komprise

The 2023 Komprise **Unstructured Data Management Report**

Unstructured Data Management in the AI Age

September 2023

Table of Contents

Executive Summary 3

Report Highlights 4

Data Storage Priorities: Preparing for AI & Cloud Cost Optimization 5

Data Management Implications for AI 6

Data Management Self Service Gains Popularity 9

Unstructured Data Management Costs, Challenges, Needs: The New Basics 11

Top 5 Takeaways 17

About Komprise 18

Top of Mind in 2023:

Preparing for AI and cloud cost optimization

©2023 Komprise, Inc. All rights reserved.

Executive Summary

In a matter of months, the technology world has drastically changed. AI business models and new products related to generative AI are proliferating everywhere. The current wave of AI is quickly creating new ways of working, significant productivity advancements and transforming the way products and services are created and distributed. The generative AI market is poised to grow to $1.3 trillion over the next 10 years from a market size of just $40 billion in 2022, according to Bloomberg Intelligence.[1]

Of course, there is a dark side and the potential dangers from this new form of AI are vast. From privacy and security risks to ethics concerns and existential threats, government and business leaders are analyzing the issues and weighing solutions for safely and successfully adopting AI in our society.

Amid these shifts, our 2023 research found that IT leaders are focusing on three core areas in the coming year:

- Preparing for AI;
- Delivering data services such as file search and tagging for departmental users and;
- Adopting cloud cost optimization strategies.

The amalgamation of these trends reflects the maturation of unstructured data management. IT and storage directors are now expected to align closely with high-priority enterprise goals for AI and data-driven initiatives.

How IT and business leaders manage data will **make or break** outcomes and ROI from AI innovations.

The 2023 Komprise Unstructured Data Management Report examines the challenges and opportunities with unstructured data in the enterprise. This report summarizes responses of 300 global enterprise storage IT and business decision makers at companies with more than 1,000 employees in the United States and in the UK. All respondents work at the director level or above, across IT/technology operations teams and C-suite executives. The survey was conducted by a third party in June 2023.

[1] Generative AI to Become a $1.3 Trillion Market by 2032, Bloomberg Intelligence, https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-mrket-by-2032-research-finds/

©2023 Komprise, Inc. All rights reserved.

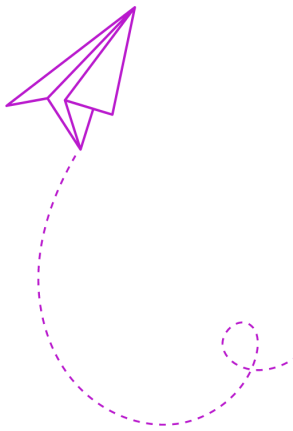
Highlights of the Report

Preparing for AI

- **Preparing for AI is the leading data storage priority in 2023 (31%)**, followed by cloud cost optimization;
- **Most organizations (90%) allow employee use of AI and the majority (65%)** have a policy in place to govern it;
- **Violation of privacy and security is the top concern for corporate AI use**, followed by data provenance and risks from inaccurate or biased data.

From Storage Management to Data Services

- **85%** say that non-IT users should have a role in managing their own data and **62%** already have attained some level of user self-service for unstructured data management;
- Monitoring and alerting for capacity issues and anomalies led the pack for important **unstructured data management capabilities (44%)**;
- The top unstructured data management challenge is moving data without **disrupting users and applications (47%)**, followed closely by **preparing for AI and cloud services (46%).**



More Data, More Spending, New Goals

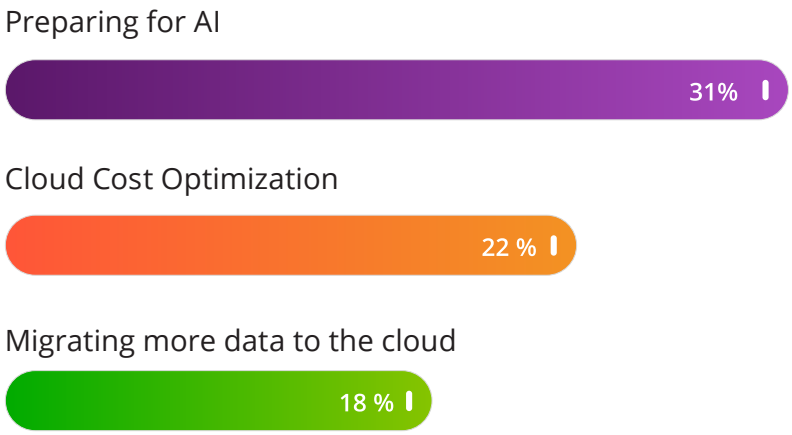
- **50% of organizations** are managing 5PB or more of data, **similar to 2022**;
- Organizations managing more than **10PB of data grew from 27% to 32% this year**, a **19% increase**.
- Nearly three-quarters **(73%) are spending 30% or more of IT budget on data storage and protection**, measurably higher than 67% in 2022;

©2023 Komprise, Inc. All rights reserved.

I. Data Storage Priorities: Preparing for AI & Cloud Cost Optimization

In 2022, we reported that a majority (65%) of organizations plan to or are already delivering unstructured data to their big data analytics platforms. While those analytics platforms likely included AI, today the landscape has shifted decidedly in favor of AI. In the blink of an eye, after ChatGTP took the world by storm, generative AI became the technology discussed in every boardroom, classroom and government office. In 2023, when asked to name their top data storage priority in the next 12 months, 31% said preparing for AI, with the next highest priority being cloud cost optimization (22%). Notably, in 2022, participants were largely focused on cloud migrations, with 56% saying it was their top initiative. Other priorities include investing in data management and mobility, buying more on-premises storage and modernizing backup and disaster recovery.

Top data storage priority in the next 12 months



Insight: The shift to AI requires cost optimization strategies

IT teams today are under pressure to reduce cloud and storage spend in an uncertain economy. Talks of cloud repatriation emerged at the beginning of 2023, after heavy spending on the cloud during the pandemic didn't always return the expected ROI. Optimizing for cloud entails tactics like monitoring spend with third-party tools and managing cloud sprawl and shadow IT through automated discovery and corporate policies. Data management is also instrumental as it gives storage and IT managers a means to view data assets across all storage and right-place data in the optimal storage solution for current needs.

Eliminating storage waste is critical to free up investment and resources for burgeoning AI initiatives. Preparing for AI from a data storage perspective means acquiring the right mix of high performance, scalable storage technology on-premises and in the cloud, which is also as cost-efficient as possible.

©2023 Komprise, Inc. All rights reserved.

II. Data Management Implications for AI

Preparing for AI, the leading data storage priority, doesn't just entail building an AI-ready storage infrastructure. Selecting the right tools is of course critical. There's no lack of affordable AI tools and services to meet nearly every industry need. Major cloud providers along with prominent enterprise software vendors are all unleashing their own flavor of generative AI-related solutions to the masses. Before deploying any tools or service, IT must consider the data management implications.

While choice, ease of use and compelling new use cases for generative AI are exciting, the other side of the equation is troubling, if not altogether frightening. From sensitive data leakage to general language learning models (LLMs)--threatening corporate IP and PII protection--to concerns with ethics, accuracy, data provenance, copyright issues with derivative works and even existential threats from malicious actors, AI has a host of valid concerns for businesses and society at large.

Insight: Adopt generative AI but with guardrails for tools and data

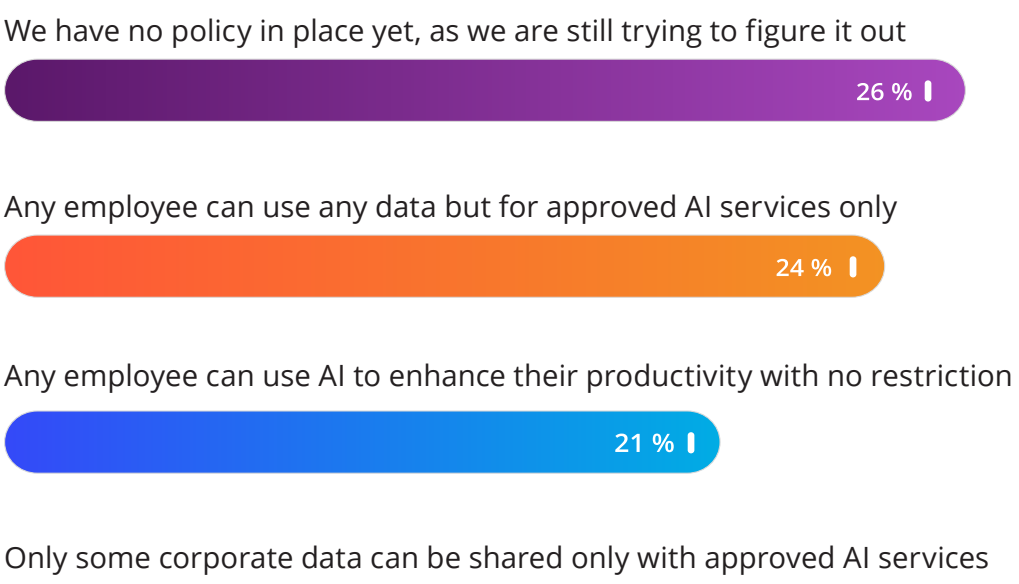
These concerns are influencing enterprise decisions for AI use but according to our survey, not dissuading business and IT leaders from pursuing the technology.

Survey participants gave mixed responses concerning generative AI plans--unsurprising given its early stage--yet the majority (44%) restrict tools and/or data that employees are allowed to use.

The Komprise 2023 survey found that only **10%** of organizations are outlawing internal use of AI tools.

©2023 Komprise, Inc. All rights reserved.

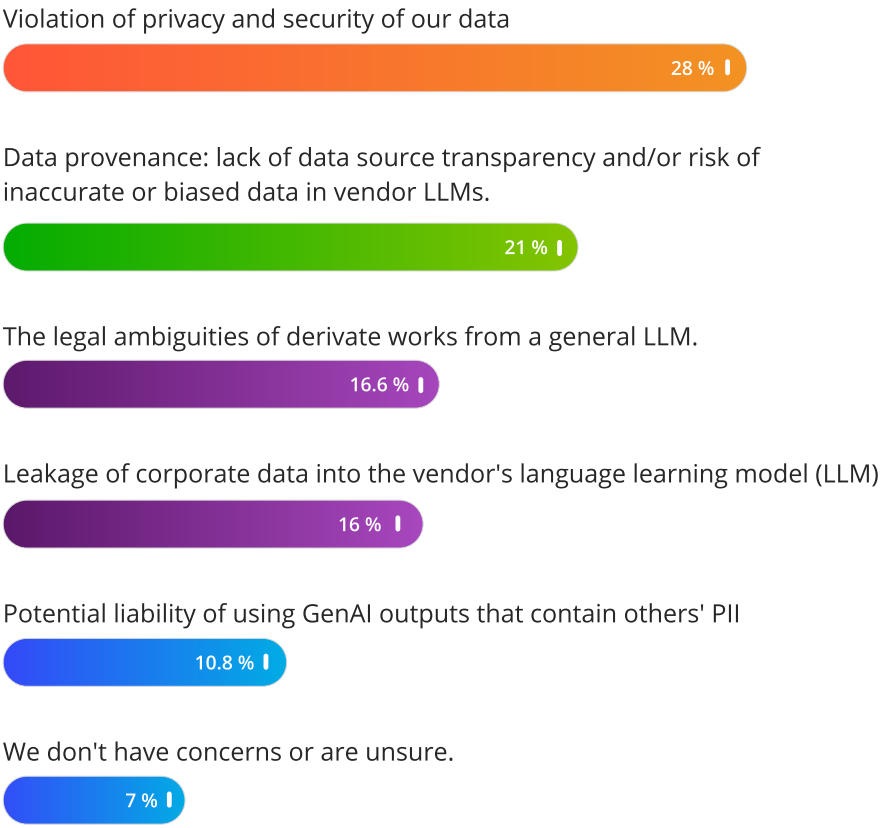
Enterprise AI plans



Top data management concerns of generative AI

IT leaders need to consider an array of privacy, security, legal and ethical issues from generative AI. They are most concerned about preventing security and privacy violations (28%) and improving data source transparency to avoid unethical, biased or inaccurate outputs from generative AI tools (22%).

©2023 Komprise, Inc. All rights reserved.



Mitigating the risks of unstructured data in AI requires several strategies—but doing nothing is not an option.

- **40%** say they will pursue a multi-pronged approach encompassing storage, data management and security tools;
- **35%** will work with their existing security/governance vendors;
- **32%** say they have capabilities in their data storage and/or unstructured data management solutions;
- **31%** have created an internal task force to develop and execute a strategy, and;
- **26%** will only work with an AI vendor that has adequate protections and controls.

Insight: Focus on governance for AI success

Clearly, the need for an unstructured data governance agenda is strong, as IT leaders cannot forsake data integrity, data protection and risk faulty or dangerous outcomes from generative AI projects. A 2022 survey from Wakefield Research and Informatica confirms this, citing data governance as the number one priority among chief data officers (CDOs)[3].

[3] CDO Insights 2023, Informatica: https://www.informatica.com/resources.

©2023 Komprise, Inc. All rights reserved.

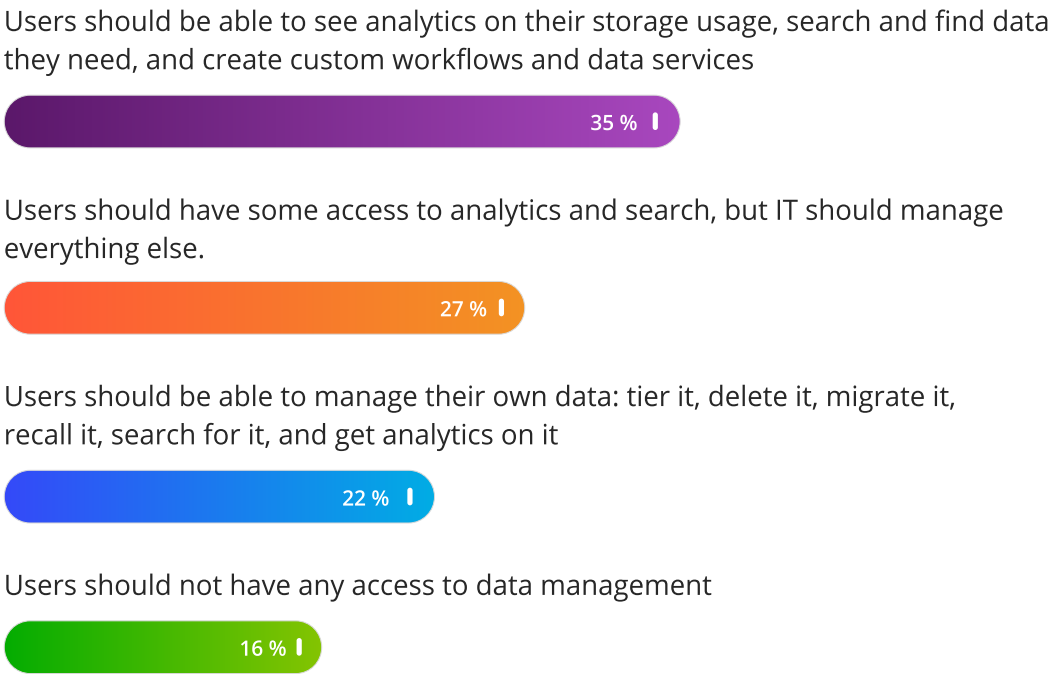
III. Data Management Self Service Gains Popularity

The general trend of IT self-service is now reaching unstructured data management. Departmental IT managers, data analysts, researchers and other data stakeholders across the business should be involved in the management of their own data. Doing so can enable improved collaboration on decisions about which data should be on high-performant storage, which data can be archived or deleted, and allow actions such as tagging files with metadata and searching across data silos.

The majority **84%** say they want users to have a role in managing their own data.

The prevailing trend (**36%**) is to allow employees to see analytics (such as data growth, file types and usage trends in their department), search for data across silos, and create custom workflows. **Yet IT wants to stay in control:** only 22% say that users should be able to fully manage their data by tiering, deleting, migrating, recalling it and more.

How do you want departmental users to manage their data?



©2023 Komprise, Inc. All rights reserved.

Insight: Data services 101: users want more control of their data

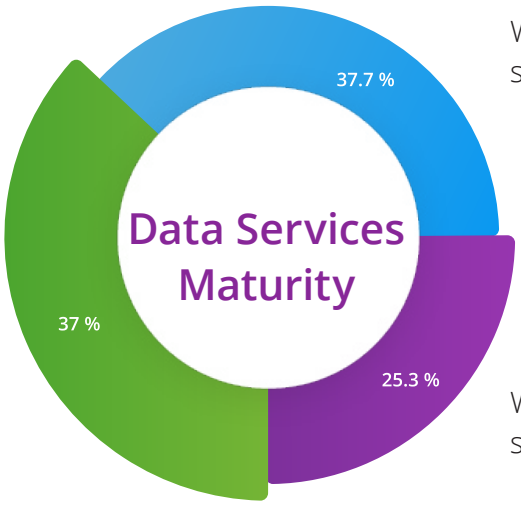
Data services involve:

- the management of data throughout its lifecycle;
- analysis and reporting on data storage growth and costs including departmental showback and data usage;
- file search and tagging and;
- data mobility use cases such as data migration, data tiering, replication and deletion.

While more than a third (38%) are still building out a data services strategy, the rest are making solid progress on self-service data management with 37% sharing departmental showback reports on data usage and spend. Self-service data management benefits IT and departmental users alike: the former can more readily meet goals for cost savings and compliance without conflicts and the latter group achieves more say in how their data is managed to meet business objectives.

A surprising **25%** have realized full user self-service for data management.

How far along are you in realizing your goals?



We are still building our data services strategy

We share showback reports on data usage and spend by department, but IT does the data management

We have achieved full user self-service for data management

©2023 Komprise, Inc. All rights reserved.

IV. Unstructured Data Management Costs, Challenges, Needs: The New Basics

Data Volumes and Costs

In 2022, **50%** of organizations were managing 5PB or more of data; that trend continues in the 2023 survey. Yet while **27%** were managing 10PB or more in 2022, this year that segment of heavy data owners has jumped to a remarkable **32%**.

What exactly is 10PB of data?

It's hard to envision but it equates to 110,000 ultra high-definition (UHD) movies, or half of the data stored by the U.S. Library of Congress[4]. The fact that one-third of respondents is storing this much data should give pause to how large unstructured data has become in the enterprise and its correlating burden on IT to manage, protect and store.

As in 2022, nearly 70% of enterprises said they will spend more this year than the year before.

In 2023, **73%** of organizations are spending more than **30%** of their IT budget on data storage, a marked increase from 67% in 2022.

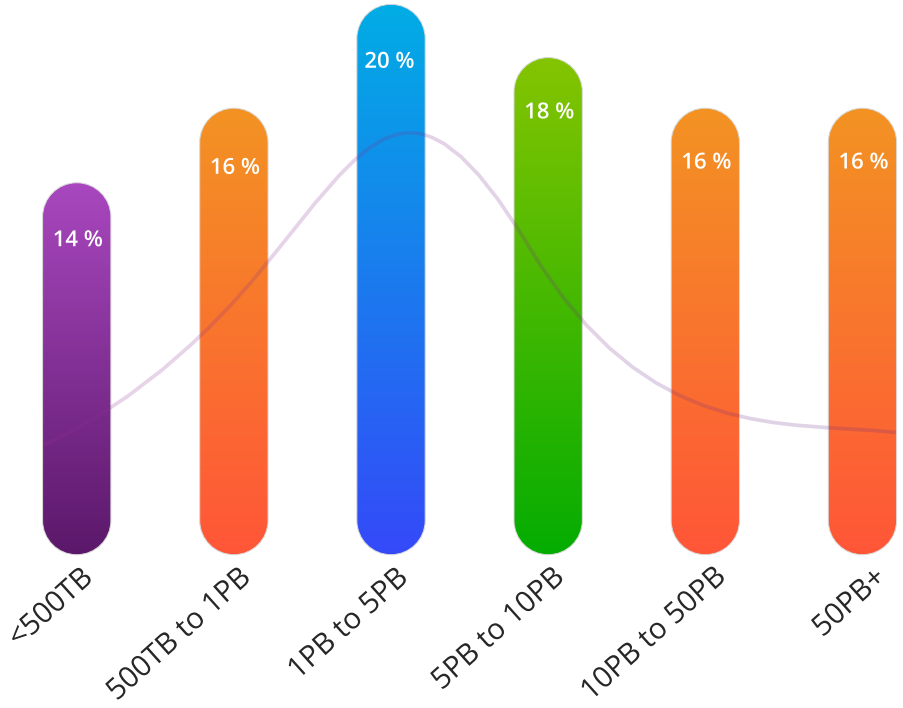
Insight: All Eyes on Cost Optimization

Despite more choice than ever in data storage and backup solutions, IT organizations are still spending a large portion of the IT budget on storage. In some respects, these costs are justified: protecting data is critical to business operations, customer success and overall growth. Yet there is also ample waste from over-procurement of storage capacity to avoid any business disruption, underutilization of cloud resources and one-size-fits-all storage strategies. Understanding data value, usage trends, data priorities and storage/cloud economics can help IT make better decisions that balance performance with savings and sustainability goals.

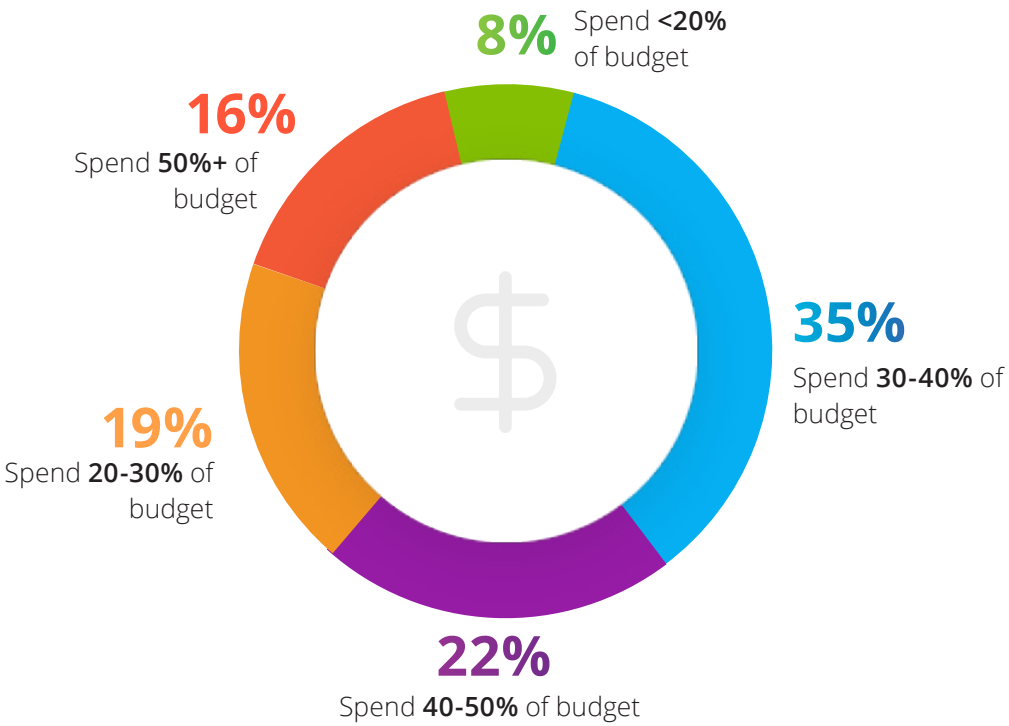
[4]Cobalt Iron, 2019: https://info.cobaltiron.com/blog/petabyte-how-much-information-could-it-actually-hold

©2023 Komprise, Inc. All rights reserved.

Current amount of data under management



Percentage of IT budget spent on data storage & protection



©2023 Komprise, Inc. All rights reserved.

Unstructured Data Management Challenges and Needs

Uncontrolled growth of unstructured data results in a lack of visibility which makes it hard to search and make decisions on how and where to store it. It can also incur legal and compliance risk from shadow data, escalating costs of storage and backups and departmental conflicts. Most organizations experience several challenges: resolving them or at least minimizing their impact is essential to meet users' needs and help generate new value from data.

In 2023, the leading struggles are moving data without disruption (also a top challenge in 2022) and preparing for AI. The former relates to a common issue where users can't find their data after it has been migrated to a new location or archived to save money. Unstructured data management solutions can help by moving data transparently so that users can simply click on the link to the file exactly as before and applications work the same way.

Preparing for AI, also a top data storage priority, is problematic because of the unknowns it brings to the table. AI also requires investment in the right high-performance, scalable and efficient storage (and compute) technology. Tirias Research forecasts that generative AI data center server infrastructure plus operating costs will exceed $76 billion by 2028[5]. In the cloud, providers like Amazon and AWS have developed services that bundle AI platforms with compute and storage capabilities to make this easier for companies--but these services may not be any cheaper over time.

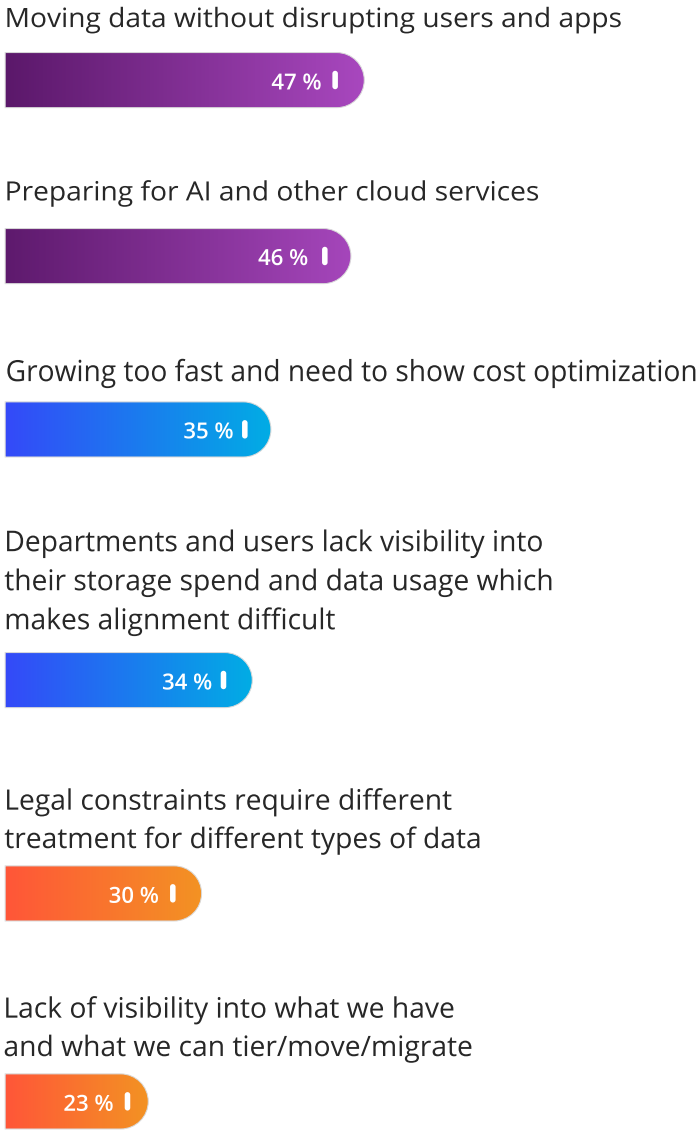
Insight: Preparing for AI requires an airtight understanding of data assets

Controlling data storage costs is tough no matter which storage solution you choose:: on premises, on the edge or in the cloud. The reason being is that data growth is not slowing down. Adopting tools and practices to analyze, classify and segment data can result in a nuanced data management strategy: lower-priority data lives on archival storage until it is needed for active use or until it can be deleted while high-priority active data is reserved for the most expensive top-tier storage. An independent unstructured data management approach can reduce annual storage, backup and disaster recovery costs by 70% or more in many organizations.

[5] Forbes, "Generative AI Breaks the Data Center", 2023: https://www.forbes.com/sites/tiriasresearch/2023/05/12/generative-ai-breaks-the-data-center-data-center-infrastructure-and-operating-costs-projected-to-increase-to-over-76-billion-by-2028/?sh=3a168f6c7c15

©2023 Komprise, Inc. All rights reserved.

Top challenges with unstructured data management



©2023 Komprise, Inc. All rights reserved.

Top Future Capabilities for Unstructured Data Management Software

The enterprise software sector of unstructured data management is morphing rapidly. It was born to address the explosive growth of unstructured data in the enterprise and the complexities of managing data across hybrid cloud storage. An unstructured data management platform should work across all storage so that IT professionals can get fast insight to make data-centric decisions.

IT organizations that are mature in unstructured data management can see annual savings on storage and backups of **70%** or more from right-placing data. They also gain the ability to deliver viable data services to users, such as easy searching and automated metadata tagging and movement of data to support a variety of use cases.

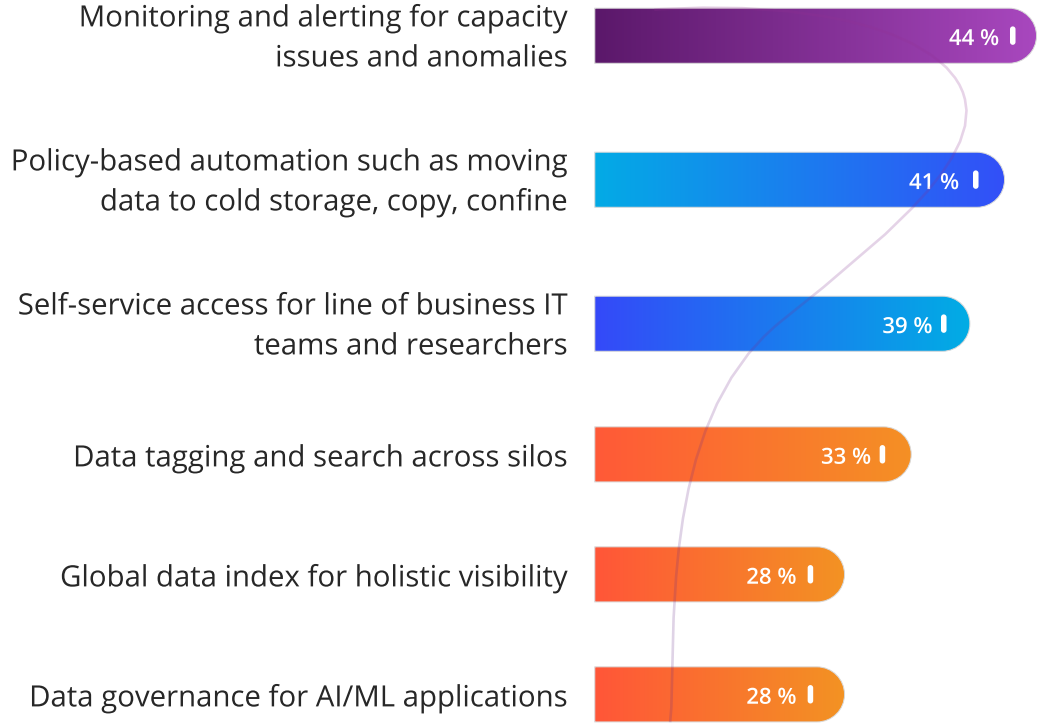
> The goal of unstructured data management is to ensure that the right data lives in the right place at the right time, with no user disruption and no vendor lock-in.

- In 2023, monitoring and alerting for capacity issues and anomalies was identified as the most important future capability of the software solutions, according to **44%** of the respondents.
- Policy-based automation such as moving data to cold storage or confinement for deletion was the second highest future software need **(41%)** followed by the familiar theme of self-service access for line of business IT teams and researchers.
- Data protection was the top new use case identified for unstructured data management in 2023—similar to the 2022 survey.

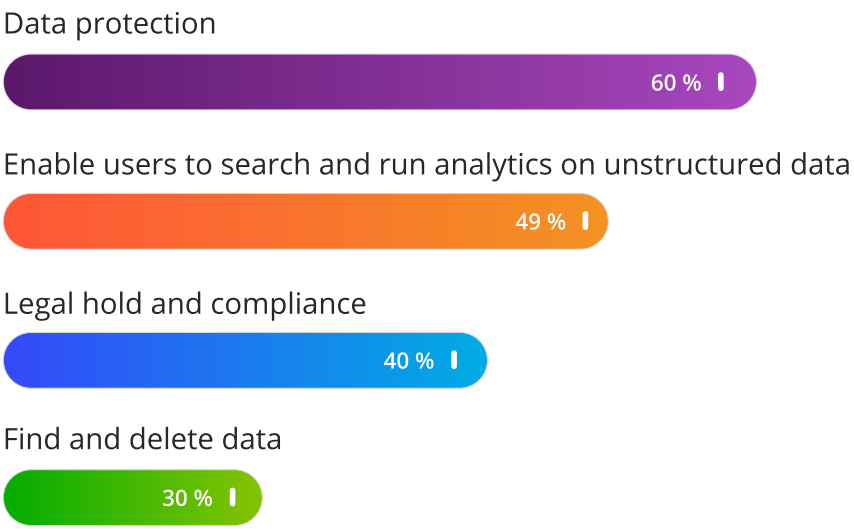
Insight: Unstructured data management solutions must meet many critical objectives from one platform: from holistic visibility across silos, to monitoring, search and tagging of files, automated policies, transparent access to data wherever it lives and data governance.

©2023 Komprise, Inc. All rights reserved.

Top future capabilities for unstructured data management

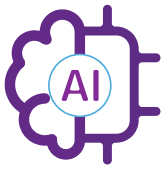


Top new use cases for unstructured data management solutions



©2023 Komprise, Inc. All rights reserved.

Top 5 Takeaways



Preparing for AI is A Top Priority and Top Challenge.

IT leaders agreed that preparing enterprise data and storage environments for AI implementations is of utmost importance but is a tough proposition given the ongoing flux in the generative AI sector. Organizations should consider how best to manage and understand the unstructured data that fuels these applications. Deep knowledge on unstructured data will be valuable regardless of how AI solutions evolve along with regulations and standards to govern it.

Leaders Want AI With Guardrails.

AI concerns range from corporate data leakage to ethics and accuracy, data transparency, copyright issues with derivative works and manipulation by malicious actors to cause harm. Despite these risks, most leaders support corporate use of AI—but want to restrict tools and/or data that employees are allowed to use. Pursuing a multi-pronged risk management approach encompassing storage, data management and security tools is popular.

A Data Services Approach Takes Hold.

The evolution from storage management to data services has been brewing since the pandemic. Digital business goals and processes are predominant. IT needs to not only protect data and make it readily available but ensure that users and department heads can understand their own data usage and participate in decision-making for its movement and management. Our 2023 research indicates that most organizations are well on their way to a mature self-service data management architecture.

Cost Optimization Tools and Tactics Grow Amid Cloud Caution.

As the cloud has proven less cost-effective than desired for some organizations, IT leaders are taking a hard look at their storage and cloud investments. This year, organizations are less focused on aggressive cloud migrations as cloud cost optimization has become a larger priority. Tools that analyze data assets and spending continuously will deliver a better understanding of data assets, and then ideally, take automated actions by policy to move data continuously to the most cost-effective storage for its use case.

Visibility, Insight Essential to Manage Data Growth.

In a competitive data storage market, cloud providers and traditional storage companies are constantly offering new cost-efficient solutions with different use cases. This reality makes it eye-opening that organizations are spending ever higher percentages of the IT budget on data storage and backups. IT managers need more insights and predictive analysis on costs, data usage and characteristics and automated methods for right-placing data. This will help free up dollars for AI, big data and other data initiatives.

©2023 Komprise, Inc. All rights reserved.

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