The Application of Big Data Technology and Data Security Guarantee in Archive Management

Bing Sun

Mulan County Labor Security Rights Protection Center Harbin, Heilongjiang 150000

Abstract: ***With the rapid development of information technology, the application of big data technology in the field of archive management is becoming increasingly widespread. Big data technology has brought unprecedented opportunities for archive management, which can optimize archive management processes, improve service quality and efficiency. However, while enjoying the convenience brought by big data technology, data security issues have become increasingly prominent. This article explores in depth the specific application of big data technology in archive management, and proposes data security protection strategies for the security risks faced, aiming to provide reference for the efficient and secure development of archive management work.***

Keywords: Archive management; Big data technology; Data security.

1. INTRODUCTION

Archives, as an important carrier of information resources, record various aspects of social development and have extremely high historical, cultural, and social value. The traditional archive management model has certain limitations in data processing capabilities, information retrieval efficiency, and other aspects, making it difficult to meet the needs of today's society for fast, accurate, and comprehensive utilization of archive information. The emergence of big data technology has brought new opportunities for archive management work. It can efficiently collect, store, analyze, and utilize massive, diverse, and complex archive data, greatly improving the efficiency of archive management. However, at the same time, the security of archival data in the big data environment faces many challenges, such as risks of data leakage, tampering, and loss, which seriously threaten the integrity, confidentiality, and availability of archival information. Therefore, it is of great practical significance to conduct in-depth research on the application of big data technology in archive management and how to ensure data security. Wu (2024) investigated large-scale parallel computing solutions for genetic disease analysis [1], while Chen (2025) explored scalable cloud architectures for autonomous driving data lakes and real-time decision making [2]. Complementing these technical studies, Wang (2024) examined the legal dimensions of enterprise naming rights and their restrictions [3], highlighting the need for balanced technological and legal frameworks. Gong et al. (2024) further contributed to enterprise applications by developing an ensemble machine learning approach for optimizing risk decision support systems [4]. Sustainable urban development has benefited significantly from AI applications. Zhou et al. (2024) improved automated garbage recognition using ResNet-50 and weakly supervised CNN models [5], demonstrating environmental applications of computer vision. In healthcare technology, Lin et al. (2025) showed how intelligent exercise monitoring enhances executive function in children with ADHD [6], while Peng et al. (2025) investigated the cognitive and sleep impacts of aerobic exercise intensity [7]. Clinical AI applications were advanced by Shen et al. (2025), who developed an LSTM-based system for anesthetic dose management in cancer surgery [8]. Cybersecurity research has seen important developments, with Xu et al. (2025) analyzing adversarial machine learning attacks and defenses [9]. Financial technology innovations were presented by Chew et al. (2025) through their AI-optimized accounting data integration and risk assessment model for e-commerce platforms [10]. Finally, in network security, Liu et al. (2025) proposed a privacy-preserving hybrid ensemble model for anomaly detection that balances security and data protection [11].

2. APPLICATION OF BIG DATA TECHNOLOGY IN ARCHIVE MANAGEMENT

2.1 Archive Data Collection and Integration

2.1.1 Multi source data collection

In the era of big data, archive data comes from a wide range of sources, including digitized electronic data of paper archives, real-time data generated by business systems, and archive related data on social media and other online platforms. Big data technology can efficiently collect multi-source archival data through various data collection tools and techniques, such as web crawlers, data interfaces, etc. For example, for the archive management of government departments, real-time access to archive data generated during administrative approval, law enforcement supervision, and other business processes can be achieved through data interfaces with various business systems, ensuring the timeliness and completeness of the data. Meanwhile, by utilizing web crawling technology, external information related to the department's business can be collected from relevant industry websites, forums, and other platforms, enriching archive data resources.

2.1.2 Data Integration and Cleaning

The collected archive data often has problems such as inconsistent format, duplicate data, and errors. The data integration and cleaning techniques in big data technology can standardize the processing of this data. By establishing a unified data standard and format, and using data conversion tools to convert data in different formats into a unified format, it facilitates subsequent storage and analysis. By utilizing data cleaning algorithms, errors, duplicates, and missing values in data can be identified and corrected, thereby improving data quality. For example, in enterprise archive management, archive data from different departments such as finance, personnel, and production are integrated and cleaned to eliminate data silos and form a complete and accurate enterprise archive data resource library, providing reliable data support for decision-making analysis of the enterprise.

2.2 Archive Data Storage and Management

2.2.1 Distributed Storage Technology

The continuous growth of archive data poses a serious challenge to storage capacity. The distributed storage technology of big data, such as Hadoop Distributed File System (HDFS), can store massive archive data in multiple storage nodes, and improve data storage reliability and read-write performance through redundant storage and data block replication mechanisms. Compared with traditional centralized storage, distributed storage has the advantages of strong scalability and low cost. For example, large archives adopt distributed storage technology, which can easily cope with the large amount of new archive data storage needs every year, and can still ensure normal access to data even when some storage nodes fail, without affecting the progress of archive management work.

2.2.2 Data indexing and retrieval optimization

Big data technology can significantly improve the retrieval speed of archival data by establishing an efficient data indexing mechanism. By utilizing techniques such as inverted indexing and full-text indexing, the content, keywords, and metadata of archival data can be indexed and constructed, enabling users to quickly locate relevant data when querying archival information. For example, in the archive management of digital libraries, after users input keywords, with the help of big data indexing technology, the system can instantly retrieve relevant books, literature, and other information from massive book archive data, greatly improving the efficiency of users in obtaining information.

2.3 Analysis and Utilization of Archival Data

2.3.1 Data Analysis and Mining Value

Big data analysis technology can deeply mine archival data, discovering hidden patterns, trends, and correlations behind the data. By using methods such as statistical analysis, clustering analysis, and association rule mining on archival data, strong support can be provided for archival management decisions. For example, the archives can analyze historical archive utilization data to understand user demand trends for different types of archives, thereby optimizing the collection and organization strategies of archive resources and better meeting user needs. By analyzing customer profile data, enterprises can explore customer consumption habits, preferences, and other information to provide reference for marketing and product development.

2.3.2 Provide personalized services

Based on the results of big data analysis, the archive management department can provide personalized services to users. By analyzing user behavior data, query history, etc., we can understand users' interests and demand characteristics, and accurately push relevant profile information to users. For example, museums utilize big data technology to recommend personalized exhibition guidance services to visitors based on their browsing history and stay time, enhancing their viewing experience. The government archives department provides customized policy interpretation, archive inquiry and other services to different enterprises and individuals based on their needs, in order to improve the pertinence and satisfaction of the services.

2.4 Optimization of archive management process

2.4.1 Automated Process Processing

The combination of big data technology with artificial intelligence, the Internet of Things, and other technologies can achieve automation of archive management processes. For example, in the process of borrowing and returning archives, IoT technology is used to automatically identify and locate archives by pasting RFID tags on them. When users borrow or return files, the system automatically records relevant information and updates the file status, without the need for manual registration, greatly improving work efficiency. In the cataloging and classification of archives, the natural language processing technology of artificial intelligence can automatically analyze and classify the contents of archives, reduce manual workload, and improve the accuracy of cataloging and classification.

2.4.2 Real time monitoring and early warning

Through big data technology, real-time monitoring of various aspects of archive management can be carried out, potential problems and risks can be identified in a timely manner, and warnings can be issued. For example, real-time collection and analysis of temperature, humidity, air quality, and other data in the archive storage environment. When the environmental parameters exceed the appropriate range, the system automatically issues warning messages to remind management personnel to take corresponding measures to ensure the safe storage of archives. Real time monitoring of archive utilization, timely warning when abnormal high-frequency query or download behavior is detected, to prevent archive data from being illegally obtained and abused.

3. DATA SECURITY PROTECTION STRATEGIES FOR THE APPLICATION OF BIG DATA TECHNOLOGY IN ARCHIVE MANAGEMENT

3.1 Technical Support Measures

3.1.1 Data Encryption Technology

Using data encryption technology to encrypt archive data, ensuring the confidentiality of data during transmission and storage. During data transmission, encryption protocols such as SSL/TLS are used to encrypt the data and prevent it from being stolen or tampered with. In data storage, sensitive data fields are encrypted, such as the user's ID number and password. Common encryption algorithms include AES, RSA, etc. The archive management department can choose the appropriate encryption algorithm and key management strategy according to actual needs. For example, in an electronic archive management system, important archive files are stored with AES encryption, and only authorized users can decrypt and view the file content using the correct key.

3.1.2 Access Control Technology

Establish strict access control mechanisms to finely manage user access permissions to archive data. Through identity authentication, authorization management, and other technologies, ensure that only legitimate users can access archive data within their authorized scope. Adopting multiple factor authentication methods, such as password+verification code, fingerprint recognition+password, etc., to improve the security of user identity authentication. In terms of authorization management, users are assigned corresponding access permissions based on their roles and responsibilities, such as read-only permission, read-write permission, etc. For example, in the enterprise archive management system, ordinary employees can only view archive data related to their own work, and only have read-only permission; And the archive administrator has read and write permissions for all archive data.

3.1.3 Data backup and recovery technology

Develop a comprehensive data backup strategy, regularly backup archive data, and store backup data in a secure location. Adopting a combination of full backup and incremental backup to improve backup efficiency and data recovery accuracy. At the same time, establish a data recovery drill mechanism and conduct regular data recovery tests to ensure timely and accurate data recovery in the event of data loss or damage. For example, the archive conducts a full backup once a week and an incremental backup once a day, storing the backup data in a remote data center. Regularly organize data recovery drills, simulate data loss scenarios, and verify the effectiveness and efficiency of data recovery.

3.2 Management Guarantee Measures

3.2.1 Improve safety management system

The archives management department should establish and improve a data security management system, clarify data security responsibilities and obligations. Develop detailed data access permission management regulations, data usage standards, data backup and recovery systems, emergency response plans, etc. For example, it is required to go through a strict approval process when handling sensitive archival data; Detailed records of data usage for traceability and auditing purposes. Regularly evaluate and update the safety management system to ensure its effectiveness and adaptability.

3.2.2 Strengthen personnel training and management

Strengthen data security training for archive management personnel, enhance their security awareness and business skills. The training content includes data security laws and regulations, security protection knowledge, operational norms, etc. Regularly organize safety knowledge assessments, link the assessment results with employee performance, and motivate employees to actively learn and comply with data security regulations. At the same time, strengthen personnel management, conduct background checks on newly hired employees, and sign confidentiality agreements; Timely revoke the access rights of departing employees to prevent data leakage.

3.2.3 Strengthen third-party cooperation management

Before cooperating with third-party organizations, strict qualification review should be conducted, including corporate reputation, safety management capabilities, technical strength, and other aspects. Clearly define data security responsibilities and obligations in the cooperation agreement, and require third-party organizations to take necessary data security protection measures, such as data encryption, access control, etc. Regularly supervise and evaluate the data security status of third-party organizations, and promptly request rectification if problems are found. For example, when collaborating with third-party data storage service providers, they are required to provide data security certification and conduct regular security checks on their storage devices and network environments.

3.3 Legal protection measures

3.3.1 Compliance with relevant laws and regulations

The archives management department should strictly comply with national and local laws and regulations on data security, such as the Cybersecurity Law of the People's Republic of China and the Data Security Law. Ensure that the collection, storage, use, and transmission of archival data comply with legal and regulatory requirements, and avoid data security issues caused by illegal or irregular behavior. For example, when collecting personal profile data, it is necessary to follow the principles of legality, legitimacy, and necessity, obtain clear authorization from users, and properly protect the data.

3.3.2 Using legal means to safeguard rights and interests

When archive data is subjected to security breaches, archive management departments should actively use legal means to safeguard their own rights and interests. Report to the public security organs in a timely manner, cooperate with law enforcement departments to investigate and collect evidence, and hold the infringer legally responsible. At the same time, the infringer shall be required to compensate for the economic losses caused by data

leakage, tampering, etc. through legal means. For example, when a hacker attacks the archive management system and steals a large amount of archive data, they should promptly report to the public security organs, pursue the criminal responsibility of the hacker according to law, and demand compensation for the economic losses caused to the unit due to data leakage.

4. CONCLUSION

The application of big data technology in archive management has brought revolutionary changes to archive management work, significantly improving the efficiency and service level of archive management. However, data security issues have always been an important aspect that cannot be ignored in the application process of big data technology. By adopting a series of measures such as technical support, management support, and legal support, data security risks can be effectively reduced, and the integrity, confidentiality, and availability of archival data can be guaranteed. In the future archive management work, with the continuous development and application of big data technology, archive management departments should continue to pay attention to data security issues, continuously improve the data security guarantee system, ensure that archive management work is efficiently carried out in a safe and stable environment, and provide strong archive information support for social development.

REFERENCES

[1] Wu, W. (2024). Research on cloud infrastructure for large-scale parallel computing in genetic disease.
[2] Chen, J. (2025). Leveraging Scalable Cloud Infrastructure for Autonomous Driving Data Lakes and Real-Time Decision Making.
[3] Wang, H. (2024). The Restriction and Balance of Prior Rights on the Right of Enterprise Name.
[4] Gong, C., Lin, Y., Cao, J., & Wang, J. (2024, October). Research on Enterprise Risk Decision Support System Optimization based on Ensemble Machine Learning. In Proceeding of the 2024 5th International Conference on Computer Science and Management Technology (pp. 1003-1007).
[5] Zhou, Y., Wang, Z., Zheng, S., Zhou, L., Dai, L., Luo, H., ... & Sui, M. (2024). Optimization of automated garbage recognition model based on resnet-50 and weakly supervised cnn for sustainable urban development. Alexandria Engineering Journal, 108, 415-427.
[6] Lin, L., Li, N., & Zhao, S. (2025). The effect of intelligent monitoring of physical exercise on executive function in children with ADHD. Alexandria Engineering Journal, 122, 355-363.
[7] Peng, Y., Zhang, G., & Pang, H. (2025). Impact of Short-Duration Aerobic Exercise Intensity on Executive Function and Sleep. arXiv preprint arXiv:2503.09077.
[8] Shen, Z., Wang, Y., Hu, K., Wang, Z., & Lin, S. (2025). Exploration of Clinical Application of AI System Incorporating LSTM Algorithm for Management of Anesthetic Dose in Cancer Surgery. Journal of Theory and Practice in Clinical Sciences, 2, 17-28.
[9] Xu, J., Wang, Y., Chen, H., & Shen, Z. (2025). Adversarial Machine Learning in Cybersecurity: Attacks and Defenses. International Journal of Management Science Research, 8(2), 26-33.
[10] Chew, J., Shen, Z., Hu, K., Wang, Y., & Wang, Z. (2025). Artificial Intelligence Optimizes the Accounting Data Integration and Financial Risk Assessment Model of the E-commerce Platform. International Journal of Management Science Research, 8(2), 7-17.
[11] Liu, S., Zhao, Z., He, W., Wang, J., Peng, J., & Ma, H. (2025). Privacy-Preserving Hybrid Ensemble Model for Network Anomaly Detection: Balancing Security and Data Protection. arXiv preprint arXiv:2502.09001.