

Migration Isn't Archiving

What's the difference and what matters?

The most surprising outcome of our [webinar](#) introducing the super-fast Komprise Elastic Data Migration feature was the questions. It was clear that the terms *migration* and *archive* were being used interchangeably by IT administrators, so we wanted to set the record straight.

Sure, there are similarities, but the two are quite different and used to achieve different goals. With today's massive influx of unstructured data leading to soaring storage costs, both of these data management operations are important. Choosing the right solution is critical to ensure you're getting the features your organization needs without painting yourself into a corner.

In this paper, we explain the differences between the use cases of archiving and migration, and when you need each, and what are the key requirements you should look for in a solution when embarking on a migration or archive project.

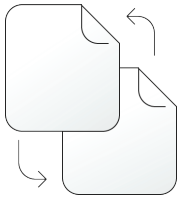
Table of Contents

Migrate vs. Archive: what's the difference	2
What to look for in a migration solution	3
What to look for in an archiving solution	5
Komprise Intelligent Data Management	7

Migrate vs Archive: what's the difference?

Definitions: Looking at the broader definitions of archiving and migration helps keep the two straight.

Migrate



1. **Animals:** Move from one region to another according to seasons.
2. **Computing:** Data migration is the process of selecting, preparing, extracting, and transforming data and permanently transferring it from one computer storage system to another.

Archive



1. **Records:** A collection of historical documents or records providing information about a place, institution, or group of people.
2. **Computing:** Data archiving is the process of moving data that's no longer actively used to a separate storage device for long-term retention. Archive data consists of older data that remains important to the organization or must be retained for future reference or regulatory compliance reasons.

When do you migrate vs archive data?

Migrate for a Storage Refresh

The primary purpose of migration is to change from one system to another. A **migration** solution is typically used when a file server or network attached storage (NAS) is being brought to “end of life,” when you’re undergoing data center consolidation, or moving to/across clouds. The contents of the old file server are transferred to a new one, and the old file server is retired.

Migrations by nature are short-term, one-off projects, and the need to migrate is greater than ever because of the following:

- Massive influx of data
- Rapid changes in technology
- More stringent requirements for a cloud migration utility
- More divergence of price/performance between fast storage and backups (e.g., Flash) and deep storage (e.g., Object and Cloud)

Archive Cold Data

An **archive** solution is used to continuously transfer files that age and haven't been accessed in a long time (cold data) from the fast, expensive primary storage to a slower, capacity-based and less-expensive secondary storage. Archive is a long-running, continuous operation. As data ages in your primary storage, it gets archived, typically either by policy or by projects. The main purpose of archiving is to:

- Delay buying more primary storage by extending existing capacity
- Keep it from being too full to ensure performance
- Reduce the cost of storing cold data
- Reduce the cost and time of backups

What Tools to Choose?

There are many available options for either migration or archiving. However, it's more cost-efficient to invest in a broader data management solution that offers reliable migration *in addition* to archiving and other data management functions for the following reasons:

- **Avoid sunk costs:** Migration is a tactical activity that needs a reliable solution. But if the migration tool is just a one-trick pony, then it becomes a sunk cost.
- **Cut storage costs:** The ability to *archive before you migrate* reduces the amount of expensive storage and backups in the new environment you need to buy.
- **Cut ongoing costs:** Once you've migrated data to the new environment, you need to ensure the new storage isn't cluttered with cold data. The ability to *continue archiving after you migrate* reduces the amount of expensive storage and backups you need to add as data grows in the new environment.

Solutions that migrate *and* archive

- **Avoid sunk costs** of point solutions
- **Save storage costs** by archiving *before* migrating
- **Deliver ongoing** savings by archiving in the new environment

What to look for in a migration solution

Migration must be accomplished reasonably fast without any data corruption or changes in the permissions or attributes of the files migrated. The cutover must be brief, efficient, and issue-free so that come Monday morning, those using the file servers aren't affected.

Below are some of the fundamental requirements to look for when evaluating or purchasing a migration tool.

1. Provide data insight to plan migrations properly:

It used to be that you could develop a migration plan without a lot of insight. That was before there were tens of thousands of shares to deal with. Today you need to know the characteristics of the data on these shares in order to:

- Properly plan the migration effort
- Understand how long it may take
- Make it more manageable by breaking up the project in parallel migrations

27X faster than *rsync*

Elastic Data Migration proved 27 times faster than *rsync* in a LAN environment and orders of magnitude faster than *rsync* on a WAN.

Read this [white paper](#) to learn more about the Komprise architecture and tests results.

Insight Matters Because File Size Matters

You might think migrating a small amount of data, say 100TB, won't take very long. But with analysis, you may find that it contains thousands of very small files, which can take days to migrate. *Size of files matters*. Conversely, the same 100TB of data could contain just a few large files, which may only take several hours to migrate. Unfortunately, most migration tools, whether free or commercial, don't provide an analytics component.

2. Make a full copy of files and their metadata:

Copying just the files is not sufficient. It's critical that the files and their metadata (attributes, permissions, and creation/modify/access times) are all copied correctly.

3. Ensure files on the source aren't altered:

Some utilities take shortcuts for speed, like not restoring the original access time of the files on the source when reading and copying them to the destination. So, if there's a failure requiring re-starting the migration operation from scratch, the access times for a subset of the copied files will be wrong. That's a problem. You'll not only end up migrating files with incorrect times, they'll be neglected by the archiving system because their access times are recent.

4. Ensure the fidelity of migrated files:

It's important that transient errors on the part of the source, destination, or the network don't corrupt a few bits on the files on the destination during migration. To ensure high fidelity, a migration utility should run a checksum¹, and then compare that with the checksum of the file that's already been written to the destination.²

5. Handle and report failures:

For long-running migrations, it's not uncommon for something to go wrong during the operation. The destination share may fill up. Network glitches might result in transient transfer failures. Manually detecting and remediating all these instances would be a daunting time-sink. Look for a utility that recovers from transient failures by automatically retrying the operation or addressing it explicitly in the next iteration. It should report and notify you of permanent failures (e.g., *destination share is full*). When running hundreds of migrations simultaneously, failure reports are critical to efficiently completing the project.

6. Provide fast performance:

For speedy migrations, a utility has to transfer the files and their metadata quickly and efficiently, which includes the ability to:

- a. *Quickly scan for files that have changed.* For large file systems, the scan time may well exceed the transfer time.
- b. *Handle small files.* If the file is very small, the overhead can easily be several times larger than the time to transfer the file. In fact, the file may be smaller than all its attributes.
- c. *Factor network latency.* If the network for migration is over a fast LAN, latencies are typically in the microseconds. WAN latencies are generally in tens of milliseconds. If the migration utility can't handle the long round-trip latency, the performance degradation will be exponentially poor.

WAN vs LAN with *rsync*

In a recent performance test, *rsync*, a popular free migration utility, took a few hours to migrate on a LAN. Migrating that same data set across a WAN with a 30ms latency, *rsync* had not completed after 48 hours.

7. Reporting:

When running many simultaneous migrations, having alerts, notifications, and reports makes for efficient management versus constantly monitoring the progress of individual migration tasks. Reports on what failed help focus the team on hot issues; completion reports can satisfy compliance and governance requirements.

8. UI for management:

Many free migration tools are command-line tools, which are ideal for a one-off, small migration project—but impractical to migrate hundreds of shares. Efficient migration management requires a UI and a console that allows you to search, filter, and monitor the progress of migration tasks and to quickly triage issues.

1. We recommend an MD5 checksum of the source.

2. Note: For compliance purposes, many companies require a report that specifies that a checksum operation was conducted on all files that were migrated. Some utilities perform checksums on chunks of the file as they transfer them to the destination, but not on the complete file, which can lead to corruptions that go undetected.

9. Automation:

Automation that dynamically leverages the processing resources to parallelize and speed the migration, to handle small files, and to mask network latencies and storage bottlenecks is critical in today's modern environments.

10. Cost:

"Free" migration tools are anything but, when you factor the time- and labor-intensive monitoring and management required to ensure the swift and successful completion of a project. Most commercial tools are expensive and come with many bells and whistles. But they're still a sunk cost because once the migration is done, *it's done*. More migrations? Get another license. A data management solution that provides migration in addition to the other key data management operations will avoid these sunk costs.

Free migration tools?

"One-and-done" tools are a sunk cost.

Another migration?
Get another license.

What to look for in an archiving solution

As defined earlier, data archiving is the process of moving data that's no longer actively used to a separate storage device for long-term retention. Archive data consists of older data, or cold data, that's still important to the organization or must be retained for future reference or regulatory compliance reasons. Given today's accelerated data growth and soaring storage costs, archiving has become a critical pillar of the data management process.

When selecting an archiving solution, it's important to consider the following capabilities to achieve greater efficiency and cost savings:

Analyze before archiving

Many vendors offer data analysis—but only after you purchase their storage. Choose a solution that allows you to:

- Analyze your existing silos of data before you act upon them
- Run what-if scenarios to determine the best ROI before you archive
- Generate reports to educate and gain end user buy-in into the cost savings effort

Critical Archiving Capabilities

Analyze before archiving

Archive first, then backup

Archive transparently

Archive natively

Archive first, backup second

With the massive influx of unstructured data, it's both technically and economically infeasible to backup and mirror everything when only a small fraction of the data is mission critical. On average, 75% of a company's data has not been touched in over a year³. By archiving that data off to a less expensive capacity storage, the backup and mirror footprint is reduced by 75%! This results in lower backup license and storage costs and mirroring license and storage costs leading to a dramatic reduction in overall storage costs.

3. Based on Komprise customer data.

Archive transparently

Most solutions archive data by simply lifting and shifting cold data from the primary store. This is enormously disruptive for the following reasons:

- Users suddenly find some of their data missing
- Applications fail because they can't access data they require, and
- Listing and searching provide incomplete, and therefore, incorrect results

This is why many archive solutions simply archive entire directories. This works for data that can be neatly organized into projects and archived off when the project completes. Unfortunately, data growth is seldom well organized. Even when there are projects that can run for years and should be archived along the way. This method is limited and doesn't provide dramatic costs savings.

What's needed is transparent archiving. This approach cherry picks cold files (as specified by policy) across files servers and replaces them with links that represent the files. The end user and applications still see the file on the primary storage exactly where the original file had been and access it seamlessly through the link. The link contains the same permissions and attributes so that listings and searches return complete results. Even backup and virus scanners work as before. Backup will back up the link. If the file needs to be restored, it simply restores the link which, will return the original file when clicked.

The benefits of transparent archiving include:

- No support tickets from frustrated users looking for their data
- More aggressive archiving (e.g., archive data older than six months instead of one year), resulting in more savings
- Enables IT to archive data without having to seek permission from the owners of the data.
- You don't lose the savings by having to "rehydrate" or bring files back for backups and migrations.
- Enables you to access the data from the original source and from the destination directly.

Archive data without disruption

Users access data like they always did

No need to rehydrate for backups or migrations

Directly access data from the original source

Transparency is vital, but it can't come at a cost. The mechanism that provides transparency must:

- *Never get in front of your hot data* thereby possibly risking or impacting the performance of access to mission critical data
- *Never require agents on the clients* (e.g., laptops accessing file systems) or the file servers as this results in more components to maintain and manage
- *Never rehydrate the data back onto the primary storage* when the archived data is accessed. Solutions that re-hydrate the data when non-NDMP backup solutions are used void all the costs savings of archiving.
- *Never require all archived data be rehydrated* when the primary file server is end-of-life'd and its contents are migrated to a new file server.

Archive natively

The way data is archived matters. The ability to write files to archive storage in its native format is critical, especially with the onset of AI/ML applications, which give organizations a competitive edge.

- **Extract data value:** If the bulk of your data is cold and you archive it in a way that inhibits its readability or access, you can't extract value from it. You may not be using AI applications or doing searches across all of your data silos today, but archiving it in native format future proofs your organization to support extracting value from all your data, wherever it is, whenever you want.
- **Offload primary storage:** Maximize the performance and reduce the load on primary storage by archiving in native format.
 - Apps that read the archived data can also be run off the archive storage
 - Searching, tagging, and copying data for AI/ML can be done against the archive storage

One solution does it all: Komprise Intelligent Data Management

We've identified the capabilities to look for in both a migration and an archiving solution. And we've noted the problem of sunk costs with point solutions that require a new license for every subsequent migration. What you need is an intelligent data management solution that provides the best of both worlds—and then some.

Komprise Intelligent Data Management is a comprehensive solution that includes analytics, archiving, migration, and replication—all without affecting user access to data. It also supports AI and Big Data projects with the ability to create a virtual data lake to search, tag, and operate on all of your data across your enterprise.

It's a smarter, analytics-driven data management approach that emphasizes savings and getting more from your data by putting you in control.

Komprise offers one powerful platform for optimal analytics, archiving, migration, cost savings, and value.

Only interested in migration?

Elastic Data Migration is also available as a standalone product, offering:

- **Analytics:** Powerful data analytics give insights that ensure better migration planning and management
- **Super-fast migration:** Specifically designed to work efficiently on small files and work across WANs to reduce migration times, its highly parallelized, multi-processing, multi-threaded approach speeds migrations
- **Cost savings:** Available at a fraction of the cost of leading commercial migration tools.
- **Easy upgrade to the full solution:** Eliminate sunk costs of a migration solution with Intelligent Data Management.

Manage your data with total insight

When you know your data before you archive it, the savings multiply. When you can archive without users noticing, the headaches disappear. And when you can do that with a vendor-agnostic, future-proof approach, you become a hero. That's the power of Komprise Intelligent Data Management.

Want to learn more?

Contact us and we'll show you the true view of all your data and how much you could be saving with smarter archiving, migrating, and much more. Go to www.komprise.com/schedule-a-demo/



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

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