Original scientific paper

# Long-term Inactive Data Retention through



Ivan Vican
Metronet telekomunikacije d.d.
Ulica Grada Vukovara 269d, Zagreb, Croatia
ivan.vican1@zg.t-com.hr

Hrvoje Stančić
Department of Information Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
hrvoje.stancic@zg.t-com.hr

**Summary**

*Increasingly the need to retain digital documents indefinitely for legal, administrative or historical purposes is simply leading to a "save everything forever" approach. The authors argue that due to the technological reasons it is much easier to preserve large amount of documents in the electronic than in the paper form. Thus the selection procedures tend to be less restrictive than they used to be. Nevertheless, for most organizations it would be impossible to sustain this data growth forever. Archives, libraries, museums, institutions holding cultural heritage, as well as other companies and firms, are implementing solutions for creating digital archives, digital libraries, digital repositories and other types of storage systems aiming at long-term preservation of digital materials. Most of the data held in such systems are inactive for a long time, i.e. only a small set of data is frequently retrieved. Therefore, due to the specific needs of every organization, the storage planning process and the technology that is going to be used for storage and long-term preservation requires individual approach. The focus of this paper is on the retention of the long-term inactive data through tape storage technology. The authors will discuss current state of the art tape storage capabilities, and their advantages and disadvantages as a long-term storage and preservation solution.*

**Key words:** long-term preservation, storage systems, tape storage, archive, library, museum**,** data, electronic material

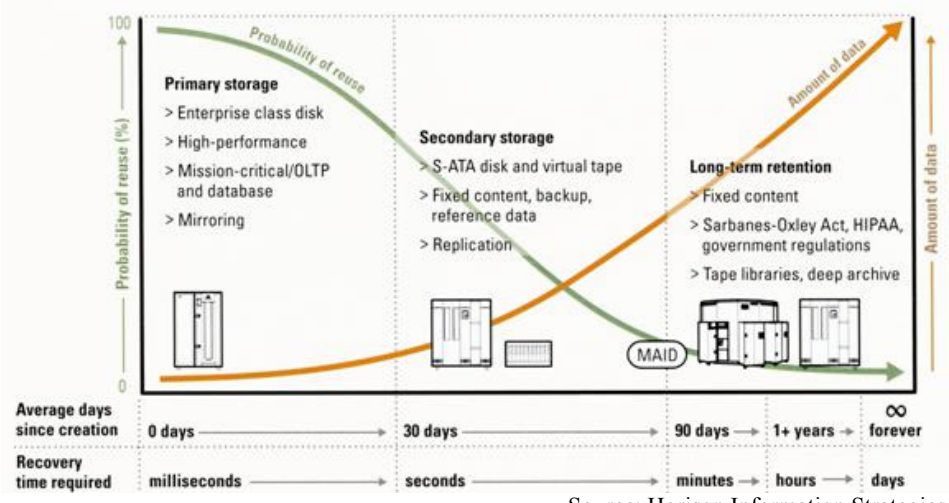
[*] The authors are solely responsible for the content of this paper. It does not represent the opinion of the institutions they work in, and those institutions are not responsible for any use that might be made of data appearing in the paper.

## Introduction

The need to retain expanding volume of digital data for regulatory, business, personal or heritage purposes is leading to a challenge on which technology data should be stored and archived. Capabilities of digital archives, libraries and repositories are multiple: faster and less complicated manipulation, concurrent access to resources, less space is required, access to high valued material… and they vary among the storage technologies. Other important variable that influences the choice of storage technology is frequency of reuse, i.e. how often data is needed.

Over the time, relevance of data is changing whereby its value is also changing. The implication is that most data will become inactive over the time and the need to access such data is declining. Data lifecycle is providing insight in data value fluctuation, which is measured by frequency in given time. This kind of approach to data lifecycle is specific to business enterprises. By expiration of retention period they will probably delete archived data (Graph 1). However, institutions such as archives, museums, libraries and institutions which focus lies on preservation of heritage need to retain data indefinitely. In other words data has a constant value. Specificity of each system implies individual approach when it comes to retention, preservation and archiving across different information systems and institutions. However, the need to archive and the storage technology that is being used are quite common.

Graph 1: Data reference over time




Source: Horison Information Strategies

As a possible answer to the rising needs of archiving digital data, tape storage technology is offered. Rapid pace of innovations in tape storage technology, especially during the last decade, is reviewing capabilities of long-term retention,

preservation and archiving through this technology. The intention of paper is to debate about advantages and disadvantages of tape storage technologies' capabilities when it comes to the long-term retention, preservation and archiving.

## Tape Storage Technology

First commercially available magnetic tape was introduced in 1952 with the capacity of 1.4 MB. Immediately after introduction, tape replaced punched cards to became the first real removable storage medium. From its very beginning, tape was connected with mainframe system in order to store bulk data. Modern usage of tape storage is mainly connected to backup and archiving systems. The first recording technology used in tape systems was linear recording technology which dominated until middle of 1980s. After that period helical record technology took primate. In the last ten years, both technologies are improved considerably. Leverage has turned to linear technology primarily because of the possibility of higher record density due to the development of the linear serpentine technology and faster data transfer rate.[1] Over the 50 years of development, the tape storage systems came out in numerous standards and formats. Prevailing standard nowadays is the linear serpentine recording technology on tape with half inch wide reel in single-hub cartridge. The half inch tape width is the most frequently used magnetic tape in history. The medium is produced with the metal particle technology and their variations like the advanced metal particle.

By the appearance of affordable disk and optical storage technologies, the magnetic tape storage was pushed down with the future not so clear. Over the last ten years, with the growing need for data storage space, tape storage is recognized as a medium that can bear with these growing challenges. This generated the explosion of tape storage technologies and formats. Among numerous formats and tape technologies, two are representing actual pinnacle in the development of magnetic tape storage: Enterprise-class tape and Linear Tape-Open formats.

### Tape Systems

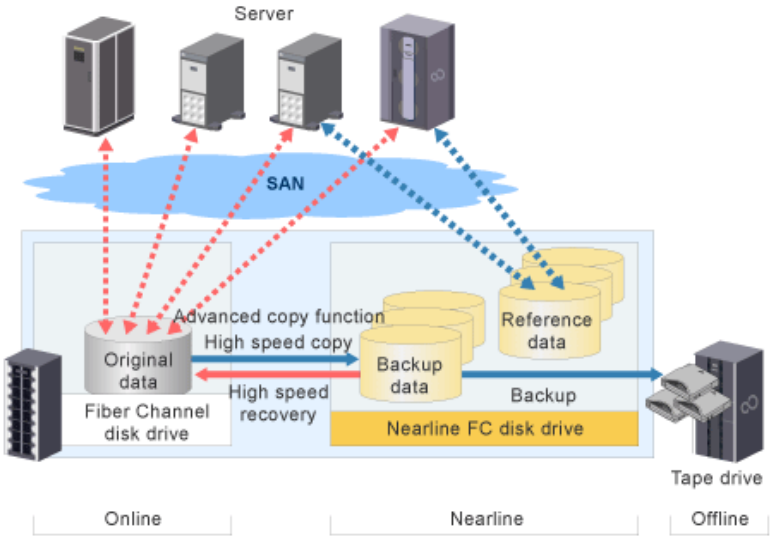
In order to utilize tape medium, proper devices are required. Such device goes by name tape system and it is divided in three types of systems: tape drive, autoloader and library. *Tape drive* represents a basic element of the system as it provides physical and logical structure for reading and writing processes.[2] It allows connection with other devices via SCSI, SAS and Fiber Channel network technologies.

[1] See: Haeusser, Babette; Kessel, Wolfgang; Silvestri, Mauro; Villalobos, Claudio; Zhu, Chen. IBM System Storage Tape Library Guide for Open Systems // IBM Redbook, Seventh Edition, 2008. http://www.redbooks.ibm.com/redbooks/pdfs/sg245946.pdf (last access: 18 August 2009).
[2] Ibid.

*Tape autoloader* consists of a tape drive and an automated tape cartridge exchange system with up to ten tape cartridges in the housing. With added automation feature, the autoloader becomes an autonomous tape drive which does not require constant human intervention in order to exchange tape cartridges.
*Tape library* is able to meet the most demanding archiving needs and because of that it is the most complicated tape system. Such systems have two or more tape drives, depending on the quantity of tape cartridges which can rise up to few thousands. The library layout permits simultaneous access to multiple tape cartridges. The exchange of cartridges is operated by a robotic mechanism and it takes only few seconds to exchange tapes.

**Tiered Storage: Position of Tape Storage**
In the traditional information system tape storage is classified as an offline (archival) tier, as opposed to the disk systems which are online (primary) or near-online (secondary) storage (Picture 1).[3] However, thanks to tape libraries, tape storage is increasingly seen as near-online tier while tape drive and tape autoloader are considered as offline tier.

Picture 1: Tape in a network storage environment




Source: Fujitsu Corporation

Hierarchy of storage classes is enabling consolidation, scalability and faster work of an information system. Storage classes are defined according to the re-

[3] See: Brooks, Charlotte; Byrne, Frank; Higuera, Leonardo; Krax, Carsten; Kuo, John. Redbook: IBM System Storage Solutions Handbook // IBM Redbook, Seventh Edition, 2006. http://www.redbooks.ibm.com/redbooks/pdfs/sg245250.pdf (last access: 20 August 2009).

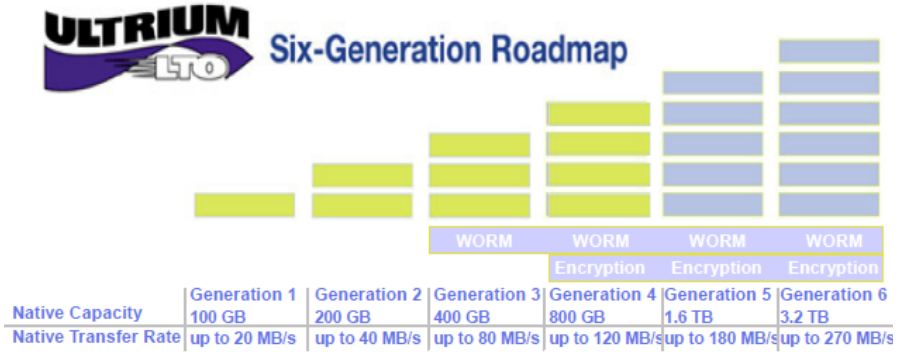
trieval speed, therefore depending to the storage technology. For example, if the data is being accessed on daily basis it will be stored at the primary disk storage tier. Predefined data policy, with the help of storage and archival software, are automating routing processes towards the designated storage device cutting down the load on network and servers. In addition, the storage area network (SAN) technology is enabling direct connection of storage devices with computer systems or with other storage devices. Thereby, it is possible to move data between tiers without the server intervention.

## Linear Tape-Open and Enterprise Tape Storage Technology

Also known as LTO, it was developed at the end of 1990s by LTO Consortium. Main goal was to define and to manufacture the first open format that will offer high-capacity, high performance of tape storage devices to midrange IT systems. The standard format of LTO is known as Ultrium.[4] From 2002 until today, LTO Ultrium is most commonly used tape ever.[5] The reason for that can be found in innovative technology and accessibility.

LTO Ultrium format has defined Six Generation roadmap for growth and scalability (Picture 2). The roadmap represents goals and there is no guarantee that these goals will be achieved. However, each available generation was released with doubled performance and capacity. The latest available generation is LTO-4 released in 2007 with native capacity of 800GB, 120MB/s of native data transfer rate and data encryption at device level. LTO-5 is coming out in 2009.

Picture 2: LTO Generations




Source: LTO Program

Data compression (DC) techniques are quite common in tape storage. LTO-DC is called Streaming Lossless Data Compression (SDLC) and it is able to pass

[4] Ibid.

[5] See: LTO Ultrium format reaches new heights with over 100 million cartridges shipped // LTO, 2008. http://lto.org/pdf/LTO%20100%20Million%20Cartridge%20Milestone.pdf (last access: 20 August 2009).

through already compressed data such as JPEG, MPEG and MP3.[6] LTO-DC algorithm is able to achieve 2:1 compression which gives LTO-4 1.6 TB of compressed capacity and 240MB/s of compressed data transfer rate (Table 1).

Table 1: LTO Tape drive specifications

| | Data transfer rate | Data transfer compressed | Native capacity | Compressed capacity | MTBF |
|---|---|---|---|---|---|
| LTO-4 | 120 MB/s | 240 MB/s | 800 GB | 1.6 TB | 250,000 hr |

Source: LTO Program

Write Once Read Many (WORM) capability was introduced in the third generation. WORM format is designed for long-term and temper-resistant data retention, which is most useful for legal regulations. This is achieved via Cartridge memory chip which holds information about specific cartridge, media in that cartridge and the data on that cartridge.[7]

Compatibility issues are common in tape generations. LTO is designed for backward compatibility for two generations according to the following rules: read/write compatible with one generation prior, read only compatible with two generation prior. For example, LTO-4 is able to perform read/write on LTO-3 and to read from LTO-2 generation. However, it is not possible for LTO-4 to expand the capacity of LTO-3.[8]

In the past, reliability was the weakest point of a tape storage technology. Tape suffered from incorrectly written data, jammed heads and short life period because of mechanical wear out. These drawbacks were solved with the following technical features: read after write verification, surface control guiding mechanism for less damage to tape, error detection/correction for data integrity, magneto-resistive head, large internal data buffer, automated cleaning system and speed matching towards host adapter.[9] Anyway, tapes should be checked once a year for medium deterioration. In case of possible data loss due to deterioration, data should be refreshed, i.e. moved to a new tape. LTO drives are automatically checking tape deterioration every time a tape is mounted.

Advancement in reliability has positively affected the availability and the predicted durability (Table 2) of tape medium. However, the tape wares off after repeated read/write operations which as an effect can have increase number of errors at tape recorded data. The LTO tape cartridge is made for 5,000 load/unload cycles.[10] With the appropriate handling and average usage of four times a week it can last approximately 30 years. This applies only to read op-

[6] See: IBM System Storage Tape Library Guide.
[7] Ibid.
[8] Ibid.
[9] Ibid.
[10] See: Sun StorageTek Linear Tape Open (LTO) Ultrium Data Cartridges // Sun Microsystems. http://www.sun.com/storage/tape_storage/tape_media/lto/specs. xml (last access: 22 July 2009).

erations. If a tape is rewritten in full once a month it will last for approximately 17 years.

Table 2: LTO-4 Tape cartridge reliability

| | Full file passes | Media durability | Archive life |
|---|---|---|---|
| LTO-4 | 260 | 5,000 load/unload cycles | Up to 30 years |

Source: LTO Program

**Enterprise Tape Storage Technology**

The start of a modern enterprise tape storage technology is dated in the first years of 1980s. The technology was primary developed for the needs of mainframe systems. [11] Today, they are still the most common tape technology attached to the mainframe systems with added interoperability towards open platforms as well. At first glance, the enterprise tape storage technology and LTO are quite similar. LTO has succeeded many technical features from the enterprise tape storage technology. WORM capability was first introduced in this technology. Differences can be found in generations of the same technical features. For example, larger data buffer and cartridge memory can be used. When it comes to mechanical components, tape drive and tape cartridge are more robust (Table 3) then LTO. The reason for that rests within the enterprise tape storage working environment.

Table 3: Enterprise class Tape cartridge reliability

| | Full file passes | Media durability | Archive life |
|---|---|---|---|
| T10000B | 360 | 15,000 load/unload cycles | Up to 30 years |
| TS1130 | 300* | 20,000 load/unload cycles | Up to 30 years |

*TS1120

Source: Sun Storage Tek, IBM Corporation

In the mainframe environment tape storage is used for transactional process with application such as LOB, OLTP, CRM and other high duty cycle applications. All this requires lots of starts and stops which puts tremendous physical stress at the tape drive and tape cartridge.

In order to achieve even faster backups and recovery processes, Virtual Tape Library (VTL) technology was developed. VTL is using disk array to emulate tape drives and tapes. Disk is a random access medium which results with higher performance rate. After some time data from virtual tapes that are spinning on disks will be migrated to the physical tapes. This is called disk-to-disk-to-tape (D2D2T). Enterprise tape technology is dominant in the VTL because it is able to sustain heavy duty cycles.

Proprietary IBM 3592 and Sun Storage Tek T10000A/B tape drives and medium are representing top of the peak in Enterprise tape storage technology.

[11] See: IBM System Storage Tape Library Guide.

Sun Storage Tek T10000B was the first available tape cartridge medium with the native capacity of 1TB. It was released in 2008 as a successor to the T10000A released in 2006. T10000A can be reformatted to T10000B capacity. The drive is not compatible with any previously released Sun/STK tape formats. T10000B tape cartridges are available in two formats: *sport cartridge*, with rapid access over less capacity, and *standard cartridge*. Both formats can feature WORM capability.[12]

Table 4: Enterprise class tape drive specification

| | Data transfer rate | Data transfer compressed | Native capacity | Compressed capacity | MTBF |
|---|---|---|---|---|---|
| T10000B | 120 MB/s | 360 MB/s | 1 TB | 2 TB | N/A |
| TS1130 | 160 MB/s | 350 MB/s | 1 TB | 2 TB | 290,000 hr |

Source: Sun Storage Tek, IBM Corporation

The IBM TS1130 represents third generation of 3592 tape technology. The first generation was introduced in 2003, while the third generation came out in 2008. TS1130 uses existing 3592 tapes and provides backwards compatibility, supporting read and write for 3592 generation 2 and read only for 3592 generation 1. Three formats of tape cartridges are available: *short-length* – providing rapid access, *standard* – providing high capacity and *extended*. Cartridges are available in WROM and rewritable format.[13]

## Conclusion and Recommendations

In general, tape storage technology is the most affordable storage technology today[14]. When it comes to archiving, both LTO-4 and enterprise tape systems are suitable. However, LTO-4 format is offering more than sufficient capacity and performances for archiving purposes at lower costs than the enterprise tape systems. In addition, LTO-4 is designed to work with the open system platforms while enterprise tape has remained primarily in the proprietary mainframe systems. Since a lot of information and storage systems in archives, museums and libraries are build using open system platforms, LTO-4 could be a more appropriate solution for such institutions.

It could be suggested to these institutions to hold dual tape systems. The primary system should consist of disk storage which complements tape storage. In that case the data is virtualized at disk storage while it is being retrieved from tape storage. The layout of system should support fast data access and retrieval which grants utilization of archive by users. There should also be the secondary

[12] See: Storage Tek T10000 Tape Drive, Operators Guide // Sun Microsystems Inc. Broomfield : Storage Technical Publications, 2009. http://dlc.sun.com/pdf/96174revEC/96174revEC.pdf (last access: 2 August 2009).

[13] See: IBM System Storage Tape Library Guide.

[14] In US $ per MB of storage.

system, which is called electronic vault and it is usually placed off site. Users should not have access to this archive. The main purpose of an electronic vault is the disaster recovery, archiving for future usage and migration to the new technologies. Only tape storage system, without disk storage, should be sufficient for the needs of an electronic vault.

The most applicable type of a large storage system for archives, libraries and museums is the tape library. Thanks to their modular design, the tape libraries can be easily reconfigured and upgraded to the new tape technologies. Entry level LTO-4 libraries are scalable up to native capacity of 20 TB, 40 TB and 2-4 tape drives. For example, a library which has 5 TB of data equals to approximately 500,000 books[15]. All this data could be stored at six LTO-4 tape cartridges without compression. If the plan is to digitalize 1TB of video content per year in the next 3 years, the library can be extended with additional three cartridges. They are also more reliable than autoloader because the system is set up that in case of one tape drive failure the other will take its place. At the same time inappropriate cartridge handling is minimized[16]. Multiple tape drives could also enable simultaneous write/read operations on multiple tapes. However, current LTO-4 technology will became obsolete in approximately six years. At that time tape drives and tapes inside the present libraries should be replaced with the new LTO generations of drives and tapes. This will be possible due the modular design of LTO libraries and the life of library can thus be extended to approximately ten years. The pricing of entry tape LTO-4 library with two tape drive and twenty tape cartridges is up to 15,000 €. This should be affordable for any institution planning serious digitization or already holding large amount of digital data on unstable media and thinking about migration.

We strongly suggest that archives, libraries, museums and other information institutions involved in the digitization of records and cultural heritage should consider recommended tape technology when building large storage systems. It could provide space and reliability for large collections thus adding to the positive perception and trust among the users and financial supporters while at the same time preparing the ground for future certification processes of the system and the applied storage and archiving procedures.

[15] Approximation: 10 MB per electronic document.

[16] The reported main reason for tape damage is its accidental dropping on the floor.

## References

Blair, Colin; Currie, Julie; Goodall, Eric; McElyea, Kevin; Miller, George; Poston, Ben. IBM Medical Archive Solution // IBM Redbook, First Edition, 2004. http://www.redbooks.ibm.com/redpapers/pdfs/redp9130.pdf (last access: 15 August 2009)

Brooks, Charlotte; Byrne, Frank; Higuera, Leonardo; Krax, Carsten; Kuo, John. IBM System Storage Solutions Handbook // IBM Redbook, Seventh Edition, 2006. http://www.redbooks.ibm.com/redbooks/pdfs/sg245250.pdf (last access: 20 August 2009)

Castets, Gustavo; McLure, Chris; Koutsoupias, Yotta. IBM TotalStorage Tape Selection and Differentiation Guide // IBM Redbook, Third Edition, 2004. http://www.redbooks.ibm.com/redbooks/pdfs/sg246946.pdf (last access: 25 July 2009)

Haeusser, Babette; Kessel, Wolfgang; Silvestri, Mauro; Villalobos, Claudio; Zhu, Chen. IBM System Storage Tape Library Guide for Open Systems // IBM Redbook, Seventh Edition, 2008. http://www.redbooks.ibm.com/redbooks/pdfs/sg245946.pdf (last access: 18 August 2009)

LTO Ultrium format reaches new heights with over 100 million cartridges shipped // LTO, 2008. http://lto.org/pdf/LTO%20100%20Million%20Cartridge%20Milestone.pdf (last access: 20 August 2009)

Reine, David; Kahn, Mike. Clipper Notes: Disk and Tape Square Off Again – Tape Remains King of the Hill with LTO-4. Wellesley : The Clipper Group Inc., 2008. http://www.dell.com/downloads/global/corporate/iar/Clipper_Tape_v_Disk_2008.pdf (last access: 17 August 2009)

Storage Tek T10000 Tape Drive, Operators Guide // Sun Microsystems Inc. Broomfield : Storage Technical Publications, 2009. http://dlc.sun.com/pdf/96174revEC/96174revEC.pdf (last access: 2 August 2009)

Sun StorageTek Linear Tape Open (LTO) Ultrium Data Cartridges // Sun Microsystems. http://www.sun.com/storage/tape_storage/tape_media/lto/specs.xml (last access: 22 July 2009)