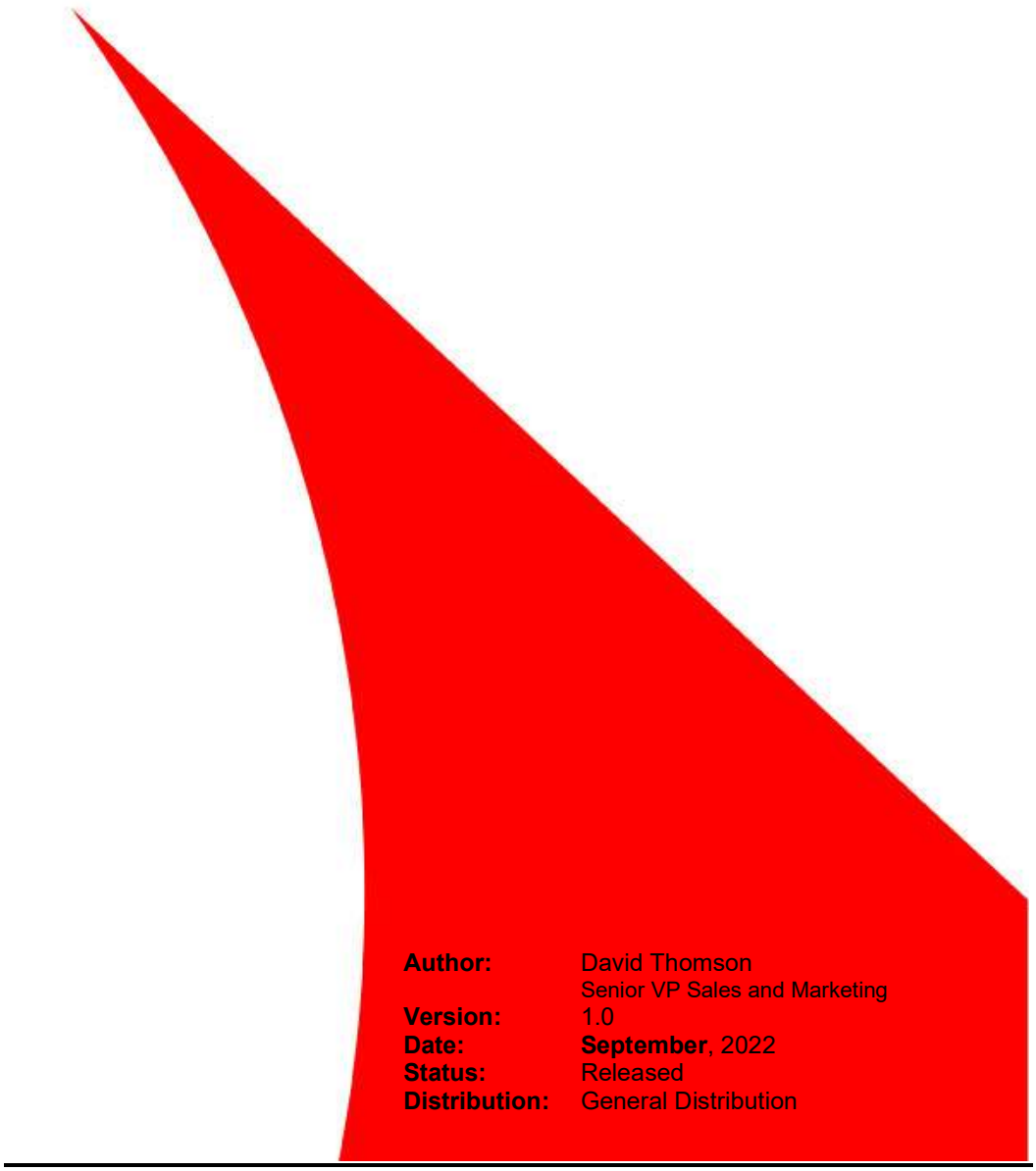
Why is Tape Storage Relevant for your Archive Today?



Author: David Thomson
Senior VP Sales and Marketing
Version: 1.0
Date: **September**, 2022
Status: Released
Distribution: General Distribution

www.qstar.com

Introduction

For decades, tape was the storage technology most organizations used for backup. Over the last 5 to 10 years, we have seen a rapid shift away from tape for backup as disk-based "Data Protection" and virtual machine environments became more prevalent. Therefore, the tape market must be decreasing… but In fact – tape media capacity is still increasing year on year.

As tape for backup has declined, tape has seen a resurgence in the archive space for many reasons that are outlined below. But first let's discuss the difference between backup and archive, as even "experts" intermingle the two as if they are the same.

A backup is an extra store of your primary storage which allows for single file, multi file or whole system restore operations based on an Administrators manual intervention. Older backups are often retained to allow for restoration of files that no longer exist on primary storage, and therefore are no longer being backed-up. This is often the confusion "experts' have with archive.

An archive is often the only store of data, preserved in a secure method for single or multi-file reading. Reading in this case does not mean that data has to be pulled back to primary storage, it can be accessed in place, without the need to consider if there is enough storage capacity available at the original location. It also means individual users can access data directly, without Administrator intervention.

Tape provides exceptional value-for-money over disk or flash-based solutions (including Cloud, which of course uses disk) for long-term archive storage, particularly when looking at power and cooling costs. Naturally, no technology is perfect, so there are downsides to using tape as well.

Let's cover those downsides first.

Why not Tape?

Tape media, with latest LTO9 generation, has an 18TB native capacity. Tape libraries allow many media and tape drives to interact together with a media exchanger doing the media movements. Of course, robotic movement takes time, often over one minute per media exchange and media load. In addition, it is possible with tape for the media needed to be out of the library (offline), in which case delays could be hours or days, particularly if that offline store is remote. So speedy (random) access will always be an issue with tape storage.

The next issue with tape is the requirement for manual intervention to export and add tape media when required. Disk systems do not include any concept of "offline" – so other than hardware maintenance, there is no requirement to employ anyone to "manage" a disk-based store.

Finally, obsolescence is a concern for tape as new generations of tape drive are not able to read all previously written tape media. This requires a media migration operation to copy data from older media to newer media, or the inclusion of older and newer drives in the same library or multiple libraries.

So why tape?

For organizations where longer-term data preservation over-rides fast access, tape can be used very successfully. To create a tape archive environment a "Gateway" is required, allowing data to be written to tape but allowing users to search for those assets that have been previously stored. In the past, gateways offered file access through SMB or NFS (Tape as NAS). Today, S3 has been added to many, providing a Cloud-like storage option. This means it is possible to write data locally through a file system, but access it remotely through S3, or to allow modern applications that have chosen S3 as their route to archive to use tape instead of Cloud (Tape as Cloud). Gateways (such as QStar Archive Manager), track all content, so when data is requested, it will be read automatically back to a disk cache within the gateway, without the need to restore to primary storage.

Tape is an extremely cost-effective technology to leverage. Other solutions require online replication or erasure coding (a specific method where data is broken into fragments and a preset number of these fragments are stored across multiple media). This eats into any total raw capacity purchased, so organizations are paying for storage they can't use. Tape uses copies of media to ensure protection and those copies are most often stored offline, therefore only incurring the small additional cost of media. Copies of media can be created "on-the-fly" using two drives (either in the same library or perhaps a second library at another site) or as a scheduled operation. It is best to use the first option when an organization is filling media very quickly (perhaps a media in under one week). The Disaster Recovery copy can then be stored offline.

In addition, first copies of tape can also be stored offline, allowing archives larger than the library purchased to be stored. For content where access is acceptable after hours or days, offlining tape media with barcode labels provides an unlimited data capacity using media only. The gateway continues to show data, and APIs, emails and / or GUI settings show the Administrator which media to fetch for retrieval of data when required.

Tape media is extremely secure, with encryption technology built into every tape drive. Many of the "facts" about tape media wear are associated with backup operations, where tape media is often re-used / re-written regularly. Archive does not use tape in the same way, once written, by definition it is not overwritten regularly (if at all). It also offers media life up to 30 years (disk is typically 3 to 5 years) and significantly fewer bit errors (tape is $1:10^{19}$ or one error every 1.25EB, compared to disk at $1:10^{15}$ or one error every 125TB, that's 1,000 times better!). Another "expert" misconception is that tape requires very specific storage requirements. Tape media should ideally be stored at temperatures between 60°F and 77°F, so a normal office environment is fine. Humidity is more of a concern, with ideal humidity up to 50% before media life begins to be impaired (offices norms are typically 40% to 60%).

A bi-product of the design of tape is in-built ransomware protection. Tape writes sequentially from the beginning of the media to the end and back again in a serpentine method. It is impossible to "overwrite" a tape, it has to be wiped clean completely before new data can be written. Subsequentially, if an organization is hit by a ransomware attack where data is re-written in an encrypted format, this can only be done as an append operation. Original data is still available with the original unencrypted data. Many gateway solutions (including QStar Archive Manager), provides a "mount on date" option to see the archive at a point before the ransomware attack.

Write performance is very high, allowing up to 400MB per second (before any compression) per LTO9 tape drive. Compare this to SATA drives at 100 to 170MB/sec and SAS at 200 to 230MB/sec. This is why most disk systems use RAID technologies to increase performance by using many drives together. Of course, RAID then has rebuild issues, particularly with large capacity drives, in that they can take many hours or days to complete, leaving the system open to second or third drives going down and therefore data loss.

To improve performance, some gateways (including QStar Archive Manager) allow grouping of tape drives together as one volume. Others have the option to use RAIT (Redundant Array of Independent Tapes) to achieve the same thing. RAIT requires as many tape drives for reading as there were initially for writing, and therefore longer delays while correct media is put in the drives. Tape drive grouping offers options to group data from the same subdirectory(ies) onto one media, so that only one media needs to be loaded for reading any content from that subdirectory or project.

Archive uses for Tape

As previously stated, tape should be used in environments where speed of retrieval is less important. Tape is used in environments with total capacities of 50TB to 100EB (so extremely wide). Tape archives can be used for individual person / department archives where files or subdirectories are manually moved from primary to archive storage once a project closes, or more often, with a data mover (such as QStar Network Migrator) that uses policies to determine when a file should be archived. This is then an automatic process where often tiny stub files replace files and are used for simple reading or retrieving of data.

Many applications have an archive option to periodically store content that is not needed immediately using SMB or NFS file protocols or more recently S3-based APIs. With the use of a tape gateway, that data can be stored to tape for lower cost and long-term preservation. Applications, such as Media Asset Managers (MAM), Healthcare (DICOM and PACS), AI and ML, High Performance Computing (HPC), and Email, plus many more, offer options to store data via SMB, NFS and S3.

Tape as Cloud (S3 connected tape) also offers cost-effective multi-cloud options. A multi-cloud environment allows organizations increased security and access options. Copying cloud data to tape via S3 provides a true 3-2-1 best archive practice as it includes the possibility the cloud provider ceases trading. 3-2-1 dictates three copies of the data are stored on two different technologies (in this case cloud and tape), one copy is offsite.

How QStar can Help

For organizations that choose to use Tape Libraries, QStar Technologies Archive Manager can provide file services (SMB and / or NFS) adding a NAS-like option for file-based archiving and cloud APIs (S3) adding a cloud-bucket approach. Archive Manager is an application that can be installed on Windows or Linux servers, using a disk cache to facilitate writes to tape and then reads from tape. All tape libraries and versions of tape (including IBM enterprise tape) are supported. Archive Manager can create a secure archive environment using retention management options and roll-back file system features to help in the event of a ransomware attack or accidental or malicious deletes or overwrites. Other options such as encryption and compression are also available. For organizations that wish to secure their data even further and potentially save significant costs, QStar Archive Replicator allows content to be replicated

to other sites and other technologies. For example, replicating data from a local tape library to Cloud storage, or to create a synchronized replica on a second tape library, locally or remotely.

For more information see www.qstar.com or email sales@qstar.com.

Summary

Many organizations have criticized tape for backup for many years. Archive software is very different and uses data verification techniques such as checksums and hashing to guarantee data has not changed.

As many organizations no longer use their tape libraries for backup, they have freely available systems that can be re-used for long-term archive, for example as Amazon Glacier / Deep Archive alternatives. Glacier / Deep Archive are other classes of cloud storage which although much cheaper than regular cloud, are still expensive compared to an internally managed tape library system.

Tape Storage for your Archive Today	
Pros	**Cons**
Low cost (online and offline)	Slow random access
Project / mass retrieval	Single file retrieval
Fast write performance	Media migration required
Ransomware Protection	Manual media management
SMB, NFS or S3 access available	
Secure for long-term	
Huge capacity options	
Tape roadmap to LTO14	