komprise

Transparent Archiving
Moving Cold Data Efficiently, Reliably, and Without Disruption

Efficiently identifying and archiving cold data is key to intelligently managing data growth and reducing significant costs—but only if it's done seamlessly. Komprise Transparent Move Technology™ (TMT) archives cold data without users or applications noticing any access difference. This whitepaper describes how TMT works to reliably move and archive cold data, seamlessly transforming files to objects and back, to and from on-premises NAS, object stores, or the cloud—all without any access headaches.

Introduction

The key to managing rampant data growth is to stop treating all data the same. Hot data—more urgently and more frequently needed files—need to be kept on faster-responding, quicker access (and more expensive) storage and backup media.

Conversely, infrequently used or less urgently needed "cold data" can be stored on slower, less expensive media, with lower levels of backup. Cold data doesn't need to be actively managed with expensive backups and DR. In short, you need to ensure the right data consumes the right level of resources.

Research shows that, across industries, about 76% of data becomes "cold" within a year of creation, yet it's often stored and managed the same as hot data (see Figure 1).

Transparent Move Technology™ (TMT)

Industry	Cold Data	YoY Growth
Banking, Insurance	66%	22%
Healthcare	82%	18%
M&E	88%	20%
Genomics	45%	300%+
EDA	61%	25%

Industry research and Komprise customer data

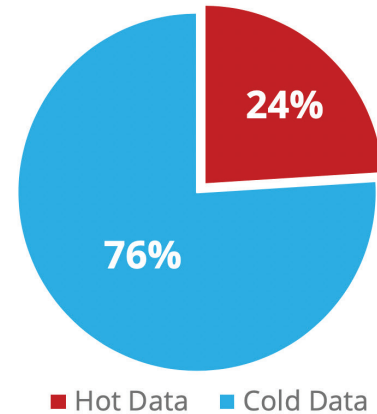24%

76%

■ Hot Data   ■ Cold Data

Figure 1: Across industries, 76% of data is cold and not accessed in over 1 year.

But archiving cold data shouldn't alter its access. Archiving data needs to be transparent, so that it won't impact users or applications. You need to be able to natively access the archived data from its new location – so you can continue to use this data even after it's moved—without paying expensive rehydration costs and without getting locked in.

Komprise TMT seamlessly extends the capacity of the original (more expensive) storage, virtually, with zero disruption. It's part of Komprise Intelligent Data Management, which enables businesses to manage their data better by understanding their data across their storage, seamlessly moving cold data with user-defined policies—without any changes to user or application access. Without an easy way to identify and move cold data seamlessly, organizations end up storing and managing cold data the same as hot data—year after year, with escalating storage costs.

Problems with Moving Data

Moving data from traditional enterprise storage systems, such as NAS environments, to cost-efficient secondary storage solutions, such as object storage on-premises or cloud, without user disruption is no easy feat. Below are the top four challenges with moving data:

1. **Access issues with different protocols and constructs**
   Cloud and object storage both use a different protocol than traditional NAS file storage. For both, the data storage construct is an object, not a file. Object storage is a flat, non-hierarchical storage mechanism. Once data is moved from NAS to the cloud or to object storage, it means that the moved data will now need to be accessed differently. This can break file-based applications, which would require an expensive re-write. It also means users who need that data now have to use some other interface instead of their regular Windows Explorer or Mac Finder tools to access it.

2. **Lack of permissions and access control**
   Each file on a NAS has permissions that restrict who can view, change, or execute the file. This is crucial to the security and privacy policies of your company. Yet, this critical element is lost when you move a file to the cloud. How do you ensure that only the users with the appropriate privileges are able to access the files in the cloud? What happens if a user wants to recall it?

How are access permissions restored and reinforced? These are important considerations and a significant shortcoming with clouds.

3. **Slower performance, longer latency**
Accessing data from the cloud, whether public or private, will not be as fast as accessing data from a local NAS. And, when you recall data from the cloud, there are egress costs to consider. Organizations must be thoughtful about what data should be retrieved from the cloud, how often, and the costs involved. Also, for some use cases, it may be beneficial to run compute in the cloud itself without having to bring any data back – so ensuring <u>native access</u> to archived data in the cloud is important to enable this.

4. **Lack of access continuity / End-user move approval**
Once the data has been moved, users and applications won't be able to access it from its original location. This is very disruptive, and, in most organizations, the decision on what should be moved resides with the end user. Since end-users don't like to move their data, it's hard to convince them to move data to a cheaper secondary storage, resulting in stalled initiatives and rising storage costs.
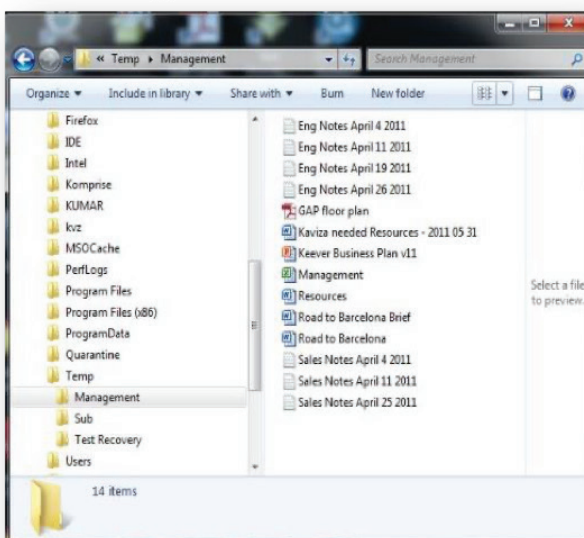
Komprise overcomes these data-move challenges with a patented technology that transparently bridges across file and object/cloud environments. The following sections detail how.

Transparent Move Technology (TMT)

Komprise Transparent Move Technology archives and moves data with the following characteristics:

1. **Moved files remain accessible** from their original location exactly as before, without any changes needed from users and apps.
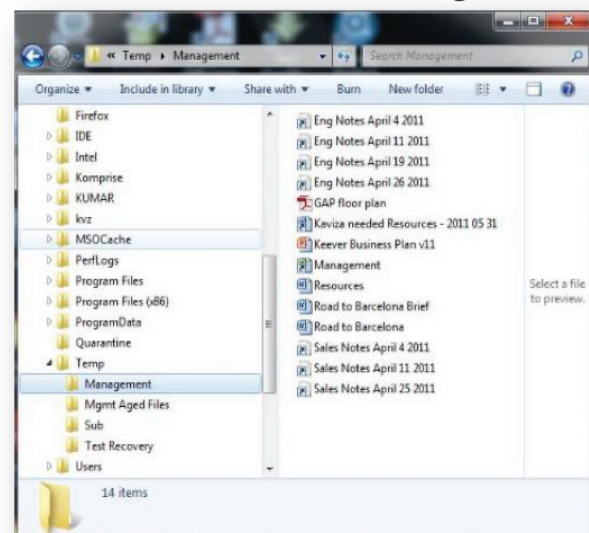
Before migration

After migration

Figure 2: Moved files appear no differently to users.

                                              Transparent Move Technology™ (TMT)

2. **Moves data without proprietary interfaces,** such as stubs or agents. These proprietary interfaces are brittle and problematic to manage. If stubs are deleted, they can leave data orphaned, creating havoc for users and applications. Instead, Komprise points to the moved data using dynamic links that are resilient and work across multi-vendor storage.

3. **Full File access** from source or target: Komprise TMT moves data at the file level with all the file metadata fully preserved at the target so the moved data can be accessed as files from both the source and from the target. Targets can be public and private clouds, LTO, and archival devices or other NAS filers.

   For example, data can be moved from an Isilon SMB share to AWS, and it can be accessed from the Isilon or read directly from AWS – either as files via Komprise or as objects from S3.

4. **Native access on target without vendor lock-in** to storage *or to Komprise*: Komprise moves data using standard protocols such as NFS, SMB, and S3 without any lock-in to any of the storage devices or to Komprise. Moved data is kept in standard object and file formats and is fully accessible from anywhere. You can turn Komprise off at any time or switch your storage vendors or clouds without lock-in. This is a major difference from storage-based tiering or "pools" solutions that do not move the entire file but rather move blocks to cheaper storage such as the cloud. With block- or storage-based tiering solutions, you cannot directly access the moved data in the cloud, thus creating vendor lock-in.

5. **Full permissions and access control** on moved data: Komprise ensures that the original permissions and access control on the files are also moved and enforced when the user tries to access the data through the NAS or through Komprise. If the data is rehydrated back to the NAS, it's brought back with identical permissions, access control lists (ACL), and attributes— nothing changes. Since users aren't affected by moving the data, control and management of the data is back in IT's hands.
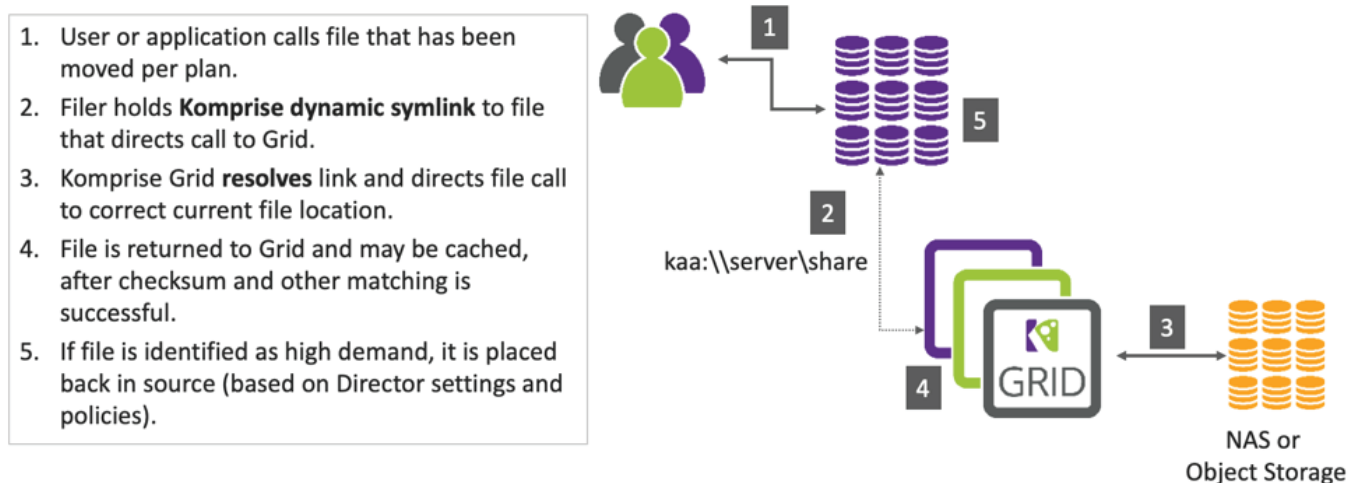

1. User or application calls file that has been moved per plan.
2. Filer holds **Komprise dynamic symlink** to file that directs call to Grid.
3. Komprise Grid **resolves** link and directs file call to correct current file location.
4. File is returned to Grid and may be cached, after checksum and other matching is successful.
5. If file is identified as high demand, it is placed back in source (based on Director settings and policies).

kaa:\\server\share

NAS or Object Storage

Figure 3: Accessing cold data

Transparent Move Technology™ (TMT)

6. **No obstruction of hot data,** optimized access to cold data: When Komprise archives cold data, it doesn't get in front hot data or metadata paths, which can cause performance degradation of hot data or metadata access

7. **Optimized recall of cold data:** With Komprise, users can set policies such as "move all data that's over 6 months old to the cloud". Users can specify exclusions based on size, location, name, and file owner. With this approach, only cold data is moved to the cloud, leaving all the hot data on the NAS where it continues to reside for best performance.

   But today's cold data can become tomorrow's hot data. When a cold file is accessed from the cloud, Komprise caches the file locally so that it's as if the file had never left the local NAS. Komprise streams files so that large files don't incur a huge latency. Specific recall policies can be set so that if the file becomes "hot" again, it can be promoted back to the primary storage.

8. **Bulk recall for projects that become active again.** Komprise also provides a bulk recall feature so that, if it's known that an old project is going to be revived, all those files can be recalled back to the NAS in advance. Figure 3 shows the logical workflow when a user or application calls cold data that has been moved by Komprise.

   Komprise runs as a hybrid cloud service with a "grid" of one or more Komprise virtual appliances, called Observers (and optionally Proxies, for SMB), deployed on premises, as shown in Figure 4. The grid is built using a highly parallelized, scale-out architecture. Observers analyze data across on-premises NAS storage, move and migrate data by policy, and provide transparent file access to data that's been moved. Another virtual appliance, called a Director, runs in the cloud and functions as the management console.
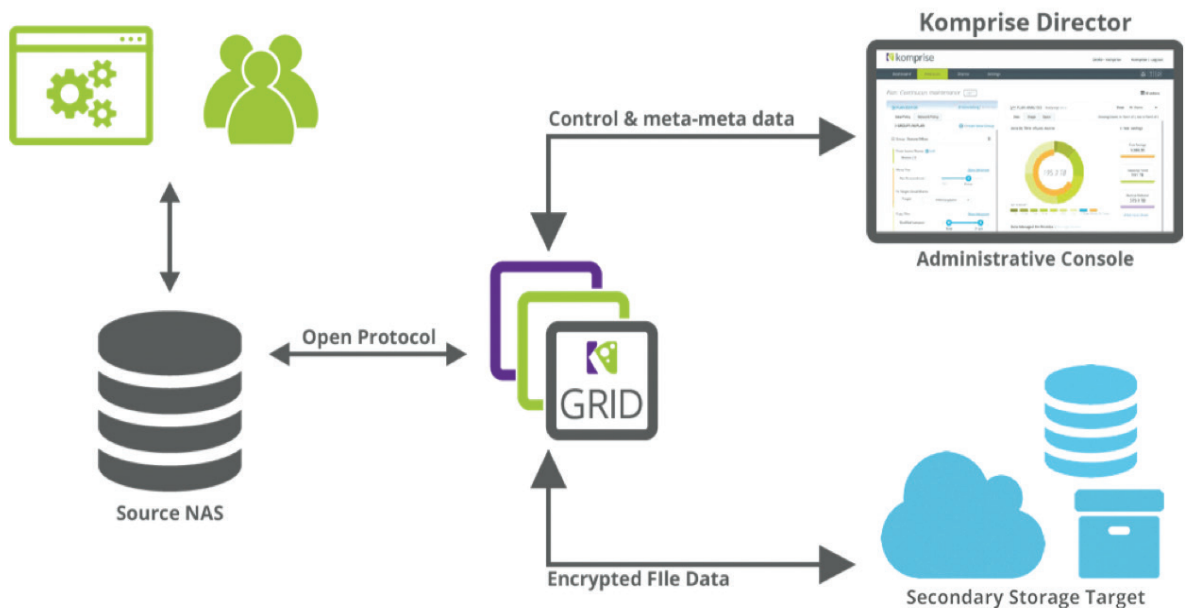

Figure 4: Komprise architecture scales with your data growth

  Transparent Move Technology™ (TMT)

Availability and Scalability

A distributed, scale-out architecture allows Komprise to scale with today's massive data growth. Komprise Observer virtual appliances are connected in a communications grid using a cooperative, distributed algorithm, which enables scalability, load balancing, and fault-tolerance for high availability. The Observers handle all data analysis and data transfers, as well as cold data access. There are no central databases or bottlenecks to limit scalability, and resiliency and high-availability are built-in without requiring any dedicated infrastructure. To learn more about the Komprise architecture, read the **Komprise Architecture Whitepaper**.

TMT in Action

In the table below, you can see how Komprise TMT has helped customers from genomics industry to higher education tackle their rising data growth costs by moving their cold data to more efficient storage.

Customer	Use Case	Benefits
Pacific Biosciences	**Massive Data Management** • Manage data across different NAS & secondary storage (NetApp E-Series, SpectraLogic) transparently	• Manage massive data growth within flat budgets • Able to move data as it goes cold—no longer have to ask for permission as file-based access preserved after data moved
Cadence Design Systems	**Future-proof with Cloud & Create Cost Efficiency** • Use Komprise to identify cold data across on premises NAS and transparently move to cloud –cut 60%+ costs	• 60%+ of data was cold and not recently accessed, moving this to cloud cut storage, DR, and backup costs • Created a path to the cloud
Leading CA University	**Reduce DR Costs** • Put DR copy in the cloud and eliminate on premises NAS mirror	• Cut 50% of NAS costs by eliminating half the footprint • Instant recoverability in the cloud • Fit with their cloud-first initiative
Online Gaming Studio	**Evolve to the Cloud by Transparently Moving Cold Data First** • Using Komprise to identify cold data across on-premises NAS and transparently move to cloud—cut 60%+ costs	• 60%+ of data was cold and not recently accessed. Moving this to cloud cut storage, DR, and backup costs • Creates a path to the cloud for future use • Moving data to the appropriate tier in the cloud, offer ongoing cloud data lifecycle management, and native cloud access

Figure 5: Sample Customer Use Cases

 Transparent Move Technology™ (TMT)

Summary

As data growth continues to explode, businesses realize that a "one-size-fits-all" approach to storage and backups is no longer tenable. Komprise is an analytics-driven intelligent data management platform that helps you easily find hot and cold data and transparently archive cold data to cost-efficient secondary storage. TMT does this transparently and seamlessly, without tying you into storage vendors or proprietary solutions, or experiencing performance degradation. With Transparent Move Technology, you can take control of rampant data growth with a strategic approach allowing you to:

- Move data without stubs or agents
- Easily access moved files same as before
- Retain full file access and permissions from source or target
- Avoid vendor lock-in to storage devices or to Komprise
- No obstruction to hot data and faster recall and access to cold

komprise

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

Transparent Move Technology™ (TMT)