

What's The Best Way To Archive Data to the Cloud?

by George Crump



Cloud Storage is an ideal target for the dormant data that is clogging up primary storage systems. For organizations looking to archive their dormant data to the cloud, there are plenty of vendors offering a solution. But these solutions are not created equal, and selecting the right one can mean the difference between fully taking advantage of the cloud or just using the cloud as a digital dumping ground. †

Why Archive to the Cloud?

The concept of archiving has been around for as long as there have been data centers. Archiving generally serves two purposes. The first is to secure and preserve data critical to the organization to meet legal or regulatory retention requirements. The second is to slow the growth of data on primary storage. Most of that growth is in the form of unstructured data like images, audio, video, as well general purpose files. These data types are ideally managed by some sort of archive process.

The historical challenge with archiving, and why most organization don't do it, is where to put all the data. It is a generally accepted belief that somewhere between 80 to 95% of an organization's data has not been accessed in the last 90 days, and thus is a candidate for archiving.

The problem is that, at least on day one, this data is already stored on storage that has been bought and paid for. It doesn't make a lot of sense for the organization to buy an archive storage system that has capacity almost the same size as the primary data center's capacity, plus buy another system to store the DR copy.

A better practice is for the organization to buy only the amount of archive capacity that it needs to avert the next storage system expansion or upgrade. Unfortunately, most archive storage solutions are sold in very large initial storage capacity. A 100TB initial purchase is not uncommon.

Enter cloud storage. Because cloud storage can be bought incrementally, a terabyte at a time, the organization can archive to the cloud as needed to prevent having to buy a storage system upgrade.

How to Archive to the Cloud

There are several methods to archive to the cloud, the most common is some sort of NAS gateway. That allows a Windows or Linux server to “see” the cloud as a SMB or NFS mount point. Another is a file-system approach that overlays the storage infrastructure creating new pools of storage that may have a mix of on-premises NAS and Cloud Object Storage. Both of these solution types typically include some form of data management capability that automatically moves data based on policies IT creates. An example policy might be, “Move all data that hasn’t been accessed in the last 90 days to cloud storage.”

The Problem with Today’s Cloud Archive Solutions

A key problem with both archive solution types is they create a separate storage area that IT needs to maintain and monitor. Data has to be moved from the primary storage systems to these storage systems prior to the policy taking effect. Or IT needs to decide to move specific shares to the secondary storage system and have it managed there. All data moving to the archive, in this case cloud storage, has to be moved through this gateway which creates the bottleneck.

Another problem is that by definition these archive solutions present a SMB or NFS mount. That means they come with all the limitations of a network file system vs. a local file system, like NTFS or one of the Linux file systems. A network mountable file system typically provides the lowest common denominator of capabilities to increase compatibility with the various devices that will access it. That means, as part of the conversion many of the attributes of the file may be lost.

Transparent recall of archived data presents another problem. Most solutions use a stub file technology when data is moved from the on-premises storage to a local archive but few support stubbing to a cloud archive. These stub files will automatically recall the file when accessed by a user or application. The stub files are fragile but very important. If they are deleted or corrupted in any way the ability to retrieve data is severely compromised. Protection of these stub files is as important as protection of the actual data.

A final problem relates to the source of the growth of much of the unstructured data set, devices, most of which don’t natively store data to SMB or NFS. Many cameras, media imaging devices and Internet of Things devices write data to a NTFS file format and in most cases this data is first loaded onto a server running that same file system. That means, to be archived it needs to be copied over so it can be subsequently managed.

What IT Needs?

IT needs a simple solution that can archive data in place extending the current file system instead of replacing it. By extending the file system, the solution does not require a separate mount point or a manual movement of data. Installation is much more seamless and assures compatibility with the existing infrastructure. Most modern file systems have the ability to have these extensions put into them. Microsoft, for example, has a filter layer that enables an archiving tool to run in between the operating system and the file system in a fully supported manner.

If the data management intelligence is put into the filter layer each server in the data center storage data has direct access to the cloud. This direct to cloud access also means performance now scales in lockstep with file server growth, essentially eliminating the potential for a bottleneck.

The file system extension approach also lowers the cost of the solution, enabling the organization to realize a quicker return on its investment. Not only does the organization save the cost of the gateway appliance, it does not have to worry about multiple copies of data. The file system becomes the sole authority.

The extension of the file system also means the data written on cloud storage is in the same format as the data written locally. In the case of Windows, it is NTFS throughout. In the case of Linux it is NFS throughout. Extending the file system means unique file system attributes and security settings are preserved.

The file system extension approach also increases compatibility with other process that might run on the server. For example, a virus scan run on a non-NTFS file system might accidentally recall all the files in

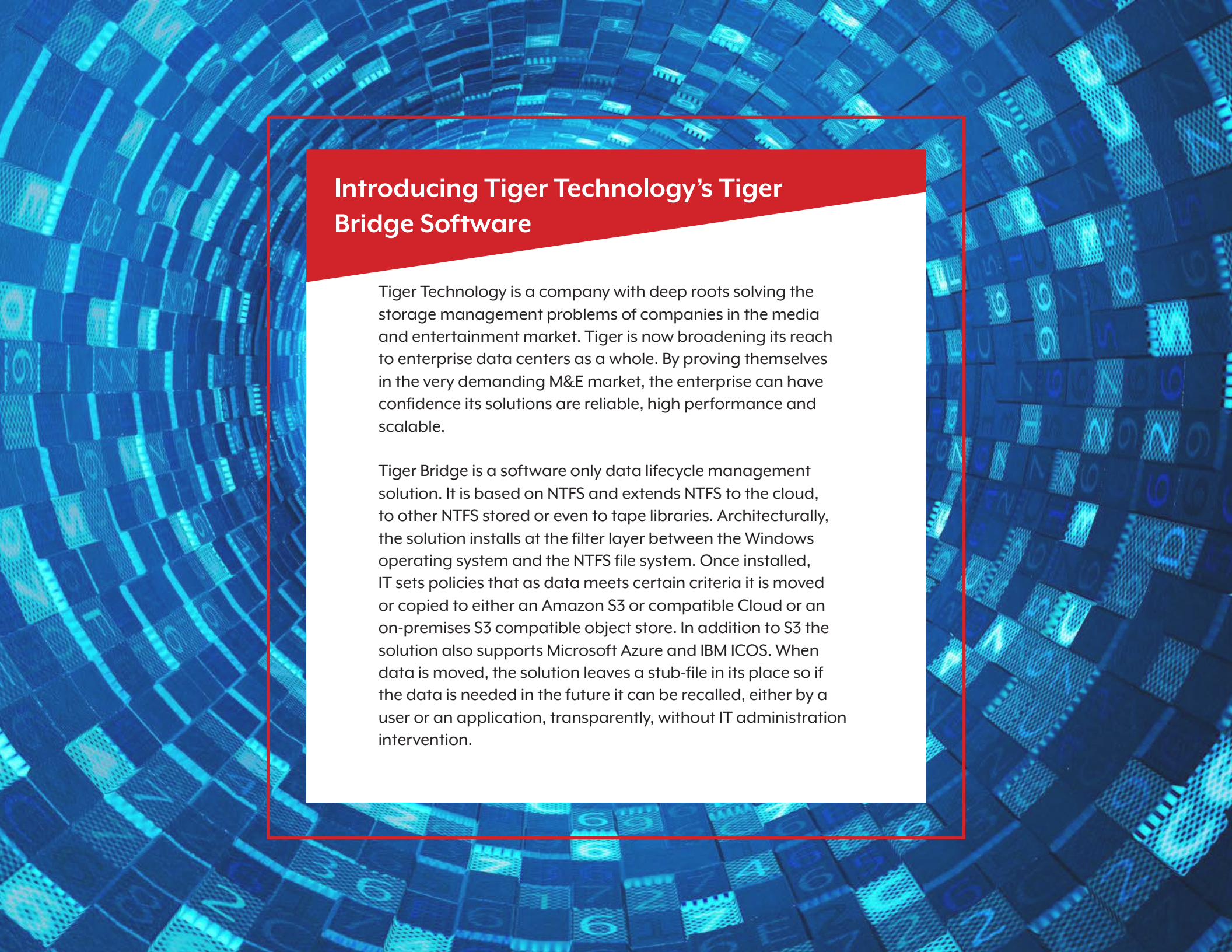
the archive as each stub file is examined. With an NTFS extension, the two software solutions should work together.

IT also needs a solution that is not myopic. Some organizations will not want to archive data to the cloud or they may want to use a combination of archive targets, creating a hybrid archive. The solution should support a variety of archive storage options including direct attached storage (DAS), network attached storage (NAS), block storage (SAN), and even tape media in a tape library.

A final need to be met is the need to aggressively protect stub files. They are the weak link in most approaches. While some archive solutions make a redundant copy of the stub data, that copy is almost always made locally and it is up to the IT professional to make sure that the metadata is moved off-site. Ideally the archive solution should make sure a copy of the stub data follows the data being archived, in this case a copy should be made in the cloud.

A cloud copy of metadata protects the organization from accidental deletion of stub file data and it also provides a powerful disaster recovery capability. In the event of a disaster, the process of recovering a server after the operating system has been installed is to simply copy down the stub files from the cloud. The new file server is ready within minutes. Then, data is retrieved as users begin to access it or IT can programmatically bring back the most active data.

“IT needs a simple solution that can archive data in place extending the current file system instead of replacing it.”



Introducing Tiger Technology's Tiger Bridge Software

Tiger Technology is a company with deep roots solving the storage management problems of companies in the media and entertainment market. Tiger is now broadening its reach to enterprise data centers as a whole. By proving themselves in the very demanding M&E market, the enterprise can have confidence its solutions are reliable, high performance and scalable.

Tiger Bridge is a software only data lifecycle management solution. It is based on NTFS and extends NTFS to the cloud, to other NTFS stored or even to tape libraries. Architecturally, the solution installs at the filter layer between the Windows operating system and the NTFS file system. Once installed, IT sets policies that as data meets certain criteria it is moved or copied to either an Amazon S3 or compatible Cloud or an on-premises S3 compatible object store. In addition to S3 the solution also supports Microsoft Azure and IBM ICOS. When data is moved, the solution leaves a stub-file in its place so if the data is needed in the future it can be recalled, either by a user or an application, transparently, without IT administration intervention.



An Extension, Not Another Silo

Tiger Bridge meets the first requirement of a modern archive solution. It does not create yet another mount point that IT needs to migrate data. With Tiger Bridge nothing needs to be mounted or provisioned. Instead, it integrates directly into the NTFS file system, extending it into the cloud or to tape, which simplifies IT's job significantly.

SOLVING THE TRANSFER BOTTLENECK

Another advantage of the extension approach is each file server can communicate directly to the cloud. While this approach does lead to more cloud storage pools it does provide scale as more file servers have the Tiger Bridge installed on them. The organization has the flexibility to decide if it wants the simplicity of a single pool of cloud storage or the scale of multiple servers writing simultaneously to the cloud.

SMALLER FAILURE DOMAIN

With the extension approach the “failure domain” is smaller. The gateway approach has a single point of cloud entry which means a failure at the appliance means all communication with the cloud stops until IT replaces the gateway. To protect against a gateway failure these solutions strongly suggest creating a cluster of gateways, which of course raises the cost.

Since each file server is communicating directly to the cloud if that server were to fail for some reason, only the data on that server would become inaccessible. The solution is an extension of Windows. That means protection from failure only requires the same protection that a typical Windows server would, which is something IT should

be very familiar with. But that protection level can be implemented on a server-by-server basis depending on its criticality to the organization.

STUB FILE PROTECTION

Since Tiger Bridge is stub file based, protecting these stub files is critical. Tiger does an excellent job here. All metadata is copied to the cloud, as well as being stored on-premises. On-premises access reduces response time and to the user makes it feel like the data is still “there.”

Metadata in the cloud means if there is a failure at the primary site, including complete site loss, a secondary site can be stood up quickly. All that requires is installing a new server with the Tiger Bridge software and copying down the metadata from the cloud. That should take only a few minutes to complete. From that point users can start accessing data as needed because it appears as if it is all there. When a file is accessed it is retrieved from the cloud. Essentially the software creates an automatic prioritized restore.

USE CASES

There are several use cases for Tiger Bridge. The easiest first step is to use the product as a solution to the unstructured data backup problem. Most filesystems today have millions, and in some cases billions, of files that can take a toll on traditional backup solutions. To get around that many modern backup solutions leverage image-based backups, but these solutions struggle managing the individual files.

Tiger Bridge allows a file by file “backup,” which essentially copies the files to the cloud but it removes the file from on-premises storage. Later, as the organization gains confidence in the solution, it can slowly start implementing policies like “remove and stub all files not accessed in the last three years”. Gradually, as confidence grows, the timeframe can be shortened until production storage is only being used for active data.

STORAGE EXTENSION

The solution can also be used to extend storage beyond the capacity of the current file server user data is on. Instead of expanding the storage system or worse buying a new server, simply install the Tiger Bridge software and move the oldest data to the cloud, creating room for new data.

VIDEO SURVEILLANCE

Most video surveillance architectures typically have several file servers getting feed data from cameras. In the past, as these systems reached capacity data was deleted from them. However, legal and corporate requirements are suggesting organizations need to retain this type of data for a significantly longer period of time.

To meet the longer retention requirements many organizations are manually copying this data to the cloud or an object storage system on a periodic basis. If they want automation, they need to copy the surveillance data to a NAS gateway, which will then move it to object storage based on policy.

The Tiger Bridge solution allows the management and moving of data directly from the Windows File Server that is originally capturing all the data. The result is a simpler data flow and a much more cost-effective architecture.



CONCLUSION

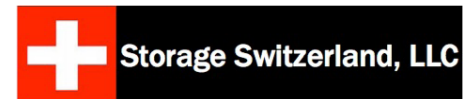
There is no question data is growing, and much of that growth is unstructured data on Windows file servers. As a result, archive and data management solutions are constantly being introduced to the market. Tiger Technology's solution extends NTFS seamlessly to the cloud and provides transparent access to migrated data while protecting the stub files it leaves behind. Another advantage Tiger Technology has is that it has vetted its solutions in the M&E market for years. That means it approaches the enterprise as a seasoned veteran; not a startup with only a few customers.

THE ANALYST

George Crump is the founder of Storage Switzerland, the leading storage analyst focused on the subjects of big data, solid state storage, virtualization, cloud computing and data protection. He is widely recognized for his articles, white papers, and videos on such current approaches as all-flash arrays, deduplication, SSDs, software-defined storage, backup appliances, and storage networking. He has over 25 years of experience designing storage solutions for data centers across the U.S.



THE FIRM



Storage Switzerland is the leading storage analyst firm focused on the emerging storage categories of memory-based storage (Flash), Big Data, virtualization, and cloud computing. The firm is widely recognized for its blogs, white papers and videos on current approaches such as all-flash arrays, deduplication, SSD's, software-defined storage, backup appliances and storage networking. The name "Storage Switzerland" indicates a pledge to provide neutral analysis of the storage marketplace, rather than focusing on a single vendor approach.

OUR PARTNER



Tiger Technology designs, markets, and supports world-class innovative storage workflow solutions that maximize capability, performance and value, for rich media, video surveillance and enterprise applications. These markets, which have demanding requirements for workflow flexibility, scalability, reliability and cost-effectiveness, value our field-based expertise in high performance network storage and storage virtualization, and have widely adopted our technologies. For further information on Tiger Technology please visit www.tiger-technology.com.