

## T-ARCHIVE: A NOVEL HSM-BASED DATA ARCHIVE SYSTEM

Mingqing Zheng<sup>1\*</sup>, Weiwei Xiong<sup>2</sup>, Dapeng Ju<sup>3</sup> and Dongsheng Wang<sup>4</sup>

<sup>1</sup>Dept. of Computer Science & Technology, Tsinghua University, Beijing, China  
Email: [zhengmq@gmail.com](mailto:zhengmq@gmail.com)

<sup>2</sup>Dept. of Computer Science & Technology, Tsinghua University, Beijing, China  
Email: [xiongww@gmail.com](mailto:xiongww@gmail.com)

<sup>3</sup>Dept. of Computer Science & Technology, Tsinghua University, Beijing, China  
Email: [judapeng@tsinghua.edu.cn](mailto:judapeng@tsinghua.edu.cn)

<sup>4</sup>Dept. of Computer Science & Technology, Tsinghua University, Beijing, China  
Email: [wds@tsinghua.edu.cn](mailto:wds@tsinghua.edu.cn)

### ABSTRACT

Rapid increases of user data from terabytes to petabytes have created new challenges in data archiving. Modern data archive systems require higher adaptivity, reliability, and performance than traditional data archive systems can provide. Recently Hierarchical Storage Management (HSM) has been applied to a data archive that stores data in a multi-level storage system according to access frequency. In this paper, we describe the design and implementation of a novel HSM-based data archive system called T-Archive, which can meet the above requirements for order-of-magnitude scaling of storage.

**Keywords:** Data archive, File migration, Hierarchical Storage Management, Metadata, File redirector

## 1 INTRODUCTION

The rapid increase of user data has brought new challenges to data storage. Research shows that 90% of all files are not used after initial creation and those that are used are normally short-lived (Gibson, Miller, & Long, 1998). These files could be archived after not being in use for a specified time. Such files may not be frequently requested, but once needed, they should be accessed immediately (Quinlan & Dorward, 2002). In that case, we must find an effective solution to access the archived files. Recently, Hierarchical Storage Management (HSM) has been applied to data archive systems. A traditional HSM-based data archive system maintains data through a three-level management system. The first level is online data, which contain the most frequently accessed data stored in a high-speed SCSI disk array. The second level is near-line data, which has infrequently accessed data stored in a slower storage platform such as an MO jukebox. The third level is offline data stored in a tape library (Lugar, 2001).

A traditional HSM-based data archive system has serious performance problems when a user requests access to a data file that has been migrated to offline storage. It is quite time-consuming to search the archived information and send a message to the system administrator to mount the requested tape for file access. To achieve high performance, the system must increase the online storage capacity. However, high-speed SCSI disks array are very expensive.

Users accessing data in an HSM system by the "classical" means of FTP or NFS will be affected by the migration because the path-names of the files change. The situation is even worse for application software where a concept of data access based on path-names has been implemented (Reuter, 1999).

In this paper, we discuss the design and implementation of a novel HSM-based data archive system named the T-Archive. It aims to reduce cost but at the same time achieve high performance. Moreover, it can keep the access process of archived data transparent to end users.

This paper is organized as follows. Section 2 gives a detailed description of the proposed T-Archive architecture and its components. Section 3 analyzes the features of the proposed system and compares it with other data archive systems. Section 4 concludes the paper.

## 2 ARCHITECTURE AND COMPONENTS

### 2.1 ARCHITECTURE

In the proposed T-Archive, we discard levels two and three of the traditional HSM and subdivide level one, which is online storage, into two sub-levels. As shown in Figure 1, the proposed level one is a high-speed SCSI disk array that stores the most frequently accessed data. A relatively slow and inexpensive SATA or IDE disk array is used as the proposed level two for storing the less frequently accessed data. Compared with the traditional HSM, the T-Archive can store more data in online storage with marginal overhead cost. Thus, higher performance can be achieved with less time-consuming offline data access.

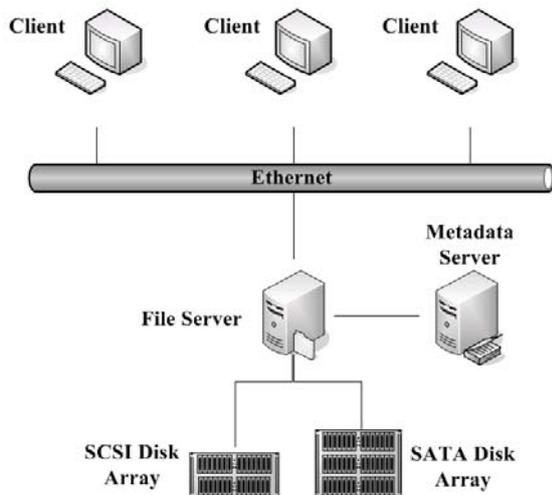


Figure 1. Architectural of the T-Archive

### 2.2 COMPONENTS

As shown in Figure 2, there are three core components in the T-Archive: the Archive Service, the Metadata Server, and the File Redirector. Clients can access data files through HTTP, CIFS, FTP, etc., which are supported by the T-Archive. When a client requests to read/write a data file, the File Redirector will check whether the file has migrated to the slower storage devices and then return the correct file handle to the file system. The Archive Service takes charge of the migration between level one and level two storage devices according to archive policies. The Metadata Server maintains the metadata, which are the file attribute information (e.g. the last access time and the access frequency). The metadata is indispensable for file migration and location.

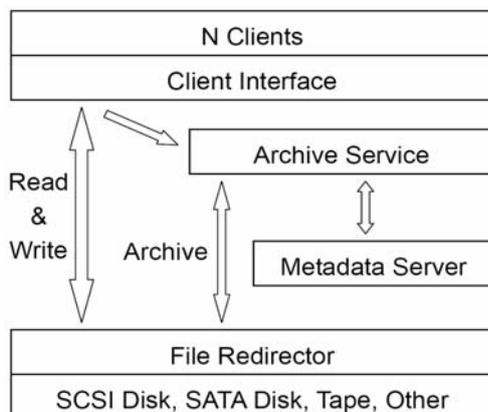


Figure 2. Components of the T-Archive

### **2.2.1 Archive Service**

The Archive Service provides a web GUI for the system administrator to configure, monitor, and control the T-Archive system. New storage devices can be added to increase the capacity of level two storage at any time if needed. The Archive Service starts the archive process automatically at a set time. It can also be started manually by the system administrator. All the archive information, such as the migrated files and the time when they are migrated, is recorded into the log file. If there is anything wrong, the system administrator can view the log file to correct the error.

If a file meets the archive requirement, it is migrated from level one storage to level two storage. After it is migrated, a stub file with the same name is created at the original location. The stub file records a flag of the stub file and the true location of the archived file.

### **2.2.2 Metadata Server**

The Metadata Server stores the metadata information of all data files. The metadata information includes the created time, the last-accessed time, the last-modified time, the access frequencies, and the true file location after migration. The T-Archive provides three access frequency thresholds for archive policy. They are the access frequencies of one month, three months, and one year.

To get accurate data for these three frequencies, we must record all data requests for one year. However, this is inefficient and impractical. As a result, the T-Archive makes use of the following approximate algorithm to calculate the three frequencies.

Assume the everyday accesses are the same in the last  $n$  days ( $n > 0$ );  $A$  is the number of times accessed today;  $S_n$  is the sum of the number of times accessed in the last  $n$  days;  $S'_n$  is the sum of the number of times accessed in the last  $n-1$  days and today. Then

$$S'_n = S_n - \frac{S_n}{n} + A = \frac{n-1}{n} * S_n + A \quad (1)$$

To calculate the number of times accessed in the last  $n$  days, we only need to record two data points. One is the sum of the times accessed in the last  $n$  days and the other is the times accessed today.

### **2.2.3 File Redirector**

The File Redirector monitors and deals with all I/O requests sent from clients to the file system. Before the file system responds to an I/O request, the File Redirector intercepts it and checks whether the file requested is a stub file. If the file is a stub file, it has been migrated to level two storage. After reading the true location of the file from the stub file, the File Redirector will translate the I/O request to that for the true file and return the real file handle to the file system. Then the file system can read or write the file requested. All of that process is transparent to the client.

## **3 FEATURES AND COMPARISON WITH OTHER DATA ARCHIVE SYSTEMS**

### **3.1 Features**

#### **3.1.1 Usability**

To provide better usability, we have added the File Redirector to realize the migration and retrieval of data file and keep the details transparent to clients. Users can use the archive system as easily as using the normal file systems. The system administrator can login into the remote system and use HTTPS protocol to control the system.

### **3.1.2 Reliability**

For reliability, the T-Archive separates metadata from data storage. This is more reliable as the metadata server can only be accessed through an authorized message interface and does not run any user programs (Watson, 2005). A redundant metadata server can be used to avoid being a single failure point.

### **3.1.3 Adaptability**

The T-Archive provides a uniform interface to support multiple standard protocols to achieve high adaptivity. Heterogeneous clients using different file systems (e.g. NFS, CIFS) and different file transport protocols (e.g. TCP/IP, IPX) can access data simultaneously. The T-Archive is not dependent on underlying storage architecture. DAS, NAS and SAN storage devices are all supported.

## **3.2 Comparison with other data archive systems**

Compared with previous HSM-based data archive systems, the proposed T-Archive subdivides online storage into two levels to increase the capacity of online storage and decrease the latency of accessing archived data. Using cheap SATA disks as level two storage, the T-Archive reduces by more than 50% the cost of the storage system while achieving almost the same performance.

Moreover, all the operations of migration and retrieval process are transparent to end users. This system is more convenient for both the system administrator and end users to use or maintain the data archive system.

## **4 CONCLUSION**

An HSM-based data archive system called T-Archive is proposed to reduce cost and increase performance. By separating the metadata server from the data file server and being transparent to users, the T-Archive can provide higher usability, reliability, and adaptability than previous HSM-based data archive systems. The proposed T-Archive has already been successfully used in a railway information system.

## **5 REFERENCES**

- Gibson, T., Miller, E., & Long, D. (1998) Long-Term File System Activity and Interference Patterns. *Proceedings of the 23<sup>rd</sup> Annual International Conference on Computer Measurement and Performance (CMG '98)*, Anaheim, CA, pp 976–987.
- Lugar, J. (2001) Hierarchical Storage Management: Leveraging New Capabilities. *IT Professional*, March/April 2001, pp 53-55.
- Quinlan, S. & Dorward, S. (2002) Venti: a new approach to archival storage. *Proceedings of the FAST 2002 Conference on File and Storage Technologies*, Monterey, California, USA
- Reuter, H. (1999) Hiding HSM systems from the user. *Proc. Sixteenth IEEE Symposium on Mass Storage Systems*, San Diego, March 1999.
- Watson, R. (2005) High Performance Storage System Scalability: Architecture, Implementation and Experience, *Proceedings of the 22nd IEEE / 13th NASA Goddard Conference on Mass Storage Systems and Technologies (MSST'05)*, April 2005.