smarsh®

Smarsh Enterprise Archive

Platform Intelligence for the Future of Communications

SOLUTION BRIEF

OVERVIEW

Smarsh Enterprise Archive users can access new archived content and return search results in less than a second, and export over a million items a minute — orders of magnitude faster than competitive solutions.

INTRODUCTION

Every organization today works with a set of communication and collaboration tools that has never been more diverse — from email and text messaging, to voice-enabled apps and AI-powered collaboration platforms.

The impact of this evolution has been profound. Each new tool consists of a unique combination of content, metadata, and context — defined by activities that take place in a persistent chat, individuals joining or leaving a meeting, or participants editing files or documents. Failing to capture the context of these communications can have severe consequences when responding to e-discovery and regulatory inquiries.

This conversational complexity also creates new IT complications in capacity planning, managing system availability, and ensuring that an archiving system will continue to perform adequately as content volume and variety continue to grow. Simply moving to "the Cloud" does not address these challenges if an archiving solution was built on traditional architectures and email-centric technologies, designed for on-premise deployment.

Archiving content, metadata, and context across multiple sources also elevates the importance of how information is stored. Reviewing a conversation that started on Microsoft Teams, continued on WeChat and ended on mobile text requires a common vocabulary that allows any source of information to be normalized across the entire archived data corpus. Leveraging a content platform designed to deliver against this content volume and variety not only improves accuracy and effectiveness in managing risk obligations, but serves as the basis for platform intelligence that can be leveraged by other applications, including content surveillance, business intelligence, and AI-enabled customer analytics.

This solution brief outlines how the Smarsh Enterprise Archive was designed to address these challenges. It is intended for IT directors and IT professionals who are responsible for selecting a cloud-based communications archive, and for those responsible for replacing a legacy email archive.

> *"Organizations continue to move away from on-premises EIA platforms as part of larger cloud strategies.*
>
> – **Gartner**, Critical Capabilities for Enterprise Information Archiving (2019)

THE PROBLEM

Legacy archiving technology is not built for today's communication tools

Electronic communications archiving has been a requirement for regulated organizations for more than two decades. Unfortunately, moving from a first generation, on-premise archive to the initial versions of archives hosted in vendor-operated data centers simply replaced one set of challenges with another. Hosted archiving service providers must consume already-scarce resources to keep the lights on, manage high availability/disaster recovery (HA/DR) configurations, and commission/decommission infrastructure as changes in archiving load arise.

While these functions are vital, they distract from higher value areas of innovation that can be incorporated into the archiving service. This has prevented firms from staying in step with the tools demanded by employees and clients today: collaboration, conferencing, voice and video, mobile apps, text messaging, etc.

Email-centric archives were not designed to handle these emerging, interactive communications. They simply process non-email content as multiple email messages, losing metadata, context, and point-in-time edits and deletions along the way. This adds significant complexity in attempting to meet discovery and regulatory demands, which are governed by rules that do not distinguish one communication type from another.

Knowing how an archive's underlying technologies and deployment models operate at the largest scale is vital to planning for datasets that will only continue to grow in volume, variety and velocity over time.

> *The limitations of today's archives are a direct result of the underlying technology that the system is constructed upon, from the platform and the index engine to the archive database. You have to build a cloud-native application from the ground up to take advantage of the scalability, performance and high availability of public cloud services.*
>
> – **Greg Vesper**, Chief Product Officer, Smarsh

THE SOLUTION

A cloud-native content platform built on today's leading cloud infrastructure

The Smarsh Enterprise Archive is the first of its kind, built with a cloud-native architecture. This distinction enables cloud-scale and multi-cloud deployment and represents a step-function gain in the enterprise information archiving market. It delivers the flexibility and scale required for global enterprises seeking a future-proofed, business-critical information platform.

By "cloud-native," we are referring to the design of the archiving platform itself. Cloud-native architecture shortens time-to-value by providing a consistent development and automated management experience across leading public cloud infrastructure, including Amazon Web Services and Microsoft Azure. These attributes include:

- Modular, distributed archiving "microservice" components that can scale independently and horizontally to avoid performance bottlenecks and single points of failure
- Continuous delivery via agile methodologies to speed deployment times into target cloud environments
- The ability to scale resources up or down dynamically across multiple geographic regions and availability zones therein, to enable quick response to business demands
- The ability to leverage industry-leading network, security and data management controls delivered by today's leading cloud infrastructure providers

The innovations of public cloud infrastructures from AWS and Microsoft mean that cloud-native archiving platforms can now deliver unparalleled availability, massive scale and flexibility as follows:

High Availability is delivered through the use of "triple-active" deployment. Triple-active means the Enterprise Archive runs simultaneously in multiple data centers distributed across three availability zones with fully elastic capacity, creating high availability and super durability. This deployment method eliminates the cost and complexity of a traditional disaster recovery site, while effectively reducing recovery time objectives (RTOs) and recovery point objectives (RPOs) to zero. If and when these events occur, firms should not face lengthy delays in fail-over to a secondary passive site, or time-consuming recovery of data once the primary site is back online.

Massive Scale is not just about delivering against a search SLA for the first page of search results; it is ensuring that large projects can be completed without delay — from start to finish. Additional resources can be easily deployed to manage the ingestion of large volumes of historical content. That content can be efficiently preserved, searched for and reviewed, and ultimately be exported to the next step in the legal or compliance workflow at whatever volume is required to satisfy the request. Adding capacity to any individual component within the Smarsh Enterprise Archive is achievable without impacting performance of other components of the system.

Flexibility typically means the ability for multi-national organizations to quickly respond to geographic or regulatory changes. This can often entail the deployment of a new archiving location to maintain data within a specific geography without forcing you to use the next closest available location. It also provides the ability to quickly enable support for content sources, enabling firms to move faster and better address client needs.

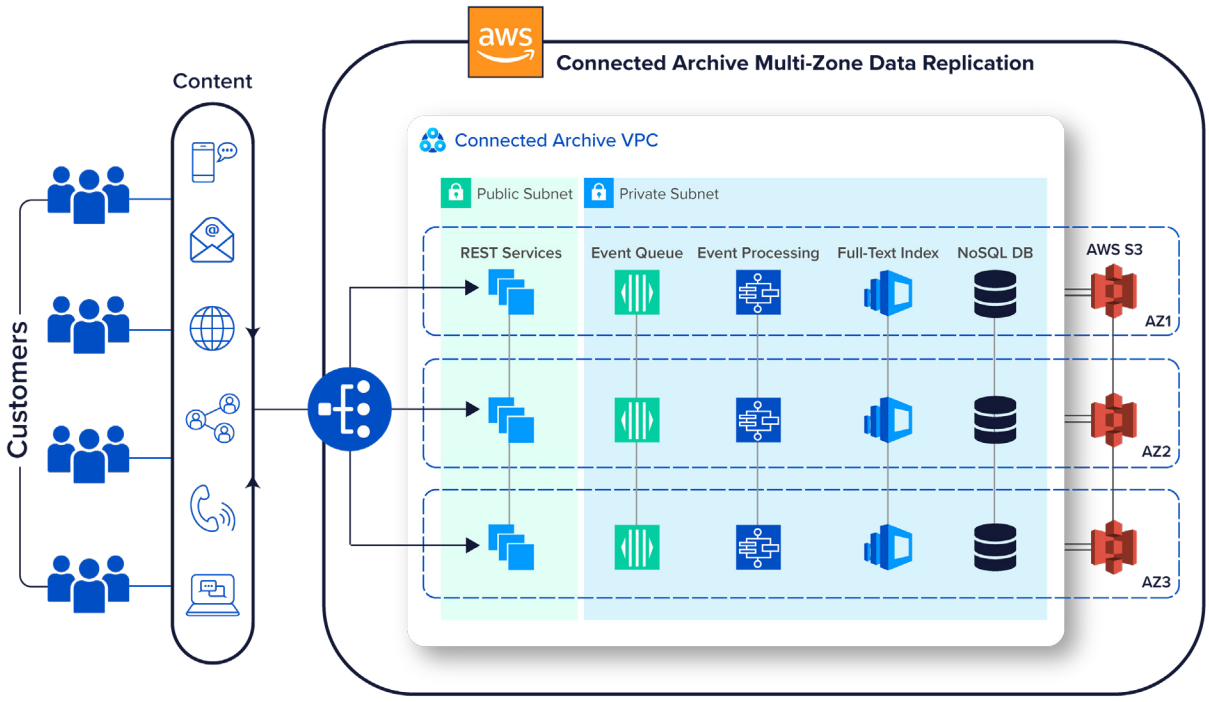


Figure 1: Smarsh Enterprise Archive deployment architecture

Figure 1 illustrates how Smarsh Enterprise Archive deploys its cloud-native archiving service within a triple-active deployment model. Content and metadata are captured from multiple content sources, and are then replicated across three unique deployment zones (AZ1-AZ3), where archive processing is distributed across multiple nodes, with additional processing capacity added as required.

Enterprise Archive: the content platform for the future of communications

Delivering a future-proofed platform for all content types that can satisfy the most complex compliance and discovery demands at the largest scale is not a simple task. It requires:

- Purpose-built architecture
- A method of capturing and playing back native conversational context
- Extensibility to support the next content source
- APIs that allow content to be used by other enterprise applications

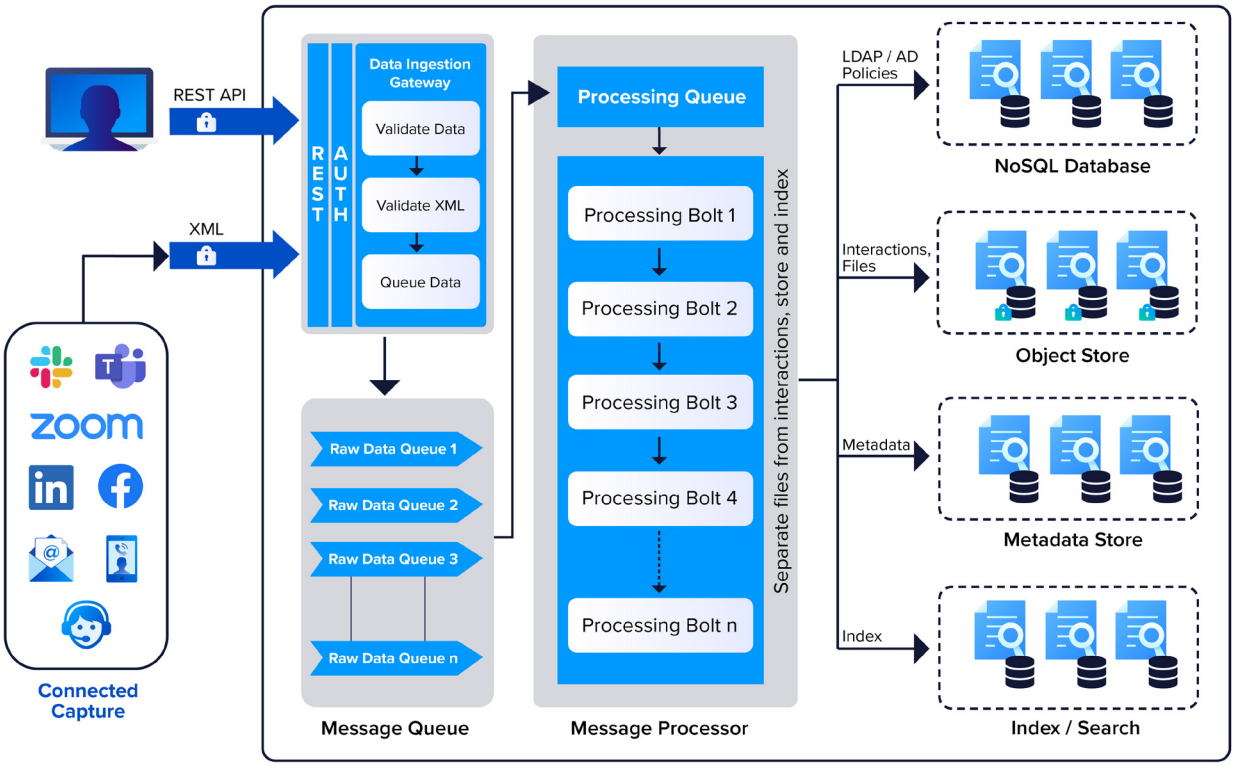
Open, scalable architecture for unsurpassed ingestion, search and throughput

The Smarsh Enterprise Archive has a distributed architecture that allows for many tasks to be executed in parallel and eliminate single points of failure. It can ingest, find and export content much faster than first generation archives, regardless of the archived content's volume. Key components of this infrastructure include:

- A scalable front-end queuing infrastructure to truly distribute processing of critical workloads
- A massively scalable content store to process metadata using the same underlying technology as today's most popular cloud applications
- Highly scalable indexing and search subsystems derived from leading open source technologies to make storing, indexing and searching more efficient
- Multi-instance infrastructure, with collections of compute and storage nodes, to which additional nodes can be added to scale horizontally
- Processing topologies that provides a configurable pipeline of tasks for ingestion, search and export in parallel

The diagram below (**Figure 2**) shows how Enterprise Archive uses distributed, parallel computing, and the elasticity of various components to achieve unsurpassed throughput for any content source.

Figure 2. Open, scalable architecture



Context-aware: a patented approach to capture & play back conversational context

The Smarsh Enterprise Archive is constructed as an object-based content platform so that organizations can archive any communication channel in its native format. Enterprise Archive uses patented technology to store data as “snapshots,” which represent content, metadata and event-based information. Snapshots can then be threaded to allow events to be combined for additional context, enabling a true context-aware archive. Snapshots are especially important to connect interactive conversations, such as persistent chats, or activities taking place within multi-modal collaborative platforms such as Microsoft Teams or Slack.

The traditional approach to archiving collaborative content is to create a separate email for every single interaction (a comment, reply, edit, etc.). For example, when someone posts a message, that generates an email. If a second individual posts a reply, it generates another email in the archive. If either person then deletes a comment, yet another email is generated, and so on.

There is simply no effective way for reviewers to understand the relationship between those various emails unless they spend time combing through metadata or looking at each individual message to try to thread them together. **Figure 3** illustrates the difference between the loss of context with an email-oriented archive, and how that same conversation is displayed by the Smarsh Enterprise Archive.

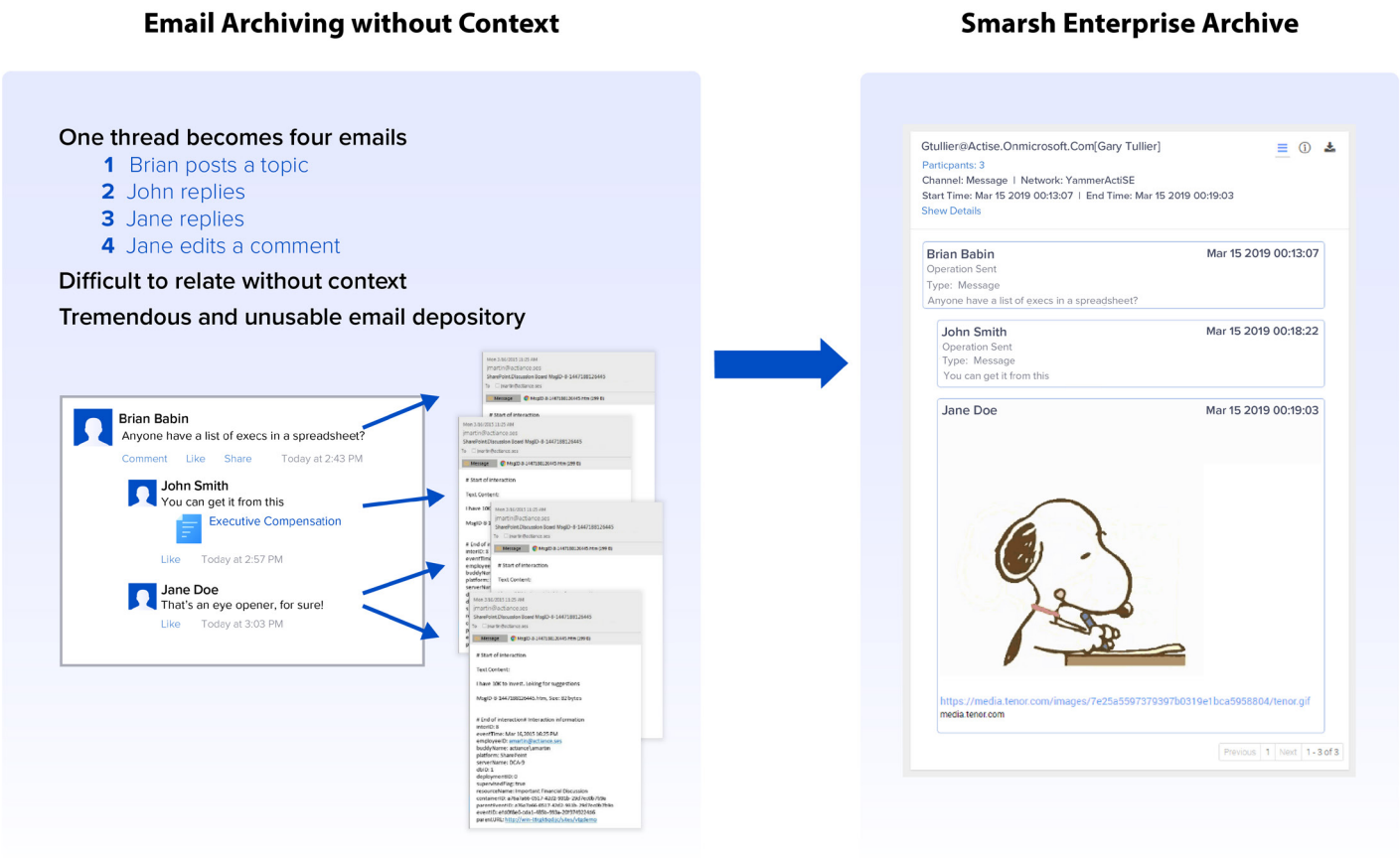


Figure 3. Context preservation with Smarsh Enterprise Archive

APIs and extensibility to connect and harness external applications and advanced analytics

Enterprise Archive ingests data as XML to enable any type of data to be stored, searched and exported without having to rewrite any part of the application. This extensibility can also be used to quickly respond when new file types and forms of communication are introduced into the market.

Smarsh delivers extensibility via full APIs that enable the Enterprise Archive to work seamlessly with other enterprise applications. This enables partners and customers to further enhance the value of their data with third party review products and services or new business intelligence applications, and to add new custom content types. These APIs include:

- **Ingestion APIs** to capture and deliver email, IM, social, mobile, voice and any other content type into an Enterprise Archive instance
- **Enrichment APIs** to enable third-party applications to send additional data into the Enterprise Archive, including transcripts, sentiment analysis and risk-based scoring
- **Export APIs** that allow clients to send data to downstream applications, allowing those applications to submit search parameters and bulk-retrieve relevant messages

With these APIs, Smarsh customers can take advantage of existing investments in best-of-breed machine learning, surveillance and analytics applications to generate additional value from their data.

Data is the lifeblood of any AI application. Accordingly, the quality of the underlying dataset is of paramount importance in any AI application. One of the most critical steps in building an AI application is to obtain and build the underlying dataset, as AI applications are generally best positioned to yield meaningful results when the underlying datasets are substantially large, valid, and current.

– FINRA report on Artificial Intelligence in the Securities Industry, June 2020

Platform Intelligence and its impact on business processes

The Smarsh Enterprise Archive was purpose-built to allow firms to capture, store and control any communications content source at the highest measures of scale. This has allowed it to be used as a system of record for highly regulated and frequently litigated firms to meet requisite obligations with unsurpassed throughput. This also creates the platform intelligence that can be leveraged by other enterprise applications and advanced analytics to fully harness archived content as a high-value information asset. The potential impact on business processes is virtually limitless, including the following use cases:

E-discovery: Legal teams can reduce the time and expense associated with e-discovery and litigation response. Playback of interactive content from collaboration platforms via patented snapshot technologies lets legal teams see conversations in their original context, thus eliminating the time and risk of reassembling multiple disjointed messages. Directionally, our platform will enable intelligent discovery by auto-classifying and tagging relevant content to reduce review time, in addition to providing data visualization and analytics. Extensibility adds to this strategy by efficiently delivering datasets to downstream early case assessment (ECA) tools that help legal teams surface additional custodians or keywords and refine legal case parameters.

Supervisory review: For firms required to supervise the communications of regulated users, responding to today's communications volume and variety creates a unique set of challenges. Relying solely on established lexicons and policies to spot potential policy infractions can often produce high false-positive rates and lead to time wasted reviewing duplicate items. In this context, the Smarsh platform intelligence can be highlighted by its echo cancellation features that allow previously reviewed items to be excluded from review sets, thus **reducing review times and volumes by over 20%, with 100% precision**. Additional noise reduction and accuracy improvements by leveraging AI to understand the semantics of communications will be added to the platform over time.

Learn more about
Echo Cancellation

ECHO CANCELLATION
SOLUTION BRIEF

Content Surveillance: Enterprise Archive's extensibility allows red flags surfaced within supervisory policies to be shared with external content surveillance technologies to examine patterns and behaviors surrounding that red flag over time. Those findings can then be shared via the enrichment APIs with the Enterprise Archive to update policies and risk-scoring models to spot emerging risks more effectively.

Other external applications: Content maintained within the Smarsh Enterprise Archive is fully accessible to feed other enterprise applications in support of know your customer (KYC), customer relationship management (CRM), or other AI-powered customer facing applications.

CONCLUSION

The Smarsh Enterprise Archive was built with the future in mind. We were driven to create a scalable content platform for any and all new communications channels, accessible via APIs. This allows users to enrich that content, adding additional layers of metadata around a message. But it also allows users to enable the content, not just for archiving retention or e-discovery or supervision, but for whatever else an organization might want to do with some of their most valuable data: the communications of their employees. This is particularly important when considering emerging advances in machine learning and communications intelligence.

The Smarsh Enterprise Archive and its platform intelligence allows firms to meet regulatory obligations, manage risk more effectively and derive increased value from archived content as a business-critical asset.

> *Where will you and your enterprise be a year from now? Two years from now? Three? Look at the explosion of data. If you build your compliance ecosystem on the wrong foundation, it will create a problem that will never go away. You will have to fix it eventually. So, think strategically, think deeply, and solve tomorrow's problem today.*
>
> – **Greg Vesper**, Chief Product Officer, Smarsh

Smarsh® is the recognized global leader in electronic communications archiving solutions for regulated organizations. The Smarsh Connected Suite provides innovative capture, archiving, e-discovery, and supervision solutions across the industry's widest breadth of communication channels.

Scalable for organizations of all sizes, the Smarsh platform provides customers with compliance built on confidence. It enables them to strategically future-proof as new communication channels are adopted, and to realize more insight and value from the data in their archive. Customers strengthen their compliance and e-discovery initiatives, and benefit from the productive use of email, social media, mobile/text messaging, instant messaging and collaboration, web, and voice channels.

Smarsh serves a global client base that spans the top banks in North America and Europe, along with leading brokerage firms, insurers, and registered investment advisors. Smarsh also enables federal and state government agencies to meet their public records and e-discovery requirements.

For more information, visit www.smarsh.com.

US: +1 (866) 762–7741 UK: +44 (0) 20 3608 1209 www.smarsh.com