Has Optical Archiving Come of Age?

Storage is growing exponentially; tape and hard drives can't keep pace

Executive Summary

Data capacity requirements continue to ratchet upwards because of both new applications that demand massive data sets as well as increasing demands for regulation and accountability. Businesses are required to store more data than ever before; some data sets may be huge but used only once or twice, and much of the data may never need to be accessed but still needs to be stored just in case. As archiving's capital costs increase quickly, operational costs are spiraling up even faster, regardless of whether this archived data is ever accessed. And when it comes to archiving, cost is king as companies struggle to find the lowest cost solution.

Tape has been the primary medium for archiving cold storage, but tape's susceptibility to damage and limited generational support have hobbled it in the face of larger data sets. While **hard drives** have often been touted as a more flexible replacement for tape, hard drives' luster starts to fade as drive capacity curves are beginning to flatten out while data growth continues its upward trajectory. Despite larger capacity drives arriving in the future, their esoteric technologies (and possibly different form factors) employed to achieve larger capacities will force higher prices and possibly lower reliability. This limits the amount of mainstream volume that traditionally drove down the cost-per-gigabyte curve over time. **Optical**, once thought to be too inefficient for true archiving, is now becoming a viable option for the storage and retrieval of these massive data sets. Optical can drive a lower overall total cost due to a lower operational cost and the ability to scale to higher capacities with zero touch over decades. Optical archiving is now poised to capture share, as forward-thinking businesses look for a better alternative to the archiving status quo.

Data Growth is Fueling New Needs

Businesses today face archiving challenges on two fronts. First, new technology is driving **huge data sets**. Second, the changing requirements around **regulatory and compliance issues** are driving the need to store more than ever before. Massive amounts of data must be captured, stored, and analyzed. Today's business models are far more data-intensive than ever before, as technologies like High Performance Computing (HPC), Big Data, analytics, cloud computing, and object storage are driving up data stores, creating management and archiving nightmares. These business models and applications are causing customers in data-intensive industries like oil & gas, healthcare, telecommunications, and financial services to create huge data repositories.

Large data sets are no stranger to archiving challenges, as HPC and analytics have created massive amounts of information that businesses cull for insight. But the sampling rate and sampling depth for data are getting larger, leading to more data points being created more quickly than traditional systems can effectively manage. For

these applications, archiving is not only about older data sets but also the loading and transfer of existing (working) data sets.

As well, regulatory and accountability concerns (especially in the financial sector) are also driving businesses to reconsider their archiving, as the cost of falling out of compliance can far outstrip the cost of collecting and storing all of that data. Compliance data needs to be retained for long periods of time, even though in most cases the data may never actually be touched. If it must be retrieved, then it must be done quickly and accurately, as some regulations even require a 48-hour window for reply. If the archive data is not easily accessible, then businesses may be forced into a mad scramble to retrieve information from offsite storage while the clock is ticking.

Operating Expenses are the Key to Storage Archiving

When dealing with an organization's primary (critical) data, often price is of no object compared to the need for availability and access. But for archived data that might never be accessed, economics will drive the decision. Clearly there is a need for handling the coldest data on the lowest tier of the storage hierarchy, and it needs to be cost effective first, with performance being a less critical consideration.

Different classes of data will have different requirements. The most important data needs to be instantly accessible and at high speed, so storage media like SSD (NAND) and enterprise-class SAS hard drives are typically deployed. But as data is accessed less frequently (if at all), tape (or potentially inexpensive SATA drives) are typically used.

As the coldest tier in the storage hierarchy, the infrequently used data is treated as a "fire and forget" archive, where administrators prefer little or no maintenance because of the low probability that the information will ever need to be accessed again in the future. Ideally, migration-free archiving is preferred where administrators do not need to touch the data or change it in any way so that it can continue to be read—should the need ever arise.

When customers contemplate archiving strategies, they need to consider the position, environment, and refresh rates for the media, which call into play all of the operating expenses that will be required in maintaining the media over time. The environmental costs of keeping tape healthy, along with the need to recalibrate tape drives, adds substantially to the TCO of any tape archiving solution.

Tape Technology is Losing Viability

Tape had always been the weapon of choice for archiving large amounts of data or housing "cold data" that needs to be stored but may never need to be accessed.

Restoring from tape is often difficult and time consuming, as both the media and the readers can be subject to environmental decay, adding complications to any restore process. As many in IT say, "Anyone can backup, but only God can restore."

Tape's two largest hurdles—the fragility of the media and limited backwards compatibility of the media / drives—increase operational costs of using tape as a medium. Unfortunately, these two key factors are not usually captured in the economic ROI models that enterprises have used when evaluating alternatives. This omission leaves customers with the impression that it is still the best option, when in fact tape is costlier in most applications.

Proper calibration and storage environmentals are critical for future tape operation, which increases the probability of storing tape offsite in a controlled environment. However, shuttling tapes offsite also increases the operational cost and shrinks the retrieve / restore window due to the recall logistics involved in the process—a factor that is also an issue when regulatory concerns are accompanied by time pressures. Archive data on tape needs to be remastered every few years to ensure that the data actually can be read and restored, which adds operating cost into the process. Because of the turnover in tape technology, not only does the archive need to be remastered, but often it is remastered onto different formats or media, which creates another set of unneeded variables. From a reliability perspective, tape requires duplicate copies for redundancy, something that other technologies can avoid with RAID or erasure coding which eliminates the need for a complete set of duplicates through parity bits.

Tape media generally stays compatible for two technology generations. Businesses not properly archiving often encounter an unpleasant surprise when a compliance audit forces them to recall archived data, only to find that they do not have the proper equipment to either complete the task or complete it within the required time window.

Hard Drives are Not the Answer Either

In light of the challenges with tape, some have suggested inexpensive hard drives, like SATA, as an alternative for archiving and cold storage. With hard drive capacities growing over time, enterprises had relied on increasing bit densities to drive the cost-per-gigabyte down. Though some saw this trend pushing hard drives into a competitive position relative to tape, the curve is flattening as 8TB hard drives are reaching the practical theoretical maximum of today's technologies, and the engineering tricks to boost capacity, like Shingled Magnetic Recording (SMR), make the drives more fragile.

As drive technologies hit the upper limits of capacity, the probability of new breakthrough technologies that will continue to drive something equivalent to a "Moore's Law of storage" have become more doubtful. In the past customers had always counted on drives doubling in size with the cost-per-gigabyte dropping, but there are practical limits to how far this technology can progress.

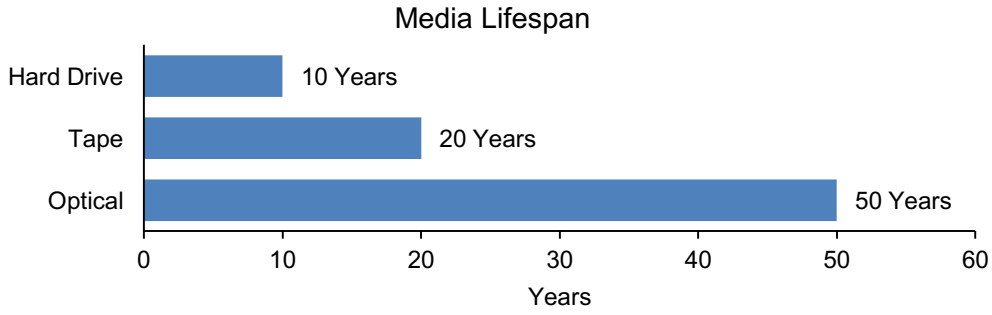
The latest generation high capacity SMR drives increases the risk for data loss; with large drive capacities there is more data per drive at risk. Additionally, as capacities increase, so does the time to rebuild a drive during a failure, growing the probability of additional failures occurring during a rebuild that could result in loss of data. The greater rebuild times push customers into more conservative strategies which increase the number of drives, as well as the total cost to archive.

But Now Optical is Changing the Rules

Optical archiving has been around for a long time, but it was never really considered as an archive or cold storage option because of the perceived cost and performance differential. Optical technology was mainly focused in the consumer space, as CD-RW, DVD-RW, and BD-RE drives were able to keep pace with archiving for client systems but were deemed inadequate for enterprise use because of the media capacity, size of data sets, and the throughput required to both read and write data. Archival Disk (AD), a standard created by Sony, was designed with capacities of 300GB through 1TB, delivering a capacity that more closely aligned with enterprise requirements. But in a single-drive environment with no robotics, the technology was not applicable for enterprise needs, because the proper throughputs still could not be reached.

The low acquisition cost and long-term reliability of optical eliminates data migration. Optical is "contactless", removing the need to reengage with media and data down the road. Tape generations change over time, forcing remastering to access archived data confidently. Optical, on the other hand, has a better track record with older media. In fact, all new optical drives being introduced today can still read from the first generation CD-ROM discs created in 1982, more than 30 years after they were produced. Optical claims the benefit of "50-50-50": accessible for **50 years** at temperatures of up to **50C** (approximately 122F) and **50% relative humidity**—all capabilities that cannot be matched by tape or hard drives. As a matter of fact, during the tsunami that hit Japan in 2011, optical media was found that had been submerged for weeks in seawater; despite the harsh environmental conditions, the media was still able to be read—a feat that would have been impossible with magnetic media or hard drives.

Figure 1: Storage Media Lifespan in Years



Optical media delivers much higher durability compared to tape or even hard drives. Because it is immune to electromagnetic interference, optical does a better job of protecting the data that has been written to it. Additionally, optical consumes less power for each gigabyte written or read, making it a much better choice for datacenters that are driving towards greater power efficiency. Many of these power-efficient datacenters are also experimenting with temperatures above normal operating ranges as well, bringing higher relative humidity which tapes despise but optical can handle easily.

The cost-per-gigabyte of optical media is very close to the cost of tape solutions, and when customers factor in the cost of the drives and robotics along with the media,

optical is on par with tape. But the real savings from optical comes not through hardware or media but through **lower operational costs**.

Similar to tape and hard drive storage, optical solutions also can be accessed by the Storage Management Component (SMC) layer in software, enabling enterprise applications to integrate seamlessly with optical just as they would with tape libraries or hard drives today. This enables customers to make a transition from tape or hard drives to optical with little or no impact to their business or operations.

While the typical optical drive features only a single head for writing or retrieving data on only one side of the media, the new generation of commercial optical archiving solutions will feature multiple heads on each side of the media to achieve a 280MB/s transfer rate, which is on par with the newest LTO-7 tape technology. This leads to much better system read and write throughput of up to 18GB/s (with 64 drives), far beyond the scalability of almost all tape libraries, while also offering an added level of redundancy.

Optical solutions can bring better scalability than competing technology. Solutions like tape introduce additional structures that limit physical density and scalability. Hard drives require power management to reduce the power consumption in times without data access, and HDDs are typically mounted in standard racks limiting the options for cost optimization for higher capacities. Optical can be implemented at a row level in the datacenter where the frontend base unit is the only piece that needs to be powered; all of the expansion for media storage is unpowered and uncomplicated. Where tape typically tops out at roughly 80 petabytes (PB) of capacity, optical can scale much higher. With 7PB of storage available in the base controller unit footprint and 9PB for each expansion unit, optical can scale up to roughly 180PB of total storage capacity—more than twice the capacity of tape—while only consuming roughly 1000W of power with all drives active per row.

The base unit has a tray with 4 enterprise class AD drives and is expandable to up to 64 total drives per system. Robotics move trays between the archiving area of the system and the drive (read or write) portion of the system. While drives are busy storing or retrieving data, the system is staging the next set of disks for reading or writing, supporting a seamless high throughput. With each media load, the AD library achieves a ~1.2GB/s incremental throughput, on par with primary storage. This throughput makes optical a much better choice for larger data sets like HPC or Big Data, where massive amounts of information need to be loaded or transferred to systems.

While optical's higher MTBF (mean time between failure) is a benefit over hard drives, one of the real advantages is that optical disaggregates the data and the reader. Should a reader fail, the data (on the optical disk) can simply be moved to another reader, and the process can continue. Hard drives lock data into physical data / reader enclosures, which creates the need for more redundancy to protect from hardware failures. Hard drives overcome this with RAID, which adds cost and reduces total addressable storage. Both hard drives and optical can use erasure coding to help protect the actual data on the devices, something tape is not capable of doing.

While today's optical solutions are focusing on write once / read many (WORM) technology, a case could be made for expanding in the future to address some customer needs around using a rewriteable cache (disk, tape, or optical)—for example, in environments like HPC where a customer may need to verify data for several months before committing it to an archive. A rewriteable media cache could be an interesting business case for some to pursue.

Optical Archive

Optical Archive, Inc., now a subsidiary of Sony Corporation, has a rich history of both enterprise expertise for handling archiving of critical data as well as scale out datacenter knowledge, which is essential for understanding the handling of large data sets. With an executive team that has spent the last few decades working with companies like Compaq, Dell, Facebook, IBM, Rackable Systems, and Silicon Graphics, the leaders carry with them input from some of the largest storage customers as well as an understanding about the problems that archiving creates for these businesses.

Through the acquisition of OAI, Sony can repurpose its existing optical IP for a rapidly growing market. Consumer Blu-ray demand has been waning with the explosion of online streaming media, but that opens up new opportunity for innovation. Ironically, it is streaming businesses that may benefit from optical archiving, when one considers the massive amounts of customer data—both streaming and click-through data—that a company like Netflix or Amazon might be generating as they drive the market from physical disks and towards streaming.

Sony creates and maintains much of the Blu-ray and Archival Disk IP, bringing opportunities to capitalize on those technologies. But more important to the solution is the IP and leverage of the robotics involved in the Optical Archive Inc. products. This solution uses the robotics components that Sony has employed for more than 20 years in the manufacturing of optical media, meaning the mechanical portion of the solution is battle-tested with years of learning and innovation already integrated into it.

Call to Action

The explosion of data is forcing many companies to reconsider their archiving strategies. Companies that need to deal with large amounts of cold data should clearly be looking at optical as a viable alternative to the current choices in the market. Optical can now surpass tape from a total system capacity and performance perspective, making it a better choice for very large data sets. Additionally, the durability, backwards compatibility, and low cost make optical a better solution for general archiving as well. Because the purest definition of "cold data" is that which may never be accessed once it is written, optical is a better solution, because it removes the need for special handling or remastering, driving down the long-term operational costs of archiving the data.

Companies with large data archiving needs should consider optical as part of their IT strategy. As a division of the industry leader Sony, Optical Archive, Inc. has access to both the knowhow and IP to make optical a successful part of any company's storage infrastructure.

Important Information About This Brief

Inquiries
Please contact us if you would like to discuss this report, and Moor Insights & Strategy will respond promptly.

Citations
This note or paper can be cited by accredited press and analysts but must be cited in-context, displaying author's name, author's title, and "Moor Insights & Strategy". Non-press and non-analysts must receive prior written permission by Moor Insights & Strategy for any citations.

Licensing

Disclosures
This paper was commissioned by Optical Archive/Sony. Moor Insights & Strategy provides research, analysis, advising, and consulting to many high-tech companies mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

DISCLAIMER
The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. Moor Insights & Strategy disclaims all warranties as to the accuracy, completeness, or adequacy of such information and shall have no liability for errors, omissions, or inadequacies in such information. This document consists of the opinions of Moor Insights & Strategy and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

Moor Insights & Strategy provides forecasts and forward-looking statements as directional indicators and not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of the date of publication for this document. Please keep in mind that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements in light of new information or future events.