

Accelerating Document Ingestion at Scale

A Technical Collaboration between Micro Strategies, Crawford Technologies and Box



Executive Summary

Technical decision makers in regulated industries face persistent challenges with legacy mainframe systems for high-volume document management. This white paper presents a cloud-native solution that integrates Box's Hyper-scaling APIs, Crawford Technologies' document processing tools, and Micro Strategies' multi-threaded import automation.

The architecture enables rapid, compliant ingestion of millions of documents—demonstrated by migrating **1 million documents in just 1.5 hours.**

This approach streamlines migration, reduces operational costs, and delivers scalable, repeatable performance to meet stringent regulatory and business requirements.

Key Takeaways:

- Replace legacy mainframe ECM systems with a modern, cloud-native solution.
- Achieve rapid, compliant migration—1 million documents in 1.5 hours.
- Lower operational costs and improve business agility.
- Meet regulatory requirements for document access and retention.

Abstract

This white paper is designed for technical decision makers and influencers tasked with modernizing enterprise document management at scale. Historically, organizations have been locked into costly, monolithic legacy mainframe systems for high-volume document processing, with few viable alternatives. This paper presents a SaaS-based cloud solution—integrating Box's Hyper-scaling APIs, Crawford Technologies' PRO Product Family, and Micro Strategies' import automation—that enables rapid, compliant ingestion of millions of documents. The solution delivers measurable outcomes: migration times reduced from months or years to hours, operational costs lowered by eliminating legacy hardware, and improved agility through scalable, cloud-native architecture.

Problem Statement

Organizations in regulated industries face persistent challenges with legacy document management systems. Before exploring the technical solution, it's important to understand the core obstacles that have prevented modernization and why these issues matter for IT leaders. As organizations pursue the modernization of document management workflows, several common challenges emerge:

Table of Contents

Abstract.....	1
Problem Statement	1
Solution Background.....	2
Why not Just Migrate the Batch Jobs off the Mainframe?	2
Mainframe Migration Hurdles	2
Mainframe Modernization—The New Trend	3
Third Party Data	3
Understanding the Documents.....	3
Where do These Documents get Used?	3
How are These Documents Generated?	4
When do Documents Need to be Available?.....	5
About the Solution Providers.....	5
Box, Inc.	6
Crawford Technologies	6
Micro Strategies Inc.	6
Ingestion at Scale	7
PRO Designer.....	7
PRO Transform	8
Micro Strategies Box Import Solution	9
Putting it all Together	10
Filling in the Gaps.....	11
What's Next for This Solution	12
Conclusion	13
Bibliography.....	13



Mainframe-generated compound documents remain prevalent across regulated industries, creating operational complexity for IT teams tasked with downstream processing and compliance.



While some legacy enterprise content management (ECM) systems support document bursting, they tend to be costly, difficult to maintain, and lack modern capabilities such as scalable automation, seamless integration, and AI-driven document management.



Regulatory requirements such as SEC Rule 17a-4, HIPAA, SOX, GDPR, and FCA—demand rapid access, long-term retention, and auditable management of documents, making traditional approaches increasingly unsustainable.



Traditional document bursting and migration approaches are constrained by performance bottlenecks, API rate limits, and manual workflows, making it challenging to meet regulatory requirements for rapid access and retention.



Technical decision makers need a scalable, flexible and compliant solution to efficiently ingest, process, and manage millions of documents from diverse sources—including mainframe outputs—into modern cloud platforms.

Solution Background

With the challenges clearly defined, this section provides context on the evolution of document management technologies and the limitations of traditional approaches. Understanding this background sets the stage for evaluating the new solution's architecture and benefits.

An understanding of the types of documents and where and how they are used within the different industries is paramount in being able to provide a solution that is flexible enough to meet the needs of this diverse crowd. Understanding what each of the parties brings to the table is equally important, as these pieces will be the foundation of the solution. However, there is one question that needs to be considered first.

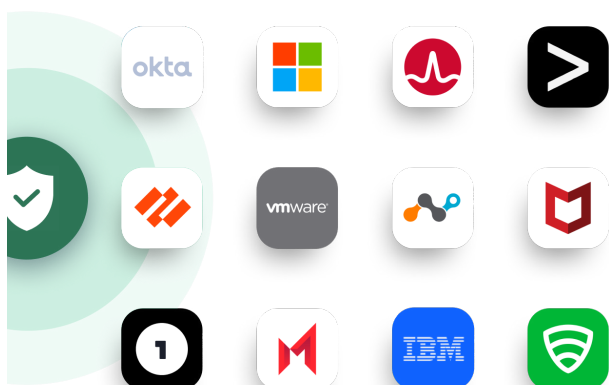
Why not Just Migrate the Batch Jobs off the Mainframe?

At first glance, this seems like a great question. However, further investigation reveals a complex and nuanced set of circumstances. Examining these circumstances in a little more detail reveals why mainframes are not only still in use but are here to stay.

Mainframe Migration Hurdles

There are many hurdles to migrating applications off the mainframe. Some of the significant ones are:

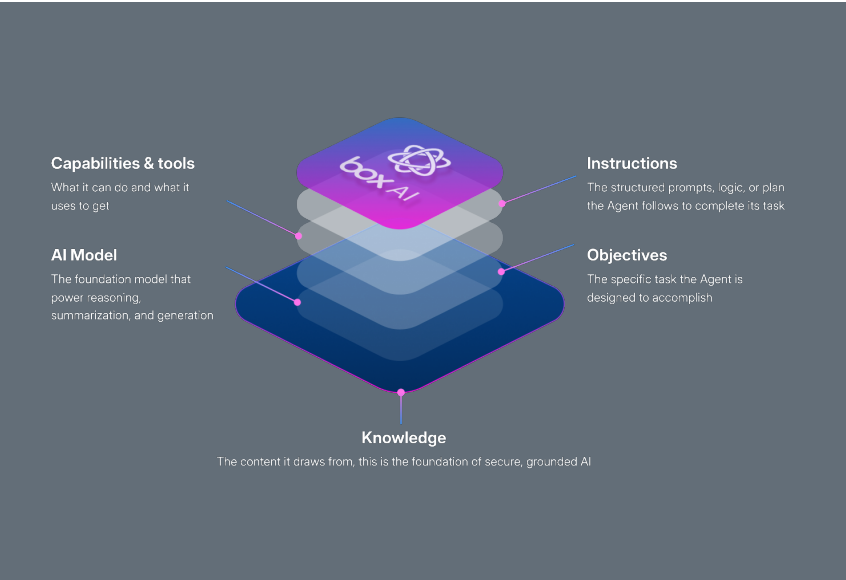
1. Legacy codebases are complex and often written in legacy languages such as COBOL, or Fortran (IBM, 2017).
2. There is a growing shortage of professionals in mainframe technologies (Shetti, 2025).
3. Mainframes support mission-critical workloads where downtime can be catastrophic to an organization (IBM, 2017).
4. The cost and time to migrate these applications can exceed the benefits (Ellis, Nelson, Alexander, & Hartig, 2024).
5. The data within these mission-critical systems is often sensitive. Given the high volume of this data and its sensitivity, securely and accurately migrating this data while maintaining compliance is a daunting task (IBM, 2017).



Understanding the Documents

This paper has mentioned compound documents numerous times. This section provides a deeper look into their structure and relevance across industries. These documents are common in sectors with long-standing reliance on legacy technologies. In many cases, the same software has been in use for decades—simply because it continues to perform reliably.

The mindset of “if it ain’t broke, don’t fix it” still resonates, especially when these systems consistently produce the complex, multi-part reports that organizations depend on.



Where do These Documents get Used?

The table below provides some examples of different types of reports that are generated for different industries.

Industry ¹	Report Type	Purpose
Banking	Daily Transaction Report	Summarizes all customer transactions for the day.
Government	Tax Filing Batch Report	Aggregates citizen or business tax data.
Healthcare	Encounter and Billing Reports	Tracks patient visits and billing codes.
Insurance	Claims Processing Summary	Lists all claims processed in a batch cycle.
Retail	Nightly Sales Report	Consolidates POS data across stores.

¹The term industry is used loosely in this paper. Multiple classification systems exist for defining industries and sectors, and this paper does not attempt to endorse or justify any particular framework.

Mainframe Modernization— The New Trend

There is another significant reason mainframes haven’t gone away. Sometimes they are the best tool for the job. The idea that mainframes are legacy is outdated.

Modern mainframes keep pace with industry and offer things such as AI-driven automation and hybrid cloud capabilities (Sharma, 2025).

Even older mainframes can outperform cloud platforms in several mission-critical scenarios. They can efficiently manage resources across the sysplex, enabling mainframes to process terabytes of data in a single operation without fragmenting it into smaller chunks.

This allows for the generation of comprehensive reports that span vast datasets.

A sysplex is a cluster of IBM mainframes acting together as a single system image.



Third Party Data

Organizations often lack control over how external data is delivered. When data originates from a third party operating on a mainframe, it frequently arrives as a large PDF or text-based report containing numerous embedded documents.

As a result, even organizations without mainframes may still need to accommodate mainframe-generated content within their workflows.

Industry ¹	Report Type	Purpose	The common theme across these examples is that they produce comprehensive reports that contain a staggering amount of data.
Telecom	Usage and Billing Summary	Details call/data usage and charges.	For example, imagine the daily transaction report for a major multi-national bank with thousands of branches or the claims processing summary from an insurance company shortly after a major natural disaster.

How are These Documents Generated?

As previously mentioned, these documents are generated by mainframe applications. The following table provides some of these applications and their primary functions.

Industry	Tool	Vendor	Primary Functions
Banking	BMC Control-M	BMC	Cross-platform job orchestration, SLA monitoring, and audit-ready reporting
	IBM Workload Scheduler for z/OS	IBM	Automates batch jobs for EOD processing, compliance, and transaction reporting
Government	CA Scheduler Job Management	Broadcom	Legacy batch job management in public sector environments
	IBM Z Batch Runtime	IBM	COBOL/PL/I batch job execution for tax, census, and benefits processing
Healthcare	IBM Z OMEGAMON	IBM	Monitoring and reporting on batch job performance in EHR systems
	Mainframe Application Tuner (MAT)	Broadcom	Performance tuning and batch reporting for healthcare systems
Insurance	CA 7 Workload Automation	Broadcom	Legacy batch job scheduling for COBOL/JCL-based systems
	IBM Tivoli Workload Scheduler (TWS)	IBM	Claims processing, policy updates, actuarial reporting
Telecom	BMC Control-M	BMC	High-volume batch processing for customer statements and usage aggregation
	IBM Tivoli Workload Scheduler (TWS)	IBM	Claims processing, policy updates, actuarial reporting

When do Documents Need to be Available?

The answer to this question depends on the types of documents, the regulations in place, and an organization's commitments. Regulations are the largest driver, as there are repercussions, often financial, if regulations aren't met. The table below shows some regulations along with their timeline constraints.

Regulation	Region	Required Timeframe
SEC Rule 17a-4	United States	Promptly furnished for immediate access for 2 years, retention for 6 years (U.S. Securities and Exchange Commission, 2023).
HIPAA Privacy Rule	United States	Access within 30 days of request. Entities are encouraged to provide access as soon as possible (U.S. Department of Health and Human Services, 2016).
Sarbanes-Oxley Act (SOX)	United States	Retention for 7 years, timely access for audits (U.S. Securities and Exchange Commission, 2002).
GDPR (Article 12)	European Union	Access without undue delay and within one month (GPDR, 2016).
FCA Recordkeeping	United Kingdom	Retention for 5 years, access on demand (Financial Conduct Authority, 2018).

It is important to note that many regulatory agencies place greater emphasis on defining document retention periods than on specifying exact timeframes for document availability. When availability is addressed, it is often described using terms like promptly, immediately, without delay, or as soon as possible—phrases that lack precise definitions. As a result, compliance officers often advocate for rapid access to documents, understanding that when a regulator requests a record, it must be readily available (Sims, 2025)².

This ambiguity results in a desire for quick turnaround of these large documents. Many organizations will say the documents have to be ready within 24 hours.

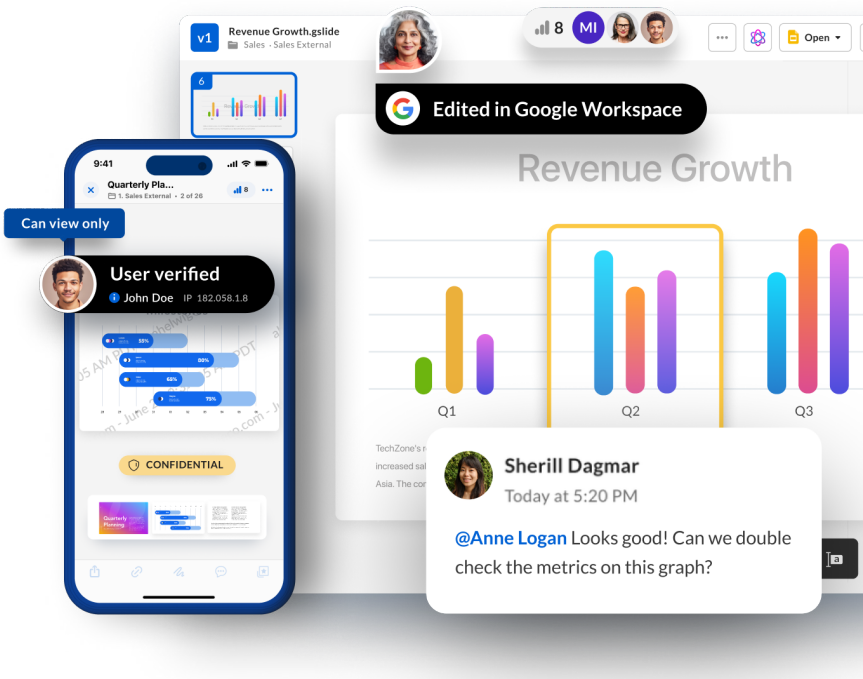
About the Solution Providers

Before diving into technical architecture, it's important to understand the key components and partners that make this solution possible.

This section introduces the core technologies and the unique value each partner brings to the integrated solution, setting the stage for how these elements work together to address the previously outlined challenges.

There are three collaborators on this solution, Box, Inc., Crawford Technologies, and Micro Strategies Inc. The following identifies the contributions to this solution by each collaborator.

²Some regulations will distinguish between electronic and physical media. The information provided is related to electronic media.



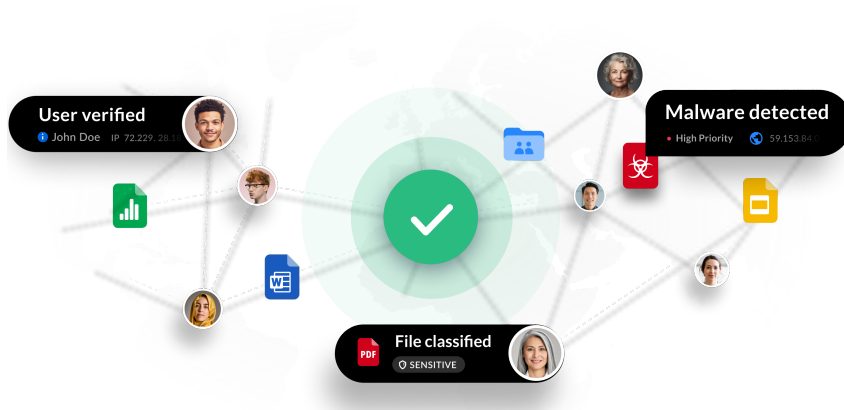
Crawford Technologies

Crawford Technologies is a global leader in enterprise document management, specializing in software solutions that optimize the delivery, storage, and accessibility of customer communications. With a strong presence across six continents and a client base that includes major banks, insurers, healthcare providers, utilities, and print service companies, Crawford Technologies empowers organizations to modernize without overhauling their existing systems. Their technologies enable seamless integration between legacy applications and modern platforms, allowing businesses to repurpose existing documents and data in innovative ways. As a strategic collaborator in this solution, Crawford Technologies brings deep expertise in bridging the gap between traditional infrastructure and future-ready communication workflows (Crawford Technologies, 2025).

Key components leveraged in this solution are **PRO Designer** and **PRO Transform**, both part of Crawford Technologies' PRO Product Family. PRO Designer is a powerful graphical interface that streamlines the configuration of processing rules for fully composed files. It enables organizations to efficiently define parameters for generating both print and digital outputs, without requiring extensive programming expertise. With its intuitive design and real-time output validation, PRO Designer significantly reduces development and approval cycles—accelerating application setup and client onboarding while lowering operational costs. Complementing this, PRO Transform provides robust capabilities for bursting composite documents into individual files, enabling high-volume document separation and preparation for downstream ingestion processes.

Micro Strategies Inc.

Micro Strategies Inc. is a strategic technology solutions provider dedicated to simplifying the complexity of modern IT environments. With a mission to empower organizations with outcome-driven solutions, Micro Strategies aligns advanced technologies with long-term business objectives to deliver measurable impact. Known for its customer-centric approach, deep industry expertise, and commitment to innovation and operational excellence, the company transforms complex systems into accessible, actionable strategies that drive scalable growth and strategic advantage. Micro Strategies' Box Import solution is a flexible, multi-threaded ingestion engine used in production environments to automate document migration and metadata mapping at scale.



Box, Inc.

Box is the Intelligent Content Cloud—a single, secure platform to manage the full lifecycle of an organization's high-value content. Beyond core ECM, Box provides advanced security, governance, workflow automation, and content intelligence that modernizes how regulated archives are stored, found, and used.

Security, governance & compliance at ingest.

With Box Governance, customers can configure classification-driven policies that apply at ingestion: granular retention/retention-lock, legal holds, and defensible disposition, plus complete audit trails. Combined with Box Shield's classification and access controls, this helps organizations meet stringent record-keeping obligations (e.g., SEC/FINRA-style retention and WORM compliance requirements) while consolidating legacy archives into a single, governable repository.

From static archives to usable data with Box AI.

Box's content intelligence, Box AI, can automatically classify documents and extract structured metadata (e.g., customer ID, policy number, statement date) to enrich files on upload. That metadata turns previously opaque archives into queryable datasets and can trigger downstream automation with Box Relay/Automate—enabling faster search, exception handling, and unlocks archived content for active business workflows, such as real-time customer service queries or predictive analytics.

High-speed ingestion with Hyper-scaling.

To make large-scale ECM migrations feasible, Box introduced Hyper-scaling—elevated API throughput for approved workloads. In this joint validation with Crawford Technologies and Micro Strategies, the team ingested 500k+ documents and demonstrated an architecture capable of scaling to millions. In benchmarking, bursting plus ingestion of 1,000,000 documents completed in roughly ~1 hour 24 minutes, a step-change from the multi-day timelines typical of standard rate limits.

Why it matters.

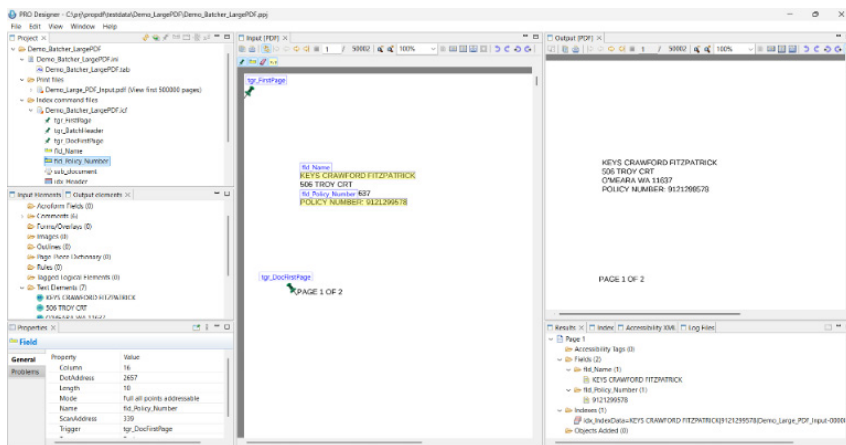
The combination of -
(1) governable, secure cloud storage,
(2) classification-driven retention and holds,
(3) AI-assisted metadata enrichment, and
(4) Hyper-scaled ingestion - allows organizations to retire costly legacy ECM, meet compliance objectives, and put archived content back to work in business workflows.

PRO Designer

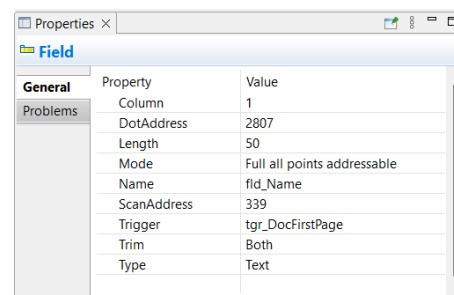
PRO Designer is a GUI application that allows users to define the configuration files for the different Crawford Technologies products. This application was used to create two configurations, the first (shown below) was a small document that was used for both initial testing and for benchmark testing and a larger document that contained over 500,000 sub-documents.

⁴The information already provided for Box's hyper-scaling solution are sufficient to meet the needs of this section and therefore no additional details are provided.

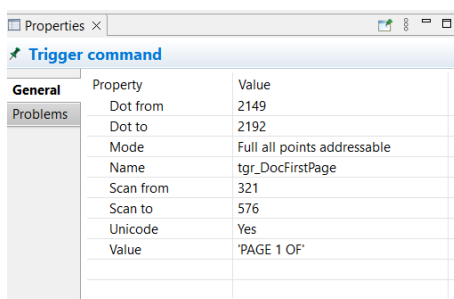
⁵Small is relative. This small document has over 50,000 pages encapsulating almost 14,000 sub-documents.



Note the different highlighting in the document view in the center panel. This represents different information captured in or about the page being shown. The first is the name field, which is found by its physical location on the page.



The name field has a trigger on `trg_DocFirstPage`. This trigger is defined similarly to the field except it has a from position and a to position and it expects the text to have a value of `PAGE 1 OF`.



⁶The terms *ScanAddress* and *DotAddress* refer to legacy print technologies, specifically Advanced Function Presentation (AFP). *Scan* indicates location scanning from left to right and *dot* indicates the number of dots up from the bottom of the page. A dot in this context refers directly to dot-based printing such as dot matrix and laser printers.

Ingestion at Scale

With the strengths of Box, Crawford Technologies, and Micro Strategies established, the next step is to see how these components work together in practice. The following section details how the integrated solution orchestrates high-speed, compliant document ingestion—demonstrating the technical workflow and performance that sets this approach apart from legacy systems. The following diagram shows the components that are used in this solution by the partner providing the components.



GUI tool
PRO Designer



application
PRO PDF



database
MongoDB

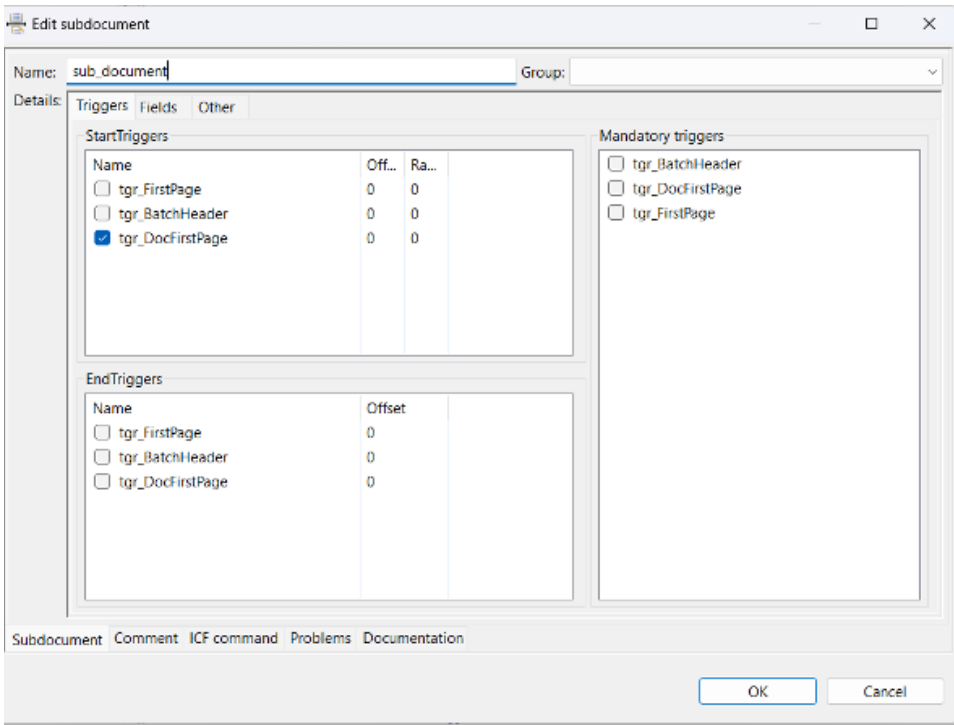


application
Box Import



API
Box REST API
with Hyper-scaling

Although each of these components have been established previously in this paper, additional details are provided in the subsections below. This section will take these components and put them together into a cohesive solution to ingest documents at a rate that meets regulatory and compliance requirements.



Subdocument commands indicate that a new document has been found.

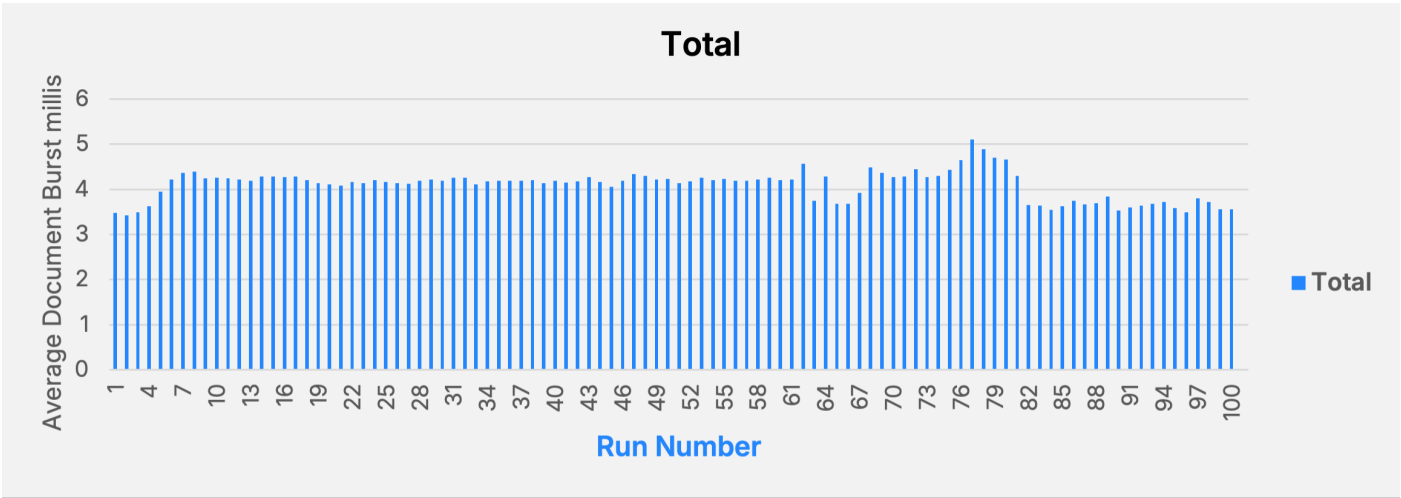
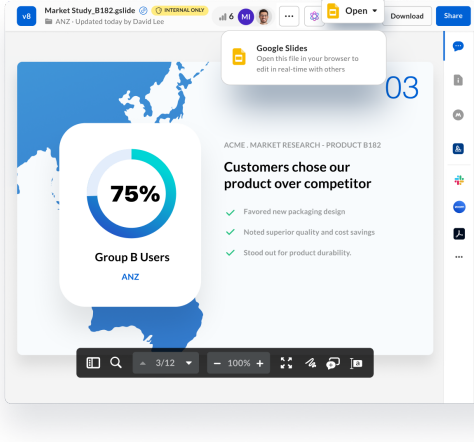
These examples highlight just a fraction of the power and flexibility that PRO Designer offers when defining configurations.

Its capabilities extend far beyond what's immediately visible, providing a robust and adaptable framework for managing complex document processing tasks with precision and efficiency.

PRO Transform

PRO Transform is an executable that uses an INI file generated by PRO Designer to perform the tasks designed there-in. In this case it uses the INI file to burst the larger PDF document into smaller documents in batches of 1,000 documents. Although the example case is simple, the speed at which documents burst is quite fast.

The chart below conveys that the average time to burst a document is between 3 and 6 milliseconds, with an average of 4.1 milliseconds across 100 runs⁷.



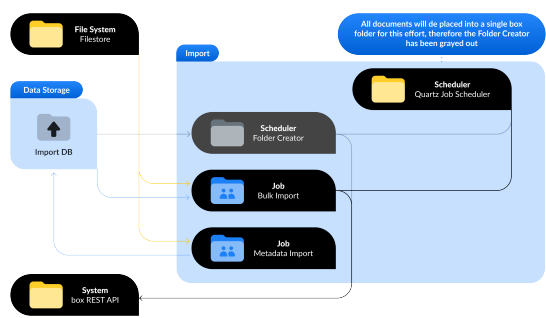
This means that the solution consistently ingests one million documents in just over one hour, meeting regulatory timelines and operational SLAs.

Micro Strategies Box Import Solution

The Box Import solution is part of a larger migration offering. This offering was built with the intention of being agnostic of either the source or the target of the documents.

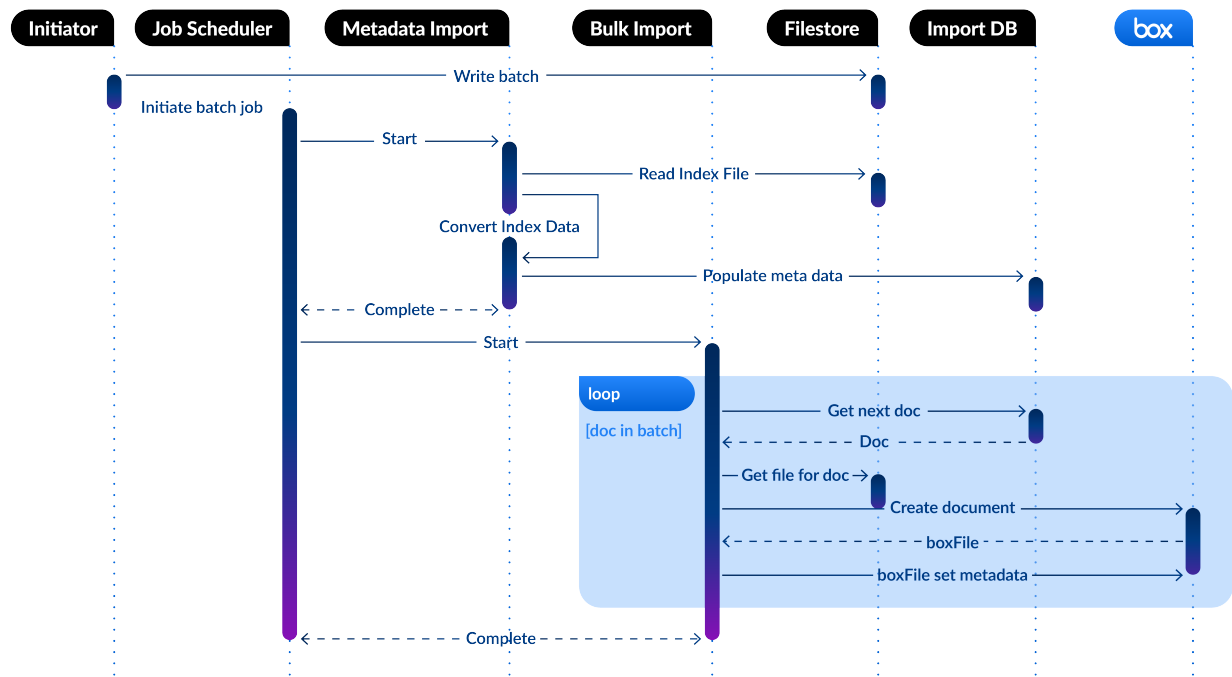
There have even been situations where multiple sources or multiple targets have been required. For example, in one situation all documents went into Box except for certain HR documents which went into Oracle's PeopleSoft.

Another situation had documents coming from OpenText InfoArchive and a database and file store and then going into Box. Due to this flexible nature the Box import could also be used independent of the migration offering and provides an excellent mechanism for high-speed multi-threaded document imports into Box. The following diagram identifies the Box Import components used within this solution.



⁷ Details can be provided upon request.

The following sequence diagram depicts the high-level flow between the components identified above:



Each job within the solution is implemented using **Spring Batch**. “Spring Batch is a lightweight, comprehensive batch framework designed to enable the development of robust batch applications that are vital for the daily operations of enterprise systems.” (Broadcom, 2025).

Spring Batch is built upon the foundational **Spring Framework**, a mature and extensible platform widely adopted in enterprise application development. Together, these frameworks provide a robust infrastructure that enables the Import solution to deliver scalable, resilient, and maintainable batch processing capabilities. Key features include:

- **Native multi-threading support**, facilitating efficient parallel execution of tasks and improved throughput for large-scale data operations.
- **Advanced retry mechanisms**, ensuring fault tolerance and operational continuity:
 - Automatic retries for transient failures, such as Box SDK connection issues, with configurable backoff strategies.
 - Flexible, application-specific retry logic enabled by Spring's declarative and programmatic support.
 - Persistent failure tracking via database logging, allowing for targeted remediation of unrecoverable errors.
- **Comprehensive logging**, utilizing standard Java utilities to capture execution details, exceptions, and performance metrics for auditability and diagnostics.
- **Monitoring and observability**, achieved through both native Java instrumentation and custom logging strategies:
 - Integration with enterprise monitoring systems via syslog error logging, as exemplified by Micro Strategies' managed environments.
 - Support for external tools and dashboards to visualize job execution, performance trends, and failure patterns.
- **Job orchestration and lifecycle management**, enabling precise control over job execution flows:
 - Modular configuration of job steps, transitions, and conditional logic.
 - Support for job scheduling, partitioning, and remote execution.
 - Built-in mechanisms for job restarting and state persistence.
- **Transactional integrity**, ensuring consistency and reliability across batch operations through Spring's declarative transaction management.

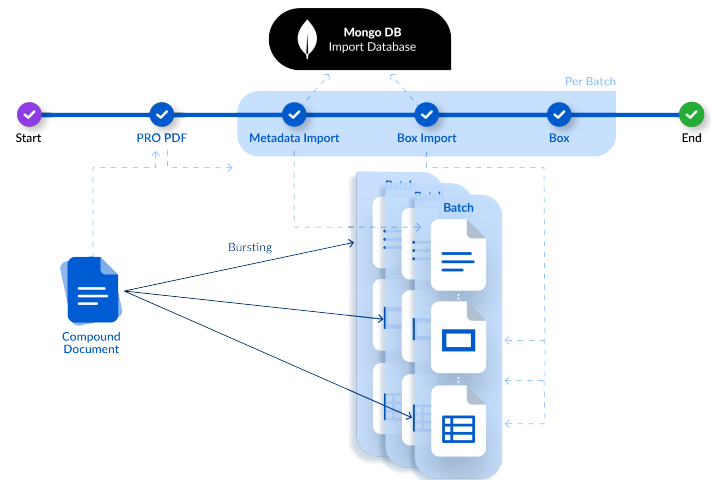
The batch jobs and their supporting framework are implemented as a **Spring Boot** application. "Spring Boot makes it easy to create stand-alone, production-grade Spring based Applications that you can 'just run'." (Broadcom, 2025). Spring Boot's streamlined packaging and deployment model enables consistent and reliable execution across a diverse range of infrastructure environments. These include:

- Physical Servers
 - Bare-metal deployments within data centers or on-premises facilities.
- Virtual Machines
 - On-premises hypervisors such as VMware and Hyper-V.
 - Cloud-hosted virtual machines, including Azure Virtual Machines, AWS EC2, and Google Compute Engine.
- Containerized Platforms
 - Standard container runtimes such as Docker.
 - Orchestrated environments using Kubernetes.
 - Serverless container services, including:
 - » AWS Fargate
 - » Google Cloud Run
 - » Azure Container Appss

This flexibility ensures that the Import solution can be deployed in alignment with enterprise infrastructure strategies, whether traditional, cloud-native, or hybrid.

Putting it all Together

With the solution architecture in place, it is essential to validate its effectiveness through real-world testing and performance measurement. This section presents the key results and benchmarks achieved during the implementation, demonstrating how the integrated approach delivers tangible improvements over legacy systems. With all the components laid out and understood the solution can be put together. The following diagram depicts both the program and data flow between the various applications within the solution. The solution begins with a compound document originating from one of the previously discussed sources. In this example, the document has already been processed using PRO Designer, and a corresponding INI file is available. A template INI file is updated with the document reference as soon as the document becomes available. PRO is then executed using the INI file to burst the compound document into subdocuments and generate an associated index file.



This is performed in batches of **1,000 subdocuments**, with each batch producing its own index file.

$$4 + (1 * 1000) = 1004$$

$$1004 \div 60 = 16.\bar{7}3 \sim 17$$

$$\frac{4 \times 1000}{60} = 66.\bar{6} \sim 67$$

$$67 + 17 = 84 = 1 \text{ hour } 24 \text{ minutes}$$

This means that the ingestion will lag the bursting by 4 seconds plus 1 second for each batch. Taking the example of one million documents, this would be one thousand batches of one thousand documents.

Therefore, ingestion will have lagged a total of
1,004 seconds or just under 17 minutes.

This added to the one hour seven minutes that PRO Transform will require to burst the documents means that one million documents will be ingested within an hour and a half (roughly 1 hour and 24 minutes)¹⁰.

¹⁰Multiple timing tests were performed, and an average was taken. These timing tests were performed on a Lenovo ThinkPad P16 Gen 2 with 128 GB RAM, Intel i7-14700HK processor, 2 TB RAID 0 storage, and network bandwidth of 80 Mbps upload/120 Mbps download. Results may vary.

Filling in the Gaps

There are things inherent in the solution that are worth noting. This section outlines some of them.

Metadata Mapping

There are two places where metadata mapping is performed within the solution. The first is when compound documents are split. It is at this time that metadata is extracted from the document and mapped to properties.

This is defined utilizing PRO Designer and performed using PRO Transform. These properties are uploaded directly into Mongo as the source for the import into Box.

The Import solution has a mapping set up program that takes a specifically formatted Excel workbook that defines the mappings of the source to Box. This program takes the workbook and creates the mapping information in the database that will be used when the import is executed.

Upon completion, each batch is queued for downstream processing. In this example, a job is configured to handle the import of these batches. The first step in this job imports metadata from the index file into a MongoDB database, which serves as the metadata store⁸. This step is highly efficient, typically completing in seconds for a batch of 1,000 documents.

Once metadata is available, the ingestion process begins⁹. This step dynamically allocates threads to leverage Box's hyper-scaling capabilities. The number of threads is determined by various factors and may be tuned over time for optimal throughput. During this phase, documents are created in Box and enriched with the metadata extracted from the index file.

The process incorporates intentional concurrency between batches to minimize total execution time. For example, while PRO Transform generates the second batch, the Box import program may concurrently import metadata and ingest documents from the first. The following outlines approximate durations for each major step, based on a batch size of 1,000 documents.

Bursting of a batch

4 seconds



Import of Index metadata

1 seconds



Ingesting of documents into Box

4 seconds



⁸The migration utility relies on a shared data format between export and import, so index files must conform to the expected structure.

⁹Folder creation is omitted, as folder taxonomy is not typically a requirement.

Security

Box has a robust security model that works on the principle of inheritance. Generally, solutions involving the splitting of compound documents work well having the parent folder define the security for any given compound document. Situations that require a more complex security model will often also require that folder creation be performed. It is within the folder creation that custom security can be applied.

Flexible Integration Options with Spring Batch and Box Platform Features

One of the key advantages of using the **Spring Batch** framework is its extensibility—custom processing logic can be seamlessly injected into the batch workflow. This flexibility enables tailored solutions that meet specific business requirements. Additionally, **Box’s webhook and event streaming capabilities** offer alternative mechanisms for integrating custom functionality, each with distinct characteristics suited to different scenarios.

When evaluating which approach to use, consider the following guidance:

Spring Batch Integration

- Executes logic for **every document** processed.
 - » Conditional logic can be applied to **skip documents** as needed.
- Can be configured to run **before or after** a document is uploaded to Box.
- Must be implemented as part of the **import process**:
 - » Development is limited to **Micro Strategies staff**.
- Requires use of **JVM-compatible languages** (e.g., Java, Kotlin).

Box Webhooks

- Triggered by **events** (e.g., file creation).
- Requires the file to **already exist in Box**.
- Provides **limited event detail**:
 - » For example, a “metadata instance updated” event signals a change but does not specify which template or what was modified.
- Allows **granular folder-level control**, which can increase maintenance complexity.
- Infrastructure is **simple to implement** (e.g., REST endpoint).
 - » Advanced setups can include **queue publishing** or **database logging** for deferred processing.
- **Event loss is possible** during infrastructure outages.
 - » Box retries delivery, but prolonged downtime (e.g., data center failure) may result in missed events.
- Supports **flexible deployment** options:
 - » Ideal for lightweight implementations using **AWS Lambda, Azure Functions, or Google Cloud Functions**.

Box Event Streams

- Also, event-driven and requires the file to **exist in Box**.
- Provides rich context for each event.
- **Event order is not guaranteed**:
 - » Events are delivered as they become available, which may result in out-of-sequence processing.
- Duplicate events may occur:
 - » De-duplication must be handled by the event consumer.
- Uses a **stream position identifier**:
 - » Enables resilience by allowing consumers to resume from the last known position.
- Supports **event filtering** to focus on relevant triggers.

What’s Next for This Solution

This solution, which continues to evolve, proves that large, compound documents can be bursted and ingested into Box at a rate that meets both compliance and regulatory requirements.

While portions of the process outlined here were initially completed manually, future deployments will incorporate automation—using tools such as Crawford Technologies’ PRO Conductor. Furthermore, enhancements to the Box Import program will be made to support the real-time nature of this solution, which differs significantly from traditional migration efforts.

Conclusion

For years, organizations have been stuck with legacy mainframe ECM systems due to a lack of modern, scalable alternatives. This integrated solution—combining Box's Hyper-scaling, Crawford Technologies' document bursting, and Micro Strategies' flexible import automation solution—replaces monolithic legacy systems with a modern, SaaS-based architecture that delivers speed, scalability and compliance. Technical teams can now achieve:



Migration of millions of documents in under two hours, compared to weeks or months with legacy approaches.



Significant reduction in operational and maintenance costs by moving to a cloud-native, SaaS platform.



Rapid, compliant access to documents to meet regulatory requirements and business SLAs.



Scalable, extensible architecture that supports future growth and integration with other enterprise systems.

This solution empowers technical decision makers to break free from legacy constraints and deliver measurable improvements in efficiency, compliance, and business agility.

Ready to modernize your ability to generate, manage and search high-volume client communications?

We invite you to schedule a High-Volume Ingestion Workshop led by Micro Strategies experts. We will review your document types (client statements, explanation of benefits, and other critical client communications) and demonstrate how the Box + Micro Strategies + Crawford architecture can replace expensive legacy system

[Schedule Your Workshop Now](#)

Bibliography

- Box. (2025, 8 3). AI-powered Box products & features. Retrieved from Box: <https://www.box.com/products-and-features>
- Broadcom. (2025, 10 16). Spring Batch Introduction. Retrieved from Spring: <https://docs.spring.io/spring-batch/reference/spring-batch-intro.html> Broadcom.
- (2025, 10 16). Spring Boot. Retrieved from Spring: <https://spring.io/projects/spring-boot>
- Crawford Technologies. (2025, 8 3). About Us. Retrieved from Crawford Technologies: <https://crawfordtech.com/about-us/>
- Deloitte. (2022, 12 6). Connect and extend: Mainframe modernization hits its stride. Retrieved from Deloitte: <https://www.deloitte.com/us/en/insights/topics/technology-management/tech-trends/2023/future-mainframe-technology-latest-trends.html>
- Financial Conduct Authority. (2018). SYSC 9.1 General rules on record-keeping. Retrieved from FCA Handbook: <https://www.handbook.fca.org.uk/handbook/SYSC/9/1.html> GDPR. (2016).
- Art. 12 GDPR. Retrieved from General Data Protection Regulation: <https://gdpr-info.eu/art-12-gdpr/>
- Hennick, C. (2023, 6 12). Why Mainframes Remain Essential for Many Organizations. Retrieved from BizTech Magazine: <https://biztechmagazine.com/article/2023/06/why-mainframes-remain-essential-many-organizations>
- IBM. (2017). Application modernization on the mainframe. Retrieved from IBM Institute for Business Value: <https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/application-modernization-mainframe>
- IBM. (2023, 06 28). Mainframe workloads: Batch and online transaction processing. Retrieved from z/OS Basic Skills: <https://www.ibm.com/docs/en/zos-basic-skills?topic=today-mainframe-workloads-batch-online-transaction-processing>
- IBM. (2023, 06 28). Mainframes working after hours: Batch processing. Retrieved from z/OS Basic Skills: <https://www.ibm.com/docs/en/zos-basic-skills?topic=today-mainframe-workloads-batch-online-transaction-processing>
- Kiplangat, B., & Dudkiewicz, L. (2025, 3 31). 60+ SaaS Statistics and Trends [2025]. Retrieved from cropink: <https://cropink.com/saas-statistics>
- Ellis, B., Nelson, L., Alexander, L., & Hartig, K. (2024, 2 29). Tackle The Overwhelming Challenge Of Mainframe Modernization. Retrieved from Forrester: <https://www.forrester.com/report/tackle-the-overwhelming-challenge-of-mainframe-modernization/RES146395>
- Fenwick, N. (2013, 1 31). Capability-As-A-Service And What It Means For Technology Vendor Strategy. Retrieved from Forrester: https://www.forrester.com/blogs/13-01-31-capability_as_a_service_and_what_it_means_for_technology_vendor_strategy/
- Salminen, M. (2025, 6 26). SaaS statistics for 2025: Market trends, growth insights, and key metrics. Retrieved from Hostingr: <https://www.hostinger.com/tutorials/saas-statistics>
- Sharma, A. (2025, 5 15). Modernizing the mainframe with IBM z17: AI-driven automation and hybrid cloud integration. Retrieved from IBM: <https://www.ibm.com/new/product-blog/mainframe-modernization-z17>
- Shein, E. (2020, 2 3). SaaS, PaaS, IaaS: The differences between each and how to pick the right one. Retrieved from TechRepublic: <https://www.techrepublic.com/article/saas-paas-iaas-the-differences-between-each-and-how-to-pick-the-right-one/>
- Shetti, M. (2025, 6 5). There's a mainframe skills gap. Here are 3 ways to address it. Retrieved from CIO: <https://modernizationwithoutdisruption.cio.com/empower-your-workforce/theres-a-mainframe-skills-gap-here-are-3-ways-to-address-it/>
- Sims, K. (2025). SEC Rule 17a-3 & FINRA Records Retention Requirements Explained. Retrieved from Pagefreezer: <https://blog.pagefreezer.com/sec-finra-books-records-retention-requirements>
- U.S. Department of Health and Human Services. (2016). Individuals' Right under HIPAA to Access their Health Information 45 CFR § 164.524. Retrieved from U.S. Department of Health and Human Services: <https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/access/index.html>
- U.S. Securities and Exchange Commission. (2002). Sarbanes-Oxley Act of 2002. Retrieved from U.S. Securities and Exchange Commission: <https://www.govinfo.gov/content/pkg/COMPS-1883/pdf/COMPS-1883.pdf>
- U.S. Securities and Exchange Commission. (2023). 17 CFR Parts 200 and 240. Retrieved from U.S. Securities and Exchange Commission.