

Mass Data Archiving Acceleration Technology for Magnetic Tape Storage

Taketoshi Yoshida

Ken Iizawa

Masahisa Tamura

Traditionally, magnetic tape storage has been used mainly for backup purposes. Owing to its large capacity and low cost, as well as recent increases in transfer speed and the spread of the linear tape file system (LTFS), which is a file system for tape that enables the handling of magnetic tape data in file units, magnetic tape storage is now expected to find increasing use for archival purposes. However, while magnetic tape storage allows fast sequential accessing of continuous areas on tape, random accessing of discontinuous locations was slow and had to be improved for magnetic tape storage to become usable for archiving large amounts of data. In response, Fujitsu Laboratories expanded the functionality of LTFS and developed methods for the virtual integration of multiple tape cartridges, data management according to tape characteristics, and access order control to boost the random access performance of magnetic tape. This article describes the mass data archiving acceleration technology for magnetic tape storage developed by Fujitsu Laboratories.

1. Introduction

In recent years, the use of big data and AI has resulted in the diversification of the formats (video, images, etc.) and sizes of the data handled as well as increases in the volume of data being handled. Indeed, the total volume of data handled has reached the petabyte class. For the management of such data, object storage, which is very convenient due to the absence of limitations on the size and number of stored data, is widely used. Object storage is provided by many cloud providers, and includes S3, offered by Amazon Web Services (AWS), and Azure Blob Storage, offered by Microsoft. Furthermore, there is a growing need to continue to store large amounts of data over long periods ranging from a few years to more than 10 years for use as evidence and for the analysis of past and future trends [1]. As a result, high-capacity archival storage with a low storage cost per unit and low access performance, such as AWS's S3 Glacier [2] and Deep Archive, and Microsoft Azure's Azure Archive Storage [3], among others, is attracting much interest [4].

However, though their cost of storage is low, cloud-based archival storage services can rack up high charges for data retrieval. Going forward, with the

progress of digital transformation (DX) by companies, the utilization of large amounts of data generated at various sites will shift into high gear and further data volume increases are expected. As an archive storage method for such large amounts of data, low-cost and large-capacity magnetic tape storage is receiving renewed attention.

Against this backdrop, Fujitsu Laboratories has developed a technology to speed up the accessing of magnetic tape storage for mass data archiving. This article describes the virtually integrated file system and its acceleration technology, developed by Fujitsu Laboratories, as well as its performance evaluation.

2. Overview of magnetic tape storage

In the utilization of large amounts of data, storage capable of holding large amounts of data inexpensively and for long periods is desired to allow analysis of changes over time and relationships among data in different fields. For such applications, magnetic tape storage presents a number of advantages over other storage media. For example, the amount of power required to hold the same size of data is a few hundredths that for a hard disk drive (HDD) [5].

Moreover, the linear tape open (LTO) standard [6] has been created, and tape capacity has been roughly doubling every two to three years. **Table 1** lists the tape capacities and transfer rates of the various generations of the LTO standard.

Software called linear tape file system (LTFS) is used to allow accessing of tape storage from applications. LTFS makes it possible for the data on a tape to be handled in file units like in the case of HDDs, and has been supported by the LTO standard since LTO5. The LTFS format and application programming interface (API) specifications were released in 2010 and became a de facto standard. They were formally established as an international standard (ISO/IEC 20919:2016) in April 2016 [5].

The LTFS format, which uses tape partitioning, consists of two partitions: an index partition and a data partition. The index information, which consists of the file names of the data, the date and time of file operations (write, delete, update), the recording positions on tape, and so on, is recorded in the index partition and is managed in a small area at the beginning of the tape [5]. The data partition is a large area where the actual data is recorded. The data is written using the write-once method, and the index information is written regularly.

Data deletion takes place at the index information level only, and the data on the data partition is not deleted. Therefore, the area on the tape where the deleted data existed cannot be reused without reformatting the cartridge. Further, when updating data, new data is added to the data partition and the index information is updated. Owing to these characteristics, tape storage is used mainly to store data that is updated with low frequency rather than data that is updated frequently.

3. Issues with conventional technology

This section describes two issues with the conventional LTFS.

1) Degradation of random read performance

Magnetic tape boasts high continuous data access performance, which reaches 360 MB/s in the latest iteration of the magnetic tape standard, LTO8. On the other hand, random read access performance for data at non-continuous locations is poor. For example, random reading of data in units of several MB results in less than one-tenth the performance of continuous reading.

Further, as the size of the data to be accessed becomes smaller, the amount of processing for aligning the magnetic head when reading data increases, resulting in degradation of access performance. For instance, when numerous small data to be accessed are intermittently recorded on magnetic tape with unnecessary data in between, intermittent reading in the direction of travel of the magnetic tape results in greatly degraded access performance.

When using magnetic tape storage for backup purposes, this is not a problem because continuous access of consecutive data is made. However, if magnetic tape is to be used for archiving, in which required data units must be accessed at the required timing, it is necessary to improve read/write performance for various data sizes and random read performance.

2) Degradation of write performance due to increase in the number of files

As mentioned in section 2, LTFS holds index information for each file. When using magnetic tape for archiving, users save files of various sizes, but if a large number of small files are written, the index information also increases, causing a drop in writing performance.

Table 1
LTO standard generations.

Year	2010	2012	2015	2017	2019 onward		
Standards	LTO5	LTO6	LTO7	LTO8	LTO9	...	LTO12
Capacity (uncompressed)	1.5 TB	2.5 TB	6 TB	12 TB	24 TB	...	192 TB
Transfer rate	140 MB/s	160 MB/s	300 MB/s	300 MB/s (HH) 360 MB/s (FH)	—	...	—

HH: Half-Height
FH: Full-Height
(FH, HH are tape drive thickness standard.)

4. Developed technology

To solve the problems described in the previous section, Fujitsu Laboratories developed a file system that virtually integrates multiple tape cartridges on top of LTFS and magnetic tape access acceleration technology (**Figure 1**). This virtually integrated file system consolidates multiple tape cartridges into one, allowing users to access the files they need without having to be aware of discrete tape cartridges. This section describes the magnetic tape access acceleration technology developed by Fujitsu Laboratories.

1) Access order control technology considering physical location

On magnetic tape, the width of the tape is divided into narrow areas called wraps, and when data has been written to the end of a wrap, data writing continues in the opposite direction onto the next wrap. The location of data when accessing data while alternating between wraps is managed by abstracted

logical addresses, separate from the physical location (physical address) on the tape. When comparing the differences between the logical address and the physical address of two data, the logical addresses may be far apart, but the physical addresses may be close to each other (**Figure 2**).

Therefore, in the virtually integrated file system, we have developed a method that accepts multiple read requests and processes requests for data in the order of their physical addresses on the tape, rather than their logical addresses. In the example of Figure 2, in management by logical address, the data on the same wrap is read in the tape travel direction. By contrast, in management by physical address, data on a different wrap close to the current head position is read, which minimizes the head travel distance.

Further, when accessing magnetic tape, it takes time to align the head with the start position. Therefore, when a request to read two files close to each other on

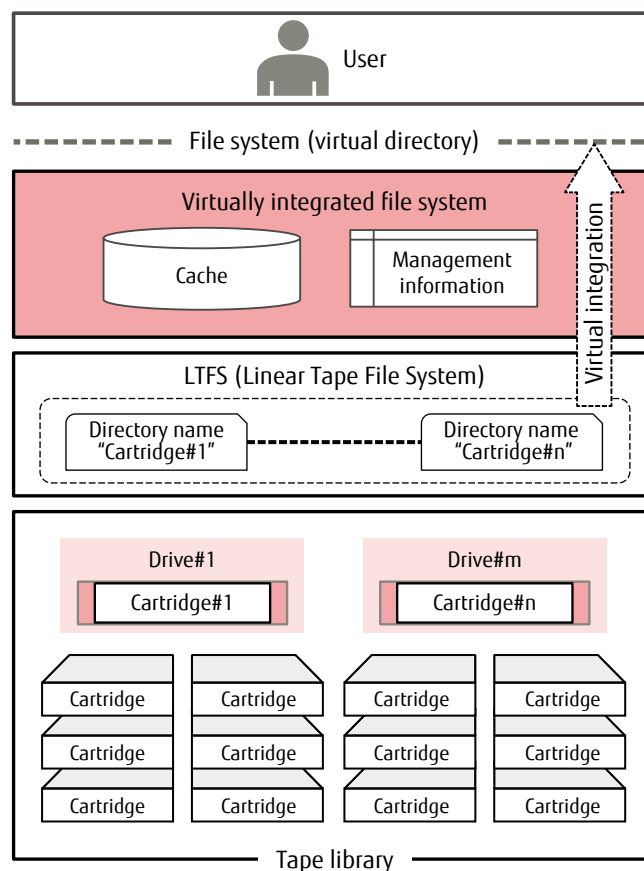


Figure 1
Image of configuration of developed system.

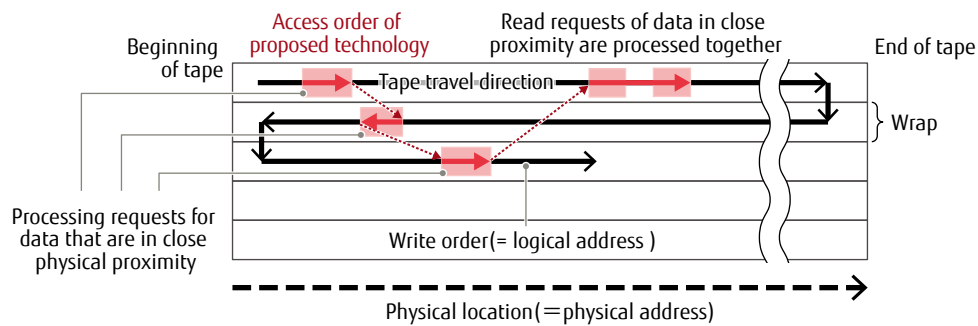


Figure 2
Access order control considering physical location.

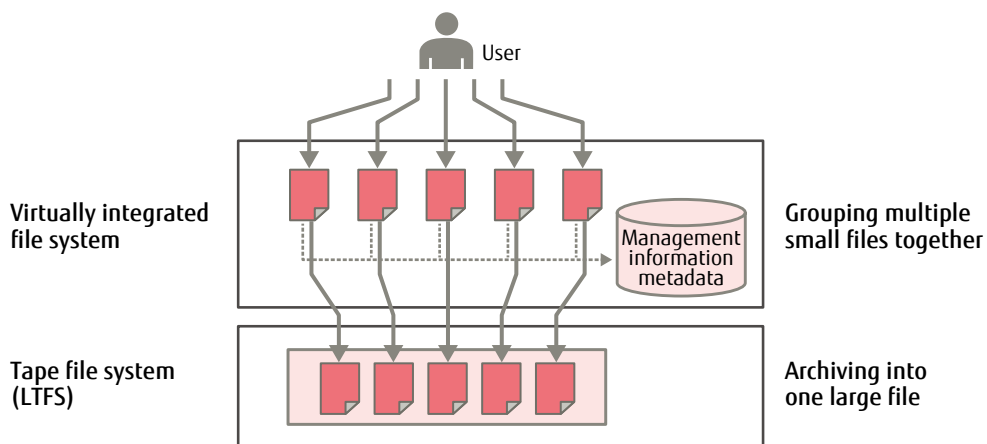


Figure 3
Multiple file aggregation function.

the same wrap is issued, each file is not read intermittently, but the group of files in between is also read and the files that are not needed are discarded to further improve speed.

2) Aggregation management technology for multiple files

We have developed a system for collectively storing small files that are less than a specified size on LTFS as large files while still displaying them as small files to users (**Figure 3**). This reduces the number of files managed by LTFS and reduces performance degradation.

5. Performance evaluation

This section describes the performance evaluation of this technology.

Using Ceph [7], which is open-source distributed storage software, we constructed two-tiered storage consisting of a storage layer composed of HDDs and

a storage layer composed of magnetic tape storage, and we evaluated the access performance achieved by using our newly developed technology.

1) Evaluation of access order control by physical location

We evaluated the time required to read 100 files at random from 50,000 files each 100 MB in size stored on magnetic tape. The readout time was confirmed to be 1,300 seconds, which is about one fourth of the 5,400 seconds required by the conventional method.

2) Evaluation of multiple file aggregation management

We evaluated the time required to move 256 1-MB files between tiered storage layers from HDD to magnetic tape. It took 2.5 seconds with the conventional method, but using the newly developed technology, we confirmed that the files could be moved in 1.3 seconds, which is about half the time required by the conventional method.

The newly developed technology makes it possible to speed up magnetic tape access such as random reading and writing of files of various sizes used for archiving. It is expected to make possible a data archiving infrastructure with an excellent price/performance ratio for long-term archiving of large amounts of data.

6. Conclusion

This article described the mass data archiving acceleration technology for magnetic tape storage developed by Fujitsu Laboratories. This technology will further the use of magnetic tape storage for archiving large amounts of data, which are increasing exponentially, by means of virtually integrated file system technology, access order control technology that takes into account physical location, and multiple file aggregation technology, and help bring about a data archiving platform with an excellent price/performance ratio.

Going forward, we will test this technology for business applications to improve its performance for various types of access.

All company and product names mentioned herein are trademarks or registered trademarks of their respective owners.

References and Notes

- [1] B. Kleyman: "Understanding Cold Storage: Best Practices around Google Nearline & Amazon Glacier," CloudBerry Lab, Aug. 2015.
- [2] Amazon S3 Glacier.
<https://aws.amazon.com/glacier/>
- [3] Microsoft: Azure Archive Storage.
<https://azure.microsoft.com/en-us/services/storage/archive/>
- [4] H. Fujimaki et al.: Big data analysis platform incorporating cold storage (in Japanese). FUJITSU, Vol. 66, No. 4, pp. 49–54 (2015).
<http://img.jp.fujitsu.com/downloads/jp/jmag/vol66-4/paper07.pdf>
- [5] Tape Storage Technical Committee, Japan Electronics and Information Technology Industries Association: Tape Storage Trends <2018 Edition>. July 2018 (in Japanese).
https://home.jeita.or.jp/upload_file/20180725112701_jYahZ50UKI.pdf
- [6] Ultrium LTO.
<https://www.lto.org/>
- [7] Ceph.
<https://ceph.io/>



Taketoshi Yoshida

Fujitsu Laboratories Ltd., ICT Systems Laboratory

Mr. Yoshida is currently engaged in research on data management systems.



Ken Iizawa

Fujitsu Laboratories Ltd., Platform Innovation Project

Mr. Iizawa is currently engaged in research on data management systems.



Masahisa Tamura

Fujitsu Laboratories Ltd., Platform Innovation Project

Mr. Tamura is currently engaged in research on data management systems.

This article first appeared in Fujitsu Technical Review, one of Fujitsu's technical information media. Please check out the other articles.

Fujitsu Technical Review

<https://www.fujitsu.com/global/technicalreview/>

