International Journal of Multidisciplinary Research and Growth Evaluation.

Data Quality and Master Data Management for Finance Data on Cloud

Suhas Hanumanthaiah
Independent Research, USA

* Corresponding Author: **Suhas Hanumanthaiah**

Article Info

ISSN (online): 2582-7138
Volume: 01
Issue: 01
March-April 2020
Received: 02-03-2020
Accepted: 01-04-2020
Page No: 148-153

Abstract
This paper explores the critical intersection of data quality and master data management (MDM) within cloud-based financial data systems. Financial institutions are increasingly migrating to the cloud for scalability and cost-efficiency, but this transition introduces challenges in maintaining data integrity and consistency. A robust data governance framework is essential to define roles, establish standards, and implement policies for data access and security. The paper examines best practices for data profiling, cleansing, validation, and monitoring, highlighting the importance of data security and privacy in the cloud. Furthermore, the study investigates MDM tools and technologies, emphasizing cloud-based solutions that offer data integration, cleansing, matching, and governance capabilities. Data quality metrics, including accuracy, completeness, consistency, timeliness, validity, and uniqueness, are discussed in the context of financial data. Integration strategies, such as real-time and batch data integration, data virtualization, and API-based integration, are also explored. The paper presents a comparative analysis of cloud tools, data quality metrics, and data size limits, along with case studies that illustrate the successful implementation of data quality and MDM initiatives in financial institutions. Future trends, such as AI/ML, blockchain, serverless computing, federated learning and data mesh architecture are also discussed.

DOI: https://doi.org/10.54660/.IJMRGE.2020.1.1.148-153

Keywords: Data Quality, Master Data Management, Cloud Computing, Finance Data, Data Governance, MDM Tools, Data Integration, Data Security, Regulatory Compliance, AI/ML, Blockchain

1. Introduction

In today's rapidly evolving digital landscape, financial institutions are increasingly adopting cloud computing to enhance agility, reduce costs, and improve scalability. However, migrating financial data to the cloud introduces significant challenges related to data quality and master data management (MDM). High-quality data is crucial for accurate financial reporting, regulatory compliance, and effective decision-making. Master data, which includes key entities such as customers, products, and accounts, must be consistent and reliable across the organization. The convergence of cloud computing, data quality, and MDM is essential for financial institutions to unlock the full potential of their data assets in the cloud. This paper explores the best practices, tools, metrics, and integration strategies for ensuring data quality and effective MDM of finance data on the cloud.

This paper aims to provide a comprehensive overview of the critical aspects of data quality and MDM in the context of cloud-based financial data management. It addresses the unique challenges and opportunities presented by the cloud environment and offers practical guidance for financial institutions seeking to leverage cloud technologies while maintaining the highest standards of data governance. The research will systematically evaluate evidence, assess methodological strengths, and consider alternative explanations to provide a balanced and insightful perspective. By comparing findings across multiple papers, analyzing similarities and differences, and synthesizing various perspectives, this paper aims to build a comprehensive understanding of the subject matter.

2. Literature Review

2.1 The Importance of Data Quality in Finance

Data quality is paramount in the financial sector, where decisions are heavily reliant on accurate and reliable information. Poor data quality can lead to flawed financial reporting, regulatory non-compliance, and operational inefficiencies. Financial institutions must adhere to stringent regulatory requirements, such as Basel II and Solvency II, which mandate formal data quality standards [1]. These regulations underscore the importance of data governance and centralized data management approaches. The financial crisis of 2007 highlighted the critical need for risk managers to address data quality concerns, particularly in the context of siloed risk data [1].

2.2 Master Data Management (MDM) in Business Transformation

Master Data Management (MDM) plays a pivotal role in business transformation programs (BTPs) by ensuring data accuracy and consistency across various lines of business and channels [2]. Effective MDM involves assigning responsibility to people, technology, and processes for capturing, maintaining, and defining data accuracy based on established rules [2]. Multiple parameters, such as change management, senior management commitment, compliance governance policies, and pre-existing data quality challenges, can impact the success of BTPs [2]. A well-designed MDM strategy enhances decision-making by providing a unified view of master data, reinforcing enterprise data governance, and safeguarding data lifecycle practices [2].

2.3 Cloud Computing and its Impact on Financial Data Management

Cloud computing offers numerous benefits for financial institutions, including scalability, cost reduction, and improved agility [3]. Cloud service providers offer various service models, such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS), which can be tailored to meet specific business needs [4]. However, migrating financial data to the cloud also introduces security and privacy concerns that must be addressed through robust data governance and compliance frameworks [5]. The economic advantages of cloud adoption, such as avoiding capital expenditures and usage-based billing, make it an attractive IT paradigm for many organizations [3].

2.4 Challenges and Opportunities of Data Quality and MDM on Cloud

While cloud computing presents significant opportunities for financial data management, it also introduces unique challenges related to data quality and MDM. These challenges include ensuring data security and privacy, maintaining data consistency across distributed cloud environments, and integrating data from various sources and formats. To address these challenges, financial institutions must implement robust data governance policies, invest in data quality tools and technologies, and adopt a centralized approach to MDM. The integration of cloud computing with big data technologies can further enhance data quality and MDM by enabling real-time data processing and analysis [6].

3. Best Practices for Data Quality in Cloud-Based Finance Data

3.1 Data Governance Framework

A robust data governance framework is essential for ensuring data quality in cloud-based finance data. This framework should define roles and responsibilities, establish data quality standards, and implement policies for data access, security, and compliance. The framework should also include processes for monitoring data quality, identifying and resolving data quality issues, and continuously improving data quality over time. According to Lepeniotis, change management and compliance governance policies are crucial parameters that can impact the success of business transformation programs [2].

3.2 Data Profiling and Assessment

Data profiling is the process of examining data to understand its structure, content, and quality. It involves analyzing data to identify inconsistencies, inaccuracies, and missing values. Data profiling should be performed regularly to assess the quality of finance data on the cloud and identify areas for improvement. The results of data profiling can be used to develop data quality rules and metrics and to prioritize data quality initiatives.

3.3 Data Cleansing and Standardization

Data cleansing involves correcting or removing inaccurate, incomplete, or inconsistent data. Data standardization involves converting data to a consistent format and structure. These processes are essential for ensuring that finance data on the cloud is accurate, reliable, and usable for reporting and analysis. Data cleansing and standardization should be performed as part of the data integration process to ensure that data from various sources is consistent and compatible.

3.4 Data Validation and Monitoring

Data validation involves verifying that data conforms to predefined rules and constraints. Data monitoring involves tracking data quality metrics over time to identify trends and anomalies. These processes are essential for ensuring that finance data on the cloud remains accurate and reliable. Data validation and monitoring should be automated as much as possible to reduce manual effort and improve efficiency.

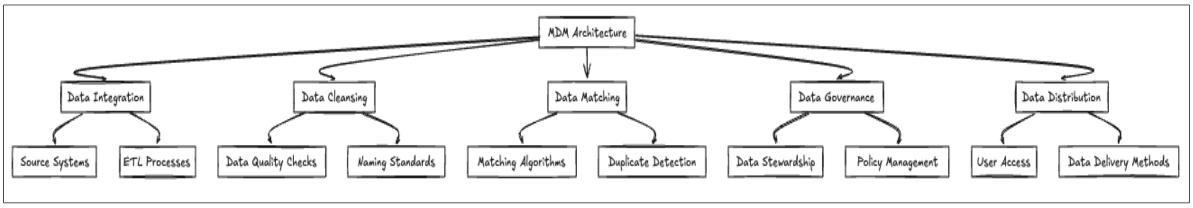
3.5 Data Security and Privacy

Data security and privacy are critical concerns for financial institutions migrating data to the cloud. Financial institutions must implement robust security measures to protect sensitive data from unauthorized access, use, or disclosure. These measures should include encryption, access controls, and data masking. Financial institutions must also comply with data privacy regulations, such as GDPR and CCPA, which require them to protect the privacy of customer data.

4. Master Data Management (MDM) Tools and Technologies

4.1 MDM Architecture and Components

A well-designed MDM architecture is essential for effectively managing master data in the cloud. The architecture should include components for data integration, data cleansing, data matching, data governance, and data distribution. Data integration involves consolidating master data from various sources into a central repository. Data cleansing involves correcting or removing inaccurate, incomplete, or inconsistent master data. Data matching involves identifying and merging duplicate master data records. Data governance involves establishing policies and procedures for managing master data. Data distribution involves disseminating master data to various systems and applications.



4.2 Cloud-Based MDM Solutions

Several cloud-based MDM solutions are available to help financial institutions manage their master data in the cloud. These solutions offer features such as data integration, data cleansing, data matching, data governance, and data distribution. Some popular cloud-based MDM solutions include Informatica MDM Cloud, Tibco EBX, and Riversand MDM. These solutions provide a centralized platform for managing master data and ensuring data consistency across the organization.

4.3 Data Integration Tools

Data integration tools are used to consolidate data from various sources into a central repository. These tools offer features such as data mapping, data transformation, and data validation. Some popular data integration tools include Informatica PowerCenter, IBM DataStage, and Talend Open Studio. These tools enable financial institutions to integrate data from various sources, including on-premise systems, cloud applications, and external data providers.

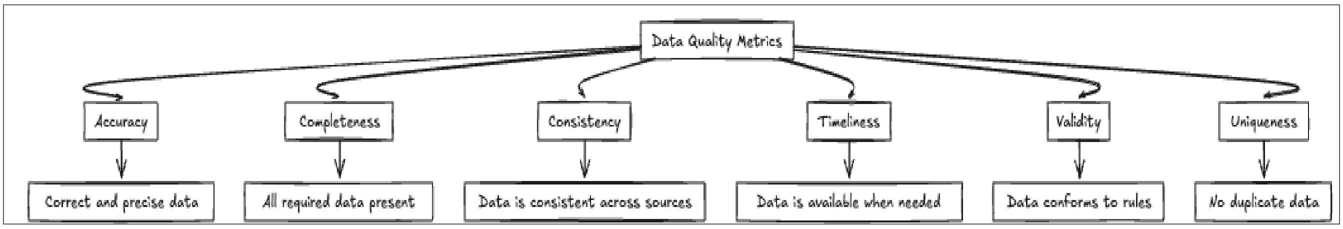
4.4 Data Quality Tools

Data quality tools are used to profile, cleanse, and monitor data quality. These tools offer features such as data profiling, data cleansing, data validation, and data monitoring. Some popular data quality tools include Informatica Data Quality, Trillium Software, and Experian Data Quality. These tools help financial institutions identify and resolve data quality issues and ensure that their data is accurate, reliable, and usable.

4.5 Data Matching and Deduplication Tools

Data matching and deduplication tools are used to identify and merge duplicate data records. These tools offer features such as fuzzy matching, probabilistic matching, and rule-based matching. Some popular data matching and deduplication tools include IBM InfoSphere Identity Insight, SAS Data Management, and Pitney Bowes Spectrum Technology Platform. These tools help financial institutions eliminate duplicate data and ensure that their master data is accurate and consistent.

5. Data Quality Metrics for Finance Data on Cloud



5.1 Accuracy

Accuracy refers to the degree to which data correctly reflects the real-world entities they are intended to represent. In the context of finance data, accuracy is crucial for ensuring that financial reports and analyses are based on reliable information. Metrics for measuring accuracy include error rates, data validation pass rates, and comparisons to trusted data sources.

5.2 Completeness

Completeness refers to the extent to which all required data is present. Incomplete data can lead to inaccurate financial reporting and flawed decision-making. Metrics for measuring completeness include the percentage of missing values, the number of records with incomplete information, and the percentage of required fields that are populated.

5.3 Consistency

Consistency refers to the degree to which data is uniform and coherent across different systems and databases. Inconsistent data can lead to discrepancies in financial reports and analyses. Metrics for measuring consistency include the number of data conflicts, the percentage of records with consistent values across systems, and the number of data reconciliation errors.

5.4 Timeliness

Timeliness refers to the availability of data when it is needed. In the financial sector, timely data is essential for making informed decisions and responding to market changes. Metrics for measuring timeliness include data latency, data refresh rates, and the percentage of reports delivered on time.

5.5 Validity

Validity refers to the degree to which data conforms to predefined rules and constraints. Invalid data can lead to errors in financial reporting and analysis. Metrics for measuring validity include the number of data validation errors, the percentage of records that pass data validation rules, and the number of data quality exceptions.

5.6 Uniqueness
Uniqueness refers to the absence of duplicate data records. Duplicate data can lead to inflated financial figures and inaccurate analyses. Metrics for measuring uniqueness include the number of duplicate records, the percentage of unique records, and the number of data deduplication errors.

6. Integration Strategies for Data Quality and MDM on Cloud
6.1 Data Integration Architecture
A well-defined data integration architecture is essential for integrating data quality and MDM processes in the cloud. The architecture should include components for data extraction, data transformation, data loading, and data governance. Data extraction involves retrieving data from various sources. Data transformation involves converting data to a consistent format and structure. Data loading involves loading data into a central repository. Data governance involves establishing policies and procedures for managing data.

6.2 Real-time Data Integration
Real-time data integration involves integrating data as it is created or updated. This approach is essential for ensuring that finance data on the cloud is always up-to-date and accurate. Real-time data integration can be achieved using technologies such as change data capture (CDC), message queues, and streaming data platforms.

6.3 Batch Data Integration
Batch data integration involves integrating data in scheduled batches. This approach is suitable for integrating data from systems that do not support real-time integration. Batch data integration can be achieved using technologies such as ETL (extract, transform, load) tools and data warehousing platforms.

6.4 Data Virtualization
Data virtualization involves creating a virtual layer that provides a unified view of data from various sources without physically moving the data. This approach can be used to integrate data quality and MDM processes without disrupting existing systems. Data virtualization tools provide a single point of access to data from various sources, enabling users to access and analyze data without having to understand the underlying data structures and formats.

6.5 API-Based Integration
API-based integration involves using application programming interfaces (APIs) to integrate data quality and MDM processes. This approach is particularly useful for integrating cloud-based applications and services. APIs provide a standardized way for applications to communicate with each other, enabling seamless data exchange and integration.

7. Cloud Tool vs. DQ Metrics, MDM Metrics, and Data Size Limits

Table 1: Here is a summary of top commercial tools and their support for DQ Metrics, MDM Metrics and Data Size limits

Cloud Tool	DQ Metrics Supported	MDM Metrics Supported	Data Size Limits
Informatica MDM	Accuracy, Completeness, Consistency, Timeliness, Validity, Uniqueness	Data Stewardship, Data Matching Accuracy, Time to Resolution	Scales to petabytes; depends on cloud infrastructure and licensing
Tibco EBX	Accuracy, Completeness, Consistency, Validity	Data Governance, Data Quality Rule Adherence, Workflow Efficiency	Scales to terabytes; depends on cloud infrastructure and licensing
Riversand MDM	Accuracy, Completeness, Consistency, Timeliness, Validity	Data Syndication, Time to Market, Data Completeness for Key Attributes	Scales to petabytes; depends on cloud infrastructure and licensing
AWS Glue	Completeness, Validity	N/A	Scales to petabytes; billed based on usage
Google Cloud Dataprep	Accuracy, Completeness, Consistency, Validity	N/A	Scales to petabytes; billed based on usage
Azure Data Factory	Completeness, Validity	N/A	Scales to petabytes; billed based on usage

8. Case Studies
8.1 Case Study 1: Improving Data Quality in a Large Financial Institution
A large financial institution implemented a data quality initiative to improve the accuracy and reliability of its finance data on the cloud. The initiative involved implementing a data governance framework, performing data profiling and assessment, and implementing data cleansing and standardization processes. The results of the initiative included a significant reduction in data errors, improved financial reporting accuracy, and enhanced regulatory compliance.

8.2 Case Study 2: Implementing MDM in a Multinational Bank
A multinational bank implemented an MDM solution to manage its customer data across various lines of business and geographies. The solution involved consolidating customer data from various systems into a central repository, cleansing and standardizing the data, and implementing data governance policies. The results of the implementation included a unified view of customer data, improved customer service, and enhanced cross-selling opportunities.

8.3 Case Study 3: Cloud Migration for a Fintech Company
A fintech company migrated its finance data to the cloud to improve scalability and reduce costs. The migration involved implementing data security measures, ensuring data consistency across the cloud environment, and integrating data from various sources and formats. The results of the migration included improved agility, reduced IT costs, and

enhanced data accessibility.

9. Challenges and Mitigation Strategies

9.1 Data Security and Privacy

Challenge: Ensuring the security and privacy of sensitive financial data in the cloud is a major concern. Cloud environments are susceptible to various security threats, including data breaches, malware attacks, and unauthorized access.

Mitigation Strategies

- Implement robust encryption techniques to protect data at rest and in transit.
- Enforce strict access controls to limit access to sensitive data based on the principle of least privilege.
- Regularly monitor and audit cloud environments for security vulnerabilities and compliance violations.
- Comply with data privacy regulations, such as GDPR and CCPA, to protect customer data.
- Implement data masking and tokenization techniques to protect sensitive data during processing and analysis.

9.2 Data Consistency and Integrity

Challenge: Maintaining data consistency and integrity across distributed cloud environments can be challenging. Data replication and synchronization issues can lead to inconsistencies and errors in financial reports and analyses.

Mitigation Strategies

- Implement robust data replication and synchronization mechanisms to ensure data consistency across all cloud environments.
- Use data validation rules and constraints to ensure data integrity.
- Regularly reconcile data across different systems to identify and resolve inconsistencies.
- Implement version control mechanisms to track data changes and ensure data lineage.
- Use data quality tools to monitor data quality and identify data quality issues.

9.3 Data Integration Complexity

Challenge: Integrating data from various sources and formats can be complex and time-consuming. Financial institutions often have data stored in various on-premise systems, cloud applications, and external data providers.

Mitigation Strategies

- Use data integration tools to consolidate data from various sources into a central repository.
- Implement a data integration architecture that supports real-time and batch data integration.
- Use data virtualization techniques to provide a unified view of data without physically moving the data.
- Use APIs to integrate data quality and MDM processes.
- Standardize data formats and structures to simplify data integration.

9.4 Data Governance and Compliance

Challenge: Establishing and maintaining data governance and compliance in the cloud can be challenging. Financial institutions must comply with various regulatory requirements and internal policies.

Mitigation Strategies

- Implement a data governance framework that defines roles and responsibilities, establishes data quality standards, and implements policies for data access, security, and compliance.
- Use data governance tools to automate data governance processes and enforce data governance policies.
- Regularly monitor and audit data governance and compliance activities.
- Provide training and awareness programs to educate employees about data governance and compliance requirements.
- Engage with regulatory agencies to understand and comply with regulatory requirements.

9.5 Skill Gap

Challenge: A shortage of skilled professionals with expertise in cloud computing, data quality, and MDM can hinder the successful implementation of data quality and MDM initiatives in the cloud.

Mitigation Strategies

- Invest in training and development programs to upskill existing employees.
- Hire skilled professionals with expertise in cloud computing, data quality, and MDM.
- Partner with cloud service providers and consulting firms to leverage their expertise.
- Establish centers of excellence to foster knowledge sharing and best practices.
- Participate in industry events and conferences to stay up-to-date on the latest trends and technologies.

10. Future Trends

10.1 Artificial Intelligence (AI) and Machine Learning (ML) for Data Quality

AI and ML technologies are increasingly being used to automate data quality processes and improve data accuracy. ML algorithms can be used to identify data anomalies, detect data quality issues, and predict data quality trends. AI-powered data quality tools can automate data profiling, data cleansing, and data validation processes, reducing manual effort and improving efficiency. Privacy-preserving machine learning (PPML) techniques can also be applied in healthcare [7].

10.2 Blockchain for Data Integrity

Blockchain technology can be used to ensure the integrity and provenance of finance data on the cloud. Blockchain provides a secure and transparent way to track data changes and prevent data tampering. Blockchain-based data quality solutions can be used to verify the accuracy and reliability of finance data and to ensure compliance with regulatory requirements. Blockchain technology offers security, anonymity, and integrity without third-party control [8].

10.3 Serverless Computing for Data Processing

Serverless computing provides a cost-effective and scalable way to process large volumes of finance data in the cloud. Serverless platforms enable developers to define their application through service functions, relieving them of infrastructure management tasks [9]. Serverless computing can be used to implement data quality and MDM processes without having to manage servers or infrastructure.

10.4 Federated Learning for Privacy-Preserving Data Analysis

Federated learning enables collaborative training of machine learning models without sharing raw data. This approach is particularly useful for analyzing sensitive finance data in the cloud while preserving data privacy. Federated learning can be used to develop data quality models and MDM solutions without compromising data security or compliance.

10.5 Data Mesh Architecture

The data mesh architecture promotes a decentralized approach to data ownership and management. In a data mesh, data is treated as a product, and each business domain is responsible for managing its own data. This approach can improve data quality and MDM by empowering business domains to take ownership of their data and implement data governance policies that are tailored to their specific needs.

11. Conclusion

Ensuring data quality and effective MDM of finance data on the cloud is essential for financial institutions to unlock the full potential of their data assets and achieve their business objectives. By implementing best practices, leveraging appropriate tools and technologies, and adopting a robust data governance framework, financial institutions can overcome the challenges and capitalize on the opportunities presented by cloud computing. The future trends of AI/ML, blockchain, serverless computing, federated learning, and data mesh architecture offer promising avenues for further enhancing data quality and MDM in the cloud. This paper provides a comprehensive guide for financial institutions seeking to navigate the complexities of data quality and MDM in the cloud and to establish a solid foundation for data-driven decision-making.

12. Reference

1. Bonollo M, Neri M. Data quality in banking: Regulatory requirements and best practices. J Risk Manag Financ Inst. 2012. https://doi.org/10.69554/mhhf7098
2. Lepeniotis P. Master data management: its importance and reasons for failed implementations. 2019. https://doi.org/10.7190/shu-thesis-00311
3. Economics of Cloud Computing. IGI Global. 2013. https://doi.org/10.4018/978-1-4666-4683-4.ch007
4. Telenyk S, Bidyuk P, Zharikov E, Yasochka M. Assessment of cloud service provider quality metrics. 2017. https://doi.org/10.1109/UKRMICO.2017.8095422
5. Jansen W, Grance T. Guidelines on security and privacy in public cloud computing. 2010. https://doi.org/10.6028/nist.sp.800-144
6. Zeng W. Big Data Architecture, Platform, Application and Trend. 2016. https://doi.org/10.12783/DTETR/SSME-IST2016/3998
7. Fritchman K, *et al*. Privacy-Preserving Scoring of Tree Ensembles: A Novel Framework for AI in Healthcare. 2018. https://doi.org/10.1109/bigdata.2018.8622627
8. Yli-Huumo J, Ko D, Choi S, Park S, Smolander K. Where Is Current Research on Blockchain Technology? A Systematic Review. PLoS ONE. 2016. https://doi.org/10.1371/journal.pone.0163477
9. Kuhlenkamp J, Werner S, Borges MC, Tal KE, Tai S. An Evaluation of FaaS Platforms as a Foundation for Serverless Big Data Processing. Proc Int Conf Util Cloud Comput. 2019. https://doi.org/10.1145/3344341.3368796