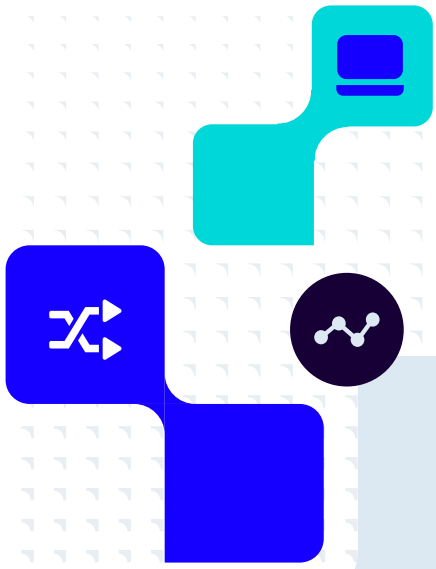
WHITEPAPER

Data Lifecycle Management

Control costs without sacrificing the speed and scale you need

smarsh

Table of contents



Introduction

The information explosion continues unabated. Despite dramatic improvements in data storage, businesses of all types and sizes struggle to keep up with exponential data growth. Data is growing faster than ever — but the real risk lies in how much of it is being stored without purpose, structure, or strategy.

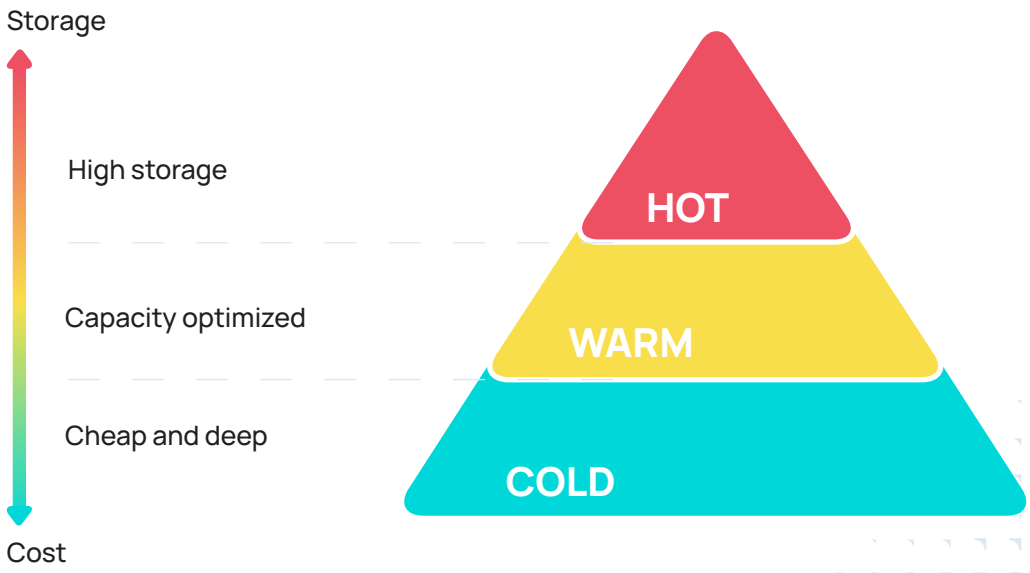
The primary culprit, however, isn't just the amount of new data. Nor is it the increasing file sizes that result from richer content. The most significant driver is an accumulation of data due to the:

- Increasing number of communication channels
- Evolving regulatory recordkeeping requirements
- Inability to determine when and how to perform the disposition of older data

As data piles up without structure or strategy, it becomes dark data — information that's stored but never used. This creates more than just inefficiency; it creates real business risk.

Enterprises need new approaches to data management to cope effectively with a rising tide of data and to avoid the risks that result from the uncontrolled accumulation of dark data.

Why existing approaches to data management fall short



In the past 20 years, organizations have adopted a range of solutions to manage exploding data volumes — often starting with tiered storage and basic archiving tools. Early systems focused on cost savings, replacing original files or emails with 'stubs,' or shortcuts to content stored on cheaper, slower storage. But as data has evolved, these approaches have fallen behind.

Modern storage platforms have introduced automated tiering to shift frequently accessed data between fast and slow storage. While useful for structured data, such as databases and indexes, these systems are typically optimized for performance rather than long-term retention or governance.

More modern data storage systems have incorporated data tiering technologies to move data among different storage devices within a pool of networked storage. However, these storage systems are primarily designed to move dynamic, frequently accessed data blocks rather than files. Self-tiering storage platforms typically focus on caching operations to improve the performance of frequently updated or accessed tables and indexes within structured data applications, rather than providing low-cost, long-term storage solutions.

Here's where purpose-built archiving comes in.

In contrast, purpose-built archiving systems incorporate technology designed to manage individual, non-changing files and to support large amounts of infrequently accessed data. These attributes are well-suited to information governance strategies. However, traditional archive platforms only provide a single class of service, with the assumption that all data within an archive is of equal value and at the final stage of its lifecycle.

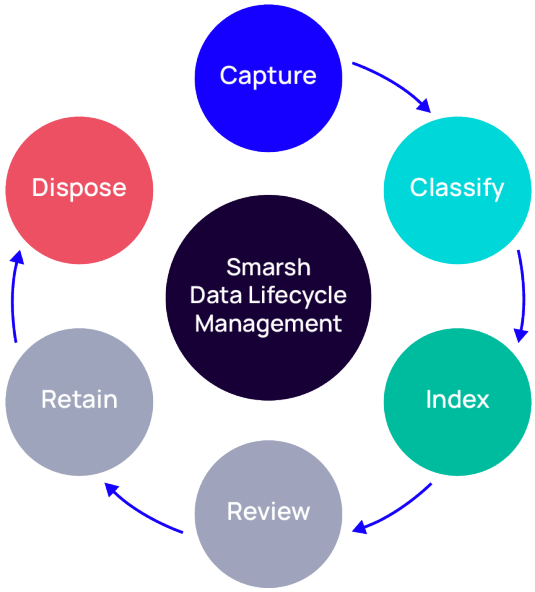
To locate specific files or messages, traditional archiving platforms utilize full-text index. But because these applications are primarily designed to reduce costs, most vendors use inexpensive indexing technology. Such index technology has proven inadequate at enterprise scale, and large organizations often suffer with:

- Poor search performance
- Failed and/or inconsistent search results
- Brittle indexes requiring constant maintenance

Smarsh Enterprise Archive

The Smarsh Enterprise Archive addresses the data lifecycle, scalability and performance challenges associated with traditional archive solutions. The platform intelligently captures, preserves, and enriches data, integrates insights, and simplifies the discovery process.

The platform's search capabilities can return millions of results in seconds. This speed and sophistication enable discovery and compliance teams to work far more efficiently, empowering them to iteratively search, filter, and refine large volumes of data in real time.



Smarsh Archive uses an extensible (XML-based) object model, which maintains a rich set of metadata with every object and supports virtually any content type. This distributed metadata strategy eliminates the need for SQL databases, which become a bottleneck in most archiving applications as the amount of data grows.

Unlike outdated platforms, Smarsh Archive utilizes a truly enterprise-class indexing technology with self-replication, self-sharding, and self-healing capabilities to create indexes for both object metadata and full text. It supports extremely sophisticated search queries with high performance.

This approach comes with a trade-off: while the Smarsh Archive offers significantly improved performance, scalability and resilience, the initial costs may be higher. However, these are offset by enabling organizations to:

- Greatly reduce the amount of data that needs to be duplicated across the various compliance and discovery applications
- Significantly lower the volume of data — and the associated legal costs — required for downstream review.
- Increase efficiency while decreasing management requirements

Smarsh Data Lifecycle Management

Our goal is to deliver the best possible total cost of ownership (TCO) for Smarsh Archive customers. One of the key elements that differentiates the Smarsh approach to data lifecycle management from the solutions described earlier is the recognition that not all data is created equal.

Email newsletters and client communications have vastly different value from the standpoint of compliance and discovery. We are developing the Smarsh Data Lifecycle Management (DLM) capability within Smarsh Archive to:

- Greatly reduce the costs of storing and managing data
- Automate the process of managing data throughout its lifecycle
- Continue to deliver high performance and other benefits that customers expect from the Smarsh Archive

Introducing Smarsh DLM

Smarsh DLM reduces the costs of operating Smarsh while continuing to provide highly differentiated, iterative search capabilities by leveraging:

- Classification
- Multiple storage types
- Granular retention parameters
- Different levels of indexing
- Legally defensible automated disposition processes

These technologies optimize data access speed and total storage costs for different types of content across their entire lifecycle.

Greater control across the data lifecycle

As data ages, it is generally less likely to be needed for investigation, discovery, or surveillance. In addition, there are some types of data (e.g., mass marketing emails with pre-approved content) that are infrequently accessed and less relevant in most searches. Reducing the level of indexing for older and less critical data significantly reduces the size of the indexes needed, which reduces the cost of providing archive services.

Smarsh DLM provides a range of tunable parameters to give customers more flexibility to manage data to address:

- Compliance mandates
- Governance mandates
- Cost metrics

Consider a company that has a compliance mandate of seven years and a governance mandate of 10 years. With an outdated archive, that company will probably keep everything on WORM for the entire 10-year period. No flexibility. No early deletions. No exceptions. With no ability to delete anything until the retention period expires, that company is stuck.

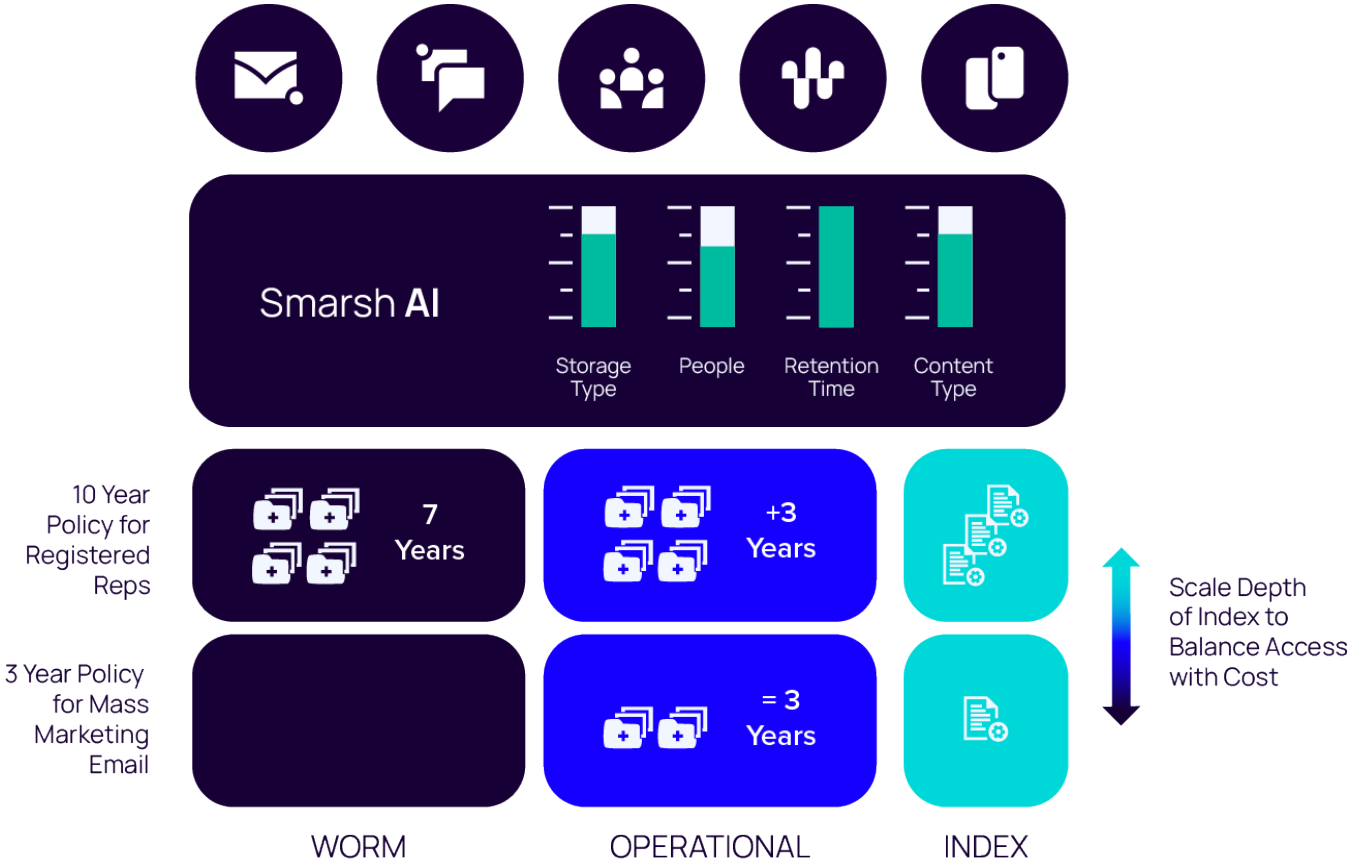


Figure 1: Smarsh DLM provides greater flexibility to address compliance and operational needs.

Smarsh DLM comes in to provide greater flexibility to span different types of data and satisfy different compliance (WORM) and governance mandates, as illustrated in the following figure.

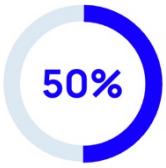
Smarsh DLM provides greater flexibility to address compliance and operational needs. For the scenario described above, when the seven-year compliance period ends, Smarsh enables the data to change from WORM to operational status. In certain instances (such as meeting GDPR requirements), the records manager can perform privileged deletion

With Smarsh, your data strategy isn't boxed in. It's built to adapt to real-world needs.

Smarsh DLM benefits

Smarsh DLM provides a set of policies and features that enable customers to capture, access, manage and dispose of data more easily while aligning costs with business priorities. It accomplishes this by partitioning data based on its importance and the customer's business needs.

This provides customers with:



Up to 50% reduced storage costs

Smarsh DLM eliminates the "one size fits all" policies of outdated archiving solutions. Users can easily define where data lives and for how long.

Improved recordkeeping

Smarsh enables enterprises to meet growing recordkeeping demands more effectively by addressing compliance (WORM) and governance (non-WORM) mandates and implementing legally defensible data disposition at scale.

Automated data classification and organization

Outdated archiving approaches are reliant on keyword search with limited retention/disposition capabilities. This approach increases 'dark data' risk. Smarsh DLM will enable AI-based classification to provide more flexible policy-based automation.

How Smarsh DLM works

Smarsh DLM uses a two-tiered approach to indexing data:

Tier 1 data includes full text indexing of both message and attachment content, providing the sophisticated, iterative search capabilities that Smarsh Archive is known for.

Tier 2 data is readily accessible, but uses a much lighter index, which significantly reduces cost. Searches of Tier 2 data are based on the metadata of the communications, such as custodian and date range. Smarsh has found that data retrieval teams use custodian and date range searches more than 80% of the time.

Search Field	Tier 1	Tier 2
Participant Details Name, Employee ID(s) Communication Aliases	✓	✓
Date Ranges Multiple Ranges, Down to the Minute	✓	✓
Group Attributes Security Groups, Distribution Lists, Custom Attributes	✓	✓
Organizational Details Division, Department & Building	✓	✓
Geographic Details Region, Country, City & State	✓	✓
Content Metadata Communications Type, Subject Line, Attachment Name & Count, Directionality	✓	✓
Content (Body, Attachments) Body, Attachments	✓	x

Table 1. Search attributes available for the tiers of Smarsh DLM.

To allow users of the system to perform global searches, Smarsh Archive provides a unified search interface that allows users to search both Tier 1 and Tier 2 at the same time. Search results from the unified search interface yield a single result set that includes data from both tiers and allows users to view and/or export content.

Archiving data directly to Tier 2 is typically used for older data being migrated from outdated archives. Some of our customers archive data directly to Tier-2 for mass marketing email and save significantly on data archival costs as nearly 60% of the data is mass marketing content. In the future Smarsh plans to expand direct archival to Tier 2 to application generated data (machine-generated data) and voice archival. By supporting data archival directly to Tier 2 for application-generated and voice data, customers can further lower archival costs. The actual savings will vary based on data distribution per customer.

In addition, Smarsh provides an Index on Demand capability, allowing Tier 2 data to be searched using the same extensive search parameters available for Tier 1 data when additional review of Tier 2 data is needed.

Discovery use case

The discovery use case illustrates how Smarsh DLM can be applied in practice, yielding cost savings with minimal or no impact to users. Consider the steps in the Electronic Discovery Reference Model (EDRM) workflow. Smarsh Archive automatically carries out the necessary information governance and identification steps, ensuring that all data is proactively retained and identified.

The Enterprise Discovery application utilizes the same unified search interface as the Smarsh Archive management application, eliminating the need to learn and access different tools for different functions.

After performing an initial search using metadata, data from the results set can then be collected into a case. The collection process includes an automated Index on Demand function that provides sophisticated search of all data using the Review interface of the Enterprise Discovery application, as illustrated in the following figure.

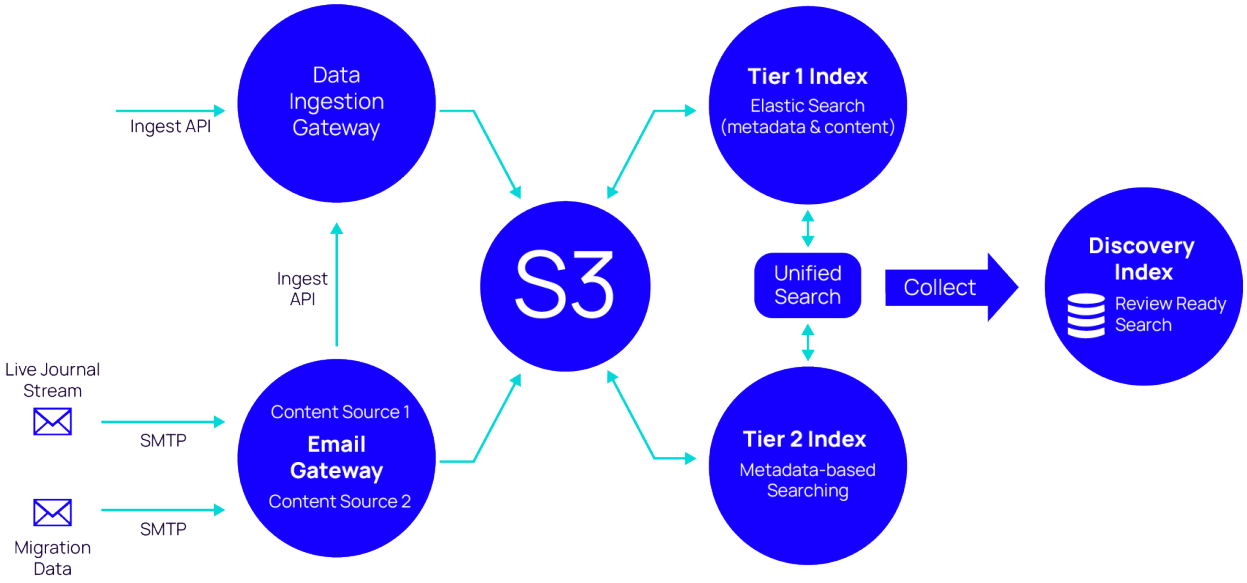


Figure 2: The Enterprise Discovery application enables a single query to collect data from both the primary and secondary tiers, and then automatically perform Index on Demand when the search results are collected into a case.

The Enterprise Discovery application enables a single query to collect data from both the primary and secondary tiers, and then automatically perform Index on Demand when the search results are collected into a case.

After an Enterprise Discovery user performs a collection action, all relevant data on the Tier 2 is elevated to Tier 1 where the message body and attachments are full-text indexed and available for full review. The following table shows various attributes of small, medium, and large cases, including time to activate.

Note that even for large cases, time to activate is typically less than four hours.

Search Field	Small Case	Medium Case	Large Case
No of Custodians	10	50	100
Date Range	6 months	12 months	2 years
No of Documents	180,000	1,800,000	7,500,000
Total Data	25 GB	250 GB	1 TB
Attributes Search (any time)	Seconds	Seconds	Seconds
Time to Activate	< 10 minutes	< 3 hours	< 8 hours
Content Search (after one-time activation)	Seconds	Seconds	Seconds

Table 2. Attribute of small, medium, and large cases showing time to activate and time to search for smarsh DLM

Smarsh DLM versus outdated archives

The following figure compares time to results for an outdated archive solution versus Smarsh Archive with Smarsh DLM. Every search with an older archive takes approximately the same amount of time to return results.

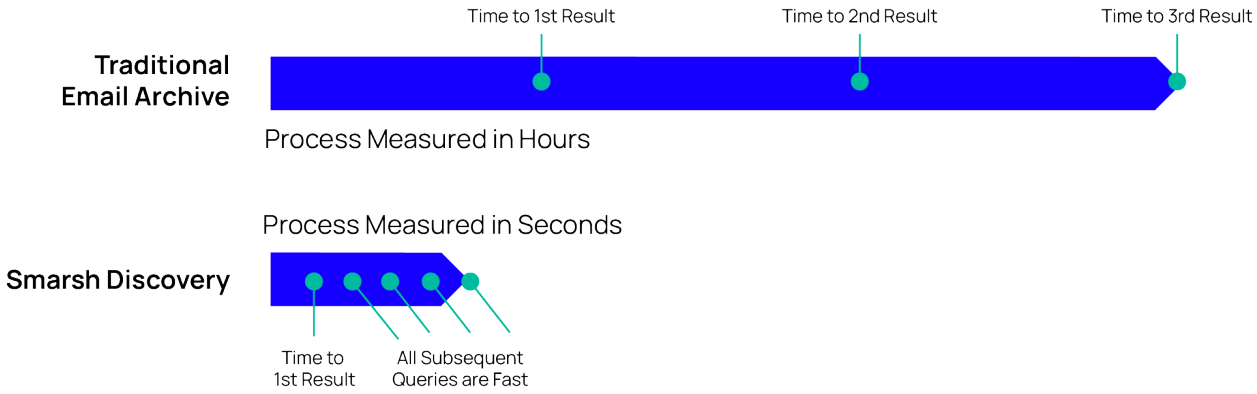


Figure 3: The tiered approach to data management provides lightning-fast search results with more sophisticated syntax and filtering.

Because searching in a traditional archive is so slow, many organizations eschew trying to perform any review within the archive. This often results in users performing only a single, basic search (e.g., custodian and date range), which leads to over-collection and preservation, and unnecessarily large exports. Exporting from traditional archives is already notoriously slow and costly. The net effect is extended time and excessive (per GB) fees for the export and import into downstream review platforms.

With Smarsh DLM, all data collected into a case is fully indexed and available for review in about the same amount of time it takes to complete an initial search in a traditional archive. Once activated, all subsequent searches return results in seconds.

Figure 4: Smarsh DLM with Index on Demand lowers costs but still provides premium search when needed.

Policy-based tiering

Smarsh DLM will also support policy-based tiering. With policy-based tiering, the Smarsh Policy Engine automatically promotes or demotes data between the two tiers according to the policies that organizations define.

For instance, a company may wish to move data for certain participants to Tier 2 after three years. Or they may wish to import data from a migration directly to Tier 2, but use the Policy Engine to ensure that the most recent data is indexed in Tier 1.

Figure 5: Smarsh DLM with Index on Demand lowers costs but still provides premium search when needed.

There are a variety of policy-based strategies that customers can use to achieve their data management and compliance objectives.

Data age

In most cases, data is accessed less and less as it ages. Aging data from Tier 1 to Tier 2 can reduce costs significantly without affecting operations. For example, for a supervision use case, an organization may want data fully indexed for six months. At that point, access rates have fallen, any potential issues have been identified and resolved, and metadata indexing is likely sufficient.

Participant group

Customers may want to archive data differently based on individuals or groups. For example, an organization might archive messages from specific users. This could be based on factors like Active Directory membership, location, etc. These messages are directed to Tier 1, while others are sent directly to Tier 2. This can be achieved by directly sending such data to the respective tiers.

Content type

An organization may want to archive data differently based on specific types of content. For example, if an organization conducts important business using Microsoft Teams or Zoom, that data may need to be retained in Tier 1 for six months, a year, or longer. This can be controlled by a demotion policy. There may be other types of content, such as spam email, mass marketing, and newsletters, that may not need to be retained in the primary tier at all. Such content could be archived directly to Tier-2.

Smarsh DLM will provide 'pre-flight' metrics when data is being collected, but before Index on Demand. Customers can use a dashboard to see how much data is in each tier and adjust policies to achieve the desired cost and performance objectives.

Retention policies and disposition

Smarsh supports the transfer of retention metadata as part of data migration (optional), simplifying Retention policy management. This enables customers to migrate the data to Smarsh Archive and then dispose of data that is no longer required. Smarsh also supports high-speed disposition with a rate of 44 million policy evaluations per second.

Classification: powered by Smarsh AI

In the future, the Smarsh Policy Engine will incorporate the classification capabilities of Smarsh AI to provide intelligent upfront routing of data to the proper tier. Smarsh AI is a mature, machine learning analytical engine currently implemented in Enterprise Conduct that analyzes communications and surfaces alerts for targeted signals. Rather than manually defining what data belongs in which tier, Smarsh AI will:

- Inspect the data
- Classify the data
- Determine where the data belongs
- Send data directly to the proper tier

This eliminates the need for full indexing of ingested data ultimately destined for Tier 2. Smarsh AI automatically classifies data as it enters the system, assigning tags such as “spam,” “financial transaction,” etc. This classification data is then used to identify and organize each data object and apply the right retention policy, providing additional intelligence and cost savings across the entire data lifecycle.

Smarsh is planning to extend the use of it’s industry-leading Cognition AI technology for even more automated DLM policy applications in the future.

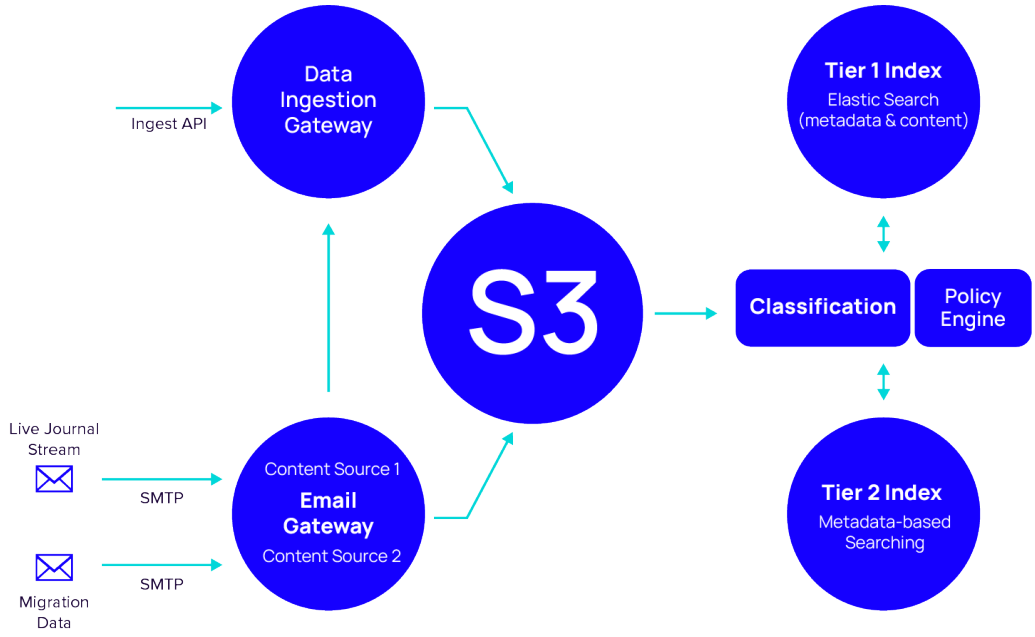


Figure 6: Smarsh AI classifies data automatically as they enter the system

Conclusion

Smarsh DLM provides enterprises with the sophisticated and high-performance search functions they need. It delivers greater reliability and resiliency than competing solutions while reducing the infrastructure costs associated with long-term data retention.

With the Smarsh Platform, Smarsh Archive, and Smarsh DLM, the data archive can become an active, intelligent, and valuable resource, not just a poorly managed dumping ground for aging data.

As data classification technologies continue to evolve, Smarsh is actively exploring ways to bring more intelligence and automation to your information governance strategy.

Contact us to learn how Smarsh DLM can help you better satisfy your compliance and data governance needs.

smarsh

Smarsh enables companies to transform oversight into foresight by surfacing business-critical signals in more than 100 digital communications channels. Regulated organizations of all sizes rely upon the Smarsh portfolio of cloud-native digital communications capture, retention and oversight solutions to help them identify regulatory and reputational risks within their communications data before those risks become fines or headlines. Smarsh serves a global client base spanning the top banks in North America, Europe and Asia, along with leading brokerage firms, insurers, and registered investment advisors and U.S. state and local government agencies. To discover more about the future of communications capture, archiving and oversight, visit www.smarsh.com.

White paper - 07/25

1-866-762-7741 www.smarsh.com @SmarshInc SmarshInc Company/smarsh