

BIOINFO Computer Engineering

BIOINFO Computer Engineering

ISSN: 2249–3980 & E-ISSN: 2249–3999, Volume 1, Issue 2, 2011, pp-10-19

Available online at: <http://www.bioinfo.in/contents.php?id=322>

EXPLORING WAYS TO OPTIMIZE SERVER PERFORMANCE BY RETRIEVAL MEANS

BIN YAN^{1*} AND FENG SHI²

¹Nanjing University of Posts and Telecommunications, 210003, Nanjing, China

²Education Science College, Shandong Normal University, 250014, Jinan, China

*Corresponding author. E-mail: 18951896220@189.cn; yanb@njupt.edu.cn, qqbaibai198807@163.com; shifeng@sdu.edu.cn

Received: August 18, 2011; Accepted: September 16, 2011

Abstract- A server is the essential equipment and retrieval is also often using mean in software system. In this paper we propose results that large amounts of data in a server were tested by retrieval means. One of results obtained shown to be unable to increase running speed of management system with RAID and another result shown to be slower the retrieval speed of the B/S than the C/S. Optimizing server performance should be primarily way to standardize design a database management system and a software system.

Keywords: RAID, Retrieval, Server Optimization, B/S, C/S

1. Introduction

We known that these such as computer network, server performance and number, large amounts of data, table structure of database are important factors of effecting server we had tested before [1]. Computerized management systems are still a kind of substitute for tradition or handwork mode and process as design aim and use functions to resolve problems of physical management software. Large amounts of data in the

server caused slowdown in running system. There are many ways to optimize a server and a software system. A way was substitute RAID for common hard disks which data saved on the hard disk is based on disk striping. The striping according to different combinations developed RAID technology. The stripe is to fragment continuous data into same size blocks of data then to be written respectively each piece of data on the array disk

and is different algorithms or methods. This technology can provide much faster and more than a single disk to read and write speed. RAID used an adaptive sector grouping approach to accessing a distributed RAID [2]. A software system for affection of a server performance has been evaluated [3]. But hardware investment whether it has a big improvement of server performance or not especially on such RAID needs to be tested to describe. Independent of computer system configuration facilitated easy changes of test algorithms and implementations [4]. Our mean was to test retrieval speed of items in management system via simulated in other server because all tests could not process in current running server. But data tested are of lack for some hardware as external devices and homepage or webpage and researchers require further study.

2. Testing RIAD attached

The library computer network of NUPT is a sub-network of the campus network and its gateway switch is managed by library technology department. There are 8 sets of servers in the department such as Web site, FTP, electronic publication (book and periodical) system, circulating system, mirror image retrieval data, CDs' management system and as on. Each floor was installed 24 ports switch that was electrical or optic fiber import port. RIAD is an abbreviation of Redundant Array of Inexpensive Disks, and bigger capacity hard disks that connected many special hard disks or normal hard disks through socket. Web site of the library was installed in

TSNGHUA TONGFANG TP200. The biggest characteristic of RIAD is data access very fast and its main function is to enhance network availability and data storage capacity and data selectively distributed across multiple disks to improve system data throughput. Based on the above information and understanding of RIAD and recognized by many experts, in order to solve effectively the large amount of data on the impact of the server performance, there was no other alternative solutions before, we decided to try it. Before installation, we tested the original system and observed server performance.

2.1 Equipment preparation

Our RIAD used was not only to increase capacity but also to improve transmission speed between hard disks and CPU. First of all, a file server was observed and tested before installed RIAD because the capacity of data occupied impressing retrieval speed. The file server had one 18.2GB hard disk installed operation system and two 32.2GB disks installed database management system as disk mirror. So the arm of test looked remain capacity of disks, the more which were, the lower effected.

The model of RAID by made USA was SANE-6000-16-FFx2-2G-3, transmission speed 2Gb/s, 512ECC cache, QLOGIC 2Gb QLA2310F, 146GB ST3146807FC disk (10 disks). The model of the main server by made Japan was Fujitsu GP7000F Model 400A type, SPARC64-GP300MHz double CPUs, 2GB memory. The operation system installing in server was Solaris Unix 7 and the database management system was Sybase 11.9.2. The double CPUs of the server account

for below 15% of the load and 2GB memory consumed about 30% in good order. One 18.2GB disk remained space 4.2GB after installed system files, accounted for 23% of the total memory space and another 18.2GB disk copied data remained space 3.7GB, 23% total memory space. A pair of 36.2G disks were data mirror and shown 10G and 10.8G, account for 28% and 30% of difference remain space. In fact a retrieval time took over 30 seconds in hundred thousands bibliography data or over one million of daily record data. The result of retrieval often showed same two records on computer screen. It was explicated that CPUs was faster than screen for data to process. And data were often not written to database or written into other data table when client end wrote data to database in server in network. The explication was above same before RAID connected to the server.

2.2 Testing plan

(1) network divided

Test respectively was carried out at a practical network which was normally online data transmission on LAN and at an ideal one which was disconnected outer with switches.

(2) network segment divided

The network in library consisted of 155 MB/s multimode fiber and 1000 MB/s single mode fiber. One switch is IBM 8271 model with 10/100MB/s 24 ports connected 155MB/s multimode fiber and used ATM protocol. Other switch is Digital China DCRS-5512GC model with 100MB/s 8 ports connected 1000 MB/s single mode fiber and used Ethernet protocol. The test method was same

network segment and straddle network segment for the influence of gateway and connecting materials.

(3) RAID capacity divided

RAID used by us with 146GB X16 disks has 2.3TB capacity memory and 4 disks of RAID used band technology 0+5 because data in file server are impossible to occupy many capacity. The reading operation of RAID 5 is faster than it of RAID 1, but the writing operation of RAID5 is more slowly than RAID1.

2.3 Testing items

(1) client and webpage

LAN in a library is consisted of client/server and browse/server models of the management system. The C/S model may simulate ideal network to test and exclude other signal interference, such as a server computer connected with a client computer directly and the B/S done not.

(2) retrieval items

Our testing was only retrieval by retrieval words, retrieval out numbers because of no written to the server affected accuracy data in. It took different time to different retrieval words and retrieval number. The different data in database fields were retrieved out different number such as book bar code, reader bar code, ISBN, which there was a piece of record in circulate system and many pieces of records in reader system. Except 2-5 seconds we need about 10 seconds to take and about 3-5 pieces of records to retrieve.

(3) testing item

- custom of circulating subsystem

Way one

Menu columns were selected book management and retrieval item was done book author. Retrieval word written "Wang Liang Yuan" and used accurate match. Number of retrieval records was 12 out.

Way two

Menu columns were selected book management and retrieval item was done book title. Retrieval word written "Advance English" and used fuzziness match. Number of retrieval records was 404 out.

Way three

Menu columns were selected daily record under statistic query and retrieval item was done book author. Retrieval word written "Wang Liang Yuan" and statistic time was from Feb. 24, 2004 to Feb. 23, 2005. Number of retrieval records was 6 out.

Way four

Menu columns were selected daily record under statistic query and retrieval item was done book author. Retrieval word written "Zhang Xin You" and statistic time was from Feb. 24, 2004 to Feb. 23, 2005. Number of retrieval records was 91 out.

Way five

Menu columns were selected daily record under statistic query and retrieval item was done book author under lending history. Retrieval word written "Zhang Xin You" and statistic time was from Feb. 24, 2004 to Feb. 23, 2005. Number of retrieval records was 170 out.

- custom of OPAC subsystem

Way six

Menu columns were selected one condition query under book query and retrieval item was done book title.

Retrieval word written "computer network" and used fuzziness match. Number of retrieval records was 370 out.

The results above six ways are shown in Table 1.

Because data in Fujitsu GP7000F Model 400A type with the increase, its retrieval speed and write data time had caused sharp down and could not meet the needs of the growing number of readers. We had acquired a set of Fujitsu PrimePower Model 450 type, 2 × SPARC64V (1.98GHz/3MB), 4 × 1GB RAM, 2 × 73GB (million RPM) hard disk. According to previous test methods, results were shown in Table 2.

2.4 Result analysis

"Way* before" was the original system test results and "Way* after" was the test results with an external disk array without any data changes. The results from the test speed were the fastest at "way two after", a difference of 1.6 seconds with way two before, and others were little different. If using disk array to improve server performance and improve data retrieval speed, you want to expend several hundred thousand yuans so the price is too big.

Four disks of RAID were used and remainder twelve disks could be used for other data management. Other data were literature databases and specialty database with a medium PC server to connect and with 0+1 disk array technology. From the original server Fujitsu Primergy MS 610 (dual CPU P III Xeon 700), 2G RAM, hard disk IDE 18.2G, 10/100 attached to ESCORT DA

3500 disk array (SCSI 73G×8, 128K cache, RIAD 0+1) was updated the SCSI card to optical card and optical interface hard disks. These have good improvement of retrieval speed and service management in the literature database system.

In table 2, due the limits of date in new server, way one checked out 18 records, way two did out 639 records, way three and way four did out same as the ole server, way five did out 31, way six did out 525. New server Fujitsu PrimePower Model 450 type was better performance than GP7000F Model 400A type in retrieval speed and function management of software system.

3. Testing Browser/Server and Client/Server

With the development and rapid spread use of the computer and the Internet network in the political, economic, social and other fields, the whole society's dependence on the network has reached an unprecedented climax. Computer networks have become the lifeblood of the country's economic base and become a powerful driving force of social development and its position becomes more and more important. Mainstream R & D companies and users were better technology either client/server (C/S) or browser/server (B/S), which could represent the trend of technology development and the note constantly of other issues. These require test results to discuss issues.

The C/S (Client/Server) structure, namely the client and server architecture and the software system architecture, allows full use of the advantages of the hardware environment of both ends, the rational allocation of tasks to the client side and server side to achieve, reducing the

system's communication overhead. Most current application software systems are Client/Server in the form of a two-tier structure. Now that the software applications are distributed to Web application development, Web and Client/Server applications can perform the same business process, application of different modules share the logical components. Therefore, the internal and external users can access the new and existing application systems and can be extended in new applications through the existing application logic. This is some computer experts believe that the current application system development.

The B/S (Browser/Server) structure, i.e. a browser and server architecture with the rise of Internet technology, was a change or improvement of the C/S structure. In this structure, the user interface is working to achieve through 3W browser, and very small part of the business logic in the front (Browser) to achieve but the main business logic or the server side (Server) to achieve the formation of the so-called three-tier structure. Thus, it may be greatly simplifying the client computer load, reducing system maintenance and upgrade costs and the workload and reduce the overall cost of the user (TCO). Look at the current technology, the establishment network applications of LAN the B/S structure and database applications through Internet/Intranet mode were relatively easy to grasp and the lower cost. It is the one place the development and achieves different people from different locations, different access methods (such as LAN, WAN, Internet / Intranet, etc.) to access and manipulate a common database. It can effectively

protect data platform and administrative access and the server database is also very safe.

3.1 Test methods

(1) the choice of network environment

Because of computer used through the LAN and Internet, the actual network choice was connected to LAN and tested the impact of data retrieval in network transmission. The ideal network was not connected with a gateway but the computer connected directly with a cable to compare influence degree of data retrieval with practical network.

(2) test equipments

Two sets of computer used Founder Wenxiang 600, P4 1.5G CPU, 256MB RAM, 40GB hard disk. A computer installed MS SQL Server7.0 with FrontPage as the web interface design and another installed PowerBuilder6.5 as a client interface design, and all of them installed software of operating system Windows2000 Server.

(3) the choice of retrieval words

The experiment was 3,748,019 records of bibliographic data which transferred to test computer from the actual running server. There were bar code of book, ISBN, book author, book call number, book status, book publisher, publishing date, book title and accession number in the original data table as retrieval items. Finally, we chose bar code of book as a retrieval item and with "252 959" as the retrieval word.

3.2 Test results

The stored data in database management system must to protect compatibility during changing date to ensure valid data into the database. The database must be

consistent multi-user concurrent updates with DBMS to ensure the users and their change not to affect the other user's job. The data updated caused much higher risks than data query, and retrieving results could illustrate more reliability and safety than writing and deleting actual data.

The test results in ideal network environment show Table 3 and the test results in actual network environment see Table 4.

After replacing the server, Fujitsu PrimePower Model 450 type, we get results as 3.05, 2.83, 3.10 and average value 2.99 via B/S model. Because the Web server was not replaced, records were retrieved out 515 instead of 525.

3.3 Results analysis

(1) comparison of the ideal and the actual environment

Test data have been listed separately in Table 3 and Table 4, including the arithmetical average value of three day test data and the total average value of three day average values under the ideal environment and the actual environment. We can be seen from the tables that retrieve time under an ideal environment was longer than under an actual environment, but the difference was not significant. It indicated that the existing network did not constitute impact to retrieve data and data transmissions of the server.

(2) retrieval speed comparison

The B/S structure uses a logical three-tier structure and the physical network structure is still the original Ethernet. In this way, they occupy the same network line between

the first layer and the second layer and between the second layer and third layer for the communication. The C/S structure is only two-tier, and network traffic only communicates between the Client and Server traffic. So even though the C/S ability to handle large amounts of information is not equal to the B/S's, the C/S in the logical structure a small layer than the B/S, the C/S speed is better fast than the B/S to complete the same task, and it makes the C/S be more conducive to handling large amounts of data.

(3) design program comparison

A test program of the B/S design after the release was 14.8KB size of the file, and program of the C/S design after the compiled was 56.0KB size of the file. It could be seen from the file size that the B/S file involved was small but took the space of server's hard disk, and C/S related file should be bigger than the B/S but the occupation was the space of client's hard disk.

(4) network impact for the B/S and the C/S

The good or bad of network for the B/S and C/S would have a direct impact, especially for the B/S. If the network where was the very poor, would directly lead to delay or failure in the search for the B/S (search time for more than 30 seconds, the system would prompt the delay or failure to search, this data was obtained in the test).

4. Conclusion and discussion

Under the current network bandwidth and network equipment management system not to cause the

retrieval bottleneck, the server hardware and application software may be the main cause problems.

4.1 Application characteristics of B/S and C/S

We can obtain some characteristics on the B/S and the C/S via the general theory and this test. The B/S can simplify application programs at client end and simplify system development and maintenance, and make installation easier, more flexible network structure. The C/S has strong interaction, a complete set of application programs that provide more secure access mode.

The C/S is a structure model of a pair of point-point and use better network security protocols on local area networks. Though the JavaScript and the VBScript provide a certain amount of interactivity for the B/S that compared with C/S, a set of client application programs is too limited. The C/S is only a two-tier structure and is better fast than the B/S to complete speed for the same task. The C/S makes more conducive handle large amounts of data.

In short, the B/S functions used should have the following characteristics:

- a wide range and flexible locations.
- frequent changes in function.
- no much security and interaction.

The C/S functions used should have the following characteristics:

- more security requirements.
- more strong interaction requirements
- small range and fixed location.

The advanced mode of the B/S and the mature mode of the C/S have people choice individual advantages of the MIS platform in the modern enterprise.

4.2 Combination the C/S mode and the B/S mode

This is a kind of more flexible combination way. Systems analysts based on the characteristics of the application system can use flexibly different MIS systems platform for different sub-systems. Many library management systems have crossed the two modes and used in parallel, which OPAC used the B/S and other sub-systems used the C/S.

4.3 The impact of else factors

We were unable to test the impact of tens to hundreds of access users for a large amount of data, but we can find from the actual phenomenon that the more access to computers, while the greater the likelihood of a field to read and write, the longer to return display information time to the client's. On the contrary, it is relatively faster and less error.

Relational database management system is a natural choice for the variety of systems software development. As a file server the hardware upgraded in any case, such as disk arrays, optical transmission interface, or the optimization software, such as using a narrow table, reducing the associated table, divide the amount of data, etc. can improve the system's data processing capabilities and optimize server performances partially.

Acknowledgments

I thank Mr. Tian-Ming Zheng, my student, in Nanjing University of Posts and Telecommunications for testing work during his undergraduate these. The project was supported by the Scientific Research Foundation of Nanjing University of Posts and Telecommunications

(NY206024).

References

- [1] Yan Bin, Chai Xiaojuan. (2005) *Beijing University of Posts and Telecommunications Press, Beijing, China.*
- [2] Hai Jin, Kai Hwang. (2001) *Cluster Computing* 4, 133-143.
- [3] Bin Yan. (2009) *Computer and Information Science.* Vol. 2, No. 3. pp. 32-39.
- [4] Samvel K. Shoukourian. (1998) *IEEE DESIGN & TEST OF COMPUTERS, APRIL-JUNE*, pp. 73-79.

Table 1- retrieval time of SCSI hard disk and RAID in server (second)

No.	Way one before	Way one after	No.	Way two before	Way two after	No.	Way three before	Way three after
1	25.21	23.31	1	27.97	26.53	1	1.52	1.76
2	25.06	23.70	2	28.97	25.90	2	2.13	1.76
3	25.18	24.32	3	26.33	26.04	3	1.74	1.59
Average	25.15	24.78	Average	27.76	26.16	Average	1.80	1.70
No.	Way four before	Way four after	No.	Way five before	Way five after	No.	Way six before	Way six after
1	1.85	1.84	1	16.50	16.05	1	2.53	2.40
2	1.75	1.69	2	16.15	16.15	2	2.60	2.43
3	1.70	1.50	3	16.24	16.76	3	2.45	2.48
Average	1.77	1.68	Average	16.30	16.32	Average	2.53	2.44

Table 2- retrieval time of RAID in old server and new server (second)

No.	Way one old	Way one new	No.	Way two old	Way two new	No.	Way three old	Way three new
1	23.31	7.08	1	26.53	7.83	1	1.76	1.03
2	23.70	7.06	2	25.90	7.97	2	1.76	1.06
3	24.32	6.98	3	26.04	7.83	3	1.59	0.91
Average	24.78	7.04	Average	26.16	7.88	Average	1.70	1.00
No.	Way four old	Way four new	No.	Way five old	Way five new	No.	Way six old	Way six new
1	1.84	0.97	1	16.05	16.49	1	2.40	2.36
2	1.69	0.98	2	16.15	18.75	2	2.43	2.37
3	1.50	0.94	3	16.76	19.77	3	2.48	1.89
Average	1.68	0.96	Average	16.32	18.34	Average	2.44	2.21

Table 3- The test results in ideal network environment (second)

	First day	Second day	Third day	First day	Second day	Third day
No.	B/S	B/S	B/S	C/S	C/S	C/S
1	31.59	32.89	31.08	23.60	23.94	20.43
2	29.36	33.39	28.14	21.73	20.98	20.24
3	42.19	36.03	31.14	26.03	28.74	21.67
average	34.38	31.78	30.12	23.78	24.55	20.78
Total average	32.09			23.04		

Table 4- The test results in actual network environment (second)

	First day	Second day	Third day	First day	Second day	Third day
No.	B/S	B/S	B/S	C/S	C/S	C/S
1	36.17	31.91	30.61	19.98	23.61	20.70
2	49.49	29.74	31.53	24.06	30.17	21.61
3	32.40	33.69	31.33	26.00	28.75	20.08
average	39.35	31.78	31.16	23.35	27.51	20.80
Total average	34.10			23.89		

About the authors



Bin Yan is a research librarian/professor and a postgraduate instructor and a former chief librarian of NUPT. He was graduated from Peking University and has engaged in library technical and management theory research for over twenty years. He has published more than forty theses on contemporary information management and information system design in domestic and abroad academic journals. His main study area is the theory and application of modern information management technology. He is a vice-director of the Modern Technology Committee of Academic Library of Jiangsu Province, China.

Feng Shi is an associate professor and works in education science college of Shandong Normal University, China. She was graduated from a speciality of chemistry education of Qufu Normal University in 1982. She has published a decade articles on contemporary management and information system in domestic academic journals. Her main study area is the theory and application of information physics.