

Digital Library Network Storage System

Liu Hongqing¹, Liu Yan²

¹Hunan Vocational College of Modern Logistics, Changsha, Hunan, 410131, China

²Hunan Mechanical & Electrical Polytechnic, Changsha, Hunan, 410151, China

Keywords: library, communication protocol, data, network storage

Abstract: As an online information resource management model to meet the information needs of the knowledge economy era, the digital library has become an important symbol to evaluate the technical level of a library. On the one hand, digital libraries use storage technology to achieve their functions. On the other hand, the development of communication protocol has promoted the development of network storage. The implementation of the paper is simple, which solves the problem of high speed transmission and remote backup of the large files of the digital library. It provides a practical method and specific process for the implementation of the iSCSI protocol. The iSCSI protocol has a great technical advantage in the digital library data transmission protocol.

1. Introduction

Digital library is a brand-new science and technology, and it is also a new social undertaking. In short, it is a digitalized information resource with multiple media contents, which can provide users with convenient and fast information service. Although it is called "library", it is not a library entity: it corresponds to the practical social activities of the management and dissemination of various public information, and is represented as a variety of new information resources organizations and information dissemination services. It uses the model of library resources organization and the help of computer network communication and other high-tech technologies. It aims at universal access to human knowledge. It creatively uses knowledge classification and accurate retrieval means to effectively carry out information ordering, so that people can obtain information consumption without space restrictions, and to a large extent it is not restricted by time.

Digital library is not a single entity. Digital library needs technology to connect more information resources of more digital libraries and information service institutions. The connection between various digital libraries and information service institutions is transparent to end-users. The goal is to realize the global inquiry of the digital library and the information service organization. The collection of the digital library is not limited to the literature substitutes: they are expanded into digital products and can not be replaced or disseminated in print format. Digital library is a digital information resource system which is supported by modern high and new technology. It is the management mode of the next generation of Internet information resources. It is an easy to use and large scale knowledge center without time and space restrictions. Generally speaking, the digital library is a virtual and non-walled library. It is an extensible knowledge network system based on the network environment. It is a large scale, distributed, easy to use, without time and space constraints, and can achieve links and intelligent inspection of the knowledge center. The digital library is not only a complete knowledge positioning system, but also an information management model for the future Internet development. It can be widely used in the public information dissemination of all social organizations such as social culture, lifelong education, mass media, commercial consulting, electronic government and so on.

2. DAS architecture

The 2 DAS system DAS is the abbreviation of Direct Attached Storage, which refers to "direct additional storage", which connects external storage devices directly to a server through a connection cable. The server structure that uses the direct additional storage scheme is like the PC

architecture. The external data storage device uses SCSI technology or FC (fiber channel) technology to connect directly to the internal bus. The data storage is part of the entire server structure, and in this case, the data and operating system are often unseparated. So DAS generally adopts SCSI protocol or FC protocol. The direct connection of DAS can solve the storage space expansion and high performance transmission requirements of single server, and the capacity of single external storage system has been developed to a number of TB. With the launch of large capacity hard disk, the capacity of single external storage system will increase. In addition, DAS can also constitute a dual machine high availability system based on raid, which meets the high availability requirement of data storage. Structure.

2.1 Storage Device

The direct additional storage equipment relies on the server host operating system for the I/O reading and writing and storage maintenance management. The data backup and recovery require the server host resources. The data flow needs to reflow the host to the tape machine (Library) connected to the server, and the data backup usually takes up the server host resource 20-30%, so the data backup usually takes up the server host resource. Many enterprise users' daily data backup often occurs at midnight or when the business system is not busy, so as not to affect the operation of the normal business system.

In addition, the connection channels between the direct additional storage equipment and the server host mostly use SCSI connection, and the bandwidth is 10MB/s, 20MB/s, 40MB/s, 80MB/s and so on. With the increasing processing ability of the server CPU, the storage hard disk space is more and more large, the number of hard disks of the array is more and more, and the SCSI channel will become the bottleneck of I/O. Under the condition of enough network bandwidth, the server itself becomes the bottleneck of data input and output. Now it is gradually unable to meet users' needs, no longer used for everyone.

For DAS, this is the simplest way of direct connection. So the protocols it supports are: SCSI protocol, NFS protocol and CIFS protocol. The main function of the SCSI protocol is to transmit commands, states and block data between the host and the storage devices. It supports the contents of multiple drives, queues, multitasking, caching, automatic driver ID recognition, and bidirectional interface operation. The role of the NFS protocol is to make the computer system access the directories and files of other computer systems through the network, as if these files were stored on the local hard disk. The CIFS protocol is similar to the NFS protocol. It is also the service that the implementation can access the files on the remote Internet computer and request the computer. The system environment they apply is different, the former is applied to the Linux/Unix environment by the Windows.

2.2 Storage Architecture

The network additional storage NAS (Network Attached Storage) is a dedicated technology for providing high performance, low cost, and high reliability for data preservation and transmission. The NAS device is designed to provide a set of secure, stable files and data preservation, easy to use and manage. It is defined as a special independent dedicated data storage server, embedded system software, and can provide NFS and SMB/CIFS file sharing. NFS and SMB/CIFS are all file sharing based on operation system. NAS is a file level data storage based on IP protocol, supporting existing network technologies, such as Ethernet, FDDI and so on. The NAS device completely takes data as the center, separates the storage device from the server completely, centrally manages the data, thus effectively releases the bandwidth, greatly improves the overall network performance, and can effectively reduce the total cost of ownership and protect the user's investment. Storing files in the same server allows different computer users to share and assemble different kinds of computers in the network, which is the main function of NAS network storage. Because of the open, industrial standard protocol of the NAS network storage system, different types of computer users run different operating systems to achieve access to the same file. So no longer care about Windows users or Unix users. They can also safely and reliably use data in NAS network storage system.

3. Storage Protocol Structure

Like any protocol, iSCSI has a clear hierarchy. According to the OSI model, the iSCSI protocol stack can be divided into five layers from top to bottom, as shown in Figure 1.

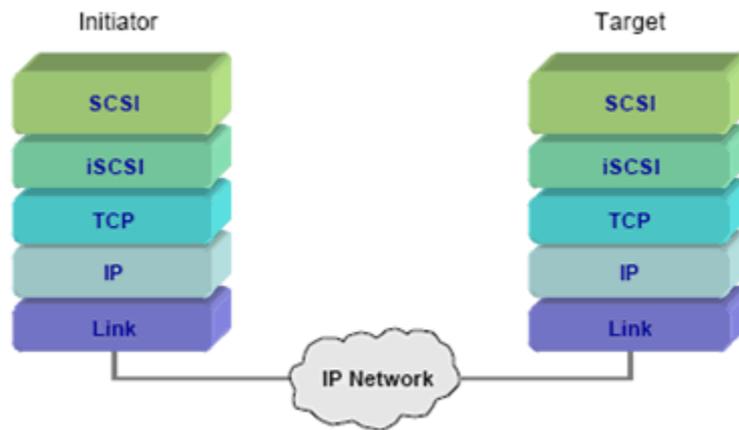


Figure 1 Graph 1 iSCSI protocol layered model

(1) SCSI layer: establish the SCSI CDB (command description block) according to the request sent by the application and pass it to the iSCSI layer; and accept the CDB from the iSCSI layer, and return the data to the application.

(2) iSCSI layer: SCSI CDB is encapsulated so that it can transmit on TCP/IP based network and complete SCSI to TCP/IP protocol mapping. This layer is the core of the iSCSI protocol.

(3) TCP layer: providing transparent and reliable transmission from end to end.

(4) IP layer: routing and forwarding to the IP message.

(5) Link level: providing point-to-point error free transmission.

Traffic control means the sender controls the speed of sending data frames to the network. The sending rate of a sender is a function of the size of buffers available on switches, routers or receivers. ISCSI adopts the end-to-end traffic control mechanism of TCP / IP protocol, and controls traffic with variable sending window. The sending window is negotiated by both parties when the connection is established, but in the process of communication, the recipient can dynamically adjust its own receiving window (can be enlarged or reduced) at any time according to the use of its own resources, and then tell the other party that the sender's sending window is consistent with its own receiving window. ISCSI adopts TCP's adaptive timeout retransmission algorithm, which can be dynamically adjusted according to the network situation. This algorithm records the time issued by each message segment and the time to receive the corresponding confirmation message segment. The difference between the two times is the round-trip delay RTT of the message section. When the data segment is sent, the corresponding timer is started. If the timer has been confirmed that the message section has not arrived, it triggers the data distribution mechanism. If the timeout is confirmed, a new round trip time delay is recorded, and the average round-trip delay RTT of the new message segment is obtained by weighted average of the back and forth time delay samples of each message segment. It is obvious that the retransmission time of the timer setting should be greater than the average round-trip delay RTT. In practical applications, the algorithm of RTT is still very complex. At present, Karm algorithm is commonly used.

4. Summary

The supporting technology of digital library mainly includes three aspects: information processing, information storage and information transmission. First, the digital library must be built on the basis of high performance computing, so that it can handle massive data information, especially the information such as voice and video. Second, the digital library needs high density storage capacity to accommodate mass data, and third, the digital library should also have the

ability of high speed transmission. Because the main information transmission channel of the digital library is the network, it is necessary to have a high-speed information transmission network for users to get the required information as soon as possible. Among them, the information storage technology is the core.

References

- [1] Arunkumar, N., Venkataraman, V., Thivyashree, Lavanya A moving window approximate entropy based neural network for detecting the onset of epileptic seizures (2013) International Journal of Applied Engineering Research, 8 (15), pp. 1841-1847.
- [2] Yijiu Zhao, Li Wang, Houjun Wang, and Changjian Liu, Minimum Rate Sampling and Spectrum Blind Reconstruction in Random Equivalent Sampling. Circuits Systems and Signal Processing, 2015; vol. 34, no. 8, 2667-2680.
- [3] Fernandes, S.L., Gurupur, V.P., Sunder, N.R., Arunkumar, N., Kadry, S.A novel nonintrusive decision support approach for heart rate measurement (2017) Pattern Recognition Letters. <https://doi.org/10.1016/j.patrec.2017.07.002>
- [4] Arunkumar, N., Ramkumar, K., Venkatraman, V., Abdulhay, E., Fernandes, S. L., Kadry, S., & Segal, S. (2017). Classification of focal and non focal EEG using entropies. Pattern Recognition Letters, vol. 94, pp. 112-117
- [5] Arunkumar, N., Ram Kumar, K., Venkataraman, V. Automatic detection of epileptic seizures using permutation entropy, Tsallis entropy and Kolmogorov complexity (2016) Journal of Medical Imaging and Health Informatics, 6 (2), pp. 526-531.