SPECTRA

HOW TO STORE AND PROTECT AN EXABYTE

Table of Contents

ELEPHANT IN THE DATA CENTER?

Although you've probably seen calculations on how much data is stored in an Exabyte, it's hard to start an e-book on Exabyte storage without setting the stage for just how much data that is. In the mid-1980s a company was started by the name of "Exabyte." Exabyte Corporation introduced an 8mm tape drive and cartridge for data storage which held an amazing 2.2 Gigabytes. Although the company's name was lofty, it would have taken over 450 million of those tapes to hold an Exabyte.

A single edition of The King James Bible is comprised of a little over 3.1 million letters. If we decided to go digital, we could hold more than 320 billion Bibles worth of text in an Exabyte. Of course a modern day analogy would be more along the lines of digital photographs. At 3 megabytes per photo, if we printed an Exabyte worth of photos and laid them side-by-side, they would stretch more than 48,000,000 miles – enough to wrap around the equator roughly 2,000 times.

No matter how excited we are over Bibles or selfies, we will probably never really need 320 billion of one or 48 million miles of the other. So do we even need to have a discussion on storing, managing and protecting an Exabyte of data?

Most of us are familiar with the phrase, "Calling out the elephant in the room." The "elephant" is that "mostly known" but "mostly ignored" problem or issue that no one wants to acknowledge. Is Exabyte storage the new "elephant" in the data center?

In mid-March of 2020, data backup provider, BackBlaze announced that it stores an Exabyte of user data. What made the announcement unusual is not the amount of data, but rather the fact that they announced how much data they manage at all. Large data centers are notoriously secretive about how much data they store or how they store it. Companies like Google, Amazon, Facebook and Microsoft rarely disclose exactly how much data they manage, but it's been theorized for years that they all hold multiple Exabytes of data.

High Performance Computing (HPC) environments regularly disclose working with hundreds of Petabytes of data. When dealing with hundreds of Petabytes, it takes counting to 10 to get to an Exabyte. And that's exactly what is happening in data centers across the globe. It's not just social media giants, cloud providers and HPC data centers. More and more organizations of every type are working with hundreds of Petabytes, rapidly on their way to a public announcement (or not) of storing an Exabyte or more.

There's never been a better time for discussing exactly how to store, manage and protect an Exabyte – whether you're there already or plan to be in the next five years.

WHAT WOULD AN EXABYTE IN AN ARCHIVE TAPE LIBRARY LOOK LIKE?

Bringing the discussion closer to home, an Exabyte is 1 quintillion bytes or a 1 with 18 zeros behind it. 1,000,000,000,000,000,000 bytes.

In the "good old days" storage products were rated in terms of compressed capacity, but storage software either pre-compresses or encrypts most content or data. The Spectra® TFinity® Exascale Tape Library is the world's largest capacity storage system. The TFinity Tape Library is capable of holding over 1 EB of data based on compressed capacity using LTO-8 or over 2 EB of data based on compressed capacity using TS Tape Technology. This was an industry first. With the announcement of LTO-9 in September 2020, the TFinity now holds an Exabyte of uncompressed data –making it the only tape library in the world capable of holding an Exabyte of uncompressed information – or as Spectra Logic calls it – 'Exascale' storage.

A 45-frame TFinity Tape library, configured with 48 LTO-9 tape drives, holds 55,990 LTO-9 tape cartridges. At 18 Terabytes of native data per tape, this equals 1.008 EB

of uncompressed data in a single tape library. Physically, this is a large library. It measures 109' long and 3.6' deep which means that it will use 391 square feet of real estate. Although the footprint is large, the data density is the best available; a TFinity of this configuration offers 2,558 TB per square foot which is extremely compact for use in a modern data center.

Just as no data center "starts" with an Exabyte of data, the TFinity can be configured as a 3-frame unit and 'transcaled' up from that point. While this e-book explores what it means to store an Exabyte, it will also show various steps along the way in configuration, size and cost.

TFinity 3-frame library footprint

TFinity 45-frame library footprint

Max capacity in tapes/drives

How Does an Exabyte in an Archive Tape Library Perform?

There are many factors to consider in determining the performance of a tape library. Examining the tape library alone – the number of tape drives, the speed of the robotics, and the internal library operating system will determine the maximum performance capabilities. The TFinity can be configured with up to 144 LTO-9 tape drives. The internal real estate of the tape library balances between the number of tape drives and tape cartridges. In our example above with 48 tape drives, the 45 frame TFinity is capable of 19.2 GB/s at the rated native throughput of 400 MB/s per drive. This means that over 69TB of data can be read/written to the storage system per hour. In one year, over 605PB can be transferred into or out of the library. If more throughput was needed, those figures could be tripled by increasing to 144 tape drives.

While disk performance is often touted as better than tape, it's important to keep in mind that an Exabyte of data under management is typically archived data vs. primary, transactional data. Hard disk drives (HDD) certainly have the performance edge on "time to first bit of data," but they would fare terribly against tape in an Exabyte-size archive. Using 16TB HDDs, it would require a bank of 62,500 HDDs to create an Exabyte archive system. The acquisition price of the HDDs alone would likely eliminate them as a contender, but when looking at ongoing costs (such as power and cooling), tape has a clear advantage for long-term archive.

In our tape library example above, with all 48 drives reading and writing at full speed and with the robotics moving at their highest speed, the tape library will still only use 4,300 watts of power. In comparison, a typical disk storage system with 40 to 60 hard drives (3.5-inch) will consume between 1,200 and 1,500 watts of power with all drives and the motherboard running. With 16TB hard drives, each of these disk storage systems will hold roughly 1PB. Four of these disk systems will hold roughly 4PB and will consume the same amount of power as the tape library which stores 250 times more data.

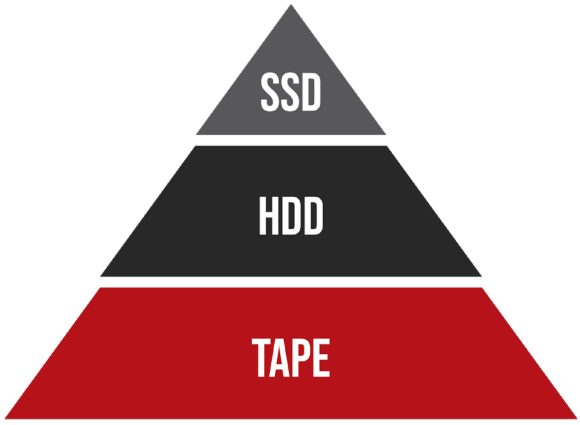
Floor life is also a major concern when discussing an Exabyte of data to archive. The floor life of disk is typically only three years. The floor life of a tape library is 10 to 12 years and the shelf life of tape media is 30 years. While all forms of storage have their place, tape is the clear winner in long-term archive projects.

TIERING FOR EXABYTES

Obviously, organizations have large amounts of data that are critical to their businesses and those data sets are growing rapidly. Often these organizations store all data, active and inactive, on an expensive Primary Tier of storage intended for active data. But upwards of 80 percent of data is typically inactive, therefore, being stored on the wrong tier --costing millions of dollars a year. As organizations approach Exabyte storage, it's important that they reexamine methods of data storage for data tiering. Older tools, such as Hierarchical Storage Management, continue to be used in very large tape archives, with modern tools such S3/Glacier for public and private clouds, and Storage Lifecycle Management, coming forth for new, large storage implementations as well as lower-end solutions.

The Traditional Storage Paradigm

The storage pyramid is one of the most commonly applied storage models in our industry. It's usually represented as three tiers, but could depict any number of tiers or combinations of storage technology entities – cache, RAM, SSD/flash, FC/SCSI disk, SAS/SATA disk, tape, optical, etc. It makes the important observation that the top of the pyramid is the most responsive, costliest, least dense, and smallest amount of storage in the ecosystem. All of those attributes flip as data proceeds down the pyramid. The lowest level of the pyramid is the least responsive, least costly, densest, and typically accounts for the largest amount of storage in the ecosystem.



Traditional Storage Paradigm

While the basic concept of the storage pyramid is as relevant today as it was 30 years ago, it's a model that doesn't address the newer challenges of modern storage – especially as we approach an Exabyte.

With the introduction of the public cloud and object storage technologies, the hierarchical nature of the traditional paradigm becomes less effective. This is especially the case when different storage technologies are used in similar roles – SSDs and HDDs both used in the top tier; disk and tape both used in backup; and tape and cloud both used for disaster recovery (DR) and offsite storage. The roles of these technologies may be similar, but there are granular differences that enable them to meet the demands of individual data centers and significantly offset costs if those differences can be accounted for.

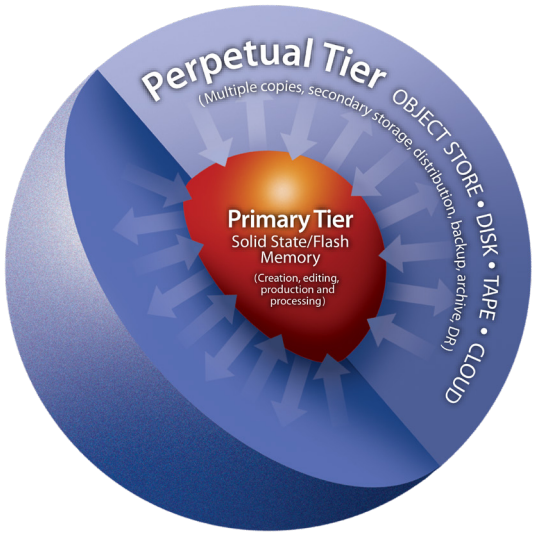
Likewise, today's storage model must consider the advent of new storage formats. As object storage enters the mainstream, there are many questions not answered by historical storage models. Does object storage apply to a single tier, or do we see block, file and object storage being used across multiple tiers and intermixed?

A Two-Tiered Storage Model

Groundbreaking storage management software has enabled a modern storage paradigm based on a two-tiered storage model. Rather than focusing exclusively on the storage medium, this model is based on the data or digital content that is actually being stored. We start by classifying data into two categories – "active," meaning it's being edited, processed or changed in some way, and "Inactive" which quite simply refers to everything else. This results in a Primary Tier for the active data and a Perpetual Tier for inactive data.

The Primary Tier holds all active data and is most commonly composed of flash / NVMe /solid state storage. By moving inactive data out of the Primary Tier and into the Perpetual Tier, organizations can significantly decrease the size of the Primary Tier. This allows administrators to better configure this tier using a combination of high-speed storage mediums in order to achieve the performance required for workflows associated with highly active data.

The Modern Storage Model



The Perpetual Tier is dedicated to inactive data and is designed to keep multiple copies of data on multiple storage mediums including NAS, object storage disk, cloud and tape. While the data is not considered "active" on the Perpetual Tier, there is quite a bit happening at this level. The Perpetual Tier is used for secondary storage; distribution; multiple copies (a responsive copy and DR copy); backup; archive; project archive; and traditional disaster recovery. The Perpetual Tier is clearly the area in which Exabyte archive should reside. As mentioned above, there are multiple mediums in the Perpetual Tier. When it comes to archiving an Exabyte or even 1 percent of an Exabyte, tape has multiple advantages.

PROTECTING AN EXABYTE

Archives are created for various reasons. While sometimes overlooked in value, they are the very essence of both our history and our future. In the words of author and poet, William Feather, “The wisdom of the wise and the experience of the ages is preserved into perpetuity by a nation’s proverbs, fables, folk sayings and quotations.” We will examine how archives are used under Use Cases later in this e-book. Regardless of the intended use – archives are meant to last, to be enduring into perpetuity.

Therefore, data reliability is paramount in digital archives. Individual tape drives have actually surpassed the bit error rate reliability of hard disks. LTO-9 tape boasts a bit error rate of 1 x 10^{19} while IBM’s TS1160 tape technology is at 1 x 10^{20}.

Undetectable bit error is another aspect of data reliability and even a greater one on which to focus. Unlike bit error rates mentioned earlier, an “undetectable” bit error is often referred to as “bit rot,” and means data corruption. Tape has a tremendous advantage over disk in this category as well. Tape boasts an undetected error rate of a single bit for every 1.6 x 10^{33} bits it reads. To put that into perspective: If you had one million tape drives and one million disk drives running simultaneously, you would get a single undetectable bit error on tape once every five times the age of the earth (total of 22.5 billion years) – in other words, 1 in 22.5 billion years. In comparison, you would have 1,577 undetected bad sectors every single year with disk.

Tape Medium	Hard Error Rate
LTO-8	1 x $10E^{19}$ bits
LTO-7	1 x $10E^{19}$ bits
TS1155	1 x $10E^{20}$ bits

Hard disk reliability is compensated with technology such as RAID. Tape reliability is significantly increased by creating multiple copies. For all practical purposes, “two copies on tape” could be considered mirroring data. While multiple copies of data on different mediums is always recommended, at half a penny per Gigabyte, tape cartridges are the most cost-effective way to assure data availability.

STaRR – A Compelling Approach

Why keep an Exabyte of archive data around? The primary reason is so organizations can meet compliance and regulatory requirements. The other purpose is that between one and five percent of disaster recoveries come from this archive, and it is almost impossible to know what percent of that one to five percent of that Exabyte of data 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.

While the cost of not being able to restore from an archive is tremendous, a well-planned archive is rarely accessed for restores. This introduces an opportunity for extreme protection at minimal cost.

Keeping multiple copies on tape eliminates the risk of data loss though the failure of a tape. Modern tape cartridges have a low failure rate of 1 out of 100,000 per year. Low as it is, it could result in a failure rate of about

one cartridge every two years in our Exabyte-scale library. Implementing a two-copies-on-tape policy decreases the odds of losing data to tape failure to 1 in ten billion per year. With each additional copy on tape, the odds decrease even further. And again, the financial burden is one of the lowest in data protection available.

The Spectra TFinity offers enough tape slots and storage capacity to house both single and multiple copies of data onsite.

> **STaRR or 'Store Twice and Read Rarely' is a compelling way to use tape for increased reliability, security and availability.**

Another advantage of tape is its "removable" nature. Sending one copy offsite provides geographic separation, further assuring business continuity in the event of natural disaster – be it flood, fire, earthquake, or other. Historically disaster recovery referred to just those sorts of disasters. Today, all organizations have to be mindful of cyberattacks. Ransomware has become the leading form of cyberattack, and the most damaging for organizations that do not have a strong disaster recovery strategy in place.

Storing a disaster recovery copy of data in the cloud creates geographic separation, but it doesn't eliminate the possibility of ransomware's encryption or destruction of data. Where there is an electronic path to data, there is also vulnerability. A true disaster recovery archive must be both offsite and offline. Tape offers a true "air-gap" between data and the cybercriminals' attempts to extort it. Even if the backup or archive server is encrypted, and they are often the first to be attacked, tape can rebuild data from scratch. It is not an easy or quick task, but it is a fail-stop measure that assures an organization can indeed recover.

User-Based Encryption

Data breaches in which data is copied or stolen continues to be a threat. Be it cyber-espionage or identity theft, data breaches are extremely costly in both soft costs and hard costs. An organization's reputation as well as stock price (if public) are significantly damaged after such breaches. These breaches are a great liability, both financially and legally. There are various industry, federal, state and world organizational laws which mandate how certain information is handled.

Just as cybercriminals use encryption to prevent organizations from being able to access their own data, organizations can use encryption to assure criminals cannot see or reproduce organizational data. The Spectra TFinity offers multiple encryption solutions to choose from. Spectra Logic is the only library manufacturer to offer encryption key management fully embedded in the tape library. This avoids the cost and complication of requiring additional servers, applications, support agreements, etc. Spectra's BlueScale software offers two versions of encryption key management – BlueScale Standard Encryption Key Management and BlueScale Professional Encryption Key Management. Standard Encryption is offered with all libraries as a function of BlueScale, free of charge.

For more complex environments, Spectra also offers externally managed encryption solutions. Our Spectra Key Lifecycle Manager (SKLM) is often selected when multiple libraries or data centers are being managed from a single area. This does involve a server or VM for hosting, but allows additional feature sets. Our SKLM solution is FIPS, KMIP, and IKEv2-SCSI compliant. A few of the additional features found in SKLM are more extensive audit trails, key grouping, assigning a single key per tape if desired, separate key states and other features which target policy-based management. Spectra libraries also support the Hewlett Packard Enterprise (HPE) encryption solution, Enterprise Secure Key Manager (ESKM).

Locking Down an Exabyte

Just as it is important to safeguard the contents of a tape via encryption, it is equally important to safeguard the tape itself. Large tape libraries may hold tens of thousands of tapes. The ability to lock these libraries is critical. The Spectra T950 and TFinity come standard with traditional locks on the back service panel. Spectra has also introduced the use of CyberLocks.

The CyberLock pictured above is best described as a "key-centric" control system designed to increase security, accountability and key control on a single or multiple libraries. It's based on a unique design of electronic lock cylinders and programmable smart keys – think "key fobs."

Spectra's T950 and TFinity libraries have the option of adding locking side windows as well. For an even greater level of security, CyberLocks can be added to all library access panels. This gives users the convenience of a mechanical key system in addition to the access permission and tracking capability of an electronic access control system.

For organizations that already use CyberLocks, the locks added to the T950 or TFinity library are easily programmed to work within their existing CyberLock infrastructure.

MAKING EXABYTE TAPE ARCHIVES EASY

New tape technologies are more reliable, offer greater capacity, greater speed and are more affordable per Gigabyte than ever before, but tape must fit into the modern data center to be viable for storing "Exascale" archives.

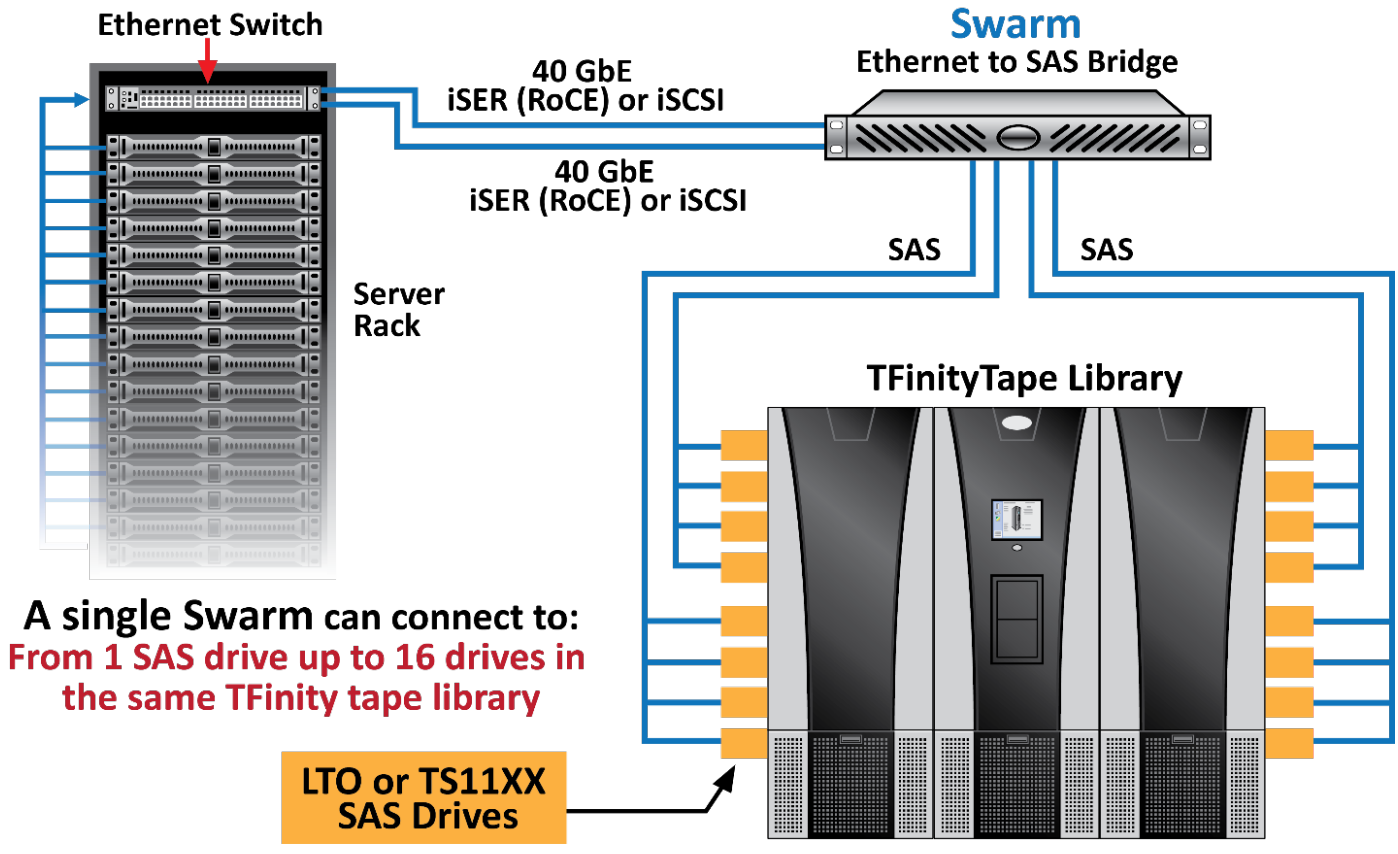
Exabytes over Ethernet? Multiple Physical Interfaces

The storage industry is filled with "end of life" predictions between various storage approaches, devices and interfaces. Ethernet vs. Fibre Channel (FC) is another such debate. While both technologies have their place, Ethernet is clearly gaining ground in some of the more rapidly growing areas of the storage market. With speeds of up to 100 Gb/s, Ethernet has closed the differential in performance with Fibre Channel, but speed is not the driving force behind the broader adoption of Ethernet.

Due in large part to the growth of social media, the internet of things (IoT) and cloud, File and Object Storage are growing at a much faster rate than block storage. Ethernet's ability to easily handle both storage types (which Fibre cannot) make it an ideal fit for the largest data growth markets today. Other developments such as hyper-converged infrastructure – which combines storage, compute and networking into a single scale-out layer of servers without the need for expensive Fibre switches – also bodes well for the versatility of Ethernet.

The lower cost of Ethernet HBAs and switches, use of existing data center infrastructure and no dedicated personnel needed for Fibre SANs are the final piece of the story. The combination of performance, versatility, ease of use and lower cost have made Ethernet a mainstay in data centers of all types and sizes – especially those working towards an Exabyte of storage.

Continuing its role in tape technology innovation, Spectra has introduced Swarm – Ethernet controllers for tape technology. The economic advantages of Ethernet are now available to tape users as well. For data centers only running Fibre Channel to support tape, moving to a single protocol and being able to share switches and other infrastructure offers significant cost savings not to mention the overall increased ease of operation and lower management overhead. Spectra Swarm supports LTO-7, LTO-8, LTO-9 full height drives and all future SAS tape technology. Spectra also supports the new IBM® TS1160 SAS tape technology which works well in the Swarm environment.



Obviously, Spectra continues to support FC and/or SAS interfaces as well. Having the option of multiple physical interfaces not only makes the Spectra libraries easy to work with today, it also assures "future- proof" access to Exabyte archives as technology evolves.

Multiple Tape Technologies

LTO and IBM TS Tape Technology are the only two viable choices for today's Exascale archives. This situation is complicated however by Oracle's departure from manufacturing the Oracle® T10000 tape drive. There are hundreds of thousands of those tape cartridges still under management with no viable way to continue to support that technology.

Spectra's TFinity is the only tape library available which can support LTO, IBM® TS tape technology and the Oracle® T10000 tape technology. As made clear already, an Exabyte of data is comprised of thousands of tapes. The TFinity makes it simple to import all existing tape, regardless of whether it's LTO, IBM® TS or Oracle® T10000. The TFinity is easily partitioned to allow for the coexistence of these very different technologies. Data can be migrated across mediums in the background while current operations are underway.

Spectra's TFinity Tape Library supports three tape technologies

Multiple Architectures

Historically, tape architecture was limited to traditional file system storage. This sometimes created difficulty in both scalability and workflows. Spectra offers File System, Object Storage and even Transparent Access architectures to make tape extremely versatile in the age of Exascale archives.

File system storage has been a mainstay of the storage industry for decades. While that may sound like "old" technology, it's still the predominant form of storage and is a proven architecture for many Exascale archives. File system storage is most commonly found in traditional backup/recovery or disaster recovery software applications (more on application support in the following section).

Object storage is a way to structure data storage, similar to a file system, but comes without the limitations to growth and performance found in traditional file systems. Object storage was a mandate for most cloud storage where trillions of files needed to be stored in a single name space. File systems simply could not scale to that level.

Objects are similar to files in that they contain a single file, document, picture, etc. They differ greatly however, by their use of a unique object ID for tracking and organizing. File systems contain not only data, they also contain the metadata, physical location, access rights and other information pertinent to that piece of information. This is very useful for transactional/high edit data, but becomes an unnecessary burden in long-term archive situations.

Object storage is ideal for Exascale archives, but until recently, object storage management was not available for tape. Spectra has changed that scenario. More on object storage for tape in the application section that follows.

Transparent access to tape is accomplished by software running in conjunction with a tape library. This comes in two broad forms, HSM software (Hierarchical Storage Management) and Data Storage Management software. Both approaches offer more seamless access to tape, further expanding the workflows which can be accomplished with tape. More on this subject will be covered as well.

Spectra's tape libraries are designed to support all of the above architectures, making it an ideal solution for archives of any size, but the best solution for archives in the multi-Petabyte to Exabyte size.

BROAD APPLICATION SUPPORT

In the previous section, we briefly introduced key architectures. In this section we look at specific solutions which run in conjunction with Spectra tape libraries to create such architectural approaches.

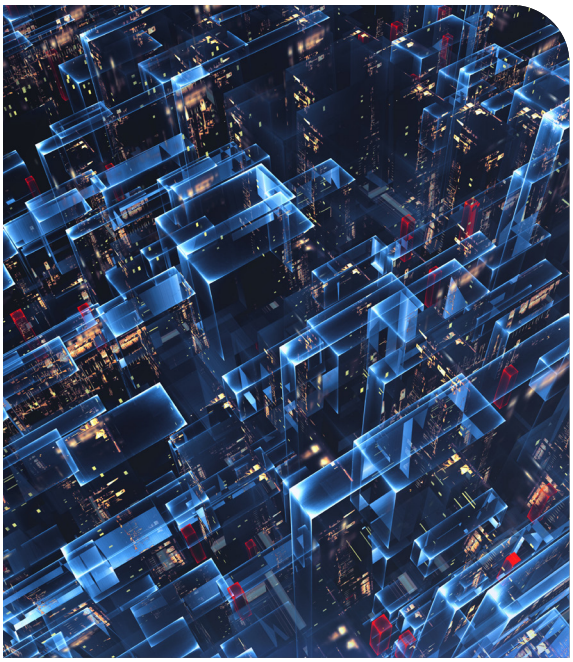
If you are working with an open systems application that supports tape, Spectra libraries most likely support that application. Spectra has the broadest range of application support available in this area. Often a single library will be used for multiple applications. As the libraries become larger, this aspect becomes more important and a significant opportunity to increase efficiencies and decrease cost.

Where Spectra differentiates from other tape approaches to large archives is in its ability to support many applications that haven't historically supported tape. More on that below.

Shown below are a small sampling of the myriad of applications that Spectra libraries support.

Software applications are sometimes hard to categorize into a single category. For purposes of exploration, we will break this discussion into the following categories.

1. Traditional Backup/Recovery & Disaster Recovery
2. Enterprise HSM Archive
3. Modern Data Management Software Archive
4. Object Storage Archive



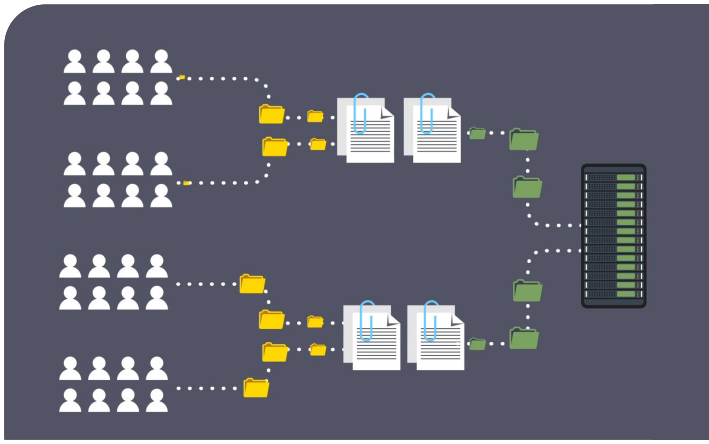
Traditional Backup and Recovery and Disaster Recovery

This is the only category listed above which does not end with "archive." Often the terms we use in discussing storage are interpreted differently by different users. In this e-book, "backup" refers to making a "copy" of the original data while leaving the original data in place. Active/transactional data should obviously remain on primary storage for greater speed of access. The term, "archive" refers to moving the original data from primary storage to another location, often referred to as migration. This approach is typically used for migrating less active data to a more appropriate storage tier such as the Perpetual Tier of storage mentioned earlier.

The backup applications discussed below often have some archiving capabilities as well, but tend to focus more heavily on the coping of data vs. migration of data.

Backup and recovery is undoubtedly the largest category of software support we will cover. In the past, tape was the mainstay of backup and recovery. In the lower end of this market, higher density disk drives and public cloud have reduced tape's presence

due to ease of recovery. As we discuss the higher end of this market, meaning greater amounts of data, tape still plays a role. However, increasingly, tape has a larger role in archive. In Exabyte archives, tape is actually the leading technology of choice.



A few of the major backup application players are Veeam, Commvault, Veritas, Dell/EMC Networker, and IBM Spectrum Protect. There are many more, but these are the names usually found in the top five for “application-based” backup.

There is a new breed of backup software we will deem “appliance-based” backup. In this scenario, appliances with high-speed storage (flash or disk) act as buffers or a cache, which then move the data to longer term storage including disk, tape and/or cloud. Rubrik and Cohesity are two of the more well-known names in this category.

All of the above solutions, and dozens more, are supported by Spectra libraries. It is not uncommon for tape libraries to be “partitioned” to appear as multiple, virtual libraries. In this manner, an Exabyte-size TFinity tape library could serve as the repository for multiple applications and workflows. A single organization could use one application to backup Windows-based systems, a second application to backup Linux/Unix-based systems, and even a third application or approach for archiving – all going to a single TFinity library. As mentioned earlier, Backup and Archiving serve different purposes, so this approach is actually not uncommon.

Enterprise HSM Archive

Today’s HSM market has fewer vendors than the backup market. The names we often see are IBM’s HPSS, HPE’s DMF, and Versity’s VFM. These solutions start our discussion on “transparent access to tape.”

HSMs try to map nearline storage to appear to be online, including tape storage. This is tricky given the sometimes long latency of nearline storage. Applications will time-out if they don’t get the data requested in the time they expect to get it. To accomplish this, HSMs typically use stub files (along with filter drivers, but we don’t need to go that deep). The stub file looks like the original file and often contains the beginning of the original file. The stub file can respond to the read request while the HSM retrieves the rest of the moved file and brings it back – this is no small feat.

HSMs become part of the file system, usually have kernel code, are very operating system (OS) dependent, have to be upgraded with the OS, and offer no ability for the application or the user to see that a file has been moved.

For the above reasons, HSMs don’t play well in all environments. They do, however, play extremely well in High Performance Computing (HPC) environments where the operating systems – like Lustre, GPFS, etc. – are more “time-out tolerant” or “HSM-aware.” Successful HSMs such as those listed above are extremely effective. They are also expensive, complex and require a lot of resources – but when you need them, nothing else will do.

In contrast, HSMs introduced for the general IT market were not as successful. We saw many such HSM solutions introduced from the late 1990s through the mid-2000s. Few if any of those HSMs exist today.

Modern Data Management Software Archive

Generally speaking, the solutions we see today in the category of Data Management Software differ from enterprise HSMs in that they require less budget, headcount and infrastructure, and sit well outside of the file system. There are some exceptions, but this is a good categorization of modern Storage Management Software.

These packages tend to be less complex and are generally compatible with a larger range of applications and use cases. There is a wide range of feature sets among this “class” or “group” of applications. Some of the more well-known names are Arcitecta’s MediaFlux, StrongBox’s StrongLink, Komprise, and Spectra’s StorCycle.

Symbolic links and/or HTML links are more likely to be used to find the moved data vs. stub files. These links work quite differently from stub files as well as from each other.

By leaving a symbolic link in place of the original file, a read can simply be redirected to where the file has been moved. This works great when moving infrequently accessed data off of primary storage (high speed disk or SSD) to a lower tier of storage like network-attached storage (NAS) disk. Most applications can tolerate the small increase in latency. However, this methodology does not work well with tape or low-response-level cloud. That’s where the HTML links come in.

Spectra created HTML links specifically to support storage mediums with longer latency, such as tape or low-response cloud levels. This is a unique feature of Spectra's StorCycle. When an HTML link is left in place of the migrated file, the user is presented with an HTML page which states that the file has been archived, gives information about when, where and how this was done, and allows the user to start a restore from tape or a recall from cloud without having to contact IT.

Again, the feature sets found in each of these solutions vary greatly. All of them support the Spectra TFinity, and any of them could be used with Spectra's TFinity to create and manage an Exascale archive. Individual use cases, IT initiatives, budget and desired management style are a consideration when evaluating the appropriate Data Management Solution for any given site.

Some commonalities in this group include the fact that they all focus on data protection as well as data archiving. While enterprise HSM focuses primarily on data tiering, data management software also incorporates the concept of "multiple copies in multiple locations" and typically works quite well with object storage.

This can be a very effective form of data protection. Having copies check-summed and stored on disk/tape/cloud (or any combination thereof) helps to ensure that a valid copy of the data can be retrieved when needed. Likewise, administrators can direct a file to be moved to multiple locations and retrieved from whichever location is most efficient.

Another commonality is that these software packages are all integrated into Spectra's BlackPearl® Converged Storage Solution. This allows for the very unique ability to bring object storage management to tape, and a significant advancement in creating and managing Exascale archives.

Object Storage for Tape

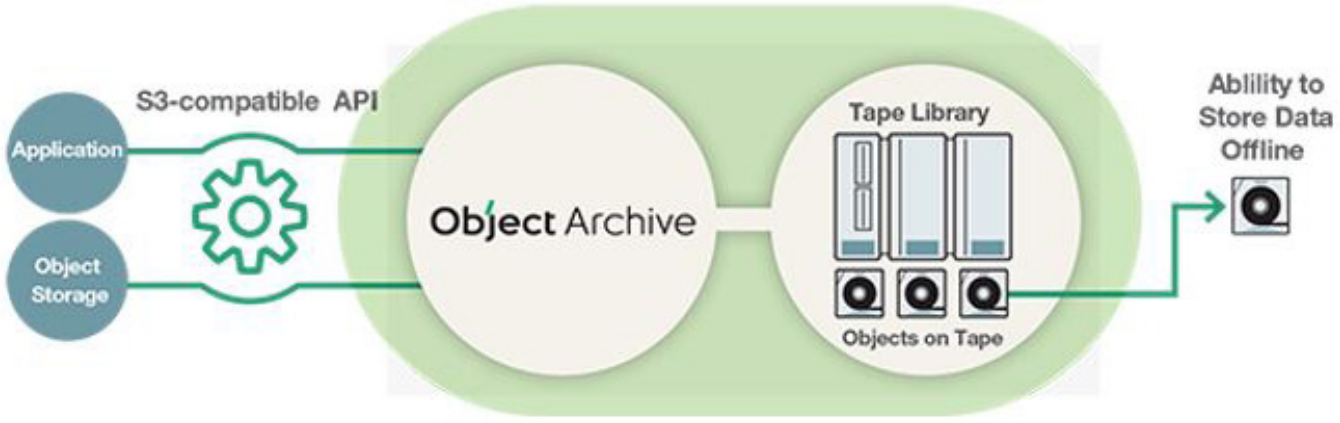
Object storage is arguably one of the most exciting developments for archiving in the past two decades. The concept came about in the mid to late 1990s and has developed rapidly since the early 2000s. An entire whitepaper could be dedicated to this topic alone, but a working definition for purposes of this discussion follows.

Object storage is best described as an architecture that manages data as objects, as opposed to other storage architectures like file systems which manage data in a file "hierarchy." Instead of embedding each new piece of data in the file hierarchy, along with other files, directories, subdirectories, folders, etc., object storage stores each piece of data independently using a unique object ID.

This creates a flat structure (vs. hierarchical) which allows for amazing scalability, independent addition of metadata, and tiering. Interfaces can be directly programmable by the application, namespaces can span multiple instances of physical hardware, and most importantly, for this discussion, data management functions like replication and data distribution can occur at this much more granular, object level.

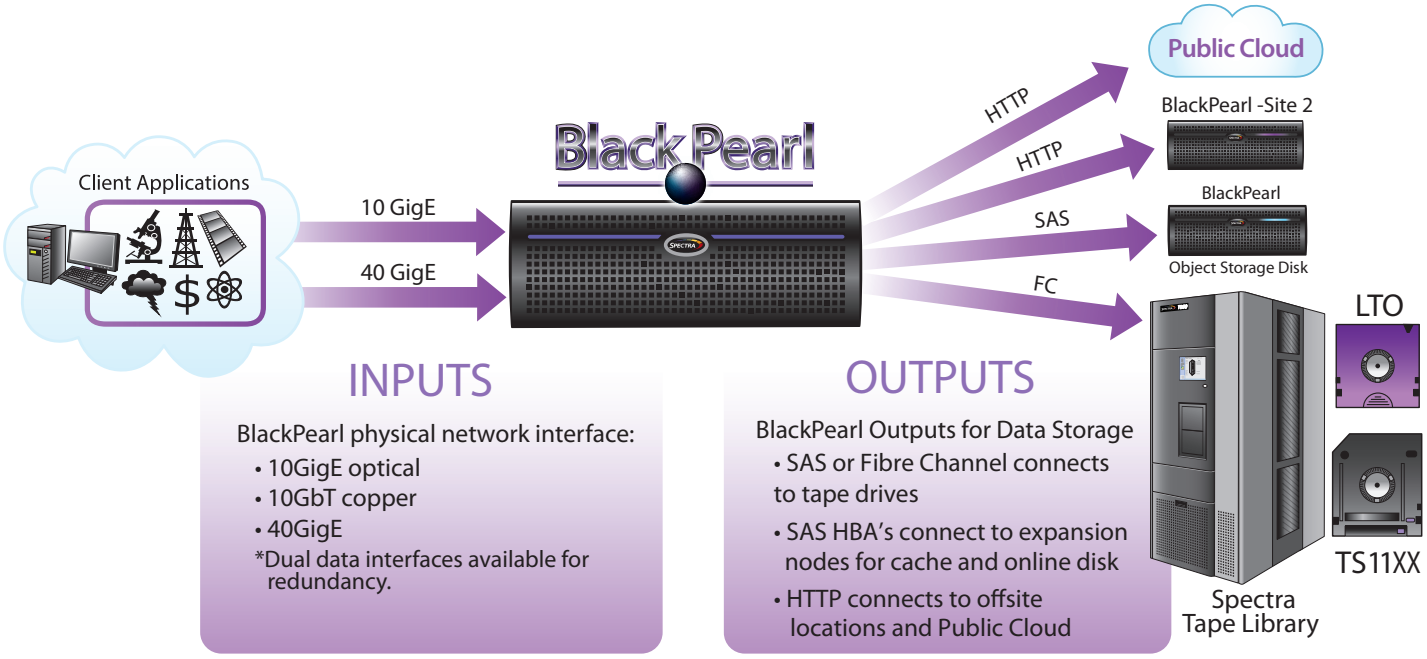
For these reasons, object storage has become the de facto standard for storing information in the cloud. It's only recently that object storage for tape has been introduced, bringing the best of both worlds together – simple, long-term, data storage management with low-cost, long-lived tape automation. This is key for Exascale tape archives.

FUJIFILM announced a new solution in May 2020 called, FUJIFILM OBJECT ARCHIVE. This software solution has a new tape format for object storage called, "OTFormat." OTFormat allows for objects and metadata to be efficiently written and read to and from tape in native form.



FUJIFILM OBJECT ARCHIVE workflow

Spectra also offers object storage via the BlackPearl Converged Storage Solution. BlackPearl is an intelligent object storage gateway and data management system that allows files to be written to tape, NAS or cloud as objects. The BlackPearl approach to object storage for tape is unique in that the files are managed as objects, yet they are written to tape using the non-proprietary LTFS format. Users are able to take advantage of scaling and fast indexing/searching – features of object storage. Simultaneously, tapes are generated under the LTFS format allowing for complete interoperability with other systems. This assures data is accessible in the future with or without a Spectra solution – providing a completely non-proprietary archive.



Data can be archived via the BlackPearl object storage solution through several paths or workflows.

Spectra provides a simple, published API for BlackPearl which allows software manufacturers and end users alike to interface to BlackPearl. The interface is called, RioBroker®. Commercial applications, such as those listed under data management software above, all offer support for BlackPearl to make writing to tape very simple in an object storage environment.

Many applications in media and entertainment have integrated support for BlackPearl as well. These edit suites have not offered direct tape integration in the past, requiring additional expensive software to be added in order to support archiving and tape archiving. For those integrated with BlackPearl, it's now as simple as selecting the "archive" button.

Over 30 commercially available applications now integrate into BlackPearl. It's so simple, users can even write their own interface to customize how they archive or backup data.

BlackPearl's RioBroker interface can be used for migrating data out of proprietary solutions into open, non-proprietary solutions as well. In the media and entertainment market segments, RioBroker will import data from Front Porch Digital's DIVA solution as well as the SGL FlashNet solution, allowing users to replace older, expensive, proprietary archives with a simple, open, modern design at significantly lower cost.

Furthermore, BlackPearl utilizes hypertext transfer protocol (HTTP). HTTP is the underlying protocol used by the World Wide Web. It defines how messages are formatted and transmitted and is capable of working with delays in response. This is an ideal protocol for moving, retrieving and sharing data with higher latency storage like tape and/or cloud. API's based on the HTTP protocol are now commonly called REST or Restful interfaces.

When building a multi-Petabyte or Exabyte archive, broad application support is critical. Spectra's TFinity offers the largest selection of supported vendors available. From commercial backup applications to HSMs to data management solutions to custom interfaces, TFinity is designed for archives in any environment imaginable. As new applications become available, TFinity's modern interface assures simple integration.

USE CASES FOR AN EXABYTE ARCHIVE

It's all too easy to become mired down in the "bits and bytes" of constructing a large archive and lose sight of the reasons the archive is being built and the value it can bring to organizations across the world. A few months prior to the writing of this paper, NASA ran the following headline in a news article: ***Earth-Size, Habitable Zone Planet Found Hidden in Early NASA Kepler Data***

The Kepler space telescope was launched by NASA in early 2009 to discover Earth-size planets orbiting other stars in or near "habitable" zones. In other words, discover planets that humankind may be able to live on. The telescope was retired in late 2018, two years ago. So why is it making headlines today?

A group of scientists from the University of Warwick created a machine learning algorithm to dig through old NASA data containing thousands of potential planet candidates. This new technology was able to identify a planet very similar to earth in size and temperature that is theorized to possibly support liquid water. Is this the beginning of a science fiction movie? Hardly. This is the result of new technology being applied to old data. And new worlds are emerging.

That is the power of archives. And because of that, archives are growing larger as more data is being generated and more and more organizations are setting their retention policies to "forever."

Not all archives are used for such "earth shattering" discoveries. From scientific research to manufacturing, from modeling to weather predictions, from entertainment to world history, from medicine to self-driving cars – digital archives help us understand our world, keep us safe, get us where we need to be, and even entertain us when we get there. The following use cases are but a few examples of how organizations in every walk of life are creating archives, breaking every previous record for size and demand.

BioPharma Research

Historically, drug research has depended on traditional, randomized clinical trials to determine the effectiveness of both new and existing drugs. As one might imagine, this research is extremely time-consuming and expensive. In many cases, this process can delay drugs from reaching the market by years.

The rapid development of data analytics now allows drug companies to study what is being termed “real-world evidence.” That’s any evidence of a drug’s effectiveness gathered outside of the traditional clinical trial setting. But where does that “evidence” or data come from?

The last two decades have seen a tremendous growth in the digitization of patent records. Now combine that with insurance records, diagnostic and genetic testing, fitness “wearables”, IoT sensors, and even social media. Data is being extracted from all of these areas in an unprecedented effort to speed the safe development and delivery of life-saving drugs.

Obviously privacy is a primary concern in using such data, but by anonymizing this information, suddenly millions of people from every walk of life can contribute to drug research. This isn’t predicted to replace clinical studies, but it has already rapidly advanced new medicines and given much greater insight to existing medicines as well as offering greater “individualized” medical treatment.

Breast cancer is not common in men, only about 1% of breast cancer cases occur in men throughout the U.S. That still accounts for roughly 2,000 cases per year in the U.S alone. Ibrance, developed by Pfizer, is known to be an effective medicine in controlling breast cancer. When the initial clinical trials were done, those trials did not include male participants. Therefore, Ibrance was not approved by insurance companies for male patients. To conduct the study in men would have taken another three to five years. Looking at the averages, that’s another 6,000 to 10,000 men who would not have access to this drug during those trials.

Pfizer was able to look at the original study results performed on women and supplement those results with data generated by male users (the real-world evidence) who had paid for their own medication. Within 12 months Pfizer was able to prove to the U.S. Food and Drug Administration that Ibrance was safe and effective for both sexes.

Additionally, by significantly lowering the cost of entry into a drug market, pharmaceutical companies are able to do drug development in markets that aren’t as lucrative to enter such as diseases that disproportionately strike poorer nations.

The role of big data in our current mission to identify, treat and eventually prevent COVID-19 can’t be underestimated. Digital modeling has been used to understand how COVID-19 spreads and where future hotspots will be. Artificial Intelligence (AI) has been used to identify the proteins associated with the virus that causes COVID-19. In April of 2020, Harvard’s School of Public Health and the Human Vaccines Project announced a new initiative to use artificial intelligence modeling to accelerate vaccine development for COVID-19.

This all depends on access to tremendous amounts of data. Genetic testing is another area that has provided incalculable information for such research. Genetic testing produces around 300 GB per person. It would take 99 Exabytes to hold the genetic testing of the U.S. population alone.

No matter how much data BioPharma Research companies have access to, more will produce faster and more effective solutions to disease prevention, control and cure. In this context, a single Exabyte archive seems small.

Large Video Archives

A global pandemic doesn't just impact those trying to combat it, it also impacts organizations trying to continue business operations through the single largest "shut-down" in human history. The Media and Entertainment industry has been severely impacted by this. Sports events have almost entirely stopped, film projects were abruptly placed on hold (impacting post and production), news events were harder to reach as travel restrictions were implemented, and virtually everything that drives this industry was impacted. The ripple effect caused by the lock down will likely affect media producers not only during 2020, but into 2021, and potentially beyond.

Archives of digital content have become a huge part of how participants in this industry have pursued business continuity, keeping viewers, quarantined at home, happy with the simple click of a button. When a media outlet owns rights to content, there are no licensing fees – and that content can be replayed without additional costs. Offering vast repositories of movies, sitcoms, documentaries and digital series keeps existing subscribers tuning in while helping to attract new audiences. Additionally, owned content can be licensed to other networks and broadcasters to create additional revenue streams.

All that said, consumers are craving new content. The demand mounts daily. As the pandemic calms and restrictions are lifted, broadcasters and producers will inevitably race to create new content to grab the attention of, and please, the masses. Recording will ensue and quickly result in an explosion of data. Many outlets currently use hi-res 4K cameras that will contribute to massive storage demand. Depending on the number of cameras used to shoot, image resolution, hours recorded, frames per second, image bit depth and compressed type used – shooting in 4K quickly leads to pressure on data storage. For example, a finished two-hour film shot at 4K with 172,000 frames at 24fps creates roughly 50MB of data per frame - approximately 10TB for just the finished film. When you look at the ratio of raw footage hours to final footage (typically 6:1 or 7:1), you are looking at around 60TB to 70TB created for each two-hour film.

One can see how a network producing a slew of reality shows could exceed an Exabyte of total storage if given the sheer number of hours it takes of raw footage to produce one hour of finished content. The ratio is unheard of – ranging from 100 to 500 raw hours to produce a mere one hour of final video. High-end ratios tout a staggering 1000 raw to 1 finished ratio. When these shows air weekly, we're talking about massive amounts of data being created!

As if the data creation associated with 4K wasn't nearly enough, 8K looms on the horizon, and the tremendous amounts of data created with it is staggering. Shooting at such high resolution allows broadcasters to show "finger-tip" finishes to their audience, creating a highly realistic experience for viewers.In fact, the Olympics are currently shot in 4K, but will soon be taking the leap to shoot in even higher-resolution 8K. All of that comes at a price. Combine 8K resolution with higher density

and higher frame rates and you have the recipe (perfect storm) that would cause any content creator to breach the realm of an Exascale archive. In comparison to shooting in 4K, 8K films nearly quadruple the amount of data created when shooting. You can expect somewhere around 200MB/frame for 8K DPX – meaning a finished film will be around 35TB. Then take into effect the raw footage to final footage ratio and we're talking around 245TB of data.

Considering the exponential storage requirements of shooting in 8K resolution, it is easy to see how quickly a broadcaster, etc., can easily reach the realm of an Exabyte archive.

To stay on top of the changing times and bring new content to consumers, it will be essential for media and entertainment organizations to plan ahead for larger archives that can store all of their invaluable raw and finished footage.

Research Data and Onsite "Glacier-like" Archives

Spectra works with a large university that supports the research efforts of over 50 different groups within the university. They offer various service level agreements (SLAs) based on the performance of the storage. Each research group is billed for the amount of storage they use based on the SLA they select.

The university has standardized on three storage performance levels: a high speed solid-state-disk-based tier; medium-speed, disk-based tier; and a NAS-based tier targeted for archive of projects. There are multiple challenges for both the university and the individual researchers that the university would like to overcome.

While NAS is the lowest cost repository in the current storage model, archiving large amounts of fixed content becomes extremely expensive over time. Researchers have asked for a lower cost solution for archiving. The data center basically acts as a "cloud" provider, but they lack the last level of storage often referred to as "glacier-like" based on AWS Glacier storage. The university would like to introduce an Exascale-sized archived based on tape, but they have no way to introduce "rule-based" file movement across the storage infrastructure to such a tape archive.

The university actively encourages the researchers to move data off of the high-speed Primary Tier with a bill-back system that offsets their costs. However, when the university runs out of high-speed storage, they don't always have the funds to acquire more before the offsets come in, which can hamper research efforts.

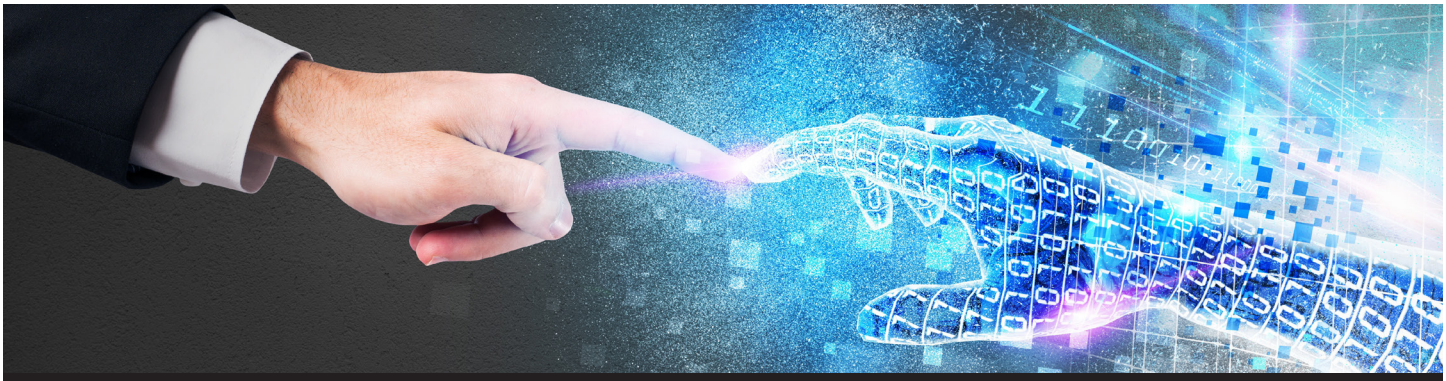
The researchers would like to move their data to a low-cost storage tier, but they have a challenge to accomplish this. The data can be human-generated, application-generated or machine-generated, and has been accumulating for years. The researchers have access to the data, but they have no way to identify what is actively used, what needs to be archived, and what data is orphaned. And if the data is manually moved, how can it be accessed after the researcher leaves the university? Grants for research often require storage of the data for periods longer than the researchers who work for the lab or university involved.

A data management software application offered the perfect solutions for both parties and enabled a tape-based archive of extreme size. The university selected Spectra Logic's StorCycle solution which offers a seamless view across all of the storage it manages. Both the university and the researchers can have access across all data managed.

Researchers can use the scanning capabilities of StorCycle to identify and target inactive data sets for archive. StorCycle's project archive will assure this situation doesn't reoccur. Project archive allows users to identify any files or directories associated with a project and archive them as a group. This can be done immediately after a large project is completed.

Archived data sets can be tagged with additional information to identify anything of importance to the project, be it grants associated with the project, researchers involved, project names, etc. This metadata can easily be searched at any point in the future. Likewise, StorCycle produces a manifest for each project archive which can be accessed as a digital file. The manifest shows exactly what was moved, where the data originated, where it was moved to, and when it was moved. It can be digitally displayed by clicking on the finished project archive and stored with other files in the project. It does not require a query into the database, and can be worked into existing workflows.

With several hundred petabytes of data currently under management, the university will now be able to deploy an affordable, glacier-like tape storage archive capable of growing to an Exabyte in size.



Machine Learning and Autonomous Driving

"Deep Blue…" "Watson…" "AlphaZero…" These are names which evoke either great fear or great inspiration in the best Chess players in the world. These are the names of the supercomputers and associated "machine-learning" or "self-learning" programs which have beat the highest ranked chess champions in existence. And therefore, these names have become synonymous with the concept and power of machine learning.

Whereas Deep Blue and Watson required hundreds of hours of programming and "teaching," AlphaZero made headlines in late 2017 for its ability to "self-learn." AlphaZero was programed with nothing more than the rules of chess. Nine hours later, it had played itself millions of times, learned from the experience, and was able to beat or draw not only the best human players, but also all prior "computer champions." This self-play learns from its success and failure and feeds the output into what is termed a "neural network."

One may think that the need for large quantities of existing information (such as Exascale archives) are not needed for machine learning. But that all depends on what the machine is expected to do. AlphaZero plays games – there are two players, complete information available to each player, and specific rules that must be obeyed. AlphaZero is incapable of making an illegal move. Nor would it be able to play another person or computer who "cheats."

In a zero-sum game – one winner and one loser – with no hidden information or elements, AlphaZero is unbeatable. But, if you need a ride to the grocery store… you might want to look for another solution.

Autonomous or "self-driving" cars hold a promising future, based on neural networks, but must be able to deal with a much different situation than found in a chess game.

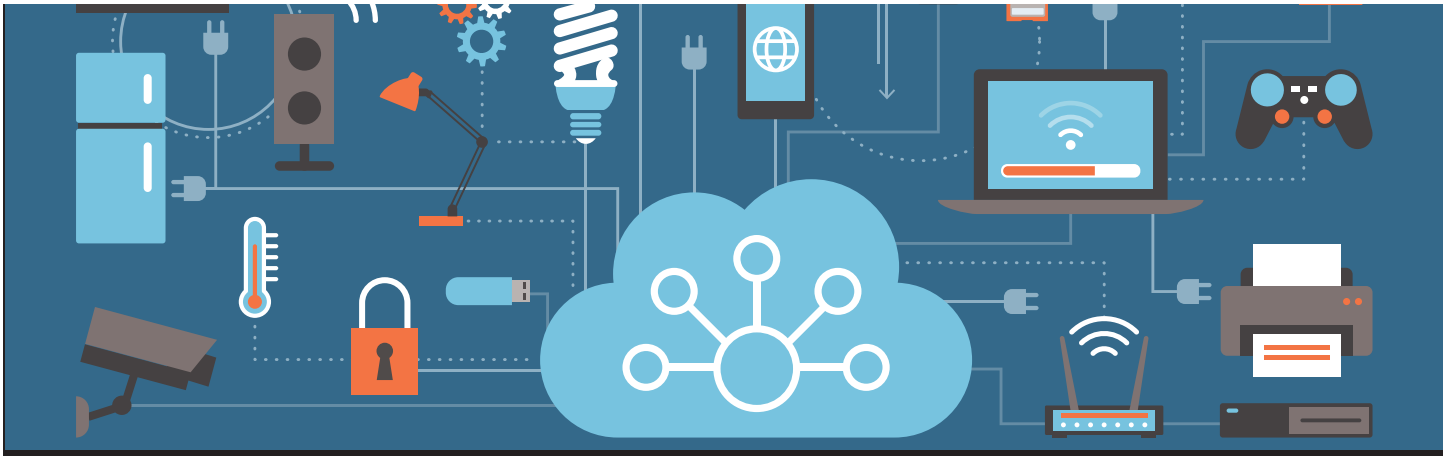
In daily, real-life scenarios, one would expect two people to go to the grocery store at the same time and both arrive home safely with their groceries – a win/win scenario. But there are many unknowns or "hidden information" floating around out there. Weather, cyclists, animals, poor drivers, drunk drivers, the list of unknowns is too large to cover.

At present there are multiple companies worldwide working on autonomous vehicles. Developing self-driving cars involves building very large neural networks (which reside in the car) which are pre-programmed to process camera, LIDAR and radar data to spot other moving vehicles, pedestrians, traffic signs and lights, animals, bicycles, motorcycles, objects in the road, and even traffic cones.

The configuration of the neural network is created and tested with "training data". This training data is collected from video cameras built into vehicles already in operation with human drivers. The more training data a manufacturer holds and processes, the safer the autonomous vehicle will be.

Depending on the level of "connectedness," autonomous automobiles will generate up to 5TB per car per day. This training data will grow into exabytes of storage.

This is AI in pure form. And it won't stop here. In the future there will be many more applications for AI, which will drive the need for training data. Autonomous cars fit into the larger category of the "Internet of Things" or IoT, our next discussion.



IoT

The Internet of Things (IoT) accounts for virtually all sensors and cameras which collect, and are capable of sharing, data. IoT devices may include traffic sensors, weather sensors, factory sensors, sensors on an oil pipeline, video surveillance cameras, your “smart” doorbell, smart watches, your home cooling/heating thermostat, your car, or even your toaster.

The advisory firm, International Data Corporation (IDC) forecasts IoT devices to grow to over 40 billion units by 2025. The idea is that millions of sensors could gather data and send it to a central location for processing. The applications for such data are endless – better traffic control, weather forecasts, manufacturing efficiencies, GPFS, health monitoring, etc. There are civilian as well as military applications. As diverse as IoT devices are, they have one thing in common: They generate data.

And generate data they will. IDC further forecasts 79.4 Zettabytes of data to be generated by IoT devices by 2025. Not all of this data will be saved, but if 0.0000125% of that data is to be saved, that alone is one Exabyte of data. It is very likely that 10 percent to 30 percent or more of this data could be used for future insight and improvement on process, product, manufacturing, security, and overall quality of life.

The IoT has become so pervasive, a new acronym is now being used – IoE, the “Internet of Everything.” When we have access to data, we have unprecedented insight on both the past and the future. Those with access to the most data are those who hold the greatest insight. Another reason Exascale storage is being sought after in every market segment conceivable, both commercial and public. IoT data plays a large role in most of the use cases we’re discussing.

Medical, BioTech and Genomics

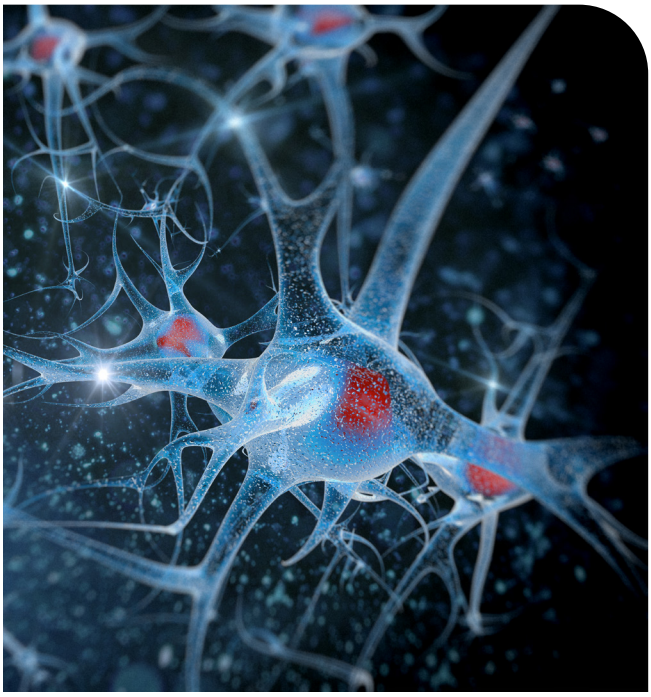
We've already covered the discussion of BioPharma and big data – drug development is being reduced by years required to reach the market – bringing more affordable, safer drugs to market more rapidly. But that is just one area of healthcare that is being revolutionized by access to large amounts of data. Computer systems are increasingly being used in diagnosing and treating disease as well.

In October of 2019, a study was published in the Lancet Journal of Digital Health which compared the performance between "deep learning" AI and healthcare professionals in detecting diseases based on medical imaging (X-rays, CT Scans, MRIs, etc.). The study looked at 14 studies published between January of 2012 and June of 2019. On average, AI was slightly better at correctly diagnosing disease than human healthcare professionals (87% correct vs. 86% correct respectively).

In certain types of cancer screening and detection, AI is even more effective. In lung cancer, Google has shown that the neural networks of AI can be trained to detect the presence of cancerous cells both earlier and more rapidly than radiologists. In reading mammograms to spot pre-cancerous and cancerous cells, computers are currently three to 10 times more accurate in detecting cancer than radiologists.

An outstanding difference is that AI can continuously train itself on any and all new data it's given. It never tires and it never retires! All it needs is more data to become even more accurate and efficient. Combining AI with the "Internet of Medical Things" – home sensors worn by patients, for example, enables health information to be gathered and processed real-time, and changes to the patient's lifestyle may be suggested.

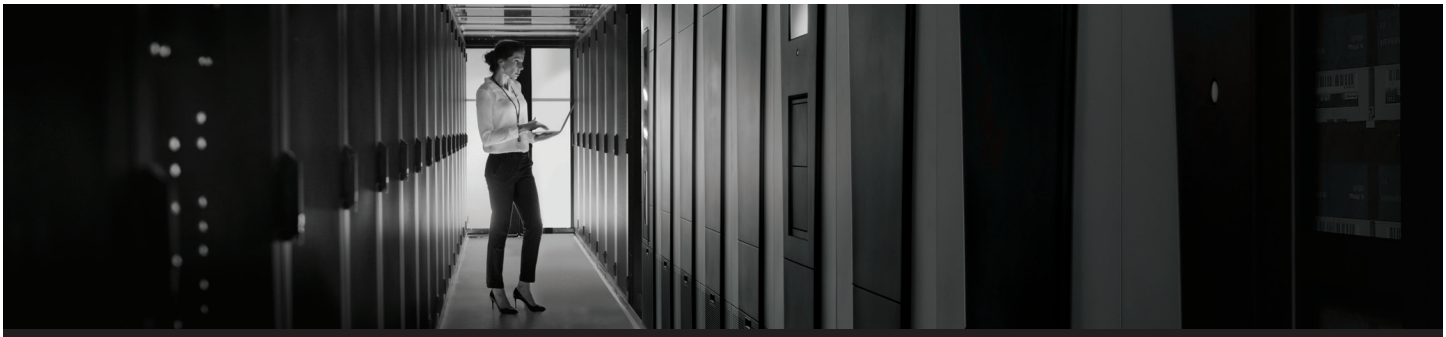
These systems are typically implemented in the form of neural networksl the more training data an entity has, the more accurate the prediction. In the future, many more AI-based healthcare systems will be created and implemented, improving disease detection.



Biological research is another area in medicine depending on and creating massive amounts of data. Cryoelectron Microscopy is a method of understanding underlying cellular mechanics. A single cell may be frozen at cryogenic temperatures, sliced into layers and then scanned by an electron microscope. This method allows samples to be examined without dyes or fixatives. Since samples remain in their native state, scientists are uncovering new information about viruses and protein complexes at their molecular level. As many as 500 layers may be sliced from a single cell – producing over 200TB in a single examination.

Genomic simulation has also opened a new door to medical understanding. Genetic testing, and the huge amounts of data it creates, are allowing individuals to understand their risks for certain diseases and use further technology (and data creation) to have the best chance of detecting and surviving what was a decade ago considered unsurvivable. We are reaching a point where medical regiments may be tailored to individuals based on their individual genomes.

All of these activities need or create petabytes to exabytes of data storage needs.



HPC Applications

The world of High Performance Computing (HPC) relies on supercomputers which are used for a wide range of computationally intensive tasks in various fields, including quantum mechanics, weather forecasting, climate research, oil and gas exploration, molecular modeling, and physical simulations (such as simulations of the early moments of the universe, airplane and spacecraft aerodynamics, automobile design, the detonation of nuclear weapons, and nuclear fusion).

As one might guess, High Performance Computing requires and creates enormous data sets. First, computational test scenarios manipulate a digital environment or situation. Data is then run through these test scenarios and the output of the tests or experiments is collected. A single experiment can create hundreds of terabytes up to multiple petabytes. Likewise, a single experiment may take months or even years to run and cost millions of dollars.

This creates a challenge for any experimental computing facility. Experiments are often repeated and reexamined. All inputs and outputs—that is, all the digital data related to a project—must be preserved. HPC sites rival data repositories of any vertical market and are probably second in size only to Internet operations.

Although data sets are regularly shared among engineers, scientists and developers who work in the HPC world, their workflows are more "manual" than those of general IT. Once the output of an experiment or modeling session is done, it does not change, and so it can be moved to a data repository requiring lower access and, hopefully, much greater density and energy efficiency for storage.

Government national labs are a large part of the HPC community. They create and store vast amounts of data for both civilian and military use. In the 2020s, each of these government labs will grow in capacity to one or more Exabytes. The tape-based, Exascale archive is already deployed at most HPC sites. The "TOP 500 list " of super computers shows the size and capacities of the world's top sites.

Weather Data

Weather prediction uses sophisticated modeling running on supercomputers, which are continuously being improved. Weather data consists of collecting data from thousands to tens of thousands of sensors. This also fits into the category of the "Internet of Things." These sensors may include ground-based sensors, satellite-based sensors, sea-water temperature sensors, aircraft collected data and occasionally weather balloons. This data is stored forever, along with the resultant forecasts. Weather scientists can tell if their models are working by evaluating the gaps and differences between actual outcome and what their model predicted for that particular time, whether currently or in the past.

Therefore, weather data archives are enormous. Most large national weather forecasters will exceed one Exabyte before 2025. There are in excess of 25 major worldwide organizations which include the National Weather Service (UD), Meteo France, The British MET, The European Center for Medium Range Forecasts, the Korean Meteorological Association, etc.

As sensors become more prevalent and higher in resolution, they will collect ever more data. As supercomputers grow in computational power, we will see greater growth in required archive storage.

High Energy Physics

When the limits of human understanding are being pushed by the desire to answer some of life's most complex questions, the amount of time and complexity of the research demands the most secure and scalable storage available. CERN, uses some of the world's largest and most complex scientific instruments to study the fundamental particles of matter, quite literally discovering the "God Particle", the missing cornerstone in our knowledge of nature.

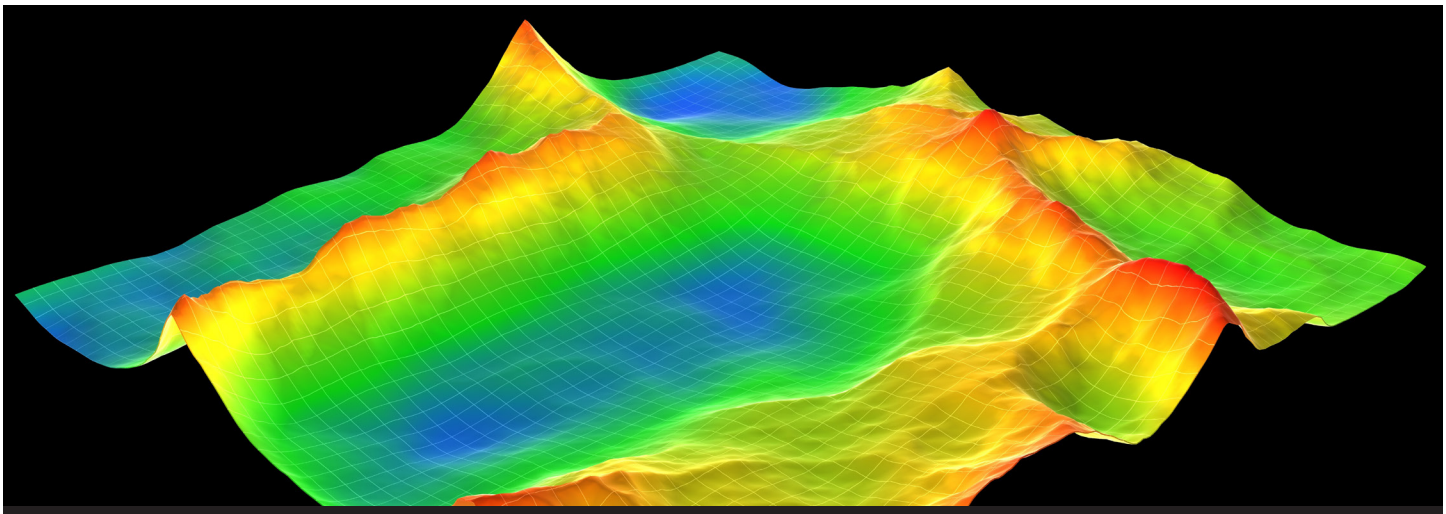
They do this with the LHC (Large Hadron Collider), the world's largest and most powerful particle accelerator. The LHC is 27 km (16.8 miles) of superconducting magnets in the shape of a ring, that sends high-energy particle beams close to the speed of light until they collide, with the goal of figuring out how the universe was formed. This research means the CERN Data Centre is producing data at an astronomical one petabyte of data per day, pushing their CERN Advance Storage System (CASTOR) to a massive 330PB of data stored on tape. According to CERN, this is equivalent to 2000 years of 24/7 HD video recording.

Currently the LHC is shut down while it gets a big performance upgrade. This upgrade will keep it shut down for two years, just long enough to upgrade their data archival software and tape system to handle the much higher data volume. Once the LHC resumes its next run, data creation is expected to double during the years 2021 to 2023. This means that they will be storing an additional 600PB or more after run three in 2023. This will put CERN very close to if not over 1 Exabyte of data on tape at the end of 2023. The LHC will then shut down again, upgrades to many of the sensors, magnets, and testing devices in the collider will be made, and when it resumes from 2026 to 2029, we can expect an increase of five times the current level of data stored. This means that during run four there will be over 1.5 Exabytes of additional data that needs to be archived to tape.

The new CERN Tape Archive (CTA) software will replace CASTOR and will store the existing 330 petabytes of data, as well as ALL new data created. It is being designed to be able to handle these massive amounts of data. The data that is produced at CERN is extremely valuable and must be preserved for future generations of physicists making tape the ideal archive storage technology to use. It is also shared worldwide. CERN has transferred 830PB

of data and 1.1 billion files to other HEPIX research organizations all over the world. This allows other physicists to conduct research and it also means that the data is archived geographically with multiple copies so that it can safely be kept forever.

CTA has the ability to store an Exabyte in native capacity and CERN is rapidly approaching the need for an Exabyte of storage. They are counting on the industry to continue to innovate to store multiple Exabytes in the coming decade to be preserved for future generations and with the roadmap of LTO tape it looks to be a partnership that will stand the test of time.

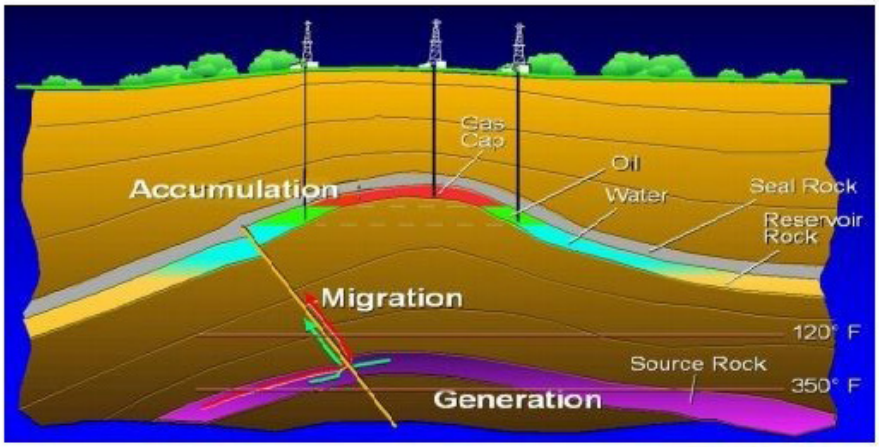


Petroleum Reservoir Storage and Modeling

Quite a bit has been said about High Performance Computing. One of the first commercial applications for HPC was seismic exploration for minerals and petroleum. What ultrasound is to doctors, seismic vibrations are to oil drillers. Seismic exploration measures the spread of waves through the earth. By measuring speed, reflection, and refraction on or near the earth's surface, data explorers can predict material composition deep underground.

Seismic data has been used for mineral exploration for decades, but it was far less effective before the invention of high-performance computing that could generate—and analyze—the tremendous amounts of data required for efficacy. Thanks to HPC, the era of drilling to determine what were dry wells and what were gushers is now as dated as the black-and-white footage that documents it.

Advances in HPC and data analysis have all but eliminated the ambiguity of finding oil, and new developments in extraction, such as horizontal drilling enables drillers to physically access oil once seismic data has located it with precision.



Simplified visualization of hydrocarbon traps recognized in producing oil and gas fields. (source: http://www.geologyin.com/2014/12/hydrocarbon-traps.html)

The important point to note here is that, although technologies for reaching and extracting oil and natural gas deposits have evolved, they still rely on seismic information gathered decades ago on deposits that were unreachable at the time.

Geologists therefore have not had to spend millions of dollars to resurvey and perform new seismic explorations. No matter its age, geological data remains as good today as it was on the day it was first recorded and can be reanalyzed using the latest and most accurate methods. This data remains invaluable as new technology continues to emerge; those approaches will be modeled against existing seismic exploration data.

The data generated from seismic exploration fits into the category of Exascale. It continues to be stored on tape, in large tape libraries.

HYBRID CLOUD FOR ARCHIVES?

One of the largest growth areas in our industry is that of hybrid cloud. “Hybrid Cloud,” is a term that’s been thrown about ever since the introduction of public cloud. In this discussion, hybrid cloud refers to the ability to keep data on premise as well as in the cloud. Some data may reside exclusively on premise, such as highly confidential information or amounts of information too large to store in the cloud indefinitely given monthly charges. Other data may reside exclusively in the cloud, such as transitive data not needed once a given calculation or compute service is completed. It may be desired to have copies of the same data stored both in the cloud and on premise simultaneously.

The ultimate definition of hybrid cloud allows for the management of data to occur from a “single pane of glass” or control mechanism which provides a universal view of the data – regardless of where that data physically resides.

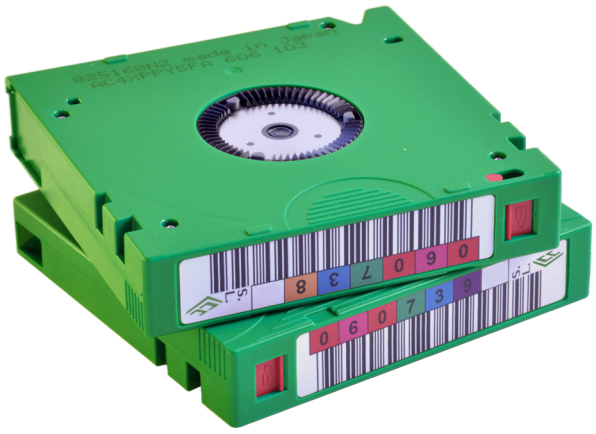
Both public cloud and on-premise storage provide certain advantages and challenges. Hybrid cloud offers the possibility of getting the best of both worlds if it’s done right. Let’s start by comparing cloud storage and on-premise storage. Cost and performance will not bode as well for cloud storage of an Exabyte, as it does for on-premise storage. That isn’t necessarily a strike against cloud storage as much as it is a reason for a hybrid approach of combining on-premise and public cloud.

Before discussing Hybrid Cloud, it’s important to understand the opportunities/challenges with “cloud only” or “on-premise only” models.

WHAT DOES IT COST TO STORE AN EXABYTE?

Tape only

Creating an Exabyte archive on tape requires capital expense as well as operational expense. The costs are more heavily weighted on the front end due to the purchase of a tape library, tapes, etc. The “pay as you go” model offered by the cloud can be replicated with Spectra’s design of the TFinity Tape Library. As mentioned earlier, the TFinity can start with as few as 3 frames, 1 tape drive and 50 enabled tape slots. Existing frames can be fully populated in the field by end users. Once a frame is filled, additional frames can be added for up to 45 frames to store and protect an Exabyte of uncompressed data.



We started this e-book with a brief description of what an Exabyte archive would physically look like in a TFinity Tape Library. We’ll now show what it looks like, including costs, to grow a large archive into a full Exabyte archive.

Obviously, the organizations we're discussing are working in a very high data-capture environment. They could be involved with Artificial Intelligence; Internet of Things; large video archives; HPC computing on an Exascale supercomputer; or capturing machine data from Cryogenic Electron Microscope, Square Kilometer Array, or another very high data output machine.

These environments can easily generate up to 200PB of data per year – reaching an Exabyte in five years. To accommodate this, we'll configure an 11-frame TFinity tape library with 48 LTO-9 full height FC tape drives, 2 high-speed Spectra BlackPearl X Object Storage systems (each capable of reading and writing to 24 tape drives at maximum native throughput), one year of support, and 50PB of media on day one. The starting cost of this complete configuration, installation and support will be around $1.5 million.

Over the five-year period, media will be purchased quarterly, 50PB per quarter, to take advantage of the media price reductions that happen on a regular basis with LTO media. At the end of the first year, the organization will spend a total of $2.68 million and will have 200PB stored on tape. The starting cost of 200PB in the AWS Deep Glacier Cloud would be roughly $2.44 million – a bit less expensive on day one.

For the next four years, no additional robotics or tape drives will be needed. Ongoing costs will include additional media expansion storage frames, media, and yearly support.

In year two, an additional 9 media expansion frames are added bringing the total library size to 20 frames. An additional 50PB of LTO-9 media will be purchased quarterly, and an additional year of support will be paid to bring the total additional spend for year two to $1.54 million.

In year three, 8 more media expansion frames are added, again 50PB of media quarterly and one year of support for a yearly spend of $1.36 million.

Year four brings another 9 media expansion frames, 50PB of media quarterly and a year of support for $1.26 million. The decreasing cost of media can be seen in the lower cost year over year for the same capacity added to the library.

The final year of use will add the last 8 media expansion frames, the final 200PB of media, and one more year of support for a total that year of $1.21 million.

This brings the total spend to $8.05 million for the five-year period and there is now 1 Exabyte of data stored in this environment. You will note in the table below, our total cost per Gigabyte of data stored is $0.008/GB.

Tape	Year 1	Year 2	Year 3	Year 4	Year 5	$/GB
Capacity	200PB	400PB	600PB	800PB	1 Exabyte	
Starting Cost	$2,679,290	--	--	--	--	
Yearly Incremental Cost	--	$1,536,882	$1,362,234	$1,264,155	$1,211,729	
Total Accrued Cost	$2,679,290	$4,216,172	$5,578,406	$6,842,561	**$8,054,290**	**$0.008/GB**

Cloud Only

Cost of storage is always a consideration, and that consideration becomes larger as the amount of data and time of retention become larger. Cloud often leads with the low cost of storage – usually shown as a "per Gigabyte / per month" charge. AWS Glacier Deep Archive offers storage as low as $0 .00099/GB (just under one-tenth of one cent) per GB/month. It's important to note that this is a "per month" charge, so it's a bit challenging to compare it in that form to the "total cost" per GB as noted above with tape.

That being said, storing 200PB of data for one year at $0.00099/GB/month (the AWS published price per GB/month for Glacier Deep Archive in North America) would cost over $2.4 million per year. As our archive grows 200PB per year until it reaches the Exabyte mark, calculations for storage alone are relatively straightforward. The total "storage" expense for an archive growing to 1 Exabyte over five years would be a little over $36 million or $0.036/GB.

Cloud Archive	Year 1	Year 2	Year 3	Year 4	Year 5	$/GB
Capacity	200PB	400PB	600PB	800PB	1 Exabyte	
Starting Cost	$2,424,000	--	--	--	--	
Yearly Incremental Cost	--	$4,848,000	$7,272,000	$9,696,000	$12,120,000	
Total Accrued Cost	$2,424,000	$7,272,000	$14,544,000	$24,240,000	**$36,360,000**	**$0.036/GB**

The word “storage” cost is emphasized above because that cost does not include charges for accessing data, various levels of storage required to “land” or “stage” data back, etc. Nor does it include the cost of bandwidth selected to move the data to and from the cloud. More on that in the sections below.

AWS also offers a seemingly inexpensive cost per Gigabyte to restore data. There are various levels, but the most affordable and least responsive is $.0025/GB to restore – about 2.5 times the amount charged to store the data. Again, with no internal movement charges being calculated, it would cost $2,500 per Petabyte to bring back the data. At Exabyte scale, even a very small read or recovery rate becomes significant. At 5% per year, read/recovery charges would be $125,000. At 5% per month, those annual charges would be $1.5 million. It would cost $2.5 million to restore the entire Exabyte archive. There would be no additional charges at all if this were being restored from a TFinity. Depending on access, this could easily save over a million a year and over $2 million for a disaster recovery or other complete recall.

Applying the cloud model of pricing to tape automation (cents/GB) gives more interesting insights. Looking at a “per Gigabyte” cost for tape automation, the total five-year archive, spanning from 200PB to 1EB, would cost roughly $0.008 (under 1 cent) per Gigabyte.

The total cost for cloud storage over the five-year period would be $.036 per GB, making it 4.5 times more expensive to create this archive on cloud vs. tape.

Archives for 10 or More Years

The above calculations are based on a five-year model to reach an Exabyte because, as stated earlier, most organizations are on their way to an Exabyte vs. being there already. Most archives are designed to last decades or even indefinitely.

What happens to costs after that first five years? Moving forward, it will cost a little over $100,000 per year to support this size system on tape, so the total spend over a 10-year period will be just over $8.5 million. This ever so slightly increases the tape archive cost from $.008/GB to $.0085/GB.

Tape	Year 6	Year 7	Year 8	Year 9	Year 10	$/GB
Capacity	1 Exabyte	1 Exabyte	1 Exabyte	1 Exabyte	1 Exabyte	
Starting Cost	**$8,054,290**	--	--	--	--	
Yearly Incremental Cost	$100,000	$100,000	$100,000	$100,000	$100,000	
Total Accrued Cost	$8,154,290	$8,254,290	$8,354,290	$8,454,290	**$8,554,290**	**$0.0085/GB**

Cloud storage will continue the monthly storage charges which will equate to $12.1 million per year for the Exabyte we hold in archive in years 6 through 10. That would put the total cost of the cloud archive at over $96 million – significantly raising the cost of storage from $.036/GB to $.097/GB -- over 11 times the cost of the same archive on tape! Whereas tape is still less than a penny per GB ($.008), cloud storage is actually running around 10 cents per GB ($.097).

Cloud	Year 6	Year 7	Year 8	Year 9	Year 10	$/GB
Capacity	1 Exabyte	1 Exabyte	1 Exabyte	1 Exabyte	1 Exabyte	
Starting Cost	**$36,360,000**	--	--	--	--	
Yearly Incremental Cost	$12,120,000	$12,120,000	$12,120,000	$12,120,000	$12,120,000	
Total Accrued Cost	$48,480,000	$60,600,000	$72,720,000	$84,840,000	**$96,960,000**	**$0.097/GB**

Furthermore, the tape archive holds no outside costs to restore data. This eliminates the shock of unplanned expenses when you least need to deal with them – during an unexpected data loss, cyberattack or natural disaster. Archives are also used to gain new insight to current problems or challenges using previously collected data. Having archives immediately available without additional, excessive costs makes future research more viable and increases the value of the archive.

As mentioned earlier, having a second copy on tape virtually assures data availability. In a high capacity tape system, about 3/4 of the total cost of the environment is in media and 1/4 is in tape library hardware and support of the system. In this scenario, making a second copy to store outside the library would cost a little over $6.2 million over the same five-year period using the same media purchasing policy and pricing. Storing two copies in the cloud simply doubles the $12.1 million for each year the second copy of data is tracked.

Pricing Models For Those "On Their Way" to an Exabyte

Similar pricing models can be done for smaller systems as well. Below are tables to show cost for organizations planning on archiving 5PB, 10PB, 25PB, 50PB, 100PB, 250PB or 500PB. In planning for a long-term tape archive, the items which account for cost will be the same as mentioned above – the tape library, support, expansion when needed, and media. Depending on the initial amount of data, Spectra recommends different models of the Spectra Tape Library series.

Since the Spectra Tape Libraries are capable of 'transcaling' up to the next size library, even starting with the smallest of Spectra libraries can eventually be grown into a TFinity while protecting your existing investment.

For the 5PB to 10PB archive, Spectra's smallest tape library would be used – the Spectra Stack. The Spectra Stack starts with a base unit and can be expanded with up to 6 expansion modules for a total of 7 modules in a single 42U, 19" rack.

For a 5PB archive, the starting point is a single Spectra Stack base unit with 2 FH tape drives. So, 250TB of media is purchased quarterly for a total of 5PB after five years. And 3 Spectra Stack Expansion modules will be added over time as needed filling 24U of a standard 42U rack.

Capacity in PB	year 1 cost	year 2 incremental cost	year 3 incremental cost	year 4 incremental cost	year 5 incremental cost	Total cost	$/GB	TB/SqFt
5	$33,822	$14,688	$14,093	$8,556	$13,520	$84,679	$0.017	769.23
10	$63,651	$23,223	$26,729	$21,109	$20,943	$155,655	$0.016	1538.5

For a 10PB archive, the starting point will be a Spectra Stack base and 1 Expansion module, 4 FH tape drive and 500TB of media per quarter. The final configuration will be the same Spectra Stack base and 6 Expansion modules filling up a 19" rack while still using the 4 FH LTO-9 drives. The costs for both these examples can be seen in the table above.

The Spectra T950 can be expanded to a total of 8 Frames allowing from 50 to 10,020 LTO-9 slots (7614 Enterprise slots), and up to 120 tape drives. With LTO-9, that would be a capacity of over 180PB (uncompressed).

For environments that may need to store 25PB or 50PB over a five-year period, the T950 tape library makes sense from a capacity and cost perspective.

For a 25PB archive over five years, a 2-frame T950 would be used with 6 LTO-9 drives. The entire library would be installed on day one and then 1.25PB of media would be added quarterly.

For a 50PB archive over five years, a 3-frame T950 would be used with 12 LTO-9 drives. While the full 3 frames would not be required on day one, it's probably more efficient to install the entire library initially. For each of the next four years, only media and yearly support will be purchased. This library will be 7.25 feet long and will occupy 26 square feet. It will also be capable of 4.8 GB/s of total throughput. The costs for these two examples can be seen in the below table.

Capacity in PB	year 1 cost	year 2 incremental cost	year 3 incremental cost	year 4 incremental cost	year 5 incremental cost	Total cost	$/GB	TB/SqFt
25	$193,532	$47,583	$42,953	$39,732	$38,424	$362,224	$0.014	1443.5
50	$291,560	$86,210	$77,211	$70,951	$68,408	$594,341	$0.012	1924.6

For the remaining archives – 100PB, 250PB and 500PB – the Spectra TFinity is recommended.

In a 100PB archive stored over five years, a 3-frame, 12-drive TFinity and a single BlackPearl X would be recommended. The TFinity will be expanded to a 6-frame library over the course of the five-year period. The total length of the library is under 15 feet.

In a 250PB archive stored over five years, a 5-frame, 24-drive TFinity and a single BlackPearl X would be recommended. The unit will grow by 50PB per year with media purchases of 12.5PB quarterly. The TFinity will be expanded to a total of 13 frames over the course of the five-year period. The costs for these two examples can be seen in the below table.

Capacity in PB	year 1 cost	year 2 incremental cost	year 3 incremental cost	year 4 incremental cost	year 5 incremental cost	Total cost	$/GB	TB/SqFt
100	$590,723	$182,704	$166,184	$154,991	$137,750	$1,232,352	$0.012	1881.37
250	$978,058	$403,844	$363,153	$335,425	$325,289	$2,405,769	$0.010	2197.40

For an organization that will need 500PB stored over a five-year period, a similar configuration of TFinity would be used. The library would start out as a 7-frame system with 48 LTO-9 tape drives and BlackPearl X object storage units. 25PB of media will be added quarterly throughout the five-year period, and, as with all of these financial projections, yearly support is also included. In year two, an additional 4 media expansion frames will be added. In year three, another 4 media expansion frames will be added. Year four requires 5 additional media expansion frames. In year five, the final 4 media expansion frames will be added for a completed TFinity with 24 frames total, capable of holding 500PB of uncompressed data.

With this size library it's roughly $0.009 per GB to store 500PB of data. A second copy of media will only add another $3.11 million, further driving down the cost per GB.

Capacity in PB	year 1 cost	year 2 incremental cost	year 3 incremental cost	year 4 incremental cost	year 5 incremental cost	Total cost	$/GB	TB/SqFt
500	$1,763,583	$761,945	$712,044	$669,422	$637,267	$4,544,262	$0.009	2392.03

For a quick comparison, a consolidated table showing costs for all configurations is listed below:

Capacity in PB	year 1 cost	year 2 incremental cost	year 3 incremental cost	year 4 incremental cost	year 5 incremental cost	Total cost	$/TB	$/GB	TB/SqFt
1000	$2,679,290	$1,536,882	$1,362,234	$1,264,155	$1,211,729	$8,054,290	$8.05	$0.008	2558.32
500	$1,763,583	$761,945	$712,044	$669,422	$637,267	$4,544,262	$9.09	$0.009	2392.03
250	$978,058	$403,844	$363,153	$335,425	$325,289	$2,405,769	$9.62	$0.010	2197.40
100	$590,723	$182,704	$166,184	$154,991	$137,750	$1,232,352	$12.32	$0.012	1881.37
50	$291,560	$86,210	$77,211	$70,951	$68,408	$594,341	$11.89	$0.012	1924.62
25	$193,532	$47,583	$42,953	$39,732	$38,424	$362,224	$14.49	$0.014	1443.46
10	$63,651	$23,223	$26,729	$21,109	$20,943	$155,655	$15.57	$0.016	1538.46
5	$33,822	$14,688	$14,093	$8,556	$13,520	$84,679	$16.94	$0.017	769.23

What Does It Take to Move an Exabyte of data?

The performance or bandwidth required of an organization's wide area network (WAN) to move data to the cloud is another consideration. Cost will vary tremendously depending on the geographic location of service and type of service. Likewise, an organization's use of WAN will be for more than cloud access alone. While it's difficult to assess the average cost of bandwidth for moving data to the cloud, it's easy to evaluate how much bandwidth it would take to accomplish archiving an Exabyte to the cloud.

Connection	Speed	TB Moved Per Day	Time to Move 1 PB	Time to Move 1 EB
Gigabit Ethernet	1 Gbs	10.8	92 days	252 years
OC48	2.5 Gbs	27	37 days	101 years
OC96	4.976 Gbs	53.74	18 days	49 years
OC192	9.600 Gbs	103.68	9.6 days	26 years

The egress charges, paid to AWS for retrieval of data, does not include the pipe customers will use for the restoration, nor does it guarantee how long it will take to actually receive the data back. The service level agreement (SLA) offered only states how long it will be before the customer has as access to their data in order to start the restoration.

Various types of internet connections offer varying speeds. Following is a chart showing examples of the performance of dedicated connections via the given connection type as well as the time it would take, given full dedicated performance, to move Terabytes, Petabytes and Exabytes of data.

The bandwidth numbers above are best-case scenario using 100 percent of available bandwidth, so any other organizational use of that bandwidth must be stopped to achieve these numbers. Even restoring a single Petabyte, 1/1000th of an Exabyte, could pose serious challenges to an organization's SLA.

	Speed	TB Moved Per Day	Time to Move 1 PB	Time to Move 1 EB
Single LTO-9 Drive (Uncompressed)	400 MBs	34.56	34,056 days	252 years
48 LTO-9 Drives (uncompressed)	19.2 GBs (aggregate)	1,659	14.3 hours	1.6 years
OC192	9.600 Gbs	103.68	9.6 days	26 years

In comparison, a tape-based, on-premise archive offers both rapid access and restoration times. Access time to data can be reduced to minutes vs. hours. More importantly, at 400MB/s uncompressed, LTO-9 restoration offers comparably rapid restoration. The below table shows the speed of a single LTO-9 as well as the aggregate speed of the 48 LTO-9 drives used in our Exabyte archive example above. The OC192 WAN speed is repeated for reference.

Even the fastest of internet connections pales compared to the aggregate speed of LTO-9 tape drives. The aggregate speed of 48 LTO-9 drives is over 16 times faster than an OC192 connection. Keep in mind, up to 144 drives can be configured in a TFinity bringing the aggregate performance to over 58 GB/s!

It's clear that cost and performance will most likely stand in the way of archiving an Exabyte of data in the cloud. But these numbers are important to consider. Few organizations "start" with an Exabyte of data, but if you look at the current growth of data, combined with longer retention rates, many organizations are already approaching archives that will eventually lead to this milestone.

The Promise of Hybrid Cloud Revisited

This section started with a brief description of Hybrid Cloud. As well noted above, cost and performance hinder Exabytes of data from being archived to the cloud. Yet the cloud offers services that an on-premise archive might be more limited with. Cloud can be very useful for distributing or sharing data. Cloud offers transcoding services to assure shared data can be consumed on multiple, disparate platforms.

The ultimate goal of data users and storage manufacturers alike should be to maximize the benefits of both public cloud and on-premise storage. How can we keep low-resolution copies of content in the cloud for editing yet have the full resolution copies remain on-premise for lower cost storage of the higher resolution files? How can we stream research data to tape for indefinite retention on-premise (creating Petabytes and Exabytes of data), yet also have accessibility via the cloud while it's still relevant to researchers across the globe?

Arguments over "end-point" storage solutions – disk vs. tape, public cloud vs. private cloud, file vs. object – have consumed too much of the storage conversation and have deterred organizations from being able to focus on the real point behind storage – meeting the desired organizational goals that the information/data/content is used for.

If it's true that we will all use the cloud in some fashion, and we believe that it is possible, we need a more consolidated approach to working within – as well as through and even around – the cloud. Would it be possible to access data via the cloud, yet store the data on-premise? Applications are under way, and we believe it will be a short matter of time before these challenges will be met.

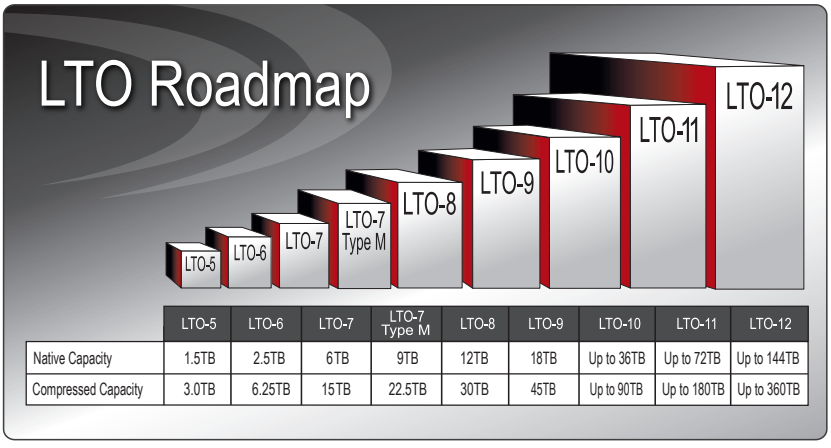
Moving towards a more "geographically independent" storage model will allow users to once again focus on using the most appropriate storage for a given role without creating independent storage silos which kill efficiencies and waste budget dollars.

In this manner, Exabyte-scale archives will become something to aspire to for the power they hold in continuing our research, manufacturing, discovery, medicine, industry and quality of life. Stay tuned!

LTO INTO THE FUTURE

In September of 2020, IBM and the LTO Consortium announced the LTO-9 full height drive which matched their latest Enterprise drive, the IBM® TS1160, in uncompressed throughput, and nearly matched the TS1160 in uncompressed capacity with 18TB vs. 20TB per cartridge. More significantly, the LTO roadmap was revised through LTO-12.

This indicates that the LTO technology in our Exabyte capacity library could increase by a factor of 8 times to to 8 Exabytes over the 10-year period we have discussed. As LTO-1 has grown to LTO-9, the growth to LTO-12 is readily possible with tweaks and improvements to the existing technology, portending a bright future.



	LTO-5	LTO-6	LTO-7	LTO-7 Type M	LTO-8	LTO-9	LTO-10	LTO-11	LTO-12
Native Capacity	1.5TB	2.5TB	6TB	9TB	12TB	18TB	Up to 36TB	Up to 72TB	Up to 144TB
Compressed Capacity	3.0TB	6.25TB	15TB	22.5TB	30TB	45TB	Up to 90TB	Up to 180TB	Up to 360TB

Assuming a 2.5:1 compression achieved with larger compression history buffer beginning with LTO generation 6 drives.

SUMMARY

In more than 40 years of creating, manufacturing and delivering storage solutions, Spectra Logic has seen many "ups and downs" of the industry. Often, one new technology threatens an older technology or requires a yet-to-be-invented technology to reach its promise. It's often difficult to determine where the market is going or how best to balance organizational mandates with technology and budget.

The one constant we see in storage is the need for more. That doesn't mean that budgets have to be broken or organizational goals need be missed.

By using the right technology, Exascale archives can be achieved at reasonable costs, often with very high return on investment. Hopefully this e-book shows that Exascale archives are created to enhance the pursuits of organizations – be it for science, time to market, health, prosperity, quality of life, education or entertainment. We have the technology today to build what we will need tomorrow. Such is the nature of the archive.

ABOUT SPECTRA LOGIC

Spectra Logic develops data storage and data management solutions that solve the problem of long-term digital preservation for organizations dealing with exponential data growth. Dedicated solely to storage innovation for over 40 years, Spectra Logic's uncompromising product and customer focus is proven by the adoption of its solutions by leaders in multiple industries 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 cloud and public cloud.

To learn more, visit www.SpectraLogic.com.

303-449-6400 • 800-833-1132 • 6285 Lookout Road • Boulder, CO 80301 USA • spectralogic.com

V2-093020