



Using a Storage Cloud for Archive



Should you use a storage cloud for your archive?

I. Summary

The cloud represents a major shift in technology and simultaneously in the cost associated with increasing storage capacity within a company. Instead of incurring significant CAPEX (capital expenditure) every few years the simple OPEX (operating expense) value represented by the cloud is intuitive. Additionally the cloud avoids the built in obsolescence of tape and the need to retain and maintain old storage hardware, software and media. The data reliability offered by an archival cloud is superior to tape and this reliability can even come at the same or lower price.

II. Overview of current choices

There are several vendors of cloud storage today. These vendors segment into two business areas namely consumer/small/medium business and enterprise. The companies targeting the lower end of the market include Amazon S3, Nirvanix, Mozy and Carbonite. S3 and Nirvanix do not offer proprietary software to utilize their services but rely on their customers to provide end-user applications and solutions. Mozy and Carbonite do provide complete end-to-end solutions allowing the more unsophisticated customer a relatively easy to use solution for their backup needs.

At the higher end of the market, the vendors include Seagate i365 and Iron Mountain's Digital offering. These higher-end solutions typically require the use of proprietary software supplied by these vendors – something that is far from ideal in the enterprise space where displacing existing tools is no small feat. The high-end solutions do offer greater security, reliability and regulatory compliance but at a higher price than the lower-end solutions.

New to the Cloud storage space is Vaultscape; Vaultscape provides an archival tape replacement cloud. Vaultscape has taken the enterprise approach to security, reliability and regulatory compliance while at the same time providing these attributes at a price competitive with tape. Vaultscape works with the existing suite of backup tools including NetBackup and CommVault meaning that enterprise customers do not have to change their workflow or tools in order to get immediate access to the Vaultscape storage archive.



III. Benefits of Cloud as Archive

IT organizations have several items to trade off when storage decisions are made. The first is Capital Expense or CAPEX – in other words how much is this solution going to cost. IT is typically provided an annual budget for hardware but even so spending decisions are monitored by the financial department to ensure the ROI (Return on Investment) of any new equipment meets the company's goals. New hardware purchases typically come with overhead. The first trade off is in personnel. Can this new storage be managed by our current staff? Is there special training required to operate and maintain? Next are the space and power requirements of this new storage: do we have the room for it in our current data center or collocation environment? Do we have sufficient power to support it? Lastly there is the life expectancy of the new storage; is it going to be reliable and supported for the life of the data to be stored on it?

When the Cloud is considered as an archival layer in the enterprise all of the above issues begin to disappear. First, there is no CAPEX requirement. The Cloud offers unlimited storage for a monthly use fee. In some instances the Cloud allows the enterprise to continue the use of existing applications and thus the training to use the Cloud is non-existent – just use your application as you always have. Next there is no space or power requirement for the Cloud which means that already cramped facilities do not need to be expanded, indeed these facilities could actually have space freed up by using the cloud. Finally, the Cloud storage is for life – the data storage technology behind the Cloud may change but these changes are carried out without interaction from the enterprise and done well will be completely transparent. The customer's files remain on line and are always available. Ultimately the Cloud is the obvious business choice for storage and especially for archive.

IV. Risks of Cloud as Archive

The real areas of risk are security, reliability and availability. Security is the paramount issue and few Cloud companies have adequately addressed this area. Most Clouds offer a public API and the customer will be moving data through the internet to and from the Cloud storage. Since the API is public, then it is immediately open to hacking and, since customer data is in transit to and from this API, then it is ultimately vulnerable to access. Of course a customer can use HTTPS which will improve security and they can also take advantage of data encryption prior to committing the data to the Cloud storage but their data is vulnerable in transit and within the Cloud storage itself if the hacker is particularly determined.

Reliability is the next issue to be addressed as a risk. Most Cloud storage is done via a single data center and the data stored on a single machine within that data center. If this is the case, then the data is vulnerable to volume damage either through a drive failure in a JBOD environment or multiple drive failures in the case of a RAID 5 or 6 volume. Should there be a physical problem at the data center, say fire or flood, then the data storage will likely be compromised. In the event of either a drive failure or a physical problem then all data will more than likely be lost.



Availability is the final risk when committing data to the cloud. The majority of the Cloud vendors have single points of failure in their systems – traffic is routed from a central core to outlying storage locations for data access. When the central core goes offline all data access is effectively blocked. File system corruption, which can happen in the database or in the normal transactional use of a file system, will in some cases corrupt the files or even make them permanently unavailable.

V. Cloud Archive v. Tape

Tape has long been the medium of choice for archive. The reasons are obvious: tape is cheap and tape is portable. The dirty little secret behind tape as all IT staff know is that it is not very reliable. Tape has also gone through several technology iterations over time. We have gone from large reel-to-reel systems, DDS, AIT, DAT, DLT and have arrived at LTO. Each tape format offered greater density and data reliability over the prior generation. Issues such as contact erasure, while they still exist, have been diminished by digital tapes. The evaporation of the lubricant from the tape surface over time both ages the drive head and wears out the tape – the older the tape, the more evaporation has occurred. As tape drive technology advanced and drives are replaced there needs to be a migration of data from the older format to the new format. Many companies choose not to perform this upgrade and retain the older drives and tapes. When the older tape does need to be read, there is risk it will not perform at all and, of course, the software that created it and the indexes to the data are lost. The tape index must therefore be recreated (assuming it can be) before the tape can be accessed. Given a large number of tapes, the data being sought may takes days or weeks to locate. Moving a tape from the machine that wrote it to another machine in order to read it may result in significant read errors; tape drives even today are notorious for the lack of interchangeability.

And the dirty little reliability secret: a Yankee Group/Sunbelt Software survey of 362 IT executives in March 2004 found 40.7 percent of respondents had been unable to recover data from tape in the last year as a result of tape unreliability. This is an important fact since, from the same report, close to a quarter of customers (23.5 percent) have had to recover data from tape systems 20 times or more in the last year, with an additional 33.5 percent recovering data 5 to 15 times during the same period, according to recent survey results. So, of these recovery attempts, 40.7% could not get the data back.

What is the advantage of tape besides cost? The data they contain can be safely moved offsite to a remote store. Of course the tapes are sometimes lost or stolen in transit (stolen for credit card data for example) or are lost or damaged at the remote storage location. Of course there is cost associated with tapes being offsite: the monthly fee for the storage and for the tapes that are effectively idle when offsite.



VI. **Cloud Archive v. VTL**

VTL (Virtual Tape Library) systems have become very popular in the last couple of years. The reason for their popularity is clear when the issues with tape are analyzed. VTL's have some tremendous values – they provide fairly reliable storage (consumer grade SATA with RAID 6 is common), they offer data online all the time, they offer power conservation when they spin down the drives and they typically support all existing applications. Where VTL's do not perform so well is partly covered in (II) above: they consume floor space in an already crowded data center, they consume power (albeit less due to their drive spin-down capabilities), they are expensive costing hundreds of thousands for a typical enterprise VTL, they age requiring maintenance contracts over time. Most VTL's offer compression typically through deduplication and deduplication has reliability issues of its own. Finally VTL's are onsite thus in a catastrophic situation the archive will be lost along with the systems it was intended to protect. Of course, multiple VTL's can be purchased and housed at different locations but this results in additional expense for the system and for its hosting.

VII. **Vaultscape Solution**

In order to provide a superior solution to tape, VTL or alternate cloud storage vendors, Vaultscape was thoughtfully designed with significant customer input.

We have four areas of focus: security, reliability, availability and low cost.

Security

Vaultscape has a private non-published API. Access to the API is only available via a contract with Vaultscape. Each user is issued a strong log-in code and two levels of password security; internally the system uses AES/TwoFish data encryption to access the storage servers, for example. Once a user opens a session via the API, they are issued a short-lived session token that is married to their IP address. Should the session token come from a different IP address than the one it was issued to, the session is immediately terminated. Session tokens expire every 24 hours or following some period of inactivity. Of course the session tokens expire when the session is formally terminated by the user. Vaultscape offers hardware VPN tunnel security to all our customers – this makes our servers as secure as if they were in the users own facility and, of course, the other security measures we deploy simply make the system even more robust. Only a small group of Vaultscape employees have access to the biometrically controlled data centers and then with electronic keys containing heavily encrypted randomly generated long passwords. Data within the storage systems is obfuscated providing no insight to the owner, file name or associated metadata.



● Reliability

Vaultscape has designed a self-healing file system that, even in the case of the catastrophic loss of the data index, will repair itself and allow data retrieval. The files uploaded by a user are stored at two geographically separate data centers where they reside autonomously from each other (more in Availability). Of course the storage systems are based on RAID 6 which inherently provides high reliability. The Vaultscape storage system has no write cache so when the API acknowledges the write is complete it means the data is safe on the disk surface and not in some limbo state as for most Cloud and physical storage environments. The storage design and implementation is so robust, Vaultscape offers a 100% Reliability SLA.

● Availability

Unlike the single point of failure storage systems available in the cloud, Vaultscape offers dual master data center design, the subject of a Vaultscape patent. These data locations, or Vaults, do not bit replicate the data as there are potential reliability issues for data in transit, for example. The methodology used to replicate the data between the dual master sites is unique and the subject of a Vaultscape patent. Since both sites have complete autonomy of each other one site can go down while the other site will simply continue. Vaultscape's design allows us to offer a 100% Availability SLA.

● Low Cost

The design of the Vaultscape file system, data center and storage all contribute to making Vaultscape the only true tape competitor when it comes to price. The underlying goal of the design was to achieve superior performance and reliability and yet to do so at a price as low as or lower than that of tape. When comparing Vaultscape to existing archival options we are several orders of magnitude less expensive.

For a Free Consultation about your archival needs, please contact us at:

(858) 217- 4848 or email info@Vaultscape.com or visit us on the web at <http://www.Vaultscape.com>

169 Saxony Road, Suite 114
Encinitas, CA 92024
858.217.4848 (Tel)
858.876.0783 (Fax)
info@vaultscape.com

