the source and target corresponds. Prior to the cutover, some representative business scenarios were chosen for end-to-end business scenario validation, and transactions were initiated from the client side to confirm that the data modifications in the new system's backend met expectations. Such verification often finds problems that field comparison does not cover, such as values being the same in the database but the results differ after application-layer calculation.

In the migration of Postal Savings Bank's new core system for corporate business, a strategy of "unified planning, step-by-step implementation and online switching" was used to achieve an almost undetectable migration of customer data. After the full-scale release of the system, all the company's customers were successfully switched online without any service interruptions and normal business operations were resumed.

5. Risk Identification and Emergency Mechanism during Migration Process

Stable operation does not mean that there are no risks. Risk management should not be improvised on the day of the launch. In advance, organise drills and contingency plans to reduce the impact.

The three kinds of risk are: One is data loss and mismatch; the second is application incompatibility; and the third is operational errors [8]. The application is not compatible either. If the database version or type changes, the original syntax of the stored procedure may no longer be supported, and its performance may also change. Often, the problems are not apparent during the migration, and they will only be discovered after the business starts operating normally.

A few full-process drills are generally conducted before the actual migration at the same time. The V3 upgrade project for the overseas core system of Industrial Bank (domestic IT innovation edition) has passed SIT and UAT tests, and five rounds of production switch drills have been carried out. Each round followed the official cut-over operation manual, and all the steps were included: data export, transmission, loading, verification, switching, rollback, etc. Drills can help to find deficiencies in the operating manual and also train the team in coordination. Every time a problem was found during the drill, it had to be recorded, the reason for the problem clarified, and then a fix implemented before re-verification, thus forming a closed-loop system.

The rollback mechanism is the final line of defence [9]. Even after all the preparations, there may still be problems in the cut-over. Rollback Conditions and Operating Procedures will be set in advance. When the data in the main table does not match, key transactions fail, or the system response is too slow, one of these triggers must trigger a rollback. Rollback will be carried out according to the above steps to restore the old system as quickly as possible. Migration logs and snapshots should also be saved for later reference.

On the night of the new core system launch for Jiangyin Bank, more than 300 people were mobilized according to the plan to carry out data migration, system switching and business verification in stages. Guolian Minsheng Securities had only two off-peak hours: May Day and National Day. They simply called it a "top leader" project and had the chairman and the president lead the response and handling of any problems directly for a quick resolution.

Migration risk sources are varied in number. At the data level, there may be field loss, format disorder and amount precision deviation; at the application level, there may be interface incompatibility and return value anomalies; at the operational level, there may be process interruption due to omitted instructions or reversed sequence. The three types of risk are interconnected, and if any one fails, it will cause all the others to fail. Therefore, the drilling should cover all the stages and uncover any problems that may have occurred during this time.

6. Coordination of the Organization and Control of the Implementation Rhythm

The construction of the facility depends on the supply of funds and other resources by the company at this time. Finally, organisational support and process management have established clear responsibilities and authorities, a stable rhythm, and smooth coordination [10].

Coordination among the departments of technology, business, operations and risk control is needed for data migration. If any link fails to cooperate, the whole process will be delayed. From the successful projects we have seen, those with good organisational support often have several similarities. The person responsible for the decision will arrange the funds for the project team. The business head of the project is responsible for all links in requirements analysis, rule confirmation and acceptance testing. They do not only appear at the core nodes. The test team is still independent, and the quality requirements are not related to the development progress.

Regulation of Rhythm is also Required. The preparation work for the system upgrade will be about three hundred days. A large number of tasks will be run concurrently. If one link is held back, the final launch time will be delayed. Over the course of 500 days, Guizhou Rural Credit Cooperative has divided the entire project into several stages, set delivery items and review standards for each stage, and conducted regular progress inspections. Finally, they achieved a 22-hour switch of 64 systems; otherwise, it would not have been possible to manage them. A report indicated that Huawei provided a series of implementation processes for the thousands of tasks in the project, and each link was standardised and visualised.

Training and knowledge dissemination should not be postponed until the product is about to go live [11]. The operation interface, business process and handling method of the new system are not the same as those of the old system. Frontline staff require many more repetitions in the simulation environment for training. Both the old system and the new system will run in parallel, and the technical team and the business department will work closely together to promptly address any operating problems or emergencies and collect feedback.

The new-generation corporate business core system of Postal Savings Bank was rolled out in a "unified planning, step-by-step implementation, online switch" mode. The pre-training and verification were both excellent; therefore, when it went live, there were no major problems of modification.

After the system went live, the project team still needed to be present at the construction site for some time to observe how the system performed, deal with data migration problems one by one, and confirm that the new system was stable before finishing the data migration. With the above organisational support, the implementation of the previous four steps has a solid foundation. We will now return to the general process of the five steps.

7. Conclusion

The five parts mentioned above are in harmony and work together.

Data cleaning addresses "what to move"; migration plans outline "how to move"; consistency verification determines "whether the move is correct"; risk management plans for "what to do in case of damage"; and organisational support determines "who will move and how to ensure its completion". The above five steps are related. If the first step is not completed correctly, the others cannot be done. If it is weak in any place, the first-investment will be wasted.

All the analysis points in this paper are based on publicly available project cases. However, each institution has different business features, system structures and available resources and time, so the solutions cannot be directly applied. For example, the field mapping rules of the banks and rural commercial banks are not the same; the window period restrictions for securities companies are also different from those at banks. It will be more suitable for one's own circumstances.

Recently, the data migration work by banks has changed from a "one-off project" to "continuous capability". The development of information-technology innovation is still advancing, the application range of domestic databases is expanding, and distributed architecture is gradually being realised. More institutions will have to go through this process eventually. The tools scripts, operation manuals, comparison templates and exercise plans from each migration will be more beneficial if they can be consolidated and used as reusable internal resources rather than just being produced for a single project.

There are some deficiencies. Based on publicly available information, many details of the project implementation are unavailable. No research was carried out at the institutions. Therefore, the analysis depth has not reached that of operating procedures. It only offers a general reference. If the following studies will be conducted to gather more detailed project logs or if horizontal comparisons of migration plans for institutions of different sizes will be made, more targeted conclusions can be drawn. The institution about to start the upgrade will be notified in advance of any potential risks to prevent a delay caused by unforeseen problems later on.

References

[1] Karkošková, S. (2023). Data governance model to enhance data quality in financial institutions. Information Systems Management, 40(1), 90–110.

[2] Ehrlinger, L., & Wöß, W. (2022). A survey of data quality measurement and monitoring tools. Frontiers in Big Data, 5, 850611.

[3] Dean, J., & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. Communications of the ACM, 51(1), 107–113.

[4] Gholami, M. F., Daneshgar, F., Beydoun, G., & Rabhi, F. (2017). Challenges in migrating legacy software systems to the cloud—an empirical study. Information Systems, 67, 100–113.

[5] Iqbal, A., & Colomo-Palacios, R. (2019). Key opportunities and challenges of data migration in cloud: results from a multivocal literature review. Procedia Computer Science, 164, 48–55.

[6] Matthes, F., Schulz, C., & Haller, K. (2011, September). Testing & quality assurance in data migration projects. In 2011 27th IEEE International Conference on Software Maintenance (ICSM) (pp. 438–447). IEEE.

[7] Manjunath, T., Hegadi, R. S., & Mohan, H. S. (2011). Automated data validation for data migration security. International Journal of Computer Applications, 30(6), 41–46.

[8] Soewito, B., Gaol, F. L., & Abdurachman, E. (2022). A systematic literature review: Risk analysis in cloud migration. Journal of King Saud University-Computer and Information Sciences, 34(6), 3111–3120.

[9] Polamreddy, V. R. (2021). Engineering reversible enterprise data migrations: A phased rollout framework for financially critical retail platforms. International Journal of Computer Technology and Electronics Communication, 4(2), 3414–3427.

[10] Gholami, M. F., Daneshgar, F., Beydoun, G., & Rabhi, F. (2017). Challenges in migrating legacy software systems to the cloud—an empirical study. Information Systems, 67, 100–113.

[11] Miake-Lye, I. M., Cogan, A. M., Mak, S., Brunner, J., Rinne, S., Brayton, C. E., et al. (2023). Transitioning from one electronic health record to another: A systematic review. Journal of General Internal Medicine, 38(Suppl 4), 956–964.