



# Data Storage Outlook 2022

**2022**



# CONTENTS

- CONTENTS ..... 2**
- EXECUTIVE SUMMARY ..... 3**
  - The Size of the Digital Universe ..... 3
  - The Storage Gap ..... 3
  - Storage Apportionment and Tiering..... 3
  - Highlights from 2022 Report..... 4
- BACKGROUND ..... 4**
- THE NEXT STORAGE ARCHITECTURE..... 5**
  - File vs. Object Storage ..... 5
  - Comparison of File vs. Object System Properties ..... 6
  - New Storage Tiers..... 7
  - The Project Tier..... 7
  - The Perpetual Tier ..... 8
  - Data Movers..... 8
- STORAGE TECHNOLOGIES..... 9**
  - Persistent Memory ..... 9
  - Performance Comparison of Optane vs. Standard DRAM ..... 10
  - Flash ..... 11
  - Magnetic Disk ..... 17
  - Tape..... 24
  - Optical ..... 28
  - Future Storage Technologies ..... 32
  - Cloud Provider Storage Requirements ..... 32
  - Cloud Versus On-Premises Perpetual Storage ..... 33
- CO<sub>2</sub> EMISSIONS OF INFORMATION TECHNOLOGY SYSTEMS ..... 37**
- THE DIGITAL UNIVERSE ..... 41**
- CONCLUSION ..... 43**
  - Data Storage Dilemma ..... 43
  - Designing with the Cloud in Mind..... 43
  - Supporting Complex Workflows ..... 44
  - Planning for the Future ..... 44
- CONTACT US ..... 44**
- APPENDIX NOTES..... 45**

Copyright ©2022 Spectra Logic Corporation. All rights reserved worldwide. Spectra and Spectra Logic are registered trademarks of Spectra Logic. All other trademarks and registered trademarks are property of their respective owners. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission. All opinions in this white paper are those of Spectra Logic and are based on information from various industry reports, news reports and customer interviews.



## EXECUTIVE SUMMARY

***“If we did all the things we are capable of, we would literally astound ourselves.”***

***-Thomas Edison***

This is the seventh annual Data Storage Outlook report from Spectra Logic. The document explores how the world manages, accesses, uses and preserves its ever-growing data repositories. It also covers the strategies and technologies designed to protect the data being created now and in the future. As Thomas Edison, the famous American inventor, said, ‘If we did all the things we are capable of, we would literally astound ourselves.’ With that in mind, the outlook for data storage looks bright, as the IT industry continues to develop new methods and technologies to store, manage, use and preserve the world’s treasury of information – enough to astound and benefit humanity for many years to come.

In 2021, the Covid-19 virus continued to impact lives, businesses and countries. In the face of the pandemic, the IT industry displayed a tremendous amount of resiliency with organizations continuing to pivot to online work to meet their business objectives. Against this backdrop, however, cyberthreat actors took full advantage of the virus and resulting virtual work to attack organizations with sophisticated ransomware. According to analyst firm IDC in their [IDC 2021 Ransomware Study](#),<sup>1</sup> approximately one third of organizations worldwide had experienced a ransomware attack or breach in the previous 12 months, making IT security a top priority for 2022 and beyond.

### The Size of the Digital Universe

- According to IDC, global data creation and replication will experience a compound annual growth rate (CAGR) of 23 percent over the 2020-2025<sup>2</sup> forecast period. And analyst firm Gartner forecasts worldwide IT spending to total \$4.5 trillion in 2022, an increase of 5.1 percent from 2021<sup>3</sup>.

### The Storage Gap

- While there will be great demand and some constraints in budgets and infrastructure, Spectra’s projections show a small likelihood of a long-term constrained supply of storage to meet the needs of the digital universe through 2031. The storage industry, like all other industries that are dependent on electronic components, has seen supplies become limited, resulting in long lead times and price increases. It is expected that this will continue through much of 2022 and possibly into early 2023.

### Storage Apportionment and Tiering

- Economic concerns will push infrequently accessed data onto lower cost media tiers. Just as water seeks its own level, data will seek its proper mix of access time and storage cost.
- Spectra continues to envision a logical two-tier architecture comprised of multiple storage types. We further envision that the first logical tier’s storage requirements will be satisfied entirely through solid-state disk (SSD) storage technologies, while the second tier requirements will be satisfied by magnetic disk, tape and cloud deployed as object storage either on-premises or in the cloud.

## Highlights from 2022 Report

- 2021 saw persistent memory based on 3D XPoint technology certified with more traditional applications such as SAP.
- The flash market, due to global supply chain factors, saw an increase of 10% to 15% in 2021. PCIe Gen 5 products are in earlier stages of announcement and promise speeds that exceed 10 GB/s.
- Seagate and Western Digital are now in production of a 20TB CMR magnetic-based disk drive. Western Digital has announced a unique method where they are using flash inside the disk system to improve capacity and performance.
- LTO-9 tape is now shipping with a capacity point of 18TB per cartridge, a 50% capacity increase over LTO-8. Multiple tape vendors are now shipping systems that support the AWS S3 Glacier storage interface, which provides tape with support for any application that supports this interface. This appears to have generated renewed interest in tape for customers who have large amounts of data and want to avoid monthly cloud storage charges.
- With increasing concern about global warming, there will be a greater focus on electrical energy consumption generated from non-renewable energy sources. As the demand for information technology is predicted to grow by six times over the next decade, the challenge will be how to satisfy this demand while, at the same time, not increasing, and preferably decreasing, the associated CO<sub>2</sub> emissions. This year we have added a new section discussing possible methods for doing so.

## BACKGROUND

Spectra Logic celebrated more than 42 years of business success last year, with Spectra's Founder and CEO Nathan Thompson at the helm. The company delivers a full range of innovative data storage and data management solutions for organizations around the world. Demonstrating continuity and longevity, Spectra Logic's vast solution set includes software-enabled storage platforms, consisting of disk, object storage and tape, as well as enterprise-class multi-cloud data management solutions.





## THE NEXT STORAGE ARCHITECTURE

Increasing scale, level of collaboration and diversity of workflows are driving users toward a new model for data storage. The traditional file-based storage interface is well suited to in-progress work but breaks down at web scale. Object storage, on the other hand, is built for scale. Rather than attempting to force all storage into a single model, a sensible combination of both is the best approach.

### File vs. Object Storage

File systems are called on to serve many purposes, ranging from scratch storage to long-term archival. Like a jack of all trades, they are a master of none, and storage workflows are exceeding the capabilities of the traditional file system. The file system interface includes a diverse range of capabilities. For example, an application may write to any file at any location. As this capability expanded to network file systems (NFS, SMB), the complexity scaled up as well – for instance, allowing multiple writers to any location within a file.

The capabilities of the file system interface make it excellent for data that is being ingested, processed or transformed. As a user creates content or modifies something, the application may quickly hop around in its data files and update accordingly. It must do this with enough performance that the user's creative process is not interrupted, and also with sufficient safety that the user's data will be intact in the event of malfunction. The file system is the user's critical working space.

Object storage is simply another way of saying “the web.” From its beginning, the web's HyperText Transfer Protocol (HTTP) was a simple method of sending an object over the public internet, whether that object was a web page, image or dynamically-generated content. Any web browser is a basic “object storage client.” HTTP has methods for getting and putting whole objects but lacks the notion of interactive, random I/O.

This simplicity, however, is a powerful enabler for object storage to operate at scale. Every object has a Uniform Resource Identifier (URI) which enables that object to be addressed -- whether it's on a server in the next room or a data logger in the remote wilderness. It doesn't matter if the network topology or storage system is involved, or whether it is traversing multiple caches and firewalls. Objects may be migrated to different storage media or even moved from a company's data center into a public cloud; so long as the URI remains unchanged, users will neither know nor care.

The cloud grew out of the web, so it's no surprise that cloud is primarily based on object storage. The first product of Amazon Web Services (AWS) — pre-dating their compute offerings — was the Simple Storage Service (S3) released in 2006. The S3 protocol is simply HTTP with minimal additions. S3 includes methods for retrieving a range of an object, or sending an object in multiple parts, but in general it maintains a very simple, high-level interface. AWS has released other storage services, including a parallel file system, but S3 remains the backbone of their cloud.

The dramatic contrast between file system and object system capabilities means that the ideal storage interface is both. The data-driven organization should use a combination of systems to fully capitalize on the strengths of each.

## Comparison of File vs. Object System Properties

Feature	File System	Object System
<b>Connection</b>	Direct-attach or local network/VPN	VPN or public internet
<b>Standardization</b>	POSIX, Windows	Lacking; AWS S3 popular
<b>Read/Write Mix</b>	Arbitrary read/write	Write-only/read-many
<b>Data Mutability</b>	Update any file in any place	Objects immutable; make new version
<b>App compatibility</b>	Broad	Limited; new applications only
<b>Network/technology independent</b>	No	Yes
<b>Transparent storage class migration</b>	No	Yes
<b>Versioned, auditable</b>	No	Yes

## New Storage Tiers

In the past, data storage usage was defined by the technology leveraged to protect data using a pyramid structure, with the top of the pyramid designated for solid-state disk to store 'hot' data, SATA disk drives used to store 'warm' data and tape used for the bottom of the pyramid to archive 'cold' data. Today, Spectra describes a two-tier architecture to replace the dated pyramid model.

The two-tier paradigm focuses on the usage of the data rather than the technology. It combines a Primary or Project Tier where in-progress data resides, which is file-based, and a second or Perpetual Tier where finished and less frequently changed data resides, which is object-based. Data moves seamlessly between the two tiers as data is manipulated, analyzed, shared and protected.

### The Project Tier

- **Data ingest, where raw data streams need to be captured rapidly.** For example, a media production may need to capture camera streams, audio streams and timecode simultaneously. Data will be raw, uncompressed, and require extremely high bandwidth. These streams may be stored to separate devices (e.g., flash cards within each camera) or captured on a central system (RAID box or filer).
- **Work-in-progress, where a user may hop around and edit content in any location.** This may include edit-in-place such as some image editing applications, where a user may work across the X/Y image plane and multiple layers. It may also include non-destructive applications, where a change stream is captured but underlying data is never changed. Regardless of technique, the application must respond instantly to user input.
- **Computation scratch space, where the volume of data exceeds RAM and/or checkpoints are saved to stable storage.** Most of it will be discarded after the job is complete; only the results will live on. Storage must have high bandwidth, as time spent waiting for a checkpoint to finish is wasted.

The file system's ability to write to any location within a file is critical for capturing data as it happens. Some applications use the file system interface directly (open a file handle and write to it) while others use software libraries such as SQLite or HDF5 to write structured data in a crash-consistent manner.

But what happens when the user is finished with editing, and the dynamically changing data becomes static? It moves to the Perpetual Tier.

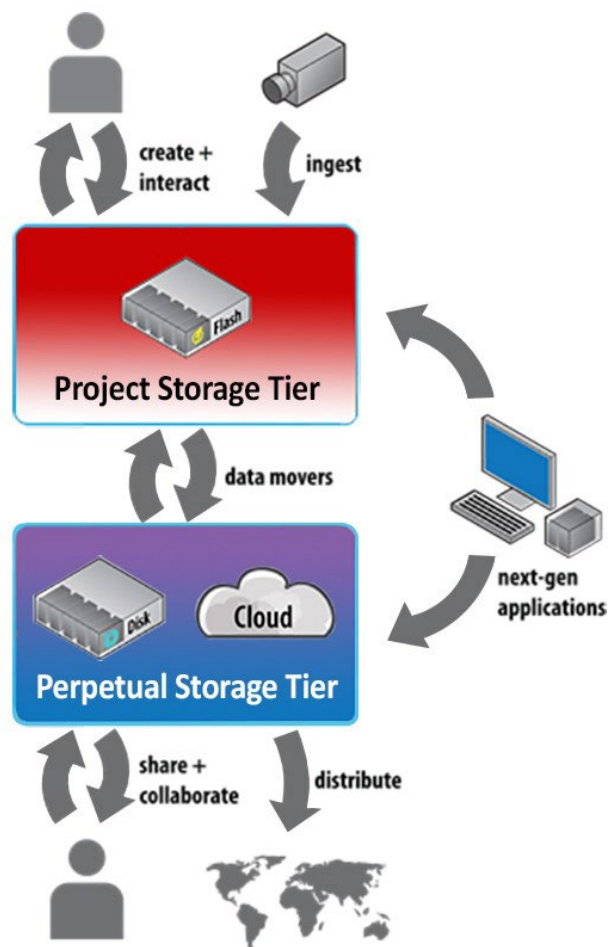


Figure 1: New Storage Tiers





## The Perpetual Tier

- **Project assets that must be shared across a team so they can be the basis for future work.** Video footage leaving production and going into post-production may need to be used by teams of editors, visual effects, audio editing, music scoring, color grading, and more. These teams may be spread across geographic regions and the object store may pre-stage copies in each region. Each source asset will have a globally-resolvable name and data integrity hash code. These are never modified. In some cases, new versions may be created, but the prior versions will be kept as well. The lifetime of raw assets is effectively forever—they are the studio’s lifeblood—and they may be migrated across storage technologies many times.
- **Completed work that must be distributed.** Object storage, along with public cloud providers, offer an ideal way to distribute data to end users across the globe. A finished media production may result in a variety of distribution files, along with a descriptive manifest, for example, MPEG-Dash as used by YouTube and Netflix. Because the objects are static, they may be cached in global content delivery networks.
- **Finished computational results to be shared across researchers.** Encryption and access controls, such as those provided in the S3/HTTP protocol, allow for sharing of sensitive data across the public internet. Storage cost may prompt users to apply a cold-storage-only policy to older data, knowing that it can be restored later if needed.


Data moves between the Project and Perpetual Tiers in both directions. Users may migrate from Project Tier to Perpetual Tier once files are complete, but migration may go the other way as well. A visual effects company may start from source files that exist in object storage in a public cloud, staging those to their Project Tier when they start work. Afterward, the finished shots are copied back to the cloud.

Whereas the software applications of the past used file systems only, next-generation applications support both tiers directly. They use a file system for their workspace and object storage (including cloud) as the source and destination for more permanent data. Additionally, some software libraries are supporting object storage natively; for example, there is a HDF5 library that can use a S3-compatible object store directly.

## Data Movers

Until applications can natively utilize both the Project and Perpetual Tier, data movers will be required in order to move data between the two tiers. Customers’ varying requirements will necessitate different types of data movers. Some customers may want the ability to move large amounts of project data over to the Perpetual Tier once a project is completed. This serves two purposes in that it frees up the Project Tier for new projects and it archives the data of the project making it available for future processing. Another set of customers may want to selectively prune the Project Tier of files that have not been accessed for a long period time. This frees up Project Tier storage such that expansion of that storage is required. Another customer may use the Perpetual Tier as a means to distribute data globally to multiple groups working on the same project. A data mover allows users to





“check out” a project by moving the data from the Perpetual Tier to a local Project Tier. Once changes to the Project Tier are made, they can be “checked in” back to the Perpetual Tier, thereby making those changes available to all sites.

## STORAGE TECHNOLOGIES

The storage device industry has exhibited constant innovation and improvement. This section discusses current technologies and technical advances occurring in the areas of persistent memory, flash, magnetic disk, magnetic tape, optical disc and future technologies, as well as Spectra’s view of what portion of the stored digital universe each will serve.

### Persistent Memory

This is the third year we have reported on the persistent storage tier that is defined by the two characteristics of persistence across power outages and performance close enough to DRAM to exist on a memory bus. Though various products, such as battery-backed DRAM, have been available for many years, they have always served a niche market. A newer technology entitled 3D XPoint™ was co-developed by Intel and Micron. This technology is revolutionary in that it retains a bit of information by altering the phase alignment of the underlying substrate associated with that bit. This bit level addressability avoids the need to perform the garbage collection operation required by zone-based devices such as NAND and SMR disk drives. Additionally, the number of binary phase changes the underlying substrate can cycle through during its lifetime is much higher than the number of times a NAND cell can be reprogrammed. These properties of consistent low latency and longevity make it ideal as persistent memory storage.

Intel introduced 3D XPoint technology in 2017 with their announcement of their Optane™ storage product line. The first generation of Optane products were exclusively provided in SSD form factors and interfaces and, therefore, directly competed with enterprise SSDs based on NAND technology. The Intel SSDs had lower latency, comparable bandwidth and higher cost per byte than that of high-end NAND-based SSD. With the release in 2019 of Intel’s second-generation Optane, the 3D XPoint technology became available as memory modules in a DIMM form factor with capacities of 128GB, 256GB and 512GB. The introduction of these, we believe, has resulted in an entirely new category of storage. This technology has the potential to be highly disruptive to the DRAM marketplace. This market of \$80 billion is dominated primarily by three players: Samsung, SK Hynix and Micron. Intel does not participate in this market, which means that any disruptive gains it makes is net new business.

Below is a performance comparison of Optane vs. standard DRAM. As can be seen, the DRAM wins every performance category by many multiples. However, performance is not the only factor that matters as Optane products wins in the categories of cost, density, power consumption and, of course, persistence. For many applications, the performance tradeoffs are worthwhile given these other factors. Micron has announced their first product based on this technology -- the Optane™ X100 SSD, which boasts read speeds of up to 9 GB/s.


The product is unavailable on the open market as Micron has indicated that their entire manufacturing output is going to system providers. In 2020, Micron said that they planned on selling their 3D XPoint manufacturing facility to focus on Flash and DRAM. Their intention was to have the facility “off their books” in 2021. This facility was purchased in July 2021 by Texas Instruments who will be retrofitting it to build other components.

## Performance Comparison of Optane vs. Standard DRAM

Latency	Optane DIMM	DRAM
Idle Sequential Read Latency	~170ns	~75ns
Idle Random Read Latency	~320ns	~80ns
<b>Per DIMM Bandwidths</b>		
Sequential Read	~7.6 GB/s	~15 GB/s
Random Read	~2.4 GB/s	~15 GB/s
Sequential Write	~2.3 GB/s	~15 GB/s
Random Write	~0.5 GB/s	~15 GB/s

Two application spaces that can derive almost immediate benefit from this technology are in-memory databases and large shared caches. In-memory databases have become more popular over the last few years as they provide the lowest latency with highest transaction rate as compared to databases running out of other storage mediums. However, these databases need as much DRAM in the system as the size of the database. This is both costly and, in some cases, just not possible given the amount of memory that can be placed into a server. Another issue is that, given the non-persistent nature of DRAM, the database needs to be “checkpointed” on a regular basis -- typically to an SSD. This is required because, if power is lost, the database can be restored to a good state. The Optane memory solves these problems with large formats, such as 512GB, which enables larger databases to be supported along with being persistent so that checkpointing is not required. For example, a database that has been checkpointed may require up to a half hour to be restored, while the same database running on Optane could be brought up in less than half a minute.

Another application that can be easily moved onto the Optane memory is that of cluster-wide caches such as Memcached. These caches hold items in-memory on each node of a cluster such that a vast majority of requests that come into the cluster can be serviced without needing to go to the back-end storage. For example, when a user logs into their account in a web application, their home page gets cached and is read one time from the back-end. As the user moves around the application, new information that is brought into the application is additionally cached. Optane memory is an excellent fit for this application as its high capacity allows for millions of items to be cached.



Besides these easy-to-convert applications, Intel reports that they are seeing substantial interest in traditional applications such as SAP. Optane memory can be presented in two modes to an application: memory mode and application-direct mode. In memory mode the Optane memory appears as DRAM. This makes it transparent to any application that is capable of using a large amount of memory; however, the persistence feature of the technology cannot be utilized as the application will be unaware of which memory is non-persistent DRAM and which memory is persistent Optane memory. In order to utilize the persistence capability of the technology, the application-direct mode is required -- and the application must be modified such that it manages the DRAM memory separately from the Optane memory. For example, a database could use DRAM as a database cache and Optane memory to hold the actual database data.


Given the disruptive nature of this technology, some of the largest DRAM producers are working on alternative solutions. Samsung, for instance, has plans to announce products based on Magnetoresistive random access memory (MRAM) that has many similar properties to that of 3D XPoint, while SK Hynix is working on technology similar to 3D XPoint. A limitation of Optane memory is that it is only compatible with Intel processors. It is clear that they are using this technology to differentiate their processors from those of AMD. Until a true competitor of 3D XPoint technology shows up that can work with AMD processors, AMD has little recourse but to support the newest generations of low latency SSDs, such as Z-NAND from Samsung and 3D XPoint-based SSDs from Micron.

## Flash

The fastest growing technology in the storage market continues to be NAND flash. It has capabilities of durability and speed that find favor in both the consumer and enterprise segments. In the consumer space it has become the de-facto technology for digital cameras, smart phones, tablets, laptops and desktop computers. As previously discussed, we predict that the Project Tier will be comprised of solid-state disk storage technologies.

Previous versions of this paper highlighted the flash vendors transition from planar (i.e., 2D) to 3D-Nand manufacturing. This transition was required in order to provide flash with a roadmap whereby increased capacities could be achieved for many years to come. During the time of this transition (years 2016 through 2017), increases in flash capacity were suppressed resulting in relatively small declines in price. In order to achieve these increases, substantial investment in existing and new flash wafer fabrication facilities as well as time was required for the 3D manufacturing process to achieve good yields. In fact, this investment over that four-year period was roughly \$100 billion. Due to high demand and supply chain issues, SSD prices rose 10 to 15% in 2021. There is some expectation that prices will drop back to where they were in 2020; however, that is highly dependent on a number of factors outside of the control of the industry.

There were five companies that owned and ran NAND fabrication lines: Samsung, Toshiba/Western Digital, Micron, SK Hynix and Yangtze Memory Technologies. The memory portion of the Toshiba conglomerate was spun out in late 2018 and is now called Kioxia. Each of the flash vendors delivered 100+ layer chips in 2020. Besides adding more layers, there are two other aspects of 3D flash that can provide greater capacity. The first is adding more voltage detection levels inside each cell. With the first flash chips produced, each cell was either charged or not charged, meaning that each represented a single binary bit referred to as single level cell (SLC). This was followed by detection of four levels, with two bits of information per cell referred to as multiple level cell (MLC).




Later, a triple level cell (TLC) holding three bits of information per cell was produced. In 2018 the first QLC parts were shipped as consumer technology, with each cell holding four bits of information. Currently QLC is prevalent in inexpensive consumer SSDs while TLC is used in higher priced enterprise SSDs. There have been some preliminary announcements from Intel and Toshiba about a five-level cell, called Penta-level (PLC); however, it is unclear if and when this technology will reach the market and what application spaces it may address when it does. As more levels of detection are added to a cell, writes take longer, the number of bits allocated for error correction at the part level increases, and the number of times that the cell can be programmed decreases. For these reasons, it may be that this technology may only be suitable for applications, such as archive, that do not overwrite data. To participate in the archive market against existing disk and tape solutions will require an order of magnitude or more of cost reduction.

The final method to increase capacity of flash is to decrease the cell size. This results in reducing signal integrity and would make it more difficult to detect voltage levels while reducing the number of bits that can be detected per cell. Given the flash roadmap, it appears that 19 nano-meters (nm) line-width is as small as the industry plans on producing. Given it is already at 20 nm, it doesn't appear that this method will yield much capacity gain. Looking at the three possible methods of increasing capacity, we conclude that the greatest opportunity will be by increasing the number of layers on a chip; however, that is now hitting technology limitations that must be overcome.

As discussed previously, all vendors have announced availability of 100+-layer parts in 2020. Of these, Samsung has delivered the first product to market consisting of single-stack 136-layer part. This is in comparison to some other vendors who are using "string stacking" in order to meet the 100+-layer goal. String stacking is a technique where multiple chips of some number of single-stacked parts are "glued" together to create a module of the desired number of layers. For example, a vendor could take four 32-layer parts and combine them to create a 128-layer part. Both techniques have their advantages and disadvantages. To do a single stack part exceeding 100 layers takes more upgrade investment in the fabrication facility than adding string stacking to a facility already producing a 64-layer part. Also, the initial yield of the 100-plus layer part is certainly going to be much lower than that of the already-established production of the lower layer part. On the other hand, the higher layer level part can win on cost when the overall manufacturing cost is reduced to a level that makes it cheaper than manufacturing multiple parts. The higher layer part also has the capability to expand into the future roadmap more easily.

For example, Samsung has indicated that they will be string stacking three 136-layer parts together to produce a 400-layer part in the next few years. To do so with a 32-layer part would require 12 or 13 parts. There are complex issues with building 100-plus layer parts. For instance, just the alignment of so many layers is problematic. For this reason and others, there are no vendors talking about building past 136-layers in a single-stack part. So, we predict that future capacity gains will be primarily achieved by string stacking parts together.




The question is what cost advantages will this yield longer term? The other question besides density is what advantages will higher layer string stack parts have over individual parts of smaller layers? For these reasons we are projecting that the price decreases in flash will slow and be more a function of the yield and manufacturing cost improvements that vendors are able to achieve.

According to the flash roadmap, the two technologies that require further explanation are Z-NAND from Samsung and 3D XPoint from Intel and Micron. The products offer lower latency (though not necessarily higher bandwidth) than all other flash products on the market. They also offer better wear characteristics making them suitable for caching applications. It is believed that Z-NAND is a combination of single level cell (SLC) flash along with improvements in the architecture of the companion flash controller. It appears that Samsung created this technology to hold off 3D XPoint from moving up in the flash space. As described previously, 3D XPoint is a completely different technology and is now mainly positioned to compete in the persistent memory segment.

A flash competitor that warrants special attention is the Yangtze Memory Technologies (YMTC). It is a China-based company that is supported by the Chinese government as flash technology is considered a key technology area. They are currently producing 20,000 64-layer wafers a month and are in the early production of a 128-layer part that was implemented by string stacking of two 64-layer parts. Given what has happened with Chinese intrusion into other markets, there is great concern as to the impact in this market. Trendforce's comment on this subject is probably the most telling when they declared: "YMTC's impact on the future market is inevitable and unstoppable."

Along with flash capacities, strides have also been made in flash controllers. A controller along with some number and capacity of flash chips packaged together comprise a solid-state drive. The current generation of enterprise drives support NVMe4 (Gen 4 PCIe) and are capable of sequential reading at greater than 7GB per second and writing at greater than 5GB per second. They can support more than 1 million random small block read and write operations per second (random I/Os). These drives currently sell for between \$150 and \$200 per terabyte. The next generation of drives will support NVMe5 (Gen 5 PCIe) and will be capable of reading and writing at 10 GB/s+. Samsung has announced their intentions of shipping Gen 5 PCIe part that will hit speeds of over 13 GB/s. Consumer drives based on the SATA interface typically use quad level cell (QLC) technology and, therefore, have poorer wear-out characteristics. They also have substantially lower performance, usually in the 500 MB/s range, and can be purchased for around \$100 per terabyte.


Flash requires the least amount of physical space per capacity of all the storage technologies. Much hype has been made regarding which technology has the largest capacity in a 2.5-inch disk form factor, with some vendors announcing capacities of up to 100TB. Those statements are misleading. The only advantage of such devices is that the cost of the controllers can be amortized over a greater amount of flash, and fewer slots are required in a chassis. But, both of these costs are trivial compared to the cost of the flash. The disadvantage of large flash SSDs is that one controller needs to service a large capacity. A better approach is to maintain the ratio of one controller to a reasonable amount of flash. In order to address this issue, new smaller form factors have been created.



These form factors allow for plugging many flash disks into a small chassis rack. These emerging systems provide high capacity, high bandwidth and low latency all housed in a 1U chassis. These new form factors are a break from flash being backward compatible with chassis designed for magnetic disk systems, and more importantly, they represent the end of magnetic disk being considered a primary storage device.

Flash drives were constrained by the physical interface as well as the electrical interface -- either SAS or SATA. These interfaces added latency to data transfers that, with magnetic disk, were “in the noise,” but with flash, became major contributors to overall performance. For this reason, the industry has moved to an interface that directly connects flash drives to the PCIe bus. The NVMe interface can be thought of as the combination of non-volatile memory (NVM) and PCIe. Though the specifications for the NVMe have been around for several years, it has only been over the last few years that adoption has started to take off. This lag was primarily caused by the desire to maintain backward compatibility with magnetic disk systems’ chassis. As a proof point that NVMe is now the predominant interface for enterprise flash, many of the current generations of enterprise flash drives only support NVMe and do not offer SAS versions. NVMe is following a parallel path and similar labelling as PCIe. For example, NVMe3 indicates NVMe running over a PCIe Gen 3 bus while NVMe4 indicates NVMe running on a PCIe Gen 4 bus.

NVMe is great for moving data from a processor to flash drives inside a chassis, but in order to fully displace SAS, it required the creation of a technology that allowed for a box of flash drives, referred to as a JBOF (just a bunch of flash) to be accessible by one or more controllers. This needed to be done without adding substantial latency on top of the NVMe protocol. The technology developed is referred to as “NVMe over fabrics” (NVMe-oF). The fabric can be PCIe (e), infiniband, SAS or Fibre Channel, but for new systems, it will predominantly be remote direct memory access (RDMA) over converged Ethernet (RoCE). With this latter technology, the physical connection between the controllers and the JBOF is commodity Ethernet. RoCE technology is becoming a commodity both at the chip and HBA level. RoCE technology will find rapid adoption for all interconnections that require high bandwidth and low latency. This includes interconnections between clients and block or file controllers, interconnections between those controllers and the shared storage JBOFs, and the interconnections between cluster members in scale-out storage solutions. Most JBOF chassis run a x4 (four lane) PCIe Gen 3 connection to each flash drive. Since each lane is capable of about 1 GB/s, an individual drive is restricted to reading or writing at 4 GB/s. Currently enterprise SSDs are designed with this limitation in mind; however, going forward, with the advent of more performance per flash chip as a result of more layers, and PCIe Gen 4 shipping in greater volume, we are now seeing NAND-based TLC SSDs exceeding 7 GB/s read and 5 GB/s write. If even higher performance is required, as mentioned earlier, Micron has announced the X100 SSD based on 3D XPoint technology that boasts 9 GB/s write and read. It is expected that this product will be priced much higher than equivalent capacity NAND-based enterprise SSDs. It is not understood, outside of a few engineers at Intel and Micron, what challenges exist in making this technology denser and cheaper. It could be that it will always be relegated to the high-performance niche part of the market.



Along with the physical interface changes described above, work is being done on developing new logical interfaces. Samsung, for example, has announced a key/value store interface. This interface is appropriate for object level systems whereby a system can take incoming objects and create some number of fixed size hunks (i.e., values) by either splitting larger objects or packing together smaller ones. These then could be distributed out to a group of SSDs that have this key/value interface thereby reducing the complexity of the application.


Another Samsung initiative involves virtualization such that a single SCSI SSD can be split into 64 virtual SSDs. This is useful in virtual processing environments where multiple virtual machines are running -- with all wanting access to some amount of independent storage from a single SSD. In this case, the hypervisor no longer needs to perform memory mapping between the virtual machine and SSD. Western Digital is taking a different approach by providing a logical interface that allows an application to manage the underlying zones (i.e., flash blocks) of an SSD directly. This effort is called the “Zone Storage Initiative” and applies to all “zone” storage device types which include flash and shingled magnetic recording disk. Regardless of the media type, a “zoned” storage device is one in which the storage is broken into equal size areas (i.e., zones) with properties that allow them to be written sequentially without being arbitrarily overwritten. In mid-2021, Samsung also announced that they will be supporting the ZNS (Zoned Namespace) on future SSD products.

## Flash Storage – Zone-Based Interface

In order to fully grasp the advantages of the zone-based interface and why it will be highly adopted by cloud providers, an understanding of the basic operations of flash controllers and the flash storage they control is required. Flash storage is broken down into blocks. Those blocks, typically 8MB in size, are the smallest segments of storage that can be erased and also can “wear-out” after some number of writes. When data, for example, a 4KB hunk, is written to a specific logical address (LBA), the flash controller algorithm decides what flash chip and which block on that chip it should be written to. If data is written to sequential LBAs there are no guarantees that the data will be placed onto the same flash chip and block. In fact, it is almost guaranteed that the data will be distributed across multiple flash chips in order to achieve the combined performance of those chips. The flash controller maintains two tables: one maps LBAs into their physical locations (which chip, which block on the chip, and what location on the block); and the second keeps information on each block (how many times written, how much free space). When a previously written LBA is rewritten, the controller writes the new data to a new block location and it updates its LBA table to point to that new location. It also updates its block table to indicate that the LBA’s old block location now contains stale data (i.e., garbage).

When the SSD is new and there are many free blocks, writes and rewrites are handled by the flash controller as described above. However, when storage blocks start becoming scarce, it becomes necessary for the controller to start the “garbage collection” (GC) process. This process involves searching the block table to find the block with the most “garbage,” reading the non-garbage from that block, writing the non-garbage to an available block, erasing the original block, and updating the LBA and block table accordingly. Once in this state, the controller attempts to balance servicing incoming data requests with the need to run GC. Besides the performance impact of running GC, it also drives faster wear out of the flash blocks as data that was written once at the interface may be moved to multiple blocks overtime. This is typically known as write amplification. To ensure that there is enough storage available to handle wear out and that GC will not have to be performed frequently, flash controllers do not present the full capacity of the flash they are managing. The percentage of storage that is “held back” is





known as overprovisioning. The amount of overprovisioning varies depending on the SSDs longevity specification, usually specified as the number of full drive writes per day (DWPD). This number specifies how many times the full SSD capacity can be rewritten, each day, for the length of the warranty. For example, an enterprise SSD that has 3 DWPD with a five-year warranty would require 20% or more of overprovisioning. A consumer SSD would have a much lower DWPD specification but would still require substantial overprovisioning because QLC flash wears out at a lower number of writes than the TLC used in enterprise controllers.

By utilizing a zone storage interface, it is possible to have very little of the SSD allocated for overprovisioning while, at the same time, completely avoiding the need for running the GC process. Consider a customer-facing application whereby, in order to support the workload, the majority of customer requests need to be serviced from a flash tier of storage; however, it would be cost prohibitive to store all customer data on flash forever. If the application profile is such that the number of accesses on data is related to the age of the data, then a two-tier system where older data is migrated to magnetic disk would be appropriate. As data enters the system, it would be “packed” into zone-size chunks that correspond to the block size of the flash device, typically 8MB. Each chunk would then be written to a zone on the SSD, specified and tracked by the software. The SSD would determine the best physical blocks and write the chunks to them while maintaining a map of the logical zone to physical block. As more data entered the system, the process would repeat and the SSD would start filling up. When migration time arrives for a specific zone, the software would read that zone and write it to magnetic disk. The software would now record that zone as available and reuse it for new incoming data. When a new 8MB chunk of data is written to that same zone, the controller selects a new available block, writes the data to that block and performs a block erase on the block that was previously associated with that zone. This process continues until the flash system starts wearing out at which point the flash controller does not accept any more zone writes. In summary, an application that utilizes the zone storage interface benefits in three ways: 1) very little storage is wasted for overprovisioning; 2) writes only occur one-time so there is no write amplification and therefore the flash exhibits longer life; and 3) there is no performance impact of GC running as a background task.

New market segments in gaming and automotive are increasing the demand for flash as they are both essentially brand-new markets for the technology. The enterprise SSD market continues to be a growth segment as IT shops continue to make the move from magnetic to solid-state drives. The client market for SSDs is also increasing as SSDs become the standard for laptop and desktop computers. The lack of performance improvements of magnetic disk will drive cloud companies to spend a greater amount of their storage budgets on flash-based technology. Consumer demands for increased capacities for phones, cameras and laptops are waning and being replaced by demands for lower costs for the same capacities. For the purposes of this document, we are only tracking the contribution to the digital universe of enterprise and client SSDs.

## Digital Universe Flash

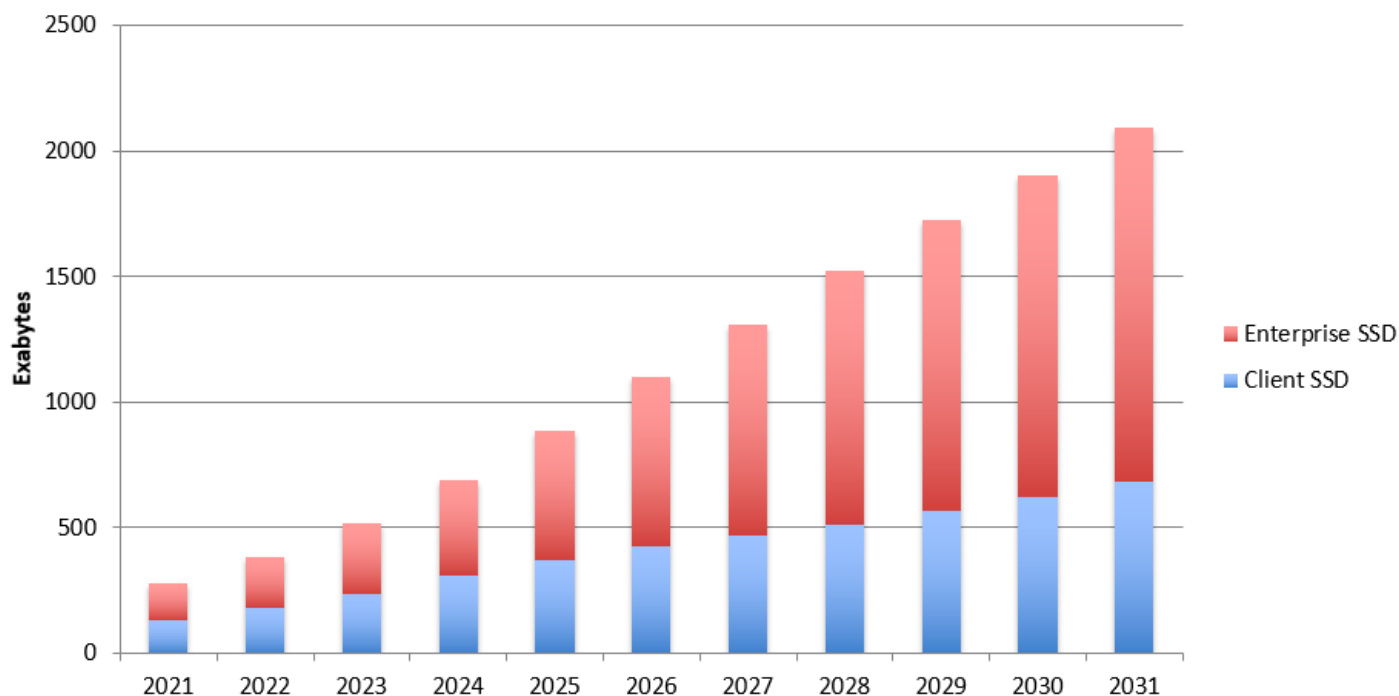


Figure 2: Digital Flash Universe

## Magnetic Disk

For many years now the disk drive industry has had three major providers: Seagate, Western Digital and Toshiba. It is interesting that two of those suppliers, Western Digital and Toshiba, also share flash fabrication facilities and, as such, are not as exposed to disk displacement by flash. However, for Seagate, it is essential that the disk segment continues to be healthy. Looking at the disk drive volume shipments from the last year, we see the volumes shipped over the last four quarters to be 259 million compared to 255 million for the prior year's four quarters. This is the first time in over a decade the volume of units, year-to-year, has not declined.

All consumer and 2.5-inch high performance categories were down. More recently, the major game console manufacturers have introduced their next-generation products that all use NAND flash rather than small hard drives. We expect this will accelerate the demise of this category of disk drives over the next few years. Accepting this, the disk manufacturers have been disinvesting in research and development of these drives as can be seen by the lack of any capacity improvements over several years. The segment that did see year-to-year increases in both capacities and volume shipments is the 3.5-inch nearline drive category. It now comprises more than 65% of all disk revenue. Developing a singular product, with a few variations, has allowed the disk companies to focus their resources, enabling them to remain profitable even as a good portion of their legacy business erodes.

### Magnetic Disk Volumes (millions)

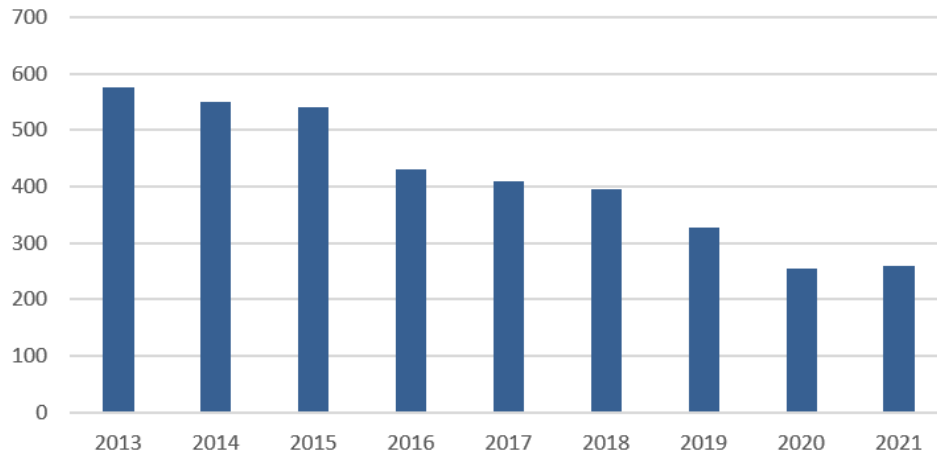


Figure 3: Magnetic Disk Volumes

More and more, the disk industry will be shipping a singular product line, that being high-capacity 3.5-inch nearline drives. These are sold predominantly to large IT shops and cloud providers. Though there will be several versions of these drives, their base technologies are identical allowing all versions to be produced off the same manufacturing line. Variation of these drives will be in the areas of single or dual actuator, shingled or conventional recording, and SAS or SATA interface. In order to sustain that market, their products must maintain current reliability, while at the same time, continue to decrease their per-capacity cost. Protection of market share requires a multiple cost differential over consumer solid-state disk technologies.

Heat Assisted Magnetic Recording (HAMR) increases the areal density of a disk platter by heating the target area with a laser. This heated area is more receptive to a change in magnetic properties (reduced coercivity), allowing a lower and more focused charge to “flip” a smaller bit. The media then immediately cools and regains its high coercivity properties thereby “locking” the bit into place such that it requires a strong magnetic field to reverse it. For this reason, this technology is thought to be able to store data for a longer time than traditional disk technology. Microwave Assisted Magnetic Recording (MAMR) uses a microwave field generated from a spin torque oscillator (STO). In this method, the STO located near the write pole of the head generates an electromagnetic field that allows data to be written to the perpendicular magnetic media at a lower magnetic field.

For many years the disk industry has been investing heavily in HAMR and MAMR technology, realizing its importance for the product roadmap. The two predominant industry leaders, Seagate and Western Digital, are taking drastically different strategies in moving to higher capacity drives. In the case of Seagate, we believe it is HAMR or bust as the entire future roadmap depends on this technology. Alternately, Western Digital is taking a more incremental approach. The first technology, called eMAMR, will be used to enable drives of 20TB and 24TB;

full-blown MAMR will be used for drives starting in the 30TB range; and HAMR will be used to achieve 50TB drives by 2026. They are also claiming that the SMR versions of these drives will have a minimum of 20% greater capacity and, as such, the 20TB drive will be 24TB in its SMR version. The 24TB drive will have a 30TB SMR version, the 30TB drive will have a 36TB SMR version, and the 50TB will have a 60TB SMR counterpart.

Western Digital has found a way to add a 10th platter to the drive and, therefore, will be shipping a 22TB CMR and 26TB SMR in 2022. They have also stated that they have the technology available to create a 30TB CMR and a 36TB SMR (assuming a 20% SMR capacity increase). <https://www.tomshardware.com/news/western-digital-30tb-hdds-incoming> Seagate has also announced that they are shipping 22TB drives to select customers with 30TB drives being available in 2023. <https://arstechnica.com/gadgets/2022/01/seagate-starts-shipping-enormous-22tb-hard-drives-to-some-customers/> These drives are not available on the open market.

Regarding all of the larger future capacity drives, there has been very sketchy information as to what the cost per gigabyte of these drives will be. Given the complexity and R&D dollars spent on developing these products, we predict that, at least for the next few years, these drives will provide a cost decrease that is less than their capacity increases. For instance, going from a 16TB to a 24TB drive yields a 50% greater capacity but may only be priced at 15% less per gigabyte. For Exascale data centers, the greater capacity does provide additional benefits in that it leads to requiring fewer racks, less power and smaller footprints -- all important considerations. A problem for these large capacity drives is one of storage density as the number of I/Os performed by the device remains essentially the same as the capacity increases. To counter this issue, both Seagate and Western Digital have introduced dual actuator disk drives with the possibility of three or four actuators in higher capacity drives.

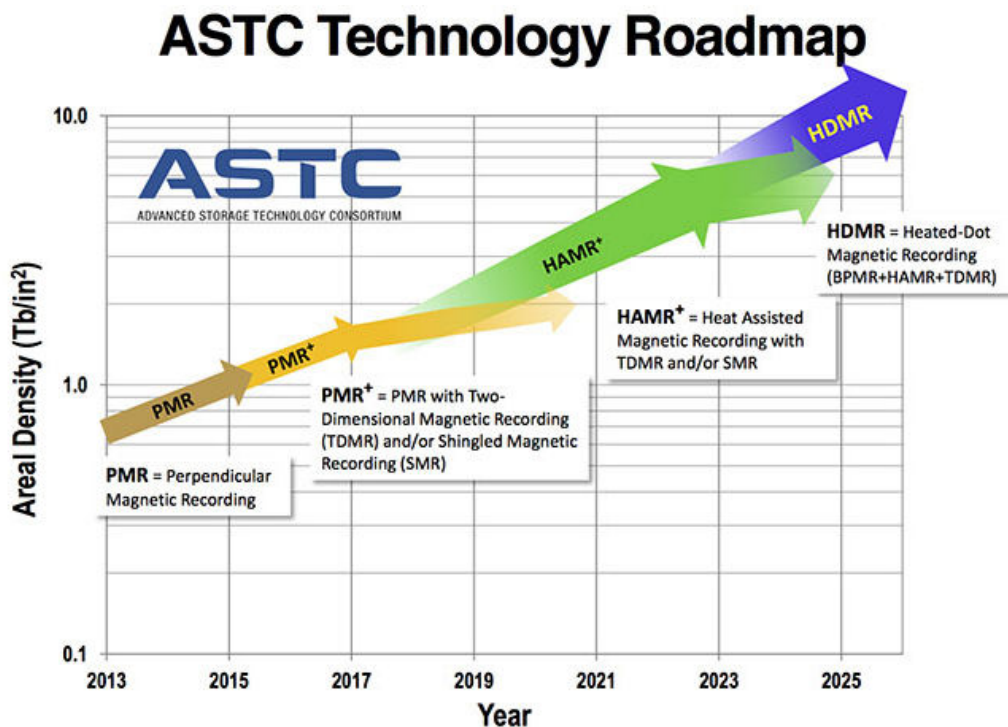



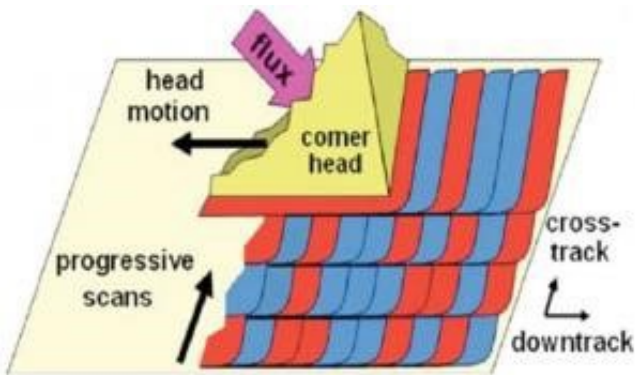
Figure 4: ASTC Technology Roadmap



Due to the delay of the HAMR technology, the Advanced Storage Technology Consortium (ASTC) substantially revised its roadmap. The new roadmap appears to be accurate thus far as it shows PMR technology having been phased out in 2021 with HAMR or MAMR being the technology driver for higher capacity drives going forward. Keep in mind that this is a technical not a product roadmap and, as such, PMR drives will be sold for many years to come. It appears that the disk drive vendors were able to add another platter and, therefore, create PMR drives of 22TB. Capacity drives greater than this will require advanced technologies such as HAMR.

An advancement that was announced by Western Digital was the addition of flash technology inside the disk drive. This initiative, called OptiNAND™, is unlike previous hybrid flash/magnetic drives whereby the flash was used as a cache for the magnetic disk. The OptiNand technology instead provides two fundamental improvements to the drive. When a track on a disk drive is written, it effects the margin for reading the tracks around it. This is called adjacent-track interference (ATI). After some number of writes, modern disk drives will read the adjacent tracks and rewrite them prior to them becoming too degraded. This process slows down the operation of the drive as a single write might involve three writes and two reads. In older generations of disk drives, this interference was only an issue after several thousand writes of a track and, therefore, the impact on performance was small. Information regarding tracks that might be of concern was kept in the DRAM of the disk drive; however, given the size of the DRAM, the information was very granular resulting in more tracks being rewritten than necessary. With the advent of much higher capacity drives, the tracks are so close together that ATI can render tracks needing to be rewritten after fewer than 10 writes of an adjacent track. Without a change, this would result in substantial performance degradation. Given that the flash component of the disk drive has a much higher capacity than the DRAM component, fine grain information can be stored in the flash component such that only tracks that need to be rewritten are rewritten. The flash component also provides improved performance for applications that require syncing of the disk drive. Disk drives have a DRAM that acts a write cache to the magnetic drive. When a sync is sent, the drive is required to respond by persisting to the magnetic medium any writes that are in the cache. This assures that, if power is disrupted, the data is not lost. With OptiNAND, on power down, any pending writes held in DRAM are automatically persisted to the flash component using the power produced by the inertia of the spinning disk. When the drive is powered back up, the writes that were held in flash are written to disk. This provides performance improvements in that a sync event can be acknowledged immediately. It also allows the drive to reorder writes in a way that reduces the mechanical distance the actuator must seek.

Given the simplification of the disk drive roadmaps into a single nearline product line, two paths for controlling these drives are emerging. One path will support the traditional disk SCSI command set thereby satisfying the current storage markets, such as Network Attached Storage (NAS). These drives will be formatted in conventional media recording (CMR) mode which will prevent the need to rewrite disk-based applications. In this mode the disk drive virtualizes the physical placement of the data from the application. The other path will be for cloud companies and is for products that are specifically designed to store fixed content. A drive that supports this interface is known as a host-managed SMR drive, which is essentially the “zoned” block interface discussed earlier in the flash section of this paper. These drives cannot be purchased on the general market as the disk vendors ensure that they are only placed into environments that have been explicitly designed to support them. SMR takes advantage of the fact that the read head of a disk drive is smaller than the write head. This allows for tracks to be written in an overlapping manner as shown in the diagram below. This leads to a capacity increase of up to 20% vs. the same drive formatted in CMR mode, as the normal track-to-track gaps are eliminated. A side effect is that a track cannot be updated as doing so would overwrite the tracks around it. For this reason, an SMR drive is broken into “zones” whereby each zone is a region on the disk, typically 256MB in length.



In the prior flash section of this paper we provided a detailed explanation of how zone-based storage can be best utilized for flash storage. The same is true for disk storage with the exceptions being that disk zones are of larger capacity and never need to be moved due to wear-out issues. Besides the capacity advantage, other advantages exist in the areas of improved write/read sequential performance and allowing the host to physically place data into zones that match the performance needs of the data. The performance of a zone corresponds to

where it exists on the physical disk. Faster zones are at the outer diameter of the disk while slower zones are at the inner diameter. The performance of the fastest zone to the slowest zone is roughly 2.5 times, which corresponds to the ratio of the circumferences of the disk at the outer edge and the inner hub. Zone-based SMR disk storage is suitable for workloads whereby large hunks of data can be grouped together, and operations such as migrations can occur on the entire group at the same time. This workload is very consistent with those found in fixed content applications. For these reasons, it is projected that the percentage of zone-based disk storage vs. conventional disk storage will steadily climb over the next three years mostly due to cloud companies moving toward purchasing only SMR drives in the future.

## Percent of Enterprise Disk Capacity

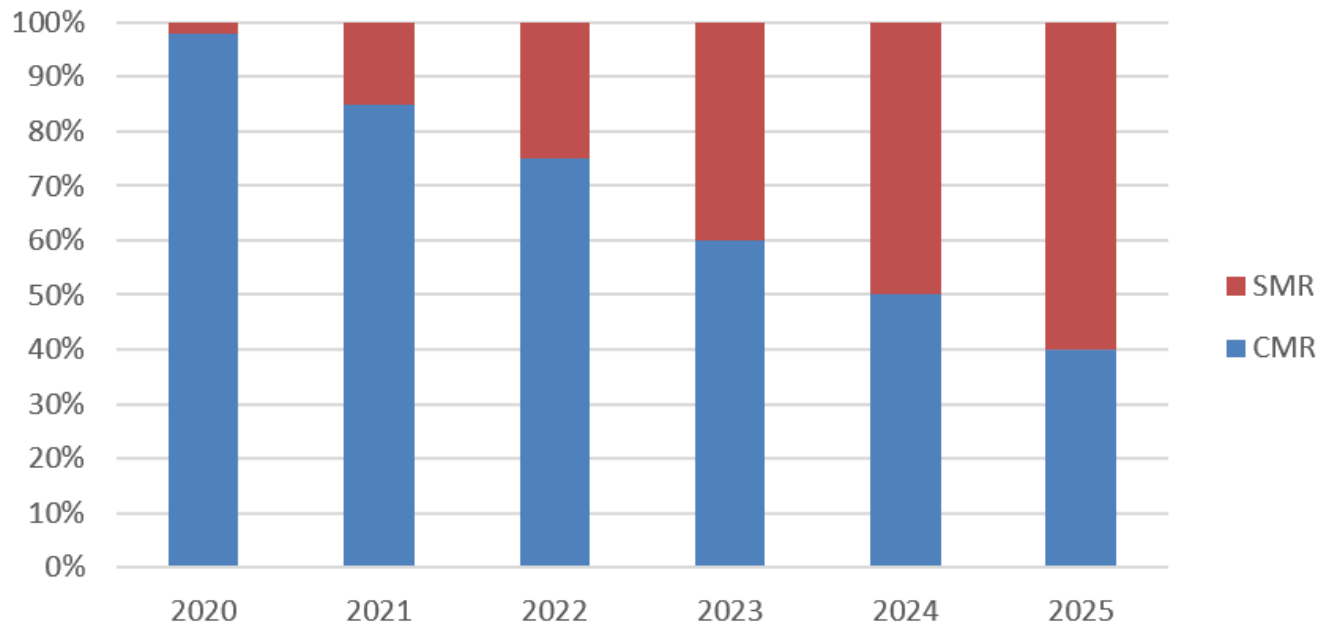


Figure 5: Disk Capacity of SMR/CMR

As noted in the architecture section of this paper, we predict that magnetic disk drives predominantly will be used as community storage. Given that community storage will contain data that is not frequently updated, a drive that has high capacity but not a particularly good random I/O performance will be adequate. However, for cloud providers, the story is completely different. Cloud providers have two-tier architectures with the first being flash and the second being magnetic disk. They have years of data collected and fully understand the workload patterns of their various applications. Many of these applications have access patterns that are time-based in that the older the data, the less frequently it is recalled. With this information, they can derive how many I/Os will need to be serviced by the flash tier vs. the disk tier. Given the large discrepancy between the I/O performance of flash and disk, it is important that the majority of I/O requests are serviced from the flash tier while it is preferable that the bulk of the data be stored on more cost-effective disk. The lower the I/O rate of the disk tier as a function of capacity (I/Os per TB), the more flash will need to be purchased to avoid backing up I/Os on the disk tier, resulting in time delays to the consumer. This is a chicken and egg problem in that if the disk industry overcomes the technical challenges associated with increasing capacity, then they have to face how to improve I/O performance at the same rate (or better) than the capacity increases.



## Digital Universe Disk

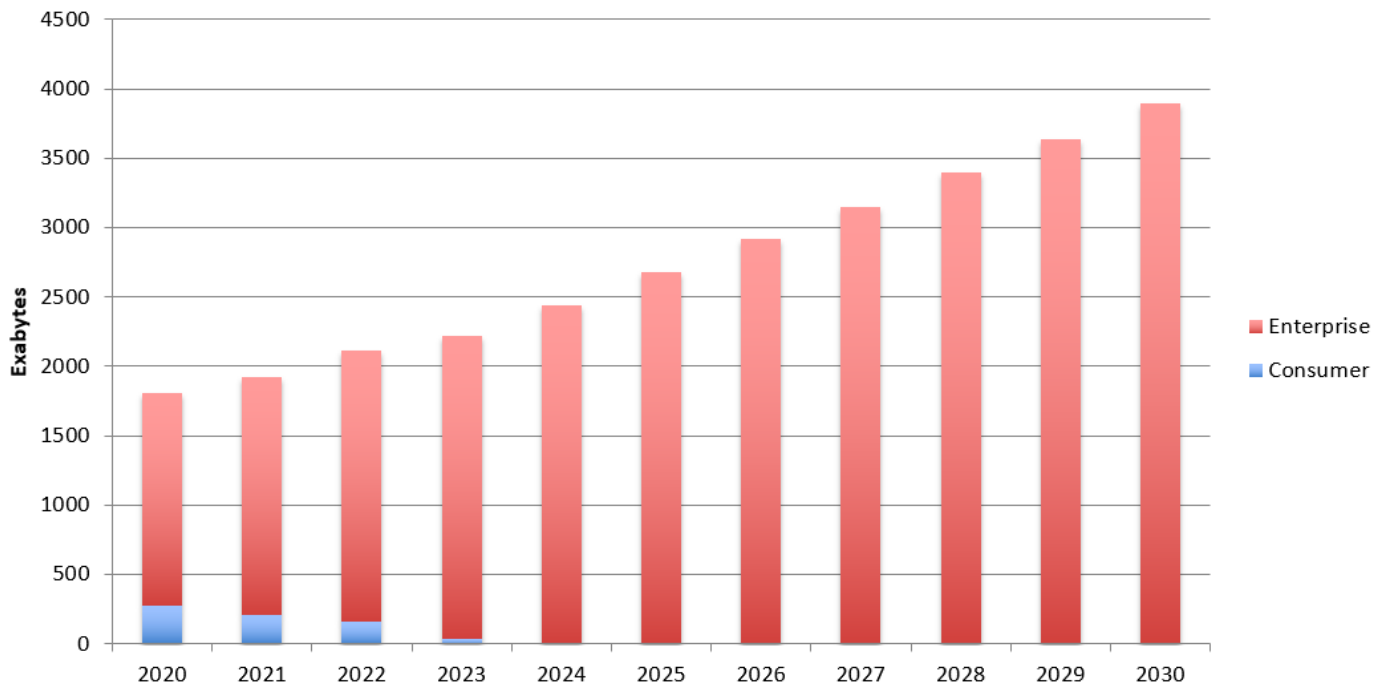


Figure 6: Digital Universe Disk

For example, if a cloud company buys 12TB drives today, and in a year, 24TB drives with the same I/O rate as the 12TB drives become available, then the organization might opt to buy only half the amount of 24TB drives to fulfill its capacity requirements, and buy more flash drives to gain the additional I/Os not provided by the disk tier. As discussed above, one technique that improves I/Os is through strategic placement of SMR zones with different speeds. A second technique being touted by both Seagate and Western Digital is to add additional independent actuators inside the drive. This is a blast from the past in that most disk drives manufactured prior to 1990 had two or more actuators. This automatically doubles the I/O performance of the disk drive; however, the cost of the second actuator could result in a 20 percent to 30 percent price increase. It is to be determined as to whether the cloud companies will see enough benefit to justify the cost. Non-cloud deployment applications that require high I/O will move to flash as it has two orders of magnitude better performance than that of even dual actuator disk drives.

As seen above, Spectra is predicting a very aggressive decrease in the aggregate shipped capacity of consumer magnetic disk as flash disk takes over that space. Capacity increases in enterprise storage will not maintain a pace that will allow the disk industry to realize volume or revenue gains.

Some reservations are warranted as to the market's ability to deliver advanced technologies and restart the historical cost trends seen in disk for decades. If the industry is unable to cost effectively and reliably deliver on this technology, the intrusion of flash into its space will be greater.



## Tape

The digital tape business for backing up primary disk systems has seen year-to-year declines as IT backup has moved to disk-based technology. At the same time, however, the need for tape in the long-term archive market continues to grow. Tape technology is well suited for this space as it provides the benefits of low environmental footprint on both floor space and power; a high level of data integrity over a long period of time; and a much lower cost per gigabyte of storage than any other storage medium.

A fundamental shift is underway whereby the market for small tape systems is being displaced by cloud-based storage solutions. At the same time, large cloud providers are adopting tape -- either as the medium of choice for backing up their data farms or for providing an archive tier of storage to their customers. Cloud providers and large scale-out systems provide high levels of data availability through replication and erasure coding.

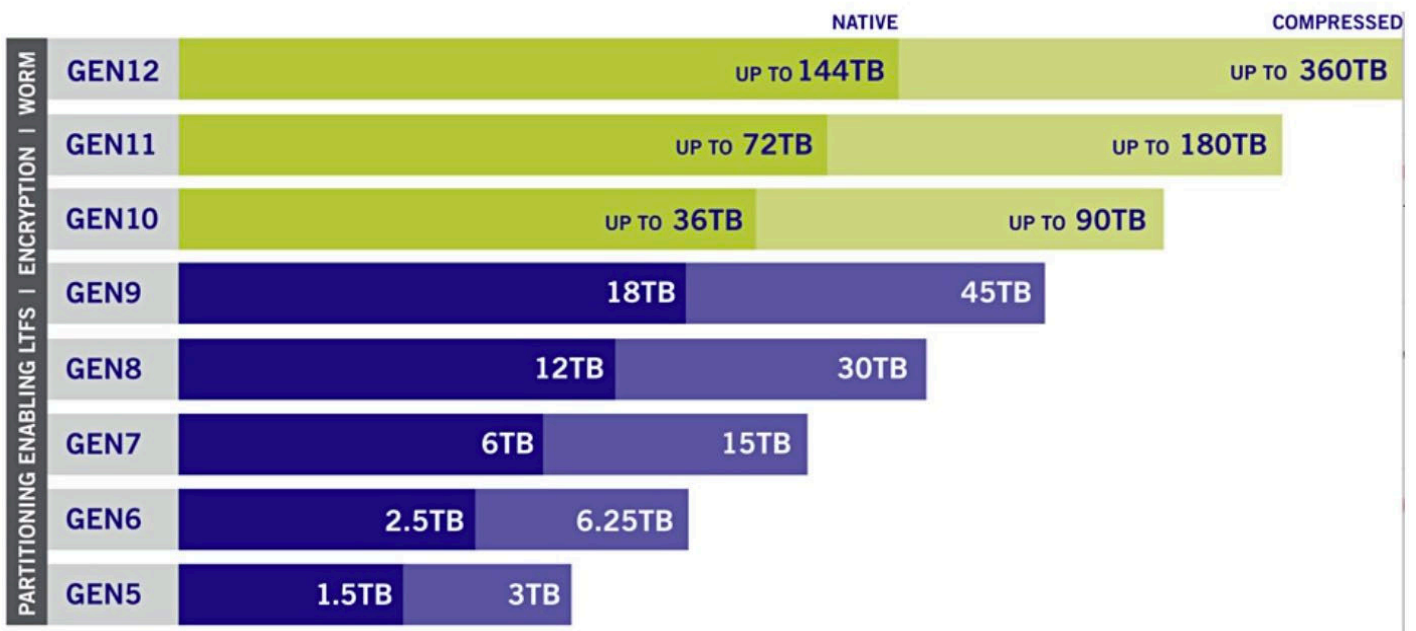
These methods have proven successful for storing and returning the data “as is.” However, if during the lifecycle of that data, it becomes corrupted, then these methods simply return the data in its corrupted form. For the tape segment to see large growth, a widespread realization and adoption of “genetic diversity,” defined as multiple copies of digital content stored in diverse locations on different types of media, is required to protect customers’ digital assets. More recently, due to ransomware and other forms of attacks, we are seeing a greater interest in using tape as a last means of defense. Tape creates an air gap, an electronically disconnected or isolated copy of data, either in a library or stored offline, that prevents the data from being infected, unlike data that resides on systems connected directly to the network.

Linear Tape Open (LTO) technology has been and will continue to be the primary tape technology. The LTO consortium assures interoperability for manufacturers of both LTO tape drives and media. In 2021, the ninth generation of this technology was introduced, providing 18TB native (uncompressed) capacity per cartridge. Each generation of tape drive has been offered in both a full height and a more cost-effective half-height form factor.

As seen in the following table, the LTO consortium is providing a very robust roadmap in terms of future products all the way to LTO-12 at a capacity point of 144TB on a single piece of media. The majority of capacity increases will be gained through adding more tracks across the tape rather than increasing the linear density of the tape. The challenges for realizing this roadmap are multi-fold.

Tape, from a capacity perspective, has a large surface area, which means it has a much lower bit density than that of current disk drives; however, as a removable media, interchange of cartridges between drives requires that the servo systems have enough margin to handle variances in drive writing behaviors. This variance is directly correlated to how precisely the tape can be moved across the tape heads at high speed. A rule of tape drive design is that the longer and heavier the tape path, the better the tape can be positioned. This presents a challenge to the half-height drive as its tape path is shorter and lighter than that of the full-height drive. This was the primary reason that LTO-9 came out at 18TB rather than the originally stated 24TB. The half-height drive just didn’t have the stability in the tape path to support the number of tracks required for an 18TB tape. LTO-9 cartridges also must be preconditioned prior to use. Customers have the choice of purchasing from vendors that sell tapes unconditioned or preconditioned, with the latter having a slightly higher cost. When an unconditioned tape is mounted into an LTO-9 drive, the drive senses that the media is unconditioned and starts the conditioning process. Depending on several factors, the conditioning process can take from a half an hour up to two hours.

The conditioning process determines the elasticity of the width of the tape and stores that information in the memory component of the tape cartridge. This information is read by the drive each time the cartridge is mounted such that the drive can compensate for individual characteristics of that cartridge. If a previously conditioned cartridge is reformatted, then the conditioning process will be performed again. The limitation of the half-height drive along with the need to determine the elasticity of the tape, indicates to us that IBM will be required to develop newer technologies in order to increase capacity as per the LTO roadmap. It is also possible that there may be a divergence between half-height drives and full-height drives unless IBM is able to solve these problems. This may result in half-height drives needing to be slowed down, resulting in lower bandwidth, or operated at lower capacity points. Another challenge for tape is that, as tape capacities have increased, the bandwidth to read and write a tape has improved at a much slower rate. For cloud companies, who measure their performance as a function of capacity, this poses a problem as they are required to purchase a larger number of drives for each successive tape generation. For example, a customer's tape system requirement might be that, for each petabyte stored, there should be 360 MB/s bandwidth available. For LTO-8, this would be satisfied by using a ratio of 84 cartridges (1000TB / 12TB per cartridge) per drive. Looking into the future and considering an LTO-11 drive that can transfer at 500 MB/s, this would result in a cartridge-to-drive ratio of 14. Given that the "true" cost of a tape cartridge is the cost of the media plus the cost of the drive divided by the number of cartridges per drive, this will erode some of the cost-per-capacity advantages of tape.



*New LTO roadmap with 18TB raw LTO-9 tape*

**Figure 7: LTO Tape Roadmap**

There are three primary methods for improving tape drive performance, each posing its own challenges. First the tape cartridge can be run across the tape head faster. As stated previously, the bigger the tape path the better the control, which in turn drives higher drive costs. We believe this method will provide very little uplift of bandwidth in future generations. Another option would be to increase the linear bit density of the tape. This requires a more advanced media formulation similar to the switch from metal particle used on LTO-6 with a linear density of 15,142 bits/mm to the barium ferrite LTO-8 at 20,668 bits/mm respectively. We believe this method will have some amount of potential upside. The last method would be to increase the number of tracks on the tape head from the current 32 to possibly 64. This would result in a more expensive and complicated tape drive, however, would provide the best method for increasing tape drive performance in a substantial way.

Customers with high duty-cycle requirements can consider using enterprise drives from IBM. IBM is now shipping IBM® TS1160 Tape Technology with a native capacity of 20TB. These tape drives use TMR technology (tunneling magnetoresistance) which should allow the capacity to double three more times and forms the basis for future LTO generations. Additionally, these drives are offered with an optional native high-speed 10GE RoCE Ethernet interface. For customers who require RoCE interface but would prefer LTO drives, there are now vendors selling RoCE to SAS bridges that have been qualified with LTO technology.

With Oracle’s exit from the tape business complete, IBM now is the only tape drive supplier. Fujifilm and Sony are the market’s two tape media suppliers. Similar to other storage technologies, when new generations of tape are introduced, the cost per gigabyte of the technology is priced higher than the older technology on the market.

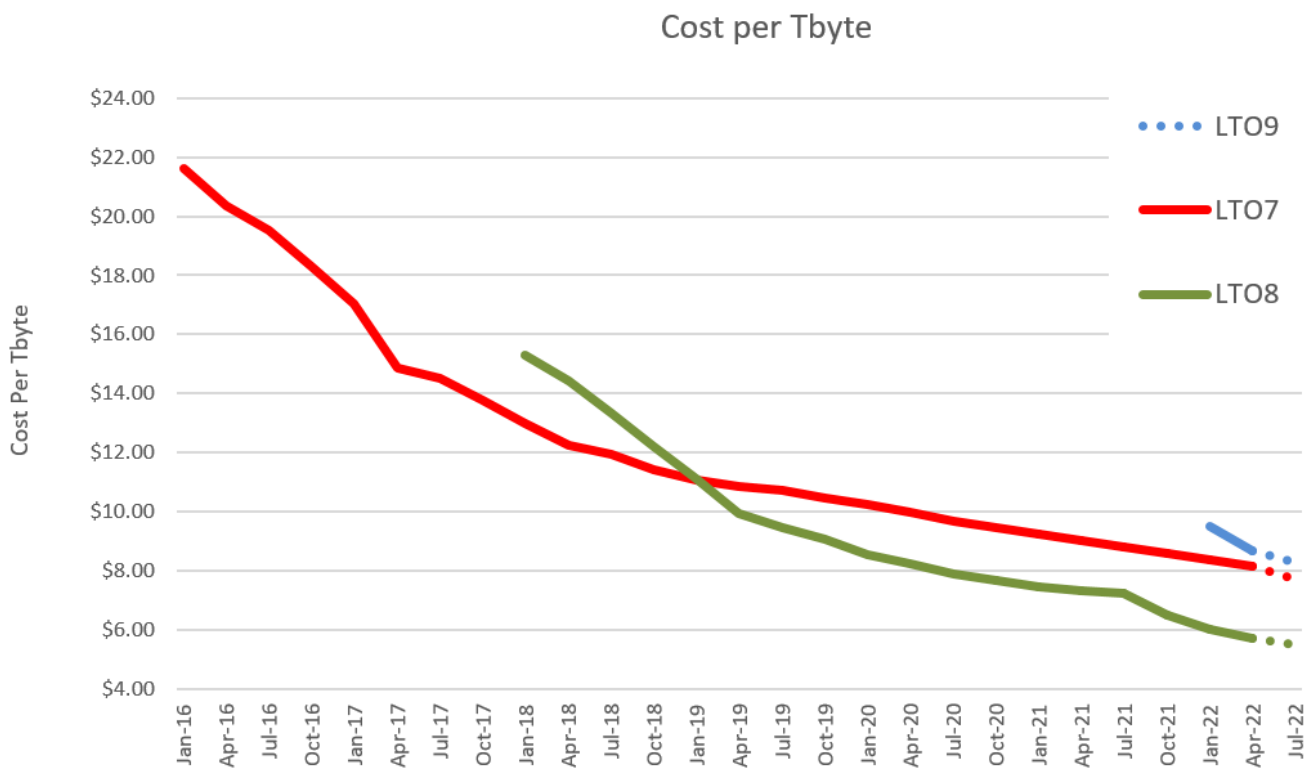


Figure 8: LTO Cost per Tbyte




Figure 8 shows the cost per TB of different LTO cartridges over time. As can be seen at the time of LTO-8 introduction, the price point of just below \$15 a TB was higher than that of LTO-7. It became cheaper than LTO-7 in January 2019. As can be seen, our projections are that LTO-9 media will be more expensive than that of LTO-8 until early 2023.

A historical issue with tape has been the perception that it is “hard to manage.” Tape has typically been supported in two ways: backup applications and hierarchical storage managers (HSMs). In the case of backup software, a substantial portion of the development of the overall product effort has to be dedicated to managing tape. This includes tracking onsite and offsite tape cartridges, interfacing with various tape libraries, and writing and reading to and from tape drives in a manner that allows the drives to perform to their streaming specifications. For these reasons, many newer backup applications have forgone tape support altogether or provide tape support only through an HSM. HSMs attempted to solve the complexity of tape by providing a standard network file interface to an application and having the HSM manage the tape system. This abstraction suffers from two major drawbacks. First, most applications are written such that when they communicate with a file interface, they have expectations of reasonably short file system access times. Given that an HSM might require several minutes to restage a file, many applications will just time-out assuming that something has gone wrong. Another drawback to HSMs is that a file system interface does not provide any information as to what comes next. For instance, a file request may occur that gets mapped into a particular cartridge. The cartridge is mounted, fast forwarded to the correct spot on the tape, the file read, easily the shortest part of the process, and the cartridge rewind and dismounted. This is a process that could take several minutes. Once complete, the application could then ask for another file on the same tape and the process could start all over. It would be beneficial if all the retrievals were known upfront such that the HSM could schedule batch retrievals from tape cartridges in the most optimum manner. This has relegated HSMs to market niches where the applications are aware that they are being backed by an HSM and not a standard disk-based network file system.

What is needed to make tape much easier to manage is an interface that accepts long retrieval times with the capability to specify that an unlimited number of data entities be retrieved at one time. It happens that a de-facto standard interface has emerged that provides this capability. The AWS S3 interface has become more or less the standard object interface to PUT (write) and GET (read) objects into either a cloud or an on-premises object store. AWS also has defined several tiers of storage that vary in their pricing structures. These tiers fit into two broad classes; online (S3-Normal, S3-Infrequent Access, etc.) and offline (S3-Glacier, S3-Deep Archive Glacier).

The offline storage classes are appropriate for storing archival data, essentially data that will be accessed infrequently but kept for a long time. Objects located on an online tier can be accessed directly, but objects in an offline tier need to be restored to the online tier prior to being available for access. The S3 RESTORE command provides the mechanism for an application to specify the objects to be restored. There is no limit to the number of parallel S3 RESTORES that can be issued at one time, and given that restores can take many hours, it is important that the application issue a restore request for each desired object upfront. This is an ideal interface for a tape system. An S3 interface would be presented to the application and all data stored on tape would be mapped as being in an offline tier. The application is hidden from any details of tape management and, at the same time, the tape system could not just manage the tape system, but could provide advanced features such as multi-copy, offsite tape management and remastering -- all done transparently to the application. By having a tape system

that supports this interface, countless S3 applications could utilize tape without need of modifications. Fujifilm, Quantum and Spectra Logic have all announced products with this capability ([www.spectrallogic.com/products/vail/](http://www.spectrallogic.com/products/vail/)).

## Digital Universe Tape

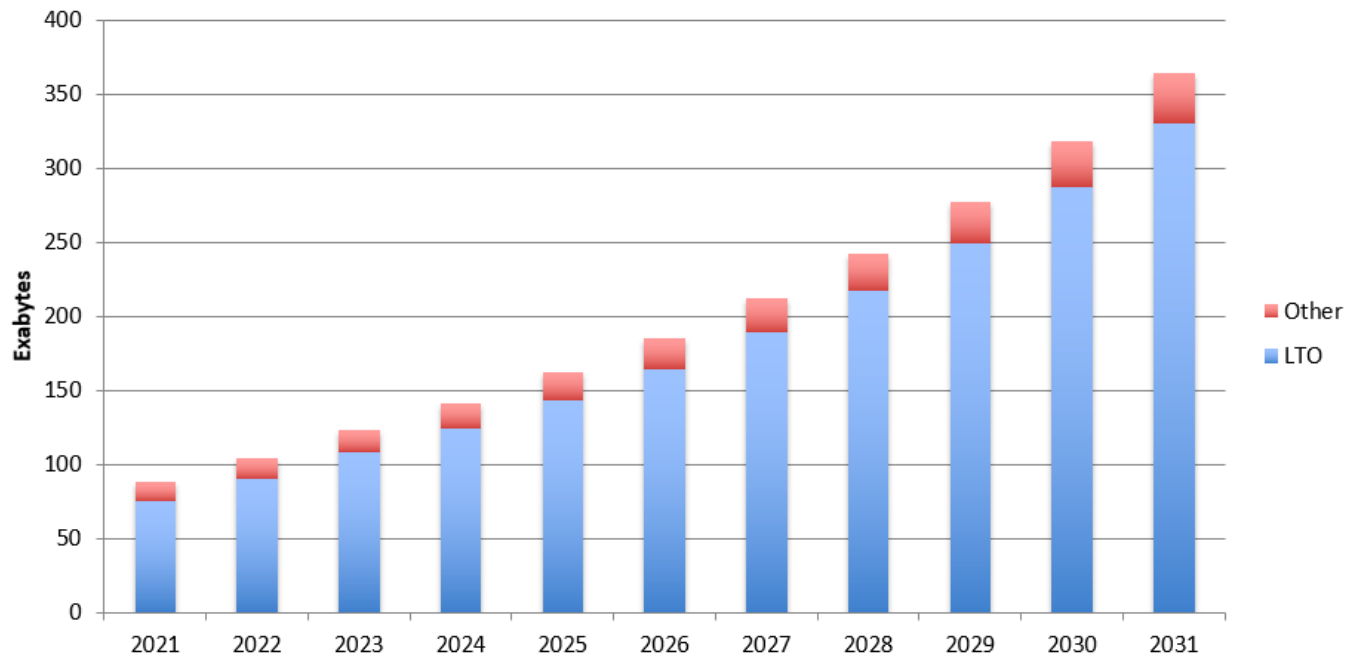
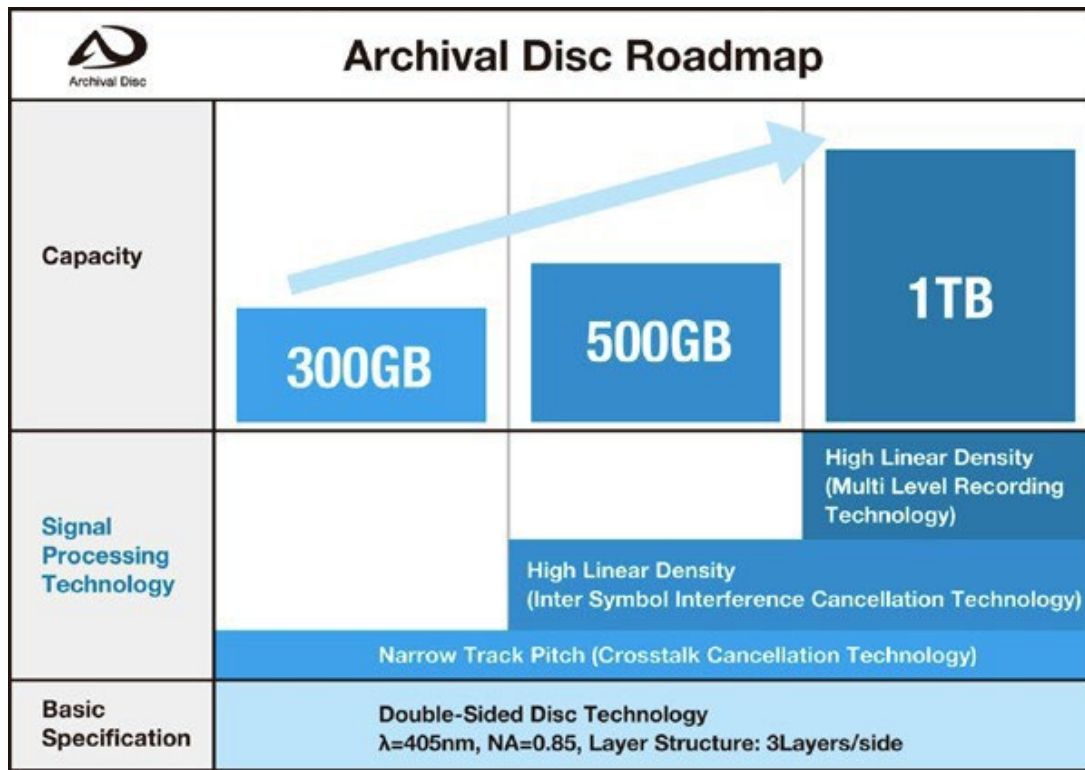


Figure 9: Digital Tape Universe

Cloud providers will mostly adopt LTO, and given their strength in purchasing overall tape technology, this will lead to a greater percentage of LTO shipments versus enterprise tape technology. The challenges for greater tape adoption with cloud providers lies partially in the environmental requirements of tape versus other equipment utilized (e.g., servers, disk, networking). Tighter controls of temperature and humidity are contrary to cloud providers' desire to be "green" by utilizing techniques that save cost, such as using external air. Tape library offerings that solve this problem efficiently without requiring the cloud provider to change their facility plan will find favor.

## Optical

In early 2014, Sony Corporation and Panasonic Corporation announced a new optical disc storage medium designed for long-term digital storage. Trademarked "Archival Disc", it will initially be introduced at a 300GB capacity point and will be write-once. An agreement covers the raw unwritten disk such that vendors previously manufacturing DVD will have an opportunity to produce Archival Disc media. Unlike LTO tape, there is no interchange guarantee between the two drives. In other words, a disc written with one vendor's drive may or may not be readable with the others. Even if one vendor's drive is able to read another's, there may be a penalty of lower performance.

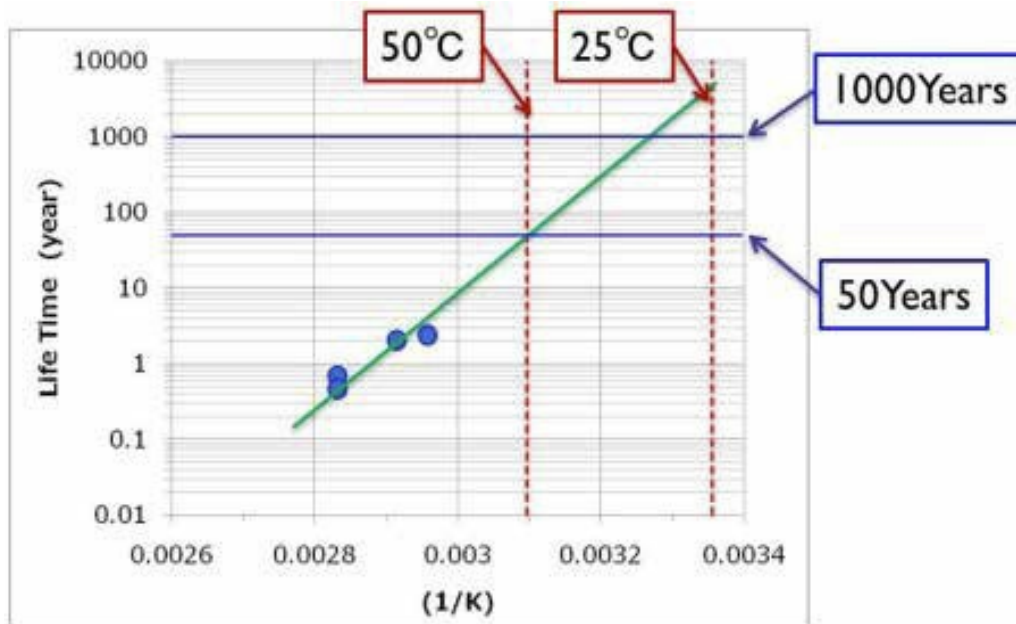


**Figure 10: Archival Disc Roadmap**  
Source: Sony & Panasonic

Sony’s solution packages 11 discs into a cartridge that is slightly larger in both width and depth than an LTO tape cartridge. The value proposition of this technology is its longevity and its backward read compatibility. As shown in the following chart, even stored at the extreme temperatures of 50 degrees Centigrade (122 degrees Fahrenheit), the disc has a lifetime of 50 years or more. Error rates are still largely unknown.

Sony and Panasonic are also guaranteeing backward read compatibility of all future drives. Unlike disk and tape, the media will not require migration to newer formats and technologies as they become available. This matches the archive mentality of writing once and storing the media for extended recoverability. Best practices in magnetic disk systems, conversely, where the failure rate of the devices considerably increases over time, indicate that data should be migrated between three to five years. In 2021, Sony shipped the third generation 500GB per disc product resulting in a cartridge capacity of 5.5TB.





**Figure 11: Long-Term Storage Reliability in Archival Disk**  
Source: Sony and Panasonic

For customers that have definitive long-term (essentially forever) archival requirements, the archive disc will find favor. The size of this particular market segment is small compared with the overall market for archival storage. The ability of this technology to achieve greater market penetration is primarily a function of the pricing of the media and, to some extent, the drives. If the media and drives were cheap enough, this would be an ideal archive media because of its longevity and its less restrictive environmental control requirements. However, this is the seventh year we have reported on this technology and the cost reduction that is required to become competitive has not occurred. Looking at the prices of different substrates that could be considered for archive, we see magnetic disk at less than \$20 per TB, tape at less than \$6 per TB while optical technology sits at just below \$33 per TB. Following is a comparison of the second and third generations of the Sony archive disc product. Note that the cost per storage (\$/TB) has not substantially changed from one generation to the next. It is presumed that the new generation of media is five layers deep. This assumption is based on the fact that the performance of the product has improved at the same ratio as that of capacity.

Product	3.3 TB (Gen 2)	5.5 TB (Gen 3)
Introduced	2017	2020
Cart Price	\$120	~\$183
\$/TB	\$36	\$33
Read BW (MB/s)	250	375
Write BW (MB/s)	125	187
Double Sided	YES	YES
Layers Per Side	3	5

**Figure 12: Optical Disc Comparison**

Since last year the price per terabyte has dropped; however, in order to achieve broader market acceptance, the price of the optical media would still need to see three times the cost reduction on a per-TB basis. This could be accomplished by a combination of increasing the capacity of the disc while at the same reducing its manufacturing costs. This type of breakthrough is being pursued by a company that was spun out of Case Western Reserve University in 2012 by the name of “Folio Photonics”. The Folio Photonics technology depends on effective use of polymer co-extrusion, a manufacturing method that creates a low-initial cost disc technology. The process uses thin, flexible polymer film that can be cut and laminated to discs, so that 64 extremely thin layers that can be read on hardware are designed for that purpose. It is unclear if and when this technology will reach the market. If it does get to market, the question remains as to whether the cost and capacity points will be attractive enough to create a disruptive change in the marketplace.

Given the number of manufacturers and the variety of products (such as pre-recorded DVDs, Blue-Ray, etc.), it is difficult to project the stored digital universe for optical disc. To remain consistent with the intent of this paper, Spectra conservatively estimates the storage for this technology at 5EB per year, and that, with the introduction of the Archive Disc, this will grow at a rate of 1EB per year for the next 10 years. This value could change in either direction based on the previous discussion regarding disc pricing.



## Future Storage Technologies

The storage industry has and will always continue to attract venture investment in new technologies. Many of these efforts have promised a magnitude of improvement in one or more of the basic attributes of storage, those being cost (per capacity), low latency, high bandwidth, and longevity. To be clear, over the last 20 years, a small portion of the overall venture capital investment has been dedicated to the development of low-level storage devices, with the majority dedicated to the development of storage systems that utilize existing storage devices as part of their solution. These developments align more with the venture capital market in that they are primarily software based and require relatively little capital investment to reach production. Additionally, they are lower risk and have faster time-to-market as they do not involve scientific breakthroughs associated with materials, light or quantum physics phenomenon.


Much of the basic research for advanced development of breakthrough storage devices is university or government funded. Once basic research has been completed, the productization of the technology needs to be executed by startups, funded by the venture capital market, or by companies who have special interest in the technology. For instance, Microsoft has interest in developing a long-term storage medium to replace tape, by writing onto glass. The research for this technology came out of the University of South Hampton but the technology effort is moving forward with Microsoft funding under the project name Silica (<https://www.microsoft.com/en-us/research/project/project-silica/>). DNA storage has progressed through various universities and is now being driven by a consortium of companies.

Though these and other efforts have the potential to revolutionize data storage, it is difficult to believe that any are mature enough at this point in time to significantly impact the digital universe through at least 2030. Historically many storage technologies have shown promise in the prototype phase, but have been unable to make the leap to production products that meet the cost, ruggedness, performance, and most importantly, reliability of the current technologies in the marketplace. Given the advent of cloud providers, the avenue to market for some of these technologies might become easier (see next section).

## Cloud Provider Storage Requirements

Over the period of this forecast, cloud providers will consume, from both a volume and revenue perspective, a larger and larger portion of the storage required to support the digital universe. For this reason, storage providers should consider whether or not their products are optimized for these environments. This brings into question almost all previous assumptions of a storage product category. For example, is the 3.5-inch form factor for magnetic disk drives the optimum for this customer base? Is the same level of the cycle redundancy check (CRC) required? Can the device be more tolerant of temperature variation? Can power consumption and the associated heat generated be decreased? Does the logical interface need to be modified in order to allow the provider greater control of where data is physically placed?

Another way to consider the requirements for these providers is to ask the reverse question, which is, 'What is it that they don't need?' Equipment that was designed for IT data centers may have substantial features that add cost and/or complexity to a product that are neither needed nor wanted by cloud providers. Additionally, systems that are managed as separate entities do not fit the cloud model because, within these operations, hundreds of identical systems may need to be managed from a central point of control.



For flash, numerous assumptions should be questioned. For example, what is the cloud workload and how does it affect the write life of the device and could this lead to greater capacities being exposed? Similar to disk, questions should be asked regarding the amount and nature of the CRC and the logical interface as well as the best form factor. Better understanding and tailoring of lower power nodes along with the need for refresh should be understood and tailored to meet cloud providers' needs.


Regarding the use of tape technology for the cloud, several questions arise, such as what is the best interface into the tape system. Given that tape management software takes many years to write and perfect, a higher level interface, such as an object level REST interface might be more appropriate for providers that are unwilling to make that software investment. When cloud providers have made that investment, the physical interface to the tape system needs to match their other networking equipment (i.e., Ethernet). Due to the fact that tape has tighter temperature and humidity specifications than other storage technologies, solutions that minimize the impact of this requirement to the cloud provider should be considered. Additionally, there are features provided by tape drives that are not needed, such as backward read compatibility, as systems stay in place until their contents are migrated into a new system. If tape capacities or time to market can be accelerated by dropping backward compatibility, it should be seriously considered.

Cloud providers have a unique opportunity to adopt new storage technologies, based on the sheer size of their storage needs and small number of localities, ahead of volume commercialization of these technologies. For example, consider an optical technology whereby the lasers are costly, bulky and prone to misalignment, and the system is sensitive to vibration. If the technology provides enough benefit to a cloud provider, it might be able to install the lasers on a large vibration-isolating table with personnel assigned to keep systems operational and in alignment. In such a scenario, an automated device might move the optical media in and out of the system. In a similar scenario where the media has to be written in this manner but can be read with a much smaller and less costly device, the media may be, upon completion of the writing process, moved to an automated robotics system that could aid in any future reads to be done.

## **Cloud Versus On-Premises Perpetual Storage**

Years ago, the Gartner Research group defined the hype cycle model. It outlines the phases that a new technology works through as its being accepted. Quite simply, a technology moves from a hype phase to a disillusionment phase and finally to a productivity phase. Only a few years ago the talk was that customers would move entirely to the cloud. They would eliminate their IT staffs and have cloud expenditures that were lower than running internal operations. Many customers tried this. In the end, they found that their expectations were not aligned with reality, resulting in disillusionment.

More recently the talk has been, even from the cloud providers, about hybrid systems -- systems that can take advantage of cloud processing capabilities when they make sense and on-premises processing capabilities when they make sense. Referring to the architecture section of this paper, the two tiers of storage are defined as the Project Tier and the Perpetual Tier. Project storage will always be resident where the data is being processed, either in the cloud or on-premises. However, with the advent of a new generation of storage solutions, the




customer will now have a choice, regardless of where the Project Tier is located, as to whether the Perpetual Tier should be located in the cloud or on-premises. The remainder of this section is intended to provide insights into what should be considered when deciding the locality of both the Project and Perpetual Tiers.

A common workflow in today's world consists of three steps: 1) ingestion of raw data; 2) manipulation of the raw data to achieve a result; and 3) storage of the raw data (and any results) forever. This simple workflow is utilized in many Industries, including media and entertainment (M&E), medical research and the Internet of Things (IoT), to name just a few. In M&E, the raw data consists of film footage and other artifacts such as special effects that comprise a project. The processing time for the film footage could be several months long and is referred to as the post-production phase of a project. The output of this phase can be daily artifacts along with the end product known as the final cut. Once the project is complete, the raw film footage, daily artifacts and all versions of the final cut can be moved to a more suitable archival medium for long-term safekeeping.

In medical research, data containing the DNA of patients is collected and an initial processing step separates candidates that are worth further study versus those who are deemed not currently applicable. For the latter, those can be archived for possible future use in other studies. A more recent example of this workflow is IoT for the automotive industry. Consider an auto manufacturer whose next-generation automobile continually sends data to 5G hot spots that collect and analyze that data. This analysis might involve separating information into data that is relevant for improving self-driving programs, data that is associated with automotive failures, and normal telemetric data that just needs to be archived. The telemetric data may be kept forever for the sole purpose of protecting the company against liability. These are just a couple examples of the workflow discussed previously whereby the processing takes place in the Project Tier and the long-term archival in the Perpetual Tier.

The first decision a customer needs to make is to determine where to perform the processing -- either in the cloud or on-premises. There are many factors that need to be weighed in making this decision, such as the total cost of ownership, the versatility each provides the business, and the business preference toward capital or operating expenses. Besides these, there are more specific questions to ask. Do my applications run all the time or do they run infrequently? Do I want to license the applications or would a pay-as-you-go model be preferable? Do one or more of my applications require specialized hardware? For example, AWS has a very robust set of M&E services that can be utilized for processing video streams. The charge for these services is based on the quantity of data processed -- not on any software licensing fees. For small M&E shops with smaller quantities of video data to be processed, this choice is quite compelling. For bigger shops that are continually processing video data, this choice may or may not be cost prohibitive when compared to licensing the software and running it on-premises. Specialized hardware sometimes is the determining factor as to whether a customer performs processing in the cloud or on-premises. For example, both Google and IBM have developed specialized hardware that is not available in the open market for performing Artificial Intelligence. Customers who want to take advantage of the capability of this hardware have no choice but to run those processes in the cloud.

Once the decision has been made to process in the cloud or on-premises or some combination of the two, the next decision the customer needs to make is where to locate the Perpetual Tier -- in the cloud or on-premises. Running processes in the cloud requires the project data to be in an online storage pool of the respective cloud provider.



As mentioned previously it needs to exist in that tier for as long as the processing of the data is performed. For an M&E project, for example, it is for the duration of the post-production phase, and for automotive it may only need to exist for the few seconds it takes to process the incoming data stream. The customer will incur a storage fee based on the amount of storage consumed and the length of time that storage is held. When the project is completed, the raw content and the resulting artifacts can easily be migrated to a lower cost cloud storage tier. The customer will then be charged a fee for retaining that data and additional fees if they need to restore it for processing. Customers may also decide to run processes on-premises, while utilizing a low-cost cloud tier of storage as a repository for raw data and project artifacts. In this scenario, the customer would assume the same long-term fees just described.

The ideal scenario might be for customers to have the option of running the Project Tier on-premises or in the cloud, while ensuring the Perpetual storage system is on-premises. This would require a next-generation storage system. Consider a future on-premises storage system whereby all the raw data is sent to it instead of the cloud and, upon receiving that data, would perform two actions. It first would “sync” the data to the cloud, in order for cloud processing to occur on that data, and secondarily, it would make an archive copy of that data to either on-premises disk or tape. Additionally, the system could be programmed to automatically delete the data in the cloud after a preset period of time or the customer could manually delete the data when processing was complete. Further, when cloud processing created data artifacts, those could be “synced” back to the on-premises storage system for archiving.

One of the high cost components of cloud storage is in the downloading of data to an on-premises location, known as data out charges. For the solution described above, there would be no costs associated with uploading the raw data to the cloud -- and the artifacts created by cloud processing, which are typically a very small percentage of the size of the raw content and therefore very small data out charges, would be incurred. If such a solution would become available, the customer could do a head-to-head comparison of a cloud versus on-premises Perpetual Tier solution.

When analyzing the advantages and disadvantage of a cloud or on-premises Perpetual Tier solution, there are several things customers should ask themselves, such as 1) How much data will be stored; 2) How long will the data need to exist; 3) How frequently and how much of the data will need to be restored; 4) How quickly will data need to be restored; 5) How committed is my organization long-term to a particular cloud vendor; and 6) Do we have the required facilities and staff to maintain an on-premises solution. Below is a table showing the prices for the lowest cost tiers of storage in the cheapest regions from the three market-leading cloud vendors. Also included is the base characteristics for an on-premises solution whose pricing model is the traditional capital expenditure upfront with an annual service charge. Below the chart a description of each category is given.

Offering (North America)	Storage Price (\$TB/Month)	Restore Price (\$TB)	Restore Time	Data Egress Price (\$TB)	Minimum Store Duration (days)	Operations Price (\$/10,000)	Escape (\$/PB)
AWS S3-Glacier Deep Archive	\$.99	Standard - \$20 Bulk - \$2.50	Standard- 12 hours Bulk – 48 hours	\$50-\$90	180	PUT - \$.50 GET (Standard) - \$1.0 GET (Bulk) - \$.25	\$52,500
Azure Archive LRS	\$.99	High Priority - \$100 Standard - \$20	High Priority– 1 hour Standard - 15 hours	\$40-\$85	180	PUT-\$.10 GET (High Priority) \$50.0 Get (Standard) - \$5.0	\$60,000
Google Archive	\$1.20	\$50.00	Sub-second	\$80-\$120	365	PUT - \$.50 GET - \$.50	\$130,000
On-Premise Perpetual Tier Storage Solution	Initial Price \$50K+	None	Tape – Minutes Disk - Seconds	None	None	None	None

- Storage Price** – An advantage to cloud storage is that there are no upfront costs. Instead customers are charged a monthly fee for storing data objects based on the capacity they utilize. For our comparison group, the one outlier is the on-premises Perpetual Tier solution that requires customers to purchase capital equipment upfront, but then has minimal ongoing costs. Given that the cloud vendors have lowered these prices drastically we believe that they will not further erode as the vendors deploy cheaper storage technologies in the future.
- Restore Price** – This is the price, per TB, that is billed to a customer’s cloud account when data is to be restored back to an online cloud pool. For AWS and Azure there are two priority levels that can be issued - each with a different cost and performance. Note that though Google Archive storage is online, there is still a fee charged for any data accessed from this pool.
- Restore Time** – Shown are the approximate times that will lapse between the initial restore request and the ability to view the restored objects. For AWS and Azure these times are dependent on the rehydration priority. For Google packets, it can start being transferred immediately. For an on-premises storage system, it is dependent on whether that system is utilizing disk or tape.
- Data Egress Price** – This fee only applies if the data is going to be brought back into an on-premises location to be processed. In the case of an on-premises storage solution backing up to the cloud, this fee would apply to any restored data back to the on-premises location.



- **Minimum Store Duration** – Cloud vendors require that objects that are put into these storage tiers remain there for a minimum amount of time. If objects are deleted prior to this time, the customer is still charged for the storage that object would have consumed -- up to the minimum duration time.
- **Operations Price** – Cloud providers charge for requests sent to their services. These fees can be substantial if the customer is dealing with millions of objects at a time. For example, a million GET requests from the Google archive storage tier would be billed at \$100.
- **Escape Price** – This is the approximate cost, per petabyte, billed to the customer if they decide to read all their data out of the cloud repository. For instance, if the customer wants to switch cloud vendors.

Customers should consider all these costs before deciding on which solutions best meet their particular needs. As an example of projects not understanding these costs, consider the NASA situation whereby \$20 million per year of egress charges were not taken into account in their contract with AWS.

[\(https://www.theregister.com/2020/03/19/nasa\\_cloud\\_data\\_migration\\_mess/\)](https://www.theregister.com/2020/03/19/nasa_cloud_data_migration_mess/)

One other possibility is that for customers who require two geographically separated copies of their data, a configuration whereby one copy is kept on an on-premises storage system and another copy kept in a cloud repository, could be quite cost effective -- using the system to direct all restore events to the on-premises storage system and only using the cloud repository as a means of defense.

## CO<sub>2</sub> Emissions of Information Technology Systems

Storage systems, like all electrical devices, add to worldwide CO<sub>2</sub> emissions generated annually. Information technology (IT) systems consist primarily of networking, processing and storage. As a rule of thumb, storage systems make up about 20% of the overall power used by these systems. Over the previous decade, though, demand for IT increased six-fold and IT power consumption remained relatively flat at around 200 terawatt hours (TWH), thereby putting storage consumption at roughly 40 TWH annually.

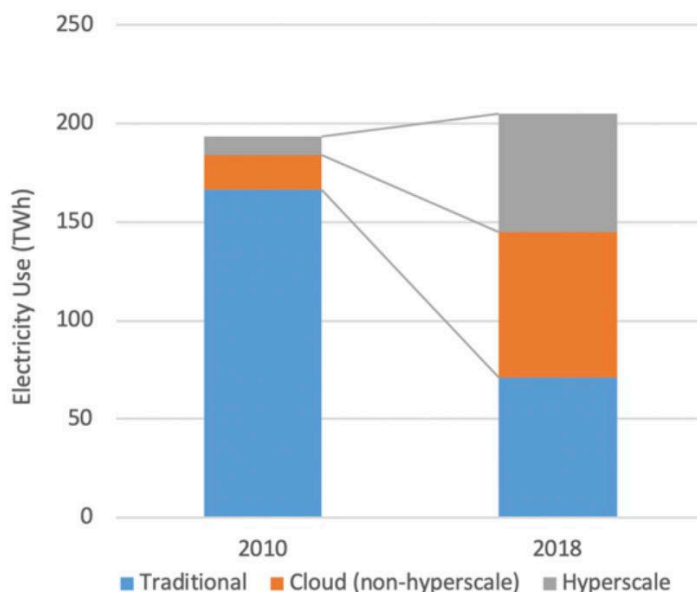



Figure 13: Estimated global data electricity use by data center type, 2010 and 2018. Source, Masanet et al. 2020




During the previous decade, there was a large shift away from traditional data centers to either cloud or hyperscale data centers. This consolidation did not naturally provide the efficiencies required to limit the growth of electrical usage given the increase in demand for services. It did provide, however, the level of scale required to fund technology efforts to reduce power consumption and provide an ROI (return-on-investment) on that funding. When considering an assumption of at least six-fold in demand over the next decade, the question is ‘Can electrical usage continue to remain flat and, more importantly, can CO<sub>2</sub> emissions be reduced over this time period?’ In order to determine the viability of this, we consider the factors that led to flat usage in the previous decade. We’ll also consider whether their impact will be as great going into the future.

**Power Usage Effectiveness (PUE)** – PUE is defined as the ratio of power coming into a data center to the amount of power that is used by the IT equipment itself (networking, processing and storage). For most IT organizations, who are usually not focused on this issue, this ratio is typically 2x or greater. In other words, at least half the power coming into these facilities is consumed by electrical conversion and cooling. Over the previous decade, cloud companies focused on reducing this ratio and announced new facilities with PEUs as low as 1.1. They have achieved this through making more efficient electrical conversions through technologies like those defined in the open compute project (<https://www.opencompute.org/>). They are also minimizing the power that is used to cool the equipment using creative water and evaporate cooling solutions. Also, by allowing the temperature and humidity within their data centers to fluctuate, they can decrease their cooling costs with the understanding that the failure rate of components will increase. But the power savings derived from not needing to maintain a more steady-state environment outweighs the lesser impact of component failure rates. Besides optimizing the use of electricity, cloud companies focus on reducing CO<sub>2</sub> emissions by using cleaner power in the way of wind, solar and hydro. Over the next decade, new hyperscale environments should consider these technologies and implement the ones that make sense for their environments. Traditional data centers do not typically have the sufficient scale to deploy these technologies.

**Server Energy Intensity** – This power intensity of a processor and its surrounding server is defined as amount of work a processor can perform per unit of electrical measurement as measured as watthours/computation. There were great strides over the last decade in increasing processor performance at a faster rate than energy consumption. Also, processors have become smarter in conserving power based on the workload they are presented. So, for instance, they will turn on and off processor cores as needed. These gains allowed the consumption of electrical energy for processing to be lower than that of the start of the decade. Unfortunately, most of the major gains have been made and additionally more modern workloads such as artificial intelligence, bitcoin mining, gaming and high-performance computing require higher energy-consuming graphical processing units (GPU).

**Server Count Per Workload** – The number of servers per workload decreased over the previous decade mainly due to virtualization technologies that enabled a server to perform processing for many applications at one time. Virtualization was also important in being able to fully utilize newer more powerful families of processors. Without virtualization there would be almost no reason to deploy these new processors. The cloud companies have led in utilizing virtualization technologies in that customers run their cloud workloads on virtual instances that emulate server hardware configurations. The actual physical hardware that is utilized by the cloud provider to provide for this emulation is usually much more powerful than the virtual instance itself; hence, many virtual instances can be provided from a single set of physical hardware. This leads to much higher processor utilization. This also allows for quicker deployment of next- generation servers as more virtual machines can be emulated by



a new hardware set thereby slowing the rate of growth servers required. Hyperscale and traditional data centers utilize virtualization to some degree, but there are opportunities to expand that use over the next decade.

**Storage** – Storage energy usage is the ratio of the power consumed by a storage device as related to the amount of capacity of that device, measured in kilowatt hours/terabytes. This means that if a storage device were to double in capacity but have the same electrical demand, the power usage factor would be cut in half. This report takes a bottom-up view of the storage industry based on what is shipped not what is utilized. It is believed that overall usage of storage devices is under 50%. Like processors, storage devices require virtualization technologies in order that they can be fully utilized. Once again, the cloud leads the way by leveraging virtualization methods to achieve a high rate of storage utilization. Cloud also has the manpower to develop technologies to support higher capacity flash and disk storage devices through implementation of the zoned storage initiative that was discussed earlier.

So far, we have only discussed the power consumption of the new demand; however, the current demand will also have to be supported. Though some portion of this demand may be retired, most will be required to run for years to come. In order to lower the electrical consumption of the remaining workload, it will need to be migrated to newer more efficient technologies. Processing and storage virtualization is the key for allowing this conversion to occur without disruption to current workloads. Processing virtualization allows newer generations of servers to be deployed that have the capability to present many more equivalent virtual machines thereby allowing many older servers to be replaced by far fewer newer classes of servers. Storage virtualization allows for the migration of data from older lower capacity storage devices to newer higher capacity devices. Unlike processing, though, storage migration needs to be handled much more carefully as it may result in lowering the performance of existing applications. As stated earlier, as storage devices have achieved higher capacities from generation-to-generation, their performance characteristics have not scaled proportionally. So blindly moving data from say a 10TB magnetic disk drive to a 20TB magnetic disk drive will result in half the performance. A better answer may be in trying to separate which existing data is “active” and which data is “cold” and then migrating that data to the appropriate storage medium. For old data that is active, it might be a good opportunity to move that data to flash technology. For example, databases that are currently running on magnetic disk will achieve a performance gain when migrated to flash technology. For older data sets, that are infrequently accessed, but require sub-second response when accessed, the data should be migrated to high-capacity enterprise magnetic disk drives. Finally, for infrequently accessed data sets, whereby long response times are acceptable, tape should be considered. Some customers are migrating to the cloud which then puts all future migrations in the hands of the cloud provider. On the positive side, this elevates the burden of migration from the customer; however, this is at the expense of not being able to take advantage of cost savings provided by future technologies. Cloud companies are continuously migrating processing and storage workloads over to new technologies without disrupting customer’s current operations. They, however, do not pass on any savings derived from these migrations. Looking backward over the previous decade, cloud processing and storage costs have declined at a much slower rate than the underlying technology has improved. In fact, the price to read data from cloud storage repositories has remained flat. Customer’s doing feasibility analysis of moving operations to the cloud should not bake into their analysis any assumptions of price reductions due to technology advances.

Whether in the cloud or on-premises, CO<sub>2</sub> emissions of storage systems are highly dependent on getting the right data into the right tier. As the chart above shows the greatest impact to CO<sub>2</sub> emissions is the “global installed storage capacity”. Though calculating the CO<sub>2</sub> emissions is complex with many variables, flash technology in general has the largest emissions, followed by magnetic disk, followed by tape with very low emissions.

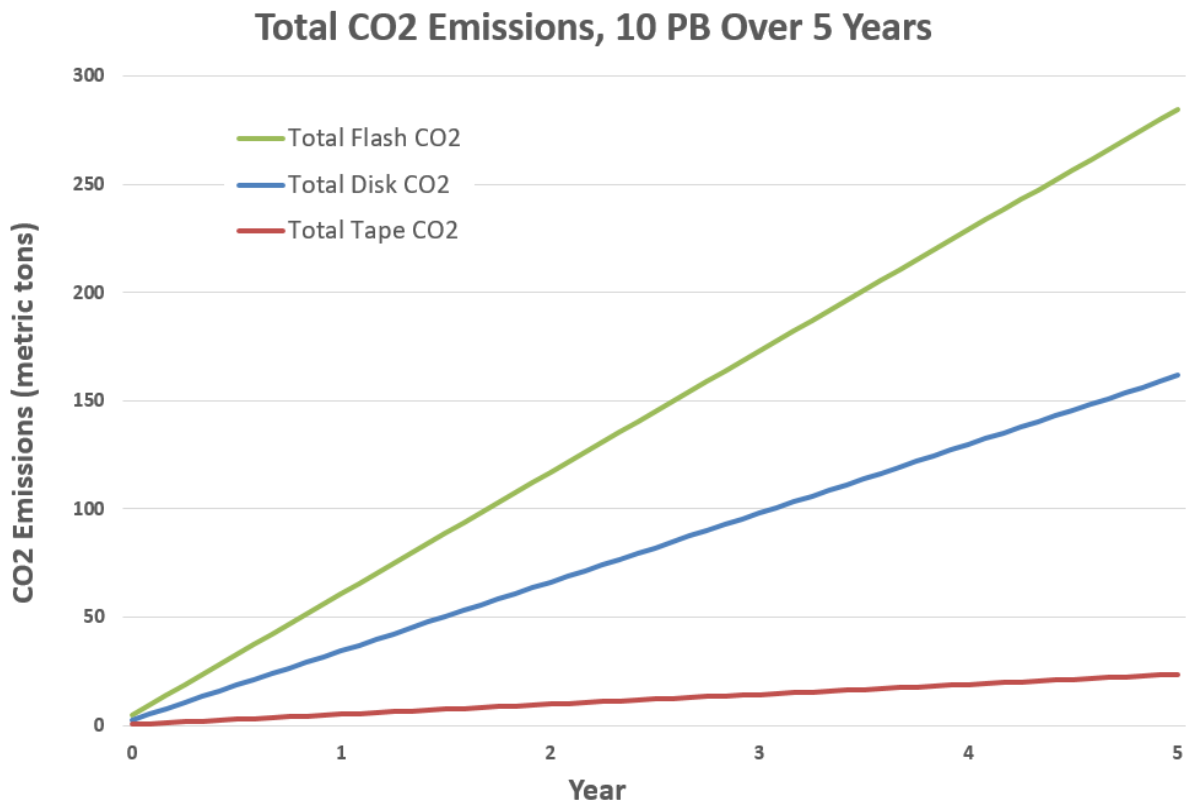


Figure 14: CO<sub>2</sub> Emissions of Different Storage Mediums

This section provided a high-level view of a very complex subject. For readers who would like a much more in-depth understanding consider watching the excellent video: <https://www.youtube.com/watch?v=-o8j5zIM0iA>



## THE DIGITAL UNIVERSE

The IDC report, published in November 2018 and commissioned by Seagate<sup>4</sup>, predicts the ‘global datasphere’ will grow by more than 175 zettabytes (ZB) by 2025. This causes many in the industry to wonder whether there will be sufficient media to contain such huge amounts of data.

The Internet of Things (IoT), new devices, new technologies, population growth, and the spread of the digital revolution to a growing middle class all support the idea of explosive, exponential data growth. Yes, 175ZB (or 175,000 Exabytes) seem aggressive, but not impossible. The IDC report took a top-down appraisal of the creation of all digital content. Yet, much of this data is never stored or is retained for only a brief time.

For example, the creation of a proposal or slide show will usually generate dozens of revisions -- some checked in to versioning software and some scattered on local disk. Including auto-saved drafts, a copy on the email server, and copies on client machines, there might easily be 100 times the original data which will eventually be archived. A larger project will create even more easily discarded data. Photos or video clips not chosen can be discarded or relegated to the least expensive storage. In addition, data stored for longer retention is frequently compressed, further reducing the amount of storage.

In short, though there might indeed be upwards of 175ZB, when a supply and demand mismatch is encountered, and there are many opportunities to synchronize:

- A substantial part of the data created will be by nature transitory, requiring little or very short retention.
- Storage costs will influence retention and naturally sort valuable data from expendable.
- Long-term storage can be driven to lower-cost tiers. Cost will be a big factor in determining what can be held online for immediate access.
- Flash, magnetic disk, and magnetic tape storage is rewritable, and most storage applications take advantage of this. As an example, when using tape for backups, new backups can be recorded over old versions up to 250 times, essentially recycling the storage media.
- The “long-tail” model will continue to favor current storage – as larger capacity devices are brought online, the cost of storing last year’s data becomes less significant. For most companies, all their data from 10 years ago would fit on a single tape today.

Spectra's analysis also differs from the larger projections by omitting certain forms of digital storage such as pre-mastered DVD and Blue-Ray disk and all flash outside of that used in solid-state disks.

### Total Digital Universe

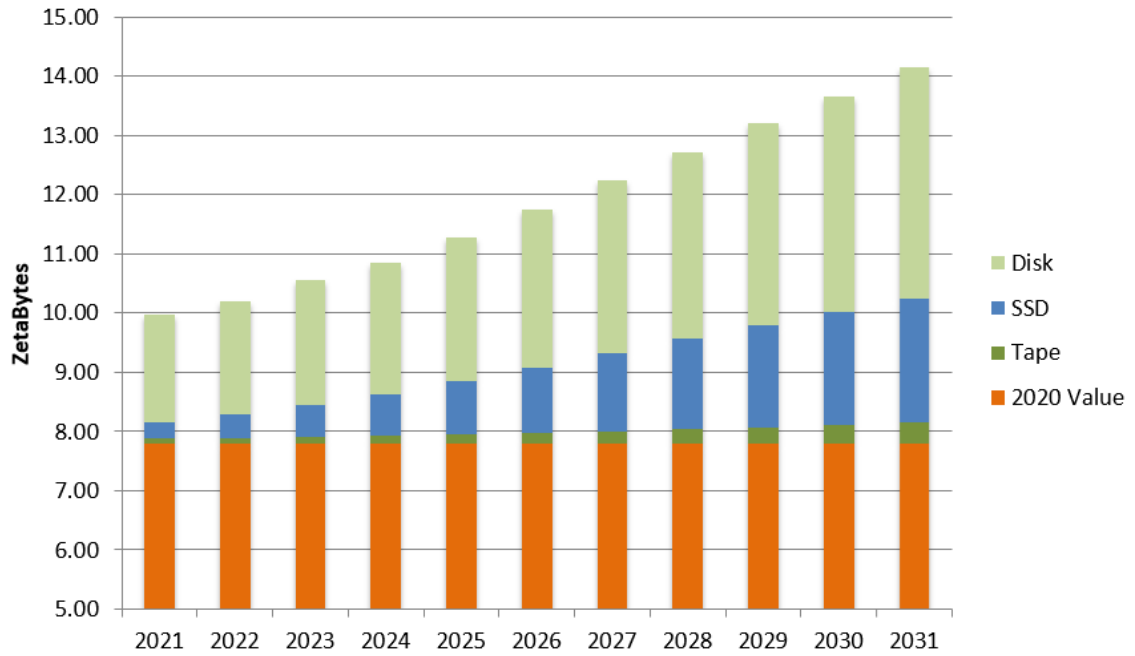


Figure 15: Total Digital Universe



## CONCLUSION

We agree with Thomas Edison's quote at the beginning of the report that states, *'If we did all the things we are capable of, we would literally astound ourselves.'* As the world economy emerges from the confines of the pandemic, the data that drives its awakening is leading the way. Whether it's ensuring the use, access, sharing and safeguarding of irreplaceable medical research, or the crunching and protecting of vast quantities of seismic computations, or the digitizing of the world's vast treasury of invaluable assets, industry leaders and IT professionals are pushing the boundaries of their capabilities to advance technology that will benefit the world.

For the foreseeable future, the storage growth requirements of customers will be fulfilled by storage device providers who continue to innovate with higher performance and higher capacities to meet increasing demand. As noted in the report, every storage category is exhibiting technology improvements. First, we see memory-hosted 3D XPoint technology becoming the latest high-performance standard for database storage. At the flash layer, 3D fabrication technology is allowing for the creation of density parts while lowering the cost per gigabyte. In the meantime, disk manufacturers are closing in on delivery of HAMR and MAMR technologies that will allow them to initially deliver disk drives of 20TB while also enabling a technology roadmap that could achieve 50TB or greater over the next 10 years. Finally, tape has enough technology headroom that it will achieve storage capacities of 100TB or higher on a single cartridge in the next decade.


### Data Storage Dilemma

Given that a singular storage technology has yet to be invented that combines the highest performance at the lowest cost, customers will continue to face the dilemma of what data, at what time, should be stored on which medium. Data that supports a project in progress one day may be suitable for archive once that project is completed. This would thereby lower overall storage costs by freeing up storage capacities for future projects. Software tools that allow customers to identify the usage patterns of their data and then provide for the movement of infrequently accessed data to lower tiers of storage have been available for quite a while; however, these tools have been priced such that most of the benefit of the storage savings are lost. A new generation of tools is required that improves data storage efficiencies while mitigating storage costs.

### Designing with the Cloud in Mind

Over the last few years, a new question has arisen for storage administrators, which is 'where' to deploy 'what' storage. More specifically, what data should be placed in the cloud, what data should be located on premises, and what data should be stored in both locations. Each location provides benefits and cost trade-offs. The demand by storage customers to use cloud-based storage prompted many legacy storage providers to 'shoehorn' basic cloud capability into their existing products. Primarily this has consisted of providing customers with the capability of making cloud disaster recovery copies of their on-premises data. This is a pattern that has been seen before such as in the adoption of flash technology into disk arrays. The first generation of storage systems to utilize flash were existing products that were designed before flash storage was available. For this reason, it was typically integrated





into these systems as extended cache because that is where it could most easily fit into these existing architectures. Customers gained some benefit, but not the full scope of the technology. Second and third generation solutions were designed with flash in mind and provided tremendous capability to the customer. Over the last few years the solutions have become the hottest segment in the storage system business.

## Supporting Complex Workflows

We consider cloud integration by on-premises storage systems to be in this first phase. Next-phase products are being designed from the ground up with the cloud in mind. These products allow for seamless integration of applications into the storage infrastructure, regardless of storage location -- whether in the cloud, multiple clouds, and/or in multiple on-premises locations. Complex customer workflows can be supported, through policies set by the customer, that allow data to be automatically moved to the right location(s), to the right storage tiers, at the right time. With this capability, customers have the freedom to decide which processes they want to run locally and which ones in the cloud – all without having to think about the underlying storage system.

There are many interesting storage ideas being pursued in laboratory settings at different levels of commercialization: storing data in DNA, 3D Ram, (5 dimension optical) hologram storage – plus many that are not yet known. Technology always allows for a singular breakthrough, unimaginable by today's understanding, and this is not to discount that possibility.

## Planning for the Future

Spectra's projections do not call for shortages or rising media costs. Due to the ongoing impact of the coronavirus pandemic there could be short-term supply-side shortages; however, it is unclear at this point whether reduced demand will result in a balanced or unbalanced market. But there are credible risks against expectations of precipitously declining storage costs. Storage is neither free nor negligible and proper designs going forward need to plan for growth and apportion it across different media types, both for safety and economy. Corporations, government entities, cloud providers, research institutions, and curators must continue to plan for data management and preservation today, evaluating data growth against projected costs.

## CONTACT US

Spectra has stepped out for its seventh year to make predictions on the data storage industry's future based on what we see today. Think these predictions are too high? Too low? Missing something important? Spectra plans to update and publish this document with new data and new thinking where needed. Please let us know your thoughts.

### Spectra Logic:

[www.spectrallogic.com](http://www.spectrallogic.com)

To obtain permission to use or copy this outlook or any of its contents, please submit your request in writing to [marcom@spectrallogic.com](mailto:marcom@spectrallogic.com).

## APPENDIX NOTES

### Footnotes:

- <sup>1</sup>Source:  
<https://www.idc.com/getdoc.jsp?containerId=prUS48159121#:~:text=NEEDHAM%2C%20Mass.%2C%20August%2012,in%20the%20previous%2012%20months>
- <sup>2</sup>Source: <https://www.businesswire.com/news/home/20210324005175/en/Data-Creation-and-Replication-Will-Grow-at-a-Faster-Rate-Than-Installed-Storage-Capacity-According-to-the-IDC-Global-DataSphere-and-StorageSphere-Forecasts>
- <sup>3</sup>Source: <https://www.gartner.com/en/newsroom/press-releases/2022-01-18-gartner-forecasts-worldwide-it-spending-to-grow-five-point-1-percent-in-2022>
- <sup>4</sup>Source: IDC Data Age 2025: the Digitization of the World from Edge to Core, November 2018  
<https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>  
(page 39)

### Charts:

**Figure 4:** Page 19, Source: ASTC <https://hexus.net/tech/news/storage/123953-seagates-hdd-roadmap-teases-100tb-drives-2025/>

**Figure 7:** Page 25, Source: The [LTO Program](#). The LTO Ultrium roadmap is subject to change without notice and represents goals and objectives only. Linear Tape Open, LTO, the LTO logo, Ultrium and the Ultrium logo are registered trademarks of Hewlett Packard Enterprises, International Business Machines Corporation and Quantum Corporation in the U.S. and other countries. Note: Compressed capacity for generation 5 assumes 2:1 compression. Compressed capacities for generations 6-12 assume 2.5:1 compression (achieved with larger compression history buffer.)

**Figure 10:** Page 29, Source: Sony and Panasonic <https://hexus.net/tech/news/storage/67165-sony-panasonic-create-archival-disc-standard/>

**Figure 11:** Page 30, Source: Sony and Panasonic [https://www.snia.org/sites/default/orig/DSI2015/presentations/ColdStorage/Yasumori\\_Archival\\_Disc\\_Technology-2.pdf](https://www.snia.org/sites/default/orig/DSI2015/presentations/ColdStorage/Yasumori_Archival_Disc_Technology-2.pdf) (slide 23)

*This preliminary calculation is based on objective data and Sony offers no guarantee that media are capable of storing data for 100 years irrespective of the environment*

**Figure 13:** Page 37, Source:  
Masanet, E., Shehabi, A., Lei, N., Smith, S., & Koomey, J. (2020). Recalibrating global data center energy-use estimates. *Science*, 367(6481), 984-986.

**All unsourced charts in this report were created by Spectra Logic.**

## About Spectra Logic

Spectra Logic develops a full range of Attack-Hardened™ data management and data storage solutions for a multi-cloud world. Dedicated solely to data storage innovation for more than 40 years, Spectra Logic helps organizations modernize their IT infrastructures and protect and preserve their data with a broad portfolio of solutions that enable them to manage, migrate, store and preserve business data long-term, along with features to make them ransomware resilient, whether on-premises, in a single cloud, across multiple clouds, or in all locations at once.

To learn more, visit [www.SpectraLogic.com](http://www.SpectraLogic.com).