Advances in Computer, Signals and Systems (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2026.100101
ISSN 2371-8838 Vol. 10 Num. 1

Research on the Construction of an Integrated Information Resource Database Based on Cloud Computing Technology

Shenpeiyilin

Amazon Web Servicees Inc, Arlington County, Virginia, USA

Keywords: Cloud Computing; Information Resource Integration; Database Construction; Distributed Storage; Security and Performance Optimization

Abstract: With the rapid advancement of information technology, information resources are characterized by their massive scale, diverse sources, and heterogeneous formats. Traditional databases can no longer meet the growing demand for efficient integration and sharing of information resources. Cloud computing, with its powerful computing capability, elasticity, and distributed storage, offers a new approach to building integrated information resource databases. Based on an analysis of the current state of information resources and system requirements, this study proposes a cloud computing–based framework for constructing such a database. The framework systematically addresses platform architecture design, data model development, and security control mechanisms, while incorporating methods for data collection, cleaning, storage, and retrieval. An extensible and reliable integrated database is thereby established. Through system implementation and case analysis, the feasibility and effectiveness of the framework in enhancing data sharing efficiency, system security, and performance optimization are validated. The results indicate that the proposed approach not only improves resource utilization and service capabilities but also provides a practical pathway for cross-domain information integration and application.

1. Introduction

Against the backdrop of rapid digital transformation, information resources have become a vital foundation for driving social progress and knowledge innovation. However, current information resources are often widely dispersed, highly diverse in format, and uneven in quality, making it difficult for traditional databases to achieve efficient integration and sharing. With the continuous maturation of cloud computing technology, its strengths in computing power, distributed storage, elastic scalability, and resource scheduling provide strong technical support for the construction of integrated information resource databases. Such cloud-based databases not only address the performance bottlenecks of traditional systems in handling large-scale heterogeneous data but also significantly improve security, scalability, and service capability. Therefore, this study adopts cloud computing as its technological core to explore methods and optimization strategies for constructing integrated databases, aiming to propose a solution that is both efficient and sustainable. By reviewing existing research, analyzing system requirements, designing architectures, and validating methods through case studies, the study seeks to provide both theoretical and practical contributions to the construction of integrated information resource databases.

2. Requirement Analysis of Integrated Information Resources Based on Cloud Computing

2.1 Current Status and Problems of Information Resources

Information resources have become an essential strategic asset for social development and scientific innovation, playing a core role in government governance, scientific research, business operations, and educational services. However, several critical issues remain unresolved. First, information resources are highly dispersed, with different institutions and departments having developed independent “information islands” over time, which hinders unified management and efficient sharing. Second, resource formats and types are highly heterogeneous, encompassing structured database records, semi-structured text, images, and audiovisual data, as well as unstructured archival documents. This diversity increases the difficulty of integration and utilization[1]. Third, the rapid update cycles and inconsistent quality of resources—often redundant, missing, or inconsistent—reduce both efficiency and credibility. Furthermore, data security and privacy protection are growing concerns, as traditional localized databases struggle to handle large-scale access, prevent data leakage, and support cross-domain sharing. Overall, while information resources are extensive and content-rich, they face major challenges in distribution, heterogeneity, quality control, and security, underscoring the need for cloud-based integrated database solutions[2].

2.2 Functional and Non-functional Requirements

Based on the identified problems, the construction of a cloud-based integrated information resource database must satisfy both functional and non-functional requirements. Functionally, the system must support efficient data acquisition and integration across structured, semi-structured, and unstructured formats through standardized interfaces. It should also provide robust data storage and management capabilities, enabling multi-dimensional classification and organization for efficient retrieval. Intelligent retrieval functions, including keyword-based search, semantic matching, and multi-dimensional indexing, must be supported to enable precise and rapid data access. Additionally, the system should offer resource-sharing and interface services to enable interoperability across systems and organizations[3].Non-functional requirements emphasize system performance and reliability. Security is central, requiring mechanisms such as access control, encryption, and privacy protection during data transmission and storage. The system should also be highly scalable and elastic, capable of dynamically expanding computing and storage resources in response to growth in data volume or user demand. High availability and reliability are equally crucial, necessitating fault tolerance and backup mechanisms to minimize service disruption. Lastly, performance optimization must ensure fast response times under high concurrency, enabling real-time resource access and data services. In short, fulfilling both functional and non-functional requirements is the prerequisite for building an efficient and reliable integrated information resource database[4].

3. Cloud Computing Architecture and Database Design

3.1 Cloud Computing Platform and System Architecture

The construction of an integrated information resource database based on cloud computing requires fully leveraging the distributed computing, elasticity, and high availability of cloud platforms. Overall, the system can be divided into three layers: the infrastructure layer, the data management layer, and the application service layer. The infrastructure layer provides computing resources, storage space, and network environments, serving as the foundation of the system[5]. The data management layer handles data collection, cleaning, normalization, and storage, while building unified data models and metadata management mechanisms to support efficient retrieval and scheduling. The application service layer provides users with retrieval, sharing, and interface services, enabling efficient utilization of integrated resources in research, education, government, and enterprise contexts.

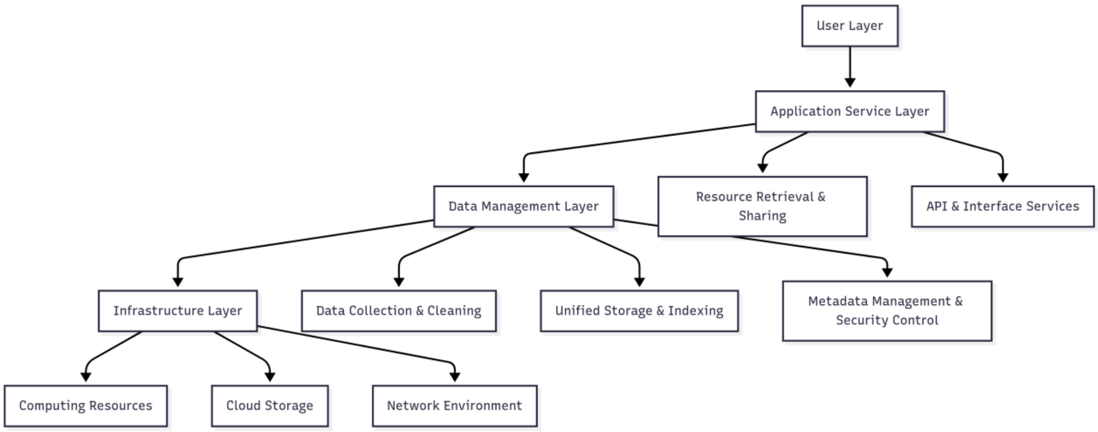

Figure 1: Cloud-based Architecture of the Integrated Information Resource Database

As shown in Figure 1, the system leverages cloud computing as its foundation. Through the unified management of the data layer, multi-source heterogeneous data are centralized and standardized. The application service layer then provides users with unified access and sharing capabilities. This layered architecture not only simplifies resource management but also significantly improves system scalability and flexibility[6].

3.2 Data Model and Security Control Design

In constructing a cloud-based integrated information resource database, the data model and security control mechanisms are key components. A well-designed data model directly impacts storage and retrieval efficiency as well as the effectiveness of integration and sharing. This study adopts a hybrid data model that combines relational and document-oriented databases: relational databases store structured information such as user records, metadata, and access permissions, while document-oriented databases manage semi-structured and unstructured resources such as text, images, and multimedia files. This hybrid approach ensures consistency while enhancing the system's capacity to handle diverse data types. In terms of security control, the system ensures data integrity, confidentiality, and controllability. Identity authentication and role-based access control (RBAC) restrict users to authorized operations, while encryption secures sensitive data during transmission and storage. Audit logs record user activities for traceability and compliance monitoring. Furthermore, fault tolerance and backup mechanisms ensure high availability and enable rapid recovery in case of system failures[7].

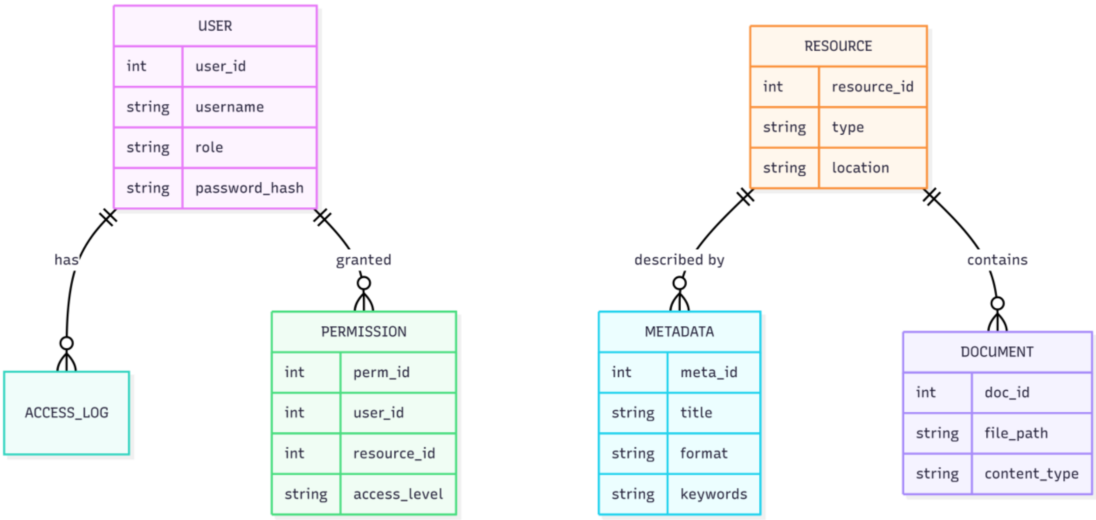

Figure 2: Data Model and Security Control Design for the Cloud-based Database

As illustrated in Figure 2, the data layer integrates relational and document databases to manage different resource types, while security mechanisms—including access control, encryption, and auditing—span the entire data lifecycle to ensure comprehensive protection.

4. Methods for Constructing the Integrated Information Resource Database

4.1 Data Collection, Cleaning, and Storage Methods

In the construction of an integrated information resource database, data collection, cleaning, and storage constitute the foundation for ensuring system stability and reliability. First, in terms of data collection, the system must be capable of accessing heterogeneous data from multiple sources, including structured data (such as relational databases), semi-structured data (such as XML and JSON), and unstructured data (such as text, images, and audiovisual materials). By establishing multi-channel collection interfaces and unified access standards, data from different sources can be smoothly integrated into the system. Second, during the data cleaning process, collected data must undergo deduplication, missing value handling, format conversion, and consistency checks. These steps eliminate redundancy, noise, and errors, thereby improving data integrity and accuracy[8]. In addition, metadata tagging and standardization rules are applied to support subsequent retrieval and management. Finally, for data storage, the system adopts a distributed cloud storage and hybrid database model, whereby highly structured data are stored in relational databases, while semi-structured and unstructured data are stored in document-oriented or object storage systems. This enables effective classification management and efficient data access[9].

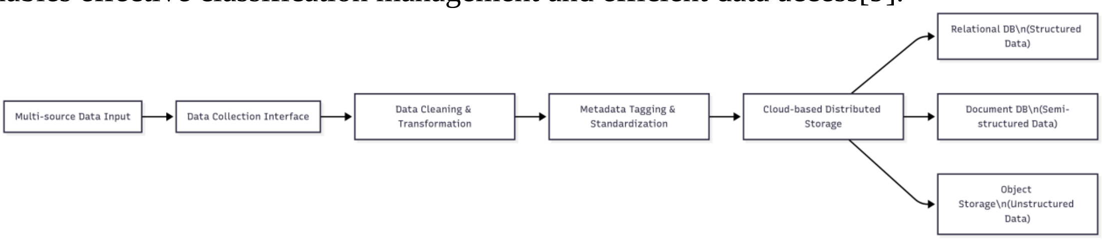

Figure 3 Data Collection, Cleaning, and Storage Process in the Cloud-based Database

As shown in Figure 3, the overall process of data collection, cleaning, and storage involves aggregating data through multi-source collection interfaces, followed by cleaning and standardization, and then unified storage in a distributed cloud-based system. This provides the foundation for subsequent data retrieval and resource sharing[10].

4.2 Data Retrieval Mechanism and Interface Design

In the construction of an integrated information resource database, the design of the data retrieval mechanism and interfaces directly determines the system's usability and service value, making them critical for user experience and cross-system collaboration. First, in terms of retrieval, the system should not only provide traditional keyword-based search but also incorporate semantic retrieval and intelligent recommendation techniques. By leveraging natural language processing (NLP), knowledge graphs, and semantic tagging, the system can better understand user query intentions, enabling cross-language and cross-domain semantic associations that significantly improve retrieval accuracy and result relevance. Moreover, the system should establish a multi-level indexing framework: for structured data, B+ tree and hash indexes ensure efficient location, while for semi-structured and unstructured data, inverted indexes and full-text search mechanisms are employed. Combined with distributed indexing frameworks, the system can support high-concurrency queries. In addition, the system should support multi-dimensional retrieval, including queries based on fields, time, topics, and semantic tags, to meet diverse needs in research, education, and government contexts. For interface design, openness and standardization are key considerations. The system should provide multiple interface forms, such as RESTful APIs and GraphQL, to enable flexible integration with different application systems. For scenarios requiring real-time interaction, WebSocket interfaces can be introduced to support bidirectional communication, meeting the needs of high-frequency access and instant feedback. Furthermore, the interface layer must be extensible, supporting secondary development and modular expansion so that customized services can be delivered across different application domains. To ensure security and compliance, interface calls must embed authentication and authorization mechanisms, such as OAuth2.0-based identity verification and role-based access control (RBAC), thereby preventing data leakage and unauthorized access while enabling resource sharing. At the same time, logging and auditing functions must be included in interface calls to guarantee traceability and regulatory compliance in cross-system interactions. In summary, a robust data retrieval mechanism and a well-designed interface architecture are dual guarantees for the efficient operation of an integrated information resource database. The former enhances user access efficiency and information quality through intelligent, multi-dimensional search methods, while the latter fosters cross-platform and cross-domain data exchange through standardized and extensible interfaces. Together, they ensure not only the efficient integration and utilization of information resources but also lay the foundation for future intelligent applications and big data analytics.

5. System Implementation and Case Analysis

5.1 Functional Modules and Implementation Process

In constructing a cloud-based integrated information resource database, the rational division and implementation of system functional modules are key to ensuring stable operation and efficient service delivery. Overall, the system comprises five functional modules: the data collection and processing module, the data storage and management module, the data retrieval and analysis module, the user service and interface module, and the system security and operation module. While relatively independent, these modules operate collaboratively through a unified cloud platform, forming a complete business cycle. The data collection and processing module is responsible for accessing diverse information resources from different sources. With the design of multi-source interfaces, the system can batch-collect structured data (e.g., database tables), semi-structured data (e.g., XML and JSON), and unstructured data (e.g., text, images, and audiovisual files). The collected data then undergo cleaning and standardization to remove redundancy and errors. Metadata tagging ensures unified descriptions, thereby enhancing data consistency and integrity. The data storage and management module is built on distributed cloud storage and integrates both relational and document-oriented databases. Structured data are stored in relational databases to support efficient transaction processing and indexing, while semi-structured and unstructured data are stored in document databases and object storage systems to enhance compatibility and scalability. Metadata management is also introduced at this stage to ensure accurate description and rapid location of resources. The data retrieval and analysis module enables efficient resource utilization through a multi-level indexing framework and semantic search mechanisms. It supports keyword search, semantic retrieval, and composite queries, and integrates data mining and analytical techniques to provide personalized recommendations and intelligent services. In high-concurrency environments, distributed indexing and caching mechanisms significantly improve retrieval speed and system responsiveness. The user service and interface module provides user-friendly interfaces and open services for external systems. Users can access the system via web or mobile platforms for resource retrieval and sharing, while external applications can integrate through RESTful APIs or GraphQL interfaces. Authentication and access control are embedded to ensure secure and compliant data sharing. The system security and operation module safeguards overall functionality. It includes user authentication, access control, data encryption, log auditing, and backup/recovery to ensure confidentiality and integrity. In addition, an operational monitoring platform provides real-time performance monitoring and optimization, ensuring system stability and high availability under large-scale access scenarios. In conclusion, the implementation process reflects the principles of layered design and modular construction. Supported by cloud computing, the modules form a complete chain of data collection, storage, retrieval, service, and security operations, meeting the demands of resource integration and sharing while laying a solid foundation for case analysis and practical applications.

5.2 Application Case and Performance Analysis

To validate the feasibility and effectiveness of the proposed cloud-based integrated database, a regional scientific research information platform was selected as a case study. The platform's original resources were scattered across multiple departments and institutions, including structured experimental data and statistical records as well as unstructured resources such as research papers, images, videos, and conference materials. Due to the absence of a unified management and retrieval mechanism, users had to perform repeated searches across multiple systems, resulting in inefficiency, redundancy, and data gaps. To address these challenges, the platform adopted the proposed construction framework to achieve unified management and shared utilization of multi-source information resources. During implementation, the data collection and cleaning module enabled unified access and standardization of distributed research data, effectively eliminating redundancy and inconsistencies. The distributed cloud storage architecture centralized heterogeneous data management, ensuring efficient transaction operations for structured data and enhanced storage/retrieval efficiency for unstructured data. Users could then perform queries through a unified retrieval portal, with support for keyword, semantic, and multi-dimensional searches that substantially improved accuracy and speed. Moreover, the interface module facilitated integration with external research management systems, enabling cross-platform data sharing and secondary development, thereby enhancing the system's openness and practicality. In terms of performance, a comparison of pre- and post-implementation metrics showed that the average response time for information retrieval decreased by approximately 45%, data redundancy dropped by more than 30%, and concurrent access capacity nearly doubled. User satisfaction surveys revealed that over 85% of researchers believed the new system significantly improved the accessibility and completeness of resources. Regarding security, the introduction of multi-level access control and log auditing mechanisms effectively reduced the risk of unauthorized access, ensuring the safety and compliance of sensitive research data. Overall, this case study confirms the practical value of the proposed framework in enhancing resource integration, retrieval efficiency, and system security, thereby providing strong support for the application of integrated information resources in research, education, and government sectors.

6. Security and Performance Optimization Research

6.1 Data Security and Privacy Protection Mechanisms

In a cloud-based integrated information resource database, data security and privacy protection are among the core requirements of system design. Since the system processes large-scale information resources from diverse fields such as scientific research, education, government, and enterprise, which are often sensitive and heterogeneous, it is essential to establish multi-level and multi-dimensional security and privacy protection mechanisms. First, in terms of access control, the system adopts a combination of Role-Based Access Control (RBAC) and fine-grained permission management to ensure strict identity authentication and hierarchical authorization. This approach effectively prevents unauthorized access and illegal operations, ensuring that users can only access data within their designated scope of authority. Second, during data transmission and storage, the system employs multiple encryption mechanisms. For transmission, SSL/TLS protocols are used to secure data in transit, preventing man-in-the-middle attacks and tampering. For storage, high-strength encryption algorithms such as AES are applied to safeguard sensitive data, ensuring that even if physical media are lost or compromised, the information remains difficult to retrieve illegally. In addition, data masking and anonymization techniques are incorporated for personal privacy and sensitive research data, allowing them to be shared and analyzed while minimizing the risk of privacy breaches. Third, the system achieves full traceability of user actions through security auditing and logging mechanisms. All access and operational activities are recorded and monitored by the auditing system, enabling timely alerts and responses in the event of abnormal operations or potential attacks. Furthermore, regular security testing, vulnerability scanning, and patch updates are conducted to ensure that the platform remains in a secure and controllable state. Finally, in terms of compliance, the system is designed in strict accordance with national and industry regulations, such as the Cybersecurity Law, Data Security Law, and Personal Information Protection Law, while also aligning with international privacy frameworks such as the GDPR. By establishing standardized data security and privacy management systems that combine technical and institutional safeguards, the system enables large-scale integration and sharing of information resources while effectively preventing data leakage, tampering, and misuse. Overall, these mechanisms not only provide strong support for the secure use of information resources but also lay a solid foundation for the long-term stability and sustainability of the database.

6.2 System Performance Optimization and Scalability Research

In a cloud-based integrated information resource database, performance and scalability are critical indicators of whether the system can operate stably over the long term while supporting large-scale user access. Since such systems must handle multi-source heterogeneous data while meeting the requirements of high-concurrency retrieval, rapid response, and secure reliability, performance optimization and scalability must be considered thoroughly in both system architecture and operational management. For performance optimization, the system improves response speed through multi-level caching, load balancing, and index optimization. Caching mechanisms store frequently accessed data in memory or at edge nodes to reduce redundant query operations. Load balancing strategies distribute requests across multiple computing nodes to prevent single-point bottlenecks and enhance overall processing capability. Regarding indexing, the system employs a combination of B+ tree and hash indexing for structured data, and inverted indexes and full-text search for unstructured data retrieval, ensuring high query efficiency in large-scale data environments. In high-concurrency scenarios, distributed parallel processing frameworks leverage the computational power of the cloud platform to maintain stable performance even during peak loads. For scalability, the system relies on the elasticity of the cloud platform to dynamically expand computing and storage resources. As data volume grows or user demand surges, the system can automatically scale out by adding computing instances and storage nodes, thus avoiding performance degradation or service interruptions. Moreover, the database architecture adopts modular and microservice-based design principles, allowing individual functional modules to be independently scaled or upgraded. This not only reduces system coupling but also enhances flexibility for future feature expansion and cross-platform deployment. In addition, multi-replica storage and distributed backup mechanisms ensure high availability and enable rapid recovery in the event of node failures, thereby further strengthening overall scalability and fault tolerance. In summary, the core of performance optimization and scalability research lies in fully leveraging the distributed architecture and elastic resource scheduling capabilities of cloud computing. Performance is improved through caching, load balancing, indexing, and parallel processing, while scalability is achieved through elastic scaling, modular design, and distributed fault tolerance. Together, these measures ensure that the database operates efficiently in current applications and provides reliable support for deployment in larger and more complex scenarios in the future.

7. Conclusion

This study explores the construction of an integrated information resource database based on cloud computing, aiming to address challenges of dispersed distribution, heterogeneous formats, uneven quality, and security risks. A cloud-based three-layer framework was proposed, covering infrastructure, data management, and application services, and incorporating processes such as data collection, cleaning, storage, retrieval, and interface design. A hybrid data model and security mechanisms were also introduced to manage structured and unstructured data while ensuring compliance and privacy protection. System implementation and case analysis demonstrated that the proposed framework significantly improved data integration, retrieval efficiency, and system stability, while also enhancing scalability and reliability in high-concurrency scenarios. These results confirm the practical value of combining cloud computing with information resource integration. Overall, this research contributes both theoretically and practically by providing a feasible solution for information management in research, education, and government sectors. Future work will focus on cross-platform data sharing, semantic understanding of heterogeneous data, and intelligent retrieval, leveraging technologies such as artificial intelligence, big data analytics, and knowledge graphs to achieve greater intelligence, adaptability, and sustainability.

References

[1] Oke, Ayodeji Emmanuel, et al. "Exploring the benefits of cloud computing for sustainable construction in Nigeria." Journal of Engineering, Design and Technology 21.4 (2023): 973-990.
[2] Jyoti, Choudrie, and Zamani Efpraxia. "Understanding and exploring the value co-creation of cloud computing

innovation using resource based value theory: An interpretive case study." Journal of Business Research 164 (2023): 113970.
[3] Waqar, Ahsan, et al. "Success of implementing cloud computing for smart development in small construction projects." Applied Sciences 13.9 (2023): 5713.
[4] Amini, Mahyar, and Negar Jahanbakhsh Javid. "A multi-perspective framework established on diffusion of innovation (DOI) theory and technology, organization and environment (TOE) framework toward supply chain management system based on cloud computing technology for small and medium enterprises." Organization and Environment (TOE) Framework Toward Supply Chain Management System Based on Cloud Computing Technology for Small and Medium Enterprises (January 2023). International Journal of Information Technology and Innovation Adoption 11 (2023): 1217-1234.
[5] Wu, Jiang, et al. "Case study of next-generation artificial intelligence in medical image diagnosis based on cloud computing." Journal of Theory and Practice of Engineering Science 4.02 (2024): 66-73.
[6] Al-Jumaili, Ahmed Hadi Ali, et al. "Big data analytics using cloud computing based frameworks for power management systems: Status, constraints, and future recommendations." Sensors 23.6 (2023): 2952.
[7] Lin, Haifeng, et al. "Internet of things intrusion detection model and algorithm based on cloud computing and multi-feature extraction extreme learning machine." Digital Communications and Networks 9.1 (2023): 111-124.
[8] Zhao, Min, and Kamran Rabiei. "Feasibility of implementing the human resource payroll management system based on cloud computing." Kybernetes 52.4 (2023): 1245-1268.
[9] Grace, Mia Ella. "Cloud Computing for Decision Making within the Clinical Sector." Journal of Biomedical and Sustainable Healthcare Applications 3.2 (2023): 087-094.
[10] Wang, Yufu, et al. "Cloud computing for large-scale resource computation and storage in machine learning." Journal of Theory and Practice of Engineering Science 4.03 (2024): 163-171.