



Keep Your Archive in the Cloud — *on Tape!*

Analyst: David Reine

Management Summary

Have you tried on your fall or winter clothes yet? If you have, you may have discovered that you have grown a bit since last year, or even since last month. Every bite that you take at the dinner table seems to go straight to your waistline or hips. The more that you consume; the more difficult it is to fit into your wardrobe. You are no longer like the proverbial teenager that can consume everything in sight and still not gain any weight. The waistline of your pants does not expand very much, even if you do. You may even be struggling to walk up that flight of stairs as your health deteriorates. No problem, *you can always take the elevator*. Perhaps, it's time for you to go on that diet, but it is too late for today. Today, you need to go shopping for more clothes, spending more money that you didn't plan on, and filling up a closet already jammed with clothes, if your significant other or roommate will allow you to use any more space than what you had been using.

Filling up your closet at home, however, pales in comparison to the troubles that you may face in the data center. Storage is expanding everyday as the enterprise tries to protect its most valuable asset: *information*. Your primary data is expanding at a precipitous rate, with projections estimating that in ten years your data will increase by ten-fold, not to mention the impact of this growth on your secondary disk backup requirements. Your IT storage requirements for both structured and unstructured data are expanding to a point that they no longer can be met by today's IT budget. Using lower-cost Tier-3 tape for short-term backup requirements usually is less expensive¹, but may be unacceptable due to the immediacy that may be required for data or file recovery. As a result, Tier-1 and Tier-2 storage acquisition costs, not to mention floor space and energy requirements, continue to rise.

Reducing the volume of data that needs to be backed-up has become critical. One way to do that is to archive rarely used data off Tier-1 and Tier-2 disk storage and onto Tier-3 tape. With recent innovations in tape hardware and software, the data center can now use tape that acts like disk. With new file systems acting as a front end to tape, along with extensions to standard file systems, the lifecycle of tape has come full circle, with tape once again gaining prominence in the data center as an extension to disk. With proven cost-effective storage media, the IT staff can deliver reasonable online access to rarely used data, using tape as an *active archive* to make enterprise data archives searchable, accessible and available, re-connecting users to lost or forgotten data, and all of this being done with a consistent and easy-to-use interface. Yes, this really is something new!

An active archive contains infrequently-accessed enterprise data, fixed or variable records that still can be retrieved online through a familiar file system interface, which facilitates ease-of-use and enables a search of all of the historical data. One of the goals of an active archive is to improve storage density and reduce footprint to lower cost and to ensure media and data integrity – in order to promote business continuity. One company that has implemented an

¹ See the issue of *Clipper Notes* dated December 20, 2010, entitled *In Search of the Long-Term Archiving Solution – Tape Delivers Significant TCO Advantages over Disk*, and available at <http://www.clipper.com/research/TCG2010054.pdf>.

IN THIS ISSUE

➤ Never-Ending Data Growth	2
➤ Who is Open Source Storage?	3
➤ The Infinidisc Cloud Archive	4
➤ Conclusion	5

effective archive solution to restore order to the data center is Open Source Storage, Inc., of Littleton, Colorado. To learn how their *Infinidisc* offering might help you, please read on.

Never-Ending Data Growth

In the past few years, data center storage has undergone an era of unprecedented growth. Between mergers and acquisitions, real-time data analysis, consolidations and virtualizations, industry regulations and government compliance, enterprise data centers are storing more data – and more kinds of data – than ever before, not only for backup and recovery, but for long-term archiving, as well, in order to take advantage of the long-term value of this data. Every enterprise is keeping information on hand for longer periods of time as the value of that data continues to grow. Have you ever seen the TV show *Hoarders*? Well, the fact of the matter is, some enterprises have a fear of deleting any data, in effect hoarding it, whether or not they know why or, even, if they are aware!

Mission-critical applications demand more and more data. Healthcare institutions seem to have a never-ending need for more patient data, in the form of both medical histories and images, such as x-rays, MRIs, and CAT scans. The entertainment industry creates all kinds of animated sequences and digital images while news agencies are saving all audio and video for possible reuse.

Data retention policies are being set based upon both data value and the requirements to meet legal compliance (such as to address eDiscovery inquiries), as well as to satisfy a more general future value as *reference source* (i.e., something that might be needed at a future time). Some data files, such as medical images, energy exploration data, and digital videos, need to be saved for reuse at a later date; they may never be deleted. All of these enterprise policies are leading to an explosion of data, both structured and unstructured, with a growth factor of ten times over the next ten years. Quite clearly, this can result in serious maintenance and administration issues.

Today, much if not most of this data is managed on Tier-1 storage, regardless of whether that is on Fibre Channel, SATA, SSD, or SAS drives, or some combination thereof, with the commensurate costs of acquisition. The vast bulk of it tends to be stored and maintained on spinning disk drives, consuming significant additional enterprise resources in the form of storage administrators, floor space, and energy. Often forgotten are the

continuing disk acquisition costs required to replace disk arrays every three-to-five years². Also, in order to protect the enterprise in the event of a storage failure, all of that data usually needs to be backed-up on additional disk storage – to be available for immediate recovery (as tape may be an inappropriate choice, in many cases, for the immediate recovery of critical information). In fact, some of this preserved data may be read only once or twice, or even not at all. The IT staff simply has got to get control of this rampant growth; they have to lower the total cost of ownership (TCO) of the storage infrastructure while meeting necessary quality-of-service requirements, usually expressed as guarantees in a Service Level Agreement (SLA).

One way to reduce the amount of data being generated every week is to archive it³, thus removing it from the disk-based data that is being backed-up on a regular basis. However, if you are moving it from local spinning media to an archive, where do you put it?

One method drawing a lot of attention lately is to put it in “The Cloud”. However, there are a lot of considerations associated with cloud storage. First, generic cloud storage tends not to be cheap, especially if the cloud vendor is using disk media for the archive. The TCO issues for the cloud vendor are the same as those for each enterprise. Many cloud vendors use a pay-for-play model based upon the number of GBs being stored and the number of copies being maintained. These vendors are trying to make a profit, not to give you a break. Secondly, what happens to your archive if the third-party cloud vendor goes down? Can you even get a guaranteed SLA? You also have to consider the latency and reliability of cloud storage. Can you access your data if the archive is out of service? Has your data been cached for low-latency access? The IT news is filled with clouds going down for hours, or even days.

What happens to your business if the cloud starts raining on your parade? One thing that you can do is to avoid the small, vulnerable cloud and stick to a cloud driven by one of the IT giants, like

² On a total-cost-of-ownership basis, it is almost always more economical to replace disk arrays after three years, because the cost of continuing maintenance on an older array usually exceeds the cost of buying an equivalent amount of storage on new arrays (which utilize newer, denser disk drives).

³ See [The Clipper Group Navigator](http://www.clipper.com/research/TCG2010029.pdf) dated June 11, 2010, entitled *Simplified, Online Access to Archived Data – Turning History into an Active Archive*, available at <http://www.clipper.com/research/TCG2010029.pdf>.

Amazon, Google, or Microsoft. *If you already have done this, how has that been working for you?* Disruptions, some for days (for example, to Google's *Gmail*, Microsoft's *Hotmail*, and RIM's *Blackberry* email) certainly do not project a secure feeling and might portend the demise of your enterprise if you depended on one of them, or any third-party cloud. The IT staff needs to protect enterprise historical data from failure.

Keeping your data on a tape archive, whether local or remote, may be your best alternative (i.e., most reasonable – operationally and in terms of cost) as long as you can keep the complexity under control. Tape is an ideal medium for the archiving of historical data for long periods of time⁴, especially data that is inert and rarely accessed, and your data center can maintain access control. The mirroring and replication of data to disk does little good if that data is corrupted; it simply propagates the corruption.

With *LTO-5*⁵ tape libraries and the *Linear Tape File System (LTFS)*⁶, the IT staff can archive your historical data effectively and efficiently. Using an archiving appliance between the Tier-1 storage and the tape archive can eliminate the complexity and enable the data center to lower the TCO of the archive infrastructure. One company that is focusing on that exact appliance is OS Storage with its *Infinidisc Cloud Archive*⁷.

Who Is Open Source Storage?

Open Source Storage, Inc., (a.k.a. “OS Storage”) is a small company out of Littleton, Colorado, that started up in 2006 developing storage management solutions for large medical imaging applications. In 2008, they began support for broadcast files. Today, they have over 30 customers with almost 900TB of storage under management, primarily in the broadcast space.

⁴ See the issue of *Clipper Notes* dated December 20, 2010, entitled *In Search of the Long-Term Archiving Solution – Tape Delivers Significant TCO Advantages over Disk*, and available at <http://www.clipper.com/research/TCG2010054.pdf>.

⁵ For more on LTO-5, see *The Clipper Group Navigator* dated January 31, 2010, entitled *LTO Consortium Announces Next Gen Tape - LTO-5 Raises the Bar for Tier-3 Storage*, available at <http://www.clipper.com/research/TCG2010002.pdf>.

⁶ For more on LTFS, see *The Clipper Group Navigator* dated April 12, 2010, entitled *IBM Enables the Data Center with LTO-5 Products – Increasing Capacity, Throughput and File Management*, available at <http://www.clipper.com/research/TCG2010017.pdf>.

⁷ The *Infinidisc Cloud Archive* also has been called the *Infinidisc Never Full Cloud Archive* and the *Infinidisc Active Archive*. Since this can get confusing, *Infinidisc Cloud Archive* or just plain *Infinidisc* will be used in this paper to refer to the tape-based offering.

Exhibit 1 — The Infinidisc Cloud Archive Ecosystem

- **The Infinidisc Appliance** – This is a local NAS appliance that resides at the enterprise (or SMB) data center and is the secure gateway for the transmission of data to a local or remote archive. The Infinidisc Appliance gives the customers the flexibility of keeping entire or partial data sets in cache for high-speed restores while maintaining a geographically-separate copy of the archive data. Infinidisc Appliances expand to suit the needs of any business. In the event that the appliance is about to become full, the local IT staff can add additional cache dynamically.
- **The Infinidisc Cloud Archive Software** – This contains all of the tools to administer both local and remote archive components. Account management, Active Directory integration, and bandwidth allocations are just a few of the things that can be managed.
- **The Infinidisc Vault** – Locally-cached data is transmitted automatically and securely to OS Storage's secure vault (which sits in The Cloud) for redundancy. This is a *private* vault at a *securely-shared facility*. While the front-end of OS Storage's data center is shared, data is secured because it is encrypted for transit and storage. One enterprise's data is never intermixed with others when stored on tape. Full offsite replication is standard for maximum data protection and preservation. OS Storage's data center is modern and fully accredited and audited for complete SAS 70 Type II compliance.

Source: Open Source Storage

OS Storage's Infinidisc Cloud Archive is a local or remote active archive ecosystem that combines high-speed local appliances with transparent access to a customer's private tape archive (which is called a *vault*). (See Exhibit 1, above, for a description of the Infinidisc Cloud Archive ecosystem.)

Whether you have tens or hundreds of terabytes of data, the Infinidisc solution delivers pre-configured appliances and vaults with infinite capacity, while also being tailored to your specific needs. By utilizing OS Storage's *Never Full* technology and moving the data to tape dynamically, leaving only a stub, this archiving appliance

can never, ever, run out of space.

OS Storage strives to ensure that the enterprise IT organization can keep its most valuable digital assets safe and secure, with the latest technological innovations and a variety of appliances, while meeting enterprise requirements. In reality, small- and medium-sized businesses (SMBs) have the same archival needs as larger enterprises. In order to support all of them, OS Storage offers both entry-level and enterprise-class solutions, along with SMB consulting services specifically designed for smaller companies that do not require larger systems and large vault accounts. The Infinidisc solution delivers enterprise quality to all.

The Infinidisc Cloud Archive

The Infinidisc Cloud Archive represents an innovative approach to data protection and preservation. It has been designed to deliver to the enterprise (or SMB) the first practical cloud-based storage ecosystem that delivers performance and complete data protection at an affordable price for customers who are trying to protect *Really Big Data*. And, when OS Storage says *Big Data*, that is exactly what they mean, even hundreds of terabytes or more.

Their appliances combine high-speed, network-attached storage with automatic data replication to a local or remote private digital vault with data encrypted in flight and using tape as the ultimate storage platform. Their appliances use standard interfaces for access, including CIFS, NFS, and FTP. The Infinidisc solution will work with most operating environments using these protocols, including *Windows*, *UNIX*, *Linux*, *Macs*.

The Infinidisc Cloud Archive provides:

- Unlimited storage capacity
- Automatic vault synchronization
- Integrated backup/archive software⁸
- High-speed local NAS storage
- Secure 256-bit AES encryption
- Secure vault access from the web

There are two Infinidisc Appliance models, based on the business' capacity requirements.

This is not an appliance for family photos or your personal music collection, but a solution for economically preserving and accessing very large

collections of typically very large files. Some of their customers have in excess of 100TBs being managed by the Infinidisc Cloud Archive, with growth continuing.

Infinidisc Cloud Archive SMB Appliance

The *Infinidisc Cloud Archive SMB Appliance* has capacities from 500GB to 2TB of RAID-1 mirrored storage. It consists of two 3.5" SATA drives, with an integrated gigabit Ethernet connection that makes the SMB appliance virtually plug and play. The SMB appliance plays well for vault sizes ranging from 5TBs to 20TBs.

Infinidisc Enterprise Appliance

The *Infinidisc Enterprise Appliance* has been designed for capacities from 4TB up to 36TB per chassis, protected with RAID-6. It delivers high performance, utilizing the latest in server technology to create very fast and very scalable active archives. Collectively, a set of these appliances dynamically scale to hundreds of terabytes and can handle the largest data sets, via up to four Ethernet ports. The Infinidisc Enterprise Appliance can expand storage capacity dynamically and seamlessly via iSCSI to hundreds of terabytes. Designed from the ground up for the most complex environments, the Infinidisc Enterprise Appliance can handle the most demanding data loads. It supports both GigE and 10GigE networks, delivering the highest performance that is needed in an archive server.

Infinidisc Vault

The *Infinidisc Vault* sits in OS Storage's secure and redundant enterprise-class data center. It has a secure key card access and locked caged environment, redundant servers, power, air conditioning and backup diesel power generators. Data can be securely transmitted into one of their Infinidisc back-end servers (which receive the data from the local, frontend Infinidisc Archive Appliances (that sit in the enterprise's data center). Data integrity check sums are performed and files are securely migrated to two tapes in *your own* vault within the OS Storage data center for long-term, secure data retention and preservation. All vaults are monitored 24x7 to insure availability.

OS Storage's data center is a SAS 70 Type II certified⁹ with 24x7 on-site staff and 24x7 telephone access. The data center uses video and card key security and the environmental infrastructure

⁸ This is automated and means that none of the archived data needs to be backed up by traditional means. This results in a significant savings on back-up licensing costs, as well as TCO.

⁹ Statement on Auditing Standards (SAS) No. 70, Service Organizations, is a widely recognized auditing standard developed by the American Institute of Certified Public Accountants (AICPA).

is continuously monitored. The data center is connected to eight Internet backbones and is fully redundant. All data is kept in cache *and* copied onto tape from cache and stored in the tape library. A second copy of the data is copied onto tape and moved to a secondary (offsite) facility for safety. That is normally an OS Storage site, but if the business requests it, the second copy can be forwarded back to them for securing in their own facility, or sent to a third-party repository, like the ones provided by Iron Mountain.

One of the challenges in building a cloud storage vault is the expense of moving data into the cloud. For moving large volumes of data, bandwidth charges can become very expensive. To address this issue, OS Storage allows customers to either ship a standard NAS device (such as an external NAS drive) containing the data to them or the customer can choose to use the Infinidisc Appliance to ferry data from its present location to the OS Storage data center. The bulk load option is very cost effective when compared to the costs of moving data across the Internet. Once the data has reached the Infinidisc Vault, then the local appliance can be connected and file systems and directories can be accessed remotely (through the appliance).

Another challenge with using cloud storage is getting your data out when you want to change vendors or stop using the cloud. OS Storage offers a *bulk-out* service, where, upon request, a customer can ask for all their data to be returned. All customer tapes will be boxed and shipped back to the customer in industry standard format¹⁰ for use in any other tape-based archiving platform. Disk-based systems, which share physical devices among many users, simply can't offer as easy a termination process. Once data is on disk in the cloud, moving it out will follow the reverse process of putting it in, and bandwidth charges definitely will apply.

The Infinidisc Vault has also improved the methods that the enterprise can use to exchange content. Often, files are sent via email attachments, but large files are cumbersome and many email servers reject larger attachments. Certain industry segments, such as broadcast and health-care, often actually copy their data onto tape and physically transport the files to their affiliate, a cumbersome process, at best.

OS Storage has found a better way. In response to an idea presented by their customers, the Infinidisc solution makes it feasible for data centers and/or remote offices to permit trusted associates to download content securely without allowing access to the entire vault. OS Storage has developed a secure, password-protected, time-stamped download utility to facilitate this user-friendly process.

Pricing

Infinidisc Active Archive Vaults, including service, are sold affordably under annual contract, with prices starting at \$.10/GB/month for a 5TB vault, and going down from there with larger models reaching as low as \$.04/GB/month for a 200TB vault. The total annual cost for a redundant 50TB Infinidisc vault is \$55,490, including a 16TB Infinidisc Appliance and 10Mb/second of dedicated bandwidth. This equals about \$0.91 per GB/month, about two thirds the raw cost of storing data at Amazon without a local cache appliance or bandwidth charges.¹¹

Conclusion

The Infinidisc Cloud Archive delivers the ultimate level of access, performance, and protection at an affordable price. It provides peace of mind with regard to data preservation, with secondary local storage plus remote vaulting.

Systems crash. Disk drives fail. Viruses attack. Mother Nature has her way with us. It is imperative that each enterprise is prepared for any of these contingencies. Perhaps you should investigate the functionality provided by the Infinidisc Cloud Archive before one of these unwanted events visits your data center.



¹⁰ Currently, the export tapes are written in TAR format, with other formats to follow. Tools to read the files are readily available for every platform, as TAR is a common data format for files on tape.

¹¹ According to an example provided by OS Storage.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

David Reine is a Senior Contributing Analyst for The Clipper Group. Mr. Reine specializes in enterprise servers, storage, and software, strategic business solutions, and trends in open systems architectures. In 2002, he joined The Clipper Group after three decades in server and storage product marketing and program management for Groupe Bull, Zenith Data Systems, and Honeywell Information Systems. Mr. Reine earned a Bachelor of Arts degree from Tufts University, and an MBA from Northeastern University.

- ***Reach David Reine via e-mail at dave.reine@clipper.com or at 781-235-0085 Ext. 123. (Please dial “123” when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, The Clipper Group Explorer, The Clipper Group Observer, The Clipper Group Captain's Log, The Clipper Group Voyager, Clipper Notes, and “*clipper.com*” are trademarks of The Clipper Group, Inc., and the clipper ship drawings, “*Navigating Information Technology Horizons*”, and “*teraproductivity*” are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosures

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

After publication of a bulletin on *clipper.com*, The Clipper Group offers all vendors and users the opportunity to license its publications for a fee, since linking to Clipper's web pages, posting of Clipper documents on other's websites, and printing of hard-copy reprints is not allowed without payment of related fee(s). Less than half of our publications are licensed in this way. In addition, analysts regularly receive briefings from many vendors. Occasionally, Clipper analysts' travel and/or lodging expenses and/or conference fees have been subsidized by a vendor, in order to participate in briefings. The Clipper Group does not charge any professional fees to participate in these information-gathering events. In addition, some vendors sometime provide binders, USB drives containing presentations, and other conference-related paraphernalia to Clipper's analysts.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.