

Why Data Growth is Not a Storage Problem

Know your data before you act to drive better choices.

Executive Summary

Explosive data growth requires a re-think of how data storage is managed. Storage capacity is running out, backups are taking longer, and budgets can't keep up with the unstructured data deluge. The focus needs to shift from a storage issue to managing the data in your storage. Because treating and storing data as if it's all the same will cost you plenty.

It's time to stop reacting to data growth and start managing it smarter. That takes data insight: how much do you have, what kind, who's using it, and how fast is it growing? You must know your data first to make the best decisions for it and your company.

Storage vendors give you fast storage and lots of capacity. Their feeds, speeds, IOPs, and throughput come compliments of their proprietary technology, which makes it challenging to easily make informed storage decisions across vendors or to easily access your moved data. Should any vendor have that much control over your data?

Komprise Intelligent Data Management puts you in control of your data. Our analytics-driven approach works across all storage vendors and backup architectures giving you a single data management pane. Get instant insight into all of your data—wherever it resides. See patterns, make decisions, make moves, and save money—all without affecting user access to any data.

TABLE OF CONTENTS

| | |
|--|----------|
| Executive Summary | 1 |
| Dynamic Data Analytics | 2 |
| Transparent Move Technology | 5 |
| Direct Data Access | 7 |
| Summary | 9 |



There are three key areas that set Komprise apart, offering a different kind of data management solution that puts data control where it belongs—with data owners:

1. **Dynamic Data Analytics** — Get data insights across all your storage to make informed storage and backup decisions and get the most value from your data.
2. **Transparent Move Technology (TMT™)** — Move your cold data to less-expensive storage without any interference to apps, users, or hot data.
3. **Direct Data Access** — Keep in control to directly access, index, and search data anytime, anywhere—independent of any storage, backup, or data management solution.

Knowledge is power, and having knowledge and power over your own data should be every company's right. Komprise provides insights into your entire data landscape, so you can make critical decisions to move and leverage the value of your data. No proprietary interfaces, no vendor lock-in—just empowered to make your own choices about your own data, now and into the future.

In this paper, we explore the three key ways Komprise provides this flexible capability in more detail.

Dynamic Data Analytics

Know your data before making data decisions.

At today's unstructured data growth rate, making decisions without first understanding more about it can be a costly undertaking. You first need data literacy, which is the ability to read, work with, analyze, and argue with data. Komprise helps you achieve data literacy by enabling you to:

- Analyze your data across all your storage to understand who, what and when
- See trends and determine how fast your data is growing
- Explore different scenarios to forecast storage capacity and strategy

Most vendors require you to buy their solution and move your data before giving you the analytics. Komprise lets you think first, then act on your data management by providing the analytics **in place**, across your storage first.

Dynamic Data Analytics provides all the information you need for data literacy, which is key to handling your

Data Literacy

- Analyze data
- See trends
- Explore scenarios

By 2020, 50% of organizations will lack sufficient AI and data literacy skills to achieve business value.¹

-Gartner

1. [A Data and Analytics Leader's Guide to Data Literacy, Kasey Panetta](#), February 6, 2019.

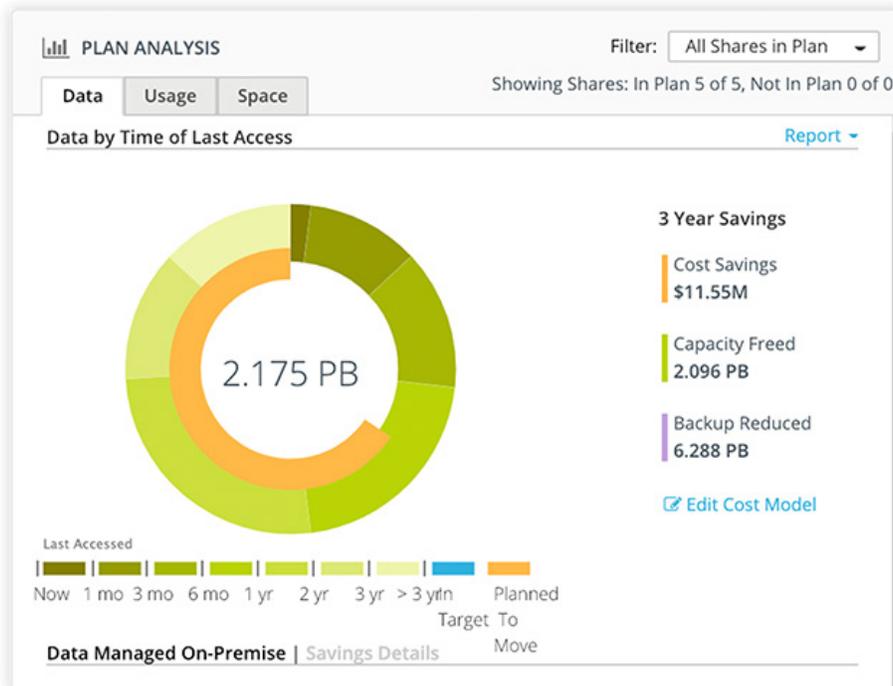


Figure 1: Easily estimate cost savings of different data storage decisions.

unstructured data deluge.

Three ways that Dynamic Data Analytics helps manage unstructured data:

1. Analyze all your data before investing, copying, or moving

Storage and backup vendors that offer analytics require your data to be on their hardware first before they'll give you these insights. This eliminates the flexibility and freedom of true data literacy because you're forced to make costly investment decisions and move or copy data *before* you've even had a chance to first understand it. (If you could, you probably wouldn't make the same decisions.)

Komprise gives you the power over your own data by letting you analyze it across your storage without first requiring a move or copy. We do this by connecting to your storage and analyzing your data via standard protocols, such as NFS, SMB or S3. This allows Komprise to analyze your data in-place without needing to import it into some proprietary format. No exclusive interfaces, agents, or clients on your storage—just a standards-based approach that gives you critical insights into your own data—wherever it resides—allowing you to make the best, most-informed decisions.

2. Analyze usage metrics in hours, not weeks

Once you can analyze your data, the next step in data literacy is discovering data trends. These trends are core to planning your storage strategy. Most data analytics solutions take weeks to crawl and index all your data before they can provide insights on billions of files and petabytes of data. But you

Run "what-if" data scenarios and get an instant cost analysis.

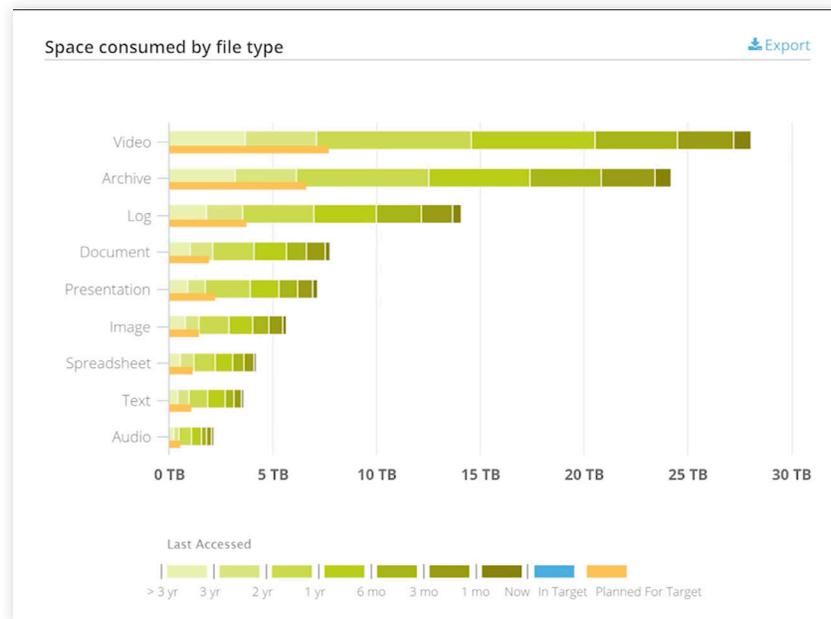


Figure 2: See what kind of data is being used how often.

don't have that kind of time, and you don't need that kind of granularity to make an intelligent storage plan for your data.

Komprise delivers Dynamic Data Analytics in hours—even on billions of files—using patented data analytics and aggregation techniques. It simplifies building your data management plan by allowing you to:

- Find file sizes, top groups, and when they were last accessed
- Run “what if” scenarios and get subsequent capacity needs and cost savings in seconds.
- Manage and move your data the way you need, to save the most

Want to know what would happen if you moved all data untouched in over a year to the cloud? Get an instant analysis based on your data, your costs, and historical data growth patterns. This is the end goal of data literacy—to know, analyze, and be able to work with and “argue” with your data.

3. Dig deeper, creating virtual data lakes with Deep Analytics

When you need to dig into your data beyond trends, Komprise Deep Analytics can help. We provide an intuitive way to search and find specific files that fit your exact criteria across all your storage.

Simply build your specific queries and Komprise Deep Analytics shows you both summary information and detailed reports on the files that fit your criteria. You can tag the data you find as well as files being created, and as you learn more about other files' content. This dynamic approach allows you to then run queries based on your tags and build real-time virtual data lakes on the fly, without having to first move the data. You can continually leverage these data lakes for applications like Big Data, AI, and ML. This is available both via a user-interface and an API.

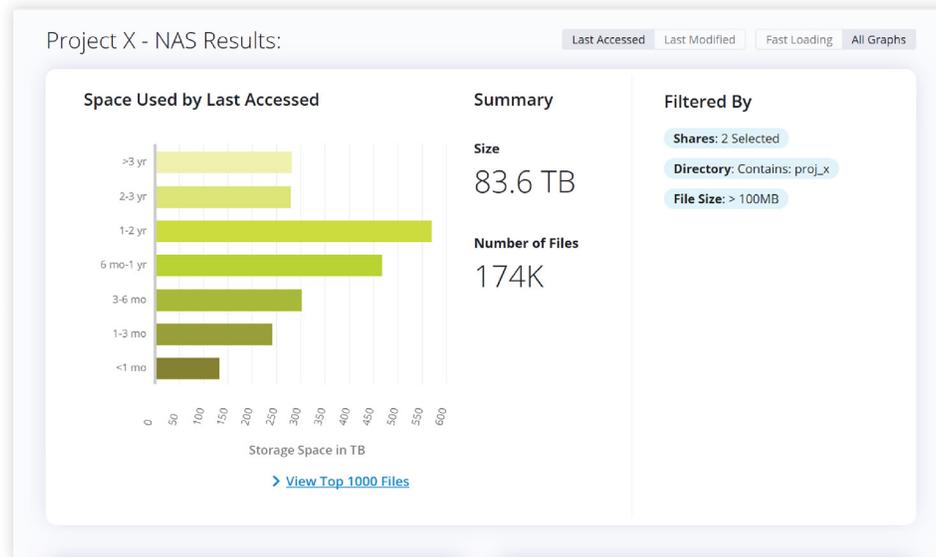


Figure 3: Drill down for detailed reports filtering by specific criteria.

Transparent Move Technology

Save costs and move data without interference to apps or hot data access.

Once Dynamic Data Analytics gives you the insight to make the right decisions, it's time to take action. Because our Transparent Move Technology (TMT™) works seamlessly with any storage and backup solution, you can make the data moves you want when you want, without disrupting user access.

TMT enables you to create the logical architecture you want without disrupting the physical architecture you have. This gives you the best of both worlds, empowering you the agility to evolve rapidly without creating user friction and hurting business productivity. For instance, you can move to a cloud architecture without eliminating your file-based NAS and without disrupting user access. Komprise achieves this virtualization without creating any interference to hot data or metadata through the patented TMT technology.

Most of our customers discover that ~75% of their data is cold. TMT lets your archive that data off primary storage so you can:

- Save on storage and backup license costs
- Cut your backup time
- Enhance performance/availability of hot data

Data rehydration eliminates storage savings.

If you're using proprietary tiering solutions, third-party storage or back up, you must rehydrate the data. You'll need the full capacity of your existing data to rehydrate, therefore significantly reducing any cost savings.

TMT makes moving cold data to less expensive storage simple. You can literally make your data moves with a click of a button, without any disruption to users, applications, data protection workflows, or access to mission-critical hot data.



Hot data access should be a given, but vendor lock-in can make this an unexpected challenge. The truth is, controlling access to moved data is part of these vendors' business model. Their proprietary data management approach requires users to access moved data by going through their hardware for all data access—and in most cases *hot data is affected*. TMT gives seamless user access to any moved data, at any time—the way it should be.

Controlling access to your moved data is part of storage vendors' business model.

TMT addresses the following key issues critical to archiving cold data:

1. No disruption to users, apps, and data protection workflows

Users shouldn't know that their data's been archived, no matter where it's been moved. They should be able to go where they've always gone—same source, same directory—and access it the same way, even if it's in the cloud.

Most data management solutions don't provide this kind of data access transparency. End users have to use a separate application or generate an IT support ticket to find their archived data, which creates an unnecessary drag on productivity and a burden on IT.

Komprise delivers transparent access by using standard protocol constructs when moving data. When you move a file, a symbolic

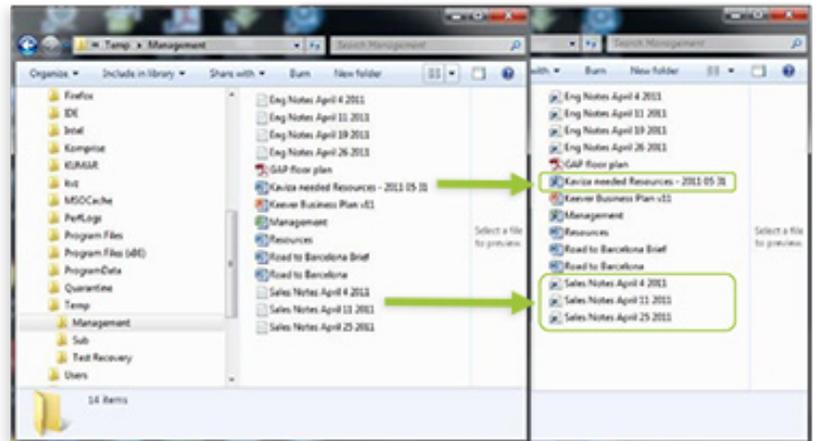


Figure 4: Users are unaffected, accessing moved data exactly as before.

link containing all the properties of the original file is left behind as a pointer. Users and apps continue to see and can open the file from the original location keeping all the permissions and access control intact. No invasive agents or stubs means no disruption to users, applications, or the data protection workflows.

This transparency applies to wherever your data resides. TMT is leveraged universally and seamlessly across all storage tiers, including cloud, object storage, and backup.

2. Works at the file level, not block level, with full metadata fidelity

Storage vendors are now using block-level tiering to move data out of the file server and into an object or cloud tier. 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.



In theory it makes sense, but the problem is that moved blocks can no longer be directly accessed from their new location, such as the cloud. The moved blocks are meaningless without all the other data blocks and file context and attributes (the file's metadata). Block-level tiering forces users to access moved data through the vendor's hardware or software—direct file access is no longer an option. It also involves the problem of rehydration, which *Direct Data Access* solves.

File-level tiering is a more advanced technology that is standards-based. It means the *file and all its metadata* moves to the new tier. It maintains full file fidelity and preserves all the attributes and metadata along with the file at each tier. Applications that rely on attributes of the file are unaffected. And users can access data directly from the target storage and be able to return it to the source storage exactly as they did before.

3. Zero access interference to hot data or hot metadata

Many solutions sit in front of your primary storage and divert requests for the cold data to another location. In general, these solutions promise some form of data virtualization or metadata offloading. But being in front of your NAS storage impacts the performance of the hot data since it introduces a middle man. A “traffic cop” now directs data access, which is a tremendous risk for all your company's data. A failure in this system creates an access nightmare. Just as if the traffic cop took a break, you'd have a major logjam—and *when data access is lost, you'll hear far worse than honking horns*.

This approach also requires that scaling based on hot data access rates rather than cold. Since hot data is 99.999% of your data access, this means that the “man-in-the-middle” device must be able to handle the massive hot data traffic requests. As this traffic increases with your data growth, the “man in the middle” must scale accordingly and still handle access spikes. If you don't plan accordingly, you'll decrease performance on your new flash storage. These types of solutions—and their issues—have been around for years losing customers due to high cost, decreased performance, and risk.

Direct Data Access

Puts you in control of your data, not your vendor.

With the advent of big data analytics, AI, and ML, the ability to index, search, and operate on all your data is a game changer. It's also a fundamental requirement for ever-increasing compliance and legal data-hold requirements. It's important to have direct data access to all your data without going through the original source.

Komprise makes this possible because it transfers your company's NAS data to a cloud or secondary storage target as a file with its complete metadata, whether that target is an on-premises storage (NAS or object storage) or cloud storage. The file is stored in the format native to the target storage device, and the metadata is stored in NFS or SMB format, depending on the protocol used to archive or copy the data.



This approach has many key benefits:

1. Access data from any tier without going to the source

Data needs to be accessed through the original source device. When you move your data with Komprise, it can be directly accessed from the target device using its native protocol.

Let's say you move data to the cloud using Komprise. You can now run applications in the cloud that use that data directly through S3 or via NFS or SMB. This ensures your data doesn't get locked into any one storage or backup solution—or to Komprise for that matter. We believe that your data should remain in your control through its lifecycle.

2. Access data in native format

At Komprise, we believe you should own your data and not be locked into a proprietary solution. When we archive or copy your data, we put it in a form that's native to that storage system. This not only enables new uses of that data, it ensures that archived data is not locked away and can be continually used to extract value. You can use third-party applications or Komprise applications to extract that value—Komprise never locks you in. We preserve all the standard metadata and the extended metadata, such as tags with the data wherever it moves, so your files retain their full context and remain usable wherever they go.

3. No rehydration necessary

Block-level tiering, used by many data management solutions, requires rehydrating archived data before it can be used, migrated, or backed up. This approach negates much of the benefit of data archiving in the first place. For instance, if you archive 75% of your data but must rehydrate it when backing up, you've saved nothing. Or if you want to end-of-life your 1PB storage system from which you've archived 3PB of data over its lifetime, you'll need to rehydrate all 3PB before you can migrate off that system.

Komprise's file-based tiering eliminates these rehydration issues.

- Symbolic links left on the source when Komprise archives files are understood by other applications.
- Backup software will backup the symbolic links without rehydrating the files they point to.
- Restores will restore the links, which still point to the same files, so third-party backup applications will function without rehydration.
- When migrating data from the source, the symbolic links get migrated without getting rehydrated.

By using industry-standard constructs, Komprise seamlessly operates with your storage and backup solutions and other applications without requiring any customization.



Summary

Komprise is an analytics-driven data management strategy that allows you to know your data before making decisions about it. Unlike others, Komprise is dedicated to data autonomy, allowing data owners to analyze, manage, and move their own data, free from proprietary technologies. We allow customers to transcend storage silos, storage vendors, and storage technologies, providing the much-needed flexibility to scale as your data grows—wherever it goes—and to be responsive as needs change. Stay in control of your data regardless of the next storage evolution. *Komprise pays for itself by always optimizing your storage and unlocking the full potential of your data.*

Your company's greatest assets are its people and its data. Komprise empowers your team with the right data insight to make the right data moves for your company's goals.

Learn More

Experience a complimentary trial of Komprise in your own environment to see the true picture of your data and how much you can save. Visit: [komprise.com](https://www.komprise.com)



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