

Arkivum LTFS Whitepaper

Whitepaper



LTFS Data Tape helps Arkivum make the difference

With an ever increasing need for organisations to retain data and the associated demands this places on equipment, power, cooling and people, it's not surprising many organisations are simply failing to cope.

At Arkivum we understand that data is valuable which is why we have created a unique data archiving service based on data tape to ensure that data is kept safe and secure – forever. Unlike any other solution in the market today, Arkivum is currently the only data archiving service to provide a 100% data integrity guarantee.

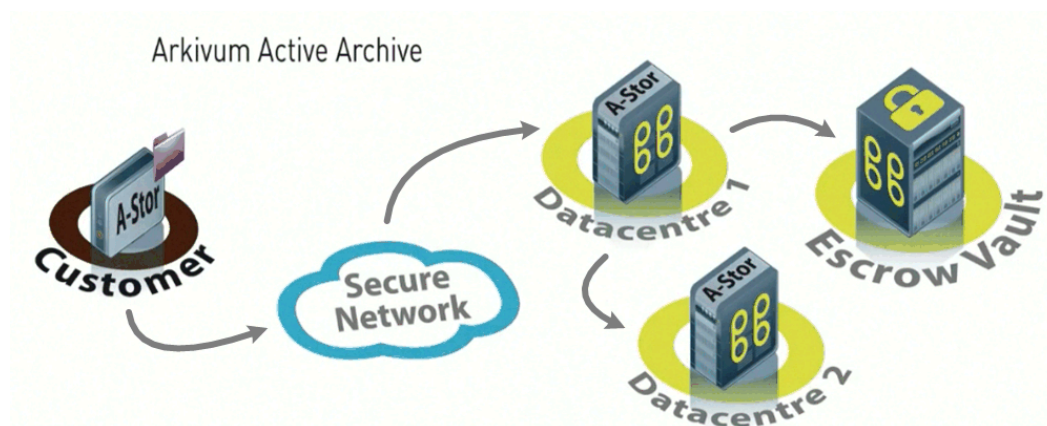
LTO data tape and LTFS form an essential part of our service. There are many advantages to data tape for archive applications, including its low cost, high reliability and very low energy consumption. These all make data tape ideal for us when providing our extremely cost effective service with fast online access to data.

LTFS makes it easy for us to create three independent and geographically separate copies of every file stored in our service, something that forms the bedrock of our data safety guarantee. All three copies are stored on data tape, with two copies always online inside tape libraries and housed in secure data centres.

The third copy uses LTFS as a key part of our 'data escrow' strategy. A copy of all customer data is stored safely and securely offline with a third-party escrow provider. Our customers have complete peace of mind that they can always retrieve their assets on data tape from the escrow provider no matter what happens to our service or business – and they can read those tapes using industry standard technology and open source tools.

So how does the service work and where does data tape fit in?

The Arkivum service all starts at the customer site with an appliance, physical or virtual, that provides a gateway to the archive service. The appliance looks like a file-system to the user, so adding files is just a case of copying them in like any other file-system. Any files added are immediately encrypted and always before they leave the user's network.



Each file is then replicated over a secured network to our first data centre. Here it is written to LTO data tape and stored in an online library for quick access. The file is then replicated to our second data centre for resilience. Every time we replicate a file we use checksums to make sure it's been safely received and stored at the destination.

Finally we create a third copy that is stored offline with an independent third-party under an escrow agreement. We use LTFS and provide open source tools for easy retrieval of data from these tapes. The user has complete peace of mind that they are in no way locked to our service. When all three copies are verified as complete then the user can choose for their local copy to be removed.

But we don't sit back and relax at this point! Arkivum will regularly migrate the data tapes to ensure data is on fresh media. We regularly check the integrity of the data to ensure all three copies are in good condition.

And throughout all of this, file access is easy - the file-system at the user site still shows every file with its original name and location - so access is just a case of copying them back out again! In essence, we've taken data tape into the cloud to provide an extremely reliable distributed storage service which we then virtualise as single file-system to make ingest and access really easy for the our users.

The benefits of LTFS

Both Arkivum and our customers benefit directly from LTFS.

For our customers, LTFS is piece-of-mind that should they ever need direct access to their data on tape then they can read that data easily using nothing more than open source tools and an LTO drive they source from whatever vendor they choose. Normally, access to data in our service is through the appliance on the customer's network. The appliance pulls the data off tape from one of the data centres we use and delivers it through the file system the appliance exposes on the local network. The magic here is that this completely hides tape as far as the customer is concerned – they just see a network file-system like any other.



There are times when other routes to the data are needed. This might be because a customer asks us to create a tape copy of their data so they can keep it on-site for DR, or it might be because the customer wants a bulk restore from the archive and shipping data tapes is faster and more cost effective than downloading data online, or it might be because the customer wants to get their data back from escrow so they can move data from our service to another.

It might seem strange for us to talk about how customers can move away from our service, but the reality is for long-term data retention applications in a fast moving world that one of the first questions you have to ask yourself is 'how do I get out of an archive system or service' and not just 'how do I start using it in the first place'. The combination of LTO data tape and LTFS is perfect for this. Try asking the same question to a cloud storage provider that uses spinning disk!

For Arkivum, LTFS makes data management a lot easier. To use tapes we simply mount them using LTFS and read/write files without having to worry about container formats such as tar or tracking in detail the whereabouts on a data tape of every bit of data we store. We just need to know what tape a customer's file is on and LTFS takes care of the rest. When we need to migrate data tapes so we can take advantage of new generations and ensure our media is up-to-date, then moving data from an old tape to a new tape is effectively just a 'copy and paste' operation. When we need to retrieve a file to check its integrity (which we do on a regular basis if the file isn't accessed by the customer), we can do this easily by requesting the specific file directly from tape using LTFS. Finally, on the extremely rare occasion that we ever need to recover from a catastrophic failure in our service then we know we can easily 'rebuild' all our internal indexes and databases from the data tapes we hold by using the LTFS index.

But perhaps most important to both us and our customers is that LTO data tape and LTFS has a clear and long roadmap ahead. Data tape capacities and performance will continue to improve year on year for at least a decade or more to come, something that is becoming a challenge for spinning disk. Industry adoption is already widespread and is increasing, which gives confidence that investment is unlikely to be in a 'flash in the pan' technology. The use of open standards and support for these from a range of vendors gives reassurance against vendor lock-in. And the backwards compatibility built into the roadmap between LTO generations means that obsolescence is far less of a concern than with many other technologies – especially ones that are proprietary.

LTO, LTFS, and seven questions to ask about any archive solution

Our service is unique in offering a 100% data integrity guarantee and sets us apart from the competition. Our guarantee of data safety isn't just about using the right technology, it's about having a highly skilled team of people with a wealth of expertise in long term data retention and data storage. When you use our service, you buy into this team and the in-depth data assurance processes we have developed.

If you're thinking about archiving we'd invite you to consider the following seven questions before

Title
Arkivum LTFS Whitepaper
Part no
ARK/MKTG/ALL/171
Version
1.0
Date
Sept 2013
Status
Release



making any decisions. If you want to know how Arkivum can say 'yes' to each and every one of these then head on over to our website for more information. If you want to know the part that LTO and LTFS has to play, then please read on.

Q1. Can you be sure of the integrity of your data?

Are you guaranteed that every single bit of data you pull back from your archive is identical to what you put in - and can you prove it?

Q2. Is your data always confidential?

Can you be sure that there is no way for anyone else to access and understand your data - and that includes any service providers used to hold the data?

Q3. Do you know the authenticity of your data?

Do you know who first put the data into the archive, who's then accessed that data, and who has handled the data inside the archive to keep it safe and secure?

Q4. Is your data available when you need it?

If you need data quickly then can it be retrieved in a matter of minutes? Can this be done for every single file you want to store?

Q5. Do you know who is in possession of your data at all times?

Do you know exactly where your data is stored and who runs the facilities? Do you know who controls how your data is stored and who is responsible for its safekeeping?

Q6. Are you confident that you can always use your data?

What happens if your vendor or service provider goes out of business or stops support? Are you locked into proprietary data formats or tools? What happens if you want to change applications that use your archive - do you need to go through a complex systems integration process?

Q7. Can you always assert the provenance of your data?

Faced with a compliance audit, the need to understand whether you're litigation ready, or the need to prove the authenticity of your data, can you provide an audit trail that says what the data is, where it's come from, and everyone and everything that's happened to it since?

The first three questions (confidentiality, integrity, authentication) are the bedrock of information security and are known as the CIA triad. The next three questions (availability, possession, utility) are further essential facets of risk management for information assets. The final question (provenance) is a pillar of digital preservation, but is probably more familiar in a business context as being able to support auditability – it's about asserting a 'chain of custody' for data assets. LTO data tape and LTFS have a role to play in answering all these questions. Data tape in tape libraries within online data centres, plus the escrow copy, allows us to deliver against the availability criteria. We store each customer's data on a dedicated set of tapes in specific secure data centre locations so you know exactly where your data is at all times and can be confident that no-one else has access.



Perhaps more interestingly, the ability of LTFS to describe what is on a tape and allow it to be seen as a file-system makes it a lot easier to create 'self-describing' media within an archive where each tape contains extra information (either as part of the LTFS index or as separate files) to provide a complete record of what is on the tape, how it got there and how to use it. For example, we use this feature for our escrow tapes so we can store extra information including checksums for verifying data integrity, information about how files have been encrypted, information on the origins of the files (where they came from, who provided them, when they were provided, and the relationship to other files) and information of the history of files e.g. migrations and integrity checks. This provides an audit trail that gets stored right alongside the data so this information is never lost.

There is of course a whole lot more to assured archiving than that, but LTO and LTFS is certainly a help along the way!

If you would like to know more, please visit www.arkivum.com or email info@arkivum.com