



State of Unstructured Data Management Report

The second annual survey shows IT priorities expanding from storage cost-cutting to delivering data services, with a focus on unstructured data analytics, end-user self-service, cloud NAS migration and emerging use cases for generating greater business value from data.

August 2022

Table of Contents

Executive Summary	3
Report Highlights	4
Data Continues to Grow, Accelerating Investments in Cloud Storage	5
Cloud File Storage Takes Off	
Benefits of Cloud Data Migrations	
Unstructured Data Management Challenges: User Disruptions	9
Unstructured Data Management Goals: Flexibility and No Lock-In	11
Future Strategies	12
New Approaches	
Unstructured Data Management Capabilities	
New Cases for Unstructured Data Management	
Conclusion: Top Takeaways	16
About Komprise	17



Executive Summary

The world of data storage and unstructured data management is in continual flux: the pandemic accelerated cloud and digital initiatives and global data volumes will reach 97 ZB in 2022. This is amplifying the pressures on enterprise IT—more than 50% are managing at least 5 PB of data today, a percentage that has grown significantly in just one year.

What to do with all this data moving forward is a pivotal question, as the wrong strategy—or no strategy at all—creates waste, introduces risks and minimizes the potential for leveraging data in new ways for organizational and marketplace gain. We attempted to uncover these challenges and opportunities in our latest industry survey: **The 2022 Komprise Unstructured Data Management Report**.

As unstructured data volumes continue to grow, IT leaders are realizing the need to manage data differently and are struggling to do so. In 2022, 87% of IT leaders rate managing unstructured data growth as a top priority, up from 70% in 2021. Leading challenges in 2022 include moving data to the cloud without disrupting users and applications, the high costs of data storage and backups, hindered visibility into data's characteristics and complying with laws and regulations.

The 2022 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 directors, VPs and C-level executives at companies with more than 1,000 employees in the United States and in the UK. The survey was conducted by a third party in June 2022.

The top priorities for unstructured data management in 2022 are moving data to big data analytics platforms, enabling self-service so business users can securely leverage data services such as cloud AI/ML and data lakes and optimizing data storage infrastructure by increasing investments in cloud storage.

Highlights of the Report

Unstructured Data Management Challenges

- **More than 50%** of enterprises are managing 5PB or more of data, compared with less than 40% in 2021.
- **Nearly 68%** are spending more than 30% of the IT budget on data storage, backups and disaster recovery.
- On-premises only environments **decreased from 20% to 12%**.
- **The largest obstacle (42%)** is moving data without disrupting users and applications.

Unstructured Data Management Priorities

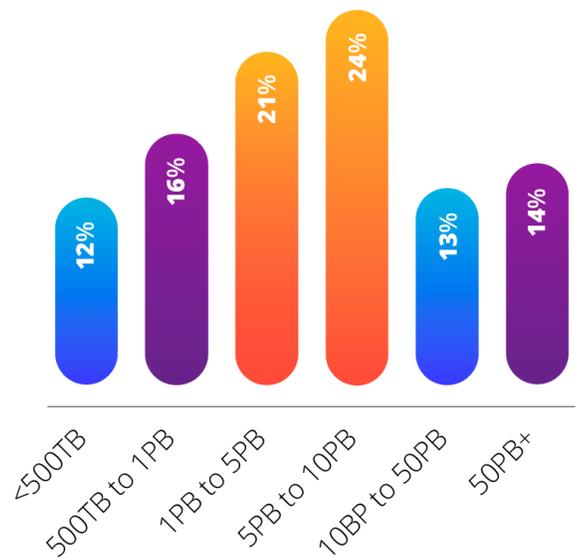
- **Nearly half (47%)** will invest in cloud NAS, followed by cloud object storage **(43%)**.
- After cutting costs, the second highest expected benefit of cloud migrations is to improve self-service for end users and departments **(43%)**.
- **A majority (65%)** of organizations plan to or are already delivering unstructured data to their big data analytics platforms.
- **The top goal (43%)** is to adopt new storage and cloud technologies without incurring extra licensing penalties and costs.
- The leading new approach for unstructured data management as identified in our survey is the ability to initiate and execute data workflows **(43%)**.
- The leading new use case for unstructured data management is protecting sensitive data **(63%)** followed by big data analytics **(41%)**.

Data Continues to Grow, Accelerating Investments in Cloud Storage

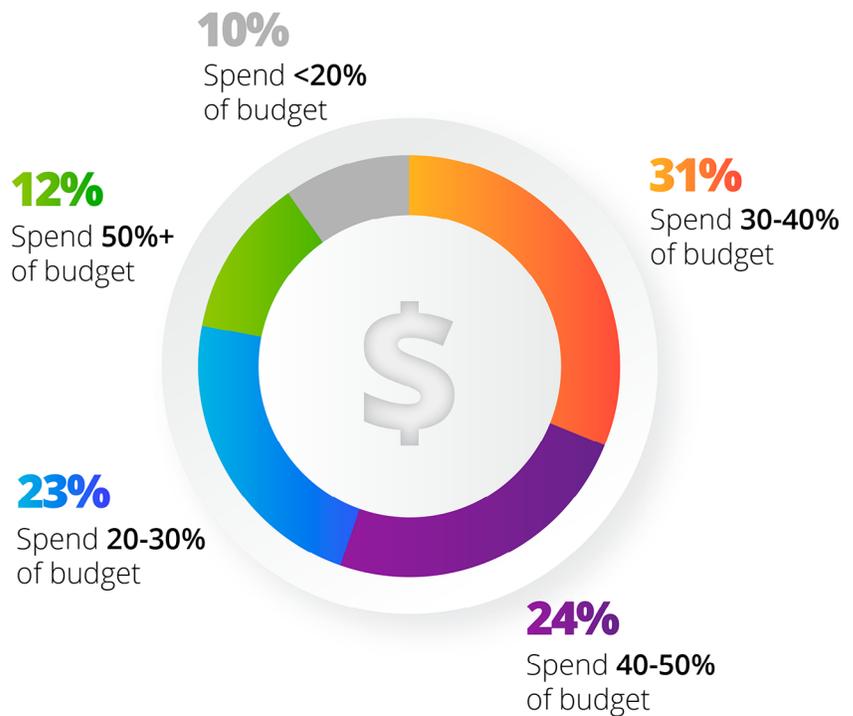
In 2021, we reported that unstructured data volumes were growing too quickly, straining IT budgets. That trend continues in 2022—with even higher volumes and spending plans.

- **More than 50%** of organizations are managing 5PB or more of data, compared with less than 40% in 2021.
- **At least 80%** of this data is unstructured, with the top file types reported as documents and user data, research data and video and audio files.
- **Nearly 68%** are spending more than 30% of their IT budget on data storage, backups and disaster recovery—similar to 2021.
- *However...* **Nearly 70%** said they would spend more on storage YoY, compared with 62% in 2021.

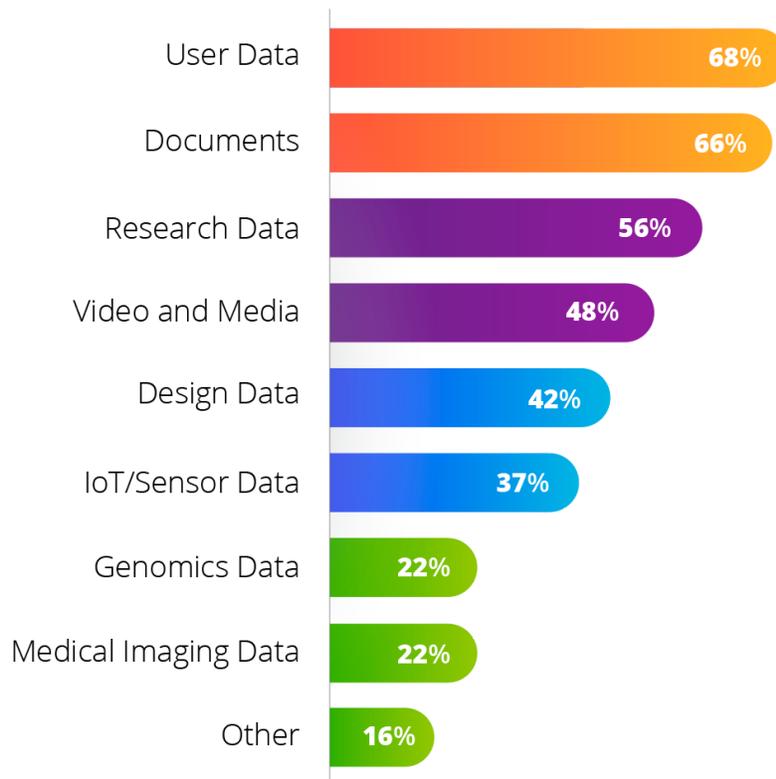
The current amount of data being managed:



Current percentage of IT budget that is being spent on data storage and data protection:



The types of unstructured data being stored:

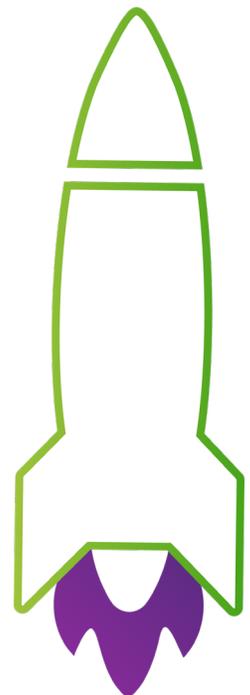


Cloud File Storage Takes Off

Hybrid cloud again persists as the primary storage architecture, **with 50% reporting a mixed storage environment across on-premises and cloud**—the same percentage as 2021. However, in just one year, the percentage of participants reporting on-premises only environments decreased from 20% to 11.9%. More than 20% of environments reported a multi-cloud environment and 15% reported a cloud-only environment.

As traditional storage vendors and cloud providers alike introduce new offerings for cloud file storage, enterprises are jumping at the opportunity to free capacity on their expensive on-premises NAS while also acquiring scalable capabilities for backup and disaster recovery.

Cloud NAS topped the list for storage investments in the next year. Cloud object storage came in second, followed by cloud archives such as AWS S3 Glacier. One-third reported they'd buy more on-premises storage.



Where most data is being stored today:



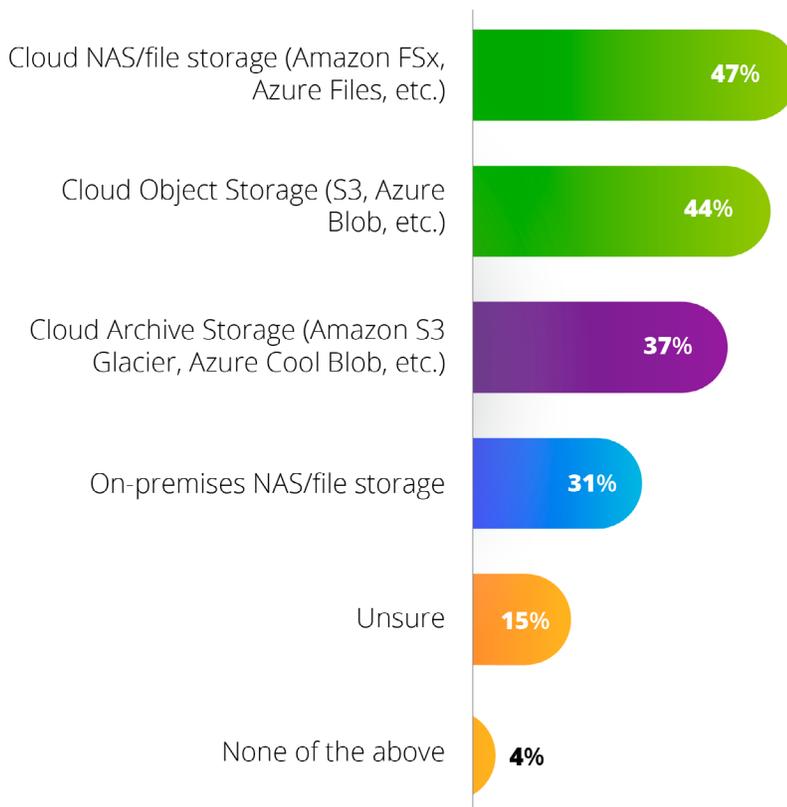
50% Hybrid cloud

23% Multi-cloud

15% Cloud only

11% On-premises

Expansion plans in the next 12 months:



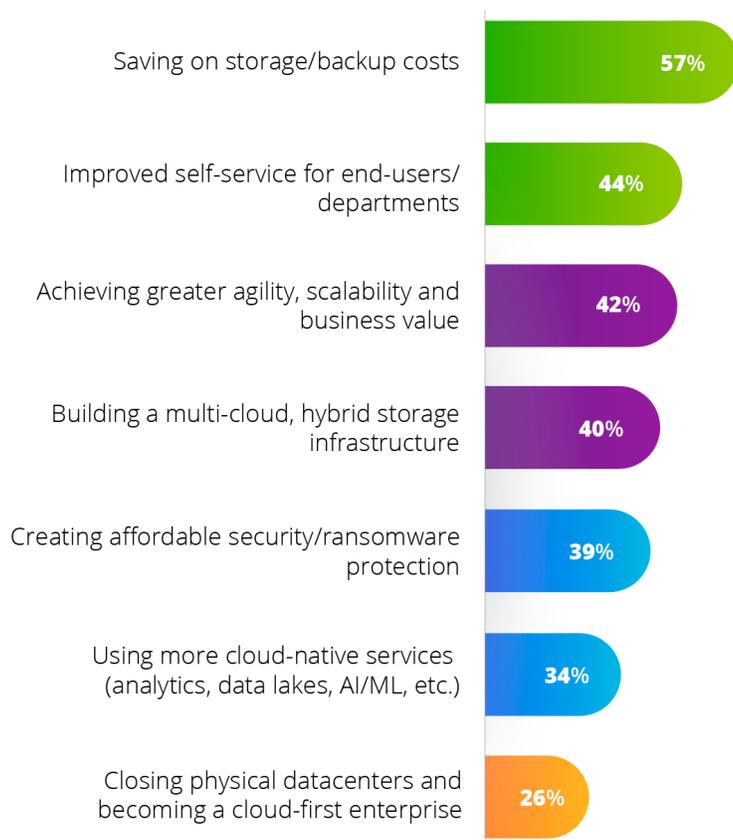
Benefits of Cloud Data Migrations

When asked what benefits they expected from moving unstructured data to the cloud, the majority (56%) of IT executives were most interested in cutting costs. **Yet the second-highest expected benefit (43%) was to improve self-service for end users and departments.**

As enterprises double down on digital initiatives to hasten a post-pandemic recovery, IT organizations are seeking ways to speed time to delivery for end-users and departments through better self-service. These strategies offer easier data access, simpler reporting, more accessible analytics via cloud data lakes and cloud ML/AI services and low-code tools for citizen development of apps and sites.

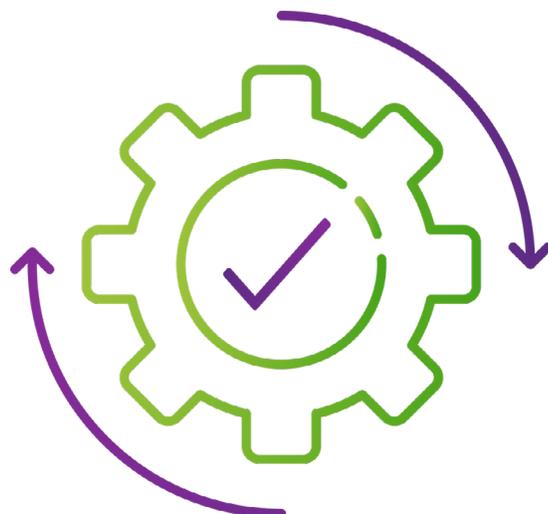
In data management, self-service typically refers to the ability for authorized users outside of storage disciplines to search, tag and enrich and act on data through automation. For instance, a research scientist could identify project files she wants to export to a cloud analytics service and then create a policy to automatically move those files as they are created.

Benefits expected from moving unstructured data to the cloud:



Data management self-service is viable with:

- Strong governance including role-based access;
- Simple search using a central console or index that looks across all files no matter where they're stored and;
- The ability to create automated workflows supporting all the steps from automated metadata tagging, search and movement of data to external platforms for analysis where it can be further enriched.



Unstructured Data Management Challenges: User Disruptions

Traditionally, controlling data storage costs and delivering reliable access to data have been the core responsibilities of storage teams. As unstructured data has grown in volume and variety, the challenges have amplified; in 2021, 30% of IT directors said that managing unstructured data wasn't a big challenge, compared with 2022 in which only 13% said the same. Progressive organizations are looking to move from a storage-centric approach focused on managing storage technologies to a data-centric approach where data drives all decisions independently of the storage platforms in use.

IT leaders need a nuanced approach to data which doesn't treat all data the same but considers its lifecycle for active access, longer-term storage, governance and future analytics potential.

Survey participants note that the largest obstacle relates to user experience: moving data without disrupting users and applications. It's critical to be able to move data to the optimal storage platforms for performance and cost yet in many cases people can't find files after they've moved—creating frustrations all around.

Unstructured data management solutions should move data transparently so that users can simply click on the link to the file exactly as before and applications work as before.

Other top unstructured data management challenges include:

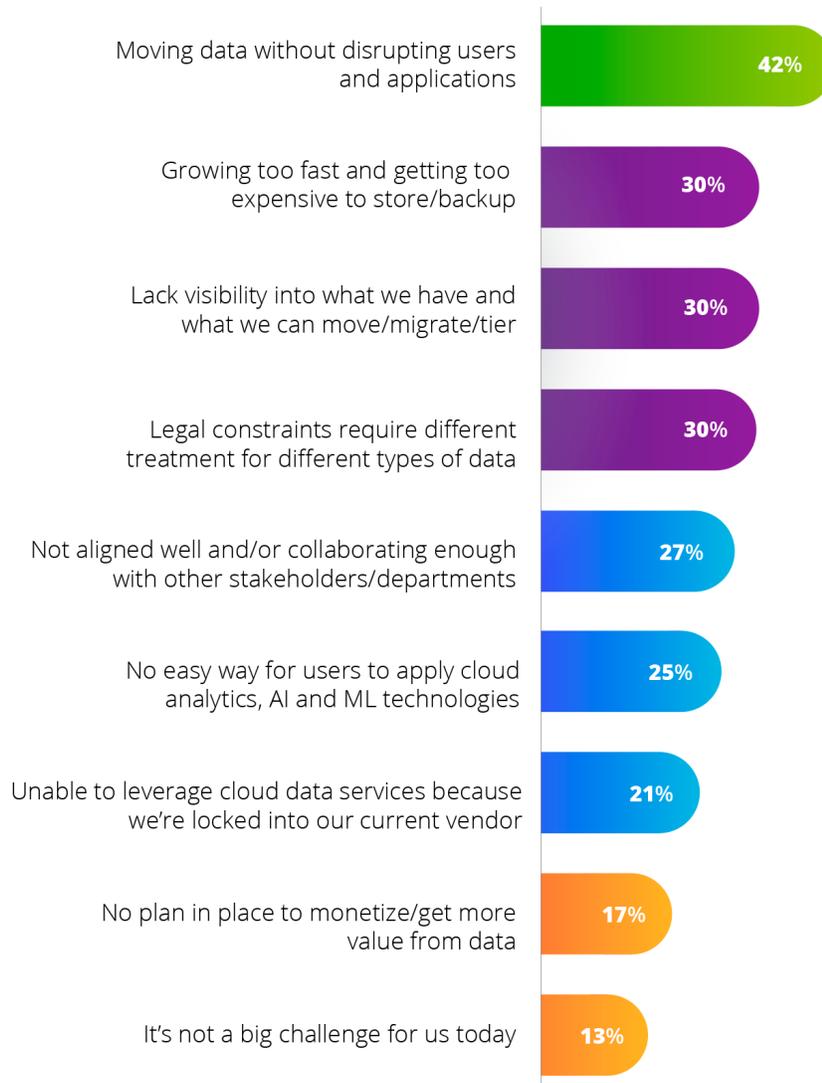
- Data is growing too fast and is too expensive to store and back up;
- A lack of visibility into data and its characteristics to inform mobility decisions and;
- The need to deploy different data management strategies to comply with legal/regulatory requirements.



Meanwhile, more than half of respondents stated that these challenges are impeding larger goals for the organization: leveraging data for new uses and supporting marketplace growth.

42% see moving data without disrupting users and applications as the top challenge.

Top challenges with unstructured data management today:



Unstructured Data Management Goals: Flexibility and No Lock-In

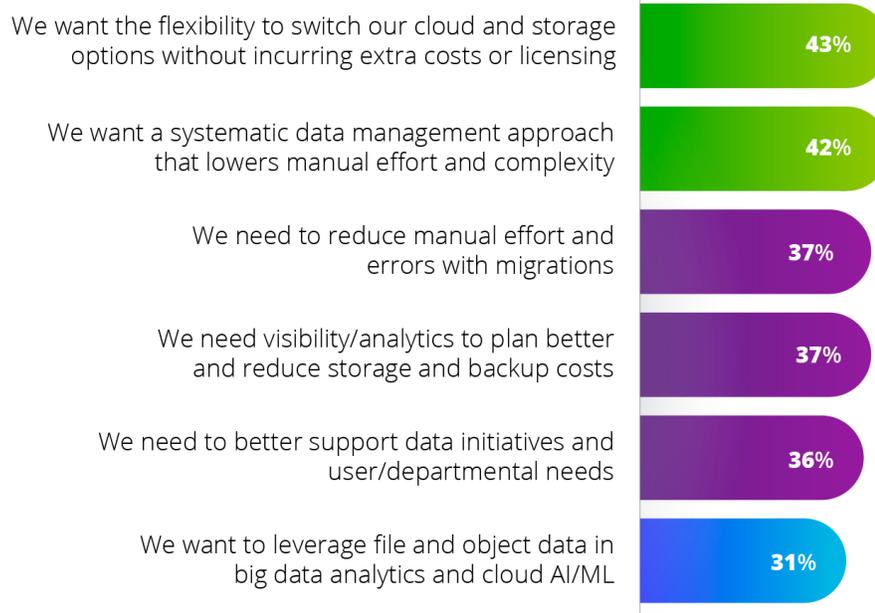
We asked IT leaders about their goals for improving their unstructured data management strategy – **and topping the list was flexibility**. Specifically, organizations wish to adopt new storage and cloud technologies without incurring extra licensing penalties and costs, such as cloud egress fees. This finding is enlightening, indicating that more than cutting costs, organizations wish to take advantage of new technologies and switch between technologies (such as cloud providers and cloud storage classes) as needed to meet shifting business goals and user requirements.

The second highest goal was attaining a systematic data management approach to reduce manual effort and complexity. This entails automated policies for data movement—such as tiering data to secondary storage after it reaches one year of age or moving research data with certain metadata tags to cloud archival storage.

IT leaders also prioritized better visibility and analytics for planning, reducing manual effort and errors for data migrations and the need to better support broader data initiatives and departmental needs.



Top objectives for improving unstructured data management strategy:

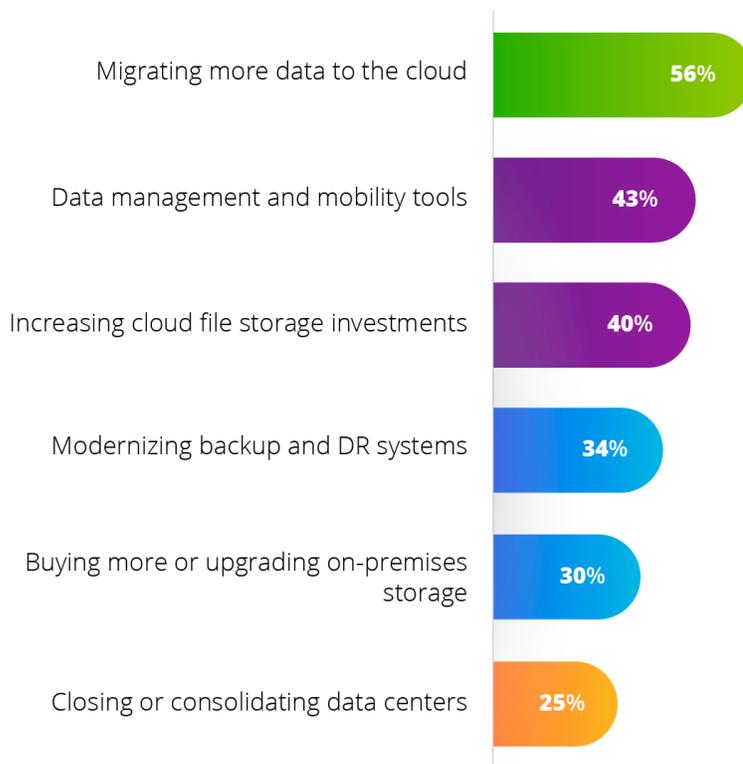


Future Strategies

We asked participants to share their plans for data storage investments, new unstructured data management approaches, emerging use cases and future needs for unstructured data management capabilities.

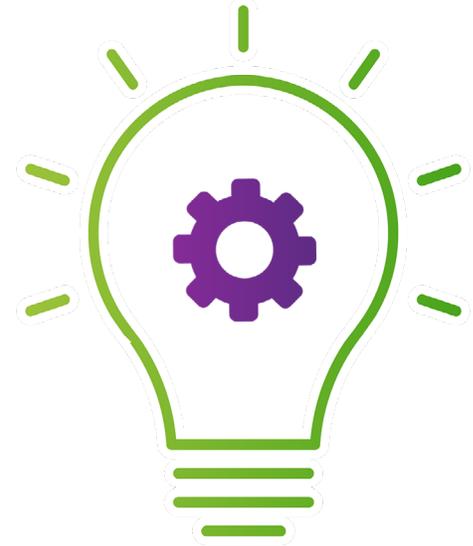
Investments: In the next 12 months, the top investment strategy for storage is migrating to the cloud (55%), followed by investing in data management and mobility tools and increasing cloud file storage investments. As IT infrastructure and business applications and processes generally move to the cloud and IT organizations require sophisticated tools and automation to control this complex environment, it is natural that storage and data management would follow suit. At the same time, nearly 30% of organizations will still invest in on-premises storage and 34% plan to modernize backup and disaster recovery systems, indicating the resilience of a multi-vendor, hybrid cloud environment.

Data storage investment priorities in the next 12 months:



New Approaches

Advances in cloud data management and analytics along with edge processing is offering more choices for IT organizations seeking to efficiently mobilize data for new uses. Earlier this year, a survey on 2022 IT spending plans conducted by Enterprise Strategy Group found that organizations are prioritizing the use of next-generation technology and the ability to handle data where it's generated. ESG also found that 62% of IT decision-makers will increase spending on AI & ML this year. Since ML models require large amounts of data to increase accuracy, incorporating unstructured data is table stakes going forward.



Our survey found that unstructured data and AI/ML are colliding, ushering in a new era for enterprise data analytics:

65% of organizations plan to or are already delivering unstructured data to their big data analytics platforms.

Another top new approach for unstructured data management is the ability to initiate and execute data workflows, such as searching and moving files with specific metadata (aka project name or customer group) to a cloud data lake (43%). Automating the process of moving specific data sets to platforms for processing and analysis is the future of unstructured data management and will speed time-to-value for big data analytics initiatives.

Enterprises are also looking to create automated data management policies—a top initiative in 2021—and understand anomalies to predict and prevent unexpected capacity needs or ransomware issues. Analytics on unstructured data is now becoming imperative to not only manage costs and performance but to protect organizations from malicious actors.

New approaches to unstructured data management being taken or planned:



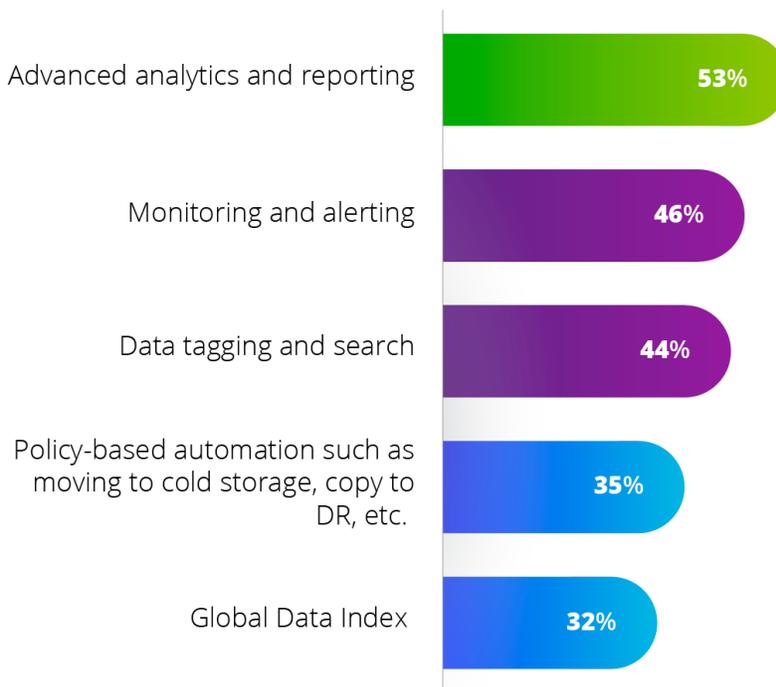
Unstructured Data Management Capabilities

Unstructured data management must be independent and agnostic from data storage, backup and cloud infrastructure technology platforms. The reason being? Most enterprises have several storage, backup and disaster recovery vendors in place. Rather than managing these technologies (and the data stored on them) in silos, an unstructured data management platform should work across all storage so that IT professionals can make data-centric decisions.

Advanced analytics and reporting will be the most important capability of these solutions (nearly 53%), according to the survey. Analytics on data is imperative for making the best decisions on where data should live and when it should move elsewhere. It can answer questions like: how much data do I have and where is it stored, how large are my files and of what type, how old is my data, what is the cost of storing it in different places, who last accessed it and which data is “active” versus which data is “cold.”

Other top requirements for unstructured data management software include monitoring and proactive alerting of key events such as running out of quota, a data service becoming unresponsive or anomalous activities, as well as data tagging and search.

The most important future unstructured data management capabilities:



New Use Cases for Unstructured Data Management

Organizations typically begin their journey with unstructured data management by cutting storage and backup costs through analytics and automation. Beyond cost savings—*primarily accomplished by moving older, less active data off of expensive on-premises storage*—enterprises are most interested in the following use cases:

63%

Protect Sensitive Data

Cybersecurity remains one of the leading priorities for IT organizations, particularly as ransomware, supply chain breaches and other sophisticated security attacks continue to increase. Individuals responsible for storage and data management are looking at how they can aid prevention and recovery through secure off-site backups including to immutable object lock storage in the cloud.

41%

Big Data Analytics

Researchers and users want to search and run analytics on unstructured data. Searching for files with specific characteristics or from specific projects across all data has historically involved Herculean manual effort. Modern data management platforms can index data, support metadata tagging for easier search and classification and automate workflows to move data to analytics platforms and further enrich the data with new tags as needed.

36%

Segment and Integrate Data for M&A

During mergers and acquisitions, data management can be an afterthought but without proper planning, data can get lost, damaged, or mismanaged during transfers between entities--creating legal and security issues. Unstructured data management solutions minimize these risks by identifying data appropriate for deletion, analyzing the optimal target storage for different data sets and ensuring migrations are error-free.

35%

Enable Deletion Policies

The elephant in the room is the fact that organizations don't like to delete data and therefore rarely do. But now, it's possible to safely delete old data that has outlived its purpose in a scheduled and monitored way, reducing management costs and eliminating the security risks of hidden and forgotten data. This also supports compliance when rules and regulations require the deletion of certain data—such as PII—after a period of time.

Conclusion

In the hybrid cloud era, enterprise IT organizations are pivoting from managing storage to managing data. With so much innovation in the cloud, IT leaders want to help users leverage cloud data services for faster collaboration, processing and AI/ML projects: *ensuring that the right data sets can always be leveraged there is a top initiative*. Key priorities include cloud file storage, self-service capabilities, non-disruptive user experience, flexibility and expanded use cases for data protection and big data analysis. With a data-centric versus storage-centric approach, IT leaders can stop the cycle of buying and managing storage technology and focus on delivering an efficient, value-driven unstructured data management practice benefiting users and the business at large.

Top 5 Takeaways:



Cloud File Storage Takes Off

As enterprises store more and more unstructured data at high costs, moving files to the cloud makes good economic sense. Cloud NAS is the preferred technology, but cloud object storage is a close follower as enterprises realize the need to tier rarely-accessed data to lower-cost storage as soon as possible.



Self-Service is a Top Goal for Cloud Data Migrations

In data management, self-service typically refers to the ability for authorized users outside of storage disciplines to easily search, tag, enrich and move data to new tools and services through automation.



Data Management Cannot Disrupt User Experience

Survey participants note that the largest obstacle to unstructured data management is moving data without disrupting users and applications, which occurs when organizations move data and files to new storage locations and users later can't find their files. Solutions that enable transparent access to files once moved are critical.



Flexibility and No Lock-In are Strategic Priorities

IT organizations wish to adopt new storage and cloud technologies and switch between technologies (*such as cloud providers and cloud storage classes*) without incurring extra licensing penalties and costs, such as cloud egress fees.



Unstructured Data Management Use Cases Expand

Enterprises will remain cost-conscious above all: saving money on storage with smart data lifecycle management practices. Yet as unstructured data management matures as a practice, new priorities are emerging—including *data protection, big data analytics, data deletion and M&A requirements for data transfers and segmentation*.



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.

Learn more at www.komprise.com