cyberhaven

Next-Gen Data Security Posture Management:

Built for the AI-Driven Data World

Table of Contents

Introduction: Why DSPM Exists

The Data Perimeter Has Collapsed

The enterprise data perimeter has irreversibly dissolved. Data no longer resides solely in centrally governed systems. Instead, it is fluid, replicated, and reshaped across an ever-expanding ecosystem of SaaS apps, cloud platforms, data warehouses, and employee endpoints. With the rise of generative AI, this fragmentation has accelerated, not only because models ingest sensitive data, but because humans use these models to create new derivatives that are then shared, copied, and stored across systems. Together, humans and AI propagate sensitive information into locations security teams may not even know exist.

Modern organizations face a new reality where every employee, system, and AI model can become a new data producer. This means that the historical assumption that data remains inside controlled repositories is inaccurate. To fix this, the DSPM category emerged to solve precisely this problem.

Defining DSPM

Data Security Posture Management (DSPM) is a technology that discovers, classifies, and protects sensitive data across an organization's ecosystem. DSPM marked the evolution from asset-centric security to **data-first security,** shifting the focus from protecting systems to protecting the data itself. It brought speed, scalability, and precision to understanding data. But until now, DSPMs have delivered only surface-level data understanding, remaining disconnected from real-time protection, and largely confined to scanning cloud environments.

First-Generation DSPM

Data Visibility

Traditional DSPMs start with automated discovery across cloud environments, particularly within leading infrastructure-as-a-service (IaaS) platforms such as Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure. They connect to cloud-native storage services and database workloads to inventory where data resides and how it is configured.

Their visibility model relies on scanning cloud resources at rest, scheduled every 90 days, for instance, to build a picture of the organization's data landscape. Support typically extends across structured and unstructured formats found in cloud stores. Because their design centers on cloud infrastructure APIs, visibility is typically strongest within IaaS-native services rather than across SaaS or on-premise environments.

Data Classification

Traditional DSPMs use pattern-based and rules-driven classification techniques, such as regex matching, keyword rules, and predefined classifiers, to identify sensitive information, similar to previous-generation data loss prevention (DLP) tools. While some vendors incorporate lightweight machine learning, the overall approach is metadata-centric and relies on scanning stored data objects rather than analyzing context or data behavior.

These classification methods can successfully detect well-defined categories such as PII, PHI, PCI data, secrets, and financial records. However, because the classification engines operate within cloud repositories and rely heavily on static rule-based approaches, they often lack deeper semantic understanding, contextual interpretation, or the ability to distinguish sensitivity based on business use cases.

Posture Assessment

Traditional DSPM platforms provide posture assessment to help organizations understand the overall risk of their cloud data environment. They evaluate misconfigurations, excessive access, unintended exposure, compliance gaps, and retention issues to highlight where sensitive data may be at risk. These assessments rely on cloud metadata, static scans, and predefined rules to surface the most common patterns of data insecurity.

To make these findings actionable, traditional DSPMs offer out-of-the-box risk dashboards. These views give security teams a structured way to interpret posture results and prioritize remediation.

Remediation

Traditional DSPM tools integrate with cloud platforms, Identity and Access Management (IAM) systems, and Information Technology Service Management (ITSM) workflows to help teams remediate identified risks. Their policy enforcement strengths include:

- Identifying misconfigured cloud storage
- Highlighting publicly exposed datasets
- Recommending least-privilege access changes
- Triggering alerts or automated remediation scripts

Remediation typically involves adjusting IAM roles, updating storage configuration settings, or applying predefined security baselines. While effective at addressing cloud configuration risks, these remediations focus on infrastructure posture rather than preventing real-time data misuse or exfiltration.

Compliance

A core function of traditional DSPM is helping organizations satisfy compliance and audit requirements. By cataloging sensitive data in cloud environments and identifying associated risks, these tools generate evidence and report on regulated data such as PII, PCI, and PHI.

This compliance-centric lens allows security and privacy teams to transform ad hoc cloud assessments into repeatable processes, though the scope remains limited to cloud-based assets and configurations.

Why First-Generation DSPM Isn't Enough Anymore

While early DSPM solutions brought much-needed visibility to cloud data stores, they focused narrowly on data at rest and often provided dashboards without meaningful actionability. Several forces now limit their effectiveness:

- **Limited coverage** focuses on IaaS with incomplete coverage of SaaS, on-prem, and endpoint devices.
- **AI-generated data** creates new derivatives that legacy DSPMs cannot track.
- **Fragmented data proliferation** outpaces scheduled scans, leaving blind spots.
- **Alert fatigue** from noisy or duplicate findings overwhelms teams.
- **Inflexible tooling** limits the ability to configure scans or customize labels for unique business needs.
- **Lack of real-time protection** prevents data from being exfiltrated or misused.

> **Fragmented data proliferation** is the rapid, uncontrolled spread of data across an organization's data ecosystem, through copying, sharing, and transformation, at a pace that outpaces traditional discovery methods. It includes shadow copies, stale backups, and AI-generated derivatives.

Security teams now require a platform that not only discovers sensitive data but also understands it, contextualizes it, and protects it in real time.

Core Requirements of a Next-Gen DSPM Platform

Next-generation DSPM must expand beyond visibility into continuous, contextual, real-time protection across all data states and environments.

AI-Driven Data Understanding

Next-gen DSPM leverages AI to deliver high-fidelity classification with drastically fewer false positives. Cyberhaven, for example, connects billions of datapoints, assigning dimensions such as category, location, lineage, and identity to produce multifaceted context. True context means understanding and connecting those dots:

- **Provenance** identifies the ownership or origin of data, such as whether it was created by the company or a public source. For example: an internal design documents or text copied from a public website.
- **Location exposure** shows who can access the data, such as only internal employees, external collaborators, or anyone on the internet. A file restricted to engineering, shared with a vendor, or accidentally made publicly accessible would each represent different exposure levels.
- **Storage medium** identifies where the data lives, whether on an endpoint, in cloud storage, within a SaaS app, or on a network drive. A file on a laptop, an S3 bucket, or a shared Google Drive folder all represent distinct contexts.
- **Data structure** describes the format of the data, such as a document, spreadsheet, graphic, or raw text, which affects how it can be protected.
- **Management status** indicates whether a system or device holding the data is managed or unmanaged by the organization. For example: a corporate-issued laptop versus a personal device or an unmonitored cloud resource.

Crucially, next-gen DSPM correlates access: **who can access data, who has accessed it, and who should access it.** It maps connections across identity, sensitivity, and exposure to create a unified understanding of risk.

Data Visibility Across Every Environment

Next-gen DSPM must inventory data across cloud providers (AWS, GCP, Azure), SaaS, and on-premise repositories. But visibility cannot stop in these environments.

A majority of sensitive data originates on employee devices, where documents are created, copied, and shared. Traditional DSPM providers do not scan endpoints at all, leaving massive blind spots. Next-gen DSPM must provide full coverage across:

- **Cloud** datastores including IaaS, SaaS, and PaaS
- **On-premises** databases and file shares
- **Endpoints** including the hundreds of thousands of devices used by employees

This also includes continuous discovery of shadow data such as stale SaaS exports, abandoned backups, and forgotten cloud buckets.

Data Visibility Across All Data States

Visibility cannot be limited to data at rest. Next-gen DSPM must track:

- **Data at rest:** where it is stored
- **Data in motion:** where it is sent
- **Data in use:** how users and systems interact with it

Traditional DSPM discovers only static data. But organizations need real-time awareness as data moves, gets duplicated, or is transformed. A dynamic system must trigger new scans when data is created or modified and maintain an up-to-date picture of the organization's data landscape.

Continuous Data Risk Monitoring & Prioritization

Rather than relying on monthly or quarterly scans, next-gen DSPM continuously evaluates exposure, monitoring for:

- Public accessibility
- Over entitlements
- Suspicious data access or movement
- Cross-border transfer risks
- Data propagated into unsanctioned repositories

Event-driven scanning ensures that risk assessment aligns with the speed of modern data creation and movement.

Data Lineage for Real Data Movement Tracking

Cyberhaven's data lineage, which tracks origin, transformation, and destination across different data environments, surpasses rudimentary lineage models that operate within a single domain. In reality, a piece of sensitive data may:

- Appear on an endpoint
- Be uploaded to a SaaS tool
- Be exported into a cloud bucket
- Be copied into a spreadsheet
- Be fed into an AI model

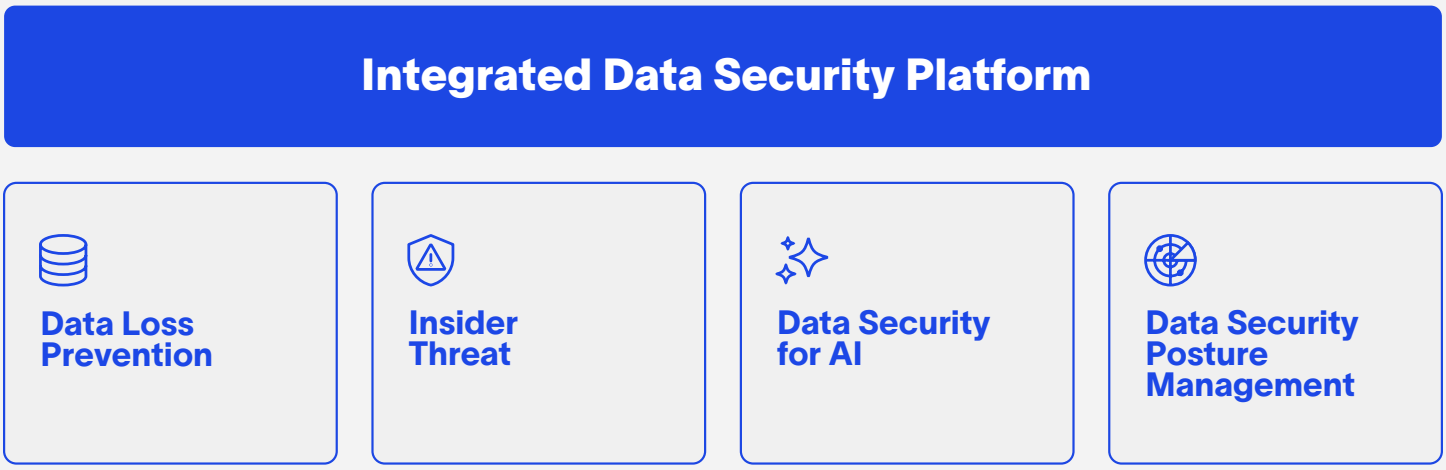
A next-gen DSPM must understand these interconnected movements to identify risks that static tools miss.

Unified Platform for End-to-End Data Security

Historically:

- **DLP** protected data in motion but lacked visibility of static data residing within repositories.
- **DSPM** discovered data at rest but couldn't intervene in real time.

A unified platform merges these strengths, offering both proactive risk reduction and real-time enforcement. This eliminates silos, reduces tooling complexity, and aligns security operations around a single, consistent understanding of data.



Implementation Roadmap: How Organizations Should Adopt DSPM

Cyberhaven supports both DSPM and DLP use cases, providing an integrated approach to data protection. Organizations can begin at any point in the lifecycle, depending on maturity and goals. Data security is not a one-time project, it is an ongoing journey of continuous monitoring, adjustment, and strengthening of the organization's security posture.

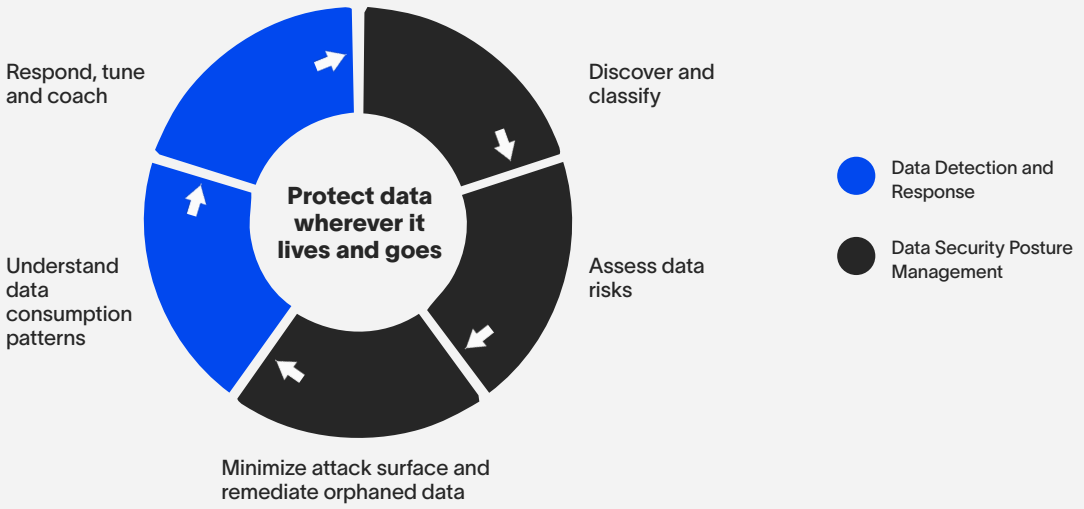
Step 1: Discover and Classify Data
Start with a comprehensive inventory of sensitive data across cloud, on-prem, and endpoints.

Step 2: Assess Data Risk
Determine whether sensitive data is publicly exposed, overly accessible, or stored improperly.

Step 3: Minimize Attack Surface
Identify and remediate shadow data, orphaned datasets, and misconfigurations.

Step 4: Understand Data Consumption Patterns
Observe how employees, systems, and AI tools use data to uncover risky workflows.

Step 5: Stop Data Breaches and Coach Users
Apply real-time controls and user education to prevent harmful data handling.



Conclusion

Data-centric security is now mission-critical. With the collapse of traditional perimeters, the rise of AI, and the exponential growth of shadow data, organizations require continuous, contextual, real-time understanding of how sensitive data is created, accessed, and used. DSPM provides this foundation, but only a next-generation, AI-powered DSPM solution can meet the demands of modern, multi-cloud, AI-driven businesses.

A data-first security model is no longer optional, ***it is the only viable path forward for the next decade.***

cyberhaven

About Cyberhaven

Cyberhaven is reimagining data security. Until now, data security products have been limited to scanning data content or looking for specific user actions. Our AI-enabled data lineage technology analyzes billions of workflows to understand every piece of data within an organization, identify when it's at risk, and take action to protect it.

To learn more, visit **cyberhaven.com**