

Extending On-Premises Network-Attached Storage to Google Cloud Storage with Komprise

This article details how you can use the Google Cloud Platform (GCP) service Cloud Storage and Komprise to actively archive and replicate data to the Google Cloud without disrupting users and applications.

The classic network-attached storage (NAS) device has been a cornerstone of the data center for over 20 years. These hardware platforms are well understood but can be an expensive tier of storage for seldom-accessed data. Enterprise NAS devices are typically refreshed at three- to five-year intervals, and the migration of large amounts of data can be complex and risky. At the same time, the majority of this data has become inactive or cold and has not been accessed in over a year or more. In some cases, moving this cold data to Cloud Storage can reduce storage costs by as much as 70 percent.

Additionally, many organizations want to keep a copy of their data in the cloud for redundancy and disaster recovery (DR) purposes. This approach can be a simple and inexpensive way to protect data that is not mission-critical.

STORAGE IT CHALLENGES

While many IT administrators know that a large segment of the data on their expensive NAS is cold, they also know that moving that cold data could be disruptive. If users or applications need access to archived data and are unaware that it was moved, operations and applications are disrupted. Asking permission from users is not any easier. Users are seldom willing to have their data archived, even when they have not accessed it in years. Even if they grant permission, identifying the correct subset of data and migrating that data to the cloud has been an extensive, cumbersome manual process that involves spreadsheets, reporting tools, and various software applications. As the data footprint grows, the problem only escalates.

Komprise coupled with Cloud Storage addresses this issue by automatically identifying and moving cold data by policy from any NAS to Cloud Storage without disruption. Data that Komprise moves still appears to users as if it is stored on the primary NAS. When a user or an application accesses this data, Komprise automatically recalls the data, preventing any disruption. This transparent bridge of object data to files enables seamless access to archived data from the NAS and is crucial to data management. IT can manage their storage farms efficiently and automatically without asking for user permission.

Another issue that IT faces is that with unstructured data being generated at such a rapid pace, backing it all up is simply too costly. In most cases, such data is never backed up. With Komprise, you can replicate this data to durable Cloud Storage, providing an automated, simple way to help protect your data and facilitate DR.



Capabilities

Komprise consists of a grid of one or more virtual appliances that are deployed on hypervisors in the data center. Install and point Komprise at the NAS shares that you want to analyze and manage. Komprise first analyzes the shares and then provides insight that the IT administrator can use to make management and capacity-planning decisions, which they codify as simple policies that direct Komprise to move and replicate data to Cloud Storage.

ANALYZE DATA USAGE & GROWTH ACROSS STORAGE

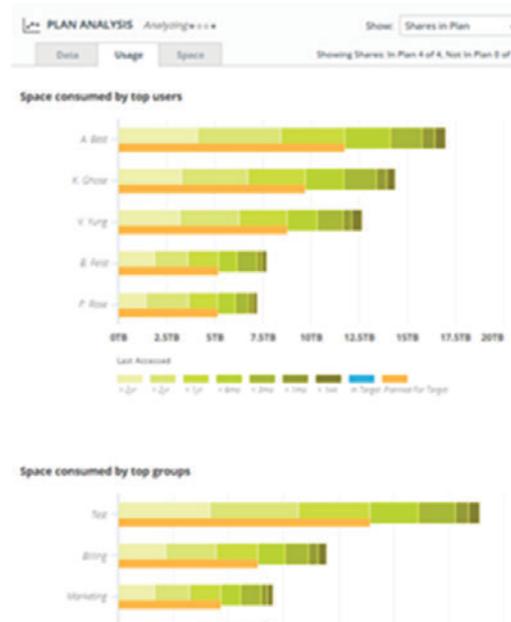
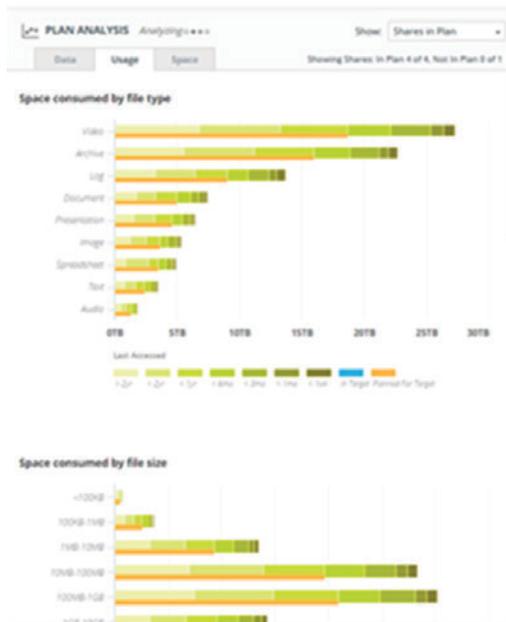
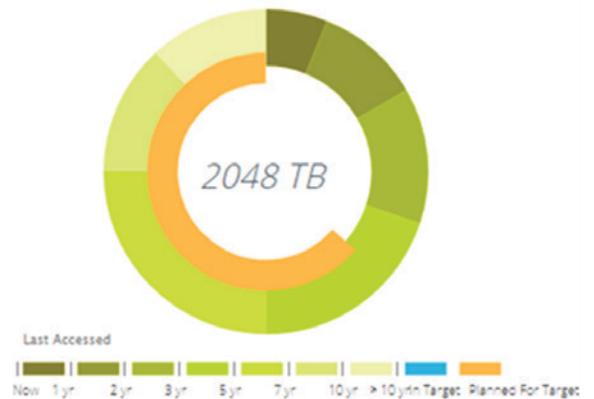
Komprise profiles the data across storage silos to identify how much data is hot and how much is cold. It then provides analytics to help answer the following questions:

- What are the types of files?
- What is the distribution of file sizes?
- Who is accessing which files?
- How fast is file storage growing?
- How much data is inactive?

Komprise provides charts to capture the data profile. The following donut chart shows that Komprise has analyzed 2 petabytes (PB) of data, and the colored buckets indicate by volume when data was last accessed, from 1 year ago to more than 10 years ago. You can customize the granularity of these buckets. The orange ring illustrates the move policy that is specified by the administrator and shows that all data that has not been accessed in more than five years is slated to be moved.

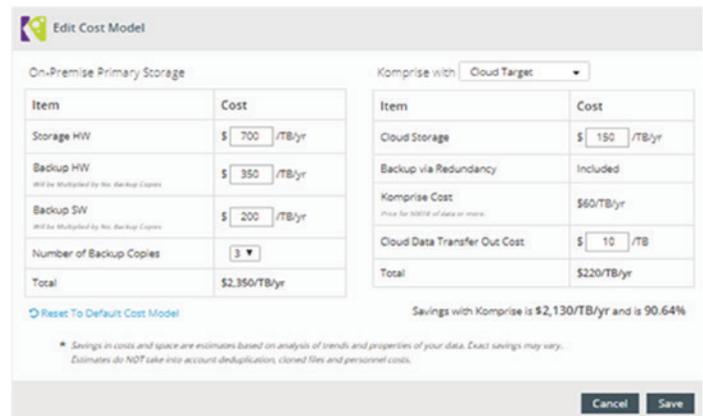
To facilitate more granular decision-making, Komprise also provides access and aging information that is based on file type, size, owner, group, and directory. Komprise also provides return on investment (ROI) information through a built-in tool.

Data by Time of Last Access



You can use the ROI tool to enter costs for NAS and backup that are specific to your environment and to determine your projected savings. By moving rarely accessed data off of your primary NAS, you can realize significant savings. Furthermore, moving that data prior to storage refresh or upgrade can offset some or all of the cost of storage acquisition when utilizing newer architectures such as Cloud Storage.

Depending on which Cloud Storage class you want to use: Regional, Multi-Regional, Nearline, or Coldline, you can modify the cost model to reflect the savings that you can realize after moving data.



MOVE & COPY DATA BY USING SIMPLE POLICIES

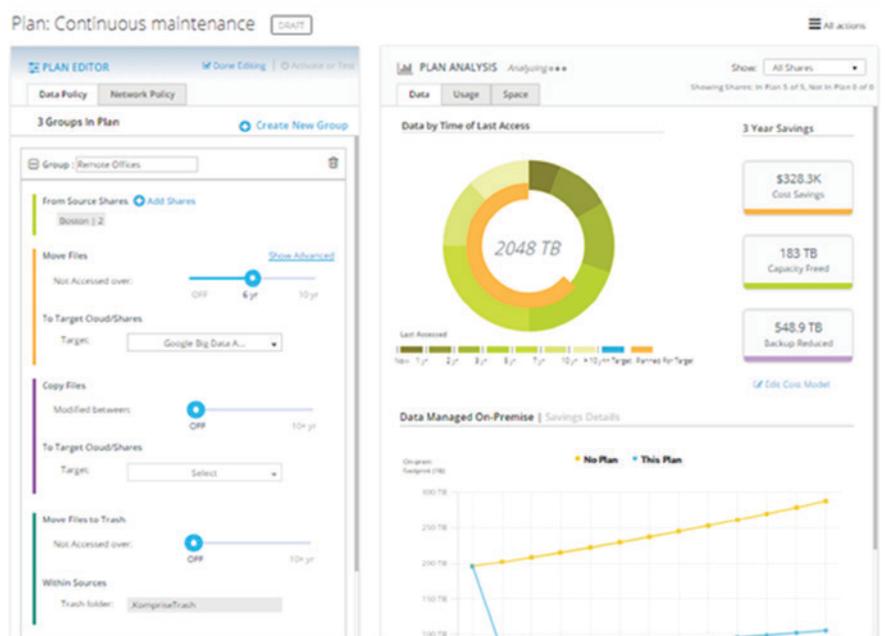
In addition to analytics, Komprise provides policy-based move and copy operations that use simple sliders and pick lists. See the following figure for an example of the policies.

- The move policy continuously moves inactive and cold data to Cloud Storage as the data ages. Identifying and moving cold data eliminates the ongoing need to increase the capacity of on-premises NAS storage.
- The copy policy facilitates copying data to the cloud for DR. You can select different conditions for copying data. For example, some users might want to replicate to the cloud only data that was modified in the last year.
- If obsolete data needs to be removed rather than moved or copied to a new storage platform, you can specify a policy to identify and move such data to a trash folder on the NAS.
- In addition, if certain data should not be moved or copied, you can define specific exclusions using file types, size, and folders.
- You can build multiple Komprise groups to set up custom policies for data that has unique needs.

After you set the move policy and save your plan, Komprise dynamically calculates the estimated capacity that will be freed up and your projected cost savings.

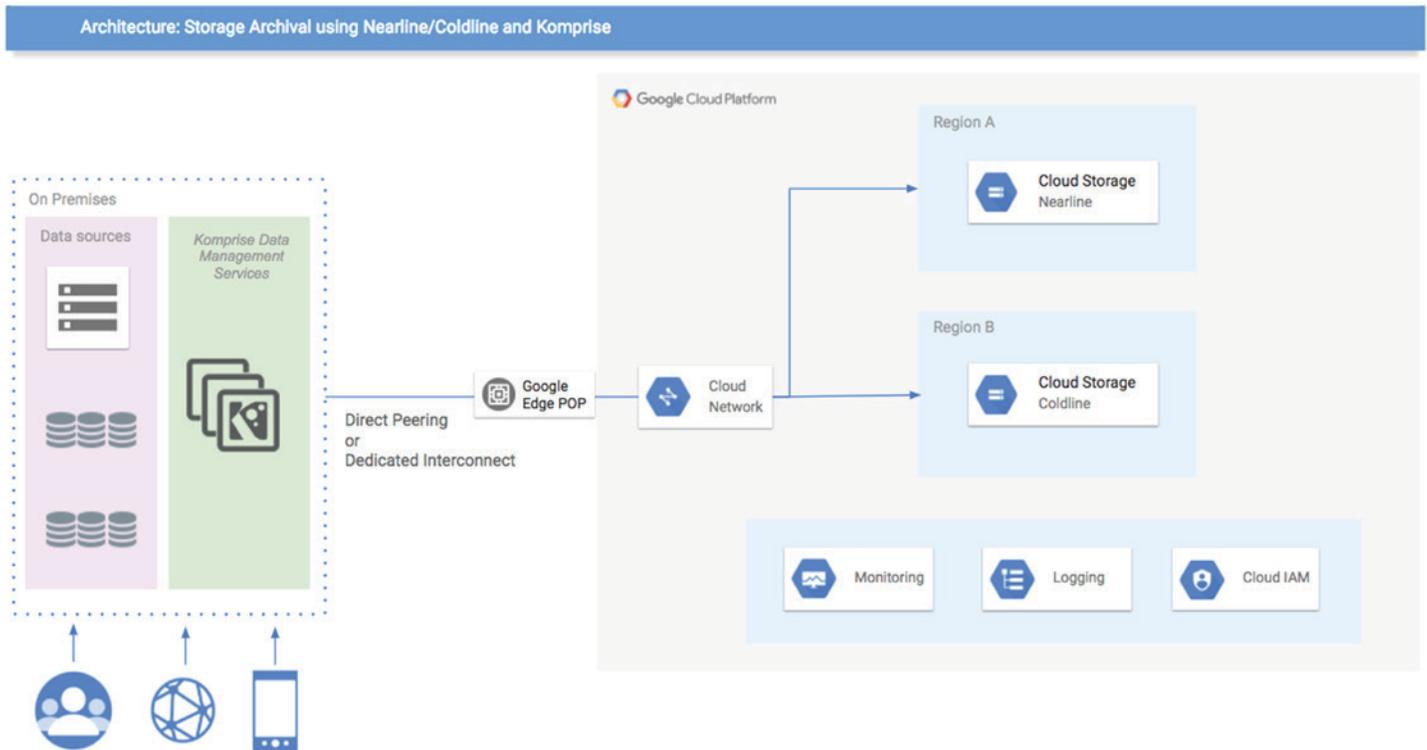
INFORMATION LIFECYCLE MANAGEMENT

Komprise uses tiered Cloud Storage to further reduce costs. Through policies that you set in Komprise, you can tier data from Nearline storage to the less expensive Coldline storage based on the age of and lack of access to the data after you have moved it to Nearline storage. Both provide similar access times, so you can reduce costs further by using Coldline storage without affecting your ability to access the data when you need it.



Architecture

The following figure depicts the architecture of a typical solution



Features

KOMPRISE DATA MANAGEMENT

Komprise runs as a hybrid cloud service with a grid of one or more Komprise virtual appliances, called Observers and Proxies, deployed on premises. The grid has a highly parallelized, scale-out architecture. Observers analyze data across on-premises NAS storage, move and replicate data by policy, and provide transparent file access to data that is stored in the cloud. Komprise Proxies encapsulate the extended Server Message Block (SMB) or Common Internet File System (CIFS) metadata and permission structure for compatibility with modern object architectures and accelerate file transfer to Cloud Storage. Finally, a Komprise Director virtual machine (VM) runs in the cloud and provides the management console.

SCALE OUT

Komprise does not require any dedicated hardware and runs as a scale-out grid of VMs that are managed as one logical unit. There are no centralized databases, which allows Komprise to grow on-demand to handle data at massive scales. The grid is highly available and so long as at least one Observer is healthy, access to all moved data remains intact. Komprise does not store data, and simply moves data through SSL to Cloud Storage, which is HIPAA-compliant.

NON-DISRUPTIVE

A typical challenge with traditional storage services is that they might disrupt end user access. Komprise preserves the directory structure as well as file attributes on the target, unlike cloud migration tools that strip data off file attributes and move blocks to the cloud that can only be accessed and understood using the application going forward. With Komprise, end users can continue accessing files with no change to their processes, because the location of data is transparent to them.

HIGH PERFORMANCE

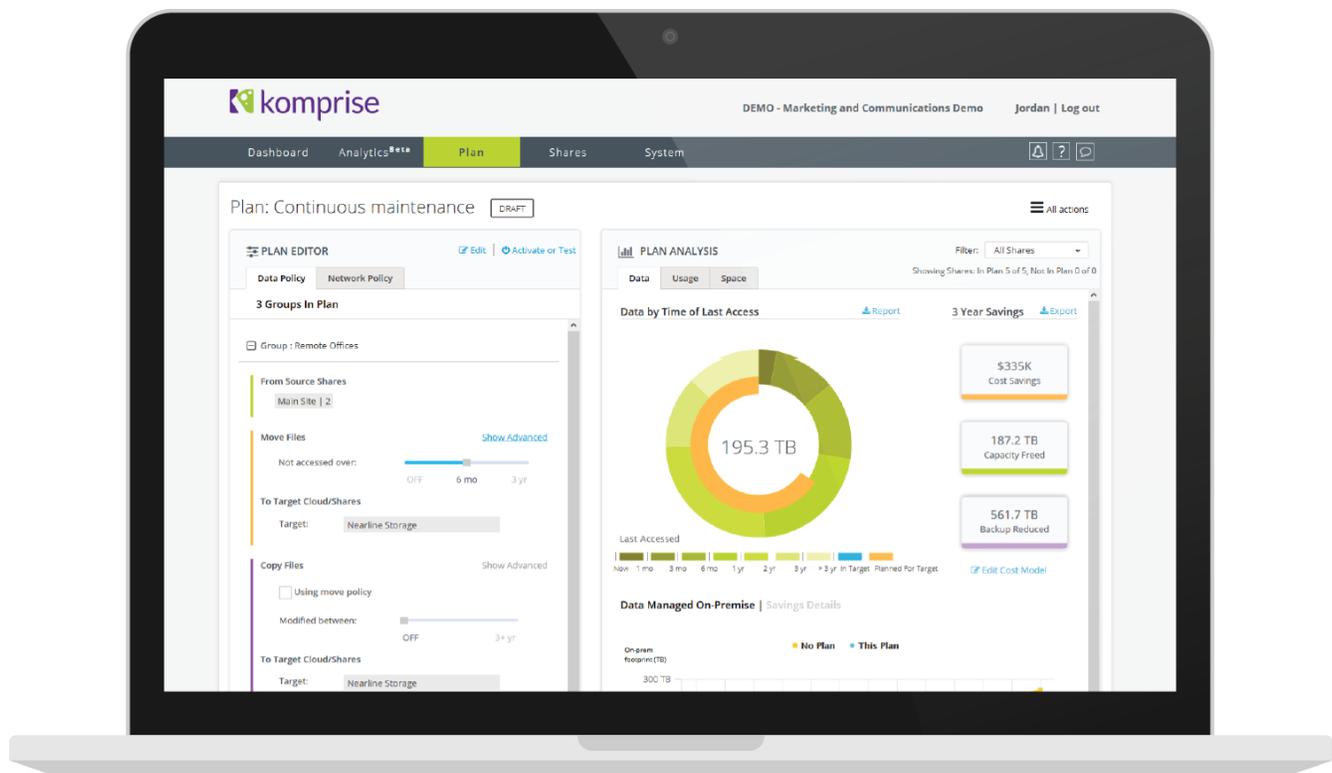
Several migration solution providers significantly reduce the performance of storage during data moves. Komprise, on the other hand, is invisible to the hot data path and does not get inline. Additionally, Komprise adaptively throttles back when the storage systems are in active use so that Komprise analytics runs non-disruptively in the background. As a result, the performance of the active data is unchanged and may even improve as the primary storage becomes less overloaded.

NO STATIC STUBS

Last-generation solutions (that is, hierarchical storage systems created prior to the advent of cloud) relied on the use of static stubs. A stub, which is a small file that contains the location to which a file has been moved, can be deleted or corrupted, orphaning the files that were moved to the target storage. As file systems grow to modern hyperscale, multi-petabyte size, managing these stubs becomes increasingly onerous, requiring large and complex database management to protect the stubs, because the loss of stubs can leave the data completely inaccessible.

Komprise overcomes this challenge by not using static stubs. Instead it uses dynamic links that point to the Komprise grid's Komprise Access Address (KAA). KAA is a DNS Hostname or IP address used to access the data transferred by Komprise servers by serving as the one address used to address the highly available grid of Observers.

The links are managed by Komprise and can be repopulated if they are accidentally deleted without losing access to the moved data. The dynamic links are backed by a mapping that provides the actual location of the file in target storage. This allows Komprise to further move the files to multiple or distinct locations by simply changing the KAA mapping and without having to change it on the source file server.



Security

Komprise ensures that data is protected and encrypted by default. Komprise provides two security options for moving data to GCP.

ENCRYPTION IN TRANSIT AND AT REST

In this mode of operation, data is transmitted between Komprise observer and GCP using SSL and Google encrypts the data using AES 256-bit symmetric key encryption using Google keys before storing the data. The keys are managed by Google and Komprise never receives the encryption keys. During access, Google decrypts the data and sends it securely over HTTPS using SSL to the Komprise observers. Data is then transferred to end users accessing the data. This is the default mode.

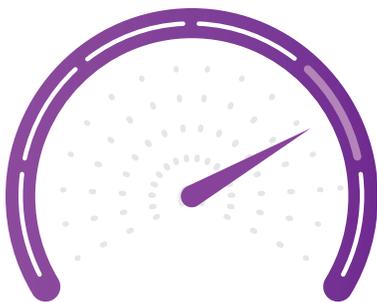
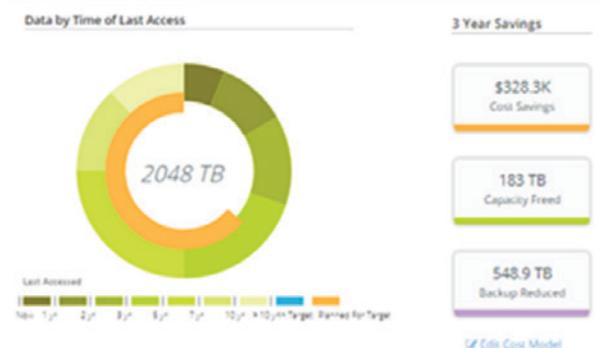
END-TO-END ENCRYPTION

In this mode of operation, data is encrypted on Komprise Observers using AES 256-bit symmetric key encryption before transferring to GCP. During access, the Komprise Observer retrieves encrypted data from Google that is transmitted in encrypted format. The Komprise Observer then decrypts the data using the Data Encryption Key and then sends it to the user. This is increased security mode where data is only available through the Komprise grid and not directly in Google. Therefore, in case of a disaster where all Observers are unavailable, the user must provide the passphrase and encryption key to decrypt and access data.

Value

COSTS

Komprise is priced by the amount of data managed. For ~\$0.005/GB/month, you can have all the features of Komprise including the analytics, data archiving, data replication, data migration, and transparent file-object data gateway. Combined with the cost-efficiency of Cloud Storage, customers can save substantially on NAS storage, DR, and backup costs.



PERFORMANCE

Unlike traditional solutions, Komprise is built from the ground up to manage data at today's scale. Our fully distributed, scale-out architecture grows with your environment—simply add more Observer virtual machines. Komprise stays out of the data, meta-data, and control paths of hot data so there is no performance impact to hot data access. Also Komprise intelligently throttles its performance to run as a background task to minimize its impact and avoid network slow-down.

Use Cases



ACTIVE ARCHIVE

Komprise moves cold data by policy transparently to Coldline without the use of stubs, agents, or any changes to users and applications. Users and applications retain file-based access to data stored as objects on Cloud Storage, continuing to access cold data from the primary NAS exactly as before — unlike array tiering solutions that require access through the source.



CAPACITY PLANNING

Komprise provides visibility across your storage silos so you can see how your data is growing, aging, and being used. Using Komprise, you can see how much of your data is hot or cold and interactively assess the savings and impact of moving data to Cloud Storage.



REPLICATION AND DISASTER RECOVERY

Cloud Storage offers an affordable, highly durable, and available alternative to traditional on-premises backup targets. Using Komprise, you can take a DR copy of only the active data and actively archive the rest. For example, you can choose to keep a DR copy on Nearline of just the data modified in the last six months and have the rest of the data archived in Coldline. Full file access, with all access control list (ACL) permissions, is preserved and honored. Users directly access the copied data from Cloud Storage in a disaster scenario.



DATA MIGRATION

Migrating NAS file data can be a nightmare — Komprise eliminates the errors and the guesswork by automating the migration to Cloud Storage with a reliable solution that is resilient against network and storage glitches. Komprise is built on a scale-out architecture so you can throttle up or down the rate of your data migration as needed and without any dedicated infrastructure.

With the rapid growth in data and its associated storage and management costs, it is imperative that an automated mechanism is used to classify and store the data based on usage profiles and in a way that is transparent to the underlying storage and to the users and applications that rely on the data.

The combination of Cloud Storage with the analytics-driven, automated management provided by Komprise offers a way to analyze and classify the data and manage and store it cost-effectively without disruption.

The combined solution can reduce storage costs by as much as 70 percent while improving the life and performance of the existing NAS and reducing or eliminating the need to buy more on-premises storage.



70%

Savings on Storage

NEXT STEPS

Complimentary Trial

Sign up for a complimentary trial of Komprise in your own environment and see how much you can save by transparently archiving, replicating, and migrating the right data using Komprise.

Contact Us

Call: 1-888-995-0290

Email: info@komprise.com

Visit: komprise.com



Komprise, Inc.
1901 S. Bascom Ave. Suite 400
Campbell, CA 95008
United States

For media requests email marketing@komprise.com
©2018 Komprise, Inc. All rights reserved.