



Komprise Hypertransfer

Migrates Data to the
Cloud **25x Faster**

[White Paper]

An abstract graphic at the bottom of the page consists of numerous curved, glowing lines in shades of blue, purple, and pink, creating a sense of motion and data flow.

Komprise Hypertransfer for Elastic Data Migration creates dedicated virtual channels across the WAN to accelerate cloud data migrations. By establishing dedicated channels to send data, Komprise Hypertransfer minimizes the WAN roundtrips, which mitigates SMB protocol chattiness and dramatically improves data transfer rates. Tests done using a dataset dominated by small files shows Komprise accelerates cloud data migration **25x faster** than other alternatives.

The Barriers to Fast Cloud NAS Migrations

Unstructured data is everywhere. From genomics and medical imaging to streaming video, electric cars, and IoT products, all sectors generate unstructured file data. Data-heavy enterprises typically have petabytes of file data, which can consist of billions of files scattered across different storage vendors, architectures and locations. And while file data growth is exploding, IT budgets are not. That's why enterprises' IT organizations are looking to migrate file workloads to the cloud. However, they face many barriers, which can cause migrations to take weeks to months and require significant manual effort. These include:

- **Billions of files, mostly small:** Unstructured data migrations often require moving billions of files, the vast majority of which are small files that have tremendous overhead, causing data transfers to be slow.
- **Chatty protocols:** Server message block (SMB) protocol workloads—which can be user data, electronic design automation (EDA) and other multimedia files or corporate shares—are often a challenge since the protocol requires many back-and-forth handshakes which increase traffic over the network.
- **Large WAN latency:** Network file protocols are extremely sensitive to high-latency network connections, which are essentially unavoidable in wide area network (WAN) migrations.
- **Limited network bandwidth:** Bandwidth is often limited or not always available, causing data transfers to become slow, unreliable and difficult to manage.

This paper reviews the results of Komprise Hypertransfer performance testing and outlines how Komprise Elastic Data Migration functionality delivers **25x performance gains** compared to other tools.

Komprise Elastic Data Migration with Hypertransfer

Komprise Elastic Data Migration is a SaaS solution available with the Komprise Intelligent Data Management platform or standalone. Designed to be fast, easy and reliable with elastic scale-out parallelism and an analytics-driven approach, it is the market leader in file and object data migrations, routinely migrating petabytes of data (SMB, NFS, Dual) for customers in many complex scenarios. Komprise Elastic Data Migration ensures data integrity is fully preserved by propagating access control and maintaining file-level data integrity checks such as SHA-1 and MD5 checks with audit logging.

As outlined in the white paper *[How To Accelerate NAS and Cloud Data Migrations](#)*, Komprise Elastic Data Migration is a highly parallelized, multi-processing, multi-threaded approach that improves performance at many levels.

First, **multi-level parallelism** maximizes the use of available resources by exploiting parallelism at multiple levels—shares, directories, files, and threads—to maximize performance. Komprise Elastic Data Migration breaks up each migration task into smaller subtasks that execute across the grid of Komprise Observers, which are one or more virtual appliances that run the Komprise Intelligent Data Management platform. This parallelism occurs automatically across the grid of Observers. The user simply creates a migration task and can configure the level of parallelism. Komprise does the rest.

Second, Komprise incorporates **protocol-level optimizations** to minimize the round trips that must be made to filers and to client file system cache. This yields significant gains over the standard uses of the protocols by operating systems and applications.

The third element of performance improvement Komprise Elastic Data Migration provides is **Hypertransfer**. The Hypertransfer update accelerates data migration by creating dedicated virtual channels between local filers and the cloud along which data is packaged efficiently to minimize time-consuming back-and-forth communications. This bundling of operations eliminates the roundtrips the file protocol needs to execute over the network during a file migration and therefore eliminates high-latency chatter, significantly alleviating the per-file overheads imposed by the file protocol. **The result is a dramatic reduction of the transfer time for small files**, which make up the vast majority of real-world workloads. Regardless of network bandwidth, small file performance is often the greatest factor in the overall time to complete data migrations. Hypertransfer is specifically designed to address the slow transfer rate associated with small SMB files. Combined with the innate parallelism of Komprise Elastic Data Migration, Hypertransfer is especially beneficial when moving data over high-latency networks such as WANs.

The rest of this paper focuses on the SMB performance gains delivered with Hypertransfer and compares Komprise Elastic Data Migration results with Robocopy, a command-line file replication tool for Microsoft Windows.

Testing Komprise Hypertransfer vs. Robocopy

Testing compared the data migration utility Robocopy with Komprise. **Figure 1** shows the architecture of a NAS migration from on-premises shares to a cloud file system using Komprise. The easy-to-deploy, fault-tolerant, distributed architecture consists of one or more **Komprise Observer** virtual appliances running at the customer site. The virtual appliances are connected to a **Komprise Director**, which provides the administrative console UI and can run either as a cloud service or on-premises. The Observers open multiple communication channels with one or more **Komprise Windows Proxies** that run in the cloud. The Observers pack multiple file operations per packet and send them to the Windows Proxy. The Windows Proxy in turn unpacks the file operations and executes them on the destination file system.

Komprise works seamlessly across on-premises, edge and cloud file and object storage through open standards (NFS, SMB, S3). When migrating data, Komprise performs automatic retries if the network or storage is unavailable. **It ensures that all permissions and file data are fully migrated and verifies it using checksums.** Performance can be enhanced in multiple ways as needed by increasing the number of threads in an Observer, by adding more Observer virtual machines as well as by adding more Windows Proxies in the cloud.

Furthermore, as can be seen in the architecture shown in **Figure 1**, **Hypertransfer actually adds security to a migration** as all the file communication passes directly from the on-premises Observers to the cloud Windows Proxies through the private Hypertransfer channel. No part of the migration goes directly from on-premises systems to the cloud filers, and therefore the cloud filers themselves do not need to be exposed to any systems or network outside the cloud if that is not desired by the customer.

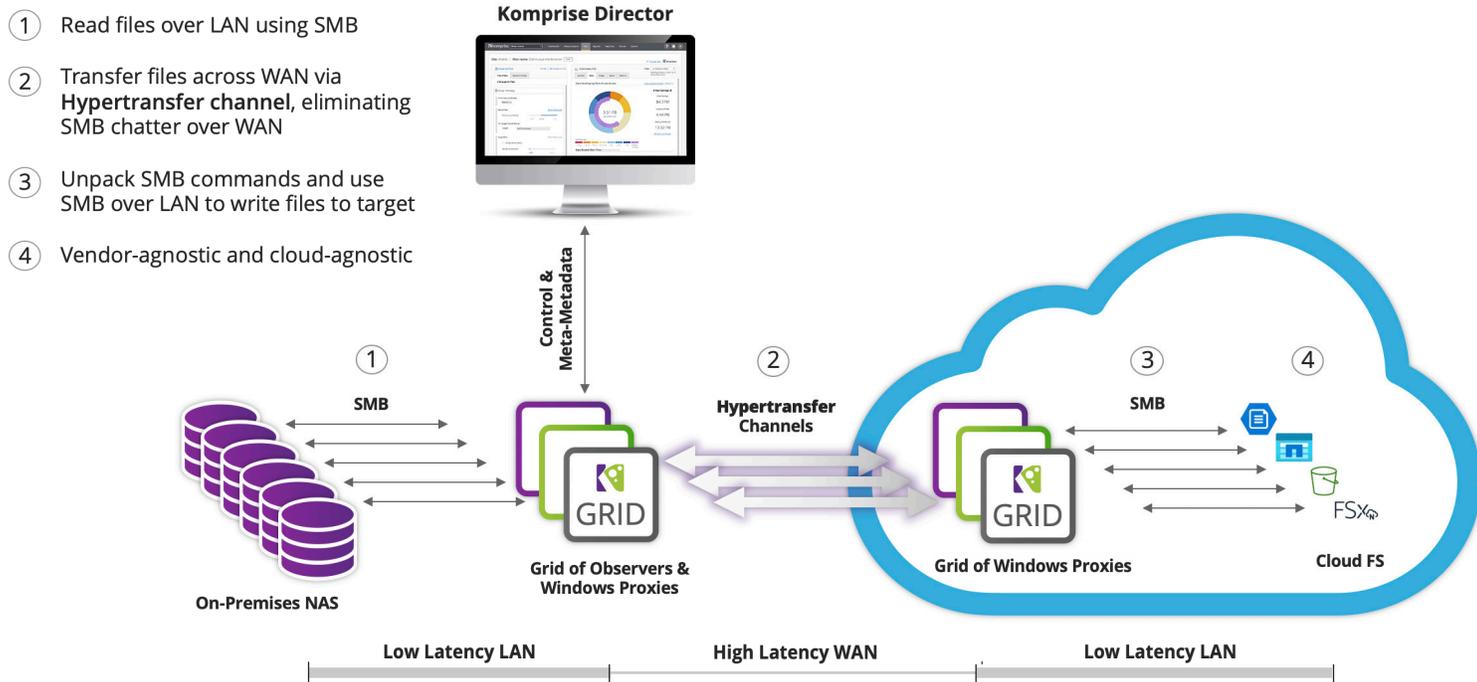


Figure 1: How Komprise Hypertransfer for Elastic Data Migration works.

The Migration Environment

We ran migrations of both a small file and the well-known Android data set using the Robocopy migration utility and the Komprise Hypertransfer Elastic Data Migration solution. The environment in which the tests were conducted consisted of a source file server consisting of 10K SAS drives. The target was an Azure File system (Premium, 4 TB provisioned). A 1 Gbps uplink connection was used to connect to Azure Files.

The Tests

Test 1: 25x Performance Improvement for Small File Migration

The large per-file SMB overhead causes a dramatic increase in migration times of small files over a WAN when compared to a local area network (LAN). In this first test, we focused purely on small files. We compared a default configuration of Robocopy with the Komprise default configuration consisting of 4 Observers and 4 Windows Proxies. The data set consisted of only small files on a single share.

Topology	Details
Data Set: small, single share	<ul style="list-style-type: none"> • Data Set Size: 335 MB, 28,344 files, 4000 folders • File Size Range: 8-20 KB • Average File Size: 10 KB

The Results¹

We ran the tests multiple times and found that **Komprise transfers small files to the cloud 25x faster than Robocopy**. This is a significant improvement and shows the transfer rate gains derived from packing multiple file operations per network communication, along with other SMB-specific protocol optimizations, when transferring over a WAN.

Komprise transfers small files to the cloud **25x FASTER** than Robocopy.



Test 2: 20x Performance Improvement for Android Open Source Project Data Set

The Android data set was chosen because it is representative of directory and file distributions that are difficult to migrate, such as those seen, for example, at EDA companies: a high file count of very small files with some large files.

The Android data set was 74 GB, with the specifics shown below:

File Count	Data Set Size	Directory Structure	File Sizes
990K	74 GB	<ul style="list-style-type: none"> 6 levels deep Files in every leaf directory Includes 139,469 directories 	Mixed, with high count of small files: <ul style="list-style-type: none"> Small: 500 B – 100 KB Medium/Large: 10 MB – 1.5 GB

Komprise Elastic Data Migration provides analytics about the data to be migrated, which is useful for planning large-scale migrations. The following figures are some of the charts from the analysis to show the characteristics of the Android data set.

The **Figure 2** shows the file types in the Android data set sorted by space consumed. “Pack” and “Archive” consume the most space, followed by binaries and source code.

Space consumed by file type

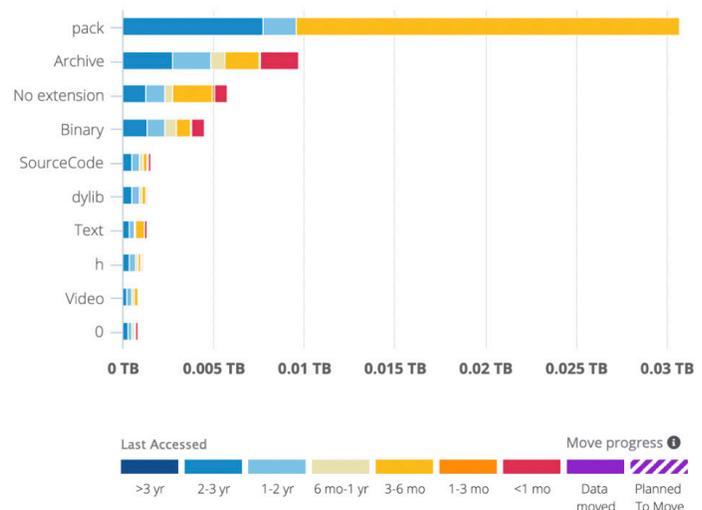


Figure 2: Types of files reported by Komprise for the Android source code.

1. Results were obtained in an isolated, high performance lab environment. Customer experiences may vary depending on a variety of factors, including but not limited to: the specific source and destination file servers, the load upon those servers, the network environment and performance, the specific virtual machines on which the Komprise virtual appliances are run, and the specific data set migrated.

The next two charts show the space consumed by the files based on their size and the count of these files.

Space consumed by file size

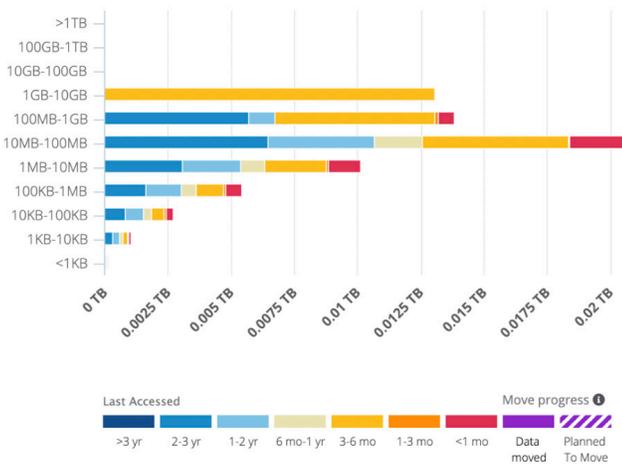


Figure 3: Komprise analysis report showing amount of space consumed by files of different sizes.

Number of files by file size

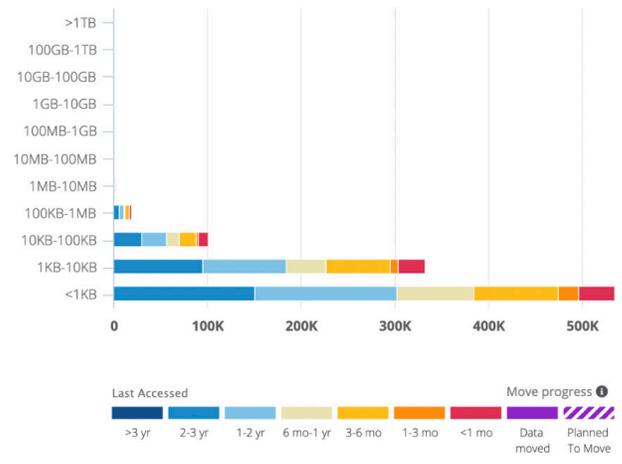


Figure 4: Komprise analysis report showing the number of files by file size.

One can see that the data is a mix of small and large files. As shown in Figure 3, most of the space is consumed by files larger than 10 MB. The amount of space consumed by files smaller than 1 KB is so low it does not register in the graph of Figure 3. Meanwhile, Figure 4 shows that most of the files are actually less than 1 KB! And the number of large files is so few that they do not register in the graph of Figure 4. Note that the Komprise analysis reports not only show the types of files, their sizes and the number of files, but also break down each category by “age” or the time they were last accessed. Komprise uses this information to tier cold files based on user-defined policies.

The Results¹

When tested over the WAN with the Android data set, **Komprise was consistently 20x faster than Robocopy**. In real-world applications there will be large and mid-sized files that do not derive the same benefits from Hypertransfer as do small files. Therefore, the resulting transfer speed gains will not be as large, but because the number of small files is typically much larger, the overall gains are still significant.

A 20x improvement can make a fundamental difference. A dataset that takes Robocopy 20 days to migrate can be **done in just 1 day with Komprise Hypertransfer**. A data set that requires 6 months to transfer via Robocopy will require **just 9 days with Komprise**. If you are paying for professional services during this time, the savings from Komprise will be dramatic. Furthermore, such drastic reductions in transfer times provides **greater flexibility** and truly makes data center consolidations and the move to the cloud feasible.

Komprise was consistently **20x FASTER** than Robocopy.



Conclusion

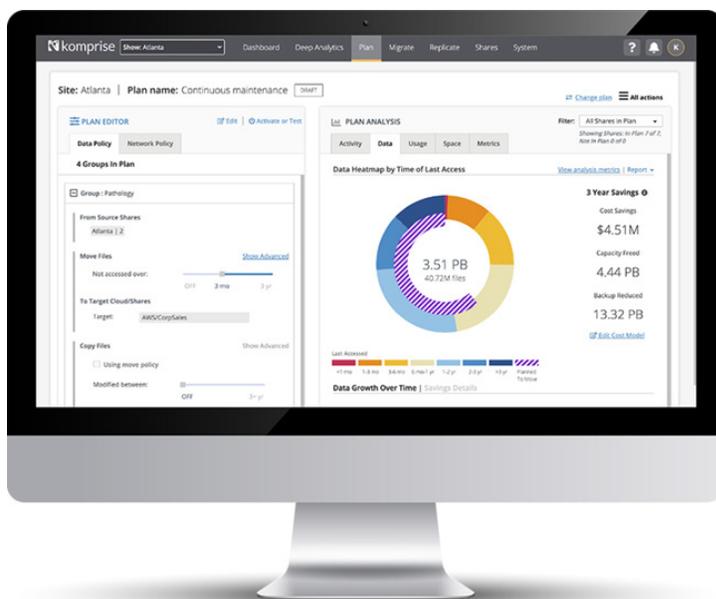
Komprise Hypertransfer for Elastic Data Migration delivers groundbreaking performance improvements for common cloud data migrations with the SMB protocol. Komprise Hypertransfer solves the vexing challenge of WAN migrations of large-scale SMB data sets: chatty protocols causing high overhead, WAN latency and network bandwidth availability issues.

Migrations that used to take 25 days to complete can now finish in a day—but this is not merely about time savings. Consider that migrations take precious person-hours and sometimes professional service hours—you can now get **to the cloud faster with a much lower investment**. Shorter migration windows lower the risk of network outages and other transient errors that make migrations a headache. Komprise also has built-in capabilities to minimize errors and data loss, such as auto-retries and checksum processes to verify that the files transfer correctly. Tools such as Robocopy do not provide this verification and as such are not ideal for large scale, enterprise grade migrations.

Cloud data migrations should be simpler and faster—with less hassle and lower manual effort. Komprise Hypertransfer can help your organization get to cloud value faster while leveraging the many advantages of the Komprise platform. Furthermore, unlike most other migration utilities, Komprise provides analytics along with migration to provide insight into the data being migrated, which allows for better migration planning and ROI.

Elastic Data Migration is available as a superior, stand-alone migration solution, and is also a part of the full Komprise Intelligent Data Management platform.

Learn more: komprise.com/elastic-data-migration



About Komprise

Komprise is a provider of unstructured data management and mobility software that frees enterprises to easily analyze, mobilize, and monetize the right file and object data across clouds without shackling data to any vendor. With Komprise Intelligent Data Management, you can cut 70% of enterprise storage, backup and cloud costs while making data easily available to cloud-based data lakes and analytics tools.



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
© Komprise, Inc. All rights reserved.