White Paper

Cloud Storage: the Next Frontier for Tape

By Mark Peters, Senior Analyst; Jason Buffington, Senior Analyst; and Monya Keane, Research Analyst

April 2013

This ESG White Paper was commissioned by Oracle and is distributed under license from ESG.

Contents

Summary

This paper explores the usage scenarios and economics of public, private, and hybrid cloud storage services—particularly in the context of recent evolutions in tape technology and usage. By taking a cohesive look at today's interrelated cloud/tape landscape, we conclude that there is a definite place for tape in the cloud services landscape.

Introduction: Storage Solutions for Service Providers

In the past decade-plus, a pendulum has swung back and forth between peoples' penchant for on-premises storage and outsourced storage services. These location and control preferences have fluctuated depending on shifts in the dominance of enterprise computers, distributed servers, PCs, appliances, and changes in the price and capabilities of storage. Cloud technology is the latest iteration, and it is delivering great flexibility to storage users.

- **Public cloud storage services,** such as Amazon's S3 (Simple Storage Service) or the vertically focused T3Media Platform provide a multi-tenant environment highly suitable for unstructured data and long-term archiving.
- **Private cloud storage services** provide a dedicated environment behind an organization's firewall. Private clouds represent an evolution over traditional infrastructure and its management, which is most appropriate for users needing customization and control over data beyond what is available in public cloud offerings.
- **Hybrid cloud storage** includes components from at least one private cloud and one public cloud infrastructure. An organization might store actively used, critical, and structured data in the private cloud, while storing unstructured and archival data in the public cloud—or it may store data that must be accessed from multiple locations in the public cloud, while keeping local data in a private cloud.

In the cloud storage scenarios above, the storage can simply be thought of as "agile capacity" offered in private/hybrid form by a central IT organization or in public/hybrid form by a service provider.

But to completely appreciate storage as it relates to a cloud architecture's abstract delivery model, one also must consider service-based software delivery (a.k.a. SaaS) and platform functionality (PaaS). With SaaS and PaaS, new production services are being consumed, which affects the agile storage within the service provider's infrastructure. Although such arrangements can spur academically interesting discussions, the idea quickly becomes semantically and pragmatically confusing. (Think of a "public cloud software-as-a-service solution" built upon private-cloud storage within a service provider's operation.)

In any case, in almost all permutations, the desirable characteristics of the cloud come down to **agility**, **capacity at scale**, and **economics**. As cloud storage service providers take these three dimensions into account for at-scale infrastructures, tape is becoming the preferred medium for long-term data retention and archiving.

Agility

Looking more closely at drivers of storage capacity (regardless of private, hybrid, or public implementation), respondents to ESG research reported that the top-five storage infrastructure areas in which they would make significant investments in 2012 were:

- Data protection (44% of respondents)
- Storage virtualization (30%)
- Data replication (28%)
- New SAN storage (25%)
- Tiered storage (24%)[1]

[1] Source: ESG Research Brief, *2012 Storage Infrastructure Spending Trends*, March 2012.

It is interesting, though not surprising, that two of the top-three areas for investment relate to data protection. The other three approaches focus on delivering storage with more agility. And again, storage agility is suited equally to private/hybrid cloud solutions managed by IT and to backend storage provided by public cloud and by public SaaS and PaaS providers (who would likely depend on all five of the usage scenarios).

Capacity at Scale: Coping with Extreme Digital Data Growth

This is no surprise to anyone in IT who's faced with data proliferation: We have entered the "era of big data" (scale and analytics) as well as the era of "bigger" (more and unstructured) data. Growth rates are especially high for digital archives of compliance-related data (both government and corporate mandated), video/multimedia files, digital images, fixed content, and social-network-related data.

As cloud storage technology evolves, it can be difficult to define what its capabilities and benefits will look like years from now. However, two areas of cloud storage already seem to have very persuasive benefits—using the cloud for storing **archives** and for storing **unstructured data**.

An appreciable amount of compliance and archive data is unstructured, and management challenges abound with regard to a lack of naming conventions, taxonomy, tags, search keys, and indexing capabilities. Unstructured data is stored as file (not block) storage, with tape and disk historically both being used in data centers to meet the unstructured-storage capacity demands.

Cause and Effect

Some archival data may be kept for decades or even "forever," although rarely (if ever) accessed. And now, archive growth is occurring in large part due to the activities of billions of data-loving people in addition to tens of thousands of traditional data-generating businesses. The situation demands effective archiving solutions that make data easy to find while lowering the cost of ownership for whatever organization is ultimately tasked with the responsibility for storing it.

Today's archival storage environments need to be robust, cheap, secure, and scalable. Fortunately, significant enhancements to tape drives, tape media, and libraries have occurred in recent years, meaning that for a lot of use cases, tape technology is now an optimal choice for staying ahead of data growth. To be specific, tape is now exceedingly appealing for use as cloud storage. In fact, some cloud storage service providers are reaching unprecedented levels of cost control, security, scalability, and availability by deploying tape in the cloud.

Storage Economic Considerations

Economic factors have a huge influence on the storage initiatives and storage infrastructure that an IT organization can deploy.

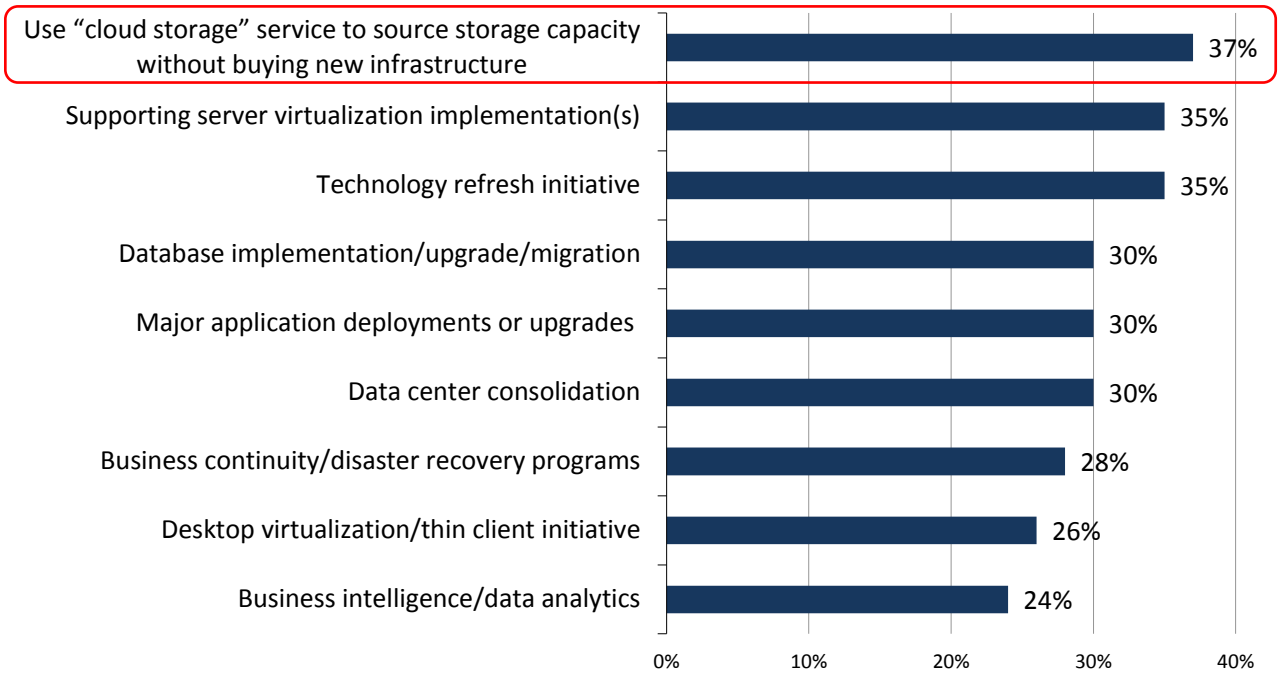
The flexibility and potential cost savings of both public and private clouds hold great appeal for several reasons. Cloud storage provides elasticity, with capacity growing on-demand as the business requires and scaling back when capacity is not needed. You pay only for what you use. Typically, public cloud storage service providers charge for the capacity used monthly, the data-transfer bandwidth, and value-added services performed in the cloud (i.e., security, encryption, deduplication, or replication to other locations for redundancy).

According to an ESG survey of IT managers, use of cloud storage services is the initiative most commonly cited as a significant source of expected storage spending in the immediate future (see Figure 1).[2]

[2] Source: ESG Research Report, *2012 Storage Market Survey*, November 2012.

Figure 1. IT Initiatives Expected to Significantly Affect Storage Spending over the Next 12-18 Months

Which of the following IT initiatives do you believe will significantly impact your organization's storage spending over the next 12-18 months? (Percent of respondents, N=418, multiple responses accepted)



Source: Enterprise Strategy Group, 2013.

Traditionally, the storage industry has used "$/GB" (dollars per gigabyte, or the total purchase price divided by total capacity) as a standard pricing metric. Cloud storage providers, on the other hand, are using cents per gigabyte per month (¢/GB/month) as a pricing metric, which can make their prices appear considerably lower.

Cloud storage *can* be less expensive than privately owned storage. After all, you don't have to pay for unused disk capacity. And the savings can be substantial when you consider that, according to one ESG survey, unused disk capacity can represent 40% to 47% of total disk capacity (see Table 1).[3]

Table 1. Storage Hardware Utilization—Lower at Organizations Citing DAS as Their Primary Disk-based Storage

To the best of your knowledge, what is your organization's overall storage hardware utilization rate?	
Organizations citing direct-attached storage (i.e., internal server storage, DAS) as their primary disk-based storage technology	**Organizations citing networked storage (i.e., NAS, Fibre Channel/iSCSI SAN, unified storage) as their primary disk-based storage technology**
53%	**60%**

Source: Enterprise Strategy Group, 2013.

However, cloud storage can become *more* expensive because "GB transferred" charges vary with the amount of data transferred each month, potentially with extreme and unpredictable fluctuations. In addition, extra hidden fees, such as connecting fees, account maintenance charges, and data access charges, can add up quickly. Most disk-based cloud storage services charge you the same amount per gigabyte to store "cold" archive data as they charge for "hot" active data, which can really add up. Understanding cloud pricing is a vitally important element in any decision to use a cloud service for data storage.

[3] Ibid.

The Increasing Relevance of 'Tape in the Cloud'

Some providers that are trying to offer capacity at scale consider using massive arrays of idle disks (MAID systems), which can approach the accessibility of production disk while spinning down economically when not in use. Other cloud service providers prefer tape rather than idle disk.

Why isn't MAID the ubiquitous choice? MAID architectures have legitimate limitations due to the spin-down/deep-sleep design of the storage for deep archiving. Many people are discovering that the 20-year-old adage that "disk always performs faster than tape" doesn't ring quite as true in today's cloud-based storage architectures, where the performance of enterprise-class tape robotics and tape formats with file-system-like accessibility may surpass the performance of MAID subsystems—at an appreciably lower price.

"Cold storage" or "deep archiving" is the primary use case for tape by cloud service providers because of the problems that using tape solves. For example, tape is often a good choice as the infrastructure behind the public-cloud side of a hybrid solution for backup or archive—where private, on-premises, short-term protection is disk-based, and overflow or longer-term, at-scale storage is housed by the cloud service provider on tape.

The tape industry is expanding the primary market applications that it serves, going beyond just backup and recovery to support compliance and long-term archival storage needs. Waves of new and strengthened data regulations together with the explosion of unstructured-data archival requirements are redefining the future for tape in important ways.

Delivering cost-effective storage, protection, and retrieval poses challenges for any large archival storage system. Oracle has responded to those challenges with tape technology that pushes up capacities, reliability standards, and data transfer speeds to impressive levels via its enterprise-class tape drives. The leap ahead represented by this next generation of enterprise tape technology has made tape a principal storage medium for data archiving and long-term preservation in the cloud.

The Era of Modern Tape Has Arrived, and Exascale Storage Capabilities Are Here

In particular, the Oracle StorageTek T10000C tape drive has a native capacity of 5TB (which equates to more than 10TB with typical 2:1 tape compression), and it supports both mainframe and open systems computing environments. With speeds comparable to traditional disk drives (with a boasted native data rate of 252MB/sec), the T10000C is well suited for transferring large data streams from applications such as those used in the media, entertainment, and broadcasting industry.

The reliability of the StorageTek T10000C tape drive—according to Oracle's accelerated lifetime and reliability testing—is three orders of magnitude higher (with a $1x10^{-19}$ uncorrectable bit error rate) than the most reliable Fibre Channel disk drive (with its uncorrectable BER of $1x10^{-16}$). The T10000C tape media also offers a data-retention lifespan of 30 years or more. Using T10000C tape drives in an Oracle StorageTek SL8500 Tape Library can provide more than one exabyte of capacity for public, private, or hybrid cloud service providers.

A 2011 study by ESG comparing an LTO-5 tape library system against a SATA disk system with deduplication showed that in several use cases for backup, using deduplicated disk resulted in a 2X to 4X higher TCO than using a tape system did.[4] And with Oracle StorageTek T10000C enterprise, that differential would be even higher. Of course, as mentioned, tape has been expanding its historical role as a short-term backup storage solution to support a much broader set of requirements, including long-term data retention for archiving.

The current trend to consolidate unstructured data into large-scale repositories in the cloud signals that another category of storage is emerging—*tape in the cloud*—which is further bolstering the economic model for the cloud as a preferred repository for archiving. It's worth noting that often, the cloud model—especially the public cloud model—replaces TCO with ROI. Tape's existing cost advantages might become even more compelling (for various affected parties) as future cloud technology developments occur.

[4] Source: ESG White Paper, *A Comparative TCO Study: VTLs and Physical Tape—With a Focus on Deduplication and LTO-5 Technology*, February 2011.

Research and development efforts at Oracle (and across the tape industry) have resulted in a situation in which a tape environment can offer higher-scale capacity than disk, with more reliability, for appreciably less cost. ESG expects this trend to continue. Tape may enable cloud storage providers to lower their internal costs and prices far below any comparable disk solutions for cloud archive services, without discernible performance differences.

How Modern Tape Is Being Used in the Marketplace

Two of tape's primary use cases are backup and archive for both onsite and offsite storage (see Table 2).[5]

Table 2. How Organizations Currently Use Tape, by Total Disk-based Storage System Capacity

"We use tape to store our …"	Less than 100TB (N=139)	100TB to 499TB (N=88)	500TB or more (N=115)
… **backup** data **offsite**	55%	55%	51%
… **archive** data **offsite**	42%	50%	54%
… **backup** data **onsite**	48%	48%	41%
… **archive** data **onsite**	25%	29%	35%

Source: Enterprise Strategy Group, 2013.

Although the terms are sometimes confused, "backup" and "archive" are not the same:

- Backup relates to the process of copying data so that the copy may be used to restore the original after a data-loss event.
- Archiving is the process of moving data that is no longer actively used to a separate data storage device for long-term retention. Archive data *is* the primary copy (moved to a new tier/location) and it should, itself, be protected.

Tape is used for backup and archive in cloud infrastructures serving the public sector and many vertical industries—healthcare, media and entertainment, oil and gas, education, and others. Several examples of use cases for tape in these markets have been publicly documented by Oracle.

Government

Many of these sites have effectively become web-based archives; the U.S. Library of Congress is a good example. It is a large Oracle tape user and is essentially now a public cloud service, with users accessing information via web browser.[6] In addition, the federal government has issued a broad mandate in general to move quickly (as a provider or consumer) to cloud services wherever possible.

Healthcare

A picture archiving and communication system (PACS) is medical imaging technology that relies heavily on tape in its storage-intensive environments. Many PACS environments are moving quickly to a cloud-based model for archiving. For example, Novant Health has implemented Oracle StorageTek T10000 tape drives with StorageTek SL8500 library systems in separate data centers that are 26 miles apart. Novant first migrated its PACS environment to the new architecture, and it has now started to evaluate other applications to add, including two cardiology imaging applications.[7]

[5] Source: ESG Research Report, *2012 Storage Market Survey*, November 2012.
[6] Source: *Archive Storage Infrastructure at the Library of Congress*, September 2011.
[7] Source: *Novant Health Uses Long-Term Archive Solution to Securely Store 910 Million Files with Combined Storage of 776 Terabytes*, undated Oracle case study.

Media and Entertainment

T3Media offers cloud-based storage, access, and licensing services for master-quality video. The company delivers large-scale archive management and "smart content" metadata as a hosted service. It uses Oracle tape to support companies such as Paramount Pictures, Sony Pictures Entertainment, *National Geographic*, *The New York Times*, and the media-focused operations of the NCAA. T3 is reselling storage capacity (on tape) as a service, and it offers a good example of how the customer base for tape is extending beyond private enterprise tape purchasers to service providers, too.[8]

Internet Cloud Service Providers

In late 2012, for the first time, Google published photographs taken inside its data centers. Google never explicitly acknowledged that the tape libraries in its photos are from Oracle (and notably, Oracle refrains from citing Google as a reference customer); still, it is readily apparent that the numerous libraries shown are Oracle SL8500 library systems packed with tapes.[9] An accompanying guided-tour video of Google's data center in Lenoir, North Carolina, showcases a bank of SL8500 tape libraries with their nameplates removed and their robotic arms working busily, as voiceover narration explains that Google uses the devices "to store another copy of the *most important* data."[10]

The Impact of the Linear Tape File System in Cloud Storage

In contemporary tape usage, particularly in cold storage environments, perhaps the biggest recent advance has been in the ability to improve user access to tape-based data by using the LTFS (Linear Tape File System) format. LTFS was unveiled by Oracle, IBM, and HP in April 2010. It is supported on midrange LTO tape technology as well as Oracle's T10000 enterprise tape technology. With LTFS, longstanding rules of tape access are changing—sequential search times for tape are replaced with disk-like access using familiar drag-and-drop techniques. The tape media's self-awareness and the open-source format of LTFS allow tape cartridges from one system to be easily accessed by another, much as one can take a USB drive from one laptop to another.

With LTFS, tape media uses two partitions: a small index partition containing file system metadata, or index data, and a partition containing the file content itself. The improved management and faster tape file access capability provided by LTFS become increasingly important as tape cartridge capacities and the number of files per tape steadily increase.

In relation to archiving, LTFS targets active archiving of unstructured data such as medical images, media and entertainment files, and some data related to compliance-heavy applications, while providing for much faster retrieval times than ever before. Incidentally, Oracle is the only vendor to support LTFS technology on enterprise tape drives such as the StorageTek T10000C.[11]

LTFS adds considerable credibility to tape as an effective archive (and deep-archive) solution, and it could open a lot of doors for tape use by cloud service providers. At this point, LTFS has barely scratched the surface of its potential, and its future looks extremely promising. Fifteen years ago, IT organizations that wanted to deploy backup solutions with better performance and differing features looked first at disk, but the backup software didn't "know" how to use disk effectively. IT then turned to virtual tape libraries (VTLs), which leverage disk that "appears to be tape" to the software.

Today, LTFS makes tape behave much like disk—albeit with improved economics—so that software solutions that work with disk storage may be "pleased" with data from tape (i.e., those solutions are unaware that they're accessing tape, not disk).

[8] Source: *T3Media Grows Global Video Management and Licensing Business with Oracle Storage*, October 2012.
[9] Source: http://www.google.com/about/datacenters/gallery/#/tech, 2012.
[10] Source: "Take a walk through a Google data center," http://www.google.com/about/datacenters/inside/streetview/, October 2012.
[11] Source: Oracle data sheet, *StorageTek Linear Tape File System*, 2011.

Improving Storage Security for the Cloud

Many prospective cloud storage users surveyed by ESG say that security concerns in general and a lack of appropriate security tools in particular are among the top factors preventing their organizations from using cloud storage more pervasively.[12] Those concerns are legitimate, and CIOs and IT managers must address them when deciding whether to move to cloud-based storage. Public clouds are, by definition, public—users must be willing to have their data reside side-by-side with data belonging to possible competitors and other completely unknown parties.

Additionally, disk arrays in the cloud use RAID algorithms to break up and spread data at the block level across multiple disks for resiliency. In multi-tenant cloud environments, the potential result is that pieces of data from multiple cloud customers could commingle on the same drive. It also means that a drive failure could affect a large number of users. Additional software and possibly encryption are required to ensure that a particular customer's data is isolated securely.

Conversely, in a tape environment, each tape cartridge is a separate object. The customer or cloud provider has control over which files go onto which tape. Users of Oracle tape libraries can also use physical partitioning, a new security capability presenting multiple virtual libraries to cloud customers while preventing inappropriate inter-accessibility. In addition, the encryption and WORM capabilities of tape provide security for data at rest, as they are fundamental to delivering a secure cloud archive strategy. StorageTek T10000C tape drives boast data integrity levels that are considerably higher than typical Fibre Channel disk drives.

Tape is also easily movable/removable. The removable/vaulting capability of tape as it relates to security centers on the fact that it can be literally locked down to prevent unauthorized access. Someone would have to physically "rob a bank" to get at the data if it is stored offline in a vault. That kind of security is not generally available with disk storage, which is one reason why storage service providers such as Iron Mountain use tape technology as the foundation for many of their offerings.

The combination of all those factors positions Oracle tape libraries with T10000C technology as an optimal, secure choice for providers looking to offer deep-archive and other cloud storage services.

Managing Tape Resources at Scale

Many types of IT environments that are designed to operate "at scale" require easy yet sophisticated management.

For many cloud service providers with storage-as-a-service offerings, the requirement may involve simple yet robust tape management, monitoring, and support capabilities. Oracle has addressed the need for effective tape monitoring and reporting tools with StorageTek Tape Analytics (STA) software, taking a step forward in making a tape environment easier to manage, monitor, diagnose, and maintain.

STA software captures tape library, drive, and media performance metrics and presents them via a single pane of glass. The software proactively identifies emerging problems while providing recommendations to prevent data loss. This product is a step forward in giving archive-as-a-service providers and other cloud service vendors readily accessible tools for tape operations and management.[13]

At a high level, the main linkage between cloud storage and high management/reporting efficiency is simply that more efficiency can lead to a better cloud storage product (equating to potentially better *revenues*) and more efficient cloud storage operations (equating to potentially better *margins*).

[12] Source: ESG Research Report, *2013 IT Spending Intentions Survey*, January 2013.
[13] Please note that STA works only with Oracle StorageTek Modular Library Systems.

The Bigger Truth

Tape, in addition to being IT's venerable "go-to-guy" for data backup, has now improved its position as the preferred choice for long-term data retention in the cloud.

One could make an argument that cloud and tape are each other's "best friends." It wouldn't be entirely appropriate to make a sweeping generalization that tape is *always* economically superior for cloud services, with no exceptions existing whatsoever. But tape—when run and used optimally in the cloud—is pretty much guaranteed to cost less per GB (OPEX and CAPEX) than equivalently well-run disk would cost.

The bottom line is that tape in the cloud offers a significantly lower total cost of ownership (primarily by leveraging low $/GB and significantly lower energy costs), better reliability, longer life, and as good or even faster streaming performance than today's disk alternatives. The tape industry needs to do more to change perceptions that are *not* in line with this reality in many circles … because too many potential users still remember tape as it was--not as it is now.

Within the tape industry, the most recent technology improvements—not to mention some intriguing future roadmaps—suggest that tape is, and will continue to be, the most cost-effective solution for many large-scale storage opportunities that lie ahead in the various manifestations of the cloud (private/public/hybrid).

Those numerous improvements to tape make it an optimal infrastructure choice for "at-scale" cloud-based archiving for the foreseeable future. It's time to realize that leveraging a cloud infrastructure can make tremendously good economic sense for an IT organization … and tape makes economic sense for the cloud providers of "at-scale" storage services. Remember that economically speaking, the difference between "user" and "provider" is that users look at TCO (whether they are doing it themselves or using a cloud), while cloud providers (public providers, certainly, but maybe private ones, too) look at ROI as well.

IT trends are always changing. And CIOs and IT managers are beginning to, once again, see the value of tape … this time, as an ideal choice for long-term data archive in the cloud.

Enterprise Strategy Group | **Getting to the bigger truth.**

20 Asylum Street | Milford, MA 01757 | Tel: 508.482.0188 Fax: 508.482.0218 | www.esg-global.com