Persistent data growth is straining IT budgets, causing more organizations to prioritize cloud data migrations but data visibility, planning and management across hybrid clouds remains a key roadblock.
# Table of Contents

- Executive Summary .......................................................... 3
- Report Highlights .................................................................................................................. 4
- Persistent Data Growth Strains IT Budgets ............................................................ 6
- Hybrid Cloud: Cloud Data Migrations a Key Strategic Priority ....................... 8
- Lack of Visibility is Slowing Hybrid Cloud Adoption ........................................... 10
- IT Leaders Want Analytics to Inform Storage Decisions .............................. 12
- Conclusion: Top Takeaways ................................................................................ 15
- About Komprise .......................................................................................................... 16
Executive Summary

Data growth, particularly of unstructured data, is continuing to accelerate, straining IT budgets and causing organizations to prioritize data migration to the cloud. As complexity continues with the dominance of hybrid cloud, enterprise IT organizations report a lack of visibility for planning—indicating a fundamental barrier to cloud data migration and management. The leading strategy for IT directors is to invest in analytics-driven tools to aid decision-making and optimize spending. Many also are looking to implement a systematic, data-centric approach versus an ad hoc, storage-centric approach to data management to be more agile and efficient. Further, as executives seek to monetize the petabytes of data in storage, IT leaders are keen on enabling data lakes in the cloud for low-cost research and analysis.

The 2021 Komprise Unstructured Data Management Report examines the challenges and opportunities with unstructured data in the enterprise—from how much data enterprises are managing, to cloud data priorities and future approaches for data management. This report summarizes responses of 300 global enterprise storage IT decision makers at companies with more than 1,000 employees in the United States and in the UK. All respondents work at the IT manager level or above, across IT/technology operations teams. The survey was conducted by a third party in June 2021.
Data Growth & Spending: Unstructured Data is Growing

- **Over 60%** of enterprises are managing more than 1PB of data and **nearly 40%** are managing more than 5PB of data.

- **65%** of organizations spend more than 30% of their IT budgets on data storage and management: One-third (33%) are spending 30-40% of IT budget on storage and data protection, while 31% are spending more than 40%.

- **62%** will spend more on storage in 2021 versus 2020.

- **One-third (33%)** of enterprises acknowledge that over 50% of their data is cold, while nearly 20% don’t know.

Storage Modernization: Getting More Data to the Cloud is a Key Priority

- **50%** of enterprises have data stored in a mix of on-premises and cloud-based storage.

- **43%** are moving to or buying more cloud storage and **40%** are investing in modernizing backup and disaster recovery systems in the next 12 months.

- **Nearly half** will spend more than 50% of IT budgets on cloud data storage in the next two years.
Unstructured Data Management Goals & Challenges:

- Will invest in analytics tools: 45%
- Wish to avoid rising costs: 44.9%
- Want better visibility for planning: 44.5%
- Are interested in tagging data for future use and enabling data lakes: 42%
- Want flexibility to switch storage vendors without lock-in: 40.5%
- Want to be agile and deliver storage-as-a-service to business units: 40%
- Want a systematic data approach that is not ad hoc: 35.9%
- Say that data is growing too fast and too expensive to store and backup: 34.9%
- Lack an effective tiering strategy and related tools: 33.6%

Cloud Data Priorities

- 56% Migrating data to the cloud
- 46% Cutting storage and data costs
- 41% Governance and security of data

©2021 Komprise, Inc. All rights reserved.
I. Persistent Data Growth Strains IT Budgets

*Unstructured data is growing too fast and getting too expensive to store and backup for many large organizations.*

Enterprise IT organizations are collecting and storing petabytes of data: nearly 63% are managing more than 1PB and nearly 40% are managing more than 5PB. Consider the size of 1 petabyte, which is equivalent to over 4,000 digital photos per day over your entire life. Most of this data is unstructured, across documents, user data, video and image files, research data, design data and other application data.

According to research from IDC, 500 million new logical applications will be created by 2023—the same number created in the past 40 years. These applications will generate massive volumes of structured, semi-structured, and unstructured data.

How much data are you currently managing?

<table>
<thead>
<tr>
<th>Data Size</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;500TB</td>
<td>16.6%</td>
</tr>
<tr>
<td>500TB to 1PB</td>
<td>20.6%</td>
</tr>
<tr>
<td>1PB to 5PB</td>
<td>23.9%</td>
</tr>
<tr>
<td>5PB to 1PB</td>
<td>14.6%</td>
</tr>
<tr>
<td>10PB to 50PB</td>
<td>14.3%</td>
</tr>
<tr>
<td>50PB+</td>
<td>10.0%</td>
</tr>
</tbody>
</table>

What types of unstructured data do you have?

<table>
<thead>
<tr>
<th>Type of Data</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>65.4%</td>
</tr>
<tr>
<td>User data</td>
<td>58.8%</td>
</tr>
<tr>
<td>Research data</td>
<td>50.5%</td>
</tr>
<tr>
<td>Video and media</td>
<td>49.2%</td>
</tr>
<tr>
<td>Design data</td>
<td>38.5%</td>
</tr>
<tr>
<td>Other application data</td>
<td>27.9%</td>
</tr>
<tr>
<td>Genomics data</td>
<td>20.9%</td>
</tr>
<tr>
<td>Medical Imaging data</td>
<td>20.3%</td>
</tr>
</tbody>
</table>
As the Pace of Digitization Continues Unabated, Enterprises are Spending High Percentages of IT budget on Data Management

While this makes sense in one regard—data is like gold in the modern enterprise—there is ample room for improvement to eliminate waste:

• The majority say that IT budget spending on storage, backups and disaster recovery is between 30 and 40% and significant numbers (21%) are spending between 40 and 50%.

What percentage of your IT budget is spent on data storage and protection? Will you spend more on data storage in 2021 versus 2020?

Key Insight:
63% of companies are already managing over 1PB of data with 30%+ IT budgets spent on data storage and backups—with most organizations expecting these costs to go up in 2021. More efficient ways to manage data is becoming critical.
II. Hybrid Cloud: Cloud Data Migrations a Key Strategic Priority

*Hybrid cloud storage is a leading priority as organizations want to move more data to the cloud, but lack of visibility is slowing them down.*

Amid this landscape of explosive data growth, enterprises have more choice than ever as to where and how to store data. They can choose from a variety of on-premises storage technologies across NAS, SAN, SDS, HCI, DAS, Object and Tape, or go to the cloud and use object storage (such as Amazon S3, Azure Blob or Google Cloud Storage) or file storage (such as Amazon EFS). Given this fragmented market and the overall acceleration to the cloud since the onset of Covid-19, hybrid strategies are becoming mainstream. IT leaders are opting to right-place data and workloads to meet distinct requirements for cost, performance and agility.

**The Komprise survey gives further evidence to the prevalence of hybrid cloud storage:**

- Half of enterprises have data stored in a mix of on-premises and cloud-based storage.
- Nearly half will spend more than 50% of IT budgets on cloud data storage in the next two years.

**Where is the majority of your data stored today?**
Hybrid storage means that IT teams can manage data in the most cost-efficient, flexible manner possible according to user and business requirements; they also can gain access to hundreds of tools for AI and machine learning in the cloud to do things with data that weren’t possible before. Yet, having a hybrid strategy also adds new storage silos, each with their own architecture, security considerations and interoperability challenges. There are even more contingencies to manage when adopting multiple clouds—despite the many benefits of a multi-cloud strategy.

Beyond primary storage, 40% are investing in modernizing backup and disaster recovery systems in the next 12 months.

IT leaders face a number of priorities across storage and data protection, which means that a multi-vendor environment is here to stay.

**Key Insight:**

The rise in hybrid cloud storage and increased use of new technologies leads to less visibility across silos. When storage managers lack detailed insight into types of data and where it is stored, at what cost and its business value, it’s difficult to make the best decisions for efficiency and performance. Costs can escalate needlessly, from staff time spent manually collecting storage metrics across environments, from the common practice of defaulting to storing most or all data on expensive Flash appliances, and from resolving governance and security issues which arise from hidden or mismanaged data.
III. Lack of Visibility is Slowing Hybrid Cloud Adoption

*Enterprises need analytics on their data to plan cloud data migrations, cut costs and best serve the organization at large.*

In our data-centric world, IT leaders are seeing the need to be more strategic in how they manage data so they can reduce risk and costs and enable new ways to monetize it. Enterprises should aim to tackle both cost and complexity while ensuring that data is easily shared, discovered and analyzed for broader benefit.

**The Komprise Survey Asked IT Directors and Managers About Their Goals for Improving Unstructured Data Management**

Respondents prioritized visibility to plan better and the desire to avoid increases in storage and backup costs. IT directors also seek a more systematic approach to data management, the ability to switch cloud and storage providers without lock-in and the ability to deliver storage-as-a-service to business units.

**What are your main objectives for improving your unstructured data management strategy?**

<table>
<thead>
<tr>
<th>Objective</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>We need visibility to plan better</td>
<td>44.5%</td>
</tr>
<tr>
<td>We want to avoid rising storage and backup costs</td>
<td>44.9%</td>
</tr>
<tr>
<td>We want the flexibility to switch our cloud and storage options without lock-in</td>
<td>40.5%</td>
</tr>
<tr>
<td>We want to be agile and deliver Storage-as-a-Service to our business units</td>
<td>40.2%</td>
</tr>
<tr>
<td>We want a systematic data management approach that is not ad-hoc</td>
<td>35.9%</td>
</tr>
<tr>
<td>Other</td>
<td>1.3%</td>
</tr>
</tbody>
</table>
Top Challenges of Unstructured Data Management

The top pain cited is that unstructured data is growing too fast and getting too expensive to store and backup (35%) followed by the lack of an effective tiering strategy.

When asked about top priorities for data in the cloud, enterprises focused on tactical goals: moving to the cloud was cited as more important than cost-cutting:

- 56%—Migrating data to the cloud topped the list
- 46%—Followed by cutting costs
- 41%—Managing governance and security of data
- 38%—Gaining visibility into business unit usage of the cloud

What are your challenges with unstructured data management today?

- Growing too fast and getting too expensive to store/backup: 34%
- We lack an effective tiering strategy and tools to move it/manage it based on age and usage: 33%
- It’s not a big challenge for us today: 29.6%
- We don’t have a plan to monetize it: 19.3%
- I don’t know what we have; we lack visibility: 18.6%

What are your top priorities for data in the cloud?

- Migrating more data to the cloud: 56.1%
- Saving cloud costs: 46.5%
- Governing our data in the cloud (security, who is doing what, etc): 41.2%
- Getting visibility into how our business units are using the cloud: 37.9%
- Other: 1.3%

Key Insight:

Most (70%) of organizations face unstructured data management challenges with rising costs, impaired visibility and lack of sufficient tools, pointing to a need for systematic data management solutions. IT leaders will need to look for solutions that offer flexibility for hybrid cloud environments while avoiding vendor lock-in so that data can move freely between storage products and services without incurring extra cost and hassle. Automation will also be required to deliver a more systematic, data-centric approach versus ad hoc, storage-centric and manual approach for core data management activities.
IV. IT Leaders Want Analytics to Inform Storage Decisions

The future of data management lies in eliminating storage waste, extending the life of unstructured data, and leveraging cloud-based services like data lakes for monetization.

Better visibility is critical to cut costs, plan migrations, prevent security incidents and understand business usage, so it makes sense that 45% plan to invest in analytics tools in the next 12 months.

The demand for analytics outweighed buying more storage on premises or in the cloud.

A Measurable Benefit of Analytics is that it Provides the Foundation for Ongoing Data Lifecycle Management

This entails understanding data usage so that infrequently accessed (cold) data can be moved to lower-cost storage, and in some cases, eventually deleted. One-third of enterprises acknowledge that over 50% of their data is cold, while nearly 20% don’t know.

What is your next investment step for storage in the next 12 months?

<table>
<thead>
<tr>
<th>Investment</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investing in analytics tools</td>
<td>45.2%</td>
</tr>
<tr>
<td>Moving to or buying more cloud storage</td>
<td>43.5%</td>
</tr>
<tr>
<td>Modernizing backup and DR systems</td>
<td>39.5%</td>
</tr>
<tr>
<td>Buying more or upgrading on-premise storage</td>
<td>37.9%</td>
</tr>
</tbody>
</table>

How much of your data today is cold?

<table>
<thead>
<tr>
<th>Data Coldness</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 25%</td>
<td>19.3%</td>
</tr>
<tr>
<td>26 to 50%</td>
<td>30.9%</td>
</tr>
<tr>
<td>51 to 75%</td>
<td>24.9%</td>
</tr>
<tr>
<td>More than 75%</td>
<td>19.3%</td>
</tr>
<tr>
<td>Unsure</td>
<td>19.3%</td>
</tr>
</tbody>
</table>
Yet, visibility is just the first step. Being able to automatically move data as it ages ensures the continual optimization of storage costs. But, less than half (46.5%) are using tools to move data to secondary storage. The rest are doing it manually or not treating cold data differently than warm or hot data.

Once an organization achieves visibility across the entire storage environment, IT and business leaders can begin to leverage data in new ways and build sophisticated data management practices. When asked about new approaches to managing unstructured data, participants were most interested in creating data management policies that can be systematically applied.

Examples of such policies include:

- When to move data off hot storage to warm or cold storage
- What data requires active backups and DR and what data can be on resilient storage
- Data archiving and tiering policies by departments and data types
- Deletion policies for files belonging to former employees
- Alerts when cloud egress fees reach a particular threshold
- Copying certain data to data warehouses in the cloud

When policies are combined with automation—to execute the action without manual effort—IT can ensure compliance to policies and eliminate errors.

How are you managing your cold data today?

- We use tools to find and archive cold data to secondary storage: 46.5%
- We manually find cold data and move it to secondary storage: 31.2%
- We are not managing it differently than warm or hot data: 22.3%

What new approaches to managing unstructured data are you taking or plan to take?

- Creating data management policies that can be systematically applied: 55.8%
- Tagging data for future use: 42.2%
- Enabling data lakes and better user search of all our data: 42.2%
- Understanding anomalies and trends for better security and ransomware detection: 40.9%
Secondarily, IT leaders expressed equal interest in strategies involving search and insights:

- **Tagging data for future use**—Data tagging entails annotating or labeling data (like text or objects in videos and images) to make it detectable and recognizable in search and/or to train AI and ML models. Tagging can show which data requires active backups and DR and which can be on resilient storage.

- **Enabling data lakes and better user search of all our data**—Virtual data lakes provide central storage and features to foster and enable Big Data, AI, and ML projects. Data search and indexing technology in data management software can automate the process of finding unstructured data based on specific criteria to feed such projects. The major cloud providers all have invested in data lakes, and large enterprises are experimenting with solid use cases. **BMW stores petabytes of data** coming from BMW vehicles around the world in an AWS-hosted data lake, allowing the company to monitor vehicle errors to identify potential issues across vehicle lines. **Pfizer** is moving petabytes of file data transparently to AWS, enabling researchers to search on the cloud objects to build virtual data lakes for further analysis.

- **Understand anomalies and trends for better security and ransomware detection**—Storage and data management tools are another weapon in the fight against hackers, if they can show trends in data usage which might indicate nefarious actions. Such insights could include an unusual volume of file downloads from cold or archived storage, high activity during off hours such as weekends, or other unusual access patterns.

**Key Insight:**

*IT needs a more precise understanding about data:* where it lives, who uses it and its organizational value. With this knowledge, IT managers can use policies and tools to automate data movement to the optimal location for both cost and performance. This data-driven approach to data management will deliver optimal efficiencies from the entire storage ecosystem and generate new life from the petabytes of unstructured data sitting idle inside corporate walls by moving it to the cloud where it can be discovered and monetized.
Don’t Compromise on Visibility
Enterprises are storing petabytes of unstructured data, and there’s no end in sight to rising storage costs. Most organizations will spend more on storage in 2021, and the majority are managing a hybrid cloud storage environment. With data growth and storage environment complexity top of mind, IT leaders will need to install a surgically precise process to understand costs for departments and regional offices across storage silos, backups and DR, including staff time to manage and provision it all. This will require tools to enable visibility across the entire organization along with detailed reports at the site level.

Right Placing Data Requires Analytics + Action
The only way to effectively manage costs and the data management lifecycle is through analytics tools which can first show you what data (file types) you have, owners, and how much it costs to store, replicate and back it up, and next, model new plans for moving data to secondary and tertiary storage. Collect data on your data so that it always lives in the right place based on organizational demand and systematically move data based on your policies.

Stop Treating all Data the Same
Most enterprises are sitting on a healthy percentage of cold data—but not all organizations are treating data differently based on age and usage. First, know your data by determining how to segment cold data—this could be 30 days for some file types and 12 months for others. Next use tiering tools to safely and automatically move that data to colder storage as it ages out. Let usage patterns dictate policies rather than asking departments to make the call on what gets moved. Conduct user surveys periodically to assess the veracity of this data-driven strategy and make adjustments as needed.

Data Management is a Foundation for Data Monetization
Once you achieve visibility across the entire storage environment, you can build a storage-agnostic data management practice that liberates older, colder data from archives or hidden file systems into data lakes for broader research and collaboration. Ensure that your data management system allows users to flexibly and easily search for specific files and file types no matter where they reside. Data tagging is key. As well, enable employees to use any AI and ML tools they wish, including those in the public cloud, to manipulate data in novel ways. Data management of the future requires a strategy that goes beyond storage efficiency. It should transcend any specific technology so that organizations can be flexible, as needs arise, to move their data to new technologies that will best serve users and align with marketplace goals.
Komprise is the industry’s only multi-cloud data management-as-a-service that frees you to easily analyze, mobilize, and access the right file and object data across clouds without shackling your data to any vendor. With Komprise Intelligent Data Management, you are able to know first, move smart, and take control of massive unstructured data growth while cutting 70% of enterprise storage, backup, and cloud costs.

Learn more at www.komprise.com

1. https://www.lifewire.com/terabytes-gigabytes-amp-petabytes-how-big-are-they-4125169#how-big-is-a-petabyte
2. IDC FutureScape 2019, IDC Datasphere 2018