Efficiency Assessment System for Resource Utilization in Commercial Bank Big Data Platform

Yanru Li*

University of Warwick, Coventry, the United Kingdom

*2482516799@qq.com

Abstract. This paper addresses the essential need for efficiently managing resource utilization within big data platforms, a critical component of modern banking operations amidst digital transformation. It aims to provide a structured methodology that not only identifies the challenges associated with escalating data volumes but also offers practical solutions. The primary objective is to enhance the overall efficiency and performance of commercial bank big data platforms by introducing an innovative efficiency scoring system tailored for this specific context. The paper amalgamates traditional library resource management approaches with the intricacies of the credit scoring system within the domain of consumer finance, resulting in a holistic efficiency scoring system. This system serves as the foundation for a comprehensive resource cost management strategy, marking a paradigm shift from ad-hoc resource governance. The proposed methodology encompasses a hierarchical structural model, incorporates a meticulous calculation model, and comprises multiple efficiency indicators. Expert consultation and feedback ensure the reliability and applicability of this framework. Our methodology offers a robust foundation for operational governance, ensuring efficient resource utilization and, consequently, enhancing the quality of operational services within big data platforms in the banking sector. It contributes to the advancement of resource governance in the context of ever-evolving digital landscapes.

Keywords: big data, cloud computing, resource utilization efficiency, banking sector, assessment model, simulated data validation.

1 Introduction

The banking sector's significant embrace of big data is not just a trend but a strategic necessity for progress amid digital transformation. Large-scale commercial banks, both domestic and international, are launching innovative digital projects, like distributed data centers and intelligent computing hubs, driven by big data platforms. These endeavors invigorate digital operations, enhancing operational service quality and fostering innovative financial models.

Yet, as banks expand their range of financial products and services, challenges surface. Firstly, there are rising data costs due to expanding data volumes and the

C. Chen et al. (eds.), *Proceedings of the 3rd International Conference on Digital Economy and Computer Application (DECA 2023)*, Atlantis Highlights in Computer Sciences 17,
https://doi.org/10.2991/978-94-6463-304-7_45

increasing importance of real-time data processing for business applications. However, expanding hardware resources to meet these demands results in higher equipment and operational costs. Secondly, there's a lack of cost management awareness in the IT sector, where while offering robust business support, there's room for improvement in cost management. Constructing big data platforms often lacks a clear cost breakdown and an efficient assessment mechanism, essential for reflecting the significant costs of daily data consumption and platform operation health. Lastly, addressing the complexity of governing redundant resources poses a challenge. Misusing redundant resources leads to resource wastage, but identifying and resolving redundant computation and storage issues is daunting. Existing tools and mechanisms struggle to automate issue recognition and facilitate effective resolutions.

In response, this article aims to establish a comprehensive methodology. This methodology not only addresses these challenges but also maintains a cyclic solution framework. It ensures ongoing effectiveness in addressing emerging challenges while preserving a sustainable solution mechanism [1].

2 Research Methodology

The analysis and design of this study employed a range of rigorous research methods to establish a comprehensive framework for assessing the efficiency of resource utilization within banking big data platforms. Firstly, a thorough Literature Analysis was conducted, utilizing various sources such as university libraries, journals, online platforms, electronic media, and industry documents. This extensive literature review enabled the creation of a robust theoretical framework for the study. Furthermore, it involved a systematic investigation and analysis of a prominent domestic commercial bank's big data platform, ensuring the theoretical aspects were firmly grounded in scientific rigor and the data accuracy was comprehensive [2]. In addition, the study utilized the Delphi Method to refine the evaluation index system, enhancing its scientific and rational nature. The Delphi Method involved soliciting expert opinions and conducting questionnaire surveys. Experts were invited to provide repeated consultations and ratings for each indicator within the evaluation index system. The selection of a well-qualified team of evaluation experts was paramount in ensuring the objectivity and accuracy of the weights derived through the Delphi Method. By employing a combination of these research methods, this study aims to establish a robust and validated framework for assessing resource utilization efficiency within banking big data platforms. The integration of literature analysis and the Delphi Method not only strengthens the theoretical foundation but also ensures the practical applicability of the proposed framework. [2-3].

3 Design Contents

While the underlying causes of operational challenges within big data platforms may vary, discernible patterns have emerged in the strategies employed to address these challenges over time. Typically, the collection of pertinent data is facilitated through tools that are responsive to specific demands and issues, followed by meticulous manual analysis and subsequent problem resolution. However, as the process of digital transformation in the banking sector continues to deepen, there has been a marked escalation in data volume, task complexity, and application proliferation. Simultaneously, the evolving landscape of business scenarios places heightened demands on data service quality and efficiency. Consequently, the conventional operational methodologies have proven inadequate in effectively confronting the multifaceted challenges posed by this evolving terrain. Considering these circumstances, our team has embarked on an innovative path, synthesizing the conventional library resource management framework with the intricacies of the individual credit scoring system within the domain of consumer finance. This innovative amalgamation has culminated in the establishment of a holistic efficiency scoring system tailored specifically for big data platforms within commercial banks. Furthermore, the utilization of this scoring system has been instrumental in crafting a meticulously structured data resource cost management strategy. This comprehensive and sustainable approach signifies a paradigm shift in resource governance for big data platforms, effectively mitigating the hitherto prevalent ad-hoc resource management [4].

3.1 Efficiency Indicator Design

The primary design objective for the efficiency score is to significantly enhance resource utilization quality, with a specific focus on governing storage resources, computing resources, and regulatory compliance. To achieve this objective, we have structured our governance framework around three core subjects: storage resources, computing resources, and regulatory compliance. This framework follows the OSM (Objective, Subjective, Management) methodology, which allows us to categorize and address efficiency-related issues comprehensively. To adapt to diverse platform characteristics and account for key influencing factors such as the number of tables, tasks, and resource consumption, we have meticulously refined the design of sub-indicators for efficiency scoring, deduction principles, and calculation rules[5]. To ensure the applicability of our efficiency scores as a foundation for operational governance, we have conducted validation using simulated data. The foundation for calculating efficiency scores in our assessment model relies on a set of efficiency evaluation indicators. These indicators are pivotal and serve as the basis for assessing the efficiency of the big data platform. Within this context, our assessment model encompasses four core governance subjects: storage resources, computing resources, resource utilization compliance, and resource utilization security. We have meticulously devised a total of 50 efficiency indicators, which are organized into nine major problem categories. Each indicator is accompanied by essential information, including its name, calculation methodology, weight, governance approach, and contribution to resource conservation. These indicators serve as critical inputs for calculating the efficiency score in our assessment model.

To ensure the comprehensiveness and reliability of our assessment model, we engaged a panel of three experts with extensive experience in big data platform operation and management, drawn from pertinent industries and technical domains. Their expertise spans various aspects of big data platform construction, maintenance, application, and promotion. Through a rigorous process of consultation, feedback, and statistical analysis, we consolidated their judgments and proposed modifications for each indicator within the preliminary indicator set. After several rounds of refinement, we arrived at a consensus on the candidate indicators[6].

The ultimate big data platform resource utilization assessment indicator system, because of this collaborative effort, comprises three primary indicators and ten secondary indicators. These indicators are detailed in Table 1 below, providing a comprehensive and formal framework for assessing and enhancing the efficiency of big data platform resource utilization.

Table 1. Big Data Platform Resource Utilization Assessment Indicator System

Primary Indicators	**Secondary Indicators**	**Example Inspection Indicators**
storage Resource Efficiency	Storage Strategy Issues	Unmanaged Table Check
		Data Table Small File Threshold Check
	Redundant Storage Issues	Data Archiving Policy Check
		Data Storage Policy Check
Computational Resource Efficiency	Task Optimization Space	Data Skew Check
		Bruteforce Scanning Check
	Task Execution Issues	Input/Output Empty Check
		Continuous Output Consistency Check
	Data Processing Chain Issues	Data Processing Pipeline Issues Check
Resource Usage Compliance	Encoding Compliance	Column Pruning Check
		Predicate Pushdown Check
	Model Consistency	Incremental Processing Check
		Data Model and Database Consistency Check
Resource Usage Security	Data Security	Access Policy Check
		Sharing Policy Check

3.2 Efficiency Assessment Model Design

3.2.1 Establishing a Hierarchical Structure Model

In accordance with the assessment index framework for the utilization of resources within a big data platform, a hierarchical structural model has been meticulously devised. This model comprises three distinct layers: the foundational Objective Layer, the intermediate Criteria Layer, and the operational Scheme Layer, as visually elucidated in Figure 1.

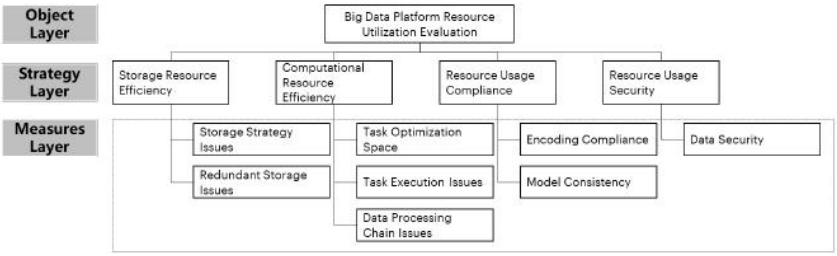

Fig. 1. Hierarchical Structure Model of Evaluation Indicators for Big Data Platform Resource Utilization

Three experts conducted independent assessments of the strategy-level indicators. Table 2 presents the judgment matrix for one of these experts, along with a consistency ratio of 0.021, indicating an acceptable level of consistency. The weights assigned to the four indicators were determined as follows: 0.358, 0.381, 0.097, and 0.164.

Table 2. Judgment Matrix of Importance Ratings for Primary Indicators by One Expert

Strategy Layer	Storage Resource Efficiency	Computational Resource Efficiency	Resource Usage Compliance	Resource Usage Security	Weights
Storage Resource Efficiency	1	5/4	3	2	0.35814618
Computational Resource Efficiency	5/4	1	4	3	0.38123771
Resource Usage Compliance	4/5	1/4	1	1/2	0.09699751
Resource Usage Security	1/3	1/3	2	1	0.1636186
Consistency Ratio	0.0210325045584869				

Utilizing the method outlined above, the expert-generated data from each judgment matrix undergoes a meticulous processing and aggregation procedure. This meticulous process culminates in the derivation of aggregated weights for all judgment matrices.

3.2.2 Calculation Model Design

Within a resource governance framework, a structured scoring system based on indicators can be designed. The main goal of this system is to identify and solve resource utilization problems and ultimately optimize the resource usage behavior of the big data platform. Based on this, we chose the deduction system as the basic calculation logic. In the banking industry, information technology (IT) plays a vital role as a business support function. When assessing IT's resource utilization efficiency, a holistic view of resource consumption must be considered. This includes the total number of forms created, the total number of tasks performed, and the resource utilization associated with the forms created.

To ensure that assessment scores remain rigorous and comparable across contexts, specific parameters need to be carefully defined. Use standardized rule sets when evaluating various dimensions within a banking organization (e.g. centre/department/product). Within each assessment dimension, all assessment objects, whether forms or tasks, are included in the assessment scope. A platform's efficiency score within a given evaluation range is defined as the sum of three independent components: storage efficiency, compute efficiency, and compliance efficiency score. These scores are calculated based on a weighted sum of the scores for their respective secondary indicators.

Considering the factors that affect deduction points, the amount of deduction points for each inspection indicator is initially determined based on the total number of objects to be evaluated within the evaluation scope. In addition, an adjustment factor is calculated based on the ratio of the resource consumption of the object being evaluated to the average resource consumption of all objects within the evaluation scope. The final penalty for a single event is the product of the penalty baseline and the adjustment factor. The deduction amount for a single indicator is the sum of the deduction points for all events related to the indicator.

To illustrate this evaluation model, consider a specific banking center using the Hadoop platform. In the storage resource efficiency category, there are m secondary indicators, each indicator is related to a weight Wj, where j=1,...,m. The score of each secondary indicator is expressed as Score(j).

In the context of table evaluation, we use the variable Oti to represent the status of table t under inspection indicator i. When a table is not marked as having any problems, we record Oti as 0. Therefore, the department can construct a hit matrix with dimensions ($T \times n$), where T represents the total number of tables and n represents the number of inspection indicators.

In our evaluation framework, "100/T" represents the penalty baseline, where "T" represents the total number of objects evaluated within the defined evaluation scope. This baseline is calculated as 100 (representing a perfect score) divided by the total number of objects to be evaluated, specifically the table in this example. On the other hand, "S_t/(Avg(T))" represents the adjustment coefficient, which is based on the resource utilization of the evaluated object being checked divided by the average resource utilization of all evaluated objects within the evaluation scope.

Combining the above points, scores can be calculated for computing resource efficiency, compliance with resource usage regulations, and security of resource utilization. These individual assessments are ultimately combined into an assessment score that represents how efficiently the platform resources are utilized.

4 Conclusion

The hierarchical analysis results reveal a distinct hierarchy of primary indicators, with computational resource efficiency holding the highest weight, followed by storage resource efficiency, resource usage security, and resource usage compliance. Within the secondary indicators, greater emphasis is placed on issues such as redundant storage, task execution problems, and data security. This underscores the consensus among experts that the effective utilization of computational and storage resources is central to the comprehensive evaluation of platform efficiency. Simultaneously, considerations regarding resource usage compliance and security remain integral [9-10].

This paper endeavors to contribute to the enhancement of daily operational optimization for big data platforms by constructing evaluation indicators for resource usage efficiency and designing an evaluation model. This model offers valuable guidance for assessing and comparing resource usage efficiency across diverse scenarios, encompassing various banking and enterprise-specific contexts. It can be tailored to accommodate varying indicators and their respective weights, enabling a nuanced evaluation of resource usage efficiency in distinct big data platform environments. This approach aligns with the formal and logical style typically associated with academic research papers.

References

1. VenkateswaraRao, M., et al. "Credit Investigation and Comprehensive Risk Management System based Big Data Analytics in Commercial Banking." 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS). Vol. 1. IEEE, 2023.
2. Wang, Yang, Sui Xiuping, and Qi Zhang. "Can fintech improve the efficiency of commercial banks?—An analysis based on big data." Research in international business and finance 55 (2021): 101338.
3. Cheng, Binqi, and Weijie Feng. "Analysis of the Application of Big Data in Banking Sector." 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). IEEE, 2021.
4. Hung, Jui-Long, Wu He, and Jiancheng Shen. "Big data analytics for supply chain relationship in banking." Industrial Marketing Management 86 (2020): 144-153.
5. Wang, Yang, Sui Xiuping, and Qi Zhang. "Can fintech improve the efficiency of commercial banks? —An analysis based on big data." Research in international business and finance 55 (2021): 101338.
6. Hung, Jui-Long, Wu He, and Jiancheng Shen. "Big data analytics for supply chain relationship in banking." Industrial Marketing Management 86 (2020): 144-153.
7. Dicuonzo, Grazia, et al. "Risk management 4.0: The role of big data analytics in the bank sector." International Journal of Economics and Financial Issues 9.6 (2019): 40-47.
8. Chen, Xihui, Xuyuan You, and Victor Chang. "FinTech and commercial banks' performance in China: A leap forward or survival of the fittest?." Technological Forecasting and Social Change 166 (2021): 120645.

9. Mikalef, Patrick, et al. "Big data and business analytics: A research agenda for realizing business value." (2019).
10. Ajah, Ifeyinwa Angela, and Henry Friday Nweke. "Big data and business analytics: Trends, platforms, success factors and applications." Big Data and Cognitive Computing 3.2 (2019): 32.