Sullivan|Strickler

The Future of Legacy Data

2020 WHITE PAPER

What is legacy data?

The definition of "legacy data" can vary depending on the audience. In its broadest form, legacy data is anything that has been created and written out for storage. Legacy certainly infers that the data is no longer actively used, or commonly accessed, but remains vital in some way.

The Business Directory describes it as "information that is stored in an old or obsolete format." As office IT systems are updated, the information stored on backups or archives is often left gathering dust.

All businesses create data as part of their operations, from emails and databases, to reports and spreadsheets. Data that is no longer actively used or refreshed is then archived to make space for new files. This data is often kept for compliance reasons, lawsuits and business or regulatory audits, or for future business use.

The advantages of tape storage

For many years, tape has been the most common medium used for the backup of legacy data because it offers a relatively low-cost and long-lasting solution. This is especially true for companies that are required to hold on to large volumes of data.

Tape storage is ideal for backups due to data being stored sequentially, not allowing existing backups to be altered or manipulated. Whether the data on a tape was written last week or 10 years ago, it is all classified as legacy data because it's not in active use.

SullivanStrickler specializes in managing legacy data for its clients. Whatever data is written to the storage medium, we can access it and make it available and usable to our customers, says Shawn Strickler, CTO and co-founder of SullivanStrickler –

"The farther you get from active use, the more complex it can be to access your data. We support everything from recently created data to things that have been archived and put on a shelf and forgotten about for years," says Strickler.

"Even when a company retires an old backup environment, they may still need to have access to that data," says Brendan Sullivan, CEO and co-founder of SullivanStrickler. "They no longer need all the functions and features of the applications for which they are paying, so to hold on to those systems and not know when they will need to access the data, is an unnecessary expense."

Do you trust the cloud?

Magnetic tape as a form of data storage has been around since the 1950s, and it remains the number one option for archiving. From the floppy disk to the cloud, there have been solutions that offer faster and more readily accessible solutions. However, none have managed to dethrone tape as the best option for legacy data.

"There was a big push to the cloud and that seemed to be all anybody was talking about for a couple of years, but you don't hear as much excitement about it anymore," says Sullivan.

"We've even seen instances of people moving off tape onto the cloud, then going back to tape. They've done this because they've discovered there a need for both."

Storing data on tape is also much more secure than it is on the cloud. Cybercrime has become a serious issue in recent years, especially when it comes to sensitive company data – **"No one ever hacked a tape,"** says Sullivan.

The cost of storing your data

There are situations where having data quickly available is important, whether it is stored locally in the organization or on the cloud. However, companies are often under obligation to retain data for tax and regulatory reasons, and over time this amount of data soon adds up.

Take, for instance, a mortgage company that provides 30-year mortgages. It needs to retain information on these mortgages for long periods of time. Storing this on an active disk or on the cloud is fine in the short term. However, as the years pass, the costs of doing this for thousands of accounts per year can soon escalate.

Storing a file in the cloud might cost fractions of a dollar on a monthly basis. But if you extend that over the course of multiple years, and for millions or tens of millions of files, as most companies have, then the costs really begin to compound. The latest backup tapes can store up to 30 Terabytes of data and can be taken out and held off site for as little as 25 cents per month.

Where is it really stored?

The industry secret that many are unaware of, is that legacy data storage in the major cloud platforms is being written out to tape. "The reason that services such as Amazon's S3 Glacier Deep Archive has a turnaround time that is a day or days is because they are having to go and find tapes that contain your data, mount those tapes and restore your data, to make it available to you," says Strickler, "We're storing the tapes for you in our high availability vault. **We'll store the tapes for you and when you have a request, we'll pull it and restore it, oftentimes in the same day."**

The reason that many companies are coming back from the cloud is that the amount of space they are consuming is significant and continues to grow. Given the avenue to store more data, most people will do it.

Running out of space

In the early days of network computing, when storage was more expensive, each employee was given an individual folder on a file share, often limited to just 500MB in size. When the mailbox was full, the solution was to delete files or move them from the server to a .PST on your computer. Employees then started copying those .PSTs to their share folders. The result was emails that the company no longer controlled through the mail server. It could no longer expire data and delete data sitting on file shares.

"Companies project, through the normal course of business, a percentage of incremental storage growth. But what they were finding is they were exceeding that by 500 percent," says Strickler.

"This was never factored into their cost calculations and now, all of a sudden, they have to force everybody to go through and delete all their stuff out of the cloud or have to just suck it up and find another solution."

Today, companies have become more intelligent about what they are sending into the cloud, and are starting to put more restrictions on what they're saving to it.

Future-proofing your storage

For many years, disk manufacturers have marketed to the world that hard disk storage was less expensive than tape. However, study after study has proven that this is not the case for legacy backup.

Hard disk drives also date back to the 1950s and do offer advantages for secondary storage, such as the ability to rewrite individual files and a faster access speed. However, the cost of tape, especially when scaled up over hundreds of terabytes, can be much less.

Tape manufacturers such as Fujifilm and Sony are now really trying to educate the market about the total cost of ownership on tape. But this cost saving isn't purely due to the physical cost of tape cartridges. It is also due to its longevity.

A report by Fujifilm on LTO tape media gives an archival life of over 30 years when stored in optimum conditions (such as a data storage center). More conservative estimates say between 15 and 30 years for tape storage, but it could be longer still.

With hard disk drives however, the average life expectancy can be as little as three to five years. A study of over 25,000 disk drives by Backblaze found that drives do start wearing out after three years, though of course this depends on how often they are spinning.

What this means on a practical level, is that you would have to replace disk drives at least three times as often as tape media. And with tape, you are only replacing the cartridge, not the drive itself, so replacement costs are much less.

Storage without moving parts

Solid-state drives (SSD don't have the same issues of moving parts that can fail, as with hard disks. However, they do slow down due to the constant rewriting of data and that eventually makes them unusable. The expected lifespan of an SSD is around 10 years but the cost is also considerably higher.

With these lifespans known, companies can plan for the depreciation and replacement costs. However, when drives fail early, it can cause problems. And if it happens enough, a company may rethink their solution.

The cloud offers a potentially infinite longevity, thanks to the way the data is held and mirrored. For this reason, backup software manufacturers were initially very keen to jump on the bandwagon of cloud storage. While many provided cloud integration into their platforms, in recent years this enthusiasm has waned. Presumably due to a lack of interest from enterprise clients.

The one exception is in email. Business solutions from the likes of Google and Office 365 have successfully provided cloud-based email management complete with archiving and compliance solutions.

Fast access to your data

Cloud accounts, barring solutions such as Amazon's Glacier Deep Archive, function similarly to local disk. You are logging in and actually viewing your data on a disk somewhere -- It just happens to be the at the cloud provider while seeming to be local.

What you are actually viewing in your browser is the metadata of your storage; the file names, the types of files, their size and when they were created and last modified.

The premise behind SullivanStrickler's proprietary metdata review tool, Invenire, is to show the metadata associated with all of your stored data. Whether it is stored in the cloud, on tape, on a live server or a forensic image of a drive.

A universal translator of backups

Different storage media and backup systems use different formats and interfaces for archiving. Backups that were created with one system normally require that software to recover it. This means keeping a live version of the software, and sometimes a subscription. SullivanStrickler offers a common interface to work with all archived backups.

"What we've done is we've created a kind of universal translator," says Strickler. **"We read the data from the tapes, regardless of the format that created it, and restore the data out as the actual application would."**

SullivanStrickler's universal translator is a software application called TRACS, which stands for Tape Restoration And Cataloging Software. This is a truly unique product that outclasses the standard solution in the market. At the front end is Invenire, a web-based portal that allows you to search your metadata from all of your archives, no matter how they were created or where they are hosted.

TRACS has the ability to make disk-based copies of tapes and place them into a container file. There are three main container formats available. The first is a tape media file, or TMF.

The content of a tape can be copied into a TMF container file, which can then be cataloged. This file can be restored back to tape at any time. However, the format offers no additional insight.

The two other formats are a tape duplicate file, or TDF, and a tape session file, or TSF. These both also benefit from being able to catalog the metadata within the container.

"This means that you can see the file names, extensions and sizes, as well as the created, accessed and modified dates for those files, as you would in a file browser," says Strickler.

With a TDF, in addition to copying the tape to a disk file, a catalog of the files located on the tape and the backup session information are appended to the end of the file. The backup session metadata allows you to identify where to find that file. It gives the exact file mark, block and offset, so you can jump straight to it and restore that file without having to restore everything that comes before it on the tape.

A TSF file allows all of the above with the added ability to consolidate the data, as well as the defensible deletion of data from within the container, which is very compelling.

Stacking tapes

Over time, the capacities of data storage tapes have vastly increased. The first generation of LTO tapes had a native storage capacity of 100GB. Today, the latest LTO-8 tapes have a native capacity of 12TB, or 30TB with compression. This means you could fit at least 120 of the original tapes on to one single new tape. And that's before compression or paring the data down.

This process is known as stacking and for companies with thousands or hundreds of thousands of tapes in archive, a significant reduction in storage. **Data that previously filled 100,000 tapes will now fit on as few as 500 of the latest versions.** This not only takes up a lot less space, but the annualized storage savings can have a rapid ROI and save significant amounts of money going forward.

Another benefit of migrating old data on to new tapes is that it resets the clock in terms of the lifespan of that data. LTO-1 tapes were first released in 2000, so even data on these cartridges could be up to 20 years old. Consolidating to new tapes not only reduces the physical footprint, but also means your storage is safe for another 30 years.

"Before LTO-7 it really didn't make a lot of sense to do stacking, as the cost benefit really wasn't there. But now we are in an era where IT managers want to spend their dollars on things that are going to help them going forward," says Strickler. "When they are paying 25 cents per tape per month, getting rid of 99,000 tapes makes sense. And they're creating new tapes every year, so the volume of tapes being stored is going up. The argument for stacking is extremely compelling."

Selectively retaining data

SullivanStrickler's Tape Session File (TSF is the most interesting container format for companies. While the volume reduction of physical media when using the TDF format is compelling for monetary reasons, the TSF also allows companies to select which data is maintained, which can also be a benefit from a legal standpoint.

In recent years, company legal teams have had a greater input into backup operations, dictating policy on what gets backed up and for how long. Companies have realized that there is risk involved in holding data that they don't need to hold.

"There's case study after case study in the legal market where a company held on to something they didn't need to, and they later got sued," says Strickler. "Had it been deleted, things would have potentially been very different in the court, but because data was retained, a smoking gun might have been there."

While most companies want to retain email backups, other servers with backups might not need to be retained. Department file shares from Accounting or HR may need to be kept, but others might not. An old marketing database or a print server from five years ago may no longer be useful. SullivanStrickler's TSF technology allows customers to choose which data to retain and which to delete during the stacking process, reducing risk AND the amount of legacy data retained.

Pick and mix your files

Tape session files give the user the ability to decide which backups to retain and which to delete on a session-by-session basis. As each session is written out to its own file, something that is kept now can be easily deleted at a later date.

"One analogy could be your cable TV channels," says Sullivan. "You paid for 100 channels but now you only want ESPN, and they won't let you climb down from the bundled contract. What we provide the client is an alternative way to use that limited feature, or to pay for just that one channel, as it were. The customer can sever the costs that they've been incurring for the last 10 years on that environment."

Legacy data is no longer restricted to just a collection of tapes in a storeroom. These TSF files can be written out to disk, the cloud or back to tape using open standard backup software. It allows users to mount a tape backup as a drive and read or write data to it, as you would an external hard drive.

"Our customers don't need to pay maintenance and support fees for software that they haven't used in five years, just in case they need to restore data from it years from now," says Strickler.

The ability to pare down data by selectively retaining only certain backup sessions has proven to have both a cost and a legal benefit. **Using TSF files, a company can reduce its footprint by 99 percent or more** -- from 100,000 tapes to fewer than 500, or may-be as little as 200 tapes if it is really being selective on what it keeps. And, if required, those files can be reduced further at a later date.

In the future, TSF files will give customers the ability to selectively delete data from within the TSF file. This will help with adhering to GDPR or the California Consumer Protection Act, where individuals have the right to be forgotten, or companies no longer have the right to hold the data.
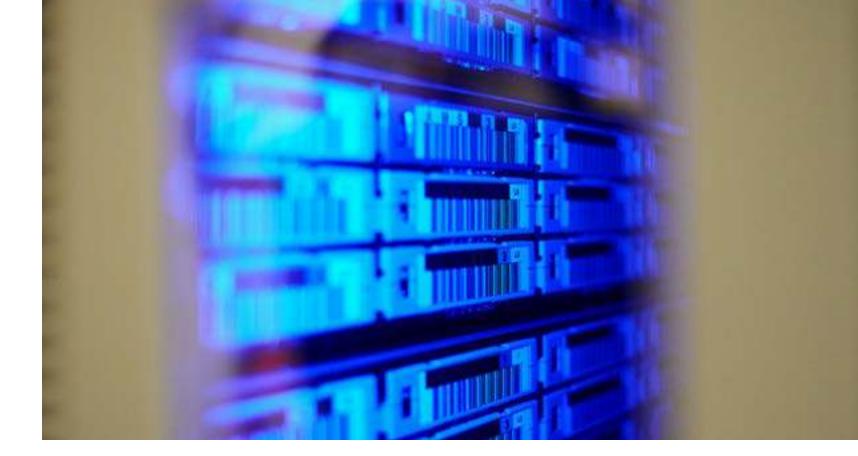
The only historical data remediation solution for data on tape is to destroy the tape entirely or restore all the tape data, delete the desired data, and then back it up again, which is a very long and potentially expensive process.

Disrupting the market

SullivanStrickler's solution for legacy data is a market disruptor in the sense that nothing else like it exists in the market.

The tape restoration and cataloging service takes the solid and reliable system of tape backup and optimizes it, making it more efficient and more accessible. Its biggest competition is perhaps the companies that retain their current more costly solutions.

"Since introducing this system, we have yet to present a price to a customer that has a total return on investment longer than two years, and often times it is less than one,"

says Strickler. "It's a very compelling story and they know it can save them money. The question is, is the status quo so painful that it's worth considering? More and more companies are saying 'Yes'."

With every new generation of tape, the capacity is doubling. Speculative capacities for future releases have the potential of 24TB for LTO-9 and 192TB by LTO-12.

"This means that our customers' ROI will increase with each subsequent generation of tape," Strickler adds.

"No other system can offer this hybrid solution, and this is just the beginning. New features are being introduced to the Invenire system very soon that will benefit not only the backup team but also the enterprise user," says Strickler.

How it all happened

SullivanStrickler was formed in 2013 by Brendan Sullivan and Shawn Strickler. The pair had worked together in the past on related projects but reunited to create groundbreaking backup and archive restoration solutions.

"The real objective of the company is to manage legacy data for clients and perform a range of services on that data," says Sullivan.

"Our skill set is getting hold of data, irrespective of what hardware or software was used to create it and making it available for the client to use it for whatever they need," he adds.

For more information, please visit: www.sullivanstrickler.com