

7 Archiving Pitfalls That Reduce Your Savings

Data archiving is a great way to keep up with data growth and keep storage costs down. When done right, finding cold data and archiving it from your expensive storage and backups can save 70% of costs. But how much you save depends on how you archive. There are distinct differences that can significantly impact the savings you get and the impact on your end users.

This brief identifies seven things to avoid when considering different data archiving solutions and lists what features to look for to optimize your cost savings.

Disruption is the Differentiator

Many vendors offer data archiving or data tiering approaches, but there's a vast difference between solutions. Some say that their data archiving is transparent to users, apps, and workflows. But when you take a closer look at how the solution archives cold data, you see a big difference—and lots of disruption—in terms of what happens when archived files get accessed, hidden costs, and vendor lock-in.

Where Archiving Gets It Wrong—7 Common Pitfalls

The common approaches to archiving are Traditional Archiving and Proprietary Transparent Archiving through HSM or storage tiering. Unfortunately, both approaches have pitfalls that can affect your total savings.

Traditional Data Archiving (aka Lift & Shift)

What is it?

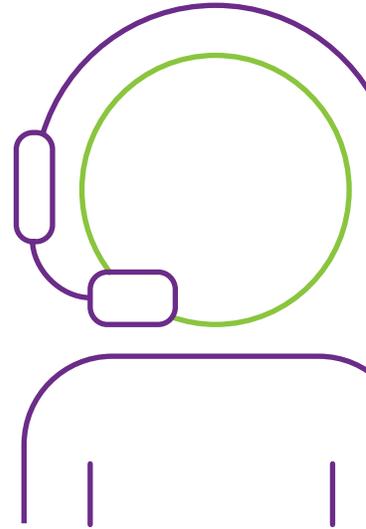
In this scenario, end users can literally wake up and find their data gone. Because, at its most basic, archiving simply moves cold data from the primary storage onto another medium. This means the archived data is no longer accessible from the original location. This is a lot easier than transparent archiving, which is why so many vendors offer it, but this simplicity comes at a cost.



What are the downsides?

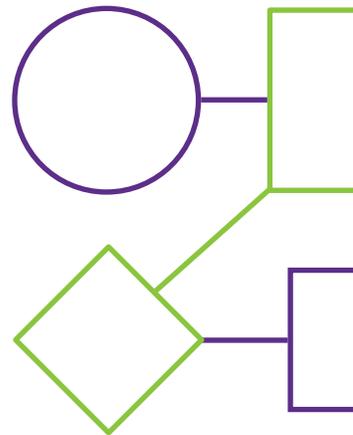
1 IT involvement in file retrieval

If users need to access a cold file or run an older application that requires accessing a cold file that's been traditionally archived, they must file a support ticket. IT administrators like these “go fetch” activities about as much as users like waiting for their files to become available. These are mutually unproductive time sinks.



2 Inefficient manual data archiving workflows

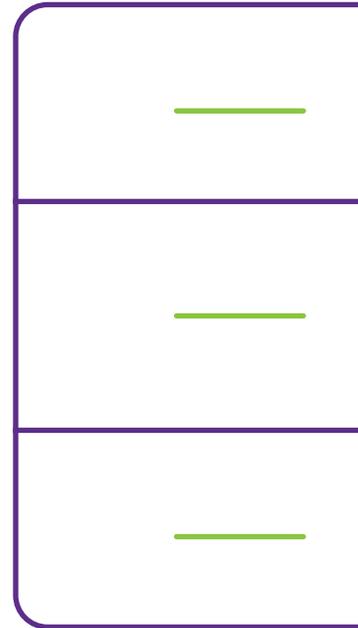
Traditional data archiving requires a manual approval process between users and IT. First to gain permission, then to painstakingly go through which files can be archived, and then repeat on an ongoing basis to keep identifying cold data to offload primary storage. Not only is this highly inefficient, but it results in only archiving less than 10% of the 70% cold data they have—a tremendous savings loss.



3 Requires entire projects to be archived

Traditional data archiving is limited to projects whose data is neatly organized into a collection of data, such as a share/volume or a directory. IT relies on users letting them know when projects are completed and when data becomes cold. Users need ample warning for project-based, or batched data archiving to avoid surprises finding their data.

Obviously, many workplace projects aren't neatly defined, and when they are, they often run for years—all time you're not saving costs. The combined manual approval process and project-based approach of traditional data archiving is not only highly inefficient but results in **<10% of all cold data being archived, which minimizes cost savings.**



Proprietary Transparent Data Archiving: HSM and Storage Tiering

HSM. What is it?

Dating back to the 70's, Hierarchical Storage Management (HSM) is one of the first attempts at transparent data archiving using proprietary interfaces, such as stubs.

What are the downsides?

4

Stubs add latency and risk

Proprietary interfaces, such as stubs, make the archived data appear to reside on primary storage, but the transparency ends there. To access data, the HSM intercepts access requests, retrieves the data from where it resides, and then rehydrates it back to primary storage. This process adds latency and increases the risk of data loss and corruption.

Stub brittleness is also problematic. When stubbed data is moved from its storage (file, object, cloud, or tape) to another location, the stubs can break. The HSM no longer knows where the data has been moved to and it becomes orphaned, preventing access. Existing HSM solutions on the market use client-server architecture and do not scale to support data at massive scale.



Storage Tiering. What is it?

An integral part of the storage array, storage tiering is used to migrate cold blocks from hot, expensive tiers to lower-cost cheaper ones. The problems arose when primary vendors began marketing these as data archiving solutions.

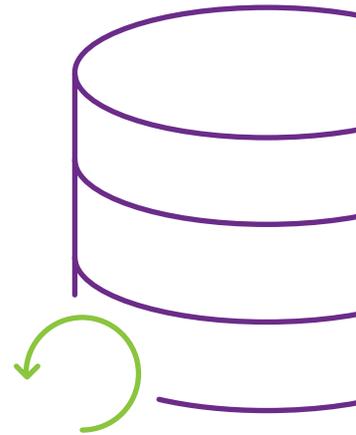
What are the downsides?

5 Erodes significant backup savings

Storage tiering is a block-level tiering technique where the primary storage stores cold blocks of files in less expensive locations, such as cheaper tiers or the cloud. And this is where the problems arise—because this is a proprietary solution not understood by most backup and third-party applications.

In the best-case scenario, the backup footprint stays **the same**, which eliminates all the savings from footprint reduction and backup licenses. *Worst case*, it can potentially *increase* the backup window since fetching all of these blocks back from the capacity storage unit will be slower.

Many storage vendors claim they can tier to the cloud, but it can result in more degraded performance and expensive retrieval and cloud egress costs if the data gets accessed.



6

Lock-in

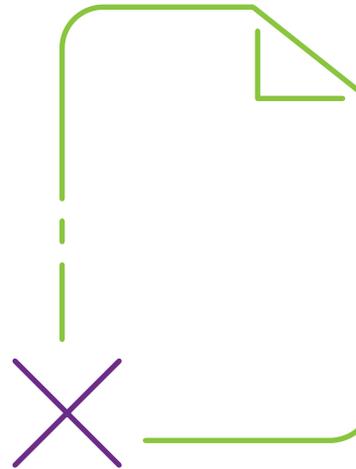
Proprietary block-based solutions limit your ability to switch storage vendors or clouds, which significantly impairs your ability to save costs. Often, all the data has to be rehydrated or brought back before you can switch vendors, which complicates vendor migrations and creates unnecessary costs.



7

No native access on target

Because only cold *blocks*, not the full *file* are moved, the data cannot be directly accessed on secondary storage. When archived files are not easily accessed, IT will be the first to hear about it from end users. This not only affects productivity but diminishes the amount of data users are willing to archive to avoid this negative experience.



What Kind of Archiving *Does* Get It Right?

Standards-based Transparent Data Archiving

What is it?

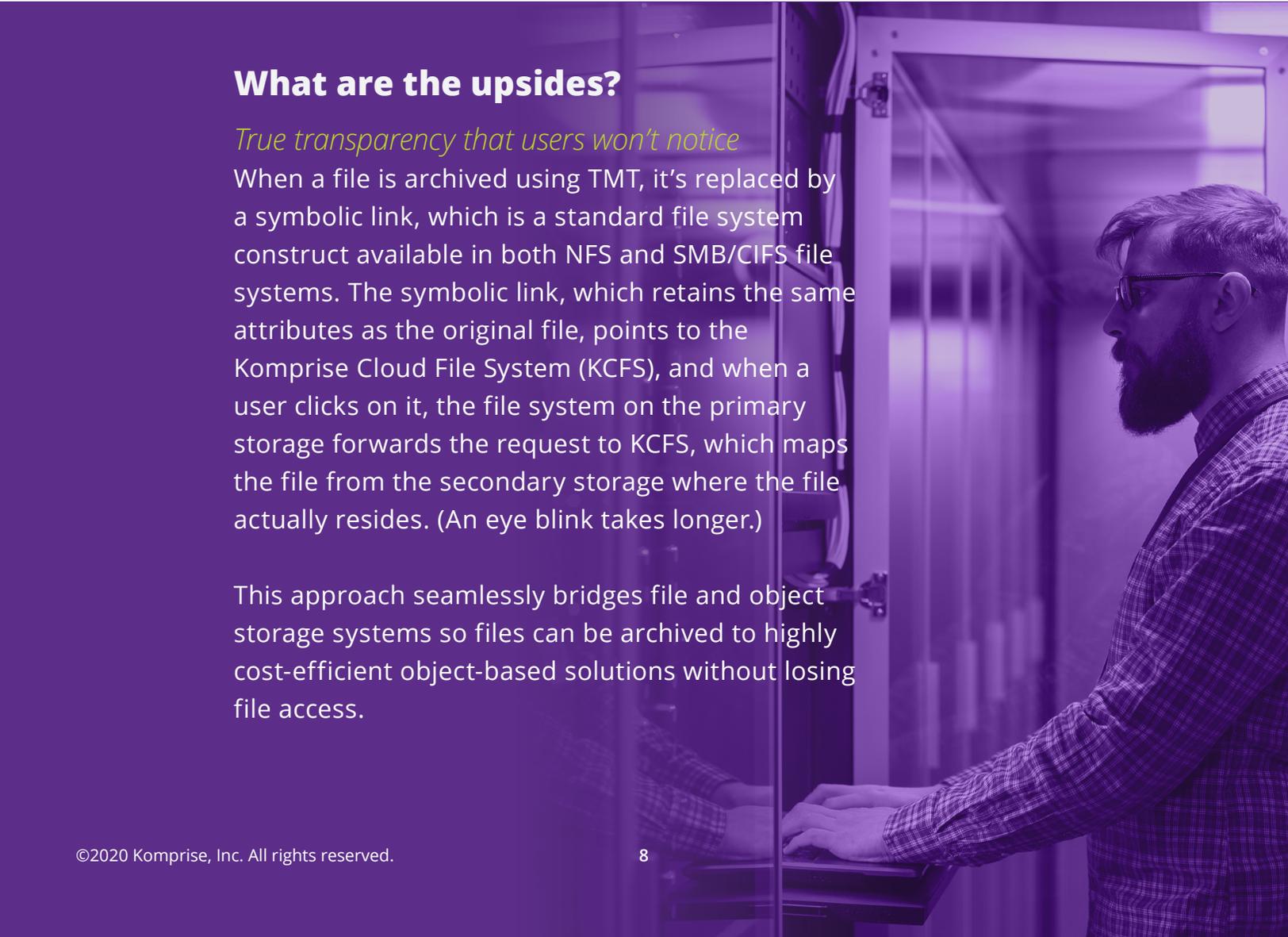
A true transparent data archiving solution creates literally no disruption, and that's only achievable with a standards-based approach. Komprise Intelligent Data Management is the only standards-based transparent data archiving solution that uses Transparent Move Technology™ (TMT), which uses symbolic links instead of proprietary stubs.

What are the upsides?

True transparency that users won't notice

When a file is archived using TMT, it's replaced by a symbolic link, which is a standard file system construct available in both NFS and SMB/CIFS file systems. The symbolic link, which retains the same attributes as the original file, points to the Komprise Cloud File System (KCFS), and when a user clicks on it, the file system on the primary storage forwards the request to KCFS, which maps the file from the secondary storage where the file actually resides. (An eye blink takes longer.)

This approach seamlessly bridges file and object storage systems so files can be archived to highly cost-efficient object-based solutions without losing file access.



Continuous monitoring maximizes savings

Komprise continuously monitors all the links to ensure they're intact and pointing to the right file on the secondary storage. It extends the life of the primary storage while reducing the amount of the primary storage required for your hot data. This savings makes it more affordable to replace your existing primary storage with a smaller, faster, flash-based primary storage.

Reduced backups

Most backup systems can be configured to just backup symbolic links and not follow them. This reduces the backup footprint and costs because only the links are backed up, not the original file. If an archived file needs to be restored, the standard restore process for that backup system needs to be followed to restore the symbolic link that is then transparently used to access the original file.

Less rehydration

To control expensive rehydration, Komprise allows you to set policies when an archived file can be rehydrated back onto the primary server. You can limit rehydration upon first file access, or when a file is accessed a set number of times within a set time. When a file is accessed for the first time, KCFS caches it thereby ensuring fast access on all subsequent access.

The 7 Archiving Pitfalls to Avoid

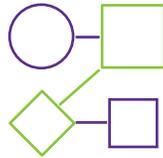
Watch for these seven archiving pitfalls when choosing your data management solution:

1



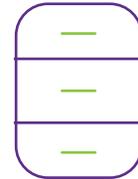
Requires IT involvement in user file retrieval

2



Creates inefficient manual archival workflows

3



Requires entire projects to be archived

Disruption comes in many forms.

4



Uses stubs that add latency and risk

5



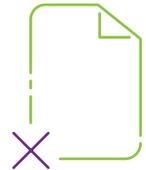
Erodes significant backup savings

6



Locks you in

7



Prevents native access to archived files on the target

Archive data smarter to save more costs

Learn how a standards-based transparent data archiving approach can help your organization. Komprise Intelligent Data Management avoids these data archiving pitfalls to enable maximum savings.

Make the Most of Data Archiving

With the growing pressure to save costs amidst soaring unstructured data growth, it's important to better understand your data archiving options. Some methods can cause unexpected types of disruption, even though they claim the ability to transparently archive.

When you factor user experience, backup footprint, rehydration, and vendor lock-in, it's clear to see why many are choosing a standards-based transparent data archiving solution. With Komprise Intelligent Data Management, you can avoid common data archiving pitfalls and achieve maximum savings without any disruption to your organization.

Learn More

Go to [Komprise.com/product](https://www.komprise.com/product) to learn more.



Komprise, Inc.
1901 S. Bascom Ave. Suite 500
Campbell, CA 95008
United States

For more information:
Call: 1-888-995-0290
Email: info@komprise.com
Visit: [komprise.com](https://www.komprise.com)

For media requests email
marketing@komprise.com
©2020 Komprise, Inc. All rights reserved.