

# Your archive contains a treasure trove of information. Use it

Analytics mean direction of data travel is no longer one way

Joseph Martins Tue 10 Nov 2020 // 07:00 UTC

---

If an historian or biographer really wants to get under the skin of an individual, an organisation or even a country, where do they start? In the archives, where what on the surface might appear to be the minutiae of day to day life can be unlocked, unpicked, and woven into a story that gives us a new and more vivid insight into the past.

Things are slightly different when it comes to enterprise data. Not because it's any less interesting. But because by having the right infrastructure, software and policies in place, the archive becomes not just a source of history, but of data and insight that can deliver value to the business as whole, helping it make decisions today and predictions about what may happen in the future.

There was once a tendency to forget about historical data once it moved off production storage and progressively worked its way through secondary storage, backup, and eventually onto tape, to be stored offline, off site, and inaccessible.

But data's direction of travel is no longer one way. Companies have embraced real time analytics, whether for serving up recommendations for ecommerce customers, automating financial trading or for making real time credit or security decisions. That superfast decision making will often require the analysis of historical data, which requires the ability to access and query that historical data in a timely manner.

Likewise, machine learning, AI and big data all require massive pools of data for analysis or training. The companies that can easily create and trawl these massive pools of their own, often historical data, will be best placed to take advantage of these new age applications.

Over a longer timescale, the analysis of historical data presupposes that it has been suitably archived, indexed and made available on an appropriate system. Your archive data most likely exists in a regulatory and compliance context, whether it is personal data covered by GDPR, intellectual property, or corporate data that is retained to conform with financial rules.

## History lessons

Even if you're not quite ready to pour your historical data into a data lake, ever tighter compliance requirements should be a good enough reason to ensure your data archives are rock solid. A long-forgotten email uncovered through e-discovery can make or break a

multi-million court case, while Europe's GDPR regulations can hit offending organisations with fines of €20m or four per cent of their annual global turnover for infringements.

Just as a reminder, as the Storage Networking Industry Association puts it, a backup is a copy of data, possibly including applications and OS, that allows for "recovery" if the original is lost, corrupted or otherwise inaccessible. Backups can be very short term – production systems will be regularly overwriting data. The primary purpose of an archive is the long-term preservation and retention of data, which has been removed from the primary storage systems.

While some immature organisations might be tempted to use an archive system as their backup, if they hit a problem, they will quickly realise that recovery time is likely to be much longer and possibly less targeted than with a dedicated backup device. Assuming it is indeed the right data.

At the time, two are complementary. If data is not needed for production purposes, it makes sense to move it off expensive, fast production storage and into an archive. This is not just a question of freeing up raw storage capacity. Once off production storage, it no longer takes up compute and network capacity or time being repeatedly cycled through backup processes. A properly constructed archive policy can therefore deliver a tangible performance boost to your storage infrastructure overall.

One final point, is that while archived data should not be confused with backup, archived data can give you an extra level of protection in case of cyber attacks, depending again on your archiving policies.

## **Ransomware**

Ransomware is currently one of the most feared attacks that companies face, precisely because it can make an organisation's data completely inaccessible or unusable. In other attacks where systems are compromised, attackers will work their way through an organisation's systems hoping to strike pay dirt, simultaneously exfiltrating or altering data as they go. In both cases, short-term data protection policies such as snapshotting and backup might not give adequate protection, as the backed up data might itself be corrupted. However, depending on organisation's archiving policy, an archive that has been held offline – that is, properly air gapped - should not be affected.

Tape was once the default media for archiving, but the choice today is broader and more nuanced. However, tape still wins out when it comes to capacity and cost. The latest generation is LTO 8, which offers 30TB of compressed data per cartridge, double the capacity of its predecessor, with transfer rates of 3.24 TB/hr per drive.

Tape suppliers are already lining up launches for next-gen LTO 9, which gives an additional 50 per cent capacity bump, for the end of this year, beginning of next.

And tape offers the very basic advantage that, depending on your policies, once the tape is full, it can be easily removed and physically transferred to a suitable remote location, whether that's a dedicated archive, a bunker or a safe on the other side of the data room. Once this physical air gap has been imposed, your biggest concern will be the physical destruction or theft of the tape itself.

But tape is not the only archive media. The cutting edge in storage is now flash-based, with production storage increasingly based on all flash, or at least hybrid, arrays. But with traditional hard disks still offering ever more capacity and performance, they still retain their place in secondary or bulk storage applications.

That increase in price performance means disk has become a much more plausible option for use in archiving appliances. Querying data for compliance and discovery is faster and much less painful when the data is held on disk. Likewise, disk makes for quicker bulk transfer back to a secondary storage system for use by analytics software or for machine learning training.

The cloud has become a third dimension in archiving. Under the traditional 3-2-1 rule for data protection, most would regard the cloud as a remote location for holding both backup and archive data, though the airgap here is not literal but virtual. It might also be the case that the cloud is a source of data that might need to be backed up and archived.

Very large organisations - or those with the need to archive exceptional amounts of data - might look to a standalone high capacity tape library scaling up into the tens of petabytes. However, mainstream enterprises will rely on appliances to handle archiving. A data management appliance might offer archiving capability alongside backup management, using tape and backup, and perhaps connectivity to the cloud. Alternatively, archiving might be handled by a separate dedicated appliance, again offering tape and perhaps disk. Also consider you should have a backup for your archive data.

### **There will be tiers**

Taking flash, bulk disk storage, tape and the cloud into account, it's clear that you have a choice of storage tiers. So, the choice of data protection software comes into view. Industrial strength data protection platforms should provide archiving, either as standard or as an add-on module. You'll want to be confident you have a solution that gives you flexibility in setting your archiving - and backup - policies, whether the end destination is tape, disk or the cloud, or a combination of all three.

You'll also want to ensure the process is as automated as possible. If you've got the right SLAs and policies for your organisation, you can then remove - or at least substantially reduce - the capacity for human error, by having as little human involvement as possible. At the same time, you'll want to ensure that your software's reporting and analytics is robust enough to ensure that you know that policies are being followed, and that should you need to find something, you can, whether it's an email thread or a vast amount of customer data that you want to retrieve and move to secondary storage for analysis.

You'll also want an architecture that can evolve with you. Data tends to beget more data, and you'll eventually hit the limits of the system you started with. So ideally, you'll scale out easily, as well as scale up, whether that's by adding more disk or tape drives to an existing appliance, adding another appliance, or tapping the cloud. Choosing a flexible, adaptable platform, and a partner you can trust, will mean you are not thrown back to square one when re-examining your data protection strategy.

The archive is no longer the end of the line for your data, but should be considered a critical part of your data protection strategy, a tier that can have an impact on your overall

data strategy, right back through backup, secondary storage and even into production. What was once the end of the line for data, is now just as likely to be the beginning.

*Sponsored by Fujitsu*

Published on "The Register", November 10,2020

[https://www.theregister.com/2020/11/10/your\\_archive\\_contains\\_a\\_treasure/](https://www.theregister.com/2020/11/10/your_archive_contains_a_treasure/)

If you want to know more about different storage solutions, check out the other articles published on "The Register" or explore the storage data jungle: [www.fujitsu.com/data-jungle/storage/](http://www.fujitsu.com/data-jungle/storage/)

The key role of storage in building a datacentric organisation

<https://docs.ts.fujitsu.com/dl.aspx?id=57cc1295-ba0b-46dc-8882-40db4fe8d188>

The persistent importance of traditional storage

<https://docs.ts.fujitsu.com/dl.aspx?id=d82ce000-1188-4d69-985a-d9b102cc14c3>

SDS or HCI? How to figure the right fit for the right workloads

<https://docs.ts.fujitsu.com/dl.aspx?id=981e167e-2f7e-4cd9-baa6-2f73a47ec2f9>

How backup fits into your data-protection strategy

<https://docs.ts.fujitsu.com/dl.aspx?id=c078abcf-2c33-4f4c-9f22-2e9cd1ed48da>