

Getting Smart about Data Growth with Intelligent Data Management

KOMPRISE IDENTIFIES HOT AND COLD DATA ACROSS MULTIVENDOR STORAGE.

With the amount of unstructured data more than doubling every two years, it's clear that organizations need to come up with new strategies to handle their data more effectively. The primary challenge is that businesses manage all of their data in the same way, regardless of importance. This results in businesses expanding their Tier 1 storage footprint, increasing their backup windows, and incurring rising infrastructure costs. How can organizations get in front of these issues without disrupting users?

The scope of the problem is hard to overstate. In a recent report, IDC predicts the collective sum of the world's data will grow from about 29 zettabytes (ZB) in 2018 to 163 ZB by 2025, a growth rate of 66% per year.¹ To put that in context, consider that a ZB is approximately equal to 1,000 exabytes, a billion terabytes, or a trillion gigabytes. Now multiply that trillion GB by 163.

IDC also found that of the 13ZB of installed storage expected to be in place in 2025, only 7.5ZB will actually store data.² As a result, businesses will over-provision storage by nearly 50% because they lack data visibility into how data is growing and being used.

As IDC notes, the amount of storage capacity will grow by 300% in the next seven years, but IT budgets are staying flat. With these flat budgets and amassing data growth, businesses can no longer treat all data the same. They need to identify hot and cold data and store them on different classes of storage. New ways to effectively manage all this data are imperative.



Sponsored Content



Why legacy solutions fall short

The existing approach to dealing with data growth has been to simply add more capacity as needed. Enterprises have continually added hardware and software over the years to accommodate their data growth needs, most likely from different vendors.

Often these systems can't adequately scale to keep up with the highly virtualized and converged (or hyperconverged) environments. The management overhead of backups and disaster recovery required for growing environments creates a drag on performance, resulting in slowdowns. Backup windows become ever larger, causing consistency concerns and reducing reliability.

Legacy data management approaches to archive cold data require users to change behavior, and users are frustrated when they can't transparently access moved data. Most important, legacy systems are costly, requiring expensive enterprise licenses and investments in an increasing amount of infrastructure—all for a solution that under-performs. It's an unsustainable approach during a time when, as the IDC study shows, data growth is orders of magnitude higher than the storage budgets.

Not all data is mission-critical

Compounding the problem is the fact that most organizations have an inordinate amount of "cold" data—or data that nobody is actively using. Within months of its creation, anywhere from 45% to nearly 90% of data becomes cold, depending on the vertical industry (see figure below). With no easy way to identify and move cold data without disrupting users, organizations end up storing and managing it in the same way as active data.



Figures aggregated from actual **Komprise** customer data

Don't replicate cold files

This has an enormous impact on storage costs because most companies back up their data multiple times. For example, if you have 1PB of data in primary storage, you're likely paying for:

- **1PB replication/mirror:** Storage is the same cost as your primary storage
- **3PB backup storage:** Typically, businesses have three to five backup copies of their primary storage data
- **1PB backup software license:** You need to license the backup software on the entire footprint

So, on 1PB of data, the typical organization pays for 4PB of storage on backup and disaster recovery, and 1PB of backup license. If you estimate these costs for your own environment, it's likely you'll find these figures ring true.

It's imperative, then, for organizations to address the vast amounts of cold data in their environments to get storage, backup, and costs under control.

¹ Source: IDC Data Age 2025 Study, sponsored by Seagate

² Ibid

Komprise has a solution

Addressing this cold data is a problem that requires advanced analytics to solve. Komprise Intelligent Data Management uses such an approach by analyzing data usage across your entire storage environment and showing you what data is hot (in active use) and what data has gone cool or cold. It enables you to conduct “what-if” scenarios to understand the projected impact on your data footprint and the return on investment (ROI) from moving inactive data to secondary storage, such as the cloud. Based on these scenarios, you define policies around data storage, and Komprise moves data based on those policies—all transparently to users, who can still see and access files in secondary storage just as they did when it was active.

Komprise Intelligent Data Management is built on three key pillars, as follows:

DATA GROWTH ANALYTICS

Data Growth Analytics identifies hot and cold data across multivendor storage environments and enables cost-effective planning and management of storage and backup. Simply download the Komprise Observer, and within 15 minutes it provides detailed analytics on how much data you have, how it’s being used, who is using it, and how quickly it’s growing.

An interactive visualization tool lets you conduct “what-if” scenarios to understand the project impact of proposed policies. You’ll quickly understand how much additional storage capacity a given change will yield, how much it will reduce backup requirements, and the resulting cost savings. It’s a no-risk way to plan capacity and determine the most effective data-management approach before actually moving any data.

TRANSPARENT MOVE TECHNOLOGY™

Once you decide on appropriate policies, Komprise’s patented Transparent Move Technology™ offloads cold files from the actively managed footprint to an appropriate secondary storage. Komprise uses no proprietary agents or static links on the storage system, mechanisms that often cause problems such as static stubs.

Komprise uses a fault-tolerant, highly available architecture that handles failures automatically, so data migrations run reliably. It preserves file hierarchy, NTFS permissions, and all metadata to ensure the integrity of the migration. Komprise also uses highly resilient dynamic links to files in secondary storage. The result is users and applications can continue to access cold files transparently just as they did before the migration, with no changes to what users see in their file systems. And, Komprise is outside the hot data and metadata paths, so there is no degradation of performance.

DATA ACCESS ANYWHERE

As your IT environment evolves to meet ever-changing business requirements, Data Access Anywhere enables your data and storage infrastructure to evolve with it, regardless of whether it’s on-premises, in the cloud, or in a hybrid environment. Data moved by Komprise can be accessed natively from anywhere without having to go back to the source storage and without requiring Komprise. This eliminates lock-in.

Komprise is also built on open standards, making it “storage agnostic” and able to work in any environment. It employs a share-nothing architecture that grows on demand. Simply add more virtual appliances as needed to keep up with growth in your managed data environment. The Komprise architecture has no centralized bottlenecks and uses no agents, static stubs, or central servers that limit scalability and present single points of failure.

Komprise Intelligent Data Management uses an approach analyzing data usage across your entire storage environment and showing you what data is hot (in active use) and what data has gone cool or cold.

Customers highlight value of Komprise

Of course, there's no better testament to the value of any IT solution than from the customers who use it.

Steve DeGroat, Enterprise Storage Manager for an Ivy League university, voices an issue any of his counterparts can relate to. "In the history of [the university], the 300 years we've been around, nobody has ever deleted a file," DeGroat says. "So we have quite a bit of data to analyze and manage."

After consolidating data storage on NetApp appliances, he brought in Komprise to help with data management and analytics. In addition to handling data migration, this helps DeGroat plan for how much capacity he needs at each storage tier. It also enables him to generate reports that show various departments how much data they're storing, how it's being used, file sizes, cold files, and more.

"This is visibility that did not exist before," DeGroat says. "It lets us partner with the business in getting the data in the best location. It saves them money, it saves us money."

For its part, Boone County, Indiana, expects to cut its storage costs by 70% over five years by transparently archiving cold data to Microsoft Azure Blob Storage. It did so after a Komprise analysis showed 83% of the county's data was cold, having not been accessed in six months. Most of that data is from body and dash cameras that drove a 3,000% increase in Boone County's evidentiary data in five years. Moving all that cold data off the SAN also enabled the county to move to a smaller, all-flash SAN for hot data storage – at 42% of the cost of the previous SAN, for a savings of 58%.

Gain efficiency, lower costs

Dealing with a 3,000% increase in data while actually lowering data storage costs is quite an achievement. But it's one that Komprise is delivering on time and again.

Komprise offers a simple way for organizations to efficiently manage data sprawl. Its analytics-driven data-management software enables you to quickly identify inactive data and assess the ROI of moving it to a lower-cost cold storage platform. It also enables the migration of the data, whether to on-premises or cloud storage, with complete transparency to end users, who still access the data the same way they always did. With no hardware to deploy, Komprise is fast and easy to install, with no storage agents or stubs to deal with. It works without creating performance degradation and can scale as needed to keep up with your data growth.

Take a fresh look at how you manage data and find out whether cold data is eating up valuable resources.

Visit www.komprise.com to learn more and schedule a free demo.

Moving all the cold data off the SAN enabled the county to move to a smaller, all-flash SAN for hot data storage – at 42% of the cost of the previous SAN, for a savings of 58%.

— Boone County, Indiana

