Computer Data Storage and Management Platform Based on Big Data

Cheng Luo1,*

1Jiangxi Vocational Technical College of Industry & Trade, Nanchang, 330038

*Corresponding author e-mail: 457960227@jxvtcit.edu.cn

Abstract. Due to the continuous development of information technology, data has increasingly become the core of the daily operation of enterprises and institutions, the main basis for decision-making development. At the same time, due to the development of network, the storage and management of computer data has attracted more and more attention. Aiming at the common problems of computer data storage and management in practical work, this paper analyzes the object and content of data management, investigates the situation of computer data storage and management in China in recent two years, and interviews and tests the data of programming in this design platform. At the same time, in view of the related problems, the research results are applied to practice. On the basis of big data, the storage and management platform is designed. The research and design adopts a special B+ tree node linear structure of CIRC tree, and the linear node structure is changed into a ring structure, which greatly reduces the number of data persistence instructions and the performance overhead. The results show that compared with the most advanced B+ tree design for nonvolatile memory, crab tree has 3.1 times and 2.5 times performance improvement in reading and writing, respectively. Compared with the previous NV tree designed for nonvolatile memory, it has a performance improvement of 1.5 times, and a performance improvement of 8.4 times compared with the latest fast-fair. In the later stage, the expansion of the platform functions is conducive to the analysis and construction of data related storage and management functions, and further improve the ability of data management.

Keywords: Big Data Drive, Computer Data, Data Storage and Management Platform, Design and Research

1. Introduction
Due to the rapid development of the Internet, the current computer information and data storage presents a continuous growth trend [1]. Therefore, the development of computer data storage technology with strong storage performance, good security performance and high efficiency has become the focus of current computer technology innovation [2-3]. Generally, the data generated in the process of data stream processing will be recorded in a certain form on the internal or external storage media of the computer, and the naming of data storage can have the value of information
feature composition. The management of data storage and change in computer system can reduce the risk of change, improve the security of system operation and improve the quality of data [4-5].

As the main data storage device, the data storage and management function of computer has received more and more attention. From computers to applications in all walks of life, data stored on computers has been facing many problems [6-7]. If the computer is deleted, formatted or restored by mistake, it will affect the data security of the computer in a simple way [8-9]. The continuous development and maturity of computers provide technical support for the storage and processing of big data, so that many users can achieve high-speed and effective data operation on different terminals [10].

In view of the common problems encountered in the actual work of data management, this paper analyzes and defines the object and content of data management, and applies the research results to practice to build a data management platform, so as to realize the centralized and transparent management of data management. From the generation, processing, storage to the application of the whole process of data, it establishes an efficient and flexible data management support tool, which can provide experience for the enterprise information system data management. In this paper, a special B+ tree node linear structure of CIRC tree is designed. The linear node structure is changed to ring structure, which greatly reduces the number of data persistence instructions and the performance overhead. In the later stage, the expansion of the platform functions is conducive to the analysis and construction of data related storage and management functions, and further improve the ability of data management.

2. Overview of computer data storage and management

2.1. Object and content of data management

Based on the leading position of big data related technology, this paper deeply studies the current situation of data management, with the relevant experience of data management at home and abroad, grasps the related data management objects in the process of data creation and processing, and stores and applies them to the current big data environment.

(1) Data management object refers to the technical object in each stage from data generation to application in information system. According to the current development of information system construction, the scope of data management should include major information systems, and data management objects include multiple application integration.

(2) The purpose of data storage and management.

Data storage and management is an information security strategy. Through the replication of the initial data, when the initial data is lost or destroyed, the initial data can be restored to the original through backup and recovery technology to ensure the normal operation of the system. All data closely related to users in computer system should be stored, not just user data in database. The purpose of data collection and management is to recover data. It can greatly reduce most of the risks of the system and protect the most important system resource data. After leaking the system, data recovery can be used to recover the data.

2.2. Key technology realization of computer data storage and management

The key technologies of data management platform are as follows:

(1) Layout algorithm based on complex visualization.

A graphical display can complete the visualization analysis and data application analysis, form an orderly analysis of each data level, determine the relationship between the independence of each level and the visual essence, and maintain the real operation results of many users. From a small level, this role is the main content of the data management platform. The platform is based on a special algorithm and adds the graphic layout of the visualization topology. Through the algorithm, users can automatically determine the order of graphics by clicking the mouse, which completely improves the
beauty of the information system, clearly displays the complex data sharing information, and speeds up the user's use speed to process data.

(2) Metadata will be collected automatically.

Metadata integration is the biggest part of housing management. Through the special research on the storage and metadata supporting methods of different pharmaceutical equipment skills, the adaptability and materials are proposed, and the automatic collection of public artificial metadata is realized. This not only supports a large number of computer devices, but also creates metadata collections for applications of large data platform related tools and companies from service bus and other products. The efficiency of data collection and automatic management is greatly reduced.

(3) Configurable workflow.

Configurable work order is the innovation breakpoint of data management platform. The platform has its own line components to define process, create, execute and control. Please note that the user can change the configuration of the worksheet according to the management method and the information system command, so as to achieve the visual interface that supports the user to draw the worksheet. Through the work example made by the user, the work can be expressed to the control level according to the configuration, which greatly improves the intelligence and operability of data management.

2.3. Computer data correlation processing

(1) Bresenham algorithm.

Bresenham algorithm is the most frequently used linear scan conversion algorithm in computer graphics. Let a straight line from the starting point \((X_1, Y_1)\) to the end point \((X_2, Y_2)\) be expressed as the equation: 

\[ y = mx + b, \quad b = y_1 - ms_1, \quad m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{d_y}{d_x}. \]

The calculation formula is as follows:

\[ y = m(x_i + 1) + b \quad \text{(1)} \]
\[ d_y = y - y_i \quad \text{(2)} \]
\[ d_x = y_i + 1 - y \quad \text{(3)} \]

(2) Rectifier linear unit (ReLU).

In cloud computing of management platform, the recommended activation function is rectilinear cell. The RELU function has the following form:

\[ \text{relu}(x) = \max\{0, x\} \quad \text{(4)} \]

3. Ideas and methods

3.1. Design ideas

Form a data management platform, according to the characteristics of each link in the data life cycle, orderly manage the whole process of the creation, processing, storage and application of large-scale data platform, data center and business system. Understand the functions of managing mixed data assets, managing the whole data process, and analyzing data across links, and provide technical support for big data management. In the business layer, users at all levels can manage and apply data from configuration management, operation monitoring, change management and information request. At the technical level, it follows the principles of precise design, progress and scalability, supports automatic collection of platform information, so as to meet the metadata collection requirements of main database, ETL tools and big data related products, and reduce the workload of maintenance manual. It provides a flexible workflow component, which can be configured to meet different
management needs. At the same time, the use of visual imaging technology to improve the user experience.

3.2. Platform architecture design
Data management platform for information system data from the collection, storage, processing and application of the whole process of centralized, transparent management.

The platform adopts standard J2EE architecture, and the front end of the platform adopts HTML technology to achieve a good human-computer interaction interface. In the field of computer data storage, a number of new B+ tree data structures are designed, which can greatly improve the performance of B+ tree structure on the new nonvolatile memory while maintaining data persistence and consistency.

4. Analysis of computer storage and management

4.1. Security analysis of data stored in computer
In order to analyze the security situation of computer data storage in China in the era of big data, we investigate the situation of computer data storage in China in recent two years, as shown in Table 1.

Table 1. The number of tampered data websites in China

| Time                  | January~March | April~June | July~September | October~December |
|-----------------------|---------------|------------|----------------|------------------|
| 2018                  | 170           | 150        | 192            | 186              |
| 2019                  | 207           | 160        | 188            | 209              |

Figure 1. The number of tampered data websites in China

According to Figure 1, there are a lot of tampered data websites in China in recent two years. The relatively advanced design of b+ tree for nonvolatile memory by crab tree has 3.1 times and 2.5 times performance improvement in reading and writing. Isle tree organizes the data of nodes into an ordered structure within a single cache block and disorder among multiple cache blocks. Every time the data is moved in a cache block, only one clflush instruction is needed to ensure the data persistence. By this design, the number of persistent instructions is reduced to optimize the write performance of b+ tree on nonvolatile memory. As an important data storage device, computer security is also paid more and more attention. From the computer to the popularization of all aspects, the storage data on the computer inevitably faces the security problem. In short, deleting, formatting or overwriting by system restore will threaten the data security of the computer. At the same time, there are more factors that will affect data security, which are generally common as follows:

(1) Hard disk damage: physical damage to the hard disk means data loss. The loss of equipment operation, storage medium failure, operation environment and human damage will cause the hard disk equipment to be affected.
(2) Human error: due to operation error, the user may delete or lose important information or change the code affecting the operation of the system, and the system is forced to stop operation due to non-compliance or improper operation.

(3) Hacker intrusion: the invasion here refers to the intruder who uses the remote network attack system.

(4) Virus: because computer viruses can cause computer systems, and often cause huge economic losses. Computer virus has a strong ability of replication and transmission, especially in the network environment, the transmission speed is very fast.

(5) Information theft: copying, deleting, or merely stealing a computer from a computer.

(6) Power failure: power system failure, short-term huge power will suddenly destroy the hard disk or storage device data.

(7) Magnetic interference: magnetic interference refers to the damage to computer data when important information contacts with magnetic materials.

4.2. Analysis of computer data storage and management platform

When we write a program, the data we want to access is random, which may be one or several bytes, or some bits in the data. Even if the width of the data is equal to the length of the data type, the starting address of the data is small, which must be an integral multiple of the length of the data type. In this way, there is a problem of data alignment when CPU accesses the data in memory. The storage size and alignment of common data types in memory are shown in Table 2.

| Kind  | Size | Alignment |
|-------|------|-----------|
| Char  | 1    | 1         |
| Short | 2    | 2         |
| Int   | 4    | 4         |
| Long  | 4    | 4         |
| Float | 4    | 4         |
| Double| 8    | 8         |

Figure 2. Data alignment in memory and processing in programming

As can be seen from Figure 2, compared with the previous NV tree designed for nonvolatile memory, it has a performance improvement of 1.5 times, and a performance improvement of 8.4 times compared with the latest fast-fair. In order to maintain the crash consistency of the B + tree designed for the traditional main memory hard disk architecture transplanted to the new nonvolatile storage device, the persistence instruction provided by the processor manufacturer should be used to ensure that the modified data in the processor cache block is persisted to the nonvolatile main memory. This process will generate a lot of performance overhead. How to reduce this kind of overhead is the focus of the current design of storage system for nonvolatile memory. To solve this problem, this paper studies and designs a special B+ tree node linear structure of CIRC tree, which changes the linear node
structure into a ring structure, greatly reducing the number of data persistence instructions and the performance overhead. The continuous development and maturity of computers provide technical support for the storage and management of big data, so that the vast majority of users can complete high-speed and efficient data operation on a variety of terminals. But there are also the following data security problems, the most common are theft, loss, redundancy and so on. Such data security problems often bring huge profit loss to users and enterprises. Strengthening hard disk protection is one of the effective means to strengthen computer data storage and management. It is very important to take measures to ensure the data security. We should reduce the incidence of hidden dangers, so we must try and apply a variety of ways to realize the high security of data storage, that is, to take computer network security measures. All aspects of computer data security measures must be organically linked with physical security and network security. Network security is to protect the security of the communication process between the systems connected to the network, especially to ensure that authentication and access control are the key factors of network security. To protect network security, we must design network platform security strategy scientifically, publish scientific and natural network security management actions, record all network activities as much as possible combined with firewall, and pay attention to the physical protection of network equipment. At the same time, it can detect the uncertainty of the network platform system and create a very valuable identification and authentication mechanism.

5. Conclusion

The in-depth application of computer storage and management platform is a long-term and arduous task. Based on the perspective of "big data", this paper studies the development and production of computer storage and management platform. This paper argues that the application of joint computer storage and management platform supported by big data not only requires researchers to have strong development ability and comprehensive quality, but also puts forward high requirements for research technology, which will enhance people's trust in computers. Make the computer storage and management level further improve, and actively promote the sustainable development process of computer storage and management platform.

References

[1] Zheng X, Fu M, Chugh M. Big data storage and management in SaaS applications [J]. Journal of Communications & Information Networks, 2017, 2(3):18-29.

[2] Lu Y T, Cheng P, Chen Z G. Design and Implementation of the Tianhe-2 Data Storage and Management System [J]. Journal of Computer Science and Technology, 2020, 35(1):27-46.

[3] Yang J, Qiao Y, Zhang X, et al. Characterizing User Behavior in Mobile Internet [J]. IEEE Transactions on Emerging Topics in Computing, 2017, 3(1):95-106.

[4] El-Erian M A. Understanding Today's Trade Friction [J]. The International Economy, 2018, 32(2):40-41.

[5] Majeed J G, Ibrahim Y K, Al-Aziz G A A. Influence of Heat Treatment on Corrosion of Low Carbon Steel in Sulfuric Acid Solutions [J]. European Journal of Engineering Research and Science, 2019, 4(4):64-69.

[6] Zhao F, Zhang L Y, Zhao M M, et al. Architecture and technical exploration of big data platform for ecological environment [J]. Chinese Journal of Ecology, 2017, 36(3):824-832.

[7] Poltavtseva M A. A Consistent Approach to Building Secure Big Data Processing and Storage Systems [J]. Automatic Control and Computer Sciences, 2019, 53(8):914-921.

[8] Zuo L. Data Transfers Using Bandwidth Reservation Through Multiple Disjoint Paths of Dynamic HPNs [J]. Journal of Network and Systems Management, 2021, 29(3):1-24.

[9] Linsi L, Daniel K. Mügge. Globalization and the growing defects of international economic statistics [J]. Review of International Political Economy, 2019, 26(3):1-23.

[10] Yuhang C. Research on application of computer aided design in environmental engineering [J]. Agro Food Industry Hi Tech, 2017, 28(1):2769-2773.