



Cloud Tiering: Storage-Based vs. Gateways vs. File-Based

Which is better and why?

[White Paper]

TABLE OF CONTENTS

Huge Growth in File Data Requires Analytics-Driven Data Management	3
Cloud Tiering Done Right	4
Built-in Storage Cloud Tiering	5
Cloud Storage Gateways	8
Storage Tiering Pool Solutions & Cloud Gateways are Proprietary	9
File-level Cloud Tiering with Komprise	9
Cloud Tiering: Proprietary vs Open	10

HUGE GROWTH IN FILE DATA REQUIRES ANALYTICS-DRIVEN DATA MANAGEMENT

Data growth is exploding, storage costs are escalating, but IT budgets are staying flat. Unstructured data is now 80% of the world's data and it's easy to understand why since in almost every facet of our lives, applications and people are constantly generating file data. Examples of unstructured file data include genomics data, self-driving car and IoT data, audio and video media, application logs, financial documents, and user documents.

Enterprise IT organizations recognize they need better data management to squeeze growth into tight budgets—especially since the majority of enterprise data gets cold and is rarely accessed within a year of creation. Cold data does not need to be managed on expensive high-performance, high-availability file storage and it does not require the same level of data protection and backups as hot data. Offloading cold data to the cloud by using cloud tiering and cloud archiving solutions can cut 80% of storage and backup costs if done correctly.

Managing file data is hard because of the volume, variety and velocity of data generation—a petabyte of file data can easily be a few billion files, each with their metadata, of varying sizes, and varying formats. Proper cloud tiering and data management of files requires analytics to understand the file data and file data management solutions to tier and archive the right data to the cloud. The approach you take to cloud data tiering could either save you millions or cost you 75%+ higher cloud costs.

By 2025:

- 175 ZB of data
- 90% unstructured

—IDC¹

¹ 2019 IDC Forecast for 2025. IDC.com

CLOUD TIERING DONE RIGHT: THE EASY PATH TO THE CLOUD THAT CAN SAVE MILLIONS

Cloud tiering (also known as cloud archiving²) enables enterprises to offload unused cold data into cost-efficient cloud storage and should yield significant savings. Most enterprises today have a corporate cloud strategy, and are now looking to move file workloads to the cloud. When done correctly, cloud tiering is an easy path to the cloud:

- **Cloud Tiering Can Leverage Existing Investments** – Transparent cloud tiering and archiving solutions move cold data from your existing NAS file storage to the cloud, so you extend the lifespan of your current investments.
- **Cloud Tiering and Archiving Done Right is Transparent to Users and Applications** – When done correctly, cloud tiering and archiving is transparent to users and applications—so they can continue to access data exactly as before, even though cold data no longer sits on expensive storage.
- **Cloud Tiering Creates a Path to the Cloud** – Since 80% of unstructured data is typically cold, by offloading cold data to the cloud, you get a bulk of your data in the cloud without disrupting existing users and applications. This creates an easy, frictionless path to the cloud and it enables you to extract value from the cold data using artificial intelligence and machine learning tools available in the cloud. One other benefit of the right cloud tiering strategy is that processing the cold data does not put stress on your high-performance on-premises storage.
- **Transparent Cloud Tiering and Archiving Can Cut 80% of Backup Costs** – Properly implementing cloud tiering and archiving can shrink the backup footprint by over 80% and correspondingly cut backup costs, since cold data is no longer part of the backup footprint. With transparent cloud tiering, backups and restores continue to function as before since cold data is accessed via links that are backed up and restored.

Cloud tiering and archiving can offer significant cost savings, a path to the cloud, and a zero-disruption solution that leverages existing investments. But, not all cloud tiering and archiving solutions are the same. ***You may end up paying more in cloud egress and storage licensing costs by picking the wrong strategy.***

This paper will compare three alternatives to move file data to the cloud:

1. **Built-in storage cloud tiering** (aka “pool” solutions)
2. **Cloud storage gateways**
3. **File-level cloud tiering**

A key technical difference between these three approaches is that both storage-based cloud tiering and cloud storage gateways use a proprietary mechanism to tier pieces of a file to the cloud, while file-level cloud tiering is a standards-based non-proprietary file data management solution that works across any vendor storage and clouds. (Read the paper, [Why Standards-Based File Tiering Matters](#))

² Cloud tiering and cloud archiving are terms that are used interchangeably in this paper and refer to the same concept, that is, the ability to identify and move cold data to the cloud. Transparent tiering or archiving is the ability to do so such that the file looks as if it is still on the source even though it is in the cloud. Transparent tiering is critical to ensure users are not disrupted and IT can tier cold data to the cloud without adversely affecting their end users. Learn more at <https://www.komprise.com/use-cases/archive-transparently/>

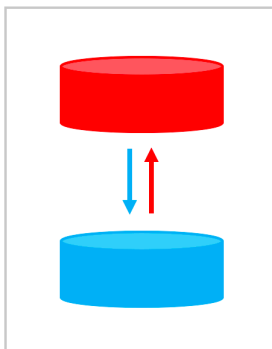
BUILT-IN STORAGE CLOUD TIERING

Block-Level Proprietary Tiering with 75%+ Higher Cloud Costs

Many **Network Attached Storage (NAS)** arrays have their own proprietary tiering solutions, some of which are positioned as “Pool” solutions that tier data to the cloud. “Automated Storage Tiering” is a block-level tiering solution that was first introduced as a technique within a storage array to make the storage box more efficient by leveraging a mix of technologies, such as more expensive high-performance flash or SAS disks to lower SATA disks. As the name implies, block-level tiering moves blocks between the various tiers to increase performance while reducing costs. Hot blocks and metadata are typically kept in the higher, faster, and more expensive storage tiers while cold blocks are migrated to lower, less expensive tiers. The cold blocks often get recalled when the data is accessed and rehydrated back to the expensive tier.

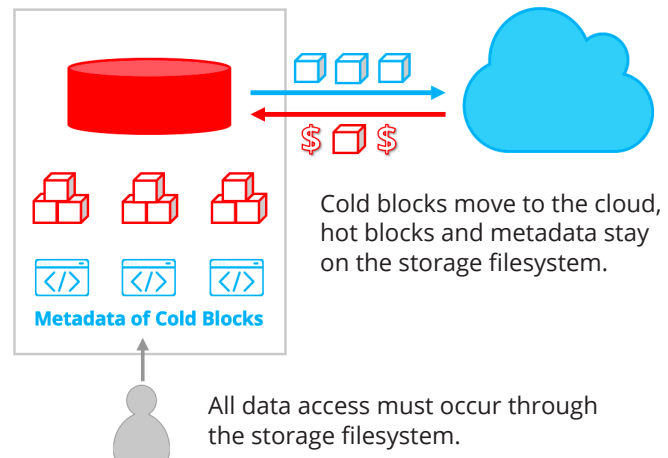
NAS vendors such as NetApp and Dell EMC have built-in tiering solutions like NetApp FabricPool and Dell EMC CloudPools to move blocks of data from their Flash storage to their object storage or the cloud. These built-in storage tiering solutions can be good for moving proprietary data off expensive Flash, but they are not a good fit for tiering files because they can degrade Flash performance and create significant cloud egress costs. A key limitation of this storage-centric, block-level tiering solution is that it is proprietary and all file access must occur through the NAS filesystem.

Storage Tiering Original Design



- Tier cold blocks to cheaper disk (e.g. SATA)
- Accessing cold blocks meant:
 - Slightly more latency, due to slower disk
 - No additional costs

Using Storage Tiering for Cloud Tiering



- Tier cold blocks to the public cloud
- Accessing cold blocks means:
 - Added WAN latency, lower SLA
 - Egress and cloud storage costs
 - No direct native cloud access

Figure 1: Storage-based tiering is a mismatch for the cloud.

Since storage tiering was built to tier within a storage array, and is not really designed to tier files to the cloud, using storage tiering solutions to tier file data to the cloud has very costly implications.

Proprietary Lock-in and Cloud File Storage Licensing Costs

Since the blocks moved to the cloud are meaningless in and of themselves, they cannot be directly accessed in the cloud. This means that the NAS filesystem must be used to access the data. In other words, even if you want to access data in the cloud, you need to run a cloud instance of the NAS filesystem. You cannot directly use native cloud services to access your data in the cloud—it has to be through the proprietary storage filesystem itself. This creates unnecessary licensing costs that customers must bear forever to access their data and creates undesirable lock-in as you cannot directly use cloud native tools without relying on the filesystem for access.

75% Higher Cloud Egress and Retrieval Costs

NAS cloud tiering creates costly egress charges when blocks are read from the cloud—which can happen with many third-party applications on both sequential and random reads. Often, the blocks have to be brought back from the cloud or rehydrated when read and accessed by third-party applications such as file-level backup software or other third-party applications. Even sequential reads such as index scans or anti-virus software can cause blocks to be read from the cloud. This can get very expensive since reading data from the cloud incurs egress charges and retrieval charges. Many cloud tiering solutions also bring blocks back for defragmentation—an unnecessary operation that is needed only because of the proprietary block-based approach.

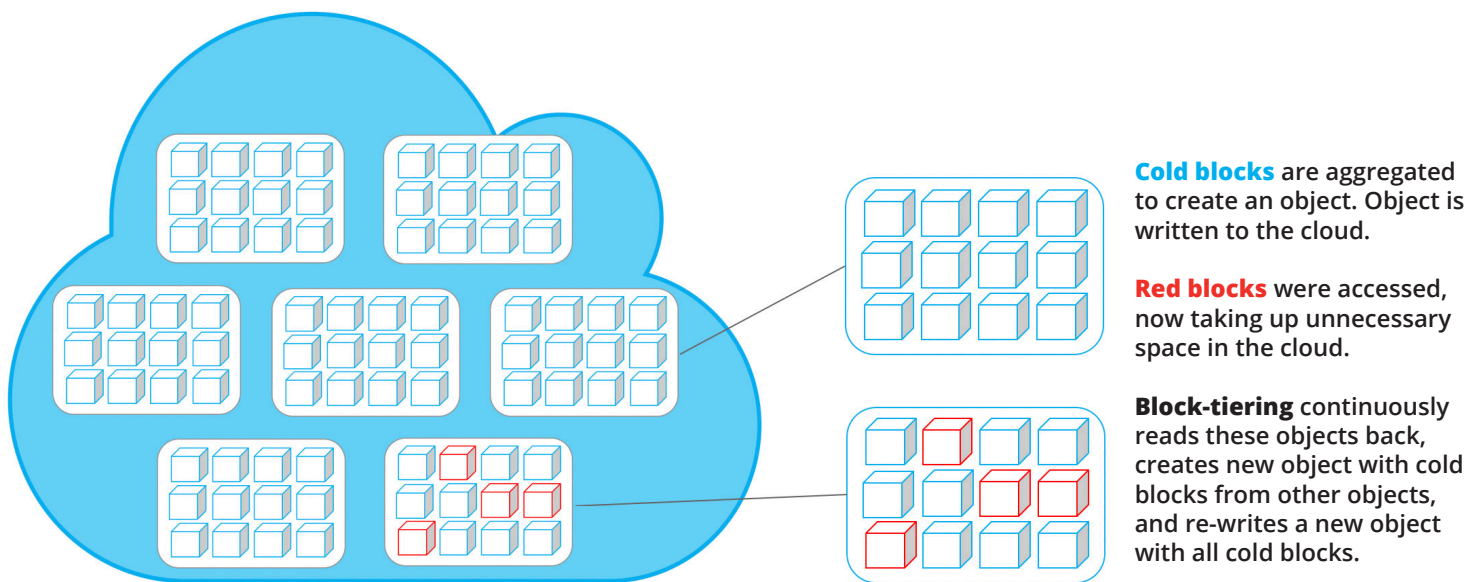


Figure 2: Defragmentation, resulting from cold blocks being accessed, leads to traffic and increased cloud costs.

Our analysis shows that on a petabyte of data, storage-array cloud tiering Pool solutions create 75% more cloud egress charges than necessary across sequential reads, random reads, and defragmentation. This is why storage vendors themselves do not recommend their cloud tiering solution for customers who have more than a few hundred terabytes of data.

As an example, here is a table comparing cloud egress costs when tiering cold data from one petabyte of source using storage array tiering (block-based) versus Komprise (file-based).

	Storage Tiering to Amazon S3 IA >6mo	File Tiering with Komprise to Amazon S3 IA and Glacier
Storage:	\$262,144	\$104,601 (S3 IA up to 4 yrs, Glacier >4 yrs)
Random Read Costs:	\$51,130	\$8,372
Anti-Virus Read*:	\$421,822	<i>Not Applicable</i>
Defragmentation:	\$32,181	<i>Not Applicable</i>
TOTAL:	\$767,276	\$112,973
SAVINGS with Komprise:	\$654,303 per PB	85.27%

* Assumes a conservative 1 full virus scan every 3 months. Third-party backup and other sequential reads may have similar costs.

Figure 3: Storage-based cloud tiering is a good solution for tiering snapshots from Flash, but is 75% more expensive than file tiering to the cloud.

CLOUD STORAGE GATEWAYS

Cloud Storage Gateways Create Expensive New Silos and Ongoing Proprietary Cloud Costs

Cloud storage gateways create a new appliance (virtual or physical) that acts as your storage at each site to cache data locally and put a golden copy in the cloud. They are useful when you are doing active file collaboration across multiple sites and do not have NAS at branch sites or do not want to use your existing NAS. But, they do not leverage existing storage investments and require data to be moved to the gateway which creates additional infrastructure costs. Cloud storage gateways store data in the cloud in their proprietary format. Similar to storage-based cloud tiering, cloud storage gateways create proprietary lock-in and unnecessary cloud gateway costs in perpetuity. And they also typically create additional on-premises costs.

- **Additional On-Premises Infrastructure:** Cloud storage gateways are typically hardware-based since they have to serve hot data from the cache. Many vendors also offer virtual appliance options for smaller deployments.
- **Duplication of Data in the Cloud:** Cloud storage gateways typically put all the data in the cloud and then cache some data locally. So, if you are using a cloud storage gateway for 100TB, then all 100TB of data is in the cloud and a subset of it (maybe 20TB or 30TB) is also cached locally. This means you may need 130TB of infrastructure to house 100TB of data. Depending on the size of the local cache, this may be larger.
- **A New Storage Silo:** A cloud storage gateway is a new storage infrastructure silo that caches some data locally and keeps all of the data in the cloud. It replaces your existing NAS. It does not work with it. It is a rip-and-replace approach.
- **Cloud Storage Gateway Licensing Charges to Access Data in the Cloud:** Cloud storage gateways lock data in the cloud with their proprietary format. This means you cannot directly access your data in the cloud—data access needs to be through the gateway software in the cloud. Many customers are surprised to learn they have to pay gateway licensing costs even to access data in the cloud, and this cost continues as long as you need your data. This lock-in limits flexibility and creates unnecessary cloud expenses. It also limits your use of the cloud as you cannot natively access your data without the gateway software.

Assuming \$700/TB/yr. of cloud storage gateway licensing costs, cloud storage gateways have 287% higher annual costs than using a file-level data management solution with the cloud. This is a recurring cost that you pay for over the lifetime of your data!

	Cloud Storage Gateway	File-Tiering with Komprise
Gateway Cost:	\$700 / TB / YR	
S3 IA Cost:	None	\$155 / TB / YR
File Data Mgmt Cost:		\$90 / TB / YR
Additional Cost/TB/Yr:	\$456 / TB / YR 287% Higher Cost per Year with Gateways	

Figure 4: Cloud Storage Gateways are good for file sync-and-share but can be nearly 300% higher than file tiering to the cloud.

STORAGE TIERING POOL SOLUTIONS AND CLOUD GATEWAYS ARE PROPRIETARY, LEADING TO HIGH COSTS, LOCK-IN

Both storage-based tiering and cloud storage gateways are proprietary cloud tiering solutions that create undesirable lock-in, and result in ongoing cloud licensing costs and high infrastructure costs. They don't allow you to leverage the benefits of the cloud and use cloud native tools to analyze and process your cold data. These solutions were designed for different use cases and not really to tier cold data seamlessly to the cloud. To realize all of the potential benefits of cloud tiering, you need a file-level cloud tiering and transparent archiving solution.

FILE-LEVEL CLOUD TIERING WITH KOMPRISE

File-level Cloud Tiering is the Easy Path to the Cloud that Maximizes Savings

Unlike proprietary tiering solutions, file-level tiering is a data management solution that works across multi-vendor file and cloud storage to move files seamlessly without lock-in. File-level tiering and transparent archiving solutions overcome the limitations of block-level solutions by keeping data in native format in the cloud while enabling transparent access to tiered data from the original filesystem.

- **File-level Cloud Tiering Eliminates Lock-In:** File-level tiering is a more advanced technology and is standards-based. File-level tiering means the file along with all of its metadata moves to the new tier. Whether you have NTFS extended attributes or POSIX ACLS, you need the ability to move the file and all of its associated metadata with high fidelity and rehydrate it back into its exact original form if needed. Moving just the file is not enough. Many applications rely on attributes of the file to operate. The file system imposes access control through basic and extended attributes. File-level tiering maintains full file fidelity and preserves all the attributes and metadata along with the file at each tier.
- **File-Level Cloud Tiering Minimizes Cloud Costs, Maximizes Savings:** The total savings is more significant since the entire file is preserved. With transparent file-level cloud tiering you can access files natively as objects in the cloud without any third-party software, so there is no need to pay an additional fee in perpetuity. Since all the file metadata is preserved, you can also access the cold data as files in the cloud and return it to the source storage exactly as it was before. Also, by transparently moving the entire file, all other operations such as third-party backup applications and migrations can be done without rehydrating the data or incurring cloud egress and retrieval costs—thus maximizing the full savings.

File-level cloud tiering enables you to maximize savings by offloading cold data to the cloud. It minimizes cloud egress costs by enabling access to data without costly rehydration, and it future-proofs your investment by writing data using standards with no lock-in.

Don't Compromise. Komprise.

Komprise delivers faster, smarter cloud tiering at the file level. Komprise Intelligent Data Management frees you to easily analyze, mobilize, and access the right file and object data across clouds without shackling your data to any vendor. With Komprise, you are able to know first, move smart, and take control of massive unstructured data growth while cutting 70% of enterprise storage, backup, and cloud costs.

We believe that data management functionality is a layer **independent of storage**. By considering data to be separate from the storage in which it resides, it is possible to manage data holistically across vendors—be they on-premises storage arrays or cloud providers—and across technologies, be they files or objects. This approach has allowed Komprise to create a data management solution that is vendor agnostic and integrates tightly with on-premises and cloud storage to create a hybrid data management platform.

This open approach to data management ensures Komprise offers cloud tiering with **75% lower** cloud egress costs, **300% lower** TCO, and provides you with full access to your data in the cloud **without lock-in**.

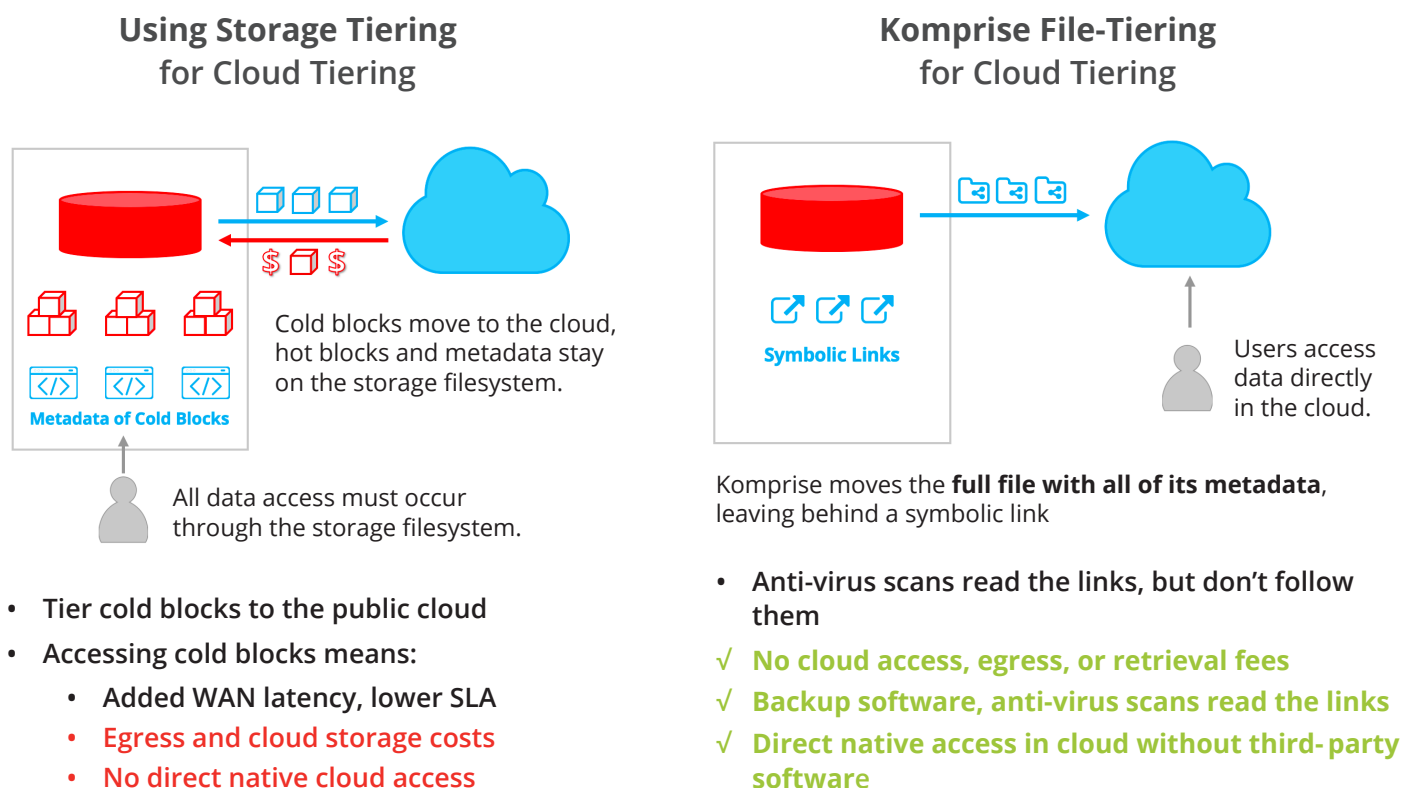


Figure 6: Komprise is designed to transparently tier data to the cloud while providing the most cost savings and enabling maximum flexibility without any vendor or data lock-in.

Komprise uses open standards to read and write data to ensure data is always stored in a format native to the storage service to ensure there is no data lock-in. This allows customers to manage their data independent of the storage devices or data management services they use. It future proofs the customer, allowing them to select and later change or update their storage devices. As an example, one of our customers tiered cold data to tape. Later, they shifted to an on-premises object store and recently moved to the cloud. Through it all, the one constant was Komprise.

CLOUD TIERING: PROPRIETARY VS. OPEN

Summary of the Differences

	Built-in Storage Cloud Tiering (aka Pool)	Cloud Storage Gateways	File-Level Cloud Tiering w/ Komprise
Approach	Block-Level	Block-Level	File-Level
Eliminates Lock-In	NO	NO	YES
Leverages Existing Infrastructure	YES	NO - 300% Higher Costs	YES
Minimizes Cloud Egress Costs	NO - 75% Higher Costs	YES	YES
Works Across File Storage	NO	NO	YES
Users Access Moved Data without Disruption	YES	NO	YES
Native Access in the Cloud	NO	NO	YES
Eliminates Third-Party Licensing Costs in Cloud	NO	NO	YES
Enables ongoing Lifecycle Mgmt in the Cloud	NO	NO	YES
Ideal Use Cases	Tier Snapshots from Flash	Global File System	Tier and Archive Cold Files to the Cloud

Figure 5: File-level cloud tiering benefits vs. cloud storage gateways and storage pool tiering.

Ask Your Vendor: Is Your Cloud Tiering Proprietary or Open?

Cloud tiering using a proprietary solution is easy for vendors to implement, but it creates cost overruns for customers and locks their data access in the cloud forever with expensive licensing costs. Ask your vendor how they tier data to the cloud:

- Can you leverage your existing investments?
- Can users access data without disruption?
- Can you use your data in the cloud without any third-party software including theirs?

If the answer to any of these questions is no, the cloud tiering solution is proprietary and unlikely to be the best long-term solution.

Cloud Tiering and Archiving at the File-Level Maximizes Savings and Future-Proofs Your Cloud Strategy

Cloud tiering at the file-level is an advanced solution that delivers a complete data management solution. It is your easy path to the cloud that preserves your existing storage investments, transparently moves files so your users do not need to change how they access data, and yet keeps data in the cloud in native format so you break the lock-in of storage filesystems.

LEARN MORE

To learn more about cloud tiering at the file-level with Komprise, visit [Komprise.com/path-to-the-cloud](https://komprise.com/path-to-the-cloud)



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
©2021 Komprise, Inc. All rights reserved.