Minimum Redundancy Distributed Storage System of Enterprise Financial Information Based on Fuzzy Control

Yan Yu(✉) and Lin Chen

Wuhan Qingchuan University, Wuhan 430204, China

Abstract. In view of the problem that the load value of traditional storage system is too small, resulting in poor distributed burst effect, this paper designs a minimum redundancy distributed storage system of enterprise financial information based on fuzzy control. In the hardware part, the storage control board is designed, and the minimum redundant data acquisition circuit is constructed to control the operation load of each hardware branch. The software uses fuzzy control to process the minimum redundant information, and uses JavaScript to write distributed algorithm to realize the storage function of the storage system. After preparing the system test environment, two kinds of traditional storage system and design storage system are prepared for experiments. The results show that the design storage system has the largest load capacity and can carry a variety of storage tasks.

Keywords: Fuzzy control · Enterprise accounting information · Minimum redundancy · Distributed storage

1 Introduction

Financial information refers to the economic information used to reflect its characteristics in combination with other information. Compared with financial information, non-financial information refers to all kinds of information that appear in the form of non-financial information and have direct or indirect connection with the production and operation activities of enterprises. Nowadays, the business activities of enterprises are more and more complex and the expectation of shareholders for financial information is higher and higher. How to determine the scale of financial information disclosure has become one of the most difficult problems in accounting practice. The disclosure of financial information changes with the change of economic environment. The following factors affect the determination of disclosure content, and then affect the cost of financial information disclosure. The use of enterprise financial information is the top priority of enterprise financial information work. Only by making good use of enterprise financial information can we give full play to the role of "barometer" and "early warning device" of enterprise financial information.

At present, high-speed storage technology has become the key technology to be solved in the era of enterprise accounting big data. Fuzzy control can adjust the extremely

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2022
Published by Springer Nature Switzerland AG 2022. All Rights Reserved
S. Liu and X. Ma (Eds.): ADHIP 2021, LNICST 416, pp. 391–402, 2022.
https://doi.org/10.1007/978-3-030-94551-0_31

high data transmission bandwidth, so its research has important practical value and application significance.

For this reason, relevant scholars have carried out research and made some progress. Literature [2] proposes a storage system of financial accounting information based on big data. Through the characteristics of big data, it plays an auxiliary role in the calculation of enterprise financial accounting work, the storage of accounting data, accounting prediction and other work, improves the quality of enterprise financial accounting work, optimizes the work process, and gives full play to the management function of accounting, But the system load of this method is too small.In reference [3], a financial information storage system based on UFIDA NC financial system is proposed. The system will store data information in the server of the enterprise headquarters. At the same time, each branch will set up an ERP system independently, and the headquarters will manage the ERP system of each branch by running UFIDA NC financial system, so as to improve the effect of financial information storage, But the effect of this system is not good.

Therefore, the research on the minimum redundancy distributed storage system of enterprise accounting information based on fuzzy control can provide reference for other high-speed storage systems. On the original hardware of the system, by adding other hardware, it can meet the needs of various functions [2]. High speed AD is added to the hardware of the system, and DSP is added to the back-end. High speed ad samples the data and transmits the data to DSP at high speed through the PCI bus. DSP is used to process the received data by compression, jigsaw puzzle and other algorithms. Based on this hardware platform, the design of high-speed data processing system based on PCI + DSP can be realized. Adding SATA hard disk on the basis of the original hardware of the system, because of the high data bandwidth and large capacity of SATA, the data can be stored into the SATA hard disk at high speed through the PCI, and the SATA hard disk can work in parallel to form a distributed storage structure.

2 Hardware Design of Distributed Storage System

2.1 Design Storage Control Board

The storage control board is composed of FPGA interface module, PCI hard core module, DDR3 memory module and six way tlk2711 interface module. The storage control board is designed as shown in the figure below (Fig. 1).

In the control board structure shown above, the FPGA is mainly used to realize the high-speed DMA read and write based on the PCIe hard core, the DDR3 controller, the design of the cache module and the interface design of the whole system. Xilinx virtex6 FPGA is adopted in the control board, which sets the working frequency to 600 MHz, supports DDR3 storage interface, provides DDR3 controller customization and rich digital clock management (DCM) resources [3]. Integrated with the PCIe hard core, it supports v2.0 protocol and X8 link width. GTX transceiver supports the transmission rate from 150mbit/s to 6.5 gbit/s. After the structure of the storage control board is built, the minimum redundant distributed data acquisition circuit structure is constructed.

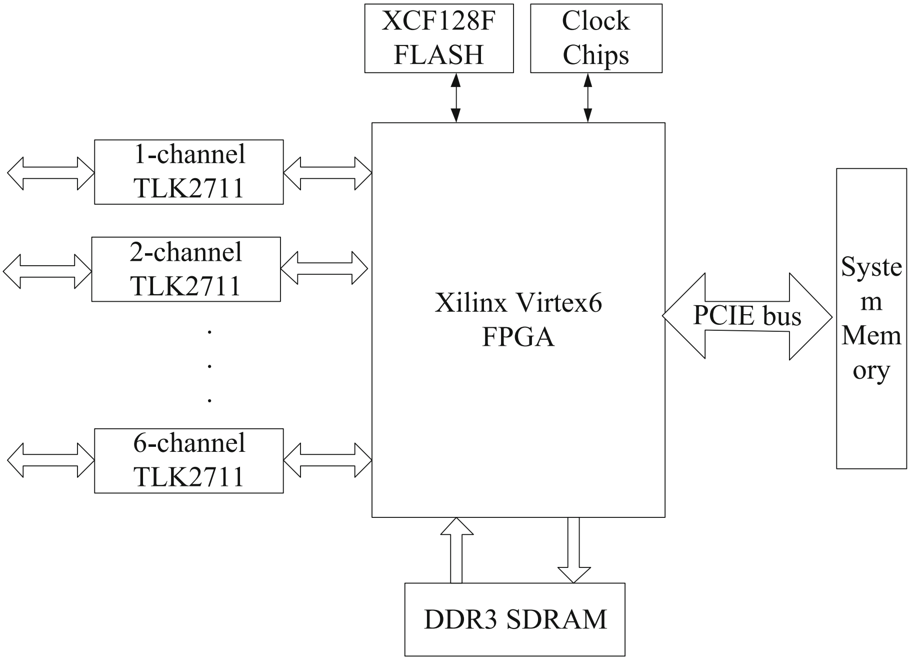

Fig. 1. Structure of storage control board

2.2 Construction of Information and Data Acquisition Circuit

Under the above hardware control, the serial port uses 422 interface chip to realize the conversion between TTL and serial port level. The control panel and serial communication select RX and TX cross connection mode [4] to realize the transmission of read-write data. The final connection circuit is shown in the figure below (Fig. 2):

According to the control circuit shown in the figure above, the plug-in serial port is used to connect the plug-in terminal, and the 9-pin serial port device terminal is controlled to run three interface units to control the voltage between the home appliance control board and the main controller to achieve the normal transmission of data. Four PNP triodes are used as the current driving unit of the relay, and four monochrome light-emitting diodes are set to monitor the working state of the furniture working unit [5]. Control MCU pin P1.0~p1.3 for low level, at this time the LED is on, diode interface Q1~Q4 is in the state of conduction, the internal coil of the relay is powered on, pull the internal switch of the relay, keep the relay in the closed state, synthesize the above circuit design, complete the hardware design of the storage system.

3 Software Design of Distributed Storage System

3.1 Using Fuzzy Control to Deal with Minimum Redundant Information

Using the distributed fuzzy controller structure, when processing the minimum redundant information, the clear physical quantity of the output and the output is converted into a

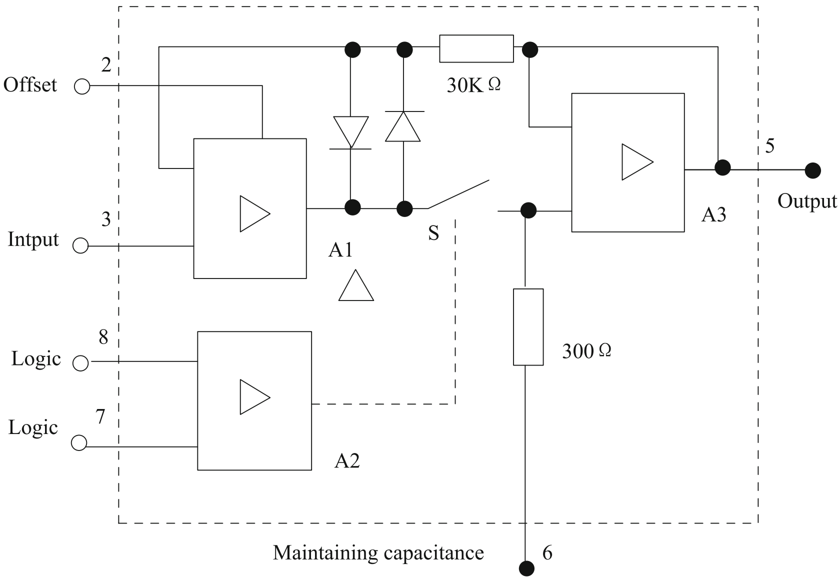

Fig. 2. Hardware control circuit

fuzzy quantity, and the physical theory domain of a component y_j of the known input variable Y_1 is:

$$Y_j = [-y, y](y > 0) \tag{1}$$

Obtain the fuzzy universe corresponding to Eq. (2) at this time as:

$$N_j = [-m_j, m_j](m_j > 0) \tag{2}$$

At this time, the fuzzy transformation coefficient k_j from the input variable Y to the variable m_j is the quantization factor, and the following conversion formula is obtained:

$$k_j = \frac{m_j}{y_j} \tag{3}$$

Among them, k_j is always positive and $k_j > 0$. According to (3), when the variable Y changes, in order to keep the control core of the fuzzy controller unchanged, only need to change k_j, so that the input quantity in the fuzzy processor can still fall into the original universe of discourse after being changed. In, the domain of discussion at this time is the use rule of fuzzy controller.

After formulating the use rules of the fuzzy controller, the approximate reasoning algorithm used by the fuzzy controller is selected [6]. Take the input element with the largest membership degree in the fuzzy set in the inference rule as the precise control quantity, adjust multiple elements to have the largest membership degree, and take the average of the membership degrees of several elements. The calculation formula is as follows:

$$\begin{cases} \Gamma = \frac{k_j}{2} L \\ \delta = \frac{L}{2} \left(\frac{d\beta_i}{d\beta_j} - \frac{\beta_j}{\beta_i} \right) \\ \beta = \frac{L(\beta_i - \beta_j)}{2} \end{cases} \tag{4}$$

Among them, Γ is the membership degree of the elements, δ is the precise control quantity, L is the number of elements, d is the applicable rules of the fuzzy set, and β_i and β_j are the maximum membership degrees of the two elements. The degree of membership is regarded as the minimum redundancy of corporate accounting information [7], which is used to realize the storage function of the system.

3.2 Storage Function Realization

When using JavaScript to write distributed algorithms and integrate them into different operation groups [8] according to the magnitude of the redundancy value, the operation grouping can be expressed as:

$$M(j) = \frac{P(j)}{Z(j)} \tag{5}$$

Among them, j represents the operation grouping, M represents the distributed function, P represents the signal power function, and Z represents the total number of enterprise accounting information operation groups. Choose a level of attribute as the storage index, and the storage process can be expressed as:

$$SE = \mu \frac{\langle E_t(t) - E_r(t) \rangle}{\langle E_t(t) \rangle} \tag{6}$$

Among them, μ represents the distribution parameter, $E_t(t)$ represents the information signal energy under the time scale, and r represents the signal data receiving parameter. When importing the storage algorithm [9], based on the system's receiving coefficient parameters, under the photon gain control, the algorithm import process can be expressed as:

$$G_r(t) = \frac{g(N_r(t) - N_0)}{(1 + sE_r)} \tag{7}$$

Among them, g represents the introduction rate, $N(t)$ represents the number of carriers, s represents the feedback coefficient, and N_0 represents the initial carrier number of the algorithm. After the distributed algorithm is imported, the storage algorithm is matched with the information attribute [10], and different distribution parameters are set according to the actual number of attributes to meet the needs of enterprise accounting information. Integrate the above hardware and software design, and finally complete the design of the distributed storage system.

4 Simulation

4.1 Experiment Preparation

Due to conditions, a pseudo-distributed cluster construction method was used when building a Hadoop cluster in the experiment, that is to say, a Hadoop cluster was constructed by creating multiple virtual machines on a single node. In the experiment, VMware is used to build a total of six virtual machine nodes, one of which is the Namenode name node, and the other five nodes are used as Datanode nodes, and the node virtual machines are built into a cluster. The hardware configuration of the experimental node is shown in the following table (Table 1 and 2):

Table 1. Hadoop cluster hardware configuration table

Serial number	Namenode node	Datanode node
Operating system	CentOS 7 64-bit system	CentOS 7 64-bit system
CPU	Intel(R) Core(TM) i5-2450 MCPU@2.S GHz	Intel(R) Core(TM) i5-2450 MCPU@2.S GHz
RAM	2 GB	1 GB
Hard disk	20 GB	20 GB

Under the control of the hardware node configured in the above table, prepare the experimental platform shown in the following table, and the parameters of the experimental platform are shown in the following table (Table 2):

Table 2. Experimental platform parameters

Serial number	Name of software	Version
1	operating system	CentOS 7 64-bit system
2	JDK	jdk-8u201-linux-x64.rpm
3	Hadoop	hadoop-2.9.0
4	Programming language	Java

Under the control of the platform parameters shown in the above table, the virtual machine is used to configure jps, and the storage system equipped with the host computer is continuously debugged. Under the control of the platform parameters shown in the above table, the virtual machine is used to configure JPS, and the upper computer is continuously debugged to carry the storage system.

In order to measure and verify the read-write performance of the onboard storage system, the DDR3 write request design interacts with the upstream FIFO, writes commands, addresses and data to the DDR3 controller, and configures the related interface timing, so as to complete the write request operation to the DDR3 memory. The sampling sequence of DDR3 memory interface write request is shown in Fig. 3.

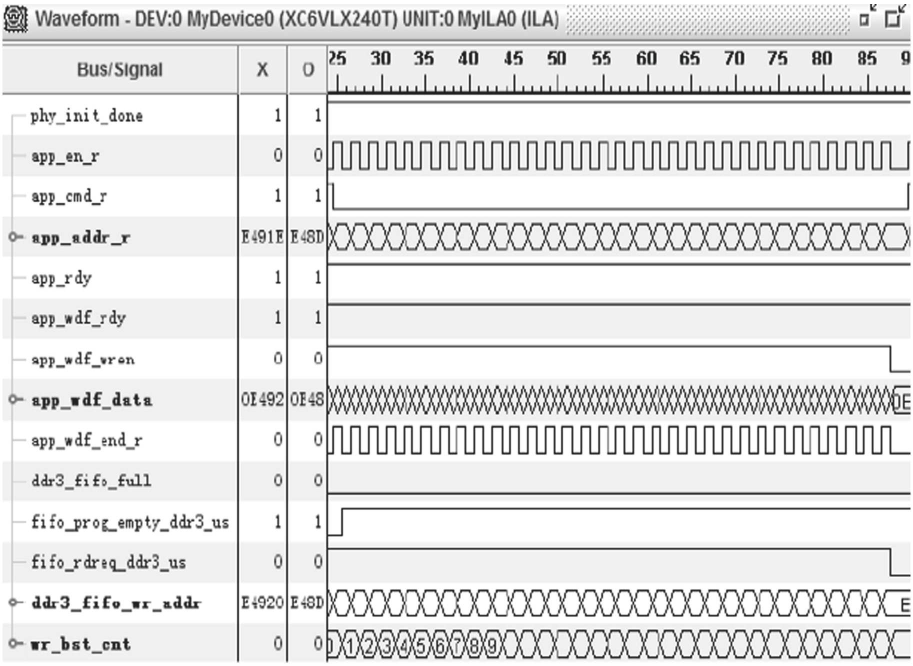

Fig. 3. DDR3 memory interface write request sampling sequence

In the process of request sampling sequence shown in the figure above, when the DDR3 controller completes the whole initialization process, it will send the signal PHY_ init_ Done set to valid. At this time, if the DDR3 memory is insufficient and the DDR3 uplink FIFO is not empty, the corresponding DDR3_ fifo_ Full signal and FIFO_ prog_ empty_ ddr3_ It's a signal. If command and address FIFO are ready to receive command and write data FIFO is ready to receive data, corresponding to app in the figure_ RDY and app_ wdf_ RDY signal, the user should initiate a write command operation, that is, the app_ Set CMD to low and give the write command address app_ addr。

When the write command is submitted successfully, the user should send the app_ Set WDF Wren to valid, and set the_ The data is submitted on WDF data, and the DDR3 memory write address is given. When the line is full, wr_ bst_ When CNT equals 32, the write operation stops. Through the analysis of the above figure, it can be seen that the sampling timing fully conforms to the DDR3 memory write design. In order to simplify the operation process, five switches are used to connect ZigBee devices, of which two switches are used as the core layer, and the remaining two switches are used as the access layer. After connecting the enterprise controlled area network, a firewall is set outside the enterprise LAN, and the final system test environment is as follows (Fig. 4):

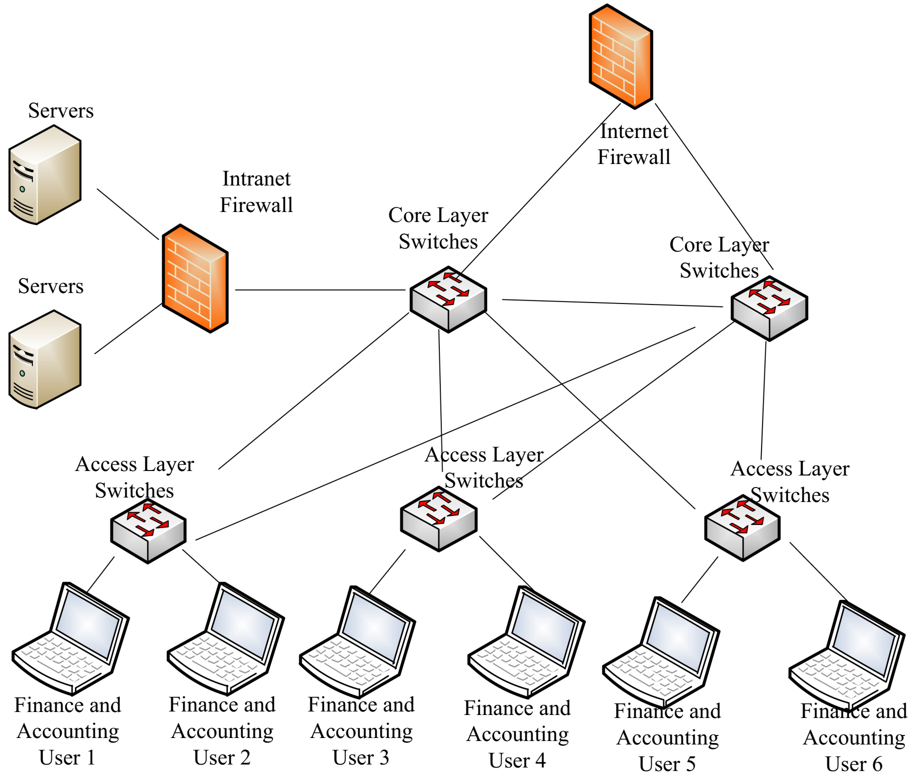

Fig. 4. Build the test environment of the system

In the test environment shown in the figure above, two ZigBee modules are used to continuously receive the information of the switch. The node of enterprise accounting information connects the minimum redundant information terminal node with the experimental PC through the hardware serial port. After starting the coordinator of the test environment, a distributed test network is built. In the above test environment, two traditional distributed storage systems and the designed reservation system are used to compare the performance of the three storage systems.

4.2 Structure and Analysis

Based on the above experimental preparation, the PCB antenna is used as the transmitting point of the minimum redundant information of the test node. The transmission power of the information is set to 5 dBm, the test data packet is 8 characters, the baud rate is set to 115000, and the minimum redundant accounting information is circularly sent for 1000 times (Table 3).

It can be seen from the packet loss rate results shown in the above table that with the increasing number of iterations, the packet loss rates generated by the three reservation systems show packet loss rates of different values. After the number of information cycles

Table 3. Packet loss rate of three storage systems

Number of times of sending information	Packet loss rate /%		
	Traditional storage system 1	Traditional storage system 2	Design storage system
50	8.6	4.5	0.4
100	8.7	4.8	0.4
150	8.7	4.8	0.6
200	8.7	4.8	0.6
250	8.7	4.8	0.8
300	8.7	4.9	0.9
350	8.8	4.9	1.1
400	8.9	5.1	1.1
450	8.9	5.2	1.2
500	8.6	5.9	0.4
550	8.6	5.1	0.7
600	8.7	5.7	0.8
650	8.8	5.5	1.0
700	8.5	5.2	0.7
750	8.2	4.2	0.8
800	8.8	5.7	0.8
850	8.9	5.8	1.0
900	8.6	5.6	1.0
950	8.6	5.8	0.4
1000	8.7	5.9	0.7

with the same accounting information number is fixed, the packet loss rate generated by the traditional storage system 1 is the largest, with a value of about 8.9%. The packet loss rate of traditional storage system 2 is about 5.2% under the same number of transmission, and the value of packet loss rate is small. However, the packet loss rate of the storage system designed in this paper is about 1.2% under the same number of transmission, and the value of packet loss rate is the smallest.

In the above experimental environment, simulate the data transmission phase of a static network without node join and exit behavior, and use three kinds of storage systems to test. If the number of information nodes participating in the regeneration process is d each time, for a single node storage process, it is necessary to start three storage processes, select d information nodes each time, and construct their transfer path to a new node Transmission path: the storage process of the next node will be started only after one node completes the storage. For a single storage information, the system storage process only needs to be started once. In this process, the transmission path from D information nodes and other two new nodes to this node is constructed for each new node. The actual available bandwidth of the three storage systems is simulated and counted. The available bandwidth values are shown in the figure below (Fig. 5):

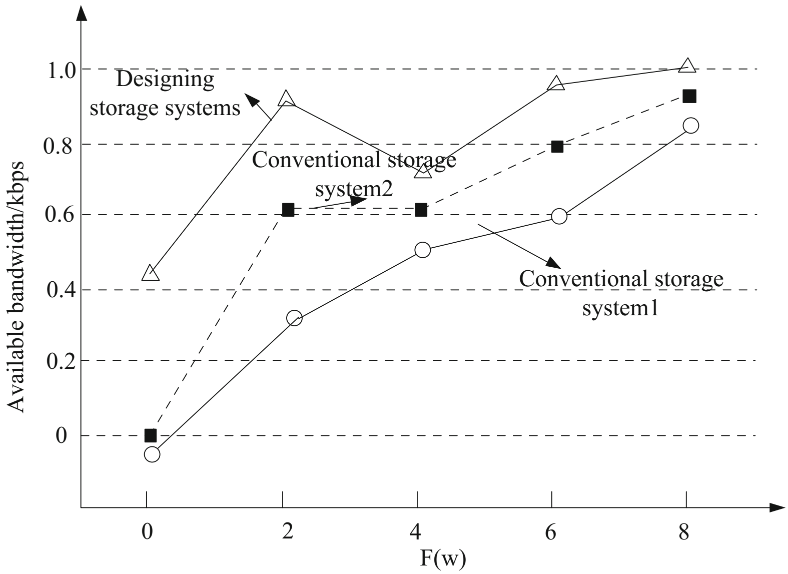

Fig. 5. Available bandwidth values of three storage systems

According to the experimental results shown in the figure above, under the control of the same information node, the available bandwidth value of traditional storage system 1 is about 0.8, and the actual available bandwidth value is the smallest. The available bandwidth value of traditional storage system 2 is about 0.9, and the actual available bandwidth of the system is larger. The available bandwidth value of the designed storage system is about 1.0, which is compared with the two traditional storage systems The designed storage system has the largest available bandwidth.

Keeping the above experimental environment unchanged, the performance runner software is used to test the three storage systems, and the test software is used to test the load, configuration, concurrency and other aspects of the performance of the three systems. The average response time of the evaluation software to the output of each module of the storage system is taken as the index of performance test, and the test results are shown in the following table (Table 4).

According to the experimental results in the table above, in the operation of each function of the system, the average reading and writing time of the storage system to the database is controlled within 1 s, and the response time of the hardware device is about 13 s. The average reading and writing time of the two traditional storage systems is more than 1 s, the response time of the traditional storage system hardware is about 17 s, and the response time of the traditional storage system hardware storage system is

Table 4. Response time test results

Storage function name User login	Average response time/ms		
	Traditional storage system 1	Traditional storage system 2	Design storage system
Video surveillance	1503.4	1268.2	731.5
Personnel positioning	864.3	659.7	452.8
Equipment information query	16083.9	14683.5	12834.5
Device delete operation	12523.4	1020.7	783.1
Issue	1010.2	869.7	616.3
Give an alarm	1120.5	986.4	817.8

about 15 s, which is larger than the storage system designed in this paper. Then, the first mock exam system is maintained, and three storage systems are tested. The number of users who store different online storage 50, 100 and 200 is simulated. The load capacity of the system is tested when the multi-user access the same module or data.

Table 5. Concurrent test results

Number of storage users	Load capacity /KBps		
	Traditional storage system 1	Traditional storage system 2	Design storage system
10	405.45	524.92	605.75
20	412.47	504.85	612.46
50	421.12	573.86	621.23
100	437.15	556.74	627.63
150	451.71	510.81	657.26
200	386.72	549.51	653.54

According to Table 5, when the number of storage users is 10, the load capacity of traditional storage system 1 can reach 405.45 kbps, the load capacity of traditional storage system 2 can reach 524.92 kbps, and the load capacity of designed storage system can reach 605.75 kbps; When the number of storage users is 150, the load capacity of traditional storage system 1 can reach 451.71 Kbps.The load capacity of traditional storage system 2 can reach 510.81 Kbps, and the load capacity of designed storage system can reach 657.26 Kbps; This method always has high load capacity.

5 Conclusion

With the rapid development of Internet and the acceleration of information digitization, the information age has come. The following is the rapid growth of data on the Internet. Traditional data storage methods are not ideal in terms of scalability, reliability and access delay, which can not adapt to the storage of massive data. Distributed storage system emerges as the times require. The designed storage system can improve the shortcomings of traditional storage system and provide theoretical support for the design and construction of storage system in the future.

References

1. Jing, W., Xuefei, Z., Shuxia, W., et al.: Construction of locally repairable codes based on fractional repetition cyclic codes in distributed storage systems. J. Electron. Meas. Instrum. **33**(04), 146–153 (2019)
2. Liang, P.: Research on the transformation from financial accounting to management accounting in the era of big data. Glob. Mark. **01**(03), 71–73 (2020)
3. Wen, H.: Financial information construction and fine management based on UFIDA NC financial system. Mod. Econ. Inf. **01**(21), 207–209 (2019)
4. Gou, Z., Zhang, X., Wu, D., et al.: Log analysis and workload characteristic extraction in distributed storage system. J. Comput. Appl. **40**(09), 2586–2593 (2020)
5. Guo, W., Xie, G., Zhang, F., et al.: Design and implementation of a mimic architecture for distributed storage system. Comput. Eng. **46**(06), 12–19 (2020)
6. Reyes-Ruiz, G., Hernández-Hernández, M.: Fuzzy clustering as a new grouping technique to define the business size of SMEs through their financial information. J. Intell. Fuzzy Syst. **40**(2), 1773–1782 (2021)
7. Popa, A., Safta, I.L., Miron, G.M., et al.: Manipulation of financial information through creative accounting: case study at companies listed on the romanian stock exchange. In: 18th RSEP International Economics, Finance & Business Conference – Virtual/Online 26–27 August 2020, Kadir Has University (2020)
8. ZHOU Yue, LI Gui-yang, JIANG Xiao-yu,et al. Piggyback Code Based on Distributed Storage Systems[J]. Journal of Chinese Computer Systems, 2020,41(05):1091–1097
9. Saghaeiannejad-Isfahani, S., Salimian-Rizi, N.: Assessment of success of financial information system in educational, health, and medical centers affiliated to Isfahan university of medical sciences. J. Educ. Health Promot. **9**(1), 128–135 (2020)
10. Frank, E.: An assessment of the conceptual linkages between the qualitative characteristics of useful financial information and ethical behavior within informal institutions. Ekonomski Horizonti **22**(2), 137–148 (2020)