

End-to-End Reconciliation: Confident Data Migrations in Regulated Industries

The Digital Safe to Enterprise Archive (DS2EA) end-to-end reconciliation provides a continuous assurance model for real-time visibility, early detection, and provable audit trails.



Executive summary

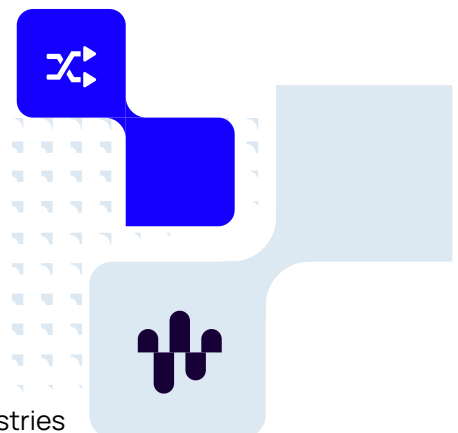
In regulated industries, migrating decades of sensitive, compliance-bound data from previous archives to modern platforms is a high-stakes process that demands more than simple transfer. It requires forensic-level preservation of content, context, and metadata. Organizations must move terabytes or even petabytes of regulated information while maintaining uncompromised data integrity. Any loss, corruption, or undocumented change poses severe regulatory, financial, and reputational risks that can impact stakeholders across the enterprise.

Smarsh Digital Safe to Enterprise Archive, or DS2EA, end-to-end reconciliation enhances migration assurance from a final checkpoint into a continuous, automated process that operates throughout the migration lifecycle. By tagging each record with a unique source identifier, capturing events at every stage, and performing message-level verification in both directions (source-to-destination and destination-to-source), the system delivers three critical capabilities:

1. **Real-time visibility** through a purpose-built reconciliation dashboard that tracks progress metrics, categorizes failures, and enables efficient remediation
2. **Early detection** of issues during migration rather than after completion, allowing teams to identify and resolve problems within weeks of ingestion
3. **Provable audit trails** through immutable event logging that documents the complete chain of custody for every message

This comprehensive approach ensures operational confidence by enabling migration teams to efficiently triage issues and maintain traceability throughout the process. It delivers audit readiness through transparent, verifiable documentation that satisfies the most stringent regulatory requirements. Most importantly, it builds customer trust by providing clear accountability at every phase of the migration lifecycle.

With Smarsh DS2EA end-to-end reconciliation, organizations can retire older systems with confidence. They'll know that every message has been accurately, completely, and verifiably migrated — meeting the strictest compliance standards without disruption to business operations.





The challenge: How can regulated organizations confidently migrate decades-worth of data?

Data is an increasingly valuable asset. It's not just about retaining and accessing data, but also analyzing, indexing, searching, automating, gaining insights, and enriching it. And, of course, it must be secure, resilient, and scalable.

This means data has outpaced old archive technologies. Organizations in regulated industries — like those in the financial services sector — are making the move to modern archives to meet the needs of their business.

But migrating decades of sensitive, compliance-bound information while maintaining preservation of the data's content, context, and metadata isn't merely a technical challenge. It's a regulatory imperative where failure can trigger significant enforcement actions, financial penalties, and lasting reputational damage for all stakeholders involved.

The complexity of these migrations stems from multiple compounding factors that create a perfect storm of technical and compliance challenges:

- **Limitations of existing systems:** Source archives often contain inconsistent, incomplete, or poorly structured records originating from systems implemented decades ago, with varying standards for data validation and storage.
- **Non-standard message formats:** Many archived messages do not fully comply with SMTP standards (such as RFC 822) yet were accepted without issue in earlier systems.
- **Diverse data sources:** Non-SMTP data sources, such as trade systems and proprietary platforms, follow entirely different data formats and protocols, requiring specialized handling.
- **Metadata inconsistencies:** Messages in previous archives may contain missing or corrupted/malformed metadata that modern platforms (with their stricter validation rules and RFC standards adherence) would typically reject.
- **Scale and scope:** The sheer volume of data — often terabytes or petabytes — combined with strict regulatory requirements creates unprecedented verification challenges.

The need: End-to-end reconciliation is a data migration requirement

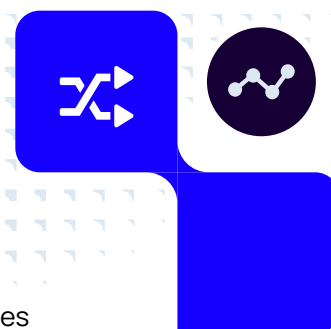
In this high-stakes environment, reconciliation emerges as the critical mechanism for proving that every message from the source archive has been successfully and accurately ingested into the destination archive.

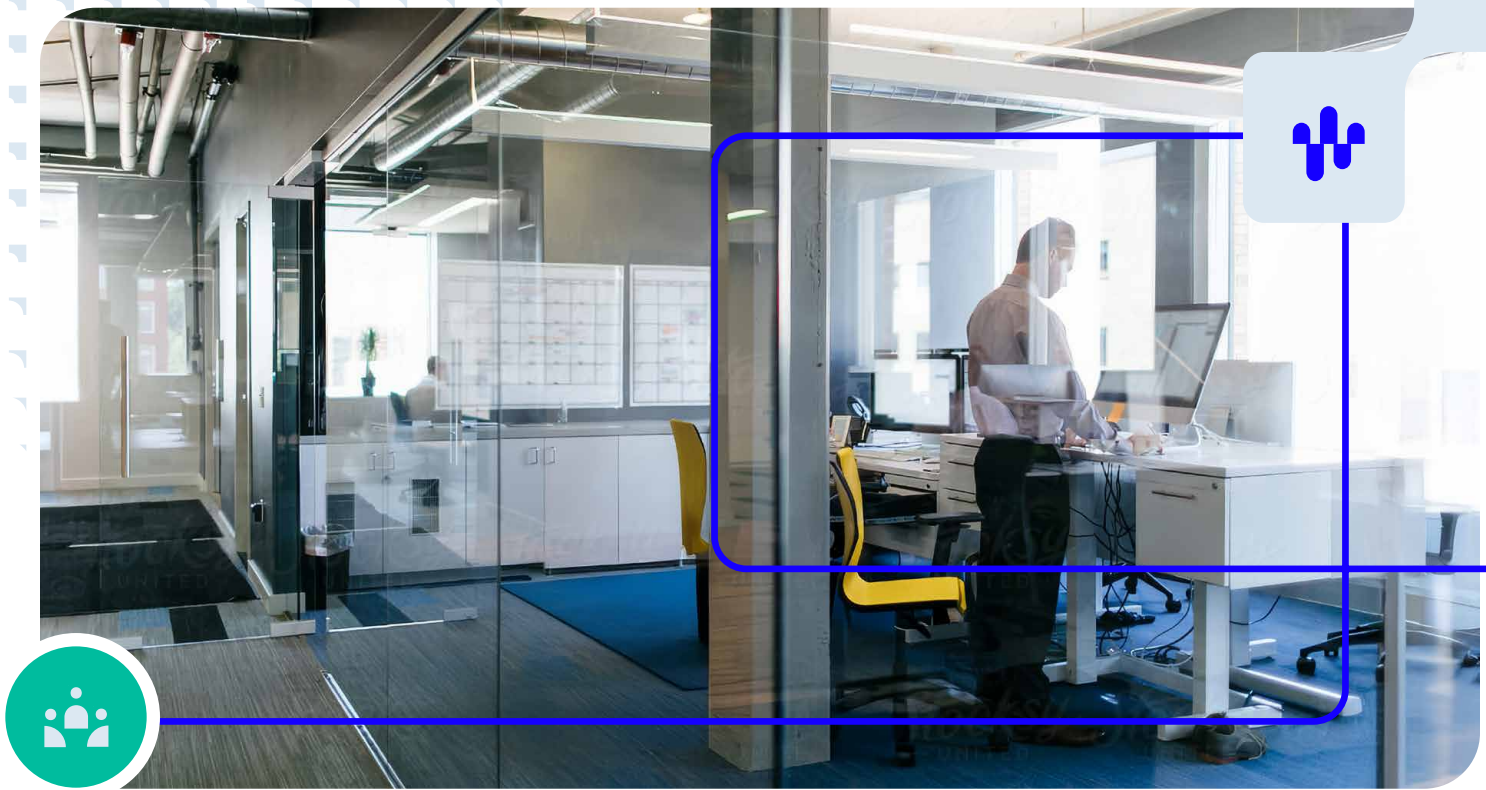
The process begins by generating authoritative data references (identifiers) from the source system and capturing the identity and metadata of every record. As data is migrated, acknowledgments from the destination archive against each source identifier are recorded and later cross-referenced against data in the destination to generate comprehensive reconciliation reports.

These reports provide transparent, auditable evidence of message-level integrity and form the foundation for formal Proof of Migration documentation. They present a clear mapping of source record counts to their confirmed counterparts in the destination, while also documenting exceptions with reference details for further analysis.

DS2EA end-to-end automated reconciliation represents a significant advancement over traditional approaches, delivering greater confidence and speed at scale. Rather than treating reconciliation as a post-migration validation exercise, this approach embeds verification throughout the migration process, enabling:

- **Incremental reconciliation** during migration rather than after completion
- **Early detection and resolution** of issues, allowing full reconciliation to be completed within weeks of ingestion
- **Comprehensive verification** ensuring that source identifiers are preserved throughout the pipeline
- **Transparent documentation** of any discrepancies, with detailed explanations for each exception
- **Evidence-backed chain of custody** that customers, legal teams, auditors and regulators can rely upon





This approach is especially critical when migrating decades of regulated data, which must remain accurate, complete, and support regulatory requirements throughout the transition.

For customers, this represents more than a technical migration — it requires placing trust in a new system, a new set of controls, and a complex process that must maintain uninterrupted adherence to regulatory standards. In such high-stakes environments, technical assurances alone are insufficient. What's required is full transparency, verifiable audit trails, and clear accountability throughout every phase of the migration lifecycle.

End-to-end reconciliation fulfills this need by enhancing migration from a leap of faith into a structured, provable operation. It enables organizations to retire older systems with confidence while ensuring regulatory obligations are maintained without disruption — a critical capability in today's increasingly regulated business environment.



Smarsh Enterprise Archive: End-to-end reconciliation

Architecture overview key terms

Archival data migration

The process of transferring data from one archive application or source to another while maintaining complete fidelity of content, context, and metadata. Unlike standard data migrations, archival migrations in regulated environments must preserve the full compliance posture of each record, ensuring it remains unaltered, secure, and audit-ready in its new location.

Destination archive

The target system where migrated data will reside after transfer. In this document, “destination” specifically refers to Smarsh Enterprise Archive, a purpose-built repository designed to maintain regulatory compliance while enabling efficient search, supervision, and discovery of communications data.

Source archive

The original system from which data is extracted for migration. Source archives may contain decades of historical messages and associated metadata, often stored in earlier formats or non-standard structures that require specialized handling during extraction and transformation.

Reconciliation

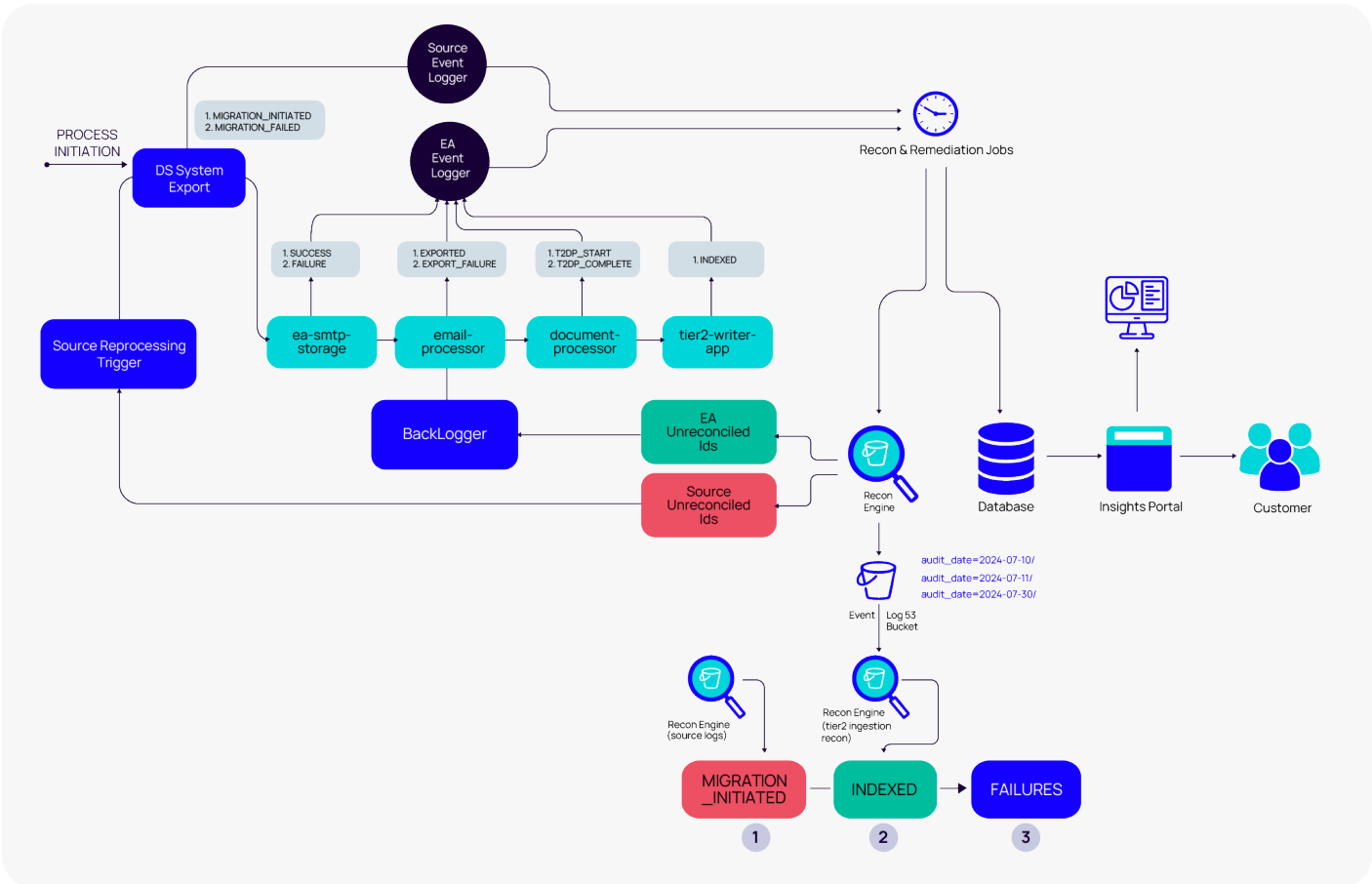
The systematic process of verifying that every message from the source has been successfully archived in the destination, with detailed error reports for any failures. Reconciliation provides clear confirmation to stakeholders that data has been accurately and completely transferred – a critical requirement for ensuring trust in the migration process and a prerequisite before the original system can be safely decommissioned.

End-to-end reconciliation

A comprehensive reconciliation approach that tracks each message from source to destination using unique source identifiers preserved throughout the migration pipeline. This methodology enables precise status tracking, accurate record validation, and message-level verification across systems, providing forensic-level assurance of data integrity and completeness.



A successful and verifiable data migration requires a well-orchestrated architecture composed of tightly integrated components that ensure fidelity, security, traceability, and auditability at every step. Although migration processes are evolving, the following subsystems form the current foundation of the migration and reconciliation process:



A. Digital Safe System Export (for data extraction from the source)

The Digital Safe System Export serves as the entry point to the migration pipeline, responsible for extracting both message content and associated metadata from the source archive.

- Transforms source data into a structured payload that complies with destination ingestion requirements
- Preserves the original message in its entirety to maintain compliance fidelity
- Tags each message with a unique, persistent source identifier to enable accurate tracking
- Transmits prepared messages downstream for storage and further processing
- Generates success events to the Event Logger upon completion of extraction
- Records failure events when issues occur to support traceability and reconciliation

This component:

This initial extraction phase establishes the foundation for all subsequent reconciliation by creating the authoritative record of what should be migrated and assigning the tracking identifiers that will follow each message throughout the process.

B. Secure transfer layer

All message data is transmitted through a secure, encrypted network channel to the destination archive. This layer:

- Ensures data-in-transit security to protect sensitive information
- Satisfies regulatory requirements for confidentiality and integrity
- Establishes a trusted conduit between source and destination systems

The secure transfer layer prevents unauthorized access or modification during transmission.

C. Destination archive ingestion workflow

The destination archive employs a multi-stage ingestion pipeline with each component responsible for a specific phase of message processing. Throughout this pipeline, events are emitted to the Event Logger to enable traceability and support the reconciliation process.

1. Ingestion endpoint (“EA-SMTP-STORAGE”)

The ingestion process begins at the destination endpoint, where messages from the source scrape system are received. This component:

- Securely stores incoming messages in WORM (Write Once Read Many) storage
- Returns immediate acknowledgment to confirm successful delivery
- Emits success events to the Event Logger to document receipt
- Generates failure events if message storage encounters issues

This initial confirmation of receipt establishes the first verification point in the destination system, confirming that messages have successfully crossed the boundary between source and destination environments.



2. Message parsing stage (“EMAIL-PROCESSOR”)

In this stage, the content of stored messages is retrieved for structural analysis and transformation. The processor:

- Retrieves messages from secure storage
- Parses content to extract relevant metadata
- Transforms messages into a standardized internal XML format
- Combines extracted metadata with original message content
- Logs success events upon completion of parsing
- Generates error events when parsing fails
- Enables technical evaluation of failures to determine recoverability
- Classifies unreadable messages as permanent failures after remediation attempts

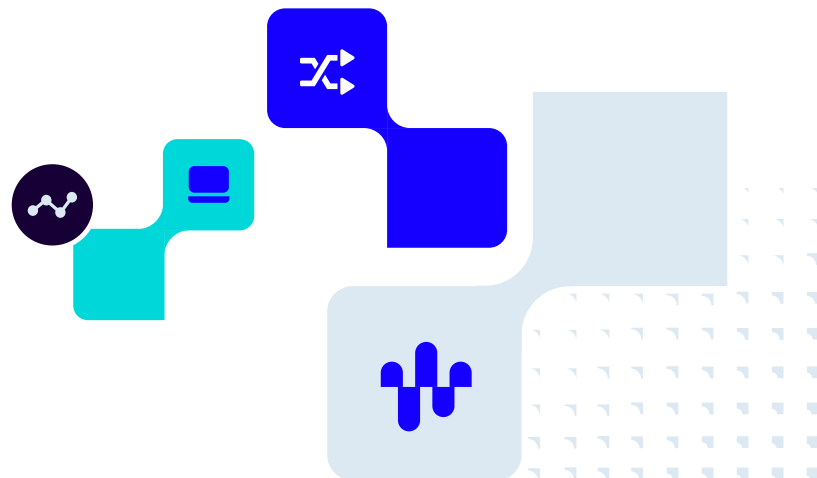
This transformation stage is critical for normalizing diverse message formats into a consistent structure that can be efficiently processed and indexed by downstream components.

3. Metadata extraction and validation (“T2DP-Processor”)

At this stage, the system performs detailed metadata analysis according to customer-specific requirements. The processor:

- Evaluates internal XML files against predefined metadata field whitelists
- Extracts and validates required metadata for compliance purposes
- Passes validated metadata to the indexing engine
- Logs both start and completion events for each message
- Implements automatic retry mechanisms to handle transient issues – no failures are expected

This component ensures that all required metadata is properly extracted and validated before messages proceed to final indexing, maintaining the compliance attributes necessary for regulatory purposes.



4. Indexing and search enablement (“Bulk-Indexer”)

In the final stage, messages are made discoverable within the destination archive. The indexer:

- Groups messages into efficient processing batches
- Indexes content and metadata into the search engine
- Makes messages searchable via the Archive UI
- Records “indexed” events to signify successful completion
- Implements automatic retry handling for transient issues — no failures are expected

Only after successful indexing can messages be considered fully migrated and available for compliance, supervision, and discovery purposes within the destination archive.

5. Event logger (Audit Layer)

The event logger creates an immutable record of all activities throughout the migration process. This system:

- Captures audit events at every transition point from extraction to archival
- Receives events via RESTful APIs with structured metadata
- Records source ID, timestamps, and migration metadata for each message
- Functions as an immutable audit ledger for the entire message lifecycle
- Provides the essential data foundation for reconciliation reporting

This comprehensive event capture enables the detailed tracking necessary for message-level reconciliation and creates the audit trail required for regulatory compliance.

6. Reconciliation engine

The reconciliation engine performs the critical analysis that verifies migration integrity. This component:

- Periodically pulls event data from the event logger
- Performs message-level comparisons across the entire pipeline
- Tracks each message’s journey from initiation to final indexing
- Uses original source identifiers to match events across components
- Detects gaps, mismatches, or missing events in the process flow
- Generates comprehensive Reconciliation Reports documenting success rates
- Identifies failures with reference details for remediation
- Provides transparent, auditable evidence of message-level integrity

The engine’s sophisticated analysis capabilities transform raw event data into actionable intelligence about migration status and issues requiring attention.

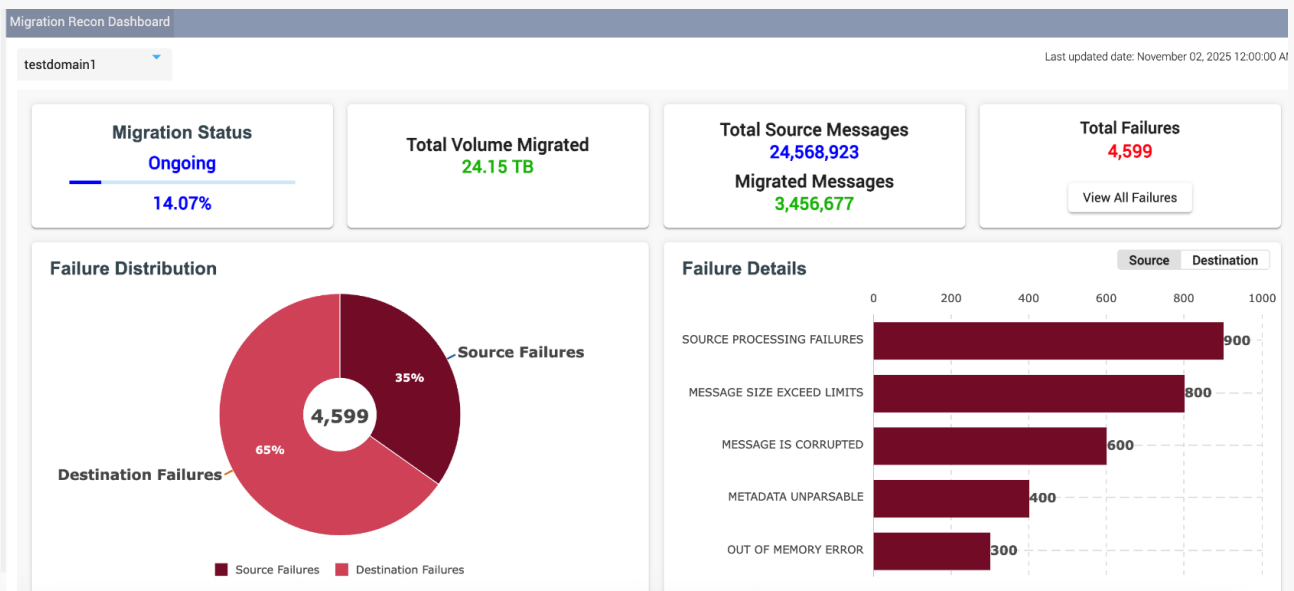


7. Reconciliation dashboard

The reconciliation dashboard serves as the information center for monitoring and managing migration activities. This interface:

- Provides real-time visibility into reconciliation status
- Displays key metrics refreshed at regular intervals
- Enables operational teams to track progress throughout the lifecycle
- Supports investigation of issues through detailed failure categorization
- Links to a dedicated Recon Console for granular analysis
- Facilitates efficient triage and remediation workflows

This dashboard transforms complex reconciliation data into accessible visualizations that support informed decision-making throughout the migration process.



Key metrics and visualizations

The Reconciliation Dashboard provides a comprehensive view of migration status through carefully selected metrics that enable teams to monitor progress, identify issues, and prioritize remediation efforts.

Reconciliation progress

A dynamic view showing the percentage of messages successfully reconciled against the total scope. This metric helps stakeholders track how far the migration has progressed from a verifiability standpoint and provides an at-a-glance assessment of overall completion status.

Total source messages:

The complete count of messages expected to be migrated, based on source profiling. This establishes the baseline against which all other metrics are measured and provides the denominator for calculating success percentages.

Successfully migrated messages

The number of messages that have been both ingested and indexed at the destination, confirming their availability for compliance and business purposes. This metric represents the fully completed migrations that have passed all validation checks.

Failure count

A running tally of messages that encountered issues during migration or reconciliation. This metric highlights potential compliance gaps requiring attention and remediation.

Failures are not only counted but also categorized to enable efficient troubleshooting and resolution. The dashboard offers a breakdown of failures by type, such as:

- **Extraction failures:** Issues occurring during the initial reading or emission of messages from the source, typically indicating problems with the source archive or extraction process
- **Parsing or processing errors:** Problems encountered during the transformation, validation, or storage of messages in the destination archive

This categorization allows teams to quickly identify patterns, assess severity, and prioritize remediation efforts based on failure types and volumes. For more granular investigation, the dashboard links to a dedicated Recon Console, where users can:

- Search for specific document references based on failure types or message identifiers
- Download failure lists for offline analysis, escalation, or remediation workflows

This integration of high-level metrics with detailed investigation capabilities supports efficient triage, improves traceability, and equips migration teams with the tools needed to respond swiftly to issues throughout the migration lifecycle.

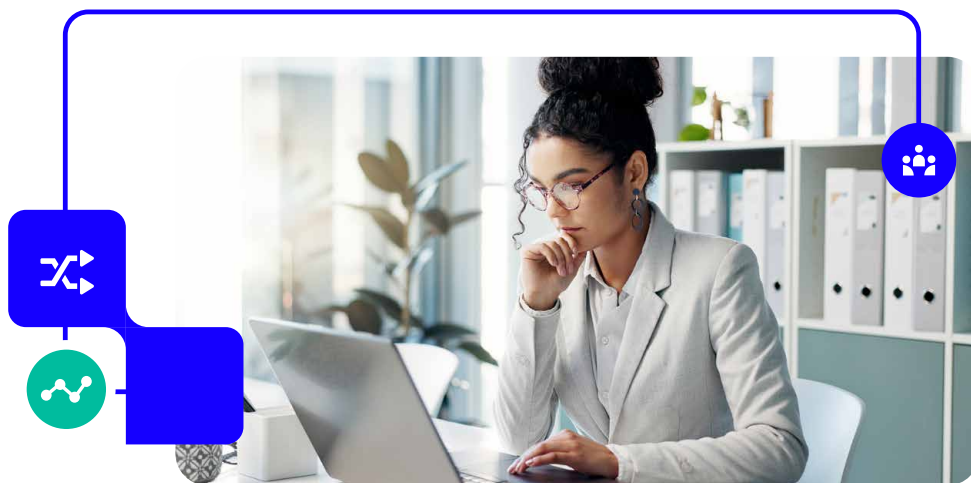
Pre-migration and setup requirements

A successful and auditable migration process depends on rigorous pre-migration planning and precise system configuration on both the source and destination sides. This section outlines the key steps required to establish a reliable foundation for data migration and end-to-end reconciliation.

Reporting entity selection

As part of the initial setup, organizations must work closely with data migration experts to identify suitable reporting entities – logical groupings that align with how their data is organized. These entities may be based on:

- Geographic regions (e.g., Region-1, Region-2)
- Business units or departments
- Functional domains (e.g., Domain-X, Domain-Y)



Choosing the right reporting entity structure is essential, as it determines how reconciliation metrics are calculated, segmented, and presented on migration dashboards and reports. A well-defined reporting entity framework ensures clarity, traceability, and accurate reconciliation outcomes throughout the migration process.

Critical requirement: Message identifiers must be unique within each reporting entity. Duplicate IDs within the same scope can lead to data inconsistencies, skewed reconciliation metrics, and inaccurate reporting. Ensuring uniqueness at this level is foundational for clean, reliable reconciliation results.

Source data profiling

Source-side data profiling provides essential visibility into migration scope and highlights potential risks early in the process. This comprehensive analysis should produce:

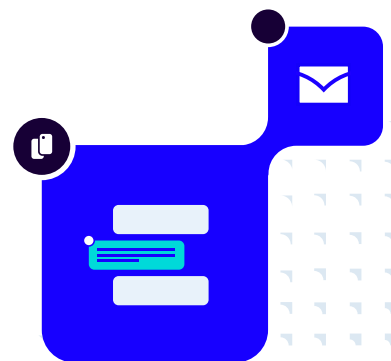
1. A detailed inventory of the data to be migrated, including:

- The total number of documents targeted for migration
- The number of logical data sets, each corresponding to a reporting entity
- A breakdown of message volumes for each value within the chosen reporting entity (e.g., Region-1: 10M messages, Region-2: 5M messages)

2. Identification of unmigratable message categories that may require special handling:

- Corrupted messages that are unreadable and may need to be excluded with appropriate documentation
- Non-standard messages, not gracefully stored in the source archive, that may raise regulatory concerns
- Large messages that may require special handling or approval for exclusion

This profiling establishes the baseline for reconciliation metrics and enables informed planning for exception handling, ensuring that all stakeholders have realistic expectations about migration completeness and potential challenges.





Reconciliation logic

Reconciliation is implemented as a continuous, automated process designed to verify, at scale, that each message ingested from the source has been successfully archived and indexed in the destination archive. Rather than being deferred until the end of the migration, reconciliation is performed incrementally throughout the migration lifecycle to improve speed, reliability, and operational visibility.

Incremental and event-driven reconciliation

The reconciliation engine leverages audit events captured during each stage of the migration pipeline. As messages are processed, events are recorded by the Event Logger — from the initial “migration initiated” event generated by the scrape system to the final “indexed” event emitted by the destination archive.

The reconciliation engine, built on top of an analytical query layer, periodically scans the event data to perform message-level comparisons. Using the source identifiers propagated throughout the pipeline, it matches each initiated event with a corresponding indexed event. This approach enables the system to:

- Detect gaps or incomplete message flows
- Identify mismatches or duplicate records
- Generate reconciliation reports that reflect the current state of progress
- Highlight failures with actionable metadata for investigation

Each run of the reconciliation process focuses on unreconciled messages, attempting to resolve them by finding their closure events (i.e., successful indexing) and updating their status accordingly. This incremental model ensures that reconciliation can scale with high-volume migrations and deliver near-real-time visibility into migration health.

Message-level tracking and duplicate detection

A core enabler of this system is message-level tracking through consistent use of source IDs. These identifiers allow the reconciliation engine to associate events across components and timeframes with high precision.

To ensure accuracy at scale, the reconciliation logic incorporates sophisticated duplication detection mechanisms. Messages with duplicate source IDs — whether due to replays, retries, or upstream anomalies — are detected and excluded from reconciliation counts to avoid skewing dashboard metrics or final reports. This approach ensures that reconciliation statistics accurately reflect the true state of migration progress, even in complex scenarios involving system recoveries or error handling procedures.

Reverse reconciliation

In addition to forward reconciliation (tracking messages from source to destination), the system performs reverse reconciliation. This is a critical verification step that starts with indexed events in the destination archive and verifies that each has a corresponding initiated event from the source.

As data is transferred from legacy systems at scale, data discrepancies are inevitable. Our reconciliation process takes this into account and provides bi-directional validation provides a deeper layer of assurance by confirming that every message present in the destination archive originated through the expected migration workflow. It is especially valuable for detecting anomalies such as:

- Missing events from the Digital Safe System Export
- Failures in event publication due to transient errors (e.g., when the Digital Safe System Export application successfully sends a message to the SMTP endpoint but goes down before publishing the event to the Event Logger)
- Unauthorized or unexpected data introduction

When such discrepancies are identified, the system can trigger a backfill of missing events from the source. Reconciliation is then re-attempted, ensuring that all indexed messages have a verifiable origin and that the entire migration is accurately accounted for in the audit trail.



Migrate to a modern archive with confidence

Data migration in regulated industries is not just a technical task; it is a regulatory obligation and a high-stakes process where precision, transparency, and accountability are non-negotiable requirements. Organizations must not only move data but also prove its integrity and demonstrate full lifecycle traceability across systems to preserve its compliance posture.

End-to-end reconciliation fulfills this need by enhancing migration from a leap of faith into a structured, provable operation. It enables organizations to retire older systems with confidence while ensuring regulatory obligations are maintained without disruption — a critical capability in today's increasingly regulated business environment.

This approach delivers multiple benefits:



Operational confidence

Issues are detected early and resolved proactively, reducing risk and rework. Migration teams can respond swiftly through efficient triage and improved traceability tools, maintaining project momentum while ensuring quality outcomes.



Audit readiness

Every message has a traceable path, supported by immutable event records and formal reports. The comprehensive audit trail satisfies the most stringent regulatory requirements, demonstrating due diligence throughout the migration process.



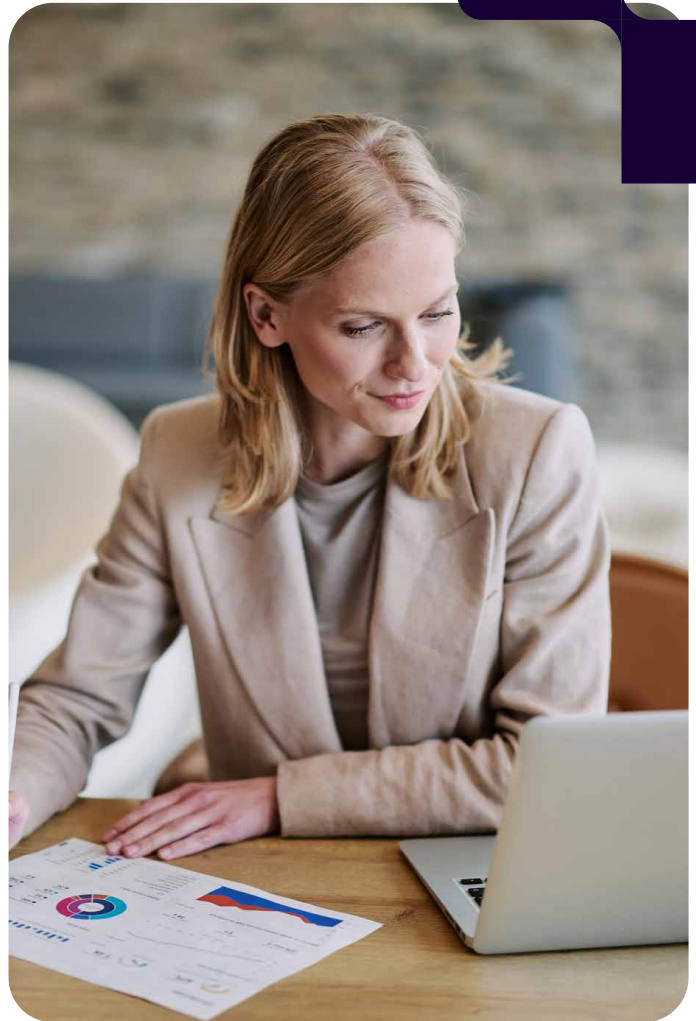
Customer trust

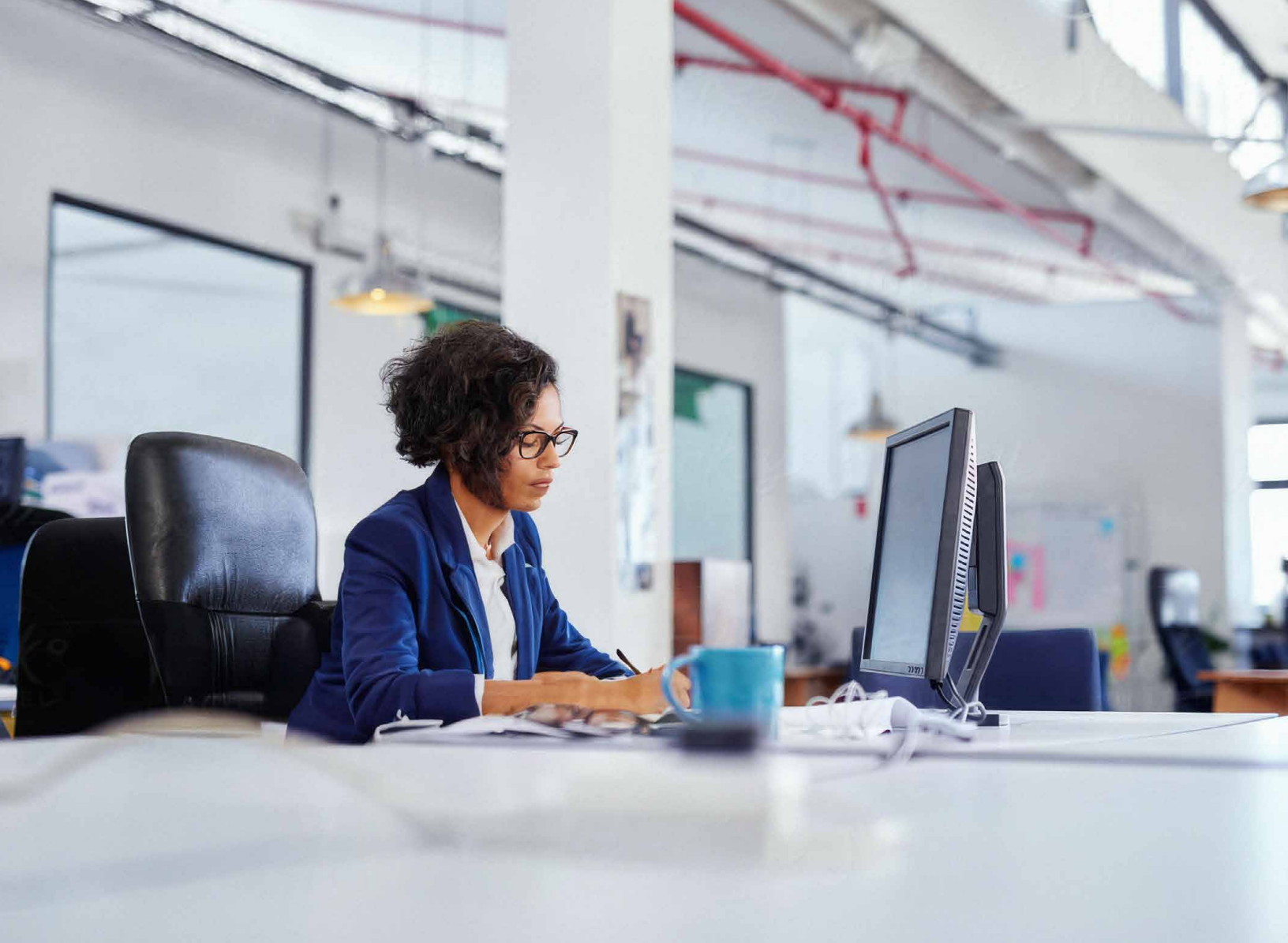
Transparency and verifiability enable organizations to retire older systems with confidence. Every message has been accurately, completely, and verifiably migrated — meeting the strictest compliance standards without disruption to business operations, thereby enabling organizations to decommission traditional platforms with peace of mind.

As data migrations become increasingly complex and regulatory requirements continue to evolve, robust end-to-end reconciliation is no longer optional — it is essential. The volume of data requiring migration continues to grow exponentially, while regulatory scrutiny of data handling practices intensifies across all jurisdictions. Organizations can no longer rely on basic verification methods or manual processes when dealing with terabytes or petabytes of sensitive, compliance-bound information.

Smarsh DS2EA end-to-end reconciliation provides the scalability, reliability, and transparency required to meet these challenges while maintaining the trust of customers, regulators, and stakeholders. The ability to retire older systems with complete confidence, backed by comprehensive verification of successful migration, enables organizations to modernize their technology infrastructure while maintaining uninterrupted compliance with regulatory obligations.

Contact us to learn how Smarsh Data Lifecycle Management can help you better satisfy your compliance and data governance needs while ensuring successful, verifiable migrations of your critical regulated data.





Smarsh® enables companies to transform oversight into foresight by surfacing business-critical signals from the most in-demand 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

Smarsh provides marketing materials for informational purposes only. Smarsh does not provide legal advice or opinions.
You must consult your attorney regarding your compliance with applicable laws and regulations.

Guide - 10/25



1-866-762-7741



www.smarsh.com



@SmarshInc



SmarshInc



Company/smarsh