

How to Accelerate NAS and Cloud Data Migrations

Komprise Elastic Data Migration

Businesses today are looking at modernizing storage and moving to a multi-cloud strategy. As they evolve to faster, flash-based Network Attached Storage (NAS) and the cloud, migrating data into these environments can be challenging. The goal is to migrate large production data sets quickly, without errors, and without disruption to user productivity. Komprise Elastic Data Migration delivers with a fast, reliable, and cost-efficient migration solution.

This white paper gives a technical overview of the Komprise Elastic Data Migration solution and shares its superior performance results compared with *rsync*.

Why data migrations are hard

Migrating unstructured file data from one NAS to another and to the cloud is no simple feat. It involves file data that consists of the file content and metadata, including permissions and attributes that aren't always represented identically by all the file servers, which creates a complex process. And since the amount of data to be migrated can be huge, it can mean long migration windows and significant effort. For companies that require minimal downtime, the disruption of long windows poses a major problem making migration speed critical for success.

Migrating files fast is easy when the migrated files are large, because the migration overhead is small compared to the time required to transfer the file over the wire. It's a different story when the files are small as the migration overhead can be a multiple of the transfer time.

To address migration to a cloud NAS or to a NAS in another datacenter, modern migration solutions also need to support migrations over WAN connections with acceptable performance. Migrations over a WAN will be much slower than over a LAN due to the higher latency. For these cases, it's important to be able to reduce the back-and-forth protocol "chatter" and automatically partition the task into multiple smaller ones that run in parallel to overcome the performance degradation from the WAN.

A strong migration product not only needs to work across different vendors' NAS solutions quickly, but it also needs an intuitive UI that lets you run, monitor, and manage hundreds of migrations simultaneously. And, of course, it has to be fast to reduce the downtime when IT cuts over from old file servers to the new ones.

Komprise Elastic Data Migration: greater speed and reliability

To address the critical migration issues IT faces today, Komprise has developed Elastic Data Migration. This super-fast migration solution is a highly parallelized, multi-processing, multi-threaded approach that works at two levels:

- **Multi-level Parallelism:** Maximizes the use of available resources by exploiting parallelism at multiple levels: shares and volumes, directories, files, and threads to maximize performance. Komprise Elastic Data Migration breaks up each migration task into smaller ones that execute across the Komprise Observers. Komprise Observers are a grid of one or more virtual appliances that run the Komprise Intelligent Data Management solution. All of 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.
- **Protocol-level Optimizations:** Reduces the number of round-trips over the protocol during a migration to eliminate unnecessary chatter. Rather than relying on generic NFS clients provided by the underlying operating system, Komprise has fine-tuned the NFS client to minimize overhead and unnecessary back-and-forth messaging. This is especially beneficial when moving data over high-latency networks such as WANs.

Analytics with migration for data insights

Unlike most migration solutions, Komprise Elastic Data Migration provides data analytics. This insight into your data makes it possible to properly plan and manage your migrations. The analytics show how much data you have on each volume (share), the age of that data (whether by last modified time or last accessed time), and for each volume charts show what types of files are present, a histogram of file sizes, and space consumed by files of different sizes.

Insights from Data Analytics

- How much data is on each share
- The age of the data
- The types of files
- How much space it takes
- And more

Ongoing value

Komprise eliminates the sunk costs of traditional migration approaches. After the migration, you can use the full Komprise data management platform to maximize your use of the new NAS file servers. Komprise analyzes data growth and usage across your storage to find cold, inactive data, and projects the ROI of moving cold data to secondary storage such as cloud/object. Komprise also moves cold data transparently based on customer-defined policies, so users continue to access the moved data the same as before. Komprise helps organizations reduce over 70% of storage costs while managing data growth.

Testing with the Android Benchmark

Testing was done to compare the data migration utility *rsync* with Komprise.

Figure 1 shows the architecture of a NAS migration 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. Komprise works across on-premises NFS and SMB/CIFS storage and object/cloud storage seamlessly.

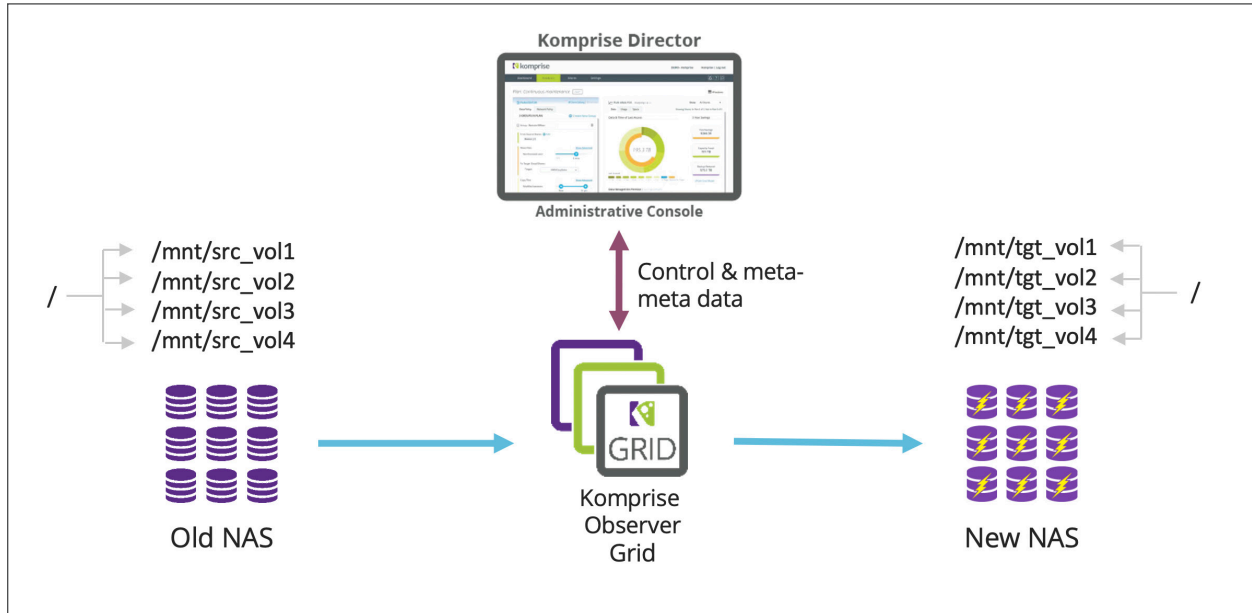


Figure 1.
Komprise Elastic Data Migration Architecture

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 MD5 checksums. Performance can be enhanced as needed by increasing the number of threads in an Observer or by adding more Observer virtual machines.

Data set: Android Open Source Project

The speed of the new Komprise migration was tested using the Android Open Source Project data set¹. The Android data set was chosen because it is representative of the files at typical 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	6 levels deep; files in every leaf directory Includes 139,469 directories	Mixed, with high count of small files and some medium/large files: Small: 500 B – 100 KB Medium/Large: 10 MB – 1.5 GB

1. Android Open Source Project data set was downloaded from <https://source.android.com/setup/build/> to an NFS volume mounted on a source file server.

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

The chart below shows the common file extensions in the Android data set. “Pack” and “archive” are the most common, followed by binaries and source code.

Space consumed by file type

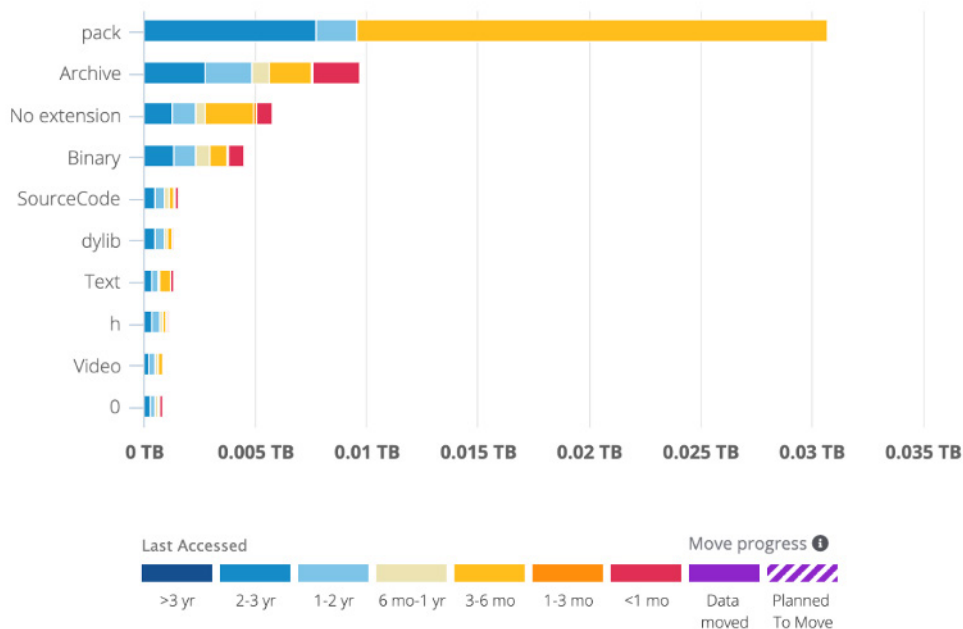


Figure 2. Types of files reported by Komprise for the Android source code

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

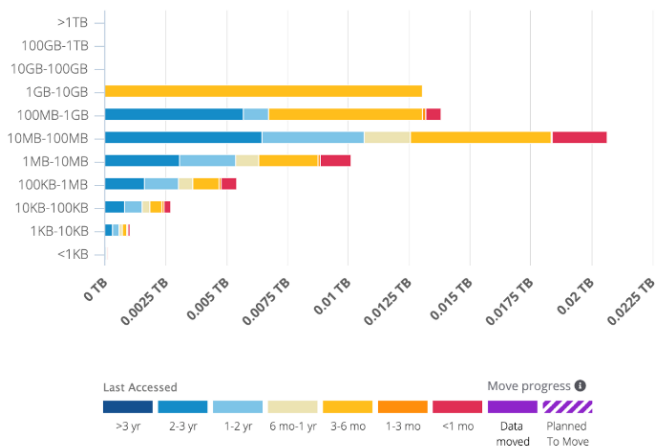


Figure 3. Komprise analytics report showing amount of space consumed by files of different sizes

Number of files by file size

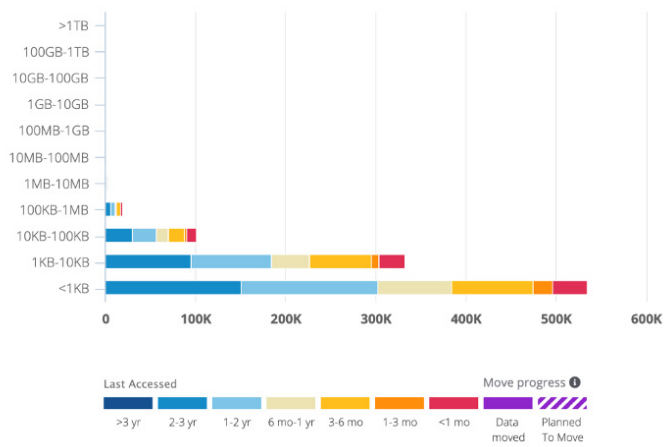


Figure 4. Komprise analytics report showing the number of files for files of each size

As shown in *Figure 3*, most of the space is consumed by files larger than 100 MB but most of the files are small (1 KB–10 KB) and almost all are less than 100 KB (*Figure 4*).

Note that the Komprise analytics reports not only show the types of files, their sizes and the number of files, the reports also break down each category by “age” or the time they were last accessed. This information is used by Komprise to archive cold files based on user-defined policies.

Migration console

Figure 5 is a sample screenshot of the UI of the migration tab. This figure shows a summary of all the ongoing migration tasks and their status. It also shows details of one particular migration task. Note that for this task, it shows details of the multiple migration iterations and details of any errors or issues found during these iterations. A CSV export providing details of a single migration task or all the tasks can be downloaded. This UI makes it easy to manage and report on hundreds of simultaneous migrations.

All migration functions, including creating a migration task, obtaining its status and error reports, and stopping or pausing it, can be done via an API. This makes it possible to integrate Komprise Elastic Data migration into an existing management platform.

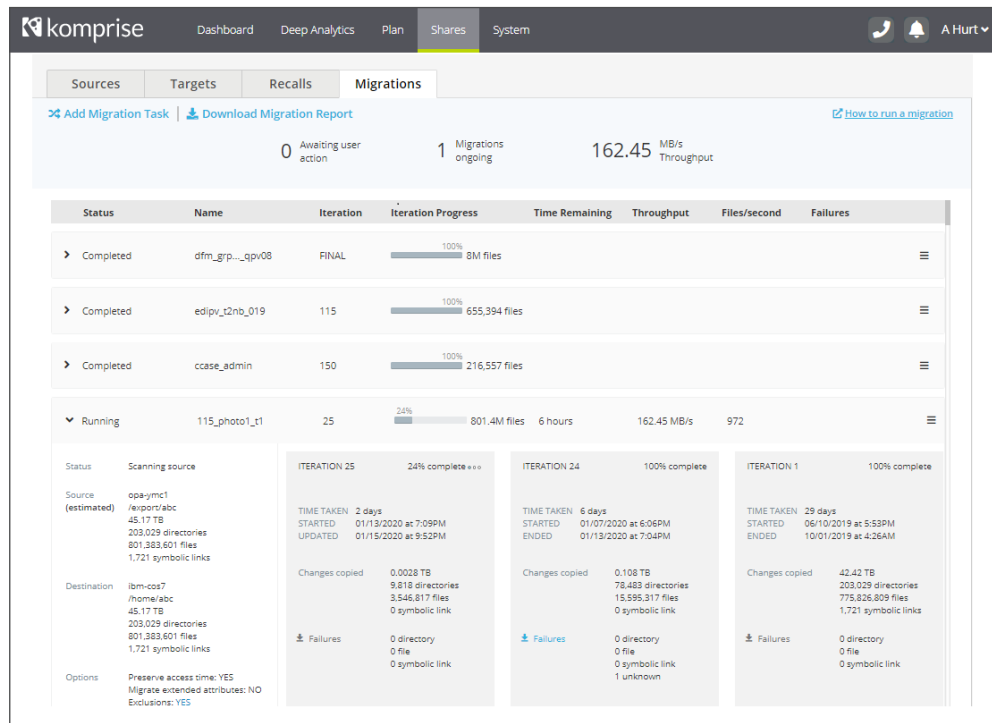


Figure 5.
A sample screenshot showing Komprise Migration tab UI

The migration environment

We ran multiple iterations of migrating the Android data set using the open source *rsync* migration utility and using the Komprise migration solution. The environment in which the tests were conducted consisted of a source file server with 10K SAS drives and a flash-based destination file server. They were connected via a 10 Gbps Ethernet link. A WAN was simulated using a WAN emulator between the migration software and the destination file server. A latency of 30 ms was used to simulate WAN conditions.

The results

When tested on the LAN with the Android data set, Komprise was consistently **27 times faster** than *rsync*². While *rsync* took several hours to complete over a LAN, Komprise completed the migration task in a few minutes.

When we ran the migration tests over the simulated WAN connection, *rsync* could not complete the test even after 48 hours. In comparison, Komprise was still able to complete the task in minutes. As the number of Komprise Observers scaled from 1 to 4, Komprise was able to **cut the degradation** due to higher latencies on WAN **by 250%**.

Komprise beat *rsync*

- 27 times faster
- Cut degradation from WAN by 250%

The table below shows the test results, both with and without using Komprise Elastic Data Migration as WAN degradation relates to the LAN performance.

# of Komprise Observers	Elastic Data Migration	Degradation due to WAN
1	Off	5X slower
2	On	3X slower
4	On	2X slower

² 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 filer 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.

Summary

Komprise Elastic Data Migration provides high-performance data migration at scale, solving critical issues that IT professionals face with these migrations. Komprise makes it possible to easily run, monitor, and manage hundreds of migrations simultaneously. Unlike most other migration utilities, Komprise also provides analytics along with migration to provide insight into the data being migrated, which allows for better migration planning.

While Elastic Data Migration is available as a superior, stand-alone migration solution, it is also a part of the full Komprise Intelligent Data Management platform. With this platform, the sunk cost of migration is eliminated. It can be used to identify and transparently archive cold data with no change in how users and applications access the cold data. This dramatically reduces storage costs without disrupting users or existing data protection workflows. Komprise Intelligent Data Management also provides Deep Analytics, which allows you to create a metadata lake of all your files, enabling company-wide search and tagging of your data.

The choice is yours, but either way, you'll never look at a NAS migration the same way again.

Learn More

Go to [Komprise.com](https://www.komprise.com) to learn more about how analytics-driven data management from Komprise can help your organization save time and costs.



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](https://www.komprise.com)

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