REINTRODUCING TAPE TO THE MODERN DATA CENTER

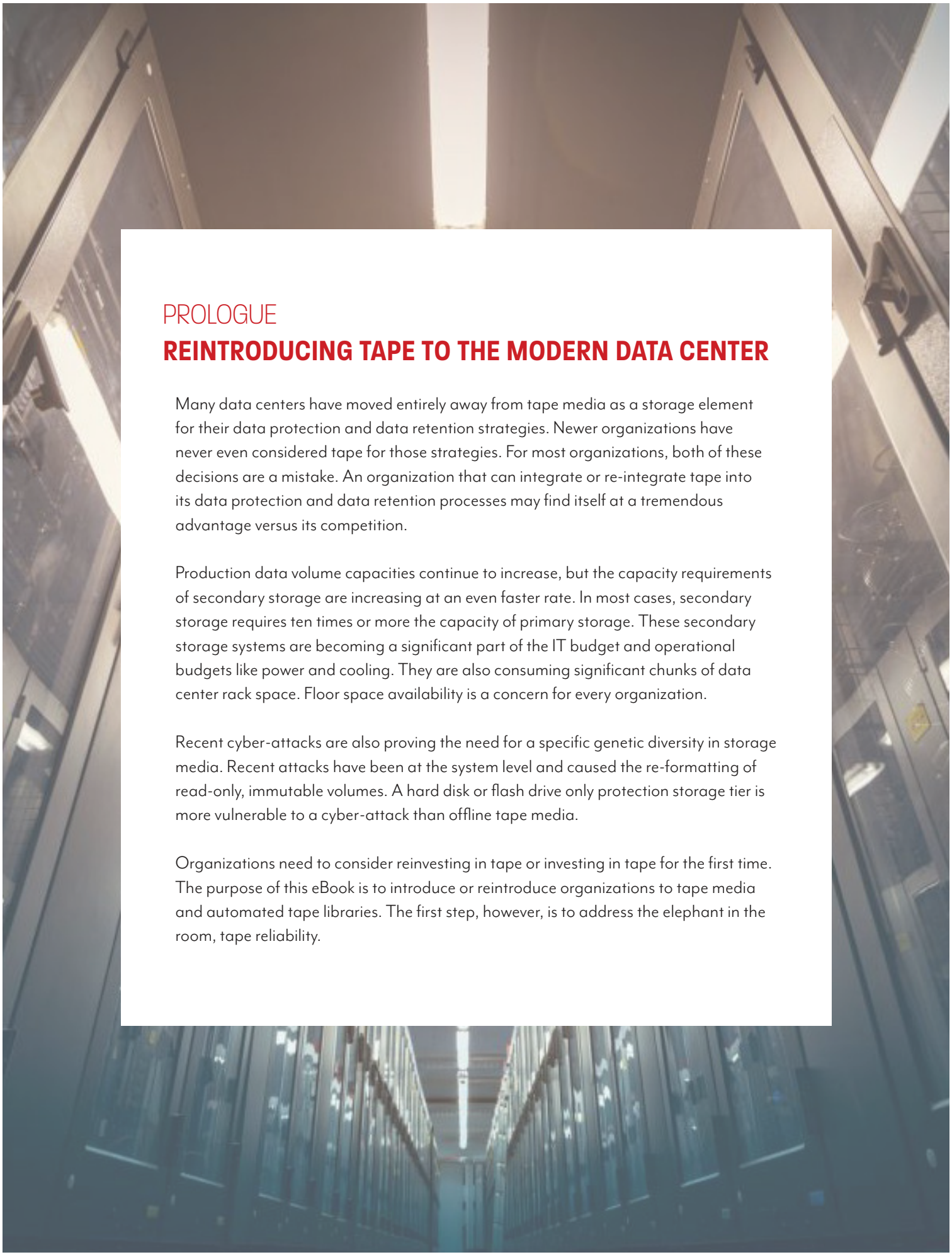
Storage Switzerland, LLC

FUJIFILM

SPECTRA

IRON MOUNTAIN®

TABLE OF CONTENTS



PROLOGUE

REINTRODUCING TAPE TO THE MODERN DATA CENTER

Many data centers have moved entirely away from tape media as a storage element for their data protection and data retention strategies. Newer organizations have never even considered tape for those strategies. For most organizations, both of these decisions are a mistake. An organization that can integrate or re-integrate tape into its data protection and data retention processes may find itself at a tremendous advantage versus its competition.

Production data volume capacities continue to increase, but the capacity requirements of secondary storage are increasing at an even faster rate. In most cases, secondary storage requires ten times or more the capacity of primary storage. These secondary storage systems are becoming a significant part of the IT budget and operational budgets like power and cooling. They are also consuming significant chunks of data center rack space. Floor space availability is a concern for every organization.

Recent cyber-attacks are also proving the need for a specific genetic diversity in storage media. Recent attacks have been at the system level and caused the re-formatting of read-only, immutable volumes. A hard disk or flash drive only protection storage tier is more vulnerable to a cyber-attack than offline tape media.

Organizations need to consider reinvesting in tape or investing in tape for the first time. The purpose of this eBook is to introduce or reintroduce organizations to tape media and automated tape libraries. The first step, however, is to address the elephant in the room, tape reliability.

CHAPTER 1

MYTH BUSTING TAPE'S SHORTCOMINGS

To many IT professionals, tape media and tape libraries seem like an ancient technology used by the "old" IT guys. The reality is tape as a technology can be of significant value to the organization. Almost every data center is experiencing explosive data growth, and tape can stem the tide. In addition, tape provides the ultimate protection from cyber-attacks. Before we explore the value of tape to a modern data center, though, let's address a few persistent myths about tape. The most common tape myths are that tape is unreliable, provides slow data access and is difficult to operate. Let's examine each of these myths one by one.

MYTH 1
TAPE ISN'T RELIABLE

Some people, maybe even you, have said that tape reliability is questionable. You might be surprised to learn that tape is actually more reliable than all other media formats, and its reliability increases when it's implemented in the correct way.

It's difficult, if not impossible, to identify a single report or study that backs up the claims of unreliability. So where does this claim come from? The myth apparently originated in the 1990s and early 2000s with emerging disk backup appliance vendors and some good old-fashioned scare tactics for the sake of sales. The sales pitch sometimes went so far as to seek credibility by citing studies from third parties like the Gartner Group and The Yankee group – studies that didn't exist. Neither organization ever concluded that tape had higher failure rates.

It's reasonable to assume that most IT professionals in that era had experienced a tape media or tape drive failure at one time or another. But it's more than likely they experienced disk media failures too. So predictable was disk failure that, by the late 1990s, data protection algorithms like mirroring or RAID were considered a requirement. Interestingly enough, tape media was generally so reliable that the concept of mirroring tape copies never really took off.

Unlike hard disk drives, tape media's biggest reliability challenge comes from its portability, which can invite the potential for classic human fallibility and clumsiness. The scene is easy to imagine—an IT admin accidentally drops a tape while walking across the data center floor. With no witnesses, he doesn't report it and places it back in the library. On a subsequent use, that damaged tape fails.

Modern tape libraries address this portability challenge by simply taking us lumbering humans out of the equation as much as possible. Libraries now have the capability to eject tapes directly into a shock absorbent carrier.

Returning to the studies – while there's nothing declaring tape to be less reliable, there are plenty of studies showing that tape is more reliable than disk media and has been for years. There are also studies indicating that tape drives are increasing in reliability over time. And no study has concluded that robot arms inside libraries have ever been a significant source of failures.

In fact, according to the 2019 INSIC report, tape's key durability benefits compared to hard disk drives can be summarized as:

- Tape has 30-year life compared to HDD's maximum five years.
- Tape has more than five orders of magnitude lower bit error rate (BER) compared to HDD.
- Tape has 50 times lower annualized failure rate (AFR) compared to HDD.

When tapes are stored and operated in the recommended environmental conditions and their usage doesn't exceed the End of Life media wear numbers, tape offers extremely high durability.

MYTH 2

TAPE HAS SLOW ACCESS

The big three media formats -- flash, hard disk and tape -- are like a rock-paper-scissors game for modern data centers. Each format is a winner in different scenarios. Tape provides excellent sequential performance so it's ideal for large data movements like a backup job or restoring multiple servers at once. Individual file restores are a less ideal use case for tape. But in situations like archiving, where time sensitivity is less of an issue, tape is more than acceptable especially when other factors like cost per TB and low power consumption are considered. Interestingly enough, flash is putting far more price per GB pressure on hard disk than it is on tape.

The data center should use each of these media formats for their designed purpose: flash for storage-intensive applications where data is active most of the time; hard disk as an intermediary storage location; and tape media as a long-term storage location.

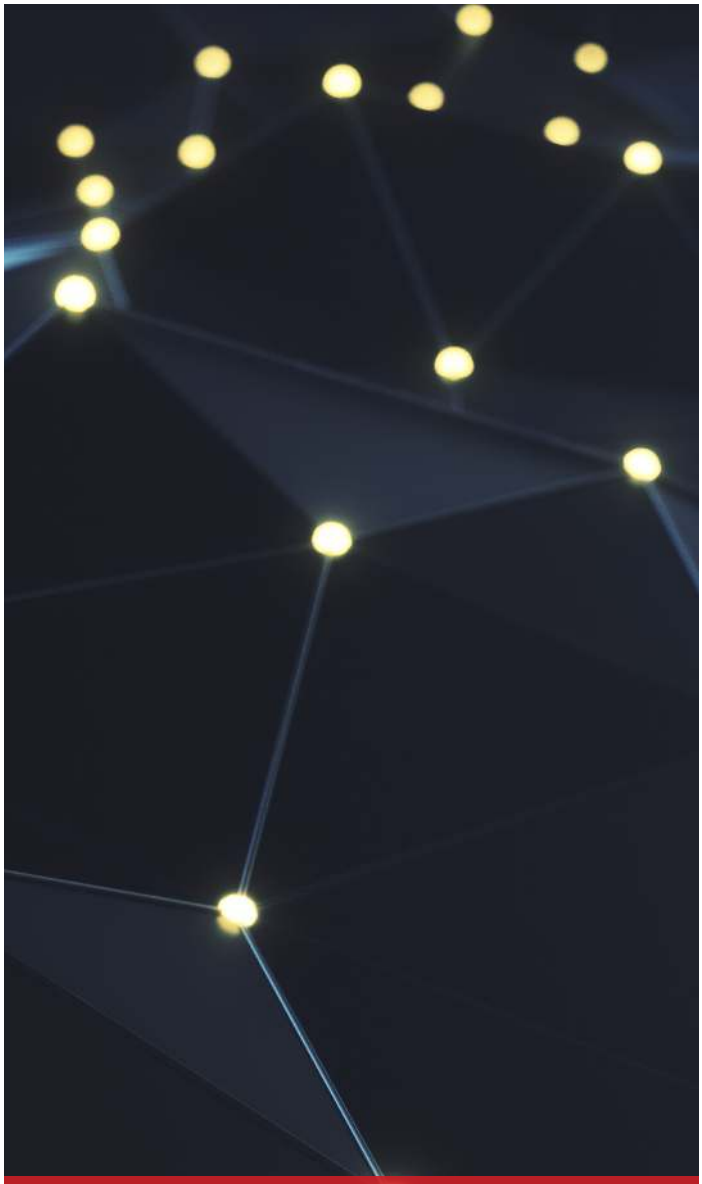
Storage Switzerland strongly believes that using tape for both archive and backup use cases is ideal. While archive is the default use case for tape, most backup data sets are stored for seven or more years. The likelihood of a time sensitive recovery from anything but the latest backup is unlikely.

MYTH 3

TAPE IS HARD TO OPERATE

The modern tape environment almost always includes a sophisticated tape library that automatically mounts and unmounts tape cartridges as they're needed. Most traditional backup applications like Commvault, Veritas, Dell Networker and IBM Spectrum Protect (formally Tivoli Storage Manager) as well as newer solutions like Veeam Backup and Replication, all support tape media and tape libraries. In fact, officials from Veeam (a company that was a disk-only backup solution) constantly state that their customers are leveraging their tape module. Most of these solutions will buffer data to disk and then can automatically copy or move that data to tape. Today's tape libraries also track details like capacity, media usage and media location.

Thanks to this automation, today's data centers can expect to interact with a tape library in much the same way that they interact with any backup software or even their regular file systems.



Tape provides excellent sequential performance so it's ideal for large data movements like a backup job or restoring multiple servers at once.

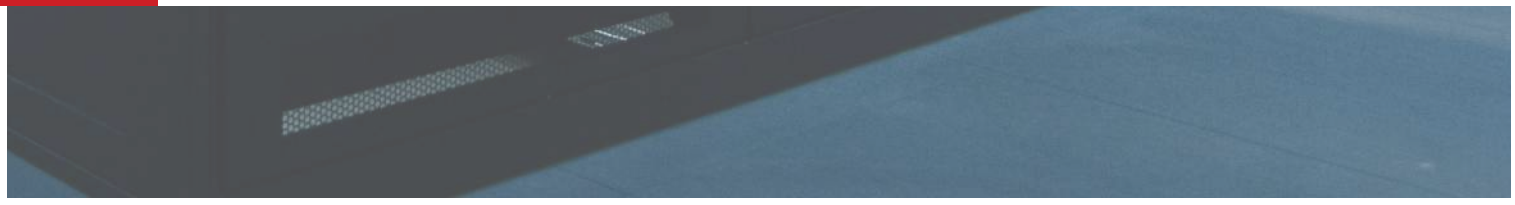
CONCLUSION

MYTH STATUS: BUSTED

The myths suggesting that tape is somehow inferior are invalid and based on twenty year-old rumors and hearsay (if not an outright conspiracy). The modern tape library can seamlessly coexist with the rest of the modern architecture and not add operational overhead. With tape in the mix, the organization is set to benefit from its many advantages including low cost, low power consumption, air gap protection, long-term shelf life and media diversity. Our next chapter starts this journey by discussing the role tape plays in the modern data protection or retention infrastructures and how modern organizations should build tape into those infrastructures.

CHAPTER 2

HOW DOES TAPE WORK IN THE MODERN INFRASTRUCTURE?



When considering integrating tape storage into their infrastructure, IT planners must consider the realities of how data is accessed. The IT planner needs to maximize tape's strengths with the goal to store as much data as possible on tape because of its lower cost and inherent air-gap advantages while not sacrificing recovery performance when it matters most.

UNDERSTANDING RECOVERY REALITIES

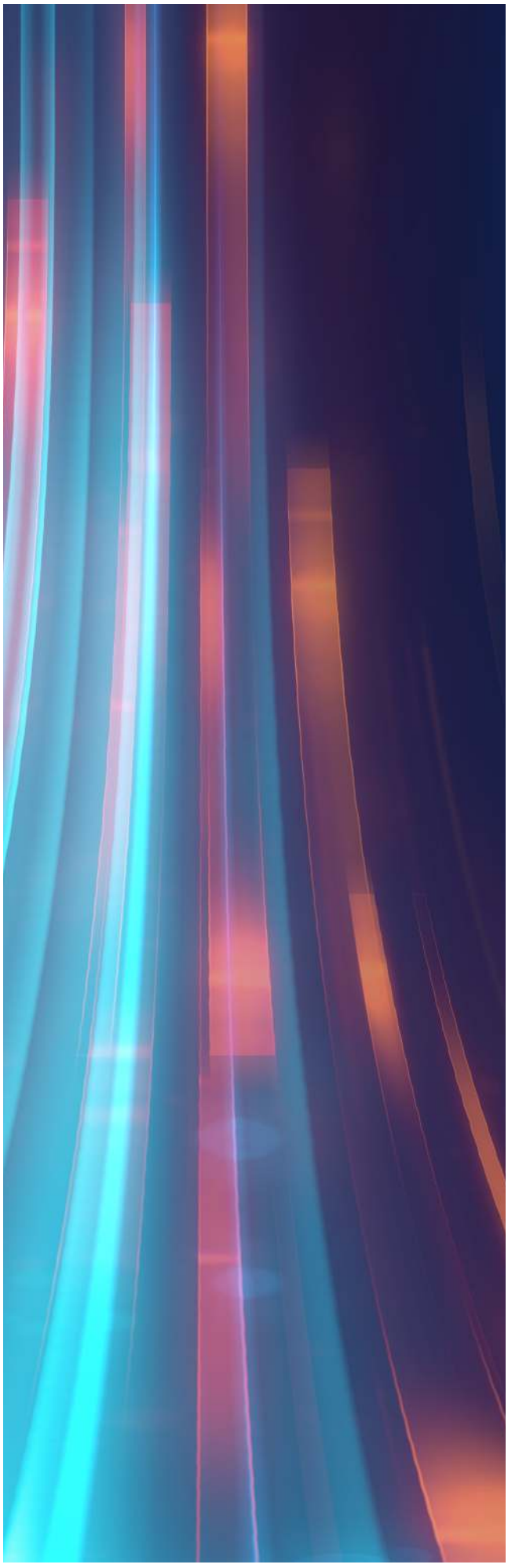
The natural inclination is to limit tape's role in the data center to the archive function. Storage Switzerland believes that tape has a significant role to play in the backup process; a case can be made that tape is at least as well suited for backup as it is for an archive. The reason for our strong position on tape's role in backup is the realities of recovery requests. Storage Switzerland continuously finds that the overwhelming number of recovery requests come from the most recent backup. In fact, as much as 95% of all recoveries come from the most recent backup. That means that 95% of backup storage capacity is rarely called on for a recovery. We also find that the backup storage capacity is approximately 5X to 10X the size of production. That means in an environment that has 100TBs of production data, they may have a petabyte of backup data, but of that backup data...but 950TBs of it will rarely be accessed.

Why keep that 950TBs of data around? The primary reason is so the organizations can meet compliance and regulatory requirements. The other purpose is that 5% of recoveries do come from this data set, and it is almost impossible to know what 5% of that 950TB will need to be recovered in response to a request. The organization is forced to keep all the data. Essentially, the capacity requirements are inversely proportional to the number of recovery requests.

It is interesting to note though that the recoveries from older backups tend not to be as time sensitive as recoveries from the most recent backup. Generally, when an organization is recovering from the most recent backup, it is because a primary system has failed, and the organization needs to get that system back into production as quickly as possible. Recoveries from older backups are typically in response to discovery requests, regulatory responses, or a need for data to analyze. While these recovery requests still need to be served promptly, they do not need to be instantly available.

Another use case for older backup sets is when a major disaster strikes the organization causing destruction of the data center. An additional and more pressing threat that can cause an equal amount of damage is cyber-attacks such as ransomware. In these situations, the organizations need data that is both offsite and 'air-gapped' so that the disaster or cyber-attack doesn't impact data quality. Organizations can provide an "air gap" by copying data onto removable media and sending that media offsite and offline, ensuring it is electronically disconnected from the network.

> 95% of backup storage capacity is rarely called on for a recovery.



DESIGNING A BACKUP ARCHITECTURE FOR THE RECOVERY REALITIES

Organizations should take advantage of the reality that not all backup data requires instant recovery and design backup architecture accordingly. An ideal architecture is to have a small and slowly growing disk (or even flash) front end coupled with a tape library-based backend. The flash or disk tier should store the most recent backup to facilitate the 95% of time-sensitive recovery requests, and the tape tier should save the rest of the data, again taking advantage of the inverse proportion of capacity as compared to recovery requests. In our example above, we'd allocate 100TBs of backup data to the flash or disk tier, and we'd allocate 900TB to the tape tier.

The cost ramifications of this design are significant. It eliminates the need for multiple or scale-out disk backup appliances, and we can make a case that even deduplication is no longer a requirement. Instead, the organization can invest in a more expensive, high-performance flash tier, with the more economical tape library providing extremely cost-effective storage capacity with almost limitless scalability and even more cost-effective power and resource consumption.

This architecture simplifies the storage capacity requirements at the remote site as well. The flash backup appliance or the backup software can replicate data to the alternate disk tier at a remote location. IT can easily contract for tapes to ship to a dedicated tape vault. In a disaster, organizations will want the latest copy of data, not data that is six months old. The organization can justify waiting a few hours for these tapes to be shipped to them while they are busy recovering the critical systems from disk. Additionally, storing these tape copies offsite, further leverages the air-gap advantage of tape as well as its long-term shelf-life of 30 years.

CONCLUSION

Once IT understands the reality of recoveries, the benefits of cost and power-efficient tape storage become immediately apparent. Combined with its natural air-gapped nature, the tape becomes the ultimate protection from disaster and cyber-attacks. Modern data protection software now includes the ability to automatically move data as it ages from the backup set to tape media. IT planners can now easily support moving data from disk to tape without adding management overhead.

CHAPTER 3

COMPARING THE PERFORMANCE OF TAPE INTEGRATED PROTECTION TO DISK ONLY PROTECTION

When vendors compare the performance of disk backup systems with the performance of tape, they often will highlight the area in which their particular product has the biggest advantage and ignore situations where performance doesn't matter. IT planners looking to optimize backup and recovery performance, as well as cost, should concentrate on designing an architecture that strikes the right balance between performance and cost. Creating a balanced architecture requires exploiting the advantages of each technology while compensating for its weak points.

Disk backup vendors tend to claim restore performance superiority, but that superiority is primarily for single file or folder-level restores, not for entire systems. Many disk backup vendors leverage technologies such as deduplication in order to compete with prices of tape based systems. However, deduplication impacts performance, especially in large restoration scenarios. In fact, ExaGrid—a disk backup vendor known for first storing data to a non-deduplicated area of the disk system—claims a significant restore performance advantage over deduplication- only systems, especially as restores get larger.

Tape-based libraries actually have an advantage in performance when doing large full system recoveries as are common in disaster recovery situations. It is also necessary to keep in mind that most organizations are protecting mission-critical systems with something other than backup, most commonly a high-availability solution like replication. For mission-critical applications, the speed of backup recovery is less of an issue since the primary recovery point, at least in a disaster, is not even part of the backup process. Additionally, the cost advantage of tape may enable the organization to leverage flash storage as the backup architecture's initial ingestion point, with tape serving as an economical secondary storage tier for multiple onsite and offsite copies, making the total architecture significantly faster and more secure for both backups and recoveries.

Creating a balanced architecture requires exploiting the advantages of each technology while compensating for its weak points.

IT'S ALL ABOUT ARCHITECTURE DESIGN

The job of the IT planner is to design an architecture that best meets the organization's current needs for capacity and performance while anticipating future needs – all while staying within budget. If organizations design their backup architectures as recommended in our previous chapter, the first tier is high performance which can be either flash or disk. In both cases though, deduplication is not as necessary since that data is tiered off to an affordable tape tier. Our recommended best practice is copying the data directly to the tape tier as soon as the backups to the ingest tier are complete. Copying data to tape immediately, enables the organization to have redundancy of backups faster. Also, when the time comes to release older data from the first tier, the backup software makes no additional copies. No deduplication on the ingest tier not only delivers faster ingestion it also delivers a faster secondary copy to tape. Regardless of the organization's stance on tape, no IT professional should consider data protected until it is on two or three separate systems, preferably on two different media types and in different geographic locations with one being offsite and offline.

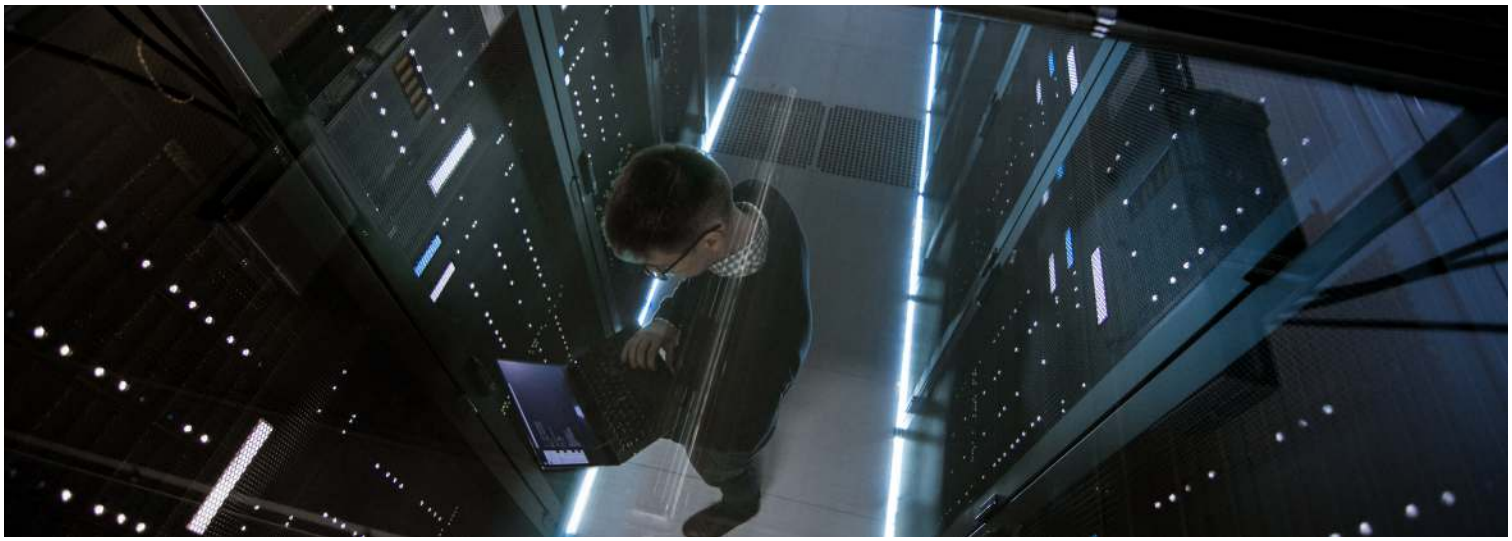
CONCLUSION

The short answer is that tape is faster than disk once it reaches its full speed. The problem is, and the reason why many disk backup vendors claim that disk is faster, is that it takes a while to get tape going at full speed and it also takes a lot to saturate tape to its full potential. When the infrastructure can't feed tape at its full potential, the technology has to slow down and wait for more data. Given the speed of the modern tape drive, many organizations have trouble creating an architecture that can maintain tape saturation.

The key for the IT planner is to realize that tape is faster and design a backup architecture to keep it properly fed. The architecture described in the last chapter is an ideal method to do that. With the tape library attached directly to the ingestion system, the chances of keeping tape running at full speed increase greatly, especially if the ingestion system is not using deduplication.

CHAPTER 4

COMPARING THE PRICE OF DISK-ONLY BACKUP TO TAPE INTEGRATED BACKUP



As discussed in our last chapter, tape, when correctly integrated into the disk architecture, does not negatively impact the performance of the backup infrastructure. In fact in some cases tape improves it while also providing a more efficient, more reliable long-term storage capability. Another advantage that organizations will realize when integrating tape into the backup architecture is a significant price reduction of the backup infrastructure.

SMALL-FAST DISK

Since the majority of restore requests are from the most recent set of backups, the disk backup staging area only needs to be large enough to store the most recent backups. Most disk backup systems, however, are actually sized to store years, if not decades, of backups. The organization either ends up buying multiple disk storage systems or is constantly forklift upgrading those systems or investing in a more complicated and expensive scale-out backup architecture. Integrating tape into the backup architecture enables the organization to eliminate all of this complexity and much of the cost. A simple scale-up NAS or low node count object storage system will more than suffice.

Organizations can use these savings to invest in a high-performance ingest tier because restore requests from the most recent backup sets also tend to be the most time sensitive. The need for a higher performance is especially evident for organizations that are leveraging the in-place/instant recovery capabilities of most software applications. A small all-flash system with basic features is ideal for this use case.

THE TAPE ADVANTAGES

In the previous chapter we demonstrated how tape could integrate into the existing backup architecture. Now let's look at the financial impact. First, tape does have an upfront cost. The organization must buy the tape library and the drives to go in that library. Depending on the requirements for making a tape copy however, the number of drives may be limited to three or four. After making that investment, the savings begin and multiply.

WHAT IS A LIBRARY?

A tape library is a chassis that contains a number of slots (shelves) that hold tape cartridges. There is at least one robotic arm, and in some larger systems two, which will automatically pick a tape cartridge when the software requests it and load that cartridge into one of the available tape drives. When the software finishes with that cartridge, it will return the tape cartridge to a slot in the library.

WHAT IS A TAPE CARTRIDGE?

A tape cartridge is a storage device, which contains a spool of magnetic tape and is normally used in an automated tape library, to store many different kinds of data. There are primarily two formats in use today LTO and IBM Enterprise Tape.

Adding additional capacity simply requires buying more tape cartridges. Typically, 12TB plus of capacity is available for less than $150. In most cases, the capacity of the tape media doubles since most drives have built in compression and greater than 2:1 compression ratios are not uncommon. Tape media today is available pre-labeled with a bar code. The software and robot use that barcode to request the right media and to track that media's use.

THE SAVINGS OF INTEGRATING TAPE

The first obvious savings is the cost of the media, especially when compression is factored in. Another savings is when a tape cartridge is not in use, which is its most normal state, since it requires no power. Idle data on hard disk drive systems use power continuously and the industry has long given up on spinning down hard disk drives.

In most cases, the organization will make two copies of the same data on two different pieces of tape media. One of these copies will go off-site and another will stay in the library. The off-site tape should be stored in a secure, temperature-controlled facility and managed for easy retrieval. Most software applications that support tape also manage the tracking of off-site media and some vaulting locations can also provide tape media management services.

The cost of creating a second copy and storing it in a secure offsite facility is far less expensive than replicating a massive disk backup storage system to a second location, which requires a second system. It also requires keeping both systems powered on. Alternatively, some disk backup vendors may offer public cloud storage as an option. While the public cloud does have lower upfront costs, the recurring cost of storing data (especially data that doesn't change) for years or decades will quickly surpass the upfront cost of the library. There are also egress fees, which can be substantial, to consider should it become necessary to retrieve any data from the cloud repositories.

It is also important to consider that most libraries remain in service for more than ten years, so the upfront cost pays off for a long time. The organization can expand tape capacity until the slots in the library are all used. Most libraries have some form of expansion that enables them to add slots. The organization has the option to choose to upgrade the library to keep more tape online or to off-site the tapes as they age. It is critical that the tapes are correctly transported and stored so that, if and when, they are needed in the future the organization can use them.

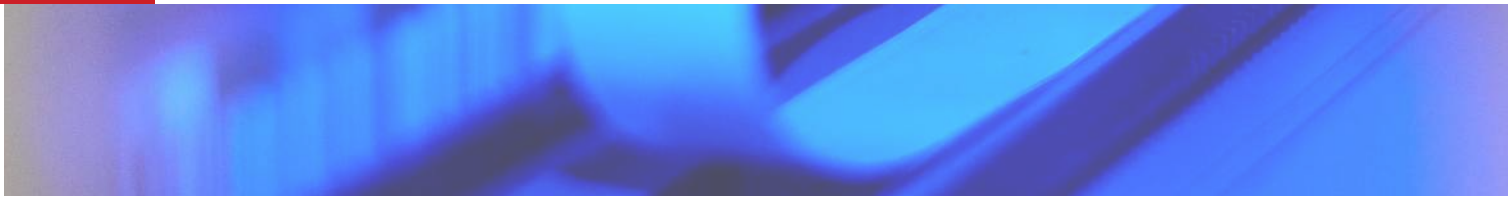
At some point in the life of a tape library, the organization will upgrade from one generation of drives to another to take advantage of increased capacity and performance. These are routine occurrences and most tape drives are backward compatible with previous generations.

CONCLUSION

The savings of implementing tape within the backup infrastructure are numerous; this chapter only scratches the surface. While there is an upfront cost of investing in tape, and potentially a small amount of additional operational overhead, the savings and long-term security of integrating tape into the backup architecture are undeniable.

CHAPTER 5

THE ULTIMATE TAPE ADVANTAGE - COLD OFFSITE STORAGE



One of the biggest advantages to integrating tape media into the backup infrastructure is the door it opens for the organization to leverage cold offsite storage. As we've discussed throughout this eBook, most data occupying capacity on a disk backup device, simply does not need to be there. The unanswered question though is where to put all this data, which the organization either wants or needs to keep.

Organizations can replicate data to a second tape library in an alternate location; they can even replicate that data to the cloud. Both options may make sense if used as an intermediary storage location. In some ways though, using a secondary location or the cloud still doesn't take full advantage of the nature of restores (95% of them come from last night's backup). The chances of data being needed in a restore request continues to decline the older it gets but again, the organization needs to keep this data accessible. Additionally, the older data gets, the less likely is the urgent need for it, with hours or even days being acceptable time frames for the restore.

One of tape's key advantages is its ability to provide 'cold storage.' The specific design of tape, unlike any other storage technology, permits removing it from a powered drive, and then setting it on a shelf to sit idle as the years go by. Thanks to the barcoding systems mentioned in the previous chapter, it is also relatively easy to determine what tapes are offsite and which are on-site.

REQUIREMENTS OF COLD OFFSITE STORAGE

While tape does have an incredible advantage in its ability to be stored offsite in a cold state, IT professionals should carefully consider what type of facility they entrust their data to. The main requirements are that the facility be secure, temperature controlled and have a track record of maintaining data assets for decades. The organization should also make sure that the vaulting service can pick up and deliver media to the client within the required timeframe.

COLD OFFSITE MANAGEMENT

Beyond the physical location and delivery of tapes, organizations need to find vaulting services that can make the process of storing tapes offsite, potentially for decades, as seamless as possible. For example, the vaulting service should leverage the barcodes on tapes to track the location of every tape they manage for the customer. Tape barcodes should be scanned in as soon as the vaulting service driver picks up the tape and as soon as it arrives at the vaulting service's storage facility. Additionally, the vaulting service should again provide exact location of every tape.

The vaulting service should also assist the organization with several consulting options including auditing of the tapes stored offsite for potential adjustment in the number of tapes stored and for how long. The vaulting service may also be able to provide an archiving consultation to expand the use of tape libraries, media and offsite vaults to production data, which actually simplifies the backup process.

One of the biggest advantages to integrating tape media into the backup infrastructure is the door it opens for the organization to leverage cold offsite storage.

HOW TO GET YOUR INFORMATION TO AND FROM COLD STORAGE: DATA RESTORATION AND MIGRATION SERVICES

Organizations may decide to outsource their tape storage to a vault facility with few, several or hundreds of tapes. Beyond the actual storage of tapes, there are additional aspects to consider.

- How do you get tapes to a vault safely and securely?
- What is the impact to the IT budget?
- How easy is it to get information back on-line for litigation or data analytic projects?

Most companies turn to vendors who offer data restoration and managed services to move tapes or other storage devices offsite. As it turns out, this process is highly secure, budget-friendly and a growing consideration point in storage management. Let's look at all three points.

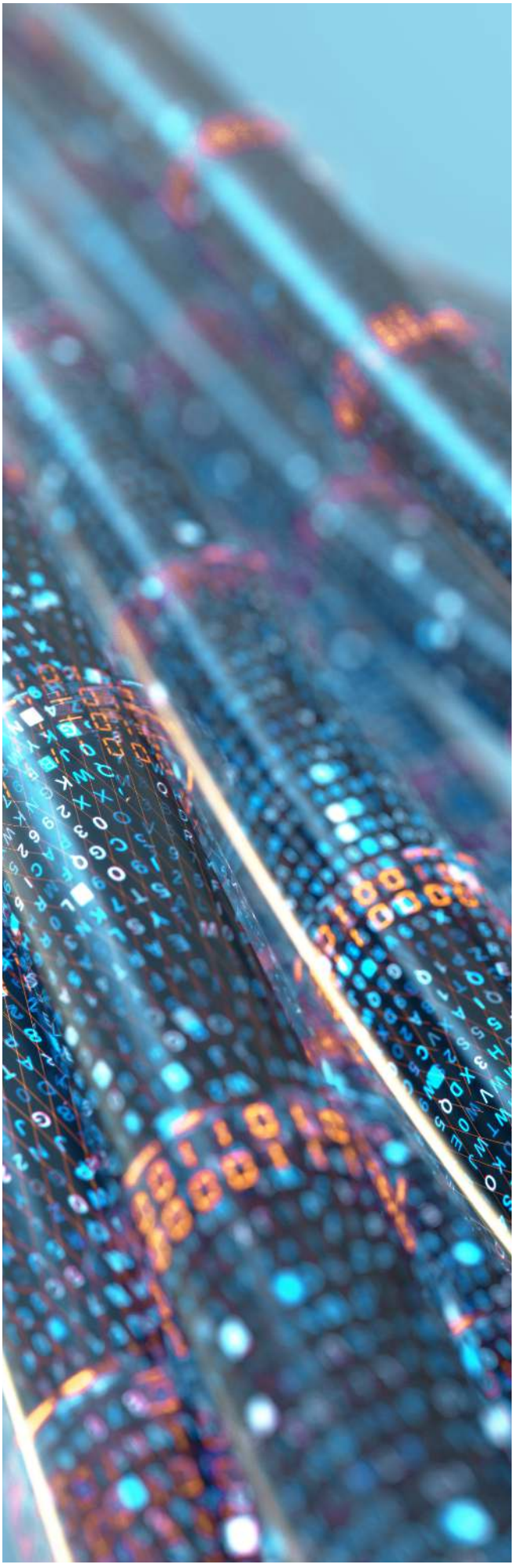
First – security. As discussed in previous chapters, offline data is often highly secure data – and a weapon against ransomware or a fast recovery point from disaster or any attack. If the organization is going to release its tapes to a 3rd party, it needs to make sure security is auditable every step of the way – from pick-up, during transit, and to delivery in a vault.

Second – understand where and how the organization can save on budget. When looking at man-hours saved on tape back-up, money saved on software licenses, and hardware equipment – organizations not only tend to recover budget – but end up in the green vs. red. Also keep in mind, through a managed service model, IT staff is freed up so they can work on more impactful projects rather than being pulled away to work on basic storage and restoration work streams. Look for vendors with track records that are able to quickly locate, restore, and deliver specific information located on tape.

The next is – if the organization chooses to vault tapes, it needs to evaluate data restoration and migration providers so it can simplify, optimize, and accelerate the way its data is not only stored, but utilized and managed.

CONCLUSION

Organizations have the opportunity to cut the cost of both secondary storage and even primary storage, dramatically by better managing the data in those tiered storage repositories. Secondary storage is "low hanging fruit" for organizations to accomplish those goals and tape is an ideal way to get there.



OUR PARTNERS

Storage Switzerland, LLC

Storage Switzerland is the leading storage analyst firm focused on the emerging storage categories of memory-based storage (Flash), Big Data, virtualization, and cloud computing. The firm is widely recognized for its s, white papers and videos on current approaches such as all-flash arrays, deduplication, SSD's, software-defined storage, backup appliances and storage networking. The name "Storage Switzerland" indicates a pledge to provide neutral analysis of the storage marketplace, rather than focusing on a single vendor approach.

FUJIFILM

FUJIFILM Recording Media U.S.A., Inc. delivers breakthrough data storage products based on a history of thin-film engineering and magnetic particle science such as Fujifilm's NANOCUBIC and Barium Ferrite technology. Our mission is to enable organizations to effectively manage the world's exponential data growth with innovative products and solutions, recognizing the social responsibility to preserve digital content for future generations.

SPECTRA

Spectra Logic develops data storage solutions that solve the problem of short and long-term digital preservation for business and technology professionals dealing with exponential data growth. Dedicated solely to storage innovation for 40 years, Spectra Logic's uncompromising product and customer focus is proven by the adoption of its solutions by industry leaders in multiple vertical markets globally. Spectra enables affordable, multi-decade data storage and access by creating new methods of managing information in all forms of storage—including archive, backup, cold storage, private and public cloud.

IRON MOUNTAIN®

Iron Mountain Incorporated (NYSE: IRM) is a global business dedicated to storing, protecting and managing, information and assets. Organizations across the globe trust us to store and protect information and assets. Thousands of local enterprises work with us, as does almost all of the FORTUNE 1000. From critical business information to geological samples, works of fine art to original recordings of treasured artists, our customers can rely on us to protect what they value and help unlock its potential.