Data Storage

Outlook 2023

Explore how the world manages, accesses, uses, and preserves its ever-growing data repositories. Get insight into the strategies and technologies designed to protect the data being created now and in the future.

www.SpectraLogic.com

Data Storage Outlook 2023

Executive Summary

"In God we trust. All others must bring data."

-W. Edwards Deming

This is the eighth annual Data Storage Outlook report published by 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.

Roughly 40 years ago, W. Edwards Deming stated, "In God we trust. All others must bring data." While Deming used data to improve management and manufacturing processes, his now famous quote applies to virtually every aspect of society today. In 2022, the world population reached 8 billion people, and it is estimated that 5.3 billion of them are using the internet.[1]

As more data is being created, more interesting ways of using it are being introduced. Strides in Artificial Intelligence (AI) have made the news repeatedly over the last 12 months. OpenAI released ChatGPT in the latter part of 2022, a "large language model" app that creates text so human-like it is virtually impossible to distinguish between machine and human-written articles. Competitors such as Google (PaLM) and Meta Platforms (LLaMa) quickly followed. This is just one example of the growing field of AI.

All forms of AI utilize machine learning algorithms, and those algorithms require enormous amounts of data to "learn." As Spectra is always careful to point out, not all data created is data stored. But as AI makes the creation and analytics of information more available to all sectors of business and individuals, we see this as another indication that the value of data continues to increase, and more data will be retained for longer periods of time.

As "all others" continue to bring their data, we feel storage vendors will continue to develop new methods and technologies to store, manage, use, and preserve the digital universe. *In data we trust.*

The Storage Gap

- Again, this year, 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. Throughout 2020 and 2021, the storage industry, like all other industries that are dependent on electronic components, saw supplies become limited, resulting in long lead times and price increases at the component level. Yet 2022 showed great normalization of earlier Covid-related supply chain issues. Flash has actually been overprovisioned by manufacturers, demand in the disk market is significantly down, and LTO-9 tape holds 18TB on a single cartridge, with LTO-10 coming next. The IBM TS1170 tape drive has just been announced to hold 50TB on a single cartridge. This is a significant development in tape technology. In the short term, we actually see a greater abundance in storage supply than in many of the previous years we've published this report.

SPECTRA

Storage Apportionment and Tiering

- Economics will continue to push infrequently accessed data onto lower-cost media tiers.
- 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. One ominous trend: If magnetic disk is unable to more rapidly improve its capacity/cost per TB over the next few years (as it has for the last few), it will be further compressed by flash for performance and tape for capacity.

2023 Report Highlights

- 2022 saw Intel cancel its persistent memory products based on 3D XPoint technology.
- The flash market entered a phase of oversupply of product, resulting in a sharp decline in prices. This oversupply resulted from reduced shipments of PCs and laptops along with less volume being purchased by cloud providers.
- Seagate, Western Digital, and Toshiba have all seen substantial reductions in magnetic disk volumes. In the consumer sector, price reductions of flash storage have obliterated the demand for 2.5-inch disk drives. In the high-capacity enterprise drive category, reduced purchases from the cloud providers have resulted in not only a reduction of drive shipments but, for the first time in history, a reduction in the total amount of storage capacity shipped. From a technology perspective, all three vendors claim to be well-positioned to start shipping 24TB+ enterprise drives during 2023 (up from 22TB in FY22).
- As the S3 interface from Amazon has become a de facto standard for object storage, there is an increasing adoption of the S3 interface among applications as well as storage devices. Both object disk and object tape can now be used for tiering data in the cloud or on-premises, allowing users to create "on-prem" Glacier storage. Three vendors, Spectra Logic, Quantum, and Point are now delivering products in this space, and IBM has announced its intention to. Likewise, more and more cloud companies are utilizing tape as a long-term storage layer.

Background

Spectra Logic develops a full range of data management and data storage solutions for hybrid cloud. 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.

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 exceed 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 writes 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 data, 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 does not 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; as long as the URI remains unchanged, users will neither know nor care.

The cloud grew out of the web, so it is no surprise that the cloud is primarily based on object storage. The first product of Amazon Web Services (AWS), pre-dating their compute offerings, was a storage service called Simple Storage Service (S3), released in 2006. The S3 protocol to interface to that service 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
Connection	Direct-attach or local network/VPN	VPN or public internet
Standardization	POSIX, Windows	Lacking; Amazon S3 popular
Read/Write Mix	Arbitrary read/write	Write-only/read-many
Data Mutability	Update any file in any place	Objects immutable; make new version
App compatibility	Broad	Limited; new applications only
Network/technology independent	No	Yes
Transparent storage class migration	No	Yes
Versioned, auditable	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 focuses on a Primary Tier where in-progress data resides, which is file-based, and a Secondary Tier where static or infrequently accessed data resides. The Secondary Tier holds data which is a combination of file-based and object-based data. Data moves seamlessly between the two tiers as data is manipulated, analyzed, shared and protected.

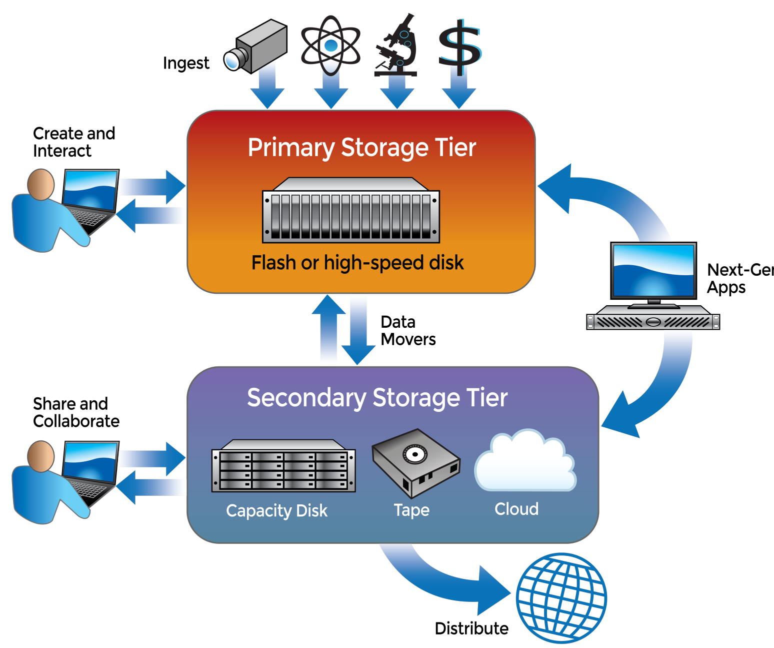

Figure 1:The multi-tiered storage model is more appropriately represented by a 2-tier system in today's modern, datacenter.

The Primary 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 timecodes simultaneously. Data will be raw, uncompressed, and require extremely high bandwidth. These streams may be stored on 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 this data will be discarded after the job is complete; only the results will live on. This 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 Secondary Tier.

The Secondary 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. Group projects, research data, experimentation output, and year-end financials would all fit into this category. There are too many examples to name. Users accessing this information may be spread across geographic regions. This is a perfect application for object storage, 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 in many cases, 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, for example, may result in a variety of distribution files, along with a descriptive manifest, like MPEG-Dash, 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 the sharing of sensitive data across the public internet. Storage costs 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 Primary and Secondary Tiers in both directions. Users may migrate data from the Primary Tier to the Secondary 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 Primary 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, including data stored as objects. They use a file system for their workspace and object storage (including the cloud) as the source and destination for more permanent data. Additionally, some software libraries support object storage natively; for example, there is an HDF5 library that can use an S3-compatible object store directly.

Data Movers

Until applications can natively utilize both the Primary and Secondary 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 Primary data over to the Secondary Tier once a project is completed. This serves two purposes: it frees up the Primary 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 Primary Tier of files not accessed for a long period of time. This frees up the Primary Tier storage such that expansion of that storage is not required. Another customer may use the Secondary Tier 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 Secondary Tier to a local Primary Tier. Once changes to the data are complete, the file(s) can be "checked in" back to the Secondary 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 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.

Dynamic Random Access Memory (DRAM)

As mentioned earlier, in a surprise move, Intel canceled its 3D XPoint™ persistent memory technology. Proceeding this by two years was a similar announcement from Micron, Intel's original partner in this technology. This technology filled a space in the storage hierarchy between DRAM and flash. Somewhat slower than DRAM but faster and with higher wearability than flash, it appeared to be a good fit for several application spaces. From a technological perspective it lived up to all its promises in terms of speed, wearability and reliability. Rumors have circulated that its production costs were higher than flash and neither Intel nor Micron could justify continuing production at a loss.

Micron's announcement is an indication that they will use the freed-up R&D resources to develop technologies associated with CXL (Computer eXpress Link). Over time, this emerging standard will revolutionize the relationship between compute and memory. To understand why this technology will eventually see mass adoption consider problems with the DRAM architecture of today:

- Memory is physically tied to a single core in a processor.
- There are numerous direct memory access (DMA) transfers that occur between CPU memory and peripheral memory tying up bus bandwidth and creating latency.
- The amount of total memory a system can contain is constrained by the size of the physical DIMM package and the number of DIMM channels provided by the processor.
- Sharing of memory across systems is not possible.

CXL is built on the PCIe physical and electrical interface and is first available with PCIe5. Industry experts more commonly believe that it will be widely adopted with the advent of PCIe6. There are currently two versions of the CXL standard, 1.1 and 2.0. The 1.1 standard addresses memory sharing within a single system while 2.0 introduces the pooling of memory resources across multiple systems. There are three types of devices supported in 1.1. The first type would typically be something like a smart NIC (network interface card) which would not contain any memory but rather be able to transfer to and from host CPU memory directly. The second type would be a graphical processing unit (GPU) which, for computational purposes, would contain its own memory; however, upon completion of processing, it could drop results into CPU memory directly. If desired the memory on the GPU could be accessed directly by the CPU. The third type consists of new memory devices that allow multiple-core processors to access a memory pool from any core. This allows for much larger capacity memory cards while lowering the need for additional DRAM slots and channels. For example, Samsung has introduced the first 512 GB DRAM card designed for CXL. Samsung has also made strategic announcements on the production of persistent memory modules that support the CXL standard.

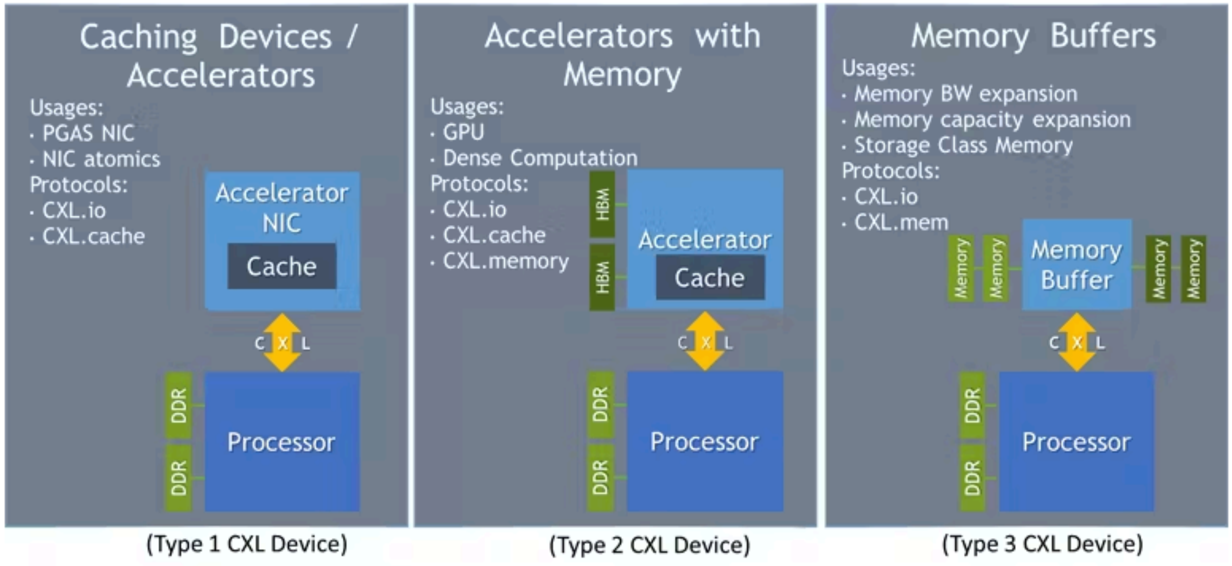

Figure 2: Three Types of CXL 1.2 devices

The CXL 2.0 specification defines CXL switching that in turn provides the benefit of multiple processors access to a common pool of memory resources. It provides pooling of those resources such that they can be sliced up in any manner desired. This pooling is dynamic in that memory allocated to one processor at one instance can be freed up upon completion of its processing and then reallocated to a different processor. A good example of this concept would be within a rack of servers a single JBOM chassis (just a bunch of memory) could exist filled with memory cards that all processors within all servers could utilize. Additionally, the JBOM chassis could contain type 1 (NICs) and type 2 (GPUs) if they are CXL-compatible devices.

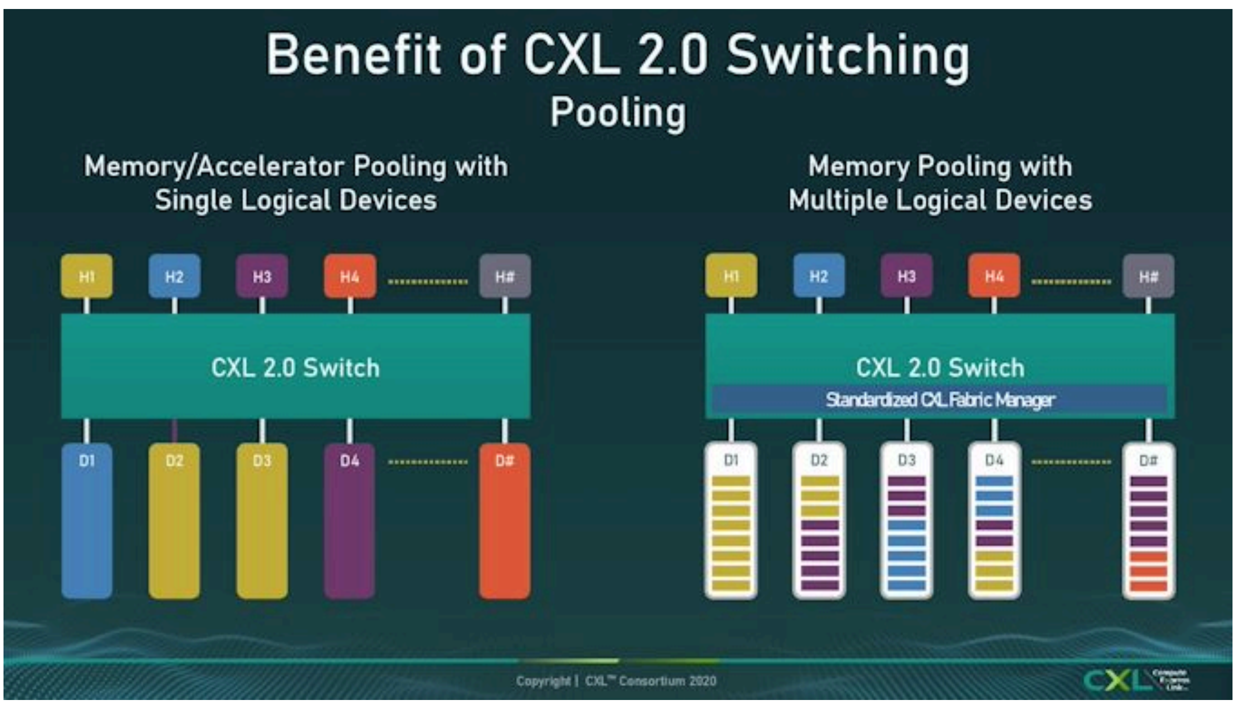

Figure 3: Memory pooling with single logical devices versus multiple logical devices

One longer-term effect of CXL technology will be in the reduction of the number of DDR channels provided by processors. With CXL technology the memory that needs to be dedicated to a processor is reduced to what needs to be pinned in place and never freed. For instance, each processor needs to run an OS (operating system) and should do so on higher-speed local memory than with slower-speed CXL shared memory. This may lead to processors being introduced with enough high-speed built-in memory such that DDR memory is no longer required.

Flash

Flash continues to be the fastest-growing technology in the storage market. 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, smartphones, tablets, laptops and desktop computers. As previously discussed, we predict that the Primary Tier will be comprised of solid-state disk storage technologies.

There are five companies that own and run NAND fabrication lines: Samsung, Kioxia/Western Digital, Micron, SK Hynix and Yangtze Memory Technologies. Several vendors delivered 176+ layer chips in 2022 using string stacking (see below). In addition to 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 a single-level cell (SLC). This was followed by the 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, this technology may only be suitable for applications that do not overwrite data, such as archiving. Participation in the archive market, directly against existing disk and tape solutions, will require an order of magnitude or more of cost reduction.

As discussed previously, all vendors have announced the availability of 176+-layer parts in 2022. Micron is shipping a 232-layer TLC part that is a "string-stack" of two 116-layer components. SK-Hynix is in production of a 236-layer part while Samsung is shipping 176-layer parts with plans to be in the 230+ layers in the near future. In theory, more layers provide for a reduction in size, cost and improvements in performance. 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 two 128-layer parts and combine them to create a 256-layer part. Both increasing the core number of layers and string-stacking techniques have their advantages and disadvantages. To do a single stack part of 256 layers takes more upgrade investment in the fabrication facility than adding string stacking to a facility already producing a 128-layer part. Also, the initial yield of the higher-level layer parts 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.

A flash competitor that warrants special attention is 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. Yangtze has been blacklisted by the U.S. government as a threat to national defense. For this reason, production equipment for chip manufacturing can no longer be sold by U.S. companies to Yangtze. It is unclear as to what effect this will have on their ability to produce higher-capacity chips in the future.

Flash SSDs had been constrained by the electrical interface -- either SAS or SATA. These interfaces added latency to data transfers that, with magnetic disk, were formerly "in the noise range" 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 commands to support very high-performance non-volatile memory (NVM) on top of a physical PCIe bus. This interface has now been fully adopted and is now standard for both enterprise and consumer technology. As a proof point, many next-generation SSDs are not offered in SATA or SAS versions. NVMe has a roadmap to the future as it is following a parallel path with PCIe. For example, NVMe3 indicates NVMe running over a PCIe Gen 3 bus while NVMe4 indicates NVMe running on a PCIe Gen 4 bus.

As 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 can read sequentially at greater than 7GB per second and write at greater than 5GB per second. They can also support more than 1 million random small block read and write operations per second (random I/Os). These drives currently sell for between $55 and $100 per terabyte. The next generation of drives will support NVMe5 (Gen 5 PCIe) and will be capable of reading and writing at 14 GB/s+. Servers and SSDs that support Gen 5 PCIe should start shipping later this year but will warrant a price premium over Gen 4 technology at a minimum through 2024. There has also been a convergence of consumer and enterprise drives. Only a few years ago, consumer drives utilized SATA interface and QLC flash. They had low write durability and even lower performance compared with enterprise drives that utilized the SAS interface and MLC (Multiple Level Cell) or TLC (Triple Level Cell) flash. Today that has all changed with new motherboards being able to accept m.2 NVMe SSDs.

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 costs are trivial compared to the flash "chip" cost. The disadvantage of large flash SSDs is that one controller needs to service a large capacity of flash chips and the performance becomes limited by the interface. A better approach is to maintain the ratio of one controller to a reasonable capacity of flash. To address this issue, new smaller form factors have been created. The most popular of these form factors is M.2 NVMe and new-generation motherboards which typically have two or more built-in M.2 NVMe slots. As previously stated, these parts can be thought of as enterprise-class rather than consumer-grade. For instance, at the time of this writing, an M.2 NVMe Gen 4 SSD that supports over 1 Petabyte of wear, along with a performance of 6 GB/s write and 7GB/s read, costs about $65 per TB. For enterprise systems that require U.2 form factor drives it may make sense to use M.2 to U.2 converter cards as there is a price premium placed on flash drives that come in the U.2 form factor.

NVMe is great for moving data from a processor to flash drives inside a chassis, but 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.

Currently enterprise storage servers support either SATA, SAS or U.2 SSD NVMe devices. As shown below each requires the server to have a unique backplane and in the case of NVMe a separate PCIe switch. This requires the storage server vendors to produce three versions of each server if they have customers who desire to use all three storage types.

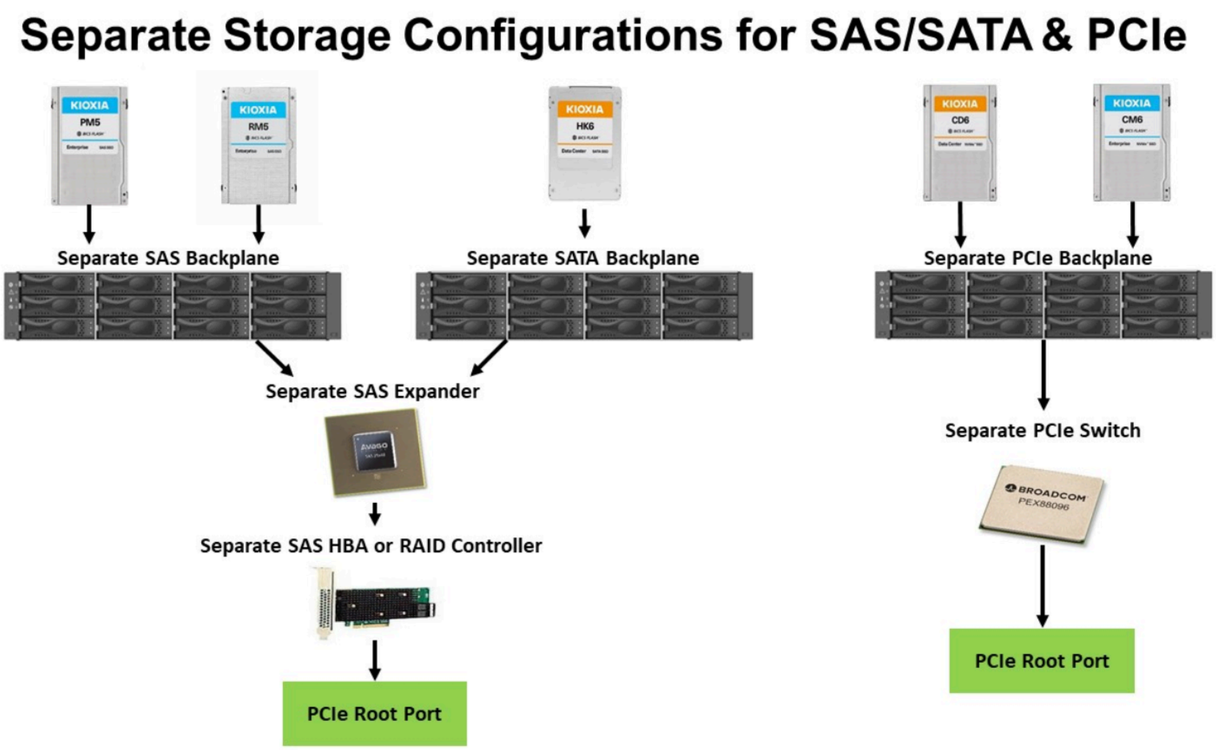

Figure 4: Server manufacturers are currently required to build 3 distinct interfaces - SAS, SATA and PCIe (as shown above). Manufacturers would prefer to standardize on a single backplane.

To address this, a new standard U.3 (pronounced U dot 3) has been created. With U.3 there is a single backplane that all SATA/SAS/NVMe U.3 storage devices can be plugged into. Pin assignments have been altered such that the backplane can identify which type of device has been plugged in. The advantage for the storage server producer is that a single tri-mode controller can be utilized which supports all three storage standards thereby reducing the variety of servers required to be created. On the downside, U.2 SSDs are not compatible with U.3 backplanes. U.3 SSDs are, however, backward compatible and will work in U.2 slots. There are also performance impacts when using the tri-mode controller that don't exist when using a standard PCIe switch. As we have also seen the convergence of consumer and enterprise SSDs and a movement away from SATA/SAS interfaces for these devices, we must question the value of U.3. In a 3.5-inch form factor, there may be some value to being able to populate some slots with magnetic disk drives and others with SSDs.

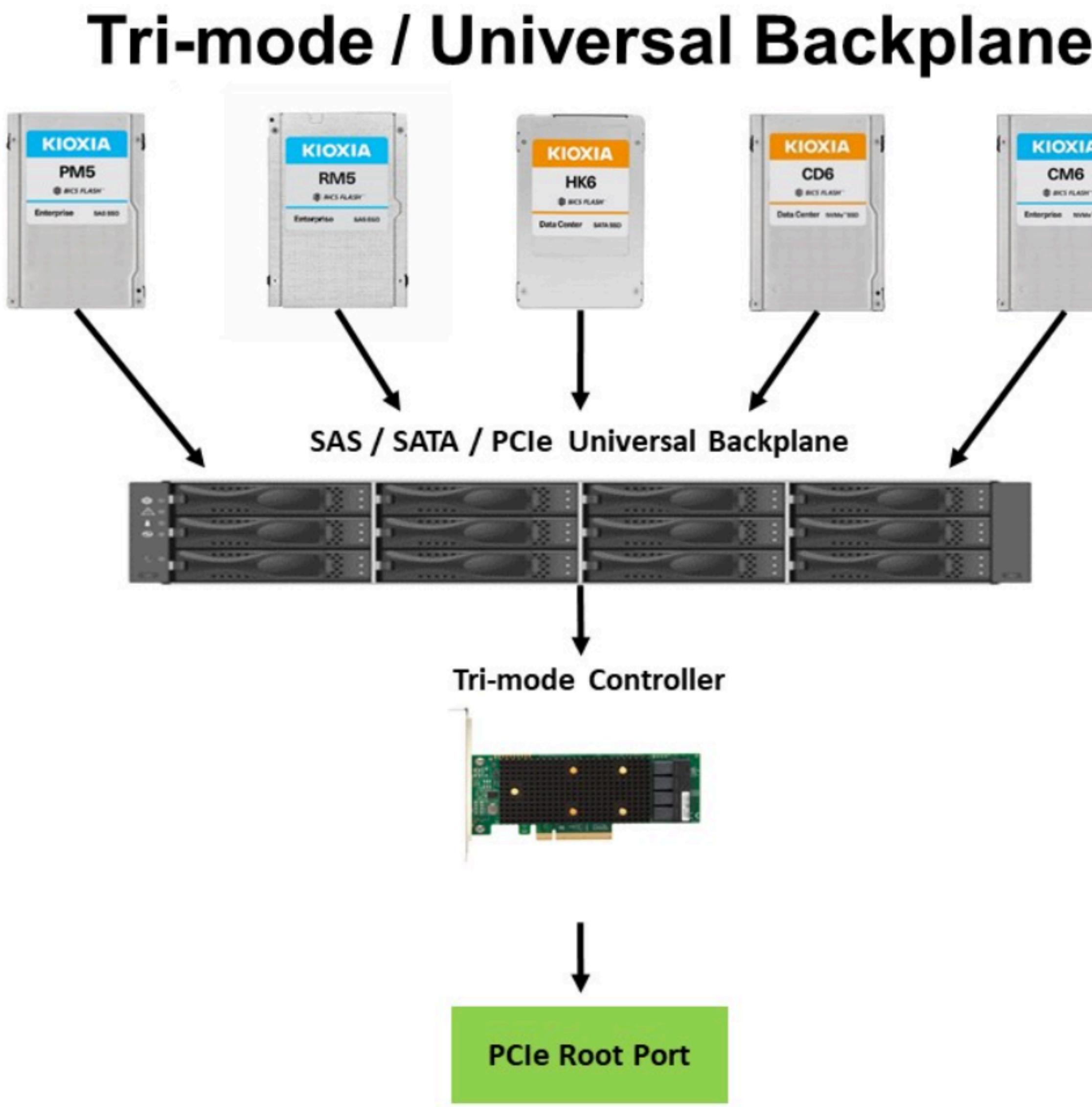

Figure 5: Unfortunately, the tri-mode controller is limited in performance compared to a PCIe controller

Along with the physical interface changes described above, work is being done on developing new logical interfaces. A Samsung initiative involves virtualization such that a single SAS 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-sized 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

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 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 to achieve their combined performance. 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 over time. 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, to support the workload, most 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 reutilize 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) 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.

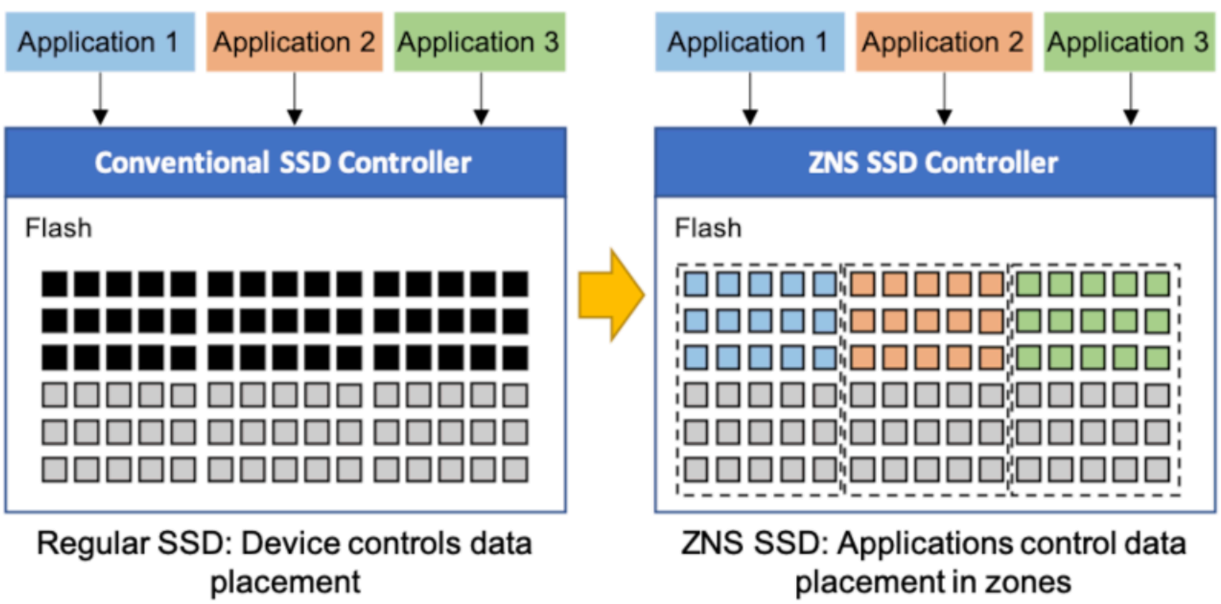

Figure 6: Conventional SSDs and ZNS SSDs internal data placement

Estimating the amount of capacity shipped of flash storage is challenging given the convergence of the enterprise SSDs and consumer SSDs so the distinction has been removed. During COVID the desktop and laptop business boomed and 2021 sales were over 50 million units. Given that over 90% of the units are now flash-based, there was great demand placed on flash production. For this reason, flash manufacturers ramped up production to meet this demand with the thought that desktop and laptop sales would continue at the same rate through 2022 which they did not. Sales fell to 9 million units in the fourth quarter of 2022 with a net loss of over 13 million units for the full year. This resulted in a glut of flash in the marketplace that resulted in substantial price declines. This glut will continue through at least half of 2023, and price increases are not forecasted for the rest of the year.

Magnetic Disk

For many years, the disk drive industry has had three major providers: Seagate, Western Digital and Toshiba. It is interesting that two of these 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 were substantially lower at 200 million compared to 255 million for the prior year's four quarters. This is the first year that the total storage "capacity" of disk drives shipped has declined over the previous year.

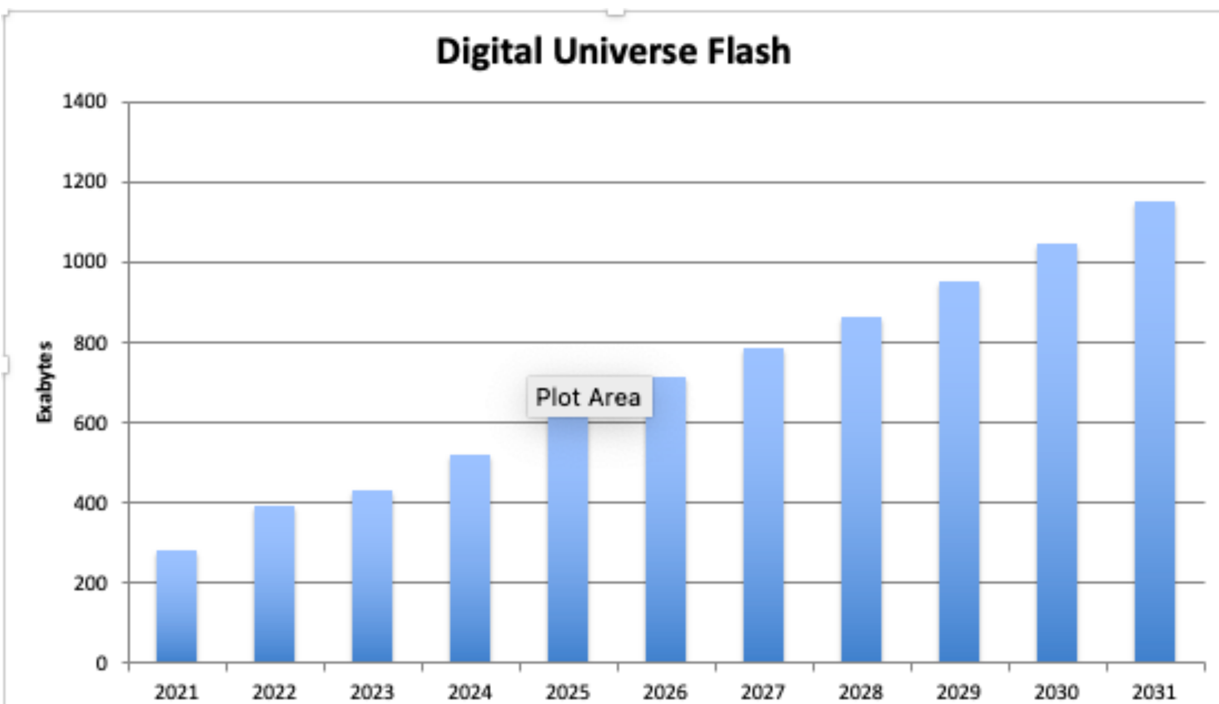

Figure 7: Digital Flash Universe

All consumer and 2.5-inch-high performance categories of disk drives 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-high capacity 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.

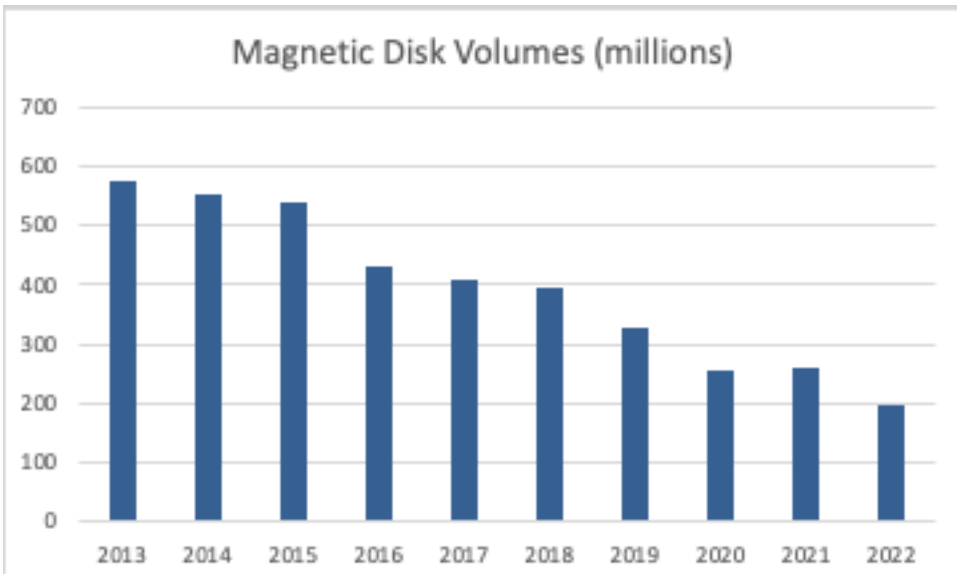

Figure 8: Magnetic Disk Volumes

Increasingly, the disk industry will be shipping a singular product line, that being high-capacity 3.5-inch nearline drives. These are predominantly sold 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. Variations of these drives will be in the areas of single or dual actuator, shingled or conventional recording and SAS or SATA interface. 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.

SPECTRA

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, was used to enable drives of 22TB CMR (Conventional Media Recording); full-blown MAMR will be used for drives starting in the 26TB CMR range in 2023; and HAMR will be used to achieve 50TB drives by 2026. They are also claiming that the Shingled Magnetic Recording (SMR) versions of these drives will have a minimum of 20% greater capacity and, as such, the 22TB drive will be 26TB+ in its SMR version. Western Digital will be shipping a 26TB CMR and 30TB Shingled Magnetic Recording (SMR) drives in 2023. 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). Seagate has announced that they have completed testing of their first generation HAMR drives and now plan on production of 30TB drives available in 2023. It is unclear if this capacity is HAMR or HAMR with SMR. Given that SMR requires Host-Managed aware software SMR drives are not sold in the open market and therefore capacities are not known. Promises on HAMR have been made and missed for many years – perhaps it is different this time?

Regarding larger capacity, future drives, there has been incomplete 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 16B 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 provides additional benefits

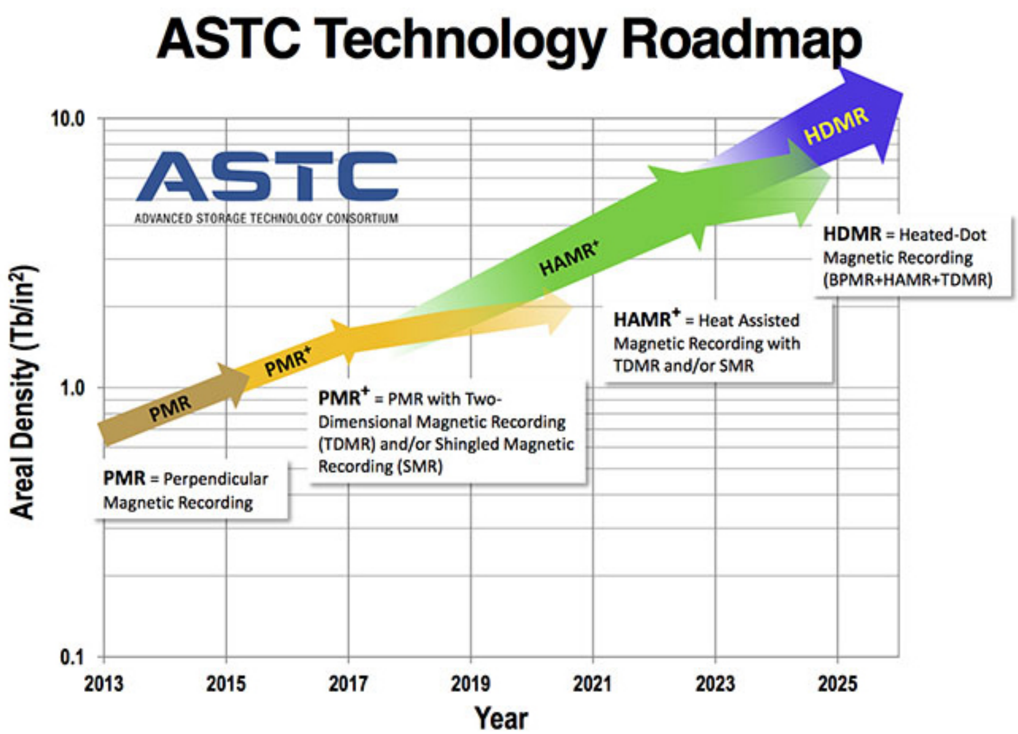

Figure 9: ASTC Technology Roadmap

in that it leads to requiring fewer racks, less power and smaller footprints -- all important considerations. A problem for these large capacity drives is 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.

Due to the delay of the HAMR technology, the Advanced Storage Technology Consortium (ASTC) significantly revised its roadmap. The new roadmap appears to be accurate thus far as it shows traditional, Perpendicular Magnetic Recording (PMR) technology being 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 a tenth platter and, therefore, create PMR drives of 22TB. Toshiba has announced that they will introduce a drive with 11 platters gaining another 10% in capacity. 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^{tm}, 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 affects the margin for reading the tracks around it. This is called adjacent-track interference (ATI). After some writes, modern disk drives will read the adjacent tracks and rewrite them before they become 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 ten 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 as 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 ensures 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.

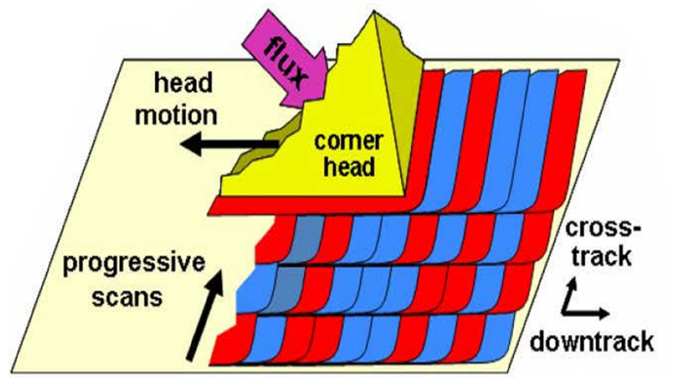

Figure 10: An example of how shingled-magnetic recording overlaps tracks

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 exception 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 matching the data's performance needs. The performance of a zone corresponds to where it exists on the physical disk. Faster zones are at the disk's outer diameter 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 chunks 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 primarily due to cloud companies moving toward purchasing only SMR drives in the future.

As noted in the architecture section of this paper, we predict that magnetic disk drives will predominantly 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 most I/ O requests are serviced from the flash tier while it is preferable that the bulk of the data be stored on a 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, they must then face how to improve I/O performance at the same rate (or better) as the capacity increases. It is fair to mention that a third tier of cloud storage is being offered for exceptionally long-term storage which often utilizes tape.

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 benefits 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.

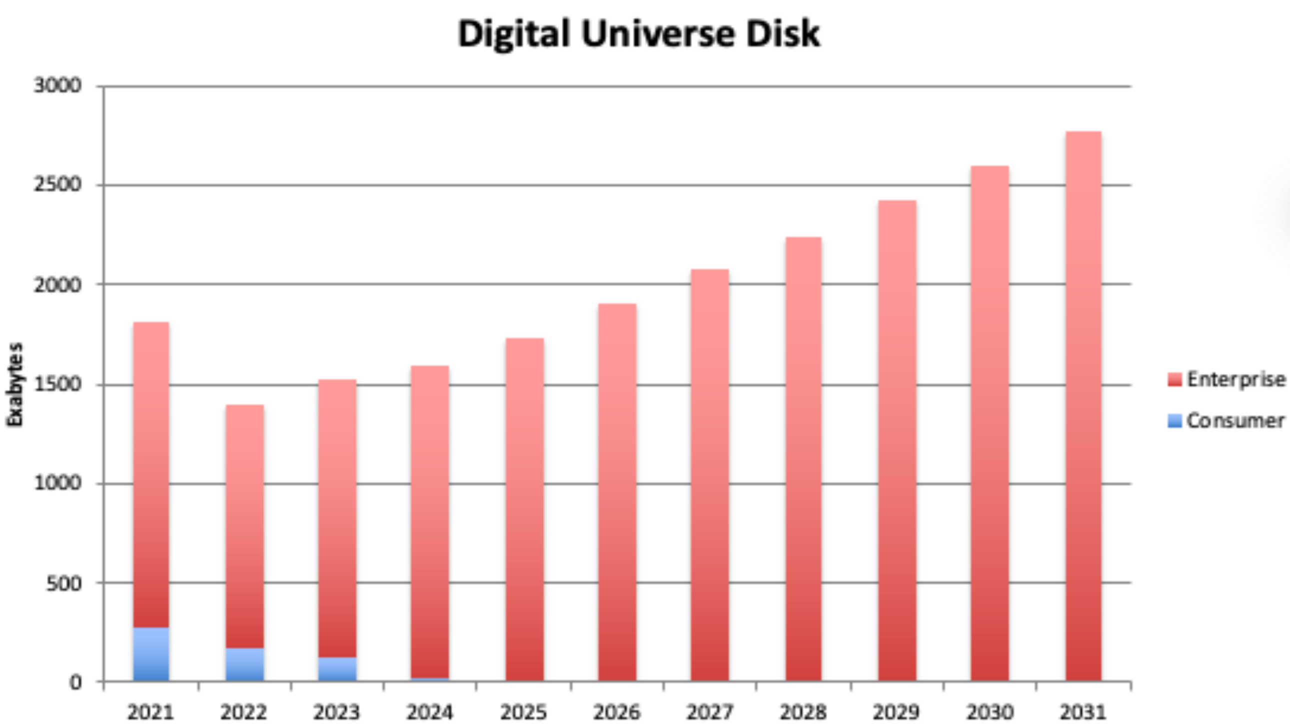

Figure 11: Looking through the end of 2031, Spectra predicts the end of consumer magnetic disk mid-decade

As seen above, Spectra is predicting the end of consumer magnetic disk over the next couple of years as flash disk takes over that space. Though the industry experienced volume declines in the "shipped number" of drives over the last decade, 2022 is the first year ever that the actual "shipped capacity" of disk storage declined year-to-year. As shown above, roughly 1.8ZB of HDD capacity was shipped in 2021 followed by roughly 1.4ZB of HDD capacity shipped in 2022. Much of this can be attributed to the increasing dependencies disk manufacturers have on cloud company purchases which now comprise over 40% of their entire businesses. Given the current economic state, cloud companies have been reducing costs on all fronts, including capital purchases. Capacity increases in enterprise storage will not maintain a pace that will allow the disk industry to realize volume or revenue gains. This could be exacerbated by the technical risk of not being able to deliver a market-ready version of HAMR or MAMR. If disk manufacturers can deliver on greater capacity units as promised, Spectra expects the yearly "capacity shipped" to increase once again by end of 2023.

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 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 a low environmental footprint on both floor space and power, a high level of data integrity over a lengthy 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 (10 slots and under) 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 up to 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.

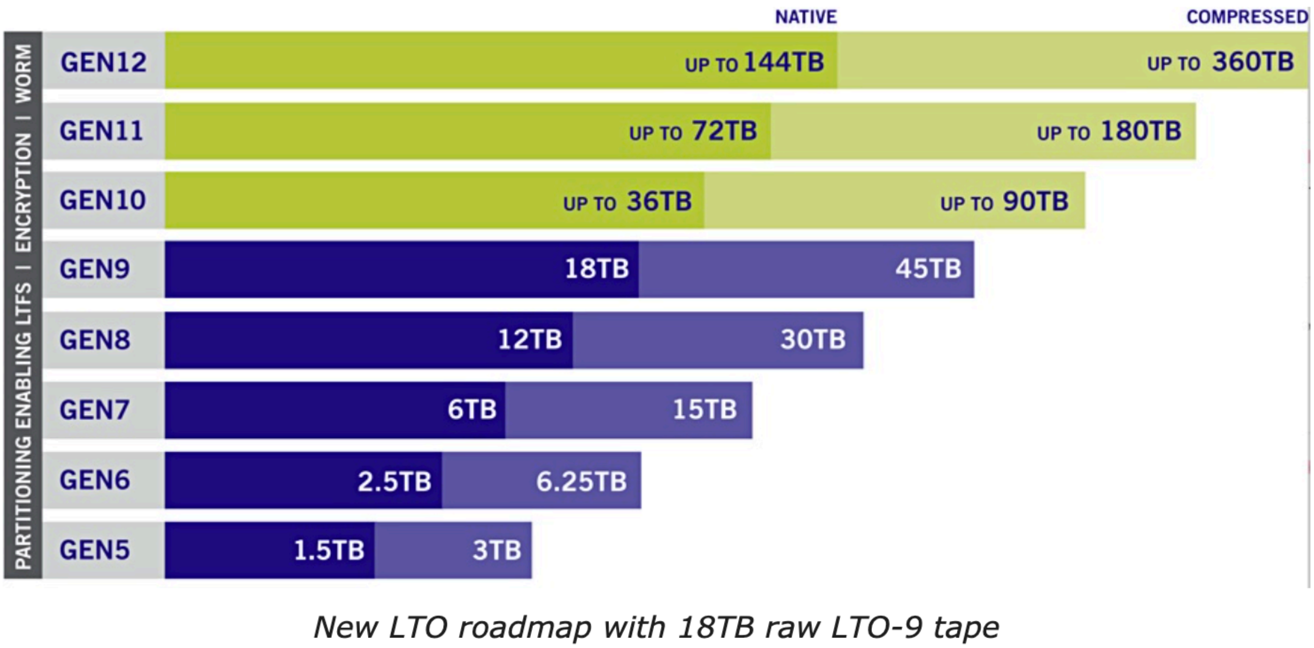

Figure 12: LTO Tape

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, the 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, while the tape path is the same length, smaller decks are subject to more vibration than that of the full-height drive. The half-height drive did not have the stability in the tape path to support the full bandwidth of the full height and therefore is 300 MB/s. LTO-9 cartridges also must be "optimized" prior to use. Customers have the choice of purchasing from vendors that sell tapes unoptimized, or pre-optimized and ready for use, with the latter having a slightly higher cost. When an unoptimized tape is mounted into an LTO-9 drive, the drive senses that the media has never been loaded into a drive and starts the optimization process. Depending on several factors, the process can take from half an hour up to two hours.

This process determines the elasticity of the width of the tape and stores that information in the RFID flash memory component of the tape cartridge. The information is read by the drive each time the cartridge is mounted, allowing the drive to compensate for the individual characteristics of that cartridge. If a previously optimized cartridge is reformatted, then the optimization 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 has or will be required to develop newer technologies to increase capacity as per the LTO roadmap. This may result in half-height drives needing to be slowed down, resulting in lower bandwidth. 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, which measure their performance as a function of capacity, this poses a problem as they must purchase more 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 more recent history and considering an LTO-9 drive that can transfer at 400 MB/s, this would result in a cartridge-to-drive ratio of 55. 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.

There are three primary methods for improving tape drive performance, each posing its own challenges. First the tape media 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 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, it 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 expected to ship IBM® TS1170 Tape Technology in 2023 with a projected native capacity of 50TB. These tape drives use TMR technology (tunneling magnetoresistance) which should allow the capacity to double two more times and form the basis for future LTO generations. The media for the drive is Strontium Ferrite (SrFe) which has higher coercivity than prior generations that were based on Barium Ferrite. For customers who require a RoCE interface there are now vendors selling RoCE to SAS bridges that have been qualified with LTO technology.

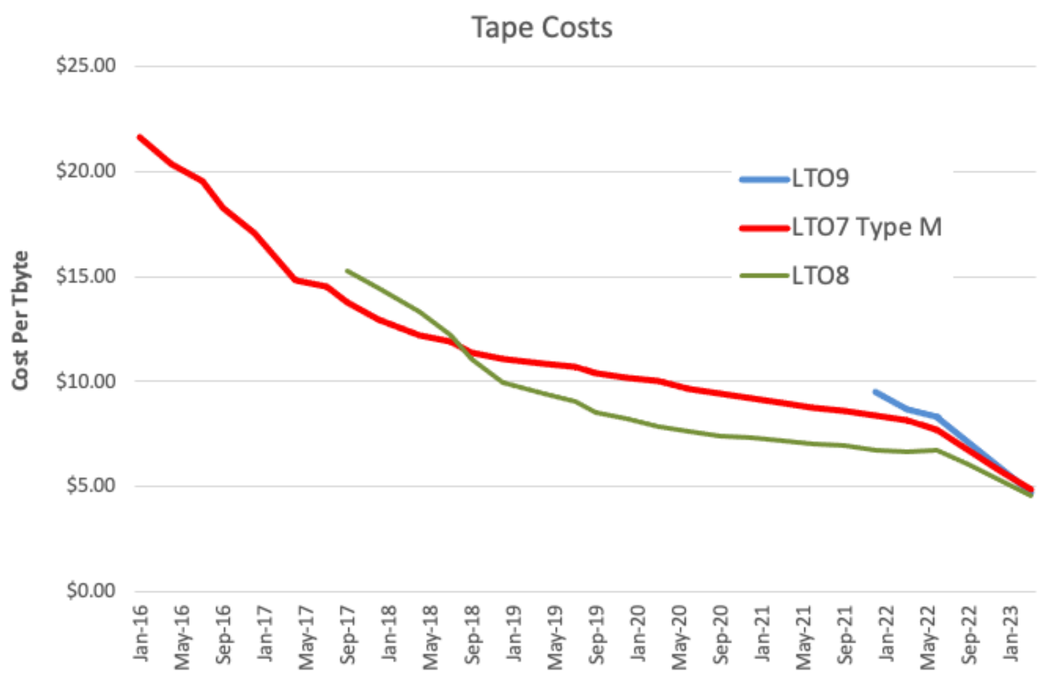

Figure 13: The introduction of new media generations drives down the cost/TB of previous generations

With Oracle's exit from the tape business complete, IBM is now the only tape drive supplier. Fujifilm and Sony are the market's two tape media suppliers. Like other storage technologies, when new generations of tape are introduced, the cost per gigabyte is priced higher than the older technology on the market. Figure 8 shows the cost per TB of different LTO cartridges over time. As can be seen at the time of the 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. Our projections are that LTO-9 media will be more expensive than that of LTO-8 until early 2024. Currently all three tape media formats have converged under $5 per TB. To put this in perspective, at the time of this writing, enterprise disk drives are around $20 per TB while enterprise flash is around $65 per TB.

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 is 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 time out assuming something went wrong.

Another drawback of 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 is mapped into a particular cartridge. The cartridge is mounted, fast forwarded to the correct spot on the tape, the file is read (easily the shortest part of the process) and the cartridge is rewound 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 Amazon S3 interface has become the standard object interface to PUT (write) and GET (read) objects into a cloud or an on-premises object-store.

AWS also has defined several tiers of storage that vary in service level and pricing structures. These tiers fit into two broad classes: online (S3-Normal, S3-Infrequent Access, etc.) and offline (S3-Glacier Instant Retrieval, S3 Glacier Flexible Retrieval and S3 Glacier Deep Archive).

Objects located on an online tier can be accessed directly. Objects in an offline tier must be restored to the online tier before being available for access. The offline storage classes are appropriate for storing archival data, which is data that will be accessed infrequently but kept for a long time. 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 only manage the tape system, but also 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 the S3 interface, countless S3 applications could utilize tape without the need of modifications. Spectra Logic, Quantum and Point have all produced products with this capability. 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 lie 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.

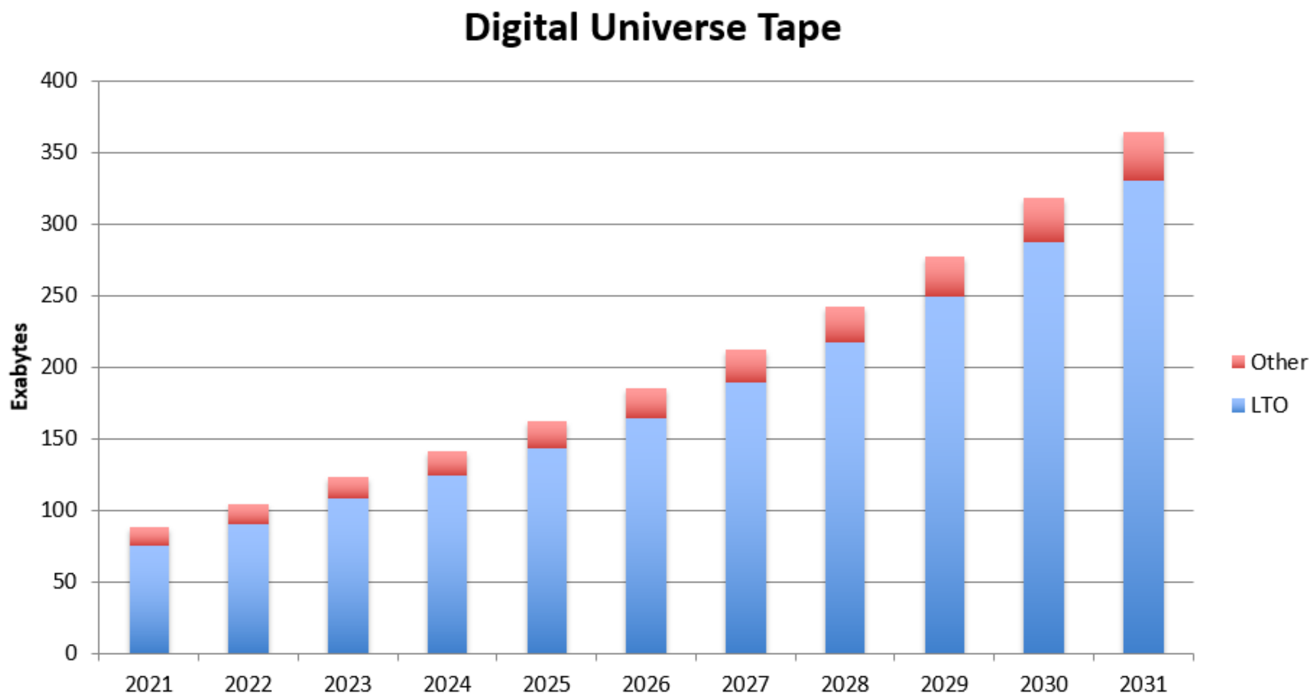

Figure 14: Spectra predicts continued growth in tape storage with LTO far exceeding that of TS tape technology in the volume of units shipped

Future Storage Technologies

The storage industry has and will always continue to attract venture investment in innovative 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. 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 phenomena.

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 a special interest in the technology. For instance, Microsoft has an interest in developing a long-term storage medium 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.

Likewise, DNA storage has progressed through various universities and is now being driven by a consortium of companies. Though these and other efforts can revolutionize data storage, it is difficult to believe that any are mature enough 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 which is discussed in the next section.

Cloud Provider Storage Requirements

According to this forecast, cloud providers will consume, from both a volume and revenue perspective, a larger portion of the storage required to support the digital universe. For this reason, storage providers should consider whether their products are optimized for these environments. This brings into question virtually 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 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 designed for IT data centers may have substantial features that add cost and/or complexity to a product 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? Like 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 the best interface into the tape system is. 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).

Because tape has tighter temperature and humidity specifications than other storage technologies, solutions that minimize this requirement's impact on 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 the small number of localities, ahead of volume commercialization of these technologies. For example, consider an optical technology whereby the lasers are costly, bulky, 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 must 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 Long-term Storage

Years ago, the Gartner Research group defined the hype cycle model. It outlines the phases that a new technology passes through as it is being accepted. In the Gartner hype cycle, 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 staff 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, even from the cloud providers, the discussion has focused on hybrid systems – systems that can take advantage of cloud processing capabilities when they make sense and on-premises processing capabilities when they make sense. We are now entering the "productivity phase" of the Gartner hype cycle.

Referring to the architecture section of this paper, the two tiers of storage are defined as the Primary Tier and the Secondary Tier. Primary 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 Primary Tier is located, as to whether the Secondary 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 Primary and Secondary 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. This phase's output can be daily artifacts along with 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 potential 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 relevant to improving self-driving programs, data 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 Primary Tier and the long-term archival in the Secondary Tier.

The first decision an organization 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 organization and the organization's 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 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 is sometimes the determining factor as to whether an organization 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 focuses on where to locate the Secondary Tier – in the cloud or on-premises. Running processes in the cloud requires the primary 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, 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.

Organizations 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 organizations to have the option of running the Primary Tier on-premises or in the cloud, while ensuring the Secondary storage system is on-premises. Consider a future on-premises storage system receiving all raw data instead of sending it to the cloud. Upon receiving that data, the on-premises storage system would perform two actions. First, it would "sync" the data to the cloud for cloud processing to occur on that data, and second, 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 creates 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 egress charges. For the solution described above, there would be no costs associated with uploading the raw data to the cloud. The output created by cloud processing, which is typically a small percentage of the size of the raw content, would result in minimal egress fees.

In such a solution, the customer could make a head-to-head comparison of a cloud versus an on-premises Secondary Tier solution.

When analyzing the advantages and disadvantages of a cloud or on-premises Secondary Tier solution, there are several things to consider:

- How much data will be stored?
- How long will the data need to exist?
- How frequently and how much of the data will need to be restored?
- How quickly will data need to be restored?
- How committed is my organization long-term to a particular cloud vendor?
- 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 least expensive regions from the three market-leading cloud vendors. Also included are 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)
Amazon S3-Glacier Deep Archive	$.99	Standard -$20 Bulk - $2.50	Standard- 12 hours Bulk – 48 hours	$50-$90	90	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-$85	90	PUT-$.10 GET (High Priority) $50.0 Get (Standard) - $5.0	$60,000
Google Archive	$1.20	$50.00	Sub-second	$80-$120	365	PUT - $.05 GET - $.05	$130,000
On-premises Secondary Tier Storage Solution	Initial investment of $50K+ (one-time charge)	None	Tape –Minutes Disk - Seconds	None	None	None	None

* Using bulk retrieval method

- **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 Secondary 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 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 to 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.

Organizations should consider all costs before deciding on which solution best meets their needs. Misunderstandings can be disastrous, as seen when NASA was surprised by millions of dollars per year in egress fees not considered in their contract with AWS.

Another option to consider is for organizations that need two geographically distinct data copies. In this setup, one copy would be stored on an on-premises storage system, while a second copy would be maintained in a cloud repository. This arrangement could prove to be a cost-efficient approach. The system would be designed to route all data restoration processes to the on-premises storage system, with the cloud repository serving as a supplementary safeguard.

CO_2 Emissions of Information Technology Systems

Storage systems, like all electrical devices, add to worldwide CO_2 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.

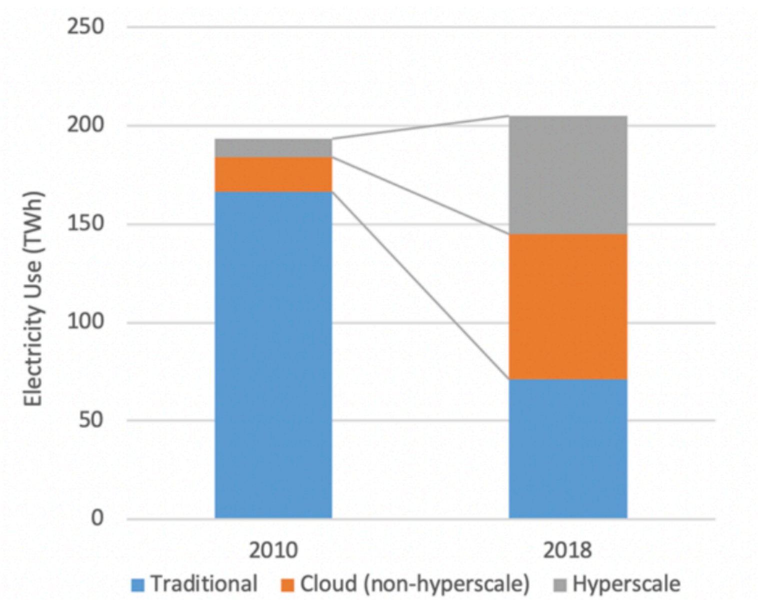

Figure 15: Estimated global data electricity use by data center type, 2010 and 2018. Source, Masanet et al. 2020

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 a six-fold increase in demand for power over the next decade, the question is, "Can electrical usage continue to remain flat and, more importantly, can CO2 emissions be reduced over this period?" To determine the viability of this, we consider the factors that led to flat usage in the previous decade. We will 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 by making more efficient electrical conversions through technologies like those defined in the open compute project.

They are also minimizing the power that is used to cool the equipment using creative water and evaporate decrease their cooling costs with the understanding that the failure rate of components will increase.

However, 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 CO2 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 sufficient scale to deploy these technologies.

Server Energy Intensity – This power intensity of a processor and its surrounding server is defined as the 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 with. 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 at 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.

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 of 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 – Total storage energy is presented as 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 the overall usage of storage devices is under 50%. Like processors, storage devices require virtualization technologies in order for them to be fully utilized.

Once again, cloud providers lead the way by leveraging virtualization methods to achieve a high rate of storage utilization. Likewise, they have the resources to develop technologies to support higher capacity flash and disk storage devices through the implementation of the zoned storage initiative that was discussed earlier.

So far, we have only discussed the power consumption of the new demand. 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. 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 to 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. If deploying tape with an S3 interface, migrations are simplified, and organizations can take advantage of savings provided by future technologies.

Whether in the cloud or on-premises, the CO_2 emissions of storage systems are highly dependent on getting the right data into the right tier. As the chart below shows, the greatest impact on CO_2 emissions is the "global installed storage capacity." Though calculating CO_2 emissions is complex with many variables, flash technology in general has the largest emissions, followed by magnetic disk, followed by tape with extremely low emissions.

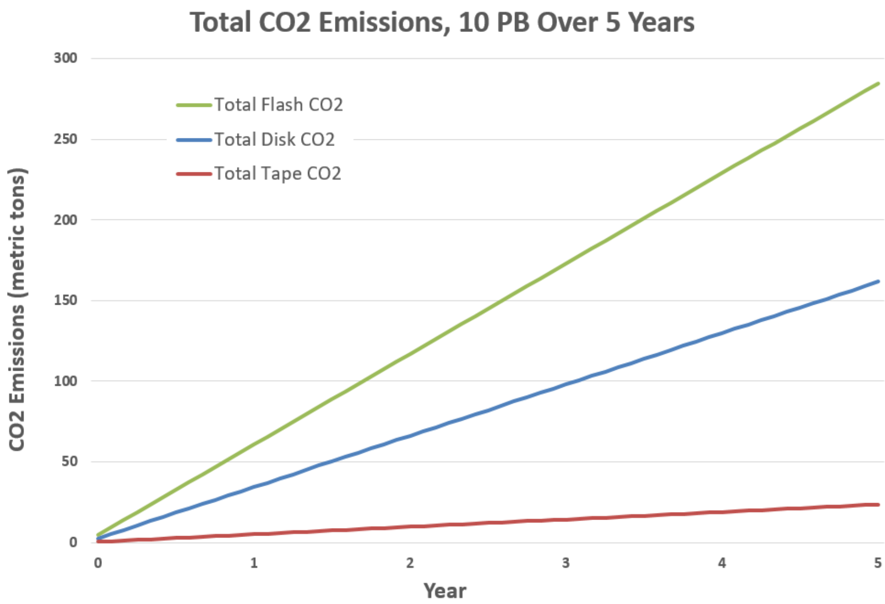

Figure 16: CO_2 Emissions of Different Storage Mediums

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, 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 massive 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) seems aggressive, but possible. The IDC report gave 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 into 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, 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 key factor in determining what can be held online for immediate access.
- Flash, magnetic disk, and magnetic tape storage are 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, 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.

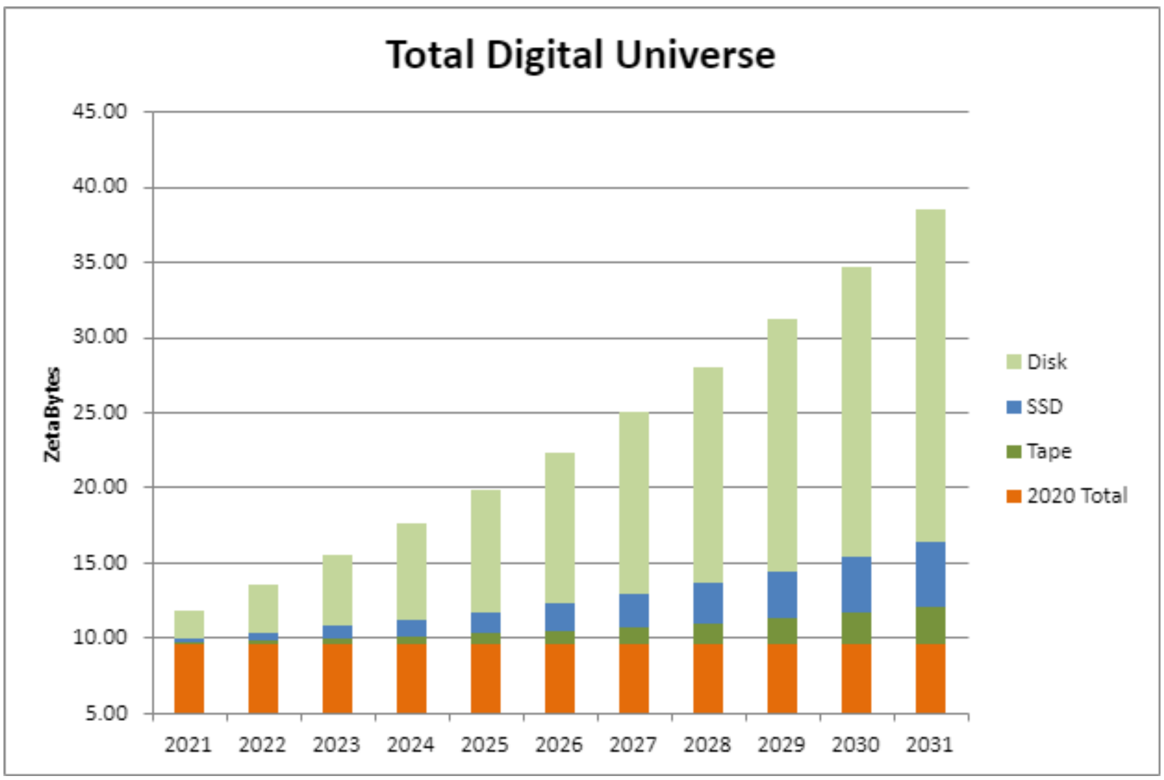

Figure 17: With no deletion of existing data, Spectra would predict 42 Zettabytes of data. Allowing for a 15% per year deletion rate, it is more realistic to expect roughly 38 Zettabytes of data by 2031.

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. With no deletion of existing data, Spectra would predict 42 Zettabytes of data being stored by 2031. Given a deletion rate of 15% per year, Spectra predicts roughly 38 Zettabytes of data actually being stored by 2031.

Conclusion

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 technological 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 allows for the creation of greater density 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 ten years. Finally, tape has enough technological 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, organizations will continue to face the dilemma of what data, and 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 organizations 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, on-premises, or 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 use flash were existing products designed before flash storage was available. For this reason, it was typically integrated into these systems as an extended cache because that is where it could most easily fit into these existing architectures.

Customers gained some benefits, 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, these flash 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 user which allows data to be automatically moved to the right location(s), to the right storage tiers, at the right time. With this capability, organizations 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 various 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. However, 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 eighth 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 updates and publishes this document yearly with new data and new thinking where needed. Please let us know your thoughts.

Appendix Notes

Footnotes:

[1] Source: Measuring Digital Development: Facts and Figures 2022. https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx#:~:text=Latest%20figures%20show%20that%20an,cent%20of%20the%20world's%20population.

Charts:

Figure 2: Source: The CXL Consortium. https://www.computeexpresslink.org/

Figure 3: Source: The CXL Consortium. https://www.computeexpresslink.org/

Figure 6: Source: Zoned Storage. https://zonedstorage.io/docs/introduction/zns

Figure 9: Source: ASTC https://hexus.net/tech/news/storage/123953-seagates-hdd-roadmap-teases-100tb-drives-2025/

Figure 10: SATA-IO. https://sata-io.org/developers/sata-ecosystem/shingled-magnetic-recording-boosting-capacity-and-lowering-costs

Figure 12: 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 a larger compression history buffer.)

Figure 15: 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 Corporation

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.

303-449-6400 · 800-833-1132 · (Fax) 303-939-8844 · 6285 Lookout Road · Boulder, CO 80301 USA
spectralogic.com