White Paper

Securing the Blind Spot: Unstructured Data Risk and Thales File Activity Monitoring

thalestct.com

THALES
Building a future we can all trust

THALES
Building a future we can all trust

Unstructured data—everything from Office documents to chat logs, Gen AI outputs, and medical images—now accounts for roughly 80 – 90 % of enterprise information assets. Analysts expect the global datasphere to exceed 175 ZB by 2025, making visibility and governance at petabyte scale a board level mandate. Yet security tooling and processes are still optimized for structured databases, leaving emails, file shares, and collaboration platforms largely unmonitored. The result is a rising wave of compliance violations, insider abuse, ransomware, and AI driven data leakage.

Thales File Activity Monitoring (FAM) closes the visibility gap—delivering unified discovery, classification, and real time monitoring across your entire data estate.

The Unstructured Data Landscape

Metric	Finding
Share of enterprise data that is unstructured	Approximately **80–90%** of enterprise data is unstructured. Edge Delta
Annual growth rate of unstructured data	Unstructured data is growing at a rate of **55% to 65% annually.** Rackspace Technology
Prevalence of AI-generated content	Experts estimate that by **2025, up to 90%** of online content could be AI-generated. Yahoo Finance
Visibility into unstructured data access	**53%** of security teams lack continuous and up-to-date visibility into data access and usage. Business Wire
Compliance concerns related to unstructured data	**59%** of organizations are very concerned about data privacy, and **47%** about regulatory compliance, in relation to unstructured data. Qlik + ETR Survey

Unstructured data represents the **fastest-growing and least-governed category of enterprise information.** According to industry research, as much as **90% of data within large organizations is unstructured**—a figure that continues to rise as businesses adopt generative AI tools, collaborative cloud platforms, and digital communications at scale.

Unlike structured data, which resides in databases and can be easily queried and governed, unstructured data includes content with no predefined format or schema. This includes:

- Business documents (PDFs, Word files, Excel sheets)
- Images and media files
- Emails and attachments
- Instant messaging/chat logs
- Code and design assets stored in Git, Jira, or Confluence
- Notes, contracts, scanned records, and AI-generated outputs

Where It Lives

What makes unstructured data particularly challenging is its distributed and fragmented nature. It resides not only in traditional file servers but across cloud object stores (e.g., Amazon S3, Azure Blob), SaaS platforms (e.g., Microsoft 365, Google Workspace), and unmanaged file shares like NFS and SMB drives. Popular repositories include:

- SharePoint Online and OneDrive
- Slack, Microsoft Teams, and Zoom chat archives
- Email servers and Gmail inboxes
- Git repositories, product design portals, customer collaboration folders
- Cloud backup archives and sync directories

Why It Matters

Unstructured data is not only voluminous—it is high value. It often contains the very information that regulators, attackers, and business stakeholders care most about:

- **PII (Personally Identifiable Information)** — employee records, customer profiles, government IDs
- **PHI (Protected Health Information)** — lab results, treatment plans, insurance records
- **Financial data** — tax forms, invoices, contracts, and M&A documents
- **Intellectual property** — product specs, code, engineering diagrams, roadmaps
- **Synthetic content** — AI-generated summaries, creative assets, or draft legal language that may carry reputational risk

This data is not only sensitive—it is critical to operations and compliance. Yet it is often invisible to traditional DLP, SIEM, or compliance systems that were built to focus on structured formats.

A Perfect Storm

The convergence of several forces is compounding this challenge:

- **Generative AI and automation:** With 90% of future online content predicted to be AI-generated, the rate of unstructured data creation is compounding—and often lacks consistent formatting or metadata to support automated governance.
- **Hybrid work and collaboration tools:** Employees now generate and share vast quantities of unstructured data across multiple platforms, often without IT oversight—creating shadow repositories beyond the reach of traditional controls.
- **Cloud-first infrastructure:** As organizations migrate to the cloud, traditional perimeter-based models fall short. Agent-only approaches struggle to scale, and visibility into access events across SaaS apps and cloud storage is inconsistent at best.

The Consequence

Security teams are flying blind when it comes to unstructured data. Without centralized visibility or file-level controls, organizations face:

- **Audit failures and compliance penalties**
- **Insider threat exposure**
- **Increased ransomware susceptibility**
- **Uncontrolled sprawl and storage costs**
- **Data leaks that remain undetected for weeks or months**

This is not just a security gap—it is a strategic blind spot.

To move forward, organizations must treat unstructured data as a first-class security asset, on par with databases. This requires tools and strategies built specifically for the scale, complexity, and sensitivity of modern data ecosystems.

Primary Security Risks

As unstructured data continues to grow in both volume and value, it exposes organizations to a broad and intensifying spectrum of security threats. Unlike structured data—centralized in databases with clearly defined access models—unstructured data is dynamic, distributed, and frequently unmanaged. Without file-level visibility and control, security leaders face a minefield of risk.

Data Sprawl & Shadow IT

In today's hybrid and cloud-first environments, employees routinely use a wide array of tools to collaborate—often beyond the boundaries of sanctioned IT systems. Documents are emailed to personal accounts, shared over Slack, stored in unauthorized Dropbox folders, or duplicated across project management tools like Jira or Notion. This uncontrolled expansion, or data sprawl, creates blind spots for security and compliance teams.

Worse, this data often contains sensitive business content—financial forecasts, engineering plans, customer PII—that now lives in environments with unknown security postures. Governance policies cannot be enforced if IT doesn't know the data exists or where it lives. Shadow IT effectively neutralizes perimeter defences, enabling data leakage through unmanaged and unmonitored channels.

Excessive & Persistent Access

Unstructured data rarely benefits from the strict access control models applied to relational databases or SaaS applications. Once access is granted to a shared folder or file path, it is seldom reviewed or revoked, even after the project ends or employees change roles.

This results in excessive privilege accumulation, where users—both internal and external—retain access to sensitive data long after they need it. In the event of credential compromise, phishing, or lateral movement, these forgotten permissions significantly increase the blast radius of any security incident. A single compromised identity could expose years of sensitive documents

Without robust file activity monitoring and access analytics, these privileges remain unchecked and unchallenged.

Ransomware & Data Exfiltration

Modern ransomware campaigns increasingly target file shares, NAS drives, and cloud storage environments—not databases. These systems are easier to access, harder to monitor, and far more likely to contain valuable unstructured content.

Attackers exploit unmonitored directories to encrypt files, steal sensitive data, and demand payment under the threat of public exposure.

In many cases, detection only occurs after damage is done, as traditional tools fail to alert on file activity patterns that precede encryption.

Organizations without file-level audit logs and behavioral baselines often struggle to investigate what was accessed, altered, or exfiltrated—hindering both incident response and regulatory reporting.

Insider Threats

Not all threats come from the outside. Contractors, temporary staff, or even long-term employees can pose a significant risk—intentionally or unintentionally. An employee planning to leave may copy gigabytes of proprietary data to a USB drive, cloud account, or personal email in minutes.

Without real-time monitoring of file access, these activities may go completely unnoticed until after the data has left the building. Even well-intentioned insiders may mishandle sensitive information, mishandling confidential files or uploading them to poorly secured platforms for convenience.

Insider threats are particularly dangerous because they often exploit legitimate access and occur within the bounds of normal workflows, making them harder to detect without fine-grained visibility and behavioral context.

Regulatory Non-Compliance

Regulations such as GDPR, HIPAA, PCI DSS, and emerging AI safety laws demand that organizations not only protect sensitive data but demonstrate how they do so. This requires detailed audit trails, role-based access control, data classification, and proof of proactive governance.

Unfortunately, unstructured data is where most compliance programs fall short. Without knowing what sensitive data exists—or where, or who can access it—organizations cannot effectively enforce or demonstrate compliance.

In the event of a breach or audit, the inability to provide comprehensive file access logs, data classification records, or retention policies can lead to regulatory penalties, reputational damage, and litigation. The reality is simple: if you can't see it, you can't secure it—or prove it was ever secured at all.

These risks are not theoretical—they're systemic. From visibility gaps and privilege sprawl to delayed threat detection and compliance exposure, unstructured data security is quickly becoming a critical domain of cyber risk management. Organizations must evolve from coarse-grained monitoring and perimeter thinking to a model that offers continuous, contextual, and content-aware controls over every file, everywhere it lives.

Impact

The risks associated with unstructured data are not abstract technical challenges—they translate directly into tangible consequences. From regulatory exposure and financial penalties to operational inefficiencies and emerging AI-related risks, the lack of governance over unstructured content can significantly erode enterprise value and resilience.

Below, we outline four critical impact areas where unstructured data security—or the absence of it—has measurable effects on enterprise operations and outcomes.

Regulatory Fines & Legal Exposure

Data privacy regulations worldwide—such as the General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA), California Consumer Privacy Act (CCPA), and Payment Card Industry Data Security Standard (PCI DSS)—require organizations to implement and prove effective controls over personal and sensitive data.

A failure to detect, report, or contain data breaches involving unstructured content (e.g., emails, documents, and chat logs) can result in severe consequences. Under GDPR Articles 33 and 34, for example, organizations must notify regulators within 72 hours of discovering a breach and demonstrate that appropriate safeguards were in place.

Without centralized visibility or complete audit trails, many organizations are unable to accurately assess what was accessed or exfiltrated. This lack of forensic evidence increases liability and can result in:

- Multi-million dollar fines
- Class-action lawsuits
- Loss of contracts or certifications
- Public brand damage

In some sectors (e.g., finance, healthcare, defense), failure to demonstrate data governance over unstructured assets may even result in regulatory shutdowns or restrictions.

Operational Disruption

The average downtime following a ransomware attack now exceeds 19 days, according to multiple industry studies. File servers and shared drives are often the first systems targeted because they are inadequately monitored, broadly accessible, and contain business-critical unstructured data.

Key operational risks include:

- **Loss of access to sensitive or regulatory-critical documents**
- **Inability to restore data without triggering backup inflation or re-infection**
- **Delayed business continuity due to unclear ownership or classification of data**

Without file-level access intelligence, organizations struggle to recover safely or prioritize what data to restore first. Additionally, when permissions models are outdated or overly permissive, malware can spread rapidly across file systems—amplifying the impact of a single breach.

AI Model Contamination

As enterprises invest in generative AI and large language models (LLMs) for internal applications—like chatbots, code generation, legal review, or content creation—they are increasingly sourcing training and prompt data from their own file repositories.

Without proper governance, these datasets may include:

- **Overshared personal data (PII/PHI)**
- **Confidential business plans or contracts**
- **Synthetic content loops generated by AI tools**

The result is model contamination—where AI systems learn from ungoverned, duplicated, or privacy-sensitive data. This can degrade model performance, introduce bias, and create legal liabilities if PII or regulated information is unintentionally embedded in outputs.

Moreover, non-classified or poorly managed training data risks becoming unreproducible or unverifiable—undermining trust in AI-driven business decisions. In regulated environments, this lack of data lineage could hinder auditability or compliance with future AI governance laws.

These risks represent more than security concerns—they are strategic inhibitors to digital transformation, innovation, and resilience. Inadequate governance of unstructured data:

- Raises compliance costs
- Reduces operational efficiency
- Exposes enterprises to disruptive events
- Limits the effectiveness of next-gen technologies like AI

As such, unstructured data security must be elevated from an IT issue to a board-level priority. The only way forward is with centralized visibility, intelligent classification, real-time monitoring, and a unified control framework—capabilities that Thales File Activity Monitoring delivers.

Technical Challenges

The fundamental difficulty of securing unstructured data lies not just in its volume or dispersion—but in its inherent complexity and variability. Unlike structured data, which conforms to schema and lives within clearly defined systems, unstructured data is chaotic, context-sensitive, and embedded in dynamic ecosystems. Protecting it at scale presents significant architectural and operational hurdles. Below, we explore four of the most pressing technical challenges.

Discovery at Scale

The challenge:
Accurately locating and identifying sensitive data across petabytes of files, messages, and cloud content—many of which lack meaningful metadata.

Why it's hard:
In traditional systems, discovery relies on metadata such as filenames, file extensions, tags, or timestamps. But in unstructured environments, metadata is unreliable or entirely absent. A file named report.pdf might contain financial projections, patient data, or source code—or nothing sensitive at all.

To discover sensitive information, systems must perform deep content inspection of billions of individual objects, often stored in fragmented repositories across cloud, hybrid, and on-premise infrastructure. This requires scalable indexing, pattern recognition, and integration with diverse file systems, object stores, and collaboration platforms—all while respecting performance constraints and access controls.

Moreover, unstructured data is not static. New files are created and shared every second. Discovery must be continuous, not one-time, to remain accurate and useful.

Real-Time Monitoring

The challenge:
Tracking file activity—who accessed what, when, how, and from where—in real time across heterogeneous systems.

Why it's hard:
File systems and SaaS applications generate massive volumes of telemetry. Protocols like SMB and NFS, or APIs from SharePoint, OneDrive, and Google Drive, produce high-frequency events for even routine activities (e.g., file opens, renames, syncs, deletions).

Forwarding this raw event stream into a SIEM or centralized logging system is prohibitively expensive, especially given how SIEM platforms often charge based on ingestion volume. Additionally, many legacy tools cannot capture the necessary granularity or context of file events—such as the sensitivity of the data accessed or whether the action violated policy.

To be effective, real-time monitoring must include:

- **Intelligent event filtering** (to reduce noise and focus on high-risk activity)
- **Correlation logic** (to distinguish between routine automation and anomalies)
- **Context enrichment** (to map actions to data classifications, user roles, and geographic origins, and encryption status)
- **Event aggregation** (to consolidate repetitive events, minimizing overhead while preserving audit fidelity)

Without these capabilities, teams either miss critical signals or drown in irrelevant alerts.

Contextual Classification

The challenge:
Accurately classifying unstructured data based on both its content and context—rather than just pattern-matching.

Why it's hard:
A string of digits matching a credit card format might appear in multiple files—but its risk level depends on context. In a test log, it may be benign. In a customer invoice, it may be highly regulated. In a design document, it may be a placeholder.

Most data loss prevention (DLP) and classification tools rely on regular expressions and rule-based detection. These methods fall short when dealing with natural language, embedded files, foreign language content, scanned documents, or AI-generated artifacts. As unstructured data becomes more diverse, context-aware classification becomes essential.

This requires classification engines to incorporate:

- **Semantic understanding of text** (via NLP or ML models)
- **File lineage and access history**
- **User and usage intent**
- **Metadata from enterprise systems** (e.g., HR, finance, CRM)

Only with this enriched context can organizations reduce false positives and focus protection where it matters most.

Unified Policy Enforcement

The challenge:
Implementing consistent, least-privilege policies across a fragmented, multi-platform environment.

Why it's hard:
Enterprises operate across a heterogeneous ecosystem: Windows file servers, Linux NAS appliances, Amazon S3 buckets, Azure Blob storage, SharePoint Online, Gmail, Slack, Git repositories, and more. Each system comes with its own access control model, audit logging format, and integration challenges.

Enforcing a unified policy—e.g., "Only Finance can access PCI data, and access must be logged and reviewed quarterly"—requires orchestration across all these platforms. But few tools are architected to span both on-premises infrastructure and cloud-native services under a single control plane.

To succeed, a unified policy engine must offer:

- **Cross-platform compatibility** (protocol-agnostic enforcement, normalization of audit data across all platforms)
- **Centralized management** (one place to define and deploy rules)
- **Granular targeting** (per user, file type, location, classification)
- **Audit and rollback capabilities**

Without this capability, security policies fragment across silos—leading to inconsistent enforcement, higher risk, and increased overhead.

Solving these challenges demands more than point solutions or patchwork scripts. It requires a platform-level approach built specifically for unstructured data at enterprise scale. Thales File Activity Monitoring (FAM) addresses these issues directly, providing intelligent discovery, real-time file monitoring, contextual classification, and unified policy management across the entire data estate.

Industry Best Practices

Across the cybersecurity landscape, industry research, compliance standards, and real-world security incidents point to a consensus: unstructured data must be managed and protected through a proactive, intelligence-driven strategy.

Regulatory guidance from GDPR, HIPAA, and NIST 800-53, as well as practitioner frameworks from the SANS Institute, ISACA, and Gartner, emphasize the importance of continuous data discovery, behavioral monitoring, least-privilege enforcement, and anomaly detection. These requirements converge on four foundational pillars—the building blocks of a modern unstructured data security strategy.

Automated Discovery & Classification

What it is:
A continuous, automated process that scans enterprise storage, endpoints, cloud services, and collaboration platforms to identify sensitive data—then assigns meaningful classifications such as "PII," "confidential," "regulated," or "public."

Why it matters:
You cannot protect what you can't find. Without accurate discovery, sensitive data remains invisible—unclassified, unmonitored, and vulnerable to exfiltration, misuse, or accidental exposure. Discovery must go beyond metadata, performing deep content inspection to identify data in various formats, languages, and containers (e.g., embedded in ZIP files, email attachments, or cloud-native documents).

Best practice features:

- Continuous background scanning across structured and unstructured repositories
- Real-time updates to classification based on data movement or sensitivity changes
- Support for regulatory and custom taxonomies (e.g., PCI, GDPR, intellectual property)

Business outcome:
Improves risk visibility, ensures regulatory alignment, and serves as the foundation for access control and policy enforcement.

File-Level Access Monitoring

What it is:
The ability to log, audit, and analyze all user interactions with unstructured data—across all file types, locations, and access methods—on a per-action basis (e.g., read, write, move, copy, delete, share).

Why it matters:
Traditional monitoring tools are often limited to directory-level events or generalized logs. They fail to provide granular insight into who accessed what file, when, from where, and under what conditions. This lack of precision leaves gaps in forensic investigations and compliance reporting.

Modern file-level monitoring enables:

- Detection of unusual behavior patterns (e.g., mass file downloads)
- Identification of privilege misuse or lateral movement
- Validation of access policies across managed and unmanaged data sources

Best practice features:

- Integration with SIEM, SOAR, and IAM platforms
- User and entity behavior analytics (UEBA)
- Alerts for policy violations, risky access, and unauthorized data sharing

Business outcome:
Accelerates threat detection, simplifies compliance audits, and reduces investigation time from days to minutes.

Data Minimization

What it is:
The practice of identifying and eliminating redundant, obsolete, or trivial (ROT) data from enterprise storage. This includes unused documents, outdated versions, temporary files, and abandoned data repositories.

Why it matters:
Excess data amplifies security risk and operational cost. ROT data increases the attack surface, consumes unnecessary storage and backup resources, and introduces legal exposure (e.g., retaining PII longer than permitted under regulations like GDPR's data retention limits).

Data minimization is a cornerstone of both cyber hygiene and privacy compliance, enabling organizations to focus protection efforts where they matter most.

Best practice features:

- Intelligent recommendations based on file age, usage frequency, access history, and sensitivity
- Automated archiving or deletion workflows
- Customizable data retention policies based on file type and classification

Business outcome:
Reduces attack surface, lowers infrastructure costs, improves backup efficiency, and supports regulatory defensibility.

AI-Assisted Analytics

What it is:
The use of artificial intelligence, including machine learning (ML) and large language models (LLMs), to interpret file activity data, identify anomalies, predict risk, and automatically generate compliance artifacts.

Why it matters:
The volume of file activity events generated in an enterprise environment is far too large for manual triage. AI accelerates insight by highlighting deviations from normal patterns, grouping related events, and recommending policy changes or remediations.

With the rise of generative AI, LLMs can now help security and compliance teams ask natural-language questions of audit data, generate custom reports, or even detect anomalies that don't match known signatures.

Best practice features:

- Pre-trained models for insider threat detection, data misuse, and risky behavior
- AI copilots for compliance reporting, incident summaries, and root cause analysis
- Adaptive learning to refine detection over time based on organization-specific patterns

Business outcome:
Improves detection speed and accuracy, reduces analyst workload, and accelerates compliance workflows.

These four pillars—automated discovery, granular access monitoring, data minimization, and AI-assisted analytics—are no longer optional. They represent the core functional requirements for any enterprise seeking to control, secure, and extract value from unstructured data.

Thales File Activity Monitoring is built around these principles, combining enterprise-scale visibility with advanced classification, intelligent threat detection, and policy-driven automation—enabling organizations to move from reactive defence to proactive data stewardship.

Thales File Activity Monitoring: Closing the Gap

Enterprises face increasing pressure to secure unstructured data across complex, hybrid environments while maintaining compliance and operational agility. Thales File Activity Monitoring (FAM) was purpose-built to meet this challenge, delivering end-to-end visibility, control, and intelligence across the unstructured data landscape. It integrates seamlessly with Thales's broader **Data Security Fabric,** creating a unified platform for structured and unstructured data governance.

Below, we break down the core capabilities of FAM, how each addresses critical risks, and what makes Thales uniquely capable of delivering at enterprise scale.

Discovery & Classification

FAM employs both agent-based and agentless sensors to scan across endpoints, NAS devices, and SaaS platforms like Microsoft 365 and Google Workspace. It performs deep content inspection to locate sensitive unstructured data—such as personally identifiable information (PII), protected health information (PHI), payment card data, or proprietary IP. Each file is mapped to relevant regulatory categories (e.g., GDPR, HIPAA, PCI DSS) and assigned a classification tag for policy enforcement and reporting.

Differentiators:

- **High-speed content fingerprinting** enables fast and scalable classification even across billions of files.
- **Continuous rescanning** ensures new data is automatically discovered and classified in real time, reducing stale or missed assets.
- **Multi-format support** includes PDFs, Office files, archives, emails, and nested attachments—ensuring complete coverage.

Why it matters:
This level of automated discovery and classification closes the visibility gap and forms the foundation for policy creation, threat detection, and compliance reporting.

Real-Time Access Monitoring

FAM logs every interaction with unstructured data—including read, write, copy, delete, and share actions—capturing not only the event but also user identity, device, location, timestamp, file sensitivity, and data source context. This granular telemetry is streamed into dashboards and analytics systems to detect policy violations, insider threats, or anomalous behavior.

Differentiators:

- Scales to billions of events per day through adaptive filtering, reducing noise while retaining forensic fidelity.
- Cross-platform monitoring includes Windows shares, Linux NAS, SharePoint Online, OneDrive, Google Drive, Gmail, and more.
- Event correlation provides contextual understanding of user behavior across systems.

Why it matters:
With real-time file activity monitoring, security teams can proactively detect suspicious actions, enforce access controls, and generate detailed audit trails to meet compliance mandates.

AI-Powered Security Assistant

Built on Large Language Models (LLMs), FAM's GenAI-powered assistant simplifies investigation, compliance, and security workflows by automatically analyzing audit logs, identifying risks, and generating human-readable reports. Users can interact with the assistant using natural language queries to quickly surface relevant insights or generate documentation.

Differentiators:

- Accepts plain English queries like: *"Show failed access attempts to PII in APAC last 90 days."*
- Auto-generates auditor-ready reports, including explanations of policy coverage and evidence of control enforcement.
- Learns from user interaction patterns to optimize recommendations and reporting output.

Why it matters:
AI drastically reduces the manual effort required to investigate incidents or prove compliance—freeing teams to focus on strategic risk management rather than rote documentation.

Compliance Dashboards & Reports

Thales FAM offers prebuilt dashboards and reporting templates mapped directly to regulatory requirements, such as:

- **GDPR Article 30** (Records of processing activities)
- **HIPAA §164.312(b)** (Audit controls for access to ePHI)
- **PCI DSS 4.0 Requirement 10** (Logging and monitoring)

Security and compliance teams can generate comprehensive evidence packs—summarizing who accessed what data, when, and under what policy—within minutes.

Differentiators:

- **One-click** export to PDF, CSV, or JSON for regulators, auditors, or internal stakeholders.
- **Immutable log retention** to ensure the integrity of audit data over multiple years.
- **Customizable views** by business unit, geography, data type, or policy domain.

Why it matters:
When audit requests arrive, organizations must respond with speed, accuracy, and confidence. FAM eliminates the scramble for evidence by making compliance reporting a continuous, automated process.

Unified Data Security Fabric

Unlike point solutions that only address unstructured content, FAM is embedded in Thales's Data Security Fabric, which spans structured and unstructured data environments—including database activity monitoring (DAM), encryption, key management, and access governance.

This integration allows enterprises to define and enforce consistent data protection policies across:

- Oracle, SQL Server, and SAP HANA databases
- Windows, Linux, and Mac file systems
- SaaS and cloud-native storage platforms

Differentiators:

- **Single policy engine** for enterprise-wide enforcement
- **Unified interface** for data classification, access control, and audit review
- **End-to-end coverage** from structured tables to SharePoint folders—no data type left behind

Why it matters:
As enterprises converge their data protection, compliance, and risk strategies, only Thales provides a holistic view and control plane across all data modalities—breaking down silos and eliminating blind spots.

Thales File Activity Monitoring doesn't just fill a gap—it redefines how organizations secure unstructured data in a cloud-first, AI-enabled world. With powerful discovery, real-time monitoring, automated compliance reporting, and AI-driven analytics, FAM transforms unstructured data from a security liability into a strategic advantage.

Capability	How FAM Addresses the Risk	Differentiators
Discovery & Classification	Agent + agentless sensors crawl endpoints, NAS, and SaaS repositories; map files to GDPR, HIPAA, PCI DSS categories.	High speed content fingerprinting; continuous rescans.
Real Time Access Monitoring	Captures every file event (read, modify, copy, delete) with user, device, geo, and sensitivity context.	Scales to billions of events with adaptive filtering.
AI Powered Security Assistant	Large language model assistant summarizes audit trails, suggests policies, and drafts auditor ready reports.	Natural language queries ("Show failed accesses to PII in APAC last 90 days").
Compliance Dashboards & Reports	Pre built templates for GDPR Article 30, HIPAA §164.312(b), PCI DSS 4.0 Req.10.	One click export; immutable log retention.
Unified Data Security Fabric	Integrates with structured data monitors (Database Activity Monitoring) for a single policy engine.	End to end coverage—from Oracle tables to OneDrive folders—under one UI.

Strategic Benefits

The success of a data protection strategy is measured not only by the ability to reduce risk but also by how well it aligns with broader business goals: agility, cost efficiency, regulatory readiness, and innovation. Thales File Activity Monitoring (FAM) equips technical executives with the visibility, intelligence, and control needed to lead secure digital transformation in an era defined by data sprawl and AI disruption.

Risk Reduction

Thales FAM dramatically reduces the likelihood, impact, and scope of data breaches by exposing and remediating "dark" unstructured data—files and assets that live outside formal governance frameworks.

By continuously discovering, classifying, and monitoring access to sensitive content, FAM enables:

- **Enforcement of least privilege** down to the file level
- **Detection of privilege misuse** and access anomalies
- **Real-time alerts on unauthorized or risky actions**

Security leaders gain a defensible position in breach scenarios, as file-level audit trails and access controls allow for precise containment and forensic reconstruction. This limits regulatory exposure and reduces the "blast radius" of both external attacks and insider threats.

Audit Agility

Preparing for regulatory audits or breach investigations often requires security teams to assemble historical access logs, generate compliance evidence, and respond to detailed auditor queries—processes that can take weeks of manual effort.

FAM eliminates this burden with:

- **Automated classification and logging**
- **Prebuilt compliance reports** mapped to GDPR, HIPAA, PCI DSS, and others
- **Searchable dashboards** that enable real-time query responses

Auditors or legal teams can ask questions like:
"Who accessed PII in the last 30 days in the EMEA region?"—and receive a precise, timestamped report in minutes.

Security leaders can move from reactive documentation to proactive compliance assurance. FAM transforms audit preparation from a high-cost fire drill into a continuous, automated process—freeing resources, reducing consultant dependence, and building institutional trust.

Cost Optimization

Unstructured data is often treated as a "dumping ground," with stale, redundant, or obsolete files consuming high-performance storage and being backed up indefinitely—leading to inflated infrastructure and licensing costs.

FAM enables cost savings through:

- **Data minimization** (identifying ROT data for archiving or deletion)
- **Cold data analytics** to support storage tiering decisions
- **Smarter backup scoping** based on file relevance and sensitivity

CIOs and infrastructure teams can reclaim storage capacity, reduce cloud egress fees, and optimize backup strategies—all while shrinking the attack surface and improving operational efficiency. These efficiencies compound at scale, delivering millions in long-term OPEX savings.

AI Readiness

With the rise of generative AI and large language models (LLMs), enterprises are racing to tap internal data for use in copilots, virtual assistants, and custom AI applications. But AI models are only as trustworthy as the data they are trained on.

FAM ensures that only governed, classified, high-quality data is surfaced for AI pipelines by:

- **Tagging and scoring unstructured data** by sensitivity and compliance relevance
- **Flagging and excluding high-risk content** (e.g., PII, legal contracts, regulated data) from training corpora
- **Providing visibility into data lineage** and model input provenance

Security and data teams can accelerate AI adoption without compromising compliance, IP protection, or data ethics. FAM becomes a gatekeeper between enterprise data and AI, ensuring that models are powered by clean, secure, and context-rich inputs—not contaminated or exposed content.

Be Ready Now

For today's technical executives, unstructured data security is no longer a niche concern—it's a critical enabler of operational excellence, regulatory posture, and innovation. Thales File Activity Monitoring empowers security leaders to operationalize governance, reduce complexity, and drive strategic outcomes across IT and business.

Conclusion: Regaining Control in a World of Unstructured Risk

The exponential growth of unstructured data represents one of the most profound—and most under-addressed—shifts in enterprise IT. As AI-generated content surges, collaboration platforms proliferate, and hybrid work becomes permanent, the traditional boundaries of data governance have eroded. Sensitive information now lives in dynamic, distributed, and often invisible formats—far beyond the reach of legacy security controls.

Yet this data is not peripheral. It includes intellectual property, regulated records, personal identities, and strategic communications. Left unmonitored, unclassified, and unmanaged, unstructured data becomes not just a blind spot, but a systemic business risk.

Across industries, research and frontline experience converge on the same core challenges: visibility, context, control, and scale. Security leaders must discover where sensitive data lives, understand who is accessing it and why, and enforce policies that evolve with business and regulatory demands.

Thales File Activity Monitoring (FAM) was built for this moment. It closes the visibility gap by combining real-time file monitoring, automated classification, and AI-powered insights into a unified solution. FAM doesn't just log events—it helps you understand them. It doesn't just enforce policy—it helps you shape it. And it doesn't just check boxes for compliance—it transforms unstructured data into a secure, strategic asset.

For CISOs, CIOs, and data protection leaders, this represents a new level of control: over risk, over operational cost, and over the future of enterprise AI. FAM enables your organization to move beyond reactive security toward intelligent governance—anchored in clarity, compliance, and confidence.

Unstructured data is not going away. But with Thales, neither is your control.

About Thales Trusted Cyber Technologies

About Thales Trusted Cyber Technologies
Thales Trusted Cyber Technologies, a business area of Thales Defense & Security, Inc., is a trusted, U.S. provider of cybersecurity solutions dedicated to U.S. Government. We protect the government's most vital data from the core to the cloud to the edge with a unified approach to data protection. Our solutions reduce the risks associated with the most critical attack vectors and address the government's most stringent encryption, key management, and access control requirements.

For more information, visit www.thalestct.com

THALES

Building a future we can all trust

Contact us

For all office locations and contact information,
please visit thalestct.com/contact-us

thalestct.com